

ТОПОЛОГИЯ ЭНЕРГЕТИЧЕСКОГО ЛАНДШАФТА МАКРОМОЛЕКУЛ В ПРОСТРАНСТВЕ ТОРСИОННЫХ УГЛОВ И ПРИНЦИП МИНИМУМА СКОРОСТИ ДИССИПАЦИИ ЭНЕРГИИ ПРИ КОНФОРМАЦИОННОЙ РЕЛАКСАЦИИ

© 2017 г. К.В. Шайтан* **

*Биологический факультет Московского государственного университета имени М.В. Ломоносова,
119991, Москва, Ленинские горы, 1/12

**Институт химической физики им. Н.Н. Семенова РАН, 119991, Москва, ул. Косыгина, 4

E-mail: shaytan49@yandex.ru

Поступила в редакцию 26.06.17 г.

Обсуждены перспективы методов молекулярной динамики в применении к сложным задачам структурной биологии, моделирования фолдинга белков и предсказания пространственной структуры биополимеров. Для энергетического ландшафта систем с конформационной подвижностью сформулировано многомерное разложение в ряд Фурье, которое является корректным с точки зрения топологии конфигурационного пространства. Для систем с выделенным глобальным минимумом на энергетическом ландшафте дано правило соответствия фаз для членов многомерного разложения, которое формально похоже на задачу дифракции на многомерной кубической решетке. Обсуждена проблема калибровки силовых полей для биополимеров и их соответствия топологически корректным энергетическим ландшафтам. Получены уравнения движения в матричной форме для релаксации репрезентативной точки на многомерных поверхностях потенциальной энергии. Показано, что решения уравнений для конформационных движений при заданной скорости релаксации потенциальной энергии (или, в данном контексте, заданной скорости сворачивания цепи) удовлетворяют принципу минимума скорости диссипации энергии.

Ключевые слова: структурная биология, молекулярная динамика, суперкомпьютерные технологии, энергетические ландшафты, топология энергетической поверхности, конформационная подвижность, функциональная динамика биополимеров, фолдинг, принцип минимума скорости диссипации энергии.

Современный этап изучения структуры и динамики биомакромолекул и их комплексов характеризуется быстрым развитием новых перспективных методов исследования [1–3]. Круг изучаемых в настоящее время биологических объектов становится весьма широким. Это и традиционные относительно крупные и хорошо сформированные белковые кристаллы, микро- и нанокристаллы, получаемые различными методами, в том числе и с использованием подходов для внутриклеточной кристаллизации [4], отдельные макромолекулы и их комплексы, включая хроматин, вирусы, нано- и липидики с включением мембранных белков [5].

Развитие структурных исследований приводит к достаточно быстрому увеличению знаний о пространственной организации белков и других биополимеров и их комплексов [6], что важно для развития биомедицины, биоинженерии, биотехнологии (включая и нанобиотехнологию). Соответствующая информация важна

для понимания физических механизмов функционирования биоструктур, для прогнозирования функциональной активности объектов *de novo* [7], для развития принципиально новых технологий, использующих принципы организации биосистем.

Развитие методов экспериментального изучения структуры и динамики биомакромолекул и их комплексов наряду с совершенствованием методов математической обработки, нацеленных на структуры с плохой периодичностью и единичные макромолекулы, способны значительно расширить класс изучаемых объектов и приближают постановку и решение новых задач структурной биологии, которые могут оказать весьма существенное влияние на уровень наших знаний о механизмах организации живой материи. Здесь имеется ряд фундаментальных проблем, которые, несмотря на огромные успехи последних десятилетий, остаются мало изученными и, более того, не привлекают пока даже

особого внимания ввиду неясности подходов к их решению. Тем не менее практическое значение решения этих проблем трудно переоценить. В краткой статье нет возможности обсудить достаточно широкий спектр этих проблем, но мы выделим одну из наиболее фундаментальных.

В настоящее время основа наших знаний в области структурной биологии – порядка 130000 пространственных структур белков, собранных в базе данных PDB [8]. На этом базируются основные представления о свойствах белков и механизмах их функционирования. Практически все структуры разрешены в условиях, весьма далеких от условий их реального функционирования в живой клетке. Белковые структуры конформационно достаточно лабильны и меняют свою конформацию в процессе функционирования, что находит некоторое отражение в базе данных PDB, где для небольшого числа объектов удается определить структуры в различных конформациях, отвечающих несколько разным внешним условиям (например, условиям кристаллизации) [9]. Это позволяет надеяться, что таким образом можно приблизительно определить пути конформационных переходов при функционировании белка [10].

Практически единственным методом, который в настоящее время реально позволяет наблюдать с атомной точностью за конформационной динамикой и «жизнью» белковой структуры, является молекулярное моделирование (молекулярная динамика) с использованием суперкомпьютерных технологий [11,12]. Метод основан на решении системы уравнений классической механики для системы материальных точек (которым соответствуют атомы белковой структуры, молекул растворителя и прочих молекулярных объектов). Силы взаимодействия между частицами задаются через систему атом-атомных потенциалов (силовое поле), которое специальным образом калибруется, чтобы определенные рассчитанные характеристики макромолекул соответствовали экспериментальным значениям. Имеется целый ряд (и это важно!) признанных силовых полей, используемых при моделировании биополимеров в различных ситуациях [13–15].

Современный уровень и возможности вычислительной техники в области молекулярного моделирования можно охарактеризовать примерно следующим образом. Частота тепловых флуктуаций отдельной частицы в конденсированной среде составляет порядка $kT/h \sim 10^{13} \text{ с}^{-1}$. Суперкомпьютер с производительностью 1 Пфлоп/с (10^{15} операций в секунду) теоретически способен рассчитывать траекторию примерно 10 атомов

со скоростью их тепловых флуктуаций, т.е. за одну секунду расчета можно в принципе проследить жизнь такой небольшой системы на интервале 1 с. Это огромная скорость. Замедление скорости расчета при увеличении числа атомов N примерно обратно пропорционально величине $N \ln N$ [11,12]. Таким образом, используя суперкомпьютер с подобными возможностями, для системы из миллиона атомов в идеальном варианте за одну секунду можно получить траекторию длиной порядка 100 нс. Для биологии принципиальный рубеж по длине траектории на четыре порядка больше и составляет порядка миллисекунд (на этих временах реализуются многие физиологически важные процессы типа открытия–закрытия каналов, фолдинга и др.). Реально использование всей мощности обычного суперкомпьютера для одного расчета маловероятно, технически трудно реализуемо и скорость набора траектории оказывается на несколько порядков ниже. Ситуация выглядит лучше на специализированных суперкомпьютерах, использующих специальную архитектуру процессоров, заточенных под алгоритмы молекулярной динамики [16,17]. Переход к экзафлопным вычислениям (прогнозируемым в среднесрочной перспективе) способен изменить ситуацию со скоростью моделирования больших биоструктур кардинальным образом. Но здесь возникает одна из наиболее фундаментальных и нерешенных проблем структурной биологии, которая пока «прячется» под ворохом текущих задач и только-только начинает обсуждаться.

Скорость экзафлопных вычислений (более 10^{18} операций/с) открывает возможность отслеживания динамики больших белковых структур в реальном окружении практически в режиме текущего времени. Например, если время сворачивания (фолдинга) полипептидной цепи в нативную пространственную структуру составляет порядка 1 с, длительность расчета траектории такой же длины займет сравнимое время. Это создаст предпосылки для переноса значительной части структурных исследований (определение пространственной структуры и динамики биополимеров в кристаллах, растворах, средах, приближенных к внутриклеточным) в режим *in silico* и высвобождение огромных экспериментальных мощностей структурной биологии для решения других, более сложных задач.

Такое положение дел, согласно фундаментальным законам механики, станет возможным при обязательном условии – корректном описании сил, действующих между атомными частицами. Как известно, информация о действующих силах содержится в функции потенциаль-

ной энергии системы U , которая зависит от координат всех атомов системы. Для реальной белковой структуры с учетом окружения эта функция зависит от миллиона и более переменных. Понятно, что это весьма сложно устроенный и сложный с точки зрения математических процедур объект. Структура таких объектов обычно изучается топологическими методами в рамках теории Морса [18–20]. Многомерная энергетическая поверхность U (или энергетический ландшафт) конформационно подвижной макромолекулярной системы кардинально отличается по структуре от аналогичной поверхности, например, для обычных кристаллов (где U соответствует многомерному параболоиду) и содержит множество относительно близких по энергии локально равновесных состояний, разделенных седловыми (критическими) точками высокой размерности (с большим индексом или большим числом отрицательных элементов диагонализированной матрицы вторых производных от U) [21–23]. Такое устройство энергетического ландшафта имеет определенный смысл для организации функционально значимых процессов переноса массы, энергии и переработки информации [22,23]. С другой стороны, согласно современным представлениям [24–26] строение энергетического ландшафта белковых структур напоминает многомерную воронку, где верхняя часть воронки соответствует случайным развернутым конформациям полипептидной цепи. В результате тепловых движений происходит релаксация макромолекулы, уменьшается потенциальная (и свободная) энергия, репрезентативная точка соскальзывает в выделенный (глобальный) минимум энергии, отвечающий нативной пространственной структуре. В такой ультрамногомерной энергетической воронке имеется экспоненциально большое число локальных энергетических минимумов (энергетических ловушек) и иных «шероховатостей», в которых репрезентативная точка может в принципе «запутаться». Поэтому принципиально важной для решения обсуждаемых задач структурной биологии *in silico* является разработка принципов корректного описания структуры всей энергетической воронки (энергетического ландшафта) и понимание правил движения по таким ультрамногомерным поверхностям [27–31]. Последнее особенно важно для решения кинетических парадоксов фолдинга белковых структур [32].

Имеющиеся в настоящее время технологии молекулярного моделирования обычно предлагают вычислять многомерную энергетическую поверхность как сумму парных атом-атомных потенциалов с параметрами, которые калибру-

ются с учетом атомного состава изучаемых систем и условий моделирования [13–15]. В настоящее время нет универсального силового поля, которое можно было бы использовать в любой ситуации. Более того, параметры силовых полей вынужденно калиброваны таким образом, что описывают свойства макромолекул, конформация которых близка к определяемой структурными методами (в некотором смысле близка к равновесной). Есть основания полагать, что такой подход не всегда продуктивен при изучении неравновесных состояний макромолекул, включая проблемы самоорганизации пространственных структур. Известны и ситуации, когда использование парных взаимодействий дает качественно неверный результат и попытки построить систему многоцентровых потенциалов (см., например, работу [33]). Эти проблемы в принципе хорошо известны, но с ними приходится пока мириться, подправляя время от времени силовые поля при моделировании различных систем и улучшая предсказательную силу молекулярной динамики для рассматриваемых объектов. Как уже отмечалось, переход к моделированию сильно неравновесных состояний больших биомолекулярных систем на длинных временах и предсказание, например, результатов динамики фолдинга полипептидных последовательностей в пространственные белковые структуры предъявляет к силовым полям дополнительные требования, связанные с определением рельефа энергетической воронки не только в окрестности минимума, отвечающего структуре, близкой к равновесной, но и для денатурированных состояний.

Рассмотрим кратко математические проблемы, связанные со структурой обсуждаемой многомерной энергетической поверхности биополимеров. Как известно, изменение взаимного положения атомов биополимера в пространстве происходит в основном за счет конформационных движений или поворотов вокруг одинарных связей. В настоящее время силовое поле традиционно задается как сумма вкладов в потенциальную энергию молекулы от валентных связей и валентных углов U_{val} , торсионных потенциалов U_{tors} , описывающих изменение потенциальной энергии при поворотах вокруг валентных связей k изолированных 1–4 фрагментов молекулы, ван-дер-ваальсовых взаимодействий U_{vdw} валентно не связанных атомов (обычно описываются суммой парных потенциалов Леннарда-Джонса, зависящих от расстояний между атомами i, j) и кулоновских взаимодействий атомов U_{Coul} , несущих парциальные электрические заряды q_i :

$$U = U_{\text{val}} + U_{\text{tors}} + U_{\text{vdW}} + U_{\text{Coul}}$$

$$U_{\text{tors}} = \sum_k u_k(\phi_k); \quad k = 1, 2, \dots, N, \quad (1)$$

$$U_{\text{vdW}} = \sum_{i \neq j} u_{ij}(|\vec{r}_i - \vec{r}_j|), \quad U_{\text{Coul}} = \sum_{i \neq j} q_i q_j \epsilon(|\vec{r}_i - \vec{r}_j|).$$

Последнее слагаемое определяется с достаточно большой погрешностью, так как приближение Кондона [34] не работает при поворотах вокруг валентных связей [35] и флуктуации парциальных зарядов атомов при конформационных изменениях в пептидных структурах могут достигать нескольких десятых долей от заряда электрона [35,36].

Изменение длин химических связей и валентных углов вносит незначительный вклад в структурную динамику этих макромолекул, поэтому вкладом U_{val} при описании конформационной динамики обычно можно пренебречь. В этом случае потенциальная энергия U макромолекулярной системы, имеющей N связей (торсионных углов ϕ_i , область значений торсионных углов от $-\pi$ до π), зависит от углов поворота вокруг валентных связей: $U(\phi_1, \phi_2, \dots, \phi_N)$, т.е. топология конфигурационного пространства, на котором задана функция потенциальной энергии, эквивалентна N -мерному тору. (Мы оставляем пока вне рассмотрения системы, включающие в явном виде растворитель или молекулярное окружение.) Заметим, что силовое поле, заданное формулами (1), в явном виде не отражает топологию конфигурационного пространства макромолекулы. Вместе с тем топология многомерного тора в силу теоремы Морса [18–20] оказывает существенное влияние на свойства функции $U(\phi_1, \phi_2, \dots, \phi_N)$, которые нужно учитывать при описании силовых полей. Поясним свойства такой функции наглядным способом.

Физически разумная функция, заданная на торе (гиперторе), может быть разложена в ряд Фурье:

$$U(\phi_1, \phi_2, \dots, \phi_N) = \sum_{n_1, n_2, \dots, n_N} A(\vec{n}) \exp(i\vec{n}\vec{\phi}), \quad (2)$$

где мы ввели обозначение N -мерных векторов

$$\vec{\phi} = (\phi_1, \phi_2, \dots, \phi_N); \quad \vec{n} = (n_1, n_2, \dots, n_N) \quad (3)$$

и ввели соответствующее скалярное произведение. Величины n_i принимают целочисленные положительные и отрицательные значения и 0. Отметим сразу ситуацию, когда в результате поворотов по углам ϕ координаты атомов могут совпасть и потенциальная энергия будет иметь

сингулярность. Эта проблема разрешается следующим образом. За любое разумное время эксперимента система практически не сможет попасть в области конфигураций с потенциальной энергией выше $30 kT$. Поэтому совершенно неважно, как ведет себя функция U в этих областях, и небольшим «шевелением» функции ее непрерывность может быть восстановлена. Разложение (2) предполагает бесконечное число членов ряда. Однако высокие гармоники не важны для описания структуры энергетического ландшафта. Дело в том, что изменения потенциальной энергии на очень мелких масштабах, сравнимых по амплитуде с нулевыми колебаниями (для углов ϕ это порядка нескольких градусов), не играют роли при описании конформационных движений, амплитуды которых имеют порядок $0,3 \text{ \AA}$ и более [37]. Поэтому значения n_i реально ограничены величинами $K \sim 10^2$ и для разложения (2) можем записать формулу в следующем виде (аналогичный результат можно получить усреднением U по мелкомасштабным осцилляциям):

$$U(\phi_1, \phi_2, \dots, \phi_N) \sim A(0) + \sum_{n_1, n_2, \dots, n_N} 2|A(\vec{n})| \cos[\vec{n}\vec{\phi} + \phi(\vec{n})]; \quad (4)$$

$$n_i = 0, \pm 1, \pm 2, \dots, \pm K; \quad \vec{n} \neq 0;$$

где $\phi(\vec{n})$ – фаза соответствующего коэффициента разложения $A(\vec{n})$. Вынося за знак суммы максимальный модуль коэффициента разложения $A(\vec{n}_0)$, получим

$$U(\phi_1, \phi_2, \dots, \phi_N) \sim A(0) + |A(\vec{n}_0)| \sum_{n_1, n_2, \dots, n_N} \{ \cos[\vec{n}\vec{\phi} + \phi(\vec{n}) + \delta(\vec{n})] + \cos[\vec{n}\vec{\phi} + \phi(\vec{n}) - \delta(\vec{n})] \}; \quad (5)$$

$$\cos \delta(\vec{n}) = |A(\vec{n})|/|A(\vec{n}_0)| \geq 0.$$

Первые члены этого разложения обычно учитываются в силовых полях как торсионные потенциалы [11–15]. Но сейчас нам важна общая математическая структура энергетического ландшафта макромолекулы, которая представляет собой, согласно уравнениям (4), интерференцию определенным образом устроенных осциллирующих функций. Каждое слагаемое можно рассматривать как вклад в $U(\phi_1, \phi_2, \dots, \phi_N)$ от соответствующего узла N -мерной кубической решетки, состоящей из $\sim K^N$ узлов. В некотором смысле это напоминает дифракцию, но только в N -мерном конфигурационном пространстве. В частности, условия формирования четко выраженного глобального минимума на энерге-

тическом ландшафте, который, как считают авторы работ [6,38], принципиально важен для формирования уникальной пространственной структуры белковой глобулы, оказываются с рассматриваемой точки зрения аналогичными условиям формирования главного дифракционного максимума (в данном случае – минимума) путем складывания осцилляций с соответствующими фазами.

Отметим также, что в рассматриваемом контексте существование единственного глобального минимума потенциальной энергии связано с хиральностью структуры биополимеров. При зеркальном отражении молекулы все значения торсионных углов будут заменены на соответствующие отрицательные значения углов. Если отраженная структура является той же макромолекулой, то это означает наличие как минимум двух глобальных минимумов потенциальной энергии. В этом отношении показательным является резко различное динамическое поведение полиглицина и, например, полиаланина [31]. Полиглицин не является хиральной структурой и не склонен к формированию элементов вторичной структуры типа альфа-спиралей. В терминах проводимого обсуждения эта молекулярная структура не имеет единственного глобального минимума потенциальной энергии. Малозначительная, на первый взгляд, замена в глицине водорода на метил делает аланин хиральной структурой, и полиаланин имеет четкую склонность к формированию устойчивых альфа-спиралей [31], что явно указывает на наличие выраженного глобального минимума на его энергетической поверхности.

Рассмотрим несколько общих соотношений для функции U , описывающей энергетический ландшафт макромолекулы, которые могут быть использованы в дальнейшем для калибровки параметров взаимодействия. Интегрирование U по любой из угловых переменных уничтожает в сумме (4) все слагаемые, кроме тех, у которых соответствующая компонента целочисленного вектора $n_i = 0$. Аналогичное уничтожение слагаемых будет происходить при последующих интегрированиях по другим торсионным углам. Таким образом, интегрируя U последовательно N раз получаем:

$$\int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} U(\phi_1, \phi_2, \dots, \phi_N) d\phi_1 \dots d\phi_N = (2\pi)^N A(0). \quad (6)$$

Величина $A(0)$ есть средняя потенциальная энергия по конфигурационному пространству, зависящая от выбора начала отсчета энергии. Выбирая начало отчета энергии в точке конфигурационного пространства $\bar{\phi}^0 = (\phi_1^0, \phi_2^0, \dots,$

$\phi_N^0)$, получим для разложения потенциальной энергии

$$\begin{aligned} U(\phi_1, \phi_2, \dots, \phi_N) &\sim 4|A(\bar{n}_0)| \sum_{n_1, n_2, \dots, n_N; \bar{n} \neq 0} \sin[\bar{n}(\bar{\phi} + \bar{\phi}^0)/2 + \\ &+ \phi(\bar{n})] \cos[\delta(\bar{n})] \sin[\bar{n}(\bar{\phi}^0 - \bar{\phi})/2] = \\ &= 2|A(\bar{n}_0)| \sum_{n_1, n_2, \dots, n_N; \bar{n} \neq 0} \{ \cos[\bar{n}\bar{\phi}^0 + \phi(\bar{n})] - \\ &- \cos[\bar{n}\bar{\phi} + \phi(\bar{n})] \} \cos[\delta(\bar{n})]. \end{aligned} \quad (7)$$

Формула (7) не является модельным потенциалом, а описывает общую структуру энергетического ландшафта макромолекулы, в которой энергии взаимодействия атомов изменяются за счет поворотов вокруг валентных связей (конформационных движений).

Формула (7) представляет также интерес с двух точек зрения. Во-первых, она задает функциональный вид энергетического ландшафта в строгом соответствии с топологией конфигурационного пространства. Во-вторых, делая разложения U в окрестностях выбранных конфигураций ϕ^0 , можно определить параметры потенциала (7) путем сравнения с результатами расчета с известными силовыми полями. Отдельный интерес представляет изучение вопроса об инвариантности получаемых таким образом параметров потенциальной функции (7) при изменении начальной конфигурации системы. Отсутствии инвариантности определяемых параметров (7) при изменении выбора начальной конфигурации будет означать необходимость внесения соответствующих поправок в силовые поля для приведения последних в согласие с общими свойствами энергетического ландшафта, которые диктуются топологией конфигурационного пространства макромолекулы.

Обсудим вопрос с общей структурой энергетического ландшафта для конформационных движений макромолекулы в применении к системам, формирующим уникальную пространственную структуру. Формула (7) задает огромное множество различных энергетических ландшафтов, который определяются разнообразными зависимостями $\phi(\bar{n})$ и $\delta(\bar{n})$. На этом множестве прослеживается тип ландшафтов, которые имеют четко выраженный глобальный минимум. Именно такой тип энергетической поверхности (воронки) считается необходимым условием формирования уникальных пространственных белковых структур. Если глобальный минимум потенциальной энергии реализуется при значениях торсионных углов

$$\bar{\phi}_m = (\phi_{1m}, \phi_{2m}, \dots, \phi_{Nm}), \quad (8)$$

то из структуры формулы (7) и неотрицательности величины $\cos[\delta(n)]$ следует, что зависимость $\phi(n)$ должна определяться простым соотношением

$$\phi(\bar{n}) = \pi - \bar{\phi}_m \bar{n} = \pi - \sum_{i=1}^N n_i \phi_{im} \quad (9)$$

и гиперповерхность потенциальной энергии с точностью до постоянного слагаемого и постоянного множителя должна иметь следующий вид:

$$U(\phi_1, \phi_2, \dots, \phi_N) \sim - \sum_{n_1, n_2, \dots, n_N; \bar{n} \neq 0}^K \cos[\bar{n}(\bar{\phi} - \bar{\phi}_m)] \cos[\delta(\bar{n})]. \quad (10)$$

Простота формулы (10) обманчива. Число слагаемых для реальных макромолекулярных структур огромно (порядка K^N), имеется также множество вариантов, определяемых зависимостями $\cos[\delta(\bar{n})]$. Отметим, что задавая различные зависимости $\delta(\bar{n})$ ($-\pi/2 < \delta(\bar{n}) < \pi/2$), можно строить и исследовать топологически корректные модели энергетической воронки для макромолекул. Формула (10) помимо очевидного глобального минимума описывает огромное число разнообразных локальных минимумов и седловых точек различной размерности (критических точек различного индекса [18–20]).

Если смотреть на энергетический ландшафт, задаваемый формулой (10), с точки зрения конструирования самособирающихся структур, то нужно также обеспечить за разумное время достижение репрезентативной точкой глобального минимума энергии. Ранее численными методами было показано, что диссипативные силы, обусловленные вязкостью окружающей среды, приводят к необходимости соблюдения принципа минимума скорости диссипации энергии при конформационных движениях и формируют коллективные конформационные степени свободы макромолекулы [27–31]. Это означает также, что многие области конфигурационного пространства макромолекулы, которые доступны энергетически, являются практически недоступными с точки зрения правил организации движения репрезентативной точки на многомерных энергетических ландшафтах [27,29]. Поэтому не все участки энергетического ландшафта, задаваемые выражением (10), важны с точки зрения организации фолдинга.

При сворачивании макромолекулярной цепи в вязкой среде (растворителе) действуют силы вязкого трения. В линейном приближении

для сил вязкого трения скорость диссипации энергии S определяется коэффициентами трения γ_i узлов цепи, их линейными скоростями $\dot{\vec{r}}_i$, а также непосредственно связана со скоростью уменьшения потенциальной энергии системы U [27]:

$$S = \sum_i \gamma_i (\dot{\vec{r}}_i)^2 = -\frac{dU}{dt}. \quad (11)$$

Известно, что силы трения для конформационных движений макромолекул в средах с вязкостью порядка вязкости воды заведомо много больше инерциальных сил [37,39] и скорость диссипации энергии S равна скорости убывания потенциальной энергии системы, как это отражено в формуле (11). Дифференцируя выражение (10), получаем

$$\begin{aligned} \frac{dU(\bar{\phi})}{dt} &= 2|A(\bar{n}_0)| \sum_{n_1, n_2, \dots, n_N; \bar{n} \neq 0}^K (\dot{\bar{n}} \bar{\phi}) \sin[\bar{n}(\bar{\phi} - \bar{\phi}_m)] \cos[\delta(\bar{n})] = \\ &= 2|A(\bar{n}_0)| \sum_{k=1}^N \dot{\phi}_k \left\{ \sum_{n_1, n_2, \dots, n_N; \bar{n} \neq 0}^K n_k \sin[\bar{n}(\bar{\phi} - \bar{\phi}_m)] \cos[\delta(\bar{n})] \right\} = \\ &= 2|A(\bar{n}_0)| \sum_{k=1}^N \dot{\phi}_k B_k(\bar{\phi}). \end{aligned} \quad (12)$$

Тонкость восприятия формулы (12) заключается в том, что для рассматриваемого процесса релаксации репрезентативной точки в минимум энергии значение производной U по времени должно быть отрицательным. Это означает, что в среднем положительным значениям $B_k(\bar{\phi})$ должны соответствовать отрицательные значения угловых скоростей и наоборот. Тем самым, уже на этом уровне прослеживается определенная знакпеременность поворотов вокруг валентных связей при конформационной релаксации. Заметим, что структура формулы (12) остается той же самой и для произвольных энергетических ландшафтов, заданных формулой (7). Все остальные выкладки также остаются справедливыми с точностью до вида функций $B_k(\bar{\phi})$.

Рассматривая конформационные движения макромолекулярной цепи с жесткими валентными связями и валентными углами, можно выразить линейные скорости атомов цепи через угловые скорости поворотов вокруг связей [27]:

$$\dot{\vec{r}}_i = \sum_{k=1}^N \vec{C}_{ik}(\bar{\phi}) \dot{\phi}_k \quad (13)$$

или

$$S = \sum_{i,j,k} \gamma_i \bar{C}_{ij}(\bar{\Phi}) \bar{C}_{ik}(\bar{\Phi}) \dot{\Phi}_j \dot{\Phi}_k = \sum_{j=1}^N \sum_{k=1}^N D_{jk}(\bar{\Phi}) \dot{\Phi}_j \dot{\Phi}_k, \quad (14)$$

где вектора $\bar{C}_{ij}(\bar{\Phi})$ определяются номером узла i и конфигурацией цепи, а элементы симметричной матрицы $D_{jk}(\bar{\Phi})$ есть сумма по всем атомам цепи соответствующих скалярных произведений этих векторов, взвешенных на коэффициенты трения:

$$D_{jk} = \sum_i^{allatoms} \gamma_i \bar{C}_{ik}(\bar{\Phi}) \bar{C}_{ij}(\bar{\Phi}). \quad (15)$$

Отметим, что $D_{kk} > 0$.

Суммируя выражения (12) и (14), получаем в общем виде соотношение для угловых скоростей, связывающее кинематику поворотов по двугранным углам цепи макромолекулы и параметры энергетического ландшафта:

$$\sum_{k=1}^N \dot{\Phi}_k \{ [\sum_{j=1}^N D_{jk}(\bar{\Phi}) \dot{\Phi}_j] + 2|A(\bar{n}_0)|B_k(\bar{\Phi}) \} = 0. \quad (16)$$

В уравнении (16) коэффициенты перед угловыми скоростями поворота вокруг связи k зависят достаточно произвольным образом от конфигурации цепи, и выполнить (16) возможно, если эти коэффициенты обращаются в 0. Тогда мы имеем систему из N дифференциальных уравнений первого порядка для траекторий движения репрезентативной точки по энергетическому ландшафту, заданному на гиперторе конфигурационного пространства макромолекулы:

$$\sum_{j=1}^N D_{jk}(\bar{\Phi}) \dot{\Phi}_j = -2|A(\bar{n}_0)|B_k(\bar{\Phi}). \quad (17)$$

Умножая каждое уравнение (17) на соответствующую угловую скорость и суммируя по всем k , получаем соответствие этой системы уравнений формулам (11), (12) и (14).

Отметим полезное соотношение для скорости диссипации энергии, которое следует из этих формул. Дифференцируя уравнение для S (14) по угловой скорости, получаем

$$\frac{\partial S}{\partial \dot{\Phi}_k} = 2 \sum_j D_{jk}(\bar{\Phi}) \dot{\Phi}_j. \quad (18)$$

Сравнивая формулы (18) и (17), получаем

$$\frac{1}{2} \frac{\partial S}{\partial \dot{\Phi}_k} = -2|A(\bar{n}_0)|B_k(\bar{\Phi}). \quad (19)$$

Формула (19) определяет изменение вклада угла k в скорость диссипации энергии при изменении скорости поворота по торсионному углу k в конформации $\bar{\Phi}$ в режиме релаксации («сползании» по стенкам энергетической воронки при сильном трении).

Ранее было показано [27,28], что в вязкой среде (растворителе) конформационные движения макромолекулярной цепи происходят таким образом, чтобы скорость диссипации энергии S (при заданной скорости уменьшения потенциальной энергии системы) была минимальной. Это ведет к разнообразным интересным следствиям, например, формированию коллективных конформационных степеней свободы за счет действия сил трения, запрету на резкие изменения потенциальной энергии за счет изменения относительно небольшого числа координат репрезентативной точки [27], практическому равномерному распределению скорости диссипации по конформационным степеням свободы [29,30]. Заметим, что требование минимальности скорости диссипации энергии при конформационных движениях в окрестности данной конфигурации (конформации) в данном случае не может быть формально определено приравниванием к 0 частных производных S по угловым скоростям (например, в формуле (19)). Мы должны рассмотреть экстремум функции S при условии выполнения соотношения (20) для заданной скорости уменьшения потенциальной энергии, которое определяется формулой (12)

$$2|A(\bar{n}_0)| \sum_{k=1}^N \dot{\Phi}_k B_k(\bar{\Phi}) = \dot{U}. \quad (20)$$

Пользуясь методом неопределенных множителей Лагранжа λ будем искать условный экстремум S при условии заданной скорости релаксации потенциальной энергии системы. Для этого рассмотрим экстремум функции G :

$$G = S - \lambda 2|A(\bar{n}_0)| \sum_{k=1}^N \dot{\Phi}_k B_k(\bar{\Phi}); \quad \frac{\partial G}{\partial \dot{\Phi}_k} = 0, \quad (21)$$

$$\frac{\partial S}{\partial \dot{\Phi}_k} - \lambda \frac{\partial}{\partial \dot{\Phi}_k} [2|A(\bar{n}_0)| \sum_{k=1}^N \dot{\Phi}_k B_k(\bar{\Phi})] = 0. \quad (22)$$

Получаем, что при $\lambda = -2$ функция G имеет минимум. То есть скорость диссипации минимальна при движении цепи, если задана ско-

рость релаксации потенциальной энергии. Это означает, что решение уравнений движения в пространстве торсионных углов (17) (из которых и следует соотношение (19)) определяет траектории с минимальной скоростью диссипации энергии). Отклонение решения от уравнений (17) может быть в случае, если некоторые угловые скорости в формуле (17) равны 0. Тогда соответствующие им уравнения (17) могут и не выполняться. Но это некий исключительный случай «заморожки» поворотов по некоторым двугранным углам. В такой ситуации просто нужно рассматривать конфигурационное пространство меньшей размерности и все выкладки остаются без изменений.

Полученный результат имеет очень простую математическую основу. Скорость диссипации энергии S является положительно определенной квадратичной формой скоростей, а производная по времени от потенциальной энергии U линейна по угловым скоростям. В релаксационном процессе эти величины должны быть согласованы (их сумма равна 0). Следствием этого согласования и являются полученные выше уравнения движения (17) и минимум скорости диссипации энергии на релаксационных траекториях (при заданной скорости сворачивания цепи). Линейность dU/dt по угловым скоростям абсолютно тривиально выглядит при записи потенциальной энергии в соответствии с топологией конфигурационного пространства макромолекулы (выражение (12)). Скорость диссипации энергии также вполне четко выражается в тех же переменных (выражение (14)), и анализ может быть доведен до конечного результата. Впрочем, в пространстве радиусов-векторов частиц эти выкладки лишь немногим сложнее.

Таким образом, полученное разложение энергетического ландшафта макромолекулы с конформационными степенями свободы с учетом топологии многомерного тора для конфигурационного пространства позволяет сделать следующие выводы. Хотя данное разложение содержит экспоненциально много членов, оно допускает в ряде случаев наглядный анализ. В частности, для систем с выраженным глобальным минимумом энергии (к которым, видимо, относятся многие белки и другие биополимеры с четко определенной пространственной структурой) необходимо согласование фаз вкладов, которое имеет достаточно простой вид скалярного произведения многомерного вектора двугранных углов на соответствующий вектор узла N -мерной кубической решетки с целочисленными компонентами. Заметим, что здесь имеется как бы элемент скрытой симметрии энергетических ландшафтов для макромолекул. Наличие

же единственного глобального минимума связано с особенностями химической структуры и хиральностью рассматриваемых макромолекул. Анализ и согласование параметров полученного разложения, рассчитываемых из имеющихся силовых полей, может быть полезен для их дополнительной калибровки. При известных параметрах топологически корректного разложения потенциальной энергии уравнения движения репрезентативной точки, которые имеют матричный вид, вполне пригодны для численных расчетов. Уравнения движения в пространстве двугранных углов позволяют наглядно получить общий результат для релаксационных траекторий – выполнение принципа минимума скорости диссипации энергии при конформационной релаксации [27] при заданном значении скорости релаксации потенциальной энергии.

Автор благодарен А.О. Иванову, Г.М. Кобелькову, Ф.Ю. Попеленскому, А.А. Тужилину и другим участникам постоянного совместного семинара по структурной биологии биологического и механико-математического факультетов МГУ имени М.В. Ломоносова за полезные обсуждения. Автор благодарен А.В. Финкельштейну за стимулирующие обсуждения, а также М.П. Кирпичникову и А.Т. Фоменко за внимание к работе и поддержку.

Работа выполнена при финансовой поддержке Российского научного фонда (гранты № 14-24-00031 (в части изучения топологической структуры энергетического ландшафта в пространстве торсионных углов для систем с выраженным глобальным минимумом) и № 14-50-00029 (в части вывода уравнений для динамики релаксации репрезентативной точки на энергетических ландшафтах в пространстве торсионных углов и принципа минимума скорости диссипации энергии для релаксационных траекторий)).

СПИСОК ЛИТЕРАТУРЫ

1. И. Н. Сердюк, Успехи биол. химии **42**, 3 (2002)
2. V. Marx, Nature Methods **11**, 903 (2014).
3. К. В. Шайтан, М. П. Кирпичников, В. С. Ламзин и др., Вестн. РФФИ, № 4 (80), 38 (2013).
4. Д. О. Синицын, Н. Г. Лойко, С. К. Гуларян и др., Хим. физика **36**, 350 (2017).
5. Д. В. Багров, Н. Воскобойникова, Г. А. Армеев и др., Биофизика **61**, 1139 (2016).
6. А. В. Финкельштейн и О. Б. Птицын, *Физика белка* (М., 2002).
7. Д. А. Долгих, М. П. Кирпичников, О. В. Птицын и В. В. Чемерис, Молекуляр. биология **30**, 261 (1996).
8. <http://www.wwpdb.org/>.

9. Г. В. Новиков, В. С. Сивожелезов и К. В. Шайтан, *Биофизика* **58**, 618 (2013).
10. G. V. Novikov, V. S. Sivozhelezov, S. S. Kolesnikov, and K. V. Shaitan, *J. Receptors and Signal Transduction* **34**, 104 (2014).
11. D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, New York, 2002).
12. Н. К. Балабаев и К. В. Шайтан, в кн. *Методы компьютерного моделирования для исследования полимеров и биополимеров* (ЛКИ, Москва, 2009), глава 2: Компьютерное моделирование молекулярной динамики, сс. 35–62.
13. R. Salomon-Ferrer, D. A. Case, and R. C. Walker, *WIREs Comput. Mol. Sci.* **3**, 198 (2013).
14. G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, *J. Phys. Chem. B* **105**, 6474 (2001).
15. <https://www.charmm.org/charmm>.
16. D. E. Shaw, M. M. Deneroff, R. O. Dror, et al., *Communications of the ACM* **51**, 97 (2008)
17. A. C. Pan, T. M. Weinreich, S. Piana, and D. E. Shaw, *J. Chem. Theory and Computation* **12**, 1360 (2016).
18. Дж. Милнор, *Теория Морса* (Мир, М., 1965).
19. А. Т. Фоменко и Д. Б. Фукс, *Курс гомотопической топологии* (Наука, М., 1989).
20. И. С. Шапиро и М. А. Ольшевецкий, *Лекции по топологии для физиков* (Ижевск, 2001).
21. K. V. Shaitan, *Macromolec. Symp.* **106**, 321 (1996).
22. К. В. Шайтан, *Биофизика* **39**, 949 (1994).
23. K. V. Shaitan, in *Stochastic Dynamics of Reacting Biomolecules*, Ed. by W. Ebeling, L. Schimansky-Gefer, and Y. M. Romanovsky (World Scientific, Singapore, 2003), pp. 283–308.
24. P. G. Wolynes, *Phil. Trans. Roy. Soc.* **363**, 453 (2005).
25. D. J. Wales, *Phil. Trans. R. Soc. A* **370**, 2877 (2012)
26. E. R. Henry, R. B. Best, and W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **110**, 17880 (2013).
27. К. В. Шайтан, *Биофизика* **60**, 843 (2015).
28. К. В. Шайтан, М. А. Ложников и Г. М. Кобельков, *Биофизика* **61**, 629 (2016).
29. К. В. Шайтан, *Биофизика* **62** (1), 5 (2017).
30. К. В. Шайтан, М. А. Ложников и Г. М. Кобельков, *Биофизика* **62**, 249 (2017).
31. К. В. Шайтан, Ф. Ю. Попеленский и Г. А. Армеев, *Биофизика* **62**, 443 (2017).
32. C. Levinthal, *J. Chem. Phys.* **65**, 44 (1968).
33. Bo Wang, Ke R. Yang, Xuefei Xu, et al., *Acc. Chem. Res.* **47**, 2731 (2014).
34. M. Lax, *J. Chem. Phys.* **20**, 1752 (1952).
35. К. В. Шайтан, А. В. Немухин, Д. А. Фирсов и др., *Молекуляр. биология* **31**, 108 (1997).
36. K. V. Wiberg, C. M. Hadad, C. M. Breneman, et al., *Science* **252**, 1266 (1991).
37. К. В. Шайтан и А. Б. Рубин, *Молекуляр. биология* **14**, 1323 (1980).
38. E. Shakhnovich, *Chem. Rev.* **106**, 1559 (2006).
39. К. В. Шайтан и С. С. Сарайкин, *Биофизика* **45**, 407 (2000).

Topology of Energy Landscapes of Macromolecules in the Torsion Angle Space and the Principle of Minimum Energy Dissipation Rate during Conformational Relaxation

K.V. Shaitan* **

*Faculty of Biology, Lomonosov Moscow State University, Leninskie Gory 1/12, Moscow, 119234 Russia

**Semenov Institute of Chemical Physics, Russian Academy of Sciences, ul. Kosygina 4, Moscow 119991 Russia

We discuss the prospects of applying molecular dynamics simulations methods to complex problems of structural biology, modeling of protein folding, and prediction of the spatial structure of biopolymers. For the energy landscape of systems with conformational mobility we formulated a multidimensional Fourier expansion, which is correct from the point of view of the topology of the configuration space. For systems with a distinguished global minimum on the energy landscape, we introduced the phase matching rule for the terms of the multidimensional expansion, which is formally similar to the diffraction problem on a multidimensional cubic lattice. The problem of the calibration of biopolymer force fields and their correspondence to topologically correct energy landscapes is discussed. Equations of motion in matrix form for the relaxation of representative point position on multidimensional potential energy surface were obtained. We showed that the solutions of the equations for conformational relaxation satisfy the principle of the minimum energy dissipation rate at a given relaxation rate of the potential energy (or folding rate).

Keywords: structural biology, molecular dynamics, supercomputer technologies, energy landscapes, topology of the energy surface, conformational mobility, functional dynamics of biopolymers, folding, the principle of minimum energy dissipation rate