

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

имени М.В. Ломоносова

ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи

Шайтан Алексей Константинович

**Компьютерное моделирование и статистический анализ  
самоорганизующихся молекулярных систем  
на основе пептидов**

02.00.06 – Высокомолекулярные соединения

ДИССЕРТАЦИЯ

на соискание ученой степени

кандидата физико-математических наук

Научный руководитель

д. ф.-м. н., академик

Хохлов Алексей Ремович

Москва – 2010

# Содержание

<b>Введение</b> . . . . .	<b>5</b>
<b>Глава 1. Принципы самоорганизации белковых и пептид-полимерных структур</b> . . . . .	<b>11</b>
1.1. Пептиды и их строение . . . . .	11
1.2. Пространственная структура белков . . . . .	15
1.3. Термодинамика формирования пространственной структуры белков . . . . .	18
1.4. Амилоиды и амилоидные фибриллы . . . . .	22
1.5. “Молекулярные химеры” и принципы их построения . . . . .	22
1.6. Гибриды тиофенов и пептидов . . . . .	24
<b>Глава 2. Некоторые методы компьютерного моделирования и анализа</b> . . . . .	<b>28</b>
2.1. Методы молекулярной динамики и механики . . . . .	28
2.2. Методы расчета свободной энергии . . . . .	32
2.3. Молекулярные поверхности . . . . .	34
<b>Глава 3. Изучение гидратации аминокислот у поверхности воды</b> . . . . .	<b>36</b>
3.1. Введение . . . . .	36
3.2. Постановка задачи . . . . .	37
3.3. Методы расчёта . . . . .	37
3.4. Результаты и обсуждение . . . . .	41
3.5. Выводы к главе . . . . .	52
<b>Глава 4. Анализ распределения звеньев в глобулярных белках</b> . . . . .	<b>56</b>
4.1. Введение . . . . .	56
4.2. Постановка задачи . . . . .	57
4.3. Создание набора белков . . . . .	57
4.4. Методы расчётов . . . . .	59
4.5. Зависимость статистики от размера белков . . . . .	61
4.6. Статистический анализ доступных поверхностей аминокислот . . . . .	64
4.7. Корреляции с экспериментальными шкалами гидрофобности . . . . .	68
4.8. Факторы, влияющие на распределения аминокислот . . . . .	79
4.9. Выводы к главе . . . . .	80

<b>Глава 5. Самоорганизующиеся тиюфен-пептидные диблок олигомеры . . .</b>	<b>83</b>
5.1. Введение . . . . .	83
5.2. Постановка задачи . . . . .	86
5.3. Рациональный компьютерный дизайн моделей . . . . .	87
5.4. Моделирование однослойных фибрилл в объёме . . . . .	95
5.5. Моделирование на подложке, виртуальная АСМ . . . . .	100
5.6. Моделирование двуслойных фибрилл . . . . .	101
5.7. Упорядочение тиюфеновых сегментов . . . . .	107
5.8. Сравнение эксперимента и моделирования . . . . .	112
5.9. Выводы к главе . . . . .	114
<b>Заключение . . . . .</b>	<b>118</b>
<b>Литература . . . . .</b>	<b>121</b>
<b>Приложение А. Типы аминокислот и их боковые цепи . . . . .</b>	<b>134</b>
<b>Приложение Б. Свободная энергия сольватации . . . . .</b>	<b>136</b>
Б.1. Введение . . . . .	136
Б.2. Свободная энергия сольватации . . . . .	136
Б.3. Постулаты классической статфизики . . . . .	137
Б.4. Связь статсуммы со свободной энергией . . . . .	137
Б.5. Сольватация, рассмотрение статфизики . . . . .	139
<b>Приложение В. Приложения к главе 3 . . . . .</b>	<b>144</b>
В.1. Профиль плотности границы раздела вода-пар . . . . .	144
В.2. Теория профилей свободной энергии . . . . .	144
В.3. К вопросу вычисления энергий гидратации . . . . .	145
В.4. Энергии адсорбции . . . . .	146
<b>Приложение Г. Приложения к главе 3: профили свободной энергии . . .</b>	<b>148</b>

## Список сокращений

BAR - bennett acceptance ratio, метод расчёта свободной энергии по Беннетту

FEP - free energy perturbation (метод расчёта свободной энергии)

PDB - Protein Data Bank (база данных пространственных структур белков)

SASA - solvent accessible surface area (поверхность, доступная растворителю)

АКО - аминокислотный остаток

АСМ - атомно-силовая микроскопия

МД - молекулярная динамика, молекулярно-динамический

МСЭ - модель случайных энергий

ПДР - поверхность, доступная растворителю

ПРГ - поверхность раздела Гиббса

ПСР - профиль средней силы

ПЭМ - просвечивающая электронная микроскопия

ПЭГ - полиэтиленгликоль

ТИ - метод термодинамического интегрирования

ЯМР - ядерный магнитный резонанс

# Введение

Диссертационная работа посвящена изучению самоорганизующихся молекулярных структур на основе поли- и олигопептидов. Проведены исследования молекул как природного (белки), так и синтетического (гибриды синтетических полимеров и пептидов) происхождения методами компьютерного моделирования и статистического анализа (анализ пространственных структур белков). Отдельная глава диссертации посвящена изучению процессов гидратации аминокислот у поверхности воды, расчёту термодинамических параметров гидратации и адсорбции.

## Актуальность работы

Актуальность выбранного направления исследований связана с повышенным вниманием мирового научного сообщества в последнее время к созданию новых материалов и соединений на основе биоинспирированных подходов, использованию принципов самоорганизации биологических макромолекул при создании умных полимеров, при разработке новых веществ и материалов. Для достижения этой цели существует два принципиальных подхода: (I) создание биомиметических синтетических полимеров путём дизайна их последовательностей и подбора свойств звеньев на основе принципов, заимствованных у природных макромолекул, и (II) создание гибридов синтетических полимеров и биополимеров, позволяющее заимствовать принципы организации и свойства биологических молекул вместе с “материальным носителем”.

Одним из серьёзных успехов биомиметического подхода явилось создание “белково-подобных” сополимеров – линейных макромолекул со специальной статистикой последовательности, способных формировать водорастворимые глобулы благодаря концентрации гидрофобных звеньев в ядре глобулы, а гидрофильных на поверхности. Создание и изучение таких сополимеров из различных типов звеньев требует описания свойств звеньев в терминах их сродства к водному окружению, поверхности, неполярному окружению. Особый интерес представляет влияние свойств звеньев на их положение в структуре полимерных глобул. Изучению обозначенных вопросов на примере природных белков и их звеньев (аминокислот) посвящены главы 3 и 4 диссертационной работы.

Для создания гибридов синтетических полимеров и биомолекул особый интерес представляют пептидные последовательности, способные к образованию бета-листов. Такие последовательности при определённых условиях могут приводить к межмолекулярной агрегации с образованием фибриллярных структур в нанометровом диапазоне подобных амилоидным фибриллам, встречающимся при ряде заболеваний (например, болезни Альц-

геймера). Сопряжение самоорганизационных свойств таких последовательностей с полупроводящими, оптическими, флюоресцентными свойствами некоторых синтетических полимеров путём их ковалентного соединения открывает большие перспективы во многих областях: микроэлектронике, сенсорных и биосенсорных технологиях и т.д. Такие гибридные соединения уже синтезированы экспериментально, однако не до конца ясными остаются вопросы, связанные со структурой получаемых агрегатов, а также представления о рациональном дизайне таких структур с заданными свойствами. Исследованию этого вопроса методами молекулярной динамики на конкретном примере посвящена глава 5 диссертационной работы.

Несмотря на бурный прогресс экспериментальных методов, методы компьютерного моделирования и статистического анализа лишь увеличивают свою актуальность. В биологии это связано с необходимостью обработки огромных массивов экспериментальных данных в автоматическом режиме: последовательностей нуклеотидов, аминокислот, пространственных структур белков и т.д. Методы компьютерного моделирования молекулярных систем из первых принципов (*ab initio*) благодаря развитию суперкомпьютеров и технологий параллельных вычислений, позволяют моделировать системы всё большего масштаба и находить новые точки соприкосновения между экспериментальными результатами и моделированием.

### **Цель диссертационной работы**

Целью диссертационной работы является изучение самоорганизующихся молекулярных структур на основе поли- и олигопептидов методами молекулярной динамики, а также статистического анализа известных пространственных структур полипептидов.

Конкретные **задачи** работы включают в себя:

- Изучение процессов гидратации аминокислот на поверхности воды.
- Разработка метода расчёта свободной энергии адсорбции молекул на границе вода/воздух на основе атомистических моделей веществ. Расчёт энергий адсорбции и гидратации для боковых цепей аминокислот.
- Изучение распределения различных типов аминокислотных остатков между поверхностью и ядром в глобулярных водорастворимых белках на основе анализа поверхности доступной растворителю для экспериментально известных пространственных структур белков.
- Расчёт статистических аналогов свободной энергии переноса аминокислотных остат-

ков из ядра глобулы на поверхность и анализ корреляций этих величин с экспериментально известными распределениями аминокислот и их аналогов между фазами различной полярности.

- Изучение процессов самоорганизации тиофен-пептидных молекул-гибридов в нановолокна. Создание молекулярных моделей нанофибрилл.
- Анализ моделей с применением метода молекулярной динамики и выявление наиболее адекватной модели, соответствующей экспериментальным данным.

### **Научная новизна**

1. Предложена методика расчёта свободной энергии адсорбции для молекул на границе вода-воздух на основе атомистических моделей.
2. Впервые в молекулярном моделировании рассчитаны значения энергии адсорбции для боковых цепей аминокислот на границе вода-воздух, изучены профили свободной энергии и термодинамика гидратации вблизи водной поверхности.
3. Создана эволюционно-независимая выборка полноатомных пространственных структур компактных глобулярных белков из 8000 экземпляров с высоким пространственным разрешением.
4. Впервые построены профили статистической свободной энергии аминокислот в глобулярных белках в зависимости от экспонированности аминокислоты растворителю, показано, что корреляция между статистическими свободными энергиями и коэффициентами распределения аминокислот между различными парами растворителей зависят от критерия разделения аминокислот на “ядерные” и “поверхностные” при подсчёте статистических энергий.
5. Впервые предложены молекулярные модели однослойных и многослойных нановолокон из гибридных молекул, состоящих из тетратиофена и олигопептидной последовательности на основе треонина и валина. Проведена интерпретация экспериментальных данных в рамках предложенных моделей.
6. Предложена методика поэтапного конструирования молекулярных моделей таких гибридных волокон на основе кросс-бета-листовой структуры с использованием метода молекулярной динамики.

## **Практическая значимость**

Разработанные методы оценки свободной энергии адсорбции молекул на границу жидкость/газ или жидкость/жидкость позволяют проводить количественные оценки соответствующих величин *in silico*, что может оказаться весьма полезным для современных научных и технических задач. Кроме того, данный метод позволит проводить дополнительную калибровку и настройку атом-атомных потенциалов (силовых полей) в методах молекулярной механики и динамики для более точного моделирования явлений на границе раздела фаз. В частности, численные результаты данной работы могут применяться для калибровки силовых полей при моделировании белков.

В ходе работы вычислены корреляции между статистическими распределениями аминокислот в белковой глобуле и экспериментальными коэффициентами распределения аминокислот и их боковых цепей между различными растворителями, а также построены профили встречаемости аминокислот в зависимости от их экспонирования к растворителю. Данные результаты могут использоваться для усовершенствования методов QSAR при оценке энергий белковых структур, в частности, при оценке выгодности аминокислотных замен, а также оценке взаимодействия иных молекул с растворителем в ходе молекулярной инженерии белков и разработки лекарств. Полученные статистические потенциалы также могут использоваться для создания огрублённых моделей в компьютерном моделировании белков.

Результаты работы в части моделирования самоорганизующихся нанофибрилл из тиюфен-пептидных гибридов представляют практический интерес с точки зрения дизайна и создания перспективных наноматериалов для нужд органической микро- и нанoeлектроники, например, создания проводящих нанопроводов, светодиодов, сенсорных систем и т.д.

## **На защиту выносятся следующие основные результаты и положения:**

1. Изучен процесс гидратации боковых цепей аминокислот вблизи поверхности воды путём вычисления профилей свободной энергии. Разработан метод оценки свободной энергии адсорбции молекул на жидкую границу раздела на основе классических полноатомных моделей молекул. Проведена оценка энергий гидратации и адсорбции на границу вода/воздух для боковых цепей аминокислот. Выявлено, что энергии гидратации боковых цепей аминокислот находятся в хорошей корреляции с абсолютными энергиями адсорбции на поверхность воды. В тоже время энергия адсорбции относительно фазы максимального сродства является независимым параметром, опи-



сывающим свойства веществ.

2. Исследовано распределение аминокислотных остатков в глобулярных белках в зависимости от их погружённости в структуру белка путём статистического анализа большого количества экспериментальных данных по пространственному строению эволюционно-независимых белков. В ходе анализа для аминокислотных остатков оценивалась поверхность, доступная растворителю, эта величина использовалась для построения статистических распределений аминокислот. Таким образом получены статистические профили свободной энергии различных типов аминокислот. С помощью различных методов классификации аминокислот на “ядерные” и “поверхностные” рассчитаны статистические энергии переноса аминокислот из ядра на поверхность, рассчитаны корреляции этих статистических энергий с экспериментальными коэффициентами распределения аминокислот между рядом растворителей. Показано, что коэффициенты корреляции сильно зависят как от выбора пар растворителей, так и от выбора критериев статистической классификации аминокислот. Обсуждаются причины отличий в различных случаях. Предложены три статистические шкалы, которые наилучшим образом коррелируют с различными экспериментальными шкалами.
3. Сконструированы различные возможные модели укладки гибридных молекул тетратиофентиофен-олигопептид ( $[\text{Thr-Val}]_3$ ) с образованием нанопибрилл. Укладки предсказаны на основании рационального конструирования с учётом имеющихся экспериментальных данных с последующей оценкой стабильности и анализом методом молекулярной динамики. Показано, что изучаемые молекулы могут формировать стабильные, упорядоченные фибриллоподобные молекулярные слои благодаря организации пептидных частей молекулы в параллельный или антипараллельный бета-слой. Показано, что однослойные структуры в объёме обладают тенденцией к образованию суперспирали. Двуслойные структуры, в отличие от однослойных, в объёме сохраняют свою линейность и, в зависимости от типа укладки, могут обладать закрученностью. Предложены различные варианты упаковки, в которых тиофеновые сегменты слоёв (лент) образуют линейные структуры по типу нанопроводов, которые, предположительно, могут обладать электропроводностью по всей длине. Наиболее вероятной упаковкой молекул в фибриллах, соответствующей экспериментальным данным, является упаковка основанная на параллельном бета-слое.

### **Апробация работы**

Основные результаты работы докладывались и обсуждались на VIII Конференции сту-

дентов и аспирантов Учебно-научного центра по химии и физике полимеров и тонких органических плёнок (Москва, Россия, 2004), Международной школе-конференции молодых учёных: Системная биология и Биоинженерия (Москва, Россия, 2005), Малом Полимерном Конгрессе (Москва, Россия, 2005), Четвёртой Всероссийской Каргинской Конференции “Наука о полимерах 21-му веку” (Москва, Россия, 2007), Европейском Полимерном Конгрессе (Москва, Россия, 2005), Третьей Конференции STIPOMAT (Лё Диаблере, Швейцария, 2007), Второй Санкт-Петербургской Международной Конференции по НаноБиоТехнологиям “НаноБио’08” (Санкт-Петербург, Россия, 2008), Пятой Международной Конференции по Биоинформатике Регуляции и Структуры Генома (Новосибирск, Россия, 2006), Первой международной летней школе - Нано2009: Наноматериалы и нанотехнологии в живых системах (Московская обл., Россия, 2009).

### **Публикации**

Материалы диссертации опубликованы в 15 печатных работах, из них 6 статей в рецензируемых журналах [1–6], 1 статья в сборниках трудов конференций [7], 1 статья в книге [8], и 8 тезисов докладов [9–16]

**Личный вклад автора** состоит в разработке оригинальных алгоритмов и моделей, проведении компьютерного моделирования и статистического анализа, интерпретации и анализе полученных результатов.

### **Структура и объём диссертации**

Диссертационная работа изложена на 150 страницах печатного текста и включает 54 рисунка. Диссертация состоит из введения, пяти глав, заключения, списка цитируемой литературы, состоящего из 147 наименований, и 4 приложений.

# Принципы самоорганизации белковых и пептид-полимерных структур

Самоорганизация биологических макромолекул лежит в основе функционирования всех живых систем. Понятие самоорганизации может трактоваться достаточно широко, начиная от соединения атомов в определённые молекулярные структуры до процессов эволюции биологических организмов и возникновения интеллекта. В контексте данной работы нас будет интересовать один аспект самоорганизации молекулярных систем – самосборка макромолекул, содержащих аминокислотные последовательности. К таким молекулам, естественно, относятся природные полипептиды, или белки. В последнее время также широкое распространение получили исследования, связанные с искусственным синтезом определённых пептидных последовательностей, а также гибридных молекул (синтетический полимер + пептид) как для биотехнологических приложений, так и для создания новых перспективных материалов.

Способность природных аминокислотных последовательностей с высокой точностью сворачиваться в уникальную пространственную структуру является образцом для подражания при разработке новых материалов в области нанотехнологий. Следует выделить ещё несколько общих свойств белковых структур, изучение которых, в том числе, весьма полезно для различных технологических приложений. Во-первых, это возможность белков самоорганизовываться в надмолекулярные структуры с поразительными механическими и прочностными характеристиками. Так, некоторые типы паутины по своим прочностным характеристикам сравнимы со сталью. При этом они обладают эластичностью, как у резины [17], благодаря самосборке молекул в структуры содержащие как кристаллические участки, так и аморфные. Во-вторых, следует отметить замечательное свойство адаптивности, которым природа наделила пептидные системы. Адаптивность относится как к подстройке свойств системы благодаря изменению пептидной последовательности (без этого не могла бы происходить эволюция живых систем), так и к изменению свойств под воздействием внешних условий среды.

### 1.1. Пептиды и их строение

Строение белков, их химическая структура и принципы организации подробно изложены во многих учебниках и монографиях (например, [18, 19]). В данном разделе мы лишь

приведём справочные факты, которые будут полезными при изложении содержательных глав диссертации.

Природные пептиды являются результатом реакции поликонденсации L-α-аминокислот (см. рис. 1.1), которые соединены между собой так называемой пептидной связью между С и N атомами, особенностью которой является ее планарность из-за размывания электронной плотности C=O связи также и на C-N-связь.

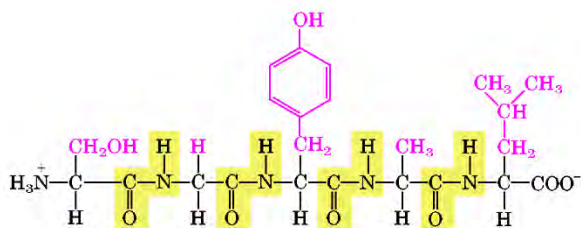


Рис. 1.1. Пример химической формулы пента-пептида, желтым цветом выделены пептидные связи между аминокислотными остатками.

Природные пептиды состоят из 20 типов аминокислот (одна из которых, пролин, является, строго говоря, иминокислотой), отличающимися друг от друга типом заместителей у центрального атома углерода, называемого C-α-атомом. Их положение в белковой цепи кодируется генами. Однако последующая модификация белка иногда увеличивает разнообразие аминокислот. Все природные аминокислоты являются L-аминокислотами, то есть обладают определенным типом стереоизомерии. Наличие в природных пептидах лишь одного типа энантиомеров имеет важное значение при формировании как пространственной, так и надмолекулярной структуры, поскольку влияет на хиральность этих структур. Список всех 20 природных аминокислот и их формулы приведены в приложении А. По принятой классификации аминокислоты разделяют по свойствам их боковых цепей на: гидрофобные, ароматические, полярные, положительно заряженные и отрицательно заряженные. Связь между хиральностью аминокислот и хиральностью формируемых из пептидов структур определяется уже энергетической выгодностью того или иного расположения. Стерически для пептидов из L-аминокислот разрешены как правые, так и левые альфа-спирали, однако в реальности наблюдаются только правые. А амилоидные фибриллы, состоящие из бета-листов, обладают свойствами лево закрученной ленты. Хиральность аминокислоты можно определить посредством мнемонического правила “CORN” (см. рис. 1.2): если смотреть на C-α-атом аминокислоты так, чтобы атом водорода был направлен от наблюдателя, то последовательность групп COOH, R, NH<sub>2</sub> должна быть расположена против часовой стрелки для L-формы и по часовой стрелке для D-формы.

Первичную последовательность аминокислотных остатков принято записывать от

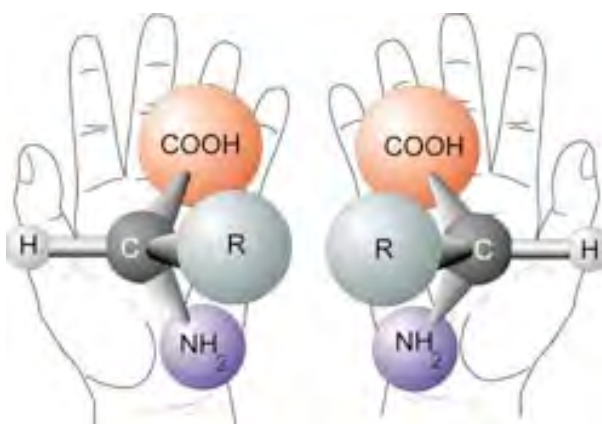


Рис. 1.2. Иллюстрация стереоизомерии аминокислот, слева аминокислота в L-форме, справа – в D-форме.

N-конца к C-концу. Конформация пептидов обычно описывается в терминах торсионных (двугранных) углов остова пептидной цепи  $\phi$ ,  $\psi$  и  $\omega$  (см. рис. 1.3). Торсионные углы вращения вокруг соответствующих связей определяются как двугранные углы между плоскостями, задаваемыми смежными тройками атомов (например, угол  $\phi$  – это двугранный угол между плоскостями  $(CNC_\alpha)$  и  $(NC_\alpha C)$ ). По договорённости для абсолютно вытянутой конформации пептидной цепи (когда все пептидные связи лежат в одной плоскости) углы  $\phi$  и  $\psi$  принимаются равными  $180^\circ$ . Опять же по договорённости, если смотреть со стороны  $C_\alpha$ -атома, то торсионные углы  $\phi$  и  $\psi$  увеличиваются, когда азот или карбонильный углерод, соответственно, крутятся по часовой стрелке. Планиарная пептидная связь, за которой закреплён угол  $\omega$ , теоретически может находиться в *cis* ( $\omega = 0^\circ$ ) и *trans* ( $\omega = 180^\circ$ ) конформациях. Однако стерическое отталкивание  $C_\alpha$ -атомов приводит к тому, что почти все пептидные связи в белках находятся в *trans* конформациях. Исключение составляет лишь пептидная связь, предшествующая пролину (имино-, а не аминокислоте): так как в пролине два почти симметричных массивных радикала при N атоме, то для этой связи *trans* конформация лишь не намного лучше, чем *cis*.

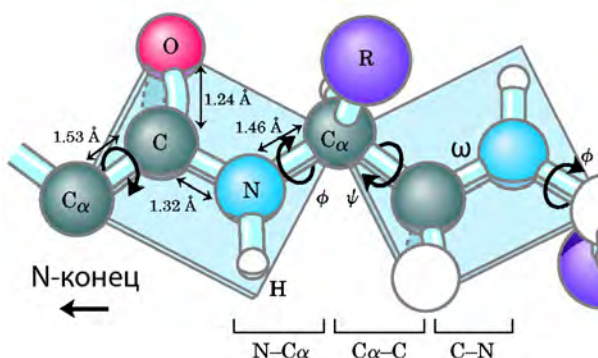


Рис. 1.3. Торсионные углы и длины связей в пептидной последовательности.

Из-за перекрываний атомов не все наборы углов  $\phi$  и  $\psi$  являются стерически разрешёнными, допустимые значения углов, а также соответствие углов определённым укладкам вторичной структуры удобно изображать на двумерной карте вероятностей, называемой картой Рамачандрана (Рис. 1.4).



Рис. 1.4. Карта Рамачандрана и области углов, соответствующие различным элементам вторичной структуры. Адаптировано из [18].

Важнейшей особенностью пептидов является возможность образования водородных связей между карбонильным атомом кислорода и атомом азота с участием ковалентно связанного с азотом атома водорода. Такое взаимодействие лежит в основе образования всех вторичных структур в белках. Какова энергия водородной связи при взаимодействии аминокислотных остатков в пептиде? В принципе, энергия водородной связи составляет порядка 23 кДж/моль (для воды [18]). Однако если вести речь об изменении свободной энергии системы полипептид + вода при образовании водородной связи между атомами полипептида, то энергетический выигрыш оказывается очень мал, так как по сути он определяется всего лишь разницей трансляционной энтропии молекулы воды и энтропийных потерь цепи при образовании связи [19]. Оценка для вклада в свободную энергию трансляционной энтропии молекулы воды составляет 10 кДж/моль.

Ключевым для понимания процессов самоорганизации является также распределение парциальных зарядов вдоль пептидного остова. Пептидная связь полярна, карбонильный атом кислорода имеет частичный отрицательный заряд, а атом азота – частично

положительный (по некоторым оценкам частичный заряд достигает  $\pm 0.5e$ ), что создаёт небольшой дипольный момент. Обычно говорят, что С-конец полипептидной цепи имеет частичный отрицательный заряд, а N-конец – положительный. При некоторых обстоятельствах (например, в альфа-спирали) дипольные моменты могут складываться, образуя большой дипольный момент.

## 1.2. Пространственная структура белков

Организацию структуры в белках принято подразделять на 4 уровня: первичная, вторичная, третичная и четвертичная структуры. *Первичной структурой* называют последовательность аминокислот в полипептидной цепи. Ко *вторичной структуре* относят локальную укладку небольших частей полипептида в альфа-спирали, бета-листы и другие типы элементов. К *третичной структуре* относят общую укладку атомов белка в пространственную структуру, включая взаимное расположение элементов вторичной структуры. Выделяют также *супервторичную структуру* (или мотивы укладок), которая представляет собой наиболее стабильные варианты укладки нескольких элементов вторичной структуры (один из таких элементов - закрученный бета слой будет обсуждаться позднее).

Некоторые белки состоят из нескольких полипептидных цепей, формирующих субъединицы белка. В этом случае говорят о *четвертичной структуре* белка как о взаимном расположении субъединиц относительно друг друга. Белки, состоящие из нескольких субъединиц также называют *мультимерами*. Мультимеры с небольшим количеством субъединиц называют также *олигомерами*. Если белок состоит из набора одинаковых частей (субъединиц или групп субъединиц), то такие части принято называть *протомерами*.

Определение пространственной структуры белков является достаточно нетривиальной задачей. Для её решения используется целый арсенал методов: от прямого экспериментального определения до построения различных компьютерных моделей. Наиболее актуальными методами на сегодняшний день являются методы рентгеновской дифракции, методы ядерного магнитного резонанса и методы электронной микроскопии. Последние зачастую позволяют определить внешнюю форму белка, но извлечь информацию о расположении атомов удаётся крайне редко. Методы двумерного и многомерного ядерного магнитного резонанса совместно с достижениями в секвенировании и синтезе полипептидных последовательностей позволяют определять пространственные структуры белков в растворе. Однако эти методы обладают ограничениями по размеру исследуемых бел-

ков. Методы рентгеновской дифракции и последующего рентгено-структурного анализа являются наиболее часто используемыми, однако для применения этих методов необходимо получить кристалл исследуемого белка, что является сложной задачей, а для многих белков (особенно мембранных) практически невозможной. Все публикуемые трёхмерные структуры белков помещаются на сегодняшний день в единую базу данных - Protein Data Bank [20]. На сегодняшний день в этой базе содержится около 50 тысяч структур, определённых методом рентгено-структурного анализа, и около 7 тысяч структур, определённых методом ЯМР.

### 1.2.1. Типы белков

Белки разделяют на группы согласно их трёхмерной структуре. Большинство белков относятся к *глобулярным*: они представляют собой плотно упакованные молекулярные структуры более или менее компактного вида. Меньшая часть белков относится к *фибриллярным*: их молекулы (обычно и надмолекулярные комплексы) в работающем состоянии представляют собой вытянутые волокна. Такие белки выполняют в основном структурные и защитные функции в живых системах (составная часть волос, когтей и т.д.). Большинство энзимов (ферментов) и регуляторных белков, наоборот, являются глобулярными белками. Глобулярные белки в свою очередь можно условно подразделить на водорастворимые и мембранные. Мембранные белки в живых системах располагаются в липидных мембранах и имеют в своей структуре трансмембранную часть, внешняя часть которой обычно представлена аминокислотами с гидрофобными боковыми цепями. Глобулярные водорастворимые белки располагаются в цитоплазме клетки, и их поверхность богата гидрофильными аминокислотами.

Полипептиды из более, чем нескольких сотен аминокислот, зачастую сворачиваются в две или более стабильных глобулярных единицы, называемых *доменами*. Пример такого белка тропонина С, кальций-связывающего белка, присутствующего в мышцах, изображён на рис. 1.5. С физической точки зрения доменное строение больших белков выгодно, поскольку с ростом размера глобулы (домена), во-первых, уменьшается скорость сворачивания белка, а, во-вторых, для формирования больших глобул необходимо было бы наличие в первичных последовательностях больших участков из гидрофобных аминокислот, которые формировали бы ядро глобулы, что несколько противоречит квазислучайной организации первичных последовательностей, которая наблюдается у глобулярных белков. То есть статистика последовательностей весьма похожа на ту, которая получалась бы в результате случайной сополимеризации аминокислот. Хотя, конечно же, последователь-



ности белков далеко не случайны. Так, по оценке ([19], стр. 255) доля белковоподобных последовательностей среди всех случайных составляет около  $10^{-8}$ .

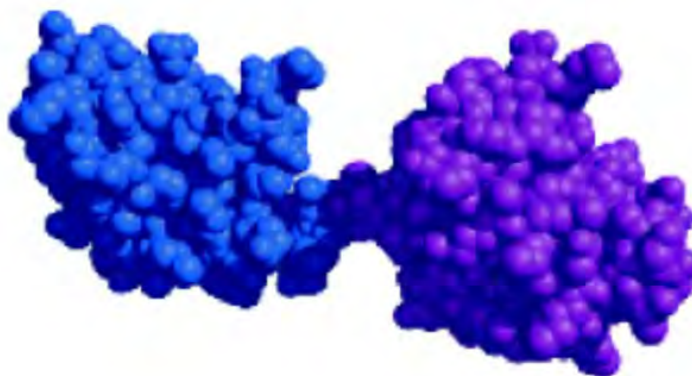


Рис. 1.5. Структурные домены в белке тропонин С (PDB ID 4TNC).

### 1.2.2. Бета-листы

Рассмотрим подробнее вопрос об организации и свойствах такого типа вторичной структуры, как бета-листы. Два возможных типа укладки полипептидной цепи в бета-слой представлены на Рис. 1.6.

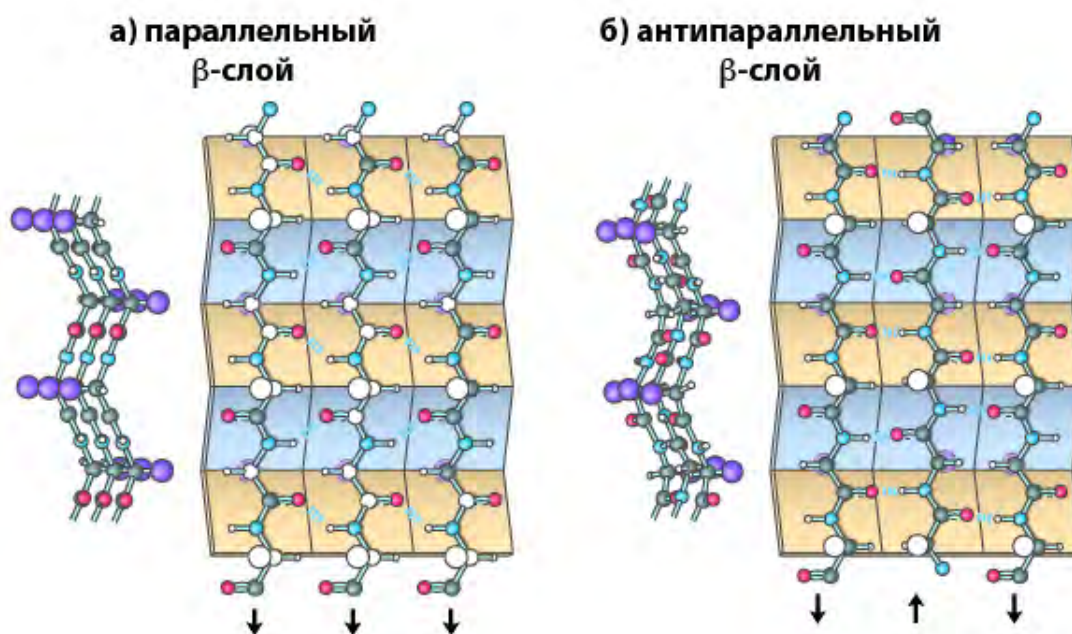


Рис. 1.6. Укладка полипептидной цепи в бета-слой.

Период укладки бета-нитей составляет около 5 Å. Антипараллельная укладка считается несколько более энергетически выгодной, в ней водородные связи образуются тремя

атомами лежащими на одной прямой. Период бета-нитей (длина вдоль пептидной цепи) составляет 6.5 Å и 7 Å для параллельного и антипараллельного случая, соответственно. В русской номенклатуре существует некоторая путаница в чётких определениях структурных элементов листа. В этой работе мы будем называть *бета-нитью* отдельный участок полипептидной цепи, набор которых формирует *бета-слой*, или *бета-лист*. Важной особенностью является хиральное поведение бета-листов. Бета-нити редко бывают абсолютно вытянутыми. Обычно они немного закручены в результате флуктуации двугранных углов. Считается, что такое поведение способствует лучшей упаковке бета-нитей в бета-лист. Закрутка бета-нити почти всегда правая. Одновременно с небольшой закруткой отдельной нити происходит и вращение плоскости всего листа в левую сторону. Однако, в литературе иногда встречается определение такого типа закрутки как “право-закрученный бета-лист”, что, по-видимому, всё-таки относится к закрутке, именно, плоскостей бета-нитей, а не всего листа. Пример закрутки на разных иерархических уровнях приведён на Рис. 1.7.



Рис. 1.7. Пример бета-листа с кручением отдельных бета-нитей и всего листа из структуры белков.

Если два и более бета-листов соприкасаются своими плоскостями, то обычно аминокислоты, участвующие в образовании интерфейса, обладают маленькими боковыми группами. Очень показательным является пример устройства кристаллических доменов натурального шёлка (белка фиброина), представленный на Рис. 1.8.

### 1.3. Термодинамика формирования пространственной структуры белков

Вопрос сворачивания белков очень активно исследуется в современной биологии и является достаточно дискуссионным, особенно в части кинетики сворачивания белков.

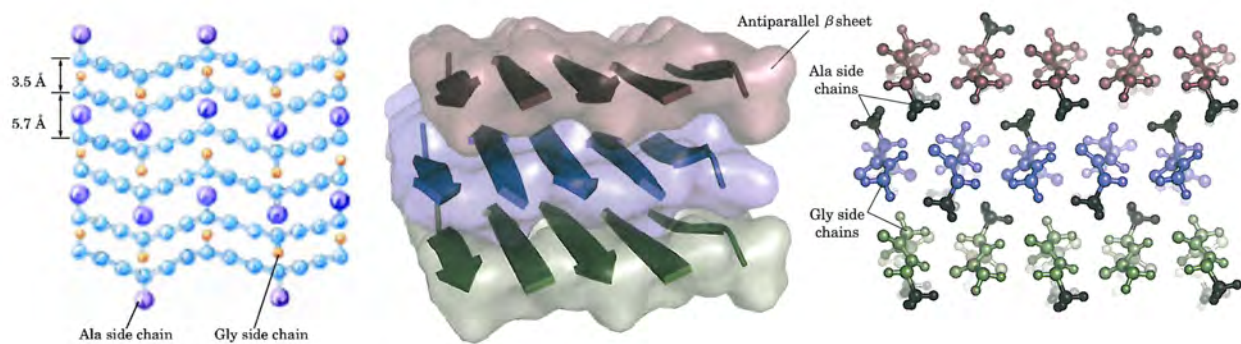


Рис. 1.8. Структура кристаллических доменов белка фиброина (натуральный шёлк). Адаптировано из [18].

Ниже, не претендуя на полноту, рассмотрим основные современные представления о термодинамике сворачивания белков.

- Согласно гипотезе Анифсена пространственная структура определяется первичной последовательностью белка, денатурация белка обратима, по крайней мере, для небольших глобулярных, водорастворимых белков. Белки обладают лишь одной стабильной конформацией, отвечающей глобальному минимуму свободной энергии. Разница в энергии между нативным и денатурированным состоянием белка составляет 20-65 кДж/моль, 5-15 kT [18], то есть нативные структуры белков “едва” стабильны. Стабильность белка определяется суммой “слабых” взаимодействий, таких как гидрофобные взаимодействия, стерические контакты, энтропийные выигрыши от образования водородных связей, ионных пар и т.д. Особо выделяют роль гидрофобных взаимодействий.
- Белковая глобула – атомарно упорядоченная структура; положения боковых групп в белке фиксировано (быть может за исключением поверхностных).
- Денатурация белка является кооперативным переходом типа “*всё-или-ничего*” и в этом плане не может быть объяснена в рамках теории перехода клубок-глобула для случайный гетерополимеров. Денатурация белка напоминает фазовый переход первого рода. При денатурации белок может проходить через ещё одно фазовое состояние, не свойственное случайным гетерополимерам, – состояние расплавленной глобулы.
- Глобальной особенностью, отличающей белок от случайного гетерополимера, является наличие энергетической щели ( $\Delta E \gg kT_C$ ), отделяющей нативное состояние от всех остальных. При этом “вымораживание” глобулы происходит скачком ещё до того, как полипептид начинает стекловаться. Приблизительная термодинамическая

оценка количества случайных последовательностей имеющих подобное свойство –  $10^{-8}$ . На Рис. 1.9 приведены формы зависимости энергии, энтропии, свободной энергии для типичных белковых глобул.

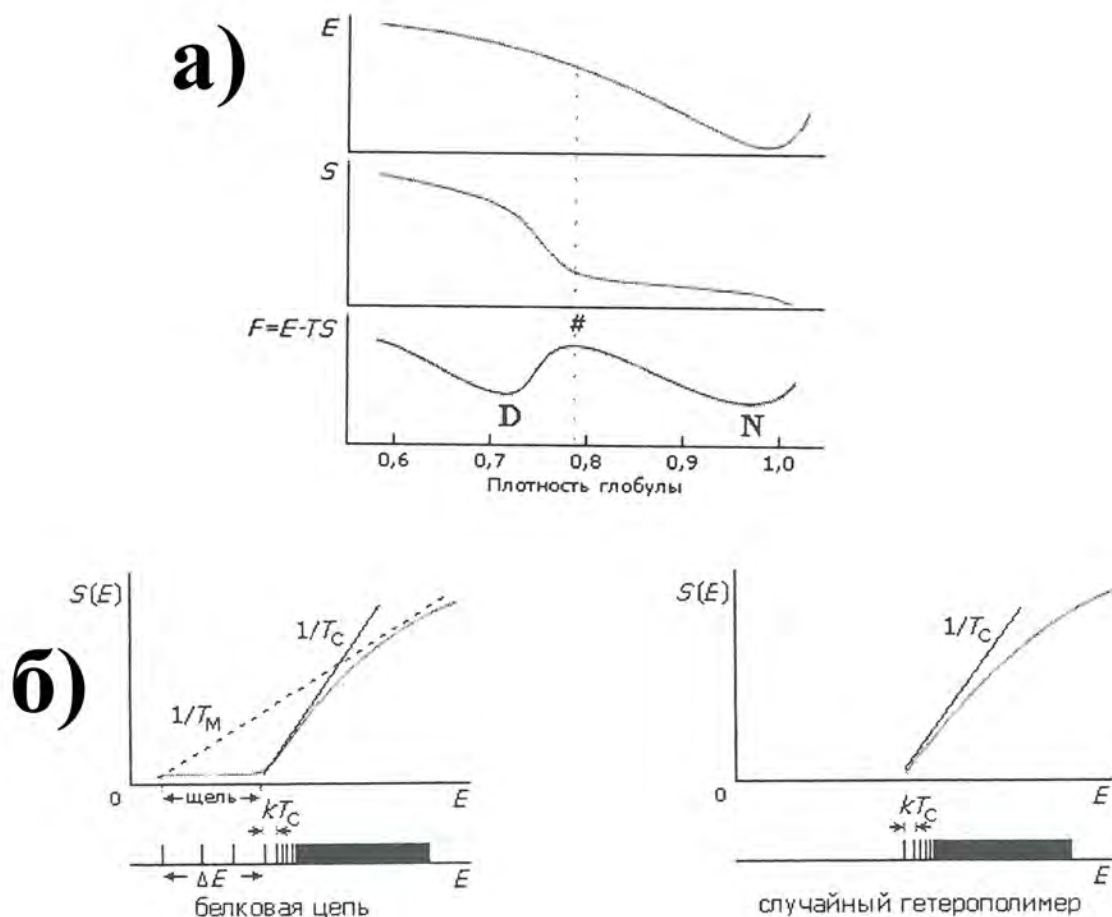


Рис. 1.9. Характерные зависимости термодинамических параметров для белковой глобулы: а) зависимость энергии, энтропии и свободной энергии от плотности. Резкий подъём энтропии связан с появлением вращательной подвижности у боковых групп. б) Зависимость энтропии от энергии, спектры энергий для белковой цепи и случайного гетерополимера. Адаптировано из [19].

### 1.3.1. Квазибольцмановская статистика распределения элементов в белках

Статистическое распределение структурных элементов в белках обладает квази-Больцмановской статистикой по отношению к оценке энергетического вклада данного структурного элемента в свободную энергию системы. Первые наблюдения такого рода касались распределения торсионных углов боковых групп и распределения аминокислот между поверхностью и ядром белка. Теоретическое обоснование такой формы статистики было дано Финкельштейном и др. в работе [21] на основе модели случайных энергий (МСЭ) для ге-

терополимера [22]. Кратко суть их выводов можно сформулировать следующим образом: когда речь идёт о белковоподобных последовательностях в рамках МСЭ, то есть о последовательностях с минимальной энергией, оказывается, что количество таких случайных последовательностей, которые содержат какую-то структурную особенность, экспоненциально зависит от энергии этой структурной особенности (речь идёт об энергии усреднённой по всем возможным компактным укладкам случайных последовательностей). Такая зависимость в силу квазислучайности белковых последовательностей, а также тем фактом, что множество белковых последовательностей можно рассматривать как подмножество последовательностей случайных гетерополимеров с минимальными по энергии компактными укладками, распространяется и на природные белки.

Согласно рассмотрению работы [21] встречаемость структурных элементов в белках определяется формулами:

$$\text{ВСТРЕЧАЕМОСТЬ} \propto \exp(-\Delta F_{select}^1/RT_c) \quad (1.1)$$

$$\Delta F_{select}^1 = (\Delta_E - \Delta_\sigma/2RT_c) - RT_c \ln M_1/M \quad (1.2)$$

где  $F_{select}^1$  это “селективная (кажущаяся, статистическая) свободная энергия” некоторого элемента структур (обозначаемого, как “1”, например, наличия аминокислоты лейцин в 72 положении на поверхности белка), ВСТРЕЧАЕМОСТЬ – это отношение количества наблюдаемых нативных укладок в белках, содержащих элемент “1”, к количеству укладок, не содержащих этот элемент.  $T_c$  – конформационная температура, типичная для замерзания случайного гетерополимера,  $\Delta_E$  – средняя (по различным компактным укладкам случайного гетерополимера) разница энергий между укладками с наличием элемента “1” и его отсутствием,  $\Delta_\sigma$  – изменение в дисперсии энергии между стабильными укладками с и без элемента “1”,  $M_1/M$  представляет долю среди всех укладок (а не только с минимальной энергией), которые содержат элемент “1”. Конформационная температура для белков находится в районе 300 К [23]. Однако Томас и Дилл [24] утверждали, что если рассматривать вопрос о распределении аминокислот между ядром и поверхностью белка, белки в базе данных PDB могут не совсем адекватно соответствовать модели случайного гетерополимера Финкельштейна и др., поскольку эффективная конформационная температура  $T_c$  зависит от длины, состава и компактности белка, в то время как для случайного гетерополимера результаты не зависят от длины и состава. Томас и Дилл показали, что для белковых наборов с различными характеристиками конформационные температуры для распределения аминокислот могут сильно различаться, однако квази-Больцмановская статистика остаётся в силе.

## 1.4. Амилоиды и амилоидные фибриллы

По современному *биофизическому* пониманию [25] *амилоидами* называют любые полипептиды, которые образуют особый тип четвертичных структур, называемый кросс-бета листом (cross-beta sheet, cross-beta spine). Кросс-бета лист представляет собой несколько бета-листов, сложенных своими плоскостями. Такая структура обладает характерным набором рефлексов при рентгеновском рассеянии. Рефлекс, соответствующий периоду укладки бета-нитей в бета-слой, обычно отвечает расстоянию в 4.8-4.9 Å (например, 4.78 Å [26]). Пример кросс-бета листовой структуры приведён на Рис. 1.10. Большой интерес к изучению амилоидов связан с их вовлечённостью во многие медицинские заболевания, связанные с накоплением и неправильной агрегацией различных белков. Заболевания, так или иначе связанные с накоплением в организме амилоидов, включают очень широкий спектр болезней: болезнь Альцгеймера, болезнь Паркинсона, сахарный диабет второго типа, коровье бешенство, атеросклероз, ревматоидный артрит, различные систематические амилоидозы, амилоидозы сердца, почек, печени и т.д.

Очень часто агрегация амилоидов приводит к образованию фибриллярных структур, называемых амилоидными фибриллами. Возможность образования фибрилл определяется как термодинамическими, так и кинетическими факторами (поскольку полимер может застекловаться ещё до обретения термодинамически равновесной структуры, что, скажем, происходит при выработке паутины пауком). Широкую популярность для качественного описания возможных структур, образующихся при агрегации амилоидов, получила теория Аггели-Бодена-Семёнова и др. [28]. На Рис. 1.11 представлены различные уровни организации амилоидов, а также фазовая диаграмма при различных параметрах модели. Уровни организации включают в себя однослойные ленты, закрученные в суперспираль, двуслойные и многослойные агрегаты, формирующие фибриллы, а также “супер фибриллы”, состоящие из переплетающихся фибрилл (модель f' на Рис. 1.11).

В амилоидных фибриллах бета-листы стремятся быть более плоскими, чем бета-листы, наблюдаемые в нативных белковых структурах.

## 1.5. “Молекулярные химеры” и принципы их построения

“Молекулярные химеры”, или полимерные биоконъюгаты, или гибриды синтетических и биополимеров, являются в последнее время очень активно исследуемой областью наук о материалах. В литературе существуют несколько обзоров, которые хорошо суммируют прогресс в этой области за последнее время [29–35]. Материалы на основе гибридов



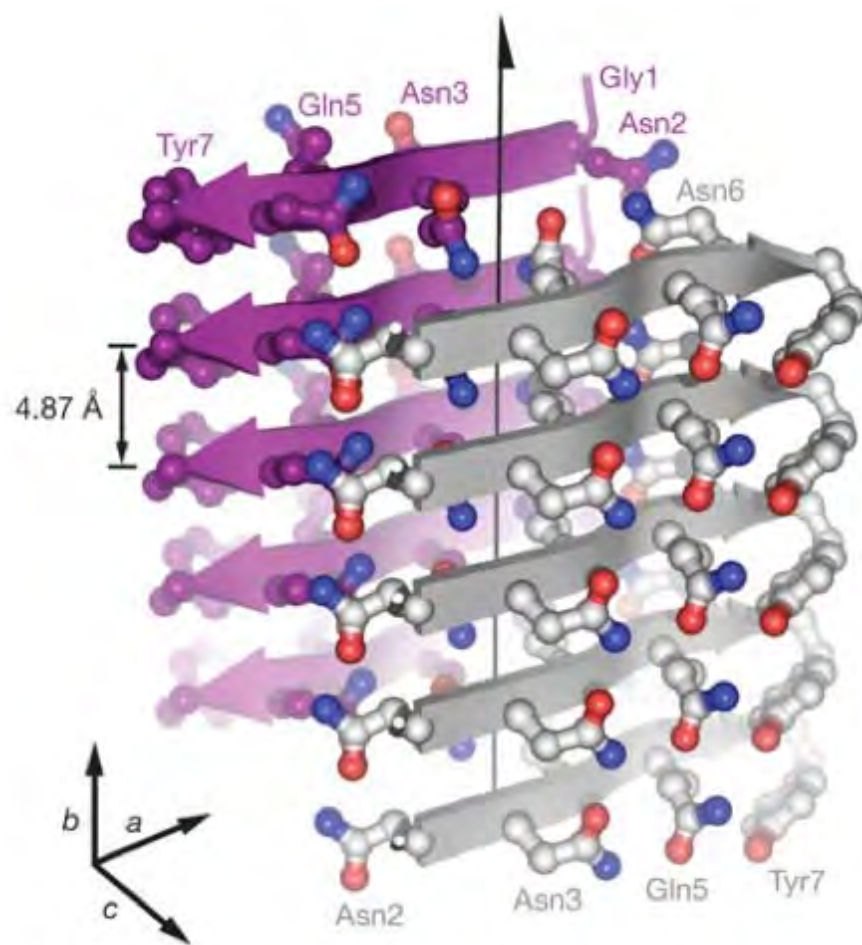


Рис. 1.10. Пример кросс-бета листовой структуры, образуемой искусственно синтезированной последовательностью [27].

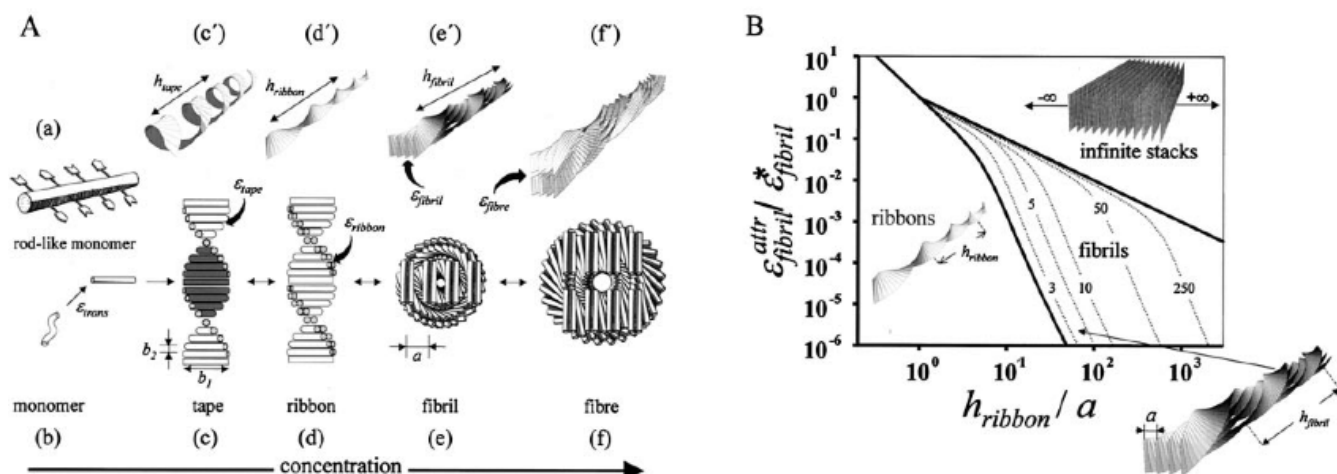


Рис. 1.11. Возможные уровни самоорганизации амилоидов в фибриллы по [28].

синтетических полимеров с пептидами обладают рядом инновационных свойств, которые относятся к (I) самосборке макромолекул за счёт агрегационных свойств пептидов, (II) возможностям изменения структуры материалов в зависимости от внешних условий, (III) широких возможностей по дизайну последовательностей и управлению свойствами получаемых структур. В области гибридов пептидов и синтетических макромолекул широко используются амилоидоподобные пептидные последовательности, а также последовательности образующие альфа-спирали в определённом диапазоне условий [34]. Важной вехой явился переход от водорастворимых полимеров к органорастворимым [31, 32], поскольку это открывает возможности для применения в различных технологических областях. На Рис. 1.12 представлены различные пептид-полимерные гибриды и их гипотетические модели. Особый интерес вызывают гибриды на основе амилоидов, способные формировать фибриллярные структуры. Обычно такие фибриллярные структуры обнаруживаются на подложке с помощью АСМ или ПЭМ. Для большинства стабильных фибриллярных агрегатов можно утверждать, что наблюдаемые на подложке фибриллы присутствуют и в растворе. Доказательству этого факта был посвящён ряд исследований с применением криоэлектронной микроскопии и малоуглового нейтронного рассеяния [36].

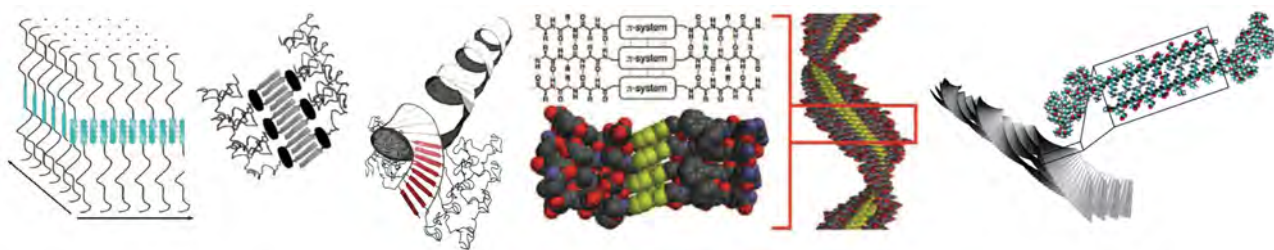


Рис. 1.12. Предполагаемые мотивы организации некоторых синтезированных пептид-полимерных гибридов на основе последовательностей способных к образованию бета-листов, по [32, 33, 37, 38].

## 1.6. Гибриды тиофенов и пептидов

Гибриды тиофенов и пептидов представляют перспективный способ дизайна новых самособирающихся функциональных  $\pi$ -систем. Из-за делокализации электронов, вызванной  $\pi$ – $\pi$ -сопряжением, политиофены относятся к классу проводящих полимеров (“синтетических металлов”)[39]. (Полу)проводящие, оптические и электролюминисцентные свойства производных политиофенов, а также удобные методы обработки полимеров (по сравнению с твердотельными полупроводниками) привлекли к исследованию тиофенов большое внимание в последние десятилетия, в том числе, из-за больших возможностей технологического применения в таких областях, как органическая микроэлектроника (включая



светодиоды [40], гибкие дисплеи [41], фотоэлектрические преобразователи [42]), дизайн сенсоров и биосенсоров [43, 44], дизайн биоэлектрических интерфейсов [45] и т.д.

Ковалентная пришивка других химических веществ к тиофенам может влиять на их структурные и электронные свойства, которые зависят от эффектов упорядочения как на молекулярном, так и на наноуровне [46]. Многообещающий подход к разработке новых инновационных функциональных производных тиофенов состоит в дизайне гибридов тиофенов и биологических макромолекул. Конъюгация олиготиофенов и биологических макромолекул может позволить создать новые вещества, которые кроме полупроводниковых, оптических и электролюминисцентных свойств тиофенов обладают способностями к самосборке, специфическому связыванию, адаптивности к внешним условиям, свойственным биологическим молекулам. Такой подход открывает новые возможности для дизайна умных полимеров и молекул на наноуровне. Современные направления исследований включают разработку самособирающихся нанопроводов [37, 47, 48], химических и биологических сенсоров со специфическим связыванием [49], биоэлектрических интерфейсов и т.д. На данный момент были синтезированы гибриды олиготиофенов с нуклеотидами [50, 51], одиночными аминокислотами [52] и пептидами [37, 47, 48]. Главной идеей является использование способности биологических компонент формировать сеть водородных связей, которые определяются последовательностью биологических мономеров, и определяют поведение наноструктуры. Из-за относительной слабости водородных связей конечная структура будет зависеть от баланса различных внутри и межмолекулярных взаимодействий (водородные связи, Ван-дер-Ваальсовы и гидрофобные взаимодействия,  $\pi - \pi$ -стэкинг), что, во-первых, даёт широкие возможности для варьирования состава и следовательно структуры получающихся агрегатов, а, во-вторых, определяет зависимость супрамолекулярной организации от внешних условий (температура, качество растворителя, pH и т.д.), что полезно для некоторых технологических приложений [53].

Примеры “молекулярных химер”, в которых один блок является олиготиофеновым, а другой пептидом, способным к амилоидной агрегации, были исследованы несколькими группами. Клок и др. [47] синтезировали гибрид Т-Р типа, где тетра-3-гексилтиофен (Т) функционализирован пентапептидом Gly-Ala-Gly-Ala-Gly (Р), который образует бета-листовые домены в натуральном шёлке. В этой работе авторы наблюдали длинные линейные нити из Т-Р молекул и тенденцию пептидных блоков к формированию бета-листов. Однако, детальной молекулярной структуры этих агрегатов пока неизвестно. Шиллингер и др. [48] изучали симметричный триблок олигомер (Р-Т-Р) алкилированного тетратиофен-пептидного гибрида, который содержал последовательность [Thr-Val]<sub>3</sub>, которая также склонна к образованию бета-листов. Пептидная часть исследовалась в планарной кон-

формации и с наличием сложноэфирного дефекта, внесившего излом в пептидную цепь. В ходе исследования был обнаружен интересный эффект: молекулы, пептидные сегменты которых содержали дефект (и следовательно излом), образовывали более длинные фибриллы с хиральной закруткой, тогда как ранее такие дефекты наоборот использовались для предотвращения агрегации молекул (дефект необратимо реизомеризуется в пептидную связь при увеличении pH). На рис. 1.13 приведены наблюдаемые фибриллы. Дигельманн и др. [37] сообщали об успешном синтезе битиофенов присоединённых к последовательностям Ala-Phe-Glu-Gln-Gln и Glu-Phe-Ala-Gln-Glu (P-T-P тип гибрида). Во всех исследованиях с помощью АСМ или ПЭМ было показано, что при определённых условиях, эти “молекулярный химеры” могут самособираться в длинные (в микрометровом диапазоне) морфологически подобные фибриллярные агрегаты с шириной волокна в диапазоне 4-50 нм и высотой от 0.3 до 6 нм. Однако, прямое определение молекулярной структуры с помощью экспериментальных методов для этих структур весьма затруднительно и предполагаются различные модели их организации.

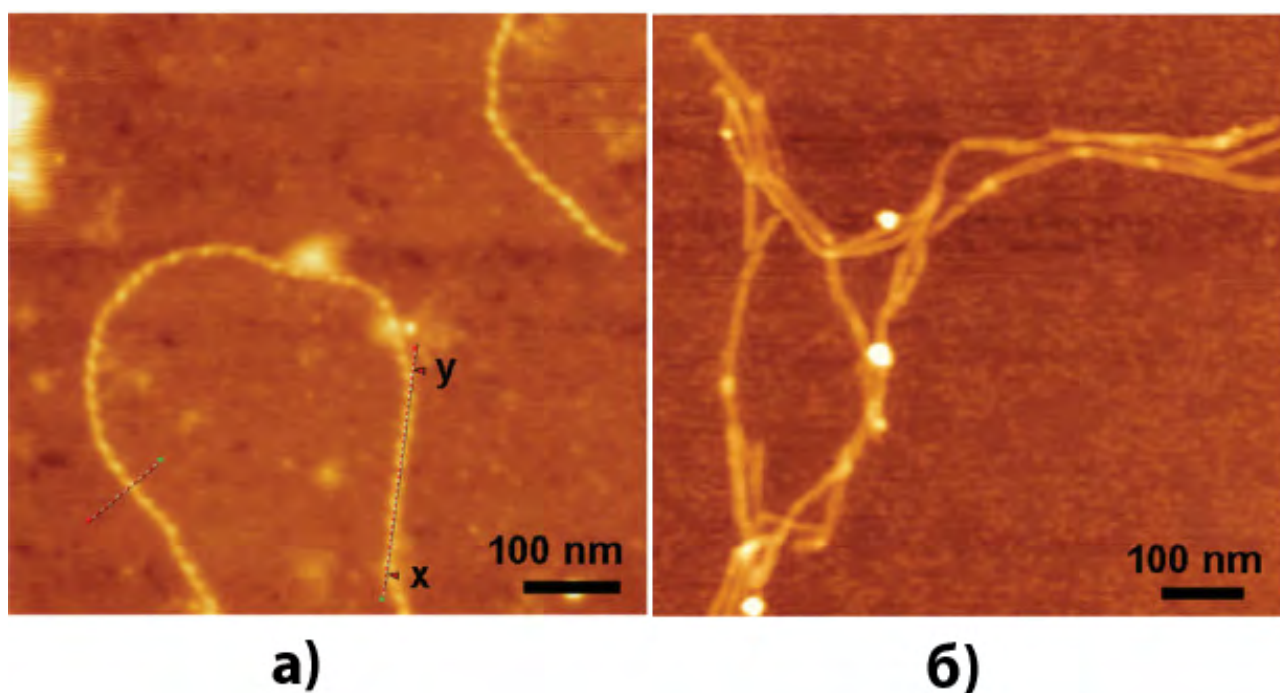


Рис. 1.13. Фибриллярные агрегаты из триблок олигомеров тетратиофена и пептидной последовательности [Thr-Val]<sub>3</sub>, наблюдаемые методом АСМ в работе [48]. а) при наличии сложноэфирного дефекта, б) при нативном состоянии пептидных сегментов

Компьютерное моделирование даёт возможность проводить изучение таких самоорганизующихся структур, различных упаковок молекул, их стабильности и баланса взаимодействий. Гуськова и др. [54–56] применяли принципы предсказания различных кристал-

лических упаковок тиофен-пептидных гибридов для последующего рационального конструирования фибриллярных агрегатов и их анализа методами молекулярной динамики. В работах были предложены возможные модели укладки молекул в нановолокна тиофен-пептидных гибридов, построенных на основе шёлкоподобной последовательности Gly-Ala-Gly-Ala-Gly.

# Некоторые методы компьютерного моделирования и анализа

С ростом возможностей компьютеров и суперкомпьютеров, развитием компьютерных технологий, в том числе параллельных вычислительных технологий, а также совершенствованием методов моделирования, последнее становится всё более более привлекательным инструментом для научных исследований, принимая во внимание также относительную незатратность моделирования по сравнению с экспериментальными исследованиями. В тоже время моделирование многих процессов *ab initio* (из первых принципов) не представляется возможным в силу чрезмерной вычислительной затратности. Поэтому широкое распространение получает концепция мультимасштабного компьютерного моделирования, в которой методы молекулярной механики и динамики занимают важное место, позволяя исследовать системы размером до миллионов атомов на временах до микросекунд.

С другой стороны в настоящее время, особенно в области наук о жизни, объёмы даже уже накопленных экспериментальных данных весьма внушительны (примером могут служить, расшифрованные геномы и структуры белков), которые опять же благодаря развитию компьютерных технологий и баз данных стали легкодоступными. В этой связи встаёт вопрос о компьютерном анализе накопленного объёма информации.

В данной главе речь пойдёт о методах молекулярной механики/динамики и о методе оценки молекулярных поверхностей.

## 2.1. Методы молекулярной динамики и механики

О методах молекулярной динамики написано множество подробных книг и монографий (см. например [57]).

В основе методов ММ/МД лежит представление о молекулярной структуре как о наборе классических частиц (материальных точек или в некоторых случаях твёрдых тел) взаимодействующих по законам классической механики. Молекулярно механическая модель обычно предполагает следующие приближения: (i) электронные степени свободы не учитываются, а учитываются лишь атомные степени свободы, (ii) молекулы полагаются находящимися в основном энергетическом состоянии, (iii) автоматически предполагается приближение Борна-Оппенгеймера. На практике при задании конкретных типов взаимодействий между атомами в виде молекулярного гамильтониана предпринимается ещё ряд

серьёзных приближений. Задача молекулярной динамики фактически заключается в решении классических уравнений движения для системы с заданным гамильтонианом. Решение уравнений движения реализуется численно с помощью разностных схем таких, как алгоритм Верле или Лип-Фрог алгоритм. Важной особенностью “хороших” алгоритмов является обратимость во времени и сохранение объёма фазового пространства. Известно также, что алгоритм Верле обладает “теневого” характеристикой, то есть точки траектории, получаемые посредством этого алгоритма, точно соответствуют (конечно, за исключением ошибки численного округления) некоторой “теневого траектории” механической системы, которая однако не проходит, через точку с начальными условиями. Иными словами, алгоритм Верле генерирует траекторию, которая соответствует экстремуму действия для некоторой “теневого траектории” ([57] разд. 4.3.5).

Для решения проблемы поверхностных эффектов в методе МД применяют периодические граничные условия. Особое внимание в последнее время обращается на тот факт, что при использовании периодических условий необходимо использовать решёточные алгоритмы для учёта дальнедействующих взаимодействий, таких как электростатические, поскольку обрезание зарядовых взаимодействий может приводить к артефактам моделирования. На данный момент хорошо разработаны достаточно быстрые решёточные алгоритмы учёта электростатики такие, как методы PME и PPPM.

Отдельной проблемой является задание ансамблей в методе молекулярной динамики. Для задания распределения по температуре существует два принципиальных подхода: стохастическая и негамильтонова динамика. В стохастическом подходе в уравнения движения вводятся случайные члены (например, такие как в уравнении Ланжевена), которые позволяют привести статистическое распределение точек МД траектории к каноническому ансамблю. Подход негамильтоновой динамики был предложен Гувером [58], он разработал такую модификацию уравнений движения, в которой предельным для точек фазового пространства является каноническое распределение. Однако в оригинальном алгоритме Нозе-Гувера были найдены проблемы при некоторых условиях, и на данный момент наиболее надёжным считается алгоритм цепей Нозе-Гувера.

### 2.1.1. Силовые поля

Рассмотрим примеры широко используемых моделей молекулярных систем, называемых, “силовыми полями”. Количество разработанных на сегодняшний день силовых полей поистине велико, они различаются как сложностью задания молекулярного гамильтониана, точностью, так и областями применимости. Для моделирования пептидных структур

широкое распространение получили такие силовые поля, как AMBER, OPLS, CHARMM, GROMOS, GROMACS и др. (эти поля иногда называют классическими силовыми полями), они сочетают относительную простоту задания энергии системы с разумной точностью. Для более широкого класса органических молекул используются поля MM2, MM3, MM4, MMFF и др. Существует достаточно большой класс поляризуемых силовых полей, некоторые из которых построены на основе классических силовых полей. К силовым полям, называемым *class II*, или силовыми полями второго поколения, относятся поля PCFF, COMPASS и др., эти поля содержат перекрёстные члены в энергии, обладают повышенной точностью, а также удобны с той точки зрения, что заряды в данных силовых полях определяются автоматически при заданных типах атомов. Поскольку вода является одним из важнейших объектов моделирования, то применительно к воде разработано большое количество молекулярных моделей, которые, хотя изначально могли разрабатываться в рамках определённых силовых полей, зачастую используют в сочетании с различными силовыми полями. Наиболее простыми и распространёнными моделями воды являются модели TIP3P, TIP4P, TIP5P, SPC, SPC/E и т.д.

### Силовое поле OPLS-AA

Силовое поле OPLS-AA [59, 60] является полноатомной версией поля OPLS (Optimized Potentials for Liquid Simulations), то есть атомы водорода рассматриваются явным образом. Параметры поля OPLS, в частности, подгонялись под экспериментальные свойства жидкостей, такие как плотность и теплота испарения. Общая структура энергии молекулярной системы представлена в уравнении 2.1.

$$\begin{aligned}
 U(r^N) &= U_{bond} + U_{angle} + U_{dih} + U_{nb} \\
 U_{bond} &= \sum_{bonds} K_r (r - r_{eq})^2 \\
 U_{angle} &= \sum_{angles} k_\theta (\theta - \theta_{eq})^2 \\
 U_{dih} &= \frac{V_1}{2} [1 + \cos(\phi)] + \frac{V_2}{2} [1 - \cos(2\phi)] + \frac{V_3}{2} [1 + \cos(3\phi)] + \frac{V_4}{2} [1 - \cos(4\phi)] \\
 U_{nb}^{ab} &= \sum_i^{on\ a} \sum_j^{on\ b} \left\{ 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \right\} f_{ij}
 \end{aligned} \tag{2.1}$$

межмолекулярные взаимодействия  $U_{nb}$  учитываются только для атомов, которые удалены на три и более связи, для 1-4 взаимодействий коэффициент скейлинга  $f_{ij} = 0.5$ , во всех остальных случаях  $f_{ij} = 1$ .

## Силовое поле PCFF

Силовое поле PCFF (Polymer Consistent Force Field) [61] основано на силовом поле CFF91 и относится к семейству силовых полей CFF (consistent force-field). Это поле создано для моделирования широкого класса органических полимеров, утверждается, что силовые поля второго поколения (к которым относится PCFF) более точно воспроизводят экспериментальные результаты, чем силовые поля первого поколения (AMBER, OPLS и т.д.). Платой за это является более сложное устройство поля и наличие перекрёстных членов в функции энергии. Общий вид энергии молекулярной системы в приближении поля PCFF приведён в формулах 2.2. На Рис. 2.1 приведено графическое пояснение к внутренним степеням свободы, являющимися параметрами для каждого члена потенциальной энергии.

$$\begin{aligned}
 U(r^N) &= U_b + U_\theta + U_\phi + U_\chi + U_{bb'} + U_{\theta\theta'} + U_{b\theta} + U_{b\phi} + U_{b'\phi} + U_{\theta\phi} + U_{\phi\theta\theta'} + U_{nb} \quad (2.2) \\
 U_b &= \sum_b K(b - b_0)^2 \\
 U_\theta &= \sum_\theta H(\theta - \theta_0)^2 \\
 U_\phi &= \sum_\phi V(1 + \cos(\phi - \phi_0)) \\
 U_\chi &= \sum_\chi K_\chi \chi^2 \\
 U_{bb'} &= \sum_{bb'} F_{bb'}(b - b_0)(b' - b'_0) \\
 U_{\theta\theta'} &= \sum_{\theta\theta'} F_{\theta\theta'}(\theta - \theta_0)(\theta' - \theta'_0) \\
 U_{b\theta} &= \sum_{b\theta} F_{b\theta}(b - b_0)(\theta - \theta_0) \\
 U_{b\phi} &= \sum_{b\phi} (b - b_0)[V_1 \cos \phi + V_2 \cos 2\phi + V_3 \cos 3\phi] \\
 U_{b'\phi} &= \sum_{b'\phi} (b' - b'_0)[V_1 \cos \phi + V_2 \cos 2\phi + V_3 \cos 3\phi] \\
 U_{\theta\phi} &= \sum_{\theta\phi} (\theta - \theta_0)[V_1 \cos \phi + V_2 \cos 2\phi + V_3 \cos 3\phi] \\
 U_{\phi\theta\theta'} &= \sum_{\phi\theta\theta'} K_{\phi\theta\theta'} \cos \phi (\theta - \theta_0)(\theta' - \theta'_0) \\
 U_{nb} &= \sum_{i>j} \left\{ \left[ \frac{A_{ij}}{r_{ij}^9} - \frac{B_{ij}}{r_{ij}^6} \right] + \frac{q_i q_j}{r_{ij}} \right\} f_{ij}
 \end{aligned}$$

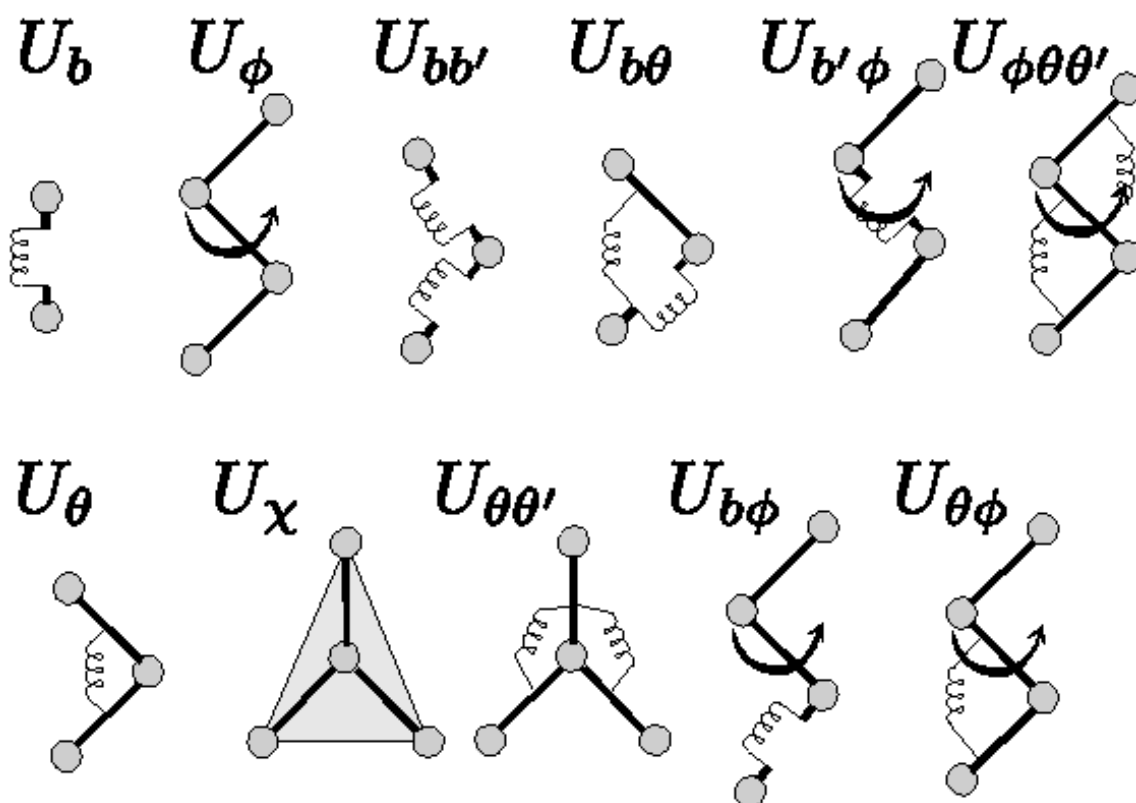


Рис. 2.1. Графическое пояснение к членам внутримолекулярной энергии уравнения 2.2.

## 2.2. Методы расчета свободной энергии

Предположим, что молекулярная система характеризуется некоторым Гамильтонианом, зависящим от параметра:  $H(\lambda)$ . Будем говорить, что каждому значению  $\lambda$  соответствует некоторое  $\lambda$ -состояние системы. Каким образом возможно рассчитать разницу в свободной энергии между двумя  $\lambda$ -состояниями?

Начиная с 1960-ых годов для практической реализации в численных экспериментах широкое применение получили такие методы, как *метод термодинамического интегрирования*, *метод возмущения свободной энергии*, *метод Беннетта*. С появлением работ Джарзинского, Крукса, Шёртса и др. [62–64] выявилась более глубокая связь между этими методами. Соотношения неравновесной статистической физики, выведенные Джарзинским (Jarzinsky equality) и Круксом (transient fluctuation theorem) предлагают более общий подход к решению задачи отыскания свободной энергии через неравновесные процессы. При этом, упомянутые ранее методы являются частными случаями в этом общем подходе, но могут также быть получены независимыми способами. В данной работе нам будут интересны эти три метода, так как они имеют широкое практическое применение.



### 2.2.1. Метод термодинамического интегрирования

Метод ТИ может быть получен из простых соображений. Свободная энергия Гиббса системы даётся в общем случае выражением 2.3.

$$G = -kT \ln \int \exp \left( -\frac{H(V, \mathbf{p}, \mathbf{q}, \lambda) + PV}{kT} \right) d\mathbf{p}d\mathbf{q}dV \quad (2.3)$$

Для метода ТИ основным является равенство 2.4.

$$\frac{\partial G}{\partial \lambda} = \left\langle \frac{\partial H}{\partial \lambda} \right\rangle \quad (2.4)$$

И разница свободной энергии между разными  $\lambda$ -состояниями даётся выражением 2.5.

$$\Delta G = \int \frac{\partial G}{\partial \lambda} d\lambda \quad (2.5)$$

Метод достаточно эффективен и часто применяется. Однако, необходимость интегрирования функции  $\frac{\partial G}{\partial \lambda}$  требует достаточно большого количества вычислений и вносит дополнительную ошибку численного интегрирования, которую трудно оценить при ограниченном наборе известных значений функции.

### 2.2.2. Метод возмущения свободной энергии

В методе возмущения свободной энергии (Free energy perturbation - FEP) разница свободной энергии между разными  $\lambda$ -состояниями может быть записана в виде 2.6, что и определяет метод отыскания разницы свободной энергии между двумя состояниями.

$$\Delta G = -kT \ln \frac{\int dV d\mathbf{p}d\mathbf{q} \exp \left\{ -\frac{H(V, \mathbf{p}, \mathbf{q}, \lambda_1) + PV}{kT} \right\}}{\int dV d\mathbf{p}d\mathbf{q} \exp \left\{ -\frac{H(V, \mathbf{p}, \mathbf{q}, \lambda_0) + PV}{kT} \right\}} = -kT \ln \left\langle \exp \frac{H_0 - H_1}{kT} \right\rangle_{\lambda_0} \quad (2.6)$$

где  $H_0 = H(V, \mathbf{p}, \mathbf{q}, \lambda_0)$ ,  $H_1 = H(V, \mathbf{p}, \mathbf{q}, \lambda_1)$ , усреднение ведётся по NPT-ансамблю соответствующему состоянию системы с  $\lambda = \lambda_0$ .

### 2.2.3. Метод Беннетта

Метод Беннетта (Bennett acceptance ratio method) позволяет наиболее эффективным образом комбинировать данные, получаемые при прямом и обратном процессе. Иными словами, мы переводим систему из состояния с  $\lambda = \lambda_0$  в состояние с  $\lambda = \lambda_1$  и при этом получаем некоторое значение работы  $W_F$ . Однако возможен и обратный процесс перевода системы из состояния  $\lambda_1$  в состояние  $\lambda_0$  с работой  $W_R$ . В 1976 году метод был предложен Беннеттом [65] как обобщение метода FEP. Однако в настоящее время стало ясно, что таким образом можно обобщить практически любые методы расчёта свободной энергии.

Этот *обобщённый метод Беннета* основывается на соотношениях Крукса и позволяет построить наилучшую статистику (оценочную функцию максимального правдоподобия) для оценки значения свободной энергии по наборам параметров  $W_F$  и  $W_R$  (см. [64]).

В простейшем случае как обобщение метода FEP метод Беннета (оригинальная работа [65], 1976 год) имеет следующий вид. Итак, допустим, что у нас есть набор работ по бесконечно быстрой трансформации системы из состояния  $\lambda_0$  в состояние  $\lambda_1$  -  $W_F$ , и обратно -  $W_R$ . Метод Беннета задаёт оценочную функцию, которая определяет значение разницы свободной энергии с минимальной дисперсией среди всех оценочных функций. Оценочная функция задаётся неявно, как решение уравнения 2.7 по параметру  $C$ .

$$\sum_m^{N_R} f(W_R^m + C) = \sum_n^{N_F} f(W_F^n - C) \quad (2.7)$$

В уравнении (2.7)  $f$ -функция Ферми-Дирака:

$$f(x) \equiv \frac{1}{1 + \exp \beta x}$$

$N_F$  и  $N_R$  - соответственно количество полученных замеров в прямом и обратном направлениях,  $\beta = 1/kT$ . Можно показать, что наилучший результат всегда достигается при  $N_F \approx N_R$ . Более того метод Беннета имеет преимущество перед простым методом FEP. Так, более эффективным является осуществление выборки значений работы в прямом и обратном направлениях, и затем применение метода Беннета, чем осуществление в два раза большей выборки в одном направлении и применении метода FEP.

#### 2.2.4. Методы расширенных ансамблей

Кроме трёх вышеприведённых методов развиваются и более современные методы расширенных ансамблей. В этих методах происходит одновременное моделирование систем при различных значениях параметра  $\lambda$  и между этими состояниями осуществляются переходы, как в методе Монте-Карло. Хорошим примером адаптивного метода расширенных ансамблей является работа [66]. Однако, на данный момент программное обеспечение, реализующее такие методы, мало распространено.

### 2.3. Молекулярные поверхности

Поверхность доступная растворителю (ПДР) (solvent accessible surface area (SASA)) – это поверхность биомолекулы, которая доступна растворителю. Обычно ПДР выражают с квадратных ангстремах. ПДР была впервые описана Ли и Ричардсом [67] и иногда

называется молекулярной поверхностью Ли-Ричардса, она определяется как поверхность которую описывает центр пробной сферы при “обкатывании” сферы вокруг молекулы. Размер сферы обычно берётся порядка размера молекулы воды. В этом плане ПДР можно воспринимать, как расширенную поверхность Ван-дер-Ваальса.

Кроме ПДР широкое распространение получила поверхность Коннолли, которая представляет собой как бы полость, которая формируется в растворителе при помещении в него молекулы. Коннолли разработал также полуаналитический метод расчёта такой поверхности [68].

На Рис. 2.2 изображены поверхности доступные растворителю и молекулярные поверхности Коннолли для модельной системы.

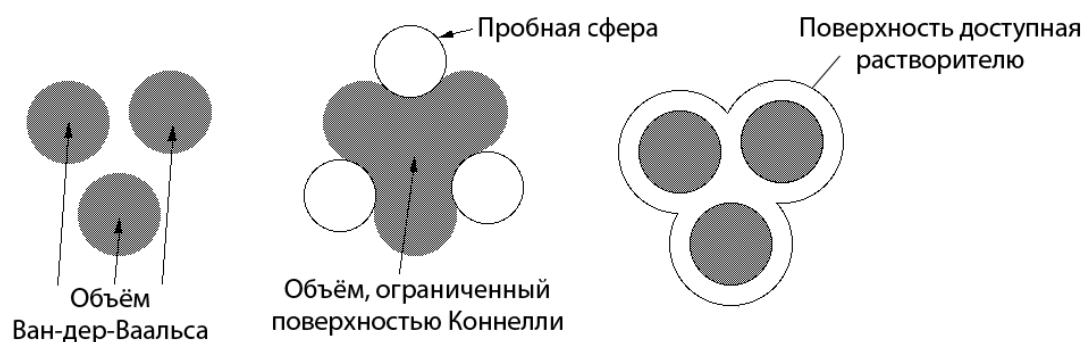


Рис. 2.2. Пояснения к определению поверхности доступной растворителю и поверхности Коннелли.

# Изучение гидратации аминокислот у поверхности воды

В данной главе речь пойдёт о изучении гидратации боковых цепей аминокислот вблизи поверхности воды путём построения профилей свободной энергии на основании классических атомистических моделей молекул. Работа также имеет методологическое значение, поскольку в ней впервые предпринята попытка высокоточной численной оценки термодинамических параметров гидратации и адсорбции. Результаты работы опубликованы полностью в статье [1].

## 3.1. Введение

Применение молекулярной динамики к анализу поверхностных явлений нашло широкое применение за последние 20 лет поскольку позволяет дополнять экспериментальные данные и получать новой информацию на молекулярном уровне [69, 70]. Величины, которые зачастую интересуют исследователя имеют статистическую природу (например, энергии связывания, коэффициенты распределения и т.д.), вычисление таких величин в компьютерном эксперименте требует серьёзной статистической выборки и значительных вычислительных ресурсов, одновременно накладывая ограничения на размеры и сложность молекулярных систем.

Методология вычисления свободных энергий (особенно с использованием стандартных силовых полей для белковых структур) получила много внимания за последнее десятилетие [71], многие работы были посвящены систематической и критической оценке и сравнению методологически, силовых полей, протоколов молекулярной динамики [72–80]. Однако эти работы в основном сводились к расчёту параметров в гомогенных системах (например, расчёту свободной энергии сольватации небольших молекул), в основном в воде, и в них не затрагивались вопросы расчёта систем с границей раздела фаз. Так было определено, что энергии гидратации боковых цепей аминокислот рассчитанные методом МД находятся в хорошей корреляции с экспериментальными данными, но систематически завышены. Разумным продолжением ряда этих работ является рассмотрение взаимодействия таких молекул не только с объёмом растворителя, но и с его границей. Если говорить о высокоточных оценках свободной энергии, то речь можно вести о достижении точности серьёзно выше 2 – 4 кДж/моль, что является пределом точности, например, при предсказании связывания лигандов и белков при разработке лекарств. Если речь идёт не о качественных, а

о количественных предсказаниях, то необходимая точность порядка 1 кДж/моль. Если же говорить, о настройке силовых полей на базе этих данных, то желаемая точность может быть около 0.1 кДж/моль.

## 3.2. Постановка задачи

Задачей данного исследования явился расчёт профилей свободной энергии для соединений аналогов нейтральных боковых цепей аминокислот на границе вода/пар с высокой точностью, вычисление энергий гидратации и адсорбции. В данной работе мы попытались обобщить и разработать наиболее точные и проверенные технологии таких расчётов, чтобы исключить зависимость результатов от методики расчётов. Для оценки качества результатов был проведён отдельный расчёт энергий гидратации молекул в объёме с использованием алгоритма Беннетта (Bennett acceptance ratio algorithm). Для изучения были выбраны широко используемые модели соединений в силовом поле OPLS-AA [81] с моделью воды SPC.

## 3.3. Методы расчёта

Для расчёта профилей использовался классический метод молекулярной динамики в комбинации с методом удерживающей силы для расчёта профилей свободной энергии. Для расчёта энергий гидратации в объёме использовался метод Беннетта. В данном разделе описывается модель силового поля, протоколы МД, а также методика расчёта. Более подробное описание нюансов протоколов и обоснование их выбора приведено в статье [1].

### 3.3.1. Силовое поле и модель

Для создания модели было выбрано полноатомное силовое поле OPLS-AA [81] и твёрдая (с фиксированными внутренними степенями свободы) модель воды SPC [82]. Основные характеристики модели воды представлены в таблице 3.1. Атомы представляются как классические частицы, соединённые связями. Невалентные взаимодействия представлены потенциалами Леннард-Джонса и Кулона:

$$U^{nb} = \sum_{i < j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i < j} \frac{q_i q_j}{4\pi r_{ij}^2} \quad (3.1)$$

Валентные взаимодействия включали угловые, торсионные и ложно торсионные члены. Параметры потенциала Леннард-Джонса и заряды для соединений были взяты из ра-

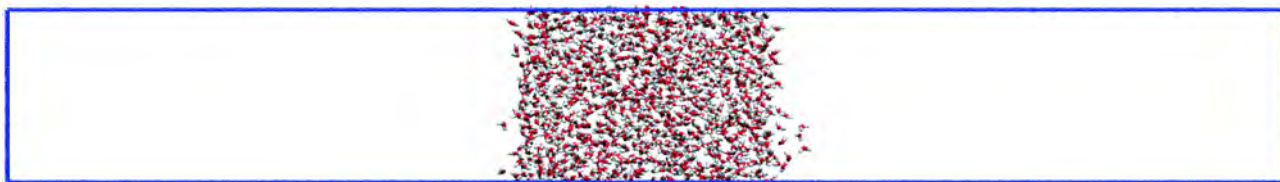


Рис. 3.1. Мгновенное изображение модели системы с границей раздела вода/пар в периодической ячейке. Наиболее длинная сторона ячейки совпадает с нормалью к границе раздела (ось Z).

боты [79]. В таблице 3.2 представлены 13 нейтральных аминокислот и соответствующие им соединения аналоги боковых цепей, которые изучались в данной работе.

### 3.3.2. Расчёт профилей свободной энергии

Для расчёта профиля свободной энергии соединений на границе раздела фаз вода/пар была сконструирована модель границы раздела, см. Рис. 3.1. Прямоугольный слой воды толщиной 4 нм помещён в расчётную ячейку ( $3 \times 3 \times 20 \text{ нм}^3$ ), чьи размеры постоянны в ходе моделирования. Количество молекул воды - 1173. Расчёты проводились в пакете GROMACS v. 3.3.1 [83]. Для расчёта профиля свободной энергии использовался метод сдерживающей силы. С помощью алгоритма поддерживалась фиксированная проекция расстояния между центрами масс молекулы и водного слоя на ось, перпендикулярную к поверхности. На каждом шаге измерялась сдерживающая сила, которую необходимо было прикладывать, чтобы это расстояние оставалось фиксированным. Для этого использовался, так называемый, GROMACS pull code, который поддерживал проекцию расстояния постоянной с использованием алгоритма SHAKE [84]. Исследуемая молекула помещалась на различных расстояниях от центра масс водного слоя, и для каждой такой системы после уравнивания проводился расчёт длиной в 4 нс. Для каждой молекулы было рассчитано 80 точек: положение от -4 нм до 4 нм с шагом 0.1 нм. Мгновенная сила записывалась на каждом шаге, затем усреднялась. Профиль средней силы затем аппроксимировался кубическим сплайном, учитывая симметрию профиля относительно  $z=0$  (см. Рис. 3.4). Интегрированием профиля средней силы вычислялся профиль свободной энергии. Соответствующие теоретические формулы приведены в приложении В.2.

### 3.3.3. Расчёт энергий гидратации

Энергии гидратации рассчитывались по методу Беннетта [85] путём постепенного “отключения” взаимодействий между молекулой растворённого вещества и растворителем в ячейке, содержащей около 900 молекул воды. Теоретические основы и детали метода

$\sigma$ , Å	$\epsilon$ , кДж/моль	$q_o$ , e	$q_h$ , e	$l_{OH}$ , Å	$\theta^\circ$	P, D
3.166	0.65	-0.82	0.41	1.0	109.47	2.27

Таблица 3.1. Основные характеристики модели воды SPC, включая параметры Леннард-Джонса  $\sigma, \epsilon$ , частичные заряды на кислороде ( $q_o$ ) и водороде ( $q_h$ ), длины связей, НОН угол ( $\theta^\circ$ ) и дипольный момент ( $P$ ).

Аминокислота	Вещество аналог боковой цепи	Химическая формула	Сокр. обозначение
аланин	метан	$\text{CH}_4$	ALA'
валин	пропан	$\text{C}_3\text{H}_8$	VAL'
лейцин	изобутан	$\text{C}_4\text{H}_{10}$	LEU'
изолейцин	н-бутан	$\text{C}_4\text{H}_{10}$	ILE'
цистеин	метанэтиол	$\text{CH}_3\text{SH}$	CYS'
метионин	метилэтилсульфид	$\text{C}_2\text{H}_5\text{SCH}_3$	MET'
серин	метанол	$\text{CH}_3\text{OH}$	SER'
треонин	этанол	$\text{C}_2\text{H}_5\text{OH}$	THR'
тирозин	п-крезол	$\text{C}_3\text{C}_6\text{H}_4\text{OH}$	TYR'
фенилаланин	толуол	$\text{C}_6\text{H}_5\text{CH}_3$	PHE'
триптофан	3-метилиндол	$\text{C}_6\text{H}_4\text{C}_2\text{HNNHCH}_3$	TRP'
аспарагин	ацетамид	$\text{CH}_3\text{CONH}_2$	ASN'
глутамин	пропионамид	$\text{C}_2\text{H}_5\text{CONH}_2$	GLN'

Таблица 3.2. Список 13 изученных нейтральных веществ - аналогов боковых цепей аминокислот, из названия, химические формулы и обозначения.

Беннетта для оценок энергии гидратации могут быть найдены в работах других исследователей [74, 78, 86]. Основные статистические соотношения подчёркнуты в приложении В.2.

Здесь мы приведём основные параметры нашего расчёта. Использовался стандартный термодинамический цикл: сначала молекула постепенно “отключалась” от среды путём зануления зарядов и параметров потенциала Леннарда-Джонса, а затем проводился аналогичный коррекционный расчёт в вакууме, чтобы измерить поправку на внутримолекулярные взаимодействия. Отключение (мутация) молекулы в обоих случаях проводилась в 10 шагов. Сначала занулялись заряды, а потом отключались взаимодействия Леннарда-Джонса. Электростатика отключалась в 2 шага, сначала заряды уменьшались в  $\sqrt{2}$  раз, а затем занулялись. Взаимодействия Леннарда-Джонса отключались в 8 шагов согласно формуле:

$$U_S^{LJ} = \sum_j \sum_i f(i, j) \lambda 4 \epsilon_{ij} \times \left( \frac{\sigma_{ij}^{12}}{(\alpha \sigma_{SC}^6 (1 - \lambda) + r_{ij}^6)^2} - \frac{\sigma_{ij}^6}{\alpha \sigma_{SC}^6 (1 - \lambda) + r_{ij}^6} \right) \quad (3.2)$$

где  $i$  пробегает по всем атомам растворённого вещества,  $j$  по всем атомам в системе,  $f(i, j)$  - ответственна за исключение взаимодействия между атомами связанными валентно.  $\alpha = 0.5, \sigma_{SC} = 0.3$ . В такой форме этапы мутации соответствовали параметрам  $\lambda = 1, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0$ . Для каждой стадии осуществлялись расчёты длиной 5 нс. Конфигурации записывались каждые 2 пс. Эти конфигурации обрабатывались по алгоритму Беннетта, чтобы получить оценку разницы свободной энергии между смежными состояниями.

### 3.3.4. МД протокол

В качестве алгоритма для получения выборок мы использовали динамику Ланжевена, согласно уравнению:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = -m_i \xi_i \frac{d\mathbf{r}_i}{dt} + \mathbf{F}_i(\mathbf{r}_i) + \hat{\mathbf{r}}_i \quad (3.3)$$

где  $\xi_i$  константа трения, а  $\hat{\mathbf{r}}_i(t)$  шумовой процесс  $\langle \hat{\mathbf{r}}_i(t) \hat{\mathbf{r}}_j(t+s) \rangle = 2m_i \xi_i kT \delta(s) \delta_{ij}$ . Ланжевенова динамика является точным методом, позволяющим реализовать изотермический ансамбль.

Использовался пакет GROMACS v 3.3.1 [83], скомпилированный с двойной точностью. Уравнения Ланжевена интегрировались с использованием лип-фрэг алгоритма третьего порядка [87] с шагом 2 фс. Для фиксации длин связей использовались алгоритмы



LINCS [88] и SETTLE [89]. Расчёты проводились при постоянной температуре 300 К. Константа трения  $\xi$  для каждого атома рассчитывалась как  $\xi_i = m_i/\tau_T$ , где  $\tau_T = 1$  пс,  $m_i$  – масса атома. Для расчётов энергии гидратации использовался алгоритм задания изобарического ансамбля Паринелло-Рамана [90] при 1 атм, время релаксации 2 пс, сжимаемость  $4.5 * 10^{-5}$  бар $^{-1}$ . Для расчёта электростатических взаимодействий использовался метод PME [91, 92] с радиусом обрезки в прямом пространстве 1.3 нм и PME-порядком 6. Максимальное расстояние в пространстве фурье (maximum Fourier spacing) задавалось равным 0.12 нм. Относительная толерантность между вкладами в энергию прямого и обратного пространств  $10^{-5}$ . Для потенциала Леннарда-Джонса использовалась свитч функция, сглаживающая потенциал между 1.0 и 1.2 нм. Коррекционная поправка (long-range correction) учитывалась при вычислении энергии и давления.

## 3.4. Результаты и обсуждение

### 3.4.1. Свойства водного слоя

Чтобы оценить правильность нашей модели, рассмотрим сначала свойства модели границы раздела фаз. Расчёт длиной в 50 нс был проведён для того, чтобы оценить профиль плотности, поверхностный потенциал и ориентационное поведение молекул воды. Профиль плотности идеально описывается уравнением В.1 с параметрами  $\sigma = 0.14$  нм,  $\rho_g = 0.2$  кг/м $^3$  и  $\rho_L = 970.5$  кг/м $^3$ , поверхность раздела Гиббса находится в точках  $z = \pm 2.007$  нм. Толщина границы, рассчитанная из теории капиллярных волн  $\sigma_{CWT} = 1.45$  Å, хорошо согласуется с данными моделирования ( $\sigma = 0.14$  нм).

Чтобы проанализировать ориентационное упорядочение молекул на границе мы построили средний косинус угла с нормалью и параметр порядка  $P_2(\cos \theta) = \langle (3 \cos^2 \theta - 1)/2 \rangle$  как функцию  $Z$  (Рис. 3.2). Видно, что дипольные моменты предпочтительно ориентированы вдоль границы раздела. Параметр порядка  $P_2(\cos \theta)$  зануляется уже на расстоянии 0.5 нм за поверхностью Гиббса. Также небольшая ориентация диполей по нормали приводит к образованию заряженного бислоя на поверхности и соответствующего среднего электрического поля (Рис. 3.3), которое достигает значений  $1.7 * 10^9$  В/м в узкой области 0.7 нм около поверхности Гиббса. Поверхностное натяжение было рассчитано на основе диагональных компонент тензора давлений по формуле  $\gamma_0 = \frac{L_z}{2}(P_{zz} - 0.5(P_{xx} + P_{yy}))$ . Получена величина  $\gamma_0 = 51.2 \pm 0.8$  мН/м. Если добавить к этой величине аналитическую поправку возникающую из-за обрезки потенциала Леннарда-Джонса, то получится значение  $\gamma = 55.4$  мН/м, это значение занижено по сравнению с экспериментальным (71.6 мН/м

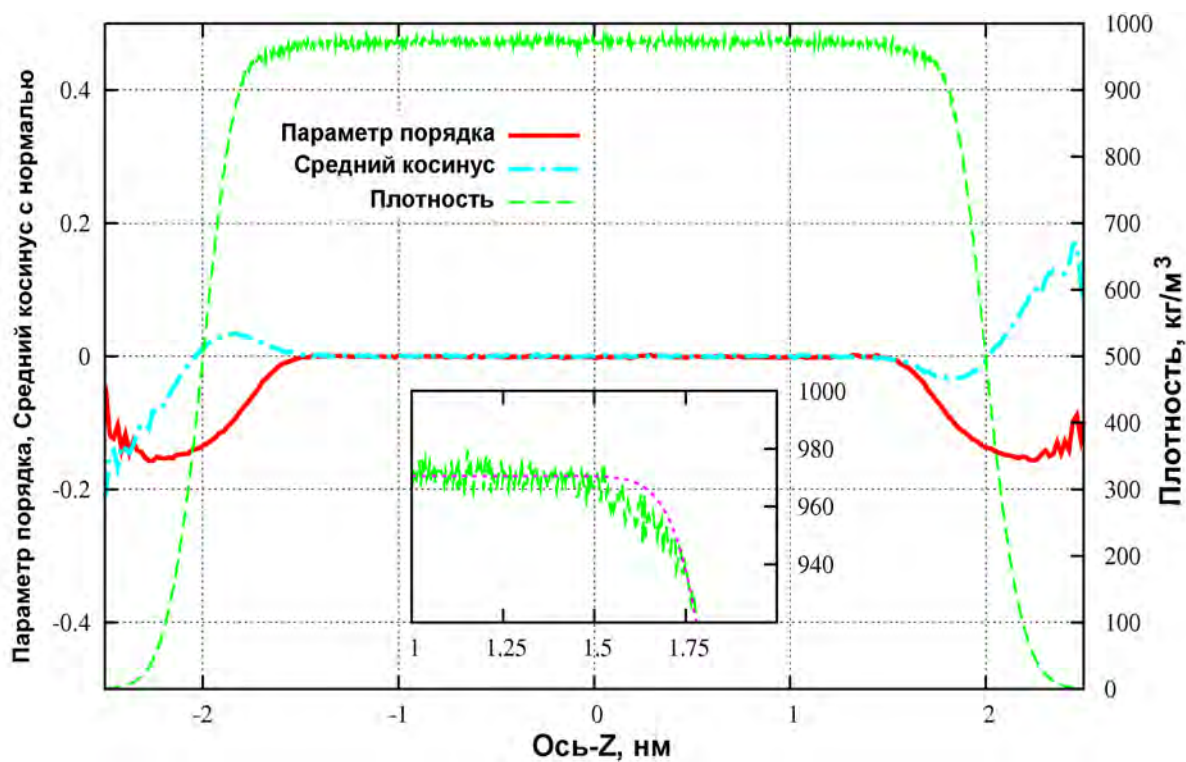


Рис. 3.2. Профиль плотности водного слоя (пунктирная линия) и средние ориентационные характеристики дипольного момента молекул воды: средний косинус угла с нормалью (пунктирно-точечная линия), параметр порядка  $P_2(\cos(\theta))$  (сплошная линия).

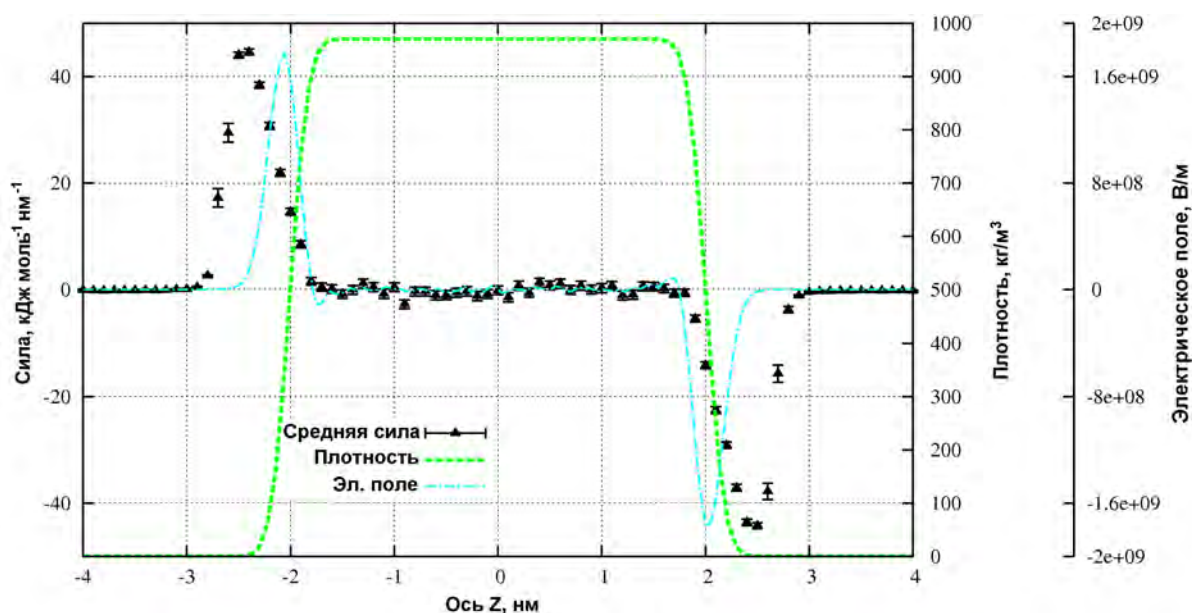


Рис. 3.3. Профиль среднего электрического поля (пунктирно-точечная линия), профиль средней силы для одной молекулы воды (треугольники с погрешностями) как функция Z. Профиль плотности дан для сравнения (пунктирная линия).

при 300K [93] ), но находится в хорошем согласии со значениями для модели SPC [94]. Также на примере молекулы воды была опробована техника расчёта профилей средней силы, на Рис. 3.3 приведён такой профиль. Энергия гидратации молекулы воды, рассчитанная из такого профиля, составляет  $-25.5 \pm 0.7$  кДж/моль.

### 3.4.2. Профили свободной энергии

Основной результат данного исследования – это набор высокоточных профилей средней силы и свободной энергии для аналогов боковых цепей аминокислот на границе вода/пар. Типичные профили средней силы, свободной энергии и концентрации исследуемых молекул представлены на Рис. 3.4. Данные для всех изученных молекул представлены в приложении Г. Симметричный вид профилей и малые значения статистических погрешностей свидетельствуют о достаточности проведённой выборки. Форма профилей средней силы похожа для всех молекул.

Рассмотрим более детально процесс сольватации с точки зрения полученных профилей. При расстоянии более 1 нм от поверхности (здесь и далее под положением поверхности понимается положение поверхности Гиббса) нет заметного взаимодействия между молекулой и водным слоем. При дальнейшем приближении к водному слою появляется притягивающая сила, которая, очевидно, обусловлена дисперсионными и электростатическими взаимодействиями. Сила притяжения достигает своего максимума приблизительно на расстоянии 0.5 нм до поверхности Гиббса и далее сходит на нет около границы раздела. Эта область, в которой действует сила притяжения, единственная вносит отрицательный (способствующий гидратации) вклад в энергию гидратации. В таблице 3.3 приведены параметры профилей средней силы для всех изученных молекул. Видно, что для молекул содержащих полярные группы (например, THR', ASN', GLN') максимальные значения силы притяжения намного выше, чем для чисто гидрофобных молекул. Приблизительно в районе поверхности Гиббса средняя сила обращается в ноль, что соответствует адсорбционному минимуму на графиках свободной энергии (см. Рис. 3.4). Визуальное рассмотрение (Рис. 3.5) показывает, что максимум силы притяжения всё ещё соответствует достаточному удалению молекулы от молекул воды, хотя уже заметно некоторое возмущение водного слоя. Когда молекула продвигается далее внутрь водного слоя, возникает выталкивающая сила, которая препятствует дальнейшему погружению молекулы в воду. Причина этой силы, очевидно, может быть приписана проявлению гидрофобного эффекта: молекула вещества нарушает сеть водородных связей молекул воды, что приводит к потере энтропии молекул воды из-за сниженной подвижности последних в округе сольватируемой моле-

кулы. Нужно отметить, что эта сила выталкивания присутствует для всех веществ (см. Таблицу 3.3), точнее во всех случаях сила выталкивания оказывается сильнее силы притяжения в области сразу за поверхностью Гиббса. Сила выталкивания пропадает при погружении молекулы на глубину 0.5-0.7 нм в зависимости от размера молекулы. После этого начинается область по своим характеристикам, неотличимая от объёма, где средняя сила взаимодействия равна нулю. Для некоторых молекул (PHE', ILE', VAL') есть подозрение, что после области выталкивания существует опять небольшая область притяжения длиной 0.2-0.3 нм, но статистическая достоверность этого утверждения не велика. Однако присутствие такой области приведёт к возникновению ещё одного максимума на графике свободной энергии, который иногда называют барьером десольватации, который кинетически затрудняет выход растворённого вещества из растворителя и обратно. Тот факт, что средняя сила обращается в ноль уже на расстоянии 0.5-0,7 нм за поверхностью раздела, говорит о том, что на этом расстоянии уже практически отсутствуют поверхностные эффекты влияющие на сольватацию, то есть приблизительно слой толщиной в две молекулы воды экранирует поверхность от молекулы.

### 3.4.3. Количественные результаты

В данном разделе обсудим численные результаты расчётов энергий гидратации и адсорбции. Энергии гидратации были определены двумя способами: как значения профилей свободной энергии в середине слоя, и с помощью метода Беннетта (BAR) из расчётов в объёме. Последние результаты, несомненно, более надёжные, поскольку мало зависят от конечномерных эффектов, дают адекватную оценку статистической погрешности, а также сам метод является методологически хорошо изученным. Поэтому результаты этого метода будут использоваться для критической оценки результатов расчёта профилей свободной энергии. Результаты обоих методов, а также экспериментальные значения приведены в таблице 3.4.

Результаты полученные методом BAR находятся в очень хорошем согласии с другими работами [78]. Обсуждения совпадения и отличий данных моделирования и эксперимента может быть найдено в работе [78]: основной вывод состоит в том, что корреляция результатов очень хорошая (в нашем случае коэффициент корреляции равен 0.997), однако абсолютные значения оказываются на 2-4 кДж/моль выше, чем экспериментальные.

Из таблицы 3.4 видно, что статистические погрешности значений рассчитанных из профилей свободной энергии несколько выше (около 1 кДж/моль). Видно, что хотя значения энергий для отдельных веществ (5 из 13) в обоих методах согласуются в пределах

	$F_{at}^{max}, \text{кДж моль}^{-1} \text{нм}^{-1}$	$F_{ex}^{max}, \text{кДж моль}^{-1} \text{нм}^{-1}$
ALA'	5.0	25.8
VAL'	10.8	33.1
LEU'	13.0	36.2
ILE'	13.1	34.9
CYS'	19.3	21.7
MET'	25.3	28.5
SER'	41.7	12.3
THR'	44.9	18.4
TYR'	51.3	27.2
PHE'	29.2	26.5
TRP'	51.2	26.5
ASN'	57.6	13.0
GLN'	59.9	20.2

Таблица 3.3. Максимальные значения средней силы притяжения и отталкивания, действующей на изучаемые молекулы по мере их прохождения через водный слой.

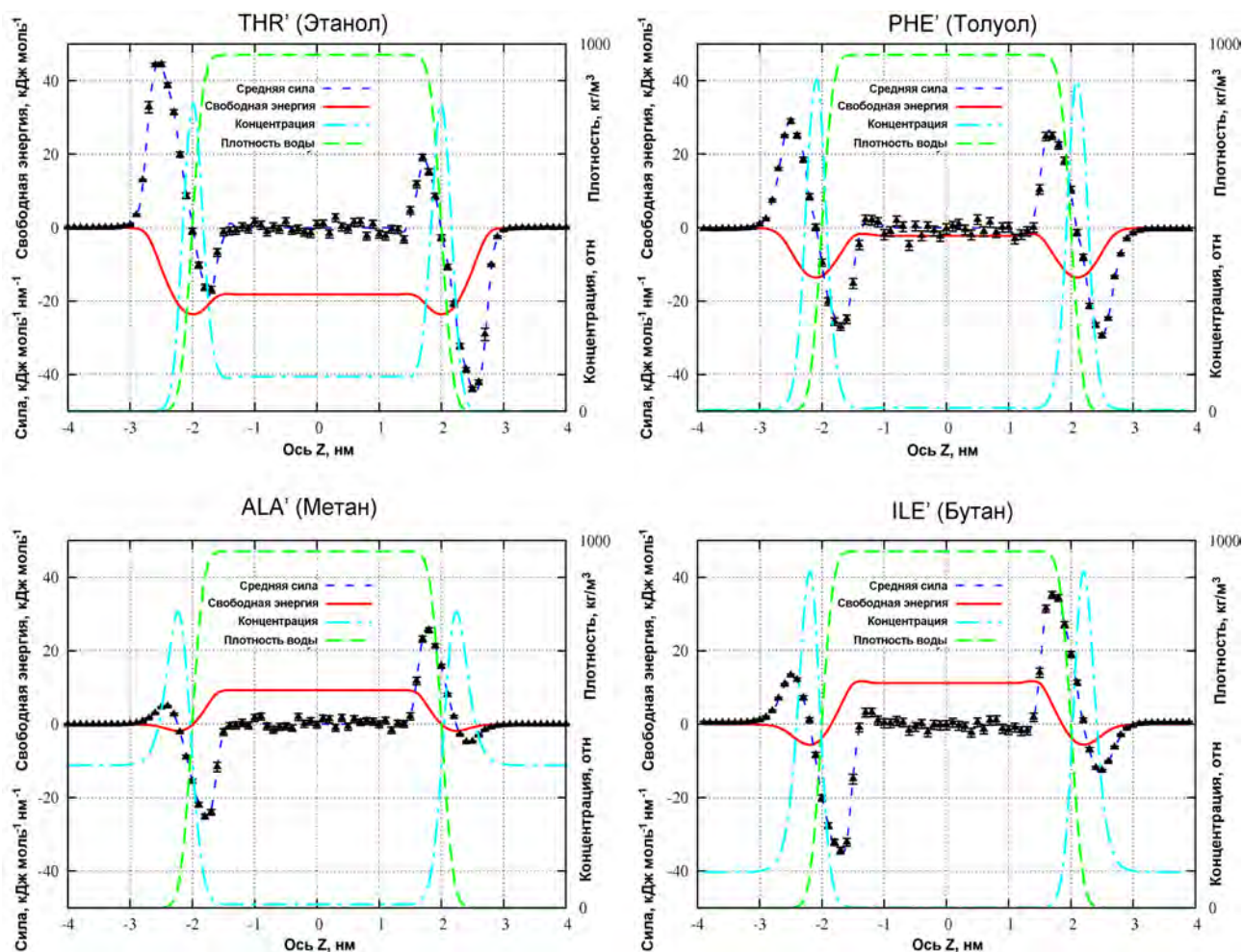


Рис. 3.4. Профили свободной энергии для нескольких изученных молекул. Пунктирные линии и точки с отмеченными погрешностями – профиль средней силы и его сплайн аппроксимация (погрешности отмечены как  $\pm$  одно стандартное отклонение). Сплошная линия - профиль свободной энергии, пунктирно-точечная линия - относительный профиль концентрации, рассчитанный по профилю свободной энергии. Для сравнения пунктирной линией отмечен профиль плотности водного слоя.

Таблица 3.4. Энергии гидратации боковых цепей аминокислот, полученные различными методами (методом Беннетта и по профилям свободной энергии), и экспериментальные результаты. Для метода Беннетта статистическая погрешность приведена как одно стандартное отклонение.

	$G_{BAR}^{hydr}, \text{кДж моль}^{-1}$	$G_{PMF}^{hydr}, \text{кДж моль}^{-1}$	$G_{exp}^{hydr}, \text{кДж моль}^{-1a}$
ALA'	$9.20 \pm 0.04$	$9.3 \pm 0.6^b$	8.12
VAL'	$10.1 \pm 0.1$	$10.9 \pm 0.7$	8.33
LEU'	$10.57 \pm 0.08$	$11.3 \pm 0.8$	9.54
ILE'	$10.5 \pm 0.3$	$11.7 \pm 0.7$	9.00
CYS'	$-2.24 \pm 0.06$	$-1.2 \pm 0.6$	-5.19
MET'	$-1.74 \pm 0.18$	$-0.5 \pm 0.7$	-6.19
SER'	$-19.2 \pm 0.1$	$-18.9 \pm 0.6$	-21.17
THR'	$-18.65 \pm 0.13$	$-18.1 \pm 0.7$	-20.42
TYR'	$-22.33 \pm 0.18$	$-20.5 \pm 0.9$	-25.56
PHE'	$-3.26 \pm 0.11$	$-1.6 \pm 0.8$	-3.18
TRP'	$-18.7 \pm 0.3$	$-15.9 \pm 0.9$	-24.60
ASN'	$-35.69 \pm 0.10$	$-33.6 \pm 0.9$	-40.50
GLN'	$-36.45 \pm 0.16$	$-32.7 \pm 0.9$	-39.26

<sup>a</sup> Экспериментальные значения из работы [95].

<sup>b</sup> Статистические погрешности не включают какие-либо систематические погрешности или погрешности силового поля и методологии.



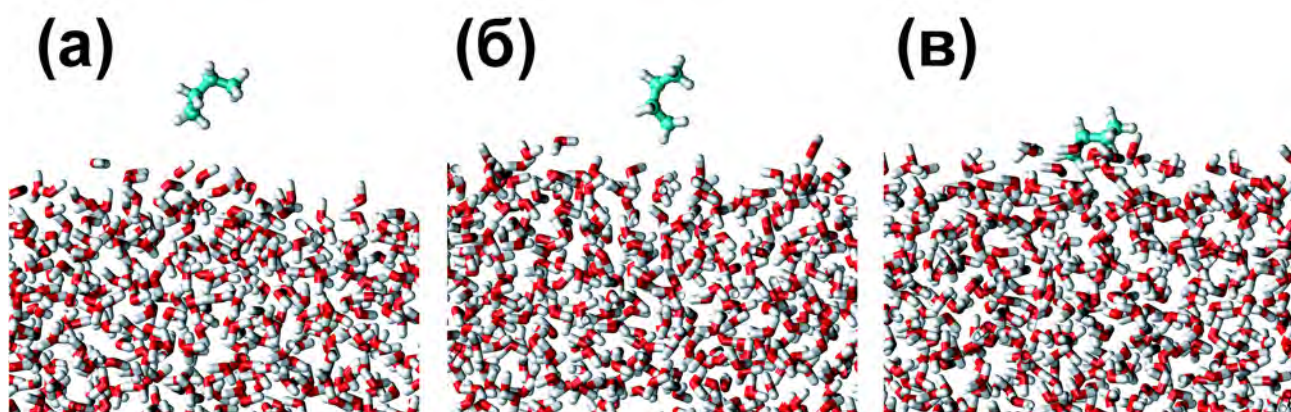


Рис. 3.5. Мгновенные изображения молекулы бутана во время расчётов на различных расстояниях от границы раздела. Положения соответствующие: (а) максимальной силе притяжения, (б) адсорбционному минимуму, (с) максимальной выталкивающей силе.

статистической погрешности, все значения полученные путём интегрирования профиля средней силы завышены относительно “настоящих” модельных значений. Несостыковки ранжируются от 0 для ALA’ до 4 кДж/моль для GLN’. Причина этих расхождений, на наш взгляд, кроется в (i) невозможности интегрирования хвоста профиля из-за наличия статистической ошибки в значениях силы, (ii) конечном размере водного слоя и (iii) обрезки потенциала Леннарда-Джонса при невозможности применения аналитической поправки (что возможно при расчётах в объёме методом BAR). Приблизительная оценка теряемой энергии при обрезке потенциала Леннарда-Джонса в нашем случае будет варьироваться от 0.4 кДж/моль для ALA’ до 3 кДж/моль для TRP’. Применение этих поправок уже сильно бы улучшило совпадения результатов. Однако пока алгоритмы учёта таких поправок для негомогенных систем не развиты.

Мы измерили энергии адсорбции двумя способами: используя подход Гиббса и путём измерения значения профиля энергии в минимуме. В подходе Гиббса сначала рассчитывался относительный профиль концентрации, предполагая концентрацию в газовой фазе равной  $1 \text{ нм}^{-3}$ . Затем рассчитывался избыток Гиббса по формуле В.12 и энергия адсорбции оценивалась по формуле В.15, предполагая толщину границы раздела равной 1 нм, поскольку это обычно была толщина, на которой варьировался профиль свободной энергии. Оба набора полученных энергий (см. таблицу 3.5) находятся в отличной корреляции ( $R=0.9996$ ), хотя заметны некоторые отклонения для веществ с низкой поверхностной активностью, для последних, очевидно, метод имеет значение. Хотя значение энергии в минимуме оказалось хорошей оценкой энергии адсорбции, предположительно метод Гиббса является более универсальным и термодинамически корректным.

Значения энергии адсорбции Гиббса в таблице 3.5 представлены относительно фа-



Таблица 3.5. Сродство изученных молекул к поверхности, характеризующееся идеальными величинами избытка Гиббса (предполагая концентрацию молекул в фазе пара  $1 \text{ нм}^{-3}$ ), энергиями адсорбции Гиббса и глубиной энергии адсорбционного минимума. Последняя колонка представляет энергию адсорбции Гиббса по отношению к фазе максимального сродства.

	$\Gamma_{excess}, \text{нм}^{-2}$	$G_{Gibbs}^{ads}, \text{кДж моль}^{-1}$	$G_{min}^{ads}, \text{кДж моль}^{-1}$	$G_{aff}^{ads}, \text{кДж моль}^{-1}$
ALA'	0.544	1.52	-1.82	1.52
VAL'	2.3	-2.08	-4.45	-2.08
LEU'	3.8	-3.33	-5.64	-3.33
ILE'	3.7	-3.27	-5.59	-3.27
CYS'	14.0	-6.61	-8.76	-4.37
MET'	57.0	-10.1	-12.35	-8.36
SER'	$2.1 * 10^3$	-19.12	-22.03	0.08
THR'	$4.7 * 10^3$	-21.1	-23.61	-2.45
TYR'	$65.0 * 10^3$	-27.63	-30.05	-5.30
PHE'	95.0	-11.37	-13.59	-8.11
TRP'	$22.6 * 10^3$	-25.0	-27.47	-6.3
ASN'	$0.74 * 10^6$	-33.71	-37.03	1.99
GLN'	$2.30 * 10^6$	-36.52	-39.15	-0.02

зы максимального сродства (как обычно делается в экспериментальных работах). Можно сказать, что все вещества являются поверхностно активными в том плане, что в профилях свободной энергии у них есть адсорбционный минимум. Однако конченные значения энергии адсорбции всегда являются относительными поскольку зависят от определения стандартных состояний и условной толщины границы раздела.

К сожалению, в литературе встречается мало экспериментальных данных, с которыми можно было бы сравнить полученные численные данные. Бул и Бриз [96] ввели шкалу энергий адсорбции для боковых цепей аминокислот, руководствуясь данными по измерению поверхностного натяжения для цвиттерионных аминокислот и применяя предположение об аддитивности вкладов различных молекулярных групп. Однако, если предположение об аддитивности достаточно хорошо себя зарекомендовало при оценках энергий сольватации, то в случае измерения энергий адсорбции справедливость такого предположения вызывает сомнения. Коэффициент корреляции шкалы Була и Бриза с модельными результатами по адсорбции составляет 0.65.

Измерения избытка Гиббса для метанола и этанола были проведены с помощью нейтронного рассеяния [97, 98] и равны  $2.5 \cdot 10^{-10}$  моль\*см<sup>-2</sup> и  $(3-5) \cdot 10^{-10}$  моль\*см<sup>-2</sup>, соответственно, при объёмной концентрации в 0.0012 моль\*см<sup>-3</sup>. Эти данные находятся в разумном согласии с результатами нашей модели равными  $1.2 \cdot 10^{-10}$  моль\*см<sup>-2</sup> и  $2.6 \cdot 10^{-10}$  моль\*см<sup>-2</sup> для этого случая. По данным работы [99] на основании измерения поверхностного натяжения энергия адсорбции метанола из газовой фазы составляет  $-15.2 \pm 0.5$  кДж/моль, в то время как из наших расчётов следует величина -19.2. кДж/моль.

Следуя идеям статьи Охупкина и др. [100], нами были построены двумерные диаграммы сродства молекул к объёмной и поверхностной фазе воды (Рис. 3.6) по двум параметрам: энергии гидратации и энергии адсорбции относительно фазы максимального сродства. Видно, что изученные молекулы покрывает достаточно большую область энергий адсорбции и гидратации. Некоторые молекулы обладают сходными характеристиками, например, (i) PHE' и MET', (ii) VAL', ILE' и LEU', (iii) ASN' и GLN'. Однако, в общем можно заключить, что корреляции между энергиями адсорбции относительно фазы максимального сродства и гидратации в нашем случае не наблюдается, эти характеристики являются независимыми и в то же время достаточно простыми параметрами, описывающими взаимодействие растворителя с другими веществами, которые могут использоваться для улучшения и уточнения молекулярных моделей и силовых полей подобно тому, как для этого используются только энергии гидратации [101].

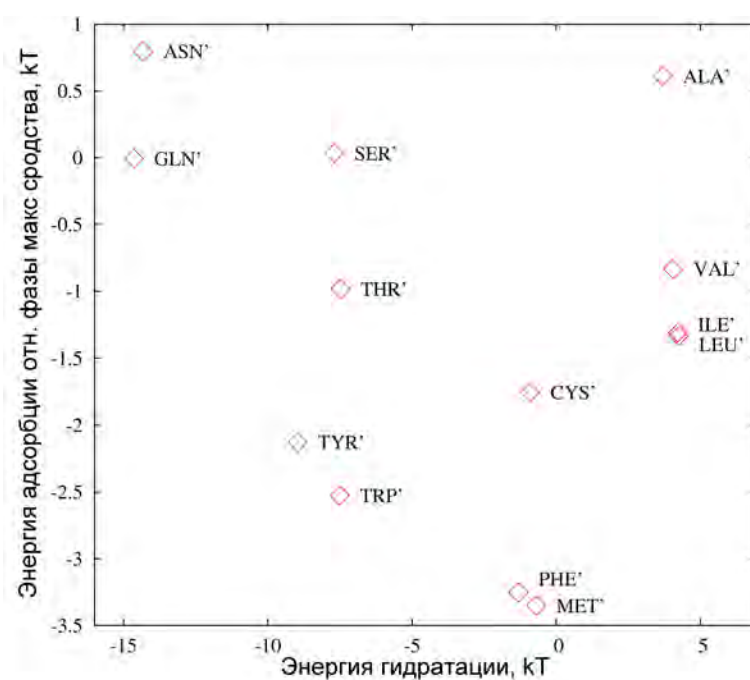


Рис. 3.6. Двумерная классификационная диаграмма для исследованных веществ в координатах энергия гидратации - энергия адсорбции относительно фазы максимального сродства.

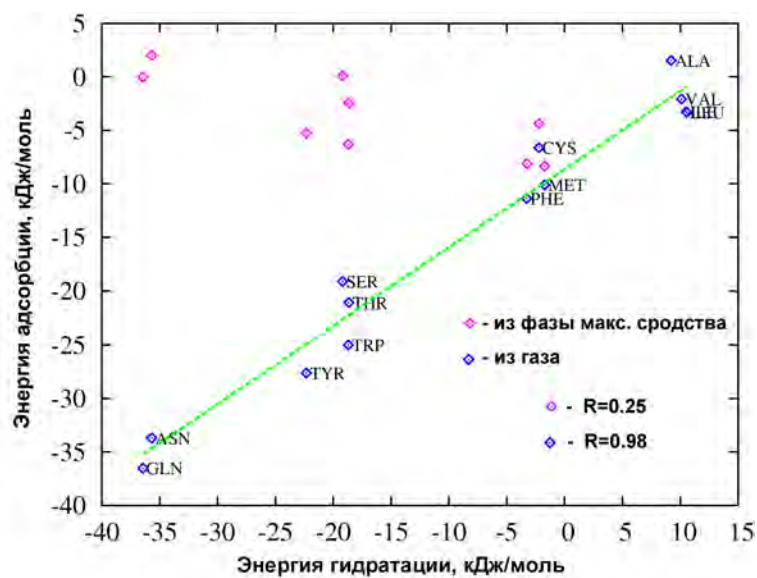


Рис. 3.7. График корреляций между энергиями гидратации и адсорбции.

Таблица 3.6. Коэффициенты корреляции энергий адсорбции ( $G_{Gibbs}^{ads}$ ), полученных в результате численного моделирования и экспериментальных шкал со статистическими данными, полученными в главе 4. Корреляции рассчитывались только для 13 незаряженных аминокислот.

	“10-20”	“50-60”	“95-105”	$G_{Gibbs}^{ads}$	$G_{affin}^{ads}$
$G_{Gibbs}^{ads}$	0.48	0.88	0.7	1	0.17
$G_{affin}^{ads}$	0.2	0.53	0.7	0.17	1
$V > W$	0.46	0.87	0.73	0.98	0.23
$CH > W$	0.19	0.91	0.92	0.81	0.56
$O > W$	0.09	0.76	0.88	0.48	0.77

#### 3.4.4. Корреляционный анализ

Поскольку на данный момент не существует систематических данных по экспериментальному измерению энергий адсорбции, изученных боковых цепей аминокислот, то представляется интересным провести анализ корреляций между полученными данными и статистическими шкалами предложенными в главе 4. Соответствующие коэффициенты корреляции, а также коэффициенты корреляции статистических шкал с экспериментальными, рассчитанные только для 13 исследованных соединений, представлены в таблице 3.6. Как видно из таблицы, по коэффициентам корреляции шкала  $G_{Gibbs}^{ads}$  достаточно хорошо совпадает со шкалой гидратации ( $V > W$ ). Это вызвано тем, что для полярных гидрофильных молекул энергия адсорбции (глубина адсорбционного минимума) не сильно отличается от энергии гидратации. Для неполярных молекул такое утверждение не справедливо, энергия гидратации там плохо коррелирует с энергией адсорбции, однако, величины этих энергий не велики и коэффициент корреляции они портят не сильно. Демонстрация этого факта приведена на рис. 3.7.

### 3.5. Выводы к главе

#### Значимость и актуальность

Рост возможностей современных компьютерных систем постепенно позволяет переходить на качественно новый уровень в моделировании молекулярных процессов: вычисления свободных энергий, энтропий и других термодинамических характеристик реальных молекулярных систем на основе их атомистических моделей с приемлемой точностью. Основные трудности таких вычислений связаны с (i) необходимостью осуществления обширной выборки состояний системы в различных статистически значимых конформациях, число ко-

торых быстро возрастает с увеличением сложности энергетического ландшафта системы, (ii) необходимостью понимания применимости существующих силовых полей, параметризованных по различным “энергетическим” характеристикам молекул, для вычислений такого рода, (iii) умением соотносить характеристики рассчитанные в системах небольшого размера с настоящими термодинамическими характеристиками и необходимостью разрабатывать соответствующие методы. Особый интерес такие методы вызывают в связи с моделированием и построением моделей различных биологических структур (в частности белков), в самоорганизации которых важную роль играют гидрофобные взаимодействия, как раз носящие энтропийный, сильно статистический характер и, без адекватного описания которых на молекулярном уровне, рассмотрение задач о самосборке и стабильности таких структур в компьютерном моделировании весьма сложно. Заметим, что выигрыш в свободной энергии при сворачивании белков весьма небольшой (10-20 ккал/моль), что накладывает дополнительные требования к качеству воспроизведения взаимодействий в моделировании. Кроме того, поскольку биологические системы являются зачастую микрорегетерогенными, (например, растворитель и ядро белка, актуальным является также корректное воспроизведение взаимодействий для отдельных молекул как в объёме фазы, так и на поверхности, что в свою очередь влияет на такую характеристику как поверхностное натяжение. В данной работе освещён один из фундаментальных и одновременно методологических аспектов компьютерного моделирования белково-подобных систем, рассмотрено взаимодействие различных типов аминокислот с границей раздела фаз вода/воздух. Результаты работы с одной стороны дают фундаментальную информацию о взаимодействии аминокислот с поверхностью воды, с другой стороны предоставляют численные результаты для оценки точности широкоиспользуемых в настоящее время силовых полей и их совершенствования.

### **Проведённые работы**

Была разработана методика моделирования, исследована структура (в том числе ориентационная) и характеристики водного слоя толщиной 4 нм. Разработана методика и вычислены профили свободной энергии для боковых цепей аминокислот вдоль нормали к границе водного слоя. Обсуждается вопрос о методах расчёта энергии адсорбции, предложено два метода и проведены оценки энергии адсорбции для боковых цепей аминокислот. Независимым способом путём “выключения” взаимодействия между молекулой и растворителем в объёме проведена оценка свободной энергии гидратации исследуемых молекул, проведено сравнение энергий получаемых двумя способами. На основании расчётов проведена оценка корреляций между энергиями гидратации и адсорбции (как абсолютных, так и относительно фазы максимального сродства), также проведена оценка корреляций

шкал с результатами статистического анализа главы 4.

### **Положение работы в контексте аналогичных работ и новизна**

Расчёты свободных энергий гидратации для боковых цепей аминокислот были проведены целым рядом исследователей для различных силовых полей. В нескольких работах авторы очень внимательно относились к возможной зависимости параметров моделирования на получаемые результаты и получали результаты с достоверной оценкой погрешностей. Следует отметить, что данная тема является достаточно широко исследуемой в связи с вышеозвученными аргументами. Расчёт профилей свободной энергии для небольших молекул (в основном газов) на границе с водой проводился некоторыми исследователями, однако, на взгляд автора, в имеющихся работах выполнен достаточно небрежно. Что же касается систематического расчёта профилей свободной энергии для боковых цепей аминокислот на границе с водой, то таких работ найдено не было, поэтому можно утверждать, что данная работа выполнена впервые. Предложенный автором способ расчёта энергий адсорбции на основе профилей свободной энергии в купе с методом молекулярного моделирования также обладает новизной.

### **Выводы**

По результатам работы можно сделать следующие выводы, обладающие научной новизной:

- Профили средней силы для боковых цепей аминокислот на границе вода-воздух приблизительно одинаковы по форме: имеют область притяжения перед поверхностью Гиббса со стороны воздуха, и достаточно узкую (0.6-0.7 нм) область выталкивания за поверхность Гиббса.
- Вывод об одинаковой форме профилей средней силы на границе вода/воздух, по видимому, можно обобщить на весь класс небольших незаряженных молекул.
- По крайней мере в использованной модели, видимые поверхностные эффекты сольватации для боковых цепей аминокислот сходят на нет, когда центр масс молекулы погружён на 0.6-0.7 нм в воду по отношению к поверхности Гиббса, что может считаться эффективной глубиной границы раздела для маленьких молекул.
- Профили свободной энергии для всех молекул обладают минимумом в районе границы раздела, который является результатом баланса между силами притяжения, доминирующими со стороны воздуха, и силами отталкивания, доминирующими со стороны воды, и имеющими энтропийную природу.

- Глубина минимума свободной энергии находится в хорошей корреляции с энергией адсорбции, рассчитанной по методу Гиббса.
- С высокой точностью вычислены энергии гидратации и адсорбции для 13 незаряженных веществ-аналогов боковых цепей аминокислот.
- Показано, что энергия адсорбции из газовой фазы для выбранного набора боковых цепей аминокислот находится в хорошей корреляции с энергией гидратации.
- Энергия адсорбции относительно фазы максимального сродства не коррелирует с энергией гидратации и может использоваться как независимая термодинамическая характеристика молекул вещества.
- Полученные данные для нескольких молекул согласуются с экспериментальными данными и могут применяться для совершенствования силовых полей, однако достоверных систематических данных по оценке энергии адсорбции для все группы веществ на границе вода-воздух пока в литературе не обнаружено.

# Анализ распределения звеньев в глобулярных белках

Результаты данной главы в подробном изложении опубликованы в работе [2].

### 4.1. Введение

Водное окружение играет ключевую роль в фолдинге и стабильности почти всех белковых структур. Гидрофобные взаимодействия неполярных аминокислот считаются главной движущей силой сворачивания белков [102]. Понимание принципов и энергетики формирования белковых структур по отношению к распределению различных типов аминокислот в ядре белка и на его поверхности весьма полезно для таких задач как: дизайн белков *de novo* [103, 104], дизайн биоинспирированных [103, 104] и умных полимеров [105]. Из более ранних работ [95, 106–115] известно, что в глобулярных белках аминокислотные остатки с заряженными и полярными боковыми цепями в среднем стремятся быть более экспонированными по отношению к растворителю, чем остатки с гидрофобными боковыми группами. Известно также, что распределение остатков между ядром и поверхностью белка находится в приблизительной корреляции с коэффициентами распределения аминокислот или их производных между водой и менее полярными растворителями. Это распределение остатков по форме похоже на Больцмановское распределение, как если бы аминокислотные остатки были бы не зафиксированы в белковой структуре, а находились в динамическом равновесии между поверхностью и ядром белка. Похожесть статистики распределения структурных элементов в глобулярных белках на Больцмановскую статистику впервые была замечена Полем [116]. Теоретическое обоснование этого факта, основанное на модели случайных энергий [22] для гетерополимерной глобулы, было предложено Финкельштейном и др [21], о чём уже шла речь в главе I. Корреляции между *кажущимися свободными энергиями переноса* (и другими статистическими параметрами) и экспериментальными энергиями переноса аминокислот между водой и менее полярными растворителями прослеживались во многих работах [95, 108–115]. Чотиа [106] впервые провёл количественное изучение поверхности, доступной растворителю, в глобулярных белках и охарактеризовал их предпочтение находиться в ядре или на поверхности, Нозаки и Тэнфорд ранее [117] ввели первую шкалу гидрофобности, основанную на распределении аминокислот между водой, этанолом и диоксаном. Волфенден и др. [108] были одними из первых, кто численно изучил корреляцию между различными экспериментальными



шкалами и статистическими данным по распределению аминокислот в белках. Это стало началом длинного периода споров относительно наиболее адекватной экспериментальной шкалы соответствующей наблюдаемым статистическим распределениям. Было предложено большое количество экспериментальных шкал, включая энергии переноса между парами растворителей вода/этанол [117], вода/пар [108], вода/циклогексан [95], вода/октанол [118, 119], энергии адсорбции на границу раздела вода/воздух [96] и вода/гексан [100], а также множество полуэмпирических шкал. Однако только несколько шкал были хорошо оттестированы и практически использованы, это шкалы вода/пар, вода/циклогексан и вода/октанол, последняя из них является наиболее широко применяемой в различных численных QSAR оценках в медицинской химии. При этом в различных более ранних работах уже подчёркивался [109] тот факт, что корреляции между экспериментальными и статистическими шкалами могут сильно зависеть от тех критериев, которые используются для статистической классификации аминокислотных остатков на “ядерные” и “поверхностные”.

## 4.2. Постановка задачи

Задача данной главы состояла в исследовании статистики распределения аминокислот в глобулярных белках в зависимости от их экспонированности по отношению к растворителю (воде). Оценка экспонированности проводилась на основе расчёта поверхности доступной растворителю. Интерес представляет также исследование корреляций между экспериментально предлагаемыми шкалами гидрофобности, иными термодинамическими характеристиками аминокислот и распределением звеньев полипептидной цепи между ядром и поверхностью.

Кроме того было решено две вспомогательные задачи: (i) создание выборки белков для анализа, (ii) исследование вопросов зависимости результатов от типа выборки и размера белков.

## 4.3. Создание набора белков

Для выполнения исследования необходимо было отобрать набор пространственных структур белков, который отвечает трём критериям:

1. набор белков должен быть не избыточным (non-redundant), то есть в нём не должно присутствовать близких гомологов, что приводило бы к “эволюционному” сдвигу

результатов и противоречило бы модели случайных энергий, применяемой для интерпретации результатов.

2. набор белков должен содержать лишь глобулярные водорастворимые белки, которые стабильны в водном окружении без взаимодействия с какими-то стабилизирующими веществами.
3. для этих белков должны быть доступны структуры с высоким разрешением в базе данных PDB.

Выбор такого набора белков в автоматическом режиме достаточно нетривиальная задача, и мы допускаем, что в результате в конечном наборе белков остались белки, которые не совсем удовлетворяют заданным в пункте 2 критериям, но их доля должна быть незначительной.

Конечный набор белков был сформирован следующим образом: был получен набор белковых цепей Non-redundant PDB chain set list (датируемый 5 сентября 2008 года)[120], поддерживаемый NCBI. Данный набор содержит список полипептидных цепей, сгруппированных по сходству их первичной последовательности оцененной с помощью алгоритма BLAST [121]. Из каждой группы гомологов со значением параметра  $p=10^{-7}$  была отобрана одна белковая цепь, помеченная как репрезентативная (с высоким разрешением и другими характеристиками). Соответствующие данной цепи пространственные структуры целых белков были загружены из базы данных WWPDDB [20]. На этом этапе наш набор содержал 11748 белковых структур. После этого к данному набору были применены методы фильтрации и модификации. 197 структур были отброшены, поскольку имели неполное описание атомной структуры для одного или более остатков. Файлы PDB фильтровались на наличие слов “MEMBRANE”, “TOXIN”, “FIBER”, “FIBROUS”, и отбрасывались в случае нахождения там таких ключевых слов. Были отброшены все структуры содержащие ДНК или РНК. Файлы, находящиеся в базе данных, обычно представляют собой лишь ассиметричные ячейки соответствующих кристаллических структур, разрешённые методом рентгеноструктурного анализа. Для создания из них биологических молекул (мультимеров) были применены преобразования BIOMT. 277 структур было отброшено из-за необычных BIOMT записей в файлах структур. Для ЯМР структур, содержащих несколько моделей, выбиралась только первая модель.

На следующем этапе решалась задача отбрасывания структур, которые содержали не компактные глобулярные структуры, а скорее большие комплексы белков (например, вирус-

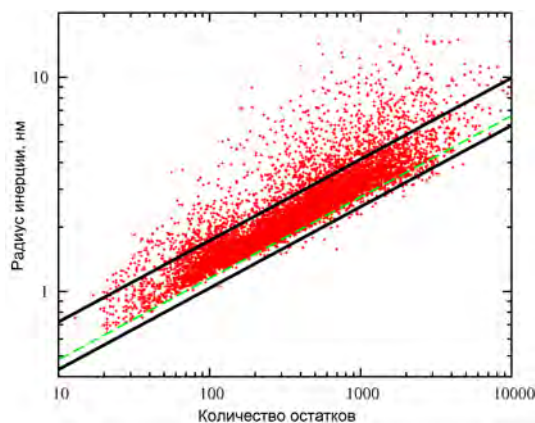


Рис. 4.1. Каждая точка представляет одну белковую структуру в координатах радиус инерции - количество аминокислот в логарифмической шкале по обеим осям. Пунктирная линия - аппроксимация нижней границы облака точек. Структуры находящиеся вне области между сплошными линиями были отброшены из дальнейшего рассмотрения.

ные капсиды, и др.). Для этого структуры фильтровались по соотношению радиуса инерции к длине первичной последовательности. На Рис. 4.1 каждая структура представлена одной точкой. Облако точек в логарифмической шкале имеет чёткую нижнюю границу, которая отвечает максимальной плотности упаковки для заданного количества остатков. Эта нижняя граница очень хорошо описывается уравнением:

$$R_g = 0.2 * N^{0.38} \quad (4.1)$$

где N-это количество остатков, а  $R_g$  выражено в нанометрах.

Степень 0.38 достаточно близка к одной трети, которая наблюдается в глобулярном состоянии гомополимеров.

Структуры белков, которые находятся вне области  $0.18 * N^{0.38} < R_g < 0.3 * N^{0.38}$  были отброшены. В результате набор белков для анализа состоял из 8022 структур. На Рис. 4.2 представлена гистограмма распределения структур по количеству аминокислотных остатков. Набор PDB-кодов белков приведён в дополнительной информации к статье [2].

## 4.4. Методы расчётов

### 4.4.1. Расчёт поверхности доступной растворителю

Поверхность доступная растворителю рассчитывалась с использованием программы NACCESS [122], которая реализует алгоритм Ли и Ричардса [67]: сфера заданного радиуса обкатывается вокруг исследуемой молекулы и поверхность, соответствующая положениям

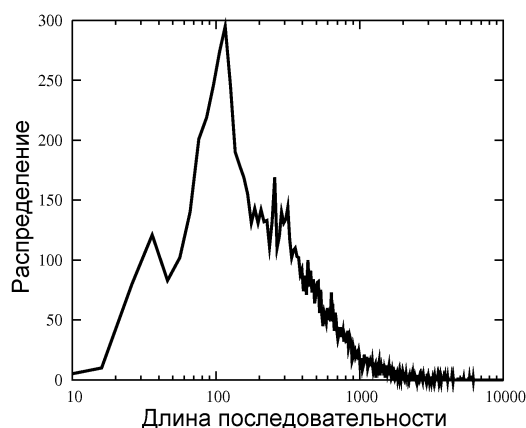


Рис. 4.2. Распределение изучаемых белковых структур по количеству остатков. Распределение сглажено методом смежного усреднения с окном 10.

центра данной сферы считается искомой поверхностью. В данном исследовании использовалась сфера радиуса 1.4 Å, соответствующая размеру молекулы воды. Индивидуальные поверхности каждого атома вычислялись, а затем складывались, чтобы получить поверхность доступную растворителю (solvent accessible surface area - SASA) каждого аминокислотного остатка, и каждой боковой цепи аминокислотного остатка.

#### 4.4.2. Экспонированность

Для описания меры доступности того или иного звена полипептидной последовательности растворителю можно использовать различные параметры. Наиболее простым и удобным является величина, которую мы назовём “экспонированность” ( $\epsilon$ ). Для каждого конкретного аминокислотного остатка в конкретной белковой структуре экспонированность определяется как отношение площади доступной растворителю ( $S_{sasa}$ ) к эталонной площади ( $S_{sasa}^0$ ) (см. ур. 4.2).

$$\epsilon = \frac{S_{sasa}}{S_{sasa}^0} \quad (4.2)$$

Эталонная площадь определялась как площадь поверхности, доступная растворителю, для заданного остатка в Ala-X-Ala трипептидах, взятых в вытянутой конформации [123].

Экспонированность выражается в процентах и может быть двух видов: (i) либо экспонированность по боковой цепи ( $\epsilon^{sch}$ ), либо (ii) экспонированность по всему остатку ( $\epsilon^a$ ), в зависимости от того берутся величины  $S_{sasa}$  и  $S_{sasa}^0$  только для боковой цепи или для всего остатка.

#### 4.4.3. “Статистическая” свободная энергия переноса

В эксперименте свободная энергия переноса молекулы между несмешивающимися жидкостями (фазами 1 и 2) рассчитывается согласно уравнению:

$$F_{transfer} = -kT \ln \frac{c_1}{c_2} \quad (4.3)$$

Аналогично “кажущаяся” свободная энергия переноса (apparent transfer free energy - устоявшийся термин в англоязычной литературе, на русский устоявшегося перевода нет, в книге “Физика белка” Птицына и Финкельштейна [19] был предложен вариант перевода - “кажущаяся энергия переноса”, в рамках данной работы будем использовать термин “кажущаяся энергия”, а также термин “статистическая энергия”) аминокислотных остатков между двумя “статистическими” фазами может быть определена. Статистическая фаза определяется как набор остатков в белковых структурах, которые удовлетворяют определённым критериям (например, их экспонированная поверхность равна 0). Введя критерий, который относит остатки к определённой фазе, можно ввести статистическую концентрацию каждого типа остатка по следующей формуле:

$$c^a = \frac{N^a}{\sum^i N^i} \quad (4.4)$$

- где  $N_a$  это количество остатков типа  $a$  в фазе. Тогда кажущаяся свободная энергия переноса может быть вычислена, используя то же уравнение (4.3) с применением уравнения (4.4) для расчёта концентраций. В данном исследовании критерии задания фаз были основаны на относительной поверхности боковых цепей аминокислот, доступной растворителю ( $\epsilon^{sch}$ ).

#### 4.5. Зависимость статистики от размера белков

Расчёт средних свойств аминокислот по ансамблю белковых структур может показаться не до конца обоснованным, поскольку конформационная статистика белков может зависеть от многих факторов, приводя к тому, что статистические параметры будут различными для различных наборов белков, что, в частности, отмечали Томас и Дилл [24]. Поскольку в данной работе мы изучаем общие закономерности влияния растворителя на глобулярные белки, то выбранный набор кажется нам вполне подходящим за исключением одной проблемы – возможной зависимости статистики от размера белков. Действительно большие белки могут иметь большие ядра, которые скрывают гидрофобные остатки, в то время как маленькие белки из-за геометрических размеров вынуждены экспонировать большую часть остатков и, следовательно, искать дополнительные взаимодействия

поддерживающие стабильность. Последний тезис может быть хорошо проиллюстрирован, если мы посмотрим на содержание аминокислоты цистеин в белках различного размера (см. Рис. 4.3). Серьёзное увеличение встречаемости цистеинов в белках размером менее 100 аминокислот подчёркивает важность дисульфидных связей в стабильности малых белков. Аналогичные феномены можно наблюдать для других остатков.

Сбор статистики по всему набору размеров белков приведёт к сомнительным результатам, поэтому нам придётся рассматривать лишь определённый набор размеров белков. На Рис. 4.3 представлена общая встречаемость различных аминокислот в белках различного размера, а также встречаемость аминокислот на поверхности и в ядре. Аминокислоты считались принадлежащими к ядру при  $\epsilon^a < 5\%$  и к поверхности в остальных случаях.

Анализ показывает, что встречаемость аминокислот у маленьких белков сильно изменяется с размером белков, но после некоторого предела остаётся приблизительно постоянной. Этот предел варьируется для разных аминокислот от 100 до 300 АКО. Избранные примеры аминокислот представлены на Рис. 4.3.

Более подробный анализ этих распределений позволяет сделать некоторые выводы относительно встречаемости различных остатков, сгруппировать остатки с аналогичными характеристиками, а также дать разумные объяснения того или иного поведения встречаемости для некоторых остатков. Такой анализ является хотя и интересным, но достаточно громоздким, поэтому мы опустим его при изложении этой главы, однако он может быть найден в статье [2].

В результате анализа можно сделать вывод о том, что все три рассмотренных характеристики встречаемости аминокислот в белках с хорошей точностью теряют свою зависимость от размера белка для белков содержащих более 300 АКО. Объяснение такого эффекта скорее всего состоит в том, что большие белки обычно имеют доменную структуру, они состоят из нескольких суб-глобул, которые обычно не превышают в размере 250-300 АКО [19]. Величина в 300 АКО может рассматриваться как предел, когда мы имеем дело с мультидоменными белками или глобулами предельного размера.

Дополнительный аргумент в пользу изменения крупномасштабного строения белков в районе 300 АКО можно привести, рассматривая зависимость поверхности белков от количества АКО (см. Рис. 4.4 ). Аппроксимируем зависимость уравнением:

$$A_S = b * N^k \quad (4.5)$$

где  $N$  количество АКО,  $b$  и  $k$  подгоночные коэффициенты,  $k$  - тангенс угла наклона кривой в логарифмической шкале. Для  $N < 200$  и  $N > 200$  наклон графика, соответственно, равно 0.74 и 0.86. Если первая величина ближе к 2/3, как для сферических тел, то вторая

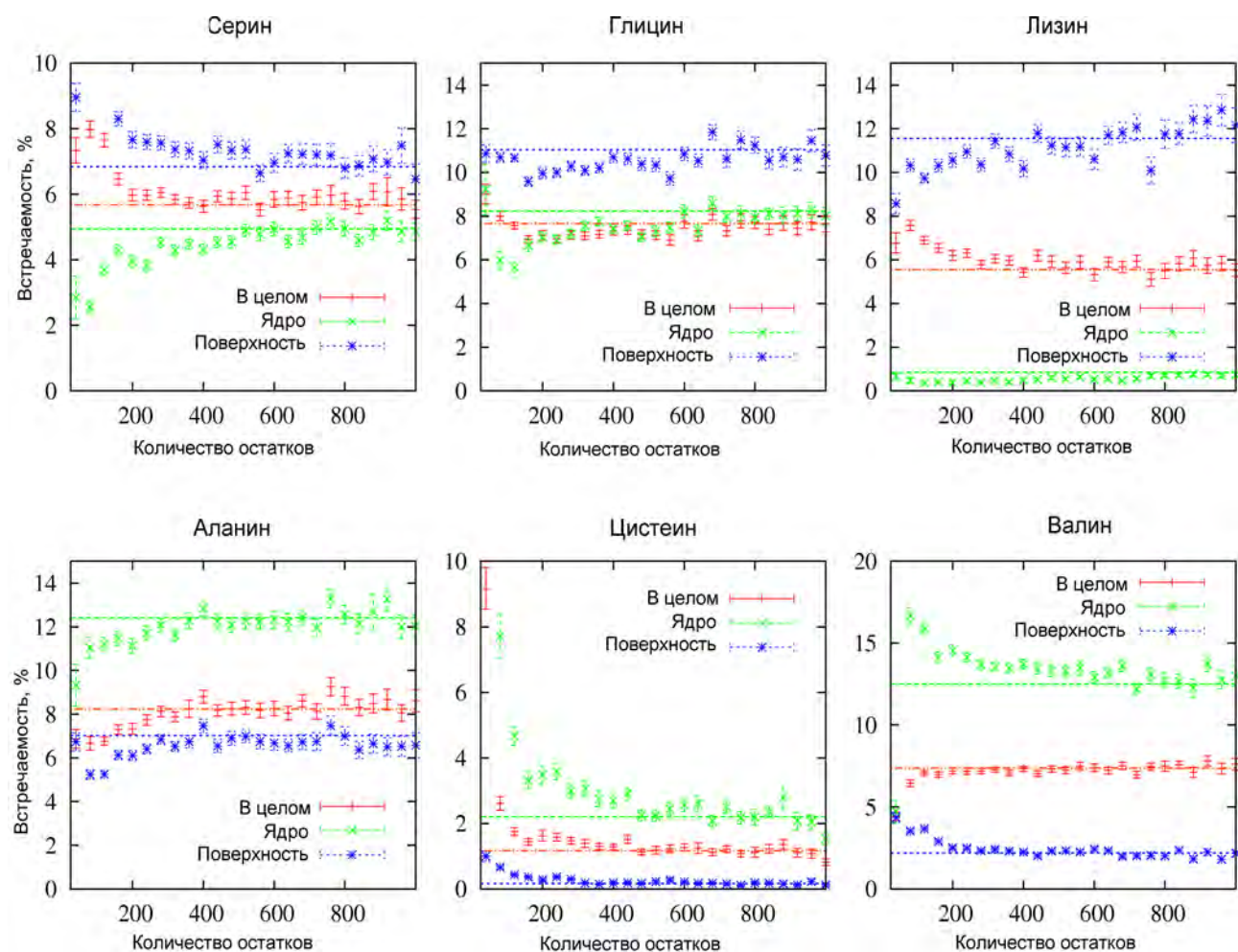


Рис. 4.3. Встречаемость аминокислот в белках различного размера, включая полную встречаемость, встречаемость на поверхности и встречаемость в ядре (см. текст). Горизонтальные линии в каждом случае представляют встречаемость, усреднённую по всем белковым структурам с размером большим 300 АКО. Каждая точка представляет данные, усреднённые по окну шириной 40 АКО. Стандартные отклонения представлены вертикальными линиями.

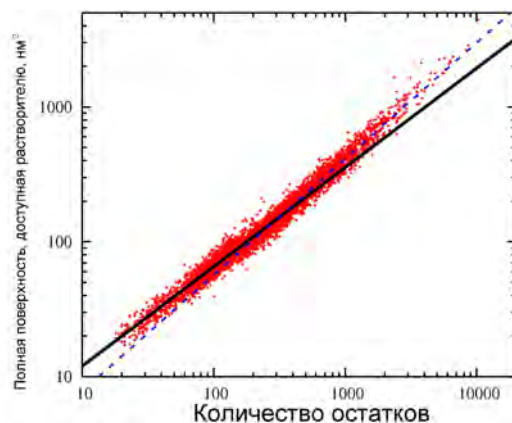


Рис. 4.4. Каждая точка - белковая структура, расположенная в координатах поверхность глобулы - размер белка в АКО. Сплошная и пунктирные линии - степенные аппроксимации для участков  $N < 200$  и  $N > 200$ .

отвечает системам с более развитой поверхностью.

В дальнейшем наш статистический анализ будет ограничиваться лишь набором белков, которые содержат более 300 АКО, такое ограничение позволит корректно провести статистический анализ и корректно интерпретировать результаты.

#### 4.6. Статистический анализ доступных поверхностей аминокислот

В данном разделе обратимся к основной цели работы – детальному изучению экспонированности различных типов аминокислот и их боковых цепей в белковых структурах. Для этого нам необходимо определить некоторый критерий, по которому можно будет оценивать предпочтение остатка иметь тот или иной уровень экспонированности, оценивать поведение данного типа остатков относительно других остатков.

Для выполнения этой задачи наиболее разумным является построение зависимости доли (концентрации) данного типа остатка среди других остатков с одинаковой экспонированностью. Эту долю можно рассматривать как концентрацию остатков в статистической фазе, состоящей из остатков с заданной относительной доступностью растворителю. Для каждого значения экспонированности  $\epsilon^{sch} = x\%$  остатки, имеющие относительную доступность своих боковых цепей в пределах  $x \pm 1\%$ , считались принадлежащими одной статистической фазе и их концентрация рассчитывалась согласно формуле 4.4. Полученные в результате зависимости представлены на Рис. 4.5. Эти графики отражают относительное поведение остатков. На каждом графике представлена дополнительная точка, соответствующая нулевой доступности, которая отражает долю данного типа остатка среди всех остатков



полностью погружённых в белковую глобулу. Многие графики, таким образом, имеют видимое отсутствие непрерывности в нуле, которое означает, что структурный состав ядра белка как целого отличается от состава статистической фазы, которая практически не экспонирована ( $0\% < \epsilon^{sch} < 1\%$ ). В то время, как любая точка на графике 4.5 представляет концентрацию остатков в бесконечно малой гипотетической фазе, точка в нуле соответствует статистике по фазе конечного размера, фазе всех остатков недоступных растворителю. Направление прыжка функции в нуле хорошо коррелирует с размером боковой цепи остатка: маленькие остатки (GLY, ALA, CYS) имеют повышенную концентрацию среди 100% недоступных остатков, в то время как более большие остатки (PHE, TRP, TYR, GLN, LYS, ASN, HIS, GLU, ASP, ARG) демонстрируют обратную тенденцию. Этот факт частично указывает на то, что чисто стерические эффекты могут быть ответственны за такое поведение – маленьким остаткам легче, чем большим, быть полностью погружёнными в глобулу.

Проведём теперь качественный анализ зависимостей представленных на рис. 4.5, и сгруппируем остатки по их сходству.

Остатки с большими гидрофобными цепями (ILE, LEU, VAL, PHE, MET, CYS) имеют монотонные экспоненциально-подобные затухающие с ростом экспонированности зависимости. Что весьма ожидаемо для гидрофобных остатков.

Для аланина (ALA) зависимость выглядит немного другим образом из-за малого размера боковой цепи.

Глицин (GLY) является местом повышенной конформационной подвижности полипептидной цепи, что сказывается на его сильной экспонированности. Такой же особенностью обладает пролин (PRO), являясь иминокислотой с нестандартными торсионными углами, он часто располагается в местах поворотов полипептидной цепи, что приводит к его увеличенной экспонированности.

Рассмотрим остатки с полярными и заряженными цепями. Такие остатки, как ASN, ASP, SER, имеют графики с монотонным ростом концентрации от экспонированности. Все эти остатки обладают маленькими боковыми группами, которые способны формировать водородные связи. Из-за конкуренции с атомами остова, которые тоже могут образовывать водородные связи, эти остатки избегают таких мест вторичной структуры как альфа-спирали и бета-листы, что опять же является фактором, приводящим к повышенной концентрации данных остатков среди остатков с высокой и очень высокой доступностью. С другой стороны другие остатки с гидрофильными боковыми цепями (ARG, GLN, GLU, LYS, HIS) имеют горбообразные зависимости, хотя энергии гидратации некоторых из этих остатков (см. табл. 4.2) превышают значения для предыдущей группы. Причины этого ско-

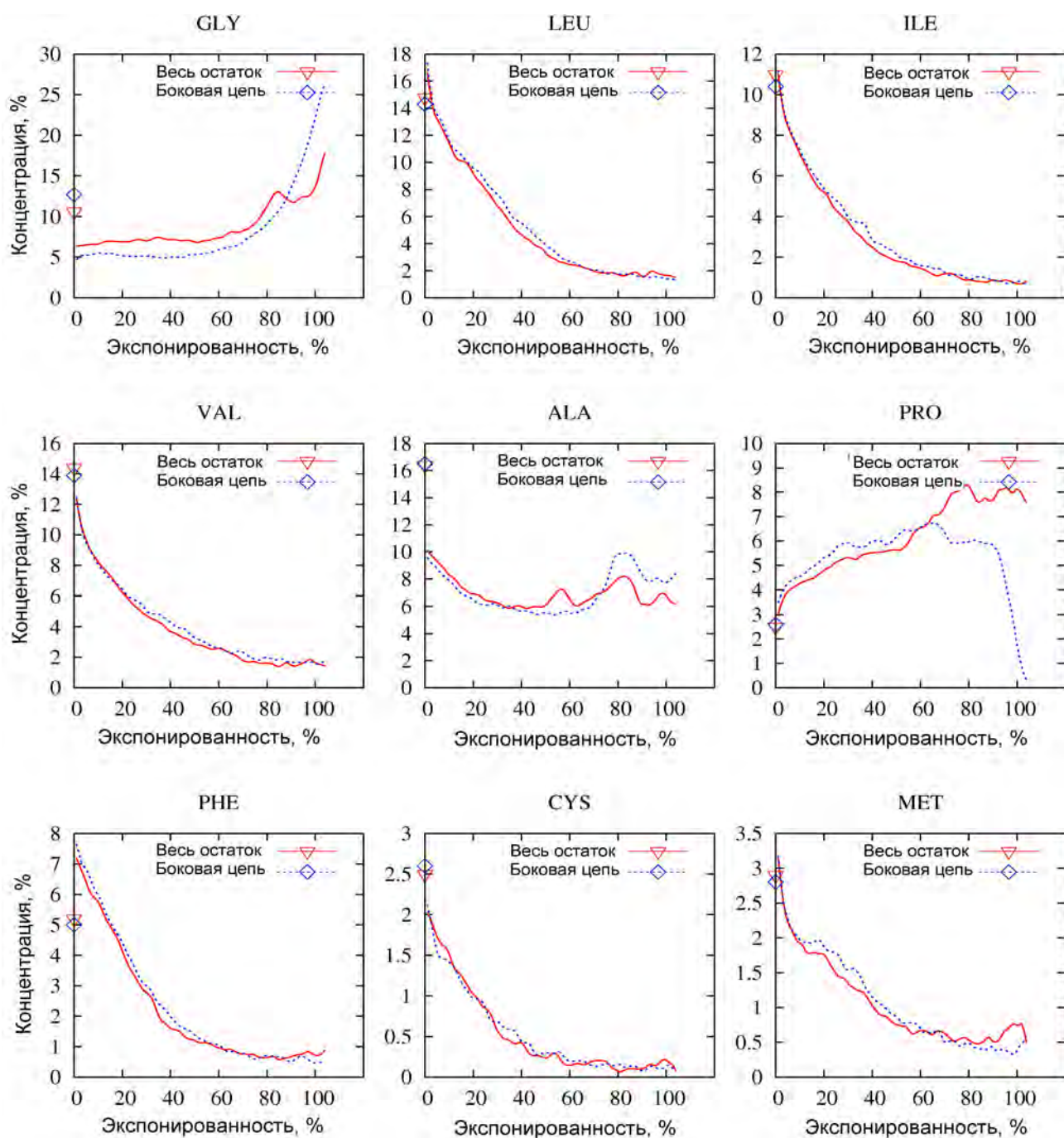


Рис. 4.5. Зависимость концентрации остатков в “статистической фазе” от относительной доступности боковых цепей аминокислот. Концентрация заданного типа остатка определяется как доля таких остатков среди всех остатков с заданной экспонированностью. Линии построены с использованием сплайн аппроксимации. Точка в нуле соответствует данным по всем неэкспонированным остаткам.

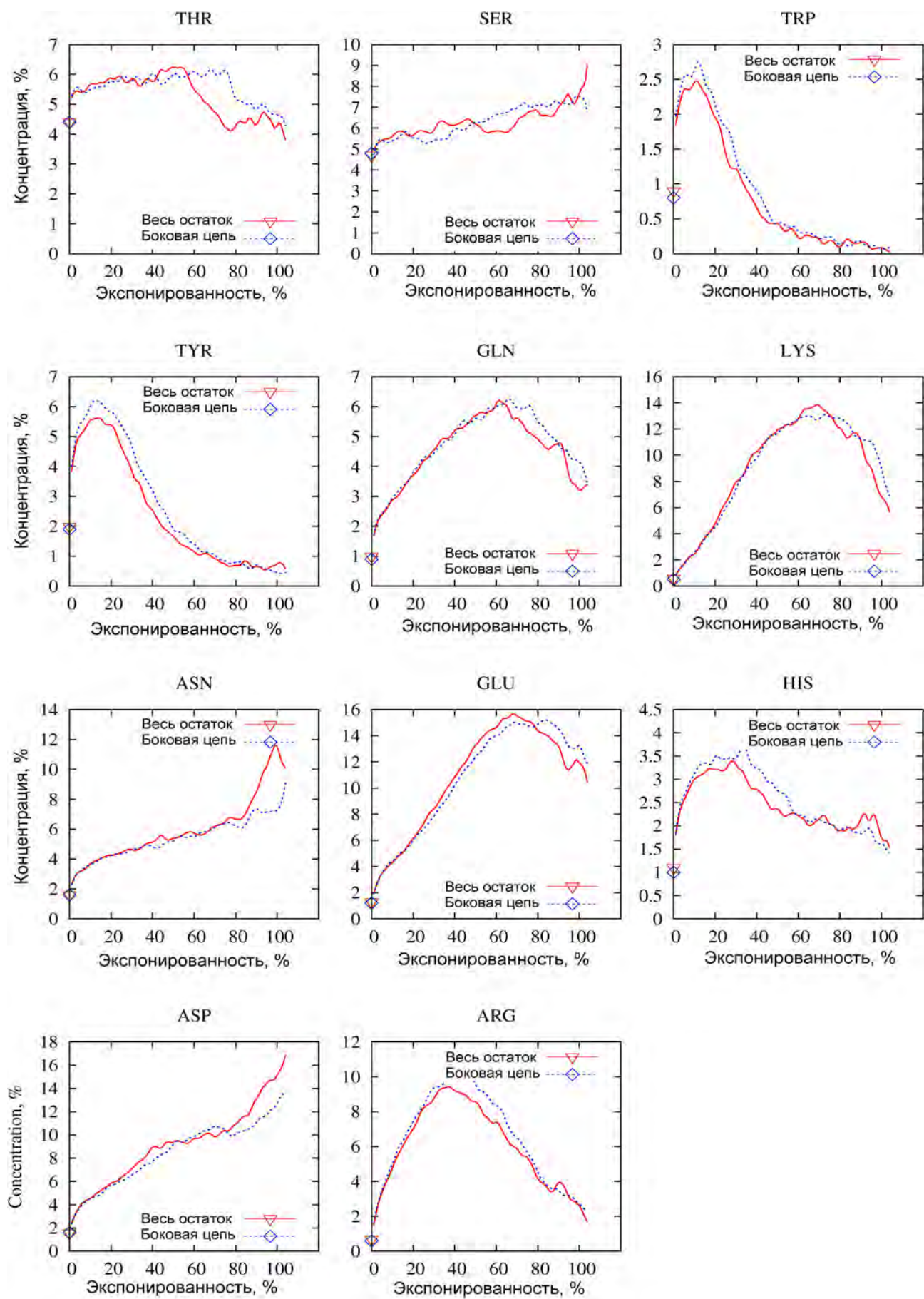


Рис. 4.5. Продолжение.  
67

рее всего состоят в том, что данные остатки имеют не только полярные атомы в своих боковых цепях, но и гидрофобные, поэтому они ищут баланс между экспонированностью полярных групп и погружённостью неполярных. Видимо, это зависит и от соотношения между группами, так LYS и ARG имеют намного более выраженную горбатость зависимости, чем GLU и GLN, которые имеют в два раза меньше атомов углерода в боковой цепи. Остатки TYR и THR имеют в составе боковой цепи в основном неполярные атомы с включением одного полярного, эти остатки часто встречаются в бета-листах, где они могут разместить свои ветвящиеся боковые цепи. Зависимости для этих остатков являются комбинацией горбо-подобных при малом экспонировании и затухающих экспоненциально-подобный при больших значениях экспонирования, что объяснимо, исходя из строения их боковых цепей.

Зависимость для треонина (THR) несколько выпадает из общего рассмотрения, но её можно рассматривать как развитие зависимости для серина при добавлении дополнительного атома углерода в боковую цепь.

Теперь у нас есть некоторое понимание, по крайней мере на качественном уровне того, как остатки ведут себя в зависимости от экспонированности. Важный вывод, который можно сделать состоит в том, что распределение аминокислотных остатков данного типа зависит не только от склонности боковых цепей к гидратации, но также и от других факторов, определяющих выгодность размещения таких аминокислот в том или ином месте структуры белка (размер боковой цепи, её способность образовывать водородные связи, влияние размера на конформационную подвижность остова и т.д.).

#### **4.7. Корреляции с экспериментальными шкалами гидрофобности**

Перейдём теперь к вопросу, озвученному в начале главы, а именно, к выявлению корреляций между экспериментально известными шкалами гидрофобности и статистическим распределением аминокислот в белках. Для проведения данного сравнения нами было отобрано три экспериментальных шкалы переноса аминокислот или их производных между следующими парами растворителей: вода/пар, вода/циклогексан, вода/октанол. Данные шкалы были выбраны по нескольким причинам: (1) эти шкалы представляют основные и простые взаимодействия между растворённым веществом и растворителем, и они легко поддаются интерпретации, (2) для этих шкал доступно много информации по их точности, надёжности, сравнению шкал и обсуждению их особенностей, (3) эти шкалы широко используются в практических приложениях.

Таблица 4.1. Список 19 аминокислот (кроме пролина), их сокращённые обозначения и R-Н аналоги.

Сокр.	1-буквенный код	Название	R-Н аналог <sup>a</sup>
GLY	G	глицин	водород
LEU	L	лейцин	изобутан
ILE	I	изолейцин	н-бутан
VAL	V	валин	пропан
ALA	A	аланин	метан
PHE	F	фенилаланин	толуол
CYS	C	цистеин	мтанэтиол
MET	M	метионин	метилэтилсульфид
THR	T	треонин	этанол
SER	S	серин	метанол
TRP	W	триптофан	3-метилиндол
TYR	Y	тирозин	4-метилфенол
GLN	Q	глутамин	пропионамид
LYS	K	лизин	н-бутиламин
ASN	N	аспарагин	ацетамид
GLU	E	глутаминовая кислота	пропионовая кислота
HIS	H	гистидин	4-метилимидазол
ASP	D	аспарагиновая кислота	уксусная кислота
ARG	R	аргинин	N-пропилгуанидин

<sup>a</sup> Названия R-Н аналогов аминокислот, где R-боковая цепь аминокислоты.

#### 4.7.1. Экспериментальные шкалы

Несмотря на тот факт, что шкалы используются достаточно часто и обсуждались в различных работах, сами экспериментальные методы и результаты очень часто содержат большое количество нюансов. В этом разделе мы кратко опишем, использованные шкалы и их особенности. Более подробное обсуждение данного вопроса и сравнение шкал изложено в статье [2].

Три экспериментальные шкалы, использованные для сравнения, представлены в таблице 4.2, обозначения аминокислот и соответствующие их боковым цепям R-Н соединения могут быть найдены в таблице 4.1.

Аминокислота пролин в дальнейшем будет исключена из рассмотрения, поскольку

Таблица 4.2. Экспериментальные шкалы свободных энергий переноса боковых цепей аминокислот и аналогичных молекул из воды в менее полярный растворитель при pH 7, выраженные в ккал/моль.

Аминокислота	Тип <sup>a</sup>	V>W <sup>b</sup>	CH>W <sup>c</sup>	O>W <sup>d</sup>	O>W <sub>occ</sub> <sup>e</sup>	CH>O <sup>f</sup>
GLY	N	2.39	0.94	0	0	0
LEU	N	2.28	4.92	2.30	2.40	1.68
ILE	N	2.15	4.92	2.46	2.27	1.52
VAL	N	1.99	4.04	1.66	1.61	1.44
ALA	N	1.94	1.81	0.42	0.65	0.45
PHE	A	-0.76	2.98	2.44	2.86	-0.40
CYS	P	-1.24	1.28	1.39(2.10)	1.17	-1.05
MET	N	-1.48	2.35	1.68	1.82	-0.27
THR	P	-4.88	-2.57	0.35	0.90	-3.86
SER	P	-5.06	-3.40	-0.05	0.69	-4.29
TRP	A	-5.88	2.33	3.07	3.24	-1.68
TYR	A	-6.11	-0.14	1.31	1.86	-2.39
GLN	P	-9.38	-5.54	-0.30	0.38	-6.18
LYS	+	-9.52	-5.55	-1.35	-1.65	-5.14
ASN	P	-9.68	-6.64	-0.79	0.30	-6.79
GLU	-	-10.24	-6.81	-2.35(-0.87)	-2.48	-5.40
HIS	P+ <sup>g</sup>	-10.27	-4.66	0.18	-1.18(1.04)	-5.78
ASP	-	-10.95	-8.72	-2.46(-1.05)	-2.49	-7.20
ARG	+	-19.92	-14.92	-1.37	-0.66	-14.49

<sup>a</sup> Тип аминокислоты: N - неполярная, алифатическая; P - полярная, незаряженная; A - ароматическая; '+'/'-' - положительно/отрицательно заряженная.

<sup>b</sup> Свободные энергии переноса R-H аналогов из воздуха в воду, по [108]

<sup>c</sup> Свободные энергии переноса из циклогексана в воду по [95].

<sup>d</sup> Энергии переноса между октанолом и водой из работы [118], отредактированная Вимли [119]. Оригинальные значения в скобках.

<sup>e</sup> Полуэкспериментальная шкала переноса из октанола в воду по [119].

<sup>f</sup> Энергии переноса между октанолом и циклогексаном.

<sup>g</sup> рК гистидина близко к 7. Шкалы переноса V>W и CH>W включают поправку на ионизацию. Данные O>W для неионизированной формы, данные O>W<sub>occ</sub> для обеих форм.

она отлична от всех остальных и не имеет разумного R-H аналога. R-H<sup>1</sup> аналоги – это простые соединения, представляющие собой боковые цепи аминокислот, в которых вместо C- $\alpha$ -атома аминокислоты присоединён атом водорода. В четвёртой колонке таблицы 4.2 представлены данные по переносу R-H аналогов из пара в воду, то есть энергии их гидратации. Эти данные были частично измерены Вольфенденом и др. [108], частично собраны из других источников, и частично рассчитаны косвенно из других измерений.

Следующая шкала - шкала переноса R-H аналогов между водой и циклогексаном, полученная Радзиком и др. [95] с использованием измерения UV-поглощения и протонного магнитного резонанса (колонка 6 таблицы 4.2), эти измерения были проведены для неионизированных веществ, а затем к данным была применена теоретическая поправка, чтобы получить оценку при pH 7. Радзика и др. [95] показали, что разница энергий переноса между шкалами вода/пар и вода/циклогексан хорошо коррелирует с полной поверхностью молекул, что может возникать в случае чисто дисперсионных взаимодействий между циклогексаном и R-H аналогами.

Последняя шкала, это шкала переноса веществ из 1-октанола в воду. Эта одна из самых используемых шкал, однако одна из самых противоречивых. Наиболее часто используется шкала Фошере и др. [118], измеренная для переноса ацетил-X-амидов, однако некоторые работы указывают на то, что данные для некоторые веществ весьма странные. Эта шкала была модифицирована Вимли и др. [119], чтобы исправить эти несостыковки, и эта исправленная шкала представлена в таблице 4.2. Строго говоря октанол не является несмешивающимся с водой, при насыщении он содержит 2.3 М воды. Некоторые исследования утверждают, что октанол имеет селективное взаимодействие с некоторыми соединениями. Так Вольфенден предупреждает [115], что в ходе экспериментов не исключён эффект затягивания воды в октанол молекулами исследуемого вещества, а также вероятно селективное притяжение между октанолом, спиртами и производными индола, такими, как TRP<sup>1</sup>.

В таблице 4.3 представлены коэффициенты корреляции между различными экспериментальными шкалами.

#### 4.7.2. Статистические энергии переноса

Расчёт статистических энергий переноса заслуживает отдельного обсуждения. Для оценки данной характеристики обычно используется бинарная классификация остатков на ядерные и поверхностные. Строго говоря, никто не выводил связь вычисляемой та-

---

<sup>1</sup> R-H аналоги аминокислот в дальнейшем будут обозначаться как трёх-буквенный код аминокислоты со штрихом.

Таблица 4.3. Коэффициенты корреляции экспериментальных шкал, представленных в таблице 4.2

	V>W	CH>W	O>W	O>W <sub>occ</sub>	CH>O
V>W	1				
CH>W	<b>0.94</b>	1			
O>W	0.68	0.86	1		
O>W <sub>occ</sub>	0.62	0.77	0.64	1	
CH>O	<b>0.97</b>	<b>0.98</b>	0.74	0.64	1

ким образом статистической энергии с параметрами  $\Delta F_{select}$ ,  $\Delta E$  уравнения 4.3 теории Финкельштейна и др., однако такой подход является весьма интуитивным и превалирует в литературе. Вопрос классификации остатков на поверхностные и ядерные обсуждался многократно [106, 107, 109, 111, 113, 124], разные исследователи предлагали разные критерии. В данной работе мы изучим сразу непрерывный набор различных критериев и выберем из них затем несколько для более подробного рассмотрения.

Чтобы исследовать набор различных критериев, мы вычислили целую серию различных наборов кажущихся энергий переноса на основе различных классификационных параметров, и построили графики корреляций между экспериментальными данными и наборами статистических энергий переноса в зависимости от параметров классификации. В первом случае (см. Рис. 4.6,а) рассматривалось распределение остатков между ядром белка ( $\epsilon^{sch} = 0$ , где  $\epsilon^{sch}$  – это относительная доступность растворителю боковой цепи АКО, выраженная в процентах) и остатками с экспонированностью в диапазоне  $X < \epsilon^{sch} < X + 1$ , в зависимости от параметра  $X$ . Как обсуждалось ранее, концентрация остатков в ядре отличается от концентрации в зоне, стремящейся к нулю доступности ( $0 < \epsilon^{sch} < 1$ ). Поскольку это отличие могло быть вызвано дополнительными специфическими взаимодействиями в ядре белка, был рассмотрен и второй тип классификации: распределение остатков между фазой, задаваемой ограничением  $0 < \epsilon^{sch} < 1$ , и фазой, задаваемой ограничением  $X < \epsilon^{sch} < X + 1$  (Рис. 4.6, б).

Рис. 4.6 позволяет лишь грубо проанализировать изменения происходящие с наборами кажущихся энергий переноса при изменении конечной фазы, определяемой доступностью  $X$ , но при этом позволяет найти области лучших корреляций с экспериментальными данными. Из рис. 4.6, а) видно, что кривая корреляции для шкалы вода/пар начинается со значения 0.7, достигает максимума при  $X=25\%$  ( $R_{max} = 0.93$ ) и затем постепенно уменьшается при больших значениях параметра доступности. Поведение шкалы вода/октанол



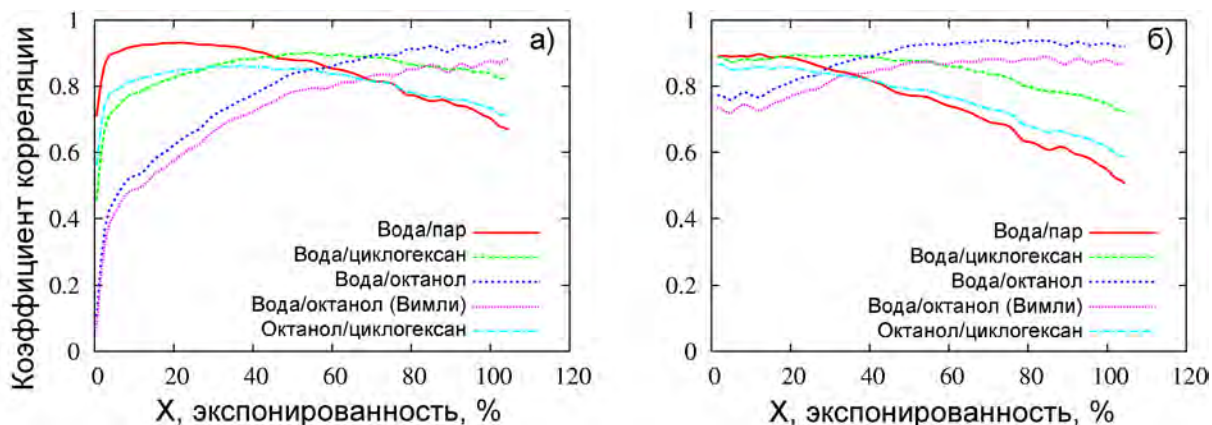


Рис. 4.6. Корреляция различных наборов статистической энергии переноса с наборами экспериментальных данных как функции классификационного критерия  $X$ . Статистические свободные энергии переноса были определены для переноса остатков из виртуальной фазы с (а)  $\epsilon^{sch} = 0$  и (б)  $0 < \epsilon^{sch} < 1$  в фазу с  $X < \epsilon^{sch} < X + 1$ , где  $\epsilon^{sch}$  относительная доступность боковой цепи остатка.

отличается: корреляции при малых значениях параметра  $X$  малы, а затем монотонно увеличиваются и достигают максимума на правом конце графика ( $R_{max} = 0.95$ ). Кривая для шкалы вода/циклогексан достигает максимума в середине шкалы ( $R_{max} = 0.906$  при  $X=55\%$ ). Предложенная шкала октанол/циклогексан плохо коррелирует со статистическими данными при любых параметрах  $X$ , то же можно сказать и о полу-теоретической шкале вода/октанол Вимли и др.[119]. Поведение кривых Рис. 4.6, а) и б) не отличается критическим образом, поэтому мы сконцентрируемся в дальнейшем на варианте а).

Для дальнейшего исследования природы корреляций мы введём 3 различных статистических шкалы, которые представляют собой виртуальный перенос остатков из ядра белка в: 1) статистическую фазу с доступностью боковых цепей в интервале  $10 < \epsilon^{sch} < 20$ , этот интервал включает плато высокой корреляции со шкалой вода/пар; 2) фазу с  $50 < \epsilon^{sch} < 60$ , плато высокой корреляции со шкалой вода/циклогексан; 3) фазу с  $95 < \epsilon^{sch} < 105$ , высокая корреляция со шкалой вода/октанол.

Полученные данные для этих трёх шкал, а также их коэффициенты корреляций приведены в таблице 4.4. Чтобы понять различия между этими шкалами, мы использовали сравнительные двумерные корреляционные диаграммы, на которых отложены статистические энергии переноса напротив экспериментальных наборов. Рис. 4.7, 4.9, 4.8.

### 4.7.3. Двумерные диаграммы

Наиболее подробное описание диаграмм было приведено нами в работе [2]. Анализ и разбор диаграмм достаточно кропотливая задача поскольку зависимости поистине многопараметрические и выявление общих закономерностей порой спекулятивно. Приведём основные выводы, которые были получены нами при анализе различных диаграмм.

Диаграмма Рис. 4.7 отражает корреляции между экспериментальной шкалой вода/пар и статистическими шкалами “10-20” и “95-105”. Коэффициент корреляции со шкалой “10-20” 0.93, а со шкалой “95-105” намного хуже ( $R=0.71$ ). По рассмотрению причин такого изменения можно сделать следующие выводы. Шкала энергий переноса вода/пар главным образом грубо схватывает основные тенденции распределения остатков между ядром и поверхностью (остатки с полярными цепями стремятся быть на поверхности, а с гидрофобными в ядре). Статистическая шкала “10-20” в сравнении со шкалой “95-105” также в основном отражает грубые тенденции, тогда как последняя более чувствительна к типу остатка и иным факторам (см. следующий раздел).

На диаграмме Рис. 4.8 исследована шкала вода/октанол. Общий вывод по этой шкале состоит в том, что шкала вода/октанол более чувствительна к небольшим вариациям в структуре и составе остатков и обладает в этом плане лучшим “разрешением”, чем шкала вода/пар. Поэтому она оказывается в хорошей корреляции со статистической шкалой “95-105” особенно в том случае, если рассматриваются гидрофобные и средне гидрофильные остатки.

Диаграмма Рис. 4.9 отражает корреляции между экспериментальной шкалой вода/циклогексан и статистическими шкалами “50-60” и “95-105”.

Общий вывод, который можно сделать по анализу этих трёх экспериментальных и трёх статистических шкал состоит в следующем. Предложенные статистических шкалы отличаются по своей “чувствительности”, или, другими словами, равномерности распределения точек на шкале. Шкала “10-20” ухватывает только грубые тенденции для АКО быть экспонированными или погружёнными в ядро, шкалы “50-60” и “95-105” более чувствительны к особенностям строения и поведения остатков. Шкала вода/пар также менее чувствительна к особенностям структуры соединений, в то время как шкалы вода/октанол и вода/циклогексан более чувствительны, но характер их чувствительности несколько различен, и поэтому они находятся в лучшей корреляции с различными статистическими шкалами. Для всех экспериментальных шкал корреляции лучше в группе гидрофобных и умеренно гидрофильных остатков, чем в группе сильно гидрофильных остатков.

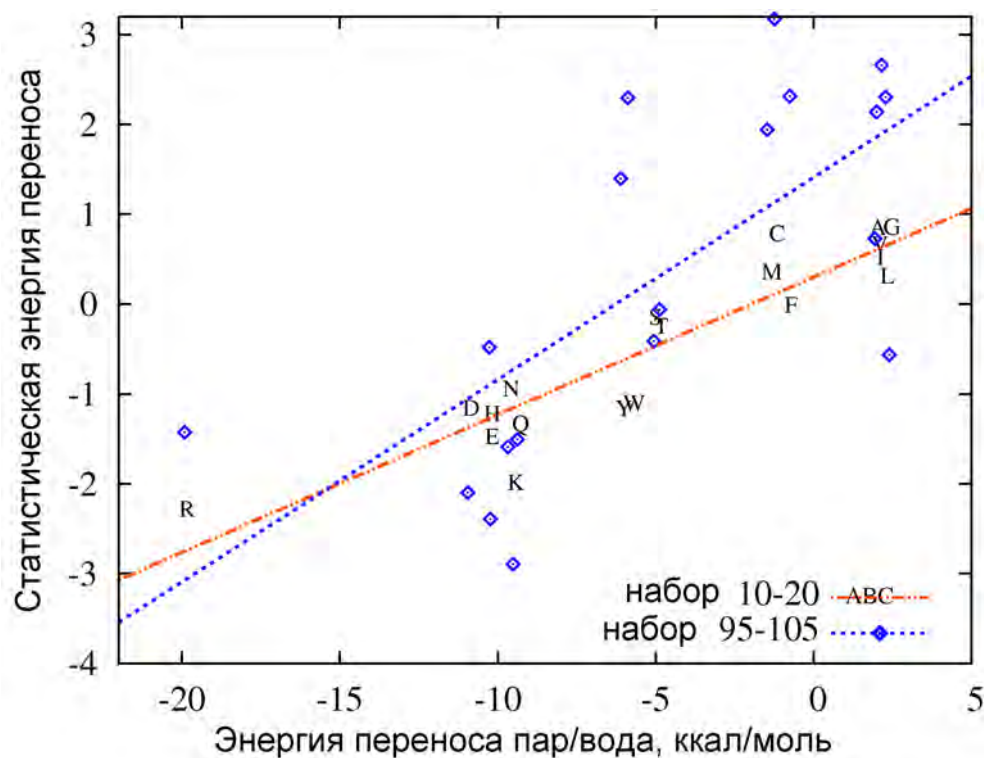


Рис. 4.7. Двумерные корреляционные диаграммы для экспериментальной шкалы вода/пар и двух статистических шкал. Шкала “10-20” – это шкала статистических энергий переноса для аминокислот из ядра белка ( $\epsilon^{sch} = 0$ ) в статистическую фазу с относительной доступностью боковых цепей в диапазоне  $10 < \epsilon^{sch} < 20$ , точки для данной шкалы отмечены символами однобуквенного кода аминокислот (см. таблицу 4.1). Пунктирная линия – это аппроксимация линейной зависимости с коэффициентом корреляции 0.93. Шкала “95-105” – это шкала статистических энергий переноса для аминокислот из ядра белка ( $\epsilon^{sch} = 0$ ) в фазу с относительной доступностью боковых цепей в диапазоне  $95 < \epsilon^{sch} < 105$ . Точки для этой шкалы изображены символами согласно обозначению на одной вертикали с однобуквенным кодом. Точечная линия – это аппроксимация линейной зависимости с коэффициентом корреляции 0.71.

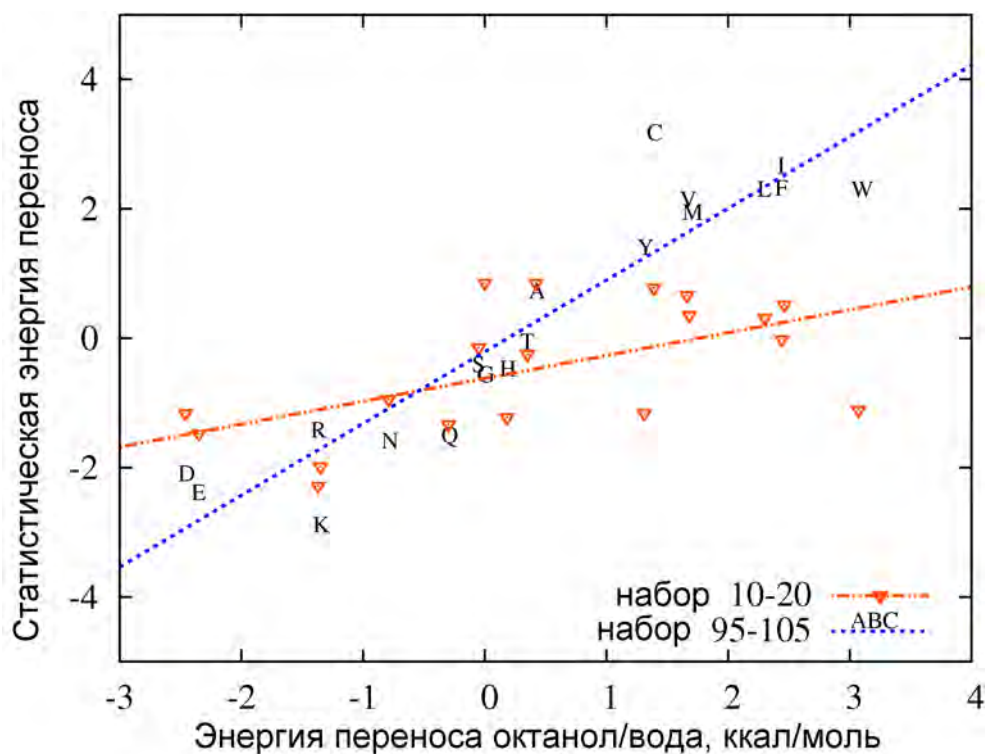


Рис. 4.8. Двумерные корреляционные диаграммы для экспериментальной шкалы вода/октанол и двух статистических шкал. Шкала “10-20” – это шкала статистических энергий переноса для аминокислот из ядра белка ( $\epsilon^{sch} = 0$ ) в статистическую фазу с относительной доступностью боковых цепей в диапазоне  $10 < \epsilon^{sch} < 20$ . Точки для этой шкалы изображены символами согласно обозначению на одной вертикали с однобуквенным кодом. Пунктирная линия это аппроксимация линейной зависимости с коэффициентом корреляции 0.57. Шкала “95-105” – это шкала статистических энергий переноса для аминокислот из ядра белка ( $\epsilon^{sch} = 0$ ) в фазу с относительной доступностью боковых цепей в диапазоне  $95 < \epsilon^{sch} < 105$ . Точки для данной шкалы отмечены символами однобуквенного кода аминокислот (см. таблицу 4.1). Точечная линия – это аппроксимация линейной зависимости с коэффициентом корреляции 0.93.

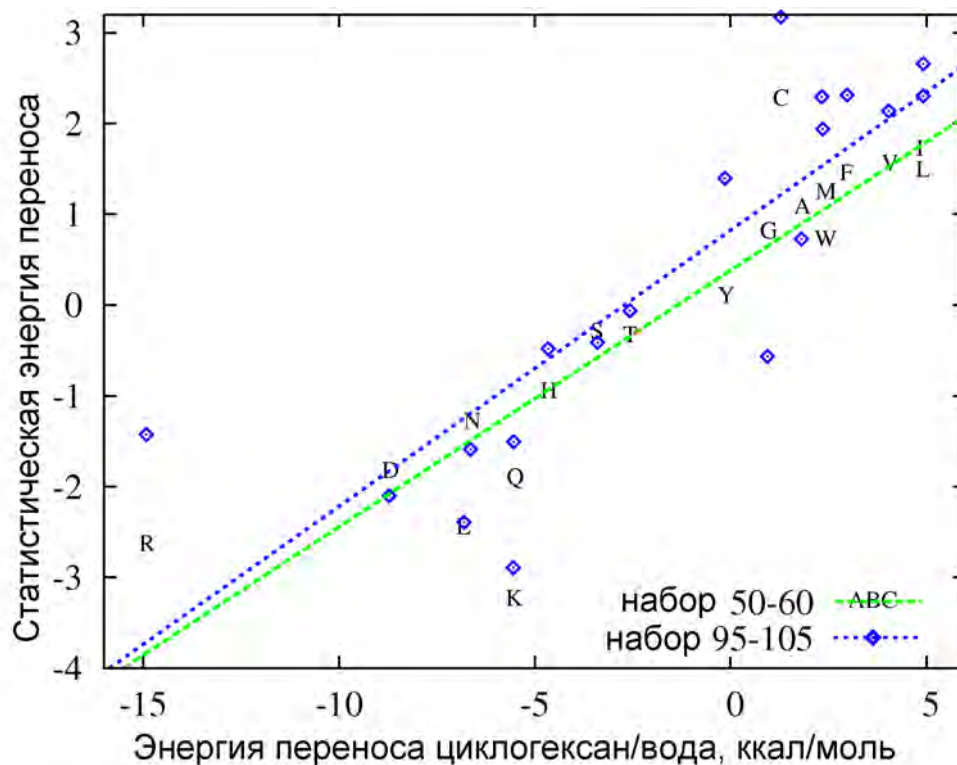


Рис. 4.9. Двумерные корреляционные диаграммы для экспериментальной шкалы вода/циклогексан и двух статистических шкал. Шкала “50-60” – это шкала статистических энергий переноса для аминокислот из ядра белка ( $\epsilon^{sch} = 0$ ) в статистическую фазу с относительной доступностью боковых цепей в диапазоне  $50 < \epsilon^{sch} < 60$ . Точки для данной шкалы отмечены символами однобуквенного кода аминокислот (см. таблицу 4.1). Пунктирная линия – это аппроксимация линейной зависимости с коэффициентом корреляции 0.90. Шкала “95-105” – это шкала статистических энергий переноса для аминокислот из ядра белка ( $\epsilon^{sch} = 0$ ) в фазу с относительной доступностью боковых цепей в диапазоне  $95 < \epsilon^{sch} < 105$ . Точки для этой шкалы изображены символами согласно обозначению на одной вертикали с однобуквенным кодом. Точечная линия – это аппроксимация линейной зависимости с коэффициентом корреляции 0.84.

Таблица 4.4. Три шкалы статистических энергий переноса боковых цепей аминокислотных остатков из ядра белка на его поверхность для виртуальных фаз с различной экспонированностью (см. текст) и коэффициенты корреляции с экспериментальными шкалами (см. табл. 4.2).

	“10-20”	“50-60”	“95-105”
GLY	0.855	0.828	-0.565
LEU	0.313	1.507	2.302
ILE	0.516	1.738	2.66
VAL	0.661	1.571	2.138
ALA	0.848	1.097	0.728
PHE	-0.017	1.465	2.314
CYS	0.779	2.292	3.171
MET	0.355	1.262	1.94
THR	-0.249	-0.315	-0.061
SER	-0.147	-0.28	-0.413
TRP	-1.11	.739	2.293
TYR	-1.168	0.116	1.397
GLN	1.335	-1.884	-1.507
LYS	-1.989	-3.22	-2.893
ASN	-0.944	-1.267	-1.59
GLU	-1.478	-2.434	-2.392
HIS	-1.224	-0.937	-0.482
ASP	-1.162	-1.81	-2.101
ARG	-2.284	-2.62	-1.426
V>W	<b>0.93</b>	0.87	0.71
CH>W	0.80	<b>0.90</b>	0.84
O>W	0.57	0.85	<b>0.93</b>
O>W <sub>occ</sub>	0.53	0.79	0.87
CH>O	0.83	0.85	0.74

Если говорить об общей предсказательной силе экспериментальных шкал и их применимости в QSAR исследованиях, то шкала вода/октанол может иметь некоторое преимущество среди других, хотя, как видно из таблицы 4.4, это предпочтение сильно зависит от используемой статистики.

#### 4.8. Факторы, влияющие на распределения аминокислот

В заключении приведём некоторые теоретические рассуждения относительно причин, по которым не достигается идеальной корреляции между экспериментальными и статистическими шкалами, а также классифицируем эти эффекты. Предположим, что статистическая свободная энергия переноса, определяемая согласно формулам 4.3, 4.4 может быть интерпретирована, как  $\Delta_E$  из теории Финкельштейна, то есть средняя разница (усреднённая по всем возможным стабильным укладкам) свободной энергии между различными укладками с экспонированной и погружённой боковой цепью аминокислоты. Очевидно, что на величину  $\Delta_E$  будут влиять компоненты, не только связанные со взаимодействием боковой цепи с растворителем и внутренностью белка, но и различные другие компоненты. Разделим статистическую свободную энергию переноса на несколько интуитивных феноменологических компонент, которые будем использовать для дальнейшей интерпретации статистических данных:

$$\Delta_E = F_t^{sch} + F_t^{bb} + F_{ii}^{sch} + F_{conf} + F_{ss} + F_{func} \dots \quad (4.6)$$

$F_t^{sch}$  – это гипотетическая энергия переноса боковой цепи с поверхности в ядро белка, эта величина должна наилучшим образом соответствовать некоторой экспериментальной шкале переноса соединений между водой и гипотетическим растворителем по полярности похожим на внутренность белка (дебаты о котором давно идут в литературе). Все остальные компоненты в уравнении 4.6, это компоненты, которые приводят к отклонению  $\Delta_E$  от наилучшей корреляции с экспериментальными энергиями переноса.  $F_t^{bb}$  – энергия переноса атомов остова,  $F_{ii}$  – член отвечающий за специфические взаимодействия боковой цепи и интерьера белка (солевые мостики между заряженными остатками, дисульфидные связи и т.д.),  $F_{conf}$  – член, учитывающий влияние размера боковой цепи на конформационную подвижность пептидной цепи (особенно важен для глицина),  $F_{ss}$  – член, учитывающий влияние предпочтительного нахождения аминокислотного остатка в тех или иных элементах вторичной структуры (например, остатки умеющие образовывать дополнительные водородные связи (SER, ASN, ASP), являются разрушителями альфа-спиралей, что увеличивает их доступность, так как им остаётся располагаться в петлях и циклах);  $F_{func}$

– учитывает возможную связь расположения аминокислот с функцией белка, всё-таки белки хотя и являются квази-случайными последовательностями, но до тех пор пока это не затрагивает их функцию (этот член может быть важен для гистидина, который часто находится в каталитических сайтах).

## 4.9. Выводы к главе

Анализ статистики распределения аминокислотных остатков между поверхностью и ядром в водорастворимых глобулярных белках представляет как фундаментальный интерес, так и практический интерес с точки зрения выявления статистических потенциалов взаимодействия аминокислотных остатков, их корреляций с экспериментальными шкалами гидрофобности и применения результатов для совершенствования различных вычислительных методик в области предсказания пространственной структуры белков, оценки различных энергетических, других характеристик методами соотношения структура-свойство, а также дизайна биомиметических сополимеров.

В данной работе был проанализирован большой набор (ок. 8000) экспериментальных пространственных структур глобулярных белков. Для каждого белка рассчитана и проанализирована поверхность доступная растворителю. Для каждого аминокислотного остатка в каждом белке также определена поверхность доступная растворителю и величина его относительной доступности (экспонированности). Рассчитанные данные были использованы для построения статистических профилей доступности аминокислотных остатков, на основании которых, применяя различные критерии разделения аминокислот на “ядерные” и “поверхностные”, были рассчитаны статистические свободные энергии переноса. Данные энергии сравнивались с различными экспериментальными шкалами гидрофобности, основанными на распределении аминокислот между различными растворителями. Было показано, что в зависимости от критерия разделения, возможны хорошие корреляции с различными экспериментальными шкалами. Были введены три различных критерия определения поверхностных аминокислот (“10-20”, “50-60” и “95-105”), для которых рассчитаны статистические свободные энергии переноса. Проанализирована также зависимость исследуемых величин от размера белков и использован набор белков, где зависимостью от размера можно было пренебречь.

Работа является далеко не первой в области исследования статистических распределений аминокислот между поверхностью и ядром белка, однако, в данной работе достиг-



нут существенный прогресс по следующим аспектам: (i) создан и использован специализированный набор белков в 10 раз больший, чем в аналогичных работах, (ii) подробно рассмотрено влияние размера белков на статистику, и получены данные независимые от размера белков, (iii) построены статистические профили распределения аминокислот в зависимости от экспонированности, (iv) использовано 3 различных критерия для разделения остатков на поверхностные и ядерные, (iv) проанализированы корреляции с 5 экспериментальными шкалами гидрофобности .

Кроме важных численных и статистических результатов, представленных в работе в виде таблиц, по результатам работы можно сделать следующие общие выводы обладающие научной новизной:

- Статистика распределения аминокислот между поверхностью и ядром белка зависит от размера (длины аминокислотной последовательности) белка. Данная зависимость выражена при небольших размерах и сходит на нет при размерах больших 300 аминокислотных остатков, что связано с формированием доменной структуры белков.
- Для небольших белков как распределение между поверхностью и ядром, так и встречаемость различных типов аминокислот зависит от размера белка. С ростом размера белка эта зависимость выходит на насыщение. Причина такого поведения, во-первых, связана с доменной структурой больших белков и, во-вторых, с тем, что по мере увеличения размера белка происходит перераспределение баланса между различными типами взаимодействий ответственными за его стабильность. Так для маленьких белков гидрофобного взаимодействия ещё не достаточно для стабилизации структуры, и важную роль играют дополнительные взаимодействия такие, как дисульфидные связи, солевые мостики и т.д. Этим фактом, в частности, объясняется повышенное содержание цистеина в маленьких белках.
- Корреляции между свободными энергиями переноса аминокислот между различными растворителями и статистическими свободными энергиями переноса зависят от критерия разделения остатков на поверхностные и ядерные, а кроме того могут быть весьма различными для подгрупп остатков сходных по своим свойствам (гидрофобные, полярные и т.д.).
- При изучении корреляций между экспериментальными шкалами гидрофобности и статистическими энергиями, полученными при анализе белковых структур (для всего набора аминокислот), оказывается, что (i) при рассмотрении переноса между ядром белка и мало экспонированными остатками наилучшие корреляции достигаются

с энергией гидратации аминокислот, а (ii) при рассмотрении переноса между ядром белка и сильно экспонированными остатками (доступность больше 80%) наилучшие корреляции достигаются со шкалой октанол/вода.

- Как статистические, так и экспериментальные шкалы можно характеризовать по равномерности распределения веществ на их шкале. Равномерность среди экспериментальных шкал выше для шкал переноса между растворителями сходной полярности. Среди статистических шкал равномерность выше у шкал переноса из ядра в более поверхностные слои, чем в менее поверхностные.

## Глава 5

# Самоорганизующиеся тиафен-пептидные диблок олигомеры

В данной главе описана работа по моделированию самоорганизующихся соединений на основе тиафен-пептидных диблок олигомеров. Описанные соединения были синтезированы и исследованы в группе профессора Бауэрли, Университет города Ульм, Германия. Показано, что они могут образовывать волокнистые агрегаты. Настоящая работа посвящена попытке предсказания и анализу молекулярной структуры наблюдаемых агрегатов.

## 5.1. Введение

Как уже упоминалось в первой главе, уникальная способность природных биополимеров к самоорганизации и самосборке в определённые пространственные структуры и комплексы может использоваться для конструирования искусственных гибридных соединений, также способных к самоорганизации. Такие гибридные соединения обычно содержат специальным образом сконструированные последовательности биополимеров и ковалентно пришитые фрагменты синтетических полимеров, обладающих необходимыми свойствами и функциями. В качестве биологического компонента в таких системах наиболее часто выступают различные пептидные последовательности, особый интерес представляют пептиды способные к самоорганизации в длинные фибриллярные агрегаты, аналогичные, амилоидным фибриллам, формирующимся в тканях человека и животных при ряде широко распространённых заболеваний (таких, как болезнь Альцгеймера и др.). Яркий пример применения такого гибридного подхода для дизайна новых соединений в области органической наноэлектроники представляют диблок олигомеры тиафена и аминокислотной последовательности [Thr-Val]<sub>3</sub>. Пептидная последовательность состоит из чередующихся аминокислотных остатков треонина и валина и известна своей способностью к формированию вторичной структуры бета-листов, а также образованию фибриллярных агрегатов. Политиафен является полупроводящим полимером, который широко используется в органической микроэлектронике, в частности, для создания органических полупроводниковых светодиодов, а также различного рода сенсоров. Создание гибридных биосинтетических структур может вывести процесс создания наноэлектронных устройств на принципиально новый уровень благодаря самоорганизационным свойствам таких соединений.

В работе [125] был синтезирован новый тип соединения – диблок олигомер тиафена и пеп-

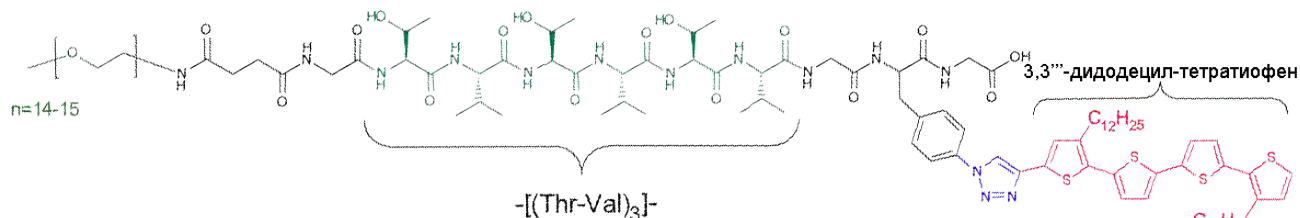


Рис. 5.1. Химическая формула синтезированной молекулы в работе [125].

тидной последовательности [Thr-Val]<sub>3</sub>, молекулярная структура которого представлена на рис. 5.1. Структура была создана путём соединения симметрично замещённого дидодецил-тетратиофена, функционализированного с одной стороны ПЭГ-пептидной последовательностью, содержащей три повторения валина и треонина.

Ниже приведены экспериментальные данные о свойствах этого соединения.

#### 5.1.1. Экспериментальные данные

Проводилось изучение раствора синтезированного соединения, а также агрегатов, осаждаемых из раствора на подложку. Учитывая различные полярности разных частей гибрида (гидрофобные свойства тиофена и полярные свойства пептидной последовательности), для получения агрегатов данного соединения использовалась стратегия самосборки, управляемая растворителем (solvent-guided strategy). Для того, чтобы предотвратить преждевременную агрегацию молекул, в молекулы соединения были введены дефекты по типу сложных эфиров вместо пептидной связи между одним из треонинов и последующим валином. Данные дефекты обладают способностью необратимо реизомеризоваться в пептидную связь при повышении pH раствора. В результате постепенного добавления метанола и NaOH (концентрацией 0.001M) к раствору соединения в дихлорметане был получен раствор (50% ДХМ + 50% метанол), в котором по данным инфракрасной спектроскопии присутствовал сигнал, соответствующий агрегации пептидных структур в бета-слои. Определить был ли это параллельный или антипараллельный бета слой не удалось. В результате исследования агрегатов методом АСМ после переноса их на слюду методом осаждения на вращающейся подложке (spin-coating) были обнаружены фибриллярные агрегаты (см. рис. 5.2). При увеличении видно, что отдельные крупные фибриллы имеют полосатую структуру и состоят, предположительно, из более мелких филаментов. На заднем фоне видны более “мелкие” фибриллы, которые не входят в состав более больших и, возможно, аналогичны филаментам, являющимся составными частями более крупных фибрилл.

“Крупные” фибриллы достигали в длину 1-2 мкм и в высоту около  $1 \pm 0.2$  нм, ширина

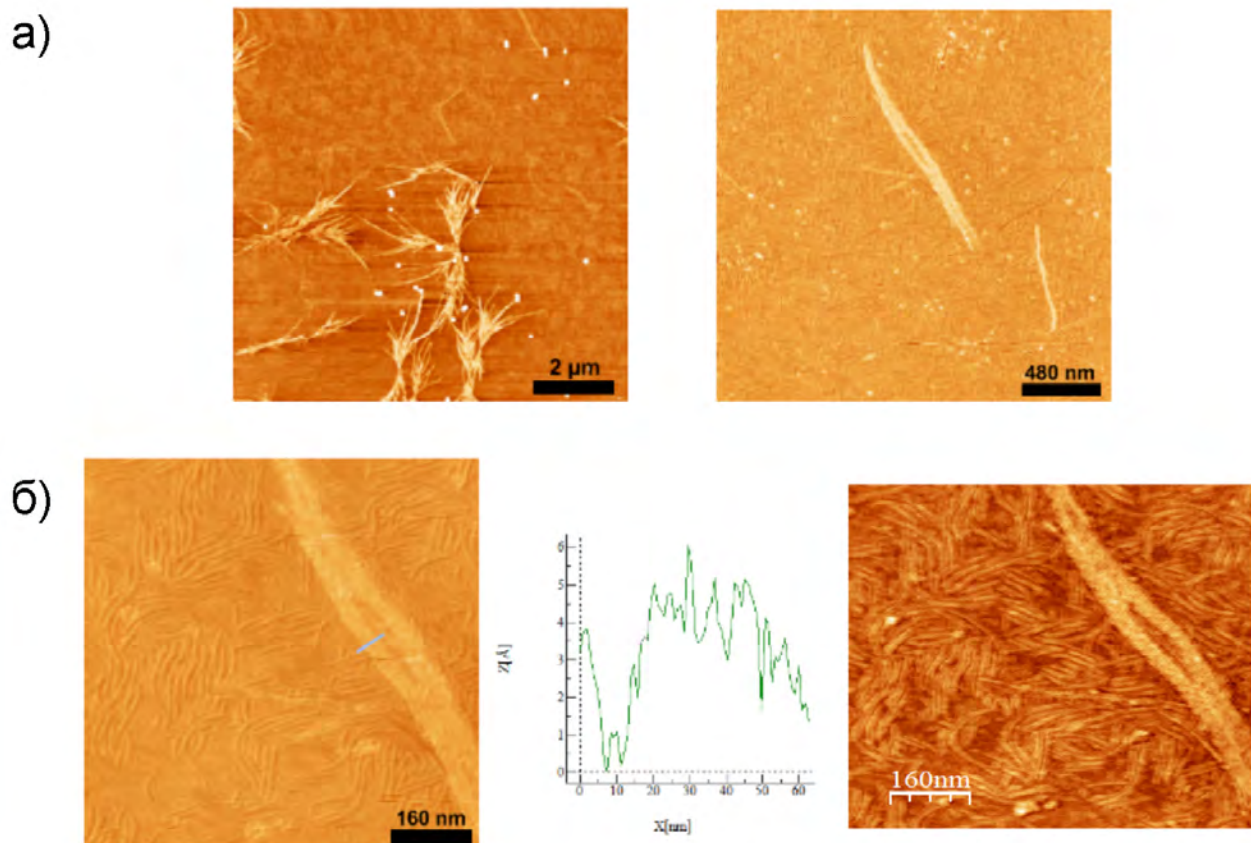


Рис. 5.2. АСМ: изображения высоты и амплитуды для гибрида тетратиофена и пептида на слюде. а) слева: АСМ изображение группы фибрилл; справа: увеличенное изображение одной фибриллы; б) слева: увеличенное изображение одной фибриллы, состоящей из более мелких филаментов; по-середине: график высоты среза фибриллы; справа: более контрастная картина, соответствующая рисунку слева, более чётко видна структура большой фибриллы.

фибрилл обычно находилась в диапазоне 15-48 нм, но могла достигать и 80 нм. “Мелкие” фибриллы обладали длиной 100-200 нм, шириной  $6-8 \pm 2$  нм и высотами  $0.4-0.7 \pm 0.2$  нм. Следует отметить, что определение высот фибрилл достаточно условное, поскольку они не находятся на чистой поверхности слюды, а находятся в слое осаждённого вещества. Существование фибриллярных агрегатов было подтверждено и при использовании ПЭМ (Рис. 5.3).

Отдельного обсуждения заслуживает вопрос соответствия агрегатов, наблюдаемых на поверхности, агрегатам в растворе. Набор исследований для аналогичных систем [36], включавший в себя криоэлектронную микроскопию и малоугловое нейтронное рассеяние, а также методы приготовления образцов (spin-coating) дают серьёзные основания полагать, что самосборка происходит, именно, в растворе.

Для раствора исследуемого соединения были проведены спектральные исследования на

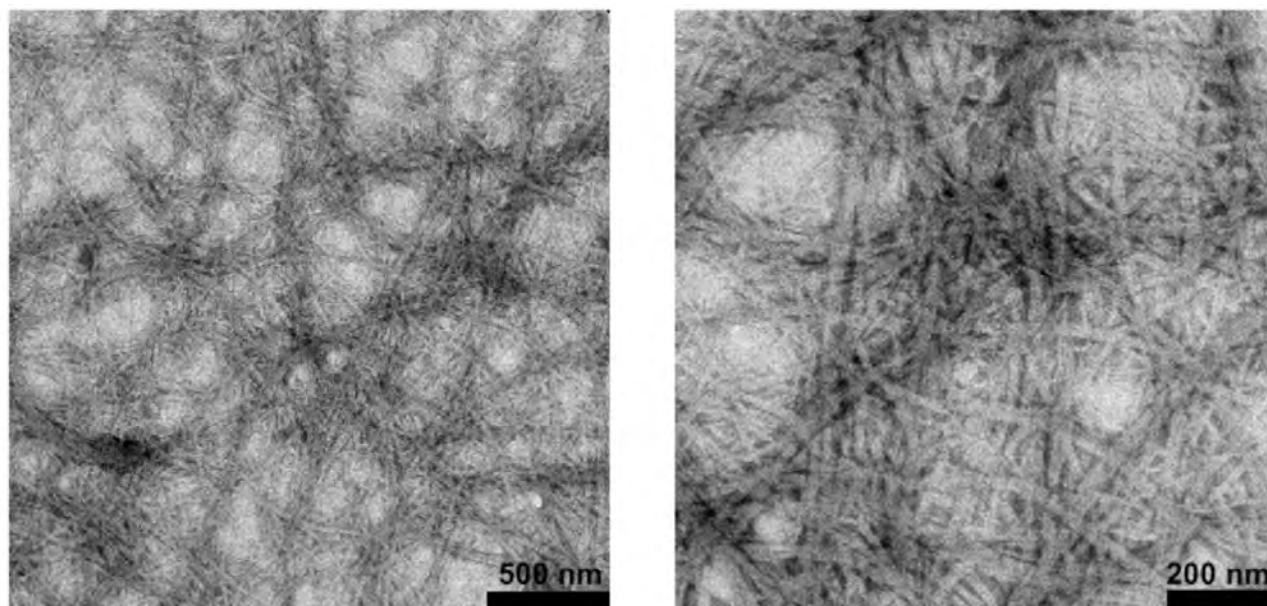


Рис. 5.3. ПЭМ изображения фибрилл.

предмет наличия  $\pi - \pi$ -стэкинга между тιοфеновыми группами молекул, поскольку известно, что тιοфены обладают способностью к самосборке благодаря  $\pi - \pi$ -стэкингу и силам Ван-дер-Ваальса. Были проведены измерения следующих величин: спектры поглощения в видимой и УФ области, спектры флюоресценции, спектры кругового дихроизма. Данные ни одного из методов не дают основания утверждать, что при агрегации возникает непосредственное взаимодействие между  $\pi$ -системами тιοфенов.

## 5.2. Постановка задачи

На данный момент для большинства самособирающихся в растворе биоконъюгатов не доступны экспериментальные методы, которые могли бы определить их внутреннее строение и структуру. В этой связи особую актуальность приобретают методы компьютерного моделирования, которые могут использоваться как для того, чтобы интерпретировать и создавать модели для уже имеющихся экспериментальных данных, так и для понимания принципов самоорганизации таких систем и дизайна на их основе новых соединений с заданными свойствами.

В данной работе задача состоит в интерпретации имеющихся экспериментальных данных, построению возможных моделей самоорганизации исследуемого соединения путём атомистического компьютерного моделирования.

## 5.3. Рациональный компьютерный дизайн моделей

### 5.3.1. Суть подхода

Золотой целью компьютерного моделирования при рассмотрении самоорганизации молекулярных структур мог бы быть метод, который позволяет однозначно предсказывать трёхмерную структуру агрегатов, зная атомистические модели составляющих их молекул и взаимодействия между ними. Однако, на данный момент, даже если предположить, что все межмолекулярные и межатомные взаимодействия описываются точно, таких универсальных методов не существует. Наблюдение самоорганизации *ab initio* в процессе молекулярной динамики для биологических систем таких, как белки и подобных им, при доступных вычислительных мощностях практически невозможно. Предпринимаются лишь отдельные успешные попытки сворачивания небольших белков *in silico*, однако и для таких систем задействуются очень значительные вычислительные ресурсы, специальные алгоритмы и речь идёт в основном о попытках воспроизведения экспериментальных структур. Вышеозвученные проблемы приводят к тому, что приходится искать иные способы анализа и моделирования таких самоорганизующихся структур. Одна из основных идей таких способов состоит в задании определённой начальной конфигурации молекулярной системы, исходя из некоторых дополнительных данных или рациональных принципов, и последующего анализа поведения и стабильности такой системы методами компьютерного моделирования, в частности, молекулярной динамики.

Один из таких подходов, основанный на предварительном предсказании возможных кристаллических упаковок молекул в рамках основных групп кристаллической симметрии, последующем отборе структур по некоторому рациональному критерию и использованию кристаллических упаковок для построения молекулярных агрегатов, был предпринят в ряде наших работ [3, 54] для анализа гибридных соединений содержащих пептидные последовательности.

Однако в случае самоорганизации молекул, рассматриваемых в данной работе, при наличии определённых экспериментальных данных, а также большого объёма экспериментальных данных по аналогичным системам и рациональных представлений о типе и силах межмолекулярных взаимодействий предпочтительным с нашей точки зрения оказался другой подход, предложенный и описанный в данной работе. Суть данного подхода состоит в последовательном конструировании молекулярных агрегатов с пошаговой оптимизацией структуры на основе представлений об иерархии взаимодействий присутствующих в таких системах и использовании экспериментальных данных при дизайне агрегатов. Проиллюстрируем суть подхода на одном из примеров данной работы – конструировании

двуслойного молекулярного агрегата. Сначала создаётся кристаллическая ячейка (ячейка с периодическими граничными условиями) из двух молекул, образующих бета-слой, с периодом 4.8 Å на один бета-лист (поскольку из эксперимента известно, что данные молекулы формируют укладку в виде бета-слоёв и известна периодичность бета-слоя), молекулы оптимизируются в периодической ячейке методами молекулярной механики и динамики, затем создаются различные двуслойные упаковки из 4 молекул, опять же находящиеся в периодических граничных условиях. Данные упаковки могут быть различных типов в зависимости от того, какими сторонами и в каком направлении соединены слои. Двуслойные агрегаты релаксируются методом молекулярной динамики, чтобы боковые цепи аминокислот нашли положение соответствующее минимуму свободной энергии. Далее среди различных двуслойных упаковок отбираются укладки, чья энтальпия образования минимальна. Наконец, периодическая ячейка, соответствующая выбранной укладке, “размножается” до необходимых размеров, и полученные агрегаты анализируются методом молекулярной динамики.

### 5.3.2. Модель молекулы и методы расчёта

В качестве начальной точки для создания компьютерной модели наблюдаемых нанофибрилл была сконструирована классическая молекулярно-механическая полноатомная модель гибридной молекулы, изображённая на рис. 5.4. Моделируемое соединение отличалось от того, которое использовалось в эксперименте отсутствием ПЭГ хвоста. ПЭГ-хвост был добавлен в соединение исключительно с точки зрения технических особенностей синтеза для увеличения растворимости соединения, и мало вероятно, что он играет структуроопределяющую роль.

Полноатомная модель соединения была создана на основе силового поля класса II PCFF (polymer-consistent force field), параметризованного для воспроизведения *ab initio* поверхностей потенциальной энергии [61, 126]. Суммарная потенциальная энергия системы представляется как сумма различных вкладов, которые подразделяются на две категории: (I) валентные члены  $U_{val}$ , включая энергетические вклады от связей, валентных углов, торсионных углов, ложноторсионных углов, а также энергетические вклады диагональных и недиагональных перекрёстных членов по внутренним координатам; (II) члены, отражающие внутри и межмолекулярные невалентные (non-bonded) взаимодействия  $U_{nb}$ , включая Ван-дер-Ваальсовы взаимодействия согласно потенциалу Леннарда-Джонса “9-6” и взаимодействия между зарядами по закону Кулона. Подробнее структура силового поля PCFF описана в главе 1. Такого количества вкладов достаточно, чтобы описывать “класси-





ный момент нет общедоступных силовых полей, которые бы явным образом учитывали такие взаимодействия, вводя дополнительные члены в потенциальную энергию, то модель тифеновой части молекулы тоже можно считать обоснованной, однако следует уделить особое внимание данной особенности модели при анализе и интерпретации результатов.

Моделирование в рамках методов молекулярной механики и динамики проводилось с использованием программного пакета LAMMPS [128] на суперкомпьютере СКИФ-МГУ. Моделирование проводилось в параллельном режиме с использованием метода *domain decomposition*, для больших систем использовался уровень распараллеливания от 64 до 128 процессоров.

### 5.3.3. Создание однослойных фибрилл

Большинство экспериментальных и теоретических данных указывают на то, что образование фибриллярных агрегатов из молекул, содержащих аминокислотные последовательности, склонные к образованию бета-листов, происходит благодаря агрегации молекул в своеобразные ленты, состоящие из бета-листов, уложенных перпендикулярно направлению роста ленты. Эти ленты в свою очередь могут участвовать в формировании более сложной структуры агрегатов. По аналогии с номенклатурой структуры белков, укладку аминокислотных последовательностей в ленты можно назвать вторичной структурой агрегатов, тогда как дальнейшие уровни организации, связанные с поведением лент, возможными вариантами их агрегации – третичной и четвертичной структурами. Укладка молекул в ленты может соответствовать либо параллельному бета-слою, либо антипараллельному. Считается, что вариант упаковки, соответствующий антипараллельному бета-слою (здесь речь идёт о белках) энергетически более выгоден, чем вариант параллельной укладки. Тем не менее в большинстве экспериментально разрешённых структур амилоидоподобных фибрилл [27, 129] преобладает структура параллельного бета слоя. Задача конструирования однослойных фибрилл из исследуемой молекулы таким образом сводится к конструированию упаковки, в которой пептидные хвосты молекул будут формировать параллельный или антипараллельный бета-слой.

В ходе работы было создано два типа упаковки молекул для конструирования однослойных фибрилл: на основе параллельного и антипараллельного бета-слоя. Задание начальной упаковки молекул в фибрилле было реализовано путём создания кристаллической ячейки из двух молекул. Фактически, значение имела лишь периодичность вдоль оси фибриллы, путём размножения ячейки по данному направлению можно получить начальную структуры фибриллы с заданной организацией пептидных сегментов.

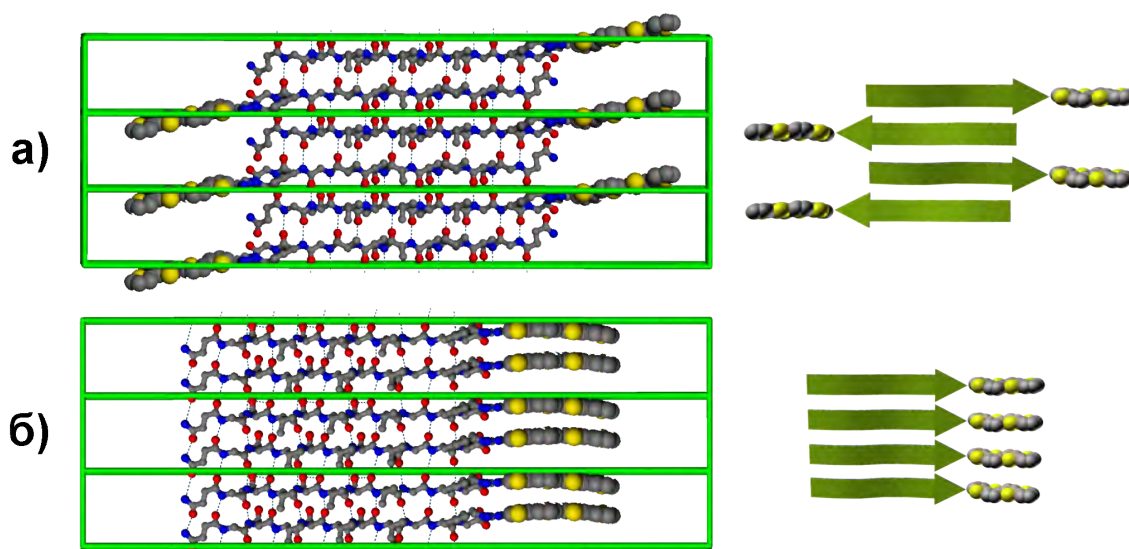


Рис. 5.5. Схематическое изображение упаковок для однослойных фибрилл (справа) и соответствующие им периодические ячейки из двух молекул (слева): а) вариант антипараллельной упаковки пептидных частей, б) вариант параллельной упаковки пептидных частей.

Два типа упаковок были созданы следующим образом. Исходная конформация молекулы была отрегулирована следующим образом: (I) конформация пептидного блока молекулы была приведена в соответствие (путём варьирования соответствующих торсионных углов пептидного остова) с конформациями соответствующими идеальному антипараллельному ( $\phi = -139^\circ$ ,  $\psi = 135^\circ$ ) или параллельному ( $\phi = -119^\circ$ ,  $\psi = 113^\circ$ ) бета-листу; (II) тиюфеновый блок, включая боковую цепь 4-азидофенилаланина и алкильные цепи брались в достаточно плоской вытянутой конформации, отвечающей локальному минимуму энергии (см. рис. 5.4, в). Две молекулы помещались в периодическую ячейку так, чтобы образовывался циклический бета-слой (см. рис. 5.5). В случае параллельной упаковки молекулы располагались чётко одна над другой, в случае антипараллельной упаковки пептидные сегменты выравнивались в плоскости перпендикулярной оси ленты так, чтобы количество водородных связей, образуемых между пептидными сегментами было максимальным. Период кристаллической ячейки вдоль роста ленты на данном этапе фиксировался на величине 9.6 Å, поскольку экспериментально известное расстояние между бета-листами в амилоидных фибриллах составляет порядка 4.8-4.9 Å[27, 129].

Далее для сконструированных молекулярных структур была проведена процедура минимизации энергии, а затем процедура релаксации методом молекулярной динамики в течение 1 нс. При минимизации и расчёте методом МД этих периодических систем использовались следующие параметры: радиус обрезания взаимодействий Ван-дер-Ваальса 10 Å, учёт электростатических взаимодействий производился с использованием метода

PPPM с радиусом обрезания в прямом пространстве  $10 \text{ \AA}$  и толерантностью  $10^{-4}$ . Шаг интегрирования в методе МД составлял 1 фс, расчёты проводились в NVT ансамбле при температуре 300 К, который был реализован за счёт добавления членов стохастической силы и силы трения в уравнения движения согласно уравнениям Ланжевена. Обратная константа трения была выбрана равной 1 пс.

#### 5.3.4. Создание двуслойных фибрилл и оценка энтальпии образования

Полученные периодические ячейки возможных однослойных упаковок фибрилл (в отрелаксированном состоянии) использовались далее для создания различных двуслойных упаковок, то есть упаковок, в которых основным структурным элементом являются два бета-листа, соприкасающиеся различными сторонами. Рассматривались четыре основных варианта двуслойных упаковок (см. рис. 5.6), в которых бета-листы соприкасались либо гидрофобными, либо гидрофобными сторонами. Преимущественное формирование гомогенных контактов (соприкосновение только гидрофобных или гидрофильных сторон листов) разумно предполагать, исходя как из энергетических принципов, так и из наблюдаемых упаковок бета-листов в белках.

При создании двуслойных упаковок была выбрана стратегия ручного конструирования начальных структур таким образом, чтобы они были по своему строению похожи на экспериментально наблюдаемое расположение молекул в амилоидо-подобных фибриллах методами рентгено-структурного анализа и ЯМР [27, 129], с последующей значительной релаксацией систем методом молекулярной динамики, которая позволяла молекулам занять конформацию, отвечающую локальному минимуму свободной энергии в заданном периодическом расположении.

Двуслойные упаковки создавались следующим образом. Для однослойной упаковки создавался её дубликат, который разворачивался на  $180^\circ$  относительно оси фибриллы, а затем выравнивался в латеральной плоскости так, чтобы пептидные сегменты двух однослойных упаковок располагались друг напротив друга (как это обычно наблюдается в экспериментальных структурах амилоидных фибрилл), формируя центр фибриллы (см. рис. 5.6). Выравнивание в латеральной плоскости возможно двумя принципиальными способами: (i) бета-листы соприкасаются поверхностями с экспонированными боковыми цепями треонинов или (ii) валинов. Пересечения фрагментов (например, алкильных цепей), если такие возникали, разрешались вручную с помощью незначительной регулировки торсионных углов, которые в любом случае принимают свои равновесные значения на этапе релаксации.

Рассмотрим подробнее свойства упаковок, представленных на рис. 5.6. В основе упаковки (I) находится два параллельных бета-слоя, сложенных гидрофобными сторонами (боковые цепи аминокислоты валин), тиюфеновые сегменты формируют стопку с каждой стороны фибриллы, при этом в каждой стопке участвуют лишь сегменты, принадлежащие одному из бета-слоёв, таким образом взаимодействие между слоями происходит лишь за счёт взаимодействия пептидных сегментов. Формированию гидрофобной середины между бета-слоями, очевидно, может способствовать наличие полярного растворителя. Упаковка (III) структурно аналогична упаковке (I), однако параллельные бета-слои сложены гидрофильными сторонами (боковые цепи аминокислоты треонин), такому типу агрегации может благоприятствовать неполярный растворитель. В упаковке (II) реализуется гидрофобный контакт между бета-слоями, однако организация самих слоёв – антипараллельная. Это приводит к тому, что тиюфеновые стопки по обоим сторонам фибриллы формируются поочерёдно чередующимися тиюфеновыми сегментами, принадлежащими различными бета-слоям. С учётом того, что тиюфеновые сегменты прикреплены к гидрофобной стороне бета-листа (к боковой цепи фенилаланина), тиюфеновые сегменты одного листа располагаются чётко между тиюфеновыми сегментами другого листа, формируя своеобразную “молнию”, которая, очевидно, дополнительным образом укрепляет взаимодействия между однослойными структурами. В упаковке (IV) в отличие от упаковки (III) бета-слои соприкасаются гидрофильными сторонами, но из-за геометрического расположения тиюфеновых сегментов (они направлены в разные стороны относительно плоскости бета-слоёв), стерического взаимодействия не образуется. Поскольку такое взаимодействие может играть существенную роль в стабильности таких структур, нами был рассмотрен ещё один тип организации молекул в двуслойный агрегат (упаковка (V)), в этой упаковке конформация пептидной цепи была изменена путём вращения торсионного угла  $\phi$  при фенилаланине таким образом, чтобы тиюфеновые сегменты были направлены в другую сторону и образовывали структуру подобную “молнии” в упаковке (IV). При таком изменении конформации пептидной цепи, однако, теряются некоторые водородные связи, участвующие в подкреплении структуры бета-слоёв, что может быть энергетически невыгодно.

Поскольку двуслойные упаковки в конформационном плане намного более сложные структуры, чем однослойные, в особенности, в плане организации контактов между двумя бета-слоями, где соприкасаются конформационно подвижные боковые цепи аминокислотных остатков, для релаксации данных структур был проведён длинный расчёт в 10 нс. Во время релаксационного расчёта методом МД происходило некоторое перестроение боко-

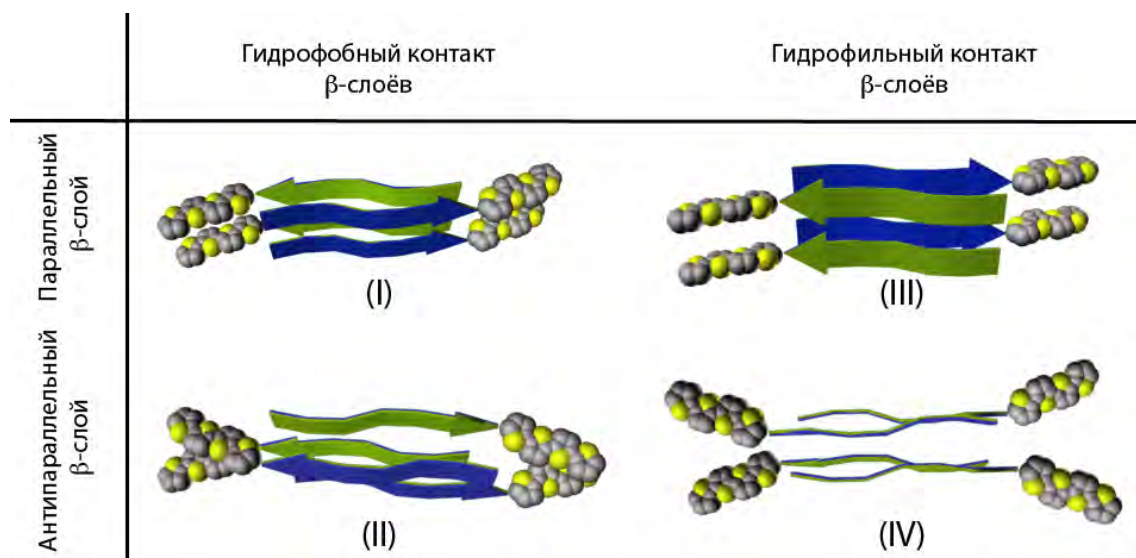


Рис. 5.6. Схематическое изображение возможных упаковок двуслойных фибрилл.

вых цепей в зоне соприкосновения бета-листов. Во время расчёта проводилось наблюдение за энергией системы.

Средняя энергия таких упаковок (по сути относительная энтальпия образования систем) является одним из критериев, по которому можно судить о вероятности реализации в реальности той или иной упаковки. Безусловно, отсутствие растворителя в данных моделях, делает их не совсем реалистичными, однако тем не менее такие оценки позволяют лучше понять баланс энергетических взаимодействий, который ответственен за образование тех или иных структур. Для пяти вышеописанных упаковок относительные энтальпии образования приведены в таблице 5.1. По результатам сравнения относительных энергий различных упаковок можно сделать следующие выводы. При сравнении двуслойных упаковок, составленных из параллельных слоёв, более энергетически выигрышной оказывается укладка с взаимодействующими гидрофильными поверхностями, что вполне ожидаемо. Среди упаковок, основанных на антипараллельных бета-листах, минимальной энергией обладает укладка с гидрофобной серединой. Можно предположить, что несмотря на то, что взаимодействие гидрофильных сторон более энергетически выгодно, оно не может компенсировать потери от отсутствия взаимодействия тиофеновых сегментов в случае упаковки (IV) и потери водородных связей вдоль бета-листа в упаковке (V). Для построения фибрилл и дальнейшего моделирования были отобраны упаковки (II) и (III) поскольку, во-первых, они обладают наименьшей энтальпией образования среди всех рассмотренных упаковок, а, во-вторых, позволяют рассмотреть поведение фибрилл на основе как параллельных, так и антипараллельных бета-листов.

Таблица 5.1. Относительные энтальпии образования различных упаковок (см. Рис. 5.6) двуслойных агрегатов.

Тип упаковки	Средняя энергия (ккал/моль) $\pm$ стандартное отклонение
(I)	$4481 \pm 1$
(II)	$4436 \pm 3$
(III)	$4463 \pm 4$
(IV)	$4483 \pm 1.3$
(V)	$4494 \pm 2$

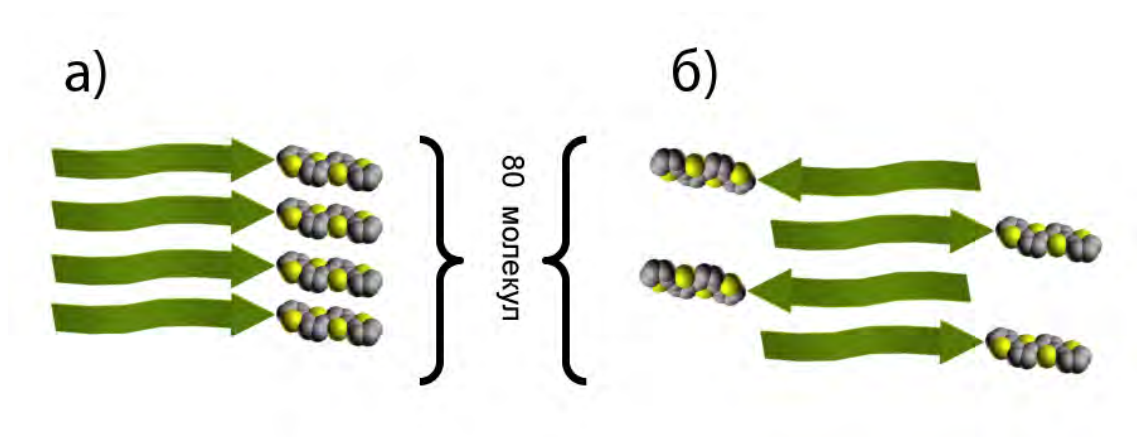


Рис. 5.7. Схематическое изображение начального состояния двух типов однослойных фибриллярных агрегатов: а) на основе параллельного бета-слоя, б) на основе антипараллельного бета-слоя.

#### 5.4. Моделирование однослойных фибрилл в объёме

Полученные однослойные упаковки были размножены в 40 раз вдоль основной оси для получения фибриллярных агрегатов длиной в 80 бета листов (см. рис. 5.7). Длина начальной конформации таких агрегатов составляла приблизительно 40 нм.

Для собранных агрегатов было проведено моделирование методом молекулярной динамики в течение 10 нс. Моделируемые системы были достаточно большими около 20 тысяч атомов (252 атома в каждой молекуле), что потребовало использование методов параллельного моделирования, использовался метод доменного разделения системы по процессорам, моделирование проводилось одновременно на 64 процессорах. Поскольку система являлась не периодической, невалентные взаимодействия (электростатические взаимодействия и взаимодействия Ван-дер-Ваальса) учитывались с радиусом обрезания 1.5 нм. NVT ансамбль при температуре 300K реализовывался благодаря использованию алгоритма Ланжевена с константой релаксации 2 пс.

#### 5.4.1. Визуальное описание

В течение первой наносекунды моделирования произошла релаксация быстрых степеней свободы фибрилл, в частности, расстояний между бета-листами (которое ранее должно было удовлетворять периодическим граничным условиям), конформации алкильных цепей и т.д. На более большом масштабе времени наблюдались крупномасштабные изменения геометрии самих фибрилл за счёт межмолекулярного переупорядочения. В течение времени моделирования (10 нс) вряд ли можно говорить о полной релаксации системы и возможности наблюдения её равновесных характеристик, поскольку времена корреляции крупномасштабных флуктуаций конформации фибрилл достаточно велики. Однако такое моделирование позволяет изучить основные принципы самоорганизации, сделать выводы о стабильности тех или иных упаковок, а также изучить основные силы и тенденции возникающие в той или иной упаковке молекул.

Мгновенные снимки начальной и конечной конформации фибрилл приведены на Рис. ???. Фибрилла, основанная на параллельной упаковке пептидных сегментов в бета-слой, в течение моделирования сохранила относительно вытянутую конформацию, однако, вдоль всей фибриллы произошло конформационное перестроение составляющих её молекул. Отчётливо видно, что пептидные сегменты молекул потеряли свою планарную конформацию, каковую они имеют в бета-слое, а некоторым образом “сложились”, однако при этом сохранили своё упорядочение. Изгиб пептидного сегмента наблюдается ближе к тиюфеновому блоку в районе фенилаланина. Кроме того, наблюдается левое закручивание фибриллы вдоль своей оси. Поведение фибриллы, основанной на антипараллельной упаковке бета-слоя, которая тоже сохраняет ближний порядок в исходном упорядочении, серьёзно отличается: бета-листы в среднем сохраняют свою планарность, при этом фибрилла в целом формирует лево закрученную спираль, которая в свою очередь закручивается в левую суперспираль. Этот факт принципиально отличает поведение данного типа укладки от параллельной укладки.

#### 5.4.2. Упорядочение пептидных частей

Чтобы подробнее охарактеризовать порядок молекул в фибриллах были построены гистограммы расстояний между центрами масс гексапептидов (Thr-Val) (см. рис. 5.10), расстояние между центрами масс этих сегментов (для антипараллельной упаковки между чётными/нечётными сегментами) по сути является оценкой периода бета-слоя. Гистограм-



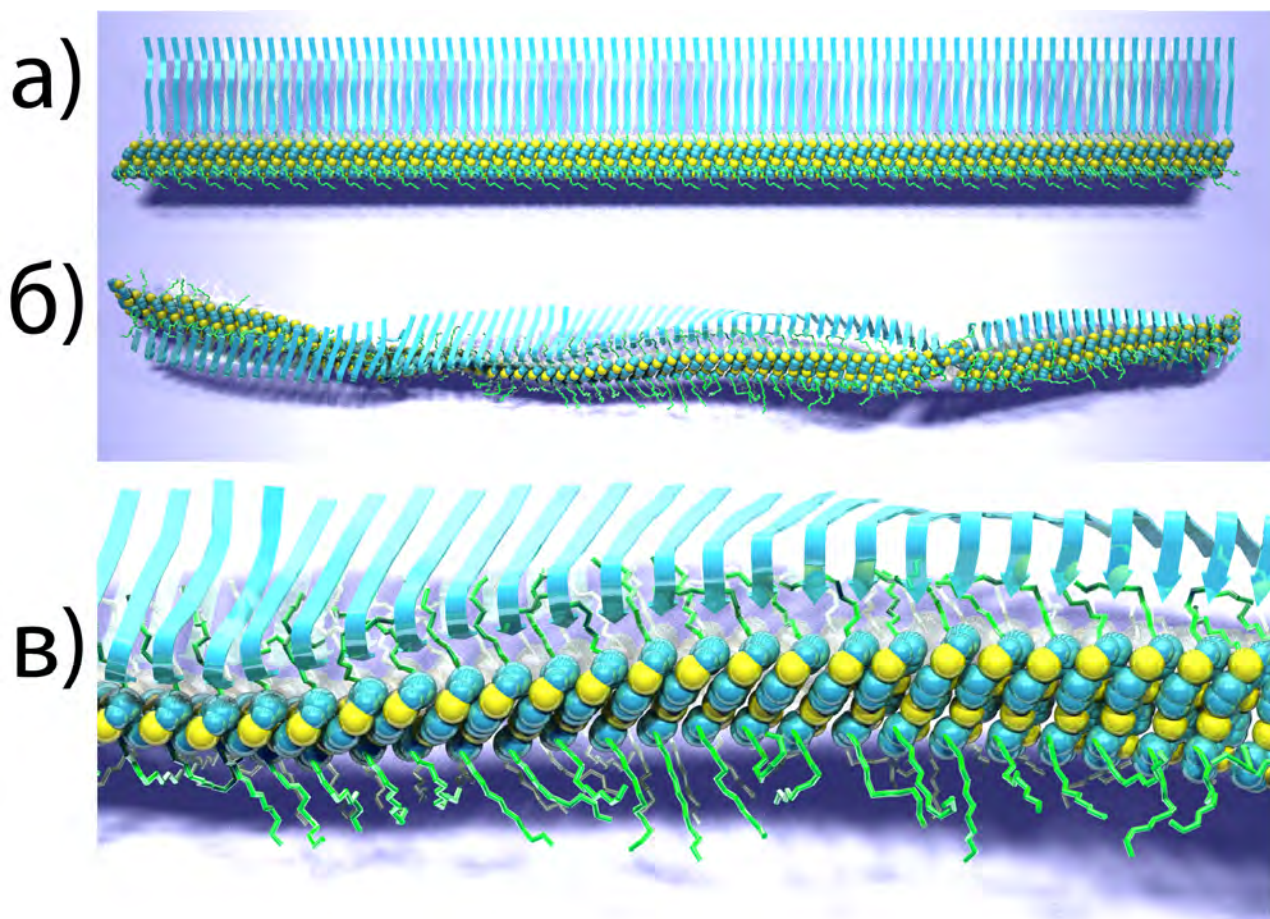


Рис. 5.8. Мгновенные изображения однослойного волокна, основанного на *параллельном* бета-слое, в ходе расчётов молекулярной динамики: а) начальная конформация, созданная из полученной ранее периодической укладки, б) система через 10 нс моделирования, в) увеличенный участок рисунка б). Пояснения к рисунку: пептидные сегменты изображены в виде листов, атомы тиафеновых сегментов в виде Ван-дер-Ваальсовых сфер (сера - жёлтый, углерод - сине-зелёный), алкильные цепи в виде зелёных трубочек.

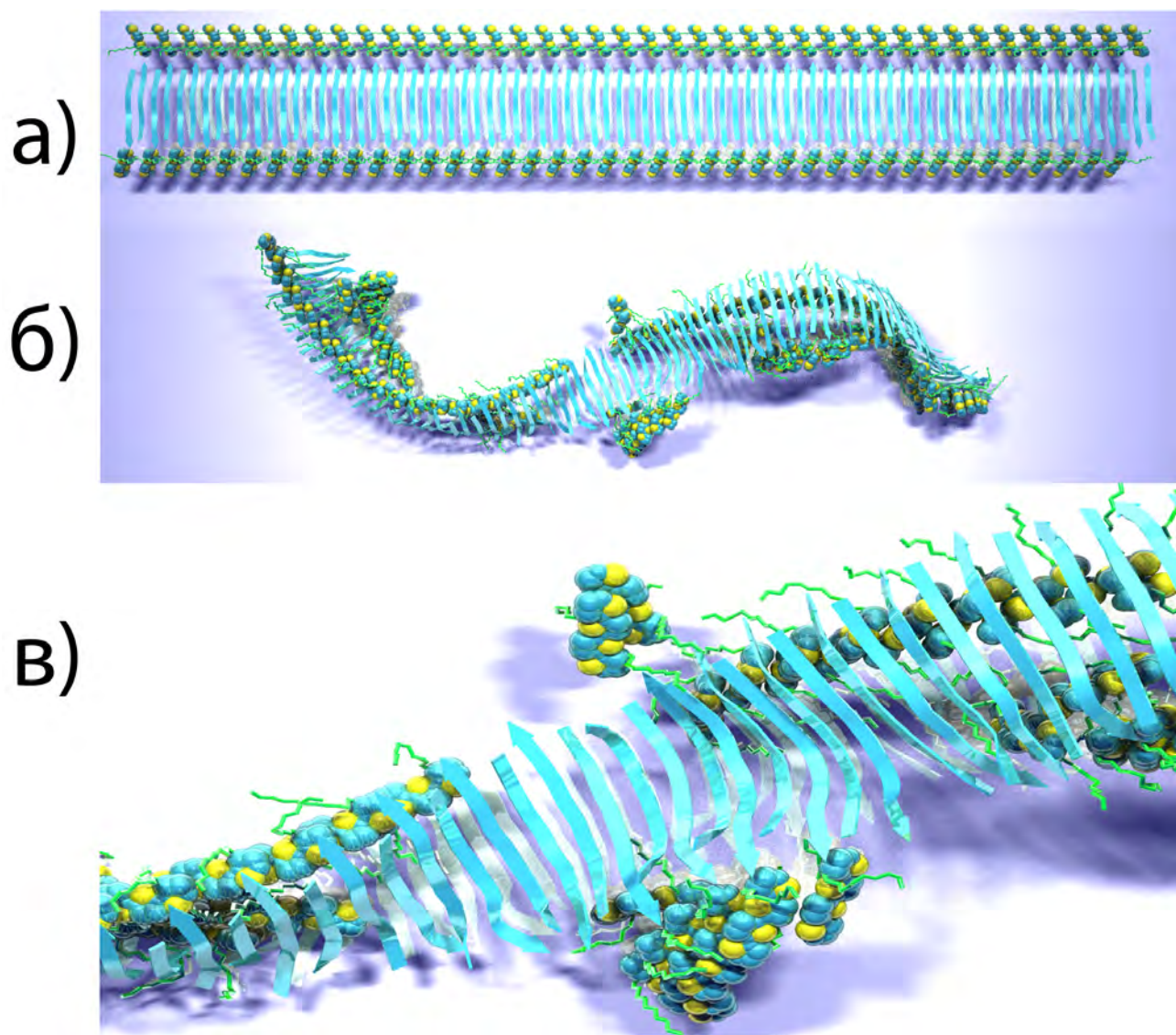


Рис. 5.9. Мгновенные изображения однослойного волокна, основанного на *антипараллельном* бета-слое, в ходе расчётов молекулярной динамики: а) начальная конформация, созданная из полученной ранее периодической укладки, б) система через 10 нс моделирования, в) увеличенный участок рисунка б). (см. также пояснения к предыдущему рисунку).

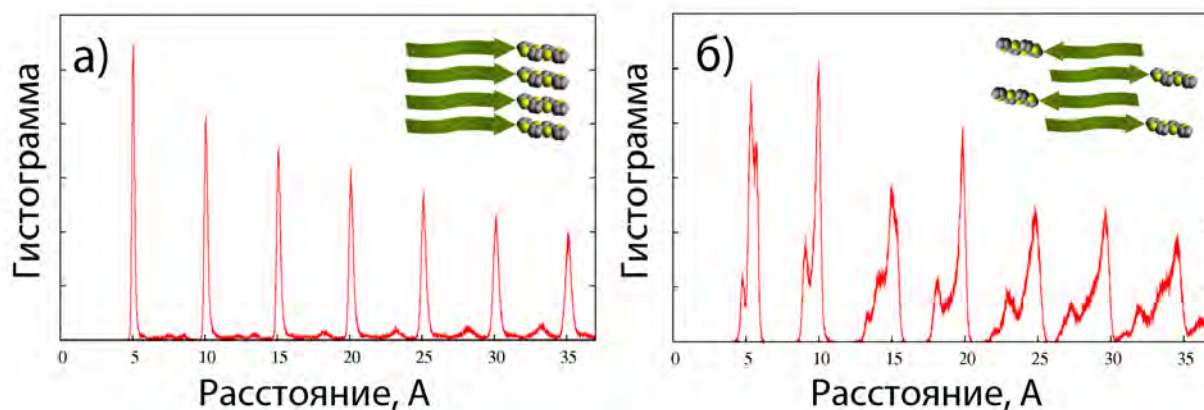


Рис. 5.10. Гистограммы расстояний между центрами масс пептидных последовательностей [Thr-Val]<sub>3</sub> каждой молекулы для однослойной фибриллы, полученные в ходе МД расчётов. а) случай параллельной укладки, б) случай антипараллельной укладки.

мы строились по последним 5 нс траектории. Гистограммы подтверждают ранее сделанные визуальные наблюдения: для параллельной упаковки наблюдается хорошее упорядочение молекул вдоль оси фибриллы, пики гистограммы достаточно острые; для антипараллельной упаковки наблюдается хороший ближний порядок, однако вдоль по цепи степень упорядоченности уменьшается быстрее, чем для параллельной упаковки, на практике, как мы уже видели, это проявляется в изгибах и закрученности фибриллы в суперспираль. Ещё один важный параметр, который можно извлечь из гистограмм, – это положение первого и второго максимума. Для антипараллельной и параллельной укладки второй большой максимум располагается на расстояниях 10.06 Å и 9.99 Å, соответственно. Отсюда делаем вывод о том, что период бета-слоя в нашей модели составляет ровно 5 Å, что очень хорошо согласуется с данными известными для бета-слоёв в белках [18].

Для более детального анализа вторичной структуры пептидных сегментов фибриллярных агрегатов был применён алгоритм STRIDE [130]. Процентный состав элементов вторичной структуры представлен на рис. 5.11. Для антипараллельной упаковки преобладают элементы бета-листов, что и следовало ожидать. Нужно также отметить, что длина всего пептидного хвоста составляла 10 аминокислот, из которых только 6 относились к последовательности [Thr-Val]<sub>3</sub>, склонной к образованию бета-структуры. Для параллельной же укладки доля бета-листов, определённых алгоритмом, оказалась намного меньше, однако, даже из визуального анализа видно, что структура пептидного остова данной фибриллы искажена и далека от канонической.



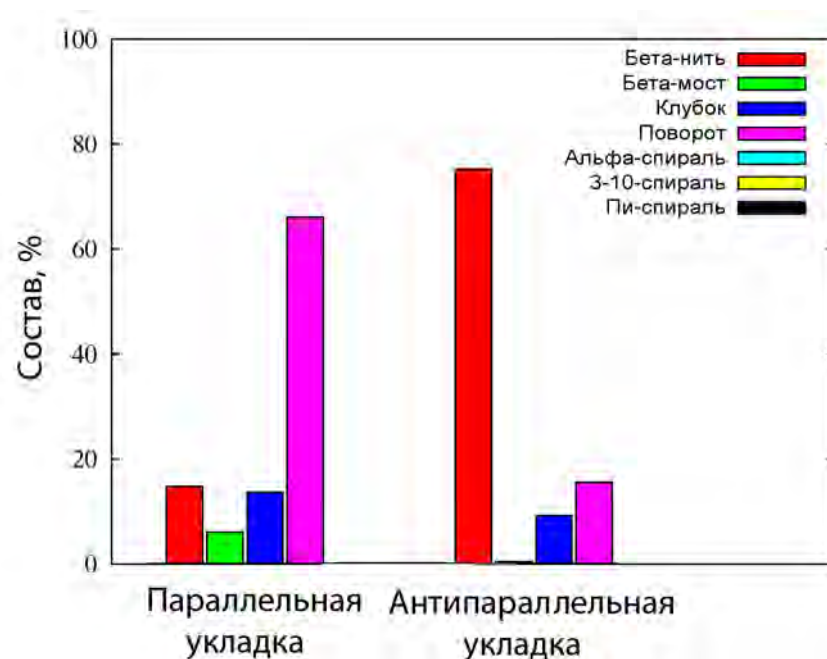


Рис. 5.11. Распределение аминокислот по элементам вторичной структуры в однослойных фибриллах, данные усреднены по последней наносекунде моделирования.

## 5.5. Моделирование на подложке, виртуальная АСМ

Для экспериментального исследования фибрилл важным методом является атомно-силовая микроскопия (АСМ), в этой связи актуальной является задача определения высоты и формы агрегатов адсорбированных на подложке. В качестве подложки в АСМ обычно используется либо слюда, либо высокоориентированный пиролитический графит (ВОПГ). В случае исследуемых молекул экспериментальные данные указывали на то, что самосборка фибрилл происходит в растворе, а в АСМ наблюдается лишь результат адсорбции уже собранных агрегатов. В тоже время актуальной является задача сравнения геометрических размеров агрегатов, наблюдаемых в АСМ и наших моделей. Для этого была собрана система - однослойный фибриллярный агрегат с антипараллельной упаковкой пептидных сегментов в бета-слой на графитовой подложке. Фибриллярный агрегат создавался путём размножения исходной периодической упаковки и состоял из 80 молекул. Графитовая подложка состояла из 4 слоёв графита и имела размер около 15 на 68 нм. Общее количество атомов в системе составило 173760. Система была отрелаксирована методом молекулярной динамики в течение 100 пс при температуре 300К. Задача периода релаксации состояла в адсорбции фибриллы на графит. Снэпшот фибриллы на поверхности графита после 100 пс релаксации приведён на рис. 5.12.

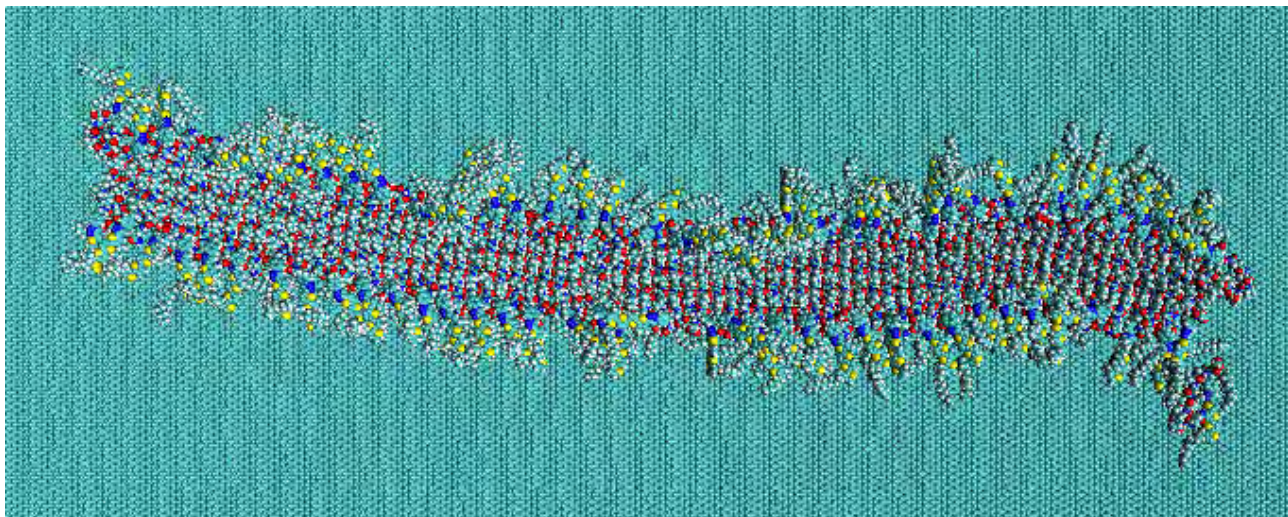


Рис. 5.12. Мгновенное изображение однослойной фибриллы на основе антипараллельного бета-слоя после 100 пс релаксации на графитовой подложке.

Моделирование фибриллы на подложке выявляет достаточно очевидный факт – адсорбция однослойной фибриллы на графите в планарной форме препятствует её закрутке и образованию спиральности. В то же время на рис. 5.12 видно, что агрегат может изгибаться в латеральной плоскости, по-видимому, с помощью персистентного механизма гибкости. Такие изгибы заметны на масштабах десятков нанометров.

Для анализа геометрических характеристик фибриллы, а главным образом высоты фибриллы, был применён разработанный алгоритм виртуальной АСМ. Суть алгоритма заключается в построении поверхности системы по следующему принципу: сфера заданного радиуса (моделирующая кантилевер) последовательно опускается в различные точки поверхности графита, как только сфера касается какого-либо атома процедура в данной точке прекращается, а Z-координата сферы на момент соприкосновения с первым атомом считается высотой системы в данной точке поверхности. Полученные таким образом модели АСМ изображений представлены на рис. 5.13. На рис. 5.14 представлен профиль высоты одного из изображений, по профилю можно чётко определить высоту фибриллы в пептидной части - около 9 Å.

## 5.6. Моделирование двуслойных фибрилл

Два типа отобранных двуслойных упаковок (упаковки типа (II) и (III)) были размножены в 40 раз вдоль основной оси для получения фибриллярных агрегатов длиной в 80 бета-листов (см. рис. 5.15). Длина начальной конформации таких агрегатов составляла приблизительно 40 нм.

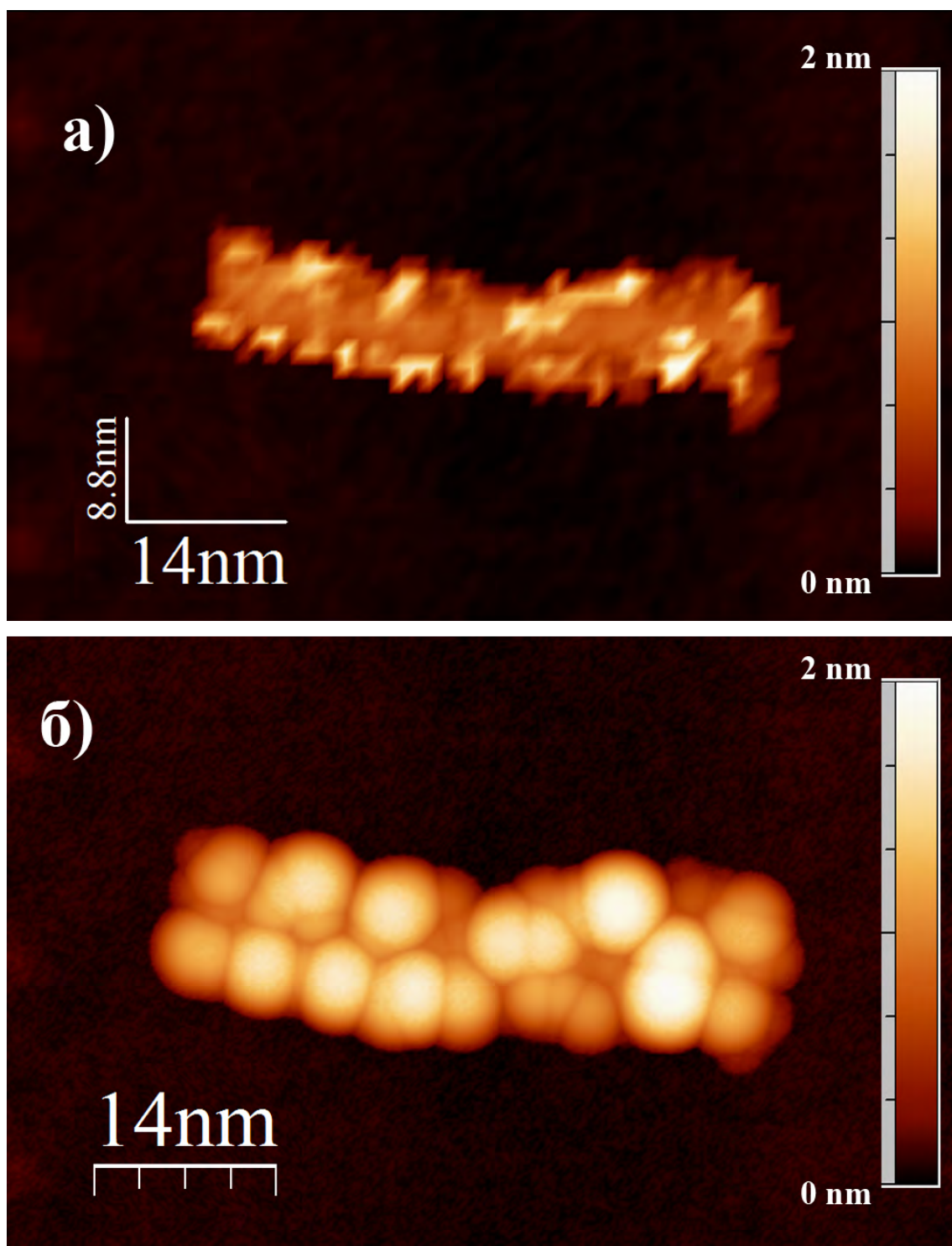


Рис. 5.13. Модель изображения АСМ микроскопа для молекулярной системы – однослойный антипараллельный фибриллярный агрегат на графите. а) радиус модельного зонда 0.5 нм, б) радиус модельного зонда 10 нм.



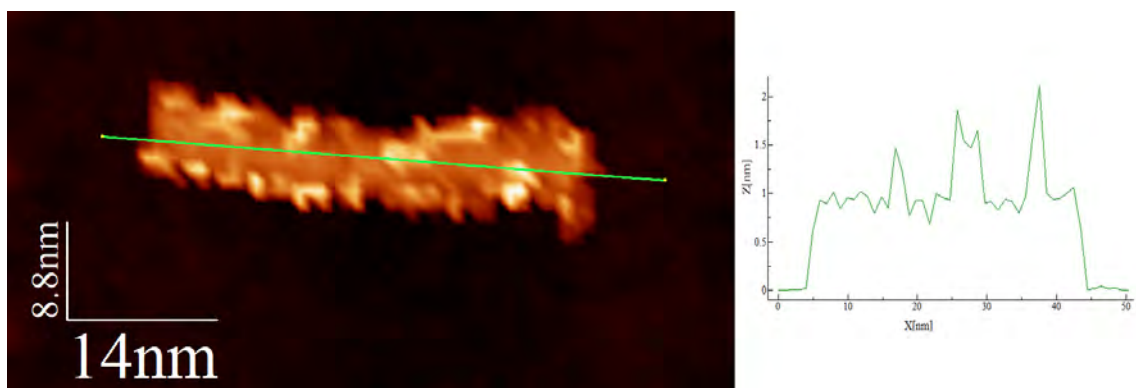


Рис. 5.14. Профиль высоты вдоль модели АСМ изображения.

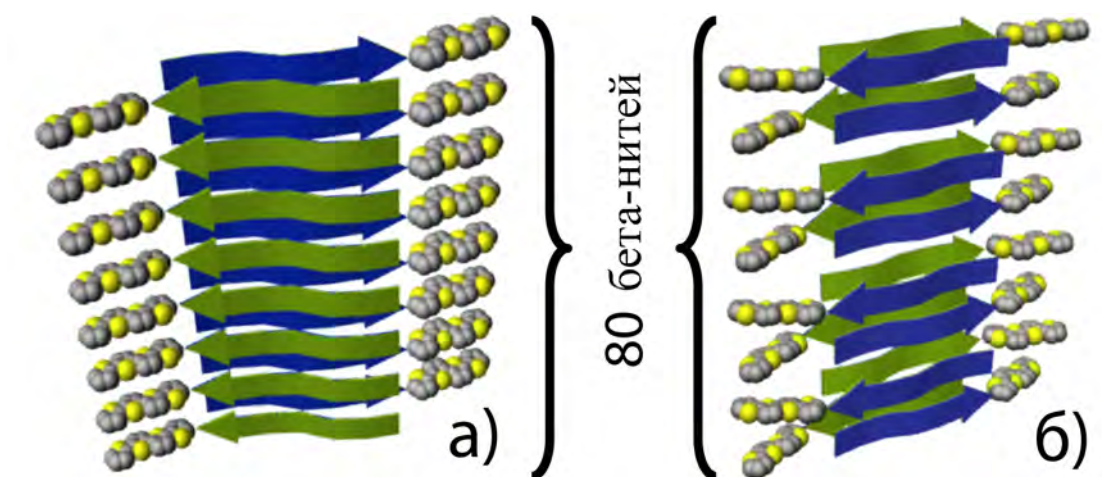


Рис. 5.15. Схематическое изображение начального состояния двух типов двуслойных фибриллярных агрегатов: а) на основе параллельного бета-слоя с гидрофильным ядром, б) на основе антипараллельного бета-слоя с гидрофобным ядром.

Для собранных агрегатов было проведено моделирование методом молекулярной динамики в течение 10 нс, аналогично тому как это делалось для однослойных агрегатов.

#### 5.6.1. Визуальное описание

Как и в случае с однослойными упаковками в течение первой наносекунды моделирования произошла релаксация быстрых степеней свободы фибрилл, в частности, расстояний между бета-листами (которое ранее должно было удовлетворять периодическим граничным условиям), конформации алкильных цепей и т.д. Однако на больших временах наблюдалось принципиальное отличие поведения фибриллы как целого. Двуслойные фибриллы (см. рис. 5.16, 5.17) оказались намного более жёсткими и стабильными в плане сохранения своей фибриллярной формы, линейная форма фибрилл практически не изменилась за исключением того, что плоскости фибрилл приобрели некоторую закрутку относительно оси, что видно на рисунках 5.16 в) , 5.17 в). При этом, если для случая, основанного на параллельной укладке бета-листов, закрутку концевых сегментов можно отнести к проявлению концевых эффектов, в то время, как сам фибриллярный слой сохраняет свою планарность, то в случае антипараллельной укладки (рис. 5.17, в)) фибрилла демонстрирует чёткую левовинтовую закрученность. Такая закрутка является проявлением специфических взаимодействий бета-листов вдоль слоя и наблюдается в том числе в структурах белков. Период закрутки фибриллы (расстояния на котором плоскость фибриллы поворачивается на  $180^\circ$ ) приблизительно равен 80 нм.

#### 5.6.2. Упорядочение пептидных частей

В основе двух изученных фибрилл лежит двойной бета-слой. Обсудим два момента, которые представляют особый интерес: во-первых, каким образом организуется межслоевой контакт, а, во-вторых, как влияет на упорядочение структуры наличие двух бета-слоёв.

На рис. 5.16 и 5.17, г) видно, что бета-листы в двух слоях располагаются с относительным смещением в латеральной плоскости; такая организация позволяет формировать плотный межслоевой контакт, когда боковые цепи верхнего бета-листа, располагаются в пространстве между боковыми цепями двух нижних бета-листов. Такой контакт может иметь гидрофобную природу, когда слои соприкасаются гидрофобными остатками, или же полярную. В последнем случае визуальное наблюдение показало образование водородных связей между ОН-группами боковых цепей треонинов, которые укрепляли межслоевой



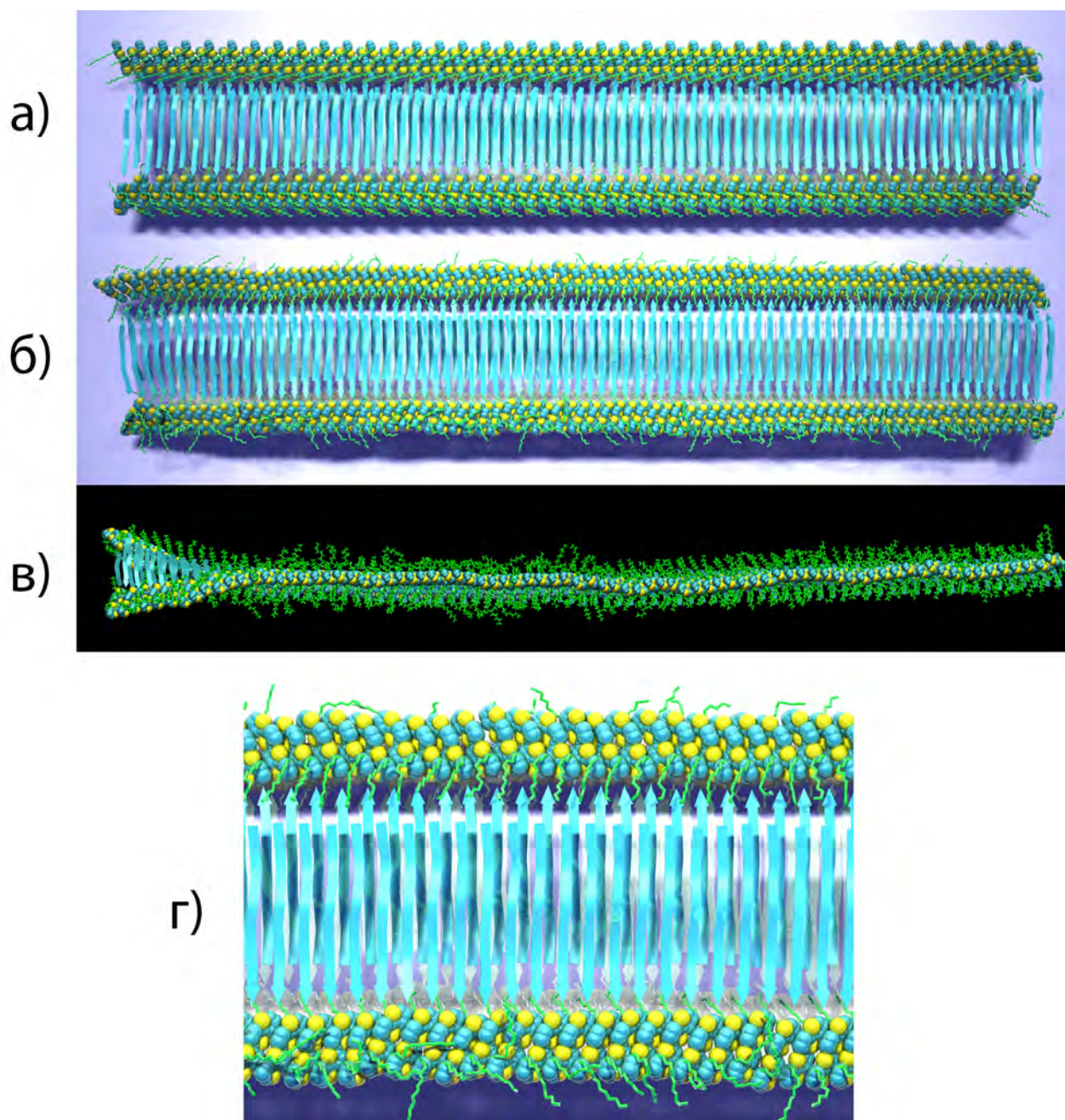


Рис. 5.16. Мгновенные изображения двуслойного волокна, основанного на двух *параллельных* бета-слоях, сложенных гидрофильными сторонами, в ходе расчётов молекулярной динамики: а) начальная конформация, созданная из полученной ранее периодической укладки, б) система через 10 нс моделирования, в) вид фибриллы в профиль через 10 нс, г) увеличенный участок рисунка б). Пояснения к рисунку: пептидные сегменты изображены в виде листов, атомы тиофеновых сегментов в виде Ван-дер-Ваальсовых сфер (сера - жёлтый, углерод - сине-зелёный), алкильные цепи в виде зелёных трубочек.



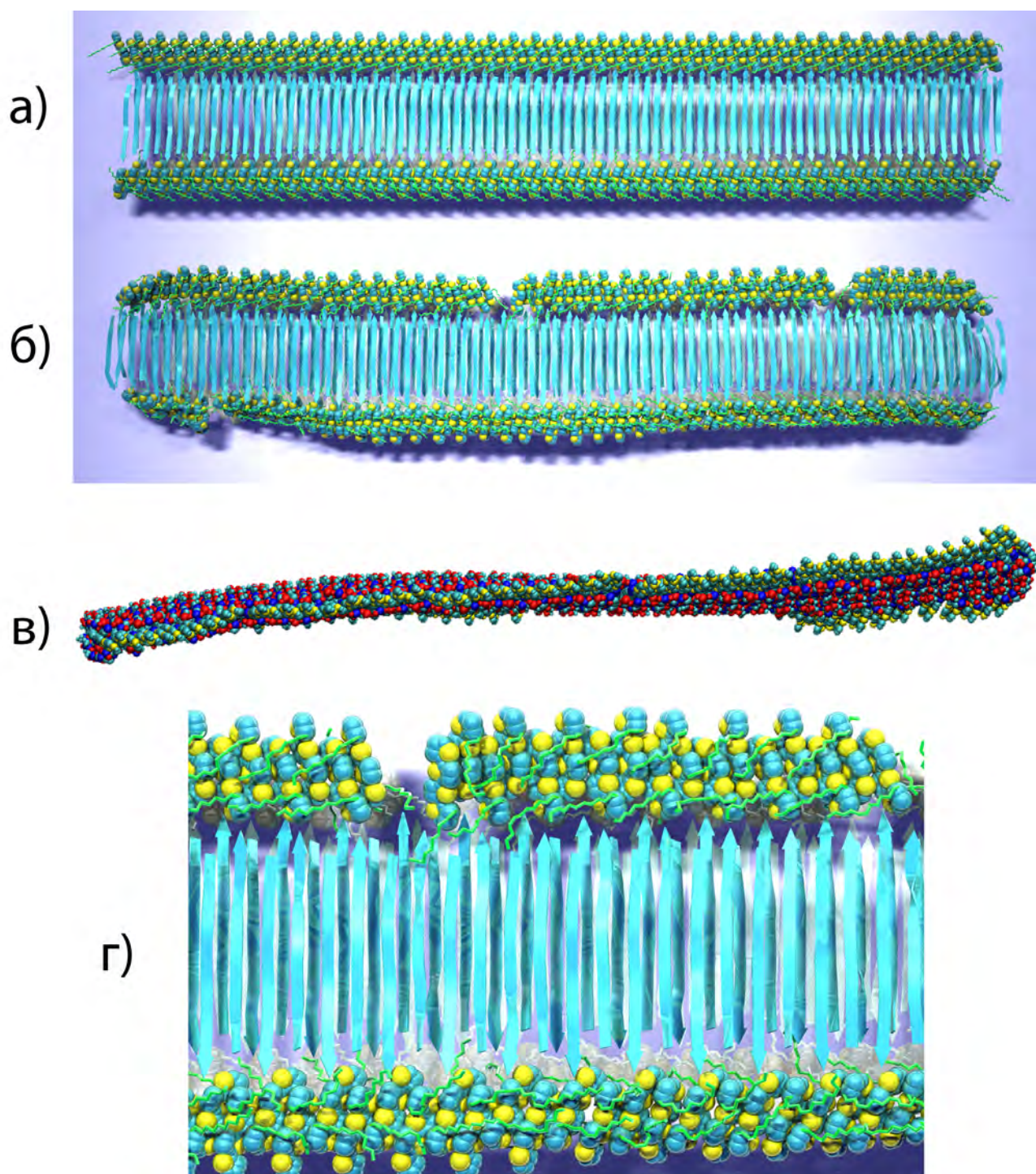


Рис. 5.17. Мгновенные изображения двуслойного волокна, основанного на двух *антипараллельных* бета-слоях, сложенных гидрофобными сторонами, в ходе расчётов молекулярной динамики: а) начальная конформация, созданная из полученной ранее периодической укладки, б) система через 10 нс моделирования, в) вид фибриллы в профиль через 10 нс, г) увеличенный участок рисунка б). Пояснения к рисунку: пептидные сегменты изображены в виде листов, атомы тиофеновых сегментов в виде Ван-дер-Ваальсовых сфер (сера - жёлтый, углерод - сине-зелёный), алкильные цепи в виде зелёных трубочек.

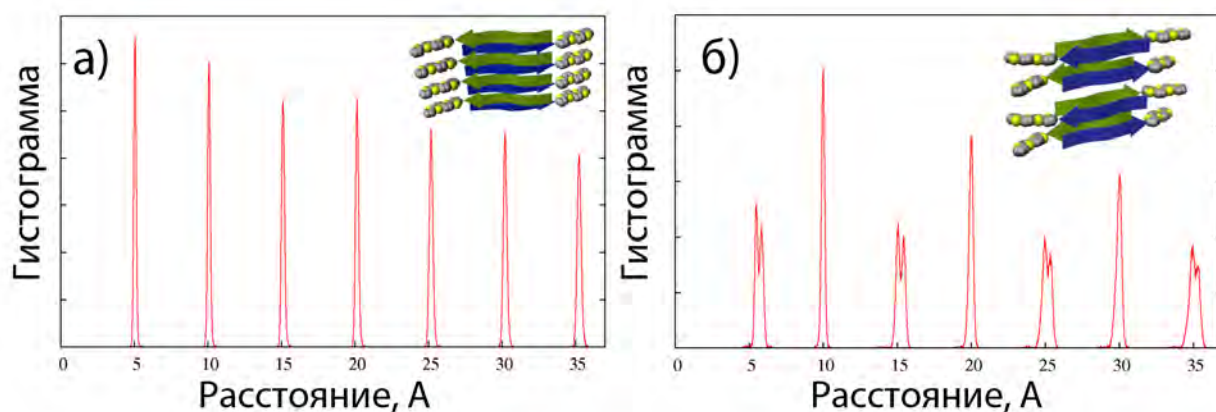


Рис. 5.18. Гистограммы расстояний между центрами масс пептидных последовательностей [Thr-Val]<sub>3</sub> в двуслойных фибриллах, полученные в ходе МД расчётов. При построении гистограмм учитывались лишь расстояния между пептидными сегментами, находящимися в одном бета-слое. а) случай параллельной укладки, б) случай антипараллельной укладки.

контакт.

Одним из основных отличий однослойных фибрилл от двуслойных является более высокая степень упорядочения пептидных сегментов, а также меньшая конформационная подвижность самих пептидных сегментов. На рис. 5.19 представлено распределение аминокислот по элементам вторичной структуры в двуслойных фибриллах. Сравнивая эти данные с аналогичным графиком для однослойных фибрилл (см. рис. 5.19), приходим к выводу, что для фибрилл на основе параллельной укладки наблюдается принципиальная разница: в случае двуслойной фибриллы алгоритм STRIDE детектирует полное преобладание бета-слоёв, чего не наблюдалось в однослойной фибрилле. Ещё более характерные особенности выявляются при рассмотрении гистограммы расстояний между пептидными сегментами (рис. 5.18). Очень заметно увеличивается дальний и средний порядок в упаковке пептидных сегментов для фибрилл на основе антипараллельных бета-слоёв при переходе от однослойных к двуслойным фибриллам. Такой же эффект в меньшей степени проявляется и для фибрилл на основе параллельной упаковки. Явная бимодальность каждого нечётного пика на рис. 5.18, б) отражает тот факт, что точный период антипараллельного бета-слоя в два раза больше, чем у параллельного.

## 5.7. Упорядочение тиюфеновых сегментов

Обсудим теперь вопрос о взаиморасположении тиюфеновых сегментов в различных типах исследованных фибрилл, так как, именно, взаимодействие тиюфенов представляет

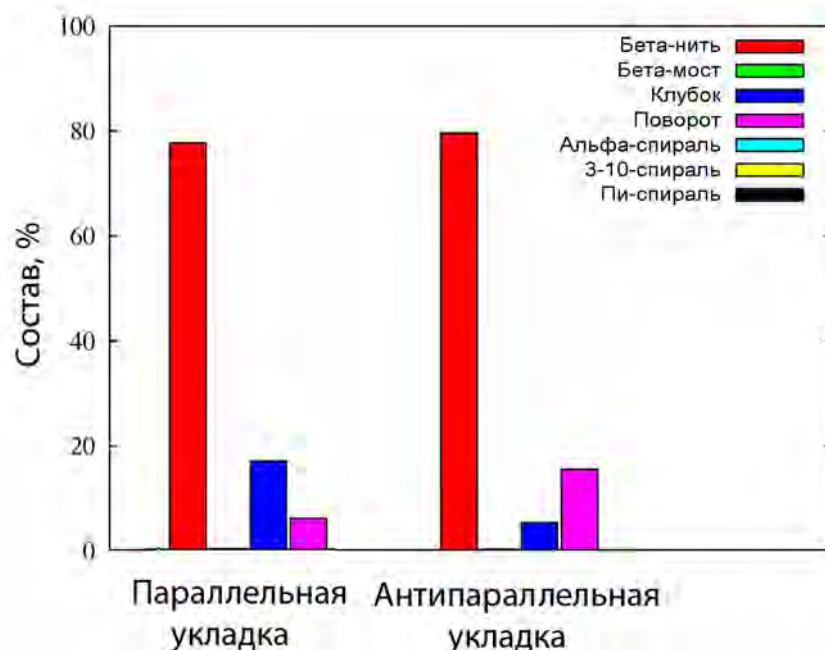


Рис. 5.19. Распределение аминокислот по элементам вторичной структуры в двуслойных фибриллах, данные усреднены по последней наносекунде моделирования.

интерес с технологической точки зрения для нужд органической микроэлектроники и других смежных областей.

### 5.7.1. Однослойные агрегаты

Рассмотрев увеличенные мгновенные снимки фибрилл после 10 нс моделирования (рис. 5.8 и 5.9, в)), видим, принципиальное различие тиофеновых частей двух типов фибрилл. В случае параллельной упаковки, тетратиофены формируют своеобразную непрерывную стопку. В случае антипараллельной укладки, непрерывного контакта тиофенов вдоль фибриллы не существует, области относительного упорядочения сменяются разрывами и зонами малого упорядочения. Ключевой вопрос с точки зрения технологических приложений, а также с фундаментальной точки зрения, – это вопрос о наличии  $\pi - \pi$ -сопряжения между электронными системами тиофенов, которое наделяет их проводящими и полупроводниковыми свойствами.

Известно, что в пептид-тиофеновых гибридах существует некоторая конкуренция между  $\pi - \pi$ -стекингом и созданием бета-слоевой структуры. Конкуренция этих двух типов взаимодействий происходит из-за того, что эти взаимодействия навязывают системе различные периоды организации супрамолекулярных комплексов. Периодичность бета-слоя составляет 4.8-5.0 Å в то время, как идеальное расстояние между плоскостями ароматических систем составляет 3.3-3.5 Å [131, 132]. Кроме того важно также латеральное расположение

$\pi$ -систем, которые обычно располагаются со сдвигом около 1.5 Å. Общее расстояние между центрами масс одинаковых сегментов находящихся в состоянии  $\pi - \pi$ -стэкинга оценивается в 3.8-4.0 Å. В некоторых экспериментальных работах [133] было показано, что при самоорганизации тиюфен-пептидных гибридов, при наличии жёсткой связи между тиюфеновым и пептидным сегментом, “побеждать” могут, в зависимости от условий, как пептидные, так и тиюфеновые взаимодействия. Причём одновременное наличие и бета-листов и  $\pi$ -сопряжённых тиюфеновых систем возможно лишь при наличии мягкого спейсера между сегментами, то есть возможности конформационной подстройки.

Вопрос о размере геометрических отклонений в структуре идеальной тиюфеновой стопки, при которых сопряжение ещё не нарушается сильно (присутствует проводимость, в спектрах регистрируется экситонное расщепление) является достаточно сложным и не до конца изученным. Ответы на такие вопросы могли бы дать дополнительные исследования методами квантовой механики. Пока в рассмотрении этого вопроса нам придётся ограничиться некоторыми грубыми оценками, а также сопоставлением результатов моделирования с экспериментальными результатами. Вероятно предположить, что при слишком больших периодах структуры, навязываемых пептидной частью тиюфеновым сегментам (такой случай имеет место в однослойной антипараллельной структуре - период 10 Å),  $\pi - \pi$ -сопряжение тиюфеновых систем будет практически полностью подавлено. В случае же параллельной укладки вопрос становится более сложным.

Для того, чтобы количественно описать упорядочение тиюфеновых сегментов в фибриллах были построены различные гистограммы и распределения для тиюфеновых сегментов и для отдельных тиюфеновых колец. На рис. 5.20 а) представлены гистограммы расстояний между тиюфеновыми сегментами для двух фибрилл. Видно, что в случае параллельной упаковки периодичность расположения тиюфеновых сегментов точно следует за периодичностью пептидного остова фибриллы. А в случае антипараллельной упаковки наблюдается некоторая хаотичность в упорядочении тиюфеновых сегментов, что говорит о малой вероятности наличия в этом случае  $\pi - \pi$ -сопряжения различных сегментов. Дополнительную информацию даёт гистограмма распределения расстояний между отдельными тиюфеновыми кольцами, не связанными ковалентно, см. рис. 5.20 б). Поскольку гистограммы дают скорее качественное представление об упорядочении системы, в то время как спектральный отклик системы зависит от (относительной) концентрации молекул находящихся в определённом состоянии, были также построены профили доли тиюфеновых колец, которые имеют хотя бы одно другое тиюфеновое кольцо в пределах заданного расстояния рис. 5.22 а). Из графика видно, что доля сопряжённых колец при маленьких расстояниях выше для антипараллельной упаковки, однако следует помнить, что эта ха-



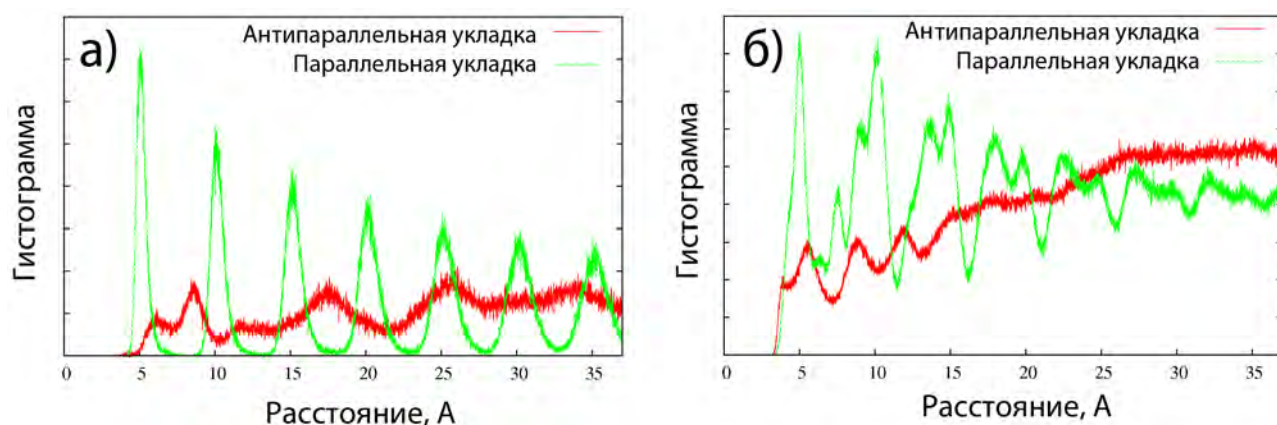


Рис. 5.20. Гистограммы расстояний между тиофеновыми сегментами в однослойных фибриллах в ходе МД расчётов: а) между центрами масс тиофеновых сегментов, б) между центрами масс тиофеновых колец, не связанных ковалентно.

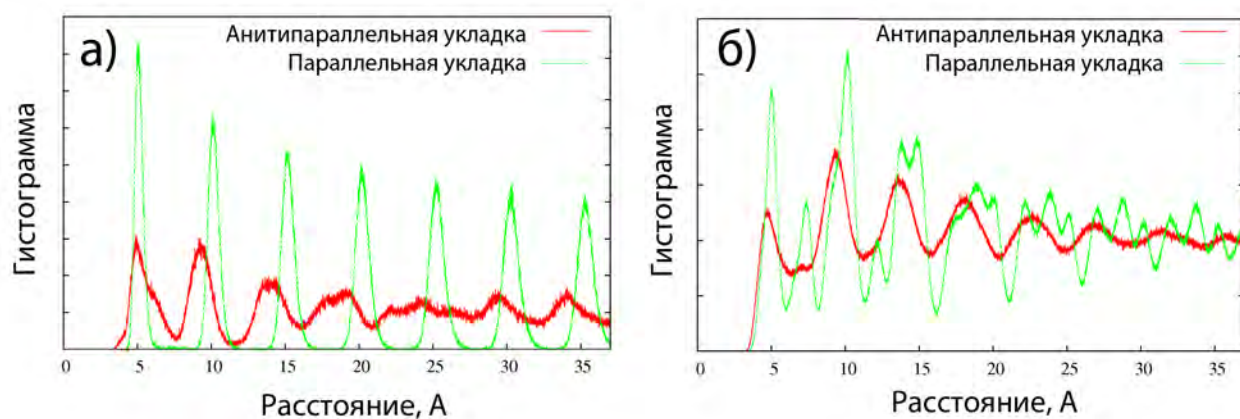


Рис. 5.21. Гистограммы расстояний между тиофеновыми сегментами в двуслойных фибриллах в ходе МД расчётов: а) между центрами масс тиофеновых сегментов, б) между центрами масс тиофеновых колец, не связанных ковалентно.

рактика отражает не столько кооперативные, сколько бинарные эффекты. Поэтому даже наличие большого количества сопряжённых колец, но в хаотическом состоянии не будет приводить к проводимости, в этом плане более важным кажется крутость графика (его ступенькообразная форма), которая больше выражена у параллельной упаковки, именно эта характеристика свидетельствует о наличии порядка в системе.

### 5.7.2. Двуслойные агрегаты

Рассмотрим вопрос об упорядочении тиофеновых сегментов в двуслойных фибриллах. Для фибриллы, основанной на параллельных бета-слоях, ситуация не отличается от такой же однослойной упаковки (см. рис. 5.8) за исключением того, что тиофеновые

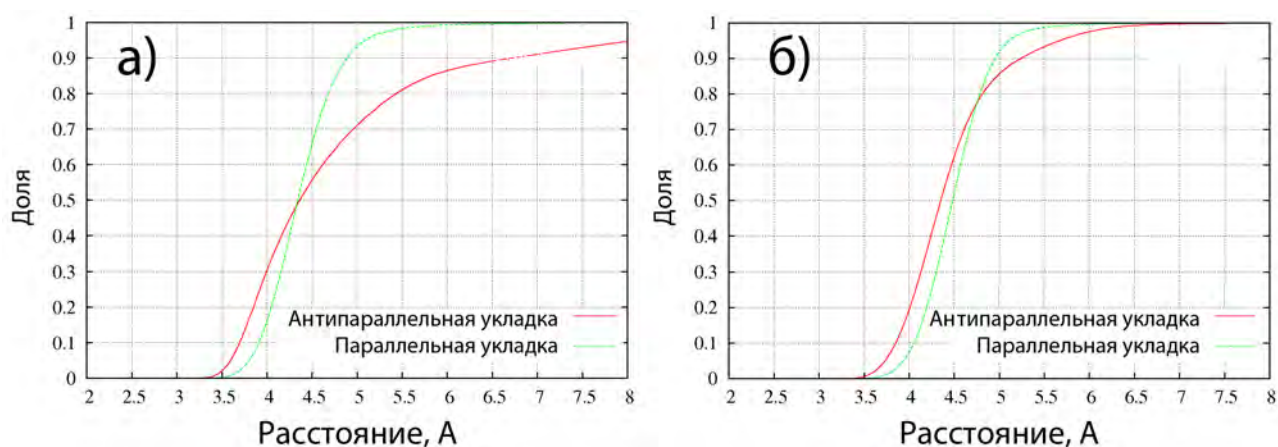


Рис. 5.22. Зависимость доли тиофеновых колец, которые имеют хотя бы одно другое ковалентно не связанное тиофеновое кольцо в пределах заданного расстояния от расстояния для исследованных (а) однослойных фибрилл и (б) двуслойных фибрилл.

стопки (“провода”) располагаются по обоим краям плоскости фибриллы. В случае же двуслойной антипараллельной упаковки наблюдается принципиальное отличие, в двуслойной упаковке: тиофеновые сегменты от двух слоёв формируют взаимопроникающие стопки. Конформация этих стопок представлена на рис. 5.17, г). Следует выделить две основных особенности такого упорядочения тиофенов: во-первых, соседние тиофены не сопоставлены “лицом к лицу”, а расположены со сдвигом вдоль направления тиофеновой цепочки, который задаётся их привязкой к пептидному остову. Во-вторых, в отличие от параллельной укладки, где тиофены формируют непрерывные стопки, в данном случае в тиофеновых стопках встречаются щели. Наблюдение за траекторией МД показало, что эти щели не являются постоянными, а диффундируют вдоль волокна благодаря тому, что отдельные тиофеновые сегменты перепрыгивают с одного края щели на другой.

На рис. 5.21 и 5.22, б) представлены гистограммы для тиофеновых сегментов аналогичные тем, что были построены при исследовании однослойных фибрилл. Заметим, что для фибрилл, основанных на параллельных бета-листах гистограммы сохраняют свой вид и становятся немного острее, что свидетельствует об увеличении упорядочения в системе. Для двуслойной фибриллы, основанной на антипараллельной упаковке, ситуация должна была поменяться серьёзным образом, поскольку плотность тиофеновых сегментов возросла в два раза, что проявляется на гистограммах появлением дополнительных пиков на небольших расстояниях. Тем не менее достаточно быстрое стремление гистограмм к постоянному значению свидетельствует о том, что тиофеновые сегменты в антипараллельной упаковке упорядочены значительно слабее, чем в параллельной.

## 5.8. Сравнение эксперимента и моделирования

Моделирование вышеописанных агрегатов было вызвано необходимостью подробнее понять экспериментальные данные и объяснить получающиеся в результате фибриллоподобные агрегаты. Моделирование представляет в том числе и самостоятельную ценность, поскольку позволяет понять закономерности образования и строения таких агрегатов. Попытаемся на основе полученных данных моделирования ещё раз осмыслить экспериментальные данные и выдвинуть гипотезы о строении наблюдаемых агрегатов.

Рассмотрим ещё раз структуры представленные на АСМ изображениях 5.2. Данные структуры образовывались при нанесении раствора исследуемого вещества концентрацией  $10^{-4}$ – $10^{-5}$ М в растворителе, состоящем на 50% из дихлорметана и на 50% из метанола на слюду (мусковит). Нанесение осуществлялось методом *spin-coating* при скорости вращения 3000 оборотов в минуту. По результатам экспериментов наиболее разумная интерпретация АСМ изображений следующая: в ходе процедуры *spin-coating* из диффузионного слоя на подложку высаживается некоторый слой вещества, который полностью покрывает плоскость подложки. Толщина этого слоя оценивается величиной до 100 нм. АСМ исследования показывают, что поверхность этого слоя вся покрыта мелкими “фибрилярными” структурами (протофибриллами). На поверхности встречаются отдельные фибриллярные образования бóльшего масштаба, в которых можно рассмотреть более тонкую структуру, похоже, что они состоят из состыкованных протофибрилл.

Любой процесс самосборки является зависимым от концентрации исходного вещества, ведь при концентрации стремящейся к нулю, термодинамически любая агрегация не является выгодной, потому что побеждает энтропийный вклад в свободную энергию. Следовательно в растворе заданной концентрации всегда будет баланс между мономерами и агрегатами различной сложности. В случае амилоидо-подобных фибрилл часто оказывается, что баланс очень сильно сдвинут в сторону образования агрегатов. В этом случае экспериментально удаётся наблюдать практически одиночные фибриллы на чистой подложке. Таких примеров можно найти много в литературе (в частности и для фибрилл из паутины или шёлка), любопытным является пример образования таких “крепких” фибрилл для соединения аналогичного нашему, являющемуся три-блок-олигомером с пептидными последовательностями, в которых оставлен сложноэфирный дефект [48] (см. рис. 1.13). В нашем случае, по-видимому, термодинамический баланс не на столько сильно смещён в сторону образования крупных агрегатов: в растворе присутствуют и маленькие прото-фибриллы, так и большие агрегаты, и, вероятно, отдельные молекулы. Все эти компоненты раствора полностью покрывают подложку.



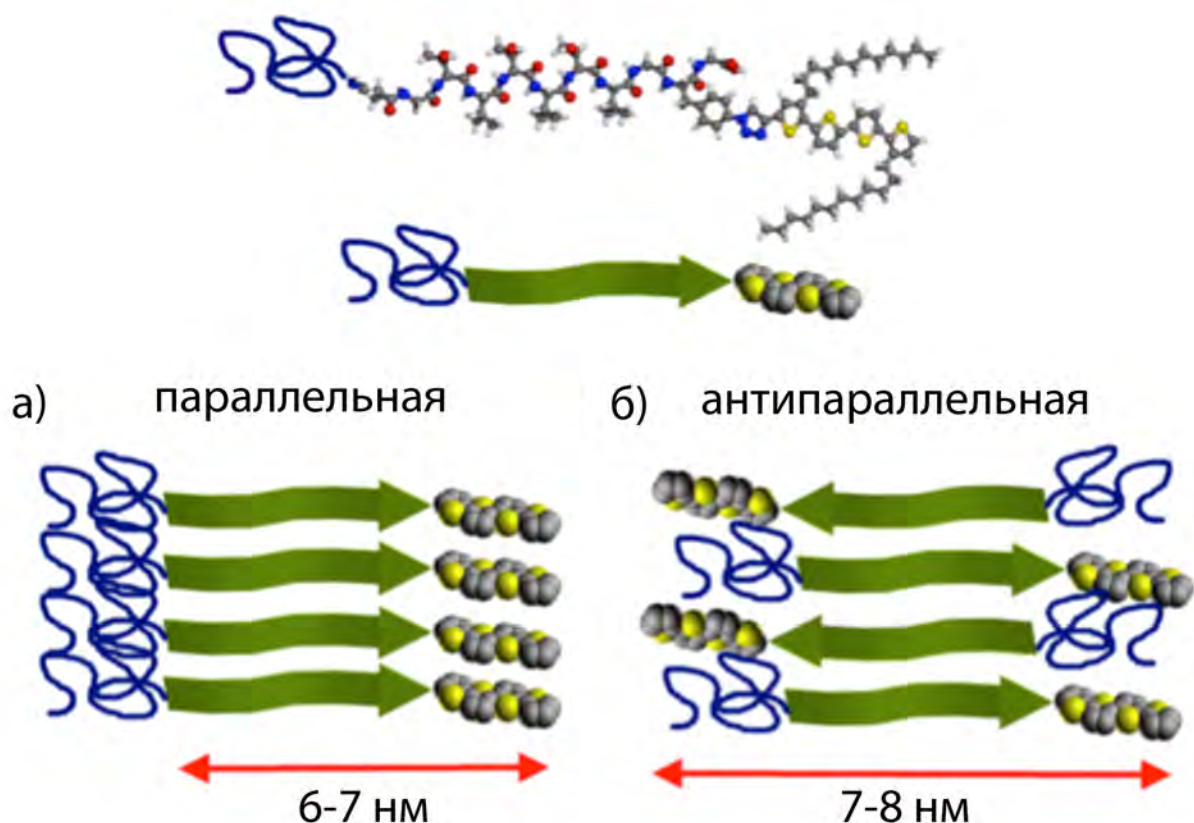


Рис. 5.23. Возможные модели укладок молекул в протофибриллах и их размеры.

Протофибриллы, которые видны на заднем плане АСМ изображений, имеют приблизительно следующие размеры: длина 100-200 нм; ширина  $6-8 \pm 2$  нм по данным АСМ (без тип-коррекции),  $6 \pm 2$  нм по данным ПЭМ. Высота протофибрилл над остальной массой вещества  $0.4-0.7 \pm 0.2$  нм. Ширина возможных укладок, рассмотренных в моделировании (см. рис. 5.23) находится в соответствии с экспериментальными данными. Высота однослойной укладки по результатам моделирования на графитовой подложке составляет 9 Å, отсюда можно сделать вывод, что наблюдаемые в эксперименте фибриллы частично погружены в слой вещества покрывающего подложку.

По спектрам инфракрасной спектроскопии не удалось однозначно определить является ли укладка бета-нитей параллельной или антипараллельной. По данным флюоресцентной спектроскопии и спектров кругового дихроизма параллельная укладка является более предпочтительной, хотя окончательного вывода сделать нельзя.

По данным флюоресцентной и КД-спектроскопии для данной системы не наблюдалось “настоящего” экситонного расщепления для уровней энергии сопряжённых систем. Данный факт в принципе можно объяснить как для параллельной, так и для антипараллельной

укладки. В первом случае взаимодействию  $\pi$ -систем препятствует навязываемая пептидным остовом периодичность тиюфенов в 5 Å. Во втором случае (см. рис. 5.23, б)) взаимодействию тиюфеновых сегментов могут мешать ПЭГ цепи.

Если предполагать параллельную организацию бета-слоёв, то в этом случае при агрегации будет возникать небольшой дипольный момент, направленный перпендикулярно оси фибриллы. Желание скомпенсировать эти дипольные моменты является одним из факторов, который может приводить к формированию фибрилл из протофибрилл путём латеральной агрегации. Другой движущей силой формирования многофиламентных фибрилл являются гидрофобно-полярные взаимодействия. В латеральной плоскости фибриллы типа рис. 5.23, а) могут взаимодействовать с одной стороны гидрофобными тиюфеновыми частями, с другой стороны полярными ПЭГ цепями. В свою очередь взаимодействия двух и более фибрилл своими поверхностями может приводить к образованию многослойных фибрилл.

В заключении, нельзя однозначно утверждать какая упаковка действительно наблюдается в эксперименте, однако более вероятным выглядит предположение о том, что наблюдаемые фибриллы основаны на параллельной укладке бета-листов.

В пользу этого предположения можно выдвинуть несколько фактов: (1) по данным моделирования фибриллы с параллельной упаковкой меньше, чем фибриллы с антипараллельной упаковкой, подвержены закрутке, отсутствие которой наблюдается в эксперименте (данные КД и АСМ), (2) данные флюоресцентной и КД-спектроскопии, хотя и неоднозначно, но указывают на бóльшую вероятность параллельной упаковки, (3) расстояние между тиюфеновыми сегментами, рассчитанное в моделировании, не противоречит отсутствию экспериментально регистрируемого экситонного расщепления уровней в КД-спектроскопии.

Однослойные протофибриллы (рис. 5.23, а)) в свою очередь могут агрегировать, образуя многослойные агрегаты, которые в свою очередь агрегируют латерально, образуя более крупные мульти-филаментные агрегаты, наблюдаемые в эксперименте.

## 5.9. Выводы к главе

Гибридные соединения, состоящие из блоков синтетических и биополимеров (полимерные биоконъюгаты), являются перспективными соединениями как для технологических, так и для биомедицинских приложений. В таких соединениях реализуется синергетический эффект от объединения свойств, присущих обоим компонентам. Фрагменты

биологических молекул такие, как пептиды или олигонуклеотиды, приносят в данные соединения такие свойства, как самосборка, биосовместимость, специфические взаимодействия (pattern recognition), зависимость структуры от внешних факторов и т.д. Одним из ярких примеров такого подхода к дизайну новых материалов являются гибриды пептидов и сопряжённых проводящих полимеров. Свойства таких гибридов открывают широкий спектр их возможных применений: построение и самосборка устройств органической микро- и нанoeлектроники, создание молекулярных сенсоров, биоэлектронных интерфейсов, самособирающихся нанопроводов и т.д. Особенный интерес представляют тиофен-пептидные гибриды, поскольку политиофены один из самых перспективных в использовании (полу)проводящих полимеров.

На данный момент в небольшом количестве экспериментальных работ [37, 47, 48] получены первые прототипы пептид-тиофеновых гибридов и показана возможность самосборки таких молекул в нанофибриллы. Однако пока ещё остаются не совсем ясными не только принципы самоорганизации таких фибрилл, но и их структура, укладка молекул в агрегатах. И если для самоорганизации простых пептид-полимерных структур в литературе достигнуто некоторое понимание на уровне возможных моделей самоорганизации, то для специфических тиофен-пептидных систем уровень теоретической проработанности весьма низок. Существенным фактором, который отличает пептид-тиофеновые системы от общего класса пептид-полимерных систем, является способность тиофенов к специфическому взаимодействию путём  $\pi - \pi$ -стэкинга, которой может в некоторых случаях быть даже сильнее, чем силы самоорганизации пептидных частей.

В данной работе на примере конкретных экспериментальных данных по самоорганизации пептид-тиофенового диблок олигомера была предпринята попытка создания полно-атомных моделей возможных агрегатов и прямого атомистического моделирования устойчивости и свойств различных типов упаковок молекул в фибриллярные агрегаты. С опорой на экспериментальные данные, а также на литературные источники, были созданы различные укладки молекул на основе параллельных и антипараллельных бета-листов. При построении различных моделей была использована оригинальная концепция иерархического построения модели, учитывая вероятную иерархию взаимодействий в таких системах (водородные связи, гидрофобные взаимодействия и т.д.). Последовательно создавались сначала простые элементы структуры (бета-слои), затем из ранее созданных конструировались более крупные элементы, при этом на каждом шаге проводилась релаксация систем методом МД, что обеспечивало поиск минимума свободной энергии не для всей системы сразу, а последовательно. Для четырёх типов упаковок были созданы модели фибрилл длиной до 40 нм, и проведено их исследование методами молекулярной динами-

ки. Изучалась стабильность агрегатов, организация пептидных и тирофеновых сегментов. Сравнительный анализ однослойных и двуслойных фибрилл позволил сделать выводы о роли тех или иных взаимодействий в устойчивости и форме агрегатов.

Кроме данной работы в литературе представлена лишь одна работа, где проводилось моделирование тирофен-пептидных гибридов методом молекулярной динамики [54]. Новизна представленной работы состоит в следующем (в порядке убывания масштаба новизны):

- Впервые проведено полноатомное молекулярное моделирование фибриллярных агрегатов, состоящих из тирофен-пептидных диблок олигомеров с последовательностью [Thr-Val]<sub>3</sub>.
- Впервые для фибрилл из полимер-пептидных гибридов построены молекулярные модели, основанные на однослойных и двуслойных бета-листах. Эти модели проанализированы методом МД.
- Предложен метод создания моделей таких агрегатов (метод последовательного иерархического конструирования).
- Предложен метод оценки высоты молекулярных структур на подложке (метод виртуальной АСМ). Этим методом построены модели АСМ изображений фибрилл и профили высоты.
- Впервые для систем такого класса проводились столь масштабные вычисления, использовались методы параллельной МД с одновременным использованием до 128 процессоров.

По результатам работы можно сделать следующие выводы, обладающие научной новизной:

- Однослойные агрегаты тирофен-пептидных олигомеров в вакууме оказываются стабильными, но не демонстрируют жёсткости структуры. В то же время двуслойные агрегаты являются намного более стабильными и способны к формированию длинных линейных структур.
- У агрегатов, основанных на антипараллельной организации бета-слоёв, плоскость фибриллы закручивается в левую сторону относительно оси фибриллы, при этом однослойные агрегаты стремятся закрутиться и в суперспираль. Для агрегатов, основанных на параллельных бета-слоях, последней тенденции не наблюдается, а закрученность вдоль оси проявляется меньше.

- По крайней мере в изученной модели взаимодействия пептидных частей доминируют над взаимодействиями тиофеновых сегментов, в результате пептидная часть навязывает тиофенам межсегментное расстояние в 5 Å.
- На основании наблюдаемых особенностей поведения различных укладок можно сделать предположение о том, что наблюдаемые в эксперименте агрегаты в своей основе имеют укладку по типу параллельного бета-слоя.

## Заключение

В диссертационной работе проведены следующие исследования и получены следующие результаты:

1. Методами молекулярной динамики изучена гидратация боковых цепей аминокислот вблизи поверхности воды. Построены профили свободной энергии молекул вблизи поверхности воды. Показано, что профили свободной энергии для всех молекул имеют схожую форму. Показано, что профили свободной энергии достигают своего постоянного значения при погружении молекул на глубину 0.6-0.7 нм (влияние поверхностных эффектов становится незначительным). Разработан метод оценки свободной энергии адсорбции молекул на границу раздела жидких фаз на основе классических полноатомных моделей молекул. Проведена оценка энергий гидратации и адсорбции на границу вода/воздух для боковых цепей аминокислот. Выявлено, что энергии гидратации боковых цепей аминокислот находятся в хорошей корреляции с абсолютными энергиями адсорбции на поверхность воды. В тоже время энергия адсорбции относительно фазы максимального сродства является независимым параметром, описывающим свойства веществ.
2. Исследовано распределение аминокислотных остатков в глобулярных белках в зависимости от их погружённости в структуру белка путём статистического анализа большого количества экспериментальных данных по пространственному строению эволюционно-независимых белков. Получены статистические профили свободной энергии различных типов аминокислотных остатков в зависимости от экспонированности. С помощью различных методов классификации аминокислот на “ядерные” и “поверхностные” рассчитаны статистические энергии переноса аминокислот из ядра на поверхность, рассчитаны корреляции этих статистических энергий с экспериментальными коэффициентами распределения аминокислот между рядом растворителей. Показано, что коэффициенты корреляции сильно зависят как от выбора пар растворителей, так и от выбора критериев статистической классификации аминокислот. Предложены три различных статистических шкалы, каждая из которых наилучшим образом коррелирует с одной из экспериментальных шкал гидрофобности (вода/пар, вода/октанол, вода/циклогексан). Проанализированы причины такого поведения.
3. Предложены различные возможные модели укладки гибридных молекул тетрапептида ([Thr-Val]<sub>3</sub>) с образованием нанофибрилл. Созданы молекулярные

модели таких укладок путём поэтапного рационального конструирования с учётом имеющихся экспериментальных данных. Созданные нанофибриллы проанализированы методом молекулярной динамики. Показано, что изучаемые молекулы могут формировать стабильные, упорядоченные фибрилло-подобные молекулярные слои благодаря организации пептидных частей молекулы в параллельный или антипараллельный бета-слой. Показано, что однослойные структуры в объёме обладают тенденцией к закручиванию, которая сильнее выражена у структур на основе антипараллельной укладки бета-слоёв, которые склонны образовывать суперспирали. Двуслойные структуры, в отличие от однослойных, в объёме сохраняют свою линейность и, в зависимости от типа укладки, могут обладать закрученностью. Предложены различные варианты однослойных и двуслойных агрегатов, в которых тиофеновые сегменты двух слоёв (лент) образуют структуры подобные нанопроводам. Показано, что наиболее вероятной упаковкой молекул в фибриллах, соответствующей экспериментальным данным, является упаковка, основанная на параллельном бета-слое.

---

## Благодарности

---

В заключение хотелось бы выразить благодарность своему научному руководителю академику А.Р. Хохлову, за постановку интересных задач, обсуждение полученных результатов и поддержку в решении научных проблем. Автор выражает особую благодарность профессору П. Г. Халатуру и доценту В. А. Иванову за помощь в работе над отдельными главами диссертации, а также всем сотрудникам, студентам и аспирантам кафедры физики полимеров и кристаллов за ценные замечания и создание приятной научной атмосферы при работе над диссертацией.

Особую признательность автор выражает родителям и друзьям.



## Литература

1. Shaytan A. K., Ivanov V. A., Shaitan K. V., Khokhlov A. R. Free energy profiles of amino acid side chain analogs near water-vapor interface obtained via MD simulations // J. Comp. Chem. 2010. Vol. 31, no. 1. Pp. 204–216.
2. Shaytan A. K., Shaitan K. V., Khokhlov A. R. Solvent Accessible Surface Area of Amino Acid Residues in Globular Proteins: Correlation of Apparent Transfer Free Energies with Experimental Hydrophobicity Scales // Biomacromolecules. 2009. Vol. 10, no. 5. Pp. 1224–1237.
3. Shaytan A. K., Khokhlov A. R., Khalatur P. G. Large-scale atomistic simulation of a nanosized fibril formed by thiophene-peptide “molecular chimeras” // опубликовано электронно // Soft Matter. 2010. URL: <http://dx.doi.org/10.1039/b918562c>.
4. Шайтан К. В., Турлей Е. В., Шайтан А. К. и др. Динамический молекулярный дизайн био- и наноструктур // Российский Химический Журнал. 2006. Т. 1, № 2. С. 53–65.
5. Шайтан К. В., Турлей Е. В., Шайтан А. К. и др. Неравновесная молекулярная динамика наноструктур, включая биологические // Химическая Физика. 2006. Т. 25, № 9. С. 31–48.
6. Шайтан К. В., Турлей Е. В., Шайтан А. К. и др. Молекулярная динамика и дизайн био- и наноструктур // Вестник биотехнологии. 2005. Т. 1. С. 66–78.
7. Shaytan A. K., Ivanov V. A., Khokhlov A. R. Peptide dynamics at water-membrane interface // The Fifth International Conference on Bioinformatics of Genome Regulation and Structure / Ed. by N. Kolchanov, R. Hofstadt. Vol. 1. Novosibirsk, Russia: Institute of Cytology and Genetics, 2006. Pp. 320–323.
8. Шайтан А. К., Халатур П. Г., Хохлов А. Р. Суперкомпьютерное конструирование биоорганических нанопроводов // Суперкомпьютерные технологии в науке, образовании и промышленности, Под ред. В. А. Садовничий, Г. И. Савин, В. В. Воеводин. Москва: Издательство Московского университета, 2009. С. 51–56.
9. Шайтан А. К. Динамика аминокислотных остатков на интерфейсе вода-мембрана // Сборник тезисов, VIII Конференция студентов и аспирантов Учебно-научный центр по химии и физике полимеров и тонких органических плёнок, Моск. Обл., Россия. 2004. С. 84.

10. Шайтан А. К. Динамика аминокислотных остатков на интерфейсе вода-мембрана // Сборник тезисов, Системная биология и Биоинженерия: международная школа-конференция молодых учёных, Моск. Обл., Россия. 2005. С. 72.
11. Шайтан А. К. Динамика аминокислотных остатков на интерфейсе вода-гексан // Сборник тезисов, Малый полимерный конгресс, Москва, Россия. 2005. С. 136.
12. Шайтан А. К., Иванов В. А., Хохлов А. Р. Определение свободной энергии гидратации боковых цепей аминокислот с использованием явной модели воды // Сборник тезисов, Четвёртая Всероссийская Каргинская Конференция "Наука о полимерах 21-му веку Москва, Россия. Т. 3. 2007. С. 364.
13. Shaytan A. K., Khokhlov A. R., Ivanov V. A. MD Simulations of Liquid-Liquid Interface of Aminoacid Solutions // Сборник тезисов, Европейский Полимерный Конгресс 2005, Москва, Россия. 2005. Р. 212.
14. Shaytan A. K., Khokhlov A. R. Conformational Statistics of Amino Acids in Proteins based on their Solvent Accessible Surface Area // Book of Abstracts, 3rd STIPOMAT Conference, Les Diablerets, Switzerland. 2007.
15. Shaytan A. K., Khokhlov A. R. Conformational Statistics of Globular Proteins with Respect to Hydrophobicity/Interfacial Activity of Amino Acid Residues // Book of Abstracts, The second Saint-Petersburg International Conference on NanoBioTechnologies "NanoBio'08", Санкт-Петербург, Россия. 2008. Р. 117.
16. Shaytan A. K. Selfassembly Simulations of Bioinspired Hybrid Systems Based on Oligothiophenes and Peptides // Сборник тезисов, Первая международная летняя школа - Нано2009. Наноматериалы и нанотехнологии в живых системах, Моск. Обл., Россия. 2007.
17. Bogush V., Sokolova O., Davydova L. et al. A Novel Model System for Design of Biomaterials Based on Recombinant Analogs of Spider Silk Proteins // J. Neuroimm. Pharm. 2009. Vol. 4, no. 1. Pp. 17–27.
18. Nelson D. L., Cox M. M. Lehninger Principles of Biochemistry. Oxford, UK: Freeman, W.H. and Company, 2008.
19. Финекльштейн А. В., Птицын О. Б. Физика белка. Книжный Дом Университет, 2005.
20. URL: <http://www.wwpdb.org>.

21. Finkelstein A. V., Badretdinov A. Y., Gutin A. M. Why do protein architectures have boltzmann-like statistics? // *Proteins: Struct., Funct., Genet.* 1995. Vol. 23, no. 2. Pp. 142–150.
22. Derrida B. Random-energy model: An exactly solvable model of disordered systems // *Phys. Rev. B.* 1981. Vol. 24, no. 5. Pp. 2613–2626.
23. Finkelstein A. V., Gutin A. M., Azat. Perfect temperature for protein structure prediction and folding // *Proteins: Struct., Funct., Genet.* 1995. Vol. 23, no. 2. Pp. 151–162.
24. Thomas P. D., Dill K. A. Statistical potentials extracted from protein structures: how accurate are they? // *J. Mol. Biol.* 1996. Vol. 257, no. 2. Pp. 457–469.
25. Fändrich M. On the structural definition of amyloid fibrils and other polypeptide aggregates // *Cell. Mol. Life Sci.* 2007. Vol. 64, no. 16. Pp. 2066–2078.
26. Croixmarie V., Briki F., David G. et al. A cylinder-shaped double ribbon structure formed by an amyloid hairpin peptide derived from the  $\beta$ -sheet of murine PrP: An X-ray and molecular dynamics simulation study // *J. Struct. Biol.* 2005. Vol. 150, no. 3. Pp. 284–299.
27. Nelson R., Sawaya M. R., Balbirnie M. et al. Structure of the cross- $\beta$  spine of amyloid-like fibrils // *Nature.* 2005. Vol. 435, no. 7043. Pp. 773–778.
28. Aggeli A., Nyrkova I. A., Bell M. et al. Hierarchical self-assembly of chiral rod-like molecules as a model for peptide  $\beta$ -sheet tapes, ribbons, fibrils, and fibers // *Proc. Nat. Acad. Sci. USA.* 2001. Vol. 98, no. 21. Pp. 11857–11862.
29. Gauthier M. A., Klok H.-A. Peptide/protein-polymer conjugates: synthetic strategies and design concepts // *Chem. Commun.* 2008. no. 23. Pp. 2591–2611.
30. Börner H. G. Strategies exploiting functions and self-assembly properties of bioconjugates for polymer and materials sciences // *Progr. Polym. Sci.* 2009. Vol. 34, no. 9. Pp. 811–851.
31. Börner H. G. Functional Polymer-Bioconjugates as Molecular LEGO Bricks // *Macromol. Chem. Phys.* 2007. Vol. 208, no. 2. Pp. 124–130.
32. Hentschel J., Börner H. G. Peptide-Directed Microstructure Formation of Polymers in Organic Media // *J. Am. Chem. Soc.* 2006. Vol. 128, no. 43. Pp. 14142–14149.
33. Börner H. G., Schlaad H. Bioinspired functional block copolymers // *Soft Matter.* 2007. Vol. 3, no. 4. Pp. 394–408.

34. Vandermeulen G. W. M., Klok H.-A. Peptide/Protein Hybrid Materials: Enhanced Control of Structure and Improved Performance through Conjugation of Biological and Synthetic Polymers // *Macromol. Biosci.* 2004. Vol. 4, no. 4. Pp. 383–398.
35. Lutz J., Borner H. Modern trends in polymer bioconjugates design // *Progr. Polym. Sci.* 2008. Vol. 33, no. 1. Pp. 1–39.
36. Borner H. G., Smarsly B. M., Hentschel J. et al. Organization of Self-Assembled Peptide-Polymer Nanofibers in Solution // *Macromolecules.* 2008. Vol. 41, no. 4. Pp. 1430–1437.
37. Diegelmann S. R., Gorham J. M., Tovar J. D. One-Dimensional Optoelectronic Nanostructures Derived from the Aqueous Self-Assembly of  $\pi$ -Conjugated Oligopeptides // *J Am Chem Soc.* 2008. Vol. 130, no. 42. Pp. 13840–13841.
38. Hamley I. W., Ansari I. A., Castelletto V. et al. Solution Self-Assembly of Hybrid Block Copolymers Containing Poly(ethylene glycol) and Amphiphilic  $\beta$ -Strand Peptide Sequences // *Biomacromolecules.* 2005. Vol. 6, no. 3. Pp. 1310–1315.
39. Schopf G., Kobmehl G. Polythiophenes - Electrically Conductive Polymers // 1997. Springer, 1997. Vol. 129.
40. Garten F., Vrijmoeth J., Schlattmann A. et al. Light-emitting diodes based on polythiophene: influence of the metal workfunction on rectification properties // *Synthetic Metals.* 1996. Vol. 76, no. 1-3. Pp. 85–89.
41. Choi M., Kim Y., Ha C. Polymers for flexible displays: From material selection to device applications // *Progr. Polym. Sci.* 2008. Vol. 33, no. 6. Pp. 581–630.
42. Cai W., Gong X., Cao Y. Polymer solar cells: Recent development and possible routes for improvement in the performance // *Solar Energy Mat. Solar Cells.* 2010. Vol. 94, no. 2. Pp. 114–127.
43. Lange U., Roznyatovskaya N., Mirsky V. Conducting polymers in chemical sensors and arrays // *Analyt. Chim. Acta.* 2008. Vol. 614, no. 1. Pp. 1–26.
44. Mcquade D. T., Pullen A. E., Swager T. M. Conjugated Polymer-Based Chemical Sensors // *Chem Rev.* 2000. Vol. 100, no. 7. Pp. 2537–2574.

45. Widge A. S., Jeffries-El M., Cui X. et al. Self-assembled monolayers of polythiophene conductive polymers improve biocompatibility and electrical impedance of neural electrodes // *Biosensors and Bioelectronics*. 2007. Vol. 22, no. 8. Pp. 1723–1732.
46. Sirringhaus H., Brown P. J., Friend R. H. et al. Two-dimensional charge transport in self-organized, high-mobility conjugated polymers // *Nature*. 1999. Vol. 401, no. 6754. Pp. 685–688.
47. Klok H.-A., Rosler A., Gotz G. et al. Synthesis of a silk-inspired peptide-oligothiophene conjugate // *Org. Biomol. Chem.* 2004. Vol. 2, no. 24. Pp. 3541–3544.
48. Schillinger E.-K., Mena-Osteritz E., Hentschel J. et al. Oligothiophene Versus  $\beta$ -Sheet Peptide: Synthesis and Self-Assembly of an Organic Semiconductor-Peptide Hybrid // *Adv Mat.* 2009. Vol. 21, no. 16. Pp. 1562–1567.
49. Mouffouk F., Brown S. J., Demetriou A. M. et al. Electrosynthesis and characterization of biotin-functionalized poly(terthiophene) copolymers, and their response to avidin // *J. Mater. Chem.* 2005. Vol. 15, no. 11. Pp. 1186–1196.
50. Jatsch A., Kopyshv A., Mena-Osteritz E., Bauerle P. Self-Organizing Oligothiophene-Nucleoside Conjugates: Versatile Synthesis via "Click"-Chemistry // *Org Lett.* 2008. Vol. 10, no. 5. Pp. 961–964.
51. Bäuerle P., Emge A. Specific Recognition of Nucleobase-Functionalized Polythiophenes // *Adv Mat.* 1998. Vol. 10, no. 4. Pp. 324–330.
52. Mucci A., Parenti F., Schenetti L. A Self-Assembling Polythiophene Functionalised with a Cysteine Moiety // *Macromol. Rapid Comm.* 2003. Vol. 24, no. 9. Pp. 547–550.
53. Ikkala O., Brinke G. Hierarchical self-assembly in polymeric complexes: Towards functional materials // *Chem. Commun.* 2004. no. 19. Pp. 2131–2137.
54. Gus'kova O. A., Khalatur P. G., Bäuerle P., Khokhlov A. R. Silk-inspired 'molecular chimeras': Atomistic simulation of nanoarchitectures based on thiophene-peptide copolymers // *Chem Phys Lett.* 2008. Vol. 461, no. 1-3. Pp. 64–70.
55. Gus'kova O., Schillinger E., Khalatur et al. Биоинспирированные гибридные системы на основе олиготиофена и пептидов (ALA-GLY)<sub>n</sub>: компьютерное моделирование адсорбционных слоёв // *Высокомолекулярные соединения, Серия А*. 2009. Vol. 51, no. 3. Pp. 1–17.

56. Gus'kova O., Khalatur P., Khokhlov A. Molecular chimeras: New strategies in the design of functional materials // *Nanotechnologies in Russia*. 2008. Vol. 3, no. 7. Pp. 481–493.
57. Frenkel D., Smit B. *Understanding molecular simulation*. Academic Press, 2002.
58. Hoover W. *Time Reversibility, Computer Simulation, and Chaos*. World Scientific, 2001.
59. Jorgensen W. L., Tirado-Rives J. The OPLS Force Field for Proteins. Energy Minimization for Crystals of Cyclic Peptides and Crambin. // *J. Am. Chem. Soc.* 1988. Vol. 110. Pp. 1657–1666.
60. Jorgensen W. L., Maxwell D. S., Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids // *J. Am. Chem. Soc.* 1996. Vol. 118, no. 45. Pp. 11225–11236.
61. Sun H. Force field for computation of conformational energies, structures, and vibrational frequencies of aromatic polyesters // *J. Comp. Chem.* 1994. Vol. 15, no. 7. Pp. 752–768.
62. Jarzynski C. Nonequilibrium equality for free energy differences // *Phys Rev Lett.* 1997. Vol. 78. Pp. 2690–2693.
63. Crooks G. E. Path-ensemble averages in systems driven far from equilibrium // *Phys. Rev. E.* 2000. Vol. 61. Pp. 2361–2366.
64. Shirts M. R., Bair E., Hooker G., Pande V. S. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods // *Phys. Rev. Lett.* 2003. Vol. 91. Pp. 140601–140601–4.
65. Bennet C. Efficient estimation of free energy differences from Monte Carlo data // *J. Comp. Phys.* 1976. Vol. 22. Pp. 245–268.
66. berg M. A., Lyubartsev A. P., Jacobsson S. P., Laaksonen A. Determination of solvation free energies by adaptive expanded ensemble molecular dynamics // *J. Chem. Phys.* 2004. Vol. 120, no. 8. Pp. 3770–3776.
67. Lee B., Richards F. M. The interpretation of protein structures: estimation of static accessibility. // *J. Mol. Biol.* 1971. Vol. 55, no. 3. Pp. 379–400.
68. Connolly M. L. Analytical molecular surface calculation // *J. App. Cryst.* 1983. Vol. 16, no. 5. Pp. 548–558.

69. Benjamin I. Molecular Structure and Dynamics at Liquid-Liquid Interfaces // *Annu. Rev. Phys. Chem.* 1997. Vol. 48. Pp. 407–451.
70. Pratt L. R., Pohorille A. Hydrophobic Effects and Modeling of Biophysical Aqueous Solution Interfaces // *Chem. Rev.* 2002. Vol. 102, no. 8. Pp. 2671–2692.
71. Chipot C., Pearlman D. A. Free Energy Calculations. The Long and Winding Gilded Road // *Molecular Simulation*. 2002. Vol. 28, no. 1. Pp. 1–12.
72. Maccallum J. L., Tieleman P. D. Calculation of the water-cyclohexane transfer free energies of neutral amino acid side-chain analogs using the OPLS all-atom force field // *J. Comput. Chem.* 2003. Vol. 24, no. 15. Pp. 1930–1935.
73. Villa A., Mark A. E. Calculation of the free energy of solvation for neutral analogs of amino acid side chains // *J. Comput. Chem.* 2002. Vol. 23, no. 5. Pp. 548–553.
74. Shirts M. R., Pande V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration // *J. Chem. Phys.* 2005. Vol. 122, no. 14. P. 144107.
75. Ytreberg M. F., Swendsen R. H., Zuckerman D. M. Comparison of free energy methods for molecular systems // *J. Chem. Phys.* 2006. Vol. 125, no. 18. Pp. 184114–184124.
76. Hess B., Vandervegt N. F. A. Hydration Thermodynamic Properties of Amino Acid Analogues: A Systematic Comparison of Biomolecular Force Fields and Water Models // *J. Phys. Chem. B.* 2006. Vol. 110, no. 35. Pp. 17616–17626.
77. Mobley D. L., Dumont E., Chodera J. D., Dill K. A. Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent // *J. Phys. Chem. B.* 2007. Vol. 111, no. 9. Pp. 2242–2254.
78. Shirts M. R., Pande V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models // *J. Chem. Phys.* 2005. Vol. 122, no. 13. P. 134508.
79. Shirts M. R., Pitner J. W., Swope W. C., Pande V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins // *J. Chem. Phys.* 2003. Vol. 119, no. 11. Pp. 5740–5761.

80. Shirts M. R., Bair E., Hooker G., Pande V. S. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods // *Phys Rev Lett*. 2003. Vol. 91, no. 14. Pp. 140601–140604.
81. Jorgensen W. L., Maxwell D. S., Rives T. J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids // *J. Am. Chem. Soc.* 1996. Vol. 118, no. 45. Pp. 11225–11236.
82. Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F. et al. Untitled // *Intermolecular Forces*, Ed. by A. Wilkinson, P. Weiner, W. F. van Gunsteren. Reidel, Dordrecht, 1981. Vol. 3. P. 331.
83. Van Der Spoel D., Lindahl E., Hess B. et al. GROMACS: fast, flexible, and free. // *J. Comput. Chem.* 2005. Vol. 26, no. 16. Pp. 1701–1718.
84. Ryckaert J.-P., Ciccotti G., Berendsen H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes // *J. Comput. Phys.* 1977. Vol. 23, no. 3. Pp. 327–341.
85. Bennett C. H. Efficient estimation of free energy differences from Monte Carlo data // *J. Comput. Phys.* 1976. Vol. 22, no. 2. Pp. 245–268.
86. Frenkel D., Smith B. Acceptance ratio method // *Understanding Molecular Simulation*. Academic Press, 2002. P. 189.
87. van Gunsteren W. F., Berendsen H. J. C. A leap-frog algorithm for stochastic dynamics // *Mol. Sim.* 1988. Vol. 1. Pp. 173–185.
88. Hess B., Bekker H., Berendsen H. J. C., Fraaije J. G. E. M. LINCS: A linear constraint solver for molecular simulations // *J. Comput. Chem.* 1997. Vol. 18, no. 12. Pp. 1463–1472.
89. Miyamoto S., Kollman P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models // *J. Comput. Chem.* 1992. Vol. 13, no. 8. Pp. 952–962.
90. Parrinello M., Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method // *J. Appl. Phys.* 1981. Vol. 52, no. 12. Pp. 7182–7190.
91. Darden T., York D., Pedersen L. Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems // *J. Chem. Phys.* 1993. Vol. 98, no. 12. Pp. 10089–10092.



92. Essmann U., Perera L., Berkowitz M. L. et al. A smooth particle mesh Ewald method // J. Chem. Phys. 1995. Vol. 103, no. 19. Pp. 8577–8593.
93. Handbook of Chemistry and Physics. CRC, Boca Raton, FL, 1986.
94. Chen F., Smith P. E. Simulated surface tensions of common water models // J. Chem. Phys. 2007. Vol. 126, no. 22. P. 221101.
95. Radzicka A., Wolfenden R. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution // Biochemistry. 1988. Vol. 27, no. 5. Pp. 1664–1670.
96. Bull B. H., Breese K. Surface Tension of Amino Acid Solutions: A Hydrophobicity Scale of the Amino Acid Residues // Arch. Biochem. Biophys. 1974. Vol. 161. Pp. 665–670.
97. Li Z. X., Lu J. R., Styrkas D. A. et al. The structure of the surface of ethanol/water mixtures // Mol. Phys. 1993. Vol. 80, no. 4. Pp. 925–939.
98. Kipling J. J. // J. Coll. Interf. Sci. 1963. Vol. 18. P. 502.
99. Donaldson D. J., Anderson D. Adsorption of Atmospheric Gases at the Air-Water Interface. 2. C1-C4 Alcohols, Acids, and Acetone // J. Phys. Chem. A. 1999. Vol. 103, no. 7. Pp. 871–876.
100. Okhapkin I., Askadskii A., Markov V. et al. Two-dimensional classification of amphiphilic monomers based on interfacial and partitioning properties. 2. Amino acids and amino acid residues // Colloid Polym. Sci. 2006. Vol. 284, no. 6. Pp. 575–585.
101. Oostenbrink C., Villa A., Mark A. E., Van Gunsteren W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6 // J. Comp. Chem. 2004. Vol. 25, no. 13. Pp. 1656–1676.
102. Bieri O., Kiefhaber T. Kinetic models in protein folding // Mechanisms in Protein Folding, Ed. by R. H. Pain. Oxford, UK: Oxford University Press, 2000.
103. Kamtekar S., Schiffer J. M., Xiong H. et al. Protein Design by Binary Patterning of Polar and Nonpolar Amino Acids // Science. 1993. Vol. 262, no. 5140. Pp. 1680–1685.
104. Desjarlais J. R., Handel T. M. De novo design of the hydrophobic cores of proteins. // Protein Sci. 1995. Vol. 4, no. 10. Pp. 2006–2018.

105. Khokhlov A. R., Khalatur P. G. Conformation-Dependent Sequence Design (Engineering) of AB Copolymers // *Phys. Rev. Lett.* 1999. Vol. 82, no. 17. Pp. 3456–3459.
106. Chothia C. Hydrophobic bonding and accessible surface area in proteins // *Nature*. 1976. Vol. 248, no. 5446. Pp. 338–339.
107. Wertz D. H., Scheraga H. A. Influence of water on protein structure. An analysis of the preferences of amino acid residues for the inside or outside and for specific conformations in a protein molecule. // *Macromolecules*. 1978. Vol. 11, no. 1. Pp. 9–15.
108. Wolfenden R., Andersson L., Cullis P. M., Southgate C. C. Affinities of amino acid side chains for solvent water. // *Biochemistry*. 1981. Vol. 20, no. 4. Pp. 849–855.
109. Guy H. R. Amino acid side-chain partition energies and distribution of residues in soluble proteins. // *Biophys. J.* 1985. Vol. 47, no. 1. Pp. 61–70.
110. Rose G. D., Geselowitz A. R., Lesser G. J. et al. Hydrophobicity of amino acid residues in globular proteins // *Science*. 1985. Vol. 229, no. 4716. Pp. 834–838.
111. Miller S., Janin J., Lesk A. M., Chothia C. Interior and surface of monomeric proteins // *J. Mol. Biol.* 1987. Vol. 196, no. 3. Pp. 641–656.
112. Lawrence C., Auger I., Mannella C. Distribution of accessible surfaces of amino acids in globular proteins. // *Proteins*. 1987. Vol. 2, no. 2. Pp. 153–161.
113. Janin J., Miller S., Chothia C. Surface, subunit interfaces and interior of oligomeric proteins. // *J. Mol. Biol.* 1988. Vol. 204, no. 1. Pp. 155–164.
114. Samanta U., Bahadur R. P., Chakrabarti P. Quantifying the accessible surface area of protein residues in their local environment // *Protein Eng., Des. Sel.* 2002. Vol. 15, no. 8. Pp. 659–667.
115. Wolfenden R. Experimental Measures of Amino Acid Hydrophobicity and the Topology of Transmembrane and Globular Proteins // *J. Gen. Physiol.* 2007. Vol. 129, no. 5. Pp. 357–362.
116. Pohl F. M. Empirical protein energy maps. // *Nat. New Biol.* 1971. Vol. 234, no. 52. Pp. 277–279.
117. Nozaki Y., Tanford C. The Solubility of Amino Acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions. Establishment of a Hydrophobicity Scale. // *J. Biol. Chem.* 1971. Vol. 246, no. 7. Pp. 2211–2217.

118. Fauchère J. L., Pliska V. Octanol-to-water solvation free energies of acetyl amino amides (Ac-X-amides) // Eur. J. Med. Chem. – Chim. Ther. 1983. Vol. 18. P. 369.
119. Wimley W. C., Creamer T. P., White S. H. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. // Biochemistry. 1996. Vol. 35, no. 16. Pp. 5109–5124.
120. URL: <ftp://ftp.ncbi.nih.gov/mmdb/nrtable/nrpdb.061207>.
121. Altschul S. F., Gish W., Miller W. et al. Basic local alignment search tool. // J. Mol. Biol. 1990. Vol. 215, no. 3. Pp. 403–410.
122. Hubbard S. J., Thornton J. M. 'NACCESS', Computer Program. Department of Biochemistry and Molecular Biology, University College London, 1993.
123. Hubbard S. J., Campbell S. F., Thornton J. M. Molecular recognition : Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors // J. Mol. Biol. 1991. Vol. 220, no. 2. Pp. 507–530.
124. Prabhakaran M., Ponnuswamy P. Spatial assignment of amino acid residues in globular proteins: An approach from information theory // J. Theor. Biol. 1980. Vol. 87, no. 4. Pp. 623–637.
125. Schillinger E.-K., Mena-Osteritz E., Bäuerle P. Синтез диблок олигомера из алкилированного тетрагидрофурана и пептидной последовательности [Thr-Val]<sub>3</sub> // не опубликовано.
126. Peng Z., Ewig C. S., Hwang M.-J. et al. Derivation of Class II Force Fields. 4. van der Waals Parameters of Alkali Metal Cations and Halide Anions // J. Phys. Chem. A. 1997. Vol. 101, no. 39. Pp. 7243–7252.
127. Grimme S. Do Special Noncovalent pi-pi Stacking Interactions Really Exist? // Ang. Chem. Int. Ed. 2008. Vol. 47, no. 18. Pp. 3430–3434.
128. Plimpton S. Fast Parallel Algorithms for Short-Range Molecular Dynamics // J Comp Phys. 1995. Vol. 117, no. 1. Pp. 1–19.
129. Luhers T., Ritter C., Adrian M. et al. 3D structure of Alzheimer's amyloid-beta(1-42) fibrils // Proc. Nat. Acad. Sci. USA. 2005. Vol. 102, no. 48. Pp. 17342–17347.

130. Frishman D., Argos P. Knowledge-based protein secondary structure assignment // *Proteins: Struct. Funct. Genet.* 1995. Vol. 23, no. 4. Pp. 566–579.
131. Rodríguez-Ropero F., Casanovas J., Alemán C. Ab initio calculations on pi-stacked thiophene dimer, trimer, and tetramer: Structure, interaction energy, cooperative effects, and intermolecular electronic parameters // *J. Comp. Chem.* 2008. Vol. 29, no. 1. Pp. 69–78.
132. Tsuzuki S., Honda K., Azumi R. Model Chemistry Calculations of Thiophene Dimer Interactions: Origin of  $\pi$ -Stacking // *J. Am. Chem. Soc.* 2002. Vol. 124, no. 41. Pp. 12200–12209.
133. Stone D. A., Hsu L., Stupp S. I. Self-assembling quinquethiophene-oligopeptide hydrogels // *Soft Matter*. 2009. Vol. 5, no. 10. Pp. 1990–1993.
134. Ландау Л. Д., Лившиц Е. М. Теоретическая физика, т.5 Статистическая физика, §87-91, Теория слабых растворов. Москва: Физматлит, 1995.
135. Ben-Naim A., Marcus Y. Solvation thermodynamics of nonionic solutes // *J Chem Phys.* 1984. Vol. 81. Pp. 2016–2027.
136. Shirts M., Pietra J., Swope W., Pande V. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins // *J Chem Phys.* 2003. Vol. 119. Pp. 5740–5761.
137. Gibbs J. *Heterogeneous Equilibrium*. 1875.
138. Jaynes E. The Gibbs Paradox // *Maximum Entropy and Bayesian Methods* / Ed. by C. Smith, G. Erickson, P. Neudorfer. Kluwert Academic Publishers, Dordrecht, Holland, 1992. Pp. 1–22.
139. Buff F. P., Lovett R. A., Stillinger F. H. Interfacial Density Profile for Fluids in the Critical Region // *Phys. Rev. Lett.* 1965. Vol. 15, no. 15. Pp. 621–623.
140. Penfold J. The structure of the surface of pure liquids // *Rep. Prog. Phys.* 2001. Vol. 64, no. 7. Pp. 777–814.
141. Senapati S., Berkowitz M. L. Computer Simulation Study of the Interface Width of the Liquid/Liquid Interface // *Phys. Rev. Lett.* 2001. Vol. 87, no. 17. P. 176101.
142. Rivera J. L., Starr F. W., Paricaud P., Cummings P. T. Polarizable contributions to the surface tension of liquid water // *J. Chem. Phys.* 2006. Vol. 125, no. 9. Pp. 094712–094712–8.

- 143. Robert M. Absence of Phase Separation for Fluids in Three Dimensions // Phys. Rev. Lett. 1985. Vol. 54, no. 5. Pp. 444–446.
- 144. Landau L. D., Lifshitz E. M., Pitaevskii L. P. Theoretical Physics: Mechanics. 3rd edition. Oxford : Butterworth-Heinemann, 2001.
- 145. Mcquarrie D. A. Statistical mechanics. Sausalito, California: University Science Books, 2000.
- 146. Kirkwood J. G. Statistical Mechanics of Fluid Mixtures // J. Chem. Phys. 1935. Vol. 3, no. 5. Pp. 300–313.
- 147. Adamson A. W. Physical Chemistry of Surfaces. 5 edition. Wiley- Interscience: New York, 1990.

## Приложение А

### Типы аминокислот и их боковые цепи

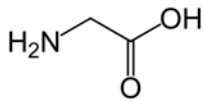

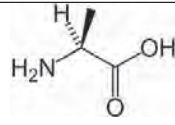
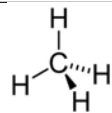
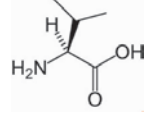

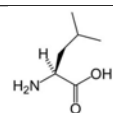
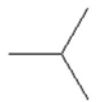
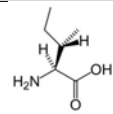

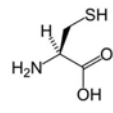
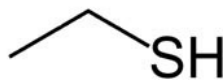
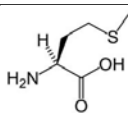
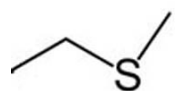
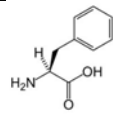
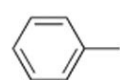
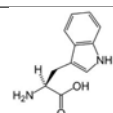
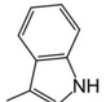
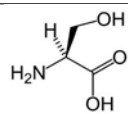
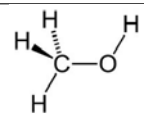
Название	Обозначение	Формула	R-Н-аналог
Глицин	GLY		
Аланин	ALA		
Валин	VAL		
Лейцин	LEU		
Изолейцин	ILE		
Цистеин	CYS		
Метионин	MET		
Фенилаланин	PHE		
Триптофан	TRP		
Серин	SER		

Таблица А.1. Аминокислоты и R-Н-аналоги их боковых цепей

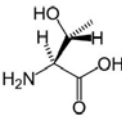
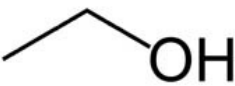
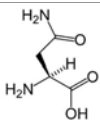
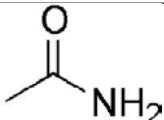
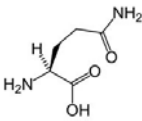
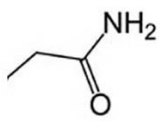
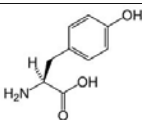
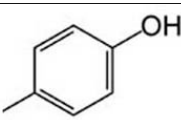
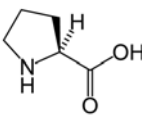
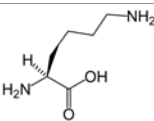
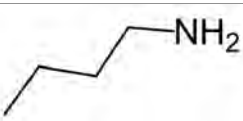
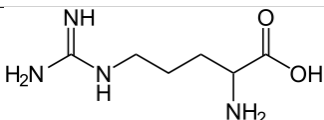
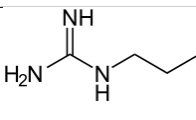
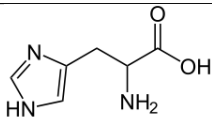
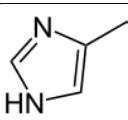
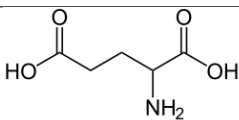
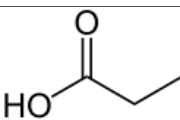
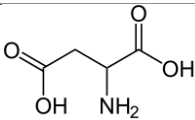
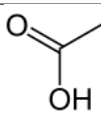
Название	Обозначение	Формула	R-Н-аналог
Треонин	THR		
Аспарагин	ASN		
Глутамин	GLN		
Тирозин	TYR		
Пролин	PRO		не определён
Лизин	LYS		
Аргинин	ARG		
Гиститдин	HIS		
Глутаминовая кислота	GLU		
Аспарагиновая кислота	ASP		

Таблица А.2. Продолжение

## Приложение Б

### Свободная энергия сольватации

#### Б.1. Введение

Классическая теория жидкостей и растворов является достаточно хорошо разработанной областью статистической физики (см. например, [134]). Развёрнутое описание процессов сольватации различных молекул с точки зрения термодинамики приведено в работе [135]. В этой же работе приводится соотношение между свободной энергией сольватации и энергией взаимодействия молекулы с окружающим растворителем. Вывод формул пригодных для вычисления энергии сольватации в компьютерном эксперименте (при использовании изобаро-изотермического ансамбля) приводится в приложении к работе [136].

В данном разделе приводится развёрнутый вывод конечной расчётной формулы Б.33. При этом, отличие от работы [136] состоит в использовании при выводе постулатов только классической статистической физики, так как, по мнению автора, использование “квантовых” подходов несколько усложняет понимание вывода, и вводит неискушённого читателя в заблуждение относительно границ применимости квантовых и классических моделей.

#### Б.2. Свободная энергия сольватации

Поясним ещё раз, что же такое свободная энергия сольватации. В большинстве экспериментальных работ свободную энергию сольватации молекулярного вещества S в жидкости В измеряют следующим образом. Допустим, у нас есть двухфазная система, состоящая из жидкой фазы (вещество В) и газообразной фазы. Кроме того, в системе присутствует вещество S в малой концентрации. Тогда, измеряя концентрации вещества S в жидкой и газообразной фазах (или парциальное давление вещества S в газообразной фазе), говорят, что величина (Б.1) является свободной энергией сольватации вещества S в жидкости В [136].

$$\Delta G_{solv} = kT \ln \frac{P_S}{c_s^B kT} \quad (\text{Б.1})$$

В литературе [135] величина  $\Delta G_{solv}$  обычно интерпретируется как свободная энергия, связанная с взаимодействием (coupling) молекулы растворённого вещества с остальной частью раствора. Из следующих параграфов станет ясно, почему используется такая интерпретация.



### Б.3. Постулаты классической статфизики

Проблема, излагаемая в данном разделе, напрямую связана с определением термодинамического параметра (свободной энергии сольватации) из микроскопических параметров наших молекул. Получением соотношений между микроскопическими параметрами системы и её термодинамическими характеристиками занимается статистическая физика. Мы будем широко использовать связь между статистической суммой и свободной энергией (Б.2-Б.3).

$$F = -kT \ln \frac{Z_V}{N!} \quad (\text{Б.2})$$

$$Z_V = \int \exp \left\{ -\frac{H(\mathbf{q}, \mathbf{p})}{kT} \right\} d\mathbf{q} d\mathbf{p} \quad (\text{Б.3})$$

Для нашего дальнейшего рассмотрения необходимо понимать в какой момент и откуда берётся множитель  $N!$  в выражении (Б.2).

### Б.4. Связь статсуммы со свободной энергией

#### Б.4.1. Квантовый подход

Изложение статфизики можно изначально строить из квантовых предположений, вводя энтропию через логарифм вырожденности энергетического уровня с энергией  $E$ . Тогда статсумма запишется как (Б.4).

$$Z_V^Q = \text{Tr} \exp\left(-\frac{\hat{H}}{kT}\right) \quad (\text{Б.4})$$

Тогда при переходе к классическому пределу по квазиклассическим формулам с некоторыми приближениями (см. например, [57]) можно получить (Б.5), где  $d$ -размерность пространства.

$$Z_Q^V \approx \frac{1}{h^{dN}} \int d\mathbf{p} d\mathbf{q} \exp \left\{ -\frac{H(\mathbf{p}, \mathbf{q})}{kT} \right\} \quad (\text{Б.5})$$

Однако, в формуле (Б.5) пока нет множителя  $1/N!$ . Появление этого множителя объясняется в литературе из принципа неразличимости частиц в квантовой механике. Так для ферми- или бозе- газа правильное выражение для квантового состояния должно быть записано как полная симметризация или антисимметризация по всем перестановкам, например, для бозе-газа (Б.6).

$$\begin{aligned}
|p_1 \dots p_N \rangle_s &= \frac{1}{N!} \sum_{\sigma} |p_{\sigma_1} \dots p_{\sigma_N} \rangle \\
|q_1 \dots q_N \rangle_s &= \frac{1}{N!} \sum_{\sigma} |q_{\sigma_1} \dots q_{\sigma_N} \rangle
\end{aligned} \tag{Б.6}$$

Можно показать, что множитель  $1/N!$  появившийся в формуле (Б.6) перейдёт в квазиклассическом приближении в выражение для статсуммы, и мы получим (Б.7)

$$Z_Q^V \approx \frac{1}{h^{dN} N!} \int d\mathbf{p} d\mathbf{q} \exp \left\{ -\frac{H(\mathbf{p}, \mathbf{q})}{kT} \right\} \tag{Б.7}$$

Итак, смысл множителя  $1/N!$  связан с учётом неразличимости частиц и отражает количество способов, которыми можно переставить одинаковые частицы. Обобщением формулы (Б.7) на случай, когда в системе присутствуют два сорта частиц будет формула, где  $N!$  заменено на соответствующее количество перестановок, то есть  $N_1! * N_2!$ .

#### Б.4.2. Подход Гиббса

Приведённые выше рассуждения, как считается в литературе, позволяют объяснить, так называемый, *парадокс Гиббса* [137]. Суть парадокса состоит в том, что энтропия смешения двух “идеальных” газов не зависит от природы этих газов, однако меняется скачком, если мы считаем газы неразличимыми. В формуле (Б.7) это будет соответствовать замене  $N_1! * N_2!$  на  $(N_1 + N_2)!$ .

В ходе дальнейшего рассмотрения методики компьютерного эксперимента мы столкнёмся с необходимостью чётко представлять себе, в каком случае нужно использовать множитель  $N_1! * N_2!$ , а в каком  $-(N_1 + N_2)!$ . Для целей данной работы, по мнению автора, квантовый подход к разрешению парадокса Гиббса не совсем прозрачен по двум причинам:

- Во-первых, мы изначально оговорили, что будем рассматривать систему как классическую. И с этой точки зрения объяснение парадокса Гиббса через квантовую механику представляет собой некоторый трюк.
- И во-вторых, в дальнейшем мы столкнёмся с необходимостью плавно изменять Гамильтониан взаимодействия одной частицы со всей остальной системой. И, разумеется, в классических вычислениях не будет никакой разницы между динамикой системы с изначальным Гамильтонианом и Гамильтонианом, отличающимся от исходного на бесконечно малую величину. Поэтому нам потребуется ясное понимание происходящих процессов, или хотя бы универсальных алгоритмов, в соответствии с которыми нужно действовать.

Замечательное рассмотрение вопросов связанных с парадоксом Гиббса приведено в работе [138]. Объяснение, излагаемое в этой работе, по-видимому, соответствует оригинальному взгляду Гиббса. Для нашего дальнейшего рассмотрения мы рассмотрим выводы работы [138].

Итак, прежде чем заниматься описанием системы с точки зрения статфизики мы должны определиться с набором макропараметров и не менять этот набор по ходу рассмотрения. Набор макропараметров зависит от тех экспериментальных фактов и возможностей, которые нам необходимо описать. Допустим, у нас есть смесь двух различных газов, однако, наши экспериментальные методики не позволяют нам их различить. Тогда на некоторых примерах можно показать, что экспериментальные результаты будут согласовываться с теорией в двух случаях:

- если мы честно записываем свободную энергию системы как сумму энергий двух различных газов

$$F/kT = N_1 \ln \frac{N_1}{V} + N_2 \ln \frac{N_2}{V}$$

- или если мы считаем газы неразличимыми и пишем

$$F/kT = (N_1 + N_2) \ln \frac{(N_1 + N_2)}{V}$$

Однако, если в нашем эксперименте существуют и используются методики, которые могут различить два газа, то описать экспериментальные результаты можно будет только с помощью первого выражения.

Хотя в [138] это показывается только при конкретном рассмотрении одного из возможных экспериментов, авторы данной работы предполагают, что может быть сделано следующее общее утверждение:

Если в рамках модели не существенно различие между частицами, одним макропараметром можно описать характеристики сразу нескольких типов частиц, то можно использовать выражение для статсуммы одинаковых частиц, если же мы ожидаем, что различие между частицами проявится в эксперименте, мы соотносим им разные макропараметры, то при теоретическом рассмотрении необходимо делить статсумму на произведение факториалов.

## Б.5. Сольватация, рассмотрение статфизики

Рассмотрим теперь эксперимент по определению свободной энергии сольватации, описанный в пункте Б.2, с точки зрения классической статфизики. Рассмотрим газ ве-

щества S с парциальным давлением  $P_s$ . Свободная энергия Гиббса системы задаётся выражением (Б.8)

$$G(N_s^g, P_s, T) = -kT \ln Z^P(N_s^g, P_s, T) \quad (\text{Б.8})$$

где статсумма  $Z^P$  определяется как

$$Z^P(N_s^g, P_s, T) = \int_0^\infty dV e^{(-\frac{P_s V}{kT})} Z^V(N_s^g, V, T) \quad (\text{Б.9})$$

Наша задача состоит в том, чтобы найти равновесные концентрации вещества S в двух фазах. Равновесие будет достигнуто, когда химпотенциалы вещества S в обеих фазах сравниваются.

$$\mu = \frac{\partial G}{\partial N} \quad (\text{Б.10})$$

$$\mu_s^{gas} = \mu_s^{liquid} \quad (\text{Б.11})$$

Займемся нахождением химпотенциалов.

## Газовая фаза

Предположим, что в газовой фазе можно пренебречь взаимодействием молекул между собой, тогда Гамильтониан системы примет вид

$$H(\mathbf{p}, \mathbf{q}) = \sum_i^{N_s} H_s$$

, где  $H_s$ -гамильтониан каждой молекулы, может быть представлен в виде движение центра масс молекулы и внутреннего Гамильтониана взаимодействий:

$$H_s(p_1, \dots, p_n, q_1, \dots, q_n) = \frac{p_{mc}^2}{2m} + H_{int}(\tilde{p}_1 \dots \tilde{p}_{n-1}, \tilde{q}_1 \dots \tilde{q}_{n-1})$$

Тогда

$$Z^V(N_s^g, V, T) = \frac{1}{h^{dN_s^g} N_s^g!} (\Lambda * V * q_s(T))^{N_s^g}$$

где

$$q_s(T) = \int \exp\left(-\frac{H_{int}(\tilde{p}_1 \dots \tilde{p}_{n-1}, \tilde{q}_1 \dots \tilde{q}_{n-1})}{kT}\right) d\tilde{p}_1 \dots d\tilde{p}_{n-1} d\tilde{q}_1 \dots d\tilde{q}_{n-1}$$

$$\Lambda = (2\pi M kT)^{3/2}$$

Подставляя эти выражения в (Б.8) получим

$$G(N_s^g, P_s, T) = -kT N_s^g \ln \left\{ \frac{\Lambda q_s(T) kT}{h^{dN_s^g} P_s} \right\} \quad (\text{Б.12})$$

Найдём также выражение для химпотенциала

$$\begin{aligned}\mu(N_s^g, P_s, T) &= G(N_s^g, P_s, T) - G(N_s^g - 1, P_s, T) \\ \mu_s^g(N_s^g, P_s, T) &= -kT \ln \left\{ \frac{\Lambda_{q_s}(T)kT}{h^d * P_s} \right\}\end{aligned}\quad (\text{Б.13})$$

## Жидкая фаза

Найдём теперь химпотенциал вещества S в жидкой фазе B.

$$\begin{aligned}G(N_s^l, N_b, P, T) &= -kT \ln Z^P(N_s^l, N_b, P, T) \\ \mu_a^l(N_s^l, N_b, P, T) &= G(N_s^l, N_b, P, T) - G(N_s^l - 1, N_b, P, T)\end{aligned}\quad (\text{Б.14})$$

теперь привлечём соображения пункта Б.4.2. Поскольку в нашем случае молекулы вещества S и B разные даже в экспериментальном смысле, веществу S соответствует отдельный макропараметр - концентрация вещества S, то для теоретического рассмотрения мы должны использовать следующее выражение для статсуммы (Б.15)

$$Z^P(N_s^l, N_b, P, T) = \frac{1}{h^{d(N_s^l + N_b)} N_s^l! N_b!} \int d\mathbf{p} d\mathbf{q} dV \exp \left\{ -\frac{H(\mathbf{p}, \mathbf{q}) + PV}{kT} \right\} \quad (\text{Б.15})$$

Тогда обозначив интеграл в выражении (Б.15) за  $\Delta(N_s^l, N_b, P, T)$  получим (Б.16)

$$\begin{aligned}Q(N_s^l, N_b, V, T) &= \int d\mathbf{p} d\mathbf{q} \exp \left\{ -\frac{H(\mathbf{p}, \mathbf{q})}{kT} \right\} \\ \Delta(N_s^l, N_b, P, T) &= \int dV \exp \left\{ -\frac{PV}{kT} \right\} * Q(N_s^l, N_b, V, T) \\ \mu_s^l(N_s^l, N_b, P, T) &= -kT \ln \left\{ \frac{\Delta(N_s^l, N_b, P, T)}{N_s^l * h^d * \Delta(N_s^l - 1, N_b, P, T)} \right\}\end{aligned}\quad (\text{Б.16})$$

Расчёт (Б.16) свёлся у нас лишь к расчёту интегралов типа  $\Delta(N_s^l, N_b, P, T)$ . Предположим, что Гамильтониан нашей системы (жидкость B + вещество S) можно записать в следующем виде (Б.17) [136].

$$H(N_s^l, N_b) = T_s(N_s^l) + T_b(N_b) + U_s(N_s^l) + U_b(N_b) + U_{b-b}(N_b) + U_{s-s}(N_s^l) + U_{s-b}(N_s^l, N_b) \quad (\text{Б.17})$$

Для простоты рассмотрим случай парных взаимодействий:

$$H(N_s^l, N_b) = \sum_{N_s^l} H_s + \sum_{N_b} H_b + \sum_{N_s^l} \sum_{N_s^l} U_{s-s} + \sum_{N_b} \sum_{N_b} U_{b-b} + \sum_{N_b} \sum_{N_s^l} U_{s-b} \quad (\text{Б.18})$$

Выделим в Гамильтониане одну из молекул растворённого вещества:

$$H(N_a^l, N_b) = \sum_{a=1}^{N_a^l-1} H_a + \sum_{b=1}^{N_b} H_b + \sum_{a=1}^{N_a^l-1} \sum_{b=1}^{N_b} U_{a-a} + \sum_{b=1}^{N_b} \sum_{b=1}^{N_b} U_{b-b} + \sum_{b=1}^{N_b} \sum_{a=1}^{N_a^l-1} U_{a-b} + \left\{ H_a + \sum_{a=1}^{N_a^l-1} U_{a-a} + \sum_{b=1}^{N_b} U_{a-b} \right\} \quad (\text{Б.19})$$

Далее сделаем следующий трюк: пусть у нас есть  $H(N_s^l, N_b, \lambda)$ , обладающий следующим свойством:

$$\begin{aligned} H(N_s^l, N_b, \lambda = 0) &= \sum_{s=1}^{N_s^l-1} H_s + \sum_{b=1}^{N_b} H_b + \sum_{s=1}^{N_s^l-1} \sum_{s=1}^{N_s^l-1} U_{s-s} + \sum_{b=1}^{N_b} \sum_{b=1}^{N_b} U_{b-b} + \sum_{b=1}^{N_b} \sum_{s=1}^{N_s^l-1} U_{s-b} + \{H_s\} \\ H(N_s^l, N_b, \lambda = 1) &= H(N_s^l, N_b) \end{aligned} \quad (\text{Б.20})$$

Простейший вариант такого Гамильтониана:

$$H(N_s^l, N_b) = \sum_{s=1}^{N_s^l-1} H_s + \sum_{b=1}^{N_b} H_b + \sum_{s=1}^{N_s^l-1} \sum_{s=1}^{N_s^l-1} U_{s-s} + \sum_{b=1}^{N_b} \sum_{b=1}^{N_b} U_{b-b} + \sum_{b=1}^{N_b} \sum_{s=1}^{N_s^l-1} U_{s-b} + H_s + \lambda \left\{ \sum_{s=1}^{N_s^l-1} U_{s-s} + \sum_{b=1}^{N_b} U_{s-b} \right\} \quad (\text{Б.21})$$

Сопоставим каждому Гамильтониану соответствующий интеграл  $Q$ . Заметим, что из-за разделения переменных при интегрировании:

$$\begin{aligned} Q(N_s^l, N_b, V, T, \lambda = 1) &= Q(N_s^l, N_b, V, T) \\ Q(N_s^l, N_b, V, T, \lambda = 0) &= Q(N_s^l - 1, N_b, V, T) * Q(N_s^l = 1, N_b = 0, V, T) \end{aligned} \quad (\text{Б.22})$$

Теперь сопоставим им интегралы типа  $\Delta$

$$\begin{aligned} \Delta(N_s^l, N_b, P, T, \lambda = 1) &= \Delta(N_s^l, N_b, P, T) \\ \Delta(N_s^l, N_b, P, T, \lambda = 0) &= \int dV \exp(-\frac{PV}{kT}) Q(N_s^l - 1, N_b, V, T) * Q(N_s^l = 1, N_b = 0, V, T) \end{aligned} \quad (\text{Б.23})$$

Рассмотрим подробнее второе выражение в (Б.24), его можно преобразовать к следующему виду (Б.27)

$$\begin{aligned} \Delta(N_s^l, N_b, P, T, \lambda = 0) &= \int_0^\infty dV \exp(-\frac{PV}{kT}) Q(N_s^l - 1, N_b, V, T) * Q(N_s^l = 1, N_b = 0, V, T) \\ &= \int_0^\infty dV \exp(-\frac{PV}{kT}) Q(N_s^l - 1, N_b, V, T) * (\Lambda * V * q_s(T)) \end{aligned} \quad (\text{Б.24})$$

$$= \Lambda q_s(T) \Delta(N_s^l - 1, N_b, P, T) \frac{\int_0^\infty dV V \exp(-\frac{PV}{kT}) Q(N_s^l - 1, N_b, V, T)}{\Delta(N_s^l - 1, N_b, P, T)} \quad (\text{Б.25})$$

$$= \Delta(N_s^l - 1, N_b, P, T) * \Lambda q_s(T) V^* \quad (\text{Б.26})$$

$$= \Delta(N_s^l - 1, N_b, P, T) * Q(N_s = 1, V^*, T) \quad (\text{Б.27})$$

где  $V^*$  обозначен последний множитель, по сути это средний объём системы состоящей из  $N_s^l - 1$  молекул вещества А и  $N_b$  вещества В при давлении Р.

Итак, проведя данные преобразования, мы можем записать химпотенциал в следующем виде (Б.30)

$$\mu_s^l(N_s^l, N_b, P, T) = -kT \ln \left\{ \frac{\Delta(N_s^l, N_b, P, T)}{N_s^l * h^d * \Delta(N_s^l - 1, N_b, P, T)} \right\} \quad (\text{Б.28})$$

$$= -kT \ln \left\{ \frac{1}{N_s^l * h^d} \frac{\Delta(N_s^l, N_b, P, T, \lambda = 1)}{\Delta(N_s^l, N_b, P, T, \lambda = 0)} \frac{\Delta(N_s^l, N_b, P, T, \lambda = 0)}{\Delta(N_s^l - 1, N_b, P, T)} \right\} \quad (\text{Б.29})$$

$$= -kT \ln \frac{1}{N_s^l * h^d} + \Delta G_{sim} - kT \ln \Lambda q_s(T) V^* \quad (\text{Б.30})$$

где  $\Delta G_{sim}$  определено (Б.31)

$$\Delta G_{sim} = -kT \ln \frac{\Delta(N_s^l, N_b, P, T, \lambda = 1)}{\Delta(N_s^l, N_b, P, T, \lambda = 0)} \quad (\text{Б.31})$$

## Равновесие

Итак, в равновесии у нас должны быть равны химпотенциалы вещества А в обоих фазах, приравнивая (Б.13) и (Б.30) получаем:

$$-kT \ln \left\{ \frac{\Lambda q_s(T) kT}{h^d * P_s} \right\} = -kT \ln \frac{1}{N_s^l * h^d} + \Delta G_{sim} - kT \ln \Lambda q_s(T) V^* \quad (\text{Б.32})$$

Произведём упрощение выражения:

$$-kT \ln \left\{ \frac{kT}{P_s} \right\} = -kT \ln \frac{1}{N_s^l} + \Delta G_{sim} - kT \ln V^*$$

Пусть  $V_l$ -объём системы, состоящей из  $N_s^l$  молекул вещества А и  $N_b$  вещества В, тогда

$$-kT \ln \left\{ \frac{kT N_s^l}{P_s V_l} \right\} = \Delta G_{sim} - kT \ln \frac{V^*}{V_l}$$

И, вспоминая (Б.1), получаем

$$-kT \ln \left\{ \frac{kT c_s^B}{P_s} \right\} = \Delta G_{sim} - kT \ln \frac{V^*}{V_l} = \Delta G_{solv} \quad (\text{Б.33})$$

Итак, мы свели задачу расчёта свободной энергии сольватации к расчёту следующего отношения интегралов

$$\Delta G_{sim} = -kT \ln \frac{\int dV d\mathbf{p} d\mathbf{q} \exp \left\{ -\frac{H(\mathbf{p}, \mathbf{q}, \lambda=1) + PV}{kT} \right\}}{\int dV d\mathbf{p} d\mathbf{q} \exp \left\{ -\frac{H(\mathbf{p}, \mathbf{q}, \lambda=0) + PV}{kT} \right\}} \quad (\text{Б.34})$$

По сути выражение Б.34 определяет разницу свободной энергии двух систем с *одинаковым числом частиц*, но разными параметрами взаимодействия. Это изменение свободной энергии уже может быть рассчитано в компьютерном эксперименте.

## Приложение В

### Приложения к главе 3

#### В.1. Профиль плотности границы раздела вода-пар

Профиль плотности границы раздела вода-пар удовлетворительно описывается полуфеноменологической капиллярно-волновой теорией (КВТ). Согласно КВТ [69, 139] профиль плотности может быть описан как некоторый молекулярный профиль, уширенный термическими капиллярными волнами. Экспериментальные результаты и результаты моделирования [140–142] допускают, что толщина границы раздела зачастую может быть описана только капиллярными волнами, то есть является практически ступенчатой на молекулярном уровне. В этом случае профиль плотности может быть описан функцией интеграла ошибок:

$$\rho_z = \frac{1}{2}(\rho_L + \rho_V) - \frac{1}{2}(\rho_L - \rho_V)\operatorname{erfc}\left[\frac{(z - z_0)}{\sigma\sqrt{2}}\right] \quad (\text{В.1})$$

где  $\rho_L$  и  $\rho_V$  плотность жидкости и пара, а  $\sigma$  толщина границы раздела.

Известной проблемой при моделировании границ раздела является зависимость толщины границы от линейных размеров поверхности  $L_x$ . Согласно КВТ имеется логарифмическая зависимость толщины от линейного размера (из-за зависимости от размера количества возможных волновых мод):

$$\sigma^2 = \frac{k_B T}{2\pi\gamma} \ln\left(\frac{L_{||}}{B_0}\right) \quad (\text{В.2})$$

где  $\gamma$  поверхностное натяжение,  $B_0$  характерная длина коротковолновой обрезки спектра, обычно полагаемая равной размеру молекулы.

В термодинамическом пределе (без учёта гравитации) толщина границы плоского раздела фаз - расходится для жидкостей только с взаимодействиями ближнего порядка при рассмотрении в трёх и менее измерениях [143].

#### В.2. Теория профилей свободной энергии

На рис. 3.1 представлена геометрия двухфазной системы с плоской границей раздела, которая перпендикулярна оси Z. В системе присутствует вода, пар и тестовая молекула S, для которой мы хотим посчитать её профиль свободной энергии вдоль оси Z. Система находится при постоянной температуре и давлении (за счёт хорошей сжимаемости пара). Предположим, что система описывается в классическом приближении, тогда система из



$N$  частиц описывается своей функцией потенциальной энергии [144]  $U(\mathbf{q})$ , зависящей от  $3N$  обобщённых координат  $\mathbf{q}$ . Чтобы отслеживать положение тестовой молекулы введём функцию  $d_z(\mathbf{q})$ , которая является проекцией расстояния между центрами масс тестовой молекулы и водного слоя на ось  $Z$ . Основные формулы статистической физики [145], которые нам нужны, это соотношения связывающие распределение вероятности, профили свободной энергии и потенциал средней силы. В нашем случае вероятность найти молекулу  $S$  на расстоянии  $z$  от центра масс водного слоя есть:

$$P(z) = \langle \delta(d_z - z) \rangle = \frac{\int \delta(d_z - z) e^{-U/kT} d\mathbf{q}}{\int e^{-U/kT} d\mathbf{q}} \quad (\text{B.3})$$

где  $\langle \dots \rangle$  усреднение по каноническому ансамблю,  $k$  постоянная Больцмана,  $T$  -температура.

Зная распределение, мы можем рассчитать профиль свободной энергии:

$$W(z) = -kT \ln P(z) + F_0 \quad (\text{B.4})$$

- где  $F_0$  произвольная константа. В эксперименте профиль вероятности будет соответствовать профилю концентрации, предполагая, конечно, что раствор является разбавленным, или идеальным. Кирквуд [146] ввёл понятие потенциала средней силы (ПСС) и показал, что профиль потенциала средней силы с точностью до константы равен профилю свободной энергии. В нашем случае средней силой является средняя сила взаимодействия тестовой молекулы и водного слоя при условии, что проекция расстояния между центрами масс на ось  $Z$  фиксирована. Средняя сила задаётся как:

$$\overline{F(z)} = \frac{\int \langle \delta(d_z(\mathbf{q}) - z) \rangle \nabla_{d_z} U(\mathbf{q}) e^{-U/kT} d\mathbf{q}}{\int e^{-U/kT} d\mathbf{q}} \quad (\text{B.5})$$

где  $\nabla_{d_z}$  оператор градиента, связанный с расстоянием вдоль оси  $Z$  между центрами масс. Можно показать, что профиль свободной энергии и потенциал средней силы – это одно и то же:

$$\frac{\partial W(z)}{\partial z} = \overline{F(z)} \quad (\text{B.6})$$

### В.3. К вопросу вычисления энергий гидратации

В случае попарно-аддитивных молекулярных взаимодействий функцию потенциальной энергии можно разбить таким образом, чтобы выделить взаимодействие между тестовой молекулой и растворителем:

$$U(\mathbf{q}, \lambda) = U_{wat}(\mathbf{q}_{wat}) + \lambda U_{int}(\mathbf{q}_{wat}, \mathbf{q}_S) + U_S(\mathbf{q}_S) \quad (\text{B.7})$$

где  $U_{wat}, U_S, U_{int}$ , потенциальные энергии молекул воды, молекулы растворённого вещества и энергия их взаимодействия, соответственно. Обобщённые координаты разбиваются удобным образом так, что  $\mathbf{q}_S$  – это координаты, описывающие конфигурацию молекулы S (молекулы растворённого вещества) в ящике с водой,  $\mathbf{q}_{wat}$  – координаты всех молекул воды.  $\lambda$  – это параметр сопряжения, вводимый для последующих выкладок. При  $\lambda = 1$  и  $\lambda = 0$  взаимодействия между молекулой S и водой включены и отключены полностью, соответственно.

Энергия гидратации, как показано в приложении Б, – это свободная энергия (работа) включения взаимодействий между молекулой S и остальным раствором в NPT ансамбле. Энергия гидратации Гиббса  $\Delta G_{hydr}$  может быть выражена:

$$\Delta G_{hydr} = -kT \ln \left( \frac{\Delta(P, T, \lambda = 1)}{\Delta(P, T, \lambda = 0)} \right) \quad (\text{B.8})$$

где  $\Delta(P, T, \lambda)$  – изотермо-изобарическая функция распределения вида:

$$\Delta(P, T, \lambda) \propto \frac{1}{N!} \int \int \int dP d\mathbf{q} \exp(-[U(\mathbf{q}, \lambda) + PV]/kT) \quad (\text{B.9})$$

и N – число молекул воды. Такая формулировка верна в термодинамическом пределе ( $N \rightarrow \infty$ ) пока у нас есть одна молекула S. Можно показать, что химический потенциал растворённого вещества (молекулы S) в идеальном растворе выражается как:

$$\mu_S = -kT \ln(1/c) + \Delta G_{hydr} \quad (\text{B.10})$$

где c – концентрация растворённого вещества.

Таким образом с экспериментальной точки зрения свободная энергия гидратации может быть получена путём измерения концентрации вещества S в воде  $c_{wat}$  и паре  $c_{gas}$ :

$$\Delta G_{hydr}^{exp} = kT \ln \left( \frac{c_{gas}}{c_{wat}} \right) \quad (\text{B.11})$$

## В.4. Энергии адсорбции

Параметры, описывающие поверхностную активность веществ, обычно описываются в терминах теории адсорбции Гиббса (хотя есть и другие: Ван-дер-Ваальса, Кана-Хилларда). В теории Гиббса [147] поверхность раздела считается отдельной двумерной фазой, расположенной на поверхности раздела Гиббса (ПРГ), и характеризуется двумерными величинами концентрации, называемыми, избытками Гиббса  $\Gamma^i$ ). Согласно Гиббсу, объёмные фазы характеризуются своими объёмными концентрациями вплоть до границы раздела, а любой избыток или недостаток вещества в районе границы раздела описывается

соответствующими двумерными избытками Гиббса. Зная профиль концентрации  $c^i(z)$ , а также величины концентрации в объёме  $c_1^i, c_2^i$ , избыток Гиббса может быть рассчитан по формуле:

$$\Gamma^i = \int_{-\infty}^{z_0} (c^i(z) - c_1^i) dz + \int_{z_0}^{\infty} (c^i(z) - c_2^i) dz \quad (\text{B.12})$$

где  $z_0$  - координата ПРД. Позиция ПРД обычно выбирается так, чтобы избыток Гиббса для растворителя был равен нулю. В этих переменных можно записать уравнение адсорбции Гиббса, которое соотносит избыток Гиббса ( $\Gamma^S$ ), концентрацию вещества в объёмной фазе ( $c^S$ ) и поверхностное натяжение ( $\gamma$ ):

$$\Gamma^S = -\frac{1}{kT} \frac{d\gamma}{d \ln c^S} \quad (\text{B.13})$$

Мерой поверхностной активности обчно считается отношение  $\Gamma^S/c^S$  в пределе бесконечно разбавленного раствора ( $c^S \rightarrow 0$ ). Это отношение можно оценить путём измерения поверхностного натяжения при различных концентрациях, что следует из (B.13):

$$\left. \frac{\Gamma^S}{c^S} \right|_{c^S \rightarrow 0} = -\frac{1}{kT} \left. \frac{d\gamma}{dc^S} \right|_{c^S \rightarrow 0} \quad (\text{B.14})$$

Понятие энергии адсорбции может быть далее введено, хотя оно остаётся до некоторой степени произвольно определённым, поскольку оно зависит от выбора стандартных состояний или толщины границы раздела (толщины поверхностного монослоя). Один из возможных вариантов состоит в определении величины толщины поверхностного слоя  $\tau$ , тогда энергию адсорбции можно записать в виде:

$$G_{ads} = -kT \ln \frac{\Gamma^S}{c^S \tau} \quad (\text{B.15})$$

Однако при сравнении экспериментальных значений энергии всегда нужно обращать внимание на то, какому определению следовали экспериментаторы.

## Приложение Г

### Приложения к главе 3: профили свободной энергии

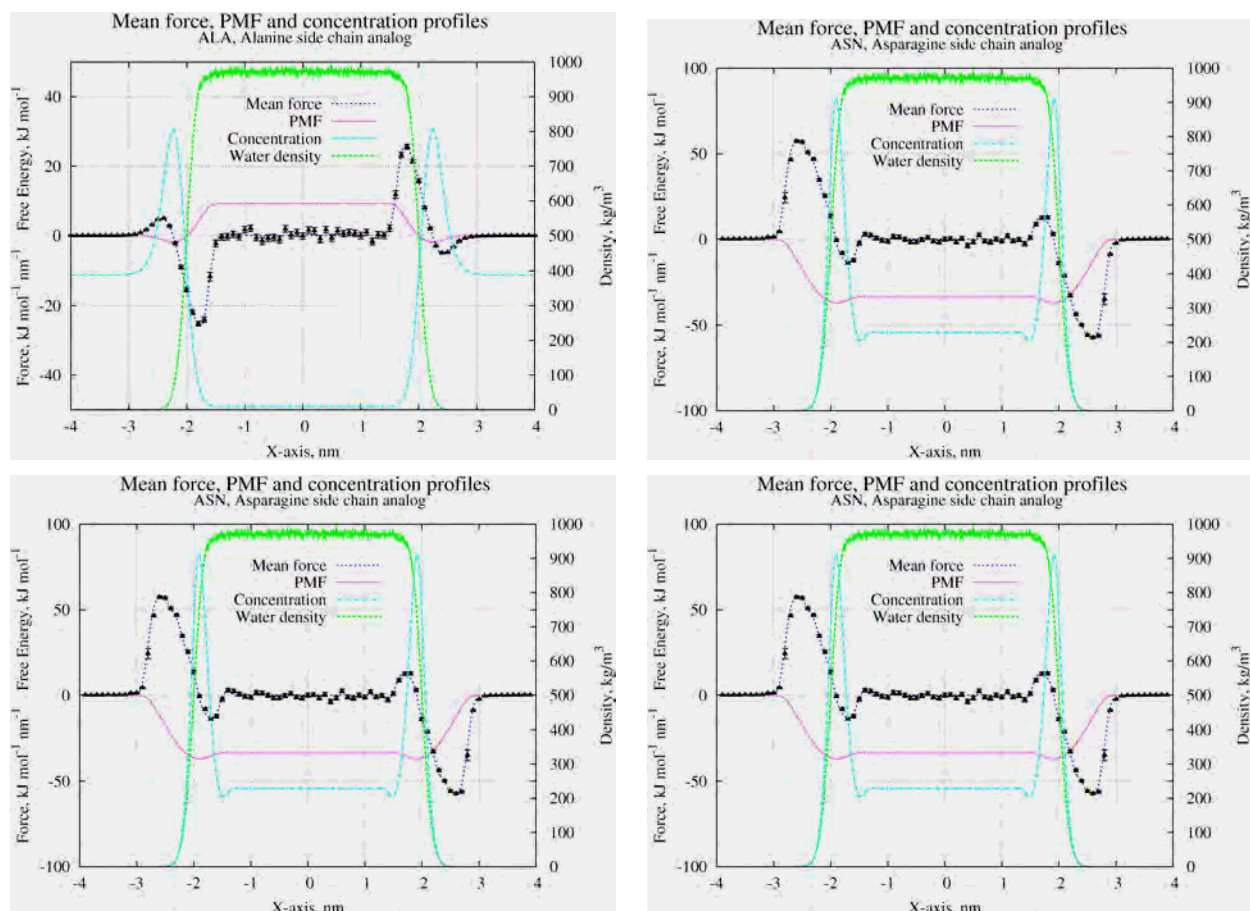


Таблица Г.1. Профили свободной энергии боковых цепей аминокислот. Комментарии к графикам аналогичны комментариям к рис. 3.4, глава 3.

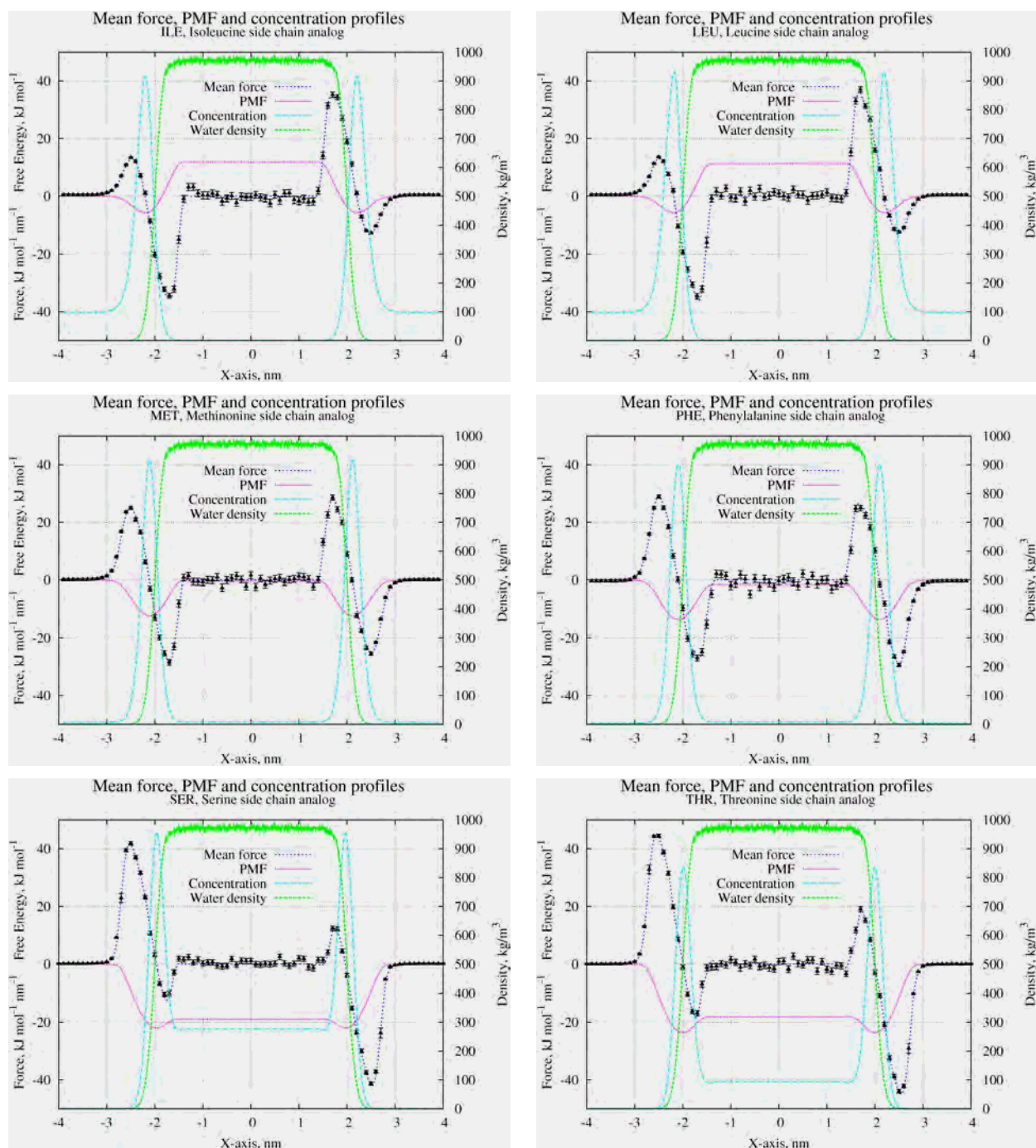


Таблица Г.2. Профили свободной энергии боковых цепей аминокислот. Комментарии к графикам аналогичны комментариям к рис. 3.4, глава 3.

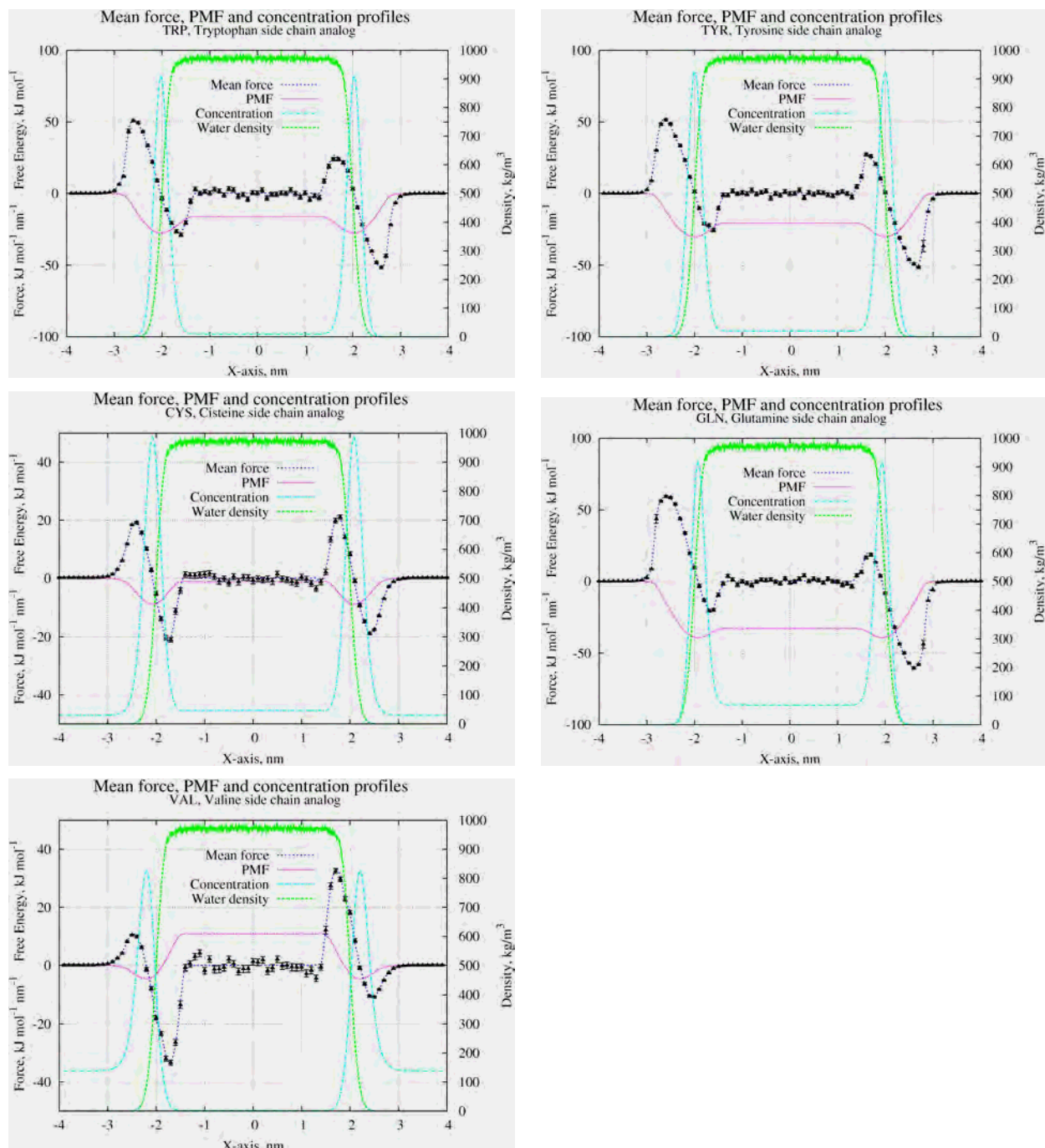


Таблица Г.3. Профили свободной энергии боковых цепей аминокислот. Комментарии к графикам аналогичны комментариям к рис. 3.4, глава 3.