

УДК 577.32; 532.5

И. В. Калгин<sup>1</sup>, С. Ф. Чекмарев<sup>1,2</sup>

<sup>1</sup> Новосибирский государственный университет  
ул. Пирогова, 2, Новосибирск, 630090, Россия

<sup>2</sup> Институт теплофизики им. С. С. Кутателадзе СО РАН  
пр. Акад. Лаврентьева, 1, Новосибирск, 630090, Россия

E-mail: chekmarev@itp.nsc.ru

## САМОПОДОБИЕ ПОТОКОВ УКЛАДКИ БЕЛКОВ

Проблема укладки белка в функциональное (нативное) состояние – одна из центральных проблем молекулярной биологии, привлекающая внимание биологов, физиков и химиков в течение многих лет. Особый интерес представляют общие закономерности процесса укладки, поскольку механизмы укладки разных белков могут сильно различаться. Ранее, при исследовании процесса укладки  $\text{fyn SH3}$  домена, нами было обнаружено, что при всем многообразии и сложности поведения индивидуальных траекторий укладки, потоки вероятностей укладки обладают выраженным свойством самоподобия, причем распределение потоков носит фрактальный характер. В данной работе мы исследуем это явление на примере другого белка –  $\text{beta3s}$ , который существенно отличается от  $\text{fyn SH3}$  домена как по структуре, так и по кинетике укладки. Кроме того, в отличие от  $\text{fyn SH3}$ , для которого использовалось огрубленное представление, мы проводим моделирование на атомном уровне разрешения. Показано, что свойства самоподобия и фрактальности потоков укладки хорошо выражены и в этом случае, что дает основание предполагать, что они являются характерными для динамики укладки белков.

*Ключевые слова:* динамика укладки белков, моделирование, потоки укладки, самоподобие, фрактальность.

### Введение

Понимание механизма укладки (фолдинга) белков в функциональное (нативное) состояние остается одной из актуальных проблем молекулярной биологии [1–4]. В начале 60-х гг. прошлого века Анфинсен показал, что трехмерная (3D) структура нативного состояния белка определяется последовательностью аминокислотных остатков, составляющих белковую цепь [5], т. е. для укладки белка в нативное состояние вовсе не обязательно участие сложной биологической машинерии в виде, например, шаперонов, контролирующих процесс укладки. Эта работа радикально расширила возможности изучения механизма укладки белков, позволив проводить исследования *in vitro* (в пробирке) и *in silico* (путем моделирования на компьютере). По-существу, белок начал рассматриваться как макромолекула, помещенная в естественную среду (раствори-

тель), поведение которой определяется взаимодействием атомов, составляющих белок, как между собой, так и с растворителем. Ярким отражением этой ситуации является тот факт, что при моделировании процесса укладки белков используются те же методы и подходы, что и при моделировании нанокластеров [6].

За последние несколько десятилетий в исследовании механизма укладки белков были достигнуты большие успехи в значительной мере благодаря моделированию, поскольку возможности экспериментального исследования процесса укладки весьма ограничены ввиду сложности самих белков (от десятков до тысяч аминокислотных остатков) и высокой скорости конформационных превращений в них (порядка пикосекунд на нижнем пределе). Методы же моделирования прогрессируют достаточно быстро – от так называемых распределенных вычислений, вовлекающих до миллиона

процессоров (проект `foldings@home` группы Панде в Стэнфордском университете<sup>1</sup>), до создания специализированных компьютеров для молекулярно-динамического моделирования белков (суперкомпьютер «Anton» лаборатории Шоу<sup>2</sup>).

Другая проблема, которую приходится решать при исследовании процесса сворачивания белков (в особенности при моделировании, где объем получаемой информации на атомном уровне чрезмерно велик), – это представление результатов в виде, доступном для анализа процесса укладки. Здесь можно выделить два основных подхода, которые активно и плодотворно используются.

Первый – это сведение многомерного конформационного пространства белка на атомном уровне к сокращенному пространству нескольких «коллективных» переменных, которые (предположительно) наилучшим образом характеризуют процесс укладки. С этой целью используются такие переменные, как радиус инерции белковой глобулы (характеризует компактность белковой структуры), число «нативных» контактов, т. е. контактов между аминокислотными остатками, которые свойственны нативному состоянию (характеризует близость к нативному состоянию), и др. [1–3]. Далее строится поверхность свободной энергии (ПСЭ) белка в зависимости от этих переменных, определяющая вероятность нахождения системы в текущей точке пространства коллективных переменных. Обычно эта поверхность имеет вид оврага, дно которого наклонено от развернутого состояния белка к нативному состоянию (что отвечает уменьшению потенциальной энергии белка), а ширина сужается (что отвечает уменьшению конформационной энтропии белка); при этом минимум свободной энергии отвечает нативному состоянию белка. Уменьшение потенциальной энергии и энтропии при следовании от развернутого к нативному состоянию происходит таким образом, что в «овраге» формируется барьер в виде седла, который разделяет частично свернутые состояния белка и состояния, близкие к нативному [1–3]. В этом отношении сворачивание белка напоминает фазовый переход, т. е. белок продолжительное

время находится либо в одном, либо в другом состоянии [3]. Как правило, ПСЭ не является гладкой; она состоит из большого числа локальных минимумов, отвечающих метастабильным состояниям белка. Некоторые из этих минимумов могут быть весьма глубокими и широкими – настолько, что будут конкурировать с нативным состоянием, как, например, минимумы, представляющие латентные состояния белка [7]. Основной недостаток ПСЭ состоит в том, что она определяет только вероятность нахождения белка в том или другом состоянии, но не указывает, в каком направлении белок эволюционирует, находясь в данной точке ПСЭ (сворачивается, разворачивается или пребывает в том же состоянии). И хотя знания ПСЭ достаточно, чтобы заключить, что белок должен сворачиваться (свободная энергия нативного состояния меньше, чем развернутого), и оценить, насколько быстро это происходит (определяется высотой барьера по свободной энергии), детальная динамика процесса остается неопределенной.

Второй подход – кинетический. Конформации белка, обнаруженные в ходе моделирования процесса укладки, собираются в кластеры, и вычисляются скорости переходов между кластерами. Кластеризация осуществляется на основе кинетической (высокая скорость переходов) или структурной (малые изменения структуры) близости конформаций. Полученные результаты оформляются в виде графов связности [6; 8; 9] и / или кинетических сетей [10–12]. Подход удобен тем, что можно не только представить, как протекает процесс укладки в целом, но и провести более детальное исследование. Зная скорости переходов, можно, например, использовать метод Монте-Карло, чтобы смоделировать типичные траектории укладки. Недостаток данного подхода состоит в том, что, во-первых, нет простых и надежных критериев для кластеризации конформаций и, во-вторых, приходится ограничиваться относительно небольшим числом кластеров, в то время как число различных конформаций обычно чрезвычайно велико (например, по известной оценке Левинтала [13], для белка в 100 аминокислотных остатков может быть до  $3^{198}$  конформаций). Поэтому нет уверенности, что детальная динамика процесса укладки, полученная путем моделирования методом Мон-

<sup>1</sup> <http://folding.stanford.edu>

<sup>2</sup> <http://www.deshawresearch.com>

те-Карло, не будет искажена, тем более, что процесс перехода между кластерами считается марковским, что не всегда выполняется. С целью исследования динамики укладки белков нами было предложено «гидродинамическое» описание процесса укладки [14]. В этом случае вероятности переходов, полученные в результате моделирования процесса укладки, используются для того, чтобы вычислить средние локальные потоки укладки в сокращенном конформационном пространстве, а затем по аналогии с гидродинамикой [15] рассчитать «векторы» и «линии тока» потоков укладки. В этой интерпретации процесс укладки представляет собой стационарный поток изображающих точек системы из развернутого состояния белка в нативное. Данный подход был применен для исследования укладки двух модельных белков – альфа-спиральной «шпильки» [14] и *fun* SH3 домена [16; 17]. В обоих случаях использовалось огрубленное представление белка, т. е. аминокислотные остатки представлялись в виде мономеров («бусинок»). Подход показал высокую эффективность. Было обнаружено, что линии тока укладки не следуют ландшафту ПСЭ. Более того, существуют области конформационного пространства, где наблюдаются вихри потоков, т. е. белок многократно частично собирается и разбирается. Дальнейшее исследование динамики укладки *fun* SH3 домена показало, что потоки

укладки для медленных траекторий обладают многими свойствами турбулентных течений жидкости, вплоть до того, что пространственные корреляционные (структурные) функции потоков согласуются с известными функциями для гидродинамической турбулентности, введенными А. Н. Колмогоровым [15].

При исследовании динамики укладки *fun* SH3 домена было также обнаружено, что при всем многообразии и сложности поведения индивидуальных траекторий укладки, потоки укладки весьма организованы – они обладают выраженным свойством самоподобия, причем распределение потоков имеет фрактальную структуру. Чтобы выяснить, насколько это свойство характерно для укладки белков, в данной работе мы проводим исследование другого белка – *beta3s*, который существенно отличается по структуре от SH3 доменов (плоский бета-лист вместо двух перпендикулярно ориентированных бета-листов) и намного короче (20 вместо 72 аминокислотных остатков), так что моделирование удастся провести на атомном уровне разрешения. Данный белок принадлежит к классу амилоидогенных белков и активно исследовался в последние годы (см., например, [12; 18; 19] и ссылки в этих работах). Кроме того, в отличие от SH3 домена, для которого характерна двухэкспоненциальная (двухстадийная) кинетика укладки [16], кинетика укладки *beta3s* – моноэкспоненциальная.

### Белок и методы исследования

Белок *beta3s* представляет собой последовательность аминокислотных остатков Thr1-Trp2-Ile3-Gln4-Asn5-Gly6-Ser7-Thr8-Lys9-Trp10-Tyr11-Gln12-Asn13-Gly14-Ser15-Thr16-Lys17-Ile18-Tyr19-Thr20 [20]. Нативная структура белка имеет вид плоского листа, состоящего из трех антипараллельных бета-тяжей (рис. 1). Для моделирования динамики укладки белка использовался пакет CHARMM [21]. Все тяжелые атомы и атомы водорода, связанные с атомами азота и кислорода, рассматривались явно. Для моделирования взаимодействия атомов внутри белка использовалось поле сил PARAM19 [22] с порогом 7,5 Å для несвязывающих взаимодействий (аминокислотных остатков, которые не являются соседями по белковой цепи). Водное окружение учитывалось не-

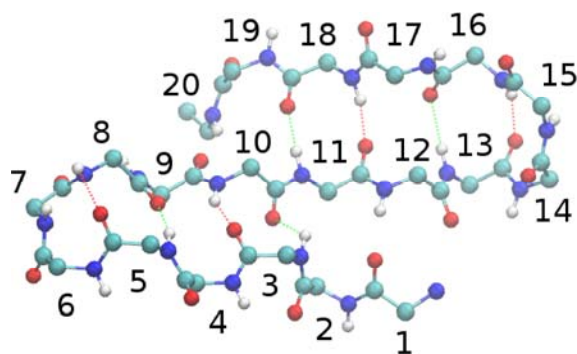


Рис. 1. Нативная структура *beta3s*. Аминокислотные остатки пронумерованы в соответствии с первичной последовательностью, указанной в тексте. Синие шарики – атомы азота, голубые – углерода, красные – кислорода, серые – водорода. Пунктирные линии – реализующиеся водородные связи. Чтобы сделать водородные связи более наглядными, боковые группы аминокислот, примыкающие к соседним с атомами азота атомам углерода (C-альфа атомам), не показаны

явно, с использованием метода поверхности, доступной для растворителя (SAS) [23]. Процесс укладки белка моделировался методом молекулярной динамики с шагом 2 фс. Моделирование проводилось при постоянной температуре  $T = 300$  К, использовался «термостат» Берендсена [24] для поддержания температуры.

Процесс укладки стартовал с развернутого состояния белка, которое готовилось путем выдерживания белка при повышенной температуре в течение  $5 \times 10^3$  временных шагов, и заканчивался при достижении нативного состояния. Для получения представительной статистики было сгенерировано 200 траекторий укладки.

В качестве исходного конформационного пространства использовалось пространство водородных связей между атомами кислорода в группах  $(CO)_i$  и атомами водорода в группах  $(NH)_j$ , где  $|j - i| \geq 2$ . Методом главных компонент [25] это пространство сводилось к пространству трех коллективных переменных  $\mathbf{g} = (g_1, g_2, g_3)$ , которые отвечают наибольшим собственным значениям ковариантной матрицы вероятностей формирования связей. Связь рассматривалась как сформированная, если длина связи не превышала  $3,6 \text{ \AA}$ , а угол ННО был не меньше  $130^\circ$ .

Зарегистрированные переходы между точками  $\mathbf{g}$ -пространства были организованы в виде локальных потоков вероятностей переходов  $\mathbf{j} = \mathbf{j}(\mathbf{g})$  [14].  $g_1$ -компонента потока в точке  $\mathbf{g}$  вычислялась как

$$j_{g_1}(\mathbf{g}) = \frac{1}{N} \left[ \sum_{\mathbf{g}^*, \mathbf{g}'(\mathbf{g} < \mathbf{g}^*)}^{g_1^* - g_1' > 0} n(\mathbf{g}^*, \mathbf{g}') - \sum_{\mathbf{g}^*, \mathbf{g}'(\mathbf{g} < \mathbf{g}^*)}^{g_1^* - g_1' < 0} n(\mathbf{g}^*, \mathbf{g}') \right], \quad (1)$$

где  $n(\mathbf{g}^*, \mathbf{g}')$  – число переходов из состояния  $\mathbf{g}'$  в состояние  $\mathbf{g}^*$ ,  $N$  – число траекторий, а символ  $\mathbf{g} < \mathbf{g}^*$  означает, что в учет берутся только переходы, для которых прямая линия, соединяющая точки  $\mathbf{g}'$  и  $\mathbf{g}^*$ , проходит в единичной окрестности точки  $\mathbf{g}$ . Компоненты потока  $g_2$  и  $g_3$  вычислялись аналогичным образом, с той лишь разницей, что выбирались переходы, пересекающие плоскости  $g_2 = \text{const}$  и  $g_3 = \text{const}$  соответственно.

Имея потоки  $\mathbf{j} = \mathbf{j}(\mathbf{g})$ , можно с помощью пассивных трассеров смоделировать линии тока. С этой целью решалось уравнение

$$\frac{d\mathbf{g}}{d\tau} = \mathbf{j}(\mathbf{g}),$$

где  $\mathbf{j} = \mathbf{j}(\mathbf{g})$  – локальный поток, определяемый уравнением (1), а  $\tau$  – параметр, имитирующий время.

## Результаты и обсуждение

На рис. 2 изображены линии тока потоков укладки beta3s при  $T = 300$  К в пространстве коллективных переменных  $\mathbf{g} = (g_1, g_2, g_3)$ ; переменные измеряются в ангстремах. Линии тока стартуют из развернутых состояний белка (большие положительные значения переменной  $g_1$ ) и заканчиваются в нативном состоянии.

Для исследования самоподобия потоков укладки мы вычисляли функцию

$$G(L) = \left\langle |G_{g_1, L} / \bar{J}_{g_1}| \right\rangle,$$

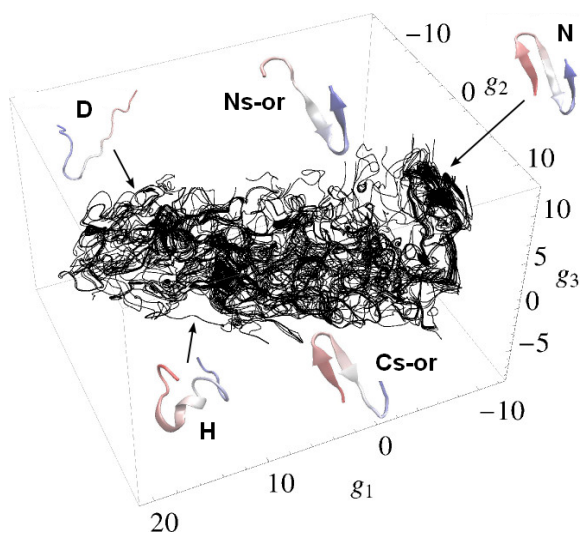


Рис. 2. Линии тока (трассеры) потоков укладки beta3s из развернутых состояний (D) в нативное состояние (N). Для иллюстрации также показаны характерные белковые структуры, наблюдающиеся в процессе укладки белка: H – конформации с участком в виде альфа-спирали, Cs-or и Ns-or – конформации, в которых соответственно N- и C-терминальные «шпильки» сформированы, а противоположные «шпильки» разобраны («or» обозначает «out of register» [12])

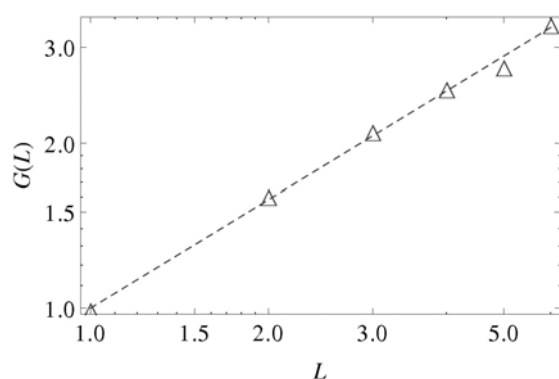


Рис. 3. Функция  $G = G(L)$ . Треугольниками показаны значения этой функции по результатам моделирования динамики укладки, а пунктирной линией – ее наилучшую аппроксимацию в виде  $G(L) \sim L^D$  ( $D \approx 0,66$ )

где  $|G_{g_1, L}|$  – абсолютная величина  $g_1$ -компоненты потока через квадрат со стороной  $L$ ,  $\bar{J}_{g_1} = \left( \sum_1^M j_{g_1, i}^2 / M \right)^{1/2}$  – средний поток в  $g_1$ -направлении,  $M$  – число элементарных ячеек, которые покрывают квадрат со стороной  $L$ , а угловые скобки обозначают осреднение по всем сечениям  $g_1 = \text{const}$  в пространстве  $\mathbf{g} = (g_1, g_2, g_3)$ . На рис. 3 показаны вычисленные значения  $G(L)$  и аппроксимация этой функции степенной зависимостью. Видно, что с хорошей точностью  $G(L) \sim L^D$ , причем показатель имеет нецелое значение ( $D \approx 0,66$ ). Это свидетельствует о том, что потоки самоподобны и носят фрактальный характер. Отметим, что показатель подобия не является универсальной величиной; для SH3 домена он варьировался от 1,53 на начальной стадии укладки до 0,96 на конечной [17].

### Заключение

Показано, что при всем многообразии и сложности поведения индивидуальных траекторий укладки beta3s в нативное состояние потоки укладки обладают выраженным свойством самоподобия с фрактальным распределением потоков, так как это было ранее обнаружено для динамики укладки fun SH3 домена [17]. Поскольку эти белки существенно отличаются по структуре и кинетике укладки, а также по модели, использо-

ванной для их представления (огрубленное представление fun SH3 домена и полноатомное представление для beta3s), можно предполагать, что свойство самоподобия потоков укладки является общим и характерным для динамики укладки белков.

### Список литературы

1. 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. Vol. 25. P. 331–339.
2. Onuchic J. N., Luthey-Schulten Z., Wolynes P. G. Theory of Protein Folding: The Energy Landscape Perspective // Annu. Rev. Phys. Chem. 1997. Vol. 48. P. 545–600.
3. Финкельштейн А. В., Птицын О. Б. Физика белка. М.: Книжный Дом «Университет», 2002. 376 с.
4. Dill K. A., Ozkan S. B., Shell M. S. et al. The Protein Folding Problem // Annu. Rev. Biophys. 2008. Vol. 37. P. 289–316.
5. Anfinsen C. B., Haber E., Sela M. et al. The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain // Proc. Natl. Acad. Sci. USA. 1961. Vol. 47. P. 1309–1314.
6. Wales D. J. Energy Landscapes. Cambridge: Cambridge University Press, 2003. 681 p.
7. Palyanov A. Yu., Krivov S. V., Karplus M. et al. A Lattice Protein with an Amyloidogenic Latent State: Stability and Folding Kinetics // J. Phys. Chem. B. 2007. Vol. 111. P. 2675–2687.
8. Becker O. M., Karplus M. The Topology of Multidimensional Potential Energysurface: Theory and Application to Peptide Structure and Kinetics // J. Chem. Phys. 1997. Vol. 106. P. 1495–1517.
9. Krivov S. V., Chekmarev S. F., Karplus M. Potential Energy Surfaces and Conformational Transitions in Biomolecules: A Successive Confinement Approach Applied to a Solvated Tetrapeptide // Phys. Rev. Lett. 2002. Vol. 88. P. 038101.
10. Krivov S. V., Karplus M. Hidden Complexity of Free Energy Surfaces for Peptide (Protein) Folding // Proc. Nat. Acad. Sci. USA. 2004. Vol. 101. P. 14766–14770.
11. Chodera J., Singhal N., Pande V. S. et al. Automatic Discovery of Metastable States for the Construction of Markov Models of

- Macromolecular Conformational Dynamics // *J. Chem. Phys.* 2007. Vol. 126. P. 155101.
12. Rao F., Caflisch A. The Protein Folding Network // *J. Mol. Biol.* 2004. Vol. 342. P. 299–306.
13. Levinthal C. How to Fold Graciously // *Mossbauer Spectroscopy in Biological Systems* / Eds. P. Debrunner, J. C. M. Tsibris, E. Münck. Urbana: Univ. Illinois Press, 1969. P. 22–24.
14. Chekmarev S. F., Palyanov A. Yu., Karplus M. Hydrodynamic Description of Protein Folding // *Phys. Rev. Lett.* 2008. Vol. 100. P. 018107.
15. Ландау Л. Д., Лифшиц Е. М. Гидродинамика. М.: Наука, 1986. 736 с.
16. Kalgin I. V., Karplus M., Chekmarev S. F. Folding of a SH3 Domain: Standard and «Hydrodynamic» Analyses // *J. Phys. Chem. B.* 2009. Vol. 113. P. 12759–12772.
17. Kalgin I. V., Chekmarev S. F. Turbulent Phenomena in Protein Folding // *Phys. Rev. E.* 2011. Vol. 83. P. 011920.
18. Krivov S. V., Muff S., Caflisch A. et al. One-Dimensional Barrier-Preserving Free-Energy Projections of a  $\beta$ -sheet Miniprotein: New Insights into the Folding Process // *J. Phys. Chem. B.* 2008. Vol. 112. P. 8701–8714.
19. Qi B., Muff S., Caflisch A. et al. Extracting Physically Intuitive Reaction Coordinates from Transition Networks of a  $\beta$ -sheet Miniprotein // *J. Phys. Chem. B.* 2010. Vol. 114. P. 6979–6989.
20. De Alba, E., Santoro J., Rico M. et al. De novo Design of a Monomeric Three-Stranded Antiparallel  $\beta$ -sheet // *Protein Sci.* 1999. Vol. 8. P. 854–865.
21. Brooks B. R., Bruccoleri R. E., Olafson B. D. et al. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations // *J. Comp. Chem.* 1983. Vol. 4. P. 187–217.
22. Neria E., Fischer S., Karplus M. Simulation of Activation Free Energies in Molecular Systems // *J. Chem. Phys.* 1996. Vol. 105. P. 1902–1921.
23. Ferrara P., Apostolakis J., Caflisch A. Evaluation of a Fast Implicit Solvent Model for Molecular Dynamics Simulations // *Proteins: Structure, Function, and Bioinformatics.* 2002. Vol. 46. P. 24–33.
24. Berendsen H. J. C., Postma J. P. M., DiNola A. et al. Molecular Dynamics with Coupling to an External Bath // *J. Chem. Phys.* 1984. Vol. 81. P. 3684–3690.
25. Jolliffe I. T. Principal Component Analysis. 2<sup>nd</sup> ed. N. Y.: Springer, 2002. 502 p.

Материал поступил в редколлегию 08.06.2012

I. V. Kalgin, S. F. Chekmarev

#### SELF-SIMILARITY OF PROTEIN FOLDING FLOWS

The problem of how a protein folds into its functional (native) state is one of the central problems of molecular biology, which attracts the attention of researchers from biology, physics and chemistry for many years. Of particular interest are general properties of the folding process, because the mechanisms of folding of different proteins can be essentially different. Previously, in the study of folding of fyn SH3 domain, we found that despite all the diversity and complexity of individual folding trajectories, the folding flows possess a well pronounced property of self-similarity, with a fractal character of the flow distributions. In the present paper, we study this phenomenon for another protein – beta3s, which is essentially different from the SH3 domain in its structure and folding kinetics. Also, in contrast to the fyn SH3 domain, for which a coarse-grained representation was used, we perform simulations on the atomic level of resolution. We show that the self-similarity and fractality of folding flows are observed in this case too, which suggests that these properties are characteristic of the protein folding dynamics.

*Keywords:* protein folding dynamics, simulation, folding flows, self-similarity, fractality.