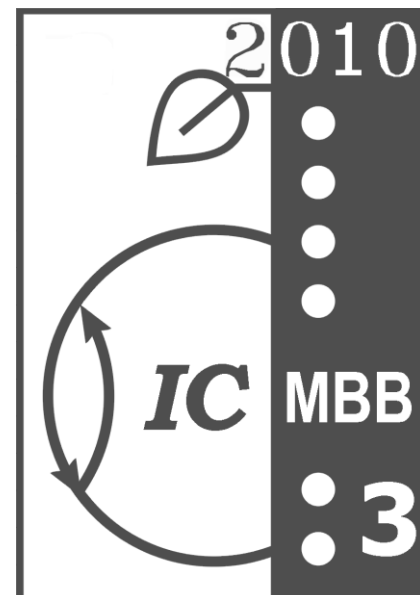
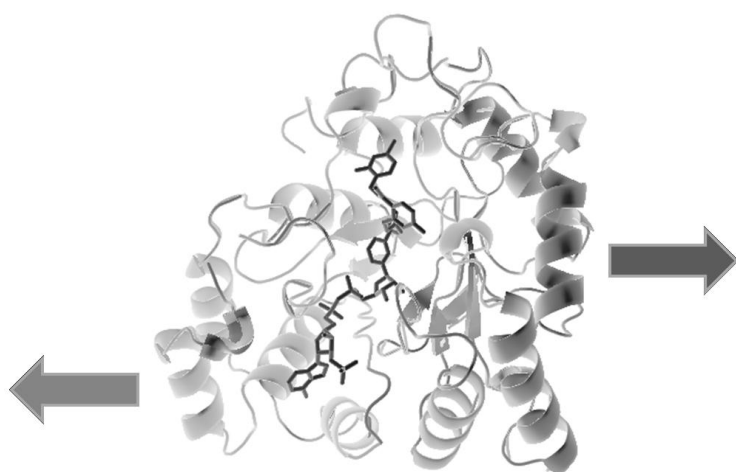




*Российская
Академия
Наук*

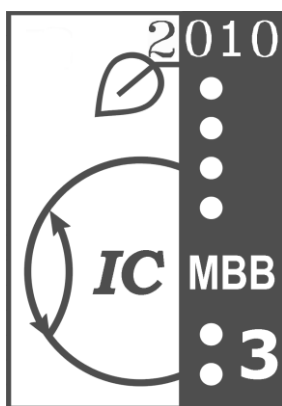
Математическая биология и биоинформатика



**РОССИЙСКАЯ АКАДЕМИЯ НАУК
ОТДЕЛЕНИЕ МАТЕМАТИЧЕСКИХ НАУК**

**УЧРЕЖДЕНИЕ РОССИЙСКОЙ АКАДЕМИИ НАУК
ИНСТИТУТ МАТЕМАТИЧЕСКИХ ПРОБЛЕМ БИОЛОГИИ РАН**

при поддержке
РОССИЙСКОГО ФОНДА ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ



МАТЕМАТИЧЕСКАЯ БИОЛОГИЯ И БИОИНФОРМАТИКА

ДОКЛАДЫ III МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ

Пушино, 10-15 октября 2010 г.

Под редакцией д.ф.-м.н. В.Д.Лахно

Пушино 2010

УДК 532
ББК 28в6
М34

**М34 Математическая биология и биоинформатика: III Международная конф., г. Пущино, 10-15 октября 2010 г.: Доклады/ Под ред. В.Д.Лахно. – 286 с.
ISBN**

В сборнике представлены доклады III-й Международной конференции «Математическая биология и биоинформатика», проводимой Научным советом по математической биологии и биоинформатике РАН и Учреждением Российской академии наук Институтом математических проблем биологии РАН. Конференция проводится при финансовой поддержке РФФИ (грант 10-07-06086-г) и Отделения математических наук РАН.

Оргкомитет

Лахно В.Д., д.ф.-м.н., – председатель
Устинин М.Н., д.ф.-м.н. – зам.председателя
Махортых С.А., к.ф.-м.н. – ученый секретарь
Александров А.А., д.б.н.
Арсеньев А.С., д.х.н.
Васин А.А., д.ф.-м.н.
Ильин В.А., д.ф.-м.н.
Козлов Н.Н., д.ф.-м.н.
Коротков А.В., д.э.н.
Коротков Е.В., д.б.н.
Марков А.В., д.б.н.
Миронов А.А., д.ф.-м.н.
Молчанов А.М., д.ф.-м.н.
Назипова Н.Н., к.ф.-м.н.
Ризниченко Г.Ю., д.ф.-м.н.
Романюха А.А., д.ф.-м.н.
Самсонова М.Г., к.б.н.
Туманян В.Г., д.ф.-м.н.
Фрисман Е.Я., д.б.н.
Шабанов Б.М., к.т.н.
Шайтан К.В., д.ф.-м.н.

Программный комитет

Журавлёв Ю.И., д.ф.-м.н., академик РАН –
председатель
Колчанов Н.А., д.б.н., академик РАН –
зам.председателя
Рудаков К.В., д.ф.-м.н., член-корр. РАН –
зам.председателя
Арчаков А.И., д.б.н., академик РАН
Бердышев В.И., д.ф.-м.н., член-корр. РАН
Дегерменджи А.Г., д.ф.-м.н., член-корр. РАН
Евтушенко Ю.Г., д.ф.-м.н., академик РАН
Жиженко А.Б., д.ф.-м.н., член-корр. РАН
Левин В.К., д.ф.-м.н., академик РАН
Матвеев С.В., д.ф.-м.н., член-корр. РАН
Рубин А.Б., д.ф.-м.н., член-корр. РАН
Федотов А.М., д.ф.-м.н., член-корр. РАН
Четверушкин Б.Н., д.ф.-м.н., член-корр. РАН
Шокин Ю.И., д.ф.-м.н., академик РАН

Редакционная группа:

Махортых С.А., к.ф.-м.н.
Назипова Н.Н., к.ф.-м.н.
Якушева А.А.

Техническую поддержку конференции осуществляет ООО «ИнтраБио»

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Лахно В.Д. Новый этап развития математической биологии	8
--	---

I. ОРГАНИЗАЦИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

Коротков А.В. О государственной политике РФ в области использования и развития информационных технологий в области науки	9
--	---

Редько В.Г. Моделирование когнитивной эволюции – перспективное направление исследований на стыке биологии и математики	10
--	----

II. МАТЕМАТИЧЕСКИЕ ОСНОВЫ НАНОБИОЭЛЕКТРОНИКИ

Каширина Н.И., Лахно В.Д. Теорема вириала и межполярное взаимодействие	12
--	----

Лахно В.Д., Коршунова А.Н. Численное моделирование движения заряда в электрическом поле в холестейновской молекулярной цепочке в отсутствие диссипации	14
--	----

Лахно В. Д., Султанов В. Б. Биполярная модель сверхпроводимости ДНК	16
---	----

Лахно В. Д., Султанов В. Б. Квазистационарные неравновесные распределения заряда в регулярных последовательностях ДНК	18
---	----

Лахно В.Д., Фялко Н.С. Развал полярона в однородном полинуклеотиде под влиянием температуры термостата	20
--	----

Лахно В.Д., Соболев Е.В., Фялко Н.С. О распараллеливании вычислений в задаче моделирования динамики переноса заряда в ДНК при конечной температуре	23
--	----

Пономарёв О. А., Шигаев А. С., Лахно В. Д. Управляемая модуляция тока во фрагментах двухцепочечной ДНК	24
--	----

Шигаев А. С., Пономарёв О. А., Лахно В. Д. Варианты нарушения структуры двухцепочечной ДНК в областях её локального расплетания	26
---	----

Bystrov V.S., Dekhtyar Yu., Paramonova E.V., Fridkin V.M., Kliem H., Kholkin A. Nanobiomaterials and Bioferroelectricity: Molecular Modeling and Nanoscale Characterization	28
---	----

Симон К.В., Тулуб А.В. Ab initio расчет активного центра гемма для различных спиновых состояний. Проблема моделирования пассивного и активного состояний бациллы туберкулеза	29
--	----

III. МАТЕМАТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЯ БИОЛОГИЧЕСКИХ СИСТЕМ. ОБЩИЕ ВОПРОСЫ ОРГАНИЗАЦИИ БИОЛОГИЧЕСКИХ СИСТЕМ. МАТЕМАТИЧЕСКАЯ БИОФИЗИКА

Лихошвай В.А., Фадеев С.И., Штокало Д.Н. Об исследовании нелинейных моделей многостадийного синтеза вещества	30
--	----

González E., Deriabina A.S., Poltev V.I., Furmanchuk A., Gorb L., Leszczynski J. Quantum mechanics approach to molecular mechanics description of approximate planarity of intrinsically non-planar nucleic acid bases	32
--	----

Атауллаханов Ф.И. Как клетка исправляет ошибки при самоорганизации веретена деления	33
---	----

Герус М.А., Коршунова А.В., Молодцов М.И., Атауллаханов Ф.И. Исследование механизма диффузии Dam1-комплексов по микротрубочке	34
---	----

Захаров П.Н., Молодцов М.И., Атауллаханов Ф.И. Динамическая модель диффузии Dam1 комплекса вдоль микротрубочки	36
--	----

Казанцев А.П. Кодирование электрокардиосигналов на основе варибельности условной энтропии конечных разностей	38
--	----

Казанцев А.П. Оценка остаточной среднеквадратической погрешности измерения электрокардиосигналов	40
--	----

Калюжный И.М. О вычислительных методах исследования автоволновых процессов	42
--	----

Зотов А.К., Зотова Т.Ю., Благонравов М.Л. Регрессионный анализ функциональной асимметрии левого и правого желудочков сердца в условиях вазоренальной артериальной гипертензии	44
---	----

Лыжко Е.В. Роль терморецепторов в явлении “afterdrop”	47
---	----

Мазуров М.Е. О функциональной роли гетерогенности синоатриального узла при формировании единого ритма сердца	49
--	----

Тихонов Д.А., Соболев Е.В. Метод псевдосредних функций в теории RISM. Температурная зависимость гидратации пептида окситоцина	51
---	----

Тихонов Д.А., Соболев Е.В. Псевдосредние в методе интегральных уравнений теории жидкостей в приближении RISM	53
--	----

Москаленко А.В. О необходимости количественного анализа спонтанного дрейфа автоволнового вихря в 2D-среде	55
---	----

Харитонов А.С. , Охнянская Л.Г. Системные принципы физиологии и симметрия процессов рассеяния и концентрации энергии в биосистемах	57
--	----

IV. МОДЕЛИРОВАНИЕ ФУНКЦИОНИРОВАНИЯ МОЗГА

Dick O.E. Analysis of the psychorelaxation efficiency by changing the brain bioelectrical activity	60
Бердникова А.П., Махортых С.А. Анализ нейрофизиологической активности по данным магнитной энцефалографии	62
Борисюк Р.М., Казанович Я.Б. Частичная синхронизация нейронной активности и обработка информации в мозге	64
Дунин-Барковский В.Л. Бугорковые аттракторы и проблема поиска "кода разума"	66
Казанович Я.Б. Сегментация изображений на основе временной корреляции нейронной активности	67
Ляховецкий В.А., Боброва Е.В. Модель кратковременной памяти человека на основе нейронной сети	69
Мурзина Г.Б. Роль процессов эндо/экзоцитоза рецепторов и их латеральной диффузии в изменении ответа нейрона	71
Покровский А.Н. Возможность выделения синаптических токов из корковых вызванных потенциалов	73
Пронин С.В., Чихман В.Н., Шелепин Ю.Е. Вейвлеты в исследовании зрительного восприятия	75
Соболев Е.В., Тихонов Д.А. Множественность решений уравнений RISM/HNC в пределе малой плотности	77
Соболев Е.В., Тихонов Д.А. Параллельный алгоритм численного решения уравнений RISM в пределе бесконечного разбавления	79
Цукерман В.Д., Еременко З.С., Кулаков С.В., Каримова О.В. Когнитивная нейродинамика навигационного поведения млекопитающих	81
Чижов А.В. Иерархия математических моделей первичной зрительной коры	83
Шац В.Н. Анализ взаимодействия нейронов на основе математической модели информационной среды	85

V. СТРУКТУРНАЯ ГЕНОМИКА И РЕГУЛЯЦИЯ БИОЛОГИЧЕСКИХ ПРОЦЕССОВ. МОДЕЛИРОВАНИЕ БИОХИМИЧЕСКИХ СИСТЕМ

Панюков В. В., Осипов А. А. Автоматизированное изучение электростатического потенциала молекул ДНК	87
Golubyatnikov V.P., Golubyatnikov I.V., Kleshchev A.G. Existence of stable cycles in some multidimensional gene network models	89
Kiselev S.S., Purtov Yu.A., Ozoline O.N. Unified promoter-search algorithm as a novel tool for annotation of bacterial genomes	91
Nikolaev S., Zubairova U., Mjolsness E., Smal P., Shapiro B., Kolchanov N. A reaction-diffusion model of shoot apical meristem compartmentalization based on CLV/WUS interplay	93
Zubairova U., Nikolaev S., Fadeev S., Mjolsness E., Kolchanov N. Modified French Flag model: instability of the structure under cell growth and division	95
Mukha D. V., Tuzikov A. V., Usanov S. A. Elucidation of the functional role of amino acids forming the substrate-binding site of human CYP51A1 by molecular modeling and sequence analysis	96
Ермаков Г.Л. Методы линейной алгебры в биохимии и метаболическом моделировании	98
Дроздов-Тихомиров Л.Н. , Васильчиков В.В., Назипова Н.Н. Расчёт реального распределения скоростей реакций в метаболической системе растущей клетки методом баланса стационарных метаболических потоков (БСМП)	100

VI. МОДЕЛИ МОЛЕКУЛЯРНОЙ ЭВОЛЮЦИИ

Bazykin G.A., Kondrashov A.S. Detecting past positive selection through ongoing negative selection	102
Кондрашов А.С. Обнаружение положительного отбора на уровне ДНК Detecting positive selection at the DNA level	103
Науменко С.А., Кондрашов А.С., Базыкин Г.А. Исследование эволюционных ландшафтов приспособленности аминокислотных сайтов при помощи реверсий	105
Bukin Y.S. Model for simulation broken gene flow in one-dimensional populations	107
Korotkov E.V. Analysis of the triplet periodicity phase shifts in genes	109
Гунбин К.В., Генаев М.А., Афонников Д.А. Компьютерная система анализа режимов молекулярной эволюции генов и белков	111
Френкель Ф.Е., Коротков Е.В. Следы мутаций в генах, приведшие к сдвигу рамки считывания	113
Матушкин Ю.Г., Лашин С.А., Суслов В.В. Компьютерное моделирование эволюции сообществ одноклеточных гаплоидных организмов	115

VII. БИОИНФОРМАТИКА

Сулимова В.В., Разин Н.И., Моттль В.В., Мучник И.Б., Куликовский К. Наиболее правдоподобное множественное выравнивание белков на основе EM-алгоритма	117
Сулимова В.В., Моттль В.В., Мучник И.Б., Куликовский К. Единая вероятностная модель эволюции аминокислот для подстановочных матриц семейств РАМ и BLOSUM	119
Поройков В.В., Веселовский А.В., Иванов А.С., Скворцов В.С., Лисица А.В., Соболев Б.Н., Филимонов Д.А., Арчаков А.И. Биоинформатика: платформа от геномов к лекарству in silico	121
Пугачева В.М., Коротков Е.В. Поиск потенциальных сдвигов рамки считывания в генах с помощью генетического алгоритма	123
Родионов И.Н., Семенов А.В. О сравнении двух подходов исследования генетических текстов	125
Тегуев Р.К., Дедус Ф.Ф., Ольшвец М.М., Панкратов А.Н., Пятков М.И., Назипова Н.Н. Использование спектрального представления для анализа последовательностей ДНК	127
Руденко В.М., Коротков Е.В. Применение метода Монте Карло для поиска сдвигов рамки считывания в кодирующих последовательностях ДНК	129
Сергеев Р.С., Тузиков А.В., Еремин В.Ф. Статистические методы анализа первичных последовательностей белков ВИЧ-1	131
Суворова Ю.М., Коротков Е.В. Изучение склеенных генов в банке данных KEGG	133
Naumoff D.G. and Naumov G.I. New Family of Isomaltase Genes IMA in <i>Saccharomyces cerevisiae</i>	135
Naumoff D.G. and Stepuschenko O.O. Sequence Analysis of the GH114 Family Glycoside Hydrolases	137
Naumoff D.G. Sequence-Based Classification of Yeast Glycoside Hydrolases: CAZy vs. Génolevures	139
Назипова Н., Теплухина Е., Тюльбашева Г., Чалей М. HETEROGENE – база достоверных неизбыточных данных о периодичности и неоднородности в геномах организмов	141
Назипова Н.Н., Зимин Ф.А., Теплухина Е.И., Устинин М.Н. Определение консенсусной последовательности для участков скрытой периодичности геномов	143
Гуменюк А.С., Морозенко Е.В., Улаев В.А., Антосюк Е.В. О выделении структурных единиц в белковых последовательностях на основе критерия целостно-завершенного текста	145
Лбов Г.С., Гусев В.А. Построение логико-вероятностной модели кодирующих последовательностей геномов	147
Угольников О.А., Пономаренко Е.А., Демич Ю.А., Лисица А.В., Корнюшко В.Ф., Швец В.И. Системный анализ процесса построения проблемно-ориентированной базы данных с использованием ресурсов Entrez	149
Усанов Н.Н., Гильванова Е.А., Усанов Н.Г. Сравнительный анализ частотных словарей цепей геномных ДНК	152
Шкурят Т.П., Шевцова В.С. Разработка и создание биоинформационной базы данных, поддерживающей метод FIAV (Fast Identification of Alleles Variation) для диагностики наследственных болезней	154
Лизунов А.В., Орлов С.В., Перевозчиков А.П. Филогения малых аполипопротеинов позвоночных	156

VIII. ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ ВЫЧИСЛЕНИЯ В МОДЕЛИРОВАНИИ БИОЛОГИЧЕСКИХ СИСТЕМ

Теплухин А.В. Многоуровневое распараллеливание алгоритмов молекулярного моделирования методом Монте-Карло	158
Лихачев И.В., Балабаев Н.К. Использование графических сопроцессоров для интерактивного анализа траекторий молекулярной динамики	160
Пятков М.И., Панкратов А.Н. О масштабируемости на многоядерных процессорах алгоритма поиска повторов в ДНК	162
Мацкевич В.А., Орлова Н.А., Воробьев И.И., Юрьев А.С., Воробьев А.И. Поиск низкомолекулярного лиганда для афинной хроматографической очистки фактора свёртывания крови VIII человека молекулярным докингом in silico	164
Феранчук С.И., Калиновский А.А., Тузиков А.В. Сервер для вычислений по биоинформатике в грид-среде	166

IX. МОДЕЛИРОВАНИЕ ПРОСТРАНСТВЕННЫХ СТРУКТУР МОЛЕКУЛ БИОПОЛИМЕРОВ И МЕЖМОЛЕКУЛЯРНЫХ ВЗАИМОДЕЙСТВИЙ.

Ефимов А.В. Моделирование структуры белков	167
Machulin A. V., Deryusheva E. I., Selivanova O. M., Serdyuk I. N. Family of Ribosomal Proteins S1 Contains Unique Conservative Domain	169
Oparina N.U., Kuznetsov E.N., Esipova N.G., Tumanyan V.G., Anashkina A.A. Key ARG residues in protein-DNA interactions	171
Sobolev O.V., Lunin V.Y., Lunina N.L., Baumstark M.W. The use of cylindrical models and cluster analysis for phasing Low Density Lipoprotein particles	173
Анищенко И.В., Кисель М.А., Андрианов А.М. Компьютерный дизайн водорастворимых аналогов	

β-галактозилцерамида – потенциальных лекарственных препаратов для терапии СПИДа	175
Глякина А.В., Галзитская О.В., Балабаев Н.К. Молекулярно-динамическое моделирование механического разворачивания белков L и G	177
Голубович В.П., Соколов Ю.А., Фигловский В.А., Шутова И.В. Аминокислотная последовательность и конформационная структура инцептина гороха огородного <i>Pisum sativum</i>	179
Aznakayev E.G., Aznakayeva D.E. Modeling of Ribosome Movement Along Molecular Chain	181
Корнилов В.В., Балабаев Н.К. Исследование взаимодействий липидов в смешанных бислоях: моделирование методом молекулярной динамики	183
Лобанов М.Ю., Галзитская О.В.. Статистический анализ и предсказание неструктурированных остатков в белковых структурах	185
Панкратов А.Н., Руднев В.Р., Куликова Л.И., Дедус Ф.Ф., Ефимов А.В. Анализ пространственной структуры глобулярных белков на основе спектрально-аналитического метода	187
Потапова У.В., Беликов С.И., Феранчук С.И. Молекулярное моделирование трехмерной структуры белкового комплекса NS2B-NS3 протеазы высококовирулентных и инаппарантных штаммов вируса клещевого энцефалита	189
Финкельштейн А. В., Переяславец Л. Б. PFFsubl и PFFol: новые поляризуемые атомные силовые поля для расчета невалентных взаимодействий в явно и неявно заданном водном окружении	191
Шутова И.В., Голубович В.П. Разработка математических методов и программных средств выявления функциональных сайтов белков	192

X. ПОПУЛЯЦИОННОЕ МОДЕЛИРОВАНИЕ

Karkach A.S., Romanyukha A.A., Yashin A.I. Adaptive trade-off between reproduction and survival in Mediterranean fruit flies induced by changing dietary conditions	194
Глушков В.Н., Саранча Д.А. Использование компьютерных технологий при популяционном моделировании	195
Дружинина О.В., Масина О.Н. Условия устойчивости нечеткой модели динамики популяций	198
Кац И.Л., Реунова Г.Д., Музарок Т.И., Журавлев Ю.Н. Изучение популяционно-генетической структуры женьшеня настоящего <i>Panax ginseng</i> C.A. Meyer на основе AFLP-маркеров с применением различных пакетов компьютерных программ	200
Колобов А.Н., Фрисман Е.Я. Моделирование межвидовой конкуренции в смешанных разновозрастных древостоях	202
Кулаков М.П. Пространственная синхронизация в моделях метапопуляций	204
Неверова Г.П. Режимы динамики двухкомпонентной модели Лефковича с плотностным лимитированием пополнения репродуктивной группы	206
Пичугин Б. Ю., Перцев Н. В., Логинов К. К. Стохастическая модель динамики популяций, развивающихся в условиях воздействия токсичных веществ	208
Ревуцкая О.Л. Определение оптимального промыслового изъятия для двухвозрастной лимитированной популяции	210
Царегородцева Г.Е. Моделирование динамики популяции, развивающейся в условиях воздействия вредных веществ на процесс репродукции особей	212

XI. МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ БИОЛОГИЧЕСКИХ ДАННЫХ. МАТЕМАТИЧЕСКАЯ БИОМЕДИЦИНА

Bogatyrev M.Y., Tuhtin V.V. Conceptual Modelling In Biomedical Knowledge Discovering	214
Молородов Ю.И., Пестунов А.И., Пестунов И.А., Смирнов В.В., Федотов А.М. Интеграция данных для разработки и анализа моделей динамики биосферы	216
Дергузов А.В., Махортых С.А. Проблемы вибрационной экологии и безопасности высотных сооружений	218
Makhortykh S.A. Computer system for vibroacoustic impact ecology control on the environment	220
Лыжко Е.В., Махортых С.А. Изучение функциональных структур головного мозга методами магнитной энцефалографии	222
Куликова Л.И., Панкратов А.Н., Дедус Ф.Ф. Дифференцирование биологических сигналов и проблема обусловленности	224
Лбов Г.С., Полякова Г.Л. Решение задачи прогнозирования в классе логических решающих функций	226
Суховеров В.С., Романов Г.А. Модель управляемого гормонами биполярного роста линейных клеточных структур растительного типа	228
Кленкова Н.А., Капустин С.И., Шмелева В.М., Хромов-Борисов Н.Н., Блинов М.Н. Применение статистических методов анализа для изучения влияния аллельного полиморфизма генов ключевых ферментов метаболизма гомоцистеина на риск возникновения гипергомоцистеинемии и атеросклеротического поражения сосудов	230
Бельская Л.В., Голованова О.А. Использование корреляционного анализа для установления механизма образования слюнных камней	232

Бельская Л.В., Голованова О.А. Применение математических методов для обработки результатов анализа ротовой жидкости	234
Дюк В.А., Сенкевич Ю.И., Толстоногов Д.А. Метод информационного структурного резонанса в задаче анализа данных ДНК-микрочипов	236
Водолагина Н.Н., Костомарова И.В., Кузнецова А.В., Малыгина Н.А., Сенько О.В. Изучение связи тяжести дисциркуляторной энцефалопатии с полиморфизмом некоторых генетических факторов методами распознавания	238
Исаева О.Г., Осипов В.А. Моделирование противоракового иммунного ответа с учетом влияния радио- и иммунотерапии	240
Устинин М.Н., Панкратова Н.М. Усреднение данных магнитной энцефалографии при отсутствии информации об опорных точках	242
Устинин М.Н., Панкратов А.Н., Рыкунов С.Д., Наурзаков С.П., Гребенкин А.П. Сравнительный анализ экспериментальных данных магнитной энцефалографии	244
Бойко А.И., Ольшелев М.М., Устинин М.Н. Оценка минеральной плотности костной ткани в системе для массовой диагностики остеопороза	246
Шевцова В.С., Соловьев А.Н., Шевцова М.С. Оптимизация топологии крыльев летающих насекомых	248
Цепков Г.В. Информативность фрагментов сигнала с позиций нервной ткани	250
Хакимов А.Г., Кузин Р.А. О собственных колебаниях круглой мембраны с утонченной центральной областью	252
XII. МАТЕМАТИЧЕСКАЯ ЭПИДЕМИОЛОГИЯ И МОДЕЛИ ИММУНИТЕТА	
Мельниченко О.А. Сравнение программ контроля туберкулеза с помощью математической модели эпидемического процесса	254
Авилев К.К. Влияние социо-экономических факторов на выявление больных туберкулезом	256
Перцев Н.В. Математические модели эпидемического процесса с произвольным описанием длительности пребывания индивидуумов в различных стадиях заболевания	258
Цурко В.В., Михальский А.И. Математическое моделирование смертности с учётом сочетанной заболеваемости	260
Сергеев В.Г., Перевозчиков А.Г. Математическая модель иммунной памяти	262
XIII. ВЫЧИСЛИТЕЛЬНАЯ ЭКОЛОГИЯ	
Khanina L.G., Smirnov V.E., Glukhova E.M. Biodiversity assessment in Decision Support Systems for sustainable forest management	264
Абакумов А.И., Силкин В.А., Паутова Л.А., Пак С.Я. Формирование видовой структуры фитопланктона в зависимости от минерального питания	266
Грабарник П.Я. Моделирование пространственно-временной структуры лесной экосистемы: сравнение моделей конкуренции	268
Гречушкина Н.А. Количественные методы оценки классификаций растительности	270
Михайлов А.В., Михайлова Н.В. Модель FGV - моделирование живого напочвенного покрова в лесной экосистеме FGV model – modeling of forest ground vegetation in forest ecosystem	272
Михайлова Н.В., Михайлов А.В. Трёхмерная упрощенная модель затенения деревьев в лесу с учетом видоспецифического коэффициента проницаемости кроны	274
Молородов Ю.И., Федотов А.М. Структура модели данных атласа «Биосубстраты Сибири»	276
Романов М. С., Мечникова С. А. Лог-линейная модель динамики размеров гнезд и величины выводков кречета на Ямале	278
Рубашко Г.Е., Ханина Л.Г., Смирнов В.Э., Глухова Е.М. Анализ динамики состава растительных группировок муравейников рода <i>Formica</i>	280
IX. ОБОРУДОВАНИЕ МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ	
Фомкина М.Г., Терновский В.И., Грохлина Т.И., Монтрель А.М. Экспресс-диагностика концентрации мочевины в сыворотке крови с помощью ультратонкого полимерного датчика	282
Авторский указатель	284

Новый этап развития математической биологии

Лахно В.Д.

Институт математических проблем биологии РАН

lak@impb.psn.ru

Представление о том, какие проблемы решаются современной математической биологией и биоинформатикой, дают предыдущие два тома международной конференции «Математическая биология и биоинформатика» [1,2]. Их практическая направленность была задана в ходе выполнения знаменитого международного биоинформационного проекта 1990 - 2000 гг - «геном Человека» математические методы становятся всё более востребованы как существующими, так и новыми научными направлениями. В качестве примера нового магистрального направления, в котором методы математической биологии играют решающую роль можно привести нанобиоэлектронику []. Пожалуй, это первое в области биологии направление, в котором биологу без предварительного применения математики делать нечего. Схемы самосборки сложных биологических электронных систем здесь предварительно рассчитываются с помощью специальных алгоритмов с использованием суперкомпьютеров и лишь затем реализуются в пробирке.

В качестве примера давно сформировавшегося направления можно привести науку о мозге. Прорывные результаты в этой области можно также ожидать скорее не от новых постановок экспериментов, возможность проведения которых сильно ограничена требованием их неинвазивности, а от применения в этой области таких математических методов как обработка данных и распознавание образов. Для примера можно привести такое новое направление как компьютерная энцефалография. Идеи постановок новых экспериментов в когнитивных науках здесь часто исходят именно от математиков.

Предлагаемые читателю труды настоящей конференции посвящены дальнейшему развитию математической биологии и биоинформатики, как уже устоявшихся, так и новых её направлений.

О государственной политике РФ в области использования и развития информационных технологий в области науки

Коротков А.В.

Московский государственный институт международных отношений (Университет),
Академия народного хозяйства при Правительстве РФ

Korotkov_a_v@mgimo.ru; korotkov@itmane.ru

Abstract

This paper reviews the State policy and Governmental programs for the productivity of scientific researches in Russia Federation. To identify the determinants of cross-region and cross-sector disparities in different scientific and research centers both at RAS and large universities, including the internet GRID penetration, we examine a panel of 47 institutions over the 2001-2010 period. We used the official statistics and own research, which was mainly based on questionnaires, produced as a part of case-study produced by the students of master program at the Moscow State Institute of International Relations under the Ministry of Foreign Affairs of Russia Federation.

Official statistics indicate that productivity of scientific research strongly dependent on penetration of the tools of IT in this specific area. The lack of computing and small penetration of broadband telecommunications, connected in GRID, makes even the very

sophisticated research useless. This situation pushes the top- and high-level scientists to move out of Russia to foreign Universities not by economic, but scientific reasons.

First, we discuss the Governmental programs in creating scientific IT networks, and their role in enhancing productivity through greater specialization, standardization, and consolidation of scientific and research processes, and a shift to higher value-added services. The influence of these programs is interpreted

as a part of restructuring and modernization, namely as the unbundling of RAS functions as well as vertical disintegration. Second, as some IT-services become more like products, both low-skilled laboratory and high-skilled scientific jobs are subjected to productivity growth through standardization and digitization.

Key words: information technology, information society, state policy, standardization.

Моделирование когнитивной эволюции – перспективное направление исследований на стыке биологии и математики

Редько В.Г.

Научно-исследовательский институт системных исследований РАН, Москва, Россия
vgredko@gmail.com

Abstract

It is interesting to analyze cognitive evolution, evolution of animal cognitive abilities by means of mathematical/computer modeling. Modeling cognitive evolution is related with foundation of mathematics, cognitive science, investigations of adaptive behavior. Initial models of elementary cognitive features are described. Further steps of modeling cognitive evolution are discussed.

1 Актуальность моделирования когнитивной эволюции

Исследование когнитивной эволюции связано с глубокой гносеологической проблемой: почему человеческое мышление применимо к познанию природы? Рассмотрим, например, физику, одну из фундаментальных естественнонаучных дисциплин. Мощь физики связана с эффективным применением математики. Но математик делает доказательства независимо от внешнего мира, используя свое логическое мышление. Почему же результаты, получаемые математиком, применимы к реальной природе? Или в более общей формулировке: почему логика человеческого мышления применима к познанию природы?

Многие исследователи познавательных процессов задумывались над близкими вопросами. Например, рассматривая познавательные процессы в приближении отдельного взрослого человека и считая их априорными, И. Кант пришел к сомнениям о возможности познания природы [2]:

«...хотя вначале это звучит странно, но тем не менее верно, если я скажу: *рассудок не черпает свои законы (a priori) из природы, а предписывает их ей*».

Возражая Канту, К. Лоренц, один из основателей этологии, подчеркивал, что познавательные способности произошли в процессе эволюции и имеют определенные эмпирические корни [1].

Но как конкретно эти способности возникали? Как и почему появились способности, позволяющие познавать внешний мир?

Естественный подход к анализу проблемы – построение математических и компьютерных моделей когнитивной эволюции, осмысление с помощью моделей эволюционного происхождения мышления человека.

Для серьезной постановки работ в этом направлении целесообразно: 1) выделить ключевую проблему, 2) указать задел исследований, 3) предложить контуры программы будущих исследований когнитивной эволюции.

Ключевая проблема – проблема происхождения правил логического вывода. Можно ли исследовать, как произошли правила логического вывода, используемые математиком? Да, можно. Одно из элементарных правил, которое использует математик в логических заключениях – правило *modus ponens*: «если имеет место А, и из А следует В, то имеет место В», или $\{A, A \rightarrow B\} \Rightarrow B$. Перейдем от математика к собаке, у которой вырабатывают классический условный рефлекс. В памяти собаки формируется связь «за УС должен последовать БС» (УС – условный стимул, БС – безусловный стимул). Когда после выработки рефлекса собаке предъявляют УС, то она, помня о хранящейся в ее памяти записи $УС \rightarrow БС$, делает элементарный «вывод» $\{УС, УС \rightarrow БС\} \Rightarrow БС$. И собака ожидает БС. Конечно, применение правила *modus ponens* (чисто дедуктивное) математиком и индуктивный вывод, который делает собака, явно различаются. Но можем ли мы думать об эволюционных корнях логических правил, используемых в математике? Да, вполне можем – умозаключение математика и индуктивный вывод собаки аналогичны. При этом результат эволюции – правила логического вывода – известны и достаточно хорошо формализованы [3]. В основе этих выводов – элементарные правила, подобные *modus ponens*.

Направление исследований «Адаптивное поведение» – задел моделирования когнитивной эволюции. Это активно развивающееся направление исследований сформировалось в начале 1990-х годов [4]. Основной подход направления – конструирование и исследование искусственных (в виде компьютерной программы или робота) «организмов», способных приспосабливаться к внешней среде. Дальняя цель этих работ (пока еще явно нереализованная) – анализ эволюции когнитивных способностей животных и происхождения интеллекта человека.

Контуры программы будущих исследований когнитивной эволюции. Предложим этапы исследований когнитивной эволюции.

- Моделирование адаптивного поведения «организмов» с естественными потребностями: питание, размножение, безопасность.
- Исследование перехода от физического уровня обработки информации в нервной системе животных к уровню обобщенных образов, уровню понятий (аналогов слов).
- Исследование процессов формирования причинных связей в памяти животных. Анализ роли прогнозов в адаптивном поведении.
- Моделирование «логических выводов», используемых животными при адаптивном поведении. Сопоставление «логики поведения» животных с логикой человеческого мышления.

Перечисленные пункты очерчивают круг исследований от моделирования простейших форм поведения к логическим правилам, используемым в математике. Далее характеризуются начальные компьютерные модели в рамках этих исследований.

2 Модели простых когнитивных способностей

Формирование обобщающих эвристик. Моделировалось поведение автономного агента в двумерной клеточной среде. В части клеток двумерного мира имелись порции пищи. Агент обладал ресурсом R , который увеличивался при питании и уменьшался при выполнении агентом действий. Агент выполнял следующие действия: питание, перемещение на одну клетку вперед, поворот направо или налево, отдых.

Выбор действий агента обеспечивался имеющейся у него системой управления. Система управления агента представляла собой набор правил вида: $S \rightarrow A$, где S и A – ситуация и действие, соответствующие этому правилу. Ситуация S определялась наличием/отсутствием пищи в ближайших к агенту клетках. Правила оптимизировались методом обучения с подкреплением.

Сформированный в результате обучения набор правил можно рассматривать как 5 обобщающих эвристик: 1) если порция пищи расположена в той же клетке, в которой находится агент, то нужно выполнить действие «питание»; 2) если пищи нет в той клетке, в которой находится агент, и есть пища в клетке впереди агента, то нужно выполнить действие «перемещение вперед»; 3,4) если в указанных выше клетках пищи нет, а есть пища в клетке справа/слева от агента, то нужно выполнить действие «поворот направо/налево»; 5) если вообще нет пищи в ближайших клетках, то нужно выполнить поисковое действие «перемещение вперед». Тем самым происходил отбор правил,

приводящих к формированию цепочек действий агента, обеспечивающих увеличение ресурса агента.

Агенты с несколькими потребностями. Исследовалось поведение простых автономных агентов, имеющих несколько естественных потребностей: питание, размножение, безопасность. Система управления агента была основана на правилах того же вида, что и в первой модели. Мир, в котором находились агенты, состоял из двух клеток: одна клетка являлась опасной для агентов, вторая – безопасной. Периодически статус клеток менялся: опасная $\leftrightarrow$ безопасная. Агент, находящийся в опасной клетке, каждый такт времени терял большой ресурс. В мире имелась восполняемая пища агентов. Агенты выполняли следующие действия: деление, питание, перемещение в другую (альтернативную из двух) клетку, отдых.

Моделирование продемонстрировало формирование естественного поведения агентов. В частности, агенты своевременно перемещались из опасной клетки в безопасную, а при эволюционной оптимизации популяции агентов важную роль играло размножение.

Взаимодействие обучения и эволюции. Построена модель агентов, которые подобны биологическим организмам, приспосабливающимся к изменению температуры T в окружающей среде. Система управления агента основана на нейросетевых адаптивных критериях и обеспечивает прогнозирование изменений T и принятие решения о перемещении агента в соответствии с изменениями температуры. Продемонстрировано, что *приобретаемые в процессе индивидуального обучения агента полезные навыки могут генетически ассимилироваться в течение 3-5 поколений дарвиновской эволюции.*

References

- [1] Lorenz K. Kant's doctrine of the a priori in the light of contemporary biology // In H. Plotkin (Ed.). Learning, Development and Culture: Essays in Evolutionary Epistemology. New York: Wiley, 1982. PP. 121-143.
- [2] Кант И. Пролегомены ко всякой будущей метафизике, могущей появиться как наука. Соч. в 6-ти томах. Т. 4, часть 1. М.: Мысль, 1965. С. 67-210.
- [3] Математическая теория логического вывода (под ред. А.В. Идельсона и Г.Е. Минца). М.: Наука, 1967.
- [4] От моделей поведения к искусственному интеллекту. Серия «Науки об искусственном» (под ред. В.Г. Редько). М.: УРСС, 2006.

Теорема вириала и межполяронное взаимодействие

Каширина Н.И.⁽¹⁾, Лахно В.Д.⁽²⁾

ИФП НАН Украины⁽¹⁾, ИМПБ РАН⁽²⁾
n_kashirina@mail.ru⁽¹⁾, lak@impb.psn.ru⁽²⁾

Аннотация

В работе рассмотрены вопросы, связанные с выполнением теоремы вириала в электронных системах с кулоновскими взаимодействиями. В качестве примера рассмотрен биполарон в полярном кристалле для произвольной величины электрон-фононного взаимодействия.

1 Введение

Вот уже 15 лет не утихает дискуссия о возможности переноса заряда в ДНК. Особую актуальность играет изучение механизмов подобного переноса [4]. В случае экспериментов с сухой ДНК в однородных цепочках, по-видимому, реализуется полярон большого радиуса [2]. В то же время, механизм переноса для поляронов малого радиуса, практически полностью локализованных в глубокой потенциальной яме на одном нуклеотиде, когда вероятность температурных скачков мала, становится проблематичным. В работе исследуются проблемы, связанные с выполнением теоремы вириала (ТВ), относящиеся к наиболее общим закономерностям поляронного и биполаронного гамильтонианов. В связи с этим исследуются эффективные функционалы, полученные на основе гамильтонианов Фрёлиха для полярона и биполарона (БП) большого радиуса. Подобные эффективные функционалы получают в результате использования различных вариационных методов, используемых для изучения энергии подобных систем.

2 Теорема вириала для полярона и биполарона сильной связи

2.1 Масштабное преобразование

В пределе сильной связи энергии основного состояния полярона и биполарона (БП) определяются как нижняя граница соответствующих функционалов сильной связи. Единственным дополнительным условием является нормируемость пробной волновой функции (ВФ). Пусть $\Psi_{Bp}(\mathbf{r})$ - ВФ БП, $\mathbf{r} \equiv \{\mathbf{r}_1, \mathbf{r}_2\}$ - полный набор электронных координат системы. В пределе сильной связи слагаемые, соответствующие кинетической и

потенциальной энергии функционалов полярона и БП, обладают свойством однородности. В связи с этим, применение масштабного преобразования $\mathbf{r} \rightarrow \lambda \mathbf{r}$ позволяет провести варьирование по параметру масштабного преобразования λ в аналитическом виде и получить простые выражения для энергии основного состояния, для которых в экстремальных точках выполняется ТВ в простейшем виде, а именно: кинетическая энергия системы равна полной энергии, взятой с обратным знаком.

2.2 Теорема Пекара 1:2:3:4 для полярона сильной связи

В применении к полюрону ТВ исследовалась в работах С. И. Пекара [5] и получила название теоремы 1:2:3:4. В данной теореме утверждается, что в пределе сильной связи справедливы следующие соотношения:

$$T_p = -F_p, E_{ph} = -2E_p, E_{el} = 3F_p, E_{int} = 4F_p \quad (1)$$

$$\text{где } F_p = T_p + \frac{1}{2}E_{int}, F_{el} = T_p + E_{int}.$$

$E_p, T_p, E_{ph}, E_{int}$ - средние значения энергии основного состояния, кинетической энергии, энергии поля фононов и энергии электрон-фононного взаимодействия соответственно. В пределе сильной связи в точках минимума полная энергия полярона $E_p = T_p + E_{int} + E_{ph}$ равна кинетической энергии, взятой с обратным знаком, а величина F_p совпадает с полной энергией, которая может трактоваться как энергия термоионизации полярона (энергия самосогласованного состояния).

ТВ для БП сильной связи исследовалась в работе [3]. Для отличного от нуля расстояния между поляронами R даже в пределе сильной связи ТВ имеет более сложный вид

$$R \frac{dE_{Bp}}{dR} + 2T_{Bp}(R) + V_{Bp}(R) = 0. \quad (2)$$

3 Теорема 1:3:4 для электронных систем при произвольной величине электрон-фононного взаимодействия

3.1 Теорема вириала для полярона при произвольной величине электрон-фононного взаимодействия

Условия выполнения ТВ для поляронного функционала в общем виде для произвольной связи исследовались в работе [1].

Соотношения (1) могут быть получены в общем виде в результате применения к поляронному функционалу теоремы Гельмана-Фейнмана, согласно которой, в том случае, когда величина $E_0 = \langle \Phi_0 | H | \Phi_0 \rangle$ зависит от некоторого параметра β , производная по параметру выражается в виде

$$\frac{dE_0}{d\beta} = \langle \Phi_0 | \frac{dH_0}{d\beta} | \Phi_0 \rangle. \quad (3)$$

В качестве параметра могут рассматриваться заряд электрона, эффективная масса, фононная энергия и др. Вывод выражений (1) с использованием теоремы Фейнмана приводится в работе [1]. Подобный приём не ограничен пределом сильной связи. В [1] приводится доказательство того, что для полярона соотношение теоремы 1:3:4 остаётся справедливым для всего диапазона изменения константы электрон-фононной связи, в то время как соотношение 1:2 выполняется только в пределе сильной связи. Соотношение 1:2 в теореме является следствием того, что основное состояние должно быть пропорционально фрёлиховской константе электрон-фононного взаимодействия α и соответствует пределу сильной связи, рассмотренному С.И. Пекаром [5].

3.2 Теорема вириала для систем с кулоновскими взаимодействиями

Метод получения теоремы вириала, использовавшийся в работе [1] для полярона может быть применён к электронной системе, содержащей кулоновские взаимодействия. В качестве объектов могут выступать, связанные на кулоновских центрах поляроны (мелкие примесные центры в ковалентных или F-центры в ионных кристаллах), свободные и связанные на кулоновских центрах биполароны (D^- или F' -центры), обменно-связанные пары мелких примесных центров или F_2 -центры), а также более сложные системы примесных центров в фононном поле. После выполнения несложных преобразований, можно показать, что для одноэлектронных и двухэлектронных (в общем случае многоэлектронных) систем в фононном поле при произвольной величине электрон-фононного взаимодействия остаётся справедливой теорема 1:3:4, имеющая в рассматриваемом нами случае более общий вид.

$$-T = \tilde{F}, \quad \tilde{E}_{el} = 3\tilde{F}, \quad \tilde{E}_{int} = 4\tilde{F}, \quad (4)$$

где введены новые обозначения, обобщающие определения, введенные в [1] на случай систем с кулоновскими взаимодействиями:

$$\tilde{F} = T + \frac{1}{2}, \quad \tilde{E}_{int} + 2E_C, \quad \tilde{E}_{el} = T + \tilde{E}_{int}, \quad (5)$$

где T - кинетическая энергия, E_{int} - энергия

электрон-фононного взаимодействия, $E_C = \langle \Phi_0 | H_C | \Phi_0 \rangle$ содержит все кулоновские взаимодействия в системе.

Вопрос о вычислении величин, входящих в выражения (4, 5) заслуживает отдельного рассмотрения. Так в методах, использующих канонические преобразования, при проверке выполнения ТВ, определяемой выражениями (4), следует сначала провести каноническое преобразование и получить выражения, соответствующие новым (перенормированным) значениям операторов кинетической энергии и электрон-фононного взаимодействия. После этого следует провести усреднение по электронным ВФ. В качестве конкретного примера нами проведена численная проверка, которая показывает, что вириальные соотношения (4) выполняются для БП, рассмотренного методом Буймистрова – Пекара, для произвольной величины электрон-фононного взаимодействия. Так же как и в пределе сильной связи, ТВ выполняется как для одноцентровой конфигурации БП, так и для двухцентровой, соответствующей мелкому побочному минимуму, который исчезает при выборе электронной ВФ с учетом электронных корреляций, связанных с прямой зависимостью ВФ от межэлектронного расстояния.

Т.о. в точках, соответствующих минимальным значениям функционала (включая побочные минимумы) ТВ выполняется автоматически.

Работа выполнена при поддержке РФФИ, проекты № 07-07-00313, № 09-07-12073-офи_м.

References

- [1] J.T. Devreese, L.F. Lemmens, *Sol. St. Comm.* 12 p. 1067-1070, 1973.
- [2] A.A. Voityk. In: *Modern Methods for Theoretical Physical Chemistry of Biopolymers*. Eds. Starikov E.B., Lewis S., Tanaka S. Amsterdam: Elsevier, 2006.
- [3] Н.И. Каширина, В.Д. Лахно, ФТТ, том 50, вып.1, с.11-16, 2008
- [4] В.Д. Лахно, В.Б. Султанов. *Математическая биология и биоинформатика*, том 4, № 2, с. 46-51, 2009.
- [1] С.И. Пекар Исследования по электронной теории кристаллов. М -Л: ГИТТЛ 1951.

Численное моделирование движения заряда в электрическом поле в холстейновской молекулярной цепочке в отсутствии диссипации

В.Д.Лахно, А.Н.Коршунова

Институт математических проблем биологии РАН,

142290 Пушкино, Московская область, Россия

alya@impb.ru

Аннотация

Проведено численное моделирование стационарного движения заряда в дискретной цепочке в электрическом поле и изучены условия перехода стационарного движения в колебательное с блоховской частотой колебаний.

являются уравнениями Шрёдингера, где b_n представляют амплитуды вероятности локализации заряда на n -ом сайте, (2) – классические уравнения движения для смещений сайтов β_n .

Для численного моделирования движения полярона нами использовались безразмерные переменные, в которых система уравнений (1), (2) имеет вид:

$$i \frac{db_n}{dt} = \eta(b_{n-1} + b_{n+1}) + \kappa \omega^2 u_n b_n + E b_n, \quad (3)$$

$$\frac{d^2 u_n}{dt^2} = -\omega^2 u_n - |b_n|^2, \quad (4)$$

где, входящие в (1), (2) размерные величины связаны с входящими в (3), (4) безразмерными величинами соотношениями:

$$\eta = \tau v / \hbar, \omega^2 = \tau^2 K / M, q_n = \beta u_n, t = \tau \tilde{t}, \quad (5)$$

$$\kappa \omega^2 = \tau^3 (\alpha')^2 / M \hbar, \beta = \tau^2 \alpha' / M, E = \Delta e a \tau / \hbar,$$

где τ – произвольный масштаб времени.

1 Введение

Интерес к моделированию движения заряженной частицы в различного типа молекулярных цепочках связан с множеством возможных практических приложений таких цепочек в наноэлектронике [1,2]. На примере дискретной холстейновской молекулярной цепочки в работе [3] было показано, что в слабом электрическом поле происходят блоховские осцилляции полярона как целого, которые в сильном электрическом поле имеют вид бризерных колебаний. В противоположность этому, в работе [4] было показано, что в континуальном пределе в электрическом поле, не превосходящем некоторого критического значения, движение полярона может быть стационарным.

2 Математическая модель

Для изучения динамики движения заряда в цепочке мы используем полуклассическое описание, основанное на том, что масса сайта M намного превосходит массу электрона. Динамические уравнения имеют вид:

$$i \hbar \dot{b}_n = v(b_{n-1} + b_{n+1}) + \alpha' q_n b_n + e \Delta a n b_n, \quad (1)$$

$$M \ddot{\beta}_n = -K \beta_n - \alpha' |b_n|^2, \quad (2)$$

где v – матричный элемент перехода заряда с n -го сайта на $n \pm 1$ -й, α' – константа взаимодействия заряда со смещениями сайтов q_n , e – заряд электрона, Δ – напряжённость электрического поля, a – постоянная решетки, M – эффективная масса сайта, K – упругая постоянная. Уравнения (1)

3 Моделирование стационарного и колебательного движения полярона в электрическом поле

В континуальном пределе решение системы уравнений (3),(4) приводит к следующему соотношению между равновесной скоростью полярона $V = va/\tau$ и напряжённостью электрического поля [4]:

$$E = 2\pi^2 \frac{\omega^4 \eta^2}{\kappa} \frac{1}{V^4} \frac{1}{\text{sh}^2(2\pi\eta\omega/\kappa V)}. \quad (6)$$

Согласно [4], стационарное движение полярона в рассматриваемой нами консервативной системе в электрическом поле оказывается возможным вследствие того, что в процессе перемещения полярона по цепочке за ним остается "хвост" из возбужденных осцилляторов цепочки. При этом работа, совершаемая электрическим полем, расходуется на возбуждение этих осцилляторов.

На Рис1.а) показаны зависимости скорости полярона V от напряжённости поля E , определяемые

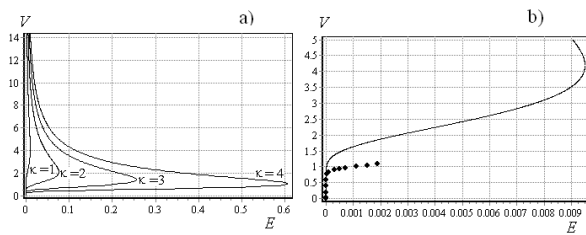


Рис.1. а)- Зависимость (6) скорости полярона V от напряжённости электрического поля E для различных значений параметра κ ($\kappa=1,2,3,4$), $\eta=1.276$, $\omega=1$. б)- Сравнение теоретической (—) (6) и численной (•) зависимости скорости полярона V от напряжённости электрического поля E при $\kappa=1$, $\eta=1.276$, $\omega=1$.

соотношением (6), для различных значений параметра κ . На Рис.1.б) приведено сравнение средней скорости стационарного движения полярона, полученное численным интегрированием системы уравнений (3), (4) с теоретической кривой, определяемой (6). Для рассматриваемых значений параметров κ и η континуальное приближение выполняется с большой точностью, поскольку радиус поляронного состояния: $\tilde{r} = r/a = 4\eta/\kappa \approx 5.1 \gg 1$.

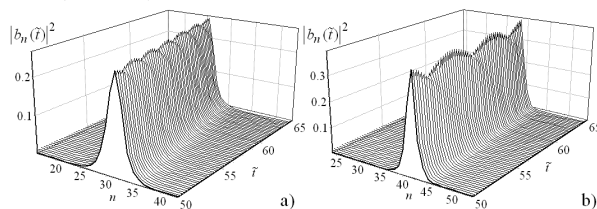


Рис.2. Движение полярона с постоянной скоростью в электрическом поле напряжённостью $E=0.01$ при $\omega=1$, $\eta=1.276$ для различных значений параметра κ : а) $\kappa=2$, б) $\kappa=3$. Хорошо заметны осцилляции формы полярона, амплитуда которых зависит от значения параметра κ .

На Рис.2 показано, что в слабом электрическом поле полярон движется равномерно, испытывая небольшие колебания своей формы с периодом $T_{PN} = 1/V$. Эти колебания связаны с дискретностью цепочки и обусловлены наличием в дискретной цепочке потенциала Пайерлса-Набарро. Амплитуда этих осцилляций обращается в ноль в континуальном пределе, когда размер полярона намного превышает расстояние между соседними сайтами. Численное сравнение периода осцилляций $T_{PN}(E)$ с величиной $1/V$ для различных значений напряжённости электрического поля E представлено на Рис.3. Для численного расчета скорости полярона V использовались соотношения: $V = dX/dt$,

$$X(t) = \sum_n n |b_n(t)|^2.$$

4 Обсуждение результатов и выводы

В заключение отметим, что в данной работе нами проведено прямое численное моделирование движения холстейновского полярона в

бездисперсной цепочке. Проведенное моделирование показало, что в электрическом поле

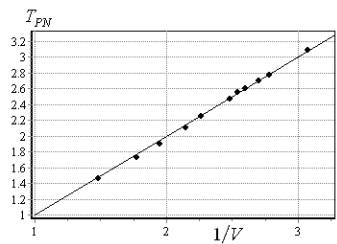


Рис.3. Численное (•) сравнение периода осцилляций $T_{PN}(E)$ с величиной $1/V$ для $E=0.02, 0.015, 0.01, 0.005, 0.003, 0.001, 0.0007, 0.0005, 0.0003, 0.0002, 0.0001$ при следующих значениях параметров: $\kappa=2$, $\eta=1.276$, $\omega=1$. Правая верхняя точка (•) на графике соответствует значению $E=0.0001$. Прямая линия соответствует графику $T_{PN} = 1/V$.

$E < E_{cr}$ полярон движется по цепочке с постоянной скоростью. В области значений параметров, при которых размер полярона становится сравним с постоянной решетки, его движение сопровождается колебаниями его формы (высоты и ширины распределения $|b_n(t)|^2$). В процессе колебаний скорость полярона V остается постоянной. В рассматриваемой нами консервативной системе постоянство скорости полярона обусловлено тем, что в процессе передвижения по цепочке полярон возбуждает в ней колебания сайтов.

При $E > E_{cr}$ равномерное движение полярона становится невозможным: энергия, получаемая поляроном в единицу времени от поля, становится больше, чем отдаваемая на возбуждение колебаний в цепочке. Равномерное движение в этом случае "срывается" и переходит в блоховские осцилляции. В отличие от блоховских осцилляций в недеформируемой цепочке, когда центр осцилляций остается на месте, в деформируемой цепочке происходит поступательное движение центра масс полярона, сопровождаемое его блоховскими осцилляциями. В сильном поле движение полярона становится подобным движущемуся бризерному состоянию.

Работа выполнена при поддержке РФФИ, проект №10-07-00112-а, №09-07-12073-офи-м.

Авторы благодарны МСЦ РАН за предоставленные вычислительные ресурсы.

Литература

- [1] V. D. Lakhno, *Int. J. Quant. Chem.*, **108**, 1970-1981, 2006.
- [2] Nanobioelectronics for Electronics, Biology and Medicine, Ed. A. Offenhäusser, R. Rinaldi, Springer, 2009.
- [3] V. D. Lakhno, A. N. Korshunova, *Eur. Phys. J. B*, **55**, 85-87, 2007.
- [4] V. D. Lakhno, *Int. J. Quant. Chem.*, **110**, 127-137, 2010.

Биполярная модель сверхпроводимости ДНК

В. Д. Лахно, В. Б. Султанов

Институт математических проблем биологии РАН

lak@impb.psn.ru

Аннотация

На примере однородных полинуклеотидных последовательностей показана возможность биполярной сверхпроводимости ДНК. При параметрах модели, характерных для ДНК, температура сверхпроводящего перехода $T_c \approx 6$ К. Биполярная модель может служить основой для объяснения экспериментов по сверхпроводимости ДНК.

1 Поляроны и биполяроны

Роль поляронов в процессе переноса заряда в ДНК интенсивно обсуждается и подтверждается как в моделях переноса заряда, так и при интерпретации посредством этих моделей экспериментов [1-3]. В большинстве работ в качестве носителей заряда в ДНК рассматриваются дырки [4,5].

В известной микроскопической модели сверхпроводимости БКШ пары электронов с противоположными спинами и импульсами образуют связанные состояния через взаимодействие с фононами. Бозе-конденсат этих пар и обеспечивает сверхпроводимость [6]. При интерпретации экспериментов с высокотемпературными сверхпроводниками в качестве основной рассматривается модель, в которой связанные состояния, биполяроны, образуют пары поляронов с противоположными спинами частиц [7,8].

В ДНК сверхпроводящее состояние наблюдалось при температуре $T \approx 1$ К в хаотических λ -ДНК [9]. Этот факт означает, что носителями тока могут быть дырки в биполярном состоянии. В хаотических цепочках бозе-конденсация биполярного газа затруднена неоднородностью нуклеотидного состава, что может являться причиной наблюдаемой низкой температуры сверхпроводящего перехода. В однородных полинуклеотидных последовательностях такие состояния могут обладать большей стабильностью.

2 Биполяроны в однородных ДНК

Волновая функция ψ_i полярона в однородной цепочке $(G)_N$ является решением уравнения (для ДНК $\eta=0.084$ эВ, $\kappa=0.5267$ эВ [10]), $i=1, \dots, N$:

$$\eta(\psi_{i-1} + \psi_{i+1}) - \frac{\kappa}{2} |\psi_i|^2 \psi_i = \lambda \psi_i, \quad \sum_i |\psi_i|^2 = 1,$$

с наименьшей энергией $E_p = \lambda + \frac{\kappa}{4} \sum_i |\psi_i|^4 + 2\eta$.

Аналогичное уравнение для биполярной волновой функции ψ_{ij} , которая задает состояние пары дырок с противоположными спинами [11]:

$$\eta(\Delta \psi)_{ij} + (-\kappa(\rho_i + \rho_j) + U\delta_{i-j})\psi_{ij} = W\psi_{ij},$$

$$\rho_i = \frac{1}{2} \sum_k (|\psi_{ik}|^2 + |\psi_{ki}|^2), \quad \psi_{ij} = \psi_{ji}, \quad \sum_{i,j} |\psi_{ij}|^2 = 1,$$

$(\Delta \psi)_{ij} = \psi_{i-1,j} + \psi_{i+1,j} + \psi_{i,j-1} + \psi_{i,j+1}$, где для ДНК кулоновское отталкивание $U \approx 1$ эВ. Энергия:

$$E_{bp} = W + \frac{\kappa}{2} \sum_{i,j} |\psi_{ij}|^2 (\rho_i + \rho_j) + 4\eta.$$

Условие стабильности биполярона – неравенство:

$$|E_{bp}| > 2|E_p|.$$

Биполяроны и поляроны в $(G)_N$ рассчитывались минимизацией функционалов $E_{bp}(\psi_{ij})$ и $E_p(\psi_i)$ с условиями симметрии и нормировки, $\eta=0.084$ эВ в области параметров κ , U , содержащей указанные выше их значения для ДНК.

На Рис. 1 показана граница области стабильности

$$|E_{bp}| - 2|E_p| = 0.$$

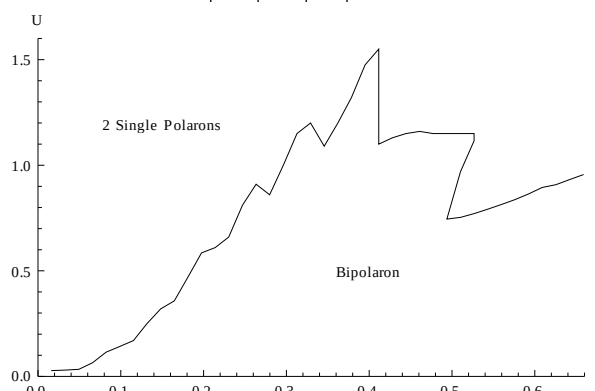


Рис. 1. Область стабильности биполяронов.

Вне области стабильности биполярон распадается на два одиночных полярона, находящихся на большом расстоянии друг от друга, энергия связи таких поляронов $\Delta = |E_{bp}| - 2|E_p| < 0$.

Примеры плотности распределения заряда $p_i = 2 \sum_j |\psi_{ij}|^2$ стабильных биполяронов, $N=31$:

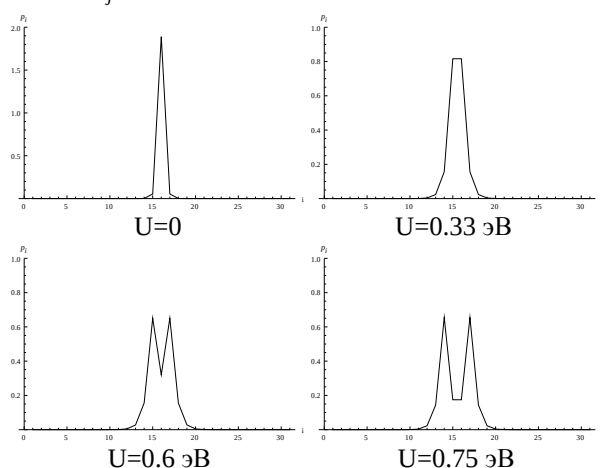


Рис. 2. Плотность заряда биполяронов, $\kappa=0.5267$ эВ.

В области стабильности связанное биполярное состояние образуют поляроны, находящиеся на расстоянии 0, 1, 2 и 3 расстояния между сайтами.

В Табл. 1 приведены значения энергий связи Δ биполяронов при $\kappa = 0.5267$ эВ, $0 \leq U \leq 0.75$ эВ:

Табл. 1. Энергии связи биполяронов Δ (эВ).

U	0	0.25	0.5	0.75
Δ	0.203	0.022	0.0018	0.00047

При описании сверхпроводимости биполярным механизмом критическая температура перехода непосредственно зависит от энергии связи Δ . При $\kappa = 0.5267$ эВ, $U=0.75$ эВ величина $\Delta \approx 5 \cdot 10^{-4}$ эВ дает температуру сверхпроводящего перехода $T_c \approx 6$ К.

Оценки величин $\kappa \approx 0.5$ эВ, $U \approx 1$ эВ являются приближенными. В растворе значение κ может сильно возрасти, согласно [12,13], примерно в 5 раз. В этом случае условия образования биполяронов с большой вероятностью будут выполняться.

3 Заключение

Полученные результаты свидетельствуют о возможности биполярных состояний в однородных полинуклеотидных цепочках и описания экспериментов по сверхпроводимости ДНК в рамках биполярной модели.

Отметим, что ранее на возможность биполярных состояний в коротких (6-сайтовых) неоднородных олигонуклеотидных цепочках обращалось внимание в [14]. Согласно [14], локализация пар в таких цепочках должна

происходить на дуплетах, триплетах и квадруплетах гуанинов.

Основной проблемой, связанной с возможностью сверхпроводящего состояния в ДНК, является необходимость существования носителей тока (электронов или дырок) в молекуле, поскольку сама по себе ДНК является диэлектриком. В эксперименте [9] при измерении проводимости молекула ДНК прикреплялась к двум рений-углеродным электродам, лежащим на слюдяной подложке. Зазор между электродами в [9] создавался выжиганием лазером полосы (толщиной $\approx 0.5 \mu\text{м}$), которая представляет собой слюду с оставшимися на ней не связанными друг с другом островками из рений-углеродных атомов. По нашему мнению появление носителей тока в ДНК могло быть обусловлено контактом прикрепленной к подложке молекулы с рений-углеродными островками зазора.

Работа выполнена при поддержке РФФИ, проекты № 10-07-00112-а; № 09-07-12073-офи-М.

Литература

- [1] E. M. Conwell, S. V. Rakhmanova, Proc. Nat. Acad. Sci. USA, v. 97, p. 4556, 2000.
- [2] E. M. Conwell, Proc. Nat. Acad. Sci. USA, v. 102, p. 8795, 2005.
- [3] C.-S. Liu, G. B. Shuster, J. Am. Chem. Soc. v. 125, p. 6098, 2003.
- [4] V. D. Lakhno, Int. J. Quant. Chem., v. 108, p. 1970, 2008.
- [5] A. Offenhäusser, R. Rinaldi (eds). Nanobioelectronics – for Electronics, Biology and Medicine, Springer, 2009.
- [6] А. А. Абрикосов. Основы теории металлов. М.: Наука, 1987.
- [7] J. T. Devreese, A. S. Alexandrov Rep. Progr. Phys., v. 72, p. 066501, 2009.
- [8] Н. И. Каширина, В. Д. Лахно, УФН, 2010, в печати.
- [9] A. Y. Kasumov et al, Science, v. 291, p. 280, 2001.
- [10] V. D. Lakhno, Int. J. Quant. Chem., v. 110, p. 127, 2010.
- [11] L. Proville, S. Aubry, Physica D, v. 113, p. 307, 1998.
- [12] A. A. Voityuk, J. Chem. Phys., v. 122, p. 204904, 2005.
- [13] A. A. Voityuk, J. Phys. Chem., v. 109, p. 10793, 2005.
- [14] V. Apalkov, T. Chakraborty, Phys. Rev. B, v. 73, p. 113103, 2006.

Квазистационарные неравновесные распределения заряда в регулярных последовательностях ДНК

В. Д. Лахно, В. Б. Султанов

Институт математических проблем биологии РАН

lak@impb.psn.ru

Аннотация

Наблюдающиеся в экспериментах на периодических последовательностях ДНК неравномерные распределения заряда могут быть получены как квазистационарные неравновесные распределения средних вероятностей в нелинейной модели. В линейной модели эти распределения стационарны и возникают вследствие образования при некоторых параметрах локализованных состояний с энергиями в запрещенных зонах спектра гамильтониана.

1 Эксперимент и модели

В [1], [2] построена модель распределения заряда в ДНК при конечной температуре, основанная на применении канонического ансамбля Гиббса к множеству всех решений стационарного дискретного нелинейного уравнения Шредингера (ДНУШ), которое в одностороннем случае имеет вид:

$$\lambda \psi_n = -\kappa_n |\psi_n|^2 \psi_n + a_n \psi_n + b_{n-1} \psi_{n-1} + b_n \psi_{n+1},$$

где $\vec{\psi} = (\psi_1, \psi_2, \dots, \psi_N)$ – вектор состояния дырки, a_n – потенциалы окисления нуклеотидов, b_n – матричные элементы перехода дырки между соседними нуклеотидами, составляющие гамильтониан H – тридиагональную матрицу; величина κ_n определяет взаимодействие дырки с колебаниями нуклеотидов на n -м сайте. Аналогичное ДНУШ имеет место и для дуплекса с гамильтонианом – блок-тридиагональной матрицей.

Расчет по построенной модели дает равномерное распределение заряда по гуанинам для однородных последовательностей $(G)_N$ и для периодических последовательностей $GG(A \dots AGG)_N$, m аденинов A . В экспериментах из работы [3] распределение на периодических последовательностях близко к равномерному по гуанинам (что свидетельствует о близости этого распределения к равновесному) только при $m \leq 2$. При $m > 2$ распределение заряда неравномерное, все более локализованное на первой паре гуанинов при возрастании m (в начальный момент в экспериментах заряд возникает на 1-м G_1).

Ниже будет показано, что неравномерные распределения, качественно согласующиеся с экспериментальными, можно получить как средние

вероятности $p_n \approx \frac{1}{t} \int_0^t |\psi_n(\tau)|^2 d\tau$, $t \rightarrow \infty$ из решений нестационарного ДНУШ с начальными условиями $\psi_n(0) = \delta_{1n}$, отвечающими экспериментам.

2 Средние вероятности, линейная модель

2.1 Односторонние последовательности

В линейном случае решение уравнения Шредингера с гамильтонианом H и начальными условиями выше $\vec{\psi}(t) = e^{-itH} \vec{\delta}_1$ выражается через собственные векторы $\vec{\psi}^k$ и спектр H : $\psi_n(t) = \sum_k e^{-it\lambda_k} \psi_n^k \psi_1^k$.

Отсюда получается формула для средних вероятностей $\frac{1}{t} \int_0^t |\psi_n(\tau)|^2 d\tau \rightarrow p_n = \sum_k |\psi_n^k|^2 |\psi_1^k|^2$, $t \rightarrow \infty$,

представляющая собой эргодическую теорему, где вместо ансамбля Гиббса вероятности усредняются по спектральной мере тридиагональной матрицы H . Для однородных последовательностей $(G)_N$ по этой формуле получаем равномерное распределение $p_n = 1/N$. У периодических матриц, непрерывный спектр которых состоит из разрешенных зон, при некоторых параметрах могут существовать дискретные точки спектра в запрещенных зонах, которым отвечают собственные векторы, локализованные на концах цепочки. Рассмотрим для простоты последовательности $G(A \dots AG)_9$, $m = 2, 3$.

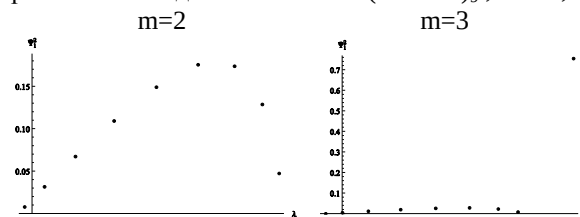


Рис. 1. Спектральная плотность.

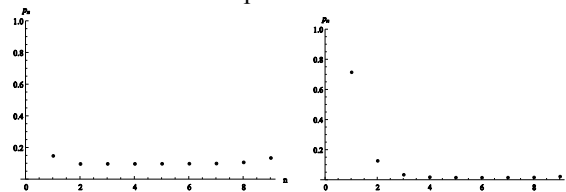


Рис. 2. Распределение p_n .

На Рис. 1 видно отделение дискретного собственного значения от разрешенной зоны спектра при переходе от $m=2$ к $m=3$. На Рис. 2 виден переход от делокализованного распределения вероятностей к

локализованному на 1-х сайтах G_n . Этот переход сохраняется при усреднении по случайно возмущенным матричным элементам с небольшой дисперсией. Поэтому такой переход есть структурное свойство периодических матриц, а не отдельного мостика $GA...AG$.

В отличие от ансамбля Гиббса, состояния, вносящие основной вклад в спектральную меру, не являются состояниями с наименьшей энергией, так как их собственные значения расположены вблизи верхней границы нижней полосы спектра. Поэтому эти распределения не могут быть равновесными.

2.2 Дуплексы

Гамильтониан даже однородных дуплексов является периодической матрицей, так как чередуются нуклеотиды на комплементарных цепях. Поэтому при некоторых параметрах на однородных дуплексах могут существовать локализованные на концах цепочки состояния [4].

Расчет p_n по формуле средних вероятностей для

периодических дуплексов $G \begin{pmatrix} A...AG \\ T...TC \end{pmatrix}_5$, число

A/T пар $m=1,2,...,7$, показывает, что существуют матричные элементы, при которых распределения качественно согласуются с экспериментами из [3]. На Рис. 3 показано распределение заряда p_n по гуанинам и точки эксперимента, пример $m=4,5,6,7$:

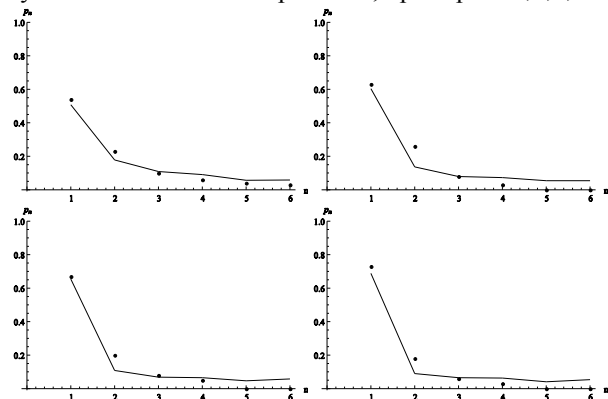


Рис. 3. Распределение p_n и точки эксперимента.

3 Нелинейная модель

В нелинейном случае средние вероятности p_n рассчитывались интегрированием на большое время нестационарного ДНУШ с начальными условиями $\psi_n(0) = \delta_{1n}$. Для однородных последовательностей $(G)_N$ получаем равномерное распределение $p_n = 1/N$. В случае периодических последовательностей, как одонитевых, так и дуплексов, в процессе динамики вначале за относительно небольшое время в результате туннелирования устанавливается неравномерное распределение вероятностей p_n , примерно соответствующее линейному случаю. Далее это распределение изменяется очень медленно, так что можно считать его квазистационарным. При $t \rightarrow \infty$ распределение

стремится к равномерному, что подтверждается расчетом на очень большое время. Пример расчета

периодических дуплексов $GG \begin{pmatrix} A...AGG \\ T...TCC \end{pmatrix}_5$ [3],

число A/T пар $m=1,2,3$:

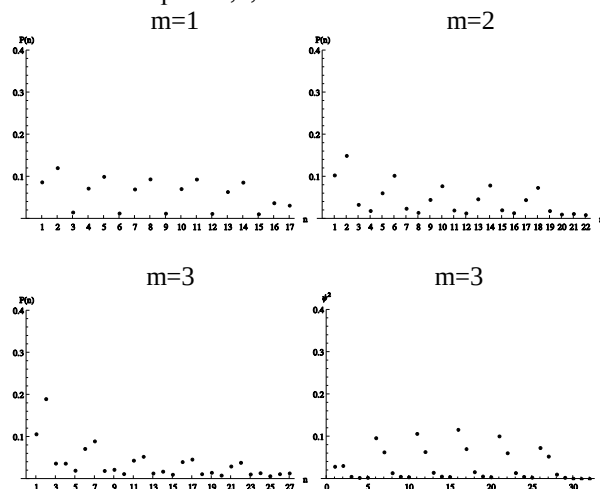


Рис. 4. Квазистационарные распределения.

Первые три Рис.4, $m=1,2,3$ – квазистационарные распределения, четвертый Рис.4, $m=3$ – равномерное распределение при $t \rightarrow \infty$.

4 Заключение

Таким образом, наличие локализованных состояний в спектре периодических гамильтонианов существенно влияет на транспортные свойства ДНК. Дополнительные факторы, такие как, например, сольватация, поскольку эксперименты [3] производились в растворе, могут еще более усилить эффект и даже привести не только к неравномерному распределению заряда, но и к полной локализации дырки на одном сайте [2], [5].

Работа выполнена при поддержке РФФИ, проекты № 10-07-00112-а; № 09-07-12073-офи-М.

Литература

- [1] В. Д. Лахно, В. Б. Султанов. Распределение заряда в ДНК при конечной температуре. ICMBB-2, тезисы, Пушино, с. 24-25, 2008.
- [2] В. Д. Лахно, В. Б. Султанов. О возможности сверхбыстрого переноса заряда в ДНК. Математическая биология и биоинформатика, т. 4, № 2, с. 46-51, 2009.
- [3] C.-S. Liu, R. Hernandez, G. B. Schuster. JACS, v. 126, p. 2877, 2004.
- [4] В. Д. Лахно, В. Б. Султанов. Перенос заряда в одно- и двухцепочечных ДНК. ICMBB-1, тезисы, Пушино, с. 33-34, 2006.
- [5] A. A. Voityuk. J. Chem. Phys., v. 122, p. 204904, 2005.

Развал полярона в однородном полинуклеотиде под влиянием температуры термостата

Лакно В.Д., Фиалко Н.С.

Учреждение Российской академии наук
Институт математических проблем биологии РАН
lak@impb.psn.ru, fialka@impb.psn.ru

Disrupting of polaron in homogeneous polynucleotide by temperature fluctuations

V.D. Lakhno, N.S. Fialko
Institute of Mathematical Problems of biology RAS

Abstract

One of possible mechanisms of charge transfer in DNA is a polaron (or soliton) mechanism. Using computational experiments, we have studied stationary polarons (the state with minimum energy) in homogeneous polynucleotide, and behaviour of them at temperature fluctuations of thermostat. The temperature of polaron disruption was estimated. It is shown that at "biological" temperatures the solvation accounting lead to formation of stable polarons.

Аннотация

В связи с возможностью использования фрагментов ДНК в нанобиоэлектронике возникает вопрос о проводящих свойствах ДНК. Одним из возможных механизмов переноса заряда в биополимерах является поляронный, или солитонный. С помощью численных экспериментов мы исследовали в однородном полинуклеотиде неподвижные поляроны, и их поведение под действием температурных флуктуаций окружающей среды. Сделана оценка температуры развала полярона в цепочках с дисперсией и сольватацией. Показано, что при «биологических» температурах эффект сольватации приводит к образованию устойчивого полярона.

1 Описание модели

Динамика распространения заряда в синтезированных (ГЦ)-полинуклеотидных ДНК (Г – гуанин, Ц – цитозин) исследуется на основе квантово-классической модели (дискретной модели Холстейна). При моделировании процесса переноса мы рассматриваем нуклеотидную цепочку как систему сайтов. Движения этой системы (и квантовые, и классические) описываются системой связанных дифференциальных уравнений. Для моделирования температуры термостата в классические уравнения движения сайтов добавляется случайная сила со специальными статистическими свойствами (уравнения Ланжевена). Система уравнений динамики заряда в

однородной цепочке из N сайтов в безразмерной форме имеет вид:

$$i\dot{b}_n = \eta(b_{n+1} + b_{n-1}) + \chi u_n b_n - \kappa |b_n|^2 b_n, \quad (1)$$

$$\ddot{u}_n = -(\omega^2 + 2\xi)u_n + \xi(u_{n+1} + u_{n-1}) - \chi |b_n|^2 + \omega' u_n + A_n(t), \quad n=1, \dots, N. \quad (2)$$

Здесь переменные b_n – амплитуды вероятности нахождения заряда на n -ом сайте, u_n – смещения сайтов из равновесного невозмущенного положения. Значения параметров для однородной гуаниновой цепочки: матричные элементы перехода заряда между сайтами $\eta = 1.276$ (в размерных величинах 0.084 эВ), константа взаимодействия квантовой и классической подсистем $\chi = 0.02$ (0.13 эВ/Å), коэффициент $\kappa = 15.5$ (1.04 эВ) учитывает сольватационный эффект; частота собственных колебаний сайтов $\omega = 0.01$ (10^{12} сек⁻¹), дисперсия $\xi = 0.000064$ (0.04 эВ/Å²). Температура окружающей среды моделируется в классической подсистеме (2) членом с трением $\omega' = 0.006$ ($6 \cdot 10^{11}$ сек⁻¹) и случайной силой $A_n(t)$, среднее квадратичное отклонение которой зависит от температуры термостата. Подробнее о выводе уравнений и величине параметров для нуклеотидных фрагментов см. [1,2].

2 Вычислительный эксперимент

Мы рассматривали 4 набора параметров.

1 случай – цепочка без учета сольватации, грубая модель «сухой» ДНК.

1.а. $\xi = 0.$, $\kappa = 0$ (простейший случай)

1.б. $\xi = 0.000064$, $\kappa = 0$ (цепочка с дисперсией)

2 случай – цепочка с сольватацией, грубая модель «мокрой» ДНК.

2.а. $\xi = 0.$, $\kappa = 15.5$ (дисперсии нет)

2.б. $\xi = 0.000064$, $\kappa = 15.5$.

При исследовании устойчивости полярона численно решается задача Коши. Задаются начальные данные – полярон в цепочке без учета температуры, и идет расчет нескольких траекторий (с разной случайной силой) на большом временном интервале. Для заданной температуры мы рассчитывали 20 реализаций на времена до $t = 10^6$ (что соответствует 10 нсек).

Результаты расчетов

В 1 случае температура, при которой полярон разрушается, очень низкая, $T < 20$ К. До разрушения полярон движется по цепочке как единое целое. При этом он «дышит» – становится то шире и ниже, то выше и уже.

1.а. При $T = 15$ К развала не обнаружено, при $T = 20$ К развал произошел в 2 реализациях, при $t \sim 250000$ (2.5 нсек) и при $t \sim 550000$ (5.5 нсек). Отметим, что в проведенных ранее расчетах [3] для 3 реализаций на временном интервале $t = 50000$ при $T = 20$ К развала не наблюдалось, а для $T = 25$ К полярон развалился на времени $t < 30000$ (300 псек). Новые результаты уточняют предыдущие.

1.б. Полярон разрушается уже при $T = 10$ К. Так как дебаевская температура для такой цепочки $\Xi \sim 8$ К, а рассматриваемая модель верна для $T > \Xi$, то для более низких температур мы расчетов не проводили.

Во 2 случае мы учитываем сольватацию заряда, коэффициент которой очень большой, порядка 1 эВ. Результат – при «биологической» $T \sim 300$ К развала полярона не обнаружено. Формально можно проводить расчеты и для температур выше температуры денатурации ДНК. Найденная таким образом температура разрушения полярона значительно больше $T = 300$ К:

2.а. Полярон разрушается при $T = 500$ К.

2.б. Полярон разрушается при $T > 1000$ К.

Обсуждение результатов

В 1 случае до развала полярона волна двигалась по цепочке как целое, максимум пика со временем постепенно становился все ниже, от полярона «отщеплялись» и размазывались по цепочке маленькие волны. Во 2 случае развал выглядел иначе. До разрушения полярон стоит на одном месте. Максимум пика слегка колеблется по времени. Затем за короткий промежуток (< 2 псек) полярон распадается на несколько волн, которые разрушаются за ~ 10 псек.

В таблице ниже приведены величины энергии полярона E для рассмотренных 4 вариантов, температура развала T и ее характерная энергия $k_B T$.

вариант	E , эВ	T , К	$k_B T$, эВ
1.а	0.0205	$15 < T < 20$	$1.29 \cdot 10^{-3} - 1.72 \cdot 10^{-3}$
1.б	0.0133	$T < 10$	$8.62 \cdot 10^{-4}$
2.а	0.467	500	0.043
2.б	0.409	$1000 < T$	0.086

Результаты для цепочки без сольватации можно объяснить на основе этой таблицы. Энергия

полярона в случае 1.а почти в полтора раза больше, чем в 1.б. Поэтому предположение, что полярон в цепочке с дисперсией 1.б разваливается при более низкой температуре (либо – при одинаковой температуре разрушается быстрее, чем в цепочке без дисперсии), выглядит довольно очевидным и разумным.

Начальные данные в 1 случае – самогласованное образование, когда на сайтах с большей вероятностью заряда есть бо́льшие смещения. При разрушении волны видно, как «поляронные» смещения сайтов уменьшаются, приближаясь к колебаниям, вызванным температурными толчками. Для случая с дисперсией волна более пологая, начальные смещения, соответствующие полюрону, меньше. Соответственно, чтобы «вытолкнуть», разрушить такую волну, хватит толчков меньшей силы.

Для случая 2 цепочки с сольватацией полученные результаты менее очевидны – энергия полярона в цепочке с дисперсией 2.б меньше, однако он более устойчив к температурным толчкам, и разрушается при большей температуре, чем полярон в цепочке без дисперсии 2.а. Отметим также, что в случае 2 при высоких температурах ($T \geq 300$ К) смещения сайтов не образуют пика там, где локализован заряд, а близки к равновесным по всей цепочке. Т.е. воздействие квантовой частицы на классическую систему слабее, чем температурные толчки, основную роль в образовании полярона играет сольватация, коэффициент которой очень велик.

По результатам моделирования можно также предположить следующее. Для «биологических» температур $T \leq 350$ К в «сухой» ДНК поляроны быстро разрушаются (или не образуются), а в «мокрой» они есть, и в одном и том же полинуклеотиде величина подвижности заряда будет сильно различаться (возможно, на порядки) в зависимости от того, находится этот фрагмент ДНК в воде или в воздухе.

Работа выполнена при поддержке РФФИ, гранты 09-07-12073, 10-07-00112. Расчеты частично проводились в МСЦ РАН, кластер MVS100k.

Литература

- [1] В.Д.Лахно, Н.С.Фиалко. Подвижность дырок в однородной нуклеотидной цепочке *Письма в ЖЭТФ*, том 78, №5, с. 786-788, 2003.
- [2] В.Д.Лахно, Н.С.Фиалко. Динамика переноса заряда вдоль олигонуклеотида при конечной температуре. *Биофизика*, том 49, №1, с. 8-12, 2004.
- [3] В.Д. Лахно, Н.С. Фиалко. Температурный развал солитона в ДНК. *I Международная конференция "Математическая биология и биоинформатика"*. Пушино: материалы конференции, с.27-28, 2006.

О распараллеливании вычислений в задаче моделирования динамики переноса заряда в ДНК при конечной температуре

Лахно В.Д., Соболев Е.В., Фиалко Н.С.

Учреждение Российской академии наук
Институт математических проблем биологии РАН
lak@impb.psn.ru, egor@impb.psn.ru, fialka@impb.psn.ru

Parallelization of calculations in task of modeling of charge transfer in DNA at prescribed temperature V.D. Lakhno, E.V. Sobolev,

N.S. Fialko

Institute of Mathematical Problems of biology
RAS

Abstract

Discrete models often are applied for modeling of excitation dynamics in quasi-one-dimensional biomacromolecules. If the nearest neighbor approximation is used to site chain describing, that for numerical integration of Cauchy problem a parallelization “by sites” may be helpful.

Аннотация

При моделировании процессов распространения возбуждения в квазиодномерных биомакромолекулах часто используются дискретные модели. Если в цепочке сайтов применяется приближение ближайших соседей, то при численном интегрировании задачи Коши для длинных цепочек становится выгодным распараллеливание «по сайтам» в одной реализации.

Мы исследовали динамику распространения заряда вдоль фрагментов ДНК заданного состава при конечной температуре окружающей среды. На основе модели Холстейна разработана дискретная квантово-классическая модель переноса заряда вдоль фрагментов ДНК. В ней заряд (электрон или дырка) рассматривается как квантовая (нерелятивистская) частица, и ее движение описывается уравнением Шредингера, а нуклеотиды (масса которых на три порядка больше) – как

классические объекты, которые описываются уравнениями классической механики. Модель учитывает сольватационные эффекты и дисперсию энергии в классической цепочке (стэкинговое взаимодействие нуклеотидов) в процессе переноса заряда. Для учета температуры окружающей раствора в уравнения модели добавлена случайная сила, имитирующая тепловые толчки окружающей среды. Миграция заряда вдоль фрагментов ДНК с заданной последовательностью исследуется с помощью вычислительных экспериментов. Система уравнений в безразмерной форме имеет вид:

$$\dot{ib}_n = \eta(b_{n+1} + b_{n-1}) + \chi u_n b_n - \kappa |b_n|^2 b_n, \quad (1)$$

$$\ddot{u}_n = -(\omega^2 + 2\xi)u_n + \xi(u_{n+1} + u_{n-1}) - \chi |b_n|^2 + \omega' u_n + A_n(t), \quad n=1, \dots, N. \quad (2)$$

Здесь переменные b_n и u_n , случайная сила $A_n(t)$ имитирует тепловые флуктуации термостата, N – длина моделируемой цепочки.

При моделировании динамики переноса проводится расчет множества реализаций (для каждой – своя сгенерированная последовательность случайных величин и начальные значения переменных); результатом являются усредненные по реализациям траектории движения. Такая постановка задачи допускает естественное распараллеливание с выполнением каждой реализации на отдельном вычислительном узле. Эффективность распараллеливания по реализациям почти 100%.

Для уменьшения времени расчета отдельных реализаций выполнено распараллеливание на узле, содержащем многоядерные процессоры, при расчете каждой реализации с использованием общей памяти и среды openMP.

Хотя при численном решении системы дифференциальных уравнений используется последовательный метод, в уравнения системы (1,2), описывающие состояния сайтов цепочки, явно входят только ближайшие соседи. Поэтому цепочка «разбивается» на более короткие части, которые интегрируются на шаге независимо на разных ядрах

узла; при этом только для крайних сайтов необходимы данные, рассчитанные другими процессами. Все процессы должны переходить к выполнению следующей итерации синхронно; на операцию синхронизации также затрачивается время.

Конечно, при этом способе чем длиннее цепочка сайтов, тем меньше относительная часть обмениваемых данных по сравнению с расчетной, и тем выше ускорение от распараллеливания $T1/Tr$ ($T1$ – время выполнения «последовательного» варианта задачи, Tr – время выполнения задачи, распараллеленной с помощью openMP на p потоков).

Нами были написаны программы на языке Си с использованием MPI и openMP. Программы протестированы на суперкомпьютере МСЦ РАН MVS-100k. Ускорение счета одной реализации на 4-ядерном узле (на MVS100k 4-ядерные процессоры, поэтому «удобное» число потоков 4) сильно зависит от длины цепочки (см. рис.1).

Эффективность распараллеливания одной реализации на узле возрастает при увеличении длины моделируемого фрагмента ДНК (при этом соответственно, уменьшается отношение обменных операций к расчетным). Для коротких цепочек такое распараллеливание невыгодно – при $N=8$ параллельная версия работает даже медленней, чем последовательная ($p=1$). Для фрагментов ДНК длиной 50-100 сайтов (нуклеотидных пар) $T1/Tr \sim 2.4$; для последовательностей из 1000 сайтов и больше $T1/Tr \sim 3.3$.

Мы использовали параллельную openMP-версию программы, в частности, для расчета отдельных

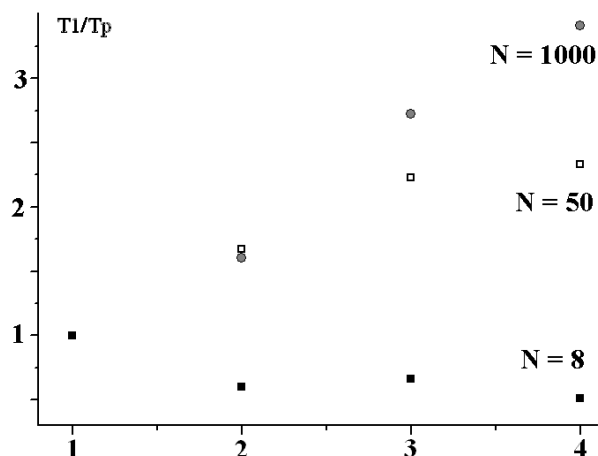


Рисунок 1. По горизонтали – число ядер в узле, по вертикальной оси – ускорение $T1/Tr$. Приведены значения ускорения $T1/Tr$, p – число параллельных потоков на процессоре ($p = 2, 3, 4$), для последовательностей разной длины: 8 (черные точки), 50(белые) и 1000(серые) сайтов.

реализаций при исследовании устойчивости полярона к температурным флуктуациям [Ляхно В.Д., Фиалко Н.С. Развал полярона в однородном полинуклеотиде под влиянием температуры термостата. III Международная конференция "Математическая биология и биоинформатика". Пушкино, Россия, 2010].

Работа выполнена при поддержке РФФИ, гранты 09-07-12073, 10-07-00112. Расчеты частично проводились в МСЦ РАН, кластер MVS100k.

Управляемая модуляция тока во фрагментах двухцепочечной ДНК

О. А. Пономарёв, А. С. Шигаев, В. Д. Лахно

Институт Математических проблем биологии РАН

olegpon36@mail.ru, shials@rambler.ru, lak@impb.ru

В работе изучаются свойства неоднородных фрагментов ДНК, специальным образом подключенных к системе потенциалов, что позволяет подробно исследовать свойства нанофрагментов ДНК. Принципиальная схема таких устройств состоит из цепочки ДНК и двух (или нескольких) локальных центров, один из которых расположен в окрестности узла n_1 , а второй – в окрестности узла n_2 цепочки ДНК, находящихся под действием специально подобранных потенциалов. Концы фрагмента ДНК находятся под постоянной разностью потенциала U . Фрагмент ДНК описывается моделью Пирарда-Бишопа. Предлагаемая схема описывается гамильтонианом

$$H = B \sum (a_n^+ a_{n+1} + a_{n+1}^+ a_n) + \sum (\varepsilon_n + A_n y_n + e \Delta U_n) a_n^+ a_n + (e U_1 + \gamma_{n_1 p_1}) a_{n_1}^+ a_{n_1} c_{p_1}^+ c_{p_1} + B_1 (c_{p_1}^+ c_{p_1+1} + c_{p_1+1}^+ c_{p_1}) + (e U_2 + \gamma_{n_2 p_2}) a_{n_2}^+ a_{n_2} c_{p_2}^+ c_{p_2} + B_2 (c_{p_2}^+ c_{p_2+1} + c_{p_2+1}^+ c_{p_2}) + \sum \left(\frac{P_n^2}{2m} + D_n [(e^{-\alpha_n \sqrt{2} y_n} - 1)^2 - 1] + \frac{k}{2} (y_n - y_{n+1})^2 \right) \text{Здесь}$$

a_n, a_n^+ - операторы электронов (или дырок) осуществляющих проводимость, $c_{p_i}, c_{p_i}^+$ - электронные операторы управляемых или управляющих центров, P_n, y_n - импульсы и координаты узлов соответственно, $y_0 = y_N = 0$, остальные величины являются параметрами.

Ограниченность системы ($N \leq 20$) позволяет численно решить колебательную часть задачи

$$\ddot{y}_n = \frac{1}{m} [-A_n a_n^+ a_n + 2\sqrt{2} D_n \alpha_n (e^{-2\sqrt{2} \alpha_n y_n} - e^{-\sqrt{2} \alpha_n y_n}) + k(y_{n-1} + y_{n+1} - 2y_n)] \quad (1)$$

с граничными условиями $y_0 = y_N = 0$ и начальными условиями $y_n(0) = \langle y_n \rangle, \dot{y}_n(0) = 0$. Из (1) видно, что проявление носителей тока приводит к сдвигу потенциальных кривых, не меняя частот ангармонических колебаний. Фиксация концов фрагментов ДНК электродами препятствует

увеличению расстояния между цепочками и разрыву Н-связи. Нелинейная Н-связь обычно слабее гармонической связи между узлами. В рассматриваемой модели невозможно разорвать спираль ДНК из-за граничных условий. «Пузырь» может образоваться на ограниченном пространстве, где граничные условия позволяют достаточное увеличение $y_n(t)$, но его образование не будет резким, похожим на фазовый переход. Два комплементарных нуклеотида, находясь даже на значительном расстоянии, не колеблются независимо, а включают энергию Н-связи в энергию собственного колебания как влияние внешнего поля, несколько увеличивая частоту колебания. Колебания не имеют мягкой моды и разрыв Н-связи происходит постепенно при увеличении $y_n(t)$. Колебательная задача (1) решена численно и аналитически в квазигармоническом приближении для $N = 4$. В последнем случае получили три собственных частоты

$$\omega_1 = \left(\frac{f_1 + 2k}{m} \right)^{\frac{1}{2}},$$

$$\omega_{2,3} = \left(\frac{f_1 + f_2 + 4k}{2m} \pm \frac{1}{2m} [(f_1 - f_2)^2 + 8k^2]^{\frac{1}{2}} \right)^{\frac{1}{2}} \text{ где}$$

$$f_n = - \frac{2\sqrt{2} D_n \alpha_n}{\langle y_n \rangle} (e^{-2\sqrt{2} \alpha_n \langle y_n \rangle} - e^{-\sqrt{2} \alpha_n \langle y_n \rangle}). \text{ Ни одна из}$$

собственных частот не становится мягкой, хотя и зависит от величины Н-связей.

Для N узлов имеем набор частот $\omega_1, \omega_2, \dots, \omega_{N-1}$, а смещения разлагаются по собственным функциям и имеют вид $y_n(t) = \sum C_{nq} \cos \omega_q t$, где C_{nq} и ω_q - известные функции, полученные в результате решения колебательной части задачи или определяются из опыта.

Введем эффективный гамильтониан для электронной подсистемы заменив колебательные координаты их значениями, считая, что колебательная подсистема влияет на электронную как внешнее переменное поле. После этого можно приступить к изучению работы управляющих центров. Центр с координатами p_1 будет

управляющим. Он приводится в необходимое состояние управляющим внешним воздействием. Центр с координатами p_2 будет управляемым, принимающим сигналы от цепочки ДНК и будет отражать свойства и строение этой цепочки. Подберем строение центров и параметры B_1 и B_2 так, чтобы время изменения состояния примесных центров было сравнимо со временем эксперимента. Тогда эти изменения можно будет измерять и проводить сопоставления с изменениями в системе ДНК. Используя эффективный гамильтониан можно получить, что заселенность точки p_1 равная единице в начальный момент изменяется со временем по закону

$$\langle c_{p_1}^+ c_{p_1} \rangle = \cos^2 \left(\frac{(eU_{1p_1} + \gamma_{n_1 p_1} \langle a_{n_1}^+ a_{n_1} \rangle)^2}{4} + B_1^2 \right)^{\frac{1}{2}} t,$$

что позволяет управлять частотой и величиной заселенности с помощью изменения потенциала U_1 .

Регистрирующий центр p_2 в начальный момент находится в равновесном состоянии и его изменение во времени зависит от состояния локального равновесия и имеет вид

$$\begin{aligned} \langle c_{p_2}^+ c_{p_2} \rangle = & \frac{1}{2} \cos^2 \left(\frac{(\gamma_{n_2 p_2} \langle a_{n_2}^+ a_{n_2} \rangle - eU_{2p_2})^2}{4} + B_2^2 \right)^{\frac{1}{2}} t \\ & + \frac{(\gamma_{n_2 p_2} \langle a_{n_2}^+ a_{n_2} \rangle - eU_{2p_2} + 2B_2)^2}{2((\gamma_{n_2 p_2} \langle a_{n_2}^+ a_{n_2} \rangle - eU_{2p_2})^2 + 4B_2^2)} \\ & \sin^2 \left(\frac{(\gamma_{n_2 p_2} \langle a_{n_2}^+ a_{n_2} \rangle - eU_{2p_2})^2}{4} + B_2^2 \right)^{\frac{1}{2}} t \end{aligned}$$

Здесь $\langle a_n^+ a_n \rangle$ - равновесная заселенность с учетом изменения потенциала вдоль цепочки ДНК.

Проведем расчет токовой функции $G_{n,n+1}(t) = \langle a_n^+(t) a_{n+1}(t) a_{n+1}^+(t+\tau) a_n(t+\tau) \rangle$ из эффективного гамильтониана с учетом явной зависимости от параметров управляющих и регистрирующих центров и в результате получаем в окрестности примесных управляющих центров

$$\begin{aligned} G_{n,n+1}(\tau, t) = & \langle a_{n_1}^+ a_{n_1} \rangle (1 - \langle a_{n_1+1}^+ a_{n_1+1} \rangle) \exp \{ i(\varepsilon_{n_1+1} \\ & + e\Delta U_{n_1+1} - \varepsilon_{n_1} - e\Delta U_{n_1})t - i\gamma_{n_1 p_1} \langle c_{p_1}^+(\tau) c_{p_1}(\tau) \rangle \tau > t \\ & + i \sum (A_{n_1+1} C_{n_1+1,q} - A_{n_1} C_{n_1,q}) \frac{1}{\omega_q} \sin \omega_q t \}. \end{aligned}$$

Для других локальных токовых функций получаются аналогичные выражения. Рассмотрено два случая: 1) когда из суммы по гармоникам главную роль играет одна частота и можно сделать разложение по функциям Бесселя и 2) когда можно сделать разложение синуса по времени. В этом последнем случае локальная проводимость между узлами равна

$\sigma_n(\tau) = e^2 a^2 B^2 \operatorname{Re} \int_0^\infty e^{-\alpha t} G_{n,n+1}(t) dt$, что приводит к локальному сопротивлению

$$\rho_{n,n+1}(\tau) = \sigma_{n,n+1}^{-1}(\tau) =$$

$$\frac{\alpha^2 + (e\Delta U_{n_1+1} - e\Delta U_{n_1} + \sum (A_{n_1+1} C_{n_1+1,q} - A_{n_1} C_{n_1,q}) - \Gamma_{n_1 p_1})^2}{\alpha(eaB)^2 \langle a_{n_1}^+ a_{n_1} \rangle (1 - \langle a_{n_1+1}^+ a_{n_1+1} \rangle)}, \quad \text{где}$$

$\Gamma_{n_1 p_1} = \varepsilon_{n_1+1} - \varepsilon_{n_1} - \gamma_{n_1 p_1} \langle c_{p_1}^+(\tau) c_{p_1}(\tau) \rangle$. Если подать постоянное напряжение на концы цепочки ДНК, то по ней пойдет переменный ток, равный разности потенциалов деленной на сопротивление цепочки

$$\begin{aligned} R(\tau) = & R(1, n_1 - 1) + \rho_{n_1-1, n_1}(\tau) + \rho_{n_1, n_1+1}(\tau) + R(n_1 + 1, n_2 - 1) + \\ & + \rho_{n_2-1, n_2}(\tau) + \rho_{n_2, n_2+1}(\tau) + R(n_2 + 1, N), \quad R(n, n_i) - \\ & \text{сопротивление участка цепи между } n \text{ и } n_i \text{ узлами.} \end{aligned}$$

Поэтому $I(\tau) = \frac{U}{R(\tau)}$, $\Delta U_n = I(\tau) \sum_1^n \rho_{n-1, n}$

$$\langle a_{n_2}^+ a_{n_2} \rangle = n_0 + I(\tau) \langle t \rangle / e = n_0 + I(\tau) / e\alpha.$$

Для упрощения выражений положим, что все ε_n одинаковы, кроме ε_{n_1} и ε_{n_2} ; $\Delta U_n = 0$, $\langle a_n^+ a_n \rangle = n_0, n \neq n_2$. Тогда имеем

$$I(\tau) = \frac{U\alpha(eaB)^2 n_0(1 - n_0)}{(N-1)\alpha^2 + 2(\varepsilon - \varepsilon_{n_1} - \gamma_{n_1 p_1} \langle c_{p_1}^+ c_{p_1} \rangle)^2},$$

$$\langle c_{p_1}^+ c_{p_1} \rangle = \cos^2 B_1 \tau \quad \text{при } eU_{1p_1} + \gamma_{n_1 p_1} n_0 = 0;$$

$$\begin{aligned} \langle c_{p_2}^+ c_{p_2} \rangle = & \frac{1}{2} \cos^2 \left(\frac{(\gamma_{n_2 p_2} I(\tau))^2}{4\alpha^2 e^2} + B_2^2 \right)^{\frac{1}{2}} \tau \\ & + \frac{1}{2} \frac{(\gamma_{n_2 p_2} I(\tau) e^{-1} \alpha^{-1} + 2B_2)^2}{(\gamma_{n_2 p_2} I(\tau) e^{-1} \alpha^{-1})^2 + 4B_2^2} \sin^2 \left(\frac{(\gamma_{n_2 p_2} I(\tau))^2}{4\alpha^2 e^2} + B_2^2 \right)^{\frac{1}{2}} \tau, \end{aligned}$$

$$\text{при } eU_{2p_2} - \gamma_{n_2 p_2} n_0 = 0.$$

Таким образом, в этой работе показано: во первых, возможность создания электронных устройств на основе единичной молекулы ДНК; во вторых, перспективность использования полевых эффектов для переключающих, запоминающих, усиливающих элементов и, в третьих, необходимость создания основ теории для получения микросхем нового типа основанных на самосборке. В дальнейшем предполагается развивать исследования в направлении изучения физических принципов функционирования наноприборов, выбор новых молекул для последующей самосборки приборов, разработка теории механизмов самосборки

Работа поддержана грантами РФФИ (проекты № 09-07-12073 офи-м; 10-07-00112).

Варианты нарушения структуры двухцепочечной ДНК в областях её локального расплетания

А. С. Шигаев, О. А. Пономарёв, В. Д. Лахно

Институт Математических проблем биологии РАН

shials@rambler.ru, lak@impb.ru

Процесс разрушения двойной спирали ДНК и разделения её на две отдельные цепи называют денатурацией или плавлением, так как самый распространённый способ денатурации ДНК – нагревание её раствора. Температура плавления ДНК зависит от её нуклеотидного состава, pH, ионной силы, и других факторов. Однако даже при значительно меньших температурах возможно образование областей локальной кратковременной денатурации дуплекса – так называемых пузырьков денатурации. Эти области имеют большое значение для функционирования генома в живой клетке. Помимо изучения роли пузырьков *in vivo*, проблема исследования их природы и поведения может являться важной составной частью проблемы переноса электрического заряда в ДНК. Локальное расплетание двойной спирали тесно связано с нарушением стэкинг-взаимодействий между основаниями, поддерживающих перекрывание их π -орбиталей. Изменение перекрывания определяет задержку миграции заряда; кроме того, может повышаться вероятность взаимодействия заряженных атомов с молекулами раствора.

В нашей работе проведён анализ литературных данных по исследованиям поведения пузырьков денатурации. На данный момент не существует ни одного экспериментального метода, который позволил бы наблюдать динамику пузырьков напрямую. Любой из имеющихся методов может дать лишь косвенную информацию о локальной кратковременной денатурации ДНК. Наиболее распространённый метод основан на гиперхромном эффекте – усилении поглощения раствором ДНК света с длиной волны 260 нм. Поглощение увеличивается в результате разрушения стэкинг-взаимодействий, которое, по данным метода, начинается при температурах 45 - 50 °C [1]. Более современным методом исследования локального плавления ДНК является корреляционная флуоресцентная спектроскопия (КФС). Метод основан на отслеживании задержки тушения флуоресценции в результате временного удаления друг от друга молекул возбуждённой светом метки и тушителя, пришитых на разные цепи ДНК. Время задержки отражает время открытого состояния ДНК. В данном методе это время следует понимать как период, в течение которого на участке дуплекса, где пришиты хромофор и тушитель, разорваны водородные связи. В отличие от данных,

полученных при помощи УФ-поглощения, по данным КФС пузырьки денатурации могут появляться уже при 30 °C [2]. При этом вероятность открытого состояния ДНК составляет примерно 0.5%, доходя до 2-4% при 40 °C. Очевидно, что помимо разной чувствительности методов, в основе такого сильного различия может лежать следующая ситуация: при разрыве водородных связей в дуплексе происходит локальное расхождение цепей ДНК без нарушения стэкинг-взаимодействий. Водородные связи между комплементарными основаниями замещаются связями с молекулами воды, попадающими внутрь спирали. Вследствие этого может образовываться устойчивая альтернативная система водородных связей, способная существовать десятки микросекунд (таково время релаксации открытых состояний ДНК по данным КФС).

В пользу этой догадки могут свидетельствовать новые данные о природе водородной связи. На основе данных О. Mishima и Н. Е. Stanley [3], М. Чаплин в своей монографии "Water Structure and Science" предположил, что эта связь имеет не один, а два энергетических минимума [4]. Менее глубокий из них имеет электростатический характер, соответствует меньшему расстоянию между центрами связи и не зависит от угла между связью О-Н и водородной. Второй (дальний) минимум связан непосредственно с образованием водородной связи: благодаря тому, что она является на 10% ковалентной и узко направленной [5], он очень ограничен в пространственном отношении. В результате такого "строения" связи её разрыв может являться сложным процессом, состоящим из нескольких стадий, каждая из которых требует относительно небольшой энергии. Видимо, здесь можно провести аналогию с термодинамикой ферментативного катализа. Для стэкинг-взаимодействий подобных эффектов нет, так как его энергетический минимум, соответствующий максимальной площади контакта плоскостей оснований, является единственным.

Возможность сохранения стэкинг-взаимодействий при разрушении системы водородных связей (вблизи температуры плавления) продемонстрирована также при помощи метода Рамановской спектроскопии в работах Мовилеану с соавт. [6]. Той же группой исследователей показана

дополнительная стабилизация структуры в АТ-трактах оснований (поли(dA)•поли(dT)) за счёт трёхцентровых водородных связей [7]. На повышенную устойчивость стэкинг-взаимодействий в АТ-трактах указывают также данные ^1H -ЯМР [8]. Исследование динамики дуплекса этим методом основано на способности протона имино-группы аденина или тимина обмениваться на дейтрон в растворе тяжёлой воды. Обмен протона происходит в присутствии катализатора (аммиака) из открытого состояния ДНК, под которым в этих исследованиях обычно понимают "выворот" (flip-out) основания из Уотсон-Криковской спирали. По данным ^1H -ЯМР, при температуре 30 °С средняя вероятность открытого состояния для основания составляет примерно 10^{-5} , а время жизни такого состояния – порядка 10^{-7} с [9]. Оценка времён закрытого состояния по результатам методов КФС и ^1H -ЯМР (для 30 °С) даёт поразительное сходство: 10 - 20 и 8 - 15 мс соответственно. Это указывает на тесную взаимосвязь коллективного разрыва водородных связей и выхода основания из стопки (стэка): первый процесс, возможно, даже является важной предпосылкой второго. Огромное (на три порядка) различие времён жизни открытых состояний по данным этих методов может говорить об очень малой вероятности обмена протона для основания, находящегося в стэке.

Тем не менее, к сравнению результатов, полученных методом ^1H -ЯМР, с результатами других исследований необходимо относиться крайне осторожно. По данным Зенга с соавт. дуплексы, длина которых ниже некоторого критического значения (около 20 пар оснований), расплетаются без образования интермедиатов (пузырьков денатурации) [1]. Длина олигомеров, исследуемых методом ^1H -ЯМР, не превышает 15-17 пар оснований (обычно 10-12). Кроме того, сильное «пропеллерное» искажение, характерное для такого короткого дуплекса, не может не влиять на параметры водородных связей и стэкинг-взаимодействий. С другой стороны, фотовозбуждение хромофора в КФС, видимо, приводит к локальным изменениям структуры связанной с ним ДНК, что может завышать регистрируемую вероятность открытого состояния.

Несмотря на то, что приведённые выше данные указывают на возможность сохранения целостности стэка при локальном разрыве водородных связей в дуплексе, вероятность подобной ситуации, в конечном счете, определяется первичной структурой ДНК. К примеру, методом Рамановской спектроскопии показано, что в полимере (поли(dA-dT)•поли(dA-dT)), каждая цепь которого состоит из чередующихся адениновых и тиминового нуклеотидов, стэкинг-взаимодействия исчезают вместе с разрушением системы водородных связей [6]. Также, по данным некоторых биохимических исследований (гидролиз эндонуклеазой S1), вероятность образования

крупных (10 и более пар оснований) пузырьков денатурации в различных областях гетерогенных ДНК может отличаться в десятки раз [10]. Это может свидетельствовать о широком диапазоне степени нарушения стэка в пузырьках различной длины и первичной структуры.

В докладе обсуждаются методы исследования пузырьков денатурации (включая приведённые выше), а также возможность количественной оценки времён жизни открытого состояния из данных по гидролизу олигомеров ДНК эндонуклеазой S1.

Работа поддержана грантами РФФИ (проекты №№ 09-07-12073 офи-м; 10-07-00112).

Литература

- [1] Y. Zeng, A. Montrichok, G. R. Zocchi, Bubble nucleation and cooperativity in DNA melting. *Journal of Molecular Biology*, v. 339, pp. 67–75, 2004.
- [2] G. Altan-Bonnet, A. Libchaber and O. Krichinsky, Bubble Dynamics in Double-Stranded DNA, *Physical Review Letters*, v. 90, p. 138101 (4 стр.), 2003.
- [3] O. Mishima and H. E. Stanley, The relationship between liquid, supercooled and glassy water. *Nature*, v. 396, pp. 329–335, 1998.
- [4] <http://www1.lsbu.ac.uk/water/abstrct.html>.
- [5] E. D. Isaacs, A. Shukla, P. M. Platzman, D. R. Hamann, B. Barbiellini and C. A. Tulk, Compton scattering evidence for covalency of the hydrogen bond in ice. *Journal of Physics and Chemistry of Solids*, v. 61, pp. 403–406, 2000.
- [6] L. Movileanu, J. M. Benevides, G. J. Thomas Jr., Temperature Dependence of the Raman Spectrum of DNA. II. Raman Signatures of Premelting and Melting Transitions of Poly(dA)•Poly(dT) and Comparison with Poly(dA-dT)•Poly(dA-dT). *Biopolymers*, v. 63, pp. 181–194, 2002.
- [7] L. Movileanu, J. M. Benevides, G. J. Thomas Jr., Determination of base and backbone contributions to the thermodynamics of premelting and melting transitions in B DNA. *Nucleic Acids Research*, v. 30, pp. 3767–3777, 2002.
- [8] U. Dornberger, M. Leijon, and H. Fritzsche, High Base Pair Opening Rates in Tracts of GC Base Pairs, *The Journal of Biological Chemistry*, v. 274, pp. 6957–6962, 1999.
- [9] D. Coman and I. M. Russu, A Nuclear Magnetic Resonance Investigation of the Energetics of Basepair Opening Pathways in DNA, *Biophysical Journal*, v. 89, pp. 3285–3292, 2005.
- [10] C. H. Choi, G. Kalosakas, K. O. Rasmussen, M. Hiromura, A. R. Bishop and A. Usheva, DNA dynamically directs its own transcription initiation, *Nucleic Acids Research*, v. 32, pp. 1584–1590, 2004.

Nanobiomaterials and Bioferroelectricity: Molecular Modeling and Nanoscale Characterization

V.S. Bystrov^(1,2), Yu. Dekhtyar⁽³⁾, E.V. Paramonova⁽¹⁾, V.M. Fridkin⁽⁴⁾, H. Kliem⁽⁵⁾, A. Kholkin⁽²⁾

Institute of Mathematical Problems of Biology RAS, Pushchino, Russia⁽¹⁾,
Dept. of Ceramics and Glass Engineering & CICECO, University of Aveiro, Aveiro, Portugal⁽²⁾,
Inst. of Biomedical Eng. & Nanotechnology, Riga Technical University, Riga, Latvia⁽³⁾,
Institute of Crystallography, RAS, Moscow, Russia⁽⁴⁾,
Institute of Electrical Engineering Physics, Saarland University, Saarbruecken, Germany⁽⁵⁾

vsbys@mail.ru

bystrov@ua.pt

Abstract

The review of recent studies on molecular modeling (including molecular dynamics run using HyperChem etc.) and characterization on nanoscale level (by AFM/PFM technique, etc.) for series of novel nanobiomaterials and nanostructures were presented in this paper. Special point is devoted to bioferroelectricity.

1 Objects

The main attention is devoted to the following molecular nanostructures: 1) ultra-thin ferroelectric polymer Langmuir-Blodgett (LB) films on the base of the polyvinylidene fluoride P(VDF-TrFE) and its nano-composite with phospholipids (1,2-Di-O-hexadecyl-*sn*-glycero-3-phosphocholine); 2) Hydroxyapatite (HAP) nanoparticles (NP) and its nano-composite with PVDF; 3) diphenylalanine (Phe-Phe) structure and its self-assembles into well-ordered di-peptide-based nanotubes (PNT). The influence of proton subsystem and hydrogen-bonds (HB) on the stable nanostructure formation are studied.

2 Methods

The molecular modelling of all molecular structures is done using HyperChem 7.52 and 8.0 tool for different methods, including molecular dynamics (MD) run, with various molecular mechanical (MM, OPLS, BIO CHARM) and first principle quantum-chemical semi-empirical (PM3) methods. Some parts of the data (precise energy calculations, particularly, for hydrogen bonds variations) were computed by Gaussian98 (using HF/6-31G(d,p) and B3LYP6-31G(d,p) basis).

Additional experimental studies were made using the photoelectron spectroscopy measurements for determining of the HAP NP work function and the atomic force microscopy (AFM) and piezoresponse force microscopy (PFM) for testing of the various molecular

component's interactions: the PVDF films, its composite with phospholipids and self-assembled di-peptides PNT physical properties – piezoelectric contrasts and local piezoresponse hysteresis. The analysis of observed bioferroelectric phenomena is presented.

3 Results

Calculations, using geometry optimization and MD run both MM, OPLS, BIO CHARM and PM3 methods, show the stable self-assembly formation of 6 molecular ring of Phe-Phe, hexagonal PNT structures and dynamical behavior of the hydrogen bonds between molecular components, influence on space orientation of Phe-Phe rings in PNT. Calculations of physical properties (dipole moment, polarization, diameters), energies (total energy, electron energy E_{HOMO} , E_{LUMO} and E_g of the forbidden gap) were carried out. Computed data predict PNT phase transition to orthorhombic phase confirmed by AFM/PFM measurements. Size and shape of the simulated HAP NP cluster has an influence on its dipole momentum, polarization and as result on the electron work function. These parameters strongly depend on concentration of protons at the surface and influence on the aggregation. Computed data allows determining the dipole moment of phospholipids and PVDF piezoelectric coefficients, which have a good agreement with our experimental AFM/PFM data. Calculations using MD run both MM (OPLS) and quantum (PM3) methods show the formation and dynamical behaviours of arisen hydrogen bonds between PVDF chain and phospholipid molecule. Especially calculations were made on the region of the lipid's head interaction with PVDF chain. Obtained data assume to conclude that beyond electrostatic interaction namely the hydrogen bonds play here the key role for stabilizing of self-organized nano-composites and arising of the observed bioferroelectric phenomena.

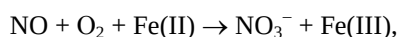
This work is partly supported by DAAD grant 325-A09/03515 and FCT grant SFRH/BPD/22230/2005.

Ab initio расчет активного центра гемма для различных спиновых состояний. Проблема моделирования пассивного и активного состояний бациллы туберкулеза

К.В.Симон, А.В.Тулуб

Санкт-Петербургский государственный университет

Выполнен неэмпирический расчет электронного строения активного центра гемма trHbN для различных строго заданных спиновых состояний. В состав системы входит железопорфирин, молекула кислорода, радикал NO и две молекулы воды. Электронная корреляция учитывается путем выделения двух наиболее существенных подпространств, в пределах которых далее использован метод CASSCF. В результате переноса заряда с железа на фрагмент (NOOO), возникают структуры типа 1) пероксинитрита и 2) нитратного аниона. Защитная реакция бациллы туберкулеза в латентном состоянии основывается на реакции образования нитратного аниона, схематически представляется как



причем нитратный анион в дальнейшем удаляется из области активного центра [1], в этом процессе принимают активное участие присутствующие в полости гемма фрагменты аминокислот.

Предполагается, что указанная реакция в известных условиях может развиваться по другому пути с образованием пероксинитрита как в его обычном основном состоянии, так и в его димерной форме [2]. Выполненный расчет фиксирует лишь возможные конечные продукты реакции, но не путь реакции. Пероксинитрит $(\text{ONOO})^-$ образуется в газовой фазе в результате взаимодействия аниона O_2^- с радикалом NO согласно $\text{NO} + \text{O}_2^- \rightarrow (\text{ONOO})^-$, перегруппировка этого аниона в нитратный анион в газовой фазе не происходит ввиду значительных активационных барьеров. Взаимодействие пероксинитрита с водой в объеме не разрушает его структуру.

Литература

- [1] Marti M. A., Bidon-Chanal A, Crespo A. et al // J. Am. Chem. Soc. 2008, 130, P.1688-1693.
- [2] Симон К. В., Тулуб А. В. // Оптика и Спектроскопия, 2009, Т.107, №1, С.53-60.

Об исследовании нелинейных моделей многостадийного синтеза вещества

Лихошвай В.А.^(1,2), Фадеев С.И.^(2,3), Штокало Д.Н.⁽²⁾

Институт цитологии и генетики СО РАН⁽¹⁾, Новосибирский государственный университет⁽²⁾, Институт математики им. С.Л. Соболева СО РАН⁽³⁾

shtokalod@gmail.com

Abstract

Multi-stage processes of synthesis of DNA, RNA, protein polymers were modeled. Efficient numerical algorithms of integration of related big dimension systems of non-linear differential equations have been developed. Numerical research confirms hypothesis on similarity of properties of non-linear system to quasi-linear system and to delay equation that model the same synthesis process with another kind of resolution level. The results help to extend class of big dimension systems that connected with delay equation and confirm ability of adequate modeling of gene networks without detailed description of DNA, RNA, protein synthesis.

$f_{n,i}^+(x_i)$ – скорость прямого процесса, переводящего вещество со стадии i на стадию $i+1$, $f_{n,i}^-(x_i)$ – скорость возврата со стадии i на $i-1$, $s_{n,i}(x_i)$ – спонтанная терминация синтеза (сток), $g(y)$ – скорость утилизации продукта.

Подобные системы часто встречаются в физике, химии, экономике. В биологии системы подобного типа являются составными частями процессов развития организмов, сетей метаболических реакций, молекулярно-генетических систем [2].

В частности в функционировании живых систем фундаментальную роль играют матричные процессы синтеза ДНК, РНК и белков – нерегулярных линейных полимеров. Данные процессы характеризуются наличием большого количества нелинейных стадий присоединения к растущим полимерам нуклеотидов или аминокислотных остатков. В результате системы обыкновенных дифференциальных уравнений, описывающие матричную стадию синтеза конечного полимера (ДНК, РНК или белка) содержат сотни, тысячи и больше уравнений.

Мы полагаем $f(t, y) = \bar{f}(y) = \alpha / (1 + \beta y^\gamma)$,
 $f_{n,i}^{+/-}(x_i) = n / \tau_i^{+/-} \cdot x_i / (1 + \rho x_i^\sigma)$,
 $s_{n,i}(x_i) = \omega_i x_i$, где $\alpha \geq 0, \beta > 0, \gamma \geq 1$,
 $\tau_i^+ > 0, \tau_i^- > 0, \rho \geq 0, \sigma > 0, \omega_i \geq 0$ – параметры.

1.2 Предельные свойства

Важное место в исследовании данного вопроса занимают т.н. предельные теоремы, согласно которым при достаточно большом числе промежуточных стадий синтеза распределение во времени продукта синтеза $y(t)$, описываемое решением автономной системы, будет близко к

1 Модели многостадийного синтеза

В данной работе рассматриваются численные методы исследования математических моделей динамических систем, содержащих последовательности следующих друг за другом стадий, каждая из которых имеет сложный нелинейный механизм функционирования.

1.1 Система уравнений

Система уравнений, соответствующая нелинейной модели в общем случае имеет вид:

$$\begin{cases} \frac{dx_1}{dt} = f(t, y) - f_{n,1}^+(x_1) + f_{n,2}^-(x_2) - s_{n,1}(x_1), \\ \frac{dx_{i+1}}{dt} = f_{n,i}^+(x_i) - f_{n,i+1}^+(x_{i+1}) - f_{n,i+1}^-(x_{i+1}) \\ \quad + f_{n,i+2}^-(x_{i+2}) - s_{n,i+1}(x_{i+1}), \quad i = 1, n-2, \\ \frac{dx_n}{dt} = f_{n,n-1}^+(x_{n-1}) - f_{n,n}^+(x_n) - f_{n,n}^-(x_n) - s_{n,n}(x_n), \\ \frac{dy}{dt} = f_{n,n}(x_n) - g(y), \end{cases}$$

решению соответствующего уравнения с запаздывающим аргументом

$$dz(t)/dt = e^{-d} f(t - \tau, z(t - \tau)) - g(z(t)),$$

где $\tau > 0$ – запаздывание, соответствующее времени синтеза, $d \geq 0$, e^{-d} – дефект от стоков.

Практически это означает, что свойства решения автономной системы уравнений большой размерности будут такими же, как и свойства соответствующего хорошо изученного уравнения с запаздывающим аргументом. Это позволяет заранее сформулировать представление о предельных состояниях, к которым будет стремиться в зависимости от параметров модели решение задачи Коши.

1.3 Методы исследования

Доказательство предельных теорем [2,3] получено для моделей синтеза с линейным описанием процесса на промежуточных стадиях

$$f_{n,i}^{+/-}(x_i) = n / \tau_i^{+/-} \cdot x_i.$$

Поскольку при этом функция $f(t, y)$, задающая описание инициации синтеза не линейна, то условно такого рода модели синтеза мы будем называть квазилинейными. Однако, если описание процесса на промежуточных стадиях не линейно, то основным способом изучения свойств модели, которую мы будем называть нелинейной, является численный эксперимент. В работе приводятся результаты численного эксперимента, свидетельствующие о том, что нелинейная модель, как и квазилинейная модель, обладает предельным свойством, т.е. определяемое из её решения распределение продукта синтеза будет близко к тому, что дает решение соответствующего уравнения с запаздывающим аргументом.

1.4 Алгоритмы численного интегрирования

Проблемы численного анализа моделей синтеза, обусловленные, в частности, большой размерностью исследуемых систем, потребовали разработки алгоритмов интегрирования, в которых существенно учитывается структура правых частей. Предложенные схемы интегрирования характеризуются гарантированной положительностью решения, устойчивостью, а также экономичностью, благодаря использованию метода прогонки [1]. В работе приведено большое число результатов вычислений в виде графиков, иллюстрирующих достоверность сформулированных гипотез о свойствах нелинейной модели.

В итоге, сочетание качественных методов исследования, на которые в работе имеются указания, и исследования в рамках численного эксперимента позволило получить достаточно полное представление о свойствах рассматриваемых моделей синтеза.

References

- [1] С.К. Годунов, В.С. Рябенкий. Разностные схемы. Москва, Наука, 1973, 400с.
- [2] В.А. Лихошвай, С.И. Фадеев, Г.В. Демиденко, Ю.Г. Матушкин. Моделирование уравнением с запаздывающим аргументом многостадийного синтеза без ветвления // Сиб. журн. индустр. математики. 2004. Т. VII, №1(17). С. 73-94.
- [3] Д.Н. Штокало. О предельном переходе к уравнению с запаздывающим аргументом в модели синтеза вещества с учетом обратимости и стоков. // Сиб. журн. индустр. математики, 12:2 (2009), 143-156.

Quantum mechanics approach to molecular mechanics description of approximate planarity of intrinsically non-planar nucleic acid bases

González E.⁽¹⁾, Deriabina A.S.⁽¹⁾, Poltev V.I.^(1, 2), Furmanchuk A.⁽³⁾, Gorb L.⁽³⁾, Leszczynski J.⁽³⁾

Physics and Mathematics Department, Benemerita Universidad Autonoma de Puebla, 72570 Puebla, Mexico⁽¹⁾, Institute of Theoretical and Experimental Biophysics, 142290 Pushchino, Russia⁽²⁾, Interdisciplinary Nanotoxicity Center, Jackson State University, Jackson, 39217 MS, USA⁽³⁾
gonzalez@fcfm.buap.mx

Abstract

Quantum mechanics calculations were performed to obtain the dependences of interaction energy between the bases of the Watson-Crick DNA pairs when buckle and twist angles were systematically changed. The results of the calculations were used to derive mathematical formulae which can be employed for molecular mechanics description of the base interactions.

1 Introduction and antecedents

Molecular mechanics method is a powerful and unique tool to construct atomic level models of nucleic acid structure and functions. However, existing force fields of molecular mechanics do not describe some details of monomer structures, such as a phenomenon of non-planarity and flexibility of amino groups of the bases. This phenomenon is important for biological functions of DNA, but there is no quantitative experimental data on amino group flexibility to calibrate force field for its description. To fill this breach in the data, we have performed extensive calculations of nucleic acid bases at the MP2/6-31G(d,p) level of *ab initio* theory for broad set of amino group configurations. Two-dimensional maps of energy and geometry characteristics as functions of two amino hydrogen torsions have been constructed [1]. We approximate the maps by polynomial expressions, which can be used as additional terms in molecular mechanics calculations. The strategy is proposed to obtain more accurate formulae using higher order *ab initio* calculations for less extensive set of amino group torsions. As additional term describing amino group flexibility favours the non-planarity of the base pairs, a necessity was revealed of a new term of quantum nature that is responsible for nearly planar arrangements of the bases in the Watson-Crick and many other base pairs.

2 Bi-dimensional energy maps obtained by quantum mechanics

To construct the term mentioned above, the calculations of base-base interaction energy as a function of the propeller twist and buckle angles in the base pairs have been performed by correlated *ab-initio* MP2/6-31G(d,p) method. The resulting energy dependence for guanine-cytosine pair on the two angles is presented in the Fig.1 as an example.

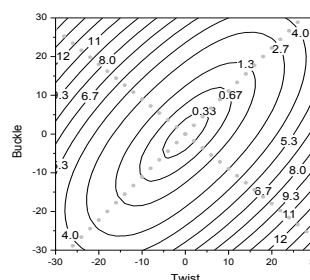


Figure 1. The bi-dimensional map for interaction energy of guanine-cytosine pair, energy is in kcal/mol.

For the same set of twist and buckle angles the calculations with different molecular mechanics methods were performed. The comparison of the results obtained by the two groups of methods enables us to suggest rather simple quadratic expression for the additional energy term. The coefficients in this expression will be adjusted after more extended molecular mechanics and higher level correlated *ab-initio* computations.

References

- [4] Poltev, V.; Gonzalez, E., Deriabina, A., Martinez; A., Furmanchuk, A.; Gorb, L.; Leszczynski J. *J. Biol. Phys.*, 2007, 33, 499-514.

Как клетка исправляет ошибки при самоорганизации веретена деления

Атауллаханов Ф.И.

Центр теоретических проблем физико-химической фармакологии РАН, Москва

fazly@hc.comcor.ru

Одной из важнейших и критических стадий деления клетки является процесс правильного разделения хромосом между дочерними клетками. Главным устройством, осуществляющим это разделение является веретено деления. Оно самоорганизуется из в клетке всякий раз, когда клетка подходит к митозу – фазе собственно деления клетки. Правильно организованное веретено к моменту расхождения хромосом – анафазе, устроено так, что микротрубочки, идущие от каждого из двух полюсов веретена прикреплены только к одной из двух сестринских хромосом. Наличие ошибочных прикреплений, когда одна хромосома связана с двумя полюсами или один полюс связан с двумя сестринскими хромосомами, всегда приводит к неправильному делению, что ведет либо к смерти клетки, либо к возникновению неконтролируемого клона клеток. В процессе самоорганизации микротрубочки имеют примерно одинаковую

хромосом. При том, что каждая хромосома в клетках человека может связать до 20 микротрубочек, получается, что неправильные закрепления (хотя бы одна трубочка прикрепилась неправильно) должны абсолютно доминировать в митозе. Однако, в нормальной клетке к моменту расхождения хромосом неправильные закрепления встречаются крайне редко, поскольку неправильные деления наблюдаются даже у дрожжей с вероятностью не больше 10^{-5} . Сегодня мы довольно много знаем о молекулярных процессах, происходящих в ходе митоза. Математические модели, основанные на этих представлениях, показывают, что, похоже, мы близки к пониманию механизмов, гарантирующих исправление ошибок с такой точностью. Предсказанные моделями новые эксперименты, видимо, позволят в ближайшие годы получить достаточно точное понимание механизмов самоорганизации правильного веретена деления.

Исследование механизма диффузии Dam1-комплексов по микротрубочке

Герус М.А.⁽¹⁾, Коршунова А.В.⁽¹⁾, Молодцов М.И.⁽¹⁾, Атауллаханов Ф.И.^(1, 2, 3)

¹ Центр теоретических проблем физико-химической фармакологии РАН, Россия

² Физический факультет, Московский Государственный Университет, Россия

³ Гематологический Научный Центр, Москва, Россия

gerus.marina@gmail.com

Abstract

Dam1 kinetochore protein complex is one of the most important candidates as a device that couples microtubule dynamics and chromosome movements in mitosis. Dam1 subunits themselves demonstrate unique ability to move along the microtubules. Here we design an experiment to study molecular mechanisms of Dam1 movement.

1 Введение

Во время митоза хромосомы разделяются по дочерним клеткам. Движение хромосом обеспечивается ростом и укорочением микротрубочек (МТ) веретена деления. Концы МТ закреплены на хромосомах в специальных местах, которые называются кинетохорами. Во время разборки связи между протофиламентами, из которых состоит МТ, разрушаются; отдельные протофиламенты изгибаются, а выделяемая при этом энергия служит для движения хромосом. Кинетохор - это большой белковый комплекс, задача которого использовать силу, развиваемую МТ, для движения хромосомы. В почкующихся дрожжах *Schizosaccharomyces pombe* был обнаружен кинетохорный белковый комплекс [3] – DASH-комплекс (или Dam1-комплекс), который образует кольцевые структуры вокруг МТ, и может играть роль связующего звена в сцеплении хромосом с МТ. Dam1-комплекс состоит из 10 белков (Ask, Dam1p, Spc34, Duo1, Spc19, Dad2p, Dad1p, Dad3p, Dad4p, Hsk3p), 16 гетеродекамеров Dam1-комплекса образуют кольцо. Было показано, что не только кольцевые комплексы, но также и более мелкие Dam1-комплексы – так называемые «пэтки» - могут диффундировать вдоль МТ. Механизм движения Dam1-комплексов разного размера является объектом данного исследования.

2 Методы

В качестве хозяина для векторов, содержащих фрагменты генов-мишеней, использовали штамм *E. coli* DH5a. Экспрессию целевого белка

проводили в штамме *E.coli* Rosetta pLysS. Штаммы поддерживали на среде LB (бакто-триптон- 10.0 г/л; дрожжевой экстракт- 5.0 г/л, NaCl-10г/л). Штамм Rosetta pLysS растили в присутствии хлорамфеникола (34 мкг/мл). Для отбора клонов, содержащих вектор, в среду LB дополнительно вносили ампициллин (100 мкг/мл). Для индукции синтеза целевого белка добавляли изопропил β-D-тиогалактопиранозид (IPTG). Твердые среды получали путем добавления агар-агара до конечной концентрации 1,5-2,0%. Выделение плазмидной ДНК проводили методом щелочного лизиса с небольшими модификациями. Последовательности праймеров конструировали с использованием программ BioEdit, Lasergene v.5.06 DNASTAR и Vector NTI 8.0. В качестве отрицательного контроля проводили ПЦР на ДНК штаммов *E.coli* Rosetta pLysS. Полимеразную цепную реакцию проводили в 20мкл реакционной смеси, содержащей однократный ПЦР буфер и стабилизатор, 0.2 mM dNTP, и 0.3 ед. Taq-полимеразы (ООО НПФ «Гентех» Россия). Амплификацию проводили в следующем режиме: денатурация ДНК (95°C – 3 мин) и 25 циклов реакции, состоящей из денатурации (95°C - 30 сек), отжига праймеров (60 °C-1 мин) и элонгации (72°C). Время элонгации устанавливали в соответствии с длиной амплифицируемого фрагмента. Полученные в ходе ПЦР фрагменты ДНК анализировали в 1% агарозном геле, окрашенном бромистым этидием. ПЦР-продукты очищали путем переосаждения ДНК раствором этанола с 0.75 М ацетата аммония при комнатной температуре. Синтез целевого белка в штамме *E. coli* Rosetta pLysS запускали в момент достижения культурой OD₆₀₀= 0,5 путем добавления IPTG до конечной концентрации 2мМ. Затем клетки растили в течение 4 часов, полученную биомассу открывали, однократно отмывали в буфере PBS и замораживали. Выделение белка проводили по стандартной методике с небольшими модификациями [4]. Концентрацию очищенного белка определяли с помощью спектрофотометра Ultrospec 1100pro (Amersham Biosystems, USA), в соответствии с рекомендациями производителя. Выделение тубулина, мечение родамином и

реакцию полимеризации МТ проводили по стандартной методике [2]. Все измерения на световом микроскопе Zeiss AxioImager.Z1 проводились при 32°C. Изображения получали с помощью CCD камеры Photometric CascadeII, используя программное обеспечение Micro-Manager-1.3. Эксперимент проводился в разработанной нами проточной камере объемом 15 мкл, используя перистальтический насос и скорость потока 20 мкл/мин. Флуоресценцию EGFP наблюдали при освещении лазером $\lambda=473\text{nm}$.

3 Результаты

Нами была создана генетическая конструкция, в который с геном белка Dam1-комплекса Ask связали ген флуоресцентного белка EGFP. Полученную генетическую конструкцию проверили с помощью ПЦР и рестрикционного анализа. Для выделения белка были взяты клоны, предварительно отобранные по интенсивности EGFP в клетках. В результате был получен Dam1-комплекс, каждый гетеродекамер которого помечен одной молекулой EGFP (Dam1[Ask-GFP]). Была проверена способность Dam1[Ask-GFP] связываться с МТ.

Для этого в разработанной нами проточной камере МТ абсорбировались на покровное стекло при помощи тубулиновых антител. После этого в камеру вливался Dam1 комплекс. На основании проведенных 10 серий экспериментов был установлен рабочий диапазон концентраций белка, который составил 40-55 нмоль. При концентрации ниже Dam1 не присоединяется к МТ. При более высоких концентрациях МТ были равномерно покрыты свящимся Dam1.

Для изучения зависимости коэффициента диффузии Dam1-комплекса от его размера мы пользовались методикой [3]. Для этого Dam1 в малой концентрации абсорбировался на покровное стекло. При этом на стекле оказывалось несколько светящихся точек. Мы записывали уменьшение яркости этих точек из-за выгорания со временем. Выгорание происходило ступенчато, что связано с дискретным количеством Dam1[Ask-GFP] субъединиц в светящейся точке.

Гистограмма распределения интенсивностей точек имеет ряд пиков, что позволяет определить

яркость одного комплекса. Она составила 300 ± 40 у.е. при мощности лазера 47 мВт (Рис.1).

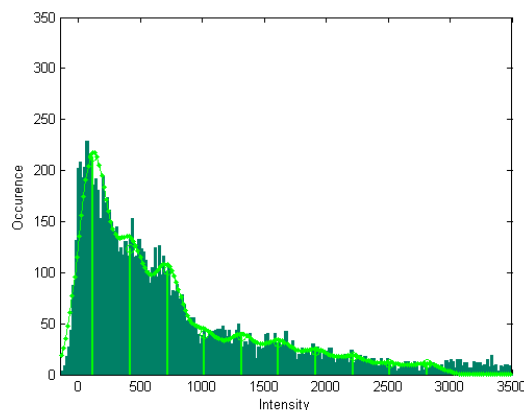


Рис 1. Гистограмма интенсивности точек Dam1[Ask-GFP] (3 экспериментальных поля, 40 точек). и аппроксимация эквидистантными Гауссами.

В результате мы определили значение светимости одной молекулы EGFP, что позволяет оценить количество гетеродекамеров в пэтчах для дальнейшей работы.

Список литературы

- [1] E. L. Grishchuk et al. Different assemblies of the DAM1 complex follow shortening microtubules by distinct mechanisms. PNAS May 13, 2008.
- [2] T. Mitchison.
<http://mitchison.med.harvard.edu/protocols.html>
- [3] H.W. Wang et al. Architecture of the Dam1 Kinetochore Ring Complex: Implications for Microtubule-Driven Assembly and Force-Coupling Mechanisms. Nat Struct Mol Biol. 2007 Aug;14(8)
- [4] S.Westermann et al. Formation of a Dynamic Kinetochore-Microtubule Interface through Assembly of the Dam1 Ring Complex. Molecular Cell, Vol. 17, 277–290, January 21, 2005.

Динамическая модель диффузии Dam1 комплекса вдоль микротрубочки

Захаров П.Н.⁽¹⁾, Молодцов М.И.⁽¹⁾ Атауллаханов Ф.И.^(1,2,3)

Центр теоретических проблем физико-химической фармакологии РАН⁽¹⁾

Гематологический научный центр РАМН⁽²⁾

Московский государственный университет им. М.В. Ломоносова, физический факультет⁽³⁾
pavell_86@mail.ru

Abstract

There are some classes of non-motor proteins moving along the microtubule without dissociating for an extended distances. Among them are +Tops which are able to track microtubule growing end and Dam1 complexes which can freely diffuse along the stable microtubule wall. It is still unknown how such proteins can diffuse along the microtubule remain bound to it while traveling for several microns. In current work dynamical model of Dam1 complex has been created to gain an insight on possible mechanism of Dam1 diffusion along microtubules.

1 Общие сведения

1.1 Структура микротрубочки

Микротрубочки – это полимеры, построенные из молекул тубулина, димеров состоящих из α и β мономеров. Размер каждого мономера – 4 нм. Внутри полимера димеры организованы в линейные протофиламенты. Несколько протофиламентов, взаимодействуя латерально, образуют цилиндрическую стенку микротрубочки диаметром 25 нм.



Рисунок 1. А) Строение микротрубочки. Б) Димер тубулина с двумя карбоксильными кончиками [3].

Мономер тубулина заканчивается небольшим, но очень подвижным фибриллярным карбоксильным кончиком (для простоты будем называть его далее С-кончиком). Он состоит из нескольких десятков аминокислотных остатков. Его длина около 4 нанометров.

1.2 Строение Dam1 комплекса

Dam1 комплекс – это олигомер, состоящий из десяти разных белков [2, 4]. Он может образовывать небольшие незамкнутые комплексы. Самый простой из них состоит всего из двух субъединиц (рис. 2).

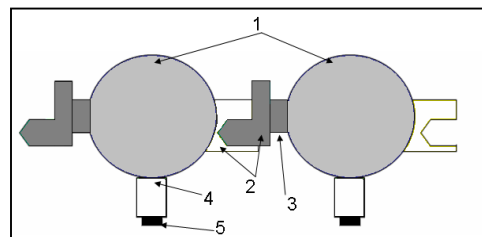


Рисунок 2. Схематическое строение Dam1 комплекса, состоящего из двух субъединиц. Обозначения: 1. тело Dam1, 2. участки латеральных взаимодействий, 3. участок, поддающийся деформации кручения, 4. линкер Dam1 комплекса, 5. участок линкера, взаимодействующий с микротрубочкой.

В экспериментах по изучению диффузии Dam1 наблюдают за перемещением окрашенного комплекса вдоль стабилизированной микротрубочки (рис. 3).

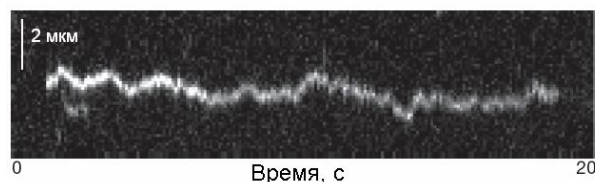


Рисунок 3. Кимограмма диффузии Dam1 [1].

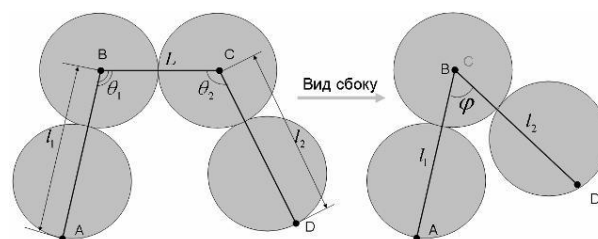


Рисунок 4. Представление Dam1 комплекса в модели. АВ и CD–линкеры, ВС–основание длины L.

2 Математическая модель

2.1 Уравнения движения

Мы рассматриваем Dam1 комплекс, состоящий из двух субъединиц. Каждая субъединица может

связываться с двумя С-кончиками микротрубочки. В нашей модели движение такого комплекса описывается как движение четырех связанных броуновских частиц сферической формы радиуса 4 нм (рисунок 4).

Положение каждого С-кончика задается двумя точками. Первая находится на поверхности микротрубочки. Это неподвижная точка, она является основанием С-кончика (точка L на рис. 1, Б). Движение второй (точка М на рис. 1, Б) описывается как движение броуновской частицы в вязкой среде под действием тепловых флуктуаций в потенциальном поле, заданным уравнением [3]

$$U_c(r) = -\frac{1}{2} k R_0^2 \lg \left(1 - \frac{r^2}{R_0^2} \right), \quad 0 < r < R_0,$$

где R_0 – длина С-кончика, k – коэффициент, определяющий крутизну потенциального барьера. Подвижная точка С-кончика взаимодействует с Dam1 комплексом.

Для описания динамического поведения системы во времени мы решаем систему уравнений Ланжевена в приближении броуновской динамики. Значение каждой координаты в момент времени $i+1$ мы находим по формуле

$$x_{i+1} = x_i + \frac{dt}{\gamma} \frac{\partial U}{\partial x} + \sqrt{2k_B T \frac{dt}{\gamma}} \cdot z_i,$$

где x_i – значение координаты на предыдущем шаге по времени; dt – шаг по времени; U – потенциальная энергия системы, k_B – константа Больцмана; T – температура; z_i – случайная величина, полученная из распределения Гаусса. $\gamma = 6\pi R\eta$ – коэффициент вязкости глобулы, рассчитанный по формуле Стокса ($R = 4$ нм для каждой из четырех глобул Dam1 комплекса и $R = 0.4$ нм для С-кончика), $\eta = 10 \cdot \eta_0 \approx 10^{-2}$ Па*с – вязкость среды, η_0 – динамическая вязкость воды при $T = 300$ К.

2.2 Потенциальная энергия системы

Потенциальная энергия системы U состоит из трех слагаемых:

$$U = U_c + U_{Dam} + U_{int};$$

где U_{Dam} – энергия деформации комплекса (см. рис.4), U_{int} – энергия взаимодействия Dam1 с микротрубочкой.

$$U_{Dam} = \frac{1}{2} K_\theta ((\theta_1 - \theta_0)^2 + (\theta_2 - \theta_0)^2) + \frac{1}{2} K_\varphi (\varphi - \varphi_0)^2 + \frac{1}{2} K_L (L - L_0)^2 + \frac{1}{2} K_I ((l_1 - l_0)^2 + (l_2 - l_0)^2);$$

$$U_{int} = K_{Dam} * \sum_{i,m} \left(1 - \exp \left\{ -\frac{\rho_{i,m}^2}{\rho_0^2} \right\} \right), \text{ где}$$

K_{Dam} – коэффициент, определяющий силу связывания, $K_\theta, K_\varphi, K_L, K_I$ – жесткости

деформации молекулы на изгиб и кручение. $\rho_{i,m}$ – расстояние от i -го сайта связывания на Dam1 комплексе ($i = 1, 2, 3, 4$) до m -го С-кончика, ρ_0 – константа, определяющая расстояние, на котором происходит взаимодействие (равна 0.24 нм – длине белок-белкового взаимодействия).

3 Результаты и выводы

Первым и главным результатом является создание полностью динамической модели диффузии Dam1 комплекса по микротрубочке. Нам удалось описать реальные времена (сотни миллисекунд) на обычном компьютере.

Мы проверили, может ли Dam1 комплекс диффундировать вдоль трубочки, связываясь не с С-кончиками, а с неподвижными участками на поверхности тубулина (по одному на каждый мономер). Оказалось, что в такой модели Dam1 вообще не способен перемещаться вдоль трубки. Отсюда следует вывод, что наличие связывания именно с подвижными С-кончиками обеспечивает движение Dam1 вдоль микротрубочки.

Также было показано, что в нашей модели вдоль микротрубочки могут диффундировать комплексы, состоящие всего из одной субъединицы Dam1.

Были посчитаны зависимости коэффициентов диффузии и времен жизни Dam1 комплекса, состоящего из одной и двух субъединиц, от энергии связывания с микротрубочкой K_{Dam} при разных длинах С-кончиков. (Под “временем жизни” мы понимаем время, в течение которого Dam1 остается связанным с микротрубочкой хотя бы одним С-кончиком).

References

- [1] D. Gestaut et al. Phosphoregulation and Depolymerization-Driven Movement of the Dam1 Complex do not Require Ring Formation. In *Nature Cell Biology*, volume 10, pages 407 - 414, 2008.
- [2] J. Miranda et al. The Yeast DASH Complex Forms Closed Rings on Microtubules. In *Nature Structural & Molecular Biology*, volume 12(2), pages 138-143, 2005.
- [3] A. Priel et al. Transitions in Microtubule C-termini Conformations as a Possible Dendritic Signaling Phenomenon. In *Eur Biophys J*, volume 35, pages 40-52, 2005.
- [4] S. Westermann et al. Formation of a Dynamic Kinetochore-Microtubule Interface through Assembly of the Dam1 Ring Complex. In *Molecular Cell*, volume 17, pages 277-290, 2005.

Кодирование электрокардиосигналов на основе вариабельности условной энтропии конечных разностей

Казанцев А.П.

Институт биологического приборостроения РАН

telemed@itaec.ru

Electrocardiosignal Encoding Based on Variability of Finite Differences' Conditional Entropy

Kazantsev A.P.

Abstract

The method for electrocardiosignal encoding based on variability of finite differences' conditional entropy is introduced. Numerical experiment results are present.

1 Введение

При создании телемедицинских систем с использованием электрокардиографии необходимо решать задачи эффективного кодирования электрокардиосигналов (ЭКС) для оптимизации их передачи по телекоммуникациям. Широко применяются методы линейно-предсказывающего кодирования [1]. Программы на их основе могут выполняться в целочисленной арифметике с минимальным количеством операций, обеспечивая необходимое быстродействие даже на микроконтроллерах. Наиболее эффективно кодирование на нескольких порядках разностей дискретных отсчетов сигнала.

2 Кодирование марковских цепей

Кодирование должно выполняться с учетом статистической взаимозависимости отсчетов ЭКС. Можно показать [2], что если представлять последовательность отсчетов $\{X_k\}_{k \geq 0}$ как простую цепь Маркова, которая в любом сечении k характеризуется ансамблем A , состоящим из N целых чисел a , то средняя условная энтропия разности 1-го порядка $H(\Delta X / X)$ не будет превышать ее собственную энтропию,

$$H(\Delta X / X) \leq H(\Delta X).$$

Это позволяет предположить, что если при разных уровнях сигнала a_i ($i=1, 2, \dots, N$) энтропия $H(\Delta X / a_i)$ различна, то при каких-то a_r должны существовать ее значения меньшие, чем среднее,

$H(\Delta X / a_r) < H(\Delta X / X)$. Можно предполагать и существование непрерывных участков сигнала, на которых выполняется это условие. Последовательность разности $\{\Delta X_k\}_{k \geq 0}$ также как сигнал может быть представлена в виде простой марковской цепи, для которой вероятно существуют значения Δx_r или даже сплошные участки, где условная энтропия разностей порядка ($m=2$) будет меньше средней, т.е. $H(\Delta^2 X / \Delta x_r) < H(\Delta^2 X / \Delta X)$. Анализ можно продолжать, рассматривая цепи разностей более высоких порядков – $m=3, 4, \dots$, обосновывая возможность существования участков сигнала, на которых

$$H(\Delta^m X / \Delta^{m-1} x_r) < H(\Delta^m X / \Delta^{m-1} X).$$

3 Численный эксперимент

Рис. 1 воспроизводит целочисленную запись ЭКС, последовательность разностей отсчетов сигнала Δ^1 – «1-я разность», и последовательность его вторых разностей Δ^2 – «2-я разность». Для сравнения последовательности на рисунке представлены в одном масштабе, и видно, как сужаются области значений при переходе от ЭКС к Δ^1 и далее к Δ^2 , что легко обеспечивает сжатие информации.

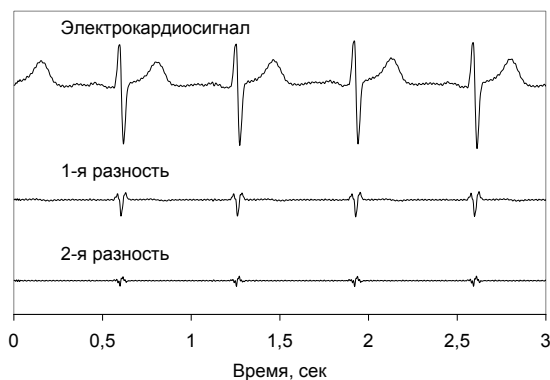


Рис. 1. Воспроизведение дискретного ЭКС и последовательностей разностей 1-го и 2-го порядка

На рис. 2 представлена диаграмма рассеяния разностных значений Δ^1 и Δ^2 попарно. Диаграмма позволяет наблюдать двумерное распределение этих величин и иллюстрирует предположение о том, что

распределения условной вероятности Δ^2 должны обладать меньшей неопределенностью, чем распределение полной вероятности этой величины.

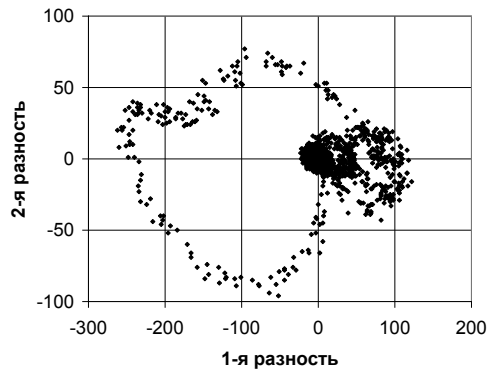


Рис. 2. Диаграмма рассеяния в системе координат $\Delta^1 \Delta^2$ (1-й и 2-й разностей отсчетов ЭКС)

Гистограмма распределения полной вероятности (рис. 3а), построенная на данных, представленных на рис. 1-2, близка к функции нормального распределения и ей соответствует энтропия выборки из всего ансамбля значений Δ^2 , $H(\Delta^2 X / \Delta X)$. В данном примере $H(\Delta^2 X / \Delta X) \approx 4,5$.

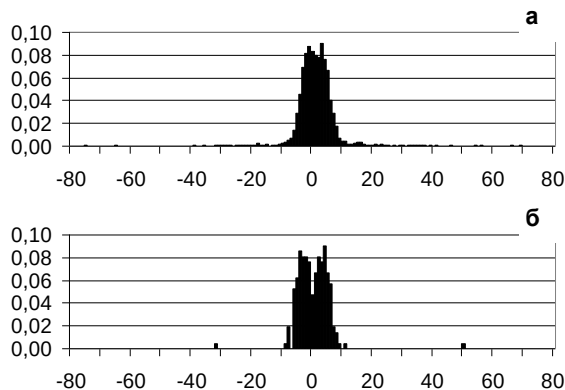


Рис. 3. а) гистограмма распределения полной вероятности Δ^2 ; б) гистограмма распределения Δ^2 при $\Delta^1 = 0$ (по данным рис. 1-2)

Гистограмма распределения условной вероятности при $\Delta^1 = 0$ для тех же данных (рис.3б), заметно отличающаяся от нормального распределения, характеризуется и меньшей условной энтропией, $H(\Delta^2 X / \Delta x = 0) < 4$.

4 Заключение

Эффект уменьшения условной энтропии конечных разностей был использован при разработке метода сжатия электрокардиосигналов для передачи в реальном масштабе времени [2]. Метод показал свою высокую эффективность и перспективность применения в медико-биологических информационных системах, в том числе и в персональных сетях с использованием беспроводной связи ограниченной пропускной способности.

References

- [1] Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. - М.: ДИАЛОГ-МИФИ, 2002. - 384 с.
- [2] Казанцев А.П., Сенин А.А. Метод сжатия электрокардиосигналов для передачи в реальном масштабе времени. // Биомедицинская радиоэлектроника. – 2008. – №7. – С.15-21.

Оценка остаточной среднеквадратической погрешности измерения электрокардиосигналов

Казанцев А.П.

Институт биологического приборостроения РАН

telemed@itaec.ru

Residual Least Square Error Estimation for Electrocardiosignals' Measurement

Kazantsev A.P.

Abstract

The method of residual least square error estimation for electrocardiosignals' optimal numerical representation is introduced. The method is based on series-mode signal model and is making use of autocorrelation and spectral analysis.

1 Введение

В телемедицинских и других приложениях цифровой электрокардиографии необходимо решать задачу оптимизации точности измерений и числового представления электрокардиосигнала (ЭКС) [1]. В [2] дано обоснование критерия точности в виде минимума среднеквадратического отклонения неустранимой погрешности на заданном множестве преобразований сигналов – остаточной среднеквадратической погрешности σ_0 .

Анализ погрешности основан на линейной модели смеси истинного ЭКС и помех с дискретным аргументом $i = \dots, -2, -2, 0, 1, 2, \dots$

$$x(i) = s(i) + y(i) + u(i) + v(i),$$

где $x(i)$ – реальный ЭКС, $s(i)$ – истинный ЭКС, $y(i)$ – флуктуации изолинии истинного ЭКС, $u(i)$ – сетевая наводка, $v(i)$ – широкополосный шум.

2 Метод оценки

В представленной работе получил развитие метод оценки σ_0 [2], использующий стабильность изолинии истинного ЭКС на сегментах ТР последовательности кардиоциклов (рис. 1а). Поскольку на этих отрезках $s(k) = C$ ($C = \text{const}$, постоянное смещение в пределах реализации сегмента; $k = 0, 1, 2, \dots, n-1$; n – длина отрезка), то $x(k) = y(k) + u(k) + v(k) + C$.

Для получения значения оценки $\tilde{\sigma}_0$ из представительного множества реальных сигналов $x(i)$ выделяется большая выборка в m сегментов ТР с наложенными помехами $y(k)$, $u(k)$, $v(k)$ (рис. 1б). Флуктуационная помеха $y(k)$ вместе со смещением C удаляется полиномиальной регрессией [3]. В результате получаются фрагменты смеси сетевой наводки и широкополосного шума $x_0(k)$ (рис. 1в),

$$x_0(k) \approx u(k) + v(k).$$

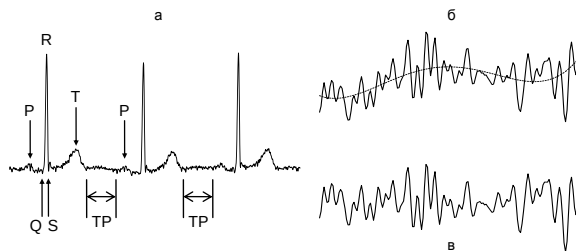


Рис. 1. Выделение смеси шума и наводки: а – реальный электрокардиосигнал; б – сегмент ТР с помехами (включая флуктуации); в – смесь шума и наводки после удаления флуктуаций

Фрагменты $x_0^{(l)}(k)$ ($l = 1, \dots, m$; $k = 0, 1, \dots, n^{(l)} - 1$) считаются отрезками реализаций некой случайной последовательности, принимаемой за стационарную в широком смысле, т.к. оценки ее математического ожидания, дисперсии и корреляционной функции не зависят от времени. Представленный метод не требует, чтобы длина n у всех фрагментов была одинаковая. Каждый l -й фрагмент предполагается принадлежащим эргодической реализации и используется для вычисления корреляционной функции вида

$$R_{xx}^{(l)}(j) = \sum_{k=0}^{n^{(l)}-1-j} x_0^{(l)}(k) x_0^{(l)}(j+k),$$

где $j = -(n^{(l)} - 1), \dots, 0, 1, \dots, n^{(l)} - 1$, $x_0^{(l)}(j+k) = 0$ при $j+k < 0$ и $j+k \geq n^{(l)}$. Далее вычисляется усредненная корреляционная функция (рис. 2а) вида

$$\bar{R}_{xx}(j) = \frac{1}{N-1} \sum_{l=1}^m R_{xx}^{(l)}(j),$$

где $N = \sum_{l=1}^m n^{(l)}$, $j = -(n_{\max} - 1), \dots, 0, 1, \dots, n_{\max} - 1$,
 $n_{\max} > n^l$ при любом l .

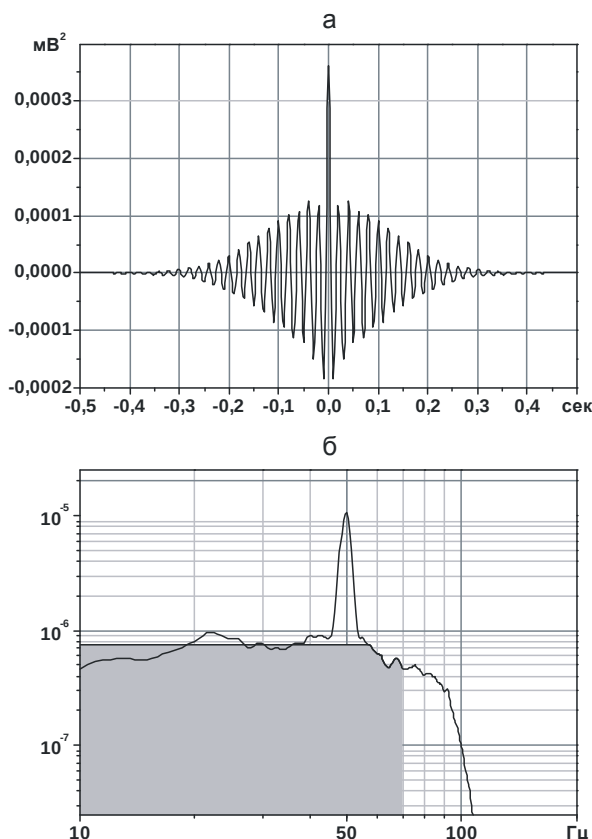


Рис. 2. Обработка выборки фрагментов смеси шума с наводкой: а – Корреляционная функция смеси; б – Спектр мощности смеси и аппроксимация спектра неустраняемого шума

Из $\bar{R}_{xx}(j)$ дискретным преобразованием Фурье в соответствии с теоремой Винера-Хинчина [4] получается спектр мощности смеси сетевой наводки и широкополосного шума (рис. 2б). Часть спектра последнего является частотным представлением неустраняемой помехи и может быть выделена численно-графическим способом.

Ограниченная диагностически значимым частотным диапазоном аппроксимация спектра неустраняемого шума (показана на рис. 2б серым фоном) определяет дисперсию оценочной остаточной погрешности $\tilde{\sigma}_0^2$, вычисляемой интегрированием в значимой области выделенного спектра шума. Отсюда получается искомая величина $\tilde{\sigma}_0$.

3 Заключение

Метод оценки минимальной погрешности измерения электрокардиосигналов успешно используется и демонстрирует свою эффективность при разработке систем мобильной телемедицины [1] – для оптимизации разрядности аналого-цифрового преобразования сигналов, разрешения устройств их графического отображения и скорости передачи данных в реальном масштабе времени.

References

- [3] Казанцев А.П., Сенин А.А., Пикуленко О.В. Архитектура и реализация мобильных телемедицинских комплексов дистанционной электрокардиографии. // Вестник Рязанской государственной радиотехнической академии. – Рязань, 2007, выпуск 20. – С. 85-90.
- [4] Казанцев А.П. Информационный анализ электрокардиосигналов и их преобразований для мобильной телемедицины. // Биомедицинская радиоэлектроника. – 2008. – №7. – С.22-28.
- [5] Казанцев А.П. Полиномиальная аппроксимация дрейфа изолинии электрокардиосигнала. // Перспективные проекты и технологии. – 2007. – Вып. 2. – С.85-89.
- [6] Сергиенко А.Б. Цифровая обработка сигналов. – СПб.: Питер, 2002. – 608 С.

О вычислительных методах исследования автоволновых процессов

Калюжный И.М.

Московский Государственный Университет Экономики, Статистики и Информатики

shahhmatist@mail.ru

Abstract

This paper contains a survey on computational methods for autowave processes and dissipative structures analyzing. Stiffness of processes in spatial and timing variables is taken into account. "Scanning method" for analyzing of autowave processes in heterogeneous environment with complex boundary is proposed.

Для математического исследования автоволновых процессов и динамических систем в настоящее время принято рассматривать системы нелинейных дифференциальных уравнений в частных производных гиперболического и параболического типов [1]

$$\frac{\partial E_i}{\partial t} = F_i(\nabla E_i, E_1, \dots, E_n) + \nabla(D_i \nabla E_i) \quad (1)$$

где E_i - переменная, F_i - нелинейная функция, D_i - коэффициент диффузии.

Для описания автономных процессов нелинейных динамических достаточно исследовать системы двух уравнений

$$\frac{\partial E_i}{\partial t} = F_i(E_1, \dots, E_n) + D_i \Delta E_i, \quad (i = 1, 2) \quad (2)$$

Точечной системой для (1) является система нелинейных дифференциальных уравнений

$$\frac{dE_i}{dt} = F_i(E_1, \dots, E_n), \quad (i = 1, \dots, n) \quad (3)$$

Специфической особенностью системы (3) при описании автоволновых процессов является то, что эта система является жесткой. Жесткость (3) приводит к тому, что решения (3), автоколебательные или потенциально колебательные, содержат участки «быстрых» и «медленных» движений, образуя на фазовой плоскости замкнутые или близкие к замкнутым незамкнутые предельные циклы. В качестве примера на рис. 1.1 приведена форма колебаний «жесткой» системы для уравнения Ван дер Поля – Фитцхью. На рис. 1.2 показана спиральная волна, порождаемая уравнением (3), где в качестве

жесткой системы использовано уравнение Ван дер Поля – Фитцхью. Существенной особенностью автоволновых процессов «жесткой» системы (3) является также жесткость по пространственным переменным x, y, z . Сказанное следует учитывать при выборе сеточных методов нахождения приближенных решений (1).

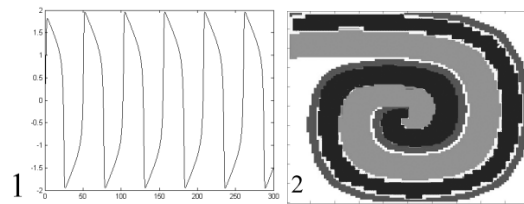


Рис. 1. Иллюстрация изменения переменных в жестких системах: 1) жесткая система по времени; 2) жесткая система по пространственным переменным x, y, z

В настоящее время широко используются: 1) явная трехточечная схема [2,3]; 2) неявная схема метода центральных разностей [2,3]; 3) сведение к системе нелинейных дифференциальных уравнений первого порядка [2,3]. Кратко отметим их особенности. При использовании явной трехточечной схемы алгоритм наиболее простой, поскольку последующие состояния переменных уравнения $U_i(t, x, y, z)$ определяются рекурсивно.

Недостатком метода являются: а) необходимость использования достаточно малого шага по времени для достижения устойчивости счета; б) необходимость еще более малого шага по времени, определяемого скоростью изменения в фазу «быстрых» движений (при этом быстрые изменения имеет одна из переменных $U_i(t, x, y, z)$). Сказанное относится и к методу центральных разностей.

Более перспективным является использование метода сведения к системе дифференциальных уравнений [2,3]. В этом случае преодолеть трудностей, связанных с жесткостью получаемых систем нелинейных дифференциальных уравнений, можно за счет использования специальных алгоритмов для приближенного решения жестких обыкновенных нелинейных дифференциальных уравнений, например, метода Розенброка [3]. Программное решение особенно эффективно в

среде MATLAB 7, использующей матричную форму представления данных.

Расчеты показали, что при использовании метода сведения к системе дифференциальных уравнений при одинаковой точности вычислений возможно увеличение скорости счета на порядок.

Погрешность вычислений сеточных методов имеет две основные составляющие: а) погрешность, обусловленную аппроксимацией частных производных; б) погрешностью, обусловленной решением систем обыкновенных дифференциальных уравнений. В явной сеточной схеме система дифференциальных уравнений приближенно решается по методу Адамса и имеет погрешность порядка $O(\Delta t)$ (Δt - шаг по времени). При использовании метода дифференциальных уравнений приближенное решение системы обыкновенных дифференциальных уравнений находится по методам типа Рунге – Кутты более высокого порядка, при этом погрешность имеет порядок $O(\Delta t)^k$ $k > 2$. При одинаковой погрешности скорость счета намного больше. При этом основную роль начинает играть погрешность аппроксимации вторых частных производных.

Для уменьшения погрешности, обусловленной аппроксимацией вторых частных производных, можно использовать ее аппроксимацию не только по трехточечной, но и по пятиточечной схеме и схемам более высокой степени точности.

Новым специфическим моментом при вычислительном исследовании автоволновых процессов и диссипативных структур является многозначность решений, особенно для диссипативных структур. Появление того или иного решения зависит от начальных условий, играющих роль бифуркационных параметров. При этом различие в топологических свойствах решений может быть большим даже при малом изменении параметра. Поэтому при вычислительном исследовании требуется: а) повышение точности вычислений; б) точный набор начальных параметров.

Различие типов решений при небольшом изменении начальных условий эксперимента иллюстрируется на рис. 2.

Рис. 2 иллюстрирует переход автоволнового процесса - спиральной волны в диссипативную структуру, составленную из частей – кусков спиральной волны.

Наличие гетерогенности, сложной структуры областей приводит к значительному усложнению программы и времени счета. Для повышения эффективности решения с учетом особенностей вычислительной среды Matlab-7 разработан алгоритм, основанный на сканировании области. Узлы счета располагаются на траекториях

Наличие гетерогенности, сложной структуры областей приводит к значительному усложнению программы и времени счета. Ведущая траектория имеет возрастающую нумерацию узлов, нумерация узлов остальных траекторий вычисляется в процессе подготовки.

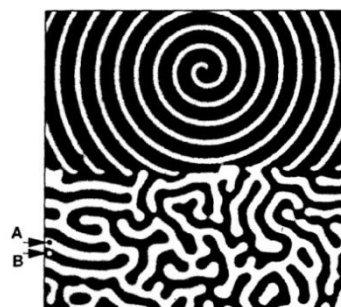


Рис. 2. Иллюстрация перехода автоволнового процесса - спиральной волны в диссипативную структуру

В итоге формируется одномерный массив данных для ввода в программу решения жестких систем дифференциальных уравнений в Matlab 7.

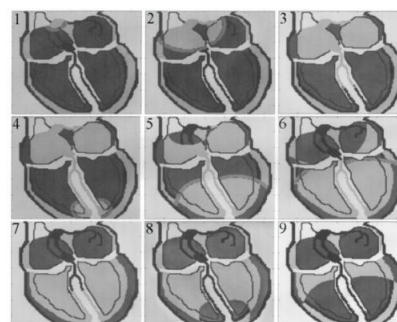


Рис. 3. Результаты расчетов распространения электрического потенциала, произведенном согласно предлагаемому методу сканирования.

На рис. 3 видно последовательное распространение возбуждения от синоатриального узла сердца, по атриовентрикулярному узлу, ножкам пучка Гиса, волокнам Пуркинью и желудочку.

Использование метода сканирования для расчета процессов в распределенной двумерной системе, имеющей сложную геометрию и гетерогенные составляющие, является эффективным, поскольку позволяет значительно упростить программирование и уменьшить время проведения вычислений.

References

- [1] М. Е. Мазуров Нелинейная синхронизация и ритмогенез в электровозбудимых системах сердца//Диссертация доктора физико-математических наук. Пушкино. 2007
- [2] Математические вопросы численного решения гиперболических уравнений//М.: ФИЗМАТГИЗ. 2001. с. 608
- [3] 4. Erwin Stein, Ren De Borst, Thomas J.R. Hughes. Encyclopedia of Computational Mechanics (3 Volume Set) free download. 2004. 233

Регрессионный анализ функциональной асимметрии левого и правого желудочков сердца в условиях вазореальной артериальной гипертензии

Зотов А.К., Зотова Т.Ю., Благоданов М.Л.

Российский университет дружбы народов, Москва

zotovat@mail.ru

Abstract

The analysis of rabbit's cardiovascular hemodynamic indices under experimental arterial hypertension had shown that the adaptation of the cardiovascular system to increased arterial BP is visualized by the study of three-dimensional correlation between intraventricle BP indices of left and right ventricles, and systolic and diastolic BP indices. With the help of empiric production functions it was shown that the intraventricle BP's contribution of left and right ventricles into formation of systolic and diastolic BP changed from one week of experiment to another, but production functions of systolic and diastolic BP had produced the constant effect of production and therefore were Cobb-Douglas production functions.

1 Материалы и методы

Экспериментальные материалы, используемые в данном исследовании, были получены на кафедре общей патологии и патологической физиологии РУДН. Данный эксперимент проводился на кроликах-самцах породы «шиншилла» общим количеством 25 и весом 2,5-3,5 кг, которые были разделены на 5 групп исследования (по 5 животных в каждой группе). Контрольную группу составили интактные кролики. В опытных группах у всех животных под общим наркозом моделировали вазореальную артериальную гипертензию путём хирургической операции сужения брюшной аорты на 1/3 от первоначального диаметра над местом отхождения от неё почечных артерий. Затем на аппаратно-программном комплексе «МИКАРД» в сроки 1, 2, 4 и 6 недель от момента операции производилось измерение параметров системной и внутрисердечной гемодинамики, а именно систолического $z(АД_c)$ и диастолического $w(АД_d)$ артериального давления, рабочего внутривентрикулярного давления левого $x(ВД_rЛЖ)$ и правого $y(ВД_rПЖ)$ желудочков сердца.

Для выявления тримерной корреляции между тройками параметров $ВД_rЛЖ-ВД_rПЖ-АД_c$ и $ВД_rЛЖ-ВД_rПЖ-АД_d$ с помощью общих коэффициентов тримерной корреляции r_{xyz} , r_{xyw} [1, 2, 4] был проведен тримерный корреляционный анализ тесноты связи между этими тройками параметров.

$$r_{xyz} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2r_{xy}r_{xz}r_{yz}}{1 - r_{xy}^2}}$$
$$r_{xyw} = \sqrt{\frac{r_{xw}^2 + r_{yw}^2 - 2r_{xy}r_{xw}r_{yzw}}{1 - r_{xy}^2}},$$

где r_{xy} , r_{xz} , r_{yz} , r_{xw} , r_{yw} - парные коэффициенты корреляции между параметрами x и y , x и z , y и z , x и w , y и w

Для оценки вкладов параметров внутривентрикулярного давления левого и правого желудочков сердца в формирование параметров систолического $z(АД_c)$ и диастолического $w(АД_d)$ давления применялся тримерный регрессионный анализ [1, 2, 3]. Сначала были рассчитаны индексы параметров $x(ВД_rЛЖ)$, $y(ВД_rПЖ)$, $z(АД_c)$, $w(АД_d)$. Затем с их использованием для параметров систолического $z(АД_c)$ и диастолического $w(АД_d)$ артериального давления были построены эмпирические производственные функции двух переменных $\hat{z} = x^{\hat{\alpha}_1} y^{\hat{\alpha}_2}$ и $\hat{w} = x^{\hat{\beta}_1} y^{\hat{\beta}_2}$, которые описывают регрессионные связи между тройками параметров: $x(ВД_rЛЖ) - y(ВД_rПЖ) - z(АД_c)$ и $x(ВД_rЛЖ) - y(ВД_rПЖ) - w(АД_d)$. Коэффициенты эластичности $\hat{\alpha}_1$, $\hat{\alpha}_2$ и $\hat{\beta}_1$, $\hat{\beta}_2$ производственных функций систолического и диастолического давления показывают, что изменение параметра $x(ВД_rЛЖ)$ на 1% приводит к изменению параметров $z(АД_c)$ и $w(АД_d)$ соответственно на $\hat{\alpha}_1\%$ и $\hat{\beta}_1\%$, а изменение

параметра $y(\text{ВД}_\text{ПЖ})$ на 1% приводит к изменению параметров $z(\text{АД}_\text{с})$ и $w(\text{АД}_\text{д})$ соответственно на $\hat{\alpha}_2$ % и $\hat{\beta}_2$ %.

Оценки коэффициентов эластичности и уровней детерминации и значимости производственных

функций по экспериментальным данным были получены методом наименьших квадратов с помощью статистической подпрограммы ЛИНЕЙН приложения Excel для Windows.

Таблица 1 Общие коэффициенты тримерной корреляции r_{xyz} и r_{xyw} троек параметров ВД и АД

Тройки параметров	в контроле	1 неделя	2 неделя	4 неделя	6 неделя
$\text{ВД}_\text{рЛЖ}-\text{ВД}_\text{рПЖ}-\text{АД}_\text{с}$ r_{xyz}	0,825 (очень сильная)	0,457 (средняя)	0,679 (сильная)	0,982 (очень сильная)	0,946 (очень сильная)
$\text{ВД}_\text{рЛЖ}-\text{ВД}_\text{рПЖ}-\text{АД}_\text{д}$ r_{xyw}	0,573 (средняя)	0,342 (слабая)	0,733 (сильная)	0,878 (очень сильная)	0,923 (очень сильная)

Таблица 2 Коэффициенты эластичности «производства» параметров АД параметрами ВД

Производственная функция	Коэффициент эластичности	в контроле	1 неделя	2 неделя	4 неделя	6 неделя
$\text{АД}_\text{с} \quad \hat{z} = x^{\hat{\alpha}_1} y^{\hat{\alpha}_2}$	$\hat{\alpha}_1$ по ЛЖ	0,83	0,67	0,82	1	1
	$\hat{\alpha}_2$ по ПЖ	0,17	0,33	0,18	0	0
$\text{АД}_\text{д} \quad \hat{w} = x^{\hat{\beta}_1} y^{\hat{\beta}_2}$	$\hat{\beta}_1$ по ЛЖ	0,84	0,73	0,95	1	1
	$\hat{\beta}_2$ по ПЖ	0,16	0,27	0,05	0	0

2 Результаты исследований

Результаты тримерного корреляционного анализа (табл. 1) показывают, что значения r_{xyz} и r_{xyw} для троек параметров $x(\text{ВД}_\text{рЛЖ}) - y(\text{ВД}_\text{рПЖ}) - z(\text{АД}_\text{с})$ и $x(\text{ВД}_\text{рЛЖ}) - y(\text{ВД}_\text{рПЖ}) - w(\text{АД}_\text{д})$ достоверно отличаются от нуля на 1% уровне значимости, как в контроле, так и в эксперименте на 2, 4, 6 неделях, а значение r_{xyz} – на 5% уровне значимости на 1 неделе. Динамика изменений общих коэффициентов тримерной корреляции по срокам эксперимента указывает на то, что адаптация сердечно-сосудистой системы к экспериментальной артериальной гипертензии может быть визуализирована при анализе тесноты тримерной корреляции показателей ВД левого и правого желудочков сердца, и показателей системного АД.

Регрессионный анализ с помощью вышеупомянутых производственных функций $\hat{z} = x^{\hat{\alpha}_1} y^{\hat{\alpha}_2}$ и $\hat{w} = x^{\hat{\beta}_1} y^{\hat{\beta}_2}$ показал, что:

1. по уровню детерминации и уровню значимости эти производственные функции адекватно моделируют тримерную регрессионную связь между тройками параметров: $x(\text{ВД}_\text{рЛЖ}) - y(\text{ВД}_\text{рПЖ}) - z(\text{АД}_\text{с})$ и $x(\text{ВД}_\text{рЛЖ}) - y(\text{ВД}_\text{рПЖ}) - w(\text{АД}_\text{д})$.

2. значения коэффициентов эластичности (табл. 2) как по ЛЖ, так и по ПЖ в контроле и на 1, 2 неделях эксперимента положительны и находятся между нулем и единицей. Это свидетельствует о том, что в контроле и на 1, 2 неделях эксперимента увеличение внутрижелудочкового давления ЛЖ или ПЖ вызывает соответствующее увеличение системного АД. На 4 и 6 неделях эксперимента увеличение внутрижелудочкового давления ПЖ не влияет (нулевое увеличение) на увеличение ни $\text{АД}_\text{с}$, ни $\text{АД}_\text{д}$.

3. наблюдается равенство сумм коэффициентов эластичности 1 ($\hat{\alpha}_1 + \hat{\alpha}_2 = 1$, $\hat{\beta}_1 + \hat{\beta}_2 = 1$) как в контроле, так и для всех сроков эксперимента. Это свидетельствует о том, что в условиях экспериментальной артериальной гипертензии,

производственные функции параметров системного АД изменяются в той же пропорции, что и параметры ВД левого и правого желудочков. Последнее означает, что при экспериментальной артериальной гипертензии производственные функции параметров системного АД сохраняют постоянный эффект «производства», т.е. являются производственными функциями Кобба-Дугласа [3].

4. во все сроки эксперимента сохраняется общая тенденция большего вклада ЛЖ (по сравнению с ПЖ) в формирование как систолического $z(AD_c)$, так и диастолического давления $w(AD_d)$.

References

- [1] Зотов А.К. Системный анализ сердечно-сосудистой гемодинамики кроликов при сезонных изменениях.//Вестник новых медицинских технологий, т. XVI, №1, с. 199 – 200, 2009.
- [2] Зотов А.К., Фролов В.А., Зотова Т.Ю. Системный анализ сердечно-сосудистой гемодинамики кроликов при дифтерии в эксперименте.// Бюллетень экспериментальной биологии и медицины, №9, с.266-269, 2009.
- [3] Доугерти В.А. Введение в эконометрию. – М.: ИНФРА-М, 2001. – 402 с.
- [4] Лакин Г.Ф. Биометрия. – М.: Высшая школа, 1980. – 293 с.

Роль терморцепторов в явлении “afterdrop”

Лыжко Е.В.^(1,2)

Институт математических проблем биологии РАН⁽¹⁾,
Пушкинский государственный университет⁽²⁾
lyzko@yandex.ru

Abstract

Temperature dynamics of the core and the skin was studied using a mathematical model based on the equations of thermal balance. Lead role of thermoreceptors was simulated warmth sense function. This allowed to simulate the phenomenon afterdrop. A comparative analysis of theoretical and experimental data showed the same temperature dynamics during warming, as well as the dependence of the afterdrop length for rewarming techniques. Afterdrop was observed during rewarming for the unsteady state (significant cooling of skin and slight cooling of core). Afterdrop wasn't observed for the steady state of core and skin.

1 Введение

В настоящее время по-прежнему актуальной остается проблема наискорейшего восстановления температуры организма после охлаждения. Однако необходимо различать охлаждение, связанное с подавлением терморегуляторных функций (охлаждение во время хирургической операции) и несвязанное (охлаждение в результате несчастного случая).

Данная работа посвящена анализу динамики температур “ядра” и “оболочки” организма, находящегося в состоянии гипотермии, в частности изучению явления “afterdrop”, которое возникает в результате быстрого охлаждения и для него характерно продолжающееся понижение внутренних температур организма вначале согревания. Однако “afterdrop” характерен не для любой гипотермии. Как отмечено в работе [4] “afterdrop” наблюдается при быстром охлаждении, продолжающемся непродолжительное время. Т.е. при таком охлаждении значительному понижению температуры “оболочки” соответствует незначительное понижение температуры “ядра”. При продолжительном охлаждении температура “ядра” и “оболочки” перестают изменяться, достигая некоторых стационарных значений, соответствующих условиям окружающей среды. При такой гипотермии во время согревания “afterdrop” не возникает. Экспериментальные данные показали [1,2,3,6], что согревание различными способами влияет на продолжительность и величину падения

температуры “ядра”. Чтобы определить наиболее оптимальный способ, необходимо изучить физические и физиологические особенности этого явления.

Попытка объяснить данное явление на основе учета инерционных свойств была предпринята в работе [4] с помощью физической модели (мешок с гелем). Однако при более детальном рассмотрении видно, что нельзя проводить прямой аналогии между этими явлениями в живом и неживом объектах. В первую очередь это несовпадение температурных градиентов.

2 Математическая модель

В данной работе была сделана попытка объяснить явление “afterdrop” как результат управляющей деятельности кожных терморцепторов. И это является вполне логичным, так как при изменении температуры окружающей среды из всех многочисленных сигналов, поступающих в центр терморегуляции, первыми изменяются сигналы, поступающие от кожных терморцепторов (управление по возмущающему фактору). Для этой цели была использована имитационная модель организма человека в виде двухслойной структуры (“ядро” и “оболочка”), представленная на рисунке 1 с соответствующими потоками тепла между слоями из внешней и во внешнюю среду.

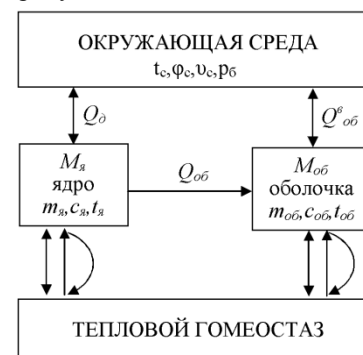


Рисунок 1. Схема потоков тепла в организме человека для записи уравнений теплового баланса

Параметрами модели являются: температура окружающей среды t_c , относительная влажность воздуха ϕ_c , скорость движения воздуха υ_c , барометрическое давление p_b . Для схемы потоков

тепла была записана система уравнений теплового баланса:

$$\begin{cases} c_{я} m_{я} \frac{dt_{я}}{d\tau} = M_{я} - Q_{\partial} - Q_{об} \\ c_{об} m_{об} \frac{dt_{об}}{d\tau} = Q_{об} + M_{об} - Q_{об}^e \end{cases} \quad (1)$$

где $c_{я}$, $c_{об}$ – теплоемкость “ядра” и “оболочки”; $m_{я}$, $m_{об}$ – масса “ядра” и “оболочки”; $t_{я}$, $t_{об}$ – температура “ядра” и “оболочки”; $M_{я}$, $M_{об}$ – теплопродукция внутренних источников “ядра” и “оболочки”; Q_{∂} – тепловой поток в окружающую среду через органы дыхания; $Q_{об}$ – тепловой поток от “ядра” к “оболочке”; $Q_{об}^e$ – тепловой поток от “оболочки” в окружающую среду (испарение, излучение, конвекция); τ – время воздействия.

Величины $m_{я}$, $m_{об}$, $M_{я}$, $M_{об}$ являются функциями $t_{об}$ соответствующей области, потоки тепла из внешней и во внешнюю среду зависят от $t_{я}$, $t_{об}$ и характеристик окружающей среды. Управляющая роль кожных терморецепторов была промоделирована функцией теплоощущения $S(t_{об})$, регулирующей процессы теплопродукции и теплоотдачи. При этом теплопродукции соответствуют процессы метаболизма, а теплоотдаче – величина кровообращения. Функциональные зависимости всех параметров системы (1) приведены в работе [5].

3 Полученные результаты

Система уравнений (1) решена численно в программе MathCad. На рисунке 2 представлены расчетные данные. Зависимости $t_{я}(\tau)$ (рис. 2а) и $t_{об}(\tau)$ (рис. 2б) получены в результате решения системы уравнений (1). Зависимостей 1 и 3 рассчитаны для одинаковой $t_c=31^\circ\text{C}$, но при различных начальных условиях, характеризующих различную степень гипотермии. Начальные условия ($t_{я}(0)=36.4^\circ\text{C}$ и $t_{об}(0)=28^\circ\text{C}$) для зависимости 1 не являются стационарными и соответствуют начальной стадии охлаждения, характеризующейся значительным понижением $t_{об}$ и незначительным понижением $t_{я}$. Для зависимости 3 начальные условия ($t_{я}(0)=33.9^\circ\text{C}$ и $t_{об}(0)=27.5^\circ\text{C}$) являются стационарными и соответствуют длительному нахождению человека при $t_c=18^\circ\text{C}$. Зависимость 2 рассчитана для $t_c=18^\circ\text{C}$ и начальных условиях ($t_{я}(0)=37.2^\circ\text{C}$ и $t_{об}(0)=34.4^\circ\text{C}$), являющихся стационарными значениями и соответствующих длительному нахождению человека при $t_c=31^\circ\text{C}$. Таким образом, зависимости 1, 3 описывают процесс нагревания, а зависимость 2 – процесс охлаждения.

Отличительной особенностью зависимости 1 (рис. 2а) является наличие “afterdrop”. Для этой зависимости в качестве начальных значений $t_{я}(0)$ и $t_{об}(0)$ были использованы значения ректальной и средневзвешенная температура (CBT) кожи соответственно, приведенные в работе [6].

Характерной особенностью экспериментальных данных было также продолжающееся понижение ректальной температуры при переходе в комфортные условия после кратковременного охлаждения. При переходе в комфортные условия СВТ кожи монотонно возрастала. Динамика изменения экспериментальных температур, СВТ кожи и ректальной [6], соответствовала расчетным значениям $t_{я}(\tau)$ и $t_{об}(\tau)$ (зависимости 1 рис. 2а, б).

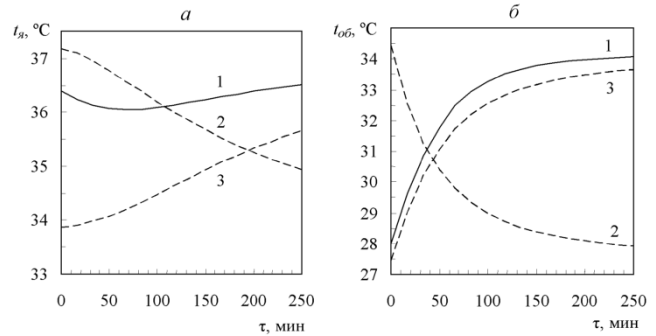


Рисунок 2. Временные зависимости $t_{я}$ и $t_{об}$, полученные с помощью модели:

1, 3 – $t_c = 31^\circ\text{C}$; 2 – $t_c = 18^\circ\text{C}$

Данная модель также показывает хорошее согласование экспериментальных и теоретических данных при моделировании различных способов согревания. Т.е. быстрое восстановление температуры кожи (моделировалось согревание в воде) позволяет уменьшить величину и продолжительность “afterdrop”.

Работа поддержана РФФИ, гранты 08-07-00353, 10-01-00609

References

- [1] M.L. Collis, A.M. Steinman, R.D. Chaney. Accidental hypothermia: an experimental study of practical rewarming methods. In *Aviat. Space Environ. Med.*, volume 48(7), pages 625-632, 1977.
- [2] G.G. Giesbrecht, G.K. Bristow. A second postcooling afterdrop: more evidence for a convective mechanism. In *Aviat. Space Environ. Med.*, volume 69(1), pages 17-22, 1998.
- [3] G.G. Giesbrecht, G.K. Bristow. A second postcooling afterdrop: more evidence for a convective mechanism. In *J.Appl.Physiol.*, volume 73(4), pages 1253-1258, 1992.
- [4] P. Webb. Afterdrop of body temperature during rewarming. In *J.Appl.Physiol.*, volume 60(2), pages 385-390, 1986.
- [5] В.А. Голиков, А.Ф. Бурденко, Ю.М. Цюпко. Математическое моделирование процессов теплообмена организма человека с окружающей средой. *Судовые энергетические установки: науч.-техн. сб.*, выпуск № 9, с. 104–115, 2003.
- [6] В.С. Кошечев. Физиология и гигиена индивидуальной защиты человека от холода. М.: Медицина, 1981

О функциональной роли гетерогенности синоатриального узла при формировании единого ритма сердца

Мазуров М.Е.

Московский государственный университет экономики, статистики и информатики

Email: Mazurov37@mail.ru

Abstract

The main mechanism of cardiac rhythm generation and competitor dynamics of this process is elucidated. The recent data on heterogeneous structure of sinoatrial node and its influence on cardiac common rhythm generation is presented.

Проблема формирования единого ритма сердца вызывала интерес физиологов и математиков более 100 лет. Значительный импульс исследованиям по этой проблеме был получен после работ Н. Винера - А. Розенблюта, И.М. Гельфанда. Согласно теории Винера - Розенблюта, Гельфанда единый ритм клеточного ансамбля пейсмейкеров должен определяться клеткой с наивысшим ритмом. Экспериментальные исследования физиологов не подтвердили выводы математиков, что значительно увеличило интерес к проблеме.

Для исследования формирования единого ритма одной пейсмейкерной клетки синоатриального узла (СУ) сердца была использована система нелинейных дифференциальных уравнений

$$\varepsilon \frac{dx}{dt} = f(x, y) + R_1(x, y), \quad \frac{dy}{dt} = g(x, y) + R_2(x, y),$$

где x, y - переменные, описывающие быстрые и медленные изменения потенциала клетки - пейсмейкера синоатриального узла; f, g - нелинейные функции; R_1, R_2 - функции, характеризующие связь пейсмейкеров. Первое уравнение описывает быстрые переменные, второе - медленные.

В качестве точечных моделей синоатриального узла использовались: 1. Модель Ван дер Поля - Фитцхьюга [1]; 2. Модель Ходжкина - Хаксли в автоколебательном режиме; 3. Модель Жанга и др. [2]. Получены количественные выражения для величины единого ритма пары пейсмейкеров и многопейсмейкерного СУ[3-9]. Полученные теоретические результаты качественно хорошо согласуются с основными экспериментальными данными [11].

Исследована нелинейная динамика формирования единого ритма синоатриального узла сердца в одномерных, двумерных и трехмерных распределенных системах. Для проверки имеющихся гипотез был выполнен численный эксперимент, идея которого схематически иллюстрируется рисунком. Рассматривается взаимодействие 10 тысяч пейсмейкеров; такое количество пейсмейкеров характерно для синоатриального узла кролика. Система описывается 20 тыс. нелинейных дифференциальных уравнений Бонхоффера - Ван дер Поля - Фитцхьюга.

Динамика развития возбуждения и формирования единого ритма в автоволновой активной среде - в двумерном синоатриальном узле из 10 тысяч пейсмейкеров с группами из различного числа пейсмейкеров с различными частотами показана на рис. 1. Цифрами около областей указано текущее время процесса.

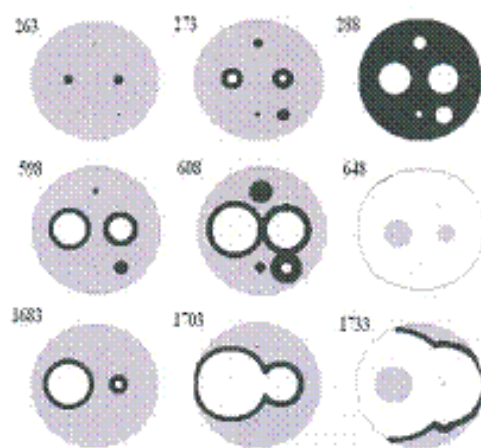


Рис. 1. Динамика развития возбуждения и формирования единого ритма в автоволновой активной среде - в синоатриальном узле из 10 тысяч пейсмейкеров с группами из различного числа пейсмейкеров с различными частотами. Цифрами около областей указано текущее время процесса. Нарастающее возбуждение показано черным

цветом, спад возбуждения показан белым цветом, состояние вблизи покоя серым

Исследована миграция водителей ритма синоатриального узла при парасимпатической регуляции. Изложены данные по исследованию миграции водителя ритма в двумерном распределенном синоатриальном узле, подкрепленные результатами оптического картирования миграции водителя ритма, полученные школой академика Л. В. Розенштрауха [12] .

Изложены экспериментальные и теоретические исследования по управлению сформированным единым ритмом сердца через парасимпатическую и адренергическую нервную системы от центральной нервной системы. Изложенные данные по исследованию управления сформированным единым ритмом в синоатриальном узле сердца, подкреплены экспериментальными результатами школы академика В.Л. Покровского[10].

Длительное время считали, что СА небольшое однородное образование правильной формы. Исследования последних лет показали, что СУ – сложное гетерогенное образование, что иллюстрируется рис. 2.

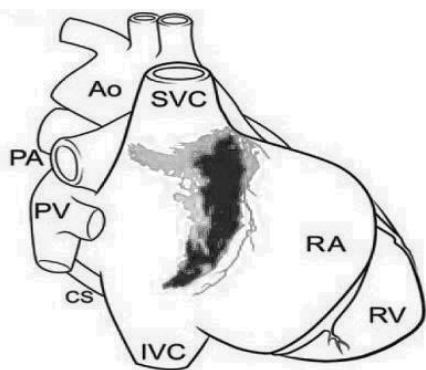


Рис. 2. Гетерогенная структура синоатриального узла кролика. Центральная часть узла показана черным цветом, периферическая часть серым

Гетерогенность СУ, согласно современным данным включает: 1) наличие неоднородной связи между пейсмекерными клетками, слабый различающийся характер связи, зависящий от местоположения; 2) неоднородность электрических свойств пейсмекерных клеток от центра к периферии, например, уменьшение частоты колебаний от центра к периферии; 3) вкрапление клеток предсердного типа в ткань узла, количество которых увеличивается от центра к периферии; 4) сложная геометрическая форма СУ и его границ с предсердиями, неоднородный объемный характер СУ. Этот список можно продолжить.

Неоднородность СУ усиливает эффективность его основной функции - формирования единого

ритма, она позволяет: 1) обеспечить активную конкурентную динамику формирования единого ритма группами элементов с наивысшим ритмом и наибольшей электрической мощностью клеток; 2) обеспечить оптимальные условия для управления единым ритмом со стороны центральной нервной системы через парасимпатическую и симпатическую нервную системы; 3) увеличить макроскопическую электрическую мощность СУ, необходимую для надежного и устойчивого возбуждения предсердий.

References

- [1] Fitz Hugh R. Computation of impulse initiation and saltatory conduction in a myelinated nerve fiber// Biophysical. J. 1962. V. 2. P. 11-21.
- [2] Zhang H., Holden A.V., Kodama I., Honjo H., Lei M., Varghese T. , Boyett M.R. Mathematical models of action potentials in the periphery and center of rabbit sinoatrial node //A. J. Physiol. 2000. V. 279. P.397-421
- [3] М.Е. Мазуров Нелинейная синхронизация и ритмогенез в электровозбудимых системах сердца /Дисс. д. ф.-м. н. Пушино. 2007.
- [4] М.Е. Мазуров Геометрический метод исследования решения диофантовых неравенств в задачах о синхронизации релаксационных колебаний //ИВУЗ. сер. Математика. 1989. №7. С. 13-20.
- [5] М.Е. Мазуров. Решение диофантовых неравенств в задачах о синхронизации релаксационных колебаний //ЖВМ и МФ АН СССР. 1991. Т. 31. №11.
- [6] М.Е. Мазуров Ритмогенез в синоатриальном узле сердца //Биофизика. 2006. Том 51. №6. С. 1092
- [7] М.Е. Мазуров Идентификация нелинейных моделей биологических мембран по методу фиксированного напряжения //Биофизика. 2006. Т. 51. №6. С. 1019-1025..
- [8] М.Е. Мазуров К проблеме формирования единого ритма в синоатриальном узле сердца //Биофизика. 2009. Том 54, выпуск 1. С. 81-88.
- [9] М.Е. Мазуров Управление единым ритмом сердца //Биофизика. 2009. Том 54, выпуск 1. С. 89-96.
- [10] В.М. Покровский Формирование ритма сердца в организме человека и животных //Краснодар: "Кубань-Книга". 2007. 143 с.
- [11] Verheijck E.E., Wilders R., Joiner R. et all. Pacemaker synchronization of electrically coupled rabbit sinoatrial node cells // J. Gen. Physiol. 1998. V. 111. P. 95-112.
- [12] Абрамочкин Д.В. Миграция водителя ритма в синоатриальном узле и её механизмы// Дисс. на соискание ученой степени кандидата биологических наук. Москва. МГУ им М.В. Ломоносова. 2009. 158 с.

Метод псевдосредних функций в теории RISM. Температурная зависимость гидратации пептида окситоцина

Тихонов Д.А., Соболев Е.В.

Институт математических проблем биологии РАН
dmitry.tikhonov@gmail.com

Abstract

The temperature dependence of the Gibbs free energy oxytocine peptide in solutions was studied by the method of pseudo averages in the frame of the RISM approach. The different functionals of the Gibbs free energy was used. It is shown that linear dependence of the excess chemical potential on the temperature obeyed the term proportional of the square of total correlation function.

1 Введение

Часто требуется весьма существенная статистика для того, чтобы учесть все возможные конформационные изменения молекулы, тем самым верно определив ее состояние в целом. Расчеты методом RISM [1] для каждой возможной конфигурации могут оказаться вычислительно громоздкими. Поэтому в работе [6] был предложен метод псевдосредних. В данной работе мы исследуем этим методом температурную зависимость свободной энергии окситоцина, используя различные функционалы для нее.

2 Функционалы свободной энергии

Свободная энергия Гиббса E молекулы в растворе есть сумма избыточного химического потенциала сольватации $\Delta\mu$ и молекулярно-механической энергии E_{MM} :

$$E = \Delta\mu + E_{MM} \quad (1)$$

В методе RISM выражения для $\Delta\mu$ есть сумма парциальных вкладов от каждого атома молекулы раствора, где каждый вклад вычисляется по тем же формулам, что и в теории Орнштейна-Цернике для одноатомных жидкостей. Существуют различные приближенные теории для оценок $\Delta\mu$. Простейшим выражением является приближение гауссовых флуктуаций (GF) [1]:

$$\Delta\mu^{GF} = \frac{\rho}{2\beta} \sum_{ia} \int 4\pi r^2 \times \\ \times [-2c_{ia}(r) - h_{ia}(r)c_{ia}(r)] dr, \quad (2)$$

где h_{ia} и c_{ia} – полные и прямые корреляционные функции атом раствора – атом растворителя, ρ – плотность растворителя, β – обратная температура.

Наиболее известное выражение для $\Delta\mu$ – формула Сингера-Чандлера (SC) [3]:

$$\Delta\mu^{SC} = \Delta\mu^{GF} + \frac{\rho}{2\beta} \sum_{ia} \int 4\pi r^2 h_{ia}^2(r) dr. \quad (3)$$

Использовались еще несколько выражений, таких как выражение, полученное в методе парциальных волн (PW) [4], и выражения с различными репульсивными поправками (RBC1) и (RBC2), учтенными по методу теории возмущений [5]. Эти же поправки, но учтенные путем термодинамического интегрирования, обозначены (INT1) и (INT2).

3 Результаты

Были сгенерированы 183 молекулярно-динамических траектории различной длины. Для каждого окружения пептида (вакуумное, водное окружение, неявный учет воды по методу MM/GBSA) были проведены молекулярно-динамические расчеты для набора температур из диапазона 280-340 К с шагом в 1 К. Во всех случаях, начальная конфигурация вначале минимизировалась, а затем медленно разогревалась до заданной температуры. Длины траекторий были одинаковыми для разных температур. Для пептида в вакуумном окружении – 15 нс (75000 фреймов), для обобщенного Борна – 11 нс (55000 фреймов), для явного учета водного окружения – 1.25 нс (6250 фреймов).

Первые 1250 фреймов каждой траектории использовались для проверки близости средних и псевдосредних корреляционных функций и термодинамических величин. Результаты сравнения для пептида в вакууме приведены на рис. 1. Сравнение показывает, что разница средних и псевдосредних полных корреляционных функций меньше, чем дисперсия среднего значения, а также симпатное поведение псевдосредних и средних

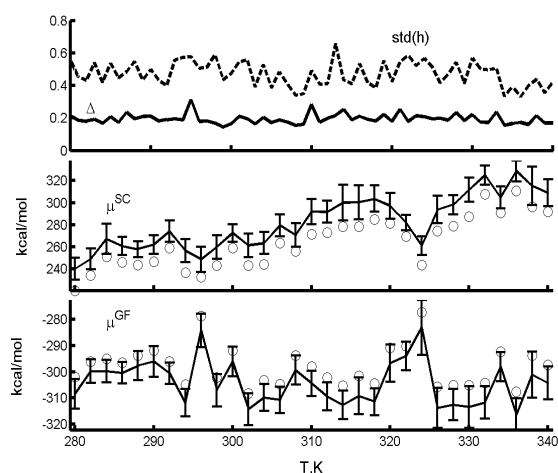


Рис. 1. Сравнение средних и псевдосредних величин. Верхний график – максимальное отклонение среднего от псевдосреднего в сравнении с дисперсией для полной корреляционной функции. Средний график – сравнение средней и псевдосредней μ^{SC} . Нижний график – сравнение средней и псевдосредней μ^{GF} .

величин $\Delta\mu$ в зависимости от температуры. При этом разница между средними и псевдосредними

величинами находится в пределах статистической погрешности. Из рисунка видно, что даже в случае такой небольшой статистики, $\Delta\mu$ для разных методов его расчета обнаруживает принципиально отличающееся поведение. Избыточный химический потенциал в методе (GF) не зависит от температуры, в то время как $\Delta\mu^{SC}$ зависит от температуры линейно. Расчеты методом псевдосредних с учетом всей статистики подтверждают этот вывод.

Также линейно зависит от температуры химический потенциал с репульсивными поправками (RBC1), (RBC2), (INT1), (INT2). Химический потенциал в методе парциальных волн (PW) не зависит от температуры, также как и (GF).

Молекулярно механическая энергия линейно растет с температурой. Для полной статистики оценивались коэффициенты линейной регрессии свободной энергии Гиббса $E(T)$. По коэффициентам мы получили разницы свободной энергии пептида между конденсированным окружением и вакуумным состоянием. На рис. 2 приведены эти разницы для модели Борна. Видно, что во всей исследуемой области температур свободная энергия вакуумного состояния пептида оказывается выше, чем пептида в конденсированном окружении. Для случая явного водного окружения картина качественно не

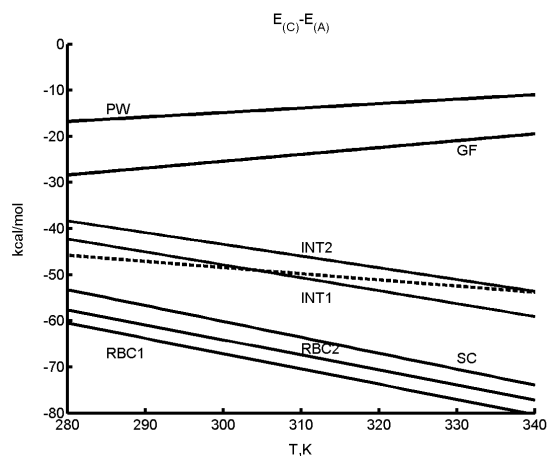


Рис. 2. Свободная энергия гидратации окситоцина относительно состояния в вакууме.

отличается. Пунктиром на рисунке отмечены оценки по методу Борна.

Можно сделать вывод о том, что член пропорциональный h^2 принципиально меняет температурную зависимость свободной энергии.

References

- [1] D.Chandler, H.C.Andersen. Optimized cluster expansions for classical fluids. II. Theory of molecular liquids. In *Journal of Chemical Physics*, volume 57(5), pages 1930-1937, 1972.
- [2] D.Chandler, Y.Singh, D.Richardson. In Excess electrons in simple fluids. I. General equilibrium theory for classical hard sphere solvents. In *Journal of Chemical Physics*, volume 81(4), pages 1975-1982, 1984.
- [3] S.J.Singer, D.Chandler. Free energy functions in the extended RISM approximation. In *Molecular Physics*, volume 55(3), pages 621-625, 1985.
- [4] S.Ten-no. Free energy of solvation for the reference interaction site model: Critical comparison of expressions. In *Journal of Chemical Physics*, volume 115(8), pages 3724-3731, 2001.
- [5] A.Kovalenko, F.Hirata. Hydration free energy of hydrophobic solutes studied by a reference interaction site model with a repulsive bridge correction and a thermodynamic perturbation method. In *Journal of Chemical Physics*, volume 113(7), pages 2793-2805, 2000.
- [6] Д.А.Тихонов, Е.В.Соболев. Псевдосредние в методе интегральных уравнений теории жидкостей в приближении RISM. *Математическая биология и бионформатика: III Международная конф., г. Пущино, 10-15 октября 2010 г.: Доклады / Под ред. В.Д. Лахно, с. xx-xx, 2010.*

Псевдосредние в методе интегральных уравнений теории жидкостей в приближении RISM

Тихонов Д.А., Соболев Е.В.

Институт Математических проблем Биологии РАН

dmitry.tikhonov@gmail.com

Abstract

A new method for studying macromolecules in solution is proposed. The method combines integral equation theory and molecular dynamics approach. It is shown that computational efficiency of the method several hundred times better than standard one with the minor discrepancy of the results.

1 Введение

В системе уравнений RISM [1] записанной в пределе бесконечного разбавления [2] предложено усреднять матрицу внутримолекулярных корреляций раствора по всем его геометрическим конфигурациям, полученным в рамках метода молекулярной динамики. Это позволяет с одной стороны учесть крупномасштабные флуктуации, а с другой стороны легко получить термодинамику сольватации, в частности свободные энергии Гиббса. Можно провести аналогию между предложенным подходом и методом PRISM [3]. В методе PRISM усреднение проводится в рамках априорных предположений о гауссовом распределении атомов полимерной цепочки в пространстве. Это позволяет свести систему нелинейных уравнений относительно корреляционных функций к единственному уравнению, поскольку предполагается, что все атомы в полимерной цепи неразличимы. Предлагаемый подход оставляет детальное описание атом-атомных взаимодействий. Однако каждое из них усредняется, исходя из распределения атом-атомных расстояний. Решения такого уравнения мы называем псевдосредними функциями.

2 Основные уравнения

Уравнения RISM для псевдосредних функций в случае бесконечно разбавленного водного раствора N-атомной молекулы могут быть записаны в виде:

$$\langle \mathbf{h} \rangle = (\mathbf{V} \otimes \bar{\mathbf{W}}) \langle \mathbf{c} \rangle, \quad (4)$$

Где $\langle \mathbf{h} \rangle$ и $\langle \mathbf{c} \rangle$ вектора, элементами которых являются Фурье-образы полных и прямых атом-

атомных псевдосредних корреляционных функций. Символом $\otimes$ обозначено прямое произведение матриц растворителя $\mathbf{V}$ и матрицы усредненных по траектории внутримолекулярных корреляционных функций $\bar{\mathbf{W}}$. В матрицу растворителя $\mathbf{V}$ входят межатомные расстояния в молекуле воды, плотность воды и полные атом-атомные корреляционные функции водного растворителя [4]. Элементы матрицы $\bar{\mathbf{W}}$ выражаются через гистограммы распределения межатомных расстояний $P_{ij}(l)$, согласно формуле:

$$\bar{w}_{ij}(k) = \int_0^{+\infty} P_{ij}(l) \frac{\sin kl}{kl} dl. \quad (5)$$

Отметим, что в случае, когда $P_{ij}(l) = \delta(l - L_{ij})$, где L_{ij} – межатомные расстояния в молекуле, уравнение для псевдосредних переходит в обычное уравнение RISM.

Вторым уравнением, которое связывает корреляционные функции в координатном пространстве и обеспечивает разрешимость системы относительно корреляционных функций, является уравнение замыкания. Например, для гиперцепного замыкания уравнение имеет вид:

$$\langle c \rangle_{ix} = \exp(-\beta u_{ix} + \langle \gamma \rangle_{ix}) - 1 - \langle \gamma \rangle_{ix}. \quad (6)$$

здесь $\beta = 1/KT$ – обратная температура, u_{ix} – потенциал атом-атомного взаимодействия i-того атома раствора с x-тым атомом молекулы воды. Функция $\langle \gamma \rangle_{ix} = \langle h \rangle_{ix} - \langle c \rangle_{ix}$ – это так называемая непрямая корреляционная функция.

Показано, что разница между псевдосредними и средними полными корреляционными функциями $\Delta = \langle \mathbf{h} \rangle - \bar{\mathbf{h}}$ выражается простой формулой:

$$\Delta = (\mathbf{V} \otimes \bar{\mathbf{W}}) \bar{\mathbf{c}} - \bar{\mathbf{h}}. \quad (7)$$

где $\bar{\mathbf{h}}$ и $\bar{\mathbf{c}}$ – средние по траектории полные и прямые корреляционные функции.

3 Результаты

Выполнены молекулярно-динамические расчеты двух коротких пептидов окситоцина (134 атома) и мет-энкефалина (75 атомов) в различных средах имитирующих растворитель. Сгенерированы 250 пс

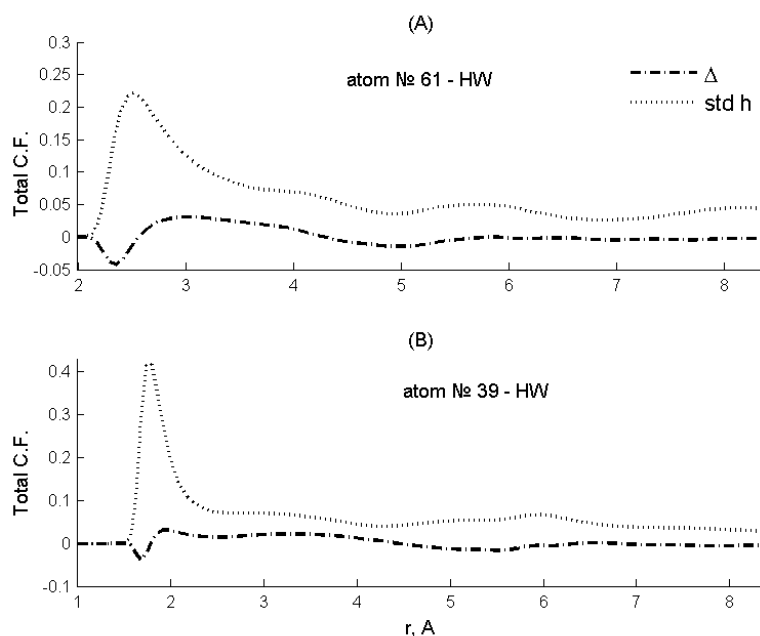


Рис. 1. Разница между средними и псевдосредними величинами Δ в сравнении с дисперсией для полной корреляционной функции: (A) – пептид мет-енкефалин в вакууме, (B) – в водной среде. Выбраны атом-атомные функции с максимальным интегралом $|\Delta|$.

траектории пептидов. В каждой из 1250 выбранных точек траекторий решалось обычное уравнение RISM с целью найти корреляционные функции атом пептида – атом воды. Вычислялись средние значения функций по траектории, и их дисперсии. Затем, по тем же точкам траектории мы вычисляли средние матрицы $\bar{W}$ и решали уравнение (4). Результаты сравнивались со средними значениями. Во всех случаях невязка Δ оказывалась меньше, чем дисперсия среднего. На рис. 1 приведено сравнение для пептида мет-енкефалина в вакууме и в явном водном окружении. Видно, что невязка значительно меньше дисперсии.

Вычислялись термодинамические величины сольватации исходя из псевдосредних корреляционных функций, и сравнивались их со средними значениями термодинамических величин полученных из уравнений RISM обычным образом. Показано, что для свободной энергии Гиббса разница зависит от выбранного функционала свободной энергии, однако все отклонения лежат в пределах дисперсии средних.

4 Выводы

Метод псевдосредних на два порядка быстрее, чем стандартные подходы. Поэтому он может оказаться единственным применимым методом в

тех случаях, когда необходимо учитывать большие объемы молекулярно-динамической информации. Типичные задачи – исследование конформационно подвижных белков, либо скрининг большого количества разных траекторий, например при изучении температурных зависимостей.

References

- [1] D.Chandler, H.C.Andersen. Optimized cluster expansions for classical fluids. II. Theory of molecular liquids. In *Journal of Chemical Physics*, volume 57(5), pages 1930-1937, 1972.
- [2] F.Hirata, P.J.Rosky, B.M.Pettitt. The interionic potential of mean force in a molecular polar solvent from an extended RISM equation. In *Journal of Chemical Physics*, volume 78(6), pages 4133-4144, 1983.
- [3] K.S.Schweizer, J.G.Gurro. Integral-Equation Theory of the Structure of Polymer Melts. In *Physical Review Letters*, volume 58(3), pages 246-249, 1987
- [4] Д.А.Тихонов. Метод интегральных уравнений теории жидкостей для изучения гидратации макромолекул. В книге *Компьютеры и суперкомпьютеры* / Под ред. В.Д.Лахно и М.Н.Устинова. – Москва-Ижевск: Институт компьютерных исследований, 2002, 528 стр., с. 209-233.

О необходимости количественного анализа спонтанного дрейфа автоволнового вихря в 2D-среде

Москаленко А.В.

Учреждение Российской академии наук
Институт математических проблем биологии РАН
cardio@avmoskalenko.ru

Abstract

The importance of quantitative analysis of autowave solutions is discussed by the example of the study of peculiarities of autowave behaviour in the bifurcation spot.

1 Автоволновые феномены

Важность автоволновых процессов в настоящее время широко осознается; особое место им отведено в кардиологии [4]. Одним из типичных автоволновых решений является *ревербератор* — автоволновой вихрь в 2D-среде, т.е. вращающаяся автоволна со свободным концом, которая часто принимает вид вращающейся спирали; отличие ревербератора от *спиральной автоволны*, достаточно подобно обсуждалось ранее [4]. Характерным поведением ревербератора является меандр — спонтанный дрейф ревербератора в однородной среде. Недавно [4, 5] был получен *автоволновой серпантин* — явление спонтанного торможения спонтанного дрейфа ревербератора в однородной среде.

Автоволновые модели показывают достаточно разнообразное поведение, причем порой даже малого изменения параметров системы достаточно для радикального изменения поведения системы, как это было показано, например, в недавних работах по изучению системы свертывания крови [3]. Эти обстоятельства делают построение параметрического портрета необходимым первым шагом при использовании той или иной автоволновой модели. Однако следует отметить некоторую странность исторически сложившейся ситуации: при исследовании автоволновых решений преобладает традиция качественного визуального анализа получаемых решений. Тем не менее недавняя практика показала [5], что использование количественных методов может оказаться весьма продуктивным: так, лишь благодаря количественному анализу дрейфа ревербератора удалось доказать, что меандр и серпантин являются существенно различными автоволновыми решениями, и в исследуемой части пространства параметров изучаемой модели была достаточно точно определена граница между меандром и серпантинном.

Автоволновой серпантин представляет собой явление так называемой *бифуркационной памяти* (БП). Под БП понимают необычный переходный

процесс, иногда наблюдаемый в поведении системы, находящейся вблизи бифуркационной границы [2, 3]. Хотя существует несколько точек зрения на природу этого феномена, тем не менее, признается то обстоятельство, что БП играет весьма существенную роль в практике, ибо система в этом случае теряет управляемость [2]. Поэтому с практической точки зрения важно выяснить ширину пограничной зоны, в которой наблюдается этот феномен. Для определенности, множество точек вдоль бифуркационной границы пространства параметров, где феномен БП прекращается, назовем *линией выхода из бифуркационной зоны*. Под *пограничной бифуркационной зоной* будем понимать множество точек между бифуркационной границей и линией выхода из бифуркационной зоны. К сожалению, использованный нами прежде [5] метод количественного анализа спонтанного дрейфа ревербератора в 2D-среде не позволил выявить ширину пограничной бифуркационной зоны.

В данной работе предлагается количественный подход, позволяющий проанализировать особенности поведения автоволновой модели в пограничной бифуркационной зоне и определить линию выхода из бифуркационной зоны.

2 Методы и результаты

2.1 Исследуемая модель

В работе используется предложенная Алиевым и Панфиловым [1] простая двухкомпонентная модель возбудимой среды:

$$\frac{\partial u}{\partial t} = \Delta u - ku(u - a)(u - 1) - uv,$$

$$\frac{\partial v}{\partial t} = \varepsilon(u, v)(-v - ku(u - a - 1)),$$

$$\varepsilon(u, v) = \varepsilon_0 + \frac{\mu_1 v}{u + \mu_2}.$$

Моделирование выполнялось для двумерных сред 128×128 с граничными условиями Неймана. Все условия моделирования и способ определения траектории кончика ревербератора были подробно описаны ранее [5, 6]. Отметим хорошо известное обстоятельство, что, хотя от способа определения кончика ревербератора положение его траектории

может несколько изменяться, тем не менее, ее существенные особенности сохраняются: в частности, сохраняется периодичность изменений ее кривизны.

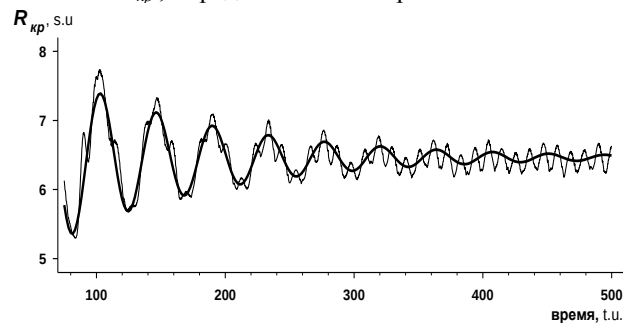
2.2 Анализ движения ревербератора

При анализе экспериментальных зависимостей одним из стандартных подходов является параметризация кривой, соответствующей результатам измерений. Прежде чем приступить к параметризации траекторий, нам представилось естественным преобразовать динамику двух пространственных переменных в динамику одной переменной — кривизны траектории. Измерение кривизны траектории осуществлено следующим образом. Для каждой точки траектории участок вблизи этой точки был аппроксимирован дугой окружности при помощи МНК. Для аппроксимации выбирался участок траектории, соответствующий фиксированному интервалу времени (окно аппроксимации). В результате для каждого момента времени получена оценка радиуса кривизны траектории.

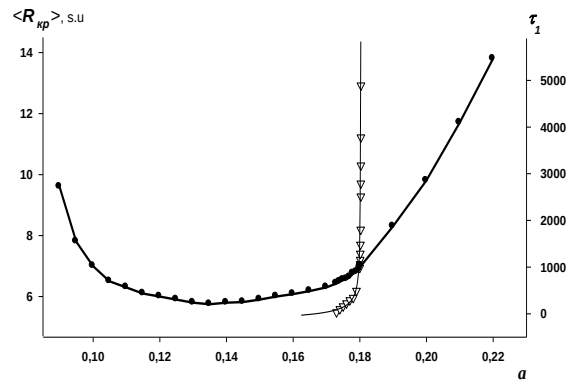
Далее соответствующие серпантину [5] зависимости радиуса кривизны траектории, $R_{кр}$, от времени, t , аппроксимированы следующей функцией:

$$R_{кр} = k_0 + k_1 \cdot e^{-x/k_2} \sin(2\pi \cdot t / k_3 + k_4),$$

где k_0 дает оценку среднего радиуса кривизны, $\langle R_{кр} \rangle$, k_1 — начальная амплитуда ее колебаний, k_2 — характерное время затухания колебаний и k_3 — период колебаний кривизны. Аппроксимация осуществлена при помощи стандартных средств, предоставляемых программой SigmaPlot. На рисунке показан пример такой аппроксимации (толстая линия) для $a = 0.170$; тонкой линией показана величина $R_{кр}$, определенная экспериментально.



В результате были получены зависимости параметров аппроксимирующей функции от изменяемого в ходе вычислительных экспериментов параметра модели a . Наиболее интересной оказалась зависимость $\langle R_{кр} \rangle$ от a ; она представлена на следующем рисунке, на котором темными кружками отмечена $\langle R_{кр} \rangle$, а светлыми треугольниками показана найденная ранее [5] зависимость характерного времени спонтанного торможения ревербератора, τ_1 , от параметра модели, a , вблизи бифуркационной границы, соответствующей значению $a_c = 0.1803$.



3 Обсуждение

На графике зависимости $\langle R_{кр} \rangle$ от a виден экстремум примерно при $a = 0.135$; он находится на некотором расстоянии от бифуркационной границы. Пока не доказано иное, естественным выглядит предположение, что этот минимум вдоль бифуркационной границы как раз и соответствует линии выхода из бифуркационной зоны.

Данная работа показала, что без применения количественных методов оценки особенностей получаемого автоволнового решения оказывается невозможным ответить на ряд важных вопросов. Нет причин сомневаться, что применение комбинации методов количественной оценки спонтанного дрейфа ревербератора позволит автоматизировать построение параметрических портретов автоволновых моделей двумерных сред.

Работа выполнена при финансовой поддержке РФФИ, гранты 08-07-00353 и 10-01-00609.

References

- [5] R. Aliev, A. Panfilov. A simple two-variable model of cardiac excitation. *Chaos, Solutions & Fractals* 1996; 7(3): 293-301.
- [6] M. Feigin, M. Kagan. Emergencies as a manifestation of effect of bifurcation memory in controlled unstable systems. *International Journal of Bifurcation and Chaos*, 2004; 14(7): 2439-2447.
- [7] Ф.И. Атауллаханов., Е.С. Лобанова О.Л Морозова, Э.Э. Шноль, Е.А. Ермакова, А.А. Бутылин, А.Н. Заикин. Сложные режимы распространения возбуждения и самоорганизации в модели свертывания крови. *УФН*. 2007. 177(1): 87–104.
- [8] Ю.Е. Елькин, А.В. Москаленко. Базовые механизмы аритмий сердца. В кн.: Клиническая аритмология. Под ред. проф. А.В. Ардашева. М.: ИД Медпрактика-М, с. 45–74, 2009.
- [9] Ю.Е. Елькин, А.В. Москаленко, Ч.Ф. Стармер. Спонтанная остановка дрейфа спиральной волны в однородной возбудимой среде. *Математическая биология и биоинформатика*, 2007; 2(1): 73-81.
- [10] А.В. Москаленко, Ю.Е. Елькин. Мономорфна ли мономорфная аритмия? *Биофизика*, 2007; 52(2): 339-343.

Системные принципы физиологии и симметрия процессов рассеяния и концентрации энергии в биосистемах

Харитонов А.С., Охнянская Л.Г.

ИМПБ РАН

kharitonov358@yandex.ru,

Abstract

A new mathematical description of an oscillation principle for self-movement of biosystems, as active search of a minimum of free energy of the formation in three classes of variables on the basis of symmetry of measures of chaos and an order and a golden section fractal is considered.

Л.А. Блюменфельд сформулировал слабость современной науки: современная биофизика не может отличить живой организм от косного тела. Подтверждением этому факту может служить анализ системных принципов физиологии и постулатов физики.

К. Бернар писал: «Живой организм - это вихрь», а В.И. Вернадский - круговорот энергии. И.М. Сеченов определил: «Организм - это единая, целостная система, включая среду его обитания». «Мир как органон» по Аристотелю имеет одну цель – это сохранение единства и целостности взаимодействия Бытия и Небытия по Платону за счет эволюции трехсущностных объектов природы к гармонии.

В. Томсон отметил два факта: 1) закон рассеяния внешних потоков энергии в живой и неживой природе является общим, 2) «тело животного работает не как термодинамическая машина». Изменение ниши существования определяет активную сущность биологической эволюции / Р.Ю. Майер/. С.А. Подолинский заметил, что эволюция биосистем определяется их способностью концентрировать солнечную энергию.

Сохранение целостности, концентрация энергии для достижения цели и механизмы обратной связи проявляются при исследовании активной подвижности живых организмов / Н.А. Бернштейн/.

В физической модели живого организма Н.А.Умов /1902г/ предложил учесть дополнительно структуру динамических элементов, резонансные взаимодействия которых позволяют понять природу

самодвижения и работу биологических систем против второго закона термодинамики.

А. А. Ухтомский в теории о доминанте раскрыл механизм активного самодвижения организма, где синхронизация и резонансы нелинейных колебательно-волновых процессов порождают целенаправленное функционирование организма.

Изучение динамики неблагоприятных колебательно-волновых воздействий, позволило сформулировать осцилляторный (колебательно-волновой) принцип активного нелинейного взаимодействия живого организма [1].

Начиная с механики Ньютона, математика описывает движение центра тяжести тел под действием внешних сил, когда активными силами тел, на которые указал еще Г.Лейбниц, можно пренебречь. И.Ньютон определил условия, когда можно пренебречь «врожденной силой инерции» и структурой тел. Эти условия соответствуют постулатам статистической физики: система состоит из одинаковых классических и тождественных квантовых частиц. В рамках применимости этих постулатов и гипотез статистической физики самодвижение не проявляется, и системы изменяют свое состояние только под действием внешних «сил», а их эволюцию можно описывать энтропией, равной мере хаоса. Ряд упрощений позволил описывать объекты природы с помощью дифференциально-интегрального исчисления в двух классах переменных. Однако эти упрощения привели к противоречию с опытом эволюции объектов природы и к отсутствию теории, способной описать указанные системные принципы физиологии. Замкнутые (консервативные) системы эволюционируют к максимальному хаосу и тепловому равновесию. Биологические же системы эволюционируют к гармонии /Л. Пачоли, Л. да Винчи, Г. Лейбниц/, активно концентрируя свободную энергию, и не существуют вечно.

Температура системы, инвариантная внешняя система отсчета и производная могут не существовать для систем, описываемых в трех классах переменных /К. Вейерштрассе/. К этому же

А.М. Молчанов указал на неполноту известных инвариантов для изучения устойчивости сложных резонансных взаимодействий. Поэтому, с одной стороны, «в науке столько истины, сколько в ней математики» /И.Кант/, а, с другой стороны, математика на бинарные аксиомах привела к искажению закона эволюции и к непознаваемости закономерностей развития биосистем.

Предлагается построение математики на алгоритмах разбиения целого (единицы) на три изменяющиеся части, так что часть может быть функционально сложнее своего целого. В таком построении математики после каждого шага разбиения целого на части возникают новые математические функции и новые свойства динамических элементов, позволяющие описывать системы, состоящие из активных динамических элементов [2,3].

Такое построение математики привело к введению новых функций и уравнений для них:

$$1 = -\sum_{i=1}^K f_i \log_K f_i + \sum_{i=1}^K f_i \log_K (K f_i) = I + G \quad (1)$$

где 1 – единица - отражает целостность системы, I – мера хаоса описывает процесс рассеяния энергии, G – мера порядка описывает процесс концентрации рассеяния энергии на множестве рассматриваемых состояний K , f_i - вероятность каждого состояния; i - последовательность этих состояний [2].

Предложена новая модели стационарного состояния систем, основанная на равновесии процессов рассеяния и концентрации энергии, которая задается равенством мер хаоса и порядка в трех классах переменных:

$$I(p, q, l) = G(p, q, l). \quad (2)$$

где K – в уравнении (1) переменная мультипликативная величина, определяемая в переменных: (p) и (q) - импульсах и координатах, к которым добавлен третий класс переменных l – набор типов степеней свободы, характеризующий структуру динамических элементов.

Условие (2) служит новым инвариантом для описания объектов природы. В таком описании статистическая энтропия равна сумме мер хаоса и порядка, и ее рост уже не противоречит факту развития систем [2].

Уравнения (1) и (2) апробированы на свойствах макромолекулы в термостате как системы, для которой постулаты об одинаковости и постоянстве свойств динамических частиц не приемлемы [3].

Свободная энергия образования системы с изменяющимися динамическими элементами определена в трех классах переменных:

$$F_{min} = E - kT \{S(p) + S(q) + S(l)\}_{max} \quad (3)$$

где F_{min} – свободная энергия образования системы, E – полная энергия системы, kT – «температура» в единицах энергии, $S(p)$ - энтропия импульсного пространства микросостояний, $S(q)$ - энтропия координатного пространства микросостояний, $S(l)$ - энтропия структурного пространства микросостояний, характеризующая структурное многообразие динамических элементов [3].

В этом случае свободная энергия образования сложной системы не только стремится к своему минимуму, но и необратимо осциллирует около положения своего минимума по внутренним причинам. Свободную энергию образования можно моделировать во внутренней системе отсчета с помощью фрактала золотой пропорции [2].

Вместо максимального значения энтропии как функции двух независимых классов переменных (p, q) нами введен новый принцип максимума суммы трех энтропий $S(p, q, l)$, так что три класса переменных оказываются взаимно зависимыми и указывают на симметрию мер хаоса и порядка в природе.

Процесс развития описывается ростом структурной энтропии, для приращения которой возникает рекуррентное уравнение:

$$\Delta S(l_n) = \Delta S(l_{n-1}) + \Delta S(l_{n-2})$$

где каждое приращение структурной энтропии $\Delta S(l_n)$ связано с ее предыдущими приращениями.

Рекуррентное уравнение порождает золотую пропорцию при $n \rightarrow \infty$, и возможность разбивать систему на три части с ее помощью тремя разными способами. Где сама золотая пропорция разбивается на части четырехбуквенным кодом, порождающим при $n \rightarrow \infty$ множество чисел удовлетворяющих теореме Пифагора и аксиомам геометрии [2].

Из такого описания следует, что живая и косная системы стремятся к одному и тому же минимуму свободной энергии, но разными способами: по-разному разбивая целое на части и по-разному организуя сбор осциллирующих комплексов.

Косные системы эволюционируют, сохраняя структуру динамических элементов, если пренебречь фазовыми переходами. Поэтому для них правдоподобны постулаты и гипотезы статистической механики, когда процессом концентрации энергии можно пренебречь.

Живые же организмы эволюционируют преимущественно за счет изменения структуры динамических элементов, которая регулирует

равновесие процессов рассеяния и концентрации энергии, порождая тем самым внутренние «силы» и самодвижение в трех классах переменных как отклонения рекуррентных параметров от золотой пропорции. Поэтому моделирование биосистем целесообразно дополнить учетом взаимодействия процессов рассеяния и концентрации энергии во внутренней системе отсчета.

Предлагается новый инвариант (2) для описания биосистем. Живой организм ищет на каждом уровне иерархии своей организации минимум свободной энергии образования по формуле (3), которую можно моделировать золотой пропорцией во внутренней системе отсчета. Математику можно строить на аксиоме разбиения единицы на три изменяющиеся части, начиная с уравнения (1).

Литература

- [1] Охнянская Л.Г., Мишин В.П., Спектор Э.Л. «А.А.Ухтомский и развитие идей теории нелинейных колебаний в области физиологии». Сб. Учение А.А. Ухтомского о доминанте и современная нейрофизиология. Ленинград. Наука. 1990. С. 60-84.
- [2] Харитонов А.С. Структурное описание сложных систем. //Прикладная физика №1, 2007, С.5 – 10.
- [3] Харитонов А.С. Структурные свойства макромолекулы в термостате. //Прикладная физика №1, 2008, С.13 – 16.

Analysis of the psychorelaxation efficiency by changing the brain bioelectrical activity

Dick O.E

Pavlov Institute of Physiology of Russian Academy of Science
glazov.holo@mail.ioffe.ru

Abstract

The EEG time series before, during and after removing a tactile pain stimulation by applying the psychorelaxation technique are examined for healthy subjects and patients with neural disorders. The degree of psychorelaxation is estimated as a change in the value of the Holder exponent corresponding to a minimal fluctuation of the brain activity gained by the continuous wavelet transform – based method of multifractal analysis. We have shown the reliability of estimating the psychorelaxation of the healthy and pathological brain by variation of the Holder exponents.

1.1 Introduction

Bioelectric activity of the human brain is rather nonstationary and has fractal properties [1]. That is why the focus from the traditional study of EEG power spectra has shifted to methods of nonlinear dynamics such as wavelet transform and fractal analysis.

The aim of the work is to examine the changes in EEG time series evoked by a tactile pain stimulation and the following psychorelaxation technique [2].

1.2 Experimental procedure

We used 60 second non-artefact segments of EEG recordings from 10 healthy subjects and 11 patients with disruptions of interrelations between cortex and subcortex on the thalamic level (first group of patients) and 12 patients with disruptions on the brain-stem level (second group).

The psychorelaxation is in switching attention from the pain sense on the perception of color spots arising in the state of concentrating on the pain locus with closed eyes. Switching attention is accompanied by decreasing the pain symptoms up to their complete disappearance.

The degree of psychorelaxation is estimated as a change in the value of the Holder exponent (h_{min}) corresponding to minimal fluctuation of the brain activity gained by the continuous wavelet transform – based method of multifractal analysis [3].

1.3 Results and discussion

For the healthy subjects relaxation is almost complete. The Holder exponents fall during pain stimulation and return up to previous values after relaxation (Fig.1).

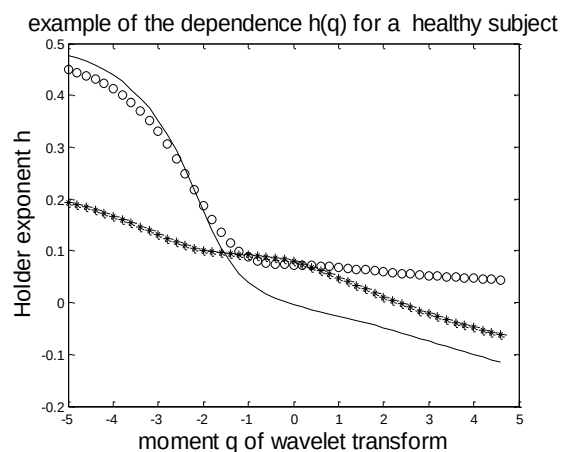


Fig.1 Curves $h(q)$ before stimulation (solid curve, $h_{min}=0.47$), during pain (* - $h_{min}=0.2$) and after relaxation (o - $h_{min}=0.45$).

The decrease of the exponents testifies the enhancement of the degree of singularity and anticorrelation in sequent EEG values under the pain stimulation and its falling during relaxation. For the power spectrum, contrastingly, there are no significant differences before, during the pain stimulation and during psychorelaxation.

For the patients with disruptions of interrelations between cortex and subcortex on the thalamic level (the first group) the complete relaxation is observed only in the case when the positive transformation of color spots (from achromatic white and black to the short-wavelength blue) takes place during concentrating on the pain locus and the following relaxation (75% patients of the group). The transformation is connected with decreasing the pain symptoms and the improvement of the psychophysiological state of the patient. For other 25% patients of the first group there is

no recovery of the Holder exponent values up to values corresponding to the healthy subjects.

For the patients with disruptions on the brain-stem level (the second group) the Holder exponents remain smaller than for the healthy subjects after relaxation (Fig.2).

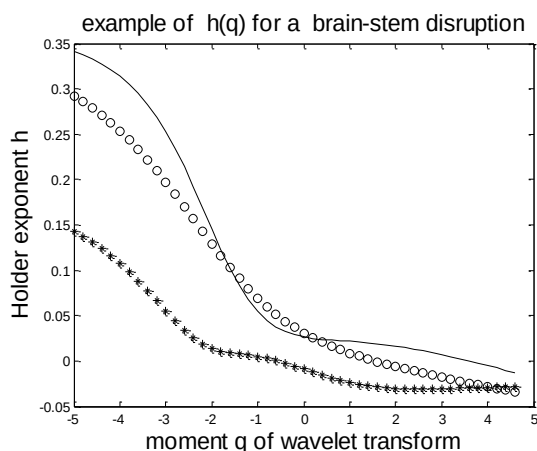


Fig.2 Curves $h(q)$ before stimulation (solid curve, $h_{min}=0.34$), during pain (* - $h_{min}=0.3$) and after relaxation (o - $h_{min}=0.15$).

That is, the sequent EEG values are more singular for this group as compared with the healthy time series. The transformation of color spots is not observed for 81% patients and for 19 % achromatic colors change into long- wavelength red or orange. The noncomplete relaxation is not accompanied by the improvement of the psychophysiological state

Changes in the power spectrum of the both groups of patients do not correlate with the degree of relaxation. For example, the power of θ -rhythm lowers *under* relaxation for 52% patients of the first group and for 100% patients of the second group.

Thus, absence of variations in power spectra of the healthy subjects correlates with the complete relaxation after the pain stimulation. However, the changes in power spectra of the patients with neural disruptions do not always reflect variations of the degree of psychorelaxation. Unlike the facts, the changes in the multifractal parameters give a good ability to estimate changes in the healthy and pathological brain activity.

References

- [11] E.T. Bullmore, M.J. Brammer, P. Bourlon et al. Fractal analysis of electroencephalographic signals intracerebrally recorded during 35 epileptic seizures; evaluation of a new method for synoptic visualization of ictal events. *Electroencephalogr. Clin Neurophysiol*, vol. 91, pages 337-345, 1994.
- [12] J.F. Muzy, E. Bacry, A. Arneodo. Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method. *Physical Review E*, vol. 47, pages 875-884, 1993.
- [13] В.А. Ишинова, И.А. Святогор, Н.Г. Зуева и др. Динамика цветового отражения боли у пациентов с тревожно-фобическими расстройствами в процессе психорелаксации. *Боль*, том 3, с. 30-31.

Анализ нейрофизиологической активности по данным магнитной энцефалографии

Бердникова А.П., Махортых С.А.

ИМПБ РАН, Пущино

makh@impb.ru

Рассматриваются вопросы изменчивости записей магнитных энцефалограмм (МЭГ) вызванной активности мозга при аудиторной стимуляции испытуемого и спонтанной активности при некоторых разновидностях болезни Паркинсона. На основе проведенного анализа динамики сигнала решается задача локализации источников биомагнитного поля для различных участков записей МЭГ. Результаты интерпретируются в рамках моделей связанных осцилляторов для наблюдаемых данных МЭГ.

Описание задачи

Одной из основных проблем при изучении работы мозга является решение так называемой обратной задачи, т.е. локализация источников биоэлектромагнетизма. В настоящее время эта задача решается как ЭЭГ [2], так и МЭГ методами [3]. Однако большинство исследователей считают метод магнитной энцефалографии более точным [4].

В исследовании рассматривались два ряда измерений: 1) у здорового испытуемого измерялась магнитная энцефалограмма при аудиторной стимуляции; 2) производилась запись спонтанной биомагнитной активности у пациента, страдающего паркинсонизмом.

Измерения биомагнитных полей были проведены в Медицинской школе Нью-йоркского университета (New York Medical School) на магнитном 148-канальном энцефалографе Magnes 2500 WH производства компании 4D Neuroimaging (США). Используемый магнитный энцефалограф обеспечивает полное кортикальное покрытие, что позволяет провести локализацию активности с миллисекундным временным разрешением и миллиметровым пространственным разрешением. Для локализации источников магнитного поля в мозге использовали магнитно-резонансную томограмму.

Измеряемый сигнал представляет собой пространственно-временную структуру: 148-мерный вектор измерений в 148 точках на поверхности головы, развернутый во временной ряд с частотой опроса датчиков 500 Гц. Исходные данные можно представить в виде двух числовых

матриц $A \in R^{148 \times T}$ и $B \in R^{148 \times Z}$, характерная величина для длины временного ряда T составляет в среднем 3×10^5 отсчетов. Интервал между временными отсчетами в соответствии с частотой опроса – 1/500 секунды. В столбце i , строке j матрицы A содержатся значения магнитного поля в момент времени t_j для i -го из 148 датчиков. В матрице B записаны трехмерные координаты 148 датчиков.

Для более точной обработки данных использовались методы так называемого функционального анализа данных – FDA (functional data analysis).

Методы классификации участков измеряемого биомагнитного сигнала подробно описаны в работах [3,6,7]. Далее приводятся только результаты использования предлагаемых процедур для двух групп испытуемых. Для второй серии измерений, связанных с заболеванием Паркинсона, было выделено четыре зоны локализации источников патологического сигнала – в мозжечке, стволе мозга (варолиевом мосту), в черной субстанции среднего мозга и в базальных ганглиях (хвостом ядре). Для моментов времени, в которых наблюдается нормальный сигнал, источники биомагнитной активности основную часть времени находятся в коре головного мозга, не имея четкой локализации.

Описание моделей

Для исследования динамики изменчивости источников МЭГ в различных участках головного мозга для спонтанной и вызванной активности ЦНС предложен класс моделей связанных осцилляторов, отвечающих отдельным нейронам или группам нейронов в структурах мозга с локализованной повышенной активностью. Частные случаи данного подхода представлены и проанализированы в работе [1]. В основу расчётной модели легла следующая система обыкновенных дифференциальных уравнений (число уравнений соответствует числу источников МЭГ):

$$\begin{cases} \ddot{x}_n = \mu_n \dot{x}_n + \varepsilon_{ni} x_i + \alpha_n f_n(x) + \gamma_n z(t - \tau_n) \\ z(t) = A(t) \cdot \cos(2\pi\omega t + \varphi) \end{cases} \quad (1),$$

Анализ динамики МЭГ при паркинсонизме свидетельствует о нелинейной стохастичности системы, при этом наблюдаются переключения от спорадической нелокализованной в коре активности к сигналу сравнительно низкой размерности [3, 6]. Для описания этой системы необходимо сохранить нелинейные члены в (1), и в качестве первого приближения используется модель Дуффинга [1]:

$$\begin{cases} \ddot{x} = -(2\pi\omega_L)^2 x - \mu\dot{x} - \beta_L x^3 - \varepsilon_{xy} y \\ \ddot{y} = -\left((2\pi\omega_H)^2 + \frac{2}{\tau} + 4\frac{(t-t_0)^2}{\tau}\right)y - \frac{t-t_0}{\tau}\dot{y} - \beta_H y^3 - \varepsilon_{yx} x \\ x(0) = x_0, \dot{x}(0) = \gamma, y(0) = y_0, \dot{y}(0) = \nu_0 \end{cases} \quad (2)$$

Параметры моделей оценивались с помощью пакетов MATLAB, FreeMat и R методами Левенберга-Маркардта и наименьших квадратов поиска оптимального решения полученных ОДУ.

Обсуждение результатов

По данным настоящего исследования у здоровых людей в состоянии спокойного бодрствования не обнаружены постоянные источники повышенной магнитной активности. В таком состоянии можно зарегистрировать только кратковременные всплески активности в коре больших полушарий.

Однако, при функциональной нагрузке (в аудиторных экспериментах) появляются источники повышенной магнитной активности в височной коре, где, как известно, находятся корковые проекции слуховой сенсорной системы [9]

Представляет интерес тот факт, что характер сигнала МЭГ несколько различается в правом и левом полушариях. При фильтрации магнитного сигнала на разных частотах источники локализуются в разных полушариях – в левом при фильтрации сигнала на частоте 10 Гц и в правом при частоте 20 Гц.

Данное явление объясняется, по-видимому, тем, что в строении слуховой сенсорной системы, также как и в строении большинства других сенсорных систем, присутствует перекрест афферентных волокон, вследствие чего информация от правого уха попадает главным образом в левое полушарие и наоборот. Однако перекресты и синапсы расположены на нескольких уровнях слуховой системы, а небольшая часть афферентных волокон не совершает перекреста, благодаря чему слуховые импульсы поступают и в ипсилатеральное полушарие [8].

Разработанные методы анализа биоманнитного сигнала мозга дают возможность выявить у больных паркинсонизмом источники повышенной магнитной активности в тех областях, которые связаны с развитием данного заболевания. Такая

активность зарегистрирована в базальных ганглиях, черной субстанции, мозжечке, т.е. структурах, относящихся к экстрапирамидной системе, нарушения в которой приводят к возникновению двигательных расстройств [9].

Предложенные модели изменчивости локализованных в различных структурах мозга источников электромагнитных полей, регистрируемых в энцефалографии, позволяют после параметрической настройки получить качественное и в ряде случаев хорошее количественное соответствие экспериментальным записям.

Работа поддержана РФФИ, гранты 08-07-00353, 10-01-00609, 09-07-12108-офи_м.

Литература

- [1] Leistritz L., Putsche P., Schwab K., Hesse1 W., Suße T., Haueisen J., Witte H. Coupled oscillators for modeling and analysis of EEG/MEG oscillations. Biomed Tech 2007; 52:83–89
- [2] Гнездицкий В.В. Обратная задача ЭЭГ и клиническая электроэнцефалография. М.: МЕДпресс-информ, 2004: 624 с.
- [3] Устинин М.Н., Махортых С.А., Молчанов А.М., Ольшеев М.М., Панкратов А.Н., Панкратова Н.М., Сухарев В.И., Сычев В.В. Задачи анализа данных магнитной энцефалографии. В кн. Компьютеры и суперкомпьютеры в биологии. М.: Институт компьютерных технологий, 2002: с.327-349.
- [4] Kristeva-Feige R., Rossi S., Feige B., Mergner T., Lücking C.H., Rossini P.M. The Bereitschaftspotential paradigm in investigating voluntary movement organization in humans using magnetoencephalography (MEG). Brain Res Brain Res Protoc. 1997 Feb;1(1):13-22.
- [5] Mazziotta J.C.. Mapping human brain activity in vivo. West J Med, 1994. 161(3): p. 273-278.
- [6] Махортых С.А., Семечкин Р.А. Спектральные методы анализа магнитных энцефалограмм. Динамика неоднородных систем. Труды Института системного анализа РАН. Том. 33, вып. 12. 2008: с. 236-250
- [7] Бердникова А.П., Махортых С.А. Распознавание повторяющихся участков в записи магнитной энцефалограммы. Нейроинформатика-2007. IX Всероссийская Научно-техническая конференция. Сборник научных трудов. Часть 2. М.: МИФИ. 2007: с. 64-71.
- [8] Carpenter M.B. Core Text of Neuroanatomy. – Baltimore: Williams & Wilkins, 1991: p. 481
- [9] Крыжановский Г.Н., Карабань И.Н., Магаева С.В., Кучеряну В.Г., Карабань Н.В. Болезнь Паркинсона (этиология, патогенез, клиника, диагностика, лечение, профилактика). – М.: Медицина, 2002: 336 с.

Частичная синхронизация нейронной активности и обработка информации в мозге

Р.М. Борисюк^{1,2}, Я.Б. Казанович¹

¹Институт математических проблем биологии РАН
Пушино, Московская область, 142290
yasha@impb.psn.ru

²University of Plymouth, Plymouth, PL4 8AA, UK
r.borisjuk@plymouth.ac.uk

Partial synchronization of neural activity and information processing in the brain

Roman M. Borisjuk^{1,2}, Yakov B. Kazanovich¹

¹Institute of Mathematical Problems in
Biology, Russian Academy of Sciences
Pushchino, Moscow Region, 142290, Russia

²University of Plymouth, Plymouth, PL4 8AA,
UK

Abstract

A model of visual selective attention is developed based on the mode of partial synchronization in a network of synaptically coupled neurons. The model correctly reproduces some observations concerning the mechanisms of attentional control, such as the coherence of spikes in the population of neurons included in the focus of attention, and the inhibition of neurons outside the focus of attention. An example of sequential selection of objects simultaneously present in the real visual scene is described.

Описывается модель зрительного внимания, основанная на режиме частичной синхронизации в сети из синаптически связанных нейронов. Модель воспроизводит ряд экспериментальных фактов, таких как синхронизация активности в популяции нейронов, представляющих объект, включенный в фокус внимания, и подавление активности нейронов вне фокуса внимания. Приводится пример

последовательного выбора в фокус внимания объектов, одновременно присутствующих в реальной зрительной сцене.

В электрической активности мозга присутствуют разнообразные ритмические компоненты, в той или иной мере проявляющиеся в различных частотных диапазонах. При этом характер колебаний коррелирует с внешними воздействиями и психологическим состоянием исследуемого организма. Устойчивые паттерны ритмической активности были обнаружены в различных структурах мозга на уровне отдельных нейронов, нейронных популяций и структур мозга. Такие экспериментальные данные получены в первичных зонах зрительной и обонятельной коры, сенсомоторной коре, в таламусе, в гиппокампе и в других структурах [1-6].

При избытии экспериментальных данных, в которых регистрируется синхронная активность, вопрос о ее роли в обработке информации в мозге остается без ясного ответа. Нет сомнения, что синхронизация играет важную роль в кодировании, переносе, запоминании и вспоминании информации, а также в координации взаимодействия различных структур мозга [7]. Однако нейронные механизмы, лежащие в основе этих функций, как правило, не известны.

Одной из когнитивных функций, которая предположительно реализуется путем синхронизации активности в различных областях коры, является внимание. Имеются экспериментальные и теоретические основания полагать, что система внимания является иерархической, то есть в ней имеется некоторая подсистема, называемая центральным элементом, которая организует включение того или иного объекта в фокус внимания [8-10]. В работах [11, 12] была выдвинута гипотеза о том, что взаимодействие

центрального элемента системы внимания с нейронными ансамблями, представляющими различные объекты, происходит по принципу фазовой автоподстройки частоты. В фокус внимания включается тот объект, признаки которого кодируются в коре осцилляторной активностью, синхронизованной с активностью центрального элемента. Недавно эта гипотеза получила экспериментальное подтверждение. В работе [13] было показано, что при зрительном внимании определенная область в префронтальной коре (так называемое фронтальное глазное поле) является источником повышенной синхронизации нейронной активности на частоте гамма-ритма в поле V4.

Основываясь на этих экспериментальных данных, мы разработали и исследовали модель внимания с центральным осциллятором, конструктивными элементами которой являются фазовые осцилляторы [14-16]. Недостатком этой модели является абстрактный характер осцилляторов, используемых в качестве элементов моделей. Это сужает возможности по воспроизведению экспериментальных данных.

В данной работе предлагается модель внимания, построенная с использованием пейсмекерных импульсных нейронов типа Ходжкина-Хаксли. Эта модель с большей правдоподобностью и детальностью отражает функционирование нейронов, что позволяет исследовать режимы синхронизации импульсной активности, ассоциированные с вниманием, воспроизводить подавление импульсной активности, соответствующей стимулам вне фокуса внимания, выбирать в фокус внимания наиболее заметные стимулы и воспроизводить частотные характеристики нейронной активности, наблюдаемые в экспериментах по вниманию [17-18]. С использованием адаптивной процедуры для модификации связей от центрального элемента к периферическим осцилляторам, кодирующим признаки зрительного объекта, в модели реализована последовательная процедура включения в фокус внимания различных объектов, одновременно присутствующих в зрительной сцене.

Литература

- [1] *Basar E.* Brain function and oscillations - New York: Springer-Verlag, 1998.
- [2] *Buzsaki G., Draguhn A.* Neural oscillations in cortical networks // *Science*, **304**, 1926-1929, 2004.
- [3] *Freeman W.J.* How brains make up their minds - New York: Columbia University Press, 2001.
- [4] *Gray C.M.* Synchronous oscillations in neuronal systems: mechanisms and functions // *J. Comp. Neurosci.*, **1**, 11- 38, 1994.
- [5] *Steriade M., Jones E.G. Llinas R.* Thalamic oscillations and signalling - New York: John Wiley & Sons, 1990.
- [6] *Vinogradova O.S.* Expression, control, and probable functional significance of the neuronal theta-rhythm // *Progr. Neurobiol.*, **45**, 523-582, 1995.
- [7] *Preti G, De Palma A.* Does the brain oscillate? The dispute on neuronal synchronization // *Neurological Sciences*, **25**, 41-47, 2004.
- [8] *Baddeley A.* Exploring the central executive // *Quarterly Journal of Experimental Psychology*, **49A**, 5-28, 1996.
- [9] *Baddeley A.* Fractionating the central executive // *Principles of Frontal Lobe Function*. Eds. D. Stuss and R.T. Knight, pp. 246-260, - New York: Oxford University Press, 2002.
- [10] *Cowan N.* Evolving conceptions of memory storage, selective attention and their mutual constraints within the human information processing system // *Psychological Bulletin*, **104**, 163-191, 1988.
- [11] *Kryukov V.I.* An attention model based on the principle of dominance // *Neurocomputers and attention I: neurobiology, synchronisation and chaos*. Eds A.V. Holden, V.I. Kryukov, pp. 319-351, - Manchester Univ. Press, 1991.
- [12] *Крюков В.И.* Модель внимания и памяти, основанная на принципе доминанты и компараторной функции гиппокампа // *Журнал высшей нервной деятельности*, **54**, №1, 10-29, 2004.
- [13] *Gregoriou G.G., Gotts S.J., Zhou H, Desimone R.* High-frequency, long-range coupling between prefrontal and visual cortex during attention // *Science*, **324**, 1207-1210, 2009.
- [14] *Kazanovich Y.B., Borisyuk R.M.* Dynamics of neural networks with a central element // *Neural Networks*, **12**, 149-161, 1999.
- [15] *Borisyuk R.M., Kazanovich Y.B.* Oscillatory neural network model of attention focus formation and control // *BioSystems*, **71**, 29-38, 2003.
- [16] *Borisyuk R., Kazanovich Y.* Oscillatory model of attention-guided object selection and novelty detection // *Neural Networks*, **17**, 899-915, 2004.
- [17] *Borisyuk R., Kazanovich Y., Chik D., Tikhanoff V., Cangelosi A.* A neural model of selective attention and object segmentation in the visual scene: An approach based on partial synchronization and star-like architecture of connections // *Neural Networks*, **22**, 707-719, 2009.
- [18] *Chik D., Borisyuk R., Kazanovich Y.* Selective attention model with spiking elements // *Neural Networks*, **22**, 890-900, 2009.

Бугорковые аттракторы и проблема поиска "кода разума"

В.Л. Дунин-Барковский

НИИ СИ РАН, Москва

В моделях поведения, интенсивно разрабатываемых в последнее время классиком теоретических нейроисследований Джоном Хопфилдом, широко используется популярная сейчас модель так называемых бугорковых (bump) аттракторов. Автор доклада со значительным опозданием понял, что эти самые аттракторы были исследованы в цикле его работ в 1984-1995 гг. В частности, их описание составляло небольшую часть докторской диссертации, защищённой на одном из последних учёных советов ИБФ АН СССР.

В докладе будет описана суть указанного теоретического явления. В настоящее время становится понятно, что бугорковые аттракторы часто встречаются (используются) в работе нервной системы высших животных и являются удобным элементом вычислений в нейронных системах. Похоже, что их описание и экспериментальное

обнаружение можно расценивать как начало этапов понимания вычислительной механики мозга. Таким образом, с большой степенью уверенности сейчас можно констатировать, что один из принципов работы нервной системы теперь понят. В докладе делается попытка сформулировать общую задачу поиска «кода разума», то есть, полного набора принципов/элементов/структур, которые обеспечивают эффективное функционирование мозга. Высказывается гипотеза, что элементов «кода разума» немного. Скорее всего, их не больше, чем элементов «кода жизни» - 64 триплетов ДНК/РНК. По мнению автора, малая эффективность поиска элементов «кода разума» до настоящего времени объясняется не сложностью задач, а недостаточной смелостью их постановок и отсутствием привычки искать простые ответы на несложные вопросы. Этот тезис аргументируется на конкретных примерах обнаружения нескольких уже известных элементов «кода разума».

Сегментация изображений на основе временной корреляции нейронной активности

Я.Б. Казанович

Институт математических проблем биологии РАН

Пушино, Московская область, 142290

yasha@impb.psn.ru

Image Segmentation Based on the Correlation of Neural Activity

Yakov B. Kazanovich

Institute of Mathematical Problems in Biology,
Russian Academy of Sciences

Pushchino, Moscow Region, 142290, Russia

Abstract

A review of neural network models for image segmentation is presented. The models are based on the synchronization of activity of neural assemblies that code individual objects simultaneously present in the visual scene. Neurobiological principles implemented in the models are formulated, architectures and functional principles of the models are described. Examples of application of the models to processing artificial and real images are presented. Capabilities of the models and perspectives of their improvement are discussed.

Работа представляет собой обзор нейросетевых моделей для сегментации изображений. В основе моделей лежит принцип синхронной активности нейронных ансамблей, кодирующих объекты, одновременно присутствующие в зрительной сцене. Приводятся нейро-биологические факты, подлежащие моделированию, описываются конструкции и принципы функционирования моделей, а также результаты их применения к искусственным и реальным изображениям. Обсуждаются сравнительные качества моделей и перспективы их улучшения.

Разбиение изображений на непересекающиеся фрагменты, соответствующие зрительным объектам

и фону, является популярной задачей искусственного интеллекта. На эту тему опубликовано огромное количество работ. Обзоры по данной области исследований можно найти, например, в работах [1, 3, 6] (см. также веб-страницы [5-7]). Сегментация считается важным предварительным шагом, существенно облегчающим последующее решение задач анализа зрительных сцен, таких как распознавание образов, запоминание, обнаружение заданных объектов и т.д.

В то время как люди в большинстве случаев успешно справляются с задачей сегментации, часто даже не фокусируя на ней внимания, компьютерная сегментация оказалась трудной задачей. В связи с этим у исследователей возникло желание лучше разобраться, как эту задачу решает мозг. Важной особенностью работы мозга является параллельная обработка информации. Это дает надежду на то, что использование биологических методов сегментации окажется не только качественным, но и быстрым.

Встречный интерес к проблеме сегментации возник и у биологов в связи с так называемой проблемой интеграции признаков объектов в цельный образ. По-английски эта проблема называется the binding problem – проблема связывания. Дело в том, что признаки, используемые при кодировании информации о зрительных стимулах в первичных зонах зрительной коры, имеют различную природу. Это могут быть геометрические, спектральные или динамические характеристики изображения. Известно, что первичная обработка различных по своей природе признаков идет в специализированных нейронных структурах коры мозга. Только в ассоциативных зонах коры формируется представление о целостных объектах. В связи с этим возникает вопрос: каковы нейронные механизмы, позволяющие сохранить информацию о принадлежности признаков к отдельным объектам, и как осуществляется объединение нужных признаков в цельный образ объекта?

Сегментацию можно рассматривать как частный случай байндинга. В самом деле, если речь идет об изображениях, заданных попиксельно, то

достаточно включить в число признаков объектов признаки отдельных пикселей, чтобы метка принадлежности пикселей различным объектам реализовывала сегментацию изображения.

Гипотеза о возможном механизме интеграции признаков в мозге была предложена в работах [2, 4] и получила название теории временной корреляции (ТВК). Идея состоит в том, что признаки объекта кодируются синфазной активностью нейронов в различных областях коры мозга. Синфазность служит в качестве метки, помечающей информацию, относящуюся к одному объекту.

Идеология ТВК хорошо сочетается с методами моделирования, опирающимися на осцилляторные нейронные сети. Согласно ТВК сегментация должна приводить к синфазной активности нейронов, кодирующих зрительный объект, в то время как активность нейронных ансамблей, кодирующих разные объекты, должна быть некогерентной.

В последние годы было предложено большое количество моделей сегментации изображений, основанных на ТВК и отличающихся степенью учета биологических данных, классами анализируемых изображений, применяемыми принципами функционирования и результатами применения [8, 9]. Некоторые авторы обращают внимание на контурные изображения, поскольку они более похожи на стимулы, используемые в нейрофизиологических экспериментах. Другие предпочитают полутоновые и цветные изображения в силу их большей практической значимости. В последнем случае рецептивные поля нейронов или нейронных осцилляторов обычно представлены пикселями на изображении, каждый из которых характеризуется интенсивностью серого (для полутоновых изображений) или интенсивностями трех-четырёх составляющих спектра (для цветных изображений).

Принятый в данной работе подход к категоризации методов сегментации является компромиссом, старающимся учесть как тип обрабатываемых изображений, так и применяемые вычислительные схемы. Сначала будут рассмотрены модели для сегментации контурных изображений. Затем для изображений, заданных попиксельно, описываются два класса моделей сегментации – с последовательным и параллельным выбором объектов. Результатом работы параллельной процедуры является одновременное существование нескольких синфазных внутри себя нейронных ансамблей, каждый из которых кодирует свой зрительный объект. При использовании последовательной процедуры синфазные ансамбли, кодирующие различные объекты, возникают один за другим. В каждый момент времени только один из представленных одновременно в зрительной сцене объектов кодируется когерентной активностью

определенного нейронного ансамбля. Активность в остальной части сети отсутствует или подавлена до низкого уровня. Далее будет рассмотрено, как сегментация может сочетаться с методами кластеризации. Затем будут описаны модели сегментации, привлекающие дополнительные когнитивные функции, такие как память и внимание. Заключение будет посвящено обсуждению результатов сегментации, достигнутых в существующих моделях, и перспективам их совершенствования.

Литература

- [1] Li J., Gray R.M. Image Segmentation and Compression Using Hidden Markov Models (The International Series in Engineering and Computer Science). - Klawer Academic Publishers, 2000.
- [2] Malsburg C. von der. The correlation theory of brain function // Internal report 81-2, Max-Planck Institute for Biophysical Chemistry, 1981 (reprinted in Models of Neural Networks, E. Domany, J.L. van Hemmen, K. Schulten (Eds.)) - Springer, New York, 1994, pp. 95-119.
- [3] Shapiro L.G., Stockman G.C. Computer Vision. - New Jersey, Prentice-Hall, 2001.
- [4] Singer W., Gray C.M. Visual feature integration and the temporal correlation hypothesis // *Ann. Rev. Neurosci.*, **18**, 555-586, 1995.
- [5] Wikipedia-Segmentation (image processing)
[http://en.wikipedia.org/wiki/Segmentation_\(image_processing\)](http://en.wikipedia.org/wiki/Segmentation_(image_processing))
- [6] Zhang A. A survey on evaluation methods for image segmentation // *Pattern Recognition*, **29**, 1235-1346, 1996.
- [7] Баринаова О., Вежневцев А. Методы сегментации изображений: автоматическая сегментация // Графика и мультимедиа,
<http://cgm.graphicon.ru/content/view/147/73/>
- [8] Казанович Я.Б. Сегментация изображений с помощью динамических нейронных сетей // Нейроинформатика-2008, X Всероссийская научно-техническая конференция. Лекции по нейроинформатике, ч. 2, под ред. Ю.В. Тюменцева, с. 37-96. Москва, МИФИ, 2008.
- [9] Казанович Я.Б., Шматченко В.В. Осцилляторные нейросетевые модели сегментации изображений и зрительного внимания // Нейроинформатика-2004. VI Всероссийская научно-техническая конференция. Лекции по нейроинформатике, ч. 1, под ред. Ю.В. Тюменцева, с. 72-125, Москва, МИФИ, 2004.

Модель кратковременной памяти человека на основе нейронной сети

Ляховецкий В.А., Боброва Е.В.⁽¹⁾

Институт Физиологии им. И.П. Павлова РАН⁽¹⁾
V_la2002@mail.ru

Whether it's correct reminiscence recall? Model of short-term human memory based on ANN

Lyakhovetskii V.A., Bobrova E.V.

Abstract

Which criterion the person should use to estimate the correctness of his reminiscence? Previously we developed the ANN model for memorizing the sequence of hand movements. This model can successfully reproduce the human error's distribution. Here we propose to use such criterion as the number of iteration necessary for network transfer to stable state. The simulation results well coincide with psychophysical ones.

1 Введение

Проблема оценки верности запомненного материала – одна из важных проблем когнитивной психологии. Некоторым образом человеку удается понять, какой материал запомнен им правильно, а какой – ошибочно. Причем отличить правильный ответ от неверного возможно и неосознанно [5]. Часто правильность ответа влияет на время, требуемое на воспроизведение запомненного, – средние времена правильных и неверных ответов испытуемого различны. При воспроизведении последовательности движений руки среднее время правильно воспроизведенных движений ниже среднего времени ошибочно воспроизведенных движений [2].

В известных моделях кратковременной памяти, основанных на теории соревновательного сцепления (*competitive queuing*), искажение при воспроизведении возникает не в результате работы модели, а искусственно вводится как действие внешнего шума [3]. По-видимому, поэтому они не уделяют внимание проблеме оценки верности запомненного. Вероятно, для такой оценки могут использоваться мета-данные, специально помещаемые в модель.

Нейросетевая модель [6], основанная на теории последовательного связывания (*chaining*), хорошо подходит для описания распределений ошибок при немедленном воспроизведении запомненных последовательностей движений руки, в частности, воспроизводит эффект новизны, наблюдаемый в психофизиологическом эксперименте. В данной

работе предпринята попытка, оставаясь в рамках внутреннего представления данных, используемого в модели [6], предложить объективный критерий, позволяющий отличить верные ответы от неверных, не обращаясь повторно к входной последовательности стимулов.

2 Модель кратковременной памяти на основе нейронной сети

2.1 Структура модели

Модель типа двунаправленная гетероассоциативная память содержит два слоя формальных нейронов, соединенных двунаправленными связями [1]. Сеть характеризуется быстро вычисляемой матрицей весов W и не нуждается в тренировке. Этап запоминания заключается в расчете матрицы W , определяющей силу связей между нейронами входного и выходного слоев. На этапе воспроизведения элементы вектора Z_i , содержащего i -й запомненный образ, вычисляются через элементы вектора Z_{i-1} и матрицы W . Двустороннее распространение активности продолжается до тех, пока не будет достигнуто устойчивое состояние.

2.2 Схемы кодирования

В психофизиологических опытах испытуемые, находящиеся с завязанными глазами, должны были запомнить и воспроизвести последовательность движений руки по 6 целям, расположенным на листе формата A4, разбитом на 24 (6x4) квадрата. Для кодирования последовательности положений рабочей точки руки в качестве элементов вектора Z выбраны [5]:

1. Смещения рабочей точки в полярной системе координат. Элементами векторов являются направление α и длина перемещения ρ . Параметры α и ρ хранятся в рамках двух подсетей, между которыми отсутствуют связи, что соответствует имеющимся в литературе сведениям о раздельном запоминании человеком направления и амплитуды движения. В этом случае искажения внутренних представлений α и ρ происходят независимо друг от друга. Для расчета симметричной матрицы весов W_ρ , используемой при запоминании последовательности длин перемещений, применялся алгоритм расчета, предложенный в [1]. Для расчета асимметричной матрицы весов W_α , используемой при запоминании последовательности направлений перемещений,

применялся алгоритм расчета, предложенный в [4]. Схема кодирования смещения рабочей точки в полярной системе координат может использоваться при векторном (относительном) способе внутреннего представления информации, т.е. при кодировании не положения, а траектории движения руки.

2. Координаты рабочей точки. Такая схема кодирования может использоваться при позиционном (абсолютном) способе внутреннего представления информации, т.е. при кодировании положения, а не траектории движения руки.

Также для обеих схем кодирования вектор Z содержит номер цели, поэтому модель способна запоминать последовательности, содержащие повторяющиеся элементы. Нейроны сети работают с бинаризованными данными, поэтому элементы векторов Z переводятся в двоичную систему счисления.

При каждой схеме кодирования с моделями проводили 10000 опытов. Регистрировалось количество итераций, за которые сеть достигнет устойчивого состояния.

3 Результаты и их обсуждение

Зависимость числа итераций, за которое нейронная сеть достигнет устойчивого состояния, от числа ошибочно запомненных битов представлена в таблице. Разница в числе итераций, необходимых для достижения устойчивого состояния при различных схемах кодирования, объясняется тем, что при векторном кодировании (схема 1) число итераций вычислялось по двум подсетям.

Таблица. Число итераций сети для различного числа ошибочно запомненных битов.

Ош. бит.	0	1	2	3	4	5	6	7	8
Схема 1	3.2	3.6	3.8	3.9	3.9	3.9	3.8	3.8	4.0
Схема 2	1.4	1.6	1.7	1.8	1.8	1.7	-	-	-

Из таблицы видно, что для обеих схем кодирования сеть достоверно быстрее приходит в устойчивое состояние при правильных ответах (0 ошибочных бит). То есть, хотя информация о числе ошибочных битов не доступна сети, для оценки правильности ответа модель может использовать количество итераций, необходимых для перехода сети в устойчивое состояние. Можно предположить, что, так как разница в числе итераций в зависимости от правильности ответа невелика (0.4 итерации для схемы 1, 0.2 итерации для схемы 2), то, основываясь на таком критерии, модель будет иногда принимать ошибочный ответ за верный. То есть, что введенный в рассмотрение критерий правильности ответа будет являться «слабым». Однако такое свойство модели не будет являться недостатком. При запоминании движений руки испытуемые достаточно часто допускают повторные ошибки, то есть при

двукратном воспроизведении запомненной серии движений дважды повторяют свою ошибку одинаковым образом (дважды подряд касаются одного и того же неверного квадрата, дважды подряд выбирают одно и то же неверное направление). Среднее время движений, связанных с повторными ошибками, значительно меньше времени движений, связанных с иными ошибками, и не отличается от времени правильно запомненных движений. Причем эта закономерность проявляется уже для первого ошибочного движения. Повидимому, испытуемые уверены в правильности таких ошибочных движений и не пытаются корректировать их при повторном заучивании последовательности движений руки, то есть не могут отличить такие неверные ответы от правильных [7].

Таким образом, предложенный критерий оценки правильности запомненного материала как числа итераций, необходимого для перехода в устойчивое состояние, хорошо согласуется с психофизиологическими данными. Такой критерий может использоваться для обучения нейросетевой модели без учителя.

Работа выполнена при поддержке РФФИ, проект №09-04-01207-а, 2009-2011.

References

- [1] B.Kosko. Bidirectional associative memories. *IEEE Trans. Systems, Man and Cybernetics*, vol.18, p. 49-60, 1988.
- [2] P.Rabbit. An elementary preliminary taxonomy for some errors in laboratory choice RT-tasks. *Acta Psychologica*, vol. 33, p.56-76, 1970.
- [3] M.P.A.Page, D.G.Norris The primacy model: a new model of immediate serial recall. *Psychological Review*, vol. 105, p. 761-781, 1998.
- [4] Z.-B.Xu, Y.Leung, X.-W.He. Asymmetric bidirectional associative memories. *IEEE Trans. Systems, Man and Cybernetics*, vol. 24, p. 1559-1564, 1994.
- [5] В.М.Аллахвердов. Методологическое путешествие по океану бессознательного к таинственному острову сознания. – СПб.: Изд-во «Речь», 2003 – 368 с.
- [6] В.А.Ляховецкий, Е.В.Боброва. Воспроизведение запомненной последовательности движений правой и левой руки: позиционное и векторное кодирование. *Журнал высшей нервной деятельности*, т.59, №1, с. 45-54, 2009.
- [7] В.А.Ляховецкий, Е.В.Боброва. Повторные ошибки при многократном воспроизведении запомненной последовательности движений руки. *III-я всероссийская с международным участием конференция по управлению движением*. Великие Луки, с. 31-32, 2010.

Роль процессов эндо/экзоцитоза рецепторов и их латеральной диффузии в изменении ответа нейрона

Мурзина Г.Б.

Институт высшей нервной деятельности и нейрофизиологии РАН, Москва, Россия

gbmurzina@mail.ru

Abstract

The receptor exo/endocytosis and lateral membrane diffusion as the basic mechanisms of number of the membrane receptor change are discussed. A reaction-diffusion equation for receptor change that takes into account these processes is proposed.

Введение

Согласно существующим экспериментальным данным изменение чувствительности нейрона к воздействию веществу происходит при изменении эффективности функционирования соответствующего типа рецепторов. Таким образом, эффективность воздействия медиатора зависит как от эффективности функционирования каждого отдельного рецептора, так и от общего числа мембранных рецепторов. Если константа диссоциации рецептора с медиатором может изменяться при фосфорилировании/дефосфорилировании рецептора, при взаимодействии с сопряженным медиатором, при изменении вязкости мембраны, либо при наличии влияния активности рецепторов одного кластера друг на друга [9], то количество мембранных рецепторов зависит от скорости эндоцитоза и экзоцитоза рецепторов, от скорости их латеральной диффузии и от процессов синтеза и деградации соответствующего типа рецепторов. Применение математического моделирования для анализа экспериментальных данных этих процессов позволило выдвинуть ряд предположений о механизмах, лежащих в их основе [6-8]. Однако, несмотря на имеющиеся экспериментальные данные и существующие математические модели, вопрос о вкладе латеральной диффузии и процессов эндо- и экзоцитоза в изменение числа мембранных рецепторов еще недостаточно изучен.

Эндо/экзоцитоз внесинаптических ацетилхолиновых рецепторов

Моделирование процессов эндо- и экзоцитоза рецепторов, приводящих к изменению числа мембранных рецепторов, осуществлялось на основании экспериментальных данных, полученных

при изучении десенситизации ацетилхолиновых рецепторов (АХР), возникающей при ритмической аппликации агониста ацетилхолина на соматические нейроны виноградной улитки [5]. Сопоставление результатов вычисления с экспериментальными данными показало, что для анализа действия различных внутриклеточных веществ на процессы десенситизации и восстановления ответа нейрона после десенситизации вместо простейшей двухкомпонентной модели состояний рецепторов, учитывающей наличие мембраносвязанных и внутриклеточных (находящихся в эндосомах) рецепторов, которая была использована для описания интернализации метаботропных АХР в клетках нейробластомы [4] и для общего анализа влияния ряда внутриклеточных веществ на скорость восстановления (рециклирования) рецепторов [10], лучше использовать трехкомпонентную модель. В этой модели учитывается нахождение некоторого количества рецепторов на мембране, в сформированных примембранных везикулах и в рециклированных везикулах. Полученные результаты вычислений и экспериментальные данные позволили прийти к заключению, что депрессия холинорецепторной чувствительности внесинаптических областей мембраны командных нейронов оборонительного поведения виноградной улитки связана с динамин-зависимым эндоцитозом холинорецепторов при участии микротрубочечного цитоскелета.

Латеральная диффузия рецепторов

Кроме процессов эндо- и экзоцитоза рецепторов изменение количества рецепторов в определенных локусах мембраны происходит вследствие их латеральной диффузии. Поэтому возникает вопрос: в какой степени изменение количества рецепторов на мембране определяется каждым из этих процессов. По степени свободы все мембранные рецепторы можно разделить на две основные группы: связанные с примембранным матриксом и свободные рецепторы. В синаптической области большая часть рецепторов вследствие взаимодействия с белками постсинаптической плотности (различными для разных типов рецепторов), с актином – белком примембранной сети микрофиламентов и в ряде случаев с адгезионными белками формирует кластеры и

находится в связанном состоянии [3]. В процессах эндо-, экзоцитоза рецепторов и латеральной диффузии участвуют только рецепторы, находящиеся в несвязанном – свободном состоянии. Согласно экспериментальным данным существует некоторая вероятность перехода рецепторов между этими состояниями, причем эта вероятность возрастает при активации синапса. Расположенная вдоль синаптической области часть мембраны является местом формирования клатриновых ямок и, следовательно, содержащих рецепторы везикул, участвующих в эндоцитозе. По-видимому, в прилегающем к этой области мембраны внутриклеточном объеме могут находиться и содержащие рецепторы рециклированные везикулы, участвующие в экзоцитозе, в то время как везикулы с вновь синтезированными рецепторами и везикулы с рециклированными рецепторами из комплекса Гольджи находятся значительно дальше – в прилегающей к шипику области дендрита. Встраивание этих рецепторов в мембрану происходит там же, и затем они диффундируют в синапс. Вследствие латеральной диффузии может происходить изменение числа мембранных рецепторов в соседних шипиках, а также на некотором расстоянии вдоль дендрита. В то же время на большие расстояния транспорт находящихся в везикулах рецепторов осуществляется по микротрубочкам.

В свободном (несвязанном) состоянии находится, по-видимому, большая часть находящихся на соме нейрона мембранных рецепторов, которые могут участвовать как в процессах эндо- и экзоцитоза, так и латеральной диффузии. Поэтому, уравнения, описывающие изменение числа мембранных рецепторов в синаптической области и на соме нейрона, несколько различаются. В настоящее время существуют математические модели, описывающие латеральную диффузию мембранных рецепторов по дендриту с учетом наличия на нем шипиков и динамики эндо- и экзоцитоза рецепторов [1,2]. Однако в них вопрос о соотношении вклада диффузии и эндо- и экзоцитоза рецепторов в изменении их числа на мембране в полном объеме не рассматривается. Не рассматривался этот вопрос и для изменения числа рецепторов на соме при воздействии на нейрон какого-либо вещества (например, при аппликации ацетилхолина). В общем виде уравнение диффузии рецепторов с учетом их эндо-и экзоцитоза имеет вид

$$\frac{\partial c}{\partial t} = D \cdot \left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} \right) + F - kc$$

где $c(x,y)$ – концентрация мембранных рецепторов, D – коэффициент латеральной диффузии, F – функция источника (моделирующая экзоцитоз рецепторов), k – константа скорости эндоцитоза

рецепторов. Решение этого уравнения (с учетом граничных условий I рода) позволяет при анализе экспериментальных данных по изучению десенситизации АХР при ритмической аппликации агониста ацетилхолина на сому командных нейронов уточнить закономерности процессов эндо- и экзоцитоза рецепторов с учетом роли латеральной диффузии в уменьшении ответа нейрона.

Без проведения точных расчетов можно приблизительно оценить какие из этих процессов и на каком расстоянии от места аппликации медиатора доминируют. Поскольку постоянная времени десенситизации рецепторов $\tau = 1/(k_{en} + k_{ex})$, где k_{en} и k_{ex} – константы соответственно скоростей эндоцитоза и экзоцитоза рецепторов, то согласно экспериментальным данным $\tau \sim 2-4$ мин. При значении $D = 0.1 \mu m^2 \cdot s^{-1}$ [9] (которое различается для шипика и сомы на порядок вследствие формирования в области синапса решетчатой структуры белками примембранного матрикса, что приводит к уменьшению D) временная шкала латеральной диффузии $t \sim r^2/D$, где r – расстояние от места подачи медиатора к мембране. Тогда на расстоянии $r_0 = 3-5 \mu m$ скорости процессов эндо-и экзоцитоза и латеральной диффузии будут одного порядка. При $r < r_0$ для свободных рецепторов большую роль будет играть латеральная диффузия (что, по-видимому, имеет место в шипиках), а при $r > r_0$ (расстояния характерные для сомы нейрона) больший вклад в изменение числа мембранных рецепторов будут вносить процессы эндо- и экзоцитоза.

Литература

- [1] Earnshaw B.A., Bressloff P.C. *J.Comput.Neurosci.*, volume 25, pages 366-389, 2008.
- [2] Holcman D., Triller A. *Biophys.J.*, volume 91, pages 2405-2415, 2006.
- [3] Newpher T.M., Ehlers M.D. *Neuron*, volume 58(4), pages 472-498.
- [4] Szekeres P.G., Koenig J.A., Edwardson J.M. *J.Neurochem.*, volume 70(4), pages 1694-1703, 1998.
- [5] Махновский Д.А., Третьякова М.С., Мурзина Г.Б., Пивоваров А.С. *Журн.Высш.Нервн. Деят.*, том 60, № 2, с.206-216, 2010.
- [6] Мурзина Г.Б., Силькис И.Г. *ДАН РАН*, том 352, № 3, с.420-423, 1997.
- [7] Мурзина Г.Б. *Журн.Высш.Нервн.Деят.*, том 53, № 2, с.184-190, 2003.
- [8] Мурзина Г.Б. *Биофизика*, том 250, вып.3, с.488-491, 2005.
- [9] Мурзина Г.Б. *Журн.Высш.Нервн.Деят.*, том 58, № 6, с.643-648, 2008.
- [10] Мурзина Г.Б. *Журн.Высш.Нервн.Деят.*, том 60, № 3, с.279-295, 2010.

Возможность выделения синаптических токов из корковых вызванных потенциалов

Покровский А.Н.

Санкт-Петербургский государственный университет
anpokr@petrodvoretz.spb.ru

Possibility of extraction of synaptic currents from cortical evoked potentials

Pokrovsky A.N./

Abstract

Simple linear approximation of the equations for extra- and intra-cellular potentials of "apical" dendrites only was used. Estimation of synaptic currents is a non-correct problem of mathematical physics. We use piecewise linear approximation of mean synaptic currents and classical method of selection of quasi-solutions. .

1 Формулировка задачи

1.1 Рабочие шипотезы и допущения

1. Вызванные потенциалы создаются только апикальными дендритами. Кора - плоская, все дендриты - одинаковые параллельные цилиндры, направлены поперёк коры, по оси z , и распределены равномерно по плоскости коры.

2.. Все синапсы каждого типа активируются одновременно. Предполагается, что все синапсы на дендритах расположены на тех же глубинах в коре, что и контакты электрода,. измеряющие вызванные потенциалы.

3. Внутриклеточные потенциалы дендритов малы по сравнению с порогом спайков, и допустимо использование линейной модели.

4. Вследствие невозможности одновременного измерения внутриклеточных и внеклеточных сигналов дендрита в модели используются безразмерные величины глубины z (в единицах характеристической длины λ дендрита) и времени (в единицах характеристического времени T мембраны дендрита), а также приведенные синаптические токи с размерностью потенциала.

1.2 Обозначения

N – число контактов измерительного электрода;

i - номер контакта; $i=1,2, \dots, N$;

z - безразмерная длина (в долях λ); z_i - глубина i -того контакта в коре;

t - безразмерное время (в долях T);

$\varphi(z,t)$ - средний внеклеточный потенциал в коре; измеряются компоненты $\varphi_i(t)$ вектора $\varphi(t)$;

$\psi(z,t) = \psi_1(t) \delta(z-z_1) + \dots + \psi_N(t) \delta(z-z_{1N})$ – это приведенный синаптический ток.

1.3. Уравнение

$$\left(\frac{\partial}{\partial t} - \frac{\partial^2}{\partial z^2} + 1 \right) \varphi(z,t) = -\psi(z,t). \quad (1)$$

содержит систему дифференциальных уравнений для вычисления вектора $\varphi(t)$ по заданному вектору $\psi(t)$; решение этой системы находится в квадратурах. Обратная задача некорректна [2]; попытки регуляризации обратной задачи дали не вполне удовлетворительные результаты в отношении формы вычисленных синаптических токов.

2 Метод подбора квазирешения

2.1 Аппроксимация синаптических токов

Используется простейшая кусочно-линейная аппроксимация синаптических токов, содержащая два линейных участка. Форма токов одного источника предполагалась одинаковой на всех уровнях отведения, отличались только амплитуды. Поскольку такая форма токов является лишь приближением к экспериментально измеряемой форме токов, результаты вычислений следует называть квазирешением обратной к (1) задачи.

2.2 Подбираемые параметры

В стандартных многоканальных электродах кончики соседних электродов расположены на равных расстояниях друг от друга. Поэтому

достаточно задать это расстояние в единицах λ . Это первый из подбираемых параметров.

Второй параметр – это постоянная времени мембраны дендрита T . Третий параметр – это момент t_0 возникновения токов синапсов [3] (если считать t_0 одинаковым для всех уровней по глубине коры; в противном случае будет необходимо подбирать N параметров $t_1, \dots, t_N$).

Кроме перечисленных трёх величин, необходимо подбирать ещё N коэффициентов амплитуд тока синапсов на всех уровнях отведения вызванных потенциалов. Итого необходимо подбирать $N+3$ параметров для тока синапсов одного типа.

2.3 Экспериментальные данные

В качестве примера была использована запись вызванных потенциалов импульсов одного нейрона из работы [1] (fig. 1, N2). Для подбора использованы 11 нижних уровней ответа (из 16-ти), с 6-го по 16-тый, на которых видны вызванные потенциалы вместе с потенциалами спайков аксона. Эта часть записи представлена на Рис. 1 слева.

Каждая экспериментальная кривая содержит быстрые колебания и более медленный процесс. Быстрые колебания вероятно являются результатом суммации внеклеточных потенциалов нервных импульсов, распространяющихся по многим ветвям аксона нейрона N2 [1]. Отрицательный более медленный процесс является результатом суммации токов синапсов этого аксона, расположенных на апикальных дендритах.

2.4 Результат подбора

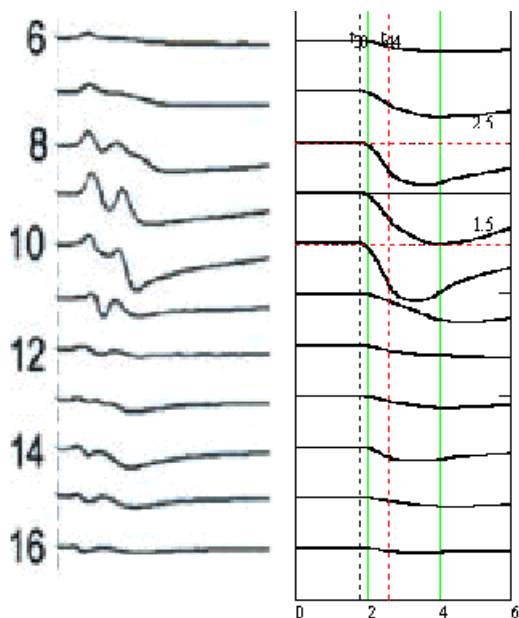


Рис. 1

На Рис. 1 слева изображены вызванные потенциалы импульсов нейрона N2 из работы [1] (11 нижних уровней, с 6-го по 16-тый). Шкала времени в миллисекундах.

Справа на Рис. 1 построены вызванные потенциалы, вычисленные по результатам подбора параметров.

На Рис. 2 слева представлены подобранные суммарные токи синапсов.

на Рис. 2 справа - разность между измеренными и вычисленными потенциалами, соответствующая суммарным внеклеточным потенциалам спайков, которые распространяются по отросткам аксона.

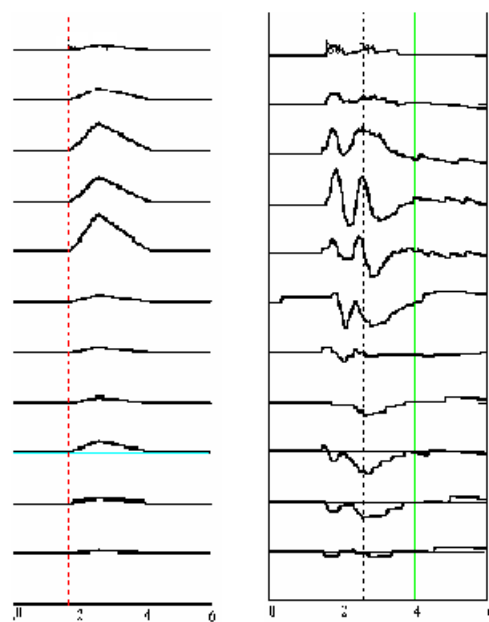


Рис. 2

Время – в миллисекундах.

3 Заключение

Для систематического использования метода подбора квазирешений задачи анализа вызванных потенциалов необходима разработка специальных алгоритмов и программ подбора параметров задачи.

Работа поддержана РФФИ проект 09-01-00473

Литературы

- [1] , Н.А. Swadlow, A.G. Gusev, T.Bezdudnaya, J.Neuroscience, 2002, 22(17):7766–7773 .
- [2] А.Н. Тихонов, В.Я. Арсенин. Методы решения некорректных задач. М.:Физматлит, 1979.
- [3] Дж. Экклс Физиология синапсов. М.: «Мир», 1966.

Вейвлеты в исследовании зрительного восприятия

Пронин С.В., Чихман В.Н., Шелепин Ю.Е.

Институт физиологии им. И.П. Павлова РАН

niv@pavlov.infran.ru

Abstract

The algorithm and program for incomplete image synthesis on the basis of wavelet transformation are described. The optimal signal for visual cortical neurons must correspond to the size of their receptive fields, and spectra with a band-pass averaging 1.4 octaves. We used wavelets, the functions of which are localized in terms of spatial and frequency domains. The program produces an adaptive algorithm. As basic wavelet we used LoG and Gabor functions. Synthesized fragmented images were employed, in the same way as in the standard Gollin-Test, in psychophysical experiments.

1 Введение

Известно, что нейроны зрительной коры с простыми рецептивными полями реагируют на стимулы, спектр пространственных частот которых занимает частотные полосы со средней шириной 1.4 октавы [1, 2]. Каким должен быть оптимальный стимул для данных нейронов, т.е. изображение, на которое реагировала бы лишь небольшая группа нейронов с близкими по характеристикам рецептивными полями? Очевидно, что такой стимул должен иметь небольшой угловой размер, не превышающий размер рецептивных полей данной группы нейронов, и его спектр должен лежать внутри полосы пространственных частот, на которую реагирует данная группа нейронов. Этим требованиям отвечают вейвлеты - функции, имеющие хорошую локализацию, как в пространственной, так и в частотной областях. В последнее время вейвлеты широко применяются для исследования зрительного восприятия [3].

В общем случае двумерный вейвлет может быть описан формулой (1), где $g(x, y)$ - функция с нулевым средним значением и нормой, равной 1, отвечающая требованию хорошей пространственной и частотной локализации; S - масштаб (размер вейвлета); x_0, y_0 - координаты центра вейвлета. Изменение масштаба S приводит к смещению спектра вейвлета, но его полуширина, измеренная в октавах, при этом остается неизменной. Важное преимущество вейвлетов заключается в том, что с их помощью можно с произвольно высокой точностью аппроксимировать любое изображение в соответствии с формулой (2),

$$w(x, y) = \frac{1}{S} g\left(\frac{x - x_0}{S}, \frac{y - y_0}{S}\right) \quad (1)$$

$$I(x, y) = \sum_{i=0}^N a_i w_i \quad (2)$$

где a_i - весовой коэффициент («контраст» вейвлета); W_i - вейвлет с параметрами S, x_0, y_0 , зависящими от индекса i .

2 Результаты

В разработанной программе вейвлет-разложения изображений для определения весовых коэффициентов a_i и параметров вейвлетов S, x_0, y_0 используется адаптивный алгоритм, предложенный С.Малла и З.Зангом [4, 5]. По сравнению с другими методами вейвлет-разложения, этот алгоритм дает наиболее компактное описание, позволяя достичь заданной точности аппроксимации при минимальном количестве элементов разложения. В общем случае в формуле разложения (2) присутствуют вейвлеты с самыми разными значениями масштаба S и координатами центров x_0 и y_0 . Оставив в формуле (2) только те вейвлеты, которые являются оптимальным стимулом для интересующей нас категории нейронов, можно исследовать процесс восприятия неполных изображений именно этими нейронами.

В качестве вейвлетов мы использовали лапласиан от гауссиана (LoG) - функцию, широко применяемую в обработке изображений для выделения перепадов яркости [6].

Для проведения психофизических экспериментов разработана программа определения порогового уровня фрагментированности изображений, при котором испытуемый правильно распознает предъявляемый ему объект. Программа осуществляет предъявление в режиме последовательного накопления фрагментов. Для формирования фрагментированного изображения из вейвлет-разложения (2) отбирается определенная часть элементов. Отбор производится в два этапа: сначала отбираются вейвлеты, которые являются оптимальным стимулом для интересующей нас категории нейронов. Например, имеющие одинаковый размер S . Затем из полученного набора N вейвлетов в случайном порядке выбирается $M < N$ элементов, из которых формируется

фрагментированное изображение. Степень его фрагментированности характеризуется отношением M/N . В вейвлет-разложении (2) коэффициенты a_i убывают по абсолютной величине с ростом i и контрасты соответствующих им вейвлетов также уменьшаются. Но для построения фрагментированного изображения (во всяком случае, для используемых нами контурных изображений) желательно, чтобы все его элементы имели близкие контрасты: в этом случае их "вклады" в результирующий образ будут приблизительно одинаковыми. Таким образом, мы отбирали определенную часть элементов, для которых разница между значениями весовых коэффициентов по абсолютной величине не отличается более чем в два раза. Т. е. из множества всех элементов вейвлет-разложения отбирается подмножество элементов с близкими контрастами, в которых самый контрастный элемент не превосходит по контрасту самый малоконтрастный более чем вдвое. В программе изображение аппроксимируется вейвлетами, имеющими только два значения размера S , отличающимися в 2 раза. Для формирования фрагментированного изображения можно использовать как вейвлеты какого-либо одного масштабного уровня, так и обоих одновременно.

Для реализации селективного воздействия на группы нейронов с различной ориентационной настройкой программа дополнена возможностью использования вейвлетов Габора. Спектр этих функций ограничен как по частотному диапазону, так и по диапазону ориентаций, поэтому они хорошо подходят в качестве стимула для нейронов с ориентационной избирательностью. Элементы Габора – вейвлеты, описываемые следующим выражением:

$$B(x, y) = g(x, y) \cos(\omega(x \cos(\theta) + y \sin(\theta)) + \varphi$$

где $g(x, y)$ – функция Гаусса, ω – пространственная частота, θ – ориентация, φ – фаза.

Спектр элемента Габора ограничен не только по диапазону частот, но и по ориентациям спектральных составляющих. Благодаря этому свойству тестовые изображения, составленные из элементов Габора, широко используются при проведении экспериментов в области физиологии зрения.

Пример изображений, полученных с помощью программы, показан на рисунке 1. Изображения предъявлялись испытуемым в режиме Голлин-теста.

Программа позволяет использовать для синтеза тестовых изображений не только контурные объекты, но и фотографии. На рис. 2 показан пример

формирования фрагментированного изображения с использованием в качестве элементов отрезков линий и краев, подвергнутых цифровой фильтрации.

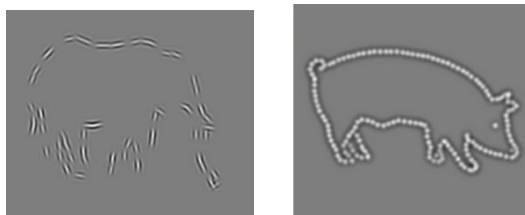


Рис. 1. Фрагментированные изображения
слева – LoG-функции, справа - вейвлеты Габора.

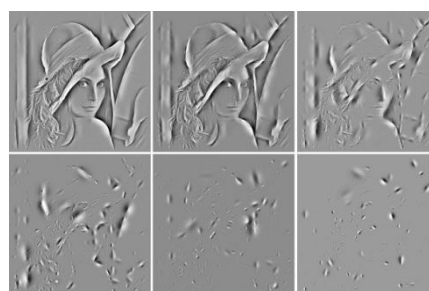


Рис. 2. Пример формирования фрагментированного изображения из фотографии.

Работа поддержана грантом РФФИ 09-07-00336.

References

- [1] DeValois, R. L., Albrecht, D. G. & Thorell, L. G. 1982, "Spatial frequency selectivity of cells in macaque visual cortex". Vision Res. 22, 545-559
- [2] Tolhurst, D. J. & Thompson, I. D. 1982, "On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat". Proc. R. Soc. Lond. B213, 183-199
- [3] D.J.Field, "Wavelets, vision and the statistics of natural scenes", Phil. Trans. R. Soc. Lond. A.1999. 357, 2527-2542
- [4] S. Mallat and Z. Zhang, 1993, "Matching Pursuit in a Time-Frequency Dictionary", IEEE Sig. Proc., 41, 3397-415
- [5] F.Bergeaud and S. Mallat, "Matching pursuit of images", Proceedings IEEE Int. Conf. Image Processing, Washington DC, October 1995, vol. I, pp. 53-56
- [6] Д.Март "Зрение. Информационный подход к изучению представления и обработки зрительных образов", М.. "Радио и связь", 1987.

Множественность решений уравнений RISM/HNC в пределе малой плотности

Соболев Е.В., Тихонов Д.А.

Институт математических проблем биологии РАН

egor@impb.psn.ru

Abstract

The molecular liquids were studied by the integral equations of the theory of liquids within the RISM approximation. Fundamental difference between molecular and simple liquids was found in the case of the HNC closure. Additional solution branch with a kink in the first bifurcation point was found. The solutions on different branches with the same temperature and density belong to different phases.

1 Введение

В работе [1,3] исследовались все связанные решения уравнений RISM для молекулярных жидкостей, и было показано, что вблизи области фазовых переходов газ-жидкость одному термодинамическому состоянию (ρ, T) соответствует несколько решений с разными значениями изотермической сжимаемости. Такое поведение соответствует поведению уравнения Орнштейна-Цернике для простых жидкостей, которое также имеет множественные решения. Однако в работе [3] для гиперцепного (HNC) замыкания была отмечена фундаментальная разница между молекулярными и простыми жидкостями. В этом замыкании была получена спинодаль (линия разрыва изотермической сжимаемости) молекулярной жидкости в диапазоне значений плотности от 6.5×10^{-3} до $10 \times 10^{-3} \text{ \AA}$. Рассчитанная спинодаль имеет вид гиперболы с асимптотой в точке $\rho=0$, в связи с этим был сделан вывод о возможном отсутствии в системе критической точки. Мы продолжили исследования, с целью выяснить существуют ли решения уравнений RISM для гиперцепного замыкания в газовом пределе.

2 Методы и параметры расчетов

Нами был развит численный алгоритм получения всех связанных решений, основанный на методе продолжения по параметру, в котором для преодоления особых точек осуществляется переход к другой параметрической зависимости уравнений RISM [1]. Алгоритм позволяет получать все

связанные решения с постоянной температурой, двигаясь по одному из параметров ρ или χ^{-1} .

Мы исследовали решения уравнений RISM для метана. Потенциал взаимодействия и геометрические параметры молекулы были взяты из работы [3]. Отметим, что результаты и выводы могут быть обобщены на любую молекулярную систему исследуемую методом RISM, поскольку особенности обусловлены самими уравнениями в гиперцепном замыкании.

3 Результаты

Мы получили несколько изотерм решений уравнений RISM для метана при высокой температуре и вплотную подходящих к нулю плотности. На рис. 1 слева изображена изотерма в плоскости (ρ, χ^{-1}) при температуре 6290 К. Видно, что первая бифуркация решения возникает даже при температуре существенно выше экспериментальной критической температуры (190.4 К). При этом в первой точке бифуркации происходит излом изотермы. Для случая простых жидкостей бифуркации решений выше критической точки не наблюдаются и являются особенностью модели RISM в гиперцепном замыкании. Принципиальным отличием полученного результата от одноатомных систем, является то, что в одной изотерме связываются два различных состояния жидкости. Первое – это состояние перегретой жидкости (флюида), ветвь которого лежит выше излома, и второе – это газовое состояние, где $\chi^{-1} < 1$. В простых жидкостях такие состояния четко разделены критической изотермой.

Справа на рис. 1 изображены атом-атомные корреляционные функции, принадлежащие найденной изотерме и соответствующие одной плотности $\rho = 5.6 \times 10^{-3} \text{ \AA}$. Как и в простых жидкостях, эти три решения отличаются в основном асимптотикой. Решения с меньшими значениями χ^{-1} являются более дальнедействующими. На соответствующем флюидной ветви решении, осцилляции, характеризующие ближний порядок, не отмечены.

4 Выводы

Используемый в работе алгоритм оказался способным найти решения, преодолев особенность

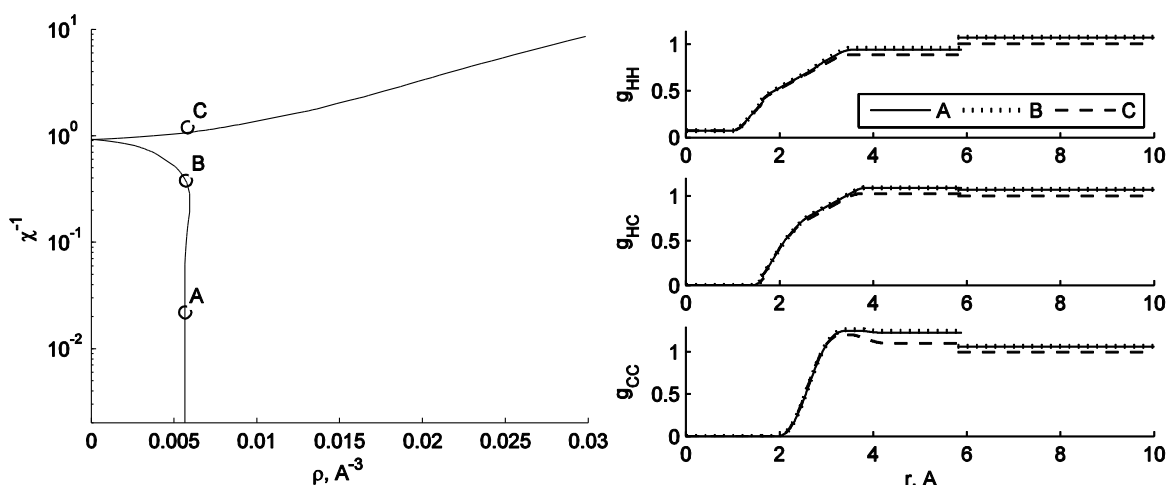


Рис. 1. Решения в приближении HNC. Слева – изотерма обратной сжимаемости при температуре 6290К, справа – бинарные функции распределения (в точке А – сплошная линия, в точке В – штриховая, в точке С – пунктирная)

в точке $(\rho, \chi^{-1}) = (0, 1)$. В результате чего и была получена ветвь решения, изображенная на рис. 1.

Результаты расчетов вновь возвращают нас к вопросу о критической точке уравнений RISM/HNC. С одной стороны мы получили решения для газовых плотностей. Следовательно, выводы работы [3], где отсутствие таких решений интерпретировалось как отсутствие критической точки, кажутся преждевременными. Но с другой стороны, полученная нами изотерма, принадлежит двум разным фазам, что также свидетельствует об отсутствии критической изотермы, разделяющей фазы.

Таким образом, необходимы дальнейшие расчеты семейства изотерм, чтобы прояснить вопрос о критическом поведении такой системы.

References

- [1] E.V.Sobolev, D.A.Tikhonov. In *Computer Research and Modeling*, volume 2 (1), pages 51-62, 2010 (Russian).
- [2] Williams D.E. Nonbonded potential parameters derived from crystalline hydrocarbons. In *Journal of Chemical Physics*, volume 47, pages 4680-4684, 1967.
- [3] E.V.Sobolev, D.A.Tikhonov. Solutions of integral equation in the theory of molecular liquids near phase transition region. *Математическая биология и бионформатика: II Международная конф., г. Пуцино, 7-13 сентября 2008 г.: Доклады / Под ред. В.Д. Лахно, стр. 50-51.*

Параллельный алгоритм численного решения уравнений RISM в пределе бесконечного разбавления

Соболев Е.В., Тихонов Д.А.

Институт математических проблем биологии РАН

egor@impb.psn.ru

Abstract

The efficient parallel algorithm to solve numerically RISM equations in the infinity dilution limit is proposed.

1 Введение

В работе [1] предложена эффективная численная схема решения уравнений RISM в пределе бесконечного разбавления. Описанный автором подход сделал возможным применение теории интегральных уравнений для изучения гидратации биологических молекул размерами до нескольких тысяч атомов, позволив получать ответ за несколько десятков часов. Сократить это время можно с помощью параллельных вычислений. Не менее важно то, что параллельный алгоритм позволяет решать системы большей размерности, потому что можно использовать оперативную память нескольких компьютеров.

В этой работе предложена схема численного решения уравнений RISM, которая основана на последовательной схеме, предложенной в работе [1]. Задача обладает параллелизмом по данным. Мы провели анализ нескольких параллельных схем и оценили их эффективность. В результате была выбрана наиболее эффективная и масштабируемая схема с наименьшим объемом коммуникационных взаимодействий. Идея алгоритма аналогична идее, которая применяется для построения параллельной схемы 2D преобразования Фурье.

2 Параллельный алгоритм

2.1 Базовые подзадачи

Рассмотрим последовательную численную схему, предложенную в работе [1]. Численное решение находится путем последовательных приближений по схеме Ньютона-Крылова начиная с некоторого $\hat{\gamma}^0$ и пока $\|\Delta^j\| < \varepsilon$:

$$\Delta^j = Z(\hat{\gamma}^j), \quad (8)$$

$$\Delta \hat{\gamma}^j = J^{-1} \Delta^j, \quad (9)$$

$$\Delta \hat{\gamma}^{j+1} = \hat{\gamma}^j + \Delta \hat{\gamma}^j, \quad (10)$$

Схема Ньютона-Крылова (или неточный метод Ньютона) предполагает нахождение $\Delta \hat{\gamma}^j$ с помощью одного из нестационарных итерационных

методов подпространств Крылова и с некоторой заранее заданной точностью η :

$$\|\Delta^j + J \Delta \hat{\gamma}^j\| \leq \eta \|\Delta^j\| \quad (11)$$

Методы подпространств Крылова не требуют предоставлять в явном виде матрицы Якоби. Для этих методов необходимо лишь задать процедуру вычисления произведения матрицы Якоби на вектор неизвестных линеаризованной системы уравнений:

$$d = Jx \quad (12)$$

Невязка Δ_{iak}^j системы уравнений на j -ой итерации определяется следующим образом:

$$\gamma_{iar}^j = \frac{1}{2\pi^2 r \Delta r} \Delta k^2 \sum_{k=0}^L \hat{\gamma}_{iak}^j \sin\left(\frac{\pi}{L} kr\right) k, \quad (13)$$

$$c_{iar}^j = \exp[-\beta u_{iar} + \gamma_{iar}^j] - \gamma_{iar}^j - 1, \quad (14)$$

$$\hat{c}_{iak}^j = \frac{4\pi}{k \Delta k} \Delta r^2 \sum_{r=0}^L c_{iar}^j \sin\left(\frac{\pi}{L} kr\right) r, \quad (15)$$

$$\Delta_{iak}^j = \sum_{\mu=1}^M \sum_{l=1}^N \hat{w}_{ilk} \hat{c}_{l\mu k}^j \hat{v}_{\mu a k} - \hat{c}_{l\mu k}^j - \hat{\gamma}_{l\mu k}^j, \quad (16)$$

где $i=1 \dots N$, $a=1 \dots M$ и $r, k=1 \dots L$. Остальные обозначения соответствуют обозначениям статьи [1].

Рассмотрим подробно каждую формулу и выделим информационно независимые вычисления.

Формулы (13) и (15) вычисляются независимо по каждой корреляционной функции. Действительно, все скалярные величины входят в эти формулы с одинаковыми первыми двумя индексами.

Формула (14) вычисляется поэлементно.

В формулу (16) все скалярные величины входят с одинаковым третьим индексом, т.е. эти вычисления выполняются независимо для каждой точки сетки дискретизации.

Таким образом, расчеты можно разбить на два этапа. Первый этап – это вычисления по формулам (13)-(15). Здесь базовой задачей являются вычисления, относящиеся к одной корреляционной функции (i, α) . Второй этап – это вычисления по формуле (16), а базовой задачей являются вычисления, относящиеся к одной точке сетки дискретизации k .

Формулы произведения матрицы Якоби на вектор неизвестных линеаризованной системы аналогичны по структуре и информационным связям, т.е. также разбиваются на два этапа с теми же базовыми задачами.

Остальные вычисления в последовательной схеме не накладывают никаких ограничений на выбор базовых подзадач. Они представлены линейными векторными операциями и операциями скалярных произведений векторов.

2.2 Масштабирование

На первом этапе базовые подзадачи могут быть распределены произвольным образом при условии, что нагрузка на процессы будет приблизительно одинаковой.

На втором этапе подзадачи распределяются по процессорам согласно условию:

$$L - \left\lfloor \frac{L}{p} \right\rfloor p = i,$$

где i – номер процесса, p – количество процессов.

В локальной памяти процесса данные каждой базовой подзадачи следует размещать последовательно. На первом этапе наиболее быстрым индексом должен быть индекс k , следующим – α , и наиболее медленным – i . На втором этапе данные должны быть транспонированы в памяти, т.е. наиболее быстрым индексом должен быть индекс α , а медленным – k .

Таким образом, во время вычисления формулы (16) распределение базовых подзадач по процессам соответствует схеме второго этапа, а все остальное время выполнения – первого. При переходе от одной схемы к другой производится параллельное транспонирование массива неизвестных, которое приводит к соответствующему перераспределению данных между процессами.

2.3 Анализ эффективности

Обозначим время выполнения последовательного алгоритма как T_1 , а время выполнения параллельного алгоритма на p процессорах – T_p . Тогда

$$T_p = \frac{T_1}{p} + T_0 + T_c,$$

где T_0 – время выполнения операций, повторяемых всеми процессами, T_c – время выполнения подзадач информационного взаимодействия. Оценим время T_c :

$$(p-1)\alpha + \frac{8NML}{p\beta} \leq T_c \leq (p-1)\alpha + \frac{8NML}{2\beta},$$

где $8NML$ – размер вектора неизвестных в байтах, α – время подготовки данных к передаче, β – пропускная способность канала передачи данных.

Оценка снизу принята в предположении, что передача данных между парами процессов может происходить одновременно, а оценка сверху – только последовательно.

Из оценки видно, что при бесконечном увеличении числа процессоров время выполнения параллельного алгоритма асимптотически стремится к линейному росту с коэффициентом α ,

но при этом $T_1 = O(N^2 M^2)$. Таким образом, для любого количества процессоров найдется такой

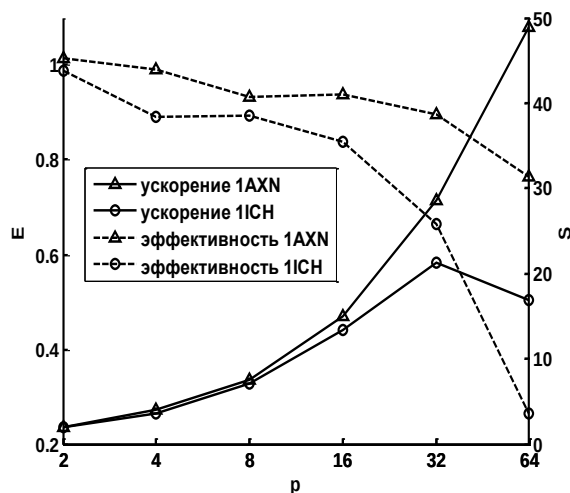


Рис. 1. Ускорение S и эффективность E параллельного алгоритма в зависимости от числа процессоров p . Экспериментальные данные для белков 1AXN (5124 атома) и 1ICH (1443 атома).

размер системы, при котором параллельный алгоритм будет эффективен, т.е. алгоритм обладает свойством масштабируемости.

3 Тестирование

На рис. 1 приведены результаты тестовых расчетов. При использовании небольшого числа процессоров ускорение почти линейное. Также подтверждается оценка масштабируемости. Для системы большего размера эффективность остается высокой при использовании большего количества процессоров.

Расчеты были выполнены на вычислительном кластере Университета Юты (Солт-Лейк-Сити, США).

4 Заключение

Предложен эффективный параллельный алгоритм для численного решения уравнений RISM в пределе бесконечного разбавления. Алгоритм обладает свойством масштабируемости и может применяться для любых систем интегральных уравнений, в которых присутствуют операции, независимые по ортогональным размерностям.

References

1. Д.А.Тихонов. Метод интегральных уравнений теории жидкостей для изучения гидратации макромолекул. В книге *Компьютеры и суперкомпьютеры* / Под ред. В.Д.Лахно и М.Н.Устинова. – Москва-Ижевск: Институт компьютерных исследований, 2002, 528 стр., с. 209-233

Когнитивная нейродинамика навигационного поведения млекопитающих

Цукерман В.Д., Еременко З.С., Кулаков С.В., Каримова О.В.

НИИ нейрокибернетики им. А.Б.Когана Южного федерального университета

[Email: vdts@krinc.ru](mailto:vdts@krinc.ru)

Abstract

The entorhinal-hippocampal system cells activity in navigation represent localized spatial patterns of the hippocampal «place cells», grid patterns in the entorhinal cortex and «head direction» cell activity in the subiculum. In this paper represented conceptual model of the organized space-time ensemble cell activity at the object localization and path integration in navigation.

векторных моделях основное управление навигацией осуществляется вектором направления головы, а кандидатом для такого вектора является клетка направления головы. При этом клеткам места отводится роль выбора того или иного вектора направления, но они не используются во время движения вдоль вектора. В векторной модели, когда животное совершает навигацию из одного места в другое, оно следует схеме направлений путей, а не списку мест.

1.2 Задачи исследования

Недавно две группы исследователей независимо друг от друга установили, что у новорожденных крыс, впервые совершающих обследование окружения, уже есть нормально функционирующие клетки направления головы и клетки места, и несколько позже появляются решетчатые клетки [4,5]. Эти факты говорят о том, что восприятие пространства в значительной мере является врожденным. Это вполне оправдывает рассмотрение моделей нейросетевой организации с «генетически предопределенной» матрицей связей нейронов, реализующих задачи пространственной навигации. Какие это задачи? 1) Определение нейродинамических кодов пространственной информации; 2) поиск нейродинамических коррелятов самолокализации объекта в пространстве; 3) интегрирование траектории и, 4) описание и математическое моделирование механизмов генерации пространственно периодических (решетчатых) полей активности. Кроме того, как известно, один из главных результатов экспериментального исследования навигационного поведения в условиях свободного поведения заключается в построении пространственно-временных частотных карт активности нейронов гиппокампа-энториальной системы. Однако, исходя из современных данных об осцилляторном характере процессов обработки пространственной информации в гиппокампа-энториальной системе [6] и наших исследований нелинейного взаимодействия быстрых (гамма) и медленных тета-ритмов [7], более перспективным представляется разработка метода комбинированного представления пространственных признаков в виде семейства фазовых кривых.

1 Введение

1.1 Общие сведения о навигации

Пространственная навигация является фундаментальной формой взаимодействия организмов с внешним окружением. Понимание механизмов, лежащих в основе генерации внутренних карт внешнего мира и их использование в форме навигационных стратегий является обширным полем исследования современной нейронауки. С другой стороны, изучение навигации в реальном мире представляет огромный интерес не только для фундаментальной, но и прикладной науки, связанной с созданием новейших технических средств навигации.

Последние годы принесли много новых нейробиологических данных об организации систем управления навигацией млекопитающих. Центральная роль в этой системе принадлежит гиппокампа-энториальной системе. Получены детальные сведения об активности специализированных «клеток места», «клеток направления головы» и «решетчатых клеток» в условиях свободного поведения животных [1-2]. На основе этих сведений предложены две модели навигации: когнитивного пространственного картирования и векторные модели. Суть первых заключается в формировании и запоминании «списка мест», последовательно посещенных организмом в процессе обследования окружения. «Когнитивная карта» является внутренним представлением окружения, которая позволяет организму выбирать наилучший путь к цели, используя при этом отношения между различными наземными ориентирами окружения.

Иную схему предлагает модель навигации, основанная на направляющих векторах [3]. В

2 Основные результаты исследования

2.1 Комбинированные фазовые представления пространственных характеристик окружения

Исследования модели ЕСИ-сети свидетельствуют, что нейросетевая популяционная активность может быть описана как карта пространственно-временной энергии. Показана возможность комбинированных фазовых представлений пространственных характеристик окружения. Так, дирекциональную настройку и пространственную частоту решеток нагляднее и естественнее представлять не в виде пространственных частотных карт, а в виде семейства кривых фазовой настройки. Левые ветви фазовых кривых при этом представляют дирекциональную настройку информационных единиц, правые – кодирование пространственной частоты всего окружения.

Каждая клетка одного и того же кольца ЕСИ-сети имеет одинаковую тета-частоту осцилляций, но различные фазы активности в референтном тета-цикле; ее частота модулируется скоростным входом вдоль определенного направления. Таким образом, все клетки кольца имеют один и тот же пространственный период (шаг решетки), одну и ту же пространственную ориентацию, но различные фазы (плоско-параллельные смещения).

2.2 Интегрирование траектории

Наша модель ЕСИ-сети, рассматривается как часть интегратора траектории, включенного в локальную навигационную систему [7]. Модель объясняет свойства генерации активности специализированных клеток и механизм, с помощью которого такие клетки могут удовлетворять вычислительным требованиям интегрирования траектории. Модификация паттернов активности основана на информации о собственном движении. Рассматривались движения двух типов: линейное (трансляционное) и/или угловое (вращательное). Показано, что паттерны активности (правые ветви фазовых кривых) чувствительны к скорости пространственного движения навигатора. Длина пути, пройденного навигатором, пропорциональна полному фазовому смещению информационных осцилляторов в тета-цикле.

2.3 Нейродинамика самоорганизации

Нейродинамическим коррелятом локализации объекта в пространстве является когерентная активность гиперансамбля ЕСИ-сети. Простое наблюдение последовательностей активации когерентных групп нейронов ЕСИ-сети (ансамблей

информационных единиц), не позволяет сказать, является ли след движения навигатора прямым или криволинейным, имеющим форму I, или U, или C. Для полного пространственного описания, информация о последовательности местоположений должна быть связана с масштабом и угловой информацией. Такая информация может быть получена добавлением к динамике полей активности информации о скорости, направлении и продолжительности движения, каждая из которых не коррелирует непосредственно с местоположениями этих полей.

3 Заключение

Существует ли принципиальная возможность интегрированных представлений пространственного положения, направления и скорости субъекта, осуществляющего навигационное поведение, нейросетевой динамикой ЕСИ-сетей? Метод картирования единственного признака– ориентацию навигатора в пространстве-времени – позволяет объяснить и упорядоченные смещения в паттернах активности информационных единиц сети, и объединение различных стимульных признаков в популяционном ответе.

References

- [1] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, E.I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, vol. 436, p. 801–806, 2005.
- [2] B.L. McNaughton, F.P. Battaglia, O. Jensen, E.I. Moser, M.-B. Moser. Path integration and the neural basis of the «cognitive map». *Nat. Rev. Neurosci.*, vol. 7, p. 663–678, 2006.
- [3] J.L. Kubie, A.A. Fenton. Heading-vector navigation based on head-direction cells and path integration. *Hippocampus*, vol. 19, p. 456–479, 2009.
- [4] T.J. Wills, F. Cacucci, N. Burgess, J. O'Keefe. Development of the hippocampal cognitive map in preweanling rats. *Science*, vol. 328, p. 1573–1576, 2010.
- [5] Langston R.F., Ainge J.A., Couey J.J., Canto C.B., Bjerknes T.L., Witter M.P., Moser E.I., Moser M.-B. Development of the spatial representation system in the rat. *Science*, vol. 328, p. 1576–1580, 2010.
- [6] C. Geisler, D. Robbe, M. Zugaro, A. Sirota, G. Buzsaki. Hippocampal place cell assemblies are speed-controlled oscillators. *PNAS USA*, vol. 104, p. 8149–8154, 2007.
- [7] В.Д. Цукерман, О.В. Каримова, С.В. Кулаков. Современные нейробиологические данные и новое в нейродинамике навигационного поведения. «Нейрокомпьютеры: разработка и применение», т.2, с.17–27, 2010.

Иерархия математических моделей первичной зрительной коры

Чижев А.В.

Физико-технический институт им. А.Ф.Иоффе РАН, С.-Петербург
anton.chizhov@mail.ioffe.ru

Abstract

To study the ability of neurons of the primary visual cortex to select the orientation of visual stimuli, a hierarchy of models is built, which provides a tool for both biophysically-detailed simulations of cortical activity and mathematical analysis. The most detailed model considers neuronal populations distributed in the retinotopic space along the cortex and receiving a thalamic input according to the canonical pinwheel architecture. The model is based on the conductance-based refractory density (CBRD) approach. The steps of model reduction are: 1) the two-compartmental conductance-based adaptive and one-compartmental fast spiking neuron models are mapped to leaky integrate-and-fire (LIF) neurons, 2) a pinwheel architecture in retinotopic space is approximated by a ring in the orientation space, 3) the second order synaptic kinetics is replaced by instantaneous kinetics, 4) the number of neuronal populations is reduced from two to one, 5) the CBRD approach for LIF neurons is replaced by the Kolmogorov-Fokker-Planck (KFP) equation describing the dynamics in voltage phase space, 6) the KFP is substituted by an artificial relaxation equation for the firing rate driven by its stationary dependence on the input conductance and current (f-g-I-function), and 7) the f-g-I-function is approximated by a threshold linear f-I-curve, thus arriving to the classical FR ring model. With the appropriate analytical expressions the parameters of the various models are mapped to four parameters of the FR ring model. The comparison shows consistency of steady-state solutions. The transient solutions reveal the roles of each of the assumptions made for different models.

1 Введение

Нейроны первичной зрительной коры организованы в колонки, отвечающие на стимул определенной ориентации [4]. Набор колонок всевозможных ориентаций составляет гиперколонку (Рис.1). Гиперколонки принимают сигналы со зрительной информацией согласно ретинотопической проекции. Простейшая модель, отражающая эффекты ориентационной селективности нейронов одной гиперколонки, основана на частотной модели и рассматривает распределенные по кольцу популяции однотипных нейронов, предпочитающих стимул с определенным углом ориентации, Firing-Rate Ring Model (FR-ring)

[3]. В настоящей же работе предлагается биофизически детальная модель, все элементы которой необходимы для воспроизведения характерных ответов нейронов зрительной коры при зрительной или внеклеточной электрической стимуляции.

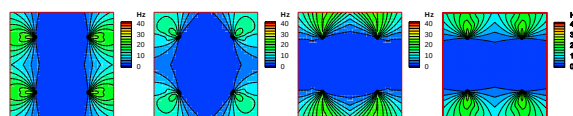


Рис.1. Пример решения, полученного с помощью CBRD-модели. Распределение спайковой активности пирамидных нейронов в 1 мм^2 первичной зрительной коры, содержащем 4 гиперколонки (pin-wheels), в разные моменты времени до (слева) и после мгновенной смены ориентации стимула с горизонтальной на вертикальную. На третьем кадре заметен «перехлест» (tilt after-effect).

Для построения и анализа подробной, но вычислительно эффективной 2-D распределенной модели взаимодействующих популяций зрительной коры был использован недавно разработанный подход для описания одной популяции нейронов типа Ходжкина-Хаксли, где под популяцией понимается бесконечное число несвязанных нейронов, получающих общий входной сигнал и различный для каждого шум [1,2,6]. Такой подход называется Conductance-Based Refractory Density (CBRD) approach – подход на основе рефрактерной плотности. CBRD-подход может быть эффективно использован для реалистичного моделирования активности связанных нейронных популяций. В случае 2-D модели коры, подход предлагает систему уравнений в частных производных с независимыми координатами пространства, фазы и времени.

С другой стороны, сильно упрощенная модель FR-ring предлагает ясное объяснение механизма усиления сигнала ориентации нейронами первичной зрительной коры, а также эффекта инвариантности ответа коры к контрасту стимула. При этом FR-ring модель обладает лишь пятью параметрами и допускает аналитический анализ решений. На плоскости основных параметров, коэффициентов внутри-корковых синаптических связей, FR-ring модель предсказывает границы области с решениями, соответствующими упомянутым эффектам. Задача настоящей работы состояла в

нахождении проекционных соотношений между биофизически подробной 2-D моделью коры и FR-ring моделью. Найденные соответствия систематизируют целую последовательность хорошо известных моделей computational neuroscience, обычно употребляющихся порознь. Тем самым результаты математического анализа простых моделей становятся применимыми для сложных, и наоборот, выводы и допущения простых моделей верифицируются с помощью сложных.

2 Методы

2.1 Подробная 2-D модель зрительной коры.

Предлагаемая модель зрительной коры рассматривает континуум нейронных популяций, распределенных в 2-D пространстве вдоль коры и получающих стимуляцию из таламуса в соответствии с архитектурой типа pin-wheels [5] (Рис.1). Каждая популяция описывалась системой уравнений CBRD-модели [1,2]. Пирамидные нейроны состояли из сомы и дендрита [6]. Учитывалась кинетика возбуждающих AMPA- и тормозных GABA-A-рецепторных синаптических проводимостей, управляемых частотой спайков пресинаптической популяции [6]. Пространственная структура связей внутри коры задавалась изотропной с экспоненциально убывающей зависимостью синаптических весов от расстояния между пре- и пост-синаптическими популяциями.

2.2 Упрощения

Последовательность упрощений такова: 1) модель пирамидного нейрона с сомой и дендритом упрощена до точечной модели нейрона-порогового интегратора (LIF); 2) архитектура связей pinwheels спроецирована на кольцо с координатой-углом ориентации стимула; 3) синаптическая кинетика заменена на мгновенную; 4) число нейронных популяций сведено к одной, распределенной по кольцу; 5) CBRD-подход редуцирован до уравнения Фоккера-Планка (FP) для функции распределения нейронов в пространстве мембранного потенциала; 6) уравнение FP заменено на релаксационное уравнение для популяционной спайковой частоты со стационарной зависимостью частоты от входного тока и проводимости для LIF с шумом; 7) последняя зависимость приближена однопараметрической кусочно-линейной функцией. Тем самым, получается классическая частотная модель кольца популяций. Для связи параметров полной и промежуточных моделей с 5 параметрами FR-ring модели найдены аналитические соотношения.

3 Результаты

Сравнение моделей подтвердило согласованность стационарных профилей распределения активности нейронов в пространстве угла ориентации. Во всех моделях воспроизводятся

эффекты усиления ориентации стимула и инвариантности к контрасту (Рис.2).

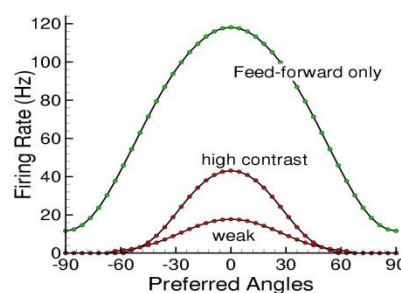


Рис.2. Эффект инвариантности к контрасту. Форма «айсберга» активности в пространстве ориентации стимула сохраняется за счёт внутрикорковых взаимодействий.

Переходные режимы активности (Рис.1) в ответ на смену ориентации стимула воспроизводятся в моделях по-разному. CBRD и FP модели реагируют быстрее чем FR, воспроизводят всплески синхронных спайковых ответов нейронов. CBRD не подтверждает задержки, якобы вызванной эффектом ложного вращения (virtual rotation) в FR. Адаптивная CBRD-модель воспроизводит адаптационные эффекты (“tilt after-effect”) (Рис.1). В целом, результаты показывают, что иерархия моделей может служить полезным инструментом для согласования математических моделей с экспериментальными данными и их анализа.

Работа поддержана грантами РФФИ-НЦНИ 07-04-92167а и РФФИ-09-01-00473-а.

References

- [1] Chizhov AV, Graham LJ. Efficient evaluation of neuron populations receiving colored-noise current based on a refractory density method // *Phys Rev E*, 2008, v. 77: 011910.
- [2] Chizhov AV, Graham LJ. Population model of hippocampal pyramidal neurons, linking a refractory density approach to conductance-based neurons // *Phys Rev E*, 2007, v. 75: 011924.
- [3] Hansel D., Sompolinsky H. Modeling feature selectivity in local cortical circuits. In *Methods in neuronal modeling: From synapses to networks*, eds. C. Koch & I. Segev, 2nd ed., Cambridge, MA: MIT Press, pp. 499–567, 1998.
- [4] Hubel D.H., Wiesel T.N. (1962) Reception fields, binocular interaction and functional architecture in the cats visual cortex. *J. Physiol.Lond.*, v. 160: 106–154;
- [5] K.Kang, M.Shelley and H.Sompolinsky. Mexican hats and pinwheels in visual cortex. *PNAS*, vol. 100(5), 2848–2853, 2003.
- [6] Чижов А.В. Модель популяций нейронов как элемент крупномасштабной нейросети. *Нейрокомпьютеры: разработка, применение*, No.2-3, стр.60-68, 2004.

Анализ взаимодействия нейронов на основе математической модели информационной среды

Шац В.Н.

Независимый исследователь, Санкт - Петербург
vlnash@mail.ru

Analysis of the neurons interaction based on a mathematical model of the information environment

Shats Vladimir N.

Independent researcher, St. - Petersburg
vlnash@mail.ru

Abstract

The model of the information environment in which dependence between the internal and external information in a local network neurons is described by the integral equation is offered. The mechanism of neurons interaction in a network and the module of networks is considered.

1. Информационные процессы в сети

В работе рассматривается взаимодействие нейронов в составе локальной сети и модулей локальных сетей. Мозг моделируется как некоторая информационная среда (ИС), которая обеспечивает взаимодействие механизмов управления, передачи и переработки информации [4]. Под ИС понимается область, содержащая структурированное множество элементов, способных получать, обрабатывать и передавать информацию другим элементам [1, 2]. Рассматривается следующая математическая модель ИС [6, 7].

Пусть множество некоторых последовательно соединенных элементов, образуют первичную цепь. К каждому из них присоединена цепь других элементов. Совокупность первичной цепи и вторичных цепей образуют ветвящуюся цепь элементов (ВЦЭ). Все элементы получают и обрабатывают информацию, а затем передают примыкающему элементу переработанную информацию. ИС содержит множество ВЦЭ. Элемент первичной цепи вместе с присоединенной к нему вторичной цепью будем именовать расчетный нейрон (РН).

Для непрерывной ВЦЭ, когда размерами элементов можно пренебречь по сравнению с длиной цепи, распространение информации описывается системой из двух дифференциальных уравнений в частных производных. При определенных значениях параметров ВЦЭ система сводится к интегральному уравнению Вольтерра первого рода.

В качестве модели локальной сети нейронов

рассмотрим дискретный вариант ВЦЭ, в которой RH_k имеет координату x_k и воспринимает компонент q_k вектора внешней информации $Q(q_1, \dots, q_M)$, $k = \overline{1, M}$. Вектор $C(c_1, \dots, c_M)$ определяет внутреннюю информацию, перемещающуюся в ВЦЭ.

Используя отмеченное уравнение, получим зависимости для вычисления величин $c_1, \dots, c_M$. Поскольку эти зависимости являются рекуррентными, то вектор C определен в системе координат, которая имеет древовидную архитектуру. Такие системы имеют преимущество для познания сложных объектов материального и духовного мира перед системой с независимыми координатами [5].

Из уравнения также следует, что c_k соответствует моменту времени $t = x_k / v$, где v - скорость распространения информации по первичной цепи ВЦЭ. Поэтому c_k можно рассматривать как оценку всей полученной сетью информации по k -ой составляющей вектора Q , которую «дает» RH_k . Соответственно нейрон, включенный в разные функциональные сети, по-разному оценивает («воспринимает») полученную сетью внешнюю информацию.

Можно предположить, что существует некоторая латентная шкала уровня значимости внешней информации для исполнения сетью своего функционального назначения, в которой отражается накопленный опыт адаптации. В соответствие с ней сеть может по M признакам отнести информацию C к тому или иному классу. Распознав класс, сеть может дать ей свою оценку в виде некоторой информации r . Согласно [3] она представляет собой «послания к высшим центрам», элементам ВЦЭ следующего иерархического уровня. Им может служить нервный центр или рецепторы исполнительных структур.

Большое число синапсов и высокая объемная плотность «упаковки» нейронов порождают иррадиацию возбуждения, вследствие которой информацию Q получает ансамбль из N сетей, имеющих разные свойства. В каждой из сетей внешняя информация преобразуется в свою

внутреннюю информацию C_n , $n = \overline{(1, N)}$, а также получает свою независимую оценку уровня значимости r_n для этой сети. Ансамбль передаст на следующий иерархический уровень оценку $R(r_1, \dots, r_N)$.

2. Учет случайных факторов

Рассмотрим некоторые особенности информации, которая поступает в локальную сеть. С этой целью отметим ряд обстоятельств, касающихся сопоставления ВЦЭ и сети в момент времени $\bar{t}$, когда сеть дает оценку r поступившей информации:

1. Величина q_k моделирует поток сигналов, которые имеют вид последовательности отдельных импульсов, поступающих в многочисленные синапсы через разнообразные функциональные сети от группы рецепторных нейронов. На перемещение сигнала между синапсом и аксоном требуется определенное время.

2. Сигнал C_k возникает только при условии, что k -ый нейрон находится в возбужденное состояние, при котором его электрический потенциал превышает пороговое значение. Поэтому в момент $\bar{t}$ все M нейронов должны находиться в возбужденном состоянии, а в сеть поступать сигналы, соответствующие информации сети $Q(\bar{t})$ о воздействии на организм внешней среды. Здесь под $Q(\bar{t})$ понимается полученная нейронами информация, которая отражает воздействие внешней среды на организм и имеет когнитивную интерпретацию для сети, а не случайный набор сигналов $q_1, \dots, q_M$.

Когерентность достигается благодаря существованию нейронов с разной скоростью проведения возбуждения, многочисленности синапсов и путей поступления сигналов в сеть, большому разнообразию медиаторов.

3. Сеть вновь окажется в режиме оценки через некоторый промежуток времени T . В момент времени $\bar{t}_1 = \bar{t} + T$ информация $Q(\bar{t}_1)$ поступит в другую группу нейронов, будет переработана в $C(\bar{t}_1)$ и получит оценку $R(\bar{t}_1)$.

Из этого анализа следует, что ВЦЭ следует рассматривать как детерминированную модель сети, а векторы Q, C - как реализацию соответствующих случайных процессов. Время T определяет некоторый случайный промежуток реального времени, который характеризуется способностью сети получить информацию, переработать ее и дать ей оценку. Следовательно, сеть получает и передает

информацию не непрерывно, а некоторыми порциями - квантами информации. Вывод согласуется с предположением о квантовой природе сознания [5].

Поэтому время $\bar{t}$ - это некоторое «внутреннее» время нервной системы, которое обеспечивает синхронизацию работы ее сетей. Существование такого времени определяется особенностью работы сознания человека, где оценивается не абсолютное время, а временные отношения событий [4].

Ансамбль сетей, которые совместно дают оценку R для внешней информации, является модулем нервной системы. Это подвижная динамическая конструкция со случайным участием в ней нейронов. Внутри нее пульсируют потоки информации. На промежутке времени $(\bar{t}, \bar{t}_1)$ нейроны сети получают сигналы, но они не составляют информацию Q для сети. Некоторые из нейронов могут войти в состав другой сети или превратиться в «молчащие», может измениться состав сетей ансамбля. Со случайной частотой модуль передает информацию на следующий иерархический уровень.

Заключение

В работе предложена модель для описания механизма взаимодействия нейронов в структурах мозга, которая согласуется с существующими данными и имеет простую численную реализацию.

Литература

- [1] Г.Н. Борисюк, Р.М. Борисюк, Я.Б. Казанович, Г.Р. Иваницкий. Моделирование динамики нейронной активности и обработки информации мозгом – итоги десятилетия. //Успехи физических наук. Т. 172(10). с. 1190 - 1214. 2002.
- [2] Н.А. Кузнецов, В.А. Любецкий, А.В. Чернавский. О понятии информационного взаимодействия, 1: допсихический уровень. //Информационные процессы. том 3. №1. с. 1 - 22. 2003.
- [3] Дж. Николлс, Р. Мартин, Б. Валлау, П. Фукс. От нейрона к мозгу, 4-е издание: Пер с англ. : УРСС. с. 672. 2003.
- [4] Физиология человека. Том 1. Под ред. В.М. Покровского, Г.Ф. Коротько. –М.: Медицина. с. 448. 1997.
- [5] А. Ю. Хренников. Моделирование процессов мышления в р-адических системах координат. — М.: ФИЗМАТЛИТ. с. 296. 2004.
- [6] В.Н. Шац. Непрерывно ветвящаяся цепь как модель биологической цепи нейронов. //Труды одиннадцатой национальной конференции по искусственному интеллекту с международным участием КИИ-2008. том 1.- М.: ЛЕНАНД. с. 164 – 170. 2008.
- [7] В.Н. Шац. Модель биологической цепи нейронов на основе непрерывно ветвящейся цепи элементов. //Сборник научных трудов XII Всероссийской научно-технической конференции «Нейроинформатика-2009». часть 2. с. 10 – 18. 2009.

Автоматизированное изучение электростатического потенциала молекул ДНК

Панюков В. В.⁽¹⁾, Осипов А. А.⁽²⁾

Институт математических проблем биологии РАН⁽¹⁾, Института биофизики клетки РАН⁽²⁾
panjukov@itaec.ru, aosypov@gmail.com

Abstract

Nucleic acids are highly charged molecules and electrostatic interactions must play an important role in many aspects of their structure and function. Here in the paper the electrostatic potential of *Escherichia coli* (*E.coli*) is studied. The program package was developed for this aim. It is found that some electrostatic potential functions have a particular shapes within promotor regions.

Цель работы

Предсказание регуляторных, и, прежде всего, промоторных районов требует интеграции разнородной информации, закодированной в последовательностях ДНК. Сюда относятся геометрия двойной спирали, изгибаемость, плавкость, динамические характеристики, и т.п.

Список величин, используемых для локализации промоторов в настоящее время, достаточно обширен. Однако электростатический потенциал для этой цели ещё не привлекался.

Цель данной работы систематическое изучение электростатического потенциала *E.coli* для вовлечения его в систему распознавания промоторов ДНК. Для этого был разработан комплекс программ, который предназначен для выявления особенностей поведения потенциала в районе промоторов.

Выбор модели и средств

Известен ряд подходов для вычисления электростатического потенциала нуклеотидной последовательности. Мы используем работу [1], где был предложен метод вычисления электростатического потенциала полноатомной молекулы ДНК, длина которой сравнима с размером генома бактерий. Далее будем говорить только об этой модели электростатического потенциала. Позднее в работах [2,3,4,6] были проведены исследования электростатических свойств некоторых геномов, которые показали важность электростатических взаимодействий промоторной ДНК и РНК-полимеразы для регуляции функций промоторов. При этом электростатические свойства промоторной ДНК характеризуются выраженными

паттернами, специфичными для различных групп промоторов, которые могут играть роль сигнальных элементов в дифференциальном распознавании соответствующих промоторов РНК-полимеразой. Поскольку эти выводы базировались на небольшой, по объёму, статистике, то было решено разработать компьютерное обеспечение для автоматизации исследования электростатического потенциала молекул ДНК.

Объект исследования

Исследуется электростатический потенциал *E.coli*. Позиции 939-ти промоторов (всего их в *E.coli* около 4000) и значение потенциала на поверхности цилиндра вдоль оси молекулы с шагом один ангстрем брались из базы данных [5]. Набор используемых промоторов обозначаем P_{pm} .

Далее под электростатическим потенциалом молекулы ДНК будем понимать среднее значение потенциала $P(x_i)$ в центре x_i каждой пары оснований, $i=1,2,\dots$. Центры пар оснований лежат на оси молекулы ДНК. Опуская индексы, мы получаем электростатический потенциал $P(x)$ как функцию аргумента x , который пробегает номера позиций оснований молекулы. В системе координат аргумент-значение мы видим потенциал как график некоторой функции.

Постановка задачи. Метод исследования

Идеальная ситуация выглядит так. График потенциала в окрестности промоторов P_{pm} имеет заметные особенности. Причём процент этих особенностей на всей молекуле близок к $(100 \times 4000) / (\text{длина молекулы } 4639675) \approx 0.9\%$. Более общо, такими свойствами обладает, в идеале, если не сам потенциал, то некоторая функция от потенциала. На деле ситуация достаточно далека от идеальной. Сформулируем задачу.

Интервал на множестве целых чисел назовём окном, а длину интервала – размером окна. Применительно к позициям оснований мы получаем окно на молекуле. Далее рассматриваются только окна на молекуле. Если промотор принадлежит окну, то такое окно назовём окном промотора. В

окне размера WinSz мы имеем WinSz значений потенциала. Рассмотрим множество окон фиксированного размера $W(\text{WinSz})$. Функцию на множестве $W(\text{WinSz})$ мы называем свойством потенциала. Множество всех функций на $W(\text{WinSz})$ обозначим как $\text{Prop}(\text{WinSz})$.

Пример, для каждого $w \in W(\text{WinSz})$ положим $f(w) = \max\{P(x) \mid x \in w\}$. Свойство $f(w)$ есть ни что иное как максимальное значение потенциала в окне.

Свойства потенциала – инструмент нашего исследования.

Особенности поведения потенциала и свойства потенциала связаны между собой следующим образом. Введём понятие популярности свойства. Пусть f некоторое свойство потенциала и $\text{Int}(f) = \max(f) - \min(f)$ – интервал значений этого свойства на промоторах Prm и пусть $\text{Dist}(f)$ распределение значений f вдоль всей молекулы ДНК. Назовём популярностью свойства f процентную частоту интервала $\text{Int}(f)$ в распределении $\text{Dist}(f)$. Если популярность свойства f невелика, то можно говорить, что свойство f определяет особенности поведения потенциала.

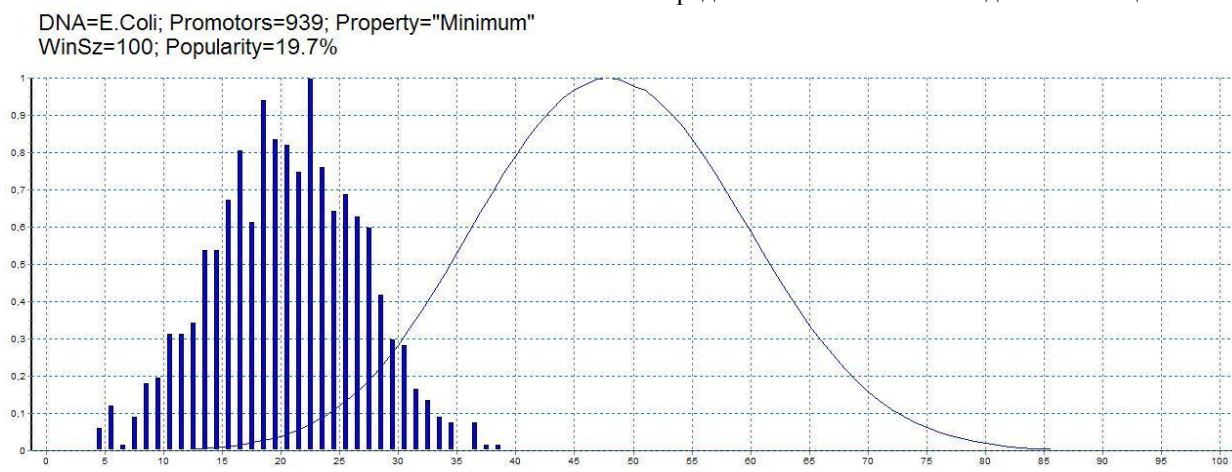


Рисунок 1. Частотные распределения значений свойства «Minimum» на промоторах и на всей E.coli (непрерывная линия). Размер окна WinSz=100. Популярность этого свойства 19,7%.

Результаты

Работа представляет начальный этап автоматизированного исследования свойств электростатического потенциала молекул ДНК на примере E.coli. Рассмотрено семнадцать свойств потенциала.

На рисунке 1 показан выходной файл программного пакета для свойства «Minimum», минимальное значение потенциала в окне для E.coli. Популярность этого свойства 19,7%. Но для фага T7 популярность этого же свойства 31,2%.

Аналогичные результаты имеют место и для других рассмотренных нами свойств.

Выводы

Рассмотренные нами свойства электростатического потенциала E.coli имеют некоторые особенности поведения. Однако эти особенности не являются универсальными для геномов бактерий как видно из T7.

Требуются дальнейшие исследования потенциала.

Работа выполнена при финансовой поддержке РФФИ, проект ,09-07-00455

Литература

- [1] Polozov R.V., Dzhelyadin T.R., Sorokin A.A., Ivanova N.N., Sivozhelezov V.S., Kamzolova S.G. Electrostatic potentials of DNA. Comparative analysis of promoter and nonpromoter nucleotide sequences. J. Biomol. Struct. Dyn. 1999. Vol.16, No.6. pp.1135-1143.
- [2] Dzhelyadi T.R., Sorokin A.A., Sivozhelezov V.S., Ivanova N.N., Polozov R.V., Kamzolova S.G. New approaches in studying RNA polymerase promoter recognition code J.Biomol.Struct.Dyn. 2001 Vol.18 pp.992-993.
- [3] Sorokin A.A., Dzhelyadin T.R., Ivanova N.N., Polozov R.V., Kamzolova S.G. The quest for new forms of promoter determinants. Relationship of promoter nucleotide sequences to their electrostatic promoter distribution J.Biomol.Struct.Dyn. 2001 Vol.18 pp.1020.
- [4] Sorokin;A.A., Beskaravainy;P.M., Osypov;A.A., Kamzolova;S.G. Electrostatic map of E.coli genome DNA. Specific features of electrostatic potential of promoter and nonpromoter regions J.Biomol. Struct. Dyn. 2005 Vol.22, No.6 pp.791-792.
- [5] DEPPDB <http://promodel.icb.psn.ru>
- [6] Джелядин Т.Р., Сорокин А.А., Иванова Н.Н., Сивожелезов В.С., Камзолова С.Г. Особенности электростатического взаимодействия РНК-полимеразы E.coli с промоторами фагаT4 Биофизика 2001 Т.46, С.972-976.

Existence of stable cycles in some multidimensional gene network models

Golubyatnikov V.P.(1), Golubyatnikov I.V.(2), Kleshchev A.G.(3)

Sobolev institute of mathematics(1), Novosibirsk state university(2),

Ugra Scientific Research Institute of Informatics Technology(3)

glbtn@math.nsc.ru

Abstract

We find sufficient conditions of existence and stability of periodic trajectories in nonlinear dynamical systems as models of functioning of gene networks with negative feedbacks. Some generalizations to the models with simplest combinations of positive and negative feedbacks are indicated as well.

1 Kinetic model of gene networks

1.1 Dynamical system

The main aim of our studies is to give mathematical explanations and some predictions for numerical experiments with gene network models regulated by negative feedbacks, where we have observed a lot of periodic trajectories. So, we consider non-linear odd-dimensional dynamical systems of chemical kinetics of the following type

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_{2k+1}) - x_1; & \frac{dx_2}{dt} &= f_2(x_1) - x_2; \\ &\dots & & \\ \frac{dx_{2k+1}}{dt} &= f_{2k+1}(x_{2k}) - x_{2k+1}. \end{aligned} \quad (1)$$

All functions $f_{j+1}(x_j)$ here are smooth, positive, monotonically decreasing. Usually, they have the form

$$f_{j+1}(x_j) = \frac{\alpha}{1 + x_j^\gamma}, \text{ see [1,5] etc., and describe the}$$

negative feedbacks in regulation of the gene networks. Let $\alpha_j = f_j(0)$ be the maximum value of the function f_j . We assume here and below that $j+1 = 1$ for $j=2k+1$.

1.2 Existence of cycles

The function $F(x_1) := f_1(f_{2k+1}(f_{2k}(\dots f_2(x_1))))$ decreases monotonically, the stationary point of (1) is defined from the equation $x_1 = F(x_1)$, hence the system (1) has exactly one stationary point $X^* = (x_1^*, x_2^*, \dots, x_{2k+1}^*)$. All trajectories of (1) enter the domain $Q = [0, \alpha_1] \times \dots \times [0, \alpha_{2k+1}]$ eventually. We decompose Q by $2k+1$ hyper-planes $\{x_j = x_j^*\}$,

$j=1,2,\dots,2k+1$. Let $\beta_j \in \{0,1\}$ be binary indices and $\{\beta_1, \beta_2, \dots, \beta_{2k+1}\} = \{X = (x_1, x_2, \dots, x_{2k+1}) \in Q \text{ such that } x_j \geq x_j^* \text{ if } \beta_j = 1, \text{ and } x_j \leq x_j^* \text{ if } \beta_j = 0\}$.

In the following diagram, one pair of adjacent indices of each of $4k+2$ small parallelepiped coincides and the others alternate. Denote by P' their union.

$\{1010\dots 01\} \rightarrow \{0010\dots 01\} \rightarrow \{01101\dots 01\} \rightarrow \dots$
 $\{1010\dots 0110\} \rightarrow \{1010\dots 100\} \rightarrow \{1010\dots 01\}$ and by d dimension of the dynamical system. As it was indicated above, the indices β_1 and β_{2k+1} are adjacent.

Direct calculations show that trajectories of the system (1) travel through this union P' cyclically from each of these domains in the diagram to the following one. So, P' is an invariant domain of the system (1).

Theorem 1. *If linearization of the system (1) at the point X^* has eigenvalues with positive real parts then this system has at least one cycle in the domain P' .*

Actually, in order to localize position of this cycle, one can construct smaller invariant domain $P'' \subset P'$, composed by $4k+2$ triangle prisms $\Delta \times Q^{2k-1}$, i.e., the products of rectilinear triangles and $(2k-1)$ -dimensional parallelepipeds. Each of these prisms is a half of corresponding parallelepiped in P' obtained by its section by an appropriate diagonal hyper-plane. Both these polyhedral domains P'', P' are non-convex.

2 Existence of stable cycles

2.1 R.Smith's approach

Following R.A.Smith [2], we represent the system

$$(1) \text{ in the vector form } \frac{dX}{dt} = A \cdot X + \Phi(X) \text{ where}$$

$A = -E - \eta \cdot \Sigma$, E is a unit matrix, $\eta > 0$ is a parameter, Σ is an orthogonal matrix of permutation of the coordinate axes, and the vector-function $\Phi(X)$ contains all non-linear terms. We compare the norm

$\max_{j,X} \left| \eta + \frac{df_j}{dx_{j-1}} \right|$ of the Jacobi matrix of $\Phi(X)$ with the norm of the transfer matrix $((i\omega - \mu) \cdot E - A)^{-1}$ of

A. Here ω are μ are real. The next result follows now from the main theorem of [2] and from the theorem on singular decomposition of a matrix.

2.2 Stability of cycles

Theorem 2. *If for some $\eta > 0$ and for all j*

$$\left| \eta + \frac{df_j}{dx_{j-1}} \right| < \eta \cdot \sin \frac{2\pi}{2k+1} \cdot \sin \frac{\pi}{2k+1}, \quad (2)$$

and the conditions of the theorem 1 are satisfied for all $X \in P''$, then the dynamical system (1) has a stable cycle in the invariant domain P'' .

In the case $k=1$ this theorem was proved in [4]. It was indicated in [2] that the main estimate is not sharp there, so the inequality (2) is just a sufficient condition of existence of a stable cycle and is not sharp as well. Stability of bifurcation cycles in the gene networks of the type (1) was discussed in [1,5].

3 Some generalizations and discussions

These theorems can be extended to the case of gene network models of the type similar to (1), when some of the functions f_j are smooth and unimodal, as the

Glass-Mackey functions $f_j(x_{j-1}) = \frac{\alpha \cdot x_{j-1}}{1 + x_{j-1}^\gamma}$. The

unimodal functions describe the simplest combinations of positive and negative feedbacks in the gene networks regulation. Combinatorial structure of the models in these cases is much more complicated than in (1), and the dynamical systems usually have several stationary points. This generates decomposition of the phase portraits of these dynamical systems to blocks of the type Q with just one stationary point inside each of these blocks. As above, we have found conditions of existence of stable cycles near all the stationary points where topological index of the velocity vector field dX/dt equals $(-1)^d$. Here the dimension d of the gene network model can be odd or even. The case $d=3$ with arbitrary unimodal functions in the right hand sides of the equations of the system (1) was studied in [3].

The theorem 1 is valid for the gene networks in presence of post-transcriptional regulation mechanism

when some of the equations in the dynamical system have the form $dx_j/dt = \alpha_j - x_j \cdot g_j(x_{j-1})$. Here the functions g_j are positive and smooth, and grow monotonically.

Theorems 1 and 2 do not state anything on uniqueness of the cycles, so our current tasks are connected with the following problems:

1. Uniqueness of the cycles in the theorems 1 and 2;
2. Evolution of the bifurcation cycles;
3. Detection of unstable cycles in even-dimensional dynamical systems of the type (1);
4. Sharpening of the condition (2).

4 Acknowledgments i

The authors are indebted to V.A.Likhoshvai and Yu.G.Matushkin for helpful discussions.

The work was supported by RFBR, project 09-01-00070, by Interdisciplinary grant 119 of SB RAS, and by ADS-Program of Development of Scientific Potential of Higher School, project 2.1.1/3707.

References

- [1] V.P.Golubyatnikov, V.A.Likhoshvai, Yu.Gaidov, E.P.Volokitin, A.F.Osipov Periodic trajectories and Andronov-Hopf bifurcations in models of gene networks. In: *Bioinformatics of genome regulation and structure II*. Springer, pages 405-414, 2006.
- [2] R.Smith. Orbital stability for ordinary differential equations. *Journ_{al} Diff. Equations*, volume 69(2), pages 265-287, 1987.
- [3] Ю.А.Гайдов, В.П.Голубятников, В.А.Лихошвай. О некоторых нелинейных динамических системах, моделирующих несимметричные генные сети, часть 2. *Вестник НГУ, серия матем.*, том 10, (№ 1), с. 18-28, 2010.
- [4] Ю.А.Гайдов. Об устойчивости периодических траекторий в некоторых моделях генных сетей. *Сиб. Журнал Индустриальной математики*, том 11, (№ 1), с. 57-62, 2009.
- [5] Системная компьютерная биология. Под ред. Н.А.Колчанова и С.С.Гончарова, Новосибирск, СО РАН, 2008.

Unified promoter-search algorithm as a novel tool for annotation of bacterial genomes

Kiselev S.S., Purtov Yu.A., Ozoline O.N.

Institute of cell biophysics, Russian Academy of Sciences

ozoline@icb.psn.ru

Abstract

A computer approach opening a way to reveal promoters in newly sequenced genomes is suggested. It is based on PWM algorithm PlatProm, initially optimized for σ^D promoters of *E.coli*. Renewed version of the software utilizes only invariant features in the structural organization of the promoter DNA.

1 Rationale

A computer-based search for promoters as the main transcription signals has now becoming a common tool for genome annotation. A large number of promoter-search protocols for both bacteria and eukaryotes have been designed. They can reveal all transcribed regions in the genome, including genes encoding proteins and untranslated RNAs. However, most promoter finders explore sequence motifs specifically recognized by DNA-binding center of RNA polymerase as the main components of the scoring system. In spite of apparent evolution stability of the transcription machinery, the context of these motifs is noticeable variable both for promoters of different bacteria and for promoters recognized by different types of RNA polymerase within one and the same bacterial cell. That means that algorithms searching promoters of particular type should be specifically adapted for the context of their conserved elements and can not be used for genomes with uncharacterized regulatory signals. This limitation hinders their utilization in annotation procedures unless universal promoter finders will be designed.

2 From species-specific towards unified promoter-search algorithm

2.1 Scoring system of PlatProm

Along with consensus elements, forming specific contacts with RNA polymerase, PlatProm takes into account conformational features of the promoter DNA in the wide area (-250/+70) nearby transcription start point [1-3]. Specifying promoter regions they include:

- regular distribution of polyA(T) tracts interacting with RNA polymerase α -subunits or stabilizing the transcription complex by a properly induced bend;
- flexible YR (Y=C=T, R=A=G) steps, supporting adaptive isomerization of the DNA helix on the protein surface;

- periodic distribution of mixed A/T-tracts putatively participating in RNA polymerase sliding along the DNA;

- direct and inverted repeats as putative targets for interaction with transcription factors.

Due to similar three-dimensional structure of different RNA polymerases, these features may be invariant in different promoter types. Unlike consensus elements, scored by position-specific weight matrices (PWM), which account expected probability in appearance of all possible base pairs, additional structural features are scored by 60 cascade matrices, which reflect the presence of dominant sequence elements, without scoring all possible variants. Only those structural modules, which occurrence frequencies in particular positions are at least five standard deviations higher than in average in the control set of non-promoter sequences were used to specify different promoter regions. An example of such simplified matrix, accounting the presence of motifs, dominant in the region -55/-52 is shown in Table 1.

Table 1.

Checking order	Position	Element	Score
1	-53	ACAT	1.56
2	-56	CACA	0.90
3	-52	CAT	0.89
4	-56	TCAT	0.70
5	-55	CAT	0.61
6	-57	ACAC	0.44
7	-57	ACA	0.18
All absent			-0.036

Positive scores represent log-probabilities to find 5'-end of each particular segment in the positions of its preferred occurrence. The presence of the most significant motif (ACAT) was first checked in the position where the probability to find it is maximal (-53), then the presence of another typical module was verified in the position -56 and so on in descending

manner. Negative scores reflect the percentage of promoters lacking all elements specifying the region of

interest.

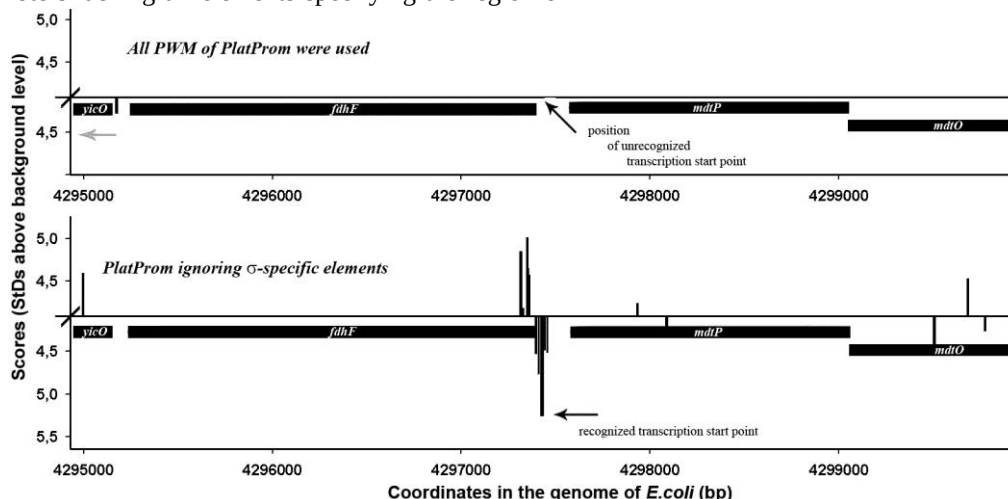


Fig. 1 Distribution of predicted by PlatProm transcription start points (bars) on the both strands (above and below the X-axis) of the *E.coli* K12 genomic DNA in the vicinity (± 2500 bp) of *fdhF*. Positions of 4 genes (all transcribed from the “-” strand) are indicated by black rectangles). Arrows show position of *fdhF* promoter.

The scoring system of PlatProm was balanced so as these additional elements compose on average 50% of the total score. Using this option our program recognizes 85,5% of well characterized *E.coli* promoters from the test compilation with $p < 0,0038$ reliability. If only structure specific elements (cascade matrices) were accounted sensitivity of the program decreases though the percentage of correctly recognized promoters still remains rather high (62,4%). Discriminative ability of structural elements provided, therefore, a chance to detect regulatory regions with different sequence-specific elements.

2.2 PlatProm ignoring scores of σ^D -specific elements accurately recognizes promoter of σ^N

Consensus sequences recognized by σ^N fundamentally differ from the σ^D -specific motifs. Promoter with σ^N signature controls expression of *fdhF* gene of *E.coli*. It was undetected by full version of PlatProm (upper panel on **Fig. 1**) but was found by unified protocol (lower panel in **Fig. 1**). Moreover, the experimentally mapped transcription start point precisely coincided with maximally scoring position (bar indicated by arrow). If σ^D -specific PlatProm recognizes only 20% of σ^N -promoters, unified version pointed out more than 70%.

Performance of renew version of PlatProm was tested on promoters of *Rhizobium etli*, *Sinorhizobium meliloti*, *Lactobacillus casei*, *Lactobacillus acidophilus*, *Corynebacterium glutamicum* and some others. The data obtained indicate that invariant features in the structural organization of promoter sites provide a chance to find them in bacterial genomes without any tuning.

If, however, bacterium with newly sequenced genome has closely-related microorganism with partly characterized regulatory sites, computer search may be further improved using provided by PlatProm procedure of “expectation-maximization”. Thus, practically

unrecognized by initial version of PlatProm ribosomal promoter of *Sinorhizobium meliloti*, was identified by unified version of PlatProm and also by protocol trained on *Rhizobium etli* promoters. Such cross-tuning (if applicable) can substantially reduce nonspecific background, which certainly is a significant lack of a unified program.

So, it is clear, that evolutionary conserved features in the structural organization of the promoter sites provide a chance to create a unified program permitting promoter mapping in genomes without information about their regulatory regions. Even for the studied genomes such a program provides new opportunities. By ignoring specific features it allows comparing normal transcription signals with pseudo-promoters, cryptic promoters and promoter-like signals within promoter islands.

The study is supported by RFBR grant 10-04-01218-a

References

- [1] A.S.Brok-Volchanski, I.S.Masulis, K.S.Shavkunov, V.I.Lukyanov, Yu.A.Purtov, E.G.Kostyanicina, A.A.Deev, O.N.Ozoline Predicting sRNA genes in the genome of *E.coli* by the promoter-search algorithm PlatProm. In: *Bioinformatics of Genome Regulation and Structure II*. (Eds. N.Kolchanov, R.Hofstaedt, L.Milanesi) Springer p. 11-20.
- [2] O.N.Ozoline, A.A.Deev Predicting antisense RNAs in the genomes of *Escherichia coli* and *Salmonella typhimurium* using promoter-search algorithm PlatProm., *J. Bioinform. Comput. Biol.*, v. 4, p. 443-454, 2006.
- [3] K.S.Shavkunov, I.S.Masulis, M.N.Tutukina, A.A.Deev, O.N.Ozoline Gains and unexpected lessons from genome-scale promoter mapping, *Nucl. Acids Res.*, v. 37, p. 4919 – 4931, 2009.

A reaction-diffusion model of shoot apical meristem compartmentalization based on CLV/WUS interplay

Sergey Nikolaev⁽¹⁾, Ulyana Zubairova⁽¹⁾,
Eric Mjolsness⁽²⁾, Pavel Smal⁽¹⁾, Bruce Shapiro, Nikolay Kolchanov⁽¹⁾

⁽¹⁾Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

⁽²⁾Computer Science Department, Center for Computational Morphodynamics, University of
California, Irvine, USA

⁽³⁾Biological Network Modeling Center, Division of Biology, California Institute of Technology,
Pasadena, USA
e-mail: nikolaev@bionet.nsc.ru

1 Motivation and Aim

There is a qualitative hypothesis of the interplay between CLV and WUS genes as a mechanism for the SAM spatial compartmentalization into central zone (CZ – stem cells), organizing center (OC), and peripheral zone (PZ), which all together form stem cell niche. The following is an important summary of the hypothesis: CLV3 expression occurs in the central cells of 3 upper layers (CZ), while WUS expression occurs in the cells of OC, just below CZ; and CLV3 by means of binding with putative receptor CLV1/CLV2 inhibits WUS expression, while WUS activates CLV3 expression. This interplay is believed to be able to regulate stem cell niche structure in the SAM.

We would like to build a reaction-diffusion model to test possibility to obtain biologically plausible compartmentalization of the SAM as a steady-state solution of the model.

2 Methods and Algorithms

Previously we developed a mathematical model of a spatially distributed molecular-genetic mechanism for such a spatial compartmentalization of the SAM [1] to test the possibility. We supposed and/or adopted the next postulates. Some hypothetical gene Y is expressing in the uppermost cells. A product of the Y gene propagates across the SAM by diffusion with simultaneous degradation. There are activation of genes CLV3 and WUS expression in dependence of Y concentration in the SAM cells. The products of the former genes diffuse and degrade in turn. And a receptor CLV1/2 is synthesized in the cells where concentration of the product of WUS gene expression exceeds some threshold. While CLV3 forms complex with CLV1/2 it concentration decreases, and activates a signal pathway that finalizes with repression of WUS expression. This can be the mechanism that provide needed compartmentalization of the SAM, as we demonstrated by performing computer simulations. The model was represented as a system of $(5 \cdot n)$ ordinary differential equations: reaction-diffusion processes for 5 substances

in n cells, arranged in a semicircular domain. Now the model has been realized in the Cellzilla package [2] – a tool in Mathematica package to develop dynamical models distributed on a cellular domain.

3 Results

Now we developed a rules to find values of the model parameters for a given geometry of the SAM compartments. The model was used for computer simulations, and a stationary solution were obtained for the spatial gene expression domains of the genes of interest. The solution is in qualitative accordance with experimentally observed data on longitudinal section of the SAM.

4 Conclusion

The described hypothesized mechanism for stem cell niche structure control in the SAM can be interpreted as a candidate mechanism to capture main features of the experimentally observed interactions between the compartments of the SAM.

References:

- [1] S.V. Nikolaev, A.V. Penenko, V.V. Lavreha, E.D. Mjolsness, N.A. Kolchanov. (2007) A Model Study of the Role of Proteins CLV1, CLV2, CLV3, and WUS in Regulation of the Structure of the Shoot Apical Meristem, Russian Journal of Developmental Biology, 38: 383–388.
- [2] B.E. Shapiro, J. Lu, M. Hucka, E.D. Mjolsness (2007) Platforms for Modeling in Systems Biology: Recent Developments in MathSBML and Cellerator, Poster and Proceedings extended abstract, ICSB 2007, Long Beach California, Oct 2-4 2007. URL: <http://www.icsb-2007.org/proceedings/abstracts/G24.pdf>

Modified French Flag model: instability of the structure under cell growth and division

Ulyana Zubairova⁽¹⁾, Sergey Nikolaev⁽¹⁾, Stanislav Fadeev⁽²⁾, Eric Mjolsness⁽³⁾,
Nikolay Kolchanov⁽¹⁾

⁽¹⁾Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

⁽²⁾Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia

⁽³⁾Computer Science Department, Center for Computational Morphodynamics, University of California, Irvine, USA
e-mail: ulyanochka@bionet.nsc.ru

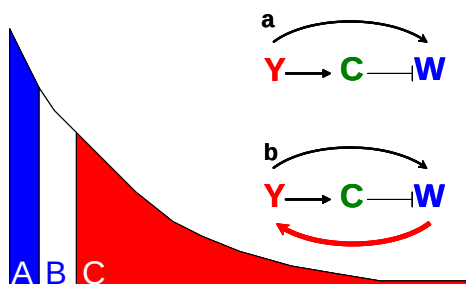
Abstract

In the paper modeling of dynamic systems with dynamic structure in the formalism of parameterized L-systems is described. By an example of model of the renewing zone in biological tissue an influence of the relation between the cell cycle characteristic time and diffusion of morphogens on stability of spatial molecular-genetic control system is considered. It is shown that an application of simulation methods of dynamic systems with dynamic structure reveals new interesting biological mechanisms

1 Motivation and Aim

A minimal 1D-model of stem cell niche structure regulation along vertical axis of the SAM was developed on the basis of a qualitative hypothesis of interplay between the CLV and WUS genes. Previously it was shown that there is a set of parameters supplying a stationary solution in qualitative correspondence with experimental observations [1]. But the question arises what will be the model dynamics under cell growth and division.

The model for a special cell group (OC) positioning on vertical axis in shoot apical meristem looks as a variant of the “French flag model” (scheme “a”) by Wolpert ([2]) because there are arguments to propose additional control from “B”-domain to “A”-domain in the case we model (scheme “b”).



2 Methods and Algorithms

Using DL-system formalism [3] we developed a mathematical model of stem cell niche structure regulation on 1D-array of growing and dividing cells. The cell divisions result in change of the system structure. The number of equations increase, and diffusion terms change. Use of a rule for cell division depends on the set of cell parameters (ϕ , y , c , w).

3 Results

In the issue the dependence of probability of the stem cell niche destruction on cell cycle duration relative to diffusion time scale was obtained. Increase of the specific cell growth rate results in monotonic increase of system destruction probability and in decrease of its mean lifetime.

4 Conclusion

A number of computer simulations were performed to study the model dynamics.

Cell divisions account for relevant perturbation in the SAM structure. The proposed “French flag model” modification may result in destruction of the SAM structure. The stem cell niche survivability depends on relations between model parameters. For example increase of the specific cell growth rate results in monotonic increase of system destruction probability and in decrease of its mean lifetime [4].

Acknowledgements:

This work was supported by the Russian Foundation for Basic Research (08-04-01214-a), SB RAS Integration project 119, RAS program №22 (project 8).

References:

- [1] Nikolaev S.V., Kolchanov N.A., Fadeev S.I., Kogai V.V., Mjolsness E. (2006) A study of a one-dimensional model for control of renewing zone size in biological tissue. Computational technologies, 11. № 2: 67-81.
- [2] Wolpert L. (1969) Positional information and the spatial pattern of cellular differentiation. J Theor Biol 25: 1-47
- [3] Prusinkiewicz P., Lindenmayer A. (1990) The algorithmic beauty of plants. N.Y.:Springer-Verlag.
- [4] Nikolaev S.V., Kolchanov N.A., Fadeev S.I., Kogai V.V., Mjolsness E. (in press) A study of a one-dimensional model for control of renewing zone size in biological tissue taking into account cell division. Sibirskii Zhurnal Industrial'noi Matematiki (in russian)

Elucidation of the functional role of amino acids forming the substrate-binding site of human CYP51A1 by molecular modeling and sequence analysis

D. V. Mukha⁽¹⁾, A. V. Tuzikov⁽²⁾, S. A. Usanov⁽¹⁾

⁽¹⁾Institute of bioorganic Chemistry of NAS, Minsk, Belarus

⁽²⁾United Institute of Informatics Problems of NAS, Minsk, Belarus

d.v.mukha@gmail.com

Abstract

Human CYP51A1 belongs to the family of sterol 14 α -demethylases, which are promising targets for developing of new anti-infection drugs. The biological role of CYP51s is participation in sterol biosynthesis. Our goal was to ascertain the topology and to learn structural and functional characteristics of the active site of human CYP51A1. In order to realize this, simulation of alanine scanning and comparison of primary sequences were applied. We identified amino acids appertaining to the substrate-binding site of CYP51A1, which are essential for substrate binding and molecular structure maintaining.

Introduction

The biosynthesis of steroids has a great importance for the metabolism of most eukaryotic and many prokaryotic organisms. Human lanosterol 14 α -demethylase (CYP51A1) belongs to the superfamily of cytochromes P450. To date, more than 174 protein sequences of CYP51 members has been described. In humans, 14 α -demethylase is involved in cholesterol biosynthesis. It is shown that a product of lanosterol demethylation affects maturation of germ cells, thus being involved in the functioning of the reproductive system [2]. The importance of steroid biosynthesis makes CYP51s the convenient targets for drug development [3]. Sterol 14 α -demethylase is crucial enzyme for synthesis of fungi cell membrane component ergosterol. Most of used today CYP51 inhibitors affect the human CYP51A1 activity. Knowledge about the structural features of human CYP51 are essential for the development of new anti-infection drugs avoiding interaction with orthologous enzyme in humans.

Methods

To clarify the role of amino acid residues composing substrate-binding site 2 methods were chosen: simulation of alanine scanning and comparison of family members' sequences.

Alanine scanning simulation was carried out in Amber 10. We selected a pool of residues, which atoms are able to form contacts with the steroid substrate molecule during molecular dynamics. The list includes the following amino acids (hereinafter, the numbering of

amino acids is provided according to the record NP_000777 of NCBI RefSeq database): V136, Y137, L140, T141, F145, V149, Y151, F158, Q161, K162, L165, F240, H242, W245, L291, M310, L314, L316, H320, T312, I383, M384, I385, M386, M387, R388, I456, M493, I494, H495, totally 30 amino acids.

Alignment of CYP51's family members protein sequences was done by MUSCLE algorithm [3-4]. Alignment consists of sequences deposited in NCBI RefSeq, having at least 36% identity with the sequence of human CYP51A1 (NP_000777) with more than 66% sequence overlapping. 174 sequences satisfied theses criteria. BLASTP 2.2.24 rating of selected sequences compared with human CYP51A1 was greater or equal to 170 units. All sequences in the alignment have no mismatch at the position of cysteine ligating iron atom in heme moiety. Phylogenetic tree based on the alignment was constructed using ClustalX v.2.11.

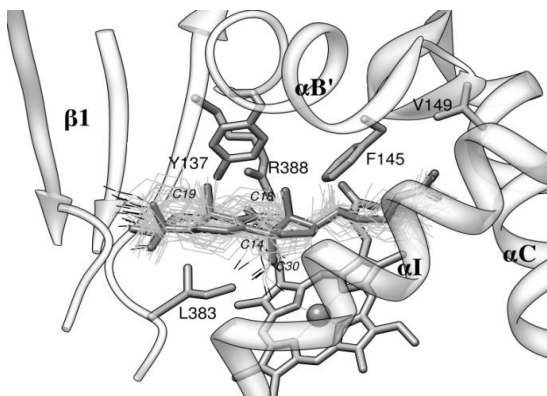


Fig. 1. Substrate-binding site of human CYP51A1 with 25 different steroid molecules docked in it (lanosterol shown in stick).

Results and Discussion

According to alanine scanning simulation, the largest magnitudes of weakening of binding enthalpy was observed in the case of substitutions Y137A, F145A, V149A, I383A, R388A. Substitutions L140A, L165A, F240A, L314A, I494A led to more moderate weakening of binding.

Substitutions of some polar amino acids from the list cause a destabilization of the native protein conformation. Also it was observed that replacement of

non-polar and charged residues with alanine favors contacts with hydrophobic ligands. However, the side chains of amino acids T141, Y151, Q161, H320, T321, H495 are involved in the formation of hydrogen bond network stabilizing the structure of the protein globule. It is noteworthy that the replacement of H320 and H495 by alanine led to an increase of affinity to the final products of the 14 α -demethylation. A magnitude of $\Delta\Delta H$ parameter for all substitutions did not exceed 4 kcal/mol.

In order to verify data on the importance of amino acid residues obtained by simulation of alanine scanning, we performed alignment of sterol 14 α -demethylases family. It is well-known, that most of the conservative residues are involved in protein fold maintenance. A conservation of residues forming the hydrophobic

cavity, designed to interact with the substrate, during evolutionary process is not a strict rule, mainly because of the interchangeability of hydrophobic amino acids. Nevertheless, a comparison of the effect of substitutions on alanine (in terms of $\Delta\Delta H$), estimated by molecular modeling, with the degree of conservativeness of amino acid residues within the family of CYP51 revealed that substitutions leading to the largest $\Delta\Delta H$ values affect the most conserved residues (figure 2). The list of essential for substrate binding residues includes Y137, F145, V149, I383 and R388. The degree of conservation of these residues is more than other residues, also comprising the substrate-binding site, which is related to their contribution to the interaction with the natural substrates of CYP51A1.

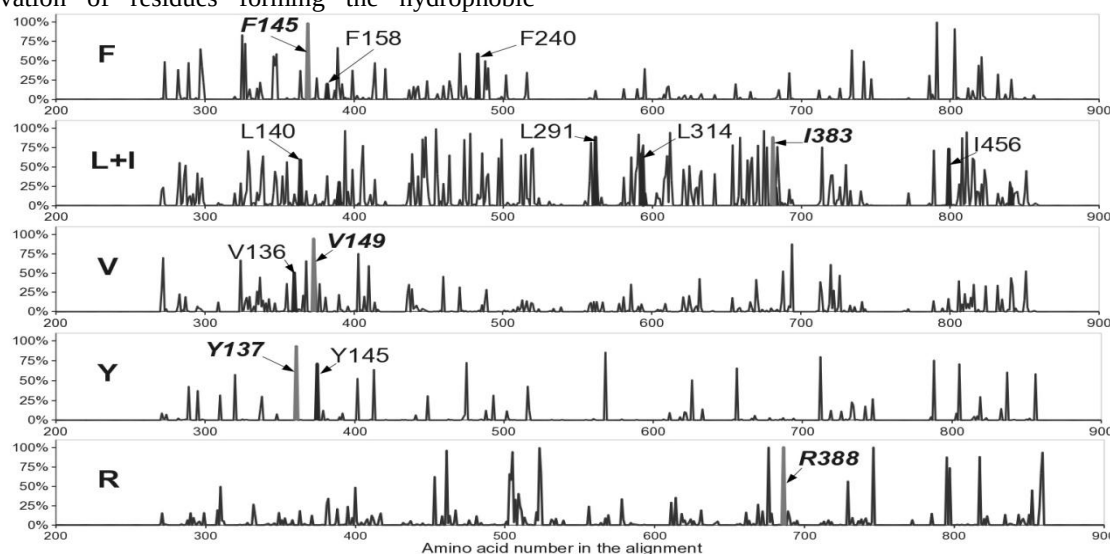


Fig. 2. Amino acid conservation in CYP51 family. Essential for substrate binding amino acids are marked by bold italic font.

Residue F145 forms hydrophobic contacts with side chain and 18th carbon atom of substrate molecule. R388 is an absolutely conservative residue over whole CYP51 family. It interacts with ring A propionate. In addition, R388 is essential for proper substrate binding. It is most likely that its charged side chain influences greatly on local electrostatic potential in substrate-binding site. L140 determines CYP51A1 selectivity toward methyl groups on 4th carbon of steroid molecule [1]. This residue is substituted to phenylalanine in CYP51 family's members preferring 4-monomethylated substrates.

Residue I383 is extremely conservative: main variation found is on the structurally similar amino acid leucine. Two prolines preceding I383 provide interruption of K-helix and juxtaposition of the 383rd residue with the active center cavity. L291 located more distantly from the substrate-binding site exhibits a high degree of conservation (figure 2) because of its structural role. This residue forms hydrophobic contacts with L165 (C-helix), L302, M310 and L311 (I-helix).

A detailed description of the properties of the active site of human CYP51A1 is essential for a rational strategy

for the development of new effective drugs. Computational methods show a capability for elucidation of a functional role of amino acid residues in enzyme's active site.

This work was supported by the program "SKIF-GRID" of Union State of Belarus and Russia.

References

- [1] Lepesheva, G. I., Zaitseva, N. G., Nes, W. D., Zhou, W., Arase, M., Liu, J., Hill, G. C., Waterman, M. R. CYP51 from *Trypanosoma cruzi*: a phyla-specific residue in the B' helix defines substrate preferences of sterol 14 α -demethylase. *J Biol Chem*, **6**, 3577-85, 2006
- [2] Stromstedt, M., Waterman, M. R., Haugen, T. B., Tasken, K., Parvinen, M., Rozman, D. Elevated expression of lanosterol 14 α -demethylase (CYP51) and the synthesis of oocyte meiosis-activating sterols in postmeiotic germ cells of male rats. *Endocrinology*, **5**, 2314-21, 1998.
- [3] Strushkevich, N., Usanov, S. A., Park, H. W. Structural basis of human CYP51 inhibition by antifungal azoles. *J Mol Biol*, **4**, 1067-78, 2010.

Методы линейной алгебры в биохимии и метаболическом моделировании

Ермаков Г.Л.

Институт теоретической и экспериментальной биофизики РАН,

142290 Московская обл., г. Пущино

ermakov-gennady@rambler.ru

Abstract

In general, living systems studied have some border separating it from the external environment in relation to some isolated parts of the system. Therefore, the metabolic network of open biochemical system can always be divided into incoming, outgoing flows and flows related to the internal part of the system. In many cases, the incoming and outgoing flows can be directly measured experimentally, or given as boundary conditions. In this case, knowing the scheme throughout the network, though presumably in a study state mode, you can find a connection between the external, known and the unknown internal flows, using the known, simple procedures of linear algebra. This work is dedicated to finding the matrix equations relating the external, known streams studied metabolic network with the desired internal flows of the network, subject to a steady state system.

Настоящая работа посвящена развитию оригинального граф-теоретический метода, выявления и анализа, ключевых структурно-динамических свойств сложных биохимических реакций, метаболических и генных сетей. Развиваемый граф-теоретический метод основан на том, что детальный механизм сложной реакции, метаболической/генной сети, может быть представлен четырьмя эквивалентными формами:

- список стехиометрических (химических) уравнений в символьной форме, с перечислением всех компонентов и стадий реакции/сети, с соответствующими стехиометрическими коэффициентами;
- система обыкновенных дифференциальных уравнений (кинетические уравнения);
- матричное/векторное представление систем стехиометрических и кинетических уравнений;
- связный ориентированный двудольный граф.

Эквивалентность четырех форм описания сложных реакций и метаболических/генных сетей

позволяет установить взаимно однозначное соответствие между двудольным графом сети реакций, который может быть построен по схеме сети, и двумя ключевыми матрицами - стехиометрической и Якобианом, которые отражают наблюдаемые структурно-динамические свойства изучаемой биохимической системы. В этом случае элементы Якобиана и стехиометрической матрицы имеют наглядный, графический смысл, который позволяет, анализируя схему/сеть путей, выявлять графические фрагменты и образы, ответственные за наблюдаемую динамику и структурные свойства, определяющие функционирование метаболической сети (сложной реакции) как целого. Таким образом, ключевые структурно-динамические свойства биохимических систем имеют графическое толкование, могут быть выявлены и перечислены. Развиваемый метод основан на работах Баландина [1], Вольперта [2], Кларка [3-5], Тайсона [6], Ивановой [7, 8], Ермакова [9-14].

Как правило, изучаемые живые системы имеют некоторую границу, по отношению к некоторой выделенной части системы, отделяющую её от внешней среды. Поэтому, в метаболической сети открытой, биохимической системы всегда можно выделить некоторое количество входящих-выходящих потоков и потоки, принадлежащие «внутренней» части сети, которая может иметь сколь угодно сложную структуру. Во многих случаях входящие и выходящие потоки могут быть непосредственно измерены экспериментально, либо заданы как граничные условия. В этом случае, зная структуру/схему/граф всей сети, хотя бы предположительно, в условии стационарного режима, можно найти связь между внешними, известными потоками и искомыми внутренними, используя простые процедуры линейной алгебры. Найденные уравнения, связывающие внешние и внутренние потоки принято называть балансными соотношениями.

В настоящей работе рассматриваются условия нахождения балансных соотношений (вид стехиометрической матрицы метаболической сети, её ранг, соотношение количества известных внешних потоков к искомым внутренним и т.д.). Используя процедуры линейной алгебры, найдены основные матричные уравнения, выражающие искомые внутренние потоки любой метаболической

сети, через внешние, известные потоки, в условии стационарного состояния системы.

Работа выполняется в рамках проекта № 6663 «Комплексное исследование неспецифической реакции организма на развитие злокачественных образований: экспериментальное и **теоретическое моделирование**» аналитической ведомственной целевой программы “Развитие научного потенциала высшей школы (2009-2010 годы)”

References

- [1] Баландин А.А. Мультиплетная теория катализа. Ч. 3. Теория гидрогенезации. Классификация каталитических органических реакций. Структурная алгебра в химии. М.: изд-во МГУ, 1970.
- [2] Вольперт А.И. (1972) Дифференциальные уравнения на графах. *Матем. сб.*, 88(130), 4(8), 578-588.
- [3] Clark B.L. (1974) Graph theoretic approach to the stability analysis of steady state chemical reaction networks. *J. Chem. Phys.*, **60**, 4, 1481-1492.
- [4] Clark B.L. (1974) Stability analysis of a model reaction network using graph theory. *J. Chem. Phys.*, **60**, 4, 1493-1501.
- [5] Clark B.L. (1975) Theorems on chemical network stability. *J. Chem. Phys.*, **62**, 3, 773-775.
- [6] Tyson J.J. (1975) Classification of instabilities in chemical reaction systems. *J. Chem. Phys.*, **62**, 3, 1010-1015.
- [7] Иванова А.Н. (1979) Условия единственности стационарного состояния кинетических систем, связанные со структурой схемы реакции. I. *Кинетика и катализ*, **20**, 4, 1019-1023.
- [8] Иванова А.Н. (1979) Условия единственности стационарного состояния кинетических систем, связанные со структурой схемы реакции. II. Открытые системы. *Кинетика и катализ*, **20**, 4, 1024-1028.
- [9] Ермаков Г.Л., Гольдштейн Б.Н. Простейшие кинетические схемы биохимических осцилляторов (2002) *Биохимия*, 67, 4, 567-579.
- [10] Ермаков Г.Л. Граф-теоретический метод поиска и анализа критических явлений в биохимических системах. II. Кинетические модели биохимических осцилляторов с участием двух и трех веществ. (2003) *Биохимия*, 68, 10, 1395-1406.
- [11] Ермаков Г.Л. Граф-теоретический метод поиска и анализа критических явлений в биохимических системах. I. Графические правила выявления осцилляторов. (2003) *Биохимия*, 68, 10, 1381-1394.
- [12] Goldstein B.N., Ermakov G.L., Centelles J.J., Westerhoff H.V., Cascante M. What makes biochemical networks tick? A graphycal tool for the identification of oscillophores (2004) *Eur J Biochem*. 271, 19, 3877-87.
- [13] Ермаков Г.Л. Стехиометрический анализ биохимических систем на графах. I. Графические правила нахождения балансных соотношений. (2007) *МАТЕМАТИЧЕСКАЯ БИОЛОГИЯ И БИОИНФОРМАТИКА*, том 2, №1, с. 36-47, [http://www.matbio.org/downloads/Ermakov2007\(2_36\).pdf](http://www.matbio.org/downloads/Ermakov2007(2_36).pdf)
- [14] Ермаков Г.Л. Стехиометрический анализ биохимических систем на графах. II. Графовые образы балансных соотношений (2007) *МАТЕМАТИЧЕСКАЯ БИОЛОГИЯ И БИОИНФОРМАТИКА*, том 2, №1, с. 48-59, [http://www.matbio.org/downloads/Ermakov2007\(2_48\).pdf](http://www.matbio.org/downloads/Ermakov2007(2_48).pdf)

Расчёт реального распределения скоростей реакций в метаболической системе растущей клетки методом баланса стационарных метаболических потоков (БСМП)

Дроздов-Тихомиров Л.Н.⁽¹⁾, Васильчиков В.В.⁽²⁾, Назипова Н.Н.⁽²⁾

ИМГ РАН⁽²⁾, ИМПБ РАН⁽¹⁾,
drozdov@img.ras.ru

Abstract

Program FLUX III is developed out, which allows calculating the real reaction rates distribution which exists in metabolic system of growing cell, if experimental data about material exchange between the cell and medium are known. The mathematical model of cell metabolism, constructed by help of the study-state metabolic flux balance method (SMFB) was used by calculation.

1 Основной раздел

Построение математических моделей метаболизма живой клетки методом БСМП (баланс стационарных стехиометрических потоков) приводит к системе линейных алгебраических уравнений для скоростей реакций, происходящих в моделируемой метаболической системе. Число переменных в такой системе уравнений, равное числу реакций, происходящих в метаболической системе (N), как правило, превышает число уравнений системы, которое равно числу метаболитов, участвующих в реакциях, протекающих в моделируемой метаболической системе (M). Не останавливаясь на причинах, которые обуславливают такое соотношение между N и M , заметим, что вследствие того, что $N > M$, система уравнений модели, очевидно, не может иметь единственное решение и, следовательно, с помощью модели, построенной с использованием метода БСМП, нельзя рассчитать реальное распределение скоростей реакций, которое устанавливается в метаболической системе клетки при росте в заданных условиях.

Это обстоятельство не мешает, однако, проводить с помощью такой модели целый ряд расчетов, имеющих большой практический и теоретический интерес. Применяя метод линейного программирования, на основе модели БСМП можно решать оптимизационные задачи, позволяющие оценивать потенциальные возможности метаболических систем различных биологических объектов.

Нами ранее сообщалось о разработанном нами пакете программ FLUX I, предназначенном для

создания математических моделей клеточного метаболизма методом БСМП и решения на его основе разнообразных оптимизационных задач для моделируемого объекта. Пакет FLUX I был успешно применён для исследования потенциальных возможностей метаболических систем клеток бактерии *E.coli* и митохондрий клеток дрожжей *Saccharomyces cerevisiae*.

В настоящем докладе мы сообщаем о создании пакета программ FLUX III, являющемуся развитием пакета FLUX I, расширяющим его возможности и позволяющим решать не только оптимизационные задачи, но и задачу расчёта реального распределения скоростей реакций в моделируемой метаболической системе.

Как было сказано выше, из-за того, что, как правило, в системе линейных уравнений для скоростей реакций, которая получается при моделировании реальных метаболических систем методом БСМП, число уравнений (M) меньше числа переменных (N), однозначное решение для скоростей реакций не может быть получено. Однако нами было показано, что за счет перехода к новым переменным, число которых будет равно числу линейно независимых уравнений, может быть осуществлена редукция системы уравнений модели, при которой матрица системы становится квадратной, что позволяет получить для новых переменных однозначное решение. Разработан алгоритм перехода к системе переменных, для которых существует однозначное решение. Показано, что среди новых переменных небольшая часть представляет собой некоторые линейные комбинации исходных переменных (т.е. представляют собой алгебраические суммы скоростей некоторых реакций), а остальные (большая часть) являются исходными переменными (т.е. скоростями той или другой отдельной реакции) не входящими в состав упомянутых выше линейных комбинаций.

Пакет FLUX I с включённым в него указанным дополнением, получивший название FLUX III, позволяет, если задаются данные о строении метаболической системы и экспериментальные данные о скорости роста и материальном обмене со средой, определять *реальные скорости* большей

части реакций, происходящих в моделируемой метаболической системе при заданных условиях, а для остальных реакций позволяет указать группы, для каждой из которых может быть однозначно определена *лишь алгебраическая сумма скоростей реакций, входящих в группу*.

Проведенное нами исследование ставит под сомнение результаты работы Палссона [1], полагающего, что моделирование метаболизма клетки дрожжей с помощью метода БССП, и использование полных экспериментальных данных о материальном обмене клетки со средой, позволило ему рассчитать реальное распределение скоростей реакций в метаболической системе клетки дрожжей *Saccharomyces cerevisiae*.

В настоящее время пакет FLUX III использован для расчёта реального распределения скоростей реакций в метаболической системе клетки *E.coli*, растущей на минимальной среде с глюкозой в аэробных условиях. Для расчёта использованы экспериментальные данные о материальном обмене со средой, взятые из литературы. Выявлены группы реакций, для которых может быть однозначно определена только алгебраическая сумма их скоростей, и определены индивидуальные скорости остальных метаболических реакций.

References

- [1] Famili I, Forster J, Nielsen J, Palsson BO. Proc Natl Acad Sci U S A. 2003 Nov 11;100(23)

Detecting past positive selection through ongoing negative selection

Georgii A. Bazykin⁽¹⁾, Alexey S. Kondrashov⁽²⁾

⁽¹⁾*Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Bolshoi Karetny pereulok 19, Moscow, 127994, Russia*

⁽¹⁾*Department of Bioengineering and Bioinformatics, M.V. Lomonosov Moscow State University, Vorbyevy Gory 1-73, Moscow, 119992, Russia*

gbazykin@iitp.ru

⁽²⁾*Life Sciences Institute and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109-2216, USA*

kondrash@umich.edu

At any given moment, positive selection which favors currently rare, derived alleles affects only a small fraction of sites in the genome and, thus, is much rarer than negative selection, which favors common, ancestral alleles. A variety of methods for detecting positive selection, both past and ongoing, are in use, and a number of unambiguous cases of positive selection-driven allele replacements have been described at the sequence level. Still, neither of these methods is perfect. Here, we propose a method for detecting past positive selection through ongoing negative selection, based on comparison of the parameters of intraspecies variation at functionally important and selectively neutral sites where the same nucleotide substitution occurred recently. A

reduced presence of recently replaced alleles at functionally important sites indicates that negative selection currently acts against these alleles and that their replacements were driven by positive selection. Application of this method to *Drosophila melanogaster* lineage shows that the fraction of adaptive amino acid replacements in it remained ~0.5 for a long time. In *Homo sapiens* lineage, however, this fraction drops from ~0.5 in the course of origin of *Catharrhini* to 0.1 in the course of origin of *Hominidae*, to 0 in the course of hominid evolution, perhaps due to the impact of a reduction of effective population size, which can mask past positive selection by elevating the level of nonsynonymous variation.

Обнаружение положительного отбора на уровне ДНК

Detecting positive selection at the DNA level

Кондрашов А.С.

Кафедра Экологии и Эволюционной Биологии Мичиганского Университета
kondrash@umich.edu

A. S. Kondrashov

Department of Ecology and Evolutionary Biology, University of Michigan

Abstract

I review a variety of methods that can be used to detect past and ongoing positive selection at the level of DNA sequences. These methods include detection of an elevated rate of evolution at functionally important sites, the MacDonald-Kreitman test, detection of clumps of allele replacements, observations of footprints of recent allele replacements in the patterns of within-population variation, detection of young high-frequency alleles, and measurements of ongoing negative selection at sites of recent allele replacements.

1 Введение

Положительный, или Дарвиновский, естественный отбор увеличивает частоту аллельных замещений и уменьшает время, в течение которого происходит отдельное аллельное замещение. Большинство методов обнаружения положительного отбора основаны на одном из этих двух эффектов.

2 Обзор методов

2.1 Высокие скорости эволюции

Если в каком-либо сегменте генома функционально важные сайты эволюционируют быстрее неважных, это безусловно указывает на действие положительного отбора [14]. Однако такие ситуации обнаруживаются редко, даже когда сравнение большого количества геномов позволяет оценивать скорости эволюции индивидуальных сайтов [8].

2.2 Тест МакДоналда-Крейтмана

Сопоставление данных о межвидовых различиях и внутривидовой изменчивости позволяет обнаруживать положительный отбор даже когда скорость эволюции важных сайтов ниже скорости эволюции неважных сайтов. Требуется лишь, чтобы отношение скоростей эволюции важных и неважных сайтов было выше отношения уровней полиморфизма в них [9, 13]. К сожалению, при этом

теряется возможность оценивать свойства отдельных сайтов, поскольку уровень изменчивости широко варьирует между сайтами. Кроме того, присутствие в популяции редких слабовредных аллелей приводит к занижению роли положительного отбора, оцененной тестом МакДоналда-Крейтмана [3], а изменения численности популяции могут приводить как к занижению, так и к завышению оценки [2, 4, 6, 12].

2.3 Сблочность аллельных замещений

В некоторых случаях можно ожидать, что положительный отбор должен привести сразу к нескольким аллельным замещениям. Поэтому наблюдение двух или более аллельных замещений, произошедших в течение короткого эволюционного интервала свидетельствует о положительном отборе. Это прежде всего относится к несинонимическим заменам внутри одного кодона, поскольку новая оптимальная аминокислота не всегда достижима за счет лишь одного замещения [1].

2.4 Следы недавних аллельных замещений

Быстрое аллельное замещение, вызванное положительным отбором, оставляет характерный след в структуре внутривидовой генетической изменчивости: уровень изменчивости падает в пределах некоторого сегмента генома, с центром в сайте замещения. Размер этого сегмента тем больше, чем ниже скорость рекомбинации и чем слабее отбор. Такие следы возникают за счет эффекта попутного транспорта и существуют время, обратно пропорциональное произведению скорости мутирования на эффективную численность популяции. Обнаружение следов быстрых аллельных замещений может использоваться для изучения положительного отбора в недалеком прошлом [10].

2.5 Молодые частые аллели

Неоконченный процесс аллельного замещения можно обнаружить по присутствию в популяции молодых но частых аллелей, таких, что их высокая частота не может быть объяснена генетическим

дрейфом с момента их возникновения. Возраст аллеля можно оценить по уровню внутриаллельной генетической изменчивости в тесно сцепленном сегменте генома. Если популяция находится в процессе адаптации к новым условиям, неоконченные аллельные замещения могут наблюдаться одновременно во многих локусах [7, 11, 15].

2.6 Отбор против предковых аллелей

После того, как процес вызванного положительным отбором аллельного замещения закончен, отбор делается отрицательным, при условии неизменности адаптивного ландшафта. Действительно, менее приспособленные предковые аллели, сделавшись редкими, оказываются под действием отрицательного отбора. Тем не менее, такие аллели продолжают присутствовать в популяции за счет постоянного действия мутационного процесса. Отбор против редких предковых аллелей в сайтах, где недавно имели место аллельные замещения, свидетельствует о том, что эти замещения были вызваны положительным отбором. Предварительные данные, полученные с помощью этого подхода, показывают, что около половины аминокислотных замен в эволюции дрозофил и в ранней эволюции гоминид происходили под действием положительного отбора.

3 Обсуждение

Обнаружение положительного отбора - трудная задача, поскольку отрицательный отбор распространен гораздо шире и, в каждый данный момент, затрагивает значительно большую долю генома [5, 13]. Несмотря на это, положительный отбор - важная из форм отбора, так как именно он отвечает за адаптивную эволюцию. Накопление данных о межвидовой и внутривидовой генетической изменчивости будет способствовать изучению положительного отбора. Нам представляется особенно перспективным новый подход к обнаружению положительного отбора в прошлом через отрицательный отбор в настоящем, который в настоящее время подвергается всесторонней проверке.

References

- [1] G. A. Bazykin, F. A. Kondrashov, A. Y. Ogurtsov, S. Sunyaev, A. S. Kondrashov, Positive selection at sites of multiple amino acid replacements since rat-mouse divergence. *Nature* **429**, 558-62 (2004).
- [2] J. Charlesworth, A. Eyre-Walker, The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. *Proc Natl Acad Sci U S A* **104**, 16992-7 (2007).
- [3] J. Charlesworth, A. Eyre-Walker, The McDonald-Kreitman test and slightly deleterious mutations. *Mol Biol Evol* **25**, 1007-15 (2008).
- [4] A. Eyre-Walker, Changing effective population size and the McDonald-Kreitman test. *Genetics* **162**, 2017-24 (2002).
- [5] A. Eyre-Walker, The genomic rate of adaptive evolution. *Trends Ecol Evol* **21**, 569-75 (2006).
- [6] A. Eyre-Walker, P. D. Keightley, Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol* **26**, 2097-108 (2009).
- [7] S. R. Grossman *et al.*, A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* **327**, 883-6 (2010).
- [8] J. P. Huelsenbeck, S. Jain, S. W. Frost, S. L. Pond, A Dirichlet process model for detecting positive selection in protein-coding DNA sequences. *Proc Natl Acad Sci U S A* **103**, 6263-8 (2006).
- [9] J. H. McDonald, M. Kreitman, Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652-4 (1991).
- [10] R. Nielsen, I. Hellmann, M. Hubisz, C. Bustamante, A. G. Clark, Recent and ongoing selection in the human genome. *Nat Rev Genet* **8**, 857-68 (2007).
- [11] J. Novembre, A. Di Rienzo, Spatial patterns of variation due to natural selection in humans. *Nat Rev Genet* **10**, 745-55 (2009).
- [12] K. Popadin, L. V. Polishchuk, L. Mamirova, D. Knorre, K. Gunbin, Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc Natl Acad Sci U S A* **104**, 13390-5 (2007).
- [13] N. G. Smith, A. Eyre-Walker, Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022-4 (2002).
- [14] Z. Yang, J. P. Bielawski, Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* **15**, 496-503 (2000).
- [15] X. Yi *et al.* Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *Science* **329**, 75-78 (2010).

Исследование эволюционных ландшафтов приспособленности аминокислотных сайтов при помощи реверсий

Науменко С.А.(1,2), Кондрашов А.С.(3), Базыкин Г.А.(1, 4)

Институт проблема передачи информации РАН(1),
Российский химико-технологический университет им. Д.И. Менделеева(2),
Мичиганский университет(3),
Московский государственный университет им. М.В. Ломоносова(4)
naumenko@iitp.ru

Abstract

Functional evolution of protein sequences takes place on a fitness landscape, and each amino acid substitution can increase or decrease fitness or leave it invariant. The rate of reversals – amino acid substitutions that give rise to an ancestral amino acid – is relevant to the shape of the fitness landscape and its dynamics. We show that reversals are very frequent in evolution of vertebrates and insects. The fact that amino acid has been present in the evolutionary history of a given amino acid site significantly increases the probability of reversal into this amino acid. The rate of reversals decreases with increased phylogenetic distance. This decrease mainly is due to decrease in fitness of the ancestral amino acid rather than increase in fitness of derived one.

1 Фиксация реверсий на филогенетическом древе

Для привязки эволюционного события реверсии к филогенетическому дереву мы использовали принцип наибольшей экономии (рис. 1). Реверсию составляют две последовательные аминокислотные замены: прямая $A \rightarrow B$ и обратная $B \rightarrow A$. Если мы наблюдаем в выравнивании 6 видов, связанных филогенетическим отношением рис.1., столбец АВВВВАА, то можно утверждать, что реверсия является наиболее вероятным эволюционным сценарием, объясняющим наблюдаемое выравнивание. Набор из 6 видов является минимальным набором для фиксации реверсии: в случае 5 видов существует равновероятный с реверсией эволюционный сценарий, реверсией не являющийся.

Для изучения зависимости частот реверсий от величины филогенетического расстояния между двумя заменами в качестве аутгрупп выбираются все более далекие от базовой группы виды.

2 Изучаемые виды

Работа выполнялась для двух филогенетических деревьев: позвоночных и насекомых (рис.2,3).

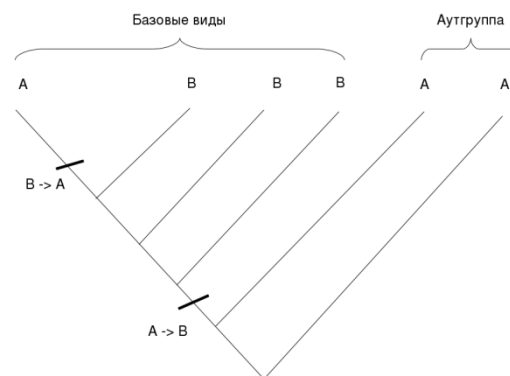


Рис. 1. Фиксация реверсии

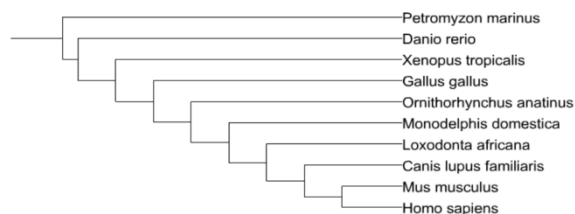


Рис.2 Дерево позвоночных

3 Метод расчета частот реверсий

В полногеномном множественном выравнивании кодирующих областей указанных видов производится поиск шаблонов, соответствующих реверсиям (рис. 4). Частота реверсий рассчитывается как $N(XYYYXX)/N(YYYYXX)$, где $N(Z)$ – количество событий типа Z. Далее частота реверсий нормируется на фоновую дивергенцию –

величину $N(XYYYY?)/N(YYYYY?)$. Полученное отношение

$$D = \frac{(N(XYYYYXX)/N(YYYYYXX))}{N(XYYYY?)/N(YYYYY?)}$$

характеризует превышение вероятности замены $Y>X$ в случае наличия в данном сайте предкового состояния X , по сравнению со случаем, когда о предковом состоянии ничего не известно.

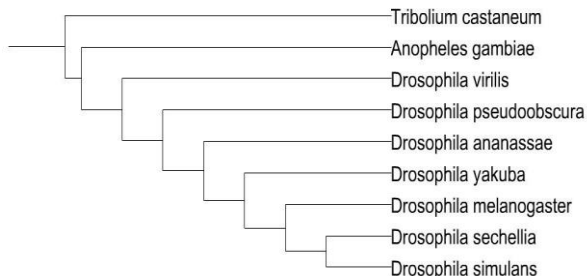


Рис.3. Дерево насекомых

4 Частоты реверсий

Таблица 1. Частоты реверсий у позвоночных

Outgroup	XXXXXX	YYYYXX	XXXXXX/ YYYYXX	XXXX??	YYYY??	XXXX?/? YYYY??	D
1 — MonDom4Ornana1	226	7795	0.029	5310	3716658	0.001	20.3
2 — OrnAna1galGal3	115	4575	0.025	5310	3716658	0.001	17.6
3 — GalGal3xenTro2	106	7315	0.014	5310	3716658	0.001	10.1
4 — XenTro2danRer5	74	8557	0.009	5310	3716658	0.001	6.1
5 — DanRer5petMar1	80	11941	0.007	5310	3716658	0.001	4.7

Таблица 2. Частоты реверсий у насекомых

Outgroup	XXXXXX	YYYYXX	XXXXXX/ YYYYXX	XXXX??	YYYY??	XXXX?/? YYYY??	D
1 — dpse — dvir	128	13735	0.0093	8798	7878972	0.0011	8.35
2 — dvir — anogam	85	12382	0.0069	8798	7878972	0.0011	6.15
3 — anogam- tricas	135	74594	0.0018	8798	7878972	0.0011	1.62

$D>1$ как для позвоночных, так и для насекомых, то есть наличие предкового состояния X в сайте существенно повышает вероятность возврата к данному состоянию. С увеличением филогенетического расстояния D монотонно убывает как у позвоночных, так и у насекомых.

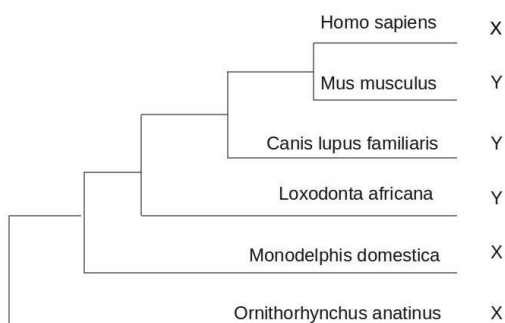


Рис.4 Схема расчета частот реверсий

5 Анализ ландшафтов приспособленности

Уменьшение относительной частоты реверсий с увеличением филогенетического расстояния может быть вызвано двумя причинами: либо увеличением приспособленности аминокислоты Y со временем (чем больше проходит времени, тем больше становится её приспособленность, тем меньше вероятность обратной замены), либо уменьшением приспособленности предкового варианта X (если приспособленность X падает, то и вероятность замены на эту аминокислоту также падает).

Чтобы определить, какой именно фактор в наибольшей степени влияет на частоту реверсий, мы сравнили частоты реверсий $X>Y>X$ с частотами “уходов в сторону”, т.е. событий $X>Y>Z$. Теоретически две разных картины изменяющегося ландшафта приспособленности дают противоположные картины отношений частоты событий $X>Y>X/ X>Y>Z$. Сценарий 1 дает монотонное увеличение этого отношения с увеличением филогенетического расстояния, а сценарий 2 – монотонный спад. Практически (таблицы 3,4) мы наблюдаем падение данного отношения как у позвоночных, так и у насекомых.

Следовательно, частота реверсий падает именно из-за уменьшения приспособленности предковой аминокислоты со временем, что является одной из особенностей изменения эволюционных ландшафтов.

Таблица 3.

Outgroup	XXXXXX	ZYYYYX	XXXXXX/ ZYYYYX
1 — MonDom4Ornana1	196	60	3.27
2 — OrnAna1galGal3	91	32	2.84
3 — GalGal3xenTro2	84	36	2.33
4 — XenTro2danRer5	49	35	1.40
5 — DanRer5petMar1	61	31	1.97

Таблица 4.

Outgroup	XXXXXX	ZYYYYX	XXXXXX/ ZYYYYX
1 — dp4 — droWil1	48	11	4.36
2 — droWil1 — anoGam1	22	11	2.00
3 — anoGam1 — apiMel3	39	18	2.17

Model for simulation broken gene flow in one-dimensional populations

Bukin Y.S.

Limnological Institute SB RAS, Russia, Irkutsk, Ulan-Batorskaya St. 3.
bukinys@lin.irk.ru

Abstract

In this work we propose model for population dynamic simulation of neutrally evolving "DNA sequences" included to each individual in the computer experiment. Using this program we simulated the process of genetic differentiation of groups in this organism taking into account isolation by distance and geographical barriers.

The study population structure of modern species using molecular genetic data is of great interest for scientist which connected the fact that the population represents a structural unit of microevolution, in which changes under the influence of gene drift and natural selection. The using genetic data, in particular molecular characteristics of organisms can most accurately define the extent of intraspecies genetic variability and geographic boundaries of populations.

Often naturally species have ribbon area of habitats. If the time spent by the organism to cross intersection of the ribbon area is less than or comparable with the duration of a generation, all the processes associated with the formation of genetic diversity will be determined by the flow of genetic information along the area. Such area of habitats, in terms of genetics, can be considered relatively one-dimensional. The boundaries of the populations on the one-dimensional area is areas where suspended or significantly slows down the flow of genes along the area. One example of a one-dimensional area are littoral zone of Lake Baikal. Flows of genetic information between different groups within a species at the littoral zone are directed along the coast of the lake.

For some species of the littoral zone of Lake Baikal and other ancient lakes on the basis of genetic data set population structure. The researcher, based on some facts, may suggest a hypothesis to explain the formation in this species of the similar population structure and genetic polymorphism. One way of testing hypotheses of this kind is mathematical-computer model that test proposed scenario under given population parameters.

In our work, we offer a program that implements a model of population dynamics on the one-dimensional area of habitat. The main characteristic of area in model is the maximum allowable density of organisms at each point. User before the numerical experiment divides the one-dimensional area into segments in each of which

specifies its maximum allowable density of organisms. Making it, the researcher models the one area that is typical for the studied species. The other parameters in the model are the mobility and rate of reproduction of organisms.

Each individual in the model is given its own "nucleotide sequence", that passed from mother to offspring (model of mitochondrial DNA). When the sequence is passed from parents to offspring with a given probability the mutation can occur. At the end of each simulation researcher can compare the molecular data model with data obtained in the study of natural populations. On the basis of comparison model and real data it is possible to do conclusions about the applicability of the hypothesis of formation of population structure of species and genetic polymorphism.

Figure 1 shows an example of the work of the program when the area of habitat of the species with x coordinates between 0 and 1 is divided into 4 segment separated by regions with unfavorable conditions for the existence of the species. The left side of the figure shows the dependence of the function $K(x)$ - the maximum possible density for the coordinates x on the area. The right side figure shows the scheme of resettlement organisms and the formation of population structure in the process of computer numerical experiment.

The algorithm of the program and the modeling method is based on an individually-based approach proposed in the papers [2,3] and then used by us in previous studies [1,4]. The algorithm consists of several parts. At the initialization step N organisms are distributed across the areal, DNA sequences are assigned. Single random sequence was generated and attributed to all individuals. Time of simulation was measured in life cycles (generations) consisting of birth, migration and death due to competition. The number of organisms born is defined as $N_r = rN$. All organisms form «reproductive pairs» so that probability of pairing between two organisms depends on the initial distance between them according to Gaussian law $C_r(x_m - x_f, d)$ where x_m and x_f are the coordinates of male and female. The pairs formed persist for only one cycle. Each pair produces random number of children so that their total number is N_r . Initial position of children is the same as position of

their mother. The mother gives to all children her nucleotide sequence, which however may mutate with probability P_m . Mutation means that at a random position the base changes equiprobably into one of the three other ones possible. At the second stage all organisms move. Each of them acquires new coordinate according to Gaussian distribution $C_d(x_i, d)$ where x_i is its initial position. The one-dimensional world was represented with a stretch of the length 1 (accordingly, all x_i were distributed between 0 and 1). The edges of this stretch were considered to be reflective. At the third stage the individuals were dying due to competition. Death probability for each individual was calculated according:

experiment and the dependence of a maximum allowable density of organisms $K(x)$ from the coordinates on the area.

$$P_i = \frac{\sum_{j=1}^{j=N+N_r} C_c(x_i - x_j)}{K(x)}$$

then the random value m between 0 and 1 was obtained. In cases when $P_i \geq m$, the organism died. The survivors started new cycle. It must be noted that parameter d on all three Gaussian functions C_c , C_r and C_d had the same value. Therefore the distance of partner choice, competition and migration directly depended on diffusion.

References

- [1] Bukin Ju.S., Pudovkina T.A., Sherbakov D.Ju., Sitnikova T.Ya., 2007. Genetic Flows in a Structured One-Dimensional Population: Simulation and Real Data on Baikalian Polychaetes M. Godlewskii. // In Silico Biology; v. 7(3), pp. 277-284.
- [2] Dieckmann U., Doebeli M., 1999. On the origin of species by sympatric speciation. // Nature vol. 400, pp. 354-357.
- [3] Doebeli M., Dieckmann U., 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. // Am. Nat. vol. 156, pp. 77-101.
- [4] Semovski S.V., Bukin Y.S., Sherbakov D.Y., 2003. Speciation and neutral molecular evolution in one-dimensional closed population. // International journal of modern physics, vol. 14, pp. 973-983.

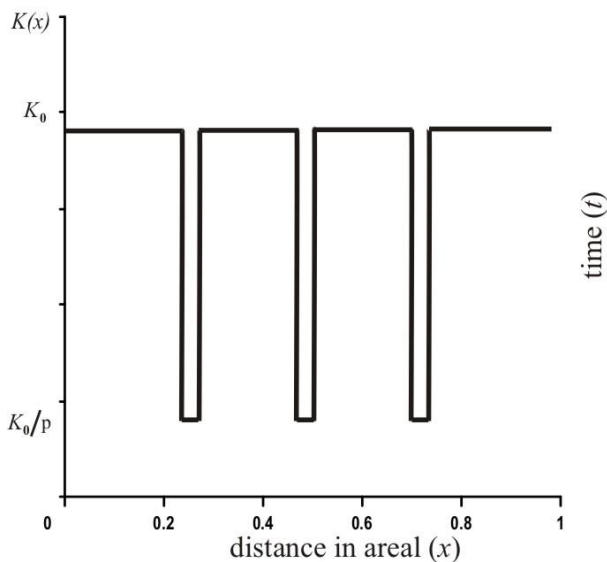


Fig. 1. The scheme of computer numerical experiment on modeling the formation of population structure of a one-dimensional heterogeneous area. The figure shows the diagram of the resettlement of organisms in the

Analysis of the triplet periodicity phase shifts in genes

Eugene V. Korotkov

Bioinformatics laboratory, Centre of Bioengineering, Russian Academy of Sciences, 117312, Prospect 60-tya Oktyabrya, 7/1

E.mail:genekorotkov@gmail.com

Abstract

The definition of a phase shift of triplet periodicity (TP) is introduced. The mathematical algorithm for detection of phase shift of TP of nucleotide sequences has been developed. Gene sequences from Kegg-46 data bank were analyzed with a purpose of searching genes with a phase shift of TP. The presence of a phase shift of triplet periodicity has been shown for 318329 genes (~10% from the number of genes in Kegg-46). We suppose that shifts of the TP phase may indicate the shifts of reading frame (RF) in genes.

1. Introduction

Mutations in gene sequences arise as substitutions, deletions and insertions of DNA bases, and as deletions, insertions and inversions of the whole DNA fragments [1-2]. Substitutions of DNA bases can induce substitutions of amino acids in protein, and a substitution of one base can change only one amino acid in amino acid sequence. These amino acid alterations very often have a strong influence on protein structure and protein ability to perform its biological function [3]. However, deletions or insertions of DNA bases can change long amino acid sequence due to the shift of a reading frame (RF) in a case when a size of deletion or insertion is not divisible by 3. The amino acid sequences downstream the point of RF shift are changed. In this sense deletions and insertions can be considered as more important evolutionary events than base substitutions. The influence of RF shifts on protein function has been studied relatively poorly due to the difficulty of detection of RF shifts. RF shifts is the cause of the TP shifts in a gene and RF shifts could be revealed with help of the TP shifts. In the present work we would like to find all genes from Kegg data base having a phase shift of TP.

2. Algorithm of searching for the TP phase shift

Let x shows the position of a base $s(x)$ in a sequence S and let x be chosen as $L_1 + 3n$, where $n=0,1,2,3,\dots(L-L_1)/3$, where L_1 is divisible by 3 and lies in the interval

from 60 to 600 bases. Let us consider the subsequence $S(x-L_1+1, x)$. For this subsequence we calculate the matrix of TP $\mathbf{M}_1(x-L_1+1, x)$ for RF of T_1 in a sequence S . The subsequences $S(x+1, x+L_1)$, $S(x+2, x+L_1+1)$ and $S(x+3, x+L_1+2)$ are also considered, and for these subsequences we calculate the TP matrices $\mathbf{M}_1(x+1, x+L_1)$, $\mathbf{M}_2(x+2, x+L_1+1)$ and $\mathbf{M}_3(x+3, x+L_1+2)$ for RF of T_1 , T_2 and T_3 , respectively, in a sequence S . If a shift of RF on 1 or 2 bases occurs just after the position x in a sequence S , then the matrix $\mathbf{M}_1(x-L_1+1, x)$ is more similar to $\mathbf{M}_2(x+2, x+L_1+1)$ or to $\mathbf{M}_3(x+3, x+L_1+2)$ matrix. If a shift of RF after position x is absent, then the matrix $\mathbf{M}_1(x-L_1+1, x)$ is more similar to the matrix $\mathbf{M}_1(x+1, x+L_1)$. Then for each of the four TP matrices another matrix was calculated which elements were the arguments of the normal distribution. Each element of such matrix was calculated using the following formula:

$$n(i, j) = \frac{m(i, j) - Lp(i, j)}{\sqrt{Lp(i, j)(1 - p(i, j))}} \quad (1)$$

$$p(i, j) = \frac{x(i)y(j)}{L^2} \quad (2)$$

where $m(i, j)$ is the element of a matrix $\mathbf{M}_1$, $\mathbf{M}_2$ or $\mathbf{M}_3$, $n(i, j)$ – normally distributed value. As a result, we obtain for each of the matrices $\mathbf{M}_1(x-L_1+1, x)$, $\mathbf{M}_1(x+1, x+L_1)$, $\mathbf{M}_2(x+2, x+L_1+1)$ and $\mathbf{M}_3(x+3, x+L_1+2)$ the matrices $\mathbf{V}_1$, $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$. The differences between the matrix $\mathbf{V}_1$ and each of the matrices $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ were calculated as:

$$D(1, k) = \sum_{i=1}^4 \sum_{j=1}^3 \left(\frac{v_1(i, j) - w_k(i, j)}{\sqrt{2}} \right)^2 \quad (3) \quad \text{for}$$

$k=1,2$ and 3 . In the capacity of a function U which allows to make the conclusion regarding the differences of two matrices of TP, we selected function $D(1, k)$. The value $D(1, k)$ is distributed as χ^2 with 6 degrees of freedom if the matrices calculated for random sequences are being compared. Then we calculated three probabilities of the fact that the random value distributed as χ^2 with 6 degrees of freedom will be greater than or equal to $D(1, k)$, $k=1,2,3$. Let us denote these probabilities P_{11} , P_{12} , P_{13} . If two matrices are similar, then the value of D equals zero and the value of P equals 1.0. In a case when two matrices being compared are different, the value of D will be greater than zero and the value of P will be less than 1.0. It is

suitable to use the values $F_1 = -\log_{10}(P_{11}/P_{12})$ and $F_2 = -\log_{10}(P_{11}/P_{13})$ for searching the shifts of TP phase. We selected two levels of F (F_0 and F_{00}) equal to 3.0 and 4.0, respectively, and calculated the number of genes which have one or more local maxima for F_1 or F_2 greater than F_0 and F_{00} , respectively. The numbers of false positives for selected F_0 and F_{00} are 6% and 18% respectively.

3. The search of genes with TP phase shift in Kegg data base.

We have analyzed 3318628 genes from Kegg data bank, release 46 (<http://www.genome.ad.jp/kegg/>). The total number of genes with F_1 or $F_2 > F_0$ was 318329, while with F_1 or $F_2 > F_{00}$ - 174879. Genes with a single shift of TP phase constituted up to 90% from the total number of genes with a shift of TP phase. Remaining 10% genes contained more than one case of TP phase shift. In the bank of random sequences we have found 58916 and 11063 shifts of TP for F_0 and F_{00} , respectively. This comparison shows that most part of TP phase shifts revealed in Kegg-46 data bank have nonrandom nature for the levels F_0 and F_{00} .

It is very interesting to see the distribution of biological functions of the revealed genes possessing the shifts of TP phase. The list of top 20 functions is shown in Tabl.1 (for F_0). Pseudogenes have the largest number of shifts of TP phase, and this number equals 9554. It is not so strange because pseudogenes do not have a functional significance and mutations in pseudogenes are neutral events for cell genome and they can accumulate rapidly [1-3]. The second position in Tabl.1 has genes in which the shift of RF was observed earlier in other investigations. We have revealed 844 of these genes. This fact is additional argument that shifts of RF in real sequences can be found by the way of searching the shifts of TP phase.

Table 1. The number of genes with TP phase shifts grouped by the identical description of biological function in Kegg data bank.

№	Number of genes	Definition
1	9554	Pseudogene
2	844	frameshift
3	348	mucin-associated surface protein (MASP), putative
4	335	translation initiation factor IF-2
5	290	trans-sialidase, putative
6	282	ABC transporter related

7	274	transposase
8	263	dispersed gene family protein 1 (DGF-1), putative
9	248	major facilitator superfamily MFS_1
10	224	PE-PGRS family protein
11	221	Protein kinase, putative
12	213	serine/threonine protein kinase
13	191	Protein kinase domain containing protein
14	185	transcriptional regulator, LysR family
15	183	PPE family protein

If we suppose that all or the most part of 10% of TP phase shift is caused by the RF shift, then this process contributes much more to the evolution of genes and proteins than it was considered earlier. This means that by no means all mutations arising in genes by RF shift are fatal for the proteins. From the functional point of view, the shifts of RF may be considered as events that can considerably change the function of the gene and the protein coded by this gene. Nevertheless, the genetic code should be adapted for these events [4,5], and new amino acid sequences should possess some new or keep an old function upon the shift of RF. Otherwise the existence of genes with RF shifts is difficult to be explained in general because these genes should lose their function and be eliminated from a genome.

This work was supported by the grant of the Ministry of the Science and Education of the Russian Federation, program "Scientific and Educational Personnel of the Innovation Russia", grant NK-97P-1.

4. References

- [1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, K. and P. Walter. Molecular Biology of the Cell. Garland Science, New York, 2002.
- [2] H. Maki, Annual Review of Genetics, 36, 279 (2002)..
- [3] J.D. Watson, T.A. Baker, S.P. Bell, A. Gann, M. Levine and R. Losick Molecular Biology of the Gene, 6th Edition. CSHL Press & Benjamin Cummings, San Francisco, 2007.
- [4] E.N. Trifonov, J. Mol. Biol., 194, 643-652, 1987.
- [5] T. Bollenbach, K. Vetsigian and R. Kishony, Genome Res., 17, 405-412, 2007.

Компьютерная система анализа режимов молекулярной эволюции генов и белков

Гунбин К.В.⁽¹⁾, Генаев М.А.⁽¹⁾, Афонников Д.А.^(1,2)

Институт Цитологии и Генетики Сибирского Отделения РАН, Новосибирск, Россия⁽¹⁾,
Новосибирский Государственный Университет, Новосибирск, Россия⁽²⁾
Email genkvg@bionet.nsc.ru

Abstract

A web-available computer system for analyzing the molecular evolution modes of genes and proteins was developed (<http://pixie.bionet.nsc.ru/samem/>).

Computations are based on the ratio of radical to conservative amino acid replacements, on the rate of amino acid changes in the course of protein evolution, and on the statistical relations between evolutionary change of all known amino acid properties and user defined phenotypical features of organisms.

1 Введение

В последнее время, адекватность использования отношения частот фиксации несинонимических замен к синонимическим (K_A/K_S) для определения режима эволюции генов поставлена под сомнение [1]. Альтернативными подходами являются анализ, отношения частот фиксации радикальных аминокислотных замен к консервативным (K_R/K_C) [2] и исследование скоростей изменения физико-химических свойств белков (V_{PC}) в их эволюции [3]. Недостатком альтернативных подходов является необходимость знать заранее “адаптивность изменения” определенного свойства аминокислот. Предложенный нами подход базируется на анализе K_R/K_C и V_{PC} , но имеет два отличия: 1) при анализе используются все известные свойства аминокислот; 2) проводится статистическое соотнесение изменения этих свойств с фенотипическими признаками организмов. На этой основе нами была создана компьютерная система анализа режимов молекулярной эволюции генов и белков.

2 Методы и алгоритмы

2.1 Конвейеры анализа

Система состоит из двух основных конвейеров, анализа эволюции генов и анализа эволюции белков, и двух дополнительных, собирающих выборки генов и белков и производящих их первичный анализ.

Основные конвейеры позволяют провести основные этапы обработки данных, комбинируя

различные методы множественного выравнивания, построения филогенетического дерева и реконструкции предковых последовательностей. Множественное выравнивание рассчитывается программами MAFFT 6.717 и KALIGN 2.04, филогения - программами FASTTREE 2.1.1 и PHYML 3.0. Предковые последовательности белков реконструируются ANCESCON, FASTML (серверная версия) и CODEML (из пакета PAML 4.4), генов – ANC-GENE, FASTML (серверная версия) и CODEML (из пакета PAML 4.4). Пользователь может рассчитать модель эволюции заданного семейства белков, используя MODELESTIMATOR 1.1, или воспользоваться обобщенными моделями. Для анализа используются данные о 531 свойстве аминокислот [4]. Общий модуль оценки взаимосвязи эволюции свойств аминокислот с эволюцией фенотипических признаков реализован на языке R.

Существенной особенностью конвейера анализа генов является возможность исследования данных разными методами оценки K_R/K_C . Пользователь может воспользоваться как оригинальным методом Жанга (программа HON-NEW) [2], так и методом Смита [5]. При анализе данных обоими методами, пользователь может задать число групп для разделения 20 аминокислот. Для каждого из 531 свойств аминокислоты группируются методом К-средних (используя R). При анализе данных методом Смита, для разбиения аминокислотных замен на консервативные и радикальные, пользователь может использовать матрицу BLOSUM. Все неотрицательные (≥ 0) значения в матрице BLOSUM трактуются нами как консервативные, все отрицательные – как радикальные. Для сравнения наблюдаемых значений K_R/K_C со значениями K_R/K_C , свойственными эволюции генов при $K_A/K_S=1$ с помощью пакета INDELible 1.03 проводится Марковское моделирование эволюции генов.

При анализе последовательностей белков существенной особенностью является доработанный нами метод анализа скоростей фиксации различных аминокислотных замен. Он основан на Марковском моделировании эволюции белков пакетом INDELible 1.03 и непараметрическом перестановочном тесте, предложенном нами в

работе [6]. INDELible 1.03 позволяет учесть детальные особенности эволюции белков (различие частот встречаемости аминокислот и частот их замен у паралогичных групп внутри одного семейства белков, вплоть до учета особенностей эволюции на отдельных ветвях дерева). Тест с перестановками [6] позволяет выявить для каждой ветви филогенетического дерева характерные типы аминокислотных замен, частота которых статистически значимо превышает частоту в смоделированных последовательностях.

2.2 Клиент-серверная платформа BioinfoWF

Система анализа режимов молекулярной эволюции генов и белков создана на основе разработанной нами клиент-серверной конвейерной платформы BioinfoWF. Серверная часть платформы реализована в виде приложения на языке Perl и отвечает за запуск и выполнение конвейера. На вход приложению подаётся схема конвейера и описание вычислительных модулей. Схема конвейера, форматы данных и взаимодействие вычислительных модулей описываются на языке XML.

Описание конвейера состоит из двух файлов: один файл описывает вычислительные модули, второй - задаёт топологию конвейера. Описание модулей состоит из следующих разделов: (1) входные файлы; (2) выходные файлы; (3) параметры (идентификатор, описание модуля и параметров, тип параметров, значения по умолчанию, внешний вид поля запроса значения для параметра на странице web-браузера); (4) правила генерации командной строки; (5) правила действия пользовательского интерфейса, позволяющие динамически изменять web-интерфейс модуля в процессе редактирования конвейера. Второй файл задаёт топологию конвейера. В нём описываются имена входных/выходных файлов для каждого модуля и значения для каждого параметра.

Приложение запускает конвейер, создавая файл с отчётом о статусе выполнения каждого узла. Серверная часть позволяет запускать ресурсоёмкие модули на вычислительном кластере.

Клиентская часть реализована в виде web-приложения. Пользователю предлагается работать с уже готовыми схемами. При этом доступны базовые опции управления конвейером: (1) установка входных файлов и параметров для каждого вычислительного модуля в конвейере; (2) старт и остановка на произвольном узле; (3) отслеживание статуса выполнения каждого вычислительного модуля в конвейере; (4) просмотр файлов для каждого этапа расчета в конвейере, (5) привязка форматов файлов к различным приложениям для их визуализации. Генерация страниц в формате HTML реализована на Perl с использованием библиотеки HTML::Template. Дополнительно используется JavaScript библиотека jQuery.2.1

3 Тестирование и доступность системы

Тестирование работоспособности системы производилось на выборке генов и белков лизоцимов приматов, представленной в работе [7]. Результаты согласуются с опубликованными ранее в работе [7] и дополнены многочисленными данными об особенностях эволюции отдельных свойств аминокислот.

Система анализа режимов молекулярной эволюции генов и белков доступна в интернет по адресу <http://pixie.bionet.nsc.ru/samem/>.

4 Выводы

Таким образом, предложенные нами улучшения позволяют полностью преодолеть недостаток альтернативных подходов анализа молекулярной эволюции белок кодирующих генов. Более того, возможность соотнесения эволюционного преобразования определенных физико-химических свойств генов и белков с адаптивными признаками организмов впервые позволяет напрямую связать молекулярную эволюцию с приспособительной эволюцией организмов.

5 Благодарности

Работа поддержана грантом РФФИ №09-04-01641-а; Интеграционными проектами СО РАН № 113, 119; Программами РАН № 22 (проект №8), 23 (проект №29), Госконтрактом Министерства образования РФ № П857.

Использованная литература

- [1] D.A. Drummond, C.O. Wilke. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*, V. 134(2), P. 341-352, 2008.
- [2] J. Zhang. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J. Mol. Evol.*, V. 50(1), P. 56-68, 2000.
- [3] T.Pupko, et al. Detecting excess radical replacements in phylogenetic trees. *Gene*, V. 319, P. 127-135, 2003.
- [4] S. Kawashima, et al. AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res.*, V. 36(Database issue), P. D202-D205, 2008.
- [5] N.G.C. Smith. Are radical and conservative substitution rates useful statistics in molecular evolution? *J Mol Evol.*, V. 57(4), P. 467-478, 2003.
- [6] K.V. Gunbin, D.A. Afonnikov, N.A. Kolchanov. Molecular evolution of the hyperthermophilic archaea of the *Pyrococcus* genus: analysis of adaptation to different environmental conditions. *BMC Genomics*, V. 10(1), P. 639, 2009.
- [7] W. Messier, C.B. Stewart. Episodic adaptive evolution of primate lysozymes. *Nature*, V. 385(6612), P. 151-154, 1997.

Следы мутаций в генах, приведшие к сдвигу рамки считывания

Френкель Ф.Е., Коротков Е.В.

Центр «Биоинженерия» РАН
felix.frenkel@gmail.com

Abstract

We present a new approach for detection of reading frame shifts in genes. Suggested method is based on phenomenon of triplet periodicity in coding regions of genes and its relative resistance to substitutions in DNA sequence. We attempted to extend 326 933 regions of continuous TP found in genes from KEGG databank by accounting possible insertions and deletions. We revealed totally 824 genes where such extension was possible and statistically significant. At a protein level 64 sequences have similarities for ancient RF only, 176 for active RF only, 3 – for both, 581 have no protein similarity at all. We suppose to have revealed lower bound for the number of genes in which a shift between RF and TP is possible.

Аннотация

Нами разработан метод, позволяющий находить сдвиги рамок считывания в высоко дивергировавших генах. Он основан на феномене наличия в кодирующих районах ДНК триплетной периодичности. Если на фоне триплетной периодичности в гене происходит сдвиг рамки считывания, то это можно заметить, так как происходит сдвиг между триплетной периодичностью и рамкой считывания. Поскольку триплетную периодичность последовательности ДНК достаточно трудно существенно изменить посредством сравнительно небольшого числа замен оснований, то такой сдвиг может сохраняться сравнительно долго. Присутствие такого сдвига между триплетной периодичностью нуклеотидной последовательностью и рамкой считывания может служить указанием на сдвиг рамки считывания в анализируемом гене.

1 Введение

Основными видами мутаций в последовательностях ДНК являются делеции, вставки и инверсии [2,5]. На аминокислотном уровне замены оснований ДНК могут приводить к заменам аминокислот, т.е. от одной замены нуклеотида меняется в аминокислотной последовательности только одна аминокислота. В этом смысле делеции и вставки оснований наряду с инверсиями участков генов являются критичными событиями, поскольку такое мутационное

изменение может привести к сдвигу рамки считывания и к изменению протяженного участка аминокислотной последовательности. Если после этого белок не потерял своей функции, то можно предполагать, что сдвиг рамки произошел в функционально малозначимом месте или же что сдвиг рамки считывания порождает похожие в функциональном смысле аминокислотные последовательности. Если же после сдвига рамки считывания белок изменил свою функцию [1], то интересно понять, как изменения аминокислотной последовательности определяют новую функцию белка.

2 Методы и результаты

Для каждого анализируемого гена из банка данных KEGG методом информационного разложения [6] были выделены районы с статистически значимой триплетной периодичностью без каких-либо делеций и вставок нуклеотидов. Всего было найдено 472 288 районов триплетной периодичности, из которых было отобрано 326 933 районов периодичности, не покрывающие весь ген. Затем была построена соответствующая матрица триплетной периодичности, которая была привязана к существующей в данном районе ДНК рамке считывания. Это означает, что первый столбец в матрице триплетной периодичности соответствовал первому основанию рамки считывания, существующей в районе ДНК с триплетной периодичностью. Далее методом модифицированного профильного анализа мы искали статистически значимое расширение участка триплетной периодичности в этом же гене за счет вставок и делеций нуклеотидов. Как результат, 824 гена содержали статистически значимый сдвиг между триплетной периодичностью и рамкой считывания, что указывает на наличие в гене сдвига рамки считывания.

Затем мы проверили, имеют ли гипотетические аминокислотные последовательности, полученные при использовании предполагаемой древней рамки считывания (определенной по триплетной периодичности), гомологию с последовательностями из банка данных Swiss-prot. Такую проверку мы делали для тех генов, у

Компьютерное моделирование эволюции сообществ одноклеточных гаплоидных организмов

Матушкин Ю.Г., Лашин С.А., Суслов В.В.

Институт цитологии и генетики СО РАН
mat@bionet.nsc.ru

Computer modeling of the evolution of unicellular haploid organisms [community](#).

Yu.G. Matushkin, S.A. Lashin, V.V.

Suslov

Institute of Cytology and Genetics SB RAS,
Novosibirsk, Russia
Abstract

The comparative modeling of populations' adaptivity and biodiversity dynamics is carried out. Populations consist of symbiotic unicellular haploid asexual organisms inhabiting in a biotope of limited volume in which the nonspecific substrate (NS) inflows externally. Organisms consume NS and it is necessary for every of them. Symbiosis is provided by intra-populational exchange of specific substrates. Using the original software package "Evolutionary Constructor" allowing varying both genetic and trophic structure of a model, we have shown that under sub-lethal conditions (NS deficiency) populations with compensatory metabolism (NS deficiency can be compensated by symbiont coevolution) have the better adaptivity (mutation in some single can save all symbiont populations) but gradually lose biodiversity. Whereas populations with non-compensatory metabolism keep biodiversity down to their extinction. The critical balance limit between population size and NS flow is obtained. Wherein it is outreached, the substitute load of normal allele by mutant one can not be hold. Advantages and disadvantages of two evolutionary strategies are discussed – strategy of biota formation on the base of several universal taxons and strategy of high biodiversity conservation.

Another model deals with origin of cells encapsulating exterior community metabolism into interior intercellular one. The origin of such cells could be the one of significant stages in evolution and possible eukaryotes origin. And, as paleontologists say, it could be result of series of horizontal genes transfer. We considered simple

ring-like symbiotic community of three members feeding each other and simulated its evolution during 15 thousand generations. During evolutionary process HGT could stochastically occur which lead to genes exchange and possible origin of a novel population. If a cell has both genes of synthesis and utilization for a particular substrate, its metabolic strategy is as follows. Cell tries to saturate itself with this substrate and only then secretes remains of these substrates, if they are. So cells could accumulate genes and enclose a part of metabolism in them, getting some benefit.

1 Программный комплекс «Эволюционный конструктор»

Мы разработали метод моделирования эволюции сообществ одноклеточных гаплоидных организмов (бактерий) и программный комплекс «Эволюционный Конструктор» (ЭК), который позволяет моделировать трофические отношения между сообществами, различные трофические стратегии организмов, влияние окружающей среды, изменения генетической структуры популяций и др. Отличительной особенностью ЭК является возможность моделирования эволюционных процессов, требующих *интенсивной* перестройки структуры модели в процессе расчёта, таких как *горизонтальный перенос генетического материала* и *видообразование*. Эти процессы могут изменять количество переменных в модели непосредственно в процессе расчета. Методика позволяет описывать полиаллельность одного или нескольких генов, число аллелей в процессе расчёта также может изменяться. Методика предполагает несколько уровней распараллеливания и является достаточно вычислительно эффективной. В процессе расчётов моделей количество сообществ доходило до 300, общая численность организмов до порядка 10^{20} особей, численность отдельного сообщества до 10^{18} , общее число различных генотипов до $10^6 - 10^7$ (в рамках одного сообщества до 4000 – 15000 различных генотипов при множественном полиморфизме 10 – 15 генов).

2. Моделирование возникновения популяций клеток, имеющих замкнутый биохимический цикл

С помощью ЭК мы смоделировали долговременную эволюцию симбиотической трофической системы. В начале расчёта система представляла собой замкнутое кольцо из трёх питающих друг друга популяций, где каждая популяция потребляла ровно один специфический (для этой популяции) субстрат и производила ровно один специфический (для этой популяции) продукт. Жизненно необходимым субстратом являлся только неспецифический субстрат, а утилизация специфических субстратов давала активирующий эффект. Между клетками популяций с определённой вероятностью (10^{-7} на поколение на клетку) мог происходить горизонтальный перенос между клетками разных популяций, в результате которого могли появляться новые популяции, потребляющие большее количество специфических субстратов, либо производящие большее количество специфических продуктов, либо и то и другое вместе. В долговременной перспективе (после 10000–15000 итераций-поколений) подавляющее преимущество получали популяции с «наиболее полным геномом», либо популяции с «почти полным геномом» (табл.). Такие популяции в долгосрочной перспективе вытесняли все остальные популяции из трофической системы (популяции-«аутсайдеры» либо погибали, либо их предельная численность колебалась в районе 10-100 клеток в среде)

Итоговое распределение численности популяций после долговременной эволюции (в течении 15 000 поколений) трофической системы. На схеме генотипа серый прямоугольник указывает на присутствие гена в данной позиции. Таблица показывает, что первое место по численности занимает популяция с «полным геномом» P_{30} , следом идут популяции с «почти полным геномом». Исходно существовавшие популяции (P_1 - P_3) – вымерли, также как еще 12 популяций.

3. Моделирование эволюции популяций с компенсаторным и некомпенсаторным типом питания

Проведено сравнительное моделирование динамики адаптивности и биоразнообразия популяций, обитающих в биотопе конечного объема, в который извне поступает неспецифический субстрат (НС), необходимый всем особям. Симбиотичность заключается в обмене между популяциями специфическими субстратами. Показано, что при сублетальном голодании по НС

популяции с компенсаторным метаболизмом обладают лучшей адаптируемостью, но постепенно теряют биоразнообразие. Популяции с некомпенсаторным метаболизмом, напротив, сохраняют биоразнообразие вплоть до вымирания. Для популяций с компенсаторным метаболизмом выявлено критическое соотношение численности популяции и притока НС, при котором субституционный груз замены нормального аллеля на мутантный выдержать невозможно. В регулярно изменяющихся условиях компенсаторные системы более стабильны с точки зрения сохранения и увеличения биомасса, тогда как некомпенсаторные системы – с точки зрения сохранения биоразнообразия. Таким образом, обе стратегии имеют собственные эволюционные преимущества/недостатки. Ни одна не имеет абсолютного преимущества.

Таблица.

Итоговое место популяции и (по численности)	Номер	Генотип (схема)	Итоговая численность (особей)
1	P_{30}		$4.86 \cdot 10^7$
2	P_{24}		$4.85 \cdot 10^7$
3	P_{28}		$2.94 \cdot 10^7$
4	P_{23}		$2.94 \cdot 10^7$
5	P_{27}		$2.94 \cdot 10^7$
---	---	---	---
18-33	P_1		0
18-33	P_2		0
18-33	P_3		0

References

- [6] S.A. Lashin, V.V. Suslov and Yu.G. Matushkin. Comparative Modeling Of Coevolution In Communities Of Unicellular Organisms: Adaptability And Biodiversity. *Journal of Bioinformatics and Computational Biology*, Vol. 8, No. 3 (2010) 627–643

Наиболее правдоподобное множественное выравнивание белков на основе ЕМ-алгоритма

Сулимова В.В.⁽¹⁾, Разин Н.И.⁽²⁾, Моттль В.В.⁽³⁾, Мучник И.Б.⁽⁴⁾, Куликовский К.⁽⁴⁾
 Тульский государственный университет⁽¹⁾, Московский физико-технический институт⁽²⁾,
 Вычислительный центр им. А.А. Дородницына РАН⁽³⁾, Университет Ратгерс, США⁽⁴⁾
 vsulimova@yandex.ru, nrmanutd@rambler.ru, vmottl@yandex.ru, muchnikilya@yahoo.com, kulikows@cs.rutgers.edu

Abstract

We propose a maximum-likelihood principle and EM procedure for finding the most probable common ancestor of a group of protein sequences in the form of its probabilistic profile jointly with the HMM-based posterior distribution over the set of possible proteins transformations. The resulting posterior distributions allows for efficient determination of the most probable multiple alignment. Results obtained on the BALiBASE 3.0 protein alignment benchmark indicate that the proposed method is generally more accurate than popular methods of multiple alignment such as CLUSTALW, DIALIGN and ProbAlign.

1 Введение

Задача множественного выравнивания группы аминокислотных последовательностей является фундаментальной проблемой современной биоинформатики. Существует множество методов множественного выравнивания, однако только немногие из них, такие как многомерное динамическое программирование [1], основаны на строгой математической формулировке задачи, имеющей неэвристическое решение. Методы, принимающие во внимание информацию об эволюции белков [2], решают NP-полные задачи и могут быть использованы только для выравнивания очень небольшого числа последовательностей [3]. Их аппроксимации и другие быстрые эвристические алгоритмы, включая широкое семейство алгоритмов прогрессивного выравнивания [4, 5] и алгоритмов, основанных на согласовании парных выравниваний [6, 7], как правило, менее адекватны с биологической точки зрения. Методы, использующие для согласования выравниваний парные или профильные скрытые марковские модели [Ошибка! Залка не определена., 8, 9], требуют, к тому же, предварительной оценки параметров модели.

В данной работе предложена принципиально новая постановка задачи множественного выравнивания, основанная на простейшей вероятностной модели происхождения совокупности белков и обеспечивающая на базе BALiBASE 3.0 в среднем лучшие результаты по сравнению с такими популярными подходами как CLUSTALW, DIALIGN и ProbAlign.

2 Модель происхождения группы белков

Пусть $A = \{\alpha^{(1)}, \dots, \alpha^{(20)}\}$ - множество аминокислот.

Гипотеза 1. Каждая из M последовательностей анализируемой совокупности $\Omega^* = \{\omega_j = (\omega_{ji} \in A,$

$t=1, \dots, N_j), N_j \geq n, j=1, \dots, M\}$ получена независимо от других из своей последовательности-прародителя $\Theta_j = (\theta_{ji} \in A, i=1 \dots n)$ известной длины n .

Гипотеза 2. Элементы прародителей θ_{ji} выбраны из алфавита аминокислот в соответствии с неизвестным наблюдателю вероятностным профилем $\bar{\beta} = (\beta_i = (\beta_i^1, \dots, \beta_i^{20}), i=1, \dots, n)$, т.е.

последовательностью распределений на множестве аминокислот $0 \leq \beta_i^l \leq 1, \beta_i^1 + \dots + \beta_i^{20} = 1$.

Гипотеза 3. Каждый белок $\omega = \omega_j$ длины $N_j \geq n$ получен из прародителя $\Theta = \Theta_j$ длины n в два этапа в соответствии со следующей моделью.

На первом этапе генерируется структура преобразования $\mathbf{v} = (1 \leq v_1 < \dots < v_n)$, определяющая позиции элементов формируемой последовательности $\omega = (\dots, \bar{\omega}_{v_1}, \dots, \bar{\omega}_{v_i}, \dots, \bar{\omega}_{v_n}, \dots)$, в которые будут преобразованы элементы прародителя $\Theta = (\theta_1, \dots, \theta_n)$.

Структура $\mathbf{v}$ выбирается из множества допустимых структур V_{Nn} в соответствии с априорным распределением вероятностей $q_{Nn}(\mathbf{v})$, учитывающим только расстояния между элементами $(v_{i+1} - v_i)$.

На втором этапе выбираются элементы формируемой последовательности в соответствии с некоторым распределением вероятностей $\eta(\omega_{v_i} | \theta_i)$, основанном на случайном преобразовании аминокислот $\psi(\alpha^{(j)} | \alpha^{(k)}), j, k = 1, \dots, 20$, принятом в модели РАМ М.Дэйхофф [10]. С учетом случайности выбора элементов прародителя оно принимает вид:

$$\eta(\omega_{v_i} | \theta_i) = \sum_{k=1}^{20} \beta_i^k \psi(\omega_{v_i} | \alpha^{(k)}).$$

Тогда функция правдоподобия происхождения j -го белка может быть записана в виде:

$$f(\omega_j | \bar{\beta}) = \sum_{\mathbf{v} \in V_{Nj,n}} q_{Nj,n}(\mathbf{v}) \prod_{i=1}^n \eta(\omega_{v_i} | \theta_i).$$

3 Оценка вероятностного профиля прародителя

Оценивать последовательность распределений вероятностей $\bar{\beta} = (\beta_i, i=1, \dots, n)$ одновременно по всей анализируемой совокупности белков $\Omega^* = \{\omega_j, j=1, \dots, M\}$ будем из условия максимизации

функции правдоподобия $F(\Omega^* | \bar{\beta})$:

$$\hat{\beta} = \arg \max_{\bar{\beta}} \ln F(\Omega^* | \bar{\beta}) = \arg \max_{\bar{\beta}} \ln \prod_{j=1}^M f(\omega_j | \bar{\beta}). \quad (1)$$

В основу решения данной задачи положена ЕМ-процедура, которая в данном случае представляет собой следующий итерационный процесс.

Пусть $\bar{\beta}_s = (\beta_{1,s}, \dots, \beta_{n,s})$ - приближение, полученное на шаге s . Тогда каждая компонента следующего приближения может быть найдена как решение оптимизационной задачи

$$\begin{cases} \beta_{i,s+1} = \arg \max_{\beta_i} \sum_{l=1}^{20} h_l^i \ln \sum_{k=1}^{20} \beta_i^k \psi(\alpha^{(l)} | \alpha^{(k)}), \\ \sum_{k=1}^n \beta_i^k = 1, \beta_i^k \geq 0, k = 1, \dots, 20, \end{cases} \quad (2)$$

где $h_l^i = \sum_{j=1}^M \sum_{t=1}^{N_j} I[\omega_{jt} = \alpha^{(l)}] p_{it}(\bar{\beta}_s, \omega_j)$, индикаторная функция $I[\omega_{jt} = \alpha^{(l)}] = 1$, если условие в скобках выполняется и $=0$, в противном случае, а $p_{it}(\bar{\beta}_s, \omega_j) = P(v_{ij} = t | p_{it}(\bar{\beta}_s, \omega_j))$ - апостериорная вероятность того, что i -й элемент прародителя преобразуется в t -й элемент j -й последовательности.

Теорема. Выбор $\bar{\beta}_{s+1} = (\beta_{1,s+1}, \dots, \beta_{n,s+1})$ в соответствии с (2) обеспечивает выполнение неравенства $F(\Omega^* | \bar{\beta}_{s+1}) \geq F(\Omega^* | \bar{\beta}_s)$ на каждом шаге s , причем $F(\Omega^* | \bar{\beta}_{s+1}) = F(\Omega^* | \bar{\beta}_s)$ только если достигнут максимум функции правдоподобия $\nabla_{\bar{\beta}} F(\Omega^* | \bar{\beta}_s) = 0$.

Задача (2) может быть легко решена методом проекции градиента, а вычисление $p_{it}(\bar{\beta}_s, \omega_j)$ является стандартной задачей теории марковских моделей, поскольку структура преобразования v_j для j -й последовательности представляет собой скрытый марковский процесс (Hidden Markov Model) ее формирования.

4 Наиболее правдоподобное множественное выравнивание

На основе апостериорных распределений $p_{it}(\bar{\beta}_s, \omega_j)$, найденных на последней итерации, можно определить наиболее вероятную структуру $v_j = (v_{j1}, \dots, v_{jn})$ преобразования прародителя в каждый из белков $\omega_j, j = 1, \dots, M$, решая M типовых задач динамического программирования:

$$v_j = \arg \max_{v_{j1}, \dots, v_{jn}} \prod_{i=1}^n p_{iv_i}(\bar{\beta}_s, \omega_j), v_i \geq v_{i-1}, i = 2, \dots, n,$$

Совокупность этих структур определяет множественное выравнивание всех анализируемых последовательностей.

5 Экспериментальное исследование

Для тестирования предложенного подхода использовались наборы относительно коротких последовательностей (с длинами от 49 до 527) из базы данных BALiBASE 3.0 [11], содержащей эталонные выравнивания, выполненные вручную.

Качество выравнивания оценивалось путем вычисления доли пар аминокислот верно размещенных в одном столбце выравнивания (SP –

Sum of Pairs) и доли верно воспроизведенных столбцов (TC – True Columns). Результаты сравнения с другими подходами приведены в таб. 1.

Таблица 1 - Средние значения TC/SP по пяти семействам коротких белков из каждого набора базы BALiBASE

наб ор	семейства и идентичность белков в них (%)	clustalw	dialign	probalign	предлаг аемый подход
RV11	laab, laboA, lbbt3, lcsy, ldox (<20%)	0.26/ 0.44	0.32/ 0.35	0.44/ 0.62	0.48 / 0.62
RV12	laxo, lfj1A, lhfh, lhpi, lkm (20-40%)	0.62/ 0.75	0.49/ 0.64	0.65/ 0.77	0.83 / 0.90
RV20	lidy, lpamA, lpgtA, ltvxA, lubi (>40% + одна далекая посл-ть, <20%)	0.18/ 0.64	0.09/ 0.23	0.22 / 0.55	0.21/ 0.61
среднее по всем выравниваниям		0.35/ 0.61	0.30/ 0.40	0.44/ 0.65	0.51 / 0.71

Из табл.1 видно, что предлагаемый подход по средним показателям превосходит другие популярные алгоритмы. В дальнейшем планируется провести более масштабное исследование.

References

1. Sankoff D, Cedergren R.J. Simultaneous comparison of three or more sequences related by a tree. In: Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison, Sankoff D, Kruskal J.B., editors., 1989, pp. 253-263. Addison-Wesley, Reading, Massachusetts.
2. Altschul S.F., Lipman D.J. *Trees, stars, and multiple biological sequence alignment*. SIAM J. Appl. Math., 49, 1989, pp. 197-209.
3. Todd Wareham H. *A simplified proof of the NP- and MAX SNP-hardness of multiple sequence tree alignments*. J. Comput. Biol., 2 4, 1995, pp. 509-514.
4. Notredame C., Higgins D.G., Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol. Biol., 302, 2000, pp. 205-217.
5. Subramanian A.R., Kaufmann M., Morgenstern B. DIALIGN-TX: Greedy and progressive approaches for segment-based multiple sequence alignment. Algorithms for Molecular Biology, 2008, 3:6.
6. Thompson, J.D., Higgins, D.G., and Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res 22: 4673-4680, 1994.
7. Roshan U., Libesay D.R. Probalign: Multiple sequence alignment using partition function posterior probabilities. Oxford University Press, 2005.
8. Durbin R., Eddy S, Krogh A., Mitchison G. Biological sequence analysis: Probabilistic models of proteins and nucleic acids. Cambridge Univ Press, 1998.
9. Pei J, Grishin N.V. PROMALS: Towards accurate multiple sequence alignments of distantly related proteins. Bioinformatics, 23, 2007, pp. 802-808.
10. Dayhoff M.O., Schwartz R.M., Orcutt B.C: A model of evolutionary change in proteins, Atlas of Protein Sequences and Structures: 5, 345-352, 1978.
11. Thompson J.D., Koehl P., Ripp R., Poch O. BALiBASE 3.0: latest developments of the multiple sequence alignment benchmark, Proteins, 2005, 61, pp:127-36.

Единая вероятностная модель эволюции аминокислот для подстановочных матриц семейств PAM и BLOSUM

Сулимова В.В.⁽¹⁾, Моттль В.В.⁽²⁾, Мучник И.Б.⁽³⁾, Куликовский К.⁽³⁾

Тулеский государственный университет⁽¹⁾, Вычислительный центр РАН, Москва⁽²⁾,
Университет Ратгерс, Нью-Джерси, США⁽³⁾

vsulimova@yandex.ru, vmottl@yandex.ru, muchnikilya@yahoo.com, kulikows@cs.rutgers.edu

Abstract

In this paper we prove that the statistical approach to computing substitution matrices of the BLOSUM family can be explained in terms of the PAM evolution model. This means that both of these approaches are based, actually, on the same evolution model, and the main difference between them lies in the different initial data for estimating unknown model parameters. We also show that all PAM and BLOSUM substitution matrices are, in terms of computer science, kernel functions by their mathematical structure, and lose their positive semi-definiteness only because of the choice of final representation.

1 Введение

Одной из фундаментальных проблем современной биоинформатики, возникающей при решении многих практических задач молекулярной биологии, таких как задача функциональной аннотации, распознавания пространственной структуры, прогнозирования взаимодействия белков и т.д., является проблема выбора количественной меры сходства или несходства аминокислотных последовательностей, образующих полимерные молекулы белков. Однако измерение сходства последовательностей неизбежно должно базироваться на сравнении самих аминокислот.

Традиционно, сходство 20-ти известных аминокислот выражается подстановочной матрицей 20×20 . Может быть предложено множество различных способов сравнения аминокислот. Однако для того, чтобы мера сходства аминокислот была полезна для решения прикладных задач анализа белков, крайне желательно, чтобы она, во-первых, аккумулировала в себе информацию о свойствах анализируемых последовательностей, во-вторых, была обоснована с точки зрения эволюции белков и, в-третьих, обладала свойствами потенциальной функции, т.е. чтобы матрица ее значений для всех аминокислот не имела отрицательных собственных чисел и могла быть интерпретирована как скалярное произведение сравниваемых аминокислот в некотором, вообще говоря, гипотетическом линейном пространстве. Последнее свойство является важным этапом для построения осмысленных потенциальных

функций на множестве последовательностей, чему посвящены многочисленные исследования [1,2,3,4,5]. Интерес к построению потенциальных функций для биологических последовательностей объясняется тем, что они, в характерной для анализа белков беспризнаковой ситуации, когда трудно сформировать информативное признаковое пространство [6], дают возможность использования удобных и эффективных методов анализа данных, разработанных для линейных пространств, таких как, например, метод опорных векторов В.Н. Вапника [7].

2 Подстановочные матрицы PAM

Одной из общепринятых мер сходства аминокислотных последовательностей является семейство подстановочных матриц PAM, предложенных Маргарет Дэйхофф [8] и основанных на одноименной модели эволюции PAM (Point Accepted Mutations). Основным математическим понятием этой модели является понятие марковской цепи эволюции аминокислоты в отдельно взятой позиции цепи при некоторой интерпретации понятия «шага эволюции», которое здесь не рассматривается. При этом предполагается, что данная марковская цепь является 1) эргодической, т.е. эволюция началась очень давно, случайный процесс мутаций успел установиться и не зависит от неизвестных начальных вероятностей состояний, и 2) обратимой, т.е. эволюция не имеет направления.

Исходными данными для оценивания параметров модели PAM являются фрагменты реализации марковской цепи эволюции аминокислот, полученные из филогенетических деревьев над близкими аминокислотными последовательностями (не менее 95% похожести).

Естественная мера сходства аминокислот, построенная на основе данной модели, имеет смысл правдоподобия гипотезы о совместном независимом происхождении сравниваемых аминокислот от одной и той же неизвестной аминокислоты в результате двух случайных преобразований. Значения данной меры сходства логарифмируются и округляются до целых. Результирующие подстановочные матрицы свойствами потенциальной функции не обладают.

3 Подстановочные матрицы BLOSUM

Подстановочные матрицы семейства BLOSUM [9], предложенные Стивенем и Джорджией Хениковыми, основаны исключительно на статистическом подходе, исходными данными для которого являются блоки консервативных регионов аминокислотных последовательностей, полученные в результате множественного выравнивания семейств существенно более далеких по сравнению с PAM белков. Это обстоятельство делает данную меру сходства более адекватной для сравнения далеких белков, однако отсутствие эволюционного обоснования является ее существенным недостатком [10]. Кроме того, подстановочные матрицы BLOSUM, как и матрицы PAM, имеют отрицательные собственные числа.

В литературе известны многочисленные попытки скорректировать подстановочные матрицы с целью сделать их неотрицательно определенными. Однако при этом, либо теряется биологический смысл меры сходства [11,12], либо нет гарантии отсутствия отрицательных собственных чисел [13].

4 Результаты

Одним из важных результатов, полученных нами в рамках данной работы, является доказательство того, что семейство подстановочных матриц BLOSUM может быть выражено в терминах той же самой модели эволюции, что и PAM. При определенной интерпретации порождения исходных данных для BLOSUM, статистики, вычисляемые Хениковыми, являются несмещенными оценками параметров модели Дэйхофф. Различие между матрицами PAM и BLOSUM объясняется исключительно разными исходными данными, по которым производится оценивание параметров одной и той же модели.

Кроме того, мы доказали, что модель эволюции М. Дэйхофф содержит в себе все необходимое для построения потенциальных функций. Более того, подстановочные матрицы семейств PAM и BLOSUM являются потенциальными функциями по своей природе и теряют соответствующие свойства исключительно вследствие неподходящего окончательного представления.

Также немаловажным является то, что любая новая подстановочная матрица, построенная на основе данной модели эволюции по технике Дэйхофф или Хениковых, также будет обладать свойствами потенциальной функции. Этот факт имеет особенно большое значение в связи с ростом интереса к построению специфических подстановочных матриц, предназначенных для решения конкретных задач [14]. Такие матрицы, в отличие от классических, строятся только по белкам определенной группы организмов или функций, и, как правило, позволяют получить лучшие результаты при их дальнейшем анализе.

Все полученные результаты с доказательствами приведены в работе [15].

References

1. Vert JP, Saigo H, Akutsu T. Local alignment kernels for biological sequences. In B. Scholkopf, K. Tsuda, and J. P. Vert, editors, *Kernel Methods in Computational Biology*. MIT Press, 2004.
2. Motil V.V., Seredin O.S., Sulimova V.V. Mathematically correct methods of similarity measurement on sets of signals and symbol sequences of different length. *Pattern Recognition and Image Analysis*, Vol.15, No. 1, 2005, pp. 87-89.
3. Caragea C., Sinapov J., Silvescu A., Dobbs D., Honavar V. Glycosylation site prediction using ensembles of Support Vector Machine classifiers. *BMC Bioinformatics*, November 2007, 8:438 doi:10.1186/1471-2105-8-438.
4. Vincent M., Passerini A., Labbe M., Frasconi P. A simplified approach to disulfide connectivity prediction from protein sequences. *BMC Bioinformatics*, January 2008, 9:20 doi:10.1186/1471-2105-9-20.
5. *Kernel methods in computational biology*. Edited by B. Scholkopf, K. Tsuda and J.-p. Vert. A Bradford book, The MIT press, Cambridge, Massachusetts, London, England, 2004, 397p.
6. Pekalska E., Duin R. P. W. *The Dissimilarity Representation for Pattern Recognition. Foundations and Applications*. Singapore: World Scientific, 2005.
7. Vapnik V. *Statistical Learning Theory*. New York: John-Wiley&Sons, Inc., 1998, 732 p.
8. Dayhoff M.O., Schwartz R.M., Orcutt B.C. A model for evolutionary change in proteins. *Atlas for Protein Sequence and Structure*, 1978, Vol. 5, 345-352.
9. Henikoff S., Henikoff J. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.*, 1992, 10915-10919.
10. Altschul S.F. The statistics of Sequence Similarity Scores, <http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html>.
11. Vanschoenwinkel B., Manderic B. Substitution matrix based kernel functions for protein secondary structure prediction. *Proceedings of the 2004 International Conference on Machine Learning and Applications*. December 16-18, 2004, pp. 388 – 396.
12. Wu F., Osion B., Dobbs D., Honavar V. Comparing kernels for predicting protein binding sites from amino acid sequence. *Neural Networks*, 2006, IJCNN'06, pp. 1612-1616.
13. Vert JP, Saigo H, Akutsu T. Local alignment kernels for biological sequences. In B. Scholkopf, K. Tsuda, and J. P. Vert, editors, *Kernel Methods in Computational Biology*. MIT Press, 2004.
14. Ng P.C., Henikoff J.G., Henikoff S. PHAT: a transmembrane-specific substitution matrix // *Bioinformatics*, 2000, 16, 760-766.
15. Sulimova V.V., Motil V.V., Kulikowski C., Muchnik I. Probabilistic evolutionary model for substitution matrices of PAM and BLOSUM families. In: *DIMACS Technical Report 2008-16*, Rutgers University, 17 p., <ftp://dimacs.rutgers.edu/pub/dimacs>.

Биоинформатика: платформа от геномов к лекарству *in silico*

Поройков В.В., Веселовский А.В., Иванов А.С., Скворцов В.С., Лисица А.В.,

Соболев Б.Н., Филимонов Д.А., Арчаков А.И.

Учреждение Российской академии медицинских наук Научно-исследовательский институт
биомедицинской химии им. В.Н. Ореховича РАМН, 119121, Москва, Погодинская ул., 10
vladimir.poroikov@ibmc.msk.ru

Abstract

Computer-aided drug discovery approaches are widely used now in pharmaceutical research and development. Examining the amino acid sequences of proteins relatively conservative antigenic determinants are revealed and used for further design of recombinant vaccines. The most prospective macromolecular targets (separate or in combinations) associated with certain pathologies can be identified by comparative analysis of genomics, postgenomics and other biomedical data. Spatial structure of the target protein, either obtained from the Protein Data Bank (PDB) or computed with molecular modelling techniques, can be used for target-based design of potential ligands or for their virtual screening among the commercially available samples. Drug-candidates may be obtained by ligand-based (Q)SAR approach applied for optimization of pharmacodynamics' and pharmacokinetics' characteristics of lead compounds. These computer-assisted methods are widely used in applied projects directed to detection of the targets and their ligands for the treatment of both infectious (HIV/AIDS, tuberculosis, etc.) and non-infectious diseases (hypertension, inflammation, etc.). The whole computational platform "From genomes to drugs *in silico*" has been developed by the Institute of Biomedical Chemistry of Rus. Acad. Med. Sci. CIS Hub for Collaborative Drug Discovery (CIS-CDD) is under development, which provides the general framework for allocation of information about the experience and potential in pharmaceutical R & D of Russian and CIS research teams, promotion of fundraising processes for organization and execution of collaborative projects; integration with the existing resources of chemical, biological and pharmaceutical information; and cooperation with the other collaborative drug discovery consortiums, etc.

1 Введение

1.1 Необходимость применения методов компьютерного конструирования лекарств в фармацевтике

Фармацевтические исследования и разработки (R&D) связаны с существенными временными и финансовыми издержками. Традиционный процесс создания одного нового лекарственного препарата в промышленно-развитых странах в среднем длится около 12 лет, а финансовые затраты составляют около 1 млрд долларов США. При этом, учитывая высокий риск получения отрицательного результата, необходимо синтезировать и протестировать около 10 тысяч структур различных химических соединений.

С целью повышения экономической эффективности создания новых лекарственных препаратов в фармацевтических R&D широко применяются методы компьютерного конструирования лекарств. В докладе будет представлено современное состояние создания и применения методов био- и хемоинформатики в фармацевтических R&D, дан обзор разработок ИБМХ РАМН и обсуждены перспективы дальнейшего развития в данной области.

2 Компьютерная платформа «от геномов к лекарству *in silico*»

2.1 Идентификация функционально-значимых участков в белках и поиск наиболее перспективных фармакологических мишеней

Путем сравнительного анализа геномных и постгеномных данных, совместно с сопутствующей биомедицинской информацией идентифицируют макромолекулы, воздействие на которые наиболее благоприятно с точки зрения терапии изучаемой патологии. Анализируя аминокислотные последовательности белков, выявляют функционально важные сайты, которые могут рассматриваться как потенциальные мишени для создания новых диагностикумов, вакцин и лекарств. Оценка консервативных и вариабельных участков в

структуре гомологичных белков из различных инфекционных агентов является основой для идентификации вероятных антигенных детерминант, а также предсказания возможных проблем с лекарственной резистентностью.

2.2 Создание потенциальных лигандов на основе структуры макромолекулы-мишени

Трехмерную (3D) структуру макромолекулы-мишени обычно определяют с помощью рентгеноструктурного анализа. Для многих белков такая информация может быть получена из банка данных белковых структур PDB. Если для изучаемого белка необходима информация в PDB отсутствует, но рассматриваемый белок имеет близкие гомологи, 3D модель может быть построена с помощью молекулярного моделирования. Механизмы белок-лигандного взаимодействия изучают с помощью молекулярного докинга. Информация об особенностях сайтов связывания и об известных лигандах используется для виртуального скрининга базовых структур новых лекарств либо для их дизайна *de novo*. Молекулярное моделирование позволяет получить оценки энергии связывания (scoring function), величина которой используется в качестве критерия для отбора наиболее перспективных базовых структур новых лекарств. Методы компьютерного конструирования лекарств, основанные на структуре макромолекулы-мишени» (target-based drug design) широко используются в настоящее время для создания лигандов, обладающих аффинностью к конкретным макромолекулам-мишеням [1, 2].

2.3 Поиск лигандов и оптимизация их свойств на основе методов (Q)SAR/(Q)SPR

В случае, если 3D структура макромолекулы-мишени не известна и не может быть в настоящее время определена с достаточной точностью методами молекулярного моделирования, применяют «дизайн, основанный на структуре лиганда» (ligand-based drug design) [3]. Используя информацию об известных лигандах (блокаторах или активаторах) в качестве обучающей выборки, анализируют взаимосвязи «структура-активность» (SAR), либо – количественные взаимосвязи «структура-активность» (QSAR). Эти зависимости в дальнейшем используют для поиска и оптимизации новых более эффективных и безопасных лигандов. Аналогичным образом оценивают различные фармакодинамические и фармакокинетические характеристики исследуемых веществ, а также их токсичность.

2.4 Практические приложения и необходимость интеграции компьютерного конструирования лекарств в единой платформе

Разработанные в ИБМХ РАМН компьютерные методы широко используются для выявления

перспективных фармакологических мишеней и их лигандов, для терапии как инфекционных (ВИЧ/СПИД, туберкулез и др.), так и неинфекционных заболеваний (артериальная гипертензия, воспалительные процессы и др.) [4].

Современные требования к разработке новых более эффективных и безопасных лекарственных препаратов обуславливают необходимость комплексного применения методов био- и хемоинформатики. В ИБМХ РАМН проводится разработка интегральной платформы, обеспечивающей всю цепочку исследований «от геномов к лекарству *in silico*».

Для организации кооперации ученых и специалистов из России и стран СНГ в разработке новых лекарств создается информационно-телекоммуникационная система CIS-CDD (CIS Hub for Collaborative Drug Discovery). CIS-CDD обеспечит обмен информацией об опыте и ресурсах различных исследовательских групп, единый «инструмент» для организации и управления совместными проектами, интеграцию доступа к существующим источникам химической, биологической и фармацевтической информации, взаимодействие с другими консорциумами в области поиска и создания новых лекарств.

3 Заключение

Представленные в докладе методы компьютерного конструирования лекарств существенно повышают экономическую эффективность и возможности поиска и создания более эффективных и безопасных лекарств.

References

- [1] А.С.Иванов, В.В.Поройков, А.И.Арчаков. Биоинформатика: путь от генома к лекарству *in silico*. *Вестн. ПГМУ*, № 4, с.19-23, 2003.
- [2] A.Ivanov, A.Veselovsky, A.Dubanov, V. Skvortsov. Bioinformatics Platform Development. From Gene to Lead Compound. *Methods Mol. Biol.*, volume 316, pages 389-431, 2006.
- [3] D.A.Filimonov, V.V.Poroikov. Probabilistic approach in activity prediction. In: *Chemoinformatics Approaches to Virtual Screening*. Eds. Alexandre Varnek and Alexander Tropsha. Cambridge (UK): RSC Publishing, pages 182-216, 2008.
- [4] A.Geronikaki, D.Druzhilovsky, A.Zakharov, V.Poroikov. Computer-aided predictions for medicinal chemistry via Internet. *SAR and QSAR in Environ. Res.*, volume 19 (1 & 2), pages 27-38, 2008.

Поиск потенциальных сдвигов рамки считывания в генах с помощью генетического алгоритма

Пугачева В.М., Коротков Е.В.

Учреждение Российской академии наук Центр "Биоинженерия" РАН

virentis@gmail.com

Search for potential frameshifts in genes using genetic algorithm

Pugacheva V.M., Korotkov E.V.

Bioengineering Centre of RAS

virentis@gmail.com

Abstract

Almost all genes contain triplet periodicity. Shifts between the reading frame and triplet periodicity can help to find frameshift mutations. Such mutations are considerable events because amino acid sequence coded by gene, changes significantly after the frameshift mutation. But it either does not lost its function, or get another biological function.

To search for such changes, we suggest a method bases on a genetic algorithm. It search for best local alignment of genes and triplet periodicity with insertions and deletions.

Введение

В настоящее время известно, что кодирующие последовательности почти всех генов содержат триплетную периодичность. Триплетная периодичность является достаточно устойчивым явлением и не может быть уничтожена в отдельных мутационных событиях.

Сдвиги между рамкой считывания гена и триплетной периодичностью могут свидетельствовать:

- о присутствии в гене мутаций сдвига рамки считывания;
- об ошибках секвенирования.

Мутаций сдвига рамки изучены мало, однако они представляют большой интерес, т.к. такие мутации существенно меняют аминокислотную последовательность, полученную из исследуемой нуклеотидной последовательности. При замене одного основания ДНК в белке может измениться только одна аминокислота. В то время как при сдвиге рамки считывания, не кратном трем основаниям, изменяются все аминокислоты, лежащие дальше сдвига. Тем не менее, белок сохраняет свою функцию либо получает новую.

Описание метода

Для поиска сдвигов рамки считывания в генах, мы разработали метод поиска, основанный на генетическом алгоритме.

Метод заключается в поиске лучшего выравнивания со вставками и делециями изучаемой последовательности с искусственной периодической последовательностью.

Метод включает использование популяции решений, которые эволюционируют путем естественного отбора. Каждое решение (особь популяции) задает некоторый тип периодичности матрицей весов (т.е. предпочтений символами позиций периодичности) и определяет выравнивание. В качестве функции приспособленности (целевой функции оптимизации) выступает статистическая значимость выравнивания. Выравнивания строятся с помощью метода циклического выравнивания для поиска скрытой периодичности [5].

Общая структура метода:

Инициализация

1. создание G_0

Оценка

2. оценка популяции в поколении n (G_n)
3. если популяция стабильна, тогда Завершение
4. выбор особей для замены

Размножение

5. выбор родителей из G_n
6. создание нового потомка
7. сохранение или отклонение нового потомка в G_{n+1}
8. переход к 6 пока все потомки не будут успешно добавлены в G_{n+1}

9. $n = n + 1$

10. переход к Оценке

Завершение

11. завершение

Первоначально, нулевое поколение G_0 создается случайно. Размер популяции поддерживается постоянным. Чтобы перейти от одного поколения к другому, потомки получаются от родителей, которые выбраны с помощью некоторой разновидности естественного отбора, основанного на их приспособляемости (т.е. чем лучше родитель, тем больше потомков он будет иметь). Для создания потомка последовательно применяются операторы кроссовера и мутации. Эти шаги повторяются многократно, поколение за поколением. Следуя

этому простому процессу, выживаемость популяции возрастает до тех пор, пока никакое улучшение не может быть сделано.

Только часть популяции заменяется в течение каждого поколения. Все особи, ранжируются в соответствии с их приспособленностью, и слабейшие заменяются новыми потомками при создании поколения $n + 1$ из поколения n . Другие индивиды (наиболее приспособленные) просто выживают как есть в течение размножения.

Не существует доказательств того, что генетический алгоритм обязательно достигнет оптимума, даже за бесконечное время. Поэтому решение о прекращении поиска принимается с использованием эвристических критериев. В нашем методе останов поиска происходит, когда не было достигнуто улучшение в течение некоторого заданного числа поколений или число поколений превысило заданный предел.

Литература

- [1] Edgar D. Arenas-Diaz, Helga Ochoterena, and Katya Rodriguez-Vazquez. Multiple Sequence Alignment Using a Genetic Algorithm and GLOCSA. *Journal of Artificial Evolution and Application* Volume 2009, Article ID 963150, 10 pages
 - [2] Holland J. Adaptation in natural and artificial systems. The University of Michigan Press, Ann Arbor. 1975
 - [3] E.V. Korotkov, M.A. Korotkova. Study of the triplet periodicity phase shifts in genes. *Journal of Integrative Bioinformatics*, 7(3):131, 2010.
 - [4] C. Gondro and B.P. Kinghorn. A simple genetic algorithm for multiple sequence alignment. *Genet. Mol. Res.* 6 (4): 964-982, 2007.
 - [5] А.А. Ласкин, Е.В. Коротков, М.Б. Чалей, Н.А. Кудряшов. Локально-оптимальный метод циклического выравнивания для обнаружения скрытой периодичности генетических текстов. NAD-связывающие сайты белков. *Молекулярная биология*. 37 (4): 663-673, 2003.
- [1] Edgar D. Arenas-Diaz, Helga Ochoterena, and Katya Rodriguez-Vazquez. Multiple Sequence

О сравнении двух подходов исследования генетических текстов

Родионов И.Н., Семенов А.В.

ОмГТУ

goruha@gmail.com avs110@inbox.ru

Comparing two methods of genetic text researching

Rodionov I.N., Semenov A.V.

Abstract

In article there are results of comparing statistic method of genetic text researching with position-interval analysis. Position-interval method formalizes and numerically shows nucleotide relative positions. Markov chains with different rank were used as a statistic method.

1 Постановка задачи

В связи с возрастающим значением биоинформатики, возникает потребность в адекватных математических моделях генетических текстов и в методах их исследования. В ряде публикаций [1, 3] упоминается об особой важности исследования взаимного расположения нуклеотидов в нуклеотидной последовательности. Одним из основных подходов к исследованию генетических текстов является статистика. При этом, в качестве модели генетического текста применяются марковские цепи [1,3]. В данной работе ставится задача сравнения, данного, статистического и позиционно-интервального подходов.

2 Позиционно-интервальный анализ

2.1 Интервалы и строй цепи

Основы аппарата интервального анализа представлены А.С. Гуменюком в работе [2]. Для наглядности, приведем понятие строя цепи на примере последовательности нуклеотидов. Пусть чтение нуклеотидной цепи осуществляется поэлементно шагами слева – направо. Заменим в ходе такого просмотра все первые встречные разные знаки алфавита (нуклеотиды) целыми числами, начиная с единицы, и далее – по возрастанию в соответствии с номером появления очередного нового элемента из алфавита {A, C, G, T}. Другие, повторяющиеся знаки данной последовательности, представим числами, установленными при первом просмотре последовательности. Построенную таким образом цепь, в которой знаки (нуклеотиды) заменены числами – номерами вхождения элементов

алфавита – назовем *строим* (рис.1) или *порядком цепи* [2].

Цепь 1	T	T	A	T	T	T	C	A	A	A	A	A	G
Цепь 2	A	A	T	A	A	A	G	T	T	T	T	T	C
Строй цепи 1	1	1	2	1	1	1	3	2	2	2	2	2	4
Цепь 3	G	C	T	C	C	T	G	C	C	A	A	T	A
Строй цепи 2	1	2	3	2	2	3	1	2	2	4	4	3	4

Рисунок 1. Нуклеотидные цепи и их строй

На рисунке 1 показано, что цепи 1 и 2 комплементарны и имеют общий строй, а цепь 3 отлична от предыдущих и соответственно имеет строй, отличающийся от строя цепи 1 и 2.

Определим *интервал «времени»* (в дальнейшем – *интервал*) числом позиций: от одного выделенного элемента до другого, такого же, ближайшего, справа (рис 2). Определим приращение номера элемента как разность между номерами указанных выше элементов. Условимся выполнять чтение текста шагами наблюдения, на каждом из которых происходит очередное событие, фиксируемое приращением – *интервалом*.

Строй	1	1	2	1	2	2	2	2	2	2	4
Интервалы элемента «2»	3		2		1	1	1	1			
	Δ_1		Δ_1		Δ_3	Δ_4	Δ_5	Δ_6			

Рисунок 2. Интервалы выделенного элемента «2».

2.2 Удаленность

В отличие от математической теории передачи сообщений, в которой за основу взята вероятность сообщения; в позиционно-интервальном анализе, среднегеометрический интервал, между эквивалентными сообщениями, является основой информационной единицей. Аналогом энтропии, в позиционно-интервальном анализе, является удаленность, определяемая как двоичный логарифм

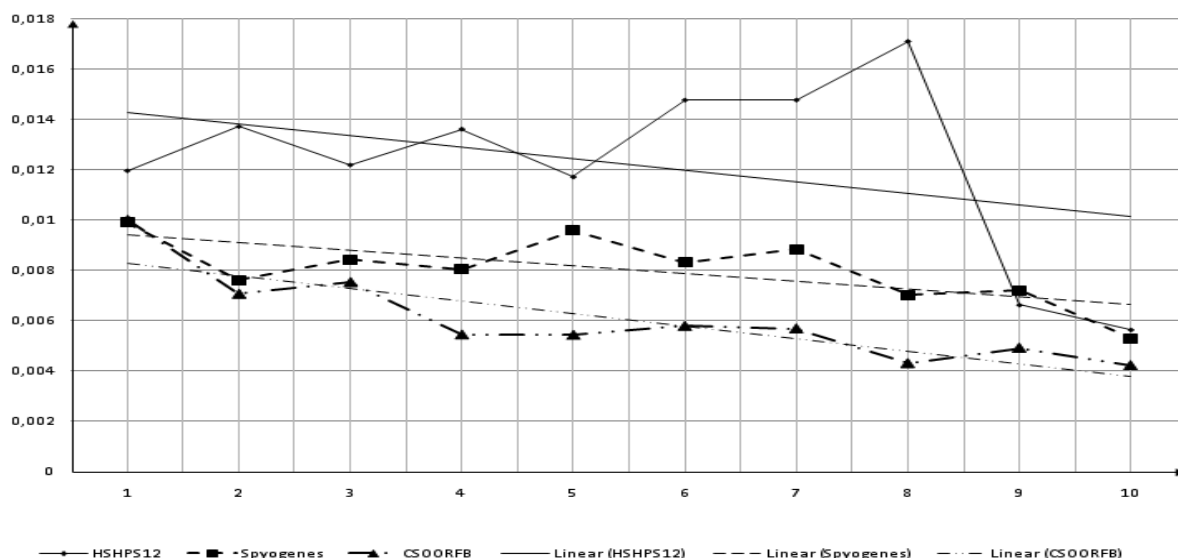


Рисунок 3. Ошибка удаленности в зависимости от порядка Марковской цепи

среднегеометрического всех интервалов информационно цепи (1).

$$g = \log_2 \prod_{j=1}^m l_j, \quad (1)$$

где m – мощность алфавита, на котром определена цепь (для нуклеотидной цепи равна 4), n – длина цепи, n_j – количество j -ого нуклеотида в цепи, Δ_{gj} – среднегеометрический интервал для j -ого нуклеотида, определяемый по формуле (2).

$$\Delta_{gj} = \prod_{i=1}^{n_j} \Delta_{ij}^{\frac{1}{n_j}}, \quad (2)$$

где Δ_{ij} – i -й интервал j -ого нуклеотида.

Удаленность чувствительна к взаимному расположению элементов в цепи; всегда меньше энтропии, и только при расположении элементов, называемом «верхним крайним» строем, приблизительно равна энтропии. Удаленность использовалась как мера измерения строя.

3 Сравнение подходов

Статистический и позиционно-интервальный подходы исследуют различные свойства информационной цепи. Статистический анализ, описывает соотношение событий, позиционно-интервальный анализ описывает их взаимное расположение. Из результатов сравнения подходов следует – насколько каждый из них способен учитывать свойства информационной цепи, не описываемые им непосредственно.

Цепи Маркова являются наиболее удобным статистическим средством для описания строя.

На данном этапе исследования, следующий ход эксперимента:

1) Для исходной нуклеотидной цепи O получили марковскую цепь i -ого порядка M_i , где $i=0$.

2) Используя M_i сгенерировали k нуклеотидных цепей $[R_{ik}]$ длиной равной исходной O .

3) Для $[R_{ik}]$ определим удаленности $[g_{ik}]$ и рассмотрим их, как реализации случайной величины. Определим её Mx_i и CKO_i .

Исследование проводилось на 250 нуклеотидных цепях. Для каждой эксперимент повторялся с изменением i от 1 до 10, при $k=500$. При анализе сравнивались массивы $[Mx_i]$ и $[CKO_i]$. Для каждого эксперимента строился график отклонения (ошибки) Mx_i от удаленности исходной цепи $g(O)$. На рисунке 3, представлены графики и тренды отклонений для некоторых нуклеотидных цепей.

4 Заключение

В результате исследования, показано, что, широко применяемые в биоинформатике, марковские цепи учитывают взаимное расположение элементов посредством переходных вероятностей. Однако, в отличие от позиционно-интервального анализа, они не предоставляют средств его компактного числового описания.

References

- [1] Александров А.А., Александров В.В. и др. Компьютерный анализ генетических текстов. // М.: Наука, 1990. – 264 с.
- [2] Гуменюк А.С., Богорад Д.М. Об исчислении взаимного расположения элементов в массивах данных. // Труды РНТОРЭС им. Попова, серия: научная сессия, посвященная Дню радио. Выпуск: LIX-2, 2004. – С. 114-116.
- [3] Садовский М.Г. Информационно-статистический анализ нуклеотидных последовательностей. Диссертация на соискание ученой степени доктора физико-математических наук. Красноярск, 2004. – 394 с.

Использование спектрального представления для анализа последовательностей ДНК

Тетуев Р.К., Дедус Ф.Ф., Ольшевец М.М., Панкратов А.Н., Пятков М.И., Назипова Н.Н.

Институт математических проблем биологии РАН,

ruslan.tetuev@gmail.com

Abstract

In this paper we discuss the ability of replacing edit distance by faster and less complicated similarity estimation of DNA-strings, based on spectral representation of signals.ⁱ

Введение

Как известно ДНК имеют линейную структуру и могут быть однозначно записаны в виде ряда букв, к примеру, АСТГА, где {А,Т,Г,С} – заглавные буквы названий нуклеиновых кислот. Поэтому говорят о ДНК строках и нуклеотидных последовательностях. В широком смысле задачи биоинформатики – это вычислительная обработка любой биологической информации, но чаще только нуклеотидных или аминокислотных последовательностей.

Нижеописанный метод основан на спектральном приближении, что означает работу не с текстом, а с функциями. Для преобразования текста в функции существует ряд количественных замен, например:

А $\Leftrightarrow$ 0; Т $\Leftrightarrow$ 1; Г $\Leftrightarrow$ 2; С $\Leftrightarrow$ 3;

однако авторы отдали предпочтение следующему:

А, Т $\Leftrightarrow$ 2

Г, С $\Leftrightarrow$ 3

Сделать выбор в пользу последнего преобразования авторов подтолкнул тот факт, что у него существует явный физический смысл. Дело в том, что двойная спираль ДНК, состоящая из двух свитых цепочек нуклеотидов, на разных участках имеет разную силу связи этих цепочек между собой. По этой причине при нагреве цепочки разъединяются в одних местах скорее, чем в других. Основным фактор, влияющий на это – количество водородных связей на данном участке и оно тем выше, чем больше на нём Г и С. Так, участки, состоящие только из Г и С, в 1,5 раза крепче тех, что содержат только А и Т. Говоря далее об участках ДНК, будем иметь в виду определённый размер W. Выпишем простую формулу количества водородных связей F на участке ДНК [x, x+W-1]:

$$F_W^{(G,C)}(x) = 2 \left(\sum_{i=x}^{x+W-1} A + \sum_{i=x}^{x+W-1} T \right) + 3 \left(\sum_{i=x}^{x+W-1} G + \sum_{i=x}^{x+W-1} C \right)$$

где $\sum A$, $\sum T$, $\sum G$, $\sum C$ – условное обозначение количества букв А, Т, Г, С на заданном интервале.

Как видим, тем самым мы предложили функцию, определяющую L-W+1 значений при рассмотрении

ДНК строки длины L. При вычислениях на ЭВМ принято использовать схожую, но более простую функцию, определяемую в следующем виде:

$$f_W^{(G,C)}(x) = \frac{\sum_{i=x}^{x+W-1} G + \sum_{i=x}^{x+W-1} C}{W} \cdot 100$$

и называемую функцией GC-содержания. Выпишем замечательные свойства этой функции:

- ✓ устойчивость к изменению аргумента;
- ✓ устойчивость к изменению параметра;
- ✓ устойчивость к замене букв в ДНК;
- ✓ устойчивость к вставке букв в ДНК;
- ✓ устойчивость к изъятию букв из ДНК;

Существуют и другие способы преобразования текстовой информации в функции, например, расчёт энтропии участка ДНК или оценка степени подобия участка с некоторым образцом. Преимущества этих подходов на фоне большей сложности их расчёта показали авторам недостаточно убедительными.

Спектральная оценка подобия ДНК

По аналогии с функцией GC-содержания введём функцию AC-содержания:

$$f_W^{(A,C)}(x) = \frac{\sum_{i=x}^{x+W-1} A + \sum_{i=x}^{x+W-1} C}{W} \cdot 100$$

Легко доказать, что при знании первых W-1 букв ДНК строки остальные L-W+1 буквы строки можно восстановить по значениям функции $f_W^{(A,C)}$ и $f_W^{(G,C)}$. Если же пренебречь влиянием начального участка при $W \ll L$, то функцию AC/GC-содержания можно даже считать новым видом записи ДНК:

$$f_W^{(AC/GC)}(x) = f_W^{(A,C)}(x) + i \cdot f_W^{(G,C)}(x)$$

Т.к. $f_W^{(A,C)}$, $f_W^{(G,C)}$ – функции дискретного аргумента, то по теореме Уитеккера-Котельникова-Шеннона придём к точному спектральному представлению:

$$f_W^{(A,C)} = \sum_{n=0}^{L-W} a_n \varphi_n(x),$$
$$f_W^{(G,C)} = \sum_{n=0}^{L-W} b_n \varphi_n(x),$$

где $\{\varphi_n(x)\}$ – функции любого ортонормированного базиса, заданного на сетке в L-W+1 значений. Тогда получим, что пара спектральных рядов $\{a_n\}$ и $\{b_n\}$ также есть новый вид записи ДНК – *спектральное представление ДНК последовательности*.

Многие задачи биоинформатики основаны на оценках схожести фрагментов строк и ранее часто решались с помощью буквенных операций, когда наименьшим рассматриваемым фрагментом строки

являлась одна буква. Такие сравнения проводились медленно, пока не были ускорены за счёт операции одновременного сравнения группы букв с бинарной оценкой все-совпадают / не-все-совпадают.

Например, программа Tandem Repeats Finder ускорена за счёт применения *кортежей* – групп из 5 букв [1]. BLAST при поиске в ДНК [2] применяет уже группы свыше 10 букв. Увеличение размера групп вызвало бы и дальнейшее ускорение, если бы не основной фактор эволюции – *мутации*. Из-за них такое сравнение даже для близких копий затруднено частыми малыми различиями. Тогда очевидно, что создание более эффективных алгоритмов сравнения требует введения оценок неточного совпадения.

Популярная оценка неточного совпадения двух групп букв называется *редакционным расстоянием* и равна минимальному числу исправлений первой группы, после которых она в точности совпадает со второй. Заметим, что вычисление данной величины требует трудоёмких, длительных операций и имеет в лучшем случае временную сложность $O(L)$ [3]. Ниже предложим новую, более быструю оценку.

Пусть для ДНК строк S_1, S_2 длины L получены функции их описания $f_1(x), f_2(x)$ длины $L - W + 1$ и их спектры $\{A'_n\}, \{A''_n\}$. Эти функции могут быть построены как профили GC- и/или AC-содержания или иначе, главное – чтобы функции удовлетворяли ранее изложенным условиям устойчивости. Тогда предложим простую меру сходства строк S_1 и S_2 :

$$c(S_1, S_2) = \sum_{x=0}^{L-W} (f_1(x) - f_2(x))^2,$$

удовлетворяющую неравенству Коши-Буняковского и условиям $0 \leq c(S_1, S_2)$ и $c(S_1, S_1) = 0$. Интересно, что утверждение $\{c(S_1, S_2) = 0 \Leftrightarrow S_1 = S_2\}$ в общем случае неверно, но для функции AC/GC-содержания оно выполняется при условии совпадения хотя бы одного из $(W - 1)$ -тибуквенных подстрок S_1 и S_2 . Далее, согласно равенству Ляпунова-Стеклова

$$\sum_{x=0}^{L-W} (f_1(x) - f_2(x))^2 = \sum_{x=0}^{L-W} (A'_n - A''_n)^2,$$

откуда получим новое выражения введённой меры:

$$c(S_1, S_2) = \sum_{n=0}^{L-W} (A'_n - A''_n)^2$$

Преимущество этого спектрального представления в том, что большую часть ряда теперь можно просто отбросить и при $N \ll L - W$ получить приближение

$$c(S_1, S_2) \approx \sum_{n=0}^N (A'_n - A''_n)^2,$$

что существенно снизит время вычислений до $O(N)$, где N выбирается опытным путём. Так, $N \approx 25$ для $L - W = 1000$ при построении *спектральной дот-матрицы*, т.е. результата сравнения всех участков длины $L - W$, представленного в виде матрицы [4].

Результаты

В работе предложена быстрая оценка подобия строк, основанная на спектральном представлении функций, описывающих ДНК последовательности. Показано ускорение алгоритмов до $O(N)$, что очень существенно при $N \ll L$, где N, L – длины спектров функций и сравниваемых участков ДНК. Ускорение произошло благодаря замене одной медленной, хотя и точной оценки подобия строк на другую быструю, хотя и приближённую оценку. Действительно:

1. если схожи две строки ДНК, то схожи и их функции описания;
2. если схожи функции описания, то схожи их первые спектральные коэффициенты;
3. тогда если схожи строки ДНК, то схожи их первые спектральные коэффициенты.

Предложенная оценка предполагает верность и обратного утверждения, т.е. первые спектральные коэффициенты функций определяют подобие строк. Такое предположение, строго говоря, неверно из-за отсутствия симметрии у рассмотренного свойства транзитивности, но практика применения оценки показала её состоятельность при малом количестве ошибок 1-го рода (ложных срабатываний).

В заключение заметим, что такие спектры можно использовать и при поиске более сложных структур в ДНК [4, 5, 6] при использовании вычислительных приёмов, изложенных в работе [7].

Работа выполнена при финансовой поддержке РФФИ, проект № 09-07-00455.

References

- [1] Benson G., Tandem repeats finder: a program to analyze DNA sequences // Nucleic Acids Research, 1999 – V.27, №2. Pp. 573–580.
- [2] Basic Local Alignment Search Tool – Nucleotide BLAST. <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.
- [3] Ukkonen E., On-line construction of suffix trees // Algorithmica, 1995 – V.14, №3. Pp. 249–260.
- [4] Pankratov A. N., Gorchakov M. A., Dedus F. F., et al., Spectral analysis for identification and visualization of repeats in genetic sequences // Pattern Recognition and Image Analysis, 2009 – V.19, №4. Pp. 687–692.
- [5] Тетюев Р. К., Назипова Н. Н., Панкратов А. Н., Дедус Ф. Ф. Поиск мегасателлитных тандемных повторов в геномах эукариот по оценке осцилляций кривых GC-содержания // Математическая биология и биоинформатика. 2010. – Т. 5, №1. С. 30–42.
- [6] Tetuev R. K., Nazipova N. N. Repbase Reports. 2010. – V. 10, №5. P. 776–776.
- [7] Тетюев Р. К., Алгебра спектральных преобразований в задачах обработки данных. 2007.

Применение метода Монте Карло для поиска сдвигов рамки считывания в кодирующих последовательностях ДНК

Руденко В.М.^(1,2), Коротков Е.В.^(1, 2)

Центр «Биоинженерия» РАН⁽¹⁾, НИЯУ МИФИ⁽²⁾
v.m.rudenko@gmail.com

Abstract

The Monte Carlo method is used for search of mutations in genes arising by reading frame (RF) shifts. Developed approach is based on comparison of triplet frequencies before and after a place of putative RF shift. We revealed more than 140 000 genes with potential RF shifts in kegg data bank version 46. The received results show, that the Monte Carlo method is more effective for detection the RF shifts in genes than analytical approaches.

1 Введение

Новые гены могут возникать вследствие различных мутаций в исходных последовательностях ДНК, в частности, посредством вставки или делеции некоторого количества нуклеотидов. Если это количество не кратно трем, то исходном гене происходит сдвиг рамки считывания. В результате этой мутации последовательность белка, транслируемая с гена, полностью изменяется ниже позиции сдвига. Новая последовательность может кодировать белок с прежней или же новой функцией. Изучение белковых последовательностей, возникших вследствие сдвига рамки считывания, имеет важное значение для теоретических исследований в области молекулярной эволюции [2]. Проведенные к настоящему времени исследования обнаружили феномен существования сдвигов, однако выявили лишь небольшое их число. Целью настоящей работы является разработка нового математического метода выявления сдвигов рамки в последовательностях оснований ДНК, более мощного, чем существующие аналоги.

2 Методы

2.1 Мера сдвига

Ранее в качестве количественной характеристики сдвига рамки считывания нами применялась статистика, основанная на степени подобия типов триплетной периодичности во фрагментах фиксированной длины w справа и слева от позиции предполагаемого сдвига [3]. Поскольку триплетная периодичность объясняется, в частности, предпочтениями организма в использовании

различных триплетов для кодирования определенных аминокислот и вырожденностью генетического кода, то можно предположить, что кодирующая последовательность без сдвига будет иметь однородное распределение частот триплетов вдоль всей своей длины. Это предположение было заложено в основу построения новой меры сдвига рамки считывания в последовательности ДНК.

Рассмотрим три возможные рамки считывания T1, T2, T3 в последовательности ДНК.

<u>DNA sequence:</u>	<u>...atggcgagagaggtgcctatagagaaattg...</u>
T1: <u>Am.acid seq1:</u>	...123123123123123123123123 123... <u>...M A R E V P I E K L ...</u>
T2: <u>Am.acid seq2:</u>	...312312312312312312312312 312... <u>...W R E R C L \$ R N</u>
T3: <u>Am.acid seq3:</u>	...231231231231231231231231 231... <u>....G E R G A Y R E I</u>

Рис.1. Три возможные рамки считывания в гене (\$ - стоп-кодон).

Пусть требуется определить наличие сдвига в позиции k (k кратно 3) последовательности. Для этого проверим гипотезу об однородности распределений частот триплетов между фрагментом последовательности с координатами $(k-w, k)$ в рамке T1 и фрагментом $(k+j, k+w+j)$ в рамке T_j, $j=1..3$, используя формулу [1]:

$$I_j = \sum_{i=1}^{64} f_i \log f_i + \sum_{i=1}^{64} v_i^j \log v_i^j - \sum_{i=1}^{64} (f_i + v_i^j) \log (f_i + v_i^j) + (N_1 + N_2) \log (N_1 + N_2) - N_1 \log N_1 - N_2 \log N_2 \quad (1)$$

w – длина фрагментов слева и справа от позиции k ;
 f_i – частоты триплетов для фрагмента последовательности с координатами $(k-w, k)$; v_i^j – частоты триплетов для фрагмента $(k+j, k+w+j)$; $N_1=N_2=w$ – количество триплетов. Если в позиции k сдвига нет, то минимум

I_j будет достигаться для $j=1$. Если I_2 или I_3 принимает малое значение (однородность), а I_1 велико, то в позиции k наблюдается сдвиг.

Величина $2I_j$ имеет распределение χ^2 с 63 степенями свободы. Однако, согласно [1] $2I_j$ хорошо согласуется со стандартным распределением только если все f_i и v_i^j больше 5. При исследовании реальных последовательностей ДНК данное требование не всегда достигается. Это не позволяет использовать стандартные формулы для расчета статистической значимости предполагаемого сдвига. Для оценки того, что величины $2I_j$ неслучайны нами применялся метод Монте Карло. Метод Монте Карло позволяет сгенерировать распределение величины $2I_j$ для каждого изучаемого фрагмента последовательности и перейти от рассмотрения статистики $2I_j$ к Z_j - отклонению наблюдаемого значения $2I_j^{ob}$ от ожидаемого $2I_j^{exp}$, рассчитанного на основании построенного распределения (2).

$$Z_j = \frac{2I_j^{ob} - 2I_j^{exp}}{\sigma_j} \quad (2)$$

здесь σ_j – корень из дисперсии случайной величины $2I_j$. Величины Z_j имеют приблизительно стандартное нормальное распределение. Обозначим за $p_j = P(N(0,1) > Z_j)$. При наличии сдвига p_j для $j=2$ или $j=3$ Z_j принимает относительно большое значение, в то время как p_1 мало. На практике удобно использовать в качестве количественной характеристики сдвига рамки считывания величины:

$$F_2 = -\log(p_1 / p_2) \quad (3)$$

$$F_3 = -\log(p_1 / p_3)$$

Если подобие частот триплетов правой части последовательности - $(k+j, k+j+w)$ с левой частью последовательности - $(k-w, k)$ наблюдается для второй рамки считывания $j=2$, то p_1 мало, а p_2 велико, что приводит к росту F_2 . Превышение F_2 некоторого порогового уровня F_0 будет свидетельствовать о том, что в позиции k наблюдается вставка $1+3n$ нуклеотидов, или же делеция $2+3n$ нуклеотидов. Аналогичным образом, при $F_3 > F_0$ можно говорить, что в позиции k имеет место вставка $2+3n$ нуклеотидов, или же делеция $1+3n$ нуклеотидов.

3 Результаты

С помощью предложенного метода был проанализирован банк данных кодирующих последовательностей ДНК Kegg-46. При сканировании последовательностей исследовались все позиции k последовательности кратные 3. Ширина окна слева и справа от k варьировалась от 150 до 600. Число случайных испытаний при реализации метода Монте Карло было равным 200. В качестве порогового уровня F_0 была выбрана величина 5.0, которая определялась методом Монте

Карло. Уровень $F_0=5.0$ соответствует вероятности ошибочной идентификации сдвигов около 5%. На этом уровне было обнаружено 140736 последовательностей со сдвигами, что составляет 7% от всего числа последовательностей банка данных.

На рис.2 изображены графики функций F_2 и F_3 для последовательности Vscen_0873, в которой был обнаружен сдвиг. Последовательность Vscen_0873 была взята из генома бактерии Burkholderia cenocepacia, кодирует lytic transglycosylase. Из рисунка видно, что в позиции 876 последовательности наблюдается вставка символа, о чем свидетельствует высокое значение функции $F_2=11.87$, в то время как $F_3 = -14.76$.

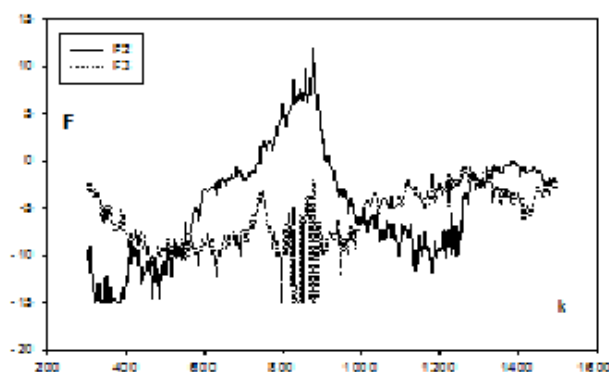


Рис.2. Вставка в последовательности Vscen_0873

Прикладные задачи, которые можно решать с помощью разработанного метода это, во-первых, выявление в последовательностях ДНК мутаций – вставок и делеций длины некротной трем. На основании этих данных можно «восстановить» древнюю аминокислотную последовательность белка, кодируемого геном, что может представлять интерес для изучения эволюционных процессов в генах. Во-вторых, метод позволяет определять наличие ошибок секвенирования в новых последовательностях ДНК. Автоматическое выявление и исправление ошибок повысит качество секвенированных последовательностей и позволит провести их корректную аннотацию.

Работа выполнена при финансовой поддержке ФЦП "Научные и научно-педагогические кадры инновационной России" на 2009-2013гг.

References

- [1] S. Kullback, Information Theory and Statistics. New York: Wiley, 1959.
- [2] J.Raes, Y.van de Peer. Functional divergence of proteins through frameshift mutations, *Trends Genetics*, volume 21, 2005, pages 428-431
- [3] Е.В.Коротков, В.М.Руденко. Сдвиг фазы триплетной периодичности в нуклеотидных последовательностях генов. *Математическая биология и биоинформатика*, том 4. № 2, с. 66-80, 2009

Статистические методы анализа первичных последовательностей белков ВИЧ-1

Сергеев Р.С.⁽¹⁾, Тузиков А.В.⁽²⁾, Еремин В.Ф.⁽³⁾

Белорусский государственный университет, Минск⁽¹⁾, Объединенный институт проблем информатики, Минск⁽²⁾, РНПЦ эпидемиологии и микробиологии, Минск⁽³⁾
roma.sergeev@gmail.com

Abstract

Many human immunodeficiency virus type-1 (HIV-1) infected persons are treated with multiple protease and reverse transcriptase (RT) inhibitors. Due to high variability of the virus strains these drug classes often lead to the development of drug resistance which is a major obstacle to the effective treatment of HIV-1 infection. The proposed method is designed to analyze HIV-1 protease and reverse transcriptase mutations, recognize whether they may lead to drug-resistance surveillance and support decisions on further treatment regimen.

1 Основной раздел

1.1 Материалы исследования

В настоящее время разработаны и применяются разнообразные схемы лечения ВИЧ-инфекции (высокоактивная антиретровирусная терапия, АРТ), которые замедляют прогрессирование ВИЧ-инфекции и ее переход в стадию СПИДа (синдрома приобретенного иммунодефицита). Однако сопротивляемость вируса к каждому из применяемых в них препаратов (обычно это ингибиторы протеазы, ингибиторы обратной транскриптазы, ингибиторы слияния, а в последнее время и ингибиторы интегразы) часто ведет к развитию кросс-резистентности [1, 2]. Тем не менее практика применения ингибиторов вирусных белков показывает, что лишь некоторые из мутаций способны вызывать появление устойчивости вируса к лекарственным препаратам.

Стоит отметить, что большинство опубликованных данных, а также программных систем для анализа мутаций ВИЧ-1 касается субтипа В. Однако в Беларуси и некоторых других странах Восточной Европы наибольшее число случаев инфицирования связано с распространением ВИЧ-1 субтипа А.

В связи с этим цели и задачи исследования на текущем этапе формулировались следующим образом:

1) Разработать методику статистического распознавания мутаций участка гена *pol* ВИЧ-1, ассоциируемых с приемом ингибиторов, а также выявления кластеров взаимосвязанных мутаций.

2) Предложить подход к практической реализации задачи по анализу мутаций в первичных последовательностях белков ВИЧ-1 субтипа А и применить его для анализа последовательностей, полученных в процессе исследования пациентов из Беларуси.

Для решения поставленной задачи был предложен подход, позволяющий получить ответы на следующие вопросы:

Какие мутации вызваны применением лекарственных препаратов?

Какие мутации являются взаимосвязанными между собой?

Какие мутации привели к появлению устойчивости к применяемым препаратам?

Краткая характеристика методов, которые применялись для решения задачи, приведена ниже.

1.2 Взаимосвязь с терапией

Для анализа возможной взаимосвязи между появлением мутаций в определенных участках последовательностей и антиретровирусной терапией используется метод анализа таблиц сопряженности. Таблица строится для каждой мутации в некоторой позиции отдельно по каждому препарату (если есть необходимость определить, каким именно препаратом была вызвана мутация). В качестве признаков выбираются наименование препарата и мутация в рассматриваемом сайте последовательности. Для оценки достоверности гипотезы о независимости этих признаков выполняется процедура проверки статистических гипотез для одного либо нескольких сайтов [3].

1.3 Выявление коррелированных мутаций

Выявление коррелированных мутаций в сайтах последовательностей белков интересно по нескольким причинам. Во-первых, совместно эволюционирующие сайты могут оказаться близко расположенными друг от друга, если говорить о пространственной структуре белковой молекулы. Во-вторых, именно совместное воздействие связанных друг с другом мутаций часто вызывает устойчивость вируса к применяемому препарату.

Для оценки коррелированности сайтов по имеющемуся множественному выравниванию восстанавливается филогенетическое дерево, например, используя метод максимального правдоподобия [4]. Затем предсказываются

последовательности во внутренних узлах этого дерева [7].

Чтобы оценить корреляции между любыми двумя выбранными сайтами выравнивания, в простейшем случае, по информации из построенного филогенетического дерева оценивается, насколько синхронно происходили мутации в этих сайтах в ходе эволюции, вычисляется коэффициент корреляции. Для определения значимости результатов строится тест проверки статистических гипотез.

Существуют альтернативные методы, которые могут быть использованы для выявления коррелированных мутаций [5, 6].

1.4. Выявление резистентных мутаций

Задачей этого этапа является на основании собранных статистических данных о результатах лечения получение ответа на вопрос, влечет ли мутация в заданной позиции развитие резистентности к выбранному лекарственному препарату. Для решения этой задачи применяется метод анализа таблиц сопряженности.

Следует отметить, для того, чтобы точно охарактеризовать мутации, которые происходят в изолятах вируса, необходимо исследовать большое число последовательностей, выделенных у людей, зараженных ВИЧ-1, которые подвергались и не подвергались лечению ингибиторами. В последнем случае обязательным является наличие информации о лекарственных препаратах, которые использовались при лечении.

2 Практическая реализация

2.1 Программная реализация

Для практической реализации предложенного подхода была разработана тестовая версия экспериментального ПО, которое включает как сторонние компоненты, так и самостоятельно разработанные модули.

Используемые для первичного анализа сторонние компоненты были предварительно установлены из исходных кодов и сконфигурированы под мультипроцессорную платформу: ClustalW для выполнения выравнивания, RAxML для филогенетического анализа, пакеты PAML и GASP.

В ходе анализа последовательностей выявляются мутации, проводится их статистический анализ. Реализован подход к анализу последовательностей на предмет наличия коррелированных мутаций, использующий эволюционную информацию, полученную из восстановленного филогенетического дерева.

Все программы установлены на сервере ОИПИ itk-115.bas-net.by, для доступа требуется установка дополнительного клиентского ПО.

2.2 Тестовая база пациентов

Параллельно с программной реализацией специалистами лабораторией диагностики ВИЧ и сопутствующих инфекций РНПЦ эпидемиологии и микробиологии выполнялось секвенирование образцов, полученных от ВИЧ-инфицированных пациентов из Беларуси, и пополнение базы последовательностей гена pol.

Для тестирования предложенного подхода на больших выборках данных были использованы последовательности из международных банков данных NCBI и Los-Alamos [8, 9]. Из этих же источников для целей генотипирования были получены и включены в выравнивания эталонные последовательности гена pol ВИЧ-1 различных субтипов.

References

- [1] S.Y. Rhee [et al.] HIV-1 Protease and Reverse Transcriptase Mutations: Correlations with Antiretroviral Therapy in Subtype B Isolates and Implications for Drug-Resistance Surveillance. *The Journal of Infectious Diseases*, 192, pages 456-465, 2005.
- [2] M.J. Gonzales, R.N. Machekano, R.W. Shafer. Human immunodeficiency virus type 1 reverse-transcriptase and protease subtypes: classification, amino acid mutation patterns, and prevalence in a northern California clinic-based population. *J. Infect. Dis.*, 184, pages 998-1006, 2001.
- [3] Y. Benjamini, Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser.*, 57, pages 289-300, 1995.
- [4] J. Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Mol. Evol.*, 17, pages 368-376.
- [5] U. Gobel [et al.] Correlated Mutations and Residue Contacts in Proteins. *Proteins: Structure, Functions and Genetics*, 18, pages 309-317.
- [6] D.D. Pollock, W.R. Taylor, N. Goldman. Coevolving Protein Residues: Maximum Likelihood Identification and Relationship to Structure. *J. Mol. Biol.*, 287, pages 187-198.
- [7] Edwards R.J., Shields D.C. GASP: Gapped Ancestral Sequence Prediction for proteins. *BMC Bioinformatics*, 5(1):123, 2004.
- [8] National Center for Biotechnology Information, 1988-2010.
<http://www.ncbi.nlm.nih.gov/>
- [9] Los Alamos HIV Database, 2005-2010.
<http://www.hiv.lanl.gov/content/>

Изучение склеенных генов в банке данных KEGG

Суворова Ю.М., Коротков Е.В.

Центр “Биоинженерия” РАН

suvorovay@gmail.com

A study of fused genes in the KEGG databank

Suvorova Y.M., Korotkov E.V.

Centre of Bioengineering, Russian Academy of Sciences

Abstract

Some of recombination events (like translocation, inversion or segmental duplication) can lead to accidental fusions of DNA or, in case of coding sequences, to formation of fused genes. In this case fused parts of a new gene in one species may be orthologous to different separated genes in another genome. This fused gene may results in a new function by the joining corresponding peptide domains into multidomain protein. The suggested method of searing for fusion events is based on the phenomenon of the triplet periodicity (TP) of the coding regions. A mathematical algorithm for detection of genes with different TP types in the sequence has been developed. This fusions of TP types may be in the same time the fusions of genes. And by the method gene sequences from the KEGG database (release 48) were analyzed with a purpose of revealing genes with fusions of TP types. With the significant level equal to 5% we found 311221 genes in KEGG-48 containing at least one such a fusion event.

1 Введение

Рекомбинационные события, такие как перестановки, инверсии и удвоения сегментов последовательности ДНК, могут приводить к образованию склеенных генов. Генов, участки которых ранее транслировались независимо, а теперь представляют собой единую кодирующую последовательность. Эти склеенные гены могут в свою очередь приводить к образованию белков с новыми функциями [5]. Таким образом, анализ этих генов может дать полезную информацию об эволюции генов и формировании новых функциональных белков.

В настоящее время при поиске склеек отталкиваются от результатов поиска подобных участков склеенных генов, как отдельных генов в других геномах [5] и от доменной архитектуры белков [6]. Такой подход имеет очевидное ограничение – если независимые гены-предки были утеряны в процессе эволюции или просто еще не содержатся в базах данных, то определение склеек

на основании сходства последовательностей невозможно. В данной работе поиск основан на структуре самой изучаемой последовательности, на изменении типа триплетной периодичности в ней. Триплетная периодичность (ТП) – это отличительное свойство последовательностей, кодирующих белки [2]. Такая специфичность позволяет использовать ее для отделения кодирующих последовательностей от некодирующих [1]. Однородность ТП также используется, например, для поиска в кодирующих последовательностях следов различных мутаций, вызывающих сдвиг исходной рамки считывания в гене [4]. Однако, существуют кодирующие последовательности, несущие участки с разными типами ТП. Поиск таких сложных по структуре периодичности генов являлся предметом рассмотрения в настоящей работе.

2 Метод

В каждой последовательности сравнивались небольшие смежные участки генов (длина участков L варьировалась от 60 до 600 нуклеотидных остатков, L кратно трем). Если рассматривается некоторая координата x последовательности (значение x кратно трем), то это соответственно участки $[x-L+1; x]$ и $[x+1; x+L]$. Для этих участков строились частотные матрицы размера 3×4 . Признаком столбца в такой матрице служит позиция периода ‘1’, ‘2’ или ‘3’, а признаком строки – символ основания ДНК (‘a’, ‘t’, ‘c’ или ‘g’, соответственно). Затем проводилась нормализация и поэлементное сравнение этих матриц. Сумма квадратов всех элементов матрицы сравнения S имеет распределение хи-квадрат с шестью степенями свободы. Анализируемое значение – величина $Z = -\log(p(S))$.

Кроме того, смена типа периодичности в первой рамке может быть вызвана мутацией (вставкой или делецией не кратной трем основанием), приведшей к сдвигу рамки считывания, но не к изменению ТП как таковой [4]. Для того чтобы исключить из результатов такие случаи, для любых двух рассматриваемых участков проводилось не одно, а три сравнения. С единственной матрицей левого участка $([x-L+1; x])$, соответствующей исходной рамке считывания в гене, сравнивались все три возможные рамки правого участка $([x+1; x+L], [x+2; x+L+1], [x+3; x+L+2])$. Такие вычисления проводились на протяжении всей последовательности гена, с шагом равным трем основаниям. Координата, соответствующая максимуму итоговой величины Z в гене может рассматриваться, как место

потенциальной склейки, если ее значение превышает некоторый порог.

3 Статистика и результаты

Для определения порогового значения величины Z был использован метод статистических испытаний Монте-Карло. Описанным выше алгоритмом были обработаны все последовательности генов банка данных KEGG-48. Для создания контрольной выборки все гены KEGG-48 были перемешаны, с сохранением триплетной структуры последовательностей, при таком перемешивании сглаживаются все существующие в гене неоднородности ТП. Из сравнения результатов обработки “случайных” и реальных генов было определено пороговое значение величины Z на уровне 5%. На этом уровне в банке данных KEGG-48 было найдено более 300 тысяч генов, содержащих минимум одну смену типа периодичности.

Часть из этих результатов была подтверждена при помощи поиска подобий. С помощью программы BLASTX для 131323 случаев было найдено подтверждающее выравнивание с одной из последовательностей базы данных Swiss-Prot: для 54406 генов найдено подобие соответствует только левому участку (до места предполагаемой склейки), для 60333 – только правому и для 16584 подобия обеих частей (существующие независимо в разных генах).

Для того чтобы определить, какая часть от общего числа потенциальных случаев склеек может

быть найдена при помощи метода, 2520 матриц классов ТП [2] попарно сравнивались при помощи описанного алгоритма. Сравнение с учетом размеров классов показало, что алгоритм на значимом уровне различает 55% сочетаний из них. Т.е. если предположить, что эти события склейки происходят равновероятно между всеми типами периодичности, то из всех генов банка KEGG-48 почти 600 тысяч генов могут представлять собой результат склейки, что составляет примерно 15% от общего объема.

Список литературы

- [1] Bernaola-Galván et.al. Finding Borders between Coding and Noncoding DNA Regions by an Entropic Segmentation Method, *Physical review*, 86 (6) 1342-1345, 2000.
- [2] F.E. Frenkel and E.V. Korotkov. Classification analysis of triplet periodicity in protein-coding regions of genes, *Gene*, 421(1-2), 52-60, 2008.
- [3] M. Kanehisa and S. Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27-30, 2000.
- [4] E.V. Korotkov and M.A. Korotkova. Study of the triplet periodicity phase shifts in genes, *Journal of Integrative Bioinformatics*, 7(3):131, 2010.
- [5] S. Pasek et.al. Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* 22 n 12, 1418–1423, 2006.
- [6] M. Serres and M. Riley. Gene fusions and gene duplications: relevance to genomic annotation and functional analysis. *BMC Genomics.*, 6:33, 2005.

New Family of Isomaltase Genes *IMA* in *Saccharomyces cerevisiae*

Daniil G. Naumoff and Gennadi I. Naumov

State Institute for Genetics and Selection of Industrial Microorganisms, Moscow, Russia
daniil_naumoff@yahoo.com

Abstract

Comparative sequence analysis of *Saccharomyces cerevisiae* S288C genome and known isomaltase gene from *S. cerevisiae* ATCC56960 allowed us to found a new family of polymeric isomaltase genes *IMA1–IMA4*, located in the subtelomeric regions of VII, XV, IX, and X chromosomes, respectively.

1 Introduction

The study of utilization of natural α -glucosides (maltose, sucrose, etc.) by yeast is of great fundamental and applied importance. The hydrolytic enzyme α -glucosidase provides the cell with carbohydrate nutrition and allows it to yield fermentation products. α -Glucosidase yeast of the genus *Saccharomyces* have been long used for fermentation of bread, kvass, and alcohol beverages as well as for production of potable and technical ethanol. The genetic systems responsible for fermentation of maltose (*MAL* genes) and sucrose (β -fructosidase *SUC* genes) have been well studied [1–3].

Biochemical and genotypic analyses showed that there are two types of α -glucosidases in *S. cerevisiae*. The first type (maltase, EC 3.2.1.20) is responsible for hydrolysis and fermentation of α -1,4-glucosides (such as maltose and turanose), whereas the second one (isomaltase/ α -methylglucosidase, EC 3.2.1.10) catalyzes hydrolysis and fermentation of α -1,6-glucosides (such as α -methylglucoside and isomaltose) [4–10]. Both enzymes have also common substrates sucrose and *p*-nitrophenyl-D-glucopyranoside.

The yeast gene encoding isomaltase have not been known until recently. The authors of [11] cloned maltase and isomaltase genes from different *S. cerevisiae* strains. This work has played a major role in the isomaltase gene identification. The international project on sequencing and annotating the genome of the *S. cerevisiae* strain S288C [12] opened new vistas for studying α -glucosidase genes.

2 Results and discussion

We carried out computer analysis of *S. cerevisiae* S288C genome and discovered a novel family of isomaltase *IMA* genes. Along with the known maltase genes *MAL12* and *MAL32* [2, 13], we discovered *IMA1–IMA4* genes in this strain.

Glycoside hydrolases (or glycosidases) are a large group of enzymes (EC 3.2.1) that catalyze the O-

glycosidic bond cleavage. On the basis of amino acid sequence homology, their catalytic domains are grouped in the CAZy classification [14] into more than one hundred of families (GH1–GH118). Sequencing the genome of *S. cerevisiae* strain S288C showed that it encodes 46 glycosidases belonging to 17 families [12]. Enzymes with α -glucosidase activities were classified into two families, GH13 and GH31.

In the CAZy classification, part of proteins of the GH13 family were combined into 36 subfamilies (GH13_1–GH13_36). Analysis of some of the unclassified proteins of this family allowed us to distinguish 10 additional subfamilies [15]. Out of nine known *S. cerevisiae* S288C proteins containing GH13 family domains, two are included into subfamilies: amylo-(1,4 $\rightarrow$ 1,6)-transglucosidase (EC 2.4.1.18) belong to the GH13_8 subfamily (encoded by the YEL011w gene, chromosome V) and amylo- α -1,6-glucosidase (EC 3.2.1.33) is in the GH13_25 subfamily (gene YPR184w, chromosome XVI). The remaining seven proteins are evolutionarily very close to one another (see below); this fact allows us to consider them as representatives of a new subfamily (GL3C0220 according to the Génolevures database [16]). This subfamily includes maltases Mal12 (YGR292w, chromosome VII) and Mal32 (YBR299w/YBR2117, chromosome II), isomaltase YGR287c (chromosome VII), as well as four biochemically uncharacterized proteins YIL172c (chromosome IX), YJL216c (chromosome X), YJL221c (chromosome X), and YOL157c (chromosome XV). The GH31 family includes α -1,3-glucosidase (EC 3.2.1.84), whose gene YBR229c is located in chromosome II of *S. cerevisiae* S288C.

It should be noted that the comparative study of amino acid sequences of yeast maltases (from the *S. cerevisiae* strain ATCC56959) and isomaltase (from the *S. cerevisiae* strain ATCC56960) with α -glucosidases from other organisms [11] made it possible to reveal diagnostically significant amino acid residues (Val in isomaltases and Thr in maltases) in a conserved site. Site-directed mutagenesis experiments confirmed the importance of this Val residue as well as the downstream residues Gly and Ser for substrate specificity of yeast isomaltase. Replacement of these three amino acid residues with Thr, Ala, and Gly (similarly to natural maltase) made it possible to change the enzyme specificity [11].

To construct the phylogenetic tree of *S. cerevisiae* S288C α -glucosidases, we performed pairwise and multiple comparison of all 10 catalytic domains of GH13 and GH31 families of this organism. For this

purpose, we used standard methods of comparative analysis of amino acid sequences [17]. For domains encoded by genes YBR229c (GH31 family) and YPR184w (GH13_25 subfamily), we managed to obtain only short local alignments with the remaining eight domains. This is not striking because the GH13_25 subfamily is one of the most divergent (along with GH13_33), and it was proposed to be considered as an independent glycoside hydrolase family [15]. YIL172c and YJL221c proteins proved to be identical and were regarded as one protein. The remaining seven domains were used to construct multiple alignment and phylogenetic tree. Analysis also included the GH13 domain of Mal62 maltase of the *S. cerevisiae* ATCC56959 strain (isomaltase from *S. cerevisiae* ATCC56960 was identical to YGR287c and, for this reason, was not included into consideration). The GH13 domain, encoded by the YEL011w gene, was considered as outgroup. Phylogenetic analysis demonstrated the existence of two distinctly isolated α -glucosidase clusters (with a more than 99% bootstrap support). One of them combines three maltases; the other one, isomaltase and its three close paralogs. In the diagnostic site, all the three maltases comprise the Thr–Ala–Gly tripeptide and four proteins from the isomaltase cluster Val–Gly–Ser. The level of identity of amino acid sequences within these clusters is at least 99% and 92%, respectively, and proteins from different clusters have approximately 71% of identical amino acid residues. The results allow all the four proteins of the isomaltase cluster to be considered as isomaltases and the genes encoding them to be regarded as representatives of one family *IMA* (see Table). The YJL216c protein shares only 60–66% of identical amino acid residues with the representatives of both clusters and is outgroup with respect to them in the phylogenetic tree. However, this protein contains the tripeptide Val–Gly–Ser in the diagnostic site, which suggests that it most likely exhibits isomaltase rather than maltase activity.

Table. Identification of the novel family of isomaltase genes *IMA1–IMA4* in *S. cerevisiae* S288C

Open reading frame	Chromosome and telomere	Suggested gene name
YGR287c	VII, R	<i>IMA1</i>
YOL157c	XV, L	<i>IMA2</i>
YIL172c	IX, L	<i>IMA3</i>
YJL221c	X, L	<i>IMA4</i>
YJL216c	X, L	<i>IMA?</i>

It should be noted that the existence of two groups of strain S288C α -glucosidase genes differing in length was first reported in [13]. However, at that time, information on the differences in the substrate

specificity of enzymes encoded by these two gene groups was absent. Only the results of studies of the strain ATCC56960 isomaltase [11] made it possible to reveal a novel family of isomaltase genes *IMA1–IMA4* in the genome of *Saccharomyces cerevisiae* strain S288C.

References

- [1] R. Needleman. *Mol. Microbiol.* **5**(9):2079-2084, 1991.
- [2] G.I. Naumov, E.S. Naumova, and C.A. Michels. *Genetics*. **136**(3):803-812, 1994.
- [3] I.V. Korshunova, E.S. Naumova, and G.I. Naumov. *Molecular Biology* (Engl. Transl.) **39**(3):366-371, 2005.
- [4] S. Hestrin and C.C. Lindegren. *Arch. Biochem.* **29**(2):315-333, 1950.
- [5] G. Terui, H. Okada, and Y. Oshima. *Technol. Repts. Osaka Univ.* **9**:237-259, 1959.
- [6] N.A. Khan and N.R. Eaton. *Biochim. Biophys. Acta*. **146**(1):173-180, 1967.
- [7] H.-Y.L. Lai and B. Axelrod. *Biochim. Biophys. Acta*. **391**(1):121-128, 1975.
- [8] R.B. Needleman, H.J. Federoff, T.R. Eccleshall, B. Buchferer, and J. Marmur. *Biochemistry*. **17**(22):4657-4661, 1978.
- [9] L.L. Spielman and D.B. Mowshowitz. *Anal. Biochem.* **120**(1):66-70, 1982.
- [10] E. Kopetzki, P. Buckel, and G. Schumacher. *Yeast*. **5**(1):11-24, 1989.
- [11] K. Yamamoto, A. Nakayama, Y. Yamamoto, and S. Tabata. *Eur. J. Biochem.* **271**(16):3414-3420, 2004.
- a. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S.G. Oliver. *Science*. **274**(5287):546, 563-567, 1996.
- [12] G. Volckaert, M. Voet, and J. Robben. *Yeast*. **13**(3):251-259, 1997.
- [13] P.M. Coutinho and B. Henrissat. Carbohydrate-Active Enzymes server (CAZy), 2010. (<http://www.cazy.org/>)
- [14] D.I. Gizatullina and D.G. Naumoff. *Proceedings of the International Moscow Conference on Computational Molecular Biology (MCCMB'09)*. July 20-23, 2009. Moscow. Russia. P.249-250. (http://mccmb.belozersky.msu.ru/2009/MCCMB09_Proceedings.pdf)
- [15] The Génolevures database, 2010. (<http://www.genolevures.org/>)
- [16] D.G. Naumoff. *Zbio*. **1**(Art.3), 2006. (<http://zbio.net/bio/001/003.html>)

Sequence Analysis of the GH114 Family Glycoside Hydrolases

Daniil G. Naumoff¹ and Olga O. Stepuschenko²

¹ State Institute for Genetics and Selection of Industrial Microorganisms, Moscow, Russia

² Kazan State University, Department of Genetics, Kazan, Russia

daniil_naumoff@yahoo.com

Abstract

Phylogenetic analysis of the GH114 family of glycoside hydrolases has been performed. Subfamily structure and interfamily evolutionary connections of this family are studied.

1 Introduction

The endo- α -1,4-polygalactosaminidase (EC 3.2.1.109) is a very rare enzyme, which has been found only in two strains of bacteria (*Streptomyces griseus* C-10 [1] and *Pseudomonas* sp. 881 [2]). Its catalytic domain belongs to the GH114 family of glycoside hydrolases (or COG3868). According to the CAZy database [3], this family is represented in 64 proteins. COG3868 is closely related to functionally uncharacterized COG2342 [4]. The latter includes proteins with TIM-barrel type of the three-dimensional structure (PDB, 2AAM). This type of 3D structure has been shown for the catalytic domains of four clans (a hierarchical group higher than family) of glycoside hydrolases [3]. Relationships within the GH114 family are still unclear and became the purpose of this work, as well as its evolutionary connections with other protein families.

2 Methods and Algorithms

Protein sequences were retrieved from the NCBI database. Multiple sequence alignment of 125 GH114 domains was made in BioEdit program (very similar and partial sequences were omitted). The phylogenetic trees were built using programs of PHYLIP package. 34 members of COG2342 family were used as outgroup. Interfamily relationships were established using PSI Protein Classifier [5]. This program analyzes results of PSI-BLAST searches. Catalytic domains of several most divergent representatives from the GH114 and COG2342 families were used as a query. All PSI-BLAST searches were performed on May 9, 2010.

3 Results and discussion

We have revealed 140 non-identical protein sequences of GH114 domains using the blast algorithm. They include representatives of several phyla of Bacteria (Actinobacteria, Aquificae, Chloroflexi, Deferribacteres, Deinococcus, and Proteobacteria), as well as some Eukaryota (Alveolata, Ascomycota, Basidiomycota, Chlorophyta, and Oomycetes). The majority of the

proteins has similar length (251–375 amino acid residues) and contains only one domain (GH114). Protein from *Hahella chejuensis* (GenPept, ABC28688.1) consists of two GH114-domains; protein from *Endoriftia persephone* (GenPept, ZP_02533448.1), in addition to GH114, has a GH9-domain. Several proteins have additional N-terminal domain of unknown function.

Multiple sequence alignment of GH114 and COG2342 proteins allowed us to find two very conserved amino acid residues (Asp and Glu). They are located on C-terminus of the fourth and sixth β -strands in the predicted TIM-barrel. These two residues are homologous to the nucleophile and proton donor of GH31-glycosidases, suggesting their involvement into the active center of endo- α -1,4-polygalactosaminidases.

Three main clusters can be recognized on the GH114 phylogenetic trees (both Neighbor-Joining and Maximum Parsimony). Two of them have high bootstrap support and include solely representatives of Actinobacteria. The third cluster is formed by proteins from Ascomycota, which are not very well separated from some bacterial proteins. The tree topology supports the important role of horizontal transfer in the evolution of GH114 proteins.

We used five most divergent representatives of the GH114 family (see Table) for iterative screening of the protein database by PSI-BLAST. Totally, we retrieved 12048 non-identical proteins during first 9 iterations. The majority of them contains domains of the following families of glycoside hydrolases: GH13 (clan GH-H), GH18 (clan GH-K), GH20 (clan GH-K), GH27 (clan GH-D), GH29, GH31 (clan GH-D), GH36 (clan GH-D), GH66, and GH114. Some obtained proteins belong to functionally uncharacterized families COG1306, COG1649, COG2342, and GHL3, as well as to several new families. Using COG2342 representative (GenPept, EEP04836.1) as a query, allowed us to obtain 7519 non-identical proteins during first 6 iterations. Among them we found members of the following families: GH13, GH31, GH36, GH66, GH114, COG1306, COG1649, COG2342, and GHL3. The majority of the mentioned families includes proteins with the experimentally determined TIM-barrel type of the three-dimensional structure. So, the data obtained support the common evolution origin of all TIM-barrel type glycoside hydrolase catalytic domains, as we have suggested earlier [6].

Table. Interfamily relationship of GH114 and COG2342 domains revealed by PSI-BLAST searches.

Family	BAG23712.1 (GH114)	EAA59176.1 (GH114)	EDK40790.2 (GH114)	EDP52951.1 (GH114)	EDX26248.1 (GH114)	EEP04836.1 (COG2342)
GH13	—	5	6	7	7	3
GH18	—	—	—	3	3	—
GH20	—	—	—	9	—	—
GH27	—	8	—	8	8	—
GH29	—	—	—	9	9	—
GH31	—	4	6	4	4	6
GH36A	—	7	9	8	7	—
GH36B	—	6	7	5	6	—
GH36E	—	—	—	—	9	—
GH36F	—	7	9	8	8	—
GH36G	—	7	8	7	7	—
GH36H	—	7	8	7	5	4
GH36J	—	9	—	9	9	—
GH36K	—	—	—	8	9	—
GH66	—	8	9	9	5	2
GH114	1	1	1	1	1	1
COG1306	—	3	3	4	8	3
COG1649	—	5	8	6	5	3
COG2342	2	1	1	1	1	1
GHL3	—	—	—	9	9	5
others	—	3	4	3	3	3

Notes. Each column corresponds to a query used for the PSI-BLAST search. The minimal number of iterations, needed to reach a family member with *E*-values below 0.005, is indicated. Former family GH36 is considered now [5, 6] as a group of several distinct families (GH36A–GH36K). COG1306 [5], COG1649 [5, 6], COG2342 [7], and GHL3 [8] are families of predicted glycoside hydrolases.

References

- [1] J.L. Reissig, W.-H. Lai, and J.E. Glasgow. An endogalactosaminidase from *Streptomyces griseus*. *Can. J. Biochem.* **53**(12):1237-1249, 1975.
- [2] J. Tamura, H. Takagi, and K. Kadowaki. Purification and some properties of the endo α -1,4 polygalactosaminidase from *Pseudomonas* sp. *Agric. Biol. Chem.* **52**(10):2475-2484, 1988.
- [3] P.M. Coutinho and B. Henrissat. Carbohydrate-Active Enzymes server (CAZy), 2010. (<http://www.cazy.org/>)
- [4] L.M. Iyer, L. Aravind, P. Bork, K. Hofmann, A.R. Mushegian, I.B. Zhulin, and E.V. Koonin. *Quoderat demonstrandum?* The mystery of experimental validation of apparently erroneous computational analyses of protein sequences. *Genome Biol.* **2**(12):research0051.1-0051.11, 2001.
- [5] D.G. Naumoff and M. Carreras. New program PSI Protein Classifier automatizes the PSI-BLAST results analysis. *Molecular Biology* (Engl. Transl.) **43**(4):652-664, 2009.
- [6] D.G. Naumoff. Development of a hierarchical classification of the TIM-barrel type glycoside hydrolases. *Proceedings of the Fifth International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2006)*. July 16-22, 2006. Novosibirsk. Russia. V.1. P.294-298. (http://www.bionet.nsc.ru/meeting/bgrs2006/BGRS_2006_V1.pdf)
- [7] D.I. Gizatullina and D.G. Naumoff. Reclassification of GH13 family of glycoside hydrolases. *Proceedings of the International Moscow Conference on Computational Molecular Biology (MCCMB'09)*. July 20-23, 2009. Moscow. Russia. P.249-250. (http://mccmb.belozersky.msu.ru/2009/MCCMB09_Proceedings.pdf)
- [8] D.G. Naumoff. GH101 family of glycoside hydrolases: subfamily structure and evolutionary connections with other families. *J. Bioinform. Comput. Biol.* **8**(3), 2010 (in press)

Sequence-Based Classification of Yeast Glycoside Hydrolases: CAZy vs. Génolevures

Daniil G. Naumoff

State Institute for Genetics and Selection of Industrial Microorganisms, Moscow, Russia
daniil_naumoff@yahoo.com

Abstract

Correspondence between CAZy and Génolevures families of glycoside hydrolases has been established. Phylogenetic analysis of the proteins allowed us to distinguish groups of orthologues in the families.

1 Introduction

On the basis of sequence similarity, catalytic domains of glycoside hydrolases (EC 3.2.1.-) have been grouped into more than 100 families (GH1–GH118, except GH21, GH40, GH41, GH60, GH69, and GH91) in the CAZy classification [1]. Various ascomycetous yeasts have about 40–50 glycosidase genes, in total representing only 32 CAZy families. The Génolevures database [2] provides classification of all proteins encoded in 9 species of hemiascomycete yeasts into about 8000 families. The purpose of the work is to clarify the correspondence between CAZy and Génolevures families and to analyze phylogeny of yeast glycoside hydrolases.

2 Methods and Algorithms

Multiple sequence alignments of glycoside hydrolases from 9 hemiascomycete yeast species were retrieved from the Génolevures site. The alignments were manually edited in BioEdit and representatives (if any) from *Schizosaccharomyces pombe* were added as outgroup. The phylogenetic trees were built using programs of PHYLIP package.

3 Results and discussion

3.1 Family corresponding: CAZy vs. Génolevures

Among the 32 CAZy families analyzed, eight are not presented in Génolevures genomes (GH10, GH45, GH71, and GH115) or represented by a single protein, all in *Debaryomyces hansenii* (GH20, GH51, GH105, and GH114). Among the remaining 24 families, seven do not include *S. pombe* members (see Table).

For 18 families we found one-to-one family correspondence. CAZy families GH15, GH17, GH27, GH28, GH32, GH37, GH38, GH47, GH63, GH65, GH72, GH78, GH81, GH85, and GH92 correspond to GL3R1486, GL3C0057, GL3R4232, GL3R3998, GL3R0923, GL3R0613, GL3R2870, GL3C0154, GL3R1048, GL3R2013, GL3R0042, GL3R4175,

GL3C0407, GL3R4002, and GL3R3768, respectively, in the Génolevures database.

Families GH2, GH31, and GH88 correspond to GL3R4262, GL3C0670, and GL3R3758, but each of the former three also includes a singleton (unclassified) protein from *D. hansenii* (GL3U4585.DEHA2E09504p, GL3U4586.DEHA2D03190p, and GL3U4585.DEHA2D00946p, respectively). GH5 includes members of GL3C0106 and GL3R0891 families, as well as a singleton from *Lachancea (Saccharomyces) kluyveri* (GL3U4585.SAKL0D15202p). GH3, GH16, GH18, and GH76 correspond to pairs of Génolevures families: GL3R0575 & GL3R4191, GL3C0130 & GL3R0461, GL3C0583 & GL3C0738, and GL3R0174 & GL3R3972, respectively. Families GL3C0220, GL3R2291, and GL3R2899 compose GH13.

3.2 Phylogenetic analysis

Phylogenetic analysis allowed us to distinguish five subclusters of hemiascomycete proteins in GH72 (GL3R0042) family. Two *S. pombe* proteins compose a sister group for one of the clusters, but the other two *S. pombe* proteins lie clearly out of the clusters. The data obtained suggest that the common ancestor of all hemiascomycete yeasts had five GH72 proteins and the corresponding duplications had happened at different stages of the evolution.

Family GH13 is one of the largest families of glycoside hydrolases. It is divided into 36 main subfamilies (GH13_1–GH13_36) in the CAZy database [1]. Subfamilies GH13_25 and GH13_33 are very divergent and we recently suggested to consider them as separate families of glycoside hydrolases [3]. Subfamilies GH13_8 and GH13_25 correspond to families GL3R2291 and GL3R2899, respectively, in the Génolevures database. Each of them is represented in the hemiascomycete yeasts by only one paralogue and is absent in *S. pombe*. Proteins of family GL3C0220 do not belong to any subfamily in the CAZy classification. Some yeast species have several paralogues but the others do not have them at all. It is impossible to distinguish groups of orthologues in this family. The majority of paralogues are a result of several recent linear-specific duplications.

The data obtained show that the hemiascomycete proteins of GH16 family represent four groups of orthologues. Three of them belong to GL3C0130 family and the fourth corresponds to GL3R0461 family. The latter includes two *S. pombe* proteins. Sequence similarity of GL3C0130 and GL3R0461 proteins is so

low, that it was impossible to prepare a common multiple sequence alignment. It looks reasonable to consider them as two separate (sub)families. The third *S. pombe* protein can be considered as a representative of an additional subfamily in GH16 family.

The hemiascomycete proteins of GH76 family can be divided into three groups of orthologues. One of them corresponds to GL3R3972 family and includes one protein from *S. pombe*. GL3R0174 family combine the other two groups, as well as, three *S. pombe* proteins, which compose a separate cluster on the phylogenetic tree.

We have distinguished five groups of orthologues in GH5 family. One of them corresponds to GL3R0891 family. The remaining four groups include members of GL3C0106 family, as well as

GL3U4585.SAKL0D15202p singleton. One of the latter groups includes a *S. pombe* protein. The other two *S. pombe* proteins also belong to GL3C0106 but lie out of the hemiascomycete groups on the phylogenetic tree.

Five, three, and two groups of orthologues can be distinguished in families GL3C0057, GL3C0154, and GL3C0407, respectively. The other Génolevures families of glycoside hydrolases are composed by only one group of orthologues each.

Our results demonstrate a high level of evolutionary plasticity of the carbohydrate utilization system in yeasts. Gene loss, as well as, multiple gene duplications have happen during the evolution of several yeast glycoside hydrolase families. These events can be only partly explained by total genome duplications.

Table. Number of glycoside hydrolase genes in yeast genomes (according to the CAZy database)

Family	SACE	CAGL	ZYRO	SAKL*	KLTH	KLLA	ERGO	DEHA	YALI	SAPO
GH2	0	0	0	0	0	1	0	2	0	0
GH3	0	0	0	1	0	3	0	4	6	1
GH5	5	5	4	5	3	3	4	6	4	3
GH13	9	2	2	7	6	4	2	4	2	12
GH15	1	1	1	1	1	1	1	1	1	2
GH16	5	5	5	4	4	4	3	9	4	3
GH17	4	5	3	4	4	4	2	5	6	1
GH18	2	2	2	4	2	2	1	5	3	1
GH20	0	0	0	0	0	0	0	1	0	0
GH27	0	0	0	1	0	0	0	1	0	1
GH28	1	0	0	1	0	0	1	0	0	0
GH31	1	1	2	1	1	1	1	3	1	4
GH32	1	0	1	3	1	1	1	1	1	2
GH37	2	2	1	1	1	1	1	1	1	1
GH38	1	1	1	1	1	0	1	1	1	1
GH47	3	3	3	3	3	3	3	4	3	2
GH51	0	0	0	0	0	0	0	1	0	0
GH63	1	1	1	1	1	1	1	1	1	1
GH65	1	1	1	1	1	1	1	1	1	0
GH72	5	6	5	6	5	6	6	4	4	4
GH76	2	2	2	2	2	2	5	3	6	4
GH78	0	0	0	0	0	0	0	1	1	0
GH81	2	1	1	2	2	2	2	2	3	2
GH85	0	0	1	1	0	0	1	0	0	0
GH88	0	0	0	0	0	0	0	5	0	0
GH92	0	0	0	1	1	1	1	1	0	0
GH105	0	0	0	0	0	0	0	1	0	0
GH114	0	0	0	0	0	0	0	1	0	0

References

- [1] P.M. Coutinho and B. Henrissat. Carbohydrate-Active Enzymes server (CAZy), 2010. (<http://www.cazy.org/>)
- [2] The Génolevures database, 2010. (<http://www.genolevures.org/>)
- [3] D.I. Gizatullina and D.G. Naumoff. Reclassification of GH13 family of glycoside hydrolases. *Proceedings of the International Moscow Conference on Computational Molecular Biology (MCCMB'09)*. July 20-23, 2009. Moscow, Russia. P.249-250. (http://mccmb.belozersky.msu.ru/2009/MCCMB09_Proceedings.pdf)

HETEROGENE – база достоверных избыточных данных о периодичности и неоднородности в геномах организмов

Назипова Н., Теплухина Е., Тюльбашева Г., Чалей М.

Институт математических проблем биологии, Пушкино, Россия

nnn@impb.ru, maramaria@yandex.ru

Абстракт

На основе спектрально-статистического подхода к выявлению достоверной и значимой неоднородности в последовательностях ДНК была создана технология поиска сильно размытых tandemных повторов в геноме. Результаты применения этой технологии к последовательностям геномов модельных организмов *A. thaliana*, *C. elegans*, *D. melanogaster* и *S. cerevisiae* представлены в базе данных HETEROGENE http://www.jcvi.ru/lp_base/. База данных HETEROGENE является достоверным и избыточным информационным ресурсом по скрытой периодичности в геномах организмов.

1 Введение

В геномах организмов выделяют две категории повторов: дисперсные повторы – транспозоны и ретротранспозоны, представители многочисленных семейств которых, случайным образом распределённые по геному, составляют до 50% его длины, и tandemные повторы – массивы последовательно повторяющихся копий некоторого исходного фрагмента последовательности ДНК (паттерна). Повреждение копий паттерна при замене исходных нуклеотидов, а также вставок и делеций, как единичных, так и нескольких нуклеотидов, ведёт к образованию размытых tandemных повторов. Размытые tandemные повторы, повреждения копий паттерна в которых ограничены только заменами нуклеотидов (нукл.), принято называть нечеткими tandemными повторами. Размытые tandemные повторы являются участками скрытой периодичности в геноме.

Применение спектрально-статистического подхода [1, 2], позволило создать комплекс программ [3], выявляющих статистически значимые размытые tandemные повторы. Теоретический предельный уровень дивергенции копий паттерна в находимых tandemных повторах составляет 50%. Комплекс программ универсален, он одинаково хорошо работает как с микро- и минисателлитами с длиной паттерна от 2 до 100, так и периодичностями с паттерном длиной до 2000 нукл.

Поскольку с алгоритмической точки зрения выявление скрытых периодичностей является очень трудной задачей, если период априори неизвестен, то для поиска сильно размытой периодичности

(максимальный уровень дивергенции 50%) был выбран метод поиска высоко значимой (на уровне $b=0,000001$) неоднородности. Для подтверждения значимости размытого tandemного повтора, содержащего небольшое количество копий (случай недостаточного статистического материала) в программном комплексе реализован специальный критерий [1]. В результате в базу данных HETEROGENE вошли только значимые, достоверные районы неоднородности, в более чем 90% случаев представленные размытыми tandemными повторами.

2 Методы и алгоритмы

Выявление значимой неоднородности осуществляется при многократном сканировании последовательности ДНК перекрывающимися окнами различного размера. Поэтому общая стратегия поиска размытых tandemных повторов напоминает shotgun-стратегию секвенирования геномов, когда сначала анализируют относительно короткие и перекрывающиеся фрагменты и затем производят их компьютерную сборку в более протяженные участки.

Для избыточного и наглядного представления результатов в базе HETEROGENE была разработана концепция логической записи как связанной группы перекрывающихся районов значимой неоднородности. Район наибольшей длины является представителем группы; участки, лежащие внутри него, рассматриваются как внутренние неоднородности. Последние либо указывают на наиболее структурированные участки района неоднородности, либо, если имеют выявленные периоды, отличающиеся от периода представителя группы, способствуют правильному прочтению данных (см. Рис. 1). В последнем случае, проанализировав состав группы, пользователь может скорректировать её размеры или рассматривать такую группу как составную. Визуальное представление последовательности ДНК в базе HETEROGENE помогает пользователю сделать окончательный вывод о том, можно ли рассматриваемый район неоднородности считать размытым tandemным повтором (т.е. можно ли вывести консенсус-паттерн повтора).

HETEROGENE DATABASE

Organism: *Arabidopsis thaliana* Chromosome: *IV*

Location	Length	Period	Exponent	Preservation Level	H-spectrum value
3239317 - 3239812	496	7	70.86	0.58	4.15

[Show spectra](#) [Show sequence](#) [Sequence Viewer](#)

INTRINSIC HETEROGENEITIES:

Location	Length	Period	Exponent	Preservation Level	H-spectrum value			
3239491 - 3239622	132	7	18.86	0.87	3.82	Show spectra	Show sequence	Sequence Viewer
3239701 - 3239760	60	7	8.57	0.9	1.95	Show spectra	Show sequence	Sequence Viewer

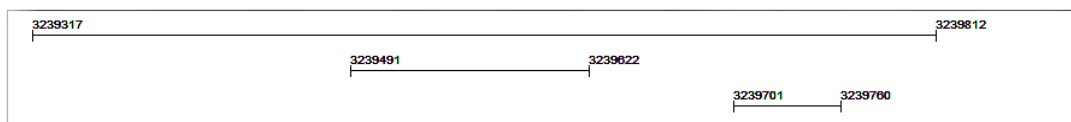


Рис. 1. Пример двухуровневой структуры записи в базе данных HETEROGENE. В верхней таблице приведены параметры последовательности, представляющей группу на неизбыточном уровне. В нижней таблице — данные об остальных последовательностях группы. Внизу располагается масштабное схематическое изображение всех последовательностей группы.

HETEROGENE является реляционной базой данных под управлением СУБД MySQL. Реализована поисковая система по всем полям базы с возможностью сортировки полученных результатов. Также предусмотрена возможность сохранения результатов запроса в виде текстового файла. В соответствии с двухуровневой структурой записей в базе (см. Рис. 1), интерфейс пользователя предлагает выбор уровня поискового запроса: неизбыточный (nonredundant) уровень, когда поиск выполняется по подмножеству представителей групп; или простой (simple) уровень, когда поиск ведется по всем записям базы данных. Для каждой последовательности в базе HETEROGENE в отдельных окнах организован: 1) просмотр спектров [1, 2], на основании которых была выведена оценка размера паттерна периодичности, и 2) просмотр последовательности в виде столбца её сегментов с длиной строки, равной оценке размера паттерна. Пользователь имеет возможность изменять длину сегментов последовательности, чтобы уточнить оценку размера паттерна в размытом тандемном повторе. С помощью графического интерфейса Sequence Viewer [4] пользователь может получить информацию о локализации и функциональном контексте рассматриваемого района ДНК на хромосоме.

3 Результаты и обсуждение

Результаты применения созданной вычислительной технологии к последовательностям геномов модельных организмов *A. thaliana*, *C. elegans*, *D. melanogaster* и *S. cerevisiae* представлены в базе

данных HETEROGENE. Сравнительный анализ данных для этих геномов из базы TRDB [5] с соответствующими данными базы HETEROGENE показал, что HETEROGENE не только содержит практически все данные, представленные в TRDB, но и существенно дополняет их данными о сильно размытых тандемных повторах и регулярных неоднородностях, описание которых является предметом будущих исследований.

Созданная база данных HETEROGENE имеет как практическое применение в качестве ресурса генетических маркеров (в основном микро- и минисателлитов), так и общее биологическое значение для дальнейших исследований скрытой периодичности в геноме.

Работа выполнена при частичной поддержке РФФИ, проект 09-07-00455.

References

- [1] Chaley M.B., Nazipova N.N., Kutyrkin V.A. *Pattern Recognition and Image Analysis*, 2009, **19**, 358-367.
- [2] Chaley M., Kutyrkin V. (2008) *Mathematical Biosciences*, **211**(1), 186-204.
- [3] Назипова Н., Теплухина Е., Тюльбашева Г., Чалей М. (2008) В: Труды II Международной конференции «Математическая биология и биоинформатика». С. 103-104. Пущино, Россия.
- [4] Sequence Viewer 2010, URL: <http://www.ncbi.nlm.nih.gov/projects/sviewer/>
- [5] Gelfand, Y., Rodriguez, A. and Benson, G. *Nucleic Acids Res.*, 2007, **35**, 80-87.

Определение консенсусной последовательности для участков скрытой периодичности геномов

Назипова Н.Н., Зимин Ф.А., Теплухина Е.И., Устинин М.Н.

Институт математических проблем биологии, Пушкино, Россия

nnn@impb.ru

1 Введение

Геном представляется символьной последовательностью, составленной из четырех видов букв, которые, на первый взгляд, разбросаны в случайном порядке. Однако некоторые участки геномов обладают выраженной периодичностью. Такие периодические участки называются тандемными (последовательно расположенными один за другим) повторами.

Тандемный повтор характеризуется длиной (длина образца) и кратностью. Каждая копия образца (и сам образец) независимо подвергается дополнительным мутациям (заменам, вставкам, удалениям отдельных элементов последовательности, новым тандемным дупликациям внутренних фрагментов и др.). Через определенное время каждая копия повтора сильно дивергирует, точный тандемный повтор становится *размытым*.

Под скрытой периодичностью в нуклеотидных последовательностях понимают размытые тандемные повторы, в частности, нечеткие тандемные повторы, не поврежденные вставками и делециями. В данной работе предполагается, что проблема поиска размытых тандемных повторов при неизвестных длине периода и кратности решена. Эта задача решается с помощью спектрально-статистического подхода к распознаванию неоднородностей в текстах ДНК [1, 2]. Создана вычислительная технология [3] нахождения участков скрытой периодичности в полных геномах. Обнаруженные участки помещаются в базу данных HETEROGENE [4]. Цель данной работы состоит в определении консенсусной последовательности для каждого участка скрытой периодичности из этой базы данных.

Консенсусная последовательность является дополнительной характеристикой тандемного размытого повтора, которая позволяет сравнивать участки периодичности между собой как внутри одного генома, так и между геномами. Для изучения найденных периодичностей необходимо группировать тандемные повторы в семейства, которые определяются на основе текстуального сходства консенсусов.

Задача нахождения консенсуса заключается в представлении участка периодичности в виде выравнивания отдельных копий тандемного повтора и последующего выявления преобладающего символа в каждом столбце выравнивания. Специфика поставленной задачи состоит в том, что найденное спектрально-статистическим методом значение длины паттерна периодичности L - это приблизительная оценка; некоторые копии образца могли укоротиться, некоторые, наоборот, удлиниться. Необходимо представить дивергировавшие тандемные повторы в виде, который давал бы наглядную картину их возможной эволюции (показать, где были возможные вставки/выпадения и замены символов).

2 Методы и алгоритмы

Для решения поставленной задачи была реализована итерационная процедура, в основу которой были положены существенные характеристики участка периодичности, использованные в спектрально-статистическом подходе [1, 2]. Этот подход, благодаря оригинальному NEP-критерию [1, 5], позволяет получать достоверные результаты в условиях недостаточного статистического материала. Поиск периодических участков реализован в виде многократного сканирования генома с использованием техники «скользящего окна». Сканирование каждого генома повторяется с учетом различных уровней дивергенции тандемных повторов, которые можно характеризовать значением величины pl ($pl \in [0,1]$), называемой *уровнем сохранности* буквы.

Метод подсчета уровня сохранности pl состоит в следующем. Анализируемая строка длины n , состоящая из букв алфавита $A = \{a_1, \dots, a_K\}$, разбивается на подстроки длины L , которые записываются друг под другом. Количество подстрок $R_L = \frac{n}{L}$ называется *кратностью*, а массив подстрок называется *L-профилем*.

L -профиль позволяет вычислить частоту π_j^i встречаемости i -той буквы алфавита A в j -той позиции (j -том столбце) профиля для $i = 1, \dots, K$ и $j = 1, \dots, L$. По матрице частот $\pi = (\pi_j^i)_L^K$ вычисляется

значение уровня сохранности паттерна pl :

$$pl = \frac{1}{L} \sum_{j=1}^L \max\{\pi_j^i : i \in 1, \dots, K\}$$

и спектрально-статистического параметра HL :

$$HL = \frac{R_L}{\chi^2_{krit}(\alpha, (K-1)(L-1))} \sum_{j=1}^L \sum_{i=1}^K \frac{(\pi_j^i - p^i)^2}{p^i(1-p^i)}, \text{ здесь } \alpha = 0.0000001, p^i - \text{ частота встречаемости } i\text{-той буквы алфавита } A \text{ во всей анализируемой строке.}$$

Описание итерационной процедуры.

Шаг 0: Дана последовательность $S = s_1 s_2 \dots s_L \dots s_{2L} \dots s_{3L} \dots s_{LSeq-1} s_{LSeq}$ длины $LSeq$ с оценкой длины периода L , значениями уровня сохранности образца и спектрально-статистического параметра соответственно $pl = pl_0$ и $HL = HL_0$.

$$s_1 s_2 \dots s_L \dots s_{2L} \dots s_{3L} \dots s_{LSeq-1} s_{LSeq} \quad pl, HL$$

Шаг 1: Вставляем поочередно в каждую позицию последовательности пробел и подсчитываем значения сохранности образца pl и спектрально-статистического параметра HL . Из всех $LSeq$ возможных позиций размещения пробела выбираем ту, вставка в которую дает максимальные приращения для обеих переменных pl и HL :

$$_s_1 s_2 \dots s_L \dots s_{2L} \dots s_{3L} \dots s_{LSeq-1} s_{LSeq} \quad (pl^1, HL^1)$$

$$s_1 _s_2 \dots s_L \dots s_{2L} \dots s_{3L} \dots s_{LSeq-1} s_{LSeq} \quad (pl^2, HL^2)$$

$$s_1 s_2 \dots s_L \dots s_{2L} \dots s_{3L} \dots s_{LSeq-1} s_{LSeq} \quad (pl^3, HL^3)$$

...

$$s_1 s_2 \dots s_L \dots s_{2L} \dots s_{3L} \dots s_{LSeq-1} s_{LSeq} \quad (pl^{LSeq}, HL^{LSeq})$$

$$pl_i = \max(pl^{i \max} : pl^{i \max} \geq pl, HL^{i \max} > HL),$$

$$HL_i = \max(HL^{i \max} : pl^{i \max} \geq pl, HL^{i \max} > HL),$$

$$i \max = 1, 2, \dots, LSeq$$

Если приращения на этом шаге нет ($pl_i < pl$ или $HL_i \leq HL$), процедура заканчивается. В противном случае в позицию с номером $i \max$, доставляющую максимумы приращений обоим функциям pl и HL на данном шаге, записываем пробел, раздвинув предварительно последовательность S :

$$S = s_1 s_2 \dots s_{i \max-1} s_{i \max} \dots s_{LSeq-1} s_{LSeq}$$

Теперь мы получаем строку длиной $LSeq = LSeq+1$ с длиной периода L , значениями уровня сохранности образца и спектрально-статистического параметра соответственно $pl = pl_i$ и $HL = HL_i$. Снова возвращаемся на шаг 1.

В случае, когда на шаге 1 появляется несколько позиций с одинаковыми максимальными значениями уровня сохранности и спектрально-статистического параметра, отрабатываются все варианты дальнейшего выравнивания с тем, чтобы

найти глобальный максимум характеристик pl и HL для данной строки.

3 Результаты и обсуждение

Разработан подход к построению выравниваний периодических участков, сильно изломанных вставками и выпадениями позиций. Благодаря тому, что реализовано промежуточное запоминание всех вариантов строк, которые могут вести к глобальному максимуму pl и HL , мы проводим обход всего множества решений, находя из них лучшее. Теоретически оно может быть не единственным.

В программе есть режим коррекции величины L . Итерационная процедура повторяется в цикле для последовательных значений длины периода, начиная с L и кончая значением $L_{\max}$, которое определяется в зависимости от того, к какому классу периодичности (микро-, минисателлит, сателлит) принадлежит обрабатываемый участок. Выбирается то значение длины, которое обеспечивает максимальное увеличение значений pl и HL . Так мы можем скорректировать длину периода, если ошибка в оценке длины паттерна была вызвана вариативностью длин копий паттерна.

Этот режим помогает также корректировать оценку длины, связанную со свойством кратности периодичности. Это свойство состоит в том, что, если есть периодичность, например, длиной 3, то есть периодичность длиной в 6, 9, 18 и т. д. В спектрально-статистическом методе заложен поиск наименьшей длины из ряда кратных длин, обеспечивающей удовлетворительные с точки зрения метода значения pl и HL . Но иногда при малых значениях кратности повторов метод может ошибаться с выбором длины периода.

Работа выполнена при частичной поддержке РФФИ, проекты 09-07-00455, 10-07-00300 и 09-07-12108.

References

- [1] М.Б. Чалей, Н.Н. Назипова, В.А. Кутыркин. *Математическая биология и биоинформатика*, 2007, том 2, №1, с. 20-35, [http://www.matbio.org/downloads/Chaley2007\(2_20\).pdf](http://www.matbio.org/downloads/Chaley2007(2_20).pdf)
- [2] Chaley M.B., Nazipova N.N., Kutyrykin V.A. *Pattern Recognition and Image Analysis*. 2009, Vol. 19, No. 2, pp. 358-367
- [3] Назипова Н., Теплухина Е., Тюльбашева Г., Чалей М. (2008) Труды II Международной конференции «Математическая биология и биоинформатика». С. 103-104. Пущино, Россия.
- [4] HETEROGENE: The Database of latent periodicity regions in genomes of model organisms, URL: http://www.jcvi.ru/lp_base
- [5] Chaley M., Kutyrykin V. *Mathematical Biosciences*, 2008, Vol. 211(1), pp 186-204.

О выделении структурных единиц в белковых последовательностях на основе критерия целостно-завершенного текста

Гуменюк А.С., Морозенко Е.В., Улаев В.А., Антосюк Е.В.

Омский государственный технический университет

morozenkoev@rambler.ru

Abstract

Any analysis on the basis of traditional scientific approach (including the non-formal expert approach) is impossible unless an object with the complex structure is shown in the form of certain primary elements.. Protein is a chain formed by the 'alphabet' of twenty amino acids. In fact, there was a description of a long molecular chain in the form of characters sequence, which can be called a genetic pseudo-text, as the 'words' in this chain are yet to be defined, and the text to be written down.

Существенный прорыв в исследовании структуры текста произошел в семидесятых годах XX века, когда советский ученый-кибернетик Ю.К. Орлов с помощью средств вероятностно-статистического и энтропийного подходов открыл феномен целостности художественного литературного и музыкального произведений, который проявлялся в хорошем совпадении реального частотно-рангового распределения слов завершенного текста с законом Ципфа-Мандельброта, представляющего гиперболическую зависимость частоты вхождения слова p_i от его ранга i [1].

Для фрагментов и конгломератов текстов такого совпадения не наблюдалось. Таким образом, было количественно установлено, что завершенный текст имеет особую структуру, в которой его словарь v , длина Z , число вхождения самого частого слова F_1 и числа вхождений одинаковых слов p_i связаны законом Ципфа-Мандельброта. Однако в работах Ю. Орлова почти не затрагивалась проблема однозначной сегментации текста на слова.

На практике, при отделении и различении элементов текста естественных языков друг от друга возникают известные трудности, в особенности, когда эту работу требуется полностью формализовать и автоматизировать. То есть, различение слов, даже отделенных друг от друга в тексте пробелами, не удается осуществлять в полностью автоматическом режиме.

Набор признаков, на основании которых производилась формальная идентификация «завершенного» текста, назван по имени ученого,

открывшего феномен целостности (завершенности) художественного произведения, критерием Орлова [2]. В состав данного критерия авторы включили следующие по приоритету факторы:

- точность совпадения теоретического v_T и фактического v объемов алфавита элементов;
- степень совпадения фактического рангового распределения с законом Мандельброта.

Важнейшим результатом исследований Ю. Орлова было открытие феномена целостно-завершенного текста, формально проявляющегося в том, что состав его элементарных компонентов (слов и т.п.) является весьма специфичным (описывается законом Ципфа-Мандельброта). На наш взгляд, открытие Орлова косвенно характеризует особую структуру (построение, взаимное расположение компонентов завершенного текста). Таким образом, если Ю. Орлов, считая очевидными элементы текста, устанавливает его завершенность, то в наших разработках была поставлена обратная задача. Полагая наблюдаемые знаковые цепи целостно-завершенными, выделить в них элементарные составляющие (компоненты), распределение которых удовлетворяет критерию Ю. Орлова.

Целью данной работы является представить модификацию известного алгоритма [3] формирования структурных единиц знаковой последовательности или ее сегментация на «слова». Данный алгоритм, называемый далее «базовым», считает выбранное испытываемое псевдослово (короткая подпоследовательность) словом такое, которое имеет неожиданно большую или малую частоту появления в тексте, в сравнении с его расчетной частотой, определенной на основе модели текста как марковской цепи. Степень различия фактической и расчетной частот задается некоторым пороговым значением. Представляемая здесь существенная модификация базового алгоритма, использовалась нами ранее для выделения склеек из фонов – элементарных акустических единиц в фонетическом представлении поэтического текста, названных созвучиями [2].

Эта модификация дополняет базовый алгоритм следующими операциями. После первой сегментации исходной знаковой цепи с помощью базового алгоритма определяется статистическое

частотно-ранговое распределение первой версии словаря, которое затем сравнивается с теоретическим распределением Ципфа-Мандельброта. Если расхождение этих распределений превышает некоторый установленный порог, то изменяется пороговое значение для выделения слов и базовый алгоритм повторно сегментирует исходную последовательность. Такая процедура сравнения теоретического и фактического распределений потенциальных словарей с последующим изменением порога для выделения слов сопровождается неоднократной сегментацией исходной знаковой цепи вплоть до установленной степени совпадения распределений. Таким образом, базовый алгоритм в структуре рассмотренной модификации является ядром во внешнем цикле, который используется для выбора наилучшего разбиения по критерию целостно-завершенного текста Ю. Орлова.

В результате эксперимента по сегментации аминокислотной последовательности **threonine protein kinase BRAF [Takifugu rubripes]** с помощью модификации базового алгоритма удалось добиться совпадения мощностей фактического и теоретического алфавитов $v = v_T = 128$ и хорошего соответствия фактического и расчетного распределений, при этом длина биологического текста $Z = 501$, число вхождений самого частого слова $F_1 = 49$, (частота $p_1 = 0,098$). Характерно, что алгоритм начал склеивать аминокислоты в группы по три и меньше.

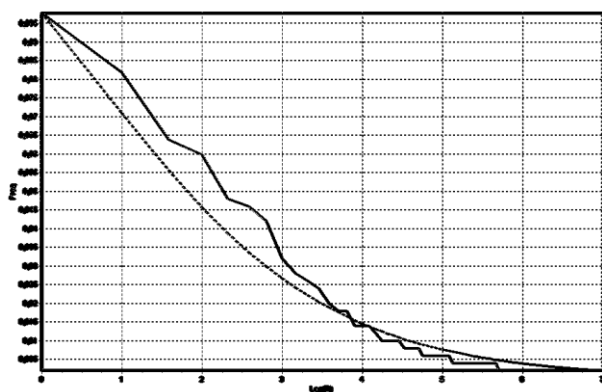


Рис. 1. Фактическое и расчетное частотно-ранговые распределения «естественных» элементов белка (аминокислотных слов) **serine/threonine protein kinase BRAF [Takifugu rubripes]**. По оси абсцисс ранг представлен в логарифмическом масштабе.

Из факта совпадения мощностей фактического и расчетного словарей данного белка и рисунка, представляющего распределения, видно, что разбиение на элементы, по критерию Орлова имеет незначительные расхождения с законом Ципфа-Мандельброта.

Знаковая цепь, отображающая последовательность белка и разбитого на аминокислотные слова в соответствии с критерием Орлова, представлена ниже. В дальнейшем с помощью этого аппарата предполагается провести исследования особенностей построения белковых последовательностей, с помощью разрабатываемых авторами средств формального анализа строя цепей [4].

M A A L S S A E S P P P V L N G D A A E R D P G R E R G
L E E L D S G F N S A C T T R V P G G A Q N E E I W N I
K Q M I K L T Q E H L E A L L D K F G G E H N P P T I Y L E A
Y E E Y T S K L D A L Q Q R E Q Q L L E A M G N G T D
F P C S P S P M P A L L E V K M G G C V P G V G A Q A P N
S L A V L Q T P T D G S R V N P R S P Q K P I V R V F L
P N K Q R T V V S A R C G M T V R D S L K K A L T M R G L
I P E C C A V Y R M Q D G E K K P I G W D T D I S W L T L E
E L H V E V L E N V P L T T H N F V R K T F F T L A F C D F
C R K L L F Q G F R C Q T C G Y K F H Q R C S T E V P L M C
V N Y D Q L D L L L V S K F F E H H P F T Q E E V S S E G T T
P V S E V C P S L P P S D S T G S I C Q S T V S P S K S I P I
P P S F R S S E E D H R N Q F G Q R D S S S A P N V H I
N T I E P V N I D D L I R V Q G L P R S D G G S T T G L S A
T P P A S L P G S L T N V K V P Q K S P C Q Q R E R K S S S
S S E D R S K M L G R R D S S D D W E I P E G Q I T L G Q R I G
S G S F G T V F K G K W H G D V A V K M L N V T A P T P Q
Q L Q A F K N E V G V L R K T R H V N I L L E M G Y T T K P
Q L A I V T Q W C E G S S L Y H H L H I I E T K F E M I K L
I D I A R Q T A Q G M D Y L H A K S I I H R D L K S N N I F L H E
D L T V K I G D F G L A T V K S R W S G S H Q F E Q L S G
S I L W M A P E V I R L Q D K N P Y S F Q S D V Y A F G I V L
Y E L M S G V L P Y S N I N N R D Q I I F M V G R G Y L S P D
L S K V R S N C P K A M K R L M A D C L K K K R E E R P L F P
Q I L A S I E L L A R S L P K I H R S A S E P S L N R A G F
Q T E D F S L Y C A S P K T P I Q A G G Y G E F S A F K

СПИСОК ЛИТЕРАТУРЫ.

- [1] Орлов, Ю.К. Невидимая гармония / Ю.К. Орлов // Число и мысль. – Вып. 3. – М.: Знание, 1980. – С. 70-105.
- [2] Gumenjuk, A., On the acoustic elements of a poem and the formal procedures of their segmentation / Gumenjuk A., Kostyshin A., Borisov K., Salnikova O. // Glottometrics. – 8 (2004). – Lüdenschied: RAM-Verl., 2004. – С.42–67.
- [3] Бородовский, М.Ю. Статистические методы анализа генетических текстов / М.Ю. Бородовский, П.А. Певзнер // Компьютерный анализ генетических текстов. – М.: Наука, 1990. С. 33–80.
- [4] Гуменюк А.С., Шпынов С.Н., Морозенко Е.В., Родионов И.Н. О средствах анализа строя нуклеотидных цепей. // Математическая биология и биоинформатика: II международная конф. – Пущино: изд-во ООО «МАКС Пресс», 2008. – С. 123-125.
- [5] Gumenjuk A.S., Morozenko E.V., Ulaev V.A., Antosjuk E.V. On the selection of structural elements in protein sequences based on the criterion holistic resulting text.

Построение логико-вероятностной модели кодирующих последовательностей геномов

Г.С. Лбов, В.А. Гусев

Институт математики им. С.Л. Соболева СО РАН, Новосибирск 630090, ул. Коптюга, 4

vgus@math.nsc.ru

Build-up of logical-and-probabilistic model genomes coding sequences

Lbov G.S., Gusev V.A.

Institute of mathematics of the Siberian branch of the
Russian Academy of Sciences,
Novosibirsk 630090, str. Koptjuga, 4
vgus@math.nsc.ru

Abstract

In operation the analysis of coding genetic sequences presented in a digital form is carried out. Logic legitimacies in analyzed 4266 genes *E.coli* are found out. The next step proposes analysis of genomes more complex forms of living systems.

Работа выполнена при частичной финансовой поддержке РФФИ (проект № 10-01-00113-а), проект № 09-07-12087-офи_м, Интеграционного гранта СО РАН (№ 83)

Обширная библиотека расшифрованных генетических последовательностей (GenBank) является в настоящее время объектом пристального внимания математиков. Как правило, работы по математическому анализу геномов посвящены применению различных математических методов для выявления регуляторных и кодирующих участков в структуре геномов [1-3]. Этот анализ основан на сопоставлении различных символьных последовательностей исследуемых геномов с паттернами ДНК, функции которых уже известны.

В работах [4-8] авторы проводили анализ структуры генетического кода, то есть кодон-аминокислотного соответствия основываясь на методах теоретико-группового анализа. Это позволило обнаружить неизвестные ранее молекулярным биологам и биохимикам алгебраические и арифметические закономерности в структуре кода.

Логично предположить, что соответствующие алгебраические и арифметические, то есть символьные и числовые закономерности генетического кода должны иметь отображения в

нуклеотидной последовательности геномов. Для поиска числовых закономерностей необходимо представить стандартную символьную последовательность из АТ и GC пар в геноме в цифровом виде. Мы воспользовались данными работы [8], в которой показано, что молекулярные массы АТ и GC пар в составе двойной спирали равны соответственно 259 и 260 независимо от ориентации, то есть принадлежности нуклеотида к конкретной нити ДНК. Таким образом, чередование АТ и GC пар в двойной спирали соответствует чередованию только двух чисел. Для анализа закономерностей в этих числовых рядах, их удобно представить в виде последовательностей нулей и единиц. Поиск возможных закономерностей в таком бинарном ряду чисел был проведен в рамках подхода к анализу эмпирической информации, изложенного в монографиях [9,10]. Подход сводится к построению логико-вероятностной модели объекта исследования. Под логико-вероятностной моделью понимается список логических закономерностей, обладающих достаточно большой прогнозирующей способностью. Целью данной работы является поиск закономерностей в бинарных последовательностях, сопоставленных кодирующим фрагментам генома в классе логических решающих функций. В дальнейшем подобного рода закономерности будем называть логическими закономерностями.

References

- [1] Орлов Ю.Л. Анализ регуляторных геномных последовательностей с помощью компьютерных методов оценок сложности генетических текстов: Дис. канд. биол. наук. - Новосибирск. - 2004. 158 с.
- [2] Abnizova, I., Schilstra, M., teBoekhorst, R., Nehaniv, C.L. // WSEAS Transactions on Computational Methods, Issue 4, V.2, October 2003, P.1188-1196
- [3] Abe, T., Kanaya, S., Kinouchi, M., Ichiba, Y., Kozuki, T., and Ikemura, T. // Genome Res. 2003. V. 13. № 4. P. 693-702.
- [4] Duplij D., Duplij S. // Biophys. Bull. 2000. V. 497. P. 1-7.
- [5] Jimenez-Montano M.A., de la Mora-Basanez C.R. & Poschel T. // BioSystems. 1996. V. 39. P. 117-125.

- [6] Jimenez-Montano M.A. // BioSystems. 1999. V. 54. P. 47-64.
- [7] Negadi T. // Int. Journ. of Quant. Chem. 2003. V. 94. P. 65-82.
- [8] shCherbak V. I. // BioSystems. 2003. V. 70. P. 187-209.
- [9] Лбов Г.С. Методы обработки разнотипных экспериментальных данных. Новосибирск: Наука, 1981, 160 с.
- [10] Лбов Г.С., Старцева Н.Г. Логические решающие функции и вопросы статистической устойчивости решений. – Новосибирск: Издательство Ин-та математики, 1999. 211 с.

Системный анализ процесса построения проблемно-ориентированной базы данных с использованием ресурсов Entrez

Угольникова О.А.⁽¹⁾, Пономаренко Е.А.⁽²⁾, Демич Ю.А.⁽²⁾, Лисица А.В.⁽²⁾, Корнюшко В.Ф.⁽¹⁾, Швец В.И.⁽¹⁾

Московская государственная академия тонкой химической технологии им. М.В.Ломоносова⁽¹⁾, Институт биомедицинской химии им.В.Н.Ореховича РАМН⁽²⁾
ugolek84@gmail.ru

System analysis of automatically creation problem-oriented databases process based on Entrez system

Ugolnikova O.A.⁽¹⁾, Ponomarenko E.A.⁽²⁾, Demich Yu.A.⁽²⁾, Lisitsa A.V.⁽²⁾, Korniyushko V.F.⁽¹⁾, Shvets V.I.⁽¹⁾

The Moscow State Academy of Fine Chemical Technology after named M.V. Lomonosov, 119571, Moscow, avenue Vernadsky, 86⁽¹⁾; Institute of Biomedical Chemistry of Russian Academy of Medical Sciences 119121, Moscow, Pogodinskaya street, 10⁽²⁾

Abstract

The paper describes technology of automatically creation a problem-oriented databases based on Entrez system. It is shown that using such technology allows creating problem-oriented databases, corresponding to the current informational priorities of the research group.

Введение

Комплекс информационных ресурсов Entrez [1] является наиболее обширным источником информации о молекулярном устройстве живых систем. В настоящее время Entrez объединяет почти 40 баз данных, содержащих данные о структуре и функции генов, белков и химических соединений, а также результаты экспериментов в области транскриптомики и протеомики. Кроме того, в Entrez входит обширная библиотека рефератов научных публикаций MEDLINE/PubMed, насчитывающая в настоящее время порядка 20 млн. статей [4].

Средства доступа к содержащейся в Entrez информации постоянно совершенствуются, однако, применяемые для этой цели информационные модели носят универсальный характер и не всегда могут в должной степени отразить частные вопросы, возникающие в узкой предметной области [2]. Актуальной становится задача разработки подходов

информационного менеджмента, которые позволяли бы на основе Entrez формировать проблемно-ориентированные базы данных, отвечающие текущим приоритетам научной группы.

Подход к созданию проблемно-ориентированных баз данных решается в данной работе с применением технологии автоматического проектирования информационных систем (CASE-технология, Computer-Aided Software Design). Данная технология позволяет анализировать информационную модель предметной области на всех этапах создания и сопровождения проблемно-ориентированного ресурса. Использование CASE-технологии позволяет достичь оптимального сопряжения программной системы с информационными потребностями пользователя, что необходимо для гибкого взаимодействия с постоянно модифицируемой информационной средой Entrez.

Метод

В работе рассматривается подход к проектированию информационной системы с применением методологии функционального моделирования IDEF0 и диаграммы потоков данных (DFD, Data Flow Diagram). В качестве примера анализируется процесс формирования реляционной базы данных согласно стандарту IDEF1X. При формировании базы данных используются следующие элементы комплекса Entrez; (i) система поиска рефератов статей PubMed, (ii) система поиска сведений о структуре и свойствах химических соединений PubChem и (iii) перечень медицинских предметных рубрик MeSH. Информация из указанных источников согласно схеме функционального моделирования (рис. 1) была объединена в состав проблемно-ориентированной базы данных, предназначенной для информационного обеспечения исследований в области систем доставки лекарств.

На рис. 1 отображена контекстная диаграмма функциональной модели построения проблемно-ориентированной базы данных. В общем виде она включает перечень ключевых терминов, определяющих целевую предметную область. Например, в случае доставки лекарств перечень может включать такие термины, как «drug delivery» и

«drug transport». Перечень накладывает условие отбора публикаций с помощью системы PubMed, которые затем составят основу базы данных.

Необходимым условием для формирования исходных данных является наличие контролируемого словаря. Контролируемый словарь может содержать названия сущностей, таких как названия генов, белков, биохимических реакций, заболеваний и др. В рассматриваемом случае в контролируемый словарь включали наименования химических соединений, загруженный из системы PubChem [5] вместе с синонимичными вариантами.

В результирующей реляционной модели каждому соединению из контролируемого словаря соответствует библиографический перечень, содержащий публикации по заданной ключевыми терминами тематики. В базе данных устанавливаются связи между объектами и публикациями, далее объекты могут динамически формировать группы в соответствии с количеством общих публикаций [3].

При декомпозиции информационной модели уточняется, что между рефератами публикаций допустимы связи, отражающие смысловую родственность текстов. Родственность текстов определяется автоматически как количество общих для двух статей медицинских предметных рубрик. Отношения родственности обеспечивают возможность взаимодействия информационной системы с экспертом. Для этого было разработано программное обеспечение, которое позволяет получить гипотезу для не содержащейся в базе данных сущности [6]. В рассматриваемом примере в качестве такой сущности фигурирует название (или структурная формула) соединения, которое затем транслируется в набор публикаций посредством запроса к системе PubMed. Набор публикаций, в свою очередь, преобразуется в профиль медицинских предметных рубрик, в соответствии с которым из базы данных отбираются аннотированные экспертом объекты. Примером результирующей гипотезы, как показано выходящей стрелкой на Рис. 1, может являться предположение о том, что вещество исследовалось в составе системы транспорта лекарств.



Рис. 1. Контекстная диаграмма A-0 функциональной модели, используемой при создании проблемно-ориентированной базы данных в области доставки лекарств.

Результаты

Тематика доставки лекарств была задана экспертом с помощью перечня из 21 ключевого термина. В соответствии с предложенным перечнем из ресурса PubMed было загружено 646 тыс. рефератов. Из системы PubChem отобрали названия 6,5 тыс. химических соединений, из которых 2358 наименований были найдены в ранее сформированной выборке рефератов.

Отобранные наименования химических соединений и соответствующую им библиографию импортировали в базу данных MySQL. Обращение к загруженной информации осуществляли с использованием системы менеджмента контента BiblioBase, реализованной в оболочке DRUPAL.

Систему менеджмента контента использовали для организации в сети Интернет работы экспертов по аннотированию базы данных. Аннотирование проводили 3 эксперта в течение 2-х месяцев. Задача аннотирования сводилась к указанию для каждого соединения метки, исследовалось ли оно в составе какой-либо системы доставки. При этом в ходе аннотирования фиксировали рефераты тех публикаций, к которым эксперт обращался для принятия решений.

Проведенная работа показала, свойства около 60% из 2358 отобранных в автоматическом режиме соединений действительно исследовались в связи со встраиванием в систему доставки лекарств. Процесс аннотирования потребовал от экспертов ознакомиться с 3,5 тыс. публикаций, причем к 20% из них эксперты обращались неоднократно. В среднем эксперт выносил решение в отношении каждого соединения после ознакомления с 2-мя публикациями. В большинстве случаев, для поиска публикаций эксперт использовал стандартные средства системы PubMed, а не режим просмотра библиографии в проблемно-ориентированной базе данных.

Выводы

Современные технологии информационного менеджмента позволяют создавать на основе комплекса биомедицинских ресурсов Entrez проблемно-ориентированные базы данных, повышающие эффективность формализации экспертных знаний в узкоспециализированной области.

Программный комплекс BiblioEngine предоставлен Научно-исследовательской компанией "Криптом" (www.cryptome.ru).

Список литературы.

- [1] A.D. Baxevanis. Searching NCBI databases using Entrez. Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]. Chapter 1:Unit 1.3; 2008.

- [2] D. Hull, S.R. Pettifer, D.B. Kell. Defrosting the digital library: bibliographic tools for the next generation web. PLoS computational biology. 4(10):e1000204; 2008.
- [3] E.A. Ponomarenko, A.V. Lisitsa, E.V. Il'gisonis, A.I. Archakov. [Construction of protein semantic networks using PubMed/MEDLINE]. Molekuliarnaia biologii. 44(1):152-61.
- [4] E. Motschall, Y. Falck-Ytter. Searching the MEDLINE literature database through PubMed: a short guide. Onkologie. 28 (10):517-22; 2005.
- [5] Y. Wang, E. Bolton, S. Dracheva, et al. An overview of the PubChem BioAssay resource. Nucleic acids research. 38 (Database issue):D255-66; 2010.
- [6] О.А. Угольникова, Ю.А. Демич, А.В. Лисица. Алгоритмический подход к группированию активных соединений с использованием ассоциативного библиометрического анализа. АМиСОД; выпуск 14, с. 184-190, 2009.

Сравнительный анализ частотных словарей цепей геномных ДНК

Усанов Н.Н., Гильванова Е.А., Усанов Н.Г.

Институт биологии Уфимского научного центра РАН
nikusa@anrb.ru

Abstract

Substrings of different length ($N \leq 12$) quantity calculation in the nucleotide sequences of the first and the second DNA strands of prokaryotic and eukaryotic genomes form the arrays of data with high degrees of the linear Pearson correlation. The phenomenon was conventionally denoted as the «linguistic symmetry of DNA strands». Correlations are degenerating or disappearing, when genomic sequences of different organisms are compared or pseudo-random DNA is analyzed. An algorithm of DNA evolution with spontaneously arising DNA symmetry strands from pseudo-random state up to “living levels” has been suggested in the computer model.

1 Введение

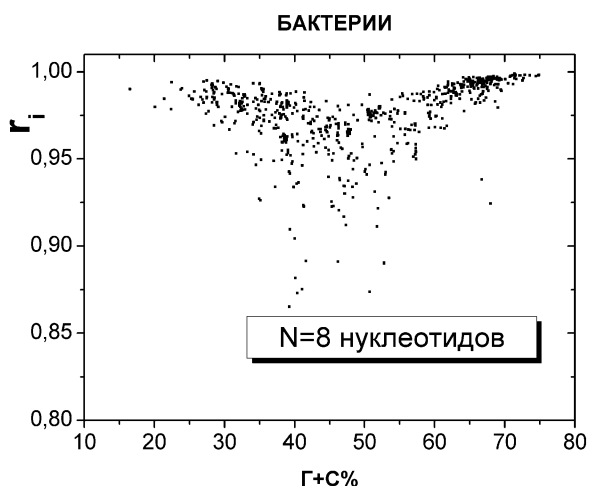
В любом геноме имеется пара комплементарных цепей ДНК, хранящих информацию об организме. Структура этих последовательностей такова, что гены равномерно располагаются на первой и второй (нумерация условна) комплементарных частях молекулы двойной спирали. На участках противоположных функциональным записям генов, (псевдогенов, регуляторных областей) находятся неактивные, комплементарные копии; между функциональными последовательностями располагаются гэнп области. Сумма всей информации, хранящаяся в одной цепи геномной ДНК может быть рассмотрена, как совокупность рабочих и инвертированных субпоследовательностей, разделенных участками с неочевидной функциональностью. Алгоритм генерации последовательности второй цепи из первой известен, но априори не позволяет сделать вывод о возможной схожести рассматриваемых последовательностей. Крайним является гипотетический случай, в котором первая цепь составлена главным образом нуклеотидами А и Г, а вторая, соответственно нуклеиновыми основаниями С и Т. Очевидно, что в данном случае частотные словари обеих последовательностей не могут иметь ничего общего. Обсуждению результатов сравнительного анализа корреляций между частотными словарями и цепей реальных геномных ДНК посвящена настоящая работа.

2 Методы

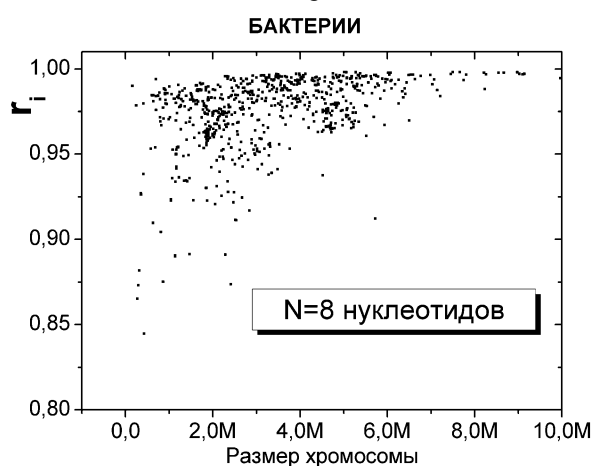
Для исследований были использованы последовательности геномов прокариот и эукариот <ftp://ftp.ncbi.nih.gov/genomes/>. Записи Генбанка рассматривались в качестве последовательностей первых цепей ДНК; (рассмотрен массив хромосом депонированных в Генбанке по состоянию на октябрь 2008). Тексты вторых цепей были сгенерированы по природному алгоритму, включающему чтение первой цепи в обратном направлении и замену нуклеиновых оснований по правилам комплементарности. В процессе анализа методом скользящей рамки по исследуемой последовательности генерировались словари толщиной $q=3 \div 12$ нуклеотидов. Оценку степени схожести словарей осуществляли вычисляя коэффициенты линейной корреляции Пирсона (r_i) [1]. Генерацию текстов псевдослучайных последовательностей проводили с помощью утилиты, разработанной на основе алгоритма Мерсенна (Mersenne twister).

3 Результаты и обсуждение

Рассчитаны частоты встречаемости идентичных субпоследовательностей ($N \leq 12$) нуклеотидов, в первой и второй цепях ДНК 752 бактериальных хромосом, 56 архей и 136 эукариот. Для каждого случая вычислены коэффициенты корреляции Пирсона. Для большей простоты изложения в дальнейшем тексте они обозначаются коэффициентами лингвистической корреляции (КЛК). Полученные данные свидетельствуют о том, что числа различных олигонуклеотидных сочетаний в первой и второй цепях хромосомальной ДНК приблизительно равны. В частности, для рамок считывания $N=3 \div 10$ (прокариоты) и $N=3 \div 12$ (эукариоты) корреляции частот встречаемости строк в парах цепей по Пирсону, в подавляющем большинстве случаев можно расценивать как сильные ($0,7 > r_i \geq 0,9$) или очень сильные ($0,9 > r_i \geq 1,0$). КЛК для бактерий в частном случае $N=8$ нуклеотидов мало зависят от ГЦ состава хромосомальной ДНК (см. рисунок).



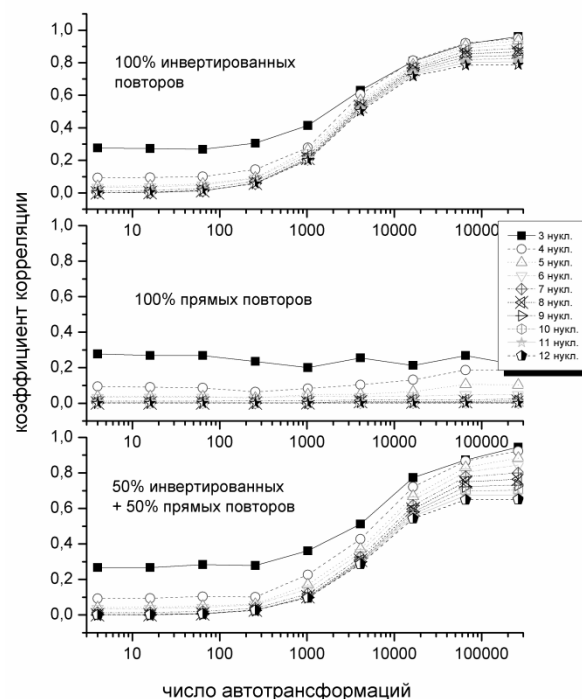
Также не наблюдается ярких зависимостей КЛК



для $N=8$ от длины хромосомальных ДНК. Аналогичные данные получены для архей и ряда эукариотов. Вместе с тем, при вычислении КЛК для пар различных организмов, отмечается резкое снижение сходства частотных словарей даже при исследовании выборки слов $N=6$. В частности, нами была рассчитана матрица кросс-соответствий 18 прокариотических организмов на уровне гексануклеотидных слов. В рамках данного текста можно привести лишь некоторые величины r_i , характеризующие сходство словарей биологически разных организмов. *Aeropyrum pernix* ÷ *Aquifex aeolicus*: КЛК=0,05317; *Bacillus cereus* ZK ÷ *Bdellovibrio bacteriovorus*: КЛК=0,23078; *Bifidobacterium longum* ÷ *Bacillus subtilis*: КЛК=0,03145; *Chlamydomonas reinhardtii* ÷ *Corynebacterium glutamicum*: КЛК=0,02014; *Escherichia coli* 536 ÷ *Clostridium tetani* E88: КЛК=0,1145.

Главным парадоксом нам представляется существование самого феномена схожести словарей первой и второй цепей хромосомных ДНК. Явление универсально практически для всех известных живых организмов но, любая нуклеотидная

последовательность, написанная случайно, обсуждаемыми параметрами не обладает. Мы предполагаем, что генерация симметрии является результатом непрерывного переноса крупных фрагментов ДНК с одной цепи на другую. Косвенно этот феномен прослеживается по наличию во всех прокариотических и эукариотических хромосомах протяженных повторов текста («эволюционных шрамов»), имеющих различную морфологию [2]. Алгоритм образования повторов был использован для генерации симметрии в модельном тексте. В качестве исходной матрицы была взята псевдослучайная ДНК последовательность размером 1 млн. п.о., имеющая КЛК $r_i < 0,35$. Матрица обрабатывалась в программе Автоморфер [1], по



алгоритмам генерации прямых и смешанных повторов фиксированной длины (300 нуклеотидов). Как следует из графиков, эффект возникновения лигвистической симметрии был наиболее активен при генерации инвертированных повторов. Мутации активно разрушали возникающую симметрию.

4 Заключение

Эволюция хромосомной ДНК определяется крупноблочными перестройками (метаморфозами), в процессе которых меняется видообразующие признаки. Мутации не являются доминантным процессом.

Литература

- [1] Н.Н.Усанов, Н.Г.Усанов Препринт ИПМ им. Келдыша РАН №43 за 2008 год.
- [2] Н.Н.Усанов, Е.А.Гильванова, А.В.Чемерис, Н.Г.Усанов; ICMBB06 2006 стр.177-178.

Разработка и создание биоинформационной базы данных, поддерживающей метод FIAV (Fast Identification of Alleles Variation) для диагностики наследственных болезней

Шкурат Т.П.⁽¹⁾, Шевцова В.С.⁽²⁾

НИИ Биологии Южного федерального университета, г. Ростов-на-Дону⁽¹⁾,
Донской государственный технический университет⁽²⁾

Barbaragen4@mail.ru

Design and development of bioinformatics databases supported the FIAV (Fast Identification of Alleles Variation) method for hereditary diseases diagnostic

Shkurat T.P., Shevtsova V.S.

Abstract

The paper proposes a fundamentally new structure of bioinformatics databases, which was designed to facilitate the work of doctors. It contains the information about SNPs causing diseases. Also this database has an integrated system of choice the primers for PCR. That allows to make the preclinical diagnosis of some diseases.

В последние годы активное развитие молекулярной биологии и смежных с ней областей, а также внедрение достижений геномики в медицину неопровержимо свидетельствует о том, что в основе патогенеза ряда заболеваний лежит тесная взаимосвязь с генетически опосредованными событиями.

Этот факт делает актуальной проблему систематизации и структурирования информации о заболеваниях, вызванных теми или иными нарушениями в геноме, а также разработки эффективных методов прогнозирования развития заболеваний на основе генотипирования пациентов. Одним из самых распространенных и удобных для работы видов хранилищ молекулярно-биологической информации являются биоинформационные базы данных (БД). Однако существующие БД ориентированы, как правило, на узкий круг пользователей (молекулярные биологи) и не могут удовлетворить потребности практической медицины в четко организованной информации. Основными причинами этого являются принципиальные различия в форматах хранения информации, что затрудняет поиск; избыточность информации, выдаваемый в результате запроса, что

требует дополнительных временных затрат на ее анализ; наконец, далеко не все БД содержат информацию о заболеваниях, вызываемых генетическими нарушениями.

Оптимальная структура медико-биологической БД, должна удовлетворять нескольким функциональным требованиям [3]:

1. Объем БД должен быть практически неограничен (т.е. ограничен только параметрами аппаратных средств).

2. БД должна быть достаточно гибкой для обеспечения перестройки по мере заполнения, так как заранее спроектировать детальную структуру динамически пополняемой базы невозможно. Это требование связано с тем, что знания в области молекулярной биологии динамично развиваются и их очень трудно описать в какой либо заранее спроектированной жесткой системе, удобной для компьютерной обработки.

3. Так как в БД должна быть интегрирована информация, представленная в различных международных базах данных в специальных форматах, база данных должна создаваться в среде гипермедиа, поддерживающей не только стандартные мультимедиа форматы, но и ряд специальных форматов (пространственные структуры молекул, химические структурные формулы и др.)

4. Эксплуатация и пополнение БД должны быть легко доступны через компьютерные сети пользователям, не имеющим компьютерной подготовки (биологи, медики).

На основании сформулированных требований и анализа существующих биомедицинских БД нами была разработана и опробована БД, ориентированная на специализированный круг пользователей и одинаково удобная, как для врачей-терапевтов, так и для молекулярных биологов. В разработанной БД содержится информация об однонуклеотидных полиморфизмах функциональных групп генов, вызывающих различные мультифакторные заболевания. Описание каждого полиморфизма включает его идентификатор (GeneID); название гена (полное и краткое), содержащего данный полиморфизм; информацию о нуклеотидной последовательности полиморфизма и мутации, приведшей к его

возникновению; информацию о функции кодируемого полиморфизмом белка, а также о клинической ассоциации полиморфизма с вызываемым им заболеванием.

Помимо функции поиска, БД оснащена интегрированной системой, осуществляющей выбор праймеров для ПЦР, что позволяет быстро формировать необходимые панели для исследований и проводить доклиническую диагностику заболеваний, с учетом этнической принадлежности, пола, места проживания, семейного анамнеза.

БД постоянно обновляется, имея интерактивные связи с крупнейшими мировыми биоинформационными БД. БД способна осуществлять поиск как по известному гену, влияющему на развитие заболевания, так и по заболеванию.

Работа над наращиванием возможностей БД продолжается.

Исследование выполнено в рамках Федеральной целевой программы «Научные и научно-педагогические

кадры инновационной России», госконтракт №02.740.11.5015.

Литература

- [1] Guy R. Cochrane and Michael Y. Galperin. The 2010 Nucleic Acids Research Database Issue and online Database Collection: a community of data resources // Nucleic Acids Research, Vol. 38, pp. 1-4, 2010.
- [2] Lewis Y. Geer et al. The NCBI BioSystems database // Nucleic Acids Research, Vol. 38, pp. 492-496, 2010.
- [3] База знаний по молекулярной и общей биологии человека (HUMBIO) <http://www.jcbi.ru/baza>
- [4] Т.П. Шкурат, И.О.Покудина, С.Р. Малхосьян и др. Новый способ FIAV для диагностики сердечно-сосудистых патологий // Геномные и постгеномные технологии, с. 57-58, 2009.

Филогения малых аполипопротеинов позвоночных

Лизунов А.В., Орлов С.В., Перевозчиков А.П.

Кафедра эмбриологии СПбГУ, НИИЭМ

Abstract

This investigation is dedicated to the phylogeny of the superfamily of small apolipoproteins. It is a superfamily of proteins, that includes in the cholesterol exchange. This group of proteins is specific for the vertebrates. In this work we used amino acid sequences of the small apolipoproteins to built dendrogram and reconstruct the evolution of the small apolipoproteins superfamily.

Введение

Литературные данные

Малые аполипопротеины позвоночных являются суперсемейством белков, объединенных структурой и функциями.

Суперсемейство малых аполипопротеинов позвоночных осуществляют перенос липидов в плазме крови, что обуславливает их антиатерогенный эффект. Исследования структуры и функций аполипопротеинов и их генов связано с вопросами исследования и поисков новых методов лечения и профилактики атеросклероза, сердечнососудистых заболеваний. Анализ структур и филогенетического родства всех аполипопротеинов позвоночных в будущем позволит найти гомологи аполипопротеинов, обладающих более сильными антиатерогенными свойствами, чем аполипопротеины млекопитающих[1-16,32-44]. Также исследование филогении аполипопротеинов имеет важное значение для понимания эволюции генов вообще и эволюции генов у позвоночных в частности

Наиболее консервативными структурами у семейства малых аполипопротеинов являются первые амфипатные структуры, состоящие из 33 аминокислотных остатков. Во вторичной структуре они образуют четкие выраженные домены на N-конце белка[2]. Также характерной чертой является то, что многие положения, например 8 и 9 положения в 33 аминокислотных доменах практически у всех аполипопротеинов занимают гидрофобные остатки. Консервативными являются и положения кислых, основных и нейтральных аминокислотных остатков[2]. Часть белка, кодируемая 4 экзоном, состоит из 22-аминокислотных остатков, состоящих из двух 11-аминокислотных сегментов[1]. 11-аминокислотные сегменты схожи между собой, практически каждый 22-аминокислотный домен, состоящий из них, начинается с остатка пролина. Хотя структурной

единицей амфипатной α -спирали является 11-аминокислотный повтор, 22-аминокислотный домен является базовой структурой для липид-связывающих доменов

Цели и задачи

Целью настоящей работы является создать схему филогении генов суперсемейства малых аполипопротеинов позвоночных, отражающую эволюцию суперсемейства малых аполипопротеинов позвоночных. Задачи: 1). Провести анализ нуклеотидных и аминокислотных баз данных позвоночных животных и составить базу данных нуклеотидных и аминокислотных последовательностей аполипопротеинов основных таксонов позвоночных. 2). Провести анализ сходства первичных последовательностей аполипопротеинов с помощью компьютерных статистических методов. 3). Составить схему филогении генов суперсемейства малых аполипопротеинов позвоночных.

Материалы и методы

Материалы

В данной работе использовались базы данных аминокислотных нуклеотидных последовательностей серверов NCBI и GenBANK. Для построения множественного выравнивания использовался пакет программ Mega4. Для расчета оптимальных модели дендрограммы использовалась программа Prottest 4. Для построения программ использовалась программа PhyML5.0

Методы

Проводился скрининг баз данных аминокислотных последовательностей. За основу были взяты последовательности аполипопротеинов человека. Так как они давно известны. Из найденных последовательностей отбирали, те, позитивная гомология с которыми составляла не менее 40%. После того, как первичная база аминокислотных последовательностей аполипопротеинов позвоночных была. Собрана, проводилось выравнивание по алгоритму clustalw с целью отбраковать все последовательности. Не являющиеся последовательностями аполипопротеинов. Все последовательности, которые давали максимальное число штрафных очков при выравнивании, повторно проверялись по базам данных аминокислотных последовательностей аполипопротеинов человека и мыши.

Полученные Последовательности аполипопротеинов выравнивались по алгоритму clustalw с помощью программы Mega4 и разбивались на домены. Шаблоном для выделения доменов служило разделение на домены, предложенное в статье Chi-Cheng Luo and Lawrence Chant, 1988. После этого проводилось определение консервативных позиций аминокислотных остатков в 33-аминокислотных доменах и 11-аминокислотных сегментах доменов 4 по классификации Chi-Cheng Luo and Lawrence Chant. После этого вероятность и значимость замен в каждой позиции рассчитывалась по Теореме Байеса. По результатам вычислений более значимыми являются замены в консервативных позициях 33-аминокислотных доменов.

Далее проводилось выравнивание 33-аминокислотных доменов по алгоритму clustalw и построение дендрограммы методом maximum likelihood в программе PhyML5.0. Для верификации параметров дендрограммы выровненные последовательности 33-аминокислотных доменов анализировались программой Prottest.

Результаты и выводы

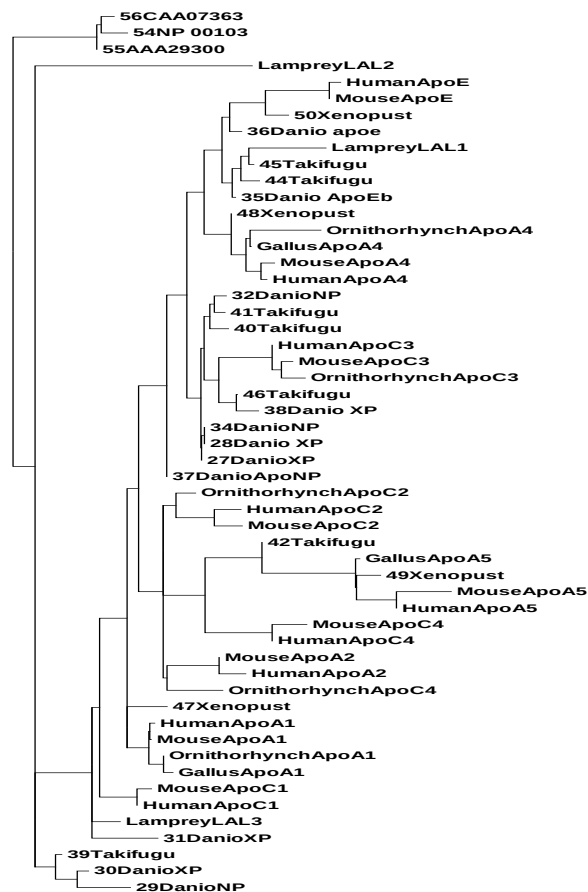
Результаты

Было идентифицировано 52 гомолога аполипопротеинов позвоночных, включая 3 гомолога аполипопротеинов *Petromyzon marinus*. Последовательности 33-аминокислотных доменов были выровнены по алгоритму clustalw. После этого была построена дендрограмма, которая была укоренена по аутгруппе. Ближайшим к предковой форме является гомолог аполипопротеинов *Petromyzon marinus* LAL2. Отдельно кластеризуются гомологи аполипопротеинов LAL3 и *Danio rerio*XP_001344822 и гомологи АпоА4,5,2,Е,С1,2,3,4. Большой подкластер разделяется на следующие кластеры: кластер АпоС3,А4,Е, и *Petromyzon marinus* LAL1, кластер АпоА1, кластер АпоА5, А2,С2,4и кластер АпоС1, образующий с остальными общий надкластер. В кластере АпоА5, А2,С2,4 АпоА5 и АпоС4 образуют общий подкластер, гомологи Апо С2 и АпоА2 образуют собственные кластеры. В кластере АпоС3,А4,Е отдельный кластер образуют гомологи АпоС3, гомологи АпоА4 и АпоЕ образуют подкластер. Дендрограмма представлена на рисунке№1.

Выводы

Аполипопротеины возникли, судя по результатам исследования, еще у первых бесчелюстных. Предковый ген дублировался с появлением двух генов. Один из них дал группу аполипопротеинов, которые на данный момент сохранились только у рыб, это группа аполипопротеинов *Danio rerio*NP_571203.1, *Danio rerio*XP_701241.1 и *Takifugu rubripes*NP_0010722100.У них в дальнейшем произошла дупликация доменов, в результате чего они приобрели современную структуру. Другой ген наоборот, претерпел делецию, в результате которой

были утеряны два 11-аминокислотных сегмента, содержащие остаток пролина на С-конце. Этот ген является предковым для большинства гомологов аполипопротеинов во всех группах позвоночных. Этот ген дублировался, с появлением гена. Предкового для LAL3 *Petromyzon marinus* и гена, давшего большой кластер аполипопротеинов – делеция одного 11-аминокислотного сегмента. Он дивергировал с появлением предкового АпоС1 и предкового гена остальных аполипопротеинов.В дальнейшем, дубликации и последующие делеции 11-аминокислотных сегментов. В результате этого появились все современные гомологи аполипопротеинов.



Рисунок№1: дендрограмма 33-аминокислотных доменов малых аполипопротеинов позвоночных. Реконструкция филогении. Метод Maximum likelihood.

Список литературы

- [1]. Wen-Hsiung Li, Masako Tanimura, Chi-Cheng Luo, Santanu Datta and Lawrence Chant, 1988. The apolipoprotein multigene family: biosynthesis, structure, structure-function relationships, and evolution. Journal of Lipid Research, Vol. 29.
- [2]. Masafumi Tanaka, Mao Koyama, Padmaja Dhanasekaran, David Nguyen, Margaret Nickel, Sissel Lund-Katz, Hiroyuki Saito, and Michael C. Phillips, 2007. Influence of Tertiary Structure Domain Properties on the Functionality.

Многоуровневое распараллеливание алгоритмов молекулярного моделирования методом Монте-Карло

Теплухин А.В.

Институт математических проблем биологии РАН, Пущино, Россия
tepl@impb.psn.ru

Abstract

An effective strategy for the molecular simulation by Monte Carlo method is presented based on three levels of the respective software parallelization using MPI. The upper level, “parallel tempering/replica exchange”, provides parallelization in the framework of configurational space of the system under consideration. At the intermediate one a domain decomposition approach is implemented to split computation work in small portions. At lowest level by-particle decomposition is used to calculate the interaction energy.

1 Введение

Компьютерное моделирование является мощным инструментом для изучения механизмов функционирования биологических объектов на атомно-молекулярном уровне. Современные ученые имеют возможность выполнять длительные вычислительные эксперименты на компьютерах, получая для исследований длинные серии «мгновенных фотографий» эволюционирующих объектов с атомным разрешением. Каждая такая «фотография» представляет собой мгновенную конфигурацию изучаемой системы в виде списка координат всех атомов в данный момент. В основе применяемых для этого компьютерных программ лежат либо детерминистские (решение уравнений молекулярной динамики), либо стохастические (реализации цепей Маркова в конфигурационном пространстве) методы получения серий мгновенных атомно-молекулярных конфигураций [1]. Искомые характеристики изучаемых систем рассчитываются далее путем усреднения по наборам мгновенных конфигураций, получаемым в ходе моделирования.

В связи с этим особую остроту принимает общая для вышеупомянутых методов моделирования проблема представительности выборок мгновенных конфигураций. Для небольших моделей (десятки атомов) эта проблема снимается путем сравнения результатов расчетов, стартовавших из разных начальных конфигураций. Однако время, требуемое

для получения одной «фотографии», с увеличением размера исследуемого объекта катастрофически возрастает. В свою очередь, усложняющаяся топология многомерного конфигурационного пространства (потенциальные барьеры, явления типа «бутылочных горловин» и др.) влечет за собой потребность в генерации значительно большего числа мгновенных конфигураций для перехода в новую представительную область. Все это делает стандартные методы моделирования непригодными для получения представительных выборок конфигураций молекулярных систем, содержащих тысячи и более атомов.

Средством преодоления этой проблемы может послужить разработка суперкомпьютерного программного обеспечения, в котором решены обе задачи: повысить эффективность «блужданий в конфигурационном пространстве» и ускорить расчет отдельной мгновенной конфигурации.

2 Основные уровни распараллеливания

2.1 Поатомная декомпозиция

Основная доля вычислительного времени уходит на расчет энергии взаимодействия молекулы/атома с окружающими частицами. Теоретически, взяв M процессоров, можно было бы M -кратно уменьшить время вычислений, закрепив за каждым процессором свою группу частиц. Однако изменение координат атомов в процессе моделирования требует своевременного обмена информацией между процессорами. На реальном оборудовании эффект ускорения вычислений с ростом числа процессоров быстро ослабевает (оптимальное количество – 32 процессора).

2.2 Пространственная декомпозиция

Согласно процедуре Метрополиса [2], выполняемой в ходе моделирования методом Монте-Карло, очередная мгновенная конфигурация получается изменением положения только одной частицы. При соблюдении некоторых условий (краткодействующие потенциалы) можно

организовать параллельное выполнение процедуры Метрополиса одновременно для нескольких частиц.

Для этого элементарная ячейка кубической формы (C -куб), содержащая моделируемую систему, представляется в виде совокупности кубов меньшего объема (P -кубы), образующих регулярную пространственную решетку. Каждый из процессоров занят расчетами движения молекул "своего" P -куба, периодически обмениваясь необходимыми данными с "соседними" (в пространстве моделируемой ячейки) процессорами о перемещениях молекул как внутри P -кубов, так и между ними.

Этот подход к моделированию молекулярных систем обладает высокой масштабируемостью и не имеет формальных ограничений ни на размер элементарной ячейки, ни на количество атомов в исследуемом объекте [3,4]. Эффективность вычислений можно существенно повысить, используя поатомную декомпозицию (см. предыд. раздел) при расчетах энергии взаимодействия частиц P -куба между собой и с частицами соседних кубов.

2.3 Декомпозиция конфигурационного пространства

Повышению результативности «сканирования» конфигурационного пространства служит группа методов, развиваемых в рамках общей идеологии под названием "параллельная закалка/обмен репликами" [5]. В соответствии с этой методикой параллельно запускаются расчеты нескольких копий изучаемой системы (старт с одной и той же конфигурации, но при разных температурах). Через регулярные интервалы «соседние» по температуре реплики могут (при выполнении некоторого условия на соотношение разностей их энергий и температур) поменяться конфигурациями. Такая процедура позволяет высокотемпературным репликам преодолевать энергетические барьеры и «забираться» в любые закоулки многомерного конфигурационного пространства, а обмен конфигурациями с «более холодными» репликами ведет к более быстрому формированию представительной выборки для целевой температуры (обычно самая холодная реплика).

Согласно статистической физике, с увеличением числа частиц в системе, соответствующие энергетические распределения имеют тенденцию «заостряться». Это ведет к падению вероятности

обменов между репликами. Чтобы не допустить этого приходится сужать температурные интервалы, и, как следствие, увеличивать число реплик, а, соответственно, и количество процессоров.

2.4 «Суперпрограмма»

В перспективе предполагается разработать программу, запускающуюся на совокупности процессоров, достаточной для размещения K групп процессов (по одной группе на реплику). Соответственно, группа состоит из P (по количеству P -кубов) доменных процессов, каждому из которых «помогают» еще по M процессов (поатомная декомпозиция). При обмене между репликами следует передавать только температурные индексы, а не массивы атомных координат. В итоге – при запуске одной задачи будет задействоваться $K \times P \times (M+1)$ процессоров/ядер.

3 Заключение

Предлагаемый подход позволит эффективно использовать тысячи суперкомпьютерных процессоров/ядер для полноатомного моделирования биомолекулярных систем размером в десятки нанометров.

References

- [1] M.P. Allen, D.J. Tildesley. *Computer simulation of liquids*. – N.Y.: Oxford University Press, 1987.
- [2] N.A. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, volume 21(6), pages 1087-1092, 1953.
- [3] А.В.Теплухин. Многопроцессорное моделирование гидратации мезоскопических фрагментов ДНК. *Матем. моделир.*, том 16, выпуск (№) 11, с. 15-24, 2004.
- [4] М.Д.Калугин, А.В.Теплухин. Изучение взаимодействия кофеина с ДНК в водном растворе методом Монте-Карло с использованием параллельных вычислений. *Ж. структ. химии*, том 50, выпуск (№) 5, с. 878-889, 2009.
- [5] Y.Sugita, Y.Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem.Phys. Lett.*, volume 314(1-2), pages 141-151, 1999.

Использование графических сопроцессоров для интерактивного анализа траекторий молекулярной динамики

Лихачев И.В., Балабаев Н.К.

Институт математических проблем биологии, г. Пущино
ilya_lihachev@mail.ru

Abstract

Data processing of big volumes of molecular trajectory files is a difficult task for modern CPUs. An application of graphical processors as computing devices for general purpose helps to develop fast interactive programs. In this work CUDA technology is applied for developing algorithms for calculating dynamical contact maps and quasiharmonic analysis of large macromolecules.

1 Введение

Программами моделирования молекулярной динамики пользуются во многих исследовательских лабораториях [1]. В настоящее время обрабатываются файлы траекторий молекулярной динамики большого объема. Программа TAMD (Trajectory analyzer of molecular dynamics – Анализатор траекторий молекулярной динамики) [2] служит для быстрого анализа траектории в целом, предоставляя возможность рассмотрения динамики молекулярной системы в виде молекулярного кино и получения простых характеристик вдоль траектории. Вместе с тем в TAMD заложены и более сложные алгоритмы анализа динамики молекулярных структур. Для больших размеров систем требуются объемные вычисления, которые при использовании стандартных алгоритмов не позволяют интерактивно обрабатывать данные в реальном масштабе времени. Решить эту проблему можно, воспользовавшись графическими процессорами как вычислительными устройствами общего назначения (GPGPU – general purpose graphical processor unit). Благодаря использованию массовой параллельности, удалось значительно ускорить наиболее требовательные к ресурсам алгоритмы Анализатора траекторий молекулярной динамики – построения динамических контактных карт и вычисления колебательных мод макромолекулы в квазигармоническом приближении.

2. Ускорение доступа к данным

Первое, что требуется от любой программы по анализу траекторий – это быстрый доступ к содержимому файла траектории. В TAMD используется буфер чтения траекторных файлов, реализованный в виде отдельного программного

модуля. Основные требования, которые к нему предъявлялись, были следующие: 1) хранить некоторое число последних кадров траекторий в оперативной памяти для последующего быстрого доступа к ним; 2) хранить в буфере случайную последовательность кадров; 3) буфер не должен замедлять линейное чтение кадров с диска.

3. Ускорение расчета контактных карт

Построение динамических контактных карт и лент контактов уже заложено в TAMD [3]. В настоящее время разработано дополнение к TAMD для ускоренного расчета контактных карт.

Контактная карта представляет собой матрицу парных контактов между атомами молекулярной системы. Обычно такая матрица представляется в виде квадратного симметричного относительно главной диагонали точечного рисунка. Интенсивность закрашки наиболее мелкой видимой единицы изображения – пикселя – связана с появлением или отсутствием контакта между единицами молекулярной системы. Под последними понимаются отдельные атомы или фрагменты молекулы (например, аминокислотные остатки) – квазичастицы.

С вычислительно-аппаратной точки зрения, процесс отображения контактной карты делится на два этапа:

- 1) нахождение матрицы расстояний между всеми парами частиц в системе;
- 2) преобразование матрицы расстояний в точечный рисунок.

В работе [3] вводятся динамические контактные карты и говорится об их преимуществе по сравнению с обычными статическими контактными картами. При исследовании больших молекулярных комплексов актуальным становится вопрос повышения быстродействия при построении динамической контактной карты. Процесс расчета контактной карты можно значительно ускорить, используя графические процессоры в качестве математического сопроцессора по технологии CUDA.

Современные графические ускорители компании NVIDIA предоставляют возможность их использования в качестве многоядерного процессора общего назначения, работающего по технологии SIMD (single instruction multiple data). Мультипроцессор способен выполнять

одновременно одну и ту же инструкцию, оперируя при этом целым набором данных. Это подобно созданию большого числа нитей (thread) внутри одного процесса операционной системы. Но на создание большого числа нитей графического процессора не тратится так много ресурсов, как на нити операционной системы. Графический процессор состоит из некоторого числа мультипроцессоров (30 для наиболее популярных моделей на момент написания статьи). Каждый мультипроцессор способен обрабатывать 768 или 1024 (для более поздних моделей) нитей по принципу SIMD. Таким образом, на графическом процессоре без потери производительности для каждой нити при грамотной организации обращения к памяти могут параллельно выполняться несколько десятков тысяч нитей. Для сравнения, современный центральный процессор содержит от 4 до 16 ядер, и позволяют одновременно обрабатывать не более 2-х потоков на каждом из ядер. Сами вычислительные ядра графического процессора уступают по быстродействию центральному процессору. Но за счет большого количества вычислительных ядер и малых затрат на создание большого числа нитей, суммарная вычислительная способность GPU (выражаемая в числе операций с плавающей точкой за 1 секунду) значительно выше, чем у центральных процессоров.

Графические процессоры необходимо рассматривать именно как сопроцессоры к центральному процессору. При решении целого ряда задач они способны показывать чрезвычайно высокое быстродействие.

3.1 Методика разработки программ с использованием технологии CUDA

Каждый элемент матрицы расстояний может быть вычислен независимо от всех остальных. Этот факт означает реализуемость алгоритма параллельного вычисления контактной карты. Составим программу, которая будет вычислять все элементы матрицы расстояний параллельно.

Программа выглядит следующим образом:

1. Задается функция расчета одного элемента матрицы расстояний, выполняемая на устройстве (device) – графическом процессоре.
2. Основная программа считывает кадр траектории.
3. Данные одного кадра копируются из основной (оперативной) памяти в память устройства.
4. Параллельно запускается функция вычисления расстояния сразу для всей матрицы расстояний.
5. Ожидается завершение работы графического процессора.
6. Данные из памяти устройства копируются в оперативную память (память хоста).

7. Данные преобразовываются в точечный рисунок.

В пункте 4 предполагается, что все функции работают параллельно. На самом деле они запускаются блоками, количество которых зависит от модели графической карты. Процесс запуска полностью контролируется технологией CUDA, и не требует от программиста дополнительных действий.

4. Квазигармонический анализ

В TAMD реализован многоступенчатый алгоритм нахождения и анализа низкочастотных колебаний молекулярной системы. Первый пункт – вычисление матрицы ковариаций молекулярной системы. Алгоритм получения последней реализован в n-процессорном (n-ядерном) режиме. Далее матрица ковариаций подается в виде входных данных в подпрограмму расчета собственных чисел и собственных векторов, реализованную с использованием GPGPU. Полученные величины определяют набор частот и соответствующих им направлений колебательных движений. Данные о колебательных модах передаются в интерактивный инструмент их просмотра – Визуализатор квазигармонических колебаний, являющийся составной частью программы TAMD.

5. Вывод

Использование современных гетерогенных вычислительных систем (многоядерных вычислительных комплексов и графических процессоров) позволяет значительно сократить время вычисления сложных характеристик молекулярных объектов и предоставить интерактивный режим анализа их поведения на траектории молекулярной динамики.

Ссылки

1. Н.К. Балабаев, К.В. Шайтан. Компьютерное моделирование молекулярной динамики. В кн.: Методы компьютерного моделирования для исследования полимеров и биополимеров, М., Издательство ЛКИ, 2009. Глава 2, с.35-62;
2. И.В. Лихачев, Н.К. Балабаев. Анализатор траекторий молекулярной динамики. Математическая биология и биоинформатика, 2007, том 2, №1, с.120-129.
3. И.В. Лихачев, Н.К. Балабаев. Построение расширенных динамических контактных карт по данным молекулярно-динамических расчетов. Математическая биология и биоинформатика. 2009. Т. 4. № 1. С. 36-44.

О масштабируемости на многоядерных процессорах алгоритма поиска повторов в ДНК

Пятков М.И.⁽¹⁾, Панкратов А.Н.^(1, 2)

Институт Математических Проблем Биологии РАН⁽¹⁾,
Московский Государственный Университет, факультет Вычислительной Математики и
Кибернетики⁽²⁾
mpyatkov@gmail.com

Abstract

This work represents a synthesis of recent research in the framework of the repeats search algorithm on the basis of a generalized spectral-analytical method.

In particular, we consider different methods of calculating the expansion coefficients, as well as study the scalability of the algorithm with different approaches to the organization of the search engine.

1 Описание алгоритма

1.1 Основные этапы

Данное исследование является продолжением работы над спектральным алгоритмом поиска повторов в геномах [1,2]. Главной задачей было изучение масштабируемости данного алгоритма на многоядерных процессорах. Далее алгоритм будет рассмотрен поэтапно.

На первом этапе работы программы нуклеотидная последовательность ДНК, состоящая из «букв» A,T,G,C, преобразуется в профиль. Профиль — функция, которая получается путем пробегания по последовательности «скользящего окна» с шагом в один нуклеотид. В этом окне на каждом шаге считается количество нуклеотидов G и C, после чего окно перемещается, и процесс повторяется до конца последовательности. На следующем этапе профиль преобразуется в вектора коэффициентов путем разложения по дискретному базису Чебышева. Для преобразования мы используем еще одно окно, которое проходит по профилю с определенным шагом. На каждом шаге профиль, попавший в окно, преобразуется в вектор коэффициентов разложения, и эти коэффициенты индексируются в виде матрицы. Следующим шагом является попарное сравнение полученных ранее векторов и оценка их близости друг к другу. На последнем этапе результаты сравнения отображаются в виде матрицы спектральной схожести, автоматический анализ которой выдает координаты повторов.

1.2 Особенности алгоритма

Отличие от традиционных подходов к поиску повторов, представленных, с одной стороны, семейством алгоритмов динамического программирования, а с другой вероятностными алгоритмами, дает как преимущества, так и недостатки. С одной стороны, когда мы используем алгоритмы динамического программирования, мы непосредственно работаем с буквенной последовательностью и сразу получаем точные координаты найденных повторов. С другой стороны, работа с протяженными последовательностями осложняется спецификой алгоритма. В основе нашего алгоритма лежит обобщенный спектрально-аналитический метод, что позволяет эффективно распараллеливать программу поиска повторов, так как при вычислении коэффициентов разложения отсутствует зависимость по данным. Такой подход также позволяет быстро обрабатывать более протяженные участки генома. Однако из-за того, что мы работаем с некоторой функцией, а не непосредственно с последовательностью, поиск точных координат повторов в буквенной последовательности является нетривиальной задачей, которая решается автоматическим анализом матрицы спектральной схожести.

2 Исследование масштабируемости

2.1 Индексация против рекуррентности

При создании программы поиска повторов учитывались возможности использования многоядерных машин. Поэтому были проведены тесты масштабируемости алгоритма. В частности, на этапе вычисления коэффициентов разложения имели место два подхода к организации вычислений. В первом случае предполагалось сначала вычислить матрицу полиномов, а затем коэффициенты разложения. Второй подход представлял собой рекуррентное вычисление коэффициентов разложения без предварительного вычисления матрицы полиномов. Исследование масштабируемости показало, что в данный момент подход с предварительным вычислением матрицы полиномов является более эффективным, но с

появлением компьютеров с большим количеством ядер он будет проигрывать подходу с постоянным вычислением, т.к является менее масштабируемым из-за низкой скорости обмена данными между процессором и оперативной памятью. Тесты проводились на машине с общей памятью(SMP). При массивно-параллельных вычислениях на кластерах подход с предварительным вычислением матрицы сложнореализуем и неоправдан, т.к отдельные узлы могут не иметь общего доступа к оперативной памяти. Подводя итог вышесказанному, можно сказать, что для одной итерации вычисления коэффициентов разложения лучше использовать подход с рекуррентным вычислением полиномов.

Если рассмотреть программу в более крупном масштабе, то мы увидим, что вопросы, относящиеся к одной итерации, справедливы и для этапа сравнения коэффициентов. А именно, стоит ли создавать матрицу, содержащую векторы коэффициентов разложения, и только после этого сравнивать найденные вектора попарно, или же обойтись без индексации и проводить вычисления и сравнения на ходу? Тщательное тестирование показало, что подход с матрицей индексации оказался более масштабируемым, и на 32 ядерной SMP машине удалось получить прирост производительности в 27 раз против 24-кратного ускорения при постоянном вычислении и сравнении векторов коэффициентов разложения.

2.2 Используемые средства

Для исследования масштабирования мы использовали стенд, предоставленный компанией Intel рамках конкурса Intel Manycore Testing Lab. Стенд представлял собой четырехпроцессорную машину с процессорами Intel Xeon X7560, где

каждый процессор обладал 8 ядрами без учета Hyperthreading, т.е в сумме реальных 32 ядра.

Для реализации многопоточности использовалась библиотека OpenMP, штатно представленная в компиляторах GCC 4.3 и Intel Compiler 11.1.

Аппаратно-ускоренные и векторные инструкции (SIMD) были реализованы с помощью библиотеки Intel IPP (Intel Performance Primitives); данные оптимизированные функции мы использовали на этапе вычисления коэффициентов разложения.

Работа выполнена при поддержке РФФИ проекты № 10-01-00609, 08-07-00353 и компании Intel.

2.3 Английское название и список авторов

Pyatkov M.I., Pankratov A.N. On the scalability on multicore processors of spectral algorithm for finding repeats in DNA.

References

1. F.F.Dedus, L.I.Kulikova, S.A.Makhortykh, N.N.Nazipova, A.N.Pankratov and R.K.Tetuev Analytical Recognition Methods for Repeated Structures in Genomes. Doklady Mathematics, Volume 74(3), pages 926-929, 2006.
2. A.N. Pankratov, M.A. Gorchakov, and F.F. Dedus, N.S. Dolotova, L.I. Kulikova, S.A. Makhortykh, N.N. Nazipova, D.A. Novikova, M.M. Olyshevets, M.I. Pyatkov, V.R. Rudnev, R.K. Tetuev, and V.V. Filippov. "Spectral Analysis for Identification and Visualization of Repeats in Genetic Sequences" . Pattern Recognition and Image Analysis, Volume 19(4), pages 687–692, 2009.

Поиск низкомолекулярного лиганда для аффинной хроматографической очистки фактора свёртывания крови VIII человека молекулярным докингом *in silico*

Мацкевич В.А.⁽¹⁾, Орлова Н.А.⁽¹⁾, Воробьев И.И.^(1,2), Юрьев А.С.⁽¹⁾, Воробьев А.И.⁽¹⁾
Гематологический Научный Центр Российской Академии Медицинских Наук⁽¹⁾
Институт Биоорганической Химии Российской Академии Наук

matskevichv@gmail.com

Finding small molecule ligand for affinity purification of human coagulation factor VIII by means of the *in silico* molecular docking

Matskevich V.A., Orlova N.A.,
Vorobiev I.I., Yuriev A.S.,
Vorobiev A.I.

Abstract

In silico docking was used to select the new small molecule affinity ligand for purification of recombinant blood coagulation factor VIII. Previously designed by Kessler and coworkers peptide ligands L3 and L4, reversibly binding factor VIII molecule, were used as reference compounds. Assumed binding sites of these compounds were located at C1 or/and C2 domains of the protein.

In the present work we have modeled the interaction of L3 and L4 compounds with theoretically proposed binding surfaces, located at apex regions of C1 and C2 domains. It was found that both compounds bind to the mapped surfaces with significant composite docking indexes. These surfaces were used for *in silico* docking of intact and modified virtual compound libraries from "Asinex" and "Chembridge" suppliers. The final docking scores of the top candidates from both libraries are of the same range and closely correspond to docking scores of reference compounds. Trimming of the initial library and modification of compounds molecules by the linker group, performed by simple Perl scripts before the docking run dramatically increased both calculation efficiency and chemical relevance of top candidates. Small subset of ligand candidates will be further tested for affinity and reversibility of factor VIII binding.

Докинг *in silico*, или, другими словами, скрининг библиотек химических соединений, является в настоящее время широко распространённым методом в поиске новых лекарственных средств. В

настоящей работе мы представляем пример использования обычных программных инструментов скрининга библиотек химических соединений для поиска наилучшего аффинного лиганда для очистки рекомбинантного фактора свёртывания крови VIII человека.

Докинг производился на 3 четырёхъядерных рабочих станциях Intel x86 под управлением ОС Linux. В процессе подготовки библиотек химических соединений и трёхмерной структуры белка к докингу использовались внутренние утилиты оболочки докинга Schrodinger Maestro, для фильтрации библиотек химических соединений и их конвертации в формат SMILES использовалась утилита Open Babel, автоматизированная модификация производилась при помощи Perl-скрипта. В работе была использована пространственная структура белка из банка данных PDB, код 2R7E [4], которая была модифицирована следующим образом: домены, не несущие (по литературным данным [2]) на своей поверхности сайтов связывания, были удалены ручным редактированием координатного файла. Далее координатный файл был подготовлен к докингу при помощи Schrodinger Maestro.

Опубликованные лиганды фактора VIII L3 и L4 [2] были использованы для проверки релевантности гипотетических сайтов связывания на поверхности доменов C1 и C2. В докинге на поверхности гипотетических сайтов связывания лигандов были использованы две библиотеки коммерчески доступных химических соединений: "Asinex" (примерно 300 тыс. химических соединений) и "Chembridge" (примерно 1 млн. химических соединений), полученных из онлайн-репозитория "ZINC" [1].

Было предположено, что вероятными областями связывания фактора VIII с лигандами L3 и L4 являются выступающие над поверхностью доменов C1 и C2 "вершины", участвующие в связывании с поверхностью липидных мембран [3]. Также на эти поверхности на вершинах доменов C1 и C2 указала утилита маппинга рецепторных сайтов связывания на поверхности белков из пакета докинга Schrodinger. Таким образом, рецепторные сетки, в

которые производятся докинг, были размещены поверх вершинных областей доменов C1 и C2.

Докинг библиотек химических соединений производится по схеме последовательного обогащения более энергетически-выгодными связывающимися с рецепторными сайтами лигандами. Первый докинг библиотеки "Asinex" начался с примерно 500 тыс. химических соединений на входе в первую стадию, продолжился в режиме докинга стандартной точности с 20 тыс. отобранных на первой стадии соединений и закончился в режиме докинга увеличенной точности с 2 тыс. отобранных на второй стадии соединений. Каждая последующая стадия докинга требовала больше вычислительной производительности от рабочей станции, так как учитывала больше факторов. Три стадии докинга длились около 12 недель чистого машинного времени на одной четырёхъядерной рабочей станции. В результате по значению композитного индекса "Glidescore", генерируемого пакетом Schrodinger для каждого соединения в ходе докинга, среди химических соединений были отобраны 40 наиболее энергетически выгодных молекул-кандидатов. Среди этих 40 молекул-кандидатов 24 имеют карбоксильную группу, и 16 из них имеют 2 и более карбоксильных групп. Величина Glidescore для этих 40 найденных кандидатов соответствовала диапазону Glidescore конформеров L3 и L4 в использованных для докинга рецепторных сетках.

Для проверки афинности и обратимости связывания этих молекул с фактором VIII их необходимо иммобилизовать на твёрдом носителе, для чего было решено использовать конъюгированные с аминокислотой L-цистеина молекулы, иммобилизованные через сульфгидрильную группу на малеимид-активированный линкер, ковалентно связанный с дном плашки.

Для повторного докинга (библиотеки "Chembridge", 2.4 млн. соединений в подготовленной для докинга форме) потребовалось бы на той же рабочей станции около 60 недель вычислительного времени, что автоматически делало докинг нецелесообразным. Вычислительная производительность была увеличена до 3 четырёхъядерных рабочих станций, сходных по характеристикам с первой, а малые молекулы было решено отфильтровать по критерию возможности их иммобилизации на малеимид-активированные линкеры или на линкеры, активированные малеиновым ангидридом (то есть, специфически взаимодействующие с первичными аминокислотными группами), после чего пакетно модифицировать, включив в их состав концевой участок соответствующего линкера, повысив химическую релевантность моделирования связывания с данными малыми молекулами *in silico*.

После проведённой в OpenBabel автоматической фильтрации карбоксил-содержащих и содержащих первичный амин лигандов из 2.4 млн. соединений библиотеки "Chembridge" осталось в сумме около 200 тыс. соединений. Эти 200 тыс. соединений были преобразованы OpenBabel в формат SMILES для последующей модификации соединений соответствующими линкерами скриптом на Perl-е. После пакетной модификации этих соединений по карбокси- и аминокислотным группам данная библиотека фильтрованных и модифицированных лигандов прошла заново процедуру подготовки к докингу в программе Schrodinger и достигла размеров примерно 300 тыс. соединений. Был проведён докинг этой библиотеки в сетки докинга вершин доменов C1 и C2 на трёх четырёхъядерных рабочих станциях. Трёхстадийный докинг этой библиотеки в две сетки в общей сложности занял около 2 недель чистого машинного времени. При анализе полученных таблиц наилучших кандидатов с третьей стадии докинга выяснилось, что поляризованные атомы линкеров в большей части соединений участвуют в образовании стабилизирующих водородных связей с атомами боковых цепей фактора свёртывания VIII, тем самым была повышена химическая релевантность моделирования *in silico* и уменьшена до приемлемой продолжительность докинга. Величины Glidescore лигандов-кандидатов также оказались в области значений этого индекса для конформеров L3 и L4, как и выбранных кандидатов докинга библиотеки "Asinex".

References

1. Irwin JJ, Shoichet BK. ZINC – a free database of commercially available compounds for virtual screening. In *Journal of Chemical Information and Modeling*, volume 45(1), pages 177-182, 2005.
2. Knör S, Khrenov A, Laufer B, Benhida A, Grailly SC, Schwaab R, Oldenburg J, Beaufort N, Magdolen V, Saint-Remy JM, Saenko EL, Hauser CA, Kessler H. Efficient factor VIII affinity purification using a small synthetic ligand. In *Journal of Thrombosis and Haemostasis*, volume 6(3), pages 470-477, 2008.
3. Ngo JC, Huang M, Roth DA, Furie BC, Furie B. Crystal structure of human factor VIII: implications for the formation of the factor IXa-factor VIIIa complex. In *Structure*, volume 16(4), pages 597-606, 2008.
4. Shen BW, Spiegel PC, Chang CH, Huh JW, Lee JS, Kim J, Kim YH, Stoddard BL. The tertiary structure and domain organization of coagulation factor VIII. In *Blood*, volume 111(3), pages 1240-1247, 2008.

Сервер для вычислений по биоинформатике в грид-среде

Феранчук С.И., Калиновский А.А., Тузиков А.В.

Объединенный институт проблем информатики НАН Беларуси
feranchuk@gmail.com

Abstract

A site for managing common bioinformatics tasks is developed in the UIIP in order to provide an easy tool to applied scientists for resource-consuming computational experiments and to adjust a flexible allocation of computational resources in a grid environment. Particular cases of the most typical bioinformatics problems were tested by researchers from Belorussian scientific groups. The covered areas include molecular dynamics, docking, quantum chemistry, protein modeling, statistical analysis of biological data. Advanced software packages from all these areas were adopted for grid calculations.

1 Введение

Для удобной работы с основными типами ресурсоемких задач биоинформатики в ОИПИ была создана грид-среда, включающая выделенный сервер как интерфейсный компьютер и настроенная для доступа к грид-сетям для постановки задач на суперкомпьютерах. Пользователи грид-среды, которыми являются биологи и медики из Беларуси, получают собственный аккаунт и доступ к командной строке и графическому интерфейсу Linux на интерфейсном компьютере. При наличии сертификата для доступа в грид-сеть для них настроена возможность пользоваться программным обеспечением промежуточного уровня (ПО ПУ) и настроенными на вычислительных узлах грид-сети прикладными пакетами программ.

2 Прикладное ПО

Прикладное ПО, настроенное на работе в грид-среде, включает пакеты молекулярной динамики

(Amber, Charmm, Gromacs, NAMD), пакеты квантовой химии (Gamess, FireFly, AbInit и др.), пакеты для докинга (Dock6, Flexx, Hex и др.), программы для обработки первичных последовательностей (Blast, Clustal, Raxml и др.), программы для моделирования структуры белков (Modeller, Jackal и др.), а так же утилиты, базы данных и клиентские программы ПО ПУ gLite и Unicore. Для доступа к интерфейсному компьютеру с пользовательских компьютеров используется протокол NX фирмы NoMachine [1].

3 Круг решаемых задач

Прикладные исследования касались выявления структурно жестких участков петли V3 ВИЧ с помощью молекулярной динамики [2], проведения докинга и определения энергии взаимодействия субстратов с белками из семейства цитохромов P450, предсказания третичной структуры белков вируса ГЛПС, статистического анализа первичных последовательностей ВИЧ, квантово-механических расчетов ячейки кристалла арсенида галлия, исследования взаимодействия кислорода с гемом в миоглобине.

References

- [1] Pfeifle K. NX – A New Stage for Network Desktop Computing http://www.linux-congress.org/2003/abstracts/index.html#3_7_2.
- [2] A. Andrianov., I. Anishchenko. Computational Model of the HIV-1 Subtype A V3 Loop: Study on the Conformational Mobility for Structure-Based Anti-AIDS Drug Design. J. Biomol. Struct. Dynam., 2009, Oct;27(2):179-93.

Моделирование структуры белков

Ефимов А.В.

Институт белка РАН, 142290 Пущино, Московская область
efimov@protnes.ru

Abstract

Protein structure modeling by A.V. Efimov, Institute of Protein Research, Russian Academy of Sciences, 142290 Pushchino, Moscow Region

Knowledge of all possible protein folds is necessary for 3D protein structure prediction. Theoretical modeling of protein 3D structures based on construction and analysis of structural trees of proteins is a promising approach to solving the problem. The approach is based on the hypothesis that at the first step of protein folding a nucleus is formed and then the remaining part of the molecule or domain is folded around it. In modeling, the structural motif having a unique overall fold and handedness is taken as the starting structure of the root structure of the tree. The larger protein structures are obtained by stepwise addition of α -helices and/or β -strands to the growing structure taking into account a restricted set of rules inferred from the known principles of protein structure. Among these rules, attention to handedness and compactness, prohibition of crossing connections, and a requirement to α -helices to be packed in α -helical layers and β -strands in β -layers are the most important. The number of allowed overall folds that can be obtained from one structural motif is limited since the rules drastically reduce the number of allowed pathways of growth of intermediate structures. A general scheme that represents the root structural motif, all the intermediate and completed structures connected by lines showing allowed pathways of structure growth, is referred to as the structural tree. To date several structural trees for the largest protein superfamilies have been constructed. Some of them are accessible at the web-site [1]. Such modeling of protein folds and folding pathways is of particular value in understanding the relationship between the 3D protein structure and the amino acid sequence.

Взаимосвязь между аминокислотной последовательностью белка и его пространственной структурой остаётся одной из интригующих и нерешённых проблем биохимии и молекулярной биологии. Тем не менее, многие закономерности строения белковых молекул установлены и широко используются при исследовании белков. Среди них – компактность молекул (результат действия принципа плотной упаковки), наличие гидрофобных ядер и полярных оболочек («подобное» взаимодействует с «подобным»), слоистость белковых молекул, способность полипептидной цепи образовывать различные регулярные вторичные структуры с насыщенной системой водородных связей, а также различные супервторичные структуры или структурные мотивы с уникальными укладками цепи и определенной хиральностью и др. На основании этих знаний разработаны десятки методов предсказания вторичной структуры белков, однако получить 100%-й результат предсказания не удается и, по-видимому, невозможно без знания информации о третичной структуре. Что касается третичной структуры, то более или менее надежное предсказание третичной структуры неизвестного белка в настоящее время возможно только методом моделирования на основе гомологии, т.е. только в том случае, когда известна пространственная структура какого-либо гомологичного белка. Таким образом, несмотря на определенные успехи, в целом проблема взаимосвязи между первичной и третичной структурами белка остается пока нерешенной. Одним из подходов к решению этой проблемы является моделирование структуры белков и путей их сворачивания с помощью построения и анализа структурных деревьев.

Структурное дерево белков – это совокупность всех разрешённых промежуточных и конечных пространственных структур, которые могут быть получены из одной корневой (стартовой) структуры путём добавления к ней других элементов вторичной структуры; возможные пути роста структур показываются линиями, которые в итоге объединяют все структуры в одно дерево. При моделировании структур каждого семейства белков в качестве стартовой (корневой) структуры берётся

соответствующий структурный мотив с уникальной укладкой цепи. Такие структурные мотивы состоят из двух и более α -спиралей и/или β -тяжей, имеют уникальную укладку цепи и хиральность и «указывают» место, где должна пристраиваться остальная часть полипептидной цепи. Пристраивание других элементов к стартовому (корневому) структурному мотиву происходит в соответствии с набором правил, выведенных из известных принципов структурной организации белков. Имеются веские основания предполагать, что корневой (стартовый) структурный мотив древа может быть зародышем в процессе сворачивания белков данного семейства, а остальная часть полипептидной цепи пристраивается к нему. По крайней мере, структурные деревья показывают возможные зародышевые структуры и пути их роста. Таким образом, структурные деревья являются удобным и перспективным инструментом, с помощью которого можно изучать структурное сходство белковых молекул, разрабатывать структурную классификацию белков, исследовать механизмы сворачивания белков, осуществлять поиск как известных, так и пока неизвестных пространственных структур белковых молекул, моделировать трёхмерные структуры белков и т.д.

К настоящему времени построены и опубликованы структурные деревья для наиболее крупных белковых суперсемейств, таких как β -белки, содержащие abcd-единицы, 3β -уголки, S-образные β -листы; двухслойные ($\alpha+\beta$)-белки, содержащие abCd-единицы; трёхслойные α/β -белки, содержащие пяти- и семисегментные α/β -мотивы; α -белки, содержащие α - α -уголки, а также белки,

содержащие ϕ -мотивы [2-6]. Основной задачей данной работы является обзор работ, посвящённых построению и анализу структурных деревьев, а также их использованию в исследованиях белков. Ознакомление с этими работами позволит читателю глубже понять основные закономерности строения белковых молекул и механизмы их сворачивания.

Работа выполнена при финансовой поддержке РФФИ (грант 10-04-00727).

References

- [1] Institute of Protein Research Web site, 2010. <http://strees.protres.ru/>.
- [2] A.V. Efimov. Structural trees for protein superfamilies. In *Proteins*, vol. 28, pp. 241-260, 1997.
- [3] A.V. Efimov. A structural tree for proteins containing S-like β -sheets. In *FEBS Letters*, vol. 437, pp. 246-250, 1998.
- [4] A.B. Gordeev, A.M. Kargatov, and A.V. Efimov. PCBOST: Protein classification based on structural trees. In *BBRC*, vol. 397, pp. 470-471, 2010.
- [5] А.В. Ефимов. Структурные деревья глобулярных белков. *Успехи биологической химии*, т. 44, с. 109-132, 2004.
- [6] А.В. Ефимов. Структурные деревья белков, содержащих ϕ -мотивы. *Биохимия*, т. 73, с. 29-35, 2008.

Family of Ribosomal Proteins S1 Contains Unique Conservative Domain

A. V. Machulin⁽¹⁾, E. I. Deryusheva⁽²⁾, O. M. Selivanova⁽³⁾, I. N. Serdyuk⁽³⁾

Institute of Cell Biophysics⁽¹⁾, Tula State University⁽²⁾, Institute of Protein Research⁽³⁾
and-4@ya.ru

Abstract

Alignment of the S1 proteins containing different numbers of domains with the S1 RNA binding domain of *Escherichia coli* polynucleotide phosphorylase provided for discovering a unique conservative domain within this family.

1 Objects of study

1.1 S1 ribosomal protein family

S1 ribosomal protein family was chosen for object of study because of its uniqueness. First, approximately 700 sequences belonging to this family are currently known; second, the number of amino acid residues in S1 proteins from different bacteria varies in a very wide range (from 111 to 863 residues in *Spiroplasma kunkelii* and *Treponema pallidum*); and, third, the proteins belonging to this family are mainly composed of repeated homologous S1 domains [1].

1.2 Polynucleotide phosphorylase

It is known that the *E. coli* PNPase contains one S1 domain at its C end; this domain is highly homologous to four initially found S1 repeats. The spatial structure of the S1 RNA binding domain of the *E. coli* PNPase determined by NMR is a barrel like fold with an additional α -helix between the third and fourth β -sheets.

This particular OB fold of the secondary structure is regarded as the main structural element of the S1 ribosomal protein family. The majority of residues on the surface of the barrel formed by PNPase S1 domain are conserved, suggesting that this domain plays an RNA binding role. However, some residues of many domains in S1 proteins are conserved (Phe 19, Phe 22, His 34, Asp 64, and Arg 68) and clustered on the surface of the domain. These residues are located in the loop between the first and second β -sheets (Phe 19), in the middle of the second β -sheet (Phe 22), at the end of the third β -sheet (His 34), and in the loop between the last two β -sheets (Asp 64 and Arg 68). These residues are spatially juxtaposed at one side of the barrel forming

an RNA binding site characteristic of the overall ribosomal protein family (Fig.1) [2].

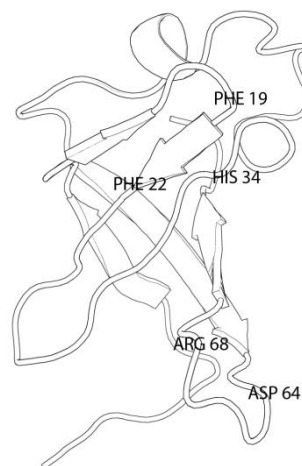


Fig.1 3D structure of PNPase from *E.coli*. PDB ID, 1SRO. Conservative residues Phe 19, Phe 22, His 34, Asp 64, and Arg 68 were marked on domain surface.

1.3 Protein families databases

It is known, that the databases, such as SMART, Pfam, and PROSITE, contain the information about protein domains, their number, and boundaries in various protein families. Each database compiling the information about protein structures, folds, and domain organization has certain operation algorithms. Since these algorithms are sometimes different and even the databases utilizing the same algorithms have additional conditions or limitations, the output data from different databases for the same object can considerably differ. We have compared the data for three databases (SMART, Pfam, and PROSITE) by the example of the S1 ribosomal protein family.

The copy numbers of S1 domains for the majority of the analyzed bacteria coincide in all three databases (SMART, PROSITE, and Pfam). For example, all bacteria from the order *Bacillales* contain four S1 repeats and those from the order *Rhodobacterales* (Alphaproteobacteria) contain six. However, some differences were found for several bacteria.

For this reason we have decided to use homology of secondary structures between phosphorylase and S1 proteins (predicted with PsiPred) in order to define the

copy numbers of S1 domains in each S1 protein sequence.

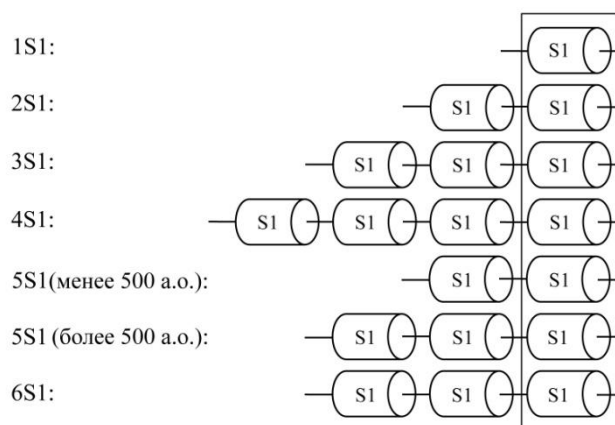


Fig.2 The S1 proteins containing different numbers of domains; the unique conserved domain is framed.

2 Results and Discussion

2.1 Definition of structural domains

We analyzed the secondary structure of the primary sequences of the S1 ribosomal protein family contained in the SwissProt database using the PsiPred program [3]. Based on only secondary structure predictions for the S1 RNA binding domain of the *E. coli* PNPase without the SMART, Pfam, and PROSITE data, we detected the structural domains in protein sequences, the number of which varied from one to six.

2.2 Discovering a unique conservative domain

Alignment with ClustalX program [4] of the S1 proteins carrying different numbers of the domains with the S1 RNA binding domain of the *E. coli* PNPase allowed us to find in this family a domain with the maximal homology to the PNPase S1 domain.

This conserved domain migrates along the chain and is localized in the proteins with different numbers of domains according to a certain pattern (Fig.2).

This domain migrates along the chain. It is the last domain in the two-, three-, and four domain proteins. In the case of five domain proteins with a length up to 500 residues, it is the second domain and with a length exceeding 500 residues as well as in the six domain proteins with a length over 550 residues – the third domain.

Note that the above mentioned residues Phe 19, Phe22, His 34, Asp 64, and Arg 68 (in some cases replaced by the corresponding conserved residues) are located in this particular conserved domain thereby once again confirming the uniqueness of this repeat and allowing it to be regarded as the strongest RNA binding site.

2.3 Model for RNA-protein interactions

All these facts suggest that found conservative domain can be regarded as the basic RNA binding domain in the S1 ribosomal protein family.

Consequently, the one domain proteins, such as, for example, the *S. kunkelii* protein, can be a simple model for RNA-protein interactions.

References

- [1] Subramanian Alap-Raman. Structure and functions of ribosomal protein S1. In *Progr. Nucleic Acids Res.*, volume 28, pages 101-142, 1983.
- [2] Bycroft M., Hubbard T., Proctor M., Freud S., Murzin G. The solution of the S1 RNA binding domain: A member of an ancient nucleic acid binding fold. In *Cell*, volume 24, pages 235-242, 1997.
- [3] Jones D.T. Protein secondary structure prediction based on position specific scoring matrices. In *J. Mol.Biol.*, volume 292, 195-202, 1999.
- [4] Jeanmougin F., Thompson J.D., Gouy M., Higgins D.G., Gibson T.J. Multiple sequence alignment with ClustalX. In *Trends Biochem. Sci.*, volume 23, 403-405, 1998.

Key ARG residues in protein-DNA interactions

Oparina N.U.⁽¹⁾, Kuznetsov E.N.⁽²⁾, Esipova N.G.⁽¹⁾, Tumanyan V.G.⁽¹⁾, Anashkina A.A.⁽¹⁾

Engelhardt Institute of Molecular Biology RAS (EIMB RAS)⁽¹⁾,

Institute of Control Sciences RAS (ICS RAS)⁽²⁾

anastasya.anashkina@gmail.com

1 Introduction

1.1 Importance of protein-DNA interactions

Most important biological processes are regulated through the interactions between DNA and proteins. Insight in these processes and subsequent approaches on gene expression regulation provide the possibility to complex diseases treatment, like cancer or hereditary diseases. Also it gives the opportunity to solve a riddle of embryogenesis when sophisticated set of events occurs when gene expression switches on molding deferent tissues and organs. Understanding of recognition process allows in principle macromolecules constructing with necessary values of binding constants for controlling of processes in living organisms. In such a way new generation drugs can be created.

1.2 Protein-DNA recognition problem

Problem of protein-DNA recognition comprises from two parts: how does protein find appropriate site in DNA and how does protein form a complex with DNA with a high affinity? We can answer only the second question that correspond to the static protein-DNA complex analysis. The process of long-scale search along the DNA macromolecule is extremely complicated. Common hypothesis suggests that recognition of DNA site starts from one-dimension diffusion of protein to concrete DNA site passing a distant way along DNA chain. Some proteins prefer three-dimensional pathways. In proximity of the binding site protein join DNA in unique kinetic way or by a number of different ways. Really pathways can contain more then one stage or include intermediate states.

1.3 Affinity and specificity

As for the difference between affinity and specificity, affinity is defined by physicochemical complementarities between protein factor and DNA site. Affinity is characteristic even for nonspecific binding. Specificity is a measure of selective binding of appointed DNA part by protein. If the protein binds DNA with high affinity, a lot of contacts between residues and nucleotides are formed [1]. Whereas amino acid residues and nucleotides have preferable partners for interactions this must be revealed in the probability to detect such pairs in complexes. So, probability code

can give solid basis for prediction of high-affinity interactions as well.

1.4 Problem statement

In this work we formulate a problem of key residues in protein-DNA combining.

2 Materials and methods

2.1 Voronoy-Delaunay tessellation

Voronoi-Delaunay tessellation [2] was used for correct detection of contacts between atoms of protein and DNA. Let notice that this rigorous mathematical approach defines atoms in contact unambiguously and in non-parametric way and bases only on coordinates of atoms.

2.2 Probability and specificity

We follow to our method of estimation of contact probabilities in protein-protein interactions study.

2.1 Protein-DNA complex set

All available protein-DNA complexes in PDB [3] were extracted for thorough contact analysis. As a result 1734 complexes were selected on May, 2010.

3 Results

3.1 Contact number distribution

Contact analysis of all 1734 protein-DNA complexes revealed that share of residues that formed contact with only one nucleotide is 40.16%, with two nucleotides is 40.55%, with three nucleotides is only 9.04% from the total amount of residues belonging interface (**Figure 1**). ARG4 from A chain of 1EA4 complex has the highest number of contacts. This arginine locates in major groove of double stranded DNA and formed 17 contacts with neighboring nucleotides (**Figure 2**).

The medium number of nucleotide neighbors per interfacial residue is 2.0. Minimal medium value is 1.6 for aspartic acid; maximal value is 2.7 for arginine.

3.2 Residue types that bond more then 6 contacts

Quantity of residues formed more then 6 contacts is less 1.3% from total amount of interfacial residues. They are mainly arginines and lysines (see **Figure 3**).

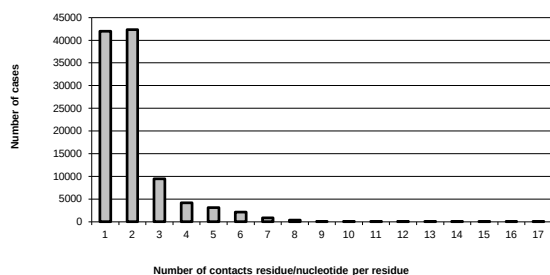


Figure 1. Contact number histogram of residue/nucleotide contacts per residue.



Figure 2. Contacts between DNA nucleotides and ARG4 residue of A chain from 1EA4. This arginine imbedded in major groove of double stranded DNA and formed 17 contacts with neighboring nucleotides.

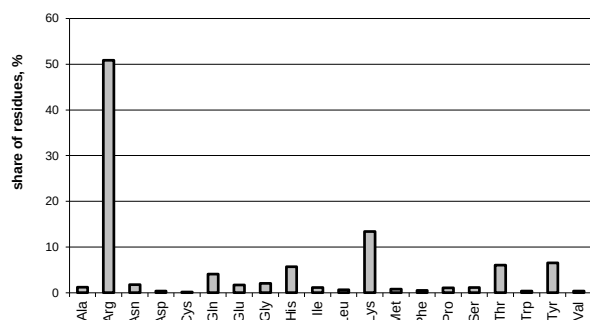


Figure 3. Amino acid composition of residues formed more than 6 contacts.

One can suppose that these residues (formed

contacts with more than 6 different nucleotides through protein-DNA interface) are the most important points in protein-DNA complex formation.

3.3 Conservatism of residues formed more then 6 contacts

Significance of revealed arginine and lysine residues for protein-DNA complex formation is open for discussion. We checked a conservatism of these residues (formed 7 and more contacts) and drew a conclusion that about 2/3 of these residues are conservative.

4 Discussion

The finding that arginine is a major participant in protein-DNA recognition seems quite reasonable because guanidine group provides possibility of various interactions and aliphatic chain supplies necessary flexibility for fitting the side chain into DNA. At the same time this observation is not trivial since the role of lysine is rather weak (see Figure 3) in spite of similarity of lysine and arginine properties. Indeed, both amino acids bear positive charge and abounded in such DNA-binding proteins as protamines and hystones. However, our study suggests that only arginine may be treated as a main agent for specific recognition. It may be proposed that as unique property as charge delocalization can serve as physicochemical basis for arginine prevalence among contacting amino acid residues.

5 Funding

Grants of Presidium RAS «Fundamental sciences for medicine» and «Molecular and Cellular biology», Grant of Education and Science Department of Russian Federation, RFBR Grant № 09-07-00503.

References

- [1] von Hippel, P.H., Protein-DNA recognition: new perspectives and underlying themes. *Science*, volume 263(5148), pages 769-770, 1994.
- [2] Anashkina, A., et al., Comprehensive statistical analysis of residues interaction specificity at protein-protein interfaces. *Proteins*, volume 67(4), pages 1060-1077, 2007.
- [3] Berman, H.M., et al., The Protein Data Bank. *Nucleic Acids Res*, volume 28(1), pages 235-242, 2000.

The use of cylindrical models and cluster analysis for phasing Low Density Lipoprotein particles

Oleg V. Sobolev⁽¹⁾, Vladimir Y. Lunin⁽¹⁾, Natalia L. Lunina⁽¹⁾, Manfred W. Baumstark⁽²⁾

Institute of Mathematical Problems of Biology RAS, Pushchino, Russia⁽¹⁾,
Medizinische Universitätsklinik Freiburg, Freiburg, Germany⁽²⁾
oleg@impb.ru

Abstract

X-ray diffraction experiment allows determining of magnitudes of complex coefficients in the decomposition of the studied electron density distribution into Fourier series. The determination of the lost in the experiment phase values poses the central problem of the method, namely the phase problem. In this work we applied cylindrical models and a cluster analysis for phasing LDL particles. An interactive computer program ClanGR was developed to perform this analysis. A result of this work was preliminary phasing of the LDL particles and determination of dimensions and the packing of particles for different LDL fractions.

1 Introduction

1.1 Phase problem

The base of X-ray experiment is diffraction of X-rays by a protein crystal. This allows to pose the problem of defining three-dimensional function, which describes a distribution of an electron density in the object. Interpreting maxima of this function as a centers of atoms allows to build an atomic model of the object. The distribution of an electron density in the crystal may be presented as a three-dimensional Fourier series. Magnitudes of a complex Fourier coefficients (in crystallography “structure factors”) may be found directly from the experiment, but the determination of phases is rather difficult, which is ‘the phase problem’ of X-ray structural analysis.

At present, there are a number of approaches to solve phase problem, but to apply them one need additional experimental data or usage of prior properties of electron density distribution. An X-ray experiment allows to determine only finite number of magnitudes of structure factors. This number depends on quality of a crystal and experimental conditions. A partial sum of a Fourier series which includes only structure factors with known magnitudes is called Fourier synthesis of electron density. Amount of experimental data is characterized by “resolution”, a value that represent a minimal size of structure details that can be observed in a graphic representation of Fourier synthesis of electron density.

Present state of macromolecular crystallography allows structure determination of many macromolecules almost automatically if high-quality crystals are available. Though there are a lot of cases where standard approaches do not give a solution. Very often this is the case when one studies a large molecule (*e.g.* a membrane protein) or macromolecular complex.

1.2 LDL

Low-density lipoprotein (LDL) belongs to the family of lipoproteins which are involved in the transport of lipids from the liver to extrahepatic cells [1] and plays a major role in the development of coronary heart disease [2]. Knowledge of the three-dimensional structure of this large macromolecular complex and its components is important in order to explain the molecular mechanisms involved in the development of atherosclerosis.

1.3 Cylindrical models

Following to our preliminary results on the shape of the LDL particle [3], we used a rough model of two nested cylinders of different densities. An outer cylinder with higher density represented a protein shell of the LDL particle and an inner cylinder with lower density represented a lipid core.

1.4. Cluster analysis

Some phasing methods at low resolution result not in the unique solution, but in a list of admissible variants. These variants may be phase sets or sets of a model parameters that allow phase calculation (rotational and translation parameters in Molecular Replacement *e.g.*) [4,5]. In such cases there is a problem of selecting a “right” variant from a list of approximately equal variants. To solve this problem one can apply a cluster analysis to this list of variants and obtain a meaningful phases. Cluster analysis methods allow splitting this list into classes of close solutions and extracting a representative or an average solution in every class.

Clusterisation technique is based on the matrix of solution-to-solution distances and the choice of a proper distance function may be essential for a success. In crystallographic applications we usually need electron density which can be interpreted rather than formal parameters. So, the choice of a metric based on the closeness of Fourier syntheses rather than the closeness of model parameters may give sometimes an advantage

[5]. Different parameters may produce similar but shifted syntheses. So, they need to be aligned before the distance calculation.

As a result of application of cluster analysis we will obtain a cluster tree. Examining the tree one can understand whether there are several similar phase sets in the list or not. Then there are two ways of obtaining an answer. The first way is to select one variant from the cluster and the second way is to calculate average upon the cluster. It is hard to foresee which way will provide a better solution, so it is reasonable to try both of them and compare results.

2 Phasing of LDL

This method was applied to phase LDL particles. We obtained X-ray data of 27Å resolution for different fraction of the LDL particles. None of routine phasing methods is applicable at such a low resolution.

A minimization of a multidimensional function in the space of cylinder parameters was carried out to fit parameters for both cylinders (position, radius, height, density). The minimized function showed the amount of the difference between the experimental data and structure factors calculated from the model. This minimization resulted not in the unique solution, but in a list of several different sets of cylinder parameters with approximately equal value of the target function. This list consisted of 20 sets of cylinder parameters that provided the best agreement with experimental data. Then 20 phase sets were calculated for these sets of cylinder parameters.

Next we carried out the cluster analysis of these phase sets and built a cluster tree. Phase sets from the best clusters were averaged.

A result of this work was preliminary phasing of the LDL particles and determination of dimensions and the packing of particles for different LDL fractions.

3 ClanGR

To perform cluster analysis and syntheses averaging described above we have developed a program ClanGR. This program allows interactive study and averaging of phase sets. ClanGR is written in Python 2.6 with wxWidgets graphic library. These tools allow fast and easy development of user-friendly and cross-platform graphic user interface. There are all necessary tools in the current version of ClanGR for processing phase sets obtained during phasing. To use this software one must have a specially formatted text file with experimental data and a list of calculated structure factors obtained during modelling.

After entering all parameters, and short calculation, user will see a main window of ClanGR (Fig. 1). There is a scale on the left side of the window, which allows to determine a distance between clusters. Selected

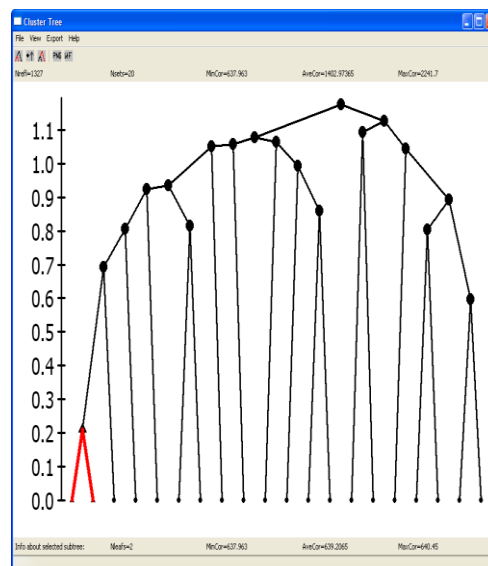


Fig. 1. ClanGR main window

cluster is red with triangle nodes. There are separate statistics for all tree and a selected subtree (cluster). For the selected cluster user can obtain a Fourier synthesis. User can save a picture of a cluster tree in .png format.

Time for a calculation of a cluster tree for a list of 1000 phase sets is not exceed 1 minute on a personal computer, and a calculation of 1 synthesis requires even less time. All these allows comfortable work with ClanGR in interactive mode.

Currently available version of ClanGR is compiled for Windows, but the next versions of ClanGR will be ported to Linux and MacOS.

This work was partially supported by RFBR grant 10-04-00254a and CNRS-RAS collaboration.

References

- [1] Brown, M. S. & Goldstein, J. L. In *Science*, volume 232, pages 34-47, 1986.
- [2] Gordon, T., Kannel, W. B., Castelli, W. P. & Dawber, T. R. In *Arch. Intern. Med.*, volume 141, pages 1128-1131, 1981.
- [3] Lunin V.Y., Lunina N.L., Ritter S., Frey I., Berg A., Diderichs K., Podjarny A.D., Urzhumtsev A. and Baumstark M.W. In *Acta Cryst. section D*, volume 57, pages 108-121, 2001.
- [4] Lunin V.Yu., Urzhumtsev A.G. and Skovoroda T.P. In *Acta Cryst. section A*, volume 46, pages 450-544, 1990.
- [5] Buehler A., Urzhumtseva L., Lunin V.Y. and Urzhumtsev A. In *Acta Cryst. section D*, volume 65, pages 644-650, 2009.

Компьютерный дизайн водорастворимых аналогов β -галактозилцерамида – потенциальных лекарственных препаратов для терапии СПИДа

Анищенко И.В.⁽¹⁾, Кисель М.А.⁽²⁾, Андрианов А.М.⁽²⁾

Объединенный институт проблем информатики НАН Беларуси⁽¹⁾,
Институт биоорганической химии НАН Беларуси⁽²⁾
anishchenko.ivan@gmail.com

Abstract

Using the cutting-edge methods of computer modeling the spatial structures for β -galactosylceramide (β -GalCer) and its water-soluble analogues were constructed followed by their molecular docking with the HIV-1 gp120 V3 loop giving rise to the virus principal neutralizing determinant. As a result, two modified forms of β -GalCer are found to present basic structures suitable for the rational anti-AIDS drug design.

1 Введение

В настоящее время особое внимание исследователей, занимающихся разработкой противовирусных препаратов, сконцентрировано на петле V3 ВИЧ-1. Этот интерес обусловлен тем, что указанный сегмент белка gp120 оболочки ВИЧ-1 представляет один из его главных “рабочих инструментов”, формируя основную мишень для нейтрализующих антител, а также детерминанты клеточного тропизма и образования синцития [3]. Согласно данным литературы [2] к числу химических соединений, способных к эффективной блокаде этого функционально значимого сегмента вируса относится гликофинголипид β -галактозилцерамид (β -GalCer), образующий на поверхности ряда чувствительных клеток первичный рецептор для ВИЧ-1, альтернативный молекуле CD4, используемой вирусом для проникновения в макрофаги и Т-лимфоциты. Для создания на основе β -GalCer эффективных противовирусных препаратов представляет интерес получить его водорастворимые формы, способные эффективно блокировать основную антигенную детерминанту вируса.

На сегодняшний день экспериментальные данные о механизме взаимодействия β -GalCer с функционально значимыми участками петли V3 отсутствуют в литературе. В этой ситуации наиболее рациональным способом получения детальной информации о принципах структурной организации комплексов гликолипидов с петлей V3 ВИЧ-1 является использование современных методов вычислительной химии. В настоящей работе с целью оценки возможностей привлечения модифицированных форм β -GalCer в качестве потенциальных терапевтических агентов выполнены следующие исследования: (i) методами

компьютерного моделирования построены математические модели, описывающие на атомном уровне структуру и динамические свойства β -GalCer и его водорастворимых аналогов; (ii) на основе расчетных данных осуществлено компьютерное конструирование структурных комплексов гликолипидов с пептидами петли V3 ВИЧ-1, имитирующими ее центральные фрагменты, и проведен анализ их стабильности.

2 Методы

Квантово-механические расчеты (RHF/6-31G*) трехмерных структур β -GalCer и его аналогов осуществляли с помощью пакета GAMESS. Конформационную подвижность гликолипидов и пептидов петли V3 ВИЧ-1 исследовали методами молекулярной динамики (МД). Молекулярный докинг пептидов петли V3 с β -GalCer и его модифицированными формами реализовали в пакете UCSF DOCK 6.3. Оценку стабильности комплексов и величин свободной энергии их образования проводили на основе анализа МД траекторий комплексов с использованием метода MM-PB/SA. МД расчеты гликолипидов, пептидов петли V3 и их комплексов осуществляли в пакете Amber 10. Во всех расчетах использовали компьютерный кластер «СКИФ-ОИПИ» ОИПИ НАН Беларуси и Национальную ГРИД-сеть.

3 Результаты и обсуждение

3.1 Конструирование исходных структур липидов и пептидов петли V3

На основе хранящейся в Банке белковых структур (Protein Data Bank) информации о родственной молекуле – α -GalCer (код 1ZT4), были построены стартовые структуры β -GalCer и трех его водорастворимых аналогов (water-soluble analogues; WSAs), изображенные на рисунке 1. После этого, используя эти конформации в качестве исходных моделей, мы осуществили квантово-механические расчеты структур гликолипидов и определили характерные заряды их атомов.

Результаты ряда экспериментальных исследований подтверждают наличие специфических взаимодействия β -GalCer с петлей

V3 белка gp120, а именно с ее центральной областью. Поэтому для изучения этих взаимодействий мы использовали центральный гексапептид петли V3, который в расчетах был представлен двумя модификациями, характерными для консенсусных последовательностей ВИЧ-1 подтипов А и В, соответственно [1,4].

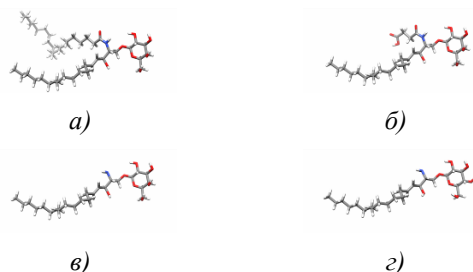


Рис. 1. Пространственные структуры β -GalCer (а), WSA1 (б), WSA2 (в) и WSA3 (г), полученные в результате модификаций α -GalCer.

3.2 МД расчеты и кластеризация

Для получения данных о конформационной подвижности липидов мы провели МД расчеты β -GalCer и его аналогов в интервале 20 нс с последующим разделением расчетной траектории на 3 структурных кластера. На рисунке 2 представлены результаты этих расчетов на примере β -GalCer-WSA1.

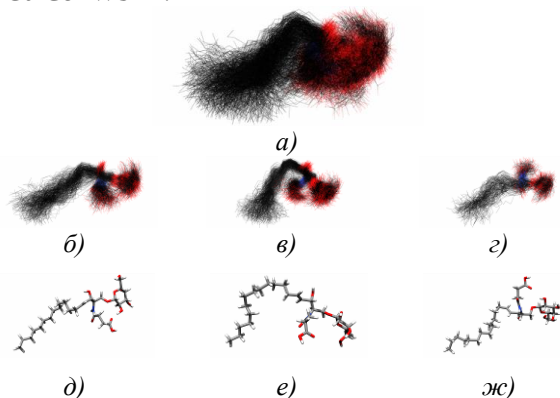


Рис. 2. (а) Структуры β -GalCer-WSA1, реализующиеся на его МД траектории. (б), (в), (г) – результаты деления МД траектории на три структурных кластера. (д), (е), (ж) – усредненные по каждому из кластеров конформации липида.

Аналогичная процедура, включающая последовательную реализацию МД расчетов и кластеризацию траектории, была использована для анализа конформационной подвижности двух пептидов петли V3.

3.3 Построение структурных комплексов и оценка свободной энергии взаимодействия гликолипидов с пептидами петли V3

Для каждой пары рецептор-лиганд проводили молекулярный докинг, после чего для каждого из комплексов генерировали МД траектории длительностью 2 нс, а затем применяли процедуру

ММ-PB/SA для оценки стабильности и определения величин свободной энергии их образования (таблица 1).

Таблица 1. Результаты расчетов свободной энергии образования комплексов пептидов петли V3 с гликолипидами.

Рецептор	Лиганд	$\langle \Delta G \rangle$, ккал/моль	$(\Delta G)_{\text{STD}}$, ккал/моль
Пептид GPGQAF	β -GalCer	-8,7	2,3
	β -GalCer-WSA1	-9,6	2,4
	β -GalCer-WSA2	-9,7	4,1
	β -GlcCer-WSA3	-6,3	2,6
Пептид GPGRAF	β -GalCer	-11,5	2,7
	β -GalCer-WSA1	-11,9	3,7
	β -GalCer-WSA2	-9,9	2,6
	β -GlcCer-WSA3	-7,8	3,8

В литературе приводятся значения свободной энергии образования комплексов белка gp120 с β -GalCer (-10,9 ккал/моль) и молекулой CD4 (от -11,9 до -11,3 ккал/моль). Сравнение этих экспериментальных величин с расчетными данными свидетельствует о том, что они имеют близкие значения для двух аналогов β -GalCer: WSA1 и WSA2.

4 Заключение

На основе анализа результатов расчетов установлено, что: (i) ключевую роль во взаимодействии гликолипидов с петлей V3 ВИЧ-1 играют остатки фенилаланина и аргинина (глутамин) иммуногенной короны петли V3; (ii) два теоретически сконструированных аналога WSA1 и WSA2 проявляют близкую к β -GalCer аффинность связывания с центральным фрагментом петли V3 и могут рассматриваться в качестве базовых структур для рационального дизайна новых эффективных противовирусных препаратов.

References

- [1] Andrianov, I. Anishchenko. In *Journal of Biomolecular Structure and Dynamics*, volume 27 (2), 179-193 (2009).
- [2] S. Bhat, S.L. Spitalnik, F. Gonzalez-Scarano, D.H. Silberberg. In *Proceedings of the National Academy of Sciences of USA* 88, 7131-7134 (1991).
- [3] O. Hartley, P.J. Klasse, Q.J. Sattentau and J.P. Moore. In *AIDS Res. Hum. Retroviruses*, volume 21 (2), pages 171-189 (2005).
- [4] Анищенко И.В., Андрианов А.М. Доклады Национальной академии наук Беларуси, 2010, Том 54, № 4, с. 66-74.

Молекулярно-динамическое моделирование механического разворачивания белков L и G

Глякина А.В.⁽¹⁾, Галзитская О.В.⁽²⁾, Балабаев Н.К.⁽¹⁾

Институт математических проблем биологии РАН, г. Пушкино⁽¹⁾, Институт белка РАН, г. Пушкино⁽²⁾
quark777@rambler.ru

Abstract

The mechanical properties of two immunoglobulin-binding domains of proteins L and G, which have the same three-dimensional structures, but differ in amino acid sequences, were investigated using the method of molecular dynamics.

1 Введение

Белки играют важную роль в различных биологических процессах. В то же время они являются весьма перспективными структурными и функциональными элементами будущих наноприборов. Функциональные возможности макромолекул в значительной степени определяются их механическими свойствами.

При растяжении белков с помощью атомно-силовой микроскопии было обнаружено, что β -структурные белки выдерживают существенно большие нагрузки, чем α -спиральные или α/β структурные белки. Считается, что упаковка элементов вторичной структуры белка является критическим фактором в определении его механической стабильности. Однако не существует простого соотношения между структурой белка и его механической стабильностью. Даже небольшие изменения в аминокислотной последовательности могут существенно повлиять на механические свойства белка. Детальное исследование путей силового разворачивания белка необходимо для понимания молекулярной природы его механической стабильности.

2 Материалы и методы

2.1 Объекты исследования

Объектами исследования являются два иммуноглобулинсвязывающих домена белков L (pdb код 2ptl, 63 аминокислотных остатка, 951 атом) и G (pdb код 1rgb, 56 аминокислотных остатков, 853 атома). Оба этих белка состоят из двух β -шпилек, расположенных на концах белка (N- и C-шпильки) и α -спирали между ними. Белки имеют одинаковые пространственные структуры (среднеквадратичное отклонение выравненных C α -атомов этих белков

составляет 1.38 Å), но по аминокислотной последовательности идентичны всего лишь на 15%.

2.2 Постановка вычислительного эксперимента

Исследование проводили с помощью метода молекулярной динамики с использованием программы ПУМА, разработанной в ИМПБ РАН. Решение системы классических уравнений движения атомов проводилось в полноатомном силовом поле AMBER-99 [1].

Для воды использовалась модель TIP3P, причем валентные связи и валентные углы не фиксировались, а задавались соответствующими потенциальными функциями.

Для поддержания постоянной температуры использовалась виртуальная столкновительная среда («столкновительный термостат») [2, 3].

Стартовые координаты атомов белков брались из банка белковых структур. Каждый белок помещался в параллелепипед, заполненный молекулами воды. Вода, которая стерически перекрывалась с атомами белка, удалялась. Таким образом, были получены гидратированные белки с 1772 молекулами воды в приповерхностном слое для белка L и с 1844 молекулами воды для белка G. Вся система белок+вода помещалась в сфероцилиндр с непроницаемыми отталкивающими стенками достаточно больших размеров, чтобы не влиять на динамику белка и прилегающую к нему воду и в то же время не позволять испарившимся молекулам воды уходить на «бесконечность», возвращая их в область моделирования.

В начальный момент времени всем атомам стартовой конфигурации системы приписывались случайные скорости, и проводилась релаксация системы при фиксированных положениях концевых атомов белка. Для каждого белка проводилась серия расчетов в независимых случайных столкновительных средах. Отрелаксированные таким образом в течение 50 пс системы были стартовыми для последующих вычислительных экспериментов, в которых использовалось дополнительное внешнее воздействие, приводящее к равномерному увеличению расстояния между концевыми атомами белков (в белке L концевыми были атомы 1N и 945C α , а в белке G – атомы 1N и 839C α).

Для каждого белка было проведено по 24 независимых вычислительных эксперимента, различающихся начальными данными (координатами и скоростями атомов) и скоростями, с которыми изменялись расстояния между концевыми атомами. Было взято три значения скорости растяжения $v=0.125, 0.0625$ и $0.005 \text{ \AA/пс}$. Моделирование проводилось при температуре 350 К.

3 Результаты

Исследование зависимости силы от расстояния между концами белков показало, что все полученные траектории можно разделить на 3 группы: траектории, на которых появляется один (в 6 случаях для белка L и в 8 для белка G), два (12 – L, 13 – G) и три пика силы (6 – L, 3 – G). Причем, два пика силы как для белка L, так и для белка G, наблюдаются в половине случаев. Наличие нескольких пиков силы на траектории говорит о том, что в процессе механического разворачивания белков появляются промежуточные состояния.

Экспериментальные работы по растяжению этих белков в атомно-силовой микроскопии [4, 5, 6] и теоретические работы [4, 7], в которых используются грубые модели белков, либо же растровитель учитывается неявно, говорят о том, что эти белки разворачиваются по простому одностадийному механизму. Но сейчас уже появились экспериментальные работы [8], по разворачиванию белка L с помощью денатуранта, в которых указывается на то, что процесс разворачивания данного белка не описывается простой одностадийной моделью.

Наличие нескольких пиков силы на траектории не зависит от величины скорости растяжения. При каждой из использованных в данной работе скоростей растяжения есть траектории, на которых появляются как один, так и два и три пика силы. От величины скорости растяжения зависит только высота пика силы, и естественно, что чем больше величина скорости растяжения, тем выше пик силы. Это объясняется тем, что при более высокой скорости растяжения не все возникающие напряжения успевают релаксировать и происходит их наложение, что в свою очередь приводит к увеличению высоты пиков.

Было обнаружено, что при двух самых высоких скоростях, оба белка характеризуются близкими значениями величины средней максимальной силы, которая возникает при таком механическом разворачивании белков, различия становятся заметны только для наименьшей из использованных скоростей растяжения. Для того чтобы развернуть белок G потребовалась большая сила, чем та, которая возникает при растяжении белка L. Из этого следует, что белок G более механически стабилен,

чем белок L. В то же время, изменения расстояния между концами белка, при которых наблюдается максимум силы, для белка L в два раза больше, чем для белка G, и работу, которую требуется совершить, чтобы развернуть эти белки на начальном этапе для белка L в 1.5 раза больше, чем для белка G.

Было выявлено три возможных способа разворачивания белков L и G. Во всех случаях белки сначала разрываются на две структурные единицы. В первом случае этими структурными единицами являются [N-шпилька+ α -спираль] и [C-шпилька], во втором – [N-шпилька] и [C-шпилька] (α -спираль не примыкает ни к одному из этих блоков) и в третьем – [N-шпилька] и [α -спираль+C-шпилька]. Причем, первый способ наблюдается наиболее часто, а третий способ наблюдался в единичном случае (как для белка L, так и для белка G) только для наименьшей из взятых скоростей растяжения ($0.005 \text{ \AA/пс}$).

References

- [1] J. Wang, A. Cieplak, P.A. Kollman. How Well a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comp. Chem.*, volume 21, pages 1049-1074, 2000.
- [2] A.S. Lemak, N.K. Balabaev. A Comparison between Collisional Dynamics and Brownian Dynamics. *Mol. Simul.*, volume 15, pages 223-231, 1995.
- [3] A.S. Lemak, N.K. Balabaev. Molecular Dynamics Simulation of Polymer Chain in Solution by Collisional Dynamics Method. *J. Comp. Chem.*, volume 17, pages 1685-1695, 1996.
- [4] D.J. Brockwell, G.S. Beddard, E. Paci, D.K. West, P.D. Olmsted, D.A. Smith, S.E. Radford. Mechanically unfolding the small, topologically simple protein L. *Biophys. J.*, volume 89(1), pages 506-519, 2005.
- [5] Y. Cao, H. Li. Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nature Mater.*, volume 6, pages 109-114, 2007.
- [6] Y. Cao, H. Li. How Do Chemical Denaturants Affect the Mechanical Folding and Unfolding of Proteins? *J. Mol. Biol.*, volume 375, pages 316-324, 2008.
- [7] J. Shimada, E.I. Shakhnovich. The Ensemble Folding Kinetics of Protein G from an All-Atom Monte Carlo Simulation. *PNAS*, volume 99, pages 11175-11180, 2002.
- [8] S.A. Waldauer, O. Bakajin, T. Ball, Y. Chen, S.J. Decamp, M. Kopka, M. Jäger, V.R. Singh, W.J. Wedemeyer, S. Weiss, S. Yao, L.J. Lapidus. Ruggedness in the folding landscape of protein L. *HFSP Journal*, volume 2, pages 388-395, 2008.

Аминокислотная последовательность и конформационная структура инцептина гороха огородного *Pisum sativum*

Голубович В.П., Соколов Ю.А., Фигловский В.А., Шутова И.В.
Институт биоорганической химии НАН Беларуси, Минск, Беларусь
golubovich@iboch.bas-net.by

Amino acid sequence and conformational structure of garden pea (*Pisum sativum*) inceptin

Golubovich V.P., Sokolov Yu.A., Figlovsky V.A., Shutova I.V.

Abstract

Amino acid sequence identification of garden pea (*Pisum sativum*) inceptin has been carried out by using BLAST and ALIGN algorithms. The peptide is a fragment of chloroplastic ATP synthase γ -subunit (cATPC). It was imported from GenBank database. Conformational structure of the inceptine has been calculated by molecular mechanics and quantum chemistry methods.

1 Введение

В последние годы были предложены различные гипотезы относительно защитных механизмов растений в ответ на атаку вредных насекомых и проведены исследования по их проверке [1-4]. В основе этих механизмов лежит понятие о пептидных элиситорах (инцептинах) – сигнальных пептидных веществах, которые специфично предупреждают растение об опасности, после чего оно запускает свои защитные биохимические реакции, которые могут снизить уровень будущих последствий такого нападения. Растения могут откликаться на угрозу либо прямой защитной реакцией, вырабатывая вещества, которые токсичны для чужеродного организма (например, ингибиторы протеиназ насекомых могут действовать как антипищевая защита, нарушая процесс пищеварения насекомых) и тем самым угрожая численности вредителей, либо непрямой защитой путём налаживания выгодных взаимодействий с полезными для растения и вредными для насекомого организмами, индуцируя, например, эмиссию летучих веществ, которые могут привлекать естественных врагов вредителей.

До настоящего времени выделены и идентифицированы ряд инцептинов [3,4], предупреждающих некоторые растения о нападении гусеницы травяной совки (*Spodoptera frugiperda*). Так для коровьего гороха (*Vigna unguiculata*) выделены и идентифицированы три инцептина ICDINGVCVDA, EICDINGVCVDA, GEICDINGVCVDA, для кукурузы (*Zea mays*) - ICDVNGVCVDA, EICDVNGVCVDA, GEICDVNGVCVDA, для фасоли (*Phaseolus vulgaris*) - ICDVNGVCIDA, EVCDINGVCIDA, GEVCDINGVCIDA, для сои (*Glycine max*) - ICDVNGVCVDA, EICDVNGVCVDA, GEICDVNGVCVDA, для шпината (*Spinacea oleracea*) - ICDINGKCVDA, EICDINGKCVDA, GEICDINGKCVDA. Заметим, что все перечисленные инцептины имеют кольцевую структуру за счёт цистеинового дисульфидного мостика, а также имеют соответственно 11, 12 и 13 аминокислотных остатков. Показано [3,4], что инцептины являются фрагментами протеолитического расщепления протеазами травяной совки регуляторной области γ -субъединицы хлоропластической АТФ-синтазы (сАТРС) растения.

Понимание биохимических механизмов защиты растений от вредных насекомых рассматривается как важнейшее звено для разработки методов борьбы с потерями сельскохозяйственных культур. Поэтому обнаружение и идентификация новых инцептинов для различных растений, исследование их структуры имеет важное научное и практическое значение, открывает дорогу разработке методов их модификации с целью улучшения видовых защитных свойств растений.

Целью настоящей работы являлась теоретическая идентификация инцептинов, предупреждающих горох огородный (*Pisum sativum*) о нападении гусеницы травяной совки (*Spodoptera frugiperda*), и исследование конформационных особенностей и электронной структуры основного из этих инцептинов.

2 Методы и результаты исследования

Теоретическая идентификация была проведена путём сравнительного анализа последовательностей уже идентифицированных инцептинов с

соответствующей частью регуляторной области γ -субъединицы хлоропластической АТФ-синтазы (сАТРС) гороха огородного (*Pisum sativum*), имеющейся в базах данных GenBank [5]. При этом использовались общедоступные программные реализации алгоритмов выравнивания первичных последовательностей BLAST и ALIGN [6]. В результате были идентифицированы следующие фрагменты ICDINGNCVDA, EICDINGNCVDA, GEICDINGNCVDA, имеющие также соответственно 11, 12, 13 аминокислотных остатков и которые высоко гомологичны упомянутым выше инцептинам для других растений. Легко заметить, что инцептиновые последовательности для различных растений отличаются только одним или двумя аминокислотными остатками. Это позволяет предположить, что найденные структуры действительно являются структурами инцептинов гороха огородного (*Pisum sativum*).

Конформационный анализ основного инцептина гороха огородного (*Pisum sativum*) ICDINGNCVDA проводился с помощью процедуры, включавшей конформационный расчет с помощью метода атом-атомных потенциалов [7], в результате чего был найден набор из 26 устойчивых конформаций. Затем было осуществлено их квантово-химическое уточнение методом РМЗ[8] (шаг по углам вращения равен 5°). Это позволило сузить набор устойчивых конформаций до трёх наиболее предпочтительных. Заметим, что по сравнению с конформацией №1, конформация №2 менее предпочтительна на $\Delta E = 6,6$ ккал/моль, №3 – на $8,4$ ккал/моль. Это означает, что населенность конформации №1-преобладающая (более 99%). Поэтому было сделано предположение, что конформация №1, по сравнению с другими, является биологически активной. Результаты расчета показали также, что конформация №1 отличается от менее предпочтительных конформаций №2 и №3, в основном, двугранными углами $\phi_2, \phi_5, \psi_5, \chi(2)_5, \chi(3)_2, \phi_6, \psi_7, \psi_8$ соответственно аминокислот C, N, G, V, C этого пептида.

Для всех трёх конформаций методом РМЗ были рассчитаны различные характеристики электронной структуры. Найдено, что распределения

эффективных зарядов для конформаций №1 – №3 достаточно близки друг к другу. Обнаружено, что предполагаемая биологически активная конформация №1 по сравнению с конформациями №2 и №3 имеет меньший дипольный момент D [Дебай], энергию высшей заполненной молекулярной орбитали EHOMO [Эв], которая коррелирует с потенциалом ионизации молекулы и характеризует её электроно-донорные свойства, а также заметно большее значение энергии электронной активации $GAP = ELUMO - EHOMO$ ($ELUMO$ [Эв] – энергия низшей свободной молекулярной орбитали, коррелирующая со сродством к электрону молекулы и характеризующая электроноакцепторные свойства), что может означать значительно менее выраженную реакционную способность, а также менее выраженные электронодонорные свойства конформации №1 по сравнению с №2 и №3.

Литература

- [1] Mackey, D. Plant responses to insect herbivory / D. Mackey [et al.] // Cell. – 2003. – Vol. 112. – P. 379-389.
- [2] Howe, G.A. Jasmonics as signals in the wound response / G.A. Howe // J. Plant Growth Regul. – 2004. – Vol. 23. – P. 223-237.
- [3] Smeltz, E.A. Fragments of ATP synthase mediate plant perception of insect attack / E.A Smeltz [et al.] // Proc. Natl. Acad. Sci. USA. – 2006. – Vol. 103. – P. 8894-8899.
- [4] Smeltz, E.A. Cowpea chloroplast ATP synthase is the source of multiple plant defense elicitors during insect herbivory / E.A Smeltz [et al.] // Plant Physiol. - Vol. 144. – P. 793-805.
- [5] <http://www.genome.ad.jp>.
- [6] <http://www.ncbi.nlm.nih.gov/blast>.
- [7] Ахрем А.А., Галактионов С.Г., Голубович С.Г. Конформации биогенных аминов. – Минск: Наука и техника, 1979. - 176 с.
- [8] Stewart, J.J.P. New quantum chemical method PM3. / J.J.P. Stewart // J. Comp. Chem. - 1989. - Vol. 10. - P. 209 – 227.

Modeling of Ribosome Movement Along Molecular Chain

Aznakayev E.G., Aznakayeva D.E.

National Aviation University, Kiev, UKRAINE

aznakayev1@yahoo.com

Abstract

Process of ribosome movement along an one-dimensional biological chain is modeling. Transition of a ribosome on other site of one-dimensional biological chain is estimated. This can mismatch the copied molecular chain sequence.

The recognized mechanism of the genetic information reading from DNA and RNA molecules assumes consecutive reading of the genetic information along a molecular chain of nucleotides. The similar mechanism is supposed for molecular chain of proteins also, where amino acids are chain elements. Linear process of information reading thus is supposed. For example, at information reading from RNA molecule the ribosome moves along a molecular chain of nucleotides and synthesizes a corresponding molecular chain from amino acids forming proteins necessary for cell activity (Fig. 1, 2). Thus, in case of correct primary sequence of elements in a forming secondary molecular chain the correct sequence of elements is synthesized. And the genetic information is not deformed. Thus, we have unequivocal conformity of elements in initial and copied molecular sequences.

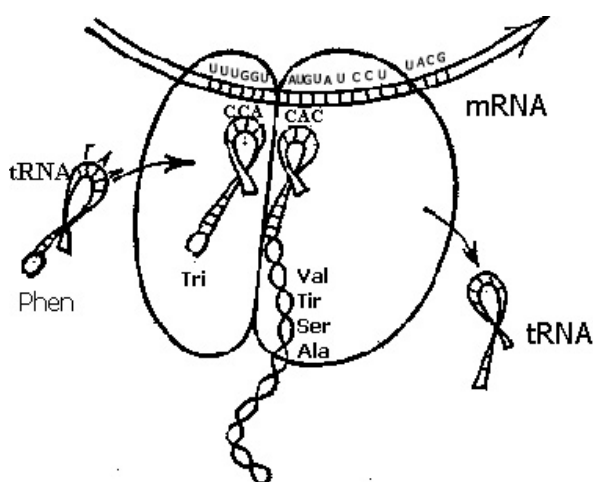


Fig. 1. Scheme of Amino acids–tRNA complex attaching to mRNA in a ribosome

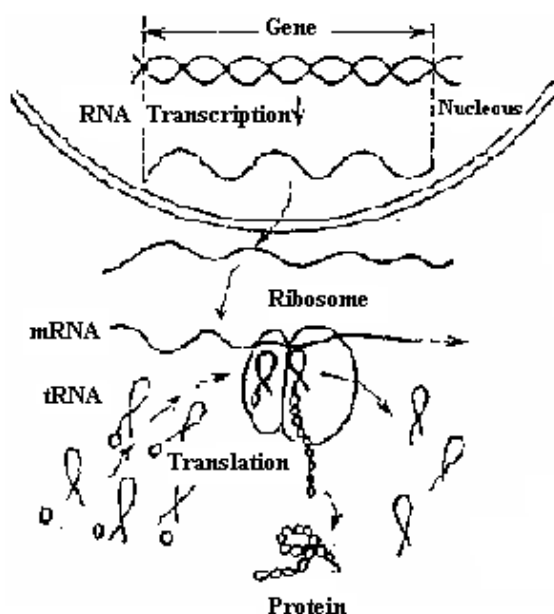


Fig. 2. A general arrangement of protein synthesis

However, in a reality the structure of biological molecular chains has complex 3D structure with set of bindings and crossings of molecular linear chains among themselves. Thus, in such binding and crossing points of initial one-dimensional chain is possible the rule infringement of linear genetic information reading and transitions of molecular structure for information reading, for example, ribosome on a new branch of an initial molecular chain. This can mismatch the copied molecular chain sequence. Such mechanism of genetic information reading can result both in losses of a part of the genetic information, and to numerous repetitions of identical elements in a copied molecular chain.

Thus, at such mechanism of genetic information reading the uncorrect genetic information reception from correct sequence is possible.

The numerous identical sites repetitions in biological sequences, received in experiments, can be caused not their real arrangement, but with such mechanism of ambiguous genetic information reading from initial biological sequence. Different stages of a ribosome movement along mRNA sequence near the crossing (transition) points of different parts of linear sequence of mRNA according to its 3D structure are

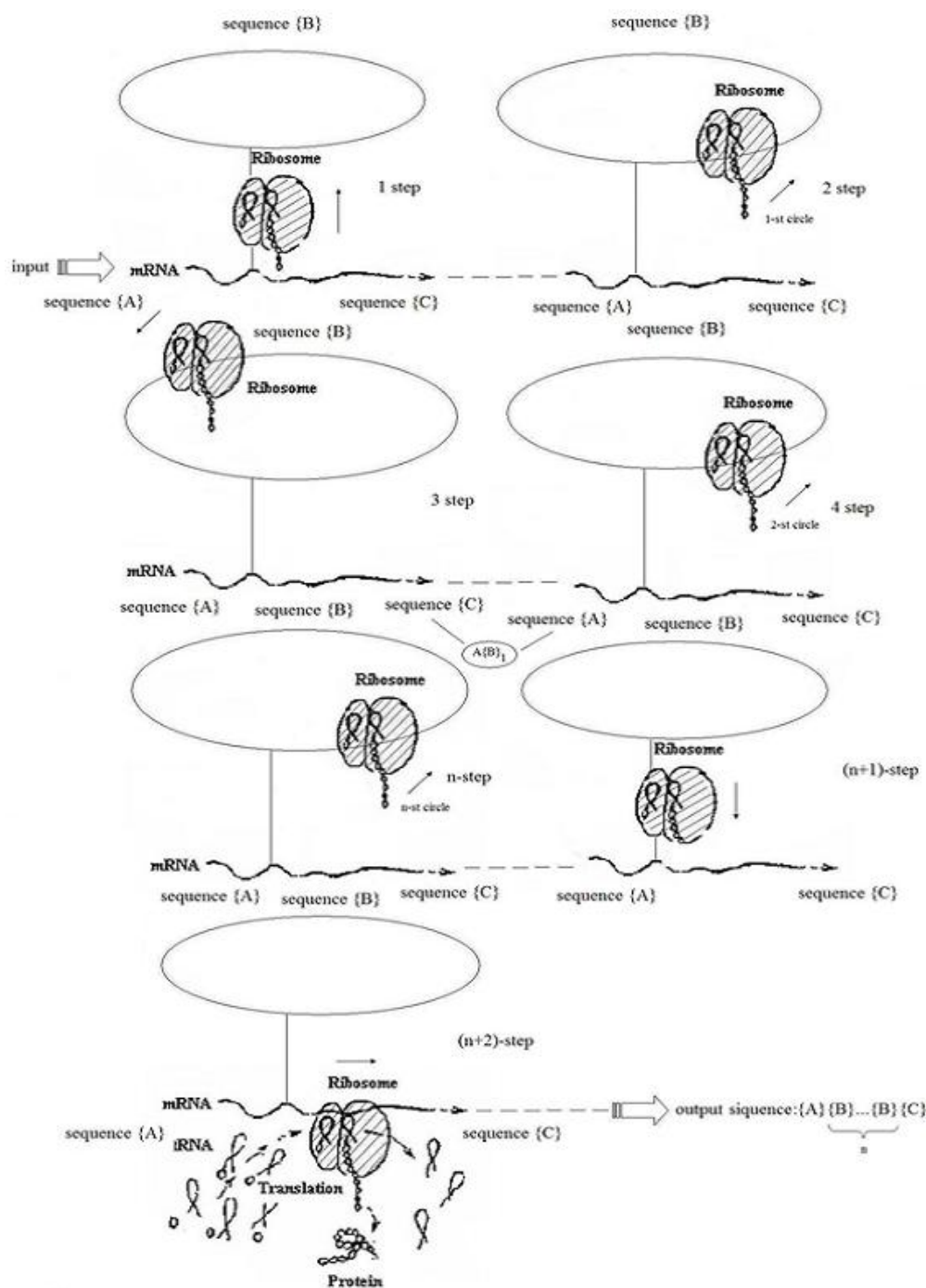


Fig. 3

shown in the Figure 3 and other below. At this ribosome movement it is made information reading about the linear nucleotides sequence in mRNA.

Conclusion

Structure of different biological molecular chains has a complex 3D structure with set of binding and crossing points of molecular linear chains. Thus, in such binding and crossing points of initial one-dimensional

chain is possible the rule infringement of linear genetic information reading and transitions arise on crossing parts of molecular structure for information reading, for example, ribosome on a new branch of an mRNA molecular chain. We suggest that such new molecular mechanism of genetic information reading can result both in losses of a part of the genetic information, and to numerous repetitions of identical elements in a copied molecular chain.

Исследование взаимодействий липидов в смешанных бислоях: моделирование методом молекулярной динамики

В.В.Корнилов, Н.К.Балабаев

Институт математических проблем биологии РАН, г. Пущино
basil@psn.ru

Abstract

Detailed molecular dynamics simulations performed to study the intermolecular interactions in hydrated bilayers containing three-component mixtures of saturated phosphatidylcholines (PC), unsaturated PC, and cholesterol.

1 Введение

Молекулярный состав липидных бислоев биологических мембран весьма разнообразен. Их основными составляющими являются фосфолипиды, различающиеся полярными головными частями, а также длиной и насыщенностью своих углеводородных цепей, и молекулы холестерина [1]. Особо важную роль в обеспечении нормального функционирования биомембран играют фосфолипиды с ненасыщенными, содержащими двойные связи, углеводородными цепями [2, 3]. Обычно углеводородные цепи фосфолипидов содержат от 12 до 24 атомов углерода и до 6 двойных связей C=C, преимущественно *cis*-конфигурации, при этом как правило только одна из двух углеводородных цепей является ненасыщенной.

Другим распространенным компонентом природных мембран различных животных клеток является холестерин. Его содержание может достигать 40%. Молекула холестерина состоит из компактного жесткого гидрофобного ядра с полярной головной частью (гидроксильной группой) и короткой углеводородной цепи. Присутствие холестерина является необходимым условием нормального функционирования биомембран. Известно, что холестерин оказывает существенное влияние на свойства липидной матрицы, увеличивая жидкость углеводородных цепей липидов при температурах ниже точки фазового перехода гель – жидкий кристалл, и уменьшая её при температурах выше этой точки [4-6].

Вопросы межмолекулярного взаимодействия липидов в мембране и влияния отдельных липидных компонентов на физико-химические

свойства липидной матрицы в целом являются объектом активного изучения, однако полного понимания механизмов влияния химического строения липидов на свойства фосфолипидных мембран ещё не достигнуто. Ценным инструментом для получения информации о взаимодействии липидов в бислоях является компьютерное моделирование биомембранных систем.

В данной работе методом молекулярной динамики (МД) проведено моделирование совокупности гидратированных липидных бислоев, состоящих из молекул дистеароил-фосфатидилхолина (18:0/18:0), молекул ненасыщенных фосфатидилхолинов (ФХ), содержащих от 1 до 6 двойных связей C=C в одной из своих углеводородных цепей (а именно, молекул 18:0/18:1(*n*-9)*cis* ФХ, 18:0/18:2(*n*-6)*cis* ФХ, 18:0/18:3(*n*-3)*cis* ФХ, 18:0/20:4(*n*-6)*cis* ФХ и 18:0/22:6(*n*-3)*cis* ФХ), и холестерина. Проведено сравнение результатов, полученных для трёх типов различных систем: а) чистых гомогенных бислоев ФХ различной степени ненасыщенности (6 разных бислоев), б) тех же бислоев ФХ, содержащих 40 мол. % холестерина, в) бислоев, содержащих тройные смеси из молекул дистеароил-фосфатидилхолина (30 мол. %), холестерина (40 мол. %) и 30% молекул ФХ различной степени ненасыщенности.

2 Модель и метод расчёта

2.1 Модель

Моделирование чистых гомогенных бислоев ФХ и бислоев с холестерином проводилось с использованием однотипной методики, подробно описанной в работах [7-10]. Моделируемые системы размещались в расчётных ячейках, имеющих форму прямоугольного параллелепипеда с периодическими по *X*, *Y* и *Z* граничными условиями (рис. 1). Ячейка содержала либо только 96 одинаковых молекул ФХ определённого вида (48 на монослой) – для чистых бислоев, либо еще 64 молекулы холестерина (32 на монослой, т.е. 40 мол. %), а также воду из расчета 24 молекулы H₂O на каждую молекулу липида (ФХ или холестерина). Первоначальная упаковка

молекул липидов в бислоях была гексагональная, молекулы липидов образовывали кристаллоподобные конфигурации с вытянутыми вдоль оси Z, нормали к поверхности бислоя, углеводородными цепями в *транс*-конформациях. Молекулы холестерина были распределены в бислоях среди молекул ФХ равномерно, таким образом, чтобы в среднем на каждые три молекулы ФХ приходилось по две молекулы холестерина.

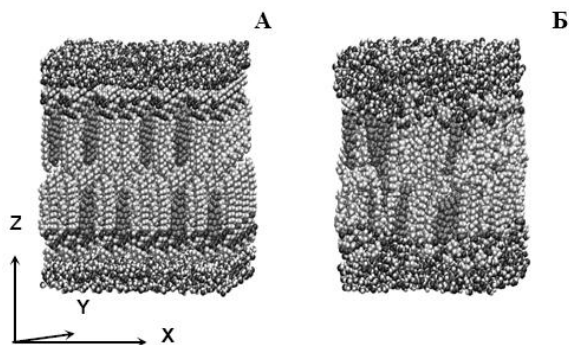


Рис. 1. Исходная конфигурация расчетной ячейки гидратированного липидного бислоя 18:0/22:6(n-3)cis ФХ с 40 моль% холестерина (А), и типичная конфигурация системы в процессе построения траектории методом молекулярной динамики (Б). Использовано объемное представление атомов всех молекул, атомы молекул холестерина показаны темным цветом.

При подготовке трехкомпонентных бислоев расстановка молекул липидов в каждом из монослоев, первоначально разбитом в соответствии с числом молекул в нем на 80 одинаковых прямоугольных ячеек, осуществлялась следующим образом: сначала в свободные ячейки случайным образом расставлялись молекулы ФХ первого типа (18:0/18:0), затем второго (одного из набора 18:0/18:0 ÷ 18:0/22:6), затем холестерин (последний размещался со сдвигом молекулы по оси Z, с тем чтобы гидроксильная группа молекулы холестерина располагалась на уровне глицериновой развилки молекул ФХ). При этом каждая из молекул дополнительно поворачивалась вокруг оси Z на случайный угол относительно своего исходного положения. Отдельно приготавливался слой воды из расчёта 30 молекул воды на 1 молекулу липида. Далее выполнялась релаксация гидратированного бислоя с использованием методов управляемой молекулярной динамики.

2.2 Метод расчёта

Потенциальная энергия каждого бислоя была представлена в виде суммы атом-атомных потенциальных функций невалентных взаимодействий (ван-дер-ваальсовых и

кулоновских) и энергий валентных связей, валентных углов, торсионных углов, неплоских отклонений атомов, примыкающих к двойным связям в цепях.

Моделирование молекулярной динамики осуществлялось при постоянном давлении (NPT-ансамбль) с использованием баростата Берендсена и столкновительного термостата. Моделирование осуществлено на основе программного комплекса PUMA (ИМПБ РАН, г. Пушкино).

3 Результаты

Вычислен ряд структурных и динамических характеристик смешанных фосфолипидных бислоев с холестерином. Определены профили плотности масс различных групп атомов, профили параметров порядка связей C-H и C-C углеводородных цепей молекул ФХ, коэффициенты латеральной самодиффузии молекул фосфолипидов и холестерина. Проведен анализ свойств бислоев в зависимости от степени ненасыщенности углеводородных цепей ФХ и наличия молекул холестерина в бислое. Проведено комплексное сравнение характеристик трёхкомпонентных бислоев с аналогичными данными для гомогенных бислоев ФХ и их смесей с холестерином. Полученные результаты способствуют развитию представлений о роли гетерогенности липидных молекул в биологических мембранах.

Ссылки

- [1] Геннис Р. Биомембраны: Молекулярная структура и функции. М.: Мир, 1997.
- [2] Рабинович А.Л., Рипатти П.О. *Успехи совр. биологии*, т. 114. № 5, с. 581-594, 1994.
- [3] Valentine R.C., Valentine D.L. *Prog. Lipid Res.*, v. 43, p. 383-402, 2004.
- [4] Ohvo-Rekilä H., Ramstedt B., Leppimäki P., Slotte J.P. *Progr. Lipid Res.*, v. 41, p. 66-97, 2002.
- [5] McConnell H.M., Radhakrishnan A. *Biochim. Biophys. Acta*, v. 1610, p. 159-173, 2003.
- [6] Bach D., Wachtel E. *Biochim. Biophys. Acta*, v. 1610, p. 187-197, 2003.
- [7] Rabinovich A.L., Ripatti P.O., Balabaev N.K., Leermakers F.A.M. *Phys. Rev. E.*, v.67, No.1, p. 011909_1-14, 2003.
- [8] Рабинович А.Л., Рипатти П.О., Балабаев Н.К. *Журн. физ. химии*, т. 78, вып. 7. с. 1160-1165, 2004.
- [9] Корнилов В.В., Рабинович А.Л., Балабаев Н.К., Бессонов В.В. *Биофизика*, т. 53, вып. 1, с.84-92, 2008.
- [10] Корнилов В.В., Рабинович А.Л., Балабаев Н.К. *Биофизика*, т. 53, вып. 2, с.250-260, 2008.

Статистический анализ и предсказание неструктурированных остатков в белковых структурах

М.Ю.Лобанов, О.В.Галзитская*

Институт белка РАН, г. Пушкино

*ogalzit@vega.protres.ru

Abstract

Получены данные по статистике встречаемости неструктурированных аминокислотных остатков на концах и в средней части белковой цепи. Вычислены частоты встречаемости неструктурированных остатков каждого из 20 типов в различных позициях белковой цепи. Создан алгоритм, основанный на модели Изинга, для предсказания неструктурированных остатков. Тестирование на двух дополнительных базах показало, что наш метод – один из лучших.

1. Введение

Предсказание структуры и функции белков является одним из главных направлений в структурной геномике. Особый интерес представляет предсказание так называемых неструктурированных участков белковой цепи (участков, не имеющих в нативном состоянии белка фиксированной пространственной структуры). Такие неструктурированные участки часто играют важную функциональную роль (см. обзоры [1, 2]). При этом неструктурированные участки структурируются только тогда, когда связываются с другой молекулой. Предполагают, что отсутствие глобулярной структуры при физиологических условиях представляет значительное функциональное преимущество для нативно-развернутых белков, так как их большая доступная поверхность при небольшом размере белка и пластичность позволяет им более эффективно взаимодействовать с белками и нуклеиновыми кислотами по сравнению с глобулярными белками того же размера, обладающими ограниченной конформационной гибкостью [1].

Поскольку неструктурированные участки белковой цепи играют важную роль в функционировании белка, то их предсказанию уделяется большое внимание. На сегодня для этих целей разработаны специализированные программы, такие как FoldUnfold, PONDR, RONN, DisEMBL, PreLINK, IUPred, GlobPlot, FoldIndex и другие. Исходя из принципа, лежащего в основе их работы, все эти программы можно разделить на две

группы. Программы FoldUnfold, PONDR, IUPred, GlobPlot, PreLINK и FoldIndex предсказывают неструктурированные участки белковой цепи, основываясь на физико-химических свойствах аминокислотных остатков в белке. В роли такого свойства может выступать локальный аминокислотный состав и гидрофильность (PONDR), число ожидаемых контактов (FoldUnfold), способность участка цепи образовывать гидрофобный кластер (PreLINK) или оценка энергетического взаимодействия между соседними аминокислотными остатками (IUPred). Вторая группа программ использует выравнивания гомологичных белковых последовательностей. Программа RONN использует нейронную сеть и сравнивает данную последовательность с рядом последовательностей, которые заранее отнесены к структурированным, к неструктурированным или к смеси того и другого. Программа DisEMBL использует нейронную сеть, натреннированную на рентгеноструктурных данных. В данной работе создан алгоритм, основанный на модели Изинга, для предсказания неструктурированных остатков. Параметры для программы получены и оптимизированы из статистики белковых структур.

2. Результаты

Под неструктурированными остатками мы понимаем остатки неразрешённые методом рентгеноструктурного анализа. А точнее остатки у которых не разрешён C_{α} атом. Рассматривались все белковые структуры, решённые методом рентгеноструктурного анализа, с разрешением лучше 3Å, опубликованные до 20-12-2008. Все 100% гомологи были сгруппированы; для них данные были усреднены. Полученная база содержит 28 727 уникальных цепей и 7 487 366 остатков. Из этих остатков 4.6% – неструктурированные. Используя эту базу, получены все статистические результаты и потенциалы. Тестовая база составлена по тем же принципам, что и основная база, но взяты белковые структуры, опубликованные после 20-12-2008. База используется для оценки созданного нами метода предсказания неструктурированных остатков. База CASP8: 8th Critical Assessment of Techniques for Protein Structure Prediction. В 2008 году проходило соревнование по предсказанию

пространственной структуры белков, в том числе по предсказанию неструктурированных остатков [3]. Эта база белков используется для сравнения наших результатов с результатами других научных групп. Из статистики встречаемости неструктурированных остатков нами показано, что на краях остатки чаще бывают развёрнуты. Чтобы результаты на краях и в центре были сопоставимыми, мы нормировали долю неструктурированных остатков конкретного типа на общую долю неструктурированных остатков. Здесь важно отметить, что любой хороший метод предсказания должен учитывать специфические особенности наблюдаемые на краях.

В нашей работе мы использовали алгоритм, основанный на модели Изинга. Каждый остаток может быть в двух состояниях: структурированном и неструктурированном. Энергия каждого остатка в том или ином состоянии зависит от типа остатка и состояния. Энергия полностью структурированного состояния была взята за ноль. Используя созданную базу, подобраны оптимальные параметры предсказания. Кроме того, мы ввели энергию границы между структурированными и неструктурированными остатками. Всего у нас рассматривалось 23 параметра: 20 значений для аминокислотных остатков быть в развёрнутом состоянии, отдельно энергии для N- и C-концов и одна энергия границы.

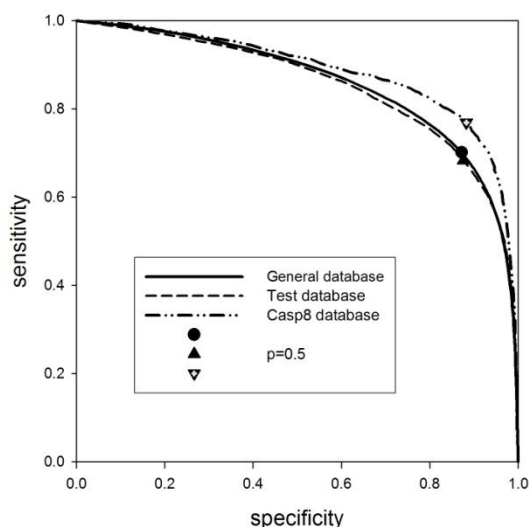


Рис. 1. ROC кривые для предсказания неструктурированных остатков нашим методом для разных баз данных.

В дальнейшем мы подсчитывали статусму всех возможных состояний цепи. А так же статусму всех состояний при заданном состоянии одного остатка. Разделив одно на другое мы получаем вероятность пребывания остатка в этом состоянии.

Мы использовали стандартные критерии оценки качества предсказания. Чувствительность (доля правильно предсказанных неструктурированных остатков), специфичность (доля правильно предсказанных структурированных остатков).

Сумма этих параметров минус 1 даёт критерий Sw, используемый более 6 лет. Варьируя вероятность, выше которой мы считаем остаток неструктурированным можно получить разные пары чувствительности и специфичности, которые и показаны на рисунке 1. Площадь под кривой даёт новый критерий качества работы программы, AUC. Этот критерий тоже используется более 6 лет.

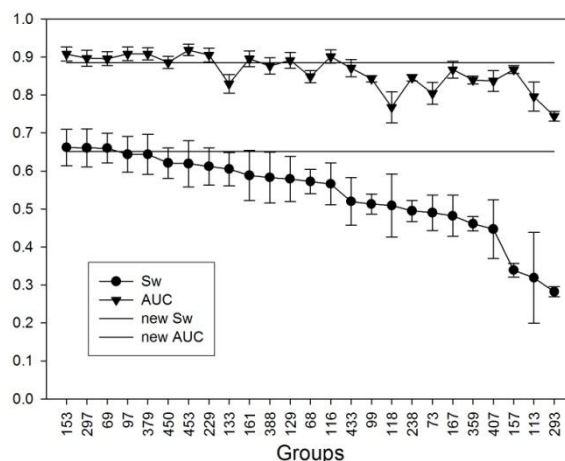


Рис. 2. Сравнение нашего метода с предсказаниями других групп, участвующих в CASP8 соревновании [3]. Горизонтальные линии показывают качество работы нашего метода.

Наши результаты можно сравнить с результатами других людей, полученных летом прошлого года. Как видно, по параметру AUC у нас результаты выше среднего, зато по Sw критерию мы одни из первых. Результаты работы нашей программы, полученные на разных базах, показывают, что наш метод даёт на тестовой базе тот же результат, что и на основной. То есть он не перетренирован.

Тестирование на двух дополнительных базах показало, что наш метод – один из лучших.

Работа выполнена при финансовой поддержке РФФИ (грант № 08-04-00561), при поддержке Российской академии наук (программа «Молекулярная и клеточная биология» и «Фундаментальные науки – Медицина»).

Ссылки

- [1] P.Tomba. Intrinsically unstructured proteins. *Trends Biochem. Sci.*, 27, 527-533, 2002.
- [2] P.E.Wright and H.J.Dyson. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, 293, 321-331, 1999.
- [3] O.Noivirt-Brik, J.Prilusky, J.L.Sussman. Assessment of disordered predictions in CASP8. *Proteins: Structure, Function, and Bioinformatics*, 77 (Suppl. 9), 210-216, 2009.

Анализ пространственной структуры глобулярных белков на основе спектрально-аналитического метода

Панкратов А.Н.⁽¹⁾, Руднев В.Р.⁽¹⁾, Куликова Л.И.⁽¹⁾, Дедус Ф.Ф.⁽¹⁾, Ефимов А.В.⁽²⁾

Институт математических проблем биологии РАН⁽¹⁾, Институт белка РАН⁽²⁾

pan@impb.ru

Abstract

Considered is the problem of recognition of the simplest super-secondary structures of proteins. Proposed is an approach to the problem of analysis of the spatial structure of proteins based on the description of the main chain of the polymer using a natural equation of the curve and the classical orthogonal bases.

Введение

Определение пространственной структуры белка является актуальной и трудоемкой задачей. Наиболее развитым и применяемым на практике является метод определения структуры на основе данных рентгеновского рассеяния. Более поздним и реже применяемым является метод определения структуры по спектрам ядерного магнитного резонанса. Теоретическое прогнозирование структуры белка оказывается возможным в среднем на 70%. Это связано с тем, что прогнозирование вторичной структуры белка (альфа и бета структур) по первичной структуре является некорректно поставленной задачей, поскольку вторичная структура зависит от третичной структуры.

Регулярные структуры (α и β) являются частью хорошо организованной пространственной структуры белка, образованной водородными связями, которые осуществляются между элементами цепи полимера в случае нативного состояния. В денатурированном (расплавленном) состоянии белок имеет другую структуру, в которой, по-видимому, все или почти все водородные связи между элементами цепи белка замещены связями с растворителем.

Значение вторичной структуры белка заключается в том, что, во-первых, возможность ее образования определяется локальными характеристиками белковой цепи и, во-вторых, она имеет относительно простую геометрическую форму. Однако, она является частью более сложной пространственной структуры, которая имеет нелокальный характер, поскольку отдаленные участки цепи полимера могут контактировать друг с другом в пространственной укладке.

Ранее были разработаны простые правила укладки белковой цепи [1], которые привели к построению структурных деревьев, которые описывают многообразие белковых структур. Как в случае с рентгеноструктурным анализом и ЯМР, дальнейшее развитие метода связано с разработкой

информационных технологий. В частности, необходимо автоматически распознавать структурные мотивы в известных структурах белков, что может способствовать построению соответствия между первичной и третичной структурой белка.

Постановка задачи

Целью данного исследования является разработка программно-алгоритмического комплекса для распознавания супервторичных структур в глобулярных белках с решенными 3-D структурами методами рентгеноструктурного анализа и ядерно-магнитного резонанса. Для достижения поставленной цели были поставлены и решены следующие задачи:

- 1) получены характеристические профили структур белков, представленных в формате PDB;
- 2) определены характеристические параметры поиска конкретных элементов супервторичных структур (длины α -спиралей, конформации перетяжек, торсионные углы, образуемые между α -спиралями, и иные геометрические особенности структуры белковой молекулы);
- 3) применен спектральный алгоритм поиска заданного паттерна супервторичной структуры в исследуемых глобулярных белках на примере α -а уголка с короткой перетяжкой.

Метод решения

Для получения уникальных белковых профилей проводилось аналитическое описание координат $C\alpha$ -атомов основной цепи белковой глобулы в виде естественного уравнения кривой, полученного с использованием методов интерполяции и аппроксимации. В качестве таких методов для получения аналитического описания кривой использовались сплайны и ортогональные многочлены. Естественное уравнение пространственной кривой включает в себя параметризованное описание кривизны и кручения в зависимости от естественного параметра кривой - длины дуги. Такое описание инвариантно по отношению к выбору декартовой системы координат. Кроме того, регулярные участки - α -спирали - представляются областями постоянных

значений кривизны и кручения. На рис.1 представлена типичная структура данного класса. На рис.2 и 3 представлены профили кривизны и кручения для этой структуры. По заданным профилям можно однозначно восстановить пространственную структуру образца с точностью до выбора системы отсчета декартовых координат.

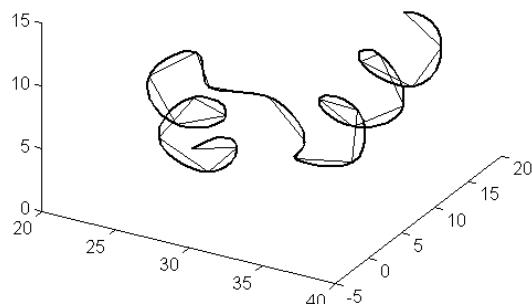


Рис.1 Супервторичная структура α - α уголок с короткой перетяжкой.

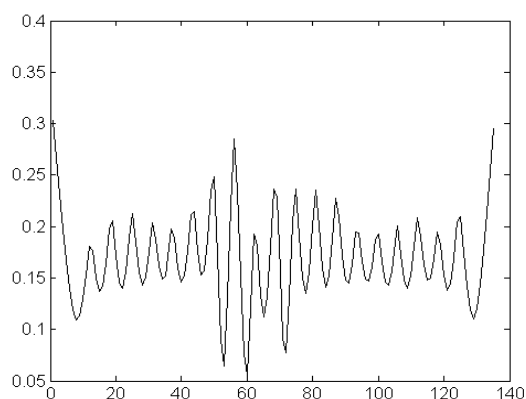


Рис.2 Профиль кривизны структуры α - α уголок с короткой перетяжкой.

Описание исследуемой простейшей структуры с помощью профиля кручения показывает, что в области перетяжки значение кручения отрицательное, в противоположность положительному значению кручения α -спиралей. Только такие структуры являются затравочными при образовании глобальной структуры белка.

Поиск образов супервторичных структур типа α - α уголков можно свести к спектральному алгоритму поиска неточных повторов в характеристических профилях [2]. Таким образом, в данной задаче методы приближенного аналитического описания применяются независимо дважды: во-первых, при

построении естественного уравнения кривой, во-вторых, при сравнении профилей аналогичных участков различных белков.

Заметим, что первая проблема является достаточно сложной, поскольку естественное уравнение чувствительно к третьей производной от координатных функций кривой.

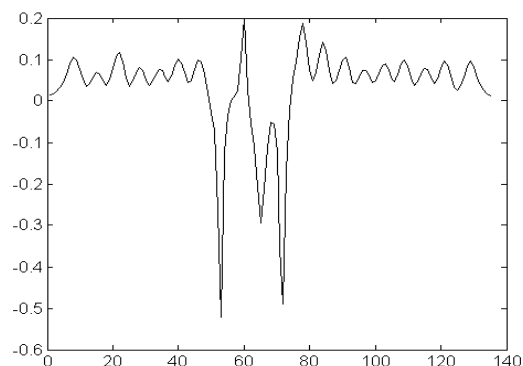


Рис.3 Профиль кручения структуры α - α уголок с короткой перетяжкой.

Результаты

В ходе проверки алгоритмов был проведён анализ различных структур белковых молекул из базы данных UniProt PDB на предмет поиска α - α уголков. Результат свидетельствует о том, что по построенным профилям возможно распознавать структуры типа α - α уголок с короткой перетяжкой. В дальнейшем планируется расширить область применения алгоритма, с возможностью распознавания широкого спектра супервторичных структур.

Работа выполнена при финансовой поддержке РФФИ, проекты № 08-07-00353, 09-07-12108, 10-01-00609, 10-07-00300.

Литература

- [1] Efimov A.V. Standard Structures in Proteins// Prog. Biophys. Molec. Biol., 1993, Vol.60, Pp.201-239.
- [2] Pankratov A.N., Gorchakov M.A., Dedus F.F., Dolotova N.S., Kulikova L.I., Makhortykh S.A., Nazipova N.N., Novikova D.A., Olshevets M.M., Pyatkov M.I., Rudnev V.R., Tetuev R.K., Filippov V.V. Spectral Analysis for Identification and Visualization of Repeats in Genetic Sequences// Pattern Recognition and Image Analysis, 2009, Vol.19, No.4, Pp.687-692.

Молекулярное моделирование трехмерной структуры белкового комплекса NS2B-NS3 протеазы высоковирулентных и инаппарантных штаммов вируса клещевого энцефалита

Потапова У.В.⁽¹⁾, Беликов С.И.⁽¹⁾, Феранчук С.И.⁽²⁾

ЛИН СО РАН⁽¹⁾, ОИПИ НАН Беларуси⁽²⁾
Shuana1983@yandex.ru

Abstract

The research is based on previously obtained data on the mutations in complete genomes of tick borne encephalitis virus strains (TBEV), isolated from patients with varying severity of disease. Three mutations previously identified led to substantial amino acid changes, which probably could affect the infection process severity. It is suggested that two associated mutations, deletion of amino acid 111 in the capsid protein C and substitution (Ser1534 - Phe) in the NS3 protein influence strictly coordinated polyprotein processing, disturbing correct arrangement of viral particles. This process can result in the development of defect viral particles, containing no RNA. To reach this aim, we constructed 3D structures of NS2B-NS3 protease in which mutations are important, by used a computer simulation method.

1 Введение

1.1 Актуальность проблемы

Клещевой энцефалит (КЭ) является наиболее значимой трансмиссивной природно-очаговой вирусной инфекцией лесной зоны Евразийского континента. Заболеваемость КЭ регистрируют более чем в 25 странах Европы и 7 странах Азии, в Российской Федерации (РФ) – на 46 административных территориях, в том числе на 18 территориях Сибири и Дальнего Востока [1].

2 Обоснование научной проблемы

2.1 Научный задел

Это исследование основано на предварительно полученных нами данных о мутациях в полных геномах вируса клещевого энцефалита (ВКЭ), изолированных от пациентов с различной тяжестью заболеваний.[5]

Из обнаруженной 21 мутации в полных геномах ВКЭ штаммов Дальневосточного субтипа, были найдены только три ключевые мутации, которые приводят к существенным изменениям аминокислот, которые, вероятно, могут влиять на тяжесть течения инфекционного процесса. Высказано предположение о том, что две сочетанные мутации, делеция 111-й аминокислоты в капсидном белке С и замена (Ser1534-Phe) в белке NS3, оказывают влияние на строго согласованный ход процессинга полипротеина, нарушая правильное формирование вирусных частиц, что может приводить к образованию дефектных вирусных частиц, не содержащих РНК. [5]

2.2 Выбор объекта исследования

Вирусный белок NS2B является основным кофактором NS3 протеазы. NS2B-NS3 протеаза необходима для ко- и пост-трансляционного процессинга вирусного полипротеина в индивидуальные функциональные белки.[2]

Мутации в NS2B-NS3 могут привести к появлению нежизнеспособных форм вирусных частиц.

3 Ход работы

3.1 Постановка задачи

Для подтверждения данной гипотезы мы построили серию трехмерных структур белков для штаммов ВКЭ, имеющих различную вирулентность и патогенность для человека.

3.2 Методы и алгоритмы

Для достижения этой цели, мы построили трехмерные структуры вирусного белка NS2B и притеазный домен вирусного белка NS3 для высоковирулентных и инаппарантных штаммов вируса клещевого энцефалита. Трехмерные модели были построены на основе метода гомологичного моделирования с использованием пакета программ Jackal, и веб-сервиса I-TASSER.[3,4] В качестве структурного шаблона для построения модели

белкового комплекса NS2B-NS3 протеазы, мы использовали NS2B-NS3 протеазу вируса Западного Нила (2FP7), являющуюся одним из ближайших гомологов для вируса клещевого энцефалита. После построения трехмерных моделей отдельных белков, была осуществлена проверка моделей с помощью метода молекулярной динамики, реализованного в программном пакете GROMMACS. После верификации трехмерных моделей белков, мы выполнили белок-белковый докинг для объединения моделей NS2B и NS3 протеазы в единый белковый комплекс. Белок-белковый докинг осуществлен с помощью программного пакета FlexX.

4 Результаты и обсуждение

4.1 Полученные результаты

Анализ трехмерных моделей показал значительные различия в конформации белков иннапарантных и высоковирулентных штаммов ВКЭ. По результатам проведенного исследования мы предполагаем что ранее найденная мутация в 1534 позиции (Ser-Phe), непосредственно влияет на формирование белкового комплекса и компактность белковой глобулы NS2B и NS3 протеазы.

4.2 Научная и практическая значимость

Практическая ценность данной работы заключается в возможности раннего прогнозирования тяжести заболевания среди населения во время сезонных вспышек инфекции, возможности выполнять картирование очагов

распространения особо опасных форм вируса в различных регионах страны. В будущем, нахождение связи между нуклеотидной последовательностью и патогенностью для человека различных штаммов ВКЭ позволяет разработать механизм молекулярного блокирования вируса, и как следствие эти данные могут быть использованы для молекулярного дизайна лекарств.

References

- [1] G.G.Onishchenko. Epidemiological Situation and the Strategy of Struggle against Tick Borne Encephalitis at the Present Stage, in Kleshchevoi i drugie virusnye entsefalit (Tick Borne and Other Virus Encephalitises), Ross. Akad. Med. Nauk, pages 5- 8, 2003
- [2] K.Wichapong, S.Pianwanit et al. Homology modeling and molecular dynamics simulations of Dengue virus NS2B/NS3 protease: insight into molecular interaction. *Molecular Recognition*, 23 (3), pages 283-300, 2009
- [3] Y.Zhang. Protein structure prediction: when is it useful?, *Structural Biology*, 19, pages 145-155, 2009
- [4] D.Petrey, Z.Xiang, B.Honig et al. Using multiple structure alignments, fast homology building, and energetic analysis in fold recognition and homology modeling, *Proteins: Struc., Func. and Genet.* 53, pages 430-435, 2003
- [5] С.И.Беликов, Г.Н.Леонова, И.Г.Кондратов, Е.В.Романова, Е.В.Павленко. Генетика микроорганизмов, том 46, выпуск (№) 3, с. 356-363, 2010.

PFFsubl и PFFol: новые поляризуемые атомные силовые поля для расчета невалентных взаимодействий в явно и неявно заданном водном окружении

А. В. Финкельштейн, Л. Б. Переяславец

Институт белка Российской Академии Наук, Пушкино, Московская обл., 142290

afinkel@vega.protres.ru⁽¹⁾

Цель нашего исследования – получение, в рамках одной и той же функциональной формы силового поля, параметров невалентных взаимодействий атомов и при явном (в силовом поле PFFsubl), и при неявном (в силовом поле PFFsol) учете водного окружения. Для получения параметров взаимодействия атомов при явном учете всех атомов (и белка, и лигандов, и растворителя) используют экспериментально полученные кристаллические структуры и энтальпии их сублимации. При неявном учете водного окружения атомы растворителя не рассматриваются, но необходима поправка энтальпии сублимации на рассчитанную с помощью констант Генри свободную энергию сольватации молекул. Обработав многие десятки структур кристаллов и термодинамические данные по их сублимации и растворению, мы получили параметры притяжения и отталкивания характерных для белка атомов (H, C, N, O, S) в различных ковалентно-связанных состояниях, а также параметры их

электростатических взаимодействий. Это сделано как в "вакуумном" (при явном учете атомов растворителя), так и в "водном" (при их неявном учете) окружении. Последний случай наиболее интересен, так как детальный расчет взаимодействий белка в воде, заданной в явном (молекулярном) виде, занимает очень много вычислительного времени. В обоих случаях все параметры ковалентных взаимодействий брали из силового поля ENCAD, а все парциальные заряды всех атомов рассчитывали квантово-механически. С оптимизированными параметрами, средняя погрешность расчета энергии связи молекул в кристалле и в "вакуумном" и "водном" окружении составила менее 10%.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, программы «Молекулярная и клеточная биология» Президиума Российской академии наук, Федерального агентства по науке и инновациям и Медицинского института Ховарда Хьюза.

Разработка математических методов и программных средств выявления функциональных сайтов белков

Шутова И.В., Голубович В.П.

Институт биоорганической химии НАН Беларуси, Минск

shutova2001@mail.ru

Abstract

With the purpose of detecting of protein's binding sites we have developed the program complex. It consists of three modules which allow building the molecular protein's surfaces, detecting the cavities and carrying out the analysis of hydrophobic areas and charge distribution on the surfaces.

1 Введение

Одной из актуальных задач молекулярной биологии и биоинформатики является распознавание сайтов связывания белков. Успешное решение этой задачи позволит с высокой точностью осуществлять прогноз белок-белковых взаимодействий, выявлять функциональные сайты белков, а также предсказывать функциональные свойства новых белков, структуры которых получены, в частности, в ходе выполнения геномных проектов.

Специфические сайты связывания белков обычно располагаются в карманах на их поверхности и образуют высоко заряженные кластеры либо обладают ярко выраженными гидрофобными свойствами. Поэтому для выявления потенциальных участков связывания белков мы считаем целесообразным исследовать гидрофобные свойства карманов и распределение заряда на поверхности белков. В данной работе мы представляем разработанный нами программный комплекс для определения и исследования структурных особенностей на поверхности белков. Программный комплекс состоит из трех модулей, которые позволяют строить молекулярную поверхность белка (модуль 1); определять карманы на построенной поверхности (модуль 2) и осуществлять анализ распределения заряда и гидрофобных областей в карманах и на поверхности белка (модуль 3) (рис. 1).

Исходными данными для всех трех модулей служат координаты атомов аминокислотных остатков, импортируемые из файлов в формате PDB. Выходные данные модуля 1 (поверхность белка) и модуля 2 (карманы) также сохраняются в формате PDB и могут служить входными данными

для следующих модулей. Потенциальные сайты связывания белков (выходные данные модуля 3) сохраняются в виде списка формирующих их аминокислотных остатков в формате *txt*.

2 Методы исследования

2.1 Построение поверхности белков

Для построения точечной молекулярной поверхности белков мы использовали алгоритм Коннолли. Исходными данными служат координаты атомов белков, импортируемые из файла формата PDB. Построенная поверхность представляется списком формирующих ее атомов аминокислотных остатков и сохраняется в файле формата PDB.

2.2 Поиск карманов на поверхности белков

Для определения карманов на поверхности белков мы выбрали алгоритм SurfNet и модифицировали его. Между каждой парой атомов A и B поверхности белка находили расстояние $R(A,B)$. Если выполнялись условия: $R(A,B) \leq R_{\max}$ (1), $R(A,B) \geq R_{\min}$ (2), то между атомами A и B строили отрезок AB , на котором в точке D располагали центр сферы $S_{\min}$ (рис. 1). Радиус сферы принимали равным минимальному радиусу определяемого кармана $R_{\min}$.

Если сфера $S_{\min}$ не пересекалась ни с одной другой сферой S_{Ni} для всех атомов Ni поверхности белка, то есть $S_{\min} \cap S_{Ni} = \emptyset, i = \overline{1, m}$ (3), то атомы A и B считали входящими в карман и переходили к следующей паре атомов поверхности белка.

Если условие (3) не выполнялось, то центр сферы $S_{\min}$ сдвигали от точки D вдоль отрезка AB в направлении атома B на величину шага $Step$ и опять проверяли условие (3).

Итерация для атомов A и B заканчивалась в том случае, если (i) выполнялось условие (3) или (ii) сфера $S_{\min}$ пересекалась со сферой S_B . Во втором случае считали, что атомы A и B не входят в карман и переходили к следующей паре атомов поверхности белка. Процесс осуществляли для всех пар атомов поверхности белка.

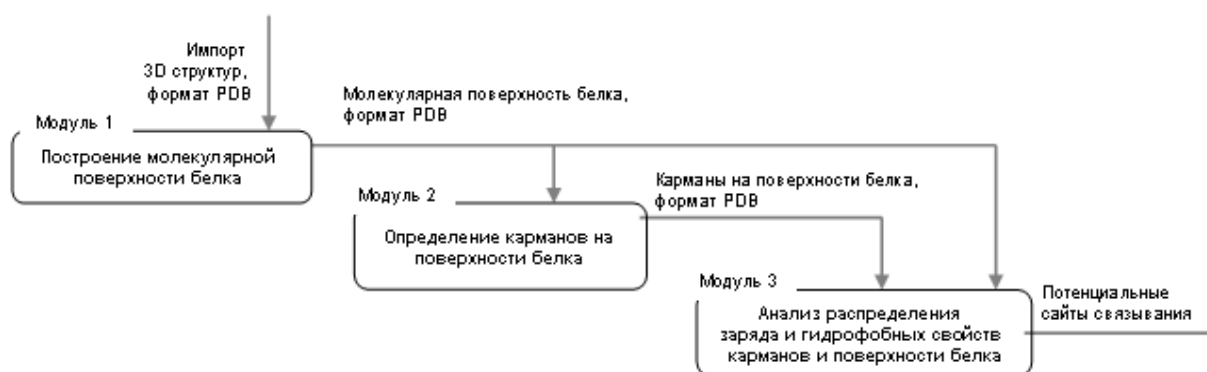


Рисунок 1. Схема программного комплекса

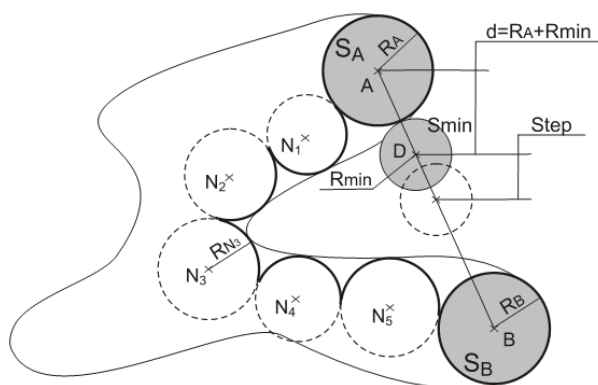


Рисунок 2. Определение карманов

2.3 Анализ распределения заряда и гидрофобных областей в карманах и на поверхности белков

На множестве всех аминокислот мы определили три подмножества. В подмножество (AA+) мы включили аминокислоты, которые имеют положительно заряженные боковые группы. В подмножество (AA-) мы отнесли аминокислоты с отрицательно заряженными боковыми группами. К подмножеству (AAN) мы отнесли аминокислоты, проявляющие гидрофобные свойства.

Входной список для модуля 3 – это список атомов аминокислот в PDB формате, который мы получаем из модуля 1 или модуля 2 (рис. 1). В первом случае из модуля 1 мы получаем список атомов молекулярной поверхности белка и проводим поиск гидрофобных и заряженных областей на всей белковой поверхности. Во втором случае из модуля 2 мы получаем списки атомов, формирующих карманы на поверхности белка, и исследуем заряд и гидрофобные свойства карманов. Каждый список в этом случае соответствует отдельному карману.

Для определения заряженных и гидрофобных областей просматривали все пары атомов входного списка. Для определения положительно заряженных областей для каждого атома AX, такого, что $X \in (AA+)$, находили все атомы AY такие, что $Y \in (AA+)$ и $R(AX,AY) < R$. Отрицательно

заряженные и гидрофобные области определяли аналогично. Атомы AX и AY из одного входного списка.

3 Результаты и их обсуждение

Построенная в нашей программе поверхность может быть экспортирована в формате PDB, который является фактическим стандартом представления пространственной структуры макромолекул и поддерживается большинством программных продуктов, осуществляющих исследование пространственной структуры макромолекул.

Для определения карманов на поверхности белков мы разработали модификацию алгоритма SurfNet, которая позволяет отсекал неперспективные для дальнейшего исследования карманы. Мы ограничили максимально возможный радиус искомых карманов, что позволило отсеять неглубокие и широкие псевдо-карманы. Такие структурные особенности представляют собой плоский участок поверхности с небольшим радиусом кривизны и не являются собственно карманами. Тем не менее, часто они имеют достаточно большую площадь поверхности и объем и распознаются большинством существующих программ как карманы. В нашей модификации алгоритма SurfNet такие псевдо-карманы отсекаются, что уменьшает количество ложноположительных определений карманов, т.е. ошибок первого рода, и увеличивает прогностическую ценность полученных данных.

Также алгоритм SurfNet идентифицирует карманы только правильной, конической формы, и не определяет структурные особенности в случаях, если поверхность кармана имеет неконическую форму с выступами внутри кармана, и, таким образом, возникает ошибка второго рода. Разработанный нами алгоритм в модуле 2 позволяет исправить этот недостаток. Модуль 3 позволяет определять гидрофобные кластеры, исследовать распределение заряда на поверхности белка и, таким образом, предсказывать потенциальные сайты связывания белков.

Adaptive trade-off between reproduction and survival in Mediterranean fruit flies induced by changing dietary conditions

Karkach A.S.⁽¹⁾, Romanyukha A.A.⁽¹⁾, Yashin A.I.⁽²⁾

Institute of Numerical Mathematics, Russian Academy of Sciences (Moscow)⁽¹⁾,
Duke University, Center for Demographic Studies (Durham)⁽²⁾

Arseny@inm.ras.ru

Abstract

Functions of living organisms depend on resources and energy and the corresponding traits are positively correlated with amount of allocated resources. An organism faces a problem of optimizing its' lifetime performance, or fitness, within the limited amount of available resources. This implies a negative relation between the main factors of fitness such as fertility and longevity. Still the fertility-longevity trade-offs are masked by various factors, are hard to detect and remain poorly understood.

Analysis of experimental data on Mediterranean fruitfly *C. capitata* shows a fertility-longevity trade-off especially pronounced in the long lived fraction of population. The presence of such trade-off suggests a possibility of a common metabolic resource shared between reproduction and somatic maintenance.

We present a model of resource acquisition, allocation and use by the main physiological functions of the fly which allows to understand the process and formulate a principle of dynamic resource allocation.

1 Introduction

The results of numerous experimental studies of aging and longevity performed with populations of laboratory animals confirm substantial role of dietary conditions in life span. In this context the results of study by Carey et al. [1] look surprising. Contrarily to traditional expectations the average life spans in the populations of Mediterranean fruit flies (medflies) did not show much variability in response to changes in the diet. Although within each population subjected to

selected dietary regime the data showed high variability in individual characteristics of reproduction and longevity, they did not provide us with the evidence of reproduction-longevity trade-off predicted by the evolutionary theory of aging; many long-lived flies also have high levels of fertility, and many low-fertility flies also have short life span.

In contrast to such a behavior of the average life span the average fertility (the total number of eggs laid by the medflies during their life course) experienced substantial variation in response to changes in dietary conditions. The detailed analyses of data with different regimes of diet revealed much higher variability in both life history traits in the long lived fraction of flies, than in the fraction of short-lived ones.

The long-lived fraction of Mediterranean fruitfly (*C. capitata*) female population demonstrates a pronounced negative relation between average life span and average total amount of eggs laid when the diet changes from sugar-only to protein-rich. This clearly adaptive feature may be a demonstration of a physiological trade-off between fertility and somatic maintenance.

2 Methods and Algorithms

We performed reanalysis of data from experimental study of fertility and longevity response to different diets in females of Mediterranean fruit fly *C. capitata* [1, 2]. In their experiments pairs of Mediterranean fruit fly females and males were subjected to different dietary regimes composed of two diets: the complete diet containing yeast hydrolyzate and sugar (diet *p*) and a sugar-only diet (diet *s*). Female life span and the number of eggs laid daily were studied in several cohorts, which received food in cycles of one day diet *p* and 1, 3, 5, 10, or 20 days diet *s*. One control cohort received diet *s* full time and one received diet *p* full time.

We construct a mechanistic mathematical model for the life history of individual female fly. It describes an evolution of its' state variables: metabolic capacity, reserves of energy (glucose and fat) and proteins in the body in response to external factors (ingestion of sugar and protein). The solution of the model is fertility and longevity of the fly. The model improves upon the earlier resource allocation model described in [3].

Results

A clear negative dependence between average fertility and longevity is observed in the long lived part of experimental cohorts as the protein content of the diet changes. At the same time the short-lived fraction does not show any trade-off, as does the whole population (mix). We suggest that such physiological trade-off allows to adapt to different environmental conditions but can only be realized when the fly has sufficient amount of metabolic resources. In order to explain the observed phenomenon we suggest the *variation principle of adaptive trade-off*. It states that among all possible ways to allocate available metabolic resource those that maximize the lifetime reproduction are realized. The result is the coordination of the decline in the two main organism's abilities – to maintain soma, and to reproduce. This coordination is manifested by simultaneous depletion of metabolic resources allocated to these functions. This phenomenon can be called *coordinated aging*.

Conclusion

The presence of a fertility-longevity trade-off suggests a possibility of existence of some resource used both by reproduction and somatic maintenance in a

fly. The trade-off may be a manifestation of metabolic machinery, processes and genetically determined laws of control which define balance between the processes of reproduction and regeneration. We propose and discuss a principle of dynamic resource allocation which explains fertility–longevity data for the long-, intermediate- and short-lived flies. Adaptive allocation of metabolic and other resources allows flies to tailor their life history parameters to the environment. Due to limitations of the physiological adaptation a significant share of the population may be genetically “preadapted” to different environmental conditions thus contributing to population stability and heterogeneity. This may be observed even in relatively homogeneous populations, such as experimental fly cohorts.

References

- [1] J.R. Carey, P. Liedo, L. Harshman, X. Liu, H.-G. Muller, L. Partridge, J.-L. Wang. (2002) Food pulses increase longevity and induce cyclical egg production in Mediterranean fruit flies, *Functional Ecology* 16: 313–325.
- [2] J.R. Carey, P. Liedo, H.-G. Muller, J.-L. Wang, J.W. Vaupel. (1998) Dual Modes of Aging in Mediterranean Fruit Fly Females, *Science*, 281: 996–998.
- [3] A.A. Romanyukha, J.R. Carey, A.S. Karkach, A.I. Yashin. (2004) The impact of diet switching on resource allocation to reproduction and longevity in Mediterranean fruitflies, *Proc. R. Soc. Lond. B.* 271: 1319–1324.

Использование компьютерных технологий при популяционном моделировании

Глушков В.Н.⁽¹⁾, Саранча Д.А.^(1, 2)

Вятский государственный университет ⁽¹⁾, Вычислительный центр РАН ⁽²⁾

saran@ccas.ru

Abstract

The complex system approach of ecological and biological objects modelling is offered. It based on a full set of operations – from an object choice substantiation, construction of an interconnected models set of different detailing degree to a hypotheses formulation about leading mechanisms of the investigated phenomenon. The given approach was used for models set construction to describe tundra populations and communities. Computer use features, methods of models filling by the information (in particular with the help of expert estimations) and questions of models adequacy are discussed.

1 Основной раздел

Стремление максимально приблизиться к описанию биологических процессов, к созданию комфортных условий для междисциплинарного диалога с неизбежностью приводит к имитационному моделированию. В данной работе решалась задача по разработке методов использования ЭВМ в качестве инструмента интеллектуальной поддержки в междисциплинарном процессе *создания и анализа математических моделей* сложных иерархических систем *при неполных, разноплановых и неизбежно искаженных данных* о свойствах изучаемого объекта и их применения для создания комплекса взаимосвязанных моделей тундровых популяций и сообществ.

Имитационные технологии снимают зависимость от «математического аппарата», позволяют выбирать зависимости «любого типа», требуемый уровень детализации. «Диалоговость, модифицируемость, наглядность», характерные для имитационных систем, позволяют эффективно подойти к пониманию сути описываемого явления и выражению этого понимания в форме математической модели. Но вместе с тем большая имитационная модель не может быть самодостаточной – это *полуфабрикат*, инструмент по переработке исходной биологической (и не только) информации, *инструмент по предварительному исследованию* объекта. Для ее

эффективного использования она должна быть компонентом набора взаимосвязанных моделей, состоящих из детальных имитационных моделей и сопряженных с ними упрощенных моделей. При этом важным аспектом является проведение комплексных исследований - согласование всех этапов моделирования.

Комплексные исследования включают в себя всю последовательность операций. Сбор, отбор, анализ и переработка исходной (биологической) информации. Обоснование выбора объекта, переменных и уравнений для его описания; выбор явления для раскрытия свойств объекта («тестирующего явления»); обоснование и построение детальных имитационных моделей; анализ их свойств. Формирование имитационной системы - набора моделей разной степени детализации; и формулирование на их основе гипотез о ведущих механизмах исследуемого явления.

Имитационное моделирование эффективно, когда нацелено на конкретные данные, на конкретную ситуацию. Это позволяет как-то бороться с «угрозой необозримости», использовать «физику процесса» при выборе упрощений, соответствующего уровня агрегирования. Успех моделирования во многом определяет удачный выбор объекта (как дрозофила для генетики). Преимуществом обладают системы, развивающиеся в некоторых экстремальных условиях, поскольку «в норме» процессы коэволюции, стремление к гомеостазу затрудняют вскрытие ведущих механизмов.

В качестве объекта моделирования была выбрана тундра. Определяющим фактором выбора является наличие «тестирующего эффекта» - наличие ярко выраженных регулярных колебаний численности животных, в частности, песцов и их основного корма – леммингов (тундровых грызунов, широко известных своими миграциями).

Рассматривалась задача изучения колебаний численности животных. При решении этой задачи было желание учесть широкую гамму свойств («принцип системности») и в то же время использовать минимально возможную математическую структуру («принцип минимальности») совместимую с доступными

экологическими данными («принцип соответствия (экологичности)»).

Анализ структуры пастбищной (наземной) части тундрового биоценоза показал возможность автономного рассмотрения сообщества «растительность - лемминги - песцы» (РЛП). Структура математической модели выбиралась в соответствии с гипотезой Вольтерра – Костицына [3]: в виде системы из трех обыкновенных дифференциальных уравнений первого порядка. Такая форма модели хорошо сочетается с подходами системной динамики Дж.Форрестера. Это позволило при моделировании сообщества РЛП создать «экологический конструктор»: динамика биомасс каждого из трех трофических уровней определялась тремя компонентами - скоростью репродукции, отчуждения и естественного отмирания, а каждая из компонент, в свою очередь, формировалась как произведение константы и соответствующих функций. «Модифицируемость» достигалась за счет их изменения.

Такой подход соответствует степени изученности, многообразию предположений и дает возможность учитывать различные представления экологов в различных модификациях модели. Это позволяет конструктивно реализовать идею А.Н.Колмогорова [10] об устранении зависимости результатов моделирования от конкретной параметризации.

При формировании базовой модели было создано несколько ее версий. Последовательно использовались следующие ведущие предположения: взаимодействия по типу "хищник - жертва", пороговая зависимость скорости размножения леммингов от биомассы корма, определяющее влияние внутривидовой динамики леммингов.

Параллельно с созданием различных версий модели сообщества началось формирование имитационной системы. Для создания более детального описания популяции леммингов с учетом возрастной структуры разумнее оказалось не расширять исходную модель (избежать эффекта «динозавра»), а писать новую независимую модель [1, 3]. Анализ результатов вычислительных экспериментов с обеими взаимодополняющими моделями позволил обосновать упрощенную модель в виде одномерного разностного уравнения, связывающего численности леммингов в двух соседних годах [1, 3]. На основе совместного анализа эколого-биологической информации и результатов вычислительных экспериментов удалось сформулировать и решить «обратную имитационную задачу» [1, 3] – выбрать такие постулаты исходной имитационной модели, которые позволили бы получить на основе исходной модели (точно, а не приближенно) указанное выше разностное уравнение. Она решается исходя из совместного анализа

результатов вычислительных экспериментов и соответствующей биологической информации, и основана на том, что в конкретный временной отрезок (сезон), изменение соответствующих переменных происходит в сравнительно узком диапазоне, что позволяет осуществить линеаризацию соответствующих функций.

Использование разностных уравнений позволило «настраивать» исходную имитационную модель в области параметров, обеспечивающие динамические режимы изменения численностей популяций, близкие к наблюдаемым в природе, а так же сформулировать гипотезу - формирование колебаний численности леммингов (и в следствие этого и песцов) определяется тремя показателями: 1) скоростью прироста биомассы в благоприятный год; 2) максимальной численностью; 3) выживаемостью в наиболее неблагоприятных условиях (или двумя безразмерными - относительной скоростью прироста популяции и долей гарантированно выживших зверьков). Первый показатель отражает баланс между процессами рождаемости и смертности в отсутствии "давления среда"; второй характеризует экосистему в целом и отражает коэволюцию леммингов и кормовой базы; третий характеризует адаптационные свойства леммингов в экстремальных условиях и во многом определяется локальными характеристиками, в частности рельефом местности в местах перезимовки.

Применение комплексного подхода позволило реализовать идею об эффективности использования имитационных технологий для обоснования упрощенных уравнений, допускающих параметрические исследования. Был обоснован особый класс моделей, учитывающих сезонность [1], а также тип дискретных отображений, для которых при определенном сценарии изменения параметров происходит изменение периодов устойчивых циклов в последовательности натурального ряда при [2]. Наличие предыдущего опыта моделирования позволило перейти на другой уровень описания, на использование индивидуально-ориентированных моделей [1].

Список литературы

- [1] Глушков В.Н., Недоступов Э.В., Саранча Д.А., Юферева И.В. Компьютерные методы анализа математических моделей экологических систем. М.: ВЦРАН. 2006. 74 с.
- [2] Недоступов Э.В., Саранча Д.А., Чигерев Е.Н., Юрезанская Ю.С. О некоторых свойствах одномерных унимодальных отображений // ДОКЛАДЫ АКАДЕМИИ НАУК, 2010, том 430, № 1. С. 1-6
- [3] Саранча Д.А. Количественные методы в экологии. Биофизические аспекты и математическое моделирование. М.: МФТИ, 1997. 283 с.

Условия устойчивости нечеткой модели динамики популяций

Дружинина О.В.⁽¹⁾, Масина О.Н.⁽²⁾

Вычислительный центр им. А.А. Дородницына Российской академии наук⁽¹⁾

Елецкий государственный университет им. И.А. Бунина⁽²⁾

ovdruzh@mail.ru, olga121@inbox.ru

The conditions of stability of fuzzy model of dynamics of populations

O.V. Druzhinina¹, O.N. Masina²

¹Dorodnicyn Computing Center of RAS

²Yelets State University by the name of
I.A. Bunin

Abstract

The question of existence of non-negative equilibria states of model of populations dynamics described by the three ordinary nonlinear differential equations is studied. The model is considered with regard to competition of species and different diffusion rates. The conditions of global asymptotic stability of equilibria states of model are obtained by the aid of methods of theory of stability and qualitative theory of differential equations. The conditions of stability of fuzzy model of dynamics of populations are obtained. The method of Lyapunov functions and the principle of reduction of the problem of stability of differential inclusion to the problem of stability of fuzzy differential equation are used for obtaining of these conditions.

1 Введение

При изучении моделей динамики численности взаимодействующих популяций возникает задача исследования устойчивости состояний равновесия [1, 2].

Отметим, что устойчивость решений классических и обобщенных моделей Лотки–Вольтерра исследовалась в [2–4] методом функций Ляпунова. В настоящей работе рассмотрено применение к исследованию устойчивости нечеткой модели динамики популяций принципа сведения задачи об устойчивости дифференциального включения к задаче об устойчивости нечеткого дифференциального уравнения. Указанный принцип описан в работах [5–7].

2 Модель динамики популяций, учитывающая конкуренцию и миграцию видов

Рассмотрена модель, учитывающая конкуренцию и миграцию видов, описываемая системой трех дифференциальных уравнений:

$$\begin{aligned}\dot{x}_1 &= x_1(1 - x_1 - qy_1) + \beta x_2 - \gamma x_1, \\ \dot{x}_2 &= x_2(1 - x_2) + \gamma x_1 - \beta x_2, \\ \dot{y}_1 &= y_1(1 - rx_1 - y_1),\end{aligned}\tag{1}$$

где x_1 и y_1 – численность конкурирующих видов x и y в ареале вида x_1 (ареале 1), x_2 – численность вида x в ареале вида x_2 (ареале 2), $q > 0$ и $r > 0$ – коэффициенты конкуренции видов в ареале 1, $\beta > 0$ и $\gamma > 0$ – скорости миграции вида x между двумя ареалами, причем $\beta \neq \gamma$. Модель (1) является обобщением модели популяционной динамики работы [3] на случай отличных друг от друга скоростей диффузии β и γ .

С помощью системы символьных, графических и численных вычислений «Mathematica 4.1» найдены состояния равновесия системы (1). Получены значения β и γ , при которых существуют неотрицательные состояния равновесия. Исследована устойчивость в смысле Ляпунова состояний равновесия.

Проведена оценка модельных параметров и построены локальные фазовые портреты системы (1) в зависимости от условий существования нетривиального состояния равновесия. Получены условия глобальной асимптотической устойчивости этого состояния равновесия.

3 Обобщения модели динамики популяций на случай дифференциального включения и на случай нечеткого дифференциального уравнения

Система (1) представима в виде векторного уравнения

$$dx/dt = f(x),\tag{2}$$

где $x = (x_1, x_2, y_1)$, $f(x) = (f_1, f_2, f_3) = (x_1(1 - x_1 - qy_1) + \beta x_2 - \gamma x_1, x_2(1 - x_2) + \gamma x_1 - \beta x_2, y_1(1 - rx_1 - y_1))$, $x \in R_+^3 = R_+ \times R_+ \times R_+$, $R_+ = [0, \infty)$, $f: R_+^3 \rightarrow R_+^3$.

Для уравнения (2) построено дифференциальное включение в виде

$$dx/dt \in F(x), \quad (3)$$

где $F(x) = \{f(x): \beta \in H, \gamma \in C, q \in Q, r \in T\}$, $H = [\beta_1, \beta_2]$, $C = [\gamma_1, \gamma_2]$, $Q = [q_1, q_2]$, $T = [r_1, r_2]$, $F: R_+^3 \rightarrow 2^{R_+^3}$.

Множества H , C , Q и T определяют множества значений нечетких параметров β , γ , q и r соответственно. Эти параметры учитывают нечеткие изменения в скоростях миграции видов, например, в летний и зимний периоды. Уровни значимости $H_\alpha = \{\beta: \mu_H(\beta) \geq \alpha\}$, $C_\alpha = \{\gamma: \mu_C(\gamma) \geq \alpha\}$, $Q_\alpha = \{q: \mu_Q(q) \geq \alpha\}$ и $T_\alpha = \{r: \mu_T(r) \geq \alpha\}$, где μ – функция принадлежности, представляют более узкие множества, полученные при учете дополнительных условий $\alpha \in (0, 1]$, влияющих на миграцию и конкуренцию видов. К таким условиям относятся, например, недостаток пищи, снижение рождаемости, климатические условия и т.д. Уравнение (2) сводится к нечеткому дифференциальному уравнению

$$d\tilde{X}/dt = \tilde{F}(\tilde{X}), \quad (4)$$

где $\tilde{X}$ – нечеткое множество, $\tilde{F}: \tilde{R}_+^3 \rightarrow P(R_+^3)$, $\tilde{R}_+^3$ – нечеткое открытое подмножество пространства R_+^3 , $P(R_+^3)$ – совокупность всех нечетких подмножеств из R_+^3 . Тогда $d\varphi/dt \in F_\alpha(\varphi)$ – соответствующее уравнению (4) дифференциальное включение при $\alpha \in (0, 1]$, и $F_\alpha(\varphi) = \{f(\varphi(t)): \beta \in H_\alpha, \gamma \in C_\alpha, q \in Q_\alpha, r \in T_\alpha\}$.

4 Устойчивость нечеткой модели динамики популяций

С помощью изложенного в работах [5–7] принципа сведения задачи об устойчивости дифференциального включения к задаче об устойчивости нечеткого дифференциального уравнения получены следующие условия устойчивости дифференциального включения (3) и нечеткого уравнения (4). Во-первых, если для замкнутого множества $M \subset R_+^3$ существует функция Ляпунова V относительно включения (3), для которой верно неравенство $D_+V(x) \leq 0 \quad \forall x \in B(M, r)$, где $D_+V(x) = \sup DV(x)$ – верхняя производная

функции Ляпунова, то множество M устойчиво относительно этого включения. Если верно неравенство $D_+V(x) \leq -w_3(e(x, M)) \quad \forall x \in B(M, r)$, где функция $w_3: B(M, r) \rightarrow R$ непрерывна и положительна вне M , то множество M асимптотически устойчиво относительно включения (3). Во-вторых, если для замкнутого нечеткого множества $\tilde{M} \subset P(R_+^3)$ существует функция Ляпунова V относительно уравнения (4), для которой при $\alpha \in (0, 1]$ верно неравенство $D_+V_\alpha(x) \leq 0 \quad \forall x \in B(M_\alpha, r)$, то множество $\tilde{M}$ α -устойчиво относительно этого уравнения. Если выполняется условие $D_+V_\alpha(x) \leq -w_{3\alpha}(e(x, M_\alpha)) \quad \forall x \in B(M_\alpha, r)$, где функция $w_{3\alpha}: (0, r) \rightarrow R$ непрерывна и положительна, то множество M α -асимптотически устойчиво относительно уравнения (4).

Работа поддержана Фондом содействия отечественной науке.

Литература

- [1] Свирежев Ю.М., Логофет Д.О. Устойчивость биологических сообществ. М.: Наука, 1978.
- [2] Пых Ю.А. Равновесие и устойчивость в моделях популяционной динамики. М.: Наука, 1983.
- [3] Zhang Xin-an, Chen L. The linear and nonlinear diffusion of the competitive Lotka–Volterra model // Nonlinear Analysis. 2007. V. 66. P. 2767–2776.
- [4] Дружинина О.В., Масина О.Н. Исследование существования и устойчивости решений дифференциальной системы экологической динамики с учетом конкуренции и диффузии // Нелинейный мир. 2009. т. 7. № 11. С. 881–888.
- [5] Шестаков А.А. Обобщенный прямой метод Ляпунова для систем с распределенными параметрами. М.: Наука, 1990.
- [6] Меренков Ю.Н. Устойчивоподобные свойства дифференциальных включений, нечетких и стохастических дифференциальных уравнений. М.: Изд-во РУДН, 2000.
- [7] Дружинина О.В., Масина О.Н. Методы исследования устойчивости и управляемости нечетких и стохастических динамических систем. М.: ВЦ РАН, 2009.

Изучение популяционно-генетической структуры женьшеня настоящего *Panax ginseng* С.А. Meyer на основе AFLP-маркеров с применением различных пакетов компьютерных программ

Кац И.Л., Реунова Г.Д., Музарок Т.И., Журавлев Ю.Н.

Биолого-почвенный институт ДВО РАН, Владивосток
Institute of Biology and Soil Science FEB RAS, Vladivostok
kats@ibss.dvo.ru

The study of genetic variability and population structure of *Panax ginseng* C.A. Meyer based on AFLP using different software programs

Irina L Kats, Galina D Reunova,
Tamara I Muzarok, Yuri N Zhuravlev

Abstract

There is a great interest in using amplified fragment length polymorphism (AFLP) markers for assessing genetic diversity because they have a wide coverage of species genomes and consequently are more reliable and reproducible as compared to RAPD and ISSR markers. Many software programs for molecular population genetics studies based on dominant markers have been developed for personal computers. POPGENE, TFGA, ARLEQUIN, HICKORY programs were chosen in order to obtain optimized parameters of genetic variability and population structure of *Panax ginseng* C.A. Meyer using dominant AFLP markers.

Ареалы многих, если не большинства видов, являются пространственно подразделенными и, в общем, могут быть описаны, как метапопуляция, состоящая из множества локальных популяций. Определение этой генетической структуры имеет важное значение для понимания эволюции видов и разработки научных основ сохранения и восстановления биологического разнообразия.

Существует много молекулярных маркеров, с помощью которых можно изучать популяционно-генетическую структуру видов. Кодоминатные маркеры, такие как аллозимы и микросателлиты, позволяют четко различать генотипы и могут быть легко проанализированы пакетами программ. Тем не менее, определенные практические ограничения, такие как: необходимость иметь свежий материал и невысокий уровень полиморфизма – у аллозимов, и

сложность и дороговизна разработки микросателлитных праймеров, служат причиной широкого использования молекулярных маркеров доминантной природы, таких как RAPD, ISSR, AFLP. Способность давать большое число маркеров, охватывающих весь геном, высокая воспроизводимость результатов и другие особенности метода определяют преимущества AFLP маркеров. С другой стороны, использование доминантных маркеров ведет к серьезным трудностям, из-за их неспособности разделить гомозиготные и гетерозиготные особи. Поэтому следует ясно понимать проблемы, возникающие из доминантной природы маркера при изучении популяционно-генетической структуры видов. По существу данные AFLP являются бинарными: для каждого индивида регистрируется «присутствие фрагмента» или «отсутствие фрагмента», и сам фрагмент должен рассматриваться как фенотипический признак.

Одним из наиболее часто используемых показателей для оценки внутривидового разнообразия является ожидаемая гетерозиготность H_{ex} , которая является эквивалентом генетического разнообразия H_s [7], а межвидовое генетическое разнообразие часто оценивают с помощью коэффициента дифференциации G_{st} [8] и F -статистик Райта [3,10]. Эти параметры легко определяются после вычисления частот аллелей на основе равновесия Харди-Вайнберга, и исследователь должен быть твердо уверен, что в исследуемых популяциях отсутствует инбридинг. На самом деле, в природе популяции пребывают в равновесии Харди-Вайнберга не часто, уровень инбридинга, как правило, неизвестен, что ведет к смещению оценок частот аллелей и других генетических параметров.

Чтобы преодолеть эти препятствия, используют косвенные оценки частот аллелей. В пакете программ POPGENE частоты аллелей вычисляются согласно Chong et al. [1] и, кроме индекса генетического разнообразия H_s , вычисляется информационная мера (или индекс) Шеннона [9], которая не требует предположения о равновесии Харди-Вайнберга. В пакете программ TFGA для

вычисления аллельных частот используется подход Lynch et al. [5]. Метод, предложенный Lynch and Milligan, подходит также и для AFLP маркеров, но, как показал Животовский [11], он ведет к смещению оценок генетических параметров в сторону занижения. Оба пакета программ не считают F-статистики Райта для доминантных маркеров.

AMOVA (пакет программ ARLEQUIN) является классическим анализом вариантов, который использует дистанции между наблюдениями как исходную матрицу. В этом анализе расчеты проводятся над фенотипическими данными. Расстояния между индивидами определяются как Эвклидовы дистанции [2, 6]. AMOVA позволяет определить межпопуляционную изменчивость F_{st} (аналог G_{st}). Но и AMOVA не вычисляет коэффициент инбридинга для доминантных маркеров.

Holsinger et al. [10] предложили вычисление параметров F_{st} (степень дифференциации гена между популяциями) и F_{is} (коэффициент инбридинга) по частотам фенотипов, используя теорему Байеса и MCMC метод (программа HICKORY).

Мы применили описанные пакеты программ к матрице 1/0 (167x282), полученной с помощью двух пар AFLP-праймеров: Eco(ACA)/Mse(CCTG) и Eco(ACA)/Mse(CCGG), которые дали 133 и 149 фрагментов на пару праймеров соответственно. Анализ фрагментов проводился на автоматическом геном анализаторе ABI-3100. В анализ были включены 167 растений редкого исчезающего вида, ценного лекарственного растения – женьшеня настоящего (*Panax ginseng* C.A. Meyer). Растения были собраны в Приморском крае Российской Федерации в десяти административных районах.

Были получены следующие показатели генетического разнообразия *Panax ginseng*.

Для TFPGA: ожидаемая гетерозиготность H_{ex} в исследованных выборках женьшеня изменялась в пределах от 0.1145 до 0.2300, со средним значением 0,1848.

Для POPGENE: ожидаемая гетерозиготность H_{ex} в популяциях женьшеня изменялась в пределах от 0,1029 до 0,2131, со средним значением 0,1729; для вида $H_{ex} = 0,2530$. Индекс разнообразия Шеннона (I_{sh}) варьировал в популяциях от 0.1602 до 0.3222 со средним значением 0.2626; на уровне вида $I_{sh} = 0.3967$. Индекс межпопуляционного разнообразия $G_{st} = 0,3257$.

Для AMOVA: в неиерархическом анализе показатель межпопуляционной изменчивости $F_{st} = 0,3553$.

Для HICKORY: Индекс межпопуляционного разнообразия $\theta_B = 0,3634$.

Полученные результаты обсуждаются.

Генетические взаимоотношения между популяциями исследовали на основе дистанций Нея методом UPGMA и с применением эвклидовых дистанций построением минимального спэннинг дерева (MST).

Все исследования (UPGMA, MST, AMOVA, Bayesian анализ) выявили две дифференцированные популяционные группы женьшеня, которые должны рассматриваться как отдельные единицы для популяционно-генетического мониторинга и демографического анализа.

Список литературы.

- [1] Chong D.R.X., R.C. Yang, and F.C. Yeh. Nucleotide divergence between populations of trembling aspen (*Populus tremuloides*), 1994
- [2] Excoffier L., P. Smouse, and J. Quattro. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Applications to human mitochondrial DNA restriction data. *Genetics* 131:479-491, 1992
- [3] Hartl DL, Clark AG. Principles of population genetics. 2nd ed. Sinauer Associates, Sunderland, MA., 1989
- [4] Holsinger K.E., P.O. Lewis and D.K. Dey. A Bayesian approach to inferring population structure from dominant markers. *Mol. Ecol.* 11:1157-1164, 2002
- [5] Lynch M., and B.G. Milligan. Analysis of population structure with RAPD markers. *Mol. Ecol.* 3:91-99, 1994
- [6] Michalakis Y., and L. Excoffier. A genetic estimation of population subdivision using distances between alleles with special interest to microsatellite loci. *Genetics* 142:1061-1064, 1996
- [7] Nei M. Genetic distance between populations. *Am Nat* 106:283-292, 1972
- [8] Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583-590, 1978
- [9] Shannon CE, Weaver W. The mathematical theory of communication. Univ. of Illinois Press, Urbana, 1949
- [10] Weir B.S. Genetic Data Analysis. Sinauer Associates, Sunderland, MA., 1990
- [11] Zhivotovsky L.A. Estimation population structure in diploids with multilocus dominant markers. *Mol. Ecol* 8:907-913, 1999

Моделирование межвидовой конкуренции в смешанных одновозрастных древостоях

Колобов А.Н., Фрисман Е.Я.

Институт комплексного анализа региональных проблем ДВО РАН, 679016 Биробиджан,
ул. Шолом-Алейхема, 4⁽¹⁾
Alex_0201@mail.ru

Abstract

The results of study and construction of a simulation model describing the growth and interaction of trees in the process of competition for light are considered. Qualitative analyses of the model behavior for different values of the parameters determining the rate of trees growth are held.

$$\begin{cases} \frac{dV}{dt} = E \cdot bV^d - cVH \\ \frac{dH}{dt} = \lambda H \cdot (H_{\max} - H) \\ D = \sqrt{\frac{4V}{\pi \cdot H \cdot f}} \end{cases}$$

где V — объем, H — высота, D — диаметр дерева, E — интенсивность фотосинтеза единицы листовой поверхности, f — видовое число, показывающее отклонение от идеального цилиндра, d — фрактальная размерность кроны.

Формула интенсивности фотосинтеза единицы листовой поверхности.

$$E = \frac{a \cdot (1 - \exp(-pV^d)) \cdot P_{\max}}{a \cdot (1 - \exp(-pV^d)) + P_{\max} \cdot pV^d}$$

где $P_{\max}$ — максимальная интенсивность фотосинтеза единицы листовой поверхности, k — коэффициент затухания светового потока проходящего сквозь крону.

1 Введение

Конкуренция между растениями в сообществе — не прямое взаимодействие растений в ходе раздела и передела, недостаточных для всех растений сообщества ресурсов — является одним из ведущих процессов в развитии любого сообщества растений. В существующих математических моделях учитываются, прежде всего, процессы конкуренции за внешние ресурсы. Описание процессов взаимодействия сводится к подсчету доли внешнего ресурса, которая приходится на отдельное дерево.

Особенность нашей модели заключается в том, что она содержит небольшое количество ведущих параметров, имеющих прямое эколого-физиологическое истолкование. Использование небольшого количества существенных параметров позволяет изучать и анализировать особенности взаимодействия различных видов деревьев в сообществе, на качественном и количественном уровне при различных вариациях этих параметров.

2 Формализация роста одиночного дерева

Для описания роста дерева в качестве исходного пункта мы использовали модель свободного роста дерева, предложенную в работе Полетаева. Согласно этой модели дерево получает энергию только путем фотосинтеза, свободная энергия расходуется на нужды фотосинтеза, на построение живой ткани и на подъем раствора из почвы. Уравнение роста записывается в форме закона сохранения энергии.

Рост дерева описывается следующей системой уравнений, позволяющей вычислять объем, высоту и диаметр ствола

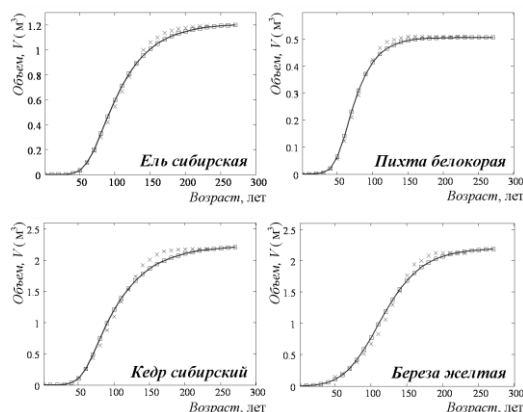


Рис. 1 Графики роста объема для некоторых видов деревьев. Точки — эмпирические данные, линии — модельные кривые.

Адекватность полученного уравнения роста дерева, была проверена на ряде таблиц хода роста. Из графиков видно, что модельная кривая роста дерева имеет S-образный вид и при этом довольно хорошо аппроксимирует эмпирические данные (см. рис. 1).

3 Рост дерева в условиях конкуренции

Уравнение роста объема i -го дерева в сообществе, учитывая влияние конкуренции со стороны окружающего древостоя, записывается в следующем виде

$$\frac{dV_i}{dt} = \frac{a_i(1 - \exp(-p_i V_i^d)) P_{i\max} Q_i}{a_i(1 - \exp(-p_i V_i^d)) Q_i + P_{i\max} p_i V_i^d} b_i V_i^d - c_i V_i H_i$$

В общем виде, при взаимодействии n деревьев, доля солнечной радиации, приходящейся на i -ое дерево при затенении окружающим древостоем, определяется выражением:

$$Q_i = \sum_{s=1}^{n-1} \left[\sum_{\substack{j_1=1 \\ j_1 \neq i}}^n \sum_{\substack{j_2=2 \\ j_2 \neq i}}^n \dots \sum_{\substack{j_s=j_{s-1}+1 \\ j_s \neq i}}^n \exp\left(-\sum_{r=1}^s k_{j_r i} x_{j_r i}\right) \cdot q_{j_1 \dots j_s i} \right] + q_{0i}.$$

Здесь q_{0i} - доля открытой части кроны i -го дерева, $q_{j_1 \dots j_s i}$ - доля кроны i -го дерева перекрываемая s раз другими деревьями, k_j - коэффициент затенения соответствующим видом дерева.

Таким образом, можно рассчитать рост дерева в условиях конкуренции, учитывая при этом видовые особенности пропуска света, что позволяет рассматривать процессы межвидовой конкуренции.

Рассмотрим случай одновидового сообщества (см. рис. 2).

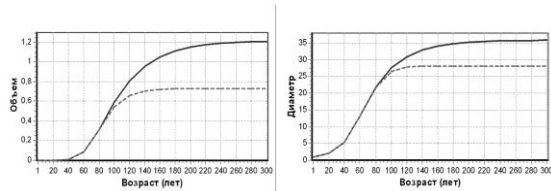


Рис. 2 Кривые роста объема и диаметра дерева при различной степени освещенности. (Верхняя кривая – полное освещение ($Q = 1$), нижняя кривая – неполное освещение ($Q < 1$))

Как видно из графиков дерево, растущее при полном освещении, реализует свои потенциальные возможности роста и, соответственно, имеет большие размеры по сравнению с деревом, произрастающим в сообществе. В результате неполного освещения, скорость роста второго дерева падает, оно раньше выходит на насыщение и как следствие имеет меньшие размеры.

4 Моделирование динамики двухвидового сообщества

При конкурентном взаимодействии видов, возможно возникновение двух сценариев поведения: либо виды сосуществуют, либо один вид оказывается более сильным и постепенно подавляет второй.

Для изучения межвидовых взаимодействий в древесных сообществах были проведены вычислительные эксперименты с некоторым набором основных лесообразующих пород Дальнего Востока (см. рис. 3). В качестве исследуемых видов, были рассмотрены: ель сибирская, пихта белокорая, кедр корейский и береза желтая.

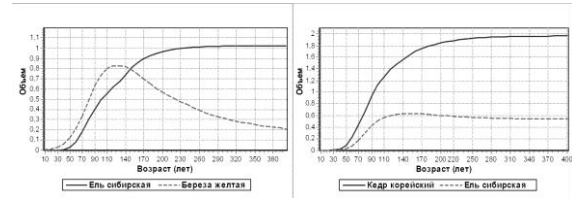


Рис. 3 Сценарии роста некоторых взаимодействующих видов деревьев: ель сибирская, кедр сибирский, береза желтая.

Береза при взаимодействии с елью оказывается в угнетенном состоянии, что видно из графиков по отрицательному приросту объема. Темнохвойные виды кедр и ель сосуществуют на одной территории, не оказывая, значительного угнетения друг на друга.

5 Выводы

В работе представлена модель конкурентного взаимодействия различных видов деревьев, построенная на идеи энергетического баланса.

Анализ рассмотренной модели показал: деревья, в условиях конкуренции взаимно подавляя друг друга, раньше выходят на насыщение и имеют меньшие размеры. Если коэффициент затухания светового потока (k_i) при прохождении через крону деревьев одного из видов, оказывается достаточно высоким, то он вытесняет остальные виды при одинаковых условиях. Параметр P_{im} , отвечающий за максимальную интенсивность фотосинтеза, являясь характеристикой теневыносливости дерева, влияет на выживаемость вида в условиях лимитирования световых ресурсов. При высоком уровне P_{im} требуется большее количество света, поэтому при неблагоприятных световых условиях, вид с меньшей интенсивностью фотосинтеза имеет более высокую вероятность выживания.

Результаты моделирования также показали, что конкурентные взаимодействия зависят не только от степени взаимного затенения, но в значительной степени определяются характером кривой хода роста дерева.

Исследования проведены при финансовой поддержке РФФИ (проект № 09-04-00146-а) и ДВО РАН (конкурсные проекты № 09-I-P23-12, № 10-III-B-06-141).

Пространственная синхронизация в моделях метапопуляций

Кулаков М.П.

Институт комплексного анализа региональных проблем ДВО РАН
k_matvey@mail.ru

Abstract

In this paper we investigated the process of spatial synchronization in the globally coupled populations. Bifurcation mechanisms formation of clusters attractors of length 2 we considered for populations whose habitat is a linear network.

1 Постановка задачи

В данной работе предпринята попытка описания простейших механизмов формирования пространственной динамики неоднородной метапопуляции, представленной связанными локальными группировками особей, обитающих в экологически идентичных обособленных местообитаниях, между которыми возможно небольшое миграционное взаимодействие. Предполагается, что все очаги определены и постоянны в пространстве и связаны в некоторые цепочки (линейные или пространственные) с глобальными связями на основе миграционных потоков.

В качестве модели пространственной динамики подобной популяции рассматривалась система связанных логистических отображений, демонстрирующих переход к хаосу через удвоение периода. Каждый элемент при этом диссипативно связан с некоторыми другими элементами с постоянной долей мигрантов, а динамика N миграционно-связанных популяций в дискретные моменты времени описывается системой

$$x_{n+1}^{(i)} = \sum_{j=1}^N M_{i,j} f(x_n^{(j)}) \quad (i=1, 2, \dots, N), \quad (1)$$

где $x_n^{(i)}$ – численность i -й популяции в n -й сезон, $M_{i,j}$ ($i \neq j$) – коэффициент диффузии (миграции) особей из j -й популяции в i -ю, $f(x)$ – функция локального воспроизводства. Диагональный элемент $M_{i,i}$ при этом полагается в виде:

$$M_{i,i} = 1 - \sum_{j=1}^N M_{j,i} \quad (i=1, 2, \dots, N), \text{ равный доле не}$$

эмигрировавших из i -й популяции особей.

Из всех возможных вариантов пространственного распределения подобных популяций изучался случай, когда все популяции выстроены в линейную цепочку, в которой i -я популяция связана с двумя ближайшими $i-1$ и $i+1$, а 1-я лишь со 2-й, а N -я лишь с $N-1$ -й. Матрица M и якобиан системы (1) при этом имеет трехдиагональную структуру. В случае равенства всех недиагональных коэффициентов матрицы, т.е. $M_{12} = M_{21} = M_{32} = \dots = M_{k,k-1} = M_{k-1,k} = m$ система (1) принимает следующий вид

$$\begin{cases} x_{n+1}^{(1)} = (1-m)f(x_n^{(1)}) + mf(x_n^{(2)}) \\ \dots \\ x_{n+1}^{(i)} = (1-2m)f(x_n^{(i)}) + mf(x_n^{(i-1)}) + mf(x_n^{(i+1)}), \\ \dots \\ x_{n+1}^{(m)} = (1-m)f(x_n^{(m)}) + mf(x_n^{(m-1)}) \end{cases} \quad (2)$$

при $m \leq 0.5$.

Если $f(x)$ – унимодальная функция, а динамика локальной популяции в отсутствии миграции может быть описана рекуррентным соотношением $x_{n+1} = f(x_n)$, с единственным нетривиальным стационарным состоянием $\bar{x} = h$, тогда система (2) имеет стационарное решение в виде $\bar{x} = (h, h, \dots, h)^T$ при условии равенства всех $M_{i,j}$ ($i \neq j$). В качестве функции воспроизводства рассматривалась зависимость запас-пополнение Рикера $f(x) = axe^{-x}$ с репродуктивным потенциалом a , для которой $h = \ln(a)$.

2 Некоторые режимы пространственной динамики

Для подобных систем ($N > 2$) описываются бифуркационные механизмы формирования и закономерности локализации кластеров синхронизации с циклами длины 2. Для системы (2) рассматривается задача распространения популяционной волны в системе идентичных популяций, где каждая локальная популяция миграционно связана с двумя соседними, а все очаги

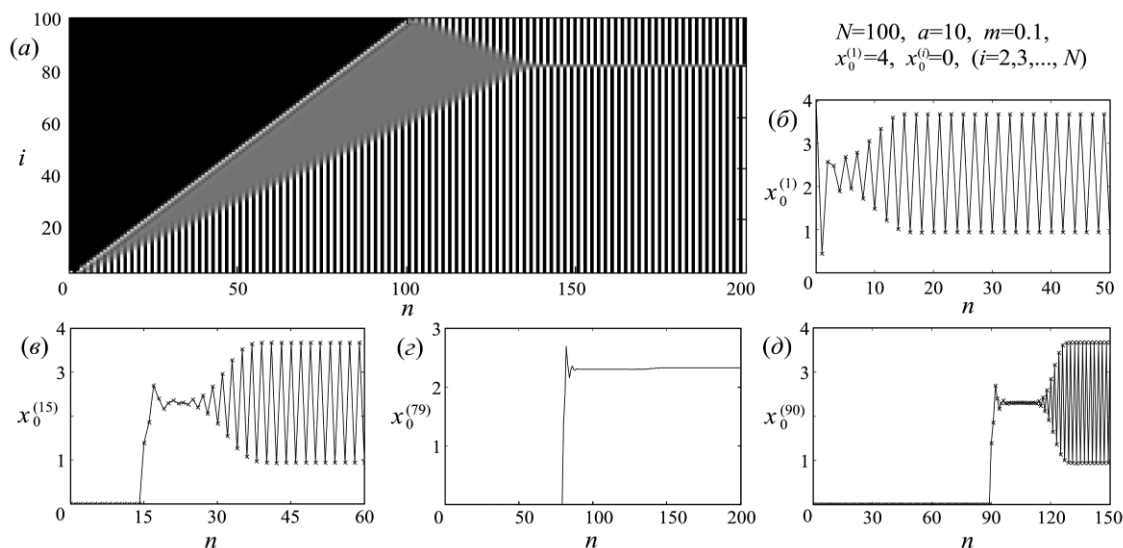


Рисунок 1. (а) Диаграмма пространственной динамики численности 100 очагов метапопуляции за 200 сезонов и (б-д) динамика отдельных очагов системы (2)

их скопления образуют в пространстве некоторую линейную цепочку.

На рисунке 1 а в плоскости номера субпопуляции и номера сезона продемонстрирован пример подобного распространения особей в системе (2) для сотни локальных очагов ($N=100$) из единственной начальной субпопуляции ($x_0^{(0)} = 4$). Черному цвету соответствует начальная нулевая численность других местообитаний, самая последняя из которых ровно через сто поколений заполняются мигрировавшими из соседних очагов особями. Сплошному пятну серого цвета треугольной формы соответствует длительный период релаксации динамики популяций, который сменяется колебаниями численности периода 2, изображенного чередующимися полосами белого (максимум) и черного цвета (минимум численности). Серая полоса, которая выступает границей между двумя кластерами синхронизации, соответствует динамики субпопуляции, в которой после периода релаксации, за счет соседних очагов с противофазной динамикой, устанавливается равновесная динамика. На рисунках 1 б-д приведены временные реализации системы (2) для отдельных субпопуляций, на которых демонстрируется характер динамики численности.

3 Результаты

В случае равенства репродуктивных потенциалов всех локальных субпопуляций матрица Якоби системы (2) в векторном виде имеет вид $J = f'(\bar{x}) \cdot M = (1-h)M$, собственные числа которой выражаются через собственные числа матрицы M (λ_M) как $\lambda = (1-h)\lambda_M$, каждому из которых соответствует собственный вектор независимый от h .

Аналитическое исследование системы (2) показало, что первой бифуркации системы соответствует потеря устойчивости ненулевой неподвижной точки и появлению устойчивого синфазного 2-цикла во всех очагах метапопуляции, что соответствует $\lambda_M = 1$ и устойчивости динамики в диагональном подпространстве. Далее при малых t с ростом a наблюдается каскад рождения вокруг неустойчивой неподвижной точки противофазных циклов периода 2 в отдельных частях метапопуляции, в которой появляется, по крайней мере, 2 кластера синхронизации. Подобные режимы реализуются на различных плоскостях, построенных на всех собственных векторах матрицы M кроме диагонального, которому они все перпендикулярны.

Из всех инвариантных подпространств системы

(2) можно выделить часть плоскости $\sum_{i=0}^N x^{(i)} = N \cdot h$

в первом октанте, при $x^{(i)} \neq x^{(j)}$ ($i \neq j$) и $x^{(i)} > 0$ ($i = 1, 2, \dots, N$, $j = 1, 2, \dots, N$). В этой части реализуются полностью противофазные колебания длины 2, а в системе (2) возникает максимальное число кластеров синхронизации равное N , когда динамика каждой локальной субпопуляции противофазно двум рядом стоящим. Было установлено, что подобная динамика неспособна сформироваться из единственной начальной популяции, однако близкое к ней поведение способно возникать, при малых коэффициентах связи, t из любого начального распределения особей изначально близкого к несинфазному.

Исследования проведены при финансовой поддержке РФФИ (проект № 09-04-00146-а) и ДВО РАН (проект №09-1-ОБН-12, № 09-III-A-09-498.).

Режимы динамики двухкомпонентной модели Лефковича с плотностным лимитированием пополнения репродуктивной группы

Неверова Г.П.

Институт комплексного анализа региональных проблем ДВО РАН, Биробиджан, Россия
Email: galina.nev@gmail.com

**Spatial synchronization in a
metapopulation models**
M.P. Kulakov

Abstract

In this paper we investigate a mathematical model of dynamics for the population number with incomplete transition of young individuals to mature age group. We have considered the case when a part of the adjoint reproductive individuals runs density limitation on the part of the entire population. The analytical and numerical analysis of the model is made.

В работе изучается процесс пространственной синхронизации в системе глобально связанных популяций. Для популяций образующих в пространстве некоторую линейную цепочку описываются закономерности формирования кластеров синхронизации с циклами длины 2.

1 Уравнения динамики

В данной работе продолжено исследование математической модели, описывающей динамику численности двухвозрастной популяции с сезонным характером размножения [4, 5]. Рассматривается случай, когда коэффициент выживаемости особей, перешедших из младшего возрастного класса в старший, является функцией численности обеих возрастных групп. Уравнения динамики популяции имеют вид:

$$\begin{cases} x_{n+1} = a \cdot y_n + b \cdot x_n, \\ y_{n+1} = x_n \cdot s(x_n, y_n) + v \cdot y_n \end{cases} \quad (1)$$

где x - численность младшего возрастного класса, y - численность старшего возрастного класса, составляющего репродуктивную часть популяции, n - номер периода размножения, a - коэффициент рождаемости, v ($0 \leq v \leq 1$) - коэффициент выживаемости взрослых особей, b - доля молодежи, оставшаяся в неполовозрелом классе, $s(x, y)$ - доля молодежи, перешедшая в старший возрастной класс,

Функция $s(x, y)$ выбрана по аналогии с моделью Рикера в виде $s(x, y) = e^{-\alpha \cdot x - \beta \cdot y}$, где α - коэффициент, описывающий интенсивность воздействия особей младшего возрастного класса, коэффициент β характеризует интенсивность воздействия особей второго возрастного класса. Предполагается, что система (1) замкнутая, поэтому справедливо неравенство $d + e^{-\alpha \cdot x - \beta \cdot y} \leq 1$ [2].

2 Результаты исследования

Проведено подробное исследование модели. Показано, что возможно единственное нетривиальное стационарное состояние системы. Определена область его устойчивости, проанализирован характер потери устойчивости и сценарии переходов динамических режимов. Удобным параметром для исследования последовательности бифуркаций является величина $c = \beta / (\alpha a)$: при такой параметризации удается построить все возможные параметрические портреты. Следует отметить, что рост числа особей, которые остаются в данном году в младшей группе, приводит к расширению области значений параметров рождаемости и выживаемости репродуктивной группы, при которых популяция развивается стабильно.

Показано, что добавление к самолимитированию выживаемости младшей возрастной группой лимитирования взрослыми особями (уменьшение выживаемости особей ставшими половозрелыми с ростом численности взрослых) приводит (при $0 < c \leq (b+2)/(b^2+b+2)$) к существенному росту области устойчивости. Потеря устойчивости может произойти только при комплексно-сопряженных корнях характеристического уравнения, при переходе $|\lambda|$ через 1, и сопровождается появлением предельных инвариантных кривых, которые при дальнейшем изменении параметров v и a от границы устойчивости разрушаются с образованием весьма сложных предельных структур.

На рис. 1 представлена бифуркационная диаграмма, системы (1) ($0 < c \leq (b+2)/(b^2+b+2)$), демонстрирующая зависимость предельного

распределения численности младшего возрастного класса (x) от величины параметра a . Диаграмма дополнена «портретом» аттрактора. Для визуализации области регулярного,

квазипериодического и хаотического поведения на рис. 1 представлен график значений старшего ляпуновского показателя (λ) и размерности аттрактора (D) [1,3].

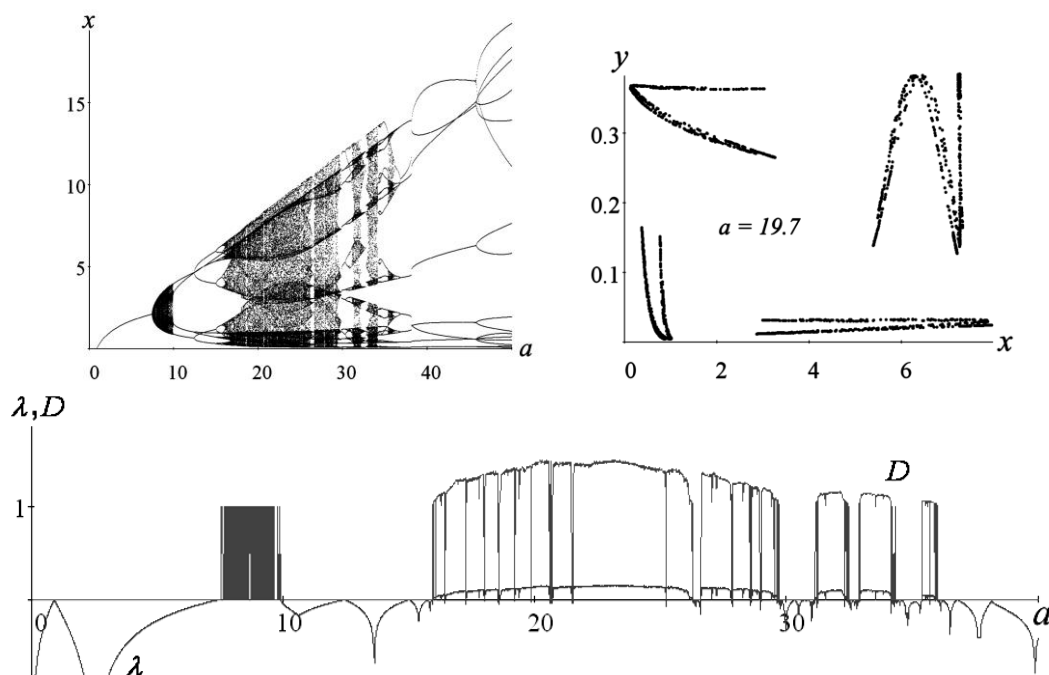


Рисунок 1. Бифуркационная диаграмма и графики старшего ляпуновского показателя (λ) и размерности (D) аттрактора в зависимости от значения параметра a при $v = 0.05$, $d = 0.05$, $\beta/\alpha = 0.01$

Бифуркационная диаграмма и график размерности хорошо дополняют друг друга. Потеря устойчивости приводит к предельному циклу (инвариантной кривой), размерности 1, а далее наблюдаются серии переходов инвариантных кривых, циклов конечной длины, аттракторов различной размерности.

Дальнейший рост ограничения выживаемости младшей группы взрослыми особями при $(b+2)/(b^2+b+2) < c < (b+3)/(b^2+2b+1)$ приводит к уменьшению области устойчивости нетривиального равновесия. При этом меняется качественная картина потери устойчивости: кроме возможной потери устойчивости при комплексных собственных числах, появляется большая зона значений параметров v и a (увеличивающаяся с ростом параметра c), попадание в которую сопровождается переходом одного из собственных чисел через -1 и появлением предельного 2-цикла. Следует отметить, что чем больше значение коэффициента d , тем меньше диапазон значений параметра c , при которых переход к хаотической динамике возможен двумя способами.

С дальнейшим ростом значений параметра c ($c \geq (b+3)/(b^2+2b+1)$) потеря устойчивости может произойти только при переходе одного из собственных чисел через -1 , т.е. по сценарию Фейгенбаума. В этом случае область устойчивости продолжает сужаться, и лимитирование выживаемости пополнения младшим возрастным классом практически не играет роли.

Исследования проведены при частичной финансовой поддержке РФФИ (проект № 09-04-00146 а) и ДВО РАН (конкурсный проект №09-I-ОБН-12)

Список литературы

- [1] Кузнецов С.П. Динамический хаос. М.: Физматлит, с.296, 2001.
- [2] Логофет Д.О., Клочкова И.Н. Математика модели Лефковича: репродуктивный потенциал и асимптотические циклы // Математическое моделирование, том 14, № 10, с. 116-126, 2002.
- [3] Неймарк Ю.И., Ланда П.С. Стохастические и хаотические колебания. М.: Наука, 1987. С.424.
- [4] Фрисман Е.Я., Неверова Г.П., Ревуцкая О.Л., Кулаков М.П. Режимы динамики модели двухвозрастной популяции // Прикладная нелинейная динамика, том 18, №2, с.113-130, 2010.
- [5] Фрисман Е.Я., Скалецкая Е.И. Странные аттракторы в простейших моделях динамики численности биологических популяций // Обозрение прикладной и промышленной математики, том 1, № 6, с. 988-1004, 1994.

Стохастическая модель динамики популяций, развивающихся в условиях воздействия токсичных веществ

Пичугин Б. Ю., Перцев Н. В., Логинов К. К.

Омский филиал Института математики им. С. Л.Соболева СО РАН
bpichugin@mail.ru, homlab@ya.ru

Abstract

We consider a stochastic model, describing dynamics of m competitive species, which are evolving under the influence of toxicants.

1 Введение

Одно из направлений в моделировании живых систем связано с изучением динамики популяций, развивающихся в нестационарных условиях среды обитания. На динамику популяций могут оказывать влияние разнообразные факторы, среди которых можно выделить температурный режим, емкость среды, наличие или недостаток ресурсов питания, загрязняющие вещества и т.д. Продолжительность жизни особей, численность и выживаемость их потомства может снижаться под воздействием разнообразных токсичных веществ, поступающих в среду обитания. Исследованию влияния этого фактора на динамику популяций посвящено достаточно большое количество публикаций (см., например, краткий обзор из работы [1]). Многие из предложенных моделей представляют собой системы дифференциальных уравнений, построенные на базе логистической модели (модель Ферхюльста–Пирла) и модели динамики конкурирующих популяций (модель Лотки–Вольтерра). Уравнения базовых моделей дополнены слагаемыми, отражающими гибель особей при непосредственном контакте с токсичными веществами, а также уравнениями для количества токсичных веществ. Настоящая работа посвящена стохастическому моделированию динамики конкурирующих популяций, подверженных воздействию токсичных веществ, поступающих в среду обитания особей.

2 Описание стохастической модели

Рассматривается биологическое сообщество, состоящее из особей популяций $A_1, \dots, A_m$. Предполагается, что динамика этого сообщества определяется следующими факторами: особи дают потомство; особи погибают вследствие конкуренции с другими особями сообщества; приток особей извне и миграция особей отсутствуют; в среду обитания особей поступает

токсичное вещество; взаимодействие особей с токсичным веществом может приводить к их гибели; токсичное вещество расходуется при взаимодействии с особями; с течением времени токсичное вещество распадается.

В качестве модели такого сообщества рассмотрим случайный процесс $Z(t) = (x(t), c(t))$, где $x(t) = (x_1(t), \dots, x_m(t)) \in \mathbf{Z}_+^m$ — вектор численностей популяций, $c(t) \geq 0$ — количество токсичного вещества в среде обитания особей в момент времени $t \geq 0$.

Параметры модели:

- $\beta_i > 0$ — интенсивность производства потомства особями популяции A_i ;
- $\lambda_i(u) = \sum_{j=1}^m \gamma_{ij} u_j$, $\gamma_{ij} \geq 0$, — интенсивность гибели одной особи популяции A_i вследствие конкуренции с особями сообщества, при условии, что численности популяций описываются вектором $u = (u_1, \dots, u_m)$;
- $r(t) \geq 0$ — кусочно-непрерывная, ограниченная на $[0, \infty)$ функция, задающая скорость поступления токсичного вещества в среду обитания особей;
- $\delta > 0$ — интенсивность снижения количества токсичного вещества за счет естественного распада и потери токсичности;
- $\mu_i > 0$ — интенсивность контактов одной особи популяции A_i с токсичным веществом;
- $F_i(w, c) \in [0; 1]$ — вероятность того, что в момент контакта особи популяции A_i с токсичным веществом будет израсходовано не более $w \geq 0$ токсичного вещества при условии, что на момент контакта общее количество токсичного вещества равно $c > 0$, $F_i(0, c) = 0$, $F_i(w, c) = 1$ при $w \geq c$;
- $\sigma_i > 0$ — пороговое количество токсичного вещества для особей популяции A_i : если в момент контакта особи с токсичным веществом израсходовано больше чем σ_i этого вещества, то особь погибает;
- $x(0) \in \mathbf{Z}_+^m$ — начальные численности популяций;
- $c(0) \geq 0$ — начальное количество токсичного вещества.

Кроме того, введем семейство детерминированных функций $c_{\tau,v}(t)$, $\tau \geq 0$, $v \geq 0$, которые описывают динамику изменения количества токсичного вещества при условии, что особи сообщества с ним не контактируют. Примем, что каждая из функций указанного семейства является решением задачи Коши

$$\frac{dc_{\tau,v}}{dt} = r(t) - \delta c_{\tau,v}, \quad t \geq \tau, \quad c_{\tau,v}(\tau) = v,$$

где $\frac{dc_{\tau,v}}{dt}$ — правосторонняя производная по t .

Определим случайную последовательность $0 = t_0 < t_1 < \dots$ моментов скачков процесса $Z(t)$ следующим образом: для всех $k \in \mathbf{Z}_+$, $u \in \mathbf{Z}_+^m$, $\tau \geq 0$, $v \geq 0$, $s \geq 0$ положим

$$P(t_{k+1} > \tau + s \mid t_k = \tau, Z(t_k) = (u, v)) = \exp\left(-\int_{\tau}^{\tau+s} Q_{\tau,u,v}(t) dt\right),$$

где $Q_{\tau,u,v}(t) = \sum_{i=1}^m (\beta_i + \lambda_i(u) + \mu_i c_{\tau,v}(t)) u_i$, $t \geq 0$.

Отметим, что последовательность $\{t_k\}$ может состоять из конечного числа элементов: если в момент $t_k = \tau$ все популяции выродились ($u = 0$), то $Q_{\tau,u,v}(t) \equiv 0$ и

$P(t_{k+1} > \tau + s \mid t_k = \tau, Z(t_k) = (0, v)) = 1$ для любого s . В этом случае $t_{k+1} = \infty$.

Для любого $t \in (t_k, t_{k+1})$ примем, что

$$Z(t) = (x(t_k), c_{t_k, c(t_k)}(t)).$$

Определим распределение скачка процесса $Z(t)$ в момент t_{k+1} для всех $k \in \mathbf{Z}_+$ таких, что $t_{k+1} < \infty$. Введем случайную величину $\theta_i(\tau')$, равную количеству токсичного вещества, которое будет израсходовано в момент τ' , если в этот момент произойдет контакт особи популяции A_i с токсичным веществом. Для всех $i = 1, \dots, m$, $0 \leq \tau < \tau' < \infty$, $u \in \mathbf{Z}_+^m$, $u \neq 0$, $v \geq 0$, $w \geq 0$ положим

$$H = \{t_k = \tau, t_{k+1} = \tau', Z(t_k) = (u, v)\},$$

$$P(Z(\tau') = (u + e_i, c_{\tau,v}(\tau')) \mid H) = \beta_i u_i Q_{\tau,u,v}^{-1}(\tau'), \quad (1)$$

$$P(Z(\tau') = (u - e_i, c_{\tau,v}(\tau')) \mid H) = \lambda_i(u) u_i Q_{\tau,u,v}^{-1}(\tau'), \quad (2)$$

$$P(\theta_i(\tau') \leq w \mid H) = F_i(w, c_{\tau,v}(\tau')),$$

$$P(Z(\tau') = (u - e_i, c_{\tau,v}(\tau') - w) \mid H, \theta_i(\tau') = w) = I(w > \sigma_i) \mu_i c_{\tau,v}(\tau') u_i Q_{\tau,u,v}^{-1}(\tau'), \quad (3)$$

$$P(Z(\tau') = (u, c_{\tau,v}(\tau') - w) \mid H, \theta_i(\tau') = w) = I(w \leq \sigma_i) \mu_i c_{\tau,v}(\tau') u_i Q_{\tau,u,v}^{-1}(\tau'), \quad (4)$$

где $e_i \in \mathbf{Z}^m$ — вектор, у которого координата номер i равна 1, а остальные координаты нулевые, $I(\cdot)$ — индикаторная функция, равная 1, когда условие в скобках выполнено и 0 — в противном случае.

Равенство (1) описывает появление новых особей, (2) — гибель особей в результате

конкуренции и самолимитирования, (3), (4) — взаимодействие особей с токсичным веществом: если в результате взаимодействия особи популяции A_i с токсичным веществом израсходовано $w > \sigma_i$ этого вещества, то особь погибает.

3 Вычислительные эксперименты с моделью

Аналитическое исследование случайного процесса $Z(t)$ представляет собой трудную задачу. Поэтому для изучения его свойств использован метод статистического моделирования.

Исследовалась динамика двух конкурирующих популяций, подверженных воздействию токсичного вещества. Параметры модели β_i , γ_{ij} , $i, j = 1, 2$, были выбраны так, чтобы выполнялись

$$\text{неравенства } \frac{\gamma_{12}}{\gamma_{22}} < \frac{\beta_1}{\beta_2} < \frac{\gamma_{11}}{\gamma_{21}}. \text{ Эти неравенства}$$

означают, что детерминированная модель Лотки–Вольтерра для численности $(y_1(t), y_2(t))$ двух популяций (без учета токсичного вещества) имеет глобально асимптотически устойчивое положение равновесия (y_1^*, y_2^*) такое, что $y_1^* > 0$, $y_2^* > 0$.

Целью экспериментов являлось изучение динамики математических ожиданий $E x_1(t)$, $E x_2(t)$, $E c(t)$ в зависимости от вариации параметров μ_1 , μ_2 , σ_1 , σ_2 модели.

Получены условия на параметры модели, при которых возможно вырождение одной из популяций, а также сосуществование обеих популяций.

Один из вариантов описанной выше модели использован для изучения динамики популяций, подверженных воздействию вредных веществ, поступающих в организм особей в составе ресурсов питания.

Все вычисления проводились на суперкомпьютере МВС-1000/128, установленном в Омском филиале Института математики им. С.Л. Соболева СО РАН. Для генерации псевдослучайных чисел на различных процессорах применялся 128 битный конгруэнтный датчик псевдослучайных чисел.

Работа поддержана РФФИ, проект № 09-01-00098-а.

References

- [1] Н. Перцев, Г. Царегородцева. Математическая модель динамики популяции, развивающейся в условиях воздействия вредных веществ // Сиб. журн. инд. матем., том 13, N.1 (41), с.109-120, 2010.

Определение оптимального промыслового изъятия для двухвозрастной лимитированной популяции

Ревуцкая О.Л.

Институт комплексного анализа региональных проблем ДВО РАН

oksana-rev@mail.ru

Abstract

The optimization problem is stated and investigated. This problem was to find the optimum quotas and a steady value number of the population providing the theoretically possible maximum of for the equilibrium withdrawal. It is shown that optimal harvesting is one out of a single age class, which is determined by the values of population parameters and the prices ratio.

1 Постановка задачи оптимального управления динамикой численности

Рассматривается динамика численности эксплуатируемой популяции, которая может быть представлена совокупностью двух возрастных классов: младшего, включающего неполовозрелых особей, и старшего, состоящего из особей, участвующих в размножении. Предполагается, что времени, протекающего между двумя последовательными периодами размножения, достаточно для развития особей младшего возраста до половозрелого состояния, а новорожденных особей – до состояния младшего возраста.

Обозначим n – номер сезона размножения; x и y – численности младшего и старшего возрастного класса, соответственно. При условии, что на развитие младшего возрастного класса оказывают влияние плотностно-зависимые факторы, уравнения динамики численности двухвозрастной промысловой популяции имеют вид

$$\begin{cases} x_{n+1} = ay_n(1-u_1) \\ y_{n+1} = (s(x_n, y_n)x_n + vy_n)(1-u_2) \end{cases}, \quad (1)$$

где a ($a>0$) – репродуктивный потенциал популяции, v ($0<v<1$) – коэффициент выживаемости взрослых особей, $s(x, y)$ – функция выживаемости неполовозрелых особей, зависящая от плотности молодежи и взрослых особей, u_1 и u_2 – доли изъятия неполовозрелых и половозрелых особей соответственно.

Доли изъятия u_1 и u_2 особей двух возрастных классов при равновесном промысле собираются в урожай, суммарный доход от которого определяется соотношением

$$I = c_1 u_1 (ay) + c_2 u_2 (s(x, y)x + vy), \text{ или } I = c_1 R_1 + c_2 R_2,$$

где c_1 и c_2 – средняя цена одной особи младшего и старшего возрастного класса, соответственно, R_i

($i=1,2$) – количество отловленных особей из соответствующей группы.

Задача оптимизации заключается в определении оптимальных долей изъятия и равновесных (оптимальных) значений численности популяции, которые бы обеспечили максимально возможный стабильный промысел, и в итоге с учетом цен давали максимальный доход от изъятия.

Поставленная задача рассматривалась для двух видов функции выживаемости неполовозрелых особей: а) $s(x, y) = \exp(-x - qy)$; б)

$s(x, y) = 1 - x - qy$, где $q = \beta/\alpha$, где α и β – коэффициенты, характеризующие интенсивности конкурентного воздействия неполовозрелых и половозрелых особей. Фактически, здесь была реализована замена переменных $\alpha x \rightarrow x$, $\alpha y \rightarrow y$, после которой новые переменные представляют собой «относительные» численности популяции.

2 Решение задачи оптимизации промысла двухвозрастной популяции

Решение задачи оптимизации равновесного промысла показало, что максимум дохода от изъятия достигается только на границах $u_1=0$ или $u_2=0$. При этом изъятие особей одновременно из двух возрастных классов не является оптимальным.

В связи с этим задача оптимизации была рассмотрена для случаев: а) $u_1=0$, то есть управляющее воздействие полностью определяется долей (u_2) изъятия половозрелых особей (y), при этом молодежь (x) экономического интереса не представляет, и б) $u_2=0$, то есть изымается только доля (u_1) особей из младшей возрастной группы (x).

Для каждого случая определены оптимальные значения численностей популяции и оптимальные доли изъятия, а также условия их существования и устойчивости. Оказалось, что «оптимальный» промысел с постоянной долей изъятия может дестабилизировать динамику системы. В силу этого была выбрана стратегия промысла, заключающаяся в том, что сначала популяция приводится к такой численности x_M , y_M , при которой ее прирост максимальный, а затем удерживается на этом уровне за счет сбора урожая (R_i , $i=1,2$) до конца процесса промысла [1]. Оптимальное управление в общем виде запишется

$$\begin{cases} R_n = z_n - z_M & \text{при } z_n \geq z_M, \\ R_n = 0 & \text{при } z_n < z_M, \end{cases} \quad (2)$$

где z_n и z_M – текущая в n -год и оптимальная численность эксплуатируемого возрастного класса. Если каждый год после промысла оставлять z_M особей, то ежегодно будем получать одинаковый максимально возможный урожай R . Это означает, что если численность z_n популяции превосходит значение z_M , то оптимальное решение состоит в изъятии ($z_n - z_M$) особей [2]. При такой стратегии оптимальное изъятие (2) всегда стабилизирует динамику системы (1).

2.1 Оптимальный сбор урожая из половозрелой части популяции

Равновесные значения численностей популяции, обеспечивающие теоретически возможный максимум равновесного улова, находятся из выражений

а) при $s = 1 - x - qy$:

$$x_M = \frac{a + v - 1}{2(a + q)}, \quad y_M = \frac{a + v - 1}{2a(a + q)}.$$

б) при $s = \exp(-x - qy)$:

$$a = \frac{(1 - v) \exp((a + q)y_M)}{1 - (a + q)y_M}, \quad x_M = ay_M.$$

При оптимальном изъятии половозрелых особей в случае линейного лимитирования равновесное значение численности снижается в два раза.

Если начальная численность популяции совпадает с оптимальной ($x_0 = x_M$, $y_0 = y_M$), то управление с самого начала становится стационарным, то есть, если в перспективе имеются n ($n = 1, 2, \dots, k$) лет эксплуатации, тогда оптимальная доля изъятия (u_2) и максимальный суммарный промысел ($R_2 = R$) находятся по формулам:

а) при $s = 1 - x - qy$: $u_2 = \frac{a + v - 1}{a + v + 1}$,

$$R_n = y_0 + \frac{(n - 1)(a + v - 1)^2}{4a(a + q)}.$$

б) при $s = \exp(-x - qy)$: $u_2 = 1 - \frac{1 - (a + q)y_M}{1 - (a + q)y_M}$,

$$R_n = y_0 + \frac{(n - 1)(1 - v)(a + q)y_M^2}{1 - (a + q)y_M}.$$

Если начальная численность эксплуатируемой части популяции больше оптимального уровня, то есть $y_0 > y_M$, тогда максимальный суммарный промысел составит

а) при $s = 1 - x - qy$:

$$R_n = y_0 + \sum_{i=2}^n (x_{i-2}(1 - x_{i-2})) + ((n - 1)(v - 1) - q \sum_{i=2}^n x_{i-2}) y_M$$

б) при $s = \exp(-x - qy)$:

$$R_n = y_0 + (n - 1)(v - 1)y_M + \exp(-qy_M) \sum_{i=2}^n (x_{i-2} \exp(-x_{i-2})).$$

2.2 Оптимальный сбор урожая из неполовозрелой части популяции

Представим результаты для случая $s = 1 - x - qy$.

Оптимальные значения численностей составляют

$$x_M = \frac{H + (a + q)(v - 1)}{q(a + q)},$$

$$y_M = \frac{(1 + a + q - v)((1 - v)(1 + q - v) - H)}{(a(v - 1) - H)((a + q)(1 + q - v) + H)},$$

где $H = (a(1 - v)(1 + q - v)(a + q))^{1/2}$.

Если $x_0 = y_M$, $y_0 = y_M$, тогда оптимальная доля изъятия (u_1) и максимальный урожай ($R_1 = R$) находятся по формулам:

$$u_1 = \frac{a(a + q) - H}{a(1 + a + q - v)},$$

$$R_n = x_0 +$$

$$+ (n - 1) \frac{(H - a(a + q))((1 - v)(1 + q - v) - H)}{(a(1 - v) + H)((a + q)(1 + q - v) + H)}.$$

Если $x_0 > x_M$, тогда максимальный суммарный урожай составит

$$R_n = x_0 + a(1 - n)x_M^2 +$$

$$+ (a((n - 1) - q \sum_{i=2}^n y_{i-2}) - n)x_M + a(y_0 + v \sum_{i=2}^n y_{i-2}).$$

В случае экспоненциального лимитирования выживаемости молоди ($s = \exp(-x - qy)$) задача оптимизации промысла молоди решается аналогично.

Таким образом, наибольший средний урожай достигается при оставляемой ежегодно такой численности, при которой осуществляется максимальный прирост популяции.

Оптимальным является изъятие только из одного возрастного класса, а какого именно определяется значениями популяционных параметров и соотношением цен.

Исследования проведены при частичной финансовой поддержке РФФИ (№ 09-04-00146-а) и ДВО РАН (№ 09-I-P15-01, №09-I-ОБН-12, № 09-II-CO-06- 006).

References

- [1] Свирижев Ю.М. О математических моделях биологических сообществ и связанных с ними задачах управления и оптимизации // Мат-лы I школы по математическому моделированию сложных биологических систем «Математическое моделирование в биологии». М.: Наука, 1975. С. 30-52.
- [2] Скалецкая Е.И., Фрисман Е.Я., Шапиро А.П. Дискретные модели динамики численности и оптимизации промысла. М.: Наука, 1979. 168 с.

Моделирование динамики популяции, развивающейся в условиях воздействия вредных веществ на процесс репродукции особей

Царегородцева Г.Е.

Омский филиал Института математики им. С. Л. Соболева СО РАН
galinatsar@ya.ru

Abstract

The mathematical model of the population developing in the conditions of consumption of harmful substances, making negative impact on reproduction individuals is presented. The equations of model are described and properties of model's solutions, including existence and stability of equilibriums are established.

1 Введение

Динамика популяций зависит от воздействия ряда факторов, влияющих на численность и выживаемость потомства, на продолжительность жизни и миграции особей. Такими факторами могут являться температурный режим, емкость среды, наличие или недостаток ресурсов питания, уровень загрязнения среды обитания особей, токсичность потребляемых пищевых ресурсов и т.д. Настоящая работа посвящена моделированию динамики популяции, подверженной воздействию разнообразных загрязняющих веществ, поступающих в организм особей из среды их обитания в составе ресурсов питания. Для описания динамики популяции строится модель вида «ресурс-потребитель», учитывающая процессы размножения, самолимитирования, естественной смертности и миграции особей, а также потребление особями загрязняющих веществ. Предполагается, что особи могут быть подвержены влиянию любого из потребляемых загрязняющих веществ и, особенно, влиянию продуктов взаимодействия этих веществ в различных комбинациях. Каждое из таких загрязняющих веществ и продукты их взаимодействия будем называть вредными веществами. Целью работы являлось построение уравнения модели и изучение характерных режимов динамики популяции в условиях негативного воздействия вредных веществ на процесс репродукции особей.

2 Уравнения модели

Будем рассматривать популяцию особей, динамика которой определяется следующими факторами: 1) особи погибают вследствие

процессов старения и самолимитирования, 2) особи подвержены миграции, 3) приток особей извне отсутствует, 4) в среду обитания особей поступают загрязняющие вещества $C_1, C_2, \dots, C_k$, которые распадаются, накапливаются в пищевых ресурсах и потребляются особями, 5) конкуренция особей за пищевые ресурсы, содержащие вещества $C_1, C_2, \dots, C_k$, отсутствует, 6) потребляемые вещества $C_1, C_2, \dots, C_k$, взаимодействуют между собой в различных комбинациях, 7) некоторые из веществ $C_1, C_2, \dots, C_k$ или продуктов их взаимодействия оказывают негативное влияние на скорость репродукции особей.

Обозначим $x = x(t)$ - численность популяции в момент времени t , а через $c_i = c_i(t)$ - количество загрязняющих веществ C_i в момент времени t , $1 \leq i \leq k$. Функция $d(x(t))$ задает интенсивность уменьшения численности популяции вследствие миграции и гибели особей. Считаем, что новые особи производятся за счет существующих особей, независимо от их возраста. Функция $b(x(t))$ описывает интенсивность роста численности популяции за счет рождения новых особей. Пусть функция $\beta(u(t))$ отражает влияние вредных веществ на репродуктивный потенциал особей. Уровень вредных веществ задается функцией $u(t) = \varphi(c_1(t), \dots, c_k(t))$, которая учитывает некоторые или все вредные вещества, а также продукты их взаимодействия. Принимаем, что в условиях воздействия вредных веществ интенсивность роста численности популяции описывается выражением $\beta(u(t))b(x(t))$. Используя сделанные предположения, можно записать, что

$$\dot{x}(t) = \beta(u(t))b(x(t))x(t) - d(x(t))x(t).$$

Зафиксируем $1 \leq i \leq k$ и примем далее, что количество загрязняющих веществ C_i в среде обитания особей изменяется по правилу:

$$\dot{c}_i(t) = r_i(t) - \theta_i(c_i(t))x(t) - \delta_i c_i(t).$$

В этом уравнении функция $r_i(t)$ и константа δ_i задают соответственно скорости поступления и распада загрязняющего вещества C_i . Функция $\theta_i(c_i)$ описывает скорость потребления одной

особью вещества C_i в составе пищевых ресурсов с учетом эффекта насыщения.

Объединяя полученные уравнения и добавляя начальные данные, приходим к модели

$$\dot{c}_i = r_i(t) - \theta_i(c_i)x - \delta_i c_i, \quad 1 \leq i \leq k,$$

$$\dot{x} = \beta(\varphi(c_1, \dots, c_k))b(x)x - d(x)x, \quad t > 0,$$

$$c_j(0) = c_j^{(0)} \geq 0, 1 \leq j \leq k, x(0) = x^{(0)} \geq 0.$$

Совокупность функций $\theta_i(c_i)$, $1 \leq i \leq k$, может участвовать в построение функции $\varphi(c_1, \dots, c_k)$. Так, например, предполагая, что вредное вещество образуется из набора только всех загрязняющих веществ, можно принять, что $\varphi(c_1, \dots, c_k) = \rho \prod_{i=1}^k \theta_i^{n_i}(c_i)$. В этой формуле константа $\rho > 0$ отражает скорость реакции взаимодействия потребленных загрязняющих веществ, а константы $n_i > 0$ определяют порядок реакции для отдельных потребляемых веществ C_i , $1 \leq i \leq k$. Функции $r_i(t)$ можно рассматривать как регулируемые или нерегулируемые скорости поступления загрязняющих веществ, попадающих в среду обитания особей из различных источников.

3 Свойства решений модели

Полагаем, что функции $b(x)$, $d(x)$ обладают следующими свойствами: $b(x)$ определена, непрерывна, неотрицательна на промежутке $0 \leq x < \infty$ и является ограниченной, т.е. $b(x) \leq \bar{b}$

для всех $x \in [0, \infty)$, $\bar{b} > 0$ - некоторая константа; $d(x)$ определена, непрерывна, положительна на промежутке $0 \leq x < \infty$, являются возрастающей и $\lim_{x \rightarrow +\infty} d(x) = +\infty$. Кроме того исключим из рассмотрения простейший случай, когда $b(x) \leq d(x)$ для всех $0 \leq x < \infty$. Для описания эффекта воздействия вредных веществ на репродуктивный потенциал особей введем следующие предположения. Функция $\beta(u)$ определена, непрерывна, положительна, убывает на промежутке $0 \leq u < \infty$, $\beta(0) = 1$, $\lim_{u \rightarrow +\infty} \beta(u) = 0$. Функция $\varphi(c_1, \dots, c_k)$ определена и непрерывна при всех $0 \leq c_i < \infty$, $1 \leq i \leq k$, не убывает по каждой переменной c_i , $1 \leq i \leq k$, $\varphi(0, \dots, 0) = 0$, $\varphi(c_1, \dots, c_k) > 0$ для всех $c_1 > 0, \dots, c_k > 0$. Полагаем, что функция $r_i(t)$ определена, неотрицательна и непрерывна при $0 \leq t < \infty$, $r_i(t) \leq \bar{r}_i$, $\bar{r}_i > 0$ - некоторая константа. Предполагаем, что $\theta_i(c_i)$ определена и непрерывна на промежутке $0 \leq c_i < \infty$

, $\theta_i(0) = 0$, является возрастающей и имеет

конечный предел $\bar{\theta}_i > 0$ при $c_i \rightarrow +\infty$.

В рамках сделанных предположений доказана теорема о существовании и единственности, неотрицательности и ограниченности решений на $[0, \infty)$. Исследованы условия существования и устойчивости положений равновесия, при $r_i(t) = r_i = \text{const} > 0$ для всех $1 \leq i \leq k$. Эти условия сводятся к анализу корней уравнения $\beta(\varphi(\bar{c}_1(x), \dots, \bar{c}_k(x)))b(x) = d(x)$, где $\bar{c}_i(x)$ решение уравнения $r_i - \delta_i c_i = \theta_i(c_i)x$, $1 \leq i \leq k$.

Изучены возможные режимы динамики популяции под влиянием регулируемого поступления загрязняющих веществ (переход с одного устойчивого уровня на другой, вплоть до вырождения популяции).

References

- [1] B. Dubey, J. Hussain. Nonlinear models for the survival of two competing species dependent on resource in industrial environments // *Nonlinear Analysis: Real World Applications*, volume 4, p.21-44, 2003.
- [2] H.I. Freedman, J.B. Shukla. Models for the effect of toxicant in single species and predator prey systems // *J. Math. Biol.*, volume 30, p.15-30, 1991.
- [3] T.G. Hallam, C.E. Clark, R.R. Lassiter. Effects of toxicants on populations: A qualitative approach. I. Equilibrium environment exposure // *Ecol. Model.*, volume 18, p.291-304, 1983.
- [4] B. Liu, L. Chen, Y. Zhang. The effects of impulsive toxicant input on a population in a polluted environment // *J. of Biol. Sys.*, volume 11, N.3, p.265-274, 2003.
- [5] Z. Ma, G. Gui, W. Wang. Persistence and extinction of a population in a polluted environment // *Math. Biosci.*, volume 101, p.75-97, 1990.
- [6] R. Naresh, S. Sundar, J. Shukla. Modeling the effect of an intermediate toxic product formed by uptake of a toxicant on plant biomass // *Appl. Math. and Comput.*, volume 182, p.151-160, 2006.
- [7] Н. Перцев, Г. Царегородцева. Математическая модель динамики популяции, развивающейся в условиях воздействия вредных веществ // *Сиб. журн. инд. матем.*, том 13, N.1 (41), с.109-120, 2010.
- [8] А.Пичугина. Интегродифференциальная модель популяции, подверженной воздействию вредных веществ // *Сиб. журн. инд. матем.*, том 7, N.4 (20), с.130-140, 2004.

Conceptual Modelling In Biomedical Knowledge Discovering

Bogatyrev, M.Y. Tuhtin, V.V.
Tula State University, Tula, Russia
okkambo@mail.ru

Abstract

Knowledge discovering technology based on applying conceptual graphs and conceptual structures is presented. It can be prepared as domain-oriented technology for biomedical texts.

1. Introduction

Overwhelming amount of biomedical knowledge recorded in texts inspired the development of text mining methods for knowledge discovering. Text mining becomes an actual branch in modern bioinformatics [1], [2], [3].

The idea of Knowledge Discovering from Data (KDD) appeared more than twenty years ago. Now still there are no universal KDD-systems capable to extract knowledge from arbitrary natural language texts. All real KDD-systems are *domain oriented* that means that they work only with texts of certain themes. Now these domain oriented texts are mainly stored in online information systems such as MedLine, BioMed, PubMed, etc. Using domain oriented data (texts) KDD-systems are also *problem oriented*: they have been applied for solving problems in systems biology, genetics, ecology, etc.

Another distinctive feature of KDD-systems is attraction of additional resources – domain oriented *dictionaries, taxonomies, ontologies* needed for practical implementing text mining methods.

In bioinformatics the most text mining systems use traditional information retrieval methods, which rely on keyword-based approaches. These approaches are based on finding patterns in the text which correspond to possible knowledge.

To get the knowledge, KDD-system has to “understand” text. In natural languages processing there are many approaches oriented on solving this problem. Among them there is one approach based on replacing analyzing text by its semantic model. This way seems to be more effective than traditional keyword-based methods because KDD-systems *can* “understand” semantic models.

2. Knowledge Discovering with Conceptual graphs

We propose a text mining approach based on applying conceptual semantic models of textual data stored as additional resource and applied in knowledge search queries.

This approach produces a KDD - technology having the following features.

1. Abstracts of scientific papers usually represent paper's content in short and clear form. They annotate information presented in full text of a paper. So the abstracts are objects of knowledge discovering.
2. To formalize abstracts, the simplest semantic model of sentence, *conceptual graph* [3], is applied. Conceptual graphs are constructive tool for solving text mining problems [5].
3. Conceptual graphs which were built for abstract's sentences then are aggregated to form *conceptual structure*. This structure being the formal semantic model of an abstract serves as a part of problem oriented ontology.
4. KDD – system based on this technology accepts queries in natural language. Executing queries, KDD – system transforms each input phrase into conceptual graph which is applied as a pattern in knowledge search technique. That makes knowledge search technique not keyword-based but graph-based.

Figure 1 illustrates the architecture of proposed KDD – system.

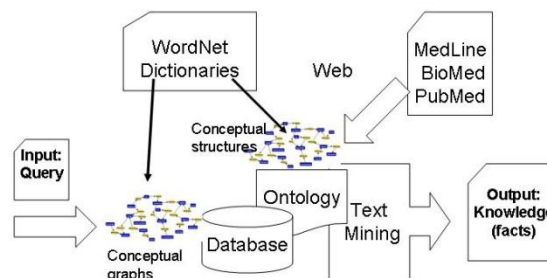


Figure 1. Architecture of KDD – system.

2. Conceptual graphs acquisition

The crucial problem in proposed technology is *conceptual graphs acquisition* from natural language

texts. This problem has no universal solution, i.e. one could not be sure to get *proper conceptual graph* from arbitrary text. There are solutions in the area of *domain oriented texts*. Biomedical texts belong to this area.

A *conceptual graph* is a bipartite graph that has two kinds of nodes called *concepts* and *conceptual relations*. Concepts are parts of speech from analyzing sentence – verbs, nouns, adjectives, etc. Relations do not belong to the analyzing sentence and constructing them is the main difficulty for conceptual graphs acquisition algorithms.

To acquire conceptual graph from a sentence we use existing approaches to realize lexical, morphological syntactic and semantic analysis. We apply *semantic roles labelling* [6] as the main instrument for building relations. Some special decisions were made for texts in Russian and English languages. Domain oriented algorithm works as a filter leaving only a sub-graph of full conceptual graph of a sentence. This sub-graph has been created with applying external resources.

For example, consider the sentence from the article [2]: “We show here that PAHX interacts with FKBP52, but not with FKBP12, suggesting that it is a specific target of FKBP52”.

Keyword-based approach [2] orients here on protein names (PAHX, FKBP52, FKBP12) and one interaction word (the verb “interacts”) only.

Our full conceptual graph for this sentence on Figure 2 is more informative.

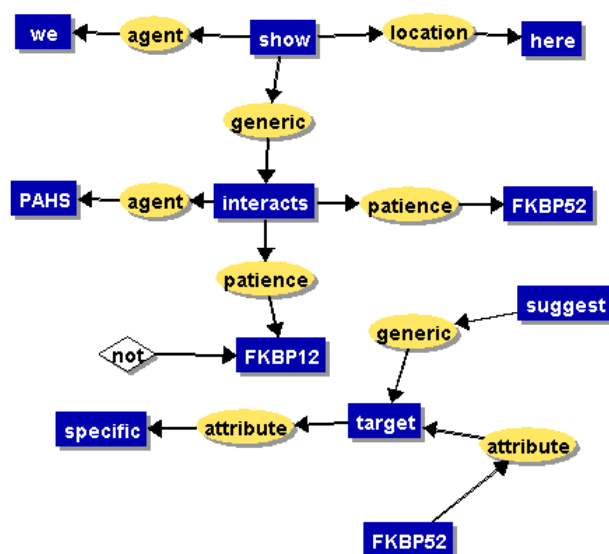


Figure 2. Full conceptual graph for the example sentence.

The portion of knowledge in the example sentence consists of three facts:

- 1) PAHX interacts with FKBP52;
- 2) PAHX does not interact with FKBP12;
- 3) it is a specific target of FKBP52.

The phrase “We show here” in the example sentence does not contain any knowledge.

The resultant filtered graph on Figure 3 additionally has some concepts in the form <type> : <entity> after

applying domain-oriented acquisition algorithm. Negation is presented as special *logical actor* “not” in the graph. Graph on Figure 3 represents all three facts as a portion of knowledge in example sentence.

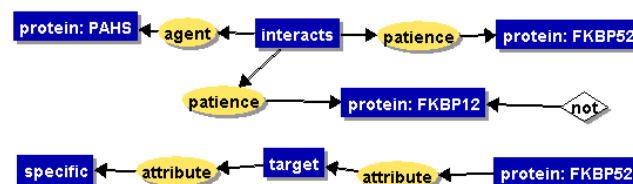


Figure 3. Filtered conceptual graph for the example sentence.

3. Current experiments

Using web interfaces we have started the series of experiments with public biomedical resource of PubMed [7]. First results show that KDD – technology described here has good perspectives in applying it as an effective text mining tool for knowledge discovering from biomedical texts.

At the same time some interesting problems in system’s developing need to be solved as interdisciplinary problems of computer science, linguistics and bioinformatics.

References

- [1] Rodriguez-Esteban R. Biomedical Text Mining and Its Applications. PLoS Comput Biol 5(12), 2009.
- [2] Chowdhary R, Zhang J, Liu JS. Bayesian inference of protein-protein interactions from biological literature. Bioinformatics, 25(12), 2009, p.p.1536-42.
- [3] Kolchanov N.A. et al. Technology For Gene Network Reconstruction and Analysis. – Proc. of the 2 nd Int. Conf. on Math. Biology and Bioinformatics, Pushchino, Russia, 2008, p.p. 86-87 (in Russian).
- [4] Sowa J. F. Conceptual structures: information processing in mind and machine. Addison Wesley, Reading, 1984.
- [5] Bogatyrev, M.Y. Tuhtin, V.V. Solving Some Text Mining Problems with Conceptual Graphs. In: Digital libraries: Advanced Methods And Technologies, Digital Collections. Proceedings of the Tenth Russian National Research Conference RCDL'2008 Dubna, JINR, 2008. p.p. 31-36. (in Russian).
- [6] Bogatyrev M. Y., Mitrofanova O. A., Tuhtin V. V. Building Conceptual Graphs for Articles Abstracts in Digital Libraries. – Proc. of the Conceptual Structures Tool Interoperability Workshop (CS-TIW 2009) at 17th International Conference on Conceptual Structures (ICCS'09) - Moscow, Russia, July 2009, - p.p. 50-57.
- [7] U.S. National Library of Medicine <http://www.ncbi.nlm.nih.gov/pubmed>

Интеграция данных для разработки и анализа моделей динамики биосферы

Молородов Ю.И.⁽¹⁾, Пестунов А.И.⁽¹⁾, Пестунов И.А.⁽¹⁾, Смирнов В.В.⁽¹⁾, Федотов А.М.^(1, 2)

Институт вычислительных технологий СО РАН⁽¹⁾,
Новосибирский государственный университет⁽²⁾
fedotov@ict.nsc.ru

Abstract

A service-oriented geo-information system for gathering, storage and processing of the satellite and ground-based data, created in the Institute of Computational Technologies of the Siberian Branch of the Russian Academy of Science is considered, based on the of Satellite Data Catalogue of the Novosibirsk Scientific Center of SB RAS.

В настоящее время в мировом сообществе информационные технологии играют все более значимую роль, формируя базовую инфраструктуру исследований, обеспечивающую эффективный междисциплинарный диалог представителей различных научных областей.

Интеграция информационных ресурсов в единую информационную среду и организация доступа к вычислительным ресурсам является одним из важнейших направлений развития современных информационных технологий.

Эти вопросы приобретают особую актуальность при разработке информационных моделей динамики биосферы, когда различные группы исследователей, разделенные географически, должны осуществлять совместную работу, обмен данными и знаниями, координировать свои действия с целью оптимизации использования данных и вычислительных ресурсов.

Следует отметить, что в настоящее время сбор данных об окружающей среде и эффективное использование полученной информации приобретают все большее значение. Данные об окружающей среде всегда уникальны, их сбор часто обходится очень дорого. По этим причинам огромную важность приобретает задача извлечения максимальной пользы от данных, полученных в результате каждого исследования.

В докладе рассматривается сервис-ориентированная геоинформационная система для сбора, хранения и обработки спутниковых и наземных данных, создаваемая в Институте вычислительных технологий СО РАН. Система создается на базе каталога спутниковых данных Новосибирского научного центра СО РАН [2],

который регулярно пополняется оперативными данными SPOT 4 и MODIS/Terra+Aqua и включает архивные данные со спутников серии LandSat на территорию РФ за 1982-2002 гг.

Система создается с учетом рекомендаций OGC на основе программных продуктов с открытым исходным кодом и работает под управлением операционной системы семейства UNIX. Подсистема картографических сервисов обеспечивает централизованный доступ к наборам векторных и растровых данных по протоколам WMS/WFS. Система позволяет публиковать пространственные данные как с локальных, так и распределенных источников. Для поиска данных по метаданным используется сервер GeoNetwork, обеспечивающий поддержку протокола Z39.50. В качестве базового инструментария для обработки и анализа данных дистанционного зондирования используются пакеты программ ESRI ENVI 4.7 и GRASS GIS с модулями расширения, созданными в ИВТ СО РАН.

Для обеспечения работ по созданию моделей динамики биосферы в эксплуатацию запущено несколько информационных ресурсов: 1) карта растительности Западной Сибири и ландшафтная карта Иркутской области, 2) карта почв бореальной зоны Западно-Сибирской равнины и соответствующая ей карта растительности, содержащая 28 различных типов растительности (М 1:7500000), 3) база данных по содержанию органического углерода в почвах Сибири, 4) тематическая карта «Содержание органического углерода в почвах Западно-Сибирской равнины» (М 1:7500000), 5) электронная библиотека, предназначенная для описания публикаций, документов, карт и других медиа-данных.

В настоящее время разрабатывается технология публикации методов обработки пространственных данных в виде веб-сервисов [1]. Она позволяет обеспечить широкому кругу потенциальных пользователей доступ к хранилищу современных наукоемких алгоритмов и вычислительным ресурсам, необходимым для оперативной обработки больших объемов информации.

Пользователями системы являются сотрудники 25 институтов СО РАН.

Работа выполнена в рамках междисциплинарного интеграционного проекта СО РАН № 50 «Модели изменения биосферы на основе баланса углерода (по натурным и спутниковым данным и с учетом вклада бореальных экосистем)», при поддержке РФФИ (грант № 09-07-12087-офи_м и 09-07-00277) и президентской программы «Ведущие научные школы РФ» (грант № НШ-931.2008.9).

Список литературы

- [8] Д.И. Добротворский, Е.А. Куликова, И.А. Пестунов, Ю.Н. Сиявский. Веб-сервисы для непараметрической классификации спутниковых данных. *Сб. матер. VI Междунар. научн. конгресса «ГЕО-Сибирь-2010»*, Новосибирск: СГГА, т. 1, ч. 2, с. 171-175, 2010.
- [9] Ю.И. Шокин, И.А. Пестунов, В.В. Смирнов и др. Корпоративная информационная система СО РАН сбора, хранения и обработки спутниковых данных. *Горный информационно-аналитический бюллетень*, отдельный выпуск “Кузбасс 2”, с. 3-10, 2009.

Проблемы вибрационной экологии и безопасности высотных сооружений

Дергузов А.В., Махортых С.А.

Пушино ИМПБ РАН

arkadid@list.ru, makh@impb.ru

Abstract

В работе предлагается метод оценки и прогнозирования передачи вибрации, вызванной техногенными, сейсмическими или метеорологическими факторами по структуре высотного здания. Используется спектрально-аналитический подход. Результаты моделирования проверялись в ходе натурных измерений. Результаты используются в системах экологического мониторинга в условиях города.

Актуальность

Задачи контроля виброакустической обстановки в городских условиях достаточно актуальны в силу многообразия действующих источников, среди которых не последнее место занимают линии метрополитена, железные дороги, строительные машины. В работе осуществляется расчет передачи вибрации от движения поездов на многоэтажные здания. Высотная застройка налагает дополнительные требования на корректность прогноза. При этом важность приобретают дополнительные факторы: сейсмические и метеорологические воздействия.

Нормативные абсолютные величины и уровни вибрации в жилых помещениях, в зависимости от среднегеометрической частоты октавных полос 16, 31.5 и 63 Гц, а также скорректированные значения, соответствующие требованиям Санитарных норм 2.2.4/2.1.8.566-9 и положениям руководства [1].

За критерий вибрации принимается абсолютная величина виброускорения.

Моделирование распространения колебаний по конструкции здания

Из уравнения для поперечных колебаний основания защемленной балки (здания) [2]

$$EJu^{IV} + m\ddot{u} = -K_y u + K_c \dot{u} \quad (1)$$

член, связанный со сдвиговой упругостью (первое слагаемое в правой части) можно опустить, если колебания балки происходят, как в нашем случае, в воздухе. Второй член имеет смысл вязкого

сопротивления колебаниям и на низких метеорологических частотах также может быть опущен. На частотах 30-100 Гц, характерных для техногенных воздействий, вязкий член сохраняется и уравнение (1) перепишется в виде:

$$EJu^{IV} + m\ddot{u} = K_c \dot{u} \quad (2)$$

Входящие в уравнение (2) величины означают следующее: J - момент инерции поперечного сечения конструкции здания, m - ее погонная масса, K_c - коэффициент затухания колебаний.

Решение уравнения (2) ищем в виде:

$$u(t, x) = f(x)e^{-i\omega t} \quad (3)$$

$f(x)$ - некоторая функция координаты вдоль высоты здания, ω - частота колебаний. Уравнение (2) для $f(x)$ перепишется в виде

$$EJf^{IV}(x) + (K_c i\omega - m\omega^2)f(x) = 0 \quad (4)$$

Результаты анализа модели

Решение полученного уравнения (4) для поперечных колебаний защемленной балки искалось с помощью численных методов математического пакета MATLAB (см. рис. 1).

Моделируемая система характеризуется следующими физическими параметрами. Габариты здания: высота здания 200 м., масса 30000 тонн, длина и ширина основания 70х30 м, акустические и статические свойства бетона рассматривались стандартными из справочников, частота внешнего поля рассматривалась в диапазоне от 1 до 500 Гц. На первом этапе для приведенных акустических и статистических характеристик системы аналитически определялось поведение вибрационного поля в отсутствии затухания (члена $K_c = 0$ в уравнении (4)). Далее рассматривалось поле вибрации с учетом затухания. Рассматривался коэффициент затухания в диапазоне $K_c = (0,01 - 0,05)m\omega$. Также проводилось исследование поведения вибрационного поля для случая совпадения вынуждающей частоты внешнего воздействия с собственной частотой системы (резонанса).

Далее осуществлялось исследование поведения вибрационного поля для случая наличия ветровой нагрузки по поверхности здания, которая определялась добавлением члена, зависящим от высоты в правую часть уравнения (2).

$$EJu^{IV} + m\ddot{u} = K_n\dot{u} + p(z) \quad (5)$$

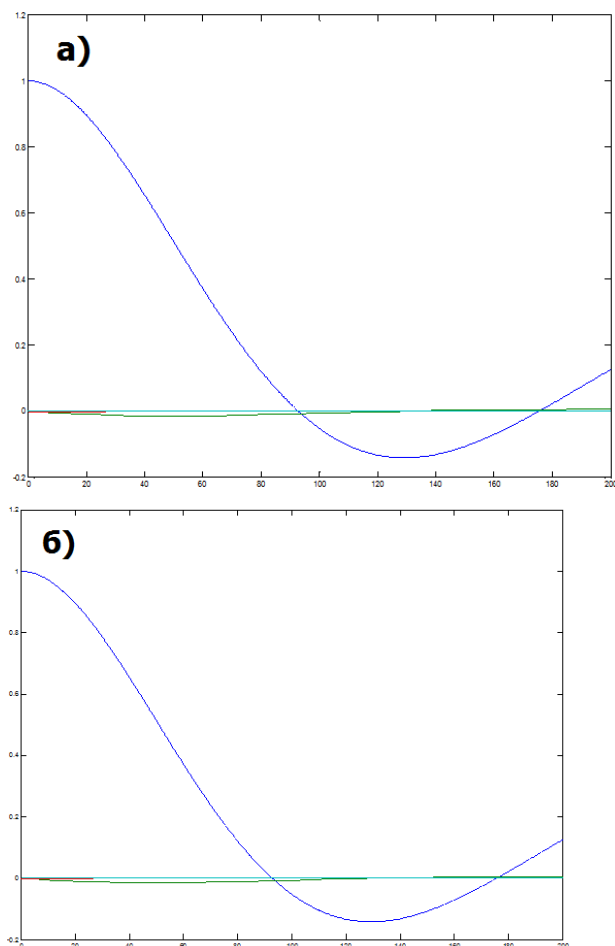


Рис.1 Вибрационное поле вдоль высоты здания (ось ординат) с учётом затухания, частота вынужденных колебаний – а) 2 Гц, и б) 64 Гц.

Выводы

Результаты выполненного моделирования вибрации высотного здания показывают, что при высоте здания большего 100 м и массе большей 20000 т (характерные для реальных сооружений величины) колебания, вызванные техногенными воздействиями в силу наличия затухания в системе быстро спадают и какие либо действия по акустической защите здания нецелесообразны [3]. Если высота здания меньше 100 м, то в этом случае требуется более детальное исследование поля вибрации с использованием более корректных для этого случая моделей по высоте здания с возможным осуществлением мер акустической защиты зданий. Однако сейсмические воздействия с возникновением резонансных явлений в здании могут оказать негативное воздействие даже для

условий Москвы [4]. Результаты моделирования вибрации с учетом ветровой нагрузки показывают, что полем вибрации пренебрегать нельзя и осуществление мер виброакустической защиты зданий является целесообразным.

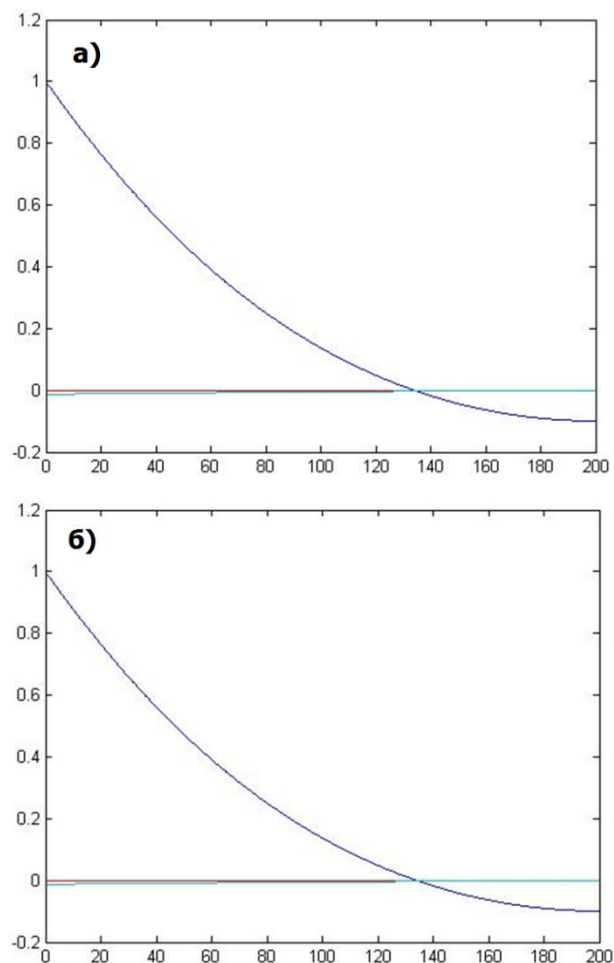


Рис.2 Поведение поля вибрации вдоль высоты здания для ветра движущегося со скоростью 30 м/с для частот а) 2 Гц и б) 64 Гц.

Работа выполнена при финансовой поддержке РФФИ, проекты 08-07-00353, 10-01-00609.

Литература

1. СП 23-105-2004. «Оценка вибрации при проектировании, строительстве и эксплуатации объектов метрополитена».
2. Л.Д.Ландау, Е.М.Лифшиц. Теория упругости. М. Наука, 1987.
3. S.A.Kostarev, S.A.Makhortykh, S.A.Rybak. Calculations of ground vibrations induced by underground sources: analytical and numerical approaches. – In Noise and vibration from high-speed trains. London, Thomas Telford Publishing, 2001. P. 397-422.
4. В.И.Уломов. О сейсмических воздействиях на высотные здания и сооружения г.Москвы. Строительные материалы, оборудование, Технологии XXI века, №2, 2008. С.62 – 65.

Computer system for vibroacoustic impact ecology control on the environment

S.A.Makhortykh

Pushchino IMPB RAS

makh@impb.ru

Abstract

In the paper an approach to calculation of vibration fields from industrial and transport sources in soil is described. This approach is based on physical aspects of acoustic waves propagation in nonhomogeneous media. Both computational and “organizational” (in obtaining exact initial geophysics media properties) are presented. Combined digital-analytic method, based on adaptive algorithms of orthogonal expansions, is proposed.

New combined digital- analytic method for analyzing of physical problems is discussed. It is based on the adaptive expansions of considered functions in adjustable orthogonal basis in appropriate function space.

Some computer realizations of described technology for acoustic and geophysics fields analysis are presented. In particular a software package for analysis of vibration and acoustic situation on the ground surface and estimations of possible measures for vibration reduction.

A problem of vibration calculations in soil

A problem of vibration levels calculation in soil is sufficiently difficult in several aspects. In the first place, soil for the city conditions is a complex geophysics medium with inhomogeneities and complicated dissipation and elastic properties. Respectively, basic models are also complex. From mathematical point of view, even the simplest model of homogeneous medium with a free surface, both shear and bulk elasticity is sufficiently complex for analysis. Computational complications often lead to losses of important wave solutions. In stratified media with layers of different elastic, density and dissipation properties we have multiple wave field with a sufficiently complex structure.

In addition to mathematical complexities we have to take into account difficulties in obtaining of exact data of morphologic and mechanical properties of realistic soil. Usually geology data are obtained in analysis of samples gained in drilling. The number of boreholes is strongly limited and our knowledge of geology data is rather rough.

For this reason the most commonly encountered methods are semiempirical.

In a deficit of reliable geology data we can use simpler models of homogeneous media, which in most cases give an acceptable accuracy and physical clarity of results and don't introduce redundant mathematical complexities.

If necessary geology and acoustic data are available, the accurate variants of Galerkins' method with approximation in orthogonal function class are strongly recommended [1]. One else promising approach is in global solution approximation in required region by finite series of orthogonal functions [2, 3]. Desired accuracy is provided by the series length choice and the minimal complexity is secured by adaptive procedure of optimal basis choice.

Orthogonal expansions into a series of classical orthogonal polynomials, in comparison with other methods of analytical data description (such as description by the least-square method, or by polynomials of best approximation, or by spline functions, atomic functions, etc.) have a number of far-reaching advantages:

- any functional dependence, experimentally obtained as a continuous curve or a digital data file, may be analytically described by a segment of an orthogonal series with an arbitrary preset accuracy in the root-mean-square sense;
- expansion into orthogonal series is rigid, that is, calculation of new expansion coefficients does not effect the previous calculations;
- classical orthogonal bases of continuous and discrete arguments have parameters, whose variation generates orthogonal bases with new properties [1, 2].

The consistent algorithmic and computational realization of the above advantages is a foundation for the development of a new technology, allowing to solve different problems of data processing and modeling from the unified position.

Software package VIBRACALC 1.2

The VIBRAtion CALCulations Computing technique permits in the simplest case of homogeneous media to obtain distribution of vibration field in soil in

the form of cylinder functions with proper boundary conditions on ground surface and casing. The latter is considered as elastic shell, in which both longitudinal and azimuthal modes are excited. Along with the bulk waves, surface waves are built into the wave field treatment.

To perfect the exactness of result for inhomogeneous case the method of finite elements is used. It is applicable also for calculations of transmission features of complex casing.

Some methods of suppression of elevated level vibration are integrated in the program.

- 1) concrete screen (or cavity) in soil on the path of acoustic waves;
- 2) isolation of casing by the soil with large acoustic decrement;
- 3) damping of casing oscillations by elastic (rubber etc) cushion under rail or between casing and soil.

Software is designed for calculation of noise level in soil at given geological conditions. The finite elements method is used for description of inhomogeneous media and complex structures. A computing of the effects of various defense measures for noise reduction on the ground surface is provided. This opportunity makes possible a choice of the most effective tool for reduction of vibration.

Programs are in the form of integrated package. An execution of required operations is accomplished by the appropriate menu and submenu selection.

Results are presented in graphics and text forms.

Menu structure

Main menu

1. File - operation with file system (printing of files and reports, specifying of work directory, current system files, termination of the task and exit to operation system).
2. Options - connection of some service programs: graphical support and exact assigning of geologic data.
3. Parameters, Experiment, Casing, Defense - input of required parameters, modification and editing of the data files, initialization of calculations (parameter input for soil, casing, basic experimental results, a choice of possible measures for vibration reduction etc).
4. Geology - a choice geology data type: homogeneous, stratified or 3-dimensional inhomogeneous media (the second and the latter type is used for accurate calculation with detailed knowledge of the actual geology).
5. Start - beginning of the calculation modules execution.

Conclusions

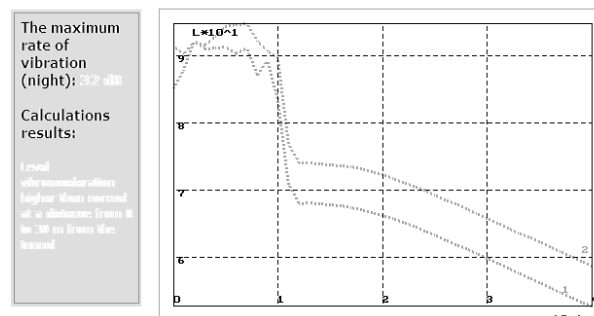
An application of described computational methods, schemes and programs for recent years in several

fundamental and applied fields shows the following complexities. Developed theoretical methods are limited by the lack of experimental data of geophysics type. By virtue of this, the possibility of results of high accuracy obtaining is also bounded. Described approach may be used, in particular, in solution of inverse problems acoustic parameters estimation for the realistic media.

Vibra
Calc 1.2

Fig. Vibration acceleration levels on the ground:

1 - x-projection, 2 - z-projection



Изучение функциональных структур головного мозга методами магнитной энцефалографии

Лыжко Е.В.^(1,2), Махортых С.А.⁽¹⁾

Институт математических проблем биологии РАН⁽¹⁾,
Пущинский государственный университет⁽²⁾
lyzko@yandex.ru

Abstract

In this paper we consider the variability of magnetic encephalogram records (MEG) evoked activity of the brain in the audio stimulation test and spontaneous activity in some cases of Parkinson's disease. The measured signal is the spatial-temporal structure: the vector of measurements in 148 points on the surface of the head with a sampling frequency of 500 Hz. The problem of localization of the biomagnetic field is solved on the basis of the analysis of dynamics of the signal for different parts of MEG records. The method of decomposition MEG data by means of spherical Legendre functions allows for quantitative analysis based on the expansion coefficients.

1 Введение

Магнитная энцефалография (МЭГ) – быстро развивающаяся область экспериментального изучения высшей нервной деятельности человека, функциональных областей мозга и диагностики различных патологий. В данной работе рассматриваются вопросы изменчивости записей магнитных энцефалограмм (МЭГ) вызванной активности мозга при аудиторной стимуляции испытуемого и спонтанной активности при некоторых разновидностях болезни Паркинсона. Измеряемый сигнал представляет собой пространственно-временную структуру: 148-мерный вектор измерений в 148 точках на поверхности головы, развернутый во временной ряд с частотой опроса датчиков 500 Гц. На основе проведенного анализа динамики сигнала решается задача локализации источников биоманитного поля для различных участков записей МЭГ.

2 Методы расчета

Аппроксимацию данных МЭГ предлагается проводить с использованием сферических функций.

В [3] приводятся результаты аппроксимации пространственного распределения биоманитного поля сферическими гармониками. Если некоторая функция $B(\theta, \varphi)$ интегрируема с квадратом на сфере, она может быть разложена в ряд по сферическим гармоникам [2,4]:

$$B(\theta, \varphi) \approx \sum_{n=0}^N \sum_{k=0}^n (a_{nk} p_n^k \cos k\varphi + b_{nk} p_n^k \sin k\varphi)$$

С учетом свойства ортогональности коэффициенты разложения a_{nk} , b_{nk} вычисляются по формулам:

$$a_{n0} \int_0^{\pi} \int_0^{2\pi} (p_n^k)^2 \sin\theta d\theta d\varphi = \int_0^{\pi} \int_0^{2\pi} B(\theta, \varphi) p_n \sin\theta d\theta d\varphi$$

$$a_{nk} \int_0^{\pi} \int_0^{2\pi} (p_n^k)^2 \cos^2 k\varphi \sin\theta d\theta d\varphi = \int_0^{\pi} \int_0^{2\pi} B(\theta, \varphi) p_n^k \cos k\varphi \sin\theta d\theta d\varphi$$

$$b_{nk} \int_0^{\pi} \int_0^{2\pi} (p_n^k)^2 \sin^2 k\varphi \sin\theta d\theta d\varphi = \int_0^{\pi} \int_0^{2\pi} B(\theta, \varphi) p_n^k \sin k\varphi \sin\theta d\theta d\varphi$$

Коэффициенты b_{n0} равны нулю в силу того, что в формулу вычисления входит $\sin(k\varphi)$.

Значения нормировочных множителей вычисляются по формулам:

$$\int_0^{\pi} \int_0^{2\pi} (p_n^k)^2 \sin\theta d\theta d\varphi = \frac{4\pi}{2n+1}$$

$$\int_0^{\pi} \int_0^{2\pi} (p_n^k)^2 \cos^2(k\varphi) \sin\theta d\theta d\varphi = \frac{2\pi(n+k)!}{(2n+1)(n-k)!}$$

$$\int_0^{\pi} \int_0^{2\pi} (p_n^k)^2 \sin^2(k\varphi) \sin\theta d\theta d\varphi = \frac{2\pi(n+k)!}{(2n+1)(n-k)!}$$

где p_n полином Лежандра, p_n^k - присоединенный полином Лежандра

Глубина разложения определяется так, чтобы сумма конечного числа членов ряда гармоник удовлетворяет выражению: $\iint [B(\theta, \varphi) - \sum] ^2 d\theta d\varphi \rightarrow \Delta$

Для вычисления сферических гармоник функции магнитного поля на поверхности головы необходимы значения функции в узлах интегрирования. Исследования исходных записей МЭГ показали, что стандартные формулы табулирования данных оказались не приемлемы, вследствие того, что датчики, расположенные на поверхности головы, не образуют “квадратную решетку”, с узлами расположенными на одинаковом расстоянии друг от друга.

Одним из определяющих свойств реальных данных, используемых в данной работе, является их

неполнота – отсутствие данных на примерно трети поверхности черепной коробки, что связано с конструкцией шлема с датчиками. Неполнота данных существенно усложняет задачу. Стандартные процедуры интерполяции не обеспечивают достаточной точности в этой области [1]. Для решения данной проблемы разработана процедура экстраполяции данных. Построен итерационный алгоритм аппроксимации на сфере $A(i)$, где A – вектор коэффициентов разложения, i – номер итерации, приводящий при росте i к близкому к реальному распределению искомой функции на сфере. Процедура протестирована на реальных данных (рис. 1).

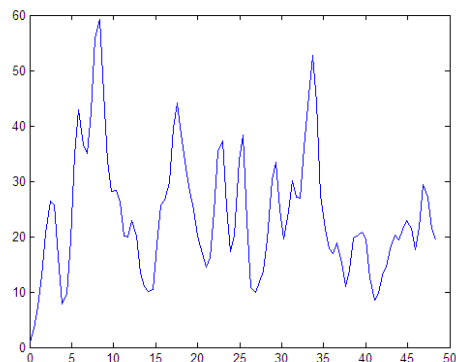


Рис. 1. Итерации экстраполяции

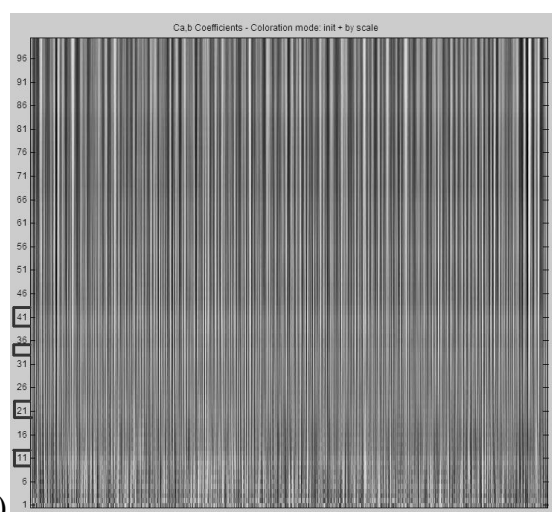
По данным настоящего исследования у здоровых людей в состоянии спокойного бодрствования не обнаружены постоянные источники повышенной магнитной активности. Однако, при функциональной нагрузке (в аудиторных экспериментах) появляются источники повышенной магнитной активности в височной коре, где находятся корковые проекции слуховой сенсорной системы. При фильтрации магнитного сигнала на разных частотах источники локализуются в разных полушариях – в левом на частоте 10 Гц и правом на частоте 20 Гц. Для исследования динамики изменчивости источников МЭГ в различных участках головного мозга для спонтанной и вызванной активности ЦНС предложен класс моделей связанных осцилляторов, отвечающих отдельным нейронам или группам нейронов в структурах мозга с локализованной повышенной активностью. Из эксперимента считаются известными $\omega_1=10$ Гц и $\omega_2=20$ Гц – эти значения используются в качестве начального приближения в модели (рис. 2). Разработанные методы анализа биомагнитного сигнала мозга дают возможность локализовать участки мозга вовлеченные в обработку внешней информации испытуемым, а также выявить у больных паркинсонизмом источники повышенной магнитной активности в областях, связанных с развитием данного заболевания.

Таким образом, комбинирование пространственно-временных спектральных

характеристик МЭГ позволяет повысить точность и детальность описания как временных, так и пространственных распределений биомагнитного поля, что, в частности, повышает точность локализации его источников (см. [5]).



а)



б)

Рис. 2. а) плотность спектра БПФ;

б) вейвлет-спектр Хаара.

Работа поддержана РФФИ, гранты 08-07-00353, 10-01-00609, 09-07-12108-офи_м.

Литература

- [1] А.В. Дергузов, С.А. Махортых. Распознавание патологической активности в записях магнитных энцефалограмм при болезни Паркинсона. *Электронный журнал "Исследовано в России"*, 149, с. 1562-1573, 2005.
- [2] Г.Джеффрис, Б.Свирлс Методы математической физики. Вып. 3, т. 3, М.: Мир, 1970.
- [3] Л.И.Куликова, С.А.Махортых. Математические операции над двумерными сигналами в базисах сферических гармоник. *Электронный журнал "Исследовано в России"*, 60, с. 598-608, 2006.
- [4] Р. Курант, Д. Гильберт Методы математической физики, т. 1, М.-Л., Гостехиздат, 1951.
- [5] С.А.Махортых, С.А.Семечкин. Локализация источников и анализ биомагнитной активности мозга при слуховой стимуляции испытуемого. *Бюллетень экспериментальной биологии и медицины*. 2009 г., 147, № 4. С.477.

Дифференцирование биологических сигналов и проблема обусловленности

Куликова Л.И.⁽¹⁾, Панкратов А.Н.⁽¹⁾, Дедус Ф.Ф.^(2, 1)

Институт математических проблем биологии РАН⁽¹⁾, Московский государственный университет им. М.В. Ломоносова⁽²⁾

kulikova@impb.psn.ru

Abstract

This work is directed to study fundamental properties of generalized spectral-analytical method. Problem solving of conditionality within the limits of the generalized spectral-analytical method at the stage of analytical description of experimental signal and at the stage of further data processing in the space of expansion coefficients is considered. Effectiveness of derivative computing by means of analytical correlations by way of expansion coefficients of initial signal and conditionality number of differentiation operator for orthogonal basis family are shown.

Обобщенный спектрально-аналитический метод (ОСАМ) имеет определенные преимущества в решении задач обработки данных биологических экспериментов. ОСАМ базируется на аппарате модифицированных классических ортогональных базисов [1], что дает возможность устранить наиболее трудные вычислительные моменты и обеспечить повышенную точность. Применение классических ортогональных полиномов позволяет решить проблему обусловленности. Причем обобщенный спектрально-аналитический метод помогает решить проблему обусловленности, как на этапе аналитического описания исследуемого сигнала, так и на этапе дальнейшей обработки сигнала в пространстве коэффициентов разложения.

В ходе решения задачи аналитического разложения сигнала в ряд Фурье необходимо вычислить коэффициенты разложения, которые несут в себе основные свойства исследуемого сигнала и определяются по методу наименьших квадратов. Метод наименьших квадратов является наиболее точным, дающим несмещенные оценки искомых параметров (согласно теореме Гаусса-Маркова) [3]. Для оценки коэффициентов разложения решается система уравнений. И если базисная функция является, например, степенной функцией, то при вычислении коэффициентов разложения приходится иметь дело с плохо обусловленными матрицами. При этом, плохая обусловленность матрицы еще более ухудшается как с увеличением степени полинома, так и с возрастанием объема выборки. Применение

ортогональных базисов значительно упрощает систему нормальных уравнений для оценивания коэффициентов разложения. Матрица нормальных уравнений в силу ортогональности полиномов превращается в диагональную матрицу.

Указанные факторы обусловили глубокое изучение свойств классических ортогональных многочленов непрерывного и дискретного аргумента и их применение для решения задач обработки данных. ОСАМ, основываясь на использовании классических ортогональных полиномов, уже на этапе аналитического описания исследуемого сигнала значительно упрощает вычисление коэффициентов разложения.

Еще одним преимуществом применения обобщенного спектрально-аналитического метода является тот факт, что ОСАМ позволяет проводить полную обработку данных в пространстве коэффициентов разложения без промежуточного восстановления исходного представления данных [1,2]. Это возможно в силу того, что члены ортогональных рядов при аналитическом описании исследуемого сигнала линейно независимы, вычисленные коэффициенты разложения несут в себе все основные характеристики и свойства исходного сигнала. И чем меньше ошибка аппроксимации, тем точнее сохраняется информация об исследуемом сигнале в коэффициентах. При правильном выборе базиса, когда необходимая точность разложения достигается при минимальной длине ряда N , свойства какой угодно длинной последовательности информации заложены в конечном ряде (длиною N) коэффициентов. Разработанные адаптивные процедуры [2] (выбор подходящего ортогонального базиса, согласование промежутков существования исследуемого сигнала и базисных функций, вариация глубины разложения для достижения требуемой точности описания сигнала) позволяют получать аналитические описания сигналов с квазиоптимальной точностью.

Для возможности проведения полной обработки данных в пространстве коэффициентов разложения разработана математическая библиотека выводов [2], включающая в себя основные математические процедуры над спектром сигнала. Такая библиотека имеется для всех ортогональных базисов.

Разработанная алгебра коэффициентов значительно упрощает и облегчает процесс обработки исследуемых массивов данных.

Особое место в проведенных работах по проверке точности и возможности использования выведенных соотношений в условиях наложенного шума различного уровня занимают аналитические соотношения для нахождения производной в пространстве коэффициентов разложения для различных ортогональных базисов. Проведенные исследования по анализу меры обусловленности [3,4] оператора дифференцирования для семейства ортогональных базисов дают возможность оценить и сравнить числа обусловленности операторов дифференцирования для различных ортогональных базисов.

Число обусловленности требуется нам для оценивания погрешности вычисления коэффициентов разложения производных, и полученная оценка позволяет указать примерное число верных цифр вычисленных значений. Число обусловленности играет важную роль: оно измеряет, насколько чувствительно полученное решение к изменениям вектора коэффициентов разложения исходной функции. Относительная погрешность вычисления коэффициентов разложения производной пропорциональна относительной погрешности вычисления коэффициентов разложения исходной функции, а коэффициент пропорциональности равен мере обусловленности оператора дифференцирования для данного базиса.

Приведем пример оператора дифференцирования для модифицированного ортонормированного базиса Лежандра при глубине разложения $n = 5$:

$$\begin{pmatrix} \sqrt{3} & 0 & \sqrt{7} & 0 & \sqrt{11} \\ 0 & \sqrt{15} & 0 & 3\sqrt{3} & 0 \\ 0 & 0 & \sqrt{35} & 0 & \sqrt{55} \\ 0 & 0 & 0 & 3\sqrt{7} & 0 \\ 0 & 0 & 0 & 0 & 3\sqrt{11} \end{pmatrix}.$$

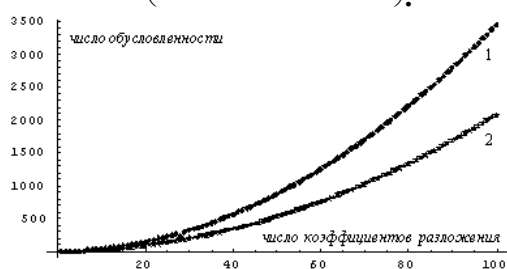


Рис. 1. Мера обусловленности оператора дифференцирования: 1 - для полиномов Чебышева, 2 - для полиномов Лежандра.

На рис. 1 продемонстрированы кривые изменения меры обусловленности для полиномов Лежандра и Чебышева в зависимости от глубины разложения исследуемого сигнала.

Логарифм меры обусловленности тоже дает полезную интерпретацию: он приблизительно равен числу значащих цифр, теряемых в решении.

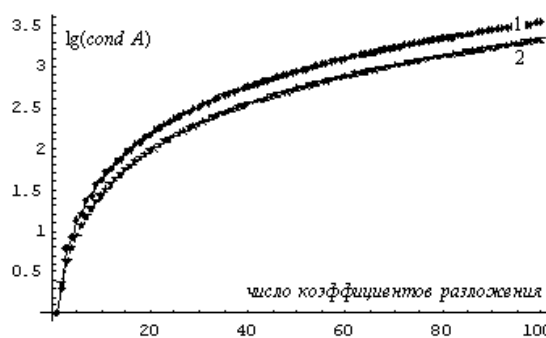


Рис. 2. Логарифм меры обусловленности

1-для полиномов Чебышева, 2 - для полиномов Лежандра.

По представленным рисункам можно сказать, что для рассмотренных базисов наиболее хорошо обусловлен оператор дифференцирования для полиномов Лежандра. Полученные выводы помогают исследователям выбрать наиболее подходящий ортогональный базис при аналитическом описании сигнала в зависимости от конкретных исходных задач обработки данных. По результатам проведенных исследований можно сказать, что оператор дифференцирования для семейств ортогональных полиномов хорошо обусловлен. Проведение адаптивных процедур при решении аналитического описания исследуемой функции, позволяющие достичь описания с требуемой точностью при минимальной глубине аппроксимации, значительно улучшает обусловленность. Результаты проведенных исследований применяются при решении широкого круга биологических задач.

Работа поддержана РФФИ, гранты 08-07-00353, 10-01-00609.

Литература

- [1] Ф.Ф. Дедус, С.А. Махортых, М.Н. Устинин, А.Ф. Дедус Обобщенный спектрально – аналитический метод обработки информационных массивов. Задачи анализа изображений и распознавания образов. // М.: «Машиностроение», 357с, 1999
- [2] Ф.Ф. Дедус, Л.И. Куликова, А.Н. Панкратов, Р.К. Тетуев. Классические ортогональные базисы в задачах аналитического описания и обработки информационных сигналов. // Учебное пособие, Издательский отдел ВМиК МГУ, М., 147с, 2004.
- [3] Каханер Д., Моулдер К., Нэш С., Численные методы и программное обеспечение. // М.: «Мир», 560 с, 1998.
- [4] Форсайт Дж, Малькольм М., Моулдер К., Машинные методы математических вычислений. // М.: Мир, 276 с, 1980.

Решение задачи прогнозирования в классе логических решающих функций

Лбов Г.С.⁽¹⁾, Полякова Г.Л.^(1, 2)

Институт математики СО РАН⁽¹⁾, Новосибирский госуниверситет⁽²⁾
lbov@math.nsc.ru, karesel2007@ya.ru

Forecasting method in class class of logical decision functions.

Gennady S. Lbov, Galina L. Polyakova

Abstract

In this paper we suggest a method using the class of logical decision functions. A search of optimal division of variable space is based on sort out similar to branch and bound method. Results of statistical analysis are represented in the form of logical rules reflected cause-effect relations of object under investigation. Effectiveness of the suggested method is shown by solving applied problems in the sphere of ecology.

Работа выполнена при частичной финансовой поддержке РФФИ (проект № 10-01-00113-а), проект № 09-07-12087-офи_м, Интеграционного гранта СО РАН (№ 83)

1 Задача прогнозирования

Пусть некоторый объект (процесс) описывается набором случайных характеристик $X = \{X_1, \dots, X_n\}$, изменяющихся по времени. Необходимо на основе анализа информации, представляющей измерения этих характеристик в ряде последовательных моментов времени (предыстории), предсказать значения набора характеристик $Y = \{Y_1, \dots, Y_m\}$ в некоторый будущий момент времени (в частности, $Y \subseteq X$). Отличительной особенностью рассматриваемых ниже задач прогнозирования является разнотипность измеряемых характеристик: набор переменных одновременно может содержать бинарные, номинальные и количественные переменные. В этом случае многомерный ряд представляет набор бинарных, символьных и числовых случайных последовательностей. Классические методы ориентированы в основном на анализ числовых последовательностей. Многие методы позволяют анализировать одномерные бинарные или символьные последовательности. Однако большое число важных прикладных задач так или иначе связано с необходимостью анализа разнотипных временных рядов. В немалом числе

задач есть основания полагать, что временные ряды — это реализация случайных процессов, вероятностные свойства которых (распределения) сохраняются по времени. В других случаях, исходя из содержательной задачи, такого предположения сделать нельзя (вероятностные свойства процесса изменяются по времени). В зависимости от указанных предположений можно предложить различные постановки задач и соответственно различные методы их решения.

Рассматривается n -мерный разнотипный случайный процесс $G = \{X_1(t), \dots, X_j(t), \dots, X_n(t)\}$. Пусть набор прогнозируемых характеристик $Y_j = X_j$, $j = 1, \dots, n$. Зафиксируем некоторые последовательные моменты времени $t_1, \dots, t_d, \dots, t_R, t_{R+1}$. Обозначим через x_j^d значение случайной переменной X_j в момент времени t_d , $x_j^d \in D_{X_j}$. Задача состоит в том, чтобы на основе данных, характеризующих предысторию ($b = \{x_j^d\}$, $j = 1, \dots, n$, $d = 1, \dots, R$), предсказать набор значений $y = (y_1, \dots, y_j, \dots, y_n)$ в некоторый будущий момент времени t_{R+1} , где $y_j = x_j^{R+1}$. Необходимо построить решающую функцию, позволяющую предсказывать набор значений $y = (y_1, \dots, y_n)$ по предыстории b .

Множество всевозможных предысторий длины R обозначим через B , а множество всевозможных наборов y — через D_Y , $b \in B$,

$y \in D_Y$, $D_Y = \prod_{j=1}^n D_{Y_j}$. Решающей функцией

прогнозирования - отображение F множества B на множество D_Y , т. е. $F: B \rightarrow D_Y$. При построении решающей функции F используется следующая гипотеза: предполагается, что условное распределение $P(y|b)$ не зависит от сдвига по времени, т. е. указанное распределение для

моментов времени $t_1, \dots, t_R, t_{R+1}$ совпадает с распределением для моментов времени $t_1 \pm \Delta T, \dots, t_R \pm \Delta T, t_{R+1} \pm \Delta T$. Если известно условное распределение $P(y|b)$, то можно найти оптимальную решающую функцию прогнозирования F_0 . Так как указанное распределение неизвестно, то решающая функция строится на основе анализа многомерного временного ряда.

2 Метод решения

Рассмотрим задачу прогнозирования для каждой переменной Y_l , $l = 1, \dots, m$, отдельно, т. е. определим решающую функцию F набором функций $\{\bar{f}_1, \dots, \bar{f}_l, \dots, \bar{f}_m\}$, где $\bar{f}_l$ — отображение из B в D_{Y_l} . Построение функции $\bar{f}_l$ осуществляется в два этапа [1, 2].

Этап I. На основе временного ряда q для фиксированного номера d , $d \in \{1, \dots, R\}$, и фиксированного номера l организуется обучающая выборка в виде таблицы данных $v_d = \{x^{\mu-d}, y_l^\mu\}$, $\mu = R+1, \dots, N$. По таблице данных v_d строим логическую решающую функцию $\langle \alpha_d, r(\alpha_d) \rangle$ с разбиением $\alpha_d = \{E_d^1, \dots, E_d^t, \dots, E_d^{M_d}\}$ множества

$$D = \bigcup_{j=1}^n D_{x_j} \text{ на } M_d \text{ подмножеств (с помощью}$$

алгоритма TEMP, см. ниже). Введем новую переменную Z_d с множеством значений $D_{Z_d} = \{1, \dots, t, \dots, M_d\}$ следующим образом: если предыстория $(x_1^{\mu-d}, \dots, x_n^{\mu-d}) \in E_d^t$, то $z_d = t$. Указанные выше вычисления проводятся для всех номеров предысторий ($d = 1, \dots, R$). Результатом первого этапа является набор разбиений $(\alpha_1, \dots, \alpha_d, \dots, \alpha_R)$ и соответствующий ему набор переменных $(Z_1, \dots, Z_d, \dots, Z_R)$. Это позволяет для любого момента времени t_μ , $\mu = R+1, \dots, N$, n -мерную предысторию длины R , представленную набором $(x^{\mu-R}, \dots, x^{\mu-d}, \dots, x^{\mu-1})$, $x^{\mu-d} \in D$, задать в виде одномерной последовательности $z^\mu = (z_R^\mu, \dots, z_d^\mu, \dots, z_1^\mu)$, $z_d^\mu \in D_{Z_d}$.

Этап II. n -мерному временному ряду $q = \{x_j^\nu\}$, $j = 1, \dots, n$, $\nu = 1, \dots, N$, ставится в

соответствие таблице данных $v = \{z^\mu, y_l^\mu\}$, $\mu = R+1, \dots, N$. По таблице v строится логическая решающая функция $\bar{f}_l$, которая позволяет прогнозировать значение переменной Y_l по предыстории длины R , $l = 1, \dots, n$. Заметим, что задача прогноза для любого набора $\{Y_{i_1}, \dots, Y_{i_m}\} \subseteq \{X_1, \dots, X_n\}$, $m = 1, \dots, n$, решается аналогично.

3 Решение прикладных задач

Метод прогнозирования, например, был использован для предсказания заболеваемости клещевым энцефалитом в зависимости от астрофизических, природных факторов. Был проведен анализ влияния 6 взаимозависимых среднемесячных показателей (температуры воздуха, относительной влажности воздуха и температуры точки росы в приземном слое, а также прямой радиации на горизонтальную поверхность, прямой радиации на перпендикулярную поверхность, отраженной радиации), измеренных на метеостанции вблизи Советского района г. Новосибирска, и среднемесячных, среднегодовых чисел Вольфа на уровень заболеваемости клещевым энцефалитом населения Советского района, г. Новосибирска в течение 18-тилетнего периода. Результаты приведены в работе [3]. В момент написания доклада описанный метод прогнозирования используется при проведении анализа иммунного ответа у иммунизированных или инфицированных вирусом клещевого энцефалита людей в зависимости от климатических и астрофизических факторов, а также при анализе соотношения клинических форм клещевого энцефалита от природных факторов. Полученные результаты будут представлены для обсуждения на конференции.

Литература

- [10] Лбов Г. С., Старцева Н. Г. Логические решающие функции и вопросы статистической устойчивости решений. // Новосибирск: Изд-во Ин-та математики, 211 С., 1999.
- [11] Лбов Г. С. Метод анализа многомерных разнотипных временных рядов в классе логических решающих функций. // Докл. РАН, т. 339, № 6, с. 27-35, 1994.
- [12] Лбов Г. С., Полякова Г. Л., Бахвалова В. Н., Морозова О. В. Исследование влияния природных факторов на заболеваемость клещевым энцефалитом. // Вестник Новосиб. гос. ун-та. Серия: Биология, клин. Медицина, № 3, с. 31-37, 2010.

Модель управляемого гормонами биполярного роста линейных клеточных структур растительного типа

Суховеров В.С.⁽¹⁾, Романов Г.А.⁽²⁾

Институт проблем управления РАН им. В.А. Трапезникова⁽¹⁾,

Институт физиологии растений им. К.А. Тимирязева РАН⁽²⁾

suhoverv@ipu.ru, gar@ippras.ru

Abstract

We propose a model for proliferative growth of a plant-type cell structure, in which cell division is governed by two activating hormones. A software package was developed for interactive model computations which showed similarity with plant-type growth.

1 Введение

Сущность биполярного роста растений состоит в том, что его стебель удлиняется вверх, а корень – вниз, углубляясь в почву. Биполярный рост обусловлен размножением клеток в специальных структурах – меристемах – расположенных в самой верхушке стебля и в кончике корня. Многообразие жизненных форм растений зависит от особенностей их роста, т.е. количества клеток по длине от верхушки побега до кончика корня. Установлено, что для деления растительных клеток необходимо наличие двух гормонов – ауксина и цитокинина [1,3]. Основной синтез этих гормонов происходит на полюсах растения: ауксина – в верхушке стебля, а цитокинина – в кончике корня. За счет транспортных каналов ауксины перемещаются вниз, а цитокинины – им навстречу [3]. Роль гормонального противотока в организации процессов роста до конца не ясна, т.к. выделить ауксино-цитокининовый контур и исследовать его функции для роста растения экспериментально невозможно. Для решения подобных задач эффективным оказывается построение моделей (М), в том числе и математических, и расчеты на них. Целью данной работы было создание М пролиферативного роста клеточной структуры. В результате разработана М, где бигормональный регуляторный контур управляет пролиферативным ростом линейной клеточной структуры растительного типа.

2. Основной раздел

2.1 Исходная М роста клеточной структуры

Предпосылки из области физиологии растений и преобраз настоящей М были опубликованы ранее [2,4], а затем представлены первые результаты

моделирования [5,6]. Предложенная в этих работах М описывает одномерную цепочку идентичных клеток, способных делиться и образующих систему транспорта двух гормонов-активаторов клеточных делений. Деление клетки происходит при условии, что концентрации каждого из гормонов в ней превышают заданные величины – пороги чувствительности: $SL_A (0 \leq SL_A \leq 1)$ и $SL_C (0 \leq SL_C \leq 1)$:

$$\left. \begin{array}{l} H_A(i) \geq SL_A \\ H_C(i) \geq SL_C \end{array} \right\}$$

Фактически каналами транспорта гормонов в М служат сами клетки (Рис. 1). Каждая из m клеток структуры характеризуется концентрациями

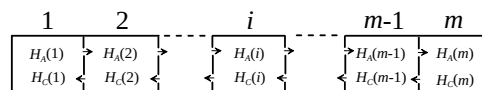


Рис. 1. Исходная М клеточной структуры

гормонов-активаторов: $H_A(i)$, $H_C(i)$, ($i=1, \dots, m$), где i – номер клетки, а m – количество клеток в клеточной структуре. Стрелками указано направление транспорта гормонов. Прототипами гормонов-активаторов М у живых растений являются ауксины и цитокинины. Гормоны в М синтезируются в конечных клетках (1-ая и m -ая клетки).

2.2 Моделирование роста клеточной структуры с учетом динамики транспорта гормонов

Используя исходную М, мы описали динамику процесса роста, т.е. синтеза, транспорта, утилизации гормонов и деления клеток надвое. М определяет условия, когда деление клетки возможно (без детализации процесса деления), и отражает количественные изменения в структуре и концентрациях гормонов. При построении М применялись известные методы теории управления. Для исходной М процессы синтеза, транспорта и инактивации гормонов, происходящие во внутренних и конечных клетках, были описаны с помощью линейных дифференциальных уравнений, в которых независимой переменной является время. Представленная таким образом М рассматривалась, как линейная стационарная динамическая система. В векторной форме уравнение состояния для такой системы:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t),$$

$$\mathbf{x}(t) \in \check{Y}^n; \mathbf{u}(t) \in \check{Y}^p; \dot{\mathbf{x}}(t) = \frac{d\mathbf{x}(t)}{dt};$$

где $\mathbf{x}(t)$ – вектор состояния; $\mathbf{u}(t)$ – вектор управления; $A = \|a_{ij}\|$ ($i=1, \dots, n; j=1, \dots, n$) – матрица системы; $B = \|b_j\|$ ($i=1, \dots, p; j=1, \dots, n$) – матрица управления. Внешнее управление $\mathbf{u}(t)$ на М не рассматривалось: $\mathbf{u}(t) = 0$, потому что управлять пролиферативным ростом растения можно либо за счет изменения наследственности, либо за счет воздействия гормональными препаратами. В продолжение развития исходной М (Рис.1) клеточную структуру можно описать с помощью компартментальной М [9] (Рис. 2), в которой каждая из m клеток состоит из компартментов. Внутренние клетки содержат два компартмента, а концевые – три. Компартменты характеризуются концентрациями гормонов-активаторов: $H_A(i), (i=1, \dots, m+1)$; и $H_C(i), (i=m+2, \dots, 2m+2)$; где i – номер компартмента, а m – количество клеток в клеточной структуре. При этом вектор состояния системы можно рассматривать как сумму двух векторов: $\mathbf{x}(t) = \mathbf{x}^A(t) + \mathbf{x}^C(t)$, где $\mathbf{x}^A(t)$ и $\mathbf{x}^C(t)$ – векторы концентрации гормонов А и С в компартментах: $\mathbf{x}^A(t) = (x_1(t), \dots, x_{m+1}(t), 0, \dots, 0)$,

$\mathbf{x}^C(t) = (0, \dots, 0, x_{m+2}(t), \dots, x_{2m+2}(t))$. Размерность вектора

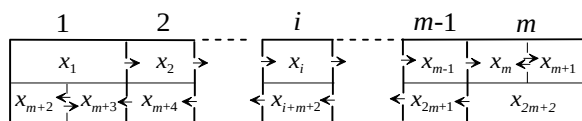


Рис. 2. Компартментальная М клеточной структуры

состояния: $n=2(m+1)$, где n – количество компартментов М, m – количество клеток структуры. Покоординатно М описывается системой линейных дифференциальных уравнений с постоянными коэффициентами. Затем полученная система уравнений была представлена схематически с помощью стандартных графических элементов (динамических звеньев), связанных между собой. Таким образом, для внутренних и концевых клеток М были созданы схемы моделирования, отражающие динамику изменения концентрации гормонов. Затем эти схемы были дополнены блоками механизма деления. Далее из полученных схем была составлена схема моделирования клеточной структуры в целом, где клетки связаны между собой каналами транспорта гормонов А и С. В соответствии со схемой моделирования была составлена система уравнений в конечных разностях. На ее основе получены вычислительные формулы модифицированного метода Эйлера для расчетов на М.

2.3 Программный комплекс для работы с М

Для проведения расчетов на М был разработан программный комплекс. Он содержит два программных модуля:

а) программный модуль собственно математической М роста клеточной структуры с возможностью изменения параметров транспорта гормонов и механизма деления клеток;

б) программный модуль экранного интерфейса, обеспечивающий интерактивную работу при вычислениях на М и функции по сохранению результатов вычислений;

3 Заключение

М адекватно воспроизводит все основные особенности роста целостного растения. Это позволяет рассматривать ее в качестве предельно упрощенной М вегетативного роста многоклеточного наземного растения, управляемого градиентами эндогенных фитогормонов. Уже сейчас на качественном уровне компьютерная М способна делать оценки влияния того или иного параметра транспорта и/или восприятия гормонов (что эквивалентно генной мутации) на фенотип растения. Такие прогнозы важны для селекционеров и агротехников, так как предсказывают вероятные фенотипические эффекты при изменении конкретных характеристик систем гормональной регуляции. В дальнейшем представляется целесообразным оценить возможности применения программного комплекса для прогнозирования темпов роста отдельных видов полезных растений на основе реальных данных.

References

- [1] H. Stals, D. Inze. When Plant Cells Decide to Divide. *Trends Plant Sci.*, 2001, vol. 6, N. 8, p. 359-364.
- [2] Romanov G.A. A Model for Bipolar Plant-type Growth: Role of Auxin-Cytokinin Countercurrent. In *Progress in Plant Growth Regulation*, Kluwer, p. 459-463, 1992.
- [3] В.В. Полевой. *Фитогормоны*. Л.: Изд-во ЛГУ, 1982.
- [4] Г.А. Романов. Модель гормонально-организуемого пролиферативного роста: аналогии с ростом растений. *Онтогенез*, т. 23, №3, с. 228-236, 1992.
- [5] Г.А. Романов, В.С. Суховеров. Исследование кинетики роста и динамики гормональных градиентов на модельных многоклеточных структурах растительного типа – компьютерных растениях. *Доклады академии наук*, т. 352, №6. с. 845-848, 1997.
- [6] В.С. Суховеров, Г.А. Романов. Моделирование гормонально управляемого биполярного роста клеточных структур растительного типа. *Проблемы управления*, №6, 2009.

Применение статистических методов анализа для изучения влияния аллельного полиморфизма генов ключевых ферментов метаболизма гомоцистеина на риск возникновения гипергомоцистеинемии и атеросклеротического поражения сосудов

Кленкова Н.А., Капустин С.И., Шмелева В.М., Хромов-Борисов Н.Н., Блинов М.Н.

ФГУ “Российский научно-исследовательский институт гематологии и трансфузиологии
Федерального медико-биологического агентства” (ФГУ “РосНИИГТ ФМБА России”)
primavista@mail.ru

Abstract

Molecular variations in genes encoding enzymes involved in homocysteine (Hcy) and folate metabolism may account for hyperhomocysteinemia (HHcy) and increased risk of atherosclerotic disorders. The main enzymes that control Hcy concentrations are methylenetetrahydrofolate reductase (MTHFR), methionine synthase (MS), methionine synthase reductase (MTRR), methylenetetrahydrofolate dehydrogenase (MTHFD). We undertook a genotype analysis of MTHFR C677T and A1298C, MS A2756G, MTRR A66G, MTHFD G1958A polymorphisms and measured the level of homocysteine in group of 225 patients with atherosclerosis (peripheral arterial disease or/and ischemic heart disease) and 120 healthy controls. To assess the influence of these polymorphisms on total plasma Hcy and risk of atherosclerosis we applied different statistical analysis. The observed results indicate that polymorphisms in the genes MTHFR, MS might play a role in susceptibility to HHcy and MTRR 66GG/MTHFD 1958GA genotype revealed increased frequency in patients vs control ($p=0,02$; $OR=2,2$; $CI:1,1-4,3$).

Сердечно-сосудистая патология, обусловленная атеросклеротическим поражением артерий, является одной из ведущих причин заболеваемости и смертности населения в развитых странах. Исследования, посвященные изучению патогенеза атеросклероза, убедительно доказали многофакторный генез данного заболевания, определяющийся неблагоприятным сочетанием генетических и экзогенных факторов. Гипергомоцистеинемия (ГГЦ) – повышенное содержание в плазме крови гомоцистеина (ГЦ) – признана одним из независимых факторов риска развития атеросклеротического поражения артерий.

Установлено, что в большинстве случаев повышение уровня гомоцистеина в плазме связано с биохимическими нарушениями в метаболическом цикле данной аминокислоты [1]. Ключевыми ферментами, вовлеченными в метаболизм гомоцистеина и фолатов являются метионин синтаза (MS), редуктаза метионинсинтазы (MTRR), метилентетрагидрофолат редуктаза (MTHFR), метилентетрагидрофолат дегидрогеназа (MTHFD). Аллельный полиморфизм в генах, кодирующих данные ферменты, может быть ассоциирован с нарушениями в функционировании ферментов, и, следовательно, оказывать влияние на концентрацию ГЦ в плазме крови и риск развития атеросклеротического поражения сосудов. В связи с этим, представляется актуальным изучить взаимосвязь между полиморфными аллельными вариантами генов MS, MTRR, MTHFR, MTHFD, концентрацией ГЦ в плазме и риском возникновения атеросклероза.

Исследуемая группа состояла из 225 пациентов с различными проявлениями артериальной патологии - ишемической болезнью сердца, атеросклерозом артерий нижних конечностей, а также сочетанием у пациентов данных заболеваний. Диагноз был верифицирован на основании данных анамнеза, электрокардиографии, велоэргометрии, эхокардиографии и ангиографии. В качестве группы контроля были обследованы 120 жителей Северо-Западного региона России, не имевших в анамнезе признаков артериальной патологии. Типирование полиморфизма генов MTHFR C677T и A1298C, MS A2756G, MTRR A66G и MTHFD G1958A было выполнено методом полимеразной цепной реакции с последующим рестрикционным анализом. Измерение уровня гомоцистеина в плазме проводилось методом жидкостной хроматографии под высоким давлением. В качестве верхней границы “нормы” базального уровня ГЦ было принято значение 13,5 мкмоль/л.

Анализ полученных биологических данных выполнялся с использованием следующих методов

математической обработки: для количественных данных (уровень ГЦ в плазме) – оценка нормальности распределения в изучаемых группах и ранговый дисперсионный анализ Краскела-Уоллиса, описательная статистика – среднее геометрическое; для категориальных данных (частоты встречаемости генотипов) – проверка закона Харди-Вайнберга по всем изучаемым локусам и оценка неравновесности по сцеплению между ними. Достоверность различий между частотами генотипов в изученных группах устанавливалась в соответствии с точным критерием Фишера. Статистическая значимость принималась при значениях $p < 0,05$. Показатель “отношение шансов” (“odds ratio” - (OR)) определялся при двустороннем доверительном интервале (CI), равном 95%. Анализ выполнялся с использованием программ GraphPad Prism4, StatGraphics Plus 5.1, Genepop 4.0.

В результате выполненной статистической обработки показано, что аллельные варианты генов

MTHFR 677TT и MS 2756AA (а так же их сочетание) ассоциированы с повышенной концентрацией ГЦ в плазме, в то время как для полиморфизмов MTRR A66G и MTHFD G1958A статистически значимых различий между генотипами по уровню ГЦ выявлено не было. Имеет место практически полное сцепление между генотипами 677TT и 1298AA гена MTHFR. Наблюдаются статистически-значимые различия по частоте встречаемости между группой пациентов с повышенным содержанием ГЦ и контрольной группой по полиморфизмам генов MTHFR 677TT и 1298AA, MS 2756AA. Вне зависимости от содержания в плазме ГЦ повышена частота встречаемости генотипа MTRR 66GG/MTHFD 1958GA в группе больных по сравнению с контролем ($p=0,02$; $OR=2,2$; $CI:1,1-4,3$). Таким образом, выполненный анализ данных может свидетельствовать о вкладе изученных полиморфизмов в риск развития ГЦ и атеросклеротического поражения сосудов.

Использование корреляционного анализа для установления механизма образования слюнных камней

Бельская Л.В., Голованова О.А.

Омский государственный университет им. Ф.М. Достоевского
LudaB2005@mail.ru

By results of microelement and amino acids composition research of a representative collection of salivary stones of Omsk region inhabitants the correlation analysis is carried out. On the basis of the revealed laws the probable mechanism of formation of salivary stones of the person is offered.

Слюнные камни – патогенные органоминеральные образования челюстно-лицевой сферы, которые формируются из слюны чаще всего в поднижнечелюстной железе и ее каналах. Сложены они, как правило, из ритмично осажженных слоев органической и неорганической компонент вокруг центрального ядра [3]. Причины и механизм образования слюнных камней до конца не изучены. К возможным причинам возникновения патогенных органоминеральных образований в организме человека можно отнести нарушения обмена веществ и условия внешней среды, в том числе и состояние среды проживания пациентов [1, 3, 5]. Значительная роль в образовании органоминеральных агрегатов принадлежит органическому веществу. Поскольку основными веществами, слагающими органическое вещество камня, являются белки и гликопротеины (58-73% от массы всей органики), целью данной работы является выявление особенностей аминокислотного состава слюнных камней и роли микроэлементов в процессе их формирования.

На первом этапе работы проанализирована коллекция слюнных камней (13 образцов) методами высокоэффективной жидкостной хроматографии (Series 200, Perkin Elmer) и рентгеновской флуоресценции с использованием синхротронного излучения (накопитель ВЭПП-3, аналитик Н.В. Максимова). В составе камней определено количественное содержание 20 микроэлементов (табл.1) и 15 аминокислот (рис.1).

Для установления закономерностей распределения аминокислот и микроэлементов в слюнных камнях проведен корреляционный анализ (Statistica 6.0, StatSoft). При числе опытов $n=13$, уровне значимости $f=0.01$, достоверной вероятности $P=0.99$ наличие линейной корреляции считают доказанным при коэффициенте корреляции, превышающим значение критического коэффициента корреляции $r=0.66$ (табл.2).

Таким образом, можно предположить один из вероятных механизмов образования слюнных камней в организме человека. Известно, что слюна представляет собой структурированную биологическую жидкость, весь объем которой

распределен между мицеллами – коллоидными образованиями [4]. Их ядра состоят из молекул фосфата кальция и окружены водно-белковыми оболочками.



Таблица 1

Микроэлементный состав слюнных камней по данным РФА СИ, $\cdot 10^4$ масс.%

Эл-т	Содержание	Эл-т	Содержание
Ti	581.4±118.9	Br	38.6±28.3
V	289.6±30.9	Sr	36.3±17.3
Cr	73.0±13.7	Zr	1.6±0.2
Fe	67.2±24.7	Sn	0.58±0.12
Ni	4.7±2.2	Sb	5.9±1.8
Cu	12.6±8.2	I	58.5±36.5
Zn	29.7±7.5	Ba	6.7±3.4
Ga	2.8±0.4	La	6.3±0.9
As	2.1±0.9	Ce	1.93±0.43
Se	2.8±1.1	Pb	5.8±4.2

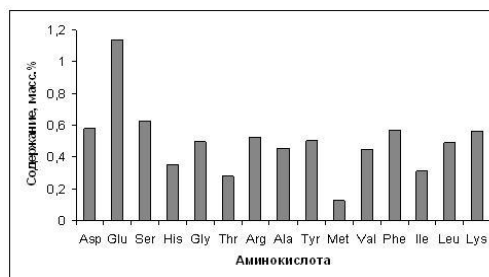


Рис.1. Аминокислотный состав слюнных камней

На первом этапе образуется органическая матрица камня в результате нарушения структуры выводных протоков, изменения состава и свойств слюны, нарушения фосфорно-кальциевого обмена и т.д. [2]. В составе белков слюнных камней преобладают следующие аминокислоты: глутаминовая кислота, глицин, серин и т.д. Накопление в камнях таких микроэлементов, как Ni, Cu, Fe, Zn, может быть результатом их высокого сродства к белкам или отдельным аминокислотам.

Для перечисленных элементов наблюдаются достоверно высокие значения коэффициентов корреляции, в качестве примера можно привести значения ступенчатых констант устойчивости хелатных комплексов никеля с глицином ($K_1=1,4 \cdot 10^6$, $K_2=8,9 \cdot 10^4$). Вероятно, при избыточном поступлении микроэлементов в состав слюны происходит образование хелатных комплексных соединений с аминокислотами, входящими в состав белка, и разрушение защитных оболочек коллоидных мицелл. При повышенной концентрации катионов металлов мицеллы теряют устойчивость и коагулируют, что приводит к нарушению структурных и минерализующих свойств слюны и образованию камней в ротовой полости человека.

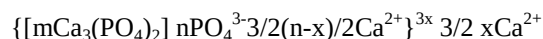
Таблица 2

Корреляционные зависимости содержания аминокислот и элементов в слюнных камнях

Аминокислота	Элемент
Аспарагиновая кислота	Na ($r = 0.80$), K ($r = 0.68$), Mg ($r = 0.87$)
Глицин	Ni ($r = 0.77$), Sr ($r = 0.80$), K ($r = 0.69$)
Треонин	Ni ($r = 0.79$), Sr ($r = 0.77$), K ($r = 0.75$)
Аргинин	Fe ($r = 0.80$), Zn ($r = 0.85$), Mg ($r = 0.76$)
Тирозин	K ($r = 0.82$), Zn ($r = 0.92$)
Метионин	Ni ($r = 0.76$), Na ($r = 0.75$), K ($r = 0.93$), Mg ($r = 0.67$)
Фенилаланин	Ni ($r = 0.86$), Cu ($r = 0.81$), Sr ($r = 0.78$), K ($r = 0.75$)
Изолейцин	Na ($r = 0.77$), Mg ($r = 0.89$)
Лейцин	Mg ($r = 0.77$)

Поскольку основным белком слюны является гликопротеид муцин, имеющий близкое родство к ионам кальция, далее возможно формирования камня за счет образования минеральной фазы на основе затравки (например, органической). Данные выводы основаны на результатах исследования состава и строения слюнных камней, согласно которым, во всех образцах выявлено наличие органического белкового вещества, содержание которого выше в центре камня. Литературными данными подтверждается возможность формирования гидроксилапатита на определенных видах белков. Минеральные слои слюнных камней, представленные в основном гидроксилапатитом, в процессе формирования и роста могут изоморфно включать в свой состав ряд микроэлементов,

например, стронций, что подтверждается близостью ионных радиусов ($r(\text{Ca}^{2+}) = 0,106$ нм, $r(\text{Sr}^{2+}) = 0,127$ нм) данных элементов. Помимо гидроксилапатита в составе камней могут встречаться брушит и витлокит, которые согласно термодинамическим расчетам являются метастабильными фазами, что, по-видимому, связано со значительными колебаниями условий среды кристаллизации слюнных камней. Как известно, состав ротовой жидкости может существенно варьировать даже в течение суток. При образовании камней происходит интенсивное выпадение кристаллов фосфата (брушит) или каплевидных шариков геля (апатит, витлокит) из пересыщенного раствора. Одновременно выделяется белок. Образовавшиеся выделения фосфатов могут агрегироваться и обволакиваться вязким белком. Таким образом, образуется слой за слоем, при этом структура камней часто зернистая или скрытокристаллическая. Такие камни имеют четко выраженную слоистую текстуру, причем разномасштабную на макро и микро уровне. При этом в слюне происходит уменьшение содержания белка за счет перехода его в состав комплексов с минеральной компонентой камня, что также способствует нарушению устойчивой коллоидной системы. На первоначально образовавшемся налете начинают активно развиваться микроорганизмы, за счет чего происходит увеличение pH среды. Подщелачивание слюны, способствуя повышению содержания ионов PO_4^{3-} приводит к изменению состава мицелл и их разрушению:



Расстройство процесса мицеллообразования обусловлено тем, что ионы Ca^{2+} и HPO_4^{2-} не могут одновременно находиться в адсорбционном слое, так как образуют трудно-растворимое соединение $\text{Ca}_3(\text{PO}_4)_2$. При этом с участием микроорганизмов, захватывающих частицы выпавшего в осадок $\text{Ca}_3(\text{PO}_4)_2$ и переносящие их на шероховатую поверхность зубной эмали, создаются благоприятные условия для образования камня.

Список литературы

- [1] А.П.Авцин, А.А. Жаворонков, М.А. Риш, Л.С. Строчкова. Микроэлементозы человека. М.: Медицина, 1991. 496 с.
- [2] В.В. Афанасьев, А.Ф. Ткаленко, М.Р. Абдусаламов. Состав ротовой жидкости смешанной слюны в зависимости от результатов лечения пациентов со слюннокаменной болезнью с помощью метода сиаолитотрипсии. Стоматология, №5, с. 36-38, 2003.
- [3] А.А. Кораго. Введение в биоминералогия. СПб: Недра. 1992. 280 с.
- [4] В.К. Леонтьев, М.В. Галиулина. О мицеллярном состоянии слюны. Стоматология, № 5, с. 17-20, 1991.
- [5] Н.А. Пальчик, Т.Н. Мороз, И.В. Леонова и др. Минеральный и микроэлементный состав слюнных камней. Журнал неорганической химии, №8, с.1353-1361, 2004.

Применение математических методов для обработки результатов анализа ротовой жидкости

Бельская Л.В., Голованова О.А.

Омский государственный университет им. Ф.М. Достоевского

LudaB2005@mail.ru

The element composition of saliva was established by the AES-ISP method. By means of mathematical statistics methods it is shown that parities Ca/P and Na/K, and also distribution of micro cells (Zn, Cu, Fe, Mn, Al) have group features that it is possible to use for diagnostics of corresponding diseases.

В последнее время в области экологии и биохимии широко используются химические и физико-химические методы анализа для изучения роли элементов, содержащихся в биологических объектах. В частности, актуальны исследования элементного состава патогенных биоминералов для выяснения возможного участия микроэлементов в формировании последних [1, 2]. Дополнительные сведения о механизмах и факторах образования камней, а также других патологических процессов, протекающих в организме человека, можно получить при изучении состава биологических жидкостей, например, слюны. В первую очередь представляет интерес исследование биогенных элементов: K, Ca, Mn, Fe, Co, Cu, Zn и др., что является перспективным для изучения биологических жидкостей с целью выявления влияния экологических условий проживания на некоторые группы заболеваний и т.д.

Цель данной работы – исследование особенностей элементного состава ротовой жидкости, применение методов математической обработки данных для выявления групповых различий при патологических процессах в ротовой полости человека.

Объектом исследования являются 28 образцов ротовой жидкости (возраст пациентов 18–30 лет), предоставленных городской стоматологической поликлиникой №1 г. Омска. Образцы слюны собирали утром, натощак, до чистки зубов. Элементный состав образцов ротовой жидкости определен методом АЭС-ИСП (Optima 2000 DV, Perkin Elmer) [3]. Чувствительность метода составляет 10^{-8} – 10^{-2} масс.%, погрешность 3–7 отн. %.

Ранее был проведен панорамный рентгено-флуоресцентный анализ с использованием синхротронного излучения (РФА СИ) для определения микроэлементного состава зубных камней пациентов Омского региона [1]. В проанализированных образцах зубных камней обнаружено 14 микроэлементов: Ti, V, Mn, Fe, Ni, Cu, Zn, Br, Rb, Zr, Ag, Sn, I, Ba. Далее были выбраны элементы, содержание которых определяли в

ротовой жидкости: Ca, P, Na, K, Mg (элементы, образующие минеральную основу камня) и Zn, Cu, Fe, Mn, Al (микроэлементы, содержание которых максимально в проанализированных образцах зубных камней).

В ходе исследования были выделены четыре группы: группа I – пациенты, имеющие зубные отложения в полости рта (8 человек, 28.6 %); группа II – пациенты с кариесом зубов (5 человек, 17.9 %); группа III – пациенты с некрозом твердых тканей зубов (9 человек, 32.1 %) и группа IV – контрольная (6 человек, 21.4 %). Взятую в качестве сравнения группу лиц, резистентных (устойчивых) к заболеваниям, составили люди с одинаковым соматическим статусом «практически здоровые», а также без заболеваний пародонта и слизистой оболочки полости рта.

Методом АЭС-ИСП определен элементный состав образцов ротовой жидкости в зависимости от принадлежности к определенной группе. Для обработки массива полученных данных использован метод дискриминантного анализа (Statistica 6.0, StatSoft). Так, по содержанию 9 определяемых элементов происходит разделение на группы, причем близки между собой и частично перекрываются группы пациентов с кариесом зубов и камнеобразованием в полости рта, а также группы с некрозом твердых тканей зуба и контрольная. Наиболее выражены различия между контрольной группой и группой пациентов с зубными отложениями в полости рта. Дополнительное исследование массива исходных данных методом многомерного корреляционного анализа позволило выявить ряд зависимостей по содержанию отдельных микроэлементов. Например, выявлены корреляции в содержании калия и фосфора ($\alpha = 0.05$, $r = 0.61$), а также железа, меди и цинка ($\alpha = 0.05$, $r = 0.69$), что дает основания для выявления закономерностей распределения данных элементов по изучаемым группам. По основному элементному составу каждой исследуемой группы были рассчитаны доверительные интервалы. Как видно из приведенных данных (табл.1), при возникновении различных заболеваний полости рта происходит существенное отклонение содержания ионов в ротовой жидкости от нормы. Так, при образовании зубных камней уменьшается содержание ионов кальция, но увеличивается содержание фосфора и электролитных компонентов – ионов натрия и калия. При наличии кариеса наблюдаются такие же

изменения, однако, содержание натрия остается в норме. При некрозе твердых тканей зубов существенных отклонений от нормы не выявлено, но при этом наблюдаются отличия в содержании микроэлементов Zn, Cu, Fe, Mn, Al. Учитывая значительную роль элементов Ca, P, Na, K в протекании физиологических процессов в полости рта, было подробно рассмотрено распределение этих элементов по исследуемым группам.

Таблица 1. Содержание элементов и значения коэффициентов Ca/P, Na/K для изучаемых групп заболеваний

Элемент	Группа I	Группа II
Кальций, мг/л	24±9,2	25±13
Фосфор, мг/л	$(2,2±0,83) \cdot 10^2$	$(2,0±0,66) \cdot 10^2$
Натрий, мг/л	$(2,7±0,71) \cdot 10^2$	$(1,4±0,22) \cdot 10^2$
Калий, мг/л	$(8,5±1,9) \cdot 10^2$	$(9,4±2,4) \cdot 10^2$
Ca/P	0,11±0,02	0,13±0,07
Na/K	0,33±0,10	0,16±0,03
Элемент	Группа III	Группа IV
Кальций, мг/л	55±15	56±8,5
Фосфор, мг/л	$(1,3±0,30) \cdot 10^2$	$(1,5±0,38) \cdot 10^2$
Натрий, мг/л	$(2,1±0,63) \cdot 10^2$	$(1,6±0,32) \cdot 10^2$
Калий, мг/л	$(6,5±1,2) \cdot 10^2$	$(6,3±1,9) \cdot 10^2$
Ca/P	0,46±0,09	0,39±0,08
Na/K	0,33±0,09	0,24±0,09

Так, в группе пациентов с зубными отложениями концентрация ионов натрия максимальна ($\alpha < 0.05$), при этом в группах пациентов с кариесом зубов и камнеобразованием в полости рта содержание ионов калия значительно выше нормы. Содержание ионов кальция в группах пациентов с зубными камнями и кариесом зубов достоверно ниже нормального ($\alpha < 0.05$), а фосфора – незначительно завышено. Поскольку в изучаемых группах концентрации кальция, фосфора, натрия и калия изменялись в противоположных направлениях, повысить значимость различий стало возможно расчетом отношений изучаемых показателей: Ca/P и Na/K. Полученные результаты после статистической обработки были выражены в виде доверительных интервалов (табл.1).

Таким образом, используя соотношение Ca/P, можно выделить группы пациентов с кариесом и камнеобразованием в полости рта по значительному отклонению параметров от нормы. Соотношение Na/K можно использовать для разделения между собой групп пациентов с кариесом и

камнеобразованием: при значении коэффициента меньше 0,2 у пациента существенный риск возникновения кариеса зубов; при значении больше 0,2 – риск образования зубных камней.

Применение метода кластерного анализа (Statistica 6.0, StatSoft) позволяет изучить распределение микроэлементов Zn, Cu, Fe, Mn, Al по исследуемым группам пациентов (рис.2). На приведенной диаграмме видно разделение на три группы: в первую группу попадают преимущественно пациенты с нормальным состоянием полости рта (№ 1–5), во вторую – с камнеобразованием в полости рта (№ 11–15), в третью – с некрозом твердых тканей зубов (№ 16–23). Пациенты с кариесом зубов (№ 6–10) равномерно распределены по трем выделенным группам. Подобное разделение исходного массива данных на кластеры позволяет показать различие между рассматриваемыми группами пациентов и предположить существенную роль микроэлементов при возникновении или течении определенного заболевания. Так, в условиях камнеобразования в полости рта повышается содержание цинка и железа, тогда как некроз твердых тканей зуба сопровождается заниженным относительно нормы содержанием всех изучаемых микроэлементов.

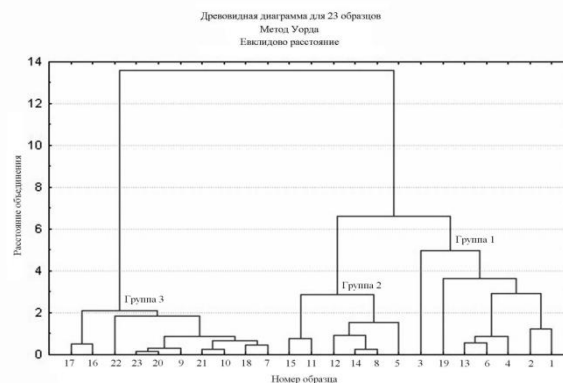


Рис.2. Дерево объединения образцов ротовой жидкости по микроэлементному составу

Таким образом, показано, что использование методов математической обработки данных позволяет выявить особенности распределения элементов по четырем группам пациентов, на основании чего предложена методика диагностики ряда заболеваний ротовой полости человека по элементному составу слюны.

Список литературы

- [1] О.А. Голованова, Л.В. Бельская, Н.Ю. Березина. Спектральный количественный анализ патогенных биоминералов у жителей Омского региона. Журнал прикладной спектроскопии, том 73, № 6, с.792–796, 2006.
- [2] А.А. Кораго. Введение в биоминералогии. СПб: Недра, 1998. 280 с.
- [3] М. Томпсон, Д.Н. Уолш. Руководство по спектрометрическому анализу с индуктивно-связанной плазмой. М.: Недра, 1988. 174 с.

Метод информационного структурного резонанса в задаче анализа данных ДНК-микрочипов

Дюк В.А., Сенкевич Ю.И., Толстоногов Д.А.

Санкт-Петербургский институт информатики и автоматизации РАН

spiiran@iias.spb.su

Abstract

The basic characteristics of the information structural resonance method are resulted. The example of application of this method in a problem of the analysis of DNA-microarray data is considered.

1 Введение

В последние годы исследованию данных биологических, в том числе, ДНК-микрочипов уделяется большое внимание. Экспоненциально растущий интерес в значительной степени мотивирован многочисленными важными практическими приложениями в медицинской диагностике, в разработке лекарственных препаратов и др.

При анализе данных микрочипов исследователи сталкиваются с весьма специфической ситуацией, когда число изучаемых генов на два порядка превышает количество имеющихся образцов. Большинство стандартных алгоритмов классификации плохо справляются с решением задач столь высокой размерности и при малом числе примеров (объектов) почти гарантированно переобучаются. Кроме того, как правило, только малая часть из огромного числа проверяемых генов актуальна в контексте решаемых задач.

В этих специфических условиях продуктивным оказалось применение подходов так называемого интеллектуального анализа данных, направленных, главным образом, на борьбу с переобучением алгоритмов и расширение их обобщающей способности.

В области интеллектуального анализа данных за последнее десятилетие произошли существенные изменения. Слово «интеллектуальный» теперь, скорее, нужно воспринимать в контексте автоматического построения классифицирующих и прогнозирующих моделей. Поиск индивидуально сильных методов и алгоритмов для основной массы специалистов ИАД стал не столь привлекательным – их интересы сместились в сторону умений работать с большими коллективами «слабых» методов и алгоритмов.

Вместе с тем, проблема построения «сильных» моделей, на наш взгляд, не потеряла своей актуальности. По-видимому, сильные модели

являются более пригодными для интерпретации, и объединяются в менее громоздкие коллективы для достижения эффективных результатов в задачах классификации и прогнозирования.

2 Основные характеристики метода информационного структурного резонанса

Нами ранее был предложен метод поиска сложных логических закономерностей в многомерных и разнотипных данных, основанный на представлениях локальной геометрии и средствах интерактивной графики [1, 2]. В этом методе используется модифицированный аппарат линейной алгебры и процедура активного формирования информационного структурного резонанса в многомерных данных. Для краткости этот метод называется методом информационного структурного резонанса (ИСР).

Основные характеристики метода ИСР: Нахождение «сильных» (наиболее полных при заданной точности) if-then правил для каждой записи базы данных; Построение и тестирование классификаторов данных на основе if-then правил; Построение «нечетких» if-then правил; Построение дендрограмм и исследование метаструктуры множества правил.

Дополнительные характеристики метода ИСР: Полиномиальная сложность; Отсутствие ограничений на тип данных; Работа в условиях любого количества пропусков в данных; Работа в условиях «засоренных» данных; Нахождение неперiodических шаблонов сложной формы в числовых и символьных рядах; Возможность распараллеливания процесса поиска if-then правил.

Исходя из приведенных характеристик видно, что особую ценность разработанный метод и его компьютерные реализации представляют для предметных областей, в которых анализируемые данные характеризуются высокой размерностью, разнотипностью описаний объектов и сложными структурными связями. Именно к таким областям относятся науки о человеке в целом и, в частности, молекулярная биология.

3 Пример использования метода ИСР в анализе данных ДНК-микрочипов

3.1 Исходные данные и критерий оценки результатов анализа

Исходные данные взяты из репозитория конкурса RSCTC'2010 Discovery Challenge (<http://tunedit.org/challenge/RSCTC-2010-A>), в котором мы приняли участие. Эти данные содержали от 20000 до 65000 признаков. Участникам конкурса было предложено на основе представленных данных разработать алгоритмы для диагностики 24 классов медицинских пациентов, сгруппированных в 6 отдельных задач.

Характерная особенность – ряд диагностических классов в обучающей (и тестовой) выборке был представлен чрезвычайно малым количеством объектов. Нередко число объектов в классе не превышало 10.

Критерием оценки решения служила сбалансированная точность, которая рассчитывается как усредненная точность классификации для каждого класса в отдельности. Таким образом, каждый класс имел одинаковый вклад в конечный результат, не зависимо от его объема. Это существенно увеличивало цену ошибки на одном объекте для малых классов.

3.2 Процедура анализа

В условиях большого количества признаков (десятки тысяч) и малого объема диагностических классов чрезвычайно важное значение придается методам предварительного отбора признаков с использованием методов кроссвалидации и различных процедур, связанных с повышением обобщающей способности алгоритмов. Для реализации этих процедур необходимо привлечение внушительных вычислительных ресурсов.

Мы не располагали в период конкурса доступом к специальным вычислительным ресурсам и, кроме того, наши аналитические программы не были объединены в единый комплекс – препроцессинг осуществлялся с помощью Матлаба, процедуры бустинга, бэггинга и генерации случайных подпространств производились в Excel, и поиск логических закономерностей в данных по методу информационного структурного резонанса делался отдельной программой.

В силу указанных причин расчет на успешность результатов анализа данных ДНК-микрочипов был основан на высокой эффективности метода информационного структурного резонанса как такового. Вместе с тем, все другие составляющие анализа были сильно загружены.

Так, предварительная селекция признаков производилась не с помощью достаточно изощренных методов, применявшихся другими участниками конкурса, а с помощью наиболее простого критерия Вилкоксона. Кроме того, из-за

ограниченности в ресурсах, для каждого диагностического класса отбиралось всего 30 наиболее информативных по этому критерию признаков. И, наконец, при реализации процедур бустинга, бэггинга и метода случайных подпространств мы ограничивались в среднем 50-ю генерациями случайных выборок.

3.3 Обсуждение результатов анализа

Наивысший результат соревнования – сбалансированная точность по 24 диагностическим классам равна 0,74. Результаты, идущие вслед за наивысшим, расположены очень плотно. Это свидетельствует о том, что в международном научном сообществе в настоящее время большое количество специалистов высокого класса владеет инструментами Data Mining и работает на близком уровне. Наш результат составил 0,67 и пришелся на моду распределения. Мы преодолели все 3 квалификационные «планки», которые установили организаторы соревнования – 0,63; 0,64 и 0,65.

После окончания соревнования организаторы открыли доступ к реальным классифицирующим меткам объектов, что позволило более детально разобраться в ошибках. Анализ показал, что были явные успехи – несколько диагностических классов на тестовой выборке были распознаны нами со 100% или близкой к этой величине точностью. Вместе с тем, как выяснилось, ошибки нашего метода были допущены на 6 малых классах, где явно нам не хватило задействованных вычислительных ресурсов.

4 Выводы

1. Несмотря на весьма упрощенную процедуру препроцессинга и предварительной селекции признаков, метод информационного структурного резонанса продемонстрировал свою эффективность и конкурентоспособность для решения задач медицинской диагностики по данным ДНК-микрочипов.

2. Ресурсы увеличения эффективности метода в задачах анализа ДНК-микрочипов прозрачны: эти задачи необходимо решать с использованием высокопроизводительных вычислительных средств, которые позволят тестировать в динамике различные виды препроцессинга исходных данных, делать предварительную селекцию значительно большего количества признаков и существенно расширить число вариантов генерации выборок в процедурах бустинга, бэггинга и генерации случайных подпространств.

References

- [1] Дюк В.А. Обработка данных на ПК в примерах. СПб, Питер, 1997. – 240 с.
- [2] Дюк В.А. Эмануэль В.Л. Информационные технологии в медико-биологических исследованиях. – СПб, Питер, 2003. – 528 с.

Изучение связи тяжести дисциркуляторной энцефалопатии с полиморфизмом некоторых генетических факторов методами распознавания

Водолагина Н.Н.⁽¹⁾, Костомарова И.В.⁽¹⁾, Кузнецова А.В.⁽²⁾, Малыгина Н.А.⁽¹⁾, Сенько О.В.⁽³⁾

Филиал ГОУ ВПО РГМУ Росздрава "НКЦ геронтологии" ⁽¹⁾, Институт биохимической физики им. Н.М.Эмануэля РАН⁽²⁾, Вычислительный центр им. А.А.Дородницына РАН ⁽³⁾

erri06@rambler.ru, azfor@narod.ru, senkoov@mail.ru

Abstract

Mathematical recognition tools for evaluating of [dyscirculatory encephalopathy](#) severity by representative set of clinical and biochemical parameters. Results of researches are represented that evaluate influence of polymorphous variants of genes coding for angiotensin-converting enzyme (ACE), lipoprotein lipase (LPL) and cholesteryl ester transfer protein (CETP) on disease severity.

1 Введение

Целью данной работы являлось создание компьютерного метода оценки тяжести течения дисциркуляторной энцефалопатии (ДЭП) по представительному набору клинических и биохимических параметров, а также исследование влияния полиморфных вариантов генов ангиотензинпревращающего фермента (АПФ), липопротеинлипазы (ЛПЛ) и белка-переносчика эфиров холестерина (CETP) на течение заболевания [2, 3].

2 Материалы и методы

2.1 База данных.

База данных содержала информацию о 358 пациентах отделения неврологии "НКЦ геронтологии". 273 женщины и 85 мужчин в возрасте от 35 до 102 лет (средний возраст 71,9). Больные были разделены на четыре возрастные группы. 1 группа - 107 человек (35-59 лет), 2 группа - 95 человек (60-75 лет), 3 группа - 33 человека (75-89 лет), 4 группа - 123 человека - старше 90 лет. Четвертая группа достоверно отличалась от трех первых по всем показателям, поэтому было решено проводить исследования связи тяжести течения ДЭП с клинико-биохимическими и генетическими показателями, исключив 4-ю группу. Все пациенты были также разделены по тяжести течения ДЭП: 1

стадия - 47 человек, 2 стадия - 195 человек, 3 стадия - 116 человек.

В базу вошли анамнестические и клинико-лабораторные показатели, результаты инструментальных методов исследования (МРТ, КТ, УЗДГ, ЭКГ, УЗИ), а также молекулярно-генетического типирования генов АПФ, ЛПЛ и CETP, полученные в лаборатории возрастной популяционной иммуногенетики "НКЦ геронтологии"[2]. Исследование липидного метаболизма проводилось в лаборатории липидного обмена "НКЦ геронтологии" Росздрава.

2.2 Методы анализа данных

Для оценки тяжести течения ДЭП использовалась универсальная система методов интеллектуального анализа данных «РАСПОЗНАВАНИЕ»[4], включающая коллекцию методов распознавания и анализа данных, основанных на различных хорошо зарекомендовавших себя подходах.

На первом этапе изучалась эффективность различных методов при диагностике тяжести в группах больных с первой и третьей стадиями ДЭП. При этом диагностика проводилась по клиническим или биохимическим показателям за исключением генетических факторов. Для достоверного оценивания диагностической способности использовался метод скользящего контроля. В результате выделялся набор методов, обладающих достаточной эффективностью. Структура алгоритма распознавания позволяет не только осуществлять диагностику в смысле отнесения пациентов к одной из групп, но производить оценивание тяжести в непрерывной шкале. В качестве оценки тяжести заболевания у пациента была использована соответствующая оценка за класс (группу) с третьей стадией ДЭП, вычисляемая распознающим оператором. Для оценивания тяжести ДЭП наряду с отдельными алгоритмами использовался также комитетный метод, вычисляющий оценку тяжести для произвольного пациента по количеству

отнесений его алгоритмами из эффективного набора к третьей стадии ДЭП. Каждый из распознающих алгоритмов, обученных по группам пациентов с первой и третьей стадиями заболевания, а также комитетный метод, далее были использованы при диагностике тяжести ДЭП у каждого из пациентов, включая пациентов со второй стадией.

На втором этапе исследовалась связь оценок тяжести ДЭП с генетическими факторами. Исследование проводилось с использованием тестов Wilkoxon-Mann-Whitney (WMW), Хи-квадрата также с помощью метода оптимальных достоверных разбиений (ОДР)[1]. Метод ОДР основан на построения разбиений областей допустимых значений объясняющих переменных, индуцирующих разбиение исходной выборки на группы объектов с максимальными различиями уровней прогнозируемой величины. Верификация выявленных закономерностей производится с помощью перестановочного теста. Метод позволяет достоверно оценивать совместное влияние на прогнозируемую величину сочетаний переменных.

Для каждого из трёх исследуемых генов и каждого генотипа производилось разбиение пациентов на две группы. При этом в первую группу входили пациенты, для которых была установлена данная аллель. Во вторую группу входили пациенты, для которых выявлялась альтернативный генотип. Таким образом было получено всего 9 дихотомических разбиений. Статистические тесты использовались для выявления статистических различий между распределениями оценок тяжести для каждой из дихотомий.

3 Результаты исследования

Исследования на выборке из 47 пациентов с первой стадией и 43 пациентов с третьей стадией ДЭП показали достаточно высокую диагностическую способность 114 клинических и биохимических показателей при использовании Линейного дискриминанта Фишера (ЛДФ) - 79%, Линейной машины - 86,6%, Статистически взвешенные синдромы (СВС) - 85,5%, Метода опорных векторов - 84,5%, Нейронной сети (классический многослойный перцептрон), Мультипликативной нейронной сети - 80 %, Метода q-ближайших соседей - 81 %. Диагностическая способность в данном случае понималась как средняя точность распознавания по двум группам. С помощью перечисленных методов, а также с помощью их комитета оценки были рассчитаны оценки тяжести течения ДЭП для 235 пациентов с первой, второй и

третьей стадиями заболевания.

С помощью теста WMW были выявлены значимые различия между компьютерными оценками тяжести ДЭП в группе больных с генотипом DD гена АПФ и группах с генотипами ID и II для целого ряда методов: ЛДФ - $p=0,008$, СВС - $p=0,03$. Для комитета алгоритмов значимость различий была оценена с помощью теста Хи-квадрат на уровне $p=0,003$.

Метод ОДР также выявил значимые различия между группами ($p=0,015$). Метод ОДР был далее применён для изучения совместного влияния генетических факторов на тяжесть течения ДЭП. С этой целью были использованы двухфакторные модели разбиений. Исследования выявили совместное влияние на тяжесть течения ДЭП полиморфных вариантов генов АПФ и СЕТР.

4 Заключение

Исследования показали достаточно высокую эффективность математических методов распознавания при компьютерной диагностике тяжести течения ДЭП. Была выявлена статистически достоверная связь между компьютерной оценкой тяжести течения ДЭП и полиморфными вариантами генов АПФ и СЕТР.

References

- [1] Senko O.V., Kuznetsova A.V. The Optimal Valid Partitioning Procedures, «InterStat», Statistics in Internet, <http://ip.statjournals.net:2002/InterStat/ARTICLES/2006/articles/0604002.pdf>
- [2] Костомарова И.В., Водолагина Н.Н., Малыгина Н.А., и др. Связь полиморфизма генов липопротеинлипазы и белков-переносчика эфиров холестерина с продолжительностью жизни больных с хронической ишемией мозга. Журн. "Клиническая медицина", 2008, №4, с.22-28.
- [3] Кузнецова А.В., Сенько О.В. Возможности использования методов Data Mining при медико-лабораторных исследованиях для выявления закономерностей в массивах данных. Журн. "Врач и информационные технологии", 2005, № 1, 49-56
- [4] Журавлёв Ю.И. Рязанов В.В. Сенько О.В. РАСПОЗНАВАНИЕ. Математические методы. Программная система. Применения. -Москва: Фазис, 2006.

Моделирование противоракового иммунного ответа с учетом влияния радио- и иммунотерапии

Исаева О.Г., Осипов В.А.

Лаборатория теоретической физики им. Н.Н. Боголюбова Объединенного института ядерных исследований 141980 г. Дубна Московской обл.

issaeva@theor.jinr.ru

Abstract

The influence of radiotherapy on the tumor-immune dynamics is studied. Simulation shows that long-term effectiveness of radiotherapy depends on the immune system condition. It is found that the use of anti-cancer vaccine is able to improve result of radiotherapy provided that vaccination is assigned during certain period of time after irradiation. Different strategies of radiotherapy have been considered. It is shown that the hyper rectification is more effective for treatment in comparison with other cures.

1 Введение

В настоящее время широко обсуждаются стратегии сочетания курсов радиотерапии и иммунотерапии. Некоторые доклинические исследования свидетельствуют об усилении противораковых эффектов в случае такого способа воздействия (см. [2] и указанные там ссылки). Поэтому при разработке оптимальных курсов радиотерапии и комбинации радио- и иммунотерапии необходимо учитывать их влияние на динамику противоракового иммунного ответа.

В данной работе эффекты радиотерапии отдельно и в сочетании с противораковой вакциной исследуются на основе недавно развитой нами модели иммунного ответа на опухолевые клетки [4]. Облучение ионизирующей радиацией приводит к значительным повреждениям клеток опухолевой ткани, в том числе клеток кровеносных сосудов. Поэтому в исходную модель дополнительно включается динамика эндотелиальных клеток, чтобы учесть задержку восстановления популяции опухолевых клеток вследствие недостатка кислорода питательных веществ.

2 Модель

Сформулируем систему обыкновенных дифференциальных уравнений, описывающих динамику популяций опухолевых клеток – $T(t)$, цитотоксических лимфоцитов (ЦТЛ) – $L(t)$, интерлейкина-2 – $I_2(t)$, эндотелиальных клеток – $E(t)$, клеток здоровой ткани – $N(t)$, ангиогенного фактора (АФ), например фактора роста эндотелия сосудов (ФРЭС) – $S(t)$:

$$\frac{dT}{dt} = -aT \ln \frac{bT}{aE} - cTL - r_1 U_1 T - r_2 U_1^2 T, \quad (1)$$

$$\frac{dL}{dt} = \frac{dE}{E + E_0} + eLI_2 - fL - r_3 U_1 L - r_4 U_1^2 L, \quad (2)$$

$$\frac{dI_2}{dt} = \frac{gT}{(T + l)(\alpha S + 1)} - jLI_2 - kTI_2 - m_1 I_2, \quad (3)$$

$$\frac{dE}{dt} = -f_1 E + qSE - r_5 U_2 E - r_6 U_2^2 E, \quad (4)$$

$$\frac{dN}{dt} = \frac{g_1 E N_0}{E + E_0} - f_2 N - k_1 TN - r_1 U_1 N - r_2 U_1^2 N, \quad (5)$$

$$\frac{dS}{dt} = \frac{p_4 T^2}{T^2 + \tau_s^2} + \frac{g_2 E_0 N_0}{N + N_0} - j_1 SE - m_2 S, \quad (6)$$

$$\frac{dU_1}{dt} = \delta_1 R(t) - \lambda_1 U_1 - \kappa_1 U_1^2, \quad (7)$$

$$\frac{dU_2}{dt} = \delta_2 R(t) - \lambda_2 U_2 - \kappa_2 U_2^2, \quad (8)$$

где переменными $U_1(t)$ и $U_2(t)$ обозначается среднее число радиационных повреждений ДНК в клетках с различной радиочувствительностью.

Первое слагаемое уравнения (1), заимствовано из [3], где считается, что «емкость» популяции раковых клеток aE/b пропорциональна числу эндотелиальных клеток. Далее предполагается, что скорость притока цитотоксических лимфоцитов (ЦТЛ) в опухолевую ткань зависит от состояния капиллярной сети (первое слагаемое в уравнении (2)). Аналогично [1], в уравнении (3) в первом слагаемом учитывается подавление активации иммунного ответа с увеличением концентрации АФ. Выражение для скорости продукции АФ опухолевыми клетками заимствовано из [1] (первое слагаемое в уравнении (6)). Таким образом, в системе учитывается стимуляция роста сосудов опухолевыми клетками, ангиогенез [1]. Помимо этого, рассматривается продукция АФ нормальными клетками.

Влияние ионизирующей радиации рассматривается в рамках известной линейно-квадратичной модели. Для этого в уравнениях (7) и (8) пренебрегается квадратичным слагаемым, аналогично [6]. Данное приближение допустимо для случая доз ионизирующих излучений, назначаемых при радиотерапии. Уравнения (7) и (8) описывают динамику радиационных повреждений в клетках соответственно двух типов: рано реагирующих и поздно реагирующих. На основе рассуждений,

приведенных в [6], выводятся выражения, описывающие скорость гибели клеток от радиационного поражения (последние два слагаемых в уравнениях (1), (2), (4) и (5)).

Некоторые параметры модели оцениваются на основе экспериментальных данных. Чтобы определить роль параметров системы, был проведен анализ чувствительности [5]. Параметры скорости гибели клеток от радиационного поражения выражены через параметры α и β линейно-квадратичной модели и известные времена репарации повреждений ДНК (подробнее см. в [5]).

Анализ стационарных состояний системы уравнений (1)–(6) без терапии ($U_1 = U_2 = 0$) показывает, что эффективность иммунного ответа зависит от уровня развития индуцированного опухолью ангиогенеза. Обнаружено, что с увеличением параметра p_4 , характеризующего скорость продукции АФ опухолевыми клетками, иммунный ответ становится неэффективным для торможения роста опухоли.

4 Численные эксперименты. Результаты

Расчет системы (1) – (8) в условиях радиотерапии и вакцинальной терапии через 35, 50 и 65 сут. после окончания облучения выполнен для двух наборов параметров P1 и P3, соответствующих различному состоянию иммунного ответа (рис. 2).

Построены дозовые зависимости вероятности излечения и биологической эффективной дозы (БЭД) для трех курсов фракционирования дозы: с

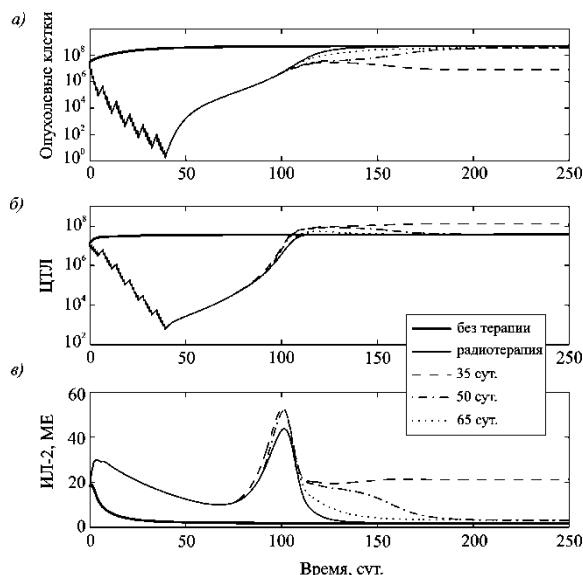


Рис. 2. Эффекты радиотерапии (30 сеансов облучения в дозе 2 Гр ежедневно по 5 мин., исключая выходные дни) и последующей вакцинальной терапии для набора параметров P1 [5]. а) опухолевые клетки б) цитотоксические Т лимфоциты и в) интерлейкин-2.

интервалом между облучениями 12 ч (гиперфракционирование), ежедневное облучение и редкие облучения с интервалом 3 дня.

5 Заключение

Сформулированная система без учета радиотерапии ($U_1(t) = U_2(t) = 0$) может использоваться для описания развития опухоли на ранних этапах ангиогенеза, когда популяция опухолевых клеток может считаться, в достаточной степени, гомогенной. В противном случае необходимо учитывать зависимость переменных и параметров модели от координат вследствие неоднородного распределения концентрации кислорода.

Численные расчеты показывают, что в случае сильного иммунного ответа радиотерапия способствует достижению режима ремиссии опухоли, тогда как в случае слабого иммунного ответа для торможения роста опухоли требуется назначение вспомогательной вакцинальной терапии не позднее определенного срока после окончания курса радиотерапии. В условиях расчета этот период составляет 35 дней.

Сравнительный анализ дозовых зависимостей позволяет сделать вывод, что режим частого облучения наиболее эффективен для лечения, если в качестве изозффекта рассматривать БЭД. С другой стороны этот режим является более щадящим, если изозффект – вероятность излечения.

References

- [1] Arciero J.C., Kirschner D.E. and Jackson T.L. A mathematical model of tumor-immune evasion and siRNA treatment. *Disc. Cont. Dyn. Syst.-B*, 4, P. 39–58, 2004
- [2] Demaria S., Bhardwaj N., McBride W.H., Formenti S.C. Combining radiotherapy and immunotherapy: a revived partnership. *Int. J. Radiat. Oncol. Biol. Phys.*, V. 63, № 3, P. 655-666, 2005
- [3] Hahnfeldt P., Panigraphy D., Folkman J., Hlatky L. Tumor development under angiogenic signaling: a dynamical theory of tumor growth, treatment response, and postvascular dormancy. *Cancer Res.*, 59, P. 4770–4775, 1999.
- [4] O.G. Isaeva, V.A. Osipov. Modelling of anti-tumour immune response: Immunocorrective effect of weak centimetre electromagnetic waves. *Computational and mathematical methods in medicine*, V.10, №3. P.185-201. 2009
- [5] Исаева О.Г. и В.А. Осипов. Влияние радиотерапии на противораковый иммунный ответ. Математическое моделирование. *Препринт ОИЯИ №P19-2009-132*. 2009 (26 с.)
- [6] Sachs R.K., Hahnfeld P., Brenner D.J., The link between low-LET dose-response relations and the underlying kinetics of damage production repair-misrepair. *Int. J. Radiat. Biol.*, V. 72, № 4. P. 351 – 374, 1997.

Усреднение данных магнитной энцефалографии при отсутствии информации об опорных точках

Устинин М.Н., Панкратова Н.М.

ИМПБ РАН
ustinin@impb.psn.ru

Abstract

The information technology of the intellectual data analysis is proposed to detect a useful signal against the background noise of high amplitude without the external information about the signal latencies. To describe the spatial behavior of magnetic field we use the Karhunen-Loeve expansion, and time series are analyzed with bandpass filtering and correlation methods. The technology is tried out in the control experiment data and makes it possible to extract the pathologic activity against the human brain spontaneous activity.

1 Задача исследования

Для получения информации о работе головного мозга необходимо применение эффективных методов обработки экспериментальных данных и постановка целенаправленных экспериментов.

Одним из современных способов диагностики является магнитная энцефалография (МЭГ). Объем данных здесь зависит от количества высокочувствительных датчиков, которые регистрируют магнитное поле на поверхности головы. В современных магнитоэнцефалографах число датчиков порядка нескольких сотен.

При анализе МЭГ производится выделение полезного сигнала на фоне общей спонтанной активности мозга, а затем решается обратная задача: по магнитному полю определяется расположение электрических источников на магниторезонансной томограмме головного мозга испытуемого и делаются научные или диагностические выводы.

Под полезным сигналом понимается пространственно-временная последовательность магнитных полей $m(k, t)$, регистрируемых в результате изучаемой активности головного мозга, например, отклика на какой-либо внешний стимул. Здесь $m(k, t)$ – магнитное поле, измеренное k -тым датчиком, $k=1, \dots, K$, в момент времени $t=t_1, \dots, t_n$, K – общее число датчиков (каналов регистрации), $(t_n - t_1)$ – длительность полезного сигнала, например, интервал между моментами подачи одинаковых стимулов.

Шумом (в целях данной работы) $\chi(k, t)$ считается вся активность мозга, не связанная с

проявлением изучаемой нами в данном эксперименте. Считается, что сигнал и шум аддитивны, так что МЭГ за время длительности отклика записывается как $m(k, t) + \chi(k, t)$, при этом $m(k, t)$ повторяется без изменений в ответ на каждый отклик, а величина $\chi(k, t)$ – случайна. Многократно повторяя стимул и регистрируя отклик на него, можно очистить сигнал усреднением:

$$m(t_i, k) + \bar{\chi}(t_i, k) = \frac{1}{L} \sum_{l=1}^L m(t_{li}, k) + \frac{1}{L} \sum_{l=1}^L \chi(t_{li}, k),$$

где t_{li} – i -я точка по времени в отклике на l -ый стимул, t_{l1} – l -я опорная точка для усреднения, k – номер канала, $i=1, \dots, n$ – номер отсчета по t в полезном сигнале. Такая процедура позволяет улучшить отношение сигнал/шум в $\sqrt{L}$ раз, хотя и требует многократного повторения стимула ($L \approx 10^3 \div 10^4$).

Для интегрированной обработки экспериментальных данных, выделения полезного сигнала и решения обратной задачи магнитной энцефалографии использовался комплекс программ MRIAN [1]. Этот же комплекс позволяет анализировать и представлять результаты на магниторезонансной томограмме испытуемого.

2 Методика

При анализе данных, полученных с поверхности головы часто возникает ситуация, когда искомым сигналом на порядок слабее спонтанной активности и лежит в той же полосе частот. Как правило, полезный сигнал выделяется усреднением по опорным точкам, которые являются моментами внешнего проявления патологической активности (миограмма, моменты подачи стимула и т.п.) Этот подход доказал свою эффективность во многих экспериментах и широко используется в энцефалографии. Однако, остается актуальной задача выделения полезного сигнала при отсутствии дополнительной информации. Особенно остро эта проблема возникает при обработке данных, снятых у пациентов с какой-либо патологией. Например, при патологии tinnitus (слуховые галлюцинации) мы не располагаем внешними проявлениями болезни.

В

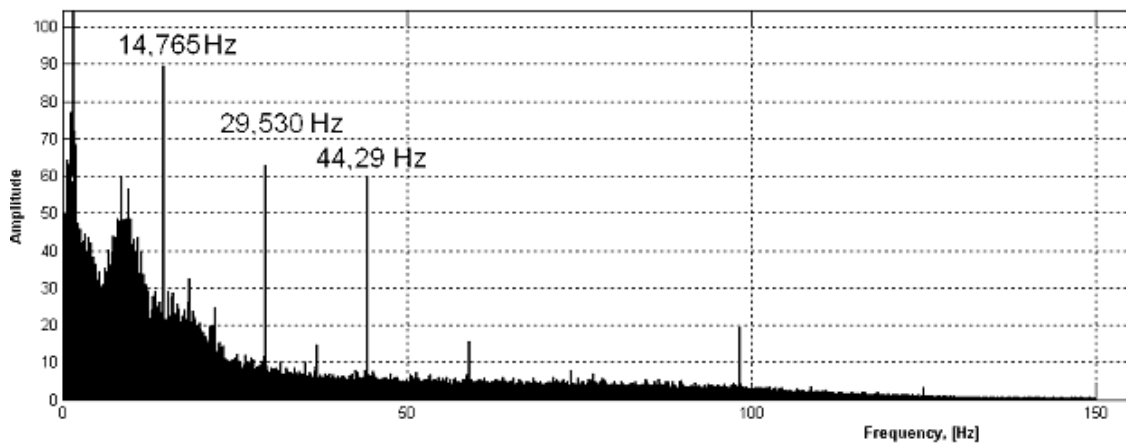


Рис.1. Спектр Фурье проекции экспериментальных данных на 3-ю моду Карунена-Лоева

настоящей работе предлагается взять в качестве опорных точек моменты обнаружения сигнала по его признакам на фоне спонтанной активности, что позволит затем очистить сигнал усреднением.

В качестве тестовой задачи использовались данные контрольного аудиторного эксперимента[1]. Здоровому добровольцу подавался акустический стимул. Частота подачи стимула примерно 7.35Гц. Предварительно, по моментам подачи стимула была получена подробная информация о структуре магнитного поля, возникающего в ответ на аудиторный стимул. Именно это пространственное распределение магнитного поля было выбрано в качестве признака для обнаружения сигнала. Было выполнено разложение экспериментального массива по собственным функциям Карунена-Лоева [1]:

$$M(k, t) = \sum_{p=1}^K C_p(t) f_p(k),$$

где K – число каналов, $t = 1, \dots, T$ – все время эксперимента.

Было найдено, что $f_3(k)$ соответствует по пространственной структуре магнитного поля аудиторному отклику, поэтому именно эта собственная функция разложения Карунена-Лоева в данном эксперименте была выбрана в качестве пространственного признака.

Проводя проецирование всего экспериментального массива, т.е. 148 каналов за все время регистрации, на пространственный признак, получаем функцию:

$$F_3(t) = \sum_{k=1}^{148} M(t, k) f_3(k)$$

Преобразование Фурье этой одномерной функции дает спектр частот (рис.1), на котором видны выделяющиеся пики на частотах равных

14.765Гц, 14.765Гц x 2, 14.765Гц x 3. Из этого спектра также видно, что ожидаемый пик на 7.35Гц отсутствует. Это объясняется тем, что длительность подаваемого стимула равна паузе перед подачей следующего, а мозг реагирует как на подачу стимула, так и на его прекращение.

Далее, была выполнена полосовая фильтрация функции $F_3(t)$. Были рассмотрены три полосы частот, включающие в себя показанные на спектре пики. На полученных после фильтрации функциях были выбраны опорные точки для усреднения экспериментального сигнала в каждой полосе частот. Опорные точки определялись методом отсечки по амплитуде. После усреднения сигнала была решена обратная задача.

3 Выводы

Решение обратной задачи показало, что источники активности (в нашем случае это отклик на стимул) в этих трех случаях лежат в слуховой зоне коры головного мозга и их координаты совпадают с найденными по внешнему стимулу [2].

Работа выполнена при финансовой поддержке РФФИ, проекты №09-07-12108, №10-07-00300, №08-07-00353.

Литература

- [1] Устинин М. Н., Махортых С. А., Молчанов А. М., Ольшеев М. М., Панкратов А. Н., Панкратова Н. М., Сухарев В. И., Сычев В. В. Задачи анализа данных магнитной энцефалографии // Компьютеры и суперкомпьютеры в биологии. Под ред. В. Д. Лахно и М. Н. Устинина. Москва–Ижевск: Институт компьютерных исследований, 2002_С. 327–348.
- [2] Устинин М. Н. Спектрально-аналитические методы обработки данных вычислительного и натурального эксперимента // Дисс. д.ф.-м. н., Пушино, 2004.

Сравнительный анализ экспериментальных данных магнитной энцефалографии

Устинин М.Н.⁽¹⁾, Панкратов А.Н.⁽¹⁾, Рыкунов С.Д.⁽¹⁾, Наурзаков С.П.⁽²⁾, Гребенкин А.П.⁽²⁾

Институт математических проблем биологии РАН⁽¹⁾, РНЦ «Курчатовский институт»⁽²⁾

ustinin@impb.psn.ru

Abstract

The comparative data analysis of the several experiments on magnetic encephalography was carried out. It was found, that the gradiometer, mounted in RSC “Kurchatov Institute”, gives reasonably high brain signal with some interference from outer sources. The method of noisy signal processing was developed in order to determine moments of the stimulus onset.

1 Задача исследования и экспериментальные данные

Задача магнитной энцефалографии (МЭГ) состоит в том, чтобы узнать, как работает мозг, по магнитному полю, измеренному у поверхности головы, то есть восстановить распределение в пространстве и во времени элементарных токовых диполей, породивших измеренные поля. Магнитные энцефалографы строятся с использованием высокочувствительных физических приборов – СКВИДов и позволяют измерять слабое магнитное поле на поверхности головы с высокой точностью. Большое количество каналов (сотни) и высокая частота регистрации (сотни герц) дают возможность получать весьма подробную пространственно-временную картину поля, отражающую электрическую активность мозга с максимальной полнотой. Однако при анализе этой картины следует учитывать, что головной мозг одновременно участвует в обработке многочисленных информационных процессов. Это проявляется в сложной структуре суммарной магнитной активности и приводит к необходимости выделения ее компонент, отвечающих разным процессам. После того, как магнитное поле, соответствующее той или иной активности, выделено, решается обратная задача по восстановлению конфигурации его источников во времени и в трехмерном пространстве, задаваемом координатами датчиков. Полученные при этом результаты приобретают физиологический смысл после их совмещения с анатомической информацией о структуре изучаемого в данном эксперименте мозга. В качестве такой информации, как правило, используется магниторезонансная томограмма.

Целью данной работы был сравнительный анализ данных МЭГ, полученных с помощью различных измерительных приборов.

В качестве опорных данных использовались результаты аудиторного эксперимента, полученные на 148-канальном магнитометре Magnes 2500 WH в больнице Бельвью, в Центре нейромагнетизма Медицинского факультета Нью-Йоркского университета [1, 2]. Датчики представляли собой высокочувствительные измерители магнитного поля на основе СКВИДов (сверхпроводящих квантовых интерферометров), погруженные в дьюар с жидким гелием. Диаметр катушки был равен 2.3 см, среднее расстояние между центрами катушек около 2.9 см. Датчики равномерно покрывали всю поверхность головы и размещались над поверхностью шлема в нижней части дьюара. Аудиторный стимул подавался с частотой 7.3825 Гц на левое ухо здорового испытуемого, велась запись стимула. Профиль стимула представлял собой плато длительностью, равной половине периода.

Проводилось сравнение этих данных с результатами измерений, выполненных на тестовой одноканальной системе, смонтированной в РНЦ «Курчатовский институт». Система представляет собой одноканальный градиометр второго порядка по оси Z с диаметром катушек 0.8 см и базой 3 см. При этом на каждом уровне число витков удвоено, так что итоговые коэффициенты при магнитной индукции, измеренной на разных уровнях, равны {2, -4, 2}. В результате прибор измеряет конечную разность магнитной индукции второго порядка, взятую вдоль оси градиометра. Это позволяет частично избавиться от внешних шумов, а также уменьшить вклад от глубоких источников спонтанной активности головного мозга. Измерения проводились в алюминиевой камере. При этом изучалась как спонтанная активность мозга, так и вызванные магнитные поля.

При экспериментах с тестовой системой ставились следующие основные задачи:

- сравнение одноканальных энцефалограмм с имеющимися многоканальными данными;
- получение данных об уровне шумов;
- создание методов выделения вызванных магнитных полей в экспериментах с внешней стимуляцией.

В перспективе использование систем с небольшим числом каналов может быть оправданным при изучении таких состояний, которые связаны с переходами мозга в квазипериодические режимы работы с высокой пространственной корреляцией между каналами.

Подобные переходы были отмечены, например, при патологиях типа tinnitus (слуховые галлюцинации) [3], или при выполнении некоторых когнитивных заданий [см., напр., 4]. Рассматриваемый здесь одноканальный градиометр является важным шагом на пути создания таких систем.

2 Сравнительный анализ МЭГ

Были рассмотрены спектры Фурье магнитоэнцефалограммы, полученной на многоканальной системе [1, 2]. Для анализа были выбраны два канала – канал 110, расположенный в височной области головы (близко к слуховой коре), и канал 41 в затылочной части (близко к зрительной коре). Оба спектра имеют характерный для работы мозга квазишумовой вид с широкими пиками в области 10 Гц (альфа-ритм) и 20 Гц (бета-ритм). При этом на спектре канала 110 («слухового») хорошо видны несколько гармоник частоты подачи стимула (7.4 - основная, 14.8, 22.1, 29.5, 36.9, 44.3). Особенно выделяются гармоники, начиная с четвертой, так как на частотах выше 25 Гц спонтанная активность мозга невелика. Отметим, что на спектре канала 41 («зрительного») гармоники стимула не наблюдаются.

Все спектры временных рядов, полученных на одноканальном градиометре РНЦ «Курчатовский институт», также имеют вид, характерный для спектров энцефалограмм. В ряде спектров хорошо виден альфа-ритм, иногда заметны низкочастотные пики (дельта-ритм). Следует отметить высокий уровень внешних по отношению к мозгу шумов, выраженных на спектрах в виде узких пиков на гармониках некоторым основным частот (15 Гц, 20 Гц, и т.п.). В дальнейшем ставится задача выяснить происхождение шумов и попытаться снизить их влияние.

3 Анализ вызванных магнитных полей

Поскольку вызванные поля в несколько раз слабее спонтанной активности, для их очистки применяется усреднение по опорным точкам. Этот подход доказал свою эффективность во многих экспериментах и широко используется в энцефалографии (см., напр., [5]). При усреднении полезным сигналом считается последовательность магнитных полей $m(t, k)$, регистрируемых в результате изучаемой активности головного мозга, а аддитивным шумом $\chi(t, k)$ считается вся активность мозга, не связанная с проявлением изучаемой нами в данном эксперименте. После L -кратного повторения стимула и регистрации отклика на него, отношение сигнал/шум увеличивается в $\sqrt{L}$ раз после усреднения:

$$m(t_i, k) + \bar{\chi}(t_i, k) = \frac{1}{L} \sum_{l=1}^L m(t_{li}, k) + \frac{1}{L} \sum_{l=1}^L \chi(t_{li}, k),$$

где t_{li} – i -я точка по времени в отклике на l -ый стимул, t_{l1} – l -я опорная точка для усреднения, k – номер канала, $i = 1, \dots, n$ – номер отсчета по t в полезном сигнале.

В поставленных нами экспериментах велись записи визуальных стимулов, подававшихся при помощи включения/выключения монитора. Оказалось, что эти записи сильно зашумлены, причем как низкочастотными, так и высокочастотными шумами. В то же время, на записях хорошо заметны моменты подачи стимула.

Для автоматического нахождения опорных точек нами был предложен следующий алгоритм. На первом этапе при помощи медианной фильтрации запись стимула очищается от точечных высокочастотных шумов. Было найдено, что при ширине окна, равной 7 точкам, такая фильтрация надежно подавляет узкие выбросы, оставляя неизменным положение передних и задних фронтов стимулов. На втором этапе от полученной функции берется первая конечная разность. Легко показать, что передние фронты стимулов теперь будут отмечены резкими минимумами, а задние фронты – максимумами, и соответствующие им моменты времени могут быть легко выделены программными средствами.

Работа выполнена при частичной финансовой поддержке Американского фонда гражданских исследований и развития (проект CRDF RB1-2027), Российского фонда фундаментальных исследований (проекты 09-07-12108, 10-07-00300, 08-07-00353), и Государственного контракта № 02.512.12.2049.

Список литературы

- [1] Llinas R., Ribary U., Jeanmonod D., Kronberg E., Mitra P. Thalamocortical dysrhythmia: A neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *PNAS*. 1999; 96: 15222.
- [2] Устинин М.Н., Кронберг Е., Филиппов С.В., Сычев В.В., Соболев Е.В., Линас Р. Анализ и динамическая визуализация экспериментальных данных магнитной энцефалографии. *Математическая биология и биоинформатика*. – 2010. – Том 5. №2. – принято в печать.
- [3] Панкратова Н.М., Устинин М.Н., Молчанов А.М., Линас Р. Математическая интерпретация переключений между режимами в сигналах электрической активности головного мозга. *Биофизика*. – 2009. – Т. 54. №5. – С. 916 – 920.
- [4] Kelso J.A.S., Fuchs A. (1995) Self-organizing dynamics of the human brain: Critical instabilities and Sil'nikov chaos. *Chaos*. – 1995. -Vol.5, No.1, P.64-69.
- [5] Михайлова Е.С., Славущая А.В., Кобышев В.А., Пирогов Ю.А., Анисимов Н.В., Шевелев И.А. Локализация диполей волны Р1 зрительного вызванного потенциала мозга человека. *Доклады Академии наук*. 2006. Т. 409. С. 702-706.

Оценка минеральной плотности костной ткани в системе для массовой диагностики остеопороза

Бойко А.И., Ольшевец М.М., Устинин М.Н.

Пушино, Институт математических проблем биологии РАН, Пушино, Россия

ustinin@impb.psn.ru

Abstract

The method of the bone mineral density estimation is proposed using the digital X-ray image of the radius. The normalized image, containing calibration step wedge with known mineral density, is obtained on the digital radiography system. The field of view distortions due to optics and X-ray flow asymmetry are corrected. Then the calibration curve is formed in order to estimate the equivalent thickness of the bone under study. The mineral density of the bone is calculated as a function of the equivalent density using the geometrical model of the bone, build from the digital image. Based on the method proposed, the hard-and-software system is developed for the osteoporosis diagnostics with the digital X-ray systems of general function.

1 Введение

В настоящее время основным методом оценки минеральной плотности костной ткани является двухлучевая абсорбциометрия, выполняемая при помощи специализированного оборудования. В то же время известно [1], что результаты таких обследований хорошо коррелируют с результатами радиографической абсорбциометрии при съемке с эталонными объектами. При цифровой рентгеновской диагностике остеопороза, в поле съемки наряду с диагностируемым объектом включается эталонный объект с нормированной минеральной плотностью. Сравнивая яркости диагностического и эталонного объектов и зная реальную плотность эталона, можно определить плотность ткани. При этом ставится задача создания оборудования для получения стандартизованных рентгеновских снимков, допускающих автоматизированный анализ без участия оператора, а также программного обеспечения для такого анализа.

Для получения цифровых рентгеновских снимков используется компьютерная приставка к медицинскому рентгеновскому аппарату [2-4]. Приставка позволяет получать снимки в двух режимах – обзорном (395x315 мм) и прицельном (155x135мм), при этом в обоих режимах разрешение снимка равно 1280x1024 пикселей с 4096

градациями яркости. Для задач денситометрии снимков предплечья используется прицельный режим. Фиксация объектов съемки в стандартном положении выполняется при помощи экспериментального стенда, включающего набор эталонных объектов с известной минеральной плотностью.

2 Программное обеспечение системы

Важной частью созданного аппаратно-программного комплекса является компьютерная программа получения, просмотра, анализа, обработки, хранения и распечатки цифровых рентгеновских изображений. Созданный программный комплекс объединяет в себе функции управления камерой, графического редактора и системы управления архивом цифровых снимков, сведений о проведенных обследованиях и личных данных пациентов [3, 4].

Программа обеспечивает выполнение следующих задач:

- получение цифрового снимка;
- хранение снимков и организация доступа к ним;
- визуализация снимков и операции цифровой диагностики;
- цифровая обработка снимков и печать на бумаге или пленке;

Относительно высокая зашумленность и низкая контрастность получаемого снимка могут приводить к появлению артефактов при неправильно выбранном сценарии обработки изображения. Поэтому наиболее широкое применение в повседневной рентгенологической практике находят алгоритмы обработки изображения из класса алгоритмов поэлементной обработки, а также некоторые алгоритмы сглаживания. В частности, имеется возможность коррекции изображения по интерактивно задаваемой пользователем передаточной функции. В состав программного комплекса включены шаблоны нескольких известных операторов свертки при различных размерах окна (гауссиан, лапласиан, и др.), пользователь может выбрать наиболее подходящую операцию, последовательно применять фильтры, задать в диалоговом режиме произвольное окно и матрицу весов фильтра и сохранить созданный

шаблон для дальнейшего исследования и анализа снимка. Также реализованы некоторые нелинейные алгоритмы обработки изображений со скользящим окном.

3 Применение ортогональных разложений для коррекции и сжатия снимков

Наряду с шумоподавлением и повышением контрастности изображения важной задачей является сжатие цифровых рентгеновских снимков. Актуальность проблемы обусловлена необходимостью хранения большого числа полноформатных цифровых изображений и их передачи по каналам связи без потерь значимой диагностической информации. Для решения обеих задач использовались разложения по wavelet-базисам (базисам всплесков) [5], представляющим собой специфические системы ортогональных функций, хорошо подходящих для обработки резко меняющихся данных. Существуют эффективные алгоритмы быстрого преобразования исходного сигнала в пространство коэффициентов разложения. Дальнейшая обработка цифровых массивов с использованием wavelet-базисов ведется в пространстве коэффициентов. Методы разложений по базисам всплесков ведут к экономному решению многих задач обработки, требующих реализации в рамках ограниченных аппаратных или вычислительных ресурсов. Для осуществления сжатия в разложении цифрового изображения по выбранному базису сохраняют только коэффициенты с амплитудой, превышающей некоторый порог. При этом частично решается и задача шумоподавления. Фактически, с выбором порога решается задача нелинейной аппроксимации по элементам базиса, вносящим наибольший вклад в разложение. В применении к задачам медицинской бесплёночной рентгенографии удастся достигнуть значительной степени сжатия хранимых цифровых рентгеновских снимков при сохранении приемлемого качества изображения.

4 Моделирование и анализ цифрового рентгеновского снимка

Для оценки требований, предъявляемых к точности измерений, а также для построения диагностической сцены проведено имитационное моделирование формирования рентгеновского снимка. Была построена модель сечения предплечья из конечных элементов объема (вокселей) с заданными коэффициентами экспоненциального поглощения. Для каждого пикселя в строке матрицы яркость формировалась как разность между входной интенсивностью и суммой интенсивностей, поглощенных всеми вокселями по ходу рентгеновского луча. По результатам моделирования был сделан вывод о том, что 12-

битная оцифровка, используемая в данной рентгеновской системе, достаточна для оценки минеральной плотности костной ткани.

Предлагаемая методика денситометрии лучевой кости предплечья основана на том, что предплечье фиксируется в стандартном положении при помощи консоли, на которой размещены эталонные объекты, соответствующие по плотности гидроксиапатиту, и имеющие различную толщину. На снимке выделяются области с заданной плотностью гидроксиапатита, измеряется их яркость (оптическая плотность) и строится калибровочная кривая. Затем по калибровочной кривой определяется эквивалентная толщина измеряемой области кости в расчете на единичную плотность гидроксиапатита.

Дальнейший анализ данных проводится с учетом геометрических характеристик, полученных из анализа снимка, а также при помощи экспериментального стенда, и дает оценку объемного содержания гидроксиапатита в кости.

5 Выводы

Полученные результаты позволяют сделать выводы о перспективности предложенного метода для диагностики остеопороза с помощью цифровых рентгеновских аппаратов общего назначения и о необходимости создания программного обеспечения для автоматизированного анализа снимка.

Работа выполнена при частичной поддержке Программы фундаментальных исследований Президиума Российской академии наук «Фундаментальные науки – медицине» и Российского фонда фундаментальных исследований, проекты 09-07-12108 и 10-07-00300.

Список литературы

- [1] Gulam M., Thornton M., Hodsmen A., Holdsworth D. Bone mineral measurement of phalanges: comparison of radiographic absorptiometry and area dual X-ray absorptiometry // Radiology. – 2000. – Vol. 216. – P.586.
- [2] Арапов Н.А., Корнев А.Н., Кулаков В.И., Никонов И.А., Санталов Б.Ф., Устинин М.Н., Фокин В.А., Яшин В.А. Аппарат для получения компьютерных рентгеновских изображений и способ получения таких изображений. пат. № 2134450 Рос. Федерация. №97112727/09: заявл.24.07.97; опубл.10.08.99, Бюл. №22. 16 с.
- [3] Устинин М.Н., Никонов И.А., Ольшевец М.М., Цифровая диагностика и телемедицина // «Компьютеры и суперкомпьютеры в биологии». Под ред. В.Д.Ляхно и М.Н.Устинина. (Институт компьютерных исследований, М.-Ижевск, 2002), сс.458-475.
- [4] Ol'shevetz M.M., Ustinin M.N., and Nikonov I.A. Software for Digital Filmless Roentgenography // Pattern Recognition and Image Analysis. – 2006. – Vol. 16. – PP. 23–25.
- [5] Daubechies I. Ten Lectures on Wavelets. – Philadelphia: SIAM, 1992.

Оптимизация топологии крыльев летающих насекомых

Шевцова В.С.⁽¹⁾, Соловьев А.Н.^(1, 2), Шевцова М.С.⁽³⁾

Донской государственный технический университет⁽¹⁾, Южный Федеральный университет, кафедра математического моделирования⁽²⁾, Южный научный центр РАН⁽³⁾
Barbaragen4@mail.ru

Topology optimization of a flying insects wings

Shevtsova V.S., Soloviev A.N.,
Shevtsova M.S.

Abstract

A problem of a carrying structure topology optimization is considered on example of a flying insect wing, that optimized during the natural evolution. New approaches to modeling of aeromechanics and mechanics of the insect wings, and to solving of structural topology optimization using the modern finite element software are presented. The statement and some results of structural optimization problem the for wing obtained by Solid Isotropic Material with Penalization-method are given and discussed.

1 Модели крыла летающих насекомых

Последние 10-15 лет пристальное внимание механиков, математиков, специалистов по проблемам управления обращено к различным аспектам полета насекомых. Главной причиной этого стало интенсивное развитие новой отрасли – разработки и проектирования беспилотных летательных объектов (MAVs & UAVs) [2]. Ключ к успешному проектированию MAVs лежит во всестороннем понимании полета насекомых, т.к. именно они обладают набором качеств, необходимых для разрабатываемых искусственных объектов (малые затраты энергии, высокая маневренность, способность к стоячему полету и др.). Во время полета насекомые активно меняют геометрию аэродинамической поверхности крыла. Кроме того, крылья насекомых - продукт многотысячелетней эволюции, создавшей их структуру оптимальной для данных способов и режимов полета. Это делает их идеальной конструкцией для отработки численных методов структурной оптимизации, которые могут иметь широчайший спектр приложений.

Обычно насекомые имеют две пары крыльев, которые могут видоизменяться или редуцироваться. Крылья представляют собой тонкие двухслойные мембранные структуры. Крылья поддерживаются

склеротизированными жилками, представляющими собой каналы в толще крыла, заполненные гемолимфой. Характер жилкования составляет существенную особенность морфологии крыльев. По расположению жилки делят на продольные и поперечные. У насекомых высших семейств и подотрядов число поперечных жилок редуцируется, а продольные жилки смещаются к переднему краю крыла, что способствует укреплению передней кромки и улучшению аэродинамических показателей крыла. Расположение жилок влияет на летные качества насекомого, т. к. от него зависят динамические характеристики конструкции крыла, но до конца значение типа жилкования до сих пор не изучено.

Из-за чрезвычайно сложной кинематики и изменения геометрии крыла в полете число степеней свободы становится бесконечным. Поэтому известные математические модели, имитирующие машущий полет, всегда рассматривают крыло абсолютно жестким. Но даже в этом, достаточно грубом приближении, задача оптимального распределения сети армирующих жилок является чрезвычайно сложной. Некоторые подходы к ее постановке и решению представлены ниже.

2. Методы и алгоритмы оптимизации топологии конструкций

Под оптимизацией топологии понимают создание оптимальной структуры подвергаемой заданным нагрузкам конструкции, обеспечивающей достижение наилучших эксплуатационных показателей (прочности, жесткости, устойчивости) при ограничениях на размеры, форму или вес. Применительно к крылу задача оптимизации состоит в определении такого набора параметров модели крыла, которые, однозначно характеризуя его топологию, при выполнении ограничений на вес, обеспечивают наибольшую жесткость конструкции, требуемые для исключения флаттера взаимные расположения центров тяжести и центров жесткости сечений, максимально высокие собственные частоты.

3. Оптимизация топологии "крыла насекомого"

В качестве объекта оптимизации в работе использовали эллиптическую область, в плоскости x-y близкую по форме крылу стрекозы. Построение модели выполняли в системе Comsol Multiphysics с

подключением и связанной работой модулей 3D Structural Mechanics (уравнения трехмерной теории упругости) и Optimization Lab. "Крыло" моделировали двумя тонкими эллипсами, жестко соединенными плоскостями. Нижний эллипс моделировал мембрану, верхний - представлял собой область проектирования, в которой необходимо было выполнить оптимизацию топологии. К нижнему, сплошному эллипсу прикладывались распределенные нагрузки (модели распределенных аэродинамических сил, которые испытывают элементы крыла в различных фазах махового движения) [3]. Левый срезанный край "крыла" закреплялся консольно. Выполнялся только статический анализ, т.е. для каждого распределения приложенных сил строилась своя картина жилкования, оптимальная для данного случая.

Задача оптимизации податливости верхнего слоя формулировалась как

$$\text{minimize } c(\rho) = \int_{\Omega} \frac{1}{2} \varepsilon^T \mathbf{D}(\rho) \varepsilon d\Omega, \quad (1)$$

где

$$\int_{\Omega} \varepsilon^T \mathbf{D}(\rho) \varepsilon d\Omega = \int_{\Omega} \mathbf{f} u d\Omega + \int_{\Gamma} \mathbf{t} u ds. \quad (2)$$

В формулах (1), (2) Ω - область проектирования, $\varepsilon, \mathbf{u}$ - векторы деформаций и смещений соответственно, $\mathbf{D}$ - матрица жесткости, $\mathbf{f}$ - массовая, и $\mathbf{t}$ - поверхностная силы соответственно. Согласно SIMP - методу матрица жесткости представляется в виде

$$\mathbf{D}(\rho, \mathbf{x}) = \rho(\mathbf{x})^p \mathbf{D}_0 \Big|_{\rho > 1, \mathbf{x} \in \Omega}, \quad (3)$$

где $\mathbf{D}_0$ - матрица жесткости данного изотропного однородного материала (жилок), и p - штрафной параметр. Переменная $\rho(\mathbf{x})$, заданная всюду в области проектирования, которую следует интерпретировать как обобщенную плотность распределения материала, должна удовлетворять следующим ограничениям

$$\begin{aligned} 0 &\leq \int_{\Omega} \rho(\mathbf{x}) d\Omega \leq V \\ 0 &\leq \rho(\mathbf{x}) \leq 1 \end{aligned}, \quad (4)$$

где V - объем материала, доступный для размещения в области проектирования. Значение параметра p , большее единицы, обеспечивает разделение зон с плотностями материала, равными 1 (материал) и равными нулю (нет материала). Согласно [1] SIMP модели материал рассматривается как двухфазный; и для упругого тела параметр p выбирался удовлетворяющим условию

$$p \geq \max \left\{ \frac{2}{1-\nu}, \frac{4}{1+\nu} \right\}. \quad (5)$$

Так, для материала с коэффициентом Пуассона 0.33 этот параметр выбирался равным 3.

Для получения приемлемой разрешающей способности создаваемой картины распределения

жилок использовалось густое конечноэлементное разбиение - 8000 элементов в мембранном слое и 16000 в оптимизируемом слое. В качестве нулевого приближения для оптимизационной задачи принималось решение упругой задачи при равномерном распределении материала жилок. Один из примеров полученной картины приведен на рисунке.



Типичная картина распределения армирующих жилок на поверхности эллиптического крыла

Варьируя распределение поверхностных, действующих на крыло в отдельных фазах маховых движений, можно было получить изменяющуюся картину распределения жилок по мембране крыла. Наложение таких картин давало качественное совпадение с реальной картиной жилкования.

Выводы

В работе показана принципиальная возможность оптимизации топологии структур типа крыльев насекомых с большой разреженностью армирующих элементов. Топология создаваемых "оптимальных" структур существенно зависела от конфигурации приложенных внешних сил и от доли поверхности, "разрешенной" алгоритму для размещения армирующих жилок. В связи с тем, что различным фазам маховых движений крыла соответствуют существенно различающиеся распределения поверхностных сил, оптимизируемая конструкция должна быть оптимальна для всех возможных внешних воздействий. Решение этой задачи требует разработки и обоснования вида целевого функционала, а также постановки задачи оптимизации на базе динамического модального и нестационарного анализа.

Литература

- [1] Bendsøe M.P., Sigmund O. Topology Optimization Theory, Methods and Applications, Springer Ed., 2004, 296 p.
- [2] Dickinson M.H. Bionics: Biological insight into mechanical design / Proceedings of the National Academy of science (PNAS), Vol.96, №25, 1999, pp.14208-14209.
- [3] T. L. Hedrick and T. L. Daniel. Flight control in the hawkmoth *Manduca sexta* / Journal of Experimental Biology №209, 2006. pp.3114-3130.
- [4] Marcelo H. Kobayashi. On a biologically inspired topology optimization method / Commun. Nonlinear Sci. Numer. Simulat. №15, 2010. pp. 787-802.

Информативность фрагментов сигнала с позиций нервной ткани

Цепков Г.В.

Международный научно-учебный центр информационных технологий и систем НАН Украины, Киев. E-mail: gtsepkov@ukr.net

Abstract

Modern methods of analogue signals sampling for their digital processing are not economical. The methods are based on uniform data sampling, which leads to equal importance of the all samples. But actually informative relevance of different fragments of a real signal is not equal, so irrelevant samples should be eliminated from the data array using a rational informativeness criterion.

Введение

Известно, что реакция нервной ткани на внешнее воздействие – стимул $u(t)$ возникает, когда это воздействие превышает порог возбуждения u_0 [1, 2]. Экспериментально доказано [1-3], что на возбудимость W ткани влияют также импульсы действия $\int u(t)dt$ – энергетическая компонента стимула и скорость $u'(t)$ изменения его амплитуды, которую следует считать информационной компонентой, ибо она описывает форму сигнала. Таким образом, согласно [1-3] имеем

$$W = W[u(t), \int u(t)dt, u'(t)], \quad (1)$$

что можно представить дифференциальным уравнением.

Эксперимент

Отметим, что в экспериментах [1-4] объектом исследований служил нервно-мышечный препарат лягушки, стимулируемый одиночными электрическими импульсами в сравнительно небольшом диапазоне длительностей $\{\tau\}$. В описанных ниже экспериментах [5,6] методика была существенно усложнена. Нервно-мышечный препарат лягушки раздражали импульсами форм: прямоугольной, треугольной, синусоидальной и пилообразной (с крутым передним фронтом). Каждый раз при задании величины τ все импульсы имели положительную полярность, одинаковую длительность и равную амплитуду. Длительность τ стимулов задавали изменением частоты

периодических колебаний f генератора в диапазоне (0,1 ... 100) Гц, используя соотношение: $\tau = 1/2f$.

Результаты и обсуждение

Было определено, что W при всех значениях τ прямоугольных и пилообразных импульсов выше, чем для стимулов иной формы, что подтверждает данные известных экспериментов. Теоретически $u'(t)$ крутых фронтов этих сигналов есть δ -функция Дирака и, хотя в действительности $u'(t) < \infty$, эта скорость существенно превышает быстроедействие механизма адаптации ткани.

Особый интерес представляют значения u_0 для синусоидальных и треугольных импульсов. Их анализ дает следующее.

Порог u_0 у этих стимулов выше, чем у предыдущих. Кроме того, зависимость «сила-длительность» $u(t)$ имеет иной характер, что, согласно теории Гилла [4], отражает эффект адаптации. В диапазоне от 10 мс до 1 с на синусоидальные импульсы W выше, чем на треугольные, т.к. у первых больше величины и $\int u(t)dt$, и $u'(t)$.

При $\tau > 1$ с ситуация изменяется: влияние синусоидального импульса на W нервной ткани оказывается ниже треугольного. Этот эффект устойчив и четко выражен. Это можно объяснить только влиянием **второй производной сигнала** $u''(t)$. В самом деле, плавное изменение амплитуд обоих стимулов до и после точки где $u'(t) = 0$ не приводит к возбуждению нервных структур вследствие нейтрализации раздражителя работой механизма адаптации. Однако, если при прочих равных параметрах в точке максимума значение $u''(t)$ для синуса величина конечная, то для треугольника при $u'(t) = 0$ теоретически $u''(t)$ будет δ -функцией Дирака. Отметим, что сокращение мышцы происходит (разумеется, с латентным периодом) именно в этот момент.

Таким образом, описанные исследования уточняют (1): на W особо влияет величина стимула и при $u'(t) = 0$, и при $u''(t) = 0$. Каждая из рассмотренных компонент имеет свой вес $\alpha, \beta, \chi, \eta$,

количественное значение которых можно определить с помощью линейных форм вида

$$\alpha u(t) + \beta u'(t) + \chi \int_0^{\tilde{\tau}} u(t) dt + \eta(t) = 1, \quad (2)$$

где значения амплитуды сигнала, производных и интеграла учитываются в момент $\tilde{\tau}$ возникновения реакции нервной ткани.

В связи с появлением нового фактора — влияния второй производной сигнала — числовые величины коэффициентов α , β , χ должны иметь иные количественные значения, чем приведенные в [2], где W исследовалась в диапазоне длительностей от 0,01 до 0,1 с. При проведении количественного оценивания, целесообразно разделить диапазон длительностей стимулов на две части. Первую часть составят сигналы, у которых $\tau < 0,1$ с, вторую — стимулы, длительность которых больше 0,1 с.

Заключение

Обобщая данные описанных экспериментов, приходим к выводу: во-первых, зависимость (1) следует дополнить компонентой $u''(t)$; во-вторых, для W нервной ткани особо значимы в стимуле $u(t)$ участки, где $u'(t) = 0$. $u''(t) = 0$.

Таким образом, получаем важный для теории и практики критерий информативности фрагментов любого случайного сигнала. Измерение амплитудно-временных значений только этих фрагментов (дискретов) сокращает избыточность дискретных данных, позволяет использовать их для эффективного решения ряда актуальных проблем.

Сжатие исходных данных можно увеличить: при обработке быстрых процессов допустимо использовать **только экстремумы**. Практика показала, что использованием для анализа параметров экстремумов эффективно решается ряд весьма актуальных задач, включающих распознавание образов [6], быстрые и точные методы спектральных преобразований [7,8], корреляционного анализа [9-11], фильтрации [12], диагностики [13,14].

Литература

- [1] Ахмеров У.Ш. Основные этапы в изучении зависимости между пороговой силой и длительностью электрического раздражения.

- В сб. *Исследования по бионике*. Казань: КазГУ. Вып.1, с. 5-18. 1960.
- [2] Антомонов Ю.Г., Котова А.Б., Пономарева И.Д. и др. Математические модели возбуждения. — К.: Наукова думка, 144 с. 1965.
- [3] Катц Б. Нерв, мышца и синапс.- М.: Мир, 220 с. 1968.
- [4] Гилл А. Два фактора времени в электрическом раздражении нерва. *Физиол. ж. СССР*. Т.19, вып. 1, с. 115-129. 1935.
- [5] Цепков Г.В. Влияние формы сигнала на возбудимость нервной ткани. В сб. *Кибернетика и выч. техника* Вып.14, с. 16-20. 1972.
- [6] Логинов В.М., Цепков Г.В., Чинаев П.И. Экономичное кодирование. К.: Техника, 174 с. 1976.
- [7] Ponomareva I.D., Tsepkov G.V. The superfast spectral analysis. *J. Automation and Information Sci.* V. 31, No 2. p.p. 93-99. 1999.
- [8] Ponomareva I.D., Tsepkov G.V. Spectral analysis of signals on the basis of extrema. *Pattern Recognition and Image analysis*. V. 8, No 4, p.p. 560-567. 1998.
- [9] А. с. № 1352510 СССР. Устройство для определения задержки при корреляционном анализе случайных сигналов/ Г.В.Цепков. *Бюл.* № 42. 1987
- [10] Пономарева И.Д. Вычисление корреляционной функции случайных аналоговых сигналов, заданных экстремумами. *Проблемы управления и информатики*, №.4, с.111-115. 1995.
- [11] Tsepkov G.V. Correlation-spectral transformation of signals in adaptive sequence basis. *Cybernetics and System Analysis*. V.36, No 3 (May-June), p.p. 450-455. 2000.
- [12] Пономарева И.Д., Карпенко А.В., Цепков Г.В. Адаптивный секвентный базис в задачах фильтрации случайных сигналов. *Проблемы управления и информатики*. № 2, с.82-87. 2001
- [13] А. с. 731927 СССР. Устройство для экспресс-диагностики функционального состояния растений /В.И Лавриненко, А.П.Пасичный, И.Д.Пonomарева, Г.В.Цепков. *Бюл.* № 17. 1980.
- [14] Бакай Э.А., Пономарева И.Д., Цепков Г.В. Анализ электрограмм мозга по параметрам экстремумов. В сб. *Кибернетика и выч. техника*. Вып. 122, с. 64-71. 1999.

О собственных колебаниях круглой мембраны с утонченной центральной областью

Хахимов А.Г.⁽¹⁾, Кузин Р.А.⁽²⁾

Институт механики Уфимского научного центра РАН⁽¹⁾,
Городская клиническая больница № 21⁽²⁾
hakimov@anrb.ru

Abstract

Using two natural lateral vibration frequencies we can determine both radius and thickness of the membrane thinned central part.

1 Введение

Диагностика барабанной перепонки по собственным частотам поперечных колебаний является актуальной задачей. Например, в случае разрыва барабанной перепонки или наличия в ней отверстия ее колебания могут нарушаться, что приводит, в свою очередь, к нарушению слуха. Распознавание закрепления кольцевой мембраны по собственным частотам ее колебаний рассматривается в [1]. По трем собственным частотам изгибных колебаний консольной балки [2], балки на шарнирных опорах [3] определяются координата надреза, его глубина и длина.

2 Постановка задачи

Рассматриваются поперечные колебания круглой мембраны. Предполагается, что в мембране имеется центральный круговой участок с меньшей толщиной (рис. 1). Этот центральный круговой участок моделирует ее повреждение. Задача состоит в определении радиуса и толщины центральной области мембраны по двум низшим частотам ее поперечных колебаний.

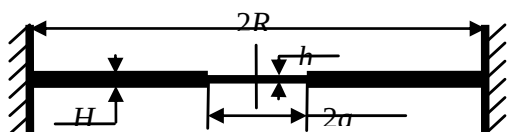


Рис. 1.

Обозначим через R , H радиус и толщину мембраны, a , h – радиус и толщину центрального кругового участка мембраны, T_0 , ρ – начальное натяжение, плотность, через w – прогиб мембраны.

Уравнение, определяющее форму поперечных колебаний мембраны постоянной толщины, имеет вид

$$\frac{\partial^2 w}{\partial r^2} + \frac{1}{r} \frac{\partial w}{\partial r} + \lambda^2 w = 0, \quad \lambda^2 = \frac{\rho H \omega^2}{T_0}, \quad (1)$$

где ω – частота. Отсчитывая координату r от центра мембраны, запишем граничные условия

$$w = 0 \quad (r = R). \quad (2)$$

Обозначая функции в областях $0 \leq r \leq a$, $a \leq r \leq R$ индексами «1», «2», соответственно, запишем условия стыкования решений при $r = a$ (условия равенства перемещений, углов поворота)

$$w_1 = w_2, \quad \frac{\partial w_1}{\partial r} = \frac{\partial w_2}{\partial r}, \quad (r = a). \quad (3)$$

Пользуясь в дальнейшем обозначениями

$$\xi = \frac{r}{R}, \quad \xi_c = \frac{a}{R}, \quad w = \frac{w}{R},$$

представим (2), (3) в виде

$$w_2 = 0 \quad (\xi = 1), \quad w_1 = w_2, \quad \frac{\partial w_1}{\partial \xi} = \frac{\partial w_2}{\partial \xi}, \quad (\xi = \xi_c). \quad (4)$$

Прогиб мембраны на двух участках представляется в виде [4]

$$w_1 = C_1 J_0(m \lambda \xi) + C_2 Y_0(m \lambda \xi),$$

$$w_2 = C_3 J_0(\lambda \xi) + C_4 Y_0(\lambda \xi),$$

$$\lambda = R \lambda^*, \quad \lambda^2 = \frac{\rho R^2 H \omega^2}{T_0}, \quad m = \sqrt{\frac{h}{H}},$$

где использованы стандартные обозначения цилиндрических функций. Согласно свойствам цилиндрических функций $C_2 = 0$.

Условия (4) в развернутом виде записываются

$$C_3 J_0(\lambda) + C_4 Y_0(\lambda) = 0,$$

$$C_1 J_0(m \lambda \xi_c) = C_3 J_0(\lambda \xi_c) + C_4 Y_0(\lambda \xi_c),$$

$$C_1 m J_1(m \lambda \xi_c) = C_3 J_1(\lambda \xi_c) + C_4 Y_1(\lambda \xi_c).$$

Таким образом, в приведенной простейшей модели повреждения фигурируют радиус ξ_c и параметр m для утонченной центральной области мембраны.

Для того, чтобы C_1 , C_3 , C_4 не были равны нулю одновременно, необходимо, чтобы определитель основной матрицы был равен нулю. Это условие дает частотное уравнение

$$\begin{aligned} & -J_0(m \lambda \xi_c) J_1(\lambda \xi_c) Y_0(\lambda) + \\ & + J_0(m \lambda \xi_c) J_0(\lambda) Y_1(\lambda \xi_c) - \\ & - m J_1(m \lambda \xi_c) J_0(\lambda) Y_0(\lambda \xi_c) + \\ & + m J_1(m \lambda \xi_c) J_0(\lambda \xi_c) Y_0(\lambda) = 0. \end{aligned} \quad (5)$$

Из полученного частотного уравнения по двум собственным частотам можно определить радиус ξ_c и параметр m для утонченной центральной области мембраны. Расчеты проведены для следующих параметров мембраны: $T_0=0.5$ Н/м, $\rho=1100$ кг/м³, $R=5$ мм, $H=0.1$ мм, $h=0.01$ мм, $a=0.5$ мм. При этом вычисленные значения первой и второй собственных частот мембраны без утончения $f_1=163.2008$ Гц, $f_2=374.6140$ Гц.

3 Прямая задача

Решение прямой задачи для мембраны с вышеприведенными параметрами дает, что $f_1 = 165.8489$ Гц, $f_2 = 387.2843$ Гц.

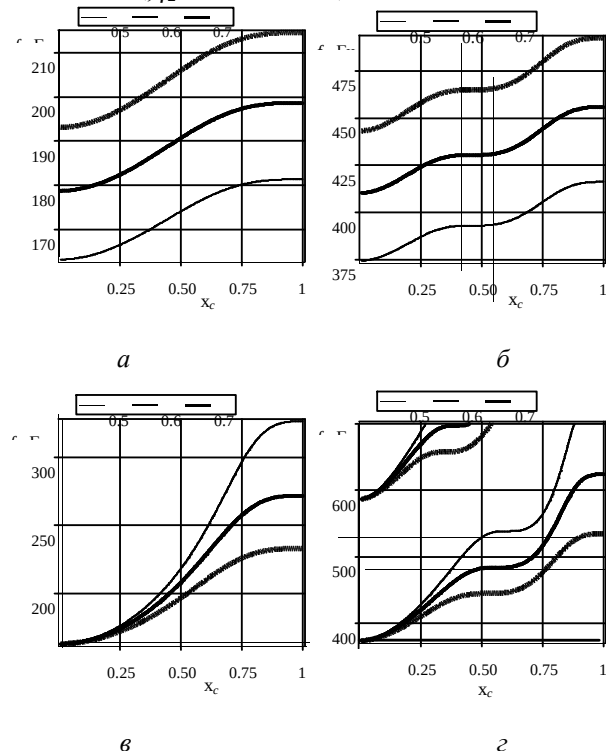


Рис. 2.

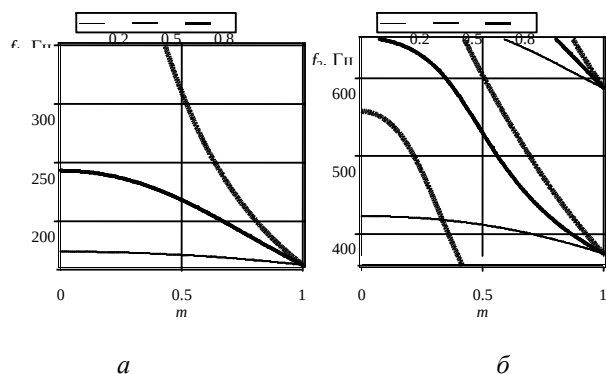


Рис. 3.

На рис. 2, а, б даются зависимости первой и второй собственной частоты поперечных колебаний от относительного радиуса ξ_c для $m = 0.9$ и различных значений начального натяжения T_0 , Н/м, а на рис. 2, в, г даются зависимости первой и второй

собственной частоты поперечных колебаний от относительного радиуса ξ_c для $T_0=0.5$ Н/м и различных значений параметра m . Из рис. 2 видно, что чем больше начальное натяжение T_0 , тем больше собственная частота и чем больше параметр m , тем меньше собственная частота поперечных колебаний. На рис. 3, а, б даются зависимости первой и второй собственной частоты поперечных колебаний от параметра m для $T_0=0.5$ Н/м и различных значений относительного радиуса ξ_c . Видно, что чем больше параметр m , тем меньше собственная частота поперечных колебаний мембраны.

4 Обратная задача

Решение обратной задачи для мембраны с вышеприведенными параметрами для $f_1=168$ Гц, $f_2 = 389$ Гц дает, что относительный радиус $\xi_c=0.3298$, параметр $m=0.9130$. На рис. 4 приводятся зависимости относительного радиуса ξ_c (рис. 4, а) и параметра m (рис. 4, б) для утонченной центральной области мембраны от первой частоты поперечных колебаний f_1 , для $f_2=388$ Гц (линия 1), $f_2=389$ Гц (линия 2), $f_2=390$ Гц (линия 3).

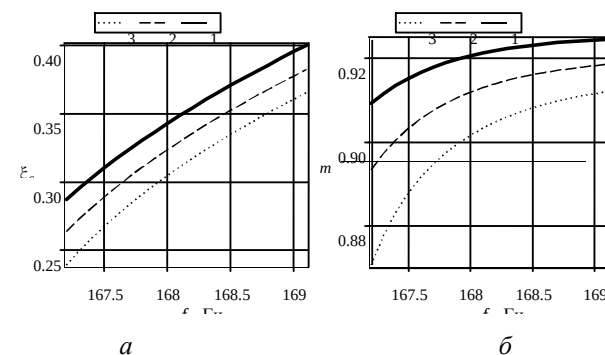


Рис. 4.

Проведенные исследования показывают, что по двум частотам поперечных колебаний мембраны можно определить относительный радиус ξ_c и параметр m для утонченной центральной ее области.

Результаты работы могут быть применены для диагностики круглых мембран, в частности, барабанной перепонки.

References

- [1] Ахтямов А.М. Теория идентификации краевых условий и ее приложения. М.: Физматлит, 2009. – 272 с.
- [2] Ильгамов М.А., Хакимов А.Г. Диагностика повреждений консольной балки с надрезом. Дефектоскопия. 2009. Том 45. № 6. – С. 83-89.
- [3] Ильгамов М.А., Хакимов А.Г. Диагностика повреждений балки на шарнирных опорах. Строительная механика инженерных конструкций и сооружений. 2010. № 2. – С. 42-48.
- [4] Гонткевич В.С. Собственные колебания пластин и оболочек. – Киев: Наукова думка, 1964. – 288 с.

Сравнение программ контроля туберкулеза с помощью математической модели эпидемического процесса

Мельниченко О.А.

Московский государственный университет им. М.В.Ломоносова,
ф-т Вычислительной математики и кибернетики
olesya.melnichenko@gmail.com

Comparison of TB control programs using a mathematical model of an epidemic process

Melnichenko O.A.

Abstract

Mathematical model of tuberculosis epidemiology in Russia is considered. The sensitivity of epidemiological indices to model parameters change is analyzed. Four TB control programs are compared using the results of sensitivity analysis. The results obtained suggest that an improvement in patients detection is the most effective control program.

$$\frac{dS}{dt} = -\beta S(B + kB_0 + \mu) + f_S(t),$$

$$\frac{dL}{dt} = (1-p)\beta S(B + kB_0) + \beta_L D + \beta_{L0} D_0 - L(\gamma + \alpha(B + kB_0) + \mu) + f_L(t),$$

$$\frac{dD}{dt} = p\beta S(B + kB_0) + L(\gamma + \alpha(B + kB_0)) + \beta_D B - (\beta_B + \beta_L + \varphi_D + \mu_D)D + f_D(t),$$

$$\frac{dB}{dt} = \beta_B D - (\beta_D + \varphi_B + \mu_B)B + f_B(t),$$

$$\frac{dD_0}{dt} = \varphi_D D + \beta_{D0} B_0 - (\beta_{B0} + \beta_{L0} + \mu_{D0})D_0 + f_{D0}(t),$$

$$\frac{dB_0}{dt} = \varphi_B B + \beta_{B0} D_0 - (\beta_{D0} + \mu_{B0})B_0 + f_{B0}(t),$$

где переменными модели являются численности следующих групп: S – восприимчивые индивиды; L – индивиды, несущие латентную инфекцию, но не развившие активных форм болезни; D – невыявленные незаразные больные туберкулезом; B – невыявленные заразные больные; D_0 – выявленные незаразные больные; B_0 – выявленные заразные больные. Подробное описание уравнений и параметров модели можно найти в [1].

С помощью теории сопряженных уравнений для рассматриваемой модели был проведен анализ чувствительности решений модели к возмущению параметров.

Было получено выражение для вариации основных эпидемиологических показателей (например, доли инфицированных индивидов, количества больных) через вариацию параметров модели, решение исходной и сопряженной задач [2].

3 Программы контроля туберкулеза

В работе рассмотрены стратегии контроля туберкулеза двух типов: направленные на повышение качества выявления больных (1 тип) и повышение качества лечения больных (2 тип).

Программа 1А ориентирована на выявление заразных больных, программа 1В – на выявление незаразных больных. Программы 2А и 2В ориентированы на лечение заразных и незаразных больных соответственно.

1 Введение

Математическое моделирование является одним из важных инструментов в эпидемиологии. Математические модели эпидемических процессов позволяют разрабатывать стратегии контроля инфекционных заболеваний, сравнивать их эффективность.

В работе рассмотрена математическая модель, описывающая основные особенности распространения туберкулеза на территории РФ [1].

Цель данной работы – сравнить несколько возможных программ контроля туберкулеза для регионов Центрального федерального округа, используя результаты настройки модели на реальные данные [1] и результаты анализа чувствительности модели к возмущению параметров [2].

2 Математическая модель

В качестве математической модели, описывающей процесс распространения туберкулеза в среде взрослого населения, рассматривается система дифференциальных уравнений

В рамках используемой математической модели программа контроля определялась изменением параметра модели, описывающего выявление или лечение интересующей нас группы больных. Значения параметров были получены в результате настройки модели на реальные данные [1].

Для 14 регионов ЦФО РФ рассчитывались изменения таких эпидемиологических показателей как численность невыявленных заразных больных (группа B), общее количество невыявленных больных (группы B и D), общее количество больных (группы B , D , B_0 и D_0) через 3 года и 20 лет после выбора каждой из 4 рассматриваемых программ. Затем вычислялось среднее изменение каждого показателя.

4 Результаты

В таблице 1 приведены средние изменения (в %) эпидемических показателей для ЦФО РФ через 3 года и 20 лет при повышении качества выявления и лечения больных на 20%; столбцы 2-5 соответствуют 4 программам, описанным ранее.

Результаты расчетов показывают, что стратегии, ориентированные на лечение больных, оказываются менее эффективными по сравнению со стратегиями, направленными на выявление.

Программы 2 типа затрагивают только группы выявленных больных, что приводит к малому изменению силы инфекции в популяции и незначительным изменениям эпидемических показателей – менее 2% через 3 года и менее 5% через 20 лет.

Стратегии 1 типа позволяют существенно снизить численность невыявленных больных (см. табл.1, столбцы 2 и 3) и, как следствие, вероятность заражения. Программа 1А позволяет в короткие сроки снизить на 14% численность невыявленных заразных больных, которые являются наиболее опасными для здоровых индивидов.

Эффективной в долгосрочной перспективе оказывается программа 1В (выявление незаразных больных): количество невыявленных заразных больных снижается на 18%, общее количество невыявленных больных – на 23%.

Таким образом, для быстрого снижения опасности заражения здоровых индивидов более

предпочтительно повышать качество выявления заразных больных. Однако в долгосрочной перспективе более ощутимые результаты приносит выявление больных на ранних стадиях болезни.

Таблица 1 Среднее изменение (в %) численности больных туберкулезом через 3 года и 20 лет для различных программ контроля для ЦФО РФ

Показатель	Программа контроля			
	1А	1В	2А	2В
3 года				
Невыявленные заразные больные	-14	-8.1	-0.1	-0.0004
Невыявленные больные	-3.5	-15	-0.3	-0.0003
Больные	-0.9	-4.5	-1.3	-1.9
Показатель	Программа контроля			
	1А	1В	2А	2В
20 лет				
Невыявленные заразные больные	-18	-18	-1.2	-0.09
Невыявленные больные	-6.5	-23	-1.6	-0.12
Больные	-2.7	-9.1	-3.1	-4.8

Литература

- [1] Melnichenko O.A., Romanyukha A.A. A model of tuberculosis epidemiology: estimation of parameters and analysis of factors influencing the dynamics of an epidemic process. // *Russ. J. Numer. Anal. Math. Modelling*, vol. 23, No. 1, pp. 63-75, 2008.
- [2] Мельниченко О.А. Модель распространения туберкулеза: анализ чувствительности // *Труды 4 Международной конференции по проблемам управления*, стр. 857-863, 2009.

Влияние социо-экономических факторов на выявление больных туберкулезом

Авилов К.К.

Институт вычислительной математики РАН

avilov@inm.ras.ru

Abstract

Avilov K.K. Impact of socio-economical factors on tuberculosis case detection.

Case underdetection is a common problem in the epidemiological data. It can be addressed with multi-stage disease models and disease-stage stratified data.

A model to analyze the impact of socio-economical factors on the tuberculosis case detection is constructed. It is built on the preceding multi-stage models that describe populational data. Personalized data is analyzed and a quantification of a qualitative experts' hypothesis is obtained.

1 Введение

При анализе эпидемиологических данных по заболеваемости различными болезнями возникает проблема соответствия между статистическими данными и реальной (полной) заболеваемостью [2,3]. Факты посмертного выявления болезни и выявления людей с остаточными изменениями, подтверждающими перенесенную болезнь, указывают на то, что выявляются далеко не все заболевшие, а поэтому оценка заболеваемости населения, основанная непосредственно на количестве выявленных больных, является заведомо заниженной.

В работах [2,3] на примере туберкулеза был предложен метод оценки полной заболеваемости, основанный на многостадийной математической модели хода болезни и данных по выявлению больных со стратификацией по стадиям болезни на момент выявления. Используемая модель описывает конкуренцию процессов выявления, гибели и спонтанного самоизлечения невыявленных больных.

Процесс выявления больных, очевидно, подвержен сильному влиянию социо-экономических факторов. В работе [1] метод, предложенный в [2], был модифицирован и с его помощью был проведен анализ полицейских данных по выявлению туберкулеза органов дыхания, содержащих информацию о социо-экономическом статусе каждого больного.

2 Математическая модель

В [1] использовалась двухстадийная модель течения туберкулеза с последовательными стадиями фиксированной длительности. Стадии разделялись по надежно диагностируемой клинической форме туберкулеза: к поздней стадии относились больные с распадом тканей легких и с клиническими формами туберкулеза, соответствующим хронизации болезни; все остальные больные относились к ранней стадии. Длительность ранней стадии принималась равной $T_1=0,5$ года. Для каждой стадии задавались суммарные удельные скорости смерти гибели и спонтанного самоизлечения ($\mu_1=0,4$ 1/год и $\mu_2=0,5$ 1/год). Процесс выявления больных задавался удельными скоростями выявления по стадиям (φ_1, φ_2). Основным нововведением модели [1] является концепция «доступности для выявления» a , зависящей от индивидуальных социо-экономических параметров больного и являющаяся мультипликативным модификатором удельной скорости выявления.

Уравнения модели описывают изменение функции дожития ℓ в зависимости от времени с начала болезни t :

$$\frac{d\ell}{dt} = -(a\varphi_1 + \mu_1)\ell, \quad t \in [0, T_1),$$

$$\frac{d\ell}{dt} = -(a\varphi_2 + \mu_2)\ell, \quad t \in [T_1, +\infty),$$

$$\ell(0) = 1.$$

Эти уравнения могут быть решены аналитически и в результате могут быть получены явные формулы для вероятности выявления на ранней и поздней стадиях для каждого больного (как функции его «доступности для выявления» a и средних скоростей выявления φ_1 и φ_2).

3 Социо-экономические факторы

Имеющиеся полицейские статистические данные описывают социо-экономический статус выявленных больных туберкулезом как набор из 10 параметров (пол, возраст, социально-профессиональная принадлежность, опыт тюремного заключения, нахождение на учете в мед.учреждениях и т.п.), каждый из которых

принимает только дискретные значения. Поэтому была принята следующая модель зависимости «доступности для выявления» a от социо-экономических параметров:

$$a = \exp\left(\sum_j w_j I_j\right),$$

где j – индекс фактора (т.е. одного из возможных значений какого-либо параметра), w_j – вес фактора, а I_j – индикатор наличия фактора у данного индивида.

4 Результаты

Настройка модели проводилась при помощи прямого метода максимального правдоподобия.

В качестве объекта настройки использовались данные из Брянской, Ярославской и Саратовской областей РФ за 2004-2006, 2001-2006 и 2004-2006 годы соответственно.

Непосредственным результатом настройки явились оценки констант скоростей выявления для этих трех областей (со стратификацией по активному и пассивному выявлению) и оценки весов для социо-экономических факторов. Полные таблицы результатов приведены в [1].

Наиболее способствующими быстрому (раннему) выявлению оказались такие факторы как проживание в учреждениях социальной защиты, нахождение на учете в учреждениях системы противотуберкулезной помощи и женский пол. Наиболее препятствующими раннему выявлению оказались хронический алкоголизм, предшествующий опыт лишения свободы и принадлежность к лицам без определенного места жительства (БОМЖ).

5 Обсуждение

Предложенная простая модель является инструментом первичного анализа данных, описывающих влияние социо-экономических факторов на выявление больных туберкулезом. Тем не менее, она позволила количественно подтвердить высказываемое экспертами-медиками предположение, что главным препятствием для выявления больных является асоциализация больного, выражающаяся в его нежелании попадать в «поле зрения» государственных учреждений вообще и медицинских учреждений в частности. Такое поведение типично для тяжелых алкоголиков, граждан с криминальным прошлым и лиц БОМЖ. С другой стороны, раннее выявление наблюдается преимущественно у людей, находящихся под постоянным медицинским контролем. Более раннее выявление туберкулеза у женщин можно объяснить как другим ходом болезни (что не учитывается в используемой модели), так и большей настороженностью по отношению к собственному здоровью и более частым посещением медицинских учреждений.

Перспективным путем развития данной модели представляется учет воздействия социо-экономических факторов на ход болезни.

References

1. K.K. Avilov. Statistical description of factors determining detected tuberculosis incidence. *Russ. J. Numer. Anal. Math. Modelling*, vol. 24(4), pages 309-324, 2009.
2. Авилов К.К., Романюха А.А. Математическое моделирование процессов распространения туберкулеза и выявления больных. В *Автоматика и телемеханика*. №9, с. 145-160. 2007.
3. К.К. Авилов. Оценка скрытой заболеваемости по реальным данным. *Доклады ICMBB-2008*, с. 172-173, 2008.

Математические модели эпидемического процесса с произвольным описанием длительности пребывания индивидуумов в различных стадиях заболевания

Перцев Н. В.

Омский филиал Института математики им. С. Л. Соболева СО РАН
homlab@ya.ru

Abstract

We consider one of the possible approach to the construction the family of deterministic and stochastic stage-dependent models using arbitrary distributions for durations of various stages of epidemic processes.

$$\begin{aligned}\frac{dx}{dt} &= k(x+y+z) - \delta x - \beta xy + \gamma z, \\ \frac{dy}{dt} &= \beta xy - (\lambda + \delta + \nu)y, \quad t > 0, \\ \frac{dz}{dt} &= \nu y - (\delta + \gamma)z, \\ x(0) &= x_0, y(0) = y_0, z(0) = z_0.\end{aligned}\tag{1}$$

1 Введение

Работа посвящена методологии построения детерминированных и стохастических моделей эпидемических процессов, предполагающих неоднородность популяции индивидуумов по их принадлежности к различным стадиям заболевания, длительность которых может описываться с помощью распределений достаточно общего вида. Как известно, экспоненциальное распределение не обладает «памятью» и позволяет строить модели эпидемических процессов в форме систем обыкновенных дифференциальных уравнений или многомерных случайных процессов рождения и гибели. Если при построении моделей применяются распределения, отличные от экспоненциального, то возникают уравнения в частных производных, интегральные или интегродифференциальные уравнения. Эти уравнения могут иметь достаточно сложный вид. Для построения стохастических моделей возникает необходимость применения полумарковских случайных процессов или марковских случайных процессов со специальным фазовым пространством. Примеры таких моделей приведены в следующих разделах.

2 Примеры интегродифференциальных моделей распространения эпидемии

Рассмотрим модель, описывающую динамику следующих переменных: $x = x(t)$, $y = y(t)$ — число восприимчивых и заболевших индивидуумов, $z = z(t)$ — число индивидуумов, которые приобрели иммунитет после выздоровления. Уравнения этой модели таковы [1]:

Обратимся к одной из возможных модификаций модели (1), позволяющей учитывать длительность заболевания τ_y и продолжительность сохранения иммунитета τ_z для отдельных индивидуумов. Положим, что распределения этих величин задаются с помощью функций

$$R_y(a) = \int_a^\infty r_y(s) ds, \quad R_z(a) = \int_a^\infty r_z(s) ds, \quad 0 \leq a < \infty, \\ r_y(s) \geq 0, \quad r_z(s) \geq 0, \quad R_y(0) = R_z(0) = 1.$$

Для фиксированного $a_0 > 0$ значения $R_y(a_0)$ и $R_z(a_0)$ определяют долю индивидуумов, которые находятся в соответствующих стадиях не менее, чем a_0 единиц времени ($\tau_y \geq a_0$, $\tau_z \geq a_0$). Обращаясь к уравнениям (1), вместо слагаемых νy и γz используем операторы

$$\rho_y(t) = \int_0^\infty e^{-(\lambda+\delta)a} r_y(a) \beta x(t-a) y(t-a) da, \\ \rho_z(t) = \int_0^\infty e^{-\delta s} r_z(s) \rho_y(t-s) ds, \quad t \geq 0.$$

Эти операторы описывают скорости перехода индивидуумов между соответствующими группам с учетом распределений величин τ_y и τ_z . Исключая из уравнений системы (1) слагаемые $\nu y(t)$, $\gamma z(t)$ и заменяя их соответственно на $\rho_y(t)$ и $\rho_z(t)$, приходим к следующей модификации модели Мэя-Андерсона:

$$\begin{aligned}\frac{dx}{dt} &= k(x+y+z) - \delta x - \beta xy + \rho_z(t), \\ \frac{dy}{dt} &= \beta xy - (\lambda + \delta)y - \rho_y(t), \quad t > 0, \\ \frac{dz}{dt} &= \rho_y(t) - \delta z - \rho_z(t), \\ x(t) &= x_0(t), \quad y(t) = y_0(t), \quad z(t) = z_0(t), \quad t \leq 0.\end{aligned}\quad (2)$$

Если положить, что $R_y(a) = e^{-\nu a}$ и $R_z(a) = e^{-\gamma a}$, то модель (2) преобразуется к модели (1). Результаты исследования модели (2) показывают, что варьирование функций $R_y(a)$, $R_z(a)$ при сохранении неизменных средних значений

$$\int_0^{\infty} R_y(a) da = 1/\nu, \quad \int_0^{\infty} R_z(a) da = 1/\gamma$$

может приводить к существенным изменениям в поведении переменных $x(t)$, $y(t)$, $z(t)$.

Описанный подход использован для модификации математической модели распространения и контроля за туберкулезом на территории России (Перельман, Марчук и др. [2]). Исходная модель в форме системы шести дифференциальных уравнений расширена до системы из восьми интегродифференциальных уравнений с учетом пребывания индивидуумов в стадии ремиссии.

3 Примеры стохастических моделей распространения эпидемии

Одним из важных достоинств стохастических моделей является то, что они оперируют с целочисленными переменными, учитывают случайный характер контактов индивидуумов и продолжительностей стадий эпидемического процесса. В ряде случаев для построения моделей применяются случайные процессы с дискретным временем, что может отражать периодически повторяющиеся (например, ежедневные) посещения индивидуумами мест их возможных контактов.

Примером такой модели является система стохастических уравнений, используемая для описания процесса распространения туберкулеза. Рассматривается шесть групп населения: S , L — восприимчивые и инфицированные индивидуумы, D , B — невыявленные больные индивидуумы без бактериовыделения и с бактериовыделением, D_0 , B_0 — выявленные больные индивидуумы без бактериовыделения и с бактериовыделением. Переменные модели $x_S(t)$, $x_L(t)$, $x_D(t)$, $x_B(t)$, $x_{D_0}(t)$, $x_{B_0}(t)$ представляют собой численности соответствующих групп индивидуумов в момент времени t . Уравнения модели содержат слагаемые, смысл которых поясним на примере уравнения для $x_L(t)$: $\bar{x}_L(t)$ — количество индивидуумов группы L , доживших от момента $t-1$ до момента t , $u_{S,L}(t)$ — количество индивидуумов, перешедших

из группы S в группу L за промежуток времени $(t-1, t]$, $f_2(t)$ — приток мигрантов в группу L за $(t-1, t]$. Система уравнений модели имеет вид [3]:

$$\begin{aligned}x_S(t) &= \bar{x}_S(t) - u_{S,L \cup D}(t) + f_1(t), \\ x_L(t) &= \bar{x}_L(t) + u_{S,L}(t) + u_{D,L}(t) + \\ &\quad + u_{D_0,L}(t) - u_{L,D}(t) + f_2(t), \\ x_D(t) &= \bar{x}_D(t) + u_{S,D}(t) + u_{L,D}(t) + \\ &\quad + u_{B,D}(t) - u_{D,L}(t) - u_{D,B}(t) - \\ &\quad - u_{D,D_0}(t) + f_3(t), \\ x_B(t) &= \bar{x}_B(t) + u_{D,B}(t) - u_{B,D}(t) + \\ &\quad + u_{D_0,B}(t) - u_{B,B_0}(t) + f_4(t), \\ x_{D_0}(t) &= \bar{x}_{D_0}(t) + u_{D,D_0}(t) + u_{B_0,D_0}(t) - \\ &\quad - u_{D_0,B_0}(t) - u_{D_0,L}(t) + f_5(t), \\ x_{B_0}(t) &= \bar{x}_{B_0}(t) + u_{B,B_0}(t) + u_{D_0,B_0}(t) - \\ &\quad - u_{B_0,D_0}(t) + f_6(t), \quad t = 1, 2, \dots\end{aligned}$$

Эта система дополняется начальными данными при $t = 0$. Условные законы распределения отдельных слагаемых опираются на мультиномиальное распределение. Для проведения вычислений использован метод Монте-Карло.

В ряде случаев модель эпидемического процесса может включать такие важные величины, как возраст индивидуумов, продолжительность периода времени, прошедшего после их инфицирования, длительность и тяжесть различных стадий заболевания и т.д. Законы распределения этих величин могут иметь достаточно произвольный характер. Здесь возникает необходимость в построении так называемых индивидуум-ориентированных или агентных моделей. В качестве примера можно привести модель распространения и контроля за туберкулезом, разработанную на базе модели сообщества взаимодействующих индивидуумов и ее компьютерной поддержки «Population Modeler» [4].

Работа поддержана РФФИ, проект № 09-01-00098-а.

References

- [1] R.M. Anderson, R.M. May. Population biology of infectious disease. Part I // Nature, volume 280, p. 361-367, 1979.
- [2] M.I. Perelman, G.I. Marchuk, et. al. Tuberculosis epidemiology in Russia: the mathematical model and data analysis // Russ. J. Numer. Anal. Math. Modelling, volume 19, N.4, p.305-314, 2004.
- [3] N.V. Pertsev, V.N. Leonenko. Stochastic model of spread of tuberculosis // Russ. J. Numer. Anal. Math. Modelling, volume 24, N.4, p.341-360, 2009.
- [4] Н.В. Перцев, Б.Ю. Пичугин. Индивидуум-ориентированная модель распространения туберкулеза // Сиб. журн. инд. матем., том 12, N.2 (38), с. 97-110, 2009.

Математическое моделирование смертности по причинам при сочетанной заболеваемости

Цурко В.В., Михальский А.И.

Институт проблем управления РАН, г.Москва
ipuran@yandex.ru

Abstract

Multiple cause mortality data are analyzed using a new mathematical model aimed to reflect dependency between cause specific mortality and associated morbidity. population.

1 Введение

В свете продолжающегося старения населения экономически развитых стран появились новые проблемы обеспечения «здоровой старости» среди населения старших возрастных групп. Улучшение состояния здоровья лиц пожилого возраста оправдано экономически и социально, продлевая период трудовой занятости, социализации пожилых, снижая нагрузку на пенсионную социальную систему. В настоящее время собраны большие объёмы статистических данных о смертности и заболеваемости пожилых людей, которые позволяют исследовать факторы, ответственные за сохранение здоровья в пожилом возрасте, оценивать вклад наследственности, окружающей среды и образа жизни.

Теоретической основой для моделирования и оценок при изучении процессов старения является геронтология, позволяющая связать физиологию, биологию, медицину с проявлениями старения и, в частности, с увеличением вероятности смерти с возрастом. Геронтологические подходы позволяют исследовать не только старение индивида, но и рассматривать проявления старения на популяционном уровне, обращая основное внимание на увеличение смертности в популяции с возрастом и на влияющие на это факторы.

В последнее время появились работы, объясняющие наблюдаемое увеличение продолжительности жизни снижением смертности в средних возрастах [1]. При этом выдвигается гипотеза, что люди, обладающие слабым здоровьем, имеют высокий запас активного долголетия и именно эти люди, пережившие период повышенного риска в молодом и среднем возрасте, имеют преимущества в выживании в старости [4]. Биологической основой этого явления может служить более эффективная система адаптации организма у людей, страдающих хроническими заболеваниями. В случае подтверждения этой гипотезы следует вывод о замедлении наблюдающегося сейчас роста продолжительности жизни и о необходимости сосредоточить

профилактические мероприятия по обеспечению «здоровой старости» на возрастных группах молодого и среднего возраста.

Косвенным подтверждением наличия связи между повышенной заболеваемостью и пониженной смертностью может служить связь между смертностью по фиксированной причине и сопутствующими при этом хроническими заболеваниями. Такие данные собираются на регулярной основе Национальным центром статистики по здравоохранению США (National Center for Health Statistics USA), предварительный анализ показал на качественном уровне, что смертность от онкологических заболеваний среди лиц, страдающих астмой, ниже, чем среди лиц, не страдающих астмой [2]. Целью настоящей работы является оценка вероятности наступления смерти по выделенной причине в данной возрастной группе и в данном году при условии наличия выделенного хронического заболевания или комплекса таких заболеваний.

2 Данные

Использовались данные о причинах смерти и ассоциированных заболеваниях, представленные Национальным центром статистики по здравоохранению США, и предназначенные для публичного доступа [3]. В данных представлены неперсонифицированные записи, содержащие год смерти, пол, возраст смерти, причину смерти в соответствии с международным классификатором заболеваний ICD10, коды сопутствующих заболеваний. Другая сопутствующая информация в настоящем исследовании не использовалась. Для демонстрации предложенного подхода из данных за 2007 год было отобрано по 100 записей, соответствующих смерти от рака в различных возрастных группах. В каждой возрастной группе строилась таблица сопряжённости 2×2 , соответствующая смерти от онкологического заболевания (ICD10 C00-C97) и наличию астмы как сопутствующего заболевания (ICD10 J45-J46). Полученные таблицы исследовались с помощью описываемой модели.

3 Математическая модель

Математическая модель, использованная при оценивании условной вероятности наступления смерти по выделенной причине при условии

наличия выделенного хронического заболевания или комплекса таких заболеваний, является рассмотрение популяции, состоящей из двух групп. Одна группа состоит из людей, для которых выделенное заболевание не влияет на смертность по рассматриваемой причине. Биологически это могут быть люди со слабо развитой системой адаптации. Ко второй группе относятся люди, для которых наличие выделенного заболевания влияет на смертность по рассматриваемой причине, это люди с развитой системой адаптации. Пропорцию людей первой группы в популяции обозначим γ , эта величина имеет смысл вероятности выбрать случайным образом из популяции представителя первой группы. Обозначив через A событие смерти по выделенной причине, а через B вероятность обнаружить в популяции человека, страдающего выделенным хроническим заболеванием или комплексом таких заболеваний, совместную вероятность событий A и B представим в виде $P(A, B) = (\gamma P(A) + (1 - \gamma)P(A|B))P(B)$. Здесь $P(A|B)$ обозначает условную вероятность смерти по выделенной причине при наличии выделенного хронического заболевания или комплекса таких заболеваний для члена второй группы. В зависимости от величины параметра γ такая модель описывает случай независимости событий A и B ($\gamma = 1$), зависимости A и B ($\gamma = 0$) и случай «не обязательной» зависимости A и B при промежуточном значении γ . Оценка параметров γ , $P(A)$, $P(A|B)$ для разных лет и различных возрастных групп позволяет выявлять влияние заболеваемости на смертность, формулировать новые биологические и геронтологические гипотезы, подлежащие экспериментальной проверке, и оценивать популяционную роль этой связи. В частности, увеличение параметра γ с возрастом может указывать на происходящие в организме человека изменение связей между различными системами жизнеобеспечения, их «разладку», что, в свою очередь, отражается в некоторых теориях старения.

Для оценки параметров модели используется метод максимального правдоподобия для гипергеометрического распределения, усреднённого по априорному распределению «мешающих» параметров. В качестве априорного распределения использовалось произведение бета-распределений

4 Анализ смертности от онкологических заболеваний среди лиц, страдающих астмой

Описанная модель применялась для анализа смертности от онкологических заболеваний при наличии и при отсутствии астмы как сопутствующего заболевания. Использовались данные о среди женщин за 2007 год в

возрастной группе 55-80 лет. Причина смерти относилась к онкологическому заболеванию, если её код согласно международному классификатору заболеваний ICD-10 лежал в интервале от C00 до C97. Считалось, что человек страдал астмой, если среди сопутствующих заболеваний находилось заболевание с кодом по ICD-10 от J45 до J46. Полученные оценки показывают, что величина параметра γ в выражении для вероятности смерти от онкологических заболеваний среди людей, страдающих астмой, не превышает величину 0.05. Это можно интерпретировать как то, что в изучаемой популяции более, чем для 95% женщин в возрастной группе 55-80 лет наличие астмы влияет на вероятность смерти от онкологических заболеваний.

5 Заключение

Представленный в докладе подход перспективен для уточнения факторов, влияющих на популяционную смертность, и выявления взаимосвязи между смертностью и заболеваемостью. Анализ с этой точки зрения данных о смертности по причинам для различных возрастных групп позволит получить новые знания, важные для геронтологии, анализа и прогноза смертности, оценки будущей динамики численности популяции.

References

- [1] M.V. Blagosklonny. Why human lifespan is rapidly increasing: solving “longevity riddle” with “revealed-slow-aging” hypothesis. *Aging*, vol. 4, p.177-182, 2010.
- [2] A.I. Michalski, S.V. Ukraintseva, K.G. Arbeev, A.I. Yashin. Investigation of old age mortality structure in the presence of co-morbidity. In *European Conference On Chronic Disease Prevention, Helsinki, Finland*, p.60, 2005.
- [3] Mortality Data, Multiple Cause-of-Death Public-Use Data Files. http://www.cdc.gov/nchs/products/elec_prods/subject/mortmcd.htm
- [4] A.I. Yashin, S.V. Ukraintseva, G. De Benedictis, V.N. Anisimov, A.A. Butov, K. Arbeev, D.A. Jdanov, S.I. Boiko, A.Z. Begun, M. Bonafe, C. Franceschi. Have the oldest old adults ever been frail in the past? A hypothesis that explains modern trends in survival. *J. Gerontol. Biol. Sci.*, vol.56, p.B432-B442, 2001.

Математическая модель иммунной памяти

Сергеев В.Г., Перевозчиков А.Г.

ГОУ ВПО «Удмуртский государственный университет»
cellbio@ya.ru

Abstract

The mathematical model of immune memory on the basis of the [system of differential equations](#) was investigated. The model assumes presence of long-living memory cells and idiotype- antiidiotypic reproduction of antigen image in the immune system. Research of model properties has shown conformity of changes of its parameters to dynamics of cellular and humoral reactions underlying the immune reaction and the immune memory.

Иммунной памятью называется способность иммунной системы отвечать более быстрыми и эффективными клеточными и гуморальными реакциями на повторное проникновение в организм чужеродного антигена. О наличии памяти к данному антигену можно судить по сохраняющемуся продолжительное время после его элиминации увеличенному титру специфических антител и/или увеличению частоты антигенспецифических клонов Т- и В- клеток «памяти». Длительность сохранения иммунной памяти у человека варьирует от нескольких месяцев до десятков лет.

В настоящее время обсуждается несколько гипотез о механизмах иммунной памяти. Первоначально предполагалось, что иммунная память обеспечивается долгоживущими лимфоцитами [2]. Однако, поскольку срок жизни лимфоцитов намного короче, чем продолжительность жизни целого организма, позже было высказано предположение о возможности постоянного возобновления клеток памяти в результате их рестимуляции персистирующими антигенами, сохраняющимися в организме после первичной иммунизации [1,5].

С другой точки зрения, клетки памяти представляют собой специфическую группу лимфоцитов, которая после первичной стимуляции антигеном не погибает как большинство лимфоцитов, а приобретает свойства стволовых клеток [3].

Альтернативная гипотеза рассматривает иммунологическую память как динамический процесс чередующейся взаимной активации комплементарных идиотип-антиидиотипических

лимфоцитарных клонов, неопределенно долго генерирующих синусоидные колебания антиген-специфических и антиидиотипических антител. [4,6].

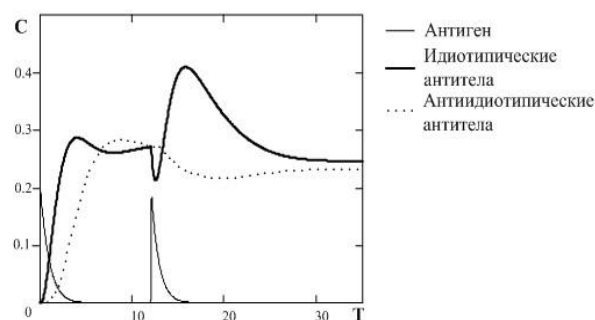
Для проверки истинности выдвинутых гипотез мы с помощью программы Matchcad 14.0 провели сравнительный анализ математических моделей иммунной памяти различной архитектуры с использованием системы дифференциальных уравнений. Адекватность математических моделей оценивали на основе соответствия их поведения известным феноменам иммунной системы: длительному увеличению числа антигенспецифических лимфоцитов и сохранению повышенного титра специфических антител после первичного иммунного ответа, а также сокращению времени элиминации антигена при его повторном введении в организм.

Сравнительный анализ продемонстрировал тот факт, что ни одна из математических моделей, описывающих выдвинутые ранее гипотезы иммунной памяти, не симулирует в полной мере известную иммунологическую феноменологию. Поэтому, нами была предложена интегральная модель иммунной памяти, сочетающая в себе наличие долгоживущих клеток памяти с повышенной реактивностью на стимулирующее действие антигена (или его «образа» в виде антиидиотипических антител), и сетевого (идиотип-антиидиотипического) способа воспроизводства в иммунной системе образа антигена и реагирующих на него лимфоцитов (Рис.1).

В предложенной модели активация антигеном (Ag) антиген-специфических наивных клеток (N_1) переводит их в состояние стимулированных клеток (S_1), которые пролиферируют и дифференцируются в клетки памяти (M_1), регуляторные клетки (R_1), эффекторные (плазматические) клетки (E_1) и долгоживущие эффекторные (плазматические) клетки (ME_1). Последние две популяции продуцируют антитела (Ab_1), которые, инактивируют антиген посредством образования иммунных комплексов ($AgAb_1$). Поскольку синтезированные в результате активации лимфоцитарного клона №1 антитела (Ab_1) являются новой антигенной формой для данного организма, они способны активировать распознающие их наивные лимфоциты (N_2). Дальнейшие события в

The diagram illustrates a two-population model with nodes $N_1, N_2, M_1, M_2, S_1, S_2, E_1, E_2, R_1, R_2, ME_1, ME_2$ and various interactions. Nodes $N_1, N_2, S_1, S_2, E_1, E_2, R_1, R_2$ are represented by single circles, while M_1, M_2, ME_1, ME_2 are double circles. Solid arrows represent transitions between these states. Dashed arrows represent interactions involving Ag (antigen), Ab_1 (antibody 1), Ab_2 (antibody 2), $AgAb_1$ (antigen-antibody 1 complex), and Ab_1Ab_2 (antibody 1-antibody 2 complex). The diagram shows a complex network of these interactions, including feedback loops and cross-population effects.

Исследование свойств данной модели продемонстрировало соответствие изменений ее параметров динамике клеточных и гуморальных реакций иммунной системы, лежащих как в основе первичного иммунного ответа, так и иммунной памяти. В частности, в моделируемой системе сохраняются повышенные уровни идиотипических антител после первичной иммунизации (стадия иммунной памяти) и наблюдается более выраженный рост продукции антител при повторном введении антигена (Рис.2).



Список литературы

- [1] D.Gray, H.Skarvall. B-cell memory is short-lived in the absence of antigen. *Nature*, volume 336, pages 70–73, 1988.
- [2] L.Hood, I. L.Weissman, W.B.Wood, J.H.Wilson. *Immunology*. Menlo Park, California: Benjamin/Cummings, 1984.
- [3] C.J.[Luckey](#), D.[Bhattacharya](#), A.W.[Goldrath](#), I.L.[Weissman](#), C.[Benoist](#), D.[Mathis](#). Memory T and memory B cells share a transcriptional program of self-renewal with long-term hematopoietic stem cells. [*Proc Natl Acad Sci U S A*, volume 103\(9\)](#), page 3304-3309, 2006.
- [4] R.Nayak, S.Mitra-Kaushik, M.S.Shaila. Perpetuation of immunological memory: a relay hypothesis. [*Mol Immunol.*, volume 46\(6\)](#), page 1250-1255, 2009.
- [5] J.G.Tew, R.P.Phipps, T.E.Mandel. The maintenance and regulation of the humoral immune response: persisting antigen and the role of follicular antigen-binding dendritic cells as accessory cells. *Immunol Rev.*, volume 53, page 175-201, 1980.
- [6] J.[Vani](#), J.[Chatterjee](#), M.S.[Shaila](#), R.[Nayak](#), N.R.[Chandra](#). Structural basis for the function of anti-idiotypic antibody in immune memory. *Mol Immunol.*, volume 46(6), page1250-1255, 2009.

Biodiversity assessment in Decision Support Systems for sustainable forest management

Larisa G. Khanina, Vadim E. Smirnov, Elena M. Glukhova

Institute of Mathematical Problems in Biology of RAS Pushchino

e-mail: lkhanina@rambler.ru

Abstract

A review of existing tools for biodiversity assessment and prognosis in the frame of Decision Support Systems (DSS) for sustainable forest management is presented. Special attention is devoted to systems with assessment and prognosis of ground vegetation dynamics.

1 Introduction

Forest biodiversity conservation is an essential element of sustainably managed forests. It is one of “Improved Pan-European Indicators for Sustainable Forest Management” and it is the first criterion in the set of Montreal Protocol Criteria and Indicators developed for the management in boreal and temperate forests [3]. Biodiversity can be considered at different levels, including the ecosystem, landscapes, species, populations and genes. In biologically diverse forests, the complexity of relationships and interactions enables organisms to adapt to continually changing environmental conditions and to maintain ecosystem functions. To consider approaches and tools for biodiversity assessment embodied in the modern DSS for forest management is an objective of this paper.

2 General background

Forests serve a multitude of purposes and multifunctional forest management planning has to involve complex decision-making processes that cannot be addressed solely by empirical methods; advanced decision-support tools are required. The function of a DSS is to organise the decision process, and provide flexible, on-demand access to the full array of prescriptive, predictive, and descriptive tools applicable to a particular problem situation. A DSS should be able to assist in carrying through the basic phases of the decision-making process. These include problem identification (diagnosing the actual issue), design (elaboration of different scenarios/ alternatives), choice (evaluation of and choice among alternatives),

implementation and monitoring. This is accomplished by integrating a database management system (DBMS) with a model base management system (MBMS) where system components accessed through a user interface. The DBMS allows the user to retrieve the needed data; the MBMS lets the user access the appropriate models and tools needed for the analysis. Within the domain of application intended for the DSS it should have the flexibility to permit the user to adapt to the particular issues at stake. Likewise, a DSS should have mechanisms for letting the user (and not only the systems expert) to bring in new data and models into the DSS.

Sustainable forest management encompasses problems of different temporal and spatial scales, operates with uncertain data from many different sources, and often involves multiple stakeholders. The spatial context of decision problems has seen the arrival of spatial DSS. Integrating the analytical capabilities of GIS into the DSS allows modelling spatial data components and conducting spatial analysis. Multi-criteria decision making techniques have been integrated with spatial DSSs to help decision makers model trade-offs between multiple and conflicting objectives and to facilitate participative decision making. The development of hybrid systems, combining the functionality of DSS and artificial intelligence approaches such as artificial neural networks, knowledge based systems and expert systems, is a promising research area. Group Decision support systems for forest management are being developed to respond to the demand to support decision-making processes in multi stakeholder contexts.

3 Biodiversity assessment

A list of 50 existence DSS for forest management is presented in the Wiki page of the FORSYS Cost Action [1]. The analysis showed that biodiversity assessment and prognosis have been rarely applied in DSS. ‘Biodiversity conservation’ has been marked only in 18 from 50 registered DSS [1], and only tree species diversity can be estimated in the majority of the systems.

A detailed review of DSS for forest biodiversity was done in the project of the USA National Commission on Science and Sustainable Forestry [4]. Over 100 tools were observed there and 30 from them were detailed analyzed as the most applicable to forest biodiversity decisions. The analysis showed that many DSS can address components of forest biodiversity, but only five systems appear to integrate capabilities for both forest and biodiversity modelling. Generally, 'biodiversity DSS' focuses either on population modelling, either on reserve selection problem. The first one deals with estimations of the size of game species populations over time, given different assumptions about breeding and available habitat. The second one involves finding the most efficient set of land parcels for conserving biodiversity, based on a landscape in which each parcel is assigned one or more biodiversity values. There was marked, while many of the 'combined DSS' model the effects of silviculture and land use change, relatively few of these systems or the 'biodiversity DSS' include tools to address the influences of fire, biological threats (pest, pathogens, invasive species) or climate change on forest habitat or biodiversity. The forest modelling systems, on the other hand, frequently consider silviculture, fire, climate change, and biological threats, but they generally do not include mechanisms to address the impacts of these disturbances on non-tree organisms. According to the review, linking DSS that focus on forest conditions to DSS that focus on wildlife has the potential to provide managers with a broad range of biodiversity indicator classes. Note, between 'non-tree organisms' game species are the main objects of biodiversity assessment and modelling in the reviewed systems. Modelling the dynamics of ground vegetation is practically absent there.

At the same time, information on ground vegetation becomes a key parameter for assessment of biodiversity and ecosystem functioning. Ground vegetation provides information about the forest types, plant species diversity, assists to estimate forest site conditions, to calculate critical loads for ecosystems, carbon and nitrogen accumulation, climate change consequences, to reconstruct forest history. A detailed review of European soil-vegetation models used to predict plant species composition as a function of atmospheric deposition of nitrogen and acidity has become recently available [2]. Dynamic soil models linked with empirical static-vegetation models (the SMART2 (-SUMO)-MOVE/NTM and MAGIC(-SUMO)-GBMOVE model chains) and a dynamic ecosystem model integrating hydrology, growth, biogeochemical cycles, and vegetation dynamics (ForSAFE-VEG) are discussed there in a view of effects of nitrogen deposition on plant species diversity. The EFIMOD-ROMUL model system [7] is another soil-vegetation dynamic model of carbon and nitrogen cycles in forest ecosystems being developed now in a frame of the decision support tool EFIMOD-DLES [1]. A part of the system is a static ground vegetation model BioCalc [5] for prognosis of ecosystem and plant species diversity based on the outputs of EFIMOD-ROMUL system.

4 BioCalc applications

BioCalc has been applied for biodiversity prognosis under different forest management regimes in several case studies located in Central European Russia [6]. Several silvicultural scenarios have been simulated: natural development without forest management, clear cutting and selective cutting regimes. Biodiversity prognosis has been also estimated under different external impacts: under different climate change scenarios and different levels of nitrogen deposition. Our results show that ground vegetation diversity can be roughly assessed on a base of standard Russian forest inventory data where only tree species composition and forest site classes are marked. Prognoses of ecosystem and plant species diversity under different forest management scenarios and external impacts have the reliable dynamics and can be applied in DSS for sustainable forest management.

This work is partly supported by «Forest management decision support system, COST Action FP0804»

References

- [1] COST Action FP0804. Available online: http://fp0804.emu.ee/wiki/index.php/Main_Page [2010, June 30]
- [2] De Vries W., Wamelink W., Van Dobben H. et al. (2010) Use of dynamic soil-vegetation models to assess impacts of nitrogen deposition on plant species composition: an overview. *Ecol. Appl.* Vol 20 pp. 60-79.
- [3] FOREST EUROPE. Biodiversity (2009). Available online: http://www.mcpfe.org/eng/What_we_work_for/Biodiversity/ [2010, June 30]
- [4] Gordon S., Johnson K., Reynolds K. et al. (2007) Decision support systems for forest biodiversity. Project A10. National commission on science and sustainable forestry. Available online: <http://ncseonline.org/ncssf/dss> [2010, June 30]
- [5] Khanina L., Bobrovsky M., Komarov A., Mikhajlov A. (2007) Modelling dynamics of forest ground vegetation diversity under different forest management regimes. *For. Ecol. Manage.* Vol. 248 pp. 80-94.
- [6] Khanina L., Bobrovsky M., Komarov A. et al. (in press) Model predictions of effects of different climate change scenarios on species diversity with or without management intervention, repeated thinning, for sites in Central European Russia. *Nitrogen Deposition, Critical Loads and Biodiversity*.
- [7] Komarov A.S., Chertov O.G., Zudin S.L. et al. (2003). EFIMOD 2 - a model of growth and cycling of elements in boreal forest ecosystems. *Ecol. Model.* Vol. 170 pp. 373-392.

Формирование видовой структуры фитопланктона в зависимости от минерального питания

Абакумов А.И.⁽¹⁾, Силкин В.А.⁽²⁾, Паутова Л.А.⁽³⁾, Пак С.Я.⁽¹⁾

Институт автоматизации и процессов управления ДВО РАН, Владивосток⁽¹⁾,
Южное Отделение Института океанологии им. П.П. Ширшова РАН, Геленджик⁽²⁾,
Институт океанологии им. П.П. Ширшова РАН, Москва⁽³⁾

abakumov@iacp.dvo.ru

vsilkin@mail.ru

larisapautova@yahoo.com
packsa@iacp.dvo.ru

Abstract

The mathematical model for dynamics of a biomass of phytoplankton in the aquatic ecosystem is presented. The structure of the dominating species is calculated.

1 Объект исследования

Для внутренних морей России характерна крайняя неустойчивость их экосистем. Нестабильность проявляется на всех трофических уровнях и, прежде всего, на уровне фитопланктона. Происходит как смена доминант среди традиционных видов, так и появление новых видов-вселенцев. Динамика структуры фитопланктона рассмотрена нами на примере Черного моря [4].

В основу положена гипотеза о том, что смена доминирующих видов является следствием изменения концентрации лимитирующих рост элементов питания или изменение их соотношения. Предлагается взять только два фактора – концентрацию азота и фосфора и с помощью модели показать, что, изменяя концентрации этих элементов, можно получить основные динамики фитопланктона, наблюдаемые в экосистеме северо-восточной части Черного моря в весенне-летне-осенний период.

Выбраны 4 вида, которые входят в лидирующие комплексы за последние 10 лет. Кокколитофориды *Emiliania huxleyi* практически всегда доминировала весной и в начале лета. Диатомеи *Pseudo-nitzschia pseudodelicatissima*, *Proboscia alata* и *Chaetoceros curvisetus* определяли биомассу фитопланктона в этот период в разные годы.

2 Математическая модель

Процесс регуляции ростовых характеристик водорослей осуществляется на двух уровнях [5]:

- скорость поглощения элемента питания i -ым видом регулируется концентрацией j -го элемента C_j в среде в соответствии с уравнением Михаэлиса-Ментен в интерпретации М.Моно [2]:

$$v_{ij}(C_j) = v_{ij}^m \frac{C_j}{K_{ij} + C_j};$$

- скорость роста i -го вида определяется содержанием j -го элемента в биомассе и выражается уравнением Друпа [1]:

$$\mu_{ij}(Q_{ij}) = \mu_{ij}^m \left(1 - \frac{q_{ij}}{Q_{ij}} \right).$$

В этих формулах: v_{ij} , v_{ij}^m – удельная и максимальная удельная скорости поглощения элемента; μ_{ij} и μ_{ij}^m – удельная и максимальная удельная скорости роста; K_{ij} – константа полунасыщения для процесса поглощения; Q_{ij} и q_{ij} – текущее и минимальное содержание элемента в биомассе. Если принять верхний продуцирующий слой как относительно изолированную систему, отрезанную от нижних слоев сезонным термоклином, можно записать нашу основную математическую модель в виде системы уравнений для биомасс N_i видов и концентрации основных элементов минерального питания C_i :

$$\begin{cases} \frac{dN_i}{dt} = \left[\min_j \mu_{ij}(Q_{ij}) - D \right] N_i \\ \frac{dC_i}{dt} = D(C_{0i} - C_i) - \sum_j v_{ij}(C_j) N_j \\ \frac{dQ_{ij}}{dt} = v_{ij}(C_j) - Q_{ij} \min_j \mu_{ij}(Q_{ij}) \end{cases}$$

В целом модель аналогична модели, описывающей поведение микроорганизмов в хемостате [3, 5]. Модель имеет конечное число локально (асимптотически) устойчивых равновесных (стационарных) решений. Основная характеристика равновесного состояния – это

структура доминирующих видов. Существенными параметрами, влияющими на набор доминирующих видов в равновесии, являются концентрации C_{0j} минеральных веществ и скорость D притока этих веществ через термоклин. Чтобы виды не вымывались, скорость притока должна быть ограничена величиной

$$\bar{D} = \min_{i,j} \frac{\mu_{ij}^m}{1 + \frac{\mu_{ij}^m}{v_{ij}^m} \cdot \frac{K_{ij} + C_{0j}}{C_{0j}}} \cdot q_{ij}.$$

Все устойчивые равновесия можно найти следующим алгоритмом. Необходимо выбрать элементы x_j столбцов матрицы $X = (x_{ij})_{i,j=1}^{m,n}$ с условием ограничения снизу выбранными элементами всех ее строк. Элементы этой матрицы вычисляются из уравнений $\tilde{\mu}_{ij}(x_{ij}) = D$, где использовано обозначение

$$\tilde{\mu}_{ij}(x) = \mu_{ij} \left(\frac{1}{D} \cdot v_{ij}(x) \right).$$

Положительные значения концентраций N_i вычисляются из подсистемы следующей системы линейных уравнений:

$$\sum_i v_{ij}(x_j) N_i = D(C_{0j} - x_j).$$

Из алгоритма, в частности, следует, что равновесное состояние сообщества не зависит от начального состояния фитоценоза.

3 Расчеты и результаты

В лабораторных экспериментах с накопительной и непрерывной культурами были оценены параметры модели.

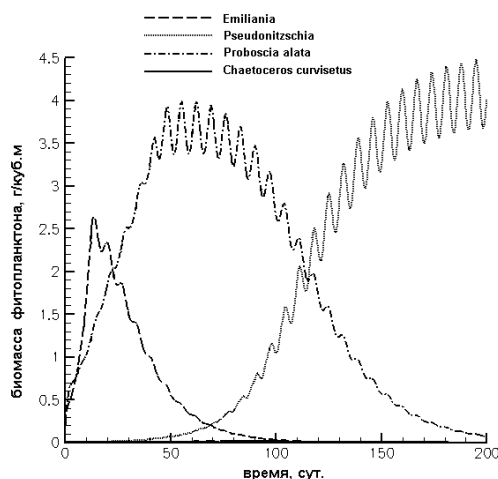


Рис.1. Динамика биомасс при $D=0,1$ (1/сут.) и периодическом изменении концентраций минеральных веществ в воде

В модельных расчетах изменяли концентрацию азота и фосфора на входе, скорость обмена, начальную концентрацию элементов в среде и начальное содержание азота и фосфора в биомассе.

Вычислительные эксперименты с моделью дали следующие результаты.

- Концентрации азота и фосфора во входном потоке оказывают существенное влияние на структуру доминирующих видов в устанавливающемся равновесии.
- Начальные концентрации азота и фосфора и начальная структура фитоценоза влияют на длительность переходных процессов, но не на финальное равновесное состояние.
- Смена доминирующих видов может происходить достаточно медленно, в течение десятков и сотен суток. Исследователь, изучающий это фитопланктонное сообщество в течение нескольких недель в природе или лабораторных условиях, может принять то или иное промежуточное состояние за финальное устойчивое равновесие.
- Неравновесная динамика вод (имитация штормовых воздействий, рис.1) влияет на структуру доминирующих видов в квазиравновесии по сравнению с равновесием в стационарной среде.

Заключение

Гидрохимический режим, как и гидрологическая структура водных масс, подвержены существенному влиянию климатических факторов. С помощью математических моделей можно рассчитать возможную структуру планктонного фитоценоза при различных трендах климатических изменений.

References

- [8] Droop M.R. The nutrient status of algal cells in continuous culture. *J. Mar. Biol. Assoc. U. K.*, volume 54, pages 825 – 855, 1974.
- [9] Monod J. The growth of bacterial cultures // *Ann. Rev. Microbiol.*, volume 111, no. 2, pages 371 – 394, 1949.
- [10] Абрсов Н.С., Боголюбов А.Г. Экологические и генетические закономерности сосуществования и коэволюции видов. Новосибирск: Наука, Сибирское отделение, 333 с., 1988.
- [11] Паутова Л.А., А.С.Микаэлян, В.А.Силкин. Структура планктонных фитоценозов шельфовых вод северо-восточной части Черного моря в период массового развития *Emiliania huxleyi* в 2002 - 2005 гг. *Океанология*, том 47, № 3, с. 408-417, 2007.
- [12] Силкин В.А., Хайлов К.М. Биоэкологические механизмы управления в аквакультуре. Л.: Наука, 230 с., 1988.

Моделирование пространственно-временной структуры лесной экосистемы: сравнение моделей конкуренции

П.Я. Грабарник

Институт физико-химических и биологических проблем почвоведения, РАН

Modelling the space-time structure of a forest ecosystem: comparison of models of competition

P. Grabarnik

Institute of Physico-Chemical and Biological Problems of Soil Science, RAS

Abstract

We analyze the spatial structure of a forest ecosystem in relation to the mode of competition. It is shown that asymmetric competition leads to a more regularized structures than interactions built on a symmetric influence.

В работе анализируется пространственная структура древостоя в зависимости от моды конкуренции. Показано, что асимметричная конкуренция приводит к более регуляризированным структурам, чем взаимодействия, построенные на симметричном влиянии

1 Введение

Изучение пространственной структуры древесного яруса растительных сообществ является важной частью изучения всего комплекса разнообразных взаимодействий элементов лесной экосистемы. Значимость пространственных взаимоотношений между деревьями обусловлена той ролью, которые они играют в процессе развития и функционирования древостоя. Математические методы, которые используются для того чтобы ответить, например, на такие вопросы: как характеризовать пространственную структуру, как учесть конкурентные взаимоотношения соседних деревьев в формировании специфического паттерна размещения деревьев, в настоящее время достаточно развиты. Многие из этих методов пришли из чисто математических дисциплин: стохастической

геометрии, теории случайных точечных процессов, математической морфологии, и сформировали основу методов пространственной статистики. Специфика методов пространственной статистики состоит в том, что, в отличие от классических методов (многомерной) статистики, данные, например, карты размещения деревьев, не могут рассматриваться как реализации независимых случайных величин и, следовательно, хорошо изученные методы здесь не работают. Поэтому для анализа пространственной структуры растительных сообществ привлекаются современные, основанные на использовании компьютеров, методы.

В данной работе анализ пространственной структуры древостоя проводился в связи с задачей различения двух типов конкуренции между растениями: симметричной и асимметричной. Вопрос, который является важным в понимании конкурентных взаимоотношений между растениями и имеет большое теоретическое значение, состоит в том, чтобы понять какие характеристики растительного сообщества могут быть чувствительны к моде конкуренции. Некоторые модели, построенные для того, чтобы ответить на этот вопрос [1], показали, что среди интегральных характеристик сообщества, например, распределения деревьев по диаметру, нет показателя, с помощью которого можно различать тип конкуренции. Задача настоящей работы состояла в том, чтобы изучить тип пространственной структуры как отклик на тот или иной характер взаимодействия между растениями.

2 Методы

В первой части работы решалась задача построения пространственно-временной модели растительного сообщества. Чтобы быть достаточно реалистичной, модель сообщества состояла из взаимодействующих пространственно распределенных дискретных элементов – деревьев, рост которых подчинялся логистическому закону. Конкуренция между растениями вводилась как локальное взаимодействие, влияющее на скорость роста растения. От того, как перераспределялся ограниченный ресурс, зависела мода конкуренции, то есть растения могли угнетать друг друга симметрично или асимметрично. Важно, что даже для такой простой системы удалось подобрать

параметры таким образом, что поведение модели довольно точно воспроизводило экспериментальные данные хода изреживания выбранного типа древостоя и усредненные характеристики древостоя.

3 Результаты

Во второй части работы анализировались пространственные структуры, возникающие при разных типах конкуренции. Для этого использовались так называемые меры второго порядка, функция Рипли и парная корреляционная функция, которые являются непараметрическими характеристиками пространственной структуры [2]. В работе были прослежены несколько временных срезов, и для каждого из них наблюдалось отличие пространственных структур в зависимости от типа конкуренции. Важный факт, найденный в процессе моделирования, который является новым в

литературе, состоит в том, что асимметричное распределение влечет более регуляризованные паттерны размещения деревьев.

Литература

- [1] Bauer S., Wyszomirski T., Berger U., Hildenbrand H., Grimm V. Asymmetric competition as a natural outcome of neighbourhood interactions among plants: results from the field-of-neighbourhood modelling approach // *Plant Ecol.*- 2004.- V. 170. – P.135–145.
- [2] B.D. Ripley. Stoyan, D., Penttinen, A., 2000. Recent applications of point process methods in forestry statistics // *Statistical Science.*-2000.-V. 15. –P. 61–

Количественные методы оценки классификаций растительности

Гречушкина Н.А.

Институт экологии Волжского бассейна РАН

grenat1976@yandex.ru

Quantitative methods for evaluation of vegetation classifications

Grechushkina N.A.

Institute of Ecology of the Volga River Basin of Russian Academy of Sciences

Abstract

Quantitative methods of an evaluation of classifications of vegetation where are diagnostic species are criterion are described. Graphic results of analyses for a data of vegetation of the Azov Sea coast of Russia are given.

Введение

Применение методов иерархической классификации растительности ставит перед исследователем две задачи: какой алгоритм лучше использовать для анализа данных, и какой уровень полученных разделений избрать.

Существуют различные методы количественной оценки качества получаемых разделений [1, 6 и др.]. Адекватными целям синтаксономии в направлении Браун-Бланке представляются методы, основанные на диагностических видах в качестве критерия классификации. Чем лучше фитоценотические группы диагностируются видами, тем проще их экологическая интерпретация. Рассмотрим два метода оценки качества классификаций по диагностическим таксонам.

Характеристика методов

Метод “оптимальной классификации” предназначен для одновременного выбора лучшего алгоритма классификации и оптимального числа групп на основе “верных” видов [3], которые в дальнейшем могут использоваться как диагностические таксоны для идентификации типов фитоценозов. Метод применяет значение P теста Фишера как меру “верности” видов к группе для данных присутствия/отсутствия. Затем измеряет количество “верных” (диагностических) видов в

кластерах анализируемого раздела. Чем больше таких таксонов, тем лучше классификация. Количество “верных” видов при этом будет зависеть от выбранного порога значения P теста Фишера. Это позволяет, например, исключить из оценки качества классификации виды с низким значением “верности” [8].

Другой метод оценки качества разделений использует расчет “четкости” классификации на основе статистической меры G . Последняя не измеряет верность видов к группам, а определяет их способность отличать кластеры в пределах данного разделения. Поэтому ее именуют “степенью разделения” видов. Значение этой степени выше среднего для всех видов анализируемого раздела (средний показатель именуется “четкостью” классификации) определяет группы, наиболее хорошо диагностируемые видами. То есть разделение с наибольшей величиной “четкости” является лучшей классификацией [2].

Оптимальный уровень разделения (количество групп) можно определить по пикам кривых на графиках “четкости” или “оптимальной классификации” [4].

Для демонстрации работы характеризуемых методов использованы 98 геоботанических описаний с данными видового обилия в процентах. Описания выполнены сотрудниками ИЭВБ РАН на российском побережье Азовского моря в 2006 г. Применяемые алгоритмы кластерного анализа и их параметры отражены в табл. 1. Анализ проведен с использованием программ JUICE 7.0 [7] и PC-ORD 5.0 [5].

Результаты

На рис. 1 показан результат сравнения различных алгоритмов кластеризации методом “оптимальной классификации”. Наибольшие количества диагностических видов с малым числом групп получены кластеризацией методом Варда с использованием данных процентов обилия (от 4 до 10 кластеров) и их квадратичным трансформированием (3 кластера). Со средним и большим количеством групп лучшие разделения получены с использованием данных процентов обилия гибким бета связыванием (мера сходства –

относительный коэффициент Сьеренсена) и UPGMA. На графике не отражены четкие пики кривых (число диагностических видов постепенно возрастает с увеличением числа кластеров). Поэтому сложно определить оптимальный уровень разделений. Можно лишь увидеть резкое возрастание количества “верных” видов.

Таблица 1
Параметры использованных алгоритмов кластерного анализа

Метод кластерного анализа	Мера сходства	Данные переменных	Краткое название алгоритма
Гибкое бета связывание	относительный коэфф. Сьеренсена	1.0; 0.5; 0.0	Flexible beta, Rel. Sorensen
Гибкое бета связывание	коэфф. Сьеренсена	1.0; 0.5; 0.0	Flexible beta, Sorensen
Невзвешенное попарное среднее	хи-квадрат	1.0; 0.5; 0.0	UPGMA, chi-squared
Метод Варда	Евклидово расстояние	1.0; 0.5; 0.0	Ward's method, Euclidean

Примечание к табл.: 1.0 – проценты видового обилия; 0.5 – квадратичное трансформирование процентов обилия; 0.0 – данные присутствия/отсутствия видов.

Для 4-х алгоритмов кластеризации определено значение “четкости” (рис. 2). Для разделения гибким бета связыванием наибольший пик отмечен для 3-х групп, что может соответствовать уровню высших синтаксонов [2], следующий меньший пик для 6-ти групп будет характерен для уровня ассоциаций. Для UPGMA высший пик соответствует разделению с 10-ю группами, меньший – с 5-ю. Метод Варда дал четкие разделения с 3-мя, 8-ю и 10-ю группами для данных процентов обилия и с 4-мя, 6-ю и 8-ю группами для данных с квадратичным трансформированием.

Закключение

Таким образом, метод “оптимальной классификации” позволяет сравнить различные алгоритмы разделения и выбрать лучший из них по количеству диагностических таксонов. Количество групп может быть выбрано по пикам кривой графика. Но если они не выражены, то целесообразнее применить расчет “четкости”, отражающий иерархию классификации, где максимальный из пиков может соответствовать более высокому иерархическому уровню.

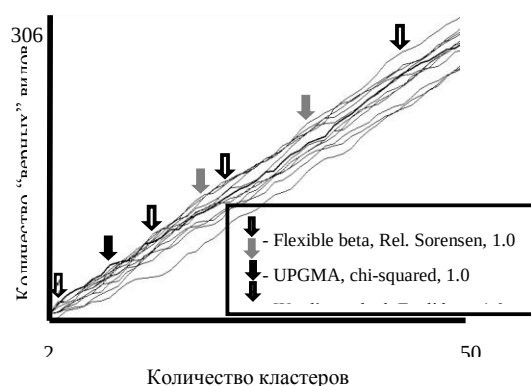


Рис. 1. Результаты анализа “оптимальной классификации”. Стрелками отмечены кривые с максимальным числом “верных” видов для данного уровня кластеризации.

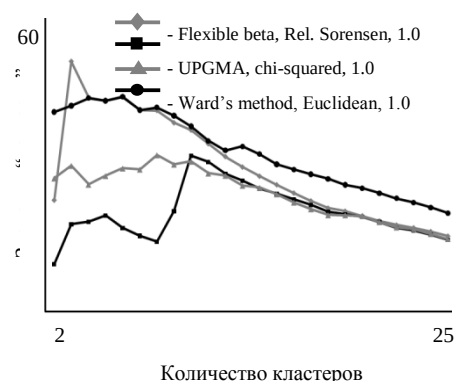


Рис. 2. Результаты расчета “четкости” классификации для 4-х алгоритмов кластеризации. Пики кривых соответствуют оптимальным уровням разделений.

Список литературы

- [1] Aho K., Roberts D.W., Weaver T. Using geometric and non-geometric internal evaluators to compare eight vegetation classification methods. *J. Veg. Sci.*, volume 19, pages 549–562, 2008.
- [2] Botta-Dukát Z., Chytrý M., Hájková P., Havlová M. Vegetation of lowland wet meadows along a climatic continentality gradient in Central Europe. *Preslia*, volume 77, pages 89–111, 2005.
- [3] Chytrý M., Tichý L., Holt J., Botta-Dukát Z. Determination of diagnostic species with statistical fidelity measures. *J. Veg. Sci.*, volume 13, pages 79–90, 2002.
- [4] Feoli E., Lausi D. Hierarchical levels in syntaxonomy based on information functions. *Vegetatio*, volume 42, pages 113–115, 1980.
- [5] McCune B., Grace J.B., Urban D.L. Analysis of Ecological Communities. *Gleneden Beach: MjM Press*, 302 p., 2002.
- [6] Mucina L., Hauser M. A new method for determining optimal number of clusters in vegetation data. *Abstracta Botanica*, volume 17, pages 147–153, 1993.
- [7] Tichý L. JUICE, software for vegetation classification. *J. Veg. Sci.*, volume 13, pages 451–453, 2002.
- [8] Tichý L., Chytrý M., Hájek M., Talbot S.S., Botta-Dukát Z. OptimClass: Using species-to-cluster fidelity to determine the optimal partition in classification of ecological communities. *J. Veg. Sci.*, volume 21, pages 287–299, 2010.

Модель FGV - моделирование живого напочвенного покрова в лесной экосистеме

Михайлов А.В.⁽¹⁾, Михайлова Н.В.⁽²⁾

Учреждение Российской академии наук Институт физико-химических и биологических проблем почвоведения РАН⁽¹⁾,

Учреждение Российской академии наук Институт математических проблем биологии РАН⁽²⁾
alexey.mikh@gmail.com

FGV model – modeling of forest ground vegetation in forest ecosystem

Mikhailov A., Mikhailova N.

Abstract

The paper presents the dynamic model of forest ground vegetation (FGV), in which the vegetation is considered as the number of functional groups. Growth of ground vegetation is depend on light condition, nitrogen availability in soil and soil moisture.

1 Введение

Модели круговорота биофильных элементов в лесных экосистемах, которые составляют больше половины суши, необходимы для оценки запасов и динамики углерода и азота, их влияния на изменения биоразнообразия и прогноза возможных изменений экосистемы при сильных внешних воздействиях.

Предлагаемая модель [1] разрабатывалась для оценки влияния абиотических факторов на развитие живого напочвенного покрова (ЖНП) в лесу и для выяснения его роли в биологическом круговороте лесной экосистемы. Модель интегрируется в модульную систему моделей DLES [2].

2 Особенности модели FGV

ЖНП представляется в виде набора функциональных групп, критериями для выделения которых являются показатели общности участия в биологическом круговороте: а) сходство биометрических показателей, б) сходное содержание основных биогенных элементов, в) одинаковость темпов развития, и г) сходная реакция на основные экологические факторы.

Модель имеет годичный шаг и позволяет исследовать динамику популяций многолетних

растений. Входными параметрами модели являются: а) освещенность, выраженная как доля света, доходящая до уровня ЖНП, б) поступающий из почвы азот, доступный растениям, в) характеристика влажности почвы, выраженная как доля вегетационного периода с переувлажнением и/или с низким увлажнением почвы. Прирост биомассы вычисляется как доля от наличной биомассы для каждой группы ЖНП в зависимости от соответствия абиотических факторов специфичным потребностям группы и лимитирующих факторов, выделенных с учетом закона Либиха.

Определение прироста биомассы и других показателей функционирования группы ЖНП

Изменение биомассы на следующий по времени шаг ΔBp вычисляем как разницу прироста биомассы I_p и опада биомассы L_p данного шага:

$$\Delta Bp = I_p - L_p. \quad (1)$$

Для определения прироста биомассы I_p выбираем минимальное значение из прироста по свету I_{p_e} и прироста по азоту I_{p_n} :

$$I_p = \min \{ I_{p_e}; I_{p_n} \}. \quad (2)$$

Прирост по свету вычисляется, используя коэффициент зависимости прироста биомассы от освещенности K_e и коэффициент зависимости прироста биомассы от влажности K_w :

$$I_{p_e} = \alpha_{\max} * B_{p_t} * K_e * K_w, \quad (3)$$

где $\alpha_{\max}$ — удельный прирост на единицу биомассы; B_{p_t} — биомасса группы ЖНП предыдущего шага по времени.

K_e , принимающий значение в интервале (0,1), определяется в зависимости от доли доступной освещенности K_{Edown} и коэффициентов A_L , B_L , определяющих реакцию группы на изменение освещенности:

$$K_e = \frac{1 - e^{-A_L(K_{Edown} - B_L)}}{1 - e^{-A_L(1 - B_L)}}. \quad (4)$$

K_w находится как произведение коэффициента зависимости от засухи K_{wd} и коэффициента зависимости от переувлажнения K_{wm} :

$$K_w = K_{wd} * K_{wm}. \quad (5)$$

Величины K_{wd} и K_{wm} принимают значения в промежутке $[0,1]$.

K_{wd} и K_{wm} определяются по формулам

$$K_{wd} = 1 - A_d * (D_{dry} - B_d)^2, \quad (6)$$

$$K_{wm} = 1 - A_m * (D_{moist} - B_m)^2, \quad (7)$$

соответственно, где A_d , B_d , A_m , B_m — коэффициенты, определяющие экологическую валентность группы по отношению к влажности, а D_{dry} и D_{moist} — доли вегетационного периода с засухой и переувлажнением.

Прирост биомассы по азоту определяем как отношение доступного для группы растений почвенного азота Nm_p к удельному потреблению азота на прирост биомассы растений данной группы Np :

$$Ip_n = Nm_p / Np. \quad (8)$$

Поступающий из почвы доступный азот, распределяется пропорционально потенциально возможному приросту по свету конкретной группы видов (см. формулу 3) и удельному потреблению азота на прирост:

$$Nm_p = Nm * \frac{Ip_e * Np}{\sum_{i=1}^{i=n} (Ip_{e_i} * Np_i)}, \quad (9)$$

где Nm — доступный для всех групп ЖНП азот; n — количество групп ЖНП на моделируемом участке.

Таким образом, группы ЖНП конкурируют между собой за доступный азот.

Количество опада для каждой группы за шаг по времени вычисляем как:

$$Lp = Bp_{tl} * b, \quad (10)$$

где Bp_{tl} — биомасса данного шага; b — доля биомассы, поступающая в опад.

При определенном наборе параметров группа видов может “исчезать”, т.е. существовать в сублетальном (квазисенильном) состоянии с минимальной биомассой.

3 Интеграция модели FGV в систему DLES

Модульная система DLES [2] — это программный продукт, ориентированный на исследовательское моделирование сложных экологических процессов с дискретным шагом по пространству и времени. DLES позволяет объединять разные модели, имеющие разный пространственный и временной масштаб в единую систему. Процесс имитации происходит на двухмерной решетке. Каждая ячейка может отличаться по набору параметров.

Используя DLES, модель FGV может интегрироваться вместе с другими моделями лесной экосистемы, так например модель разложения органического вещества почвы (почвенная модель) может быть источником такого рода данных, как содержание почвенного азота в доступных формах для минерального питания растений. В свою очередь FGV модель будет вычислять количество опада, при отмирании частей растений. Опад является одним из важных входных параметров почвенной модели. Таким образом, можем получить систему, состоящую из 2-х моделей с обратными связями. Интеграция модели FGV с пространственной моделью древесной растительности даст возможность отследить влияние древесного яруса на напочвенный покров через затенение последнего и конкуренцию деревьев с живым напочвенным покровом за элементы минерального питания.

References

- [1] Михайлов А. В. Модель динамики биомассы живого напочвенного покрова в лесу Математика. Компьютер. Образование. вып.8, 2001
- [2] Михайлов А., Шанин В., Безрукова М. Компонентный подход к построению моделей. Лесоведение, 2, с. 69-76, 2010.

Трехмерная упрощенная модель затенения деревьев в лесу с учетом видоспецифического коэффициента проницаемости кроны

Михайлова Н.В.⁽¹⁾, Михайлов А.В.⁽²⁾

Учреждение Российской академии наук Институт математических проблем биологии РАН,
г. Пущино, Россия⁽¹⁾

Учреждение Российской академии наук Институт физико-химических и биологических
проблем почвоведения, г. Пущино, Россия⁽²⁾
natalia.mikh@gmail.com

Three-dimensional simple model of trees shadowing in forest taking into account species-specific crown light interception

Mikhailova N.V.⁽¹⁾, Mikhailov A.V.⁽²⁾

Abstract

The simple three-dimension simulation model was created for calculation of tree's crown shadowing. The shadowing was calculated for particular tree in uneven-aged mixed tree stand.

1 Введение

При моделировании лесных экосистем необходимо учитывать световой режим в лесу. Конкуренцию за свет между деревьями можно имитировать различными способами: от простых коэффициентов конкуренции (Сизов, Грабарник, 2008) до сложных пространственных моделей, учитывающих 3-х мерную структуру лесного полога, траекторию движения солнца, длительность светового дня и др. параметры (Чумаченко, 1993). В рамках данной работы стояла задача создать максимально простую 3-х мерную модель конкуренции за свет, вычисляющую долю ФАР, доступную каждому дереву в условиях затенения ближайшими деревьями в древостое. Данная модель является компонентом системы EFIMOD/LES (Михайлов и др., 2009), разрабатываемой на основе предыдущей версии модели EFIMOD (Komarov et al. 2003).

2 Описание модели

В простейшем варианте модели крона дерева рассматривается как шар. Количество света, поглощаемое кроной отдельного дерева (назовем рассматриваемое в конкретный момент времени дерево фокальным), рассчитывается как сумма световых потоков пучка лучей, идущих из ряда направлений к центру кроны (рис.).

Подобный подход к расчету светового режима так же используется в модели SLIM (Gregoire, 2002).

Если некоторый луч пересекает крону другого дерева, то его световой поток уменьшается соответственно коэффициенту проницаемости кроны этого дерева. Древостой может быть представлен деревьями, принадлежащими к разным биологическим видам, коэффициент проницаемости кроны у разных видов деревьев различен. При пересечении нескольких крон одним лучом уменьшение светового потока данного луча рассчитывается как произведение коэффициентов проницаемости кроны затеняющих деревьев. Тогда для фокального дерева коэффициент затенения рассчитывается по формуле:

$$K_E = \frac{\sum (R_i * \prod \lambda_j)}{\sum R_i};$$

где K_E – коэффициент затенения, R_i – интенсивность светового потока, λ_i – доля ФАР, поглощенная кроной затеняющего дерева. В данной версии модели длина пересечения лучом кроны затеняющего дерева не учитывается.

Лучи в зависимости от зенитного угла (направления по зениту) могут иметь разную интенсивность светового потока. Так, чем ближе к зениту, тем интенсивность светового потока больше. При расчете прохождения лучей учитывается не вся модельная территория, а только деревья в радиусе воздействия, являющегося параметром модели.

По умолчанию в алгоритме используется 8 азимутальных направлений и 18 зенитных углов. Таким образом, для сканирования пространства вокруг фокального дерева рассматривается 144 луча. Однако для разных целей моделирования может понадобиться уменьшение или увеличение точности расчета. Следует понимать, что увеличение количества лучей повлечет за собой замедление работы модели, но для определенных целей это может быть необходимо.

Модель позволяет рассматривать для фокального дерева не одну точку, а совокупность точек,

расположенных на поверхности кроны, при этом форма кроны может быть любой. Так деревья, принадлежащие к разным видам, могут быть

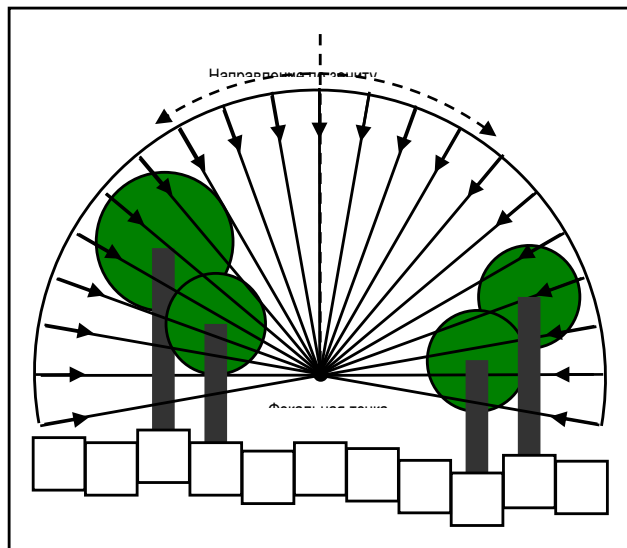
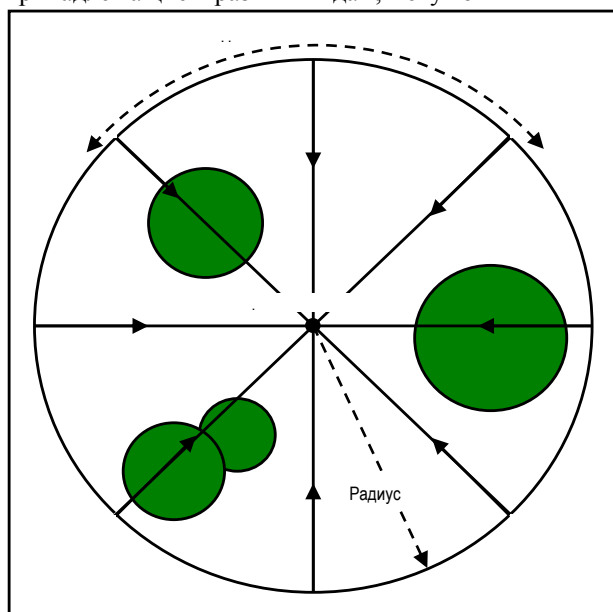


Рисунок. Схема прохождения лучей в горизонтальной и в вертикальной плоскости

аппроксимированы разными геометрическими фигурами.

Коэффициент затенения, являющийся результатом предложенной модели, может быть использован в других лесных индивидуально-ориентированных моделях для редукции потенциального прироста дерева в зависимости от затенения соседними деревьями и для расчета условий отмирания дерева.

Благодаря описанию алгоритма через понятие освещенности фокальной точки, модель позволяет так же рассчитывать затенение для напочвенного покрова с любой заданной топографией исследуемой местности.

References

- [1] Gregoire V., Degi H. SLIM software: a simple light interception model for multi species, multi-strata forests // Bois et forets des tropiques. 2002. 272 (2). P. 97-100
- [2] Komarov A.S., Chertov O.G., Zudin S.L., Nadporozhskaya M.A., Mikhailov A.V., Bykhovets S.S., Zudina E.V., Zoubkova E.V. EFIMOD 2 – A model of growth and elements cycling of boreal forest ecosystems // Ecological Modelling. 2003. 170. P. 373-392.
- [3] Михайлов А.В., Безрукова М.Г., Шанин В.Н., Михайлова Н.В. Компонентный подход к моделированию лесных экосистем // Материалы конференции «Математическое моделирование в экологии» (ЭКОМАТМОД-2009).
- [4] Сизов И.Е., Грабарник П. Сравнение индексов конкуренции с помощью имитационной модели светового режима древостоя // Принципы и способы сохранения биоразнообразия: материалы III Всероссийской научной конференции. /Отв. редактор Л.А. Жукова. Я (2008) -Йошкар-Ола, Пушкино: Марийский гос.ун-т, С.488-489
- [5] Чумаченко С.И. Базовая модель динамики многовидового разновозрастного лесного ценоза // Вопросы экологии и моделирования лесных экосистем. Науч. труды. Вып.248. М.: МГУЛ, 1993. С.147-180

Структура модели данных атласа «Биосубстраты Сибири»

Молородов Ю.И.⁽¹⁾, Федотов А.М.^(1, 2)

Институт вычислительных технологий СО РАН⁽¹⁾,
Новосибирский государственный университет⁽²⁾
yumo@ict.nsc.ru, fedotov@sbras.ru

Abstract

The model data and the structure of the atlas "biological substances in Siberia", intended for storing and processing data on the chemical composition of biological substances in the blood of representatives of the peoples of Siberia and the Far North.

1 Введение

Для оценки влияния геохимических и антропогенных факторов на состояние и качество жизни различных популяций может быть использована информация об элементном составе некоторых биосубстратов человека, животных, компонентов их пищи, растительного покрова и почв. Химические элементы, распределенные в объектах окружающей среды, являются важным природным явлением, которые отражают интенсивность глобальных и региональных биохимических циклов. В настоящее время с расширением глобальной сети экологического мониторинга загрязнения биологических и природных объектов ведется поиск биоиндикаторов, которые дают возможность адекватной оценки экологической ситуации. Разработанная модель данных, на основе которой построен атлас «Биосубстраты Сибири» позволяет объединить и структурировать накопленный материал по многоэлементному составу биосубстратов и объектов геохимической среды Сибири и других регионов и предоставить средства их математической и статистической обработок.

2 Постановка задачи

Изучение процессов взаимодействия живых организмов с окружающей средой является предметом многолетних исследований, немаловажную часть которых занимают вопросы влияния загрязнения окружающей среды на здоровье представителей животного мира и человека в частности. Одной из приоритетных задач является поиск взаимосвязи концентраций следовых элементов в объектах окружающей среды (атмосферные аэрозоли, вода) с многоэлементным составом тканей живых организмов.

В течение длительного времени в Сибирском отделении РАН ведутся комплексные исследования

по изучению свойств атмосферных аэрозолей (АА) и их роли в различных биохимических циклах биосферы. В ходе регулярных экспедиций проводились заборы образцов биосубстратов в различных районах Сибири, Хакасии, Бурятии, Среднего Урала и Крайнего Севера, с последующим измерением их многоэлементного состава путем рентгенофлуоресцентного элементного анализа.

В итоге был собран большой объем данных элементного состава объектов биогеохимической среды, в частности:

- Пробы крови и волос, плаценты, продуктов питания человека;
- Растительная масса лесного биоценоза;
- Рационы (содержимое желудков), печень, костные ткани мелких млекопитающих различных трофических уровней;
- Образцы шерсти и костных тканей крупных млекопитающих.

Для сохранения этого уникального эмпирического материала и обеспечить доступность для широкого круга специалистов нужно использовать идеи, заложенные в современных информационных технологиях. Нужно разработать электронные атлас, или информационно-вычислительную систему (ИС), позволяющую специалистам экологической и биохимической областей науки получать доступ к экспериментальным данным многоэлементного состава объектов биогеохимической среды, средствам ее обработки и анализа.

2.1 Структура данных

Исходные данные представляют собой временные ряды скалярных функций, связанные с географическими координатами места проведения забора. Они могут быть описаны набором метаданных:

- Координаты места проведения забора;
- Измеряемая величина;
- Объект измерения;
- Класс объекта измерения;
- Прибор, с помощью которого измерялась величина;
- Метод предварительной обработки данных.

Описанной структуре данных можно поставить в соответствие логическую структуру модели данных (рис.1.)

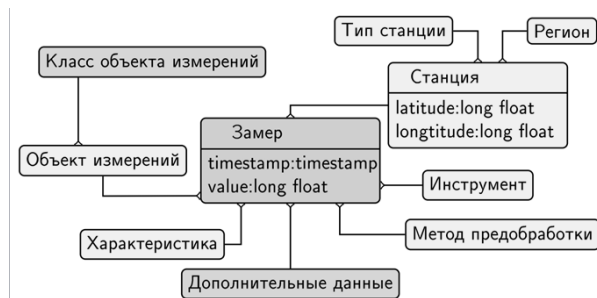


Рис.1. Структура модели данных

2.2 Модули математической обработки

Для анализа элементного состава биосубстратов человека, животных, компонентов их пищи, растительного покрова и почв необходимо разработать математические алгоритмы обработки данных на основе методов многомерного статистического анализа (регрессионный, кластерный, факторный, дискриминантный анализы) и вейвлет-анализа.

Для решения задач, связанных с исследованием взаимосвязи атмосферных аэрозолей и биосубстратов были разработаны компоненты-обработчики включающих в себя

- набор препроцессоров для:
 - сплайн-аппроксимации;
 - усреднения рядов данных различными характеристиками (математическое ожидание, дисперсия, и пр.);
- набор постпроцессоров для:
 - регрессионного анализа;
 - кластерного анализа на основе иерархической кластеризации bootstrap resampling;
 - дискриминантного анализа (линейный, квадратичный);
 - факторного анализа (PCA, Maximum Likelihood, критерий каменистой осыпи)
 - статистической проверки гипотез (критерии Стьюдента).

В качестве источников математических алгоритмов использовались:

- язык программирования **R**;
- итерфейсы взаимодействия между языками **R** и **Python** — **RPy** и **RPy2**;
- библиотеки **Numpy** и **Scipy**;
- реализации алгоритмов метода наименьших квадратов.

2.3 Результаты обработки

На рис.2 представлены образцы проверки гипотезы о влиянии атмосферного аэрозоля на состав крови ненцев. Здесь уместно заметить, что атмосферный аэрозоль представляет собой сложный конгломерат их химических элементов.

3. Заключение

На основе ORM (объектно-реляционного менеджера) SQLAlchemy спроектирована и реализована унифицированная модель данных, а на ее основе информационная Система. В рамках задачи реализованы ввод, просмотр и редактирование многоэлементного состава биосубстратов, собранных на территории Сибири.

Для полного решения задачи разработан интерфейс подключения алгоритмов математической обработки указанных данных: статистики, Вейвлет-анализ и др. Важным элементом является представление результатов обработки в табличном и графическом виде.

Работа частично финансировалась Президентской программой НШ 6068.2010.9, РФФИ (грант № 09-07-00277), и Интеграционным проектом СО РАН № 50, Проект IV.31.2.1. программы фундаментальных исследований СО РАН на 2010 - 2012 гг. Проект No.4 программы междисциплинарных интеграционных исследований СО РАН на 2009 - 2011 гг.

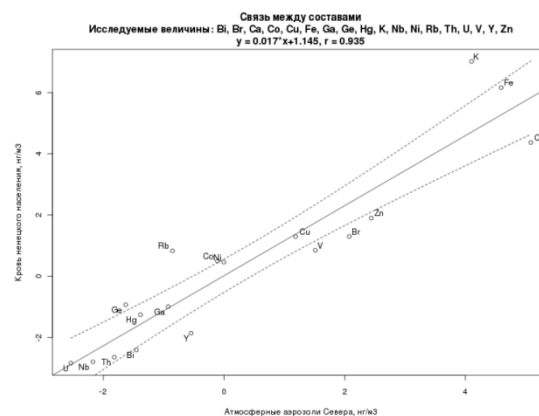


Рис.2. Соотношение аэрозолей в крови ненцев

Список литературы

- [1] Synchrotron Radiation X-Ray Fluorescence Analysis (SRXRF) for the Measuring the Multielement composition of samples of Biogenic Nature / Koutzenogii K., Savchenko T., Chankina O. et al. // Journal of Trace and Microprobe Techniques. — 2003. — Vol. 21, no. 2.
- [2] Разработка информационной системы для хранения, обработки и представления данных экологических наблюдений. П.Е. Ширшов. — НГУ, 2009.
- [3] Ю.И. Молородов, А.М. Федотов Методология интеграции данных для атласа "Атмосферные аэрозоли Сибири".- Вестник НГУ. Сер. Информационные технологии.- Т.5 Вып. 2, 2007 г. Стр. 58-66
- [4] The Comprehensive R Archive Network. <http://cran.r-project.org/>.
- [5] Murrell Paul. R Graphics Paul. R Graphics (Computer Science and Data Analysis). — Chapman & Hall/CRC, 2005. — Aug.
- [6] Ефимов В.М., Ковалева В.Ю. Многомерный анализ биологических данных // Изд-во Томского университета. — 2005. — Т. 56.

Лог-линейная модель динамики размеров гнезд и величины выводков кречета на Ямале

Романов М. С.⁽¹⁾, Мечникова С. А.⁽²⁾

Институт математических проблем биологии РАН, Пущинский государственный университет⁽¹⁾, Государственный Дарвиновский музей⁽²⁾
michael_romanov@inbox.ru

Abstract

In Yamal Peninsula (Russia), the Gyrfalcon nest mainly in trees. In recent 40 years, its average brood size decreased from 3.1 down to 2.2 chicks per active nest. At the same time gyrfalcons began to occupy smaller nests, in which chick mortality is higher. Log-linear analysis table of contingency of 3 factors—time interval, nest size, brood size ('time', 'nest', 'chicks') shows that the best model was $TN + NC$, which suggest that, probably, decrease of the brood size in recent time can be explained by nesting in smaller nests. Why the Gyrfalcon began to use smaller nests is still unclear. Possible reasons of this phenomenon are discussed.

1 Введение

Юг полуострова Ямал – один из уникальных районов нашего Севера, где уже много лет (с 1970-х гг.) ведется наблюдение за состоянием популяции и гнездовой экологией вида. В настоящей работе анализируются тенденции изменения величины выводка кречета и размеров используемых им гнезд за период 1973-2008 гг.

2 Методы и материалы

2.1 Район исследования

Исследования проводились на крайнем юге полуострова Ямал, в бассейнах рек Щучья и Хадыга-яха.

2.2 Методы исследования

Было осмотрено 63 активных гнезда и получена дополнительная информация еще о 36 гнездах, предоставленная рядом авторов в виде устных сообщений и публикаций. Всего обобщена информация по 103 активным гнездам кречета.

При находках гнезд регистрировали их размерный класс: крупные (средний диаметр более 80 см), средние (51-80 см), мелкие (50 см и менее).

Взаимодействие между временным фактором, размерами гнезд и числом птенцов в выводках кречета исследовалось с помощью метода лог-линейных моделей. В качестве меры согласия

модели с реальными данными применяли статистику G^2 (G -тест) Вилкса [2].

3 Результаты

Анализ динамики среднего размера выводка кречета (числа успешно выращенных птенцов на 1 активное гнездо) с 1973 по 2008 гг. выявил его постепенное снижение. Так, в 1973-1989 гг. среднее число птенцов на активное гнездо было 3,1 ($n = 28$), в 1990-1997 гг. — 2,8 ($n = 33$); в 1998-2008 гг. — 2,2 ($n = 22$).

Уменьшение величины выводка происходило «на фоне» перехода к использованию кречетами более мелких гнезд на деревьях. До середины 1980-х гг. они занимали почти исключительно крупные гнезда (более 80 см), построенные орланом-белохвостом. Со второй половины 1980-х гг. стали наблюдаться случаи использования соколами гнезд среднего и небольшого размера, построенных на деревьях зимняком и вороном. С 1990-х гг. птицы все чаще стали занимать средние и даже мелкие гнезда, и все реже — крупные. В 1992-2006 гг. было отмечено по крайней мере четыре случая, когда кречеты использовали даже старые гнезда серых ворон (*Corvus cornix*), что вообще совершенно не характерно для этого вида [1].

Сравнение величины выводка в гнездах разного типа и размера показывает, что наиболее стабильно и продуктивно кречеты размножаются в гнездах на скалах. Для таких гнезд, независимо от их размера и состояния, в основном характерно «нормальное» число птенцов: 3-4 ($\bar{X} = 3,3$; $n = 16$). В крупных и средних гнездах на деревьях размер выводка несколько меньше — соответственно $\bar{X} = 3,0$ ($n = 34$) и $\bar{X} = 2,5$ ($n = 20$). В мелких гнездах, расположенных на деревьях, число молодых перед вылетом не превышало двух ($\bar{X} = 1,6$; $n = 12$).

Можно предположить, что уменьшение величины выводков у кречета в многолетнем ряду наблюдений вызвано именно переходом к размножению в мелких и средних гнездах.

Для того, чтобы проверить эту гипотезу и вскрыть взаимосвязи между факторами, был выполнен лог-линейный анализ трехмерной таблицы сопряженности, в которой гнезда сгруппированы по трем признакам: период времени T , размер гнезда N , размер выводка C (от слов 'time', 'nest', 'chicks').

Число категорий каждого признака было сокращено до двух во избежание появления в ячейках таблицы нулевых значений.

В результате такой классификации получена таблица сопряженности трех дихотомических признаков, состоящая из 8 ячеек (табл. 1).

Таблица 1. Таблица сопряженности размеров выводка кречета с размером гнезда и периодом времени*

Период, T	Размер гнезда, N	Число птенцов, C	
		Мало (0-2)	Много (3-5)
I (1973-1989)	Небольшие	3	1
	Крупные	5	17
II (1990-2007)	Небольшие	18	10
	Крупные	4	8

*Примечание: В модель включены гнезда на деревьях, для которых были известны размеры и число выращенных птенцов ($n = 66$).

Иерархическая структура моделей представлена на рис. 1. Вначале была создана насыщенная модель TNC , наиболее сложная и учитывающая все возможные взаимодействия. Трехфакторное взаимодействие TNC оказалось практически незначимым ($G^2 = 0,540$; $df = 1$; $p > 0,05$), что дало основание исключить его и принять модель $TN + NC + TC$.

На следующем шаге модель $TN + NC + TC$ была подвергнута попыткам дальнейшего упрощения путем поочередного исключения каждого из двухфакторных взаимодействий. Выяснилось, что взаимодействия TN и NC высоко значимы ($G^2 = 20,241$; $df = 1$; $p < 0,001$ и $G^2 = 10,467$; $df = 1$; $p < 0,01$, соответственно); кроме того, попытки исключить их приводят к потере согласия модели с реальными данными. Взаимодействие же TC , напротив, является незначимым ($G^2 = 0,087$; $df = 1$; $p > 0,05$), что дает основание исключить его, приняв модель $TN + NC$.

Попытки дальнейшего упрощения модели $TN + NC$ путем исключения одного из взаимодействий, однако полученные модели $NC + T$ и $TN + C$ не согласовывались с реальными данными (что и ожидалось, поскольку взаимодействия TN и NC , как было показано выше, являются значимыми) и были отвергнуты.

Таким образом, наилучшей оказалась модель $TN + NC$, которая хорошо согласуется с наблюдаемыми частотами ($G^2 = 0,627$; $df = 2$; $p = 0,731$) и

предполагает наличие взаимодействий NC и TN , а также отсутствие взаимодействия TC . Это т. н. модель условной независимости, которая интерпретируется следующим образом: фактор T напрямую не связан с фактором C , однако связан с ним опосредованно, через фактор N . Следовательно, снижение успеха размножения кречета произошло именно вследствие перехода к использованию небольших по размеру (и, соответственно, менее прочных) гнезд.

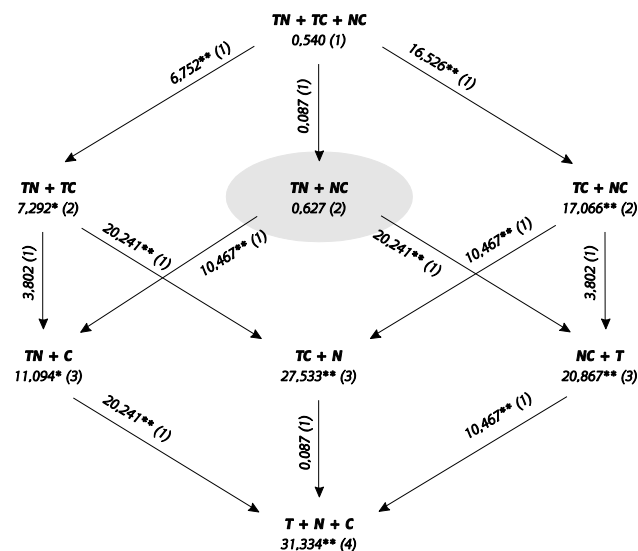


Рис. 1. Отображение иерархии лог-линейных моделей в виде графа. В узлах графа — модели, соответствующие различным сочетаниям факторов T , N и C и их взаимодействий. Цифры в узлах — значения статистики G^2 (степень отклонения ожидаемых частот от наблюдаемых), цифра в скобках — число степеней свободы. Наилучшая модель отмечена серой заливкой. Числа вдоль стрелок — разность между G^2 двух соседних моделей. Значение этой статистики характеризует значимость параметра, исключенного из модели. Числа в скобках — число степеней свободы, связанное с данным параметром. Звездочками отмечены значимые G^2 : при $p < 0,05$ (*) и $p < 0,01$ (**).

Список литературы

- [1] Potapov E., Sale R. The Gyr Falcon. Yale University Press, 2005. 288 pp.
- [2] Sokal R. R., Rolf F. J. Biometry: the principles and practice of statistics in biological research. 3-rd ed. N. Y.: W.H. Freeman and Co. 1995. 887 p.

Анализ динамики состава растительных группировок муравейников рода *Formica*

Рубашко Г.Е., Ханина Л.Г., Смирнов В.Э., Глухова Е.М.

Учреждение Российской Академии Наук Институт математических проблем биологии РАН
grubashko@rambler.ru

Abstract

Dynamics of plant species composition and structural diversity in view of *Formica* sp. activity have been studied in pine forest (*Peucedano-Pinetum*) located in Bryansk region. Analysis showed the increase of plant species diversity owing to the increase of ecological variability and meadow species incoming.

1 Введение

Согласно концепции мозаично-циклической организации экосистем, ведущим фактором динамики наземных экосистем являются природные нарушения, а живой покров представляет собой мозаику эндогенных микро-, мезо- и макросукцессий [3]. Поддержание максимального биологического разнообразия возможно, если в спонтанном режиме возникают участки нарушений разного масштаба, которые сменяют друг друга во времени и пространстве. Развитие этой концепции связано, в частности, с детальными исследованиями мозаик разного ранга, создаваемых животными и растениями. К животным, преобразующим растительный покров на внутриапарцеллярном уровне организации биогеоценозов, относятся муравьи, кроты, мелкие грызуны, кабаны. Создаваемые животными «нарушения» рассматриваются как своеобразные микроместообитания (микросайты). На них формируются своеобразные комплексы растений и животных – микрогруппировки, которым присущи динамические процессы спонтанного развития растительности – микросукцессии. Целью настоящей работы является изучение видового состава и структуры растительных группировок, обусловленных деятельностью муравьев рода *Formica*, их вклада в разнообразие лесной растительности.

2 Материал и методика

Полевой материал был собран в 2004-2005 гг. в сосняках-зеленомошниках, расположенных в задровых местностях Неруссо-Деснянского полесья [7]. Части муравейника (купол, вал) на разных стадиях существования рассматривали как отдельные микросайты; совокупность растений в микросайтах рассматривали как растительные

группировки, расположенные внутри фонового фитоценоза – сосняка-зеленомошника. Были исследованы 4 типа микросайтов: 1) купол активного муравейника, 2) земляной вал активного муравейника, 3) купол брошенного муравейника, и 4) вал брошенного муравейника. В качестве пятого (контрольного) типа микросайта рассматривался «фон» – участки сообщества сосняка-зеленомошника, на которых не наблюдалось явных признаков зоогенных изменений. Оценивалось проективное покрытие основных ярусов растительности, а также видовой состав и покрытие-обилие видов в каждом ярусе по методике Браун-Бланке. Были исследованы 21 активный муравейник, 17 брошенных муравейников, сделано 20 описаний фонового сообщества. Всего было проанализировано 96 геоботанических описаний.

Для каждого типа микросайта были определены списки доминантных и константных видов. Также рассчитаны следующие показатели: а) средние и модальные значения проективного покрытия травяно-кустарничкового и мохово-лишайникового ярусов; б) среднее и общее число видов растений; в) число и доля видов различных эколого-ценотических групп; г) число и доля видов разных жизненных форм и типов популяционного поведения. Характеристики сосудистых растений взяты из базы данных [5]. Значимость различий между микросайтами по числу видов с различными биологическими характеристиками оценивали путем проведения непараметрического дисперсионного анализа Краскела-Уоллиса [4] с последующими попарными сравнениями. Флористическое сходство пар микросайтов оценивали по коэффициенту Жаккара. Расчеты проводили по полным спискам видов растений, отмеченных в каждом микросайте. Значимость различий между микросайтами по растительности в целом с учетом баллов покрытия-обилия видов оценивали по методу MRPP [2]. Основные градиенты варьирования растительности оценивали путем ординации геоботанических площадок методом DCA [2], который проводили с удалением редких видов и уменьшением «веса» видов с низкой встречаемостью. Для интерпретации осей ординации оценивали корреляцию между координатами описаний и их экологическими характеристиками с построением на диаграмме соответствующих векторов [2]. Экологические характеристики растительности были получены как средневзвешенные баллы, рассчитанные по

соответствующим характеристикам видов растений в индикационных экологических шкалах [5]. Рассчитывали средние и диапазоны значений экологических факторов для выделенных микросайтов. Для расчетов использовались программы PC-ORD [2], EcoScaleWin [6] и SpeDiv [9].

3 Результаты

Анализ показал, что растительные группировки во всех парах исследованных микросайтов значимо различаются между собой ($p < 0,001$) за исключением пары «купол – вал» брошенных муравейников. Для этой пары наблюдается самое высокое флористическое сходство (коэффициент Жаккара 0,60). Основными экологическими градиентами варьирования растительности являются богатство почвы элементами минерального питания, увлажнение почвы и освещенность (Рис. 1).

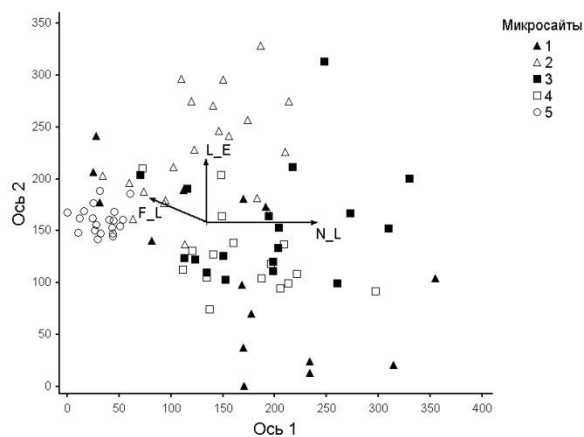


Рис. 1. Ординация DCA геоботанических площадок. F_L, L_E, N_L – векторы, отражающие корреляцию почвенного увлажнения, освещенности и богатства почвы элементами минерального питания с осями DCA. Номера микросайтов см. в тексте.

Практически по всем биологическим характеристикам видов растений фоновое сообщество и группировки растительности на куполах активных муравейников значимо отличаются от всех остальных микросайтов. Для подавляющего большинства характеристик этот результат практически очевиден, т.к. связан с тем, что в фоновом сообществе было зарегистрировано максимальное среднее число видов на площадке (17,1), а на куполах активных муравейников это число было минимальным (1,5). Однако, для двух параметров - число видов лугово-опушечной и свежелуговой эколого-ценотических групп - попарное сравнение дало неочевидный результат, подробно описанный в [8]. Анализ показал высокое варьирование экологических условий на микросайтах, обусловленных деятельностью муравьев, а также тенденцию увеличения богатства почвы элементами минерального питания на этих микросайтах по сравнению с фоновым сообществом. Показано, что появление новых микросайтов, с

отличными от фоновых экологическими условиями, приводит к увеличению бета-разнообразия растительности и увеличению флористического разнообразия территории. Из 57-ми видов, произрастающих на «мирмикогенных местообитаниях», 30 видов являлись эндемичными для них, т.е. они были встречены только на специфичных микросайтах и не были встречены в описаниях фоновых сообществ. В основном это лугово-опушечные виды: сухо- и влажнолуговые, неморально-опушечные. Такие «эндемики» составили 37% от общего списка видов исследованных сосняков-зеленомошников. Таким образом, хотя муравейники занимают относительно небольшую часть площади исследованного биогеоценоза, общее флористическое богатство микросайтов, связанных с деятельностью муравьев, сопоставимо с богатством фоновых сообществ: оно составляет 52 и 57 видов растений на муравейниках и «на фоне», соответственно.

4 Заключение

Исследование количественно показало, что появление специфических микросайтов, создаваемых муравьями рода *Formica*, способствует увеличению емкости среды обитания, увеличению бета-разнообразия территории, возрастанию ее флористического разнообразия.

Литература

- [1] McCune B., Grace J.B. *Analysis of ecological communities*. MjM Software Design. 300 pp., 2002.
- [2] McCune B., Mefford M.J. *PC-ORD. Multivariate analysis of ecological data*. Version 5.10. MjM Software Design. 237 pp., 2006.
- [3] Remmert H. The mosaic-cycle concept of ecosystems – an overview. In H. Remmert (ed.). *The mosaic-cycle concept of ecosystems*, pages 1–21, 1991.
- [4] Sokal R.R., Rohlf F.J. *Biometry: the principles and practice of statistics in biological research*. W.H. Freeman and Co. 887 pp., 1995.
- [5] База данных «Флора сосудистых растений Центральной России», 2010.
- [6] <http://www.jcbi.ru/eco1/index.shtml>
- [7] Грохлина Т.И., Ханина Л.Г. Автоматизация обработки геоботанических описаний по экологическим шкалам. В: *Принципы и способы сохранения биоразнообразия*. МАРГУ, с. 87-89, 2006.
- [8] Евстигнеев О.И. 2004. Сукцессионные процессы в растительном покрове задровых местностей. В: *Восточноевропейские леса*, том 2. Наука, с. 245-260, 2004.
- [9] Рубашко Г.Е., Ханина Л.Г., Смирнов В.Э. Анализ динамики растительных группировок муравейников *Formica rufa*. *Зоологический журнал*, в печати.
- [10] Смирнов В.Э. SPEDIV – программа для анализа разнообразия растительности В: *Принципы и способы сохранения биоразнообразия*. МАРГУ, с. 142-143, 2006.

Экспресс-диагностика концентрации мочевины в сыворотке крови с помощью ультратонкого полимерного датчика

Фомкина М.Г.^{(1)*}, Терновский В.И.⁽²⁾, Грохлина Т.И.⁽³⁾, Монтрель А.М.^(1,4)

Учреждение Российской академии наук Институт теоретической и экспериментальной биофизики РАН⁽¹⁾, Учреждение Российской академии наук Институт биофизики клетки РАН⁽²⁾, Учреждение Российской академии наук Институт математических проблем биологии РАН⁽³⁾, ГОУ Московский государственный университет им. М.В. Ломоносова⁽⁴⁾

mfomkina@mail.ru

Express-diagnosis of urea concentration in the blood serum by ultrathin polymeric sensor

Fomkina M.G.^{(1)*}, Ternovsky V.I.⁽²⁾, Grokhlina T.I.⁽³⁾, Montrel A.M.^(1,4)

Institute of Theoretical and Experimental Biophysics RAS⁽¹⁾, Institute of Cell Biophysics RAS⁽²⁾, Institute of Mathematical Problems of Biology RAS⁽³⁾, Lomonosov Moscow State University⁽⁴⁾

Abstract

There is proposed a method for determining urea in the blood serum and urine, based on potentiometric enzyme sensor, manufactured by polymer nanotechnology. We developed a program for calculating the concentration of urea in the blood serum, having its own buffering capacity. It is shown that the proposed biosensor can be used in clinical medicine for express-diagnosis of urea in biological fluids.

1 Введение

Создание биосенсорных устройств, как правило, предполагает практическую задачу. В частности, в клинической медицине необходимо проведение анализов, определяющих концентрации мочевины в моче и крови. В предыдущей работе [2] было показано, что в модельных тестовых растворах с разной ионной силой, температурой, буферной емкостью, определение концентрации мочевины

проводилось с высокой точностью. В настоящей работе предлагается экспресс-метод для диагностики мочевины в биологических жидкостях.

2 Материалы и методы

В работе использовался способ создания чувствительного слоя на основе полимерных нанотехнологий, в котором в качестве биодатчика применялся полимерный материал, состоящий из полиэлектролитных слоев между которыми были заключены микрокапсулы из этих же полиэлектролитов, содержащие функционально-активные молекулы уреазы [1,2]. Сигнал регистрировался потенциометрическим методом с помощью модифицированного рН-электрода.

Чувствительный слой наносился на шарик стеклянного рН-электрода, который служил в качестве трансдьюсера потенциометрического сенсора. Толщина полиэлектролитной оболочки, отделяющей молекулы фермента в микроячейках покрытия от исследуемого раствора, составляла менее 100 нм. Полимерная оболочка проницаема для низкомолекулярных веществ, например, для субстрата уреазы – мочевины и непроницаема для высокомолекулярных молекул фермента. Ответ электрода получали в течении 10-20 сек. Модельные тестовые растворы содержали 1 – 5 мМ Трис-цитратный буфер, рН около 6, 20 - 100 мМ NaCl или KCl.

3. Результаты и обсуждение

Проведение анализов сыворотки крови с помощью потенциометрического биосенсора мочевины затруднено, так как кровь имеет собственную буферную емкость, которая регистрируется рН-электродом наряду с измерением

концентрации мочевины. Причем, как выяснилось, у разных пациентов кровь имеет различную буферную емкость. Для повышения точности определения концентрации мочевины в сыворотке крови нами был опробован следующий метод «двойных добавок». После фиксированной добавки анализируемой жидкости в исследуемый раствор, в этот же раствор последовательно вносили 3 добавки мочевины с известной концентрацией. Все ответы рН-электрода регистрировались. Затем по следующей формуле:

$$f = \frac{ax^b}{c^b + x^b}, \text{ где } a, b, c - \text{const}$$

искали x_1 по экспериментально полученным значениям функции f в четырех точках - x_1, x_2, x_3, x_4 , где

$$x_2 = x_1 + \Delta_1$$

$$x_3 = x_1 + \Delta_2$$

$$x_4 = x_1 + \Delta_3$$

Значения Δ_i , т.е. значения добавок, были заданы.

Параметры функции a, b, c вычисляются методом последовательных приближений путем минимизации суммы отклонений результатов расчета от наблюдаемых значений в этих 4-х точках.

Время, потраченное программой на нахождение x_1 , после получения значений функции f в четырех точках, составляло несколько секунд в зависимости от заданного числа вычислений.

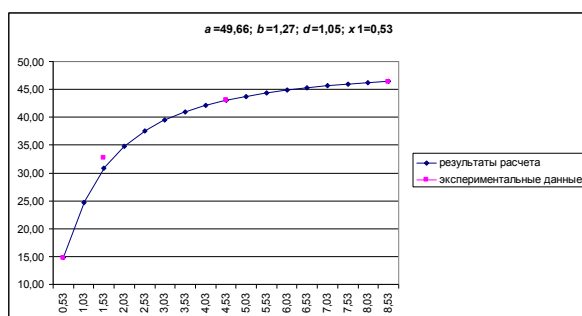


Рис.1

Результаты эксперимента				Результаты вычислений			
				x_1	Параметры кривой		
					a	b	c^b
y_1	y_2	y_3	y_4				
14,70	32,70	43,00	46,30	0,53	49,66	1,27	1,05

На рис.1 представлен график зависимости ответа электрода (мВ) от концентрации мочевины (ммоль/л) и расчетная кривая. В таблице представлены результаты экспериментальных данных и теоретических вычислений. Оработка результатов проводилась с помощью известных концентраций мочевины в растворах с различной буферной емкостью. Проверка работы биосенсора и программы расчетов проводилась на сыворотке крови с известной концентрацией мочевины, полученной стандартным спектральным методом в биохимической лаборатории поликлиники РАН г. Пушкино. Как видно из приведенных данных, с помощью этого метода можно проводить измерения концентрации мочевины в сыворотке крови. При этом достоверный результат получается при различных значениях буферной емкости крови. Предварительная калибровка электрода в этом случае не нужна. Все измерения вместе с расчетами составляют примерно 5 мин., что является вполне удовлетворительным показателем для экспресс-диагностики.

Список литературы

1. Монтрель М.М., Терновский В.И., Фомкина М.Г., Петров А.И. Патент на изобретение № 2333231, Бюлл. № 25, 2008г.
2. Терновский В.И., Чернохвостов Ю.В., Фомкина М.Г., Монтрель М.М. Биофизика – 2007. - Т.52. - №5. - С.825-829.

* Работа поддержана Министерством образования и науки Российской Федерации, государственный контракт № П 1417 от 03.09.2009 г.

Авторский указатель

Anashkina A.A.	171	Smal P.	93
Aznakayev E.G.	181	Smirnov V.E.	264
Aznakayeva D.E.	181	Sobolev O.V.	173
Baumstark M.W.	173	Stepuschenko O.O.	137
Bazykin G.A.	102	Tuhtin V.V.	214
Bogatyrev M.Y.	214	Tumanyan V.G.	171
Bukin Y.S.	107	Tuzikov A. V.	96
Bystrov V.S.	28	Usanov S. A.	96
Dekhlyar Yu	28	Yashin A.I.	194
Deriabina A.S.	32	Zubairova U.	93, 95
Deryusheva E.I.	169	Абакумов А.И.	266
Dick O.E.	60	Авилов К.К.	256
Esipova N.G.	171	Андрианов А.М.	175
Fadeev S.	95	Анищенко И.В.	175
Fridkin V.M.	28	Антосюк Е.В.	145
Furmanchuk A.	32	Арчаков А.И.	121
Glukhova E.M.	264	Атауллаханов Ф.И.	33, 34, 36
Golubyatnikov I.V.,	89	Афонников Д.А.	111
Golubyatnikov V.P.	89	Базыкин Г.А.	105
González E.	32	Балабаев Н.К.	160, 177, 183
Gorb L.	32	Беликов С.И.	189
Karkach A.S.	194	Бельская Л.В.	232, 234
Khanina L.	264	Бердникова А.П.	62
Kholkin A.	28	Благонаправов М.Л.	44
Kiselev S.S.	91	Блинов М.Н.	230
Kleshchev A.G.	89	Боброва Е.В.	69
Kliem H.	28	Бойко А.И.	246
Kolchanov N.A.	93, 95	Борисюк Р.М.	64
Kondrashov A.S.,	102	Васильчиков В.В.	100
Korotkov E.V.	109	Веселовский А.В.	121
Kuznetsov E.N.	171	Водолагина Н.Н.	238
Leszczynski J.	32	Воробьев А.И.	164
Lunin V.Y.	173	Воробьев И.И.	164
Lunina N.L.	173	Галзитская О.В.	177, 185
Machulin A. V.	169	Генаев М.А.	111
Makhortykh S.A.	220	Герус М.А.	34
Mjolsness E.	93, 95	Гильванова Е.А.	152
Mukha D. V.	96	Глухова Е.М.	280
Naumoff D.G.	135, 137, 139	Глушков В.Н.	195
Naumov G.I.	135	Глякина А.В.	177
Nikolaev S.	93, 95	Голованова О.А.	232, 234
Oparina N.U.	171	Голубович В.П.	179, 192
Ozoline O.	91	Грабарник П.Я.	268
Paramonova E.V.	28	Гребенкин А.П.	244
Poltev V.I.	32	Гречушкина Н.А.	270
Purtov Yu.A.	91	Грохлина Т.И.	282
Romanyukha A.A.	194	Гуменюк А.С.	145
Selivanova O. M.	169	Гунбин К.В.	111
Serdyuk I.N.	169	Гусев В.А.	147
Shapiro B.	93	Дедус Ф.Ф.	127, 187, 224
		Демич Ю.А.	149

Дергузов А.В.	218	Малыгина Н.А.	238
Дроздов-Тихомиров Л.Н.	100	Масина О.Н.	198
Дружинина О.В.	198	Матушкин Ю.Г.	115
Дунин-Барковский В.Л.	66	Махортых С.А.	62, 218, 222
Дюк В.А.	236	Мацкевич В.А.	164
Еременко З.С.	81	Мельниченко О.А.	254
Еремин В.Ф.	131	Мечникова С. А.	278
Ермаков Г.Л.	98	Михайлов А.В.	272, 274
Ефимов А.В.	167, 187	Михайлова Н.В.	272, 274
Журавлев Ю.Н.	200	Михальский А.И.	260
Захаров П.Н.	36	Молодцов М.И.	34, 36
Зимин Ф.А.	143	Молородов Ю.И.	216, 276
Зотов А.К.	44	Монтрель А.М.	282
Зотова Т.Ю.	44	Морозенко Е.В.	145
Иванов А.С.	121	Москаленко А.В.	55
Исаева О.Г.	240	Моттль В.В.	117, 119
Казанович Я.Б.	64, 67	Музарок Т.И.	200
Казанцев А.П.	38, 40	Мурзина Г.Б.	71
Калиновский А.А.	166	Мучник И.Б.	117, 119
Калюжный И.М.	42	Назипова Н.Н.	100, 127, 141, 143
Капустин С.И.	230	Науменко С.А.	105
Каримова О.В.	81	Наурзаков С.П.	244
Кац И.Л.	200	Неверова Г.П.	206
Каширина Н.И.	12	Ольшевец М.М.	127, 246
Кисель М.А.	175	Орлов С.В.	156
Кленкова Н.А.	230	Орлова Н.А.	164
Колобов А.Н.	202	Осипов А. А.	87
Кондрашов А.С.	103, 105	Осипов В.А.	240
Корнилов В.В.	183	Охнянская Л.Г.	57
Корнюшко В.Ф.	149	Пак С.Я.	266
Коротков А.В.	9	Панкратов А.Н.	127, 162, 187, 224, 244
Коротков Е.В.	113, 123, 129, 133	Панкратова Н.М.	242
Коршунова А.В.	34	Панюков В. В.,	87
Коршунова А.Н.	14	Паутова Л.А.	266
Костомарова И.В.	238	Перевозчиков А.Г	262
Кузин Р.А.	252	Перевозчиков А.П.	156
Кузнецова А.В.	238	Переяславец Л. Б.	191
Кулаков М.П.	204	Перцев Н. В.	208, 258
Кулаков С.В.	81	Пестунов А.И.	216
Куликова Л.И.	187, 224	Пестунов И.А.	216
Куликовский К.	117, 119	Пичугин Б. Ю.	208
Лахно В.Д.	8, 12, 14, 16, 18, 20, 23, 24, 26	Покровский А.Н.	73
Лашин С.А.	115	Полякова Г.Л.	226
Лбов Г.С.	147, 226	Пономаренко Е.А.	149
Лизунов А.В.	156	Пономарёв О.А.	24, 26
Лисица А.В.	121, 149	Поройков В.В.	121
Лихачев И.В.	160	Потапова У.В	189
Лихошвай В.А.	30	Пронин С.В.	75
Лобанов М.Ю.	185	Пугачева В.М.	123
Логинов К. К.	208	Пятков М.И.	127, 162
Лыжко Е.В.	47, 222	Разин Н.И.	117
Ляховецкий В.А.	69	Ревуцкая О.Л.	210
Мазуров М.Е.	49	Редько В.Г.	10

Реунова Г.Д.	200	Усанов Н.Н.	152
Родионов И.Н.	125	Устинин М.Н.	143, 242, 244, 246
Романов Г.А.	228	Фадеев С.И.	30
Романов М. С.	278	Федотов А.М.,	216, 276
Рубашко Г.Е.	280	Феранчук С.И.	166, 189
Руденко В.М.	129	Фиалко Н.С.	20, 23
Руднев В.Р.	187	Фигловский В.А.	179
Рыкунов С.Д.	244	Филимонов Д.А.	121
Саранча Д.А.	195	Финкельштейн А. В.	191
Семенов А.В.	125	Фомкина М.Г.	282
Сенкевич Ю.И.	236	Френкель Ф.Е.	113
Сенько О.В.	238	Фрисман Е.Я.	202
Сергеев В.Г.	262	Хакимов А.Г.	252
Сергеев Р.С.	131	Ханина Л.Г.	280
Симон К.В.	29	Харитонов А.С.	57
Скворцов В.С.	121	Хромов-Борисов Н.Н.	230
Силкин В.А.	266	Царегородцева Г.Е.	212
Смирнов В.В.	216	Цепков Г.В.	250
Смирнов В.Э.	280	Цурко В.В.	260
Соболев Б.Н.	121	Цукерман В.Д.	81
Соболев Е.В.	23, 51, 53, 77, 79	Чалей М.	141
Соколов Ю.А.	179	Чижов А.В.	83
Соловьев А.Н.	248	Чихман В.Н.	75
Суворова Ю.М.	133	Шац В.Н.	85
Сулимова В.В.	117, 119	Швец В.И.	149
Султанов В.Б.	16, 18	Шевцова В.С.	154, 248
Суслов В.В.	115	Шевцова М.С.	248
Суховеров В.С.	228	Шелепин Ю.Е.	75
Теплухин А.В.	158	Шигаев А.С.	24, 26
Теплухина Е.	141, 143	Шильников А.Л.	48
Терновский В.И.	282	Шкурят Т.П.	154
Тетуев Р.К.	127	Шмелева В.М.	230
Тихонов Д.А.	51, 53, 77, 79	Шпынов С.Н.	123
Толстоногов Д.А.	236	Штокало Д.Н.	30
Тузиков А.В.	131, 166	Шутова И.В.	179, 192
Тулуб А.В.	29	Юрьев А.С.	164
Тюльбашева Г.	141		
Угольников О.А.	149		
Улаев В.А.	145		
Усанов Н.Г.	152		