

ОБЗОРЫ АКТУАЛЬНЫХ ПРОБЛЕМ

Неупорядоченные полимеры

А.Ю. Гроссберг

Рассматривается одиночная макромолекула с такими формами беспорядка, как разветвления, узлы и гетерогенные последовательности химически различных мономерных звеньев. Во всех случаях рассматриваются математически простейшие подходы, достаточные для достижения физического понимания. Для разветвленных полимеров изложена простая теория типа Флори, демонстрирующая различие классов универсальности для полимеров с замороженными и размороженными ветвлениями. Для узлов рассмотрена другая теория типа Флори, описывающая как набухание и/или коллапс полимерного кольца с замороженной топологией, так и распределение по узлам для размороженного случая. Для рассмотрения гетерополимеров систематически используется так называемая модель случайных энергий (МСЭ) Дерриды, заимствованная из теории спиновых стекол. Она позволяет описать просто и вместе с тем строго как стеклование глобулы гетерополимера со случайной последовательностью, так и конструирование последовательностей с энергетически оптимизированной конформацией основного состояния. Наряду с аналитической теорией рассматриваются компьютерные тесты стеклования пространственной структуры и конструирования последовательностей. Гипотеза о том, что в ходе биологической и/или предбиологической эволюции первичные структуры белков прошли стадию конструирования, ведет к предсказанию определенного типа корреляций в последовательностях белков. Изложенные в работе статистические тесты действительно выявляют корреляции предсказанного типа.

PACS numbers: 36.20.-г, 64.70.-р, 87.10.+е, 87.10.-v

Содержание

1. Введение (130).

1.1. Типы беспорядка в полимерной макромолекуле. 1.2. Беспорядок и биополимеры. 1.3. Гетерополимеры и проблема сворачивания белков. 1.4. Замечание о терминологии. 1.5. Что не рассматривается в этом обзоре.

2. Беспорядок разветвлений (132).

2.1. Формула Зимма – Штокмайера. 2.2. Замороженные и размороженные разветвления. 2.3. Теория типа Флори для замороженного разветвленного полимера. 2.4. Теория типа Флори для размороженного разветвленного полимера. 2.5. Условия приготовления полимера для замороженного случая. 2.6. Обсуждение и сравнение с данными других теоретических подходов и с методом Монте-Карло.

3. Топологический беспорядок: узлы и зацепления (135).

3.1. При чем здесь топология? 3.2. Топология и беспорядок. 3.3. Общая формулировка топологической задачи в физике полимеров. 3.4. Коротко об инвариантах узлов. 3.5. Что известно об энтропии узла? 3.6. Скомканная глобула. 3.7. Раздувание узла. 3.8. Затянутые узлы.

4. Фазовый переход стеклования: глобула гетерополимера со случайной последовательностью (143).

4.1. Глобулярный гетерополимер. 4.2. Модель случайных энергий (МСЭ). 4.3. Применима ли МСЭ как модель стеклования гетерополимера? 4.4. Размороженный гетерополимер. 4.5. Фазовый переход стеклования. 4.6. Компьютерные тесты стеклования.

5. Конструирование гетерополимеров (153).

5.1. Конструирование последовательностей с помощью канонического ансамбля. 5.2. Конструирование и сворачивание с разными взаимодействиями.

6. Статистика последовательностей реальных белков (161).

6.1. Индивидуальный подход или статистический? 6.2. Представление случайными блужданиями для изучения статистики последовательностей. 6.3. Конструирование последовательностей и модели эволюции.

7. Заключение (164).

Список литературы (165).

1. Введение

Для упрощения чтения статьи начинаем ее с краткого толкового словаря основных терминов.

Беспорядок замороженный — совокупность всех степеней свободы, которые фиксируются при приготовлении системы и не меняются в ходе теплового движения.

Беспорядок размороженный — элементы беспорядка, принимающие участие в тепловом движении и релаксирующие к термодинамическому равновесию.

Гетерополимер — полимер с разными сортами звеньев.

А.Ю. Гроссберг. Институт биохимической физики, Россия, 117977 Москва
Department of Physics and Center for Materials Science and Engineering,
Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, USA. E-mail: shura@gels.mit.edu

Статья поступила 3 сентября 1996 г.

Гомополимер — полимер, цепи которого состоят из одинаковых звеньев.

Конструирование последовательностей — процесс целенаправленного формирования последовательности звеньев гетерополимера так, чтобы полимер приобретал желаемую конформацию (подобно самоорганизации конформаций белков).

Конформация — пространственная форма полимерной цепи; в статистической механике полимеров играет роль микросостояния, статистическая сумма полимерной цепи представляет собой сумму по конформациям.

Модель независимых взаимодействий (МНВ) — модель гетерополимера, в которой энергия взаимодействия каждой пары мономеров считается независимой от остальных пар случайной величиной.

Модель случайных энергий (МСЭ) — модель неупорядоченной системы, в которой энергия каждого микросостояния считается независимой от остальных энергий случайной величиной.

Сворачивание белков — самоорганизация уникальной нативной пространственной структуры макромолекулы белка.

Стеклование — фазовый переход в гетерополимерной цепи между двумя глобулярными (т.е. компактными) фазами, в одной из которых, подобно глобуле гомополимера, экспоненциально много конформаций участвуют в равновесии, а в другой равновесие определяется одной или немногими конформациями.

Узел — конформация кольцевого замкнутого полимера в трехмерном пространстве.

— . . —

За 45 лет своей научной деятельности И.М. Лифшиц плодотворно работал во многих областях физики конденсированного состояния, но теория неупорядоченных систем принадлежит, несомненно, к самой сердцевине его научного наследия. В частности, и к полимерам он пришел со стороны биополимеров (см. название его первой работы о полимере [1]), и в работах о полимерах он всегда подчеркивал аспект неупорядоченности, неполного равновесия, как, например, в его концепции линейной памяти. В этом, кстати, подход И.М. Лифшица отличался от подходов, например, С.Ф. Эдвардса [2] и П.Ж. де Жена [3], для которых теоретико-полевой аспект и аспект критических явлений были главными. Когда в середине 1960-х гг. несколько видных физиков и математиков стали принимать участие в школах и семинарах по молекулярной биологии и многие из них "переключились на биологию", друзья пытались отговорить И.М. Лифшица от такого переключения. Беспокойство друзей было напрасным: Илья Михайлович никуда не переключался, а нашел настоящий физический подход к биополимерам, и его работы по биополимерам стали "аналитическим продолжением" его "обычной" физики.

Автор этих строк был свидетелем того, что Илья Михайлович буквально до последнего дня своей жизни сохранял жгучий интерес к этой теме и энтузиазм по поводу физического подхода к вопросу о том, что же все-таки позволяет биополимерам функционировать так, как они это делают. В последние годы был достигнут определенный прогресс в попытках ответить на этот вопрос. Представляется совершенно естественным обсу-

дить этот прогресс в выпуске УФН, посвященном И.М. Лифшицу.

1.1. Типы беспорядка в полимерной макромолекуле

На рисунке 1 схематически показаны различные типы беспорядка, которые могут существовать в полимерной макромолекуле и которые будут предметом обсуждения в этой работе. Речь идет о беспорядочных разветвлениях, заузлениях и гетерогенных последовательностях (первичных структурах). В каждом случае необходимо четко различать режимы замороженного и размороженного беспорядка.

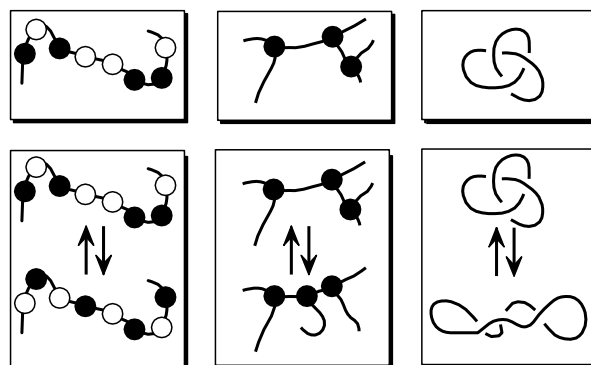


Рис. 1. Типы беспорядка в полимерной макромолекуле. Верхний ряд — замороженный беспорядок, нижний — размороженный. Слева — беспорядочные последовательности, в середине — ветвления, справа — узлы.

Хотя эта концепция хорошо известна и имеет общую значимость для статистической физики, полезно напомнить основные положения:

- 1) замороженные элементы формируются при приготовлении системы и после этого не меняются за время существования системы;
- 2) размороженные элементы принимают участие в тепловом движении.

На теоретическом жаргоне частенько говорят, что разница в том, что надо усреднять: для замороженного случая надо усреднять свободную энергию (что трудно и интересно), для размороженного случая надо усреднять статистическую сумму (что гораздо легче). Хотя это, конечно, верно, но это лишь половина правды (а полуправда — это булгаковская осетрина второй свежести). Прежде всего свободную энергию нужно усреднять не (только) потому, что это трудно, но главным образом из-за принципа самоусреднения [4]: для замороженной системы распределение свободной энергии по реализациям беспорядка близко к нормальному, среднее значение близко к наиболее вероятному и, значит, в среднем доминируют обыкновенные типичные реализации. Напротив, распределение статистической суммы экспоненциально шире, и в нем доминируют очень специальные реализации — те, которые система отбирает сама при размораживании, когда она получает возможность свободно и эргодически перебирать состояния.

1.2. Беспорядок и биополимеры

Воззрение, что усреднять свободную энергию — это единственное, что можно осмысленно делать для замороженной системы, возникло из опыта с "нормальными"

физическими системами вроде спиновых стекол. Действительно, мы никогда не можем экспериментально управлять всеми деталями при приготовлении образца спинового стекла, мы можем только контролировать какие-то грубые средние статистические характеристики вроде среднего количества парамагнитных атомов на единицу объема и т.п. Биологические системы дают физике совершенно другие возможности. Например, последовательность звеньев в цепной молекуле белка играет роль элемента замороженного беспорядка, поскольку последовательность остается неизменной и независимой от теплового движения, пока молекула белка существует как таковая. Используя синтетический аппарат живой клетки, экспериментатор может приготовить макроскопическое количество точно *идентичных* копий этой неупорядоченной системы. Это напоминает фантастическую способность точно воспроизвести после отжига пространственное расположение парамагнитных центров в образце спинового стекла. Хотя мы только вскользь коснемся этих вопросов в данной статье и в основном будем усреднять свободную энергию, надо иметь в виду, что другие постановки вопросов тоже возможны и осмысленны для биологических систем.

По существу, речь идет о том, что биополимеры, хотя и, безусловно, замороженные, но не вполне "неупорядоченные". Действительно, то, что является замороженным в структуре данного биополимера, само по себе на большем масштабе времени есть продукт (биологической и/или предбиологической) эволюции. Это ведет к глубоким физическим вопросам, некоторые из них мы обсудим ниже.

Львиная доля обзора посвящена гетерополимерам, которые считаются хорошей моделью для изучения сворачивания белков. Представляется целесообразным сделать здесь несколько предварительных замечаний по этому поводу.

1.3. Гетерополимеры и проблема сворачивания белков

Явление сворачивания белков состоит в способности цепочки белка к ренатурации: одиночная макромолекула белка в подходящих условиях (температура, pH и т.п.) в растворе способна сама найти "правильную" нативную трехмерную конформацию. Многие белки (хотя и не все) не нуждаются для этого ни в какой помощи и способны к такой самоорганизации в разбавленном растворе [5]. Это явление представляет серьезный вызов для физики (и физиков). Это видно наиболее ясно из следующего *парадокса Левинтала*. Полимер с N мономерами имеет порядка $\exp(\omega N)$ разных конформаций (ω — число порядка единицы). Чтобы перебрать их все, нужно астрономическое время. Как же умудряется цепочка найти правильную конформацию, если правильная конформация только одна, а полный перебор всех конформаций невозможен?¹

За последнее десятилетие в теоретическом понимании этой проблемы был достигнут заметный прогресс. Он связан, в основном, с концепцией "стеклования" гетерополимера: это есть фазовый переход цепочки гетерополимера между двумя глобулярными фазами, из которых

одна представлена экспоненциально большим числом конформаций ($O(\exp(N))$), тогда как в другой всего несколько из них ($O(1)$) дают заметный вклад. Для описания этого перехода используются идеи и концепции, заимствованные из статистической физики спиновых стекол, прежде всего, модель случайных энергий (МСЭ), предложенная Дерридой [6].

Имеются два подхода, в известном смысле параллельных. Один из них восходит к известной работе Брингельсона и Вулинеса [7], он использует достаточно простую математику, но, к сожалению, недостаточно ясен в том, что касается формулировки рассматриваемой модели. Не удивительно, что в этом подходе очень трудно, если не совсем невозможно, найти условия применимости результатов или разработать какие-либо обобщения (см. также [8]). Другой подход, начавшийся с известной работы Шахновича и Гутина [9], рассматривает безукоризненно ясно сформулированную модель, но, к сожалению, использует столь громоздкий аппарат, что он затеняет физическое существо дела. Неудивительно, что эта теория, хотя и считается очень важной, остается мало известной, кроме лишь некоторых качественных выводов.

Обе названные теории, как показывает анализ, глубоко связаны с МСЭ, которая, в свою очередь, очень проста как по физической природе, так и по математическому аппарату. Следовательно, возможно разработать подход, сочетающий простоту работы Брингельсона–Вулинеса со строгостью теории Шахновича–Гутина. Такой подход был недавно найден [10], и он будет изложен в настоящем обзоре.

В первых моделях гетерополимеров последовательность звеньев (или первичная структура) считалась, конечно, случайной. Кроме теоретической простоты, такое предположение казалось разумным также и потому, что статистически последовательности белков выглядят достаточно похожими на случайные [11]. Больше того, простые оценки времени и материала эволюции показывают, что последовательности и не могли очень далеко уйти от случайных. Тем не менее при ближайшем рассмотрении выяснилось, что случайные последовательности не могут все же служить сколько-нибудь разумной моделью белков. В частности, хотя в ансамбле цепочек со случайными последовательностями есть много таких, у которых основное состояние не вырождено (и, значит, уникально), это состояние не "грубое" — при малейшем ничтожном изменении параметров взаимодействия мономеров (например, за счет изменения условий в растворителе) основное состояние совершенно кардинально меняется. Больше того, для большинства последовательностей с уникальной пространственной структурой сворачивание в эту структуру оказывается кинетически не быстрым и не надежным. Все это, конечно, не так уж удивительно, потому что эволюция ведь все-таки что-то делала с белками. На качественном (словесном?) уровне было предположено [7], что результаты эволюции можно сформулировать как "принцип наименьшей фрустрации". Реальные модели процедур конструирования последовательностей, претендующих быть моделью эволюционно отобранных белков, были предложены совсем недавно: Шахнович и Гутин предложили в рамках компьютерного эксперимента разморозить последовательность при замороженной конформации и провести оптимизацию энергии

¹ Есть белки, которым для сворачивания требуется помощь, и клетка умеет такую помощь оказать. С точки зрения физики, однако, важно, что есть белки, сворачивающиеся сами. Объяснить это — вот в чем первоочередной вопрос.

методом Монте-Карло [12]; другую реализацию очень похожей идеи (предложенной независимо) можно назвать "моделью отпечатка" [13]. Довольно интересно, что практически сразу же, как только эти модели были предложены, на их основе было сделано предсказание о том, какого характера отклонение от случайности должно существовать в белковых последовательностях, и, когда стало ясно, что именно нужно искать, предсказанные корреляции были немедленно найдены [14]. Все эти идеи будут обсуждаться в данной работе.

Заметим, что проблема стеклования стоит не только для белков, но и для синтетических гетерополимеров и привлекает большое внимание в связи с потенциалом промышленных и биомедицинских приложений. Пытаясь понять белки, мы, возможно, добьемся какого-то прогресса и в направлении искусственных *белковоподобных* полимеров.

1.4. Замечание о терминологии

Хотя английская терминология по этому вопросу установилась и стала общепринятой (замороженный и размороженный случаи называются соответственно *quenched* и *annealed*), соответствующая русская терминология, видимо, еще формируется. По поводу термина "замороженный беспорядок" разногласий, судя по всему, нет, но тот беспорядок, который в этой статье называется размороженным, иногда называют также динамическим, подвижным, плавающим или "беспорядком при отжиге". Использовать последний термин в применении к полимерам и вообще к "мягким" конденсированным системам представляется особенно нелепым. Все другие термины, не исключая и предлагаемое здесь слово "размороженный", кажутся тоже довольно неуклюжими. С сожалением жертвуя изяществом стиля, автор надеется, что данное выше объяснение, по крайней мере, четко указывает, о чем идет речь.

Другая терминологическая проблема связана с названием фазового перехода, который выше был описан как стеклование. Поскольку большинство читателей знакомо с англоязычной литературой, имеет смысл отметить, что установившееся английское название этого явления *heteropolymer freezing transition*. Этот переход назван здесь стеклованием, хотя это слово очень перегруженное и его использование часто только множит ложные ассоциации и недоразумения. Единственный выход состоит в том, чтобы постараться возможно точнее описать, что подразумевается.

1.5. Что не рассматривается в этом обзоре

Есть много других примеров неупорядоченных полимерных систем и явлений, которые здесь не затрагиваются: полимерные гели (см. недавнюю замечательную работу Панюкова и Рабина [15]), блок-сополимеры с их способностью образовывать доменные структуры (см. работы [16–18] и цитированную в них литературу), полимеры в неупорядоченных внешних полях, адсорбция гетерополимеров на поверхности [19–25] или на селективной фазовой границе [26–29], плавление гетерополимерной ДНК [30–32], неравномерное распределение гибкости вдоль ДНК [33, 34], полиамфолиты [35–39] и многие другие. Коллапс размороженного гетерополимера был рассмотрен в работе [40], недавно результаты были вновь получены в [41]. Нельзя объять необъятное, в настоящей

статье мы сконцентрируемся на разветвленных полимерах, узлах и, в основном, на гетерополимерах.

2. Беспорядок разветвлений

Разветвленные полимеры представляют значительный интерес как для химии высокомолекулярных соединений, так и для биофизики. Подавляющее большинство синтетических полимеров в той или иной степени разветвленные. Молекулы РНК образуют структуры типа так называемого клеверного листа; в сущности, — это разветвленный полимер с кусками спиралей в качестве ветвей. Еще лучший пример — это суперспирализованная ДНК [42, 43]; об этой системе читатель может получить неплохое представление, покрутив телефонную трубку и глядя, как шнур скручивается сам собой (как обычно, наиболее интересное поведение наблюдается, к сожалению, только в области значительного риска для исправности телефона).

2.1. Формула Зимма–Штокмайера

Для начала мы пренебрежем объемными взаимодействиями и выясним, чему равен размер идеального случайно разветвленного полимера R_{id} . Ответ на этот вопрос был дан в 1949 г. в работе Зимма и Штокмайера [44]. Их результат имеет вид $R_{id} \propto N^{1/4}$. Есть несколько способов вывести эту формулу; один из них описан в книге [32]. Он основан на оценке "химического диаметра" разветвленной структуры, т.е. химического интервала или контурного расстояния между двумя произвольными концами структуры. Это делается путем помещения разветвленной структуры на специальном графе, который называется деревом Кэли (рис. 2). Выбирая на дереве произвольную начальную точку O , разветвленную структуру можно представить как одномерное случайное блуждание с возможными направлениями

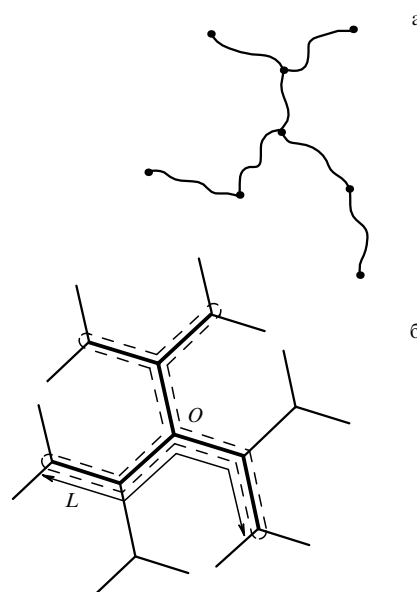


Рис. 2. Разветвленная структура без циклов (а) может быть размещена на дереве Кэли (б). На графе образ исходного разветвленного полимера окружен "кольцевым полимером" (штриховая линия), его характерное расстояние от начальной точки O пропорционально химическому диаметру структуры L .

элементарных шагов "к" и "от" точки O . Это приводит к оценке, что химический диаметр L пропорционален квадратному корню из числа N мономеров в макромолекуле; более аккуратно, результат имеет вид

$$L \propto g \sqrt{\frac{N}{g}} \propto \sqrt{Ng}, \quad (1)$$

где g — характерное число мономеров в линейной части между двумя соседними разветвлениями. Так как каждый диаметр представляет собой просто гауссову линейную цепочку, то получается

$$R \sim aL^{1/2} \sim a(Ng)^{1/4}, \quad (2)$$

где a — характерный мономерный масштаб.

Таким образом, в трехмерном пространстве ($d = 3$) разветвленный полимер очень компактный и плотный, поэтому эффект исключенного объема очень сильный.

2.2. Замороженные и размороженные разветвления

Говоря об эффекте исключенного объема для разветвленных полимеров, необходимо различать два предельных случая. Чтобы объяснить их, представим себе разветвленный полимер как древоподобный граф, погруженный в реальное пространство. В одном предельном случае структура этого графа зафиксирована, в тепловое равновесие ни с чем не приходит, и тепловое движение касается только того, как граф погружен в пространство — веточки извиваются, но не перестраиваются. В противоположном пределе как структура графа, так и его пространственная форма флуктуируют, участвуют в тепловом движении и/или равновесии. Первый случай отвечает замороженным, второй — размороженным ветвлениям. Оказывается, что замороженный и размороженный разветвленные полимеры принадлежат двум разным классам универсальности и характеризуются разными критическими показателями.

Физическая причина такого различия между замороженными и размороженными ветвлениями проста. Для замороженного случая исключенный объем ведет только к растяжению линейных участков макромолекулы, так как для полимера это единственная возможность набухать. Напротив, в размороженном случае дополнительное набухание можно осуществить за счет перестроек системы разветвлений. Это изменяет характер соответствующей конформационной энтропии и, соответственно, класс универсальности задачи [45].

2.3. Теория типа Флори для замороженного разветвленного полимера

Теории флориевского типа никогда не претендуют на точный результат, но они подкупающе просты. Кроме того, известно, что для линейного полимера теория Флори дает очень хорошее приближение для величины критического показателя. Теория Флори основана на предположении, что размер полимера R определяется уравниванием таких двух факторов, как энтропийная упругость полимера и расталкивание мономеров; другими словами, R определяется минимизацией свободной энергии

$$F(R) = F_{\text{elast}}(R) + F_{\text{int}}(R), \quad (3)$$

где $F_{\text{elast}}(R)$ и $F_{\text{int}}(R)$ означают, соответственно, вклады полимерной энтропии и энергии объемных взаимодей-

ствий. Что касается взаимодействий, соответствующий вклад можно легко написать во втором вириальном приближении так же, как это делается для линейных полимеров:

$$F_{\text{int}} \sim T \frac{vN^2}{R^d}, \quad (4)$$

где v — исключенный объем мономера. Предположим пока, что $v \approx a^d$.

В полной аналогии с теорией Флори для линейного полимера представляется естественным записать энтропийный вклад в виде

$$F_{\text{elast}}(R) = T \frac{R^2}{R_0^2}, \quad (5)$$

где $R_0 \sim a(gN)^{1/4}$ — невозмущенный размер полимера. Из уравнений (4) и (5) автоматически вытекает

$$R \sim aN^{v_{\text{qu}}} g^{\mu_{\text{qu}}} \quad (6)$$

с критическими показателями

$$v_{\text{qu}} = \frac{5}{2(d+2)}, \quad \mu_{\text{qu}} = \frac{1}{2(d+2)}. \quad (7)$$

Этот результат был получен в работах [46, 47]. Очевидно, что вывод этого результата опирается на предположение о том, что характер разветвлений не меняется с набуханием (например, потому что невозмущенный размер зависит, вообще говоря, от степени разветвленности). Это предположение верно только для замороженного случая. Любопытно, что это обстоятельство не было даже отмечено в [46, 47].

2.4. Теория типа Флори для размороженного разветвленного полимера²

Чтобы принять во внимание возможность дополнительных перестроек ветвей, необходимо заметить, что характерный химический диаметр L удовлетворяет уравнению Зимма–Штокмайера (1) только для невозмущенного полимера. Для полимера с исключенным объемом и размороженной структурой перестройка ветвлений может изменить число мономеров между концами.

Для описания этого фактора мы снова прибегнем к теории Флори, на этот раз применим ее к полимеру, расположенному на дереве Кэли (см. рис. 2). Сколько есть конфигураций с химическим диаметром L ? Как уже упоминалось, все эти конфигурации можно отобразить на траектории одномерных случайных блужданий (с шагами к точке O и от нее вдоль дерева Кэли), поэтому ответ на поставленный вопрос дается энтропией линейного полимера растянутого так, что его расстояние между концами по порядку величины равно L . Действительно, если дерево Кэли — это пространство, то L играет роль пространственного размера. Таким образом, мы получаем

$$F_{\text{br}} = T \frac{(L/g)^2}{N/g} = T \frac{L^2}{Ng}. \quad (8)$$

Теперь наша теория Флори похожа на матрешку. Во внутренней куколке имеются два флориевских вклада в

² Материал этого раздела основан на работе [45].

упругую часть свободной энергии:

$$T \frac{R^2}{a^2 L} + T \frac{L^2}{Ng}. \quad (9)$$

Любопытно, что оба члена здесь одного и того же типа, но L играет роль длины цепи в первом члене и пространственного размера — во втором. Минимизируя по отношению к L , получаем [45]

$$L \sim \left(\frac{R}{a}\right)^{2/3} (Ng)^{1/3} \quad (10)$$

и

$$F_{\text{elast}}^{\text{ann}} \sim T \frac{(R/a)^{4/3}}{(Ng)^{1/3}}. \quad (11)$$

Что же касается внешней куколки нашей матрешки, то мы должны теперь минимизировать свободную энергию

$$F = F_{\text{int}} + F_{\text{elast}}^{\text{ann}} \sim Tv \frac{N^2}{R^d} + T \frac{(R/a)^{4/3}}{(Ng)^{1/3}} \quad (12)$$

по отношению к R . Это дает

$$R \sim a N^{v_{\text{ann}}} g^{\mu_{\text{ann}}} \quad (13)$$

с

$$v_{\text{ann}} = \frac{7}{3d+4}, \quad \mu_{\text{ann}} = \frac{1}{3d+4}. \quad (14)$$

Интересно также записать соотношение между L и N , которое включает два новых критических показателя ρ и σ :

$$L \sim N^\rho g^\sigma, \quad \rho = \frac{d+6}{3d+4}, \quad \sigma = \frac{d+2}{3d+4}. \quad (15)$$

Очевидно, все это может быть справедливо только при условии $L < N$, или

$$g < g^* = N^{(1-\rho)/\sigma} \sim N^{2(d-1)/(d+2)}; \quad (16)$$

если $g > g^*$, то полимер настолько слабо разветвлен, что уже небольшая перестройка делает его практически линейным, с $L \simeq N$. Неудивительно поэтому, что для случая $g = g^*$ уравнение (13) дает

$$R \sim a N^{v_{\text{lin}}} \sim a N^{3/(d+2)}$$

точно так же, как уравнение (6) для замороженного случая при $g = N$.

2.5. Условия приготовления полимера для замороженного случая

Если подумать чуть глубже о замороженных разветвленных структурах, то становится ясно, что они могут быть очень разными: от почти линейного регулярного гребня с $L \approx N/2$ до регулярно ветвящейся структуры типа Кэли с $L \sim \ln N$. Полученный выше результат (6) должен быть справедлив для случайно разветвленной структуры, потому что для подавляющего большинства структур верна оценка (1) и для них $L \sim N^{1/2}$. Тем не менее можно действительно приготовить разные разветвленные структуры и естественно поставить вопрос об их размерах. Простой способ подойти к этому вопросу — сказать, что приготовление разветвленной структуры состоит в том, что ее в размороженном состоянии помещают в один растворитель, потом замораживают и переносят в другой.

Чтобы такую ситуацию рассмотреть, вспомним, что исключенный объем зависит на самом деле от природы и типа растворителя. Обозначим исключенный объем в растворителе приготовления и в растворителе наблюдения (где измеряется размер замороженного полимера) соответственно $v = \tau_p a^d$ и $v = \tau a^d$. В общем случае для замороженной структуры с произвольным химическим диаметром L флориевская свободная энергия (4) и (5) дается формулой

$$\frac{R^2}{La^2} + \frac{vN^2}{R^d}, \quad (17)$$

что после минимизации приводит к

$$R \sim a(\tau L N^2)^{1/(d+2)}. \quad (18)$$

Что касается L , грубо говоря, эта величина дается уравнением (15), потому что в условиях приготовления наш полимер — размороженный. Более аккуратно, проследив за фактором τ_p , мы получаем

$$L \simeq \tau_p^{2/(3d+4)} N^\rho g^\sigma \quad (19)$$

и, таким образом,

$$\begin{aligned} R &\simeq \tau_p^{2/(3d+4)(d+2)} \tau^{1/(d+2)} N^{v_{\text{ann}}} g^{\mu_{\text{ann}}} \simeq \\ &\simeq \left(\frac{\tau_p}{\tau}\right)^{2/(3d+4)(d+2)} R_{\text{ann}}. \end{aligned} \quad (20)$$

Как и следовало ожидать, критический показатель, описывающий зависимость от N , оказался таким же, как для размороженного случая. Впрочем, это верно лишь до тех пор, пока τ и τ_p не зависят от N . Если же использовать для приготовления полимера, например, растворитель, в котором исключенный объем очень большой (τ_p велико), то можно получить структуры с набуханием весьма отличным как от простого замороженного, так и от размороженного случаев.

Основной смысл данного маленького раздела состоит в том, что размораживание можно использовать для приготовления различных замороженных конфигураций. Мы вернемся к этой идее в применении к гетерополимерам.

2.6. Обсуждение и сравнение с данными других теоретических подходов и с методом Монте-Карло

Можно легко построить графики зависимостей критических показателей v от размерности пространства d как для замороженного и размороженного случаев, так и для линейного полимера (для которого $v = 3/(d+2)$). Известно [46], что $d = 8$ — верхняя критическая размерность для разветвленных структур. Неудивительно поэтому, что теория Флори дает невозмущенные значения показателей v при $d = 8$ как для замороженного, так и для размороженного случаев.

Для $d = 3$ известен точный теоретико-полевой результат [48] для размороженного случая: $v_{\text{ann}} = 1/2$. Теория Флори дает это же значение $1/2$, но для замороженного случая $v_{\text{qu}} = 0,5$; для размороженного режима она дает $v_{\text{ann}} = 7/13 \approx 0,54$. Это характеризует степень точности подхода типа Флори. Интересно, что ни в одной из работ [46, 47] даже не упоминалось, что в них рассматривается на самом деле только замороженный режим; аналогично, в работе [48] не было сказано, что в ней речь идет

о размороженной ситуации. Считалось поэтому, что результаты этих работ замечательно согласуются друг с другом, что, конечно, бессмысленно (см. также [49, 50]).

Что важно — это то, что всегда $v_{qu} < v_{ann}$. Это ясно физически: размороженный полимер имеет некоторую дополнительную свободу и может ее использовать для того, чтобы ослабить "тяготы" исключенного объема. Это согласуется и с результатами первого порядка ($\epsilon = 8 - d$)-разложения:

$$v_{ann} \simeq \frac{1}{4} \left(1 + \frac{\epsilon}{9} + \dots \right), \quad v_{qu} \simeq \frac{1}{4} \left(1 + \frac{\epsilon}{10} + \dots \right)$$

[51, 52].

Проведенное недавно компьютерное моделирование [52] дало результаты $v_{ann} = 0,49 \pm 0,01$ и $v_{qu} = 0,45 \pm 0,01$. Эти результаты сходны с теорией Флори в том, что касается разницы между этими двумя показателями.

Различие между замороженным и размороженным разветвленными полимерами проявляется в множестве разных физических обстоятельств. Разными оказываются осмотическое давление полуразбавленного раствора, проницаемость через узкие капилляры и т.д. Многие из этих эффектов можно описать с помощью стандартной техники скейлинга [3] (см. [45]).

Для линейного полимера кроме набухания за счет исключенного объема и самоизбегания есть и противоположный режим коллапса, или перехода клубок — глобула. Это существенно при $d > 2$, потому что при $d = 2$ гауссов полимер имел бы не зависящую от N плотность порядка единицы, и ему некуда было бы коллапсировать. Аналогично, для разветвленного полимера коллапс существует при $d > 4$. Теория Флори может быть обобщена для этого случая, и также более строгая теория [53] может быть разработана по образу и подобию теории И.М. Лифшица для перехода клубок — глобула в линейном полимере [1].

3. Топологический беспорядок: узлы и зацепления

3.1. При чем здесь топология?

Естественный образ для полимера — это веревка или нитка. Эта аналогия наталкивает на мысль, что полимеры должны запутываться. Это действительно происходит. Проявления этого факта многочисленны — от вязкоупругости полимерных жидкостей до специальных ферментов, созданных в живой клетке для распутывания ДНК.

Насколько известно автору, самые ранние работы, связанные с узлами в физике, принадлежат перу Вильяма Томсона (позже — лорд Кельвин) [54], в их обсуждении горячее участие принимал Максвелл. Речь шла о попытке ответить на вопрос, откуда берется дискретность химических элементов. Идея состояла в том, чтобы каждому химическому элементу поставить в соответствие узел, и тогда дискретность получалась бы естественно. Хотя мы теперь знаем ответ квантовой теории на этот вопрос, идее по-прежнему нельзя отказать в привлекательности. В недавнее время различные концепции, связанные с узлами, нашли применение в статистической физике, теории поля, гидродинамике, магнитной гидродинамике, астрофизике и т.д. [55–58]. В этих областях, однако, узлы появляются как плоды достаточно

абстрактных математических построений. Напротив, узлы в полимерах — это самые обыкновенные узлы в простом механическом смысле слова, как на веревке. В контексте полимеров и биополимеров узлы впервые обсуждались в работах Фриша и Вассермана [59] и Дельбрюка [60].

Все без исключения топологические свойства полимеров возникают из того факта, что две цепочки, или два куска одной цепи, пересекаясь, не могут пройти друг сквозь друга. Возьмем теперь кольцевой полимер: раз самопересечения невозможны, то остающийся набор конформаций, которые система эргодически перебирает в ходе теплового движения — это те, которые можно получить друг из друга путем *непрерывных* деформаций, а это точно подпадает под математическое определение топологии. Таким образом, кольцевой полимер может произвольно менять свою геометрическую форму, но с условием сохранения неизменного топологического класса.

Надо сказать, что, вообще, слово "топология" часто (и по непонятной для автора причине) используется в физике без всякой связи с топологией, как таковой (например, об общем виде белковой структуры говорят почему-то, как о топологии белка). В данном случае это не так, и речь идет действительно о топологии.

Ясно, что топологические ограничения очень сильно влияют на все (как статические, так и динамические) свойства полимеров. Многие молекулы ДНК работают в клетке в виде замкнутых колец. Недавно узлы были найдены в некоторых белках [61].

В математике словом узел обозначается произвольная замкнутая кривая, погруженная в трехмерное пространство. То, что мы бы нормально назвали кольцом без узла, называется тривиальным узлом. Аналогично, зацепление — это несколько колец, погруженных в одно и то же пространство, и тривиальное зацепление — это такое, в котором кольца друг с другом реально не зацеплены. Ясно, что эти концепции прямо и полностью применимы к полимерам.

Заслуживает упоминания и то, что макроскопическая полимерная сетка, составляющая полимерный гель (то же самое, что желе), дает пример экстремально сложной топологии.

3.2. Топология и беспорядок

Топологические свойства полимеров похожи на свойства неупорядоченных систем в том отношении, что топология, скажем, полимерного кольца определяется в момент его приготовления и после этого накрепко запоминается системой. Как и в других случаях, необходимо различать случаи замороженного и размороженного беспорядка:

1. Замороженный топологический беспорядок существует в обычном полимерном кольце, системе нескольких колец или сетке.

2. Размороженный режим можно реализовать в растворе ДНК в присутствии специального фермента (называемого топоизомеразой II) и достаточного количества молекул АТФ, нужных для работы этого фермента.

3. В системе линейных (незакольцованных) полимеров топология не накладывает строгих ограничений, но перестройки "запутанности" происходят медленно, поэтому временные топологические ограничения отчасти похожи на размороженный беспорядок.

3.3. Общая формулировка топологической задачи в физике полимеров

Сформулируем теперь в самом общем виде, из каких этапов должно состоять решение задачи о топологических свойствах полимеров. Начнем со случая замороженной топологии. Статистическая сумма, скажем, полимерного кольца может быть записана в виде

$$Z = \int_{\text{контуры данной топологии}} \exp\left(-\frac{E}{T}\right) d\Gamma = \int \exp\left(-\frac{E}{T}\right) \delta(G(\Gamma) - G_0) d\Gamma, \quad (21)$$

где интегрирование по $d\Gamma$ означает суммирование по конформациям. В первой строчке это суммирование ограничено конформациями определенного топологического типа, а во второй суммирование формально происходит по всем конформациям, но на самом деле нужная топология отбирается δ -функцией: $G(\Gamma)$ — это такая величина, которая имеет различные значения для конформаций Γ , если они топологически различны, но остается неизменной, пока топология конформации не изменяется. Такая величина называется топологическим инвариантом.

Таким образом, проблема в целом должна решаться в два шага: во-первых, нужно найти подходящий топологический инвариант; иными словами, нужно классифицировать узлы; во-вторых, когда топологический инвариант построен, нужно найти статистическую сумму (21).

В простейшем случае, если между частями полимера вообще нет объемных взаимодействий (кроме, конечно, совсем короткодействующего отталкивания, которое, собственно, и запрещает пересечения), последняя проблема сводится к вопросу о доступном для заданного узла $\mathcal{K}$ фазовом объеме, или к вопросу об *энтропии узла* $S(\mathcal{K})$. Этот вопрос, в свою очередь, эквивалентен вопросу о вероятности получить данный узел при случайном замыкании кольца: $p(\mathcal{K}) \sim \exp[S(\mathcal{K})]$.

3.4. Коротко об инвариантах узлов

Первая попытка классификации узлов была предпринята Тэйтом (P.G. Tait) [77]. Он просто рисовал узлы и старался (с успехом) подметить закономерности. В наше время значительные усилия затрачиваются на более систематическое рисование узлов и составление таблиц узлов (см. [76, 78]). Обычно в этих таблицах узлы группируются в соответствии с минимальным числом пересечений $n(\mathcal{K})$, которые может иметь двумерная проекция данного узла $\mathcal{K}$. К сожалению, число возможных узлов растет экспоненциально с n , и поэтому неудивительно, что "даже наиболее мощные компьютеры быстро теряют энтузиазм" ("even the most powerful computers quickly run out of enthusiasm") [79]. Другое направление развития связано с идеями алгебраической топологии и восходит к Александру [80]. В последние годы на этом пути был достигнут замечательный прогресс, были изобретены новые полиномиальные топологические инварианты, включая полиномы Джонса, Кауфмана, Васильева и др. [81], и, вообще, проблема классификации узлов стала гораздо яснее [57].

Хотя новые полиномы гораздо сильнее полиномов Александра, они гораздо менее удобны для вычислений. Для узла с n пересечениями на проекции необходимое число элементарных компьютерных операций растет как

n^3 для полиномов Александра и обычно как $\exp n$ для других полиномов. Считалось, что это делает применение новых полиномов практически невозможным. Выход из этой трудности был предложен в недавней работе [72], ее результаты обсуждаются ниже.

Еще один подход к классификации узлов связан с идеей так называемой "энергии узла": представим себе, что наш полимер равномерно заряжен вдоль всего контура, и будем постепенно увеличивать плотность заряда. Поскольку заряды одноименные, полимер будет растягиваться и в конце концов примет какую-то форму, которая отвечает максимальному растяжению совместимому с заданной топологией (контурная длина полимера при этом, конечно, предполагается неизменной). Правдоподобно, что эта максимально растянутая форма является инвариантом узла — зависит от его топологии, но не от конкретной геометрической формы. Более простой инвариант есть просто кулоновская энергия этой максимально растянутой формы:

$$E = \min_{\mathbf{r}(s)} \left\{ \frac{1}{L} \oint \oint ds ds' \left[\frac{1}{|\mathbf{r}(s) - \mathbf{r}(s')|} - \frac{1}{|s - s'|} \right] \right\}, \quad (22)$$

где s — есть длина вдоль контура полимера, L — полная длина полимера, и вектор-функция $\mathbf{r}(s)$ задает подлежащую оптимизации форму контура. Второй член под интегралом предназначен для того, чтобы ликвидировать паразитическую расходимость кулоновской энергии при $s \rightarrow s'$; множитель $1/L$ введен так, чтобы сделать результат не зависящим от линейного масштаба, а только от топологии.

Имеются и другие попытки построить разные "энергии узлов", но они совсем не так наглядны, как формула (22).

Еще одна идея для классификации узлов была выдвинута совсем недавно, она связана с возможностью "максимально надутого" [73] или "идеального" [82] представления узла. Мы обсудим это более подробно в разделе 3.7.

3.5. Что известно об энтропии узла?

Несмотря на прогресс в классификации узлов и математической теории топологических инвариантов, имеется очень мало строгих результатов относительно энтропийных свойств узлов [62, 63], и дальнейшие попытки в этом направлении встречают серьезные математические трудности [64–66]. В этой ситуации приходится апеллировать либо к упрощенным точно решаемым моделям [67–69], либо к компьютерному моделированию [70–72], либо искать какой-то упрощенный подход типа теории Флори [73]. Что касается точно решаемых моделей, таких как наведение полимера на одиночное бесконечное прямолнейное препятствие или удержание полимера в решетке препятствий, они подробно описаны в работах [74, 75]. Превосходный обзор о компьютерном моделировании узлов был опубликован в УФН [76]. Поэтому здесь мы лишь кратко коснемся новых компьютерных результатов и перейдем к теории Флори.

3.5.1. Для длинной цепи вероятность тривиального узла экспоненциально убывает с числом сегментов. Один из немногих строгих результатов для энтропии узла был получен в работах [62, 63]. Рассмотрим ломаную линию из N сегментов, каждый сегмент единичной длины. Это, конечно, модель свободно сочлененного полимера без

исключенного объема. Чему равна вероятность $\mathcal{P}_0(N)$ того, что при случайном соединении концов ломаной получится тривиальный узел ($\equiv$ кольцо без узлов)? Результат работ [62, 63] состоит в том, что названная вероятность экспоненциально убывает с ростом N :

$$\mathcal{P}_0(N) \propto \exp\left(-\frac{N}{N_0}\right), \quad (23)$$

где N_0 — соответствующий характерный масштаб (это просто ни от чего не зависящее безразмерное число, вроде мировой постоянной)³.

Хотя строгое доказательство соотношения (23) не простое, можно проиллюстрировать его рассуждением, поясняющим, почему вероятность (23) спадает по экспоненте. Поделим наш полимер из N сегментов на N/g блоков по g сегментов каждый. Чтобы не было узлов на всей цепи, их не должно быть ни в одном блоке, и это дает

$$\mathcal{P}_0(N) \leq [\mathcal{P}_0(g)]^{N/g},$$

что и приводит к экспоненциальной зависимости от N .

Важное следствие результата (23) состоит в том, что стандартная гауссова модель полимерной цепи, представляющая полимер как траекторию броуновского движения, абсолютно неприемлема для топологических задач. Действительно, броуновская траектория должна рассматриваться как предел ломаных, когда длина сегмента стремится к нулю, $l \rightarrow 0$, а число сегментов стремится к бесконечности, $N \rightarrow \infty$, но так, что контурная длина остается неизменной: $L = Nl = \text{const}$. Естественно, вероятность узла зависит от N , а не от L , поэтому в броуновском пределе вероятность узла равна единице. Это обычная "ультрафиолетовая катастрофа" — бесконечное запутывание броуновской траектории на бесконечно малых масштабах. Для осмысленной постановки вопроса об энтропии или вероятности узла необходима какая-то "гранулярность" — сегменты конечной длины, или конечная толщина, или система блужданий по решетке вместо непрерывного пространства.

Чему равна величина N_0 ? В настоящее время нет никаких способов даже оценить ее аналитически, и приходится обращаться к данным моделирования.

3.5.2. Некоторые данные метода Монте-Карло. Есть довольно много работ о компьютерном моделировании узлов, начиная с пионерской работы [70], включая статью [71] и самую новую из известных автору в настоящее время работ [72] (см. также прекрасный обзор [76]). Общая схема моделирования проста:

1. Сначала нужно сгенерировать замкнутые петли либо на какой-нибудь решетке, либо в виде цепи сегментов. Чтобы заставить концы встречаться есть специальный алгоритм, выбирающий следующий шаг блуждания из *условного* распределения вероятностей с условием, что целая траектория представляет собой петлю (или броуновский мост).

2. Когда петля готова, нужно определить, что это за узел. Эта часть требует наибольших вычислений. В большинстве случаев используется полином Александра, но в работе [72] был впервые применен полином Васильева.

Некоторые данные суммированы на рис. 3. Раньше данные приводились для вероятности нетривиального узла $\mathcal{P}(N) = 1 - \mathcal{P}_0(N)$, так как само появление узлов считалось удивительным. Теорема (23) показывает, что узлы, наоборот, типичны для длинных цепей. Данные, представленные на рис. 3, очень хорошо согласуются с асимптотикой (23) и приводят к оценке $N_0 \approx 335 \pm 5$.

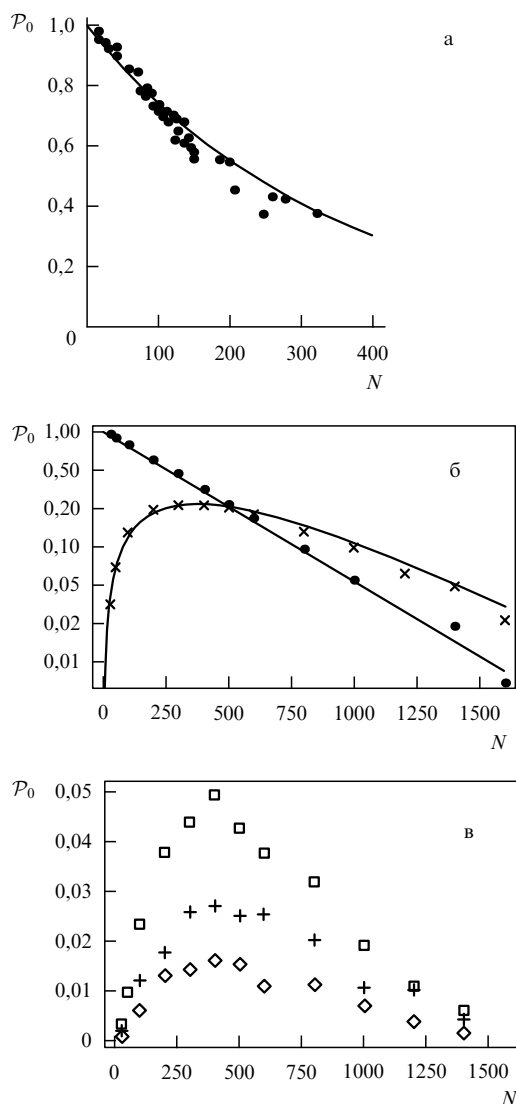


Рис. 3. Данные метода Монте-Карло для вероятности получить тривиальный и некоторые нетривиальные узлы при случайном замыкании полимера из N сегментов нулевой толщины: (а) данные из обзора [76] для тривиального узла, (б, в) данные работы [72]. Интерполяционные кривые построены в соответствии с уравнением (24). Обозначения: $\bullet$ — данные для тривиального узла, $\times$ — для трилистника 3_1 , $\square$ — для 4_1 , $+$ — для 5_1 , $\diamond$ — для 5_2 .

Характерная длина цепи оказывается довольно большой! Остается неясным, почему она так велика. Напрашивается сравнение этого числа с параметром зацеплений N_c , фигурирующим в теории репаций [3, 2]; экспериментальное значение этого параметра обычно лежит в интервале 50–500. Соотношение этих двух чисел остается загадкой.

³ Более точно, утверждение работ [62, 63] такое: существует конечный предел $\lim_{N \rightarrow \infty} [N^{-1} \ln \mathcal{P}_0(N)]$.

Более подробные компьютерные модели включают определение вероятностей для некоторых специальных нетривиальных узлов [72]. Как и следовало ожидать, с ростом числа сегментов растет разнообразие получаемых узлов. Что же до вероятности $\mathcal{P}_K(N)$ получить какой-то определенный тип узла, скажем, K , о ней можно высказать такие догадки:

1. Пока N мало, $\mathcal{P}_K(N) \approx 0$, за исключением случая тривиального узла $K = 0$, поскольку короткая петля вообще не может сформировать нетривиальный узел.

2. В пределе больших N , вероятность $\mathcal{P}_K(N)$ должна убывать экспоненциально, причем с той же характерной длиной N_0 (23). В самом деле, с ростом N любой узел K в конце концов становится "маленьким" и "локальным". Для столь больших N вероятность $\mathcal{P}_K(N)$ представлена в терминах произведения вероятности узла K , локализованного на относительно короткой части цепи N_1 , на вероятность тривиального узла $\mathcal{P}_0(N - N_1)$ для остальной цепи. В пределе больших N последний фактор лидирует, и получается экспоненциальное убывание на масштабе N_0 независимо от узла K .

3. Из двух приведенных аргументов вытекает, что вероятность $\mathcal{P}_K(N)$ должна иметь максимум при какой-то длине цепи N_K , которая зависит, конечно, от типа узла. Разумно предположить, что N_K увеличивается, когда узел K становится более сложным.

В работе [72] эти утверждения нашли полное подтверждение. В пределах точности результатов, было найдено справедливым следующее интерполяционное выражение:

$$\mathcal{P}_K(N) = C_K N^{v(K)} \exp\left(-\frac{N}{N_0}\right). \quad (24)$$

Некоторые численные данные работы [72] приведены в табл. 1, где $N_{\max}$ — длина цепи, для которой вероятность найти данный узел, максимальна, $\mathcal{P}_{\max}$ — соответствующая максимальная вероятность; использованы стандартные обозначения узлов [78].

На основе численных данных [72] было предположено, что для составного узла $K = K_1 \# K_2$ справедливо соотношение $v_K = v_{K_1} + v_{K_2}$. Насколько известно автору, это соотношение не было пока ни доказано, ни опровергнуто.

3.5.3. Роль толщины полимера. Хотя это и не доказано математически, не вызывает сомнений, что почти все конформации (бесконечно) длинного полимерного кольца заузлены — для случая цепочки с исключенным объемом, так же как и для рассмотренного выше случая бестелесной цепи. Однако толщина может существенно повлиять на характерный масштаб заузливания, и это действительно происходит.

Что касается полимеров умеренной длины, толщина сильно подавляет их способность завязывать узлы. Это было показано методом Монте-Карло для свободно сочлененного полимера с сегментами длины l и толщины d [83]. Очевидно, вероятность $\mathcal{P}$ нетривиального или $\mathcal{P}_0 = 1 - \mathcal{P}$ тривиального узла зависит теперь как от $N = L/l$, так и от d/l : $\mathcal{P}_0 = \mathcal{P}_0(N, d/l)$. Численные результаты хорошо интерполируются формулой:

$$\mathcal{P}_0\left(N, \frac{d}{l}\right) \approx \exp\left[-\frac{N}{N_0 \exp(27d/l)}\right]. \quad (25)$$

Как видно, толщина цепи приводит к очень значительному увеличению характерной длины заузливания сверх уже и без того большой величины $N_0 \approx 335$ для бестелесной цепи. Для цепи данной длины это означает, конечно, сильное подавление тенденции к заузливанию.

Сам факт, что толщина подавляет узлы, хорошо установлен, но это не значит, что он хорошо понят. В принципе, два разных эффекта могли бы здесь играть роль. Один состоит в том, что исключенный объем заставляет весь полимер в целом набухать. Второй — в том, что толщина мешает кончику растущей цепи пронизывать маленькие петельки ранее выросшей части цепи. Можно было бы попробовать разделить эти два эффекта, рассматривая бестелесный полимер и заставляя его набухать в подходящем внешнем поле (например, вида $\phi(x) \sim -\chi x^2$, с закреплением одного звена в начале координат). Результаты попыток подобного типа пока не привели к ясному ответу.

3.5.4. Экспериментальное наблюдение узлов. Имеется уже довольно долгая история наблюдения узлов в ДНК начиная с 1976 г. [84]. Недавнее достижение [85, 86] состоит в количественном измерении вероятностей для некоторых конкретных узлов, таких как 3_1 , 4_1 , 5_1 и 5_2 . Для сравнения данных реальных и компьютерных экспериментов было решающе важно, что в последних была принята во внимание толщина цепи. В экспериментах [85, 86] эффективную толщину полимера регулировали просто ионной силой раствора. Экспериментально измеренные зависимости вероятностей узлов 3_1 , 4_1 , 5_1 и 5_2 от ионной силы оказались в очень хорошем согласии с данными метода Монте-Карло для тех же узлов в зависимости от толщины полимера d (о стандартных обозначениях узлов см. [78]).

3.5.5. Роль сжатия полимера. В обычной физике полимеров изучается набухание полимеров с исключенным объемом, но известен и противоположный эффект коллапса, о котором часто говорят, что он вызывается отрицательным исключенным объемом $B < 0$. В действительности, нулевым (в Θ -точке) или отрицательным

Таблица 1. Численные данные для некоторых узлов

Тип узла	v_K	C_K	$N_{\max} = N_0 v_K$	$\mathcal{P}_{\max}$
0	≈ 0	$1,03 \pm 0,03$	≈ 0	1
3_1	$1,11 \pm 0,05$	$(1,1 \pm 0,3) \times 10^{-3}$	372 ± 15	0,26
4_1	$1,34 \pm 0,09$	$(6,6 \pm 3,0) \times 10^{-5}$	450 ± 30	0,06
5_1	$1,35 \pm 0,15$	$(1,9 \pm 1,5) \times 10^{-5}$	450 ± 50	0,02
5_2	$1,40 \pm 0,11$	$(2,2 \pm 1,3) \times 10^{-5}$	470 ± 30	0,03
$3_1 \# 3_1$	$2,3 \pm 0,1$	$(2,9 \pm 2,1) \times 10^{-7}$	770 ± 30	0,13
$3_1 \# 4_1$	$2,5 \pm 0,2$	$(3,4 \pm 4,2) \times 10^{-8}$	840 ± 70	0,06
$3_1 \# 3_1 \# 3_1$	$3,7 \pm 0,2$	$(1,5 \pm 1,5) \times 10^{-11}$	1240 ± 70	0,1

(ниже температуры Θ) может стать соответствующий вириальный коэффициент, но никак не толщина полимера, и это важно при разговоре о топологии. Вопрос, таким образом, касается вероятности того или иного узла для полимера, который, несмотря на ненулевую положительную толщину, поддерживается в сжатом состоянии либо за счет притяжения звеньев, достаточно сильного, чтобы пересилить эффект исключенного объема, либо за счет сжимающего внешнего поля.

Данные метода Монте-Карло [76, 87] говорят о сильном росте шансов найти нетривиальный узел при сжатии цепочки. Вероятность тривиального узла для глобулы убывает с длиной полимера заметно быстрее, чем для клубка.

Хотя сам факт сильного заузливания в глобуле хорошо виден, его довольно трудно описать количественно. Проблема также очень трудна и для аналитических подходов, есть только одна попытка теоретической оценки вероятности узлов для глобулы [88]. Как обычно в среднеполевых оценках для глобул, определяется *изменение* свойств полимера за счет коллапса. В данном случае оценка работы [88] представляет собой дополнительный множитель, который описывает дополнительное подавление тривиального узла в глобуле:

$$\mathcal{P}^{\text{glob}}(R) \sim \exp\left[-\frac{Nl^2}{R^2}\right] \sim \exp\left[-N^{1/3}\left(\frac{l}{d}\right)^{4/3}\right], \quad (26)$$

где R — размер локализации полимера, а последняя оценка написана для максимально компактного полимера объемом R^3 порядка собственного объема полимера Nld^2 . Не вполне ясно, однако, законно ли написать вероятность тривиального узла просто в виде произведения $\mathcal{P}_0(N, d/l)\mathcal{P}^{\text{glob}}(R)$. В любом случае получается, что даже для полимера скромной длины в глобулярном состоянии почти все конформации очень сильно заузлены. Это обстоятельство имеет важные следствия.

3.6. Скомканная глобула

3.6.1. Коллапс "недозаузеленного" полимера. Довольно неожиданно получается, что обилие узлов в коллапсированном состоянии играет особенно важную роль для тех систем, где образование узлов по той или иной причине невозможно, так как для них сильно обедняется доступное фазовое пространство. Таких систем довольно много: субцепи полимерной сетки; раствор незацепленных колец; длинная линейная полимерная цепь на начальной стадии коллапса, до того, как концы успеют пронизать глобулу и наделать равновесное количество узлов.

Для подобного рода систем встает вопрос: какова структура сжатого, но не заузленного (или заузленного недостаточно, "недозаузеленного") полимера? Ответ был предложен в работе [89]. Согласно оценкам этой работы, сжатый незаузеленный полимер принимает конформацию, скомканную в том смысле, что она не подчиняется теореме Флори: если взять кусок цепи длиной, скажем, k мономеров, то этот кусок окажется сам по себе коллапсированным размером порядка $k^{1/3}$. Это, конечно, резко контрастирует с обычной глобулой, где кусок из k мономеров имеет (гауссов) размер порядка $k^{1/2}$, пока k короче одного гауссова прохода цепи сквозь глобулу, и становится независимым от k при больших k . Это показано схематически на рис. 4. Высказанное утвержде-

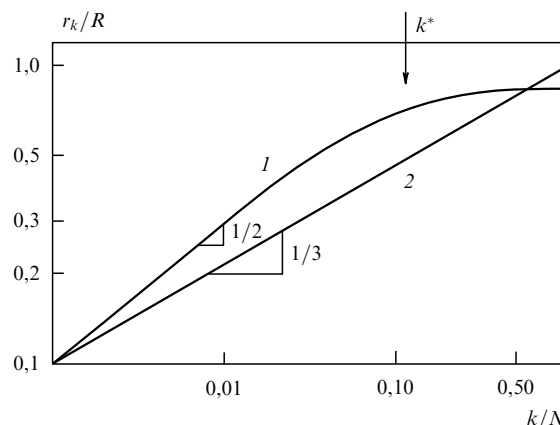


Рис. 4. Схематический дважды логарифмический график зависимости размера k -звенного куска цепи от k в обычной глобуле (кривая 1) и в скомканной глобуле (2). Характерный масштаб k^* для обычной глобулы связан с гауссовым проходом сквозь глобулу, т.е. $a(k^*)^{1/2} \sim R \sim (N/n_0)^{1/3}$, или $k^* \sim (N/a^3 n_0)^{2/3}$, где n_0 — плотность глобулы. Для максимально компактного полимера $n_0 \simeq 1/v$.

ние довольно сильное, хотя отсутствие узлов — это локальное условие, оно отбирает конформации со своеобразными локальными фрактальными свойствами.

Идея фрактальной скомканной глобулы была предложена в 1987 г., и с тех пор она не была ни подтверждена, ни опровергнута. Некоторые не прямые свидетельства в ее пользу были получены недавно из моделирования [90, 91] и из экспериментов [92, 93], а также теоретически [94, 95]. В целом идея остается в статусе гипотезы.

3.6.2. Скомканная глобула как модель нативной упаковки ДНК.

Простая оценка объема показывает, что нативная ДНК должна быть глобулой, а не клубком. Это, конечно, безумно сложная глобула, где взаимодействия перенормированы белками и включают такие явления, как распознавание каким-то белком какого-то участка ДНК и т.п. Тем не менее заключение об обилии узлов в большинстве компактных конформаций не зависит от всего этого и должно быть верно для ДНК.

С другой стороны, представляется невероятным, чтобы нативные структуры были столь сильно заузлены, так как зацепления, как кажется, подавили бы всякую способность реагировать на какие бы то ни было биохимические факторы: при обилии зацеплений это было бы весьма жесткое стекло. Руководствуясь этой логикой, авторы работы [96] высказали предположение, что *статистически глобула ДНК почти не заузлена*. При этом предполагалось, что наличие топоизомераз не меняет дела, так как ферменты действуют локально и не способны даже распознать глобальную топологию ДНК.

Модель скомканной глобулы естественно объясняет наблюдаемую реально иерархию уровней структурной организации ДНК, начиная с двойной спирали, включая нуклеосомы и все другие структуры вплоть до хроматина в целом. Все эти уровни естественно сопоставляются с глобулярными складками различных масштабов.

Альтернативный взгляд представлен в работах [97, 98], где предложена модель фантомной ДНК. Идея состоит в том, что ферментов топоизомераз может быть очень много и снабжение их АТФ может быть

неограниченным. Если ситуация в самом деле такова, то пересечения ДНК могли бы происходить без затруднений, и тогда ДНК *in vivo* могла бы выглядеть как обычная глобула.

Скорее всего, истина лежит где-нибудь посередине, и ДНК почти фантомна на сравнительно малых масштабах и скомкана на больших. Однако в целом диспут не разрешен и должен, очевидно, быть разрешен экспериментом.

3.6.3. Замечание о белках. Своеобразные фрактальные свойства скомканной глобулы побуждают думать, что они могут иметь какое-то отношение к фрактальным свойствам белков (существование или не существование которых само по себе является предметом разгоряченных дискуссий) или даже к природе вторичной структуры. Здесь эти идеи только упоминаются, и мы не будем их обсуждать более подробно.

3.7. Раздувание узла⁴

3.7.1 Топологический инвариант: максимально надутое представление узла. Чтобы построить для узлов простую теорию, была предложена следующая конструкция [73, 82]. Рассмотрим полимерную цепь в какой-то пространственной конформации и обозначим через L ее контурную длину. Заклучим контур полимера в очень тонкую трубку — настолько тонкую, чтобы можно было определить понятие топологии трубки как целого и быть уверенным, что эта топология такая же, как у цепи. Теперь станем надувать эту трубку так, чтобы ее длина L оставалась неизменной, а поперечное сечение было примерно одинаковым всюду вдоль всей трубки (мы, конечно, предполагаем, что части трубки друг в друга не проникают). Такое надувание рано или поздно кончится тем, что трубка заполнит по порядку величины весь объем внутри всех петель. Обозначим через D диаметр максимально надутой трубки. Мы утверждаем, что осевое отношение максимально надутой трубки

$$p \equiv \frac{L}{D} \quad (27)$$

есть топологический инвариант узла. Это, видимо, довольно слабый инвариант в том смысле, что может быть много узлов с одним и тем же значением p ⁵. Тем не менее p есть инвариант в том смысле, что если мы возьмем две замкнутые кривые в трехмерном пространстве так, что геометрически они будут различны, но топологически одинаковы, и проделаем нашу операцию надувания трубки, то результаты будут одинаковыми. Как видно из рис. 5, при раздувании все "случайные" черты будут стерты" и максимально надутая форма, а с нею и величина p будут зависеть только от топологии. Так получается потому, что "запасенная" длина может свободно перераспределяться вдоль контура, и, соответственно, надувание не встречает никаких фрустраций спин-стекольного типа и ведет к хорошо определенному

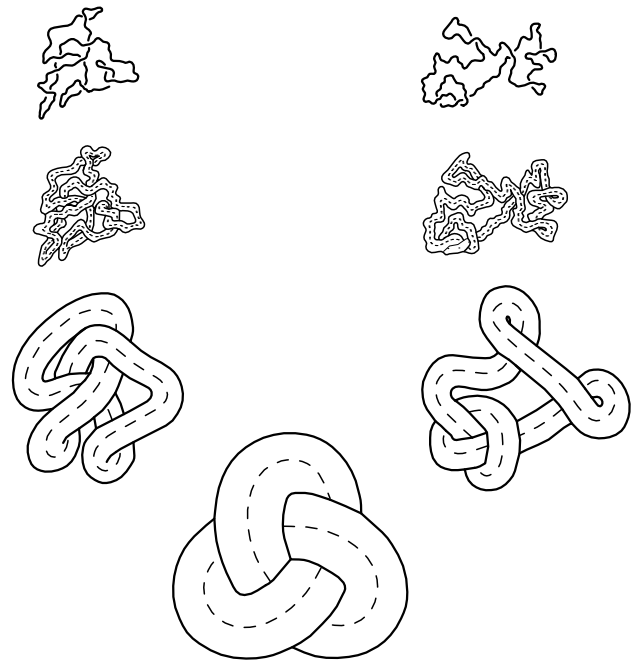


Рис. 5. Иллюстрация раздувания трубки. Сначала показаны две разные, но топологически эквивалентные конформации. Тоненькая трубочка, центрированная на полимере, очень напоминает сам полимер. По мере надувания, при сохранении контурной длины, все мелкомасштабные "вихляния" полимера постепенно вытягиваются, и в конце получается максимально раздутая форма, которая от начальной конформации не зависит, а определяется только топологией узла. Рисунок по просьбе автора любезно подготовил Габриел Беритц.

оптимуму. Упомянем, что похожий инвариант обсуждался в работе [58] в гидродинамическом контексте.

Если кольцо не заузлено, т.е. это тривиальный узел, то надувание приведет его к форме простого тора с $D \sim L$ и $p \simeq 1$. С другой стороны, чем сложнее узел, тем меньше возможность его раздувать. Поскольку реальный полимер всегда имеет конечную толщину, всегда есть максимальная достижимая сложность узла такая, при которой никакое дополнительное надувание сверх толщины полимера уже невозможно. Если полимер сделан из мономеров длиной порядка толщины полимера, то самое большое возможное значение топологического инварианта будет порядка числа мономеров: $p = L/D \simeq N$. Таким образом,

$$1 \leq p \leq N, \quad (28)$$

и можно сказать, что величина p дает грубую меру сложности узла: более сложный узла соответствуют обычно большие значения p . Чтобы это проиллюстрировать, полезно заметить, что p имеет свойство аддитивности: для составного узла $K = K_1 \# K_2$, топологический инвариант дается суммой

$$p(K_1 \# K_2) = p(K_1) + p(K_2). \quad (29)$$

Заметим, что топологический инвариант p тесно связан с примитивным путем для цепи в решетке препятствий [65], если представить себе решетку препятствий как своего рода самосогласованное поле, моделирующее топологические ограничения в узле. Примитивный путь можно сконструировать путем аффинного растяжения решетки (аналог раздувания трубки) при

⁴ Материал этого раздела основан на работе [73].

⁵ В самом определении p есть, конечно, внутренние неточности. Что такое "диаметр" для изогнутой трубки? Как гнется ось трубки при надувании? Например, для ломаных линий из n сегментов величина p для одного и того же узла может (слегка) зависеть от n . Тем не менее на практике p оказывается, видимо, довольно хорошим инвариантом [82].

сохранении длины полимера. Максимальное расширение решетки будет достигнуто, когда полимер вытянется вдоль своего примитивного пути. Соответственно, p можно интерпретировать как длину полимера, измеренную в единицах постоянной решетки в ее максимально растянутой форме. Из этой аналогии вытекает оценка числа топологически различных узлов, отвечающих данному значению p : поскольку решетку препятствий можно отобразить на дерево Кэли [69], то названное число должно экспоненциально расти с p :

$$K(p) \sim \exp(\lambda p), \quad (30)$$

где λ — просто численная константа.

3.7.2. Теория типа Флори для набухания и коллапса узла. Рассмотрим полимерное кольцо в состоянии, характеризуемом заданной величиной p , и допустим, что этот полимер имеет пространственный размер R . Наша цель заключается в том, чтобы найти равновесное значение R при заданных условиях растворителя и заданной топологии узла p . Мы будем делать это с помощью теории типа Флори, т.е. будем считать, что R определяется минимизацией свободной энергии, имеющей два вклада, связанных с полимерной упругостью и с объемными взаимодействиями:

$$F = F_{\text{elast}} + F_{\text{interact}}. \quad (31)$$

Будем сначала считать, что все мономеры более или менее равномерно распределены в объеме R^3 , хотя впоследствии мы увидим, что это далеко не всегда так. Если это предположение выполнено, то энергию взаимодействий можно записать в обычном виде:

$$\frac{F_{\text{interact}}}{T} \simeq \frac{BN^2}{R^3} + \frac{CN^3}{R^6} \simeq \frac{B}{a^3} \frac{\sqrt{N}}{\alpha^3} + \frac{C}{a^6} \frac{1}{\alpha^6}, \quad (32)$$

где $\alpha = R/a\sqrt{N}$, B и C — второй и третий вириальные коэффициенты, a — размер мономера.

Трудность проблемы в том, чтобы написать правдоподобное выражение для энтропийной (упругой) части с учетом замороженной топологии. Способ решить эту задачу был предложен в работе [73]. Он состоит в следующем. Возьмем максимально надутую трубку, занимающую объем порядка LD^2 , и деформируем ее аффинно (т.е. просто подвергнем гомотетии или преобразованию подобия) так, чтобы она занимала объем R^3 . Назовем эту трубку " R -трубкой". Если L_R и D_R — длина и диаметр R -трубки соответственно, то обе эти величины легко найти: так как R -трубка геометрически подобна максимально надутой трубке, то $L_R/D_R = p$; с другой стороны, по определению, $L_R D_R^2 = R^3$. Поэтому находим

$$L_R \simeq R p^{2/3}, \quad D_R \simeq R p^{-1/3}. \quad (33)$$

Теперь главное предположение: чтобы оценить полимерную энтропию можно считать, что полимер фантомный, но заключен в R -трубку. В поддержку этого предположения говорят результаты моделирования [82]; похоже, что тепловое движение заузленного полимера представляет собой флуктуации вокруг осевой линии соответствующей трубки. Если наше предположение принять, то энтропию оценить совсем легко, так как она не зависит от того, как трубка погружена в трехмерное пространство. Можно даже просто считать трубку прямым цилиндром, с концами цепи, закрепленными в донцах цилиндра. Тогда получается

$$\frac{F_{\text{elast}}}{T} \simeq \frac{L_R^2}{Na^2} + \frac{Na^2}{D_R^2} = \alpha^2 p^{4/3} + \alpha^{-2} p^{2/3}, \quad (34)$$

где первый и второй члены описывают растяжение вдоль трубки и сжатие поперек нее (см., например, [32]). Мы теперь минимизируем свободную энергию с членами (32) и (34) и получаем уравнение для α :

$$\alpha^5 p^{4/3} - \alpha p^{2/3} - \frac{B}{a^3} \sqrt{N} - \frac{C}{a^6} \frac{1}{\alpha^3} = 0. \quad (35)$$

Это уравнение очень похоже на аналогичное уравнение для линейной цепи (см. [32], формула (13.5)), за исключением факторов, зависящих от p . Это хорошо: выполнен "принцип соответствия", при $p = 1$ мы получаем старый результат для фантомной цепи. Для простоты будем представлять результаты прямо в терминах $R \sim \alpha a N^{1/2}$ и положим $B \simeq a^3 \tau$ и $C \simeq a^6$, где τ означает безразмерное отклонение от Θ -точки. Глядя на уравнение (35), легко найти следующие режимы:

Режим хорошего растворителя получается, когда $\tau > (p/N)^{1/2}$; ведущими факторами в этом случае являются растяжение цепи вдоль трубки (первый член формулы (35)) и двухчастичное отталкивание (третий член). Это дает

$$R \sim a N^{3/5} \tau^{1/5} p^{-4/15}. \quad (36)$$

Зависимость размера от N ($R \sim N^{3/5}$) и от температуры остались без изменений по сравнению с фантомной цепью, появился новый множитель, показывающий, что цепь набухает меньше, если на ней завязан сложный узел с большим значением p . Интересно, что размер цепи в этом режиме может быть даже меньше, чем гауссов размер фантомного полимера $a N^{1/2}$; это так при $p^{1/2} N^{-1/2} < \tau < p^{4/3} N^{-1/2}$, или, другими словами, для достаточно сложных узлов с $p > (\tau N^{1/2})^{3/4}$. Только сравнительно простой узел (по сравнению с гауссовым фантомным клубком) в очень хорошем растворителе ($p < (\tau N^{1/2})^{3/4}$) набухает.

При меньших τ или больших p система плавно переходит к квазигауссову режиму.

Квазигауссов режим имеет место при $-(p/N)^{1/2} < \tau < (p/N)^{1/2}$; в этом случае отталкивание имеется только трехчастичное и его вклад оказывается такого же порядка, как вклад сжатия поперек трубки, а противоположная тенденция по-прежнему представлена растяжением вдоль трубки. В результате

$$R \sim a N^{1/2} p^{-1/6}, \quad (37)$$

т.е. зависимость от N получается гауссова, но коэффициент уменьшается для сложных узлов.

При еще меньшем τ или меньшем p система плавно переходит в режим плохого растворителя.

Режим плохого растворителя имеет место при $\tau < -(p/N)^{1/2}$; в этом режиме ведущие факторы — это двухчастичное притяжение и трехчастичное отталкивание. Мы учтем также первую поправку, связанную с сжатием поперек трубки, что дает

$$R \simeq a |\tau|^{-1/3} N^{1/3} \left[1 + |\tau|^{-4/3} \left(\frac{p}{N} \right)^{2/3} \right]. \quad (38)$$

Главный член здесь очевиден, так как при сильном притяжении должна получиться глобула с не зависящей

от N плотностью (соответственно, этот член такой же, как и для обычной глобулы). Поправка показывает, что цепочка несколько менее компактна по сравнению с фантомной: полимер оказывается "недозаузленным" для такого сильного сжатия.

Режим *затянутого узла* новый, не имеет аналога для фантомной цепи. Он реализуется при $p \sim N$, и при нем размер цепи не зависит от качества растворителя или от взаимодействий, цепочка получается всегда компактной:

$$R \simeq aN^{1/3}. \quad (39)$$

Наши результаты для критических показателей (36), (37) согласуются с данными моделирования: в работе [99] изучались размеры колец $R(K)$ для нескольких разных узлов K , включая тривиальный узел, трилистник, восьмерку и двойной трилистник. Было найдено, что критические показатели $\nu(K)$ в формуле $R(K) \sim N^{\nu(K)}$ в пределах статистических ошибок одинаковы для всех изученных узлов K , а именно, близки к 0,6 при наличии исключенного объема и к 0,5, если он очень мал.

Изложенная простейшая теория не показывает разницы между скомканной и обычной глобулой. Этот недостаток можно исправить. Для этого нужно считать, что в R -трубке захвачена не фантомная, а реальная, но незаузленная цепь. Исключение узлов ведет к дополнительному члену в упругой свободной энергии (34), который пропорционален α^{-6} . В самом деле, выражение (34) должно быть исправлено для случая, когда цепочка сжимается в обоих направлениях — и вдоль трубки, и поперек. Скейлинговую форму свободной энергии для этого случая можно установить из условия, что производимое ею осмотическое давление зависит от N и от объема трубки $L_R D_R^2 \sim R^3$ только через плотность N/R^3 . В результате дело сводится к переопределению коэффициента C/a^6 в уравнении (35), что не влияет на скейлинговую форму основных результатов (36)–(39). Более delicate поправки могут потребоваться, если вместо простой скомканной глобулы окажется правильной или фрактальной структура [100].

Вернемся теперь к нашему предположению, что цепь заполняет объем R^3 равномерно, и обсудим еще раз режим хорошего растворителя. Рассмотрим конформацию, где все узлы собрались вместе на участке цепи длиной порядка p , и в пространстве, естественно, эта часть цепи достаточно плотная. Остальная часть цепи из $N - p$ мономеров свободно набухает в хорошем растворителе без всяких узлов. Если только $p \ll N$, наша теория дает для такого "фазового расслоения" свободную энергию порядка p для коллапсированной части с узлами плюс член порядка $(N - p)^{1/5} \approx N^{1/5}$ для свободно набухающей петли, т.е. в итоге — около $N^{1/5}$ (при предположении $p \ll N^{1/5}$). С другой стороны, однофазное состояние имеет более высокую свободную энергию: подставляя (36) в (32) и (34), получаем $N^{1/5} p^{4/5}$ ($p \gg 1$). Получается, что равномерное распределение узлов термодинамически выгодно только при $N^{1/5} \ll p \ll N$, а при $p \ll N^{1/5}$ узлы должны отслаиваться. Конечно, если узел составной, то его компоненты могут блуждать вдоль цепи независимо, и в этом смысле наш результат о фазовом расслоении смыкается с идеей локальных узлов, обсуждавшейся в работе [101].

Было бы интересно проверить предсказание о сегрегации узлов путем моделирования или в эксперименте.

3.7.3. Вероятности для узлов. Напрашивается идея использовать оценку упругой свободной энергии F_{elast} (34), чтобы найти распределение вероятностей для узлов, получаемых при случайном слипании концов линейного полимера. Такое же распределение отвечает и фантомному кольцу. Действительно, F_{elast} определяется фазовым объемом, доступным системе с замороженной топологией узла, но этот же объем определяет и вероятность получить данный узел в размороженной системе, которая свободно флуктуирует от одного узла к другому. Таким образом, вероятность получить узел с данным p можно написать в (нормированном) виде

$$\mathcal{P}_p(R) = \mathcal{Q}_p(R) \left[\int_1^N dp \mathcal{Q}_p(R) \right]^{-1},$$

$$\mathcal{Q}_p(R) = K(p) \exp \left[-\frac{F_{\text{elast}}(R)}{T} \right]. \quad (40)$$

К сожалению, сюда нельзя прямо подставить выражение (34) для $F_{\text{elast}}(R)$ из-за проблемы сегрегации узлов. В самом простом приближении можно считать, что участок цепи длиной p вообще не может набухать; другими словами, эффективная длина полимера уменьшена до $N - p$, и поэтому для растянутого полимера с $R > aN^{1/2}$ мы должны написать

$$F_{\text{elast}}(R) \simeq \frac{TR^2}{a^2(N - p)}.$$

Это дает

$$\mathcal{Q}_p(R) \sim \begin{cases} \exp \left[\lambda p - \frac{R^2}{a^2(N - p)} \right], & R \gg a\sqrt{N}, \\ \exp [\lambda p - \alpha^2 p^{4/3} - \alpha^{-2} p^{2/3}], & R \ll a\sqrt{N}, \end{cases} \quad (41)$$

где мы использовали оценку (30) для числа $K(p)$ разных узлов с данным p .

Формула (41) показывает, что для каждого размера полимера R имеется некоторая оптимальная степень заузленности, при которой распределение вероятностей (40) имеет пик:

$$p_{\text{opt}}(R) \sim \begin{cases} N - \frac{R}{a\sqrt{\lambda}}, & R \gg a\sqrt{N}, \\ N \left(\frac{Na^3}{R^3} \right)^2, & R \ll a\sqrt{N} \end{cases} \quad (42)$$

(если $\lambda > 4\sqrt{2}/3 \approx 1,9$). Таким образом, наиболее вероятная степень заузленности растет с компактизацией цепи и достигает максимальной величины $p \sim N$ для максимально компактной глобулы, где $R \sim aN^{1/3}$.

Если мы оценим нормировочный фактор в формуле (40) с помощью метода перевала (что, конечно, просто дает $\mathcal{Q}_{p_{\text{opt}}}(R)$), то получим

$$\mathcal{P}_p(R) \propto \begin{cases} \exp [\lambda(p - N)], & R \gg a\sqrt{N}, \\ \exp \left[-p^{2/3} \frac{Na^2}{R^2} - \lambda N \left(\frac{Na^3}{R^3} \right)^2 \right], & R \ll a\sqrt{N}. \end{cases} \quad (43)$$

Этот результат хорошо согласуется с формулой (23), так как он дает экспоненциальное убывание вероятности с N (для этого нам был нужен нормирующий множитель). Больше того, мы видим теперь возможную связь параметров $N_0 \approx 335$ и λ . Наш результат согласуется с численными данными (24) в области больших N (с

нашей грубой теорией мы не могли претендовать на получение степенных поправок к экспоненциальному члену). Наконец, мы воспроизвели и оценку (26).

3.8. Затянутые узлы

Узлы на веревках склонны затягиваться. Это неприятно на рыбалке, но полезно для ботиночных шнурков. Де Жен высказал гипотезу, что нечто подобное может происходить и с полимером [102]: если запасенная длина вся вытянута из узла, то рептационная релаксация подавлена, и релаксация получается необычайно медленной. Это могло бы быть важно, например, для полимерных кристаллов. Сколько звеньев нужно, чтобы затянуть узел на цепи полимера? Конечно, это зависит от локальной геометрии данного химического вещества. На модели [102] получается, что для цепи полиэтилена нужно примерно 33 звена. Любопытно, что это число очень близко к минимальной длине линии, которая может образовать узел на кубической решетке; это число нашел в свое время Дельбрюк [60], и оно равно 27.

В недавней работе [103] была предложена возможность затянуть узлы на линейном полимере в ходе деколлапса глобулы. Действительно, если начать с глобулы и дать полимеру отрелаксировать в этом состоянии, то, как мы уже знаем, в нем будет избыток узлов. Если же мы теперь быстро поместим полимер в хороший растворитель, то узлы не успеют распутаться и затянутся. Может получиться очень необычное состояние почти компактных глобул, долго живущих в хорошем растворителе. Было бы интересно проверить эту идею экспериментально.

4. Фазовый переход стеклования: глобула гетерополимера со случайной последовательностью⁶

4.1. Глобулярный гетерополимер

С этого момента мы переходим к гетерополимерам, где беспорядок представлен неоднородной последовательностью химически различных мономерных звеньев. Мы будем рассматривать проблему одной цепи. Рассматривая различные упрощенные модели, мы будем иметь в виду проблему сворачивания белков как своего рода сверхзадачу.

Мы ограничимся гетерополимерами в состоянии максимально компактной глобулы. Максимальная компактность означает, что плотность близка к плотности упаковки, а потому распределена в пространстве равномерно и флуктуировать не может. Объемное приближение [1] применимо для этого случая. Для простой решеточной модели условие максимальной компактности означает, что цепь из N мономеров занимает часть решетки с ровно N вершинами и посещает каждую вершину один и только один раз (например, цепь из 27 звеньев занимает участок $3 \times 3 \times 3$ на кубической решетке). Гамильтониан, который включает в простейшей форме все основные аспекты проблемы, как то: последовательность мономеров из определенного химического "алфавита", произвольный набор взаимодействий между ними и конформационную гибкость цепи,

можно записать в виде

$$\mathcal{H}(\text{seq}, \text{conf}) = \sum_{I,J}^N B_{s_I s_J} \Delta(\mathbf{r}_I - \mathbf{r}_J), \quad (44)$$

где заглавные латинские буквы в индексах означают номера звеньев вдоль цепи, $s_I \in \{1, \dots, q\}$ изображает "химическую породу" мономера номер I (так что $\{s_I\}$ представляет "последовательность"), q есть число химически различных типов мономеров, вектор $\mathbf{r}_I$ указывает положение мономера I в пространстве (и, тем самым, $\{\mathbf{r}_I\}$ описывает конформацию цепи). Функция $\Delta(r)$ сконцентрирована там, где могут находиться ближайшие пространственные соседи мономера; на решетке, например, $\Delta(a) = 1$ и $\Delta(r > a) = 0$, где a — размер ячейки решетки. Таким образом, наша модель имеет простой смысл, что мономеры взаимодействуют, если в данной конформации они соседи в пространстве, и энергия их взаимодействия определяется тем, какие это мономеры, в соответствии с "матрицей взаимодействий" B_{ij} . Мы предполагаем, хотя и не пишем этого явно в формуле (44), что положения мономеров могут быть только такими, что сохраняются все условия связности цепи, или линейной памяти (точки $\mathbf{r}_I$ и $\mathbf{r}_{I+1}$ всегда соседи в пространстве для всех I), исключенного объема ($\mathbf{r}_I \neq \mathbf{r}_J$ для всех $I \neq J$) и плотной упаковки.

Как ни обща модель, представленная формулой (44), это, конечно, все еще только приближение. Например, мы пренебрегли гетерогенными трех- и более частичными взаимодействиями, а матрица взаимодействий на самом деле представляет уже соответствующие свободные энергии, неявно учитывающие роль растворителя. Мы также предполагаем, что "мономеры" в формуле (44) — это уже перенормированные "квазимономеры" [32]. Последнее, в частности, означает, что многие мелко-масштабные детали "вымыты" из модели. Например, модель не может даже претендовать на описание низкотемпературных свойств белков, где известен совсем другой переход стекловидного типа [104, 105], связанный с вымерзанием малых колебаний вокруг нативной конформации.

Тем не менее модель (44) включает, как мы думаем, все основные факторы, определяющие крупномасштабное поведение глобулы гетерополимера типа белка, в том числе само явление правильного сворачивания белков. Мы ограничимся этой моделью.

Что касается матрицы взаимодействий B_{ij} , природные белки включают $q = 20$ типов мономеров. Матрица B_{ij} для них должна быть 20×20 . Есть много попыток оценивать энергии взаимодействия между аминокислотными остатками, т.е. матричные элементы B_{ij} . Среди специалистов нет согласия ни о хороших оценках энергий, ни о том, как построить разумную модель матрицы взаимодействий меньших размеров. Так называемая матрица Миязавы–Джернигана (MJ) имеет размер 20×20 и найдена из статистики базы данных для белков [109] (хотя сама процедура извлечения энергий из такой статистики очень не проста по идее и противоречива; см., в частности, [106–108]). Есть и другие попытки вывести реалистические оценки для этих энергий [110]. Конечно, они, в принципе, не могут быть совсем удовлетворительными, так как само понятие "быть в контакте" — довольно неопределенное для таких громоздких молекул, как аминокислотные остатки. Модели другого типа

⁶ Материал этого раздела, а также разделов 5.1.5 и 5.2 основан на работе [10].

часто обосновывают тем, что гидрофобный эффект является главным фактором, обеспечивающим компактность белковых глобул. Соответственно, можно грубо считать, что есть всего два типа мономеров (черно-белая модель, ЧБМ). Как мы увидим, это еще не определяет модель полностью, и есть целая совокупность разных ЧБМ. Несколько особняком стоит модель независимых взаимодействий (МНВ), ищущая упрощений не в уменьшении, а в увеличении q : в этой модели число химических разновидностей считается равным числу мономеров в цепи ($q = N$). В этом случае для каждой конформации любой матричный элемент B_{ij} встречается в сумме (44) не более одного раза. Поэтому предполагается, что все эти матричные элементы — взаимно независимые случайные величины. Как мы увидим, эта модель во многих отношениях очень удобна и полезна.

Основные модели взаимодействий с некоторыми комментариями собраны в табл. 2

4.2. Модель случайных энергий (МСЭ)

В этом разделе мы отвлечемся от полимеров и рассмотрим свойства МСЭ в общем виде. Если читателю это скучно, то можно все пропустить до раздела 4.2.5, где дана краткая сводка свойств МСЭ. МСЭ была впервые предложена Дерридой [6] (см. также [117]), имеется трудно обозримое количество работ об этой модели. В нашем изложении мы следуем подходу работы [10].

4.2.1. Что такое МСЭ? Формально, чтобы вычислить статистическую сумму какой-то системы, необходимо и достаточно знать список всех микросостояний (конформаций) $1, 2, \dots, M$ вместе с соответствующими энергиями $E_1, E_2, \dots, E_M$. Число микросостояний M огромное, оно растет экспоненциально с числом частиц N :

$$M \simeq \exp(\omega N), \quad (45)$$

где $\omega \sim 1$ в полимерном случае определяется доступными конформациями, т.е. гибкостью цепи, условиями упаковки, для решеточных моделей — геометрией

решетки и т.д. Для неупорядоченной системы все энергии $E_1, E_2, \dots, E_M$ зависят, вообще говоря, от конкретной реализации беспорядка, т.е. для гетерополимера — от последовательности. Определяющим для МСЭ является предположение, что распределения вероятностей для энергий всех микросостояний (конформаций) одинаковы и *статистически независимы друг от друга*. Обозначим через $P(E)$ распределение вероятностей для энергии какой-то одной конформации, тогда МСЭ предполагает, что для всех других конформаций распределение такое же и

$$P(E_1, E_2) = P(E_1)P(E_2). \quad (46)$$

Обычно также предполагается, что распределение $P(E)$ гауссово:

$$P(E) = (2\pi N\mathcal{E}^2)^{-1/2} \exp\left[-\frac{E^2}{2N\mathcal{E}^2}\right], \quad (47)$$

где $\mathcal{E}$ — характерная энергетическая ширина. Подчеркнем, что предположение о гауссовом характере распределения (47) несравненно менее важно, чем предположение статистической независимости (46), которое определяет МСЭ.

4.2.2. Плотность состояний для МСЭ. Очень легко можно генерировать на компьютере отдельные реализации спектра энергий МСЭ. Два примера показаны на рис. 6. Как видим, типичный спектр состоит из плотной области с огромным количеством состояний сравнительно высокой энергии и дискретной части при низких энергиях. Больше того, непрерывная часть выглядит одинаковой для всех реализаций, а дискретная, напротив, абсолютно индивидуальна.

Свойства МСЭ можно понять, рассматривая плотность состояний $n(E)$ (где $n(E)\Delta E$ — число состояний с энергиями между E и $E + \Delta E$). Легко найти математическое ожидание для $n(E)$:

$$\langle n(E) \rangle = MP(E). \quad (48)$$

Таблица 2. Распространенные модели гетерополимерных взаимодействий. Для гетерополимерных приложений все матрицы 2×2 удобно параметризовать в терминах угла θ (см. раздел 5.2.3)

Название	Число букв	Матрица	Ссылка
МЈ	20	Реалистические энергии для аминокислот	[109]
МНВ	N	Независимые случайные энергии	[9]
Поттс	q	$B_{ij} = 1 - 2\delta_{ij}$	[111]
ЧБМ	2	$B_{ij} = \bar{B} + \frac{\delta B}{2} \begin{pmatrix} -\sqrt{2}\cos\theta - \sin\theta & \sin\theta \\ \sin\theta & \sqrt{2}\cos\theta - \sin\theta \end{pmatrix}$	Раздел 5.2.3
ГП	2	$\theta = \arccos\left(\sqrt{\frac{2}{3}}\right) \approx 35,3^\circ, \quad B_{ij} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$	[112]
Изинг	2	$\theta = \frac{\pi}{2} = 90^\circ, \quad B_{ij} = -\sigma_i\sigma_j = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$	[113]
	2	$\theta = 0, \quad B_{ij} = \sigma_i + \sigma_j = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}$	[114]
NEC	2	$\theta \approx 0,092 \approx 5,27^\circ, \quad B_{ij} = \begin{pmatrix} -2,3 & -1 \\ -1 & 0 \end{pmatrix}$	[115]
Электрослабое смешивание	2	$\theta \approx 0,485 \approx 27,8^\circ, \quad B_{ij} = \begin{pmatrix} 0,762 & 0,479 \\ 0,479 & -1,720 \end{pmatrix}$	[116]

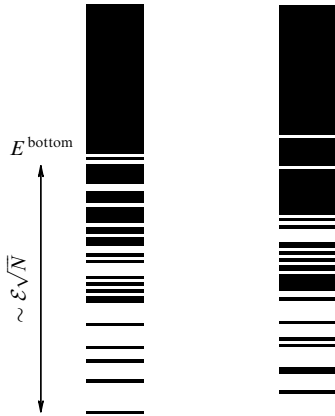


Рис. 6. Два типичных спектра для МСЭ. Каждый состоит из непрерывной и дискретной частей. Сравнивая две показанных реализации, можно убедиться, что непрерывная часть спектра универсальна, а дискретная индивидуальна.

Благодаря множителю M эта величина огромная, если только энергия E не очень далека от центральной части спектра; другими словами, если только вероятность (47) не очень мала, то астрономическое число состояний образует квазинепрерывный спектр. Для такой большой плотности состояний ее флуктуации от реализации к реализации пренебрежимо малы, т.е. $n(E) \simeq \langle n(E) \rangle$. Это справедливо лишь до тех пор, пока $MP(E) > 1$, или $E > E^{\text{bottom}}$, где

$$MP(E^{\text{bottom}}) \sim 1 \rightarrow E^{\text{bottom}} \simeq -N\epsilon\sqrt{2\omega}. \quad (49)$$

При энергиях ниже порога $E < E^{\text{bottom}}$ ожидаемое число уровней в разумном интервале ΔE становится меньше единицы. Это значит, что для некоторых реализаций там есть один уровень, а для некоторых — ни одного:

$$n(E) = \begin{cases} MP(E), & \text{если } E > E^{\text{bottom}}, \\ \text{случайные пики}, & \text{если } E < E^{\text{bottom}}. \end{cases} \quad (50)$$

Как мы и ожидали, непрерывная часть спектра от реализации не зависит, а дискретные, наоборот, для разных реализаций совсем различные.

4.2.3. Типичные и не типичные реализации в МСЭ. Чтобы понять свойства МСЭ глубже, интересно посмотреть на уровни низкой энергии. Прежде всего легко найти распределение вероятностей для энергии основного состояния. В самом деле, чтобы состояние с энергией E было основным, необходимо и достаточно, чтобы энергии всех остальных состояний были выше; для каждого из $M-1$ остальных состояний соответствующая вероятность равна

$$\int_E^\infty P(E) dE,$$

а так как (наиболее важный пункт!) все состояния статистически независимы, то мы получаем

$$P_{\text{ground}}(E) = MP(E) \left[\int_E^\infty P(E) dE \right]^{M-1}. \quad (51)$$

Поскольку же число M астрономически велико, (51) сводится к⁷

$$P_{\text{ground}}(E) \simeq \begin{cases} MP(E) & \text{при } E < E^{\text{bottom}}, \\ 0 & \text{при } E > E^{\text{bottom}}. \end{cases} \quad (52)$$

Таким образом, для подавляющего большинства реализаций основное состояние находится по энергии на величину порядка $\epsilon N^{1/2}$ ниже границы непрерывного спектра E^{bottom} , и в этом интервале лежат все дискретные уровни. Разница между ними, следовательно, пренебрежимо мала в термодинамическом пределе: все эти уровни соответствуют состояниям, которые энергетически почти тождественны, поэтому кинетически любое из них может стать ловушкой. Заметим, что сама граница непрерывного спектра на величину порядка ϵN ниже средней энергии спектра (см. (49)).

Решающе важно для дальнейшего, что наряду с такими типичными реализациями имеется также некоторое (сравнительно очень маленькое) количество редких реализаций, для которых спектр выглядит по-другому. В частности, есть и такие реализации, хотя их доля и экспоненциально мала, для которых энергия основного состояния на величину порядка N ниже порога E^{bottom} ; они очень важны для дальнейшего рассмотрения.

4.2.4. Термодинамика МСЭ. Рассматривая термодинамику МСЭ, можно было бы захотеть вычислить статистическую сумму и свободную энергию для каждой реализации беспорядка, но практически это невозможно. Поэтому необходимо воспользоваться принципом самоусреднения свободной энергии [4], который утверждает, что среднее значение свободной энергии близко к наиболее вероятному, и потому в нем доминируют обычные типичные реализации. Таким образом, мы рассмотрим среднее значение вида

$$F(T) = \langle F_{\text{seq}}(T) \rangle = -T \langle \ln Z_{\text{seq}}(T) \rangle. \quad (53)$$

Усреднять логарифм трудно — это одна из причин, почему неупорядоченные системы столь сложны для теоретического анализа и используется знаменитый метод реплик [117]. К сожалению, репличные теории не только громоздки, но и не вполне ясны в том, что касается их фундамента. Большое достоинство МСЭ в том, что для ее описания столь сложные методы не нужны.

В самом деле, статистическая сумма для данной реализации беспорядка, $Z_{\text{seq}}(T)$, есть просто сумма по всем состояниям $i = 1, 2, \dots, M$, и ее всегда можно переписать в терминах плотности состояний:

$$Z_{\text{seq}}(T) = \sum_{i=1}^M \exp\left(-\frac{E_i}{T}\right) = \int_{-\infty}^{\infty} n(E) \exp\left(-\frac{E}{T}\right) dE. \quad (54)$$

При высоких температурах в этой сумме доминируют члены с большим $n(E)$ (большая энтропия), а в этой

⁷Поскольку $N \gg 1$, мы используем метод перевала и пишем $\int_{E^{\text{bottom}}}^\infty P(E) dE \simeq 1 - k \exp[-(E^{\text{bottom}})^2/2N\epsilon^2] \simeq 1 - k \exp(-N\omega)$ (k — численная постоянная). Эта величина очень близка к единице, и ее (отрицательное) отклонение от единицы (по абсолютной величине) порядка $1/M$. Поэтому сразу же над порогом E^{bottom} величина $\int_{E^{\text{bottom}}}^\infty P(E) dE$ опускается ниже единицы настолько, что ее степень M практически обращается в нуль.

области спектр непрерывный и не зависит от последовательности. Следовательно, никакие трудности, связанные с различием между разными реализациями, при этих температурах вообще не возникают, и беспорядок в этой области вообще, по сути, не важен. В самом деле, если седловая точка в интеграле

$$Z = \int_{-\infty}^{\infty} \mathcal{M}P(E) \exp\left(-\frac{E}{T}\right) dE \quad (55)$$

$$\simeq \mathcal{M}P(E_{\text{saddle}}) \exp\left(-\frac{E_{\text{saddle}}}{T}\right)$$

лежит в области квазинепрерывного спектра $E > E^{\text{bottom}}$, то можно использовать первую строку формулы (50), что приводит к не зависящему от беспорядка значению статистической суммы. Для гауссова распределения (47), где $E_{\text{saddle}} = -N\mathcal{E}^2/T$, получаем, что условие $E_{\text{saddle}} > E^{\text{bottom}}$ выполнено при $T > T_{\text{glass}} = \mathcal{E}(2\omega)^{-1/2}$. Таким образом, при $T > T_{\text{glass}}$ можно спокойно усреднять по беспорядку *статистическую сумму* (так как она вообще от реализации не зависит!). Конечно, соответствующая свободная энергия тоже не зависит от беспорядка.

Все обстоит совершенно иначе при более низких температурах, где главный вклад в статистическую сумму дают состояния с низкой энергией, принадлежащие дискретному спектру. Казалось бы, свободная энергия могла бы очень сильно зависеть от реализации в этом случае, но это не так для типичных реализаций МСЭ, потому что различие энергии между типичными низкоэнергетическими состояниями пропорционально только $N^{1/2}$, т.е. пренебрежимо в термодинамическом пределе.

Поскольку свободная энергия должна быть непрерывной функцией температуры, мы получаем следующее важное заключение для МСЭ (см. также [118]):

$$F(T) = \begin{cases} -T \ln \langle Z_{\text{seq}}(T) \rangle, & T > T_{\text{glass}}, \\ -T_{\text{glass}} \ln \langle Z_{\text{seq}}(T_{\text{glass}}) \rangle, & T \leq T_{\text{glass}}. \end{cases} \quad (56)$$

Мы объясним более подробно в разделе 4.4.2, что усреднить статистическую сумму — значит взять "размороженное среднее". Поэтому уравнение (56) говорит на самом деле, что для МСЭ при температуре выше T_{glass} замороженное среднее значение свободной энергии совпадает с размороженным. Это важное заключение (верное только для систем типа МСЭ) мы будем активно использовать в дальнейшем.

4.2.5. Свойства МСЭ: краткая сводка. Здесь мы просто перечислим основные свойства МСЭ:

1. Определяющее основное свойство МСЭ — это статистическая независимость микросостояний (46).

2. Энергетический спектр МСЭ состоит из квазинепрерывной части, которая практически не зависит от реализации беспорядка, и нескольких дискретных уровней, конкретное расположение которых индивидуально для каждой реализации.

3. Для типичных реализаций основное состояние МСЭ находится на величину порядка $\sqrt{N}$ ниже границы непрерывного спектра, которая, в свою очередь, на величину порядка N ниже середины спектра. Энергетическая разница между дискретными состояниями тоже порядка $\sqrt{N}$.

4. Для МСЭ имеется особая температура T_{glass} такая, что при $T > T_{\text{glass}}$ система имеет малую энтропию и перебирает по ходу теплового движения состояния квазинепрерывного спектра, тогда как при $T < T_{\text{glass}}$ система имеет на выбор всего несколько индивидуальных дискретных состояний.

5. Свободная энергия МСЭ дается формулой (56).

Заметим, что все эти свойства независимы от гауссова характера распределения (47).

4.3. Применима ли МСЭ как модель стеклования гетерополимера?

Вопрос, поставленный в названии раздела, рассмотрен более подробно в работе [119]. Если читатель готов довериться МСЭ, то можно пропустить этот раздел.

4.3.1. МСЭ не может быть точной для гетерополимеров.

Брингельсон и Вулинес в работе [7] *постулировали* применимость МСЭ для белковых глобул. Шахнович и Гутин в работе [9] *привели аргументы* в пользу того, что МСЭ должна быть применима для компактного гетерополимера с независимыми взаимодействиями. Тем не менее утверждение о применимости МСЭ часто встречается с понятным недоверием. Действительно, МСЭ, очевидно, не может быть точной. Чтобы это увидеть, возьмем две конформации, скажем, α и β , отличающиеся друг от друга только маленькой локальной перестройкой. Энергии E_α и E_β как суммы по парам контактирующих мономеров будут иметь одинаковыми большинство членов и различаться будут только вкладами от той маленькой области, где отклоняются друг от друга α и β . Ясно, что эти энергии статистически сильно зависимы.

Простейшая количественная мера статистической зависимости между энергиями двух заданных конформаций α и β — это корреляция

$$\langle E_\alpha E_\beta \rangle - \langle E_\alpha \rangle \langle E_\beta \rangle = \langle \delta E_\alpha \delta E_\beta \rangle,$$

где $\langle \dots \rangle$ означает усреднение по последовательностям. Для усреднения мы берем вероятность реализации каждой последовательности в виде произведения

$$P_{\text{seq}}^{(0)} \equiv P^{(0)}(\{s_i\}) = \prod_{i=1}^N p_{s_i}, \quad (57)$$

что соответствует независимому появлению мономеров $\{i\}$ с вероятностями $\{p_i\}$ (см. также формулу (62))⁸. Мы определяем также среднее $\bar{B}$ и разброс δB^2 для матрицы взаимодействий:

$$\bar{B} = \sum_{ij} p_i B_{ij} p_j,$$

$$\delta B^2 = \sum_{ij} p_i (B_{ij})^2 p_j - \left(\sum_{kl} p_k B_{kl} p_l \right)^2 =$$

$$= \sum_{ij} p_i (B_{ij} - \bar{B})^2 p_j. \quad (58)$$

Непосредственное вычисление для гамильтониана (44) тогда дает

$$\langle \delta E_\alpha \delta E_\beta \rangle = \langle \delta B^2 \rangle Q_{\alpha\beta} + K_{\alpha\beta} \sum_{ijk} p_i \delta B_{ij} p_j \delta B_{jk} p_k, \quad (59)$$

⁸ Вероятности p_i пропорциональны $\exp(\mu_i/T)$, где μ_i — химические потенциалы мономеров в резервуаре, где готовился полимер.

где

$$Q_{\alpha\beta} \equiv \sum_{I \neq J} \Delta(\mathbf{r}_I^\alpha - \mathbf{r}_J^\alpha) \Delta(\mathbf{r}_I^\beta - \mathbf{r}_J^\beta) \quad (60)$$

можно определить как перекрытие между конформациями α и β , эта величина пропорциональна числу одинаковых межмономерных контактов в этих двух конформациях, и

$$K_{\alpha\beta} = \sum_{I \neq J \neq K} \Delta(\mathbf{r}_I^\alpha - \mathbf{r}_J^\alpha) \Delta(\mathbf{r}_J^\beta - \mathbf{r}_K^\beta). \quad (61)$$

Для максимально компактного полимера, если каждый мономер имеет одно и то же число z соседей в пространстве, величина $K_{\alpha\beta}$ в объемном приближении вообще не зависит от конформаций и равна просто Nz^2 . Для некоторых других случаев $K_{\alpha\beta}$ играет роль важного параметра порядка.

Таким образом, между состояниями есть действительно какая-то зависимость, и формула (59) показывает, что характер корреляции зависит от взаимодействий (B_{ij}), от пространства доступных конформаций ($Q_{\alpha\beta}$) и от последовательности (p_i). Первый, конформационно-зависимый член коррелятора (59) соответствует в точности нашим качественным аргументам, так как он пропорционален числу одинаковых контактов в двух конформациях.

Взаимодействия входят в конформационно-зависимый член в формуле (59) через энергетический разброс δV , поэтому этот член существует для всех (гетерополимерных) типов взаимодействий. Во втором члене формулы (59) взаимодействия играют менее тривиальную роль. Для многих моделей этот член вообще равен нулю, но допустим, например, что есть какой-то один или несколько сортов мономеров (скажем, гидрофобные), которые сильно взаимодействуют со всеми остальными. Тогда две конформации оказываются коррелированными даже в отсутствие перекрытия, без общих пар мономеров, при нулевом Q , если они имеют только единичные одинаково расположенные мономеры. Важный пример такого типа — гидрофобно-полярная (ГП) модель (см. табл. 2).

Для моделей с равномерным составом ($p_i = 1/q$) и симметричными взаимодействиями, например, для моделей Поттса или МНВ [9] конформационно-независимый член в формуле (61) равен нулю, и тогда корреляции связаны только с перекрытием конформаций.

4.3.2. Почему же МСЭ все-таки может быть хорошим приближением? Хотя МСЭ и неточна, но во многих случаях является хорошим приближением. Ее применимость связана с характером пространства конформаций, которая позволяет осуществить лишь очень небольшое количество локальных конформационных перестроек, недостаточное, чтобы серьезно нарушить МСЭ. Это случается из-за стерических ограничений, накладываемых максимальной компактностью, что особенно очевидно для решеточных моделей. Формально, чтобы установить применимость или неприменимость МСЭ, нужно вычислить, сколько есть пар состояний (конформаций) α и β с данным значением $Q_{\alpha\beta}$; поскольку это делается обычно методом Монте-Карло, то об этом числе говорят, как о вероятности получить данное значение $Q_{\alpha\beta}$. МСЭ разумна, если эта вероятность бимодальна, с пиками, сосредоточенными на малых и

на максимальных Q , и с провалом при промежуточных Q . МСЭ не разумна в других случаях.

Больше того, поскольку фазовый переход стеклования гетерополимера происходит как переход первого рода между макросостояниями с большим числом (почти) не перекрывающихся конформаций и с малым числом сильно перекрывающихся конформаций, то наличие некоторого количества конформаций с умеренными перекрытиями не сыграет никакой роли в термодинамике. Отметим, однако, что эти состояния играют важную роль в кинетике сворачивания белков.

4.3.3. Примеры МСЭ- и не-МСЭ-интуиции. На первый взгляд кажется, что вопрос о применимости МСЭ — это вопрос чистой теории или даже технико-теоретического жаргона. На самом деле это не так. Например, предположение типа МСЭ о статистической независимости состояний делалось неявно в нескольких *экспериментальных* работах. Так, в работах по конструированию белков *de novo* [120] неявно делалось предположение, что понижение энергии желаемой конформации за счет изменения последовательности в надлежащем месте не приведет к сильному изменению энергий других конформаций, которые, в принципе, могли бы в результате стать основными. Это типичное предположение типа МСЭ.

Но есть и примеры *противоположной* интуиции. Например, многие работы посвящены попыткам компьютерного определения для данной последовательности состояния низкой, но не самой низшей энергии [121]. При этом, конечно, предполагается, что это состояние как-то похоже на основное; но если МСЭ верна, то нахождение такого состояния вообще ничего не говорило бы об основном состоянии.

В этой статье мы будем предполагать, что МСЭ справедлива. Вопрос об отклонениях от МСЭ, как и роль поправок к МСЭ для кинетики, мы обсудим в другом месте [119, 122], а здесь ограничимся случаем МСЭ.

4.4. Размороженный гетерополимер

4.4.1. Что такое размороженный гетерополимер? Основной результат (56) раздела 4.2.4 включает усреднение статистической суммы по всем возможным последовательностям. Рассмотрим эту величину более внимательно:

$$\langle Z_{\text{seq}}(T) \rangle = \frac{1}{\mathcal{N}} \sum_{\text{seq}} Z_{\text{seq}}(T). \quad (62)$$

Здесь $\mathcal{N}$ — общее число последовательностей. Так как $Z_{\text{seq}}(T)$ представляет собой сумму по конформациям, то большая сумма $\sum_{\text{seq}} Z_{\text{seq}}(T)$ имеет физический смысл статистической суммы гипотетической системы, в которой и конформация, и последовательность принимают участие в тепловом движении на равных правах. Такая система называется размороженным гетерополимером. Хотя существуют реальные физические системы, в большей или меньшей степени похожие на размороженный гетерополимер [40], интерес к этой модели связан в первую очередь не с ними.

Заметим, что гамильтониан размороженного полимера дается той же формулой (44), что и для замороженного. Различие состоит не в гамильтониане, а в том, как мы вычисляем статистическую сумму — суммируя

только по конформациям для реального замороженного полимера или по конформациям и последовательностям — для размороженного.

Размороженный полимер, в принципе, гораздо проще замороженного. Именно поэтому формула (56) так важна, она выражает свойства сложной системы через свойства более простой. Хотя даже для размороженного полимера нет способа точно вычислить статистическую сумму или свободную энергию в терминах B_{ij} , но соотношение (56) позволяет прибегнуть к разного рода упрощенным или феноменологическим моделям для размороженной системы. Это очень похоже на подход, выдвинутый и пропагандировавшийся И.М. Лифшицем в теории полимеров. Действительно, полимерная жидкость *a priori* кажется более сложной системой по сравнению с низкомолекулярной жидкостью, для которой, в свою очередь, теоретическая физика почти беспомощна в том, что касается каких бы то ни было вычислений из первых принципов, но очень богата всевозможными феноменологическими, интерполяционными и тому подобными подходами. Соответственно, для полимеров И.М. Лифшиц предложил [1] поставить задачу так: пусть мы знаем все о жидкости разорванных звеньев (квазимономеров [1, 32]), что мы можем тогда узнать о жидкости цепей? Мы хотим использовать подобный подход здесь: допустим, что мы знаем все о размороженной системе и постараемся на этой основе понять свойства замороженной. Подчеркнем, что ограничения применимости полученных на этом пути результатов связаны не с грубостью тех интерполяций, которые мы будем использовать для размороженной системы (их можно было бы и усовершенствовать), а с МСЭ.

В соответствии с этой программой, начнем с выбора подходящего приближения для размороженной средней свободной энергии.

4.4.2. Размороженная средняя свободная энергия для модели независимых взаимодействий (МНВ). Во избежание путаницы подчеркнем, что модель случайных энергий (МСЭ) и модель независимых взаимодействий (МНВ) — вещи разные. В первой случайно выбираются макроскопические энергии системы в целом, во второй — отдельные микроскопические взаимодействия. Автор чувствует определенную вину за такую терминологическую громоздкость, но, как кажется, корни этой трудности лежат в сложной природе предмета рассмотрения.

Единственная модель, для которой размороженную среднюю свободную энергию можно вычислить точно, — это МНВ (см. табл. 2), так как дело сводится к своего рода идеальному газу. Напомним, что в МНВ число химических разновидностей равно числу мономеров в цепи (или больше его), $q \geq N$, и попарные энергии взаимодействия выбираются *независимо* из гауссова распределения:

$$P(B_{ij}) = \frac{1}{\sqrt{2\pi\delta B^2}} \exp\left[-\frac{(B_{ij} - \bar{B})^2}{2\delta B^2}\right], \quad (63)$$

где $\bar{B}$ и δB^2 играют роль просто параметров распределения. В данном случае размороженное среднее находится с помощью следующего простого аргумента. Возьмем сначала зафиксированную (формально — замороженную) конформацию и рассмотрим суммирование (или усреднение) по последовательностям. В каждом члене

такой суммы мы должны были бы присвоить "химическое имя" каждому мономеру. В МНВ вместо этого мы должны каждой связи присвоить какое-то значение B , причем делается это для всех связей независимо. При зафиксированной конформации мы имеем зафиксированную систему таких связей, как показано на рис. 7, и дело сводится к идеальному "газу" параметров B на такой своеобразной решетке. Усреднение по последовательностям эквивалентно усреднению энергии каждой связи согласно соотношению

$$\exp\left(-\frac{B_{\text{eff}}}{T}\right) = \int \exp\left(-\frac{B}{T}\right) P(B) dB; \quad (64)$$

для гауссова распределения (63) получаем

$$B_{\text{eff}} = \bar{B} - \frac{\delta B^2}{2T}. \quad (65)$$

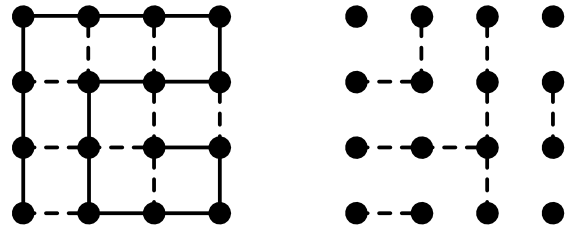


Рис. 7. Схематическая двумерная иллюстрация того, что фиксированной конформации отвечает фиксированный набор связей (или контактов) между мономерными звеньями.

Поскольку мы рассматриваем только максимально компактные конформации, общее число связей во всех конформациях одинаковое, каждая связь имеет энергию B_{eff} , так что суммирование по всем M конформациям тривиально, и получается

$$\begin{aligned} F_{\text{ann}}^{\text{av}} &= -T \langle Z_{\text{seq}}(T) \rangle = -T \ln \left[M \exp\left(-Q \frac{B_{\text{eff}}}{T}\right) \right] = \\ &= Q \left[\bar{B} - \frac{\delta B^2}{2T} \right] - TN\omega, \end{aligned} \quad (66)$$

где $Q = \sum_{I \neq J} \Delta(\mathbf{r}_I - \mathbf{r}_J) \equiv Q_{xx}$ — (независимое от конформации α) число связей (или контактов) между мономерами в любой компактной конформации, $\omega = -\ln M/N$ — полимерная энтропия на мономер.

К тому же ответу можно прийти и формально, вычисляя гауссовы интегралы по B_{ij} в выражении

$$\langle Z_{\text{seq}}(T) \rangle = P(B_{IJ}) \sum_{\text{confs}} \exp\left[-\sum_{I,J} \frac{B_{IJ}}{T} \Delta(\mathbf{r}_I - \mathbf{r}_J)\right]. \quad (67)$$

4.4.3. Размороженная средняя свободная энергия в терминах высокотемпературного разложения. Для других моделей, кроме МНВ, точное определение даже размороженной свободной энергии невозможно. Вместо этого мы прибегнем к высокотемпературному разложению. *A priori* это может показаться странным, так как стеклование, к которому мы подходим, кажется низко-, а не высокотемпературным эффектом. На самом деле стеклование в замороженной системе с фрустрациями происходит при такой температуре, которая еще вовсе не низка по масштабам размороженной системы, поэтому высо-

котемпературное разложение законно, оно оправдано сравнением температуры стеклования для замороженной системы T_{glass} с температурой соответствующего фазового перехода T_c в размороженной системе. Мы ограничимся первым нетривиальным членом высокотемпературного разложения, будут также понятны случаи, когда необходимы следующие.

Так же, как мы делали для МНВ выше, начнем с усреднения по последовательностям при заданной конформации. Выполняя высокотемпературное разложение и сохраняя члены вплоть до $O(1/T^2)$ включительно, для размороженного среднего получаем

$$\begin{aligned} -T \ln W^{\text{conf}}(T) &= -T \ln \left[\left\langle \exp \left[-\frac{\mathcal{H}(\text{seq}, \text{conf})}{T} \right] \right\rangle \right] \simeq \\ &\simeq \langle \mathcal{H} \rangle - \frac{1}{2T} \left[\langle \mathcal{H}^2 \rangle - \langle \mathcal{H} \rangle^2 \right] = \\ &= Q \left[\bar{B} - \frac{\delta B^2}{2T} \right]. \end{aligned} \quad (68)$$

Это усреднение выполняется так просто, потому что, как и для МНВ, результат не зависит от того, о какой именно (компактной) конформации идет речь, так как число контактов Q для них всех одинаковое. Суммирование (статистической суммы, конечно) по всем M компактным конформациям дает

$$F_{\text{ann}}^{\text{av}} = Q \left[\bar{B} - \frac{\delta B^2}{2T} \right] - TN\omega, \quad (69)$$

что очень похоже на результат для МНВ (66), но теперь мы используем более общие определения (58) для $\bar{B}$ и δB . Таким образом, в этом порядке мы фактически аппроксимируем $P(B_{ij})$ и сводим модель к МНВ. С другой стороны, здесь выявляется и смысл МНВ как эквивалента первого члена высокотемпературного разложения. Отклонения от МНВ будут появляться в следующих порядках высокотемпературного разложения.

Во избежание недоразумений отметим также разницу между свободной энергией размороженной системы и "размороженным средним"; в первом случае мы суммируем по последовательностям, во втором — усредняем. Соответственно, энтропия $T \ln M$ присутствует в первом случае и отсутствует во втором, и это — единственная разница.

Имеется несколько общих свойств, которые мы более подробно рассмотрели в другом месте [123], а здесь просто повторим:

1. Гетерополимерные эффекты не зависят от среднего значения B_{ij} .

2. Изменение разброса энергий B_{ij} эквивалентно изменению температуры, именно δB является характерным энергетическим масштабом в гетерополимерной задаче.

3. **Теоремы сведения.** Можно придумать матрицы взаимодействия, которые формально различны, но физически одинаковы. Например, можно формально создать целый новый "клон" мономеров, но если они взаимодействуют со всеми так же, как старые, то ничего не должно измениться. Было формально показано, что это действительно так.

4.5. Фазовый переход стеклования

Мы теперь, наконец, полностью подготовлены к тому, чтобы перейти к основному предмету данного раздела —

стеклованию глобулы гетерополимера со случайной последовательностью. Наши главные инструменты — это формула (56), выражающая замороженную свободную энергию в терминах размороженного среднего, и (69), выражающая размороженную среднюю свободную энергию в терминах высокотемпературного разложения. Для дальнейших ссылок мы выписываем здесь соответствующее выражение для замороженной средней свободной энергии (определенной в соответствии с (53)):

$$F(T) \simeq \begin{cases} Q \left[\bar{B} - \frac{\delta B^2}{2T} \right] - TN\omega, & T > T_{\text{glass}}, \\ Q \left[\bar{B} - \frac{\delta B^2}{2T_{\text{glass}}} \right] - T_{\text{glass}}N\omega, & T \leq T_{\text{glass}}. \end{cases} \quad (70)$$

4.5.1. Переход стеклования в МСЭ. Наш анализ МСЭ наводит на мысль, что что-то важное происходит, когда при уменьшении температуры средняя энергия опускается ниже границы квазинепрерывного спектра. Выше соответствующей температуры T_{glass} , система перебирает много (порядка $O(\exp N)$) состояний и ведет себя практически независимо от реализации беспорядка: побеждает энтропия. Ниже этой температуры главный вклад дают несколько индивидуальных дискретных состояний: побеждает энергия. Температура T_{glass} называется температурой стеклования, соответствующий переход мы здесь называем стеклованием. Хотя очень легко найти выражение $T_{\text{glass}} = \mathcal{E}(2\omega)^{-1/2}$ для примера гауссова распределения (47), нам сейчас было бы более интересно применить идею стеклования к гетерополимеру. Для этого мы замечаем, что точка стеклования характерна тем, что в ней энтропия становится порядка единицы, т.е. пренебрежимо мала в термодинамическом пределе. Значит, чтобы найти температуру стеклования для гетерополимера, мы можем просто использовать соотношение

$$S(T_{\text{glass}}) = - \left. \frac{\partial F}{\partial T} \right|_{T=T_{\text{glass}}} = 0. \quad (71)$$

4.5.2. Стеклование случайного гетерополимера. Используя соотношение (70), находим для энтропии замороженного гетерополимера

$$S(T) = - \frac{dF}{dT} \simeq N\omega - Q \frac{\delta B^2}{2T^2}, \quad T \geq T_{\text{glass}}. \quad (72)$$

Значит, $S(T) = 0$ при температуре T_{glass} такой, что

$$T_{\text{glass}}^2 = \frac{\delta B^2}{2s}, \quad (73)$$

где $s \approx \ln a^3/v$ — конформационная энтропия в расчете на связь (она связана с энтропией на мономер ω : $s = N\omega/Q$). Этот результат, как и поправки к нему за счет дальнейших членов высокотемпературного разложения, полностью согласуется с предшествующими результатами (несравненно более трудных) репличных вычислений [9, 123, 124]. Формула (73) также вполне ясна физически: гетерополимерная глобула стеклует, потому что выигрыш энергии за счет удачного расположения притягивающихся мономеров превышает проигрыш полимерной конформационной энтропии. Поскольку мономеры связаны в цепь, что вызывает фрустрацию и не дает им

расположиться оптимально, то максимум, что они могут выиграть в энергии на мономер, — это величина порядка δB , потери же энтропии определяются параметром s .

Полезно теперь еще раз переписать соотношение (70) для свободной энергии глобулы, на этот раз в терминах T_{glass} вместо ω или s ; используя формулу (73), получаем

$$F(T) \simeq \begin{cases} Q \left[\bar{B} - \frac{\delta B^2}{2T} \left(1 + \frac{T^2}{T_{\text{glass}}^2} \right) \right], & T \geq T_{\text{glass}}, \\ Q \left[\bar{B} - \frac{\delta B^2}{T_{\text{glass}}} \right], & T \leq T_{\text{glass}}. \end{cases} \quad (74)$$

4.5.3. Параметр порядка для перехода стеклования. Более деликатной характеристикой стеклования является параметр $x(T)$, который можно определить для каждой последовательности с помощью соотношения

$$x_{\text{seq}}(T) = 1 - \sum_{\gamma=1}^M \mathcal{P}_{\gamma}^2, \quad (75)$$

где $\mathcal{P}_{\gamma}$ — больцмановская вероятность состояния (конформации) γ для данной последовательности; $\mathcal{P}_{\gamma} \sim \exp(-\mathcal{H}_{\gamma}/T)$. Можно определить и среднее по последовательностям $\langle x(T) \rangle_{\text{seq}} \equiv x(T)$. С одной стороны, эта величина получает простой физический смысл в теории реплик (она описывает группирование реплик из-за спонтанного нарушения репличной симметрии; из n реплик получается n/x групп с x репликами в каждой). С другой стороны, эта величина связана с числом термодинамически существенных состояний $M_{\text{seq}}(T)$: $x_{\text{seq}}(T) = [1 - M_{\text{seq}}(T)]^{-1}$. Именно величина x играет роль параметра порядка для стеклования [117]. Для МСЭ можно показать, что

$$x(T) = \begin{cases} 1, & T \geq T_{\text{glass}}, \\ \frac{T}{T_{\text{glass}}}, & T \leq T_{\text{glass}}. \end{cases} \quad (76)$$

4.5.4. Свойственно ли стеклование только каким-то специальным типам взаимодействий? В заключение подчеркнем, что полимер можно считать гетерополимером, если он составлен из мономеров, которые по-разному взаимодействуют. Математически это выражается условием $\Delta \hat{B} \neq 0$, где $\hat{\Delta}_{ij} = p_i p_j - p_i \delta_{ij}$. При любых взаимодействиях, удовлетворяющих этому условию, полимер должен стекловаться ниже определенной температуры. Таким образом, детали взаимодействий не существенны ни для самого факта стеклования, ни для качественных свойств этого перехода. Вместе с тем, они очень важны для определения структуры застеклованной фазы. Стеклование — это весьма грубое свойство практически любого гетерополимера, если только разнообразие взаимодействий достаточно, чтобы оно происходило при температуре в пределах разумного.

4.6. Компьютерные тесты стеклования

4.6.1. Почему компьютерное моделирование нужно? Есть несколько причин, по которым аналитические результаты предыдущих разделов нужно дополнить компьютерным моделированием:

1) реальные белки не очень велики, а аналитические методы позволяют изучать только термодинамический предел;

2) применимость МСЭ, на которой основаны наши аналитические результаты, не очень надежно установлена;

3) интересно рассмотреть более сложные модели.

4.6.2. Перечисление гамильтоновских путей и других конформаций⁹. Чтобы моделировать стеклование на компьютере, необходимо для начала иметь полный список конформаций. Действительно, раз система застывает в одной конформации, то случайный поиск методом Монте-Карло в пространстве конформаций может легко эту единственную конформацию пропустить. Для максимально компактной глобулы это сводится к проблеме перечисления гамильтоновских путей на графе, получаемом как участок, скажем, кубической решетки.

В общем случае путь по ребрам графа называется гамильтоновским, если он посещает каждую вершину графа один и только один раз. Шахнович и Гутин в работе [125] впервые показали, что на участке кубической решетки размером $3 \times 3 \times 3$ имеется 103346 гамильтоновских путей (исключая повторения за счет вращательной или зеркальной симметрий). Один из них показан на рис. 8. Использование параллельного суперкомпьютера со 128-ю процессорами позволило в работе [126] перечислить гамильтоновские пути на суперрешетках $3 \times 3 \times 4$ и $3 \times 4 \times 4$. Результаты представлены в табл. 3.

На рисунке 9 представлен график зависимости натурального логарифма числа гамильтоновских путей M

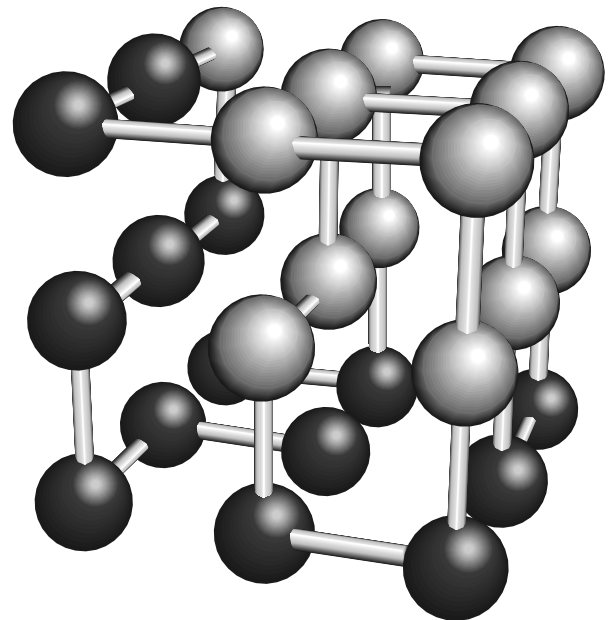


Рис. 8. Одна из компактных конформаций 27-мера на участке кубической решетки размером $3 \times 3 \times 3$. В силу доступности для компьютерных расчетов, эта система, впервые предложенная в [125], стала очень популярной моделью в работах по сворачиванию белков. Показан 27-мер с двумя типами мономеров (белыми и черными), показанная конформация отвечает одному из состояний наинизшей энергии для изинговских взаимодействий.

⁹ Материал этого раздела основан на работе [126].

Таблица 3. Сводка данных перечисления конформаций (N — число мономеров цепи, M — число максимально компактных конформаций (кроме связанных симметрий) или гамильтоновских путей, M_{tot} — общее число конформаций)

N	Комментарий	M	M_{tot}
18		1,085	5,577,317,124
27		103,346	
36		84,731,192	
48	64 часа процессорного времени на СМ-5	134,131,827,745	
26	С "карманом"	564,368	
64	Скомканные	261,496,832	

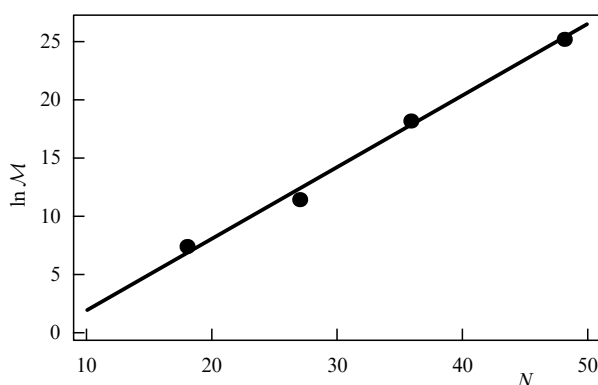


Рис. 9. Логарифмическая зависимость числа гамильтоновских путей M от длины пути N . Интересно, что точка для $N=27$ лежит несколько ниже линии; возможно, это связано с правильной кубической формой соответствующего куска решетки, если это верно, то точка для $N=64$ будет опять ниже прямой.

от N . Как видно, точки хорошо ложатся на прямую

$$M \approx ab^N \approx 0,02 \cdot (1,86)^N \quad (77)$$

(точнее, $\ln a = -4,3 \pm 1,2$ и $\ln b = 0,62 \pm 0,04$). Эта интерполяция хорошо работает в области $N \leq 48$.

С другой стороны, среднеполевая теория Флори для полимерного расплава [127] вполне применима для оценки числа компактных глобулярных конформаций в области $N \rightarrow \infty$. Действительно, теория Флори основана на оценке числа способов добавить новый мономер к концу растущей цепи с учетом ограничений, накладываемых уже занятыми узлами. Такой подход одинаково применим для расплава или для большой глобулы: эти две системы отличаются энтропией смешения отдельных цепей, которая есть в расплаве и которой нет в глобуле, и поверхностными эффектами, которые в глобуле гораздо больше, чем в расплаве. Первый фактор можно просто из формулы Флори выбросить, а вторым для большой глобулы можно пренебречь. В итоге в асимптотике $N \rightarrow \infty$ получается

$$M \approx \left(\frac{z-1}{\exp 1} \right)^N \approx (1,84)^N, \quad (78)$$

где z — координационное число решетки, и мы взяли $z=6$, так как рассматриваемая нами решетка кубическая.

Получается, что формулы (77) и (78) согласуются очень хорошо; разумно предположить, что уравнение

(77) можно использовать для любого N и надеяться на логарифмическую точность.

В работе [128] утверждается, что формула Флори (78) должна быть исправлена и что правильная формула имеет вид $M \simeq (z/\exp 1)^N$; численные данные [126] согласуются с формулой Флори (78), но не согласуются с усовершенствованной формулой.

Для $N=64$ формула (77) дает $M \approx 2 \times 10^{15}$. Это значит, что перечисление гамильтоновских путей для субрешетки $4 \times 4 \times 4$ на несколько порядков превосходит современные компьютерные возможности. Больше того, хотя для случая $N=48$ гамильтоновские пути удалось перечислить, это потребовало столько компьютерного времени, что никакого применения к моделированию пока не обещает. Лишь случаи $N=18$, $N=27$ и $N=36$ можно использовать для моделирования. Интересно, что уже случай $N=36$ представляет нечто существенно новое даже по сравнению с $N=27$; например, при $N=36$ в глобуле впервые появляются траектории с узлами (точнее, с псевдоузлами, так как цепь линейная).

Есть всего несколько других примеров, когда удается полностью перечислить конформации. Полное перечисление всех (компактных и не компактных) конформаций, насколько известно автору, удалось осуществить только для $N \leq 18$. Упомянем также перечисление всех компактных конформаций 26-мера в субрешетке $3 \times 3 \times 3$ (где одна вершина остается пустой — модель активного центра фермента) [13]. Наконец, можно перечислить все скомканные конформации в субрешетке $2^k \times 2^k \times 2^k$ для произвольного $^{10} k$. Для этого случая очень хорошее приближение в асимптотике $k \rightarrow \infty$ дается формулой

$$M_k^{\text{crumpled}} \simeq \gamma A^N, \quad (79)$$

где $\gamma \approx 0,87300$, $A \approx 1,3565$; например, для $N=64$ приближенная формула дает $M_2^{\text{crumpled}} \approx 2,605 \times 10^8$, а точный ответ есть $2,61496832 \times 10^8$.

Используя оценку (78), мы можем найти долю скомканных среди всех компактных конформаций; как и следовало ожидать, она экспоненциально убывает с N :

$$\frac{M^{\text{crumpled}}}{M} \approx \left(\frac{A \exp 1}{z-1} \right)^N \approx \exp(-0,3N).$$

4.6.3. Стеклование, действительно, наблюдается при моделировании. Первый компьютерный тест стеклования был выполнен в работе [125] для модели независимых взаимодействий. Стеклование наблюдалось путем слежения за величиной $\chi(T)$ (75). Предсказание, выраженное формулой (76), было подтверждено для $N=27$ в работе [125], а после этого — еще в нескольких работах для $N=18$, 27 и 36. Разумеется, поскольку системы конечные, настоящего излома зависимости $\chi(T)$ не могло наблюдаться, но получаются зависимости, действительно, очень резкие.

¹⁰ Скомканные конформации (см. раздел 6.6) на решетке определяются так: куб размером $2^k \times 2^k \times 2^k$ представляем в виде 8 меньших кубов размером $2^{k-1} \times 2^{k-1} \times 2^{k-1}$, каждый из которых можно опять разбить таким же образом, и т.д. вплоть до кубов $2 \times 2 \times 2$. Теперь мы строим скомканную траекторию так, чтобы она переходила в новый куб данного масштаба тогда и только тогда, когда она уже посетила все вершины в старом.

Больше того, сам факт, что $x(T)$ становится меньше единицы, говорит прямо, что энтропия системы резко уменьшается (если бы речь шла о бесконечной системе и термодинамическом пределе, можно было бы сказать, что энтропия становится нулевой), а это, собственно, и означает стеклование.

4.6.4. Компьютерные тесты применимости МСЭ. Один из наиболее естественных вопросов, которые можно поставить перед компьютерным моделированием, — это проверка применимости МСЭ. Напомним, что мы в нашем подходе поставили МСЭ в основу всех аналитических выкладок. В других случаях это не делается столь явно, но МСЭ появляется в большинстве случаев в том или ином обличии (например, как одноступенчатое нарушение репличной симметрии). Поэтому проверить применимость МСЭ важно. Впервые это было сделано в работе [125], где вычислялось распределение вероятностей для параметра перекрытия между $Q_{\alpha\beta}$ случайно выбираемыми парами конформаций:

$$P(Q) = \sum_{\alpha\beta} P_{\alpha} P_{\beta} \delta(Q - Q_{\alpha\beta}). \quad (80)$$

Если МСЭ справедлива, то это распределение должно быть бимодальным, с пиками при очень маленьком и при полном перекрытии. Действительно, ведь главный вклад в распределение (80) дают состояния с низкой энергией, которые в МСЭ предполагаются независимыми, т.е. не перекрывающимися. Строго говоря, в МСЭ все состояния предполагаются независимыми, но можно было бы думать, что какая-то зависимость для состояний с большой энергией не так важна, раз эти состояния все равно почти не встречаются. Независимо от того, верна ли эта логика, можно сначала посмотреть на состояния с низкой энергией, что и делается с помощью распределения (80). В МСЭ для всех пар с $\alpha \neq \beta$ имеем $Q_{\alpha\beta} = 0$, и все эти пары образуют пик распределения $P(Q)$ при $Q = 0$. С другой стороны, все пары с $\alpha = \beta$ имеют, очевидно, $Q = Q_{\max}$. В итоге получается, что МСЭ предсказывает распределение вероятностей вида

$$P(Q) = x\delta(Q) + (1 - x)\delta(Q - Q_{\max}),$$

где величина x зависит от температуры и сигнализирует о том, как глубоко застыла система; можно показать, что эта величина в точности равна определенному выше параметру $x(T)$ (75).

В работе [125] распределение $P(Q)$ было действительно найдено бимодальным. Конечно, максимумы не были расположены точно при $Q = 0$ и при $Q_{\max}$, что естественно для системы конечных размеров. Поэтому заключение работы [125] состояло в том, что МСЭ справедлива для стеклования гетерополимеров.

Недавно в работе [119] вопрос о применимости МСЭ был подвергнут более внимательному анализу. Было найдено, что МСЭ во многих случаях действительно справедлива, но далека от того, чтобы считаться применимой универсально. МСЭ хорошо работает для максимально компактных конформаций с независимыми взаимодействиями типа МНВ (для которых МСЭ и была исходно подтверждена в работе [125]), но не для других случаев. Например, конформационный аспект иллюстрируется рис. 10. На этом рисунке представлено распределение перекрытий для всех пар конформаций,

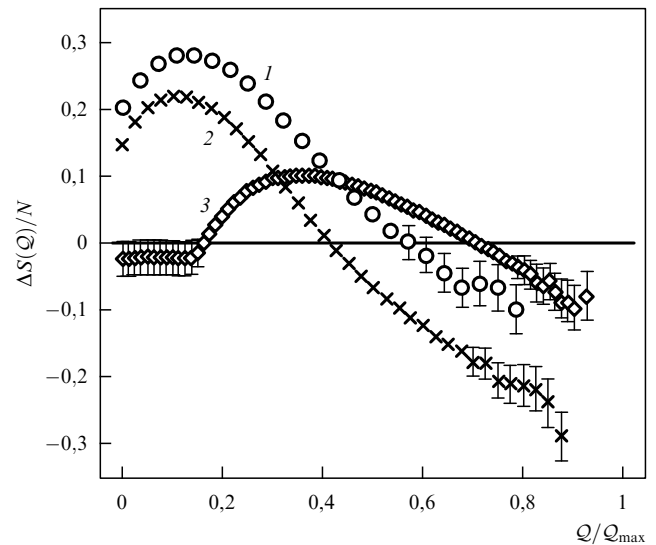


Рис. 10. Зависимости $\Delta S(Q) \equiv \ln[P(Q)/P(Q = Q_{\max})]$ для компактных 27-меров (1), компактных 36-меров (2) и компактных скомканных 64-меров (3). Граница дискретной области меняется существенно: $Q_d/Q_{\max} \approx 0,6; 0,4; 0,7$ соответственно для 27-меров, 36-меров и для скомканных 64-меров. Если рассматривать этот график справа налево, то его можно понимать как зависимость "площади сферы" в конформационном пространстве от радиуса сферы. В самом деле, величина $1 - Q/Q_{\max}$ отмеряет расстояние от данной конформации ("центра"), и мы смотрим, сколько есть других конформаций на этом расстоянии от данной. На нулевом расстоянии всегда есть одна конформация, а именно данная, поэтому всегда $P(Q_{\max}) = 1$. Если расстояние маленькое, т.е. $Q > Q_d$, то обычно в этой области вообще нет конформаций, лишь для некоторых центральных конформаций находится что-то, поэтому после усреднения получается, что в этой области $\Delta S < 0$. Только уже на больших расстояниях "площадь сферы" начинает расти с радиусом. (Сравнительно маленькие значения при совсем малых Q связаны просто с тем, что "пространство кончается" — эффект конечной системы.)

т.е. формально берется распределение (80) при $T = \infty$. Как видим, распределение $P(Q)$ для всех изучавшихся пространств конформаций действительно имеет пик при малом Q . Когда Q растет, то $P(Q)$ экспоненциально убывает (график построен в логарифмическом масштабе) и когда Q превышает определенную величину Q_d , то остаются в лучшем случае всего несколько ($O(1)$) конформаций с таким большим Q (см. подпись под рисунком). Таким образом, все зависит от величины Q_d : если в данном пространстве конформаций Q_d мало, т.е. очень мало пар с большим перекрытием, тогда МСЭ справедлива. Например, для скомканных конформаций величина Q_d наибольшая, что неудивительно, так как они дают наибольшую свободу для локальных конформационных перестроек; для скомканных конформаций применимость МСЭ проблематична. К сожалению, автору не известен никакой аналитический способ даже грубой оценки величины Q_d .

У вопроса о применимости МСЭ кроме геометрии конформаций и конформационных пространств есть и другие аспекты. В частности, для многих моделей с "дискретными" взаимодействиями типична проблема множественного вырождения основного состояния [129]. Это совсем не обязательно нарушает МСЭ, если вырожденные состояния наинизшей энергии не перекрываются. Именно так получается для 27-меров и для 36-меров с поттсовскими взаимодействиями (см. табл. 2). Для них

можно построить распределение $P(Q)$ только для вырожденных основных состояний (т.е. формально взять распределение (80) при $T=0$) и оно по виду оказывается неотличимым от распределения по всем парам конформаций [125, 129]. В свете сказанного выше можно думать, что МСЭ справедлива для этих случаев потому, что величина Q_d достаточно мала. Действительно, для скомканных 64-меров распределение $P(Q)$ для основных состояний при изинговских взаимодействиях оказывается очень сильно смещенным в сторону больших Q , что сигнализирует о нарушении МСЭ [119].

Хотя мы показали, что применимость МСЭ вовсе не универсальна и не гарантирована *a priori* даже в трехмерном случае и, например, нарушается для скомканных конформаций, будем продолжать рассмотрение в рамках МСЭ, так как это приближение все же во многих случаях справедливо.

5. Конструирование гетерополимеров

5.1. Конструирование последовательностей с помощью канонического ансамбля

5.1.1. Почему необходим отбор последовательностей?

Когда был открыт переход стеклования для случайных гетерополимеров [9], многие подумали, что это уже хорошая модель для белков, так как получалось, что даже случайная последовательность дает уникальное пространственное состояние с заметной (главное, не зависящей от N) вероятностью. Довольно скоро, однако, было понято, что это состояние, хотя формально и уникальное, но совершенно не достаточно "грубое" и устойчивое. Действительно, поскольку различие энергий низколежащих состояний в МСЭ пропорционально только $N^{1/2}$, и это несмотря на то, что эти состояния структурно абсолютно различны (с нулевым или малым перекрытием), то достаточно ничтожного ($\sim N^{-1/2}$) изменения параметров взаимодействий (например, за счет небольшого изменения растворителя), чтобы конформация основного состояния перестроилась полностью, до неузнаваемости [130]. Это, конечно, совсем не то, что случается с природными белками, у которых нативные состояния, наоборот, замечательны своей устойчивостью.

Другая сторона проблемы состоит в том, что выбор основного состояния для случайной последовательности, естественно, никак нельзя контролировать. Было бы проблематично найти последовательности, нативные состояния которых обладают какими-то желаемыми свойствами.

Можно провести такую аналогию, которая оказывается неожиданно глубокой. Сворачивание макромолекулы белка можно сопоставить с чтением какого-то послания; больше того, имеется в виду чтение с *пониманием*. Когда мы понимаем прочитанное, то в нашем сознании появляется какой-то образ, навеянный ("индуцированный") смыслом того послания, которое мы читали, и опосредованный всей совокупностью наших предшествующих знаний, настроений и т.п. Тривиальное замечание состоит в том, что этот образ по своей физической природе — объект совершенно отличный от исходной системы, скажем, чернильных знаков на бумаге. Аналогично, сворачивание белковой цепи означает появление трехмерной структуры, которая хоть и

"навеяна" последовательностью, но представляет собой нечто новое, совершенно отличное от одномерной "строки химических букв". Что касается чтения, то для понимания нужно знать язык. Другими словами, чтобы послание можно было прочесть и понять, оно сначала должно было быть как-то написано, желательно со знанием того же языка, на котором его предполагается читать. Случайную последовательность букв читать трудно. Эта аналогия наводит на мысль, что гетерополимерные последовательности тоже должны быть написаны, или сконструированы. Больше того, для конструирования нужно использовать тот же язык, который работает при "чтении = сворачивании", или хотя бы похожий язык.

Что же играет роль языка для сворачивания белков? Если понимать "язык читателя" как инструмент для перекодировки послания как такового в образ, связанный со смыслом послания, то становится понятно, что эту роль для белков играют взаимодействия между мономерами. Поскольку "язык писателя" — это инструмент кодировки какого-то понимания (которое у писателя, предположительно, должно быть) в форме послания, то мы приходим к мысли, что конструирование последовательностей должно использовать в качестве основного инструмента взаимодействия между мономерами.

На что же должны быть направлены усилия конструирующего последовательности? Поскольку цель состоит в получении более устойчивого основного состояния, необходимо отбирать нетипические реализации, у которых основное состояние "оттянуто" возможно дальше вниз, к более низким энергиям, от границы непрерывного спектра. Желательно добиться, чтобы энергия основного состояния была на величину порядка N ниже этой границы. Соответствующие процедуры были предложены в работах [12, 13]; они называются, соответственно, методом размораживания последовательностей и методом отпечатка. Эти два подхода были сначала разработаны независимо, но потом было понято, что они эквивалентны, по крайней мере, с точностью приближения самосогласованного поля. Кстати, об обоих можно при желании сказать, что они являются реализациями принципа "минимальной фрустрации" [7] (см. также [131]). Вместо чисто формального подхода, понятнее начать с простого описания обеих процедур.

5.1.2. Метод отпечатка. Идея метода поясняется рис. 11.

Чтобы начать процесс, необходимо подготовить две компоненты. Во-первых, мы берем систему не связанных мономеров в таком количестве и в таком наборе, какой мы желаем получить длину и состав полимерной цепи. Мы смешиваем мономеры, заключаем их в "коробку" (или внешнее поле), где они поддерживаются в концентрации близкой к плотности упаковки, и позволяем им достичь термодинамического равновесия при какой-то "приготовительной" температуре T_{des} (этап 2 на рис. 11). Разумеется, и это решающе важно, в ходе этой релаксации мономеры взаимодействуют с энергиями V_{ij} в соответствии с тем, что мы говорили выше о языке. Во-вторых, независимо от того, что происходит с мономерами, мы выбираем конформацию, которую хотели бы видеть как нативную у будущего полимера (этап 3 на рис. 11); мы называем эту конформацию $*$. На рисунке все это показано даже в более амбициозном варианте,

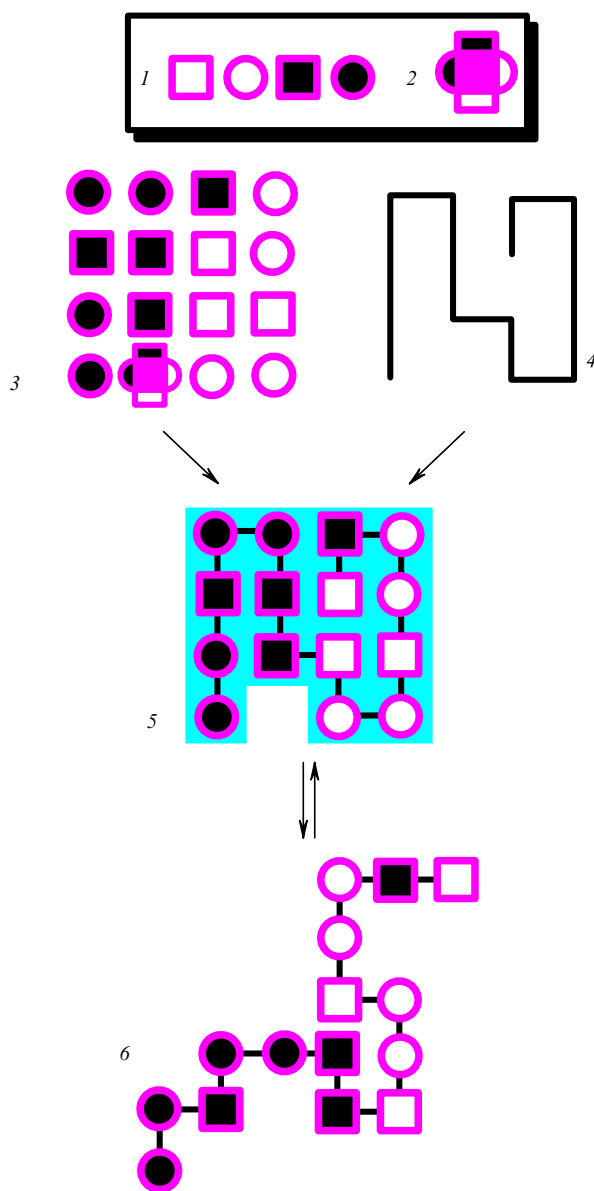


Рис. 11. Схематическое двумерное представление метода отпечатка. Допустим, что у нас есть четыре сорта мономеров (1) и мы хотим сконструировать полимер, который будет специфически распознавать некоторый субстрат (2). Пусть взаимодействия в системе таковы, что притягиваются друг к другу одинаковые формы (круглое к круглому, прямое к прямому) и одинаковые цвета (белое к белому, черное к черному). Возьмем смесь мономеров и молекулы субстрата и дадим этой смеси отрелаксировать (3). Отдельно выберем какую-нибудь конформацию $\star$ (4). Теперь произведем полимеризацию мономеров в смеси 3 вдоль конформации 4. В результате получится глобула 5, имеющая в самом деле желаемый "карман" похожий до известной степени на активный центр фермента. Если теперь наша глобула развернется, или "денатурирует" (6), то она сможет сама вернуться в свое нативное состояние $\star$.

когда мы хотим получить конформацию с активным центром, готовым специфически распознавать желаемую молекулу субстрата. Эта идея в больших подробностях разработана в [13], здесь мы не будем на ней останавливаться.

Когда оба названных "ингредиента" готовы, мы приступаем к изготовлению "блюда": мы мгновенно и необратимо прошиваем мономеры вдоль пути, обозна-

ченного выбранной конформацией $\star$. Таким образом, появляется полимерная цепь, и мы считаем, что ее последовательность в момент приготовления замораживается и более не меняется.

Надежда состоит в том, что $\star$ окажется основным состоянием. Почему мы на это надеемся? Потому что, пока мономеры не были соединены цепью и релаксировали под влиянием взаимодействий B_{ij} , они могли принять конфигурацию с низкой энергией, т.е. с большим количеством низкоэнергетических (на теоретическом жаргоне — "хороших") связей, особенно, если температура приготовления T_{des} была достаточно низкой. Если теперь полимер приходит к конформации $\star$, то у него получается много связей низкой энергии. Конечно, некоторые из таких связей выводятся из игры, когда по ним проходит путь полимеризации. Таких связей, однако, сравнительно немного. Например, на кубической решетке каждая частица имеет шесть связей, из них две уходят на полимеризацию, и, значит, $2/3$ связей остаются. Для безрешеточной системы, особенно, если мономеры крупные, эта доля может быть еще больше. В общем случае доля связей вдоль цепи среди всех связей управляется параметром v/a^3 , который хорошо известен в теории полимеров [32]. Приближение среднего поля отвечает наинизшему порядку по параметру v/a^3 , поэтому по крайней мере среднеполевая теория предсказывает, что конформация $\star$ будет в самом деле очень хорошо оптимизирована для полимера.

Название метода связано с тем, что молекула субстрата на рис. 11 как бы отпечатывается в последовательности.

5.1.3. Метод размораживания последовательности. Хотя и не очень ясно, как осуществить метод отпечатка на практике, но все же он сформулирован так, что по крайней мере человек, не знающий химии, может подумать, что осуществить его можно. (По мнению автора, эта задача трудная, но не безнадежная.) Альтернативный метод размораживания последовательности был предложен в работе [12] и формулируется в более абстрактном или, если угодно, в более честном виде как рецепт для компьютерного моделирования. В этом подходе мы начинаем с полимера и закрепляем его в желаемой конформации $\star$, а после этого разрешаем мономерам обмениваться местами при закрепленной конформации хребта полимера. Другими словами, конформация считается замороженной, а последовательность — размороженной, в точности наоборот по сравнению с обычной физической ситуацией. Повторим, что представить себе такой процесс химически трудно, если не невозможно, но мы на это и не претендуем. Важно, что конформация $\star$ — глобулярная, мономеры в ней контактируют не только и не столько с соседями по цепи, сколько совсем с другими мономерами. При этом, как и в методе отпечатка, мономеры взаимодействуют с энергиями B_{ij} . Мы позволяем им отрелаксировать, перепрыгивая вдоль цепи с места на место, при какой-то температуре T_{des} и затем останавливаем процесс и замораживаем получившуюся последовательность.

Можно заключить, что оба метода очень похожи и единственная разница между ними состоит в том, что в методе отпечатка мономеры релаксируют в обычном пространстве, а в методе размораживания — в пространстве с выколотыми полимерными связями (см. рис. 7).

5.1.4. Микроканоническое и каноническое конструирование последовательностей. Теперь мы готовы к тому, чтобы обсудить конструирование последовательностей с более общей точки зрения. Поскольку, как уже говорилось, цель конструирования состоит в том, чтобы отобрать последовательности с оптимизированной энергией E_{NS} основного (нативного) состояния, то можно было бы попробовать так и сформулировать рецепт: отобрать последовательности с данным значением E_{NS} . Такой ансамбль последовательностей был бы похож на микроканонический ансамбль состояний в статистической механике. Работать с таким ансамблем, как и с микроканоническим, неудобно.

В статистической механике технически гораздо удобнее канонический ансамбль, где вместо энергии фиксируется температура. Мы используем похожую идею для конструирования последовательностей. Мы не считаем, что энергия E_{NS} зафиксирована, а стараемся управлять ею с помощью некоторой искусственной температуры T_{des} . Другими словами, мы выбираем "целевую" конформацию $\star$ и связываем ее энергию $E_\star = \mathcal{H}(\text{seq}, \star)$ множителем Лагранжа $1/T_{des}$. В таком каноническом ансамбле каждая последовательность появляется с гиббсовской вероятностью:

$$P_{\text{seq}}^\star = P_{\text{seq}}^{(0)} \exp \left[-\frac{\mathcal{H}(\text{seq}, \star)}{T_{des}} \right] \left\{ \sum_{\text{seq}} P_{\text{seq}}^{(0)} \exp \left[-\frac{\mathcal{H}(\text{seq}, \star)}{T_{des}} \right] \right\}^{-1}, \quad (81)$$

где $P_{\text{seq}}^{(0)}$ — вероятность появления последовательности при полностью случайном приготовлении (57). Очевидно, как метод отпечатка, так и размораживание последовательностей — частные реализации более общей идеи канонического конструирования.

Таким образом, мы характеризуем канонический ансамбль последовательностей температурой T_{des} : при низкой T_{des} моделируем последовательности с очень эффективно оптимизированным основным состоянием, при высокой T_{des} практически имеем неизменный ансамбль случайных последовательностей.

Мы оставили в стороне вопрос о том, все ли конформации $\star$ одинаково допускают конструирование соответствующих последовательностей. Этот вопрос обсуждался в работах [106–108, 132], где он был поставлен следующим образом: почему многие белки, совсем разные как по эволюционному происхождению, так и по выполняемой функции, тем не менее имеют похожие конформации, хотя бы в грубых чертах? Были высказаны аргументы в пользу идеи, что выбираются те конформации, для которых легче сконструировать последовательность. Недавно этот вопрос стал предметом впечатляющего своим объемом компьютерного эксперимента [115]: для одного определенного типа взаимодействий (обозначенного как NEC в табл. 2), выбор которого, впрочем, не был как-либо серьезно обоснован, авторы вычислили энергии всех 2^{27} возможных двухбуквенных последовательностей 27-мера во всех 103346 его компактных конформациях. Результаты указывают на то, что, действительно, разные конформации имеют весьма различное количество последовательностей, для которых они служат основными состояниями. Теоретический анализ этого вопроса показывает, что эффект обусловлен либо конечностью системы, либо скопленным характером конформаций [132], либо непол-

ной компактностью глобулы. В любом случае этот эффект выходит за рамки нашего аналитического рассмотрения; мы будем рассматривать все (компактные) конформации $\star$ "демократически", на равных правах.

5.1.5. Энергия нативного состояния. Этот раздел посвящен вычислению энергии нативного (целевого) состояния $\star$, усредненного по последовательностям с распределением (81). Имеем

$$\begin{aligned} \langle E_\star(\text{seq}) \rangle &= \sum_{\text{seq}} P_{\text{seq}}^\star E_\star(\text{seq}) = \\ &= \frac{\sum_{\text{seq}} P_{\text{seq}}^{(0)} \exp \left[-\frac{\mathcal{H}(\text{seq}, \star)}{T_{des}} \right] \mathcal{H}(\text{seq}, \star)}{\sum_{\text{seq}} P_{\text{seq}}^{(0)} \exp \left[-\frac{\mathcal{H}(\text{seq}, \star)}{T_{des}} \right]}. \end{aligned} \quad (82)$$

Сама структура этой формулы наводит на мысль использовать простой трюк (очень похожий на то, что всегда делается на первых нескольких страницах любого учебника статистической механики). Давайте вернемся к определению размороженной статистической суммы (68) и рассмотрим ее как функцию температуры:

$$\begin{aligned} W^\star(T) &= \left\langle \exp \left[-\frac{\mathcal{H}(\text{seq}, \star)}{T} \right] \right\rangle = \\ &= \sum_{\text{seq}} P_{\text{seq}}^{(0)} \exp \left[-\frac{\mathcal{H}(\text{seq}, \star)}{T} \right]. \end{aligned} \quad (83)$$

Теперь мы можем написать

$$\langle E_\star(\text{seq}) \rangle = -\frac{1}{W^\star} \frac{\partial W^\star}{\partial (1/T)} \bigg|_{T \rightarrow T_{des}} = -\frac{\partial \ln W^\star}{\partial (1/T)} \bigg|_{T \rightarrow T_{des}}. \quad (84)$$

На данном этапе мы даже не использовали никаких приближений. Чтобы получить конкретный результат, обратимся к формуле (68), выражающей размороженное среднее свободной энергии в терминах первого члена высокотемпературного разложения. В этом приближении находим, что $\ln W^\star(T)$ не зависит от $\star$ и равняется

$$-\ln W^\star(T) = \frac{Q\bar{B}}{T} - \frac{Q\delta B^2}{2T^2}. \quad (85)$$

Следовательно, величина $\langle E_\star \rangle$ равна

$$\langle E_\star \rangle = -\frac{\partial \ln W^\star}{\partial (1/T)} \bigg|_{T \rightarrow T_{des}} = Q \left[\bar{B} - \frac{\delta B^2}{T_{des}} \right]. \quad (86)$$

Это очень важный результат. Мы прямо видим, как температура конструирования влияет на энергию основного состояния. Заметим, что "восприимчивость" целевого состояния к попыткам конструирования определяется разбросом энергий взаимодействия δB : естественно, что это чисто гетерополимерный эффект.

В МСЭ предполагается весьма своеобразный характер энергетического спектра для сконструированных последовательностей. Действительно, в МСЭ, где состояния независимы, конструирование влияет только на энергию целевой конформации $\star$, тогда как статистика всех остальных уровней энергии остается неизменной. В этом смысле и надо понимать теоретическое жаргонное выражение, что конструирование означает просто "оттягивание вниз" основного энергетического уровня. Эта идея иллюстрируется рис. 12. Конструирование означает отбор нетипичных реализаций из ансамбля реализаций МСЭ (см. раздел 4.2.3).

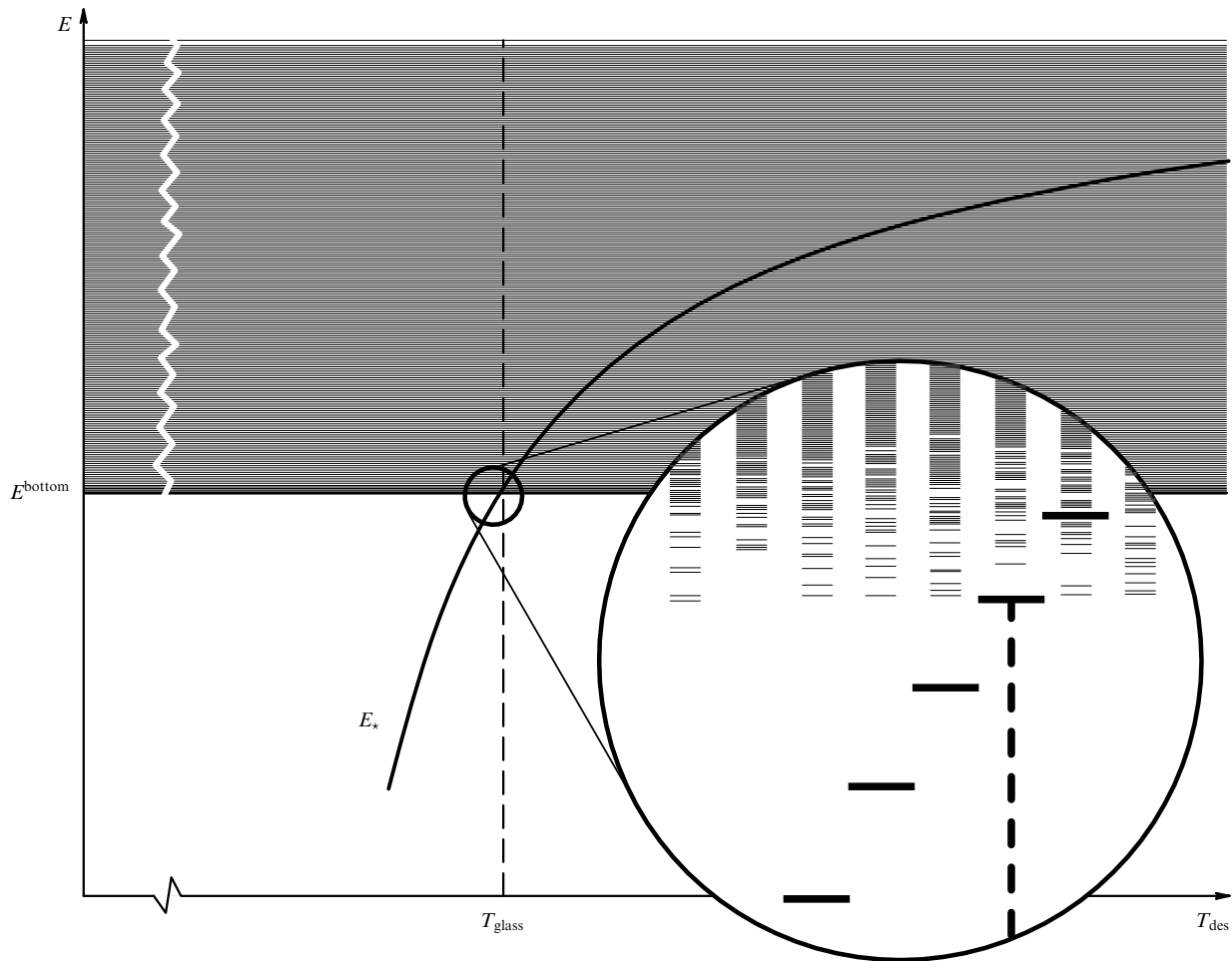


Рис. 12. Типичные энергетические спектры для последовательностей, сконструированных при различных температурах T_{des} . В области $T_{\text{des}} \approx T_{\text{glass}}$ (показанной в увеличенном масштабе), как видим, энергия E_* близка к границе непрерывного спектра E^{bottom} , или к средней энергии основного состояния для случайной цепочки. Это связано с фазовым переходом между свернутой (нативной) и стеклообразной фазами (см. фазовую диаграмму на рис. 13).

Отметим, что в более реалистическом приближении за рамками чистой МСЭ выясняется, что наряду с основным состоянием конструирование влияет также на целую область конформационного пространства (особенно на состояния с $Q > Q_d$), что важно для кинетики.

5.1.6. Свернутая (нативная) фаза и температура сворачивания. Своеобразный характер энергетического спектра приводит к тому, что стеклование для сконструированных последовательностей происходит тоже необычно. Действительно, если мы конструируем при бесконечной температуре ($T_{\text{des}} = \infty$), то никакого отбора последовательностей не осуществляем ни в смысле их энергии в целевом состоянии $\star$, ни в каком бы то ни было другом смысле. Но когда мы понижаем T_{des} , то уже выбираем последовательности, которым конформация $\star$ доставляет низкую энергию. При каком-то значении T_{des} энергия конформации $\star$ становится ниже энергии типичного основного состояния МСЭ и тогда практически наверняка $\star$ оказывается основным состоянием, и мы ожидаем, что система будет застывать в состоянии $\star$. Это происходит при $T_{\text{des}} < T_{\text{glass}}$, так как основное состояние МСЭ доминирует при температуре T_{glass} . Больше того, последовательности, сконструированные

при $T_{\text{des}} < T_{\text{glass}}$, застывают при температуре выше T_{glass} , потому что их основное состояние более устойчиво и имеет более низкую энергию.

Таким образом, при $T_{\text{des}} < T_{\text{glass}}$ целевая конформация будет лучше оптимизирована, чем основное состояние типичной случайной последовательности, и поэтому вместо обычного стеклования при температуре T_{glass} полимеры с такими последовательностями будут сворачиваться в наперед заданную конформацию $\star$ при более высокой температуре T_{fold} . Мы можем найти эту температуру T_{fold} , сравнивая энергию целевого состояния со свободной энергией случайной глобулы (напомним, что когда система сворачивается в целевую конформацию, то энтропия ее равна нулю).

Итак, сравним целевую энергию $\langle E_\star \rangle$ (86) со свободной энергией случайной глобулы, для которой наиболее удобно воспользоваться выражением в первой строке формулы (74). Приравняв эти две величины, находим следующее выражение для температуры сворачивания T_{fold} :

$$\frac{1}{T_{\text{fold}}^2} + \frac{1}{T_{\text{glass}}^2} = \frac{2}{T_{\text{fold}} T_{\text{des}}} . \quad (87)$$

Заметим, что это соотношение не зависит от специальных свойств матрицы энергий $\hat{B} = B_{ij}$, и только разброс δB^2 энергий входит через T_{glass} . Так получилось из-за отбрасывания дальнейших членов высокотемпературного разложения. Это приближение разумно, если число типов мономеров велико и корреляции элементов матрицы $\hat{B}$ слабы. Для других случаев необходимо удерживать дальнейшие члены [133, 134].

Результаты этого раздела суммированы на фазовой диаграмме на рис. 13. Все фазовые границы определены прямо по размороженной статистической сумме, что позволяет просто рассмотреть даже довольно деликатные модели, если только верна МСЭ.

Подчеркнем, что МСЭ не только помогает делать вычисления, но и определяет, в принципе, возможность применить нашу процедуру для конструирования последовательностей. Именно статистическая независимость состояний позволяет манипулировать одним из них, не затрагивая остальных.

Фазовая диаграмма, показанная на рис. 13, позволяет дать новую формулировку принципу оптимизации последовательностей: можно сказать, что оптимизация состоит в выборе возможно большего отношения двух характерных температур, $T_{\text{fold}}/T_{\text{glass}}$. Эта формулировка становится очень популярной. Однако, по мнению автора, она мало полезна: нет же способа определить T_{fold} или T_{glass} , глядя на последовательность. Чтобы определить их, нужно выполнить расчет или эксперимент в трехмерном пространстве, а в этом случае можно уже прямо определить способность цепи правиль-

но сворачиваться, не обращаясь ни к какому косвенному критерию.

5.1.7. Где же реплики? Хотя в этой работе мы целенаправленно избегаем использования метода реплик, некоторые читатели могут быть с ним знакомы и могут интересоваться связью нашего подхода с более стандартным, хоть и более громоздким репличным. Если читатель этим вопросом не интересуется, то может перейти к разделам 5.2 и 6.

Чтобы вскрыть параллелизм с репликами, рассмотрим размороженную среднюю статистическую сумму $W^*(T)$ при специальной температуре, определенной соотношением

$$\frac{1}{T_{\text{eff}}} = \frac{1}{T_{\text{des}}} + \frac{n}{T}, \quad (88)$$

и после этого перепишем уравнение (84) в виде

$$\left\langle \frac{E_*(\text{seq})}{T} \right\rangle = - \left. \frac{\partial W^*(T_{\text{eff}})}{\partial n} \right|_{n \rightarrow 0}. \quad (89)$$

Как и следовало ожидать, W по форме аналогична репличной статистической сумме Z^n . Действительно, в методе реплик мы имеем размороженное среднее с участием n репличных гамильтонианов. В нашем методе мы не должны интерпретировать n как реплики, а просто как внешнее поле, которое в конце считаем равным нулю (подобно стандартному определению восприимчивости в нулевом поле). Неудивительно поэтому, что результаты нашего метода точно совпадают с результатами соответствующих репличных вычислений [134].

Конечно, обход трудностей никогда не дается бесплатно. В данном случае мы оплачиваем обход трудных репличных вычислений широким использованием МСЭ. Фактически мы можем работать с нашим методом только выше температуры стеклования и можем определить температуру перехода как границу той области, где наш подход еще годится. Мы не можем изучить структуру застеклованной фазы. В частности, мы не могли вывести результат (76). Представляется, однако, что такое ограничение возможностей — приемлемая цена за радикальное упрощение теории.

В последовательной репличной теории можно пытаться выйти за рамки МСЭ [136, 137]. Впрочем, и наш подход допускает некоторые обобщения [122].

5.1.8. Компьютерные тесты конструирования. Есть по меньшей мере две причины, по которым полученные выше заключения стоит проверить моделированием:

1. Все результаты существенно опираются на МСЭ, применимость которой не очевидна.

2. Как всегда в статистической механике, мы используем термодинамический предел. Реальные белки длинные, но все же числа частиц порядка 100 не только меньше того, к чему привыкли в обычной термодинамике, но даже меньше, чем у многих полимерных цепей. Интересно посмотреть, как работает конструирование последовательностей для сравнительно малых систем.

Обе схемы конструирования последовательностей [12, 13, 138] интенсивно изучались путем моделирования. Компьютерные тесты в целом говорят, что конструирование работает хорошо. Если говорить более

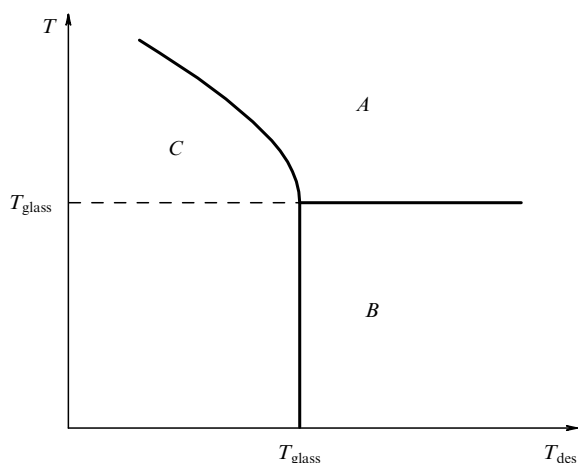


Рис. 13. Диаграмма состояний глобулярного гетерополимера. Отмечены три фазы (еще раз подчеркнем, что все три — глобулярные): *A* — случайная: эта фаза очень похожа на гомополимерную глобулу, в ее равновесии участвуют экспоненциально много состояний; *B* — стеклообразная: эта фаза реализуется для тех последовательностей, которые не слишком хорошо оптимизированы (достаточно высокая T_{des}), одна или порядка одной конформации доминируют в равновесии ниже T_{glass} , но эти состояния никак не связаны с целевым, не допускают какого-либо контроля и крайне неустойчивы; *C* — свернутая (нативная): целевая конформация $\star$ доминирует в равновесии; при $T_{\text{des}} < T_{\text{glass}}$ конформация $\star$ оптимизирована лучше, чем основное состояние типичной случайной последовательности. Легко понять, почему граница между стеклообразной и свернутой (нативной) фазами вертикальна: поскольку энтропия равна нулю в обеих фазах, невозможно вызвать переход одной из них в другую путем изменения температуры.

конкретно, то что должно быть проверено на модели? Если последовательность приготовлена (в компьютере), то вот вопросы, подлежащие проверке:

1. Является ли конформация $\star$ основным состоянием, т.е. верно ли, что энергии *всех* остальных состояний больше?

2. Если состояние $\star$ — основное, то является ли оно также невырожденным?

3. Наконец, если состояние $\star$ — невырожденное основное состояние, то доступно ли оно кинетически?

Неудивительно, что ситуация существенно зависит от характера взаимодействий B_{ij} . В работах [13, 138] подробно изучен случай поттсовских взаимодействий с q сортами мономеров, где $B_{ij} = J(1 - 2\delta_{ij})$ (мономеры одного типа притягиваются с энергией $-J$, мономеры разных типов отталкиваются с энергией $+J$). Заметим, что как случай $q = 1$, так и случай $q = N$ отвечают гомополимеру (в последнем случае все от всех одинаково отталкиваются)¹¹. Следовательно, должен быть оптимум при каком-то промежуточном значении q . Было найдено, что оптимальное значение q равняется 7 для 27-меров и 9 для 36-меров. При оптимальном q больше половины всех сконструированных последовательностей имели целевую конформацию $\star$ в качестве невырожденного основного состояния.

Если предположить, что оптимальное число q зависит линейно от длины цепи N , и экстраполировать результаты до длины типичного белкового домена (порядка 100), то получается воодушевляющая оценка $q \approx 20$, что в точности равно числу аминокислот, имеющихся в природе. Конечно, это скорее всего просто совпадение (а совпадение, по словам Н.В. Тимофеева-Ресовского, это просто падение сов ...).

Методом Монте-Карло активно изучалась и кинетика сворачивания гетерополимеров со сконструированными последовательностями. В целом сконструированные гетерополимеры показывают хорошую способность сворачиваться в соответствующую целевую конформацию $\star$. Конечно, для этого нужно подобрать соответствующую температуру.

Впечатляющий результат был достигнут Шахновичем в работе [135]. Автор работал с довольно длинными цепями — до 100 мономеров. Для таких длин нет никакого способа точно знать конформацию, отвечающую основному состоянию, так как перечислить состояния невозможно (см. раздел 4.6.2). Тем не менее конструировать последовательности вполне можно и можно выполнить кинетический эксперимент методом Монте-Карло, начиная с развернутой клубковой конформации. Было найдено, что цепочки действительно релаксируют к целевой конформации $\star$, и происходит это за вполне обозримое время.

Таким образом, данная компьютерная модель умеет как-то обходить парадокс Левинталя (компьютер не мог бы перепробовать все состояния 100-мера). Это значит, что идея конструирования последовательностей — правильный шаг на пути к разрешению парадокса.

Другой интересный результат такой. Уже отмечалось, что 36-мер может иметь конформации с узлом

(точнее, с квазиузлом, так как цепь линейная и имеет свободные концы). Интересы ради мы сконструировали последовательность, для которой основное состояние есть конформация с узлом (для 36-мера мы можем перечислить все компактные конформации и знаем основное состояние точно). Сможет ли такая цепь сложиться в заузленное основное состояние? Смогла! И время релаксации осталось вполне обозримым, всего примерно вдвое дольше обычного.

Здесь мы не касаемся никаких деталей методов моделирования и приводим только один рис. 14, который показывает хорошее согласие результатов моделирования с аналитическими результатами для фазовой диаграммы. Стоит отметить, что в этом сравнении нет ни одного подгоночного параметра.

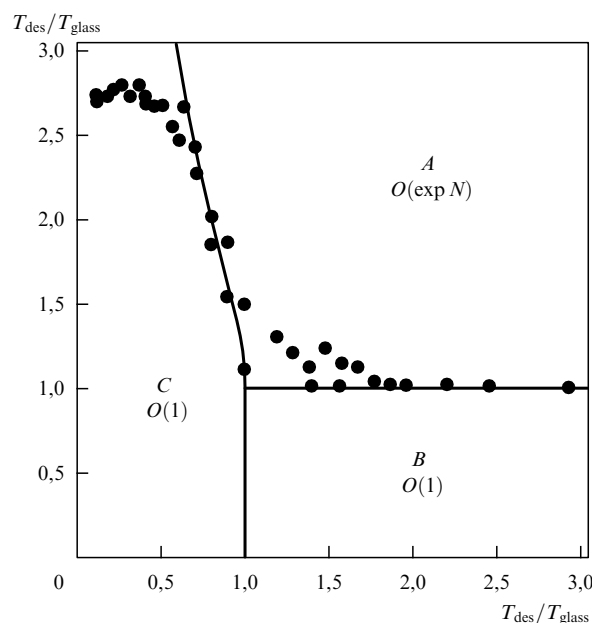


Рис. 14. Фазовая диаграмма сконструированного гетерополимера с семью сортами мономеров и поттсовскими взаимодействиями: данные компьютерного моделирования для компактного 27-мера на решетке $3 \times 3 \times 3$ и предсказание аналитической теории. Компьютерная модель использует размораживание и осуществляет его методом Монте-Карло при данном значении T_{des} . Когда последовательность готова, то статистическая сумма для нее вычисляется точно (так как известен список всех 103346 компактных конформаций). Температура сворачивания или стеклования определяется из условия $x(T) = 0,9$. Так как система конечна, то температура сворачивания становится постоянной (не зависящей от T_{des}) при $T_{\text{des}}/T_{\text{glass}} < 0,5$, поскольку мы просто достигаем максимально возможной степени оптимизации для 27-мера. В области $T_{\text{des}}/T_{\text{glass}} \sim 1$ тоже видны значительные флуктуации. Чтобы состыковать аналитические и компьютерные результаты, измеряем T_{glass} для случайных последовательностей прямо из модели при высокой температуре T_{des} . Используя измеренное значение T_{glass} и вычисленную величину δB , строим теоретические фазовые границы, беря только один член высокотемпературного разложения $O(1/T^2)$.

Модель Поттса с 7 типами мономеров была выбрана для этого рисунка потому, что она "наиболее гетерополимерна" [13] для 27-меров. (Напомним, что для поттсовских взаимодействий мы получаем гомополимер как при $q = 1$, так и при $q = 27$: оптимум должен быть где-то в этом интервале.) Для других моделей поведение похожее, но, например, для модели Поттса с $q = 2$ (т.е. для

¹¹ Для простоты мы не говорим здесь об общем фоне среднего притяжения, которое всегда предполагается и которое держит полимер в состоянии максимально компактной глобулы.

модели Изинга) необходимо удерживать больше членов высокотемпературного разложения, чтобы добиться количественного согласия.

5.2. Конструирование и сворачивание с разными взаимодействиями

5.2.1. Почему взаимодействия могут быть разными?

Допустим, что природные белки эволюционно прошли, в самом деле, процесс конструирования последовательностей более или менее того типа, который мы описали. Очевидно, этот процесс управлялся природными реальными взаимодействиями. Представим себе теперь, что мы пытаемся разработать компьютерный алгоритм для предсказания пространственной структуры белков по последовательностям. По существу это значит, что мы пытаемся сворачивать белки, используя компьютерные модели, где взаимодействия по необходимости не точно соответствуют реальным. Если наши приблизительные энергии хорошо приближают реальность, то все, наверное, будет в порядке, а если — не очень хорошо, то наши шансы на успех проблематичны. (В действительности компьютерное предсказание структур имеет много и других трудностей помимо вопроса о точности потенциалов; для целей нашего обсуждения мы можем от них отвлечься.) Можно представить себе и обратную постановку: мы пытаемся научиться писать (на бумаге или в компьютере) последовательности так, чтобы синтезированные по нашему предписанию реальные цепочки сворачивались в желаемую конформацию. В обоих случаях получается, что конструирование последовательностей и сворачивание цепочек управляется несколько различными взаимодействиями. Подобная ситуация возникнет и тогда, когда последовательности будут конструироваться в пробирке, но условия растворителя при этом будут несколько отличаться от тех, при которых готовый полимер сворачивается.

Для всех таких случаев возникает несколько вопросов:

1. Насколько различны могут быть потенциалы взаимодействий, чтобы все еще обеспечивать правильное сворачивание?

2. Как следует характеризовать "сходство" или различие взаимодействий?

3. Какова фазовая диаграмма системы?

Чтобы подойти к этим вопросам мы используем два разных гамильтониана: конструируем последовательно-сти с помощью

$$\mathcal{H}^{\text{des}}(\text{seq}, \text{conf}) = \sum_{I,J}^N B_{s_I s_J}^{\text{des}} \Delta(\mathbf{r}_I - \mathbf{r}_J) \quad (90)$$

и потом предполагаем, что цепочки сворачиваются, имея гамильтониан

$$\mathcal{H}^{\text{fold}}(\text{seq}, \text{conf}) = \sum_{I,J}^N B_{s_I s_J}^{\text{fold}} \Delta(\mathbf{r}_I - \mathbf{r}_J). \quad (91)$$

Мы используем канонический ансамбль (81) с гамильтонианом $\mathcal{H}^{\text{des}}$, чтобы конструировать последовательности для состояния $\star$.

5.2.2. Энергия целевого состояния. В полной аналогии с тем, что мы делали раньше в разделе 5.1.5, мы можем усреднить энергии целевого состояния $\star$ по последова-

тельности:

$$\begin{aligned} \langle E_{\star}(\text{seq}) \rangle &= \sum_{\text{seq}} P_{\text{seq}}^{\star} E_{\star}(\text{seq}) = \\ &= \frac{\sum_{\text{seq}} P_{\text{seq}}^{(0)} \exp[-\mathcal{H}^{\text{des}}(\text{seq}, \star)/T_{\text{des}}] \mathcal{H}^{\text{fold}}(\text{seq}, \star)}{\sum_{\text{seq}} P_{\text{seq}}^{(0)} \exp[-\mathcal{H}^{\text{des}}(\text{seq}, \star)/T_{\text{des}}]}. \end{aligned} \quad (92)$$

Единственное отличие от формулы (82) состоит в том, что теперь задействованы два гамильтониана, $\mathcal{H}^{\text{des}}$ и $\mathcal{H}^{\text{fold}}$. Тем не менее мы по-прежнему можем использовать тот же трюк: определим размороженную среднюю статистическую сумму (аналог (83)) для эффективного гамильтониана $\mathcal{H}_{\text{eff}}$:

$$\begin{aligned} W^{\star} &= \left\langle \exp \left[-\frac{\mathcal{H}_{\text{eff}}(\text{seq}, \star)}{T} \right] \right\rangle = \\ &= \sum_{\text{seq}} P_{\text{seq}}^{(0)} \exp \left[-\frac{\mathcal{H}_{\text{eff}}(\text{seq}, \star)}{T} \right], \end{aligned} \quad (93)$$

где

$$\frac{\mathcal{H}_{\text{eff}}(\text{seq}, \star)}{T} = \frac{\mathcal{H}^{\text{des}}(\text{seq}, \star)}{T_{\text{des}}} + n \frac{\mathcal{H}^{\text{fold}}(\text{seq}, \star)}{T}. \quad (94)$$

Тогда легко проверить, что справедливо следующее соотношение:

$$\frac{\langle E_{\star}(\text{seq}) \rangle}{T} = -\frac{1}{W^{\star}} \frac{\partial W^{\star}}{\partial n} \bigg|_{n=0} = -\frac{\partial \ln W^{\star}}{\partial n} \bigg|_{n=0} \quad (95)$$

(ср. с (89)).

Последняя формула представляет собой полное (в рамках МСЭ) решение задачи, так как позволяет хотя бы в принципе использовать различные приближения или феноменологические интерполяции для размороженной средней статистической суммы $W^{\star}$. Как и выше (68), мы для конкретности используем младший член высокотемпературного разложения (68) (конечно, с новой матрицей взаимодействий):

$$\frac{\hat{B}_{\text{eff}}}{T} = \frac{\hat{B}^{\text{des}}}{T_{\text{des}}} + n \frac{\hat{B}^{\text{fold}}}{T}. \quad (96)$$

В главном порядке по $1/T$ и по $1/T_{\text{des}}$ размороженная средняя свободная энергия не зависит от конформации $\star$ (опять-таки потому, что все компактные конформации имеют одинаковое число контактов между мономерами Q) и равна

$$-\ln W^{\star} = \frac{Q \overline{B}_{\text{eff}}}{T} - \frac{Q \delta B_{\text{eff}}^2}{2T^2}, \quad (97)$$

где

$$\frac{\overline{B}_{\text{eff}}}{T} = \frac{\overline{B}^{\text{des}}}{T_{\text{des}}} + n \frac{\overline{B}^{\text{fold}}}{T}, \quad (98)$$

$$\begin{aligned} \frac{\delta B_{\text{eff}}^2}{T^2} &= \sum_{ij} p_i \left[\frac{B_{ij}^{\text{des}} - \overline{B}^{\text{des}}}{T_{\text{des}}} + n \frac{B_{ij}^{\text{fold}} - \overline{B}^{\text{fold}}}{T} \right]^2 p_j = \\ &= \frac{\delta B_{\text{des}}^2}{T_{\text{des}}^2} + 2n \sum_{ij} p_i \frac{\delta B_{ij}^{\text{des}}}{T_{\text{des}}} \frac{\delta B_{ij}^{\text{fold}}}{T} p_j + n^2 \frac{\delta B_{\text{fold}}^2}{T^2}, \end{aligned} \quad (99)$$

и, следовательно,

$$\begin{aligned} \langle E_* \rangle &= Q \left[\overline{B^{\text{fold}}} - \frac{1}{T_{\text{des}}} \sum_{ij} p_i \delta B_{ij}^{\text{des}} \delta B_{ij}^{\text{fold}} p_j \right] = \\ &= Q \left[\overline{B^{\text{fold}}} - \frac{\delta B_{\text{eff}}^2}{T_{\text{des}}} \right], \end{aligned} \quad (100)$$

где

$$\delta B_{\text{eff}}^2 = \sum_{ij} p_i \widehat{\delta B}_{ij}^{\text{des}} \widehat{\delta B}_{ij}^{\text{fold}} p_j. \quad (101)$$

Здесь, конечно, напрашивается аналогия с формулой (86). Однако характерный разброс энергии определяется теперь своего рода корреляцией элементов двух матриц взаимодействия. Этот коррелятор имеет интересную геометрическую природу.

5.2.3. Геометрическая интерпретация. Структура полученных формул наводит на мысль о следующей интерпретации. Будем трактовать матрицы $\widehat{B}$ как векторы, хотя их компоненты B_{ij} и нумеруются парными индексами. Вычитание среднего $\delta B_{ij} = B_{ij} - \overline{B}$ в такой геометрической картине означает, что вектор δB_{ij} имеет нулевую проекцию на "главную диагональ" (т.е. на вектор $(1, 1, \dots, 1)$) в этом векторном пространстве. Для любых двух векторов, скажем, A и B , можно определить скалярное произведение:

$$A \cdot B = \sum_{ij} p_i A_{ij} B_{ij} p_j.$$

Подчеркнем, что это скалярное произведение не имеет ничего общего с произведением соответствующих матриц в привычном смысле.

Можно теперь сказать, что и температура перехода стеклования T_{glass} (73), и "восприимчивость" энергии целевого состояния к конструированию (86) определяются квадратом длины вектора δB . Это, конечно, не очень интересная переформулировка.

Более интересен смысл формулы (100): когда матрицы сворачивания B^{fold} и конструирования B^{des} различны, все зависит от угла между δB^{fold} и δB^{des} . В самом деле, этот угол определяется условием $\cos \theta = g$, где

$$g \equiv \frac{\widehat{\delta B}^{\text{des}} \widehat{\delta B}^{\text{fold}}}{\sqrt{(\widehat{\delta B}^{\text{des}} \widehat{\delta B}^{\text{des}})(\widehat{\delta B}^{\text{fold}} \widehat{\delta B}^{\text{fold}})}}. \quad (102)$$

"Параллельные" ($\theta = 0$) взаимодействия идеально скоррелированы, тогда как "перпендикулярные" ($\theta = 90^\circ$) полностью нескоррелированы.

Эта геометрическая интерпретация объясняет параметризацию матрицы взаимодействий для модели с двумя сортами мономеров ("черно-белая" модель, ЧБМ) в терминах угла (см. табл. 2). Вообще говоря, симметричная матрица 2×2 задается тремя параметрами. Выбирая среднее равным нулю и характерный разброс (58) — единице, мы остаемся с единственной степенью свободы. Параметризация ее в терминах угла θ (для случая $p_1 = p_2 = 1/2$)

$$B_{ij}(\theta) = \overline{B} + \frac{\delta B}{2} \begin{pmatrix} -\sqrt{2} \cos \theta - \sin \theta & \sin \theta \\ \sin \theta & \sqrt{2} \cos \theta - \sin \theta \end{pmatrix} \quad (103)$$

обеспечивает выполнение очень удобного соотношения

$$\widehat{\delta B}(\theta_1) \widehat{\delta B}(\theta_2) = \delta B^{(1)} \delta B^{(2)} \cos(\theta_1 - \theta_2). \quad (104)$$

Таким образом, угол действительно определяет метрику в пространстве матриц взаимодействия. Формулу (103) можно обобщить для произвольных p_i , хотя результат становится заметно более громоздким.

Интересно теперь поставить такой вопрос: насколько хороши или разумны модели с двумя сортами мономеров, и какая из них лучше? Мы теперь можем ответить на этот вопрос количественно. Действительно, раз сходство матриц взаимодействия измеряется углом между ними, то нужно вычислить угол $\alpha(\theta)$ между матрицей ЧБМ, отвечающей параметру θ (103), и реалистической матрицей Миязавы–Джернигана. (Для этого нужно сначала воспользоваться теоремой сведения в "обратную сторону" — записать вместо ЧБМ-матрицы размером 2×2 матрицу 20×20 , но такую, что у нее первые десять строк одинаковые и последние десять строк одинаковые и так же со столбцами.) В итоге подсчета можно убедиться, что угол $\alpha(\theta)$ наименьший при $\theta \approx 0$, т.е. наилучшая модель, как и ожидалось, — гидрофобно-полярная (ГП). Однако даже наименьшее значение $\alpha(\theta)$ оказывается около 60° и трудно себе представить, чтобы было особенно хорошим приближение, в котором считаются параллельными векторы с углом 60° .

5.2.4. Диаграмма состояний. Чтобы выяснить, свернется ли правильно полимер, если его последовательность сконструирована с помощью матрицы взаимодействий B^{des} , а сворачивание происходит при взаимодействиях B^{fold} , нужно сравнить энергию целевого состояния (100) со свободной энергией случайной глобулы (74). Приравняв их, находим температуру соответствующего перехода T_{fold} :

$$\frac{\widehat{\delta B}^{\text{fold}} \widehat{\delta B}^{\text{fold}}}{T_{\text{fold}}^2} + \frac{1}{T_{\text{glass}}^2} = 2 \frac{\widehat{\delta B}^{\text{fold}} \widehat{\delta B}^{\text{des}}}{T_{\text{fold}} T_{\text{des}}}. \quad (105)$$

Исходная фазовая диаграмма (рис. 13) модифицируется, следовательно, простым переопределением T_{des} .

5.2.5. Насколько точные потенциалы необходимы для успеха в моделировании сворачивания белков? Статья под таким названием была недавно опубликована [139]. В свете упоминавшейся аналогии с языком, другое название могло бы быть таким: "Насколько косноязычным может быть писатель, чтобы его все-таки понял читатель?" Смысл ответа состоит в том, что сворачивание остается правильным ("читатель понимает") до тех пор, пока T_{des} и T отвечают фазе нативного состояния ниже фазовой границы (105). С ростом угла ошибки это условие рано или поздно нарушается. Таким образом, максимально допустимое значение ошибок зависит как от T_{des} , так и от T , но, что важно, не зависит от N . В этом принципиальное достоинство сконструированных последовательностей. Для случайных последовательностей основное состояние меняется при изменении взаимодействий на величину порядка $N^{-1/2}$ [130].

Если характеризовать степень оптимизации белка параметром $T_{\text{fold}}/T_{\text{glass}} = x$, а условия работы белка параметром $T/T_{\text{glass}} = y$, то необходимый уровень параметра g дается условием

$$g \geq \frac{1+y^2}{y} \frac{x}{1+x^2}. \quad (106)$$

Например, если $x = 1,1$ и $y = 1,05$ (что по оценкам правдоподобно), то получаем $g \geq 0,997$. Даже если $x = 1,2$ и $y = 1,1$, то $g \geq 0,988$. Соответствующие углы не должны тогда превышать примерно 1° или 9° . Главное, конечно, не в числах (которые достаточно произвольны), а в физике: необходимая точность контролируется, с одной стороны, тем, насколько далек белок в рабочих условиях от точки стеклования (y), а с другой стороны, насколько хорошо оптимизирована последовательность (x).

6. Статистика последовательностей реальных белков

6.1. Индивидуальный подход или статистический?

6.1.1. Статистический подход не означает, что последовательности генерируются побуквенно. Если сополимеры синтезируются химически, то получающиеся последовательности хорошо соответствуют представлению о марковской цепи [127]. Разумеется, марковская статистика означает, что получается много разных последовательностей; средние составы и локальные корреляции у них более или менее одинаковы, но в деталях они различны. С другой стороны, азбучная истина молекулярной биологии состоит в том, что последовательности биополимеров находятся под строжайшим генетическим контролем. Если не говорить об относительно редких (хотя и важных) ошибках, то все цепочки, скажем, гемоглобина человека одинаковы, имеют точно одинаковые последовательности. Они специфицированы в той же степени и в том же смысле, что последовательности букв в книгах: в "Войне и мире" — одна, в "Преступлении и наказании" — другая. Так возникает подход, принятый, судя по всему, большинством биологов: изучать каждую последовательность отдельно, индивидуально. Многие верят, что изучение простых моделей не дает ничего полезного для биологии, потому что биологическая функция чрезвычайно чувствительна к сложным молекулярным деталям.

С другой стороны, когда о биополимерах думает физик, то ему кажется естественным обратиться к методам и подходам, которые проявили себя столь успешно в физике систем с беспорядком, таких, как спиновые стекла. При этом основная идея точно противоположна тому, что только что было сказано о биополимерах. Индивидуальная структура образца стекла не может быть проконтролирована, не может быть (в деталях) воспроизведена и не представляет никакого интереса. Осмысленные экспериментальные и теоретические вопросы касаются только соответствующих средних значений: мы интересуемся теми свойствами, которые будут найдены в любом образце стекла, приготовленном при данных условиях.

В предыдущих разделах мы старались показать, что "усредняющий" подход может быть плодотворным для биополимеров. Безусловно, его возможности не следует преувеличивать. Тем не менее статистический подход представляется единственным способным ответить на общие вопросы типа, какие последовательности способны сворачиваться и какие нет, сколь-

ко всего есть сворачивающихся, как трудно их эволюционно отобрать и какими общими свойствами они обладают?

6.1.2. Конструировать последовательности невозможно, если смотреть только на локальные корреляции вдоль цепи. Статистический подход предполагает, что последовательности случайно "выуживаются" из какого-то ансамбля. Если этот ансамбль состоит из всех возможных последовательностей, то случайный выбор из такого ансамбля эквивалентен тому, чтобы последовательно и без корреляций выбирать букву за буквой, как в знаменитом примере с обезьяной за пишущей машинкой¹². Мы показали выше, что подавляющее большинство таких последовательностей разумными моделями белков служить не могут, и предложили другой ансамбль, точнее, целую серию таких ансамблей (81) — последовательности, получающиеся в результате канонического конструирования при той или иной температуре T_{des} . Подчеркнем, что процесс конструирования происходит в трехмерном пространстве и потому не может никаким образом быть описан в терминах простых локальных корреляций вдоль цепи. Тем не менее ансамбль сконструированных последовательностей отличается, конечно, от ансамбля всех, т.е. случайных, последовательностей, даже если на них смотреть, только как на одномерные последовательности, не думая о сворачивании в трехмерном пространстве.

Какие же предсказания о характере последовательностей можно сделать на основании модели конструирования? На качественном уровне предсказания очень простые. Прежде всего предсказание состоит в том, что корреляции должны наблюдаться между теми свойствами аминокислотных остатков, которые связаны с характером их взаимодействий (в трехмерном пространстве — не вдоль цепи) с обобщенными "зарядами". Более специально, если вспомнить, например, модель отпечатка, то можно понять, что те мономеры, которые притягиваются в пространстве, имеют повышенную (по сравнению со средней) вероятность оказаться поблизости друг от друга на цепи. Верно и обратное: если мономеры отталкиваются в пространстве, то они имеют пониженную вероятность быть рядом друг с другом на цепи. Эти предсказания кажутся поистине тривиальными: притягивающиеся имеют повышенную вероятность быть соседями, отталкивающиеся — пониженную. На самом деле предсказание абсолютно нетривиальное. В самом деле, речь идет о биологических последовательностях, сформированных в ходе эволюции и подвергавшихся давлению таких далеких от физики факторов, как борьба видов за существование и пр. Предсказание же состоит в том, что вопреки всему этому последовательности выглядят так, будто они сформированы в результате простого *физического* процесса.

В следующем разделе мы перейдем к проверке этого предсказания.

6.1.3. Последовательности белков: случайны ли они? У вопроса о случайности белковых последовательностей

¹² Кстати, такого опыта с реальной обезьяной, насколько известно, никто не делал. Между тем есть основания думать, что обезьяна-то как раз будет производить не случайную, а фрактальную последовательность.

длинная история. Поначалу верилось, что белки идеально приспособлены к выполнению соответствующих функций и потому, видимо, каждый белок — уникальное и незаменимое творение природы. Вместе с тем, было понятно, что ни времени, ни материала во Вселенной не могло хватить на выбор уникального наилучшего варианта. Это обстоятельство столь элементарно, что объясняется даже в популярной литературе [140]. Тем не менее потребовалось время, чтобы четко сказать: раз оптимизация не могла быть полной, значит, она и не была полной, и белки не так уж незаменимы, а, наоборот, ближе к случайным. Этот взгляд нашел выражение в концепции белка как "слегка отредактированного" ¹³ случайного сополимера", где "редактирование" касается в основном окрестности активного центра [11] (см. также по этому поводу интересную философскую книгу [141]). Это заключение было в согласии не только с общими воззрениями нейтралистской теории эволюции, но и с простыми статистическими тестами, которые подтверждали близкий к случайному характер последовательностей. Впрочем, определить случайность или неслучайность последовательности, вообще говоря, очень нелегко [142, 143]. Единственный надежный способ состоит в том, чтобы знать, какого характера отклонения от случайности можно ожидать. Именно это мы теперь знаем исходя из модели конструирования последовательностей.

Итак, вопрос, который мы рассмотрим в следующем разделе, такой: выглядят ли случайными последовательности белков в своей совокупности, или они несут на себе отпечаток пройденного ими этапа эволюционного конструирования?

6.2. Представление случайными блужданиями для изучения статистики последовательностей

Очень плодотворная техника изучения последовательностей состоит в том, чтобы отображать их на траектории соответствующих случайных блужданий. Насколько известно автору, эта идея впервые была предложена И.М. Лифшицем в контексте теории перехода спираль — клубок в ДНК [144]. Для последовательностей ДНК реализация идеи проще (потому что алфавит состоит из четырех букв, а не из двадцати, а последовательности гораздо длиннее, что обеспечивает лучшее усреднение). Поэтому мы сначала объясним технику на (интересном) примере ДНК.

6.2.1. Техника случайных блужданий для последовательностей ДНК. Для отображения последовательности ДНК на траекторию одномерного случайного блуждания используется следующая процедура. Пусть мы делаем шаги единичной длины направо и налево. В момент времени t мы шагаем направо, если мономер номер t в нашей ДНК — пурин (Pu), и шагаем налево, если он — пиримидин (Py); другими словами, определим $\xi_t = -1$, если мономер t — аденин (A) или гуанин (G), и $\xi_t = +1$, если t — цитозин (C) или тимин (T). Получается траектория ¹⁴; расстояние между ее положениями в

моменты s и s' равно

$$x(s, s') = \sum_{t=s}^{s'} \xi_t, \quad (107)$$

где $\xi_t = \pm 1$ — определенные выше "вклады мономеров". Полезно рассмотреть функцию

$$F(l) = \left[\langle x^2(s, s+l) \rangle_s - \langle x(s, s+l) \rangle_s^2 \right]^{1/2}, \quad (108)$$

где $\langle \dots \rangle_s$ означает усреднение вдоль всей ДНК, $1 \leq s \leq L$ — так называемое "усреднение по скользящему окошку". Разумеется, $F(l)$ — просто среднеквадратичное различие между количествами пуринов и пиримидинов в участке ДНК длиной l .

Много *разных* четырехбуквенных последовательностей ДНК соответствуют одному и тому же в пуриново-пиримидиновом представлении. Взаимно однозначно можно отобразить последовательность ДНК на двумерную траекторию, или траекторию на комплексной плоскости: например, шаги направо, вниз, налево и вверх, т.е. $\xi_t = +1, -i, -1, +i$ могут отвечать T, G, A, C, соответственно.

В работе [145] было сделано сильное утверждение, что

$$F(l) \sim l^z, \quad (109)$$

причем были найдены два типа последовательностей: у одного типа $z \approx 0,5$, у другого $z = 0,6-0,7$. Это утверждение, если оно верно, переносит всю проблему последовательностей ДНК, их структуры, функции и эволюции на почву фракталов.

6.2.2. Техника броуновского моста для последовательностей белков¹⁵. Поскольку в белках 20 сортов мономеров, их последовательности можно отображать на траектории очень многими способами. Другими словами, можно по-разному определять вклады ξ_i для мономеров химической природы i . Другая сторона проблемы в том, что белки — мономеры сравнительно короткие, усреднение со скользящим окошком невозможно — окошко некуда скользить. Надо, однако, помнить, что мы ищем отклонение от случайностей не в одной какой-то последовательности, а в ансамбле последовательностей разных белков, поэтому усреднять нужно не вдоль одной последовательности, а по разным белкам из базы данных [146]. Так вопрос был поставлен в работе [14].

Были исследованы три отображения, связанные с основными типами объемных взаимодействий в белках — гидрофобными (A), водородными связями (B) и кулоновскими (C). Шаги соответствующих блужданий для каждого отображения определялись так:

A. $\xi_i = +1$, если мономер номер i сильно гидрофобный (Lys, Arg, His, Asp, Glu) или $\xi_i = -1$ во всех других случаях.

B. ξ_i может быть $+1$ или -1 для мономеров способных (Asn, Gln, Ser, Thr, Trp, Tyr) или не способных (все остальные) образовывать водородные связи.

C. ξ_i может быть $+1, -1$ или 0 для положительно (Lys, Arg, His) или отрицательно (Asp, Glu) заряженных или для нейтральных (остальные) мономеров i .

Чтобы искать корреляции, сравнивая, как выглядят траектории, необходимо исключить зависимость от

¹³ "Поэт старается напоить читателя из чистого родника поэзии, но не может этого сделать, пока там не выкупается редактор" (М. Светлов).

¹⁴ Заметим, что для двух комплементарных нитей величины ξ_t отличаются знаком, а, соответственно, $F(l)$ совпадают.

¹⁵ Материал этого раздела основан на работе [14].

общей длины белка, его общего состава и среднего шага блуждания. Это делается путем следующих переопределений. Взяв белок, мы декодируем его в виде последовательности $\{\xi_1, \xi_2, \dots, \xi_L\}$ согласно одному из вариантов А, В или С и затем строим траекторию

$$x(l) = \sum_{i=1}^l \xi_i. \quad (110)$$

У траектории (110) может быть сильный дрейф (за счет состава белка), тогда главный член в $x(l)$ будет линейным по l . Раз мы не интересуемся составом, то дрейфовый член вычитаем:

$$y(l) = x(l) - \frac{l}{L} x(L), \quad (111)$$

где L — общее число звеньев в рассматриваемом белке. Как видно, y -траектория возвращается в начало координат в конце и поэтому называется "броуновским мостом".

Чтобы собрать данные в сравнимой форме, перенормируем траекторию еще раз, с дисперсией распределения ξ :

$$z^2(\lambda) = \frac{y^2}{L(\xi - \bar{\xi})^2}, \quad (112)$$

где $\overline{(\dots)}$ означает среднее по последовательности данного белка, т.е.

$$\bar{\xi} = \frac{1}{L} \sum_{i=1}^L \xi_i, \quad (113)$$

и переопределяем координату вдоль цепи (l) с учетом общей длины белка (L) согласно $\lambda = l/L$, где $0 \leq \lambda \leq 1$.

С траекториями $z^2(\lambda)$ уже можно проделать усреднение по ансамблю белков:

$$r(\lambda) = \langle z^2(\lambda) \rangle_{\text{ensemble}}. \quad (114)$$

Собирая вместе уравнения (110)–(114), получаем

$$r(\lambda) \equiv \left\langle \left(\sum_{i=0}^{[Lp]} \frac{\Delta \xi_i^{(p)}}{\sigma^{(p)}} \right)^2 \right\rangle_p, \quad (115)$$

где индекс p нумерует белки в базе данных, $\langle \dots \rangle_p$ означает усреднение по множеству белков, $[\dots]$ — взятие ближайшего большего целого числа, и L_p есть число аминокислот в белке p .

Как же судить о случайности или неслучайности последовательностей, глядя на их траекторные представления? Чисто случайная траектория должна уходить от начала координат на расстояние порядка $\sigma\sqrt{L}$ и для нее легко найти

$$r_{\text{rand}}(\lambda) = [\lambda^{-1} + (1 - \lambda)^{-1}]^{-1}. \quad (116)$$

Чтобы уйти дальше, необходимо упорно повторять шаги в одном и том же направлении до самого момента $\lambda = 1/2$, когда нужно сменить направление и теперь сохранять уже это противоположное направление, так как в итоге нужно ведь вернуться в начало координат. Значит, максимально возможное расстояние порядка $\sigma L/2$. Аналогично, чтобы держаться все время поближе

к началу координат, нужно компенсировать каждый шаг в одном направлении немедленным шагом в противоположном. Из сказанного вытекает, что последовательности с положительными, персистентными (ферромагнитными, если использовать физический лексикон) корреляциями будут характеризоваться броуновскими мостами выше случайного, а последовательности с альтернирующими (антиферромагнитными) корреляциями — ниже.

Реальные траектории $r_A(\lambda)$, $r_B(\lambda)$, $r_C(\lambda)$ для множества глобулярных белков прокариотических организмов представлены на рис. 15, где также показана и траектория $r_{\text{rand}}(\lambda)$ (116) для чисто случайного блуждания. Как видно, мосты $r_A(\lambda)$ и $r_B(\lambda)$ заметно выше $r_{\text{rand}}(\lambda)$, что говорит о персистентных корреляциях как в распределении гидрофобности, так и в способности образовывать водородные связи. С другой стороны, мост $r_C(\lambda)$, совершенно очевидно, ниже моста $r_{\text{rand}}(\lambda)$, что свидетельствует об альтернирующих корреляциях в распределении зарядов. Эти заключения из рис. 15 полностью соответствуют предсказаниям модели конструирования последовательностей.

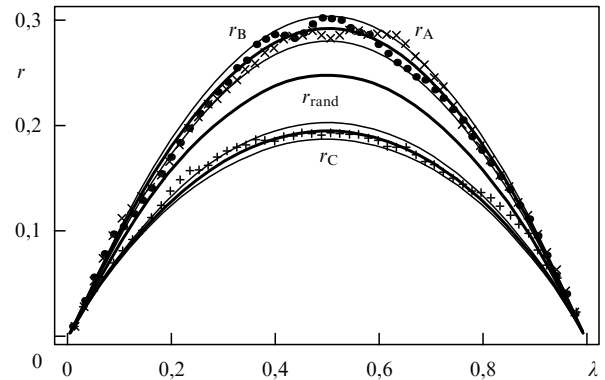


Рис. 15. Броуновские мосты для гидрофильного ($\times$), отвечающего водородным связям ($\bullet$) и кулоновского ($+$) представлений последовательностей тех прокариотических белков, которые имеют каталитическую активность и, значит, глобулярную структуру. Средняя линия отвечает полностью случайной последовательности r_{rand} (116).

Интересно, что данные для r_A , r_B и r_C очень хорошо аппроксимируются феноменологической формулой

$$r(\lambda) = \frac{L_0^{2\alpha-1}}{\lambda^{-2\alpha} + (1 - \lambda)^{-2\alpha}}, \quad (117)$$

где мы выбрали $L_0 = 110$ и получили $\alpha_A = 0,520 \pm 0,05$, $\alpha_B = 0,520 \pm 0,05$, $\alpha_C = 0,470 \pm 0,05$. Коридоры, показанные на рис. 16 (и заключающие внутри себя все экспериментальные точки) показывают, что величины α действительно определяются со столь малыми ошибками. По своей формально-алгебраической природе величины α в формуле (117) похожи на фрактальный показатель в уравнении (109). Тем не менее фрактальная интерпретация представляется невозможной, точнее — бесплодной, так как формула (117) выражает усреднение по большому числу разных белков. Тем не менее близость всех трех показателей α к 0,5 говорит о том, что последовательности белков лишь немного отличаются от случайных, они были отредактированы действительно только "слегка". Неудивительно, что отклонения от случайности было

нелегко обнаружить. Тем более важно, что они несут предсказанный характер.

В работах [14, 147] приведены также броуновские мосты для белков из разных других групп организмов. Хотя общий характер корреляций остается неизменным, есть много других обстоятельств, возможно, представляющих биологический интерес. Некоторые дальнейшие шаги в этом направлении рассмотрены в [148].

6.3. Конструирование последовательностей и модели эволюции

Одно из интригующих приложений модели конструирования последовательностей — в вопросе о возможных сценариях ранней предбиологической эволюции. Уже довольно давно было ясно, что понимание сворачивания белков — важный ключ к проблеме предбиологической эволюции, т.е. к вопросу о том, как в первичном бульоне могли возникнуть первые "биополимеры" [11, 149, 150]. Не вступая глубоко в эту интересную и волнующую область, мы здесь останемся на уровне простейших моделей. В этих рамках, как обсуждалось в работе [151], интересно посмотреть, какие последовательности могли бы быть отобраны, если к ним предъявить два *конфликтующих* требования компактности и растворимости. Требования эти действительно противоположно направленные: наилучший способ обеспечить растворимость — это иметь поменьше гидрофобных мономеров, но тогда полимер не будет компактным, и наоборот. Эта идея была материализована в компьютерном эксперименте [152]. Для этого была взята ГП-модель (с двумя видами мономеров: гидрофобными и полярными) и в ней осуществлялись случайные мутации с условием, чтобы процент гидрофобных мономеров не превосходил некоторого заранее выбранного предела (это условие моделировало требование растворимости). Для каждой мутации методом Монте-Карло изучалась динамика полимера, и, если она быстрее вела к более компактному состоянию, чем до мутации, то такая мутация принималась, или мутация отвергалась в противном случае. Результатом такой эволюции неизменно оказывались последовательности с хорошо оптимизированной (т.е. низкой) энергией основного состояния. Конечно, нужны дальнейшие усилия, чтобы проверить, действительно ли данная простейшая модель правильно представляет то, что могло бы происходить в первичном бульоне.

Упомянем также еще один аспект, в котором конструирование последовательностей помогает понять эволюционные вопросы. Допустим, мы конструируем последовательности для нативного состояния *. Хотя мы говорили больше о решеточных моделях, но можно, конечно, использовать и просто реальные конформации. Так или иначе, на решетке или без мы можем, конечно, приготовить много разных последовательностей для одной и той же конформации *. Имея такой ансамбль последовательностей, можно теперь посмотреть, какие мономеры меняются, а какие остаются всегда или почти всегда неизменными. Можно далее сравнить это с данными о вариабельности аминокислот в разных известных мутантах одного и того же белка. Такая программа была недавно осуществлена [153], и результаты показывают многообещающее сходство экспериментального и модельного профилей вариабельности.

7. Заключение

В заключение вернемся к вопросу о соотношении между рассмотренными здесь простыми моделями и биополимерами. Прежде всего повторим, что эти модели, в принципе, не рассматривают никаких мелкомасштабных деталей, как бы они ни были важны [104, 105]: предполагается, что все необходимые перенормировки уже сделаны, и (квази)мономер — это уже точки. В частности, решеточный полимер нужно понимать как своего рода перенормированный белок. Хотя и неизвестно, как это технически осуществить, но и вторичные структуры должны быть включены в перенормировки, так что конформацию решеточного полимера нужно понимать как огрубленное представление о расположении элементов вторичной структуры. С этой точки зрения 27-мер [154] или, может быть, даже 18-мер, видимо, могут считаться адекватными для однодоменных белков. Хотя типичный домен имеет, скажем, от 50 до 200 остатков, но не надо думать, что моделью белка из 200 аминокислот должен быть обязательно полимер из 200 мономеров. Больше того, *de novo* приготовленный "белок" из четырех α -спиралей [120] похож даже на решеточный 12-мер ($2 \times 2 \times 3$). В то же время решеточные 36- и 48-меры, возможно, должны уже рассматриваться как модели многодоменных белков [155]. Это можно сопоставить с фактом, что 36-мер — кратчайший из решеточных полимеров, для которых в компактном состоянии возможны квазиузлы. По связанной причине 36-мер также самый короткий, для которого глобула может состоять из двух хорошо сегрегированных доменов по 18 мономеров, заполняющих соответствующие области $2 \times 3 \times 3$. Таким образом, термодинамический предел — отнюдь не обязательно самый интересный режим.

Безусловно, простые модели не могут объяснить многого, а для некоторых вопросов их применимость не ясна. Например, известно, что белок функционирует как машина [156] и его функционирование требует, чтобы у него была вполне определенная конструкция [157]. Возникает вопрос: если мы сконструируем последовательность полимера так, как это описано выше, будет ли этого достаточно, чтобы полимер уже имел конструкцию необходимую для функционирования, или потребуется что-то еще? Вопрос интересный; ответ (автору) пока не известен. Тем не менее кажется, что рассмотренные модели схватывают правильно какую-то важную часть биополимерной специфики, по крайней мере, в том, что касается сворачивания белков. Например, в эксперименте (хотя и компьютерном) установлено, что решеточный 100-мер с подходящим образом сконструированной последовательностью способен к ренатурации [135] в той же степени, как и реальные белки. Это само по себе не значит, что мы ответили на парадокс Левинталя; сворачивание решеточного полимера столь же парадоксально, как и реального, так как компьютер не может успеть пребрать все варианты, и как же тогда находится единственно правильный? Трудно себе представить, чтобы разрешение парадокса было разным для моделей и для белков. Поэтому модели наверняка очень интересны.

Разрешен ли парадокс Левинталя на уровне моделей? Есть мнение, что да, но автор этих строк полагает, что мнение это слишком оптимистичное. Есть все основания

думать, что идея конструирования последовательностей — шаг в нужном направлении. Обнаружение предсказанных корреляций в белковых последовательностях свидетельствует в пользу такого мнения. Тем не менее процедура конструирования, как мы видим ее теперь, оптимизирует только термодинамические свойства полимера. Компьютерные модели показывают, что при этом и кинетика оказывается в какой-то мере оптимизированной, но полного понимания этого факта (у автора этих строк) пока нет. Нужно думать, что интересные новые работы не за горами. Можно высказать предположение, что более явный учет вторичной структуры и/или топологических ограничений будут для понимания важны.

Благодарности

Данный обзор в значительной мере основан на работах, которые автор выполнил в соавторстве с А.М. Гутиным, В.С. Панде, И. Рабином, Т. Танакой и Е.И. Шахновичем. Автор глубоко признателен всем им за энтузиазм и знания, которые они вкладывали в эти работы.

Список литературы

1. Лифшиц И М *ЖЭТФ* **55** 2408 (1968); см. также Лифшиц И М *Избранные научные труды. Электронная теория металлов. Полимеры и биополимеры* (М.: Наука, 1994)
2. Doi M, Edwards S F *The Theory of Polymer Dynamics* (Oxford: Oxford University Press, 1987)
3. Де Жен П Ж *Концепция скейлинга в физике полимеров* (М.: Мир, 1982)
4. Лифшиц И М, Гредескул С А, Пастур Л А *Введение в теорию неупорядоченных систем* (М.: Наука, 1982)
5. Anfinsen C *Science* **181** 223 (1973)
6. Derrida B *Phys. Rev. Lett.* **45** 79 (1980)
7. Bryngelson J D, Wolynes P G *Proc. Nat. Acad. Sci. USA* **84** 7524 (1987)
8. Wolynes P G, in *Spin Glass Ideas in Biology* (Ed. D Stein) (Singapore: World Scientific, 1991) p. 1
9. Shakhnovich E, Gutin A *Biophys. Chem.* **34** (3) 187 (1989)
10. Pande V S, Grosberg A Yu, Tanaka T *Biophys. J.* (in press)
11. Pitsyn O B, Volkenstein M V *J. Biomolec. Struct. and Dynamics* **4** 137 (1986)
12. Shakhnovich E, Gutin A *Proc. Nat. Acad. Sci. USA* **90** 7195 (1993)
13. Pande V S, Grosberg A Yu, Tanaka T *Proc. Nat. Acad. Sci. USA* **91** 12976 (1994)
14. Pande V S, Grosberg A Yu, Tanaka T *Proc. Nat. Acad. Sci. USA* **91** 12972 (1994)
15. Panyukov S, Rabin Y *Phys. Rep.* **269** 1 (1996)
16. Bates F, Fredrickson G H *Ann. Rev. Phys. Chem.* **41** 525 (1990)
17. Erukhimovich I, Khokhlov A *Polymer Sci.* **35** 1522 (1993)
18. Binder K *Adv. Polymer Sci.* **112** 181 (1994)
19. Гросберг А Ю, Каганова Е М, Молчанов С А *Биофизика* **29** 30 (1984)
20. Grosberg A Yu *J. Stat. Phys.* **38** 149 (1985)
21. Гросберг А Ю, Каганова Е М *ДАН СССР* **294** 838 (1987)
22. Гросберг А Ю, Шахнович Е И *ЖЭТФ* **91** 837 (1986)
23. Гросберг А Ю, Шахнович Е И *ЖЭТФ* **91** 2159 (1986)
24. Гросберг А Ю, Шахнович Е И *Биофизика* **31** 1045 (1986)
25. Joanny J-F *J. de Phys II* **4** 1281 (1994)
26. Garel T et al. *Europhys. Lett.* **8** 9 (1989)
27. Grosberg A Yu, Izrailev S F, Nechaev S K *Phys. Rev. E* **50** 1912 (1994)
28. Sommer J-U, Daoud M *Europhys. Lett.* **32** 407 (1995)
29. De Gennes P G "Chrysanthemums": *Weak Micellisation* Preprint (1995)
30. Веденов А А, Дыхне А М, Франк-Каменецкий М Д *УФН* **105** 479 (1971)
31. Azbel M Ya *Biopolymers* **19** 61, 81, 95, 1311 (1980); Azbel M et al. *Biopolymers* **21** 1687 (1982)
32. Гросберг А Ю, Хохлов А Р *Статистическая физика макромолекул* (М.: Наука, 1989)
33. Trifonov E N, Tan R K Z, Harvey S C, in *Structure and Expression* (Eds M H Sarma, R H Sarma) (Schenectady, NY: Adenin Press, 1988) p. 243
34. Bednar J et al. *J. Mol. Biol.* **254** 579 (1995)
35. Higgs P G, Joanny J-F *J. Chem. Phys.* **94** 1543 (1991)
36. Kantor Y, Li H, Kardar M *Phys. Rev. Lett.* **69** 61 (1992); Kantor Y, Kardar M *Europhys. Lett.* **27** 643 (1994); Kantor Y, Kardar M, Li H *Phys. Rev. E* **49** 1383 (1994); Kantor Y, Kardar M *Phys. Rev. E* **51** 1299 (1995); Kantor Y, Kardar M *Phys. Rev. E* **52** 835 (1995); Ertas D, Kantor Y *Phys. Rev. E* (1996) (in press)
37. Wittmer J, Johner A, Joanny J F *Europhys. Lett.* **24** 263 (1993)
38. Dobrynin A V, Rubinstein M J. *de Phys. II* **5** 677 (1995)
39. Gutin A M, Shakhnovich E I *Phys. Rev. E* **50** R3322 (1994)
40. Гросберг А Ю *Биофизика* **29** 569 (1984)
41. Garel T, Leibler L, Orland H *J. de Phys. II* **4** 2139 (1994)
42. Marko J F, Siggia E D *Macromolecules* **27** 981 (1994); *Science* **265** 506 (1994); *Phys. Rev. E* **52** 2912 (1995)
43. Франк-Каменецкий М Д *Самая главная молекула* (М.: Наука, 1989)
44. Zimm B H, Stockmayer W H *J. Chem. Phys.* **17** 1301 (1949)
45. Gutin A M, Grosberg A Yu, Shakhnovich E I *Macromolecules* **26** 1293 (1993)
46. Isaacson J, Lubensky T C *J. de Phys. Lett.* **41** L469 (1980)
47. Daoud M et al. *Macromolecules* **16** 1833 (1983)
48. Parisi G, Sourlas N *Phys. Rev. Lett.* **46** 871 (1981)
49. Daoud M, Joanny J-F *J. de Phys.* **42** 1359 (1981)
50. Gaunt D, Flesia S J. *Phys. A* **24** 3655 (1991); Janse van Rensburg E, Madras N J. *Phys. A* **25** 303 (1992)
51. Shi-Min Cui, Zheng Yu Chen *Phys. Rev. E* **52** 5084 (1995)
52. Shi-Min Cui, Zheng Yu Chen *Phys. Rev. E* **53** 6238 (1996)
53. Grosberg A Yu, Gutin A M, Shakhnovich E I *Macromolecules* **28** 3718 (1995)
54. Kelvin, Lord *Trans. Roy. Soc. of Edinburg* **25** 217 (1868)
55. Wu F Y *Rev. Mod. Phys.* **64** 1099 (1992); **65** 577 (1993)
56. Kauffman L H *Knots and Physics* (Singapore: World Scientific, 1991)
57. Атья М *Физика и геометрия узлов* (М.: Мир, 1995)
58. Moffatt H K *Nature* (London) **347** 367 (1990)
59. Frisch H L, Wasserman E J. *Am. Chem. Soc.* **83** 3789 (1961)
60. Delbrück M *Mathematical Problems in the Biological Sciences, Proc. of Symposium in Applied Mathematics* (American Mathematical Society, Providence RI) **14** 55 (1962)
61. Liang C, Mislow K J. *Am. Chem. Soc.* **116** 11189 (1994); **116** 3588 (1994); *J. Math. Chem.* **15** 245 (1994)
62. Sumners D W, Whittington S G J. *Phys. A* **21** 1689 (1988)
63. Pippenger N *Discrete. Appl. Math.* **25** 273 (1989)
64. Nechaev S K *J. Phys. A* **21** 3659 (1988)
65. Nechaev S *Statistics of Knots and Entangled Random Walks* (Singapore: World Scientific, 1996)
66. Nechaev S K, Grosberg A Yu, Vershik A M *J. Phys. A* **29** 2411 (1996)
67. Edwards S F *Proc. Phys. Soc.* **91** 513 (1967)
68. Prager S, Frisch H L *J. Chem. Phys.* **46** 1475 (1967)
69. Khokhlov A R, Nechaev S K *Phys. Lett. A* **112** 156 (1985)
70. Вологодский А В, Лукашин А В, Франк-Каменецкий М Д *ЖЭТФ* **67** 1875 (1974)
71. Koniaris K, Muthukumar M *Phys. Rev. Lett.* **66** 2211 (1991)
72. Deguchi T, Tsurusaki K *J. of Knot Theory and its Ramifications* **3** 321 (1994)
73. Grosberg A Yu, Feigel A, Rabin Y *Phys. Rev. E* (in press)
74. Grosberg A Yu, Nechaev S K *Adv. Polymer Sci.* **106** 1 (1993)
75. Drossel B, Kardar M *Phys. Rev. E* **53** 5861 (1996)
76. Вологодский А В, Франк-Каменецкий М Д *УФН* **134** 641 (1981); Vologodskii A V *Topology and Physics of Circular DNA* (Boca Raton: CRC Press, 1992)
77. Tait P G *Scientific Papers* **1** 273 (1898)
78. Rolfsen D *Knots and Links* (Berkeley: Publish or Perish Press, 1976)
79. Atiyah M *Rev. Mod. Phys.* **67** 977 (1995)
80. Кроуэл Р, Фокс Р *Введение в теорию узлов* (М.: Мир, 1965)
81. Jones V F R *Bull. Am. Math. Soc.* **12** 103 (1985)
82. Katritch V et al. *Nature* (1996) (in press)
83. Klenin K V et al. *J. Biomolec. Struct. and Dynamics* **5** 1173 (1988)
84. Liu L F, Depew R E, Wang J C J. *Molec. Biology* **106** 439 (1976)
85. Rybenkov V V, Cozzarelli N R, Vologodskii A V *Proc. Nat. Acad. Sci. USA* **90** 5307 (1993)
86. Shaw S Y, Wang J C *Science* **260** 533 (1993)
87. Tesi M C et al. *J. Phys. A* **27** 347 (1994)

88. Grosberg A Yu, Nechaev S K *J. Phys. A* **25** 4659 (1992); *Europhys. Lett.* **20** 613 (1992)
89. Grosberg A Yu, Nechaev S K, Shakhnovich E I *J. de Phys.* **49** 2095 (1988)
90. Jianpeng Ma, Straub J E, Shakhnovich E I *J. Chem. Phys.* **103** 2615 (1995)
91. Timoshenko E et al. *J. Chem. Phys.* (in press)
92. Chu B, Ying Q, Grosberg A Yu *Macromolecules* **28** 180 (1995)
93. Chu B, Ying Q *Macromolecules* **29** 1824 (1996)
94. Grosberg A Yu, Kuznetsov D V *Macromolecules* **26** 4249 (1993)
95. Ganazzoli F, Ferla R La, Allegra G *Macromolecules* **28** 5285 (1995)
96. Grosberg A Yu et al. *Europhys. Lett.* **23** 373 (1993)
97. Sikorav J-L, Jannink G *C.R. Acad. Sci. Paris* **316** 751 (1993)
98. Sikorav J-L, Church G M *J. Mol. Biol.* **222** 1085 (1991)
99. Janse van Rensburg E J, Whittington S G *J. Phys. A* **24** 3935 (1991)
100. Cates M E, Deutsch J M *J. de Phys.* **47** 2121 (1986)
101. Semenov A N *J. de Phys.* **49** 175 (1988)
102. De Gennes P G *Macromolecules* **17** 703 (1984); reprinted in *Simple Views on Condensed Matter* (Singapore: World Scientific, 1992) p. 215
103. Rabin Y, Grosberg A Yu, Tanaka T *Europhys. Lett.* **32** 505 (1995)
104. Frauenfelder H, Wolynes P G *Physics Today* **47** (2) 58 (1994)
105. Goldanskii V I *Tunneling Phenomena in Chemical Physics* (New York: Gordon and Breach Sci. Publ., 1989)
106. Finkelstein A V, Gutin A M, Badretdinov A Ya *FEBS Lett.* **325** 23 (1993)
107. Finkelstein A V, Badretdinov A Ya, Gutin A M *Proteins* **23** 142 (1995)
108. Finkelstein A V, Gutin A M, Badretdinov A Ya *Proteins* **23** 151 (1995)
109. Miyazawa S, Jernigan R *Macromolecules* **18** 534 (1985)
110. Godzik A, Kolinski A, Skolnick J *Protein Sci. (BNW)* **10** 2107 (1995)
111. Goldenfeld N *Lectures on phase transitions and the renormalization group* (Reading, Mass.: Addison-Wesley Press, 1992)
112. Lau K F, Dill K A *Macromolecules* **22** 3986 (1989)
113. Obukhov S P *Phys. Rev. A* **42** 2015 (1990)
114. Garel T, Leibler L, Orland H *J. de Phys. II* **4** 2139 (1994)
115. Li H et al. *Science* **273** 666 (1996)
116. 27,8° — правильное значение угла электрослабого смешивания. К белкам это, конечно, не имеет абсолютно никакого отношения: автор надеется на читательское чувство юмора ...
117. Mezard M, Parisi G, Virasoro M *Spin Glass Theory and Beyond* (Singapore: World Scientific, 1987)
118. Koukiou F J *Phys. Rev. A* **26** L1207 (1993)
119. Pande V S, Grosberg A Yu, Tanaka T *Phys. Rev. Lett.* **76** 3987 (1996)
120. Hecht M H et al. *Science* **249** 884 (1990)
121. Editorial commentary *Science* **269** 1821 (1995)
122. Pande V S, Grosberg A Yu, Tanaka T *J. Chem. Phys.* (in press)
123. Pande V S, Grosberg A Yu, Tanaka T *Phys. Rev. E* **51** 3381 (1995)
124. Sfatos C D, Gutin A M, Shakhnovich E I *Phys. Rev. E* **48** 465 (1993)
125. Shakhnovich E, Gutin A J. *Chem. Phys.* **93** 5967 (1990)
126. Pande V S et al. *J. Phys. A* **27** 6231 (1994)
127. Flory P *Principles of Polymer Chemistry* (Ithaca: Cornell University Press, 1953)
128. Garel T, Orland H, Thirumalai D Preprint (1995)
129. Gutin A M, Shakhnovich E I *J. Chem. Phys.* **98** 8174 (1993)
130. Bryngelson J D *J. Chem. Phys.* **100** 6038 (1994)
131. Yue K, Dill K A *Proc. Nat. Acad. Sci. USA* **89** 4163 (1992)
132. Abkevich V, Gutin A, Shakhnovich E *J. Mol. Biol.* **252** 460 (1995)
133. Ramanathan S, Shakhnovich E I *Phys. Rev. E* **50** 1303 (1994)
134. Pande V S, Grosberg A Yu, Tanaka T *Macromolecules* **28** 2218 (1995)
135. Shakhnovich E I *Phys. Rev. Lett.* **72** 3907 (1994)
136. Derrida B *J. de Phys. Lett.* **46** L401 (1985)
137. Plotkin S S, Wang J, Wolynes P G *Phys. Rev. E* **53** 6271 (1996)
138. Pande V S, Grosberg A Yu, Tanaka T *J. Chem. Phys.* **101** 8246 (1994)
139. Pande V S, Grosberg A Yu, Tanaka T *J. Chem. Phys.* **103** 9482 (1995)
140. Гросберг А Ю, Хохлов А Р *Физика в мире полимеров* (М.: Наука, 1989)
141. Monod J *Chance and Necessity: an Essay on the Natural Philosophy of Modern Biology* (New York: Vintage Books, 1971)
142. Турыгин А Ю, Чечеткин В Р *ЖЭТФ* **106** 335 (1994)
143. Chechetkin V R, Turugin A Yu *Phys. Lett. A* **199** 75 (1995); *J. Theor. Biol.* **175** 477 (1995); **178** 205 (1996)
144. Лифшиц И М *ЖЭТФ* **65** 1100 (1973); см. также: Лифшиц И М *Избранные научные труды. Электронная теория металлов. Полимеры и биополимеры* (М.: Наука, 1994)
145. Peng C-K et al. *Nature* **356** 168 (1992)
146. Bairoch A, Boeckmann B *Nucleic Acids Res.* **20** 2019 (1992)
147. Pande V S Ph. D. Thesis (MIT, 1995)
148. Strait B J, Dewey G *Phys. Rev. E* **52** 6588 (1995)
149. Eigen M, Schuster P *The Hypercycle, a Principle of Natural Self-Organization* (Berlin, New York: Springer Verlag, 1979); Eigen M, Winkler R *Laws of the Game: How the Principles of Nature Govern Chance* (New York: Harper & Row, 1983); Eigen M, Winkler-Oswatitsch R *Steps Towards Life: a Perspective on Evolution* (Oxford New York: Oxford University Press, 1992)
150. Dyson F *Origins of Life* (Cambridge: Cambridge University Press, 1985)
151. Гросберг А Ю, Хохлов А Р, Препринт НЦБИ (Пушино, 1985)
152. Gutin A, Abkevich V, Shakhnovich E *Proc. Nat. Acad. Sci. USA* **92** 1282 (1995); Abkevich V, Gutin A, Shakhnovich E *Proc. Nat. Acad. Sci. USA* **93** 830 (1996)
153. Shakhnovich E, Abkevich V, Ptitsyn O *Nature* **379** 96 (1996)
154. Onuchic J N et al. *Proc. Nat. Acad. Sci. USA* **92** 3626 (1995)
155. Shakhnovich E *Phys. Rev. E* (in press)
156. Романовский Ю М, Степанова Н В, Чернавский Д С *Математическая биофизика* (М.: Наука, 1984)
157. Blumenfeld L A, Tikhonov A N *Biophysical Thermodynamics of Intracellular Processes* (Berlin: Springer-Verlag, 1994)

Disordered polymers

A.Yu. Grosberg

Institute for Biochemical Physics

117977 Moscow, Russia

Department of Physics and Center for Materials Science and Engineering,

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

E-mail: shura@gels.mit.edu

A single polymer macromolecule is considered with disorder types such as branches, knots, and heterogeneous sequences of chemical units. In all cases, simple theoretical approaches are employed to gain useful physical insights. For branched polymers, a simple Flory-type theory is described by means of which the difference between the universality classes for molecules with quenched and annealed branches is demonstrated. For knots, another Flory-type theory is suggested to describe the swelling and/or collapse of a quenched topology ring or the size distribution for the annealed case. To consider heteropolymers, the Random Energy Model borrowed from the spin glass theory is systematically employed. This allows a simple yet rigorous description of both the freezing transition of a random sequence globule and the use of the canonical ensemble for designing sequences with energy-optimised ground state conformation. Along with the analytical theory, computer tests for the freezing and design processes are discussed. The sequence design scheme is shown to yield a specific prediction concerning the character of correlations in protein sequences. Statistical tests confirming this prediction are described.

PACS numbers: **36.20.-r**, **64.70.-p**, **87.10.+e**, **87.10.-v**

Bibliography — 157 references

Received 3 September 1996