

50+ ЛЕТ САМООРГАНИЗАЦИИ БЕЛКОВ

©2018 г.

А. В. ФИНКЕЛЬШТЕЙН

Институт белка РАН, Пущино, Московская область

I. Введение. II. Исторический экскурс. III. Физические теории и их результаты. IV. Заключение.

I. ВВЕДЕНИЕ

Самоорганизация белка – физический процесс, при котором белковая цепь «сама собой» приобретает свою нативную (биологически функциональную) трехмерную структуру.

Явление спонтанной самоорганизации было открыто в группе Анфинсена в 1961 году [1] на примере спонтанного восстановления биохимической активности и «правильных» S–S-связей *in vitro* в бычьей рибонуклеазе А, – после её полного, с разрывом всех S–S-связей, разворачивания денатурантом – и последующего помещения её в «нативные» условия. Это открытие было впоследствии подтверждено на множестве других белков, не подвергшихся существенной посттрансляционной модификации [2, 3].

В живой клетке белок синтезируется на рибосоме и «созревает под опекой» специальных белков-шаперонов. Биохимический синтез белковой цепи занимает порядка секунд или минут, да и все изготовление «готового», свернутого белка, занимает примерно столько же – эксперимент не видит разницы [2–4].

Естественно было предположить, что сворачивание белка *in vivo* может начинаться еще на рибосоме, еще до окончания синтеза всей белковой цепи. По-видимому, так и обстоит дело в крупных, многодоменных белках. Так, люцифераза (с цепью из ~540 аминокислотных остатков, свернутой в два, как минимум, домена) активна сразу после биосинтеза [4]. Однако сворачивание столь большого белка *in vitro* может протекать за биологически-разумное время (минуты; см. рис.

Адрес для корреспонденции: afinkel@vega.protres.ru

Первая часть этой работы была поддержана грантами Медицинского Института Ховарда Хьюза и программой «Молекулярная и клеточная биология» РАН (гранты № 01200957492, 01201358029), а вторая ее часть – грантом № 14-24-00157 Российского научного фонда.

5, 9 ниже) только при «подоменном» образовании его структуры – и такому «подоменному» сворачиванию может способствовать постепенный выход белковой цепи из рибосомы. Известно также, что сравнительно небольшая (из ~150 аминокислотных остатков) цепь глобина способна связывать свой лиганд (гем) когда рибосома синтезировала только немного более её половины [5]. Эти и другие аналогичные факты и привели к предположению, что котрансляционное (и шаперон-зависимое) сворачивание белковой цепи *in vivo* очень сильно отличается от её сворачивания *in vitro*.

Однако принципиального различия котрансляционного сворачивания *in vivo* от ренатурации *in vitro* не наблюдается в малых однодоменных белках. Согласно недавним работам [6–8], в таких белках (которые, будучи мечеными ^{15}N и ^{13}C изотопами, видны ядерному магнитному резонансу на фоне рибосом и прочей клеточной машинерии), «полипептиды [на рибосоме] остаются неструктурированными во время элонгации, но сворачиваются в компактную, нативоподобную структуру, когда появляется вся цепь» [6, 7], и «... ко-трансляционное сворачивание ... идёт через компактное, но не нативное состояние [т.е., видимо, через что-то вроде расплавленной глобулы – АФ], ... перестраиваясь в нативоподобную структуру сразу же после того, как полная последовательность домена появилась из рибосомы» [8] – то есть не до конца синтезированная рибосомой цепь однодоменного белка *in vivo* ведет себя так же, как укороченная на несколько С-концевых аминокислот цепь *in vitro*: она не образует определенной трехмерной структуры [9].

Из котрансляционного поведения небольших белков можно извлечь два вывода.

Во-первых, – что оно (как и поведение связывающей гем N-концевой половины цепи глобина, см. выше) напоминает поведение «нативно-неупорядоченных» белков [10, 11], многие из которых имеют вид расплавленной глобулы, и обретают определённую пространственную структуру, только дополнительно связавшись с лигандом.

Во-вторых, – что и *in vivo*, и *in vitro* нативные структуры появляются только в полных аминокислотных последовательностях белков (или белковых доменов, цепей которых обычно достаточно для образования их правильной нативной структуры [12]).

То же самое отсутствие принципиального различия в сворачивании относится к влиянию на него шаперонов (основной задачей которых является борьба с последствиями агрегации белков в густом «клеточном супе» [13]). Открытие шаперонов навело на мысль, что они могут обладать «структурообразующей» каталитической актив-

ностью (см., например, [14] и цитируемую там литературу), и потому образование структуры белка может идти *in vivo* и *in vitro* совсем по-разному. Однако анализ данных, представленных в работе [14], показал, что наиболее изученный шаперон, GroEL, не убыстряет сворачивание белка [15]; скорее, эти данные подтверждают сделанный ранее [16, 17] вывод, что GroEL служит временной «ловушкой», которая связывает имеющиеся в избытке сворачивающиеся белковые цепи, и тем самым предотвращает их необратимую агрегацию.

Итак: открытие Анфинсеном спонтанной самоорганизации *in vitro* [1, 18] (впоследствии подтвержденное на множестве других белков), отсутствие котрансляционного образования нативных структур в не полностью синтезированных белках [6–8], а также возможность чисто химического синтеза белковой цепи, спонтанно сворачивающейся в активный белок (опыты Меррифилда и др. [19]) – всё это позволяет, в первом приближении, отделить спонтанное образование структуры белка (по крайней мере – однодоменного) от процесса его биосинтеза.

И данная статья сфокусирована именно на спонтанном образовании структуры однодоменного глобулярного белка – т.е. на его самоорганизации *in vitro*.

II. ИСТОРИЧЕСКИЙ ЭКСКУРС

Способность белковой цепи спонтанно свернуться в сложную пространственную структуру долго озадачивала исследователей: ведь эта цепь должна найти свою нативную (и, судя по тому, что находит она её и *in vitro*, и при биосинтезе, т.е. из разных стартовых условий – наиболее стабильную) структуру среди несметного числа других – в течение тех нескольких минут или секунд, что отведены ей на это биологией.

А число это действительно огромно [20, 21]: по крайней мере 2^{100} структур, но скорее 3^{100} или 10^{100} или даже 100^{100} для цепи из 100 остатков, – так как у каждого остатка цепи есть не менее 2 («верная» и «неверная»), но скорее 3 (α , β , «клубок») или 10 [22] или даже $(10_{\text{ по } \varphi}) \times (10_{\text{ по } \psi}) = 100$ конформаций [21]; так что их «тупой» перебор занял бы порядка $\sim 2^{100}$ или 3^{100} или 10^{100} или 100^{100} пикосекунд, кладя всего пикосекунду (время теплового колебания) на переход из одной конформации в другую, т.е. $\sim 10^{10}$ или 10^{25} или 10^{80} или даже 10^{180} лет. А достаточно полный перебор конформаций, считал Левинталь, может быть только «тупым», ведь белок может «почувствовать» стабильность конформации, только попав прямо в нее, так как отклонение даже на 1 \AA может сильно повысить энергию цепи в плотной белковой глобуле.

Этот «парадокс Левинталя» оказал сильнейшее влияние на структурных биологов и биофизиков (Е. Шахнович сравнил это с воздействием «Великой теоремы Ферма» на начинающих математиков). Более того: под влиянием гипотезы, что котрансляционное (и шаперон-зависимое) сворачивание белковой цепи *in vivo* очень сильно отличается от её сворачивания *in vitro*, некоторые предположили, что должно быть *два* решения этого парадокса: (1) для *in vitro* и (2) для *in vivo* сворачивания – и, может быть, еще одно, для сворачивания моделей белковых цепей *in silico* [В.К. Lee, выступление на семинаре в НИИ]!

В попытке ответить на свой парадокс, Левинталь предположил, что нативная структура белка определяется *не* стабильностью, т.е. *не* термодинамикой, а кинетикой, т.е. что белок следует по некоему специфическому, «быстрому» пути сворачивания, и его нативная укладка – просто конец этого пути, независимо от того, является ли она наиболее стабильной или нет. Другими словами, Левинталь предположил, что нативная структура соответствует не глобальному, а просто быстро достижимому минимуму свободной энергии цепи.

Однако компьютерные эксперименты с решеточными моделями белковых цепей убедительно показали, что цепи складываются в свою самую стабильную структуру, то есть, что «нативная структура модели белка» обладает самой низкой энергией, и, значит, сворачивание белка находится под термодинамическим, а не кинетическим контролем [23, 24].

Тем не менее, большинство существующих гипотез о сворачивании белка основано на «гипотезе кинетического контроля».

Филлипс [25] в 1966 г. (то есть, еще до Левинталя) предположил, что ядро сворачивания белка образуется на N-конце его растущей цепи, а остальная цепь обертывается вокруг этого ядра. Однако позже было показано, что успешное сворачивание в пробирке многих однодоменных белков и белковых доменов вовсе не начинается с их N-конца [26, 27].

Из личных воспоминаний: Статья Филлипса в *Scientific American* попала мне на глаза в Ленинке в том же 1966 г. Увидев там – впервые – картинку трёхмерной атомарной структуры белка, я, третьекурсник Физтеха, сказал себе: «Ну, уж с этим-то ужасом мне никогда не придется иметь дела»; и – ошибся...

Несколько лет спустя, Ветлауфер [28] выдвинул гипотезу, что ядро сворачивания состоит из близких друг к другу по цепи остатков. Однако потом эксперименты *in vitro* показали, что это далеко не всегда так [29].

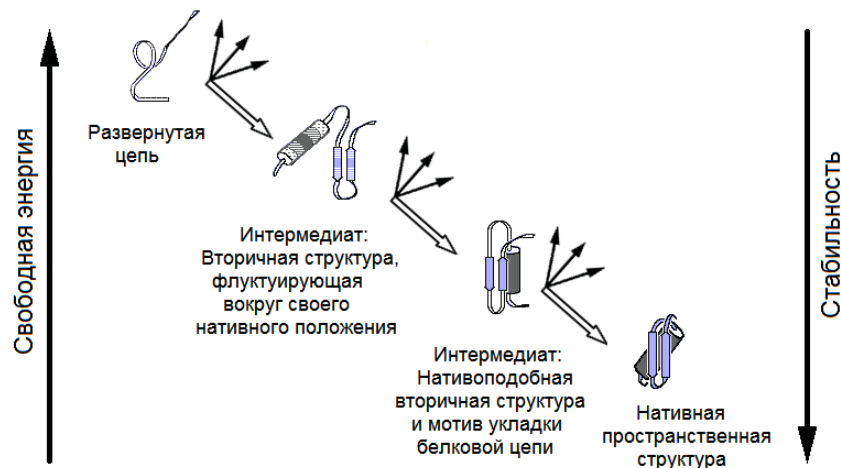


Рис. 1. Стадийная модель Птицына [30]. Выделены вторичные структуры – α -спирали (цилиндры) и β -участки (стрелки). Оба предсказанные интермедиата впоследствии были найдены на опыте и названы: первый – «предрасплавленной глобулой», а второй – «расплавленной глобулой» [31].

Одновременно, Птицын [30] предложил модель иерархического сворачивания (рис. 1) с постадийным вовлечением различных взаимодействий и постадийным формированием различных промежуточных состояний.

Согласно этой модели, сворачивание белка проходит в несколько этапов, на каждом из которых формируется наиболее стабильная (на данном этапе) форма молекулы, служащая исходной точкой для последующей стадии самоорганизации белка. В показанном на рисунке 1 примере, выбор одной – нативной – финальной структуры из, скажем, $\sim 2^{100}$ возможных структур сводится к трем стадиям. На каждой из них выбирается одна структура из «лишь» $\sim 2^{100/3}$ возможных, т.е. при стадийном механизме выбор финальной структуры осуществляется в $\sim 2^{100}/(3 \times 2^{100/3}) \sim 10^{20}$ (!) раз быстрее, чем при полном переборе.

Нельзя не заметить, однако, что это колоссальное ускорение достигается за счет необратимости выбора оптимального интермедиата на каждом этапе – и, значит, сильного (на много $k_B T$, где k_B – постоянная Больцмана, а T – температура) падения свободной энергии на каждой стадии. Поэтому предполагаемое Птицыным быстрое сворачивание белка может иметь место только тогда, когда нативное состояние белка несравненно (на много-много $k_B T$) стабильнее его развернутого

состояния. Но: такая «обязательная сверхстабильность» нативной структуры белка в корне противоречит исследованной Приваловым [22] термодинамике фазовых переходов в белковых молекулах...

Из личных воспоминаний: Я хорошо помню жаркие споры Птицына и Привалова о существовании интермедиатов сворачивания белка и их термодинамической и кинетической роли. Сейчас, по прошествии лет, ясно, что выведенные «на кончике пера» Птицыным в 1973 г. и отвергаемые тогда Приваловым интермедиаты сворачивания действительно существуют; они обнаружены в массе белков, – как в кинетических, так и в термодинамических опытах [31, 32]. Однако *не* подтвердилась основная идея Птицына о *необходимости* существования стабильных интермедиатов для самоорганизации белка. Во многих небольших белках стабильные интермедиаты сворачивания не наблюдаются вообще [33], а в более крупных белках они обычно наблюдаются тогда, когда нативное состояние много стабильнее денатурированного, т.е. вдали от точки термодинамического равновесия этих двух состояния (в которой, однако, самоорганизация белка также происходит, хоть и заметно медленнее) [33–35].

Итак, каждый из двух спорщиков был полностью прав в одном пункте – и неправ в другом...

Ближе к концу этой статьи мы рассмотрим модель Птицына более подробно и увидим, что на её основе – но при несколько другой её трактовке – действительно можно понять причину быстроты самоорганизации белка.

Заключая обзор предложенных подходов к решению загадки самоорганизации белков, надо отметить, что, начиная с 1990-х годов, популярными моделями для иллюстрации и разъяснения быстроты самоорганизации белков стали «воронки сворачивания» [36–39]. Однако и они не позволили оценить – хотя бы ориентировочно – потребные для спонтанного сворачивания различных белков времена. А согласно эксперименту (см. ниже) эти времена, для однодоменных глобулярных белков, лежат в диапазоне от микросекунд до часов.

Вообще, сложность проблемы самоорганизации белков – при практически бесконечном числе возможных их структур – заключается в том, что её нельзя решить чисто экспериментально. Действительно, предположим, что у белковой цепи есть другая, «ненативная», кинетически крайне медленно достижимая, но еще более стабильная укладка. Как её найти, если сам белок её не находит? Ждать результата в течение 10^{10} (или даже 10^{180}) лет?

С другой стороны, вопрос о том, что – кинетика или термодинамика – определяет укладку белковой цепи, постоянно возникает на пути решения разных прикладных задач. Он возникает, когда речь идет о предсказании структуры белка по его аминокислотной последовательности (надо знать, что предсказывать: самую стабильную

или самую быстро сворачивающуюся его структуру). Он возникает и тогда, когда речь идет о дизайне новых, не встречающихся в природе белков (надо знать, что делать: максимально усилить стабильность желаемой структуры или пролагать максимально быстрый путь к ней).

Однако действительно ли существует противоречие между «структурой стабильной» и «структурой быстро сворачивающейся»? Может быть, стабильная структура *автоматически* является фокусом «быстрых» путей сворачивания и потому *автоматически* обладает свойством сворачиваться быстро?

Прежде чем приступить к исследованию этих вопросов, т. е. прежде, чем рассматривать *кинетические* аспекты сворачивания белков, вспомним ряд уже хорошо изученных фундаментальных фактов из области их *термодинамики* (здесь всюду речь идет об относительно небольших, однодоменных белках, т. е. о белках из 50–200 аминокислотных остатков). Эти факты помогут нам понять, какие условия протекания процесса сворачивания мы должны рассматривать. Термодинамические факты таковы:

1) Денатурированная форма белков, во всяком случае, небольших белков, развернутых сильным и концентрированным денатурантом, часто является неупорядоченным клубком [40].

2) Разворачивание белка обратимо [18]; более того, денатурированное и нативное состояние могут находиться в кинетическом равновесии [41], и переход между этими состояниями происходит как переход «*всё или ничего*» [22]. Последнее означает, что в точке денатурации белка только две формы белковой молекулы – «нативная» и «денатурированная» – присутствуют в заметных количествах, а все прочие («полусвернутые» и «неверно свернутые» формы) практически отсутствуют. Такой переход возможен только в цепях, аминокислотная последовательность которых обеспечивает достаточно большую «энергетическую щель» между основной массой структур и самой стабильной из них [22, 42–45]; она обеспечивает надежность функционирования белка – по принципу «*всё или ничего*», как у электрической лампочки, – которая либо полностью годна, либо совсем нет.

3) В нормальных физиологических условиях нативная форма белка лишь немногим (на несколько ккал/моль [22]) стабильнее его денатурированной формы (а в средней точке перехода обе эти формы имеют, естественно, одинаковую стабильность); при этом нативная структура белка стабильна благодаря своей низкой энергии, т. е. благодаря сильным взаимодействиям в нативной структуре, а развернутая – благодаря своей высокой конформационной энтропии, т. е. благодаря огромному числу разных развернутых конформаций.

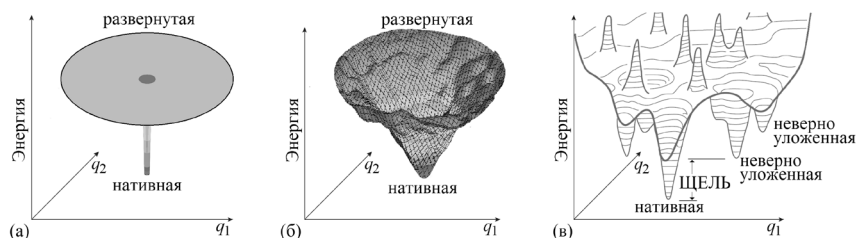


Рис. 2. Иллюстрация базовых моделей энергетического ландшафта белковой цепи: «поле для гольфа» Левинталя (а) и «воронка» Леопольда и др. (б); обе имеют своим центром самую низкоэнергетическую («нативную») структуру. (в) Более реалистичное изображение ухабистого энергетического ландшафта белковой цепи. Широкая (шириной в много $k_B T_{\text{melt}}$, где k_B – постоянная Больцмана, а T_{melt} – температура плавления белка) энергетическая щель между глобальным минимумом энергии и другими энергетическими минимумами необходима для обеспечения перехода «всё или ничего» при распаде стабильной структуры белка [22, 42, 43]. На иллюстрации можно изобразить только две координаты (q_1 и q_2), в то время как конформация цепи определяется сотнями координат.

Необходимое пояснение: как принято в литературе, термин «энергия» здесь означает, строго говоря, *всю свободную энергию взаимодействий*, в том числе взаимодействий цепи с растворителем (например, «энергия» гидрофобных взаимодействий определяется энтропией растворителя [40]). Термин же «энтропия» здесь охватывает лишь конформационную энтропию цепи, но не энтропию растворителя. Такая терминология принята, чтобы, оставив растворитель за скобками, сосредоточиться на главной проблеме – как белковая цепь находит «свою» пространственную структуру среди гигантского числа возможных.

Отмеченный выше переход «всё или ничего» означает, что нативное (N) и денатурированное (U) состояния разделены высоким барьером свободной энергии.

Именно высота этого барьера ограничивает скорость такого перехода, и именно эта высота должна быть оценена, чтобы решить парадокс Левинталя.

Для начала, однако, уместно выяснить, является ли «парадокс Левинталя» в самом деле парадоксом? Уже Брингельсон и Волинес [46] отметили, что этот «парадокс» основан на абсолютно плоской (и, следовательно, нереалистичной) модели «поля для гольфа» для описания поверхности потенциальной энергии белка (рис. 2а).

Несколько позже Леопольд и др. [36], следуя идеям Го и Абе [47], рассмотрели более реалистичную модель энергетической поверхности (с наклоном к нативной структуре белка), и ввели «воронки сво-

рачивания» (рис. 2б), которые, казалось бы, ликвидируют «парадокс Левинталя».

Однако не все так просто...

Проблема огромного времени поиска самой стабильной структуры существует: было математически доказано, что, несмотря на воронки и т.д., поиск такой структуры является так называемой «NP-трудной задачей» [48, 49], которая, грубо говоря, требует экспоненциально-большого времени для своего решения (и укладываемойся цепью, и человеком).

Так или иначе, различные модели «воронок» стали популярным средством объяснения и иллюстрации сворачивания белка [37, 50, 51]. Лежащая в центре воронки структура с наименьшей энергией (которая формируется самыми мощными взаимодействиями в цепи) окружена структурами с более высокими энергиями, содержащими только часть этих взаимодействий. «Энергетическая воронка» направляет движение в сторону структуры с наименьшей энергией, – что, кажется, должно помочь белковой цепи избежать «Левинталева перебора» всех конформаций.

Тем не менее, можно показать, что энергетические воронки сами по себе *не* решают парадокс Левинталя. Анализ [52] строго сформулированных моделей воронок [39, 53] показывает, что в точке равновесия между свернутой и развернутой формами цепи эти модели не могут одновременно объяснить обе основные особенности, наблюдаемые при сворачивании белка: (1) не-астрономическое время сворачивания, и (2) переход «всё или ничего», то есть сосуществование нативных и развернутых форм белковых молекул в процессе сворачивания.

Кстати, как уже было отмечено выше, стадийный механизм сворачивания белка [30], взятый сам по себе, также не может [54] одновременно объяснить обе эти главные особенности, наблюдаемые при сворачивании белка.

Таким образом, ни стадийный механизм, ни «воронки» не могут решить проблему Левинталя, хотя они дают намек на то, что может ускорить сворачивание белков.

Принципиальное решение парадокса обеспечивается особым типом воронок, учитывающих разделение развернутой и нативной фаз в сворачивающейся цепи [55, 56] (см. также обзор [57]).

Этому решению будет посвящена следующая часть нашего обзора.

III. ФИЗИЧЕСКИЕ ТЕОРИИ И ИХ РЕЗУЛЬТАТЫ

ФИЗИЧЕСКАЯ ОЦЕНКА ВЫСОТЫ СВОБОДНО-ЭНЕРГЕТИЧЕСКОГО БАРЬЕРА, РАЗДЕЛЯЮЩЕГО НАТИВНОЕ И РАЗВЕРНУТОЕ СОСТОЯНИЕ ЦЕПИ: ВИД НА БАРЬЕР СО СТОРОНЫ НАТИВНОГО СОСТОЯНИЯ

Чтобы разрешить «парадокс Левинталя» и показать, что самую стабильную структуру белковой цепи можно (или нельзя?) найти за разумное время, мы можем в первом приближении рассматривать только скорость перехода «*все или ничего*» из клубка в самую стабильную структуру цепи. При этом нам достаточно рассмотреть случай, когда самая стабильная укладка цепи так же стабильна, как клубок (или лишь чуть-чуть более стабильна, чем он), а все прочие формы белковой цепи термодинамически нестабильны. При этом рассматривать сворачивание белка наиболее просто, так как здесь нет стабильных интермедиатов сворачивания: они появляются лишь тогда, когда нативная структура становится много стабильнее клубка; при этом сворачивание белка достигает максимальной скорости [33], но его анализ усложняется. Поэтому мы для начала ограничимся окрестностью точки равновесия нативной структуры с клубком, где белок сворачивается пусть не максимально быстро, но максимально просто.

Из личных воспоминаний: Я хорошо помню, что, обсуждая загадку самоорганизации белков в 1970-х, 80-х и даже 90-х годах, почти все мы, так или иначе, сосредотачивались в основном на тех условиях среды, в которых этот процесс протекал наиболее быстро (считалось, что так имитируются «физиологические», внутриклеточные условия); см., например, [30, 31, 33, 58]. С точки зрения биологии, такой особый интерес к «физиологическим» условиям был вполне оправдан. Однако основной физический вопрос – как белок ухитряется найти свою структуру за не-астрономическое время? – при этом затенялся массой второстепенных (как я понимаю теперь) для ответа на этот вопрос подробностей, типа вопроса о судьбе метастабильных интермедиатов сворачивания (которого просто не существует, когда сворачивание белка рассматривается вблизи точки термодинамического равновесия его нативной структуры и клубка), и всё внимание должно быть сосредоточено на переходном состоянии [33]...

Не могу не отметить, что пристальное внимание к точке равновесия нативной структуры с клубком отличает развиваемый нами сейчас подход от тех, что превалировали с 1960-х до середины 1990-х годов.

Так как переход «все или ничего» требует большой энергетической щели между наиболее стабильной структурой и остальными [22, 42–45] (рис. 2в), мы будем полагать, что рассматриваемая аминокислотная последовательность обеспечивает такую щель. Наша цель состоит в том, чтобы оценить скорость перехода «все или ничего»

и доказать (если это возможно), что наиболее стабильная структура белка или домена нормального размера (из ~100 аминокислотных остатков) может возникать в течение нескольких секунд или минут.

Для того чтобы доказать, что самая стабильная белковая структура должна сворачиваться быстро, достаточно показать, что к ней всегда ведет по крайней мере один «быстрый» путь сворачивания. Наличие не одного, а многих путей реакции может только ускорить процесс...

А вот путей сворачивания не нативной, а прочих (и значит, при наличии «щели» – высокоэнергетических) структур мы можем не рассматривать вовсе! Они не могут сбить нашу цепь с пути истинного. Ведь вблизи точки перехода типа «*все или ничего*» между самой стабильной структурой цепи и клубком *никакие* «полу-» и «неверно свернутые» состояния не могут служить ловушками: они не могут «впитать» сворачивающиеся цепи просто в силу малости своей суммарной стабильности. Полезная аналогия здесь – просачивание воды через трещины в стенке, разделяющей два бассейна: если «емкость» трещин мала, т. е. они неспособны впитать всю воду, то каждая новая трещина в стенке может только ускорить наполнение второго бассейна, так что, рассмотрев просачивание воды через одну трещину, мы оценим минимальную скорость его наполнения.

Чтобы путь сворачивания был быстр, каждый шаг на этом пути должен проходиться быстро, таких шагов должно быть не слишком много, и – главное! – этот путь не должен преграждаться «барьером» в виде очень высокой свободной энергии на одной из стадий сворачивания.

Так как время фиксации одного звена мало (наносекунды, судя по измеренной скорости роста α -спиралей в полипептидных цепях [59]), то белок, фиксируя одно свое звено за другим, сворачивался бы мгновенно (100-звенная цепь – не более чем за ~1000 нс), *если бы* при этом он не должен был преодолевать свободно-энергетический барьер. Сворачивание белка занимает несколько секунд или минут, а не микросекунды, из-за свободно-энергетического барьера: большая часть времени тратится на восхождение на этот барьер и падение назад, а не на движение по пути сворачивания.

Ключевую роль в этом процессе принадлежит «переходному состоянию», т.е. наименее стабильному («барьерному») состоянию на пути реакции. Согласно классической теории переходных состояний [60–62], время многоэтапного процесса преодоления барьера оценивается как

$$\text{ВРЕМЯ} \sim \tau \times \exp(+\Delta F^\ddagger/k_B T) \quad (1)$$

где τ – характерное время одного шага процесса (обычно – порядка наносекунды), а ΔF^\ddagger – высота свободно-энергетического барьера.

Что касается ΔF^\ddagger , то главный вопрос, на который надо ответить: высок ли барьер ΔF^\ddagger на пути, ведущем к самому стабильному состоянию белковой цепи?

Сворачивание белковой цепи ведет к падению как ее энтропии (из-за роста упорядоченности цепи), так и энергии (из-за образования контактов между сближающимися звеньями цепи). Падение энергии понижает, а падение энтропии повышает свободную энергию цепи.

Если, по ходу сворачивания, цепь должна очень близко подойти к своей финальной структуре перед тем, как начнут возникать стабилизирующие эту структуру контакты (т. е. цепь должна потерять почти всю свою энтропию *перед* тем, как начнет выигрываться энергия), то повышение свободной энергии на первом этапе сворачивания будет пропорциональным числу звеньев в цепи, т. е. очень большим, а сворачивание цепи – страшно медленным. Именно такая картина (проигрыш всей энтропии *до* начала выигрыша энергии) лежит в основе парадокса Левинталя, утверждающего, что белковая цепь никак не может, даже за время жизни Вселенной, найти свою самую стабильную структуру.

Напротив, если путь сворачивания таков, что по ходу сворачивания падение энтропии практически тут же компенсируется падением энергии [47], то он не перекрыт высоким свободно-энергетическим барьером, и сворачивание идет довольно быстро. Именно такая картина, как мы убедимся, и имеет место.

Рассмотрим изменение энергии ΔE , энтропии ΔS и результирующей свободной энергии $\Delta F = \Delta E - T\Delta S$ по ходу изображенного на рис. 3 *последовательного* сворачивания белка (или, как говорят, сворачивания по нуклеационному [26] механизму). На каждом шаге этого пути одно звено цепи извлекается из клубка и занимает то положение, в котором оно находится в финальной (самой стабильной) структуре глобулы.

Такой процесс может показаться несколько искусственным (откуда звено знает свое место в финальной структуре?). Но это впечатление исчезает, если заметить, что мы, таким образом, просто просматриваем «кино» распада стабильной структуры белка в обратном направлении, и вспомнить, что, в соответствии с хорошо известным в физике принципом детального равновесия [63], прямая и обратная реакции следуют по одному тому же пути и имеют одинаковую скорость, когда оба конечных состояния имеют равную стабильность.

Таким образом, можно использовать принцип детального равновесия, чтобы найти переходное состояние для сворачивания путем нахождения оптимального переходного состояния для разворачива-

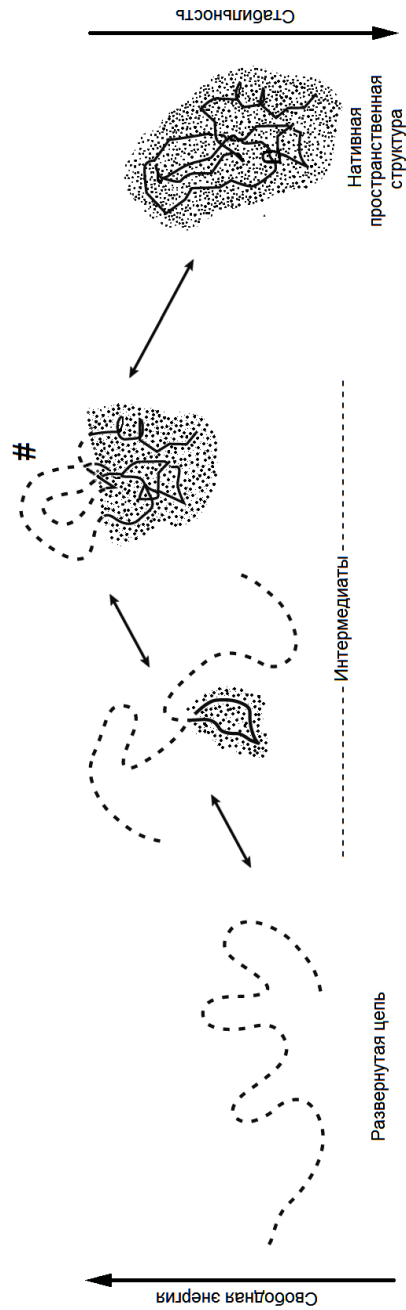


Рис. 3. Один из возможных путей последовательного сворачивания белка с разделением развернутой и глобулярной фаз [56]. Свободная энергия интермедиатов сворачивания повышена из-за имеющейся в них поверхности раздела фаз. Значком # отмечено максимально нестабильное (переходное) состояние. Выделенная точками область соответствует уже обретшей свою финальную конформацию части белковой глобулы. Сплошной линией выделена главная цепь в ней, а боковые группы не показаны для упрощения рисунка. Пунктирная линия показывает все еще развернутую цепь. Глобулярная часть «полуразвернутых» структур, лежащих на «оптимальном» (подходящем через интермедиаты с невысокой свободной энергией) пути, должна быть компактной, т.е. имеющей минимальную границу между свернутой и развернутой фазами.

ния белка. Преимуществом анализа пути разворачивания является то, что он гораздо проще: для любой глобулярной структуры можно легко найти путь последовательного разворачивания, проходящего через наименее нестабильные, т.е. отличающиеся наименьшим по площади разделением глобулярной и «развернутой» фаз компактные полуразвернутые состояния (рис. 3) [55, 56, 64–66].

Из личных воспоминаний:

1) Насколько я помню, разворачивание белка, в отличие от сворачивания, никогда не считалось «парадоксальным», хотя было давно и хорошо известно, что нативное состояние может находиться в кинетическом равновесии с развернутым. По-моему, никто никогда не задавал вопроса, дополнительного к вопросу Левинталя, – как белок может, опять-таки за «неастрономическое» время получить ту огромную энергию, что нужна для его разворачивания... Отсутствие самого такого вопроса показывает, насколько легче представить себе, как развернуть любую белковую структуру, чем как свернуть ее...

2) С проблемой разделения фаз в белке мы впервые столкнулись, выясняя причины фазового («всё–или–ничего») перехода при плавлении белка [67]. Однако тогда, к сожалению, мы не догадались исследовать влияние этого эффекта не только на фазовый характер, но и на кинетику этого перехода – иначе парадокс Левинталя был бы решен на 15 лет раньше...

Рассматривая путь последовательного сворачивания (реконструированного из пути последовательного разворачивания) – видим, что, по мере роста финальной глобулы, в ней одно за другим восстанавливаются взаимодействия, стабилизирующие финальную структуру.

Если растущая структура все время остается более или менее компактной, как на рис. 3, 4а (именно такого типа оптимальные пути нас должны интересовать), то число этих взаимодействий будет расти (а их энергия, соответственно, будет падать) почти пропорционально числу n фиксируемых в глобуле звеньев (рис. 4в).

Правда, в начале сворачивания падение энергии несколько замедлено, так как прилипание звена к поверхности маленькой глобулы дает в среднем меньше контактов, чем прилипание к поверхности большой. В результате возникает нелинейный поверхностный (т.е. пропорциональный $\approx n^{2/3}$) член в энергии ΔE растущей глобулы. Поэтому максимальное отклонение от линейного падения энергии составляет величину $\sim L^{2/3}$ (где L – число звеньев в цепи) на пути, идущем через компактные интермедиаты, рис. 4а [55, 56, 65] (и больше – если интермедиаты не компактны, рис. 4б). Это отклонение, очевидно, мало по сравнению с полным падением энергии при сворачивании цепи, которое составляет величину порядка L .

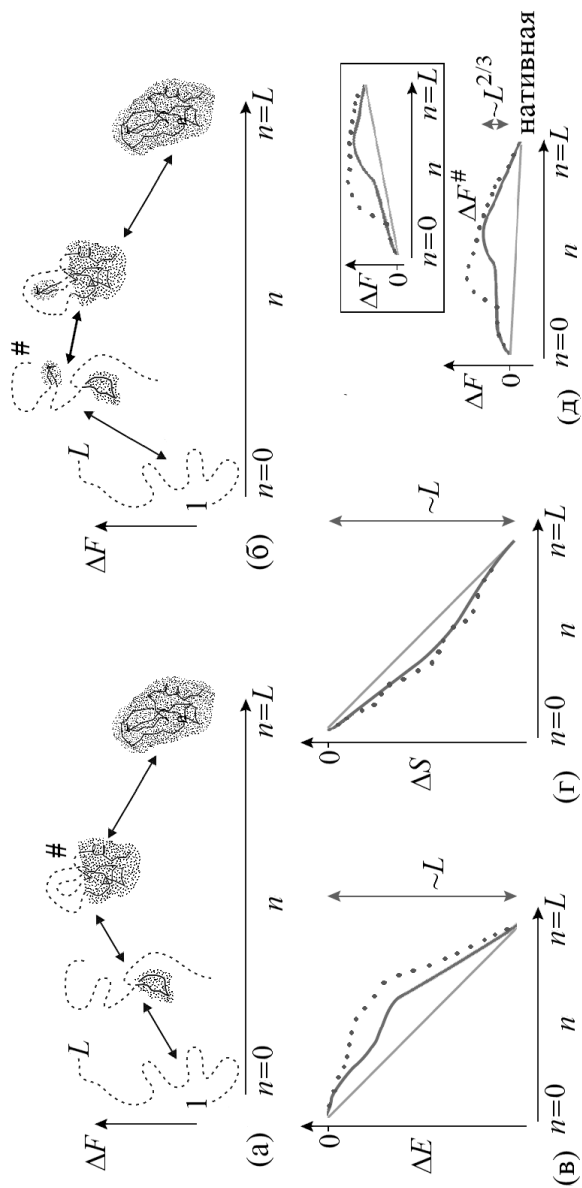


Рис. 4. Последовательное сворачивание/разворачивание с компактными (а) и не-компактными (б) интермедиатами, и изменение энергии (в), энтропии (г) и свободной энергии (д) вдоль этих путей вблизи точки термодинамического равновесия между клубком ($n=0$) и конечной структурой ($n=L$): все L звеньев цепи уложены в глобулу). Полное изменение энергии, $\Delta E(L)$, и энтропии, $\Delta S(L)$, примерно пропорциональны L . Жирные прямые на (в) – (г) показывают линейную (пропорционально числу n уже уложенных звеньев) часть $\Delta E(n)$ и $\Delta S(n)$. Нелинейные части $\Delta E(n)$ и $\Delta S(n)$ связаны, главным образом, с поверхностью глобулярной части молекулы (сплошные линии – для пути с компактными интермедиатами; пунктирные – для пути с некомпактными). Максимальные отклонения $\Delta E(n)$ и $\Delta S(n)$ от линейных зависимостей пропорциональны $L^{2/3}$. В результате $\Delta F(n) = \Delta E(n) - T\Delta S(n)$ также отклоняется от линейной зависимости (жирной прямой линии) на величину $\sim L^{2/3}$ для компактных интермедиатов (и больше – для некомпактных). Таким образом, в точке равновесия (где $\Delta F(n=0) = \Delta F(L)$), максимальный избыток (“барьер”) $\Delta F^{\#}$ на пути с компактными интермедиатами также пропорционален лишь $L^{2/3}$. Изменение $\Delta F(n)$ на пути к другим структурам выглядит примерно так же (см. врезку в (д)), но этими путями можно пренебречь, т.к. все эти структуры нестабильны, имеют $\Delta F(n=0) < \Delta F(L)$ при наличии энергетической щели между самой стабильной и прочими глобулами, и перехода «все или ничего» между развернутым и самым стабильным глобулярным состоянием цепи. Взято из [56].

С ростом глобулы *энтропия* цепи тоже падает примерно пропорционально числу фиксированных в глобуле звеньев (рис. 4г). Правда, в начале сворачивания энтропия может падать несколько быстрее из-за образования замкнутых петель, торчащих из растущей глобулы (рис. 3, 4а, 4б).

В результате возникает нелинейный (поверхностный) член в энтропии ΔS этой растущей глобулы, составляющий (так же, как и поверхностный член в ΔE) величину порядка $L^{2/3}$ [55, 56]. (Точнее, $\sim L^{2/3} \ln(L^{1/3})$, так как развернутая петля, исходящей из поверхности наполовину свернутой глобулы, имеет длину $\sim L^{1/3}$, и энтропия петли такой длины составляет $\sim \ln(L^{1/3})$ [68, 69], а $\ln(L^{1/3}) \sim 1$ при $L \sim 100$; более тщательно сделанную оценку см. ниже, после формулы (3). Это падение (см. [55, 56] и более поздние строго-математические статьи [70, 71]) существенно меньше, чем само падение энтропии, составляющее (как и падение энергии) величину порядка L .

И линейные, и поверхностные компоненты ΔS , ΔE входят в свободную энергию $\Delta F = \Delta E - T\Delta S$ растущей (или разворачивающейся) глобулы. Когда завершена глобула находится в термодинамическом равновесии с клубком, большие линейные члены взаимно аннигилируют в разности $\Delta E - T\Delta S$ (так как $\Delta F(n=0) = \Delta F(L)$), и остаются только поверхностные члены: если бы не они, $\Delta F(n)$ была бы равна нулю по всей длине пути.

Таким образом, свободно-энергетический барьер (см. рис. 4д) на пути последовательного сворачивания с компактными интермедиами связан только с относительно небольшими эффектами, связанными с поверхностью глобулярных интермедиатов, а его высота пропорциональна не L (как подразумевает оценка Левинталя), а лишь $L^{2/3}$.

В упрощенной форме (детальные выкладки см. в [55, 56, 65, 72]), свободная энергия барьера оценивается следующим образом.

Самый быстрый путь сворачивания – тот, что имеет самый низкий свободно-энергетический барьер. Барьер, на каждом данном пути, соответствует интермедиату с самой высокой свободной энергией, то есть, имеющему максимальный, на этом пути, интерфейс свернутой и развернутой фаз. Для компактных интермедиатов, такой интерфейс содержит около $L^{2/3}$ остатков.

Энергетическая составляющая $\Delta E^\#$ свободной энергии барьера возникает из-за взаимодействий, утраченных находящимися на интерфейсе остатками; она равна примерно

$$L^{2/3} \cdot 1/4 \varepsilon, \quad (2)$$

где $\varepsilon \approx 1.3 \text{ ккал/моль} \approx 2k_B T_{\text{melt}}$ – средняя теплота плавления аминокислотного остатка в белке [22] (это – первый эмпирический параметр,

используемый теорией), и $\approx 1/4$ – доля взаимодействий, в среднем теряемая остатком на интерфейсе. Таким образом,

$$\Delta E^\# / k_B T_{\text{melt}} \approx 0.5 L^{2/3}. \quad (2a)$$

Энтропийная составляющая $\Delta S^\#$ свободной энергии барьера связана с потерей энтропии замкнутыми петлями, выступающими из глобулярной фазы в развернутую (см. рис. 3, 4а, 4б).

Верхний предел $\Delta S^\#$ равен нулю (если интерфейс не содержит таких петель).

Нижний предел $\Delta S^\#$ – около

$$(\Delta S^\#)_{\text{lower}} = 1/6 L^{2/3} \cdot [-5/2 k_B \ln(3L^{1/3})], \quad (3)$$

где $1/6 L^{2/3}$ – максимальное число замкнутых петель на оптимальном (минимально покрытом петлями) интерфейсе глобула/клубок. На самом деле, это – среднее число для одного сечения глобулы (рис. 3, 4а), т. к. остаток интерфейса может иметь 6 направлений: 4 вдоль поверхности, 1 внутрь, и только 1 – наружу, а используемый для сворачивания интерфейс должен быть покрыт минимальным, то есть не более, чем это среднее, числом петель. $3L^{1/3} \equiv (L/2) / (1/6 L^{2/3})$ – среднее число остатков в такой петле (оно равно числу развернутых остатков, разделенному на число петель), а $-5/2 k_B \ln(3L^{1/3})$ – энтропия, потерянная такой замкнутой петлей (внутренние части которой не проникают внутрь глобулы, что меняет обычный коэффициент Флори, $3/2$ на $5/2$ [55, 56]). При $L \sim 100$ (на самом деле, это приближение хорошо для всего диапазона $L = 10-1000$),

$$(\Delta S^\#)_{\text{lower}} \approx -k_B L^{2/3}. \quad (3a)$$

В результате, время и образования, и разворачивания самой стабильной структуры цепи растет с числом остатков цепи L не «по Левинталю» (то есть, не как 2^L , или 10^L , или любая экспонента от L), – а, в условиях середины перехода, как

$$ВРЕМЯ \sim \tau \times \exp[(1 \pm 0.5)L^{2/3}] \quad (4)$$

где $\tau \approx 10$ нс [59] (это – второй и последний эмпирический параметр теории).

Оцененное таким образом время зависит как от размера, так и от формы (определяющей, через $\Delta S^\#$, фактор 1 ± 0.5 ; см. выше) нативной структуры белка.

Физическая причина этой «не-Левинталевой» оценки – (1) то, что уменьшение энтропии при сворачивании почти сразу и почти полностью компенсируется за счет уменьшения энергии вдоль пути последовательного сворачивания (и, точно так же, увеличение энергии

почти сразу же и почти полностью компенсируется увеличением энтропии вдоль того же пути последовательного разворачивания), и (2) то, что свободная энергия барьера возникает из-за увеличения свободной энергии, связанного только с поверхностными эффектами, которые относительно слабы.

Наблюдаемые времена сворачивания белка охватывают (рис. 5) 11 порядков величины (различие, как между продолжительностью жизни комара и возрастом Вселенной).

В середине перехода (при $\Delta F = 0$) эти времена действительно (рис. 5) находятся в теоретически очерченном диапазоне $10 \text{ нс} \times \exp(0.5L^{2/3})$ — $10 \text{ нс} \times \exp(1.5L^{2/3})$. При «более физиологических» условиях («в воде», где $\Delta F < 0$), $L^{2/3}$ заменяется на $L^{2/3} + 0.4\Delta F/k_B T$ [65] (см. Обсуждение), но во всех остальных отношениях диапазон остается тем же.

Полученная формула (4) и рис. 5 показывают, что цепь из $L \approx 80$ – 90 остатков всегда найдет свою самую стабильную структуру за минуты даже при «небиологических» условиях середины перехода, где, как известно [33, 41], сворачивание происходит наиболее медленно. Значит, нативные структуры таких небольших белков находятся под термодинамическим контролем: они – наиболее стабильные среди всех структур таких цепей. Нативные структуры больших белков (из 90–400 остатков) находятся под дополнительным «структурным» контролем в том смысле, что слишком «запутанные» укладки таких длинных цепей не могут быть достигнуты в течение дней или недель, даже если они термодинамически устойчивы; и действительно, сильно запутанные укладки длинных белковых цепей никогда не наблюдаются [65] – видимо, они исключены из репертуара белковых структур. Это объясняет также, почему крупные белки должны быть либо далеко не сферичными, либо состоять из отдельных доменов: в противном случае, цепи из более чем 400 остатков уложатся слишком медленно для работы в клетке. Это – «структурный контроль», действие которого напоминает действие «кинетического контроля» Левинтала, хотя и на другом уровне и только для больших белков. Приведенные выше оценки (80–90 и ≈ 400 остатков) несколько увеличиваются, если свободная энергия ΔF нативной структуры меньше, чем у развернутой цепи (см. ниже), но, по существу, остаются теми же [65].

И еще одно. «Квази-Левинталев» поиск правильного образования узлов (рис. 6) может, в принципе, ограничивать скорость сворачивания, так как узел не может быть изменен без распада глобулярной части. Однако, как показывают компьютерные эксперименты, один узел

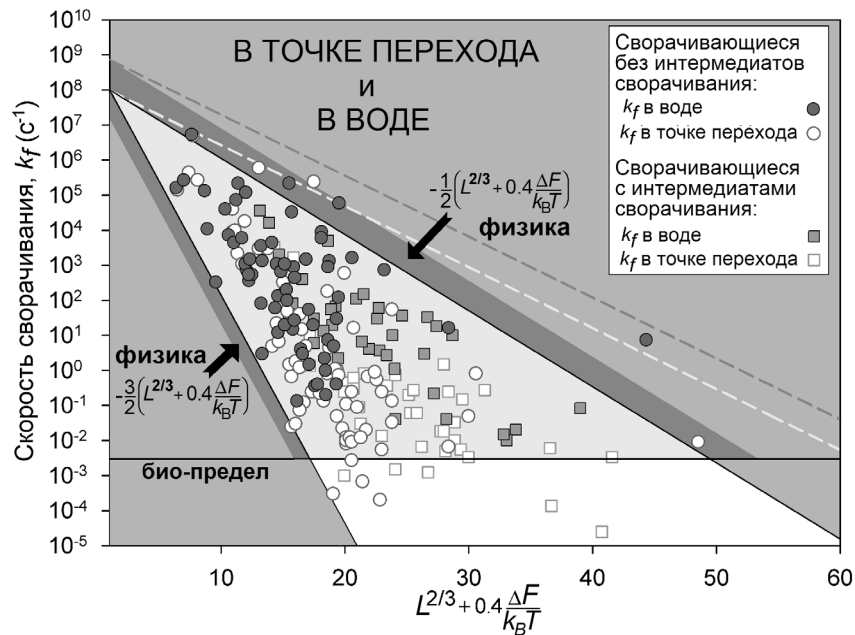


Рис. 5. Экспериментально измеренные *in vitro* константы скорости сворачивания белков в воде (в приблизительно «биологических» условиях) и в середине перехода (в присутствии небольших количеств денатуранта) для 107 однодоменных глобулярных белков (или отдельных доменов) без SS связей и ковалентно связанных лигандов (но скорости сворачивания белков с SS связями и без них не различаются принципиально [73]). Треугольник: разрешенная физикой область; его светло-серая часть (с более темной лентой) соответствует биологически приемлемому времени сворачивания (≤ 10 мин); большие времена (т.е. меньшие скорости) сворачивания наблюдаются (для некоторых белков) только в середине перехода, т.е. при небиологических условиях. Светлый пунктир ограничивает область скоростей для сплюснутых (1:2) и продолговатых (2:1) глобул в середине перехода; более темный пунктир – при «биологически-нормальных» условиях. L – число аминокислотных остатков в белковой цепи. $\Delta F \leq 0$ – разность свободной энергии между нативным и развернутым состоянием цепи. Взято из [65].

включает в себя около ста остатков цепи (см. ссылки в [72]), так что поиск правильного «заузливания» может быть лимитирующим только для очень длинных цепей [72], которые все равно не могут уложиться за разумное время (в соответствии с формулой (4)).

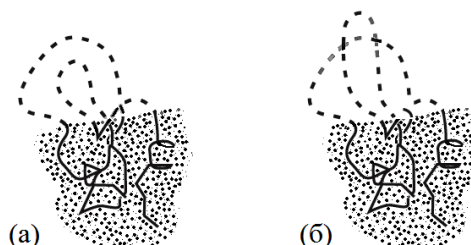


Рис. 6. (а) Компактный интермедиат сворачивания белка с торчащими из него неупорядоченными петлями. Рост интермедиата соответствует смещению границы между фиксированной (глобулярной) и неупорядоченной (клубковой) частями белковой цепи. Успешное сворачивание интермедиата требует правильного образования узлов в его цепи: полусвернутая структура с неправильным заузлением (б) не может прорасти до правильно свернутого белка, ей сперва надо развернуться. Однако в цепи из ~ 100 звеньев может быть лишь один-два узла, так что перебор интермедиатов с разной заузленностью петель не должен лимитировать скорость сворачивания белков [72].

ОЦЕНКА ЗАВИСИМОСТИ ОБЪЕМА ПОИСКА НАТИВНОЙ СТРУКТУРЫ БЕЛКА В КОНФОРМАЦИОННОМ ПРОСТРАНСТВЕ ОТ ЕГО РАЗМЕРА: ВИД НА БАРЬЕР СО СТОРОНЫ РАЗВЕРНУТОГО СОСТОЯНИЯ

Выше приведенная оценка времени сворачивания белка основана на анализе его *разворачивания*, а не сворачивания, потому что легче наметить хороший путь разворачивания любой структуры (и оценить потребное для этого время, см. выше), чем хороший путь, ведущий к сворачиванию структуры с наименьшей энергией, в то время как свободно-энергетический барьер – один и тот же на обоих путях.

Другими словами, мы рассматривали свободно-энергетический барьер между развернутым и свернутым состоянием (рис. 4д), сфокусировавшись на той его стороне, что отвечает увеличению энергии на пути от жерла вулкана до края его кратера (рис. 7; ср. с рис. 4д), и не рассматривали ту сторону барьера, что связана с потерей энтропии на пути от развернутого состояния к краю кратера.

Поскольку скорости прямой и обратной реакций одинаковы в середине перехода (как то следует из физического принципа «детального равновесия»), то здесь обе стороны барьера имеют одинаковую высоту, и, следовательно, рассмотрение только одной стороны («стороны разворачивания») достаточно для того, чтобы оценить высоту барьера.

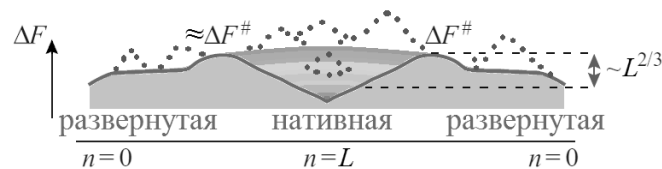


Рис. 7. Этот, чисто иллюстративный рисунок показывает, как энтропия преобразует энергетическую воронку (иллюстрируемую рис. 2б) в «вулканообразный», как он теперь называется [74], ландшафт свободной энергии со свободно-энергетическими барьерами (см. рис. 4д) на каждом пути, ведущем из развернутого к глобулярному состоянию. Любой путь от развернутого состояния к нативному сначала идет вверх, – до барьера, т.е. до края кратера вулкана, и только потом начинает спускаться в «воронку», т.е. в кратер вулкана. Гладкий ландшафт свободной энергии соответствует компактным полусвернутым промежуточным структурам (показанным на рис. 4а), а «скалы» (очерченные пунктирными линиями) представляют собой пейзаж, включающий и некомпактные полусвернутые интермедиаты (показанные на рис. 4б). Более корректная, хоть и менее красивая схема свободно-энергетического ландшафта дана на рисунке 2 в работе [64].

Тем не менее, для полноты анализа надо посмотреть и на другую сторону барьера, – ту, что больше всего всех интересует биологов: она связана со сворачиванием (т.е. с потерей энтропии). Так мы получим другой ракурс головоломки самоорганизации белка.

Для анализа сворачивания надо проанализировать выбор стабильной конформации белковой цепи.

Полный объем пространства всех конформаций цепи, оцененный Левинталем [21] на уровне аминокислотных остатков, воистину огромен: от 3^{100} до 100^{100} конформаций для цепи из 100 остатков.

Однако, – должна ли цепь опробовать все эти 100^{100} конформаций в поисках наиболее стабильной из них? Нет: конформационное пространство покрыто локальными энергетическими минимумами, каждый из которых окружен локальной энергетической воронкой, обеспечивающей быстрый спуск в этот минимум (см. рис. 2б, а также [75] и рис. 6 в [76] в этом же номере *этого журнала*), так что при поиске глобального энергетического минимума огромный «Левинталев перебор» *всех*, в основном некомпактных конформаций цепи, можно заменить на гораздо меньший перебор компактных глобулярных структур, отвечающих глубоким энергетическим минимумам.

Таким образом, для оценки объема пространства поиска, надо оценить количество локальных минимумов энергии (а также время перехода от одного минимума энергии к другому). В каком-то смысле это похоже на перебор возможных «топомеров» белковой цепи [77,

78], но наша цель – не в том, чтобы рассчитать скорость сворачивания белка: нам надо только оценить нижний предел этой скорости, что принципиально отличается от поиска топомеров.

Обзор белковых структур показывает, что взаимодействия в них в основном связаны с вторичными структурами [79–81]. Таким образом, возникает вопрос о том, насколько велико общее количество энергетических минимумов на уровне формирования и сборки в глобулу вторичных структур, – то есть на уровне, который был рассмотрен Птицыным [30] в его модели стадийной самоорганизации белка.

Оказывается, что это число на много порядков меньше, чем число конформаций аминокислотных остатков, т.е. 100^L или 10^L или 3^L , и составляет $\sim L^N$ в L -звенной цепи с N элементами вторичной структуры [82]. N много меньше L , и это является основной причиной резкого уменьшения конформационного пространства.

Оценка порядка L^N получена следующим образом (рис. 8).

Число архитектур (т.е. типов плотной укладки вторичных структур) мало (см. [79, 80, 83]), – как правило, ~ 10 или меньше для данного набора вторичных структур (рис. 8а), так как архитектуры – это укладки слоев вторичной структуры (каждый из которых содержит несколько структурных элементов), и, следовательно, комбинаторика слоев очень мала по сравнению с комбинаторикой гораздо более многочисленных элементов вторичной структуры (см. ниже).

Максимальное количество упаковок, т.е. всех комбинаций позиций N элементов в данной белковой архитектуре составляет $N! \equiv N \times (N-1) \times \dots \times 2 \times 1$ (рис. 8б).

Максимальное число топологий, т.е. всех комбинаций направлений этих элементов, не может превышать 2^N (рис. 8в).

Поперечные сдвиги и наклоны элемента внутри плотной упаковки запрещены (рис. 8г).

Сдвиги и повороты элементов вторичной структуры в пределах плотной упаковки тесно связаны (это показано на рис 8д на примере β -листа, где эта связь наиболее очевидна, но это также относится к α -спиралям – вспомним плотные, «выступы во впадины», упаковки Крика [84]); в результате, каждый α или β элемент может иметь приблизительно L/N (т.е. порядка средней длины элемента) возможных его «сдвиго-вращений» в глобуле, образованной N структурными элементами в L -звенной цепи.

Все это ограничивает число минимумов энергии в конфигурационном пространстве величиной $\sim 10 \times (L/N)^N \times 2^N \times N!$; что (с использованием приближения Стирлинга $N! \sim (N/e)^N$) дает, в главном члене при $L \gg N \gg 1$ [82]:

$$\text{ЧИСЛО подлежащих перебору энергетических минимумов} \sim L^N. \quad (5)$$

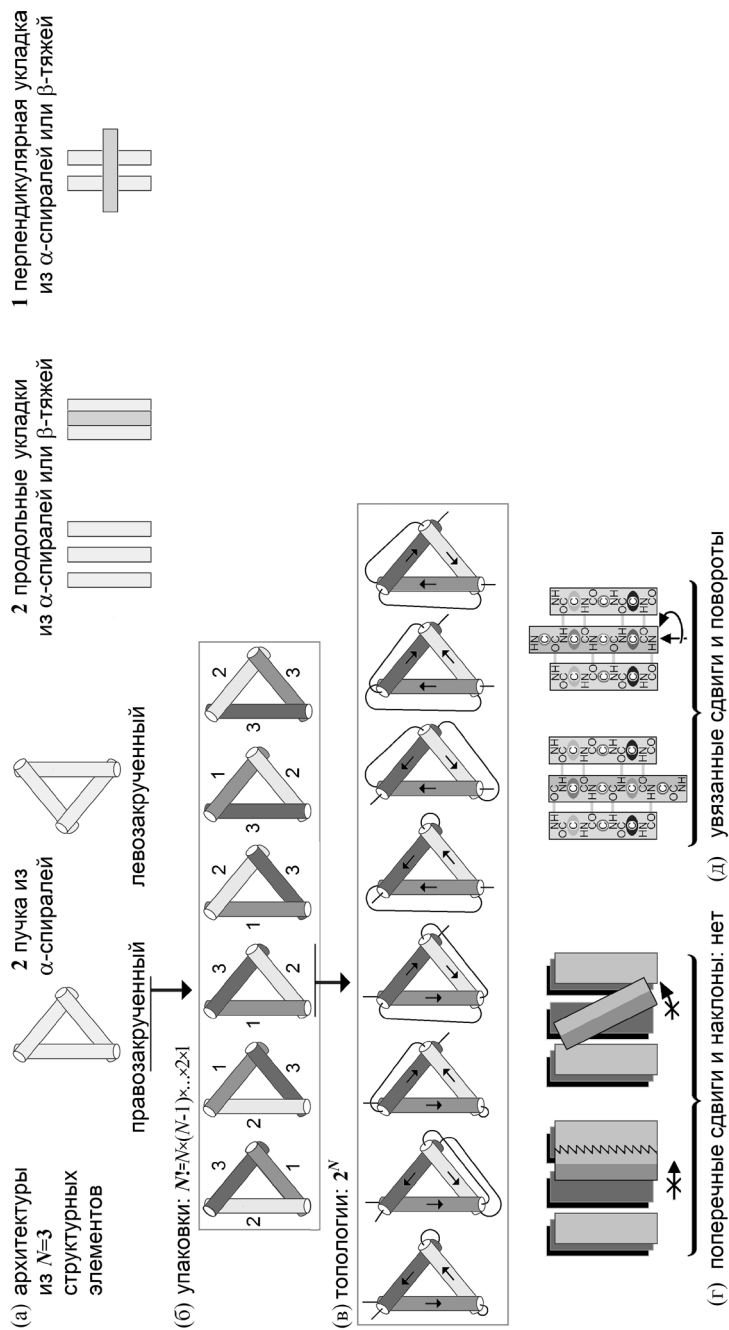


Рис. 8. Схема оценки объема конформационного пространства на уровне всевозможных упаковок структурных элементов. Взято из Приложения к [82].

Это число может быть снижено симметрией глобулы; кроме того, α -спираль не может занять место в β -листе независимо от перегруппировки других элементов, и наоборот, потому что β -тяжу нужен партнер, другой β -тяж, для образования водородных связей, а α -спираль избегает такого партнерства. Сверх того, короткие или пересекающиеся петли между структурными элементами могут не позволить им принять произвольные положения и направления в глобуле, и т.д. [85]. Однако, это снижение не важно для нас, потому что наша цель состоит в том, чтобы оценить *верхний предел* числа конфигураций.

Здесь возникает вопрос: как цепь знает, где сформировать вторичную структуру и какую? Ответ на него представляется следующим: большинство вторичных структур определяется локальными аминокислотными последовательностями [30, 86, 87]. Так или иначе, выбор «to be or not to be» элементу вторичной структуры добавляет только одно состояние к уже учтенному числу L/N его возможных сдвиг-поворотов, а преобразование $\alpha \leftrightarrow \beta$ лишь удваивает его, что не очень существенно; см [88].

В не слишком малой компактной глобуле, длина элемента вторичной структуры должна быть пропорциональна диаметру глобулы, т.е. $\sim L^{1/3}$. Конкретнее: объем глобулы $\approx 150 \text{ \AA}^3 \times L$ (и, значит, его диаметр $\approx 5 \text{ \AA} \times L^{1/3}$), в то время как сдвиг на один остаток – 1.5 \AA в α -спирали и $\approx 3 \text{ \AA}$ в вытянутой цепи [81]. Таким образом, α -спираль состоит из $\approx 3L^{1/3}$, а β -тяж или петля – из $\approx 1.5L^{1/3}$ остатков, т.е.

$$\begin{aligned} \text{ЧИСЛО элементов «вторичная структура + петля» } N &\approx \\ &\approx L^{2/3}/4.5 \text{ — } L^{2/3}/3, \end{aligned} \quad (6)$$

а ожидаемая величина L^N (т.е. оценка величины полного перебора) лежит в пределах

$$\sim L^{2/3/4.5} \equiv \exp([\ln(L)/4.5] \times L^{2/3}) \text{ — } \sim L^{2/3/3} \equiv \exp([\ln(L)/3] \times L^{2/3}). \quad (7)$$

Аналогичная зависимость величины L^N от L получается [70, 71] из строгого математического анализа степени сложности задачи поиска.

Так как $\ln(L)/4.5 \approx 1$ и $\ln(L)/3 \approx 1.5$ при $L \approx 80\text{--}90$, то вышеуказанные пределы близки к верхнему пределу, даваемому уравнением (4).

С другой стороны, значение L/N (т.е. количество остатков, в среднем приходящееся на элемент вторичной структуры с сопровождающей его петлей) составляет 15 ± 5 согласно белковой статистике [74]; это тоже приводит к величине L^N , численно весьма близкой к вышеуказанной оценке.

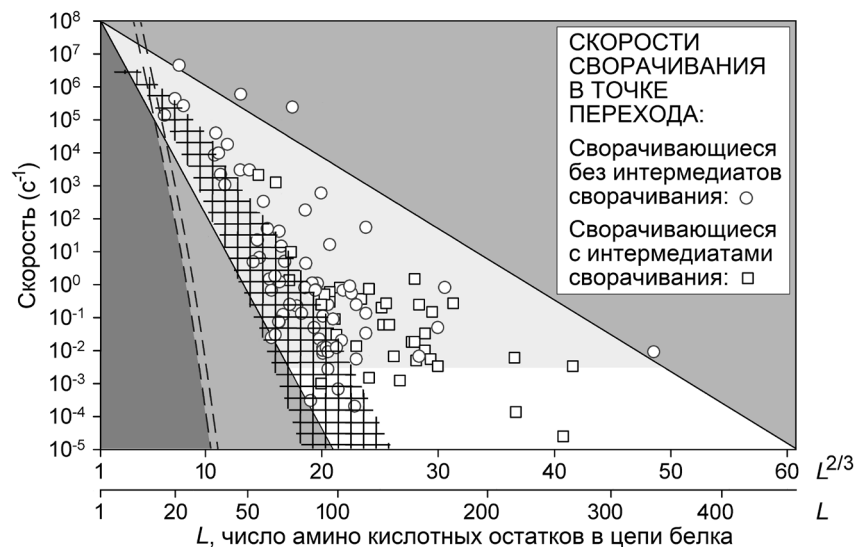


Рис. 9. Скорость перебора и скорость сворачивания.

Скорости сворачивания (кружки и квадратики) показаны для белков, экспериментально изученных *in vitro* в середине перехода (т.е., при равной стабильности их нативных и развернутых состояний); большой светлый треугольник – прогнозируемый (из анализа разворачивания!) диапазон этих скоростей (ср. с рис. 5). Сетчатая штриховка – теоретическая оценка минимальной скорости исчерпывающего перебора, при сворачивании, всевозможных укладок вторичных структур (α -спиралей и β -тяжей). Предел «скорости перебора по Левинталю» ($10^{12} \text{ s}^{-1} / 3^L$ при всего трех возможных состояниях остатка: α , β , клубок) показан двойным пунктиром, а скорости подобных переборов при 10 или 100 состояниях остатка лежат гораздо ниже (в темно-серой зоне слева). Рисунок взят из [91].

Принимая, из экспериментов по сворачиванию самых маленьких белков [89, 90], микросекунду в качестве приблизительной оценки времени «просмотра» одной конфигурации, и, взяв $L/N=15 \pm 5$ из белковой статистики, – видим, что время, теоретически необходимое для зондирования всего конформационного пространства на уровне формирования и сборки вторичной структуры примерно совпадает (рис. 9) с верхним пределом экспериментальных времен сворачивания, наблюдаемых для небольших ($L \lesssim 80-90$ остатков) белков (выбор структуры которых диктуется, как мы уже знаем, стабильностью).

Вышесказанное не означает, что белковая цепь должна просмотреть *все* свои конформации на уровне формирования и упаковки вторичных структур (хотя цепь из 80–90 и менее остатков может сделать

это за минуты, как то показывает, для некоторых белков, рис. 9). Оно означает только то, что «энергетическая воронка», ведущая к нативной структуре, начинает свою работу от уровня вторичных структур, и потому должна ускорить процесс сворачивания лишь на несколько порядков (как то показывает рис. 9 для большинства белков), а не десятки и сотни порядков, – что было бы, *если бы* она начала работать с уровня аминокислотных остатков (ср. с теорией поиска топомеров у Дебе и др. [77] и Макарова и Плакско [78]). Рисунок 9 показывает, что ускорение особенно заметно для цепей из > 100 остатков, но даже и тогда основная работа осуществляется вторичными структурами.

С высоты птичьего полета, полученные оценки (4)–(7) числа конформаций цепи, подлежащих перебору при поиске ее наиболее стабильной структуры, выглядят так. Это число определяется, в главном члене, размером поверхности глобулы – числом поверхностных остатков или – что почти то же – числом вторичных структур N (оба эти числа пропорциональны $L^{2/3}$). Физическая причина – в том, что все независимые степени свободы в плотной глобуле связаны с ее поверхностью, т.к. плотная глобула запрещает независимые перегруппировки остатков внутри нее [45, 92] – так же, как то делает вторичная структура. С этой точки зрения, используемые нами здесь вторичные структуры не являются необходимыми для полученных принципиальных оценок (отметим, что аналогичные оценки Фу и Ванга [70] и Штейнхофеля и др. [71], а также наши предыдущие оценки [55, 56, 64, 65, 93], не использовали вторичных структур), – хотя эти структуры действительно образуют ядро белковой глобулы и они удобны для уточнения полученных принципиальных оценок.

IV. ЗАКЛЮЧЕНИЕ

Мы осмотрели путь через иллюстрируемый рисунком 7 «вулканообразный» свободно-энергетический ландшафт сворачивающейся белковой цепи как от подножья вулкана в кратер, так из жерла до края кратера. Так мы исследовали обе стороны барьера, разделяющего свернутые и развернутые состояния цепи, перейдя барьер туда и обратно и ознакомившись с двумя, решающими парадокс Левинталя, аспектами сворачивания/разворачивания белка.

Сторона барьера, обращенная к нативной структуре, проще для анализа, потому что легче наметить разумный путь разворачивания любой структуры, чем путь сворачивания структуры, еще неизвестной

самой цепи. Анализ разворачивания, т.е. вид изнутри воронки сворачивания, дал нам оценку диапазона времен разворачивания, – и, далее, принцип детального равновесия дал нам диапазон времен сворачивания (или, точнее, – перебора). Анализ же сворачивания дал нам только верхнюю границу этого диапазона.

Заметим, что та же схема может быть, в принципе, применима к образованию нативной структуры белка не только из клубка (который мы использовали здесь для простоты), но и из расплавленной глобулы или из другого состояния. Но для таких процессов (где, впрочем, опыт не показывает, см. рис. 5 и 9, принципиального ускорения самоорганизации) все оценки были бы гораздо сложнее из-за сложности анализа денатурированного состояния. Поэтому мы не будем выходить за рамки простого случая – образования нативной глобулы из клубка.

Отметим, что «проблема Левинталя» существует при кристаллизации (напоминающей самоорганизацию белка – там тоже разные атомы должны выбрать одну из очень многих конфигураций в «пока неизвестном» для них кристалле); однако, насколько мы знаем, там она не привлекала там такого внимания, как в белках (см. [94, 95]).

Несколько заключительных замечаний:

1. Наша оценка *числа* подлежащих рассмотрению ансамблей вторичных структур (т.е. числа соответствующих им минимумов энергии белковой цепи) не зависит (см. формулы (5) – (7)) от стабильности этих ансамблей. Влияние стабильности (ΔF) нативного состояния на *время* сворачивания рассматривается ниже.

2. Наша оценка (4) времени сворачивания до сих пор относилась к точке равновесия развернутого и нативного состояний, где $\Delta F = 0$, а наблюдаемое время сворачивания максимально – но оно может на порядки величины превышать время сворачивания при нативных условиях [33].

Уместен вопрос: как изменится время сворачивания, когда нативное состояние станет более стабильным, чем клубок (т.е. при $\Delta F < 0$)? И эксперимент (см [33]), и теоретический анализ [56] показывают, что при небольшой (порядка нескольких $k_B T$, так что стабильные интермедиаты не возникают) величине $-\Delta F$ время сворачивания падает с ростом стабильности, и, теоретически [65], оно может быть оценено как

$$\text{ВРЕМЯ} \sim \tau \times \exp[(1 \pm 0.5) \times (L^{2/3} + 0.4 \times \Delta F/k_B T)], \quad (8)$$

где множитель 0.4 соответствует теоретической оценке доли цепи, вовлеченной в ядро сворачивания, так что $0.4 \times \Delta F$ – оценка изменения свободной энергии ядра. Это уравнение дает оценку скоростей сворачивания, происходящего при разных условиях (см. рис. 5).

Для случая очень высокой стабильности нативного состояния ($-\Delta F \gg k_B T$), Тирумалаи [96] предложил закономерность $\ln(\text{ВРЕМЯ}) \sim L^{1/2}$, отличную от формулы (4), но сходную с ней. В этом случае быстрое сворачивание белка идет «под гору» (по энергии); но «энергетический склон» имеет (из-за гетерогенности белка) ухабы с энергией, пропорциональной $L^{1/2}$. Однако численные эксперименты с решеточными моделями белков показали [35, 58], что при температуре, обеспечивающей наиболее быстрое сворачивание, время его растет с длиной цепи как $\ln(\text{ВРЕМЯ}) \sim A \times \ln(L)$, где коэффициент A равен 6 для «случайных» цепей и 4 для цепей, «отредактированных» для наиболее быстрого сворачивания и имеющих большую энергетическую щель между наиболее стабильной укладкой и остальными. Это еще раз подчеркивает зависимость скорости сворачивания от экспериментальных условий и от величины энергетической щели [44, 57].

3. Некоторые белки «метаморфны» [97]: наблюдаются в двух разных структурах. Наиболее интересны нам некоторые, очень немногие из них (например, серпин), сначала обретающие «нативную» структуру, работающую в клетке или пробирке около часа, а затем переходящие в другую, не работающую, но более стабильную структуру [98]. Важно, что этот переход не связан с изменением окружения белка (т.е. с агрегацией, как у амилоидов, или с формированием каких-то комплексов). Так что цепь такого белка имеет две стабильные структуры: первая складывается быстрее, другая – более стабильна. Видимо, такие белки и должны быть очень редки: теоретические оценки [35, 81] показывают, что аминокислотная последовательность, кодирующая одну стабильную структуру (энергия которой отделена большой щелью от энергий других) – редкость сама по себе, а последовательность, кодирующая две стабильные структуры – чудо в квадрате...

4. Уравнения (4), (8) оценивают диапазон возможных скоростей сворачивания, а не скорости сворачивания индивидуальных белков, которые, даже для белков одинакового размера, могут (рис. 5) отличаться друг от друга на несколько порядков. Влияние формы укладки конкретной белковой цепи на время сворачивания можно оценить с помощью феноменологического параметра «контактного порядка» (CO%) [99]. CO% равен среднему расстоянию по цепи между остатками, контактирующими в нативном белке, деленному на длину цепи (см. также [51, 100]). Большой CO% отражает наличие множества длинных замкнутых петель в нативной глобуле, в то время как высокое значение коэффициента (1 ± 0.5) в формулах (4), (8) отражает наличие их на поверхности полусвернутой глобулы

(рис. 6). Поэтому CO% более или менее пропорционален фактору (1 ± 0.5) [101]. CO% сам по себе хорош для сравнения скоростей сворачивания белков одинакового размера, но не для сравнения скоростей сворачивания небольших и крупных белков, так как CO% падает примерно пропорционально $L^{-1/3}$ с ростом длины цепи L [65, 101, 102] (что отражает малую «запутанность» цепей крупных белков), – в то время как скорость сворачивания падает (а его *ВРЕМЯ* растет) с увеличением размера белка (рис. 5).

Поэтому хорошо предсказывает скорость сворачивания белков величина $AbsCO = CO\% \times L$, которая растет с длиной цепи L как $L^{2/3}$ [101] и сочетает эффект, следующий из формы укладки белковой цепи [99, 102], с основным эффектом, следующим из размера белка.

5. В этом обзоре, я позволил себе оставить в стороне многочисленные попытки «биоинформатически» предсказывать скорости сворачивания белков, так как, согласно «слепому» тестированию [103], они приводят к неудовлетворительным результатам.

6. Возвращаясь к парадоксу Левинтала, можно заключить, что он решен для белковых цепей из менее чем 100 аминокислотных остатков (при условии, что их последовательности обеспечивают значительную стабильность только одной их укладки); Это связано с тем, что (а) эти цепи могут преодолеть барьер свободной энергии на пути к их наиболее устойчивым укладкам независимо от их сложности (рис. 5), и (б) они могут просмотреть все свои укладки на уровне формирования и сборки ансамблей вторичных структур (рис. 9), и найти наиболее стабильную из них.

Что касается более крупных цепей, они в состоянии просмотреть только относительно простые (не слишком «запутанные») укладки, так что для них остается открытым вопрос о том, может ли какая-либо другая (очень «запутанная») их укладка быть более стабильной, чем нативная (что действительно наблюдается для некоторых «особых» белков, подобных серпину, имеющему 400 остатков в цепи).

БЛАГОДАРНОСТИ

Я благодарен О. Б. Птицыну, А. М. Гутину и Е. И. Шахновичу за множество плодотворных дискуссий, и моим соавторам по работам, посвященным теории сворачивания белков – А. Я. Бадретдинову, О. В. Галзитской, Д. Н. Иванкову, Н. С. Богатыревой и С. А. Гарбузинскому.

ЛИТЕРАТУРА

1. Anfinsen, C.B., Haber, E., Sela, M., White, F.H., Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA*, **47**, 1309–1314.
2. Страйер Л. (1984) *Биохимия*. Т. 1, Гл. 2, 9. М: Мир.
3. Creighton, T.E. (1991) *Proteins*. 2-nd edn. Chs. 2, 7. N. Y.: W.H. Freeman & Co.
4. Kolb, V.A., Makeev, E.V., Spirin, A.S. (1994) Folding of firefly luciferase during translation in a cell-free system. *EMBO J.*, **13**, 3631–3637.
5. Komar, A.A., Kommer, A., Krashe-ninnikov, I.A., Spirin, A.S. (1997) Cotranslational folding of globin. *J. Biol. Chem.*, **272**, 10646–10651.
6. Eichmann, C., Preissler, S., Riek, R., Deuerling, E. (2010) Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA*, **107**, 9111–9116.
7. Han, Y., David, A., Liu, B., Magadan, J.G., Bennink, J.R., Yewdell, J.W., Qian, S.-B. (2012) Monitoring co-translational protein folding in mammalian cells at codon resolution. *Proc. Natl. Acad. Sci. USA*, **109**, 12467–12472.
8. Holtkamp, W., Kobic, G., Jäger, M., Mittelstaet, J., Komar, A.A., Rodnina, M.V. (2015) Cotranslational protein folding on the ribosome monitored in real time. *Science*, **350**, 1104–1107.
9. Flanagan, J.M., Kataoka, M., Shortle, D., Engelman, D.M. (1992) Truncated staphylococcal nuclease is compact but disordered. *Proc. Natl. Acad. Sci. USA*, **89**, 748–752.
10. Wright, P.E., Dyson, H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, **293**, 321–331.
11. Uversky, V.N., Gillespie, J.R., Fink, A.L. (2000) Why are 'natively un-folded' proteins unstructured under physiologic conditions? *Proteins*, **41**, 415–427.
12. Petsko, G.A., Ringe, D. (2004) *Protein Structure and Function*. Ch. 1. London: New Science Press Ltd.
13. Ellis, R.J., Hartl, F.U. (1999) Principles of protein folding in the cellular environment. *Curr. Opin. Struct. Biol.*, **9**, 102–110.
14. Libich, D.S., Tugarinov, V., Clore, G.M. (2015) Intrinsic unfoldase/foldase activity of the chaperonin GroEL directly demonstrated using multinuclear relaxation-based NMR. *Proc. Natl. Acad. Sci. USA*, **112**, 8817–8823.
15. Marchenko, N.Y., Marchenkov, V.V., Semisotnov, G.V., Finkelstein, A.V. (2015) Strict experimental evidence that apo-chaperonin GroEL does not accelerate protein folding, although it does accelerate one of its steps. *Proc. Natl. Acad. Sci. USA*, **112**, E6831–E6832.
16. Марченков В.В., Соколовский И.В., Котова Н.В., Галзитская О.В., Бочкарева Е.С., Гиршович А.С., Семисотнов Г.В. (2004) Взаимодействие шаперона GroEL с ранними кинетическими промежуточными состояниями ренатурирующих белков ингибирует формирование их нативной структуры. *Биофизика*, **49**, 987–994.
17. Marchenko, N.Y., Garbuzynskiy, S.O., Semisotnov, G.V. (2009) Molecular chaperones under normal and pathological conditions. *Molecular Pathology of Proteins*. / Ред. D.I. Zabolotny. New York: Nova Science Publishers, 57–89.
18. Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science*, **181**, 223–230.
19. Gutte, B., Merrifield, R.B. (1969) The total synthesis of an enzyme with ribonuclease A activity. *J. Am. Chem. Soc.*, **91**, 501–502.

20. Levinthal, C. (1968) Are there pathways for protein folding? *J. Chim. Phys. Chim. Biol.*, **65**, 44–45.
21. Levinthal, C. (1969). How to fold graciously. *Mössbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois* / Ред. P.Debrunner, J.C.M.Tsibris, E.Munck. Urbana-Champaign, IL: University of Illinois Press, 22–24.
22. Privalov, P.L. (1979) Stability of proteins: small globular proteins. *Adv. Protein Chem.*, **33**, 167–241.
23. Šali, A., Shakhnovich, E., Karplus, M. (1994) Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. *J. Mol. Biol.*, **235**, 1614–1636.
24. Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I. (1994) Specific nucleus as a transition state for protein folding: evidence from the lattice model. *Biochemistry*, **33**, 10026–10031.
25. Phillips, D.C. (1966) The three-dimensional structure of an enzyme molecule. *Sci. Am.* **215**, 78–90.
26. Goldenberg, D.P., Creighton, T.E. (1983) Circular and circularly permuted forms of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.*, **165**, 407–413.
27. Grantcharova, V.P., Riddle, D.S., Santiago, J.V., Baker, D. (1998) Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat. Struct. Biol.*, **5**, 714–720.
28. Wetlaufer, D.B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA*, **70**, 697–701.
29. Fulton, K.F., Main, E.R.G., Daggett, V., Jackson, S.E. (1999) Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J. Mol. Biol.*, **291**, 445–461.
30. Птицын О. Б. (1973) Стадийный механизм организации белковых молекул. *Докл. Акад. Наук СССР*, **210**, 1213–1215.
31. Ptitsyn, O.B. (1995) Molten globule and protein folding. *Adv. Protein Chem.*, **47**, 83–229.
32. Privalov, P.L. (1996) Intermediate states in protein folding. *J. Mol. Biol.*, **258**, 707–725.
33. Fersht, A. (1999) *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*. Chs. 2, 15, 18, 19. N.Y.: W.H. Freeman & Co.
34. Melnik, B.S., Marchenkov, V.V., Evdokimov, S.R., Samatova, E.N., Kotova, N.V. (2008) Multi-state protein: Determination of carbonic anhydrase free-energy landscape. *Biochem. Biophys. Res. Commun.*, **369**, 701–706.
35. Finkelstein, A.V., Ptitsyn, O.B. (2016) *Protein Physics. A Course of Lectures*. 2nd Edition. Chs. 7, 10, 13, 18, 19–21. Amsterdam • Boston • Heidelberg • London • New York • Oxford • Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo: Academic Press, an Imprint of Elsevier Science.
36. Leopold, P.E., Montal, M., Onuchic, J.N. (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA*, **89**, 8721–8725.
37. Wolynes, P.G., Onuchic, J.N., Thirumalai, D. (1995) Navigating the folding routes. *Science*, **267**, 1619–1620.
38. Dill, K.A., Chan, H.S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol.*, **4**, 10–19.
39. Bicout, D.J., Szabo, A. (2000) Entropic barriers, transition states, funnels, and exponential protein folding kinetics: A simple model. *Protein Sci.*, **9**, 452–465.
40. Tanford, C. (1968) Protein denaturation. *Adv. Protein Chem.*, **23**, 121–282.
41. Creighton, T.E. (1978) Experimental studies of protein folding and unfolding. *Prog. Biophys. Mol. Biol.*, **33**, 231–297.
42. Shakhnovich, E.I., Gutin, A.M. (1990) Implications of thermodyna-

- mics of protein folding for evolution of primary sequences. *Nature*, **346**, 773–775.
43. Gutin, A.M., Shakhnovich, E.I. (1993) Ground state of random copolymers and the discrete random energy model. *J. Chem. Phys.*, **98**, 8174–8177.
 44. Galzitskaya, O.V., Finkelstein, A.V. (1995) Folding of chains with random and edited sequences: similarities and differences. *Protein Eng.*, **8**, 883–892.
 45. Shakhnovich, E.I. (2006) Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. *Chem. Rev.*, **106**, 1559–1588.
 46. Bryngelson, J.D., Wolynes, P.G. (1989) Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.*, **93**, 6902–6915.
 47. Gō, N., Abe, H. (1981) Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers*, **20**, 991–1011.
 48. Ngo, J.T., Marks, J. (1992) Computational complexity of a problem in molecular structure prediction. *Protein Eng.*, **5**, 313–321.
 49. Unger, R., Moult, J. (1993) Finding the lowest free energy conformation of a protein is an NP-hard problem: proof and implications. *Bull. Math. Biol.*, **55**, 1183–1198.
 50. Karplus, M. (1997) The Levinthal paradox: yesterday and today. *Fold. Des.*, **2**, Suppl. 1, S69–S75.
 51. Nölting, B. (2010) *Protein Folding Kinetics: Biophysical Methods*. Chs. 10, 11, 12. N.Y.: Springer.
 52. Bogatyreva, N.S., Finkelstein, A.V. (2001) Cunnings simplicity of protein folding landscapes. *Protein Eng.*, **14**, 521–523.
 53. Zwanzig, R., Szabo, A., Bagchi, B. (1992) Levinthal's paradox. *Proc. Natl. Acad. Sci. USA*, **89**, 20–22.
 54. Finkelstein, A.V. (2002) Cunnings simplicity of a hierarchical folding. *J. Biomol. Struct. Dyn.*, **20**, 311–313.
 55. Финкельштейн А. В., Бадретдинов А. Я. (1997) Физические причины быстрой самоорганизации стабильной пространственной структуры белков: Решение парадокса Левинтала. *Мол. Биол.*, **31**, 469–477.
 56. Finkelstein, A.V., Badretdinov, A.Ya. (1997) Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. *Fold. Des.*, **2**, 115–121.
 57. Wolynes, P.G. (1997) Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proc. Natl. Acad. Sci. USA*, **94**, 6170–6175.
 58. Gutin, A.M., Abkevich, V.I., Shakhnovich, E.I. (1996) Chain length scaling of protein folding time. *Phys. Rev. Lett.*, **77**, 5433–5436.
 59. Zana, R. (1975) On the rate determining step for helix propagation in the helix-coil transition of polypeptides in solution. *Biopolymers*, **14**, 2425–2428.
 60. Eyring, H. (1935) The activated complex in chemical reactions. *J. Chem. Phys.*, **3**, 107–115.
 61. Полинг Л. (1974) *Общая химия*. Гл. 16. М.: Мир.
 62. Эмануэль Н. М., Кнорре Д. Г. (1984) *Курс химической кинетики*. 4-е изд. Гл. III, V (§§ 2, 3). М.: Высшая школа.
 63. Ландау Л. Д., Лифшиц, Е. М. (1964) *Статистическая физика*. §§ 7, 8, 150. М.: Наука.
 64. Galzitskaya, O.V., Finkelstein, A.V. (1999) A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl. Acad. Sci. USA*, **96**, 11299–11304.
 65. Garbuzynskiy, S.O., Ivankov, D.N., Bogatyreva, N.S., Finkelstein, A.V. (2013) Golden triangle for folding

- rates of globular proteins. *Proc. Natl. Acad. Sci. USA*, **110**, 147–150.
66. Finkelstein, A.V., Badretdin, A.J., Galzitskaya, O.V., Ivankov, D.N., Bogatyreva, N.S., Garbuzynskiy, S.O. (2017) There and back again: Two views on the protein folding puzzle. *Phys. Life Rev.*, **21**, 56–71.
 67. Шахнович Е.И., Финкельштейн, А.В. (1982) К теории кооперативных переходов в белках. *Докл. АН СССР*, **267**, 1247–1250.
 68. Jacobson, H., Stockmayer, W. (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Chem. Phys.*, **18**, 1600–1606.
 69. Флори П. (1971) *Статистическая механика цепных молекул*. Гл. 3. М.: Мир.
 70. Fu, B., Wang, W. (2004) A $2^{O(n^{1-1/d} \cdot \log(n))}$ time algorithm for d-dimensional protein folding in the HP-model. *Lecture Notes in Computer Science*, **3142**, 630–644.
 71. Steinhofel, K., Skaliotis, A., Albrecht, A.A. (2006) Landscape analysis for protein folding simulation in the H-P model. *Lecture Notes in Computer Science*, **4175**, 252–261.
 72. Finkelstein, A.V., Badretdinov, A.Ya. (1998) Influence of chain knotting on the rate of folding. ADDENDUM to Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. *Fold. Des.*, **3**, 67–68.
 73. Galzitskaya, O.V., Ivankov, D.N., Finkelstein, A.V. (2001) Folding nuclei in proteins. *FEBS Lett.*, **489**, 113–118.
 74. Rollins, G.C., Dill, K.A. (2014) General mechanism of two-state protein folding kinetics. *J. Am. Chem. Soc.*, **136**, 11420–11427.
 75. Finkelstein, A.V. (2017) Some additional remarks to the solution of the protein folding puzzle: Reply to comments on «There and back again: Two views on the protein folding puzzle». *Phys. Life Rev.*, **21**, 77–79.
 76. Бычкова В.Е., Семисотнов Г.В., Балобанов В.А., Финкельштейн А.В. (2018) Расплавленная глобула: 45 лет спустя. *Усп. биол. хим.*, **83**, 67–100.
 77. Debe, D.A., Carlson, M.J., Goddard, W.A., 3rd. (1999) The topomer-sampling model of protein folding. *Proc. Natl. Acad. Sci. USA*, **96**, 2596–601.
 78. Makarov, D.E., Plaxco, K.W. (2003) The topomer search model: A simple, quantitative theory of two-state protein folding kinetics. *Protein Sci.*, **12**, 17–26.
 79. Levitt, M., Chothia, C. (1976) Structural patterns in globular proteins. *Nature*, **261**, 552–558.
 80. Chothia, C., Finkelstein, A.V. (1990) The classification and origins of protein folding patterns. *Ann. Rev. Biochem.*, **59**, 1007–1039.
 81. Финкельштейн А.В., Птицын О.Б. (2012) *Физика белка. Курс лекций с цветными и стереоскопическими иллюстрациями и задачами*. 4-е дополненное издание. Гл. 7, 10, 13, 18–21. М.: Книжный Дом «Университет».
 82. Finkelstein, A.V., Garbuzynskiy, S.O. (2015) Reduction of the search space for the folding of proteins at the level of formation and assembly of secondary structures: A new view on solution of Levinthal's paradox. *ChemPhysChem*, **16**, 3373–3378.
 83. Murzin, A.G., Finkelstein, A.V. (1988) General architecture of α -helical globule. *J. Mol. Biol.*, **204**, 749–770.
 84. Crick, F.H.C. (1953) The packing of α -helices: Simple coiled coils. *Acta Crystallogr.*, **6**, 689–697.
 85. Ptitsyn, O.B., Finkelstein, A.V. (1980) Similarities of protein topologies: evolutionary divergence, functional convergence or principles of folding? *Quart. Rev. Biophys.*, **13**, 339–386.
 86. Птицын О.Б., Финкельштейн А.В. (1970) Связь вторичной структуры глобулярных белков с их первичной структурой. *Биофизика*, **15**, 757–767.

87. Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202; современная версия программы: <http://bioinf.cs.ucl.ac.uk/psipred/>.
88. Финкельштейн А.В., Гарбузинский С.А. (2016) Решение парадокса Левинтала возможно на уровне формирования и упаковки вторичных структур белков. *Биофизика*, **61**, 5–10.
89. Muñoz, V., Thompson, P.A., Hofrichter, J., Eaton, W.A. (1997) Folding dynamics and mechanism of beta-hairpin formation. *Nature*, **390**, 196–199.
90. Mukherjee, S., Chowdhury, P., Bunagan, M.R., Gai, F. (2008) Folding kinetics of a naturally occurring helical peptide: Implication of the folding speed limit of helical proteins. *J. Phys. Chem. B*, **112**, 9146–9150.
91. Finkelstein, A.V. (2015) Two views on the protein folding puzzle. <http://atlasofscience.org/two-views-on-the-protein-folding-puzzle/>.
92. Shakhnovich, E.I., Gutin, A.M. (1989) Formation of unique structure in polypeptide-chains theoretical investigation with the aid of a replica approach. *Biophys. Chem.*, **34**, 187–199.
93. Финкельштейн А.В. (2014) *Физика белковых молекул*. Гл. 9. М. – Ижевск: АНО «Ижевский институт компьютерных исследований».
94. Уббелоде А. (1969) *Плавление и кристаллическая структура*. Гл. 2, 5, 6, 10–12, 14, 16. М.: Мир.
95. Slezov, V.V. (2009) *Kinetics of First-Order Phase Transitions*. Chs. 3–5, 8. Weinheim: Wiley-VCH.
96. Thirumalai, D. (1995) From minimal models to real proteins: Time scales for protein folding kinetics. *J. Phys. I. (Orsay, Fr.)*, **5**, 1457–1469.
97. Murzin, A.G. (2008) Metamorphic proteins. *Science*, **320**, 1725–1726.
98. Tsutsui, Y., Cruz, R.D., Wintrobe, P.L. (2012) Folding mechanism of the metastable serpin $\alpha 1$ -antitrypsin. *Proc. Natl. Acad. Sci. USA*, **109**, 4467–4472.
99. Plaxco, K.W., Simons, K.T., Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.*, **277**, 985–994.
100. Nölting, B., Schälike, W., Hampel, P., Grundig, F., Gantert, S., Sips, N., Bandlow, W., Qi, P.X. (2003) Structural determinants of the rate of protein folding. *J. Theor. Biol.*, **223**, 299–307.
101. Ivankov, D.N., Garbuzynskiy, S.O., Alm, E., Plaxco, K.W., Baker, D., Finkelstein, A.V. (2003) Contact order revisited: Influence of protein size on the folding rate. *Protein Sci.*, **12**, 2057–2062.
102. Ivankov, D.N., Bogatyreva, N.S., Lobanov, M.Yu., Galzitskaya, O.V. (2009) Coupling between properties of the protein shape and the rate of protein folding. *PLoS One*, **4**, e6476.
103. Corrales, M., Cuscó, P., Usmanova, D.R., Chen, H.C., Bogatyreva, N.S., Filion, G.J., Ivankov, D.N. (2015) Machine learning: How much does it tell about protein folding rates? *PLoS One*, **10**, e0143166.