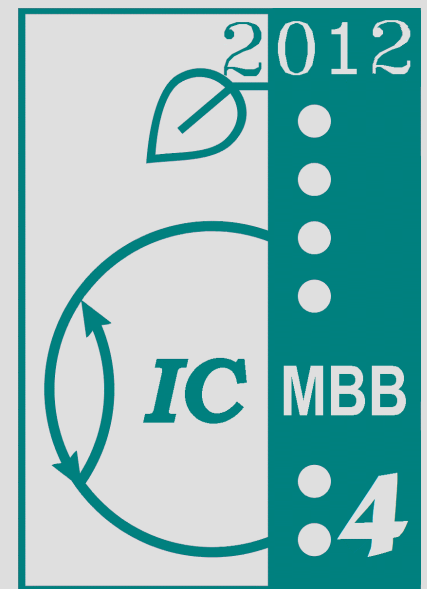
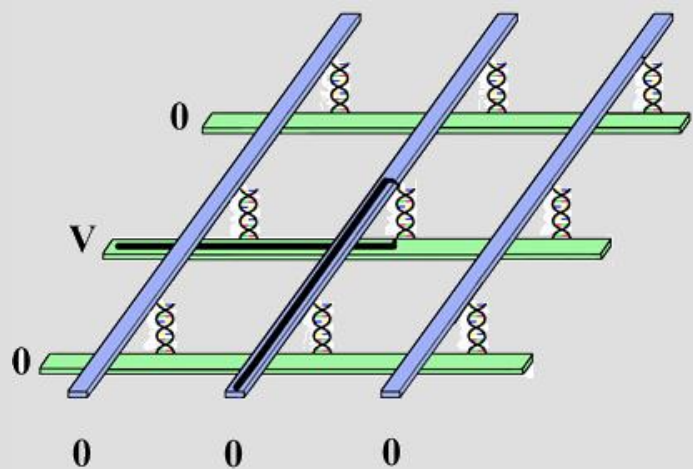


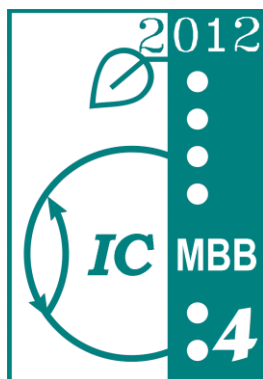


*Российская
Академия
Наук*

Математическая биология и биоинформатика



**РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ МАТЕМАТИЧЕСКИХ ПРОБЛЕМ
БИОЛОГИИ**



**МАТЕМАТИЧЕСКАЯ БИОЛОГИЯ И
БИОИНФОРМАТИКА**

ДОКЛАДЫ IV МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ

Пушино, 14-19 октября 2012 г.

Под редакцией д.ф.-м.н. В.Д.Лахно



МОСКВА - 2012

Математическая биология и биоинформатика: Доклады IV Международной конференции.

14-19 октября 2012 г., г. Пущино

В сборнике представлены доклады IV-й Международной конференции «Математическая биология и биоинформатика», проводимой Научным советом по математической биологии и биоинформатике при Отделении биологических наук РАН и Институтом математических проблем биологии РАН. При финансовой поддержке РФФИ (грант 12-07-06052).

Состав Оргкомитета

Лاخно 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 Е.В., д.б.н.
Марков А.В., д.б.н.
Миронов А.А., 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.

Состав Программного комитета

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-корреспондент
РАН - зам.председателя
Арчаков А.И., д.б.н., академик РАН
Бердышев В.И., д.ф.-м.н., академик РАН
Дегерменджи А.Г., д.ф.-м.н., академик РАН
Евтушенко Ю.Г., д.ф.-м.н., академик РАН
Жижченко А.Б., д.ф.-м.н., академик РАН
Левин В.К., д.ф.-м.н., академик РАН
Матвеев С.В., д.ф.-м.н., член-корреспондент
РАН
Рубин А.Б., д.ф.-м.н., член-корреспондент
РАН
Федотов А.М., д.ф.-м.н., член-корреспондент
РАН
Четверушкин Б.Н., д.ф.-м.н., академик 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.
Махортых Е.С.

ОГЛАВЛЕНИЕ

Предисловие

7

I. МАТЕМАТИЧЕСКИЕ ОСНОВЫ НАНОБИОЭЛЕКТРОНИКИ

<i>Лахно В.Д., Коришунуова А.Н.</i> Блоховские осцилляции полярона в холстейновской молекулярной цепочке в постоянном электрическом поле.	8
<i>Лахно В.Д., Пономарев О.А., Свиначев А.И., Шигаев А.С.</i> Вольт-амперные характеристики фрагментов ДНК.	10
<i>Лахно В.Д., Султанов В.Б.</i> Возможность биполярной высокотемпературной сверхпроводимости в Poly A / Poly T дуплексах ДНК.	12
<i>Лахно В.Д., Фиалко Н.С.</i> Возможные механизмы проводимости в однородных ДНК.	14
<i>Shirmovsky S.Eh., Boyda D.L.</i> Mathematical model of charge transfer mechanisms in DNA.	16
<i>Соболев Е.В., Тихонов Д.А., Фиалко Н.С.</i> Численное решение уравнений дискретной модели Холстейна в задаче о моделировании переноса заряда в ДНК.	18
<i>Шигаев А.С., Пономарёв О.А., Лахно В.Д.</i> Термодинамические и кинетические свойства различных типов открытого состояния ДНК.	20
<i>Bystrov V.S., Paramonova E.V., Bdikin I.K., Kopyl S., Heredia A., Kholkin A.L.</i> Bioferroelectricity in the di-phenylalanine peptide molecular nanostructures: computational modeling and nanoscale physical properties.	22
<i>Kopyl S., Bystrov V.S., Maiorov M., Maevskij E.I., Sousa C.M.A.</i> Filling of carbon nanotubes with magnetic nanoparticles for biomedical application.	24
<i>Bystrov V.S., Coutinho J., Costa E., Dekhtyar Yu., Bystrova A.V., Santos C., Almeida M., Kopyl S., Bystrova V.V., Paramonova E.V.</i> Computational study of Hydroxyapatite nanostructures, properties and its surface interactions with various species in different conditions.	26
<i>Каширина Н.И., Лахно В.Д.</i> Применение методов квантовой теории поля для описания взаимодействия заряженной частицы с фононами.	28

II. МАТЕМАТИЧЕСКАЯ БИОФИЗИКА

<i>Коваленко И.Б.</i> Компьютерное моделирование образования комплексов фотосинтетических электрон-транспортных белков.	30
<i>Голомысова А.Н., Пантелеев М.А.</i> Математическое моделирование кальциевой сигнализации в тромбоците.	32
<i>Айду Э.А.И., Трунов В.Г., Титомир Л.И.</i> Изменение векторкардиограммы при деформации миокарда.	34
<i>Быстрова В.В., Парамонова Е.В., Быстров В.С.</i> Структура и свойства льда при динамометаморфизме и его моделирование.	36
<i>Якушевич Л.В.</i> От маятника к ДНК	37
<i>Правдин С.Ф., Бердышев В.И., Панфилов А.В., Кацнельсон Л.Б., Соловьёва О.Э., Мархасин В.С.</i> Трёхмерная математическая модель электромеханической функции левого желудочка сердца человека.	39
<i>Мазуров М. Е.</i> О решении проблемы общего ритма в автоколебательной активной среде Винера-Розенблюта.	41
<i>Морнев О.А.</i> Физика и математика движения автоволн в нелинейных реакционно-диффузионных системах: от парадоксов к пониманию.	43
<i>Твердислов В.А., Попцова М.С.</i> О востребованности математической модели для знакопеременных хиральных иерархий.	45
<i>Курсанов А.Г., Викулова Н.А., Кацнельсон Л.Б., Соловьёва О.Э., Мархасин В.С.</i> Роль взаимодействия электрической и механической волн в функции миокарда (одномерная континуальная модель).	46

III. БИОИНФОРМАТИКА

<i>Кутыркин В.А., Чалей М.Б.</i> Регулярная статистическая организация кодирования в последовательностях ДНК.	48
<i>Чалей М.Б., Кутыркин В.А.</i> Профильная периодичность генов и структурно-функциональные свойства кодируемых белков.	50

Беликов С.И., Гусев В.Д., Мирошниченко Л.А., Титкова Т.Н. Сравнительный анализ геномов вируса клещевого энцефалита: дифференциация по степени вирулентности.	52
Поздниченко Н.Н., Гуменюк А.С., Шпынов С.Н. Таксономия живых организмов по характеристикам строя их ДНК.	54
Поздниченко Н.Н., Гуменюк А.С. Определение характеристик пространственной зависимости компонентов нуклеотидных цепей.	56
Голышев М.А., Коротков Е.В. Аннотирование генов методом создания функциональных групп.	58
Дюк В.А. Поиск паттернов в последовательностях на основе эффекта информационного структурного резонанса.	59
Суворова Ю.М., Коротков Е.В. Анализ распределения триплетной периодичности между генами одного генома.	61
Пугачева В.М., Коротков Е.В. Метод поиска сдвигов фазы периодичности в минисателлитах.	63
Пятков М.И., Панкратов А.Н. Спектральные методы в задаче поиска различных типов повторов в геномах.	65
Озолин О.Н., Киселев С.С., Панюков В.В. Новые элементы бактериальных геномов – промоторные «островки», как маркеры чужеродной ДНК.	66
Наумов Д.Г. Структура и эволюция альфа-L-рамнозидаз.	68
Gorbunov K.Yu., Lyubetsky V.A. Modelling co-evolution of regulatory systems, genes and species with supercomputers.	70
Matushkin Yu.G., Orlov Yu.L., Likhoshvai V.A., Levitsky V.G. Correlation between some properties of the primary structure of promoter regions and gene coding sequences of yeasts.	72
Козлов Н.Н. Расчет генетического кода.	74
Леоненко В.Н., Перцев Н.В. Применение вычислений на графических процессорах для оценки эффективности выявления индивидуумов, предрасположенных к колоректальному раку.	76
Зимин А.А., Микулинская Г.В., Нугматуллина Л.Ф., Назипова Н.Н. Сравнительный анализ аминокислотных последовательностей отдельных доменов белков Нос бактериофагов подсемейства Teequatrovirinae.	78
Назипова Н., Теплухина Е., Тюльбашева Г., Чалей М. Совершенствование автоматизированной технологии похромосомного сканирования геномов для нахождения районов скрытой периодичности.	80

IV. МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ И АНАЛИЗА БИОЛОГИЧЕСКИХ ДАННЫХ

Кузнецова А.В., Костомарова И.В., Водолагина Н.Н., Малыгина Н.А., Сенько О.В. Анализ связи биохимических и клинических показателей с возникновением нарушения мозгового кровообращения (преходящего и острого) у пациентов с хронической ишемией головного мозга.	82
Загоруйко Н.Г., Борисова И.А., Дюбанов В.В., Кутненко О.А. Использование алгоритма FRiS-GRAD для анализа активности генов при решении 9 медицинских задач.	84
Богатырев М.Ю., Вакурин В.С. Технология концептуального моделирования биомедицинских данных.	86
Зотов А.К., Зотова Т.Ю., Черномордик О.М. Анализ отличия вероятностных распределений значений параметров гемодинамики сердечно-сосудистой системы кроликов на примере их сезонных изменений.	88
Цыганкова И.А. Эволюционный метод классификации биологических объектов на основе принципа ранжирования.	90
Парамонова Е.В., Быстрова Н.К., Маевский Е.И., Кокорев Л.С., Быстров В.С. Программа обработки кардиоинтервалов для графического представления динамики показателей variability сердечного ритма.	92
Осовец С.В. Применение метода ортогональной регрессии для решения некорректных задач в биологии.	94
Сенкевич Ю.И. Лингвистическая форма представления электрофизиологических сигналов.	96
Журавлев Ю.И., Назаренко Г.И., Рязанов В.В., Константинова М.В. Дискретные методы анализа прецедентных данных и их применение в задачах неврологии.	98
Махортых С.А. Ортогональные разложения в задачах томографии и акустической экологии.	100
Быстров В.С., Камышилов В.В. Формальные операции над элементами аналитических функций, исследование областей сходимости Kml-функций от двух комплексных переменных.	102
Панкратова Н.М., Устинин М.Н. Усреднение экспериментальных данных по двум и более признакам искомого сигнала.	104
Цепков Г.В. О спектральном методе прогнозирования сейсмической активности	106

V. МОДЕЛИРОВАНИЕ МЕТАБОЛИЧЕСКИХ ПРОЦЕССОВ, ГЕННЫЕ СЕТИ И РЕГУЛЯЦИЯ КЛЕТОЧНЫХ ПРОЦЕССОВ

<i>Акбердин И.Р., Казанцев Ф.В., Ермак Т.В., Ри Н.А., Тимонов В.А., Хлебодарова Т.М., Лихошвай В.А.</i> Компьютерное моделирование динамики функционирования молекулярно-генетических систем <i>Escherichia coli</i> .	108
<i>Шилов Е.Б., Беловолова О.О., Пантелеев М.А., Атауллаханов Ф.И.</i> Регуляция метаболизма гликогена в клетках скелетных мышц.	110
<i>Пучкова А.И., Орлов М.В.</i> Моделирование оптимального распределения ресурсов микробом в процессе усвоения двух типов питательных веществ.	111
<i>Дроздов-Тихомиров Л.Н., Ковалёв П.В., Александров А.А., Уральский Л.И.</i> Программа SBS-I для исследования сетей метаболических реакций в больших полиферментных системах, представленных списком образующих их ферментов.	113
<i>Крылов А.К.</i> Модель сопряжения генерации энергии и биосинтеза в клетке.	114
<i>Зубаирова У.С., Николаев С.В.</i> Моделирование CLV-WUS регуляции структуры ниши стволовых клеток в апикальной меристеме побега.	116
<i>Ри Н.А., Хлебодарова Т.М., Когай В.В., Фадеев С.И., Лихошвай В.А.</i> Математическое моделирование метаболизма нитрита при культивировании клеток <i>Escherichia coli</i> в проточном хемостате.	117
<i>Трифонов Е.А., Лихошвай В.А., Акбердин И.Р., Гайнова И.А., Макашева В.А., Шорина А.Р.</i> Построение и анализ математической модели mTOR сигнального пути.	119

VI. ПОПУЛЯЦИОННОЕ МОДЕЛИРОВАНИЕ. МОДЕЛИРОВАНИЕ В ИММУНОЛОГИИ И ЭПИДЕМИОЛОГИИ

<i>Абакумов А.И., Израильский Ю.Г.</i> Распределение по глубине фитопланктона в океане.	121
<i>Саранча Д.А., Юрезанская Ю.С.</i> Аналитическое решение неавтономной системы дифференциальных уравнений при моделировании динамики численности леммингов.	123
<i>Саранча Д.А., Трапезов Р.В.</i> О некоторых способах построения и анализа моделей экологических систем.	125
<i>Пичугина А.Н., Перцев Н.В.</i> Интегродифференциальная модель динамики конкурирующих популяций, подверженных воздействию вредных веществ.	127
<i>Люлякин О.П., Саранча Д.А.</i> Изучение динамики численности популяции животных с использованием индивидуально-ориентированного моделирования.	129
<i>Никольский М.С., Молчанов А.А.</i> Исследование одной управляемой модели, связанной с лечением хронического миелолейкоза.	131
<i>Монтиле А.И., Давыдов О.Д., Монтиле А.А., Марчук Ю.В.</i> Стабилометрическая модель поддержания равновесия при отсутствии целенаправленного движения.	133
<i>Анонин Ю.М., Анонина Е.А.</i> Математическое моделирование эволюционных процессов в микробных популяциях на протоке.	135
<i>Абакумов А.И., Силкин В.А., Паутова Л.А.</i> Динамика и стабильность фитопланктонных сообществ.	137
<i>Перцев Н.В.</i> Исследование асимптотического поведения решений некоторых моделей эпидемических процессов.	139
<i>Лихошвай В.А., Хлебодарова Т.М.</i> Математическое моделирование клеточного цикла прокариот: согласование темпов роста объема клетки и репликации ДНК.	141

VII. МОДЕЛИРОВАНИЕ ПРОСТРАНСТВЕННЫХ СТРУКТУР МОЛЕКУЛ БИОПОЛИМЕРОВ И МЕЖМОЛЕКУЛЯРНЫХ ВЗАИМОДЕЙСТВИЙ

<i>Sobolev O.V., Lunin V.Y.</i> Analysis of atomic shifts in crystallographic unrestrained refinement can indicate alternative conformations of amino acid residues.	143
<i>Руднев В.Р., Панкратов А.Н., Куликова Л.И., Дедус Ф.Ф., Тихонов Д.А., Ефимов А.В.</i> Распознавание и анализ устойчивости структурных мотивов в глобулярных белках.	145
<i>Медведев К.Е., Афонников Д.А., Воробьев Ю.Н.</i> Исследование влияния высокого давления и высокой температуры на структуру белка Nip7 архей рода <i>Rugosoccus</i> методами молекулярной динамики.	147
<i>Куликова Л.И., Панкратов А.Н., Руднев В.Р., Дедус Ф.Ф.</i> Выбор оптимальных условий при реализации спектрального алгоритма распознавания структурных мотивов в глобулярных белках.	149
<i>Крауклис И.В., Соколова Е.В., Хамеева Н.А., Тулуб А.В.</i> Пересечение синглет-триплетных поверхностей и их возможная роль в процессе встраивания атомов в химическую связь.	151

<i>Shutava I., Junaid M., Lapins M., Wikberg J. ES.</i> Japanese encephalitis virus protease: homology modeling and structural comparisons.	153
<i>Галзитская О.В., Glyakina A.V.</i> Prediction of protein folding rate based on nucleation mechanism: How important is the size of folding nucleus?	155
<i>Гарбузинский С.А., Иванков Д.Н., Богатырёва Н.С., Финкельштейн А.В.</i> «Золотой треугольник» для скоростей сворачивания глобулярных белков.	157

VIII. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ФУНКЦИОНИРОВАНИЯ СТРУКТУР МОЗГА

<i>Ершов Н.М., Кравчук А.В.</i> Моделирование нейронных сетей с помощью стохастических клеточных автоматов.	158
<i>Покровский А.Н.</i> Три математические модели основных компонентов вызванных потенциалов коры мозга.	160
<i>Редько В.Г., Сохова З.Б.</i> Модель автономного агента-охранника.	162
<i>Лахман К.В., Бурцев М.С.</i> Анализ структурных коррелятов процессов поддержания кратковременной памяти в поведении нейросетевых агентов.	164
<i>Мурзина Г.Б.</i> Моделирование динамики мембранных рецепторов в свободном состоянии.	166
<i>Faber J., Novák M.</i> Interactions of EEG and NIRS signals in the prefrontal neo-cortex operation balancing.	168
<i>Svíttek M.</i> Quantum models for brain network.	170
<i>Votruba Z., Novák M.</i> System alliances models for prefrontal cortex operation.	172
<i>Цукерман В.Д., Еременко З.С., Кулаков С.В.</i> О фундаментальной природе пространственных представлений в мозге.	174
<i>Казанович Я.Б., Бурилко О., Борисюк Р.М.</i> Анализ устойчивости режимов синхронизации для одной системы фазовых осцилляторов.	176
<i>Лыжко Е.В., Махортых С.А.</i> Корреляционный анализ данных магнитной энцефалографии.	178
<i>Рыкунов С.Д., Сычев В.В., Устинин М.Н.</i> Система прямого моделирования экспериментальных данных магнитной энцефалографии	180

IX. ВЫЧИСЛИТЕЛЬНАЯ ЭКОЛОГИЯ

<i>Иванова Н.В., Шашков М.П.</i> Опыт создания открытой web-ориентированной информационной системы о распространении редкого лишайника лобария легочная (<i>Lobaria pulmonaria</i> (L.) Hoffm.) на территории России.	182
<i>Трофимова Г.Ю.</i> Структурные инварианты видового богатства растений дельты Аму-Дарьи.	184
<i>Романов М.С., Мастеров В.Б.</i> Матричная модель популяции белоплечего орлана на Сахалине в 2004-2011 гг.	186
<i>Безрукова М.Г., Грабарник П.Я., Смирнов В.Э., Алейников А.А.</i> Анализ пространственной организации лесного ценоза: демографический аспект.	188
<i>Голубев С.Н.</i> Фрактальный анализ горизонтальной структуры напочвенного покрова.	190
<i>Мироненко Л.М.</i> Малый биологический круговорот как динамический итог сукцессионных процессов в экосистеме: математическая модель.	192
<i>Галицкий В.В.</i> Динамика биомассы ветвей высших порядков дерева. Модельный анализ и натурные данные.	194
<i>Шалаумова Ю.В.</i> Модель представления и методы обработки данных в фотоландшафтной базе данных, используемой для изучения динамики лесотундровых сообществ Полярного Урала.	196
<i>Шашков М.П.</i> Информационная система поддержки мониторинга состояния крон деревьев международного экологического мониторинга ICP Forests на территории России.	198
<i>Галицкий В.В.</i> Виртуальные деревья, секции и вертикальное распределение биомассы дерева.	200
<i>Ханина Л.Г., Романов М.С., Бобровский М.В., Лукина Н.В., Грозовская И.С., Исаева Л.Г., Глухова Е.М., Михайлова Н.В., Рубашко Г.Е.</i> Базы данных для моделирования динамики живого напочвенного покрова лесных экосистем.	200
<i>Смирнов В.Э.</i> Статистическое оценивание некоторых параметров видового разнообразия по проблемным экологическим данным.	204
<i>Дергузов А.В., Махортых С.А.</i> Оценка вибрационного воздействия транспортных средств на экологическую обстановку в городе	206
<i>Давлетишина М.Р.</i> Концепция базы данных почв Башкортостана.	208

<i>Алфавитный указатель авторов</i>	210
-------------------------------------	-----

Математическая биология и биоинформатика

В этом издании представлены материалы Четвертой международной конференции по математической биологии и биоинформатике (ICMBB12). Конференция проводится в год 40-летия Института математических проблем биологии РАН и посвящена этой юбилейной дате.

Развитие математического моделирования, компьютерной техники и вычислительных методов сделало возможным детальное описание систем, состоящих из многих миллионов атомов, а успехи молекулярной биологии привели к накоплению большого объёма экспериментальных данных. Эти предпосылки положили начало новым научным дисциплинам — математической биологии и биоинформатике (МББ), изучающим работу живых систем, опираясь на точные сведения об их структуре. Основными прикладными задачами этих дисциплин являются компьютерное конструирование лекарств, нанобиоэлектроника и анализ индивидуальной генетической информации. В сочетании с технологией массового секвенирования генома человека математическая биология и биоинформатика составляют основу национальной программы по развитию персональной медицины.

В результате выполнения проекта "Геном человека", который был запущен в 1988 г. по инициативе американских учёных Уильяма Гилберта и Джеймса Уотсона, были определены следующие задачи МББ, которые остаются актуальными и в наше время [1]:

- распознавание белок-кодирующих участков в первичной структуре биополимеров (сравнительный анализ первичных структур биополимеров);
- расшифровка пространственной структуры биополимеров и их комплексов (рентгеноструктурный анализ, методы ядерного магнитного резонанса — ЯМР);
- моделирование структуры (3D-фолдинг) и динамики биомакромолекул;
- создание и сопровождение специализированных баз данных.

В кратком предисловии невозможно дать полный анализ всех направлений МББ. Более полное представление о математической биологии и биоинформатике можно получить из материалов первых трёх конференций, проведённых в Пущине [2 - 4]. Многообразие представленных в них направлений исследований, фундаментальных и прикладных задач позволяют сделать вывод о том, что МББ — это наука XXI столетия.

- [1] В.Д.Лахно. Математическая биология и биоинформатика. Вестник Российской академии наук, 2011, том 81, № 9, с. 812-818.
- [2] Математическая биология и биоинформатика. Доклады I Международной конференции. Пущино, 9-15 октября 2006 г. 221 с.
- [3] Математическая биология и биоинформатика. Доклады II Международной конференции. Пущино, 7-13 сентября 2008 г. 254 с.
- [4] Математическая биология и биоинформатика. Доклады III Международной конференции. Пущино, Пущино, 10-15 октября 2010 г. 286 с.

Председатель Оргкомитета Конференции, д.ф.-м.н., проф. В.Д.Лахно

Блоховские осцилляции полярона в холстейновской молекулярной цепочке в постоянном электрическом поле

Лахно В.Д., Коршунова А.Н.

Институт математических проблем биологии РАН,
142290 Пущино, Московская область, Россия
alya@impb.psn.ru

It is shown that in a constant electric field, the movement of the polaron can occur with Bloch oscillations. And Bloch oscillations are observed in the fluctuations polaron as a whole, and at the stationary motion of a polaron in the chain. In this case, low-amplitude components of the polaron demonstrate Bloch oscillations.

1 Математическая модель

Динамическое поведение полярона при наличии постоянного внешнего поля в однородной молекулярной цепочке моделируется системой связанных квантово-классических динамических уравнений с диссипацией. В используемой нами модели ДНК рассматривается как однородная цепочка, составленная из N сайтов. Каждый сайт представляет собой нуклеотидную пару, которая рассматривается как гармонический осциллятор [1]. Динамика квантовой частицы в классической цепочке описывается системой нелинейных дифференциальных уравнений, которая в безразмерных переменных имеет вид :

$$\begin{aligned} i \frac{db_n}{d\tilde{t}} &= \eta(b_{n-1} + b_{n+1}) + \kappa \omega^2 u_n b_n + E n b_n, \\ \frac{d^2 u_n}{d\tilde{t}^2} &= -\omega' \frac{du_n}{d\tilde{t}} - \tilde{\omega}^2 u_n - |b_n|^2, \end{aligned} \quad (1)$$

где b_n – амплитуды вероятности локализации заряда на n -ом сайте, ω – частота колебаний n -го сайта, ω' – коэффициент трения, u_n – смещения сайтов из равновесных положений, η – матричные элементы перехода по сайтам, E – напряжённость электрического поля, κ – константа связи, $\tilde{t} = t/\tau$, $\tau = 10^{-14}$ сек (произвольный масштаб времени).

2 Моделирование движения полярона в электрическом поле.

Для моделирования движения полярона в постоянном электрическом поле мы поместили в цепочку полярон, соответствующий стационарному решению уравнения (1) в отсутствие внешнего поля.

Центр полярона поместили на сайт цепочки с номером n_0 . Начальные значения $b_n(0)$ выбрали в виде растянутого или нерастянутого обратного гиперболического косинуса:

$$|b_n(0)| = \frac{\sqrt{2}}{4} \sqrt{\frac{\kappa}{\xi|\eta|}} \operatorname{ch}^{-1} \left(\frac{\kappa(n - n_0)}{4\xi|\eta|} \right). \quad (2)$$

Начальные значения x_n^0 и y_n^0 ($b_n = x_n + i y_n$) берутся не гладкими функциями n вида:

$$x_n^0 = |b_n(0)|(-1)^n / \sqrt{2}, \quad y_n^0 = |b_n(0)|(-1)^{n+1} / \sqrt{2}.$$

Длина цепочки выбирается такой, чтобы полярон не оказался слишком близко к концу цепочки до окончания вычислений. Значение n_0 (центр начального обратного гиперболического косинуса) в (2) выбирается так, чтобы на начальном этапе вычислений полярон находился достаточно далеко от концов цепочки. Поле включается “мгновенно” в начальный момент времени.

Выбраны следующие значения параметров: $\eta=2.4$, $\kappa=4$. При таких параметрах цепочки устоявшееся и аналитическое распределение заряда (см. [1]) располагаются на достаточно большом количестве сайтов. Также при этих параметрах существует равномерное движение полярона для некоторого интервала напряжённости электрического поля $0 < E < E_{\max}$ см. [2,3]. Для $E > E_{\max}$ равномерного движения заряда не наблюдается (см. Рис.1), начальная форма заряда “разваливается”, заряд, осциллируя, движется по цепочке по направлению поля. В последнем случае осцилляции полярона соответствуют блоховским и сохраняются таковыми достаточно продолжительное время.

На Рис.1 представлены графики функции $X(\tilde{t})$, описывающей поведение центра масс частицы, $X(\tilde{t}) = \sum_n |b_n(\tilde{t})|^2 n$, для различных значений напряжённости электрического поля E . На Рис.1 графики, соответствующие равномерному движению полярона, представляют собой прямые линии. Блоховские осцилляции на Рис.1 наблюдаются только при достаточно больших значениях напряжённости электрического поля E , когда начальная форма заряда “разваливается”.

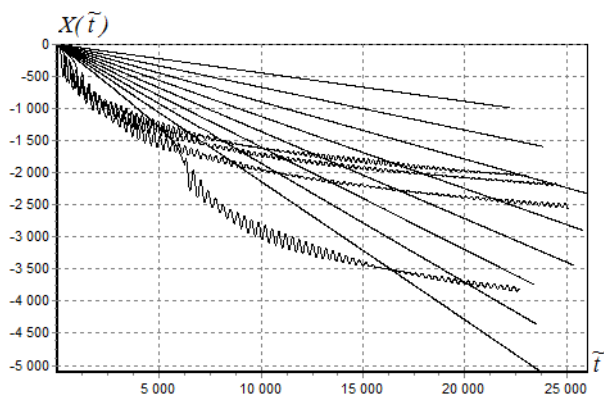


Рис.1. Графики функции $X(\tilde{t})$ для $E=0.004$ (верхний график), 0.006, ... 0.018 – равномерное движение полярона, далее для $E=0.02, 0.022, 0.024, 0.026$ – движение неравномерное. Значения параметров $\omega=1, \omega'=1, \xi=1$, длина цепочки 10001.

Но если рассмотреть малоамплитудные составляющие волновой функции электрона при равномерном движении заряда по цепочке, то и в этом случае наблюдаются блоховские осцилляции, хотя заряд сохраняет свою форму при движении. При этом перед собственно поляроном, сохраняющим при движении форму близкой к (2), наблюдается малоамплитудное возбуждение. Амплитуда колебаний сайтов в этом возбуждении зависит от напряжённости электрического поля E .

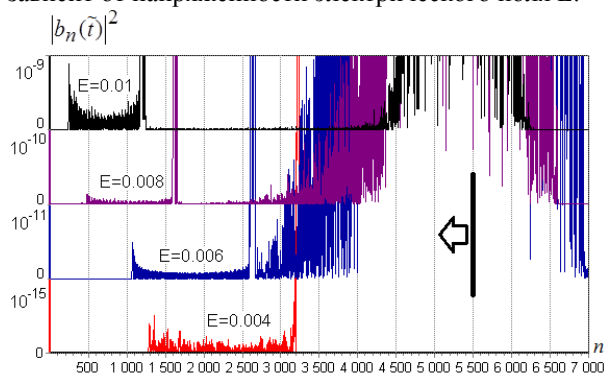


Рис.2. Графики функции $|b_n(\tilde{t})|^2$ для различных значений напряжённости электрического поля E в момент времени $\tilde{t} \approx 40000$, $n_0 = 5500$.

Ширина распределения по сайтам этого возбуждения так же зависит от E и примерно равна максимальной амплитуде блоховских осцилляций для конкретной величины E . На Рис.2 представлены графики функций $|b_n(t)|^2$ для различных значений напряжённости электрического поля в разных масштабах для каждого значения E . Максимальная амплитуда блоховских колебаний $A_{BL} = 4\eta/E$. Для $E = 0.004, 0.006, 0.008, 0.01$ $A_{BL} = 2400, 1600, 1200,$

960 соответственно. При равномерном движении полярона для выбранных выше значений параметров $\omega=1, \omega'=1$ графики функций $X(\tilde{t})$ выглядят в данном масштабе прямыми линиями, и блоховских осцилляций не видно. Но если взять значения $\omega=0.01, \omega'=0.006$ то можно наблюдать блоховские осцилляции полярона с периодом $T_{BL} = 2\pi/E$ и различными (максимальными и не максимальными – в зависимости от выбранных параметров) амплитудами функций $X(\tilde{t})$ при равномерном движении – см. Рис.3.

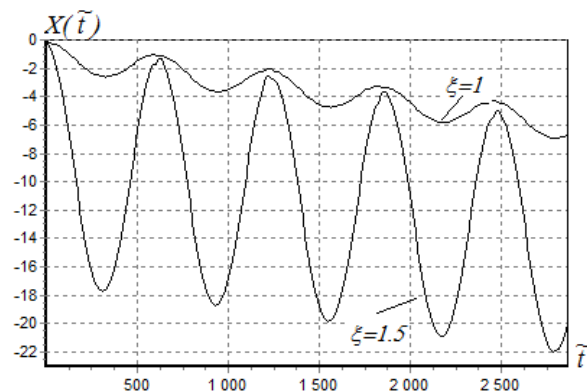


Рис.3. Графики функции $X(\tilde{t})$ для $E=0.01, \omega=0.01, \omega'=0.006$. $T_{BL} = 628$.

3 Выводы

При выбранных нами начальных значениях параметров частица (в целом) движется в направлении поля. При “мгновенном” включении внешнего постоянного электрического поля на начальном этапе движения динамическое поведение полярона может представлять собой равномерное движение по цепочке с сохранением формы или колебательное движение с непродолжительным сохранением формы или сразу с потерей формы (в зависимости от начальных условий) с дальнейшим распределением заряда по цепочке. Во всех случаях наблюдаются блоховские осцилляции.

Работа выполнена при поддержке РФФИ, проекты № 11-07-12054 и № 10-07-00112.

Литература

- [5] Лахно В.Д., Коршунова А.Н.. Математическая биология и биоинформатика, 2010, 5, №1, 1-29.
- [6] V. D. Lakhno, Int. J. Quant. Chem., 2010, **110**, 127-137.
- [7] V. D. Lakhno, A. N. Korshunova, 2011, *Eur. Phys. J. B*, **79**, 147-151.

Вольт-амперные характеристики фрагментов ДНК

Лахно В.Д., Пономарев О.А., Свиначев А.И., Шигаев А.С.

Институт математических проблем биологии РАН
olegpon36@mail.ru

We report theoretical calculation of electrical transport through GAGG fragment DNA molecules. The result suggest that electrical transport through DNA molecules occurs by polaron hopping. The volt-ampere curves calculate at various field on DNA trapped between two metal electrodes.

Появление методов атомно-силовой микроскопии (АСМ) и сканирующей зондовой микроскопии (СЗМ) позволило продвинуться в разработке эффективных чипов ДНК. Можно использовать «Solver P47» и контроллеры производства ЗАО «NT-MDT», (г. Зеленоград, Россия). В качестве подложки можно взять свежие сколы слюды, пирографит, графен, кремний и т.п. Однако несмотря на это результаты измерения коэффициентов переноса в ДНК, выполненные в различных научных группах, не согласуются между собой и причина этого пока не понята. Выяснение причин этого важно, так как молекула ДНК считается перспективным кандидатом для молекулярной электроники из-за благоприятного соотношения параметра «длина/диаметр», возможностью синтеза и двухнитевой структурой позволяющей самосборку.

Целью данного сообщения является изложение результатов изучения электропроводности олигонуклеотидов методом СЗМ и определение их вольт - амперной характеристики (ВАХ). Изучение электропроводности биомолекул чрезвычайно важно, так как может внести вклад в развитие амперометрических биосенсоров, необходимых в нанобиотехнологиях и в легкой промышленности.

Нами исследован фрагмент ДНК вида GAGG с параметрами (в эВ)

1.24	0.09	1.69	0.05	1.24	0.08	1.24
G ₁₁	----	A ₂₁	----	G ₃₁	----	G ₄₁
I		I		I		I
I 0.05		I		I		I
I		0.03		0.05		0.05
		I		I		I
C ₁₂	----	T ₂₂	----	C ₃₂	----	C ₄₂
1,90	0.10	1.90	0.08	1.90	0.04	1.90

Горизонтальные и вертикальные связи имеют резонансные интегралы порядка 0.1-0.03 эВ, наклонные (не нарисованы) – 0.00-0.02 и ими мы пренебрегаем. Цифры около букв это потенциалы

окисления соответствующих нуклеотидов, а около связей являются их энергиями. Расчеты показали, что ВАХ сильно нелинейны по полю.

Для двухнитевой ДНК с металлическими контактами СЗМ с G₁₁ и G₄₁ имеем гамильтониан (n=1,2; m = 1, 2, 3, 4) вида

$$H = \sum_{m,n} \varepsilon_{mn}(t) a_{mn}^+ a_{mn} + \sum_{m,n,gx,gy} B_{mn;m+gx;n+gy} a_{mn}^+ a_{m+gx;n+gy} + \sum_q \omega_q b_q^+ b_q + \sum_E E a_E^+ a_E + \sum_E C_{1E} (a_{11}^+ a_E + a_E^+ a_{11}) + \sum_E C_{4E} (a_{41}^+ a_E + a_E^+ a_{41}) + \sum_m B_{m1;m2} (a_{m1}^+ a_{m2} + a_{m2}^+ a_{m1}),$$

где

$$\varepsilon_{mn}(t) = \varepsilon_{mn} + \alpha_{mn} \frac{V}{6d} (m-2.5) \cos \Omega t + \sum_q \omega_q A_{mnq} (b_q + b_q^+)$$

и выполнено условие

$$\sum_{m,n} a_{mn}^+ a_{mn} + \sum_E a_E^+ a_E = 1. \text{ Величина } d - \text{ расстояние}$$

между сайтами, V - электрический потенциал между концами фрагмента. Этот гамильтониан приводит к уравнениям движения вида

$$i \frac{da_{11}^+}{dt} = -\varepsilon_{11}(t) a_{11}^+ - B_{11;21} a_{21}^+ - \sum_E C_{1E} a_E^+ - B_{11;12} a_{12}^+$$

$$i \frac{da_E^+}{dt} = -E a_E^+ - C_{1E} a_{11}^+ - C_{4E} a_{41}^+$$

$$i \frac{da_{41}^+}{dt} = -\varepsilon_{41}(t) a_{41}^+ - B_{41;31} a_{31}^+ - \sum_E C_{4E} a_E^+ - B_{41;42} a_{42}^+$$

$$i \frac{db_q^+}{dt} = -\omega_q b_q^+ - \sum_{mn} \omega_q A_{mnq} a_{mn}^+ a_{mn}$$

Дифференциальная проводимость равна $C = dI / dV$ (тогда ток $I = \int C(V) dV$) и может быть получена по формуле Кубо в приближении Ландауэра – Буттикера [1, 2]: слева это будет

$$C_1 \approx \text{Re} \int_0^\beta d\lambda \int_0^\infty dt \int dEC_{1E}^2 \langle a_E^+ (-i\lambda) a_{11} (-i\lambda) a_{11}^+(t) a_E(t) \rangle, \text{ а}$$

справа

$$C_4 \approx \text{Re} \int_0^\beta d\lambda \int_0^\infty dt \int dEC_{4E}^2 \langle a_E^+ (-i\lambda) a_{41} (-i\lambda) a_{41}^+(t) a_E(t) \rangle.$$

Выберем нулевое значение потенциала на середине фрагмента. Тогда металл слева будет иметь дополнительную энергию $(-eV/2)$, а справа $(eV/2)$. Сайт m=1 и m=2 имеют соответственно потенциалы -0.25 V и $-1/12 \text{ V}$, а сайты 3 и 4 - $(1/12)\text{V}$ и 0.25 V .

Вычисляя дифференциальную проводимость, получим интегрированием ее по потенциалу ток и ВАХ. ВАХ несимметрична относительно смены

направления тока и характеризует концевые сайты фрагмента.

Если дырка перемещается слева направо, то ее перемещение описывается КФ $\langle a_E^+ a_{11} | a_{11}^+(t) a_E(t) \rangle$.

Это означает, что дырка (отсутствия электрона) из металла перемещается во фрагмент, если ее нет у фрагмента ДНК. Перемещаясь по фрагменту, она выходит с правой стороны и уходит в металл, если ее нет для данной энергии в металле. Обратное явление описывается КФ вида $\langle a_{41}^+ a_E | a_E^+(t) a_{41}(t) \rangle$.

Для указанных КФ имеем уравнения (с учетом малости резонансных интегралов для поляронов):

$$i \frac{d}{dt} \langle a_E^+ a_{11} | a_{11}^+(t) a_E(t) \rangle = (E - \varepsilon_{11}(t)) \langle a_E^+ a_{11} | a_{11}^+(t) a_E(t) \rangle$$

$$i \frac{d}{dt} \langle a_{41}^+ a_E | a_E^+(t) a_{41}(t) \rangle = (\varepsilon_{41}(t) - E) \langle a_{41}^+ a_E | a_E^+(t) a_{41}(t) \rangle$$

Решения этих уравнений имеют вид

$$\langle a_E^+ a_{11} | a_{11}^+(t) a_E(t) \rangle = \langle a_E^+ a_E \rangle (1 - \langle a_{11}^+ a_{11} \rangle) \exp \{ i(\varepsilon_{11} - \omega A_{11}^2 - \frac{eV}{4} - E + \frac{eV}{2})t - A_{11}^2(2N+1) + A_{11}^2(N+1)e^{i\omega t} + A_{11}^2 N e^{-i\omega t} \},$$

$$\langle a_{41}^+ a_E | a_E^+(t) a_{41}(t) \rangle = \langle a_{41}^+ a_{41} \rangle (1 - \langle a_E^+ a_E \rangle) \exp \{ i(E - \omega A_{41}^2 + \frac{eV}{4} - \varepsilon_{41} + \frac{eV}{2})t - A_{41}^2(2N+1) + A_{41}^2(N+1)e^{i\omega t} + A_{41}^2 N e^{-i\omega t} \}.$$

Вычисляя все интегралы в выражениях для дифференциальной проводимости, получим [3]

$$C_1 \approx \beta \exp(-A_{11}^2(2N+1)) \sum_n \left(\frac{N}{N+1} \right)^{\frac{n}{2}} I_n(2A_{11}^2 \sqrt{N(N+1)}) \left(\frac{\pi}{2} - \arctg \left(\frac{\varepsilon_F - 0.25eV - \varepsilon_{11} + n\omega + A_{11}^2\omega}{\gamma} \right) \right),$$

$$C_4 \approx \beta \exp(-A_{41}^2(2N+1)) \sum_n \left(\frac{N}{N+1} \right)^{\frac{n}{2}} I_n(2A_{41}^2 \sqrt{N(N+1)}) \left(-\frac{\pi}{2} + \arctg \left(\frac{\varepsilon_F + 0.25eV - \varepsilon_{41} + n\omega + A_{41}^2\omega}{\gamma} \right) \right).$$

Величина C_1 имеет «плечо» при $eV = 4(\varepsilon_F - \varepsilon_{11} + n\omega + A_{11}^2\omega)$, а величина C_4 - при $eV = 4(-\varepsilon_F + \varepsilon_{41} - n\omega - A_{41}^2\omega)$.

Из этих выражений видно, что происходит значительное замедление зарядового обмена вдоль фрагмента ДНК из-за сильного фононного взаимодействия. Для системы $G_1A_2G_3G_4$ выберем безразмерные параметры взаимодействия дырок с колебаниями в виде $A_1 = 3$, $A_2 = 8$, $A_3 = 11$, $A_4 = 2$ и частоты колебаний $\omega = 0.019B$. Сравнение теоретических значений заселенности

при этих параметрах и экспериментальных значений для заселенности из [4,5] при $T=300K$ показало хорошее согласие их между собой (рис.1).

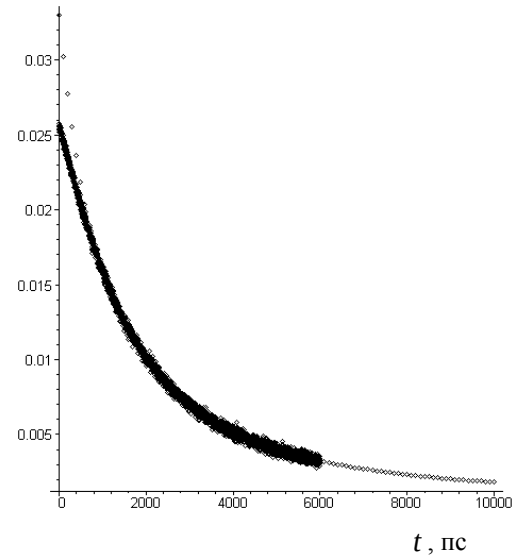


Рис. 1. Теоретическая и экспериментальная (размытая) зависимость заселенности 11 узла от времени для $T=300K$.

Это свидетельствует о разумности выбранных параметров. Следует заметить, что как и для всякой обратной задачи удовлетворительный выбор параметров может быть не единственным.

Используя параметры и приведенные выше формулы, можно рассчитать вольт - амперные характеристики фрагмента GAGG ДНК для металлических электродов.

Список литературы

- [1] Y. Meir, N. Wingreen. Phys. Rev. Lett., 1992, 68, 2512.
- [2] A. Jauho, N. Wingreen, Y. Meir. Phys. Rev. B, 1994, 50, 5528.
- [3] С. И. Кубарев, А.П. Кирьянов и др. Спин-селективные процессы в наноструктурах и методы определения их параметров. М.: МГУДТ, 2009.
- [4] F.D. Lewis, J. Liu, X. Liu et al. Angew. Chem. Int. Ed., 2002, 41, 1026.
- [5] F.D. Lewis, X. Liu, J. Liu et al. J. Am. Chem. Soc., 2000, 122, 12037.

Возможность биполярной высокотемпературной сверхпроводимости в Poly A / Poly T дуплексах ДНК

В. Д. Лакно, В. Б. Султанов

Институт математических проблем биологии РАН

lak@impb.psn.ru

Расчет биполяронов на однородных ДНК-дуплексах по модели Холстейна-Хаббарда приводит к очень стабильному биполярону на Poly A / Poly T дуплексе с энергией связи -0.13 эВ, в котором дырки локализованы на нуклеотидах А и Т среднего сайта дуплекса. Из свободных энергий биполярона и полярона получена оценка температуры сверхпроводящего перехода $T_c \approx 350\text{K}$. Предложен способ конструирования сверхпроводящих нанопроводов на основе ДНК.

Введение

Возможность сверхпроводимости органических молекул рассматривалась в середине прошлого века [1,2]. В применении к ДНК сверхпроводимость на основе теории БКШ рассматривалась в [3,4]. Хотя сверхпроводимость в органических молекулах была обнаружена давно, в ДНК сверхпроводящее состояние наблюдалось совсем недавно при температуре $T \approx 1\text{K}$ в хаотических λ -ДНК [5,6]. Этот факт означает, что носителями тока могут быть дырки в биполяронном состоянии. При описании сверхпроводимости в ДНК мы используем модель, в которой связанные состояния, или биполяроны, образуют пары полярона с противоположными спинами частиц и это ведет к Бозе конденсации в реальном пространстве [7-9].

Ранее [10,11] мы предложили такую модель с нуклеотидами в качестве осцилляторов.

В хаотических цепочках бозе-конденсация биполяронного газа затруднена неоднородностью нуклеотидного состава, что может являться причиной наблюдаемой низкой температуры сверхпроводящего перехода. В однородных полинуклеотидных дуплексах биполяроны могут обладать большей стабильностью. В данной работе мы рассмотрим условия образования стабильных биполяронов в синтетических Poly G / Poly C и Poly A / Poly T дуплексах ДНК с парами Уотсона-Крика в качестве независимых осцилляторов.

1 Модель Холстейна-Хаббарда

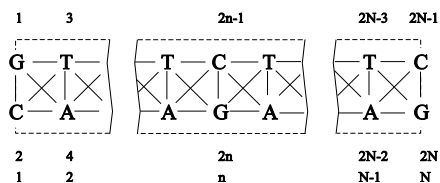


Рис. 1. Дуплекс ДНК.

В показанной на Рис. 1 нумерации нуклеотидов $j=1, \dots, 2N$ и сайтов $n=1, \dots, N$ гамильтониан модели [12]

$$\hat{H} = \sum_{(i,j),\sigma} \eta_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \sum_n \hbar \omega \left(a_n^\dagger a_n + \frac{1}{2} \right) + \sum_n g \hat{n}_n (a_n^\dagger + a_n) + \sum_j U \hat{n}_{j\uparrow} \hat{n}_{j\downarrow}$$

Электронный гамильтониан $\hat{H}_e = (\eta_{ij})_{i,j=1}^{2N}$ – блочно-тридиагональная матрица с блоками 2-го порядка.



Рис. 2. Переходы η_{ij} дырки в эВ [11].

В адиабатическом приближении $G = \frac{1}{4}(\hbar \omega / 2g)^4 \ll 1$ получаем функционалы энергии

$$\Phi_p(\vec{\psi}) = \sum_j (\hat{H}_e \vec{\psi})_j \psi_j - \frac{\kappa}{4} \sum_n (|\psi_{2n-1}|^2 + |\psi_{2n}|^2)^2,$$

$$\Phi_{bp}(\vec{\psi}) = \sum_{i,j} (\hat{H}_e \vec{\psi}_i + \hat{H}_e \vec{\psi}_j)_{ij} \psi_{ij} + U \sum_i |\psi_{ii}|^2 - \frac{\kappa}{2} \sum_{i,j} (\rho_n + \rho_m) |\psi_{ij}|^2$$

$$\rho_n = \frac{1}{2} \sum_k (|\psi_{2n-1,k}|^2 + |\psi_{k,2n-1}|^2 + |\psi_{2n,k}|^2 + |\psi_{k,2n}|^2), \quad \kappa = \frac{4g^2}{\hbar \omega},$$

[13], гамильтониан $\hat{H}_e$ действует на 1-ю и 2-ю координаты волновой функции $\vec{\psi} = \{\psi_{ij}\}_{i,j=1}^{2N}$. Для ДНК $\kappa=0.5267$ эВ, $U \approx 1$ эВ [11]. Биполяроны и поляроны рассчитывались минимизацией функционалов $E_{bp} = \text{Min}[\Phi_{bp}(\psi_{ij})]$ и $E_p = \text{Min}[\Phi_p(\psi_i)]$ с условиями симметрии и нормировки, в области параметров κ , U , содержащей указанные их значения. Условие стабильности биполяронных состояний определяется неравенством:

$$E_{bp} < 2E_p \quad (1)$$

На Рис. 3 показана граница области стабильности биполяронов на дуплексе G/C. Вне области стабильности биполярон распадается на два одиночных полярона, находящихся на большом расстоянии друг от друга. Как видно из Рис. 3, значения параметров $\kappa=0.5267$ эВ и $U=0.75$ эВ лежат в области стабильности. Энергия связи биполярона равна $\Delta = E_{bp} - 2E_p \approx -1.5 \cdot 10^{-4}$ эВ. Биполярон локализован на двух гуанинах G_{N-2} и G_{N+2} , соседних со средним сайтом дуплекса.

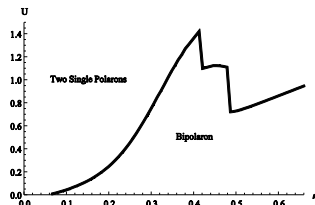


Рис. 3. Область стабильности на дуплексе G/C.

На A/T дуплексе биполяроны оказываются очень стабильными: неравенство (1) выполняется во всей рассматриваемой области параметров κ , U . При $\kappa=0.5267$ эВ и $U=1$ эВ энергия связи биполярона равна $\Delta = E_{bp} - 2E_p \approx -0.13$ эВ. Соответствующий биполярон локализован на сайте A_N , T_{N+1} в середине дуплекса – Рис. 4 ($p_j = 2 \sum_i |\psi_{ij}|^2$, $N=21$).

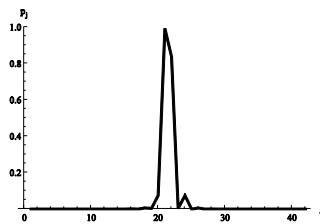


Рис. 4. Плотность заряда p_j на дуплексе A/T.

2 Биполяроны при температуре $T \neq 0$

При температуре $T > 0$ в качестве условия стабильности биполярона примем неравенство

$$F_{bp}(T) < 2F_p(T) \quad (2)$$

где $F(T) = -T \ln Z(T)$ – свободная энергия биполярона и полярона, рассчитанная по полуклассической статистической сумме

$$Z(T) = \int \dots \int \prod_n d\beta_n \left(\sum_j e^{-\lambda_j/T} \right) e^{-\sum_n \frac{K}{2} \beta_n^2/T},$$

где $K = 0.062$ эВ/Å² [14], $\lambda_j = \lambda_j(\beta_1, \dots, \beta_N)$ – собственные значения гамильтонианов соответственно биполярона и полярона при заданных смещениях осцилляторов $\{\beta_n\}$. Интеграл вычисляется методом Монте-Карло как среднее подынтегральной функции по равномерной на гиперкубе $[-R, R]^N$ случайной выборке точек $\{\beta_n\}$. Для каждой точки $\{\beta_n\}$ решается линейная задача на собственные значения, $2N$ -мерная для полярона и $(2N)^2$ -мерная для биполярона. На Рис. 5 представлен расчет $F_{bp}(T)$ и $2F_p(T)$ для дуплексов G/C и A/T, $N=5$, $R=5$, выборка $\approx 10^4$ точек, $U=0.75$ эВ для дуплекса G/C и $U=1$ эВ для дуплекса A/T. Из Рис. 5 видно, что для дуплекса G/C неравенство (2) нарушается при $T_c \approx 0$, для дуплекса A/T неравенство (2) нарушается при $T_c \approx 0.03$ эВ ≈ 350 К.

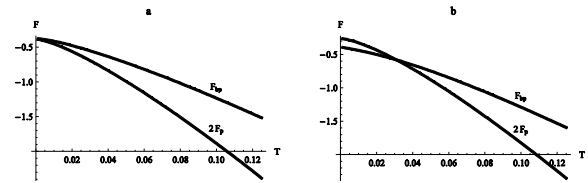


Рис. 5. Энергия связи а: дуплекс G/C б: дуплекс A/T.

При $T \rightarrow 0$ величина энергии связи $\Delta(0) = F_{bp}(0) - 2F_p(0) = E_{bp} - 2E_p$ согласуется с рассчитанной выше (см. п. 1) по энергии основного состояния биполярона и полярона $\Delta = -1.5 \cdot 10^{-4}$ эВ для дуплекса G/C и $\Delta = -0.13$ эВ для дуплекса A/T.

3 Экспериментальная проверка теории

Рассмотренная нами модель биполяронных состояний в ДНК близка к теоретической модели Андерсона [15], в которой электроны в паре локализованы на соседних сайтах (в случае Poly A / Poly T дуплекса – на аденине А и тимине Т одной из Уотсон-Криковских пар). В этом случае кулоновское отталкивание оказывается сильно ослабленным, что приводит к образованию очень стабильного биполярона, который распадается при температуре $T_c \approx 350$ К, соответствующей температуре сверхпроводящего перехода.

Высокотемпературная сверхпроводимость в Poly A / Poly T дуплексе открывает возможность создания сверхпроводящих нанопроводов электролитическим осаждением атомов на поверхность дуплекса так, чтобы атомы служили как акцепторная примесь для р-дуплексов и как донорная примесь для п-дуплексов, по аналогии с р- и п- типами полупроводников.

Работа выполнена при поддержке РФФИ, проекты № 11-07-12054-офи-М-2011, 10-07-00112.

Литература

- [1] F. London, Superfluids, N.Y., 1950, 1.
- [2] W.A. Little, Phys. Rev. 134, A1416, 1964.
- [3] J. Ladik, G. Biczio, J. Redly, Phys. Rev. 188, 1969
- [4] J. Ladik, A. Bierman, Phys. Lett., 29A, 636, 1969.
- [5] A. Y. Kasumov et al, Science, 291, 280, 2001.
- [6] A.D. Chepelianskii et al, N. J. of Phys., 13 2011.
- [7] A.S. Alexandrov, J. Ranninger, Phys. Rev. B, 24, 1164, 1981.
- [8] A.S. Alexandrov, JETP Lett., 46 (S1), S107, 1987.
- [9] A.S. Alexandrov, J. Ranninger, S. Robaszkiewicz, Phys. Rev. B, 33, 4526, 1986.
- [10] V.D. Lakhno, V.B. Sultanov, Biophysics, 56, 2011.
- [11] V.D. Lakhno, V.B. Sultanov, Chem Phys Lett, 503, 2011.
- [12] L. Proville, S. Aubry, Physica D 113, 307, 1998.
- [13] В.Д. Лахно, В.Б. Султанов, в сб. “Современные проблемы математического моделирования”, Абрау-Дюрсо, 2011.
- [14] V.D. Lakhno, Int. J. Quant. Chem. 110, 127, 2010.
- [15] P.W. Anderson, Science, 316, 1705, 2007.

Возможные механизмы проводимости в однородных ДНК

Лакно В.Д., Фиалко Н.С.

Федеральное государственное бюджетное учреждение науки Институт математических проблем биологии Российской академии наук

lak@impb.psn.ru, fialka@impb.psn.ru

1 Введение

Возможные механизмы проводимости в однородных ДНК – зонный и поляронный. В настоящее время сложилось представление, что носителями тока в ДНК являются поляроны или солитоны. Этому убеждению в значительной мере способствует тот факт, что ДНК является квазиодномерной системой, в которой избыточные заряды – электроны или дырки – в случае их сильного взаимодействия с колебательными степенями свободы молекулы переходят в поляронные состояния.

В экспериментах Yoo et al. [1] были измерены вольт-амперные характеристики в однородных полинуклеотидах при разной температуре T , и на основе этих данных получена температурная зависимость проводимости $\sigma(E)$. Эта зависимость хорошо описывается в рамках прыжковой модели полярона малого радиуса (ПМР).

Мы провели сравнение с экспериментом, не обращаясь к теоретически полученным для ПМР температурным зависимостям подвижности $\mu(T)$. Зависимость $\mu(T)$ в рассматриваемом нами подходе рассчитана в результате непосредственного численного моделирования, основанного на холстейновской модели для нуклеотидной цепочки.

2 Модель

В простейшем случае, когда заряд движется по одной из нитей ДНК (т.е. в пренебрежении перехода между комплементарными нитями), его движение вдоль однородной цепочки моделируется уравнениями [2]:

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

$$M\ddot{u}_n + \omega^2 u_n - \alpha' |b_n|^2 + \gamma \dot{u}_n + A_n(t) = 0, \quad (1б)$$

где b_n – амплитуда вероятности нахождения заряда на n -ом сайте, α' – константа взаимодействия заряда со смещением вдоль водородной связи n -ой уотсон-криковской пары, $\omega = \tau\sqrt{K/M}$, M – эффективная масса комплементарной пары, τ – характерное время колебаний. Член с трением $\gamma = \tilde{\gamma}\tau/M$ ($\tilde{\gamma}$ – коэффициент трения) и случайная сила $A_n(t)$ со свойствами $\langle A_n(t) \rangle = 0$, $\langle A_n(t) A_k(t') \rangle \sim T\gamma \delta_{kn} \delta(t-t')$ моделируют окружающую среду с заданной температурой T (уравнение Ланжевена). Длина цепочки N сайтов, $n = 1, \dots, N$.

Система уравнений (1) описывает диффузию заряда по цепочке. Коэффициент диффузии D находится из соотношения $\langle X^2(t) \rangle = 2Dt$, где $\langle X^2(t) \rangle = \left\langle \sum_{n=-[N/2]}^{[N/2]} |b_n(t)|^2 n^2 \right\rangle$ – среднеквадратичное смещение, рассчитанное по множеству реализаций. Зная D , можно оценить подвижность заряда μ по формуле

$$\mu = eD/k_B T \quad (2)$$

при температуре T , e – заряд электрона.

При моделировании динамики заряда в отдельных реализациях мы рассматривали два случая начальных данных.

В случае 1) в начальный момент времени ($t = 0$) смещения и скорости сайтов соответствуют равновесному распределению для заданной температуры, а заряд локализован на сайте в центре цепочки. Такие начальные условия можно сопоставить с экспериментами, в которых избыточный заряд при $t = 0$ создается на одном из сайтов цепочки.

В случае 2) для начальных данных предварительно было рассчитано поляронное распределение заряда и смещений в цепочке без случайной силы (при $T=0$). Это стационарное состояние с наименьшей энергией, при котором скорости сайтов нулевые. При расчете отдельных реализаций в момент $t = 0$ распределение заряда по сайтам поляронное, смещения сайтов равны сумме поляронных и случайных величин по равновесному распределению для заданной температуры, скорости сайтов – по равновесному распределению для заданной температуры.

В случае 2) при расчете динамики для больших температур поляронное состояние разрушается довольно быстро, и D одинаково для случаев 1) и 2) (см. рис. 1). При понижении T «время жизни» полярона растет, причем нелинейно, и при дальнейшем понижении температуры поляроны не разрушаются во всех реализациях на всем времени счета. Вопрос – это состояние равновесное или квазиравновесное – сложный и отдельный; тем не менее мы можем рассчитанный по этим данным коэффициент диффузии считать оценкой D снизу.

3 Результаты

На рис.1 приведены результаты расчетов k_D для polyA/polyT цепочки.

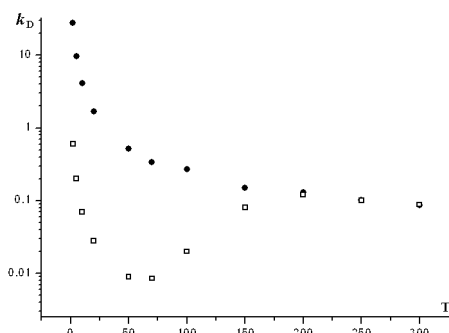


Рис.1. Рассчитанные коэффициенты диффузии дырки для polyA/polyT цепочки в зависимости от температуры. Черные кружки – из начальных условий 1), квадраты – из начальных условий 2). Ось ординат в логарифмическом масштабе.

Из рис. 1 видно, что при задании «поляронных» начальных условий 2) кривая зависимости $k_D(T)$ имеет типичный холстейновский вид. Из рис. 1 также следует, что за все время численного эксперимента заряд в случае 1) не переходит в поляронное состояние. Т.е. кривая 1) соответствует температурной зависимости подвижности зонного электрона.

Полученные результаты можно применить для вычисления проводимости σ polyA/polyT цепочки по формуле

$$\sigma = en\mu, \quad (3)$$

где μ рассчитано по формуле (2). Согласно [1], носителями тока в polyA/polyT являются электроны, концентрация n которых

$$n = n_0 \exp(-\tilde{E}_a / k_B T) \quad (4)$$

задается наличием доноров электронов с концентрацией n_0 . На рис. 2 показано сравнение рассчитанной и экспериментальной [1] температурных зависимостей относительных проводимостей σ/σ_0 (мы выбрали $\sigma_0 = \sigma(T=300K)$). Примерно одинаковое согласие с экспериментом Yoo et al. [1] для высоких температур ($T > 100K$) получается при выборе энергии активации донорных уровней $\tilde{E}_a \approx 0.220$ эВ для случая 2), т.е. «поляронного» начального распределения, и $\tilde{E}_a \approx 0.225$ эВ для случая 1), т.е. в случае зонного механизма.

Отметим, что при начальных условиях 1) коэффициент диффузии зависит от температуры как $D \sim T^{-1.2}$, при низких температурах ($T < 50K$) для начальных данных 2) зависимость такая же. Т.е. $\mu \sim T^{-2.2}$, главный вклад в величину σ (3) дает экспонента (4), и расчетная σ стремится к нулю при убывании T гораздо быстрее, чем найденная экспериментально σ_{Yoo} . Соответственно, для низких температур $T < 100 K$ расхождение σ и σ_{Yoo} порядка самой величины σ_{Yoo} (рис. 2, врезка).

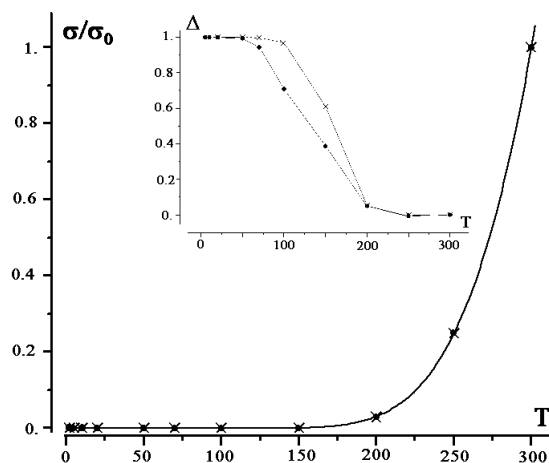


Рис.2. Сравнение аппроксимации экспериментальной зависимости (непрерывная линия) σ_{Yoo} ($E_a=0.18$ эВ) [1] с рассчитанными значениями относительных проводимостей (кружки – из начальных данных (1), $\tilde{E}_a \approx 0.225$ эВ, крестики – из начальных данных (2), $\tilde{E}_a \approx 0.220$). Врезка – относительная ошибка аппроксимации $\Delta = |\sigma_{Yoo} - \sigma|/\sigma_{Yoo}$

Такая большая ошибка при низких T может быть связана с вопросом, откуда берутся носители тока в ДНК, т.е. насколько применимы формулы (3), (4) для оценки проводимости. В эксперименте [1] носители тока инжектируются в молекулу с металлических электродов. Так как сама по себе молекула ДНК является диэлектриком, то эффективный размер области, в которую могут быть инжектированы электроны из металлических электродов в ДНК составляет несколько нанометров, в то время как расстояние переноса составляет ~ 100 нм.

Таким образом, можно сделать вывод, что экспериментальные данные Yoo et al. [1] не позволяют сделать однозначного вывода о механизме проводимости в однородных полинуклеотидных цепочках.

Работа выполнена при поддержке РФФИ, гранты 10-07-00112, 11-07-00635 и 11-07-12054-офи-м

Список литературы

- [1] К.-Н. Yoo, D.H. Ha, J.-O. Lee et al. *Phys. Rev. Lett*, 2001, **87** (19), 198102.
- [2] Лахно В.Д., Фиалко Н.С. Моделирование процессов переноса заряда в ДНК // В книге "Динамические модели процессов в клетках и субклеточных структурах" под ред. Ризниченко Г.Ю. и Рубина А.Б., М.-Ижевск: НИЦ "Регулярная и хаотическая динамика", ISBN 978-5-93972-783-9, 2010, с.11-67

Mathematical model of charge transfer mechanisms in DNA

Sergey Eh. Shirmovsky^{1,2}, Denis L. Boyda¹

¹Far Eastern Federal University,
8 Sukhanov St., Vladivostok, 690950, Russia
shirmovsky@sns.dvfu.ru

²Physics, mathematics cathedra, Vladivostok State Medical University, 2 Ostryakov St.,
Vladivostok, Russia

The problem of charge transfer mechanisms in DNA is quite topical nowadays. It is primarily connected with the fact that a charged particle's impact on the polynucleotide strand triggers mechanisms which play an important part in organism's life activity. Thus, it is now evident that many human diseases are caused by mutations, which are either inherited or acquired by human beings during their lives [1]. It can be assumed that one of the initial reasons for acquired mutations may be a charge impact on DNA: the charge acquired by a DNA molecule as a result of such interaction can migrate through the bases of the chain and cause a DNA molecule lesions resulting in the mutations. Thus, a problem of DNA conducting properties is put in the forefront.

We investigate the charge (hole) - DNA interaction on the basis of a nonlinear dynamic model [2]. In our case the fact that a DNA molecule is formed by two polynucleotide strands is taken into consideration. Stacking interaction through either DNA strand is also accounted for, and the hydrogen bonds are determined [3]. Since point mutations of gene p53 are considered to be one possible cause of cancer in human beings, we modeled the point pre-mutation effect (gene's lesions) in the DNA sequence TGTGGGTTG of gene p53 (14402-14500 position) as an example [4]. Figure 1 shows that in our case the hole was on the first guanine G (second bottom graph) at the initial moment of time (PROBABILITY=1). Then, the probability of its location on the guanine drops considerably: the charge migrates from the first guanine to the second, the third and the fourth one, oscillating between them and bypassing the thymine.

The hole localization in the guanine should be followed by local point-like DNA damage. Some assumptions about the nature of the damage can be made if we consider the dynamics of $\text{Shift} = y_{1,n} - y_{2,n} + l$ of the sequence TGTGGGTTG base pairs, where $y_{1,n}$ and $y_{2,n}$ are coordinates of the bases located on two different DNA strands, l - horizontal distance between the bases in equilibrium. We proceed from the assumption that $y_{1,n} < 0$ and $y_{2,n} > 0$. The case $\text{Shift} = 0$ Å corresponds to the base equilibrium at a distance of $l = 7$ Å, the difference $y_{1,n} - y_{2,n} = -7$ Å. A deviation from equilibrium ($\text{Shift} \neq 0$ Å) should result in local DNA disruption.

Figure 2 shows time-dependent Shift of the sequence pairs. Consider the pair G - C (line 2) on which the

charge was localized at the initial moment of time, and the pairs G-C (lines 4-6), on the bases of which the charge is localized in the course of transfer. It is assumed that the first G - C pair is displaced horizontally from equilibrium at the initial moment of time. We assumed that the bases of the first G - C pair had the coordinates $y_{1,n} = -3.20$ Å and $y_{2,n} = 4.00$ Å. In other words, the bases of the pair, on one of which a charge was located, were at a distance 7.2 Å from each other at the initial moment of time. In this case the first base pair G - C is shifted by - 0.2 Å from equilibrium ($\text{Shift} = -0.2$ Å while the shift is equal to 0 for all the rest base pairs of the strand). In about between 5.40×10^{-13} s and 2.46×10^{-12} s the charge migrates from the first guanine G, as a result of which the bases of the first pair G - C shift from equilibrium by -2.9 Å. At the following moments of time, the process stabilizes at a $\text{Shift} = -1.35$ Å with the distance between the bases of the pair being 8.35 Å, suggesting a possible hydrogen bond G - C disruption. Leaving the first guanine G, the charge is localized on the GGG unit (lines 4-6). The process of localization is followed by a considerable deviation - of over - 4 Å on the average - from equilibrium. Thus, the hole localization on the guanines results in a considerable shifting of the pair bases, which makes the disruption of G - C hydrogen bonds of the base pairs highly probable. In virtue of stacking bonds, the shift of the pairs under consideration should be followed by that of their neighboring pairs. Fig. 2 shows that the first pair G - C shift results - through a vertical bond - in the T - A pair shifting (line 1). The first pair T - A shift is less significant. The third pair T - A (line 3) of the strand shifts more noticeably - $\text{Shift} = -2.2$ Å - due to an additional vertical bond with the bases of the fourth pair G - C. The - 2 Å shift of the seventh pair T - A (line 7) is considerable enough at the expense of a similar bond. The shift of the last two pairs T - A (line 8) and G - C (line 9) is less noticeable and decreases as it recedes from the charge localization site. A considerable shift difference between the pairs with a charge localized on them (4-6 G-C pairs) and the neighboring third (line 3) and seventh (line 7) pairs T - A suggests some possible single- or double- (one- or two-strand) disruptions between them. Developing vertical disruptions are bound to result in pre-mutation state, i. e. GGG unit dropping out of the strand under consideration.

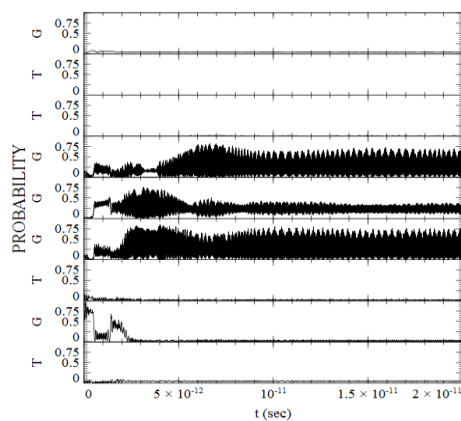


Fig. 1.

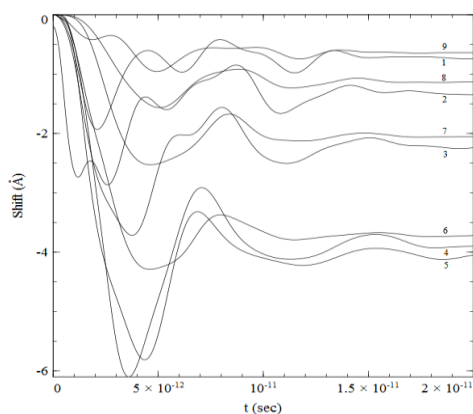


Fig. 2.

References

- [1] <http://www.ncbi.nlm.nih.gov>
- [2] N.S. Fialko, V.D. Lakhno, Phys. Lett. **A 278**, 108 (2000).
- [3] O. V. Belov, D. L. Boyda, S.Eh. Shirmovsky, 12th Int. Workshop on Radiation Damage to DNA, 2nd-7th June, Prague, 43 (2012).
- [4] <http://www.ncbi.nlm.nih.gov/nucleotide/56157802?>

Численное решение уравнений дискретной модели Холстейна в задаче о моделировании переноса заряда в ДНК

Соболев Е.В., Тихонов Д.А., Фиалко Н.С.

Институт математических проблем биологии РАН
egor@impb.psn.ru

Введение

В модели Холстейна распространение заряда вдоль цепочки сайтов и движения классических сайтов цепочки описывается самосогласованной системой ОДУ [3]. Для учета термостата при заданной температуре движение классических сайтов описывается уравнением Ланжевена. В безразмерных величинах система записывается как

$$\begin{aligned} i\dot{b}_n &= \eta_{n-1,n}b_{n-1} + \eta_{n,n}b_n + \eta_{n,n+1}b_{n+1} + \chi u_n b_n, \\ \dot{u}_n &= v_n, \end{aligned} \quad (1)$$

$$\dot{v}_n = -\omega_0^2 u_n - \gamma v_n - \chi |b_n|^2 + Z_n(t),$$

где $Z_n(t)$ – случайная гауссова величина.

Чтобы найти средние величины, необходимо рассчитать множество траекторий системы. Например, для вычисления обобщенных частотных спектров необходимо усреднять по миллионам реализаций. Поэтому требуется эффективная численная схема.

1 Численная схема

Особенность системы (1) заключается в том, что характерные времена ее квантовой и классической части отличаются более чем на два порядка. Для решения квантовой части в работе об адиабатическом приближении модели Холстейна [4] использовался метод матричной экспоненты. В адиабатическом приближении классические уравнения не зависят от квантовых переменных $\chi |b_n|^2 \equiv 0$. Поэтому их интегрировали полностью отдельно. Затем с помощью экспоненты матрицы интегрировали квантовую часть с шагом по времени h , значительно превышающим ее характерное время.

$$\begin{aligned} \mathbf{b}^{(k+1)} &= \exp(-hi\mathbf{A}^{(k)})\mathbf{b}^{(k)}, \\ \mathbf{A}^{(k)} &= \left\{ a_{nm}^{(k)} \right\} = \mathbf{N} + \chi \left\{ \delta_{nm} \left(u_n^{(k)} + \frac{1}{2} h v_n^{(k)} \right) \right\}, \end{aligned} \quad (2)$$

где $\mathbf{N}$ – трехдиагональная матрица с элементами η_{nm} , δ_{nm} – символ Кронекера.

Мы развили изложенный подход и разработали согласованную схему для решения полной системы (1), в которой учитывается влияние ее квантовой части на классическую часть, с шагом по времени, достаточным для интегрирования классической

части системы. Действительно, разностная схема требует, чтобы на выбранном шаге интегрирования классические переменные изменялись медленно. Следовательно, можно использовать экспоненту матрицы, чтобы выразить решение квантовой части на этом же шаге и получить новое значение правой части классических уравнений.

Интегрировать классическую часть системы (1) будем двухшаговым методом второго порядка с одной гауссовой величиной [2]. Для этого введем две вспомогательные переменные:

$$g_n^{(k)} = f(u_n^{(k)}, v_n^{(k)} + \sqrt{h}Z_n^{(k)}, b_n^{(k)}),$$

$$p_n^{(k)} = f(u_n^{(k)} + h u_n^{(k)}, v_n^{(k)} + g_n^{(k)}, b_n^{(k+1)}),$$

где

$$f(u, v, b) = -\omega_0^2 u - \gamma v - \chi |b|^2.$$

Тогда значение классических переменных на следующем шаге будет:

$$v_n^{(k+1)} = v_n^{(k)} + \frac{1}{2} h (g_n^{(k)} + p_n^{(k)}) + \sqrt{h} Z_n^{(k)},$$

$$u_n^{(k+1)} = u_n^{(k)} + h v_n^{(k)} + \frac{1}{2} h^2 g_n^{(k)}.$$

2 Вычисление экспоненты матрицы

Решая (2), важно эффективно вычислять матричную экспоненту. Рассчитать экспоненту симметричной трехдиагональной положительно полуопределенной матрицы можно с помощью рациональной аппроксимации [1].

Аппроксимацию экспоненты рациональной функцией $R_m(z)$ представим в виде суммы простейших дробей:

$$\exp(-z) \approx R_m(z) \equiv \frac{p_m(z)}{q_m(z)} = a_0 + \sum_{j=1}^m \frac{a_j}{z - \theta_j}. \quad (3)$$

Матрица $\mathbf{A}^{(k)}$ является трехдиагональной по условию задачи. Чтобы гарантированно работать с положительно полуопределенной матрицей, нужно найти минимальное собственное значение $\lambda_1^{(k)}$ матрицы $\mathbf{A}^{(k)}$ и выразить экспоненту матрицы $\mathbf{A}^{(k)}$ через экспоненту матрицы $\mathbf{T}^{(k)} = hi(\mathbf{A}^{(k)} - \lambda_1^{(k)}\mathbf{I})$:

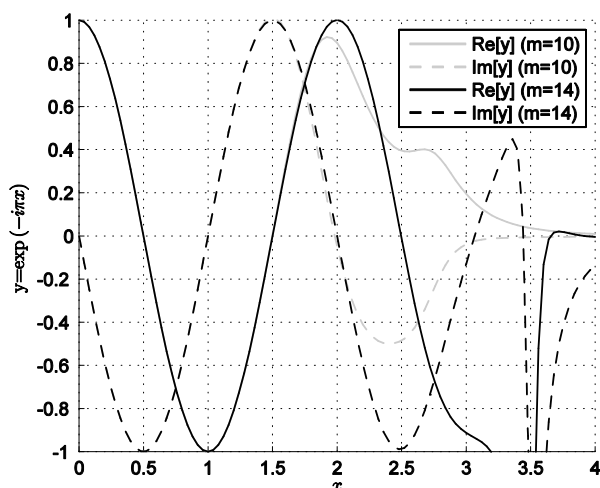


Рис. 1. Рациональная аппроксимация Чебышева функции $v(x) = e^{-inx}$ ($x \in \mathbb{R}$) для $m = 10$ и $m = 14$.

$$\begin{aligned} \exp(-hiA^{(k)})b^{(k)} &= \\ &= \exp(-hi\lambda_1^{(k)})\exp(-hi(A^{(k)} - \lambda_1^{(k)}I))b^{(k)} = \\ &= \exp(-hi\lambda_1^{(k)})\exp(-T^{(k)})b^{(k)} \end{aligned}$$

Тогда

$$\exp(-T^{(k)})b^{(k)} \approx a_0x^{(k)} + \sum_{j=1}^m a_j y_j^{(k)}$$

где векторы $y_j^{(k)}$ являются решениями систем линейных уравнений

$$(T^{(k)} - \theta_j I)y_j^{(k)} = x^{(k)}, j = 1, \dots, m \quad (4)$$

Так как $T^{(k)}$ является трехдиагональной матрицей, то для решения (4) относительно $y_j^{(k)}$ можно воспользоваться методом прогонки.

Для случая рациональной аппроксимации Чебышева, которая минимизирует максимальную ошибку аппроксимации экспоненты на интервале $[0; \infty)$, коэффициенты a_j и полюса θ_j приведены в работе [1] для $m = 10$ и $m = 14$. Для действительного аргумента z при $m = 14$ на всей положительной полуоси ошибка этой аппроксимации не превышает $2 \cdot 10^{-14}$. На рис. 1 приведены результаты аппроксимации для мнимых значений. Видно, что (3) справедливо только в ограниченной области z , а точнее ошибка не превышает 10^{-7} только на интервале $[0; 0.75\pi]$. Матрицы в экспонентах уравнений (2) являются мнимыми, поэтому важно выбирать шаг h так, чтобы норма матрицы

$T^{(k)} = hi(A^{(k)} - \lambda_1^{(k)}I)$ оставалась в допустимом диапазоне. Это легко осуществить, если найти максимальное собственное $\lambda_n^{(k)}$ значение матрицы $A^{(k)}$ и выбрать шаг так, чтобы произведение $h(\lambda_n^{(k)} - \lambda_1^{(k)})$ было меньше верхней границы интервала, на котором аппроксимация экспоненты матрицы дает достаточную точность. Например, чтобы всегда получать точки равномерно с заранее заданным шагом τ , адаптивный шаг будет $h = \tau / M$, где M – количество промежуточных шагов:

$$M = \left\lceil \tau \frac{\lambda_n^{(k)} - \lambda_1^{(k)}}{0.75\pi} \right\rceil.$$

В системе (1) существует первый интеграл – полная вероятность нахождения заряда в цепочке всегда постоянная:

$$S = \sum_{n=1}^N |b_n|^2 = 1.$$

Это свойство позволяет контролировать точность расчета и корректировать текущие амплитуды на каждом шаге для сохранения нормировки: $b_n^{(k)} / S^{(k)}$.

Заключение

Предложенная смешанная схема на три порядка эффективнее схем, которые требуют шаг по времени меньше характерного времени квантовой части системы.

Работа выполнена при частичной финансовой поддержке РФФИ, проекты 10-07-00112-а, 11-07-00635-а, 11-07-12054-офи-м-2011.

Список литературы

- [1] G. Gallopoulos, Y. Saad. SIAM Journal on Scientific and Statistical Computing, 1992, **13**, 1236–1264.
- [2] H.S. Greenside, E. Helfand. Bell System Technical Journal, 1981, **60**, 1927–1940.
- [3] В.Д. Лахно, Н.С. Фиалко. Письма в ЖЭТФ, 2003, **78**, 786–788.
- [4] Д.А. Тихонов, Е.В. Соболев, В.Д. Лахно, Н.С. Фиалко. Математическая биология и биоинформатика, 2011, **6**, 264–272.

Термодинамические и кинетические свойства различных типов открытого состояния ДНК

Шигаев А.С.⁽¹⁾, Пономарёв О.А.⁽¹⁾, Лакно В.Д.⁽¹⁾

Институт математических проблем биологии РАН⁽¹⁾

shials@rambler.ru

The probabilities and lifetimes of DNA open states by data of two different experimental techniques (NMR and FCS) are compared. The discrepancy of 3 orders is explained by several different underlying conditions. Among them are the different lengths and nucleotide sequence of examined DNA, energy transfer and localization in long duplexes and the different types of the open state, observed with the two methods.

Диссоциацию двухцепочечной ДНК на отдельные цепи обычно называют денатурацией, или плавлением, так как самый простой способ разделения цепей – нагревание раствора дуплекса. Температура плавления ДНК $T_{пл}$ зависит от целого ряда факторов, среди которых важную роль играет её нуклеотидная последовательность.

Даже при температурах, намного меньших $T_{пл}$, в ДНК могут возникать так называемые пузырьки денатурации – локальные области, в которых её цепи разделены, а стэкинг-взаимодействия между соседними основаниями нарушены. Способность ДНК к образованию пузырьков играет важную роль в функциях генома, а также влияет на динамику переноса катион-радикала в ДНК, так как последняя сильно зависит от целостности стэкинг-взаимодействий.

В нашей работе проведён анализ литературных данных, а также сравнение пузырьков денатурации с другими типами флуктуаций дуплекса с точки зрения термодинамики. Единственным методом, позволяющим напрямую изучать кинетику коллективных открываний пар оснований в ДНК, является флуоресцентная корреляционная спектроскопия (ФКС) модифицированных дуплексов, называемых маячками. В маячках флуорфор и тушитель пришиты на разные цепи ДНК, друг напротив друга. Поэтому открытое состояние того участка, в середине которого они находятся, является причиной задержки тушения флуоресценции, в отличие от закрытого, в котором флуорфор и тушитель могут беспрепятственно контактировать [1].

Тем не менее, данные о временах жизни открытого состояния ДНК, полученные при помощи

ФКС, существенно отличаются от результатов ЯМР, согласно которым это время в среднем ниже на целых 3 порядка [2,3]. В силу значительно большего объёма ЯМР-данных, а также ряда других причин, достоверность кинетической картины, полученной на спектроскопии, подвергают сомнению [4]. Мы проанализировали данные двух методов с точки зрения кинетики и термодинамики. Показано, что хотя активационные барьеры свободной энергии для одиночных и коллективных открываний нуклеотидных пар имеют схожую величину, они в корне отличаются по энтальпийно-энтропийному соотношению. Другими словами, одиночное и коллективное открывание – процессы, термодинамически совершенно разные.

Кроме того, активационный барьер для коллективного открывания в экспериментах по ФКС сильно снижен из-за очень специфической первичной структуры исследуемого домена: 18 идущих подряд АТ-пар. Длинная АТ-богатая область способствует локализации энергии, что серьёзно повышает вероятность открытого состояния. Последние теоретические результаты, полученные в модифицированной модели Пейярда-Бишоп-Доксуа, указывают на то, что в подобных областях сильные флуктуации приводят не только к повышенной вероятности открытых состояний, но и к очень большим временам их жизни [5,6]. В экспериментах показано также влияние областей с сильно выраженными флуктуациями на динамику удалённых участков [7].

В докладе обсуждаются различия двух типов открываний, возможность сохранения стэкинг-взаимодействий при коллективном открывании, а также принципиальная разность флуоресцентного подхода и ЯМР с точки зрения чувствительности к разным типам открытых состояний. Также обсуждается зависимость динамики открываний от длины ДНК и её первичной структуры и влияние флуктуаций сахарофосфатного остова, усиливающих при нагревании.

Работа выполнена при частичной финансовой поддержке РФФИ, проекты 10-07-00112-а, 11-07-12054-офи-М и 12-07-00279-а.

Список литературы

- [1] G. Altan-Bonnet, A. Libchaber and O. Krichevsky, *Physical Review Letters*, 2003, **90**, 138101.
- [2] D. Coman, I.M. Russu, *Biophysical Journal*, 2005, **89**, 3285-3292.
- [3] U. Dornberger, M. Leijonh, and H. Fritzsche, *The Journal of Biological Chemistry*, 1999, **274**, 6957–6962.
- [4] M. Peyrard, S. Cuesta-López, D. Angelov., *Journal of Physics: Condensed Matter*, 2009, **21**, 034103.
- [5] B.S. Alexandrov, V. Gelev, Y. Monisova, L.B. Alexandrov, Bishop, K.O. Rasmussen, A. Usheva, *Nucleic Acids Research*, 2009, **37**, 2405–2410.
- [6] B.S. Alexandrov, V. Gelev, S.W. Yoo, L.B. Alexandrov, Y. Fukuyo, A.R. Bishop, K.O. Rasmussen, A. Usheva, *Nucleic Acids Research*, 2010, **38**, 1790–1795.
- [7] S. Cuesta-Lopez, H. Menoni, D. Angelov, M. Peyrard, *Nucleic Acids Research*, 2011, **39**, 5276–5283.

Bioferroelectricity in the di-phenylalanine peptide molecular nanostructures: computational modeling and nanoscale physical properties

Bystrov V.S.^(1,2), Paramonova E.V.⁽¹⁾, Bdikin I. K.⁽³⁾, Kopyl S.⁽³⁾, Heredia A.⁽²⁾, A. Kholkin A.L.⁽²⁾

Institute of Mathematical Problems of Biology RAS, Pushchino, 142290, Russia ⁽¹⁾,
Department of Mater.Cer.Eng. & CICECO, University of Aveiro, 3810-193 Aveiro, Portugal ⁽²⁾
Department of Mech.Eng. & TEMA, University of Aveiro, 3810-193 Aveiro, Portugal ⁽³⁾
vsbys@mail.ru bystrov@ua.pt

The results of computational molecular modeling of di-phenylalanine (FF) nanostructures revealed its ferroelectric properties and predicted order-disorder phase transition are presented in this work. The theoretical computational data are corroborated by experimental studies of piezo- and ferroelectric properties of FF peptide nanotubes (PNT) at the nanoscale using atomic and piezo-force microscopy techniques.

Кластером называем группу нейронов, между отростками которых имеются поры (синцитиальные анастомозы). В модели кластер рассматривается как один геометрический граф с уравнениями на нём; учитывается как синаптическое, так и внесинаптическое химическое управление медиаторами.

1 Введение

1.1 Геометрические особенности нейронов и моделей

Нейроны имеют форму деревьев или кустарников с ветвями конечной толщины. Поверхностная мембрана нейронов изолирует цитоплазму от межклеточной среды электрически и химически; её удельная ёмкость $C = 1 \text{ мкф/см}^2$. В некоторых структурах мозга имеются поры между отростками соседних нейронов [1]. Группу связанных порами нейронов назовём кластером; между нейронами разных кластеров нет пор.

Проведём в каждой ветви среднюю линию, а места разветвления (и сому) заменим точками, и будем рассматривать задачу на геометрическом графе Γ , состоящем из отрезков гладких кривых (рёбра графа) и точек (вершины графа). Множество граничных вершин (концы отростков) обозначаем $\partial\Gamma$, множество внутренних вершин J .

Поры в кластере заменяются очень короткими рёбрами с узлами на каждом конце ребра. Поэтому граф кластера не является просто объединением графов входящих в кластер нейронов $\bigcup_{n=1}^N \Gamma_n$; в него добавляются рёбра пор и внутренние вершины

концов пор, в результате чего те рёбра графов нейронов, которые образовали пору, разбиваются (каждое из них) на два более коротких ребра.

1.2 Электрические потенциалы и токи

Внутриклеточный потенциал кластера обозначим $V(x,t)$, $x \in \Gamma \in \mathbb{R}^3$, $t \geq 0$, где x – одномерная переменная на рёбрах Γ . Внеклеточный потенциал обозначим как $\varphi(x,t)$, $x \in \mathbb{R}^3$. Дополнительные фазовые переменные обозначим $c(x,t)$, $p_l(x,t)$, $q_l(x,t)$; $p_l, q_l \in [0,1]$; $l=1,2,3$; $c \geq 0$. Токи в модели нейрона и кластера могут быть двух типов: токи через поверхность (мембрану) нейронов будем обозначать буквой I с нижними индексами; токи вдоль рёбер будем обозначать буквой i с нижними индексами.

2 Управление электрическим состоянием нейронов и кластеров

2.1 Ионные каналы, управляемые электрическим полем

Ограничимся классической системой уравнений Ходжкина-Хаксли. В этой системе есть натриевые и калиевые каналы, управляемые полем, и каналы утечки. Суммарные токи обозначаются $I_{Na}(x,t)$, $I_K(x,t)$ и $I_L(x,t)$. Для этих токов имеем уравнения:

$$I_{Na} = g_{Na}(q_1)^3 q_2 (V - \varphi - V_{Na}) ; \quad I_K = g_{Kp3} (q_3)^4 (V - \varphi - V_K); \\ I_L = g_L (V - V_L), \text{ и дифференциальные уравнения} \\ (q_l)_t = \alpha_l (V - \varphi) - \gamma_l (V - \varphi) \cdot q_l; \quad l=1,2,3. \quad (1)$$

$$(V)_t - (\varphi)_t - (V)_{xx} + I_{Na}(x,t) + I_K(x,t) + I_L(x,t) + I_s = 0. \quad (2)$$

Уравнения (1) и (2) заданы на рёбрах графов. Во внутренних вершинах графов выполняется условие непрерывности $V(x,t)$ и закон Кирхгофа для токов. В граничных вершинах модели ($x \in \partial\Gamma$) задано условие $(V)_x = 0$.

2.2 Прямое химическое управление ионными каналами

Управляющими сигналами являются специальные химические вещества ("медиаторы"). Обычно нейроном управляют несколько медиаторов одновременно, но далее рассмотрим только один из них. Точки $x_j \in \Gamma$, в которых в моменты $t_{j,k}$, $k=1,2, \dots$

выделяется порция молекул медиатора, называются синапсы. Медиатор вызывает синаптический ток [2]

$$I_s(x_j, t - t_{j,k}) = g_s(t - t_{j,k}) \cdot (V - \varphi - V_s),$$

и в уравнении (2)

$$I_s = \sum_j \delta(x - x_j) \sum_{t_{j,k}} I_s(x_j, t - t_{j,k}). \quad (3)$$

2.3 Параметрическое химическое управление ионными каналами

Не все молекулы из выделившейся порции медиатора вызывают синаптический ток; часть молекул удаляется из-за диффузии во все стороны и постепенно уничтожается. Часть молекул медиатора

достигает точек $x_m \in \Gamma$, в которых находятся "метабоотропные" рецепторы [3]. Эти рецепторы, соединяясь с медиатором, генерируют в Γ поток "вторичного посредника 1", с концентрацией $p_1(x, t)$, диффузия которого в Γ описывается уравнением

$$(p_1)_t - D_1(p_1)_{xx} + \varepsilon_1 p_1 = \sum_m \delta(x - x_m) \mu c(x_m, t). \quad (4)$$

Для концентраций $p_2(x, t)$ и $p_3(x, t)$ веществ 2 и 3 имеем уравнения на каждом ребре:

$$(p_2)_t - D_2(p_2)_{xx} + \varepsilon_2 p_2 = p_1; \quad p_3(x, t) = f(p_2(x, t)), \quad (5)$$

где $f(p)$ – монотонная функция, $f(0)=1$, $f(\infty)=0$.

Увеличение концентраций p_1 и p_2 приводит к уменьшению коэффициента p_3 в формуле для I_K , то-есть к уменьшению проводимости для ионов калия и тока этих ионов. Поэтому проведение синаптических сигналов от отдаленных дендритов улучшается.

2.4 Концентрация медиатора во внеклеточной среде

Внеклеточная среда в нервной системе – это совокупность тонких щелей между клетками и их отростками, сообщающихся между собой и заполненных межклеточной жидкостью с высокой электропроводностью. Математический аппарат теории дифференциальных уравнений на стратифицированных множествах недостаточно разработан. Поэтому для первоначальной оценки концентрации $c(x, t)$ медиатора во внеклеточной среде в (4) используем уравнение диффузии в R^3 при $x \square = (x_{\square 1}, x_{\square 2}, x_{\square 3}) \in R^3$; $\{x_{j1}\} \supset \{x_j\}$; $\{t_{i1,k}\} \supset \{t_{i,k}\}$:

$$(c)_t - D \cdot \Delta c + \mu c = c_0 \cdot \sum_{i1} \delta(x - x_{i1}) \sum_{t_{i1,k}} \delta(t - t_{i1,k}). \quad (6)$$

2.5 Поры

В отличие от рёбер, соответствующих ветвям дендритов и аксонов, для "вырожденного" ребра поры нет уравнений вида (1)-(3), а используется

только закон Ома. Если обозначить узлы поры как $n1$ и $n2$, а внутриклеточные потенциалы как V_{n1} и V_{n2} , то ток через пору $I_{n1,n2}$ равен

$$I_{n1,n2} = g_{n1,n2}(V_{n2} - V_{n1}).$$

При этом проводимость поры $g_{n1,n2}$ может меняться в широких пределах в зависимости от её диаметра. Влияние поры на функционирование кластера нейронов также может меняться от незначительного до полной синхронизации импульсов, уходящих по аксонам.

3 Заключение

Уравнения (1) - (3) определяют прямое управление электрическим состоянием в моделях нейрона и кластера. Уравнения (4), (5), (6) определяют параметрическое управление. Подчеркнем, что в прямом управлении участвуют только те источники химического сигнала, которые находятся на Γ ; в то же время в параметрическом управлении участвуют все источники, расположенные в объёме, занимаемом кластером Γ .

Выходным сигналом нейрона является электрический импульс, возникающий и распространяющийся в неуправляемой ветви называемой "аксоном", а выходным сигналом кластера - синхронные импульсы по пучку аксонов. При стационарных управляющих сигналах решение уравнений (1) - (6) может быть близким к постоянному (при устойчивости в малом постоянного решения). При потере устойчивости может возникнуть последовательность пачек импульсов. Потеря устойчивости и генерация импульсов кластером Γ возможны даже в том случае, когда Γ вообще не получает медиатора через синапсы, но молекулы медиатора выделяются на близлежащих нейронах.

Работа поддержана грантами РФФИ 08-04-90033-Бел_а, 07-04-92167а.

Литература

- [1] Сотников О.С., Рыбакова Г.И., Соловьева И.А. Проблема слияния отростков нейронов// Морфология, 2007, т. 132, № 5. С. 18-22.
- [2] Экклс Дж. Физиология синапсов. М.: изд. МИР, 1966. -395 с.
- [3] Радченко А.Н.. Информационные механизмы мозга. СПб.: Изд. Геликон Плюс, 2007. – 240 с.

Filling of carbon nanotubes with magnetic nanoparticles for biomedical application

Kopyl S.⁽¹⁾, Bystrov V.S.^(2,3), Maiorov M.⁽⁴⁾, Maevskij E.I.⁽⁵⁾, Sousa C.M.A.⁽¹⁾

Department of Mech. Eng. & TEMA-NRD, University of Aveiro, 3810-193 Aveiro, Portugal ⁽¹⁾

Institute of Mathematical Problems of Biology RAS, Pushchino, 142290, Russia ⁽²⁾,

Department of Mater.Cer.Eng. & CICECO, University of Aveiro, 3810-193 Aveiro, Portugal ⁽³⁾

University of Latvia, Institute of Physics, Salaspils, 2169, Latvia ⁽⁴⁾

Institute of Theoretical.& Experimental Biophysics RAS, Pushchino, 142290, Russia ⁽⁵⁾

svitlanakopyl@ua.pt

Magnetic carbon nanotubes composites were obtained by filling carbon nanotubes with paramagnetic iron oxide particles. Detailed investigation of magnetic behaviour of resulting composites was done at different temperatures. Measurements indicate that these functionalized nanotubes are superparamagnetic at room temperature; however, no superparamagnetism was observed at 125 K and 80 K. The blocking temperature T_B was estimated at 145 K. These magnetic carbon nanotubes have the potential of being used in a wide range of applications, in particular, the production of nanofluids, which can be controlled and steered by appropriate magnetic fields.

1 Introduction

The research field of carbon nanotubes (CNTs) has received a continuously growing interest since their discovery in 1991 [1] due to their unique and highly desirable electrical, thermal and mechanical properties [2]. Functionalizing CNTs with magnetic nanoparticles can combine the features of magnetic nanoparticles and CNTs, which may result in materials with novel chemical and physical properties, and thus promising applications. Our motivation for creating a CNT/magnetite composite stems from the goal of creating a new generation of nanofluids, i.e. a suspension of nanoparticles, in the present case magnetic carbon nanotubes, in a basefluid [3], which have very high thermal conductivity and are amenable to be controlled by appropriate magnetic fields. This can be achieved using appropriately placed and designed external magnets; therefore, this system is practically maintenance free, which makes it very attractive for numerical application.

Recent reports described the attachment of various inorganic nanoparticles to the internal surface of the CNT cavity through several experimental methods. One of these methods based on spontaneous penetration of fluid into wettable capillaries by capillary action. This method of filling nanotubes was pioneered by Ajayan and Iijima [4].

The present study largely follows the work of Gogotsi's group (Drexel University, USA), Korneva *et al.* [5], whom demonstrated that it is possible to fill CNTs with a diameter of 300 nm using commercially available ferrofluid. An aqueous suspension of magnetic

nanoparticles is drawn by capillary suction into an open ended nanotube. Subsequent evaporation of the solvent resulted in nanotubes loaded with magnetic nanoparticles. To produce a nanofluid with magnetic CNTs in suspension, this method seems to have some potential, however, the level of CNTs filling, in terms of the ratio between the magnetic particles volume by that of the nanotube, is relatively low (approximately 11%).

Besides, it is technologically important to retain the desirable magnetic properties of the magnetic nanoparticles in CNTs through control over the particle size and uniformity of packed particles. In this paper it is presented a modification of the procedure CNTs production proposed in [5], which can increase the nanotubes filling percentage to levels that are adequate for the production of nanofluids with magnetic CNTs as nanoparticles and detailed magnetic characterization of CNTs filled with ferroparticles.

3 Experimental details and results.

2.1 Experimental

Carbon nanotubes were produced in the NRD laboratory (TEMA, University of Aveiro) by a non-catalytic chemical vapour deposition technique based on the pyrolysis of ethylene. Before the growth of CNTs, the alumina template membranes (13 mm diameter, 60 μm thick, and 0.2 μm pore size purchased from Whatman® Inc., England) were placed vertically inside the quartz tube. The nanotubes were formed in straight cylindrical pores, which run through the membrane thickness. The pore diameter and thickness of the membrane determine the dimensions of the nanotubes. The diameter of pore in our experimental results is in the range of 200 ± 15 nm. The resulting CNTs have at least one open end, and their walls are highly disordered and amorphous - this characteristic makes viable filling the nanotubes with both organic and water-based fluids, as noted in [5]. The average length of the CNTs is 6 μm and the majority of carbon nanotubes have a diameter close to 124 nm [6].

The organic based ferrofluid EMG 911 (Ferrotec Corporation) was used in this experiment (the surface tension of EMG 911 is 68 mN/m; the average diameter of Fe_3O_4 nanoparticles is 7.5 nm) [6]. To fill the carbon nanotubes sitting in the template with ferrofluid the

procedure reported by Korneva *et al.* [5] was modified. In the present study, the loading was conducted by making multiple additions (up to 100 times by weight) of the ferrofluid to the system; no magnetic field was used in this filling procedure. The ferrofluid overflow enters the tubes at a relatively fast pace, and the flow was kept by making small additions of ferrofluid to the overflow. When it is apparent that no further ferrofluid enters the tubes, the system is left still for 24 hours to allow impregnation (see more details in ref. 6) The nanoparticle-filled CNTs were characterized using scanning transmission electron microscopy (TEM) mode in the SEM Hitachi SU-70 operated at 30 kV. This scanning electron microscope is equipped with an energy-dispersive X-ray spectrometer (EDS) (University of Aveiro).

2.2 Results

A typical TEM image of CNTs with Fe₃O₄ nanoparticles from an organic-based ferrofluid is shown in Fig. 1.

In the image, it can be easily identified the nanotubes filled with the Fe₃O₄ nanoparticles. Fig. 1 (inset) results from the EDS measurements of the CNTs filled Fe₃O₄ nanoparticles; these measurements confirm the presence of atomic iron in the CNTs.

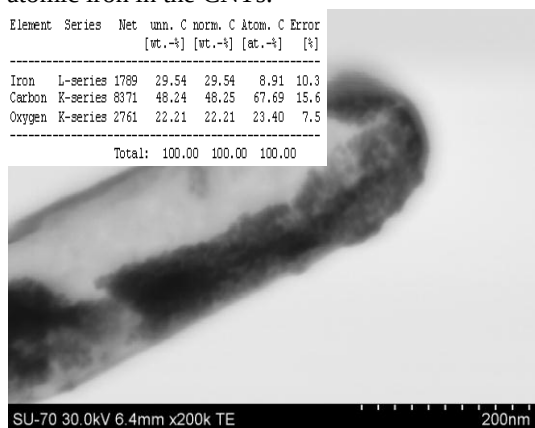


Fig.1. TEM image of CNTs filled with Fe₃O₄ nanoparticles. Inset: EDS results for CNTs filled with Fe₃O₄.

The original carbon nanotubes do not present any ferromagnetic properties. The samples of CNTs filled with Fe₃O₄ indicate that ferromagnetic properties do occur. The saturation magnetization depends on the temperature, and it varies from 17.53 Am²/kg to 20.10 Am²/kg; this range is in good agreement with the values reported in the literature [7]. The magnetic CNTs show no hysteresis loop in the presence of a low magnetic field at room temperature and at 273 K, however, hysteresis is clearly observed at low temperature: at 125 K and 80 K. Detailed analysis of the magnetic behaviour of CNTs with magnetic particles was done. The some parameters such saturation magnetization M_s, coercivity H_C, remanent magnetization M_R and ratio R=M_R/M_s at different temperatures are presented in Table 1.

At low temperatures the coercivity of a system of non-interacting and randomly oriented particles is described by specific relation [8], allow us to determine the blocking temperature T_B. At temperatures well below T_B, the hysteresis appears and, consequently the superparamagnetism disappears, as the thermal energy is no longer sufficient to overcome the magnetic anisotropy energy. In our case the value of T_B ~ 145 K.

TABLE 1. SOME CHARACTERISTICS OF CNTs WITH MAGNETIC NANOPARTICLES

Temperature T, K	H _C , A/m (Oe)	M _s , Am ² /kg	M _R , Am ² /kg	Ratio R = M _R /M _s
293	0	17.53	0	0
273	0	17.78	0	0
175	0	19.26	0	0
125	217.7 (2.75)	19.78	0.2	0.01
80	794.4 (9.98)	20.10	0.63	0.03

The filling level of carbon nanotubes with magnetic particles was calculated. This corresponds to 46 % filling by volume of the Fe₃O₄ particles in the nanotubes, which is a very promising result.

CNTs can be used for nonspecific and specific sorption. Carbon nanotubes themselves are excellent sorbent due to a developed surface. Magnetite particles have increased affinity for the heme-containing molecules, and if necessary can be covered by specific ligands such as antigens, antibodies or low molecular weight substrates. The composition CNTs can easily be used as a magnetic sorbent, for example, for cleaning biological fluids from the extracellular hemoglobin and myoglobin, which appear in the blood plasma during pathology (hemolysis of red blood cells and massive damage of muscle tissues). At present, we determine on model systems the selectivity of sorption with magnetic CNTs for purification of blood plasma and whole blood from the hemoglobin and myoglobin. For cleaning of the liquid from CNTs loaded with toxins, we have developed special device - a magnetic separator.

REFERENCES

- [1] Iijima, S., *Nature*, 1991, **354**, pp.56-58.
- [2] Dresselhaus, M.; Dresselhaus, G.; et al. *Annu. Rev. Mater. Res.* 2004, **34**, 247-278.
- [3] Godson, L.; Raja, B.; et al. *Renewable and sustainable energy reviews*, 2010, **14**, 629-641.
- [4] Ajayan, P. M. and Iijima, S. *Nature* 1993, **361**, 333-334.
- [5] Korneva, G.; Ye, H.; Gogotsi, Y. et al. *Nano letters* 2005, **5**, 5, 879-884.
- [6] Kopyl, S.; Bystrov, V.; Maiorov, M.; et al. *J. Mater. Chem.* 2012 (submitted).
- [7] Tan, F.; Fan, X.; Zhang, G.; et al. *Materials letters* 2007, **61**, 1805-1808.
- [8] Fonseca, F.; Goya, G.; Jardim, R.; et al. *PHYSICAL REVIEW B* 2002, **66**, 104406.

Work supported FCT PTDC/EME-MFE/105031/2008

Computational study of Hydroxyapatite nanostructures, properties and its surface interactions with various species in different conditions

Bystrov V.S.^(1,2), Coutinho J.⁽³⁾, Costa E.⁽²⁾, Dekhtyar Yu.⁽⁴⁾, Bystrova A.V.^(4,5), Santos C.^(2,6), Almeida M.⁽²⁾, S. Kopyl⁽⁷⁾, Bystrova V.V.⁽⁸⁾, Paramonova E.V.⁽¹⁾

Institute of Mathematical Problems of Biology RAS, Pushchino, 142290, Russia⁽¹⁾,
Department of Mater.Cer.Eng. & CICECO, University of Aveiro, 3810-193 Aveiro, Portugal⁽²⁾
University of Aveiro, Department of Physics & I3N, Aveiro, Portugal⁽³⁾
Riga Technical University, Biomedical Engineering and Nanotechnology Institute, Riga, Latvia⁽⁴⁾
Institute of Theoretical & Experimental Biophysics RAS, Pushchino, 142290, Russia⁽⁵⁾
Instituto Politécnico de Setúbal, EST, Estefanilha, Setubal, Portugal⁽⁶⁾
Department of Mech.Eng. & TEMA, University of Aveiro, 3810-193 Aveiro, Portugal⁽⁷⁾
Moscow State University, Faculties of Geography & Geology, Moscow, Russia⁽⁸⁾
vsbys@mail.ru bystrov@ua.pt

The results of computational modeling for Hydroxyapatite (HAP) nanostructures and properties exploration are presented in this review. HAP was studied from first principles approaches using Local Density Approximation (LDA) method in combination with various quantum-chemical (QM), including Density Functional Theory (DFT) methods, and molecular mechanical (MM, BIO CHARMM) methods from HyperChem 7.5/8.0 package. Different conditions were explored on HAP surface properties and sizes/shapes of HAP clusters/nanocrystals. Obtained data then were used for studies of interactions of HAP clusters with various species (citrate, living cells).

1 Introduction

Hydroxyapatite (HAP) is a naturally occurring mineral form of calcium apatite with the formula $\text{Ca}_5(\text{PO}_4)_3(\text{OH})$, but is usually written $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$ to denote that the crystal unit cell comprises two entities. The OH^- ion forms inner channels along *c*-axis, allowing for transfer of protons and can be replaced by fluoride, chloride or carbonate, producing fluorapatite or chlorapatite. It crystallizes usually in the hexagonal crystal system ($P6_3/m$), but exists in monoclinic structure ($P2_1/b$) too. If OH^- ions are ordered in one parallel direction HAP has a non-centrosymmetric structure, which could reveal piezoelectric properties. Geological pure HAP powder is white. Naturally occurring apatites can, however, also have brown, yellow, or green colorations, comparable to the discolorations of dental fluorosis. Up to 50% of bone by weight is made up of a modified form of HAP (known as bone mineral) [1]. Citric acids and citrates attached on HAP surface and influence on HAP nanoparticles size and shape formation as well as protons content in media (pH) [1,2]. Polarized HAP (charged surface) attracts more living cells (osteoblasts) [3]. Carbonated calcium-deficient HAP is the main mineral of which dental enamel and dentin are

comprised. HAP nanocrystals are also found in the small calcifications (within the pineal gland of brain and other structures) known as corpora arenacea or 'brain sand' [4] and have piezoelectric properties [5]. All these effects urgently need further studies. In present work we explore HAP using *ab initio*/ other methods.

2 Main computational details and results of models

2.1 Computational details

Beyond HyperChem widely used as in our previous studies [1] as well now, additionally in this work we use calculations by density functional theory (DFT) at local density approximation (LDA). We use the standard valence configurations for hydrogen (1s1), for phosphorus (3s23p3) and for oxygen (2s22p4). But for calcium we use two different valence configurations: 1) Ca2 with 2 electrons (4s2) shell (WF1), and most accurate 2) Ca10 with 10 electrons (3s2_3p6_4s2) shell (WF2). The cutoff for the kinetic energy for bulk calculations was at 300 a.u. The Monkhorst–Pack *k*-point mesh for the Brillouin Zone (BZ) integration of hexagonal and monoclinic primitive cells were found that the best are $2 \times 2 \times 4$.

The initial lattice unit cell for hexagonal HAP structure consists of 44 atoms, while for monoclinic HAP nanostructures it consists of 88 atoms, because the lattice unit cell is doubled in this last monoclinic case along *b* axis. In all calculations—except the elastic constant calculations—we allow for full relaxation of the cell, including changing the volume and adjusting the atomic positions and cell shape. When calculating the elastic constants, we only allow for the relaxation of ionic positions within the strained cells. Obtained optimized data for each HAP structure, were taken for further HyperChem modelling in usual *.hin format.

2.2 Results

Main computed data are presented in Table below.

HAP type	Lattice parameters (Å)	Total energy E (a.u.)	E gap (eV)	Bulk modulus (GPa)
<i>WF1 model: for Ca with 2 electrons</i>				
Hexag P6 ₃	a=9.4732 c=6.9986	-467.09923	5.6	81.6
Monocl P2 ₁ (OH parall)	a=18.9795 /2=9.4898 b=6.9893 c=9.4801	-934.19833 /2 -467.09917	5.48	79.6
$\Delta E = E(P2_1) - E(P6_3)$		+1.63 meV		
Monocl P2 ₁ /b (OH oppo)	a=18.9500 /2=9.4897 b=6.99709 c=9.47401	-934.20024 /2 -467.10012	5.8	82.7
$\Delta E = E(P2_1/b) - E(P6_3)$		-24.22meV		
<i>WF2 model: for Ca with 10 electrons</i>				
Hexag P6 ₃	a=9.35125 c=6.91222	-826.67621	5.44	102.3
Monocl P2 ₁ (OH parall)	a=18.7493 /2=9.3747 b=6.9032 c=9.35468	-1653.3521 /2 - 826.6761	5.99	99.83
$\Delta E = E(P2_1) - E(P6_3)$		+3.92 meV		
Monocl P2 ₁ /b (OH oppo)	a=18.7372 /2=9.3686 b=6.90648 c=9.35381	-1653.3525 /2 - 826.6762	5.96	101.3
$\Delta E = E(P2_1/b) - E(P6_3)$		-2.03 meV		

Our computed data show that monoclinic phase with opposite oriented OH⁻ions in the neighboring channels have lowest energy on ~24 meV for Ca2 model and on ~ 2 meV for Ca10 model. This result confirmed by [6-8] as well the computed values of bulk modulus and forbidden energy gap Eg. The ordered P2₁ form with parallel oriented OH⁻ions in channels have some higher energies: +1.63 meV for Ca2 model, and +3.92 meV for Ca10 model. In any cases these values are not much and lie in the frame of thermal energy at room temperature ~25 meV. It means that all phase structures can co-exist and easy transfer one to other, depending of the slow environmental changes. It leads to the opportunities of the local fluctuation of the phase in space and to the variations of the surface high profile due to the occasional changes of the neighboring phase along ones. This evidence was recently observed by AFM [9]. For monoclinic ordered phase P2₁ the piezo- response at the local places was observed too [19, 10] on the thin films samples with ~ 500 nm width.

Calculation of DOS allow us to study also the influence of lattice charges on the shift and changes of Eg, which are agreed with experimental explorations of living cells attachment to charged HAP surface [1, 3] and predict some new properties [11]. Of course, it need also studies of the HAP surface changes (deformations, atoms positions shifting, vacancies, etc). It will be made in our further investigations. One of the possible opportunities is arising of the hydrogen vacancies on the HAP surface

(inducing the corresponding charges) under microwave irradiations, due to soliton mechanism of protons propagation along OH-channels inside HAP similar to the proton dynamics in the water chains [12].

Influence of the citrates was studied by model with the fixed effective atomic charges of HAP lattice. Using different lattice parameters and citrates contents, which were corresponding to experimental data [3], we obtain followings: 1) the interaction of citrates has two main types – with longitudinal (along one axis) and parallel orientations; 2) the interactions of free and attached on HAP surface citrates have a very different character. If free citrates are oriented along one axis its push away one from other, if its in parallel positions – its attract one to other and arise the hydrogen bond (HB, E_{HB} ~ 5 kcal/mol) with minimum bonding position at ~ 6 Å. For HAP attached citrates in both cases we have very complexly modulated profile with deep wholes. If citrates have a low content it prefer to attach to HAP surface (at the specific positions for Ca atoms) in longitudinal orientation stabilizing HAP nanocrystals with most short lattice parameters and lower total length (~100 nm [3]). Another situation arise for citrates having most high content, when its feeling all HAP surface, but with parallel and with HB orientation. These situations correspond to growth of most long (~1000 nm) stable HAP nanocrystals [3]. Obtained data clarifying situation of HAP surface interactions with various species and give us new insight for most deep further studies of HAP structural, physical and interfacial peculiarities.

Список литературы

- [1] V.S. Bystrov, E. Paramonova et al. *J.Phys.: Condens. Matter.*, 2011, **23**, 065302 (10pp).
- [2] M. Martins, C. Santos, M. Almeida, E. Costa, *J. Colloid and Interface Scienc.*, **2008**, 210-216.
- [3] Yu. Dekhtyar, V. Bystrov, et al, In: The Society for Biomaterials 2011 Annual Meeting. - April 13-16, 2011, Orlando, Florida, USA.
- [4] G. Bocchi, G. Valdre G, *J. Inorg. Biochem.*, 1993, **49** (3), 209–20.
- [5] S. Lang, A. Marino et al, *Bioelectrochem.Bioenerg.*, 1996, **41**, 191-195.
- [6] A. Slepko and A.A. Demkov, *Phys. Rev.*, 2011, **B 84**, 134108.
- [7] N. Y. Mostafa, P. W. Brown, *J. Phys. Chem. Solids* 2007, **68**, 431.
- [8] S.A.M. Tofail, D. Harvety, K.T. Stanton, J.B. McMonagle, *Ferroelectrics*, 2005, **319**:1, 117-123.
- [9] I. Bdikin, M. Wojtas, 2012, private comm. unpubl.
- [10] S.Lang, S.A.M. Tofail et al, *Appl. Phys.Lett.*, 2011, **98**, 123703.
- [11] V. Bystrov, E. Costa, Yu. Dekhtyar et al, No **115**, In: Proceedings of ISAF-ECAPD-PFM 2012. 9-13 July, 2012, Aveiro, Portugal.
- [12] В.С. Быстров. *Биомед. радиоэлектроника*, 2000, **3**, 34-40.

Work is supported by FCT SFRH/ BPD/22230/2005

Применение методов квантовой теории поля для описания взаимодействия заряженной частицы с фононами

Каширина Н.И.⁽¹⁾, Лахно В.Д.⁽²⁾

ИФП НАНУ⁽¹⁾, ИМПБ РАН⁽²⁾

n_kashirina@mail.ru

The polaron functional obtained the quantum field theory methods is studied. Analytical expressions of polaron functional and numerical calculations of polaron energy are carried out for intermediate and strong electron-phonon coupling.

1 Постановка задачи

Теория полярона, посвящённая изучению вопросов, связанных с описанием поведения заряженной частицы в поле фононов является одним из примеров математического описания взаимодействия заряженной частицы с бозонами [1-6]. Многочисленные приложения данной теории в физике, биофизике, биологии привели к тому, что данный раздел науки к настоящему времени стал одним из наиболее востребованных, а методы, применяемые для расчёта энергетического спектра частицы в фононном поле широко применяются не только в физике.

В [6], была построена последовательная теория полярона методами квантовой теории поля. В пределе сильной связи в [6] были получены результаты пекаровской теории полярона [5]. В пределе слабой и промежуточной со стороны слабой связи Тулубом получены результаты близкие в фейнмановским (метод интегрирования по траекториям). В [3] сообщалось о том, что метод, предложенный Тулубом при больших фрёлиховских константах электрон-фононного взаимодействия ($\alpha \geq 30$) даёт наиболее низкие значения поляронной энергии по сравнению с другими методами. Для расчетов энергии связи использовались приближённые выражения, справедливые, строго говоря, в пределе $\alpha \rightarrow \infty$.

Значительный интерес представляет получение точных математических выражений для энергии полярона, полученных в рамках полевой теории. Помимо получения аналитических выражений, необходимых для нахождения минимума

функционала полярона при заданном выборе пробной вариационной функции, необходимо провести численные расчёты для области $5 \leq \alpha \leq 20$, которая представляет значительный интерес как для поляронной, так и для биполяронной теории, построенной на основании применения метода Тулуба [1, 4].

В нашей работе исследуется поляронный функционал Тулуба [6] для двухпараметрической пробной функции

$$f_k = -NV_k \exp(-k^2/2a^2), V_k^2 = 4\pi\alpha/k^2\sqrt{2}, \quad (1)$$

где N и a – вариационные параметры, входящие в функционал Тулуба, k – волновой вектор фонона. В работе Тулуба исследовались приближённые выражения, полученные в пределе $\alpha \rightarrow \infty$, нам удалось получить точные аналитические выражения интегральных форм, входящих в функционал

2 Функционал Тулуба для полярона

Аналитическое выражение для функционала полярона, полученное в [6] для некоторой пробной функции f_k имеет вид:

$$E = \Delta E + 2 \sum_k V_k f_k + \sum_k f_k^2 \quad (2)$$

Первое слагаемое в (2) может трактоваться как кинетическая энергия электрона в фононном поле

$$\begin{aligned} \Delta E &= \frac{1}{2\pi^2} \int_0^\infty dk k^4 f_k^2 \omega_k \frac{1}{2\pi i} \int_C \frac{\sqrt{s}}{(s - \omega_k^2)^2} \frac{1}{D(s)} ds \\ &\equiv \frac{1}{2\pi^2} \int_0^\infty dk k^4 f_k^2 \omega_k F(\omega_k), \quad \omega_k = 1 + \frac{k^2}{2}. \end{aligned} \quad (3)$$

Для вариационной функции (1) подынтегральное выражение можно значительно упростить. Многие интегралы могут быть вычислены в аналитическом виде, интегралы, имеющие полюсы на вещественной оси вычисляются в смысле главного значения:

$$\Delta E(a, N, \alpha) = \Delta E_0(a, N^2\alpha) + \Delta E_1(a, N^2\alpha), \quad (4)$$

$$\Delta E_0 = \frac{3a^2}{16} \frac{1}{1 + \frac{1}{\lambda} - \frac{\sqrt{\pi}}{a} \exp(-\frac{4}{a^2}) \left(\operatorname{erf}\left(\frac{2}{|a|}\right) - 1 \right)},$$

$$\Delta E_1(a, N, \alpha) = \frac{3a^2}{16} \cdot \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{p^2 + 2/a^2}{p^2 + 4/a^2} \frac{\exp(-p^2)(1 - \Omega(p))}{(v(p) + 1/\lambda)^2 + \pi p^2 \exp(-2p^2)/4} dp$$

$$\lambda = \lambda(a, \alpha N^2) = 4 \frac{(\alpha N^2)a}{3\sqrt{2\pi}},$$

$$\Delta E_1(a, N, \alpha) \approx \frac{3a^2}{16} q(a, N^2 \alpha), \text{ для } a, \alpha \gg 1 \Delta E_0(a, N, \alpha) \approx \frac{3a^2}{16},$$

$$v(y) = 1 - ye^{-y^2} \int_0^y e^{t^2} dt - \xi(y) e^{\xi(y)^2} \int_{\xi(y)}^\infty e^{-t^2} dt, \quad \xi(y) = \sqrt{y^2 + \frac{4}{a^2}} \quad (5)$$

$$\Omega(y) = 2y^2 \left[\left(1 + 2\xi(y)^2\right) \xi(y) e^{\xi(y)^2} \int_{\xi(y)}^\infty e^{-t^2} dt - \xi(y)^2 \right] \quad (6)$$

2.2 Предел сильной связи

Для больших констант связи $y \approx \xi(y)$ и выражения (5) и (6) можно заменить на:

$$v(y) = 1 - ye^{-y^2} \int_0^y e^{t^2} dt - ye^{y^2} \int_{\xi(y)}^\infty e^{-t^2} dt \quad (7)$$

$$\Omega(y) = 2y^2 \left[\left(1 + 2y^2\right) ye^{y^2} \int_y^\infty e^{-t^2} dt - y^2 \right] \quad (8)$$

$$(p^2 + 2/a^2)/(p^2 + 4/a^2) \approx 1 \quad (9)$$

2.3 Исследование подынтегральной функции

Рассмотрим, как ведёт себя выражение (5).

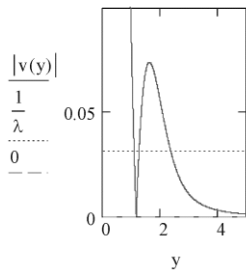


Рис. 1. $\alpha = 200, N = 1$

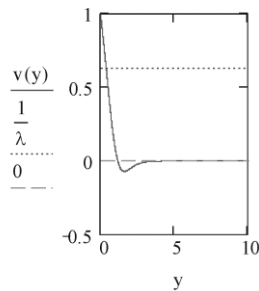


Рис.2. $\alpha = 10, N = 1$

Минимальное значение функции (для параметров, приведенных на рис.1.) $v(y_m) \equiv v_m = -0.07388$ достигается для $y_m = 1.625$. Это означает, что при $1/\lambda \leq v_m$ или $\alpha \leq 21,8$ (для параметра минимизации $a = a_m \approx 0.3\alpha$) выражение $1/\lambda + v(y)$ дважды обращается в ноль. При

$1/\lambda \rightarrow 0$ первый ноль стремится к нулю функции $v(y)$, а второй приводит к появлению расходимости подынтегрального выражения при $y \rightarrow \infty$ (рис.2). Если учесть, что реальный фононный спектр обрезается максимальным волновым вектором k_m , второй минимум может быть опущен. Численные расчёты показали, что предел сильной связи хорошо работает при $\alpha \geq 35$. В пределе сильной связи $a_m/\alpha \approx 0.315268$, $N_m = \sqrt{2}$, $E(a_m)/\alpha^2 \approx -0.125784$. (10)

2.4 Минимизация двухпараметрического функционала для промежуточных значений α

Численная минимизация двухпараметрического функционала (2) для параметров $5 < \alpha < 35$ проводилась без использования приближений (7)-(9). Использование выражений (5), (6) при значениях $5 < \alpha < 8$ (см. Таблицу), наиболее актуальных для вычисления энергии связи и области существования биполярона [1, 4], даёт отклонение от выражения (10), полученного для предела сильной связи, в сторону более глубоких энергий полярона. Однако приведенные значения выше тех, что получены фейнмановским методом интегрирования по траекториям [2]. Поэтому пробная функция, использовавшаяся нами для сравнительно небольших значений параметра электрон-фононного взаимодействия должна быть выбрана в более общем виде.

Таблица

α	a	N	$-E_p(\alpha, a)$
5.0	1.50096	0.762628	4.827907
5.5	2.35223	0.765811	5.333817
6.0	2.36237	0.769500	5.844284
6.5	2.37202	0.773536	6.359477
7.0	2.38137	0.777800	6.879523
7.5	2.39053	0.782201	7.404512
8.0	2.39960	0.786669	7.934505
9.0	2.41767	0.795614	9.009629
10.0	2.43586	0.804359	10.10497

Работа выполнена при поддержке грантов № 11-07-12054 и 10-07-00112.

Список литературы

- [1] V.D.Lakhno. Sol. St. Com. **152** (7), 621-623, 2011.
- [2] R.P. Feynman. Phys. Rev. **97**(3), 660-665, 1955.
- [3] Н.И. Каширина, В.Д.Ляхно, А.В. Тулуб, ЖЭТФ **141**(5), 994- 996, 2012
- [4] В.Д.Ляхно. ЖЭТФ **137**(5), 926- 931, 2010
- [5] Пекар С. И. Исследование по электронной теории кристаллов, Гостехиздат, 1951.
- [6] А. В. Тулуб. ЖЭТФ, **41**(6), 1828-1837, 1961.

Компьютерное моделирование образования комплексов фотосинтетических электрон-транспортных белков

Коваленко И.Б.⁽¹⁾

Московский государственный университет имени М.В.Ломоносова⁽¹⁾
ikovalenko78@gmail.com

We developed a new method for the multiparticle 3D computer simulation of the diffusion and interactions of proteins which gives the possibility to simulate interactions of many protein molecules simultaneously proceeding in a cell. This method was used to study the processes of transient complex formation of the proteins taking part in photosynthetic electron transport chain: plastocyanin and cytochrome f at a variety of ionic strength values in solution and in the chloroplast thylakoid lumen with the account of the electrostatic interactions of proteins with the charged surface of the thylakoid membrane. The method was applied to study the effect of electrostatic interactions on the rates of complex formation between Photosystem 1 and mobile carrier proteins plastocyanin and ferredoxin and to study kinetic characteristics of interaction of *Anabaena* ferredoxin and ferredoxin-NADPH:oxidoreductase. Complex formation of plastocyanin with Photosystem 1 and cytochrome f was also simulated in the case of cyanobacteria. Besides, this method was applied to study the direct electron transfer from cytochrome f to P700 in Photosystem 1 via protein plastocyanin in the luminal space of a chloroplast thylakoid.

1 Введение

Разработан метод прямого компьютерного многочастичного моделирования взаимодействия белков, позволяющий описать образование временных комплексов белков как в растворе, так и в клетке. Целью моделирования является изучение регуляции направленного потока электронов в клетке на основании знания структуры молекул-переносчиков электронов и реакционного объема, а также представлений об элементарных физических процессах, обеспечивающих взаимодействие компонентов системы. В основе предлагаемого подхода лежит метод броуновской динамики. Однако, в методе броуновской динамики рассмотрение ограничено взаимодействием только двух молекул в растворе [8]. Особенностью и новизной нашего метода является возможность с его помощью исследовать взаимодействия нескольких

молекул белков одновременно. Это позволяет моделировать процесс образования большого количества комплексов так, как это происходит в растворе или компартменте клетки, а также наблюдать реальную кинетику этого процесса во времени.

2 Методы исследования

В предлагаемом методе моделирование производится в виртуальном трехмерном реакционном объеме, в котором случайным образом распределены молекулы белков. Для описания движения используется уравнение Ланжевена, описывающее изменение каждой координаты со временем под действием случайной и внешней сил. Для вычисления коэффициентов вязкого трения в модели форма молекулы белка аппроксимируется эллипсоидом вращения, а не сферой, как это обычно принято в моделях броуновской динамики.

Трехмерные модели молекул белков строятся по данным Protein Data Bank. При расчете столкновений белков используется описание формы белков с помощью небольшого количества (10–300) сфер, что дает достаточно реалистичное представление поверхности молекулы для расчета столкновений с другими молекулами и одновременно сокращает время расчетов по сравнению с методами броуновской динамики, в которых используется атомное разрешение поверхности белка. Электростатические взаимодействия между белками учитываются при сближении белков на расстояние менее 35 Å. Белок представлен как область с диэлектрической постоянной $\epsilon = 2$ и пространственно распределенными парциальными зарядами, а для окружающего раствора $\epsilon = 80$. Электростатическое поле, создаваемое зарядами на поверхности белков, было рассчитано по уравнению Пуассона-Больцмана, что позволило учесть различные значения диэлектрической проницаемости белков и раствора [7].

В модели имитируется движение и взаимодействие нескольких сотен молекул белков, что дает возможность наблюдать непосредственно кинетику взаимодействия белков, учитывая взаимодействие нескольких молекул одновременно, и изучать процесс формирования предварительных

комплексов в зависимости от геометрических размеров и формы реакционного объема.

3 Результаты

3.1 Моделирование взаимодействия белков в растворе

Метод применен для описания взаимодействия электрон-транспортных белков в растворе: пластоцианина и цитохрома *f* [1], ферредоксина и ФНР [3], пластоцианина и ферредоксина с ФС1 [2]. Получены зависимости скорости реакции взаимодействия белков от ионной силы и pH раствора, а также точечных мутаций белков. Показано, что электростатические взаимодействия могут играть роль в регуляции скорости фотосинтетического транспорта электронов. Скорость образования комплексов электрон-транспортных белков зависит от ионной силы раствора немонотонно и достигает максимального значения при значениях ионной силы 30-50 мМ [9].

С использованием предлагаемого метода проведено сравнение скорости образования комплексов белков фотосистемы 1 и пластоцианина, цитохрома *f* и пластоцианина в цианобактериях и высших растениях [6]. Показано, что в цианобактериях электростатические взаимодействия играют меньшую роль, чем в высших растениях, так как в цианобактериях скорость образования комплексов пластоцианина с фотосистемой 1 и цитохромом *f* слабо зависит от ионной силы. Электростатические взаимодействия могут значительно ускорить процесс формирования комплекса: скорость реакции связывания пластоцианина с фотосистемой 1 в высших растениях на порядки больше, чем в цианобактериях. Наблюдаемая разница в характерах зависимостей от ионной силы для цианобактерий и высших растений и в абсолютных значениях скоростей реакций обусловлена разным электростатическим потенциалом, создаваемым заряженными аминокислотными остатками на поверхностях белков.

Предлагаемый метод был также применен для определения областей связывания белков и структуры предварительного комплекса и позволил значительно упростить задачу последующего точного моделирования и предсказания структуры образованного финального комплекса [5]. Построенная модель позволяет проводить моделирование взаимодействий белков при различном распределении зарядов на белках, различной ионной силе и pH среды и таким образом предсказывать области связывания для различных белков при разных условиях.

3.2 Моделирование взаимодействия белков в тилакоиде хлоропласта

Построена имитационная компьютерная модель

диффузии и взаимодействия пластоцианина и цитохрома *f* в люминальном пространстве тилакоида хлоропласта [4] с учетом электрического заряда мембраны [7]. С помощью модели исследована зависимость константы скорости взаимодействия белков пластоцианина и цитохрома *f* от размера люминального пространства и сделан вывод о возможности ингибирования реакции при гиперосмотическом стрессе. Показано влияние электрического поля тилакоидной мембраны на процесс диффузии и образования комплекса пластоцианин-цитохром *f* в люмене тилакоида хлоропласта.

Построена модель направленного переноса электрона белком пластоцианином от цитохромного *bf* комплекса на фотосистему 1 в люминальном пространстве тилакоида с учетом неравномерного распределения белковых комплексов на мембране. Проведено сравнение полученных характерных времен реакций с экспериментальными данными, полученными на изолированном хлоропласте шпината. Расчеты показали, что модель корректно воспроизводит процессы образования белок-белковых комплексов, перенос электрона и диссоциации комплексов после короткой вспышки света. С использованием модели исследовали влияние концентрации молекул пластоцианина и геометрической формы тилакоидной мембраны на кинетические характеристики процессов восстановления пигмента P700 фотосистемы 1 и окисления цитохрома *f* цитохромного *bf* комплекса. Показано, что броуновская диффузия и дальнедействующие электростатические взаимодействия в сложном интерьере тилакоидной мембраны обеспечивают физические условия для направленного транспорта электронов при фотосинтезе.

Работа выполнена при поддержке грантов РФФИ: 11-04-01019, 11-04-01268 и 12-04-90834.

Список литературы

- [1] I.B. Kovalenko, A.M. Abaturva et al. *Phys. Biol.*, 2006, **3**, 121–129.
- [2] I.B. Kovalenko, A.M. Abaturva et al. *BioSystems*, 2011, **103**, 180–187.
- [3] I.B. Kovalenko, A.N. Diakonova et al. *Phys. Biol.*, 2010, **7**, 026001.
- [4] И.Б. Коваленко, А.М. Абатурова и др. *Биофизика*, 2008, **53**(2), 261–270.
- [5] И.Б. Коваленко, А.М. Абатурова и др. *ДАН*, 2009, **427**(5), 696–698.
- [6] И.Б. Коваленко, О.С. Князева и др. *ДАН*, 2011, **440**(2), 272–274.
- [7] О.С. Князева, И.Б. Коваленко и др. *Биофизика*, 2010, **55**(2), 259–268.
- [8] D.C. Pearson, E.L. Gross. *Biophys. J.*, 1998, **75**, 2698–2711.
- [9] G.Yu. Riznichenko, I.B. Kovalenko et al. *Biophys. Rev.*, 2010, **2**(3), 101–110.

Математическое моделирование кальциевой сигнализации в тромбоците

Голомысова А.Н.⁽¹⁾, Пантелеев М.А.^(1, 2)

Московский государственный университет имени М.В.Ломоносова, физический факультет, 119991, ГСП-1, Москва, Ленинские горы, 1/2⁽¹⁾, Центр теоретических проблем физико-химической фармакологии РАН, 119991, Москва, ул. Косыгина, 4⁽²⁾
agolomy@yandex.ru

Calcium is in the heart of platelet's activation. Different agonists induce calcium release from intracellular stores, which results in various answers of the cell. Weak activation results in irregular calcium spiking while strong activation leads to the sustained elevation of the cytoplasmic calcium. Here we present a kinetic mathematical model of calcium signaling in platelet, which comprises calcium exchange between several compartments of the cell and includes store-operated calcium entry system. The model allows correct description of calcium concentration pattern for normal activation and shows that for strong activation the platelet, if alive, will have not sustained elevation but very frequent oscillations of calcium concentration.

1 Введение

Тромбоциты представляют собой безъядерные фрагменты клеток, циркулирующие в кровотоке. При появлении активаторов тромбоциты изменяют форму и начинают слипаться друг с другом и со стенками сосуда. Избыточная или недостаточная активация тромбоцитов приводит к серьезным патологиям свертывания крови. Большинство агонистов передает сигнал внутрь тромбоцита посредством увеличения концентрации кальция в цитоплазме, при этом сила сигнала, скорее всего, кодируется частотой и амплитудой колебаний концентрации кальция [5]. Однако механизмы, отвечающие за установление кальциевого сигнала определенной формы в тромбоците, изучены недостаточно [14]. Основным депо кальция внутри тромбоцита является ЭПР¹ и все активаторы вызывают выход ионов кальция из этого компартмента посредством инозитолтрифосфата (IP₃). При этом накапливается все больше информации о роли каналов плазматической мембраны и митохондрий в установлении формы кальциевого сигнала [3, 12]. В частности, при сильной активации ключевую роль играет вход внешнего кальция в цитоплазму через систему SOCE (store-operated calcium entry), состоящую из сенсора на мембране ЭПР – Stim1, связанного с каналом на плазматической мембране – Orai1, который открывается при опустошении ЭПР [4, 14].

В настоящей работе представлена математическая модель кальциевой сигнализации в тромбоците, описывающая обмен ионами кальция между цитоплазмой, ЭПР, митохондриями и внеклеточным пространством. Схема реакций, включенных в модель, представлена на рис. 1.

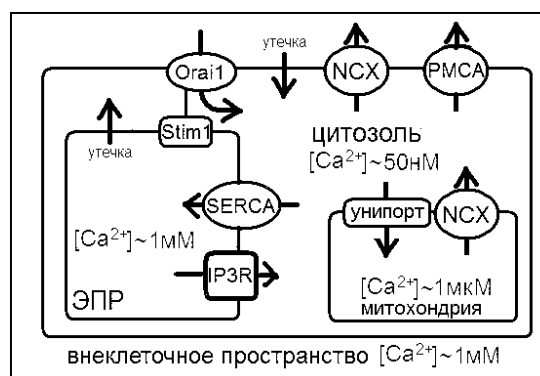


Рис. 1. Схема основных путей обмена ионами кальция между внеклеточным пространством, цитоплазмой, ЭПР¹ и митохондриями. NCX – натрий-кальциевый обменник, IP₃R – рецептор к инозитолтрифосфату, SERCA и PMCA – кальциевые АТФ-азы внутриклеточной и плазматической мембран, соответственно; Orai1 и Stim1 – система входа внешнего кальция при опустошении ЭПР (см. текст).

2 Основные принципы построения математической модели

Математическая модель представляет собой систему обыкновенных дифференциальных уравнений, описывающих изменения концентраций IP₃ в цитозоле и кальция в четырех представленных на рис. 1 компартментах. Рецептор к IP₃ активируется IP₃ при низкой концентрации кальция в цитоплазме и ингибируется кальцием в высоких концентрациях; в тромбоците в основном присутствует тип II [2], кинетика которого описана в работе [13]. Поддержание постоянной концентрации кальция в цитозоле определяется АТФ-азами, расположенными на мембране ЭПР – SERCA; в тромбоците в равных количествах представлены типы 2b и 3a [6], кинетика обоих типов описывается уравнением Хилла [9]. АТФ-аза, находящаяся в

плазматической мембране, PMCA, также может быть описана уравнением Хилла [1]. В настоящей работе использовалась такая же система обмена ионами кальция между митохондриями и цитозолем, как в работе [10], в которой основным каналом для выхода кальция считается натрий-кальцевый обменник, а основным каналом для входа – кальцевый унипорт. Натрий-кальцевый обменник также расположен на плазматической мембране тромбоцита, однако там он отвечает за выход кальция из цитоплазмы [12]. В нашу математическую модель также включена система кальцевых каналов, открывающихся при опустошении кальцевых депо (SOCE), кинетика описана в работе [8].

3 Результаты и обсуждение

Построенная математическая модель позволяет исследовать ответ тромбоцита при активации различной силы. В отличие от существующих математических моделей внутриклеточной сигнализации в тромбоците Purvis et al. [11] и Lenoci et al. [7], в нашей математической модели при возникают осцилляции концентрации кальция в цитоплазме в широком диапазоне значений концентрации IP_3 (от 0.2 до 1 мкМ). Из экспериментов с одиночными тромбоцитами [5] известно, что при увеличении концентрации активатора или применении более сильного активатора частота и амплитуда колебаний концентрации цитоплазматического кальция возрастают. Такое же поведение наблюдается и при увеличении параметра концентрации IP_3 в нашей модели, при этом возрастание амплитуды колебаний от 150 нМ при слабой активации до 400 нМ при сильной численно совпадает с наблюдаемой в экспериментах [5]. Согласно проведенным расчетам в отсутствие IP_3 (покоящийся тромбоцит) концентрация кальция в цитоплазме остается постоянной и равной 10 нМ, что хорошо отражает известные экспериментальные данные (до 50 нМ) [5].

Наиболее интересным результатом расчетов является поведение системы при очень сильной активации. В этом случае происходит сначала установление концентрации кальция на значении 110 нМ, через 60 с снова переходящее в осцилляции высокой частоты (сплошная кривая на рис. 2). Такое поведение системы объясняется тем, что открываются дополнительные каналы (Orai1) для входа внешнего кальция при опустошении внутриклеточных хранилищ (пунктирная кривая на рис. 2). Результаты расчетов показывают, что при сильной активации не происходит полного опустошения внутриклеточных запасов кальция, так как они пополняются за счет тока внешнего кальция и работы кальцевых насосов на мембране ЭПР. Таким образом, постоянно высокая концентрация

внутриклеточного кальция не является физиологическим состоянием тромбоцита, а может возникнуть только при нехватке АТФ для работы насосов, либо при отсутствии достаточного количества кальция вне клетки.

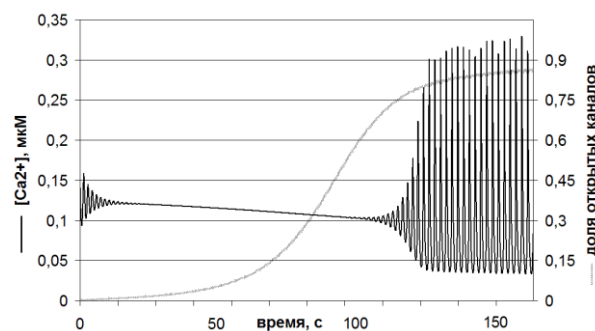


Рис. 2. Расчетные зависимости от времени концентрации кальция в цитоплазме (сплошная линия) и доли открытых каналов Orai1 (штрихованная линия). Начальная концентрация IP_3 равна 1 мкМ.

Список литературы

- [1] Baker H.L., Errington R.J., et al. *Biophys J*, 2002, **82**, 582-590.
- [2] El-Daher S.S., Patel Y., et al. *Blood*, 2000, **95**, 3412-3422.
- [3] Jobe S.M., Wilson K.M., et al. *Blood*, 2008, **111**, 1257-1265.
- [4] Harper M.T., Poole A.W., et al. *Cell Calcium*, 2011, **50**, 351-358.
- [5] Heemskerk J.W.M., Hoyland J., et al. *Biochem J*, 1992, **283**, 379-383.
- [6] Kovacs T., Felfoldi F., et al. *Biochem J*, 2001, **358**, 559-568.
- [7] Lenoci L., Duvernay M., et al. *Mol BioSyst*, 2011, **7**, 1129-1137.
- [8] Liu W., Tang F., et al. *Mathematical Biosciences*, 2010, **228**, 110-118.
- [9] Lytton J., Westlin M., et al. *J Biol Chem*, 1992, **267**, 14483-14489.
- [10] Magnus G., Keizer J. *Am J Physiol Cell Physiol*, 1997, **273**, C717-C733.
- [11] Purvis J.E., Chatterjee M.S., et al. *Blood*, 2008, **112**, 4069-4079.
- [12] Roberts D.E., Matsuda T., et al. *Br J Pharmacol*, 2011, **165**, 922-936.
- [13] Sneyd J., Dufour J.F. *Proc Natl Acad Sci U S A*, 2002, **99**, 2398-2403.
- [14] Varga-Szabo D., Braun A., et al. *J Thromb Haemost*, 2009; **7**, 1057-1066.

¹ ЭПР – эндоплазматический ретикулум

Изменение векторкардиограммы при деформации миокарда

Айду Э.А.И., Трунов В.Г., Титомир Л.И.

Федеральное государственное бюджетное учреждение науки
Институт проблем передачи информации им. А.А. Харкевича Российской академии наук
aidu@iitp.ru

Simple biophysical models that explain changes in the shape of consecutive cardiocycles studied on a large vectorcardiographic database are discussed. In normal cases, the depolarization part of vectorcardiogram (VCG) was invariant under linear transformations for each individual, and only in several cases of cardiac abnormalities, the VCG changes were more complex and nonlinear. In the series of cardiocycles, QRS loop alterations represented as a series of linear transformations reflect changing position, shape, and volume of heart ventricles and are useful for diagnostic analysis of VCG.

1 Введение

При анализе векторкардиограмм (ВКГ), записанных во время параболических полетов самолета-лаборатории Airbus-300 (Novespace-CNES-ESA), было установлено [1], что изменение гравитации приводит к линейным преобразованиям (вращению и дилатации) петли QRS, и только в нескольких случаях с нарушениями сердечной деятельности изменения ВКГ были более сложными и нелинейными (см. раздел 2). Анализ ряда других последовательностей ВКГ подтверждает эти выводы.

Поворот петли QRS можно объяснить поворотом вектора сердца, вызванным таким же поворотом миокарда. В разделе 3 рассматриваются простые биофизические модели, объясняющие дилатацию петли QRS как результат дилатации миокарда, и приводятся соотношения между этими двумя дилатациями. Анализируется первое, линейное приближение, что имеет косвенное экспериментальное подтверждение: в исследованиях [2, 4] на открытом и неповрежденном сердце собаки было установлено, что изменения формы миокарда хорошо приближаются линейными преобразованиями.

2 Анализ последовательностей ВКГ

При анализе последовательности ВКГ изменения петли QRS представлены линейными преобразованиями. Одна из петель QRS выбирается в качестве стандартного образца (эталона),

остальные петли QRS аппроксимируются преобразованиями эталонной петли.

Петля QRS для каждого испытуемого, как кривая в трехмерном векторкардиографическом пространстве, остается практически неизменной при подходящем линейном преобразовании с точностью в пределах измерительных шумов. Почти во всех случаях относительная ошибка аппроксимации была достаточно малой – менее 0,05 (Рис. 1А). Только в нескольких случаях с нарушениями сердечной деятельности относительная ошибка значительно превышала 0,05 (Рис. 1Б).

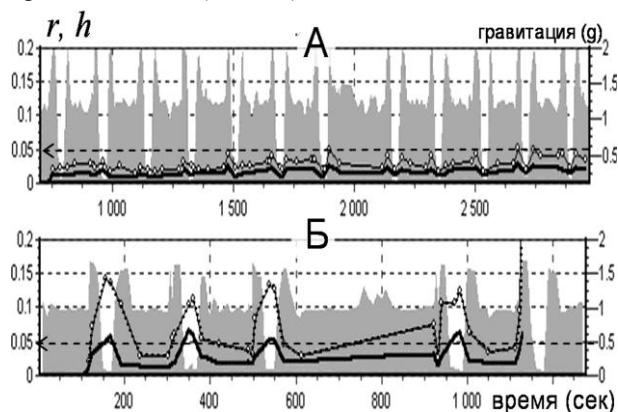


Рис. 1. Относительные ошибки аппроксимации r (жирная линия) и нормализованные расстояния Хаусдорфа h (тонкая линия) для двух человеко-полетов. Приведена диаграмма изменения гравитации во время полета (закрашена серым цветом). А — один из типичных нормальных случаев; Б — случай с нарушением сердечной деятельности, отмечалось нарушение ритма, приведшее к синкопе.

3 Биофизические модели

3.1 Бидоменная модель

Предположим, что миокард (область $\mathcal{M}$ на рис. 2) подвергся некоторой линейной деформации или линейному преобразованию T :

$$\rho = Tr, \quad r = (x, y, z)' \in \mathcal{M}, \quad \rho = (\xi, \eta, \zeta)' \in \mathcal{M}_T \quad (1)$$

ρ — новое положение точки r миокарда после преобразования T .

Рассмотрим изменение вектора сердца для одного момента времени кардиоцикла при преобразовании T . Пусть d , $J(r)$ и d_T , $J_T(\rho)$ —

соответственно, дипольные моменты и плотности источников тока до и после преобразования T . Вектор дипольного момента равен интегралу плотности источников:

$$d = \int_M J(\rho) dv_\rho, \quad d_T = \int_{M_T} J_T(\rho) dv_\rho. \quad (2)$$

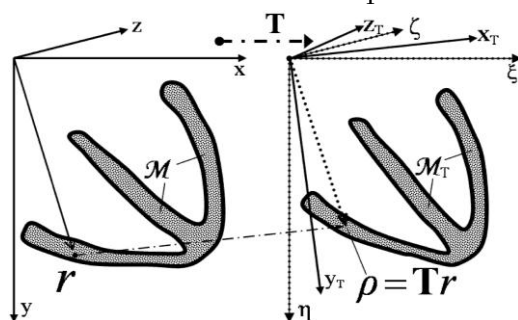


Рис. 2.

Линейная деформация миокарда.

В бидоменной модели возбудимой среды плотность источников тока определяется градиентом трансмембранного потенциала [3], тогда плотность источников тока до и после преобразования выражается следующим образом:

$$J(r) = -\sigma_i \nabla_r U(r), \quad J_T(\rho) = -\sigma_i \nabla_\rho U_T(\rho), \quad (3)$$

где σ_i – внутриклеточная проводимость, $U(r)$, $U_T(\rho)$ – трансмембранные потенциалы до и после преобразования, $\nabla_r = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)$; $\nabla_\rho = \left(\frac{\partial}{\partial \xi}, \frac{\partial}{\partial \eta}, \frac{\partial}{\partial \zeta} \right)$.

Предположим далее, что преобразование T , изменив положение точки, оставило неизменным значение трансмембранного потенциала:

$$U_T(\rho) = U_T(Tr) = U(r). \quad (4)$$

После подстановки (3) в формулу (2) для d_T , замены переменной $r = T^{-1}\rho$ в подынтегральном выражении и простых преобразований получаем равенство, отражающее изменение дипольного момента при линейной деформации T :

$$d_T = (T')^{-1} |T| d, \quad |T| = \det(T). \quad (5)$$

3.2 Двойной электрический слой

Линейное преобразование T миокарда представим в виде произведения ортогонального O и симметрического S преобразований (полярное разложение): $T = SO$. Преобразование S характеризуется тремя ортогональными собственными векторами e_1, e_2, e_3 и соответствующими собственными числами $\lambda_1, \lambda_2, \lambda_3$ – коэффициентами дилатации вдоль собственных векторов.

Пусть двойной электрический слой (или мембрана клетки) пересекает единичный куб в координатах e_1, e_2, e_3 (рис.3). Если интенсивность двойного электрического слоя (или трансмембранный потенциал) остается неизменным, то коэффициенты дилатации дипольного момента в координатах e_1, e_2, e_3 равны соответствующим

площадям граней единичного куба после его преобразования:

$$\sigma_1 = \lambda_2 \lambda_3, \quad \sigma_2 = \lambda_1 \lambda_3, \quad \sigma_3 = \lambda_1 \lambda_2. \quad (6)$$

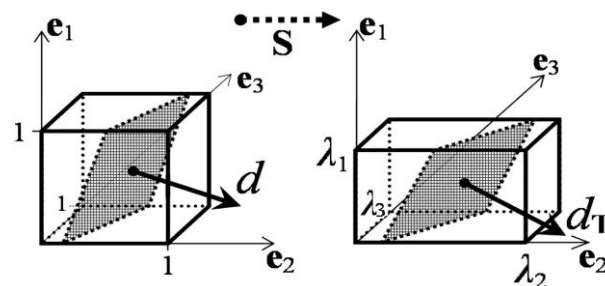


Рис. 3. Дилатация поверхности при преобразовании S .

Те же соотношения будут верны и для суммарного дипольного момента всего двойного электрического слоя. Из соотношений (6) легко вывести равенство (5) и обратно.

4 Заключение

Дилатация (растяжение-сжатие), или линейная деформация возбудимой среды приводит к дилатации дипольной компоненты при условии топологической неизменности процесса возбуждения. И обратно: линейное преобразование петли QRS при сравнении двух кардиоциклов одного и того же индивидуума отражает поворот и деформацию миокарда при отсутствии резких изменений в процессе деполяризации желудочков сердца.

Изменения формы петли QRS в последовательности кардиоциклов, представленные последовательностью линейных преобразований, отражают изменения положения, формы и объема желудочков сердца и полезны при диагностическом анализе ВКГ.

Работа поддержана Российским фондом фундаментальных исследований (проект № 11-01-00806)

Список литературы

- [1] Aidu E.A.I., Trunov V.G., Titomir L.I., Capderou A., Vaïda P. Transformation of vectorcardiogram due to gravitation alteration. *Measurement Science Review*, 2003, 3(2), 29-32.
- [2] Grover M., & Glantz S.A. Endocardial Pacing Site Affects Left Ventricular End-Diastolic Volume and Performance in the Intact Anesthetized Dog. *Circ. Res.*, 1983, 53, 72-85
- [3] Miller W.T., Geselovitz D.B. Simulation studies of the electrocardiogram. 1. The normal heart. *Circ. Res.*, 1978, 43, 301-315.
- [4] Walley K. R., Grover M., Raff G.L., Bengt J.W., Hannaford B. Glantz S. A. (1982). Left ventricular dynamic geometry in the intact and open chest dog. *Circ. Res.* 50, 573-589.

Структура и свойства льда при динамометаморфизме и его моделирование

Быстрова В.В.⁽¹⁾, Парамонова Е.В.⁽²⁾, Быстров В.С.⁽²⁾

Географический факультет Московского государственного университета им.

М.В.Ломоносова, Ленинские горы, Москва, Россия ⁽¹⁾,

Институт математических проблем биологии РАН, Пущино, Россия⁽²⁾

vvbystrova@mail.ru

The structure and properties of ice under dynamic metamorphism are reviewed in this presentation. The several computational approaches for ice modeling with combined first principal, semi-empirical and molecular dynamics methods are discussed.

Исследование структуры и свойств льда при различных условиях и динамических нагрузках представляет значительный интерес как для общих наук о Земле, геологии, так и для практики освоения арктических территорий России. Компьютерное моделирование является здесь одним из важнейших инструментов в проведении этих исследований. В геологии динамометаморфизмом льда называется преобразование его петрографической структуры в процессе деформирования и движения под влиянием внешних динамических воздействий, собственного веса и колебаний температуры. Этот динамометаморфизм играет существенную роль при развитии подземных и поверхностных льдов. При повышенных напряжениях в таких промерзающих и мерзлых грунтах формируются внутригрунтовые льды со сложной композитной структурой. Физика данного процесса такова, что при нагрузках на лед, превышающих его предел упругости, развиваются вязкопластические и хрупкие деформации. Экспериментальные структурно-петрографические наблюдения механизмов деформации льда проводились достаточно давно. Однако, до сих пор остается неизученным вопрос об условиях перехода от пластических деформаций к хрупким, имеющий довольно резкий и зачастую непредсказуемый характер, что крайне важно для практики.

На Земле лёд представлен в основном в гексагональной кристаллической модификации I_H , в котором каждая молекула воды окружена четырьмя ближайшими к ней молекулами, находящимися на одинаковых расстояниях ($2,76 \text{ \AA}$) и размещённых в вершинах правильного тетраэдра. Эти молекулы в

решетке кристалла льда связаны как водородными, так и химическими связями, а также слабыми силами Ван-дер-Ваальсовых взаимодействий. В природе лёд встречается в различном виде - материкового, плавающего, морского и подземного льда. Механические свойства льда особенно важны, т.к. под действием собственного веса лёд приобретает пластические свойства и текучесть. Лёд может содержать и механические примеси: твёрдые частицы, капельки растворов, пузырьки газов. Наличием кристалликов солей и капелек рассола определяется особенность морского льда, который имеет специфический характер ползучести, выраженный в развитии определенных, резко разграниченных стадий деформации. Поведение льда под нагрузкой определяется общим принципом минимума свободной энергии: напряженные тела релаксируют путем перекристаллизации с изменением внешней формы и объема, а также формы, ориентировки, размеров кристаллов, либо путем частичного или полного разрушения.

Моделирование структуры и свойств льда при различных условиях позволяет выявить особенности этих процессов. Моделированию льда посвящено много работ. Однако особенности механических свойств при различных деформациях пока еще изучены крайне недостаточно. Как видно из существующих экспериментальных данных, здесь много тонких переходных процессов. В данной работе мы применяем моделирование различными методами, заложенными в программный пакет HyperChem, а также используем данные моделирования из первых принципов в сочетании с методами молекулярной динамики. Это позволяет оценить важные особенности переходных процессов. Данные результаты необходимы для изучения и прогнозирования динамики метаморфных процессов морского подземного льда.

От маятника к ДНК

Якушевич Л.В.

Институт биофизики клетки РАН
yakushev@icb.psn.ru

The models of pendulum, of two couple pendulums and of several coupled pairs of pendulums are used to study oscillation dynamics of one isolated base, of a base pair and of several base pairs coupled by stacking interactions.

1 Введение

В этой работе представлены полученные недавно результаты по моделированию колебательной динамики коротких фрагментов молекулы ДНК, начиная от отдельных оснований и пар оснований и заканчивая короткими (длиной в несколько пар оснований) фрагментами двойной цепи ДНК (рис. 1, слева).

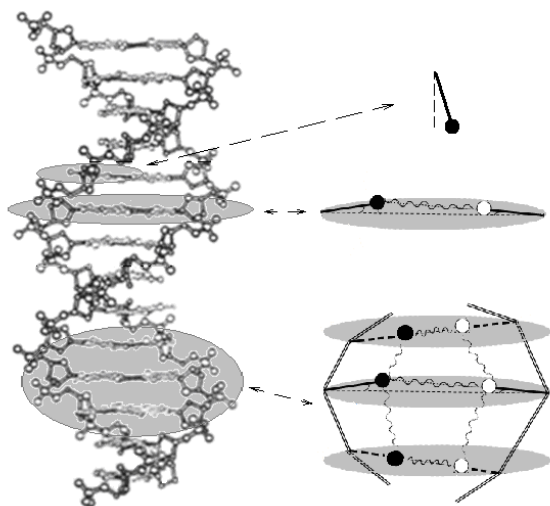


Рис. 1. Двойная спираль ДНК с выделенными фрагментами (слева) и модели, имитирующие колебательную динамику этих фрагментов (справа).

Материал расположен таким образом, чтобы проследить развитие моделирования от простейших моделей, таких как модель маятника или модель двух связанных маятников до более сложных моделей, таких, например, как модель трех пар связанных маятников, расположенных по спирали (рис. 1, справа).

2 Маятник как модель колебаний отдельного основания

Выделим небольшой элемент структуры ДНК, содержащий только одно азотистое основание (рис.

1, слева). Пренебрежем внутренними степенями свободы основания, и будем рассматривать его как твердое тело массы m . Расстояние от центра массы основания до сахаро-фосфатной цепочки обозначим буквой R . Сахаро-фосфатную цепочку, к которой крепится основание, промоделируем горизонтальной нитью, а потенциал, в котором движется основание, - гравитационным полем Земли. Тогда, задачу о колебаниях азотистого основания ДНК можно свести к задаче о колебаниях маятника (рис. 1, справа) и записать Лагранжиан и уравнение колебаний в виде:

$$L = \frac{I}{2} \dot{\varphi}^2 - V(1 - \cos \varphi); \quad (1)$$

$$I\ddot{\varphi} + V \sin \varphi = 0; \quad (2)$$

где $\varphi(t)$ - угол отклонения азотистого основания от положения равновесия, I - момент инерции основания, V - коэффициент, характеризующий потенциал внутреннего поля вблизи этого основания. Так как уравнение (2), а значит и его решения, инвариантны относительно смещения на период 2π , а также симметричны относительно оси ординат, можно ограничиться рассмотрением функции φ в интервале $-\pi < \varphi \leq \pi$.

Определим координаты особых точек на фазовой плоскости $\{\varphi; \dot{\varphi}\}$ в указанном выше интервале [3]:

$$Q_1 = \{0; 0\}; Q_2 = \{\pi; 0\}.$$

Легко показать, что первая из этих точек устойчивая, а вторая неустойчивая. Фазовые траектории системы в окрестности особых точек приведены на рис. 2. Расчеты проведены для аденина. Наиболее интересен для биофизики ДНК случай движения модельной системы вблизи неустойчивой точки. Начальное отклонение аденина на угол $\varphi(0) = \pi$, $\dot{\varphi} = 0$ можно трактовать как раскрытие пары АТ и образование открытого состояния. Такие состояния являются важным и необходимым элементом процессов белок-нуклеинового узнавания.

3 Пара связанных маятников как модель колебаний пары оснований ДНК

Выделим небольшой элемент структуры ДНК, содержащий пару азотистых оснований (рис. 1, слева). Пренебрежем внутренними степенями свободы оснований, и будем рассматривать их как

твердые тела массы m_1 и m_2 . Расстояние от центров масс оснований до соответствующих сахаро-фосфатных цепочек обозначим буквами R_1 и R_2 .

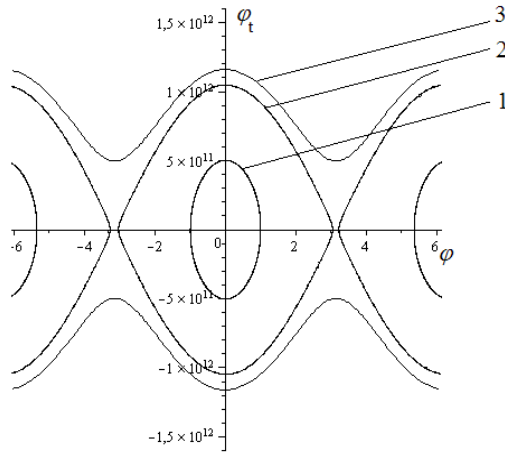


Рис. 2. Фазовый портрет основания (аденина). Начальные условия: $\varphi = 1$, $\varphi_t = 0$ (кривая 1); $\varphi = \pi - 0.1$, $\varphi_t = 0$ (кривая 2); $\varphi = \pi - 0.1$, $\varphi_t = 5 \cdot 10^{11}$ (кривая 3).

Расстояние между центрами масс оснований в положении равновесия ($\varphi_1 = \varphi_2 = 0$) обозначим буквой $\underline{a}$. Взаимодействие между основаниями внутри пары промоделируем пружинкой с жесткостью K_{12} . Тогда задачу о колебаниях пары азотистых оснований ДНК можно свести к задаче о колебаниях двух связанных маятников (рис. 1, справа) и записать функцию Лагранжа и уравнения движения для такой модельной системы в виде [1]:

$$L = \frac{I_1}{2} \left(\frac{d\varphi_1}{dt} \right)^2 + \frac{I_2}{2} \left(\frac{d\varphi_2}{dt} \right)^2 - [U_1(1 - \cos \varphi_1) + U_2(1 - \cos \varphi_2) + U_{12}(1 - \cos(\varphi_1 + \varphi_2)) + U_0] \quad (3)$$

$$I_1 \frac{d^2 \varphi_1}{dt^2} + U_1 \sin \varphi_1 - U_{12} \sin(\varphi_1 + \varphi_2) = 0;$$

$$I_2 \frac{d^2 \varphi_2}{dt^2} + U_2 \sin \varphi_2 - U_{12} \sin(\varphi_1 + \varphi_2) = 0; \quad (4)$$

где $\varphi_1(t)$ и $\varphi_2(t)$ - углы отклонений оснований, I_1 и I_2 - моменты инерции этих оснований, $U_1 = K_{12}R_1(R_1 + R_2 + \underline{a})$, $U_2 = K_{12}R_2(R_1 + R_2 + \underline{a})$, $U_{12} = K_{12}R_1R_2$,

$$U_0 = \frac{K_{12}\underline{a}^2}{2}.$$

Такая модельная система обладает четырьмя особыми точками на фазовой плоскости в области, ограниченной условием $-\pi < \varphi_1, \varphi_2 \leq \pi$, [1]:

$$Q_1 = \{0; 0; 0; 0\}, Q_2 = \{\pi; 0; 0; 0\},$$

$$Q_3 = \{0; \pi; 0; 0\}, Q_4 = \{\pi; \pi; 0; 0\}.$$

Первая из этих точек устойчивая, а остальные три точки неустойчивые. На рис. 3 показана траектория движения модельной системы в конфигурационном пространстве $\{\varphi_1; \varphi_2\}$ с начальными условиями, взятыми вблизи второй точки. Расчеты выполнены для пары АТ.

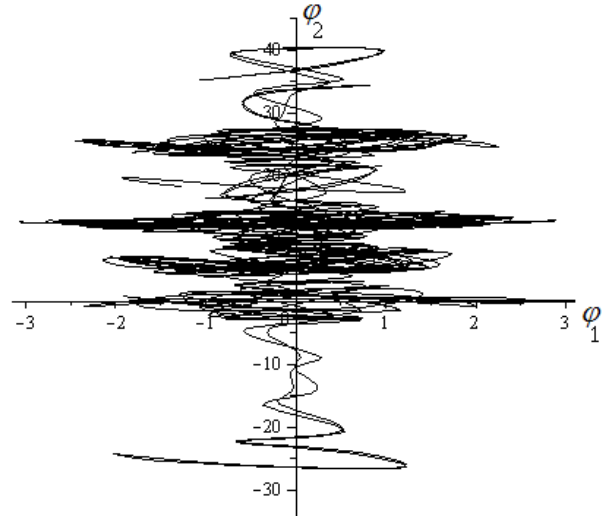


Рис. 3. Траектория движения пары АТ в конфигурационном пространстве $\{\varphi_1; \varphi_2\}$. Начальные условия: $\varphi_1 = \pi - 0.01$; $\varphi_2 = 0$; $\varphi_{1,t} = \varphi_{2,t} = 0$.

4 Три пары связанных маятников как модель небольшого фрагмента ДНК

Действуя аналогичным образом легко построить модель небольшого фрагмента ДНК, состоящего, например, из трех пар оснований (рис. 1, слева). Мы показали, что такой фрагмент ДНК можно промоделировать тремя парами маятников, повернутыми на 36° друг относительно друга и связанных вертикальными и горизонтальными пружинками (рис. 1, справа). Горизонтальные пружинки использовались для имитации водородных связей, а вертикальные - для имитации стэкинговых взаимодействий. Были рассчитаны координаты особых точек такой системы [2] и построены фазовые траектории в окрестности неустойчивых особых точек.

Список литературы

- [1] L. V. Yakushevich, S. Gapa., J. Awrejcewicz. *International Journal of Bifurcation and Chaos*, 2011, **21**(10), 3063-3071.
- [2] L.V. Yakushevich. *Journal of Nonlinear Mathematical Physics*, 2011, **18**(2), 449-461.
- [3] А.П. Кузнецов, С.П. Кузнецов, Н.М. Рыскин *Нелинейные колебания*. М.: Физматлит, 2002, 315 с.

Трёхмерная математическая модель электромеханической функции левого желудочка сердца человека

Правдин С.Ф. (1, 2, 3), Бердышев В.И. (1, 4), Панфилов А.В. (3),
Кацнельсон Л.Б. (2, 4), Соловьёва О.Э. (2, 4), Мархасин В.С. (2, 4)

Институт математики и механики Уральского отделения РАН (1),
Институт иммунологии и физиологии Уральского отделения РАН (2)

Гентский университет, Бельгия (3),
Уральский федеральный университет (4)
o.solovyova@iip.uran.ru

We developed a model of geometry and anisotropy, electrophysiology and mechanics of the left ventricle (LV) of the human heart. The morphology part is based on an analytical description of the LV muscular layers and fibers. The electrophysiological model is the phenomenological model of Aliev and Panfilov. And the mechanical part includes Jarrousse's method of 'Virtual Hexagon', Hunter's law of passive myocardium, and a space-uniform blood pressure law. The initial electrical stimulus was given to lower-half of the LV subendocardium. One LV cycle consists of 4 phases, two of them are isovolumetric, and the other two are auxotonic ones. We calculated one LV beat and measured the ejection fraction, which was 57 %. We observed the typical LV twist/torsion during the systole-diastole.

Мы разработали модель геометрии и архитектоники ЛЖ сердца человека, основанную на представлении ЛЖ в виде тела вращения, заполненного поверхностями (мышечными слоями), каждая из которых, в свою очередь, заполнена кривыми (мышечными волокнами). Заполнение тела слоями и волокнами, имеющими нулевую толщину, непрерывно. Все слои представляют собой идентичные поверхности, а заполнение тела достигнуто путём вращения этой поверхности вокруг вертикальной оси симметрии модельного ЛЖ.

Направление волокна в точке ЛЖ – это направление касательной к кривой, проходящей через данную точку.

Гарантируется, что через все точки, не лежащие на оси симметрии, проходит ровно одна поверхность и ровно одно волокно, имеющее касательную.

Расчёты электрофизиологии мы провели в некоторой специальной системе координат, где уравнения поверхностей ЛЖ записываются особенно просто.

Сетка была равномерной по этим специальным координатам, но неравномерной по декартовым координатам. Граничное условие было стандартным: $nD \text{ grad } u = 0$ (здесь n – нормаль к поверхности, D – матрица диффузии, u – потенциал).

Электрическую анизотропию мы смоделировали путём использования двух различных коэффициентов диффузии: вдоль волокна (12) и поперёк (4/3).

При моделировании электрофизиологической и механической активности ЛЖ мы использовали известную модель Алиева–Панфилова [1], содержащую лишь 3 фазовые переменные. Сетка для электрических расчётов имела шаг по пространству около 0,3-0,7 мм, а для механических – в 4 раза больше.

Начальное условие в электрофизиологической части задачи: нижняя половина эндокардиального слоя узлов имеет повышенное значение потенциала, в остальных точках – потенциал покоя.

В модели была учтена только прямая электромеханическая связь, поэтому электрическая сетка была оставлена неподвижной.

При моделировании механической активности ЛЖ мы учли следующие силы: силу тяжести, силу вязкого трения, пассивную жёсткость миокарда [2], активную силу, действующую вдоль волокон, и силу давления крови со стороны полости ЛЖ.

Механическая сетка состояла из тетраэдров, при расчёте активных и пассивных сил мы использовали метод «виртуального шестигранника» [3].

Давление крови было принято постоянным (по пространству) во всей полости ЛЖ. Зависимость давления крови от времени включала в себя следующие 4 фазы.

Фаза пассивного наполнения. Длительность её фиксирована (300 мс), давление мы положили линейно растущим со временем от 0,01 до 0,03. Эта фаза завершалась в момент задания повышенного значения потенциала в соответствующих узлах ЛЖ.

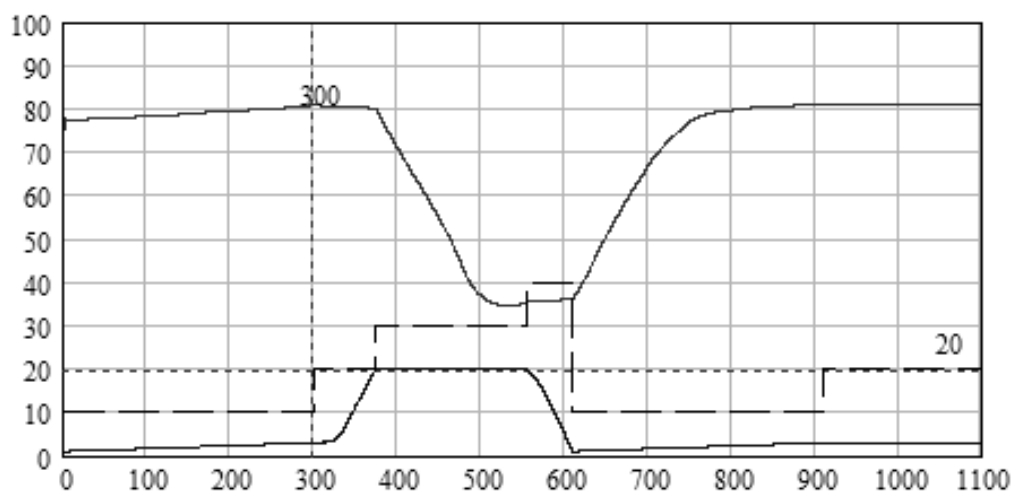


Рис. 1. Объем полости ЛЖ (в мл, красная линия), давление крови в полости ЛЖ (в Па*10, синяя линия) в зависимости от времени (в мс). Чёрная пунктирная линия – номер фазы модельного сердечного цикла (x10). Отметка «300» – момент установки повышенного значения потенциала в узлах нижней части субэндокарда ЛЖ. Отметка «20» – уровень давления крови в аорте

Фаза изоволюмического напряжения. Кровь здесь была представлена как трудно сжимаемая среда, эта фаза продолжалась, пока давление не достигло порога 0,2.

Фаза изгнания. Сопротивление крови и артерий мы не учитывали, давление крови было зафиксировано, на каждом шаге по времени мы вычисляли объем полости ЛЖ. Изгнание крови заканчивалось в тот момент, когда объем полости начинал расти.

Фаза изоволюмического расслабления. Кровь здесь снова была представлена как трудно сжимаемая среда. Эта фаза продолжалась, пока давление не падало ниже порога 0,01. После окончания этой фазы вновь начиналось пассивное наполнение полости ЛЖ кровью.

График зависимости объема полости ЛЖ и давления крови в ней от времени представлен на рис. 1. Фракция выброса составила 57 %.

Также мы отслеживали положение нескольких волокон миокарда в процессе движения ЛЖ. Укорочение волокон менялось от 5 до 30 %.

Работа поддержана грантами: 12-M-14-2009 (Президиум УрО РАН), 1F2B8M/JDW/2010-2011/10-BTL-RUS-01 (Фламандское Сообщество Бельгии), 01SF1511 (Гентский Университет).

Список литературы

- [1] R.R. Aliev, A.V. Panfilov. *Chaos, Solitons and Fractals*, 1996, 7(3), 293-301.
- [2] P. Hunter, M.P. Nash, G.P. Sands. *Computational Electromechanics of the Heart*. In: *Comput. Biology of the Heart*. Chichester: John Wiley & Sons, 1997.
- [3] O. Jarrousse. *Modified Mass-Spring System for Physically Based Deformation Modeling*. PhD Dissert. Karlsruhe, 2011.

О решении проблемы общего ритма в автоколебательной активной среде Винера-Розенблюта

Мазуров М.Е.

Московский государственный университет экономики, статистики и информатики, Москва,
Нежинская, 7
mazurov37@mail.ru

According to the assumption of N. Wiener - A. Rosenbluet, acad. I.M. Gelfand in the active self-oscillatory medium united rhythm is determined by element with the highest rhythm. However, the theorem about the united rhythm was not formulated and proven, it formulated and proven in this work. The proof of theorem is based on the use of theory of the synchronization of relaxation hunting systems, theorem of Kronecker, which is been base theorem in the theory of almost periodic functions. Theorem shows that there is a denumerable set of versions of rhythmical activity. In the particular case united rhythm is caused by the high-frequency element. Theorem shows that the very simple system can possess very complex properties.

Автоколебательная активная среда Винера - Розенблюта была предложена в 1945 году [1] и использована для решения проблемы формирования единого ритма сердца, известной более 100 лет [7]. Ритм сердца создается в синоатриальном узле, состоящем из десятков тысяч автоколебательных клеток, каждая из которых генерирует свой собственный ритм. Ритмы всех клеток синхронизируются и возникает единый ритм для всей совокупности. Экспериментально установлено, что доминирующую роль играют клетки с наивысшим ритмом [7]. Проблема единого ритма существенна в хронобиологии для синхронизации биологических часов; в социальных средах при формировании фигуры лидера; в распределенной экономике при формировании динамики рыночной конкуренции [4]. Обоснование доминирующей роли элемента с наивысшим ритмом впервые предложено Винером – Розенблютом в аксиоматической теории возбудимой среды, которая задается аксиомами [1]:

1. Каждая точка среды находится в одном из трех состояний: а) покоя; б) возбуждения; в) рефрактерности. 2. Возбуждение распространяется по покоящейся среде с постоянной скоростью. 3. Точки среды могут обладать автоколебательной активностью (дополнительная аксиома).

Согласно предположению Н. Винера - А. Розенблюта [1], акад. И.М. Гельфанда и М.Л. Цетлина в 1960 г. в активной автоколебательной среде единый ритм определяется элементом с наивысшим ритмом [1, 7]. Впоследствии многие авторы ссылались на это утверждение, воздвигнутое в ранг парадигмы. Однако, в работах Н. Винера - А. Розенблюта, И.М. Гельфанда теорема о едином ритме не только не доказана, но даже не сформулирована [1, 7]. Теорема о едином ритме возбудимой автоколебательной среды сформулирована и доказана в данной работе [4].

Определение (единого ритма и его периода). Пусть один из автоколебательных элементов с периодом T_k синхронизирует все остальные элементы с периодами T_i ($i=1,...,n; i \neq k$) в соотношении m_i/n_i . Единый ритм n автоколебательных элементов Винера - Розенблюта с периодами T_i ($i=1,...,n$) равен $T_{aa} = \max m_i T_k$ - наибольшему периоду колебаний «неправильной формы»

Теорема. Пусть возбудимая среда обладает свойствами 1,2, и состоит из n автоколебательных возбудимых элементов с периодами $T_1, ..., T_n \in \{T\}$. Рассмотрим систему диофантовых неравенств

$$|n_i T_i - n_j T_j| < \varepsilon; \quad i, j \in [1, ..., n]; \quad n_i, n_j \in \{N\}.$$

Пусть $n_{01}, ..., n_{0l}$ - минимальное решение диофантовых неравенств. Тогда период единого ритма на интервале ε -почти-периода T_ε находится в соответствии с определением.

Следствие. 1. Если $n_1 = n_2 = ... = n_n = 1$;

$$T_1 = \min_n T_n \text{ и } T_1 \in \bigcup_{i=2}^n [T_i - \varepsilon, T_i], \text{ то единый ритм}$$

$T_0 = T_1$. 2. Если $n_1 \neq 1$; $n_2 = n_3 = ... = n_n = 1$;

$$T_2 = \min_n T_n \text{ и } T_2 \in \bigcup_{i=3}^n [T_i - \varepsilon, T_i], \text{ то единый ритм}$$

$T_0 = T_2$ и т.д. Возможен случай, когда единый ритм определяется элементом с наименьшим ритмом.

Доказательство теоремы основано на использовании теоремы Кронекера, являющейся базовой теоремой в теории почти-периодических функций [3, 4], поскольку в общем случае периоды $T_i \in \{T\}$ являются несоизмеримыми. Из неравенств выше следует

$$|t - n_i T_i| < \varepsilon; \quad (i = 1, \dots, n); \quad n_i \in \{N\}$$

По следствию из теоремы Кронекера эти неравенства имеют всюду плотное множество ε -почти-периодов и соответствующих решений. Если существуют начальные сдвиги T_{0i} автоколебаний элементов активной среды, то стационарной синхронизации предшествует переходный процесс, заканчивающийся с одновременным возбуждением всех элементов активной среды [3, 4].

Для иллюстрации возможностей теоремы рассмотрены различные варианты формирования единого ритма в вычислительном эксперименте. Была сформирована автоколебательная аксиоматическая активная среда в области, размещенной в узлах прямоугольной решетки с бесконечно малым шагом. В первом эксперименте точечные водители ритма 1, 2, 3 с периодами автоколебаний T_1, T_2, T_3 возбуждают аксиоматическую активную среду с периодом автоколебаний T_4 . Рассмотрены 4 случая. В первом случае при $T_1 = 200, T_2 = 210, T_3 = 215, T_4 = 220, \varepsilon = 25, \tau = 1$, т.е. $T_1 < T_2 < T_3 < T_4$, выполняется следствие 1 теоремы и единый ритм формируется элементом 1 с наивысшим ритмом $T_1 = 200$. Во втором случае при $T_1 = 184, T_2 = 210, T_3 = 215,$

$T_4 = 220$, выполняется следствие 2 теоремы и единый ритм формируется элементом 2 с ритмом меньше наивысшего с периодом $T_2 = 210$. В случае 3 при $T_1 = 163, T_2 = 189, T_3 = 215, T_4 = 220, \varepsilon = 25, \tau = 1$ единый ритм формируется элементом 3 с наименьшим ритмом $T_3 = 215$. В случае 4 при $T_1 = 140, T_2 = 166, T_3 = 192, T_4 = 220, \varepsilon = 25, \tau = 1$ единый ритм определяется ритмом элементов области и равен $T_4 = 220$.

Теорема о едином ритме в автоколебательной активной среде Н. Винера - А. Розенблюта показывает, что существует счетное множество вариантов ритмической активности, оно может быть охарактеризовано диаграммой Кантора. В частном случае единый ритм обусловлен наиболее высокочастотным элементом. Эта теорема показывает, что очень простая система может обладать очень сложными свойствами.

Список литературы

- [1] N. Wiener, A. Rosenblueth *Arch. Inst. Cardiologia de Mexico*. 1946. V. 205. № 16. 3-4.
- [2] И. М. Гельфанд, М. Л. Цетлин *ДАН СССР*. 1960. Т. 131. № 6. С. 1242-1245.
- [3] Б. М. Левитан Почти-периодические функции ГИТТЛ. 1953. 396 с.
- [4] М. Е. Мазуров *Экономика, Статистика и Информатика*. Вестник УМО. 2011. №2. С. 191-195.
- [5] М. Е. Мазуров *ИВУЗ. сер. Математика*. 1989. №7. С. 13-20.
- [6] М. Е. Мазуров *ЖВМ и МФ АН СССР*. 1991. Т. 31. № 11. С. 1619-1636.
- [7] М. Г. Удельнов *Успехи совр. биол.* 1957. Т. 44. №. 3 (6).

Физика и математика движения автоволн в нелинейных реакционно-диффузионных системах: от парадоксов к пониманию

Морнев О.А.

Институт теоретической и экспериментальной биофизики РАН
mornev@mail.ru

The physical mechanisms supporting propagation of autowaves in nonlinear active media with chemical reactions and diffusion are investigated. The paradoxes that arise at studying of movement of such waves are explained.

В большинстве физически интересных случаев описание движения фронта плоской концентрационной автоволны, распространяющейся вдоль длинного цилиндрического реактора, заполненного подверженной диффузии химически активной автокаталитической средой, сводится к паре уравнений $u_t + (J_D)_x = f(u)$, $J_D = -Du_x$; здесь $u = u(x, t)$ и $J_D = J_D(x, t)$ – концентрация и диффузионный поток активатора в точке x реактора в момент времени t , D – коэффициент диффузии активатора, $f(u)$ – N -образная кинетическая функция с тремя нулями u_- , u_b , u_+ ($u_- < u_b < u_+$, $f'(u_-) < 0$, $f'(u_b) > 0$, $f'(u_+) < 0$), описывающая зависимость скорости производства активатора в единице объёма реактора (или в другой нормировке – на единице его длины) от концентрации u .

Первое из написанных уравнений есть уравнение неразрывности с источниками, а второе выражает известный феноменологический закон диффузии; исключение переменной J из обоих уравнений приводит к известному реакционно-диффузионному уравнению $u_t = Du_{xx} + f(u)$, являющемуся основой описания концентрационной динамики активатора в реакторе. Эта динамика является динамикой распределённой бистабильной системы с порогом. Действительно, в рассматриваемом случае реактор имеет два устойчивых пространственно-однородных стационарных состояния, определяемых нулями u_- и u_+ функции $f(u)$ (им соответствуют концентрационные профили $u(x, t) \equiv u_-$ и $u(x, t) \equiv u_+$), которые отделены друг от друга неустойчивым пороговым пространственно-однородным стационарным состоянием, отвечающим промежуточному нулю u_b ,

указанной функции (ему соответствует концентрационный профиль $u(x, t) \equiv u_b$).

Профиль автоволны в бистабильном реакторе имеет вид бегущего фронта – концентрационного перепада, по левую сторону которого концентрация активатора экспоненциально близка одному из двух указанных устойчивых стационарных значений (например, высокому значению, равному u_+), а по правую сторону – другому устойчивому стационарному значению (например, низкому значению, равному u_-). В этом случае автоволна, движущаяся вдоль реактора слева направо, переключает последний из состояний с низким значением концентрации активатора в состояние с высоким значением этой величины.

Каков же физический фактор, поддерживающий движение автоволны и управляющий её скоростью? Стандартный ответ: этим фактором является диффузионный поток активатора, индуцируемый наличием градиента концентрации u в окрестности фронта. Пояснение к ответу: упомянутый поток переносит активатор в сторону убывания градиента u (в нашем случае слева направо), что приводит к повышению значения u перед фронтом до надпороговой величины и к запуску автокаталитического процесса производства активатора справа от фронта. В результате концентрация u перед фронтом нарастает, и профиль последнего претерпевает перестройку, проявляющуюся в виде автомодельного сдвига фронта в направлении диффузионного потока. И чем сильнее диффузионный поток (т. е. чем выше крутизна фронта), тем быстрее движется автоволна.

Однако изложенные аргументы сталкиваются с парадоксом: они не могут объяснить ни движение фронта автоволны в сторону, противоположную направлению диффузионного потока, ни существование стоячих концентрационных фронтов (и те, и другие реализуются в бистабильных реакционно-диффузионных средах при некоторых значениях параметров кинетической функции $f(u)$).

Для разрешения указанного парадокса детально проанализированы физические механизмы, поддерживающие распространение автоволн.

Показано, что физическим фактором, непосредственно управляющим скоростью и направлением движения фронта автоволны, является не диффузионный поток, а *дисбаланс мощностей пространственных зон производства и деструкции активатора*, измеряемый разностью $\Delta P = P_+ - P_-$, где P_+ (P_-) есть суммарное количество активатора, появляющееся (исчезающее) в единицу времени на всём участке реактора, охватываемом зоной производства (зоной деструкции). Указанный дисбаланс определяет величину и направление *потока наполнения среды активатором* (коротко – *наполняющего потока*) J_F , численно равного суммарному количеству активатора, появляющегося/исчезающего в единицу времени на всей длине реактора. Наполняющий поток, как нетрудно показать, удовлетворяет двум соотношениям

$$J_F = Av, \quad J_F = \Delta P \equiv P_+ - P_-,$$

где $A = u_+ - u_-$ есть амплитуда автоволны, v – её скорость. Из этих соотношений вытекает формула

$$v = A^{-1} J_F = A^{-1} \Delta P,$$

прямо указывающая на тот факт, что величина и знак скорости автоволны управляются дисбалансом мощностей зон производства/деструкции активатора, предопределяющим величину и направление наполняющего потока.

Теперь становится понятной ошибочность наивного ожидания, что движение автоволнового фронта поддерживается и управляется именно (и только лишь) диффузионным потоком. Но тогда возникает вопрос о взаимном балансе потоков J_F и J_D : как соотносятся оба потока? Ответом служит общее *соотношение дуальности диффузионного и наполняющего потоков*

$$J_F \langle J_D \rangle_u = D \int_{u_-}^{u_+} f(u) du,$$

где $\langle J_D \rangle_u$ – среднее значение диффузионного потока в интервале концентраций (u_-, u_+) , охватывающем изменения u на профиле фронта.

Из этого соотношения, в частности, следует, что если изменения параметров кинетической функции, сохраняя интеграл в правой части написанного соотношения, вызывают усиление диффузионного потока J_D , то при этом наполняющий поток J_F будет ослабляться. Поэтому управляя кинетическими параметрами химической подсистемы реактора подходящим образом, можно добиться того, что крутизна фронта автоволны и генерируемый фронтом диффузионный поток будут неограниченно возрастать, тогда как скорость фронта будет падать до сколь угодно малых величин – контринтуитивный эффект, кажущийся парадоксальным. Парадокс разъясняется просто. В

только что рассмотренной ситуации мощный диффузионный поток, текущий в узкой пространственной зоне, охваченной крутым концентрационным фронтом, играет роль лишь "конвейера", который переносит почти весь активатор, порождаемый зоной производства, в зону его (активатора) деструкции, где последний исчезает. А то ничтожное количество активатора, которое при этом ежесекундно отдаётся наполняющей реактор химически активной среде, способно поддерживать лишь очень медленную "сдвиговую" перестройку фронта автоволны – тем более медленную, чем выше сила J_D (следствие соотношения дуальности).

Отметим, наконец, случай, когда движение фронта автоволны целиком поддерживается одним лишь диффузионным потоком. Этот случай реализуется тогда, когда мощность одной из зон (например, зоны производства) много больше мощности другой зоны (например, зоны деструкции). В этом случае величина наполняющего потока близка средней величине генерируемого фронтом диффузионного потока: $\langle J_D \rangle_u \sim J_F$.

Действительно, при $P_+ \gg P_-$ диффузионный поток отдаёт зоне деструкции лишь ничтожные количества активатора, и поэтому практически весь активатор, порождаемый зоной производства и диффундирующий из этой зоны, участвует в формировании наполняющего потока. В этой ситуации движение фронта автоволны поддерживается одним лишь диффузионным потоком; соответствующую скорость фронта можно оценить, сделав в соотношении дуальности замену $\langle J_D \rangle_u \rightarrow J_F = Av$. Это приводит к формуле

$$v = \frac{1}{A} \sqrt{D \int_{u_-}^{u_+} f(u) du},$$

которая совпадает с классической формулой Зельдовича – Франк-Каменецкого, оценивающей в теории горения скорость движения фронта пламени.

Предпринятый анализ имеет прямой выход в электрофизиологию. Из него, например, следует, что перед фронтом бегущего нервного импульса всегда движется низкоамплитудная волна ионного тока, который выходит из нервного волокна через его мембрану, унося из аксона электрические заряды ("зона деструкции" внутриклеточных зарядов). И лишь позади этой волны мембранный ионный ток реверсирует своё направление на входящее, поставляя заряды внутрь аксона ("зона производства" внутриклеточных зарядов).

Последнее замечание, кстати, подсказывает, как следует откорректировать известную модель Маркина – Пастушенко – Чизмадзе, предложенную для качественного описания динамики ионных токов при движении нервного импульса, чтобы сделать её физически адекватной.

О востребованности математической модели для знакопеременных хиральных иерархий

Твердислов В.А.⁽¹⁾, Попцова М.С.⁽¹⁾

МГУ имени М.В.Ломоносова, физический факультет,
кафедра биофизики ⁽¹⁾
tverdislov@mail.ru

Сформулирована синергетическая закономерность, имеющая общий физико-химический и биологический характер: эволюционирующая система, обладающая запасом свободной энергии и элементами хиральной асимметрии, способна в процессе самоорганизации изменять тип симметрии внутри одного иерархического уровня, повышая свою «сложность», а при переходе на более высокий уровень – изменять масштаб и знак хиральности со сменой функциональной роли энантиоморфа [1-3]. Приведены примеры из области неживой природы, в биомакромолекулах отмечено чередование знака хиральных структур в последовательности от мономеров до суперспиралей: D-L-D-L для ДНК, L-D-L-D для белков. Приведенные иерархии отражают сквозные взаимодействия в макромолекулярных структурах клеток. Востребована математическая модель эволюции хиральных структур, формирующих знакопеременные иерархии. Известные фрактальные модели характеризуются сохранением

знака хиральности в разномасштабных инвариантах.

Список литературы

- [1] Твердислов В.А., Сидорова А.Э., Яковенко Л.В. От симметрий – к законам эволюции. I. Хиральность как инструмент стратификации активных сред/Биофизика, 2012, том 57, вып. 1, с. 146–154
- [2] Сидорова А.Э., Твердислов В.А. Самоорганизация в иерархии активных сред как движущая сила эволюции биосферы / ВМУ. Серия 3. Физика. Астрономия. 2012, №2, с. 65-69
- [3] С.В.Стовбун, А.М.Занин, А.А.Скоблин, А.И.Михайлов, А.А.Берлин. Феноменологическое описание спонтанного образования макроскопических струн в низкоконцентрированных хиральных растворах и формирования анизометрических гелей/Доклады Академии наук, 2012, том 442, № 5, с. 645-

Роль взаимодействия электрической и механической волн в функции миокарда (одномерная континуальная модель)

Курсанов А.Г.⁽¹⁾, Викулова Н.А.⁽¹⁾, Кацнельсон Л.Б.^(1, 2), Соловьева О.Э.^(1, 2), Мархасин В.С.^(1, 2)

ФГБУН Институт иммунологии и физиологии Уральского отделения Российской академии наук⁽¹⁾, Уральский федеральный университет⁽²⁾
alexander.kursanov@gmail.com

Contribution of the muscle fiber length to the electrical and mechanical activity of myocardial tissue has been analyzed within the framework of a 1D continuous mathematical model.

1 Введение

В настоящей работе (в рамках математической модели одномерного волокна) исследуется взаимодействие электрической волны возбуждения и механической волны сокращения-расслабления, возникающей в процессе последовательной активации волокна. Сложный характер локальных деформаций волокна вследствие последовательности его активации в силу механоэлектрических обратных связей существенно влияет на структуру электрической волны. В данной работе исследуется, как длина волокна (в данном исследовании - количество кардиомиоцитов) влияет на характеристики электрической и механической функции кардиомиоцитов и волокна в целом.

2 Методы и результаты

2.1 Математическая модель

Разработанная нами математическая модель волокна сердечной мышцы основана на континуальном описании миокардиальной ткани. С одной стороны, миокардиальная ткань рассматривается как сплошная среда, по которой распространяется электрическая волна возбуждения и механическая волна деформаций. С другой стороны, каждой точке этой среды ставится в соответствие кардиомиоцит, функция которого детально описывается математической моделью электромеханической активности одиночной клетки миокарда, разработанной нами ранее [2].

Схема модели представлена на рисунке 1, включая реологическую схему кардиомиоцита, как элементарного сократительного звена (точки) в модели ткани. В этой схеме представлены активный сократительный элемент (CE) модели кардиомиоцита – саркомер, а также пассивные упругие (PE, SE) и вязкий элементы (VS), определяющие не только пассивно-упругие свойства

клетки, но и влияющие на его активную механическую функцию [1].

Наряду с вариантом реологической схемы кардиомиоцита, где VS параллелен CE (рис. 1), мы исследовали 1D модель волокна с другой реологической схемой клетки, где VS был параллелен PE.

Особенностью нашей модели является подробное описание функции сократительных белков и их кальциевой активации. Эти механизмы обеспечивают воспроизведение в рамках модели широкого круга экспериментальных данных о влиянии механических условий сокращения миокарда (в том числе, начального растяжения саркомеров, нагрузки на миокард и др.) не только на собственно механическую активность сердечной мышцы, но и на кинетику внутриклеточного свободного кальция и генерацию потенциала действия (феномены механо-электрической обратной связи в миокарде).

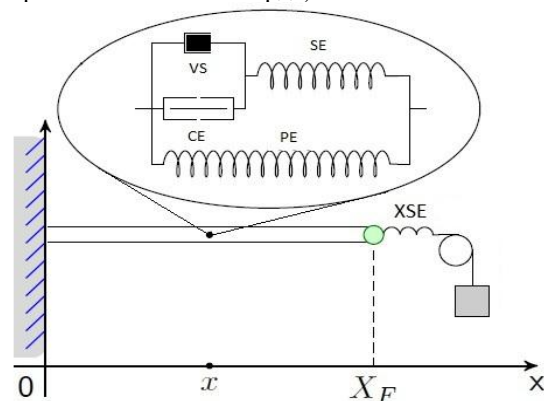


Рис. 2. Одномерный многоклеточный препарат миокардиальной ткани.

Электрическая волна возбуждения волокна описывается уравнением реакционно-диффузионного типа для распространяющегося в среде потенциала действия.

2.2 Анализ электромеханических свойств волокна

Мы исследовали изометрические сокращения (левый и правый концы волокна закреплены) однородного волокна (т.е. использовались модели

кардиомиоцитов с одинаковыми параметрами вдоль волокна) при различных длинах волокна в состоянии провиса. При этом начальное относительное растяжение волокон, и, следовательно, начальные длины саркомеров (здесь - 2.1 мкм) в кардиомиоцитах всех рассмотренных виртуальных препаратов были одинаковыми, так что волокна разной длины отличались только количеством последовательно соединенных клеток. Левый край волокна электрически стимулировался извне, а далее волна возбуждения автоматически распространялась вдоль волокна слева направо. Были рассмотрены 3 типа моделей, в разной степени учитывающих электрическое и механическое взаимодействия между клетками. Первый тип – электрическая модель (ЭМ), – учитывал только электротоническое взаимодействие. Второй тип – механическая модель (ММ), – учитывал только механическое взаимодействие, при этом последовательность возбуждения задавалась вручную. Третий тип – полная электро-механическая модель ЭММ, – включал и электротоническое, и механическое взаимодействия.

2.3 Результаты

Континуальная модель однородного мышечного волокна воспроизводит: распространение электрической волны возбуждения вдоль волокна с физиологически адекватной скоростью (0.46 м/с); процессы укорочения-растяжения и генерации силы в сегментах волокна по ходу распространения волны возбуждения; изменение концентраций основных внутриклеточных ионов, ответственных за изменение мембранного потенциала и сократительную функцию в клетках.

Последовательная активация клеток в ЭММ волокна при распространении волны возбуждения приводит к возникновению существенной неоднородности функциональных свойств кардиомиоцитов. В частности, на рисунке 2 иллюстрируются градиенты длительности потенциала действия (ДПД) на уровне 90% реполяризации в клетках ЭММ в волокнах с разной длиной провиса от 20 мм до 55 мм. ДПД в клетках на левом конце волокна возрастали по сравнению с ДПД в изолированной клетке (ДПД_{кл} = 236 мс), и оказались тем больше, чем больше была длина волокна (на 8, 13, 24 и 33 мс соответственно, см. рис. 2). Во всех рассмотренных моделях ДПД в клетках волокна уменьшаются с увеличением расстояния, на котором клетки расположены от левого края волокна (т.е. с ростом задержки времени прихода электрического стимула). При этом ДПД в клетках на правом краю волокон падают даже ниже ДПД_{кл} (до ~220 мс). Таким образом, дисперсия реполяризации в исходно однородной ЭММ увеличивается с увеличением протяженности волокна.

Наряду с градуальным изменением ДПД в

клетках волокна в ЭММ возникают градиенты пиков концентрации кальция и числа поперечных мостиков. Примечательным оказался тот факт, что несмотря на одинаковые начальные длины саркомеров в волокнах различной длины, пик силы уменьшался, а время его достижения увеличивалось при увеличении длины препарата.

В отличие от ЭММ в ЭМ, где механическое взаимодействие отсутствует, наблюдался лишь незначительный монотонный спад ДПД по сравнению с ДПД_{кл} вследствие электротонического взаимодействия кардиомиоцитов.

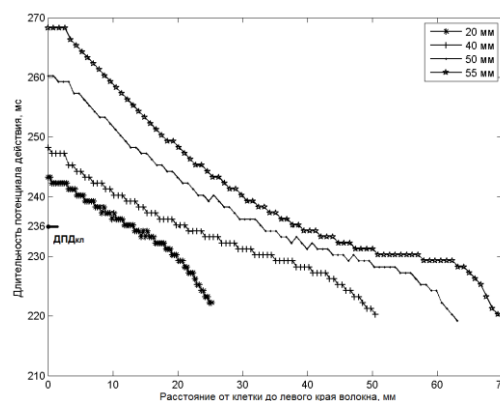


Рис. 3. ДПД в клетках при различных длинах волокна.

Результаты, полученные в численных экспериментах на 1D моделях с отличающимися реологическими схемами кардиомиоцитов, качественно не различались, при этом количественные эффекты были несколько более выраженными в ЭММ, представленной на рис. 1. Это сходство результатов позволяет предположить, что обнаруженные нами феномены являются существенным свойством активного миокарда и не зависят от реологической модели клетки.

3 Обсуждение

Анализ результатов, полученных в рамках ЭМ, ММ и ЭММ показал, что описанные выше феномены возникают в результате механического взаимодействия динамически деформирующихся элементов и, в частности, благодаря механо-электрическим обратным связям во внутриклеточных процессах и в межклеточном взаимодействии.

Таким образом, на разработанных нами 1D моделях волокна сердечной мышцы нами впервые показан существенный вклад механических факторов на макро и микро уровнях в формирование функциональной геометрии сердечной мышцы.

Работа поддержана грантами РФФИ и УрО РАН (12-М-14-2009, 12-П-4-1067, 12-П-4-1036, 10-04-96075, 11-4-НП-467).

Список литературы

- [1] L. B. Katsnelson, et al., *Journal of Theoretical Biology*, 2011, 272(1), 83-95.
- [2] T. B. Sulman, et al., *Bulletin of Mathematical Biology*, 2008, 70(3), 910-49.

Регулярная статистическая организация кодирования в последовательностях ДНК

Кутыркин В.А.⁽¹⁾, Чалей М.Б.⁽²⁾

МГТУ им. Н.Э. Баумана, Москва⁽¹⁾, ИМПБ РАН, Пущино⁽²⁾
vkutyркиn@yandex.ru

Original spectral-statistical methods were developed to recognize a new type of latent periodicity in DNA, called latent profile periodicity, or latent profility. Searching for latent profility allows the detection of different levels of information coding in genes and local DNA segments.

1 Введение

Понятие скрытой профильной периодичности (скрытой профильности) в последовательностях ДНК, было введено в [1]. Известные методы, применявшиеся для распознавания размытых тандемных повторов, не могут быть использованы для выявления этого нового типа скрытой периодичности.

Скрытая профильность, выявленная в кодирующих районах ДНК, транслируется в особенности структуры первичной последовательности белка. Непосредственное выявление таких особенностей является достаточно сложной задачей, поскольку цель поиска не известна априори. Этим обусловлена актуальность распознавания скрытой профильности в последовательностях ДНК.

В настоящей работе предлагаются методы распознавания скрытой профильности и регулярности в последовательностях ДНК, основанные на спектрально-статистическом подходе [2, 6].

2 Методы

В качестве модели скрытой профильности рассматривается мульти-полиномиальная схема испытаний. Эта схема индуцируется L последовательными полиномиальными схемами с четырьмя исходами (4 буквы алфавита ДНК). Следовательно, в каждом испытании реализуется случайная буква, которая описывается вероятностным распределением букв алфавита ДНК. В первом испытании этой мульти-полиномиальной схемы реализуется первая случайная буква (полиномиальная схема), во втором – вторая, ..., в L -том – L -ая случайная буква (полиномиальная схема). Затем, такая последовательность L испытаний повторяется. На заключительном этапе в

последовательности испытаний могут реализоваться менее L таких случайных букв (полиномиальных схем). Следовательно, моделью скрытой профильности является совершенный тандемный повтор, паттерн которого представлен случайной строкой из независимых случайных букв. Анализируемая последовательность ДНК рассматривается как реализация такого тандемного повтора. Поэтому скрытая профильная периодичность в текстовой последовательности может быть выявлена только с помощью статистических критериев. При анализе последовательности ДНК в этих критериях используются различные статистики [1, 2, 6], имеющие вид функциональных зависимостей (спектров) от тестируемых периодов последовательности. Тестируемый период (тест-период) – это натуральное число, не превышающее половины длины анализируемой последовательности ДНК.

3 Результаты и обсуждение

Спектр последовательности ДНК, названный характеристическим, позволяет получить оценку периода скрытой профильной периодичности в анализируемой последовательности ДНК. На примере последовательностей из базы данных TRDB (Tandem Repeat Database) [5] показано, что в размытом тандемном повторе, первый ярко выраженный пик характеристического спектра наблюдается на тест-периоде, соответствующем периоду этого повтора (см., например, рис. 1, а). Аналогичное свойство может проявляться и для характеристических спектров последовательностей ДНК, в которых выявляется скрытая профильная периодичность [2, 3, 6]. Следовательно, первый тест-период, на котором проявляется ярко выраженный максимум этого спектра, может служить оценкой периода скрытой профильной периодичности (см. рис. 1, б).

На основе полученной оценки периода исследуется спектр отклонения от скрытой профильной периодичности с таким периодом. Если

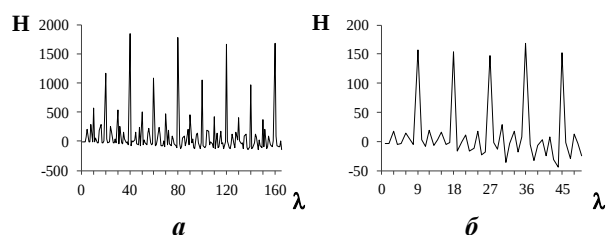


Рис. 1. Характеристические спектры последовательностей ДНК со скрытой периодичностью: **а** — размытый tandemный повтор на хромосоме I человека (TRDB, Indices 1943319 – 1945814, Venter 2007); **б** — кодирующий район гена фактора некроза опухоли (KEGG, hsa:338872, 1002 нукл.).

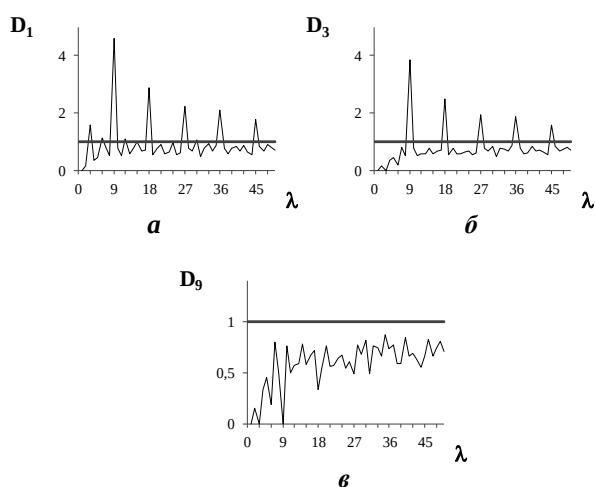


Рис. 2. Спектры D_L отклонения от L -профильности (**а** — $L=1$, **б** — $L=3$, **в** — $L=9$) для кодирующего района гена фактора некроза опухоли человека (KEGG, hsa:338872, 1002 нукл.).

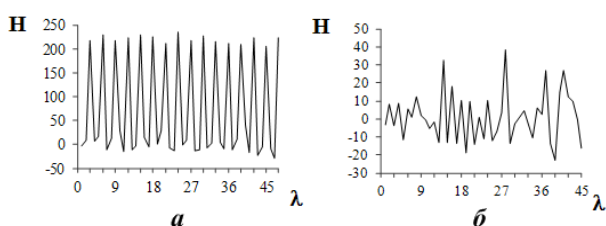


Рис. 3. Характеристические спектры кодирующих и некодирующих последовательностей ДНК: **а** — кодирующий район гена трансмембранного белка человека (KEGG, hsa:80757, 960 нукл.); **б** — интрон гена UCHL1 человека на хромосоме IV, 917 нукл.).

значения этого спектра не превышают единицы, то принимается гипотеза о том, что в рассматриваемой последовательности ДНК проявляется скрытая профильная периодичность [2, 3, 6]. Этот критерий

основан на стандартной статистике Пирсона. В частности, если в последовательности проявляется скрытая профильная периодичность с единичным периодом (тривиальная периодичность), то эта последовательность признаётся однородной, в противном случае — неоднородной. Поэтому спектр отклонения от 1-профильности называется спектром отклонения от однородности. Нетривиальная скрытая профильная периодичность ищется только для неоднородных последовательностей. В качестве примера на рис. 2 приведены спектры отклонения от 1-, 3- и 9- профильности для последовательности из базы данных KEGG (Kyoto Encyclopedia of Genes and Genomes) [4] для кодирующего района ДНК, характеристический спектр которого на рис. 1, б даёт оценку периода скрытой профильной периодичности в 9 нуклеотидов (нукл.).

В характеристических спектрах кодирующих районов ДНК, как правило, наблюдается регулярное чередование пиков через два нуклеотида (см. рис. 1, б и рис. 3, а). Такое явление было названо 3-регулярностью характеристического спектра [2, 3, 6]. В характеристических спектрах интронов 3-регулярность, как правило, не наблюдается (см., например, рис. 3, б).

В характеристическом спектре кодирующего района может наблюдаться следующая картина. На фоне 3-регулярности проявляется ярко выраженный пик (или периодическое повторение выраженных пиков), указывающий на наличие скрытой профильности в этом районе (см. рис. 1, б). Следовательно, характеристический спектр может проявлять двух-уровневую организацию кодирования, когда период скрытой профильности превышает 3 нукл. Первый уровень обусловлен триплетным кодированием аминокислот, второй — скрытой профильностью кодирующего района ДНК.

Список литературы

- [1] M. Chaley, V. Kutyrrkin *Math. Biosci.*, 2008, **211**, 186-204.
- [2] M. Chaley, V. Kutyrrkin *DNA Res*, 2011, **18**, 353-362.
- [3] M. Chaley, V. Kutyrrkin *Moscow Univ. Biol. Sci. Bull.*, 2010, **65**(4), 133-135.
- [4] Kyoto Encyclopedia of Genes and Genomes Web site, 2012. URL: <http://www.genome.jp/kegg/>
- [5] Tandem Repeats Database Web site, 2012. URL: <http://tandem.bu.edu/cgi-bin/trdb/trdb.exe>
- [6] В.А. Кутыркин, М.Б. Чалей *Вестник МГТУ им. Н.Э.Баумана, Сер. «Естественные науки»*, 2011, Спец. выпуск «Математическое моделирование», 200-215.

Профильная периодичность генов и структурно-функциональные свойства кодируемых белков

Чалей М.Б.⁽¹⁾, Кутыркин В.А.⁽²⁾

ИМПБ РАН, Пущино⁽¹⁾, МГТУ им. Н.Э. Баумана, Москва⁽²⁾

maramaria@yandex.ru

New type of latent periodicity called latent profile periodicity has been revealed in the genes of apolipoprotein, zinc finger and fibronectin protein families. Correlation of the latent periodic gene structures with domain organization of the proteins is discussed.

1 Введение

В настоящей работе методы спектрально-статистического подхода [1, 5] к распознаванию нового типа скрытой периодичности (скрытой профильной периодичности) в последовательностях ДНК применяются для выявления регулярной структуры в кодирующих районах геномов различных организмов. Эти методы позволяют выделить два уровня в организации кодирующих районов ДНК [1, 2, 5]. Первый уровень обусловлен триплетным законом кодирования аминокислот, второй – скрытой профильной периодичностью. В работах [1, 2, 5] было показано, что профильная периодичность в кодирующих районах коррелирует со структурно-функциональными свойствами транслируемых белков. Как правило, методы Фурье-анализа оказываются нечувствительными к выявлению существующей в последовательностях ДНК скрытой профильной периодичности. Выявление второго уровня скрытой структурной организации в кодирующих районах может быть полезным для анализа и классификации генов из различных организмов с целью предсказания особенностей структуры и функциональных свойств их белковых продуктов. Взаимосвязь скрытой профильной периодичности в кодирующих районах ДНК и структурно-функциональных особенностей транслируемых белков продемонстрирована для семейств белков аполипопротеинов, цинковых пальцев и фибронектинов.

2 Методы

При применении спектрально-статистического подхода [1, 5] к исследованию скрытой структуры последовательности ДНК используются функциональные зависимости (спектры) её статистических свойств от тестируемых периодов. Количественный анализ этих спектров позволяет распознавать существование в анализируемой последовательности скрытой регулярной структуры.

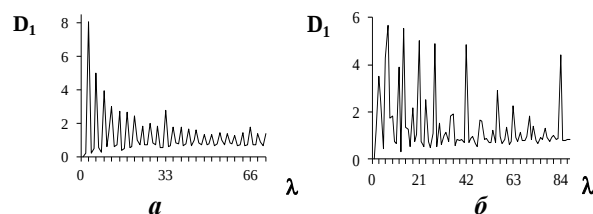


Рис. 1. Спектры отклонения от однородности кодирующих районов генов для белков из семейств **а** — аполипопротеинов (A-I) рыбы дорады (GenBank, AF013120, 783 нукл.) и **б** — цинковых пальцев человека (KEGG, has:26974, 1173 нукл.)

Нетривиальная структура скрытой периодичности может наблюдаться только в неоднородных последовательностях ДНК. Спектр D_1 отклонения от однородности, введённый в спектрально-статистическом подходе [1, 5], служит для выявления таких последовательностей на выбранном уровне доверия 95% (уровне значимости 5%). Если на всех тестируемых периодах значения спектра отклонения от однородности не превышают единичного значения, то последовательность признаётся однородной, в противном случае – неоднородной. На рис. 1 приведены спектры отклонения от однородности для кодирующих районов отдельных генов белков семейств аполипопротеинов из базы данных Pfam [4] и цинковых пальцев из базы данных KEGG [3].

Для оценки периода скрытой профильной периодичности в неоднородных последовательностях ДНК используется спектр, ранее названный характеристическим [1, 2, 5]. Первый тестируемый период, на котором проявляется ярко выраженный максимум этого спектра может служить оценкой периода скрытой профильной периодичности. На рис. 2 приведены характеристические спектры последовательностей ДНК, для двух из которых (рис. 2, а, б) спектры отклонения от однородности приведены на рис. 1. Согласно рис. 2, а, б, в, оценками периодов скрытой профильной периодичности являются периоды $L=33$, $L=84$ и $L=291$, соответственно.

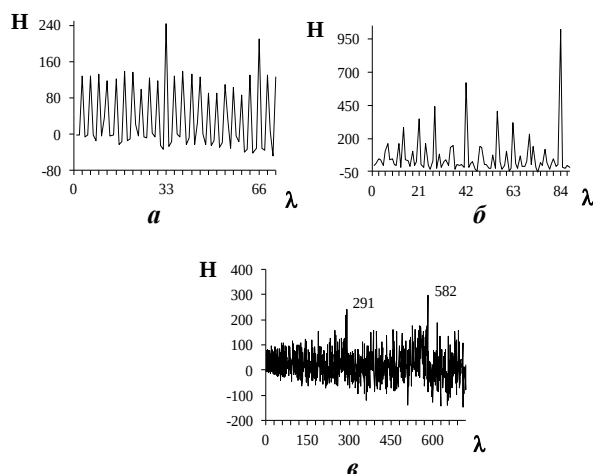


Рис. 2. Характеристические спектры кодирующих районов генов из семейств: **а** — аполипопротеинов (А-І) рыбы дорады; **б** — цинковых пальцев человека; **в** — фибронектинов III мыши (KEGG, mму:319448, 3597 нукл.).

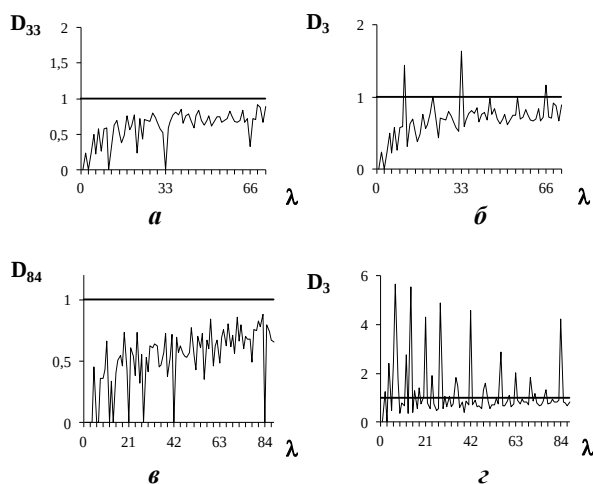


Рис. 3. Спектры D_L отклонения от L -профильности для кодирующих районов генов: аполипопротеина (А-І) рыбы дорады (**а** — $L=33$, **б** — $L=3$) и белка из семейства цинковых пальцев человека (**в** — $L=84$, **г** — $L=3$).

Пусть L — оценка периода скрытой профильной периодичности. Тогда введенный в спектрально-статистическом подходе спектр D_L отклонения от L -профильности позволяет подтвердить или отвергнуть эту оценку. Если на всех тестируемых периодах значения спектра D_L не превышают единичного значения, то признаётся, что в последовательности проявляется скрытая L -профильная периодичность, в противном случае — наличие скрытой L -профильной периодичности отрицается. На рис. 3 приведены спектры D_L для отдельных генов белков из семейств

аполипопротеинов и цинковых пальцев. Согласно этому рисунку, в последовательностях генов выявляется скрытая L -профильность ($L=33$ для аполипопротеинов и $L=84$ для цинковых пальцев). Для достоверного подтверждения L -профильности в спектрально статистическом подходе [1, 5] предложены и другие методы.

3 Результаты и обсуждение

Спектрально-статистический подход позволил выявить наличие скрытой профильности в 33 нукл. (рис. 3, а) в кодирующих районах генов семейства аполипопротеинов PF01442 из базы Pfam [4]. Основные функции аполипопротеинов — транспорт липидов. Семейство содержит более 800 последовательностей белков из 100 различных видов.

Наличие скрытой профильной периодичности в 84 нукл. (см. рис. 3, б) в кодирующей последовательности ДНК соответствует в белке повторяющемуся домену «цинкового пальца» (как правило из ~20 аминокислот), содержащему альфа-спираль и две антипараллельные бета-структуры. Основной группой белков с цинковыми пальцами являются ДНК-связывающие факторы транскрипции.

Характерный размер паттерна скрытой профильной периодичности ~290 нукл (рис. 2, в). наблюдается в характеристических спектрах генов семейства фибронектинов, содержащих домены III-типа. Фибронектин — один из белков межклеточного матрикса, гликопротеин, построенный из двух идентичных полипептидных цепей. Каждая цепь состоит из ~10 доменов. Предполагают, что пространственная структура доменов содержит семь антипараллельных β -листов. Размер паттерна в 291 нукл. коррелирует с размером (~90-100 аминокислот) повторяющихся характерных доменов в белках этого семейства.

Список литературы

- [1] M. Chaley, V. Kutyrkin *DNA Res*, 2011, **18**, 353-362.
- [2] M. Chaley, V. Kutyrkin *Moscow Univ. Biol. Sci. Bull.*, 2010, **65**(4), 133-135.
- [3] Kyoto Encyclopedia of Genes and Genomes Web site, 2012. URL: <http://www.genome.jp/kegg/>.
- [4] The Pfam protein families database Web site, 2012. URL: <http://pfam.sanger.ac.uk/>.
- [5] В.А. Кутыркин, М.Б. Чалей *Вестник МГТУ им. Н.Э.Баумана, Сер. «Естественные науки»*, 2011, Спец. выпуск «Математическое моделирование», 200-215.

Сравнительный анализ геномов вируса клещевого энцефалита: дифференциация по степени вирулентности

Беликов С.И.⁽¹⁾, Гусев В.Д.⁽²⁾, Мирошниченко Л.А.⁽²⁾, Титкова Т.Н.⁽²⁾

г. Иркутск, Лимнологический институт СО РАН⁽¹⁾,
г. Новосибирск, Институт математики им. С.Л. Соболева СО РАН⁽²⁾
gusev@math.nsc.ru

The system of “fine” differences revelation between close (allied) genomes on the nucleotide level was developed and realized. It was approved in the problem of tick-borne encephalitis virus genomes differentiation by degree of virulence.

1 Введение

Вирус клещевого энцефалита (ВКЭ) активно изучается как возбудитель опасного инфекционного заболевания. Различают три субтипа ВКЭ: дальневосточный, европейский и сибирский. Геном ВКЭ представлен молекулой однонитчатой РНК длиной около 11000 оснований, кодирующей полипептидную цепь, содержащую около 3400 а.о. Нетранслируемые 5'- и 3'-концы составляют незначительную часть генома. Полипептидная цепь кодирует (без перекрытия) 3 структурных (С, PrE_M, E) и 7 неструктурных (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) белков. Они образуются в результате расщепления полипротеина протеазами вирусного или клеточного происхождения.

В пределах отдельного генотипа разные штаммы ВКЭ характеризуются различной активностью. Известно, что на степень вирулентности влияют точечные замены аминокислот в отдельных генах (E, NS1 и др.) и, возможно, мутации в 5'- и 3'-некодирующих областях. В связи с этим представляет интерес сравнительный анализ полногеномных последовательностей ВКЭ, значимо отличающихся по степени вирулентности. Рассматриваются два класса последовательностей: низковирулентные (НВ), выделенные от людей с инаппарантной (без клинических проявлений инфекции) формой заболевания и высоковирулентные (ВВ), изолированные из крови или тканей умерших от КЭ людей. В [1] проведен анализ НВ- и ВВ-штаммов дальневосточного субтипа на аминокислотном уровне. По его итогам выделен ряд маркерных позиций, различающих по составу аминокислот рассматриваемые классы последовательностей.

Целью данной работы является исследование возможностей дифференциации НВ- и ВВ-геномов

ВКЭ на нуклеотидном уровне. Это позволяет обнаруживать новые (не проявляющие себя на аминокислотном уровне) маркеры для различения НВ- и ВВ-штаммов. Кроме того, информация о наличии структурных особенностей на нуклеотидном уровне в местах расположения аминокислотных маркеров может прояснить механизмы влияния отдельных аминокислотных замен на степень вирулентности.

2 Исходные данные и методы анализа

Анализировались полные геномы 38 штаммов ВКЭ дальневосточного субтипа с зафиксированной формой протекания заболевания (НВ или ВВ). Значительная часть их депонирована в GenBank'e, остальные секвенированы (относительно недавно) сотрудниками Лимнологического института СО РАН (г. Иркутск). Кодирующие части всех геномов имеют практически одинаковую длину и высокий уровень попарного сходства (порядка 95% и выше). По степени близости кодирующих частей подборку можно разделить на две группы из 22-х и 16 геномов соответственно. Это разбиение в значительной мере коррелирует с априорными оценками степени вирулентности штаммов из подборки, однако имеются исключения. В частности, штаммы Primorye-90 (Pr-90) и Primorye-86 (Pr-86) выделены от умерших людей, но по нуклеотидной последовательности они ближе к инаппарантным штаммам, в большинстве своем попавшим в первую группу (22 штамма). В то же время штаммы Pr-739, Pr-52 и Pr-196 выделены от людей с инаппарантной формой протекания заболевания, но на нуклеотидном уровне они ближе к ВВ-штаммам. Такого рода исключения могут быть обусловлены особенностями иммунной системы инфицированных людей, но подобная информация отсутствует.

Для выявления «тонких» различий между группами НВ- и ВВ-штаммов в целом, а также между штаммами-исключениями в каждой группе и остальными ее представителями, мы используем два подхода: L-граммный анализ [2] и сложностной [3]. Первый дает представление о том, из каких цепочек символов длины L (L-грамм) составлен текст. Фиксируется многообразие и частоты цепочек при

Сложностной анализ осуществляет представление текстов в терминах повторов (прямых, симметричных, комплементарных симметричных). Чем больше длины повторов и их количество, тем меньше сложность текста [3]. Фрагменты текста с аномально низкой сложностью отличаются высокой структурированностью и часто встречаются в регуляторных областях геномов. Одиночные мутации приводят к усилению или ослаблению регуляторной структуры, что в нашем случае может повлиять на степень вирулентности.

Таксономия живых организмов по характеристикам строя их ДНК

Поздниченко Н.Н., Гуменюк А.С., Шпынов С.Н.

ОмГТУ
nick670@yandex.ru

В работе проведены таксономии 29 живых организмов по характеристикам их ДНК, взятых в разных представлениях.

Введение

Использование при таксономии непосредственно символьных последовательностей, отображающих ДНК, затруднено неопределённостью построения соответствующего им признакового пространства. В связи с этим для классификации живых организмов используются разного рода характеристики, вычисляемые на основе знакового представления нуклеотидных цепей. Совокупность этих характеристик составляет пространство информативных признаков, на котором строится классификация объектов.

В настоящее время практически отсутствуют адекватные средства формального анализа структуры нуклеотидных последовательностей, которая, по нашему мнению, наиболее естественно отображается оригинальным порядком следования элементов, представляющим организованное целое.

В дальнейшем будем называть **строём цепи** событий (сообщений, знаков и т.д.) особого рода кортеж (упорядоченное множество), в котором каждому компоненту данной цепи поставлено в соответствие натуральное число, причем идентичные по выбранному признаку компоненты отображены одним и тем же числом. Самый первый компонент такого кортежа – единица, а все остальные первые встречные разные натуральные числа (представляющие вместе с единицей **алфавит строя**) возрастают на единицу. В соответствии с теоретико-множественным определением вектора назовем такой специфически организованный кортеж **«вектор строя»** [1].

Ниже, в первой строке таблицы 1, приведена произвольная знаковая последовательность нуклеотидов; во второй строке – полный строй этой цепи; в третьей строке выделена однородная цепь цитозина; в четвертой строке представлена соответствующая однородная цепь строя.

Определим значения «интервалов» Δ_{ji} между соседними знаками в выделенной однородной цепи ($\Delta_{21}=5$; $\Delta_{22}=1$; $\Delta_{23}=4$; $\Delta_{24}=1$; $\Delta_{25}=1$).

Таблица 1

G	C	T	A	A	A	C	C	T	A	G	C	C	C
1	2	3	4	4	4	2	2	3	4	1	2	2	2
-	C	-	-	-	-	C	C	-	-	-	C	C	C
-	2	-	-	-	-	2	2	-	-	-	2	2	2

Для описания строя полной неоднородной знаковой цепи интервалы выделяются для всех j -ых однородных цепей (где $j = 1, 2, \dots, m$, где m – мощность алфавита знаковой цепи; в случае нуклеотидных цепей $m = 4$). Значения этих интервалов используются для вычисления разнообразных характеристик строя, представленных в [1]. В данной работе интервалы используются для вычисления **среднего геометрического интервала, средней удалённости и регулярности** одинаковых компонентов в знаковой цепи, которые определены соответственно в виде

$$\Delta_g = \sqrt[n]{V}, \quad g = \log \Delta_g, \quad r = \Delta_g / D,$$

где V – объем цепи, D – число описательных информаций (по Мазуру [2]), определяемые в виде

$$V = \prod_{j=1}^m \prod_{i=1}^{n_j} \Delta_{ji}, \quad D = \prod_{j=1}^m \frac{n_j^n}{n_j},$$

где n_j – число вхождений j -го компонента. Кроме того, используется характеристика глубины цепи $G = \log V$.

2 Таксономия

2.1 Филогенетические деревья

В настоящее время при построении классификаций живых организмов всё чаще используются филогенетические деревья, основанные на сравнении нуклеотидных последовательностей и отражающие родственные связи организмов, вместо фенотипических классификаций, основанных на внешних признаках организмов. При этом для построения автоматических классификаций требуются средства

для сравнения генетических последовательностей и нахождения «расстояния» между ними.

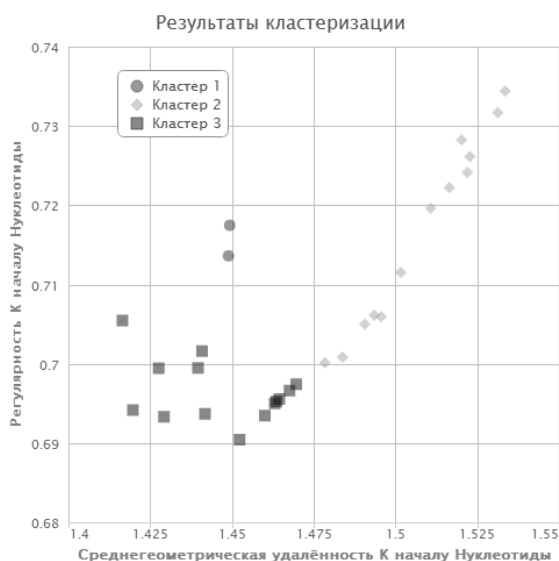
В настоящее время широко распространены три способа сравнения и соответствующие им меры сходства: статистические методы, информационно-энтропийные и редакционного расстояния. Как отмечено выше в нашей работе используются числовые характеристики взаимного расположения элементов в цепи и сформулированные на их основе меры сходства этих цепей, которые учитывают удалённость, регулярность и глубину.

2.2 Алгоритм таксономии

Для таксономии организмов по их ДНК использовался алгоритм кластеризации λ -KRAB, описанный в [3], отличительной особенностью которого является использование нелинейного нормированного λ -пространства, учитывающего не только расстояние между объектами, но и характеристику локальной плотности объектов.

3 Исследования

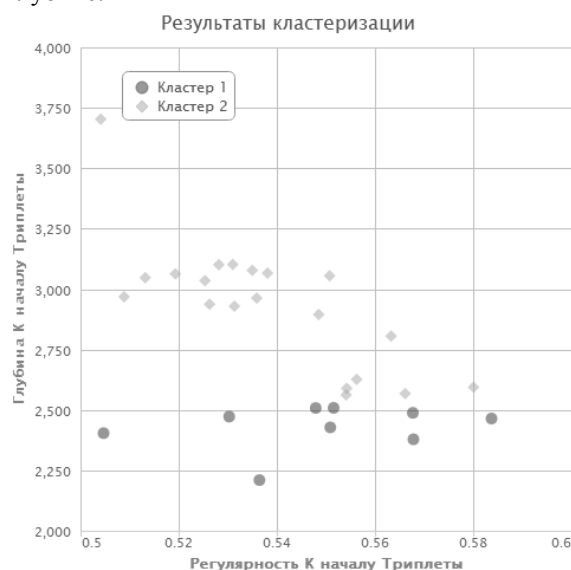
Были выполнены разбиения на таксоны с опорой на нуклеотиды и с опорой на триплеты выборки из 29 организмов.



1 опыт	Название объекта	2 опыт
2	Candidatus N.m – бактерия	1
2	B.anthraxis - Сибирская язва	1
1	Th.thermophilus – микроорганизм	1
1	Th.thermarum – микроорганизм	1
2	S.pyogenes – Стрептококк	1
2	P.humanus cap – блоха	1
2	N.g гонорея	1
2	M.pneumoniae - атипичных пневмоний	1
3	M.musculus – мышь	1
2	I.persulcatus - Искодовые клещи	2
3	H.s – человек	2
3	G.gallus – курица	2
3	C.crocodylus – Крокодил	2

3	C.familiaris – Собака	2
2	B.burgdorferi – боррелиоз	2
3	A.calva – рыба	2
2	Zebrias zebra – Рыба	2
3	Sus scrofa – Кабан	2
3	Rattus norvegicus - Серая крыса	2
3	Mus musculus - Домовая мышь	2
2	Kareius bicoloratus - двухцветная камбала	2
3	Homo sapiens - Человек разумный	2
3	Gallus gallus - Банкивская джунглевая курица	2
3	Erinaceus europaeus - Обыкновенный ёж	2
2	Crocodylus niloticus - Нильский крокодил	2
3	Cricetulus griseus - серый хомячок	2
3	Bos taurus - Дикий бык	2
2	O.moubata – клещи	2
2	M.domestica – муха	2

Ниже приведена проекция таксономии по трём характеристикам: удалённости, регулярности и глубине.



Характеристики триплетного представления хотя и дают другую картину разбиения, но при этом представляют дополнительную информацию о свойствах организмов и взаимосвязях между ними.

Данные предварительные исследования позволяют сделать вывод о том, что классификация с опорой на разные структурные единицы даёт разные представления таксономии. По нашему мнению, для получения правильной классификации, необходимо выбрать адекватные структурные единицы естественных цепей. Открытым остаётся вопрос поиска таких структурных единиц (естественных слов) для генетических текстов.

Список литературы

- [1] Гуменюк А.С. Алгоритмы анализа структуры сигналов и данных: монография / А.С. Гуменюк, Ю.Н. Кликушин, В.Ю. Кобенко, В.Н.Цыганенко; под науч. ред. д-ра техн. наук Ю.Н. Кликушина. – Омск: Изд-во ОмГТУ, 2010. – 272 с.
- [2] Мазур М. Качественная теория информации, Москва: Мир, 1974.
- [3] Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: Изд-во ин-та математики, 1999. – 270 с.

Определение характеристик пространственной зависимости компонентов нуклеотидных цепей

Поздниченко Н.Н., Гуменюк А.С.

ОмГТУ
nick670@yandex.ru

В данной работе определены пространственные зависимости двух уровней: между элементарными компонентами ДНК – нуклеотидами, и между группировками нуклеотидов, которые представляют собой «естественные слова» генетического текста.

1 Алгоритм выделения слов в генетических текстах

Алгоритм сегментации, называемый далее «базовым», считает выбранное испытуемое псевдослово словом такое, которое имеет неожиданную частоту появления в тексте, в сравнении с расчетной, определенной на основе марковской модели текста. Степень различия фактической и расчетной частот задается некоторым пороговым значением.

Модификация предполагает неоднократное применение базового алгоритма (с направленным изменением порога) для выбора такого разбиения на слова данной цепи, которое наилучшим образом отвечает критерию целостно-завершённого текста Ю. Орлова [1].

2 Вычисление зависимостей между компонентами знаковых цепей

В работе [2] на основе наглядных графических образов (диаграмм) представлены характеристики зависимости пары однородных цепей (j -ой и l -ой), взятых из состава данной неоднородной цепи. Это позволило соотнести субъективное суждение о степени «пространственной зависимости» элементов сопоставляемых цепей с формальной числовой характеристикой, представляющей такую зависимость.

Для наглядности в первой строке таблицы 1 представлен строй полной неоднородной цепи, в составе которого выделены две однородные j -ая и l -ая (в строках 2 и 3). На основе этих цепей составлена бинарно-однородная (далее бинарная) цепь (строка 4), в состав которой включены только те пары разных знаков j и l сравниваемых однородных цепей, которые определяются знаками l , «ближайшими справа» относительно знаков j . Предполагается, что наличие бинарной цепи свидетельствует о такой «причинно-следственной» связи цепей, при которой хотя бы некоторые компоненты (следствия) l -ой цепи расположены

вслед за компонентами (причинами) j -ой цепи на определенных интервалах.

Таблица 1

1	2	3	2	3	3	3	2	1	3	2	1	2	1	3	2
1								1			1		1		
	2		2				2			2		2			2
1	2							1		2	1	2	1		2

Связь такого типа названа «правосторонней пространственной зависимостью» событий одной l -ой цепи от другой j -ой цепи и обозначаемой (l/j) .

Обозначим: $\Delta(l/j)_i$ – интервал между i -ми ближайшими справа знаками l по отношению к знакам j ; $n(l/j)$ – число пар знаков j и l , связанных интервалами $\Delta(l/j)_i$. В таблице: $\Delta(l/j)_1 = \Delta(l/j)_3 = 1$; $\Delta(l/j)_2 = \Delta(l/j)_4 = 2$; Среднее геометрическое значение всех $n(l/j)$ установленных интервалов (бинарной цепи) $\Delta(l/j)_i$ между «смежными» элементами двух однородных цепей определено в виде

$$\Delta(l/j)_{cp} = n(l/j) \sqrt[n(l/j)]{\prod_{i=1}^{n(l/j)} \Delta(l/j)_i}.$$

Интервалы только между теми знаками выделенной однородной цепи l , которые являются «ближайшими справа» относительно знаков цепи j , обозначены $\Delta(l)_i$. В примере: $\Delta(l)_1 = 9$; $\Delta(l)_2 = 2$; $\Delta(l)_3 = 3$; $\Delta(l)_4 = 1$; Среднее геометрическое значение интервалов $\Delta(l)_i$ между выделенными элементами l_j в данной однородной цепи определено в виде

$$\Delta(l_j)_{cp} = n(l/j) \sqrt[n(l/j)]{\prod_{i=1}^{n(l/j)} \Delta(l)_i}.$$

При условии $\Delta(l_j)_{cp} > \Delta(l/j)_{cp}$, разность вида

$$\nu(l_j) = \left(1 - \frac{\Delta(l/j)_{cp}}{\Delta(l_j)_{cp}} \right)$$

представляет «избыточность» l -ой цепи, зависимой от j -ой. Некоторые элементы такой цепи «связаны справа» с элементами j -ой цепи. В противном случае, при $\Delta(l_j)_{cp} \leq \Delta(l/j)_{cp}$, данная разность свидетельствует об отсутствии избыточности. На основе отмеченного в [4] введен коэффициент частичной зависимости l -ой однородной цепи (от j -ой цепи) в виде

$$K_1(l/j) = (n(l/j)/n_l) \cdot \nu(l_j).$$

Отметим, что соотношение $n(l/j)/n_l$ представляет собой условную вероятность события, состоящего в появления пары знаков l и j , связанных

пространственной зависимостью, от появления знаков цепи l . В частном случае, когда $n(l/j) = n_l$ имеем *полную зависимость* этих цепей.

Кроме того, удобно использовать нормированный коэффициент частичной зависимости, учитывающий частоту встреч тех или иных пар – $n(l/j)/(n/2)$, определяемый в виде

$$K_{ln}(l/j) = (2n^2(l/j)/n_l \cdot n) \cdot v(l_j)$$

Нормированные числовые характеристики зависимостей позволяют эффективнее отделять бинарные цепи, содержащие большее количество пар от цепей, содержащих меньшее количество пар.

3 Исследования

Просмотр двух матриц зависимостей однородных цепей (нуклеотидов и слов) позволил выделить наиболее зависимые цепи для каждого из четырёх организмов.

3.1. *Th.thermarum* (микроорганизм-термофил)

Из матрицы первого типа видно, что однородные нуклеотидные цепи мало различаются по зависимости, при этом наибольшие зависимости наблюдаются от аденина.

Первый элемент \ Второй элемент	A	G	C	T
A	0	0,1294	0,1096	0,1020
G	0,0553	0	0,0642	0,0483
C	0,0646	0,0820	0	0,0750
T	0,0797	0,0895	0,0885	0

Представленный ниже фрагмент матрицы второго типа показывает наиболее зависимые цепи. При этом в целом зависимость между словами слабее чем между нуклеотидными.

	GATA	ATACA	TAA	CCGG	AAAGC
AAG	0,0355	0,0156	0,0203	0,0272	0,0220

Ниже показано расположение зависимых слов GATA от AAG в данной цепи:

...**AAG**AGCCCTTTGAG**GATA**...**AAG**AGCCCTTTGAGGATA
 CAAACTACCGCCAA**GATA**...**AAG**CCGCTAGCAAAAATACAT
 AGGCTTTAACGGGGCTGAAATAGCCCAAA**GATA**...**AAG**AGAC
 CCGTTTCTAATGTCCTTAGAACCG**GATA**...**AAG**GGCCGGAAC
 GATGACATTCCACCTTGGCTTCAAATGGACTC**GATA**

3.2. *Candidatus N.m.* (бактерия)

	A	G	C
T	0,1132	0,1051	0,1139

	TAGTT	CCCG	GTTTAA
A	0,0233	0,0159	0,0162
TTC	0,0141	0,0208	0,0151
GGTA	0,0056	0,0037	0,0161

3.3. *Rattus norvegicus* (Серая крыса)

	A	G	C
T	0,0950	0,0971	0,1016

	GAA	GCGTTA	GGGCATC
T	0,0249	0,0209	0,0095
GCGTTA	0,0138	0,0000	0,0162

3.4. *Homo sapiens* (Человек разумный)

	A	G	C
T	0,1154	0,1074	0,1133

	AAAATTTAACA	TC
ACC	0,0210	0,0126
GAAC	0,0144	0,0204
AT	0,0193	0,0148

4 Заключение

По результатам исследований малой выборки можно заметить, что у всех организмов (кроме бактерии-термофила) при понуклеотидном рассмотрении ДНК наибольшая зависимость элементов от тимина. Лишь немногие однородные цепи, представленные словами, зависимы друг от друга, для остальных же зависимости отсутствуют.

Список литературы

- [1] Гуменюк А.С. О выделении структурных единиц в белковых последовательностях на основе критерия целостно-завершенного текста. // Математическая биология и биоинформатика: III международная конф. – Пушино: изд-во ООО «МАКС Пресс», 2010
- [2] Гуменюк А.С. Алгоритмы анализа структуры сигналов и данных: монография / А.С. Гуменюк, Ю.Н. Кликушин, В.Ю. Кобенко, В.Н.Цыганенко; под науч. ред. д-ра техн. наук Ю.Н. Кликушина. – Омск: Изд-во ОмГТУ, 2010. – 272 с.

Аннотирование генов методом создания функциональных групп

Голышев М.А.⁽¹⁾, Коротков Е.В..⁽¹⁾

Центр «Биоинженерия» РАН, 117312, Москва, проспект 60-летия Октября, 7/1 ⁽¹⁾
mixagol@gmail.com, genekorotkov@gmail.com

We developed an information system for gene annotation that uses an approach from comparative genomics. This method is based on creation of phylogenetic profiles for new genes and genes with known functions. The result of prediction is a sorted list of possible functions. Our realization exactly predicts 80% of functions of annotated genes of the *Salmonella enterica* with the first five possible functions and 73% of functions of non-annotated genes.

1 Постановка задачи

Успехи секвенирования геномов предоставляют для анализа огромное количество информации. К настоящему времени доступно несколько тысяч геномов бактерий. Одной из важнейших задач биоинформатики на сегодняшний день является определение биологических функций нуклеотидных и аминокислотных последовательностей (аннотирование генов). Наибольшее распространение получили методы, основанные на подобию и НММ, однако они предсказывают лишь 60% всех секвенированных генов, в то время как функции оставшихся 40% остаются неизвестными. Предлагаемый подход не использует сравнение генов между собой, а учитывает взаимную встречаемость групп генов в определённых геномах.

Предполагается, что гены, имеющие одинаковые функции, будут иметь сходные филогенетические профили, построенные по множеству референтных геномов. Это следует из того, что ген, как правило, выполняет свою функцию не в одиночку, а совместно с другими генами в составе метаболического пути. Такая функциональная группа в процессе эволюции наследуется различными организмами, что в итоге приводит к формированию сходных профилей генов из одной группы.

2 Описание метода

Для построения филогенетических профилей генов с известной функцией из банка данных GenBank были выбраны 1200 референтных геномов и 3.7 млн. генов. Процесс создания базы данных

профилей состоял из нескольких этапов. На первом этапе происходил поиск значимых локальных выравниваний всех генов во всех референтных геномах с помощью программ для быстрого поиска локальных выравниваний. После чего, найденные локальные выравнивания расширялись до глобальных выравниваний, и для них методами Нидлмана-Вунша и Монте-Карло рассчитывалась статистическая значимость. Далее, для каждого гена формировался бинарный вектор: “1”, если статистическая значимость глобального выравнивания с соответствующим референтным геномом превышала заданный уровень, и “0”, если подобия нет, или оно не превышало заданный уровень. Таким образом, для каждого гена был рассчитан вектор бинарных значений размерности N, где N - число референтных геномов. Процесс аннотации нового гена выглядит следующим образом: для гена так же формируется бинарный вектор, после чего происходит поиск похожих векторов среди аннотированных генов, используя вероятностный подход. Функции тех генов, вероятности которых превысили заданный минимальный порог, участвуют в определении потенциальной функции аннотируемого гена. Результатом предсказания является отсортированный по вероятности список возможных функций аннотируемого гена.

3 Полученные результаты

Для проверки качества работы разработанной информационной системы было проведено аннотирование всех генов организма *Salmonella enterica* с учётом того, что гены этого организма не были использованы для создания базы данных профилей аннотированных генов. Для 3684 аннотированных генов (83% всех генов) были предсказаны функции для 3349 из них (91%), для остальных 9% возможных функций не обнаружено. Из 3349 генов была точно предсказана функция для 2682 (80%), если использовать список только из 5 наиболее вероятных функций. Из 735 не аннотированных генов были предсказаны возможные функции для 525 (71%) из них.

Поиск паттернов в последовательностях на основе эффекта информационного структурного резонанса

Дюк В.А.

Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук
v_duke@mail.ru

Examples of search of patterns with jokers in DNA sequences without procedure of their alignment on the basis of an information structural resonance effect are resulted.

1 Введение

Одна из стандартных задач компьютерного анализа молекулярно-генетических последовательностей состоит в том, чтобы выделить воспроизводимые структурные особенности и связать их с соответствующими функциями. Отбор и фиксация таких связей в ходе молекулярной эволюции позволяют использовать подход, основанный на сравнительном анализе последовательностей в базах данных.

2 Эмпирико-статистические методы исследования последовательностей

Наиболее обширная группа методов основана на препроцессинге последовательностей молекулярно-биологических данных с применением так называемого множественного выравнивания. После применения процедуры множественного выравнивания для анализа полученной матрицы «объект-признак» может быть применен весь арсенал методов прикладной статистики, в том числе большой набор инструментов технологии Data Mining. Некоторые инструменты требуют перевода символьных значений в числовую форму (перекодировки исходных данных). В тоже время, как известно, ряд методов Data Mining способен работать с разнотипными признаками. Применению этих методов в биоинформатике посвящено много современной литературы (например [1-3]).

Вместе с тем, процедура множественного выравнивания способна вносить существенный произвол в представление информации для дальнейшего анализа. Особенно это становится заметным при больших длинах исследуемых символьных рядов, когда различные варианты множественного сравнения приводят к близким значениям функционала качества. В связи с этим наше внимание, как и ряда других исследователей, было сконцентрировано на проблеме поиска регулярностей в молекулярно-биологических

последовательностях без использования процедуры множественного выравнивания.

Для поиска регулярностей в молекулярно-биологических последовательностях наиболее часто применяются методы, берущие свое начало в анализе временных рядов. Продуктивным направлением анализа рядов сегодня является подход, связанный с преобразованием ряда в матрицу с помощью однопараметрической сдвиговой процедуры «Гусеница» [4].

Построенную матрицу можно рассматривать как M -мерную выборку объема k или M -мерный временной ряд, которому соответствует M -мерная траектория (ломаная в M -мерном пространстве из $k - 1$ звена). Матрица представлена в традиционном для прикладной статистики виде «строка – объект, столбец – признак». Для ее дальнейшей обработки можно применять различные методы из богатого арсенала математического аппарата многомерного анализа.

3 Примеры

Для поиска логических правил в экспериментальных данных использовалась технология SRMD (Structural Resonance in Multidimensional Data), основанная на представлениях локальной геометрии [5]. В этой технологии используется аппарат линейной алгебры и процедура активного формирования информационного структурного резонанса в многомерных данных.

Основные характеристики технологии SRMD: Нахождение «сильных» (наиболее полных при заданной точности) «ЕСЛИ–ТО» правил для каждой записи базы данных; Построение и тестирование классификаторов данных на основе «ЕСЛИ–ТО» правил; Построение «нечетких» «ЕСЛИ–ТО» правил. С помощью технологии SRMD удастся создавать компактные и эффективные экспертные системы для получения высокоточных диагностических и прогностических решений [6]. Программная реализация технологии – система Argos Data Mining (Свидетельство об официальной регистрации программы для ЭВМ №2007612080).

Первый пример относится к поиску паттернов в ДНК *e-coli* (кишечной палочки). В рассмотренном примере во фрагменте ДНК удалось найти шаблон

AxAxA, где в позиции «х» может стоять любой из четырех символов **A, C, G, T**. Обращает на себя внимание то, что найденный шаблон появляется в последовательности ДНК через различные по длине интервалы (8, 8, 7, 8). Аналогичные шаблоны встречались и в других фрагментах ДНК *e-coli*.

Для следующего более сложного примера использованы данные из базы данных Eukaryotic Promoter Database (EPD) [7].

Из базы были извлечены 1871 последовательностей нуклеотидов *Homo sapiens* (man) в диапазоне -100 +100 от сайта старта транскрипции. Каждая последовательность нуклеотидов была преобразована в матрицу с помощью процедуры «Гусеница» (длина гусеницы равна 80 позиций). К матрице ряда была «пристыкована» матрица «шума» – случайно организованных последовательностей, количество которых в 5 раз превышало количество строк в матрице исходного ряда. При этом для генерации случайных последовательностей использовалась специальная процедура, особенность которой заключалась в том, что матрица «шума» образовалась из исходной матрицы реальной последовательности путем случайной перестановки в ней элементов столбцов. Тем самым, в матрице «шума» соблюдалась полная идентичность частот символов частотам символов в столбцах реальной матрицы, тогда как связи между элементами строк становились разрушенными.

В результате анализа подготовленных описанным способом данных с помощью программы Argos Data Mining в каждой из таблиц удавалось найти примерно от 3-х до 10 логических правил, характерных только для «реальных» частей таблицы данных. Частота выполнения логических правил на «реальных» частях таблицы была невысока. Она не превышала 10 % от размера этих частей.

В ряде случаев выявлялась картина, которая в значительной мере перекликается с паттерном для ДНК *e-coli*. В ДНК *e-coli* был найден паттерн **AxAxA**, в котором как бы в пакете идет символ «A». Для последовательностей нуклеотидов *Homo sapiens* наблюдалась аналогичная ситуация только уже с символами «G» и «C». Для этих символов выявлялись идентичные паттерны **GxGxGxGG** и **CxCxCxCC**, где «х» – джокер. Подобные паттерны, более или менее выраженные, нередко встречались и в других последовательностях ДНК человека с промоторами из базы EPD. Этот занятый факт конечно пока нельзя принимать как открытие нового

явления, он нуждается в многократной перепроверке.

3 Выводы

1. Поиск сложных неперiodических шаблонов в последовательностях символов методами локальной геометрии с использованием эффекта информационного структурного резонанса представляет интерес для целого ряда областей, связанных с анализом временных и иных рядов в биологии, медицине, технике и экономике.

2. Особую ценность данные методы, по-видимому, имеют в современных молекулярно-генетических исследованиях, в которых наступил этап выяснения функционального смысла различных участков секвенированной ДНК.

3. В ряде проанализированных последовательностей ДНК (*e coli* и *Homo sapiens*) с помощью технологии информационного структурно резонанса удалось обнаружить относительно сложные шаблоны с джокерами, следующие с изменяющимся периодом, которые могут представить интерес для теоретической геномики.

Список литературы

- [1] Wang, J.T.L., Zaki, M.J., Toivonen, H.T.T., Shasha, D.E. Data Mining in Bioinformatics. – 2005, Springer, 340 p.
- [2] Eibe Frank, Mark Hall, Len Trigg, Geoffrey Holmes, and Ian H. Witten. Data mining in bioinformatics using Weka. Bioinformatics, Oct 2004; 20: 2479 – 2481
- [3] J. E. Gewehr, M. Szugat, and R. Zimmer. BioWeka extending the Weka framework for bioinformatics. Bioinformatics, March 1, 2007; 23(5): 651 - 653.
- [4] Голяндина Н.Э. Метод «Гусеница»-SSA: анализ временных рядов: Учеб. пособие. СПб: Изд-во СПбГУ, 2004. 76 с.
- [5] Дюк В.А., Толстоногов Д.А. Опыт применения метода информационного структурного резонанса в задачах анализа данных ДНК микрочипов // Биотехносфера, 2010 № 5-6 (11-12), 2010. С. 52-59.
- [6] Дюк В.А., Эмануэль В.Л. Информационные технологии в медико-биологических исследованиях. СПб.: Питер. 2003. 525 с.
- [7] Eukaryotic Promoter Database – <http://www.epd.isb-sib.ch/>.

Анализ распределения триплетной периодичности между генами одного генома.

Суворова Ю.М., Коротков Е.В.

Центр “Биоинженерия” РАН
suvorovay@gmail.com

Triplet periodicity is distinguishing property of coding sequences of all living organisms. In our work we investigated how triplet periodicity distributed between genes of one genome (on example of the *Escherichia coli* genome). We performed a pairwise comparison of all *E.coli* genes and then computed the distribution of difference between them. From comparison the distribution of real genes with two models one can conclude that most of *E.coli* genes have similar triplet periodicity type.

1 Введение

Триплетная периодичность – отличительное свойство последовательностей, кодирующих белки. ТП ассоциируется с действующей рамкой считывания в гене [2]. Триплетная периодичность (ТП) используется для отделения кодирующих последовательностей от некодирующих [3], поиска мутаций сдвига рамки считывания [5] и склейки в генах [6]. Возникновение ТП объясняют триплетной структурой генетического кода [7], а также предпочтение определенных триплетов для кодирования некоторых аминокислот [1,8]. Мы провели исследование, насколько гены внутри одного бактериального генома однородны с точки зрения триплетной периодичности. Изучение вопроса об однородности триплетной периодичности внутри генома, а также сравнение распределений ТП в разных геномах представляет интерес с точки зрения эволюции генов и геномов.

2 Материалы и методы

2.1 Алгоритм сравнения

В первую очередь было проведено сравнение генов одного генома (на примере генома *Escherichia coli*) друг с другом с точки зрения триплетной периодичности. Кодирующие последовательности были загружены с сайта базы данных KEGG [4].

Рассмотрим кодирующую последовательность S длины L (L кратно трем). Для этой последовательности целиком заполняется частотная матрица M . Элемент матрицы $m(i,j)$ равен числу элементов типа i ($i=1$ для ‘a’, $i=2$ для ‘t’, $i=3$ для ‘g’ и $i=4$ для ‘c’), находящихся в позиции j кодона ($j=1,2,3$). Например, элемент $m(1,2)$ равен числу символов ‘t’, стоящих на вторых позициях кодонов в

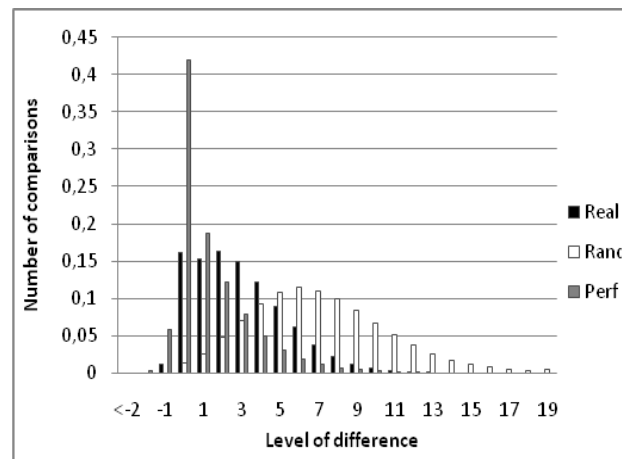
рассматриваемом гене. Затем каждая матрица была приведена к нормальному виду, с использованием преобразования

$$n(i, j) = \frac{m(i, j) - Lp(i, j)}{\sqrt{Lp(i, j)(1 - p(i, j))}}$$
$$p(i, j) = \frac{(\sum_{i=1}^4 m(i, j)) \cdot (\sum_{j=1}^3 m(i, j))}{L^2}$$

Полученные матрицы попарно сравнивались между собой, и для каждой пары было рассчитано значение различия между ними.

$$D = \sum_{i=1}^4 \sum_{j=1}^3 \left(\frac{n_1(i, j) - n_2(i, j)}{\sqrt{2}} \right)^2$$
$$N = \sqrt{2D} - \sqrt{2df - 1}$$

Полученное распределение величины различия N между матрицами генов генома *Escherichia coli* (в отношении к общему числу сравнений) представлено на Рисунке черным цветом.



2.2 Построение модельных выборок

В качестве контроля мы использовали дополнительные распределения, полученные с использованием двух искусственных выборок. Для построения первой выборки последовательность каждого гена *Escherichia coli* была случайным образом перемешана. Затем путем случайных перестановок символов, уровень триплетной периодичности полученной последовательности постепенно восстанавливался до уровня исходного гена. Эта выборка соответствует гипотезе о том, что триплетная периодичность между генами одного генома распределена случайно и независимо. Для

последовательностей этой выборки было проведено сравнение частотных матриц такое же, как для реальных генов. На рисунке это распределение показано белым цветом.

Для получения второй контрольной выборки были созданы последовательности с совершенной периодичностью одного типа (повтор триплета “acT”), длина которых равна длине исследуемых генов. Затем периодичность этих последовательностей постепенно размывалась, путем случайных замен до уровня периодичности исходного гена. Этот тип моделирования соответствует гипотезе о том, что все гены генома изначально соответствуют одному типу периодичности. Было построено распределение величины различия генов полученной выборки, оно представлено на Рисунке серым цветом.

3 Результаты

Из сравнения трех полученных распределений видно, что хотя не все реальные гены имеют матрицу одного типа, их распределение ближе ко второй гипотезе, чем к первой. Т.е. большая часть генов *Escherichia coli* имеет триплетную периодичность похожего типа. Из распределения также можно заметить, что около 10% генов генома *Escherichia coli* все же значительно отличаются по триплетной периодичности.

Рассмотрение других бактериальных геномов показало, что общая картина сохраняется, т.е. большая часть генов в них имеют похожую

триплетную периодичность. Однако геномы отличаются между собой в проценте неоднородных генов.

Список литературы

- [1] Antezana M.A., Kreitman M. The nonrandom location of synonymous codons suggests that reading frame-independent forces have patterned codon preferences. *J. Mol. Evol.* 49, 36–43 1999
- [2] Bernaola-Galván et.al. Finding Borders between Coding and Noncoding DNA Regions by an Entropic Segmentation Method, *Physical review*, 86 (6) 1342-1345, 2000.
- [3] 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.
- [4] M. Kanehisa and S. Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27-30, 2000.
- [5] E.V. Korotkov and M.A. Korotkova. Study of the triplet periodicity phase shifts in genes, *Journal of Integrative Bioinformatics*, 7(3):131, 2010.
- [6] Suvorova, Y.M., Rudenko, V.M., Korotkov, E.V., Detection change points of triplet periodicity of gene. *Gene*. 491, 58-64. 2012
- [7] Trifonov E.N.. Elucidating sequence codes: three codes for evolution. *Annals New York Acad. Sci.* 870, 330–338. 1999
- [8] Zoltowski M. Is DNA Code Periodicity Only Due to CUF – Codons Usage Frequency? Conf. Proc. IEEE Eng. Med. Biol. Soc. 1, 1383–1386. 2007

Метод поиска сдвигов фазы периодичности в минисателлитах

Пугачева В.М.⁽¹⁾, Коротков Е.В.⁽¹⁾

Учреждение Российской академии наук Центр «Биоинженерия» РАН⁽¹⁾
virentis@gmail.com

Mini and micro satellites in DNA are repeats of several time repeat of some pattern, periodical sequence with substitutions and indels. Mini and microsatellites may play a variety of regulatory and evolutionary roles. Most of known methods for searching tandem repeats have restrictions related to pattern, size of pattern, mutations in pattern and especially indel mutations. We present a new method for finding phase shifts in periodical sequences, i.e. tandem repeats with substitutions and indels.

1 Введение

Одним из важных механизмов изменчивости генома эукариот являются геномные перестройки, в частности тандемные дубликации и мультипликации. Микро- и минисателлиты могут привести к возникновению новых промоторных областей, сайтов альтернативного сплайсинга и т.д. После возникновения сателлитной последовательности возможно изменение последовательности в результате точечных мутаций, что может привести к невозможности его идентификации методами динамического программирования. Таким образом, требуется разработка метода поиска сильно вырожденных, тандемно повторяющихся мотивов в последовательностях ДНК.

Согласно существующей теории [8] при возникновении микро- и минисателлиты не содержат ошибок (замен, делеций и вставок), но в силу происходящих в процессе биологической эволюции точечных мутаций сателлитные последовательности их накапливают. Т.о. необходимо учитывать возможность нуклеотидных замен, вставок и делеций.

В настоящий момент существуют различные программы для поиска периодических структур в последовательностях ДНК. Но пока нельзя сказать, что существует единственно возможный правильный алгоритм поиска. Каждый алгоритм нацелен на определенную задачу и класс определяемых повторов; степень их вырожденности, длину периода и другие параметры. Часть таких программ нацелена на определение повторов ограниченных типов, причем в основном это микросателлитные повторы и некоторые

минисателлитные повторы ограниченной длины. Это методы поиска поиска точных коротких повторяющихся фрагментов или точных тандемных повторов с заранее заданными мотивами, с помощью этих программ возможно искать только невырожденные тандемные повторы [2,4,5,7].

Другая часть таких программ использует эвристические подходы и позволяет искать вырожденные повторы. Но они не лишены недостатков: не учитывают вставки и делеции, либо при поиске повторов используются общие статистических свойства присущие повторам, что может приводить к тому, что многие статистические особенности последовательностей становятся слабо выраженными, что может значительно ухудшать мощность таких методов. Многие хорошо известные программы поиска тандемных повторов в ДНК [1,3,6] ищут вырожденные тандемные повторы, где мотив повторяется с некоторым уровнем вариаций, определенным весовой системой. Например, подход, используемый для определения повторов в хорошо известной программе Tandem Repeat Finder [3], можно сравнить с известным алгоритмом BLAST. Он основан на первоначальном получении информации о коротких точных повторяющихся фрагментах, и дальнейшего удлинения этих фрагментов в соответствии с неким статистическим критерием уже для вырожденных тандемных повторов.

2 Метод поиска сдвигов фазы периодичности различной длины

Предлагаемый метод поиска сдвигов фазы периодичности позволяет находить тандемные повторы в ДНК учитывая не только замены, но и вставки и делеции. Анализируя последовательность ДНК, мы меняем координату с заранее заданным шагом и проверяем для каждой следующей координаты x наличие сдвига фазы периодичности. Для длины периода $period$, для каждой координаты x мы рассматривает участок последовательности слева от x в 1-ой фазе и $period$ участков последовательности справа от x в фазах от 1-ой до $period$. Для сравнения всех пар левого и правого от x участков мы ввели меру подобию $Z(i)$ соответствующих частотных матриц этих участков (участка слева и i -го участка справа). С ростом $Z(i)$ увеличивается вероятность сдвига фазы периодичности после координаты x на i позиций вправо. Затем мы оценивали статистическую

значимость найденных сдвигов фазы. Для определения порогового значения мы использовали метод Монте-Карло. Так как статистическая значимость отражает вероятность найти структуру подобную данной в случайном тексте, сгенерированном в соответствии с определенной моделью, мы перемешивали исходную последовательность определенным образом (с учетом анализируемой длины периода, т.е. отдельно мешали последовательности составленные из первых, вторых и т.д. до длины периода позициях). После анализа исходной и перемешанных последовательностей мы выбрали пороговое значение Z_0 , сдвиги фазы имеющие $Z(i) > Z_0$ считали значимыми.

Мы проанализировали геном организма *C.elegans* для ряда длин периодов (2-11, кроме кратных 3). В качестве уровня значимости было выбрано значение, при котором число сдвигов фазы периодичности в перемешанных последовательностях было менее 5% от числа сдвигов фазы периодичности в последовательности генома *C.elegans*. Затем мы сравнили полученные результаты с данными Tandem Repeat Data Base, результат сравнения приведен в таблице 1. Т.о. мы обнаружили более 10 тысяч, не найденных ранее, тандемных повторов со сдвигами фазы периодичности.

Таблица 1.

Период	Повторов сдвигами фазы	Повторов Tandem Repeats Data Base	Не известные ранее повторы
2	1424	587	1365
4	4049	143	4008
5	3226	67	3213
7	8045	504	7919
8	30314	485	30125
10	46463	1083	46182
11	28754	2081	28055

Список литературы

- [1] C. Abajian, *Sputnik*. 1994. URL: <http://abajian.net/sputnik/>
- [2] G. Benson and M.S. Waterman. *A method for fast database search for all k-nucleotide repeats*. Nucleic Acids Res, 1994. 22(22): p. 4828-36.
- [3] G. Benson. *Tandem repeats finder: a program to analyze DNA sequences*. Nucleic Acids Res., 1999. 27: p. 573-578.
- [4] A.T. Castelo, W. Martins and G.R. Gao. *TROLL--tandem repeat occurrence locator*. Bioinformatics, 2002. 18(4): p. 634-6.
- [5] A.M. Hauth and D.A. Joseph. *Beyond tandem repeats: complex pattern structures and distant regions of similarity*. Bioinformatics, 2002. 18 Suppl 1: p. S31-7.
- [6] R. Kolpakov, G. Bana, and G. Kucherov. *mreps: efficient and flexible detection of tandem repeats in DNA*. Nucleic Acids Res., 2003. 31: p. 3672-3678.
- [7] E. Rivals, O. Delgrange, J.P. Delahaye, M. Dauchet, M.O. Delorme, A. Henaut and E. Ollivier. *Detection of significant patterns by compression algorithms: the case of approximate tandem repeats in DNA sequences*. Comput Appl Biosci, 1997. 13(2): p. 131-6.
- [8] C. Schlotterer. *Evolutionary dynamics of microsatellite DNA*. Chromosoma, 2000. 109(6): p. 365-71.

Спектральные методы в задаче поиска различных типов повторов в геномах

Пятков М.И.⁽¹⁾, Панкратов А.Н.^(1, 2)

Институт Математических Проблем Биологии РАН⁽¹⁾,
Московский Государственный Университет, факультет Вычислительной Математики и
Кибернетики⁽²⁾
mryatkov@gmail.com

The paper presents recent results in developing an approach to find long repeats in nucleotide sequences based on the generalized spectral-analytical method. In particular, shows the results of a simultaneous search for different types of repeats, the analysis of nucleotide sequences at different scales and increased filtration by embedding additional features.

Одним из самым значимым достижений последнего десятилетия является секвенирование генома человека и основных модельных эукариотических организмов. Появившиеся в распоряжении исследователей огромные хранилища генетической информации позволяют проводить полномасштабные исследования этих геномов и продвигаться вперед в понимании их структурно-функциональной организации. Одной из легко формализуемых задач, связанных с секвенированными последовательностями, является исследование повторяющихся элементов (повторов), изучение их структуры и распределения в геномах.

Авторами рассматривается задача поиска протяженных неточных повторов и предлагается подход для решения данной задачи, в основе которого лежат спектральные методы. Отличительной особенностью предлагаемого алгоритма, по сравнению с существующими аналогами, является высокая эффективность работы за счет использования многоуровневого параллелизма, анализ нуклеотидных последовательностей в разном масштабе, а также возможность поиска различных типов повторов (прямых, инвертированных, tandemных) одновременно.

В данной работе представлено дальнейшее развитие вычислительного метода [1,2] для поиска протяженных повторов в нуклеотидных последовательностях, наиболее полное описание и приложение алгоритма представлено в работе [3].

В ходе настоящего исследования был получен ряд результатов направленных на улучшение качества распознавания, а именно:

- 1) Повышена устойчивость распознавания повторов за счёт введения дополнительной кривой GA-содержания, которая делает информационное описание последовательности более полным;
- 2) Выведено усовершенствованное решающее правило, инвариантного к выбору масштаба;
- 3) Определены условия аппроксимации, при которых ранее найденные повторы [4,5] отображались на матрице гомологии в виде квадрата, что позволило найти новые подобные повторы [6];
- 4) Построен алгоритм поиска инвертированных повторов в пространстве коэффициентов разложения.

Работа выполнена при поддержке гранта РФФИ №11-07-00716.

Список литературы

- [1] Дедус Ф.Ф., Куликова Л.И., Махортых С.А., Назипова Н.Н., Панкратов А.Н., Тетюев Р.К. Аналитические методы распознавания повторяющихся структур в геномах. Доклады Академии Наук. 2006. 411. 599--602.
- [2] A.N. Pankratov, R.K. Tetuev et al. Spectral Analysis for Identification and Visualization of Repeats in Genetic Sequences. Pattern Recognition and Image Analysis. 2009. 19(4). 687--692
- [3] Панкратов А.Н., Пятков М.И., Тетюев Р.К., Назипова Н.Н., Дедус Ф.Ф. Поиск протяженных повторов в геномах на основе спектрально-аналитического метода. Математическая биология и биоинформатика. 2012.7(2). 476-492.
- [4] Tetuev R.K., Nazipova N.N.. Repbase Reports. 2010. 10(5). 776.
- [5] Tetuev R.K., Nazipova N.N., Dedus F.F. Repbase Reports. 2010. Vol.10(8). 1185.
- [6] Pyatkov M.I., Filippov V.V., Pankratov A.N. Repbase Reports. 2012. Vol.12(3). 256.

Новые элементы бактериальных геномов – «промоторные островки», как маркеры чужеродной ДНК

Озолин О.Н.^(1, 2), Киселев С.С.^(1, 2), Панюков В.В.⁽³⁾

Институт биофизики клетки РАН⁽¹⁾, Пушкинский государственный естественно-научный институт⁽²⁾, Институт математических проблем биологии РАН⁽³⁾
ozoline@icb.psn.ru

In this study we used a unified version of the promoter finder PlatProm (PlatPromU) so as to find genomic regions with a high density of potential promoters with mixed σ -specificity (“mixed promoter islands”, MPIs). We found that as well as “promoter islands” revealed by the σ^{70} -specific promoter finder PlatProm, they are often located nearby the foreign genes or overlap with alien DNA. This observation provides an opportunity to consider the genomic regions enriched by transcription initiation signals as reliable signs of horizontally acquired genes.

1 Introduction

Gene expression in all organisms is carried out by RNA polymerase(s) and is controlled by multiple σ -factors or regulatory proteins that cooperate in the transcription initiation by interacting with specific genomic regions – promoters. Most of the promoters are located in front of the sequences encoding proteins or untranslated RNAs. In addition to them, all promoter finders predict transcription initiation signals within genes, where they may be involved in the synthesis of antisense or alternative RNAs [8]. Though the biological significance of many such RNAs is not yet clear, their synthesis from the intragenic regions is confirmed by modern experimental techniques, which make it possible to map the transcription start points (TSPs) in the living cell [1]. Except for the normal and intragenic promoters with one or several points of transcription initiation, all bacterial genomes investigated so far contain areas of anomalously enriched by potential promoters [8]. Thus, our promoter finder PlatProm, configured to search for promoters recognized by the major σ -factor (σ^{70}), found in the chromosome of *E.coli* 78 such regions. We called them “promoter islands” or PIs [8]. All 78 PIs form the transcription complexes and initiate the synthesis of short abortive RNAs, but are defective in the normal transcription. It has been shown that this suppression at least in part is mediated by the protein H-NS, which represses the expression of foreign genes and 73 PIs are associated with horizontally acquired DNA. However Lawrence et al. [5] have found 706 foreign genes in the genome of *E.coli*, and Nakamura et al. [6] even more -

1038. Therefore in order to consider the possibility that PIs can mark alien DNA, we evaluated the transcription activity and the degree of association with the foreign genes for the “mixed promoter islands” (MPIs), revealed by the unified promoter finder PlatPromU [4], which ignores the σ^{70} -specific modules and can find the promoters, recognized by all seven σ -subunits of *E.coli*.

2 Methods

2.1 Genome and promoter compilation

MPIs were searched for in the genome of *E.coli* MG1655 (NC_000913 in NCBI database). Their properties were compared with 352 promoter regions, non-overlapping with MPIs and containing at least one promoter recognized by σ^{70} . Appropriate compilation was created using RegulonDB [2].

2.2 Expression analysis

Functional activity of MPIs was assessed using the 5'-end specific RNA-seq data [1], which contain 8967903 44 bp long sequence reads (samples). Raw data were evaluated by the software RNAMatcher, which maps samples to the genomic DNA, estimates their number, and reports if evaluated sample has multiple matching in the genome. The samples, perfectly matching to the sequences of genomic DNA, were considered as representatives of normal transcription. The samples, which have only 11 nucleotides at the 5'-end, coinciding with the sequences of genomic DNA, and contain adapter sequence at the 3'-part were considered as abortive products.

2.3 Overlap with horizontally acquired genes

Association with the alien DNA was determined using a list of foreign genes, compiled on the basis of codon usage bias [5] or/and obtained by the methods of comparative genomics [6], summarized in [7].

3 Results

3.1 Genome of *E.coli* possesses 460 MPIs

PlatProm assesses the probability for a given genomic position to be the transcription start point

(TSP). It operates with position weight matrices, which score the contribution of σ^{70} -specific hexanucleotides, and with more than 50 cascade matrices, which take into account the specific structural modules, favoring adaptive isomerisation of promoter DNA in the transcription complex. The evolutionary conservatism of the transcription machinery offers an opportunity to find promoters recognized by different σ -factors using the unified program PlatPromU, which operates only with cascade matrices.

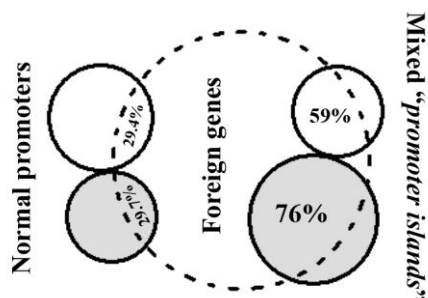


Fig.1 Degree of overlap of normal promoters and MPIs with foreign genes (dashed circle). Portions of promoters and MPIs, for which long RNAs were not registered is shadowed.

Using PlatPromU and the threshold providing $p < 0.00004$, defined as described in [4], we found several thousand of promoter regions. Following the same criteria as described in [1] (≥ 8 TSPs on either DNA strand in a sliding window 100 bp along at least 300 bp) we identified 600 MPIs. All previously defined PIs were included in this set. Their minimum length was 360bp, i.e. promoters with different σ -specificity can be found in one MPI. To make this set comparable in size to the compilation of normal promoters we increase the cut-off length up to 360 bp, which gave 460 MPIs for comparisonⁱ.

3.2 Transcription activity of MPIs is similar to PIs

Only 156 MPIs (34%) produce RNAs with length ≥ 44 nucleotides, which is less than in the case of normal promoters (54%). Moreover, the total number of sequence reads, matching to the MPIs was 2-fold lower than expected by chance, while in the case of normal promoters the number of samples matching to the ± 50 bp regions around TSPs was, by contrast, 27-fold higher. For all that 94% of MPIs and only 65% of normal promoters produced abortive RNAs. Thus, 460 MPIs show the same functional properties as 78 PIs.

3.3 Most MPIs are associated with foreign genes

MPIs or normal promoter were considered as associated with alien DNA if they overlap or are located nearby horizontally acquired genes, according to

Nakamura et al. [6] and/or Lawrence et al. [5]. Less than 30% of normal promoters passed this criterion, regardless of their productive activity. In the case of MPIs situation is quite different: 59% of the MPIs, initiating synthesis of normal RNAs, or 76% of “islands”, producing only abortive products are associated with foreign genes (Fig. 1). Thus, “promoter islands” with a mixed σ -specificity can be involved in evolution of horizontally acquired genes.

4 Conclusion

Previously it has been shown that the extended (4000 – 48000bp) “Genomic islands”, consisting of alien DNA and often containing pathogenic genes, are characterized by high frequency of transcription initiation signals [3]. Here we have shown that the “promoter islands”, containing an extremely large number of TSPs, are associated with both “Genomic islands” and with foreign genes, distributed throughout the bacterial chromosome. The mechanisms, underlying adaptation of foreign genetic material into the regulatory networks of host cells, remain unknown. “Promoter islands” can take part in these processes by offering an ensemble of suitable promoters for beneficial genes or by suppressing expression of dangerous genes, which were integrated by chance.

The study was supported by the Russian Foundation for Basic Research (grants No.10-04-01218 and No. 12-07-00530) and Russian Ministry of education and science (project: “Multiple promoters as indicators of overlapping transcription and alternative coding”).

References

- [1] J.E. Dornenburg et al. *mBio* 2010, **1**, e00024-10.
- [2] S. Gama-Castro et al. *Nucl. Acids Res.* 2010; doi: 10.1093/nar/gkq1110
- [3] Q. Huang, X. Cheng, M.K. Cheung, S.S. Kiselev, O.N. Ozoline, H.S. Kwan, *PLoS One*, 2012, **7**, e33759.
- [4] S.S. Kiselev, O.N. Ozoline, *Mathematical Biology and Bioinformatics*, 2011, **6**, t1-t13, URL: [http://www.matbio.org/2011/Ozoline2011\(6_t1\).pdf](http://www.matbio.org/2011/Ozoline2011(6_t1).pdf)
- [5] J.G. Lawrence, H. Ochman, *Proc. Natl Acad. Sci. USA*, 1998, **95**, 9413–9417.
- [6] Y. Nakamura et al. *Nat. Genet.*, 2004, **36**, 760–766.
- [7] T. Oshima et al. *DNA Research*, 2006, **13**, 141–153.
- [8] K.S. Shavkunov, I.S. Masulis et al. *Nucl. Acids Res.*, 2009, **37**, 4919 – 4931.

Структура и эволюция альфа-L-рамнозидаз

Наумов Д.Г.

Институт микробиологии им. С.Н. Виноградского РАН, Москва 117312
daniil_naumoff@yahoo.com

α -L-Rhamnosidases catalyze the hydrolysis of the terminal α -L-rhamnose residues in various carbohydrates. The catalytic domains in most of these enzymes belong to the families GH78 and GH106 of glycoside hydrolases. More than 1700 non-identical protein sequences of GH78 and GH106 domains have been revealed using the blast algorithm. They represent several phyla of Bacteria, as well as some Archaea and Eukaryota (mainly fungi). Proteins from the phyla *Acidobacteria*, *Bacteroidetes*, and *Verrucomicrobia* were overrepresented compared to the phyla *Actinobacteria*, *Firmicutes*, and *Proteobacteria*. Phylogenetic analysis of GH78 and GH106 families suggested multiple events of lateral gene transfer, as well as the gene duplications and eliminations. In particular, 22 acidobacterial hypothetical α -L-rhamnosidases belong to 13 distinct clusters on the phylogenetic trees. The majority of their genes evolved as the result of several independent gene transfers from the *Bacteroidetes*. Iterative screening of the protein database allowed us to reveal relationship of GH78 with several other families of glycoside hydrolases having the $(\alpha/\alpha)_6$ -toroid type of the 3D structure: GH15 (clan GH-L), GH37 (GH-G), GH63 (GH-G), GH65 (GH-L), GH92, GH94, GH95, GH100, and GH116. PSI-BLAST searches using GH106 domains as queries found their relationship with the $(\beta/\alpha)_8$ -barrel domains belonging to GH2 (clan GH-A), GH5 (GH-A), GH13 (GH-H), GH35 (GH-A), GH39 (GH-A), GH42 (GH-A), and GH133 families of glycoside hydrolases.

1 Введение

α -L-Рамнозидазы (К.Ф. 3.2.1.40) – широко распространённая и промышленно значимая группа ферментов, ответственных за отщепление с нередуцирующего конца терминальных остатков α -L-рамнозы от углеводов и их производных. На основании гомологии каталитических доменов почти все ферменты этой группы отнесены к семействам GH78 и GH106 гликозил-гидролаз [1]. Домены семейства GH78 имеют структуру $(\alpha/\alpha)_6$ -тороида (PDB, 2OKX и 3CIN), а пространственная структура GH106-доменов остаётся неизвестной. Согласно базе данных CAZy

[1], эти семейства содержат 463 и 85 белка соответственно. Ферменты из этих двух семейств обладают только α -L-рамнозидазной активностью, но несколько различаются по субстратной специфичности. Настоящая работа посвящена исследованию эволюционных взаимоотношений между белками в пределах семейств GH78 и GH106, а также поиску их дальних гомологов.

2 Методы и алгоритмы

Поиск белков, содержащих домены семейств GH78 и GH106, проводили с помощью программы blastp, используя 25 и 69 представителей этих семейств соответственно. Множественное выравнивание 358 и 343 доменов семейств GH78 и GH106, соответственно, проводили вручную с помощью программы BioEdit. Филогенетические деревья белковых семейств строили с помощью программ Neighbor-Joining и Maximum Parsimony из пакета PHYLIP. Эволюционное родство между семействами белков устанавливали с помощью программы PSI Protein Classifier [2], позволяющей анализировать результаты последовательных и/или независимых итераций программы PSI-BLAST. Каталитические домены нескольких наиболее дивергентных представителей семейств GH78 и GH106 использовали в качестве запросов при итеративном скрининге базы данных.

3 Результаты и обсуждения

3.1 Поиск белков и их доменная структура

Нам удалось обнаружить 1222 и 486 белков, содержащих домены семейств GH78 и GH106 соответственно. При этом в обоих семействах преобладают белки из бактерий, хотя есть представители эукариот (главным образом грибы) и архей. Среди отделов бактерий оказались перепредставленными *Acidobacteria*, *Bacteroidetes* и *Verrucomicrobia* по сравнению с *Actinobacteria*, *Firmicutes* и *Proteobacteria*. Анализ белков обоих семейств показал, что большинство из них наряду с каталитическим доменом также содержит дополнительные домены, некоторые из которых консервативны у многих белков в пределах каждого из семейств. Это указывает на доменную структуру предковых белков.

3.2 Филогенетический анализ

Построение множественных выравниваний и последующий филогенетический анализ выявил многочисленные факты дупликаций и горизонтальных переносов генов, кодирующих белки семейств GH78 и GH106. Например, 22 гипотетических α -L-рамнозидазы из бактерий отдела *Acidobacteria* оказались принадлежащими 13 различным кластерам на филогенетических деревьях. При этом большинство из этих белков кодируются генами, которые по всей видимости попали в ацидобактерии путём горизонтального переноса из бактерий отдела *Bacteroidetes*.

Геном *Clostridium methylpentosum* DSM5476 (GenBank, ASEC000000000.1) кодирует по меньшей мере 83 белка, содержащих домены семейств GH78 или GH106, в то время как большинство доступных геномов других штаммов *Clostridium* таких белков не кодирует. В геноме протеобактерии *Glaciecola* sp. HTCC2999 (GenBank, NZ_ABST01000019.1) закодировано по меньшей мере 12 белков этих двух семейств, а геномы двух других штаммов этого рода (GenBank, CP002526.1 и CP003060.1) таких белков не кодируют.

3.3 Дальние эволюционные связи

Итеративный скрининг базы данных аминокислотных последовательностей с помощью доменов семейства GH78 позволил выявить их эволюционное родство с каталитическими доменами нескольких других семейств гликозил-гидролаз, также имеющих структуру $(\alpha/\alpha)_6$ -тороида: GH15 (клан GH-L), GH37 (GH-G), GH63 (GH-G), GH65 (GH-L), GH92, GH94, GH95, GH100 и GH116. Аналогично удалось проследить эволюционные связи доменов семейства GH106 с доменами семейств GH2 (клан GH-A), GH5 (GH-A), GH13 (GH-N), GH35 (GH-A), GH39 (GH-A), GH42 (GH-A) и GH133 гликозил-гидролаз, имеющих структуру $(\beta/\alpha)_8$ -бочонка. Это позволяет предположить, что домены семейства GH106 имеют такой же тип пространственной структуры.

Список литературы

- [1] Carbohydrate Active Enzymes server. 2012. (<http://www.cazy.org/>).
- [2] Д.Г. Наумов, М. Карперас. Новая программа PSI Protein Classifier автоматизирует анализ результатов программы PSI-BLAST. Молекулярная биология, 2009, **43**(4), 709-721.

Modelling co-evolution of regulatory systems, genes and species with supercomputers

K.Yu. Gorbunov, V.A. Lyubetsky

Institute for Information Transmission Problems of the Russian Academy of Sciences
(Kharkevich Institute), 19 Bolshoy Karetny, 127994 Moscow, Russia
gorbunov@iitp.ru

1 Abstract

An original model is developed to describe the co-evolution of regulatory systems, genes and species (ref. to sections 2-5). The model was extensively tested on simulated and biological data. We will present the underlying supercomputer techniques and the obtained results.

2 Reconciliation of a set of gene trees

A long recognized problem is inference of a tree S that reconciles a set of input trees G_i , with leaves in each G_i being assigned homologous sequences from an i -family (homologous genes or regulatory regions with or without the regulated gene, etc.). Usually the tree S is a tree of species or other taxonomic units. Assume leaves in S are labelled with species names, leaves in G_i – with pairs of gene-species names; paralogs are allowed. We further develop a traditional approach to find the tree S such that it minimizes the total cost of mappings of individual trees into S , [6]. Let us call S a supertree, each G_i – a gene tree, each sequence – a gene. A mapping of G_i into S implies a fixed set of evolutionary events. The *standard set* contains only gene duplications and losses. The *extended set* additionally contains horizontal gene transfers, gains, etc. The *total cost* is the sum of “individual” costs for all G_i mappings into the tree S , and is similar to the cost from [6] in case of the standard set. Is computing total/individual costs possible in small polynomial time? This question is tackled below.

Under a traditional approach, the supertree building problem is NP-hard, i.e., any algorithm to solve it correctly must possess exponential complexity. Numerous heuristics exist, but they generally do not find the global minimum of the total mapping cost. We proposed a reformulation of this problem that allows a computationally effective deterministic algorithm and meets many biological prerequisites. Namely, the supertree S is sought for such that it contains the majority of clades from input trees G_i . With the standard event set, the algorithm is mathematically correct and possesses the running time of $O(n^3 \times m^3)$, where n is the number of gene trees, and m is the total number of species [2]. For simplicity, here we assume that the average number of leaves in input trees is a multiple of m . With an extended event set, the algorithm is heuristically correct and cubic in complexity. The

authors are unaware of analogous approaches in published literature. A relevant biological discussion of our approach is provided in [3].

The algorithm realization is available at <http://lab6.iitp.ru/en/super3gl/>.

3 Reconciliation of gene and species trees

Edges in S may be broken by inserting additional nodes, thus formally producing another tree S_0 , with nodes producing two descendant edges or only one edge. It imposes time slices such that horizontal transfers are allowed only within one slice, see [4]; in particular cases $S=S_0$. With the extended event set, we developed an algorithm that reconciles any gene tree G and S_0 , i.e., correctly computes the mapping of G into S_0 and its cost in time $O(|G| \times |S_0|)$, which gives $O(|S|^3)$. Here $|S|$ is the number of nodes in a tree. A mathematical proof is given in [4, 5] (refer also to a later study [1], which has used [4, 5]). Refresh that a phylogenetic net of genes or species is an acyclic directed graph with one vertex (the “root”) that can be connected by a path with any other node, and terminal vertices (the “leaves”); the leaves are labelled with species names or species-gene name pairs. An important special case is a binary net, where for each node, except for the root, one of the following is true: the node possesses only one incoming edge and no outgoing edges, or two outgoing and one incoming edge, or two incoming and one outgoing edge. In a species net, introducing time slices ordinarily generates vertices with one incoming and one outgoing edge. The definitions and costs of mapping of a gene net into a species net and mapping of a gene tree into a species tree are identical.

The algorithm realization is available at <http://lab6.iitp.ru/en/embed3gl/>.

4 Mapping a gene tree to a species tree and its cost

Thus, Problems from the sections 2-3 are reduced to the extended event set case. We proposed a possible definition in [4, 5] and formulate the idea below; informal motivation can be found in [4].

Below e runs over all edges of G , and d – over all edges (also referred to as *tubes*) of S_0 . The root is pictured at the top in all. A formal edge (the so-called *root edge*) enters the root from above. By definition,

$e' < e$ if an edge e' is strictly below e , and $e' \leq e$ if $e' < e$ or $e' = e$. One of at most 15 evolutionary events can occur on fixed edge e in tube d ; such events are marked by index i . Two tubes d and d' belong to the same (temporal) slice, by definition, if the lengths of path from the root tube to d and d' are equal, correspondingly. Assume d is a terminal tube, and e is a terminal edge. If gene e belongs to species d , then the mapping $f_{\langle e, d \rangle}(e) = \langle d, fin \rangle$ and the cost $c(\langle e, d, fin \rangle) = 0$. Otherwise, if e belongs to species d' , then the mapping $f_{\langle e, d \rangle}(e, fin) = \langle d, tr(d, d') \rangle$ and the cost $c(\langle e, d, fin \rangle) = 13$, where 13 is the cost of gene transfer without retention (all costs are conditional). Here $tr(d, d')$ indicates that gene e is transferred from tube d to tube d' such that no copy of e remains in d , and 0 is the index of such event. Thus, mapping $f_{\langle e, d \rangle}(e, i)$ is defined by induction. Now we exemplify further induction steps $f_{\langle e, d \rangle}(e') = \langle d', mark \rangle$ where $e' < e$ and d' is in a slice later or equal then the slice d .

Assume d is a tube with single descendent tube d_1 and $pass = \arg_i \min c(e, d, i)$. Then $f_{\langle e, d \rangle}(e) = \langle d, pass \rangle$, $f_{\langle e, d \rangle}(e') = f_{\langle e, d_1 \rangle}(e')$, and $c(\langle e, d, pass \rangle) = \min_i c(\langle e, d_1, i \rangle)$. Here, the “pass” indicates the survival of gene e down to the next tube.

Assume d is a tube with two descendent tubes d_1 and d_2 and $passl = \arg_i \min c(e, d, i)$. Then $f_{\langle e, d \rangle}(e) = \langle d, passl \rangle$, $f_{\langle e, d \rangle}(e') = f_{\langle e, d_1 \rangle}(e')$, and $c(\langle e, d, passl \rangle) = \min_i c(\langle e, d_1, i \rangle) + 2$, where 2 is the cost of loss. Here, the “passl” indicates the survival of gene e into tube d_1 and loss of its copy in tube d_2 . Symmetric cases are not discussed everywhere.

Denote e_1 and e_2 two descendent edges of non-terminal edge e and $passlr = \arg_i \min c(e, d, i)$. Then $f_{\langle e, d \rangle}(e) = \langle d, forklr \rangle$, $f_{\langle e, d \rangle}(e') = f_{\langle e_1, d_1 \rangle}(e')$ if $e' \leq e_1$, and $f_{\langle e, d \rangle}(e') = f_{\langle e_2, d_2 \rangle}(e')$ if $e' \leq e_2$; $c(\langle e, d, passlr \rangle) = \min_i c(\langle e_1, d_1, i \rangle) + \min_i c(\langle e_2, d_2, i \rangle)$. Here, the “forklr” indicates divergence of gene e into e_1 in tube d_1 and e_2 in tube d_2 .

For non-terminal edge e and $dupl = \arg_i \min c(e, d, i)$, $f_{\langle e, d \rangle}(e) = \langle d, dupl \rangle$, $f_{\langle e, d \rangle}(e') = g_{\langle e_1, d_1 \rangle}(e')$ if $e' \leq e_1$, and $f_{\langle e, d \rangle}(e') = f_{\langle e_2, d_2 \rangle}(e')$ if $e' \leq e_2$; $c(\langle e, d, dupl \rangle) = \min_i c(\langle e_1, d_1, i \rangle) + \min_i c(\langle e_2, d_2, i \rangle) + 3$, where 3 is the cost of duplication. Here, the “dupl” indicates the duplication of gene e within tube d into e_1 and e_2 .

For non-terminal edge e assume $\exists d' [tr1(d, d') = \arg_i \min c(e, d, i)]$, where d' be a tube in the same slice with d that differs from d and minimizes the value $\min_i c(\langle e_1, d', i \rangle)$ over d' . Then $f_{\langle e, d \rangle}(e) = \langle d, tr1(d, d') \rangle$, $f_{\langle e, d \rangle}(e') = f_{\langle e_1, d' \rangle}(e')$ if $e' \leq e_1$, and $f_{\langle e, d \rangle}(e') = f_{\langle e_2, d \rangle}(e')$ if $e' \leq e_2$; $c(\langle e, d, tr1(d, d') \rangle) = \min_i c(\langle e_1, d', i \rangle) + \min_i c(\langle e_2, d, i \rangle) + 11$, where 11 is the cost of transfer with retention. Here, the “tr1(d, d’)” indicates that gene copy e_1 is transferred from tube d into tube d' , and gene copy e_2 remains in d .

5 Gene tree reconstruction

The authors propose the problem restatement and a polynomial algorithm that correctly infers the gene tree

in special cases. In simulations of a general case the algorithm was shown to be very fast and in about 75% cases reconstruct a tree very close to that produced by PhyloBayes v.3.3 in much longer time. The tree is sought for among trees consisting of clades from a prebuilt set P that possesses the following property: each set in P can be split into two subsets also from P , and so on until singlet sets are obtained that correspond to alignment rows (ref. to further as rows). The solution found by the algorithm depends heavily on the “correctness” of set P .

In the algorithm, selected “non-informative” columns of the initial alignment X are omitted as in [7], dynamic programming is then applied to the refined alignment for a maximum likelihood inference of one tree that contains all clades from P per alignment column. The obtained set of trees is reconciled using a supertree building algorithm described in section 2 and in [2, 3]. Set P is constructed with induction. All singlet sets (i.e. consisting of a single row) are included in P ; at each induction step, if sets X_1 and X_2 from P do not intersect, and for their union X the conditions $S(X) \geq l(|X|)$, $m^2(X) + \beta > M^2(X)$ are true then X is included in P . Define $I(X)$ the sum of self-similarities of each row from X divided by the total length of all rows from X . Define $S(X)$ the sum of similarities of all row pairs from X divided by the product of total pairs length and $I(X)$. Define $m(X)$ the minimal value of S over all pairs from X ; $M(X)$ its maximal value over all pairs that contain one row from X and another – from its complement. Let $l(m)$ be a linear function, where m decreases from $|X|$ to 2, $l(|X|) = S(X)$, and the derivative equals α , the algorithm parameters are α and β .

The algorithm realization is available at <http://lab6.iitp.ru/en/gene3gl/>.

References

- [1] J.-P. Doyon et al. An Efficient Algorithm for Gene/Species Trees Parsimonious Reconciliation with Losses, Duplications and Transfers. *Lecture Notes in Bioinformatics*, 2010, **6398**, 93-108.
- [2] K. Gorbunov, V. Lyubetsky. The tree nearest on average to a given set of trees. *Problems of Information Transmission*, 2011, **47**(3), 274-288.
- [3] K. Gorbunov and V. Lyubetsky. Fast Algorithm to Reconstruct a Species Supertree from a Set of Protein Trees. *Molecular Biology (Mosk)*, 2012, **46**(1), 161-167.
- [4] K. Gorbunov and V. Lyubetsky. Reconstructing the evolution of genes along the species tree. *Molecular Biology (Mosk)*, 2009, **43**(5), 881-893.
- [5] K. Gorbunov and V. Lyubetsky. An algorithm of reconciliation. *Information Processes*, 2010, **10**(2), 140-144 (in Russian).
- [6] R. Guigo et al. Reconstruction of ancient molecular phylogeny. *Molecular Phylogenetics and Evolution*, 1996, **6**, 189-213.
- [7] V. Lyubetsky et al. Removing noise in a multiple protein alignment. *Information processes*, 2005, **5**(5), 380-391 (in Russian).

Correlation between some properties of the primary structure of promoter regions and gene coding sequences of yeasts

Matushkin Yu.G.⁽¹⁾, Orlov Yu.L.⁽¹⁾, Likhoshvai V.A.^(1,2), Levitsky V.G.^(1,2)

Institute of Cytology and Genetics SB RAS⁽¹⁾, Novosibirsk State University⁽²⁾
mat@bionet.nsc.ru

Elongation Efficiency Index (EEI) was suggested earlier in [3] to estimate gene expression efficiency by nucleotide context of coding sequence in unicellular organisms. We have analyzed association between EEI and nucleosome formation potential (NFP) in 5' regulatory regions upstream translation initiation site (TIS) from two yeast species. Theoretical estimations of NFP based on DNA sequence were obtained by Recon method [2]. Experimental estimation of nucleosome occupancy was obtained by high-throughput sequencing data of nucleosomal DNA in *S.cerevisiae* [1]. For the sample of all genes correlation coefficient was calculated between two vectors: vector of NFP values for fixed position relative to TIS and vector of EEI values. Profiles of correlation coefficients of NFP and EEI were counted in (-600; +600) regions relative to TIS for gene sequences extracted from GenBank. We found regions of strong negative dependence between NFP and EEI for all genes as well as for 10% highly expressed genes in *S.pombe* (10% of EEI-highest genes). At the same time we found positive dependence between NFP and EEI for all genes and for low expressed genes in *S.cerevisiae* (10% of EEI-lowest genes). The association between NFP and EEI could be explained by evolutionary selection of context characteristics of nucleotide sequences for gene expression optimization.

1 Materials and Methods

S.cerevisiae and *S.pombe* open reading frames of length exceeding 30 nt were extracted from GenBank database (<http://www.ncbi.nlm.nih.gov/>) to calculate EEI. Two final whole-genome sequence sets consisting of respective 5'-flanks, 5'-untranslated regions and coding sequences of total length 1200 nt (-600 to +600 relative to TIS). Extraction was done based on primary gene annotation removing introns. Sequences with undefined nucleotides (N) were ignored. We have calculated elongation efficiency index (EEI) for the extracted sequences disregarding gene functions. 6301 genes of *S.cerevisiae* and 4681 of *S.pombe* were used in analysis. Additionally, top 10% and bottom 10% genes ranked by EEI were selected from these lists.

Experimental data for genome-wide sequencing of *S.cerevisiae* nucleosomal DNA at different growth conditions were taken from [1]. We used the data only for 5418 genes to keep consistency between genome releases and gene annotations used earlier [3]. Experimental nucleosome density profile was

constructed from short sequence tags mapped to the reference yeast genome (GEO: GSM351492). Unique sequence reads (Illumina Genome Analyzer II) were mapped to the yeast genome and extended to the length 146 bp. We used raw mapping data for mapping nucleosomal DNA *in vivo* at optimal yeast growth conditions (YPD) to construct nucleosome density profile. Each position in genome has corresponding nucleosome density value as integer number of overlapping nucleosomal sequence reads.

2 Results and Discussion

2.1 Correlation between NFP & EEI for *S.cerevisiae*

As we have shown earlier [3], EEI is adequate for *S.cerevisiae* genes. Correlation coefficient between IEE and experimentally defined expression efficiency is 0.8 at $p < 0.00001$.

Correlation coefficients of NFP and EEI were calculated for *S.cerevisiae* and *S.pombe* between vectors: NFP vectors for each given position of sequences phased relative to TIS in the (-600; +600) nt interval and EEI values vector, calculated in advance for gene coding sequences. Profiles of correlation coefficients were obtained for (-525; +525) regions relative to TIS.

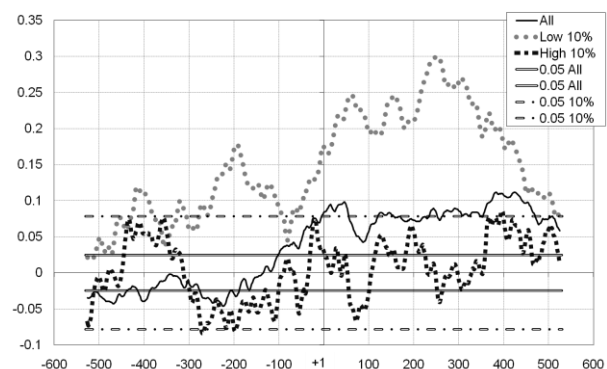


Fig.1. Averaged correlation coefficients (axis Y) between EEI of coding sequences and NFP for 5' regulatory regions upstream of the translation initiation site in *S.cerevisiae*. All – all sequences; Low 10 % and High 10 % – bottom 10% and top 10% of all sequences by EEI rank, respectively Position +1 corresponds to the translation initiation site. Horizontal lines mark significance levels ($p < 0.05$) by Fisher criteria. Each point of profile presents the correlation coefficient between EEI and average NFP for ten consecutive positions.

2.2 Correlation between NFP & EEI for *S.pombe*

As one can see from Figure 2, the significant negative correlation between EEI and NFP of upstream region for sequences with highest EEI is in interval (-400;-200) relative to TIS. I.e., the higher elongation efficiency of coding sequences, the weaker nucleosome packing in 5'-region. There is also significant positive correlation for EEI-lowest sequences and even greater and significant negative correlation for EEI-highest sequences in the beginning of coding sequence (+1; +400). Apparently, the selection went not only to facilitate transcription initiation, but for facilitation of both elongation and transcription for EEI-highest sequences. Though, possible, it is a consequence of codon selection in coding sequence.

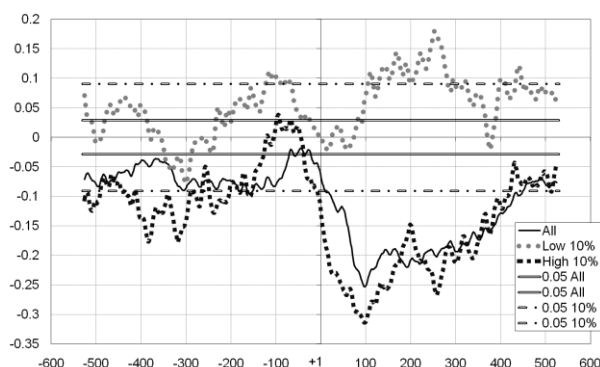


Fig.2. Correlation coefficients (axis Y) between EEI of coding sequences and NFP of 5' regulatory regions upstream of the translation initiation site for *S.pombe*.

2.3 Analysis of nucleotide context of *S.cerevisiae* and *S.pombe*

To explain the relation detected for EEI and NFP at the level of context features we have investigated nucleotide sequences upstream translation starts of *S.cerevisiae* and *S.pombe* genes in details. We found that proximal promoter gene region (-300;+1) is enriched by poly(A) tracts (length from 5 to 9).

The analysis has shown that for *S.cerevisiae* genes in the region (-200;-100) fraction of genes with poly(A) tracts is up to 40% higher than expected fraction based on dinucleotide frequencies. For the both remaining regions of *S.cerevisiae* genes the excess over the expected fraction is very limited. At the same time for *S.pombe* genes such effect of poly(A) tracts enrichment is absent. One would suggest that this enrichment is related to location of (-200;-100) region in critical position for gene transcription initiation.

2.5 Conclusions

We have shown inter-relation between nucleosome localization in promoter regions and elongation efficiency index. Therefore, we have confirmed the suggested hypothesis. Elongation efficiency negatively correlates with nucleosome formation potential. Therewith the selection may lead both to NFP decrease for high-expressing sequences (for

S.pombe), and to NFP increase for low-expressing ones (for *S.cerevisiae*). Apparently this is the cause of distinct distributions of poly(A) tracts in upstream regions for two yeast species.

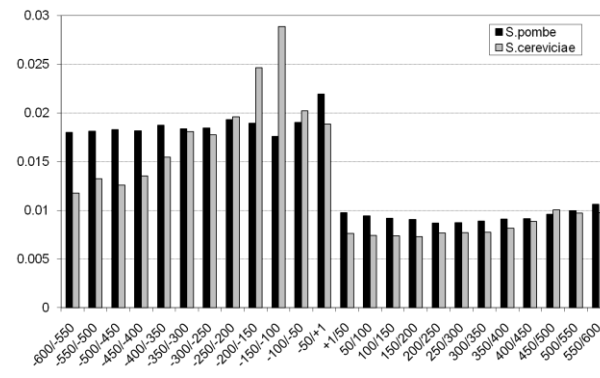


Fig.3. The distribution of poly(A) tracts is different for two yeast species. We suggest that this difference leads to distinctions in correlations between NFP and EEI for *S.cerevisiae* and *S.pombe*. Sharp distinction in correlation trend from TIS to 3'-end we explain by the classes of elongation optimization: *S.cerevisiae* belongs to the first class of elongation efficiency index optimization – the selection went only by codon frequencies, but *S.pombe* belongs to the fourth class of index optimization – selection went by codon frequencies as well as by length and number of local complementary structures [3]. Possibly, this explains the difference in nucleosome positioning mechanisms of these related.

2.4 Correlation between EEI and experimental data on nucleosome density

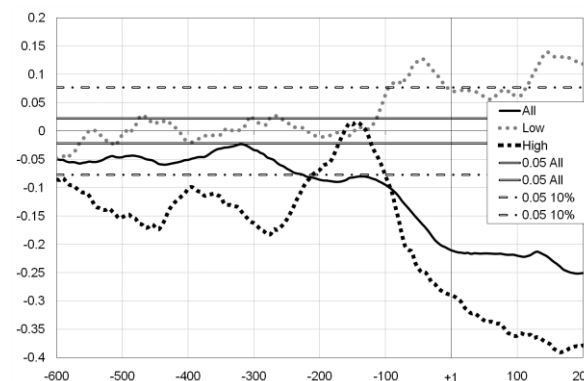


Fig. 4. Correlation coefficients (axis Y) between EEI of coding sequences and experimental data on nucleosome density obtained by direct sequencing of nucleosomal DNA *S.cerevisiae* [1].

References

- [1] Kaplan N., Moore I.K., Fondufe-Mittendorf Y., Gossett A.J., Tillo D., Field Y., LeProust E.M., Hughes T.R., Lieb J.D., Widom J. & Segal E. (2009) The DNA-Encoded Nucleosome Organization of a Eukaryotic Genome *Nature* 458(7236), 362-366.
- [2] Levitsky V.G. (2004) RECON: a program for prediction of nucleosome formation potential. *Nucl Acids Res* 32, 346-349 (2004).
- [3] Vladimirov N.V., Likhoshvai V.A. & Yu.G. Matushkin. (2007) Correlation of codon biases and potential secondary structures with mRNA translation efficiency in unicellular organisms. *Mol Biol (Mosk)* 41(5), 926-933.

Расчет генетического кода

Н.Н. Козлов

Институт прикладной математики им. М.В. Келдыша РАН, Москва

Математический анализ генетических расшифровок проводился автором в течение ряда лет. Причем одним из главных объектов исследования были перекрывающиеся гены, открытые в 1976 году. Это случаи, когда один и тот же участок ДНК кодирует две или более белковых последовательностей. Многочисленные случаи выявленных генетических перекрытий, позволили поставить ряд математических задач, которые были успешно решены. Подробное изложение этих задач приводится в авторской монографии [1]. Такие задачи дали возможность достаточно глубоко проникнуть в структуру генетического кода и ее взаимосвязи с перекрывающимися генами. В итоге была поставлена новая задача: расчет генетического кода на основе аминокислотных последовательностей, записывающих перекрывающиеся гены. Об одном подходе к решению этой задачи говорится в данной работе.

Как известно, генетический код был окончательно установлен экспериментально в 1966 году. Сразу после этого началась расшифровка генов с использованием генетического кода. Главный объект исследований автора также гены, но перекрывающиеся, открытые в 1976 году и обнаруживаемые в настоящее время в значительных количествах в больших геномах. Перекрывающиеся гены это случаи, когда один и тот же участок ДНК кодирует более одной (вплоть до 6-и) белковой последовательности. В ходе исследований были решены ряд задач для области ДНК с перекрытиями генов: степень жесткости связи белковых последовательностей и этой области, роль нерегулярностей кода, востребованность каждого из смысловых кодонов, способ построения любого перекрытия пар генов из одной цепи ДНК, уменьшение размера генома, за счет отклонений кода от его стандартной структуры, математический анализ всех 5-и способов парных перекрытий генов для стандартного, а также всех 14-и девиантных кодов, введение интегральных характеристик кода как для перекрывающихся, так и для не перекрывающихся генов и установление их взаимосвязи, установление математической

аналогии генетических перекрытий и стеблей вторичной структуры матричных РНК, теория элементарных генетических перекрытий, а также роль неоднозначностей во множествах таких перекрытий. Подробное изложение этих задач в их взаимосвязи представлено в авторской монографии [1]. Решения некоторых из этих задач, важных для дальнейшего, представлены в докладе. В ходе решения указанных задач возникла мысль попытаться расчетным путем установить структуру генетического кода, или найти такой тип математических преобразований, на основе которых можно было бы вычислить генетический код. Об одном подходе к решению этой задачи говорится в данной работе.

Ранее нами было введено понятие элементарного перекрытия применительно к перекрытиям пар генов. Обобщим это понятие для трех генов, принадлежащих одной и той же цепи ДНК [2]. Под термином элементарное перекрытие понимаем, перекрытие для кодонов одиночных аминокислот по максимальному числу позиций. На рис.1А для аминокислоты Ama , кодируемой триплетом $p_1p_2p_3$ указаны альтернативные аминокислоты Ama_1 и Ama_2 , кодировки которых $p_0p_1p_2$ и $p_2p_3p_4$ соответственно, образованы сдвигами на -1 и $+1$ нуклеотид в той же цепи ДНК ($\rightarrow$). Предполагается при этом, что все значения $p_0 - p_4$ принадлежат каноническому набору из четырех нуклеотидов. На основе рис.1А можно построить три вида сочетаний аминокислот, представленных на рис.1Б и обозначенных соответственно u_1 , u_2 , u_3 : одно u_1 для перекрытия по одной позиции, два u_2 для перекрытия по двум позициям, и одно u_3 для перекрытия 3-х аминокислот. Следует отметить, что в генетических экспериментах более 10 лет назад нами были обнаружены все допустимые генетическим кодом элементы u_1 , u_2 , u_3 , которые будем рассматривать соответственно как элементы множеств U_1 , U_2 , U_3 . Число элементов U_2 равно 160, а элементов $U_3 - 307$; полный их перечень представлен в [3-4]. Что касается множества U_1 , то в ходе решения были использованы лишь небольшая часть элементов этого множества и его обобщенные характеристики, которые приводятся.

Постановка задачи. Пусть имеем набор из 4 букв: $N: a, b, c, d$, а также триплеты – любые тройки из этих букв, всего их 64. При этом каждая из 20 канонических аминокислот может кодироваться

произвольным сочетанием таких триплетов. Задача состоит в поиске всех генетических кодов, отвечающих всем элементам, обозначенных выше трех множеств U_1 , U_2 , U_3 , соответствующих генетическим экспериментам.

В докладе предлагается впервые представить метод расчета генетического кода, который основывается на исследовании свойств, обнаруженных для специального представления главного из множеств - сжатого множества U_3 . Процедуры вычисления кода по этой методике были представлены ранее [2-4].

(А)

$$\begin{array}{c} Ama_2 \\ Ama \\ Ama_1 \\ p_0p_1p_2p_3p_4 \end{array}$$

(Б)

$$\begin{array}{ccccc} & & Ama & & \\ Ama_2 & & Ama_1 & & Ama_2 \\ Ama_1 & & & & Ama \\ & Ama_2 & & Ama_1 & \\ & Ama & & & \\ u_1 & u_2 & & u_3 & \end{array}$$

Рис.1

А. Для аминокислоты Ama , кодируемой триплетом $p_1p_2p_3$, имеют место 2 альтернативных аминокислоты Ama_1 и Ama_2 . Кодировки их

образованы сдвигами -1 , $+1$ в той же цепи ДНК ($\rightarrow$). Кодон $p_0p_1p_2$ для Ama_1 перекрывается с кодоном $p_1p_2p_3$ для Ama – перекрытие содержит два нуклеотида p_1p_2 ; кодон $p_2p_3p_4$ для Ama_2 перекрывается с кодоном $p_1p_2p_3$ для Ama – перекрытие содержит два нуклеотида p_2p_3 . В итоге тройное перекрытие содержит всего одну общую позицию p_2 . Каждая из приведенных трех аминокислот принадлежит тем наборам A^0 из (1), кодировки которых допускают такие перекрытия.

Б. Элементы множеств сочетаний аминокислот, образованные на основе элементарного перекрытия из рис.1А. Слева – один элемент множества U_1 , в центре – два элемента множества U_2 и справа – один элемент множества U_3 .

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (коды проектов 10-01-00145, 11-01-00110).

Список литературы

- [1] Козлов Н.Н. Математический анализ генетического кода. Монография. М., БИНОМ, 2010, 223 с.
- [2] Козлов Н.Н. Расчет генетического кода. ДАН 2010. Т.433. №.2. С.158-162.
- [3] Козлов Н.Н. Расчет генетического кода. Препринт ИПМ им. М.В.Келдыша РАН 2010, N 26, 23 с.
- [4] Козлов Н.Н. О расчете генетического кода. Математическое моделирование, т.23, N 6, с.3-17, 2011.

Применение вычислений на графических процессорах для оценки эффективности выявления индивидуумов, предрасположенных к колоректальному раку

Леоненко В.Н.⁽¹⁾, Перцев Н.В.⁽¹⁾

Омский филиал Федерального государственного бюджетного учреждения науки
Института математики им. С. Л. Соболева СО РАН⁽¹⁾
homlab@ya.ru

The aim of this work was to create a model for the efficiency estimation of detecting the individuals predisposed to colorectal cancer and to implement the simulation program for it by means of NVIDIA CUDA technology.

1 Введение

Колоректальным раком (КРР) называется злокачественное образование, вызываемое перерождением полипов на внутренних стенках толстой кишки. Ежегодная заболеваемость КРР достигает 1 миллиона случаев, а ежегодная смертность превышает 500 тысяч случаев. В настоящее время КРР занимает второе место по смертности среди злокачественных образований [4]. Поскольку заболевание не является наследственным и на его ранних стадиях симптомы отсутствуют, большинство больных КРР не обращаются своевременно за медицинской помощью.

Для снижения смертности от КРР практикуется массовое профилактическое обследование индивидуумов группы риска (мужчины и женщины старше пятидесяти лет), направленное на выявление и удаление полипов. Наиболее чувствительным методом обследования является фиброколоноскопия (ФКС). Конкуренцию ей составляет метод виртуальной колоноскопии (ВКС), который при меньшей чувствительности является более щадящим для пациентов. В связи с появлением метода ВКС в медицинских учреждениях Российской Федерации возникла задача сравнения эффективности массовых обследований населения регионов России с применением ФКС и ВКС.

Авторами настоящей работы были построены стохастические модели выявления индивидуумов, предрасположенных к КРР, для популяции с фиксированной численностью (в соавторстве с Е. Ю. Хомутовой) [3] и для популяции с переменной численностью [2]. В обеих моделях используется многопараметрическое описание индивидуумов, позволяющее отразить особенности развития заболевания и его выявление с помощью ФКС или ВКС с высокой степенью детализации. Вычислительные алгоритмы опираются на методы Монте-Карло генерации дискретных и непрерывных

случайных величин с использованием датчиков псевдослучайных чисел.

Из-за больших объемов обрабатываемых данных последовательные версии моделирующих программ для индивидуум-ориентированных моделей отличаются невысоким быстродействием. В связи с этим затрудняется проведение вычислительных экспериментов с популяциями большой численности, а также настройка моделей на реальные данные. Одним из способов решения указанной проблемы является создание моделирующих программ, использующих технологии параллельных вычислений.

Целью данной работы является разработка вычислительного алгоритма и моделирующей программы для модели выявления индивидуумов в фиксированной популяции на основе технологии CUDA, позволяющей проводить параллельные вычисления на графических процессорах NVIDIA.

2 Описание модели

Рассматривается группа из N индивидуумов возраста 40 лет (предполагается, что начиная с этого возраста имеется ненулевая вероятность возникновения полипа). Продолжительность времени жизни каждого индивидуума описывается случайной величиной с распределением Мейкхем-Гомперца. Время модели дискретное, с шагом в один год.

У каждого индивидуума в течение года может возникнуть полип, вероятность этого события зависит от возраста индивидуума. Считается, что индивидуум в каждый момент времени может иметь не более одного полипа. Начальный размер возникшего полипа r_0 является случайной величиной. Если предположить, что полип у индивидуума появился в момент времени $t = 0$, то его последующий рост описывается следующим законом:

$$r(t+1) = r(t)[1 + \beta(r(t))], \quad t = 0, 1, 2, \dots, \quad r(0) = r_0,$$

где $r(t)$ – размер полипа в мм на момент времени t , величина $\beta(r(t)) \geq 0$ отражает темп прироста полипов за один год в зависимости от текущего размера.

Моделирование развития полипа происходит вплоть до момента естественной гибели

индивидуума или его перехода в состояние предрака (невывявленный полип более 10 мм).

Индивидуум может попасть на обследование при самостоятельном обращении к врачу по причинам, напрямую не связанным с развитием полипа (например, при наличии патологии органов брюшной полости), либо принять участие в массовом обследовании населения (скрининге). В последнем случае индивидуум приглашается на обследование через равные промежутки времени, например, каждые десять лет, начиная с возраста пятьдесят лет. Вероятность согласия индивидуума на прохождение обследования зависит от используемого метода (ФКС или ВКС). При проведении обследования полип может быть обнаружен с вероятностью, зависящей от размера полипа и метода обследования. Обнаруженный полип размером более 6 мм удаляется.

Для сравнения эффективности предотвращения развития КРР с помощью ФКС и ВКС были выбраны следующие показатели:

- x - количество индивидуумов гипотетической популяции, попавших в группу риска по развитию КРР (полип достиг размера 10 мм и более);
- y - количество индивидуумов группы высокого риска по развитию КРР, выявленных при первом обращении к врачу.

В ходе вычислительного эксперимента изучались законы распределения и числовые характеристики показателей x и y при варьировании параметров, отвечающих за продолжительность жизни индивидуумов, и разных методах обследования (ВКС и ФКС). Статистические вычисления проводились по выборке из ста значений показателей (x, y) , полученных при реализации алгоритма с разными стартовыми значениями датчика случайных чисел.

3 Вычисления на графических процессорах

Структура модели выявления индивидуумов, предрасположенных к развитию КРР, в фиксированной популяции является удобной для реализации моделирующей программы с использованием технологии CUDA. Программа на CUDA должна задействовать как центральный (CPU), так и графический (GPU) процессор, при этом работа GPU заключается в одновременном применении заданного набора функций («ядра») к большому массиву входных данных [1]. В нашем случае ядром является набор функций, моделирующих «судьбу» одного фиксированного индивидуума, включая появление, развитие и выявление полипа; массив входных данных формируется из стартовых значений датчика случайных чисел. Выбор ядра в указанном виде обладает следующими преимуществами:

– обмен данными между CPU и GPU ограничивается пересылкой данных на GPU в начале работы и получением результатов в конце работы;

– при характерных численностях моделируемых популяций $10^5 - 10^7$ индивидуумов объём массива данных, на котором происходит выполнение ядра, достаточен для эффективного использования ресурсов GPU.

В табл. 3.1 приведены характерные времена выполнения одной реализации алгоритма для гипотетической популяции, состоящей из 10^7 индивидуумов, при использовании скрининга на основе ФКС или ВКС для различных версий программ. В версии программы для CPU использовался мультипликативный датчик случайных чисел, в версии для GPU – датчик на основе линейного конгруэнтного метода, выполняющийся на графическом устройстве. Расчёты проводились на ПЭВМ с CPU Intel Core2 Quad 2,66 ГГц и GPU NVIDIA GeForce 8600 GT.

Таблица 3.1. Время выполнения расчётов, с.

Программа/Метод	ФКС	ВКС
Программа для CPU	133,1	117,4
Программа для GPU	4,9	5,0

Проведённые эксперименты показали, что версия программы для GPU позволяет получить ускорение вычислений до 20–25 раз в зависимости от выбора параметров и датчика случайных чисел.

Наличие моделирующих программ для CPU и GPU позволяет совместно использовать ресурсы центрального и графического процессоров, что существенно ускоряет проведение вычислительных экспериментов.

Работа поддержана Сибирским отделением РАН (интеграционный проект № 47, 2012–2014 гг.).

Список литературы

- [1] А. В. Боресков, А. А. Харламов. Основы работы с технологией CUDA. М.: ДМК Пресс, 2010, 232 с.
- [2] В.Н. Леоненко, Н.В. Перцев. Анализ эффективности программ выявления индивидуумов, предрасположенных к колоректальному раку, на основе имитационного моделирования // Управление большими системами. 2011. N.35. С.207–236.
- [3] Н.В. Перцев, Е.Ю. Хомутова, В.Н. Леоненко. Применение математического моделирования для оценки эффективности выявления индивидуумов, предрасположенных к колоректальному раку // Медицинская визуализация. 2011. N.2. С.104–108.
- [4] Скрининг колоректального рака. Практическое руководство Всемирного гастроэнтерологического общества (ВГО) и Международного союза по профилактике рака пищеварительной системы. URL: http://www.worldgastroenterology.org/assets/downloads/ru/pdf/guidelines/06_colorectal_cancer_screening_ru.pdf

Сравнительный анализ аминокислотных последовательностей отдельных доменов белков Нос бактериофагов подсемейства *Teequatrovirinae*

Зимин А.А.⁽¹⁾, Микулинская Г.В.⁽²⁾, Нигматуллина Л.Ф.⁽¹⁾, Назипова Н.Н.⁽³⁾

ИБФМ РАН⁽¹⁾, ФИБХ РАН⁽²⁾, ИМПБ РАН⁽³⁾
nnn@impb.ru

Authors present the results of the comparative study of immunoglobulin-like domains in proteins of bacteriophages related to T4-coliphage (bacteriophages of subfamily *Teequatrovirinae*). A hypothetical scheme of the evolutionary origin of repetitive amino acid sequences in Hoc proteins is presented.

Белки, содержащие иммуноглобулинподобные домены, относят к суперсемейству иммуноглобулинов (IgSF). Иммуноглобулинподобные домены найдены у организмов всех типов животного мира, у бактерий и вирусов. У разных организмов они выполняют различные функции, начиная от клеточного узнавания, связывания, агрегации до проявления эластичности. В данной работе представлен результат сравнительного исследования иммуноглобулинподобных доменов в белках бактериофагов, родственных бактериофагу T4 (т. назв. бактериофагов подсемейства *Teequatrovirinae*).

Бактериофаг T4 - один из самых изученных вирусов, бактериофаг, поражающий бактерию *E. coli*. Имеет геномную ДНК порядка 169—170 тысяч пар нуклеотидов, упакованную в икосаэдрическую головку, окруженную белковой оболочкой — капсидом. Капсид построен из множества одинаковых, геометрически правильных, белковых структур — капсомеров. Капсомеры точно подобраны и подогнаны друг к другу, образуя икосаэдр. К центру капсомера, присоединяется белок Нос, который содержит четыре tandemных иммуноглобулинподобных домена [1].

С помощью криоэлектронной микроскопии было показано, что молекула белка Нос имеет форму гантели. Вероятно, два шара гантели представляют собой два функциональных модуля. Один из них, консервативный, взаимодействует с поверхностью капсида, а удаленный от капсида вариативный модуль взаимодействует с поверхностью бактериальной клетки.

Семейство бактериофагов, родственных бактериофагу T4, является одной из самых

многочисленных групп бактериальных вирусов. В него входят бактериофаги, инфицирующие различные бактерии. Просеквенированный 31 геном этого семейства, составляют небольшую часть всех, имеющихся в природе. Подсемейство T4-подобных фагов *Teequatrovirinae* было поделено по сходству аминокислотных последовательностей [6] на 2 группы (фаги T4-типа и KVP40-подобные вирусы). Мы исследовали первую из этих групп, которая содержит 4 подгруппы (T4-type, 44RR2.8-type, RB43-type, RB49-type phages).

Белки Нос у бактериофагов подсемейства *Teequatrovirinae* гетерогенны по длине и могут содержать разное число доменов, сходных по аминокислотной последовательности и по структуре [2]. Природа гетерогенности белков Нос, а также происхождение белков может быть исследовано путем сравнительного анализа аминокислотных последовательностей отдельных доменов. Такое исследование может пролить свет на пути эволюции как белков Нос, так и других белков, содержащих повторяющиеся иммуноглобулинподобные домены.

Сравнение аминокислотных последовательностей отдельных доменов белков Нос бактериофагов T4 и RB49 с базами данных белковых последовательностей проводилось программой PSI-BLAST [3]. Мы осуществили сравнительный филогенетический анализ отдельных доменов 31 белка Нос бактериофагов подсемейства *Teequatrovirinae*, предварительно выполнив выравнивание последовательностей с помощью программы ClustalX [4]. Было построено филогенетическое дерево с помощью пакета программ Mega4 [5]. На основе полученных данных сформулированы гипотетические схемы эволюционного происхождения повторяющихся аминокислотных последовательностей в белках Нос.

Было показано, что число иммуноглобулинподобных доменов в белке Нос разных бактериофагов подсемейства *Teequatrovirinae* варьирует от одного до пяти и определяет антигенные свойства белка. Белки Нос этого семейства бактериофагов можно разделить на 6 групп:

Группа I: однодоменные белки Нос

Группа II: двухдоменные белки Нос

Группа III: трехдоменные белки Нос

Группа IV: четырехдоменные белки Нос

Группа V: четырехдоменные белки Нос с С-концевым удлинением третьего домена

Группа VI: пятидоменные белки Нос

Филогенетическое дерево отдельных доменов в белках - продуктах генов *hos* бактериофагов, родственных T4, образует три основные ветви. Это ветвь С-концевых доменов, ветвь N-концевых доменов и ветвь промежуточных доменов.

Обязательное присутствие С-концевого домена во всех белках Нос указывает на его функциональную и структурную значимость для формирования белка и его прикрепления к капсиду фага. Сходство этого домена с адгезивными белками – селектинами – и наличие в нем консервативной белковой последовательности ESRNG также косвенно свидетельствует о его участии в присоединении к капсиду.

При анализе первых трех доменов белка Нос бактериофага T4 было обнаружено выраженное сходство с белками-компонентами иммунной системы млекопитающих. Это может указывать на то, что кровь является нормальной средой для существования в ней фага. В связи с этим возможна его молекулярная мимикрия под компоненты иммунной системы. Наличие подобных доменов в белках отростков других бактериофагов может говорить о широком распространении такой молекулярной мимикрии среди фагов.

Сравнительный анализ отдельных доменов белка Нос бактериофага RB49 показал, что он, вместе с белками Нос бактериофагов Phi1 и JSE, образует обособленную группу.

Для этой группы характерно отсутствие сходства второго и третьего доменов белка Нос с иммуноглобулинами и иммуноглобулинподобными

белками, а также с соответствующими доменами белков Нос фага T4 и ряда других родственных бактериофагов. Это, вероятно, говорит о существенном отличии в антигенных и других свойствах белка Нос фага RB49 (а также Phi1 и JSE). Сравнительный анализ доменной организации Нос позволяет выделить три возможных пути эволюции его генов: а) внутригенная дупликация промежуточных доменов, б) перемешивание доменов при рекомбинации между генами *hos* различных родственных фагов и в) дупликация гена двухдоменного варианта белка Нос.

Работа была частично поддержана грантом РФФИ №12-07-00530.

Список литературы

- [1] Bateman A., Eddy S.R., Mesyanzhinov V.V. A member of the immunoglobulin superfamily in bacteriophage T4. *Virus Genes*. 1997, **14**, 163–165.
- [2] Fokine A., Islam M.Z., Zhang Z., Bowman V.D., Rao V.B., Rossmann M.G. Structure of the three N-terminal immunoglobulin domains of the highly immunogenic outer capsid protein from a T4-like bacteriophage. *J. Virol.*, 2011, **85**(16), 8141-8148.
- [3] Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
- [4] Jeanmougin F., Thompson J.D., Gouy M., Higgins D., Gibson T.J. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.*, 1998, **23**, 403–405.
- [5] Tamura K., Dudley J., Nei M., Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 2004, **24**, 1596–1599.
- [6] Lavigne R., Darius P., Summer E.J. et al. Classification of *Myoviridae* bacteriophages using protein sequence similarity. *BMC Microbiology*, 2009, **9**, Paper No. 224.

Совершенствование автоматизированной технологии похромосомного сканирования геномов для нахождения районов скрытой периодичности

Назипова Н., Теплухина Е., Тюльбашева Г., Чалей М.

Институт математических проблем биологии РАН, Пущино, Россия
nnn@impb.ru

Automatic technology for revealing latent periodicity regions which refines the former method, based on preliminary searching for heterogeneous blocks of the nucleotides, is presented. New approach permits finding the fragments of the repeats, that are not tied to sliding window borders, and therefore obtaining true picture of hidden periodicity in the genomes of the eukaryotes. This technology allows improving data presentation in the HETEROGENE database.

На основе спектрально-статистического подхода [1, 2] к выявлению достоверной и значимой неоднородности в последовательностях ДНК была создана технология поиска сильно размытых tandemных повторов в геноме [3, 4]. Результаты применения этой технологии при анализе геномов модельных эукариотических организмов *A. thaliana*, *C. elegans*, *D. melanogaster* и *S. cerevisiae* были представлены в базе данных HETEROGENE (http://www.jcbi.ru/lp_base/).

Теоретический предельный уровень дивергенции копий паттерна в находимых tandemных повторах составляет 50%. созданный 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 4000 нукл.

Поскольку с алгоритмической точки зрения выявление скрытых периодичностей является очень трудной задачей, если истинная длина периода *a priori* неизвестна, то для поиска сильно размытой периодичности (максимальный уровень дивергенции 50%) был выбран метод поиска высоко значимой (на уровне $\alpha=0,000001$) неоднородности [2]. Для подтверждения значимости размытого tandemного повтора, содержащего небольшое количество копий (случай недостаточного статистического материала) в программном комплексе был использован специальный статистический критерий [1, 2].

При создании базы данных HETEROGENE применялась двухэтапная технология многократного сканирования хромосомной ДНК перекрывающимися окнами различного размера, подобная shotgun-стратегии секвенирования геномов, когда на первом этапе анализируют

короткие перекрывающиеся фрагменты, а на втором производят их компьютерную сборку в более протяженные участки. Анализ содержимого базы данных HETEROGENE, полученного с применением этой технологии, показал, что для размытых tandemных повторов технология этапа сборки более протяженных фрагментов из первоначальных коротких участков, и методика выявления исходных данных требуют совершенствования. Цель настоящей работы состояла в разработке и технологической реализации подхода к наиболее точному определению границ размытых tandemных повторов. Для этого применялась модель эволюции tandemного повтора путём последовательных дупликаций смежных копий текстового паттерна.

При поиске участков скрытой периодичности используются две характеристики – значение *уровня сохранности* паттерна периодичности *pl* ($pl \in [0,1]$) и значение *параметра HL*, показывающего значимое отклонение от однородности (на уровне $\alpha=0,000001$) на тестируемом периоде. Максимальное значение параметра *pl* указывает истинную длину периода при условии, что параметр *HL* гарантирует неоднородность нуклеотидной последовательности на таком периоде.

Метод подсчета уровня сохранности $pl(L)$ состоит в следующем. Анализируемая строка длины *n*, состоящая из букв алфавита $A = \{a_1, \dots, a_K\}$, разбивается на подстроки длины *L*, равной длине тест-периода. 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* для значения тест-периода *L*:

$$pl(L) = \frac{1}{L} \sum_{j=1}^L \max\{\pi_j^i : i \in 1, \dots, K\}$$

и спектрально-статистического параметра *HL*:

$$HL(L) = \frac{R_L}{\chi_{krit}^2(\alpha, (K-1)(L-1))} \sum_{j=1}^L \sum_{i=1}^K \frac{(\pi_j^i - p^i)^2}{p^i(1-p^i)},$$

здесь $\alpha = 0.000001$, p^i - частота встречаемости i -той буквы алфавита A во всей анализируемой строке.

Хромосомная ДНК многократно сканируется скользящим окном, длина окна равна удвоенной длине тест-периода, а шаг смещения окна берется примерно равным 0.1 от длины тест-периода, но не меньше 2 н.п. Количество проходов такого сканирования зависит от диапазона длин тест-периодов. В тестовых расчетах мы использовали диапазон от 2 до 4096.

Если уровень сохранности двух одинаковых по длине смежных участков в окне оказывается не меньше 0.875 (т.е. эти участки отличаются между собой не более, чем в одной четверти позиций), мы считаем эти участки размытыми повторами с кратностью 2. Затем перекрывающиеся участки с одной и той же длиной тест-периода анализируются с целью формирования из них путем слияния одного участка, характеризующего периодичность более высокой кратности на участке хромосомы, покрытой набором участков попарного сходства. Для этого анализа мы строим доверительный интервал для среднего значения pl , те участки, значения уровня сохранности входят в этот доверительный интервал, мы объединяем. Таким образом, крайние участки

большого района периодичности, которые «захватываются» автоматической процедурой поиска и «портят» картину периодичности, удаляются из рассмотрения, Технология слияний и удалений участков повторяется многократно с различными параметрами и условиями. В результате мы получаем качественный материал для второго этапа технологии поиска районов скрытой периодичности, разработанного ранее [4].

Работа выполнена при поддержке РФФИ, проект № 12-07-00530.

Список литературы

- [1] Chaley M.B., Nazipova N.N., Kutyrtkin V.A. *Pattern Recognition and Image Analysis*, 2009, **19**, 358-367.
- [2] Chaley M., Kutyrtkin V. (2008) *Mathematical Biosciences*, **211**(1), 186-204.
- [3] Назипова Н., Теплухина Е., Тюльбашева Г., Чалей М. В: *Труды II Международной конференции «Математическая биология и биоинформатика»*. Пущино, 2008, С. 103-104.
- [4] Назипова Н., Теплухина Е., Тюльбашева Г., Чалей М. В: *Труды III Международной конференции «Математическая биология и биоинформатика»*. Пущино, 2010, С. 141-142.

Анализ связи биохимических и клинических показателей с возникновением нарушения мозгового кровообращения (преходящего и острого) у пациентов с хронической ишемией головного мозга

Кузнецова А.В. ⁽¹⁾, Костомарова И.В. ⁽²⁾, Водолагина Н.Н. ⁽²⁾,
Малыгина Н.А. ⁽²⁾, Сенько О.В. ⁽³⁾

Институт биохимической физики им. Н.М. Эмануэля ⁽¹⁾, Филиал ГБОУ ВПО РНИМУ им.Н.И. Пирогова Минздравсоцразвития "НКЦ геронтологии" ⁽²⁾, Вычислительный центр им. А.А. Дородницына РАН ⁽³⁾
azfor@narod.ru, erri06@rambler.ru, senkoov@mail.ru

Effects of different biochemical, clinical and instrumental indices on temporary cerebrovascular disorder are studied in patients with chronic cerebral ischemia. Regularities existing in groups with or without acute stroke and in anamnesis are analyzed with the help of pattern recognition methods and data mining technique that is based on optimal valid partitioning of explanatory variables space.

1 Введение

Представленное исследование посвящено изучению влияния биохимических, клинических, инструментальных показателей и генетических полиморфизмов на развитие нарушения мозгового кровообращения в группах пациентов с хронической ишемией головного мозга (ХИГМ). [1]. В поставленные цели входило исследование влияния изучаемых показателей на возникновение преходящего нарушения мозгового кровообращения (ПНМК), а также косвенная оценка совместного влияния ПНМК и других показателей на возникновение острого нарушения мозгового кровообращения (ОНМК – ишемический инсульт). В ходе исследований использовали методы интеллектуального анализа данных (Data Mining) [2], оценивали эффективность новых модификаций метода анализа данных, основанных на построении оптимальных достоверных разбиений признакового пространства [3].

2 Материалы и методы

2.1 База данных

База данных содержала информацию о 358 пациентах отделения неврологии "НКЦ геронтологии", включая 273 женщины и 85 мужчин в возрасте от 35 до 102 лет (средний возраст 71,9). Исследования проводились для следующих групп

пациентов: I группа (117 больных) с ХИГМ без ПНМК и ОНМК в анамнезе (ср.возраст 59,4); II группа (53 больных) ПНМК в анамнезе (ср.возраст 62,0); III группа - 24 пациента с перенесенными ПНМК и ОНМК (ср.возраст 64,1) и IV группа – 41 больной с перенесенным ОНМК (ср.возраст 63,4), но без ПНМК.

В базу вошли анамнестические, клинко-лабораторные показатели и результаты инструментальных методов исследования (МРТ, КТ, УЗИ, ЭКГ, УЗИ), а также молекулярно-генетического типирования по генам ангиотензинпревращающего фермента (I/D полиморфизм), липопроteinлипазы ((Hind III полиморфизм) и белка-переносчика эфиров холестерина (Taq I полиморфизм), полученные в лаборатории клинической генетики "НКЦ геронтологии" [1,4]. Исследование липидного метаболизма проводилось в лаборатории липидного обмена Филиала РНИМУ "НКЦ геронтологии" Минздравсоцразвития.

2.2 Методы анализа данных

Для поиска закономерностей, связывающих развитие ПНМК с различными показателями, использовали метод Оптимальных Достоверных Разбиений (ОДР), основанный на построении разбиения признакового пространства. Была исследована новая модификация метода ОДР, предназначенная для оценки статистической достоверности различий между закономерностями, характерными для различных группах пациентов.

Для прогноза развития ПНМК у пациентов с ХИГМ использовали универсальную систему методов интеллектуального анализа данных «РАСПОЗНАВАНИЕ», включающую коллекцию методов распознавания и анализа данных, основанных на различных хорошо зарекомендовавших себя подходах [5].

3 Результаты исследований

На первом этапе были проведены исследования влияния оцениваемых показателей на возникновение ПНМК у пациентов с ХИГМ. С использованием метода ОДР, для этой группы были выявлены закономерности, связывающие развитие ПНМК с биохимическими, клиническими и инструментальными показателями. Исследования показали достоверную и выраженную связь развития ПНМК с определенными уровнями, в крови пациентов, липопротеидов высокой, низкой и очень низкой плотности, общего холестерина, триглицеридов, фибриногена, общего билирубина, аспартат и аланин-аминотрансферазы, а также наличием у пациентов патологии щитовидной железы в виде аутоиммунного тиреоидита и узловых изменений. Методы распознавания использовали для сравнения группы пациентов без ПНМК, в анамнезе, с группой пациентов, перенесших ПНМК. Исследования на выборке из 53 пациентов с перенесенным ПНМК и 117 пациентов без ПНМК показали достаточно высокую диагностическую способность 14 биохимических показателей и трёх показателей характеризующих состояние щитовидной железы, при использовании Линейного дискриминанта Фишера (ЛДФ) - 73%, Линейной машины - 72%, Статистически взвешенные синдромы (CBC) - 73%, Метода опорных векторов - 72%, Диагностическая способность в данном случае понималась как средняя точность распознавания по двум группам.

Далее, было произведено исследование влияния изучаемых показателей на риск развития ОНМК:

- методом оптимально достоверных разбиений (ОДР) выявлены закономерности, связывающие риск развития ОНМК с уровнями значений изучаемых показателей;

- с использованием новой модификации метода ОДР проведено сравнение выявленных закономерностей с закономерностями выявленными для группы без перенесенного ОНМК.

Проведённые исследования выявили существенные различия между двумя наборами закономерностей. Данные различия связаны с одновременным влиянием перенесенного ПНМК и ряда других показателей на риск развития ОНМК.

4 Заключение

Исследования показали статистически достоверную и выраженную связь развития ПНМК с содержанием липопротеидов различной плотности и рядом других биохимических показателей, а также наличием аутоиммунного тиреоидита и узловых изменений щитовидной железы. Информативность полученной группы признаков подтверждается достаточно высокой точностью диагностического алгоритма, созданного на основе методов распознавания.

Список литературы

- [1] 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.
- [2] 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
- [3] 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>
- [4] Кузнецова А.В., Костомарова И.В., Водолагина Н. Н., Малыгина Н. А, Сенько О. В. Изучение влияния клинико-генетических факторов на течение дисциркуляторной энцефалопатии с использованием методов распознавания. Матем. биолог. и биоинформ., 2011, том 6, выпуск 1, стран
- [5] Журавлёв Ю.И., Рязанов В.В., Сенько О.В. РАСПОЗНАВАНИЕ. Математические методы. Программная система. Применение.- Москва: Фазис, 2006.

Использование алгоритма FRiS-GRAD для анализа активности генов при решении 9 медицинских задач

Загоруйко Н.Г.^(1, 2), Борисова И.А.^(1,2), Дюбанов В.В.⁽¹⁾, Кутненко О.А.^(1, 2)

Институт Математики СО РАН⁽¹⁾, Конструкторско-технологический институт
вычислительной техники СО РАН⁽²⁾
biamia@mail.ru

In tasks of modern biology the quantity of attributes often on orders exceeds quantity of objects. For the decision of such tasks the method Data Mining based on use the new measure of similarity between objects in the form of Function of Rival Similarity (FRiS) is offered. On this basis the methods of a quantitative estimation of compactness of patterns, of a construction decision rules and of feature selection are developed. All these techniques are realized in FRiS-GRAD algorithm. High efficiency of the algorithm is illustrated by results of the decision of nine tasks of recognition of diseases on microarray dataset.

1 Введение

В связи с появлением и активным использованием ДНК-микрочипов при решении различных задач в медицине, биоинформатике и молекулярной биологии, усилилась потребность в алгоритмах Data Mining, способных обрабатывать задачи, в которых число анализируемых объектов на порядки превышает число описывающих признаков. Однако большинство из существующих ныне алгоритмов изначально не предназначались для решения подобных сложных, плохо обусловленных задач. Нами был разработан подход, основанный на идее конкурентного сходства[1], который позволяет разрабатывать алгоритмы, лучше приспособленные для этих целей. Одним из таких алгоритмов является алгоритм FRiS-GRAD, который одновременно решает задачу распознавания и задачу выбора системы информативных признаков.

Для проверки эффективности предложенного алгоритма в частности и идеи конкурентного сходства в целом, нами было проведено его масштабное тестирование на девяти медицинских задачах, объектами в которых выступали пациенты с различными заболеваниями, а признаками – активности тысяч генов. Результаты работы нашего алгоритма сравнивались с более ранними результатами, полученными четырьмя наиболее часто используемыми при решении такого рода задач алгоритмами распознавания (метод опорных векторов, межгрупповой анализ, Байесов классификатор и метод k-ближайших соседей) и десятью алгоритмами выбора информативных признаков (significance analysis of microarrays,

analysis of variance (ANOVA), empirical Bayes t-statistic, template matching, maxT, between group analysis, area under the ROC curve, the Welch t-statistic, fold change и rank products).

Результатам проведенного сравнения, а также краткому описанию алгоритма FRiS-GRAD посвящена данная работа.

2 Алгоритм FRiS-GRAD

2.1 Функция конкурентного сходства

Понятие функции конкурентного сходства возникло из идеи, что для оценки близости между объектами произвольной природы необходимо учитывать конкурентную ситуацию. Так, при распознавании двух образов, для оценки конкурентного сходства объекта z с первым образом необходимо учитывать не только расстояние r_1 от z до этого образа, но и расстояние r_2 до ближайшего конкурирующего образа (в случае двух образов это будет расстояние от z до второго образа). В качестве расстояния от объекта до образа может использоваться расстояние от этого объекта до ближайшего объекта этого образа либо до ближайшего типичного представителя этого образа.

Нормированная величина конкурентного сходства при этом вычисляется по следующей формуле:

$$F_{1/2}(z) = (r_2 - r_1) / (r_2 + r_1)$$

При этом, если речь идет об объекте из обучающей выборки, то для него вычисляется величина конкурентного сходства со «своим» образом. Таким образом, r_1 может рассматриваться как расстояние до «своего» образа, а r_2 – как расстояние до ближайшего конкурирующего образа.

Среднее значение конкурентного сходства (FRiS-функции), вычисленное по всей выборке A , является эффективной мерой компактности образов, представленных этой выборкой, и может использоваться в качестве критерия Q выбора информативной системы признаков.

$$Q = \frac{1}{|A|} \sum_{z \in A} F_{1/2}(z)$$

2.2 Алгоритм FRiS-GRAD

Алгоритм FRiS-GRAD состоит из двух блоков. Один блок, ответственный за выбор информативной

системы признаков, основан на использовании алгоритма GRAD[2]. В нем сначала формируются информативные системы признаков малой размерности (не больше трех признаков), а затем они используются в комбинации прямого и обратного поисков, реализуемого алгоритмом AdDel.

Качество $Q(Y)$ каждой из рассматриваемых систем признаков Y оценивается во втором блоке алгоритма FRiS-GRAD. Для этого вычисляется среднее значение FRiS-функции в пространстве Y этих признаков, а в качестве расстояния от объекта до образа используется расстояние до типичного представителя этого образа, далее называемого столпом. Множество столпов, в свою очередь, выбирается таким образом, чтобы максимизировать величину $Q(Y)$ используя как можно меньше столпов. В результате в столпы попадают объекты, которые располагаются в центрах зон сгущения объектов выборки.

Более подробно алгоритм описан в [1].

2 Сравнение эффективности алгоритмов

Сравнение эффективности алгоритмов распознавания и выбора информативных признаков с алгоритмом FRiS-GRAD проводилось на 9 различных медицинских задачах среди которых были задача распознавания рака простаты, рака прямой кишки, лимфомы, миеломы и лейкемии.

Таблица 1. Штрафные баллы алгоритмов выбора признаков в итоговом сравнении

Алгоритмы выбора признаков	Баллы
Fold change	47
Between group analysis	43
Analysis of variance (ANOVA)	43
Significance analysis of microarrays	42
Rank products	42
Welch t-statistic	39
Template matching	38
Area under the ROC curve	37
maxT	37
Empirical Bayes t-statistic	32
FRiS-GRAD	12

Для оценки качества работы алгоритмов использовалась перекрестная проверка – 50% обучающей выборки использовалась для обучения, а на оставшихся 50% процентах оценивалась надежность распознавания и индекс RCI. Все результаты, кроме относящихся к FRiS-GRAD, брались из статьи [4].

После решения каждой задачи алгоритмы упорядочивались по качеству полученных решений. При этом для оценки качества методов выбора

признаков использовались лучшие результаты, лучшего из четырех тестируемых алгоритмов распознавания. Победитель получал 1 штрафной балл, следующий по качеству – 2 балла и так далее. Затем эти штрафные баллы суммировались по всем задачам. Итоговые баллы по различным алгоритмам выбора информативных признаков приводятся в Таблице 1, а по алгоритмам построения решающего правила – в Таблице 2.

Таблица 2. Штрафные баллы алгоритмов распознавания в итоговом сравнении

Алгоритмы распознавания	Баллы
Between group analysis (BGA)	35
K-nearest neighbours (kNN)	32
Naïve bayes classification (NBC)	25
Support vector machines (SVM)	19
FRiS-GRAD	12

Согласно данному сравнению алгоритм FRiS-GRAD оказался несомненным фаворитом как при выборе информативных признаков, так и при построении решающего правила.

3 Выводы

Проведенное сравнение демонстрирует серьезное преимущество подхода, основанного на использовании FRiS-функции, при решении задач с малым количеством объектов и большим количеством признаков, типичными представителями которых являются задачи анализа данных ДНК-микрочипов.

Работа выполнена при финансовой поддержке РФФИ, проект №11-01-00156.

Список литературы

- [1] N. G. Zagoruiko, I. A. Borisova, V. V. Dyubanov, and O. A. Kutnenko. *Int. J. Pattern Recognition and Image Analysis*, 2008, **18** (1), 1-6.
- [2] N. G. Zagoruiko, O. A. Kutnenko, and A. A. Ptitsin. Algorithm GRAD for Selection of Informative Genetic Feature, *Proceedings of the International Conference on Computational Molecular Biology*, 2005, Moscow.
- [3] Higgins Laboratory Web site, 2006 . <http://www.bioinf.ucd.ie/people/ian/>
Jeffery I., Higgins D., Culhane A. (2006), Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data, *BMC Bioinformatics*, 7:359P.

Технология концептуального моделирования биомедицинских данных

Богатырев М. Ю., Вакурин В. С.

Тульский государственный университет
okkambo@mail.ru

A Text Mining technology based on Formal Concept Analysis and oriented on modeling of biomedical data from texts of articles is presented. It uses conceptual graphs acquired from texts as a source for formal concepts and concept lattices. The technology has been applied in experiments of information retrieving from PubMed.

1 Введение

Концептуальное моделирование применяется в технологиях *извлечения знаний из данных* (KDD – технологиях) для построения семантических моделей. Такие модели относятся к классу *семантических сетей* и предоставляют собой *мультиграфы* [2], содержащие объекты с определенным смысловым содержанием и отношения между ними.

В биоинформатике концептуальное моделирование используется, по крайней мере, в двух направлениях. Первое направление — это создание баз знаний, для организации которых применяют самые масштабные концептуальные модели — *онтологии*. Исследуемая в данной работе система PubMed [5] использует онтологии. Второе направление, описываемое термином Text Mining, — это извлечение знаний непосредственно из текстов, известное также как «*понимание текстов*». Здесь применяются несколько видов концептуальных моделей, среди которых весьма эффективными являются *концептуальные графы* и *решетки понятий* [2].

Данная работа относится ко второму направлению применения KDD – технологий в биоинформатике. Она продолжает исследования, представленные в работе [1], в которой описано применение концептуальных графов для моделирования семантики биомедицинских тестов.

Основное внимание в работе [1] уделено особенностям построения корректных концептуальных графов для биомедицинских тестов. Целью данной работы является развитие технологии концептуального моделирования путем применения в ней моделей в виде решеток понятий.

Это позволяет строить компактные представления понятий, содержащихся в текстовых данных, и выявлять связи, существующие между понятиями.

2 Принцип концептуального моделирования биомедицинских данных

В биоинформатике задача извлечения знаний из текстов известна давно и имеет ряд законченных решений, например [3]. Вместе с тем, главная цель, поставленная в первых работах по Text Mining в области биоинформатики, — заменить дорогостоящие биологические эксперименты, например, по построению геномов, знаниями, полученными на основе анализа уже существующих результатов, - пока не достигнута. В настоящее время становится очевидно, что прогресс в этом направлении обусловлен, главным образом, совершенствованием моделей знаний, применяемых в методах Text Mining, и эффективностью их реализаций в KDD – технологиях.

Принцип концептуального моделирования, лежащий в основе разрабатываемой нами информационной технологии, заключается в следующем.

1. Обработываемые данные представляют собой тексты, имеющую определенную тематику, отраженную в их *контекстах*. Контексты представляют собой объекты, внешние по отношению к текстам — их *метаданные*.
2. Для каждого предложения текста может быть построен его концептуальный граф — двудольный граф, содержащий в качестве узлов *концепты* и *отношения*. Концепты моделируют отдельные слова предложений, а отношения — связи между ними. Технология построения концептуальных графов описана в [4]. Число концептуальных графов совпадает с числом предложений текста и может быть весьма велико. Как и некоторые предложения, концептуальные графы могут не содержать полезной информации. Контекст, определяющий тематику текста, позволяет задать ограничения на число и состав концептуальных графов.
3. Концептуальная модель более высокого уровня — решетка понятий - строится с целью обобщения информации, содержащейся в концептуальных графах. При этом обобщается информация из нескольких текстов, что важно при практической реализации принципа концептуального моделирования.

3 Реализация и применение технологии

Принцип концептуального моделирования является универсальным, не связанным с определенной тематикой текстовых данных. Реализация данного принципа в конкретной технологии уже должна опираться на подобную тематику и в целом решать связанные с ней задачи. В данной работе рассматривается технология решения следующей задачи.

В настоящее время биомедицинские информационные ресурсы активно наращивают свои масштабы и расширяют функциональные возможности. Примером служит широко известная англоязычная система PubMed. Текстовый запрос к данной системе, например, 'Gene expression', приводит к формированию ссылок на 863436 статей (по состоянию баз данных системы на июль 2012 г.), в текстах которых содержится данное словосочетание. Прочитать все найденные статьи невозможно, поэтому необходима дополнительная обработка полученной информации.

Концептуальное моделирование позволяет связать найденные статьи через используемые в них общие понятия. Объектом, реализующим такие связи, является решетка понятий. Ее контекст строится на основе анализа аннотаций статей, которые доступны вместе с найденными системой ссылками.

Контексты аннотаций задаются множествами словосочетаний разной длины, определяющих тематику статьи, аннотация которой анализируется. Соответственно, на множестве концептуальных графов, построенных на предложениях аннотаций, решается задача поиска словосочетаний методом, описанным в [4]. Данный метод позволяет находить словосочетания в текстах в случаях, когда образующие их слова не являются соседними, что является главным преимуществом метода.

Разработанная технология реализована на платформе Java 2EE и имеет открытую модульную структуру.

3.1. Эксперименты в системе PubMed

Разработанная технология опробована на экспериментах с системой PubMed в режиме on-line доступа к ресурсам системы. Первоначальная настройка на контекст каждой статьи выполнялась по прилагаемым к статьям ключевым словам и словосочетаниям. На этапе формирования формального контекста решетки понятий использовался интерактивный режим отбора

объектов и их атрибутов с целью исключения заведомо неинформативных пар и сокращения размеров решетки.

В результате экспериментов получен ряд нетривиальных результатов, позволяющих судить об эффективности метода концептуального моделирования. Так рассмотренное в примере родительское понятие 'Gene expression' порождает в решетке понятий множества дочерних понятий, в которых объединены другие словосочетания, например, 'Translational Control' и 'Cell Division'. Словосочетания, образующие понятия, принадлежат разным статьям, что позволяет находить семантические связи между статьями.

4 Выводы

Разработанная технология концептуального моделирования биомедицинских данных представляет собой эффективный инструмент анализа данных, содержащихся в текстах научных публикаций. Практическое значение данной технологии состоит в том, что она позволяет использовать существующие информационные ресурсы для создания локальных тематических баз знаний, основанных на концептуальных структурах.

Работа выполнена при поддержке РФФИ: грант № 11-07-97542.

Список литературы

- [1] Bogatyrev, M.Y. Tuhtin, V.V. Conceptual Modelling In Biomedical Knowledge Discovering. – Proc. of the 3 rd Int. Conf. on Mathematical Biology and Bioinformatics, Pushchino, Russia, 2010. – p.p. 214-215.
- [2] Ganter, B., Wille, R.: Formal Concept Analysis. Mathematical Foundations, Berlin. Springer Verlag (1999).
- [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] Michael Bogatyrev and Alexey Kolossoff. Using Conceptual Graphs for Text Mining in Technical Support Services. - Lecture Notes in Computer Science, 2011, Volume 6744/2011, p.p. 466-471. Springer-Verlag, Heidelberg, 2011.
- [5] U.S. National Library of Medicine <http://www.ncbi.nlm.nih.gov/pubmed>.

Анализ отличия вероятностных распределений значений параметров гемодинамики сердечно-сосудистой системы кроликов на примере их сезонных изменений

Зотов А.К., Зотова Т.Ю., Черномордик О.М.

Российский университет дружбы народов, Москва
Zotovat@mail.ru

The analysis of seasonal changes of probabilistic distribution's of cardiovascular system's hemodynamic parameters was carried out with the help of nonparametric q -criterion of probabilistic distribution's difference. The obtained results had shown that adaptive answers of cardiovascular system to seasonal changes took place not only on the level of parameter values, but on the level of parameter values probabilistic distributions.

1 Материалы и методы

Экспериментальные материалы получены на кафедре общей патологии и патологической физиологии РУДН и используются в данном исследовании ретроспективно [1]. В этом хронобиологическом эксперименте было задействовано 480 интактных кроликов-самцов породы «шиншилла» массой 2,6-3,5 кг. Каждый сезон 21 числа соответствующего месяца с 0 часов в течение 3 суток с интервалом 3 часа производились замеры параметров систолического ($АД_с$) и диастолического ($АД_д$) артериального давления, рабочего внутрижелудочного давления правого ($ВД_рПЖ$) и левого ($ВД_лПЖ$) желудочков сердца, а также максимального внутрижелудочного давления правого ($ВД_мПЖ$) и левого ($ВД_мЛЖ$) желудочков сердца. В данном сообщении рассматривается возможность применения непараметрического критерия для проверки отличия вероятностных распределений значений вышеупомянутых параметров, основанного на статистике Колмогорова-Смирнова.

$S(n,m) = \sup_{-\infty < x < +\infty} |F_n(x) - G_m(x)|$
где $F_n(x) = (\text{число}\{x_i < x\})/n$, $G_m(x) = (\text{число}\{y_i < x\})/m$ - эмпирические функции распределения, построенные соответственно по случайным выборкам

$$X = \{x_1, x_2, \dots, x_n\}, Y = \{y_1, y_2, \dots, y_m\}$$

из генеральных совокупностей с непрерывными функциями распределения $F(x)$ и $G(x)$.

Ставится задача статистической проверки гипотезы

$$H_0: F(x) = G(x)$$

Статистика критерия – это некоторая функция от наблюдений $f(X,Y)$. Критерий значимости строится следующим образом: в предположении H_0 находится функция (точного) распределения статистики $f(X,Y)$:

$$P_0(a) = P(f(X,Y) < a | H_0)$$

Для некоторого достаточно малого числа Q , которое называется *уровнем значимости (критерия)*, находится *критическое значение r (критерия)*

$$1 - P_0(r) = P(f(X,Y) \geq r | H_0) \leq Q$$

решающее правило критерия: если наблюдаемое значение статистики $f^*(X,Y)$ попадает в критическую область критерия $\{f^*(X,Y) \geq r\}$, то гипотеза H_0 отвергается, в противном случае (нет оснований отвергать гипотезу) – гипотеза H_0 принимается.

Функция $P_0(a)$ (точного) распределения статистики $f(X,Y)$ не зависит от конкретного вида функций распределения $F(x)$ и $G(x)$, входящих в формулировку гипотезы $H_0: F(x) = G(x)$, поэтому критерий называется непараметрическим, в противном случае (например, если предполагается нормальность («гауссовость») этих функций) - критерий называется параметрическим.

Следует также отметить, что, во-первых, указанный способ построения критерия значимости основан на простом соображении – если H_0 гипотеза справедлива, то маловероятное событие не может произойти, и, во-вторых, как правило, критические значения заранее вычисляются для конкретных объемов выборок (n,m) и конкретных уровней значимости (Q).

Получаемые таблицы неудобны в использовании. Более удобно использовать эмпирический уровень значимости

$$q = 1 - P_0(f^*(X,Y)) = P(f(X,Y) \geq f^*(X,Y) | H_0)$$

В этом случае вместо описанного выше решающего правила, связанного с использованием табличных критических значений, используется эквивалентное решающее правило

$$q = 1 - P_0(f^*(X,Y)) \leq Q$$

Оно более простое в том смысле, что можно использовать компьютерную программу для вычисления эмпирического уровня значимости q по исходным данным на входе – значениям выборок (все равно приходится вычислять по программе наблюдаемое значение статистики $f^*(X,Y)$ и поэтому нелогично сначала вычислить по программе наблюдаемое значение статистики, а затем «искать» в таблице критическое значение). Кроме того, само значение эмпирического уровня значимости q дает исследователю дополнительную информацию – насколько уверенно он может отвергнуть гипотезу H_0 .

Компьютерная программа, по вычислению эмпирического уровня значимости q , разработана на основе результатов статьи [2]. Программа выдает значение q в %, которое для двух независимых случайных выборок позволяет показать вероятность отличия соответствующих вероятностных распределений генеральных совокупностей равную $1 - \frac{q\%}{100\%}$.

2 Результаты исследований

Полученные нами данные демонстрируют сезонные отличия вероятностных распределений параметров системной гемодинамики сердечно-сосудистой системы кроликов (AD_c , AD_d) между сезонами Весна-Лето, Весна-Осень, Весна-Зима, Лето-Зима и Осень-Зима, по крайней мере, на 7,2% уровне значимости для параметра AD_c , между сезонами Весна-Осень, Весна-Зима, Лето-Зима и Осень-Зима, по крайней мере, на 0,34% уровне значимости для параметра AD_d . Это свидетельствует о выраженных отличиях между вероятностными распределениями значений параметров AD_c и AD_d для упомянутых пар сезонов.

Аналогично данные по внутрисердечной гемодинамике ($VD_{рПЖ}$, $VD_{рЛЖ}$, $VD_{мПЖ}$, $VD_{мЛЖ}$) указывают на сезонные отличия вероятностных распределений значений параметров внутрисердечной гемодинамики сердечно-сосудистой системы кроликов между сезонами Весна-Лето, Весна-Осень, Весна-Зима, Лето-Осень и Лето-Зима, по крайней мере, на 0,0014% уровне значимости для параметра $VD_{рПЖ}$, между сезонами Весна-Лето, Весна-Зима, Лето-Осень, Лето-Зима и Осень-Зима, по крайней мере, на 0,423% уровне значимости для параметра $VD_{рЛЖ}$, между сезонами Весна-Лето, Весна-Осень, Весна-Зима, Лето-Осень и Лето-Зима, по крайней мере, на 1,54% уровне значимости для параметра $VD_{мПЖ}$, между сезонами Весна-Зима, Лето-Осень, Лето-Зима и

Осень-Зима, по крайней мере, на 0,002% уровне значимости для параметра $VD_{мЛЖ}$. Это свидетельствует о выраженных отличиях между вероятностными распределениями значений параметров внутрисердечной гемодинамики для перечисленных выше пар сезонов.

Полученные результаты также свидетельствуют о том, что адаптивные ответы сердечно-сосудистой системы кроликов к сезонным изменениям происходят не только на уровне изменений значений самих параметров [1], но и на уровне изменений их вероятностных распределений. Аналогичный вывод для адаптивных ответов сердечно-сосудистой системы кроликов также получен нами в случае дифтерийной интоксикации [3] и экспериментальной артериальной гипертензии [4].

Список литературы

- [1] Зотов А.К. Системный анализ сердечно-сосудистой гемодинамики кроликов при сезонных изменениях. *Вестник новых медицинских технологий*, 2009, XVI(1), 199 – 200.
- [2] Черномордик О.М. Об одном непараметрическом критерии однородности нескольких выборок. *Теория вероятностей и ее применение*, 1980, 25(1), 197-200.
- [3] Зотов А.К., Фролов В.А., Зотова Т.Ю. Системный анализ сердечно-сосудистой гемодинамики кроликов при дифтерии в эксперименте. *Бюллетень экспериментальной биологии и медицины*, 2009, 147(9), 266-269.
- [4] Фролов В.А., Благонравов М.Л., Зотов А.К., Зотова Т.Ю. Анализ стадийности изменений параметров гемодинамики при формировании экспериментальной артериальной гипертензии у кроликов. *Бюллетень экспериментальной биологии и медицины*, 2011, 151(1), 13 -16.

Эволюционный метод классификации биологических объектов на основе принципа ранжирования

Цыганкова И.А.

Санкт-Петербургский институт информатики и автоматизации РАН
pallada-ltd@infopro.spb.su

Method of classification of biological objects is suggested which is based on the evolutionary approach to the solution of the extremal problem of multivariable function. Method is based on a ranking of objects in the multidimensional feature space. Method is implemented with the help of modified genetic algorithm, does not require to reduce the dimensionality of the feature space and to restore the missed values. It allows creating the hierarchic class of algorithms, which model the obtaining of the decision classification function. Originally it is oriented at the use of parallel computing.

Введение

Типичной задачей, которую необходимо решать в процессе разработки новых биологически активных веществ и методов диагностики, является классификация (распознавание) объектов по геномной информации. Характерными особенностями геномных данных являются высокая размерность пространства признаков и малый объем выборки объектов.

В настоящее время известно большое количество методов классификации, эффективность которых существенно зависит от специфики предметной области, в которой эта задача сформулирована, и особенностей исходной информации. Анализ существующих методов показал, что их использование для классификации объектов по геномной информации требует или предварительного снижения размерности пространства признаков, или фактическое снижение размерности происходит уже в процессе формирования классифицирующего правила. Специфика геномной информации делает снижение размерности признакового пространства принципиально недопустимым, так как это может привести к потере значимой информации о неизвестных взаимных связях между генами, что чрезвычайно важно при создании новых методов диагностики и лекарственных средств.

Постановка задачи

Имеется конечное множество объектов $S = \{s_1, \dots, s_i, \dots, s_N\}$, которое разделено на два

непересекающихся подмножеств (классов): K_0 и K_1 . Каждый объект описывается набором параметров $s_i = \{X_i, y_i\}$, где $X_i = (x_1, x_2, \dots, x_j, \dots, x_p)$ – вектор входных параметров объекта, которые могут принимать значения из некоторого множества допустимых значений действительных чисел, y_i – классифицирующий (целевой) параметр, определяющий принадлежность объекта к одному из классов. Значения отдельных параметров x_j для некоторых объектов s_i могут быть не определены, т.е. таблица данных имеет пропущенные значения. Размерность признакового пространства R^p существенно больше объема выборки, т.е. $p \gg N$.

Требуется, не снижая размерность признаков, предложить метод, позволяющий с приемлемой точностью классифицировать объект q , заданный вектором $X_q = (x_1, x_2, \dots, x_j, \dots, x_p)$.

Метод классификации

Рассматриваемая задача классификации является плохо формализованной в силу того, что вся информация об объектах представлена лишь набором входных и выходных параметров, о которых нельзя сколько-нибудь определенно сказать, что они полны, непротиворечивы, и не искажены. В такой ситуации наиболее эффективными становятся методы [1-4], базирующиеся на эволюционном подходе к решению экстремальных задач функции многих переменных, которые в отличие от традиционных методов поиска оптимального решения, ориентированы на получение приемлемого решения, лучшего по сравнению с полученным ранее или заданным в качестве начального.

Для решения сформулированной задачи предлагается метод, основанный на предположении, что в многомерном признаковом пространстве R^p существует некоторый базовый элемент, относительно которого формируется ранжированная последовательность объектов, сепарирующая обучающую выборку на два класса. Предполагается, что базовым элементом могут быть такие объекты многомерного пространства как: точка, плоскость, прямая или ломаная линия. Метод реализует двойное ранжирование объектов относительно базового элемента: упорядочение объектов по классам и упорядочение объектов по возрастанию

расстояния от базового элемента внутри классов. Тогда принадлежность объекта q к одному из классов определяется рангом t_q классифицируемого объекта в упорядоченном ряду объектов обучающей выборки.

Поиск базового элемента проводится с использованием эволюционного подхода, реализуемого с помощью модифицированного генетического алгоритма, где в качестве исходной популяции рассматриваются сами объекты обучающей выборки. Поиск базового элемента при заданном типе его представления в пространстве R^p и выбранном варианте упорядочения классов в формируемой последовательности объектов является, по сути, самостоятельной задачей, что позволяет достаточно легко организовать параллельную реализацию алгоритма обучения.

Для оценки эффективности предлагаемого метода был создан программный комплекс, в состав которого входят: база данных, пакет программных модулей, пользовательский интерфейс. Программные модули реализованы в среде объектно-ориентированного языка программирования *C++Builder*. С помощью разработанного программного комплекса был проведен численный эксперимент с использованием реальных экспериментальных данных об уровне экспрессии генов атлантического лосося. В качестве обучающей выборки рассматривалась выборка из 100 рыб. Каждая рыба описывалась вектором числовых параметров, которые являлись результатами инструментальных измерений уровня

экспрессии различных генов у атлантического лосося. Размерность вектора параметров составляла 967. Обучающая выборка была разделена на классы K_0 и K_1 . Класс K_0 включал 50 рыб, зараженных инфекционным лососевым вирусом анемии, а класс K_1 – 50 рыб, незараженных этим вирусом.

Заключение

Проведенные расчетные исследования показали высокую эффективность предлагаемого метода классификации. Достоверность полученных результатов была подтверждена расчетами на контрольной выборке.

Метод изначально ориентирован на использование параллельных вычислений, что позволило оптимизировать вычислительные ресурсы и существенно сократить время поиска решения.

Список литературы

- [1] Evolutionary optimization / Ed. by R. Sarker, M. Mohammadian, X. Yao. – Boston etc.: Kluwer acad. publ., 2002.
- [2] A.A. Freitas. Data Mining and Knowledge Discovery with Evolutionary Algorithms. – Berlin etc.: Springer, 2002.
- [3] Z. Michalewicz. Genetic algorithms + data structures = evolution programs. – Berlin etc.: Springer, 1996.
- [4] Д. Рутковская, М. Пилиньский, Л. Рутковский. Нейронные сети, генетические алгоритмы и нечеткие системы. М.: Горячая линия, 2008.

Программа обработки кардиоинтервалов для графического представления динамики показателей variability сердечного ритма

Парамонова Е.В.⁽¹⁾, Быстрова Н.К.^(1, 2), Маевский Е.И.⁽²⁾,
Кокорев Л.С.⁽³⁾, Быстров В.С.⁽¹⁾

Институт Математических Проблем Биологии РАН, Пушкино, 142290, Россия⁽¹⁾,
Институт Теор. и Экспериментальной Биофизики РАН, Пушкино, 142290, Россия⁽²⁾,
Военно-медицинская Академия МО РФ, Санкт-Петербург, 194044, Россия⁽³⁾
ekatpl@yandex.ru

The program for cardio-processing with graphical representation for the data of the dynamica of heart rate variability is presented in this work. Program was realized on the C# with graphical part created using Visual studio express. The main algorithm and features of cardio-processing was used from our previous publications.

1 Введение

В данной работе представлены основные положения и описание программной реализации алгоритма графического представления и анализа показателей ритмокардиограмм, предложенного в работе [1]. Данный подход необходим, нагляден и эффективен для отслеживания не только динамики изменения индивидуальных показателей сердечной деятельности, но и прогноза изменения их функциональных состояний. Это особенно важно для контроля состояния у лиц, работающих в условиях стресса. Кроме того, данный подход может быть эффективен при анализе и сравнении динамики поведения различных групп, а также при оценке динамики воздействия различных фарм-препаратов.

Программа реализована на основе использования базы данных «Варикарда» [2], и может работать как самостоятельно, так и в виде дополнительного приложения к основной программе «Варикарда», что расширяет его возможности.

2 Основные положения и описание программы

Программа реализована на языке C#. Графическая часть программы создана с помощью Visual studio express. Для построения графиков достаточно один раз за сеанс работы загрузить базу данных и, после выбора новых испытуемых, нажимать «Draw».

На Рис.1 представлена часть окна с уже открытой базой данных.

В левой части указаны имеющиеся в базе данных записи обследуемых с порядковым номером, именем и датой обследования. Справа выведен список возможных параметров, по умолчанию строятся MxDMn, CV, Mo, SI.

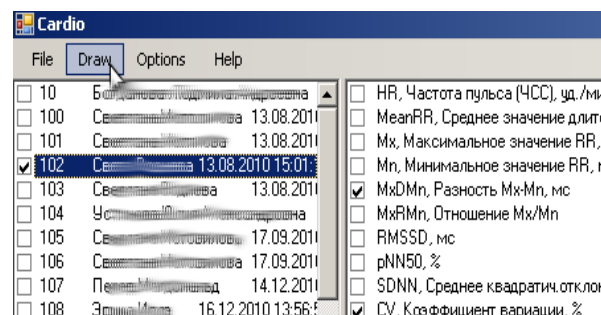


Рис. 1. Внешний вид программы Cardio для графического представления динамики показателей сердечного ритма.

Выбрать можно любой другой набор параметров, а также большее число обследуемых (или результаты одного испытуемого, полученные в разное время). Результаты для трех обследуемых, выведенные на один график, представлены на Рис.2

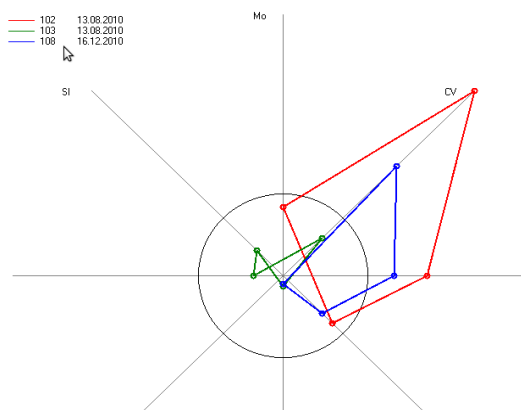


Рис. 2. Результаты для трех обследуемых, представленные в одном масштабе на одном графике.

По умолчанию на график выводится «единичный» круг (круг нормы параметров), названия осей, «легенда» с номером испытуемого (и датой исследования) и соответствующим ему на графике цветом, значения параметров (с отступом вниз и вправо от подписанной точки). В некоторых случаях подробный вывод загромождает рисунок, тогда

можно отключить вывод «легенды» и/или значений параметров. В меню «Options» выбрать отключаемую на вывод опцию: «Legend» и/или «Parameters values» и нажать «Draw». Программа снабжена подробным указателем “Help” в *.pdf формате.

Список литературы

- [1] Н.К. Быстрова, Е.И. Маевский, Е.В. Пармонова. Алгоритм прогноза развития неоптимальных функциональных состояний у лиц, работающих в условиях высокого психо-эмоционального напряжения. 1. Графическое представление и анализ выбранных показателей ритмограмм. - Математическая биология и биоинформатика. 2009. Т. 4. № 2. С. 81.
URL: [http://www.matbio.org/downloads/Bystrova2009\(4_81\).pdf](http://www.matbio.org/downloads/Bystrova2009(4_81).pdf)
- [2] Комплекс для анализа variability сердечного ритма «ВАРИКАРД». Руководство пользователя ЮИМНЗ. 941111.001 ПС. – Институт внедрения новых медицинских технологий «РАМЕНА». Рязань, 2004.

Применение метода ортогональной регрессии для решения некорректных задач в биологии

Осовец С.В.

Южно-Уральский институт биофизики, г.Озерск Челябинской обл.,Россия,
clinic@subi.su

In this paper major algorithms of orthogonal regression in relation to linear and non-linear cases are studied. It was established, that algorithms obtained can be applied for effective processing of different relations based on the empirical data, as well as for solving incorrect problems in biology. For example, the current method is applicable to parametrical assessment of empirical distribution, and to recovering the parameters of the sum of exponents in the analysis of biological data.

Введение

Несмотря на то, что метод ортогональной регрессии, как естественное обобщение метода наименьших квадратов, был впервые предложен для линейной зависимости Вальдом [2] еще в середине прошлого века, его дальнейшее развитие как в теоретическом, так и прикладном плане оставляет желать лучшего. В отечественной литературе всего в двух монографиях [3,7], кратко упоминается метод ортогональной регрессии. В фундаментальной монографии Гаральда Крамера [5] также дано лишь краткое описание линейного случая ортогональной регрессии. К сожалению, в настоящее время все эти книги являются библиографической редкостью.

С другой стороны, не вызывают сомнений, что потенциальные возможности метода ортогональной регрессии (особенно для нелинейного случая) не исчерпаны и он может быть успешно применен не только для нахождения эффективных оценок параметров регрессионных моделей [4], но и для решения так называемых некорректных задач [9], возникающих при обработке экспериментальных данных в биологии.

Линейная ортогональная регрессия

В методе ортогональной регрессии в отличие от обычной линейной регрессии при минимизации суммы квадратов отклонений сами отклонения (а точнее их квадраты) берутся по нормали к теоретической кривой, а не вдоль оси ординат. Поэтому для линейной *ортогональной регрессии* ($y=ax+b$) сумма квадратов отклонений (обозначим ее как S_{ort}) выглядит следующим образом:

$$S_{ort} = \sum_{i=1}^n \left(\frac{(y_i - ax_i - b)}{\sqrt{1 + a^2}} \right)^2 \Rightarrow \min, \quad (1)$$

здесь $\{x_i, y_i\}$ – эмпирические значения абсцисс и ординат экспериментальных точек; n – число экспериментальных точек; a, b – параметры линейной регрессии.

Дифференцируя функционал (1) по параметрам a и b , получаем систему *нелинейных* уравнений, из которых для параметров a и b можно получить соответствующие рабочие формулы [7].

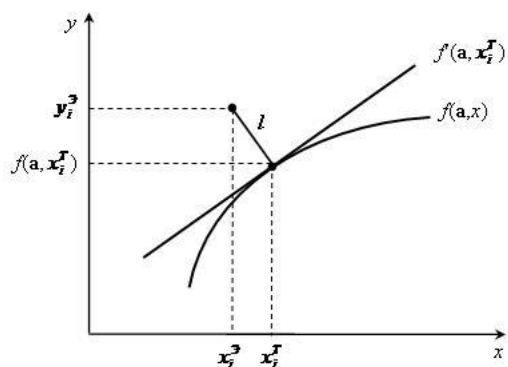
Эта простейшая линейная ортогональная регрессия была успешно использована для описания линеаризованных дозовых зависимостей некоторых эффектов, возникающих при остром аварийном облучении работников атомного предприятия [1]. Эффективность полученных оценок (по критерию суммы квадратов отклонений) в сравнении с обычным методом нелинейных квадратов была в среднем выше на 20-30%.

Далее на примере решения линейной некорректной задачи по спектрометрии смеси нуклеотидов с помощью четырехпараметрической ортогональной регрессии показана эффективность метода по сравнению с ранее разработанным методом статистической коррекции [8].

Нелинейная ортогональная регрессия

Нелинейная ортогональная регрессия в отличие от линейной имеет ряд особенностей: *во-первых*, - ее параметры можно оценить только численно, используя различные методы минимизации задаваемого функционала; *во-вторых*, алгоритм нахождения параметров можно описать только в общем виде, так как конкретный вид функционала и вспомогательных соотношений зависит от вида нелинейной регрессии, ее производной и т.д.

Введем следующие обозначения: пусть $f(a, x)$ – нелинейная регрессия с набором параметров $a = \{a_1, a_2, \dots, a_m\}$, m – число параметров; $\{x_i^3, y_i^3\}$ и $\{x_i^T, y_i^T\}$ – соответственно наборы эмпирических и теоретических вычисляемых значений пар точек $i = 1, 2, \dots, n$ – число пар точек (при этом обязательно выполняется условие $n > m$). Для получения основных формул и построения обобщенного алгоритма нахождения параметров методом нелинейной ортогональной регрессии воспользуемся, графической интерпретацией представленной на следующем рисунке.



Нелинейная ортогональная регрессия

Из графика представленного на рисунке легко получить уравнение перпендикуляра восстановленного в точке касания x_i^T и проходящего через i -тую экспериментальную точку $(x_i^Э, y_i^Э)$:

$$x_i^T = x_i^Э + f'(a, x_i^T)[y_i^Э - f(a, x_i^T)] \quad (2)$$

здесь $f'(a, x_i^T)$ – значение производной функции $f(a, x)$ в точке x_i^T все остальные обозначения прежние.

Нелинейное уравнение (2) позволяет для каждой i -ой экспериментальной точки итерационным методом Ньютона найти значение x_i^T и затем вычислить $y_i^T = f(a, x_i^T)$.

Учитывая, что для i -той точки квадрат расстояния между эмпирической точкой $\{x_i^Э, y_i^Э\}$ и вычисляемой теоретической точкой $\{x_i^T, y_i^T\}$ будет равен:

$$l^2 = (x_i^T - x_i^Э)^2 + (y_i^T - y_i^Э)^2, \quad (3)$$

и затем, обобщая полученный результат с учетом (2) на n эмпирических точек, получаем необходимый функционал с помощью которого можно найти искомые параметры нелинейной ортогональной регрессии:

$$S = \sum_{i=1}^n [1 + f'^2(a, x_i^T)] [y_i^Э - f(a, x_i^T)]^2 \Rightarrow \min. \quad (4)$$

Таким образом, обобщенный алгоритм нахождения параметров нелинейной ортогональной регрессии состоит из двух основных этапов: вначале задаются исходные значения параметров регрессии и с помощью нелинейного уравнения (2) вычисляются все неизвестные величины x_i^T и y_i^T , после этого производится минимизация по

параметрам функционала (4) и в результате находятся уточненные значения параметров. Все вычисления повторяются до тех пор, пока с заданной точностью ε не будут найдены параметры искомой нелинейной регрессии.

В нашем исследовании эффективность метода ортогональной регрессии в докладе будет показана на примере решения известной некорректной задачи, – проблеме восстановления параметров суммы трех экспонент К.Ланцоша [6], а также на примере аппроксимации функции и плотности эмпирического распределения с помощью закона Вейбулла.

Рассмотренный метод ортогональной регрессии потенциально не исчерпан и требует дальнейшего развития как в теоретическом, так и прикладном плане.

Литература

- [1] Osovets S.V., Azizova T.V., Day R.D., Wald N., Druzhinina M.B. Assessment of risk and dose thresholds for some effects of acute exposure // Health Physics, 2011, v.100, N2, pp 176-184.
- [2] Wald A. Statistical decision functions, New York, John Wiley, 1950.
- [3] Айвазян С.А. Статистическое исследование зависимостей. – М: Металлургия, 1968.
- [4] Дрейпер Н., Смит Г. Прикладной регрессионный анализ (т.1,2). – Финансы и Статистика, 1986.
- [5] Крамер Г. Математические методы статистики. – М: Мир, 1975.
- [6] Ланцош К. Практические методы прикладного анализа. – М: Государственное Издательство Физико-математической Литературы, 1961 г.
- [7] Линник Ю.В. Метод наименьших квадратов на основе математико-статистической теории обработки наблюдений. – М: Государственное издательство физико-математической литературы, 1958.
- [8] Обеснюк В.Ф., Осовец С.В. Статистическая коррекция результатов многокомпонентной спектрофотометрии нуклеотидов по априорной информации // Биохимия т.56, вып.8, 1991. – С.1503-1508.
- [9] Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. – М: Наука, 1986.

Лингвистическая форма представления электрофизиологических сигналов

Сенкевич Ю.И.

Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук
spiiran@ias.spb.su

The transformations method for the presentation of physiological signals in the form of the codes sequences is presented.

Поиск эффективных и корректных способов обработки физиологических сигналов в интересах формирования пространства признаков является одним из ключевых вопросов поиска семантической значимости получаемых значений этих признаков или их связанной совокупности. Известно, что сигналы живых систем имеют известный ряд особенностей в отличие от систем искусственного происхождения. Так выраженная нестационарность сигналов физиологических сигналов является серьезным препятствием к применению спектральных и статистических методов обработки. Однако многие медицинские эксперты используют в своей практике для постановки диагноза метод обнаружения признаков заболевания, состоящий из обычного визуального анализа записанного физиологического сигнала. Сущность метода состоит в поиске фрагментов определённой характерной формы, которые могут неоднократно встретиться на протяжении рассматриваемого сигнала достаточной длительности. Такие фрагменты в определённом смысле можно рассматривать как некоторые графические символы, отражающие выраженные реакции функционирования организма. Возникает гипотеза о том, что такие графические символы могут составлять основу более сложных сообщений, отражающих изменение функционального состояния живой системы. Развивая названный подход к изучению сигналов живых систем, можно принять модель их формирования в виде некоторого фиксированного набора правил, формирующих грамматику определённого символического языка.

Для проверки данной гипотезы потребовалось разработать специальные инструменты обработки и анализа сигналов живых систем, что и стало Задачей исследования.

В рамках решения поставленной задачи физиологический сигнал рассматривается с позиции теории информации как кодовая последовательность, которую можно анализировать как символическое сообщение, записанное на некотором условном языке. Такой подход дает возможность проводить манипуляции с кодом и анализировать сообщение подобно грамматическому

разбору обычного текста. Методика проведения исследования состоит из преобразования континуального сигнала в дискретную кодовую последовательность без существенной потери переносимой исходным сигналом информации, и, далее, выделить значимую информацию из полученного кода.

Если в ходе анализа сообщения рассматривать только значения амплитуд в экстремумах и значения интервалов между ними, то можно утверждать, что сохраняется наиболее весомая информация об амплитудно-фазовых характеристиках исходного сигнала. Для обработки выделяется совокупность, состоящий из пар чисел, определяющих положение каждого экстремума относительно ряда соседних ему экстремумов. Результаты сравнения представляются в виде матрицы. Матрица представляет определенный образ выбранного экстремума в сигнале, характеризующий его амплитудное и временное положение по отношению к соседним экстремумам на глубину соседних ближайших экстремумов. Матрица несет в себе конструктивную нагрузку, соответствующую амплитудному и фазовому значению для каждой особой точки, которую она описывает. Для поиска подобных матриц-образов экстремумов в сигнале используется следующее простое правило: если матрицы-образы, составленные для различных экстремумов, встречаются в сообщении более чем два раза, то такая матрица-образ описывает символ. Преобразуя сигнал с использованием созданного правила, получим некоторое закодированное символами сообщение.

Для обоснования поиска признаков связи обнаруживаемых символов и функциональными изменениями в организме живой системы строится лингвистическая модель функционального состояния биосистемы. Алфавитом сообщения назовём множество выделенных из сигнала символов. Размером алфавита назовем общее число обнаруженных символов алфавита в сообщении. Частота появления того или иного символа в сообщении различна и указывает на события, которые наиболее часто повторяются в генерирующей сигнал биосистеме. Если принять, что в некотором функциональном состоянии набор событий характеризующих это состояние системы не изменяется на протяжении определённого

времени, то состав алфавита A воспринимаемого сообщения, отражает набор событий.

Выдвигается гипотеза о том, что в рамках каждого устойчивого состояния система способна генерировать только фиксированный набор символов алфавита определенной размерности. Если произойдет изменение состояния системы, это должно повлечь за собой изменение набора событий, что, в свою очередь, должно вызвать качественные и количественные изменения в составе алфавита, которые можно обнаружить, применяя аппарат теории распознавания образов.

Представленные положения лингвистического анализа физиологических сигналов, были

разработаны и исследуются с 1998г. К настоящему времени для практической апробации разработанного аппарата лингвистического анализа физиологических сигналов создано несколько компьютерных программ, которые практически используются для мониторинга процесса адаптации полярников Российских антарктических экспедиций.

Лингвистический подход к анализу сигналов открывает перспективу широкого использования известных методов обработки и анализа текстовой информации для обработки и анализа физических сигналов.

Дискретные методы анализа прецедентных данных и их применение в задачах неврологии

Журавлев Ю.И.⁽¹⁾, Назаренко Г.И.⁽²⁾, Рязанов В.В.⁽¹⁾, Константинова М.В.⁽²⁾

Федеральное государственное бюджетное учреждение науки Вычислительный центр им. А.А. Дородницына Российской академии наук⁽¹⁾, Медицинский Центр ЦБ РФ⁽²⁾

zhur@ccas.ru, rvv@ccas.ru, mkdoc@yandex.ru

The mathematical methods of analysis of primary data of precedent are presented, which may be useful in the initial stages of research in poorly formalized fields. The practical results of their application to the analysis of some data from the neurology are shown.

1 Введение

Исследование многих проблем, возникающих в различных плохо формализованных областях, обычно включает два последовательных этапа создания качественной и количественной модели изучаемых объектов, процессов, ситуаций или явлений. Для создания качественной модели необходимо выделить наиболее важные параметры, общие закономерности и связи. Далее при создании точной количественной модели применяют подходы математического моделирования. Следует отметить, что успешное создание математической модели предполагает наличие предварительных исследований, исходной качественной модели, достаточного объема исходной информации и точности ее вычисления. В условиях быстрого роста объема информации (прежде всего благодаря развитию компьютерных технологий и средств Интернета) и появления новых проблем, создание моделей исследования и анализа первичной информации является актуальным и важным. В настоящей работе рассматриваются методы дискретного анализа прецедентных данных и результаты их применения при решении одной задачи неврологии.

2 Исходные данные и основные обозначения

Пусть известно описание выборки объектов $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ из некоторой генеральной совокупности. Для простоты будем считать, что $\mathbf{x}_i \in R^n$, где значение x_{ij} выражает величину некоторого числового признака, который мы можем измерять. Предполагается, что существует скрытый параметр $y = f(\mathbf{x}) \in M$, значения которого

известны для X : $y_i = f(\mathbf{x}_i), i = 1, 2, \dots, m$. Требуется по обучающей выборке $\{y_i, \mathbf{x}_i\}, i = 1, 2, \dots, m$ построить алгоритм вычисления $y = f(\mathbf{x})$ для произвольного $\mathbf{x}$. В случае $M = R$ говорят о задаче регрессии. Мы будем рассматривать случаи задач распознавания или классификации с учителем, когда $M = \{1, 2, \dots, l\}$. В последнем случае мы будем говорить о принадлежности объекта $\mathbf{x}_i$ классу (множеству объектов) $K_j, j = 1, 2, \dots, l$, если $y_i = j$.

3 Дискретный анализ данных обучения и алгоритм распознавания

В настоящее время существуют различные подходы для решения задач распознавания по прецедентам: логические, нейросетевые, статистические, геометрические, метод опорных объектов и др. Мы использовали логические модели распознавания (алгоритмы частичной прецедентности или алгоритмы вычисления оценок) [1,2] со специальным видом опорных множеств алгоритмов [3]. Принципиальная схема распознавания здесь состоит в следующем. По данным обучения для каждого класса K_j вычисляется множество $P_j = \{P_t\}$ предикатов (логических закономерностей) вида $P_t = \bigwedge_{k \in \Omega} (a_{tk} \leq x_k \leq b_{tk})$, $\Omega \subseteq \{1, 2, \dots, n\}$ обладающих свойствами: $P_t(\mathbf{x}_i) = 1$ для “нерасширяемого” подмножества класса K_j обучающей выборки, $P_t(\mathbf{x}_i) = 0, \forall \mathbf{x}_i \in X, \mathbf{x}_i \notin K_j$. При распознавании нового объекта $\mathbf{x}$ вычисляется степень близости объекта (оценка за класс, число голосов) к каждому из классов $\Gamma_j(\mathbf{x}) = \frac{1}{|P_j|} \sum_{P_t \in P_j} P_t(\mathbf{x}), j = 1, 2, \dots, l$. Объект $\mathbf{x}$

относится в класс K_j , если $\Gamma_j(\mathbf{x}) > \Gamma_\lambda(\mathbf{x}), \lambda = 1, 2, \dots, l, \lambda \neq j$. В противном случае происходит отказ от распознавания. Существующие способы вычисления оценок и опорных множеств приведены, например, в [1-3].

Нахождение множеств логических закономерностей (ЛЗ) классов позволяет (наряду с решением задачи распознавания) вычислять многие характеристики исходных данных, имеющие практическую ценность. Приведем некоторые из них.

1. Логические описания классов $F_j(\mathbf{x}) = \bigvee_{P_i \in P_j} P_i(\mathbf{x}), j = 1, 2, \dots, l$, обладающие свойствами $F_j(\mathbf{x}_i) = 1, \forall \mathbf{x}_i \in K_j$, $F_j(\mathbf{x}_i) = 0, \forall \mathbf{x}_i \notin K_j$ для непротиворечивых данных обучения.

2. Кратчайшие логические описания классов $F_j^{short}(\mathbf{x}) = \bigvee_{P_i \in P'_j \subseteq P_j} P_i(\mathbf{x}), j = 1, 2, \dots, l$, $|P'_j| \rightarrow \min, F_j^{short}(\mathbf{x}_i) = F_j(\mathbf{x}_i), \forall \mathbf{x}_i \in X$.

3. Веса признаков $0 \leq p_i \leq 1, i = 1, 2, \dots, n$.

Возможно вычисление наиболее информативных (типичных) и неинформативных объектов, логических корреляций признаков, минимальных признаковых подпространств [4].

4 Дискретный анализ данных на примере одной задачи в неврологии

Исходную информацию составила выборка из 257 больных среди которых 134 человека контрольной группы (пациенты без симптомов заболевания сосудов головного мозга или инсультов в анамнезе – класс K_1) и 123 пациента перенесших ишемический инсульт (класс K_2). Признаковое описание каждого пациента включало 20 факторов риска: результаты клинического, инструментального и лабораторного обследования, традиционные факторы риска, генетические маркеры. Для анализа данных использовалась программная система РАСПОЗНАВАНИЕ [5]. Точность распознавания в режиме скользящего контроля в модели вычисления оценок составила 78.3%, а при голосовании по большинству (использовалось 9 разнотипных алгоритмов) – 81% правильных ответов. Наиболее информативными признаками оказались: “артериальное давление”, “курение”, “стеноз б/ц”, “холестерин”, “гомоцистеин”, “триглицерид”, “ЛПВП”, “ЛПНП”, “ЛПОНП”. Признаки выражали наличие, отсутствие или степень выраженности соответствующего фактора риска. Найдено 65 логических закономерностей классов (33 для

первого и 32 – для второго). Наилучшая ЛЗ первого класса (пол=1)&(уровень АД≤3)&(3.34≤холестерин≤8.21)&(4.0≤гомоцистеин≤31)&(0.36≤триглицерид≤5.52)&(1.23≤ЛПВП≤2.78)&(1.26≤ЛПНП≤5.33)&(0.16≤ЛПОНП≤2.5)&(АроЕ4=1)&(инфаркт миокарда=0)&(фибрилляция предсердий=0)&(51≤возраст≤79) выполнялась на 37.6% объектов первого класса, а наилучшая ЛЗ второго класса (АГ=1)&(2≤уровень АД)&(ИБС=1)&(2≤х8)&(2.50≤холестерин≤8.86)&(7.3≤гомоцистеин≤29.2)&(0.84≤триглицерид)&(ЛПВП≤1.94)&(1.36<ЛПНП<3.61)&(0.38≤ЛПОНП)&(47≤Возраст≤90) выполнялась для 41.5% объектов второго класса.

Следует отметить, что множества логических закономерностей являются важным источником скрытых в обучающей выборке знаний. Они позволяют находить интервалы изменения небольшого числа факторов риска, характерные для существенного числа обучающей выборки некоторого класса. Данная задача решается неоднозначно в результате поиска локально-оптимальных решений задачи дискретной оптимизации. Система РАСПОЗНАВАНИЕ позволяет отсеивать статистически малозначимые ЛЗ (что определяется числом признаков, объектов, качеством отделимости классов). При плохой отделимости классов число статистически значимых ЛЗ, как правило, является малым, как и качество ЛЗ. ЛЗ классов имеют хорошую интерпретируемость. Они могут не только «подтвердить» известные детали, но и выдвинуть новые гипотезы, обосновать вклад каких-то факторов в процесс классификации.

References

- [1] Журавлев Ю.И., Об алгебраическом подходе к решению задач распознавания или классификации. *Проблемы кибернетики*. М.: Наука, 1978. Вып.33. С.5-68.
- [2] Журавлев Ю.И., *Избранные научные труды*. М.: Издательство Магистр, 1998. 420 с.
- [3] Рязанов В.В. Логические закономерности в задачах распознавания (параметрический подход), *Журнал вычислительной математики и математической физики*, 47(10), 2007, с.1793-1808.
- [4] Ю.И.Журавлев, И.Б.Петров, В.В.Рязанов. Дискретные методы диагностики и анализа медицинской информации. В кн.: *Медицина в зеркале информатики*. М. Наука, 2008. стр.5-21.
- [5] Ю.И.Журавлев, В.В.Рязанов, О.В.Сенько, *РАСПОЗНАВАНИЕ. Математические методы. Программная система. Практические применения*. Изд.во «ФАЗИС», Москва, 2006, 168 с.

Ортогональные разложения в задачах томографии и акустической экологии

Махортых С.А.

Институт математических проблем биологии РАН, Пущино, Россия

An approach to describe the data and to solve parameter identification systems in the vibroacoustic ecology, based on the orthogonal expansions is proposed. The method uses adaptive tunable approximation digital arrays of truncated orthogonal series. All subsequent analysis procedure is carried out in the space of the expansion coefficients. This approach can significantly reduce the computation time and the effective use of analytical methods in data processing and to improve an acoustics ecology forecast quality.

В задачах виброакустической экологии предлагается подход к описанию данных и параметрической идентификации систем на основе ортогональных разложений. Метод использует адаптивную настройку аппроксимации цифрового массива усеченным ортогональным рядом. Все последующие процедуры анализа проводятся в пространстве коэффициентов разложения. Этот подход позволяет существенно уменьшить время вычислений и эффективно использовать аналитические методы в задачах обработки данных.

Аналитическое описание данных отрезками ортогональных рядов имеет постоянную структуру. Поэтому можно легко получать аналитические зависимости для нахождения различных оценок и характеристик. В качестве базисных функций можно использовать как обычные тригонометрические ряды Фурье, так и более широкий набор функций. Ниже будут использоваться классические ортогональные многочлены [1]. Последние обладают преимуществами, обусловленными многообразием наборов базисных функций, наличием параметров, позволяющими легко модифицировать описание и подстраивать его к конкретным сигналам, возможностью получения точных аналитических зависимостей в пространстве коэффициентов разложения.

Используемые системы базисных функций являются полными и замкнутыми. Поэтому любая функция $x(t)$, принадлежащая пространству L^2 , может быть представлена точно бесконечным ортогональным рядом из любых ортонормированных классических полиномов или

функций $x(t) = \sum_{n=0}^{\infty} A_n \varphi_n(t)$, где $\varphi_n(t)$ -

ортонормированные функции из выбранного базиса, A_n - коэффициенты разложения $x(t)$ по выбранному

базису. Они вычисляются по известной формуле

$$A_n = \int_0^T x(t) \cdot \varphi_n(t) \cdot \rho(t) dt, \quad \rho(t) - \text{весовая функция.}$$

В реальных условиях число членов ряда конечно и задача, как уже было сказано, состоит в получении заданной точности аппроксимации выражением наименьшей сложности. Выполнение условия $N=N_{\min}$ требует подбора такого базиса, первые функции которого по форме будут близки или совпадать с формой изучаемого сигнала. Для интегральной оценки формы входного сигнала вводится понятие “коэффициента формы” в виде

$$K_s = \frac{\int_0^T x(t) \eta(t) dt}{\int_0^T x(t) \xi(t) dt}, \quad (1)$$

где $x(t)$ - аппроксимируемый сигнал, $\eta(t), \xi(t)$ - известные функции, выбираемые с учетом от того, какие общие свойства сигнала $x(t)$ необходимо учесть. Получаемые значения коэффициента формы для пары $\eta(t), \xi(t)$ можно откалибровать и связать с тем или иным базисом и использовать его в расчетах для выбора оптимальной базисной системы.

При моделировании волнового поля и решении задачи параметрической идентификации в спектральной формулировке будем исходить из следующего выражения:

$$y(x) = M(a(x), \alpha_i(x)), \quad (2)$$

где $y(x)$ и $a(x)$ - выходной и входной сигналы, $\alpha_i(x)$ - набор варьируемых параметров, M - заданный оператор. Если входной сигнал и параметры известны, имеем прямую задачу. В противном случае получаем обратную задачу оценки параметров $\alpha_i(x)$ сигнала $a(x)$ по известным экспериментальным данным $y(x)$. При этом решение обратной задачи в ряде случаев можно свести к решению прямой задачи на основе ее многократного решения (что и было использовано выше).

Решение прямой задачи с использованием ортогональных разложений иллюстрируется ниже на простом примере. Пусть имеется система, описываемая с помощью одномерного (зависящего от единственной независимой переменной x) линейного дифференциального оператора

$$\hat{L} = \frac{d^k}{dx^k} + a_1 \frac{d^{k-1}}{dx^{k-1}} + \dots + a_{k-1} \frac{d}{dx} + a_k : \hat{L}y = f \quad (3)$$

с граничным условием $\hat{g}y(s) = f_0$ на концах интервала s (в точках x_0 и x_1).

При этом $y(x)$ – искомая функция одной переменной. Для простоты будем считать величины a_i , b_i константами. Общая структура алгоритма решения поставленной задачи сводится к следующему. Записываются в общем виде разложения известных и искомых функций в выбранном ортогональном базисе, например в базисе полиномов Чебышева 1 рода $\{T_i(x)\}$

$$y = \sum_{i=0}^N A_i T_i, \quad f = \sum_{i=0}^N B_i T_i \quad \text{и} \quad \sum_{i=0}^N A_i \hat{g}T_i \Big|_s = f_0 \quad (4)$$

(для решения $y(x)$ и всех известных функций из (3)). Разложения (4) подставляются в уравнения (3), а затем используются известные аналитические выражения (например, переход от коэффициентов разложения функций к коэффициентам разложения производных, произведений и т.д.). В результате получается арифметическая система линейных или нелинейных уравнений относительно неизвестных коэффициентов разложения A_i с известными B_i , решение которой дает аппроксимацию (4) решения задачи (3).

Производная сигнала (функции) разлагается в ряд Фурье $y'_N(t) = \sum_{i=0}^N D_i^{(1)} T_i(t)$, при этом $\mathbf{D}^{(1)} = \mathbf{D}_N \mathbf{A}$. Здесь $\mathbf{D}^{(1)}$ и $\mathbf{A}$ – соответственно N и $N+1$ -мерные векторы-столбцы коэффициентов разложения производной и исходной функции (верхний индекс 1 означает отнесение соответствующего обозначения к первой производной функции), а $\mathbf{D}_N$ – $(N+1) \times N$ матрица, представляющая собой аппроксимацию оператора дифференцирования в выбранном функциональном базисе $\mathbf{D}_N = \{d_{ij}^{(1)}\}$.

Процедура, аналогичная описанной для решения прямой задачи, может быть применена для получения разложения функции, описывающей неизвестный параметр (например, распределения скорости упругих волн по глубине), т.е. в обратной задаче. В этом случае коэффициенты Фурье для входного и выходного сигналов задаются, а искомый спектр для $\alpha_i(x)$ вычисляется с использованием аппроксимации модели типа (3). Достаточно низкая параметрическая сложность задачи обеспечивается надлежащим выбором базиса аппроксимации.

Звуковое давление p в мягком грунте (приближение жидкой среды) описывается уравнением

$$\frac{d^2 p}{dz^2} + \omega^2 \left[\frac{1}{c^2(z)} - \frac{1}{c_\phi^2} \right] p = 0, \quad (5)$$

в рассматриваемом случае распределение скорости $c(z)$ представлено на рис. 1. Здесь z – вертикальная координата и c_ϕ – фазовая скорость в волноводе. Интенсивность источника подбирается из условия соответствия измеренным значениям.

Для аппроксимации решения уравнения (5) используются модифицированные полиномы Якоби $R_n^{\alpha\beta}(x)$ [1], определенные на интервале $(0, +\infty)$. Этот базис позволяет получить хорошую аппроксимацию для $p(z)$ рядом длиной $N=4$. Исходное распределение скорости звука и получаемое при обращении уравнения (5) показано на рис. 1. Его аппроксимация выполняется единственным многочленом с номером $n=2$ и значениями параметров $\alpha=2$, $\beta=1$.

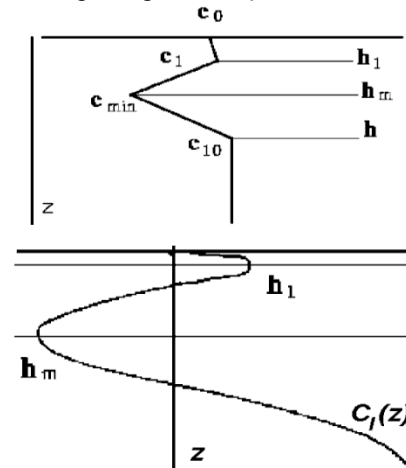


Рис. 1

Во многих случаях для вычисления требуемых коэффициентов Фурье итерационная процедура является более эффективной, чем прямое использование уравнений (5, 6). Ее главная идея состоит в использовании известного начального приближения для коэффициентов $A_i^{(0)}$. Выбирая начальное разложение $y_0(t) = \sum_{i=0}^n A_i^{(0)} T_i(t)$, можно найти более точное выражение $y_1(t) = \sum_{i=0}^n A_i^{(1)} T_i(t)$, если исходная модель может быть представлена в виде $y(t) = F(y(t), t)$, F – явным образом заданная функция или оператор.

Работа поддержана проектами РФФИ 11-07-00519 и 10-01-00609.

Список литературы

- [1] Ф.Ф.Дедус, С.А.Махортых, М.Н.Устинин, А.Ф.Дедус. Обобщенный спектрально-аналитический метод обработки информационных массивов. М.: Машиностроение, 1999, 356 с.

Формальные операции над элементами аналитических функций, исследование областей сходимости Kml-функций от двух комплексных переменных.

Быстров В.С., Камышлов В.В.

Институт математических проблем биологии РАН, Пущино, Россия
gatorado@rambler.ru

In the report the question of algebraic operations over infinite power series is considered. Regularities for such operations are received. On the basis of these regularities, functions which express roots of trinomial and quadronomial polynoms are received. These functions are investigated, areas of convergence of power series are received.

1 Введение

Одним из основных понятий в теории функции комплексной переменной является понятие кругового элемента аналитической функции. Существует определенная классификация круговых элементов аналитических функций, с которыми оперируют в комплексном анализе. Очень интересным свойством круговых элементов является то, что при выполнении над круговыми элементами различных алгебраических операций (возведение в степень, извлечение корня, умножения и деления), получаем также круговой элемент, некоторой аналитической функции, то есть имеет место свойство замкнутости круговых элементов относительно этих операций [1]. При этом, коэффициенты полученных круговых элементов, связаны с коэффициентами исходного элемента, определенной функциональной зависимостью.

Установление этой функциональной зависимости, ее математическое представление в ясной, компактной, удобной для вычисления и общей для всего класса круговых элементов форме, является основной целью этого доклада.

На базе, формальных операций над элементами аналитических функций, можно построить функции, которые выражают корни полиномов второй, третьей и более высоких степеней. Исследованию областей сходимости этих функций, посвящена вторая часть доклада.

2 Возведение в степень, извлечение корня, умножение и деление кругового элемента аналитической функции

Пусть имеется обычный круговой элемент некоторой аналитической функции с центром в точке $z=0$.

$$f(z) = a_0 + \sum_{n=1}^{\infty} a_n z^n,$$

$$a_0 = f(0) \neq 0,$$

$$a_n = \frac{f^{(n)}(0)}{n!}.$$

Необходимо выполнить над этим элементом операцию возведения в k -ю степень. После выполнения этой формальной операции, получим другой круговой элемент.

$$\varphi(z) = b_0 + \sum_{n=1}^{\infty} b_n z^n,$$

$$b_0 = \varphi(0) \neq 0,$$

$$b_n = \frac{\varphi^{(n)}(0)}{n!}.$$

$$\varphi(z) = f^k(z)$$

Коэффициенты b_n находятся путем последовательного дифференцирования исходного ряда

$$\varphi'(0) = kf^{k-1}(0),$$

$$\varphi''(0) = f^{k-2}(0) \cdot \Delta_2 = f^{k-2}(0) \cdot \begin{vmatrix} kf'(0) & kf''(0) \\ -f(0) & (k-1)f'(0) \end{vmatrix};$$

$$\varphi'''(0) = f^{k-3}(0) \cdot \Delta_3 = f^{k-3}(0) \cdot \begin{vmatrix} kf'(0) & kf''(0) & kf'''(0) \\ -f(0) & (k-1)f'(0) & (2k-1)f''(0) \\ 0 & -f(0) & (k-2)f'(0) \end{vmatrix};$$

Используя метод математической индукции, получим

$$\varphi^{(n)}(0) = f^{k-n}(0) \cdot \Delta_n,$$

где Δ_n - почти треугольный определитель n-го порядка с элементами

$$a_{ij} = [kc_{j-1}^{i-1} - c_{j-1}^{i-2}] f^{(j+i-1)}(0),$$

$$c_m^n = \begin{cases} 0, & \text{если } m < n < 0 \\ 1, & \text{если } m = n \end{cases}$$

тогда,

$$f^k(z) = a_0^k \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{z}{a_0} \right)^n \cdot \Delta_{nk} \right],$$

где Δ_{nk} - почти треугольный определитель n-го порядка с элементами

$$a_{ij} = [k(j-i+1) + (1-i)] a_{j-i+1}$$

В частном случае, $k=-1$ (операция деления единицы на ряд), получим

$$\frac{1}{f(z)} = \frac{1}{a_0} \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{z}{a_0} \right)^n \cdot \Delta_{n,-1} \right],$$

$$a_{ij} = -j \cdot a_{j-i+1}, \quad f(z) \neq 0.$$

В другом частном случае, $k=1/m$ (операция извлечения корня m-ой степени), получим

$$\sqrt[m]{f(z)} = \sqrt[m]{a_0} \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{z}{a_0} \right)^n \cdot \Delta_{n, \frac{1}{m}} \right],$$

$$a_{ij} = [(1/m) \cdot (j-i+1) + (1-i)] \cdot a_{j-i+1}$$

Аналогичным образом производятся операции по возведению в степень и извлечение корня с полярными элементами аналитических функций. Полученные выражения имеют следующий вид

$$f^k(z) = a_0^k \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{z}{a_0} \right)^n \cdot \Delta_{nk} \right],$$

$$\frac{1}{f(z)} = \frac{1}{a_0} \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{z}{a_0} \right)^n \cdot \Delta_{n,-1} \right], \quad f(z) \neq 0,$$

$$\sqrt[m]{f(z)} = \sqrt[m]{a_0} \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{z}{a_0} \right)^n \cdot \Delta_{n, \frac{1}{m}} \right]$$

И наконец, в случае критических элементов аналитических функций, которые выражаются рядами следующего вида

$$f(z) = \sum_{n=\mu}^{\infty} a_n z^{\frac{n}{p}},$$

$$a_{\mu} \neq 0;$$

при возведении в степень получим выражение следующего вида

$$f^k(z) = a_{\mu}^k \cdot z^{\frac{k\mu}{p}} \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{z^{\frac{1}{p}}}{a_{\mu}} \right)^n \cdot \Delta_{nk} \right].$$

При умножении и делении круговых элементов аналитических функций, члены нового ряда, также выражаются, через члены исходных рядов, при помощи почти треугольных определителей. Закономерности для вычисления таких определителей, получены.

3 Выводы

Полученные закономерности, позволяют проводить обращение степенных рядов, и в частном случае, полиномов. Для полиномов, становится возможным, выразить их корни, через коэффициенты [2]. Такие выражения получены. Для трехчленного полинома, независимо от его степени, корни выражаются функцией одной комплексной переменной. Эти функции исследованы, области сходимости рядов найдены [3]. Для четырехчленного полинома, корни выражаются функцией двух комплексных переменных, для этой функции также получены области сходимости рядов.

Список литературы

- [1] А. А.И. Маркушевич. Теория аналитических функций. Гостехиздат 1950, с.70-89.
- [2] Ф.Г.Кравченко. Представление корней полиномов с помощью Kml-функций. ДАН УССР, 1968, серия А, 8, с.692-696.
- [3] Ф.Г.Кравченко. Элементарные Kml- функции их свойства. Вычислительная и прикладная математика. Межведомственный научный сборник, 1970, 11, с.171-185.

Усреднение экспериментальных данных по двум и более признакам искомого сигнала

Панкратова Н.М., Устинин М.Н.

ИМПБ РАН
pnm@impb.ru

Offers an information technology of data mining to detect the desired signal from the background noise of large amplitude without using external information about its inception. The method to improve the signal-to-noise ratio is proposed, exploiting two or more features of the desired activity of the brain.

1 Задача исследования

При обработке данных магнитной энцефалографии (МЭГ) возникает проблема обнаружения полезного сигнала на фоне шумов большой амплитуды без использования внешней информации о моментах его возникновения. На этих данных искомая активность, как правило, имеет амплитуду на порядок ниже амплитуды спонтанной активности головного мозга. Ранее было показано, что для исследования данных, имеющих не только временную зависимость, но и пространственную структуру, необходимо находить и использовать пространственный признак искомого сигнала [1]. В качестве пространственного признака нами используется одна из базисных функций разложения Карунена-Лоэва (КЛ) конкретного эксперимента МЭГ. Для улучшения результата усреднения используются два признака искомого сигнала: основной и дополнительный. Для проверки предлагаемого подхода к обработке данных большого объема с пространственным распределением мы использовали данные контрольного аудиторного эксперимента, в котором здоровому человеку подавался периодический акустический стимул с частотой подачи 7,35 Гц, запись проводилась в течении 5 минут.

2 Методика

В ходе предварительной обработки по моментам подачи стимула был выделен аудиторный отклик и получена информация о структуре его магнитного поля на поверхности головы. В качестве основного пространственного признака была взята базисная функция Карунена-Лоэва (в нашем случае это была функция с номером 3), структура которой наиболее соответствует максимуму аудиторного отклика. Так же, как в работе [2], методом отсечки проекции (1) экспериментального массива на выбранный признак был получен набор опорных точек для усреднения.

$$x_i(l) = \sum_{k=1}^K h_k^l v_k^{(i)}, \quad (1)$$

где $x_i(l)$ - проекция экспериментального массива $H = \{h_k^l\}$ на $v(i)$ - i -ю базисную функцию КЛ - преобразования.

Усреднение по опорным точкам одного признака дает хороший результат (Рис.1) и позволяет выделять искомую активность, однако этот сигнал содержит шумовую компоненту, что размывает истинную полосу частот и несколько искажает решение обратной задачи (локализацию источника искомой активности).

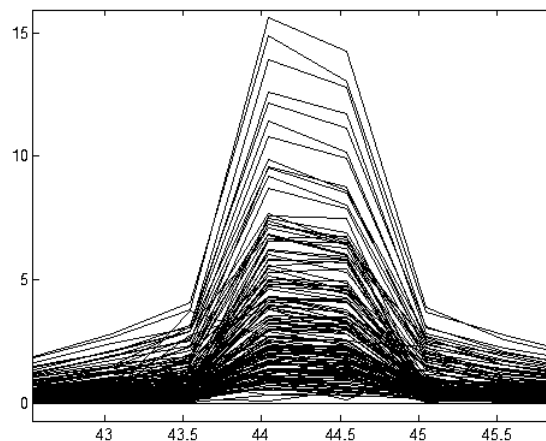


Рис.1. Полоса спектра сигнала, усредненного по опорным точкам полученным на проекции экспериментального массива на 3-ю моду КЛ.

Для того, чтобы очистить полученный набор точек, предлагается использовать дополнительный признак.

В результате деления спектра экспериментального массива и спектра предварительно выделенного отклика на стимул в зависимости от номера функции Карунена-Лоэва, был получен спектр, на котором видно, что наряду с 3-й модой КЛ, на интересующих нас частотах хорошо выделяется и 7-я мода КЛ. Так же был получен второй набор опорных точек. На Рис.2 представлен фрагмент спектра сигнала, усредненного по второму набору.

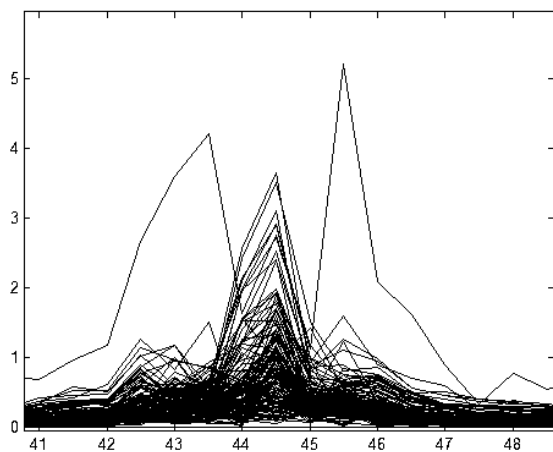


Рис.2. Полоса спектра сигнала, усредненного по опорным точкам полученным на проекции экспериментального массива на 7-ю моду КЛ.

Для улучшения соотношения сигнал-шум в выделенном сигнале, из массива основного признака для усреднения выбрали только те моменты времени, в малую окрестность которых попадали моменты возникновения дополнительного признака. После проведения усреднения по новому набору точек, был получен сигнал с более узкой полосой на спектре (Рис.3).

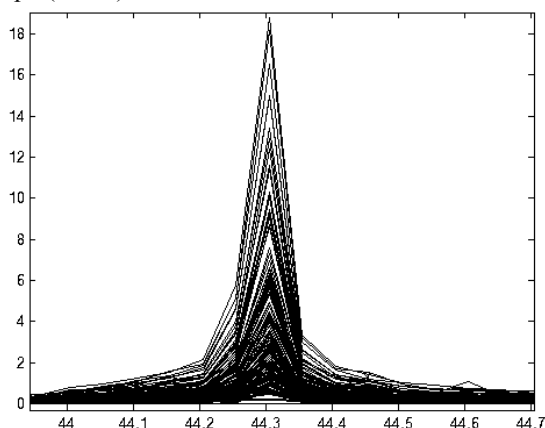


Рис.3. Полоса спектра сигнала, усредненного по уточненному набору опорных точек.

3 Выводы

В случае использования только одного пространственного признака, очистка сигнала не является идеальной и требует уточнения, для которого можно использовать дополнительные признаки искомого сигнала. Предложенный подход использования в качестве такого дополнительного признака базисной функции Карунена-Лоева с подходящей частотной характеристикой показал качественное улучшение очистки выделенного сигнала. Таким образом, можно говорить о переходе в пространство признаков искомой активности. Технология проверена на данных, полученных в контрольном эксперименте, и позволяет выделять патологическую активность на фоне спонтанной активности головного мозга человека.

Работа выполнена при частичной финансовой поддержке РФФИ, проекты 10-07-00300, 11-07-12033, 11-07-00577, 11-07-00716, 11-01-00765.

Список литературы

- [1] Панкратова Н.М., Устинин М.Н., Молчанов А.М., Линас Р. (2009) Математическая интерпретация переключений между режимами в сигналах электрической активности головного мозга. Биофизика, т.54, №5, с.916-920.
- [2] Устинин М.Н., Панкратова Н.М. (2010) Усреднение данных магнитной энцефалографии при отсутствии информации об опорных точках. Математическая биология и биоинформатика: III Международная конференция, г.Пушкино, 10-15 октября 2010г.: Доклады. С. 242-243.

О спектральном методе прогнозирования сейсмической активности

Цепков Г.В.

Кибернетический центр НАН Украины, Киев

gtsepkov@ukr.net

Proposed in the article is a method of spectral analysis for seismograms processing having a number of unique positive features, namely: it ensures obtaining of informative characteristics for evidently non-stationary processes and theoretically allows to analyse signals of infinite duration with frequency band $(0 \dots \infty)$ Hz. Proposed are as biotectors for earthquake prediction plants reacting on soil ingredient state changes by change of biopotentials shape.

экстремумах заключено не менее 97% существенной информации [1]. Таким образом, в сжатые сроки вычисляются корреляционные и спектральные характеристики сигналов, обеспечивается быстрая и эффективная фильтрация. Отметим, что высокая информационная значимость моментов, когда $x'(t) = 0$ и $x''(t) = 0$, была выявлена в экспериментах с нервной тканью [2].

1. Введение

Математический аппарат анализа сигналов при физических измерениях в сейсмологии составляют, в основном, методы вероятностного оценивания, где главную роль играют корреляционные и спектральные преобразования. Сейсмологи полагают, что землетрясения связаны с колебаниями сверхнизкой частоты, а надежность прогнозирования катаклизмов зависит от обнаружения и точности измерения параметров этих колебаний. Проблема заключается в том, что полезные сигналы о мощных процессах регистрируются приемником (сейсмографом) как слабые вследствие значительной удаленности их от генератора колебаний - ядра Земли. В то же время из-за относительной близости к приемнику твердых пород (базальт, гранит) моменты их разрушения (форшоки) фиксируются аппаратурой как инфразвуковые колебания большей амплитуды и они вместе с другими аддитивными помехами очень затрудняют выделение из сейсмограмм нужных данных.

Современные устройства спектральной обработки сейсмограмм не способны эффективно функционировать в широком спектральном диапазоне и вообще малоприспособлены для анализа слабых сверхнизкочастотных колебаний, которые к тому же не стационарны. Это означает, что при эксплуатации таких устройств обеспечивать качественное прогнозирование землетрясений невозможно.

Сейчас известны элементы новой информационной технологии, где анализу подлежат только экстремумы $\{x_i\}$ случайных аналоговых сигналов $\{x(t)\}$. Этим достигается резкое сокращение избыточности исходных данных, ибо в

2. Анализ экстремумов сейсмограмм

Поскольку в сейсмографии главную роль играет спектральная обработка данных, кратко изложим возможности пионерского метода спектроанализа [3]. Теоретическую основу этого метода составляет адаптивный секвентный базис [4,5]. Поскольку базисные функции здесь формируются с использованием амплитудно-временных параметров самого сигнала $x(t)$, теоретически метод пригоден для обработки сигналов любого класса. В этом методе исключен слепой поиск частотных компонент, т.к. интервалы между экстремумами x_i, x_{i+1} выявляют в $x(t)$ полуволну j -ой частоты: $\Delta t_i = t_{i+1} - t_i = (2f_j)^{-1}$, а об амплитуде можно судить по приращению: $\Delta x_i = |x_{i+1} - x_i|$. Суммирование и усреднение $\{\Delta x_i\}$, которые относятся к одинаковым по длине Δt_i , дает оценку амплитуды a_j спектральной компоненты сигнала.

Сейсмологам, как известно, требуется осуществлять мониторинг колебаний с длительностью не ниже 24 часов. Но все известные (кроме [3]) устройства цифрового спектрального анализа дискретов $x(t)$ требуют их предварительного запоминания, затрачивая на это «мертвое» для анализа время t_z , которое тем больше, чем длиннее волна. Это означает, что землетрясение может произойти раньше t_z .

Метод [3] допускает исключение t_z , если в анализаторе данных применяется 2 параллельно работающих процессора. Тогда обеспечивается обработка теоретически бесконечно длинных $x(t)$,

а значит и обнаружение любых колебаний в диапазоне частот $(0... \infty)$ Гц. Обработка экстремумов $x(t)$ позволяет получать пригодные для диагностирования спектральные характеристики явно нестационарных сигналов [6]. А измерение **приращений** $\{\Delta x_i\}$ обеспечивает высокую устойчивость метода к аддитивным помехам.

3. БЭР для сейсмологии

Известно, что при значительном разрушении твердых пород в мантии и литосфере Земли в почву и водоемы поступает ощутимое количество таких веществ как радон, гелий, пары ртути. Геохимики доказали [7], что резкие изменения концентрации этих элементов могут служить предвестниками землетрясений. Однако точность химических методов анализа недостаточна для уверенного прогнозирования. К тому же химические исследования трудоемки, не обладают экспрессностью, а это препятствует проведению непрерывного контроля обстановки.

В решении этой проблемы наряду с геофизическими и геохимическими исследованиями важную роль может сыграть анализ биоэлектрической реакции (БЭР) растений. Известно, что растения чутко реагируют на быстрые изменения в среде обитания температуры, влажности и другие метеорологические показатели. Так, при скачкообразном изменении освещенности в листе растения возникают колебания электрического потенциала с длительностью полуволн не больше нескольких минут. При этом БЭР здорового и угнетаемого растения очень разнятся.

Ухудшение функционального состояния растений вызывается и изменением их минерального питания, например, недостатком калия, азота или фосфора [8]. Причем особенно чувствительна корневая система растения к ингредиентам, вредным для ее питания. Это означает, что путем контроля и диагностики состояния растений, используя БЭР, можно оперативно определять наличие в

окружающей среде неблагоприятных факторов. Резонно предположить: появление в почве радона, гелия и т. п. вызовет ощутимые изменения параметров БЭР, а это может дать ценную дополнительную информацию для предсказания экологической катастрофы.

Список литературы

- [1] Дедус Ф.Ф., Пономарева И.Д., Цепков Г.В. Восстановление аналогового сигнала по экстремумам// Изв. ВУЗ. "Приборостроение". 1989, том 32, № 1, с. 7-10.
- [2] Цепков Г.В. Влияние формы сигнала на возбудимость нервной ткани// Кибернетика и вычислит. техника. Киев: Наукова думка, 1972, выпуск 14, с. 16-20.
- [3] Авторское свидетельство № 845600 СССР. Способ определения спектра аналогового сигнала / В.И. Скурихин, И.Д. Пономарева, П.М. Сиверский, Г.В. Цепков. — 1981.
- [4] Пономарева И.Д., Цепков Г.В. Сверхбыстрый спектральный анализ // Проблемы управления и информатики. 1998, № 1, с. 107-114. (Перевод: Ponomareva I.D., Tsepkov G.V. The superfast spectral analysis // J. Automation and Information Sci. 1999, v. 31, No 2, p. 93-99).
- [5] Ponomareva I.D., Tsepkov G.V. Spectral analysis of signals on the basis of extrema // Pattern recognition and image analysis. 1998, v. 8, No 4, p. 560-567.
- [6] Цепков Г.В., Яковенко И.Н. Анализ нестационарных сигналов в адаптивном секвентном базисе// Контроль. Диагностика. 2010, № 4, с. 26-31.
- [7] Барсуков В.Л., Беляев А.А. Серебрянников. Вестники беды. М.: Наука, 1989, 115 с.
- [8] Исследование биоэлектрической реакции озимой пшеницы при недостаточном минеральном питании/ М.Н. Зражевский, В.И. Лавриненко, И.Д. Пономарева, Г.В. Цепков // Доклады АН Укр.ССР. 1977, № 4, с. 338-340.

Компьютерное моделирование динамики функционирования молекулярно-генетических систем *Escherichia coli*

Акбердин И.Р.⁽¹⁾, Казанцев Ф.В.⁽¹⁾, Ермак Т.В.⁽³⁾, Ри Н.А.⁽¹⁾, Тимонов В.А.^(2,3),
Хлебодарова Т.М.⁽¹⁾ и Лихошвай В.А.^(1,2)

Институт цитологии и генетики СО РАН⁽¹⁾, Новосибирский государственный университет⁽²⁾,

Сибирский государственный университет телекоммуникаций и информатики⁽³⁾
akberdin@bionet.nsc.ru

In this paper, we have presented a hierarchical modular approach of constructing mathematical models of molecular-genetic systems developed in the ICG SB RAS on the example of the study of molecular genetic mechanisms of the pyrimidine nucleotides' biosynthesis pathway in *E. coli*. We have also demonstrated developing databases on mathematical models and kinetic data for *E. coli*. The mathematical model of ribopyrimidines biosynthesis in *E. coli* was developed and numerically explored. The model takes into account nonlinear effects of the negative regulation of the enzymes activity by the final products. These enzymes catalyze the earlier stages of the process. The mechanisms and parameters of enzymatic reactions were clarified.

1 Введение

Прогресс в секвенировании геномов различных видов организмов поставил в разряд актуальных, проблему скоростной функциональной расшифровки генетических программ, закодированных в геномах. В связи с этим, в биологии все более востребованным становится метод эксперимента *in silico*, обладающий практически неограниченным исследовательским и прогностическим потенциалом. Особенно остро эта потребность ощущается при попытке создания полномасштабной модели биологической системы типа клетки - например, *E.coli*, которая является классическим модельным объектом. Ведь именно на *E.coli* выявлены основные принципы структурно-функциональной организации генома и исследованы механизмы основных фундаментальных клеточных процессов. Но для решения этой сверхзадачи, в первую очередь, необходимо создавать и непрерывно развивать: 1) современные компьютерные подходы для анализа сложной иерархической организации молекулярно-генетических систем и позволяющие проводить ее декомпозицию на элементарные подсистемы; 2)

базовые математические модели элементарных подсистем; 3) различные базы данных с целью накопления и систематизации информации о структурно-функциональной организации этих систем, о динамике функционирования и кинетических параметрах исследуемых процессов. В данной работе мы представляем программные разработки для решения этой проблемы и демонстрируем их применение на примере метаболического пути биосинтеза рибопиримидиновых нуклеотидов, обязательных компонентов РНК, ДНК и компонентов энергетического обеспечения процессов жизнедеятельности у *E.coli*. Путь *de novo* биосинтеза пиримидиновых нуклеотидов включает 19 ферментативных реакций, в результате которых синтезируются рибо- и дезоксирибонуклеотиды УТР, СТР, dСТР и ТТР. Активность семи ферментов, контролирующих различные этапы этого пути, подвержена положительной и отрицательной регуляции по механизму обратной связи продуктами биосинтеза. В теоретических исследованиях показано, что модели генетических и биохимических систем, регулируемых по механизму обратной связи, демонстрируют сложные режимы функционирования, в том числе, осциллирующие и хаотические [1–3]. Система биосинтеза пиримидинов у *E. coli* также находится под контролем достаточно большого числа отрицательных регуляторных связей различной длины.

2 Материалы и методы

Для разработки баз математических моделей и кинетических данных использованы язык программирования «Java» и свободно распространяемые среды разработки «JDeveloper» и «Eclipse». В качестве хранилища данных использована СУБД «Oracle».

Модели элементарных подсистем ферментативных реакций, генетических систем

регуляции экспрессии, представленные в базе MGSmodelsDB, разработаны на основе закона действующих масс, методов Михаэлис-Ментен и Кинга-Альтмана, а также обобщенных функций Хилла [4]. Значения части параметров моделей взяты из литературы, суммированные данные о которых представлены в разработанной авторами базе кинетических данных KiNET [http://modelsgroup.bionet.nsc.ru/?page_id=25].

Базовая модель биосинтеза рибопиримидинов и ее различных модификаций автоматически собирались из элементарных моделей базы MGSmodelsDB, с помощью встроенного в него программного модуля, реализующего алгоритм суммирования локальных скоростей.

Значения дополнительных параметров базовой модели, возникающие на стадии сборки, оценены с использованием численной подгонки и метода градиентного спуска. Анализ индивидуального вклада каждой отрицательной регуляторной петли в функционирование системы биосинтеза рибопиримидинов проводился с помощью программного пакета STEP+ [5]. Для численного интегрирования использовался метод Гира, для анализа устойчивости стационарных решений – метод Годунова-Булгакова, реализованные в STEP+.

3 Результаты

Для интеграции литературных данных о кинетике ферментативных реакций (условия эксперимента, значения кинетических параметров, концентрации ключевых метаболитов, ферментов) в *E. coli* разработана база KiNET. В настоящее время база содержит кинетические данные для, более чем, 30 метаболических путей и описание более 500 экспериментов.

Для хранения моделей элементарных подсистем разработана база MGSmodelsDB [<http://modelsgroup.bionet.nsc.ru/MGSmodelsDB/>].

База имеет модульную структуру, позволяющую хранить: (1) структурные модели элементарных процессов; (2) математические модели элементарных процессов; (3) наборы значений параметров индивидуальных математических моделей. База снабжена встроенным программным модулем автоматической генерации математических моделей сложных метаболических и молекулярно-генетических систем на основе информации о субстратах и продуктах в отобранных элементарных процессах. В настоящее время MGSmodelsDB содержит более 100 совместимых элементарных моделей, описывающих процессы синтеза

нуклеотидов и дыхания в клетке *E.coli*. Собранные математические модели могут быть представлены в различных форматах, в том числе в формате SBML. С помощью MGSmodelsDB сгенерирована базовая модель биосинтеза рибопиримидинов и её модифицированные варианты, отличающиеся между собой количеством регуляторных связей. Проведен численный анализ моделей. Изучен индивидуальный вклад регуляторных связей в динамику функционирования системы. Исследованы лимитирующие стадии синтеза рибопиримидинов, параметрические области существования в системе стационарных и колебательных режимов функционирования. Исследованы аллостерическая и конкурентная гипотезы регуляции активности фермента UMP киназы и выявлены особенности поведения системы, которые могут быть проверены в эксперименте.

4 Заключение

В данной работе представлены базы математических моделей и кинетических данных для *E. coli*. Разработана математическая модель биосинтеза рибопиримидиновых нуклеотидов *E.coli*. Численно исследовано влияние нелинейных эффектов отрицательной регуляции активности ферментов и альтернативных механизмов регуляции активности ферментативной реакции фосфорилирования UMP.

5 Благодарности

Работа выполнена при частичной финансовой поддержке грантов РФФИ [10-01-00717-а, 12-04-31119 мол_а] и интеграционного проекта Президиума СО РАН №80. Работа также частично поддержана грантами научной школы № 5278.2012.4, программ Президиума РАН (6.8), (30.29), а также МинОбрНауки (контракт №P857)

Список литературы

- [1] Decroly O., Goldbeter A. *PNAS*, 1982, **79**(22), 6917-21.
- [2] Pigolotti S., Krishna S., Jensen M.H. *PNAS*, 2007, **104**: 6533-6537.
- [3] Xiao M., Cao J. *Math Biosci.*, 2008, **215**: 55-63.
- [4] Likhoshvai V.A., Ratushny A.V. *J. Bioinform. Comput. Biol.*, 2007, **5**, 521-531.
- [5] Fadeev S.I. *et al. Proc. of the 5th International Conference on BGRS*, 2006, **2**, 118-120.

Регуляция метаболизма гликогена в клетках скелетных мышц

Шилов Е.Б.⁽¹⁾, Беловолова О.О.⁽²⁾, Пантелеев М.А.⁽³⁾, Атауллаханов Ф.И.⁽³⁾

Центр теоретических проблем физико-химической фармакологии РАН⁽¹⁾

Физический факультет МГУ им. Ломоносова⁽²⁾

Гематологический научный центр МЗ РФ⁽³⁾

perseeker@gmail.com

Skeletal muscle metabolism is regulated by blood adrenalin levels. In particular, adrenergic stimulation affects glycogen metabolism in myocytes, the corresponding signal mechanisms being studied for a long time. However, the role of adrenergic stimulation in this case is not clear, since glycogen metabolism is also regulated by cell energy charge. Both factors play some role in forming muscle response to a physiological

stress – intensive exercise, both accelerate glycogen degradation. It is not clear, why adrenalin affects glycogen metabolism, since for most cells the energy charge sensitivity is enough.

To address these questions, we have formulated a kinetic model of glycogen regulation system. Reaction mechanisms and parameters were taken from existing literature data.

Моделирование оптимального распределения ресурсов микробом в процессе усвоения двух типов питательных веществ

Пучкова А.И., Орлов М.В.

МГУ им. М.В. Ломоносова
apuchkova@gmail.com, orlov@cs.msu.su

The nonlinear dynamic system, describing the growth process of microbial cells and resource allocation in the growth process between two types of nutrients uptake mechanisms, is considered. The aim of control process is to maximize structural biomass of the cell. Two control regimes are studied and compared in sense of maximizing the biomass.

1 Постановка задачи

Рассматривается следующая нелинейная динамическая система

$$\begin{cases} \dot{x}_1 = \omega_1 y \varphi(x_1, x_2), \\ \dot{x}_2 = \omega_2 (1 - y) \varphi(x_1, x_2), \\ \dot{y} = (u - y) \frac{d}{dt}(\varphi(x_1, x_2)), \\ x_1(\tau_0) = x_{10}, x_2(\tau_0) = x_{20}, \\ y(\tau_0) = y_0 \in (0, 1), \end{cases} \quad (1)$$

где ω_1, ω_2 - положительные константы, удовлетворяющие условию $\omega_1 + \omega_2 = 1$, $\varphi(x_1, x_2) \equiv \min\{x_1, x_2\}$. При определенных предположениях (см. [2]) система (1) моделирует процесс роста клеток микроба и распределение ресурсов в процессе роста между двумя типами механизмов усвоения питательных веществ. Функция $\varphi(x(t)) = \varphi(x_1(t), x_2(t))$ характеризует объем структурной биомассы в момент времени t , $y(\cdot)$ регулирует распределение внутренних ресурсов между механизмами усвоения питательных веществ, а функция управления $u(\cdot)$, $0 \leq u \leq 1$, отвечает за распределение вновь синтезированных ресурсов («строительных блоков»). Цель процесса управления – максимизировать $\varphi(x(T))$. Более детально биологическая интерпретация модели (1) описана в [1].

2 Скорейший выход на «сбалансированный путь» роста

В естественном предположении, что клетка микроба старается максимизировать свою среднюю скорость роста, нетрудно показать (см. [1]), что в случае, когда $x_1(\tau_0) = x_2(\tau_0)$ и $y(\tau_0) = \omega_2$, функция $u(t) \equiv \omega_2 \quad \forall t \geq \tau_0$, является наилучшим режимом управления. Легко проверить, что при таком законе управления $x_1(t) = x_2(t)$, $y(t) = \omega_2$ для всех $t \geq \tau_0$. Ситуация, когда $dx_2/dx_1 = 1$, известна в микробиологии, как «сбалансированный рост». Таким образом, логично рассмотреть задачу выхода клетки микроба на «сбалансированный путь» роста в кратчайшее время. Это приводит к следующей задаче оптимального управления в общем случае, когда либо $x_1(\tau_0) \neq x_2(\tau_0)$, либо $y(\tau_0) = \omega_2$, как можно быстрее добиться выполнения условий выхода на «сбалансированный путь» роста, т.е.

$$\begin{cases} x_1(T) = x_2(T), \\ y(T) = \omega_2, \\ T - \tau_0 \rightarrow \min_{u(\cdot)}, \end{cases} \quad (2)$$

где $T > \tau_0$ - некоторый момент времени. После момента T оптимальное управление становится очевидным.

Найдена замена переменных, которая сводит трехмерную задачу (1)-(2) к двумерной. Решение последней задачи получено и доказана его оптимальность.

3 Сравнение двух режимов

В работе также сравниваются два режима управления. Первый $\tilde{u}(t)$ решает задачу выхода клетки на «сбалансированный путь» роста в кратчайшее время (см. раздел 2). Второй режим

$$\bar{u}(x) = \begin{cases} 1, & x \in I_1, \\ 0, & x \in I_2, \end{cases}$$

где $I_1 = \{x : 0 < x_1 < x_2\}$, $I_2 = \{x : 0 < x_2 < x_1\}$, является наиболее целесообразным с биологической точки зрения, он означает, что клетка способна точно знать, какое питательное вещество в избытке в данный момент времени.

Показано, что при достаточно больших T режим $\tilde{u}$ оказывается лучше режима $\bar{u}$ с точки зрения максимизации структурной биомассы φ .

Список литературы

- [1] H.A. van den Berg, Y.N. Kiselev, S.A.L.M. Kooijman, M.V. Orlov. Optimal allocation between nutrient uptake and growth in a microbial trichome. *Math. Biol.*, 1998, **37**, 28-48.
- [2] Х. ван ден Берг, Ю.Н. Киселёв, М.В. Орлов. Исследование траекторий одной нелинейной системы дифференциальных уравнений. *Дифференциальные уравнения*, 2001, **37**, N 11, 1-10.
- [3] Ю.Н. Киселёв, М.В. Орлов. Задача быстрогодействия для одной нелинейной управляемой модели. *Вестник Моск. ун-та, Сер. 15, Вычисл. матем. и киберн.*, 2002, N 2, 23-30.

Программа SBS-I для исследования сетей метаболических реакций в больших полиферментных системах, представленных списком образующих их ферментов

Л.Н. Дроздов-Тихомиров, П.В. Ковалёв, А.А.Александров, Л. И. Уральский

Институт молекулярной генетики РАН
drozdov@img.ras.ru

Program “SBS” (STEP by STEP) was created, which enable one to investigate the topology of metabolic net in multienzyme systems presented by the list of enzymatic reactions, which may process in it. The program was used by building of mathematical model of the growing *S. cerevisiae* cell metabolism.

Разработана программа «SBS-I» (Step By Step), позволяющая проследить последовательное увеличение пула метаболитов в сети метаболизма, определяемой заданным списком возможных в системе реакций.

Прослеживается по-этапно изменение пула от его заданного начального состава до конечного, достигаемого при исчерпании заданного списка возможных в системе реакций.

На каждом этапе рассматриваемого виртуального процесса состав пула увеличивается за счёт добавления продуктов той реакции из списка, для протекания которой текущий состав пула метаболитов уже содержит необходимые субстраты.

Программа SBS-I позволяет определить - может ли, в принципе, полиферментная система, обеспечивающая возможность протекания перечисленных в характеризующем её списке реакций, обеспечить образование некоторого заданного набора целевых продуктов, если в качестве исходного материала для их синтеза будет использован некоторый заданный начальный набор субстратов.

Если выясняется, что осуществить такое превращение в заданной полиферментной системе возможно, то программа показывает какие реакции из заданного списка будут при этом работать, и в какой последовательности.

В том случае, если обнаруживается, что в исследуемой полиферментной системе полный

набор заданных целевых продуктов не может образоваться из заданных субстратов, программа показывает те метаболиты, которые могут в этой системе образоваться из заданных субстратов при полном исчерпании возможностей превращений, перечисленных в списке, и те цепочки реакций, которые приводят к их образованию.

При математическом моделировании метаболизма роста той или другой клетки в качестве основы используются, как правило, экспериментальные данные о том, какие биохимические реакции в этой клетке протекают.

Однако такие списки могут быть неполными, или, напротив, могут содержать избыточные данные, относящиеся к второстепенным процессам, которыми при моделировании основного (первичного) метаболизма можно пренебречь.

Программа SBS-I может быть эффективно использована для исследования списков метаболических реакций, составленных на основе экспериментальных данных, определения их пригодности для построения математической модели и указания на их необходимую коррекцию

Разработанная программа была использована в при построении математической модели метаболизма клетки *S.cerevisiae* на этапе составления списка реакций первичного метаболизма клетки, обеспечивающих её самовоспроизведение.

Список реакций первичного метаболизма в моделируемой клетке был получен с помощью SBS-I путём обработки большого исходного списка реакций, составленного на основе экспериментальных данных, представленных в литературе.

Модель сопряжения генерации энергии и биосинтеза в клетке

Крылов А.К.

Институт психологии РАН
neuru@mail.ru

It is proposed that the synthesis of the Krebs cycle intermediates takes place after the energy is generated enough, thus increasing the speed of the further energy generation. Then the energy and the intermediates are spent to the biosynthesis. The model of the processes shows cyclic behavior of the variables: the energy charge, the size of the intermediates pool, the calcium concentration and the biosynthesis activation.

1 Введение

Процесс генерации энергии в митохондриях использует те же субстраты, что и процесс биосинтеза – промежуточные вещества цикла Кребса («интермедиаты»: альфа-кетоглутарат, сукцинат, малат и др.) [1]. Вопрос о роли промежуточных веществ цикла Кребса является нерешенным и актуальным [2, 3]. Пониманию причин активации полезного для клетки процесса биосинтеза [6], может помочь моделирование сопряжения генерации энергии и биосинтеза [4]. Предлагаемая нами модель учитывает общность субстрата этих двух процессов.

Мы предполагаем, что чем больше пул промежуточных веществ в цикле Кребса, тем мощнее идет процесс генерации энергии (рис.1, реакция 1). С другой стороны, промежуточные вещества (интермедиаты) цикла Кребса могут выходить из цикла (рис.1, реакция 3), чтобы быть использованы для биосинтеза [1, 2] – для синтеза аминокислот, а в дальнейшем для синтеза белков. Поэтому мы предполагаем ключевую роль анаплероза (рис.1, реакция 2) для активации биосинтеза (рис.1, реакция 3), т.е. для репарационных процессов в клетке и ее развития.

Другой механизм, возможно, усиливающий биосинтез может быть связан с потреблением глутамата. Глутамат может преобразовываться в альфа-кетоглутарат и входить в цикл Кребса [1]; аналогично и ГАМК, преобразуясь в сукцинат, становится компонентом цикла Кребса (рис.1, реакция 4). Поэтому, глутамат, войдя в цикл Кребса в форме альфа-кетоглутарата может сначала усилить генерацию энергии в клетке, а затем быть использован как субстрат для биосинтеза.

2 Описание модели

Взаимосвязь между процессами генерации энергии (Э) в цикле Кребса, генерации промежуточных веществ цикла Кребса (П), и затратами на биосинтез (Б), может быть записана следующим образом:

$$dЭ/dt = П*f1(Э) - f2(Э,П) - f3(Э,П)$$

$$dП/dt = f2(Э,П) - f3(Э,П) + f4$$

$$dБ/dt = f3(Э,П)$$

Номера функций f_1 , f_2 , f_3 , f_4 соответствуют номерам реакций на рис 1. Энергия (Э) генерируется в реакции 1 и затрачивается в реакциях 2 и 3. Промежуточные вещества цикла (П), усиливающие реакцию 1, генерируются в реакции 2 и затрачиваются в 3. Скорость генерации энергии в реакции 1 пропорциональна количеству промежуточных веществ (П) цикла Кребса, в которые могут трансформироваться медиаторы глутамат или ГАМК (рис. 1, реакция 4). В такой версии модели считаем, что приток субстрата (глюкоза, лактат, пируват) – достаточный.

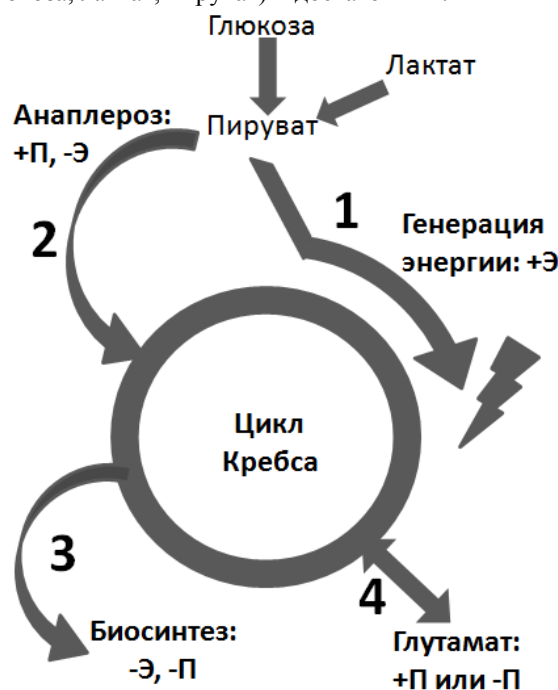


Рис. 4. Взаимосвязь процессов генерации энергии (Э) в цикле Кребса (1), генерации промежуточных веществ (П) цикла (2), их обмена с глутаматом (4) и их затратами на биосинтез (3).

3 Результаты

В этой модели мы получили колебательный характер процессов генерации энергии и биосинтеза (рис. 2), причем эти два процесса идут в противофазе – для старта биосинтеза необходимо накопление достаточного количества энергии «Э» и промежуточных веществ «П», которые потом расходуются в процессе биосинтеза «Б» и затем цикл повторяется (рис. 2, 3).

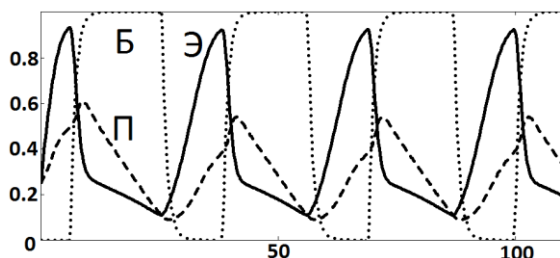


Рис. 5. Пример работы модели. Видна циклическая смена процессов генерации энергии (Э) и биосинтеза (Б). По гориз. оси – время в усл.ед.

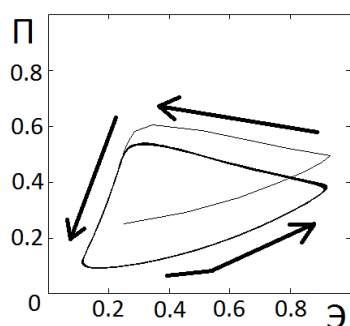


Рис. 6. Фазовый портрет колебаний количества энергии (Э) в клетке и величины промежуточных веществ (П) цикла Кребса.

Управляющей переменной для колебательного процесса в модели является концентрация кальция: при высокой концентрации кальция активируются дегидрогеназы митохондрий и усиливается процесс генерации энергии (Э) в цикле Кребса, а при низкой концентрации кальция активируется биосинтез (Б) при условии наличия достаточной энергии (Э) и субстрата (П) для него. Поэтому колебания кальция и биосинтеза происходят в противофазе (рис. 4).

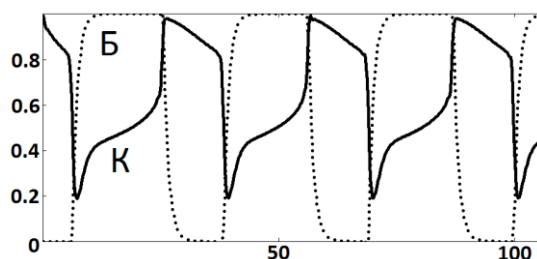


Рис. 7. Колебания кальция (К) и биосинтеза (Б) в модели. По горизонтали – время в усл.ед.

Реализуемые моделью закономерности (рис. 2, 4) взаимосвязи колебательных процессов генерации энергии, биосинтеза и динамики концентрации кальция соответствуют биологическим данным [5, 6] и могут быть использованы для объяснения: причин увеличения синтеза промежуточных веществ цикла Кребса при функциональной нагрузке и их значение для адаптации к нагрузке [3]; роли общего субстрата процессов генерации энергии и биосинтеза [1, 2]; колебательного характера уровня энергии в клетке и колебаний биосинтеза [5, 6].

Исследование выполнено при поддержке гранта РГНФ № 11-06-00482а и гранта Президента РФ для поддержки ведущих научных школ НШ-3010.2012.6.

Список литературы

- [1] J.M. Berg, J.L. Tymoczko, L. Stryer. Biochemistry. The 5th edition, W.H. Freeman & Company, 2002.
- [2] O.E. Owen, S.C. Kalhan, R.W. Hanson. The key role of anaplerosis and cataplerosis for citric acid cycle function. *J Biol Chem*, 2002, **277**, 30409–30412.
- [3] M.E. Walton, D.J.W. Ebert, R.G. Haller. Relative rates of anaplerotic flux in rested and contracted rat skeletal muscle measured by ¹³C NMR spectroscopy. *J Physiol*, 2003, **548.2**, 541-548.
- [4] С.Н. Гринченко, С.Л. Загускин. Механизмы живой клетки: алгоритмическая модель. М., Наука, 1989, -232с.
- [5] С.Л. Загускин. Околочасовые ритмы клетки и их роль в стимуляции регенерации. // Бюлл. экспер.биол.и мед. 1999, (7), С.93-96.
- [6] С.Л. Загускин. Ритмы клетки и здоровье человека. Ростов-на-Дону, Изд-во ЮФУ, 2010.-296с.

Моделирование CLV-WUS регуляции структуры ниши стволовых клеток в апикальной меристеме побега

Зубаирова У.С., Николаев С.В.

Институт цитологии и генетики СО РАН
ulyanochka@bionet.nsc.ru

In spite of many years searching of the mechanism of stem cell niche maintenance in shoot apical meristem (SAM) its particular details are not clear enough. There is a qualitative hypothesis of the interplay between *CLV3* and *WUS* genes which are believed to be able to regulate the SAM spatial compartmentalization into central zone (CZ – stem cells) and organizing center (OC). The following is key moment of the hypothesis: *CLV3* expression occurs in the CZ-cells of 3 upper layers, while *WUS* expression occurs in the OC-cells just below CZ; and *CLV3* by means of binding with putative receptor *CLV1/CLV2* inhibits *WUS* expression, while *WUS* activates *CLV3* expression. Several models have been

proposed that ascribe roles of these genes. Principle feature of our model is interaction between the SAM zones represented as sequence of the activating influences *WUS-Y-CLV3*, where *Y* is a hypothetical gene expressing in the uppermost cells of the SAM. Along with this regulation, a negative feedback *CLV3-WUS* was also implemented as interaction between CZ and OC. It was shown that the considered mechanism based on isotropic diffusion determines possible geometries of zones in the SAM, and dictates spacial locations and/or diffusion parameters of some activators and repressors to obtain the observed zone geometry.

Математическое моделирование метаболизма нитрита при культивировании клеток *Escherichia coli* в проточном хемостате

Ри Н.А.⁽¹⁾, Хлебодарова Т.М.⁽¹⁾, Когай В.В.⁽²⁾, Фадеев С.И.^(2,3), Лихошвай В.А.^(1,3)

Институт Цитологии и Генетики СО РАН⁽¹⁾, Институт Математики им. С.Л.Соболева СО РАН⁽²⁾, Новосибирский Государственный Университет⁽³⁾,
likho@bionet.nsc.ru

1 Резюме

Представлены результаты исследования модели, описывающей систему утилизации нитрита клетками *E.coli* при культивировании их в проточном хемостате. Система представляет интерес с точки зрения анализа механизмов, которые позволяют клетке использовать для дыхания токсичные для нее субстраты. Показано, что в системе существует возможность установления двух различных стационарных значений концентраций нитрита, т.е. возникновения бистабильности. Определен интервал скоростей притока нитрита в хемостат, при которых модель обладает свойством бистабильности. На основании модели предсказывается возможность спонтанной гибели культуры при переходе из одного стационарного состояния в другое.

2 Введение

В анаэробных условиях культивирования наиболее энергетически выгодным донором электронов для *E.coli* является нитрат. В процессе его утилизации образуется токсичный для клеток нитрит, который утилизируется двумя редуктазами – периплазматической NrfA и цитоплазматической NirB. Оба фермента метаболизируют нитрит до аммония, однако, при низких концентрациях нитрата более эффективна NrfA редуктаза, при промежуточных обе, а при высоких – NirB [1]. Концентрация нитритредуктаз в клетке зависит от уровня нитрита в среде, который контролирует экспрессию оперонов *nrf* и *nir*, кодирующих субъединицы ферментов, через транскрипционные факторы (ТФ) NarP и NarL.

Особенности регуляции экспрессии *nrf* оперона на генетическом уровне имитируют субстратное ингибирование ферментативной реакции, которое, как показано, может приводить к бистабильности, а при некоторых дополнительных условиях и к цикличности [2,3]. Для исследования этих возможностей нами была разработана модель

культивирования клеток *E.coli* в проточном хемостате при дыхании на нитрите. В модели принято, что клетки *E.coli* мутантны по NirB редуктазе и утилизация нитрита осуществляется только NrfA ферментом. Жизнеспособность таких клеток доказана экспериментально [4].

3 Модель утилизации нитрита клетками *E.coli*.

Модель состоит из двух уравнений и описывает изменение концентрации нитрита в хемостате (x_1) и концентрации NrfA фермента (x_2) в клетке:

$$\begin{cases} \frac{dx_1}{dt} = v - C \cdot V(x_1) \cdot x_2 - k_2 \cdot x_1, \\ \frac{dx_2}{dt} = k_s \cdot f(x_1) - k_d \cdot x_2; \end{cases} \text{ где } V(x_1) = \frac{k_{cat} \cdot x_1}{K_m + x_1}$$

В первом уравнении: v – скорость притока нитрита в хемостат, $k_2 \cdot x_1$ – скорость его выведения, $V(x_1)$ – функция, которая описывает скорость утилизации нитрита в клетке согласно кинетике реакции $\text{NO}_2^- + 7 \text{H}^+ \rightarrow \text{NH}_3 + 2 \cdot \text{H}_2\text{O}$, описываемой уравнением Михаэлиса-Ментен [5], C – коэффициент, учитывающий плотность культуры в хемостате. Во втором уравнении: $k_s \cdot f(x_1)$ и $k_d \cdot x_2$ – скорости синтеза и деградации NrfA фермента в клетке, соответственно.

В модели принято, что скорость синтеза фермента пропорциональна активности *nrf* оперона, которая управляется функцией $f(x_1)$. Функция $f(x_1)$ описывает процессы позитивной $f_1(x_1)$ и негативной $f_2(x_1)$ регуляции оперона *nrf* низкими и высокими концентрациями нитрита, которые зависят от эффективности связывания факторов NarP и NarL со своими сайтами и от расположения этих сайтов в структуре промотора. Удаленность высоко и низко аффинных сайтов связывания ТФ NarL в промоторе *nrf* оперона позволяет предположить независимость процессов активации и ингибирования его экспрессии, что отражено произведением функций $f_1(x_1)$ и $f_2(x_1)$. Уравнения, описывающие эти процессы, приведены ниже:

$$f(x_1) = f_1(x_1) \cdot f_2(x_1),$$

$$f_1(x_1) = \frac{1 + \delta_1 \cdot \left(\frac{x_1}{k_{d1}}\right)^{h_1}}{1 + \left(\frac{x_1}{k_{d1}}\right)^{h_1}}, f_2(x_1) = \frac{1 + \delta_2 \cdot \left(\frac{x_1}{k_{d2}}\right)^{h_2}}{1 + \left(\frac{x_1}{k_{d2}}\right)^{h_2}}$$

где δ_1 и δ_2 – коэффициенты активации и ингибирования скорости синтеза фермента NrfA через механизм взаимодействия ТФ NarP и NarL с высоко и низко аффинными сайтами в промоторе оперона *nrf*, соответственно; k_{d1} и k_{d2} – константы эффективности связывания ТФ с активирующими и ингибирующим сайтами, соответственно; h_1 и h_2 – коэффициенты нелинейности активирующего и ингибирующего влияния нитрита на экспрессию *nrf* оперона, соответственно.

Значения параметров модели оценены из экспериментальных данных. Модель адаптирована к эксперименту, в котором $v = 1.7 \cdot 10^{-4}$ мМ/сек [1].

4 Результаты и заключение

Анализ поведения модели проведён методом отображения параметра. В результате установлена область бистабильности $1.47 \cdot 10^{-3} \text{ мМ/сек} \leq v \leq 2.59 \cdot 10^{-3} \text{ мМ/сек}$. Из рис. 1 видно, что если изменять скорость притока нитрита в хеостат от малых значений ($v \lesssim 1.47 \cdot 10^{-3}$) к более высоким, то стационарная концентрация нитрита в хеостате плавно возрастает и при $v \sim 2.59 \cdot 10^{-3}$ достигает значения $x_1 \sim 2$ мМ (рис. 1 слева, черные стрелки). При дальнейшем увеличении скорости притока нитрита в хеостат происходит скачкообразное более чем пятикратное увеличение концентрации нитрита в хеостате (рис. 1 слева, серые стрелки). При этом новое значение концентрации нитрита превышает порог физиологической устойчивости штамма *E.coli*, используемого в работе [1]. Учитывая высокую токсичность нитрита для клетки, полученные результаты предсказывают возможность спонтанной гибели культуры при скоростях поступления нитрита в хеостат выше 0.0025 мМ/сек (рис.1 слева).

Ранее было показано, что если кинетика простой ферментативной реакции негативно зависит от концентрации субстрата, то в системе возможно возникновение явления бистабильности [2]. В исследованной системе нет прямой зависимости активности NrfA фермента от концентрации субстрата, однако эффективность деградации нитрита контролируется им самим через регуляцию экспрессии оперона, кодирующего структуру NrfA фермента. Причем, при высоких концентрациях нитрита эффективность синтеза фермента резко падает, а вместе с этим падает и эффективность деградации нитрита, что имитирует ситуацию субстратного ингибирования ферментативной реакции, но не на уровне активности фермента, а на

уровне эффективности его синтеза. Согласно модели эти особенности регуляции экспрессии NrfA фермента *E.coli* могут создать условия для установления двух различных стационарных значений концентраций нитрита в хеостате, одно из которых, в силу токсичности высоких концентраций нитрита для клетки, может привести к гибели культуры.

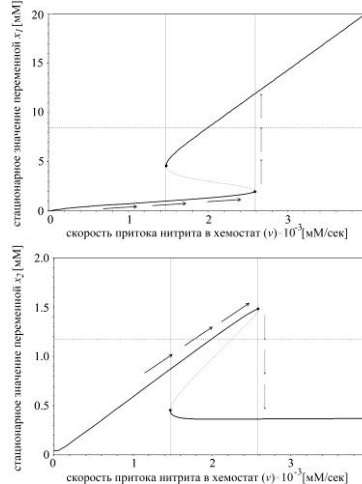


Рис. 1. Зависимость стационарного значения концентрации нитрита в хеостате x_1 (слева) и концентрации NrfA фермента в клетке x_2 (справа) от скорости притока нитрита в хеостат (v). Серый цвет кривой соответствует области неустойчивого состояния системы, черный – стабильного. Вертикальными линиями отмечена область множественности стационарных состояний; горизонтальная линия показывает примерную границу физиологической устойчивости штамма *E.coli*, используемого в работе [1].

Работа выполнена при финансовой поддержке РФФИ (проект 10-01-00717), Программ Президиума РАН «Молекулярная и клеточная биология» и «Биологическое разнообразие» (проекты 6.8 и 30.29), СО РАН (интеграционный проект №80) и гранта НШ-5278.2012.4

Список литературы

- [1] H. Wang, R.P. Gunsalus. *J. Bacteriol.*, 2000, **182**, 5813-5822.
- [2] J.F. Hervagault, S.Canu. *J. Theor. Biol.*, 1987, **127**, 439-449.
- [3] P. Shen, R.Larter. *Biophys.J.*, 1994, **67**, 1414-1428.
- [4] J.A. Cole et al. *J. Gen. Microbiol.* 1980, **120**, 475-483.
- [5] G.L. Kemp et al. *Biochem. J.* 2010, **431**, 73-80.
- [6] S.I. Fadeev et al. *Proc. of the IC. on BGRS*, 2006, **2**, 118-120.

Построение и анализ математической модели mTOR сигнального пути

Акбердин И.Р.⁽¹⁾, Трифонова Е.А.⁽¹⁾, Гайнова И.А.⁽³⁾,
Макашева В.А.⁽⁴⁾, Шорина А.Р.⁽⁴⁾, Лихошвай В.А.^(1, 2)

Институт цитологии и генетики СО РАН⁽¹⁾, Новосибирский государственный
университет⁽²⁾, Институт математики им. С.Л.Соболева СО РАН⁽³⁾,
Новосибирский областной детский клинический психоневрологический диспансер⁽⁴⁾

gajnova@math.nsc.ru

Autism Spectrum Disorders (ASD) is likely polygenic in origin in most cases, but we presently lack an understanding of the relationships between ASD susceptibility genes and the neurobiological as well as behavioral phenotypes of ASD. The mTOR/PI3K signaling pathway and serotonin pathway are both potential biomarkers for ASD diagnosis and treatment. Biochemical evidence exists for an interaction between these pathways; however, the relevance of this for the pathogenesis of ASD is unclear. A simple mathematical model describing the dynamics of the mTOR/PI3K/serotonin signaling in terms of the pseudo-bimolecular reactions, according to the law of mass action, has developed on the basis of publications for the last five years.

1 Введение

На начало 2012 года в Новосибирской области было официально зарегистрировано 2 105 детей, страдающих расстройствами аутистического спектра (РАС), что составляет приблизительно 0,5% детского населения. Зарубежная статистика (США) по данному типу патологии развития в настоящее время – около 1% детского населения.

В настоящее время показано, что расстройства аутистического спектра имеют полигенную природу, но конкретной связи между нарушениями экспрессии отдельных генов и поведенческим фенотипом пока не выявлено. В современной научной литературе одной из наиболее вероятных причин расстройств аутистического спектра считается гиперактивация mTOR/PI3K/серотонин сигнального пути.

2 Цель исследования и методы

Наследственные болезни с интеллектуальными нарушениями вносят большой вклад в структуру

психоневрологической патологии детского возраста и требуют специализированного медико-генетического консультирования. В этой связи является целесообразным более широко использовать современные генетические диагностические технологии и междисциплинарные подходы.

Недостаток экспериментальных данных и высокая стоимость проведения медико-генетических экспертиз во всех группах детей, страдающих расстройствами аутистического спектра, побуждают применять методы математического моделирования. Одним из важных преимуществ таких методов является возможность изучения побочных эффектов лекарственных препаратов в теории, а не на практике.

3 Построение математической модели

На основе публикаций последних нескольких лет [1-7] была разработана простейшая, на данном этапе исследований, математическая модель динамики функционирования геной сети, представляющей mTOR/PI3K/серотонин сигнальный путь.

Математическая модель, представляющая собой псевдо-линейную систему из 14 обыкновенных дифференциальных уравнений, состоит из элементарных блоков - подсистем, которые описывают кинетику псевдо-бимолекулярных реакций согласно закону действующих масс.

Целью проведения качественного анализа модели является создание системы прогнозирования того, как изменения экспрессии каждого конкретного гена-участника сигнального пути будут влиять на mTOR PI3K/серотонин сигнальный путь, а также исследование механизма действия лекарства – рапамицина, как mTOR-ингибитора, и эффективности лечения рапамицином в случае мутации каждого конкретного гена-участника.

4 Заключение

По данной модели были проведены расчеты на программном комплексе STEP+ [8]. Рассматривалось пять режимов состояния системы. Полученные количественные оценки позволяют судить о характере изменения концентраций белков, участвующих в сигнальном пути, при различных значениях концентрации рапамицина в крови.

Для дальнейшего развития математической модели предполагается усложнение уравнений, описывающих механизмы фосфорилирования (с помощью введения нелинейных членов в терминах обобщенных функций Хилла) [9] и расширение системы для описания динамики функционирования на молекулярно-генетическом уровне с учетом генетической регуляции.

Список литературы

- [1] J.E. Aslan, G.W. Tormoen et al. *Blood*, 2011, **118(11)**, 3129-36.
- [2] S.A. Connor, C.A. Hoeffler, E. Klann, P.V. Nguyen. *Learn Mem.*, 2011, **18(4)**, 207-20.
- [3] P.J. De Vries. *Neurotherapeutics*, 2010, **7(3)**, 275-82.
- [4] M. Rosner, K. Hofer, M. Kubista, M. Hengstschlager. *Oncogene*, 2003, **22(31)**, 4786-98.
- [5] S. Subramaniam, F. Napolitano, et al. *Nat Neurosci.*, 2011, **15(2)**, 191-3.
- [6] J. Veenstra-VanderWeele, C.L. Muller et al. *Proc Natl Acad Sci USA*, 2012, **109(14)**, 5469-74.
- [7] X. Zha, Z. Hu, S. He, F. Wang, H. Shen, H. Zhang. *Cancer Lett.*, 2011, **313(2)**, 211-7.
- [8] I.R. Akberdin, S.I. Fadeev, I.A. Gainova, F.V. Kazantsev, V.K. Korolev, V.A. Likhoshvai and A.E. Medvedev. *Proc. of the 7th ISAAC Congress*, 2010, 563-569.
- [9] V. Likhoshvai and A. Ratushny. *Journal of Bioinformatics and Computational Biology*, 2007, **5(2b)**, 521-531.

Распределение по глубине фитопланктона в океане

Абакумов А.И.^(1,2), Израильский Ю.Г.^(1,2)

Институт автоматики и процессов управления ДВО РАН, Владивосток⁽¹⁾,
Дальневосточный федеральный университет, Владивосток⁽²⁾
abakumov@iacp.dvo.ru

We propose a model describing the dynamics of phytoplankton biomass density distribution in a horizontal layer of water. This model is driven by external actors – mineral nutrition and photo synthetically active solar radiation, and water temperature. The model is applied to the satellite data over Peter the Great Gulf (sea of Japan).

1 Введение

Характеристики состояния и функционирования фитопланктона во многом определяют водные экосистемы в целом. Для морских и океанических экосистем существенное значение имеет спутниковая информация. Возникает проблема оценки состояния фитопланктона по спутниковым данным. Это задача определения того состояния фитопланктона в его распределении по глубине, которое «обеспечивает» наблюдаемые со спутника характеристики поверхностного слоя. Решение этой задачи требует сосредоточения на основных закономерностях жизнедеятельности фитопланктона, связанных с его распределением в толще воды.

Существенное значение для фитопланктона имеет его пространственное распределение, которое обладает высокой степенью неоднородности [1, 5]. Эта неоднородность определяется как внутри- и межвидовыми отношениями в фитопланктоне, так и условиями внешней среды.

2 Постановка задачи. Модель динамики биомасс

Модель описывает динамику биомасс основных видов фитопланктона в зависимости от условий внешней среды и при учете пространственного распределения по глубине в столбе воды. Энергия солнца и минеральные вещества в водной среде – основные источники пополнения биомассы. Учитываем освещенность поверхности океана и ее распределение $I(t, x)$ по глубине x . Все процессы рассматриваются во

времени t . Через $z(t, x)$ обозначена плотность массы минеральных веществ. Вектор плотности массы групп видов фитопланктона по глубине x обозначен $y(t, x) \in G \subset R_+^m$. Фитопланктон разбит на группы по сходству реакций на условия внешней среды. Множество G предполагается инвариантным относительно всех операций над вектором $y(t, x)$. Изменение биомассы фитопланктона зависит и от температуры $\theta(t, x)$ водной среды.

Наша модель имеет вид:

$$\begin{cases} \frac{\partial y_i}{\partial t} = k_i^{(y)} \frac{\partial^2 y_i}{\partial x^2} + \mu_i(t, x, z, I) y_i - e_i(y) y_i \\ \frac{\partial z}{\partial t} = k^{(z)} \frac{\partial^2 z}{\partial x^2} - \sum_{i=1}^m \gamma_i \mu_i(t, x, z, I) y_i + \\ \quad + \sum_{i=1}^m \beta_i \gamma_i e_i(y) y_i \\ \frac{\partial I}{\partial x} = -\alpha(t, x, y, z) I \end{cases}.$$

Минеральные вещества и планктон пассивно перемещаются за счет диффузии с коэффициентами диффузии $k_i^{(y)}$, $k^{(z)}$. Параметр α означает удельную скорость падения освещенности в водной среде. Функционирование фитопланктона описывается параметрами удельной скорости μ_i роста биомассы фитопланктона за счет потребленных минеральных веществ и удельной скорости e_i элиминации биомассы фитопланктона из-за выедания зоопланктоном и иных причин смертности. Параметры β_i характеризуют долю возвращаемых минеральных веществ при отмирании органики после бактериального разложения, а γ_i – усредненную долю учтенных минеральных веществ в растительных организмах.

3 Расчеты и результаты

Предварительно осуществлено моделирование годового цикла жизнедеятельности фитопланктона с учетом сезонных изменений условий среды обитания. В стационарных условиях среды наблюдается установление устойчивого

уравновешенного состояния. Проведенные расчеты позволили обоснованно применить модель для анализа динамики содержания хлорофилла в заливе Петра Великого (Японское море) в 2008 – 2010 годах. В результате модельных расчетов усредненное по поверхности содержание хлорофилла под единицей площади поверхности оказалось изменяющимся согласно кривым рис. 1. Прерывистые кривые указывают доверительные интервалы изменений показателей с доверительной вероятностью 0,95 (в предположении, что отклонения распределены нормально).

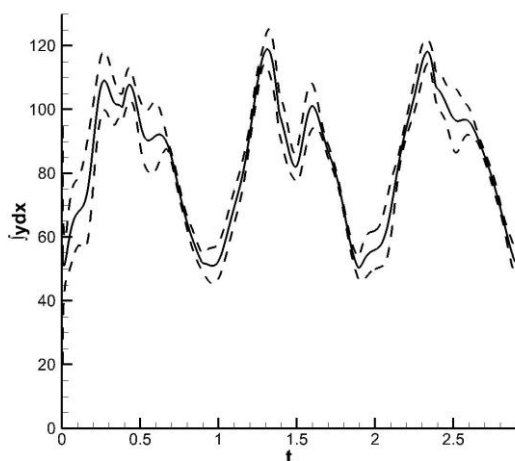


Рис. 1. Расчетная динамика содержания хлорофилла под единицей площади ($\text{мг}/\text{м}^2$) в среднем по заливу ПВ. По горизонтальной оси указано время в годах.

Содержание хлорофилла по спутниковым данным на поверхности и по модельным расчетам во всем фотическом слое воды меняются в противофазе. Это может объясняться тем, что летом фитопланктон распределен на большую глубину с уменьшением его концентрации на поверхности, зимой же он «прижат» к поверхности. Такое распределение хлорофилла получается при модельных расчетах годового цикла жизнедеятельности фитопланктона.

4 Заключение

В работе моделируется жизнь фитопланктона в толще воды. Исследуется влияние трех основных внешних факторов: солнечного света, температуры и питания на динамику биомассы фитопланктона. Найдено базовое вычислительно стабильное уравновешенное состояние

фитопланктона. Получены оценки содержания хлорофилла в заливе Петра Великого (Японское море) в 2008 – 2010 годах с помощью математической модели динамики биомассы фитопланктона по спутниковой информации о поверхностном слое моря. На основе модельных оценок содержания хлорофилла можно рассчитать первичную продукцию в заливе.

Модель используется нами также для анализа распределений фитопланктона в фотическом слое и изучения закономерностей его функционирования с учетом особенностей видового состава и сезонных факторов среды.

Работа поддержана грантом РФФИ, проект № 11-01-98517-р_восток_a, грантом ДВО РАН № 12-I-П15-02 по программе № 15 Президиума РАН, грантом ДВО РАН конкурса интеграционных проектов с институтами СО РАН № 12-II-CO-01M-010. Спутниковые данные предоставлены лабораторией спутникового мониторинга ИАПУ ДВО РАН.

Список литературы

- [1] G. Fu, K.S. Baith, and C.R. McClain. SeaDAS: The SeaWiFS Data Analysis System. *Proceedings of "The 4th Pacific Ocean Remote Sensing Conference"*, Qingdao, China, 1998, 73-79.
- [2] L. Mei, X. Zhang. On a nonlocal reaction-diffusion-advection system modeling phytoplankton growth with light and nutrients. *Discrete and continuous dynamical systems. Series B*, 2012, **17**(1), 221-243.
- [3] А.И. Абакумов, Ю.Г. Израильский. Влияние условий среды на распределение фитопланктона в водоеме. *Математическая биология и биоинформатика*, 2012, **7**(1), 274-283.
- [4] В.А. Левин, А.И. Алексанин, М.Г. Алексанина, С.Е. Дьяков, И.В. Недолужко, Е.В. Фомин. Разработка технологий спутникового мониторинга окружающей среды по данным метеорологических спутников. *Открытое образование*, 2010, **5**, 41-49.
- [5] Р.В. Озмидов. Диффузия примесей в океане, Л.: Гидрометеиздат, 1986, 280.
- [6] Г.Ю. Ризниченко, А.Б. Рубин. Математические модели биологических продукционных процессов, М.: Изд-во МГУ, 1993, 301.
- [7] В.А. Силкин, К.М. Хайлов. Биоэкологические механизмы управления в аквакультуре, Л.: Наука, 1988, 230.

Аналитическое решение неавтономной системы дифференциальных уравнений при моделировании динамики численности леммингов

Саранча Д.А.⁽¹⁾, Юрезанская Ю.С.⁽¹⁾

ВЦ РАН⁽¹⁾

july@ccas.ru

A nonlinear dynamical system vegetation – lemmings is considered, its nonlinearity is manifested in the alternation of zones of stable cyclic fluctuations in numbers of lemmings to the zones of chaos in the bifurcation diagram. Depending on the season differential equations contain a different number of terms defined by the processes of growth, the natural death and the alienation of lemmings and vegetation. The simplest form of the solution is in winter, the most difficult is in summer (the solution is expressed in terms of Bessel functions). Solutions from each previous season are used for initial data in the subsequent season.

1 Описание модели

Рассматривается нелинейная динамическая система растительность (V) – лемминги (L), её нелинейность проявляется в чередовании зон устойчивых циклических колебаний численности леммингов с зонами хаоса на бифуркационной диаграмме. Приведём систему дифференциальных уравнений, описывающих данную динамическую систему [1].

$$\begin{cases} \frac{dV}{dt} = f_V(V, L, t), \\ \frac{dL}{dt} = f_L(V, L, t). \end{cases} \quad (1)$$

Здесь t – модельное время в месяцах.

В зависимости от времени года правые части дифференциальных уравнений содержат разное количество слагаемых, определяемое процессами прироста (R_L , R_V) отчуждения (D_L , D_V) и естественного отмирания (M_L , M_V) леммингов и растительности. Наиболее простой вид решения системы зимой, наиболее сложный летом (решение выражается через функции Бесселя [2]). Решения с каждого предыдущего сезона используются для начальные данных в последующем сезоне.

2 Аналитическое решение в зависимости от сезона

2.1 Аналитическое решение зимой

Прирост, отчуждение и естественное отмирание параметризуются следующим образом: $R_L=0$, $R_V=0$, $D_V=C_2L$, $D_L=0$, $M_V=C_1V$, $M_L=C_3L$, где C_1 , C_2 , C_3 – некоторые константы. Естественная смертность леммингов M_L зависит от сезона и наличия кормов. Но с математической точки зрения функция M_L имеет всегда один и тот же вид (разница в значении константы C_3). При этом здесь и далее предполагается, что для упрощения решения системы функция смертности леммингов в зависимости от сезона, входящая в параметризацию M_L , постоянна в каждом месяце. Система (1) приобретает следующий вид:

$$\begin{cases} \frac{dV}{dt} = C_1V + C_2L, \\ \frac{dL}{dt} = C_3L. \end{cases}$$

Решения данной системы следующие:

$$L(t) = C_4 e^{C_3 t}, V(t) = a e^{C_1 t} + \frac{C_2 C_4}{C_3 - C_1} e^{C_3 t}. \text{ Здесь}$$

a , C_4 – некоторые константы, являющиеся функциями C_1 и C_2 .

2.2 Аналитическое решение весной

Прирост, отчуждение и естественное отмирание параметризуются следующим образом: $R_L=HA_1A_2L$, $R_V=0$, $D_V=CL$, $D_L=0$, $M_V=0$, $M_L=C_5L$, где H , A_1 , A_2 , C_5 – некоторые константы. При этом здесь и далее предполагается, что функция оптимальной плотности леммингов и функция, формализующая снижение ценности кормов при их дефиците, заменены на A_1 и A_2 для упрощения решения системы. Система (1) приобретает следующий вид:

$$\begin{cases} \frac{dV}{dt} = CL, \\ \frac{dL}{dt} = B_2L. \end{cases}$$

Здесь B_2 – константа, зависящая от H, A_1, A_2, C_5 .

Решения данной системы следующие:

$L(t) = Ae^{B_2 t}, V(t) = F_0 e^{B_2 t} + C_0$. Здесь A, F_0, C_0 – некоторые константы.

2.3 Аналитическое решение летом

Прирост, отчуждение и естественное отмирание параметризуются следующим образом: $R_L = H^0 A^1 A^2 L$, $R_V = C^1 \left(1 - \frac{V}{C^2}\right) \cdot V$, $D_V = C^4 L$, $D_L = 0$, $M_V = 0$, $M_L = C_6 L$, где $H^0, A^1, A^2, C^1, C^2, C^4, C_6$ – некоторые константы.

Система (1) приобретает следующий вид:

$$\begin{cases} \frac{dV}{dt} = C^1 V - KV^2 + C^4 L, \\ \frac{dL}{dt} = B^2 L. \end{cases}$$

Здесь $K = \frac{C^1}{C^2}$, B^2 – константа, зависящая от H^0 ,

A^1, A^2, C_6 . Очевидно, что функция $L(t) = M_1 e^{M_2 t}$, где M_1 и M_2 – некоторые константы. Далее решение системы сводится к решению уравнения Риккати:

$$\frac{dV}{dt} = C^1 V - KV^2 + M_1 e^{M_2 t}. \quad (2)$$

В [2] представлено решение данного уравнения, выраженное через функции Бесселя, но при $M_2 = 1$. При помощи правила дифференцирования сложной функции уравнение (2) сведено к виду,

опубликованном в [2] при произвольном M_2 . Таким образом, получено решение для V с использованием [2].

3 Выводы

Имитационная модель растительность-лемминги, подробно описанная в [3], использовалась для получения аналитических решений в различные сезоны. Данные аналитические решения могут быть использованы при тестировании программ, реализующих численные методы решения системы (1).

Работа поддержана программой ОМН-03.

Список литературы

- [1] Э.В. Недоступов, Д.А. Саранча, Е.Н. Чигерев, Ю.С. Юрезанская. О некоторых свойствах одномерных унимодальных отображений. Доклады академии наук, 2010, **430**(1), 23-28.
- [2] В.Ф. Зайцев, А.Д. Полянин. Справочник по нелинейным дифференциальным уравнениям, М.: Наука, 1993.
- [3] Д.А. Саранча. Биомоделирование. Материалы по количественной экологии. Математическое моделирование и биофизические аспекты. М.: ВЦ РАН, 1995.

О некоторых способах построения и анализа моделей экологических систем

Саранча Д.А.⁽¹⁾, Трапезов Р. В.⁽²⁾

ВЦ РАН⁽¹⁾, ИФПБ РАН⁽²⁾
tslav85@mail.ru

The method of complex research, including a full set of operations, uniting formal and informal methods, simulation and analytical approaches is proposed. This method is used in construction and analysis a set of models of tundra communities: vegetation, lemmings, polar foxes (VLF), and simplified model in the form of difference equation. The formulas relating parameters of the model VLF and difference equation are obtained, hypotheses of the leading mechanisms determining oscillations in numbers of animals are formulated. The analysis of properties of difference equations and their expression in a model of RLP is conducted.

1 Метод комплексных исследований

Эколого-биологическая область может считаться одним из лидеров в поисках эффективных способов применения точных методов в междисциплинарных исследованиях. Уже первые модели А.Лотки и В.Вольтерра показали свою эффективность в решении биологически значимых задач. Но ограниченные возможности аналитических (докомпьютерных) методов не позволили развить этот успех.

Появление «системной динамики» Дж.Форрестера [3] – метода создания имитационных моделей существенно увеличило возможности междисциплинарных взаимодействий. Имитационная технология позволяет учитывать практически все пожелания экспертов в количественной и/или качественной форме, обеспечить необходимый уровень детализации, эффективно подойти к пониманию сути описываемого явления и выражению этого понимания в форме математической модели.

Но с помощью чисто имитационных средств затруднительно получить удовлетворительное описание изучаемого явления, выделить его ведущие механизмы даже при создании идеальных условий для междисциплинарных взаимодействий. Привлекательно выглядит сочетание имитационных и аналитических подходов.

Поиск способов реализации таких сочетаний привел к формированию комплексных исследований (КОИС), включающих в себя полный набор операций:

- Сбор и анализ исходной биологической информации;
- Построение детальной имитационной модели в диалоге с экспертами, и анализ её свойств;
- Построение аналитической модели на основе аппроксимации вычислительных экспериментов с имитационной моделью, анализ её свойств;
- Формирование имитационной системы – набора взаимосвязанных моделей разной степени детализации, включающей в себя упрощенные - аналитические модели;
- Формулирование на основе этих моделей гипотез о ведущих механизмах исследуемого явления.

2 Модель растительность – лемминги – песцы

Данный метод был реализован при построении модели растительность – лемминги – песцы (РЛП) [2].

В результате вычислительных экспериментов с моделью РЛП были получены разнообразные динамические режимы, характерные для тундры колебания численности леммингов и песцов.

Неудовлетворенность традиционным окончанием исследований имитационных моделей – прогнозом динамических характеристик модели при различных сценариях внешних воздействий [2], и стремление приблизиться к пониманию механизмов формирования динамики численности тундровых животных привели к созданию модели популяции леммингов [3], определяющей характер колебаний численности животных тундрового сообщества. Это дало возможность обосновать в качестве упрощенной модели одномерное разностное уравнение, связывающее численности леммингов (ведущего блока в модели РЛП) в двух соседних годах [2].

3 Разностное уравнение

На основе анализа имитационной модели РЛП, и с помощью дополнительных предположений удалось получить разностное уравнение (1).

$$\tilde{L}_{n+1} = \begin{cases} P\tilde{L}_n, & \tilde{L}_n \leq 1/P, \\ 1 - r(\tilde{L}_n - 1/P), & 1/P < \tilde{L}_n \leq \tilde{B}, \\ d, & \tilde{L}_n > \tilde{B}. \end{cases} \quad (1)$$

Разностное уравнение позволило определить в исходной имитационной модели области параметров, обеспечивающие динамические режимы изменения численностей популяций близкие к наблюдаемым в природе. И сформулировать гипотезы о ведущих механизмах, определяющих колебания численности тундровых животных. Особая роль упрощенной модели в исследовании колебаний численности популяций тундровых животных привела к поиску более тесной связи разностного уравнения и исходной (имитационной) модели РЛП. На основе совместного анализа эколого-биологической информации и результатов вычислительных экспериментов удалось сформулировать и решить «обратную имитационную задачу». Она состоит во введении таких дополнительных предположений, которые позволили получить формулы, связывающие параметры исходной модели сообщества с параметрами разностного уравнения.

4 Сформулированные гипотезы

Анализ разностного уравнения позволил выделить три главных показателя, которые формируют динамические режимы колебаний численности леммингов. Такими показателями являются скорость прироста биомассы в благоприятный год P , максимальная численность V и выживаемость в наиболее неблагоприятных условиях d (или же два безразмерных показателя - относительная скорость прироста популяции и доля гарантированно выживших зверьков). Все эти три показателя сформировались в результате коэволюции характеристик леммингов - физиологических и экологических, и параметров среды. Первый показатель характеризует баланс между процессами рождаемости и смертности во всех фазах развития, когда нет "давления среды". Второй - характеризует экосистему в целом и

выступает в основном показателем коэволюции 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.

Полученные выводы хорошо согласуются с одной из распространенных гипотез о том, что формирует колебания численности популяций не какой-то отдельно взятый фактор, а некоторая их комбинация [1]. При этом указаны эти комбинации и показано (количественно) как они влияют на формирование динамики численности животных.

Кроме того, удалось получить формулы перехода между моделью РЛП и разностным уравнением, что обеспечивает совместимость этих моделей. Таким образом, упрощенное разностное уравнение можно использовать для настройки модели РЛП (и других более детальных моделей) на необходимые режимы функционирования, а детальные модели для более подробного изучения процесса.

Работа поддержана программой ОМН-03.

Список литературы

- [1] Орлов В.А., Саранча Д.А., Шелепова О.А. Математическая модель динамики численности популяции леммингов (*Lemmus Dicrostonyx*) и ее использование для описания популяций Восточного Таймыра // Экология. 1986. №2. С. 43-51.
- [2] Саранча Д.А. Количественные методы в экологии. Биофизические аспекты и математическое моделирование. М.: МФТИ, 1997. 283 с.
- [3] Форрестер Дж. Мировая динамика. М.: Мир, 1978. 234 с.

Интегродифференциальная модель динамики конкурирующих популяций, подверженных воздействию вредных веществ

Пичугина А.Н.⁽¹⁾, Перцев Н. В.⁽²⁾

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

Омский государственный университет им. Ф.М. Достоевского⁽¹⁾

Омский филиал Федерального государственного бюджетного учреждения науки Института
математики им. С. Л. Соболева СО РАН⁽²⁾

We study the systems of differential and integrodifferential equations describing the dynamics of competing populations exposed to harmful substances into the environment of individuals.

1 Введение

Работа посвящена построению и исследованию математических моделей динамики популяций, развивающихся в условиях воздействия вредных веществ, поступающих в среду обитания особей и потребляемых в составе пищевых ресурсов.

В отличие от [1, 2], здесь представлены модели, описывающие динамику популяций $A_1, \dots, A_m$, особи которых потребляют общий для всех пищевой ресурс, содержащий вредное вещество. Целью работы является исследование характерных режимов динамики численности популяций $A_1, \dots, A_m$, формируемых за счет воздействия на особей вредных веществ.

2 Модель в форме системы дифференциальных уравнений

Полагаем, что динамика численности популяций $A_1, \dots, A_m$ определяется следующими факторами:

- особи дают потомство;
- особи погибают вследствие конкуренции с другими особями;
- приток особей извне отсутствует;
- в среду обитания особей поступает вредное вещество, потребление которого может приводить к гибели особей;
- конкуренция особей за пищевой ресурс, содержащий вредное вещество, отсутствует;
- с течением времени вредное вещество распадается или теряет опасные для особей свойства.

В основу модели положена система

дифференциальных уравнений Лотки–Вольтерра, описывающая динамику численности популяций $A_1, \dots, A_m$ [5], и модель «ресурс–потребитель», учитывающая потребление особями вредных веществ [6].

Обозначим через $x_i = x_i(t)$, $c = c(t)$ – численность популяции A_i и количество вредного вещества в среде обитания особей в момент времени t , $i = 1, \dots, m$. Система уравнений модели имеет вид:

$$\begin{aligned} \frac{dx_i(t)}{dt} &= (\beta_i - \sum_{j=1}^m \gamma_{ij} x_j(t) - \sigma_i \theta_i(c(t))) x_i(t), \\ \frac{dc(t)}{dt} &= r(t) - \sum_{k=1}^m \theta_k(c(t)) x_k(t) - \delta c(t), \quad t > 0, \end{aligned} \quad (1)$$

$$x_i(0) = x_i^0 > 0, \quad c(0) = c_0 \geq 0, \quad i = 1, \dots, m.$$

Параметры системы (1) удовлетворяют условиям: все $\gamma_{ij} \geq 0$, $\sigma_i \geq 0$, $\beta_i > 0$, $\gamma_{ii} > 0$, существуют $\sigma_i > 0$ и $\delta > 0$. Для фиксированного $k = 1, \dots, m$ функция $\theta_k(c) \geq 0$ определена, не убывает, ограничена сверху, имеет непрерывную производную при $c \geq 0$ и $\theta_k(0) = 0$. Функция $r(t) \geq 0$ определена, непрерывна, ограничена сверху при $t \geq 0$.

В условиях полного отсутствия вредного вещества ($c(t) \equiv 0$) система (1) представляет собой модель Лотки–Вольтерра [5]. Предполагается, что параметры β_i , γ_{ij} таковы, что система уравнений Лотки–Вольтерра имеет локально или глобально асимптотически устойчивое положение равновесия

$$\bar{x}_1 > 0, \dots, \bar{x}_m > 0. \quad (2)$$

Исследование режимов динамики популяций $A_1, \dots, A_m$ в условиях воздействия вредного вещества опирается на несколько подходов. Для случая $r(t) = r^* = \text{const} > 0$ изучается устойчивость положений равновесия $x_1^* \geq 0, \dots, x_m^* \geq 0$, $c^* \geq 0$ модели (1), которые находятся из системы уравнений

$$(\beta_i - \sum_{j=1}^m \gamma_{ij} x_j - \sigma_i \theta_i(c)) x_i = 0, \quad i = 1, \dots, m, \quad (3)$$

$$r^* - \sum_{k=1}^m \theta_k(c) x_k - \delta c = 0, \quad x_1 \geq 0, \dots, x_m \geq 0, \quad c \geq 0.$$

Отыскание всех решений системы (3) и их последующий анализ на асимптотическую устойчивость представляет собой очень трудную задачу, несмотря на приведенное выше предположение относительно (2). Наиболее полное исследование этой задачи проведено авторами при $m=1$ и $m=2$. В случае $r(t) = \text{const}$ можно рассмотреть упрощающее предположение о том, что $c(t)$ является «быстрой» переменной. Тогда дифференциальное уравнение для $c(t)$ следует заменить на алгебраическое с последующим применением свойств неявных функций. Если $r(t) \neq \text{const}$ и существует $\lim_{t \rightarrow \infty} r(t) = r^* > 0$, то используются теоремы об устойчивости решений при постоянно действующих возмущениях. Для исследования поведения решений модели на конечных промежутках времени применяются численные методы решения задач Коши для систем дифференциальных уравнений.

3 Модель в форме системы интегродифференциальных уравнений

Естественным обобщением модели (1) является система уравнений, учитывающая возрастную структуру популяций $A_1, \dots, A_m$. В условиях полного отсутствия вредного вещества ($c(t) \equiv 0$) вместо системы (1) может быть использована интегральная модель Лотки–Вольтерра [3, 4]. Примем далее, что вредное вещество потребляется особями любого возраста каждой популяции и потребление этого вещества может приводить к гибели особей. Введем обозначения ($i = 1, \dots, m$):

$$J_i(t, u) = e^{-\int_u^t \lambda_i(x(s), c(s)) ds}, \quad \lambda_i(x, c) = \sum_{j=1}^m \gamma_{ij} x_j + \sigma_i \theta_i(c),$$

$$x_i^0(t) = \int_t^\infty L_i(a) \varphi_i(a-t) da, \quad b_i^0(t) = \int_t^\infty \mu_i(a) L_i(a) \varphi_i(a-t) da.$$

Здесь $0 \leq L_i(a) \leq 1$, $\mu_i(a) \geq 0$ – функции дожития и рождаемости особей популяции A_i , $a \geq 0$ – возраст особей, $L_i(a) = 0$ при $a \geq \tau_i$, где $0 < \tau_i < \infty$ – максимальная продолжительность жизни особей популяции A_i ; $\varphi_i(s) \geq 0$ – скорость рождения первоначально существующих особей популяции A_i (особей, рожденных при $t \leq 0$). Полагаем, что $L_i(a)$, $\mu_i(a)$, $\varphi_i(a)$ определены и непрерывны на промежутке $0 \leq a \leq \tau_i$, $i = 1, \dots, m$. Обозначим через $b_i(t)$ скорость рождения особей популяции A_i в момент времени $t > 0$, $i = 1, \dots, m$.

Система уравнений модели имеет вид:

$$b_i(t) = J_i(t, 0) b_i^0(t) + \int_0^t \mu_i(a) L_i(a) J_i(t, t-a) b_i(t-a) da, \quad t > 0,$$

$$x_i(t) = J_i(t, 0) x_i^0(t) + \int_0^t L_i(a) J_i(t, t-a) b_i(t-a) da, \quad i = 1, \dots, m,$$

$$\frac{dc(t)}{dt} = r(t) - \sum_{k=1}^m \theta_k(c(t)) x_k(t) - \delta c(t), \quad c(0) = c_0 \geq 0. \quad (4)$$

Параметры модели (4) заданы так, что для каждого $i = 1, \dots, m$ существует единственный корень $\beta_i > 0$

уравнения $\int_0^\infty \mu_i(a) L_i(a) e^{-\beta_i a} da = 1$, называемый

внутренней скоростью роста популяции A_i .

Система (4) сводится к системе, аналогичной (1). В новой системе уравнение для $x_i(t)$ таково:

$$\frac{dx_i(t)}{dt} = (\beta_i - \sum_{j=1}^m \gamma_{ij} x_j(t) - \sigma_i \theta_i(c(t))) x_i(t) + \varepsilon_i(t), \quad t > 0,$$

$$x_i(0) = x_i^0(0), \quad i = 1, \dots, m.$$

Каждая функция $\varepsilon_i(t)$ является непрерывной и стремится к нулю при $t \rightarrow \infty$, $i = 1, \dots, m$. Возникающая система уравнений позволяет использовать модель (1) и теоремы об устойчивости решений при постоянно действующих возмущениях для исследования модели (4).

Список литературы

- [1] А.Н. Пичугина. Интегродифференциальная модель популяции, подверженной воздействию вредных веществ // Сиб. журн. инд. матем. 2004. Т.7. N.4(20). С.130-140.
- [2] Н.В. Перцев, Г.Е. Царегородцева. Математическая модель динамики популяции, развивающейся в условиях воздействия вредных веществ // Сиб. журн. инд. матем. 2010. Т.13. N.1(41). С.109-120.
- [3] Н.В. Перцев. Интегральная модель Лотки–Вольтерра // Сиб. журн. инд. матем. 1999. Т.2. N.2(4). С.153-167.
- [4] Н.В. Перцев, А.Н. Пичугина, Б.Ю. Пичугин. Поведение решений диссипативной интегральной модели Лотки–Вольтерра // Сиб. журн. инд. матем. 2003. Т.6. N 2(14). С. 95-106.
- [5] Ю.М. Свирежев, Д.О. Логофет. Устойчивость биологических сообществ. М.: Наука, 1978.
- [6] П.В. Фурсова, А.П. Левич. Математическое моделирование в экологии сообществ // Проблемы окружающей среды (обзорная информация ВИНТИ). 2002. N.9.

Изучение динамики численности популяции животных с использованием индивидуально-ориентированного моделирования.

Люлякин О.П.⁽¹⁾, Саранча Д.А.⁽¹⁾

ВЦ РАН имени Дородницына⁽¹⁾

In this work it was studied dynamics of animal populations with using individual-based modeling. Also for the study of population dynamics were constructed difference equations describing the relationship the number of animals in two adjacent years. Using differential equations, it was possible to reproduce three-year dynamics similar to individual-based modeling.

1 Индивидуально-ориентированная модель

В данной работе рассматривается задача разработки инструментария проведения междисциплинарных исследований при неполной информации о свойствах моделируемого объекта с активным использованием ЭВМ.

Прогресс в развитии количественной экологии сдерживали ограниченные возможности (докомпьютерных) методов исследования. Появление «системной динамики» Дж.Форрестера[5] предоставило соответствующий инструментарий – метод создания имитационных моделей.

Данный подход использовался при анализе колебаний численности животных с помощью модели тундрового сообщества «растительность-лемминги-песцы»[4]. В результате исследований были сформулированы гипотезы о ведущих механизмах, определяющих колебания численности тундровых животных. Проведенные исследования позволили перейти на новый уровень описания, на использование метода индивидуально-ориентированного моделирования.

Эта модель используется для изучения популяции с точки зрения отдельно взятого животного, у которого выделяют параметры и процессы жизнедеятельности. В качестве объекта исследования была выбрана популяция копытных леммингов.

В этой модели учитывались следующие параметры: возраст особи, координаты норы, координаты особи и потенциал жизнестойкости, который означает, что, если его значение становится меньше нуля, то считается, что особь погибает.

Характеристики пространства: данная модель

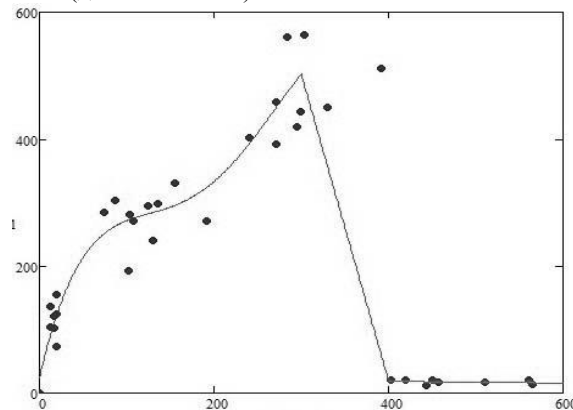
относится к классу пространственно-распределенных, все моделируемые объекты ассоциированы с определенным положением в пространстве, пространство – дискретное, двумерное, описывается в виде дискретной сетки, состоящей из ячеек.

Были разработаны алгоритмы для следующих процессов: движение и питание, рост, стычки с другими особями, изменение потенциала, беременность, рождение потомства, смерть.

Используя эти параметры и вводя начальное распределение особей по пространству, была написана программа. В результате вычислительных экспериментов получилось воспроизвести колебания численности в 3 и 4 года, характерные для копытных леммингов.

2 Использование разностных уравнений

Разностные уравнения можно использовать, чтобы связывать численность популяции в двух соседних годах. Выбираем некоторый момент времени, к примеру, середину декабря, и после выполнения численного эксперимента из индивидуально-ориентированной модели получаем точки. Разбиваем ось абсцисс условно на три участка. Далее строим, используя метод наименьших квадратов, кривую (для первого участка) и две прямых (для остальных).



Ось абсцисс – численность в некоторый момент времени, ось ординат – численность популяции в тот же момент времени на следующий год.

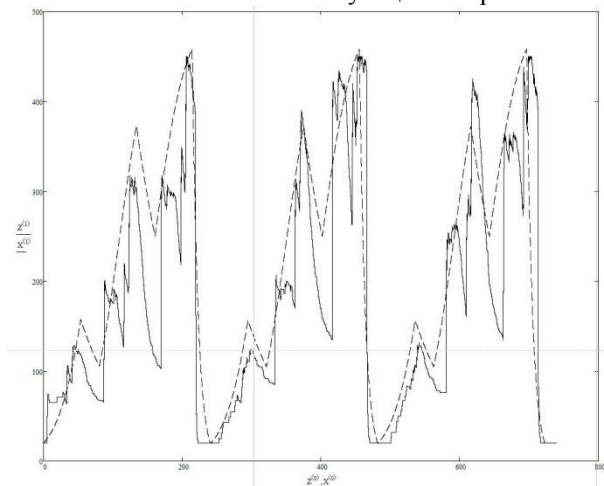
Разностные уравнения позволяют предсказывать численность популяции через определенные промежутки времени.

3 Использование дифференциальных уравнений

Напишем дифференциальные уравнения для каждого из сезонов так, чтобы воспроизвести циклы численности популяции длиной в три года.

Для сезона размножения воспользуемся логистическим уравнением Ферхюльста, а для зимовки уравнением Мальтуса.

Далее параметры подбираются таким образом, чтобы получались циклы в 3 года, похожие на результат из индивидуально-ориентированной модели. Проводя численный эксперимент, используя метод Рунге-Кутты четвертого порядка, получим зависимость численности популяции от времени.



На данном графике показана зависимость численности популяции от времени. Пунктирная линия получена с помощью дифференциальных уравнений, сплошная — индивидуально-ориентированная модель.

Из этого графика видно, что динамика численности, полученная из дифференциального уравнения и полученная из индивидуально-ориентированной модели, сравнительно хорошо совпадают при высокой численности в конце сезона. По совпадению значительно хуже в области низкой численности.

В этой модели есть минусы по сравнению с индивидуально-ориентированной моделью. Она не учитывает в явном виде различные параметры: возраст, потенциал жизнестойкости и так далее. Индивидуально-ориентированное моделирование более приближено к реальности.

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.]
- [2] Перминов В. Д., Саранча Д. А. Об одном подходе к решению задач популяционной экологии // Математическое моделирование. 2003. №11. С.45-53.
- [3] Саранча Д.А., Сорокин П.А., Фролова А.А., Математическое моделирование динамики численности популяции животных. М.: ВЦРАН. 2005. 27 с.
- [4] Саранча Д.А. Количественные методы в экологии. Биофизические аспекты и математическое моделирование. М.: МФТИ, 1997. 283 с.
- [5] Форрестер Дж. Мировая динамика. М.: Мир, 1978. 234 с.

Исследование одной управляемой модели, связанной с лечением хронического миелолейкоза

Никольский М.С.⁽¹⁾, Молчанов А.А.⁽¹⁾

Московский государственный университет им. М. В. Ломоносова,
кафедра оптимального управления⁽¹⁾
mni@mi.ras.ru, amolchanov@cs.msu.su

In this paper we present a qualitative study of a controllable mathematical model of chronic myelogenous leukemia for a single patient. The model consists of a set of non-linear differential equations with two control parameters, which makes it possible to apply control theory results to study the behaviour of the system's solutions and formulate optimization problems based on it. However, several facts need to be checked prior to any further research. Our goal is to examine the system analytically and prove some facts, which reveal fundamental properties of the model and generally lead to existence of solutions of optimal control problems, based on this model.

1 Введение

Доклад посвящён одной из моделей хронического миелолейкоза, управляемый вариант которой был предложен в статьях зарубежных авторов. Основной задачей доклада является предоставление теоретической базы для изучения данной модели методами оптимального управления.

2 Описание модели

Здесь и далее мы следуем обозначениям из статей [1,2]. За $C(t)$ обозначена концентрация раковых клеток, $T_n(t)$, $T_e(t)$ обозначают соответственно концентрации «необученных» и «эффекторных» Т-лимфоцитов. Закон изменения этих концентраций во времени выражается следующей системой дифференциальных уравнений:

$$\begin{aligned}\frac{dT_n}{dt} &= s_n - u_2(t)d_n T_n - k_n T_n \left(\frac{C}{C + \eta} \right) \\ \frac{dT_e}{dt} &= \alpha_n k_n T_n \left(\frac{C}{C + \eta} \right) + \alpha_e T_e \left(\frac{C}{C + \eta} \right) - \\ &\quad - u_2(t)d_e T_e - \gamma_e C T_e \\ \frac{dC}{dt} &= (1 - u_1(t))r_c \ln \left(\frac{C_{\max}}{C} \right) - u_2(t)d_c C - \gamma_c C T_e\end{aligned}$$

Считается, что начальные условия $T_n(0) > 0$, $T_e(0) > 0$, и, кроме того, $0 < C(0) \leq C_{\max}$. Функции $u_1(t)$ и $u_2(t)$ используются для моделирования двух различных типов лекарственных препаратов, причём случай $u_1(t) \equiv 0, u_2(t) \equiv 1$ означает отсутствие какой-либо терапии. Все параметры, обозначенные строчными латинскими и греческими буквами считаются известными и положительными, их значения выбираются в процессе идентификации модели.

В данной модели $u_1(t)$ и $u_2(t)$ – управляющие параметры, которые соответствуют двум различным типам терапии. $u_1(t)$ описывает воздействие препарата, специфичного к раковым клеткам (иматиниб в оригинальной статье), а $u_2(t)$ – воздействие лекарств широкого цитотоксического спектра действия (цитарабин). Значения управляющих функций ограничены: $0 \leq m_1 \leq u_1 \leq M_1 < 1$, $1 \leq m_2 \leq u_2 \leq M_2$, $M_2 > 1$. Решения системы рассматриваются в области $C > 0$, управления $u_1(t)$ и $u_2(t)$ $t > 0$ – измеримые по Лебегу функции.

3 Основные задачи

Приведённая модель допускает постановку нескольких качественных вопросов, важных для применения в дальнейшем результатов теории оптимального управления:

1. Положительность решений модели при положительных начальных условиях и любых допустимых управлениях
2. Оценки на компоненты решения системы при положительных начальных условиях и любых допустимых управлениях
3. Продолжимость решений системы на $[0, +\infty)$ для положительных начальных условий и любых допустимых управлений.

4 Теоретические результаты

Основными теоретическим результатом доклада являются несколько утверждений, опирающихся на

аппарат дифференциальных неравенств и обыкновенных дифференциальных уравнений. В частности, было доказано, что при $T_n(0) > 0$, $T_e(0) > 0$ и $0 < C(0) \leq C_{\max}$ при любых допустимых управлениях решения системы продолжимы на $[0, +\infty)$ и верны следующие оценки

$$0 \leq C(t) \leq C_{\max}$$

$$0 < T_n(t) \leq T_n(0) + s_n t$$

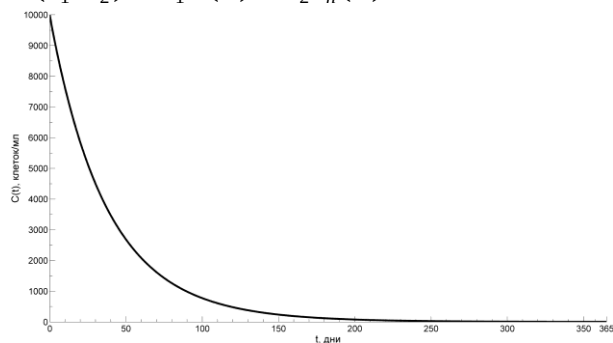
$$0 < T_e(t) \leq e^{\alpha_e t} \left(T_e(0) + \int_0^t e^{-\alpha_e s} (T_n(0) + s_n s) ds \right)$$

5 Численное моделирование

Для иллюстрации возможных постановок задач было проведено несколько расчётов с различными типами функционалов. Параметры для моделирования были взяты из статей [1,2]. В отличие от оригинальных авторов, был взят более длинный временной промежуток ($T=365$), другой вид интегральной части функционала и рассмотрен отдельно случай терминального функционала.

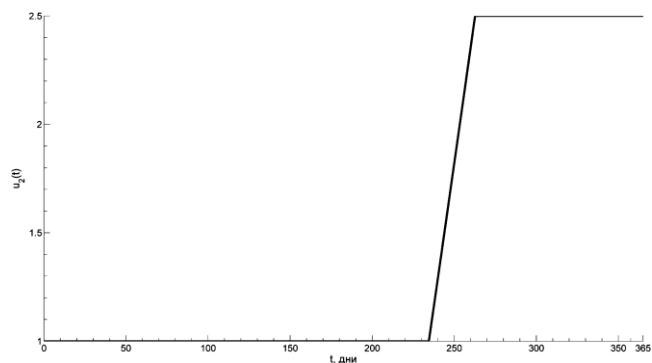
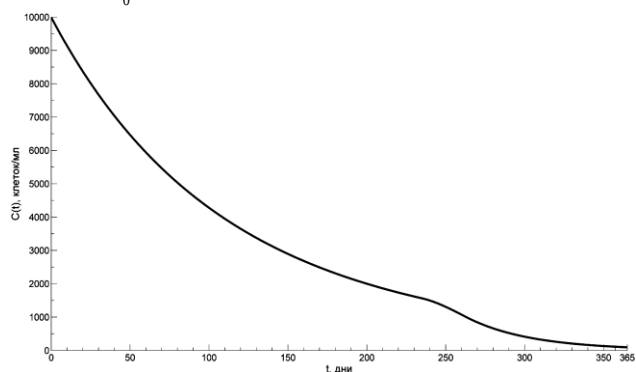
5.1 Терминальный функционал

$$J(u_1, u_2) = B_1 C(T) - B_2 T_n(T) \rightarrow \min$$



5.2 Смешанный функционал

$$J(u_1, u_2) = \int_0^T (B_1 u_1(t) + B_2 u_2(t)) dt + B_3 C(T) - B_4 T_n(T) \rightarrow \min$$



Список литературы

- [1] H. Moore, N.K. Li. A mathematical model for chronic myelogenous leukemia (CML) and t cell interaction. *Journal of Theoretical Biology*, 2004, **227 (4)**, 513-523.
- [2] S. Nanda, H. Moore, S. Lenhart. Optimal control of treatment in a mathematical model of chronic myelogenous leukemia. *Mathematical Biosciences*, 2007, **210 (1)**, 143-156.
- [3] F. Walter. *Differential and Integral Inequalities*, 1970
- [4] Э.Б. Ли, Л. Маркус. *Основы теории оптимального управления*, 1972
- [5] Л.С. Понтрягин, В.Г. Болтянский, Р.В. Гамкрелидзе, Е.Ф. Мищенко. *Математическая теория оптимальных процессов*, 1983
- [6] Ф. Хартман. *Обыкновенные дифференциальные уравнения*, 1970

Стабилометрическая модель поддержания равновесия при отсутствии целенаправленного движения

Монтиле А.И.⁽¹⁾, Давыдов О.Д.⁽²⁾, Монтиле А.А.⁽³⁾, Марчук Ю.В.⁽⁴⁾

Уральский государственный лесотехнический университет⁽¹⁾, Екатеринбург, Россия,

Уральский научно-исследовательский институт травматологии и ортопедии

им. П.Д. Чаклина⁽²⁾, Екатеринбург, Россия

Ботанический сад УрО РАН⁽³⁾, Екатеринбург, Россия

Научно-практический центр «Бонум»⁽⁴⁾, Екатеринбург, Россия

amontile@gmail.com, davod09@yandex.ru, org17@mail.ru, yura-mak@yandex.ru

The formal model of stable equilibrium, based on a new system of indicators for stabilometric research, is offered. These indicators and the model allowed us to develop new methods of diagnosis, differential diagnosis and evaluation of the effectiveness of therapeutic interventions.

Введение

Актуальность рассматриваемой тематики, на фундаментальном уровне относящейся к разработке математических моделей физиологии движений [1,2], обусловлена в первую очередь потребностями клинической практики, связанными с потенциальной возможностью использовать результаты стабилометрических исследований для решения задач диагностики, дифференциальной диагностики и оценки эффективности лечебных мероприятий. Проблема заключается в невозможности решения вышеперечисленных задач на базе единой методики [2] обработки данных стабилометрических тестов, хотя сам факт наличия нарушений «функции равновесия» в настоящее время успешно выявляется [4]. Применительно к травматологии проблема усугубляется: поражаются структурные и функциональные элементы костно-суставной, мышечной и нервной систем, при этом диагностика травм предполагает различение патологических процессов и определение степени влияния каждого из повреждений на состояние человека.

Объекты и методы

Объекты исследований - пациенты травматолого-ортопедического и неврологического профиля, а также условно здоровые («норма»). Предмет исследований - способность сохранять баланс тела, изучение которой является центральной задачей постурологии [2]. Измерения проводятся с помощью стабилометрической платформы (АПК «Стабилан 01-2»). С заданной частотой измеряются значения фронтальной (ось X) и сагиттальной (ось Y) координат центра давления тела человека на

платформу стабилографа.

Авторами разработан дополнительный к системе базовых и векторных показателей [3] набор признаков, основанный, во-первых, на учете дискретности процессов (фактически выявленной на всех уровнях организации движения [1,2]), и, во-вторых, на смещении относительно кинематической модели Берн-штейна - Гельфанда [1] акцента моделирования с пары: мышечная и костно-суставная система (система регулирования ограничивает количество степеней свободы); на пару: мышечная система – нервная система (костно-суставная система рассматривается как «источник» ограничений).

С помощью специализированной программы определяются: временные интервалы движения с постоянной скоростью по фронтальной и сагиттальной координатам, и на плоскости (для статокинезиограммы в целом); расстояния и углы, проходимые за каждый из интервалов; величины мгновенных изменений скоростей на границах интервалов; количества интервалов каждой продолжительности с конкретной скоростью. Значения признаков заносятся в базу данных и передаются отдельно или в составе формируемых выборок для дальнейшей обработки в статистические пакеты или пакет обработки биометрических данных.

Результаты и обсуждение

Применение новой системы показателей при клинической апробации обеспечило решение всех предложенных практических задач, подробно рассматриваемых в докладе. В большинстве случаев использовалось графическое отображение различных двумерных сечений пространства признаков, что обеспечило наглядность и понятность предложенных методик для врачей. Каждая из визуализаций для трех основных сечений (скорость - V , интервал времени - Δt ; скорость, длина интервала движения с постоянной скоростью - L ; скорость, мгновенное изменение скорости) структурно одинакова для всех обследованных пациентов при явно стохастическом характере исходных стабilo-грамм и статокинезиограммы.

Параметрические (и очевидные на графиках) отличия, связанные со значениями признаков по осям в основных сечениях, обеспечили решение приведенного во введении перечня задач в конкретных предметных контекстах. Пример одного из сечений $(V, \Delta t)$ приведен на рис. 1.

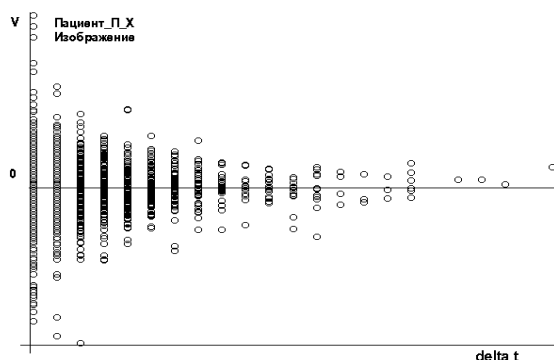


Рис.1. Сечение $(V, \Delta t)$ для фронтальной стабилограммы одного из пациентов.

В общем случае уменьшение количества дискретов по оси Δt свидетельствует о функциональных нарушениях всей системы регуляции; отсутствие дискретов и не разовые выбросы за сечение гиперболы по уровням – о нарушении регуляции в контуре с периодом цикла Δt ; отклонения от гиперболы внутрь – о патологии опорно-двигательного аппарата.

Важным, как с точки зрения формального анализа, так и с точки зрения физиологической интерпретации результатов, является определение для каждой точки бинарного сечения пространства признаков очевидного отображения в другое сечение. Так, например, $V \cdot \Delta t = L$, откуда следует, что гипербола, ограничивающая сверху и снизу множество точек на рис.1, непосредственно определяется расстоянием до границы зафиксированного на уровне синергий [1] образа области опоры.

Одним из существенных, по мнению авторов, результатов, стало выявление трех пространственных конфигураций, связанных с частотами появления пар признаков. Основная из которых (см. рис. 2), характеризует частоту – μ появления интервалов неизменного движения определенной длительности Δt с конкретной скоростью V .

Выявленная зависимость может быть аппроксимирована следующим образом:

$\mu(V, \Delta t) = k_1 \cdot \exp(-k_2 \cdot (V \cdot \Delta t)^2) / \Delta t$, где k_1 и k_2 определяются индивидуально. Из ряда соображений следует, что k_1 и k_2 взаимосвязаны. При подтверждении этой гипотезы зависимость при-

нимает вид: $\mu(V, \Delta t) = k \cdot \exp(-2 \cdot \pi^2 \cdot k^4 \cdot (V \cdot \Delta t)^2) / \Delta t$, где k определяется как функция расстояния от центра площади опоры пациента до ее границы.

Использование функции в качестве модели идеального поддержания равновесия при отсутствии целенаправленного движения снимает неопределенность понятия «нормы».

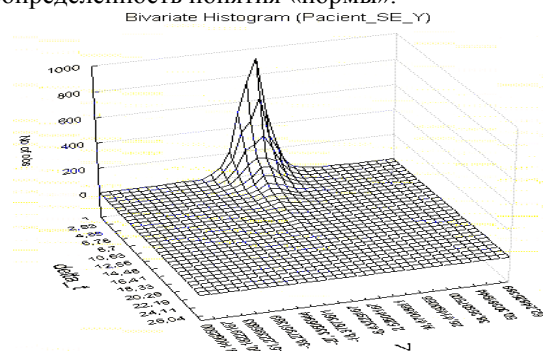


Рис.2. Поверхность гистограммы для $(V, \Delta t)$

Стабилометрическая модель используется для верификации динамической модели переходов из точки в точку в пространстве признаков (процесс без последействия). Проверка гипотез о виде зависимости интервала Δt , изменения направления и величины скорости в момент t , от положения относительно фактических границ площади опоры, направления и величины скорости в предыдущий момент $t - \Delta t$, выявила необходимость переопределений фазового пространства (перехода от одного сечения пространства признаков к другому) при моделировании различных патологий, в связи с различными способами адаптации организма к повреждениям: параметрической и структурной.

Список литературы

- [1] Бернштейн Н.А. Физиология движения и активность. – М: Наука, 1990
- [2] Гаже П.М., Вебер Б. Постурология. Регуляция и нарушения равновесия тела человека. – СПб.: Изд. Дом СПбМАПР, 2008
- [3] Скворцов Д.В. Клинический анализ движений, стабилометрия. – М.: Антидор, 2000
- [4] Усачёв В.И., Говорун М.И., Голованов А.Е., Кузнецов М.С. Динамическая стабилизация вертикального положения тела человека // Известия Южного федерального университета. Технические науки. 2010. Т. 110. № 9. С. 164-169.

Математическое моделирование эволюционных процессов в микробных популяциях на протоке

Апонин Ю.М., Апонина Е.А.

Институт математических проблем биологии РАН
yuma@impb.psn.ru

A historical sketch of the investigations in mathematical modeling of the evolution of microbial populations is given. A general model of microbial evolution in continuous culture is suggested. The method of parametric and phase portraits is used for investigation of some special cases of the general model.

Идея эволюции старше самой науки. Зародившись в глубокой древности, эта идея, в различных вариантах, проходит через всю историю человеческой цивилизации. Утвердившись в биологии XIX в., эволюционная идея в наши дни проникает во все области знания, становится парадигмой современной науки, проявляясь как концепция глобального эволюционизма или универсальной эволюции: всё течёт, всё в движении возникает, развивается, а затем разрушается и исчезает, причём результаты эволюционных процессов могут быть поняты не из их мгновенного состояния, а лишь на основе предыстории их возникновения и развития [13].

Однако в биологии и сегодня эволюционное учение проявляется как теория со «многими лицами», вызывая многочисленные обсуждения и споры. Примером является дискуссия на тему “Существует ли естественный отбор?”, представленная в недавно опубликованном сборнике [9]. Одна из причин этих дискуссий, видимо, связана с недостаточностью исследований экспериментальных моделей эволюции. Такие модели можно конструировать в лаборатории, например, на основе взаимодействующих бактериальных популяций и микробных сообществ [10, 11, 12].

Бактерии – самые древние обитатели Земли являлись главными конструкторами биосферы в течение первых двух третей её существования и остаются одним из основных факторов биосферной эволюции [8]. Идея использования популяций микроорганизмов как экспериментальных моделей для изучения закономерностей биологической эволюции впервые высказана ещё в 1900 г. голландским исследователем Бейеринком. Однако систематическое изучение эволюции микробных популяций в контролируемых лабораторных условиях началось лишь в середине XX в. после того

как Моно, Новик и Сцилард разработали математическую теорию проточного культивирования микроорганизмов.

Системы проточного культивирования микроорганизмов с первых лет после их математического обоснования стали использоваться как лабораторные “эволюционные машины”, удобные для изучения закономерностей эволюции как отдельных микробных популяций, так и сложных микробных сообществ с различными типами межпопуляционных отношений. При этом вместе с экспериментальным моделированием развивалось и математическое моделирование эволюционных процессов [10, 11, 12].

Одна из первых математических моделей эволюции бактериальных популяций, в которых учитывается мутационная изменчивость бактерий, построена в 1957 г. Мозером [1]. При построении модели Мозер основывался на результатах экспериментальной работы Новика и Сциларда [2]. В этой работе при хемостатном культивировании бактерий *E. coli* наблюдались следовавшие друг за другом “шаги эволюции” в виде серии перестроек генетической структуры популяции, вызванных появлением в культуре новых быстрорастущих мутантов. Предложенная Мозером математическая теория хорошо описывала полученные Новиком и Сцилардом экспериментальные данные. Эта теория получила развитие в ряде последующих работ, в которых предлагались её обобщения, применимые не только к хемостату, но и к другим системам проточного культивирования [10, 11, 12].

Однако быстрое развитие молекулярной генетики в течение 1960 – 1980 годов привело к открытию немутационной изменчивости бактериального генома, связанной с существованием подвижных генетических элементов в клетках бактерий (вирусной ДНК, плазмид, транспозонов, IS – элементов и др.). Оказалось, что геном бактериальной клетки может изменяться вследствие не столько мутаций генов, сколько утраты или добавления генетического материала. Например, утрачиваются гены, содержащиеся в плазмидах, при появлении бесплазмидных дочерних клеток после деления материнской клетки, содержащей плазмиды. С другой стороны, число генов в бактериальной клетке может увеличиваться за счёт конъюгационного переноса плазмид или вследствие

автономной репликации плазмид, существенно повышающей их копиюность в клетке. Адаптационная пластичность клеточного генома обеспечивается также мобильными IS – элементами, которые при своих перемещениях изменяют структуру генома, повышают частоту мутаций некоторых генов и влияют на механизмы регуляции их активности.

Многие математические модели эволюции микробных популяций на протоке, в которых наряду с мутациями учитываются и немутационная изменчивость генома (см., например, работы [6, 7] и указанную там литературу), представляют собой системы ОДУ следующего вида:

$$\begin{cases} \dot{s} = D(s_0 - s) - \sum_{i=1}^n l_i \mu_i(s, x) x_i \\ \dot{x}_i = [\mu_i(s, x) - D] x_i + f_i(s, x), \quad i = 1, \dots, n. \end{cases} \quad (1)$$

Здесь s и s_0 – концентрации лимитирующего субстрата в ферментёре и в свежей питательной среде, поступающей в ферментёр, $x = (x_1, \dots, x_n)$ – вектор концентраций различных генетических вариантов микробной популяции, присутствующих в ферментёре, D – удельная скорость протока, l_i – стехиометрический коэффициент, характеризующий пищевые потребности субпопуляции генетического варианта i , а её удельная скорость роста обозначена как μ_i . Функции $f_1, \dots, f_n$ характеризуют скорости процессов генетической изменчивости микробной популяции, например, мутационных потоков.

Система ОДУ (1) представляет собой общую математическую модель микроэволюции микробной популяции на протоке, т.к. функции μ_i, f_i не заданы явно в виде формул, а подчиняются лишь некоторым весьма общим условиям (см. [7]), допускающим широкий выбор конкретного вида этих функций. Выбирая конкретный вид функций μ_i, f_i , можно получить все известные математические модели микроэволюции при длительном хемотратном культивировании микроорганизмов (см., например, [1, 3-6, 10-12]).

Для общей модели (1) построена функция Ляпунова, с помощью которой установлено существование в фазовом пространстве положительно инвариантного компактного полиэдрального множества. Дано аналитическое описание этого полиэдра как множества решений конечной системы линейных неравенств. Доказано, что этот полиэдр является устойчивым, глобально притягивающим множеством и, следовательно, содержит в себе траектории всех возможных в модели установившихся процессов эволюции.

Методом параметрических и фазовых портретов проведено исследование некоторых частных случаев общей модели (1), в которых функции μ_i, f_i заданы

явно в виде аналитических формул. На сравнительно простой модели показано, что горизонтальный перенос генетического материала в микробной популяции порождает целый ряд эволюционно значимых эффектов. К ним относятся такие эффекты, как множественность аттракторов и многовариантность путей эволюции, повторяющиеся ускорения и замедления эволюционных процессов, их незатухающие колебания при постоянных условиях культивирования, неоднократные исчезновения и затем вновь появление некоторых биотехнологически ценных признаков в ходе эволюции, быстрые перестройки генетической структуры микробной популяции в течение нескольких генераций – шаги эволюции и др. Несомненно, понимание и анализ этих эффектов необходимы для развития биотехнологий на основе использования микробных популяций и сообществ.

Список литературы

- [1] Moser H. *Gold Spring Harbor Symposia on quantitative biology*, 1957, **22**, 121-137.
- [2] Novick A., Szilard L. *Proceedings Nat. Acad. Sci. USA*, 1950, **36**, 708-719.
- [3] Stewart F.M., Levin B.R. *Genetics*, 1977, **87**(2), 209-228.
- [4] Апонин Ю.М. *Популяционная динамика бактериальных плазмид в условиях хемотратного культивирования*. Препринт. Пушкино: ОНТИ НЦБИ АН СССР, 1982.
- [5] Апонин Ю.М., Аполина Е.А. В сб.: *Исследования по математической биологии*. Пушкино: ОНТИ ПНЦ РАН, 1996, 32-48.
- [6] Апонин Ю.М., Аполина Е.А. *Биофизика*, 2008, **53**(4), 638-645.
- [7] Апонин Ю.М., Аполина Е.А. *Компьютерные исследования и моделирование*, 2011, **3**(2), 177-190.
- [8] Заварзин Г.А. *Лекции по природоведческой микробиологии*, М.: Наука, 2003.
- [9] *Идея эволюции в биологии и культуре*. Сб., ред. Баксанский О.Е., Лисеев И.К., М.: Канон⁺ РООИ «Реабилитация», 2011.
- [10] Печуркин Н.С., Терсков И.А. *Анализ кинетики роста и эволюции микробных популяций*, Новосибирск: Наука, 1975.
- [11] Печуркин Н.С. *Популяционная микробиология*, Новосибирск: Наука, 1978.
- [12] Печуркин Н.С., Брильков А.В., Марченкова Т.В. *Популяционные аспекты биотехнологии*, Новосибирск: Наука, 1990.
- [13] Эбелинг В., Энгель А., Файстель Р. *Физика процессов эволюции*, М.: Эдиториал УРСС, 2001.

Динамика и стабильность фитопланктонных сообществ

Абакумов А.И.^(1,4), Силкин В.А.⁽²⁾, Паутова Л.А.⁽³⁾

Институт автоматики 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⁾,
Дальневосточный федеральный университет, Владивосток⁽⁴⁾
abakumov@iacp.dvo.ru, vsilkin@mail.ru, larisapautova@yahoo.com

The model of a biomass dynamics for a competitive community of microorganisms is investigated. The model is applied to research of phytoplankton in north-east part of the Black Sea. Situations of a dominating of species in the community and situations of species invasion are studied.

1 Введение

Исследуемая модель с теми или иными вариациями уже использовалась для теоретического изучения сообществ микроорганизмов, а также при анализе непрерывных культур фитопланктона [3, 6]. Модель применена к изучению фитопланктонных сообществ в море. Зона обитания фитопланктона летом ограничена снизу термоклином, и если она рассматривается в небольшом районе поверхности моря, то может быть описана подобной «точечной» моделью.

2 Описание модели конкурентного сообщества

Содержание элементов минерального питания группы j в клетке вида i обозначим q_{ij} . Этот параметр называют клеточной квотой. Поглощение элементов минерального питания водорослями фитопланктона является функцией их концентрации z_j в среде осуществляется с удельной скоростью $v_{ij}(z_j, q_{ij})$, а рост растительной биомассы может происходить с потенциальной удельной скоростью $\mu_{ij}(q_{ij})$ в зависимости от вектора $z = (z_j)_{j=1}^n$ концентрации минеральных веществ во внешней среде и матрицы $q = (q_{ij})_{i,j=1}^{m,n}$ содержания питательных веществ в клетках растений. Принципиальным здесь является использование уравнения Друпа для определения зависимости $\mu_{ij}(q_{ij})$ [1].

Модель динамики масс системы имеет вид [4]:

$$\begin{cases} \frac{dy_i}{dt} = (\mu_i(q_i) - D)y_i \\ \frac{dz_j}{dt} = D(z_j^{(0)} - z_j) - \sum_{i=1}^m v_{ij}(z_j, q_{ij})y_i \\ \frac{dq_{ij}}{dt} = v_{ij}(z_j, q_{ij}) - \mu_i(q_i) \cdot q_{ij} \end{cases}$$

для $i = 1, \dots, m; j = 1, \dots, n$.

Реализуемая удельная скорость μ_i роста биомассы вида является непрерывной функцией потенциальных скоростей $(\mu_{ij})_{j=1}^n$. Под q_i понимается вектор $q_i = (q_{ij})_{j=1}^n$. В нашем случае эта зависимость определяется на основе принципа «узкого места» Либиха.

В модели изучены общие свойства решений, построен алгоритм поиска равновесных стационарных решений, получены условия их устойчивости.

3 Моделирование для Черного моря

В последние два десятилетия основные изменения в структуре фитопланктонного сообщества и, как следствие, его продуктивности происходят в системе «диатомовые – кокколитофорида».

Среди диатомовых доминирующими видами в зависимости от сезона и года становятся такие виды как *Pseudo-nitzschia pseudodelicatissima*, *Skeletonema costatum*, *Chaetoceros curvisetus*, *Proboscia alata*, кокколитофорида представлены практически одним видом *Emiliania huxleyi* [5].

С 2005 г в северо-восточной части Черного моря были зафиксированы новые для Черного моря виды фитопланктона *Chaetoceros trondsenii* и *Chaetoceros minimus*, относящиеся к группе диатомей [2].

Вычислительные эксперименты с моделью показали, что эти виды водорослей могут сосуществовать с другими в верхнем перемешанном слое при относительно низких скоростях обмена водной массы – около $0,1 \text{ сут}^{-1}$. Начальные плотности биомасс вселенцев в расчетах

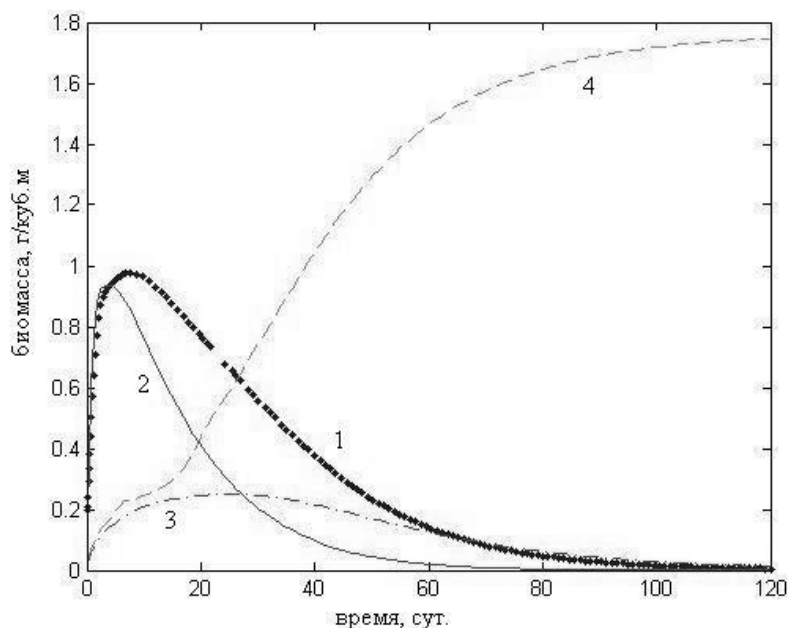


Рис. 1. Сукцессия видов при $N/P=1$. Виды пронумерованы в порядке упоминания: 1 - *Pseudo-nitzschia pseudodelicatissima*, 2 - *Emiliana huxleyi*, 3 - *Caetoceros trondsenii*, 4 - *Chaetoceros minimus*.

Список литературы

предполагались на один-два порядка меньшими плотностей видов-аборигенов.

Оказалось, что эти виды особенно чувствительны к соотношению азота и фосфора (N/P) в термоклине. При N/P больше 8 диатомея *Pseudo-nitzschia pseudodelicatissima* вначале сосуществует с кокколитофорицей *Emiliana huxleyi*, а затем успешно завоевывает пространство. При соотношении меньше 8, но больше 4 наблюдается более сложная сукцессия, где вначале доминируют кокколитофорида и диатомея, позже вид-вселенец *Chaetoceros trondsenii* начинает вытеснять другие виды. Наконец, при N/P меньше 4 поочередно доминируют *P. pseudodelicatissima* и *E. huxleyi*, внедряется *C. trondsenii*, но в итоге всех вытесняет вид-вселенец *Chaetoceros minimus* (рис. 1). Времена этих сукцессий зависят от начальных плотностей видов, но качественно картины от начального состояния планктона не зависят.

Таким образом, существуют условия, при которых закрепление и даже доминирование видов-вселенцев становится реальным. Это может привести к перестройкам экосистемы в целом, последствия которых сейчас трудно предсказать.

- [1] M.R. Droop. The nutrient status of algal cells in continuous culture. *J. Mar. Biol. Assoc. U. K.*, 1974, **54**, 825-855.
- [2] V.A. Silkin, A.I. Abakumov, L.A. Pautova, A.S. Mikaelyan, V.K. Chasovnikov, T.A. Lukashova. Co-existence of non-native and the Black sea phytoplankton species. Invasion hypotheses discussion. *Russian Journal of Biological Invasions*, 2011, **2**(4), 256-264.
- [3] D. Tilman. Resource competition between planktonic algae: an experimental and theoretical approach. *Ecology*, 1977, **58**, 338-348
- [4] А.И. Абакумов, В.А. Силкин, Л.А. Паутова, С.Я. Пак. Третья международная конференция "Математическая биология и биоинформатика", тезисы докладов. Пушино, 2010, 266-267.
- [5] Л.А. Паутова, А.С. Микаэлян, В.А. Силкин. Структура планктонных фитоценов шельфовых вод северо-восточной части Черного моря в период массового развития *Emiliana huxleyi* в 2002 - 2005 гг. *Океанология*, 2007, **47**(3). 408-417.
- [6] В.А. Силкин, К.М. Хайлов. Биоэкологические механизмы управления в аквакультуре. Л.: Наука, 1988.

Исследование асимптотического поведения решений некоторых моделей эпидемических процессов

Перцев Н. В.

Омский филиал Федерального государственного бюджетного учреждения науки
Института математики им. С. Л. Соболева СО РАН
homlab@ya.ru

We consider the problem of asymptotic behavior ($t \rightarrow +\infty$) for the solutions of the certain class of mathematical models of epidemic processes.

1 Введение

В работе изучается устойчивость положений равновесия математических моделей эпидемических процессов в форме систем нелинейных дифференциальных уравнений с последствием и высоко-размерной системы нелинейных обыкновенных дифференциальных уравнений. Модель первого типа представляет собой SIRS модель с иммиграцией, основанной на работах [1–3]. Приведено несколько вариантов уравнений модели, различающихся законами распределения времени пребывания индивидуумов в стадиях I и R инфекционного процесса. В основу модели второго типа положена работа [4], посвященная изучению динамики распространения ВИЧ-инфекции в России с учетом фактора социальной дезадаптации индивидуумов. Специфика модели второго типа обусловлена описанием неоднородности популяции индивидуумов в форме произвольного числа групп, в которые могут входить индивидуумы, а также осуществлять переходы между ними.

Ниже представлены уравнения моделей и приведены условия асимптотической устойчивости положений равновесия, при которых численности больных индивидуумов равны нулю (отсутствие инфекции в популяции).

2 SIRS модель

Рассмотрим одну из версий SIRS модели с иммиграцией, описывающую динамику следующих переменных: $x(t)$, $y(t)$, $z(t)$ – численности в момент времени $t \geq 0$ групп соответственно восприимчивых индивидуумов (S), больных индивидуумов (I), индивидуумов, которые приобрели иммунитет после выздоровления (R). Предполагается, что скорость появления больных индивидуумов пропорциональна числу контактов индивидуумов группы S и группы I. Группа S восприимчивых индивидуумов может пополняться

за счет внешнего источника со скоростью f .

Используя стандартный подход, запишем систему уравнений модели:

$$\begin{aligned}\frac{dx(t)}{dt} &= f - \delta x(t) - \beta x(t)y(t) + \mu z(t), \\ \frac{dy(t)}{dt} &= \beta x(t)y(t) - (\gamma + \delta + \eta)y(t), \quad t > 0, \\ \frac{dz(t)}{dt} &= \eta y(t) - (\delta + \mu)z(t),\end{aligned}\quad (1)$$

$$\frac{dz(t)}{dt} = \eta y(t) - (\delta + \mu)z(t),$$

$$x(0) = x_0, \quad y(0) = y_0, \quad z(0) = z_0.$$

В системе (1) все параметры положительны, начальные данные неотрицательны.

Обратимся к модификации модели (1), позволяющей учитывать длительность заболевания τ_y и продолжительность сохранения иммунитета τ_z для отдельных индивидуумов. Систему уравнений модели запишем в виде:

$$\begin{aligned}\frac{dx(t)}{dt} &= f - \delta x(t) - \beta x(t)y(t) + \rho_z(t), \\ \frac{dy(t)}{dt} &= \beta x(t)y(t) - (\gamma + \delta)y(t) - \rho_y(t), \quad t > 0, \\ \frac{dz(t)}{dt} &= \rho_y(t) - \delta z(t) - \rho_z(t),\end{aligned}\quad (2)$$

$$\frac{dz(t)}{dt} = \rho_y(t) - \delta z(t) - \rho_z(t),$$

$$x(t) = x_0(t), \quad y(t) = y_0(t), \quad z(t) = z_0(t), \quad t \leq 0.$$

В модели (2) функции $\rho_y(t)$ и $\rho_z(t)$ задают скорости переходов индивидуумов между соответствующими группами; начальные данные неотрицательны. Далее приведены два варианта модели (2), отражающие различные формы задания функций $\rho_y(t)$ и $\rho_z(t)$.

В первом варианте модели (2) предполагается, что τ_y и τ_z – постоянные величины: $\tau_y = \omega_1 > 0$, $\tau_z = \omega_2 > 0$. В этом случае:

$$\rho_y(t) = e^{-(\gamma+\delta)\omega_1} \beta x(t-\omega_1)y(t-\omega_1),$$

$$\rho_z(t) = e^{-\delta\omega_2 - (\gamma+\delta)\omega_1} \beta x(t-\omega_1-\omega_2)y(t-\omega_1-\omega_2).$$

Во втором варианте модели (2) распределения величин τ_y и τ_z задаются с помощью функций

$$r_y(s) \geq 0, \quad r_z(s) \geq 0, \quad s \geq 0,$$

таких, что

$$r_y(s) = 0, \quad s \geq \omega_1, \quad r_z(s) = 0, \quad s \geq \omega_2,$$

$$\int_0^{\infty} r_y(s) ds = 1, \quad \int_0^{\infty} r_z(s) ds = 1,$$

где $\omega_1 > 0$, $\omega_2 > 0$ – заданные константы. Тогда

$$\rho_y(t) = \int_0^{\infty} e^{-(\gamma+\delta)a} r_y(a) \beta x(t-a) y(t-a) da,$$

$$\rho_z(t) = \int_0^{\infty} e^{-\delta s} r_z(s) \rho_y(t-s) ds.$$

Обозначим через x^* , y^* , z^* неотрицательные положения равновесия моделей (1) и (2). Для каждой из этих моделей существует положение равновесия

$$x_1^* = f/\delta, \quad y_1^* = 0, \quad z_1^* = 0. \quad (3)$$

Кроме того, возможно существование еще одного положения равновесия

$$x_2^* > 0, \quad y_2^* > 0, \quad z_2^* > 0. \quad (4)$$

Положение равновесия (4) существует только в том случае, когда $x_2^* < x_1^*$.

Утверждение 1. Если для модели (1) или модели (2) существует только одно положение равновесия (3), то оно асимптотически устойчиво; если существует положение равновесия (4), то (3) не устойчиво.

3 Модель распространения ВИЧ-инфекции

Рассмотрим одну из модификаций модели распространения ВИЧ-инфекции, предложенной в [4]. Представим население некоторого региона в виде групп индивидов $A_1, \dots, A_n$, $B_1, \dots, B_n$. Полагаем, что группы $A_1, \dots, A_n$ состоят из восприимчивых к ВИЧ-инфекции индивидов, различающихся по уровню социальной дезадаптации. Группы $B_1, \dots, B_n$ используются для описания ВИЧ-инфицированных индивидов с учетом уровней социальной дезадаптации. Обозначим через $x_i = x_i(t)$, $y_i = y_i(t)$ численности индивидов групп A_i , B_i в момент времени $t \geq 0$.

Для описания динамики этих переменных используем следующую систему уравнений:

$$\begin{aligned} \frac{dx_i}{dt} &= f_i + \sum_{k=1, k \neq i}^n \gamma_{k,i} x_k - \sum_{k=1}^n \gamma_{i,k} x_i - \sum_{j=1}^n \beta_{i,j} y_j x_i, \\ \frac{dy_i}{dt} &= \sum_{k=1, k \neq i}^n \alpha_{k,i} y_k - \sum_{k=1}^n \alpha_{i,k} y_i + \sum_{j=1}^n \beta_{i,j} y_j x_i, \quad t > 0, \\ x_i(0) &= x_i^0, \quad y_i(0) = y_i^0, \quad i = 1, \dots, n. \end{aligned} \quad (5)$$

В системе (5) начальные данные и параметры неотрицательны. Для каждого $i = 1, \dots, n$ $\gamma_{i,i} > 0$, $\alpha_{i,i} > 0$ и $\beta_{i,i} > 0$ при

некоторых $j = 1, \dots, n$. Кроме того, $\sum_{i=1}^n f_i > 0$.

Пусть x_i^* , y_i^* – неотрицательные положения равновесия модели (5), $i = 1, \dots, n$. Одно из этих положений равновесия имеет вид

$$x_i^* = u_i, \quad y_i^* = 0, \quad i = 1, \dots, n, \quad (6)$$

где $u_1 \geq 0, \dots, u_n \geq 0$ однозначно находятся из системы

$$f_i + \sum_{k=1, k \neq i}^n \gamma_{k,i} u_k - \sum_{k=1}^n \gamma_{i,k} u_i = 0, \quad i = 1, \dots, n.$$

Определим две матрицы:

$$D = \text{diag}(d_{1,1}, \dots, d_{n,n}), \quad G = (g_{i,j}),$$

$$d_{i,i} = \sum_{k=1}^n \alpha_{i,k} > 0, \quad g_{i,i} = u_i \beta_{i,i} \geq 0, \quad i = 1, \dots, n,$$

$$g_{i,j} = u_i \beta_{i,j} + \alpha_{j,i} \geq 0, \quad i, j = 1, \dots, n, \quad i \neq j.$$

Положим $H = D - G$. Матрица $H = (h_{i,j})$ является матрицей специального вида, так как ее элементы $h_{i,j} \leq 0$, $i, j = 1, \dots, n$, $i \neq j$.

Утверждение 2. Положение равновесия (6) модели (5) асимптотически устойчиво тогда и только тогда, когда H является невырожденной М-матрицей.

Одним из следствий утверждений 1, 2 является следующий результат. Пусть параметр f моделей (1), (2) и параметры $f_1, \dots, f_n$ модели (5) имеют достаточно малые значения. Тогда положения равновесия (3) и (6) асимптотически устойчивы. Малость значений этих параметров указывает на невысокие темпы пополнения групп восприимчивых индивидов, подверженных инфицированию и развитию заболевания. Численность групп таких индивидов будет относительно невысока, что обеспечит затухание инфекционного процесса.

Работа поддержана Сибирским отделением РАН (интеграционный проект № 80, 2012 – 2014 г.г.).

Список литературы

- [1] S. Busenberg, K. Cooke. The Effect of Integral Conditions in Certain Equations Modelling Epidemics and Population Growth // J. Math. Biol. 1980. N.10. P.13-32.
- [2] K. Cooke, J. York. Some equations Modelling Growth Processes and Gonorrhea Epidemics // Math. Biosci. 1973. V.16. P.75-101.
- [3] H.W. Hethcote, H.W. Stech, P. van den Driessche. Stability analysis for models of diseases without immunity // J. Math. Biol. 1981. V.13. P.185-198.
- [4] А.А. Романюха, Е.А. Носова. Модель распространения ВИЧ-инфекции в результате социальной дезадаптации // Управление большими системами. Сборник трудов ИПУ РАН. 2011. № 34. С.227-253.

Математическое моделирование клеточного цикла прокариот: согласование темпов роста объема клетки и репликации ДНК

Лихошвай В.А.^(1,2), Хлебодарова Т.М.⁽¹⁾

Институт Цитологии и Генетики СО РАН⁽¹⁾, Новосибирский Государственный Университет⁽²⁾,
likho@bionet.nsc.ru

1 Резюме

Представлены результаты исследования моделей клеточного цикла прокариот, реализующих различные стратегии регуляции инициации репликации ДНК. Показано, что решение проблемы согласования процессов роста клетки и репликации ДНК зависит от законов роста клетки. Для клеток, рост которых подчиняется экспоненциальному закону, согласование может быть достигнуто только за счет механизмов негативной регуляции инициации репликации. В моделях, реализующих линейный закон роста клетки, наблюдается автоматическая подгонка темпов роста клетки под заданный темп репликации. В этом смысле, в клетках, растущих по линейному закону, проблемы согласования процессов роста и репликации не существует.

2 Введение

Несмотря на прогресс в области исследования механизмов репликации генома у разных видов прокариот, проблема согласования темпов роста клетки и скорости репликации генома до сих пор стоит на повестке дня [1]. Исторически этой проблеме более 40 лет, когда были разработаны первые феноменологические модели клеточного роста прокариот [2,3], в которых процессы роста клетки и репликации генома рассматривались как процессы экспоненциальные, контролируемые пороговой концентрацией активатора или репрессора. В «репрессорной» модели процессы клеточного роста и репликации ДНК были скоординированы на основе предположения, что инициация репликации ДНК осуществляется репрессором, который контролирует экспрессию собственного гена по механизму авторепрессии, и расположен недалеко от локусов инициации репликации ДНК (ORI) [2], что позволяет установить прямую зависимость между скоростью репликации и объемом клетки. Однако эта модель не получила развития, возможно потому, что к настоящему времени у прокариот не обнаружено

репрессора, который обладал бы всеми этими характеристиками [4]. В «инициаторной» модели [3] процессы клеточного роста и репликации ДНК были скоординированы на основе экспериментально установленного факта, что отношение массы клетки к количеству ORI в момент инициации репликации является величиной постоянной [5]. Этот параметр был назван «инициаторной массой» и его введение в модель позволяет имитировать процесс репликации генома при разных скоростях роста клетки. Он до сих пор успешно используется для симуляции деления клетки *E.coli* [6]. В последующие годы развитие получили именно такого типа модели. В них в явном виде вводятся параметры, связывающие каким-либо образом рост и репликацию, что снимает проблему согласования, но не решает ее.

3 Метод моделирования

Для построения моделей использован блочный подход, в соответствии с которым клеточный цикл разбит на подсистемы, описывающие процессы репликации ДНК, рост и деление клетки, а также процессы регуляции экспрессии генов, белки которых участвуют в регуляции инициации репликации генома. Для каждой подсистемы разработаны индивидуальные подмодели, в которых реализованы различные гипотезы и предположения, возникающие в связи с проблемой согласования роста клетки и репликации генома. Для описания подсистем использованы непрерывный, дискретный и стохастический подходы. Модели клеточного цикла были построены из моделей подсистем, совокупность которых определялась текущей решаемой задачей. Расчет моделей проводился методом, который идейно восходит к методу дробных шагов Яненко [7].

Было разработано восемь моделей с разными вариантами механизмов контроля инициации репликации и законами роста клетки.

4 Согласование темпов роста и репликации в экспоненциально растущих клетках

Пороговый негативный механизм регуляции репликации, предложенный в 1969 году, вполне обеспечивает решение проблемы согласования темпов репликации ДНК и роста клетки [2]. Однако авторы не рассматривали эту проблему отдельно от проблемы синхронизации актов репликации между собой. Поэтому заранее было неясно, являются ли условия, заложенные в работе [2], важными для решения только проблемы согласования. Мы полагали, что если рассматривать проблему согласования как самостоятельную цель, для ее успешного решения, некоторые, если не все, условия могут быть существенно ослаблены.

Для выяснения этих условий, была разработана модель экспоненциально растущей клетки, в которой максимально ослаблен пороговый эффект влияния концентрации репрессора на инициацию репликации и снижена чувствительность рассматриваемого механизма от дозы сайтов репликации в клетке.

Результаты расчета модели на временном интервале, в течение которого происходит больше двух тысяч циклов деления, показывают отсутствие неограниченного роста массы генома (рис.1) или объема клетки в последовательном ряду растущих и делящихся клеток.

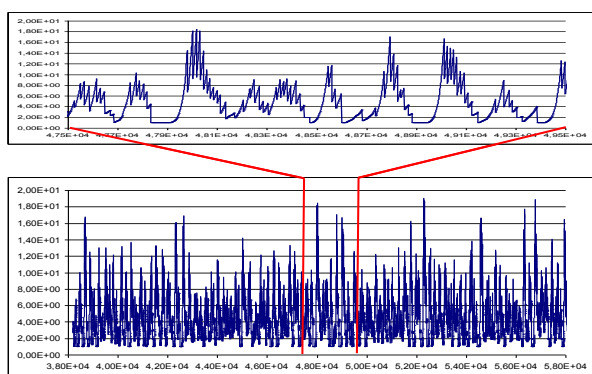


Рис. 1. Изменение массы генома (в геномных эквивалентах) в процессе роста и деления клеток, рассчитанное по простейшей «репрессорной модели».

Из этих данных следует, что контур негативной регуляции процесса инициации репликации, успешно согласует темпы экспоненциального роста объема клетки и геномной массы уже при самых минимальных предположениях. Механизм действия может быть самым простым, генетический носитель функции может располагаться в произвольном месте генома, также не накладывается никаких дополнительных условий на механизмы регуляции экспрессии гена-репрессора и стабильность белка.

Что касается «инициаторного» механизма, то в трех вариантах моделей, разработанных на этом принципе, даже с элементами негативного контроля, не удалось выявить область устойчивого

согласования темпов репликации с темпами экспоненциального роста объема клетки. Наблюдалось либо неограниченное накопление геномной массы в ряду делящихся клеток, либо прекращение инициации репликации.

5 Согласование темпов роста и репликации в клетках, растущих по линейному закону

Этот анализ был проведен на моделях «инициаторного» типа, в которых был изменен только закон роста клеток. Для всех вариантов моделей наблюдалось автоматическое согласование темпов роста объема клетки и массы генома (рис. 2).

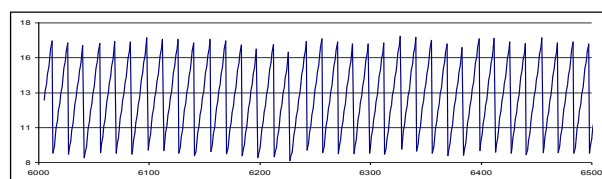


Рис.2. Изменение массы генома в ряду поколений, рассчитанное по простейшей «инициаторной» модели.

Что касается «негативных» моделей, то, как и ожидалось, замена экспоненциального закона роста клетки на линейный, не сказывалась на возможности негативного механизма согласовывать темпы роста объема клетки и репликации генома.

6 Заключение

Таким образом, в моделях с линейным законом роста клетки все механизмы контроля инициации репликации, независимо от его типа, негативный или инициаторный, являются допустимыми в том смысле, что они обеспечивают устойчивый рост и деление клеток в ряду поколений. В противоположность этому, только негативные механизмы оказались совместимыми с экспоненциальным законом роста объема клетки.

Список литературы

- [1] H. Merrikh, A.D. Grossman. *Mol Microbiol.* 2011, **82**, 434-446.
- [2] R.H. Pritchard et al. In: *Microbial Growth, Symp. Soc. Gen. Microbiol.* 1969, **19**, 263-297.
- [3] S. Cooper, C.E. Helmstetter. *J. Mol. Biol.* 1968, **31**, 619-644.
- [4] T. Katayama et al. *Nat. Rev. Microbiol.* 2010, **8**, 163-170.
- [5] W.D. Donachie. *Nature.* 1968, **219**, 1077-1079.
- [6] A. Zaritsky et al. *Microbiology.* 2011, **157**, 1876-1885.
- [7] N.N. Yanenko. In: *The method of fractional steps.* NY.: Springer-Verlag, Ed. by M. Holt, 1971.

Analysis of atomic shifts in crystallographic unrestrained refinement can indicate alternative conformations of amino acid residues

Sobolev O.V., Lunin V.Y.

Institute of Mathematical Problems of Biology of RAS, Pushchino, Russia
oleg@impb.psn.ru

Unrestrained reciprocal space refinement is stable for the vast majority of atoms when working at atomic resolution. Nevertheless geometrical restraints should be held in refinement for residues that are present in several (alternative) conformations in the crystal used for the X-ray experiment; in the other case such residues deteriorate significantly. We believe that on the contrary, large distortion of a residue in unrestrained refinement may hint the presence of alternative conformations for this residue. To get these hints in a routine way we developed two methods to analyze shifts of atomic centers resulted from several cycles of unrestrained refinement. The first is a simple diagram plotting the values of atomic shifts against the residue number. The second method use decision-making procedures for composing a list of residues that are likely to be present in alternative conformations or disordered and so should be checked thoroughly with Fourier syntheses, and included into the model with alternative conformations, when necessary.

1 Intruduction

The base of X-ray experiment is diffraction of X-rays by a protein crystal. In an ideal crystal, the content of different unit cells is assumed to be identical, and every amino acid residue has the same conformation in all copies of the unit cell. In real protein crystals different kinds of static and dynamic disorder may be present resulting in a small number (two or three) of alternative conformations (AC) for some residues. To model this situation, coordinates of several AC are included into an atomic model with partial occupancies. AC can be detected for side chains and for parts of the main chain. The modeling of AC increases the accuracy of the calculated structure factor phases and results in more precise electron density maps in the whole unit cell. Furthermore in some cases the presence of AC is related to protein function. Detection of residues that are present in AC is time-consuming and demands visual analysis of electron density maps by an experienced crystallographer.

Generally, a target function for crystallographic refinement of biological macromolecules contains two

parts. The first part of this criterion ensures correspondence between the model and X-ray data. The second part of the criterion puts restraints on model geometry. The presence of the second part compensates for the relatively low data-to-parameter ratio and stabilizes refinement. Nevertheless, well-ordered parts of the structure can be refined without stereochemical restraints at last steps of refinement if the resolution of experimental data is high enough. At the same time it was noted long ago that even at atomic resolution poorly ordered residues deteriorate significantly in unrestrained refinement while the ordered residues are stable [1].

Here we discuss whether the deterioration of a residue in high resolution unrestrained refinement can be used to detect residues those are present in AC in the studied crystal.

2 Trial unrestrained refinement (TUR) and composition of databases.

To analyze a particular model, we offer to apply the Trial Unrestrained Refinement (TUR). The TUR consists of 3 macro-cycles of unrestrained refinement with *phenix.refine* [2] following by calculation for each atom of its shift, i.e. the distance between positions of the center of the atom before and after refinement. The set of atomic shifts forms the output of the TUR. This output may be used to visualize TUR results or generate in an automated mode a list of problematic parts of the model that should be checked with a caution.

To study the mobility of atoms in unrestrained refinement and derive parameters for decision-making procedures two databases were composed corresponding to two different resolution zones. The first database corresponded to resolution between 1.1 Å and 1.2 Å (135 PDB structures); the second database corresponded to resolution better than 1.1 Å (68 PDB structures). We set limits for R-factor values of the selected models to choose high-quality structures which were studied carefully, so that the alternative conformations were assigned reliably.

The process of preparing the database consisted of three stages. The first is model preparation consist of removing AC and standard refinement. The second stage is TUR. The third stage is calculation of atomic

shifts resulted from the TUR. They were included into the databases together with information identifying the atom and some complementary information. All refinements were carried out with *phenix.refine*.

3 Visual analysis of atomic shifts

Visually, the results of TUR can be displayed as a diagram presenting atomic shifts vs. residue number (Fig.1). In this diagram each dot corresponds to one non-hydrogen atom and the dots corresponding to the same residue are grouped in a column. High columns reveal residues that had significant atomic shifts in unrestrained refinement and thus are unstable and suspected to be present in alternative conformations. The diagram can be plotted separately for main chain and side chain atoms; this can simplify the search for unstable parts of the main chain. The diagram gives a hint which residues may have alternative conformations, but the proper choice of cut-off level to select high columns is not always obvious [3].

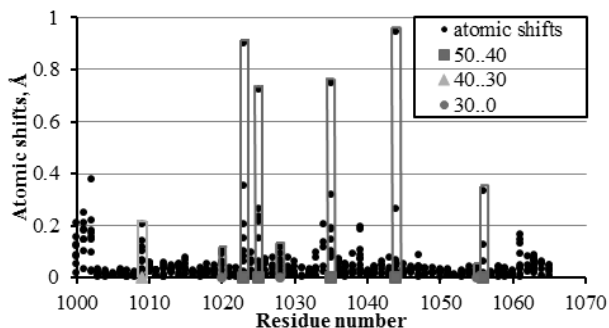


Fig.1. Atomic shifts for 1hg7 structure [4].

4 Atomic mobility in unrestrained refinement and decision-making procedures

The analysis of atomic shifts obtained in the unrestrained refinement revealed different mobility for different kinds of atoms (*e.g.* ones in main and side chains, or atoms in the core and at the surface of the molecule). Moreover, atomic mobility depends on crystallographic *R*-factor which reflects the overall quality of the model (Fig.2). These distinctions were taken into account when constructing decision making procedures.

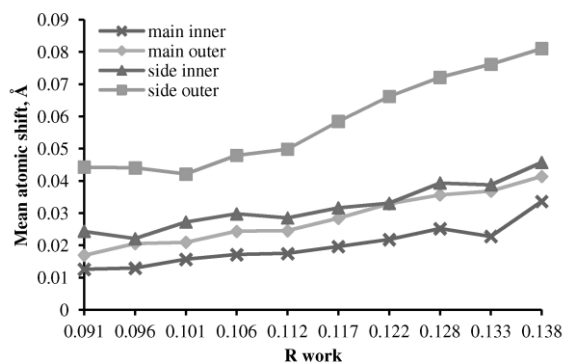


Fig.2. Atomic shifts for different atom groups.

To classify every residue as 'single conformation' (SC) or 'alternative conformations' (AC) in an

automated mode several decision making procedures were developed and tested [5]. Each procedure uses the TUR results for generating a list of residues suspected of presence in alternative conformations. The decision is made either by comparing the observed atomic shifts to a predetermined threshold, or by comparing probabilities of these shifts for residues possessing of single and multiple conformations. The threshold values and probability distributions for atomic shifts were defined differently for different kinds of atoms and for different *R*-factor bins.

The testing of automatic decision-making procedures showed that, as it was expected, the results are statistical in nature and contain errors of both types (false positive and false negative).

The five tested decision-making procedures showed near to similar prediction quality. The prediction accuracy achieved at the moment does not give a reason to decide on the advantage of any of tested procedures. Generally, the quality of prediction approximately matches results of CASP competitions devoted to "Identifying disordered regions in target proteins" [6].

5 Conclusion

Trial unrestrained refinement, followed by an analysis of atomic shifts, highlights the most unstable parts of the structure. These parts are prime candidates for the introduction of alternative conformations, and they should be tested first with electron density maps. The proposed formal decision-making procedures do not allow predicting exactly which residues are present in alternative conformation, but provide an opportunity to reduce the amount of work for the visual analysis of the maps, identifying the most likely candidates. At its current implementation the method is applicable at the resolution better than 1.2 Å. The application of the method for a worse resolution is problematic and demands further improvement.

The programs for analysis of TUR results are available from http://www.impb.ru/lmc/programs/ac_prediction/ or from the authors on request.

This work was supported by RFBR, grant № 10-04-00254-a.

References

- [1] Dauter, Z., Sierer, L.C. & Wilson, K.S. (1992). *Acta Cryst.* B48, 42-59.
- [2] Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. & Adams, P.D. (2012). *Acta Cryst.* D68, 352-367.
- [3] Sobolev O.V., Lunin V.Y. (2008). *Math. Biol. Bioinf.* 3, 50-59 (rus).
- [4] Antson, A.A., Smith, D.J., Roper, D.I., Lewis, S., Caves, L.S., Verma, C.S., Buckley, S.L., Lillford, P.J. & Hubbard, R.E. (2010). *J. Mol. Biol.* 305, 875-889.
- [5] Sobolev O.V. & Lunin V.Y. (2012). *Acta Cryst.* D68. (in press).
- [6] Noivirt-Brik, O., Prilusky, J. & Sussman, J.L. (2009). *Proteins*, 77 (Suppl. 9), 210-216.

Распознавание и анализ устойчивости структурных мотивов в глобулярных белках

Руднев В.Р.⁽¹⁾, Панкратов А.Н.^(1,2), Куликова Л.И.⁽¹⁾, Дедус Ф.Ф.^(2,1), Тихонов Д.А.⁽¹⁾,
Ефимов А.В.⁽³⁾

Институт математических проблем биологии РАН⁽¹⁾, Факультет вычислительной
математики и кибернетики МГУ имени М.В.Ломоносова⁽²⁾, Институт белка РАН⁽³⁾
volodyarv@mail.ru

Рассматривается задача распознавания структурных мотивов белков на примере α - α -углов. Предложен комбинированный подход к решению задач анализа пространственной структуры белков на основе аналитического описания основной цепи белковой глобулы и спектрального метода распознавания повторов. Нами найдено 110 α - α -углов в базе данных PDB, соответствующих заданному эталону. Методом молекулярной динамики показано, что α - α -углы как автономные структуры довольно устойчивы в водной среде.

Введение

Объем информации о белках очень велик и постоянно растет: в базе данных UniProt доступны последовательности более чем трех миллионов белков, что значительно превышает число известных структур. В базе данных PDB доступны структуры более 60000 белков. Поэтому разработка теоретических методов предсказания трехмерной структуры белка представляется крайне актуальной и необходимой. Наиболее надежное предсказание третичной структуры белка дает моделирование на основе гомологии (или сравнительное моделирование), то есть только в том случае, когда известна пространственная структура какого-либо гомологичного белка.

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 научиться с высокой точностью распознавать структурные мотивы в известных структурах белков, что может способствовать построению соответствия между первичной и третичной структурой белка. Известны структурные мотивы, состоящие из двух и более элементов вторичной структуры и имеющих уникальные укладки полипептидной цепи в пространстве (α - α -углы, α - α -шпильки, L-образные и V-образные структуры и др.). Одним из часто встречаемых структурных мотивов в гомологичных и негомологичных белках являются α - α -углы. Эта супервторичная структура образована двумя соседними по полипептидной цепи α -спиралями, связанными между собой перетяжкой, и упакованы ортогонально (крестообразно). В белках α - α -углы встречаются в форме левой суперспирали. Их последовательности имеют определенное расположение в цепи гидрофобных, гидрофильных и глициновых остатков. Выдвинута гипотеза, что структурные мотивы с уникальными укладками могут быть зародышами при сворачивании белка, а остальные участки цепи пристраиваются к ним в соответствии с простыми принципами. Однако, независимо от того, по какому механизму происходит сворачивание белка структурные мотивы могут быть использованы в качестве стартовых структур для поиска возможных укладок полипептидной цепи [1].

Методы решения

В работе рассмотрены три основные задачи:

1. Подготовка обучающей выборки супервторичных структур белковых молекул. На этом этапе решается задача распознавания супервторичных структур в базе данных PDB по заданному шаблону эталона, методом спектрально-аналитического сравнения пространственных структур белковых молекул. Алгоритм распознавания α - α -углов в белковых молекулах сводится к

спектральному алгоритму поиска повторов [2].

2. Поиск характерных признаков супервторичных структур в аминокислотных последовательностях на обучающей выборке.
3. Верификация супервторичных структур методом молекулярной динамики. На этом этапе используется свойство автономной устойчивости супервторичных структур для проверки результатов распознавания.

Описанный подход реализован в виде программно-алгоритмического комплекса «Protein Reviser». Поиск α - α -углов осуществлялся по эталону, представленному на рис.1 в базе данных PDB.

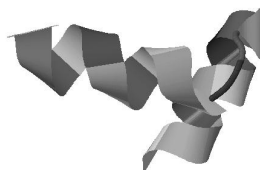


Рис.1. Ленточная модель α - α уголка с короткой перетяжкой.

Результаты

В ходе исследований были получены следующие основные результаты:

На первом этапе подготовки обучающей выборки найдено 110 α - α -углов в базе данных

PDB, соответствующих заданному эталону 1D1L C α : 15-37.

На втором этапе были выявлены характерные признаки α - α -углов в аминокислотных последовательностях, что позволяет осуществлять распознавание α - α -углов в первичных структурах белковых молекул.

На третьем этапе проведён анализ автономной устойчивости α - α -углов методом молекулярной динамики в воде. Эксперимент ставился на выборке из 57 α - α -углов. Моделирование производилось средствами Amber 11. Исследование показало, что α - α -уголок является автономно устойчивой структурой, и это свойство может рассматриваться как дополнительный верифицирующий признак в исследованиях.

Работа выполнена при поддержке грантов РФФИ: №№10-01-00609, 11-07-00716.

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.

Исследование влияния высокого давления и высокой температуры на структуру белка Nip7 архей рода *Pyrococcus* методами молекулярной динамики

Медведев К.Е.⁽¹⁾, Афонников Д.А.^(1, 2), Воробьев Ю.Н.⁽³⁾

Институт Цитологии и Генетики СО РАН, г.Новосибирск⁽¹⁾,
Новосибирский Государственный Университет, г.Новосибирск⁽²⁾,
Институт Химической Биологии и Фундаментальной Медицины СО РАН, г.Новосибирск⁽³⁾
kirill-medvedev@yandex.ru

To investigate mechanisms of protein structure adaptation to high pressure environment we compare the dynamic properties of Nip7 proteins from hyperthermophilic archaea, *Pyrococcus abyssi* (deep sea habitat) and *Pyrococcus furiosus* (shallow water habitat) at different pressures (0.1-300 MPa) and temperatures (300, 373K). We perform search for Specificity Determining Positions (SDPs) in homologous archaeal proteins with respect to the organism's habitat depth (deep/shallow water) and compare results with molecular dynamics simulation analysis.

The structure of *P.furiosus* protein was obtained by homology modeling using Nip7 from *P.abysssi* as template (PDBID 2p38). MD simulations and structure analysis were performed using GROMACS [3].

It is shown that the structure of Nip7 N-terminal domain is more stable and has smaller structure fluctuations than C-terminal (RNA-binding) domain. We demonstrated that protein from *P.abysssi* is more stable under extreme conditions than from *P.furiosus*. *P.abysssi* protein has larger polar solvent accessible surface area in comparison with *P.furiosus* protein. Most of detected SDPs in Nip7 homologs display significant changes in side chain polarity between deep/shallow water organisms.

In general, these data demonstrate the importance of water-protein interactions for the protein stability under high pressure.

Давление – один из важных факторов внешней среды. Большинство живых организмов обитают при атмосферном давлении ~0,1 МПа, а воздействие на них повышенного давления нарушает работу таких систем как репликация, транскрипция, трансляция и синтез белка, транспорт веществ через мембраны, и приводит к гибели клеток. Однако существуют организмы-пиезофилы, которые способны жить при давлениях в тысячу раз больших атмосферного, например, на дне Марианской

впадины (~100 МПа). Изучение механизмов адаптации этих организмов к высоким давлениям позволит выявить факторы, обеспечивающие устойчивость живых систем в экстремальных условиях, а их белки могут быть использованы в биотехнологии.

Высокое давление приводит к денатурации белка в результате проникновения молекул воды через поры внутрь гидрофобного ядра белка [1]. Один из перспективных подходов изучения этого процесса является подход молекулярно-динамического моделирования белковых структур при повышенном давлении.

В данной работе представлены результаты исследования молекулярной динамики влияние высоких давлений на структуру белка Nip7 гипертермофильных архей рода *Pyrococcus*, как глубоководных (*P.abysssi*, живущих на глубине 2200 м), так и мелководных (*P. furiosus*, живущих на глубинах порядка 100 м) видов. Эти белки вовлечены в биогенез рибосом, и участвуют в процессинге 27S пре-рРНК и в образовании 60S субъединицы рибосом [2].

В работе проведено сравнение компьютерных моделей трехмерных структур белков Nip7 из *P. abyssi* и *P. furiosus* при давлениях от 0,1МПа до 300МПа и температуре 300К и 373К, полученных при моделировании с помощью пакета программ GROMACS [3] в течение 30 наносекунд [2]. Показано, что деформация структуры Nip7 мелководной *P. furiosus* в целом больше, чем глубоководной *P. abyssi*. Кроме того, анализ компьютерных моделей белков Nip7 из *P. abyssi* и *P. furiosus* показал, что с увеличением давления площадь поверхности белка, доступной растворителю уменьшается. Для моделей Nip7 *P. abyssi* эта площадь меньше и ее относительные изменения меньше, чем у Nip7 из *P. furiosus*. Полученные данные свидетельствуют о том, что РНК-связывающий домен *P.abysssi* Nip7 белка более устойчив к воздействию высокого давления. Кроме того, анализ компьютерных моделей Nip7 белков из *P. abyssi* и *P. furiosus* показал, что площадь поверхности белка доступная растворителю уменьшается с увеличением давления. Данная площадь меньше для моделей Nip7 *P. abyssi* и ее

относительное изменение меньше, чем у *P. furiosus*. Также были проанализированы такие структурные характеристики белков как радиус гирации, объем и плотность белковой молекулы и расстояние между центрами масс двух доменов данного белка.

В целом, эти данные согласуются с важностью гидрофобных взаимодействий для формирования белковой глобулы [4] и предположительным механизмом денатурации белковых структур под высоким давлением [1].

Список литературы

- [1] Hummer, G., Garde, S., Garcia, A.E., et al. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins, *Proc.Natl. Acad. Sci. USA*, 1998, **95**,1552–1555.
- [2] P.P. Coltri et al. Structural insights into the interaction of the Nip7 PUA domain with polyuridine RNA, *Biochemistry*, 2007, **46**, 14177-14187.
- [3] D.Van der Spoel et al. GROMACS: Fast, Flexible and Free. *J. Comp. Chem.* 2005, **26**, 1701-1718.
- [4] Kauzmann, W. Some factors in the interpretation of protein denaturation, *Adv. Protein Chem.* 1959, **14**, 1-63.

Выбор оптимальных условий при реализации спектрального алгоритма распознавания структурных мотивов в глобулярных белках

Куликова Л.И.⁽¹⁾, Панкратов А.Н.^(1,2), Руднев В.Р.⁽¹⁾, Дедус Ф.Ф.^(2,1)

Институт математических проблем биологии РАН⁽¹⁾, Московский государственный университет им. М.В.Ломоносова⁽²⁾

Рассматривается задача распознавания структурных мотивов белков на примере α - α -углов. В рамках предложенного спектрального подхода к решению задач анализа пространственной структуры белков исследуются разложения по различным базисным полиномам: Чебышева, Лежандра, Бернштейна, сплайны и др. Оптимизация по выбору базиса и глубине разложения ведется с целью повышения точности распознавания супервторичных структур.

Определение пространственной структуры белка является актуальной и трудоемкой задачей. Решение этой проблемы имеет огромное значение, как для фундаментальной, так и для прикладной науки, поскольку это позволило бы целенаправленно исследовать взаимосвязь структуры и функции белков, конструировать новые белки с заданными свойствами, тем самым обеспечив прорыв в области биотехнологии и медицины. На практике для определения структуры белка применяются два метода: метод на основе данных рентгеновского рассеяния и метод определения по спектрам ядерного магнитного резонанса. На сегодняшний день известны последовательности более чем трех миллионов белков, что значительно превышает число известных структур. В базе данных PDB доступны структуры более 60000 белков. Поэтому разработка теоретических методов предсказания трехмерной структуры белка представляется крайне актуальным и необходимым. Успешное решение проблемы связано с внедрением и эффективным использованием информационных технологий.

Для предсказания пространственной структуры белков необходимо знать все возможные конформации полипептидной цепи и структуры, а также принципы укладки полипептидной цепи в пространственные структуры. Поэтому в настоящее время объектом пристального внимания исследователей являются элементы пространственной структуры белков, структурные мотивы, состоящие из двух и более элементов вторичной структуры и имеющих уникальные

укладки полипептидной цепи в пространстве (α - α -уголки, α - α -шпильки, L-образные структуры и др.). Такой интерес к структурным мотивам обусловлен уникальностью структур и их возможной ролью в процессе сворачивания белков. Существует гипотеза, что структурные мотивы с уникальными укладками могут быть зародышами при сворачивании белка, а остальные участки цепи пристраиваются к ним в соответствии с простыми принципами. Однако, независимо от того, по какому механизму происходит сворачивание белка, структурные мотивы могут быть использованы в качестве стартовых структур для поиска возможных укладок полипептидной цепи [1]. Одним из часто встречаемых структурных мотивов в гомологичных и негомологичных белках являются α - α -уголки [1]. Эта супервторичная структура (само понятие «супервторичные структуры белков» впервые было постулировано Л. Полингом и Р. Кори в 1951г.) образована двумя соседними по полипептидной цепи α -спиралями, связанными между собой перетяжкой, и упакована ортогонально (крестообразно). В белках α - α -уголки встречаются в форме левой суперспирали. Их последовательности имеют определенное расположение в цепи гидрофобных, гидрофильных и глициновых остатков.

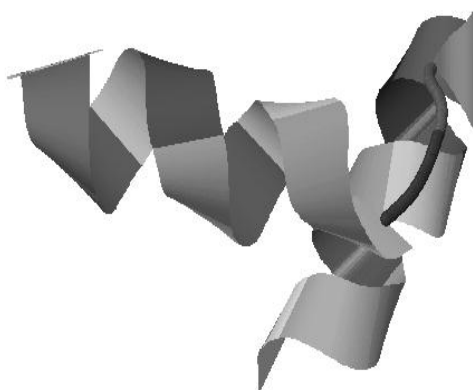
Вся полученные знания использовались при создании программно-алгоритмического комплекса для распознавания супервторичных структур в глобулярных белках с решенными 3-D структурами методами рентгеноструктурного анализа и ядерно-магнитного резонанса. Для достижения поставленной цели были получены характеристические профили структур белков, представленных в формате PDB, а потом применен спектральный алгоритм поиска заданного паттерна [2] супервторичной структуры в исследуемых глобулярных белках на примере α - α уголка с короткой перетяжкой.

Для получения уникальных белковых профилей проводилось аналитическое описание координат S_{α} -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 различные базисы. Исследовались классические ортогональные многочлены [3], сплайны, полиномы Бернштейна. Одновременно с этим решалась задача по оптимизации глубины разложения.

На рис.1 представлена пространственная структура классического α -уголка и её аналитическое описание.

а)



б)

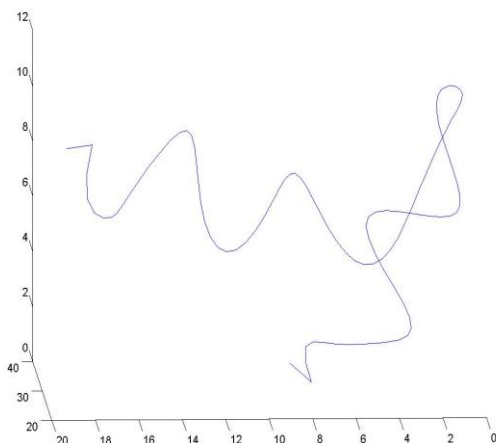


Рис.1. Супервторичная структура α - α уголок с короткой перетяжкой (а) и её аналитическое описание по полиномам Лежандра (б).

Поиск образов супервторичных структур типа α - α -углов сводится к спектральному алгоритму

поиска неточных повторов в характеристических профилях.

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, при этом при решении второй задачи возникают свои трудности, но в целом она является достаточно изученной и проработанной.

В ходе проверки алгоритмов был проведён анализ всех структур белковых молекул из базы данных UniProt PDB на предмет поиска α - α уголков. В результате достоверно были распознаны более 100 структур по заданному профилю. В дальнейшем планируется расширить область применения алгоритма, с возможностью распознавания широкого спектра супервторичных структур.

Работа выполнена при поддержке грантов РФФИ: №№10-01-00609, 11-07-00716.

Список литературы

- [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.
- [3] 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.

Пересечение синглет-триплетных поверхностей и их возможная роль в процессе встраивания атомов в химическую связь

И.В. Крауклис, Е.В. Соколова, Н.А. Хамеева, А.В. Тулуб

Санкт-Петербургский государственный университет, физический факультет
198504, г. Петродворец, ул. Ульяновская, д.1

Реакция прямого металлизации органического галогенпроизводного известна на примере магния, она протекает в простом эфире и заканчивается образованием реактива –реактива Гриньяра. Была осуществлена реакция взаимодействия метилгалогенидов с высокоактивным атомом магния в триплетном состоянии, полученным в результате лазерной абляции [1]. Подобный подход открывает в принципе возможность синтеза металлоорганических соединений с участием других металлов, например цинка, которые невозможно получить обычным путем. Эффект встраивания атомов в химическую связь представляет интерес и в биохимических реакциях, существенно выяснение условий, способствующих такому процессу. В этой связи расчет пути реакции имеет смысл исследовать методом функционала плотности, был выбран функционал B3PW91/6-31G(d).

Взаимодействие атома магния с метил- и фенилгалогенидами с образованием реактивов Гриньяра RMgHal изучено в литературе. При расчете их аналогов RZnHal ($\text{R}=\text{CH}_3, \text{C}_6\text{H}_5$; $\text{Hal}=\text{F}, \text{Cl}, \text{Br}$) согласно синглет-триплетному механизму [2,3] были найдены переходные состояния как на триплетной, так и синглетной поверхностях в комплексах RMgHal^* и RZnHal^* , они характеризуются одной мнимой колебательной частотой, связывающей исходные реагенты и продукты реакции. Был вычислен путь реакции $\text{RHal} + \text{Mg(Zn)} \rightarrow \text{RMg(Zn)Hal}$ вдоль внутренней координаты реакции с шагом, равным в масс-взвешенных координатах $0.01 \text{ (a.e.m.)}^{1/2}$. Бор. На рис.1 представлены зависимости полной энергии изучаемых систем от внутренней координаты реакции (IRC) для синглетных и триплетных состояний. Эффект псевдопересечения двух триплетных кривых отражен в методе DFT в виде пика на триплетной кривой. В начальной и конечной участках синглетного пути реакции энергия синглетного состояния всегда ниже триплетного. Активационный барьер вдоль

синглетного пути достаточно высок: 1.0 эВ для CH_3MgF^* , 1.44 эВ для CH_3MgCl^* , 1.18 эВ для CH_3MgBr^* , при замене магния на цинк активационные барьеры соответственно равны 1.46 эВ, 1.64 эВ и 1.26 эВ. В области переходного состояния в комплексах CH_3MgCl^* и CH_3MgBr^* триплетная кривая лежит заметно ниже по энергии, чем синглетная (на 0.2-0.4 эВ), что естественно связать в плане установления корреляций с возможностью образования реагента Гриньяра.

В случае переходных комплексов CH_3ZnF^* и CH_3ZnCl^* триплетная кривая находится выше синглетной, и только для CH_3ZnBr^* эти кривые пересекаются в области переходного состояния, как и у системы CH_3MgF^* . Таким образом, с ростом атомного номера галогена наблюдается понижение триплетной кривой по отношению к синглетной, одновременно возрастает энергия спин-орбитального взаимодействия. Переход системы с синглетной ППЭ на триплетную запрещен в нерелятивистском приближении. При замене метильного радикала на фенильный наблюдаются те же закономерности. Отсутствие пересечения триплетной и синглетной ППЭ указывает, что в таких случаях в рамках моноатомного приближения получить конечные продукты, по всей видимости, проблематично. Однако это не исключено в рамках кластерного подхода [4].

Список литературы

- [1] W.D. Bare, and L. Andrews // J. Am. Chem. Soc. 1998. V. 120. P. 7293 - 7301.
- [2] А.А.Тулуб // ЖОХ. 2002. Т. 72. С. 948 - 955.
- [3] А.В.Тулуб, В.В.Порсев, А.А.Тулуб // Докл. Акад. наук. 2004. Т. 398. С. 502–505.
- [4] Yu.V. Barsukov, V.V. Porsev, A.V. Tulub // Intern.J.Quant.Chem. 2012, DOI: 10.1002/qua.24187.

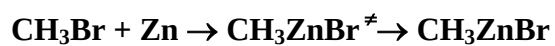
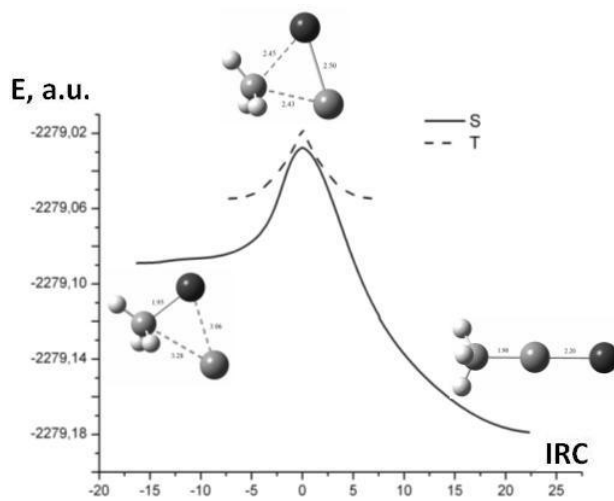
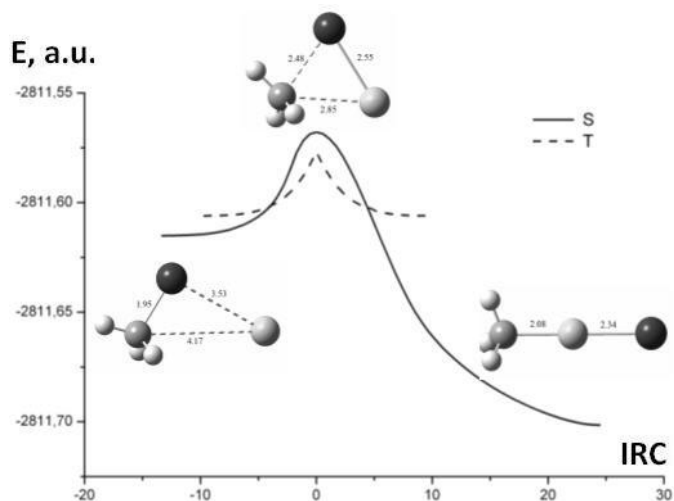
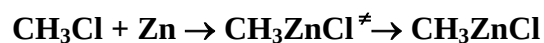
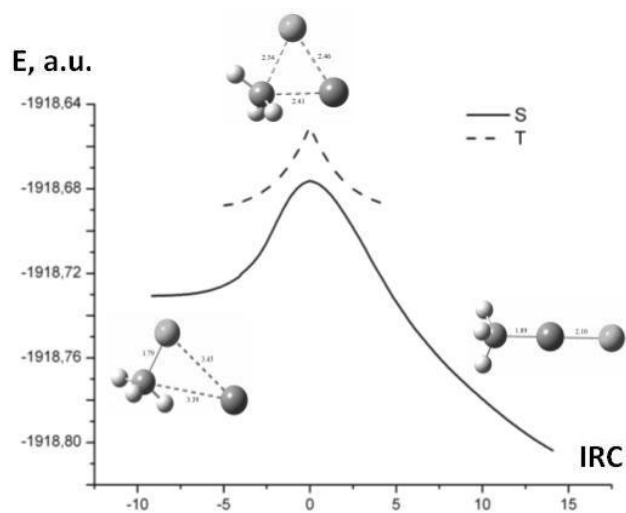
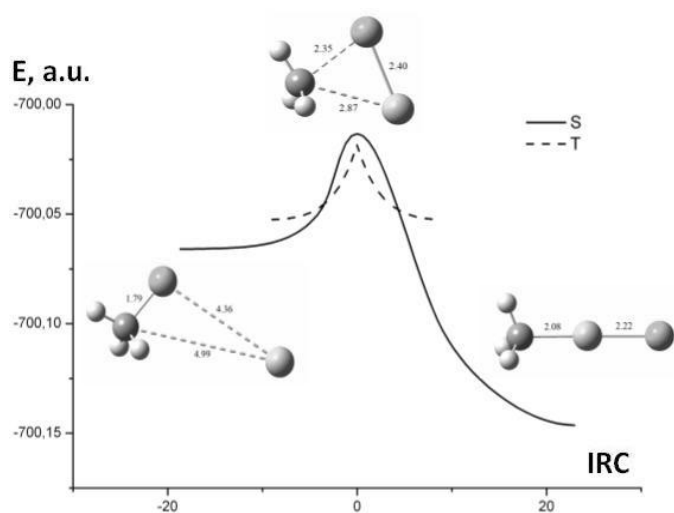
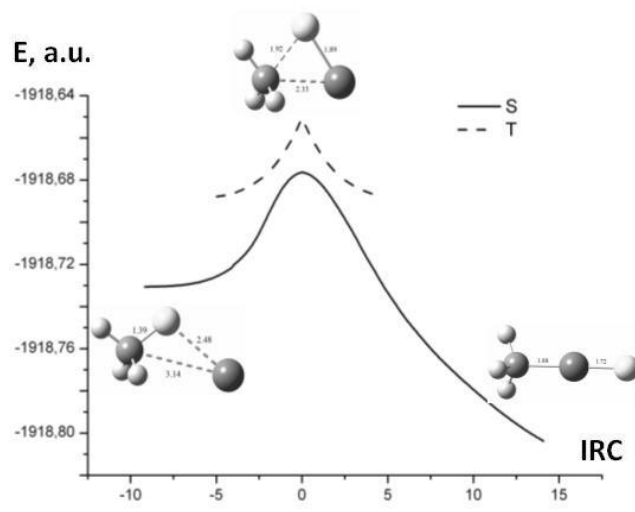
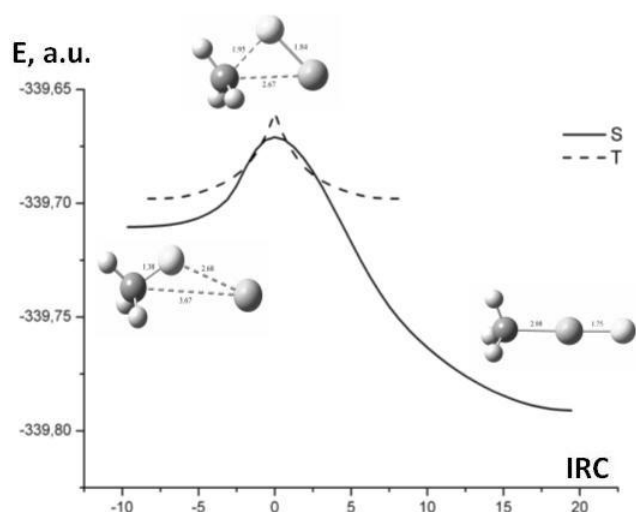


Рис.1. Полная энергия как функция внутренней координаты реакции $\text{CH}_3\text{Hal} + \text{Mg}(\text{Zn}) \rightarrow \text{CH}_3\text{Mg}(\text{Zn})\text{Hal}$ для синглетного (S) и триплетного (T) состояний согласно ВЗР91 расчетам.

Japanese encephalitis virus protease: homology modeling and structural comparisons

Shutava I. ⁽¹⁾, Junaid M. ^(1, 2), Lapins M. ⁽¹⁾, Katzenmeier G. ⁽²⁾, Wikberg J. ES ⁽¹⁾

Department of Pharmaceutical Biosciences, Uppsala University, Sweden ⁽¹⁾

Institute of Molecular Biosciences, Mahidol University, Thailand ⁽²⁾

Iryna.Shutava@farmbio.uu.se

Multiple sequence alignment of the NS2B-NS3 proteases of West Nile virus, Japanese encephalitis virus, dengue virus and yellow fever virus with a homology modeling of the Japanese encephalitis virus NS2B-NS3 protease allowed identifying differences in the putative binding pockets of the proteases.

1 Sequences alignment and homology modeling

Amino acid sequences of the NS2B-NS3 of West Nile virus (WNV), Japanese encephalitis virus (JEV), dengue virus (DEN) and yellow fever virus (YFV) (UniProt IDs: P06935, P27395, P2999 and Q6DV88) were imported from the Universal Protein Resource, UniProt [1]. Multiple sequence alignment of the four NS2B-NS3 proteases was done with ClustalW, realized in UniProt.

The crystal structures of the WNV protease/inhibitor complex (NS2B(H)-NS3-protease-Bz-nKRR-H, PDB ID: 2FP7) was used as template to build a model of the JEV NS2B(H)-NS3-protease. A pairwise sequence alignment was first done for the WNV and JEV NS2B-NS3 proteases using the ClustalW algorithm of UniProt. Homology models were then built using Modeller 9.9 [2]. The model with the lowest DOPE Score was chosen as the best.

2 Structural comparisons of JEV protease with WNV, DEN and YFV proteases

The alignment of the JEV, WNV DEN2 and YFV polypeptide sites cleaved by NS2-NS3 proteases shown that the JEV and WNV sites are the most similar; they all contain Lys-Arg at the P2-P1 positions, while at the P'1 position a Gly is present at four out of the five sites (the exception is the NS3/NS4A-junction, which contains a Ser). The DEN2 and YFV sites are less similar with those of JEV, however; in six cases out of ten, these sites have Arg-Arg at the P2-P1 positions, while the four remaining sites have different positively charged residues (Lys-Arg, Arg-Ser, Gln-Arg and Arg-Lys) at these positions. It is also notable that the five cleavage sites of JEV contain five different amino acids at P3; the same holds true for P4.

The multiple sequence alignment of the WNV, JEV,

DEN2 and YFV NS2B-NS3 revealed that the JEV protease is closest to WNV, with a sequence identity of 76.3% (93.7% sequence similarity), while the DEN2 and YFV proteases show only 50.2% and 45.5% sequence identity (79.5% and 76.9% sequence similarity) with the JEV protease, respectively.

The X-ray crystal structure of the WNV protease/inhibitor complex NS2B(H)-NS3-protease-Bz-nKRR-H [3] indicates that the WNV NS2B-NS3 protease has four distinct substrate binding pockets, termed S4-S1, which accommodate the P4-P1 residues of protease substrates [4]. The multiple sequence alignment shows that JEV and WNV share the same amino acid at 17 of the 22 indicated positions; JEV and DEN2 share the same amino acid at 11 positions and JEV, and YFV at 13 positions. However, there are positions in each one of S4-S1 pockets, which contain physico-chemically quite different amino acids. Residue 84 of NS2B contributes to the S2 pocket, and is Asp, Asn, Ser and Glu in, respectively, JEV, WNV, DEN2 and YFV. Residue 86 of NS2B is part of the S3 pocket and contains also four different amino acids in the four viruses (His, Gln, Ser, and Lys). Positions 132 and 155 of NS3, which belong to S1 and S4 pockets, respectively, are also different in the different proteases. These amino acid differences are thus likely candidates to contributing to the differences in kinetics of substrates for the different proteases.

In order to cast further light into this possibility, we built a homology model of the JEV NS2B-NS3 protease, using the crystal structure of a WNV NS2B(H)-NS3-protease-Bz-nKRR-H inhibitor complex as template. The model for the JEV protease is shown superimposed on the WNV protease/inhibitor complex in Fig. 1, panel A. Pairwise 3D structure alignment showed that the JEV and WNV proteases share large structural similarity (RMSD=0.172 Å) with a large structural conservation in the enzymes' active sites. However, clear differences are seen in the substrate-binding pockets at NS2B residues Asn84, Gln86, and NS3 residues Thr132, Ile155 (WNV NS2B-NS3 protease numbering, corresponding to NS2B residues Asp84, His86, and NS3 residues Arg132, Glu155 in the JEV protease). This is visualized in Fig. 1, panels B and C. Side chains of all four residues are in close proximity with the ligand and are capable of forming multiple hydrogen bonds (except Ile 155 in the WNV protease). The superimposition shows that the

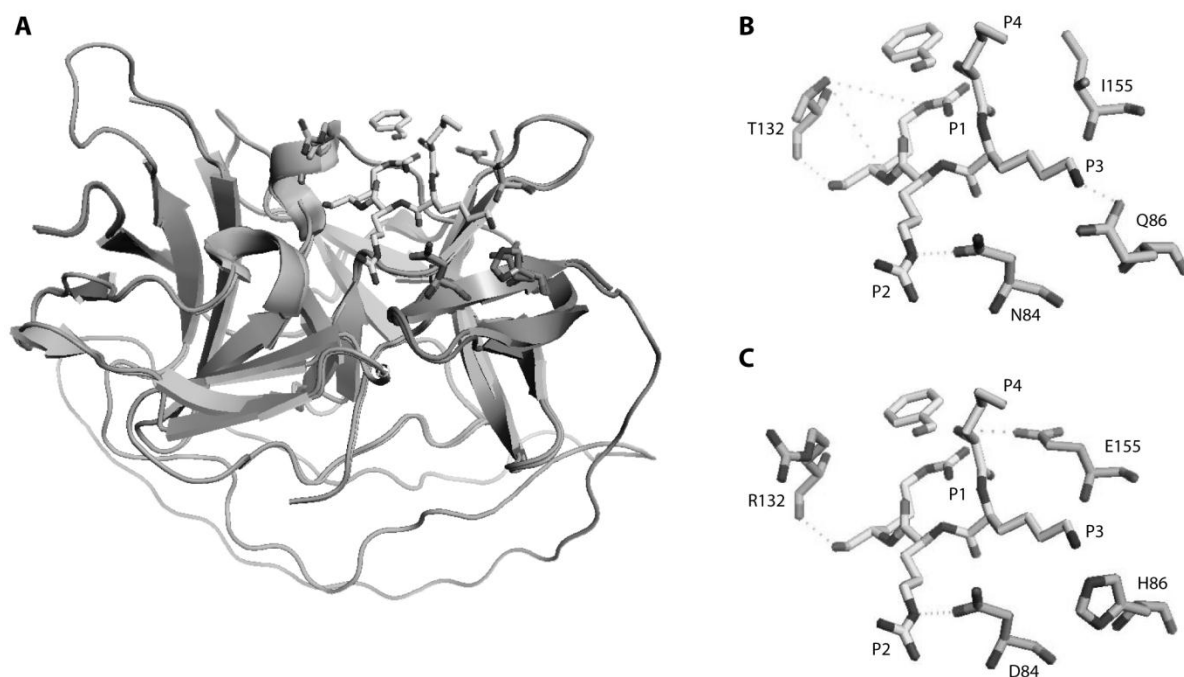


Figure 1. Structures of JEV and WNV proteases bound to Bz-nKRR-H. (A) Superimposition of the WNV X-ray structure (PDB ID 2FP7) on the JEV homology model. The WNV structure is shown in blue (NS2B) and cyan (NS3); the JEV structure in orange (NS2B) and green (NS3). The Bz-nKRR-H inhibitor in the WNV structure is shown in stick representation in yellow. (B) Detailed interactions of Bz-nKRR-H with WNV (from X-ray structure). (C) Superimposition of the Bz-nKRR-H X-ray structure (PDB ID 2FP7) on the JEV homology model. Potential hydrogen bonds indicated by dashed lines. Putative interacting amino acid residues of WNV and JEV are represented in cyan and green, respectively, and Bz-nKRR-H in yellow (figure from [7]).

change of these residues between JEV and WNV leads to a major rearrangement of the H-bond network in the substrate-binding pocket, which very likely will affect the cleavage preferences of the two proteases. Moreover, while all these residues are uncharged in the S1-S4 pockets of the WNV protease, the corresponding residues in the S2 and S4 pockets of the JEV protease are acidic (Asp and Glu) while in the S1 pocket it is basic (Arg). All of this may influence both the ionic interactions with substrates and the geometries of the binding pockets.

All of this makes the proteases of the two viruses quite different in terms of their binding mode, affinity and cleavage preferences and are likely reasons for the observed differences in the cleavage efficiency of substrates for the JEV and WNV enzymes.

An earlier study employing hybrids between NS2B and NS3 of DEN and JEV showed that only the (DEN)NS2B-(JEV)NS3 hybrid could efficiently process the JEV polypeptide, whereas the (JEV)NS2B-(DEN)NS3 hybrid was inactive, supporting the notion that the NS3 proteases have a major impact on the structure and substrate affinity of the proteases [5, 6]. This also indicates that there are likely differences in the conformational space between the NS2B-NS3 proteases of different flaviviruses.

This work was supported by grants from the Thailand Research Fund (TRF) BRG4980008 (to GK) and the Swedish International Cooperation Development Agency (SIDA). A PhD scholarship from the University of Malakand, Pakistan, to MJ is gratefully acknowledged. Dr. Iryna Shutava was supported by a postdoctoral scholarship from the Swedish Institute under the Visby programme.

References

- [1] The UniProt Consortium (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 40: D71-D75.
- [2] Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234: 779-815.
- [3] Erbel P, Schiering N, Arcy AD, Renatus M, Kroemer M, et al. (2006) Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat Struct Mol Biol* 13: 372-373.
- [4] Wichapong A, Pianwanit S, Sippl W, Kokpol S (2009) Homology modeling and molecular dynamics simulations of Dengue virus NS2B/NS3 protease: insight into molecular interaction. *Journal of Molecular Recognition* 23: 283-300.
- [5] Lin CW, Huang HD, Shiu SY, Chen WJ, Tsai MH, et al. (2007) Functional determinants of NS2B for activation of Japanese encephalitis virus NS3 protease. *Virus Research* 127: 88-94.
- [6] Jan LR, Yang CS, Trent DW, Falgout B, Lai CJ (1995) Processing of Japanese encephalitis virus non-structural proteins: NS2B-NS3 complex and heterologous protease. *J Gen Virol* 76: 573-580.
- [7] Junaid M, Chalayat C, Sehgelmeble Torrejon A, Angsuthanasombat C, Shutava I et al. (2012) Enzymatic analysis of recombinant Japanese encephalitis virus NS2B(H)-NS3pro protease with fluorogenic model peptide substrates. *PLoS One* 7(5): e36872.

Prediction of protein folding rate based on nucleation mechanism: How important is the size of folding nucleus?

Galzitskaya O.V.⁽¹⁾, Glyakina A.V.⁽²⁾

Institute of Protein Research, RAS, Pushchino⁽¹⁾, Institute of Mathematical Problems of Biology,
RAS, Pushchino⁽²⁾
ogalzit@vega.protres.ru

The problem of protein self-organization is one of the most important problems of molecular biology nowadays. Despite the recent success in the understanding of general principles of protein folding, details of this process are yet to be elucidated. Moreover, the prediction of protein folding rates has its own practical value due to the fact that aggregation directly depends on the rate of protein folding. The transition state free energies and the size of folding nuclei for 84 proteins and peptides with known experimental data at the point of thermodynamic equilibrium between unfolded and native states have been calculated using the dynamic programming method where each residue is considered to be either folded as in the native state or completely disordered. The rates of folding for 84 proteins and peptides are in good correlation with the experimentally measured folding rates at the mid-transition point (the correlation coefficient is 0.75). We were the first to demonstrate that the correlation coefficient between the folding nucleus size and the experimentally found folding rate is not high (-0.57). The comparison of predictions of folding rates by using different methods has been done.

1 Introduction

The current study is focused on folding of separate protein molecules and their structural properties depending on this event.

We studied 82 single-domain proteins and two peptides whose folding rates and tertiary structures are known [1, 2]. Among them, 26 proteins exhibited multi-state kinetics, and 58 proteins showed two-state kinetics (see <http://phys.protres.ru/resources/compact.html>); 24 belong to the class of all- α proteins (class *a*), 28 – to all- β proteins (class *b*), 9 – to α/β proteins (class *c*), and 23 – to $\alpha+\beta$ proteins (class *d*).

Using our model [3] we calculated the values of free energy barriers ($F_{eff}^{\#}/RT$, $F_{min}^{\#}/RT$) and the size of folding nuclei (L_{nuc}) for 84 proteins, and predicted folding times ($t_{1/2}$) for 67 proteins using the Monte Carlo algorithm [4]. We modeled the folding process using a relatively simple theoretical approach based on the nucleation mechanism. The dynamic programming method was applied to investigate free-energy landscapes of protein folding and the size of folding nuclei. The proposed model has proved to be reliable and fit with the size of the data set.

2 Results and Discussion

2.1 Estimation of the size of folding nucleus and the protein folding rate from calculations of the free energy barrier

With the assumption that the folding rate k_{mt} is proportional to $\exp(-F^{\#}/RT)$ and the possibility to calculate the transition state free energy $F^{\#}$, we can estimate the correlation of computed $F_{eff}^{\#}/RT$ with experimentally found folding rates k_{mt} (or rather, $\ln(k_{mt})$). The correlation coefficient between these values is 0.72. It should be noted that for the proteins under study, the experimentally found folding rates range over about 11 orders of magnitude, which coincides with the range of estimated folding rates (folding barrier heights) (except for two large proteins).

Furthermore, the calculated minimal free energy barriers ($F_{min}^{\#}/RT$) correlate somewhat better with the logarithm of experimentally measured folding times at the point of thermodynamic equilibrium: for all 82 proteins, the correlation coefficient is 0.75.

We were the first to estimate the sizes of the folding nuclei which include about 60% of all the residues of the studied proteins. The correlation coefficient reaches only 0.57 as k_{mt} . However, it is high (0.79) for correlation with free energies predicted from the proposed model.

2.2 Comparison of different parameters with experimentally found folding rates for two-state and multi-state folders

A separate analysis (used as convenience) of two-state- and multi-state folders showed that the former have a higher correlation coefficient of $F_{eff}^{\#}/RT$ and $F_{min}^{\#}/RT$ with folding rates at mid-transition (0.72 ± 0.09 and 0.77 ± 0.09 , respectively) than the latter (0.58 ± 0.16 and 0.55 ± 0.17 respectively). Such a tendency was observed for all parameters of the 84 proteins studied. However, in mid-transition conditions all single-domain proteins demonstrated a two-state (i.e., “all-or-none”) transition mode in terms of both thermodynamics and kinetics.

The values of characteristic first passage time $t_{1/2}$ computed (for mid-transition) by Monte Carlo modeling for 67 are in a good correlation with $\ln(k_{mt})$: the correlation coefficient is -0.82 ± 0.06 . Note that proteins

with two-state kinetics have a higher correlation coefficient (-0.83 ± 0.08) than proteins with multi-state kinetics (-0.68 ± 0.15).

The coefficients of correlation between folding rates and a set of structural parameters of the 84 studied proteins (contact order, length and its various dependents) were calculated. It is worth mentioning that no correlation was observed between folding rates and contact orders for both two-state and multi-state folders. As to correlation with the length (L), the coefficient is as high as 0.70. Surprisingly, correlation between folding rates and various length dependents ($\ln(L)$, $L^{2/3}$, $L^{1/2}$) is high, which suggests that the protein size is an important and determinative parameter of protein folding [1, 5]. L and $\ln(L)$ have a high correlation coefficient with folding rates at mid-transition, especially for proteins with two-state kinetics.

2.3 Comparison of different parameters with experimentally found folding rates for different structural classes of proteins

The average structural properties give a high correlation with folding rates in each structural class. In case of $F_{eff}^{\#}/RT$ (and $F_{min}^{\#}/RT$), the correlation coefficient is sufficiently high for two classes: -0.74 (-0.79) for class *b*, and -0.72 (-0.75) for class *d*. It should be noted that no structural class displayed high correlation between its folding rates and the size of folding nucleus.

The correlation between the logarithm of the first passage times in Monte Carlo simulations and the logarithm of folding rates for 67 proteins from different structural classes has sufficiently high coefficients: -0.74 ± 0.14 for class *a*, -0.83 ± 0.11 for class *b*, -0.84 ± 0.12 for class *d*.

2.4 The effect of energy-entropy compensation on minimizing of the free energy barrier

The most efficient kinetic pathway has the minimal (among other pathways) transition state free energy. To clarify the relationship between the size of the optimal folding nucleus with the minimal free energy barrier and conformational entropy of this state, we sought to calculate this value. We divided the minimal free energy barrier into entropic and energy components. During the folding, a notable loss in conformational entropy was expected due to native contacts newly gained by the chain. The entropic and energy components correlated with the logarithm of experimentally measured folding times at the point of thermodynamic equilibrium as 0.49 and 0.55, respectively. The effect of energy-entropy compensation led to a free energy barrier ($F_{min}^{\#}/RT$) that correlated somewhat better with the logarithm of experimentally measured folding times at the point of thermodynamic equilibrium: the correlation coefficient for all 82 proteins was 0.75. The coefficient of correlation between the minimal free energy barrier and the minimal size of folding nucleus was 0.69, while for the pair conformational entropy – the size of optimal folding nucleus it was very high (0.99). The correlation coefficient between the effective size of folding nucleus

and conformational entropy of the folding transition state was 0.76. The high correlation level reported here is in agreement with previous findings that topological dependence of the folding rate is related to conformational entropy of the folding transition state [6]. These results indicate that considerable entropy–energy compensation is less effective for slowly folding proteins than for fast folding proteins.

It was of interest to demonstrate the positions of native-like regions in optimal folding nuclei of four similar-sized proteins from each of the mentioned four structural classes. Although the "best" transition states have nearly equal free energies, they demonstrate notable difference in location and even in size of their native-like parts ("folding nuclei" on different pathways).

3 Conclusions

The proposed theoretical approach is based on a search for free energy saddle points in networks of protein unfolding pathways for prediction of folding/unfolding nuclei in three-dimensional protein structures. It allows estimating the folding nucleus size and free energy barriers, which is its advantage over other empirical and bioinformatical approaches. Our nucleation-based calculations of folding rates of 84 proteins and peptides show a good correlation with the experimentally measured folding rates at mid-transition both for proteins from different structural classes and for two-state and multi-state folders. We were the first to demonstrate that the correlation coefficient between the folding nucleus size and the experimentally found folding rate is not high (-0.57). The established good correlation (0.75) between the estimated free energy barriers and the measured folding rates for 84 proteins and peptides proves that our model is reliable and fits with the size of the considered data set. Also, the offered nucleation model implicates the physical aspect of the process of protein folding.

This study was supported in part by the Russian Foundation for Basic Research (11-04-00763), Russian Academy of Sciences (programs "Molecular and Cell Biology" (01200959110) and "Fundamental Sciences to Medicine").

References

- [1] D.N. Ivankov, N.S. Bogatyreva et al. *PLoS ONE*, 2009, **4**, e6476
- [2] O.V. Galzitskaya. *Curr Protein Pept Sci*, 2010, **11**, 523-537.
- [3] O.V. Galzitskaya, A.V. Finkelstein. *PNAS USA*, 1999, **96**, 11299-11304.
- [4] J.W. Moore, R.G. Pearson. *Kinetics and Mechanism*. New York: Wiley, 1981.
- [5] O.V. Galzitskaya, S.O. Garbuzynskiy et al. *Proteins*, 2003, **51**, 162-166.
- [6] S. Wallin, H.S. Chan. *J Phys: Condens Matter*, 2006, **18**, S307–S328.

«Золотой треугольник» для скоростей сворачивания глобулярных белков

Гарбузинский С.А., Иванков Д.Н., Богатырёва Н.С., Финкельштейн А.В.

Институт белка РАН, 142290, г. Пушкино Московской области, Россия
afinkel@vega.protres.ru

Ранее, на основе полимерной теории, нами было теоретически показано [1-3], что: (1) скорость сворачивания однодоменного глобулярного белка *in vitro*, в точке термодинамического равновесия его нативного и денатурированного состояний, должна лежать в пределах $\sim 10^8 \text{ c}^{-1} \times \exp\left(-\frac{1}{2} L^{2/3}\right) \div \sim 10^8 \text{ c}^{-1} \times \exp\left(-\frac{3}{2} L^{2/3}\right)$, где $\approx 10^8 \text{ c}^{-1}$ – определенная

на опыте скорость конформационной перестройки одного аминокислотного остатка, а L – число остатков в цепи белка; (2) если ΔG – разность свободной энергии нативного и денатурированного состояния белка, то скорость его сворачивания должна падать примерно в $\exp(-\Delta G / 2RT)$ раз, где T – температура, а R – газовая постоянная. Кроме того, по понятным биологическим причинам, скорости сворачивания белков не должны быть меньше, чем $\sim 10^{-3} \div 10^{-4} \text{ c}^{-1}$. Очерченные выше условия создают «золотой треугольник», в который, теоретически, должны попадать скорости сворачивания однодоменных глобулярных белков любой величины и стабильности.

Теперь мы показываем, что все наблюдаемые на опыте скорости сворачивания белков – а их около двух сотен – действительно, попадают в этот «золотой треугольник». Кроме того, мы предсказываем (1) предел величины белкового

домена, самоорганизация которого находится под чисто термодинамическим контролем, (2) максимально возможный размер однодоменного глобулярного белка или белкового домена.

Работа выполнена при поддержке Программы "Молекулярная и клеточная биология" Президиума РАН (грант № 01200957492), Федерального агентства по науке и инновациям (грант № 02.740.11.0295), РФФИ (грант № 10-04-00162а), Грантов Президента РФ для государственной поддержки молодых российских учёных (МК-4894.2009.4, МК-5540.2011.4) и Фонда "Династия".

Список литературы

- [1] Finkelstein A.V., Badretdinov A.Ya. / Fold. Des., 1997. V.2. P.115–121.
- [2] Финкельштейн А.В., Бадретдинов А.Я. / Мол. Биол., 1997. Т.31. С.469–477.
- [3] Finkelstein A.V. In "Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter", J.-L. Barrat, M. Feigelman, J. Kurchan, J. Dalibard, Eds. (EDP Sciences, Les Ulis, Paris, Cambridge & Springer-Verlag, Berlin). 2003. С.649–704.

Моделирование нейронных сетей с помощью стохастических клеточных автоматов

Ершов Н. М.^(1, 2), Кравчук А. В.⁽²⁾

ВМК МГУ⁽¹⁾, Университет «Дубна»⁽²⁾
ershovnm@gmail.com

A new approach to simulation modeling of neural networks based on the use of stochastic cellular automata is suggested. The notion of two-dimensional Markov automata is introduced. The model of an excitable medium and the mechanism of inhibition are considered. Simple feedforward neural network, realizing the XOR function, is constructed and numerically analyzed.

1 Марковские автоматы

1.1 Марковские системы

В качестве инструмента для имитационного моделирования нейронных сетей [1] в настоящей работе используются двумерные клеточные автоматы специального типа, построенные на основе марковских систем [2]. *Марковская система* представляет собой набор подстановок вида $\alpha \rightarrow_{[p]} \beta$, где α и β цепочки символов в заданном алфавите, p – вероятность применения подстановки. Набор может включать в себя несколько подстановок с одинаковой левой частью, но их суммарная вероятность не должна превышать единицы.

Процесс применения подстановок к заданной цепочке символов состоит в последовательном выполнении следующего алгоритма: 1) текущая цепочка случайным образом разделяется на подцепочки; 2) каждая полученная подцепочка заменяется на новую согласно заданному набору подстановок с учетом их вероятностей; 3) все подцепочки, полученные в результате второго шага, соединяются и образуют новую текущую цепочку.

Если правая и левая части каждого правила заданной системы подстановок имеют равные длины, то такую марковскую систему можно трактовать как блочный стохастический клеточный автомат, в котором состояния клеток меняются согласованно в рамках каждого блока, а сами блоки формируются случайным образом. Такой клеточный автомат будем называть *марковским автоматом*.

1.2 Двумерные марковские автоматы

Будем представлять пространство двумерного марковского автомата в виде прямоугольной

решетки (матрицы) размера $n \times m$, в каждую ячейку которой помещен символ заданного алфавита.

Состояние такого автомата меняется со временем по следующему алгоритму: на каждом шаге эволюции с вероятностью 1/2 выбирается способ разбиения матрицы символов на цепочки – вертикально (на столбцы) или горизонтально (на строки), после чего, каждая цепочка преобразуется стандартным (вышеописанным) образом.

2 Моделирование возбудимой среды

2.1 Массовые марковские системы

Если набор подстановок системы содержит в себе перемешивающие подстановки вида $ab \rightarrow_{[\theta]} ba$, и вероятности применения всех остальных подстановок много меньше θ , то такую систему будем называть *массовой*. Состояние массовой системы в каждый момент времени описывается системой дифференциальных уравнений, аналогичных кинетическим уравнениям, используемых при описании динамики химических реакций.

2.2 Моделирование возбуждения

Рассмотрим следующую систему подстановок:

$$\Sigma_1: \begin{cases} x\lambda \leftrightarrow_{[\theta]} \lambda x, \\ x\lambda \rightarrow_{[p]} x\lambda x, \\ x\lambda \rightarrow_{[q]} \lambda\lambda. \end{cases}$$

При условии, что вероятности p и q много меньше вероятности перемешивания θ , данная система обладает массовым поведением и ее эволюция может быть описана дифференциальным уравнением вида:

$$\dot{x} = ax^2(1-x) - bx(1-x),$$

параметры которой определяются вероятностями p и q . Нетрудно убедиться, что это дифференциальное уравнение имеет три точки покоя: $x_1 = 0$, $x_2 = 1$ и $x_3 = b/a$. Если выполнено условие $b < a$, то первые две точки являются устойчивыми, а последняя – неустойчивой.

При значительной степени перемешивания для гарантированного возбуждения системы Σ_1 необходимо существенно превысить порог возбуждения, причем это превышение должно носить

глобальный характер. Однако, если перемешивание производится не столь активно, то возможна ситуация, когда локальное превышение порога в некоторой ограниченной подобласти приводит к возникновению волны возбуждения, которая переводит всю систему в возбужденное состояние.

2.3 Моделирование торможения

Наряду с процессами возбуждения в нейронных сетях важную роль играют процессы торможения, которые фактически меняют (повышают) порог возбуждения нейрона. Для моделирования торможения добавим к системе Σ_1 символ y (ингибитор) и следующие две подстановки:

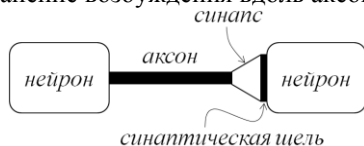
$$yx \rightarrow [r] y\lambda, \quad y \rightarrow [s] \lambda.$$

Символы y «гасят» имеющееся возбуждение (первая подстановка), при этом постепенно исчезают сами (вторая подстановка). Поэтому следует ожидать, что наличие символов y в системе будет приводить к повышению порога возбуждения, при условии, что тормозящий импульс будет располагаться вблизи от возбуждающего импульса.

3 Моделирование нейронной сети

3.1 Искусственный нейрон

Будем представлять искусственный нейрон в виде некоторой двумерной области, состоящей из символов λ (возбудимая среда). Нейрон передает свое возбуждение другим нейронам с помощью аксона, представленному цепочкой символов α . Распространение сигнала по аксону описывается подстановками $x\alpha \rightarrow x\tilde{x}, \tilde{x}\alpha \rightarrow \tilde{x}\tilde{x}$, первая из которых инициирует возбуждение аксона, вторая отвечает за распространение возбуждения вдоль аксона.



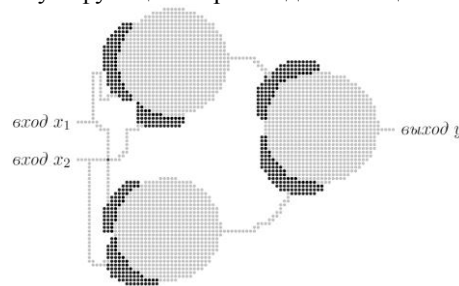
Передача возбуждения с аксона на другой нейрон выполняется с помощью синапса, представляющего собой некоторую область, отделенную от нейрона прослойкой символов s , моделирующей синаптическую щель. Синапс заполнен символами A или B , представляющих собой нейромедиаторы в неактивной форме.

После того, как возбуждение аксона достигло синапса, все его нейромедиаторы переводятся в активное состояние a или b и переносятся в нейрон через синаптическую щель, возбуждая нейрон или тормозя его возбуждение. Активные нейромедиаторы постепенно разрушаются (истощаются), это гарантирует, что процесс передачи импульса (возбуждающего или тормозящего) с одного нейрона на другой будет ограничен во времени.

3.2 Нейронные сети прямого распространения

Хотя нейроны в предложенной модели способны только возбуждаться, а сбрасывать возбуждение не могут, тем не менее, из таких нейронов можно строить нейронные сети прямого распространения [3], например, многослойные перцептроны (со ступенчатой функцией активации нейронов).

В качестве примера была построена простейшая сеть, состоящая из трех нейронов и реализующая логическую функцию строгой дизъюнкции.



В следующей таблице показаны результаты тестирования рассматриваемой нейронной сети. Первые два столбца соответствуют двум входам (1 – есть сигнал на входе, 0 – сигнала нет), третий – ожидаемому выходу нейросети. В последнем столбце показан процент C полученных правильных ответов.

x_1	x_2	y	C
0	1	1	78%
1	0	1	87%
1	1	0	95%

Случай $x_1 = x_2 = 0$ не рассматривался, т. к. выход сети в этом случае гарантированно будет нулевым.

4 Заключение

В результате выполненной работы получены следующие результаты: введено понятие двумерного марковского автомата; построена и проанализирована бистабильная марковская система с двумя устойчивыми состояниями, рассмотрено влияние пространственной организации на процесс возбуждения системы, предложен механизм торможения возбуждения; описана модель искусственного нейрона; рассмотрен и численно исследован пример простой нейронной сети прямого распространения.

Список литературы

- [1] С. Хайкин. Нейронные сети: полный курс. М.: Изд. дом «Вильямс», 2006.
- [2] Н. М. Ершов. Имитационное моделирование с помощью Марковских систем. *Тихоновские чтения, МГУ, 14 июня 2011 г.: Тезисы докладов*, 2011, с. 35.
- [3] P. D. Wasserman. Neural computing: theory and practice. Van Nostrand Reinhold, 1989.

Три математические модели основных компонентов вызванных потенциалов коры мозга

Покровский А.Н.

Санкт-Петербургский государственный университет
anpokrovski@gmail.com

Three geometrical surfaces: a cylinder, a cone and a parabolic cone were used in models of main evoked potentials' components in cerebral cortex. Differential equations of intracellular (and extracellular) potentials were derived for each geometrical model.

1 Введение

Вызванные потенциалы (ВП) коры мозга регистрируются (как функции времени и глубины коры) с помощью системы внеклеточных электродов. Максимальная амплитуда ВП регистрируется в средних слоях коры как внеклеточные потенциалы апикальных дендритов пирамидных нейронов. В данной модели ([1], [2]) пренебрегаем всеми остальными дендритами и нейронами, считаем кору плоской, апикальные дендриты одинаковыми цилиндрами или конусами, направленными поперёк коры, по оси z . Пологаем, что уровень $z=0$ – верхняя граница апикальных дендритов пирамидных нейронов; на уровне z_m находятся тела пирамидных нейронов.

Предполагаем, что поверхность коры – плоскость $z=z_0<0$; плотность распределения синапсов зависит только от z . Обозначим радиус апикального дендрита как $\rho(z)$; ранее детально рассматривался только случай $\rho(z)=\rho=\text{const.}$.

Средний внеклеточный потенциал $\phi(z,t)=\alpha V(z,t)$ приблизительно пропорционален внутриклеточному потенциалу $V(z,t)$ дендрита ([1], [2]), $\alpha \ll 1$. Поэтому в первом приближении функцию $V(z,t) - \phi(z,t) = (1 - \alpha)V(z,t)$ и её производную по времени t можно заменить на V и $\partial V/\partial t$.

2 Варианты уравнений

2.1 Цилиндрические дендриты

В случае $\rho=\rho_m=\text{const.}$ основное уравнение имеет вид [2]:

$$(\rho/2R)\partial^2 V/\partial z^2 = C*\partial V/\partial t + I(z, V, t), \quad (1)$$

где R – удельное сопротивление протоплазмы в дендрите, C – удельная ёмкость мембраны дендрита. Функция $I(z, V, t)$ зависит от управляющего («синаптического») сигнала, а также от решений дифференциальных уравнений,

не связанных с геометрией апикального дендрита.

Вычислялось решение $V(z,t)$ уравнения (1) при заданных токах синапсов и $I(z, V, t)$ («прямая задача», «корректная»), а так же решение обратной задачи вычисления токов синапсов при заданной функции $V(z,t)$ («некорректная» [3] обратная задача). Все вычисления используются только при $0 \leq z \leq z_m$.

2.2 Первый вариант конического дендрита

Площадь поперечного сечения дендрита линейно зависит от z при $0 < z < z_m$. Тогда $\rho(z)/\rho_m = (z/z_m)^{1/2}$, и основное уравнение имеет вид

$$(\rho_m/R)*[(4z*z_m + \rho_m^2)^{-1/2}]* \{ \partial [z*(\partial V/\partial z)]/\partial z \} = C*\partial V/\partial t + I(z, V, t). \quad (2)$$

Здесь $0 \leq z \leq z_m$; правая часть уравнения (2) совпадает с правой частью уравнения (1). Прямая задача корректная, как и для уравнения (1), но её решение не выписывается в явном виде [2], как в случае уравнения (1). Обратная задача «некорректная» и требует регуляризации [3].

2.3 Второй вариант конического дендрита

Радиус поперечного сечения дендрита линейно зависит от z , то-есть при $0 \leq z \leq z_m$

$$\rho(z)/\rho_m = z/z_m.$$

В этом случае уравнение имеет вид:

$$(\rho_m/2R)*[(z_m^2 + \rho_m^2)^{-1/2}]* z^{-1}* \{ \partial [z^2*(\partial V/\partial z)]/\partial z \} = C*\partial V/\partial t + I(z, V, t), \quad (3)$$

и правая часть уравнения (3) совпадает с правой частью уравнения (1). Прямая задача корректная, как и для уравнения (1), но её решение не совпадает с решением прямой задачи в случае уравнения (1).

Обратная задача «некорректная» и требует регуляризации [3], как и в предыдущих случаях.

3 Заключение

Для сравнения этих трёх вариантов математической модели вызванных потенциалов необходимо выяснить, какой из вариантов лучше аппроксимирует экспериментальные данные, хотя

бы на одном примере. Сравнение с экспериментальными данными для варианта (1) уравнений (случай $\rho = \rho_m = \text{const.}$) уже было сделано в [2]. Аналогичные (вероятно, ещё более громоздкие вычисления) необходимо будет выполнить и для вариантов (2) и (3) уравнений.

С целью выбора наиболее адекватного варианта уравнений необходимо будет решать прямые и обратные задачи для каждого варианта уравнений, то-есть устойчивые (корректные) и неустойчивые («некорректные») задачи. Для решения неустойчивых задач можно использовать известные методы из [3].

В коре высших животных различают две популяции пирамидных нейронов, тела которых находятся в разных слоях коры. Для таких случаев придётся рассматривать одновременно два варианта наилучшей математической модели вызванных потенциалов с различными наборами

параметров модели для каждого варианта.

Кроме рассмотренных моделей вызванных потенциалов, для верхних и нижних слоёв коры мозга, где нет апикальных дендритов, необходимо будет использовать другие, ещё не проверенные варианты математических моделей вызванных потенциалов коры мозга.

Список литературы

- [1] А.Н. Покровский Процессы управления в нервных клетках. Уч. пособие. Л., 1987. 85 с.
- [2] А.Н. Покровский / *Нейроинформатика-2010, часть I*. М.: НИЯУ МИФИ, 2010. С. 52-59.
А.Н. Тихонов, В.Я. Арсенин Методы решения некорректных задач. М.: Наука, 1974.

Модель автономного агента-охранника

Редько В.Г., Сохова З.Б.

НИИ системных исследований РАН, Москва.
vgredko@gmail.com, zarema_s@mail.ru

The computer model of an autonomous agent-guard, which is functioning in a circular ring that is divided into sectors, is designed and investigated. Two versions of the model for the agent-guard are analyzed by means of computer simulations.

1 Введение

В настоящей работе построена и проанализирована модель автономного агента на примере агента-охранника. Исследованы два типа модели. В первой модели обучение агента производится по *факторам потребностей*, во второй – обучение производится по *общей энергии агента-охранника*. Работа развивает модели [1, 2], в которых было начато исследование агентов, обладающих естественными потребностями и мотивациями.

Функция агента-охранника состоит в поиске нарушителей и поддержании своего внутреннего ресурса. Нарушители появляются на охраняемой территории случайно. При обнаружении нарушителя агент ударяет его, и нарушитель покидает территорию. Качество поведения агента-охранника оценивается по числу нарушителей, которых он не пропустил на территорию.

2 Модель агента-охранника с двумя потребностями

2.1 Описание модели

Рассматривается один агент-охранник, обладающий внутренним ресурсом $R(t)$. Время t дискретно. Моделью мира агента является круговое кольцо, разбитое на шесть секторов. Кольцо представляет собой границу охраняемой территории. В четных секторах находятся розетки. В каждый такт времени в любом из секторов с вероятностью p_1 появляется нарушитель.

Считаем, что агент-охранник имеет две потребности: *питания* и *охраны территории*. Если ресурс агента R меньше порога r_{th1} , то ведущей потребностью является питание, иначе ведущей является потребность охраны.

Потребностям агента соответствуют два фактора: фактор питания F_f и фактор охраны территории F_p .

Фактор питания пропорционален ресурсу агента: $F_f = k_F R(t)$. Фактор охраны территории увеличивается при выполнении агентом действия «удар» и наличия

нарушителя в одной клетке с агентом на ΔF_p и уменьшается на 1 в других ситуациях.

Удовлетворение ведущей потребности является положительным подкреплением при обучении агента.

2.2 Система управления агента-охранника с двумя потребностями

Система управления агента основана на наборе правил вида: $S_k \rightarrow A_k$, где S_k – ситуация, A_k – действие, k – номер правила. Согласно правилам в ситуации S_k нужно выполнить действие A_k . Каждое правило имеет свой вес W_k . Веса правил изначально случайны, а затем модифицируются методом обучения с подкреплением [3].

Ситуация S_k определяется 1) наличием или отсутствием розетки в текущем секторе агента и в двух соседних секторах, 2) наличием или отсутствием нарушителя в текущем секторе и в двух соседних секторах, и 3) ведущей потребностью.

В каждый такт времени агент может выполнять одно из следующих действий A_k : 1) *питание*, 2, 3) *перемещение на один сектор по или против часовой стрелки*, 4) *удар*, 5) *отдых*. Если действие «питание» вырабатывается в секторе, где имеется розетка, то ресурс агента $R(t)$ увеличивается на r_1 . При выполнении действий «перемещение» (в любом из двух направлений), удар и отдых ресурс агента уменьшается на величины r_2, r_4, r_5 .

Каждый такт времени с вероятностью $1-\varepsilon$ выполняется то действие, для которого вес W_k соответствующего ему правила для текущей ситуации максимален, с вероятностью ε выполняется случайное действие.

2.3 Схема обучения

Используется схема обучения с подкреплением [3]. Подкреплением является изменение фактора ведущей потребности F_f или F_p :

$$\Delta W(t-1) = \alpha [F_L(t) - F_L(t-1) + \gamma W(t) - W(t-1)],$$

где $F_L(t)$ – фактор ведущей в такт времени t потребности, $W(t)$ – вес правила, примененного в такт t , α – параметр скорости обучения, γ – дисконтный фактор.

2.4 Результаты моделирования

Параметры компьютерного моделирования составляли: $\Delta F_p = 5$, $k_F = 0.2$, $\varepsilon = 0.05$, $\gamma = 0.9$, $\alpha = 0.1$, $r_{th1} = 50.0$, $r_1 = 50.0$, $r_2 = 1$, $r_4 = 1$, $r_5 = 5$, $p_1 = 0.1$.

Моделирование показало, что агент-охранник обучается поддерживать внутренний ресурс на уровне не ниже заданного порога. Количество

нарушителей также уменьшается. Зависимости ресурса агента и количества нарушителей, остающихся в мире, от номера такта времени представлены на рис. 1, 2.

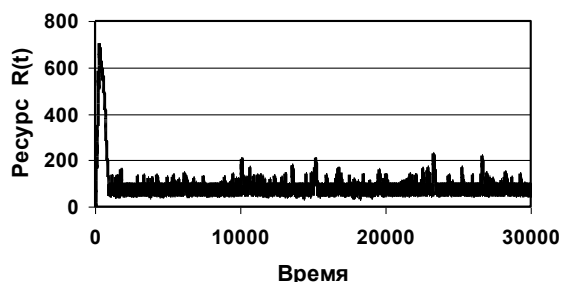


Рис. 1. Динамика ресурса агента-охранника $R(t)$

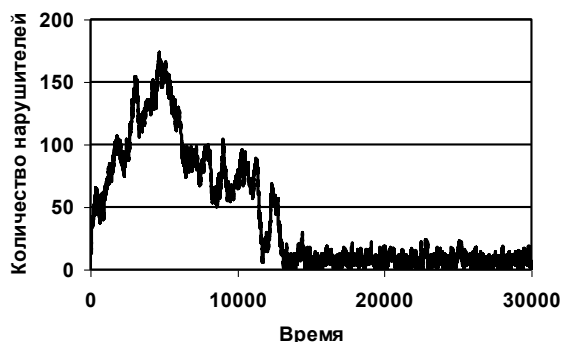


Рис. 2. Динамика количества нарушителей

3 Модель агента-охранника без потребностей

3.1 Описание модели

Система управления агента аналогична системе управления агента с потребностями, но в отличие от модели, описанной выше, не содержит ведущую потребность.

В данной модели под ресурсом агента $R(t)$ понимается *общая энергия*, которую агент накапливает при выполнении любых действий. При выполнении действия «питание» и наличии розетки в текущем секторе ресурс агента увеличивается на dR_f . При выполнении действия «удар» и наличии нарушителя в секторе ресурс агента увеличивается на dR_s . При питании, перемещении, ударе и отдыхе

ресурс агента уменьшается соответственно на r_0 , r_1 , r_2 , r_3 .

3.2 Схема обучения

В данном варианте модели подкреплением является изменение ресурса $R(t)$:

$$\Delta W(t-1) = \alpha [R(t) - R(t-1) + \gamma W(t) - W(t-1)],$$

где $W(t)$ – вес правила, примененного в такт t , α – параметр скорости обучения, γ – дисконтный фактор.

3.3 Результаты моделирования

Параметры компьютерного моделирования составляли: $dR_f = 1$, $dR_s = 3$, $\varepsilon = 0.05$, $\gamma = 0.9$, $\alpha = 0.3$, $r_0 = 1$, $r_1 = 1$, $r_2 = 1$, $r_3 = 1$, $p_1 = 0.1$. Моделирование показало, что обученный агент справляется со своей задачей, количество агентов-нарушителей резко уменьшается, при этом ресурс агента растет.

4 Заключение

Построена и исследована модель автономного агента-охранника с потребностями и без потребностей. Компьютерные эксперименты показали, что после обучения агента в модели без потребностей количество нарушителей остающихся в мире незначительно меньше, чем в модели с потребностями.

В модели с потребностями агент-охранник показывает более разумное поведение, так как он обучается поддерживать ресурс на определенном уровне.

Список литературы

- [1] А.Г. Коваль, В.Г. Редько. Поведение модельных организмов, обладающих естественными потребностями и мотивациями. *Математическая биология и биоинформатика*, 2012, Т. 7. № 1, 266-273.
- [2] В.Г. Редько, Г.А. Бесхлебнова. Моделирование адаптивного поведения автономных агентов. *Нейрокомпьютеры: разработка, применение*, 2010. № 3, 33-38.
Р.С. Саттон, Э.Г. Барто. *Обучение с подкреплением*, М.: Бином, 2011.

Анализ структурных коррелятов процессов поддержания кратковременной памяти в поведении нейросетевых агентов

Лахман К.В., Бурцев М.С.

Национальный Исследовательский Центр «Курчатовский Институт»
klakhman@gmail.com

Animals behave adaptively in the environment with multiply competing goals. Understanding of the mechanisms underlying such goal-directed behavior remains a challenge for neuroscience as well for neuromorphic artificial intelligence research. To address this problem we developed an evolutionary model of adaptive behavior in the multigoal stochastic environment. Proposed neuroevolutionary algorithm is based on neuron's duplication as a basic mechanism of agent's recurrent neural network development. Results of simulation demonstrate that in the course of evolution agents acquire the ability to store the short-term memory and, therefore, use it in behavioral strategies with alternative actions. We found that evolution discovered two mechanisms for short-term memory. The first mechanism is integration of sensory signals and ongoing internal neural activity, resulting in emergence of cell groups specialized on alternative actions. And the second mechanism is slow neurodynamical processes that makes possible to code the previous behavioral choice.

1 Введение

Понимание механизмов обучения и поддержания эффективного поведения в стохастических средах со сложной иерархией целей является одним из ключевых вопросов в изучении принципов нейрональной переработки информации и для разработки био-подобного искусственного интеллекта. Тем не менее, полное понимание этих механизмов невозможно, если мы не рассматриваем историю их возникновения в эволюции и индивидуальном развитии.

Использование альтернативных действий критически необходимо для адаптивности поведения большинства животных. С точки зрения задачи моделирования для решения данной проблемы управляющий алгоритм должен реализовывать не-марковский процесс. Другими словами для выбора различных действий в одном и том же состоянии среды агент должен иметь память о прошлых действиях. Феномен кратковременной памяти активно обсуждается в области исследований рекуррентных нейронных сетей с

точки зрения реверберации сигнала [1] и повторяющихся последовательностей активаций [2]. Теоретики рассматривают реверберацию в качестве важного механизма кратковременной памяти [3]. Тем не менее, принципы, лежащие в основе генерации текущего поведения с использованием кратковременной памяти, на данный момент не до конца изучены.

В данной работе мы исследовали появление кратковременной памяти в процессе эволюции популяции нейросетевых агентов в модельном мире и ее роль в адаптивном поведении.

2 Модель среды и агента

Текущее состояние среды, в которой существует агент, кодируется бинарным вектором. В любой момент времени агент может изменить один бит вектора состояния. Поведение агента управляется многослойной нейронной сетью произвольной топологии. В среде задана иерархия целей, которые определяются как последовательность изменений вектора состояния среды. С каждой целью среды ассоциируется награда, которую агент накапливает, достигая различные цели за фиксированное время. Набранная награда влияет на репродуктивный успех агента в процессе эволюции. В качестве эволюционного алгоритма для популяции нейроморфных агентов используется оригинальный алгоритм, основанный на дубликации нейронов. Данный алгоритм, как и широко известный NEAT [4], позволяет осуществлять топологические мутации нейронной сети. Используемая модель подробно описана в работе [5].

3 Экспериментальное исследование

3.1 Поведенческая эволюция

Типичное поведение агента, вырабатываемое в ходе эволюции, состоит из двух фаз: предварительная фаза движения к основному поведенческому циклу и дальнейшие повторения этого цикла. В конкретном запуске обе этих фазы зависят от начального состояния среды, из которого агент начинает реализовывать поведение.

Результаты моделирования демонстрируют, что резкое увеличение средней набранной награды по популяции обычно совпадает с появлением

большого количества новых циклов и их динамической конкуренции. При этом новые стратегии формируются как расширение появившихся ранее циклов. Впоследствии одна из стратегий доминирует в поведении большинства агентов популяции. Далее эволюция происходит путем увеличения количества состояний, из которых доминирующий цикл может быть реализован.

3.2 Альтернативное поведение

Анализ поведенческих стратегий, появляющихся в моделируемой эволюции, показал, что агенты приобретают способность использовать в поведении кратковременную память. Наличие памяти подтверждается тем, что агенты могут реализовывать политики поведения, основанные на альтернативных действиях, когда из одного и того же состояния среды совершаются различные действия в зависимости от предыдущей истории поведения. Мы обнаружили два механизма обеспечения альтернативных действий. Первым механизмом является интеграция текущей сенсорной информации и внутреннего состояния сети, с помощью которой динамически появляются группы специализированных клеток. Данные группы значимо меняют свои выходы и влияют на принятие решения в состояниях, которые ассоциированы с альтернативными ситуациями. Наиболее глубокая память, которую мы наблюдали у проанализированных агентов, составляет как минимум 4 прошлых состояния (нижняя граница).

Примитивное альтернативное поведение в модели может быть также базировано на медленном нейродинамическом процессе. Механизм, лежащий в основе данного типа поведения, - это осцилляторная динамика выхода нейрона. В то время, как выход нейрона выше порога, и значит передается по исходящим связям, агент реализует первую часть поведения. В определенный момент значение выхода опускается ниже порога передачи и нейрон больше не вытормаживает другой нейрон, чья активность необходима для реализации второй ветки поведения.

Способность использовать кратковременную память делает возможным реализацию значительно более сложного поведения и аккумуляцию большей награды в течение «жизни» агента.

3.3 Поведенческая роль нейронов

Анализ роли элементов нейронной сети, управляющей поведением агента, показал, что можно выделить отдельные нейроны, ответственные за выход на основную поведенческую стратегию, и нейроны, обеспечивающие ее реализацию. В случае удаления нейронов первого типа сохраняется возможность выхода на основной цикл, но из

значительно меньшего количества начальных состояний среды. На рис.1 представлены области притяжения поведенческих стратегий, реализуемых одним из агентов, после удаления такого нейрона. До удаления поведение агента сходилась к одному циклу вне зависимости от начального состояния.

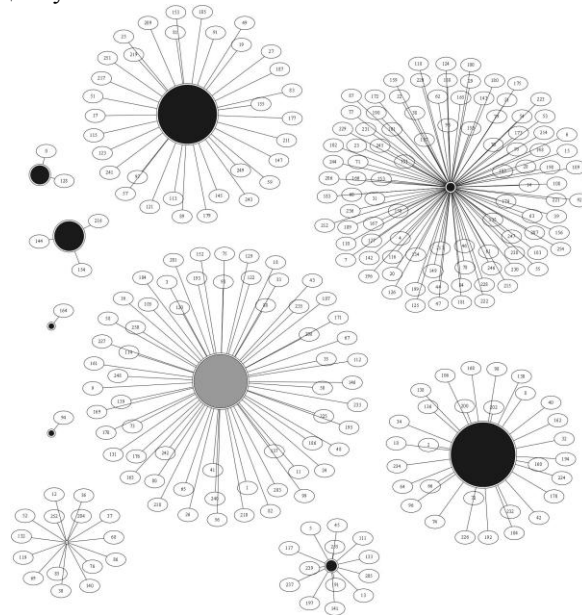


Рис. 1. Области притяжения поведенческих циклов (черные круги) после удаления нейрона, обеспечивающего выход на основное оптимальное поведение (серый круг). Размер круга пропорционален длине цикла - кол-ву совершаемых действий. Белыми овалами обозначены начальные состояния, из которых запускается агент.

Стратегии, которые агент начинает реализовывать после удаления, можно найти в поведении агентов более ранних поколений эволюции. Данный факт говорит о том, что удаляемые нейроны корректируют поведение агента таким образом, чтобы он мог выполнять более успешную политику поведения.

Список литературы

- [1] S. Hochreiter, J. Schmidhuber et al. *A Field Guide to Dynamical Recurrent Neural Networks*, 2001, IEEE Press, 237–243.
- [2] M.M. Botvinick, D.C. Plaut. *Psychological Review*, 2006, vol. 113, no. 2, 201–233.
- [3] S. Grossberg. *Studies in Applied Mathematics*, 1973, vol. 52, no. 3, 213–257.
- [4] S. Kenneth, R. Miiikkulainen. *Evolutionary Computation*, 2002, vol. 10, no. 2, 99–127.
- [1] К.В. Лахман, М.С. Бурцев. *Нейроинформатика – 2012*, ч. 2, 134–143.

Моделирование динамики мембранных рецепторов в свободном состоянии

Мурзина Г.Б.

Институт высшей нервной деятельности и нейрофизиологии РАН, Москва, Россия

Email: gbmurzina@mail.ru

Cluster and “uncluster” receptors as major types of postsynaptic receptors are discussed. Inter receptor relations may result in activation of the majority of cluster receptors, when the mediatory concentration is small then to activate “uncluster” receptors. The analysis of receptor lateral diffusion, receptor endocytosis and exocytosis influence on change in number of free membrane receptors was performed by the means of mathematical model.

Введение

Согласно современным представлениям мембранные рецепторы, определяющие чувствительность нейрона к воздействующим веществам (медиаторам, гормонам), существуют в двух основных состояниях: свободном (несвязанном) состоянии и в кластерном (фиксированном) состоянии. Кластерное состояние формируется при взаимодействии рецепторов с примембранным актиновым цитоскелетом либо с белками постсинаптической плотности (комплексом белков, прилегающим к постсинаптической мембране и имеющим толщину порядка 30 нм [3]). В кластерном состоянии рецепторы находятся преимущественно на постсинаптической мембране, а в свободном – на остальной мембране нейрона и в небольших количествах на синаптической мембране. Переход рецепторов из кластерного состояния в свободное и обратно происходит постоянно с небольшой скоростью. При воздействии медиатора этот процесс активизируется. Формирование рецепторных кластеров сопровождается увеличением синаптической эффективности и возникновением длительной потенциации, а при их распаде и уменьшении количества мембранных рецепторов наблюдается длительная депрессия. При возникновении длительной депрессии изменяется количество рецепторов в свободном состоянии на определенном локусе мембраны, что приводит к увеличению их эндоцитоза внутрь клетки и латеральной диффузии в другие области мембраны. Изучение депрессии холиночувствительности командных нейронов виноградной улитки на клеточном аналоге привыкания [6,9] и анализ экспериментальных кривых с помощью математической модели позволяет определить соотношение процессов эндоцитоза/экзоцитоза мембранных рецепторов (ЭнР/ЭкР) и их латеральной диффузии (ЛД), и таким образом

исследовать динамику мембранных рецепторов в свободном (несвязанном) состоянии.

Кластерное состояние рецепторов

Формирование рецепторных кластеров приводит к существованию значительного различия концентраций рецепторов на синаптической и внесинаптической мембранах (для АХ-рецепторов эта величина составляет соответственно $10^4/\text{мкм}^2$ и $10/\text{мкм}^2$ [5]). Рецепторы, формирующие один кластер, могут оказывать влияния друг на друга через рецептор-рецепторное или рецептор-белковое взаимодействие, либо вследствие конформационных изменений окружающих их липидов мембраны и примембранных белков. Экспериментально обнаружено существование взаимодействующих между собой рецепторов, активируемых различными медиаторами: A_1 аденозиновых и D_1 дофаминовых рецепторов, A_{2A} аденозиновых и D_2 дофаминовых рецепторов, D_1 и NR1 субъединицы НМДА-рецепторов [2]. Результаты вычислений показали [7], что если предположить наличие синхронизирующего влияния активности рецепторов одного кластера, друг на друга, то активация таких рецепторов будет происходить при концентрации медиатора меньшей, чем у некластерных рецепторов.

Динамика рецепторов в свободном состоянии

Относительно перемещения внесинаптических некластерных мембранных рецепторов делались выводы на основании анализа экспериментальных кривых, полученных при ритмической локальной аппликации агониста ацетилхолиновых рецепторов на сомму командных нейронов виноградной улитки и применении блокаторов различных протеинкиназ (ПК) и протеинфосфатаз (ПФ). Для анализа использовалась математическая модель, в которой сома нейрона рассматривалась в виде плоской поверхности, в центр которой ($r=0$) апплицировался медиатор [8]. В модели учитывались наличие ЛД рецепторов, их транспортировка по мембране к области формирования клатриновых ямок, ЭнР и ЭкР, которые определяют интернализацию и рециклирование рецепторов. Концентрация мембранных рецепторов $u(r,t)$ и концентрация рецепторов в примембранной области $v(r,t)$ в

зависимости от расстояния r от места аппликации медиатора в момент времени t описывались системой дифференциальных уравнений. Параметры модели задавались с учетом результатов работ [6,9], согласно которым скорости ЭНР и ЭкР изменяются при аппликации медиатора, т.к. в этих процессах участвуют ПК и ПФ, активация которых зависит от увеличения внутриклеточной концентрации Ca^{2+} . Для соответствия расчетных и экспериментальных данных моделировалась серия ритмических подведений медиатора, осуществляемых через промежутки времени $\Delta t = 3$ мин. Результаты вычислений позволяют получить изменения $u(r,t)$ с течением времени на любом локусе мембраны. Согласно результатам моделирования основное изменение концентрации мембранных рецепторов наблюдается в точке подведения медиатора и небольшие изменения – в прилегающих областях. При повторяющейся аппликации медиатора увеличение коэффициента ЛД рецепторов приводит к увеличению величины ответа нейрона, т.е. к уменьшению стационарного уровня депрессии АХ-тока. Сравнение динамики изменения $u(0,t)$ и $v(0,t)$ при значениях $\Delta t = 1-6$ мин показало, что уменьшение уровня депрессии АХ-тока при увеличении промежутка времени между аппликациями медиатора определяется доминированием процесса ЭкР, вследствие чего большее количество рецепторов «возвращается» на мембрану к моменту следующей аппликации медиатора.

Поскольку из полученных экспериментальных данных следовало, что ряд ПК и ПФ влияют преимущественно на ЭНР и ЭкР, транспорт рецепторов в лизосомы и транспорт вновь синтезированных рецепторов к мембране [6,9], была исследована зависимость концентрации мембранных рецепторов $u(r,t)$ от скоростей этих процессов. В ходе моделирования было обнаружено, что в случае отсутствия перемещения внутриклеточных везикул с интернализированными рецепторами вдоль мембраны или внутрь нейрона, т.е. при наличии только примембранных ЭНР и ЭкР, может возникать эффект «постстимульной отдачи». В этом случае после прекращения стимуляции число рецепторов в прилегающих к месту аппликации медиатора локусах мембраны в течение некоторого времени превышает число рецепторов на этих участках

мембраны до начала стимуляции. Амплитуда такого эффекта зависит от величины коэффициента ЛД рецепторов в мембране. Изменение коэффициента ЛД рецепторов, зависящего не только от вязкости мембраны, но и от состояния примембранного матрикса (сформированного преимущественно актиновыми нитями и образующего примембранную двухмерную решетку, внутри которой находятся рецепторы [1]), может происходить вследствие воздействия различных ПК и ПФ на различные актинсвязывающие белки, которые осуществляют взаимодействия актинового цитоскелета с мембраной клетки [4]. Деполимеризация актина будет приводить к увеличению мобильности мембранных рецепторов, и, следовательно, к увеличению коэффициента ЛД рецепторов, а стабилизация примембранного матрикса, наоборот, – к уменьшению скорости ЛД рецепторов. Полученные расчетные данные позволяют говорить о необходимости учета ЛД мембранных рецепторов в свободном (несвязанном) состоянии наряду с ЭНР и ЭкР при изучении процессов, лежащих в основе изменения реакции клетки на воздействующие вещества (медиаторы).

Работа выполнена при поддержке РФФИ (грант № 12-04-00209-а).

Список литературы

- [1] Allison D.W., Chervin A.S., et al. *J.Neurosci.*, 2000, **20**, 4545-4554.
- [2] Franco R., Canals M. et al. *Trends Bioch.Sci.*, 2003, **28**(5), 238-243.
- [3] Kennedy M.B. *Science*, 2000. **290**, 750-754.
- [4] Nakamya N., Oshiro N. et al. *Gen.Cell*, 2000, **5**, 571-576.
- [5] Sanes J.R., Lichtman J.W. *Ann. Rev. Neurosci.*, 1999, **22**, 389-442.
- [6] Махновский Д.А., Мурзина Г.Б. и др. *ЖВНД*, 2011, **61**(4), 459-475.
- [7] Мурзина Г.Б. *ЖВНД*, 2008, **58**(6), 643-648.
- [8] Мурзина Г.Б. *Биофизика*, 2012 (в печати).
- [9] Пивоваров А.С., Васильева Н.А. и др. *ЖВНД*, 2012, **62**(2), 222-232.

Interactions of EEG and NIRS signals in the prefrontal neo-cortex operation balancing

J. Faber⁽¹⁾, M. Novák⁽¹⁾

Faculty of Transportation Sciences,
Czech Technical University of Prague ⁽¹⁾
mirko@fd.cvut.cz

EEG and oxidation measurement appear to be mutually complementary methods to a certain extent. A comparison of EEG, NIRS graph and physiological (hyperventilation) as well as psychological (Raven) activities brings out the following facts: increased attention, while pronounced in EEG (alpha drop in particular), has little effect on the NIRS curve. Increased mentation takes the form of a delta rise in EEG and quite a striking oxydation increase in the NIRS record.

1 Introduction

Many higher brain functions are realized in the prefrontal parts of neo-cortex which interacts with majority of other, evolutionary older brain parts. This concerns also the sub-cortical areas, hypothalamus and amygdale. While in neo-cortex the important rational functions prevail, these sub-cortical parts influence the particular subject emotional stages before all. For well operation of the subject, these interactions must be kept in balance, which is a complicated procedure, influenced by many internal and also external factors, both of the physical and pschical nature. In this contribution we present some results reached by systematic application of combined measurements :EEG electroencephalography) and NIRS (Near Infrared Spectral Analysis).

2 Significant activities of prefrontal cortex and other brain parts

The human prefrontal cortex is known as the seat of signals, which are of the quasi-periodic and quasi-stationary nature and are very complicated. The analysis of them needs the use of special mathematical procedures. Unfortunately, one cannot distinguish, from which layer certain component of EEG signals comes.

constructive one) has to be considered as the key to general abilities of neural networks with which deal in their cortices the higher animals.

The EEG signals recorded on the surface of human head, involve the mixture of frequency components from about 0,5 to more than 80 Hz. If one provides the deeper advanced analysis of some EEG signals components, one finds that e.g. in the so called α region (8 to 13 Hz) the signal carries additional low frequency modulation, the character of which is sketched in Fig.1.

Here one sees the increasing and decreasing amplitudes of low frequency modulation, including about 7-8 α waves. The lower part of Fig.1 shows the schematic explanation of couplings among the vertical

and lateral paths from cortex to thalamus in which this modulation origins.

Beside sensory nuclei, the thalamus houses also motor nuclei which control muscle activity, in particular movement of the extremities, trunk, mimic muscles, respiration, speech etc. Most of the thalamic nuclei have yet another microstructure and function: groups of neurons called rhythmic thalamic generators (RTG) which generate mainly the already mentioned EEG α and β waves. Most of the outputs from the particular RTGs are aimed at the cortical microstructures, i.e., at the columns (echelons) their number is estimated approximately to 10 million.

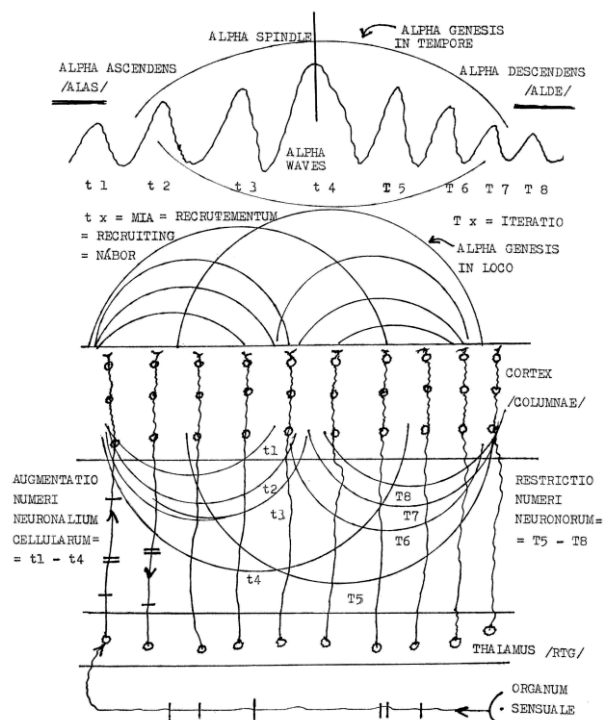


Fig.1. The low frequency spindles (fusi) existing on EEG α waves

3 Hypothesis on α waves

In our opinion, it is exactly reverberation in the TCRS that stands for iteration represented in the EEG curve by fussy-form alpha activity with it ALAS and ALDE parts. The neuronal network in ALAS operates in accordance with the Ivachnenko multilayered iterative algorithm MIA algorithm, while ALDE represents typical iteration, i.e., a more or less stereotype process aimed at the desired end.

Alpha waves of the highest amplitude with the height of the waves declining from there on. In the ALDE phase is alpha wave amplitude decreasing in the second half of the alpha spindle due to the neuronal population decline. This can be likened to the process of neuron selection in an artificial neuronal network (ANN) as devised by Ivakhnenko. The ALDE phase also coincides with the task solution proper running on the rules of the ordinary iteration process, i.e., with the computation aimed at the desired results

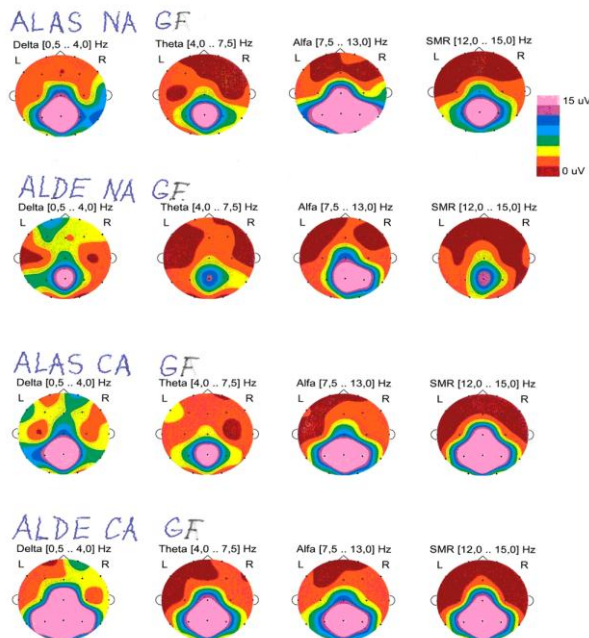


Fig.3. Text: This graph shows 4 series of brain mapping (BM) diagrams of No. XVIII. from spectral GF computations in four frequency bands (delta, theta, alpha and "SMR or slow beta activity"); on the extreme right is a scale for easier orientation from 0 to 15 micro V in pseudo-colors or in different shades of grey. Series 1 are EEG curve computations of the native state (NA) in the ascending part of the spindle (ALAS), series 2 - computations in the descending part (ALDE), showing decreased alpha and delta values in the ALDE phase (as well as in theta and "SMR", though these bands proved to be of lesser significance in other tested persons, which was why we refrained from studying them systematically). Series 3 are GA computations of EEG in the process of calculation (CA) in ALAS, and - series 4 - in the ALDE phase. Neither alpha nor "SMR" show any differences there, although there was a significant build-up of delta (and theta) activity during the arithmetic task.

4 Interactions with NIRS measurements

EEG optical description was within norm in 12 juniors and 2 seniors of the total 23 investigated persons, the rest of the cohort had slightly abnormal graphs with episodes of theta, especially in response to HV. The EEG curve was further analysed using frequency spectrum and coherence function. The EEG analyzed interval usually lasts 4-9 seconds. The EEG and NIRS records occurred

during experiment simultaneously and continually.

The EEG analysis corroborated our earlier experience which can be summed up in a few sentences.

Attention is inversely proportional to alpha activities (8 -13 Hz), i.e., the higher level of attention we show, the lower is the output of the alpha band.

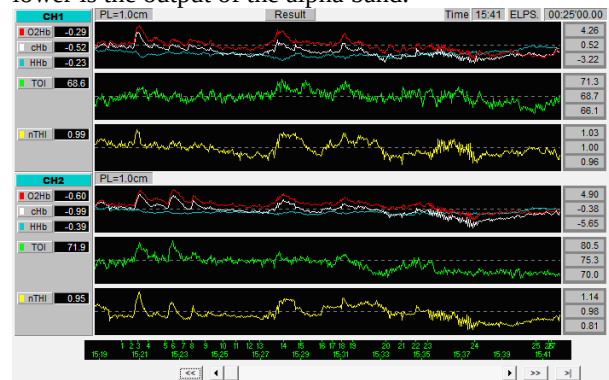


Fig. 4. Global view (15 minutes, 41 seconds) to oxymetric (NIRS) and electroencephalographic (EEG) measurement in 28 years high educated female. Upper part is left, bottom part is right hemisphere in prefrontal localization (Fp). Red curve is oxyhaemoglobin (O2Hb), blue curve si deoxyhemoglobin (CO2Hb), white curve is total haemoglobin. Tissue Oxygenation Index (TOI in %) represents the brain saturation and equals fraction: $k \text{ O2Hb} / k \text{ O2Hb} + k \text{ HHb}$, k = constant scattering contribution, green curve.

5 Discussion

EEG and oxidation measurement appear to be mutually complementary methods to a certain extent. EEG Increased mentation takes the form of a delta rise in EEG and quite a striking oxydation increase in the NIRS record.

References

- [1] Faber J., Pěkný J., Pieknik R., Tichý T., Faber V., Bouchner P., Novák M.: Simultaneous recording of electric and metabolic brain activity, Neural Network World, vol.20, 2010, No.4, pp.539-557
- [2] Faber J., Novák M.: Thalamocortical reverberation in the brain produces alpha and delta rhythms as iterative convergence of fuzzy cognition in a uncertain environment Neural Network World, vol. 21, 2011, No.2, , pp. 169-192
- [3] Faber J.: QEEG (Qualitative and quantitative EEG analysis), Galén, Prague, 2005, pp..191
- [4] Walter W.G., Cooper R., Aldridge V.J., et al. : Contingent negative variation, electric sign of sensori-motor association and expectancy in the human brain, Nature London, 203, 1964, 380-384
- [5] Timsit M., Koninckx N., Dargent J.: et al.: Variations contingentes négative en psychiatrie. Electroenceph. clin. Neurophysiol., 28, 1970, 41-47
- [6] Howard R.C., Fenton G.W., Fenwick P-B-C.: Event-related brain potentials in personality and psychopathology: a pavlovian approach, Research

Quantum Models for Brain Network

Miroslav Svítek

Czech Technical University in Prague, Faculty of Transportation Sciences,
Konviktska 20, Prague 1, Czech Republic
svitek@fd.cvut.cz

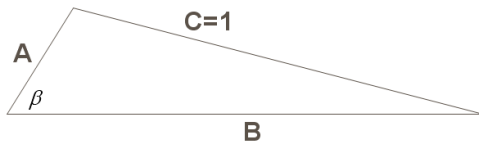
Quantum models have been discovered in the field of quantum information science. We will illustrate these principles with concrete examples of quantum information science for human brain modeling.

1 Interconnection of observers

Let us imagine that we are flipping a coin, so that every toss comes out as heads or tails. Now let us assume that we are observing the same phenomenon of coin tossing, but now with two observers who are not consulting each other about their observations. There might be two persons, one of whom watches for and registers only tosses of heads and the other only tails.

Let us deal with the geometric interpretation of the errors of our two observers. Because the condition must be fulfilled that the length of the hypotenuse of a right triangle must always equal one, we can model the error rates of our observers using the angle between the triangle's legs, so that the square root of the probability determined by the first observer (including his or her errors) will be depicted on the x-axis and the square root of the probability found by the second observer (including that observer's errors) will be depicted on the y-axis.

Tossing heads:	H	Tossing experiment:	H,T,T,H,H,T,T,H,T
Tossing tails:	T	Number of registered tossing:	10
P observer registered:	2	O observer registered:	5

$$p(H) = \frac{2}{10} \quad p(T) = \frac{5}{10} \quad \beta = 1.076$$


$$C^2 = 1 = A^2 + B^2 \pm 2 \cdot A \cdot B \cdot \cos(\beta) = |A + B \cdot e^{j\beta}|^2$$

$$p(H) + p(T) \pm 2 \cdot \sqrt{p(H)} \cdot \sqrt{p(T)} \cdot \cos(\beta) = |\sqrt{p(H)} + \sqrt{p(T)} \cdot e^{j\beta}|^2 = 1$$

Fig. 1 In this non-right-angled triangle, in our case of coin tosses, 'C' still must equal 1. 'A' represents the probability of tossing heads as registered by the first observer $p(H)$, and 'B' is the probability of tossing tails as registered by the second observer $p(T)$. The angle β models the errors of the observers.

Let us now examine the parallel between signal breakup into harmonic components and probability

theory. Probability values have analogies to energies and can be modeled as the squares of the values assigned to individual phenomena (concrete values). By the square roots of the probability of event phenomena, one may interpret how dominant a given phenomenon is in a random process, or how often the phenomenon occurs. In this conception, phase indicates the degree of structural links between the individual phenomena, i.e. by analogy the shift with respect to the defined beginning. This beginning may be a phenomenon with a zero phase, to which we relate all of the structural links of the other phenomena.

2 Quantum entanglement

We can take a closer look at *entanglement* phenomenon using our simple example with coin tossing. Suppose we have two parallel systems, and that the outcome of each system is either throwing heads or tails. If the probability from measuring the first system is a fifty-percent chance of throwing heads or a fifty-percent chance of throwing tails, the output from measurement of the second system is determined entirely by the value of measurement of the first system. Let us imagine that thanks to a complex wave probability function, a situation might arise where we would be monitoring the probability of the union of several phenomena, i.e. that either the first phenomenon will occur, OR the second will occur, OR the third will not occur etc., and that this probability works out to equal zero. Naturally, this situation cannot arise under the classical theory of probability, because there probabilities are merely added together, and at the most, repeating overlaps of phenomena are subtracted. In the newly introduced area of complex wave probability functions, there can also occur through the influence of the existence phases the subtracting of probabilities, and under certain conditions it is possible to find such a constellation of phenomena, that their union works out to zero probability. This, however, automatically means that the inversion phenomenon (intersection) for the given union (in our case, this inversion phenomenon would mean that the first phenomenon does not occur, AND at the same time the second phenomenon does not occur, AND the third phenomenon does occur) will occur with 100% probability, regardless of how the phenomena are arranged spatially.

In my opinion, quantum entanglement is caused by the resonance of complex wave functions. Among the ways this resonance manifests itself is that thanks to it we arrive from a purely probabilistic world to a

completely deterministic world, where there is disruption of the probabilistic characteristics of various phenomena, and the links between the entangled phenomena become purely deterministic events that even show up in different places, and for that reason they are also often designated as spatial distributed system states. Similarly, one may arrive at the conclusion that thanks to the principle of resonance, selected distributed states absolutely cannot occur in parallel, and this leads to an analogy with the Pauli exclusion principle.

3 Why is the word quantized

Let us now assume that we have a certain random phenomenon characterized by its amplitude and phase function, where these functions typically depend on time or on space. For the sake of clarity, let us assume that our function does not depend on time, but depends rather only on its (x,y) position, i.e. a selected point in an x-y plane. We can also imagine every point (x,y) in the plane in polar coordinates as a module and phase assigned to that point.

If we change the phase in an x-y plane by multiples of 360 degrees, we are still at our point (x,y), so the resultant amplitude and phase probability function assigned to the point (x,y) must remain the same for all variant multiples of 360 degrees. If we accept this completely logical requirement for the unambiguousness of the assignment of complex probability to a given point (x,y), we arrive at an important condition for our wave probability function.

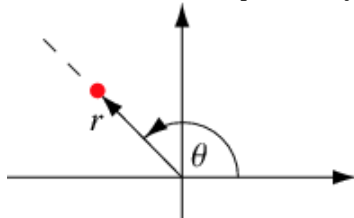


Fig. 2 The representation of a point (x,y) with polar coordinates using module r and phase θ . Ambiguousness of phase means that for all natural numbers 'n' there is also an identical point for the phase $(\theta + n \cdot 360)$.

It now suffices to imagine as a trivial example that the phase function of wave probability linearly depends on the phase representation of the position of our (x,y) point in polar coordinates. This means that the phase parameter of a probability function 'k' is given by the multiple of the spatial phase assigned to the point (x,y). If the linear multiplying constant 'k' is not a natural number, in the phase of the probability wave function we get 'k' multiples of 360 degrees, which lead each

time to a different value. These different values of the phase of the wave function mean the ambiguous assigning of a probability wave function to point (x,y). Only in the case that 'k' is a natural number do 'k' multiples of 360 such as 720 or 1080 represent the same value of the phase of the wave probability function.

By the extension of the ideas stated above from an (x,y) plane to an (x,y,z) space, there arises a similar analogy of unambiguousness, but it has more degrees of freedom and leads to expansion of the quantization of spin. On an (x,y) plane, whole numbered quantum spin or multiples of the constant 'k' were sufficient, but in an (x,y,z) space, we must introduce halved number values of quantum spin (multiplying the constant 'k'). The principle of quantization and its justification, however, are identical.

Using the conclusions defined above, it is just one more step to define phenomena that have similar properties, such as *fermions* or *bosons* in quantum physics. Bosons are characterized by whole-numbered spin, while fermions have spins with halved numbers, so our constant 'k' corresponds either to the whole numbers 1,2,... or to halved numbers such as 1/2, 3/2, etc.

4 Conclusion

The quantum theory of systems enables a description of complex information systems where information fermions can create knowledge structures and information bosons secure the information flow with possibilities for the arising of mutual resonances of wave probability functions, which lead to the formation of quantum processes (entanglement among individual phenomena). This principle can introduce new approach into modeling of complex brain networks.

References

- [1] Svítek, M.: Quantum System Theory – Principles and Applications, VDMVSG Saarbrücken, 140 pages, ISBN: 978-3-639-23402-2 (in English).
- [2] Svítek M: .Wave probabilistic functions for quantum cybernetics, 2012, IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS, PART C-APPLICATIONS AND REVIEWS, Volume: 42, Issue: 2 , pp. 233-240, 2012. (in English)
- [3] Svitek M.: Applying Wave Probabilistic Functions for Dynamic System Modeling, IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS, PART C-APPLICATIONS AND REVIEWS, Volume: 41, Issue: 5, pp. 674-681, 2011 (in English).

System alliances models for prefrontal cortex operation

Z. Votruba⁽¹⁾, M. Novák⁽¹⁾

Faculty of Transportation Sciences,
Czech Technical University of Prague ⁽¹⁾
Votruba@fd.cut.cz, Mirko@fd.cvut.cz

Many people suffer from influences of various unpleasant psychical stages, resulting in their non-adequate behavior. System models of brain parts based on alliance approach can help to explain respective mechanisms.

1 Introduction

Among unpleasant brain stages dominate the fear, anxiety, stress and aggressive behavior. All these stages cause a lot of problems and especially in transportation activities, which are necessary for each human community survival result in many accidents and catastrophes leading to in extreme high losses of lives, health and money.

The last results presented in literature (see [1,2] e.g.) show that while at normal situation the tendency to fall in some of these problematic stages is minimized by control of the prefrontal neo-cortex activity of our brain, the influence of some stressing factors impacting on the subject can disturb the brittle balance existing in prefrontal neo-cortex. Then some internal evolutionary much older parts of brain, namely the hypothalamus and amygdale, become prevailing which can result in some kind of non-rational, emotive and principally dangerous reactions.

2 Artificial alliances operation

System alliances stand in front of interest of system theorists since the threshold of millennium. Alliance of two or more parts / modules originates as:

- a product of random encounter
- an outcome of processes of contamination and immunity
- a construct, see e.g. [6]

The concept of system alliances appeared to be very useful for understanding the function and operation of certain important complicated heterogeneous complexes, namely in transportation area. Later on we have tried to solve some problems, related to such system sensitivity of their parameter changes, to their reliability in interaction with human subjects and to their operation dynamics. Some reached results can be found e.g. in [9].

This approach can also be used for an improvement of our understanding of a function of complex neural networks and other information structures operating within human body and for modeling these objects see [8]. It also allows tackling the aspects of time and other independent variables and the aspects of sensitivity of respective alliance functions to the changes of their parameters as well.

The principles of the alliance forming have been explained through utilizing the concepts of information power (IP) and of multilingual translation efficiency respectively. An illustration of basic phenomena resulting in the emergence of alliance could be based on the concepts of interface sharing (IS), and of irregularities conjugation (IC) as well, see [7] e.g.

The language model of the alliance makes the quantitative evaluation of processes possible. The basic ideas of the respective construct are as follows:

Any module/system taking part in the alliance is expressed in respective multi-grammar language. The relations within these modules are expressed through mutual translations of particular languages of modules elements.

The alliances of artificial systems have to be resistant to various kinds of troubles, resulting from the probability that a certain object (system, system alliance or the whole network of system alliances) can fail in the period of its operation. This is expressed by various hazards and risks. Depending on the particular risk analyzed, the measurement of risk can be carried out with a greater number of different variables and factors, depending inter alia on the complexity of the chain of impact, on the number of impact factors, and also on the desired level of precision.

The most complicated artificial systems operate usually as a system alliance (objects), consisting of a set of partial systems, the interaction and mutual cooperation of which is supervised by one or more alliance information tools (AIT). All the partial systems can have either their own inputs and outputs or two or more of them can share some common input/outputs ports (places in which external signals and stimuli can enter into the respective system or in which some output signal can be reached). Besides this, in the respective alliance structure, the AIT creates an area (one or more) where the general control activity is provided and where also the necessary general data are stored.

The detail structures of system alliances can differ significantly, of course, being influenced namely by the kind and purpose of their operation.

The operation of all artificial system alliances can however, be expressed by the set F of N by them realized system functions F_k ,

$F = \Xi \{F_k\}, k=1 \dots K$. The operator Ξ represents the actual kind of composition of partial system functions F_k into the set F , representing the properties and operation of the whole alliance.

Each system function F_k depends on three main phenomena:

- the system structure S_p ,

- the vector X of system parameters $x_i, i = 1 \dots N$
- the vector of independent variables $P = \{p_j\}, j = 1 \dots J$.

The system parameters x_i are supposed usually to have real values only (if they are vectors we suppose that they can be separated in real vector components).

Each component of a system vector value depends on independent variables P . The components p_j of independent variable vector P can be of various physical nature representing the influence of temperature, atmospheric pressure, humidity, level of radiation, electric or magnetic field intensity, gravity, etc. All these variables except one can be reversible, i.e. we can influence their values. The only exception is time $p_j = t$, which is absolutely irreversible. Time is the only absolute independent variable on which we have no influence.

3 Brain structures dominant for alliance models

Brain as the whole represents an extremely complex system alliance or the whole network of system alliances. At present our knowledge does not allow to construct the models of the whole brain, however some specific parts can be modeled considerably well.

Fig.1 shows the very simplified basic structure of system alliance model representing the interaction among the pre-frontal neo-cortex (PFNC), evolutionary ancient brain parts (APB), mainly the hypothalamus (HT) and amygdale (A), thalamus (T) and visual cortex (VC). These parts interaction was found to be significant for analysis of appearance of fear, anxiety and aggressive behavior, which all can be a reason for various accidental stages of road vehicle drivers. Among these brain parts many feed-forward and feed-back paths exist (solid arrows). Through thalamus and out of pre-frontal neo-cortex input signals (EI) and outputs to effectors function (green and red arrows).

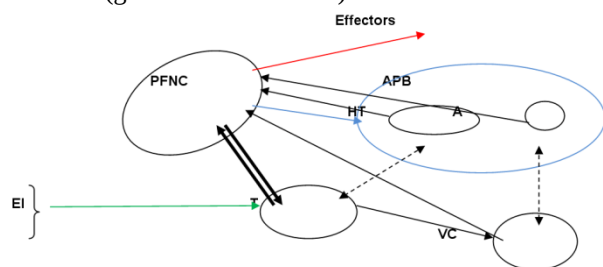


Fig1. Main interactions among pre-frontal neo-cortex PFNC and ancient parts of brain (APB).

Among all such alliance parts the signal communication is provided by interactions which are realized through specific interfaces, mainly known as synapses.

The prefrontal neo-cortex and the APB parts operate mainly in brittle balance, in which the emotive aspects typical for APB is kept at acceptable low level by the rationally based functions of prefrontal neo-cortex.

However, when the particular subject is exposed to influence of significant stress, fear, mood or injury, the emotive activities of APB can prevail and the respective human subject can fall into panic, fairy or aggressive behavior modes. These can be very dangerous – according to our estimation can cause about 25% from all by human factor coming accidents.

4 Time dependences and open problems

When modeling these brain system alliances functions, one has however to take into account their time dependences and the influences of various independent variables. The whole regulation procedures sketched in Fig.1 have various dynamic phases, in which either the rational or emotive components can temporary dominate. The system operates therefore in dynamic stable regime, moving in the respective parameter space between various attractors, among them the selection of the actual one can be influenced by various external physical and/or psychical factors. The people influenced by strong fears, stress, physical or psychical mood can have higher tendency to fall is some of the unstable positions.

The modeling of the very complicated system alliances operating prefrontal neo-cortex and APB parts is still in very beginning. This concerns especially the problems of such alliances stability and their time dynamics. The natural multidimensionality of such models requires very careful reduction of the number of parameters which have to be taken into account in particular model. Too simple models suffer from to be too far from reality, while too complicated ones are hard to analyze.

6 References

- [1] Arnsten A., Mazure C.M., Sinha R.: This is your brain in meltdown, Scientific American, vol. 306, April 2012, No.4, 36-43
- [2] Novák M., Votruba Z., Brandejský T.: Alliances in heterogenous networks (in Czech: Aliance v heterogenních sítích), res. Rep. No. V – 1003/2007, and LSS 315/07, Prague, 2007
- [3] Novák M., Votruba Z.: Dynamics of Sensitivities in Transportation Systems Alliances, Research report No. V – 1020/2008, and LSS 325/08, Prague, July 2008
- [4] Wimzig C., Schneider S., Schoerner G.: The time course of saccadic decision making: Dynamic field theory, Neural Networks, 19, 2006, 1059-1074
- [5] Frank M.J.: Hold your horses: A dynamic computational role for the sub-thalamic nucleus in decision making, Neural Networks, 19, 2006, 1120-1136
- [6] Vlček J.: Fabric of Constructive Theory of Hybrid Systems (in Czech: Návrh struktury konstruktivní teorie hybridních systémů), Vlček's Seminars CTU, Faculty of Transportation Sciences, CTU, Prague, May 2001.
- [7] Novák M., Votruba Z.: Complex Uncertain Interfaces, Conf. ACMOS '05, March, 13-15, 2005, Prague
- [8] Votruba, Z. - Novák, M. - Brandejský, T. - Fábera, V. - Bouchner, P. - et al. Theory of System Alliances in Transportation Science Praha: Ústav Informatiky AV ČR, v.v.i., 2009. 162 p. ISBN 978-80-87136-08-9.
Svítek, M. - Votruba, Z. - Moos, P. Towards Information Circuits In: Neural Network World. 2010, vol. 20, no. 2, p. 241-247. ISSN 1210-0552.

О фундаментальной природе пространственных представлений в мозге

Цукерман В.Д., Еременко З.С., Кулаков С.В.

НИИ нейрокибернетики им. А.Б. Когана Южного федерального университета
vdts@krinc.ru

In the report neural correlates formations of spatial representations in a brain and conceptual neurodynamical model of spatial processing allocentric and egocentric signals of an environment, on the basis of neural network architecture are presented. For the first time it is shown that interferential oscillator and neuron network attractor models, can be incorporated for generation of phenomena of cell's spatial activities in the neuron network organization.

1 Введение

В последние годы в гиппокампальной формации мозга идентифицирован целый ряд пространственно активных клеток (*place cells*, *grid cells*, *border cells* и др.) в условиях свободного поведения животных. В ряде работ указывается на то, что пространственно настроенные нейроны, в частности, *grid cells* могут производить свою позиционную активность из фазовой интерференции собственных тета-частот осцилляций, которые модулируются скоростью и направлением трансляционных движений [1–2, 6, 9]. Сообщается, что в соответствии с осцилляторной интерференцией частота пачек тета клеток варьирует как функция косинуса направления движения крысы и эта дирекциональная настройка подвергается влиянию ориентационных источников [8, 11–12].

Другая точка зрения основана на непрерывных аттракторных моделях, в которых относительная пространственная фаза активности клеток устанавливается и стабилизируется аттракторной динамикой, которая приводит к формированию паттерна [3, 5, 10]. В этом классе моделей паттерны активности формируются в сети рекуррентно связанных, периодически активных нейронов пропорционально скорости и направлению движения животного. Таким образом, несмотря на то, что в обоих классах моделей решетчатые паттерны возникают посредством интегрирования сигналов скорости и направления движения, механизмы получения пространственной периодической активности клеток у них различны [7].

Ранее нами предложена модель и опубликованы результаты вычислительного моделирования свободно масштабируемых сетей с четным

циклическим торможением (ECI-networks, подробное описание см. в работе Цукерман с сотр., [13]), объединяющих в себе механизмы генерации пространственной активности клеток обоих классов моделей и, одновременно, лишенных недостатков каждой из них. Прежде всего, в отличие от существующих частотных моделей кодирования пространственных переменных, нами предложен частотно-фазовый механизм, обладающий значительно более высокой временной и пространственной точностью кодирования, а главное биологически более обусловленным характером обработки сигналов в аллоцентрической и эгоцентрической системах координат. В настоящем докладе, нами впервые представлена концептуальная модель микросхемной сетевой организации, реализующей различные типы пространственно активных нейронов, являющихся нейронными коррелятами пространственного познания.

2 Концептуальная модель формирования пространственных представлений

Важнейший вопрос формирования пространственных представлений в мозге заключается в генезисе и функциональном взаимодействии многочисленных пространственно активных клеток гиппокампальной формации. В докладе представлена концептуальная модель различных уровней организации нейронов, в формировании пространственных представлений в мозге. Модель имеет архитектуру нейронной сети с четным циклическим торможением (ECI-сети) [13]. Сеть представляет собой активную среду слабосвязанных нелинейных осцилляторов, генерирующих гамма-пачки осцилляций мембранного потенциала в широкополосном диапазоне низкочастотных тета-ритмов. Рекуррентные тормозные связи между осцилляторами, образуют многочисленные квадратные решетки и осцилляторные кольца многослойной сети. ECI-сеть обладает двумя подсистемами осцилляторных единиц: референтной и информационной, чередующихся между собой в определенном порядке, и реализует частотно-фазовый способ кодирования сигналов пространственного окружения. Ниже приводятся основные положения, лежащие в основе

концептуальной модели пространственной обработки сигналов:

- 1) нейронные представления основываются на векторах преимущественных направлений;
- 2) природа сигналов, используемых в интегрировании траектории пространственного движения — угловые и линейные (трансляционные) сигналы скорости;
- 3) скорость и направление трансляционных движений модулируют частоту информационных осцилляторов в диапазоне тета-частот;
- 4) глубина тета-частотной модуляции осцилляторов, модулируемых скоростью (ОМС) зависит от двух параметров: длины вектора преимущественного направления и скорости пространственного движения;
- 5) кривые фазовых ответов ОМС определяют аллоцентрические и эгоцентрические представления информационных единиц ЕСИ-сети;
- 6) каждое местоположение организма в пространстве кодируется специфическим паттерном фазовых отношений между информационными осцилляторами;
- 7) фазовая интерференция информационных осцилляторов приводит к формированию функции огибающей, которая независима от траектории пространственного движения организма.

3 Нейродинамические корреляты пространственных представлений

Вычислительные экспериментальные исследования концептуальной модели, описанной выше, показали ряд нейродинамических феноменов, характеризующих клеточный (*place cells*, *grid cells*, *head direction cells* и т.д.), групповой (фазовые квартеты), ансамблевый (дирекциональные и топологические) и сетевой уровни организации нейродинамических коррелятов пространственных представлений. Показано, что: а) информационные ОМС фазового квартета могут иметь вектор преимущественного направления одной и той же ориентации, одинаковой длины, но различную фазовую настройку; б) информационные ОМС различных фазовых квартетов, могут иметь различные вектора преимущественных направлений и различных длин; в) ансамбли информационных единиц независимо от скорости и направления

пространственного движения в среде могут иметь один и тот же пространственный период (шаг), одинаковую ориентацию, но различные пространственные фазы; г) таких фаз три, и они сохраняются в процессе движения, что приводит к формированию пространственных гексагональных решеток с тремя осями симметрии; это подтверждается известными данными о том, что энторинальная кора у людей, в целом более активна, когда мы движемся «вперед» либо «назад» вдоль трех главных осей симметрии наших энторинальных решеток [4].

В заключении отметим, что эгоцентрические и аллоцентрические представления об окружающем пространстве существуют в мозге параллельно, комбинируясь при решении поведенческих задач. В наших модельных экспериментах удалось показать возможность одновременных представлений обоих типов в одной и той же нейронной сети.

Список литературы

- [1] N. Burgess, C. Barry et al. *Hippocampus*, 2007, **17**, 801–812.
- [2] N. Burgess. *Hippocampus*, 2008, **18**, 1157–1174.
- [3] Y. Burak, I.R. Fiete. *PLoS Comput Biol.*, 2009, **5**, e1000291.
- [4] C. Doeller, C. Barry et al. *Nature*, 2010, **463**, 657–661.
- [5] M.C. Fuhs, D.S. Touretzky. *J. Neurosci.* 2006, **26**, 4266–4276.
- [6] C. Geisler, D. Robbe et al. *PNAS USA*, 2007, **104**, 8149–8154.
- [7] L.M. Giocomo, M.-B. Moser et al. *Neuron*, 2011, **71**, 589–603.
- [8] M.E. Hasselmo, L. Giocomo et al. *Hippocampus*, 2007, **17**, 1252–1271.
- [9] A. Jeewajee, C. Barry. *Hippocampus*, 2008, **18**, 1175–1185.
- [10] B.L. McNaughton, F.P. Battaglia et al. *Nat.Rev. Neurosci.*, 2006, **7**, 663–678.
- [11] E.A. Zilli, M.E. Hasselmo. *J. Neurosci.*, 2010, **30**, 13850–13860.
- [12] A.C. Weldon, I.G. Shlifer et al. *J. Neurosci.*, 2011, **31**, 16157–16176.
- [13] В.Д. Цукерман, Еременко З.С., Каримова О.В., Сазыкин А.А., Кулаков С.В. «Математическая биология и биоинформатика». Пуцзино, РАН, 2012, **7** (1), 87–124.

Анализ устойчивости режимов синхронизации для одной системы фазовых осцилляторов

Я.Б. Казанович¹, О. Бурилко², Р.М. Борисюк^{1,3},

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

²Институт математики Национальной академии наук Украины
Терещенская 3, Киев, 01601, Украина

³Университет г. Плимут, Плимут, PL4 8AA, Великобритания

Рассматривается система фазовых осцилляторов Курамото с радиальной архитектурой связей (сеть с центральным элементом). Исследуются условия, при которых периферические осцилляторы "соревнуются" за синхронизацию с центральным осциллятором. Обсуждается применение полученных результатов для моделирования внимания.

Многие явления, связанные с синхронизацией колебательной активности в физике, химии и биологии могут быть промоделированы с помощью взаимодействующих фазовых осцилляторов Курамото (см., например, обзорные статьи [1,2]). Мы рассматриваем специальную систему фазовых осцилляторов, используемую для моделирования внимания и других функций мозга [3-6]. Это так называемая система с центральным элементом, известная также как система с радиальной архитектурой связей. В такой системе взаимодействие между осцилляторами реализуется через центральный осциллятор (ЦО), имеющий прямые и обратные связи со всеми остальными элементами сети, которые называются периферическими осцилляторами (ПО).

Полный бифуркационный анализ системы, содержащей два или три ПО был предпринят нами в работе [7]. В данной работе мы исследуем устойчивость режима, при котором ПО "соревнуются" за синхронизацию с ЦО при произвольном числе ПО. В результате соревнования k ПО синхронизируются с ЦО, в то время как остальные $n-k$ ПО работают в противофазном режиме по отношению к ЦО. Основным результатом заключается в аналитическом описании областей параметров, в которых режим соревнования реализуется при произвольных n и k .

Пронумеруем осцилляторы индексами $i = 0, 1, \dots, n$, где 0 соответствует ЦО, а $1, \dots, n$ являются индексами ПО. Динамика системы описывается следующими ОДУ:

$$\frac{d\theta_0}{dt} = \omega_0 + \sum_{i=1}^n f(\theta_i - \theta_0), \quad (1)$$

$$\frac{d\theta_i}{dt} = \omega_i + g(\theta_0 - \theta_i), \quad i = 1, \dots, n, \quad (2)$$

где $(\theta_0, \theta_1, \dots, \theta_n) \in T^{n+1}$ - фазовые переменные на $(n+1)$ -мерном торе, $\theta_i \in [0, 2\pi)$, ω_i - собственные частоты осцилляторов, $f(x), g(x)$ - функции взаимодействия: нечетные, непрерывные, 2π -периодические, удовлетворяющие условиям

$$g(\pi) = f(\pi) = 0, \quad i = 1, \dots, n.$$

Вычитая (1) из (2), получаем

$$\frac{d\varphi_i}{dt} = \Delta_i - \sum_{j=1}^n f(\varphi_j) - g(\varphi_i), \quad i = 1, \dots, n, \quad (3)$$

где $\varphi_i = \theta_i - \theta_0$, $\Delta_i = \omega_i - \omega_0$. Ограничимся исследованием (3) для случая одинаковых собственных частот

$$\omega_0 = \omega_i = \omega, \quad \Delta_i = 0, \quad i = 1, \dots, n.$$

Обозначим

$$f'(0) = a_1, \quad f'(\pi) = a_2, \quad g'(0) = b_1, \quad g'(\pi) = b_2.$$

Будем предполагать, что $n \geq 2$.

Точки $\Phi = (\varphi_1, \dots, \varphi_n)$ с координатами $\varphi_i \in \{0, \pi\}$ ($i = 1, \dots, n$) являются стационарными точками системы (3) ((3) может иметь и другие стационарные точки). Наша цель состоит в том, чтобы в терминах параметров a_1, a_2, b_1, b_2 найти условия асимптотической устойчивости точек Φ_k с k координатами, равными 0, и $(n-k)$ координатами, равными π . Линеаризуя (3), можно вычислить характеристические значения для точек Φ_k :

$$\begin{aligned} \lambda_{1, \dots, n-1} &= -b_2, \quad \lambda_n = -na_2 - b_2 \quad (k=0), \\ \lambda_{1, \dots, n-2} &= -b_2 \quad (n \geq 3), \quad \lambda_{n-1, n} = z \quad (k=1), \\ \lambda_{1, \dots, k-1} &= -b_1, \quad \lambda_{k, \dots, n-2} = -b_2, \quad \lambda_{n-1, n} = z \\ &\quad (2 \leq k \leq n-2), \end{aligned}$$

$$\lambda_{1,\dots,n-2} = -b_1 \quad (n \geq 3), \quad \lambda_{n-1,n} = z \quad (k = n-1),$$

$$\lambda_{1,\dots,n-1} = -b_1, \quad \lambda_n = -na_1 - b_1 \quad (k = n),$$

где

$$z = -z_1 \pm \sqrt{z_2^2 + z_3},$$

$$z_1 = [na_2 + k(a_1 - a_2) + (b_1 + b_2)] / 2,$$

$$z_2 = [-na_2 + k(a_1 + a_2) + (b_1 - b_2)] / 2,$$

$$z_3 = k(n-k)a_1a_2.$$

Условия асимптотической устойчивости заключаются в требовании, чтобы все характеристические значения были меньше 0. Бифуркационные поверхности, ограничивающие области асимптотической устойчивости, имеют вид

$$L_1: kb_2a_1 + (n-k)b_1a_2 + b_1b_2 = 0,$$

$$L_2: ka_1 + (n-k)a_2 + (b_1 + b_2) = 0$$

Архитектуры с центральным элементом типичны для организации взаимодействия структур мозга. В частности, система внимания предположительно функционирует как система с центральным элементом [8]. Недавние экспериментальные данные показывают, что механизм взаимодействия между центральным управляющим элементом (ЦУЭ) системы внимания и первичными зонами коры основан на синхронизации (на частоте гамма-ритма) активности между ЦУЭ и нейронными ансамблями, представляющими объекты в фокусе внимания [9]. Вне фокуса внимания такая синхронизация отсутствует. Полученные в данной работе результаты позволяют целенаправленно выбирать параметры системы (1)-(2) при ее использовании в качестве модели внимания, чтобы получить нужный режим синхронизации. В случае $k=1$ мы получаем осцилляторный аналог процедуры winner-take-all, а

$k > 1$ может соответствовать случаю разделенного внимания. Дальнейшее развитие исследований связано со снятием ограничения на одинаковость собственных частот ЦО и ПО.

Список литературы

- [1] Acebron J.A. et al. The Kuramoto model: a simple paradigm for synchronization phenomena. *Rev. Modern Physics*, **77**, 137-185, 2005.
- [2] Arenas A. et al. Synchronization in complex networks. *Physics Reports*, **469**, 93-153, 2008.
- [3] Крюков В.И. Модель внимания и памяти, основанная на принципе доминанты и компараторной функции гиппокампа // *Журнал высшей нервной деятельности*, **54**, №1, 10-29, 2004.
- [4] Borisyuk R.M., Kazanovich Y.B. Oscillatory neural network model of attention focus formation and control // *BioSystems*, **71**, 29-38, 2003.
- [5] Borisyuk R., Kazanovich Y. Oscillatory model of attention-guided object selection and novelty detection // *Neural Networks*, **17**, 899-915, 2004.
- [6] Borisyuk R., Chik D., Kazanovich Y. Visual perception of ambiguous figures: Synchronization based neural models. *Biological Cybernetics*, **100**, 491-504, 2009.
- [7] Burylko O., Kazanovich Y., Borisyuk R. Bifurcations in phase oscillator networks with a central element. *Physica D*, **241**, 1072-1089, 2012.
- [8] Baddeley A. Exploring the central executive // *Quarterly Journal of Experimental Psychology*, **49A**, 5-28, 1996.
- [9] 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.

Корреляционный анализ данных магнитной энцефалографии

Лыжко Е.В., Махортых С.А.

Институт математических проблем биологии РАН, Пущино, Россия
lyzko@yandex.ru

This paper considers the problem of finding diagnostic features for data analysis of magnetic encephalography (MEG) healthy subjects (control group) and patients with Parkinson's disease. For this purpose, the coefficients of spherical functions were calculated for the measured MEG data at each time. Correlation matrices were constructed for these coefficients and the number of well-correlated coefficients is calculated for a healthy subjects and a patient with Parkinson's disease. It was shown that the number of well-correlated coefficients for these groups of subjects is significantly different.

1 Введение

В настоящее время для изучения работы мозга и диагностирования заболеваний используют структурные (рентгенография, рентгеновская компьютерная томография, магнито-резонансная томография) и функциональные методы (электроэнцефалография, магнитоэнцефалография, электрокортикография, реоэнцефалография). В данной работе для выявления диагностических признаков заболевания будут использоваться данные магнитоэнцефалографии (МЭГ).

Отклонения энцефалограммы от нормы, выявляемые при типичных расстройствах ЦНС, как правило, не обладают выраженной нозологической специфичностью (за исключением эпилепсии) и чаще всего при визуальном анализе сводятся к нескольким основным типам [1].

Изменения в записях МЭГ происходят и при нормальной нейрофизиологической активности мозга, связанной с переключением внутренних состояний в соответствии с текущей деятельностью и анализом ЦНС внешних воздействий. Помимо изменений во временной структуре сигналов меняются также их пространственные характеристики. Возникают нарушения нормальной пространственной структуры распределения поля, например, грубая межполушарная асимметрия. Разрабатываются специальные методы обработки сигналов для диагностики патологий [1-3,5].

2 Построение корреляционных матриц

В работе рассматривается задача поиска диагностических признаков для анализа данных МЭГ здоровых испытуемых (контрольная группа) и

пациентов с заболеванием Паркинсона. С этой целью был предложен корреляционный подход анализа пространственно-временных характеристик сигнала.

2.1 Построение корреляционной матрицы по массиву измерительных датчиков

В первом случае построение корреляционной матрицы осуществлялось для 148 точек, в которых располагались датчики, и для 10173 временных отсчетов.

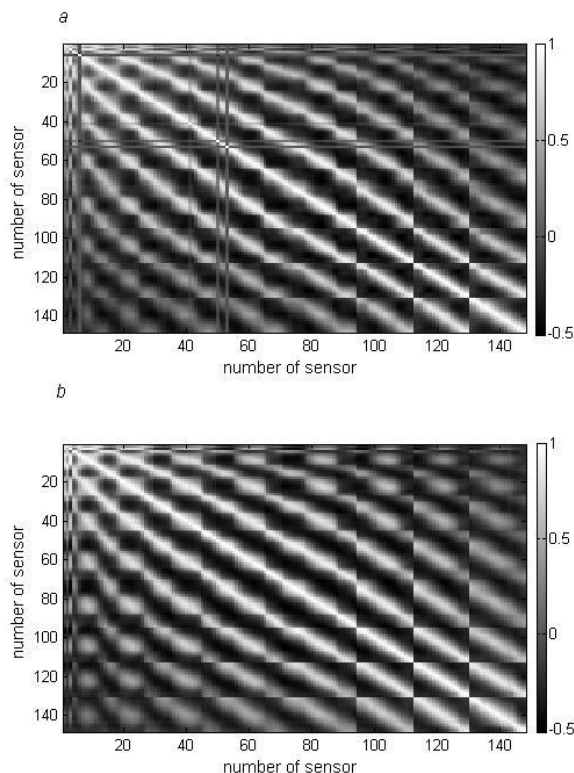


Рис.1. Корреляционная матрица измеренного сигнала а) для здорового испытуемого ($N_{corr}=1372$), б) для пациента с заболеванием Паркинсона ($N_{corr}=1650$)

На рис.1 представлена матрица парных коэффициентов корреляции:

$$r_B = \begin{pmatrix} 1 & r_{B^1 B^2} & \dots & r_{B^1 B^{148}} \\ r_{B^2 B^1} & 1 & \dots & r_{B^2 B^{148}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{B^{148} B^1} & r_{B^{148} B^2} & \dots & 1 \end{pmatrix}$$

где $r_{B^i B^j}$ - коэффициент корреляции между i и j столбцами измеренных данных.

Вычисление коэффициента парной корреляции $r_{B^i B^j}$ осуществляется по следующей формуле:

$$r_{B^i B^j} = \frac{\sum_{k=1}^N [(B_k^i - \bar{B}^i)(B_k^j - \bar{B}^j)]}{\sqrt{\sum_{k=1}^N (B_k^i - \bar{B}^i)^2 \sum_{k=1}^N (B_k^j - \bar{B}^j)^2}}$$

где B_k^i - значение магнитной индукции, измеренное на i датчике в k момент времени;
 $\bar{B}^i$ - среднее значение магнитной индукции, измеренное на i датчике для N временных отсчетов;
 $k = \overline{1, 148}$ - номер временного отсчета;
 $i = \overline{1, 148}$, $j = \overline{1, 148}$ - номера измерительных датчиков.

2.2 Спектрально-аналитический метод описания данных

Для второго случая были рассчитаны коэффициенты сферических функций Лежандра.

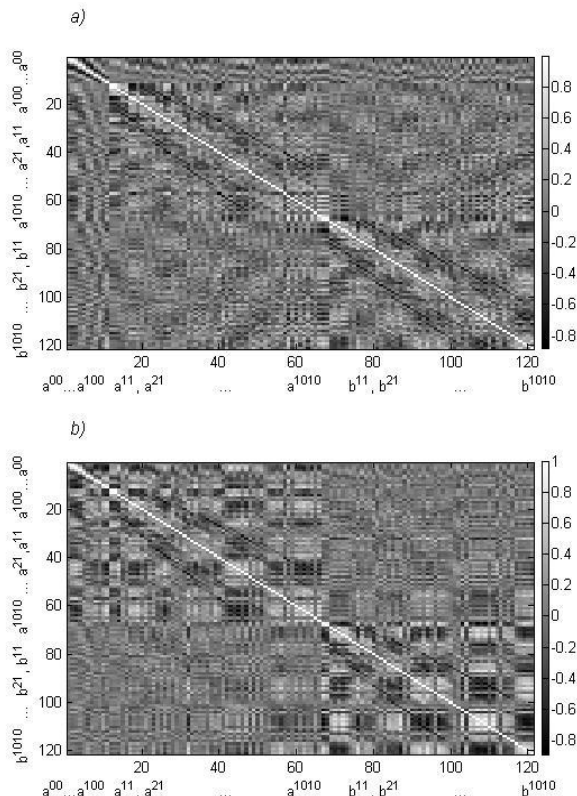


Рис.2. Корреляционная матрица коэффициентов разложения (a_{nk}, b_{nk}) измеренного сигнала а) для здорового испытуемого ($N_{corr}=88$) б) для пациента с

заболеванием Паркинсона ($N_{corr}=264$)

Расчет спектрально-аналитических характеристик данных (a_{nk}, b_{nk}) производился по известным формулам [4,6]. Коэффициенты разложения (a_{nk}, b_{nk}) вычислялись на 10173 временных отсчетах, для которых строились корреляционные матрицы (121x121). На рисунке 2 представлена матрица парных коэффициентов корреляции:

$$r_{ab} = \begin{pmatrix} 1 & r_{a^{00}a^{10}} & \dots & r_{a^{00}a^{11}} & \dots & r_{a^{00}b^{1010}} \\ r_{a^{11}a^{00}} & 1 & \dots & r_{a^{11}b^{11}} & \dots & r_{a^{11}b^{1010}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{a^{1010}a^{00}} & r_{a^{1010}a^{11}} & \dots & r_{a^{1010}b^{11}} & \dots & r_{a^{1010}b^{1010}} \\ r_{b^{11}a^{00}} & r_{b^{11}a^{11}} & \dots & 1 & \dots & r_{b^{11}b^{1010}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{b^{1010}a^{00}} & r_{b^{1010}a^{11}} & \dots & r_{b^{1010}b^{11}} & \dots & 1 \end{pmatrix}$$

3 Выводы

Корреляционный анализ данных МЭГ позволяет получить усреднение оценки пространственно-временного распределения областей, связанных с источниками активности. Отличительной особенностью для пациента с заболеванием Паркинсона является наличие существенной асимметрии в корреляции коэффициентов сферических функций a_{nk} и b_{nk} (рис. 2б), в отличие от результатов, полученных для здоровых испытуемых (рис. 2а). Для обоих случаев, представленных на рисунках 1 и 2, количество хорошо коррелирующих коэффициентов для здорового испытуемого меньше, чем их число для пациента с заболеванием Паркинсона.

Работа поддержана проектами РФФИ 11-07-00519 и 10-01-00609.

Список литературы

- [1] А.П. Бердникова, С.А. Махортых, С.А. Семечкин. Анализ нейрофизиологической активности по данным магнитной энцефалографии. Нейроинформатика, 2010, часть 1, с. 318-328.
- [2] А.В. Дергузов, С.А. Махортых. Комплексная диагностика паркинсонизма по данным магнитной энцефалографии. Электронный журнал "Исследовано в России", 2006, 65, с. 646-659.
- [3] А.В. Дергузов, С.А. Махортых. Распознавание патологической активности в записях магнитных энцефалограмм при болезни Паркинсона. Электронный журнал "Исследовано в России", 2005, 149, с. 1562-1573.
- [4] Г.Джеффрис, Б.Свирлс Методы математической физики. Вып. 3, т. 3, М.: Мир, 1970.
- [5] Л.И.Куликова, С.А.Махортых. Математические операции над двумерными сигналами в базисах сферических гармоник. Электронный журнал "Исследовано в России", 2006, 60, с. 598-608.
- [6] Р. Курант, Д. Гильберт Методы математической физики, т. 1, М.-Л., Гостехиздат, 1951.

Система прямого моделирования экспериментальных данных магнитной энцефалографии

Рыкунов С.Д., Сычев В.В., Устинин М.Н.

ИМПБ РАН

stanislavrykunov@gmail.com

This work is aimed at the creation of software to model complicated multidimensional time series to be obtained during the experiments studying the human brain. The software under development will provide the possibility to perform the whole-cycle computational experiments on magnetic encephalography and to predict their results, up to making a decision about the appropriateness of the particular expensive studies. Model encephalograms will permit to create and debug optimal protocols of data cleaning.

1 Задача исследования

В последнее десятилетие интерес научного сообщества к изучению механизмов работы мозга сильно возрос. Вследствие этого было создано достаточно большое число программных продуктов для анализа и моделирования данных ЭЭГ и МЭГ. Основным же направлением является именно анализ экспериментальных данных, в то время как моделированию уделяется меньшее внимание.

Далее представлен краткий обзор программных продуктов с возможностью прямого моделирования магнитных полей.

Brainstorm.[1] Программный комплекс Braistorm написан на связке Java/Matlab. В основном он предназначен для обработки и визуализации экспериментальных данных МЭГ и ЭЭГ. Для большей точности и информативности используется сопряжение данных с данными МРТ. Для моделирования магнитных используется модель точечного токового диполя в проводящей сфере [2]. Пакет обладает развитым графическим интерфейсом и возможностью подключения дополнительных пользовательских плагинов.

MRIAN. Программный комплекс для среды Matlab, разработанный в Институте Математических проблем биологии РАН. Он предназначен для обработки магнитоэнцефалограмм и локализации источников на магниторезонансной томограмме. MRIAN представляет собой графическую среду для визуализации и анализа магниторезонансных томограмм головного мозга и магнитоэнцефалограмм (МЭГ). Программный комплекс MRIAN может работать с тремя типами приборов: 148-канальным магнитометром и 275-ми и 7-ми канальными градиометрами.

Основным направлением работы является временной и спектральный анализ экспериментальных данных. Для решения обратной

задачи используется многократное решение прямой задачи по методу диполя в проводящей сфере [2,4]. Пакет обладает развитым графическим интерфейсом. Помимо обработки экспериментальных данных MRIAN может использоваться для моделирования магнитного поля в виде расчета мгновенных значений. Рассчитанные значения поля отображаются в виде карты напряженностей.

Настоящая работа посвящена созданию системы прямого моделирования магнитных энцефалограмм с целью разработки эффективных методов их анализа. Создаваемая система позволит проводить полные вычислительные эксперименты по магнитной энцефалографии с целью прогнозирования их возможных результатов и для формирования оптимальных протоколов очистки данных[3].

2 Методика

Модельная МЭГ формируется как сумма любого количества расчетных энцефалограмм, регистрируемых на модельном приборе, так как магнитные поля от разных источников аддитивны. В качестве спонтанной активности и внешних шумов могут с любой амплитудой использоваться экспериментальные записи, полученные на существующих или перспективных приборах.

В состав программы входит графическая среда для моделирования полезного сигнала. Под полезным сигналом здесь понимается последовательность магнитных полей, порождаемых изучаемым в эксперименте процессом в головном мозге (например, предполагаемым откликом на внешний стимул).

Графическая среда позволяет:

1. Задавать практически неограниченное количество источников поля, их расположение и амплитудно-временную зависимость. Данная зависимость может задаваться как внешним временным рядом, так и при помощи редактора внутри среды.
2. Решать прямую задачу магнитной энцефалографии, т.е. вычислять распределение магнитного поля, порождаемого заданными источниками, регистрируемое модельным прибором.
3. Моделировать экспериментальные данные. Моделирование сводится к созданию ряда указанной длины, по которому заданным образом

распределены полученные ранее решения прямой задачи.

4. Объединять смоделированные данные эксперимента с записями спонтанной активности мозга.

Внешний вид графической среды представлен на рисунке 1.

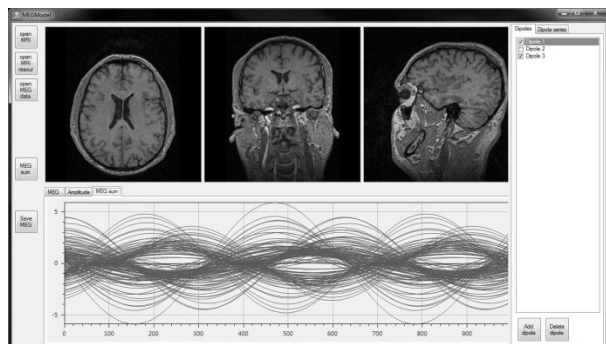


Рис.8. Основное окно программы с модельным полем, порождаемым двумя диполями.

3 Модели

Модели источников поля. При решении прямой задачи МЭГ используется модель точечного токового диполя в проводящей сфере [2]. Датчик, находящийся в точке $\mathbf{r}_k$ и имеющий направление $\mathbf{n}_k$ регистрирует от токового диполя амплитуды Q , находящегося в точке $\mathbf{r}_0$, магнитное поле B_k :

$$B_k(\mathbf{r}_0, Q) = \frac{\mu_0}{4\pi F^2} ((F(Q \times \mathbf{r}_0) - (Q \times \mathbf{r}_0, \mathbf{r}_k) \nabla F), \mathbf{n}_k),$$

где $F = a(a\mathbf{r}_k + \mathbf{r}_k^2 - (\mathbf{r}_0, \mathbf{r}_k))$, $a = \mathbf{r}_k - \mathbf{r}_0$, $a = |a|$, $\mathbf{r}_k = |\mathbf{r}_k|$, $|\mathbf{n}_k| = 1$, $\mu_0 = 4\pi \cdot 10^{-7}$

Модели структуры головного мозга. В качестве трехмерной модели головного мозга используются T1-взвешенные магниторезонансные томограммы. Эти томограммы охватывают всю голову, сняты с шагом около 1 мм и представляют структуру мозга с точностью, вполне достаточной для моделирования магнитных полей. Размещение модельных диполей в пространстве головы осуществляется интерактивно в графической среде пакета MRIAN [4].

Модели приборов. В качестве моделей приборов используется описание расположения и направления датчиков модельных приборов, заданные в координатной системе головы. Параметрами модельных приборов являются пространственное положение, количество катушек в датчиках и ориентация катушек. При этом в каждом вычислительном эксперименте используются параметры реальных приборов, для которых нужно моделировать данные. В их число входят: 148-канальный магнитометр [3] и 275-канальный градиометр первого порядка.

[18]

4 Реализация

Программа написана на языке C# с использованием математической библиотеки Math.NET Numerics. Использование этой библиотеки позволяет легко распараллеливать вычисления и применять архитектурно-оптимизированные пакеты вычислений, такие как IntelMKL.

5 Результаты

В результате работы был создан программный продукт для моделирования сложных многомерных временных рядов, описывающих распределение магнитных полей, порождаемых несколькими источниками. Для проверки результатов моделирования использовался программный комплекс MRIAN – бралась пара источников с известным пространственным расположением, для них задавались амплитудно-временные зависимости и рассчитывалось результирующее магнитное поле. Полученная в результате магнитная энцефалограмма загружалась в пакет MRIAN и для нее решалась обратная задача, т.е. проводилась локализация источников поля. В результате расчётов точность локализации составляла 1.0, что означает полное соответствие между заданным и найденным расположением источников.

Работа выполнена при частичной финансовой поддержке РФФИ, проекты 10-07-00300, 11-07-12033 и 11-07-00577.

Список литературы

- [14] Francois Tadel, Sylvain Baillet, John C. Moshier, Dimitrios Pantazis, and Richard M. Leahy Brainstorm: A User-Friendly Application for MEG/EEG Analysis. Computational Intelligence and Neuroscience, vol. 2011, Article ID 879716, 13 pages, 2011. doi:10.1155/2011/879716
- [15] Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problems. Phys. Med. Biol., 32, 11–22.
- [16] Устинин М.Н., Поликарпов М.А., Панкратов А.Н., Рыкунов С.Д., Наурзаков С.П., Гребенкин А.П., Панченко В.Я. Сравнительный анализ экспериментальных данных магнитной энцефалографии. Математическая биология и биоинформатика. 2011. Т. 6. №1. С. 63 – 70
- [17] Устинин М.Н. Спектрально-аналитические методы обработки данных вычислительного и натурального эксперимента. Дисс. д.ф.-м.н. – Пушкино, 2004.

Опыт создания открытой web-ориентированной информационной системы о распространении редкого лишайника лобария легочная (*Lobaria pulmonaria* (L.) Hoffm.) на территории России

Иванова Н.В.⁽¹⁾, Шашков М.П.⁽²⁾

ФГБУ ВПО «Пушкинский естественно-научный институт»⁽¹⁾, ФГБУН «Институт физико-химических и биологических проблем почвоведения РАН»⁽²⁾
Natalya.dryomys@gmail.com⁽¹⁾, Max.carabus@gmail.com⁽²⁾

The author's development the open information system for organizing and visualizing data on the distribution of *Lobaria pulmonaria* describes in this synopsis.

Изучение редких видов – один из важнейших процессов в исследованиях биоразнообразия.

Объект данной работы – редкий лесной эпифитный лишайник лобария легочная (*Lobaria pulmonaria* (L.) Hoffm.). Этот вид внесен в Красные книги многих европейских стран, в том числе – в Красную книгу РФ [7].

По литературным данным *Lobaria pulmonaria* имеет обширный ареал в северном полушарии и встречается в бореальной и умеренной зонах, нежарких тропических лесах, а также в горных областях и районах с морским климатом. В течение XX века происходило катастрофическое сокращение численности лишайника в большинстве районов Центральной Европы, и к настоящему времени сохранились лишь небольшие по численности фрагментированные популяции лишайника [6].

Известно, что на территории России *L. pulmonaria* встречается в Европейской части страны, Западной и Восточной Сибири, Урале и Дальнем Востоке [7, 8]. Однако, все существующие на данный момент обзоры по изучению распространения редкого лишайника, имеют региональный характер. Сводки по ареалу лобарии для всей территории России не существует.

Цель данной работы – разработка единой открытой информационной web-ориентированной системы для систематизации и визуализации полевых, гербарных и литературных данных о распространении лобарии легочной на территории России. Для достижения поставленной цели в работе решались следующие **задачи**: 1) разработка структуры информационной системы; 2) выбор программного обеспечения для её реализации; 3) реализация ИС.

Структура информационной системы LLA (*Lobaria Lichen Area*) включает в себя: базу данных и web-интерфейс. Схема базы данных представлена на рис. 1. Все ПО, используемое для разработки ИС,

распространяется по различным вариантам открытых лицензий – GNU GPL [3], BSD [4] и другим. ИС LLA реализована на основе серверного ПО, работающего под управлением операционной системы Linux Ubuntu 12.4, web-сервера Apache 2.2.20, СУБД PostgreSQL 9.1.1, интерпретатора языка PHP 5.3.6.

Ядром ИС является база данных о распространении лобарии, разработанная на основе СУБД PostgreSQL [5]. База данных состоит из одной рабочей таблицы и справочников. Одной записи в рабочей таблице соответствует одно дерево, заселенное лобарией, что соответствует системе счетных единиц, применяемых при популяционных исследованиях лишайников [8]. Первоначально при разработке ИС в базе данных предполагались справочные таблицы, в которых содержатся сведения о: субъектах Российской Федерации; административных районах субъектов РФ; видах деревьев и кустарников, на которых может поселяться лишайник (потенциальных форофитов лобарии). По мере наполнения базы данными, возникла потребность в вынесении в отдельные справочные таблицы источники информации о распространении лишайника и список авторов его находок. Кроме того, потребовалась унификация авторских названий сообществ, которые также необходимо вынести в отдельный справочник.

Web-интерфейс ИС LLA состоит из двух компонентов: интерактивной карты распространения лобарии и формы, отражающей результаты пользовательских запросов о находках лишайника. Web-интерфейс размещен на сайте, разработанном для исследований лобарии легочной (www.lobaria.ru). Сайт построен на основе системы управления содержимым (CMS) Drupal 7.9 [2]. Интерфейс базы написан на скриптовых языках PHP и JavaScript с применением технологии асинхронного обмена данными AJAX.

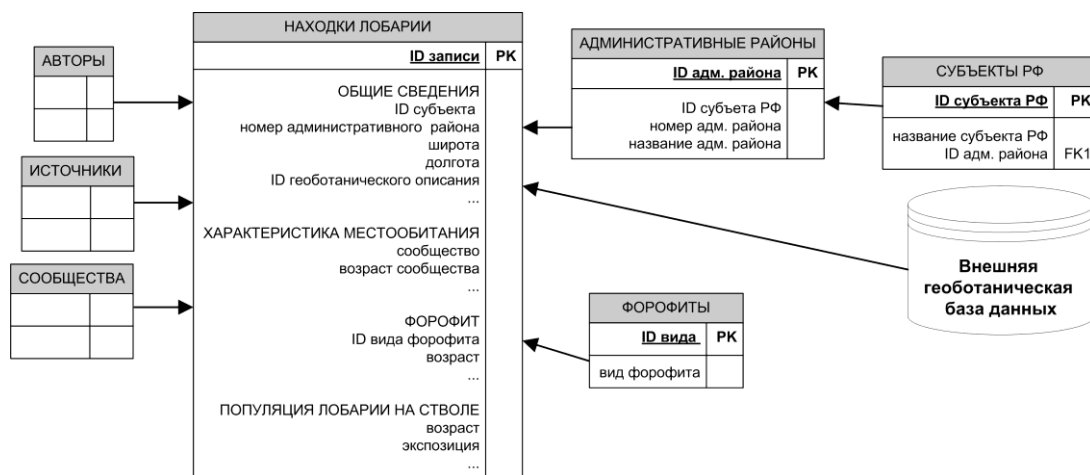


Рис. 9. Схема информационной системы LLA

Для визуализации точной географической привязки генерируется точечный векторный файл, для привязки уровня административных территорий из соответствующего слоя выбираются полигоны, на территории которых отмечены находки. В качестве средства визуализации пространственной информации используется картографический сервис Google Maps (через API) [1] и векторные слои, представленные в формате KML [9].

К настоящему времени в базе содержатся 204 записи о распространении лобарии. Информация о находках охватывает территорию 26 субъектов РФ. Данные собраны по материалам гербарных коллекций Ботанического института им. Комарова, на основе региональных Красных книг, а также публикаций по лишенологической тематике.

В ближайшее время предполагается разработка веб-формы для внесения данных о находках редкого лишайника пользователями сайта.

Список литературы

- [1] API Карт Google – Google Code. URL: <http://code.google.com/intl/ru/apis/maps> (accessed November 2011).
- [2] Drupal – Open Source Content Management System. URL: <http://drupal.org> (accessed November 2011).

- [3] GNU General Public License. URL: <http://www.gnu.org/copyleft/gpl.html> (accessed November 2011).
- [4] Open Source Initiative OSI – The BSD License. URL: <http://www.opensource.org/licenses/bsd-license.php> (accessed November 2011).
- [5] PostgreSQL – The world's most advanced open source database. URL: <http://www.postgresql.org> (accessed November 2011).
- [6] Scheidegger Ch, Frey B., Walser J.-Cl. Reintroduction and augmentation of populations of the endangered *Lobaria pulmonaria*: methods and concepts / Lobarion lichens as indicators of the primeval forests of the Eastern Carpathians (Darwin International Workshop, 25–30 May 1998, Kostrino, Ukraine). – p. 33–52.
- [7] Красная книга РСФСР (растения). М., 1988. – 591 с.
- [8] Михайлова И. Н. Анализ субпопуляционных структур эпифитных лишайников (на примере *Lobaria pulmonaria* (L.) Hoffm.). / Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Биология. №1. С. 124–134

Справочное руководство KML (Keyhole Markup Language) URL: <http://code.google.com/intl/ru/apis/kml/documentation/kmlreference.html> (accessed November 2011).

Структурные инварианты видового богатства растений дельты Аму-Дарьи

Трофимова Г.Ю.

Институт водных проблем РАН
t_g_yu@mail.ru

The relationship between the elements of the plant species richness structure in the Amu Darya delta is described by Eq.(1), and the relationship between the elements of the species richness structure of herbaceous plant, by Eq.(2). For the succession of ecosystems of the same type in the Amu Darya delta equations are structural invariants and show the self-similarity typical of natural systems.

1 Введение

Поиск структурных закономерностей в природных системах достаточно сложен, хотя само их существование не вызывает сомнений [3, 5–6]. Среди инвариантов видовой структуры следует указать на полученную степенную зависимость видового богатства от численности сообщества [2, 4], а также степенную зависимость, связывающую количество трофических связей с видовым богатством [1].

Вклад автора в копилку инвариантов видовой структуры в данной статье представлен структурными инвариантами видового богатства растений и видового богатства травянистых растений дельты Аму-Дарьи.

Выбор в качестве объекта исследования флоры дельты Аму-Дарьи обусловлен тем, что с начала 1960–х гг. экосистема дельты оказалась в условиях резко изменяющегося гидрологического режима территории, а в 1980–е гг. гидрологические условия в низовьях Аму-Дарьи ухудшились настолько, что Южное Приаралье было объявлено зоной экологического бедствия. В экосистеме дельты Аму-Дарьи, характеризующейся повышенной ранимостью, произошли необратимые нарушения в составе, структуре и её функционировании.

2 Материалы и методы

2.1 Речной сток – основная составляющая водного фактора в дельте Аму-Дарьи

Речной сток Аму-Дарьи, определявший обводненность дельты, уровень залегания грунтовых вод и в конечном итоге – состав и структуру экосистемы дельты, являлся основной составляющей водного фактора в дельте с 1944 по 1989 г. [7]. На основе анализа гидрологических данных по створам дельты Аму-Дарьи была

выделена и охарактеризована последовательность временных периодов. При этом учитывались как стабильность поступления водных масс в дельту и направленность тренда на протяжении всего периода, так и обобщённый анализ состояния почвенного и растительного покровов дельты по литературным источникам.

№ п/п	Период (гг.)	Средний сток (км ³ /год)
1	1944–1960	45,6
2	1961–1970	36,1
3	1971–1977	18,5
4	1978–1981	12,3
5	1982–1989	5,0

Каждый выделенный период был рассмотрен как зона нормальной жизнедеятельности (зона оптимума) для некоторого конечного набора видов растений из множества всех видов, зафиксированных в дельте Аму-Дарьи в 1947–1989 гг. исследователями растительного покрова. Поэтому наряду с последовательностью временных периодов будет использована также последовательность однотипных экосистем дельты, различающихся по количеству водных масс, поступавших в дельту через створ Саманбай.

2.2 Динамика флоры дельты Аму-Дарьи в условиях изменяющегося гидрологического режима территории

Флористический список дельты Аму-Дарьи, зафиксированный в базе данных о природной среде Южного Приаралья с 1947 по 1989 г., состоит из 265 видов [10]. Все виды растений дельты были классифицированы по жизненным формам в соответствии с упрощенной системой жизненных форм, которая наиболее часто используется при анализе растительного покрова низовьев Аму-Дарьи. В ней выделены следующие жизненные формы: деревья (дер.), кустарники (кт.), кустарнички (ктч.), полукустарники (пкт.), полукустарнички (пктч.), травы многолетние (мнл.), травы однолетние (одн.) и травы двулетние (двул.).

В качестве структурных составляющих флоры дельты Аму-Дарьи автором рассматриваются следующие [8]:

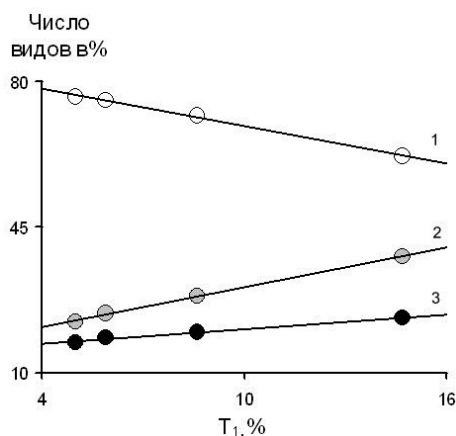


Рис. 1. Зависимости видового богатства структурных составляющих флоры дельты Аму-Дарьи от видового богатства полудревесных растений по данным 1944–1981 гг.: 1 – H ; 2 – T ; 3 – T_2 .

- древесно-кустарниковые виды растений (дер. + кт. + ктч. + пкт. + пктч);
- полудревесные виды растений (пкт. + пктч.);
- древесные виды растений (дер. + кт. + ктч.);
- травы (одн. + двул. + мнл.);
- одно- и двулетние виды трав (одн. + двул.);
- многолетние виды трав (мнл.).

Обозначим через T – число видов (видовое богатство) древесно-кустарниковых растений в каждом из выделенных временных периодов. Обозначим число видов в остальных структурных составляющих в соответствии с приведённым выше списком, а именно: T_1, T_2, H, H_1, H_2 . Тогда видовое богатство растений экосистемы дельты Аму-Дарьи каждого периода можно представить в виде суммы $T + H$, где $T = T_1 + T_2$, а $H = H_1 + H_2$. Динамика видового богатства структурных составляющих флоры дельты Аму-Дарьи представлена на рис. 1-2.

Данные, представленные на рис. 1, аппроксимированы с помощью линейных трендов, величина достоверности которых изменяется от 0,9936 до 0,9991 [8]. Согласно рис. 1, структурные составляющие видового богатства растений дельты Аму-Дарьи каждого периода связаны соотношением [9]:

$$\frac{T}{H} = \frac{T_1}{T_2}. \quad (1)$$

Согласно рис. 2, структурные составляющие видового богатства травянистых растений дельты Аму-Дарьи каждого периода связаны следующим соотношением:

$$\frac{\min(H_1, H_2)}{\max(H_1, H_2)} \approx \frac{2}{3} \quad (2)$$

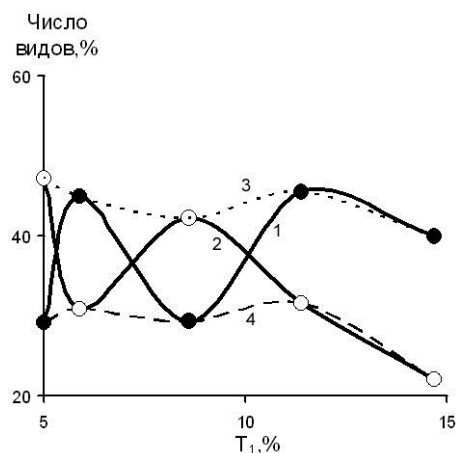


Рис. 2. Зависимости видового богатства структурных составляющих травянистой флоры дельты Аму-Дарьи от видового богатства полудревесных растений по данным 1944–1989 гг.: 1 – H_1 ; 2 – H_2 ; 3 – пунктирная линия – $\max(H_1, H_2)$ по периодам; 4 – штриховая линия – $\min(H_1, H_2)$ по периодам.

3 Заключение

Формулы (1), (2) не зависят от числа видов растений в экосистеме дельты Аму-Дарьи в выделенных временных периодах и тем более от численности каждого вида. Для последовательности однотипных экосистем дельты Аму-Дарьи, различающихся по величине речного стока, поступавшего в неё через створ Саманбай, формулы (1), (2) являются структурными инвариантами и демонстрируют свойство самоподобия, характерное для природных систем.

Список литературы

- [1] U. Brose, A. Ostling et al. *Nature*, 2004, **428**, 167-171.
- [2] Д.И. Иудин, Д.Б. Гелашвили и др. *ДАН*, 2003, **389**(2), 279-282.
- [3] Б. Мандельброт. Фрактальная геометрия природы. М.: Ин-т компьютер. исслед., 2002.
- [4] Р. Маргалеф. Облик биосферы. М.: Наука, 1992.
- [5] И. Пригожин, И. Стенгерс. Порядок из хаоса. М.: Прогресс, 1986.
- [6] В.Б. Сочава. Введение в учение о геосистемах. Новосибирск: Наука, 1978.
- [7] Г.Ю. Трофимова. *География и природ. ресурсы*, 2008, 1, 169-174.
- [8] Г.Ю. Трофимова. *География и природ. ресурсы*, 2009, 4, 82-87.
- [9] Г.Ю. Трофимова. *ДАН*, 2009, **426**(3), 427-429.
Г.Ю. Трофимова. Эколого-географическая база данных Южного Приаралья. М.: РАСХН, 2003.

Матричная модель популяции белоплечего орлана на Сахалине в 2004-2011 гг.

Романов М. С.⁽¹⁾, Мастеров В. Б.⁽²⁾

Институт математических проблем биологии РАН⁽¹⁾, Московский государственный университет⁽²⁾

michael_romanov@inbox.ru

Matrix model was developed for the Steller's sea eagle population in Sakhalin, Russia. The results show that the population is probably declining ($\lambda \leq 0.986$). Also, connection of demographic parameters with the age structure of the population was studied. We found that ratio "young/adults" is tightly related to λ and, therefore, can be used for diagnostics of the population state.

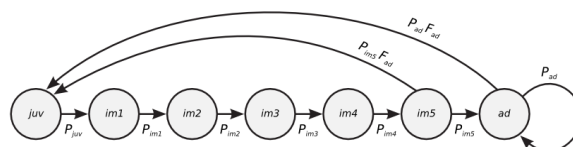


Рис. 1. Жизненный цикл белоплечего орлана, представленный в виде графа. Окружности — возрастные классы, стрелки — демографические переходы [3].

1 Введение

Данная работа является обобщением и развитием ранее выполненных исследований по моделированию популяционной динамики белоплечего орлана на Сахалине, начатых в 2004 г. [3]. К 2012 г. накопился значительный объем данных о продуктивности и возрастной структуре белоплечего орлана за 8 сезонов, отличающихся погодными, кормовыми и др. условиями, что позволило получить уточненный и более сглаженный «портрет» возрастной структуры популяции. Задачами исследования было 1) проверить выводы, полученные ранее, с учетом новых данных; 2) исследовать свойства проекционной матрицы при разных значениях выживаемости возрастных классов.

2 Материалы и методы

В основе модели лежат данные о численности и продуктивности белоплечих орланов на севере Сахалина в 2004-2011 гг.

Моделирование выполнено в среде R. Инструментом моделирования была линейная матричная модель [2], описывающая динамику популяции на основе уравнения $N' = AN$, где N , N' — вектор-столбцы численности возрастных классов в начале и в конце временного периода, A — проекционная матрица, задающая правила перехода между классами. Доминантное собственное число матрицы (λ) соответствует темпам роста популяции, правый собственный вектор — устойчивой возрастной структуре. Временной шаг модели равен 1 году; размножение происходит в конце шага; продуктивность рассчитывается на 1 особь; возраст наступления половой зрелости — 6 лет. Жизненный цикл орланов представлен на рис. 1.

3 Результаты

3.1 Возрастная структура популяции

Под возрастной структурой популяции мы понимаем соотношение численности трех основных хорошо различимых групп — взрослые особи, молодые (1-5-летнего возраста), а также слетки (подростки птенцы перед вылетом из гнезда).

Возрастная структура популяции отличается преобладанием взрослых особей: их доля в среднем составляет 0.72, в то время как доля молодых равна 0.15, доля слетков — 0.13 (рис. 2).

3.2 Выживаемость возрастных групп

Прямые оценки выживаемости возрастных классов отсутствовали, поэтому поведение матричной модели исследовали при разных значениях выживаемости взрослых особей, начиная с 0.95, которое соответствует выживаемости в популяциях близкородственных видов [3]. Выживаемость в течение 1-го года (P_{juv}) определяли по возрастной структуре, предположив, что выживаемость молодых особей после первого года жизни (P_{im1}) возрастает до максимума (т. е. до уровня взрослых).

3.3 Результаты моделирования

Была построена объединенная модель за весь период исследований, а также модели за каждый сезон, чтобы оценить разброс результатов. Результаты моделирования представлены в табл. 1. Почти во все годы (за исключением 2004 г.) $\lambda < 1$, что соответствует сокращающейся популяции. В объединенной модели $\lambda = 0.986$, что соответствует сокращению популяции вдвое за 50 лет.

Таблица 1. Входные параметры модели и результаты моделирования

Параметр	2004	2005	2006	2007	2008	2009	2010	2011	Все годы
N_{juv}	0.15	0.11	0.14	0.18	0.14	0.19	0.07	0.15	0.13
N_{im}	0.21	0.12	0.12	0.14	0.11	0.12	0.18	0.14	0.15
N_{ad}	0.64	0.77	0.74	0.68	0.75	0.69	0.75	0.71	0.72
P_{ad}	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
F_{ad}	0.23	0.14	0.18	0.26	0.18	0.27	0.10	0.21	0.19
S	0.33	0.16	0.16	0.21	0.15	0.17	0.24	0.20	0.21
P_{juv}	0.32	0.23	0.19	0.17	0.17	0.14	0.53	0.20	0.23
λ	1.014	0.978	0.980	0.986	0.977	0.982	0.991	0.985	0.986

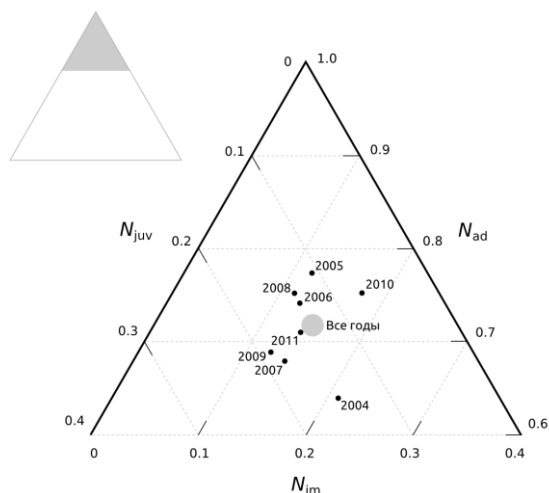


Рис. 2. «Портреты» возрастной структуры популяции белоплечего орлана в различные сезоны с 2004 по 2011 гг. в треугольной системе координат. N_{ad} — доля взрослых особей, N_{im} — доля молодых особей, N_{juv} — доля слетков.

Анализ пространства всех возможных значений P_{juv} и F_{ad} при $P_{ad} = 0.95$ показал, что распределение λ коррелирует с отношением молодые/взрослые особи, и при этом практически не зависит от доли слетков (рис. 3). Смоделировав поведение λ при разных значениях P_{ad} , мы вывели следующее уравнение: $\lambda^5 = P_{ad}^5(1 + S)$, где P_{ad} — выживаемость взрослых, S — отношение численности молодых особей к взрослым.

4 Заключение

То, что доля молодых особей в популяции жестко связана с λ , в то время как доля слетков, определяющая параметр продуктивности, не вошли в данное уравнение, оказалось неожиданностью, так как считается, что доля молодых особей — параметр ненадежный [1].

Полученное уравнение упрощает некоторые расчеты. Например, вычислим P_{ad} , при которой популяция была бы стабильной ($\lambda = 1$). Для данной популяции ($S = 0.21$) она равна 0.963. Так как

реальное значение, по нашему мнению, находится в интервале 0.94-0.95, это подтверждает вывод, что за период исследования популяция сокращалась.

Однако точное значение λ определить без знания выживаемости невозможно, поэтому в дальнейшем усилия следует направить на определение выживаемости возрастных групп.

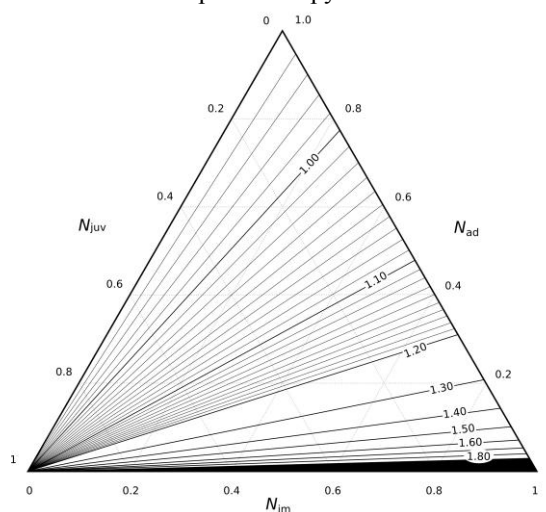


Рис. 3. Связь λ со структурой популяции при $P_{ad} = 0.95$.

Список литературы

- [1] Stalmaster, M. V. 1987. The Bald Eagle. Universe Books, New York, New York, USA. 227 pp.
 - [2] Логофет Д. О., Белова И. Н. 2007. Неотрицательные матрицы как инструмент моделирования динамики популяций: классические модели и современные обобщения. Фундаментальная и прикладная математика, 2007. 13(4). 145–164.
- Романов М. С., Мастеров В. Б. Матричная модель популяции белоплечего орлана *Haliaeetus pelagicus* на Сахалине. Математическая биология и биоинформатика, 2008, т. 3, No2, с. 36-49.

Анализ пространственной организации лесного ценоза: демографический аспект

Безрукова М.Г.⁽¹⁾, Грабарник П.Я.⁽¹⁾, Смирнов В.Э.⁽²⁾, Алейников А.А.⁽³⁾

Институт физико-химических и биологических проблем почвоведения РАН⁽¹⁾, Институт математических проблем биологии РАН⁽²⁾, Центр по проблемам экологии и продуктивности лесов РАН⁽³⁾
grya@rambler.ru

Analysis of spatial pattern of a forest stand has been conducted. We study patterns in location of young trees with respect to adult trees. A new testing procedure for this is suggested.

1 Введение

Анализ пространственной структуры экологических систем и процессов, влияющих на ее изменения, становится в последнее время одной из важнейших задач экологических и биогеоценотических исследований. Интерес к изучению пространственных отношений между растениями в сообществе связан с пониманием того, что пространственная структура является существенным детерминантом развития популяции растений, определяя судьбу растения как на ранних стадиях, так и в течение всей жизни.

К настоящему времени накоплен большой экспериментальный материал и сформулированы биологические гипотезы, объясняющие механизмы, влияющие на скорость роста и выживаемость растений в сообществе. Предложены модели, описывающие популяционные механизмы, регулирующие отношения деревьев в сообществе. Эти теоретические концепции позволяют приблизиться к пониманию основ устойчивого развития лесных экосистем. Например, популярная гЭП-теория построена на предположении о следующих этапах динамики древостоя: 1) образовании «окна» в пологе леса за счет усыхания или вывола отмершего дерева, 2) появления в «окне» многочисленного подроста 3) развитие подроста, заканчивающееся тем, что одно или несколько молодых деревьев вырастают в полог, закрывая «окно», 4) отмирание оставшегося подроста (Yamamoto, 1992). В случае смешанного широколиственно-темнохвойно-кедрового леса на основе результатов дендрохронологического и пространственного анализа было показано (Ухваткина, Омелько, 2011), что расположение подроста не определяется размещением деревьев верхнего полога, а связано с размещением существовавших в прошлое время «окон» в древостое. Таким образом, развитие хвойного подроста согласно выводам Ухваткиной и Омелько

(2011) происходит в несколько этапов: вначале подрост накапливается под пологом древостоя, а затем последовательно выходит во второй и первый полог, при этом периоды быстрого роста чередуются с длительным угнетением. Эта концепция подтверждается анализом карт расположения деревьев на лесной территории методами статистического анализа, основанного на применении мер второго порядка.

В данной работе анализируется пространственная организация ельнично-пихтового древостоя. Биологический интерес представляет задача анализа пространственной структуры подроста, которая позволяет высказать гипотезы о процессе возобновления лесов данного типа. С математической точки зрения интерес представляет задача, связанная с различием эффектов первого и второго порядков. Эффектами первого порядка называют нарушения однородности пространственных случайных процессов, описывающих размещение подроста по лесной территории, в то время как эффекты второго порядка связаны с корреляционной структурой случайной системы.

2 Материалы и методы

На Рис.1 представлены планы расположения деревьев верхнего полога и подроста. Материал собран в Печоро-Илычском заповеднике (средняя тайга) в елово-пихтарнике крупнопоротниковом. Для анализа точечных структур, состоящих из точек нескольких типов, обычно используется частная (или кросс-) K -функция Рипли и парная корреляционная функция $g(r)$, которые определяются с помощью среднего числа пар различных точек, расположенных друг от друга на расстоянии меньше, чем r [3,1].

Вместо K -функции Рипли удобнее работать с ее модифицированной формой - L -функцией: , которая имеет как графические так и статистические преимущества, в частности, ее дисперсия почти постоянна в широком диапазоне шкал. Стандартной задачей, в которой используется кросс L -статистика является проверка "независимости" типов точек при фиксированном расположении объектов. В нашем случае мы интересуемся значимостью отличия

наблюдаемой L -функции от типичных L -функций, которые соответствуют паттернам с фиксированным расположением деревьев верхнего полога и случайным расположением подроста.

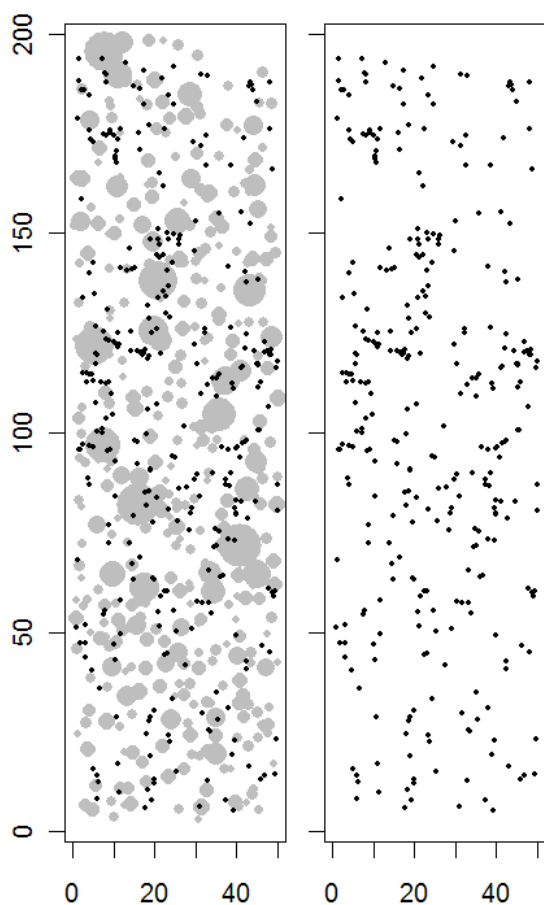


Рис.1. План расположения деревьев верхнего полога и подроста (слева) и отдельно подроста (справа).

3 Результаты

В результате проведенного анализа (см. Рис.2) можно говорить о том, что возможности возобновления у растений расположенных на расстояниях до 3-4 метров в данном типе леса ограничены, так как подрост не получает достаточного ресурса для роста.

Доля подроста расположенного на периферии крон деревьев верхнего полога относительно высокая главным образом из-за относительно низкой плотности подроста на открытых участках. Форма кривой L -функции на расстояниях от 8 метров и больше свидетельствует о наличии эффектов

первого порядка, которые выражаются в повышенной доле участков с низкой плотностью деревьев. Это подтверждается также статистическими расчетами, основанными на мерах распределения расстояний до ближайшего объекта (не показано здесь). С биологической точки зрения наблюдаемый эффект объясняется низкой приживаемостью древесных видов растений из-за высокой конкуренции со стороны травянистых видов.

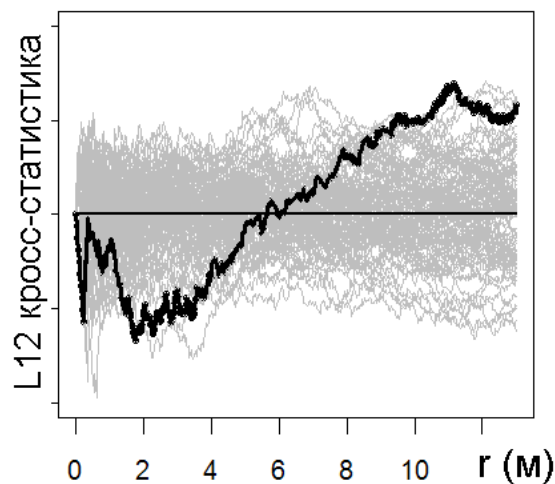


Рис. 2. Оценка кросс L -функции для деревьев верхнего полога и подроста и области принятия нулевой гипотезы о «случайности» размещения подроста.

Список литературы

- [1] Грабарник П.Я. Анализ горизонтальной структуры древостоя: модельный подход. *Лесоведение*, 2010, **2**, 77-85.
- [2] Ухваткина О.Н, Омелько А.М. Структура подроста и естественное возобновление в смешанном хвойно-широколиственном лесу Южного Сихотэ-Алиня. *Journal of Siberian Federal University. Biology*, 2011, **3**, 266-280.
- [3] Illian J., Penttinen A., Stoyan H. and Stoyan D. *Statistical Analysis and Modelling of Spatial Point Patterns*. J. Wiley & Sons, Chichester. 2008.
- [4] Yamamoto S.-I. The gap theory in forest dynamics. *Journal of Plant Research*, 1992, **105**, 375-383.

Фрактальный анализ горизонтальной структуры напочвенного покрова

Голубев С.Н.

ООО «Лаборатория фрактального анализа, экологии и программирования»

Schwejk-rpnt@rambler.ru

Developed a new method the analysis of the horizontal structure of the ground cover by photographs. The connection of the horizontal structure of the characteristics of plant communities.

1 Введение

Анализ структуры напочвенного покрова необходим при изучении диссипативных свойств растительности. Сложность – важнейший критерий при таком анализе, но метод, позволяющий выразить ее численно, отсутствует. Нет данных о связи сложности структуры с условиями произрастания.

2 Цель работы

Цель работы - разработать метод оценки сложности горизонтальной структуры напочвенного покрова. Проследить связь сложности структуры напочвенного покрова с условиями произрастания.

3 Методика

Для оценки сложности структуры применен новый критерий, основанный на фрактальном анализе фотоизображений напочвенного покрова.

Заложены пробные площади в г. Санкт-Петербурге (Лесопарк «Пискаревка» - 13 пробных площадей) и Мурманской области (Мончетундры и Хибины – 21 пробная площадь). Фотографии сделаны на учетных площадках внутри пробных площадей (всего 383 учетных площадки). Для каждой учетной площадки определено повидовое и общее проективное покрытие травяно-кустарничкового и мохово-лишайникового яруса. Размер учетной площадки 0,5х0,5 м и 1х1 м.

К настоящему дню обработаны полевые материалы по малым (0,5х0,5 м) учетным площадкам (80 площадок).

При анализе данных описанные виды травяно-кустарничкового яруса делились на 4 эколого-ценотические группы:

Лесные виды, характерные для ненарушенного леса черничной серии типов, такие как *Vaccinium myrtillus*, *Oxalis acetosella*, *Trientalis europaea* и др.;

Луговые виды (*Trifolium pratense*, *Lathyrus pratensis* и др.);

Рудеральные виды (*Plantago major*, *Taraxacum officinale* и др.);

Опушечные, неморальные, прибрежные и прочие виды, такие как *Geum urbanum*.

Для каждой из эколого-ценотических групп рассчитывалась степень доминирования - индекс Симпсона [5].

На первом этапе обработки фотоизображений, снимки обрезаны по контуру учетной площадки и переведены в формат bmp (256-цветовая палитра) с разрешением 22,97 x 22,97 см (650 x 650 пикселей). Затем полученное изображение сохранялось в виде негатива формата bmp (черно-белый растр). При этом травы и кустарнички отображаются в виде закрашенных контуров (злаки отображаются в виде узких полос, толщиной 1-3 пикселя). Мхи отображаются в виде группы крупных точек размером 4-15 пикселей. Неоднородности подстилки (сборки, разрывы) отображаются в виде отдельных точек, размером 1-4 пикселя. Однородные участки подстилки отображаются в виде белого фона.

После этого полученное изображение с помощью специального программного обеспечения покрывалось сетью клеток определенного масштаба. Подсчитывалось количество клеток, в которых граничили между собой черный и белый пиксель

Изображение покрывали 10 раз клетками различного масштаба.

Зависимость между площадью одной клетки и количеством клеток содержащих в себе черный и белый пиксель аппроксимировали степенным законом (коэффициент аппроксимации выше 0,98) $y=ax^{-D}$, где D - размерность Хаусдорфа-Безиковича (фрактальная размерность).

Для природных объектов характерно множество размерностей (мультифракталы). Показатель размерности при их изучении зависит не только от сложности анализируемой структуры, но и от параметров клеточного метода [3]. Обычно для анализа мультифракталов применяют кривую спектра фрактальных размерностей [1], [4], [6]. Чтобы выразить этот спектр численно, применена авторская формула отношения области охваченной мультифрактальным спектром к площади, ограниченной топологическими размерностями, между которыми заключен фрактальный спектр:

$$Z = \frac{\int_b^a f(x)dx}{(Dt*b)-(Dt*a)}$$

где Z - показатель сложности структуры (назван интегральной размерностью покрытия), $f(x)$ - функция, наилучшим способом аппроксимирующая фрактальный спектр, Dt – топологическая

размерность; a – минимальная площадь клеток покрытия; b – максимальная площадь клеток покрытия.

4 Результаты

Наиболее простая структура (значение интегральной размерности покрытия – 0,65-0,7) наблюдается при высокой сомкнутости древостоя (80%) низкой мощности подстилки, малом проективном покрытии. Видовое разнообразие на таких площадках низкое (в среднем 3 вида травяно-кустарничкового яруса), представлены преимущественно виды лесных местообитаний (*Oxalis acetosella*, *Vaccinium vitis-idaea*, *Vaccinium myrtillus*, *Majanthemum bifolium*).

При снижении сомкнутости древостоя увеличивается видовое разнообразие и проективное покрытие видов травяно-кустарничкового яруса. Снижается степень доминирования лесных видов, за счет усиления роли видов опушенных, неморальных и прочих местообитаний. Появляются рудеральные виды (*Taraxacum officinale*). Увеличивается мощность подстилки (более 1 см). Значение интегральной размерности покрытия для структуры напочвенного покрова составляет 0,7-0,8.

В условиях минимальной сомкнутости древостоя проективное покрытие и видовое разнообразие достигают самых больших значений. Мощность подстилки в этих условиях максимальна. В травяно-кустарничковом ярусе из лесных видов сохраняется только седмичник европейский. Появляется *Poa pratense* (типичный луговой вид). Доминируют виды переходных местообитаний. Структура напочвенного покрова имеет наибольшую сложность. Интегральная размерность покрытия для таких площадок выше 0,9.

Интегральная размерность покрытия зависит от трех факторов: проективного покрытия ($r^2=0,64$), видового разнообразия ($r^2=0,83$) травяно-кустарничкового яруса и мощности подстилки на учетной площадке ($r^2=0,94$).

Связь сложности структуры напочвенного покрова и проективного покрытия наилучшим образом описывается степенным законом, поскольку общее проективное покрытие имеет верхний предел (100 %). Связь интегральной размерности покрытия и видового разнообразия травяно-кустарничкового яруса линейна, поскольку максимальная граница видового разнообразия отсутствует.

Сложность структуры напочвенного покрова связана со степенью доминирования основной эколого-ценотической группы. При смене доминирующей эколого-ценотической группы от лесных видов к сорным, сложность структуры повышается, достигает максимума при равном доминировании разных эколого-ценотических групп и в дальнейшем снижается при усилении доминирования новой эколого-ценотической группы. Связано это с дугообразным изменением видового разнообразия при смене доминирующей эколого-ценотической группы [2].

5 Заключение

Разработанный метод подходит для оценки сложности горизонтальной структуры напочвенного покрова. С увеличением освещенности и смене доминирующей эколого-ценотической группы горизонтальная структура напочвенного покрова усложняется.

Список литературы

- [1] Божокин С.В., Паршин Д.А. Фракталы и мультифракталы. – Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001, 128 с.;
- [2] Голубев С.Н., Изменение видового разнообразия травяно-кустарничкового яруса при смене доминирующей эколого-ценотической группы в л/п «Пискаревка» // сборник материалов международной студенческой конференции: в 3 ч. / Ч.3. – Йошкар-Ола, 2009, с.43.;
- [3] Иванов А.В., Короновский А.А., Минюхин И.М., Яшков И.А. Определение фрактальной размерности овражно-балочной сети города Саратова // Изв. вузов «ПНД», т. 14, № 2, 2006. с. 64-74;
- [4] Мандельброт Б. Фрактальная геометрия природы. — М.: «Институт компьютерных исследований», 2002. 656 с.;
- [5] Методы изучения лесных сообществ – СПб.: НИИХимии СПбГУ, 2002. 240 с.;
- [6] Шурганова Г. В., Иудин Д. И., Гелашвили Д. Б., Якимов В. Н. Мультифрактальный анализ видового разнообразия зоопланктоценозов Чебоксарского водохранилища сб. докладов Всероссийской конференции "Актуальные проблемы водохранилищ", 2002, Борок, ИЭВВ РАН.

Малый биологический круговорот как динамический итог сукцессионных процессов в экосистеме: математическая модель

Мироненко Л.М.

Институт физико-химических и биологических проблем почвоведения РАН, Пущино,
Московская область, 142290, Россия
lmir@yandex.ru

The equations of model of a small biological turnover are obtained from equations of ecological succession model. The results of applying of the first model to analysis of quantitative nitrogen datas for different ecosystems are considered.

1 Введение

Настоящий доклад посвящен дальнейшему исследованию возможностей модели экологической сукцессии (ЭС), представленной в работе [1]. Модель ЭС была успешно использована при анализе процессов в лабораторной микрэкосистеме, где она работала и как модель ЭС и как базовый элемент модели разложения растительных остатков одновременно [2]. В настоящем докладе показано, что из указанной модели ЭС следует модель малого биологического круговорота (МБК). Обсуждаются результаты анализа соответствующих литературных данных [3] при помощи данной модели МБК.

2 Модель малого биологического круговорота

Как известно, основу МБК составляют сопряженные процессы созидания живого вещества из неживого и возвращение живого вещества в состояние неживого (рис.1d). Складывая уравнения ЭС (рис. 1a) для биомассы (Y) и субстрата (X) отдельно, получим уравнения модели замкнутого МБК (рис. 1c), соответственно приведенному выше определению.

Стабильность МБК предполагает постоянство величин биомассы и субстрата в экосистеме. Согласно модели ЭС (рис. 1b, графики) это постоянство носит динамический характер. Составляющие живое вещество сообщества организмов со временем отмирают и на их место приходят новые, видовой состав биомассы

непрерывно меняется. Начиная с некоторого момента это происходит уже при практически постоянных величинах биомассы Y и субстрата X. Т.е., МБК существует в некотором динамическом равновесии.

Модель МБК была использована при анализе соответствующих литературных данных [3]. Были рассмотрены количественные данные по круговороту азота для нескольких экосистем – от тундры до пустыни. Полагалось, что экосистемы находятся в стабильном состоянии. Принимались во внимание стабильные запасы азота в растительных биомассе и мортмассе, содержания азота в потоках потребления и отмирания и в почве. Удалось установить следующее.

Модель МБК хорошо работает для микрэкосистем, описанных в [2], в то время как в природных экосистемах при анализе процессов на самом общем уровне в схему процесса МБК (рис. 1d) необходимо ввести дополнительный блок мортмассы. Это связано с медленным разложением и, соответственно, накоплением мортмассы (сухих деревьев, ветоши и пр.) до известного предела, в результате чего величина мортмассы заметно превышает величину биомассы, и это превышение возрастает от юга к северу. Необходимым субстратом здесь является почва. Вместе с тем константы скоростей потоков потребления и отмирания находятся между собой примерно в том же соотношении, что и в работе [2], что свидетельствует о наличии определенного единства в процессах лабораторных и природных экосистем. Исследование тормозится большими методическими ошибками при определении азота в компонентах экосистем, на что указывают сами исследователи [4]; это исключает многие литературные данные из рассмотрения; существует дефицит данных. Однако перспективы получения новой информации из количественных данных наблюдений при помощи рассмотренных

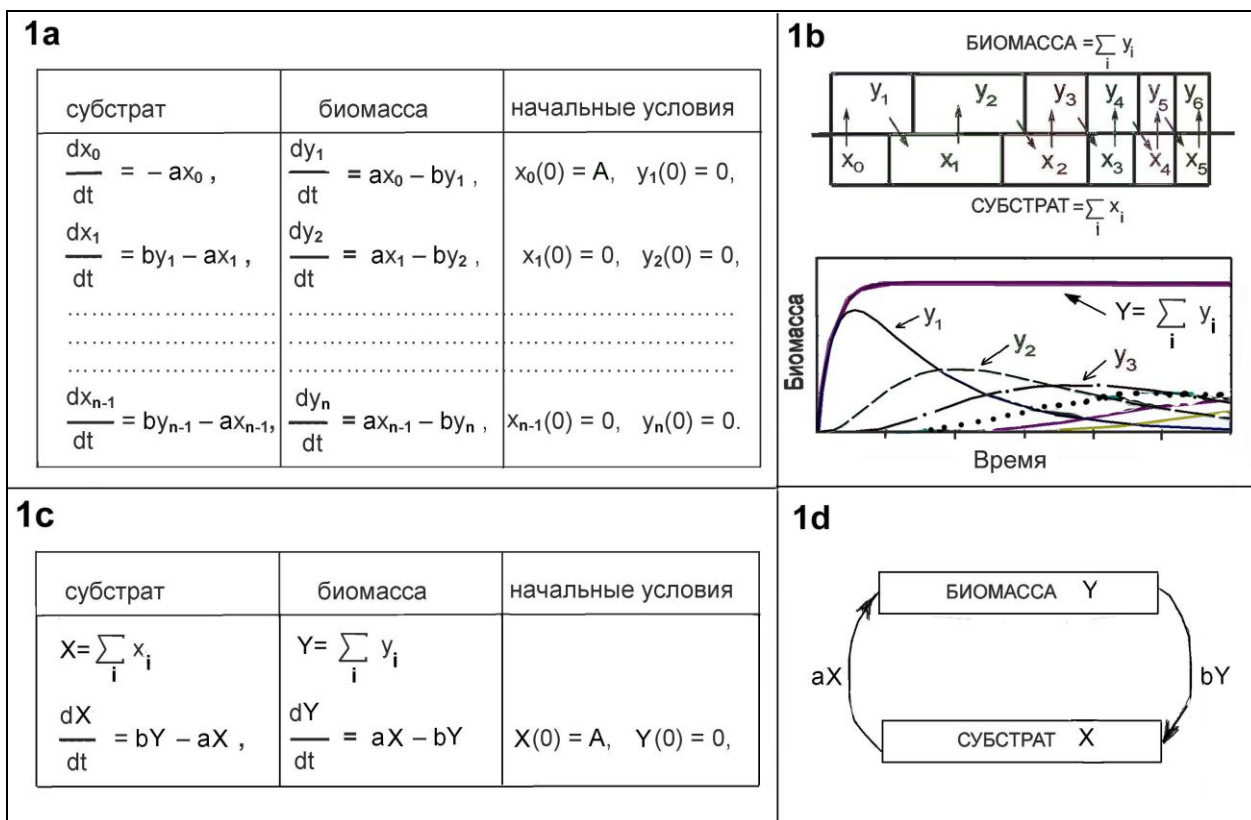


Рис. 1. Модели экологической сукцессии и сопутствующего ей биологического круговорота: 1a – уравнения модели ЭС, A – константа [2]; 1b – схема сукцессионного процесса и расчетные кривые биомассы [2]; 1c – уравнения модели МБК; 1d – схема процесса МБК согласно модели.

моделей существуют. Работа продолжается.

3 Заключение

Обсуждаемые в докладе модели ЭС и сопутствующего ей МБК чрезвычайно просты. И вместе с тем использование этих моделей в качестве инструмента анализа количественных данных об экосистемах и других экспериментальных данных позволило получить новую информацию. Исследователь экосистемных процессов, находясь внутри изучаемого объекта много больше его самого, поневоле фиксирует и моделирует частности. Возможно, стоит попытаться начать с наиболее общих формулировок и представлений.

«Наука в этом смысле чрезвычайно хлопотлива и полна суеты. В хаосе и неразберихе эмпирически спутанных, текучих вещей надо уловить идеально-числовую, математическую закономерность, которая хотя и управляет этим хаосом, но сама-то не есть хаос, а идеальный, логический строй и порядок» [5].

Список литературы

- [1] Л.М. Мироненко. Математическая модель экологической сукцессии. *Математика. Компьютер. Образование. Сб. научных трудов.* 1999. Вып. 6. Ч. 2. С. 516–525.
- [2] Л.М. Мироненко. Два подхода к математическому моделированию процессов в лабораторных микрэкосистемах: взгляд извне и изнутри. *Математическая биология и биоинформатика.* 2012. Т 7. № 1. С.182–196. URL: <http://www.matbio.org/article.php?lang=rus&id=105>.
- [3] Н.И. Базилевич, О.С. Гребенщиков, А.А. Тишков. Географические закономерности структуры и функционирования экосистем. М.: Наука, 1986. 297 с.
- [4] Н.И. Базилевич, А.А. Титлянова, С.Я. Кудряшова, Н.П. Косых.. *Известия АН, сер. географическая.* 2003, 4, 8-16.
- [5] А.Ф. Лосев. Диалектика мифа. М.: «Правда». 1990.

Динамика биомассы ветвей высших порядков дерева. Модельный анализ и натурные данные

Галицкий В.В.

Институт физико-химических и биологических проблем почвоведения РАН
galvv@rambler.ru

The sectional model of biomass dynamics of tree branches' system shows bell-shaped dynamics of a biomass of branches and boundedness of number of branches' orders. The important element of model is the inter-verticil green biomass. Known peculiarities of growth of a spruce – an initial growth inhibition and presence of inter-verticil branches are considered. Good approximation of natural data is received.

1 Модель системы регулярных ветвей дерева

В [4,6] была представлена секционная структура дерева и секционная модель динамики его биомассы. Модель, основанная на представлении о виртуальном дереве и секции как разности двух виртуальных деревьев, позволяет по двум зависимостям от возраста – зеленой биомассы и высоты реального дерева – получить динамику распределения биомассы по высоте дерева. На примере ели секцию можно отождествить с междоузлем (часть ствола) и соответствующей мутовкой из пяти скелетных 1-ветвей. Ниже эта модель расширяется для описания динамики биомассы и времен жизни j -ветвей высших порядков ели обыкновенной, которые состоят из образуемых виртуальными ветвями секций с двумя регулярными ветвями каждая.

Биомассу ${}^j b_i$ k -секции ветви j -го порядка мутовки i -секции дерева можно записать как

$${}^j b_i(T-T_i - (j+k)\Delta T) = {}^j br_i(T-T_i - (j+k)\Delta T) - {}^j br_i(T-T_i - (j+k+1)\Delta T), \quad (1)$$

где в правой части представлена разность биомасс ${}^j br_i$ смежных виртуальных k - и $(k+1)$ -ветвей j -ветви i -секции дерева. Тогда динамику биомассы ${}^{j+1} br_i$ ветви порядка $j+1$, являющейся частью мутовки k -секции j -ветви, можно представить в виде рекуррентного выражения

$${}^{j+1} br_i(T-T_i - (j+k+1)\Delta T) = ({}^j br_i(T-T_i - (j+k)\Delta T) - {}^j b_i(T-T_i - (j+k)\Delta T)) / 2. \quad (2)$$

Динамика зеленой биомассы реального дерева представляет собой монотонно растущую функцию с насыщением (σ -функция) [6]. Поэтому из (1) следует, что динамика биомассы секции дерева

описывается колоколообразной положительной функцией, как и динамика скелетных ветвей (1-го порядка) этой секции, которые, однако, при некоторой межмутовочной биомассе ${}^j b_i$ в (2) могут отмирать (см. ниже). Ветви 2-го и следующих порядков, согласно (2), в принципе имеют, в отличие от 1-ветвей, конечное время жизни и соответственно возможный максимальный порядок ветвей невелик.

2 Приложение к ели европейской

Описанная модель системы регулярных ветвей дерева была применена к натурным данным для ели европейской *Picea abies* (L.) Karst. из [10]: 1) максимальный порядок ветвей – 4; собственные длительности $t_{D,j}$ ([год]) жизни от скелетных (1-го порядка) ветвей до 4-ветвей – (30,13,8,3); суммарная масса хвои 1-ветви максимальна в 10 лет. Использовались зависимости: изометрическая [8] для высоты дерева $H_{pl}(T)$ и аллометрическая [6] для биомассы B_i виртуального i -дерева, зависящей от высоты H_i его появления

$$H_{pl}(T) = H_m \operatorname{th}(T/A_1), \quad B_i(T) = B_{m,i}(H(T)/H_m)^\mu, \\ B_{m,i} = B_m(1-H_i/H_m)^r. \quad (3)$$

Поиск (с применением алгоритма SA [1]) значений в пространстве параметров (A_1, r, μ) описанной выше модели показал, что эта модель регулярной системы ветвей дерева не только не позволяет получить в соответствующих диапазонах параметров ни одного значения $t_{D,j}$, близкого к требуемым, но не показывает существования ветвей порядка выше первого. Модель была дополнена двумя субмоделями, которые учитывают известные ростовые и морфологические особенности ели обыкновенной.

2.1 Начальное торможение роста

Для ели характерен весьма медленный рост до 10-15 лет, обычно начинающийся под березовым пологом, т.е. в довольно густом травяном покрове. Эта особенность учитывается добавлением в выражении (3) для $H_{pl}(T)$ множителя $k_w(T)$

$$H(T) = H_{pl}(T) k_w(T), \quad k_w(T) = \operatorname{th}(T/\tau_w) \quad (4)$$

где w и τ_w – дополнительные параметры модели.

2.2 Межмутовочные ветви

Эти ветви появляются на междоузлиях ели при внезапном увеличении освещенности [2,9], приводящего к увеличению B_m в (3). Регулярные ветви возникают из пазушных почек, закладываемых в сезон, предшествующий появлению ветвей. Свойства ветвей (т.е. величина B_m в (3)) жестко заданы условиями этого сезона и соответствуют условиям на момент закладки почек. При улучшении условий (т.е. увеличении B_m) для сохранения “гармонии” с транспортом ассимилятов регулярные ветви должны бы изменить свои свойства, чего они сделать не могут. В отличие от регулярных межмутовочные ветви образуются из придаточных почек. Модель (1)–(2) при учете придаточных почек подсказывает возможный способ решения этой проблемы, объясняющий также и функцию “межмутовочных” ветвей. Как можно видеть из выражения (2), межмутовочная биомасса есть член $j b_i$. Субмодель реализована как 1-ветвь, аналогичная регулярной ветви с биомассой $B'_m = \beta B_m$ и появляющаяся через t_B после i -секции ствола (β и t_B – параметры). В работе [9] отмечено, что отмирание регулярных ветвей и рост межмутовочных происходят одновременно.

2.3 Результаты

На рис.1 приведена динамика биомассы j -ветвей ели для полной модели ($i=10$). Полученные значения времен отмирания j -ветвей весьма близки к натурным данным. Включение только одной из двух субмоделей не дает удовлетворительных результатов. На рис. 2 представлена динамика распределения биомассы по высоте дерева при найденных параметрах, штриховые линии – без добавления субмоделей ($T=20, 30, 50, 100, 200$).

2.4 Избыток ассимилятов инициирует придаточные почки?

Результаты моделирования вместе с [5] позволяют предположить, что пробуждение адвентивных почек, дающих начало межмутовочным ветвям, инициируется повышением давления ассимилятов во флоэмной системе дерева при увеличении интенсивности освещенности и невозможности адекватного увеличения пропускной способности регулярной флоэмы. О подобной ситуации свидетельствуют также данные эксперимента по декапитации ветвей *Wollemia nobilis* (Araucariaceae) – пазушные почки и затем побеги появлялись среди группы клеток коры, обогащенных крахмалом [3].

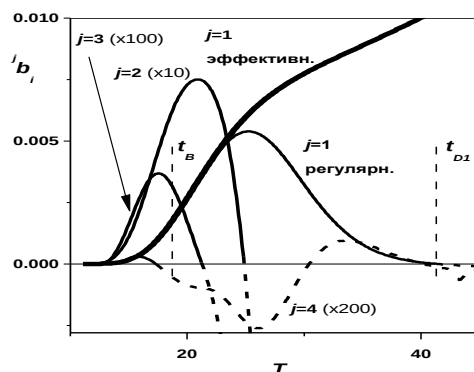


Рис.1

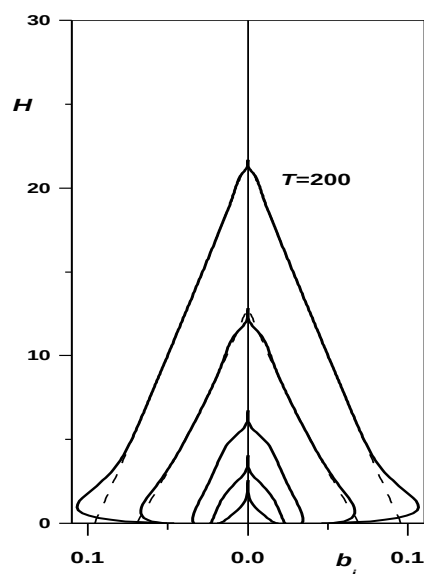


Рис.2

Список литературы

- [1] A. Corana et al. ACM Trans. Math. Soft, 1987, **13**(3), 262–280.
- [2] P.J. Kramer, T.T. Kozlowski. Physiology of woody plants. NY.: Acad. Press, 1979.
- [3] P.B. Tomlinson, B.A. Haggett. Ann.Bot., 2011, **107**, 909–916.
- [4] В.В. Галицкий. ДАН. 2006, **407**(4). 564–566.
- [5] В.В. Галицкий. Мат. биол. и биоинформатика, 2009, **4**(1), 1–20.
- [6] В.В. Галицкий. ЖОБ, 2010, **71**(1), 19–29.
- [7] Н.И. Казимиров. Ельники Карелии, Л.: Наука, 1971.
- [8] И.А. Полетаев. Пробл. киберн. (16), М.: Наука, 1966, 175–177.
- [9] П.П. Трескин. В: Структура и продуктивность еловых лесов южной тайги. Л.: Наука, 1973, 222–240.
- [10] Ю.Л. Цельникер. Лесоведение, 1994, (4), 35–44.

Модель представления и методы обработки данных в фотоландшафтной базе данных, используемой для изучения динамики лесотундровых сообществ Полярного Урала

Шалаумова Ю.В.

Уральский государственный лесотехнический университет
yulyash@gmail.com

The model of representation and methods of search and ensure consistency in the sets of descriptions of landscape photographs are offered. They are the basis of the information system that provides the solution of problems related to assessing the dynamics of high mountain plant communities.

Введение

В последнее время всё больший интерес представляет изучение динамики высокогорных растительных сообществ, поскольку они наиболее чутко реагируют на изменения факторов внешней среды и могут рассматриваться в качестве индикаторов таких изменений. Одним из методов, применяемых для оценки динамики высокогорной растительности, является сопоставление повторных ландшафтных фотоснимков [3]. При этом если на одиночном ландшафтном фотоснимке представлен срез реальности, т.е. зафиксировано состояние растительных сообществ на момент времени, то при сопоставлении разновременных фотоизображений визуально фиксируемые изменения выступают таким же объективным свидетельством, которое непосредственно позволяет подтвердить или опровергнуть гипотезу (теорию) о происходивших процессах, т.е. динамике растительных сообществ, как и измеряемые числовые данные.

Актуальной проблемой является использование накопленных, но разобщенных и малодоступных для научного сообщества архивов ландшафтных фотоснимков, решение которой возможно осуществить на базе информационной системы. При этом основные сложности возникают при выборе модели хранения описаний визуальных объектов, а также методов поддержания непротиворечивости и поиска в наборах описаний фотоснимков.

Цель работы – организация удобного хранения и быстрого доступа к коллекции ландшафтных фотоснимков и их описаний в фотоландшафтной базе данных для решения прикладных задач, связанных с оценкой пространственно-временной динамики высокогорной растительности.

Объекты и методы

Объектом исследований является древесная растительность и крупные кустарники, произрастающие на верхней границе своего распространения в высокогорьях Полярного Урала, их состояние и динамика, определяемые по ландшафтными фотоснимкам.

Для отображения описаний фотоснимков в фотоландшафтной базе данных разработана формальная модель представления знаний. Для этого в качестве базовой была выбрана модель представления и обработки наборов описаний слабоструктурированных объектов [1, 2]. Основные модификации исходной модели заключаются, во-первых, в учете специфики визуальных объектов, а, во-вторых, в повышении эффективности процедур пополнения и поиска за счет разработки моделей и методов, совмещающих процесс обработки порции данных, поступающей от пользователя, на двух уровнях: описания отдельного объекта и набора описаний объектов.

Результаты и обсуждение

В формальной модели под описанием ландшафтного фотоснимка понимается совокупность, в общем случае структурированная, значений данных различных типов. Алгоритмы работы формальной модели позволяют использовать поступающую информацию для проверки имеющихся в системе данных на непротиворечивость и, в соответствии с намерениями пользователя, для поиска и/или пополнения.

Модель позволяет обеспечить унифицированный механизм учета произвольности выбора значений переменных из областей их допустимых значений при формировании описания экземпляра объекта или поискового образа. Для этого в состав модели включены формальные объекты, называемые обусловленностями, являющиеся конъюнкциями произвольного количества формул типа импликации и эквиваленции, которые связывают между собой утверждения о принадлежности значений переменных некоторым подмножествам множеств допустимых значений.

В модели непротиворечивость данных поддерживается посредством процедуры сужения, обеспечивающей динамическое формирование областей допустимых значений одних переменных при означивании других, связанных с ними обусловленностями. Если представить ход процесса итерационным, то при появлении в переменной значений, не попадающих в ранее сформированную область допустимых значений, возникает ситуация несоответствия, которую можно решить в ходе диалога с пользователем. Если противоречия не были обнаружены, то область допустимых значений второй переменной будет сужена, благодаря задействованным обусловленностям.

Предлагается совместить в одной процедуре, с одной стороны, операции поиска-идентификации на наборах описаний и, с другой стороны, операции сужения областей допустимых значений еще не означенных атрибутов для отдельных кортежей с помощью обусловленностей по мере появления значений других атрибутов в данном кортеже. При поступлении текущего описания осуществляется, с одной стороны, проверка введенных данных на непротиворечивость, и, с другой стороны, расширение поискового образа, в рамках единой процедуры. Расширение обеспечивается за счет использования процедуры сужения как для поискового образа, так и для описаний из набора, что позволяет включать в операции процедуры поиска не только множества имен атрибутов, имеющих значение одновременно в текущем и обрабатываемом описаниях, но и множества имен атрибутов, связанных с ними обусловленностями.

На расширенном с помощью системы обусловленностей и непротиворечивом наборе означенных атрибутов инициируется процедура

поиска-идентификации. Обозначим через S_i' множество атрибутов, сформированное по результатам процедуры сужения, обрабатываемое в момент i , $i = 1, 2, \dots, m$. Определим через Q имеющееся на момент времени множество описаний. В качестве названия подмножества кандидатов на идентификацию на момент i с расширенным описанием, будем использовать

$Q_u^{(i)}$, $Q_k^{(i)} \subset Q$. Процедура формирования

подмножества $Q_u^{(i)}$ будет содержать следующие итерации:

$$\text{Шаг 1 } Q_k^{(1)} = K(Q, S_1'), \quad Q_u^{(1)} = U(Q, S_1') \quad (1, 2)$$

$$\text{Шаг } i \quad Q_k^{(i)} = K(Q_k^{(i-1)}, S_i'), \quad (3)$$

$$Q_u^{(i)} = K(Q_u^{(i-1)}, S_i') \cup U(Q_k^{(i-1)}, S_i'). \quad (4)$$

где K – операция выделения подмножества описаний в базе данных, таких, что значения всех означенных атрибутов совпадают с введенными

пользователем; U – операция выделения подмножества кортежей, в которых хотя бы по одному атрибуту значения совпадают с введенными, а по остальным – не определены. Итак, модель представления и обработки описаний ландшафтных фотоснимков обеспечивает непротиворечивость данных, а также ускорение поиска-идентификации за счет увеличения количества не удовлетворяющих и уменьшения количества не противоречащих поисковому образу описаний из коллекции.

С помощью функции поиска, реализованной в фотоландшафтной базе данных, была решена задача классификации параметров, характеризующих климатически обусловленные изменения лесотундровой растительности на Полярном Урале. Для целей анализа, планирования наблюдений и прогнозирования динамики высокогорных растительных сообществ в районе исследования выявлены интервалы времени между фотосъемками, которые могут быть информативными для оценки параметров, характеризующих пространственно-временную динамику лесотундровых сообществ при использовании метода сравнения разновременных ландшафтных фотоснимков. Проведенная оценка информативности повторных фотоснимков позволила выделить периоды времени для определения таких групп параметров, как морфологические изменения древостоев и отдельных особей древесной и кустарниковой растительности, изменения в структуре древостоев и пространственном положении древесной растительности.

Таким образом, фотоландшафтная база данных позволяет хранить фотоснимки и их описания, удобно и эффективно формулировать запросы к данным и получать необходимые, достаточные и не избыточные наборы фотоснимков, обеспечивающие решение конкретных прикладных задач, связанных с оценкой динамики растительных сообществ.

Работа выполнена при поддержке РФФИ (грант № 11-04-00623-а).

Список литературы

- [1] А.И. Монтиле, В.А. Попов. *Кибернетика и вуз. Межвузовский научно-технический сборник*, 1991, **26**, 69-78.
- [2] А.И. Монтиле, Е.Ю. Щербина. *Кибернетика и вуз. Межвузовский научно-технический сборник*, 1989, **24**, 142-150.
- [3] С.Г. Шиятов. Динамика древесной и кустарниковой растительности в горах Полярного Урала под влиянием современных изменений климата. Екатеринбург: УрО РАН, 2009.

Информационная система поддержки мониторинга состояния крон деревьев международного экологического мониторинга ICP Forests на территории России.

Шашков М.П.

ФГБУН «Институт физико-химических и биологических проблем почвоведения РАН»

max.carabus@gmail.com

The experience of developing the information system for crowns condition in ICP Forests ecological monitoring system in Russia described in this paper. The system based on open DBMS PostgreSQL and used to organize data of many years monitoring. The organized data array is served as a source of information for reports, GIS and modeling of ecosystems.

Экологический мониторинг ICP Forests стартовал в 1985 году. Этому предшествовали наблюдения повреждения европейских лесов, вызванных неизвестными причинами. Целью проекта стало наблюдение за влиянием загрязнений воздуха на леса. Постепенно были поставлены задачи разного уровня сложности [1]. На территории России формирование сплошной мониторинговой сети началось в 2007 году.

Данная работа посвящена описанию информационной системы обработки данных первого уровня мониторинга по состоянию крон деревьев.

1 Организация мониторинга

Основа мониторинга – приобретение знаний о пространственной и временной изменчивости состояния лесов. Так как все экосистемные процессы отследить чрезвычайно трудоёмко, мониторинг разделен на три уровня (экстенсивный, интенсивный и интегрированный мониторинг).

Мониторинг проводится на основе регулярной транснациональной сети *постоянных пунктов наблюдений* (ППН) с шагом 16x16 км (32x32 км для бореальных лесов). Мониторинг первого уровня – это ежегодное наблюдение за преждевременным опадением листвы (дефолиация) и изменением ее цвета (дехромация), а также подробное описание симптомов и повреждений на модельных деревьях.

На территории Ленинградской, Псковской, Новгородской, Калининградской, Мурманской областей и Республики Карелия сформирована сеть из более чем 700 ППН, каждая из которых представляет собой кластер из 4х точек учета (ТУ). Всего на каждой ППН описано и промаркировано как минимум 24 модельных дерева.

При закладке ППН для каждого модельного дерева измеряется периметр ствола и определяется происхождение (семенное или порослевое), а также для всей ППН производится таксация древостоя. Далее эти работы проводятся раз в 5 лет, а состояние крон описывается ежегодно. В процессе полевой работы заполняются бумажные ведомости, которые являются источником первичных данных.

2 Информационная система

Для обработки данных состояния крон с первого года нового этапа мониторинга была начата разработка специализированной информационной системы (ИС). Задачи, решаемые данной информационной системой:

1. Проверка первичных данных на корректность и связность с ранее введенными данными
2. Систематизация и собранных данных.
3. Вычисление расчетных параметров (средние показатели дехромации, дефолиации, категории санитарного состояния, а также параметры таксации древостоя) на разных уровнях организации мониторинга (ППН, лесничество, субъект).
4. Подготовка к полевым работам: генерация карточек со списками модельных деревьев для внесения данных при обследовании ППН, напоминание о необходимости проведения на ППН тех или иных работ.
5. Предоставление пространственных данных для визуализации в ГИС
6. Формирование запросов (представлений) для экспорта данных во внешние базы данных.

Для основы информационной системы была выбрана открытая клиент-серверная СУБД PostgreSQL [2]. Кроме базы данных в ИС имеется пользовательское приложение.

а. База данных

База данных состоит из 8 рабочих таблиц и 18 справочников. Справочники организованы согласно

Руководству международной методики проведения исследований по проекту ICP Forest [3].

Иерархия системы мониторинга отражена в структуре базы

1. Мониторинговая сеть
2. Постоянный пункт наблюдений (ППН)
3. Точка учета (ТУ)
4. Модельное дерево (МД)

Рабочие таблицы связаны между собой иерархическим ключом, включающим в себя код субъекта, номер ППН в данном субъекте, код точки учёта, номер модельного дерева и две цифры года мониторинга. В зависимости от таблицы (блока данных) этот ключ может иметь разную длину.

Кодировка справочников соответствует Руководству, субъектов – федеральной кодировке.

Первичная информация в базе объединена в четыре логических блока:

1. Описание ППН
2. Данные таксации древостоя
3. Оценка состояния крон
4. Описание симптомов и повреждений.

Для вывода расчетных параметров используется система запросов. По данным таксации древостоя рассчитываются формула древостоя, бонитет, полнота и страта (для этого используются справочники бонитетов, возрастов рубки (для определения группы возраста), типов леса, полнот и запасов). При вычислении параметров используются функции (хранимые процедуры): группа возраста, бонитет, возраст рубки, средняя дефолиация на ППН, средняя дехромация на ППН и другие.

2.2 Пользовательское приложение

Пользовательский интерфейс реализован IDE Turbo Delphi. Приложение содержит формы для проверки, ввода, импорта, просмотра и редактирования данных.

Вводить данные в базу можно как через визуальную форму, так и путем импорта с карточек, заполненных в виде файлов Excel (аналогичных бумажным ведомостям).

Для вывода расчетных данных предусмотрена форма работы с запросами. Полученные данные можно сохранить в виде таблицы CSV.

3 Практическое применение и перспективы развития ИС

ИС является источником данных для формирования отчетов по мониторингу ICP Forests на территории России. С 2008 года, данные из ИС импортируются в интегрированную базу данных мониторинга ЦЭПЛ РАН. ИС также является источником данных для верификации модели динамики лесных экосистем EFIMOD [4].

Пространственная информация, имеющаяся в базе, позволяет использовать её (базу) как источник данных для геоинформационных систем (ГИС). Пространственная информация представлена в виде абсолютных данных (географические координаты) и относительных (указания на субъекты и лесничества). В первом случае по данным базы можно строить точечный векторный слой, во втором случае для визуализации надо использовать уже готовый полигональный слой (административного или лесохозяйственного территориального деления).

Применение данной разработки разнообразно и затрагивает деятельность нескольких организаций, поэтому в дальнейшем необходима разработка web-приложения для авторизованного доступа к данным через Internet. Также желательно дополнение системы приложением для ввода первичных данных, работающего на базе КПК для оптимизации полевых работ.

Список литературы

- [1] Руководство по методам и критериям согласованного отбора проб, ... Часть I. Структура и функции ICP Forests. Available from <http://www.icp-forests.org/pdf/Rmanual1.pdf>. Accessed May 2012.
- [2] PostgreSQL: The world's most advanced open source database. <http://www.postgresql.org>
- [3] Manual of methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests. Part II. Visual Assessment of Crown Condition. Updated 06/2006. Available from http://www.icp-forests.org/pdf/Chapt2_compl06.pdf. Accessed June 2012.x
- [4] Chertov O., Komarov A., Loukianov A., Mikhailov A., Nadporozhskaya M., Zubkova E. . The use of forest ecosystem model EFIMOD for research and practical implementation at forest stand, local and regional levels // *Ecological modeling*, 194, 2006. P. 227-232.

Виртуальные деревья, секции и вертикальное распределение биомассы дерева

Галицкий В.В.

Институт физико-химических и биологических проблем почвоведения РАН
galvv@rambler.ru

The rhythmically growing tree's model based on conception about a virtual tree and section as elements of the tree architecture is offered. The model explains a number of situations observed in the nature and demonstrates real distributions of a green biomass along tree height.

1 Секционная модель динамики биомассы дерева

1.1 От двумерных моделей к трехмерным

Двумерные модели не всегда дают адекватное описание, например, конкуренции в сообществах деревьев [1,3]. Верхние части конкурирующих деревьев растут свободно, тогда как более низкие части активно участвуют в конкуренции за территорию. Трехмерное рассмотрение дерева облегчается использованием секции в качестве элемента морфологической структуры дерева. Биомасса секции ассоциируется с зеленой биомассой скелетных ветвей на междоузлии ствола. Ниже описывается способ разложения динамики биомассы ритмично (с шагом ΔT) растущего дерева на сумму динамик секций, распределенных по его высоте.

1.2 Виртуальное дерево и секция

Для описания динамики биомассы секции используется представление о виртуальном дереве – верхней части реального дерева, начинающейся с рассматриваемой секции [2]. Каждой i -секции сопоставляется свое виртуальное i -дерево с биомассой B_i ($i=0,1,\dots$). Динамика биомассы i -секции дерева, появившейся в T_i , есть разность биомасс смежных виртуальных деревьев, вложенных друг в друга

$${}^0b_i(T - T_i) = B_i(T - T_i) - B_{i+1}(T - T_i - \Delta T). \quad (1)$$

И.А. Полетаевым [4] была предложена балансовая модель динамики высоты H дерева, аллометрически связываемая с его биомассой B

$$H(T) = H_m \operatorname{th}(T/A_2), B(T) = B_m \operatorname{th}^u(T/A_1) \quad (2)$$

Выражения (2) используются для вычисления высоты появления виртуального i -дерева H_i и его биомассы при замене B_m на $B_{m,i}$

$$B_{m,i} = B_m f_0(x), x = H_i / H_m, f_0(0) = 1, f_0(1) = 0, \quad (3)$$

$f_0(x)$ – монотонно уменьшающаяся функция [4].

2 Результаты

С использованием (1) и предположения о монотонной ограниченной динамике (2) биомассы дерева были продемонстрированы и объяснены некоторые наблюдаемые в природе явления [2].

2.1 Акропетальное оголение ствола

Непосредственное использование (1) и (2) показывает, что динамика биомассы секции имеет колоколообразную форму и со временем должна отмирать, а ствол дерева акропетально оголяться. Такая динамика биомассы скелетных ветвей ели обыкновенной была фактически показана в [6] и акропетальное оголение наблюдается повсеместно.

2.2 “Зонтик” как предельная форма кроны

Колоколообразная форма динамики биомассы секции и ограниченность высоты дерева [4] приводят к заключению, что предельная с возрастом форма свободно растущего дерева может быть подобна зонтику. Некоторые формы деревьев, например, в саванне указывают, что этот эффект можно встретить в реальности.

2.3 Неограниченность с возрастом фитомассы

Фитомасса (физиологически пассивная часть массы) является результатом деятельности биомассы, которая, судя по сказанному выше, ограничена по величине. Несложные рассуждения [2] приводят к выводу, что фитомасса дерева с возрастом монотонно растет.

2.4 Распределение биомассы по секциям

Для многих видов акропетальное оголение стебля значительно замедляется, если не заканчивается на некоторой высоте (в некотором возрасте). Соответственно, зонтик не всегда может являться предельной формой дерева. Очевидно, что это может быть, если “колокол” ${}^0b_i$ не стремится к нулю с возрастом. Реализация этого с помощью $f_0(x)$ в (3)

а) $f_0(x) = (1-x)^r$, б) $f_0(x) = ((1-x)/(1+x))^r$, в) $f_0(x) = (1-x^r)$, (4)
приводит к различным формам распределения биомассы секций по высоте дерева, которые могут быть ассоциированы с некоторыми видами деревьев. Форма распределения биомассы по

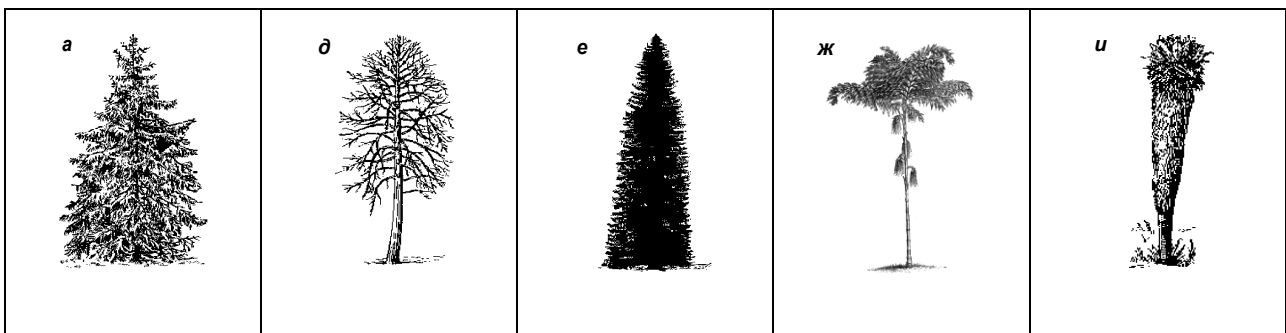
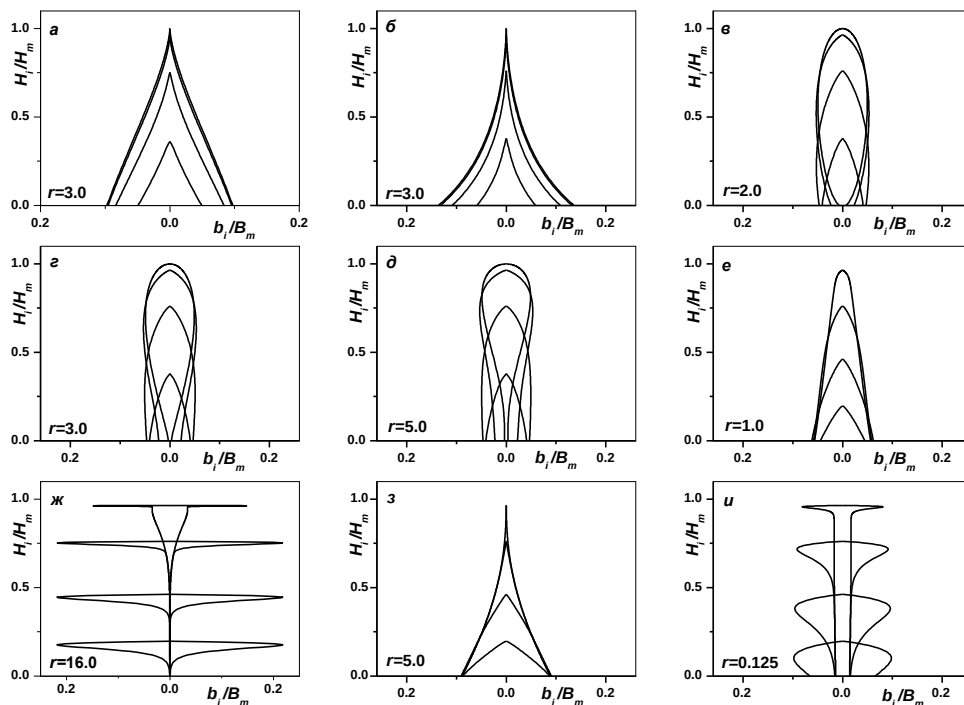
высоте может также изменяться при изменении отношения масштабов времени динамики биомассы дерева A_1 и высоты дерева A_2 .

На рис. 1 приведены примеры модельных динамик распределения биомассы для нескольких типов $f_0(x)$ и отношений A_1/A_2 : $a - (4a)$, $A_1=A_2=100$, $б - (4б)$, $A_1=A_2=100$, $в, з, д - (4в)$, $A_1=A_2=100$, $е - (4б)$, $A_1=100$, $A_2=200$, $ж - (4в)$, $A_1=5$, $A_2=200$, $з - (4а)$, $A_1=50$, $A_2=200$, $и - (4а)$, $A_1=25$, $A_2=200$, и для нескольких значений параметра r ; $B_m=1$, $H_m=100$, $\mu=2.87$. Распределения для каждого варианта приведены для четырех значений возраста дерева T : 40, 100, 200, 400, максимальная высота с возрастом увеличивается. На рис. 2 показаны примеры силуэтов реальных деревьев, которые можно ассоциировать с одноименными типами распределения, представленными на рис. 1: ель

обыкновенная (тип a), липа ($д$), ель Шренка ($е$), пальма ($ж$), эспелеция ($и$) [6].

Список литературы

- [1] V.V. Galitskii. For. Ecol. & Manag. 2003, **183**, 95-111.
- [2] В.В. Галицкий. ЖОБ, 2010, **71(1)**, 19-29.
- [3] В.В. Галицкий. Мат. биол. и биоинформатика, 2012, **7(1)**, 54-80..
- [4] И.А. Полетаев. Пробл. киберн. (16), М.: Наука, 1966, 175-177.
- [5] Т.И. Серебрякова и др. Ботаника с основами фитоценологии: Анатомия и морфология растений. М.: Академкнига. 2006.
- [6] Ю.Л. Цельникер. Лесоведение, 1994, (4), 35-44.



Базы данных для моделирования динамики живого напочвенного покрова лесных экосистем

Ханина Л.Г.⁽¹⁾, Романов М.С.⁽¹⁾, Бобровский М.В.⁽²⁾, Лукина Н.В.⁽³⁾, Грозовская И.С.⁽²⁾,
Исаева Л.Г.⁽⁴⁾, Глухова Е.М.⁽¹⁾, Михайлова Н.В.⁽¹⁾, Рубашко Г.Е.⁽¹⁾

ФГБУН Институт математических проблем биологии РАН⁽¹⁾,
ФГБУН Институт физико-химических и биологических проблем почвоведения РАН⁽²⁾
ФГБУН Центр по проблемам экологии и продуктивности лесов РАН⁽³⁾
ФГБУН Институт проблем промышленной экологии Севера КНЦ РАН⁽⁴⁾
lkhanina@rambler.ru

Modelling of forest ground vegetation is a modern challenge of the ecological modeling. Databases on plant species ecology, biomass and productivity, chemical element composition, etc. have been developed aiming the further model development.

В последние годы в области экологического моделирования стала насущной задача учета роли живого напочвенного покрова в общей структуре и функционировании лесных экосистем. Почвенный и живой напочвенный покров лесных экосистем (под последним понимается травяно-кустарничковый и мохово-лишайниковый ярусы лесного растительного сообщества) долгое время находился «вне актуальных потребностей» лесоводов и экономистов. Для планирования лесозаготовок и ведения лесного хозяйства в стационарных условиях достаточно было использования обычных расчетов и довольно простых моделей роста древостоя. В настоящее время ситуация изменилась в связи с повышенным вниманием общества к глобальным процессам изменения климата, увеличения атмосферных азотных выпадений, к вопросам сохранения биоразнообразия, проблемам устойчивости наземных экосистем к катастрофическим природным и антропогенным нарушениям. Моделирование динамики живых систем в различных условиях является основным инструментом решения этих вопросов. Учет вклада и функционирования почвенного и напочвенного покровов в процессах, связанных с устойчивостью и биоразнообразием лесных экосистем, глобальной экосистемной динамикой, стал насущно необходим.

Моделирование динамики живого напочвенного покрова (ЖНП) лесных экосистем можно осуществлять с использованием различных подходов. Один из них – моделировать во времени функционирование лесной экосистемы (состоящей из древостоя, почвы и ЖНП) с реализацией круговорота основных химических элементов [1].

Моделируемый круговорот должен зависеть от свойств видов и групп видов ЖНП, от внешних факторов (таких как температура воздуха и почвы, количество доступного азота, тип условий местопроизрастания, ценоотическое окружение лесного участка и др.), а также учитывать антропогенные и природные катастрофические нарушения (такие как рубки, пожары, ветровалы). Моделирование круговорота элементов (прежде всего, углерода, азота, кальция) в системе древостой-почва уже реализовано в системе моделей EFIMOD-ROMUL [1,6], функционирующей в настоящее время в рамках интегрированной модульной среды DLES [3]. Первая попытка моделирования динамики напочвенного покрова *per se* была предпринята ранее [2]. Следующим приближением была реализация статической модели динамики разнообразия ЖНП на основе результатов работы модели EFIMOD, использования эколого-ценоотических групп видов растений, баз данных геоботанических описаний и применения правил типа IF-THEN [7]. Нынешний этап построения модели динамики ЖНП состоит в разработке модели, встроенной в общую систему DLES и связанной с моделями EFIMOD и ROMUL как прямыми, так и обратными связями. Одной из задач, возникающей при разработке такой модели, является доступность данных по биомассе и продуктивности ЖНП, химическому составу видов и скорости разложения их опада, потребностям видов в элементах минерального питания, воде и свете. Разработка комплекса баз данных по ЖНП с целью последующего использования собранной и структурированной информации в модели динамики и круговорота элементов является целью настоящей работы.

На данный момент в коллективе имеются четыре базы данных по экологическим свойствам видов (экологические шкалы Л.Г.Раменского с соавторами, Д.Н.Цыганова, Г.Элленберга и Э.Ландольта), подробно описанные в [5]. Как

правило, в шкалах содержится балльная оценка экологического оптимума (или интервал экологической толерантности) видов по факторам увлажнения почвы, освещенности, почвенного богатства и кислотности, а также температурного режима и континентальности климата. Кроме того, в новой редакции шкал Э.Лан-дольта [8], помимо климатических и почвенных индикаторов, содержатся данные по жизненным формам, популяционным стратегиям видов, репродуктивным особенностям и способам распространения растений, длительности жизни семян в почве, глубине залегания корней и другие. Также в коллективе имеются обобщенная база данных по эколого-ценотическим свойствам видов [4] и база данных геоботанических описаний растительности Европейской России FORUS [9].

В настоящее время существующий информационный комплекс дополняется новыми базами данных с условными названиями «Биомасса» и «Элементы». В первой базе содержатся данные по биомассе и продуктивности различных компонентов ЖНП. Вторая база содержит данные по элементному составу видов и групп видов ЖНП. Источниками данных первой базы данных являются сведения из научных публикаций и результаты оригинальных научных исследований, а также другие базы данных, интернет-страницы и т.п. Информация во второй базе данных является оригинальными данными, полученными совместно ЦЭПЛ РАН и ИППЭС Кольского научного центра РАН.

Основной особенностью разработки базы данных «Биомасса» является то, что существующие данные очень неоднородны. Последнее столетие происходило накопление фактического материала по биомассе видов растений и продуктивности ЖНП. Однако этот материал разрознен, исследования выполнены по разным методикам, с разными подходами и для разных целей. Например, существуют десятки способов членения ЖНП на компоненты – функциональные группы, микрогруппировки, фракции и т.п., по которым собирались данные по биомассе ЖНП. В связи с этим, структура базы должна быть по возможности расширяемой. Помимо самих данных, необходимо хранить сведения об источнике информации, географической, экологической и ценотической привязках материала. Практически любой фрагмент данных является независимым от других, но в то же время требуется группировка и обобщение данных. В настоящее время база «Биомасса» содержит 28 таблиц, включая 7 основных, 3 списка флоры (сосудистые растения, мохообразные, лишайники) и 18 вспомогательных таблиц. Собственно измерения (Biomass) выполнены в пределах пробных площадей (Plots), которые, свою очередь, можно позиционировать в пределах районов исследования (Study areas). Также пробные площади имеют

привязку к литературным источникам (References). Таблицы References и Study areas представляют собой наиболее крупный иерархический уровень информации, они связаны друг с другом отношением «многие-ко-многим» (через таблицу Plots), так как в одной публикации могут быть представлены сведения из разных районов исследований, а одному району исследований может быть посвящено несколько публикаций. На настоящий момент в базе имеется 3770 единиц информации о биомассе или продуктивности ЖНП.

Структура базы «Элементы» состоит пока из 2х таблиц, но в будущем она усложнится, так как еще не выполнена работа по классификации элементов ЖНП. На настоящий момент в базе имеются сведения о содержании элементов (N, C, Ca, Mg, K, Al, Fe, Mn, Zn, Ni, Cu, P, S) в 2610 объектах ЖНП.

Следующим шагом работы является исследование баз данных на предмет выделения функциональных групп видов растений, которые будут использоваться при моделировании динамики ЖНП.

Работа выполнена при поддержке Программы № 30 фундаментальных исследований Президиума РАН «Живая природа: современное состояние и проблемы развития», подпрограмма «Динамика лесных экосистем», проект «Экосистемные функции живого напочвенного покрова бореальных лесов Европейской России».

Список литературы

- [1] Комаров А.С., Чертов О.Г. и др. Моделирование динамики органического вещества в лесных экосистемах. М.: Наука. 2007. 380 с.
- [2] Михайлов А.В. В: Г.Ю.Резниченко (ред.) *Математика. Компьютер. Образование*, 2001, 8(2), 651-655.
- [3] Михайлов А.В., Безрукова М.Г. и др. *Известия Самарского научного центра РАН. Сер. Биологическая*, 2009, 11(1/7), 1522-1527.
- [4] Смирнов В.Э., Ханина Л.Г., Бобровский М.В. *Бюлл. МОИП. Сер. Биологическая*, 2006, 111(2), 36-47.
- [5] Ханина Л.Г., Заугольнова Л.Б. и др. В: *Оценка и сохранения биоразнообразия лесного покрова в заповедниках Европейской России*. М.: Научный мир, 2000.
- [6] Chertov O.G., Komarov, A.S. et al. *Ecological Modelling*, 2001, 138(1-3), 289-308.
- [7] Khanina L., Bobrovsky M. Et al. *Forest Ecology and Management*, 2007, 248, 80-94.
- [8] Landolt E., Bäumler B. et al. *Ecological Indicator Values and Biological Attributes of the Flora of Switzerland and the Alps*. Haupt-Verlag, Bern. 2010. 376 pp.
- [9] Smirnova O., Zaugol'nova L. et al. В: В.Д. Лахно (ред.) *Математическая биология и биоинформатика*, 2006, 150-151.

Статистическое оценивание некоторых параметров видового разнообразия по проблемным экологическим данным

Смирнов В.Э.

Институт математических проблем биологии РАН
vesmirnov@rambler.ru

Species richness is one of the main criteria when assessing state and/or dynamics of natural ecosystems. However, numbers of species are only comparable if sampling efforts are standardized in an appropriate way. A case study of the Voronezh nature reserve vegetation illustrates: 1) how one may apply resampling methods to bring samples of different sizes to a common ground; 2) how one may test significant differences in species richness; 2) how some properties of ecological data, especially the number of rare species only represented in one or two sample units (singletons and doubletons), can be used to help estimating and interpreting species richness correctly.

1 Введение

Видовое разнообразие складывается из трех иерархически связанных компонентов (уровней): альфа-, бета- и гамма-разнообразие [3]. Под ними, соответственно, понимается: 1) разнообразие в одной выборке стандартного размера (локальное разнообразие); 2) вариация разнообразия в наборе выборок (скорость изменения ценоза вдоль пространственного или временного градиента или просто степень ценотической неоднородности безотносительно какого-либо градиента); 3) общее разнообразие в наборе выборок (как правило, разнообразие на ландшафтном уровне). Все три уровня разнообразия возможно оценить отдельно, с помощью различных количественных мер. Настоящая работа посвящена некоторым проблемам, возникающим при измерении первого и третьего уровней разнообразия растительности, и в качестве оценок взяты основные, наиболее часто используемые меры: для альфа-разнообразия – видовая насыщенность (среднее число видов на единицу площади), для гамма-разнообразия – общее число видов на территории.

При оценке многолетней динамики разнообразия растительности исследователь может столкнуться (и, как правило, сталкивается) со следующими проблемами в данных.

1. Описания растительности выполнены разными авторами, ставящими различные задачи и выполняющие полевые работы с различной степенью «надежности».

2. В зависимости от задач могут варьировать как схемы закладки геоботанических площадок для проведения описаний, так и их размеры.

3. За тот или иной временной период доступно неодинаковое число описаний.

Все указанные факторы могут смещать оценки разнообразия (проблемы 1 – 2 – более всего видовую насыщенность, 3-я – оценку гамма-разнообразия) и, таким образом, делать данные за разные периоды несопоставимыми. В настоящей работе предпринята попытка выявить средствами статистического моделирования возможное влияние указанных факторов на оценки разнообразия и при необходимости попытаться скорректировать эти оценки.

2 Методы

Исследование проводили на основе анализа 1047 геоботанических описаний, собранных на территории Воронежского заповедника в разных растительных сообществах разными исследователями на площадках разного размера с 1929 по 2011 гг.. Около половины всех площадок (546) были расположены на геоботаническом профиле – трансекте, состоящей из двух рядов кварталов общей протяженностью 22 км и шириной 1 км, пересекающей заповедник в наиболее широкой его части с запада на восток. Подробное описание геоботанических данных и растительности заповедника приведены в работе [4]. Динамику разнообразия растительности оценивали по трем временным периодам: первый – 1930-1940 гг., второй – середина 1960-х гг., третий – середина 1980-х – по настоящее время.

Все расчеты выполнялись в среде статистического программирования R с подключением библиотек *rich*, *labdsv*, *vegan*, *BiodiversityR* и др. [1].

Значимость различий между видовой насыщенностью проверялось с помощью рандомизационных тестов [2]: описания сравниваемых групп случайно перемешивались, каждый раз рассчитывалась разность между групповыми средними. Наблюдаемая разность затем сравнивалась с полученным распределением. Выполнялось по 999 пермутаций, что соответствует минимально возможному уровню значимости $\alpha = 0.001$; нулевые гипотезы отклонялись при $p < 0.05$. Использовалась функция *c2m* библиотеки *rich*.

Модельные оценки для гамма-разнообразия (общего числа видов в заповеднике) для второго и

третьего периодов получены методом бутстрепа [2] реализованного в виде оригинальной функции. Процедура основана на случайном выборе без возвращения 127 описаний 10000 раз (127 – объем доступных данных по первому периоду, минимальный из трех) и получения распределения общего числа видов, ожидаемых в подвыборках такого объема. В качестве оценок брались средние модельных распределений и их перцентильные доверительные интервалы (доверительная вероятность 0.95).

3 Результаты

1. Влияние площадок разного размера. Рандомизационный тест показал, что значительно отличаются по видовой насыщенности описания, выполненные на площадках менее 100 кв. м и выполненные на площадках размером в 100 кв.м. Площадки малого размера (всего 133) были удалены из дальнейшего анализа. Видовая насыщенность площадок большего размера, чем стандартный, значительно не отличалась, и такие площадки были оставлены.

2. Сравнение гамма-разнообразия по периодам. Наблюдаемое число видов: для первого периода – 311, для второго – 335, для третьего – 425. Эти оценки несравнимы из-за разного числа описаний, доступных за каждый период. Методом бутстрепа получены модельные оценки общего числа видов для второго и третьего периодов на подвыборках размером в 127 описаний: 272.1 и 297.4, соответственно. Доверительный интервал для третьего периода захватывает наблюдаемое число видов для первого периода и ожидаемое число видов второго периода; доверительный интервал для второго периода не перекрывается с доверительным интервалом для третьего (смещен влево) и не накрывает наблюдаемое число видов для первого периода.

3. Влияние «авторства описаний». Во втором периоде, помимо самой низкой оценки гамма-разнообразия, наблюдается также и наименьший процент редких видов (суммарно singletons и doubletons второго периода составляют 28% против 36% в других периодах). Дополнительный анализ показал, что «недобор» видов, а также аномально низкая видовая насыщенность (15.7) имеет место в описаниях экспедиции Института географии АН СССР (ИГАН). Без этих описаний характеристики видового разнообразия второго периода становятся близки к оценкам первого и третьего периодов. Насыщенность возрастает до 20.3, процент редких видов – до 38, и если к модельной оценке общего числа видов второго периода добавить разницу в доле редких видов (10%), то ожидаемое гамма-разнообразие увеличивается до 300, что практически соответствует модельной оценке третьего периода. Таким образом, нет оснований предполагать реальное уменьшение гамма-разнообразия заповедника во втором периоде и значимое уменьшение гамма-разнообразия за исследуемое время в целом.

4. Сравнение схем закладки площадок. Описания, сделанные на профиле и вне его, не отличаются значимо по видовой насыщенности для всех трех периодов (для второго – после удаления описаний ИГАН). Статистический вывод позволяет далее сравнивать видовую насыщенность между периодами без разделения описаний, сделанных на профиле и вне его.

5. Сравнения видовой насыщенности между периодами. Наблюдается значимое уменьшение в ряду от первого к третьему периоду (для второго – без учета описаний ИГАН), которое подтверждается таким же закономерным уменьшением доли описаний типов сообществ, богатых видами (лугово-опушечные, степные) и увеличением доли описаний лесных (прежде всего, неморальных) сообществ, а также уменьшением видовой насыщенности во всех типах сообществ.

4 Заключение

Проведенный анализ показывает принципиальную возможность оценки многолетней динамики разнообразия растительности по доступным данным. Вместе с тем, любой из перечисленных во введении факторов может смещать оценки как альфа-, так и гамма-разнообразия, причем в любую сторону. Для выявления и коррекции такого влияния необходимо проводить статистическое (компьютерное) моделирование современными методами, основанными на процедурах монте-карловского типа (бутстреп, рандомизационные тесты) с учетом: 1) общих представлений о сукцессионном процессе; 2) специфики растительности конкретной территории; 3) особенностях ее описания.

Для рассматриваемого в работе примера – динамика разнообразия растительности Воронежского заповедника – предложенная методика статистического моделирования позволила получить сравнимые оценки видового разнообразия за три периода проведения исследований. Показано, что за 80 лет, прошедших с начала заповедания, гамма-разнообразие растительности в целом сохранилось на прежнем уровне; альфа-разнообразие уменьшилось. Наши результаты в целом подтверждают, а в деталях уточняют описанный ранее процесс уменьшения разнообразия на первых стадиях восстановления растительности после нарушений [5].

Список литературы

- [1] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2011. Vienna, Austria. <http://www.R-project.org/>.
- [2] Sokal, R. R. and F. J. Rohlf. Biometry: the principles and practice of statistics in biological research. 2011. NY: W.H. Freeman and Co. 937 pp.
- [3] Whittaker R.H. *Taxon.*, 1972, 21, 213-251.
- [4] Стародубцева Е.А., Ханина Л.Г. *Растительность России*, 2009, 14, 63-141.
- [5] Смирнова О.В. (ред.) *Восточноевропейские леса*. 2004. М.: Наука. Т. 2. 575 с.

Оценка вибрационного воздействия транспортных средств на экологическую обстановку в городе

Дергузов А.В., Махортых С.А.

Институт математических проблем биологии РАН, Пушкино
arkadiid@list.ru

Vibration and noise are important physical parameters that have an impact on the environment in an urban setting. An estimation method for vibro-acoustic impact of transport on the environment. An example is given for a specific case in Moscow.

1 Методы оценки вибрации и шума

Линии метрополитена являются источником повышенной вибрации зданий, расположенных в зоне их влияния. В связи с этим возникает задача прогнозирования уровней вибрации в жилой застройке, прилегающей к проектируемым и строящимся линиям. Другой, не менее важной задачей является оценка эффективности различных мероприятий по защите зданий и сооружений как от проектируемых, так и существующих линий метрополитена.

Проблема расчета уровней вибрации в грунте в городских условиях представляется довольно сложной по нескольким причинам. Грунт является довольно сложной геофизической средой с присущими ему гетерогенностью, сложными поглощающими и упругими свойствами. Часто проблема осложняется также нехваткой данных о геометрических характеристиках и упругих свойствах слоев грунта. Данная проблема решается в рамках экспериментальной методики с последующим решением обратных задач томографии грунта с использованием спектрально-аналитических подходов [1].

Вследствие сложности вибрационного поля и отсутствия достаточно подробных данных о геометрических и физических свойствах слоев грунта в условиях нехватки геофизических данных используется однородная модель грунта, которая обеспечивает приемлемую точность и физическую прозрачность результатов, не приводя к избыточной математической сложности.

Проведены теоретические исследования колебаний обделок тоннелей метрополитена различного типа, рассматриваемых как упругие оболочки, создана математическая модель упругих волн в грунте, учитывающая распространение продольных, поперечных, а также поверхностных волн Рэлея. Учет последних в практике акустических расчетов, связанных с метрополитеном, производится впервые. Показана

важность роли поверхностных волн при передаче энергии вибрации к жилой застройке.

Учитываются также особенности передачи вибрации по структуре здания. Предлагаемые в данном случае подходы адаптируются также к задачам метеорологического воздействия на высотные сооружения.

Достоверность полученных в работе результатов подтверждается натурными экспериментами на действующих линиях метрополитена, а также многочисленными расчетами и практикой применения разработанных методик в Москве, Санкт-Петербурге, Екатеринбурге, Нижнем Новгороде.

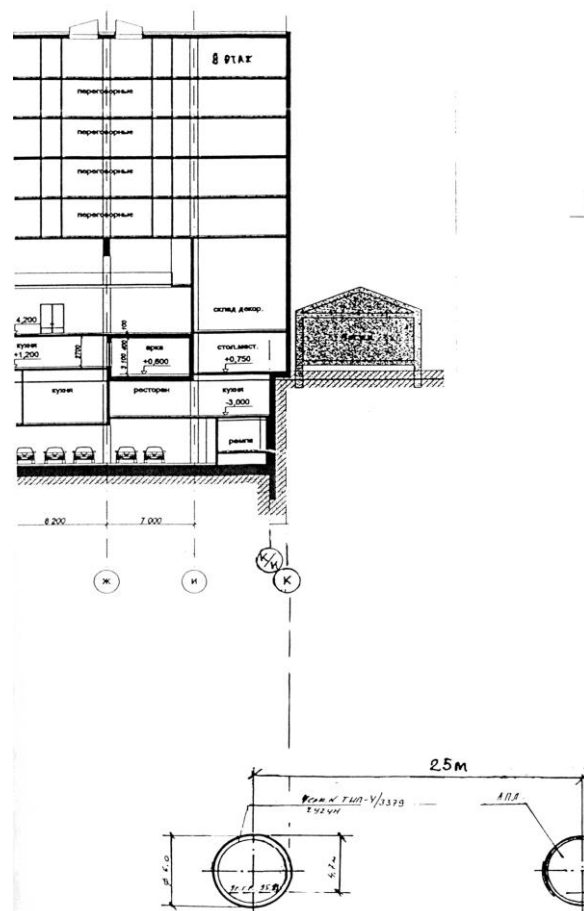


Рис.1 Взаимное расположение линий метрополитена и здания.

В качестве примера оценки параметров вибрационного поля на структуру здания приводится случай воздействия вибрации и шума от движения электроподвижного состава на участках

Арбатско-Покровской и Филевской линий метрополитена от ст. метро Арбатская до ст. метро Смоленская вблизи проектируемого многофункционального здания по ул. Арбат, вл. 24 и 26/2/3 стр. 1, Большой Новопесковский пер., вл.4, стр.2, 5. На Рис.1 представлено взаимное расположение линий метрополитена и здания.

Дополнительно были проведены измерения и выполнены расчеты в соседнем со строительством зданием кассовых залов театра им. Вахтангова для оценки влияния вибрации от движения составов метрополитена на состояние строительных конструкций существующих строений.

Средства измерения: Шумомер-анализатор спектра фирмы "Larson-Davis" типа LD2900A АО-332 (в составе: микрофон типа 2540 №2012 с предусилителем типа 900B №2531; акселерометр типа PCB 353B33 заводской номер №11092; акселерометр типа PCB 353B33 заводской номер №11093); калибратор фирмы "Brüel & Kjær" типа 4294 заводской номер №1145111 и пистонфон фирмы "Brüel & Kjær" типа 4220 заводской номер №284904.

Для расчетов уровней вибрации на поверхности грунта на примагистральной территории использовалась модифицированная версия

программы VIBRACALC 2.1, разработанная в Институте математических проблем биологии РАН для расчетов вибрации, возбуждаемой поездами метрополитена. В модифицированной версии учитывается динамическое воздействие на поверхность земли, а также конкретное строение несущей конструкции пути. Используя вышеперечисленные выкладки и описанные программы были проведены расчеты возбуждаемых уровней вибрации вблизи тоннеля на примагистральной территории. В основу расчета положены параметры метropolездов, характерные для действующего в России подвижного состава метрополитена. В ходе процедуры вычислений производится перерасчет уровней вибрации на стандартной обделке (кругового сечения) в уровни вибрации на реально рассматриваемых обделках перегонов Арбатская – Смоленская Арбатско-Покровской и Филевской линий с заданными грунтовыми условиями. В настоящее время в качестве основы для нормирования вибрации в помещениях жилых и административных зданий используются СН 2.2.4/2.1.8566. Результаты расчета и измерения уровней виброускорения в трех расчетных точках представлены в таблице 1.

Таблица 1. Максимальные уровни виброускорения в кассовом зале театра

	Уровни виброускорения в октавном диапазоне, дБ						
	2 Гц	4 Гц	8 Гц	16 Гц	31,5 Гц	63 Гц	125 Гц
Расчет	69,0	67,8	67,6	70,6	82,6	80,8	86,2
Измерения	71,6	70,2	69,3	70,6	78,7	84,7	86,4
Предельно допустимые уровни виброускорения	115	118	121	123	126	129	132

2 Выводы

Произведено сравнение измеренных и рассчитанных значений вибрации с требованиями санитарных норм. Произведен также расчет ожидаемых уровней структурного шума от движения поездов метрополитена в помещениях третьего этажа здания (зал драмтеатра) и на уровне нижнего жилого этажа. Анализ результатов расчета позволяет сделать вывод о том, что предлагаемая методика является достаточно эффективной, позволяя получить приемлемое соответствие с реально измеряемыми в натурных экспериментах величинами

вибрации и шума.

Работа поддержана проектами РФФИ 11-07-00519 и 10-01-00609.

Список литературы

- [1] Рыбак С.А., Махортых С.А., Костарев С.А., Дергузов А.В. Методы оценки акустических параметров грунта. *Акустич.журнал.* 2004, т.50, №3, с. 1-9.
- [2] С.А.Костарев, С.А.Махортых, С.А.Рыбак. Свод правил «Оценка вибрации при проектировании, строительстве и эксплуатации объектов метрополитена». СП 23-105-2004. Госстрой России, М.: 2004.

Концепция базы данных почв Башкортостана

Давлетшина М.Р.

Башкирский государственный аграрный университет
davletshina@yandex.ru

Почва представляет собой пространственно-временную функциональную систему, которая динамически изменяется, следовательно, актуальными являются вопросы создания новых и корректировки существующих почвенных карт.

Глобальные закономерности, как правило, выявляются на основе комплексного анализа многолетних наработок сведенных в единую базу данных. В лаборатории почвоведения Института биологии УНЦ РАН и кафедре почвоведения БГАУ активные исследования ведутся с 60-х годов. В области почвенных исследований накоплен большой объем информации – это материалы экспедиций и лабораторных опытов, результаты исследований многих поколений. В течение этих лет информация накоплена в различных формах, часто в не систематизированном виде в картотеках лаборатории и в публикациях сотрудников. С 1995 года разрабатываются актуальные вопросы биогеоценологии – круговорот и трансформация органического вещества в различных биогеоценозах. Основное внимание уделяется соединениям углерода, азота, содержанию тяжелых металлов.

Применение ГИС-технологий для решения задач по систематизации и обработке данных дает возможность эффективного и оперативного использования, организации упорядоченных информационных потоков. Они сочетают в себе достижения машинной графики и баз данных, обеспечивая оценку и визуализацию пространственно-распределенных данных. Исследования в этом направлении успешно проводятся во многих регионах. Однако, в Республике Башкортостан на сегодняшний день вопросы применения информационных моделей в оценке почв недостаточно изучены. Поэтому становится очевидным, что для многосторонней характеристики агроэкологических свойств почвы, ранжирования участков почвенного покрова и осуществления мониторинга почв районам РБ необходимо автоматизированное картографирование почвенного покрова на базе геоинформационных технологий. Основным преимуществом ГИС является наиболее естественное для человека представление как собственно пространственной информации, так и любой другой информации, имеющей отношение к объектам, расположенным в пространстве (так называемой атрибутивной информации). Способы

представления атрибутивной информации различны: это может быть числовое значение, таблица из базы данных о характеристиках объекта, его фотография.

С этой целью ставятся следующие задачи:

- составить карту продуктивности пахотных земель Башкортостана;
- построить многофакторные модели взаимосвязи плодородия почв и продуктивности почв сельскохозяйственного назначения в различных зонах республики;
- оставить крупномасштабные почвенные карты охраняемых территорий Республики Башкортостан;
- полученный за многолетний период обширнейший и разнообразный объем данных по различным характеристикам почв Башкортостана требует систематизировать и унифицировать;
- для работы с геоданными спроектировать базу данных «Почвы Башкортостана».

На первоначальной стадии разработки почвенной геоинформационной системы необходимо провести большую работу по созданию базы данных, в которой интегрируются картографическая, аналитическая, текстовая информация об объекте исследования. Основное назначение БД как главного звена ГИС заключается в обслуживании информационных потребностей пользователя. Создание БД предполагает структуризацию, классификацию информации по уровням иерархии. База данных должна:

- содержать строго классифицированные данные без избытка и противоречивости;
- учитывать вновь вносимые изменения данных;
- связывать различные информационные блоки по общим параметрам.

Главной особенностью геоинформационных данных является их послойное представление.

ГИС позволит моделировать отдельные почвенные ареалы и связывать с картографическими объектами некоторую атрибутивно - семантическую информацию, хранящуюся в базе данных. К такой информации относятся: наименование подобъектов их морфометрические характеристики, качественные и количественные характеристики и т.д. Каждый объект слоя может иметь сложную форму (структуру) и описываться разными стилями, например, на разных масштабах карты или даже в

пределах одного масштаба.

БД позволяет пользователю выбрать объект на электронной карте и сразу же получить о нем все его характеристики. Впоследствии при изменении данных, появлении новых экспериментальных показателей БД корректируется, дополняется.

В качестве инструментального средства для создания БД может быть выбрана одна из современных систем управления базами данных (СУБД). С этой целью предлагается провести исследования по созданию единой цифровой базы данных. В целях обобщения и систематизации многолетних данных о почвах Башкортостана собранных лабораторией почвоведения УНЦ РАН и кафедрой почвоведения и общего земледелия БГАУ, предлагается на базе стандартных СУБД создать оболочку для базы данных «Почвы Башкортостана»

Блок ввода данных должен содержать 10-15 позиций ввода, часть которых представляют собой почвенные характеристики, другая часть – сельскохозяйственное использование, и несколько дополнительных.

В целях удобства работы и стандартизации вводимых данных ввод должен осуществляться с помощью выпадающих списков вариантов ввода. Данные списки вариантов автоматически корректируются по мере ввода информации, что позволит оперативно вводить большие объемы информации. В качестве основы можно выбрать региональную классификацию почв, представленную в монографии «Почвы Башкортостана» 2 Т.. Подобная стандартизация предоставит возможность конвертировать БД в любую российскую либо международную классификацию. Дополнительные данные могут содержать такую информацию как место и год закладки разреза, а также ссылка на публикацию, позволят при необходимости в любой момент найти исходные данные. Для каждого разреза необходимо ввести достаточное количество горизонтов. Для каждого горизонта, в свою очередь, должны вводиться соответствующие физико-химические, химические параметры и данные мех состава.

При построении базы данных необходима формализация экспериментального материала: классификация информации, ее структурирование и унифицирование. Классификация материала определяет логическую схему создания файлов по отдельному признаку – территориальному, тематическому, масштабному. Возможно использование БД, составленной в разных форматах: Excel, Access и др. В условиях Республики Башкортостан при создании тематической карты на базе ГИС-технологий большое значение имеет составление тематических

электронных слоев рельефа, растительности, гидрологии, содержания органического вещества почвы и др. На первом этапе в хозяйстве создаются цифровые модели территории (ЦМТ) с привязанными данными и БД для последующего моделирования. В качестве исходных данных необходимо использовать базу данных почв хозяйства с характеристиками агроэкологических свойств. Это позволит моделировать динамику природных экосистем в условиях антропогенных нагрузок (Дитц, Смоленцев, 2002).

Контроль состояния почвы включает комплекс измеряемых параметров и показателей, которые содержатся в БД. В наших исследованиях предполагается конкретная информация БД для отдельных почвенных разрезов в определенный момент времени. В этот блок входят следующие данные:

- определение потерь запасов и содержания гумуса с изменением мощности гумусовых горизонтов в связи с развитием водной эрозии;
- выявление участков с нарушенным балансом основных элементов питания;
- выявление участков почвы, загрязненных тяжелыми металлами;
- контроль за загрязнением почвы нефтепродуктами;
- контроль за водно-физическими свойствами.

Выводы

Разработана информационная модель почвенной системы, обеспечивающая на единой основе интеграцию данных, а также оперативное и качественное представление информации для принятия решений при управлении почвенными ресурсами. Особенностью разработанной БД является комплексный и системный подход, учитывающий иерархию и индивидуальную привязку отдельных элементов к конкретным географическим условиям. Проектируемая БД дает возможность для комплексного анализа почвенных ресурсов, используя компьютерное моделирование и прогноз возможных изменений почвы.

Список литературы

- [1] Дитц Л.Ю., Смоленцев Б.А. Геоинформационная система в почвенной картографии. – Новосибирск: Наука, 2002. -78 с.
- [2] Симакова М.С. О компьютерной картографии почв // Почвоведение. – 2002. -№2. С 133-139.
- [3] Шапиро, М.Б. Новая методика составления крупномасштабных агроэкологических почвенных карт [Текст]/ Почвоведение. 2008.№5. С 626-628.

Алфавитный указатель авторов

- Almeida M. 26
Bdikin I.K. 22
Boyda D.L. 16
Bystrov V.S. 22, 24, 26
Bystrova A.V. 26
Bystrova V.V. 26
Costa E. 26
Coutinho J. 26
Dekhtyar Yu. 26
Faber J. 168
Glyakina A.V. 155
Gorbunov K.Yu. 70
Heredia A. 22
Junaid M. 153
Kholkin A.L. 22
Kopyl S. 22, 24, 26
Lapins M. 153
Levitsky V.G. 72
Likhoshvai V.A. 72
Lunin V.Y. 143
Lyubetsky V.A. 70
Maevskij E.I. 24
Maiorov M. 24
Matushkin Yu.G. 72
Novák M. 168, 172
Orlov Yu.L. 72
Paramonova E.V. 22, 26
Santos C. 26
Shirmovsky S.Eh. 16
Shutava I. 153
Sobolev O.V. 143
Sousa C.M.A. 24
Svítek M. 170
Votruba Z. 172
Wikberg J. ES. 153
Абакумов А.И. 121, 137
Айду Э.А.И. 34
Акбердин И.Р. 108, 119
Алейников А.А. 188
Александров А.А. 113
Апонин Ю.М. 135
Апони́на Е.А. 135
Атауллаханов Ф.И. 110
Афонников Д.А. 147
Безрукова М.Г. 184
Беликов С.И. 52
Беловолова О.О. 110
Бердышев В.И. 39
Бобровский М.В. 198
Богатырев М.Ю. 86
Богатырёва Н.С. 157
Борисова И.А. 84
Борисюк Р.М. 176
Бурилко О. 176
Бурцев М.С. 164
Быстров В.С. 36, 92, 102
Быстрова В.В. 36
Быстрова Н.К. 92
Вакурин В.С. 86
Викулова Н.А. 46
Водолагина Н.Н. 82
Воробьев Ю.Н. 147
Гайнова И.А. 119
Галзитская О.В. 155
Галицкий В.В. 194, 200
Гарбузинский С.А. 157
Глухова Е.М. 202
Голомысова А.Н. 32
Голубев С.Н. 190
Гольшев М.А. 58
Грабарник П.Я. 188
Грозовская И.С. 202
Гуменюк А.С. 54, 56
Гусев В.Д. 52
Давлетшина М.Р. 209
Давыдов О.Д. 133
Дедус Ф.Ф. 145, 149
Дергузов А.В. 206
Дроздов-Тихомиров Л.Н. 113
Дюбанов В.В. 84
Дюк В.А. 59
Еременко З.С. 174
Ермак Т.В. 108
Ершов Н.М. 158
Ефимов А.В. 145
Журавлев Ю.И. 98
Загоруйко Н.Г. 84
Зимин А.А. 78
Зотов А.К. 88
Зотова Т.Ю. 88
Зубаирова У.С. 146
Иванков Д.Н. 157
Иванова Н.В. 182
Израильский Ю.Г. 121
Исаева Л.Г. 202
Казанович Я.Б. 176
Казанцев Ф.В. 108
Камышлов В.В. 102
Кацнельсон Л.Б. 39, 46
Каширина Н.И. 28
Киселев С.С. 66
Ковалёв П.В. 113
Коваленко И.Б. 30
Когай В.В. 117
Козлов Н.Н. 74
Кокорев Л.С. 92
Константинова М.В. 98
Коротков Е.В. 58, 61, 63
Коршунова А.Н. 8
Костомарова И.В. 82
Кравчук А.В. 158
Крауклис И.В. 151
Крылов А.К. 114
Кузнецова А.В. 82
Кулаков С.В. 174
Куликова Л.И. 145, 149
Курсанов А.Г. 46
Кутненко О.А. 84
Кутыркин В.А. 48, 50
Лахман К.В. 164

- Лахно В.Д. 8, 10, 12, 14, 20, 28
 Леоненко В.Н. 76
 Лихошвай В.А. 108, 117, 119, 141
 Лукина Н.В. 202
 Лыжко Е.В. 178
 Люлякин О.П. 129
 Маевский Е.И. 92
 Мазуров М. Е. 41
 Макашева В.А. 119
 Малыгина Н.А. 82
 Мархасин В.С. 39, 46
 Марчук Ю.В. 133
 Мастеров В.Б. 186
 Махортых С.А. 100, 178, 206
 Медведев К.Е. 147
 Микулинская Г.В. 78
 Мироненко Л.М. 192
 Мирошниченко Л.А. 52
 Михайлова Н.В. 202
 Молчанов А.А. 131
 Монтиле А.А. 133
 Монтиле А.И. 133
 Морнев О.А. 43
 Мурзина Г.Б. 166
 Назаренко Г.И. 98
 Назипова Н.Н. 78, 80
 Наумов Д.Г. 68
 Нигматуллина Л.Ф. 78
 Николаев С.В. 116
 Никольский М.С. 131
 Озолин О.Н. 66
 Орлов М.В. 111
 Осовец С.В. 94
 Панкратов А.Н. 65, 145, 149
 Панкратова Н.М. 104
 Пантелеев М.А. 32, 110
 Панфилов А.В. 39
 Панюков В.В. 66
 Парамонова Е.В. 36, 92
 Паутова Л.А. 137
 Перцев Н.В. 76, 127, 139
 Пичугина А.Н. 127
 Поздниченко Н.Н. 54, 56
 Покровский А.Н. 160
 Пономарев О.А. 10, 20
 Попцова М.С. 45
 Правдин С.Ф. 39
 Пугачева В.М. 63
 Пучкова А.И. 110
 Пятков М.И. 65
 Редько В.Г. 162
 Ри Н.А. 108, 117
 Романов М.С. 186, 202
 Рубашко Г.Е. 202
 Руднев В.Р. 145, 149
 Рыкунов С.Д. 180
 Рязанов В.В. 98
 Саранча Д.А. 123, 125, 129
 Свиначев А.И. 10
 Сенкевич Ю.И. 96
 Сенько О.В. 82
 Силкин В.А. 137
 Смирнов В.Э. 188, 204
 Соболев Е.В. 18
 Соколова Е.В. 151
 Соловьёва О.Э. 39, 46
 Сохова З.Б. 162
 Суворова Ю.М. 61
 Султанов В.Б. 12
 Сычев В.В. 180
 Твердислов В.А. 45
 Теплухина Е. 80
 Тимонов В.А. 108
 Титкова Т.Н. 52
 Титомир Л.И. 34
 Тихонов Д.А. 18, 145
 Трапеев Р.В. 125
 Трифонова Е.А. 119
 Трофимова Г.Ю. 184
 Трунов В.Г. 34
 Тулуб А.В. 151
 Тюльбашева Г. 80
 Уральский Л.И. 113
 Устинин М.Н. 104, 180
 Фадеев С.И. 117
 Фиалко Н.С. 14, 18
 Финкельштейн А.В. 157
 Хамеева Н.А. 151
 Ханина Л.Г. 202
 Хлебодарова Т.М. 108, 117, 141
 Цукерман В.Д. 174
 Цыганкова И.А. 90
 Чалей М.Б. 48, 50, 80
 Черномордик О.М. 88
 Шалаумова Ю.В. 192
 Шашков М.П. 182, 198
 Шигаев А.С. 10, 20
 Шилов Е.Б. 110
 Шорина А.Р. 119
 Шпынов С.Н. 54
 Юрезанская Ю.С. 123
 Якушевич Л.В. 37