

УДК 577.322.6+532.517.4

ДИНАМИКА УКЛАДКИ БЕЛКА В ПРОСТРАНСТВЕ ЭКСПЕРИМЕНТАЛЬНО ИЗМЕРЯЕМЫХ ПЕРЕМЕННЫХ: ЯВЛЕНИЯ ТУРБУЛЕНТНОСТИ

С. Ф. Чекмарев

Институт теплофизики им. С. С. Кутателадзе СО РАН, 630090 Новосибирск, Россия
Новосибирский национальный исследовательский государственный университет,
630090 Новосибирск, Россия
E-mail: chekmarev@itp.nsc.ru

С использованием метода молекулярной динамики исследован процесс укладки молекулы белка villin subdomain HP-35. При описании конформаций белка введены две переменные, соответствующие расстояниям между флуорофорами в экспериментах по исследованию динамики укладки белков с помощью метода ферстеровского резонансного переноса энергии. В результате моделирования показано, что поле потоков вероятностей переходов между состояниями белка насыщено вихрями. Обнаружено, что в рассмотренном случае, в отличие от ранее исследованных случаев гидродинамической турбулентности и турбулентности укладки белков в трехмерном конформационном пространстве, зависимость всех структурных функций потоков от величины инкремента в конформационном пространстве является линейной. Предложено объяснение этой зависимости на основе β -модели. Показано, что она не обусловлена выбором переменных для описания динамики укладки.

Ключевые слова: динамика укладки белков, молекулярная динамика, ферстеровский резонансный перенос энергии, потоки вероятностей переходов, турбулентность, структурные функции.

DOI: 10.15372/PMTF20180508

Введение. Исследование динамики укладки (фолдинга) белков является одной из основных проблем молекулярной биологии [1]. В природе белок синтезируется на рибосоме в клетке в виде цепочки аминокислотных остатков, последовательность которых кодируется соответствующим геном и задает свойства белка. Для функционирования молекула белка должна свернуться в определенную (так называемую нативную) конформацию, специфичную для каждого белка. Ошибки в сворачивании приводят к различным нейродегенеративным заболеваниям как у животных, так и у людей (“коровье бешенство”, болезнь Крейтцфельда — Якоба и др.). В природе укладка белков происходит с участием сложной белковой машинерии (шаперонов). Однако, как показано в работе [2], нативная конформация белка полностью определяется аминокислотной последовательностью. Поэтому для правильной укладки белка белковая машинерия не требуется и процесс укладки можно изучать как *in vitro* (в пробирке), так и *in silico* (на компьютере).

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 18-04-00013).

© Чекмарев С. Ф., 2018

В настоящее время структура белков исследована достаточно хорошо: более чем для 4 000 000 белков известны последовательности аминокислотных остатков (база UniRef [3]) и более чем для 100 000 белков подтверждены их нативные конформации (Protein Data Bank [4]).

Значительно труднее исследовать динамику укладки белков, определяющую ее правильность [5–8]. Для каждого белка пути укладки могут существенно различаться. При этом время укладки, как правило, варьируется в широком диапазоне (среднеквадратичное отклонение имеет величину порядка времени укладки). Поэтому для получения детальных сведений о динамике укладки необходимо исследовать не ансамбли белковых молекул, а отдельные молекулы. Среди современных экспериментальных методов исследования динамики укладки белков наиболее эффективным является метод ферстеровского резонансного переноса энергии (Förster resonance energy transfer (FRET)), который позволяет отслеживать изменение расстояния между флуорофорами, расположенными на различных участках молекулы одиночного белка [9–11]. От одного из флуорофоров (донора), возбуждаемого лазером, излучение передается другому флуорофору (акцептору) за счет безызлучательно-го диполь-дипольного переноса энергии, после чего регистрируется излучение акцептора. Интенсивность передачи энергии E между флуорофорами зависит от расстояния между ними r : $E = 1/[1 + (r/R_0)^6]$ (R_0 — критический радиус, при котором $E = 0,5$), что позволяет измерять расстояние между донором и акцептором в процессе укладки белка. В настоящее время при исследовании динамики укладки белков методом FRET [9–11] используется только пара хромофоров, которые обычно размещаются на концах молекулы белка. В этом случае удается построить профиль свободной энергии, характеризующий вероятность реализации определенного расстояния между участками белковой цепи, где закреплены донор и акцептор. Однако одномерные профили реакции не всегда адекватно описывают процесс укладки, поскольку проецируют существенно различные пути укладки на одну координату. Более информативны двумерные представления процесса укладки, например в виде двумерных поверхностей свободной энергии или двумерных поверхностей потоков вероятностей укладки. В экспериментах, проводимых с помощью метода FRET, можно измерять несколько расстояний между характерными участками молекулы белка, если использовать несколько хромофоров, размещенных вдоль белковой цепи [11]. Это позволяет получать двумерные картины укладки белков.

Одно из недавно обнаруженных существенных свойств динамики укладки белков заключается в том, что поведение потоков вероятностей укладки подобно поведению турбулентных потоков жидкости [12, 13]. В частности, поле потоков укладки белков насыщено вихрями, а структурные функции потоков имеют тот же вид, что и колмогоровские структурные функции в случае однородной изотропной турбулентности [14], т. е. функции второго и третьего порядков изменяются в зависимости от инкремента в конформационном пространстве l по законам $C_2(l) \sim l^{2/3}$ и $C_3(l) \sim l$, причем функция третьего порядка отрицательна. Представляет интерес выяснить, проявляются ли эти свойства турбулентности в FRET-экспериментах по исследованию динамики укладки белков, и если проявляются, то в какой мере.

Исследуемый белок и моделирование процесса укладки. Изучалась динамика укладки белка villin subdomain HP-35 (1wy4.pdb [15]), который представляет собой молекулу длиной 35 аминокислотных остатков, состоящую из трех α -спиралей (рис. 1). Использовалось огрубленное представление белка, что позволило значительно уменьшить объем вычислительной работы. В частности, использовалась C_α -модель молекулы белка, в которой каждый аминокислотный остаток заменяется мономером (“зерном”), помещенным на место C_α -атома в этом остатке. Потенциал взаимодействия мономеров выбран в виде модели Го [16], в которой доминирующими считаются нативные контакты. В рас-

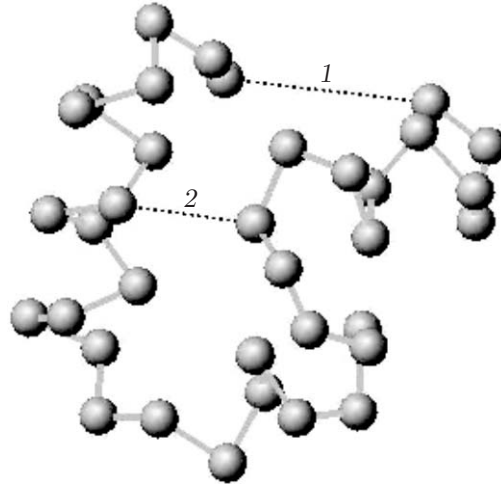


Рис. 1. Нативная структура белка villin subdomain HP-35 (1wy4.pdb [15]) в C_α -представлении:
1 — пара мономеров 1 и 35, 2 — пара мономеров 10 и 28

смаатриваемом случае этот потенциал имеет вид [17]

$$U = \sum_1^{N-1} [k_1(r_{i,i+1} - d_0)^2 + k_2(r_{i,i+1} - d_0)^4] + \sum_{i+1 < j}^{nat} 4\varepsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i+1 < j}^{non} 4\varepsilon \left[\left(\frac{\sigma_0}{r_{ij}} \right)^{12} - \left(\frac{\sigma_0}{r_{ij}} \right)^6 + \frac{1}{4} \right] \Delta(r_{ij} - d_{nat}), \quad (1)$$

где N — число мономеров; r_{ij} — расстояние между i -м и j -м мономерами; ε — характерная энергия притяжения мономеров. Первый член уравнения описывает жесткость цепи, причем $d_0 = 3,8\lambda$ ($\lambda = 10^{-8}$ см), $k_1 = \varepsilon/\lambda^2$ и $k_2 = 10^2\varepsilon/\lambda^4$, второй и третий члены — вклад нативных (nat) и ненативных (non) контактов соответственно. Контакт между i -м и j -м мономерами считался нативным, если $|i - j| > 1$ и расстояние между этими мономерами в нативной конформации меньше значения $r_{cut} = 7,8 \cdot 10^{-8}$ см. Во втором члене $\sigma_{ij} = 2^{1/6}d_{ij}$ (d_{ij} — расстояние между i -м и j -м мономерами в нативной конформации), в третьем — $d_{nat} = \langle d_{ij} \rangle$, $\sigma_0 = 2^{1/6}d_{nat}$, $\Delta(r_{ij} - d_{nat})$ — функция Хевисайда. Число нативных контактов равно $N = 99$.

Моделирование укладки белка проводилось методом молекулярной динамики на основе уравнения Ланжевена, обеспечивающего постоянную температуру T :

$$m \frac{d^2 \mathbf{r}_i}{dt^2} + \gamma \frac{d\mathbf{r}_i}{dt} = -\frac{\partial U}{\partial \mathbf{r}_i} + \Phi_i(t). \quad (2)$$

Здесь \mathbf{r}_i — радиус-вектор i -го мономера; m — масса мономера; Φ_i — случайная сила, действующая со стороны окружающего растворителя; γ — коэффициент трения “зерен” молекулы белка относительно растворителя. Случайные силы удовлетворяют гауссову распределению с нулевым средним и дисперсией $\langle \Phi_i^j(t) \Phi_{i'}^{j'}(t + \tau) \rangle = 2\gamma k_B T \delta_{ii'} \delta_{jj'} \delta(\tau)$, где k_B — постоянная Больцмана; нижний индекс у величины Φ обозначает номер мономера, верхний — компоненту вектора; $\delta_{ii'}$, $\delta_{jj'}$ — символы Кронекера; $\delta(\tau)$ — дельта-функция Дирака. Уравнение (2) интегрировалось численно с использованием разностной схемы [18] с шагом по времени $\Delta t = 0,0125\tau$ и коэффициентом $\gamma = 3m/\tau$. При масштабе длины

$l = 7,8 \cdot 10^{-8}$ см и характерной энергии притяжения мономеров $\varepsilon = 1,529 \cdot 10^{-20}$ Дж = 2,2 ккал/моль масштаб времени равен $\tau = (ml^2/\varepsilon)^{1/2} \approx 2,7 \cdot 10^{-12}$ с ($m = 1,862 \cdot 10^{-25}$ кг = 110 а.е.м.).

В соответствии с условиями проведения FRET-экспериментов [9] рассматривались “равновесные” условия, т. е. моделировалась одна длинная молекулярно-динамическая траектория (10^9 временных шагов), на которой молекула белка многократно сворачивалась до состояния, близкого к нативному, и разворачивалась. Расчеты проводились при $T = 300$ К. В качестве двух пар флуорофоров были приняты мономеры 1 и 35, расположенные на концах молекулы белка (пара 1), и мономеры 10 и 28, расположенные в середине цепи (пара 2) (см. рис. 1). При моделировании вычислялись расстояния d_1 между мономерами в паре 1 и d_2 в паре 2, которые определяли текущую конформацию белка. Поскольку расстояния d_1 и d_2 нельзя рассматривать в качестве независимых переменных, пространство $\mathbf{d} = (d_1, d_2)$ было преобразовано в пространство ортогональных переменных $\mathbf{g} = (g_1, g_2)$ с помощью метода главных компонент [19].

При вычислении потоков вероятностей переходов между состояниями белка в пространстве $\mathbf{g} = (g_1, g_2)$ использовалось гидродинамическое представление процесса укладки белка [20]. В соответствии с этим представлением поток вероятности переходов вдоль оси g_1 вычисляется по формуле

$$j_1(\mathbf{g}) = \frac{1}{\Delta g_2 N \Delta t} \left[\sum_{\mathbf{g}', \mathbf{g}''}^{g'_1 - g''_1 > 0} n(\mathbf{g}'', \mathbf{g}') - \sum_{\mathbf{g}', \mathbf{g}''}^{g'_1 - g''_1 < 0} n(\mathbf{g}'', \mathbf{g}') \right], \quad (3)$$

где $n(\mathbf{g}'', \mathbf{g}')$ — число переходов из состояния \mathbf{g}' в состояние \mathbf{g}'' через элемент поверхности g_2 , равный Δg_2 ; N — общее число переходов. Поток вдоль оси g_2 , т. е. поток $j_2(\mathbf{g})$, вычисляется аналогично, но в этом случае учитываются переходы через элементы поверхности g_1 , равные Δg_1 . В расчетах полагалось $\Delta g_1 \approx 0,2$ и $\Delta g_2 \approx 0,1$. Зная распределение потоков $\mathbf{j}(\mathbf{g})$, можно построить линии тока укладки

$$\Psi(g_1, g_2) = \int_0^{g_2} j_1(g_1, g'_2) dg'_2.$$

Результаты исследования и их обсуждение. На рис. 2,а представлены линии тока укладки молекулы белка. Видно, что поле потоков заполнено мелкими вихрями. На рис. 2,б показан увеличенный фрагмент поля течения. Из рис. 2,б следует, что типичный размер вихрей составляет приблизительно 10^{-8} см, т. е. по порядку величины он равен расстояниям d_0 и r_{cut} .

Для того чтобы охарактеризовать потоки, рассчитаны пространственные корреляционные функции потоков, т. е. структурные функции. В данном случае, в отличие от случая классической гидродинамической турбулентности, пространственная плотность молекул белка, которая представлена вероятностью их пребывания в данной точке пространства $\mathbf{g} = (g_1, g_2)$, существенно изменяется от точки к точке (на несколько порядков [12, 13]). Поэтому при вычислении структурных функций целесообразно использовать не скорости переходов между состояниями белка, а потоки вероятностей переходов, определяемые уравнением (3) [12, 13].

Рассмотрим продольные структурные функции

$$C_p(l) = \langle \delta j_{||}(l)^p \rangle,$$

где p — порядок структурной функции; $\delta j_{||}(l) = [\mathbf{j}(\mathbf{g} + \mathbf{l}) - \mathbf{j}(\mathbf{g})] \cdot \mathbf{l}/l$; \mathbf{l} — инкремент в пространстве \mathbf{g} ; угловые скобки обозначают осреднение по ансамблю состояний белка. На

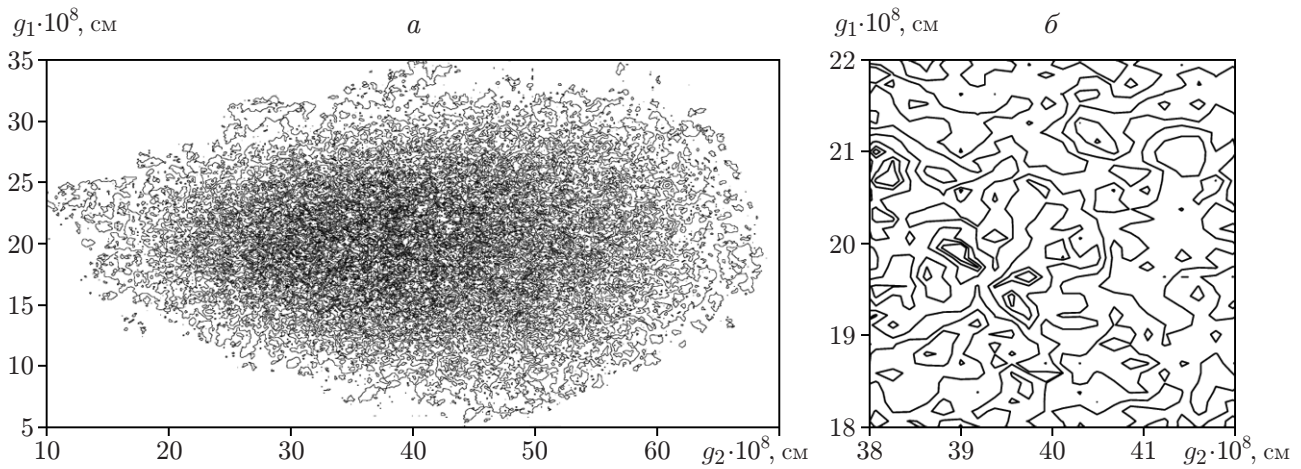


Рис. 2. Поле потоков сворачивания молекулы белка в форме распределения линий тока (*а*) и его увеличенный фрагмент (*б*)

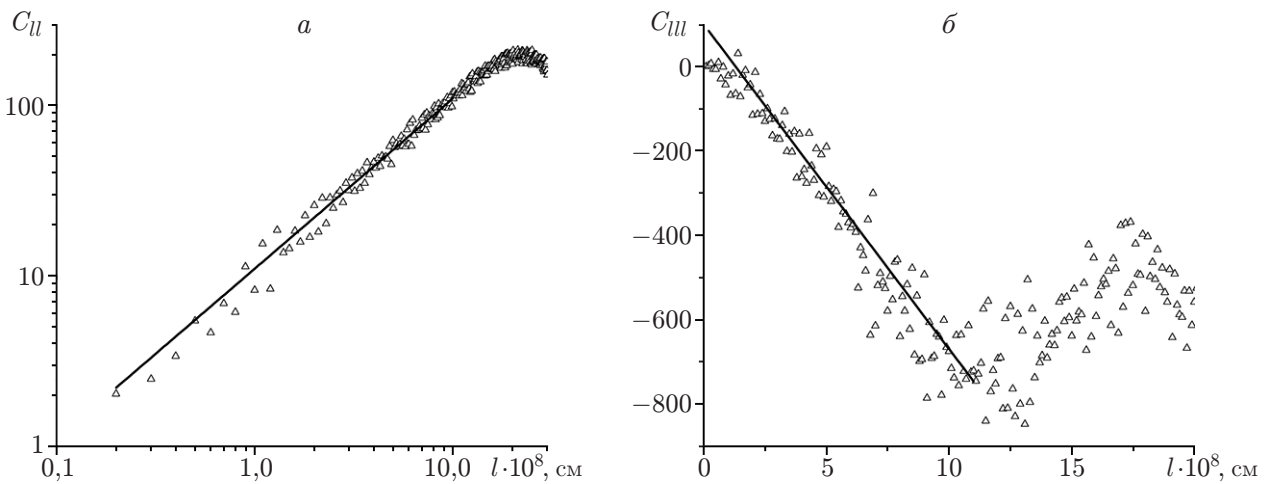


Рис. 3. Структурные функции второго (*а*) и третьего (*б*) порядков для потоков вероятности переходов в конформационном пространстве $\mathbf{d} = (d_1, d_2)$: точки — результаты моделирования, линии — аппроксимирующие линейные функции

рис. 3 показаны структурные функции второго и третьего порядков, нормированные на их значения при $l = 0,1$. Видно, что в обоих случаях зависимость структурных функций от l является практически линейной, т. е. показатель подобия приближенно равен единице. При этом структурная функция третьего порядка отрицательна и соответствующая ей кривая имеет отрицательный наклон. Те же особенности наблюдаются для функций более высокого четного и нечетного порядков (вычислялись функции третьего и четвертого порядков). Линейная зависимость структурных функций от l ограничена величиной $l \approx 10$. Это значение существенно меньше характерных размеров поля потоков (от $g_2 = 35$ до $g_1 = 70$ (см. рис. 2)), поэтому его можно рассматривать в качестве аналога верхней границы инерционного интервала в теории гидродинамической турбулентности [14]. Такое поведение структурных функций отличается от их поведения в случае “неравновесного” сворачивания молекул ряда белков, когда рассматривался ансамбль траекторий укладки из развернутого состояния белка в нативное и поток вычислялся в трехмерном пространстве коллективных переменных. В этом случае $C_2(l) \sim l^{2/3}$, что соответствует

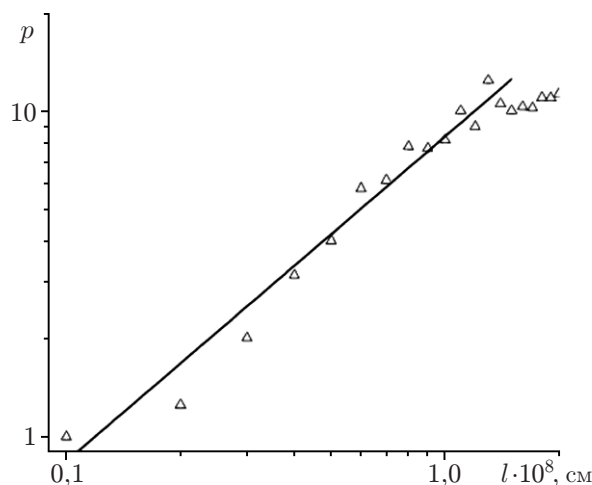


Рис. 4. Вероятность распределения вихрей по линейному размеру для потоков вероятности переходов в конформационном пространстве $\mathbf{d} = (d_1, d_2)$: точки — результаты моделирования, линии — аппроксимирующая линейная функция

теории А. Н. Колмогорова для инерционного интервала в случае однородной изотропной турбулентности [14]. Возможная причина такого отличия состоит в том, что в трехмерном и двумерном пространствах степень и характер насыщения поля потоков вихрями, в том числе степень фрактальности распределения вихрей по размеру, различны. Чтобы оценить вклад фрактальности, используем известную β -модель [21], в соответствии с которой степенной показатель зависимости функции p -го порядка от инкремента l определяется формулой

$$\zeta_p = p/3 + (D - K)(1 - p/3), \quad (4)$$

где D — размерность пространств; K — индекс фрактальности распределения вихрей по размеру. При этом множитель $D - K$ равен показателю Q в зависимости вероятности распределения вихрей по размеру l $p(l) \sim l^Q$. На рис. 4 показана эта зависимость, рассчитанная с использованием результатов моделирования (величина p нормирована на ее значение при $l = 0,1$). Размер вихря оценивался как расстояние между двумя точками, в которых достигаются экстремальные значения потока (для каждой составляющей потока с последующим осреднением по составляющим), т. е. как расстояния между минимумом и следующим максимумом, между этим максимумом и следующим минимумом и т. д. При этом флуктуациями, составляющими 20 % максимального изменения потока, пренебрегалось. На рис. 4 видно, что $Q \approx 1$. Тогда из уравнения (4) следует $\zeta_p \approx 1$, поэтому зависимость структурных функций любого порядка от инкремента l является линейной.

Следует отметить, что обнаруженная линейная зависимость структурных функций любого порядка от пространственного инкремента не обусловлена выбором переменных, в данном случае — расстояний между донорами и акцепторами, размещенными вдоль белковой цепи. Подобные зависимости имеют место и в том случае, если в качестве переменных используются также другие характеристики молекулы белка, в частности широко применяемые в теории укладки белков радиус инерции молекулы белка R_g , определяющий текущую компактность молекулы, и среднеквадратичное отклонение от нативной конформации σ , характеризующее степень близости конформации молекулы к нативной. Как и в рассмотренном выше случае, перед вычислением потоков конформационное пространство (σ, R_g) преобразовывалось в пространство ортогональных переменных (g_1, g_2) с помощью метода главных компонент. Результаты вычисления структурных функций показаны на рис. 5 (величины C_{II} , C_{III} нормированы на их значения при $l = 0,1$).

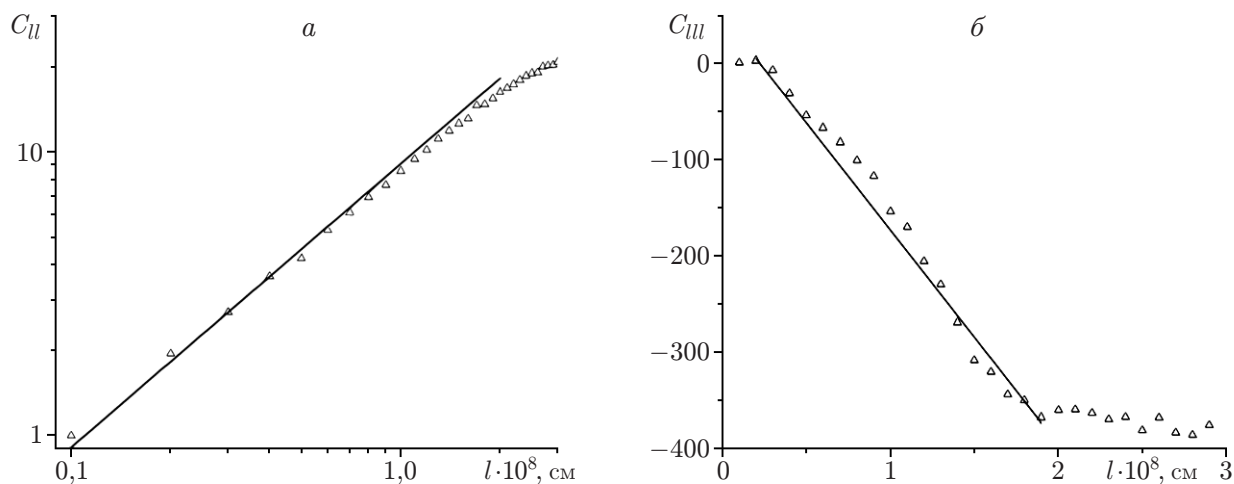


Рис. 5. Структурные функции второго (*а*) и третьего (*б*) порядков для потоков вероятности переходов в конформационном пространстве (σ, R_g): точки — результаты моделирования, линии — аппроксимирующие линейные функции

Заключение. С использованием метода молекулярной динамики проведено моделирование динамики укладки молекулы α -спирального белка в равновесных условиях, когда она многократно сворачивается и разворачивается. При моделировании FRET-экспериментов по исследованию динамики укладки белков в качестве переменных, характеризующих текущие конформации белка, выбрана пара расстояний между различными участками молекулы белка. Показано, что поле потоков вероятностей переходов между состояниями белка, определяемыми этими переменными, насыщено вихрями с характерным размером порядка 10^{-8} см. Обнаружено, что зависимость структурных функций потоков любого порядка от пространственного инкремента является линейной. Предложено объяснение этой зависимости на основе β -модели, учитывающей фрактальность распределения вихрей по размеру. Показано также, что полученная линейная зависимость структурных функций от пространственного инкремента не обусловлена выбором переменных: подобные зависимости имеют место и в том случае, если используются такие переменные, как радиус инерции молекулы белка и среднеквадратичное отклонение от нативной конформации.

ЛИТЕРАТУРА

1. **Финкельштейн А. В.** Физика белка / А. В. Финкельштейн, О. Б. Птицын. М.: Кн. дом "Университет", 2012.
2. **Anfinsen C. B.** Principles that govern the folding of protein chains // Science. 1973. V. 181. P. 223–230.
3. **Base UniRef.** [Электрон. ресурс]. Режим доступа: ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/uniref/.
4. **Protein Data Bank.** [Электрон. ресурс]. Режим доступа: <https://www.rcsb.org/>.
5. **Dinner A. R., Šali A., Smith L. J., et al.** Understanding protein folding via free-energy surfaces from theory and experiment // Trends. Biochem. Sci. 2000. V. 25. P. 331–339.
6. **Onuchic J. N., Luthey-Schulten Z., Wolynes P. G.** Theory of protein folding: The energy landscape perspective // Annual Rev. Phys. Chem. 1997. V. 48. P. 545–600.
7. **Dobson C. M., Šali A., Karplus M.** Protein folding: A perspective from theory and experiment // Angew. Chem. Intern. Ed. 1998. V. 37. P. 868–893.

8. **Dill K. A., Ozkan S. B., Shell M. S., Weikl T. R.** The protein folding problem // Annual Rev. Biophys. 2008. V. 37. P. 289–316.
9. **Chung H. S., McHale K., Louis J. M., Eaton W. A.** Single-molecule fluorescence experiments determine protein folding transition path times // Science. 2012. V. 335. P. 981–984.
10. **Schuler B., Hofmann H.** Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales // Current Opin. Struct. Biol. 2013. V. 23. P. 36–47.
11. **Lerner E., Cordes T., Ingargiola A., et al.** Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer // Science. 2018. V. 359. eaan1133.
12. **Kalgin I. V., Chekmarev S. F.** Turbulent phenomena in protein folding // Phys. Rev. E. 2011. V. 83. 011920.
13. **Andryushchenko V. A., Chekmarev S. F.** Turbulence in protein folding: Vorticity, scaling and diffusion of probability flows // PLoS One. 2017. V. 12. e0188659.
14. **Ландау Л. Д.** Теоретическая физика: В 10 т. Т. 6. Гидродинамика / Л. Д. Ландау, Е. М. Лифшиц. М.: Наука, 1986.
15. **Chiu T. K., Kubelka J., Herbst-Irmer R., et al.** High-resolution X-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein // Proc. Nat. Acad. Sci. USA. 2005. V. 102. P. 7517–7522.
16. **Gō N.** Theoretical studies of protein folding // Annual Rev. Biophys. Bioengng. 1983. V. 12. P. 183–210.
17. **Hoang T. X., Cieplak M.** Molecular dynamics of folding of secondary structures in Go-type models of proteins // J. Chem. Phys. 2000. V. 112. P. 6851–6862.
18. **Biswas R., Hamann D. R.** Simulated annealing of silicon atom clusters in langevin molecular dynamics // Phys. Rev. B. 1986. V. 34. P. 895–901.
19. **Jolliffe I.** Principal component analysis. N. Y.: Springer Verlag, 2002.
20. **Chekmarev S. F., Palyanov A. Yu., Karplus M.** Hydrodynamic description of protein folding // Phys. Rev. Lett. 2008. V. 100. 018107.
21. **Frisch U., Sulem P.-L., Nelkin M.** A simple dynamical model of intermittent fully developed turbulence // J. Fluid Mech. 1978. V. 87. P. 719–736.

Поступила в редакцию 7/VI 2018 г.
