

ПРЕДСКАЗАНИЕ КОНФОРМАЦИОННОЙ ПОДВИЖНОСТИ БЕЛКОВ И ОЦЕНКА ЕГО ДОСТОВЕРНОСТИ С ПОМОЩЬЮ МАЛОУГЛОВОГО РЕНТГЕНОВСКОГО РАССЕЯНИЯ

© 2015 г. С.Н. Князев*, В.Ю. Калякин*, И.Н. Дерябин*, Б.А. Федоров*, А.В. Смирнов*, Е.О. Степанов* ** ***, Ю.Б. Порозов*

*Университет ИТМО, 197101, Санкт-Петербург, пр. Кронверкский, 49;

**Санкт-Петербургское отделение Математического института им. В. А. Стеклова РАН, 191023, Санкт-Петербург, наб. Фонтанки, 27;

***Санкт-Петербургский государственный университет, 198504, Петергоф, Санкт-Петербург, Университетский пр., 28

E-mail: porozov@ifc.cnr.it

Поступила в редакцию 17.08.15 г.

Представлена «крупнозернистая» модель конформационной подвижности белков. Проведено сравнение предсказанной с помощью этой модели траектории конформационного движения для пяти белков с движением, полученным на основе метода «ближайшего соседа», опирающегося на степень различия кривых рентгеновского малоуглового рассеяния. Показано, что для всех рассмотренных белков последовательность конформаций, рассчитанных как на основе «крупнозернистой» модели, так и на основе метода «ближайшего соседа», хорошо совпадает, хотя имеются исключения, причина которых для каждого белка требует отдельного рассмотрения.

Ключевые слова: конформационная подвижность белка, крупнозернистые модели, рентгеновское малоугловое рассеяние, транспортная задача, траектория конформационного движения, предсказание движения.

В настоящее время общепризнанным является факт зависимости функции белка от его структуры. Это, а также развитие компьютерной техники и экспериментальных методов исследования, в первую очередь ЯМР и рентгеноструктурного анализа, в последние десятилетия позволило изменить взгляд на белки как на «жесткие», неподвижные молекулы с фиксированной структурой [1]. И, как следствие, появились многочисленные методы расчета молекулярной динамики белков [2–5]. Дальнейшее развитие молекулярной биологии и, в частности, открытие белков, не обладающих постоянной, фиксированной структурой [6–9], привело к изменению парадигмы «структура–функция». В работах [10,11] подчеркивается, что ключевым фактором, своеобразным «мостом» от структуры белка к его функциям является молекулярная подвижность.

Второй важнейший вопрос в молекулярной биологии белков и изучении клеточных метаболических каскадов – это определение очеред-

ности занятия белком своих конформаций, выстраивание их порядка и восстановление траекторий конформационной подвижности, связанных с осуществлением белком его функций. В настоящее время крайне мало данных, посвященных этому вопросу, и в подавляющем большинстве случаев эта проблема рассматривается в приложении к фолдингу – процессу сворачивания белка, принятия им своей нативной конформации [12]. При помощи метода ЯМР удается получить несколько (до тысяч) конформаций одного белка в растворе, но в базу данных PDB попадают лишь несколько десятков конформаций, причем записанных без всякого порядка.

Вместе с тем подвижность макромолекул – это процесс, который можно рассматривать в различных и довольно сильно отстоящих друг от друга временных масштабах [13]. Так, колебания атомов, представляющих боковые цепи аминокислот, происходят с периодами в диапазоне 10^{-15} – 10^{-12} с и не превышают по амплитуде 0,2 Å. Между тем характерное время конформационной функциональной подвижности

С.Н. Князев и В.Ю. Калякин внесли равный вклад в работу.

и движения доменов белков лежит в миллисекундном диапазоне.

Для моделирования динамики белков методы, базирующиеся на уравнениях молекулярной динамики и квантовой механики, в большинстве случаев применяются только для предсказания движений на коротких (пико-, наносекунды) и средних (микросекунды) временных промежутках. Сложность использования этих методов для моделирования на длинных (миллисекунды) временных интервалах связана с невозможностью точно учесть действие сил, определяющих динамику молекулярной системы на этих временных интервалах [14,15]. Методы моделирования движений, основанные исключительно на внутренней энергии молекулы (возможно, с заданным полем внешних сил), могут неправильно предсказывать межконформационное движение на таких длительных интервалах времени. Более адекватными для этих целей могут быть грубые и кажущиеся нефизичными модели межконформационных движений, называемые крупнозернистыми моделями. Как правило, такие модели белка представляют молекулу в виде набора шаров, заменяющих отдельные атомы аминокислот или аминокислоты целиком, а в расчетах используют «заменители» реальных энергий.

В этой работе мы проведем сравнение конформационных движений, предсказанных при помощи крупнозернистой модели подвижности белка [16,17], с данными, полученными на основе сравнения кривых рентгеновского малоуглового рассеяния, на примере пяти белков.

Способ упорядочивания промежуточных конформаций путем попарного сравнения кривых малоуглового рентгеновского рассеяния основан на следующем предположении: при обычном наблюдаемом в природе конформационном движении белка наиболее вероятной является такая последовательность промежуточных состояний, при которой распределение электронной плотности меняется плавно. Интенсивность рассеяния $I(\mathbf{q})$ для вектора рассеяния \mathbf{q} пропорциональна квадрату модуля соответствующей Фурье-гармоники пространственного распределения электронной плотности $\rho(\mathbf{r})$ в рассеивающем объекте [18]. Кривая рассеяния $I(q)$ является результатом усреднения интенсивности $I(\mathbf{q})$ по всем возможным направлениям вектора \mathbf{q} при фиксированном значении модуля q этого вектора. Таким образом, кривая $I(q)$ однозначно определяется пространственным распределением электронной плотности $\rho(\mathbf{r})$. Причем вид кривой зависит именно от относительного распределения электронной плотности в разных

частях молекулы и не чувствителен к переносам и поворотам молекулы как целого относительно системы координат. Если использовать подходящую норму (например, L_2) для обеих функций, переход от $\rho(\mathbf{r})$ к $I(q)$ является непрерывным преобразованием. Следовательно, если верно предположение о плавном изменении распределения электронной плотности при конформационном движении, то кривая рассеяния при таком движении также будет изменяться плавно.

Заметим, что при исследовании рентгеновского малоуглового рассеяния каждый рассеивающий атом можно считать точечным рассеивающим центром с амплитудой рассеяния, пропорциональной зарядовому числу атома. Зарядовое число пропорционально массе атома для всех атомов белка за исключением водорода. Если не учитывать незначительные отклонения, обусловленные присутствием атомов водорода, то пространственное распределение масс практически совпадает с пространственным распределением заряда. Следовательно, можно ожидать, что последовательность конформаций, реализующая наиболее плавное преобразование для пространственного распределения масс, должна хорошо коррелировать с последовательностью, дающей наиболее плавное изменение кривых рассеяния.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ МОРФИНГА СТРУКТУРЫ БЕЛКА

В настоящей работе мы используем одну из крупнозернистых моделей динамики белка, предложенную в работе [16] и развитую в работе [19]. В основе этой модели – представление белковой молекулы как механической системы (шарнирного механизма), в процессе движения сохраняющей основные известные в физике белка ограничения на возможное изменение пространственной структуры молекулы: неизменность или малый диапазон изменения длин ковалентных связей и планарных углов в основной цепи, ограничения на возможные изменения некоторых торсионных углов, отсутствие самопересечений при движении. Движение находится путем минимизации некоторого интегрального функционала типа функционала механического действия. Лагранжиан подобран таким образом, чтобы с точки зрения минимизации было невыгодно перемещать тяжелые статические структуры молекулы и выгодно перемещать легкие структуры. При этом, как правило, не изменяются наблюдаемые статические части молекул, такие как α -спирали и β -структуры. Расчеты, проведенные для различных белков,

показали, что использование похожих моделей [20] дает вполне физически приемлемые движения для межконформационных переходов. Для белка сукцинат-полуальдегид дегидрогеназы (human succinic semialdehyde dehydrogenase, SSADH) нами было показано [21], что движения, получаемые с помощью разных моделей морфинга из предложенного набора, весьма похожи и близки к движениям, получаемым при помощи другого метода морфинга MinAction-Path, предложенного в работе [20]. Последний метод основан на качественно другой интерполяционной модели, в которой также не учитывается воздействие внешних сил.

Используемая в настоящей работе модель морфинга показала себя вполне физически адекватной и достаточно простой для численной реализации, что важно для предсказания возможных перестроек в больших многодоменных белках. Белковая молекула представляется в виде основной цепи из n узлов. Текущее положение цепи задается вектором из $3n$ пространственных координат узлов, масса каждого из которых m_j ($j = 1, 2, \dots, n$) равна сумме масс атома в этом узле и атомов сопряженных боковых цепей. Искомое движение представляется как абсолютно непрерывная функция $\gamma: [0, 1] \rightarrow \mathbb{R}^{3n}$, параметризованная на отрезке $[0, 1]$, и такая, что:

1) $\gamma(0)$ – заданная начальная конформация x_0 , а $\gamma(1)$ отличается от заданной конечной конформации x_1 на некоторое заранее неизвестное твердотельное движение, определяемое оператором T (сдвиг и/или поворот всей молекулы целиком), т.е. $\gamma(1) = Tx_1$;

2) в процессе движения соблюдаются все физические ограничения геометрического характера на длины ковалентных связей, планарные углы и некоторые из торсионных узлов.

Заданные начальная и конечная конформации x_0 и x_1 , как и все «промежуточные» конформации $\gamma(t)$, являются просто векторами из $3n$ пространственных координат узлов. Возможный межконформационный переход γ и твердотельное движение T ищутся как минимум функционала

$$G_2(\gamma, T) = \int_0^1 \sum_{j=1}^n m_j |\dot{\gamma}_j|^2(t) dt, \quad (1)$$

где γ и T удовлетворяют требованиям (1) и (2), а $\gamma_j: [0, 1] \rightarrow \mathbb{R}^3$ – кривая, описываемая j -м узлом основной цепи в движении, задаваемым функцией γ . Таким образом, если заданные начальная и конечная конформации x_0 и x_1 от-

личаются только на сдвиг и поворот всей молекулы целиком, то движение γ производит просто сдвиг и разворот молекулы без изменения торсионных углов, т.е. без изменения относительного положения отдельных частей молекулы.

Минимизация специальным образом дискретизированного функционала G_2 производится при помощи численного метода, предложенного в работе [16]. Получаемое дискретизированное движение хорошо аппроксимирует функцию γ , являющуюся вместе с некоторым твердотельным движением T локальным минимумом исходного функционала G_2 . Затем достраиваются боковые цепи, ликвидируются возможные стерические конфликты (самопересечения отдельных частей молекулы), если они возникают, и производится дополнительная точная «подстройка» положений атомов боковых цепей при помощи минимизации по алгоритму GRO-MACS.

Рассмотренный выше метод дает возможность построения траекторий крупномасштабных конформационных движений белков. Иными словами, на каждом этапе конформационных изменений можно определять структурное состояние белка вплоть до описания координат всех его атомов. Это обстоятельство позволяет обратиться к широко известному методу малоуглового рентгеновского рассеяния, который дает возможность рассчитывать интенсивности рассеяния различных структур с известными координатами атомов. Как следует из дальнейшего анализа, сравнение между собою кривых рассеяния для всех построенных конформационных состояний данного белка дает возможность не только с высокой точностью подтвердить предсказанную траекторию конформационного перехода, но и определять последовательность конформационных изменений, если эта последовательность заранее не была известна.

РАСЧЕТ ИНТЕНСИВНОСТИ МАЛОУГЛОВОГО РЕНТГЕНОВСКОГО РАССЕЙЯНИЯ ДЛЯ ЧАСТИЦ С ИЗВЕСТНЫМИ КООРДИНАТАМИ АТОМОВ

Для расчета интенсивности $I(q)$ малоуглового рентгеновского рассеяния частицей с известными координатами составляющих ее атомов и свободно ориентированной относительно первичного рентгеновского пучка можно воспользоваться двумя способами.

Первый из них состоит в применении широко известной формулы Дебая:

$$I(q) = \sum_{i,j}^N f_i f_j \frac{\sin(qr_{ij})}{qr_{ij}}, \quad (2)$$

в которой $q = |\mathbf{q}| = \frac{4\pi}{\lambda} \sin(\vartheta/2)$ – модуль вектора

рассеяния в обратном пространстве (ϑ – угол рассеяния, λ – длина волны рентгеновских лучей), f_i – атомный фактор i -го атома (при малоугловом рассеянии он равен количеству электронов в атоме), r_{ij} – расстояние между i -м и j -м атомами, N – общее число атомов в молекуле.

Несмотря на достаточную простоту, формула Дебая обладает существенным недостатком: если N велико, то ее применение требует значительного времени расчета, поскольку количество членов в сумме (2) $\sim O(N^2)$ и, кроме того, необходимо $\frac{N(N-1)}{2}$ раз вычислять расстояние r_{ij} .

Другой способ расчета $I(q)$ носит принципиально иной характер. Предположим, что в начале частица имеет фиксированную ориентацию относительно первичного рентгеновского пучка. Тогда ее амплитуда рассеяния в произвольной точке \mathbf{q} обратного пространства будет

$$F(\mathbf{q}) = \sum_{j=1}^N f_j e^{i(\mathbf{q}\mathbf{r}_j)}, \quad (3)$$

(\mathbf{r} – радиус-вектор j -го атома), а интенсивность в той же точке \mathbf{q}

$$I(\mathbf{q}) = |F(\mathbf{q})|^2, \quad (4)$$

причем ее расчет требует совершения $2N$ операций.

Следующий этап работы – переход к интенсивности рассеяния $I(q)$ путем усреднения $I(\mathbf{q})$ по всем точкам обратного пространства, расположенным на расстоянии q от ее начала. С этой целью вся поверхность сферы радиуса q разбивается на малые участки с одинаковой площадью, для каждого участка рассчитывается интенсивность $I(\mathbf{q})$, которая затем усредняется по всем участкам разбиения. Так как элемент поверхности сферы $q^2 \sin\theta d\theta d\varphi$ (φ , θ – углы азимутальный и полярный соответственно), то при равном выборе углов $\Delta\varphi$ равномерность разбиения сферы на равные площади сводится к утверждению $\sin\theta \cdot \Delta\theta = \text{const}$. Легко показать, что при этом проекция ΔZ дуги $q \cdot \Delta\theta$ на ось Z есть величина постоянная. Таким образом, для нахождения полярного угла искомого малого участка поверхности необходимо разделить ось

Z внутри сферы на n равных участков, найти расстояние Z_i -го участка от центра сферы ($Z_i = \Delta Z \cdot i$, $1 < i < n$) и рассчитать $\theta_i = \arccos(Z_i/q)$. Так как $\Delta\varphi = 2\pi/m$ (m – число разбиений в азимутальной плоскости), то $\varphi_j = \Delta\varphi \cdot j$ ($1 < j < m$) и

$$q_x = q \sin\theta_i \cos\varphi_j, \quad q_y = q \sin\theta_i \sin\varphi_j, \quad q_z = q \cos\theta_i. \quad (5)$$

Рассчитав интенсивность рассеяния $I(\mathbf{q})$ в каждой выбранной точке поверхности сферы и усреднив ее по всем точкам i и j , можно определить искомую интенсивность рассеяния

$$I(q) = \frac{1}{mn} \sum_{i=1}^n \sum_{j=1}^m I(q_x, q_y, q_z). \quad (6)$$

Как показывает опыт, при расчете интенсивности рассеяния даже сравнительно мелкими частицами (например, молекулами белков с 150–200 аминокислотами) второй метод может обеспечить достаточно высокую точность при существенном выигрыше во времени.

УПОРЯДОЧИВАНИЕ ПОСЛЕДОВАТЕЛЬНОСТИ КОНФОРМАЦИОННЫХ ПЕРЕХОДОВ В БЕЛКАХ ПО ДАННЫМ РЕНТГЕНОВСКОГО МАЛОУГЛОВОГО РАССЕЯНИЯ. МЕТОД «БЛИЖАЙШЕГО СОСЕДА»

Рассмотрим какой-либо белок, способный изменять свою конформационную структуру, при этом будем считать, что координаты атомов в каждой конформации известны, но неизвестна последовательность перехода от одной конформации к другой. Ниже представлен метод, основанный на сравнительном анализе кривых малоуглового рассеяния, который может решить задачу определения последовательности этих переходов.

Пусть $I(q, K_\alpha)$ – интенсивность рассеяния одной из выбранных конформаций, рассчитанная в диапазоне от q_{\min} до q_{\max} для равноотстоящих значений модуля вектора рассеяния q_i ($i = 1, 2, \dots, N$). Введем норму интенсивности рассеяния как

$$\|I(q, K_\alpha)\| = \sqrt{\sum_{i=1}^N (I(q_i, K_\alpha))^2}. \quad (7)$$

В качестве меры различия двух конформаций белковой молекулы K_α и K_β используем относительное различие $\Delta(K_\alpha, K_\beta)$ между соответствующими этим конформациям интенсивностями $I(q, K_\alpha)$ и $I(q, K_\beta)$ малоуглового рентге-

Белковые структуры, используемые для сравнения алгоритмов

Белок	Количество аминокислот	Переход (PDB id)
Сукцинат-полуальдегид дегидрогеназа	487	2W8O:1 → 2W8N:1
Кальмодулин	148	1CFC:7 → 1CFC:15
Циановирин	101	1J4V:A:1 → 1L5B:A:1
Ацетил-СоА-трансфераза	392	4N44:A:1 → 4N45:A:1
Мальтозосвязывающий периплазматический белок	370	1LZO:1 → 1LZO:2

новского рассеяния. Это относительное различие, используя норму (7), определим следующим образом:

$$\Delta(K_\alpha, K_\beta) = \frac{\|I(q, K_\alpha) - I(q, K_\beta)\|}{\sqrt{\|I(q, K_\alpha)\| \cdot \|I(q, K_\beta)\|}} \quad (8)$$

Мера различия конформаций (8) позволяет упорядочить произвольный набор $[K_{\alpha=1}^{\alpha=M}]$ конформаций белковой молекулы. Предлагаемый алгоритм такого упорядочения состоит из двух этапов. На первом этапе, используя каждую из конформаций в качестве начальной, все остальные конформации выстраиваются в последовательность $K_{\alpha_1}, K_{\alpha_2}, \dots, K_{\alpha_M}$ по методу «ближайшего соседа». Метод состоит в том, что для выбранной начальной конформации K_{α_1} находится «ближайшая» к ней конформация K_{α_2} , для которой различие $\Delta(K_{\alpha_2}, K_{\alpha_1})$ минимальное, затем конформация K_{α_3} , для которой минимальным будет различие $\Delta(K_{\alpha_3}, K_{\alpha_2})$ и т.д. Для каждой такой последовательности, начинающейся с данной конформации K_{α_1} , вычисляется сумма различий между членами последовательности:

$$\Delta_{\alpha_1} = \sum_{k=1}^{M-1} \Delta(K_{\alpha_{k+1}}, K_{\alpha_k}), \quad \alpha_1 = 1, 2, \dots, M. \quad (9)$$

На втором этапе из полученных M последовательностей выбирается та, для которой сумма (9) является минимальной. Первый этап упорядочивания конформаций фактически реализует жадный алгоритм, впервые примененный к кривым рентгеновского малоуглового рассеяния белковыми молекулами.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В данной работе рассмотрены пять белков: сукцинат-полуальдегид дегидрогеназа, кальмодулин, циановирин, ацетил-СоА-трансфераза, мальтозосвязывающий периплазматический белок. Данные о начальных и конечных конформациях для этих белков, взятые из Банка белковых данных [22], представлены в таблице.

На основе описанного выше алгоритма морфинга и оптимизации крупномасштабных конформационных движений для каждого белка были построены траектории конформационных переходов между начальной и конечной конформациями. Для всех полученных конформаций каждого исследуемого белка были рассчитаны кривые малоуглового рентгеновского рас-

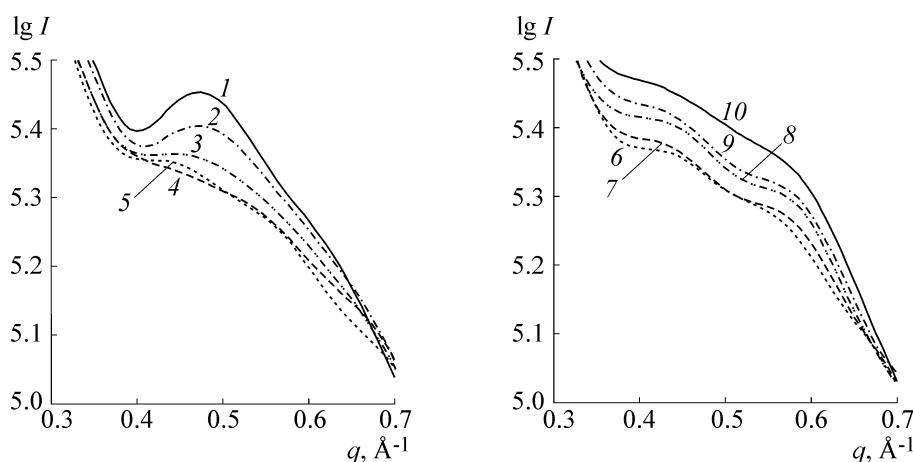


Рис. 1. Кривые малоуглового рентгеновского рассеяния белка кальмодулина для 10 конформаций, последовательность которых найдена на основе развитого в работе алгоритма морфинга.

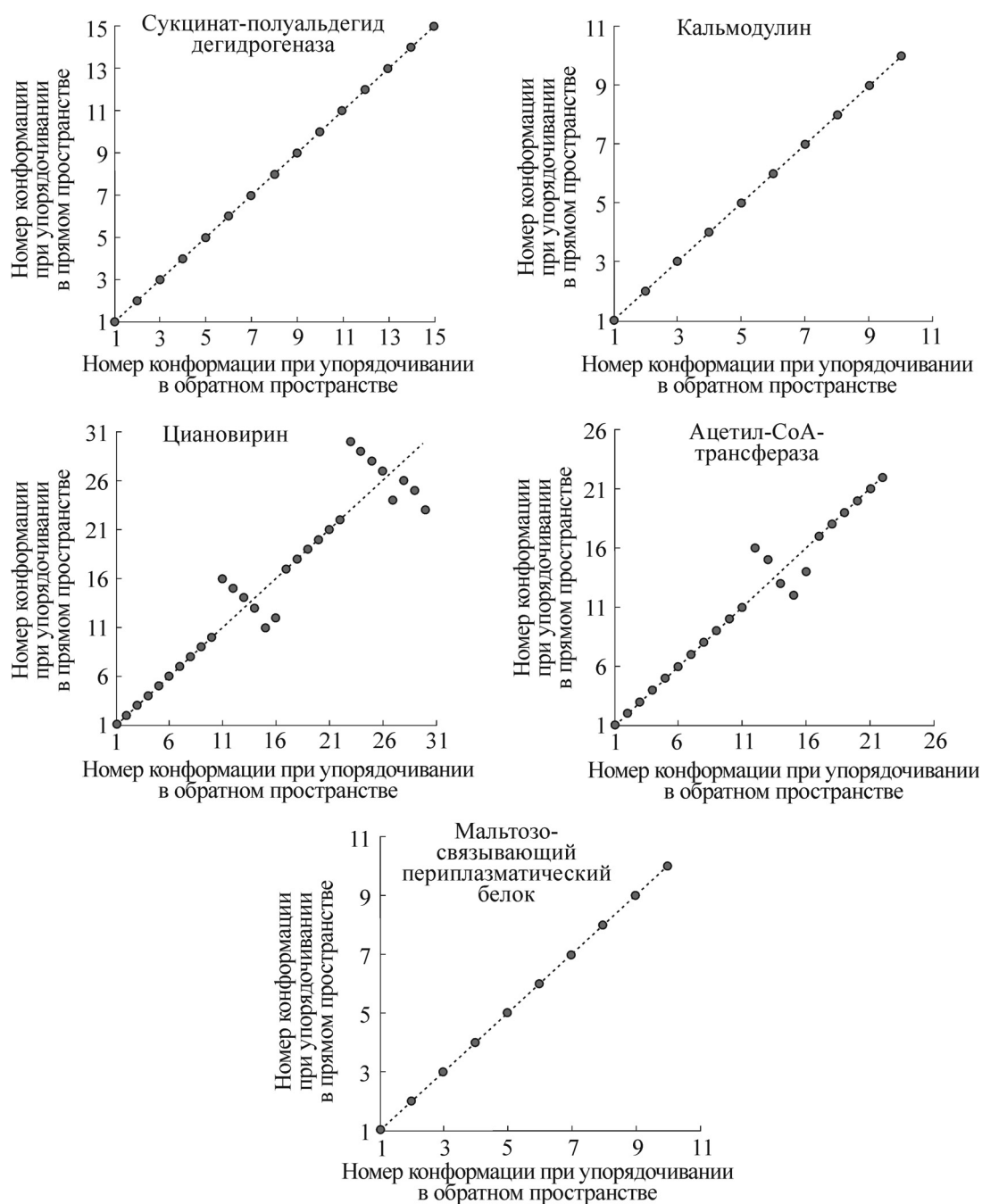


Рис. 2. Сравнение конформационных последовательностей, упорядоченных в обратном и прямом пространствах, для пяти белков. Названия белков указаны на диаграммах. По оси абсцисс отложены порядковые номера в последовательности, полученной после упорядочивания конформаций по степени близости кривых рентгеновского малоуглового рассеяния. По оси ординат – порядковые номера тех же конформаций в последовательности, полученной при морфинге по методу, предложенному в работе [16].

сеяния в вакууме в диапазоне модуля вектора рассеяния $q = (0,001 \dots 1,2) \text{ \AA}^{-1}$. Далее, по методу «ближайшего соседа» была найдена та последовательность конформаций, для которой величина Δ_{α_1} (9) минимальна. В некоторых случаях на основе сравнения кривых рассеяния выделялись две обратные друг к другу после-

довательности конформаций с одинаковым минимальным значением величины Δ_{α_1} . Из этих двух последовательностей выбирали ту, которая начиналась с конформации, использованной в качестве начальной для морфинга.

На рис. 1 представлены кривые малоуглового рентгеновского рассеяния для белка кальмо-

дулина в диапазоне $q = (0,3 \dots 0,7) \text{ \AA}^{-1}$ (в этом диапазоне кривые различаются наиболее заметно). Из рисунка видно, как плавно изменяются кривые, соответствующие конформациям 1–5 и 6–10.

Последовательность конформационных переходов, найденную с помощью алгоритма морфинга и оптимизации крупномасштабных конформационных движений, назовем «упорядоченной в реальном пространстве». Последовательность переходов, упорядоченную на основе степени различия кривых рентгеновского малоуглового рассеяния, назовем «упорядоченной в обратном пространстве». Сравнение конформационных переходов, упорядоченных в реальном и обратном пространствах для пяти рассмотренных белков, представлены на рис. 2 в виде точечных диаграмм. В качестве идентификатора конформации для данного белка выбран порядковый номер этой конформации в последовательности, упорядоченной в реальном пространстве. На каждом из графиков рис. 2 для указанного белка представлена зависимость этого номера-идентификатора от порядкового номера в последовательности, упорядоченной в обратном пространстве. В идеале, если последовательность конформаций белка, упорядоченная в обратном пространстве, совпадает с последовательностью, упорядоченной в реальном пространстве, график должен иметь вид биссектрисы прямого угла. Как видно, обе последовательности в большинстве случаев хорошо совпадают, хотя имеются исключения, причина которых для каждого белка требует в дальнейшем своего анализа.

С целью иллюстрации того факта, что совпадение последовательностей конформационных переходов в реальном и обратном пространствах не может быть случайным, был проведен следующий расчет. Построены 102 последовательных конформации белка SSADH. Алгоритм генерации основан на методе получения начального приближения из работы [16]. Единственным отличием было то, что для торсионных углов основной цепи промежуточных конформаций были использованы произвольные значения. Далее были рассчитаны интенсивности малоуглового рассеяния для всех полученных конформаций и в соответствии с методом «ближайшего соседа» определена последовательность конформаций, упорядоченная в обратном пространстве. Как видно из рис. 3, нет ни малейшей корреляции между последовательностями, упорядоченными в реальном и обратном пространствах.

ЗАКЛЮЧЕНИЕ

В представленной работе предложен метод упорядочивания конформаций белковой моле-

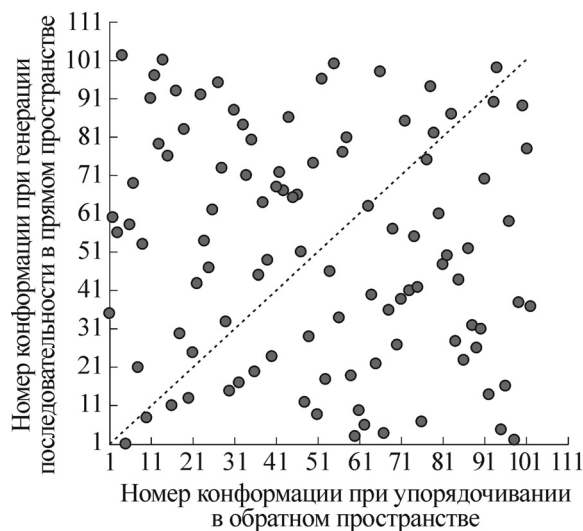


Рис. 3. Сравнение случайной последовательности из 102 конформаций белка SSADH, с последовательностью тех же конформаций, упорядоченной в обратном пространстве. По оси абсцисс отложены порядковые номера конформаций в последовательности, упорядоченной по степени близости кривых рентгеновского малоуглового рассеяния. По оси ординат – порядковые номера конформаций, полученные при первоначальной генерации последовательности путем произвольного вращения элементов главной цепи белка.

кулы на основе сравнения кривых рентгеновского малоуглового рассеяния и проведено сопоставление последовательностей конформаций, найденных с помощью морфинга, и последовательностей, упорядоченных по степени различия кривых рассеяния. Для каждого из пяти рассмотренных белков последовательности конформаций, полученные с помощью обоих методов, практически полностью совпадают.

При минимизации функционала (1) применялся ряд эвристик, основная из которых – оптимизация функционала только на движениях основной цепи (с искусственно «утяжеленными» на массы соответствующих боковых цепей атомами основной цепи), а затем «подстройка» боковых цепей при помощи алгоритма GROMACS. В то же время при вычислении кривых рассеяния сразу учитывается взаимное расположение всех атомов белковой молекулы. То, что предсказанное с помощью использованных эвристик конформационное движение дает наиболее плавное изменение кривой рассеяния, говорит в пользу того, что примененные эвристики не противоречат интуитивному предположению о плавной трансформации пространственной структуры белка (в смысле плавного изменения распределения электронной плотности), и полученные движения, по-видимому, мо-

гут реализовываться в природе. Более того, можно сформулировать следующую гипотезу: в «длинных» межконформационных движениях исследованных белков, по-видимому, определяющим является движение основной цепи, а боковые цепи оказывают только вторичное влияние.

Авторы выражают благодарность Г.С. Тамазяну за консультации по отдельным вопросам решения оптимизационных задач.

Работа выполнена при государственной финансовой поддержке ведущих университетов Российской Федерации (субсидия 074-U01), а также Санкт-Петербургского государственного университета (гранты 6.38.670.2013 и 6.38.223.2014), Правительства РФ (грант НШ-1771.2014.1), и Российского фонда фундаментальных исследований (грант РФФИ 14-01-00534)..

СПИСОК ЛИТЕРАТУРЫ

1. K. E. Tang and K. A. Dill, *J. Biomol. Struct. Dyn.* **16** (2), 397 (1998).
2. H. A. Carlson and J. A. McCammon, *Mol. Pharmacol.* **57** (2), 213 (2000).
3. А. В. Финкельштейн и О. Б. Птицын, *Физика белка* (Книжный дом «Университет», М., 2002).
4. К. В. Шайтан и К. Б. Терёшкина, *Молекулярная динамика белков и пептидов*. Методическое пособие (МГУ, М., 2004).
5. C. Bystroff, *Protein. Eng.* **14** (11), 825 (2001).
6. A. K. Dunker, J. D. Lawson, C. J. Brown, et al., *J. Mol. Graph. Model.* **19** (1), 26 (2001).
7. K. Gunasekaran, C. J. Tsai, S. Kumar, et al., *Trends Biochem. Sci.* **28** (2), 81 (2003).
8. K. Gunasekaran, B. Ma, and R. Nussinov, *Proteins* **57** (3), 433 (2004).
9. A. L. Fink, *Curr. Opin. Struct. Biol.* **15** (1), 35 (2005).
10. M. J. Forster, *Micron* **33** (4), 365 (2002).
11. I. Bahar, T. R. Lezon, L. W. Yang, et al., *Annu. Rev. Biophys.* **39**, 23 (2010).
12. V. Frappier, M. Chartier, and R. J. Najmanovich, *Nucl. Acids Res.* (2015).
13. A. Shehu, L. E. Kavrakli, and C. Clementi, *Biophys. J.* **92** (5), 1503 (2007).
14. Y. Phillip and G. Schreiber, *FEBS Lett.* **587** (8), 1046 (2013).
15. B. Macdonald, S. McCarley, S. Noeen, et al., *J. Phys. Chem. B.* **119** (7), 2956 (2015).
16. A. A. Koshevoy, E. O. Stepanov, and Yu. B. Porozov, *Biophysics* **59** (1), 28 (2014).
17. S. Knyazev, S. Tarakanov, V. Kuznetsov, et al., in *Proc. of 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* (2014), pp. 260–262.
18. Б. К. Вайнштейн, *Дифракция рентгеновских лучей на цепных молекулах* (Изд-во АН СССР, 1963).
19. M. Buzdalov, S. Knyazev, and Yu. B. Porozov, in *Proc. 13th Intern. Conf. on Machine Learning and Applications (ICMLA)* (2014), pp. 35–40.
20. J. Franklin, P. Koehl, S. Doniach, and M. Delarue, *Nucl. Acids Res.* **35**, W477 (2007).
21. G. Tamazian, J. H. Chang, S. Knyazev, et al., *Proteins: Structure, Function, and Bioinformatics*. Принята в печать. doi:10.1002/prot.24937 (2015).
22. H. M. Berman, J. Westbrook, Z. Feng, et al., *Nucl. Acids Res.* **28**, 235 (2000).

Prediction of Protein Conformational Mobility and Evaluation of Its Reliability Using Small-Angle X-ray Scattering

S.N. Knyazev*, V.Y. Kalyakin*, I.N. Deryabin*, B.A. Fedorov*,
A.V. Smirnov*, E.O. Stepanov* ** ***, and Yu.B. Porozov*

*ITMO University, Kronverkskiy prosp. 49, St. Petersburg, 197101 Russia

**St. Petersburg Department of Steklov Institute of Mathematics, Russian Academy of Sciences,
nab. Fontanki 27, St. Petersburg, 191023 Russia

***St. Petersburg State University, Universitetskiy prosp. 28, Peterhof, St. Petersburg, 198504 Russia

The «coarse-grained» model of protein conformational mobility is presented. We compared the trajectories of conformational motions predicted for five proteins using this model with the motion obtained by the method of the «nearest neighbor», based on small-angle X-ray scattering data. It is shown that for all studied proteins the sequence of conformations calculated on the basis of «coarse-grained» model and on the basis of the «nearest neighbor», coincides well, although there are exceptions. Some separate consideration should be given to each protein to discern the causes of these exceptions.

Key words: conformational motions of proteins, coarse-grained models, small-angle X-ray scattering, transportation modelling, conformational paths, motion simulations