

## РЕШЕНИЕ ПАРАДОКСА ЛЕВИНТАЛЯ ВОЗМОЖНО НА УРОВНЕ ФОРМИРОВАНИЯ И УПАКОВКИ ВТОРИЧНЫХ СТРУКТУР БЕЛКОВ

© 2016 г. А.В. Финкельштейн, С.А. Гарбузинский

*Институт белка РАН, 142290, Пущино Московской области, ул. Институтская, 4*

*E-mail: afinkel@vega.protres.ru*

Поступила в редакцию 17.09.15 г.

Полный объем конформационного пространства белковой молекулы на много порядков меньше на уровне укладки элементов вторичной структуры, чем его объем на уровне конформаций аминокислотных остатков. Последний, по оценке Левинталя, растет как  $\sim 10^{2L}$  с ростом числа  $L$  остатков в цепи, в то время как первый, на уровне укладки элементов вторичной структуры, растет не быстрее, чем  $\sim L^N$  с ростом числа  $N$  участков вторичной структуры, причем  $N$ , как показывает статистика белковых структур, близко к  $L/15$ . Столь значительное уменьшение показателя степени (до  $L/15$  вместо  $2L$ ) резко уменьшает число подлежащих перебору состояний, что и объясняет, почему перебор конформаций на уровне элементов вторичной структуры не делает невозможным нахождение белковой цепью ее наиболее стабильной укладки.

*Ключевые слова: сборка вторичных структур, формирование вторичных структур, объем пространства конфигураций, сворачивание белка, скорость сворачивания белка, белковые структуры.*

Способность белковой цепи спонтанно свернуться в сложную пространственную структуру долго озадачивала исследователей. Много лет назад О.Б. Птицын предложил рассматривать сворачивание белка как процесс формирования и сборки элементов вторичной структуры [1]. Развивая модель Птицына, Дж. Роллинз и К. Дилл недавно смогли подобрать такие величины для внутренней стабильности вторичных структур и для свободной энергии их связывания и такие кинетические коэффициенты для всех шагов процесса, чтобы получить разумные, хоть и приближенные, оценки для времени сворачивания разных белков [2].

В этой статье мы покажем, что обобщение и упрощение модели Птицына–Роллинза–Дилла дает простое решение известного парадокса Левинталя. Как известно, этот парадокс утверждает [3], что сворачивание белков сталкивается с астрономическим числом конфигураций цепи ( $\sim (10^2)^L$ , где  $L$  – число аминокислотных остатков в цепи (обычно  $L \sim 100$ ), а  $10$  – число микросостояний по каждой из двух ( $\phi$  и  $\psi$ ) степеней свободы аминокислотного остатка). Таким образом, «по Левинталю», если белок должен был бы достичь своей правильной конфигурации путем полного перебора всех возможных конформаций цепи, это потребовало бы времени большего, чем возраст Вселенной.

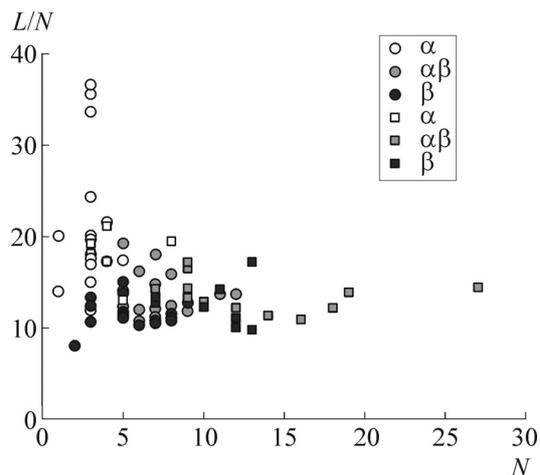
Ниже мы показываем (см. также [4]), что полный перебор, невозможный на уровне аминокислотных остатков, становится возможным на уровне формирования и упаковки элементов вторичной структуры.

Основной причиной этого является то, что число  $N$  элементов вторичной структуры ( $\alpha$ -спиралей и  $\beta$ -тяжей) и, следовательно, число их степеней свободы гораздо меньше, чем число  $L$  аминокислот в цепи. Численно, как показывает статистика белковых структур,  $N$  близко к  $L/15$  (рис. 1). Одно это на много порядков уменьшает объем конфигурационного пространства, рассматриваемого на уровне вторичных структур, по сравнению с объемом конфигурационного пространства, рассматриваемого на уровне аминокислотных остатков цепи.

Еще одна причина состоит в том, что возможные положения каждого элемента вторичной структуры, а тем более сочетания таких положений, сильно ограничены в белковых глобулах (структурный каркас которых, как известно (см. работы [5–9] и цитируемую там литературу), сложен из вторичных структур).

Оценим число возможных конфигураций – или, строже говоря, число локальных минимумов энергии для белковой глобулы.

При этом нужно принять во внимание следующее.



**Рис. 1.** Число  $L/N$  аминокислотных остатков, приходящихся на один элемент вторичной структуры в  $\alpha$ -,  $\beta$ - и  $\alpha\beta$ -белках ( $L$  – число аминокислотных остатков в цепи;  $L/N$  включает не только среднюю длину  $\alpha$ - и  $\beta$ -участков, но также и среднюю длину петель между этими участками [2]); в среднем по всем типам белков,  $L/N = 15 \pm 5$  ( $20 \pm 6$  в  $\alpha$ -белках,  $12 \pm 2$  в  $\beta$ -белках,  $13 \pm 2$  в  $\alpha\beta$ -белках). Кругами отмечены те белки, что, на опыте, сворачиваются в одну стадию; квадратиками – те, что в несколько стадий (видно, что между  $\alpha$ - и  $\beta$ -участками в белках этих двух типов нет принципиальной разницы). Данные взяты из таблиц S3 и S4 вспомогательной информации к работе [2].

1. Число всех возможных комбинаций позиций  $N$  элементов в сложной из них архитектуре белковой глобулы не может превышать  $N! \equiv N \times (N - 1) \times \dots \times 2 \times 1$ , т.е. первый элемент может занимать любую из  $N$  позиций в архитектуре, второй – любую из оставшихся  $N - 1$  позиций и т.д. (см. рис. 2). Отметим, что здесь говорится «не может превышать», а не «равно», так как некоторые из таких комбинаций могут включить  $\alpha$ -спираль в  $\beta$ -лист, что физически невозможно [7–10], другие комбинации могут требовать слишком длинных петель и т.д. Число же комбинаций направлений элементов (каждый из которых может иметь два направления, см. рис. 2д) не может превышать (по указанным выше причинам – «не может превышать», а не «равно»)  $2^N$ .

2. В архитектуре глобулы, т.е. в плотной упаковке  $\alpha$ - и/или  $\beta$ -структур, оба поперечных сдвига и оба поперечных наклона каждого элемента вторичной структуры практически запрещены плотной упаковкой, т.е. каждый элемент сохраняет только две степени свободы: сдвиг вдоль своей оси и вращение вокруг этой оси (см. рис. 3). Но и эти два перемещения не являются независимыми в плотноупакованной глобуле, где каждый сдвиг элемента вторичной

структуры требует (при переходе от одного локального минимума энергии к другому) определенного его поворота, а каждый поворот – сдвига. В результате каждый элемент может иметь порядка  $L/N$  (столько, сколько на него приходится остатков цепи) возможных сдвигов/поворотов в глобуле.

Вместе с приведенной выше верхней оценкой числа возможных позиций и ориентаций вторичных структур в глобуле ( $N! \times 2^N$ ) допустимые сдвиг/повороты вторичных структур ограничивают количество возможных их конфигураций числом порядка  $(L/N)^N \times N! \times 2^N$  – при заданной архитектуре глобулы, состоящей из  $N$  заданных вторичных структур в цепи из  $L$  аминокислотных остатков.

Это число,  $(L/N)^N \times N! \times 2^N$ , может быть несколько снижено симметрией глобулы вследствие того, что  $\alpha$ -спираль не может быть включена в  $\beta$ -лист, а также того, что короткие петли между вторичными структурами могут не позволить им занять какие-то позиции и иметь какие-то направления в глобуле, и т.д. Однако это снижение не столь важно для нас, потому что наша цель – оценить верхний предел числа возможных конфигураций.

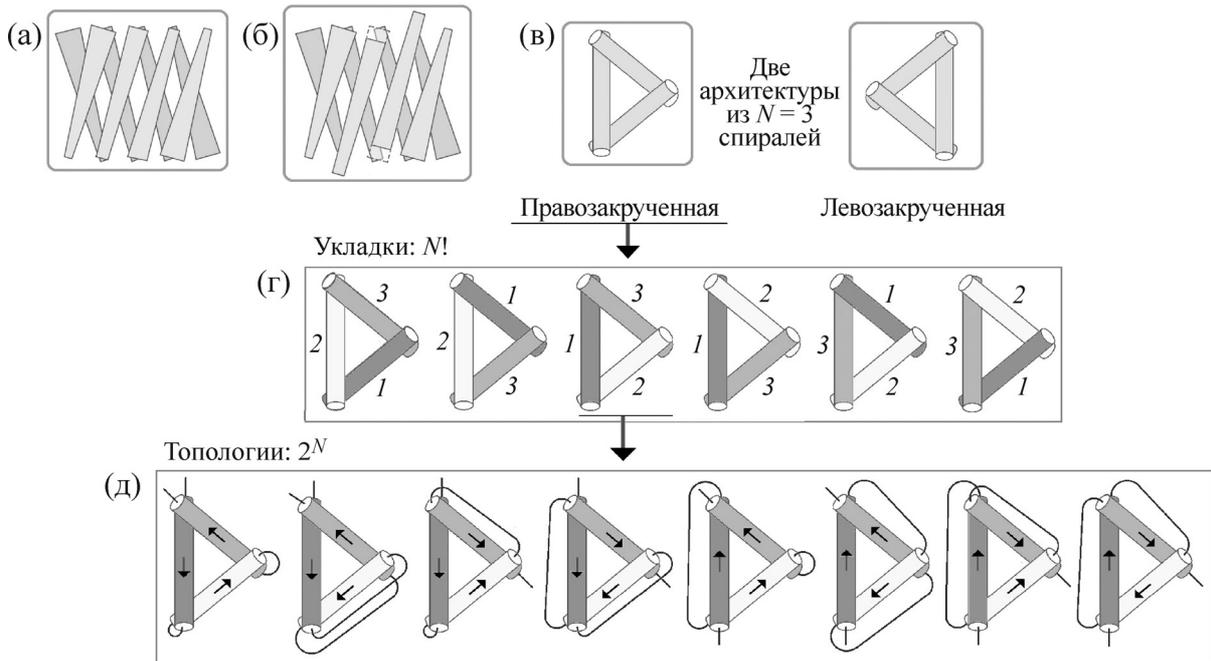
Число архитектур белковой глобулы – для заданного набора структурных элементов – сравнительно невелико, порядка 10 (так, есть две стандартные архитектуры для глобул, состоящих из трех  $\alpha$ -спиралей, см. рис. 2; есть 8–10 архитектур для компактных глобул из четырех–шести  $\alpha$ -спиралей [6] и т.д.).

Поэтому число всех конфигураций глобулы из  $N$  структурных элементов может быть оценено сверху, по порядку величины, как  $\sim 10 \times (L/N)^N \times N! \times 2^N$ .

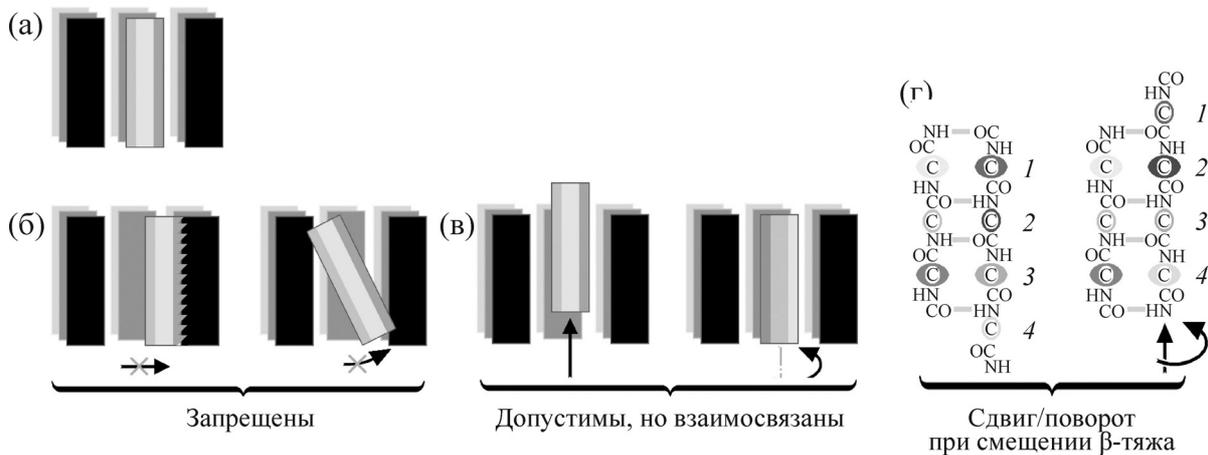
Вторичные структуры белковой цепи довольно хорошо определяются локальными аминокислотными последовательностями [10,12,13], хотя участок цепи может еще не иметь своей вторичной структуры до вхождения в белковую глобулу. Поэтому при оценке объема конфигурационного пространства надо учесть, что каждый элемент вторичной структуры белка может быть в двух состояниях – «уже свернутом» и «еще развернутом».

Это, однако, не сильно увеличивает число подлежащих перебору структур цепи.

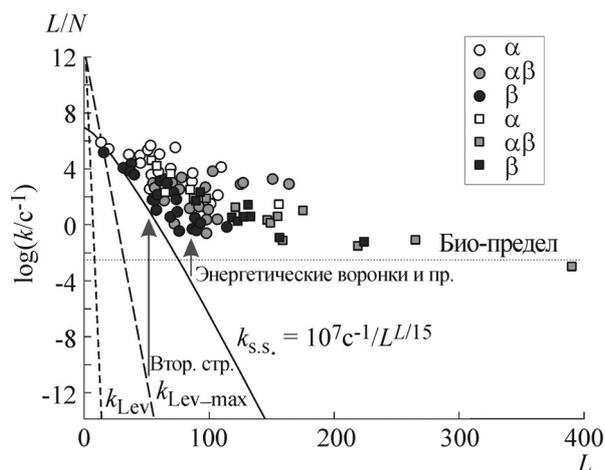
В самом деле, рассмотрим один из  $N!$  (рис. 2) возможных способы упаковки  $N$  элементов вторичной структуры в данную архитектуру белка. Число комбинаций всех возможных направлений, сдвигов и поворотов составляет  $2^N \times (L/N)^N$ , если все элементы вторичной структуры



**Рис. 2.** (а) – Одна из двух типичных архитектур  $\beta$ -структурного белка – коллинеарная укладка  $\beta$ -листов [5]. (б) – Та же архитектура, где, в изображенной структуре, два центральных  $\beta$ -тяжа одного  $\beta$ -листа сдвинуты (по сравнению с рисунком (а)) вдоль своих осей. (в) – Идеализированное изображение обеих возможных архитектур глобулы, сложенной из  $N = 3$   $\alpha$ -спиралей (пучки спиралей: право- и левозакрученный) [6]. (г) – Все  $3! = 3 \times 2 \times 1 = 6$  способов укладки, т.е. взаимного расположения элементов вторичной структуры в одной (подчеркнутой) из трехспиральных архитектур. Боковые цепи и петли, соединяющие структурные элементы, не показаны на рисунках (а) – (г); на рисунках (а) – (в) схематично показаны только обводы глобул. (д) – Все  $2^3 = 8$  топологий, т.е. путей цепи через одну (подчеркнутую) из изображенных укладок. Петли, соединяющие  $\alpha$ -спирали, показаны очень схематично. Боковые цепи не показаны.



**Рис. 3.** (а) – Схематичное изображение белковой архитектуры (т.е. места элементов вторичной структуры в плотной укладке). (б) – Плотная укладка запрещает поперечный сдвиг и поперечный наклон структурного элемента (для простоты это показано только для одного – центрального – элемента; сдвиги и наклоны элемента в перпендикулярной рисунку плоскости запрещены тоже). (в) – Допустим сдвиг вдоль оси структурного элемента, допустим его поворот вокруг этой оси, но они взаимосвязаны: сдвиг требует определенного поворота, и наоборот. (г) – Сочетание сдвига и поворота при смещении  $\beta$ -структурного участка (на рисунке – правого; номера при его  $C^\alpha$ -атомах позволяют следить за его смещением). Эллипсами отмечены  $C^\alpha$ -атомы тех аминокислотных остатков, боковые группы которых ориентированы – до смещения и после него – в сторону читателя (сами эти боковые группы не показаны на рисунке); кружками обведены  $C^\alpha$ -атомы тех аминокислотных остатков, боковые группы которых глядят в обратном направлении. Водородные связи между  $\beta$ -тяжами показаны серыми линиями. Аналогичная картина сочетания сдвига с поворотом характерна и для плотной упаковки  $\alpha$ -спиралей [11], но изобразить его намного сложнее.



**Рис. 4.** Зависимость скорости  $k$  (измеренной в  $\text{с}^{-1}$ ) сворачивания белков (кружки и квадраты) и полного перебора структур (линии) от числа аминокислотных остатков  $L$  в цепи. Экспериментальные данные взяты из таблиц S3 и S4 дополнительной информации к работе [2]. Сплошная линия соответствует перебору на уровне вторичных структур:  $k_{s.s.} = k_r L^{L/15}$ , где  $k_r = 107 \text{ с}^{-1}$  получена из уравнения  $k_r / 14^{14/15} = 0,8 \times 10^6 \text{ с}^{-1}$  для скорости образования спирали в пептиде из 14 остатков [17] (согласно опыту –  $0,8 \times 10^6 \text{ с}^{-1}$ ). Пунктирные линии (только их небольшие части уместаются на рисунке) соответствуют исходной оценке Левинталя [3],  $k_{Lev} = 10^2 \cdot 10^{12} \text{ с}^{-1} / 10^{2L}$ , и «максимальной квази-Левинталевой оценке»  $k_{Lev-max} = 3 \times 10^{12} \text{ с}^{-1} / 3^L$ ; последняя исходит из того, что у аминокислотного остатка есть не  $10^2$  микросостояний, а только 3 ( $\alpha, \beta$ , «клубок»); в обоих этих оценках считается, что один остаток структурируется за 1 пикосекунду. Стрелки схематически показывают роль вторичных структур и энергетических воронок в ускорении сворачивания белков по сравнению с полным перебором на уровне конформаций аминокислотных остатков. Точечная линия ориентировочно показывает «биологически приемлемую» скорость сворачивания [16].

уже свернуты. Если же один из  $N$  элементов вторичной структуры развернут (и не входит в укладку), то суммарное число комбинаций всех возможных направлений, сдвигов и поворотов остальных (структурированных) элементов составляет  $N \times 2^{N-1} \times (L/N)^{N-1}$ . Если развернуты два элемента вторичной структуры, то соответствующее число есть  $[N(N-1)/2] \times 2^{N-2} \times (L/N)^{N-2}$  и так далее. Таким образом, учет сочетаний всевозможной комбинаторики разворачивания  $n = 0, 1, \dots, N$  из  $N$  структурных элементов приводит к суммарному числу конфигураций, равному

$$\sum_{n=0}^N \frac{N!}{n!(N-n)!} \left(\frac{2L}{N}\right)^{N-n} = \left(1 + \frac{2L}{N}\right)^N$$

(что не сильно отличается от ранее полученного  $(2L/N)^N$ , когда  $L \gg N$ ).

Поэтому число всех конфигураций цепи из  $N$  структурных элементов может быть оценено, по порядку величины, как  $\sim (2L/N)^N \times N! \times 10$ , или (так как  $N! \sim (N/e)^N$  по формуле Стирлинга) как  $\sim L^N$  в основном члене при  $L \gg N \gg 1$ . Если  $N \approx L/15$  (см. рис. 1), то  $L^N \approx (15N)^N \approx L^{L/15}$ .

(Необходимо заметить, что для очень больших длин  $L$  на один участок вторичной структуры в глобуле должно приходиться не  $L/15$ , а больше – порядка  $L^{1/3}$  – аминокислотных остатков (учитывая, что вторичная структура стабильна тогда, когда она идет от одного края глобулы до другого [10]). Поэтому число вторичных структур в такой длинной цепи – порядка  $L/L^{1/3} \sim L^{2/3}$ , а число всех конфигураций этих структур – порядка  $L^{L^{2/3}}$ . Эта оценка в основном члене (показателе степени, равно  $\approx L^{2/3}$ ) соответствует полученному нами ранее результату [14–16]. Однако, как показывают наши прикидки, существенное превосходство  $L^{L/15}$  над  $L^{L^{2/3}}$  проявится только в колоссальных (и не наблюдаемых в глобулярных белках) цепях из тысяч аминокислотных остатков.)

Таким образом, число конфигураций у белковой цепи из  $L = 100$  остатков составляет, на уровне элементов вторичной структуры, около  $10^{13}$ , что несравненно меньше, чем оценка Левинталя ( $\sim 10^{200}$ ) для той же цепи на уровне аминокислотных остатков. Резкое снижение показателя степени (с 200 до 13) делает просмотр всего конфигурационного пространства на уровне вторичных структур в принципе допустимым для белка из 100 остатков (хотя и весьма длительным – учитывая, что время формирования одной  $\alpha$ -спирали в цепи из  $\approx 20$  остатков составляет  $\sim 10^{-6}$  с [17], а шпильки из двух  $\alpha$ -участков в цепи той же длины –  $\sim 10^{-5}$  с [18]).

Конечно, полное время просмотра конфигурационного пространства больше, чем фактическое время складывания структуры, поскольку к нативной структуре белка ведет, как сейчас принято говорить, «энергетическая воронка» [2, 16, 19, 20], позволяющая, из-за падения энергии при сворачивании, избежать перебора всех вариантов в поисках энергетически-оптимальной структуры. И хотя «эффект воронки» не позволяет исключить перебор [16, 21], как то первоначально полагалось [19, 20], он позволяет его сильно уменьшить. Поэтому остающийся перебор только замедляет (по сравнению, например, со складыванием одной  $\alpha$ -спирали), но не предотвращает попадание в самую низкоэнергетичную укладку (если белок – или домен большого белка – содержит не более чем  $\sim 100$  аминокислотных остатков [16]).

Сравнительная роль «воронок» и уменьшения перебора наличием вторичных структур в сворачивании белка иллюстрируется на рис. 4.

На этом рисунке экспериментально измеренные скорости сворачивания белков сравниваются с теоретической оценкой зависимости скорости полного (на уровне элементов вторичной структуры) перебора конформаций от длины белковой цепи. Эта зависимость вычислена как  $k/L^{L/15}$ , где  $L^{L/15}$  – оцененное выше число конфигураций, а элементарная скорость  $k_e$  выбрана такой, чтобы соответствовать экспериментально измеренной скорости образования одной  $\alpha$ -спирали из 14 остатков [17]. Обращает на себя внимание то, что экспериментально измеренные скорости сворачивания белков начинают резко отрываться от скорости полного (на уровне вторичных структур) перебора, когда в цепи содержится более 80–100 аминокислот. Это – те же длины (и те же скорости), при которых «термодинамический контроль» (который в принципе может потребовать полного перебора всех структур цепи) становится возможным лишь для относительно простых (не слишком запутанных) нативных укладок белковых цепей [16], потому что структуры, время образования которых превосходит «био-предел» (многие минуты и более) не могут успеть образоваться в клетке.

То, что наша оценка числа конфигураций делается на уровне элементов вторичной структуры, а не аминокислотных остатков (как это делал Левинтал), оправдывается тем, что: (а) вторичные структуры довольно четко определяются локальными аминокислотными последовательностями [12], и (б) экспериментально наблюдаемое время формирования  $\alpha$ -спирали и  $\beta$ -шпильки только на два–три порядка больше, чем время конформационной перестройки одного аминокислотного остатка (ср. работы [17,18] с работой [22]); эти два–три порядка не имеют существенного значения для принципиального рассмотрения скоростей сворачивания белков, охватывающих – на опыте – диапазон в девять порядков величины [16]).

В заключение надо отметить, что на уровне элементов вторичной структуры объем подлежащего перебору конфигурационного пространства,  $\sim L^N \approx L^{L/15}$ , получен в этой работе непосредственно и в тех же простых комбинаторных терминах, в которые сформулирован парадокс Левинталя, в отличие от предыдущих работ [2,14–16], где для решения этого пара-

докса были использованы вспомогательные энергетические оценки.

Работа выполнена при финансовой поддержке Российского научного фонда (грант № 14-24-00157).

#### СПИСОК ЛИТЕРАТУРЫ

1. О. Б. Птицын, ДАН СССР **210**, 1213 (1973).
2. G. C. Rollins and K. A. Dill, J. Am. Chem. Soc. **136**, 11420 (2014).
3. C. Levinthal, in *Biological Systems. Proc. of Meeting Held at Allerton House* (Monticello, Illinois, 1969), pp. 22–24.
4. A. V. Finkelstein and S. O. Garbuzynskiy, Chem-PhysChem (in press).
5. C. Chothia, Annu. Rev. Biochem. **53**, 537 (1984).
6. A. G. Murzin and A. V. Finkelstein, J. Mol. Biol. **204**, 749 (1988).
7. A. V. Finkelstein and O. B. Ptitsyn, Progr. Biophys. Mol. Biol. **50**, 171 (1987).
8. C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland Publ., Inc. New York, 1999).
9. А. В. Финкельштейн, *Физика белковых молекул* (АНО «Ижевский институт компьютерных исследований», М.–Ижевск, 2014).
10. А. В. Финкельштейн и О. Б. Птицын, *Физика белка* (Книжный Дом «Университет», М., 2002).
11. F. H. C. Crick, Acta Crystallogr. **6**, 689 (1953).
12. D. T. Jones, J. Mol. Biol. **292**, 195 (1999).
13. A. V. Finkelstein, A. M. Gutin, A. Ya. Badretdinov, in *Subcellular Biochemistry. Proteins: Structure, Function and Protein Engineering*, Ed. by B. B. Biswas and S. Roy (Plenum Press, New York, 1995), vol. 24, pp. 1–26.
14. A. V. Finkelstein and A. Ya. Badretdinov, Fold. Des. **2**, 115 (1997).
15. А. В. Финкельштейн и А. Я. Бадретдинов, Молекуляр. биология **31**, 469 (1997).
16. S. O. Garbuzynskiy, D. N. Ivankov, N. S. Bogatyreva, and A. V. Finkelstein, Proc. Natl. Acad. Sci. USA **110**, 147 (2013).
17. S. Mukherjee, P. Chowdhury, M. R. Bunagan, and F. Gai, J. Phys. Chem. B **112**, 9146 (2008).
18. V. Mucoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton, Nature **390**, 196 (1997).
19. P. E. Leopold, M. Montal, and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **89**, 8721 (1992).
20. P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
21. N. S. Bogatyreva and A. V. Finkelstein, Protein Eng. **14**, 521 (2001).
22. R. Zana, Biopolymers **14**, 2425 (1975).

## **Solution of Levinthal's Paradox Is Possible at a Level of Formation and Assembly of Protein Secondary Structures**

**A.V. Finkelstein and S.A. Garbuzynskiy**

*Institute of Protein Research, Russian Academy of Sciences,  
ul. Institutskaya 4, Pushchino, Moscow Region, 142290 Russia*

The complete volume of the protein conformation space is, by many orders of magnitude, smaller at the level of secondary structure elements than that at the level of amino acid residue. The latter, according to Levinthal's estimate, scales up as  $\sim 10^{2L}$  with  $L$  being the number of residues in the chain, while the former, at the level of secondary structures, scales no faster than  $\sim L^N$  with  $N$  being the number of the secondary structure elements.  $N$  is approximately  $L/15$  according to statistics of protein structures. This drastic decrease in the exponent (down to  $L/15$  instead of  $2L$ ) considerably reduces sampling space, and explains why sampling of the conformation space at the level of secondary structure elements does not stop the ability of the protein chain to find its most stable fold.

*Key words: assembly of secondary structures, formation of secondary structures, volume of configuration space, protein folding, protein folding rate, protein structures*