Федеральное государственное бюджетное образовательное учреждение высшего образования
«Курский Государственный Университет»

На правах рукописи

Нерух Дмитрий Александрович

Сложная динамика больших биомолекулярных систем в водных растворах

1.5.4. Биохимия
1.5.8. Математическая биология, биоинформатика

Диссертация на соискание учёной степени
доктора биологических наук

Москва — 2026

Оглавление

Стр.

Введение 6

Глава 1. Гибридное атомистически-гидродинамическое представление жидкостей и растворов 12

1.1 Введение 12

1.2 Теория 18

1.2.1 Модель флуктуационной гидродинамики для описания макроскопических потоков 19

1.2.2 Сопряжение атомистического и гидродинамического представлений 21

1.2.3 Динамика: законы сохранения массы и импульса 24

1.2.4 Реализация законов сохранения массы и импульса 25

1.2.5 Дифференциальная форма: основные уравнения модели двусторонней связи 27

1.2.6 Упрощенная модель односторонней связи 32

1.3 Модели жидкости 33

1.3.1 Модель «Mercedes Benz» 33

1.3.2 MD реализация модели MB 38

1.3.3 Модель MB: Соответствие реальной воде и гидродинамике 43

1.3.4 3D модели: аргон, вода и пептид в воде 48

1.3.5 Диаланин в воде 48

1.3.6 Вирус PCV2 в воде 50

1.4 Особенности компьютерной реализации гибридного моделирования 54

1.4.1 Реализация гидродинамики двухмерной модели воды «Mercedes Benz» . . . 54

1.4.2 Детали численного решения уравнений гидродинамики сплошной среды и связь с MD частью решения: аргон и вода 55

1.4.3 Конструирование s-функции 57

1.4.4 Диаланин в воде 58

1.4.5 Свиной цирковирус в воде 60

1.4.6 Высокопроизводительное молекулярное моделирование: MDGRAPE 62

1.5 Результаты 66

1.5.1 Вода в модели MB 66

1.5.2 Гибридное моделирование аргона 70

1.5.3 Гибридное моделирование воды 77

1.5.4 Гибридное моделирование диаланина в воде 79

1.5.5 Вирус PCV2 86

1.6 Выводы 87

Глава 2. Сложность растворов биомолекул с точки зрения нелинейной динамики . . . 92

2.1 Количественная оценка долговременной памяти атомистических траекторий молекулярных систем 92

2.1.1 Методология оценки памяти системы . . . 93
2.1.2 Результаты и обсуждение . . . 96
2.1.3 Выводы . . . 99
2.2 Оценка молекулярной сложности траекторий с точки зрения динамического хаоса 100
2.2.1 Вычислительная механика имеет дело с определенным разбиением молекулярного фазового пространства . . . 103
2.2.2 Численная реализация . . . 107
2.2.3 Результаты . . . 110
2.2.4 Выводы . . . 116
2.3 Статистическая сложность позволяет отличить динамический хаос от стохастического процесса . . . 117
2.3.1 Изученные системы и методы анализа . . . 120
2.3.2 Численные эксперименты . . . 120
2.3.3 Гипотеза о свойстве Марковости . . . 121
2.3.4 Обсуждение . . . 124

Глава 3. Структура и динамика вирусов в моделях с атомистическим разрешением . . 127
3.1 Обзор: существующие работы по полноатомному моделированию целых вирусов . 127
3.2 Изучение капсида бактериофага **MS2** с использованием полноатомной молекулярной динамики: распределение зарядов на капсиде и динамика ионов из окружающего раствора . . . 133
3.2.1 Детали моделирования . . . 134
3.2.2 Результаты и обсуждение . . . 136
3.2.3 Выводы . . . 141
3.3 Моделирование атомной молекулярной динамики целого капсида вируса **PCV2** выявило критически важную роль правильного распределения ионов в стабильности капсида . . . 141
3.3.1 Моделируемая система . . . 143
3.3.2 Результаты . . . 145
3.3.3 Поток ионов и воды через стенку капсида PCV2 . . . 150
3.3.4 Выводы . . . 150
3.4 Подробности распределения заряда в стабильном вирусном капсиде **PCV2** . . . 152
3.4.1 Детали моделирования . . . 153
3.4.2 Результаты и обсуждение . . . 154
3.4.3 Выводы . . . 159
3.5 Молекулярная динамика капсидного белка **PCV2** при постоянном значении pH раскрывает механизм сборки капсида . . . 161
3.5.1 Экспериментальная часть . . . 163
3.5.2 Результаты . . . 166
3.5.3 Обсуждение результатов . . . 170

Глава 4. Белки и взаимодействие белок-лиганд 177
4.1 Немарковская модель динамики пептидов 177
4.1.1 Молекулярная система 178
4.1.2 Методология моделирования 179
4.1.3 Результаты 180
4.1.4 Выводы 185
4.2 Вода управляет конформационными переходами пептидов 187
4.2.1 Анализируемые величины: конформационные переходы в диаланине 187
4.2.2 Поведение воды 189
4.2.3 Выводы 195
4.3 Самосборка молекул трегалозы на поверхности лизоцима: гипотеза разбитого стекла 196
4.3.1 Детали моделирования 197
4.3.2 Результаты 198
4.3.3 Обсуждение 199
4.3.4 Выводы 201
4.4 Диффузия лиганда на поверхности белка, наблюдаемая при моделировании молекулярной динамики 201
4.4.1 Моделируемая система 203
4.4.2 Результаты 204
4.4.3 Выводы 206
4.5 Энергия связывания серии новых веществ с альбумином 206
4.5.1 Калориметрическое титрование 207
4.5.2 Подготовка белковых структур BAT2 207
4.5.3 Подготовка структур лиганда 208
4.5.4 Прогоны молекулярной динамики (МД) для расчёта ABFE в BAT2 209
4.5.5 Результаты 210
4.5.6 Выводы 210
4.6 Константа скорости диссоциации белок-лиганд, полученная с помощью полноатомного моделирования 212
4.6.1 Теория 213
4.6.2 Молекулярная модель и детали моделирования 216
4.6.3 Результаты и обсуждение 218
4.6.4 Выводы 219

Заключение 232

Словарь терминов 234

Список рисунков 271

Список таблиц 286

Приложение А. Предел возможностей метода Молекулярной Динамики 287

Приложение Б. Вывод уравнений движения с учетом сохранения массы и импульса . 288
Б.1 Сохранение массы 288
Б.2 Сохранение импульса 290
Б.3 Гибридная вода модели MB 292

Приложение В. Вывод гибридных уравнений движения в дифференциальной форме . 293

Приложение Г. Вычислительная механика 296

Приложение Д. Моделирование вируса MS2 297
Д.1 Детали моделирования вируса MS2 297
Д.2 Подготовка ячейки моделирования 297

Приложение Е. Детали распределения ионов на внутренней поверхности капсида PCV2 299

Приложение Ж. Подробности моделирования PCV2 300

Приложение З. Моделирование белков PCV2 про постоянном pH 303
З.1 Сходимость результатов 303
З.2 Сходимость значений pKa 303
З.3 Репрезентативные структуры 304

Приложение И. Константа скорости диссоциации белок-лиганд, полученная с помощью полноатомного моделирования 318
И.1 Подробности моделирования молекулярной динамики 318
И.2 Вероятность выживания 318

Введение

Актуальность темы. Современные суперкомпьютеры способны моделировать молекулярные системы размером в сотни миллионов атомов в рамках модели классической молекулярной динамики (МД), описывающей атомы как точечные массы, взаимодействующие по закону Ньютона. И хотя общедоступные модели состоят из десятков, максимум сотен тысяч атомов, рекордные результаты опубликованы для системы размером больше миллиарда (10^9) атомов. Такие масштабы позволяют говорить о перспективах полноатомного моделирования *целой живой клетки* (оценочные размеры – от половины до шести миллиардов атомов), а также целого ряда *биохимических процессов*, представляющих интерес для экспериментальной биологии. И, действительно, моделирование, например, целых вирусных капсидов или клеточных органелл в последние годы не является уже чем-то экстраординарным, хотя и по прежнему требует значительных вычислительных ресурсов.

Сложность, однако, заключается в том, что такие впечатляющие масштабы достижимы только по пространству, в то время как по времени рост значительно скромнее. Это связано с принципиальными ограничениями физической концепции МД, которая позволяет увеличивать количество одновременно моделируемых атомов через распределение атомов по параллельным вычислительным ядрам, что невозможно сделать для временных шагов: миллионный шаг моделирования может быть выполнен только после последовательного выполнения всех предыдущих шагов. Поэтому на сегодняшний день максимально достижимые времена моделирования не превышают нескольких микросекунд, максимум – десятков микросекунд, даже для скромных по количеству атомов систем.

Важно подчеркнуть, что это – принципиальное ограничение, связанное как с алгоритмом МД, так и с принципами, на которых построены компьютеры. Чтобы проиллюстрировать ситуацию, рассмотрим идеализированный “компьютер”, состоящий из 100 миллионов вычислительных ядер, которые обмениваются данными с памятью со скоростью света (максимально возможная скорость), а расстояние между “ядром” и “памятью” – 1 см. Также будем считать, что один МД шаг состоит из 4 элементарных операций. Тогда расчёт *одной секунды* жизни следующих систем займёт:

– самая маленькая биологическая клетка ($3 \cdot 10^8$ атомов): *7 дней*,

– клетка человека (10^{14} атомов): *6342 лет*,

– условная “кошка” (10^{10} клеток или 10^{18} атомов): *190 258 752 лет*.

Подробности расчётов этих оценок приведены в приложении к диссертации. Т.е. атомистическое моделирование макроскопических сложных объектов на данный момент *принципиально невозможно*.

Очевидно, что необходимо другое описание системы, не атомистическое, по крайней мере в тех частях системы, где это возможно. Например, вода и водные растворы, занимающие превалирующую часть биологических систем, может быть описана в рамках более экономной гидродинамики на макро- и мезо-масштабах. Несмотря на то, что гидродинамика была создана и развивалась для строго макроскопических систем (“бесконечное” число атомов в каждом

“элементарном” объёме гидродинамического поля – фундаментальное предположение в гидродинамике), современные обобщения уравнений классической гидродинамики позволяют описывать микрообъёмы, содержащие всего несколько сотен молекул. Это позволяет распространить гидродинамическое описание до тех же масштабов, на которых действует атомистическая молекулярная динамика, и заменить атомистическое описание воды в тех частях системы, где молекулярные подробности несущественны.

Таким образом, актуальной становится разработка подходов, сочетающих атомистическое и гидродинамическое описание системы в рамках одной модели, с физически обоснованным сопряжением этих фундаментально разных описаний вещества. Такие “гибридные” модели активно разрабатываются последние 20-30 лет, пожалуй наиболее развиты модели с разного рода “огрубляющим” или “крупнозернистым” представлением атомов, когда группа атомов объединяется в одну квази-частицу, что снижает вычислительные затраты. Более предпочтительным представлением на больших масштабах является именно гидродинамика, вернее её обобщение – флуктуационная гидродинамика, предложенная Ландау и Лифшицем и активно развивающаяся по сей день. Здесь, в отличие от разного рода “мегаатомов”, соблюдены все механические, термодинамические и кинетические фундаментальные принципы: от законов сохранения вещества и энергии до корректно воспроизводимых коэффициентов переноса.

Также очевидно, что при разработке таких гибридных моделей ключевую роль играет корректное воспроизведение физико-химических свойств вещества, как в микроскопической атомистической, так и в макроскопической гидродинамической частях модели. Этот аспект предъявляет ещё более сложные требования к модели, особенно в частях системы, где эти два описания “сопрягаются” или переходят друг в друга. Здесь необходимо воспроизведение не только основных физических свойств вещества, таких как плотность, давление, температура и т.д., но и более тонких параметров, таких как коэффициенты переноса, динамические характеристики или физхимические параметры, как константы ассоциации химических реакций, биофизических и *биохимических* характеристик. Как показывает практика, именно эти последние свойства вещества особенно сложны для расчёта в компьютерных моделях.

Степень разработанности темы. Несмотря на активную разработку гибридных моделей в мировой литературе, эта задача весьма нетривиальна и на сегодняшний день достаточно далека от завершения. Как правило, существующие модели обеспечивают выполнение только одного или нескольких основополагающих законов, оставляя прочие за пределами внимания. Обычно это связано с конкретной областью, в которой модель вводится и используется. Например, в “гранулированных” моделях стараются, чтоб выполнялся закон сохранения энергии (в силу “Молекулярно Динамического” генезиса этих моделей), оставляя без внимания закон сохранения импульса (локально) и иногда даже вещества. Последнее связано со сложностью “перетока” массы между атомами и мезочастицами, в силу нефизичности последних. С другой стороны, модели, основанные на обобщении гидродинамических подходов, часто упускают из виду закон сохранения энергии, который, как правило, выполняется автоматически для гидродинамических потоков в жидкостях при комнатных физико-химических условиях.

Следует также подчеркнуть, что разработка физически обоснованной гибридной модели неизбежно требует высокопроизводительного компьютерного расчёта с использованием очень больших МД систем на самых высокопроизводительных суперкомпьютерах.

Целью данной работы является разработка гибридных атомистически-гидродинамических методов на основе высокопроизводительного компьютерного моделирования больших молекулярных систем в водном растворе, а также применение этих методов наряду с традиционными методами молекулярного моделирования для получения биохимических характеристик ряда биомолекулярных систем с точностью, количественно совпадающей с экспериментальными биохимическими данными.

Для достижения поставленной цели необходимо было решить следующие **задачи**.

1. Разработать теорию гибридной модели “Молекулярная Динамика - Гидродинамика” с получением соответствующих уравнений движения.
2. Воплотить разработанную теорию в компьютерном коде, как в специально для этой цели созданной программе, так и встроив её в общеупотребительный пакет GROMACS, включающем параллельные вычисления для высокопроизводительных суперкомпьютеров.
3. Исследовать физико-химические и биохимические свойства таких гибридных систем для нескольких моделей жидкостей и биомолекулярных растворов, включая свойства этих систем с точки зрения нелинейной динамики и динамической сложности.
4. Провести моделирование полноатомной модели капсида для двух вирусов – PCV2 и MS2, а также исследовать их физико-химические и биохимические свойства в растворе как для чисто Молекулярно Динамического представления, так и гибридного, описанного выше.
5. Проанализировать динамику молекулярных систем с точки зрения нелинейной динамики и теории сложности. Связать эти свойства с особенностями движения молекул воды и биомолекул в растворе.
6. Исследовать несколько моделей белка в растворе с целью изучения таких свойств как их конформационная динамика и механизмы процессов ассоциации и диссоциации малых молекул с белком.
7. Доказать возможность применения разработанных гибридных методов для решения прикладных задач биохимии: моделирования поведения белков в водных растворах, их взаимодействия с другими белками и небелковыми биомолекулами, моделирования механизмов инфицирования клетки вирусами, моделирования взаимодействия лигандов с белковыми мишенями.

Научная новизна.

1. Впервые разработана и реализована в программном пакете гибридная модель жидкости, связывающая на основе физически обоснованных фундаментальных принципов атомистическое и континуальное, гидродинамическое описания; связь осуществлена в двух вариантах – как в одном направлении от гидродинамики к атомам, так и в обоих направлениях самосогласованно между представлениями.
2. Впервые построена полноатомная модель капсидов двух вирусов, PCV2 и MS2, проведено Молекулярно Динамическое моделирование этих капсидов, изучены физико-химические и биохимические свойства системы, включая характеристики зарядов на

капсиде и распределение и потоки ионов из раствора; показана критическая роль правильного распределения ионов между внутренним пространством капсида и внешним раствором в стабильности капсида.

3. Динамика воды и водных растворов проанализирована с точки зрения теории нелинейных динамических систем, получены характеристики хаотического движения молекул воды, количественно оценена сложность этой динамики.
4. Получены количественные характеристики влияния воды на конформационную динамику белков; рассчитана константа диссоциации малой молекулы-лиганда изониазида и белка каталазы с количественным совпадением с экспериментальной величиной этой константы; по результатам моделирования получена свободная энергия взаимодействия ряда новых производных тиазолопиримидина с альбумином, показано количественное совпадение с экспериментально измеренными значениям.

Теоретическая и практическая значимость состоит в разработке и апробации нового подхода в моделировании жидкостей и растворов биологических макромолекул, который позволяет описывать моделируемую систему одновременно на масштабах, различающихся на несколько порядков, как по пространству, так и по времени. При этом исследователь может произвольно определять области, в которых атомистическое разрешение необходимо (например, вблизи макромолекул), описывая остальное пространство с помощью экономной гидродинамики без ущерба для качества описания атомистических областей. Подход реализован в общеупотребительном пакете Молекулярной Динамики GROMACS и доступен для свободного использования. Также получены несколько моделей капсидов целого вируса с атомистическим разрешением. Эти модели особенно важны для обработки результатов современной криоэлектронной микроскопии, которая включает подобное моделирование как необходимый этап в обработке измеренных данных. При этом сложность заключается в неравномерном разрешении экспериментально измеренных электронных плотностей, что требует нетривиальных подходов в построении атомистической модели по измеренным данным. Наконец, очень важна валидация полученной модели, что можно сделать только имея согласованную МД модель и проведя моделирование на существенных временах, что было сделано для описанных биомолекулярных систем.

Результаты работы имеют важное значение для решения прикладных задач биохимии, медицинской химии и экспериментальной медицины: моделирования поведения и взаимодействия белковых и небелковых молекул в водных растворах. Например, полученная константа диссоциации изониазида из активного центра каталазы совпала с экспериментально измеренной величиной в пределах погрешности эксперимента и расчета. Это – достаточно редкий случай в моделировании такого рода величин, где обычно совпадение в пределах 2-3 порядков величины считается хорошим.

Методология и методы исследования. Теоретическая часть исследования включала разработку уравнений движения с использованием подходов из фундаментально разных областей физики жидкостей: континуальной гидродинамики, описывающей жидкость с помощью полей плотности, скорости и внутренней энергии и дискретной Молекулярной Динамики, оперирующей конечным набором точечных масс. Сопряжение этих двух разнородных подходов с выполнением

фундаментальных законов сохранения (массы, импульса и энергии) составляло суть теоретической части работы.

Практическое воплощение полученных уравнений движения было проведено путём встраивания новых уравнений в пакет GROMACS, что требовало нетривиальных методов кодирования с использованием языка программирования C++ и методов параллельных вычислений, адаптированных для высокопроизводительных суперкомпьютеров.

Наконец, расчеты проводились на самых быстрых (на момент расчетов) суперкомпьютерах в мире: универсальных K-computer и Fugaku (RIKEN Center for Computational Science, Japan) и специализированном суперкомпьютере MDGRAPE-4 (RIKEN BDR Institute, Japan), построенном для быстрых расчетов с использованием Молекулярной Динамики. Также расчеты производили на других суперкомьютерах: Athena, Cirrus, Sulis (HPC Midlands+ Consortium, Великобритания) и нескольких локальных кластерах в университете Астон (Бирмингем, Великобритания) и Санкт-Петербургском Научно-Исследовательском Институте Фтизиопульмонологии.

Результаты расчетов валидировались экспериментальными биохимическими и биологическими методами и в большинстве случаев совпадали количественно.

Основные положения, выносимые на защиту.

1. Теоретическая концепция, связывающая атомистическую Молекулярную Динамику на микромасштабах и флуктуационную гидродинамику на макромасштабах, с плавным переходом одного описания в другое при изменении масштаба и выполнением законов сохранения на всех промежуточных масштабах; полученные уравнения движения в рамках этой концепции.
2. Программное воплощение этих уравнений движения, встроенное в пакет GROMACS, адаптированное для расчётов на высокопроизводительных суперкомпьютерах.
3. Результаты моделирования в рамках этой концепции ряда жидких систем: от простых жидкостей до растворов малых молекул, пептидов, белков и целых вирусных капсидов; валидация подхода с использованием экспериментальных данных и результатов моделирования с использованием классической Молекулярной Динамики.
4. Результаты моделирования полноатомного представления вирусных капсидов бактериофага MS2 и вируса PCV2 и анализ их физических и биохимических свойств, в том числе распределения зарядов, самих капсидов, а также водного раствора внутри и вокруг них; сравнение результатов этого моделирования с результатами моделирования с использованием гибридной модели, описанной выше; анализ влияния протонирования ключевых аминокислотных остатков белков капсида PCV2 на процесс самосборки капсида на основании моделирования с использованием метода Молекулярной Динамики при постоянном pH.
5. Анализ атомистических траекторий воды и водных растворов с точки зрения нелинейной динамики и хаоса, в том числе с использованием подхода “Computational Mechanics” для количественной оценки статистической сложности этих траекторий; демонстрация возможностей метода различить хаотическую и стохастическую траекторию, а также наличие долговременных нетривиальных корреляций в хаотической траектории воды, в отличие от статистически неотличимой суррогатной стохастической траектории.

6. Анализ конформационной динамики пептида в водном растворе, выявление немарковского характера этой динамики, анализ роли молекул воды в такой динамике.
7. Биохимическое моделирование сродства нескольких новых соединений-кандидатов в противоопухолевые препараты к альбумину; количественное соответствие полученных термодинамических характеристик для этих веществ с экспериментально измеренными калориметрическим титрованием.
8. Моделирование динамики малых молекул во взаимодействии с белком в водном растворе, в том числе расчёт констант ассоциации и диссоциации с получением значений, количественно совпадающих с экспериментальными данными.

Достоверность полученных результатов обеспечивается множественными сравнениями с результатами как эксперимента, так и других компьютерных моделей, в том числе “de facto” стандартной классической Молекулярной Динамики и классической гидродинамики. Результаты для сравнения были получены как из литературы, так и из работ автора.

Апробация работы. Основные результаты работы докладывались на многочисленных международных конференциях и форумах. Например:

- “CECAM 2024 - Perspectives and challenges of future HPC installations for atomistic and molecular simulations”, Berlin (Germany), 2024,
- “9th Annual CCPBioSim Conference”, Leeds (UK), 2023,
- “Грановские чтения”, Санкт Петербург, 2023,
- “WATOC 2020, 12th Triennial Congress of the World Association of Theoretical and Computational Chemists”, Vancouver (Canada), 2022,
- “Faraday Discussion: Challenges in biological cryo electron microscopy”, Sheffield (UK), 2022,
- “Современные проблемы биомедицины”, Санкт Петербург, 2021,
- “EMLG2019”, Kutna Hora (Czech Republic), 2019,
- “EMLG2018”, Nagoya (Japan), 2018,
- “WATOC 2017, 11th Triennial Congress of the World Association of Theoretical and Computational Chemists”, Munich (Germany), 2017.

Личный вклад. Автор лично разработал концепцию и теорию описанных новых подходов и методов. Автор лично проводил компьютерные расчеты на суперкомпьютерах, а также обрабатывал результаты расчетов на локальных рабочих станциях и личных компьютерах. Автор инициировал и принимал участие в обсуждении результатов расчетов с коллегами-экспериментаторами. Все научные положения, выводы и практические рекомендации сформулированы лично автором.

Публикации. Основные результаты по теме диссертации изложены в 59 печатных изданиях, 49 из которых изданы в журналах, рекомендованных ВАК при Минобрнауки России.

Объем и структура работы. Диссертация состоит из введения, 4 глав, заключения и 10 приложений. Полный объём диссертации составляет 321 страницу, включая 162 рисунка и 14 таблиц. Список литературы содержит 394 наименования.

Глава 1. Гибридное атомистически-гидродинамическое представление жидкостей и растворов

1.1 Введение

Классические методы молекулярной динамики развиты до такого уровня, что они позволяют воспроизводить не только макроскопические (термодинамические) и некоторые микроскопические (например, функции радиального распределения, автокорреляционные функции и др.) свойства простых жидкостей, для которых они изначально были предназначены, но и обеспечивают качественное, а иногда и количественное описание сложных биомолекулярных структур и их функциональность [1; 2]. Полученные атомистические детали воспроизводят экспериментально измеренные структурные и динамические свойства различных систем от небольших пептидов [1] до белков среднего размера [3], мембран клеток [4—6], вплоть до целых клеточных органелл и целых вирусов [7; 8]. Такие полноатомные, «ab initio» результаты позволяют исследовать системы на более крупных пространственных и временных масштабах, в то же время обеспечивая описание на экспериментально недоступных промежуточных масштабах между атомистическим и макроскопическим уровнями, приводящие к появлению новых видов молекулярных объектов (сложные структуры «молекулярного аппарата» клетки, ее сложные функциональные движения, коллективная динамика наборов молекул, и т. д).

Кроме того, часто необходимо учитывать несколько различных масштабов одновременно, в иерархии уровней дающих целостную картину молекулярной системы. Сложная система переходов масштабов с уровня на уровень при корректном описании дает новое глобальное понимание физических свойств всей системы отталкиваясь от элементарных нижних уровней взаимодействия атомов. Важность такого описания признается в научном сообществе и многомасштабные модели в последнее время разрабатываются очень активно [9; 10; 2]. Приложения охватывают широкий спектр систем в биологии, химии, материаловедении и других областях [11—15; 3].

Должны ли жидкости описываться как континуум или как совокупность атомов зависит от пространственного и временного масштабов рассмотрения. Одно описание сменяет другое когда масштабы постепенно уменьшаются от макроскопических (метры, секунды) до микроскопических (ангстремы, пикосекунды). Но существует и диапазон масштабов (нанометры, наносекунды), при которых оба описания верны. На этих масштабах гидродинамика приобретает существенные флуктуации, в то время как в атомистической динамике становятся важны коллективные явления с участием сотен атомов. Хотя этот факт не нов, лишь недавно стало возможно исследовать процессы на этих перекрывающихся масштабах, либо с использованием вычислительного эксперимента (высокопроизводительная молекулярная динамика, флуктуационная гидродинамика) либо экспериментально (вся область «нано-» наук).

Однако до сих пор существует концептуальный пробел в теоретическом описании этих явлений поскольку континуальная и атомистическая теории не переходят друг в друга. С одной

стороны, статистическая механика сходится к макроскопическим законам только в «макроскопическом» пределе, с другой, строгого математического перехода от гамильтоновой динамики атомов к гидродинамике, подчиняющейся уравнениям Навье–Стокса (НС) (см., например, [16]) до сих пор не существует [17]. То же самое верно и для движения по шкале масштабов вниз: гидродинамические флуктуации не становятся атомами в пределе.

Разработка многомасштабных методов для молекулярных систем чаще всего ассоциируется с идеей увеличения зернистости (coarse graining, CG) [4; 18—20]. В этом случае вводятся новые объекты большего масштаба, которые аппроксимируют группы атомов как единые частицы. Метод диссипативной динамики частиц - хорошо известный пример CG [21—24], широко применяемый для биологических объектов и реализованный в популярных программах, таких как GROMACS [25] (силовое поле MARTINI), и другие [5; 26—29]. Сложный вопрос корректного взаимодействия между шкалами исследуется, например, авторами [6] где изучается взаимосвязь между представлениями Молекулярной Динамики (MD) и CG, с использованием Марковских процессов. Еще один пример сопряжения «мелкозернистой» и «крупнозернистой» фазы даны в [26; 27], где связывание осуществляется плавно через межфазный параметр λ.

Одним из основных недостатков методов CG является их сильная зависимость от выбора межчастичных CG потенциалов. Основной целью в разработке методов CG является построение адекватного потенциала взаимодействия между отдельными частицами, рассматриваемые просто как «большие атомы» или «капли» (хотя и более сложные, чем обычные атомы). Несмотря на возможные связи со статистической механикой, например, между multi-coarse-grained методом и теорией жидкого состояния [30], CG процедура нетривиальна и сильно влияет на конечное описание физики процессов в системе.

Для «простых» жидкостей, таких как вода при нормальных условиях, процедуры CG хорошо зарекомендовали себя и могут успешно использоваться для многомасштабного моделирования в рамках подхода геометрической декомпозиции области, основанной на Лагранжевых межчастичных методах. Например, в [31—33] предлагается семейство методов адаптивного разрешения (AdResS, H-AdresS), где полноатомное моделирование проводилось в некоторой области, окружающий растворитель был представлен в упрощенном описании CG, а в «буферной» области между ними атомы постепенно уменьшали количество степеней свободы, (DOF) становясь компьютерными «каплями». В оригинальной работе Прапротника и др. [31], для подавления нефизичного повышения давления и плотности в гибридной буферной зоне использовался особый термостат. Эффект этого специального термостата позднее был заменен на так называемый компенсационный член свободной энергии в модели Espanol и др. [33], который сделал метод консервативным по энергии ценой потери свойства сохранения импульса.

Другой класс многомасштабных методов основан на представлении части системы как бесструктурного континуума. В MD сообществе они известны как модели «неявного растворителя» и используются для экономного моделирования воды и других растворители окружающих интересующую молекулу. Исторически первые попытки связать разные масштабы в молекулярных системах используя эту идею позволяли атомам покинуть или зайти в континуальную часть системы. Серьезной концептуальной проблемой здесь является существование границы между атомистической и континуальной (гидродинамической) частями. Достижение правильного балан-

са потоков массы и импульса через эту границу без внесения артефактов в чисто атомистическую часть моделирования, которая очень чувствительна к расположению раздела между атомистическим и гидродинамическим представлениями одной и той же жидкости, – очень нетривиальная задача [14; 4]. В так называемой схеме переменных состояния [9; 11; 15], связь между чисто атомистической и гидродинамической областями устанавливается методом «частица в ячейке». В таком методе лагранжева (MD-частицы) и эйлерова (континуум) части системы связаны через зону перекрытия конечного размера, которая обеспечивает сохранение объемных потоков массы и импульса. Использование такой зоны перекрытия обеспечивает более плавный переход между двумя представлениями по сравнению с методом сопряжения потоков через границу раздела. В схеме переменных состояния всегда используется параметр – «переключатель» интерполяции. Смысл этого параметра в гибридной «буферной» зоне между чисто атомистической и чисто континуальной частями обычно неясен.

В работах Маркштейна с соавторами [2] и Павлова с соавторами [3] был предложен другой подход к сопряжению с использованием переменных состояния между молекулярно-динамическим и гидродинамическим представлениями одной и той же жидкости. По сравнению с другими подходами в литературе по многомасштабному моделированию, наш метод использует физическую аналогию для задания условий сопряжения в «буферной» зоне между атомистической и гидродинамической областями. Методы основанные на физической аналогии для связывания моделей различного разрешения использовались в гидродинамике в течение нескольких десятилетий. Классическим примером является аналогия с акустикой предложенная Лайтхиллом [34], которая была введена в гидродинамику чтобы соединить масштабы охватывающие 3-4 порядка между звуковыми волнами в диапазоне слышимых частот и турбулентных структур течения генерирующие звук. Начиная с оригинальной работы Лайтхилла [34], различные гибридные методы такого рода были разработаны с общей идеей переписать основные уравнения Навье-Стокса в форме неоднородных линейных уравнения распространения звука («крупнозернистая» модель) с нелинейным источником («мелкомасштабная» модель). В наиболее продвинутых подходах этого типа (например, [35—37]), нелинейный источник напрямую связан со свойствами мелкомасштабного решения (пространственный и временной масштабы турбулентности). По аналогии для многомасштабного моделирования жидкостей в атомистическом и гидродинамическом масштабах в работе Маркштейн и др. [2] рассматривалась классическая модель фильтрации Бакли-Лаверетта [38] по аналогии с двухфазным потоком, реализованная для 2D-жидкости, аргона при высоком давлении. В работе Щукина с соавторами [5] аналогия с двухфазным потоком была распространена на двухмерную модель воды, где использовалась модель «Mercedes-Benz» [39; 6] для MD части решения. Идея гибридного метода заключается в рассмотрении двух представлений одной и той же жидкости как частиц (атомистическая) и как эйлеровых контрольных объемов (континуум) одновременно. Частичная и континуальная части решения рассматриваются как «фазы» одной и той же жидкости в соответствии с законами сохранения. Связь контролировалась определяемой пользователем функцией, зависящей от координат и времени $s\,(x,t)$, описывающей влияние представлений друг на друга и имеющей смысл частичной концентрации «фаз» по аналогии с двухфазным потоком. По сравнению с детерминистическими уравнениями Навье-Стокса исходной модели Бакли-Лаверетта, здесь уравнения флуктуационной гидродинамики Ландау-

Лифшица (LL-FH) [40; 41] представляют континуальную часть решения в многомасштабной модели, основанной на аналогии с двухфазным потоком.

Уравнения LL-FH допускают правильное статистическое описание коллективных свойств жидкостей, включая тепловые флуктуации. Будучи стохастическими дифференциальными уравнениями в частных производных, уравнения LL-FH сложнее интегрировать численно по сравнению с детерминистическими уравнениями Навье-Стокса. Примечательно, однако, что уравнения LL-FH все еще поддаются решению с помощью метода конечных разностей [42—44], конечных объемов [14; 45; 46] или метода решеток Больцмана [47; 48].

Для корректного перехода между MD и гидродинамическими описаниями требуется соответствие статистики флуктуаций в обоих представлениях. Эти вычисления более требовательны к ресурсам, так как некоторые характеристики жидкости сходятся особенно медленно (например, изотермическая сжимаемость, рассчитанная с использованием производной давления по объему). Это особенно верно на этапе разработки, так как уравнения должны быть решены одновременно на всех уровнях для проверки решения. Чтобы преодолеть вычислительные трудности возможны два пути. Первый путь использует недорогие игровые платформы, второй – суперкомпьютеры. В нашей работе использовались оба пути при разработке метода решения.

Широкое распространение компьютерных игр приводит к быстрой эволюции графических процессоров, разработка которых идет быстрее, чем разработка обычных процессоров. Основной сложностью в приложении к молекулярному моделированию является потребность в кодировании программного обеспечения, которое требует специализированных навыков. Тем не менее тенденция такова, что, возможно, графические процессоры скоро станут одной из самых привлекательных альтернатив для дешевых и относительно высокопроизводительных вычислений. Наши результаты показали высокую эффективность вычислений на GPU при решении уравнений LL-FH.

Разработка специализированных петафлопных платформ сложнее, но у нее есть потенциал для фундаментальных прорывов в нишевом моделировании, таких как молекулярное моделирование, поскольку они могут достигать временных масштабов, невозможных на любом другом оборудовании. Современные тенденции в пета- (и экза-)флопных высокопроизводительных вычислениях можно разделить на два класса, Рисунок 1.1. В первом используются процессоры общего назначения и достигается высокая производительность за счет использования очень большого количества вычислительных узлов. Примерами такой машины являются K-computer и Fugaku, созданные в RIKEN. Второй класс основан на специализированных ускорителях, в которых молекулярные вычисления, занимающие наибольшее время, выполняются на аппаратном уровне, например, компьютеры MDGRAPE или ANTON, Рисунок 1.1. В нашей работе были использованы оба типа суперкомпьютеров при разработке нашего гибридного подхода.

Разрешение микроскопических деталей в макроскопическом потоке важно для множества практических применений. Одним из таких приложений является быстрая смена химической среда вокруг белковой молекулы, которая заставляет ее сворачиваться. Так, в одном из примеров использовался микромиксер для диффузии белка в гидродинамически сфокусированном набегающем потоке для создания условий в которых очень быстро изменяется концентрация денатуранта [49]. Типичный смеситель имеет тройной поток (два канала входа и один выход). Похожая идея

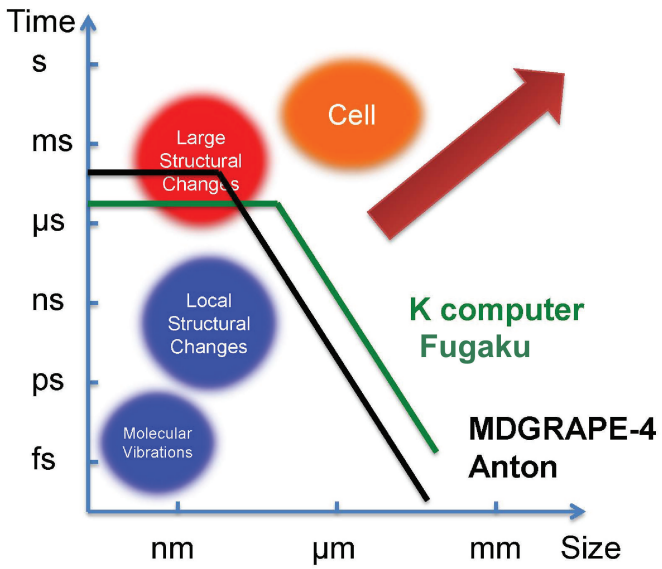


Рисунок 1.1 — Временные и пространственные масштабы биомолекулярных процессов и пределы возможностей MD моделирования.

была применена в [50] с микроканалом, состоящим из двух входов и одного выхода, где изготовление мембраны было достигнуто за счет контролируемого сдвига напряжения и межфазной реакция сшивки между двумя потоками жидкости. В другом примере система с двумя каналами впуска и одним выпуска [51] оптимизировала геометрию потока для контроля времени смешивания денатурирующего агента с целью контроля сворачивания пептидов. Большинство описанных в литературе моделей для этого типа задач ограничиваются довольно простым моделированием, основанным на уравнении Стокса с учетом пассивного поля концентрации. Это не включает микроскопические флуктуации, несмотря на то, что для этих процессов (например, сворачивании белков) молекулярные детали имеют решающее значение. Для чего можно использовать моделирование с помощью Молекулярной Динамики. Метод очень эффективен для моделирования наножидкостей и сложных течений (см., например, [52]), но остается слишком затратным даже для суперкомпьютеров. Так, характерный шаг по времени в MD составляет около фемтосекунды, что на 9 порядков меньше по сравнению с характерным временем смешивания в лучших конструкциях микрожидкостных смесителей (см., например, [51; 53]). Это не только не позволяет промоделировать реалистичное устройство с помощью чистой MD, но также не имеет большого смысла, так как атомарное разрешение требуется в очень ограниченных областях пространства.

Существует направление исследований, где изучается эффект динамики сплошной среды. Например, когда представляет интерес коллективное движение молекул в чистой жидкости или биомолекулярном растворе как динамика белков в воде. Для первого, стекловидная динамика является предметом интенсивных исследований, где обнаружено, что большие домены, включающие сотни молекул, движутся как единое целое с очень разной динамикой по сравнению с молекулами между доменами [54]. Аналогично, когерентное поведение крупных конгломератов молекул воды образующих микроскопические «водяные вихри» на поверхности белков было исследовано в [55]. Это согласуется с гипотезой [56], что вся динамика белков «закодирована» в коллективной динамике воды. В последнее время эта гипотеза была косвенно подтверждена [1], где было показано, что флуктуации водного континуума вокруг молекулы пептида имеют реша-

ющее значение для динамики самого пептида. Один из возможных способов решения проблемы коллективных явлений в сложных атомистических системах заключается в использовании гибридного моделирования.

Одной из важнейших проблем при использовании многомасштабного моделирования является проверка правильности результатов. Чтобы выполнить проверку рассматриваемой системы должны быть использованы экспериментальные данные, или система должна быть смоделирована на самом мелком масштабе. В последнем случае для проверки правильности модели на гидродинамических масштабах моделирование должно приближаться к гидродинамической шкале времени, которое сначала должно быть промоделировано в атомистическом разрешении. Т.е. гидродинамическое представление системы может быть промоделировано и подтверждено по сравнению с атомистическими результатами «ab initio». Сложность заключается в том, что результаты «ab initio» практически никогда не доступны на временах необходимых для гидродинамики.

Для проверки необходимы эффективные модели жидкостей, которые менее требователен к вычислительным ресурсам, но в то же время достаточно подробны, чтобы описывать атомистическую природу жидкости. Для атомистического представления воды может быть использована так называемая двумерная модель «Mercedes Benz» (BN2D), разработанная Ben-Naim [57] в силу высокой вычислительной эффективности модели. Действительно, объем системы в 2D увеличивается как L^2 по сравнению с L^3 в 3D, где L — размер системы в каждом измерении. С другой стороны, объем системы масштабируется с количеством дискретных частиц в системе, N, которое, в свою очередь, является аргументом в законе масштабирования, типичном для Молекулярной Динамики, которая в лучшем случае масштабируются как $NLog(N)$. Кроме очевидных преимуществ уменьшенного размера в 2D, BN2D не включает в себя дальнодействующих кулоновских вычисления, требующих значительных вычислительных ресурсов. Использование таких методов как суммирование Эвальда [58; 59] добавляет большое количество операций и иногда приводит к артефактам.

Таким образом, двумерные модели воды вызывают интерес в подобных исследованиях [60—66]. Такие модели требуют меньших вычислительных затрат, но, в то же время, они обеспечивают качественное согласие с экспериментом [67]. Подробное изучение модели [60; 68—72] продемонстрировало, что модель имитирует реальные свойства воды. Компьютерная реализация 2D модели Mercedes Benz для MD была описана в [65; 66], где исследовался 2D белковый раствор. Модель воды MB также использовалась для изучения гидрофобных эффектов и денатурации при понижении температуры [73].

Наконец, следующая по сложности система – полноразмерный вирус. Современные экспериментальные методы достигли достаточно высокого разрешения, чтобы измерять атомистическую структуру целых вирусов [74—78]. Это дает уникальную возможность моделировать их с помощью классической MD и получать очень важную биомолекулярную информацию о деталях функционирования вируса, что, в свою очередь, может быть использовано для разработка методов манипулирования и уничтожения вирусов. Такая модель «ab initio» представляет собой крайнюю точку на другом конце спектра для проверки гибридных методов: это большая, сложная молекулярная система.

К сожалению, эксперимент обычно ограничен измерением только статических и симметричных частей биомолекулярной структуры, в то время как динамика вирусных структур важна для представления специфических взаимодействий между вирусными частицами, раствором и клеточной мембраны. Кроме того, экспериментальные данные получены при низких температуры и не отражают влияние физиологического раствора на биологический обьект. Вода, однако, критически важна для динамики белка, т.к. молекулы воды контролируют его движение [7].

Наш подход поможет понять логику некоторых вирусных процессов, скрытых в динамике, взаимодействии с раствором, происходящих за короткие промежутки времени. Этот взаимодополняющий экспериментальный и теоретический подход поможет сэкономить время и ресурсы экспериментальных исследований и предложить пути разработка эффективных методов лечения вирусных инфекций.

Для поверки нашей гибридной методологий был выбран капсид свиного цирковируса, тип 2 (PCV2), поскольку это – самый маленький из известных вирусов диаметром всего около 20 нм. [79]. Капсид PCV2 стабилен без генома и его атомистическая структура измерена с помощью рентгеновской кристаллографии [74]. Однако структура N-концов белка капсида не была измерена в эксперименте, скорее всего, из-за их подвижной или неупорядоченной структуры. Сочетание этих фактов делает этот вирусный капсид удобной моделью для полноатомного MD моделирования.

Были промоделированы два случая полного вирусного капсида в атомистической разрешении: с N-концами (назовем их «хвостами») и без них. Обе структуры стабильны в физиологических условиях и хорошо согласуются с экспериментальными данными. В нашей статье [7] было показано, что значение среднеквадратичного отклонения, RMSD, – мало для этих структур по сравнению с рентгеновской структурой.

1.2 Теория

В рамках нашей модели объединены два представления жидкости: 1) стандартное классическое представление MD, где речь идет о наборе точечных масс, взаимодействующих друг с другом посредством эмпирически заданного потенциала и 2) континуум флуктуационной гидродинамики Ландау-Лифшица (LL-FH), обобщение гидродинамики Навье-Стокса со стохастическими источниками, добавленными для статистической имитации тепловых флуктуаций на малых масштабах. В качестве модели рассматривается условная двухфазная жидкость представляющая одно и то же химическое вещество (Рисунок 1.2). «Фазы» смешаны на самых мелких масштабах так, что поверхностное натяжение не учитывается, и обе «фазы» одновременно занимают один и тот же контрольный объем. Парциальные концентрации «фазы» MD и «фазы» LL-FH равны s и $1 - s$ соответственно, где s — параметр модели $0 \leqslant s \leqslant 1$. В общем случае s — задаваемая пользователем функция пространства и времени, которая определяет сколько атомистической информации требуется в конкретной области моделирования, Рисунок 1.2.

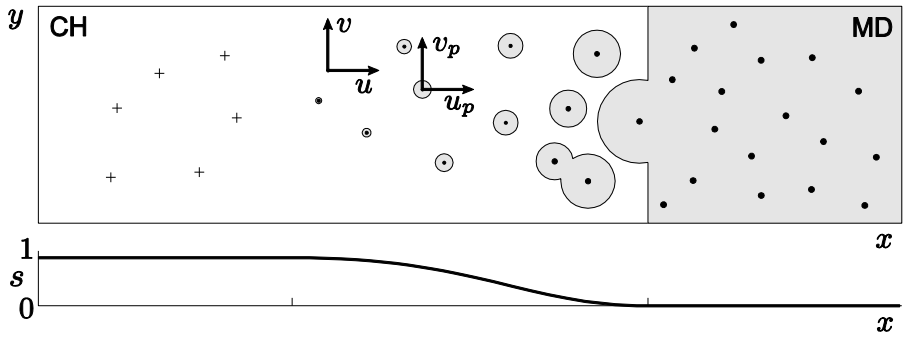


Рисунок 1.2 — Схематическое представление гибридной модели; форма функции s (внизу) определяет чисто MD частицы (справа), которые постепенно трансформируются в пассивные метки, следующие за гидродинамическим потоком (слева).

1.2.1 Модель флуктуационной гидродинамики для описания макроскопических потоков

Уравнения флуктуационной гидродинамики Ландау-Лифшица (LL-FH) представляют собой обобщение детерминированных уравнений Навье-Стокса (НС) для микроскопических потоков с учетом флуктуирующих стохастических источников в уравнениях импульса и энергии, как описано ниже.

$$\begin{cases} \frac{\partial \rho}{\partial t} + div\left(\rho \mathbf{u}\right) = 0 \\ \frac{\partial \rho u_i}{\partial t} + div\left(\rho u_i \mathbf{u}\right) = \nabla_j \left(\mathbf{\Pi}_{ij} + \tilde{\mathbf{\Pi}}_{ij}\right), \; i = 1{,}3 \\ \frac{\partial \rho E}{\partial t} + div\left(\rho E \mathbf{u}\right) = \nabla_j \left(\left(\mathbf{\Pi}_{ij} + \tilde{\mathbf{\Pi}}_{ij}\right) \cdot u_i\right) + div\left(\mathbf{q} + \tilde{\mathbf{q}}\right) \end{cases} \tag{1.1}$$

где $E = C_V T + \frac{1}{2}\mathbf{u}^2$ — плотность удельной энергиии, $C_V = \frac{R}{\gamma - 1}$ – теплоемкость при постоянном объеме, $R = \frac{k_B}{m}$ – газовая постоянная, члены в правой части состоят из усредненной величины и стохастического слагаемого, обозначаемого тильдой.

Усредненные члены в правой части уравнений (1.1) - стандартный тензор напряжения

$$\mathbf{\Pi}_{\alpha,\beta} = -\left(p - \zeta div \mathbf{u}\right)\delta_{\alpha,\beta} + \eta\left(\partial_\alpha u_\beta + \partial_\beta u_\alpha - 2D^{-1} div \mathbf{u} \delta_{\alpha,\beta}\right) \tag{1.2}$$

и поток тепла

$$q_\alpha = \kappa \cdot \partial_\alpha T, \tag{1.3}$$

где η и ζ — сдвиговая и объемная вязкости, D — размерность пространства, κ — коэффициент теплопроводности, и T – температура.

Что касается детерминированных уравнений Навье-Стокса, то уравнения для усредненных по времени величин замыкаются с помощью уравнения состояния, которое может быть получено, например, методом молекулярной динамики [80].

Флуктуирующая часть уравнений LL-FH соответствует континуальному описанию тепловых флуктуаций скорости, чья статистика подчиняется флуктуационно-диссипационной теореме

[81]. Стохастический тензор напряжений $\tilde{\mathbf{\Pi}}_{\alpha\beta}$вводится как случайная гауссова матрица с нулевым средним значением и ковариацией равной

$$\left\langle \tilde{\mathbf{\Pi}}_{\alpha\beta}\left(\mathbf{r}_1,t_1\right)\tilde{\mathbf{\Pi}}_{\gamma\delta}\left(\mathbf{r}_2,t_2\right)\right\rangle = 2k_BT\left[\eta\left(\delta_{\alpha\delta}\delta_{\beta\gamma}+\delta_{\alpha\gamma}\delta_{\beta\delta}\right)+\left(\zeta-\frac{2}{D}\eta\right)\delta_{\alpha\beta}\delta_{\delta\gamma}\right]\delta\left(t_1-t_2\right)\delta\left(\mathbf{r}_1-\mathbf{r}_2\right). \tag{1.4}$$

Используя эту ковариацию, можно явно выразить стохастический тензор напряжений как [82]:

$$\tilde{\mathbf{\Pi}}_{\alpha\beta} \cong \sqrt{\frac{2k_BT}{\delta t\delta V}}\left(\sqrt{2}\cdot\sqrt{\eta}\cdot\mathbf{G}^S_{\alpha\beta}+\sqrt{D}\cdot\sqrt{\zeta}\cdot\frac{tr[\mathbf{G}]}{D}\cdot\mathbf{E}_{\alpha\beta}\right), \tag{1.5}$$

где $\mathbf{G}_{\alpha\beta}$ — случайная гауссова матрица с нулевым средним значением и ковариацией $\langle\mathbf{G}_{\alpha\beta}\mathbf{G}_{\gamma\delta}\rangle = \delta_{\alpha\gamma}\delta_{\beta\delta}$, $\mathbf{G}^S_{\alpha\beta} = \frac{\mathbf{G}_{\alpha\beta}+\mathbf{G}^T_{\alpha\beta}}{2} - \frac{tr[\mathbf{G}]}{D}$, $\mathbf{E}_{\alpha\beta}$ — случайная симметричная матрица с нулевым следом, $\mathbf{E}_{\alpha\beta}$ — единичная матрица, и $tr\left[\mathbf{G}\right]$ — след матрицы $\mathbf{G}_{\alpha\beta}$. Стохастический тепловой поток также имеет нулевое среднее, а ковариационная составляющая определяется выражением

$$\langle\tilde{\mathbf{q}}_\alpha\left(\mathbf{r}_1,t_1\right)\tilde{\mathbf{q}}_\beta(\mathbf{r}_2,t_2)\rangle = 2k_B\kappa T^2\delta_{\alpha\beta}\delta\left(t_1-t_2\right)\delta\left(\mathbf{r}_1-\mathbf{r}_2\right), \tag{1.6}$$

или явно:

$$\tilde{\mathbf{q}}_\alpha \cong \sqrt{\frac{2k_B\kappa T^2}{\delta t\delta V}}\mathbf{G}_\alpha, \tag{1.7}$$

где $\mathbf{G}_\alpha$ — случайный Гауссов вектор с нулевым средним и ковариацией $\langle\mathbf{G}_\alpha\mathbf{G}_\beta\rangle = \delta_{\alpha\beta}$. Кроме того, стохастический тензор напряжений и стохастические тепловые потоки не коррелированы.

Модель LL-FH позволяет точно описывать статистические свойства атомистической системы с правильным стандартным отклонением массы, импульса и кинетической энергии для контрольных объемов размером от 5 до 10 атомов на сторону ячейки [46]. В пределе больших объемов модель LL-FH естественным образом переходит к обычным уравнениям Навье-Стокса.

Уравнения LL-FH могут быть эффективно решены в Эйлеровом представлении, соответствующие методы хорошо развиты в области вычислительной гидродинамики (CFD), например, в рамках схемы Time Asynchronous Relative Dimension In Space (TARDIS) [83]. Идея TARDIS заключается в обобщении современных мульти-шаговых алгоритмов на мульти-масштабных сетках, разработанных в CFD [84], на многомасштабное моделирование в гидродинамике, когда размер системы сам по себе является параметром и влияет на основные уравнения, как в случае уравнений LL-FH. TARDIS основан на одновременной трансформации уравнений гидродинамики в пространстве и времени, и последующем решении преобразованных уравнения на равномерной декартовой сетке с соответствующими условиями на границах сетки. Например, в одномерном случае, рассмотрим координату $x = x_0$, которая соответствует границе отделяющей большие физические масштабы в пространстве и времени $l_s(x) = L_s$ и $t_s(x) = T_s$ для $x \geqslant x_0 + \Delta x/2$ от малых $l_s(x) = l_s$ и $t_s(x) = t_s$ для $x < x_0 - \Delta/2$. Тогда правило преобразования в пространстве и времени есть

$$\bar{x} - x_0 = (l_s(x)/l_s)^{-1}(x - x_0); \bar{t} - t_0(t) = (t_s(x)/t_s)^{-1}t \tag{1.8}$$

и алгоритм TARDIS решает уравнения в преобразованных пространственно-временных переменных $\bar{x}$ и $\bar{t}$. С помощью масштабируемой функции увеличения $(l_s(x)/l_s)^{-1} = (t_s(x)/t_s)^{-1} \equiv f(x)$,

TARDIS может быть явно адаптирован к желаемому пространственно-временному отображению, лежащему в основе теоретической модели.

В качестве иллюстрации приведем применение TARDIS для разрешения тепловых флуктуации скорости потока в интересующей области микроканала на сетке 200x200 контрольных объемов. На рисунке 1.3 показан профиль скорости через два поперечных сечения по высоте канала. Эти графики показывают что флуктуации наиболее выражены в центре области, соответствующей малым масштабам (CS2). В крупномасштабной части области (CS1) флуктуации становятся почти неотличимыми от фонового профиля потока Пуазейля. На рисунке 1.4 показаны результаты применения TARDIS в моделировании тепловых флуктуаций плотности воды при термодинамическом равновесии, когда пространственный масштаб длины контрольного объема колеблется от 8 атомов в центре домена до 2190 атомов вблизи границы. Для этого расчета сетка TARDIS составляла 256x10 контрольных объемов. Рассматриваемая задача соответствует большому каноническому ансамблю объема V_c (размер контрольного объема) и температуры T, для которых верно следующее линейное соотношение между стандартным отклонением плотности флуктуации и обратной изотермической скорости звука c_T (1510 м/с для воды при нормальных условиях) [80]: $stdev(\rho') = c_T^{-1}\sqrt{\frac{\rho k_B T}{V_c}}$, где $\rho' = \sqrt{\langle(\rho - \langle\rho\rangle)^2\rangle}$.

Для всего рассматриваемого диапазона масштабов стандартное отклонение плотности вычисляемой из решения LL-FH совпадает с теоретическим соотношением в пределах погрешности в несколько процентов, что находится в пределах статистического шума, Рисунок 1.5.

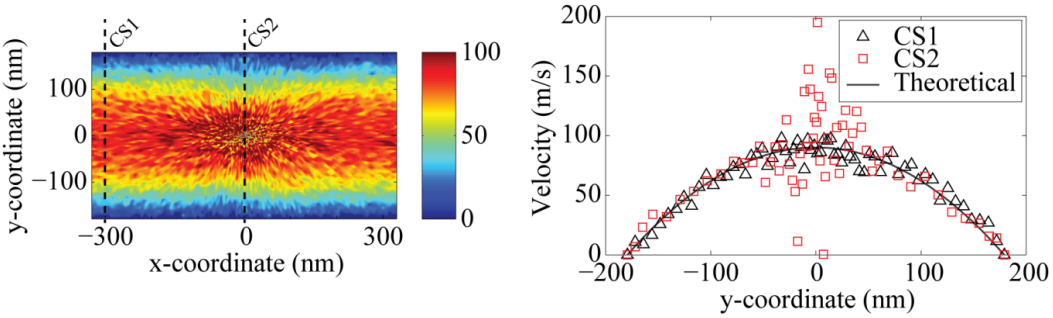


Рисунок 1.3 — Результаты моделирования с флуктуирующей гидродинамикой, сравнение с теорией; CS2 соответствует мелкомасштабной части домена, а CS1 соответствует крупным масштабам.

1.2.2 Сопряжение атомистического и гидродинамического представлений

Наиболее общие уравнения справедливые как для непрерывной так и для дискретной (частичной) частей гибридной модели являются интегральные законы сохранения, определяемые с помощью связи переменных состояния. Интегральные законы сохранения сформулированы так, что масса смеси, содержащей две «фазы» строго сохраняется, а также удовлетворяет второму закону Ньютона таким образом, что изменение полного импульса системы равно сумме всех сил. Альтернативный подход может состоять в том, чтобы сформулировать ту же модель в

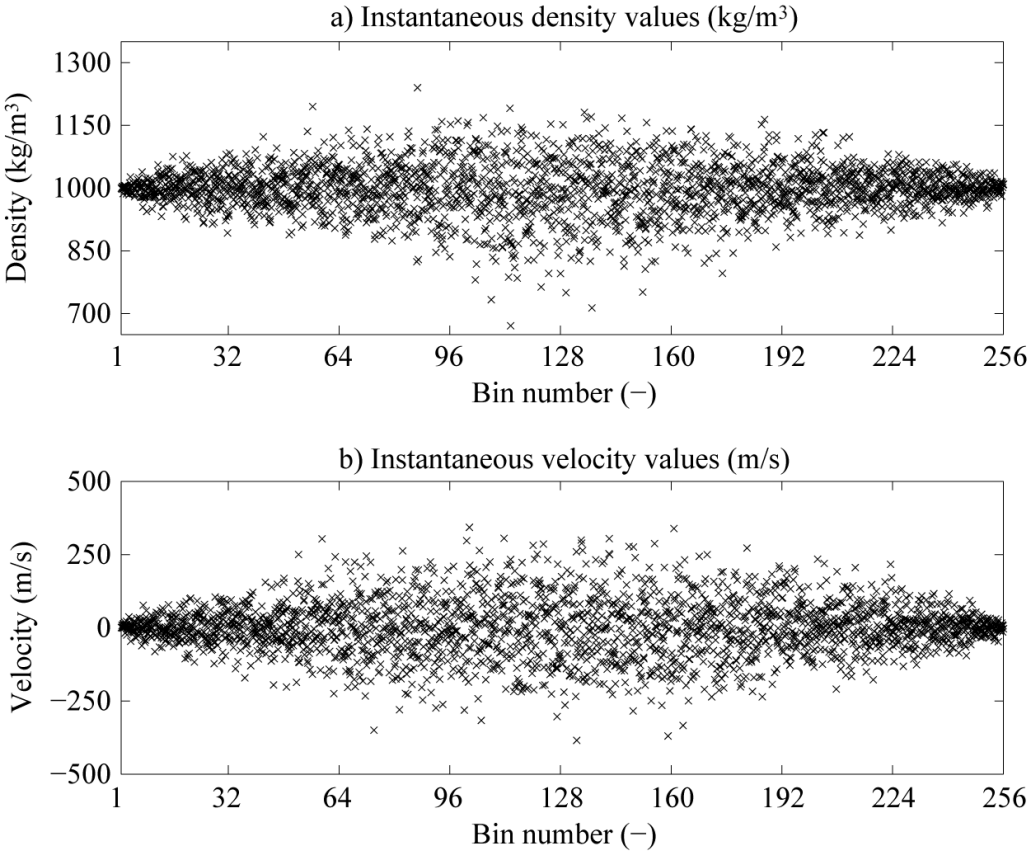


Рисунок 1.4 — Результаты моделирования тепловых флуктуаций в воде при равновесии для ряда контрольных объемов (ячейки 1 и 256 соответствуют наибольшим контрольным объемам, а ячейка 128 соответствует наименьшей): мгновенная плотность (а) и x составляющая скорости (б) флуктуаций.

рамках подхода обобщенных функций, где дискретные частицы представлены как функции Хевисайда (например, путем представления атомов как частиц конечного размера, соответствующего характерной ширине потенциальной ямы взаимодействия), их производные становятся дельта функциями Дирака и т. д. Вывод модели аналогичен выводу уравнений Навье-Стокса в дискретных переменных [85].

Следуя стандартному подходу двухфазного моделирования, рассмотрим смесь двух идеально смешивающихся жидкостей [86]. Одна фаза соответствует Лагранжевому представлению (MD частицы), а другая — Эйлеровому (континуум FH).

Связь между фазами вводится за счет обмена массой и импульсом между фазами. Параметр, который количественно определяет распределение массы и импульса между фазами, обозначается как s и математически представляет собой плавно изменяющуюся функцию пространства (и, возможно, времени), описывающую концентрацию каждой фазы. Значение s варьируется от 0 до 1, например, в центре системы может быть превалирующим описание MD ($s \approx 0$), а по краям – FH ($s \approx 1$), см. Рисунок 1.6 для 2D модели, Рисунок 1.7 для 3D.

Введем следующие определения. Фаза MD описывает частицы, свойства которых (плотность и скорость) рассчитываются в среднем по ячейке на регулярной сетке, и образует MD «жидкость» с плотностью $\sum_p \rho_p$ и импульсом $\sum_p \rho_p u_{ip}$, где $\rho_p = \frac{m_p}{V}$, u_{ip} — плотность MD ча-

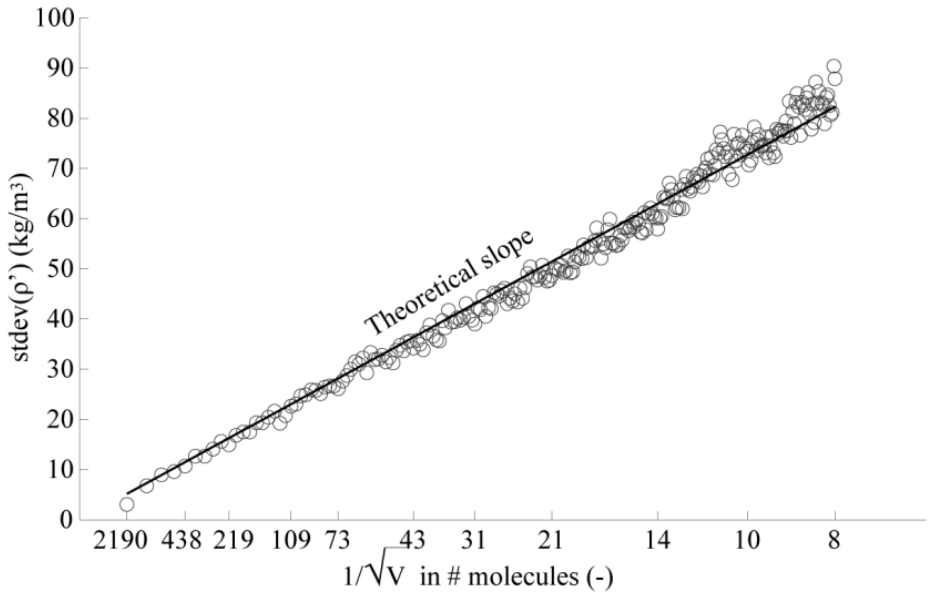


Рисунок 1.5 — Результаты моделирования тепловых флуктуаций в воде при равновесии для диапазона контрольных объемов: стандартное отклонение флуктуаций плотности в зависимости от размера системы.

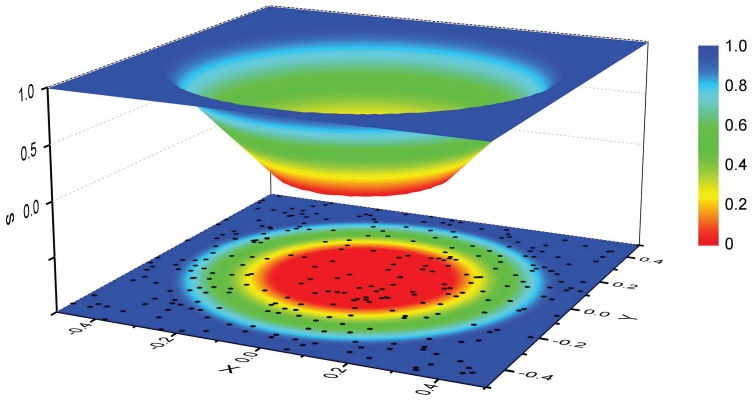


Рисунок 1.6 — В плоскости xy область 2D моделирования показана черными точками, представляющими частицы MD; фаза FH отображается в виде цветного поля; параметр связи s в этом примере имеет круговую симметрию, где в центре $s = 0$ а по краям $s = 1$, и его проекция на плоскость xy показана цветом.

стицы и скорость пространственной составляющей i соответственно, m_p — это масса частицы, а V — объем ячейки. Используется терминология трехмерного случая, поэтому «объем» рассматривается как площадь в двумерном случае. Эйлерова фаза (FH), соответствующая средним по ячейкам на регулярной сетке, формирует «жидкость» FH с плотностью ρ и импульсом ρu_i, по одному значению на ячейку.

Смесь этих двух жидкостей имеет плотность

$$\tilde{\rho} = s\rho + (1 - s) \sum_p \rho_p \tag{1.9}$$

и импульс

$$\tilde{u_j}\tilde{\rho} = su_j\rho + (1 - s) \sum_p \rho_p u_{jp}. \tag{1.10}$$

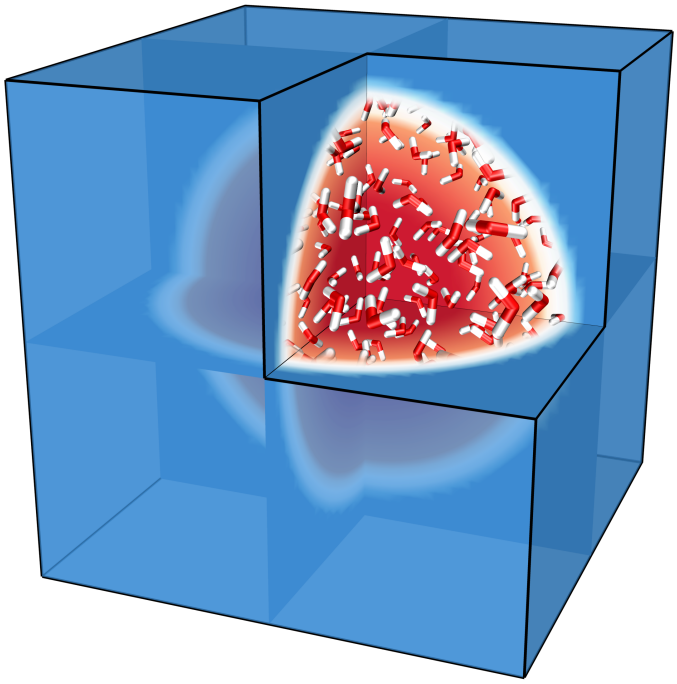


Рисунок 1.7 — Переменный параметр s и MD-сфера внутри вычислительной ячейки. Красная зона — это чистая область MD ($s = 0$), а синяя — область флуктуирующей гидродинамики ($s = 1$) для сферически симметричного случая s-функции; молекулы воды схематически показаны для примера (не в масштабе).

Суммирование $\sum_p$ выполняется по ячейке и включает $N(t)$ частиц находящихся в ячейке.

Потребуем сохранение массы и импульса смеси FH и MD «жидкостей» при всех значениях концентрации s. В частности [14], в случае жидкости в изотермических условиях, для которых коэффициент адиабатического тепловыделения приближается к единице, достаточно рассмотреть макроскопические законы сохранения массы и импульса, поскольку уравнение энергии отделяется от основных уравнений Навье-Стокса.

1.2.3 Динамика: законы сохранения массы и импульса

Сохранение массы

Законы сохранения массы каждой из фаз имеют вид:

$$\frac{\partial}{\partial t} s\rho + \frac{\partial}{\partial x_i} \tilde{u}_i s\rho = J^\rho, \tag{1.11}$$

$$\frac{\partial}{\partial t}(1-s)\sum_p \rho_p + \frac{\partial}{\partial x_i}(1-s)\sum_p u_{ip}\rho_p = -J^\rho, \tag{1.12}$$

где J^{ρ} — источник, обменивающий массу между фазами, и скорость транспорта в (1.11) – сохраняемая скорость $\tilde{u}_i$, то есть средняя скорость смеси.

Сохранение импульса

Фаза FH моделируется с помощью уравнений флуктуирующей гидродинамики Ландау-Лифшица [82] (LL-FH), учитывающие флуктуирующие стохастические источники, обусловленные микроскопическим движением молекул.

Уравнение сохранения импульса для «жидкости» FH можно записать в виде

$$\frac{\partial}{\partial t} s\rho u_j + \frac{\partial}{\partial x_i}\tilde{u}_i u_j s\rho = sF_j + J^u \tag{1.13}$$

а для MD «жидкости» в виде

$$\frac{\partial}{\partial t}(1-s)\sum_p \rho_p u_{jp} + \frac{\partial}{\partial x_i}(1-s)\sum_p u_{ip}u_{jp}\rho_p = \frac{(1-s)}{V_{cell}}\sum_p F_{jp} - J^u, \tag{1.14}$$

где J^u — скорость обмена импульсом,

$$F_j = \frac{\partial}{\partial x_i}\left(\Pi_{ij} + \tilde{\Pi}_{ij}\right), \tag{1.15}$$

F_j — гидродинамическая сила (на единицу объема), V_{cell} — объем ячейки, F_{jp} — межмолекулярная сила, действующая на каждую частицу в «жидкости» MD.

Как следует из [82], детерминированный тензор гидродинамических напряжений рассчитывается с использованием (1.2), а тензор стохастических напряжений рассчитывается с использованием (1.4).

1.2.4 Реализация законов сохранения массы и импульса

Сохранение массы

Для сохранения массы введем следующий динамический закон:

$$\frac{D}{Dt_0}\left(\tilde{\rho} - \sum_p \rho_p\right) = L^{\rho}\cdot\left(\tilde{\rho} - \sum_p \rho_p\right), \tag{1.16}$$

где $\frac{D}{Dt_0} = \frac{\partial}{\partial t}\cdot + \frac{\partial}{\partial x_i}\tilde{u}_i\cdot$ and L^{ρ} — оператор, подгоняющий среднюю плотность $\tilde{\rho}$ к правильному значению $\sum_p \rho_p$ внутри зоны, где $0 < s < 1$ и равен нулю на границах $s = 0$ и $s = 1$.

Для текущей реализации был использован оператор

$$L^{\rho}\cdot\left(\tilde{\rho}-\sum_{p}\rho_{p}\right)=\frac{\partial}{\partial x_{i}}\left[s\left(1-s\right)\alpha\frac{\partial}{\partial x_{i}}\left(\tilde{\rho}-\sum_{p}\rho_{p}\right)\right], \tag{1.17}$$

управляемый параметром $\alpha > 0$.

Уравнение сохранения массы в виде

$$\frac{\partial}{\partial t}\sum_{p}\rho_{p}+\frac{\partial}{\partial x_{i}}\sum_{p}\frac{dx_{ip}}{dt}\rho_{p}=0, \tag{1.18}$$

дает модифицированные скорости $\frac{dx_{ip}}{dt}$ MD частицы, которые включают источник J^{ρ}.

Таким образом, скорость отдельной частицы в «жидкости» MD становится равной

$$\frac{dx_{ip}}{dt}=u_{ip}+s(\tilde{u}_{i}-u_{ip})+s(1-s)\alpha(x)\frac{1}{\rho_{p}N(t)}\frac{\partial}{\partial x_{i}}\left(\tilde{\rho}-\sum_{p}^{N(t)}\rho_{p}\right), \tag{1.19}$$

где $N(t)$ — количество частиц в ячейке.

Для «жидкости» FH сделанные выше предположения дают следующее уравнение сохранения массы:

$$\frac{\partial}{\partial t}\tilde{\rho}+\frac{\partial}{\partial x_{i}}\tilde{u}_{i}\tilde{\rho}=\frac{\partial}{\partial t}\sum_{p}\rho_{p}+\frac{\partial}{\partial x_{i}}\tilde{u}_{i}\sum_{p}\rho_{p}+\frac{\partial}{\partial x_{i}}\left[s(1-s)\alpha\frac{\partial}{\partial x_{i}}\left(\tilde{\rho}-\sum_{p}\rho_{p}\right)\right]. \tag{1.20}$$

Подробности вывода см. в приложении Б.1.

Сохранение импульса

Процедура, аналогичная процедуре сохранения массы, была применена к уравнению сохранения импульса.

Используя

$$\frac{D}{Dt_{0}}\left(\tilde{u}_{j}\tilde{\rho}-\sum_{p}\rho_{p}u_{jp}\right)=L^{u}\cdot\left(\tilde{u_{j}}\tilde{\rho}-\sum_{p}\rho_{p}u_{jp}\right)+sF_{j}, \tag{1.21}$$

где оператор связи равен

$$L^{u}\cdot\left(\tilde{u_{j}}\tilde{\rho}-\sum_{p}\rho_{p}u_{jp}\right)=\frac{\partial}{\partial x_{j}}\left[s\left(1-s\right)\beta\frac{\partial}{\partial x_{j}}\left(\tilde{u_{j}}\tilde{\rho}-\sum_{p}\rho_{p}u_{jp}\right)\right], \tag{1.22}$$

и $\beta > 0$, уравнения сохранения импульса в виде

$$\frac{\partial}{\partial t}\sum_{p}\rho_{p}u_{jp}+\frac{\partial}{\partial x_{i}}\sum_{p}\frac{dx_{ip}}{dt}u_{jp}\rho_{p}-\sum_{p}\rho_{p}\frac{du_{jp}^{N}}{dt}=0, \tag{1.23}$$

приводят к модифицированным силам в MD «жидкости», $\frac{du_{jp}^{N}}{dt}$, которые включают скорость обмена импульсом J^{u}.

Таким образом, получаем новые силы для MD «жидкости»:

$$\frac{du_{jp}^N}{dt} = \frac{(1-s)}{\rho_p V_{cell}} F_{jp} + \frac{1}{\rho_p N(t)} \frac{\partial}{\partial x_i} \left[s(1-s)\alpha(x) \frac{\partial}{\partial x_i} \left(\tilde{\rho} - \sum_p \rho_p \right) \frac{\sum_p u_{jp}}{N(t)} \right] - \tag{1.24}$$
$$- \frac{1}{\rho_p N(t)} \frac{\partial}{\partial x_i} \left[s(1-s)\beta(x) \frac{\partial}{\partial x_i} \left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp} \right) \right].$$

Для жидкости FH используется следующее уравнение сохранения:

$$\frac{\partial}{\partial t}\tilde{\rho}\tilde{u}_j + \frac{\partial}{\partial x_i}\tilde{u}_i\tilde{\rho}\tilde{u}_j = \frac{\partial}{\partial t}\sum_p \rho_p u_{jp} + \frac{\partial}{\partial x_i}\tilde{u}_i \sum_p \rho_p u_{jp} + sF_j + \tag{1.25}$$
$$+ \frac{\partial}{\partial x_i} \left[s(1-s)\beta \frac{\partial}{\partial x_i} \left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp} \right) \right].$$

Подробности вывода см. в приложении Б.2.

1.2.5 Дифференциальная форма: основные уравнения модели двусторонней связи

В гибридном представлении жидкости, описанном выше предполагается, что как координаты и скорости атомов, так и поля плотности и импульса гидродинамической части являются независимыми переменными. Другими словами, дискретная и континуальная подсистемы влияют друг на друга динамически, что описывается приведенными выше уравнениями (в упрощенной, одномерной форме). В данном разделе выведены уравнения движения в трехмерном представлении. Также, они получены в дифференциальной, а не в интегральной форме, как в предыдущем разделе. Дифференциальная форма значительно удобней для численных расчетов, описывая, естественно, те же физические процессы что и интегральная.

Рассмотрим область решения V_0, которая разбита на элементарные эйлеровы кубические ячейки объема V. Каждая ячейка имеет 6 граней $\gamma = 1,..,6$ и заполнена континуальной жидкостью и, в то же самое время, MD-частицами, которые соответствуют дискретному представлению одного и того же химического вещества. Предполагается, что континуальная часть номинально двухфазной жидкости имеет такую же транспортную скорость как и смесь. В изотермическом состоянии эта номинально двухфазная жидкость кроме макроскопического уравнения состояния удовлетворяет также следующим макроскопическим законам сохранения. Для массы:

$$\delta_t(sm) + \sum_{\gamma=1,6} (s\rho\bar{\mathbf{u}})\, d\mathbf{n}^\gamma \delta t = \delta_t J^{(\rho)} \tag{1.26}$$

для фазы LL-FH и

$$\delta_t \left((1-s) \sum_{p=1,N(t)} m_p \right) + \sum_{\gamma=1,6} \left((1-s) \sum_{p=1,N_\gamma(t)} \rho_p \mathbf{u}_p \right) d\mathbf{n}^\gamma \delta t = -\delta_t J^{(\rho)} \tag{1.27}$$

для MD-фазы, где m и $\rho = m/V$ - масса и плотность континуальной «фазы» элементарного объема V, m_p — масса частицы, $\mathbf{u}_p$— скорость MD, $\bar{\mathbf{u}}$ — средняя скорость «смеси» $\bar{u}_i = \left[s\rho u_i + (1-s) \sum_{p=1,N(t)} \rho_p u_{ip} \right] / \bar{\rho}$, u_i — скорость континуума LL-FH «фазы», $\bar{\rho} = s\rho + (1-s) \sum_{p=1,N(t)} \rho_p$, $N(t)$ — это количество частиц в объеме V. $N_\gamma(t)$ — количество частиц, пересекающие грань ячейки γ^{th} с нормалью $d\mathbf{n}^\gamma$, $\rho_p = m_p/V$ – эффективная плотность MD частицы p, которая занимает объем V, а $\delta_t J^{(\rho)}$ – член описывающий источник массы, который отвечает за преобразование массы между «фазами», δ_t описывает изменение количества во времени δt, например, счетчик массы и импульса частиц в ячейке V накопленный с течением времени δt.

Для импульса это:

$$\delta_t(smu_i) + \sum_{\gamma=1,6} (s\rho u_i \bar{\mathbf{u}})\, d\mathbf{n}^\gamma \delta t = s \sum_{j=1,3} \sum_{\gamma=1,6} \left(\Pi_{ij} + \tilde{\Pi}_{ij} \right) dn_j^\gamma \delta t + \delta_t J_i^{(\mathbf{u})}, \tag{1.28}$$

$$\delta_t \left((1-s) \sum_{p=1,N(t)} m_p u_{ip} \right) + \sum_{\lambda=1,6} \left((1-s) \sum_{p=1,N_\gamma(t)} \rho_p u_{ip} \mathbf{u}_p \right) d\mathbf{n}^\gamma \delta t = (1-s) \sum_{p=1,N(t)} F_{ip} \delta t - \delta_t J_i^{(\mathbf{u})}, \tag{1.29}$$

где $\mathbf{\Pi}$ и $\tilde{\mathbf{\Pi}}$ — это детерминированная и стохастическая части тензора напряжений Рейнольдса в модели LL-FH, F_{ip} — сила MD, действующая на частицу p обусловленная парным потенциалом взаимодействий, а $\delta_t J_i^{(\mathbf{u})}$ — член LL-FH/MD, соответствующий компоненту импульса i.

Суммы потоков

$$\sum_{\gamma=1,6} \left((1-s) \sum_{p=1,N_\gamma(t)} \rho_p \mathbf{u}_p \right) d\mathbf{n}^\gamma \delta t$$

и

$$\sum_{\lambda=1,6} \left((1-s) \sum_{p=1,N_\gamma(t)} \rho_p u_{ip} \mathbf{u}_p \right) d\mathbf{n}^\gamma \delta t$$

являются соответствующими счетчиками массы и импульса частиц, пересекающих границы ячейки $\gamma = 1,..,6$.

Теоретически члены потока могут быть рассчитаны из распределения частиц в каждой точке границы ячейки. На практике для вычисления значений на границах ячейки можно использовать метод интерполяции используя распределения частиц в центрах смежных объемов V, например, методом конечных объемов.

Суммируя уравнения для масс, (1.26) и (1.27) и предполагая сохранение потоков, обращающихся в ноль на границах домена, по теореме Остроградского-Гаусса, закон сохранения массы смеси выполняется точно, $\bar{m}(t+\delta t) = \bar{m}(t), \quad \bar{m} = \bar{\rho} V$. Аналогичным образом, комбинируя уравнения импульса (1.28) и (1.29), видно, что второй закон Ньютона, приравнивающий изменение полного импульса $\bar{m} \cdot \bar{\mathbf{u}}$ приложенной силе, $\bar{F}_i = s \sum_{j=1,3} \sum_{\alpha=1,6} \left(\Pi_{ij} + \tilde{\Pi}_{ij} \right) dn_j^\alpha \delta t + (1-s) \sum_{p=1,N(t)} F_{ip}$, выполняется . Последнее выражение для силы, приложенной в гибридной системе, похоже на интерполяцию, используемую в методе AdResS [31] для взаимодействия частиц.

В уравнениях (1.26–1.29), $\partial_t J^{(\rho)}$ и $\partial_t J_i^{(\mathbf{u})}$ — определяемые пользователем функции, которые необходимо указать для самосогласования модели. Эти функции можно получить, задав как быстро средние значения смеси $\bar{\rho}$ и $\bar{u}_i\bar{\rho}$ приближаются к усредненными параметрами ячейки из MD «фазы» , $\sum\limits_{p=1,N(t)} m_p$ и $\sum\limits_{p=1,N(t)} u_{ip}m_p$:

$$
\begin{aligned}
&D_t\left(\bar{m} - \sum_{p=1,N(t)} m_p\right) = L^{(\rho)} \cdot \left(\bar{m} - \sum_{p=1,N(t)} m_p\right), \\
&D_t\left(\bar{u}_i\bar{m} - \sum_{p=1,N(t)} u_{ip}m_p\right) = L^{(u)} \cdot \left(\bar{u}_i\bar{m} - \sum_{p=1,N(t)} u_{ip}m_p\right) + s\sum_{j=1,3}\sum_{\gamma=1,6}\left(\Pi_{ij} + \tilde{\Pi}_{ij}\right) dn_j^{\gamma}\delta t,
\end{aligned}
\tag{1.30}
$$

где

$$
\begin{aligned}
&D_t\left(\bar{m} - \sum_{p=1,N(t)} m_p\right) = \delta_t\left(\bar{m} - \sum_{p=1,N(t)} m_p\right) + \sum_{\gamma=1,6}\left(\bar{\rho} - \sum_{p=1,N_\gamma(t)} \rho_p\right)\mathbf{u}d\mathbf{n}^{\gamma}\delta t, \\
&D_t\left(\bar{u}_i\bar{m} - \sum_{p=1,N(t)} u_{ip}m_p\right) = \delta_t\left(\bar{u}_i\bar{m} - \sum_{p=1,N(t)} u_{ip}m_p\right) + \sum_{\gamma=1,6}\left(\bar{u}_i\bar{\rho} - \sum_{p=1,N_\gamma(t)} u_{ip}\rho_p\right)\mathbf{u}d\mathbf{n}^{\gamma}\delta t
\end{aligned}
\tag{1.31}
$$

являются интегральными аналогами полных консервативных производных в случае гладких переменных поля по теореме Остроградского-Гаусса, и, используя ту же теорему, операторы в правой части

$$
\begin{aligned}
L^{(\rho)}.\quad &\left(\bar{m} - \sum_{p=1,N(t)} m_p\right) = \\
&\sum_{k=1,3}\sum_{\gamma=1,6}\left(s(1-s)\cdot\alpha\cdot\tfrac{1}{V}\left(\sum_{\lambda=1,6}\left(\bar{\rho} - \sum_{p=1,N_\lambda(t)}\rho_p\right)dn_k^{\lambda}\right)\right)dn_k^{\gamma}\delta t, \\
L^{(u)}.\quad &\left(\bar{u}_i\bar{m} - \sum_{p=1,N(t)} u_{ip}m_p\right) = \\
&\sum_{k=1,3}\sum_{\gamma=1,6}\left(s(1-s)\cdot\beta\cdot\tfrac{1}{V}\left(\sum_{\lambda=1,6}\left(\bar{u}_i\bar{\rho} - \sum_{p=1,N_\lambda(t)} u_{ip}\rho_p\right)dn_k^{\lambda}\right)\right)dn_k^{\gamma}\delta t,
\end{aligned}
\tag{1.32}
$$

являются интегральными аналогами соответствующей производной диффузии второго порядка. $\alpha,\beta > 0$ — числовые настраиваемые параметры, которые характеризуют «скорость диффузии» разницы между двумя «фазами». Последние параметры характеризуют насколько быстро две «фазы» уравновешиваются к одному и тому же макроскопическому состоянию, т. е. сходятся к одной и той же жидкости, которую они описывают. Характерное время релаксации, связанное с эти параметрами

$$
\tau^{diff} \sim \Delta x^2/\alpha \sim \Delta x^2/\beta, \tag{1.33}
$$

где $\Delta x = \sqrt[3]{V}$— масштаб длины, связанный с ячейкой объема V, должен быть сопоставим с временным шагом для частиц $\tau^{diff} \sim \tau^{MD}$ так, чтобы процесс релаксации повлиял на траекторию частицы в характерном для них масштабе времени Молекулярной Динамики.

Вывод приведенных выше уравнений можно найти в [2; 8], однако подчеркнем следующее.

1. Средние по ячейке значения полей $\bar{\rho}$, $\bar{\mathbf{u}}$, $\sum\limits_{q=1,N(t)} \rho_q$ и $\sum\limits_{p=1,N(t)} \rho_q\mathbf{u}_q$ в правой части уравнений выше, соответствуют центрам объемов ячеек и должны быть аппроксимированы

для каждого положения x MD частицы p, например используя билинейную интерполяцию в пространстве для положения частицы между центрами ячейки и интерполяции во времени в случае когда гидродинамическое решение опережает во времени с большим временным шагом по сравнению с Молекулярной Динамикой.

2. Суммы в правой части соответствуют средним градиентам поля (первые производные и вторые производные в $\frac{d\mathbf{x}_p}{dt}$ and $\frac{du_{ip}}{dt}$ соответственно), которые могут быть получены, например, используя теорему Остроградского-Гаусса для исключения одной производной, что может быть реализовано методом конечных объемов.
3. Для области чистой гидродинамики ($s = 1$) в соответствии с уравнениями выше, частицы становятся пассивными маркерами, переносимыми полем гидродинамики в соответствии с внешними граничными условиями налагаемыми гидродинамикой (периодические границы в данном случае), в этом случае нет необходимости вычислять потенциалы молекулярной динамики в этой области, что позволяет избежать самой затратной части MD моделирования.

Особое внимание следует уделить термостатированию системы. В широко используемых термостатах масштабирующих скорости, в стандартной реализации мгновенная температура вычисляется по скоростям определенной группы атомов (обычно всей системы). Затем скорости этих атомов равномерно перемасштабируются согласно соотношению между мгновенной и нужной температурами. Однако в нашем случае гибридных частиц система делится на существенно разные регионы с точки зрения термостатирования. Температура чисто гидродинамической области определяется только полем FH, таким образом, атомы находящиеся в данной области в данный момент времени не должны учитываться при расчете мгновенной температуры для термостата, т.е. они не должны быть термостатированы. Наоборот, атомы в гибридной зоне контролируются полем FH лишь частично, поэтому они должны иметь измененные скорости. Поэтому в нашем подходе осуществляется раздельное термостатирование для чистой области MD и двух половин гибридной области ($s \leqslant 0.5$ и $s > 0.5$, Рисунок 1.2); область чистой гидродинамики остается не термостатированной. Для каждой зоны мгновенная температура и масштабирование скорости рассчитывались индивидуально.

Чтобы самосогласовать модель, уравнения (1.26–1.29) сочетаются со следующими уравнениями для массы и ускорения частиц в каждой эйлеровой ячейке

$$\begin{aligned}
&\delta_t \sum_{p=1,N(t)} m_p + \sum_{\gamma=1,6} \left(\sum_{p=1,N_\gamma(t)} \frac{d\mathbf{x}_p}{dt} \rho_p \right) d\mathbf{n}^\gamma \cdot \delta t = 0, \\
&\delta_t \sum_{p=1,N(t)} m_p u_{ip} + \sum_{\gamma=1,6} \left(\sum_{p=1,N_\gamma(t)} \frac{d\mathbf{x}_p}{dt} \rho_p u_{ip} \right) d\mathbf{n}^\gamma \cdot \delta t = \sum_{p=1,N(t)} m_p a_{ip} \delta t, \qquad a_{ip} = \frac{du_{ip}}{dt},
\end{aligned} \tag{1.34}$$

которые определяют условия источника и модификацию MD уравнения для скорости и ускорения частиц, $\frac{d\mathbf{x}_p}{dt} = \mathbf{u}_p, \frac{d\mathbf{u}_p}{dt} = \mathbf{F}_p$ как:

$$
\begin{aligned}
&\frac{d\mathbf{x}_p}{dt} = \mathbf{u}_p + s(\bar{\mathbf{u}} - \mathbf{u}_p) + s(1-s)\cdot\alpha\cdot\frac{\sum\limits_{\gamma=1,6}\left(\bar{\rho} - \sum\limits_{q=1,N_\gamma(t)}\rho_q\right)d\mathbf{n}^\gamma}{\sum\limits_{q=1,N(t)} m_q},\\
&\frac{du_{ip}}{dt} = (1-s)F_{ip}/m_{ip} +\\
&+\sum_{k=1,3}\sum_{\gamma=1,6}\left(s(1-s)\cdot\alpha\cdot\sum_{q=1,N_\gamma(t)}\rho_q u_{iq}\cdot\left(\frac{\sum\limits_{\lambda=1,6}\left(\bar{\rho} - \sum\limits_{q=1,N_\lambda(t)}\rho_q\right)dn_k^\lambda}{\sum\limits_{q=1,N(t)} m_q}\right)\right)dn_k^\gamma / \sum_{q=1,N(t)} m_q\\
&+\sum_{k=1,3}\sum_{\gamma=1,6}\left(s(1-s)\cdot\beta\cdot\frac{1}{V}\left(\sum_{\lambda=1,6}\left(\bar{\rho}\cdot\bar{u}_i - \sum_{q=1,N_\lambda(t)}\rho_q u_{iq}\right)dn_k^\lambda\right)\right)dn_k^\gamma / \sum_{q=1,N(t)} m_q, \qquad i = 1,3,
\end{aligned}
\tag{1.35}
$$

где макроскопические поля $\bar{\rho}$, $\bar{\mathbf{u}}$, $\sum\limits_{q=1,N(t)}\rho_q$ и $\sum\limits_{p=1,N(t)}\rho_q\mathbf{u}_q$ соответствуют средним значениям по ячейке в каждом положении x MD частицы p. Для практических расчетов значения этих полей могут быть определены интерполяцией, так же как потоки через грани ячеек в уравнениях (1.26–1.29). Подробности вывода уравнений приведены в приложении B.

Модифицированные уравнения MD (1.35) зависят только от сохраняющихся переменных для смеси и значений MD, усредненных по ячейкам. Для численной реализации удобно решать (1.35) вместе с уравнением сохранения для плотности и импульса смеси, (1.28–1.29), а не исходные уравнения (1.26–1.27), которые вырождаются в пределе $s = 0$ или $s = 1$. Из-за включённых стохастических членов для напряжений, уравнения (1.28–1.29) являются стохастическими уравнениями в частных производных, подобным уравнениям флуктуационной гидродинамики Ландау-Лифшица (LL-FH), которые есть их предельный случай. Действительно, в случае, когда «фаза» континуума – единственная часть гибридной модели (т.е. когда $s = 1$ и когда нет MD частиц), уравнения (1.28–1.29) для плотности смеси и импульса сводятся к классическим уравнениям флуктуационной гидродинамики Ландау-Лифшица.

Помимо сохранения массы и импульса «смеси», модель (1.28–1.29) также непосредственно удовлетворяет флуктуационно-диссипациионной теореме (FDT) в предельных состояниях при $s = 0$ и $s = 1$, то есть для чистого MD и чистых уравнений LL-FH. В гибридной области MD/LL-FH, если предположить, что две части решения полностью релаксированы до одного и того же макроскопического состояния, члены диффузии, пропорциональные расхождению между MD и LL-FH частями решения, обращаются в ноль и члены связи гибридной модели (1.35) просто становятся линейной комбинацией LL-FH и усредненных по ячейке скоростей и сил MD, следовательно, будучи линейными, удовлетворяют FDT. На практике предположение о полной релаксации двух «фаз» к одному и тому же макроскопическому состоянию нуждаются в апостериорном подтверждении, что и показано ниже.

1.2.6 Упрощенная модель односторонней связи

Как частный, более простой случай, рассмотрим только макроскопически стационарные жидкости в отсутствие каких-либо гидродинамических градиентов, вдали от жестких границ. При таких предположениях тепловые флуктуации являются единственным источником макроскопических флуктуаций в жидкостях, описываемых моделью LL-FH.

Поэтому будем считать, что уравнения «двухфазной смеси» (1.28) и (1.29) полностью отвязаны от «фазы» MD и значения сохраняющихся переменных, $\bar{\rho}$ и $\bar{u}_i$, которые влияют на уравнения MD (1.35), можно получить из независимо рассчитанной гидродинамики. Для простоты рассмотрим случай жидкостей, для которых коэффициент адиабатического тепловыделения приближается к единице, тогда первые два уравнения для массы и импульса в уравнениях LL-FH (1.1) можно решить отдельно от уравнения для энергии [80]. Уравнения LL-FH, которые необходимо решить в этом случае, равны

$$\begin{array}{l} \frac{\partial \bar{\rho}}{\partial t} + div\left(\bar{\rho}\cdot\bar{\mathbf{u}}\right) = 0, \\ \frac{\partial(\bar{\rho}\cdot\bar{u}_i)}{\partial t} + div\left(\bar{\rho}\cdot\bar{u}_i\cdot\bar{\mathbf{u}}\right) = \sum\limits_{j=1,3}\nabla_j\left(\Pi_{ij}+\tilde{\Pi}_{ij}\right),\ i=1,2,3 \end{array} \tag{1.36}$$

где уравнение состояния (EOS), $\bar{p} = \bar{p}(\bar{\rho})$ и коэффициенты сдвиговой и объемной вязкостей, η и ζ, входящих в тензор напряжений Рейнольдса $\mathbf{\Pi}$ и его флуктуирующая составляющая $\tilde{\mathbf{\Pi}}$,

$$\begin{array}{l} \Pi_{i,j} = -\left(\bar{p}-\zeta div\,\bar{\mathbf{u}}\right)\delta_{i,j} + \eta\left(\partial_i\bar{u}_j + \partial_i\bar{u}_j - 2D^{-1}div\,\bar{\mathbf{u}}\delta_{i,j}\right), \\ \tilde{\Pi}_{i,j} = \zeta div\,\tilde{\mathbf{u}}\delta_{i,j} + \eta\left(\partial_i\tilde{u}_j + \partial_i\tilde{u}_j - 2D^{-1}div\,\tilde{\mathbf{u}}\delta_{i,j}\right), \qquad i,j=1,2,3 \end{array} \tag{1.37}$$

должны быть определены в соответствии с моделью MD. Тензор стохастических напряжений $\tilde{\mathbf{\Pi}}$ описывается как в (1.4).

Для нашей реализации односторонней связи гибридной многомасштабной модели, уравнения LL-FH (1.1–1.2) решаются совместно с MD уравнениями (1.35). Частицы MD присутствуют в растворе во всей области, включающей зону с преобладанием гидродинамики, где заданы периодические граничные условия. При большом размере зоны LL-FH наша реализация все еще может быть эффективной по сравнению с полноатомным моделированием, поскольку затраты на моделирование LL-FH по сравнению с MD – незначительны. Например, в MD вычисление потенциалов взаимодействия масштабируется как $N\log N$, где N — количество MD-частиц. Снижение вычислительных затрат для гибридной модели по сравнению с моделированием всех атомов может доходить до $\log N$. Еще большая экономия вычислительных ресурсов может достигается введением пространственно-переменных масштабов в моделирование с ростом от мелких атомистических до крупных гидродинамических масштабов где частицы MD теряют свою подвижность из-за малых тепловых флуктуации в больших объемах ячеек и, таким образом, могут быть заключены в небольшой области решения.

Чтобы самосогласовать модель, необходимо определить функцию «частичной концентрации» $s = s(x,y,z)$. Рассматриваются два типа функции s: (i) постоянное поле во всей системе, как в [2; 3] для 2D-моделирования и (ii) функция масштабирования, позволяющая изменять разрешение модели на определенной пользователем геометрической основе $s(x,y,z)$, которую можно

рассматривать как 3D вариант круглой зоны, рассмотренный в [5]. Для последних были использованы два типа переменной s-функции. В первом s изменяется в направлении только x: $s = s(x)$, а в другом имеет сферическую симметрию (Рисунок 1.7):

$$s\,(x,y,z) = \begin{cases} S_{\min}, & r \leqslant R_{MD}; \\ \frac{r-R_{MD}}{R_{FH}-R_{MD}}\,(S_{\max}-S_{\min})+S_{\min}, & R_{MD} < r < R_{FH}; \\ S_{\max}, & r \geqslant R_{FH}. \end{cases}, \tag{1.38}$$

где $r = (x - L/2)^2 + (y - L/2)^2 + (z - L/2)^2$, L - размер ячейки моделирования, $x,y,z \in [0,L]$, $S_{\min} = 0$, $S_{\max} = 1$.

На рисунке 1.7, красная область с молекулами воды — чисто атомистический домен, который постепенно меняется на белый (гибридная атомистически-континуальная область) до синего (область чисто флуктуационной гидродинамики).

1.3 Модели жидкости

1.3.1 Модель «Mercedes Benz»

Двумерные модели в молекулярно-динамическом моделировании имеют преимущество из-за квадратичного масштабирования числа частиц с ростом размера системы. Этот может иметь решающее значение для моделирования в нано- и микромасштабах, когда система приближается к гидродинамическим размерам. При моделировании воды важно получить аномальные свойства, такие как минимум давления, отрицательный коэффициент теплового расширения при низких температурах, высокая теплоемкость и минимум изотермической сжимаемости. Эти свойства определяются водородными связями, важной особенностью воды. Трехмерные модели воды дают хорошее количественное согласие с экспериментом, однако качественное согласие может быть быть достигнуто и для двухмерных (2D) моделей. Двумерная модель воды, имитирующая свойства реальной воды, называется «Mercedes Benz» (BN2D) [57]. Эта модель изучалась с использованием метода Монте-Карло, который не позволяет получить динамические свойства.

На удивление, модель «Mercedes Benz» (MB) не была реализована в MD. Существует исследование [87], в котором изучается 3D-версия модели, однако в нем получено мало физических свойств модели.

В нашей работе были получены подробности реализации MD воды MB, выведены формулы термодинамических свойств модели, и проанализированы численные а также физические свойства модели.

Модель MB или BN2D [88] представляет собой двухмерную модель воды, где молекулы попарно взаимодействуют через потенциал Леннард-Джонса и потенциал, зависящий от ориентации молекулы (зависящий в свою очередь от соответствующей ориентации «плеч» молекул),

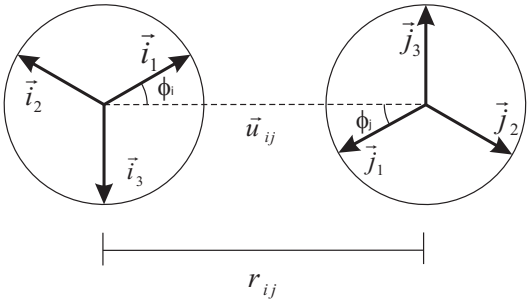


Рисунок 1.8 — Модель воды «Mercedes Benz» [57].

Рисунок 1.8. Такого взаимодействия оказывается достаточно для воспроизведения термодинамических свойств реальной воды. Общий потенциал имеет вид

$$\Phi = \Phi_{LJ} + \Phi_{HB}, \tag{1.39}$$

где потенциал Леннард-Джонса Φ_{LJ} определяется обычным образом

$$\Phi_{LJ} = 4\ \varepsilon_{LJ} \left(\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right) \tag{1.40}$$

и член Φ_{HB} представляет водородную связь, определяемый как

$$\Phi_{HB} = \varepsilon_{HB} \cdot G(r_{ij} - r_{HB}) \sum_{ij}^{N} G(\vec{i}_k \cdot \vec{u}_{ij} - 1) G(\vec{j}_l \cdot \vec{u}_{ij} + 1), \tag{1.41}$$

где G — функция Гаусса

$$G(x) = e^{\frac{-x^2}{2\sigma^2}},$$

$\vec{u}_{ij}$ — единичный вектор, соединяющий центры молекул, $\vec{i}_k$ и $\vec{j}_l$ — единичные векторы ориентации плеч, r_{ij} — расстояние между молекулами, а угол между плечами молекулы равен $120°$.

Термодинамика модели MB

Поскольку MD-моделирование выполняется в микроканоническом ансамбле, где энтропия $S(N,V,E)$ зависит от количества частиц (N), объема (V) и энергии (E), термодинамические свойства должны быть записаны в терминах средних по времени параметров, которые могут быть оценены в процессе моделирования.

Используя первый закон термодинамики

$$dE = TdS - PdV + \mu dN, \tag{1.42}$$

и правило дифференцирования сложной функции для энтропии

$$dS = \frac{dS}{dE}dE + \frac{dS}{dV}dV + \frac{dS}{dN}dN,$$

получаем следующие зависимости для температуры (T) и давления (P):

$$\left(\frac{dS}{dE}\right)_{V,N} = \frac{1}{T}, \tag{1.43}$$

$$\left(\frac{dS}{dV}\right)_{E,N} = \frac{P}{T}. \tag{1.44}$$

Изохорная теплоемкость определяется как

$$C_V = \left(\frac{dE}{dT}\right)_{V,N} = \left[\left(\frac{\partial^2 S}{\partial E^2}\right)_{V,N}\right]^{-1}, \tag{1.45}$$

изотермическая сжимаемость как

$$\beta_T = -V\left(\frac{\partial P}{\partial V}\right)_{T,N} = -V\left[\left(\frac{\partial P}{\partial V}\right)_{V,N} + \frac{T}{C_V}\left(\frac{\partial P}{\partial T}\right)^2_{V,N}\right], \tag{1.46}$$

коэффициент теплового расширения как

$$\alpha_P = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_{P,N} = -V\left(\frac{\partial P}{\partial V}\right)_{T,N}\left(\frac{dP}{dT}\right)_{V,N} = \beta_T\left(\frac{dP}{dT}\right)_{V,N}, \tag{1.47}$$

где $\frac{dP}{dT}$ можно вычислить численно, используя (1.43) и (1.44).

Индекс N может быть опущен в уравнениях (1.43 – 1.47), так как μ (химический потенциал) считается постоянным.

Объем фазового пространства Ω для микроканонического ансамбля определяется как интеграл по всем координатам и импульсам фазового пространства ($\vec{q}$ и $\vec{p}$) [89]:

$$\Omega(N,V,E) = M\int \delta(E - H[\vec{q},\vec{p}])\, d\vec{q}\, d\vec{p}, \tag{1.48}$$

где $H[\vec{q},\vec{p}]$ — Гамильтониан системы. Плотность фазового пространства Σ определяется как [89]

$$\Sigma(N,V,E) = M\int \Theta(E - H[\vec{q},\vec{p}])\, d\vec{q}\, d\vec{p}, \tag{1.49}$$

где Θ — функция Хевисайда, а энтропия может быть определена как

$$S = k_B\, ln\Sigma. \tag{1.50}$$

Зависимость между Ω и Σ задаётся как

$$\Omega(N,V,E) = \frac{\partial}{\partial E}\Sigma(N,V,E).$$

Интегралы (1.48) и (1.49) не могут быть рассчитаны явно из-за их размерности 6N и 6N-1.

Предполагая разделение гамильтониана системы на потенциальную и кинетическую энергии

$$H[q^N,p^N] = \Phi(q^N) + K(p^N),$$

где q^N и p^N — компоненты вектора фазового пространства, преобразование Лапласа (1.48) можно найти, вводя коэффициент нормализации $M = \frac{E_0}{N!h^{3N}}$:

$$L(\Omega) = M \int dp^N \int dq^N \int dE\, e^{-\beta E} \delta(E - \Phi(q^N) - K(p^N)) = \tag{1.51}$$
$$= M \int dp^N \int dq^N e^{-\beta K(p^N)} e^{-\beta \Phi(q^N)} =$$
$$= M \int dq^N\, e^{-\beta \Phi(q^N)} \int dp^N\, e^{-\beta K(p^N)},$$

где h — постоянная Планка, E_0 — толщина энергетического слоя.

В декартовой системе координат кинетическая энергия модели МВ имеет три компоненты (K_x, K_y, K_φ): две поступательных

$$K_x(p_x) = \frac{p_x^2}{2m},$$

$$K_y(p_y) = \frac{p_y^2}{2m}$$

и одну вращательную

$$K_\varphi(\varphi) = \frac{\omega^2}{2I},$$

где p_x, p_y — импульсы вдоль x и y соответственно, ω — скорость вращения. Части интеграла (1.51) независимы, поэтому их можно взять отдельно, выполнив замену $p'^2 = \frac{p^2}{2m}$, $p = p'\sqrt{2m}$ и $dp = dp'\sqrt{2m}$ (обратите внимание, что $dp^N = dp^1\, dp^2 ... dp^N$).

Кинетическая часть равна

$$\int dp^N\, e^{-\beta K(p^N)} = (2m\pi)^{3N/2} \beta^{-3N/2},$$

поскольку

$$\int dp'\, e^{-\beta p'^2} = \sqrt{\frac{\pi}{\beta}},$$

и преобразование Лапласа (1.51) упрощается до

$$L(\Omega) = M \int dq^N\, e^{-\beta \Phi(q^N)}\, (2m\pi)^{3N/2} \beta^{-3N/2} = M\, (2m\pi)^{3N/2} \int dq^N\, e^{-\beta \Phi(q^N)}\, \beta^{-3N/2}. \tag{1.52}$$

Применяя обратное преобразования Лапласа к приведенному выше интегралу

$$L^{-1}(L(\Omega)) = L^{-1}\left(\int dq^N\, e^{-\beta \Phi(q^N)} \beta^{-3N/2} \right) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} dq^N \beta^{-3N/2} e^{\beta(E - \Phi(q^N))} d\beta, \tag{1.53}$$

где

$$\int \frac{1}{p^\nu} e^{-ap} dp = \frac{(E-a)^{\nu-1}}{\Gamma(\nu)}, \tag{1.54}$$

получаем

$$\Omega(N,V,E) = M\, (2m\pi)^{3N/2} \int \frac{(E - \Phi(q^N))^{3N/2-1}}{\Gamma(3N/2)} \Theta(E - \Phi(q^N)) dq^N. \tag{1.55}$$

Используя соотношение (1.51), плотность фазового пространства становится равна

$$\Sigma(N,V,E) = M\ (2m\pi)^{3N/2} \int \frac{(E-\Phi(q^N))^{3N/2}}{\Gamma(3N/2+1)}\Theta(E-\Phi(q^N))dq^N. \tag{1.56}$$

Зависимость от объема неявно содержится в пределах интегрирования, таким образом, заменой переменных в интегралах (1.55) и (1.56), пределы делаются независимыми от размера системы. С использованием замены координат и импульсов в фазовом пространстве

$$q = A^{\frac{1}{2}} q^*$$

и

$$p = A^{\frac{1}{2}} p^*$$

интегралы фазового пространства становятся

$$\Sigma(N,V,E) = M_1 A^N \int dq^{*N} \frac{(E-\Phi((A^{1/2}q^*)^N))^{3N/2}}{\Gamma(3N/2+1)}\Theta(E-\Phi((A^{1/2}q^*)^N)) \tag{1.57}$$

и

$$\Omega(N,V,E) = M_1 A^N \int dq^{*N} \frac{(E-\Phi((A^{1/2}q^*)^N))^{3N/2-1}}{\Gamma(3N/2)}\Theta(E-\Phi((A^{1/2}q^*)^N), \tag{1.58}$$

где $M_1 = M(2m\pi)^{3N/2}$.

Используя приведенные выше соотношения, термодинамические величины записываются через объем и плотность фазового пространства.

Температура

$$T = \frac{1}{k_B}\frac{\Sigma}{\Omega}, \tag{1.59}$$

давление

$$P = \frac{1}{\Omega}\frac{\partial \Sigma}{\partial V}, \tag{1.60}$$

изохорная теплоемкость

$$\frac{1}{C_V} = \left(\frac{\partial T}{\partial E}\right)_V = \left(\frac{\partial\left(\frac{1}{k_B}\frac{\Sigma}{\Omega}\right)}{\partial E}\right)_V = \frac{1}{k_B}\left(\frac{1}{\Omega}\frac{\partial\Sigma}{\partial E} - \frac{\Sigma}{\Omega^2}\frac{\partial\Omega}{\partial E}\right) = \frac{1}{k_B}\left(1 - \frac{\Sigma}{\Omega^2}\frac{\partial\Omega}{\partial E}\right), \tag{1.61}$$

изотермическая сжимаемость

$$\frac{1}{\beta_T} = -V\left(\frac{\partial P}{\partial V}\right)_T = -V\left(\frac{\partial\left(\frac{1}{\Omega}\frac{\partial\Sigma}{\partial V}\right)}{\partial V}\right)_T = -V\left(\frac{1}{\Omega}\frac{\partial^2\Sigma}{\partial V^2}\right)_T + V\left(\frac{1}{\Omega}\frac{\partial\Sigma}{\partial V}\frac{1}{\Omega}\frac{\partial\Omega}{\partial V}\right)_T. \tag{1.62}$$

Среднее по ансамблю наблюдаемого A, описываемого функцией фазового пространства $a(q^N,p^N)$, равно [58]

$$A = \frac{\int a(q^N,p^N)\delta(E-H[q^N,p^N])dq^N dp^N}{\int \delta(E-H[q^N,p^N])dq^N}. \tag{1.63}$$

Используя уравнения (1.59–1.62) с уравнением (1.63) термодинамические свойства получаются из усреднения параметров системы по ансамблю/времени:

Температура

$$k_B\,T = \frac{\Sigma}{\Omega} = \frac{2}{3N}\,\langle K\rangle\,, \tag{1.64}$$

где $\langle\rangle$ — усреднение по ансамблю/времени, k_B — постоянная Больцмана.

Давление

$$P = \frac{1}{\Omega}\frac{\partial\Sigma}{\partial A} = \frac{N}{A}\frac{2}{3N}\,\langle K\rangle - \left\langle\frac{\partial\Phi}{\partial A}\right\rangle = \rho\;k_BT - \left\langle\frac{\partial\Phi}{\partial A}\right\rangle, \tag{1.65}$$

где $\rho = N/A$ — плотность числа частиц.

Изохорная теплоемкость

$$\frac{C_V}{Nk_B} = \left(\frac{2}{3}\,\langle K\rangle\left\langle K^{-1}\right\rangle + N(1 - \langle K\rangle\left\langle K^{-1}\right\rangle)\right)^{-1}. \tag{1.66}$$

Первый член в уравнении для изотермической сжимаемости (1.46) равен

$$\begin{aligned}\left(\frac{\partial P}{\partial A}\right)_S = \Bigg(&\frac{N(N-1)}{A^2}k_BT - \left\langle\frac{\partial^2\Phi}{\partial A^2}\right\rangle - 2\rho\left\langle\frac{\partial\Phi}{\partial A}\right\rangle + \\ &+ (3N/2-1)\left\langle K^{-1}\left(\frac{\partial\Phi}{\partial A}\right)^2\right\rangle\Bigg) - \\ &- \left(\rho - (3N/2-1)\left\langle K^{-1}\frac{\partial\Phi}{\partial A}\right\rangle\right)\,P.\end{aligned} \tag{1.67}$$

1.3.2 MD реализация модели MB

Производные потенциалов в модели MB

Для интегрирования уравнений движения MD, нужно получить силы действующие на каждую частицу. Для консервативных потенциалов сила, действующая на атом i, рассчитывается как минус градиент потенциала. Для модели MB помимо поступательной силы производная по углу φ создает крутящий момент τ:

$$\tau = -\frac{\partial\Phi}{\partial\varphi}. \tag{1.68}$$

Дополнительная степень свободы φ дает угловую скорость ω. Соответствующий момент инерции I следует ввести чтобы охарактеризовать вращение молекул воды (I — эквивалентно массе для вращения) и шаг по времени должен быть выбран соответствующим образом.

Формулы для термодинамических свойств зависят, помимо кинетической энергии, только от производных потенциала. Производные первого и второго порядка потенциальной энергии по площади равны:

$$\frac{d\Phi}{dA} = \frac{1}{2A}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} dx\frac{d\Phi_{ij}}{dx} + dy\frac{d\Phi_{ij}}{dy}, \tag{1.69}$$

и

$$\frac{d^2\Phi}{dA^2} = \frac{1}{4A^2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} dx^2\frac{d^2\Phi_{ij}}{dx^2} + 2\cdot dxdy\frac{d^2\Phi_{ij}}{dxdy} + dy^2\frac{d^2\Phi_{ij}}{dy^2}, \tag{1.70}$$

где Φ_{ij} — потенциал, dx и dy — расстояния вдоль направлений x и y между частицами i^{th} и j^{th} соответственно.

Потенциал Леннард-Джонса

Производные потенциала Леннард-Джонса (LJ) для частицы i приводят к

$$F_{LJ} = -\nabla_q\Phi_{LJ} = -\frac{\Phi_{LJ}}{dr_{ij}}\frac{dr_{ij}}{dq}, \tag{1.71}$$

где $r_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$ и q обозначают координаты x_i или y_i, обозначение ∇_q — либо $\frac{\partial}{\partial x_i}$, либо $\frac{\partial}{\partial y_i}$.

Используя уравнение (1.40) получаем

$$\frac{d\Phi_{LJ}}{dr_{ij}} = -24\cdot\varepsilon_{LJ}\cdot\sigma^6\left(2\,\sigma^6 r_{ij}^{-14} - r_{ij}^{-8}\right)\cdot r_{ij},$$

где производные расстояния по x_i и y_i равны

$$\frac{dr_{ij}}{dx_i} = -\frac{x_j - x_i}{r_{ij}},$$

и

$$\frac{dr_{ij}}{dy_i} = -\frac{y_j - y_i}{r_{ij}}$$

соответственно.

Производные второго порядка потенциала LJ по компонентам x_i и y_i равны

$$\frac{d^2\Phi_{LJ}}{dq^2} = 24\cdot\varepsilon_{LJ}\cdot\sigma^6\left(\frac{\left(26\,\sigma^6 - 7\,r_{ij}^6\right)\cdot\left(\frac{dr_{ij}}{dq}\right)^2}{r_{ij}^{14}} + \frac{r_{ij}\left(-2\sigma^6 + r_{ij}^6\right)\left(\frac{d^2r_{ij}}{dq^2}\right)}{r_{ij}^{14}}\right), \tag{1.72}$$

где

$$\frac{d^2r_{ij}}{dx_i^2} = -\frac{(x_j - x_i)^2}{r_{ij}^3} + \frac{1}{r_{ij}}$$

и

$$\frac{d^2r_{ij}}{dy_i^2} = -\frac{(y_j - y_i)^2}{r_{ij}^3} + \frac{1}{r_{ij}}.$$

Крутящий момент потенциала Леннард-Джонса равен нулю.

Член, описывающий водородную связь

Производная члена HB потенциала по x_i, y_i и φ_i выводится из уравнения (1.41) как

$$\frac{1}{\varepsilon_{HB}}\nabla_q\Phi_{HB} = \tag{1.73}$$
$$= \nabla_q G(r_{ij}-r_{HB})\sum_{ij}^{N} G(\vec{i}_k\cdot\vec{u}_{ij}-1)G(\vec{j}_l\cdot\vec{u}_{ij}+1)+$$
$$+G(r_{ij}-r_{HB})\sum_{ij}^{N}\nabla_q G(\vec{i}_k\cdot\vec{u}_{ij}-1)G(\vec{j}_l\cdot\vec{u}_{ij}+1)+$$
$$+G(r_{ij}-r_{HB})\sum_{ij}^{N} G(\vec{i}_k\cdot\vec{u}_{ij}-1)\nabla_q G(\vec{j}_l\cdot\vec{u}_{ij}+1).$$

Введя новые обозначения $h_i=\vec{i}_k\cdot\vec{u}_{ij}$ и $h_j=\vec{j}_l\cdot\vec{u}_{ij}$ скалярные произведения h_i и h_j становятся

$$h_i=\frac{(x_j-x_i)\cdot cos\ \varphi_i+(y_j-y_i)\cdot sin\ \varphi_i}{r_{ij}},$$

и

$$h_j=\frac{(x_j-x_i)\cdot cos\ \varphi_j+(y_j-y_i)\cdot sin\ \varphi_j}{r_{ij}}$$

соответственно.

Используя обозначения, приведенные выше, разложим члены в (1.73) в

$$\nabla_q G(r_{ij}-r_{HB})=\frac{dG(r_{ij}-r_{HB})}{dr_{ij}}\frac{dr_{ij}}{dq},$$

где

$$\frac{dG(r_{ij}-r_{HB})}{dr_{ij}}=-\frac{(r_{ij}-r_{HB})}{\sigma^2_{HB}}G(r_{ij}-r_{HB}).$$

Члены, зависящие от ориентации, становятся

$$\nabla_q G(h_i-1)=\frac{dG(h_i-1)}{dh_i}\frac{dh_i}{dq}=-\frac{(h_i-1)}{\sigma^2_\theta}G(h_i-1)\frac{dh_i}{dq},$$

и

$$\nabla_q G(h_j+1)=\frac{dG(h_j+1)}{dh_i}\frac{dh_j}{dq}=-\frac{(h_j+1)}{\sigma^2_\theta}G(h_j+1)\frac{dh_j}{dq},$$

где

$$\frac{dh_i}{dx_i}=-\frac{cos\ \varphi_i}{r_{ij}}-\frac{(x_j-x_i)\cdot cos\ \varphi_i+(y_j-y_i)\cdot sin\ \varphi_i}{r^2_{ij}}\frac{dr_{ij}}{dx_i},$$
$$\frac{dh_j}{dx_i}=-\frac{cos\ \varphi_j}{r_{ij}}-\frac{(x_j-x_i)\cdot cos\ \varphi_j+(y_j-y_i)\cdot sin\ \varphi_j}{r^2_{ij}}\frac{dr_{ij}}{dx_i},$$
$$\frac{dh_i}{dy_i}=-\frac{sin\ \varphi_i}{r_{ij}}-\frac{(x_j-x_i)\cdot cos\ \varphi_i+(y_j-y_i)\cdot sin\ \varphi_i}{r^2_{ij}}\frac{dr_{ij}}{dy_i},$$
$$\frac{dh_j}{dy_i}=-\frac{sin\ \varphi_j}{r_{ij}}-\frac{(x_j-x_i)\cdot cos\ \varphi_j+(y_j-y_i)\cdot sin\ \varphi_j}{r^2_{ij}}\frac{dr_{ij}}{dy_i}.$$

Крутящий момент HB для частицы i^{th} равен

$$\nabla_{\varphi_i}\Phi_{HB} = \varepsilon_{HB}\cdot G(r_{ij}-r_{HB})\cdot\sum_{ij}^{N}\nabla_{\varphi_i}G(h_i-1)G(h_j+1),$$

где

$$\frac{dh_i}{d\varphi_i} = \frac{-(x_j-x_i)sin\ \varphi_i + (y_j-y_i)cos\ \varphi_i}{r_{ij}}.$$

Вводя новые обозначения $A = G(r_{ij}-r_{HB})$, $B = G(h_i-1)$, $C = G(h_j+1)$ производные второго порядка потенциала по координате для члена HB равны

$$\frac{d^2\Phi_{HB}}{dq^2} = \frac{d^2\left(A\sum B\cdot C\right)}{dq^2} = \nabla_q^2 A\sum B\cdot C + A\sum\nabla_q^2 B\cdot C + A\sum B\cdot\nabla_q^2 C + \tag{1.74}$$
$$+2\left(A\sum\nabla_q B\cdot\nabla_q C + \nabla_q A\sum B\cdot\nabla_q C + \nabla_q A\sum\nabla_q B\cdot C\right).$$

Производная второго порядка функции Гаусса $G(f(q)) = e^{-\frac{f(q)^2}{2\sigma^2}}$ равна

$$\frac{d^2G(f(q))}{dq^2} = \frac{G(f(q))\left((-\sigma^2+f(q)^2)\cdot\left(\frac{df(q)}{dq}\right)^2 - \sigma^2\cdot f(q)\cdot\left(\frac{d^2f(q)}{dq^2}\right)\right)}{\sigma^4}, \tag{1.75}$$

в то время как

$$\frac{d^2h_i}{dx^2} = (y_i-y_j)\left(\frac{sin(\varphi_i)(-2(x_i-x_j)^2+(y_i-y_j)^2)}{r_{ij}^5} + \frac{3\ \cos(\varphi_i)(x_i-x_j)(y_i-y_j)}{r_{ij}^5}\right), \tag{1.76}$$

$$\frac{d^2h_i}{dy^2} = (x_i-x_j)\left(\frac{cos(\varphi_i)((x_i-x_j)^2-2(y_i-y_j)^2)}{r_{ij}^5} + \frac{+3\ \sin(\varphi_i)(x_i-x_j)(y_i-y_j)}{r_{ij}^5}\right), \tag{1.77}$$

$$\frac{d^2h_j}{dx^2} = (y_i-y_j)\left(\frac{sin(\varphi_j)(-2(x_i-x_j)^2+(y_i-y_j)^2)}{r_{ij}^5} + \frac{+3\ \cos(\varphi_j)(x_i-x_j)(y_i-y_j)}{r_{ij}^5}\right), \tag{1.78}$$

$$\frac{d^2h_j}{dy^2} = (x_i-x_j)\left(\frac{cos(\varphi_j)((x_i-x_j)^2-2(y_i-y_j)^2)}{r_{ij}^5} + \frac{+3\ \sin(\varphi_j)(x_i-x_j)(y_i-y_j)}{r_{ij}^5}\right). \tag{1.79}$$

Перекрестные производные потенциалов рассчитываются численно используя конечные разности из-за громоздскости формул.

Единицы измерений

Были использованы безразмерные единицы измерения [60]:

$$T = \frac{\varepsilon_{HB}}{k_B}T^*, \tag{1.80}$$

$$r = r_{HB}\cdot r^*, \tag{1.81}$$

$$m = m_{H_2O} \cdot m^*, \tag{1.82}$$

$$t = \sqrt{m \cdot r_{HB}^2 / \varepsilon_{HB}} \cdot t^*, \tag{1.83}$$

$$P \cdot V = \varepsilon_{HB} \cdot P^* \cdot V^*, \tag{1.84}$$

где * обозначает безразмерные переменные. Временной шаг выбирали $\Delta t^* = 0.001$ достаточно малым, чтобы учесть вращение молекул, которое значительно быстрее, чем перемещение.

Параметры моделирования перечислены в таблице 1. Значение σ^*_{HB} должно быть небольшим, чтобы была сформирована только одна водородная связь [60].

Переменная	Реальное значение	Единицы	Безразмерное значение
m_{H_2O}[90]	$18 \cdot 10^{-3}$	$\frac{kg}{mol}$	1
r_{HB}[90]	2.95		1
ρ [90]	10^3	$\frac{kg}{m^3}$	0.9
ε_{HB}[60]	12.742	$\frac{kJ}{mol}$	-1
ε_{LJ}[60]	1.274	$\frac{kJ}{mol}$	0.1
σ_{HB} [60]	0.25		0.085
σ_{LJ} [60]	2.065		0.7
I[90]	$2.938 \cdot 10^{-47}$	$kg \cdot m^2$	0.001126
k_B	$1.38 \cdot 10^{-23}$	$\frac{J}{K}$	1

Таблица 1 — Переменные, использованные в моделировании MD. r_{HB} — длина водородной связи, m_{H_2O} — масса молекулы воды, ε_{HB} — энергия водородной связи, ε_{LJ} — энергия Леннард-Джонса, σ^*_{LJ} — $0.7\ r^*_{HB}$.

Рассчитанные MD величины

Обычное моделирование MD проводится при постоянном числе молекул, объеме, и энергии, т.е. с использованием ансамбля NVE. Существуют различные методы интегрирования уравнений движения. Был использован метод Верле с масштабированием скорости в качестве алгоритма термостатирования [58]. Кинетическая энергия для модели MB равна

$$K = \sum_{i=1}^{N} \frac{m\vec{v}_i^2}{2} + \frac{I\omega_i^2}{2}, \tag{1.85}$$

где I — момент инерции, $\vec{v}_i$ и ω_i — поступательная и угловая (вращательная) скорости.

Структура воды MB характеризуется функцией радиального распределения (RDF)

$$g_r^{(2)}(r) = \frac{2V}{N^2} \left\langle \sum_{i<j} \delta(r - |\vec{u}_{ij}|) \right\rangle. \tag{1.86}$$

Динамические характеристики описываются корреляционными функциями скорости, которые не могут быть получены методом Монте-Карло. Автокорреляционная функция скорости (VACF) равна

$$f_v(\tau) = \langle \vec{v}(t) \cdot \vec{v}(t+\tau) \rangle \,, \tag{1.87}$$

где $\vec{v}(t)$ и $\vec{v}(t+\tau)$ — поступательные скорости в моменты времени t и $t+\tau$. Если процесс стационарный, f_v не зависит от t. Корреляционная функция скорости вращения определена для вращательной скорости $\omega(t) = \frac{\partial \varphi}{\partial t}$:

$$f_\omega(\tau) = \langle \omega(t) \cdot \omega(t+\tau) \rangle \,. \tag{1.88}$$

Моделировалась система из 2500 частиц МВ.

Гибридная связь представлений

Связи между представлениями дает новые уравнения движения, (1.19) и (1.24), для моделирования MD, которые интегрируются вместо стандартных уравнений движения MD.

Поскольку расстояние между центром и плечами частицы МВ мало, считалось, что внешнее силовое поле, обусловленное связью, не меняется на этом расстоянии. Поэтому была применена техника связи только к поступательной скорости частиц МВ.

1.3.3 Модель МВ: Соответствие реальной воде и гидродинамике

Модель МВ двумерна, следовательно, имеет три степени свободы (x, y и φ). Полученные уравнения для термодинамических свойств имеют вид согласующийся с предыдущими работами [89], несмотря на разницу в третьей степени свободы в нашей модели.

Как было показано выше, термодинамические свойства состоят из комбинаций кинетической энергии K и производных первого и второго порядка от потенциала по площади, $\frac{\partial \Phi}{\partial A}$ и $\frac{\partial^2 \Phi}{\partial A^2}$.

Сходимость изотермической сжимаемости – самая медленная из всех анализируемых свойств. Это связано с флуктуационными членами в уравнении (1.67), особенно с первой производной потенциала по площади.

Флуктуации первой производной потенциала по площади для различных размеров системы представлены на рисунке 1.9, поэтому для малых систем время усреднения следует брать достаточно большим, чтобы обеспечить сходимость среднего значения этих флуктуаций. Для системы в 2500 частиц сходимость флуктуирующего члена $\frac{d\Phi}{dA}$ представлена на рисунке 1.10.

Динамические свойства воды характеризуются поступательной (VACF) и вращательной (RVACF) автокорреляционными функциями. Полученные автокорреляционные функции, Рисунок

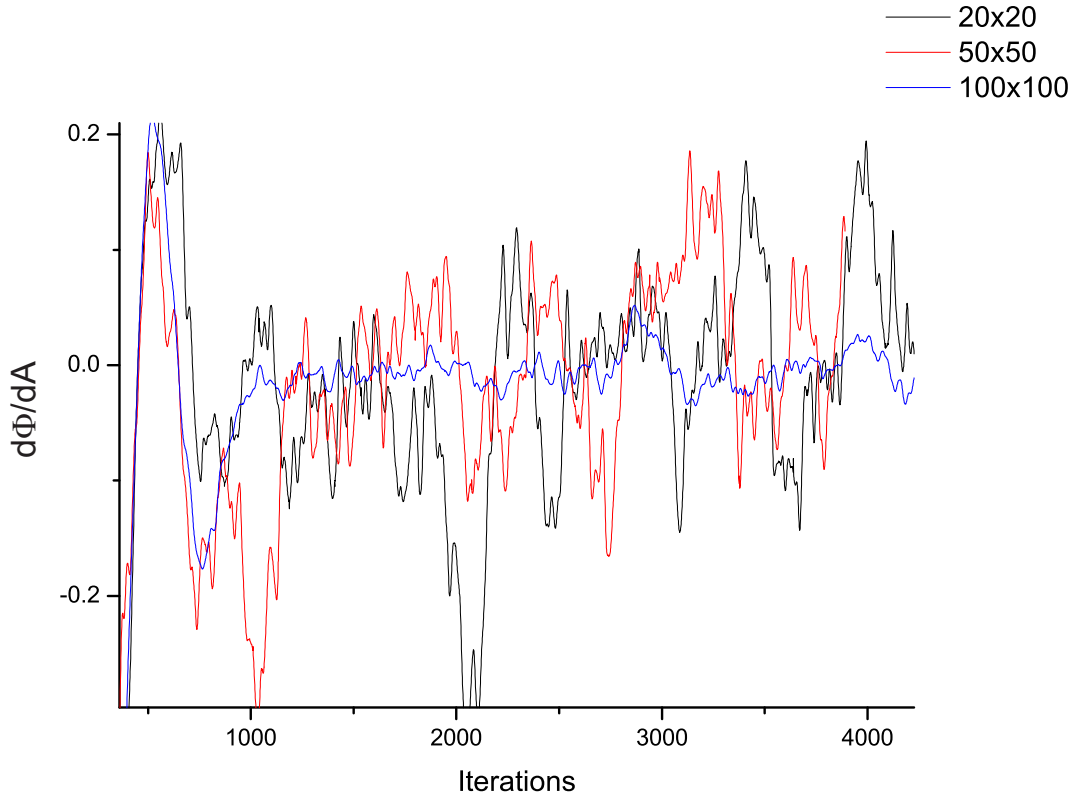


Рисунок 1.9 — Флуктуации производной первого порядка по площади для разных размеров системы (указано количество молекул), $T^* = 0.21$.

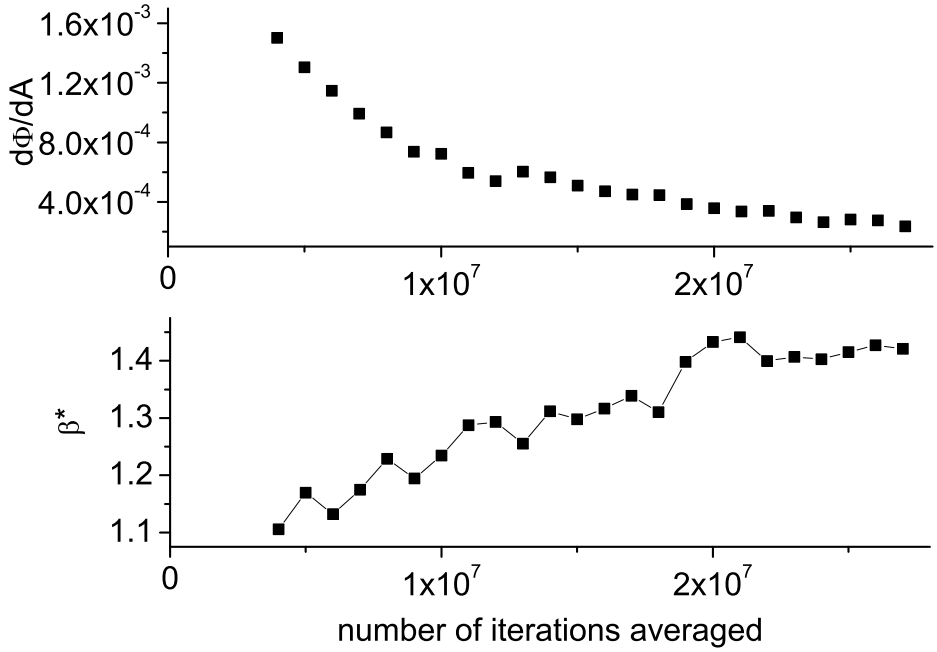


Рисунок 1.10 — Сходимость производной первого порядка потенциала МВ по площади и изотермической сжимаемости, $T^* = 0.21$.

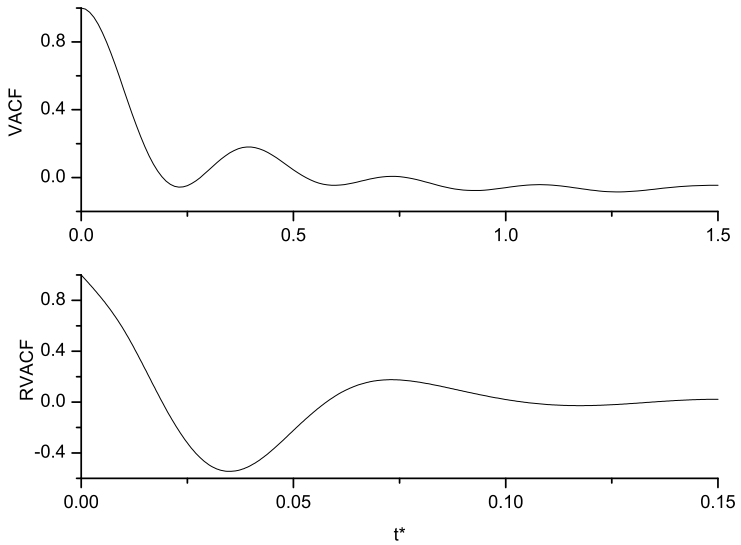


Рисунок 1.11 — Автокорреляционные функции поступательной и вращательной скоростей модели МВ.

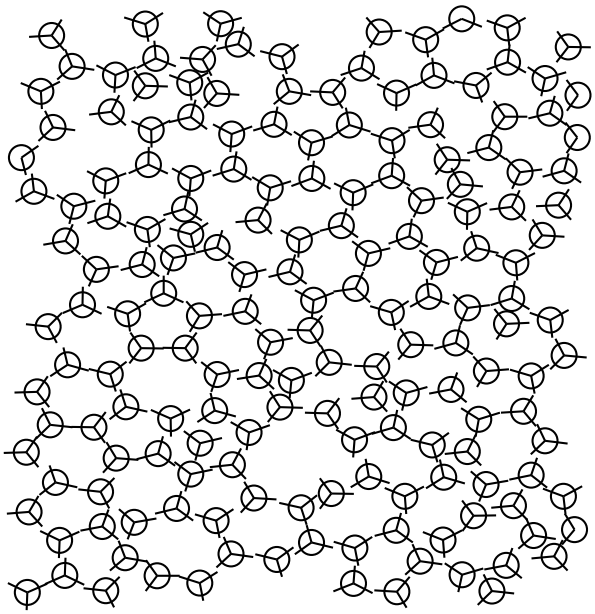


Рисунок 1.12 — Мгновенный снимок структуры воды МВ, $T^* = 0.15$.

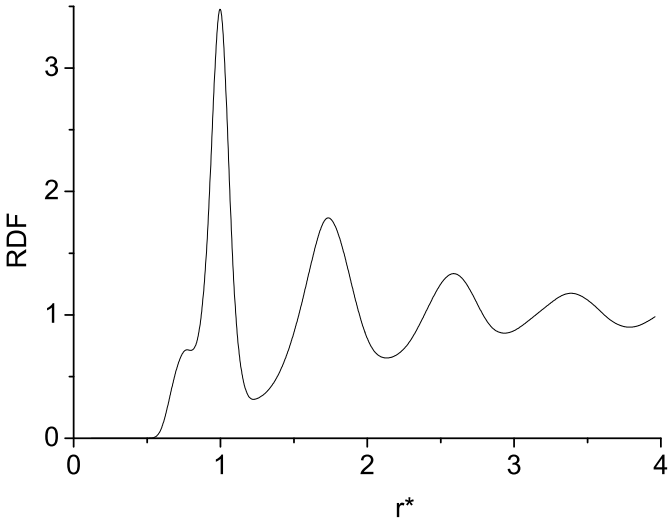


Рисунок 1.13 — Функция радиального распределения для воды МВ.

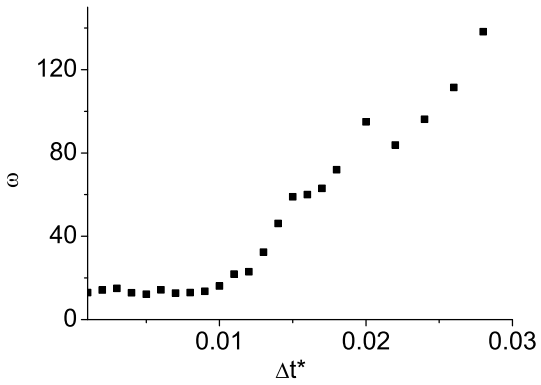


Рисунок 1.14 — Стандартное отклонение скорости вращения с временным шагом Δt^*.

1.11, 1.16, совпадают с таковыми в 3D модели, например, SPC [91]. Минимумы VACF и RVACF расположены примерно в правильных положениях, что подтверждает правильность значений момента инерции и массы.

Структура, образованная моделью MB, Рисунок 1.12 и Рисунок 1.13, 1.17, похожа на реалистичную воду. Она также воспроизводит результаты Бен-Наима [88]. Максимумы на RDF указывают на сольватационные оболочки разного порядка. Максимумы примерно соответствуют гексагональной структура льда, а небольшой максимум на ≈ 0.7 образован «междоузельными» молекулами воды (см. Рисунок 1.12).

Результаты моделирования сходятся, если временной шаг меньше $\Delta t = 0.01$. Например, стандартное отклонение скорости вращения быстро растет с шагом по времени, Рисунок 1.14. Шаг по времени меньше влияет на поступательные скорости и другие свойства. Таким образом, максимальный временной шаг ограничен вращением молекулы и выбран равным $\Delta t^* = 0.001$, что соответствует $t = 0.35$ fs.

Уравнение состояния (УС), Рисунок 1.15, было получено из моделирования MB воды при температуре $T^* = 0.196$ и различной плотности с использованием MD.

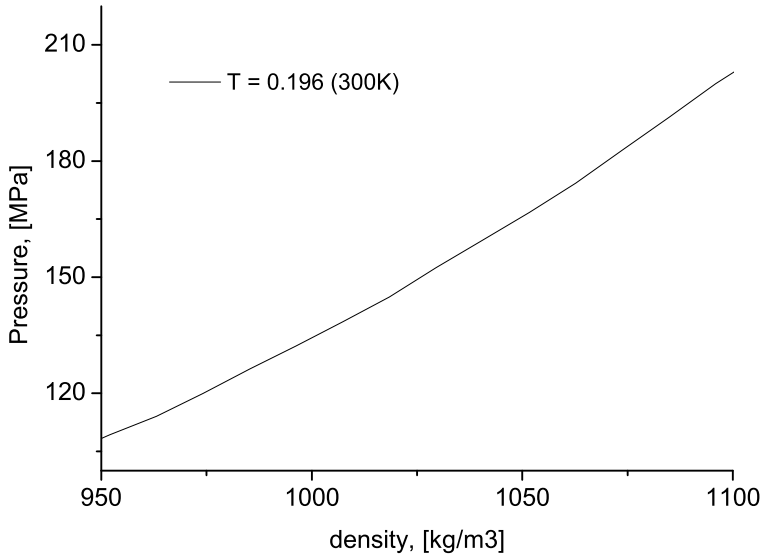


Рисунок 1.15 — Уравнение состояния для модели воды MB.

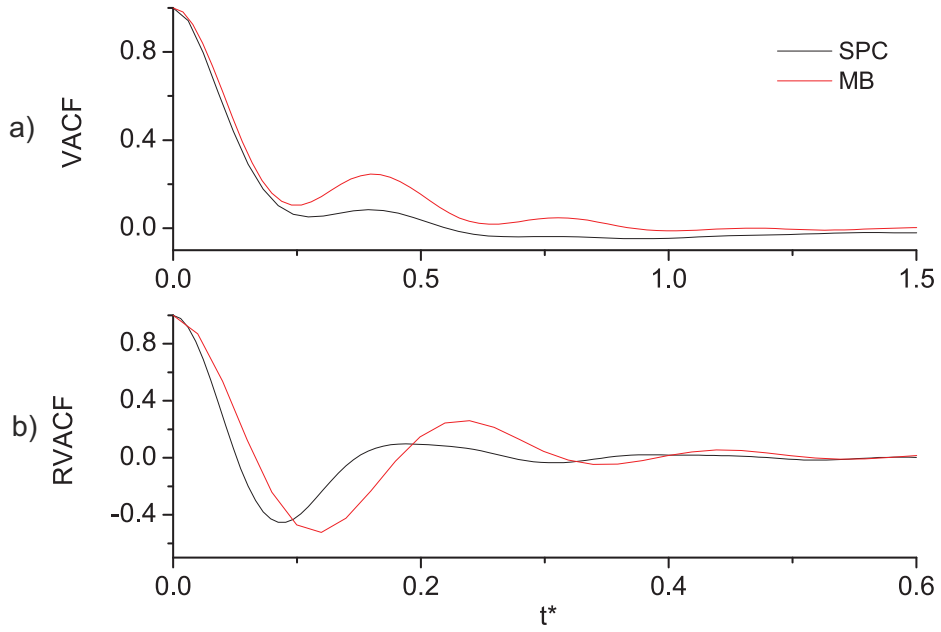


Рисунок 1.16 — а) Поступательная и б) вращательная автокорреляционные функции в $T^* = 0.195$ для моделей MB и SPC.

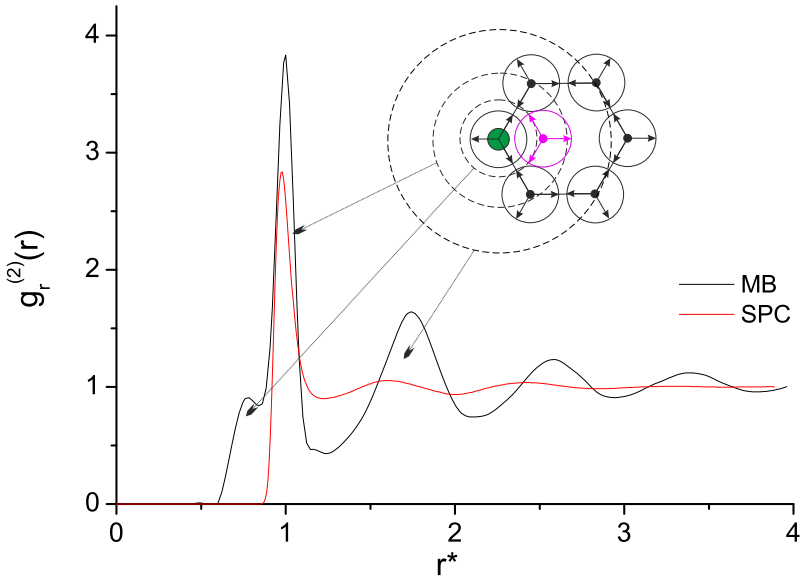


Рисунок 1.17 — Функция радиального распределения при $T^* = 0.16$. Центральная молекула показана зеленым цветом. «Междоузельная» вода выделена пурпурным цветом.

Моделирование LL-FH требует макроскопических транспортных параметров для континуальной фазы, таких как объемная и сдвиговая вязкости, а также уравнение состояние для расчета давления, которое было получено из отдельного моделирования чистой MB.

Обычно используемый метод определения вязкости – использование соотношения Грина-Кубо [92], где объемная вязкость рассчитывается как интеграл по времени

$$\eta = \frac{V_0}{k_B T} \lim_{t\to\infty} \int_0^t B_{ACF}(t')dt', \tag{1.89}$$

где V_0 — объем системы,

$$B_{ACF}(t') = \langle \delta P(t')\delta P(0)\rangle, \tag{1.90}$$

$\delta P(t) = P(t) - P_{avg}$ — флуктуации давления.

Для сдвиговой вязкости использовалось следующее соотношение:

$$\xi = \frac{V_0}{k_B T} \int_0^t S_{ACF}(t')dt', \tag{1.91}$$

где

$$S_{ACF}(t') = \langle P_{xy}(t') \cdot P_{xy}(0) \rangle. \tag{1.92}$$

Недиагональные компоненты тензора давления (в двумерном случае) вычисляются с помощью

$$P_{xy} = \frac{1}{V} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(m \cdot u_i^x \cdot u_i^y + f_{ij}^x \cdot r_{ij}^y \right), \tag{1.93}$$

где u_i — скорость частицы, f_{ij} — межмолекулярная сила, r_{ij} — расстояние между молекулами, а верхние индексы обозначают компоненты x и y.

Гибридное моделирование проводилось на регулярной сетке в 2D-квадрате, разделеном на 5 на 5 ячеек с 10000 молекулами воды и 25 узлами FH при температуре $300[K]$ и плотности $\rho = 1054 \left[\frac{kg}{m^3}\right]$.

1.3.4 3D модели: аргон, вода и пептид в воде

Оба уравнения LL-FH (1.1) и модифицированные уравнения MD (1.35) были реализованы как внутренние процедуры пакета GROMACS 5.0 для трех случаев: аргон, вода SPC/E и система пептид в воде. Общие параметры, используемые в MD и LL-FH частях приведены в таблице 2. Для всех моделей использовался ансамбль с постоянной температурой (NVT) и с термостатом Nosé-Hoover [93; 94], реализованным в GROMACS. Граничные условия во всех случаях были периодическими. Для водных моделей для ограничения длины связей использовался алгоритм LINCS [95].

Для тестирования нашего подхода был исследован жидкий аргон в равновесном состоянии при температуре T=2.4 и средней плотности ρ =0.607. Использованы приведенные единицы измерения, Таблица 3.

1.3.5 Диаланин в воде

Диаланин (L-аланил-L-аланин) — наименьший белок, состоящий только из двух аминокислотных остатков. Это рабочая система для биомолекулярных исследований, из-за своей простоты и легкости анализа. В цвиттерионной форме противоположные концы пептида заряжены, что заставляет их сильно взаимодействовать с окружающей водой. Взаимодействие осуществляется

Таблица 2 — Параметры моделирования, используемые в GROMACS для аргона и SPC/E воды и значения вязкости, используемые в коде LL-FH.

	Аргон	Аргон (тест на прохождение) акустической волны)	SPC/E вода
Количество атомов (молекулы)	64000	32000	91125 (30375)
Молекулярная масса (г моль$^{-1}$)	39.948	39.948	18.015
Температура, (K)	300	300	298.15
Объем ячейки, (нм^3)	16.21x16.21x16.21	32.424x8.106x8.106	9.686x9.686x9.686
Шаг по времени MD, (пс)	0.01	0.01	0.001
α, β, (nm^2 ps^{-1})	1000	1000	5000
Размер капли, (нм)	0.28	0	0.18
Средняя плотность (amu нм$^{-3}$)	600.24	600.0	602.18
Сдвиговая вязкость (amu nm^{-1} ps^{-1})	54.74	54.74	409.496
Объемная вязкость (amu nm^{-1} ps^{-1})	18.23	18.23	933.41

Таблица 3 — Приведенные единицы для жидкого аргона.

Время	2.1502e-12 sec
Плотность	1.6410e3 kg/m^3
Скорость	159.61 m/s
Температура	122.4 K
Сила	4.9240e-12 N
стресс	41.805 MPa
Число атомов	10000
Число ячеек LL-FH	5x5 с 400 атомами на ячейку

через водородные связи с молекулами воды, поэтому молекулярное представление воды в непосредственной близости от молекулы пептида имеет решающее значение. В то же время, в этом участвует множество водных слоев вокруг молекулы пептида, поэтому необходима правильная связь с объемом жидкости. Динамика биомолекулы не может быть правильно промоделирована в изоляция от относительно большого объема воды.

Перед расчетом коэффициента диффузии гибридная модель диаланина в воде была проверена на сохранение массы и флуктуаций импульса внутри зоны чистой молекулярной динамики s=0. Этот было необходим для подтверждения того, что атомы, соответствующие объему, привязанному к движущемуся центру масс пептида имеют правильную статистику в соответствии с теорией для чистой воды в равновесных условиях (пренебрегая влиянием пептида на относительно большой объем воды внутри сферы). Здесь можно вспомнить, что только флуктуации

комбинированной смеси автоматически сохраняются в модели односторонней связи и решение, описывающее MD частицы должно быть проверено апостериори. Примечательно, что необходимость апостериорного анализа эффективности текущего метода LL-FH/MD подобна другим численным подходам с открытыми границами в литература по моделированию.

На рисунке 1.4, 1.18 показаны гистограммы вычисленных функций плотности вероятности флуктуаций массы, $\frac{\sum_{p=1,N} m_p - \rho_0 V_0}{\rho_0 V_0}$ и импульса $\frac{\sum_{p=1,N} m_p u_p}{\rho_0 V_0 c_T}$, внутри движущейся MD сферы с пептидом, где ρ_0 — средняя плотность воды, V_0 — объем зоны MD, c_T — изотермическая скорость звука в воде, m_p, u_p – масса и x-компонента скорости частицы p.

Теоретические гауссовы распределения массы и импульса, соответствующие заданным температуре, обьему и скорости звука приведены на графике для сравнения.

Теоретические распределения соответствуют следующим функциям:

$$f(x) = A \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{1.94}$$

где A — константа калибровки, σ — дисперсия, так что $\sigma_m = \frac{std(\rho)}{\rho_0}$ для массы и $\sigma_{\rho u} = \sigma_m\sqrt{2}$ для импульса и

$$std(\rho) = \frac{1}{c_T}\sqrt{\rho_0 k_B \frac{T}{V_0}}, \quad \frac{std(u)}{c_T} = \frac{std(\rho)}{\rho_0}. \tag{1.95}$$

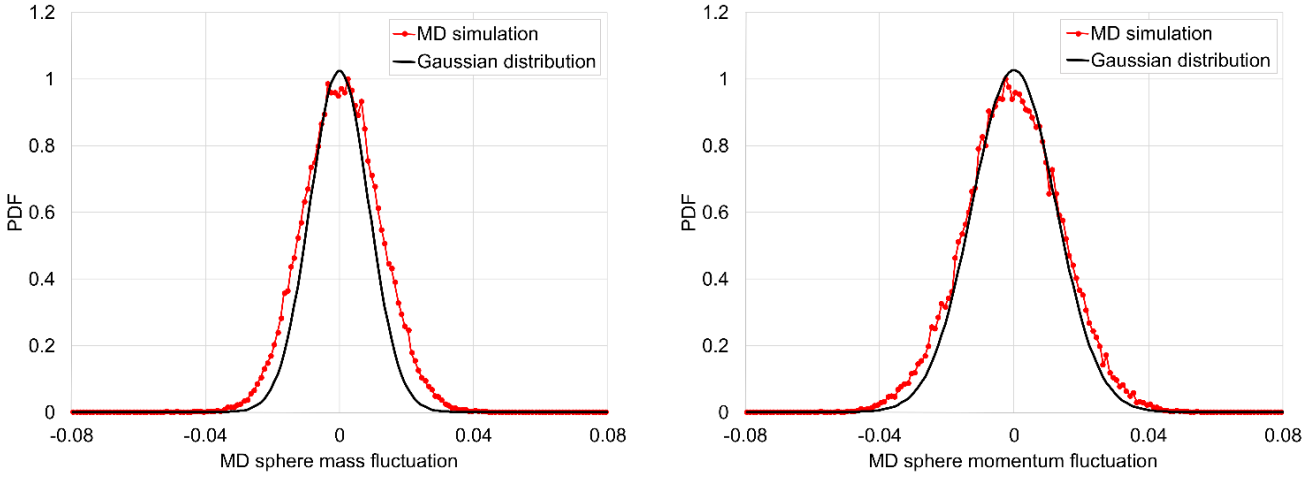


Рисунок 1.18 — Распределение функции плотности вероятности для массы (слева) и x-компоненты импульса (справа), теоретические распределения в каждом случае показаны сплошными черными линиями.

1.3.6 Вирус PCV2 в воде

Цирковирус свиней типа 2 (PCV2) связан с различными заболеваниями свиней, включая мультисистемный синдром истощения после отъема (PMWS) и синдром дерматита и нефропатии

(PDNS), в общем называемый свиным цирковирус-ассоциированным заболеванием (PCVADs), вызывающим значительные экономические проблемы в свиноводстве по всему миру [96; 97]. Как член семейства Circoviridae, рода Circovirus, PCV2 имеет небольшой безоболочечный икосаэдрический капсид с одноцепочечной кольцевой ДНК генома внутри. Геном содержит две основные метки считывания (ORF), где ORF1 кодирует белки репликазы (Rep, Rep') а ORF2 кодирует капсидный белок (Cap) [98; 99]. Капсидный белок может спонтанно собираться в пустые PCV2-подобные частицы. Пустые капсиды с морфологией, сходной с вирусом PCV2 были обнаружены экспериментально [99; 100].

Нам не известны примеры моделирования вируса PVC2 с полноатомным разрешением, но эта тема привлекает заметное внимание в литературе. Одним из ключевых вопросов, решаемых в этих работах, является стабильность капсида вируса в растворе. Например, в предыдущей публикации по полноатомному моделированию вируса спутника табачной мозаики [101] сообщается, что после 10 нс моделирования пустой капсид деформировался. Смоделированная система содержала 932 500 атомов, включая пустой капсид вируса (без РНК) погруженного в воду, где сам капсид составлял около 132000 атомов. Авторы утверждают, что коллапс вируса может быть вызван отсутствием РНК.

Моделирование всех атомов также использовалось для изучения капсида вирус спутника некроза табака [102], где аналогичные изменения в капсиде были связаны с недостатком ионов кальция. В данном случае система состояла примерно из 1 200 000 атомов.

Важность достаточно большой вычислительной области в случае моделирования полноатомного вируса была подтверждена в [103], где капсиды полиовируса в водных растворах оказались стабильными в отличие от предыдущих работ. Моделирование включало 60 молекул сфингозина, 1 884 218 молекул воды, 5310 ионов натрия, 136 ионов калия и 5206 ионов хлора. общее количество атомов, рассматриваемых в этом моделировании, составляло 6 480 236.

В нашей работе изолированный капсид цирковируса был помещенный в большой водный обьем при нормальной температуре и давлении. В таких условиях, согласно экспериментальным наблюдениям [99; 100] известно, что структура вируса остается стабильной. Детальная кристаллографическая структура белковой субъединицы капсида была приведена в [74] с атомистическим разрешением (запись PDB 3R0R). У нас структура вирусного капсида была взята из последнего эксперимента, для построения начальной структуры вируса без учета ДНК внутри (Рисунок 1.19). При погружении в воду, форма вируса будет меняться, чтобы достичь минимума свободной энергия системы, а затем скорректированная форма капсида должна в дальнейшем сохраняется при моделировании, как это наблюдается в эксперименте. Поэтому при атомистическом моделировании любое несоответствие результатов начальному состоянию следует отнести к неточностям численного метода, которые происходят из урезания расчетной области и аппроксимирования обьема периодическими граничными условиями.

Полноатомное моделирование MD было выполнено с явной моделью воды с помощью GROMACS. Для моделирования была использована рентгеновская структура [74] из базы Protein Data Bank (3R0R), и последовательность N-конца, которая не разрешена рентгенологически, вероятно из-за ее подвижности или неупорядоченности. Поэтому структура недостающего N-конца

была реконструирована с использованием гомологического метода MODELLER [104], а затем 60 белков были собраны в цельный капсид, используя Viperdb [105].

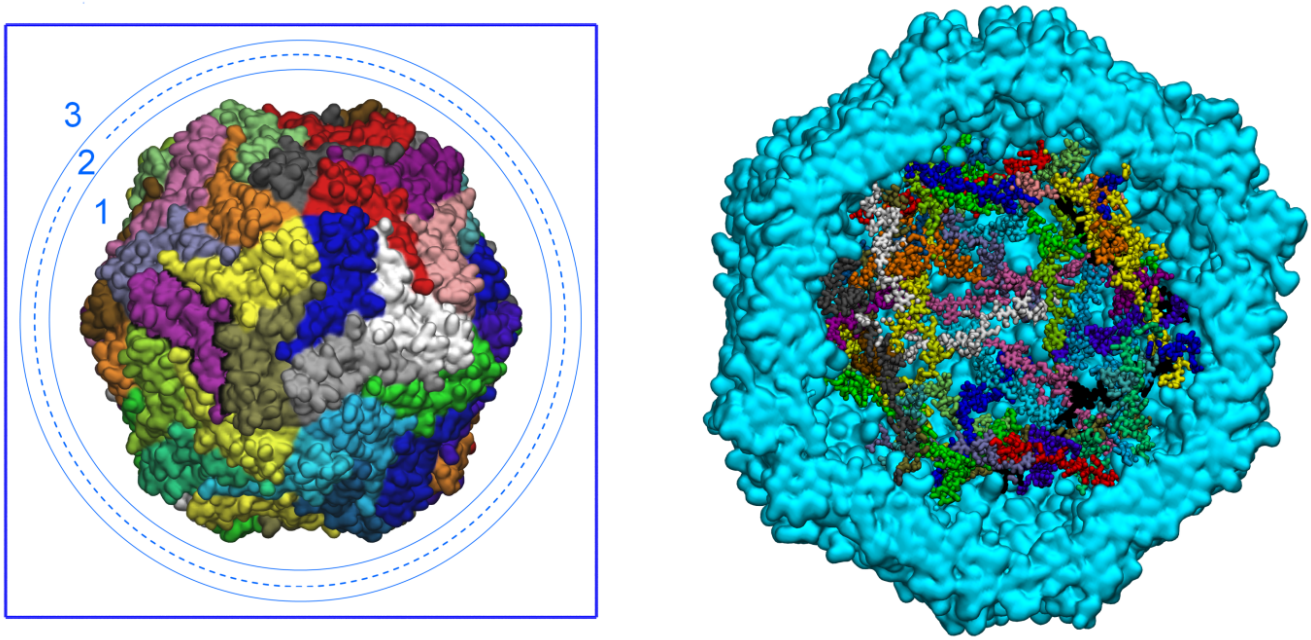


Рисунок 1.19 — Исходное строение капсида: слева – общий вид с границами области моделирования и разделение системы на области: 1 — чистая MD, 2 — гибридная область, 3 — чистая гидродинамика; пунктирная линия показывает подразделение гибридной области для термостатирования; справа – срез, показывающий «хвосты» (система Б); отдельные белки и хвосты показаны разным цветом.

Былио подготовлены две исходные структуры капсида: первая была без N-конца (Рисунок 1.19 слева) и вторая – с N-концом (Рисунок 1.19 справа) (будем называть их системами «А» и «В» соответственно).

Капсид заряжен сильно положительно (360 e), а хвосты заряжены отрицательно (-3 e на хвост). Система разделена капсидом на две части: полость и внешний раствор, что ведет к необходимости правильной нейтрализации заряда всей системы. Для этого было оценено радиальное распределение заряда капсида путем разделения капсида на концентрические сферические слои с центрами, расположенными в центр масс (ЦМ). Увеличивающийся заряд слоев строился в зависимости от радиуса наибольшего слоя, Рисунок 1.20. Как видно, интегральный заряд капсида увеличивается с расстоянием до ≈7,8 нм и затем уменьшается. Этот факт указывает на то, что капсид сильно поляризован с внутренней поверхностью имеющей большой положительный заряд, и отрицательно заряженной внешней поверхности. Поэтому для правильной нейтрализации капсида каждая поверхность была нейтрализована отдельно. Количество ионов Cl^-, равное высоте пика на графике заряда, помещали внутрь капсида, а к внешнему раствору добавляли ионы Na^+ в количестве, равном разнице между общим зарядом капсида и зарядом внутренней поверхности. Эти значения перечислены в таблице 4. Наконец, дополнительные 1720 ионов Na^+ и Cl^- были случайным образом распределены по ячейке, что соответствовало физиологическому раствору (0,9% мас. NaCl). Подчеркнем, что суммарный заряд обеих ячеек моделирования был равен нулю.

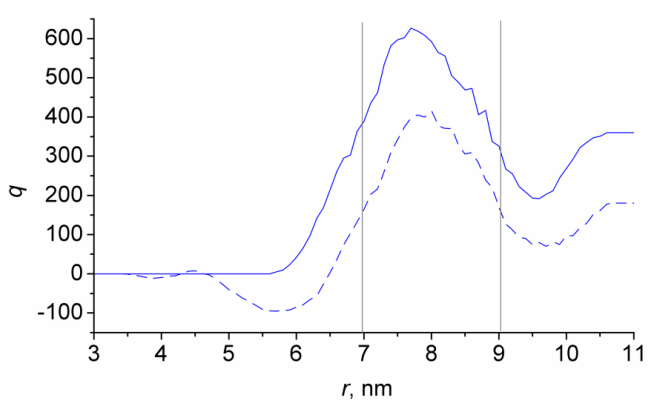


Рисунок 1.20 — Зависимость интегрального заряда капсидных слоев от радиуса наибольшего слоя; сплошная кривая соответствует системе А, штриховая кривая соответствует системе Б; вертикальные линии обозначают примерные границы стенки капсида.

Таблица 4 — Количество ионов, добавленных к капсиду для нейтрализации заряда.

capsid	added ions	
	внутри	снаружи
с хвостами	407 Cl^-	227 Na^+
без хвостов	606 Cl^-	246 Na^+

Чистое моделирование MD было выполнено с использованием силового поля AMBER03 с моделью воды TIP3P. Общее количество атомов в системе составило 1 898 573 для системы А и 1 897 998 для системы В. Смоделированные системы содержали только белковый капсид, воду и ионы без нуклеиновых кислот.

После минимизации энергии капсид был зафиксирован ограничениями координат и расчет проводился при T =200 K за которым следовало увеличение температуры до 300 K. Через 1 нс этой процедуры ограничения были отключены, и было выполнено MD-моделирование системы при 300 K в течение 10 нс. Параметры моделирования были следующими: PME обработка электростатических сил, обрезание Ван-дер-Ваальсовых взаимодействий на 1.0 нм, термостат с масштабированием скорости с постоянной 0.1 пс.

Расчет потоков воды и ионов через стенку капсида

Был использован следующий алгоритм для расчета количества ионов и молекулы воды внутри капсида. Сначала капсид был разделен на 20x20 секторов с вершинами в центре масс капсида. Затем в каждом секторе было рассчитано расстояние от каждого атома до ЦМ, и положение границы между полостью и внешним раствором определялось как среднее расстояний от ЦМ до атома капсида:

$$r_{border} = \frac{1}{N}\sum_{i=1}^{N} r_i.$$

Такой подход делает значение r_{border} устойчивым к колебаниям положения отдельных атомов. Атомы хвоста не образуют стенку капсида и, таким образом, они не учитывались. Считалось, что атом находится внутри капсида, если его расстояние до ЦМ было меньше, чем r_{border}. Используя этот критерий, в каждом секторе были вычислены индексы молекул воды и ионов, находящиеся внутри капсида, и их число просуммировано.

Для оценки потоков между полостью и внешним раствором сравнивали индексы молекул воды и ионов, находящихся внутри капсида между двумя моментами времени. Молекулы, находящиеся внутри капсида в первый (второй) момент и снаружи во второй (первый) считались вышедшими (вошедшими); и сумма вышедших и вошедших молекул давала суммарный поток через стенку капсида в интервале времени между первым и вторым моментами времени.

1.4 Особенности компьютерной реализации гибридного моделирования

1.4.1 Реализация гидродинамики двухмерной модели воды «Mercedes Benz»

Моделирование FH

Модифицированные уравнения флуктуационной гидродинамики (1.20) и (1.25) решаются численно для плотности и импульса, как указано в уравнениях (Б.28) и (Б.29) приложения Б.3. Дифференциальные уравнения в частных производных для флуктуаций (Б.30)-(Б.33) дискретизуются с помощью конечно-разностной схемы Central Leapfrog для членов адвекции в левой части и конечных разностей в правой. Общий порядок приближения равен двум в пространстве и во времени. Для усиления численной устойчивости, используется шахматная схема Central Leapfrog путем введения отдельных переменных для центров ячеек и граней ячеек вместе с малодиссипативной коррекцией нелинейного потока. Подробная информация о нелинейной схеме Central Leapfrog для уравнений переноса, а также ее реализация для классической флуктуационной гидродинамики Ландау-Лифшица дана в [106; 107].

Связь представлений

Для реализации предлагаемого метода проводилось MD моделирование с использованием уравнений (1.19) и (1.24) и моделирование LL-FH с использованием уравнений (1.20) и (1.25). Каждое моделирование требует информацию как из MD, так и из LL-FH, поэтому необходимо обновление/обмен усреднённых по времени плотностей $\sum_p \rho_p$ и импульсов $\sum_p \rho_p u_{ip}$ MD «жид-

кости» с плотностью $\tilde{\rho}$ и импульсом $\tilde{\rho}\tilde{u}_i$ «жидкости» FH. Важным критерием является то, что шаг по времени в MD и LL-FH соответствуют разнице во временных масштабах этих представлений. Здесь было использовано соотношение между временными шагами FH и MD $dt_{FH}/dt_{MD} = 10$, обновление/обмен усредненных по значениям ячеек проводилось на каждой 10-ой итерации MD.

Параметры связи были выбраны $\alpha,\beta = 2$, значения выбиралось таким образом, чтобы сила связи была достаточно слабой, чтобы избежать нефизичного поведения и в то же время достаточно сильной, чтобы связь имела эффект. Кроме того, эти параметры определяют числовую устойчивость симуляции. Значения s не имеют особого значения, были использованы $0.1/0.5/0.8$ для демонстрации влияния на концентрацию фазы FH в смеси фаз FH и MD.

1.4.2 Детали численного решения уравнений гидродинамики сплошной среды и связь с MD частью решения: аргон и вода

Уравнения LL-FH решаются в консервативной форме (1.37) с двухуровневой модификацией схемы Central Leapfrog из [44]. В модифицированной схеме Central Leapfrog используется малая поправка на диссипативный нелинейный поток для стабильности. Он прост в реализации но, несмотря на это, корректно предсказывает правильное значение тепловых флуктуаций на уровне наиболее точных трехэтапных методов Рунге-Кутты, такими как описанны в [43; 108]. Однако по сравнению с последней, вычислительные затраты одноступенчатой схемы Central Leapfrog примерно в 3 раза меньше.

Для решения уравнений LL-FH конкретное уравнение состояния (EOS) важно для связи уравнений континуума с решением MD [44; 9; 10]. Однако, в отличие от континуального домена LL-FH, EOS для домена MD представляет собой результат моделирования, а не соотношение, предписанное в качестве входной модели. Поэтому, чтобы обеспечить одинаковое поведение в обоих доменах, было использовано отдельное MD моделирование для определения EOS исследуемой жидкости. В нашей работе используется метод определения EOS из [44], который состоит из следующих шагов: 1) выполняется несколько полноатомных MD моделирований с разной средней плотностью; 2) в них давление в системе рассчитывается по Ирвингу и Кирквуду [109]; 3) выполняется полиномиальная аппроксимация полученной кривой зависимости давления от плотности; 4) полиномиальная подгонка подставляется в тензор напряжений Рейнольдса уравнений LL-FH.

Помимо EOS, есть и другие важные параметры, которые необходимо учитывать в области LL-FH, а именно значения сдвиговой и объемной вязкости. Как и в случае с EOS, вязкость жидкости MD является результатом моделирования, а не вводится как параметр. Следовательно, значение полученное из моделирования MD используется в качестве входного параметра для LL-FH домена. Расчет коэффициентов вязкости воды требует особого внимания. Действительно, как известно из литературы, вязкость вычисляемая с использованием моделей воды не совпадает с экспериментальным значением [110—114]. Следовательно, была использована вязкость полученная для конкретных моделей воды MD, а не экспериментальные значения, как указано в таблице 2.

Значения для сдвиговой и объемной вязкости аргона менее чувствительна к моделированию MD и для этой работы они взяты из [45].

Вычисление новых координат и скоростей атомов в (1.35) состоит из двух этапов: (i) получение усредненных по ячейкам переменных поля $\bar{\rho}$ and $\bar{u}_i$ из решения уравнений LL-FH, а также соответствующие усредненные по ячейке величины $\sum_{p=1,N(t)} \rho_p$ и $\sum_{p=1,N(t)} u_{ip}\rho_p$ от MD частиц, которые также усреднены по времени, чтобы быть совместимыми с гидродинамическими переменными и (ii) реконструкция непрерывно меняющихся распределений переменных поля внутри каждой ячейки LL-FH, которые будут использоваться в уравнениях для частиц MD (1.35).

Для согласования с численным решением уравнений сплошной среды (1.34) и (1.35), которые соответствуют определенному шагу по времени, или усреднение по времени в соответствии со шкалой времени гидродинамики, в 10 раз большей, чем временной шаг MD для текущей гибридной модели, усредненные по ячейкам поля MD $\sum_{p=1,N(t)} \rho_p$ и $\sum_{p=1,N(t)} u_{ip}\rho_p$ также должны быть соответственно усреднены по времени. Хотя это стандартная практика, когда информация о континууме извлекается из MD моделирования, несоответствия возникают, когда в ячейке слишком мало атомов или размер ячеек слишком мал [46]. Основная причина этого в том, что выборка учитывает только координату центра масс каждого атома, которая непосредственно переводится в (один) дискретный индекс ячейки LL-FH. Это имеет смысл с молекулярной точки зрения, поскольку на ядро (а не на электронное облако) приходится почти вся масса атома и, следовательно, большая часть массы будет находиться в одной ячейке LL-FH. Однако, когда исследуются флуктуации для каждой ячейки с использованием такой простой методики, статистика флуктуаций не совпадает с величинами в континууме («BLOB» фильтр) [46; 115]. Простой метод, позволяющий получить правильные величины для континуума, заключается в использовании так называемых методов отображения [46; 115]. В таком методе, каждый атом или молекула берется как кубический объем центр которого соответствует центру масс атома или молекулы и имеет соответствующую длину стороны d, которую можно настроить таким образом, чтобы вычисленные величины совпадали с континуальными. Однако, как поясняется в [115], для различных величин континуума (масса и импульс) должны использоваться разные размеры таких объемов. Кроме того, масштабирование и оптимальный размер обьема также зависит от типа атома или молекулы, например аргон или вода [115].

Использовался метод для аргона и воды SPC/E, описанный в [46]. Для каждой молекулы вычисляется центр масс и скорость центра масс. Далее вычисляется доля C каждого кубического обьема размером d в ячейках LL-FH равной рассчитаной, где центр обьема совпадает с центром масс молекулы. Это означает, что вклад каждого обьема в плотность и импульс ячейки прямо пропорциональны доле C обьема. Фильтр работает в трехмерном случае и предполагает периодические граничные условия. Размер кубического обьема для аргона принимался равным 0.28 нм, а размер кубического обьема для воды SPC/E – 0.18 нм. Эти значения дали наилучшие результаты в нашем случае и находятся в пределах диапазона значений, указанный в [115]. Однако значения немного отличаются от оптимальных, указанных в [115]. Причина разницы может быть в том, что уравнения, решаемые здесь, – в изотермических условия, т. е. уравнение для энергии не решается в явном виде.

Когда переменные плотности и скорости континуума, включая поля MD, получены как средние по ячейкам, соответствующие непрерывные поля необходимо реконструировать внутри каждой ячейки для решения (1.35). Непрерывность восстановленных полей важна для гидродинамической силы, действующей на частицы MD, чтобы они оставались конечными на границах ячейки LL-FH. Для текущей реализации использовалась трикубическая интерполяция, которая гарантировала, что восстановленные решения не только непрерывны, но и плавны, так что силы, пропорциональные градиенту решения не только ограничены, но и непрерывны на границе ячейки.

1.4.3 Конструирование s-функции

В работе [8] модель односторонней связи была реализована для меняющегося в пространстве разрешения со следующей трехмерной функцией $s = s\,(x,y,z)$ с центром в начале области решения:

$$s\,(x,y,z) = \begin{cases} S_{\min}, & r \leqslant R_{MD}; \\ \frac{r-R_{MD}}{R_{FH}-R_{MD}}\,(S_{\max}-S_{\min})+S_{\min}, & R_{MD} < r < R_{FH}; \\ S_{\max}, & r \geqslant R_{FH}. \end{cases}, \tag{1.96}$$

где $r = (x-L/2)^2 + (y-L/2)^2 + (z-L/2)^2$, L – размер вычислительной ячейки, $x,y,z \in [0,L]$, $S_{\min} = 0$, $S_{\max} = 1$ с центром в центре зоны MD ($r = 0$).

Эта реализация была основана на следующем алгоритме:

- Задаем s-функцию, которая разделяет пространство между MD и гидродинамической областями.
- для каждого положения частицы находим соответствующий контрольный объем, интерполируем гидродинамические поля, полученные в результате решения (1.30)-(1.32), вычисляем значение s-функции в точке положения частицы и интегрируем уравнения (1.35) за один временной шаг.
- Обновляем положения всех частиц и повторяем.

Предположение о стационарности s-функции не обязательно, так как «двухфазные» уравнения (1.30)-(1.32) были сформулированы с учетом пространственно-временной переменной s-функции, и вывод уравнений (1.35) не требует стационарности s-функции. Более того, s-функция в принципе может быть и не односвязной: можно задать и динамически отслеживать более одной интересующей зоны MD.

Таким образом, следующая модификация исходного алгоритма адаптивно отслеживает интересующую атомистическую область:

- Указываем интересующий обьект, например, центр масс сложной молекулы, которую необходимо рассматривать с полноатомным разрешением.
- Фиксируем центр s-функции на этом обьекте, чтобы координаты центра обновлялись на каждом временном шаге.

– Для каждого положения частицы, в определенный момент времени, находим контрольный объем, которому он принадлежит, интерполируем гидродинамические поля из (1.30)-(1.32) в его положение, рассчитываем пространственно-временную s-функция в этом месте и интегрируем уравнения (1.35) за один временной шаг, затем обновляем положения всех частиц, перемещаем центр s-функции в соответствии с кинематикой обьекта, и повторяем.

Текущая модель односторонней связи с привязкой центра s-функции к центру области интереса MD была реализована в пакете с открытым исходным кодом GROMACS. GROMACS — универсальный код C/C++, позволяющий проведение высокопроизводительного моделирования, то есть решение ньютоновских уравнений движения для систем до нескольких миллионов частиц. Наша реализация связанной модели MD/FH состоит из двух основных частей: решатель MD (предоставляется GROMACS) и решатель LL-FH (независимая функция C). Уравнения флуктуационной гидродинамики решаются на эйлеровой равномерной ортогональной сетке с использованием консервативной малодиссипативной конечно-разностной схемы, разработанной в [106]. Решатель MD основан на алгоритме Верле, позволяющим легко включать дополнительное члены в уравнения движения для координат и скоростей каждой частицы на каждом временном шаге в соответствии с (1.35). Дополнительные члены соответствуют полям, полученным из решения LL-FH, определенного на сетке Эйлера с шагом, определенным в соответствии с размером контрольного объема. Чтобы получить значения полей LL-FH в положениях частиц между точками сетки используется интерполяция. Реализация поддерживает внутренний параллелизм GROMACS, что означает, что моделирование с гибридной моделью MD/FH может быть выполнено с использованием CUDA, MPI или GROMACS Thread MPI протоколов.

1.4.4 Диаланин в воде

Диффузию пептида в воде моделировали в соответствии с гибридной моделью [8]. Силовое поле GROMOS96 53a6 [116] использовалось для пептида, а модель SPC/E использовали для воды. Условия для расчета молекулярной динамики были следующим: T = 298 K (поддерживался термостатом Нозе-Гувера с постоянной времени 0.5 пс), ρ =993.3 г/см^3, шаг по времени – 1 фс, длина обрезания электростатики составляла 1.0 нм, диэлектрическая проницаемость – 78, длина обрезания потенциала ван Дер-Ваальса составляла 1.0 нм. Вся вычислительная область в том числе гидродинамическая область соответствовала кубу с ребром 7.2 нм (~12 тысяч молекул воды) с периодическими граничными условиями и расположением чистых зон MD и LL-FH как показано на рисунке 1.21.

Для исследования влияния гидродинамической области (s > 0) на чисто атомистическую область ($s = 0$), рассчитывали коэффициент диффузии для молекулы пептида D. Результат сравнивали с эталонным решением, полученным в полноатомном моделировании, проводимым в той же области с периодическими границами. Для расчета D использовалось соотношение Эйнштейна $MSD\left(t\right) = \left\langle \Delta r\left(t\right)^2 \right\rangle = A + 6Dt$, где A — произвольная константа, а MSD — среднеквадратичное

отклонение. формула справедлива в пределе большой статистической выборки, соответствующей длинным временам t.

Для стационарной функции (1.96), поскольку пептид уходит от центра домена, только относительно короткая часть его траектории проходит внутри «полезной» чисто атомистической области, что приводило к относительно короткому времени расчета доступного для постобработки. Чтобы компенсировать короткое время моделирования, необходима сложная обработка молекулярных траекторий, что неизбежно снижает точность конечного результата [8]. В данной работе центр функции $s = s(x, y, z, t)$ связан с центром масс пептида, так что молекула пептида все время остается внутри атомистической области. Это позволяет увеличить время моделирования на порядок и, таким образом, непосредственно вычислить траекторию пептида без каких-либо промежуточных предположений. В свою очередь, это приводит к получению оценки коэффициента диффузии пептида в воде с гораздо меньшим загрязнением статистической постобработкой.

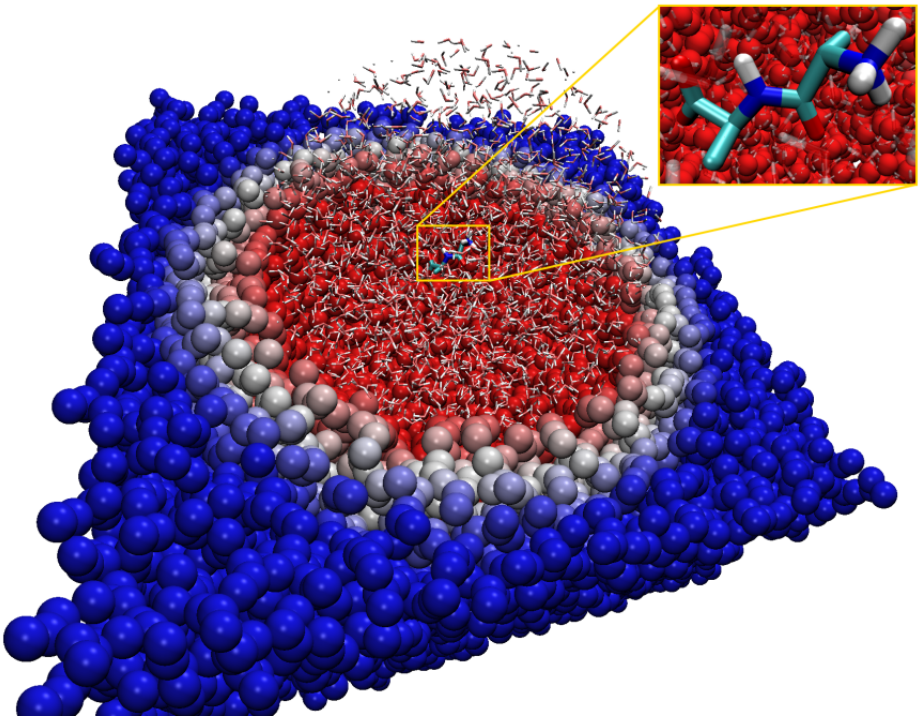


Рисунок 1.21 — Диаланин в гибридной MD/FH воде; цвет частиц постепенно меняется с синего (область FH) на белый (область смешанной MD-FH) и на красный (чистая область MD).

Для диффузии пептидов в воде длительность гибридного моделирования составляла 1 нс. Всего было проведено 50 запусков, начиная с различных начальных конфигурации для лучшего статистического усреднения. Для каждого вычислялось среднеквадратичное отклонение $MSD(t)$, и окончательный результат получался усреднением всех 50 траекторий. На рисунке 1.22 показано сравнение MSD результирующей траектории, полученной гибридным методом, где результат чистого MD также включен.

Для расчета коэффициента диффузии D использовался приблизительно линейный интервал от 100 до 400 пс. Для оценки неопределенности значения исходный интервал в 300 пс был разбит на несколько подинтервалов по 100 пс каждый с перекрытием в 50 пс. Каждый интервал давал свой собственный коэффициент диффузии D_i в соответствии с локальным градиентом и половина разницы между максимумом и минимумом D_i была принята как статистическая погрешность ΔD.

Окончательное значение коэффициента диффузии D оказался равным (0,83±0,08)$\cdot 10^{-5}$см^2/с, что очень хорошо согласуется с эталонным значением из моделирования чистым MD, 0,86$\cdot 10^{-5}$ см^2/с.

Можно отметить, что в отличие от MD моделирования гибридное решение испытывает скачкообразное изменение траектории в начальный момент времени. Этот скачкообразное изменение, скорее всего, связано с более сильными начальными переходными процессами в случае гибридного моделирования из-за того, что расчет начинается с начальных условий, которые труднее сделать полностью согласованными для гибридной модели по сравнению с полноатомной моделью.

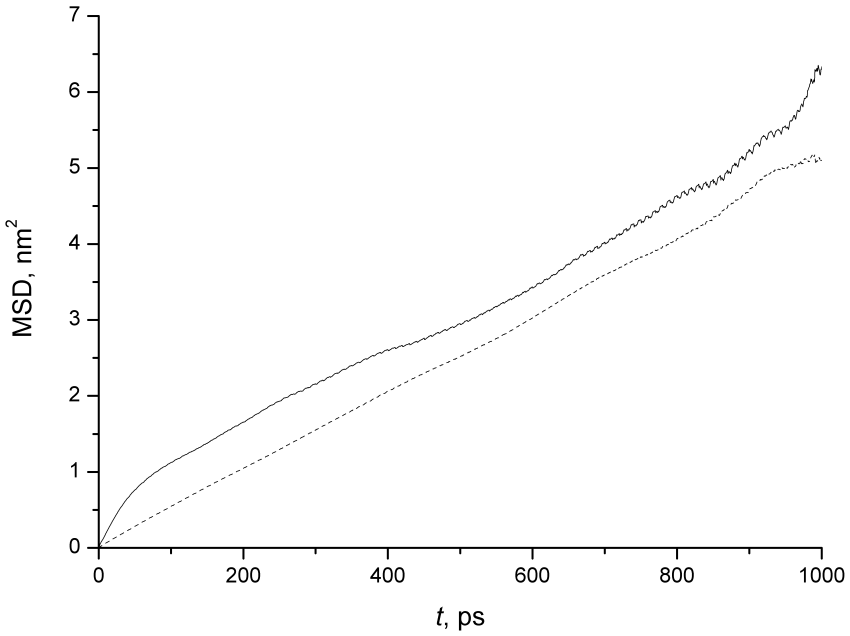


Рисунок 1.22 — Вычисленная функция $MSD(t)$ для задачи гибридного моделирования LL-FH/MD диффузии пептида в воде (сплошная линия) и в чистом MD (штриховая линия).

1.4.5 Свиной цирковирус в воде

При том же размере вычислительного домена, в котором находится капсид вируса, окруженного водой, были применены два типа граничных условий. Первый – стандартные периодические граничные условия для полноатомного моделирования, а второй соответствует гибридной модели LL-FH/MD (1.35).

Капсид имеет диаметр ~18-20 нм и состоит из 126180 атомов, погруженных в кубическую ячейку с водой с длиной стороны 26.68 нм. Для модели LL-FH/MD чистая область MD была связана с центром масс капсида, который мог свободно перемещаться внутри вычислительной области, аналогично моделированию диаланина. Внутренняя MD область диаметром 22.68 нм также включала слой молекул воды, взаимодействующий с капсидом. Общее число атомов в зоне MD

составило примерно 590000, и 1 837 905 частиц во всей расчетной области, в том числе в промежуточной зоне LL-FH/MD и области FH (Рисунок 1.23).

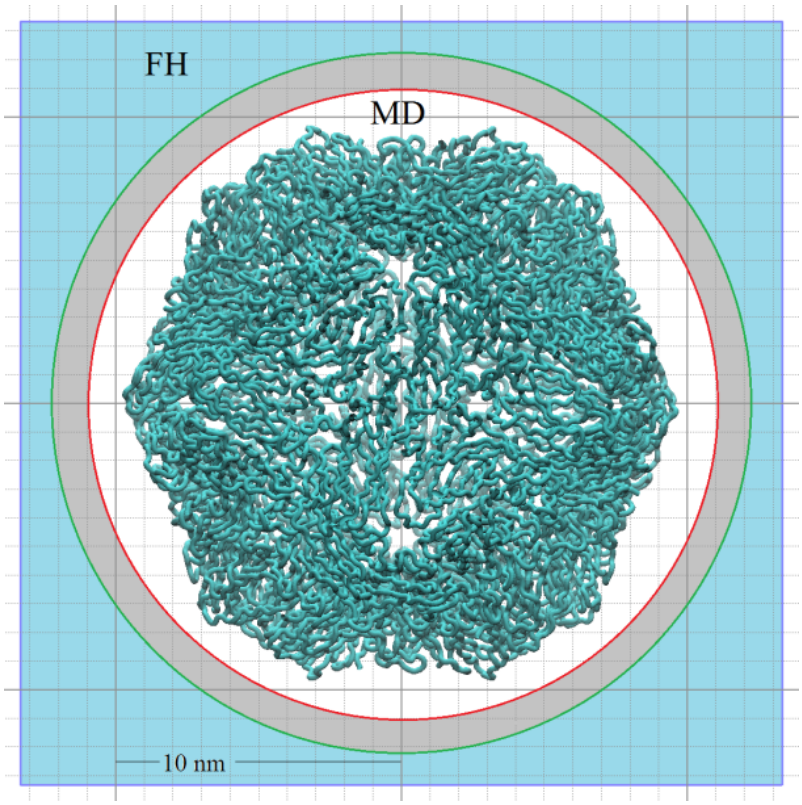


Рисунок 1.23 — Капсид PCV2 в гибридной воде LL-FH/MD; капсид находится внутри области чистой MD (белая), которая постепенно изменяется через промежуточную LL-FH/MD область (серый) в область FH (синий); мелкие клетки имеют размер 1 нм что примерно соответствует размеру ячейки FH.

Все MD расчеты проводились с силовым полем GROMOS96 53a6 [116] и SPC модели воды. Гибридное моделирование LL-FH/MD было выполнено с помощью термостата Nose-Hoover (T =300K), электростатикой поля реакции с длиной обрезания 1.0 нм и диэлектрической проницаемостью 78, длина обрезания потенциал Ван-дер-Ваальса была равна 1.0 нм. Временной шаг MD составлял 1 фс, а временной шаг для решателя FH был в 10 раз больше (10 фс). Общее время моделирования составило ~2 нс (220000 шагов MD в день на 256 ядрах MPI).

Для полноатомного моделирования использовалась та же атомистическая модель вируса внутри той же ячейки, но с явным атомистическим моделированием всей окружающей воды и периодическими граничными условиями на границах ячейки. Общее число атомов в полноатомной модели равнялось числу частиц в модели LL-FH/MD (1837905).

Параметры моделирования MD приведены в таблице 5.

Моделирование молекулярной динамики проводили при постоянном объеме с использованием значения параметра vdwtype равного длине обрезания, электростатика Particle-Mesh Ewald, v-rescale термостат и схема обрезания Верле. Для поноатомного моделирования использовались периодические граничные условия во всех трех направлениях. Расстояние между капсидом и его периодическим изображением было выбрано достаточно большим в соответствии с [117], где было

Таблица 5 — Параметры моделирования MD: L – размер кубической ячейки, t – время работы, Δt – временной шаг интегрирования, τ определяет постоянную времени для температурной связи, а T - целевая температура для температурной связи, R_{vdw} и R_{coul} расстояния обрезания (для электростатики Леннарда-Джонса и Particle-Mesh Ewald для кулоновских соответственно).

Параметр	Величина	Единицы
L	26.68	nm
t	2000	ps
Δt	0.002	ps
τ	0.1	ps
T	300	K
R_{vdw}	1.0	nm
R_{coul}	1.0	nm

предложено использовать как минимум 2 нм слоев воды для обеспечения пренебрежимо малого взаимодействия белок-белок. При подготовке начальных условий после минимизации энергии MD расчет проводился с ограничениями координат капсида с теми же параметры MD, которые использовались в основных расчетах (Таблица 4). Через 1 нс ограничения были сняты, а прогон уравновешивания производился до тех пор, пока энергия системы не стала постоянной.

1.4.6 Высокопроизводительное молекулярное моделирование: MDGRAPE

Суперкомпьютеры общего назначения и специализированные суперкомпьютеры

В многомасштабном моделировании скорость вычислений ограничена скоростью вычислений для частей с мелким масштабом из-за больших различий во времени и пространстве. Базовый временной шаг поноатомного моделирования MD составляет 1-2 фс, в то время как в гидродинамике – порядка сотен пикосекунд. Таким образом, необходимо заполнить пробел в несколько порядков во временной шкале. Этот пробел был заполнен с помощью HPC (High Performance Computing).

Наиболее популярны обычные суперкомпьютеры на базе CPU, такие как уже списанный K-компьютер или современный Fugaku, эффективные для моделирования больших молекулярных систем, состоящих из миллионов атомов. В этом случае скорости вычислений хорошо растет с ростом числа процессоров, Рисунок 1.24. Напротив, специализированные компьютеры позволяют более быстрые расчеты для небольших систем размером в несколько тысяч атомов, Рисунок 1.1. Наши коллеги разработали специализированные компьютерные комплексы для MD

моделирования, которые называются MDGRAPE [118], Рисунок 1.25. MDGRAPE компьютеры – это ускорители для расчетов нековалентных связей.

В последнее время исследовательским центром D. E. Shaw был разработана машина ANTON, которая также является специализированной для MD моделирования [119]. Эта группа интегрировала ускорители сил нековалентных связей, ядра универсального процессора, память и сетевые интерфейсы в одном кристалле (SoC). Такая интеграция минимизирует задержки между этими вычислительными элементами. В результате это обеспечивает 100 кратное или более увеличение производительности для небольших систем, по сравнению с обычными кластерами, и высокая производительность всей системы может быть достигнута для систем размером до 10K атомов. Таким образом, это позволяет проводить моделирование 100 пикосекунд в секунду для подобных систем.

В этом случае особенно полезна возможность быстрого моделирования небольших белковых систем а также многомасштабного моделирования. Цель настоящего исследования - разработка машины четвертого поколения, MDGRAPE-4, Рисунок 1.26. MDGRAPE-4 ориентирован на быстрое моделирование малых системы. Он также основан на технологии SoC, аналогичной ANTON. Целевая производительность составляет 100 пс моделирования в секунду для системы со 100 тыс. атомов. Программное обеспечение основано на GROMACS и содержит набор библиотек C для доступа к ядру GP.

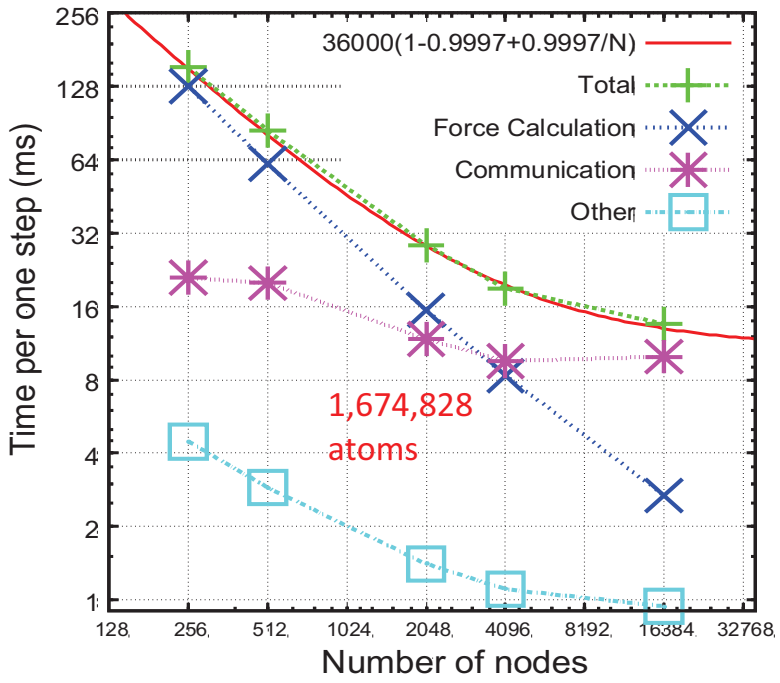


Рисунок 1.24 — Масштабирование с количеством вычислительных узлов для моделирования MD с помощью K-компьютера в RIKEN.

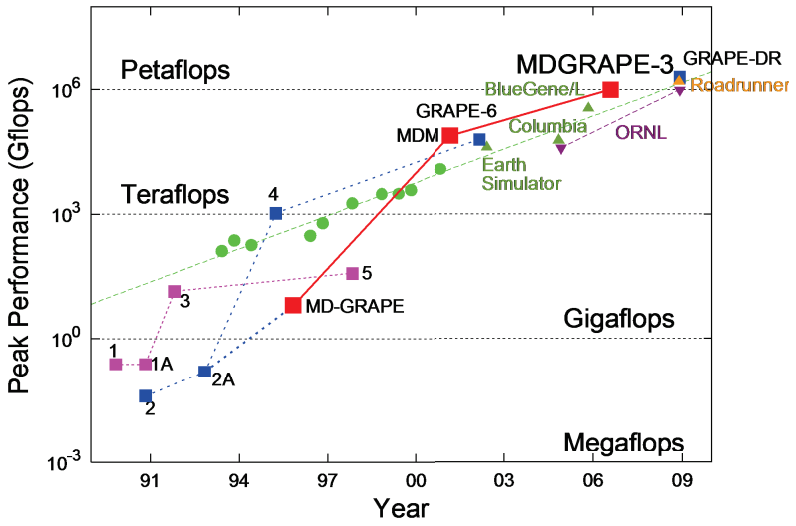


Рисунок 1.25 — Разработка ускорителей семейства GRAPE (Gordon Bell Призы, выигранные в 1995, 1996, 1999, 2000 (двойной), 2001, 2003 и 2006).

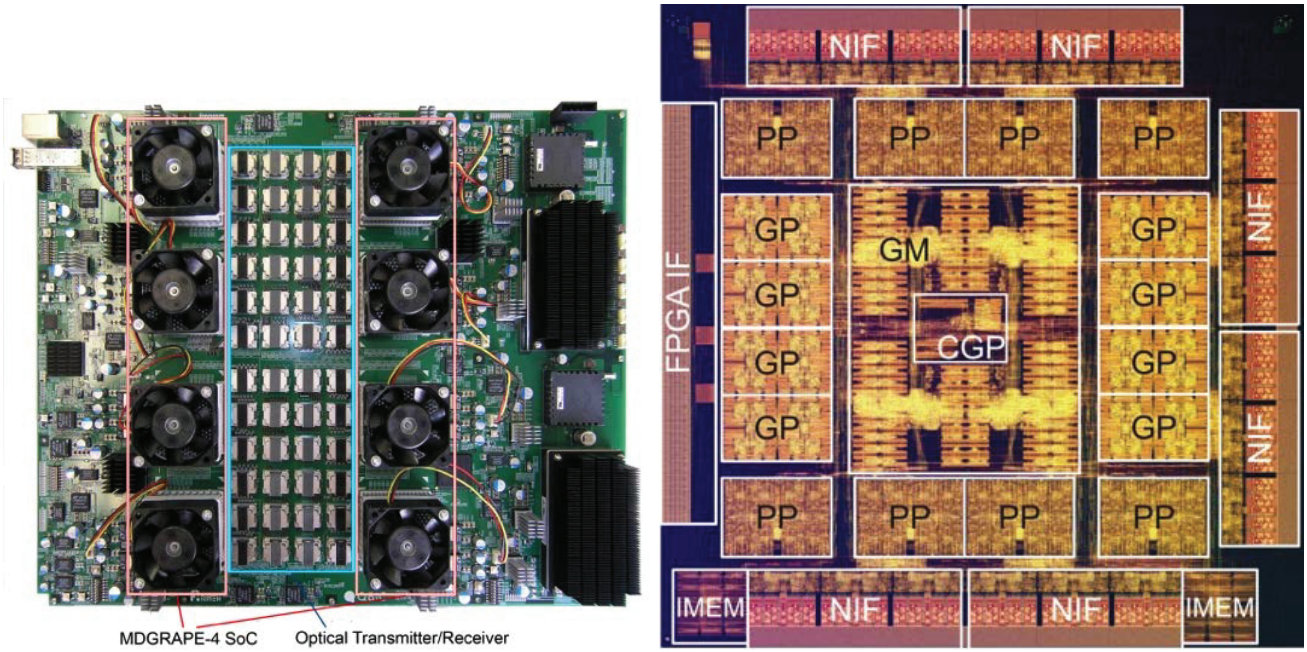


Рисунок 1.26 — Системная плата MDGRAPE-4 и система на кристалле. На плате расположено 8 SoC и 48 оптических передатчиков/приемников для работы в сети. SoC имеет 8 модулей расчета нековалетных сил (PP), 8 модулей процессоров общего назначения (GP), каждый из которых имеет 8 ядер, память (GM) и сетевые интерфейсы (NIF).

Вычисления на основе графического процессора

Другая ветвь реализации HPC нашей инфраструктуры использует расчеты на основе GPU. По природе стохастического моделирования, важно статистическое поведение уравнений LL-NS а не их мгновенное решение. Для статистической сходимости на низких частотах, характерных для гидродинамических флуктуаций, требуется достаточно длительное время моделирования. Напри-

мер, решение задачи стохастической диффузии сходится как $\frac{1}{\sqrt{N_t}}$, где N_t — количество временных шагов моделирования, которое обычно имеет порядок в несколько миллионов, чтобы уменьшить погрешность флуктуаций до нескольких процентов.

Для трехмерных задач это приводит к значительным вычислительным затратам если расчет производится последовательно на основе одного процессора. С другой стороны, флуктуационная гидродинамика уравнений LL-NS хорошо подходит для реализации в CUDA от NVidia. Он позволяет разгрузить вычисления на графический процессор (GPU) с выполнением массивно-параллельных вычислений на основе сотен вычислительных потоков.

Основная идея вычислений на GPU, если использовать терминологию CUDA, основана на том факте, что определенные фрагменты кода (ядра) могут выполняться одновременно в нескольких блоках, каждый из которых содержит сотни потоков. Ключом к оптимизации является ограничение запроса на одну и ту же область памяти, что приводит к высокой задержке по времени. В случае уравнений LL-NS каждый блок ячеек (например, 16x16) может оцениваться независимо, где только один слой граничных ячеек должен запрашивать пространство памяти совместно используемое с другими блоками. Это похоже на простую декомпозицию блочного домена. Помимо разного подхода к доступу к вычислительным массивам, сам (C++) код требует немного изменений.

Графический процессор, используемый в нашем моделировании жидкого аргона в состоянии равновесия — NVidia Tesla C2075. Рисунок 1.27 показывает количество времени, требуемого для вычисления одного цикла двухуровневого алгоритма Central Leapfrog для трехмерных уравнений LL-NS в зависимости от числа ячеек.

Рабочая нагрузка возрастает линейно с объемом ячеек, т. е. имеет линейное масштабирование при достижении баланса нагрузки. Также интересно сравнить это масштабирование с нагрузкой на основе одного процессора. Рисунок 1.27(b) показывает, как время вычислений графического процессора сравнимо с временем вычислений на основе одного ядра Intel Xeon E5-2609. На графике сравнения процессора и графического процессора есть два «режима».

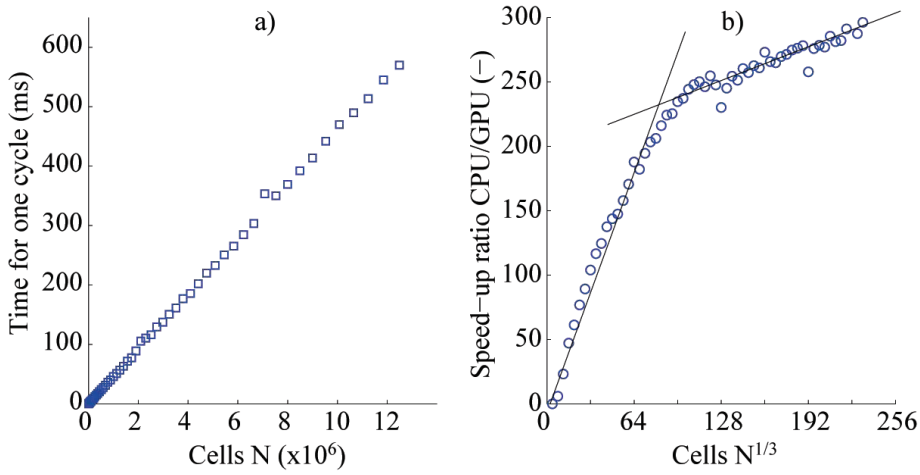


Рисунок 1.27 — a) Количество времени, которое занимает один цикл GPU как функция количества ячеек (N); b) отношение времени, необходимого для выполнения одного цикл на процессоре по сравнению со временем на графическом процессоре как функция количества ячеек в одном измерении кубической области.

В первой части кривой 1.27 соответствующей относительно небольшим вычислениям, кривая круто поднимается вверх, что свидетельствует о резком росте эффективности, когда только часть потоков графического процессора используется полностью и их конкуренция за общую память низка. После того, как все потоки GPU становятся полностью заняты, скорость роста кривой стагнирует. Последнее, вероятно, вызвано тем, что система достигает предела доступной оперативной памяти (в случае GPU выделено 5 ГБ по сравнению с CPU, совместно используемым с системой 6 ГБ). Тем не менее, как следует из этого графика, GPU может выполнять вычисления до 300 раз быстрее по сравнению с одним ядром CPU, или в лучшем случае примерно в 75 раз быстрее по сравнению с четырехъядерным процессором CPU.

Из-за значительного ускорения за счет реализации графического процессора LL-NS моделирование флуктуационной гидродинамики заняло несколько часов, а не недель, на сетках до 10 миллионов вычислительных ячеек.

1.5 Результаты

1.5.1 Вода в модели MB

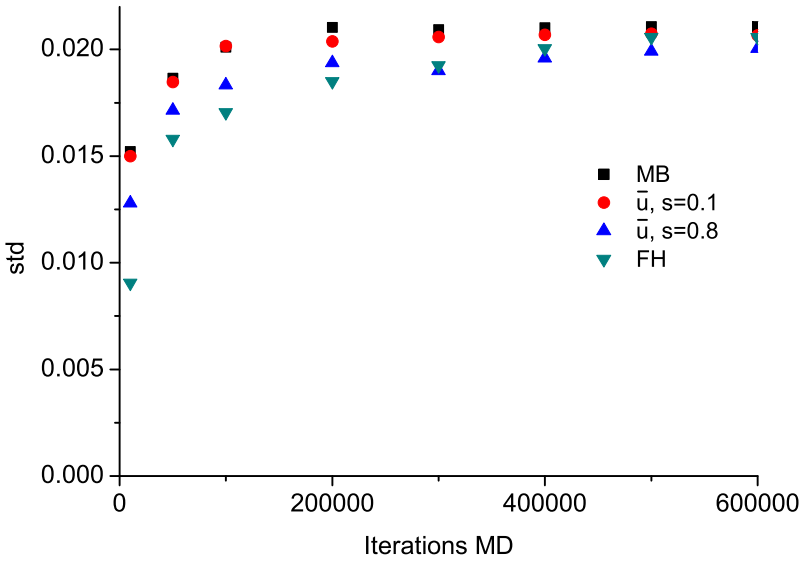


Рисунок 1.28 — Стандартные отклонения средних скоростей (компонента x) в чистом моделировании MB, моделировании LL-FH и связанных системах.

Стандартные отклонения (*std*) для колебаний средней скорости $\tilde{u}$ для связанных систем и чистых систем MB, LL-FH, Рисунок 1.28, подтверждает правильность нашей схемы так как все они схожи друг с другом и с правильными установленными значениями. *std* для $\tilde{u}_x$ чистого MB и $s = 0.1$ на рисунке 1.28 выше, чем фаза FH и $s = 0.8$. Это связано с предположением, что молекула воды MB является точечной частицей, что не совсем так, так как молекула занимает некоторую площадь «эффективного» объема. Это приводит к пульсации плотности и скорости, когда частицы

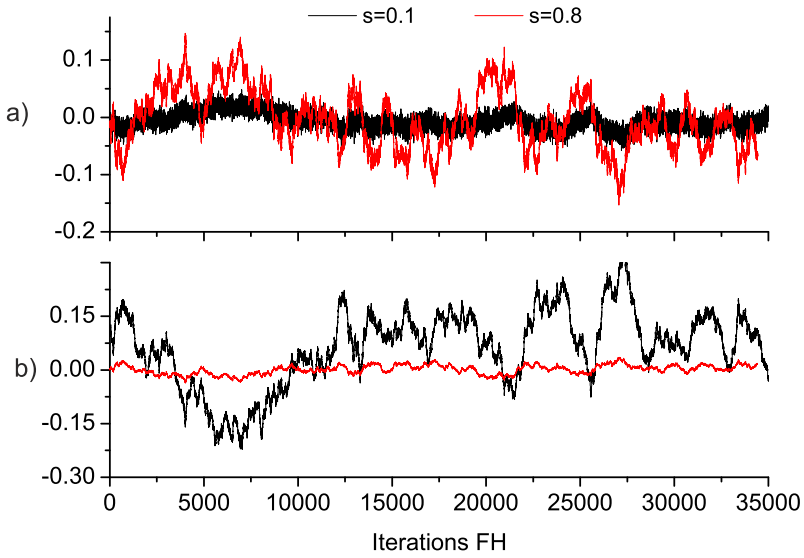


Рисунок 1.29 — Разность скоростей в смеси и в отдельных фазах a) $\tilde{u}_x - \sum_p u_{xp}$и b) $\tilde{u}_x - u_x$ для $s = 0.1$ и $s = 0.8$.

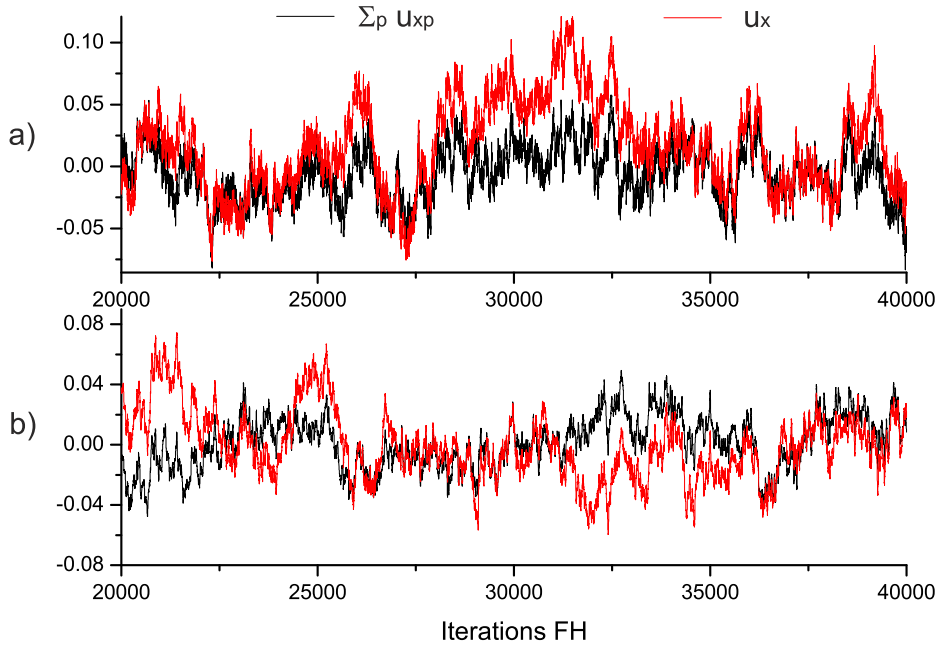


Рисунок 1.30 — Профили компонент скорости x фазы MD $\sum_p u_{xp}$ и фазы FH u_x; a) $s = 0.1$, b) $s = 0.8$.

мигрируют из одной ячейки в другую. Эти пульсации можно устранить, если вклады молекулы в плотность и импульс вычисляются как часть его объема в разных ячейках [120].

Разница между скоростью смеси $\tilde{u}_x$ и MD фаз $\sum_p u_{xp}$, Рисунок 1.29 а также Рисунок 1.30, показывает, что когда $s = 0.1$ разница невелика, это означает, что скорости следуют друг за другом в обоих фазах, тогда как для $s = 0.8$ разница значительно больше и имеет место сложная динамика, означающая в значительной степени независимую эволюцию скорости во времени. Эффект связи, Рисунок 1.29 и Рисунок 1.30, показывает, что $\tilde{u}$ наследует колебания фазы FH, когда $s \to 1$ и MD фазы, когда $s \to 0$.

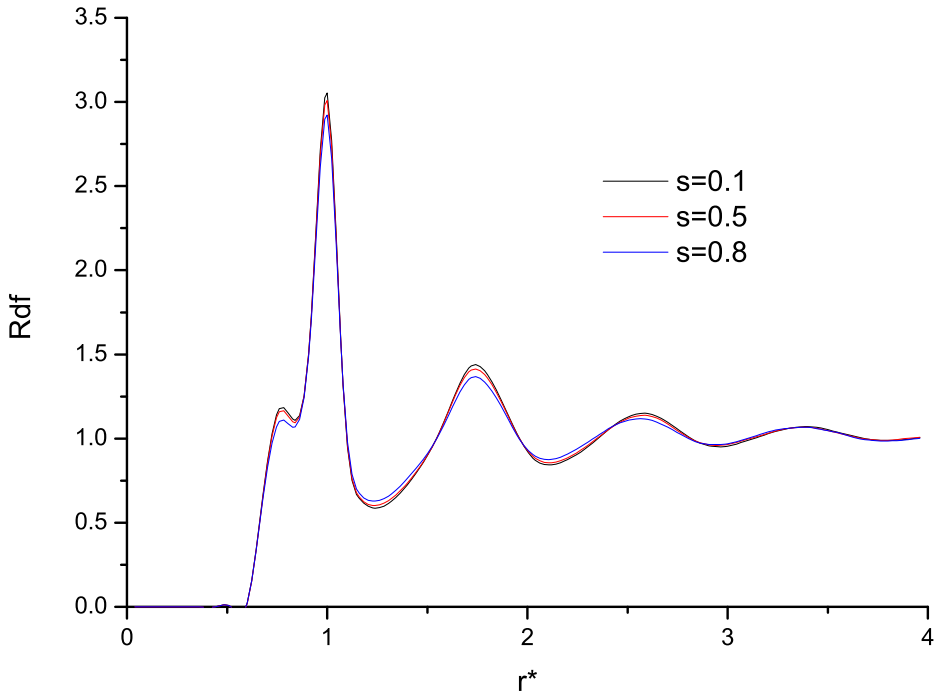


Рисунок 1.31 — Функция радиального распределения для различных значений s.

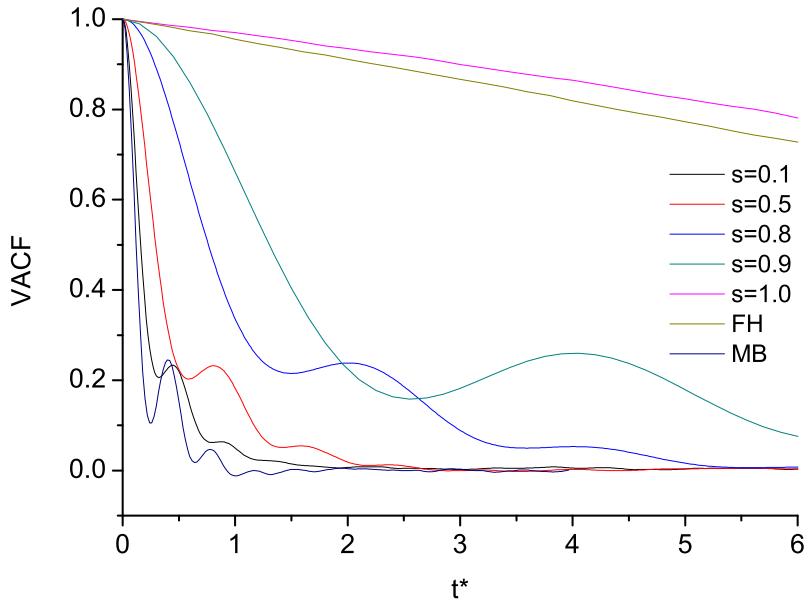


Рисунок 1.32 — Функция автокорреляции поступательной скорости для различных значений s, рассчитанных из скорости частиц.

Функция радиального распределения молекул, Рисунок 1.31, остается такой же, как и для модели MB без связи. Этот означает, что внешняя сила, «связь», не влияет на распределение молекул MB в локальной окрестности.

Были связаны поступательные скорости частиц MB со скоростями 2D LL-FH, оставляя вращательную степень свободы несвязанной. Поскольку в макроскопическом описании компоненты скорости, усредненные по контрольному обьему, вносят вклад в законы сохранения LL-FH. Из-за усреднении требуется учитывать именно скорости центров масс каждого MB диска. Угловая скорость частиц MB используется только для управления изменением масштаба угловой скорости, чтобы поддерживать постоянную температуру частиц MD.

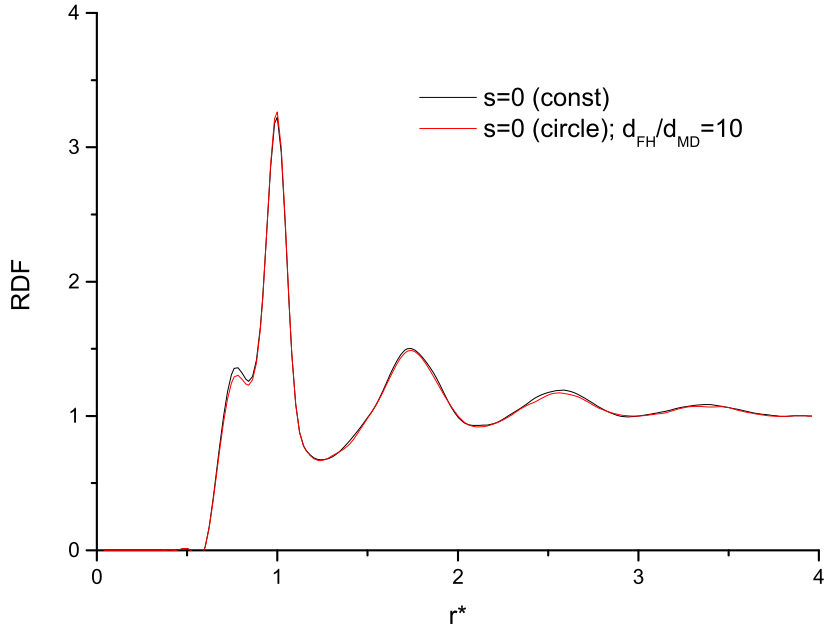


Рисунок 1.33 — Функция радиального распределения MB частицы в центре гибридной области параметра s, где $s = 0$.

Рассчитана автокорреляционная функция поступательной скорости (VACF) используя реальную скорость частицы $\frac{dx}{dt}$, (1.19), показана на рисунке 1.32. Для больших значений s VACF становится более растянутой во времени и в пределах $s \to 0$ и $s \to 1$ она приближается к соответствующим VASF чистого MD и чистого FH.

Можно рассмотреть более общий случай, когда параметр связи s изменяется в пространстве:

$$s(r) = \begin{cases} S_{min}, & r < R_{MD}; \\ \frac{r - R_{MD}}{R_{HD} - R_{MD}}(S_{max} - S_{min}) + S_{min}, & R_{MD} < r < R_{HD}; \\ S_{max}, & r > R_{HD}, \end{cases}$$

где $S_{min} = 0$, $S_{max} = 0.99$, R_{MD} — это 0.4 и R_{HD} — это 0.9 области моделирования, которая формирует профиль круглой формы для параметра связи s, Рисунок 1.6.

Таким образом, в центре $s = 0$ присутствует только чистая MD-фаза, а с краев $s = 0.99$, таким образом, FH является основной фазой. Переменный параметр связи s показывает аналогичные тенденции в отношении величин VACF и RDF. RDF не претерпевает резких изменений независимо от значения s, Рисунок 1.31 и 1.33, и соответствуют чистому MB.

В области, где s велико, VACF вытянуты по оси времени (движение частицы становится «студенистым») и при малых значениях s VACF напоминают чистый MB. В центре круга, где $s = 0$ растяжение также сохраняется 1.34, хотя это не так заметно, как в зоне $s \neq 0$. Это связано с наличием фазы FH на краях области $s = 0$, влияющие на движение частиц MB в центре окружности не смотря на то, что $s = 0$.

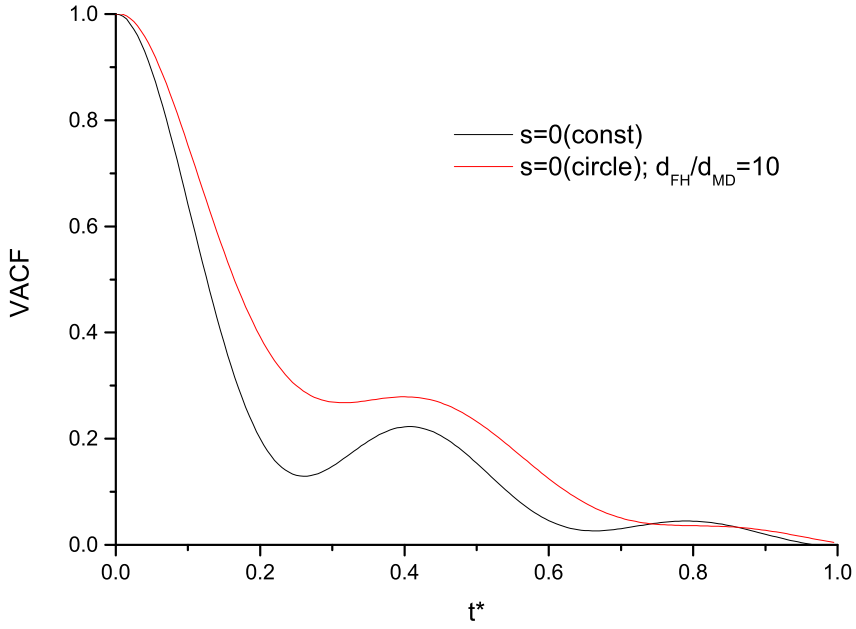


Рисунок 1.34 — Автокорреляционная функция поступательной скорости частиц MB в центре гибридной области s, где $s = 0$.

Из-за этого эффекта необходимо аккуратно учитывать размер круга, а также временной шаг гидродинамической части гибридной модели так, чтобы она постепенно меняла границу зон от MD до FH масштабов.

В целом это показывает, что переход между представлением FH и MD плавный и начинается в области, где FH отсутствует («граница» между фазами MD и FH). Кроме того, если используется круговой профиль параметра s, то частицы MB вблизи границы $s = 1$ могут прилипать к фазе FH и оставлять вакансию в прежнем положении. В районе $s = 1$ MB частицы не взаимодействуют между собой, поэтому они увлекаются фазой FH без сопротивления. Следовательно, когда частицы MB прилипают к фазе FH в совокупности появляются пустые места в области где $s \neq 1$. Хотя через некоторое время частицы MB возвращаются обратно в область где $s \neq 1$. Этот эффект не виден в анализе MD частиц на близких расстояниях (RDF), так как это макромасштабный эффект.

1.5.2 Гибридное моделирование аргона

Постоянная s-функция

Результаты моделирования жидкого аргона в условиях высокого давления представлены первыми. Одна очень важная характеристика для проверки [44; 108] — стандартное отклонение плотности и скорости для LL-FH и MD частей решения. В соответствии с теорией стандартные

отклонения флуктуации скорости и плотности, соответствующие уравнениям (1.35) равны:

$$STD(\bar{\rho}) = c_T^{-1}\sqrt{\rho k_B \frac{T}{V_{cell}}}, \tag{1.97}$$

$$STD(\bar{u}) = \sqrt{k_B \frac{T}{\rho V_{cell}}}, \tag{1.98}$$

где V_{cell} — объем ячейки, T и ρ — температура и плотность, а c_T — изотермическая скорость звука.

На рисунке 1.35 показано стандартное отклонение (STD) колебаний скорости для аргона. Рассматриваемые случаи: чистое MD ($s = 0$) без BLOB фильтра, часть MD решение для двух промежуточных случаев с константами $s = 0.1$ и $s = 0.8$ и раствор LL-FH. Как видно из рисунка, колебания скорости во всех случаях сходятся к одному и тому же значению (~0,011 нм/пс) в течении приблизительно 3 нс. Отсюда также видно, что BLOB фильтр не требуется чтобы получить лучшее согласование колебаний скорости. Теоретическое значение скорости STD для аргона согласно (1.98) составляет 0,0110 нм/пс.

На рисунке 1.36 показано стандартное отклонение флуктуаций плотности аргона для тех же случаев что и для чистого MD с включенным BLOB фильтром. Опять же, во всех случаях, кроме чистого MD без BLOB фильтра, значения сходятся к тому же значению (~10,7 а.е.м./нм^3), которое находится в хорошем согласии с теоретическим значением 10,52 а.е.м./нм^3 для аргона согласно уравнению (1.97). Как и ожидалось, значение флуктуаций плотности без BLOB-фильтра значительно завышены, что наглядно показывает необходимость фильтра для того, чтобы получить правильные значения колебаний плотности.

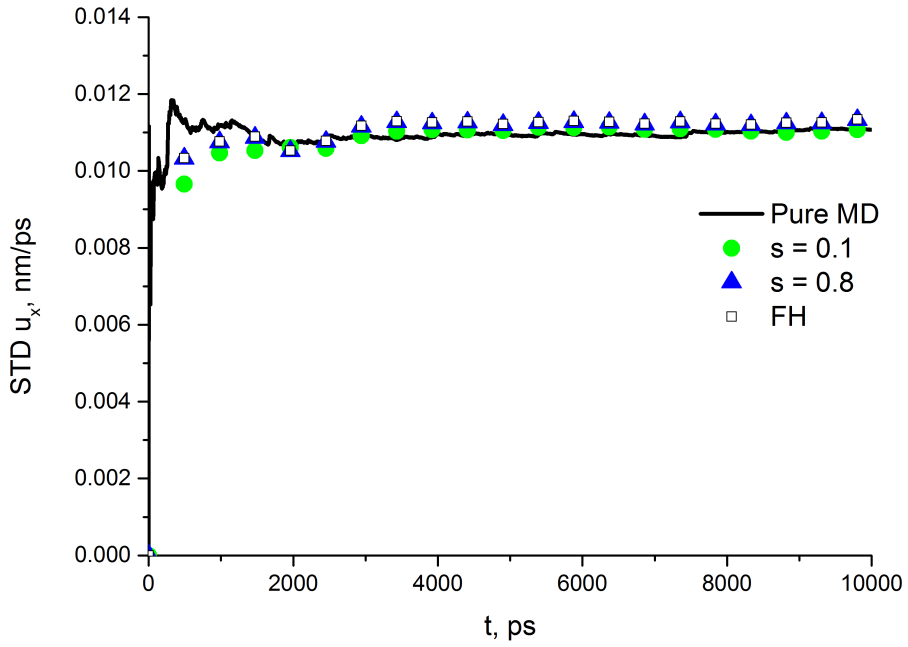


Рисунок 1.35 — Стандартное отклонение флуктуаций скорости для аргона для четырех разных случаев; чистое решение MD ($s = 0$) без BLOB фильтра, MD часть решения для $s = 0.1$ и $s = 0.8$ и решение LL-FH.

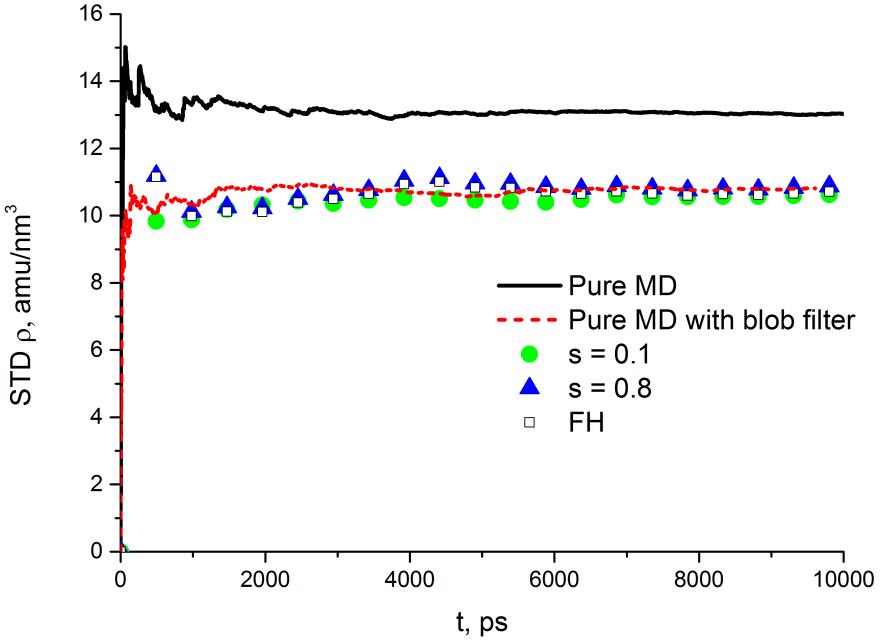


Рисунок 1.36 — Стандартные отклонения флуктуаций плотности аргона для пяти разных случаев; чистое решение MD ($s = 0$) без и с BLOB фильтром, часть решения MD для $s = 0.1$ и $s = 0.8$ и решение LL-FH.

Значение изотермической скорости звука c_T можно получить из соотношения между уравнениями (1.98) и (1.97):

$$c_T = \rho \frac{STD(\bar{u})}{STD(\bar{\rho})}. \tag{1.99}$$

Согласно уравнению (1.99) и результатов моделирования, полученная скорость звука для аргона составляет примерно 0,617 нм/пс. Теоретическое значение можно определить из уравнения состояния [44], которое оценивается как 0,63 что находится в 2-процентном согласии с численной оценкой.

Помимо флуктуаций, необходимо изучить еще две важных характеристики: функцию радиального распределения (RDF) и автокорреляционную функцию скоростей (VACF). На рисунке 1.37 показаны RDF, полученный для различных MD/LL-FH случаев, здесь s заменяется константой $s = 0$ (чистый MD), константой $s = 0.1$, 0,5 и 0,9 и, для сравнения, также меняется в соответствии с (1.38) где использовались только атомы внутри сферы MD. Как видно из рисунка, на точность RDF не влияет введение гидродинамической компоненты. Поскольку RDF можно связать с плотностью на микромасштабах, это означает, что текущая процедура гибридизации не влияет на распределение атомов, сохраняя плотность жидкости независимо от концентрации «фазы» континуума.

На рисунке 1.38 представлены автокорреляционные функции скорости (VACF) для различных констант s для аргона, чистый MD ($s = 0$) и чистый LL-FH ($s = 1$). Видно, что для случаев преобладания гидродинамики (большие s), скорости сильно коррелированы в отличие от случаев

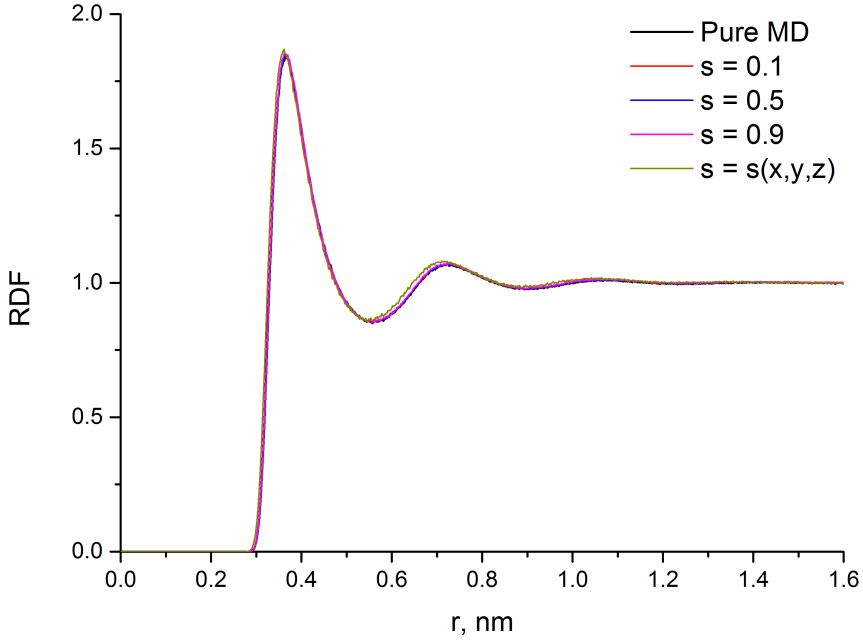


Рисунок 1.37 — Функции радиального распределения для различных значений s для аргона.

атомистического доминирования (маленький s). Кривые для промежуточных s плавно переходят от чистого решения MD к чистому LL-FH, в то время как s увеличивается с 0,1 до 1.

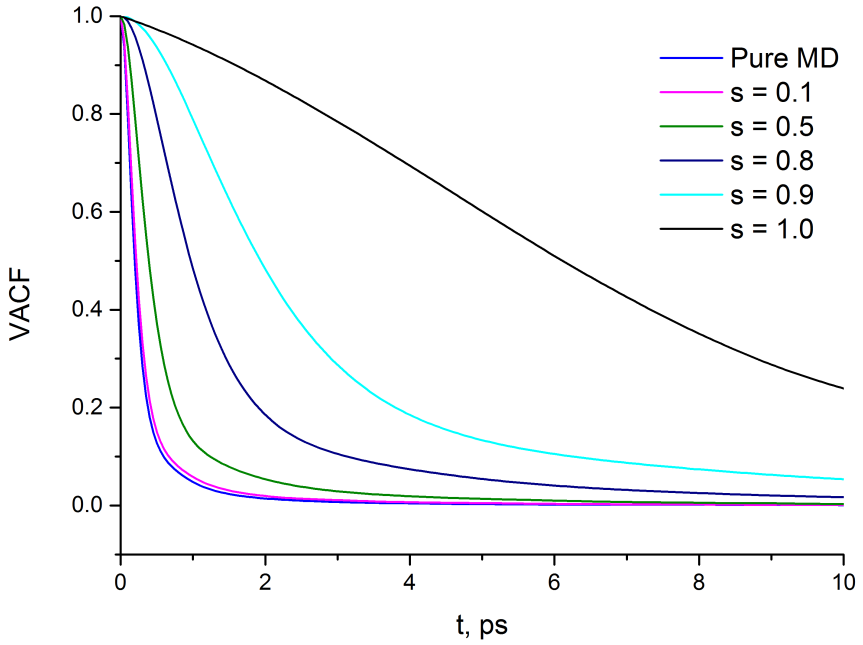


Рисунок 1.38 — Автокорреляционные функции скорости для различных постоянных s значений для аргона.

Переменная s-функция

Чтобы исследовать влияние гибридной зоны MD/LL-FH на точность атомистической части модели для случая переменной s в сферической области, описываемой уравнением (1.38), функция автокорреляции скорости (VACF) далее рассчитывали для аргона. Только те частицы, которые находятся внутри MD сферы, где $s = 0$, учитываются при расчете VACF. На рисунке 1.39 показано, что результирующая кривая VACF имеет форму, аналогичную чистой MD VACF. Скорости внутри области MD выглядят немного более коррелированными, чем в эталонном чистом MD. Скорее всего это эффект относительно небольшой сферы MD и высокого влияния области LL-FH которые могут индуцировать коллективные движения частиц MD в области MD.

Чтобы уменьшить эти движения, дальнейшая работа будет посвящена моделированию большего обьема LL-FH, который соответствует большей буферной зоне между чистой сферой MD и областями LL-FH, а также введение более плавного перехода от мелких пространственно-временных атомистических масштабов в центре к большим пространственно-временным гидродинамическим масштабам на границах домена LL-FH, например, используя подход пространственно переменного времени, описанного в [83].

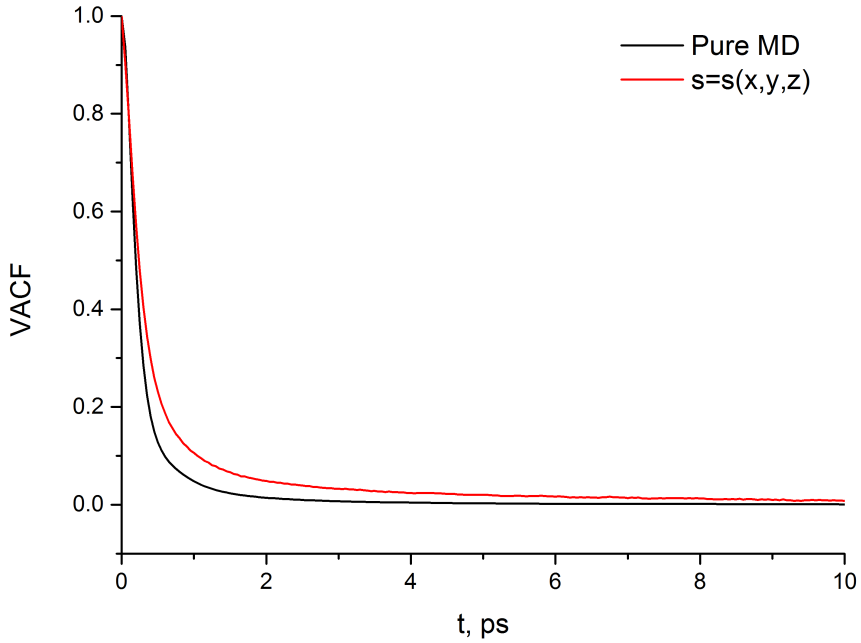


Рисунок 1.39 — Автокорреляционные функции скорости в случае переменной s для аргона.

Далее, непрерывность полей плотности и импульса, а также их флуктуации в гибридной зоне MD/LL-FH $0 < s < 1$ исследованы и сравнены с теоретическим решением. Для смеси, переменные текущей модели односторонней связи, $\bar{\rho}$ и $\bar{u}_i$, которые соответствуют решению уравнений LL-FH (1.37), гарантируют непрерывность и правильные флуктуации. Следовательно, именно MD часть решения остается в центре внимания современных исследований.

На рисунках 1.40 и 1.41 показаны изменения плотности, импульса и их стандартных отклонения для MD-части решения, $\rho_{MD} = \sum_{p=1,N(t)} \rho_p$ и $u_{MD} = \sum_{p=1,N(t)} u_{ip}\rho_p/\rho_{MD}$, построенной как функции расстояния от центра гибридной области моделирования, где s изменяется от 0 до 1. Вертикальные пунктирные линии обозначают границы гибридной зоны и значения теоретического решения обозначены индексом 0. Можно видеть что плотность сохраняется в пределах 0,1% и импульс сохраняется в пределах 0,5% от средней плотности и произведения средней плотности и скорость звука соответственно. Последний способ нормализации импульса был выбран, поскольку средняя скорость потока равна нулю.

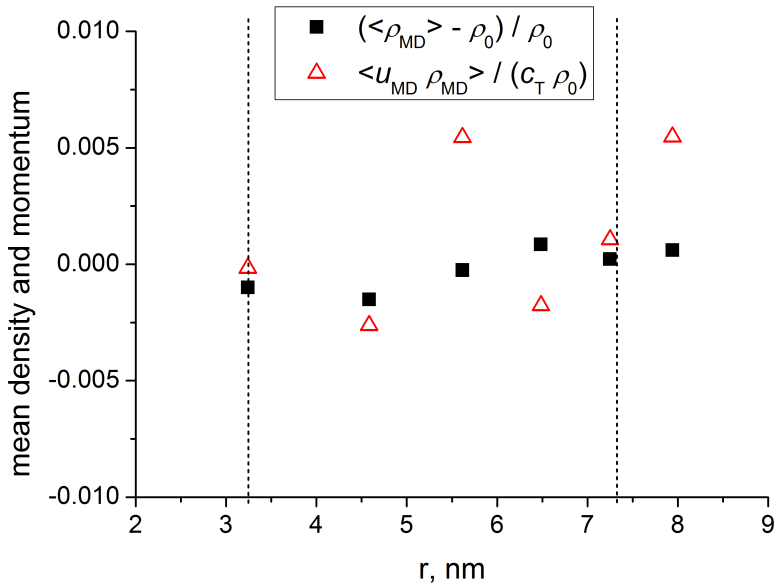


Рисунок 1.40 — Изменение средней по времени плотности и импульса в гибридной зоне $0 < s < 1$ показана вертикальными пунктирными линиями (слева $s = 0$ и справа $s = 1$) для аргона.

Стандартные отклонения плотности и скорости MD части гибридного решения по гибридной зоне колеблется как 5-7% от теоретического значения. Результаты получены для аргона и остаются аналогичными для случая модели воды, рассмотренной в следующем разделе.

Чтобы подтвердить, что часть решения MD удовлетворяет флуктуационно- диссипационной теореме, нужно показать, что (i) автокорреляция амплитуды плотности и скорости правильные и (ii) плотность и флуктуации скорости не коррелированы, т.е. автокорреляции обоих близки к дельта-функции в пределах уровня шума. Сохранение правильных колебаний в гибридной зоне MD/FH продемонстрированы на рисунке 1.41. Рисунок 1.42 показывает соответствующие автокорреляции плотности и x-флуктуации составляющей скорости MD-части решения, соответственно, где начало координат x_0 берется в середине гибридной зона MD/FH, которая соответствует $s = 0.5$. Обе автокорреляционные функции резко затухают при ненулевых пространственных расстояниях $|\delta_x| > 0$. Шумовой фон автокорреляционных функций, вероятно, связан с недостаточным временным усреднением. Обратите внимание, что пространственное усреднение, чтобы компенсировать отсутствие временной сходимости статистики, не проводилось.

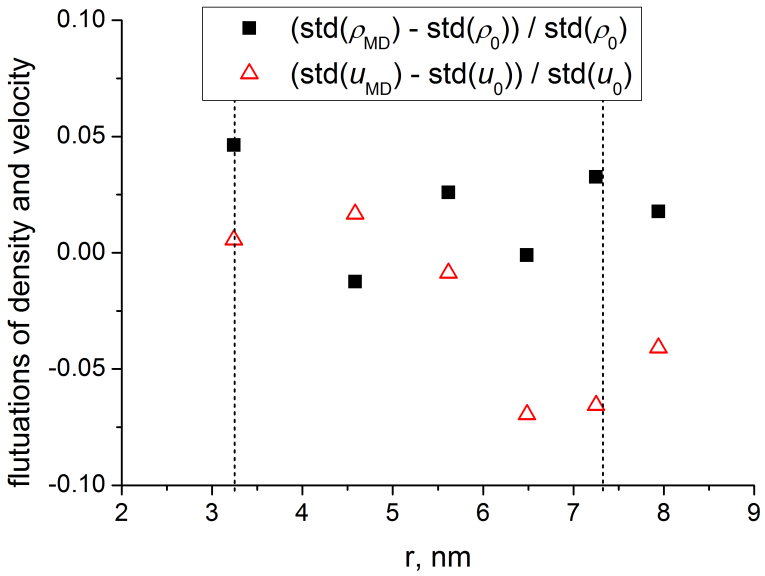


Рисунок 1.41 — Изменение стандартного отклонения плотности и импульса в гибридной зоне $0 < s < 1$, показано вертикальными пунктирными линиями (слева, $s = 0$ и справа, $s = 1$) для аргона.

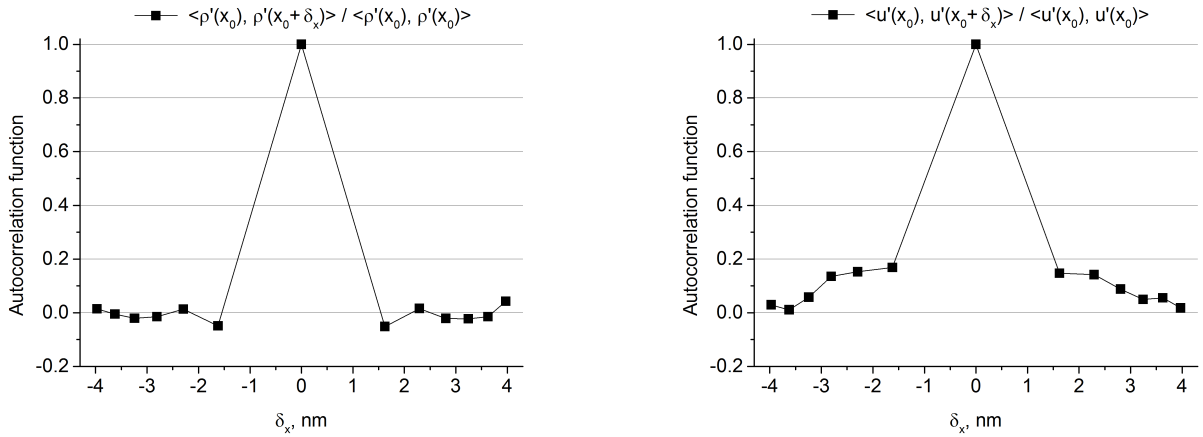


Рисунок 1.42 — Автокорреляции (а) плотности и (б) x-скорости компонент для аргона для положения x_0 в центре гибридной зоны MD/LL-FH.

В заключение этого раздела настоящий гибридный метод исследуется на предмет его способность корректно проводить акустическую волну через зону MD/LL-FH. Тест на распространение акустических волн, по существу, должен показать, насколько хорошо гибридная схема сохраняет импульс в нестационарном течении.

Для теста область решения, состоящая из 20 x 5 x 5 ячеек (x на y на z), которая в сумме содержит 32000 атомов аргона с «чистым» зоной MD в центре между двумя зонами LL-FH на входе и выходе. Решение периодической акустической волны задано как входное граничное условие области решения так, чтобы ее длина волны была в точности равна длине области решения в направлении x и периодические граничные условия для частиц оставались в силе. Граничное условие акустической волны было реализовано путем добавления аналитических членов источников в управляющие уравнения LL-FH во входных граничных ячейках. Исходные члены соответствуют

производным по времени от плотности и скорости набегающей акустической волны малой амплитуды, распространяющейся по заданному постоянному полю среднего течения LL-FH. Сигнал флуктуаций плотности вычисляется в пробной точке, расположенной в центре «чистой» области MD и сравниваемой с аналитическим решением.

На рисунке 1.43 сравниваются сигналы флуктуирующей плотности с аналитическим решением. Из-за очень низкого отношения акустического сигнала к тепловому шуму (~0,01), исходный сигнал полностью маскируется наличием тепловых флуктуаций плотности. Однако, в соответствии с диссипационно-флуктуационной теоремой, темповые флуктуации плотности некоррелированны, а после фазового усреднения и дополнительного пространственного усреднения в плоскости нормали к распространению акустической волны $(y-z)$, сигнал флуктуирующей плотности раствора MD становится очень близким к заданному аналитическому решению (Рисунок 1.43 справа). Примечательно, что несоответствие между расчетными колебаниями плотности в зоне MD после усреднения и аналитическим решением - того же порядка, что и уровень шума в той же гибридной модели без акустической волны.

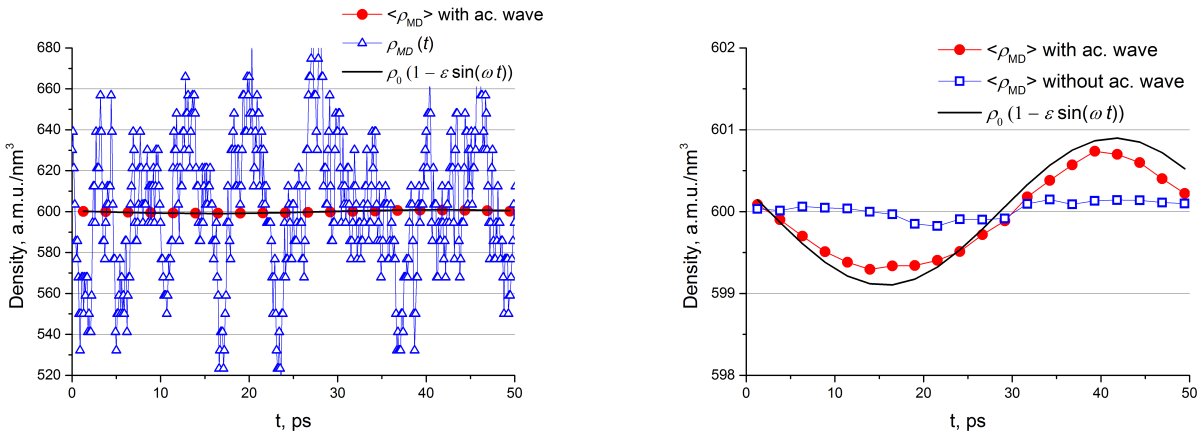


Рисунок 1.43 — Сигнал флуктуирующей плотности, полученный в MD части решения по сравнению с аналитическим решением (слева), включая исходный MD сигнала без фазового и пространственного усреднения, (справа) увеличенный вид с включением эталонного фазового и усредненное по пространству решение без учета акустической волны.

1.5.3 Гибридное моделирование воды

Далее представлены результаты моделирования SPC/E воды. Во-первых, сравним флуктуации: на рисунках 1.44 и 1.45 представлены стандартные отклонения флуктуации скорости и плотности для воды SPC/E. STD скорости для чистого MD, LL-FH и части MD гибридного решения сходятся к значению приблизительно 0.023-0.024 нм/пс. Что очень хорошо совпадает с теоретическим значением 0.0238 нм/пс согласно (1.98). С другой стороны, как и в случае с аргоном, STD плотности для чистой воды MD заметно завышено. С введением BLOB фильтра

размером 0.18 нм, согласование становится намного лучше: приблизительно 9.8-10.2 а.е.м./нм^3 по сравнению с теоретическим значением 10.06 а.е.м./нм^3.

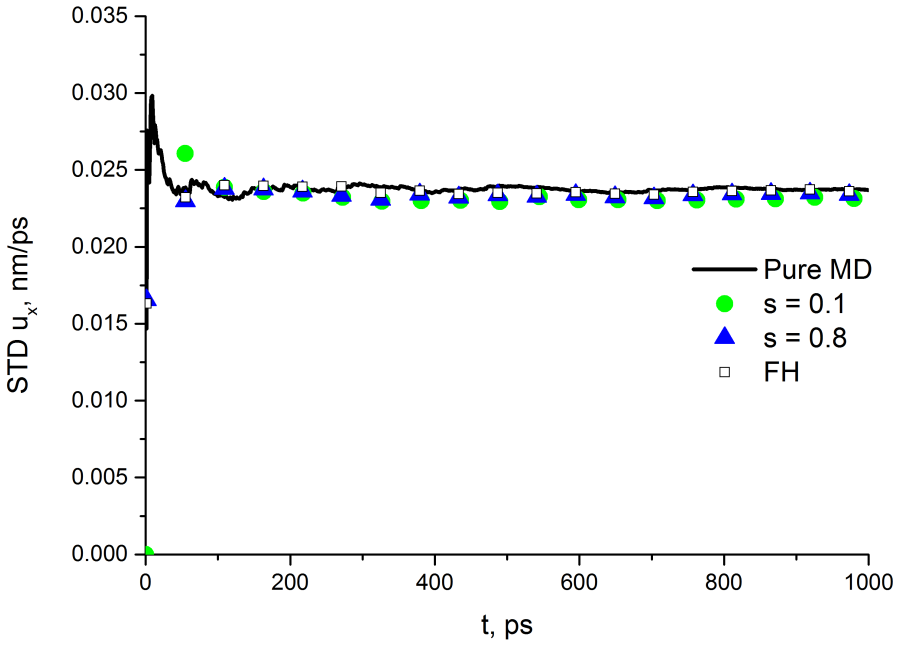


Рисунок 1.44 — Стандартные отклонения пульсаций скорости для SPC/E воды для четырех различных случаев; чистый MD (s =0), два варианта MD/LL-FH с $s = 0.1$ и $s = 0.8$ и чистый LL-FH ($s = 1$).

На рисунке 1.46 представлены автокорреляционные функции скорости для воды для различных постоянных параметров s. Здесь, в отличие от случая с аргоном, VACF воды имеют локальные минимумы и максимумы. Для значений s, приближающихся к 1, VACF становятся вытянуты по оси времени и стремятся к кривой LL-FH.

Далее, рассчитываем RDF и VACF для SPC/E воды для случая переменной s в соответствии с (1.38). На рисунке 1.47 показан снимок SPC/E воды для этого случая. Разрез, где центральная сфера представляет область MD с белыми и красными молекулами воды окруженная синими сферами, которые представляют более тяжелые и медленные частицы с преобладанием гидродинамики (водяные капли). Обратите внимание, что частицы воды с преобладанием гидродинамики на рисунке представлены сферами для ясности в отображении переходной зоны между MD и LL-FH. В реальности эти капли также являются молекулами воды, где относительный вес уравнения MD или LL-FH задаются значением s в соответствии с гибридной моделью MD/LL-FH.

На рисунках 1.48 и 1.49 показаны радиальные функции распределения для воды (пар O-H и O-O) для различных параметров s, включая случай с переменной s по сравнению с моделированием чистой MD. Для случая переменной s показано соответствующее решение внутри сферы MD ($s = 0$). Аналогично результатам для аргона, можно сделать вывод, что пространственное распределение атомов воды хорошо сохраняется как для постоянной s, так и для области MD, где s переменная.

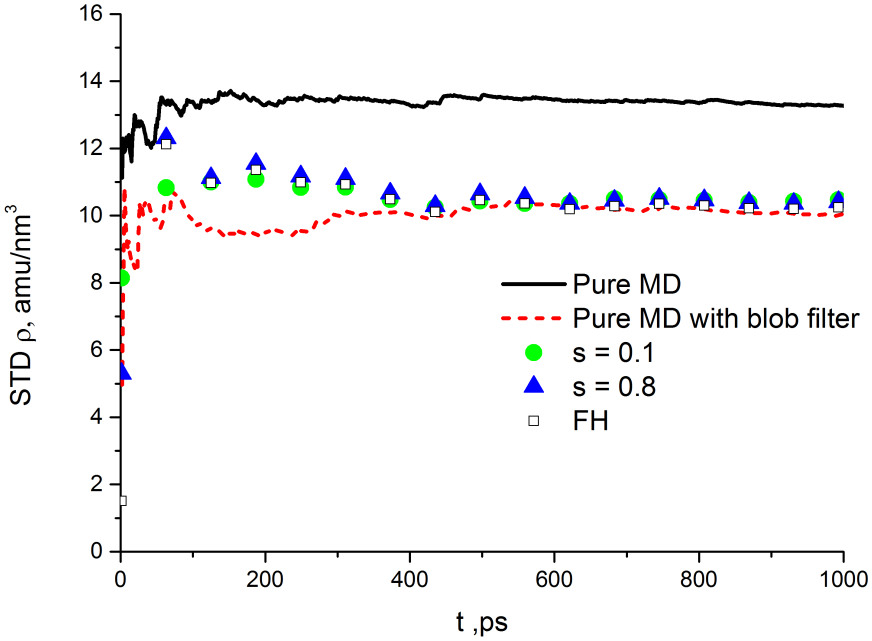


Рисунок 1.45 — Стандартные отклонения флуктуаций плотности для SPC/E воды для пяти разных случаев: чистый MD ($s = 0$) без BLOB фильтра, часть решения MD для s =0,1 и s =0,8 и решение LL-FH.

VACF воды внутри области MD в случае переменной s по сравнению со случаем чистой MD показаны на рисунке 1.50. Можно увидеть что для первых 0.06-0.07 пс функции одинаковы. Через 0.07 пс кривые немного отличаются из-за конечного размера сферы MD, но их форма очень похожа. Это означает, что для переменной s в сфере MD с $s = 0$ микроскопические статистические свойства воды сохраняются: как с точки зрения RDF так и VACF.

1.5.4 Гибридное моделирование диаланина в воде

Следующий шаг – добавление небольшой пептидной молекулы, цвиттерионной формы диаланина, в воду. Одна молекула пептида изначально помещается в центр MD-сферы (область $s = 0$, изображенная на рисунке 1.51) и в окружении нашей гибридной модели SPC-E/гидродинамической воды. Начальная конфигурация изображена на рисунке 1.51.

В случае неподвижной гибридной зоны моделирование останавливается когда макромолекула достигает гибридной зоны (1.38). Чтобы продлить время моделирования, вводится нестационарная зона MD связанная с движением центра масс пептидной системы, так что параметр связывания в (1.35) становится функцией пространства и времени, $s = s\,(x,y,z,t)$.

Чтобы проверить влияние области с преобладанием гидродинамики ($s > 0$) на область MD ($s = 0$) вычисляли коэффициенты поступательной самодиффузии D как для воды, так и для молекул пептида в этой области и сравнивали их с величинами, полученными в чистом MD. Для

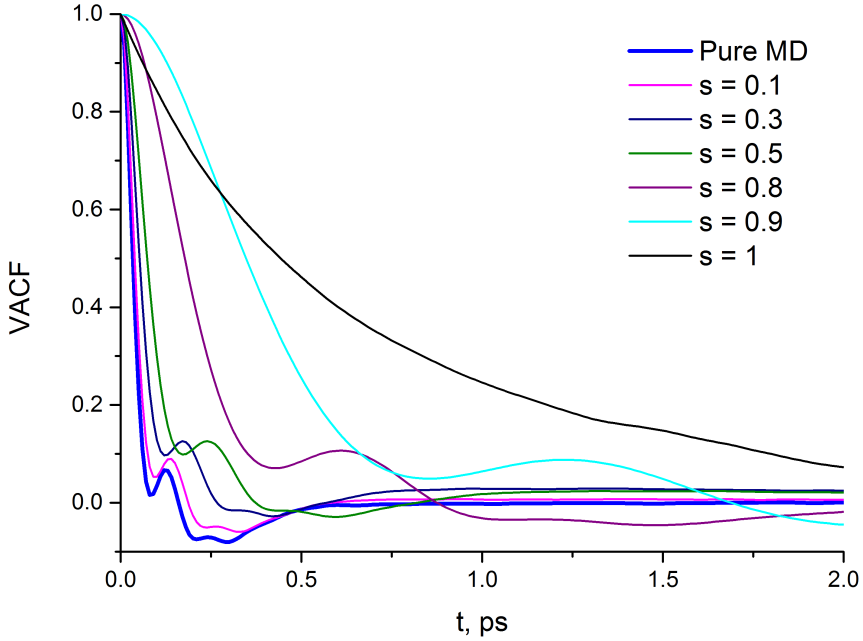


Рисунок 1.46 — Автокорреляционные функции скорости для различных постоянных значений s для SPC/E воды.

расчета использовалось соотношение Эйнштейна D: $MSD(t) = \langle \Delta r(t)^2 \rangle = A + 6Dt$, где A — это произвольная постоянная. Важно отметить, что эта формула верна только на больших временах (и точна на бесконечно больших временах). На практике должна использоваться часть кривой, которая может быть удовлетворительно аппроксимирована прямой линией. Другими словами, следует учитывать локальный наклон траектории среднеквадратичного отклонения на больших временах, а начальный участок на коротких временах следует отбрасывать.

Все расчеты (чистые и гибридные) были выполнены с одними и теми же параметрами: $T = 298$ K (термостат Nose-Hoover); постоянная плотность $\rho = 999.15$ g/cm^3 для воды и 992.92 g/cm^3 для раствора пептида; временной шаг MD $\Delta t = 1$ fs; электростатика поля реакции с длиной отсечки 0,9 нм и диэлектрической проницаемостью 78; отсечка Ван-дер-Ваальса 0,9 нм.

При оценке D проводилось моделирование для чистого MD длительностью 10 нс и моделирование раствора пептида длительностью 40 нс. Графики MSD(t) были рассчитаны для интервалов в 1 нс, а затем усреднены (Рисунок 1.52). Полученные значения D: D(вода) $= 2.68 \cdot 10^{-5}$см^2/с, D(белок) $= 0.86 \cdot 10^{-5}$ см^2/с.

Однако вычисление D в случае гибридной многомасштабной модели нецелесообразно просто потому, что, в отличие от одномасштабной MD, тестовая молекула покидает внутреннюю сферу MD, за которой находится область с преобладанием гидродинамики, где диффузия отдельных молекул представлена неправильно. Поэтому были приняты специальные меры для корректного расчета MSD. Они не отличаются от специальных мер, которые необходимо принять для проверки других многомасштабных алгоритмов, которые проходят несколько масштабов по сравнению с задачей одного масштаба [11].

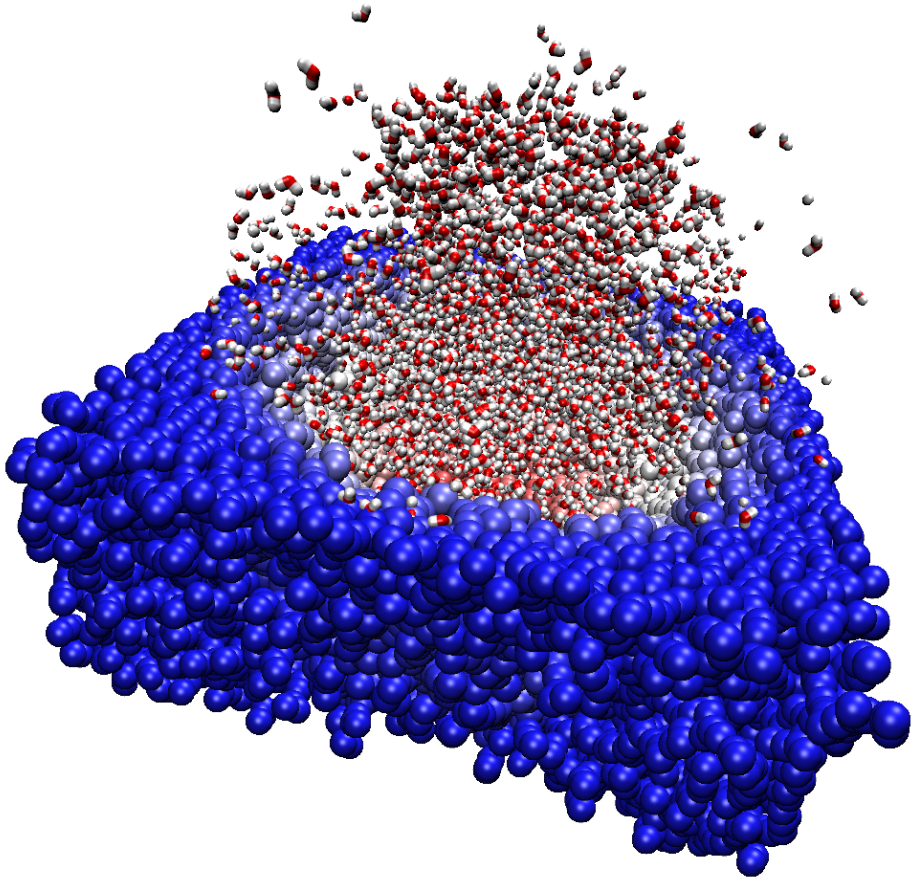


Рисунок 1.47 — Разрез, показывающий SPC/E воду в случае переменной s; белые и красные молекулы - молекул в MD-области, синие сферы — капли воды в области с преобладанием гидродинамики.

Поскольку молекуле требуется относительно короткое время, чтобы достичь гибридной зоны, нет смысла проводить длительные симуляции, а типичный алгоритм, аналогичный упомянутому выше, становится неприемлемым. Поэтому, получали множество коротких траекторий одиночных молекул MSD(t) и усредняли их для накопления статистически достоверных данных. С одной стороны длинные траектории необходимы, чтобы удовлетворить условию ограничения по времени для соотношения Эйнштейна, с другой стороны, чем дольше моделирование, тем меньше молекулы остаются в области $s = 0$. Следовательно, необходимо: (i) разместить интересующие молекулы в центре коробки в начале моделирования и (ii) исключить молекулы, посетившие область $s > 0$ из усреднения графиков MSD(t).

Для удовлетворения этих требований был использован следующий алгоритм:

1. *Подготовка исходных ячеек*. Во время начального уравновешивания молекулу пептида фиксировали в центре коробки наложением ограничений на пептидную связь. Всего приготовили 70 ячеек с водой и 182 ячеек с раствором пептида, начиная с разных исходных конфигурации.
2. *Сбор данных*. Прогоны гибридного MD имели длину 200 пс. Начиная с 80 пс, каждый 20 пс кадр записывался для дальнейшего анализа. График MSD(t) рассчитывался для каж-

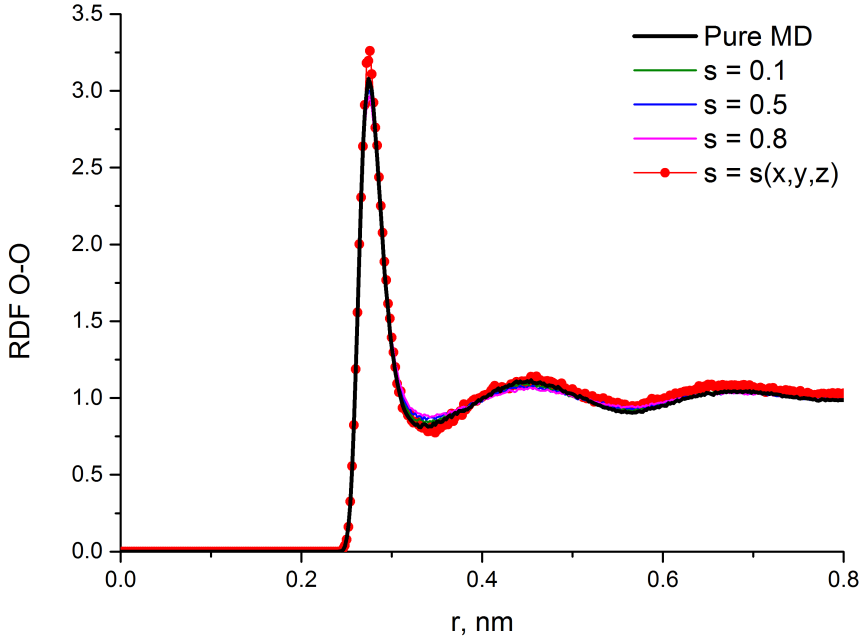


Рисунок 1.48 — Функции радиального распределения кислород-кислород для разных значений s для SPC/E воды.

дой траектории. В каждом моделировании чистой воды, использовались те молекулы, которые уже находились в центре исходной ячейки (1-5 молекул на ячейку).

3. *Фильтрация*. Анализируя записанные кадры, траектории с тестовыми молекулами, посетившими $s > 0$ регион были определены и исключены из выборки. Критерием фильтрация являлось расстояние между геометрическим центром молекулы и центром ячейки. Они должны были быть меньше радиуса обрезанияR_c. Было исследовано несколько радиусов обрезания R_c в диапазоне от 1,3 до 2,9 нм.
4. *Оценка D*. Полученные графики MSD(t), Рисунок 1.53, 1.54, были усреднены по графикам MSD одиночной молекулы, полученным после фильтрация. Графики MSD для воды и пептида почти одинаковы ниже некоторого значения R_c (2,1 нм для воды и 1,5 нм для пептида), которое предполагает, что результаты достаточно независимы от R_c и дальнейшее уменьшение обрезания R_c не требуется. Подгонка для D была сделана с промежутками в середине периодов (40–80 пс для воды и 45–70 нс для раствора пептида). На малых временах график не удовлетворяет требованию отношения Эйнштейна, в то время как конечные интервалах близость гибридной зоны влияет на результат. Окончательный D рассчитывался как среднее значение по трем наименьшим радиусам обрезания R_c (Таблицы 6 и 7).

Погрешность D оценивалась как сумма двух погрешностей, которые предполагались некоррелированным:

1. Во всем интервале подгонки 20 пс (белок) или 15 пс (вода) извлекались периоды со сдвигом 10 пс. Несколько значений D_i были рассчитаны из этого интервала путем подгонки,

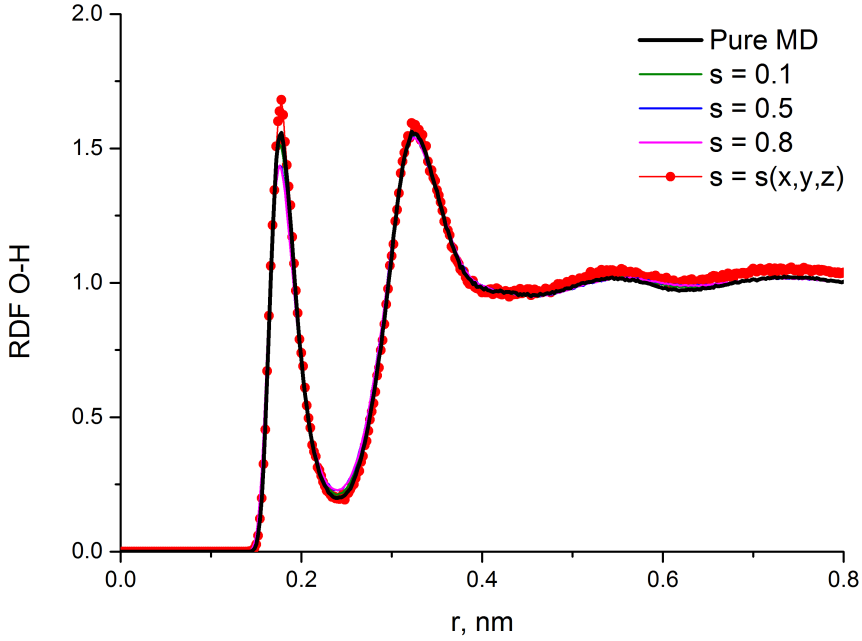


Рисунок 1.49 — Функции радиального распределения кислород-водород для разных значений s для SPC/E воды.

Таблица 6 — Коэффициент диффузии D получен в гибридном моделировании.

R_c, нм	D, 10^{-5} см^2/с
вода	
2.3	2.25
2.1	2.15
1.9	2.01
белок	
1.7	1.53
1.5	1.42
1.3	1.37

а половина диапазона между самой маленькой и самой большой D_i были приняты как погрешность наклона ΔD_{sl}.

2. Зависимость от R_c учитывалась с неопределенностью ΔD_c, полученной по формуле:

$$\Delta D_c = \left|\frac{\partial D}{\partial R_c}\right|, \Delta R_c \approx \left|\frac{D(R_{c2}) - D(R_{c1})}{R_{c2} - R_{c1}}\right| \Delta R_c,$$

где ΔR_c было принято равным 0,2 нм, так что общая неопределенность

$$\Delta D = \sqrt{\Delta D_{sl}^2 + \Delta D_c^2}.$$

Окончательные коэффициенты диффузии D перечислены в таблице 7. Наш гибридный метод несколько завышает D для молекул воды и занижает для пептида. С учетом неопределенностей

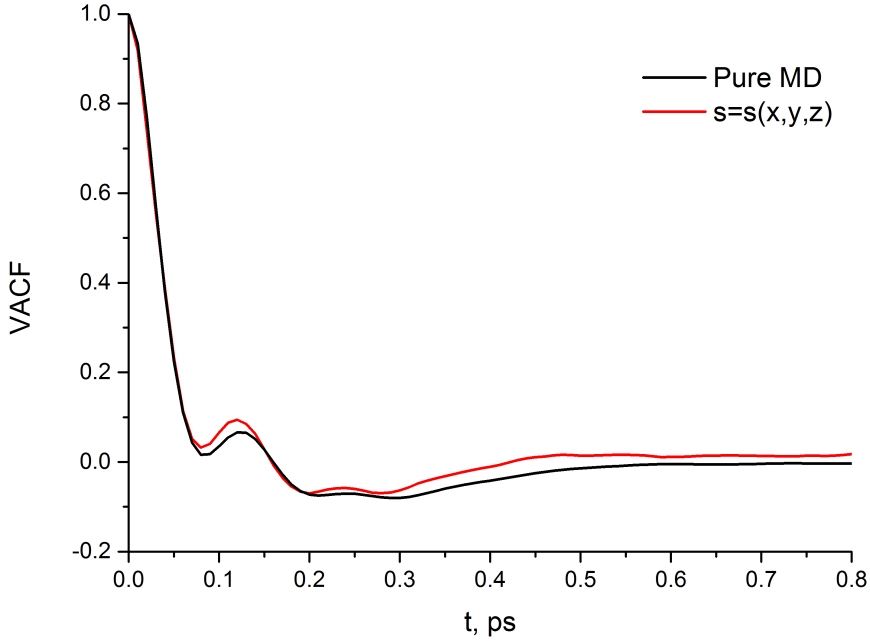


Рисунок 1.50 — Автокорреляционные функции скорости в случае переменной s для SPC/E воды.

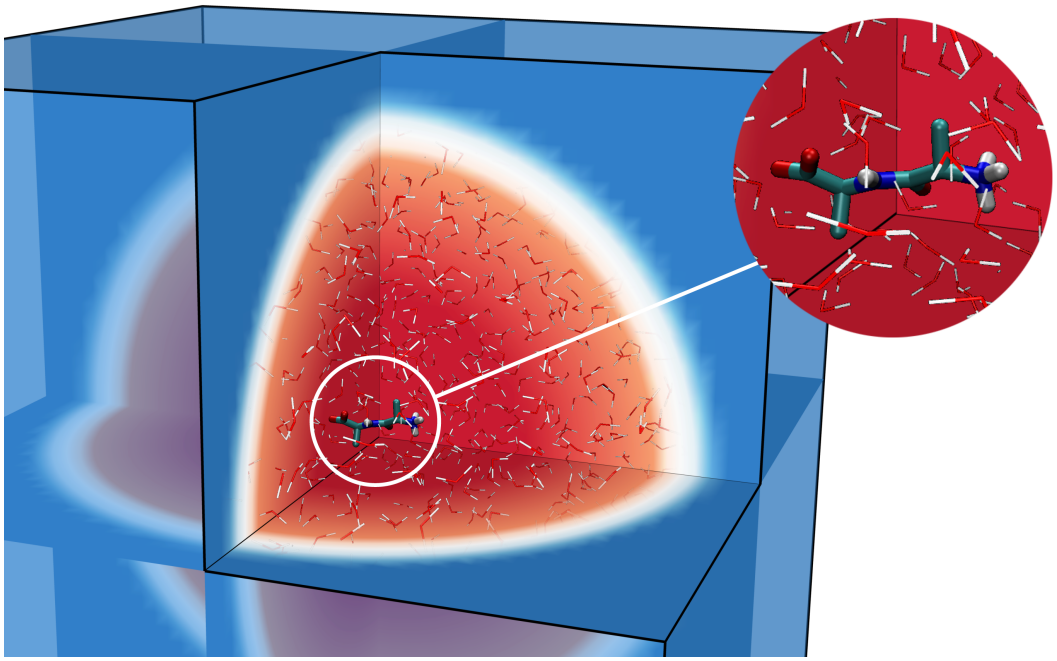


Рисунок 1.51 — Переменный параметр s и MD-сферас небольшой пептидной молекулой, цвиттерионной формой диаланина, внутри в центре сферы. Красная сфера — область чистой MD ($s = 0$), синий — область флуктуационной гидродинамики (s =1).

расхождение между гибридной моделью и эталонным моделирование MD составляют около 20% для воды и 30% для пептида соответственно. Это можно связать с малостью зоны MD, в которой рассматриваются тестовые молекулы, и ее близость к гибридной зоне. Первый эффект уменьшает статистическую выборку, доступную для обработки для определения такой нелокальной величины, как молекулярный коэффициент диффузии, а второй приводит к артефакту взаимодействия

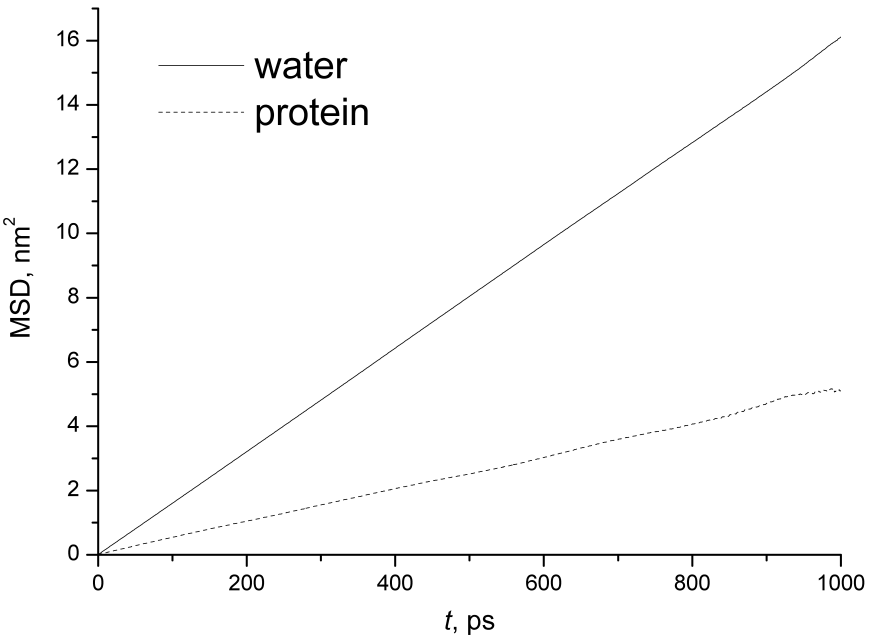


Рисунок 1.52 — График MSD(t)) для воды SPC/E и белка для чистого MD.

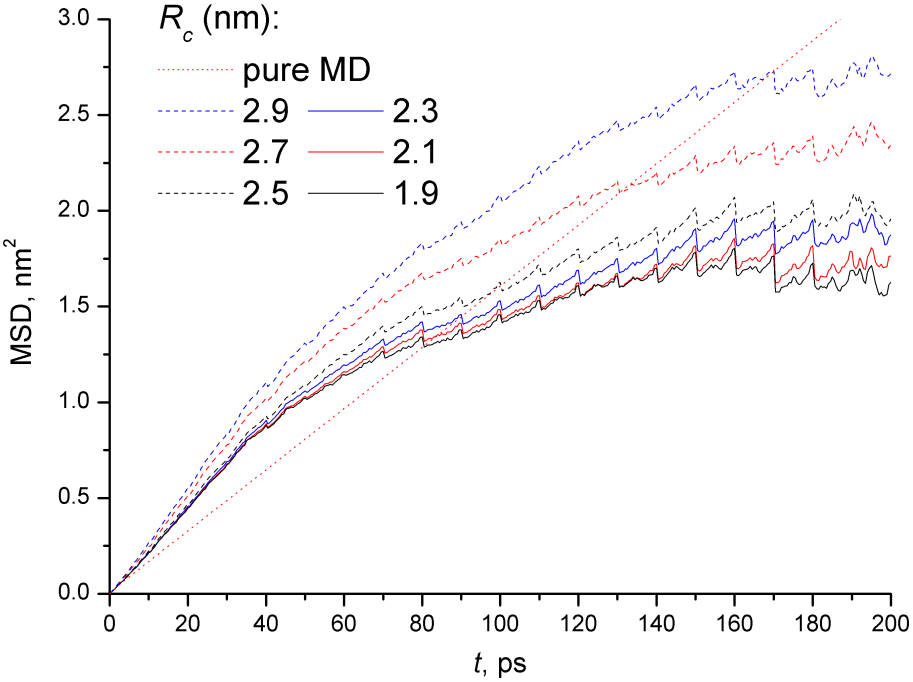


Рисунок 1.53 — Графики MSD(t) для воды (гибридные MD), рассчитанные с разными радиусами обрезания. График чистог MD приведен для сравнения.

между тестовыми молекулами в зоне MD где преобладает гидродинамика. Чтобы смягчить эти эффекты, большая MD зона и более широкая гибридная зона будут реализованы в будущем.

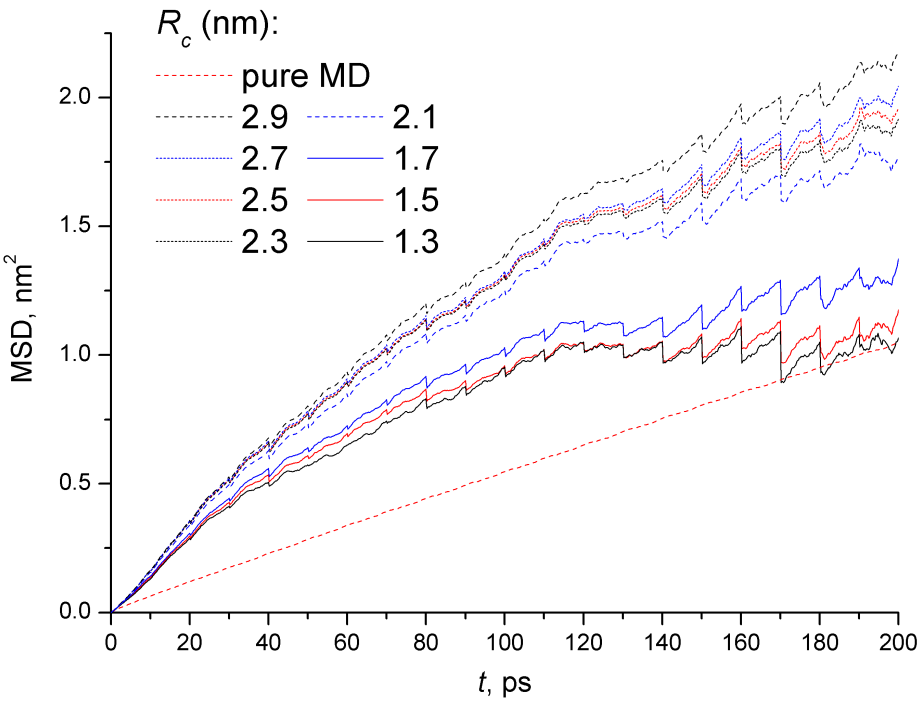


Рисунок 1.54 — Графики MSD(t) для белка (гибридная MD), рассчитанные с разными радиусами обрезания; график для чистого MD приведен для сравнения.

Таблица 7 — Окончательные значения D (погрешность для результатов чистого MD пренебрежимо мала)

	D, 10^{-5} см^2/с	
	чистая MD	Гибридная
Вода	2.68	2.1 ±0.12
пептид	0.86	1.4 ±0.16

1.5.5 Вирус PCV2

Полноатомный капсид вируса PCV2 при старте моделирования из кристаллографической структуры (Рисунок 1.55) схлопывается после приблизительно 1 нс времени моделирования (Рисунок 1.56), при этом он остается стабильным в гибридном моделировании (Рисунок 1.57). Можно обьяснить этот эффект артефактами стандартных периодических граничных условий в случае полноатомного моделирования.

Для количественной оценки сохранения формы вирусного капсида был использован корень среднеквадратичного отклонения (RMSD) атомов белка по сравнению с исходной структурой в обоих случаях: для полноатомного моделирования и для гибридного подхода LL-FH/MD. Рисунок1.58 показывает, что RMSD атомов белков в случае моделирования чистой MD продолжает расти, что указывает на то, что система все еще не стабилизирована. В случае LL-FH (Рисунок 1.58 внизу) расчеты значений RMSD асимптотически стремятся к константе после ~60 пс. Это остаточное среднеквадратичное значение составляет около 0.4 нм, что достаточно мало, что-

бы предположить, что структура вируса стабильна в соответствии с литературой [121] и [101], где утверждается, что значения RMSD около 0.3 нм приемлемы для стабильности крупных белков.

Интересно, что коллапс капсида вируса наблюдается в чистом MD моделировании, даже если размер системы увеличен до 4 000 000 атомов, то есть содержит гораздо больше молекул воды. В противоположность, для модели LL-FH/MD расчетная область вдвое больше, что соответствует вдвое большему размеру зоны чистого MD (примерно 1 x 180 000 атомов) и отодвигает «гибридную границу» от вируса, находящегося в центре чистой зоны MD, приводя к правильной симметричной структуре очень похожей на экспериментальную. Результатом этого теста было то, что форма вируса практически не изменилась, независимо от размера используемой вычислительной области. В частности, RMSD атомов основной цепи капсида составляли 0,4 нм для меньшей вычислительной области и 0,5 нм для большей, что демонстрирует практически неизменную структуру по сравнению с кристаллографическими экспериментальными данными.

Приведенные выше результаты позволяют предположить, что для сохранения вируса в полноатомной модели со стандартными периодическими граничными условиями размер расчетной области должен быть увеличен, следовательно, число атомов воды и связанные с этим вычислительные затраты должны быть значительно увеличены. Введение модели LL-FH/MD, в свою очередь, позволяет проводить моделирование задачи в такой же достаточно небольшой вычислительной системе без заметных числовых артефактов.

Более того, так как расчет сил MD выключен внутри FH области модели LL-FH/MD, расчет ускоряется в несколько раз (в зависимости от объемного соотношения MD/FH) по сравнению с полноатомным моделированием для того же размера вычислительной ячейки. Например, в нынешнем моделировании область FH содержит примерно 1/2 часть от общего числа частиц в ячейке. Предполагая, что расчет MD потенциала является наиболее затратной в вычислительном отношении частью численного алгоритма (расчет сил занимает до 90% всего вычислительного времени), можно получить ускорение вычислений в два раза, когда расчет сил MD производится только для частиц внутри чистой зоны MD и промежуточной области LL-FH/MD. Преимущества гибридного подхода становятся еще более явными по сравнению с полноатомным моделированием для больших гидродинамических (FH) областей.

1.6 Выводы

Был разработан новый самосогласованный гибридный подход молекулярной динамики (MD) и континуальной модели гидродинамики (HD), основанный на аналогии с двухфазной гидродинамикой. В этом подходе уравнения MD имеют ограничения такие, что их макроскопическое поведение удовлетворяет граничным условиям в гибридной флуктуационной гидродинамике Ландау-Лифшица (LL-FH) со строгим сохранением макроскопической массы и импульса во всей области решения.

Параметр детализации представления, s, вводится в основные уравнения с использованием аналогии с двухфазной гидродинамикой, позволяющей учитывать различные масштабы в

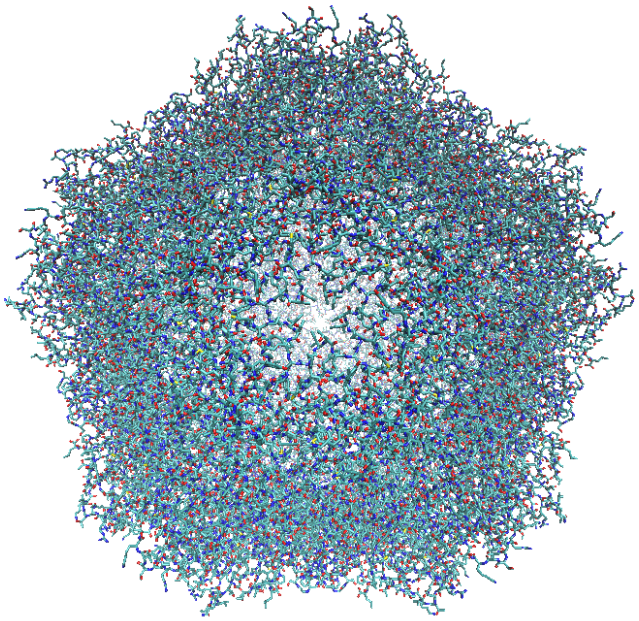


Рисунок 1.55 — Начальная конформация капсида PCV-2 по экспериментальным рентгеновским данным.

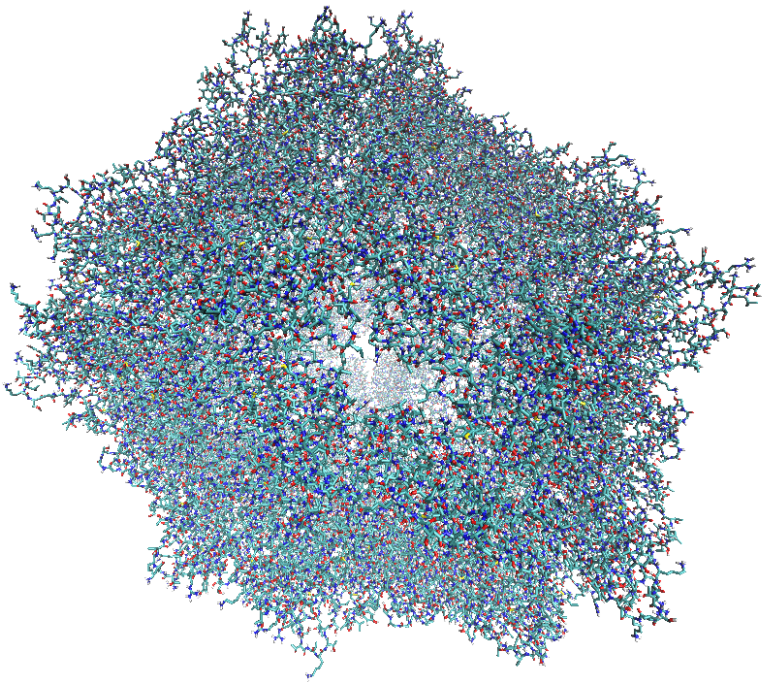


Рисунок 1.56 — Капсид PCV2 нестабилен в чистом MD.

представлениях MD и LL-FH. Модель LL-FH для макроскопического описания позволяет воспроизвести статистику пульсаций тепловой скорости для широкого диапазона контрольных объемов от 8 до 2190 атомов на длину ячейки. Его реализация может быть настолько же эффективна, как и обычный метод детерминированной вычислительной гидродинамики (CFD) с использованием неравномерных сеток Эйлера и асинхронного шага по времени. Микроскопические флуктуации, которые можно ввести с помощью LL-FH модели, не только не пренебрежимо малы, но и составля-

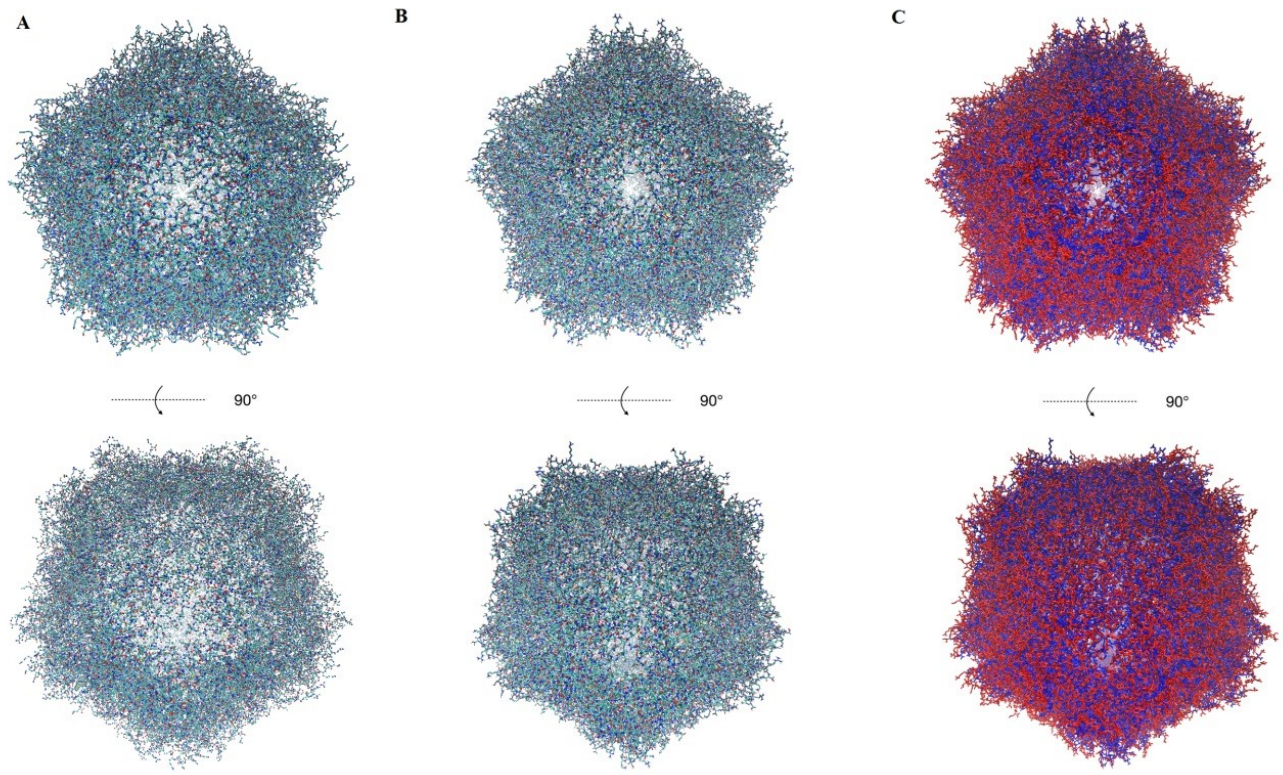


Рисунок 1.57 — Капсид PCV-2 стабилен в гибридном моделировании LL-FH/MD: исходная форма вируса после уравновешивания, соответствующая экспериментальная структура; B — смоделированная структура в гибридном LL-FH/MD; C это сравнение A и B (синий - моделирование, красный - исходный вирус после уравновешивания).

ют суть молекулярных явлений, что было продемонстрировано при моделировании цвиттерион диаланина в растворе.

Новый гибридный метод LL-FH/MD реализован для моделирования несколько жидкостей, таких как вода MB, жидкий аргон, вода SPC/E, растворы пептида и вирусной частицы в равновесии. Показано, что этот метод позволяет получить сходящееся решение, которое правильно воспроизводит макроскопические параметры, такие как стандартные отклонения тепловых пульсаций и средней скорости потока. Это достигается без введения какой-либо специальной обработки частиц MD, такой как «изменение масштаба» решения MD.

Было продемонстрировано следующее:

(i) для постоянного параметра s текущая трехмерная реализация гибридного метода правильно воспроизводит макроскопические флуктуации плотности и скорость в соответствии с литературными данными;

(ii) для переменного параметра s, несмотря на некоторую чувствительность к размеру гибридной зоны MD/LL-FH, гибридный метод сохраняет важные структурные функции жидкостей, такие как функция радиального распределения, а также автокорреляционную функцию скорости в атомистической части решения; изменение структурных функций происходит постепенно при изменении разрешения представления при введении влияния гидродинамики на MD;

(iii) показано, что масса и импульс MD-части решения сохраняются в гибридной зоне MD/LL-FH в пределах 0,5%;

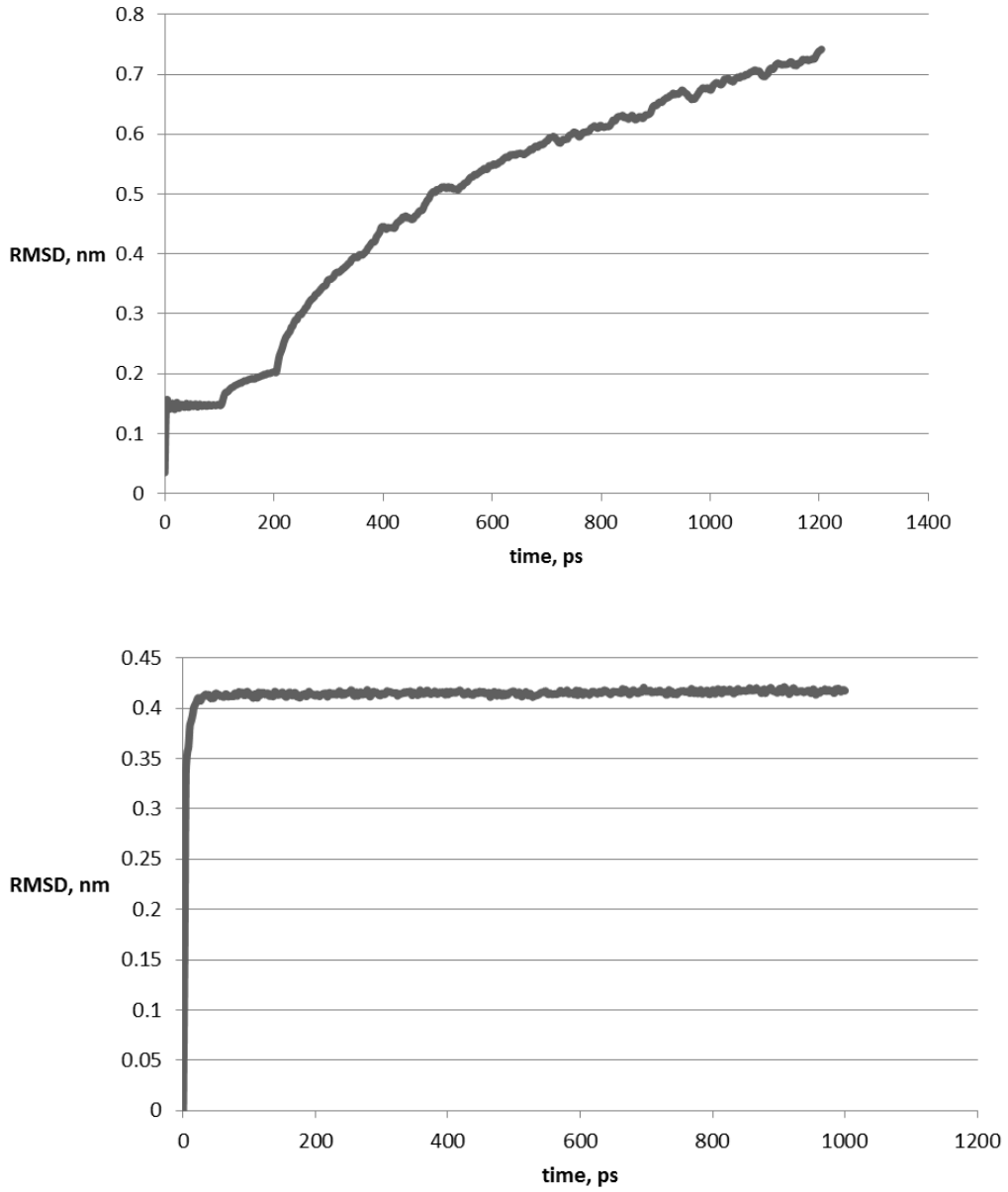


Рисунок 1.58 — Скользящие средние значения среднеквадратичного отклонения (RMSD) атомов основной цепи от их исходной структуры для расчета чистой MD (вверху) и гибридной модели LL-FH (внизу).

(iv) автокорреляции плотности и скорости MD части решения правильно сохраняются в гибридной зоне MD/LL-FH в соответствии с флуктуационно-диссипациионной теоремой;

(v) результаты моделирования бегущей акустической волны через гибридный MD/LL-FH слой продемонстрировали способность метода правильно передавать импульс в нестационарном течении с точностью до статистического шума;

(vi) предварительные результаты гибридного метода молекулярной диффузии воды и диффузия диаланина в воде показывают разумное согласие с эталонным моделированием MD.

Гибридный метод LL-FH/MD [2; 8] был расширен для динамического отслеживания особенностей атомистического разрешения, которые могут развиваться в потоке. Было выполнено моделирование двух очень разных биомолекулярных систем: диффузии пептида в воде и вза-

имодействие капсида цирковируса с водой. Было показано, что гибридная модель в состоянии сохранить важные свойства молекулярной части системы. При диффузии пептида в воде, флуктуации массы и импульса зоны чистой молекулярной динамики, окружающие пептид, были найдены в отличном согласии с теорией. Полученный коэффициент диффузии из гибридного решения был найден в очень хорошем согласии со всеми эталонными моделированиями. Для моделирования капсида вируса в воде, по сравнению с полноатомным моделированием, существующий гибридный метод сохраняет структуру вируса и приводит к корректному среднеквадратичному отклонению (RMSD) структуры атомов основной цепи. Это асимптотическое значение достаточно мало, что предполагает стабильность структуры, и также достаточно мало чувствительно к размеру расчетной области в случае гибридного метода. В отличие от полноатомного моделирования со стандартными периодическими условиями, которые не могут обеспечить постоянное среднеквадратичное отклонение, а также не могут сохранить стабильность вирусного капсида даже при увеличенном размере вычислительной области.

Дальнейшее развитие гибридного метода будет включать его расширения до алгоритмов двусторонней связи в 3D, чтобы включить обратную связь молекулярной динамики на гидродинамику. Дополнительные разработки будут включать расширение текущего подхода, который использует постоянные объемы интегрирования для гидродинамики и постоянными шагами по времени для MD частиц независимо от значения s-параметра, который играет роль «крупнозернистости» (от MD до FH) до нескольких пространственно-временных алгоритмов, таких как TARDIS [83] для дальнейшего повышения вычислительной эффективности.

Глава 2. Сложность растворов биомолекул с точки зрения нелинейной динамики

2.1 Количественная оценка долговременной памяти атомистических траекторий молекулярных систем

Принято считать, что эффекты памяти в жидкостях можно охарактеризовать и, следовательно, понять, вычислив различные молекулярные корреляционные функции, чаще всего атомистические функции скорости или автокорреляционные функции координат. Они указывают на память длительностью не более нескольких пикосекунд, после чего все процессы в простых молекулярных жидкостях выглядят совершенно случайными. Однако недавно было показано с нескольких разных точек зрения, что молекулярные системы имеют память, которая сохраняется значительно дольше [12—15; 122; 16—18]. Это особенно актуально для жидкой воды, которая, как было показано, демонстрирует медленную релаксацию, явление, связанное с зависимостью спектра флуктуаций энергии от частоты $1/f$ [123; 124].

Очевидно, такая простая статистическая мера, как усредненное по времени (или ансамблю) произведение разделенных во времени значений – двухточечная корреляционная функция – неадекватна для представления сложных многоточечных эффектов памяти в молекулярных жидкостях. С другой стороны, классические молекулярные статистические ансамбли, будучи детерминированными динамическими системами, могут быть описаны с альтернативной точки зрения с применением геометрических и статистических методов, недавно разработанных в области нелинейной динамики и теории информации.

Описание жидкостей как нелинейных динамических систем имеет то преимущество, что изучение траекторий происходит в правильно выбранных проекциях фазового пространства а не просто координат и скоростей одного атома или ограниченного набора атомов. Полноразмерная траектория фазового пространства содержит, по крайней мере теоретически, всю возможную информацию о системе. Таким образом, анализируя различные проекции траектории фазового пространства с использованием статистических методов, которые могут извлекать существенную неслучайную информацию, можно выявить скрытую память в жидкостях.

Однако прямое применение методов нелинейной динамики, таких как анализ размерностей, энтропий, различных мер сложности, экспонент Ляпунова и т.д., к траекториям фазового пространства наталкивается на определенные трудности, так как большинство этих методов могут иметь дело только с низкоразмерными системами. Поэтому невозможно исследовать полное фазовое пространство траекторий даже малых молекулярных систем с помощью этих методов, не говоря уже о реалистичных моделях жидкости, состоящих из тысяч атомов. Таким образом, новые теоретико-информационные и статистические методы для проведения анализа многомерных молекулярных траекторий крайне востребованы.

В этой главе используется подход, анализирующий молекулярные траектории в терминах динамики символов. Молекулярная траектория преобразуется в последовательность символов из алфавита состоящего всего из нескольких символов. Полученная символьная последовательность может быть проанализирована с помощью различных статистических методов. Важным моментом является тот факт, что траектория анализируется не как отдельные символы, а как последовательность символьных строк (слов). Это позволяет извлечь подробную информацию из изначально непрерывной траектории, несмотря на кажущееся очень грубое представление с использованием всего нескольких символов [16; 19—21].

Была проанализирована динамика жидкой воды, с использованием такого подхода и была обнаружена удивительно медленная сходимость рассчитанных статистических показателей. Статистика временных рядов для воды ведет себя принципиально подругому, по сравнению с псевдослучайными последовательностями (например, цифры числа π) или со случайным суррогатным сигналом, имеющим корреляционную функцию (и, следовательно, спектр мощности), идентичную корреляционной функции воды. Более того, динамика воды напоминает поведение, наблюдаемое в простой хаотической системе, такой как отображение Чирикова-Тейлора или стандартное отображение. Таким образом, выдвинуто предположение, что такие неслучайные свойства молекулярных траекторий могут носить детерминированный хаотический характер.

2.1.1 Методология оценки памяти системы

Молекулярные траектории были получены с использованием классической молекулярной динамики. Объемная вода (периодические граничные условия) состоящая из 392 SPC [125] молекул была промоделирована с использованием пакета `GROMACS` [126]. Температура системы поддерживалась постоянной при 300K с использованием термостата Berendsen [127]. Перед сбором данных для анализа было выполнено достаточное уравновешивание. Скорость атома водорода одной из молекул воды использовались в качестве временного ряда для анализа. В местах, где кривая скорости пересекает плоскость xy, были сгенерированы точки двумерного отображения (сечения), которые использовались в качестве исходного временного ряда для анализа (Рисунок 2.1). Анализируемая траектория длиной 1μс давала приблизительно $3 \cdot 10^7$ точек данных.

Двумерное сечение траектории затем символизировалось с использованием угловой переменной в полярной системе координат. Угол 2π был разделен на три равных сегмента и точкам данных были присвоены символы «0», «1» или «2» в зависимости от сегмента, куда попадали значения угла (Рисунок 2.1). Такой способ деления фазового пространства и преобразования значений вектора скорости (вещественные числа) в символы из трехсимвольного алфавита имеют асимптотически однородную структуру символических последовательностей, то есть таких, которые имеют одинаковую частоту появления для каждого символа.

Заметим, что выбор подходящего метода символизации данных — нетривиальная задача. В идеальном случае необходимо получить так называемое генерирующее деление, т.е. такое, которое обеспечивает уникальное представление исходного вещественного временного ряда с

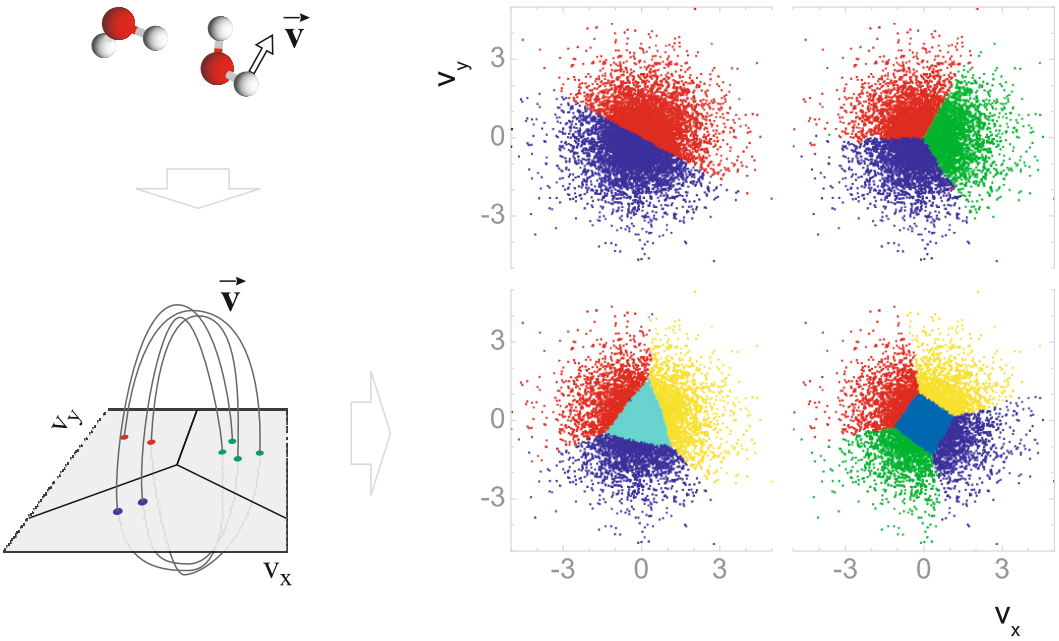


Рисунок 2.1 — Процесс преобразования непрерывного сигнала скорости атомов $\mathbf{v}$ в символьную последовательность; справа показана символизация с 2, 3, 4 и 5 символами; 3-символьный алфавит использовался во всех последующих расчетах.

помощью строки символов. Расчет такого деления, как известно, – сложная задача даже для сравнительно простых случаев двумерных отображений когда уравнения движения известны явно [128; 129]. Процедура получения генерирующего деления становится все более сложной для систем более высоких размерностей или когда его необходимо найти непосредственно из экспериментальных данных, то есть тогда, когда уравнения движения неизвестны [129]. Наш выбор деления основан на наших предыдущих экспериментах [16] где был использован метод, описанный в [129] для анализа статистической сложности траекторий [130]. Аппроксимация генерирующего деления, полученная методом из работы [129] совпадает с той, которая была использована в настоящей работе. Другим аргументом в пользу такого выбора разбиения является тот факт, что оно приводит к максимально возможному значению топологической энтропии, $h_T = \log(3)$, определяемой использованным трехсимвольным алфавитом. Известно, что неправильный выбор деления может привести к снижению значений топологической энтропии, по крайней мере, в случае низкоразмерных динамических систем [131]. Также стоит заметить, что наше деление порождает временной ряд для траектории воды неотличимый ни от шума, ни от стандартного отображения с точки зрения топологической энтропии.

Наш дальнейший анализ касается статистических свойств коротких подпоследовательности (слов), входящих в символьную последовательность, полученную из полной траектории. В частности, анализируются распределения условных вероятностей символа, следующего за заданным словом

$$p(a_{i+1}|\{a_{i-n}\dots a_{i-1}a_i\}), \tag{2.1}$$

где a_i — символ в момент времени t_i, а i изменяется от n до количества точек в траектории. n — длина символического слова $s_i \equiv \{a_{i-n}\dots a_{i-1}a_i\}$. Этот параметр был выбран равным 9 во всех

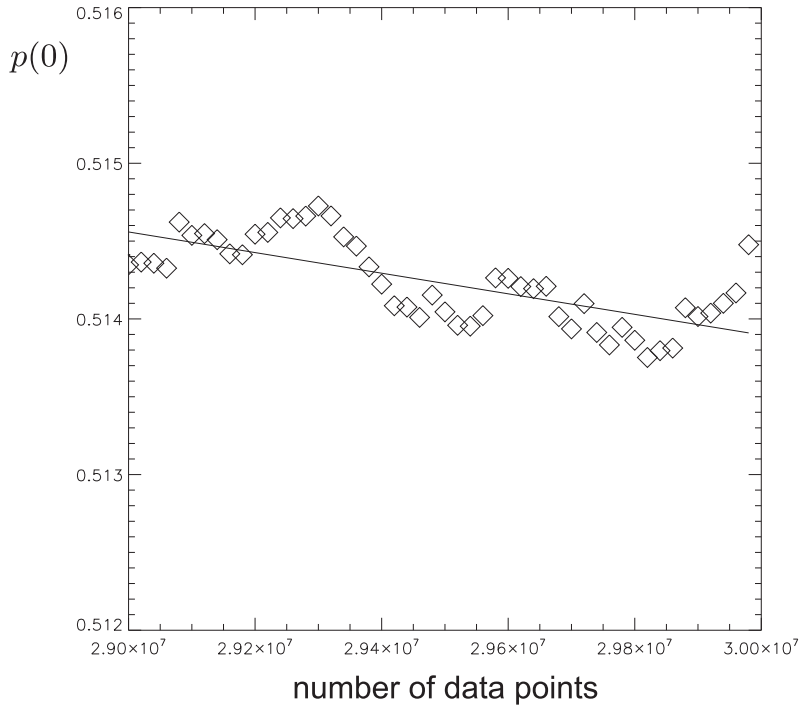


Рисунок 2.2 — Сходимость условной вероятности $p(0) = p(0|\{010200100\})$ с длиной временного ряда для произвольно выбранного символьного слова ‘010200100’; прямая показывает линейную аппроксимацию результата расчета значения $p(0)$ для различной длины символьной последовательности; коэффициент $\overline{p}(0)$ подгонки определяет наклон линии, тогда как среднеквадратичное отклонение точек от линейной подгонки обозначается как $\sigma_{p(0)}$.

результатах, представленных ниже. Выбор n=9 был обусловлен необходимостью иметь достаточно большое количество слов для получения надежной статистики.

Нас интересовало поведение (2.1) для длинных временных траекторий, изучая сходимость условных вероятностей со временем в пределе $t \to \infty$. Для этого был рассчитан наклон $\overline{p}$ линейного тренда на конечном участке длинной траектории, а также колебания σ_p среднего значения вокруг линии тренда (см. Рисунок 2.2).

Чтобы сравнить результаты, полученные при моделировании воды с аналогичными характеристиками обнаруженными в других системах с хорошо известными свойствами, были рассчитаны те же статистические показатели для (i) случайного сигнала, полученного с помощью стандартного генератора случайных чисел или по более сложному алгоритму генерации цифр троичного разложения числа π, (ii) случайного суррогата, полученного из молекулярного сигнала и (iii) двумерной последовательности точек, созданной стандартным отображением.

Для случайных сигналов были использованы псевдослучайные последовательности генерируемые специально разработанными алгоритмами. Хотя такие сигналы можно рассматривать как детерминированно-хаотичные, они генерируются численными процедурами эквивалентными многомерным динамическим системам. Поэтому они практически неотличимы в компьютерном эксперименте от подлинных случайных последовательностей получаемых из физических (аппаратных) источников шума, таких как, например, тепловые генераторы шума. Самый простой пример — случайное перемешивание строки символов, полученной из временного ряда молекулы воды. В качестве другого примера было взято троичное разложение числа π, которое хотя и гене-

рируется детерминированным алгоритмом, может быть рассмотрено как источник шума высокого качества из-за сложности алгоритма. Оба типа случайных сигналов приводили к идентичным результаты в терминах описанных условных вероятностей.

Суррогатный сигнал был получен с использованием стандартной методики [132] преобразования анализируемого временного ряда в ряд с тем же спектром мощности, но принадлежащий к классу линейных гауссовых процессов. Вычислительная процедура в простейшей реализации включает три этапа:

– выполнение преобразования Фурье данных;
– рандомизация фаз компонентов Фурье, сохранение амплитуд неизменными;
– обратное преобразование Фурье.

Выводы об отличии исходного ряда данных от сгенерированных компьютером суррогатов могут быть затем получены с помощью подходящего статистического анализа [132].

Временные ряды стандартного отображения были созданы с помощью пары детерминированных уравнений

$$\begin{aligned} P_{n+1} &= P_n + K\sin\theta_n, \\ \theta_{n+1} &= \theta_n + P_n + K\sin\theta_n, \end{aligned}$$

где P и θ вычисляются по модулю 2π и K — положительный параметр, определяющий поведение системы. Пример хаотической траектории в этой системе при значении $K = 6{,}908745$ показан на рисунке 2.3, где динамика содержит большую хаотическую область с двумя островами устойчивости, расположенными симметрично относительно начала координат. Бесконечное количество более мелких островов также существует рядом с большим, но они не различимы при таком разрешении рисунка.

Присвоение символов точкам траектории осуществлялось путем деления фазового пространства на три сегмента таким образом, что $a_i \equiv 0$ для $\theta < -\frac{\pi}{3}$, $a_i \equiv 1$ для $-\frac{\pi}{3} \leqslant \theta < \frac{\pi}{3}$ и $a_i \equiv 2$ для $\theta \geqslant \frac{\pi}{3}$.

2.1.2 Результаты и обсуждение

На рисунке 2.4 показана диаграмма рассеяния условных вероятностей (2.1), их тренд и дисперсия для случайного сигнала. Результаты для случайного суррогата, полученные из молекулярного сигнала, показаны на рисунке 2.5. Здесь статистические корреляции создают характерную структуру на графике $p(0) - p(1)$. Вращательная симметрия графика соответствует используемому трехсимвольному алфавиту. Отличие от простого случайного поведения заключающееся в появлении трех (приблизительно гауссовых) кластеров точек на рисунке 2.5 вместо одного кластера, наблюдаемого на рисунке 2.4, которые определяются наличием коротких и промежуточных временных корреляций в суррогате или, на языке спектра мощности, конечной эквивалентной шириной полосы шума, представляющего суррогатный временной ряд. Видны большие вариации

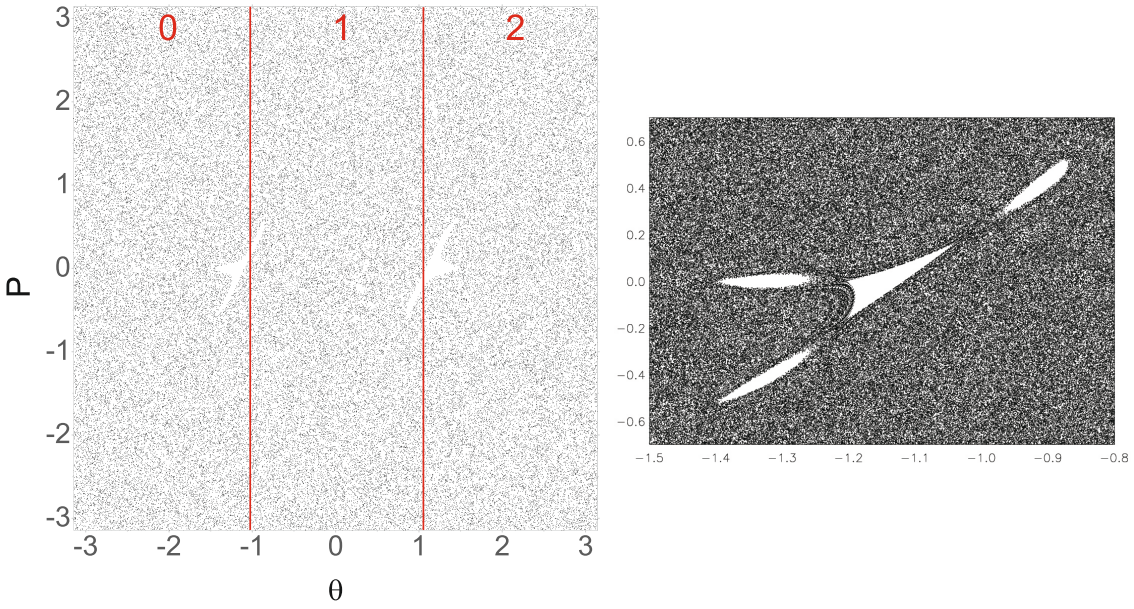


Рисунок 2.3 — Стандартное отображение; символизация с помощью трехсимвольного алфавита (слева); два периодических острова встроены в хаотичное море; один из них в увеличенном виде показан справа.

тренда и дисперсии по сравнению с некоррелированным случайным сигналом. Также наблюдается тенденция к тому, что более вероятные символические слова имеют меньшую дисперсию (правая панель на рисунке 2.5). Такое поведение может быть интерпретировано как естественное следствие статистической оценки иметь меньшую дисперсию для больших наборов данных.

Принципиально иную картину представляет молекулярная система, Рисунок 2.6. Здесь гораздо более сложная картина формируется точками условной вероятности. Распределение тренда и дисперсии существенно отличаются от найденных для случайного сигнала и суррогатного временного ряда. Кроме того, в отличие от суррогатного сигнала, здесь имеется небольшое количество слов, которые существенно более вероятны чем остальные последовательности. Эти более вероятные последовательности также имеют меньший тренд и дисперсию по сравнению с остальными последовательностями.

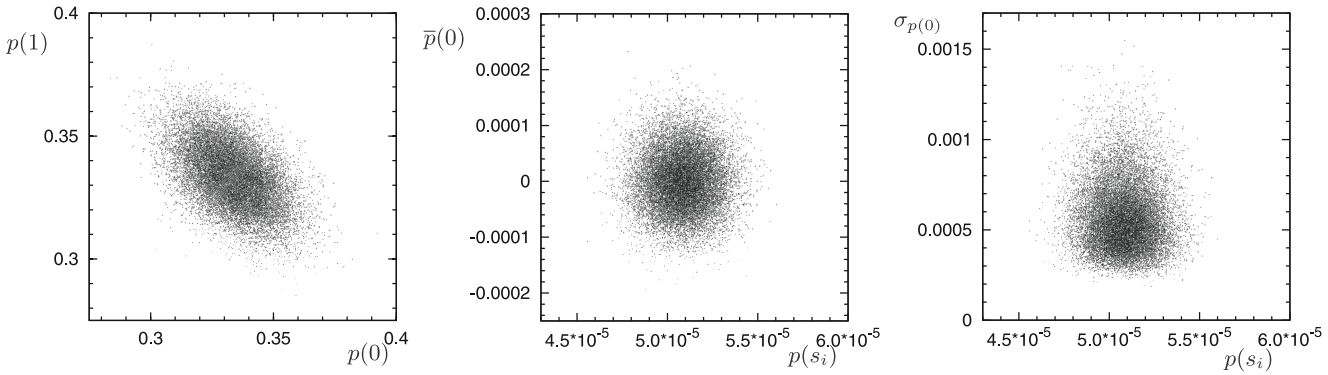


Рисунок 2.4 — Условные вероятности ‘0’, $p(0|\{\})$, и ‘1’, $p(1|\{\})$ (слева) и их тренд $\overline{p}(0)$ (в середине) и дисперсия $\sigma_{p(0)}$ (справа) как функция вхождения слова $p(s_i)$, вычисленного для символьной строки длиной $3 \cdot 10^7$ для чисто случайного процесса (троичное разложение числа π).

Результаты для стандартного отображения представляют совершенно другой случай, Рисунок 2.7 Прежде всего, сами условные вероятности образуют очень сложную картину, которая даже теряет тройную симметрию несмотря на равные вероятности появления символов в анализируемой символьной строке данных. Тренд и дисперсия показывают изменения на несколько порядков. Наша интерпретация этого результата состоит в проявлении хаотической динамики, которая производит значения условных вероятностей с большой дисперсией и, кроме того, приводит к медленной сходимости. Известно, что хаотическое движение стандартного отображения имеет участки с существенно разными статистическими свойствами, когда траектория заходит в области, близкие к островам стабильности. Динамика становится «прилипчивой» в том смысле, что траектория долгое время находится в окрестности периодических островов если динамика приводит траекторию достаточно близко к одному из них. «Прилипчивость» периодических островов, таким образом, приводит к очень длительному времени сходимости для условных вероятностей, а также сильно разбросанные значения вероятностей представляют собой сложную картину, показанную на левой панели рисунка 2.7

Наблюдаемые структуры не зависят от того как была введена символизация данных. Структура, показанная на рисунке 2.7 была получена путем введения разбиения фазового пространства, показанного на рисунке 2.3 на три вертикальные полосы одинакового размера. Сдвиг границы вдоль оси θ или изменение ориентации разбиения не вносят качественных изменений в структуры представленной на рисунке 2.7. Например, на рисунке 2.8 построенны те же распределения, но для случая разбиения фазового пространства на три горизонтальные полосы одинакового размера.

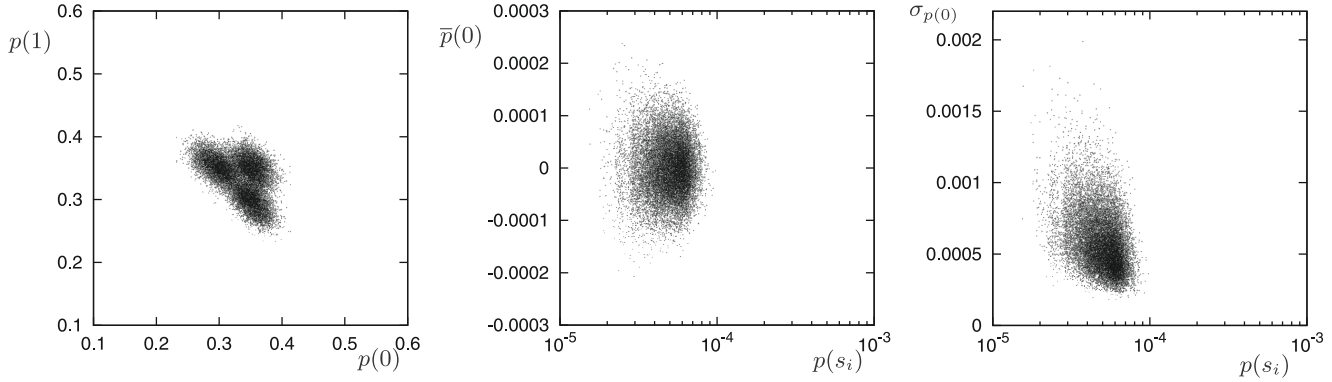


Рисунок 2.5 — То же, что и предыдущая картинка, но рассчитано для суррогатного временного ряда.

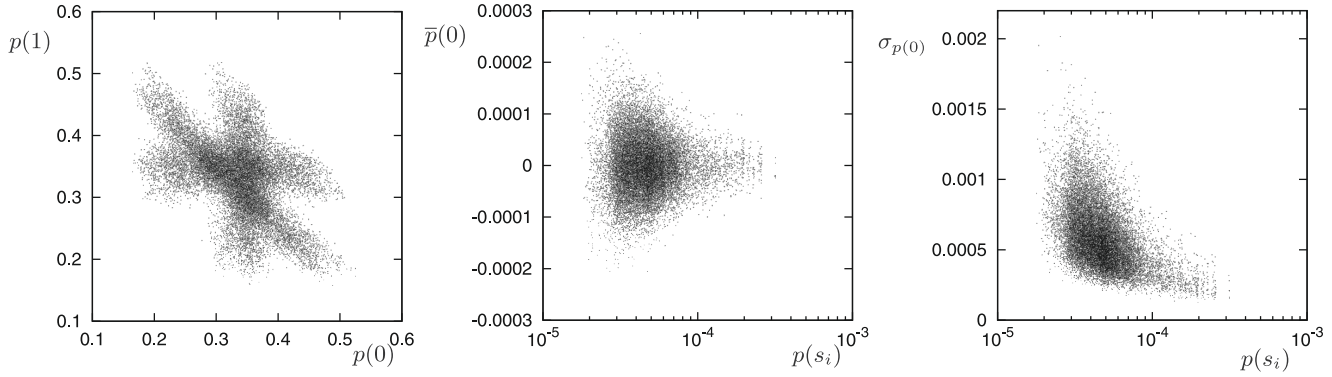


Рисунок 2.6 — То же, что и на предыдущих двух рисунках, но для сигнала от воды.

Форма закономерностей на рисунках остается качественно неизменной, что подтверждает нашу гипотезу о его происхождении из детерминированной динамики отображения.

2.1.3 Выводы

Было проанализировано поведение траекторий молекулярного фазового пространства в моделировании жидкой воды с использованием классической молекулярной динамики, символической динамики и статистики. Несмотря на однородный характер системы состоящей из идентичных молекул, были обнаружены значительные отклонения от «простого» случайного поведения на временах порядка 1μc. С другой стороны, наблюдалось определенное сходство между траекториями рассчитаными из моделирования динамики воды и динамики классического двумерного отображения (Стандартное отображение). Наш статистический подход был основан на поиске признаков хаотической динамики в такой многомерной динамической системе, как статистический ансамбль взаимодействующих молекул. Был сделан вывод, что статистические характеристики траекторий в молекулярной системе занимают промежуточное положение между случайным суррогатом и хаотическим отображением. По нашему предположению, отклонения от хаотичности в молекулярной системе вызваны его детерминированной хаотичной динамикой, включающей в себя «липкие» области фазового пространства. Липкость в отображениях с сохранением фазового обьема (Гамильтоновы отображения) — это хорошо известное свойство динамики, хорошо документированное для низкоразмерных систем [133]. Однако существует очень мало исследований этого явления для систем с многими степенями свободы.

Когда размерность системы становится большой, транспортные свойства уже не определяются непроницаемыми барьерами, образованными торами, а в результате распада тора возникают принципиально новые особенности, такие как диффузия Арнольда. Разрушенные торы имеют сложную структуру, а некоторые из них неустойчивы (аналогично седловым точкам в случае Стандартного отображения). Хаотические траектории могут быть захвачены такими структурами, поэтому наблюдаемые статистические свойства произвольной хаотической траектории могут сильно зависеть от их присутствия в фазовом пространстве. Например, было показано, что лип-

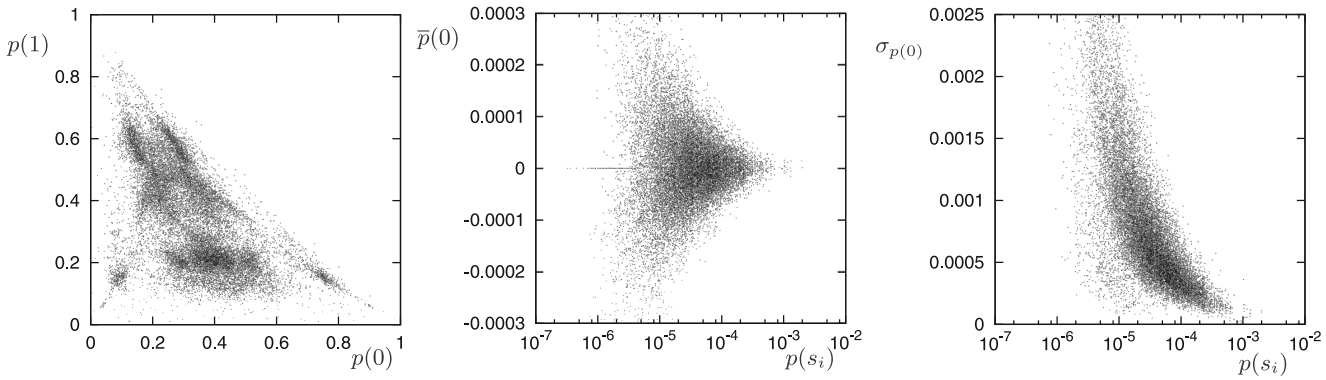


Рисунок 2.7 — То же, что и предыдущие три рисунка, но для стандартного отображения.

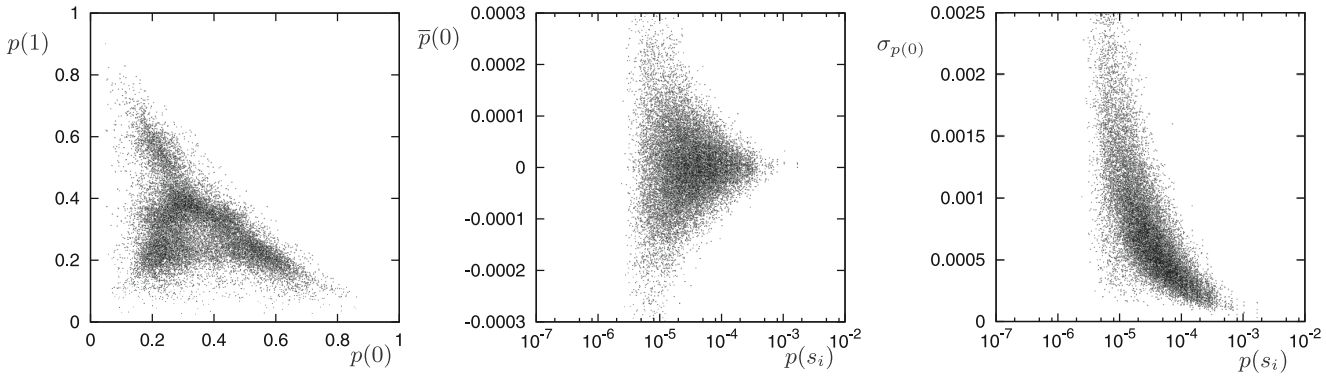


Рисунок 2.8 — То же, что и предыдущий рисунок, но для другого разбиения стандартного отображения

кость существует в бильярде Синая, хотя барьер имеет нулевую меру в фазовом пространстве [134]. Более того, статистика времени повторения, аналогичная той, что обнаружена в стандартном отображении, недавно была получена для случая диффузии Арнольда в 5-мерной гамильтоновой системе [135]. Этот результат дает возможность предположить, что свойства диффузии Арнольда могут сильно зависеть от явления «липкости».

2.2 Оценка молекулярной сложности траекторий с точки зрения динамического хаоса

Траектории атомов и молекул в жидкостях можно описать с помощью обыкновенных дифференциальных уравнений движения Ньютона. Поэтому любые сложные структуры, образованные молекулами в результате их взаимодействия, имеют геометрические аналоги в фазовом пространстве, определяемом их координатами и скоростями. Проблема идентификации и классификации закономерностей, а также прогнозирования их появления имеет решающее значение, поскольку они в конечном итоге определяют функциональность систем и могут дать ключи к пониманию фундаментальных свойств таких, например, как сворачивания белков. Однако в динамической картине молекулярных систем имеется серьезная трудность, связанная с высокой размерностью их фазового пространства. Обычно используемые подходы из нелинейной динамики, такие как показатели Ляпунова, размерности и энтропии, в большинстве случаев не работают когда движение происходит в фазовом пространстве размерности выше $\approx$10. Поэтому для многомерных систем необходимо разрабатывать новые, концептуально иные методологии.

Альтернативное описание с точки зрения вероятности и статистики может быть успешно применено во многих ситуациях к системам со слишком сложным поведением. Однако из-за особенностей построения статистики, а именно аксиоматического предположения об априорно заданных функциях распределения, она имеет ограниченный потенциал для понимания динамических закономерностей в открытых системах, демонстрирующих весьма сложное нестационарное поведение.

Недавно была предложена вычислительная механика (ВМ) — новая многообещающая концепция, направленная на построение статистического и одновременно динамического описания. Она объединяет хорошо разработанную теоретическую основу обобщенных цепей Маркова, называемых ε-машины, с концепцией краткосрочной предсказуемости, характерной для динамических систем.

Поскольку типичные движения молекул в конечном итоге определяют их конформационные перестройки, полный количественный анализ закономерностей траектории, предоставляемый ВМ, дает новое представление о молекулярных механизмах. Таким образом, наша цель — найти устойчивые структуры в фазовом пространстве образованном траекториями и интерпретировать типичное поведение таких структур с точки зрения как статистической теории, так и подхода динамических систем. Были проанализированы траектории систем моделируемых методом молекулярной динамики, в которых координаты и импульсы атомов могут быть получены с любой разумной точностью. Использование мер основанных на сложности [136; 137] и анализ вероятностных свойств символьных последовательностей, соответствующих траекториям фазового пространства, являются многообещающими альтернативами для обнаружения структур в фазовом пространстве.

Одной из самых сложных проблем при анализе многомерных молекулярных траекторий является определение понятия «структуры» или «кластера» в фазовом пространстве. Здесь эта проблема рассматривается в широком статистическом смысле, рассматривая отклонения от равномерного заполнения фазового пространства типичной траекторией как кластеры. Наличие структур в фазовом пространстве динамических систем можно интерпретировать как существование неоднородностей в инвариантной мере [138]. Последняя определяет вероятности посещения различных частей фазового пространства траекториями или, в предположении эргодичности, типичной траекторией, наблюдаемой в течение достаточно длительного периода времени. Кластеры появляются в фазовом пространстве из-за наличия большого количества резонансов, возникающих в результате нелинейных взаимодействий между атомами. Известно, что границы резонансных областей «прилипчивы» в том смысле, что любая траектория длительное время находится вблизи них. Это контрастирует с другими, нерезонансными областями, где траектории развиваются случайным образом, заполняя фазовое пространство почти равномерно.

В простых гамильтоновых системах, например, в низкоразмерных отображениях сохраняющих фазовый обьем, резонансные области появляются как острова устойчивости в фазовом пространстве. Они известны, с одной стороны, как источники неоднородности инвариантной меры, а с другой, они приводят к нарушению эргодичности из-за образования непроницаемых и «липких» барьеров в их окрестности [139]. Острова, как правило, имеют фрактальную структуру, и чем мельче масштаб субостровов, тем более «липкими» являются их границы для траекторий, то есть типичная траектория, попав в такую структуру, остается там очень долгое время.

Количественное описание неравномерности покрытия фазового пространства траекториями можно осуществить с помощью теории возвратов Пуанкаре [140]. Рассмотрим небольшой элемент $\Delta\Gamma$ фазового пространства Γ гамильтоновой системы, расположенной в небольшой окрестности точки $\mathbf{x}$. Траектория блуждает в хаотической области, посещая элемент $\Delta\Gamma$ время от времени (возвращаясь к нему). Обозначив время между последовательными посещениями как τ можно

ввести функцию распределения вероятностей повторения времен $P(\Delta\Gamma,\mathbf{x},\tau)$, которая зависит от фазового объема и положения элемента $\Delta\Gamma$, а также значения τ. Если движение эргодическое, то зависимость τ от координат $\mathbf{x}$ становится несущественной и можно ввести функцию распределения

$$P(\tau) = lim_{\Delta\Gamma\to 0} P(\Delta\Gamma,\tau)/\Delta\Gamma. \tag{2.2}$$

Для типичной хаотической траектории справедливо следующее асимптотическое соотношение:

$$P(\tau) = \frac{1}{\langle\tau\rangle} exp(-\tau/\langle\tau\rangle), \tag{2.3}$$

где $\langle\tau\rangle$ - среднее время повторения по распределению $P(\tau)$. Уравнение (2.3) может быть использовано, в принципе, для распознования областей с хаотическим движением от областей, близких к липким, путем введения разбиения фазового пространства на неперекрывающиеся объемы и анализа распределений $P(\tau)$ для каждого из них. Отметим также, что задача выбора размеров и форм элементов разбиения не является тривиальной и в общем случае распределение времен возврата Пуанкаре может зависеть от расположения и формы области $\Delta\Gamma$.

Важно также, что липкость резонансных областей приводит к аномальным транспортным свойствам траекторий в фазовом пространстве. Этот вопрос привлекает большое внимание в последнее время [139] и было продемонстрировано, что ключевые сведения о деталях транспортировки можно получить с помощью теоремы Пуанкаре о возвратах.

В этом разделе показано как анализ молекулярных траекторий с точки зрения ε-машины, связанный с понятием статистической сложности (SC), обеспечивает связь от чисто статистического описания с моделированием типа цепи Маркова к теории динамических систем, основанной на анализе рекуррентности Пуанкаре. Одним из актуальных вопросов при анализе траекторий очень высокой размерности с помощью возвратов Пуанкаре является выбор областей для исследования. Другими словами, неясно, как создать разбиение в фазовом пространстве, которое дало бы осмысленное описание динамики с точки зрения времен возвратов траекторий. Показано, что обусловливающие состояния, составляющие ядро ε-машины обеспечивают «естественное» разбиение в фазовом пространстве в том смысле, что для каждого обусловливающего состояния времена возврата Пуанкаре распределяются в соответствии с уравнением (2.3). Поэтому каждое обусловливающее состояние можно рассматривать как элемент фазового пространства, содержащего набор хаотических циклов Пуанкаре. Более того, учет отклонения от уравнения (2.3) позволяет идентифицировать различные типы движения в фазовом пространстве, то есть находить обусловливающие состояния с особыми свойствами, которые соответствуют резонансным областям внутри хаотического моря в фазовом пространстве. Таким образом, цель количественного различения между областями с качественно различной динамикой может быть достигнута путем сравнения времен затухания возвратов Пуанкаре с обусловливающими состояниями.

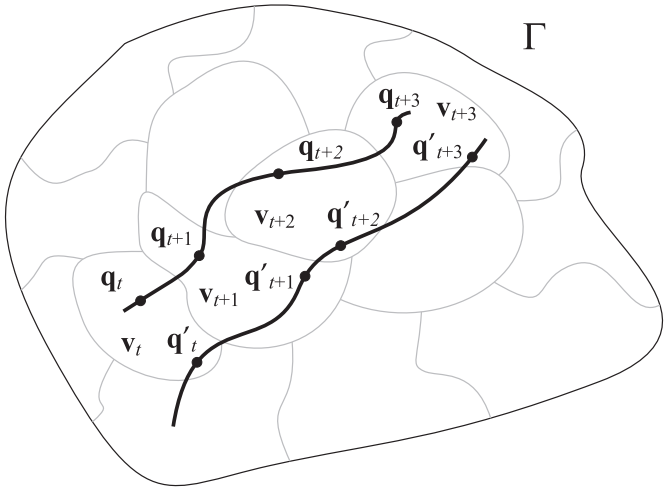


Рисунок 2.9 — Иллюстрация вырожденности макронаблюдаемой проекции траектории полномерного фазового пространства. Та же последовательность наблюдаемой (скорости) $\{\mathbf{v}_t\mathbf{v}_{t+1}\mathbf{v}_{t+2}\mathbf{v}_{t+3}\}$ генерируется двумя различными частями траектории фазового пространства $\{\mathbf{q}_t\mathbf{q}_{t+1}\mathbf{q}_{t+2}\mathbf{q}_{t+3}\}$ и $\{\mathbf{q}'_t\mathbf{q}'_{t+1}\mathbf{q}'_{t+2}\mathbf{q}'_{t+3}\}$.

2.2.1 Вычислительная механика имеет дело с определенным разбиением молекулярного фазового пространства

Первоначальная символизация создает очень грубое разбиение

Молекулярная траектория, полученная в ходе вычислительного эксперимента, представляет собой ряд точек $2N$-мерного фазового пространства $\mathbf{q}_i$, где N — число степеней свободы системы, т.е. число атомов, умноженное на 3, за вычетом различных ограничений, таких как фиксированные длины связей, углы и т.д. Для реалистичного моделирования MD N имеет порядок нескольких тысяч. Таким образом, молекулярная траектория представляет собой очень многомерный объект. Точки генерируются системой вдоль траектории в фиксированные моменты времени (Рисунок 2.9).

Для анализа данных неизбежно рассматриваются низкоразмерные наблюдаемые величины, например скорость атома $\mathbf{v}$. Это проекция полноразмерной траектории на низкоразмерную наблюдаемую, в данном случае $\mathbf{v}$. Эта низкоразмерная проекция траектории фазового пространства является вырожденной, то есть существует очень много различных реализаций траектории производящих один и тот же ряд значений низкоразмерной проекции $\mathbf{v}$ (Рисунок 2.9). Это вызвано (i) дискретностью временной выборки траектории, (ii) конечной погрешностью измерений $\mathbf{v}$, и (iii) независимостью $\mathbf{v}$ в каждый отдельный момент времени от некоторых других степеней свободы, например от положений и скоростей атомов находящихся на большом расстоянии. Следовательно, все фазовое пространство Γ разбивается на области таким образом, что на каждой из них макроскопически наблюдаемая $\mathbf{v}$ принимает единственное значение, в то время как полноразмерные точки $\mathbf{q}_i$ могут иметь разные значения (Рисунок 2.9).

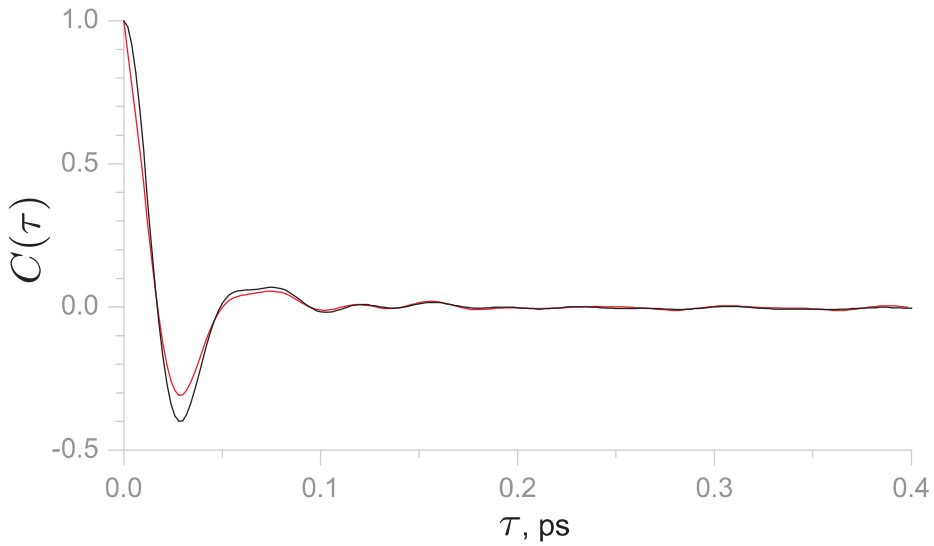


Рисунок 2.10 — Автокорреляционные функции $C(\tau) \equiv \frac{1}{T}\sum_t^T \mathbf{v}_t \cdot \mathbf{v}_{t+\tau}$ для исходной скорости водорода в объемной воде (черный) и сигнала, состоящего из 27 символов (красный).

Значения наблюдаемых переменных дискретны и конечны. Другими словами, мы имеем дело с набором счетного числа *символов* . Например, в случае представления чисел с плавающей точкой на компьютере количество символов велико, но ограничено и определяется точностью, используемой при моделировании (одинарная, двойная и т. д.). Конечная точность $\mathbf{v}$ приводит к конечному (но большому) набору возможных значений. Более того, легко проверить, что даже очень грубое представление $\mathbf{v}$ дает практически одинаковые характеристики анализируемых молекулярных сигналов. На рисунке 2.10 показана одна из таких характеристик – общая функция автокорреляции скорости для сигнала, где координаты скорости заменяются только тремя значениями в $\mathbf{v}_x$,$\mathbf{v}_y$, и $\mathbf{v}_z$, так что $\{x \equiv -1$, если $x < -1; x \equiv 0$, если $-1 \leqslant x < 1; x \equiv 1$, если $x \geqslant 1\}$, где x – это $\mathbf{v_x}$,$\mathbf{v}_y$ и $\mathbf{v}_z$. Общее число возможных значений результирующего крупнозернистого вектора составляет $3^3 = 27$, то есть сигнал может быть представлен всего 27 символами. Тем не менее, автокорреляционная функция этого сигнала очень похожа на исходную, рассчитанную на основе значений $\mathbf{v}$ двойной точности.

Такое представление динамики в терминах символов из алфавита конечного размера называется «символьной динамикой» и является предметом изучения в математической области с таким же названием [141].

Подводя итог, можно сказать, что путем преобразования непрерывной траектории в символьную последовательность создается крупнозернистое разбиение фазового пространства.

Динамика делает разбиение более точным

Эволюция точек фазового пространства $\mathbf{q}$, выбираемых в моменты времени t, управляется оператором $\mathbf{T}$: $\mathbf{q}_{t+1} = \mathbf{T}\mathbf{q}_t$. Из-за детерминизма динамика $\{\mathbf{q}_t\}$ образует цепь Маркова. Рассматривая ансамбль таких динамических систем, обозначим случайную величину, представляющую

текущее микросостояние, как $\mathbf{Q}$, то есть набор всех возможных значений точек фазового пространства, имеющих вероятности, генерируемые динамикой $\mathbf{T}$.

Макроскопическая наблюдаемая переменная A является функцией f микросостояния $\mathbf{Q}$ (например, мгновенная температура $\frac{1}{Nk}\sum_i m_i \mathbf{v}_i^2$, где N — число степеней свободы, k — постоянная Больцмана, m_i — массы атомов, а $\mathbf{v}_i$ — их скорости). Как обсуждалось ранее, функция f разбивает фазовое пространство Γ на взаимоисключающие и совместно полные множества, в каждом из которых f принимает уникальное значение. Обозначим разбиение Γ индуцированное f как $\mathcal{F}$. Наблюдаемый процесс $A_t = f(\mathbf{Q})$ не обязательно является марковским (Рисунок 2.11).

Что же происходит с этим разбиением, если вместо отдельных значений A рассматривать последовательности A_t? Рассмотрим момент времени t, A_t. Соответствующий набор точек в Γ есть $\mathcal{F}_t$. Для последовательности из двух наблюдений в текущий и предыдущий моменты времени набор точек имеет вид

$$\mathcal{F}_t \cap \mathbf{T}\mathcal{F}_{t-1}, \tag{2.4}$$

что является уточнением разбиения $\mathcal{F}$. Эту процедуру можно повторять любое счетное число раз, тем самым получая уточненные разбиения для историй макронаблюдаемого A. Таким образом, динамика создает начальное разбиение, вызванное макронаблюдаемым, тем точнее, чем длиннее последовательность $\{A_t\}$ («история») тем мельче разбиение, сгенерированное последовательностью.

Вычислительная механика огрубляет разбиение

Следующий шаг — применение специальной статистики, называемой вычислительной механикой (ВМ) [130], к наблюдаемому A. Строгое определение ВМ дано в Приложении Г. Здесь приведено описание части подхода, необходимой для ответа на основной вопрос, сформулированный во введении.

Рассматриваются все прошлые A_i^- и будущие A_i^+ половины бесконечных в обе стороны последовательностей макронаблюдаемой величины, центрированные на моменты времени i. Два прошлых события A_1^- и A_2^- определяются как эквивалентные, если условные распределения по их будущим событиям $P(A^+|A_1^-)$ и $P(A^+|A_2^-)$ равны. *Обусловливающее состояние* $\varepsilon(A_i^-)$ — это набор всех прошедших значений, эквивалентных A_i^-: $\varepsilon_i \equiv \varepsilon(A_i^-) = \{\lambda : P(A^+|\lambda) = P(A^+|A_i^-)\}$. В данный момент система находится в одном из обусловливающих состояний и переходит в следующее с вероятностью, заданной матрицей перехода $T_{ij} \equiv P(\varepsilon_j|\varepsilon_i)$. Матрица перехода определяет асимптотические вероятности обусловливающих состояний как ее левый собственный вектор $P(\varepsilon_i)T = P(\varepsilon_i)$, где $\sum_i P(\varepsilon_i) = 1$. Совокупность обусловливающих состояний вместе с вероятностями перехода определяет *ε-машину*. *Статистическая сложность* является информационной мерой размера ε-машины: $C_\mu = H[P(\varepsilon_i)]$, где P — вероятности обусловливающих состояний, а H — энтропия Шеннона распределения случайной величины ν, $H[P(\nu)] \equiv -\sum_\nu P(\nu)\log_2 P(\nu)$.

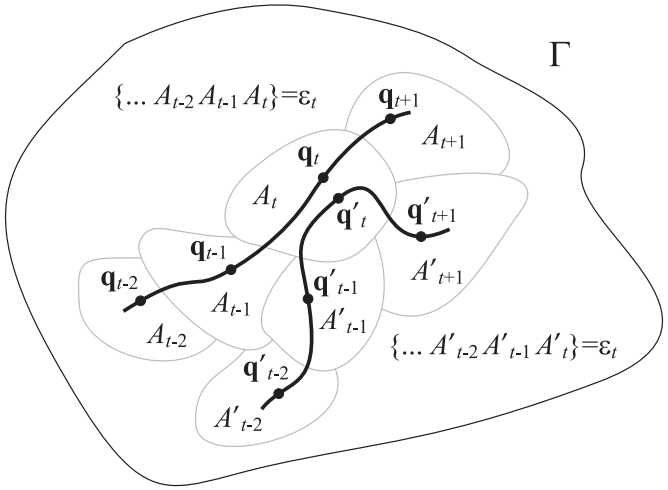


Рисунок 2.11 — Схематическое изображение последовательностей, используемых для определения формулы (2.5); точки фазового пространства $\{\mathbf{q}\}$ и $\{\mathbf{q}'\}$ двух частей траектории образуют последовательности Маркова соответствующие последовательности наблюдений A и A' не являются марковскими, так как одно и то же значение A_t приводит к разным A_{t+1} и A'_{t+1} в зависимости от предыдущих значений A_{t-1} и A'_{t-1}; однако, если обе истории $\{\dots A_{t-2}A_{t-1}A_t\}$ и $\{\dots A'_{t-2}A'_{t-1}A_t\}$ принадлежат к одному и тому же обусловливающему состоянию ε_t, следующее обусловливающее состояние ε_{t+1} определяется без знания ε_{t-1}, таким образом делая $\{\varepsilon\}$ последовательностью Маркова.

Таким образом, суть ВМ заключается в группировке историй $\{A_t\}$ в обусловливающие состояния. С точки зрения разбиений фазового пространства это соответствует объединению ячеек Γ, вызванные динамикой. Важно отметить, что новые ячейки представляют собой марковский процесс, построенный на основе наблюдаемого процесса A_t путем построения ε-машины на A. По определению ε-машинные последовательности обусловливающих состояний $\{\varepsilon\}$ создают цепь Маркова (Рисунок 2.11).

Разбиение, созданное вычислительной механикой, является наиболее информативным

Shalizi и Moore [142] показали, что в этой ситуации статистическая сложность имеет ясный физический смысл: она количественно определяет объем информации о микросостоянии

$$C_\mu = I[\mathbf{Q}; \varepsilon], \tag{2.5}$$

содержащейся в новом сконструированном макронаблюдаемом процессе $\{\varepsilon_i\}$, где I - взаимная информация между случайными величинами X и Y: $I[X;Y] = H[X] - H[X|Y]$; а $H[X|Y]$ - условная энтропия X при заданном Y: $H[X|Y] = -\sum P(X) \sum P(X|Y) \log_2 P(X|Y)$.

Это происходит потому, что знание микросостояния точно определяет макронаблюдаемое: $H[\varepsilon|\mathbf{Q}] = 0$, потому что все истории содержатся в ε_t и соответствующее разбиение $\mathbf{Q}$ однозначно

определяет следующее состояние ε_{t+1} (определение ε-машины). Из этого равенства и уравнений $H[X] + H[Y|X] = H[Y] + H[X|Y]$, (2.5) следует:

$$\begin{aligned} H[\mathbf{Q}|\varepsilon] + H[\varepsilon] &= H[\varepsilon|\mathbf{Q}] + H[\mathbf{Q}] \\ H[\mathbf{Q}|\varepsilon] + C_\mu &= H[\mathbf{Q}] \\ C_\mu &= H[\mathbf{Q}] - H[\mathbf{Q}|\varepsilon] \\ C_\mu &= I[\mathbf{Q};\varepsilon]. \end{aligned}$$

Из-за свойств ε-машины это – максимальная информация, которую можно извлечь из выбранного макронаблюдаемого и его начального разбиения.

Три стадии символизации

Подводя итог, можно сказать, что разбиение фазового пространства, которое было использовано в численных экспериментах, конструируется в три этапа.

I. Наблюдаемая макропеременная индуцирует начальное (обычно очень грубое) разбиение фазового пространства, определяемое процедурами проецирования, ошибками измерений и символизацией.

II. Элементы этого разбиения уточняются динамикой, когда рассматриваются слова (истории) вместо отдельных символов (2.4). Также следует отметить, что рассмотрение слов вместо символов аналогично реконструкции многомерного фазового пространства из скалярного временного ряда с помощью процедуры Такенса [143]. С этой точки зрения истории соответствуют разным точкам фазового пространства, а длина истории l равна размерности подпространства.

III. Уточненные элементы разбиения (истории) далее группируются с помощью процесса реконструкции ε-машины, тем самым обеспечивая окончательный разбиение, являющееся минимальным, уникальным и наиболее информативным (учитывая начальное разбиение Г).

2.2.2 Численная реализация

Моделирование молекулярной динамикой

В последующих разделах разработанная теоретическая основа была применена к анализу динамики в статистическом ансамбле взаимодействующих молекул воды. Молекулярная динамика — это метод численного решения уравнений Ньютона, описывающих изменения

координат атомов во времени $\mathbf{x}$ и скорости $\mathbf{v} = \dot{\mathbf{x}}$: $\dot{\mathbf{v}} = -\frac{1}{m}\mathbf{F}$. Сила $\mathbf{F}$ выводится из заданного потенциала межатомного взаимодействия V (также называемого «силовым полем»): $\mathbf{F} = -\nabla V(\mathbf{x}_i)|_{i=1..N}$, который есть функция всех координат атомов. Обычно используемые силовые поля представляют собой эмпирические функции, представляющии собой тщательный баланс между сложностью воспроизведения реалистичных межатомных взаимодействий и вычислительной эффективностью. Параметры силовых полей калибруются для воспроизведения либо строгих квантово-механических расчетов, либо экспериментальных термодинамических данных.

В этом разделе объемная вода (периодические граничные условия), состоящая из 392 или 878 SPC или SPC-E [125] молекул были промоделированы с использованием пакета молекулярной динамики `GROMACS` [126]. Температура систем поддерживалась постоянной на уровне 300K с использованием термостата Берендсена [127] или Nose-Hoover [94], с различными константами связи. Перед сбором данных для анализа было выполнено достаточное уравновешивание. В качестве анализируемого сигнала использовалась скорость атома водорода одной из молекул воды. В местах, где скорость пересекает плоскость xy, были сгенерированы точки двумерного отображения, которые использовались в качестве исходного непрерывного сигнала для анализа.

Было обнаружено, что результаты не зависят от параметров молекулярного моделирования таких как силовое поле, температура, тип термостата, количество молекул и т. д. [16].

Символизация

Как описано выше, наилучшее возможное начальное разбиение для преобразования данных временного ряда с плавающей точкой двойной точности в символьную строку может быть достигнуто с помощью генерирующего разбиения (GP). Хотя точно задать его невозможно, существуют методы его приблизительного вычисления. GP обеспечивает разбиение, которое сохраняет всю информацию о сигнале. Поэтому, чем ближе используемое приближение к GP, тем больше информации переносится из непрерывного сигнала в символьную последовательность.

Для начального приближения к GP было выбрано разбиение, полученное путем применения метода, описанного в [129]. Пример расчетов этим методом для двумерного сечения наших (трехмерных) данных о скорости с использованием 2, 3, 4 и 5 элементов разбиения показан на рисунке 2.1. Для всех случаев полученные приближения к GP являются центрально-симметричными (вероятно, из-за центральной симметрии распределения точек данных). Симметрию двумерного набора точек можно дополнительно проиллюстрировать, преобразовав данные в полярные координаты $(x,y) \to (\rho,\varphi)$ и оценив плотность вероятности $w(\varphi)$ для случайной величины φ. Гистограмма, соответствующая такому $w(\varphi)$ распределению показана на рисунке 2.12. Очевидна почти идеальная однородность функции распределения, что оправдывает выбор центрально-симметричных разбиений, которые были использованы во всех последующих расчетах. Результаты для трех и более разбиений (количества символов) качественно одинаковы, см. [16]. Во всех представленных

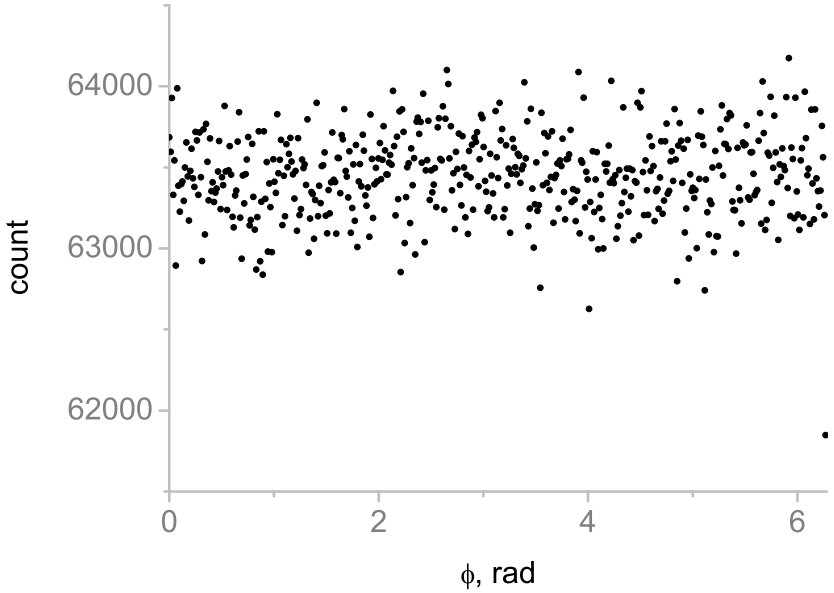


Рисунок 2.12 — Гистограмма случайной величины φ иллюстрирующая равномерное симметричное распределение точек в плоскости поперечного сечения $v_z = 0$ траектории скорости водорода.

результатах был использован трехсимвольный алфавит. Промоделированная траектория длинной в 1μs дала примерно $3 \cdot 10^7$ точек данных (символов).

Реконструкция ε-машины: CSSR

На следующем этапе анализа описание переходило с рассмотрения отдельных символов в символьной строке на изучение историй (слов символов конечной длины) и построение ε-машины. Для этой цели был использован метод, разработанный Shalizi с соавторами, которые также предложили алгоритм реконструкции ε-машина из заданного ряда данных [144]. В общем случае КМ формулируется с использованием предположения о бесконечно длинных прошлых и будущих подисториях. На практике необходимо выбрать конечную длину истории l, и это один из настраиваемых параметров алгоритма CSSR. Число возможных историй растет экспоненциально с длиной истории. Поэтому для длинных историй также необходимо экспоненциальное увеличение числа точек данных.

Вторым параметром алгоритма CSSR является уровень значимости σ используется при сравнении распределений $P(\overrightarrow{s}|\overleftarrow{s}_i)$ для группировки историй в обусловливающие состояния по их предсказательным свойствам (используется критерий Колмогорова-Смирнова). Слишком большие значения σ (слишком строгий порог для того, чтобы два распределения считались эквивалентными) приводят к искусственному появлению слишком большого количества обуславливающих состояний. Это эквивалентно недостаточно большой выборке историй. Та же ситуация имеет место при слишком большой длине истории, поскольку число возможных историй слишком

велико, а для умеренно длинных экспериментальных временных рядов распределения $P(\overrightarrow{s}|\overleftarrow{s}_i)$ становятся статистически незначимыми.

Поэтому для получения надежных результатов представляется необходимым провести анализ ε-машины как функции этих двух параметров. Слишком длинная история или слишком большой значение σ приводит к статистически неверным результатам. Как рекомендуют авторы CSSR, величину σ следует выбирать таким образом, чтобы существовало «плато» в числе обусловливающих состояний как функции l. Если таких значений l несколько, то следует выбрать наименьшее (согласно принципу минимальности ВМ). Это постоянное значение l является «истинным» значением длины истории для стабильной архитектуры ε-машины.

Наш анализ сходимости алгоритма CSSR с длиной истории представлен на рисунке 2.13. Здесь сравниваются временные ряды, полученные в результате MD моделирования воды, с так называемыми суррогатными данными, которые имеют идентичный спектр мощности и, следовательно, функцию автокорреляции, но являются стохастическими в смысле отсутствия динамических корреляций между соседними точками [132]. Результаты показывают, что выбор плато — нетривиальная задача. Суррогатные данные (левая панель) демонстрируют четкое плато при любом заданном значении длины временного ряда, но минимальное значение l растет от ≈ 3 в ≈ 6 при увеличении длины данных с 60нс ($\approx$ 2 000 000 символов) к 1μс ($\approx$ 30 000 000 символов). Плато для сигнала воды (правая панель на рисунке 2.13) менее выражено, но все еще видно при низких значениях σ что соответствует, например, $l = 10$ для длины данных выше $\approx 0{,}45\mu$. Очень медленная сходимость ε-машины с размером данных для молекулярных сигналов обсуждалась в [16; 17; 20]. Результаты для $l = 10$ находятся на пределе статистической надежности для использованных длин данных. Поэтому в обсуждаемом здесь анализе были использованы $l = 9$. Однако стоит отметить, что основной результат этой работы, то есть разделение обусловливающих состояний на две группы, не зависит от значения l, начиная со значения $l \approx 6$.

2.2.3 Результаты

Как было указано в разделе 2.2.1 анализируемые символьные слова (истории) соответствуют элементам разбиения фазового пространства на непересекающиеся области. Дальнейшее объединение историй в обусловливающие состояния приводит к более грубому разбиению, которое обладает определенными марковскими свойствами и определяет ε-машину через функцию распределения их частот появления $P(\varepsilon_i)$. Чтобы получить более глубокое представление о связи между ε-машиной и динамикой были проанализированы распределения времен возврата для множества обусловливающих состояний, рассматривая их как элементы разбиения фазового пространства. Чтобы ввести времена возврата рассматриваются временные интервалы между последовательными появлениями обусловливающего состояния в символьном временном ряду. Для всех проанализированных данных сначала был определен набор обусловливающих состояний, а затем проанализированы гистограммы времени возврата (периодов) для каждого из них.

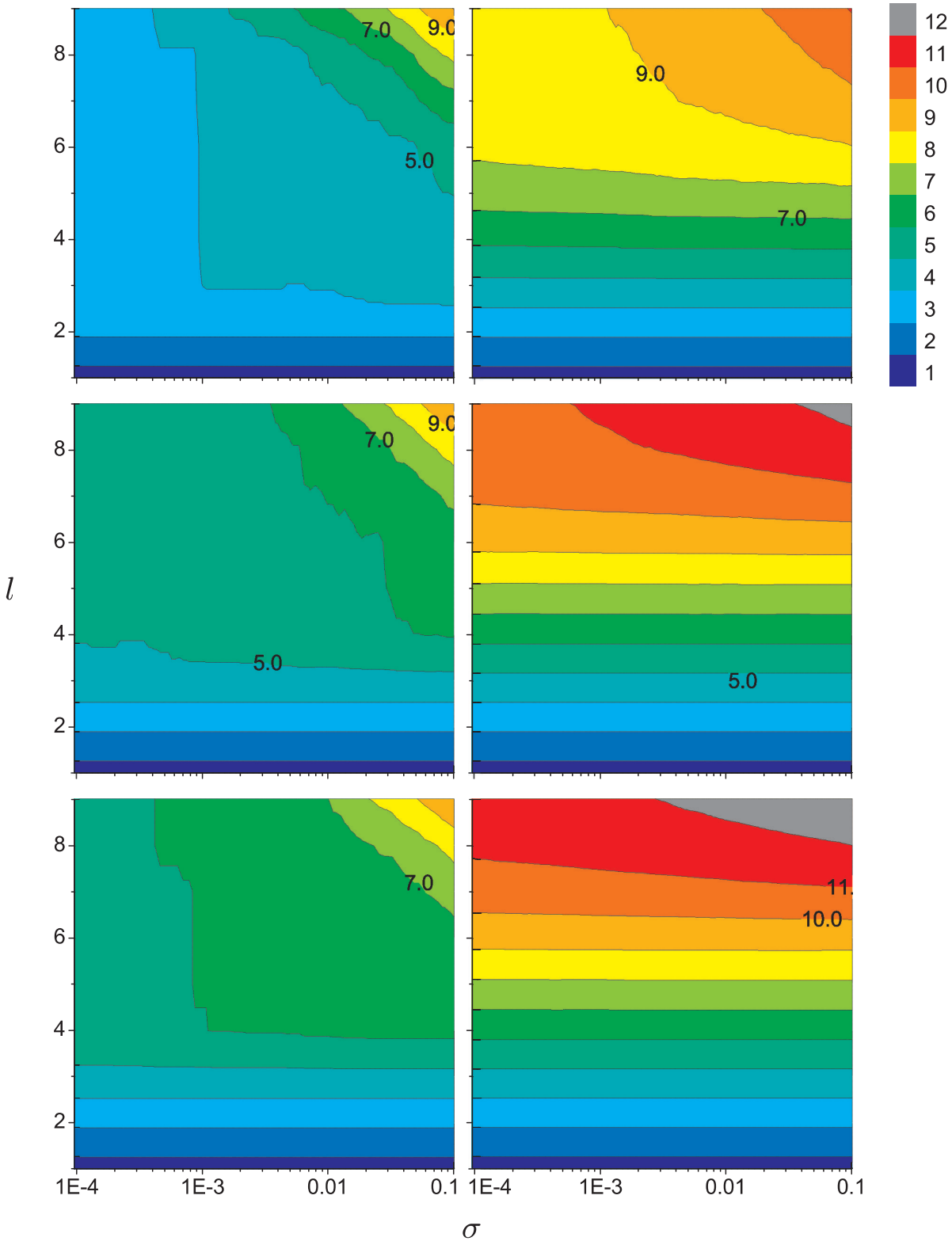


Рисунок 2.13 — Логарифм числа обусловливающих состояний $\log_2 n_{st}$ в ε-машине как функция длины истории l, погрешности σ и длительности временного ряда; слева: суррогатный временной ряд Фурье [132], справа: молекулярный сигнал; вверху: длительность временного ряда 60 нс, посередине: 450 нс, внизу: 1μs.

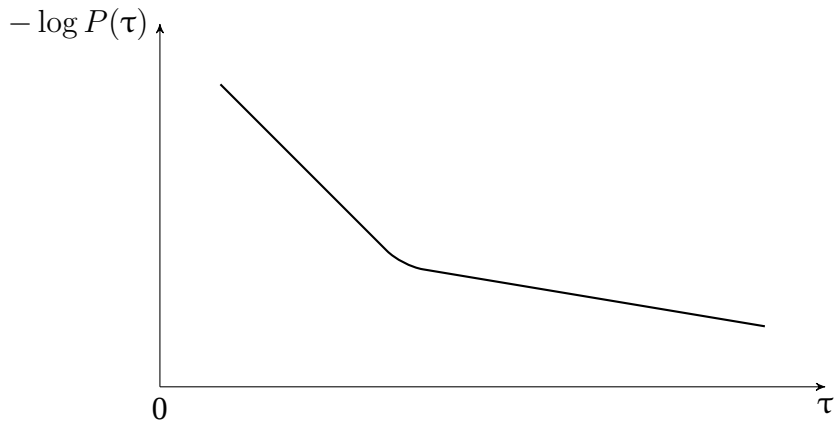


Рисунок 2.14 — Схема функции распределения вероятностей для времени возврата Пуанкаре τ. Экспоненциальный спад при малых τ изменяется на, по-видимому, более медленный закон спада на больших временах.

Как следует из теории Пуанкаре, времена возврата для хаоса в системах, где существуют только хаотические движения, распределены в соответствии с пуассоновским распределением (уравнение (2.3)). Оказывается, что в случае хаотической динамики в сохраняющих обьем отображениях с разделенным фазовым пространством все обстоит иначе [145], области хаоса сосуществуют с периодическим поведением (периодические острова). В таких системах экспоненциальный спад функции распределения переходит в степенной закон

$$P(\tau) \propto \tau^{-\gamma} \tag{2.6}$$

на очень больших временах, как показано на рисунке 2.14.

В численном анализе функция $P(\tau)$ изображенная на рисунке 2.14 обычно вычисляется путем усреднения соответствующих функций распределения по многим траекториям, выбранным случайным образом путем задания произвольных начальных условий. Это соответствует анализу времени повторения во многих случайно расположенных небольших областях, сосредоточенных вокруг случайно выбранных центров. В наших расчетах предполагалась эргодичность системы и вычислялись подобные распределения, рассматривая только один временной ряд. Также были выбраны области для анализа возвратов Пуанкаре, совпадающие с разбиением фазового пространства, налагаемым набором обусловливающих состояний. Удивительно, но было обнаружено, что для каждого обусловливающего состояния распределение времен возврата при достаточно больших временах хорошо аппроксимируется экспоненциальной функцией, т.е. каждое обусловливающее состояние содержит набор статистически независимых циклов Пуанкаре.

Для того, чтобы сравнить наши результаты относящиеся к анализу траекторий MD в воде с имеющимися в литературе, была выбрана хорошо документированная двумерная системя сохраняющая обьем, известная как Стандартное отображение (иногда называемая отображением Тейлора-Грина-Чирикова) [146]. Оно определяется как преобразование плоскости в саму себя

$$\begin{aligned} P_{n+1} &= P_n + K \sin \theta_n, \\ \theta_{n+1} &= \theta_n + P_n + K \sin \theta_n, \end{aligned}$$

где P и θ вычисляются по модулю 2π и K — положительный параметр, контролирующий различные виды поведения, которые может демонстрировать система. Для примера хаотической траектории в этой системе был рассчитан временной ряд при значении параметра $K = 6{,}908745$ где фазовый портрет имеет большую хаотическую область, содержащую два острова устойчивости, симметрично расположенных относительно начала координат. По соседству с крупными островами также существует бесконечное множество более мелких островов, что усложняет динамику в их окрестностях. Кроме того, было применено простое равномерное разбиение переменной θ на три равных интервала, т.е. трехсимвольный алфавит аналогично случаю анализа временных рядов воды.

Наши численные эксперименты как для временных рядов воды, так и для стандартного отображения показывают, что обусловливающие состояния демонстрируют четкое разделение на два класса, которые будем называть «периодическими» состояниями (т.е. теми, которые определяются временами возвратов Пуанкаре, затухающими гораздо медленнее по сравнению с уравнением (2.3)) и «хаотические» (демонстрирующие экспоненциальный распад в соответствии с законом Пуанкаре). Для количественной оценки разницы между двумя классами введем безразмерный параметр D, равный расхождению между показателем затухания λ рассчитаному на основе гистограммы времен возврата и ее «нормального» значения $\frac{1}{\langle\tau\rangle}$ определяемого уравнением (2.3)

$$D = \frac{1}{\lambda\langle\tau\rangle} - 1, \tag{2.7}$$

где λ — показатель степени, определяющий форму функции распределения

$$P(\tau) \propto exp(-\lambda\tau), \tag{2.8}$$

рассчитанный численно. Большие значения D указывают на сильное расхождение между рассчитанным значением показателя степени в уравнении (2.8) и ожидаемоего значения $1/\langle\tau\rangle$.

Хвост степенного закона в распределении времен возвратов (уравнение (2.6)) можно наглядно представить для случая Стандартного отображения если построить соответствующее распределение, усредненное по всем обуславливающим состояниям, составляющим ε-машину (Рисунок 2.15a). Видны два сегмента в функции распределения, соответствующей нормальному хаотическому поведению, определяемому уравнением (2.3) и аномально длинные повторения в хвосте. Однако в случае временных рядов воды не наблюдается явного различия между двумя типами поведения, как показано на рисунке 2.15b. В простейшем приближении это можно интерпретировать как отсутствие периодических островов в фазовом пространстве воды из-за разрушения всех инвариантных торов, которое происходит из-за взаимодействия между резонансами в системе с несколькими степенями свободы. Однако следует отметить, что расчет средней гистограммы по всем обусловливающим состояниям эквивалентен усреднению частот возврата по всем доступным областям фазового пространства. Поэтому движение в окрестности нескольких периодических островов, посещаемых траекторией умеренно часто, может маскироваться более частыми хаотическими движениями. Таким образом, наш анализ ансамбля обусловливающих состояний с точки зрения параметра D обеспечивает альтернативный подход, позволяющий обнаруживать периодические острова в хаотическом море, проводя более тонкое различие между периодической и хаотической фазами движения.

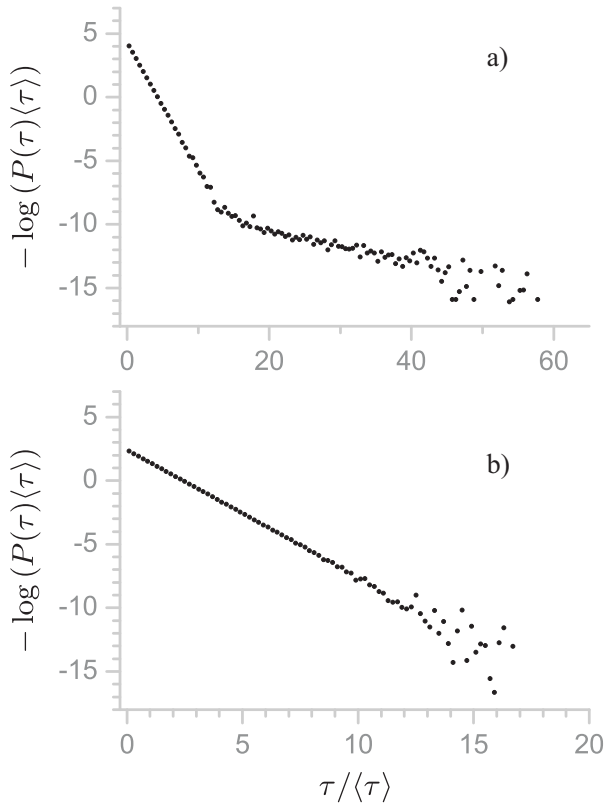


Рисунок 2.15 — Распределение времен возврата Пуанкаре, усредненное по всем обусловливающим состояниям; a) Стандартное отображение; b) временной ряд воды, соответствующий скорости атома водорода.

На рисунке 2.16 показаны диаграммы рассеяния, представляющие кластеризацию обусловливающих состояний в два класса по параметру D. Горизонтальная ось аппроксимирует частоту появления (или вероятность $P(\varepsilon_i)$) обуславливающих состояний, то есть для каждого из них было подсчитано количество его появлений в символьном временном ряду и оценена вероятность $P(\varepsilon_i)$, разделив его на общую длину символьного ряда.

Интересно также отметить, что гистограмма для «периодических» состояний имеет ярко выраженный пик на значении около 0,1 пс (см. Рисунок 2.17b), тогда как для остальных обуславливающих состояний они характеризуются всего лишь простой экспоненциальной функцией (Рисунок 2.17e, f). Сочетание пика на коротких временах с медленно затухающим хвостом на длинных временах на гистограмме свидетельствует о том, что траектория довольно редко посещает окрестности периодических островов, но, как только она попадает в ловушку липкой границы, она испытывает последовательность из нескольких коротковременных возвратов. Обуславливающие состояния, характеризующиеся низким значением D, имеют строго экспоненциальное распределение времен возврата и не обладают выраженной периодичностью низкого порядка.

Кратковременные возвраты обуславливающих состояний с высоким значением D обусловлены их квазипериодическим характером, очевидным в символических последовательностях, составляющих эти состояния. Вот они: ‘000000000’, ‘222222222’, ‘001001001’, ‘010010010’, ‘100100100’, ‘122122122’, ‘212212212’, ‘221221221’. Во всех них наблюдается повторение с максимальным периодом 3.

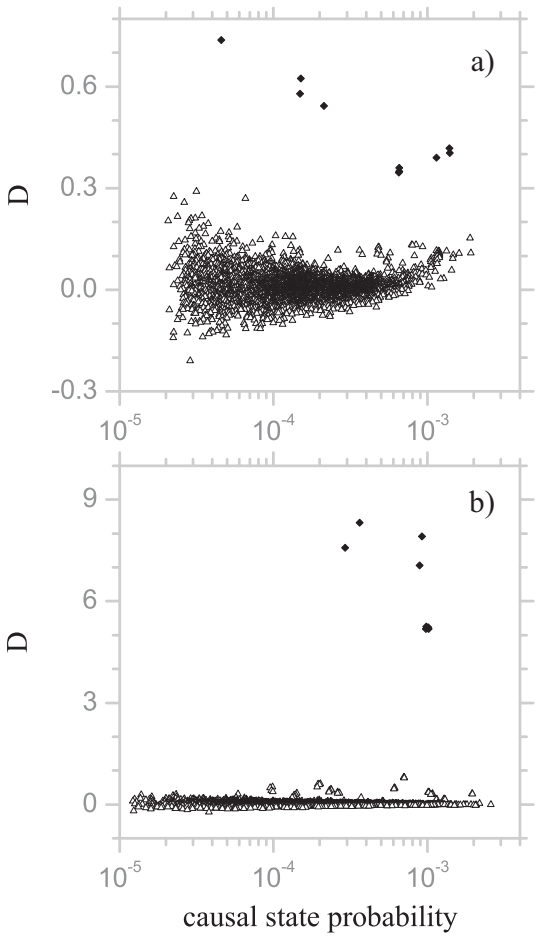


Рисунок 2.16 — Кластеризация обусловливающих состояний для временного ряда скорости атома водорода (a) и стандартного отображения (b) на «периодические» (ромбы) и «хаотические» (треугольники) классы; параметр D построен в виде графика в зависимости от частоты возникновения обуславливающих состояний.

Дополнительную иллюстрацию разделения множества обуславливающих состояний на два качественно различных класса можно получить с помощью анализа Фурье. Для каждого из обуславливающих состояний был сгенерирован бинарный временной ряд, содержащий «1» в те моменты времени, когда наблюдалось данное состояние, и «0» в другие моменты. Вычисляя спектры мощности для бинарных временных рядов, соответствующих каждому из обуславливающих состояний, получаем альтернативное указание на разницу между «периодическими» состояниями и другими состояниям. «Периодические» состояния имеют сравнительно высокий уровень спектральной плотности выше характерного периода $\approx$ 1 пс, а также около $\approx$ 0,03 пс, где соответствующая автокорреляционная функция достигает своего первого нулевого значения. «Хаотические» состояния имеют спектр мощности типа «белого шума» с приблизительно равномерной функцией спектральной плотности, Рисунок 2.17d-f. Это наблюдение позволяет предположить, что процессы с характерными временными масштабами $\approx$ 0,03 пс, что соответствует первому нулю корреляционной функции, а также $\approx$ 1 пс, соответствующие пику спектра мощности, в основном определяются «периодическими» обусловливающими состояниями.

Подводя итог, можно сказать, что в анализируемых временных рядах, таких как скорости атома водорода или Стандартного отображения присутствуют два класса «периодических» и «ха-

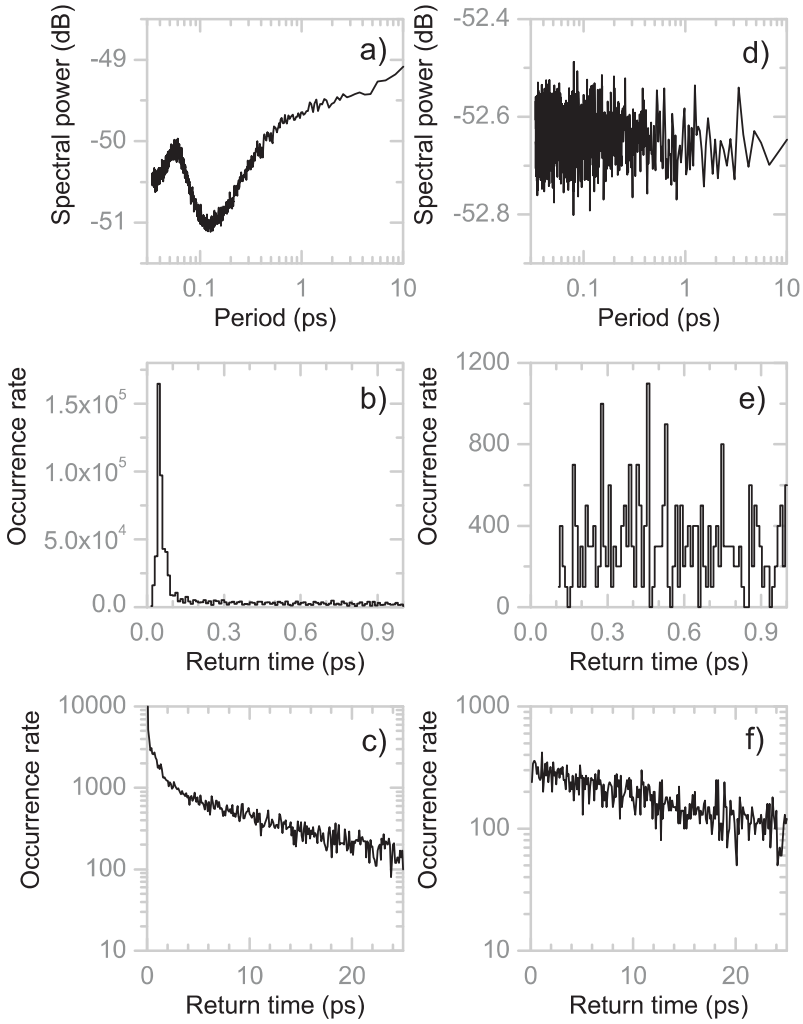


Рисунок 2.17 — Анализ Фурье временных рядов воды; спектры мощности (a, d) и гистограммы времен возврата (b, c, e, f) для типичных обуславливающих состояний, относящихся к разным типам: «периодическому» состоянию (a-c) и «хаотическому» состоянию (d-f); гистограммы на рисунках (c, f) представляют собой увеличенные и сглаженные фрагменты гистограмм, показанных на рисунках (b, e); спектры в (a, d) являются функциями обратной частоты.

отических» состояний. «Хаотические» состояния ε-машины представляет собой долгосрочные процессы смешивания, описывающие способ, которым система исследует фазовое пространство. Число «хаотических» состояний велико, что свидетельствует о преобладании областей хаотических движений (хаотическое море) над периодическими компонентами (резонансные острова), что является довольно типичной картиной, ранее наблюдавшейся только в низкоразмерных нелинейных динамических системах [139].

2.2.4 Выводы

Было проанализировано применение вычислительной механики к гамильтоновой динамике молекулярных систем. Проведен анализ концептуально важной связи обуславливающих

состояний ε-машины, построенной на символизированной траектории, с областям фазового пространства, оптимальным в смысле прогнозирования поведения траектории. Было показано, что области в фазовом пространстве, определяемые обуславливающим состояниями, обладают особыми свойствами в динамическом смысле, то есть их распределения времени возврата следуют закону Пуанкаре с двумя различными показателями степени. Это позволяет классифицировать обуславливающие состояния на квазипериодические и хаотические типы. Сравнивая наши результаты с хорошо изученным случаем Стандартного отображения, можно сделать вывод, что наш подход дает новую количественную характеристику, позволяющую разделить движение в фазовом пространстве на два отдельных класса.

Наш результат по распределению частот повторения по ансамблю обуславливающих состояний предполагает, что фазовое пространство динамической системы, соответствующей воде, имеет более сложную структуру, чем можно заключить из среднестатистического анализа времен возврата. Параметр D, введенный как индикатор отклонения от закона Пуанкаре, таким образом, обеспечивает более тонкое различие между периодическими и хаотическими фазами движения по сравнению с анализом одной гистограммы, представленным на рисунке 2.15. Некоторые обуславливающие состояния демонстрируют гораздо более медленную скорость спада, чем можно было бы ожидать из закона Пуанкаре. Этот факт свидетельствует о наличии областей в фазовом пространстве, где траектория проводит больше времени по сравнению с остальным доступным объемом. Такие области нелегко обнаружить другими методами, скорее всего, из-за обилия резонансных областей в многомерном фазовом пространстве, что затрудняет четкое различие между хаотическими и квазипериодическими движениями.

С другой точки зрения, наш метод также имеет важное значение для проблемы количественной оценки свойств переноса в высокоразмерных молекулярных системах, поскольку он выявляет (небольшое) количество областей в фазовом пространстве, играющих решающее значение для движения частиц в фазовом пространстве. Поиск таких областей на основе анализа отдельных скалярных временных рядов может быть очень полезным в численных экспериментах с большим количеством взаимодействующих частиц, которые обычно генерируют огромные объемы данных. Таким образом, извлечение необходимой информации из траектории отдельной тестовой частицы выглядит многообещающим подходом, например, при моделировании процесса сворачивания белка или динамики сложных биомолекул.

2.3 Статистическая сложность позволяет отличить динамический хаос от стохастического процесса

Статистическая сложность — это мера, введенная J. Crutchfield и K. Young в 1989 году [130]. Она оказалась полезной для описания различных сложных систем, в том числе с сотнями степеней свободы [147]. Согласно нашей работе [19] статистическая сложность многомерных траекторий, генерируемых динамикой статистического ансамбля молекул воды, возрастает до временного

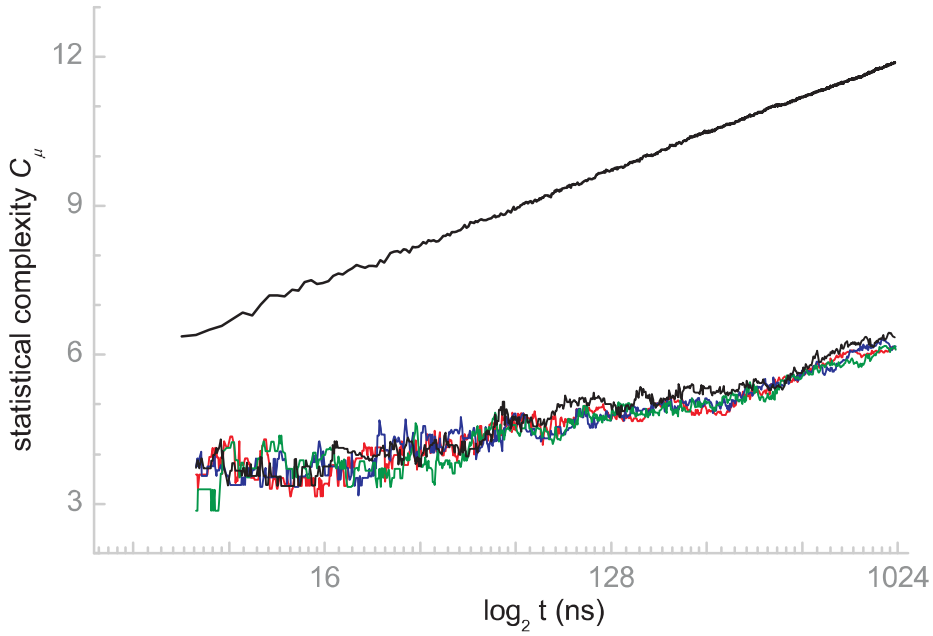


Рисунок 2.18 — Статистическая сложность в зависимости от (логарифма) длины интервала анализа для временного ряда скорости водорода (верхняя кривая) и четырех суррогатных временных рядов (нижние кривые): три независимые реализации алгоритма фазовой перестановки (красный, зеленый и синий) и один временной ряд белого шума, пропущенного через линейный фильтр нижних частот (черный) (из [19]); обратите внимание, что значение статистической сложности для данных, полученных с помощью простого генератора случайных чисел, близко к нулю и не зависит от длины временного ряда при достаточно большом значении последней (не показано).

масштаба 1 микросекунды, т.е. чрезвычайно длительного временного интервала для типичного моделирования молекулярной динамики. Более того, это свойство гораздо менее выражено для так называемых суррогатных временных рядов, которые имеют точно такой же спектр мощности и, следовательно, автокорреляционную функцию, что и исходный временной ряд.

Например, на рисунке 2.18 построен график зависимости статистической сложности от длины временного ряда для символьных данных, полученных из сечения Пуанкаре трехмерных скоростей, описывающих движение атома водорода в ансамбле из 392 молекул воды [19]. Подробную информацию о вычислении атомных траекторий, а также метод, используемый для разбиения фазового пространства и получения символьной строки из изначально данных с плавающей точкой, можно найти в [16]. На этом же рисунке показаны кривые, рассчитанные для так называемых суррогатных временных рядов с перестановкой фаз [132]. Данные имеют идентичную автокорреляционную функцию и, следовательно, спектр мощности, как и исходные траектории скорости. Можно заметить существенные различия между статистической сложностью физических и искусственно созданных данных.

В данной работе предложена гипотеза о том, что это свойство – высокое значение статистической сложности – может быть использовано для распознования детерминированных и стохастических систем (см. также [148]). Феномен роста сложности с длиной временного ряда, обеспечивающий различие между случаями детерминированного и стохастического поведения, до сих пор остается неизученным. Для того, чтобы выяснить механизм, который делает значение

сложности высоким, были проведены численные эксперименты со Стандартной отображением (известным также как отображение Чирикова-Тейлора) [146], одной из наиболее изученных парадигмальных моделей в нелинейной динамике. Было замечено, что статистическая сложность действительно была высокой в случае стандартного отображения и имела гораздо более низкое значение для суррогатного временного ряда, будучи близкой к нулю в случае некоррелированного шума от генератора случайных чисел.

Для оценки статистической сложности символического временного ряда использовался алгоритм CSSR [144] который был отмечен как эффективное, надежное и простое в использовании программное обеспечение. Алгоритм создает ε-машину, цепь Маркова с памятью на l шагов, которая представляет собой вероятностную модель для анализируемого ряда данных. Статистическая сложность измеряет информационное содержание ε-машины через ее энтропию Шеннона. Однако наш анализ показывает, что, хотя алгоритм CSSR всегда хорошо сходится и выдает конечное значение сложности, в некоторых случаях аппроксимация данных с помощью модели типа цепи Маркова оказывается недостаточной, что делает значение сложности зависимым от длины анализируемых данных. В итоге приходим к выводу, что, по крайней мере в случае Стандартного отображения, основной причиной роста сложности является свойство липкости периодических островов в хаотическом море, общее явление в Гамильтоновых системах [149]. Как было отмечено в [150], из-за свойства прилипания регулярного компонента в докритической области ($K < 0{,}9716$) динамика стандартного отображения является субдиффузионной, что может быть хорошо аппроксимировано с помощью модели случайного блуждания в непрерывном времени. Аномальные свойства временного поведения неэкстенсивной энтропии, обобщения обычной энтропии Больцмана-Гиббса, также были проанализированы в [151].

В настоящей работе в основном изучалась область $K \gg 1$, где площадь, занимаемая периодическими островами, мала, и можно ожидать, что хаотическое движение будет сильно перемешивающим и эргодическим. Тем не менее, как показывают наши результаты, наличие липкости по-прежнему является важным фактором, определяющим долгосрочные статистические показатели. С точки зрения алгоритма CSSR свойство липкости нарушает независимость точек данных, разделенных длительным интервалом времени, что делает приближение цепи Маркова необоснованным.

Наконец, обсуждается гипотеза о том, что свойство немарковости ε-машины и рост статистической сложности могут быть использованы для распознавания детерминированного и стохастического поведения. Проблема обнаружения детерминизма в хаотичном временном ряду, выглядящем как шум, существует уже давно. Подробный обзор вопросов, связанных с различием между хаосом и шумом, а также с присущими им трудностями, возникающими в случаях высокой размерности, можно найти в [152]. Можно предположить, что есть существенная разница между статистической сложностью гамильтонова хаоса и цветного шума с идентичным спектром мощности, основная причина которой заключается в наличии долговременной памяти во временных рядах, полученных из гамильтоновых систем. Это свойство возникает из-за липкости периодических островов, которые в изобилии встречаются в хаотичном море из-за множественных резонансов, возникающих в фазовом пространстве.

Также стоит отметить, что поскольку фазовое пространство гамильтоновых систем имеет сложную структуру хаотических областей, перемежающихся с периодическими островами, это приводит в некоторых случаях к необходимости различать хаос и сложное квазипериодическое движение. Мера, называемая *орбитальная сложность* была введена для этой цели в контексте анализа орбитальных движений планет [153—155]. Эта мера, хотя и основана на вычислении энтропии Шеннона (но в спектральной области), имеет совершенно иное значение, цель и сферу применимости по сравнению со статистической сложностью.

2.3.1 Изученные системы и методы анализа

Стандартное отображение определяется как

$$\begin{aligned} p_{n+1} &= p_n + K \sin \theta_n \bmod 2\pi, \\ \theta_{n+1} &= \theta_n + p_{n+1} \bmod 2\pi, \end{aligned}$$

где K — единственный параметр, определяющий динамику этой системы. Во всех приведенных ниже расчетах значение параметра K было выбрано равным $K = 6{,}908745$.

Во-первых, на этапе, называемом «символизацией», исходный действительный временной ряд преобразуется в символьную последовательность путем введения подходящего разбиения фазового пространства (Рисунок 2.3)

На следующем этапе последовательность символов преобразуется в последовательность историй, причем строки l-символов представляют собой уточнение разбиения в фазовом пространстве [19]. Реконструкция с помощью ε-машины требует группировки историй по «обуславливающим состояниям» на основе анализа предсказательных свойств каждой истории на один шаг вперед во времени. Наконец, статистическая сложность рассчитывается из ε-машины как энтропия Шеннона распределения вероятностей причинных состояний:

$$C_\mu = -\sum_{i=1}^{N_c} p_i \log p_i,$$

здесь p_i - вероятности обусловливающих состояний в ε-машине, а N_c — общее число обуславливающих состояний.

2.3.2 Численные эксперименты

Была рассчитана статистическая сложность, используя алгоритм CSSR [144] для стандартного отображения, и построены графики зависимости сложности от объема данных (длины символьной последовательности). Было также изучено, как вычисленные значения зависят от способа разбиения отображения, начальных условий и параметра K системы.

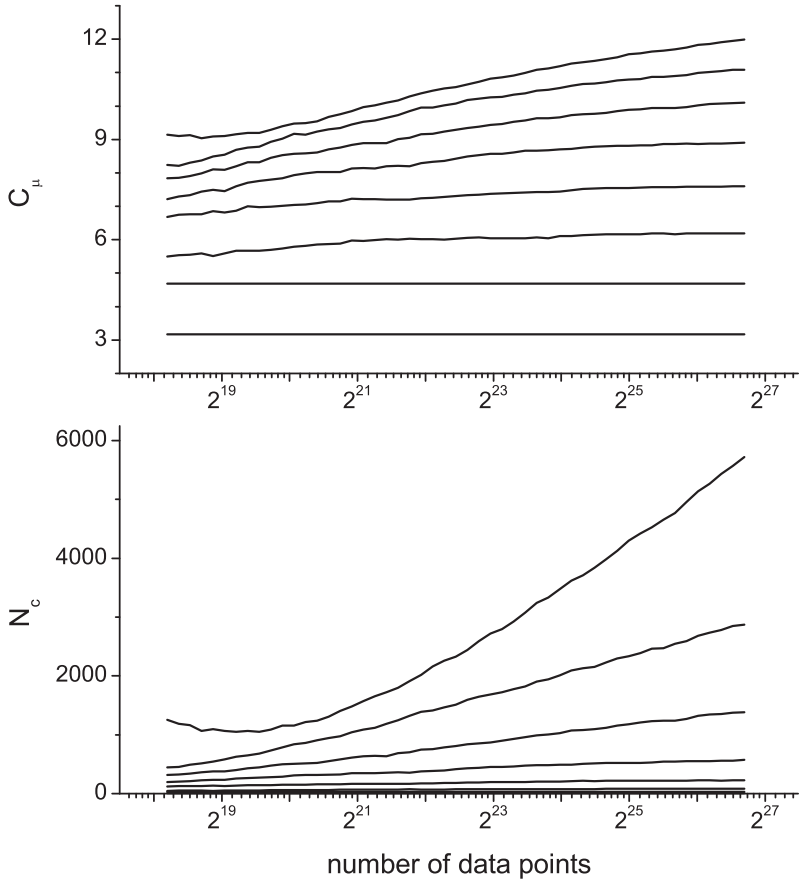


Рисунок 2.19 — Значения C_μ и количество обуславливающих состояний для различных длин истории (снизу вверх $l = 2 \ldots 9$) для стандартной траектории отображения как функции длины данных.

Типичный график статистической сложности C_μ и количество обуславливающих состояний для длины истории $l = 2 \ldots 9$ показаны на рисунке 2.19.

Результаты для суррогатных данных, полученных с использованием той же траектории Стандартного отображения показаны на рисунке 2.20. Изменение начальных условий, типа разбиения фазового пространства на этапе символизации и/или значения параметра K приводит к качественно одинаковым результатам, т.е. наблюдается существенная разница между значениями сложности, рассчитанными для данных, полученных из отображения, и для суррогатного временного ряда. Однако следует отметить, что значение сложности умеренно увеличивается с параметром K, как показано на рисунке 2.21. Такое поведение похоже на поведение других характеристик, используемых в нелинейной динамике, таких как показатели Ляпунова или мера хаотической области, описанная в [156].

2.3.3 Гипотеза о свойстве Марковости

В этом разделе показано, что большие значения сложности, наблюдаемые в случае Стандартного отображения, обусловлены наличием определенных сегментов в хаотической траектории (которые становятся историями после символизации), которые не обладают свойством,

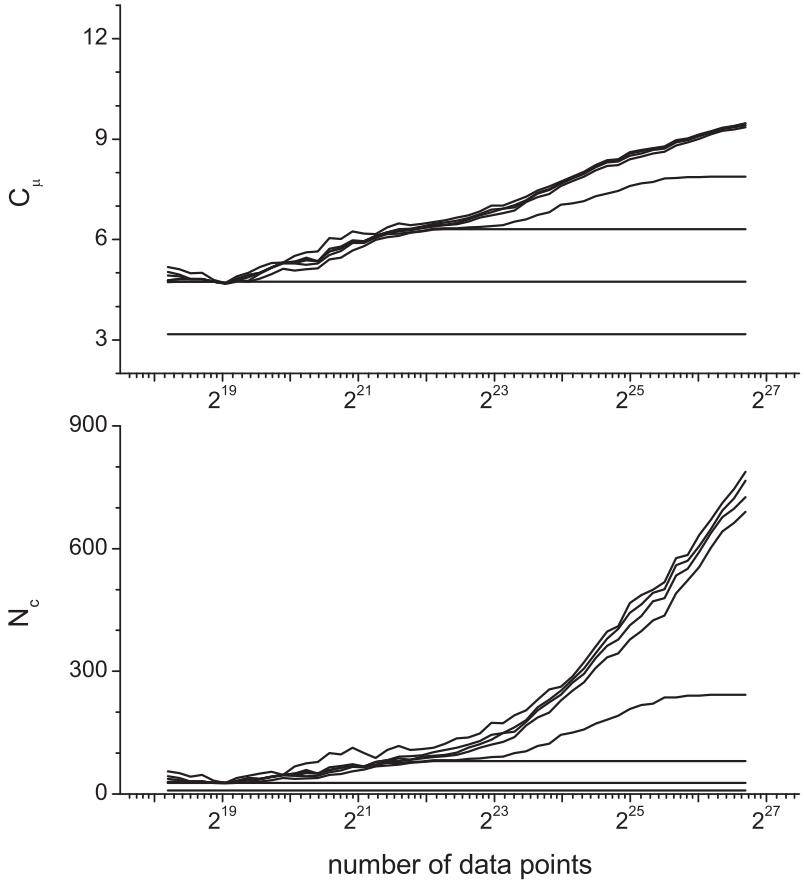


Рисунок 2.20 — То же, что на рисунке 2.19 но для случайного суррогата.

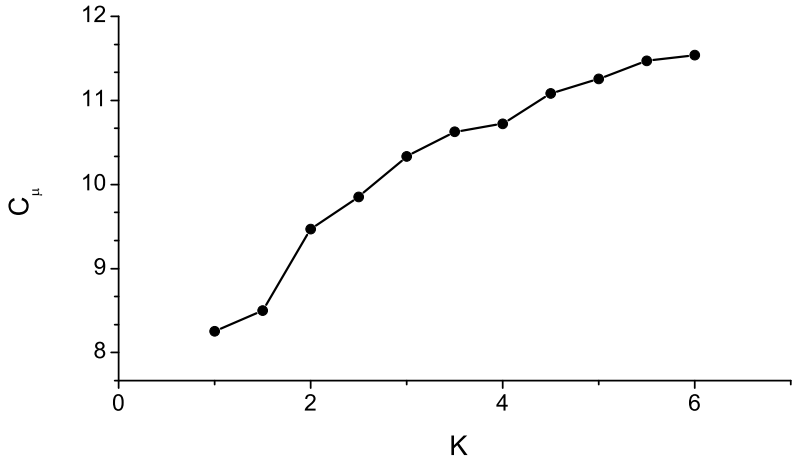


Рисунок 2.21 — Зависимость статистической сложности от параметра K (количество точек данных 10^8).

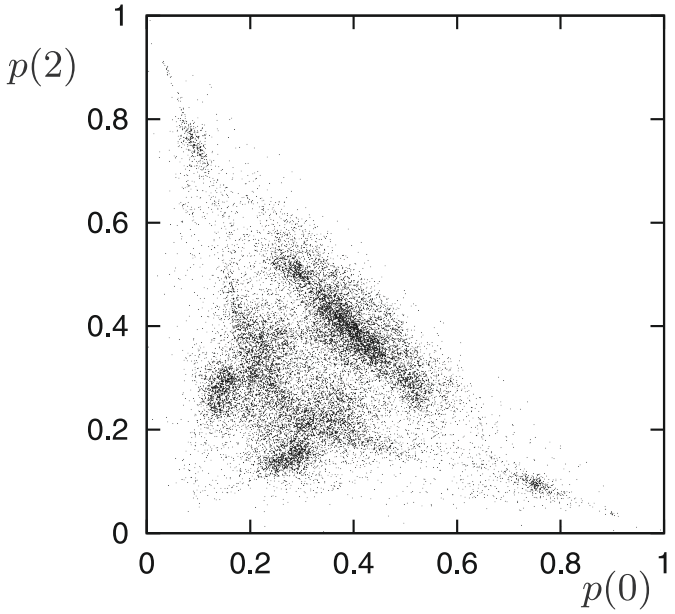


Рисунок 2.22 — Условные распределения следующего символа для всех историй при длине истории $l = 8$. Каждая точка представляет собой историю. Общее количество точек (историй) составляет 3^8.

необходимым для построения цепи Маркова из данных. Рассмотрим этап, когда символьная строка преобразована в последовательность историй, т.е. символьных слов длины l. Цепь Маркова (ε-машина) может быть построена из такой последовательности, если условное распределение вероятностей следующего символа в символьной последовательности зависит только от строки в l-символов, предшествующих символу, и не зависит от предыдущего символа, т. е. символа, который появился на $l+1$ шагов раньше. Другими словами, если рассмотреть условное распределение вероятностей для данной истории, то оно не должно измениться (в статистическом смысле), если увеличить длину истории на один символ в прошлое.

На рисунке 2.22 представлена диаграмма рассеяния, которая демонстрирует распределение условных вероятностей для каждой истории при длине истории $l = 8$. Каждая точка на диаграмме соответствует одной истории. Большой разброс вокруг точки с координатами (1/3,1/3) свидетельствует о существенном различии по сравнению со случаем равномерного распределения. Изменение с $l = 8$ на $l = 9$ не меняет общую картину распределения точек, показанную на рисунке 2.22. Однако анализ условных вероятностей для отдельных историй выявляет огромные изменения в положении точек в зависимости от дополнительного символа, добавленного в начало истории. На рисунке 2.23 показано условное распределение для следующего символа для двух историй: одной, которая показывает сильное отклонение от марковского свойства, и «нормальной», т.е. случайно выбранной истории. О большом отклонении в распределении условных вероятностей можно судить по большому расстоянию между вершинами верхнего треугольника (распределение условных вероятностей при $l = 9$) и крестом, соответствующим условной вероятности при $l = 8$. Вероятности при $l = 9$ вычисляются путем добавления одного из трех символов (012) в начале истории $l = 8$.

Наконец, участки хаотических траекторий, соответствующие истории с большим отклонением от марковского свойства, расположены в областях фазового пространства близких к

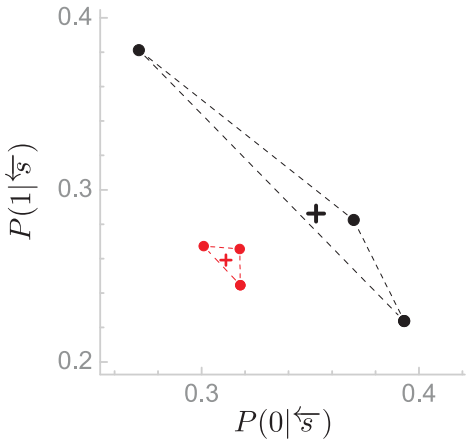


Рисунок 2.23 — Условное распределение вероятностей для истории с наибольшим отклонением от марковского свойства (черный треугольник); добавление символа в историю кардинально меняет положение точки на диаграмме; три круга соответствуют добавлению ‘0’, ‘1’ или ‘2’ к истории $l = 8$ (крестики); красный треугольник представляет собой случайно выбранную историю с явным марковским свойством, т.е. распределение вероятностей не зависит от добавленного символа.

периодическим островам. График на рисунке 2.24 только точек, соответствующих историям с большим отклонением от свойства цепи Маркова. Сравнение с рисунком 2.3 показывает, что история с большим отклонением в распределениях расположена близко к периодическим островам. Поэтому можно предположить, что нарушение марковскости можно интерпретировать как проявление известного феномена «липкости» [149] траекторий в областях, близких к периодическим островам. Длительное блуждание по траектории вокруг острова эквивалентно существованию долговременной памяти в соответствующих отрезках хаотического временного ряда. Рисунок 2.24 следует также сравнить с рисунком 2.25, на котором представлены отрезки хаотической траектории, соответствующие истории, обладающей марковским свойством. По-видимому, она не имеет никакого отношения к периодическим островам. Такие истории составляют подавляющее большинство в ансамбле 3^{10} историй, а немарковские истории составляют лишь долю процента.

2.3.4 Обсуждение

В данном разделе показано, что статистическая сложность, по-видимому, является полезной мерой для распозанвания гамильтонова хаоса в системах низкой и высокой размерности от коррелированного шума с идентичной функцией автокорреляции. Его значение для символьного временного ряда, рассчитанного на основе динамики гамильтоновых систем, существенно больше, чем для временного ряда белого шума или временного ряда, полученного с помощью суррогатного алгоритма перетасовки фаз. Наше объяснение происхождения этого явления в терминах теории цепей Маркова состоит в разложении марковского свойства с помощью символьных последовательностей, полученных из Гамильтоновых систем.

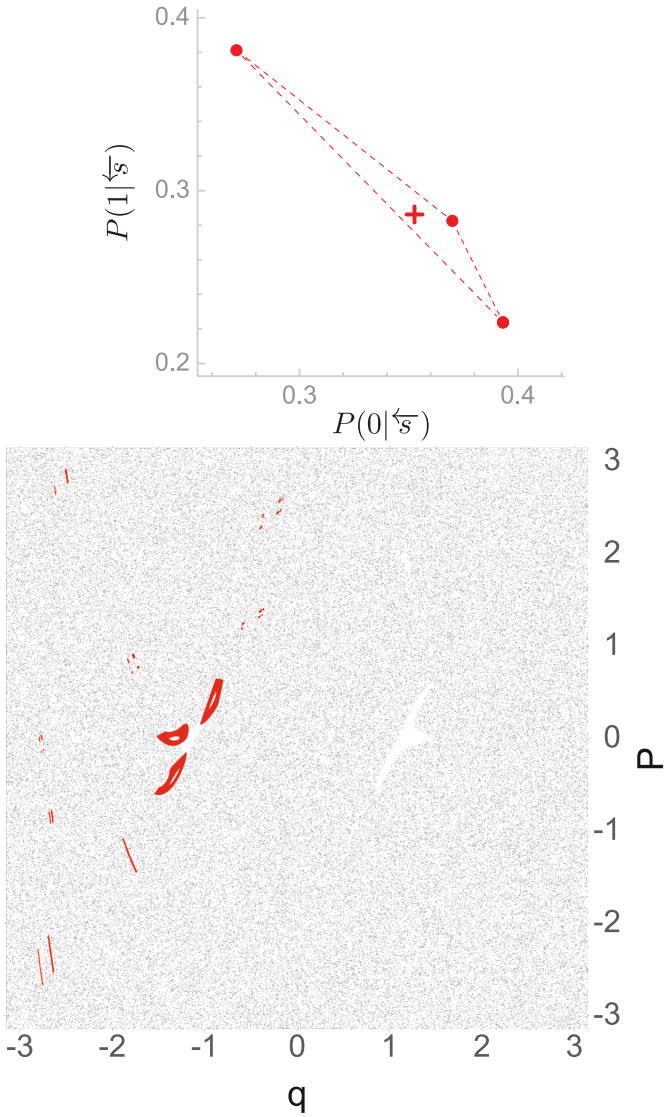


Рисунок 2.24 — Участки хаотической траектории, соответствующие истории с большим отклонением от свойства цепи Маркова; отображаются только точки, соответствующие трем центральным символам в истории; по-видимому, история включает в себя один из двух периодических островов, показанных на рисунке 2.3.

Полагаем, что большое значение сложности, наблюдаемое в наших численных экспериментах, определяется наличием периодических островов с липкими границами в фазовом пространстве Гамильтоновых систем. Липкость определенных областей в фазовом пространстве приводит к долговременным эффектам памяти, которые ответственны за нарушение статистической независимости будущих состояний от прошлых. Это, в свою очередь, делает неустойчивой процедуру группировки историй в обусловливающие состояния, составляющую ядро алгоритма CSSR. В результате алгоритм находит все больше и больше обусловливающих состояний, необходимых для построения ε-машины как цепи Маркова, а значение сложности растет с числом обуславливающих состояний.

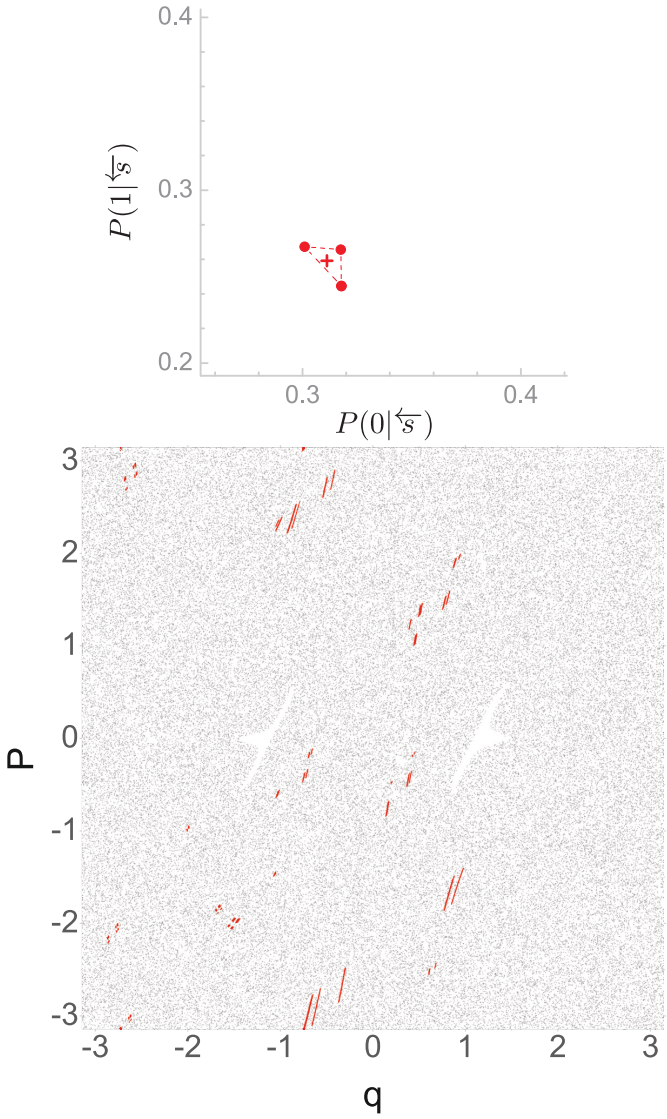


Рисунок 2.25 — Участки хаотической траектории, соответствующие истории без отклонений от свойства цепи Маркова; отображаются только точки, соответствующие трем центральным символам в истории; по-видимому, история не включает ни один из двух периодических островов, показанных на рисунке 2.3.

Глава 3. Структура и динамика вирусов в моделях с атомистическим разрешением

3.1 Обзор: существующие работы по полноатомному моделированию целых вирусов

Современные экспериментальные возможности для измерения атомистической структуры очень больших молекулярных систем (недавний пример представляет собой измеренную с помощью крио-ЭМ структуру вирусного капсида диаметром 125 нм [157]) в значительной степени полагаются на сложные методы моделирования, которые сопоставляют известные химические структуры с трехмерными картами плотности, полученными из экспериментальных данных. В настоящее время эти методы моделирования в основном касаются *структурной* подгонки, которая особенно сложна для тех частей плотности, где возможна неоднозначность из-за более низкого разрешения или отсутствия данных. Однако атомистическая *динамика* очень важна как для лучшего качества подгонки (например, путем указания гибких областей молекул, где экспериментальные данные по определению имеют низкое качество), так и для последующего использования полученных атомистических моделей (например, предоставления короткоживущих конформаций, невидимых в эксперименте). «Моделирование атомистической динамики» означает классическую молекулярную динамику (MD). Таким образом, MD моделирование очень больших (по современным стандартам MD) систем полезно и важно, несмотря на технические трудности, вызванные огромным количеством атомов в системе.

Среди различных крупных молекулярных систем с известной атомистической структурой вирусы особенно удобны для моделирования MD. Это связано с тем, что в растворе они существуют как изолированные молекулярные образования, взаимодействующие только с водой и ионами. Сложные и малоизученные взаимодействия с другими вирусами, клеточными мембранами, другими клеточными органеллами и т. д. можно исключить из моделирования, по крайней мере, на некоторых этапах жизненного цикла вируса. Также размер вирусов варьируется в большом интервале: от диаметра в ~17 нм, что поддается обычному моделированию MD, до сотен нанометров, что в настоящее время недостижимо для MD. Несмотря на разнообразие структур и функций, можно составить набор вирусов с общим сходным строением. Например, все они могут состоять только из капсида, образованного повторяющейся копией одного (или небольшой группы) белка среднего размера и генома относительно короткой цепи ДНК или РНК, упакованного внутри этого капсида. Набор таких вирусов разных размеров может служить удобными моделями для настройки и проверки моделирования MD.

Этот раздел посвящен полноатомному моделированию MD целых вирусов или их полных капсидов. Важно отметить, что рассматриваются только работы по атомистическому моделированию, исключая (гораздо более многочисленные) публикации по моделям вирусов с разрешением ниже атомистического. Акцент делается на физико-химической стороне моделирования, а не на

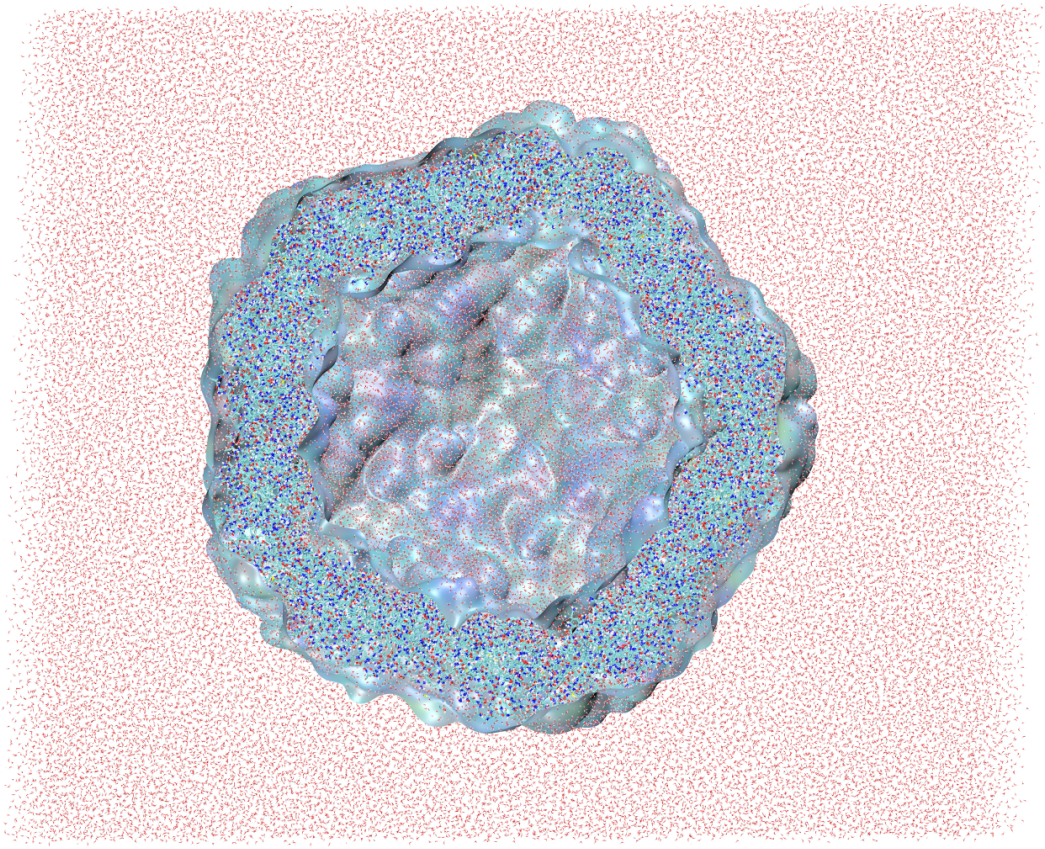


Рисунок 3.1 — Разрез полноатомной модели капсида PCV2 без генома, погруженного в ячейку, заполненную атомистической водой [7].

биологических аспектах изучаемых вирусов. С этой точки зрения взаимодействие с водой и ионами особенно важно, и ему обычно уделяется значительно меньше внимания (если вообще уделяется) в биологической литературе. Поэтому были выбраны в основном симуляции с атомистическим растворителем (Рисунок 3.1), то есть молекулы и ионы воды включены в модель в атомистическом представлении [1], а не как бесструктурный континуум или любая другая внешняя коррекция.

Лишь несколько лет назад начали появляться публикации по полноатомному моделированию вирусов с атомистическим растворителем. Первая работа была опубликована в 2006 году группой Клауса Шультена [101] где был промоделирован вирус сателлит табачной мозаики (STMV). Этот вирус состоит из РНК-генома, заключенного в икосаэдрический капсид диаметром приблизительно 17 нм. STMV — это «безоболочечный» вирус, то есть оболочка вируса состоит только из одного слоя, состоящего из белков, в отличие от «оболочечных» вирусов, у которых дополнительно имеется еще один внешний слой из липидов и белков. Капсид был построен из 60 копий одного белка, первоначальная структура которого была взята из кристаллографических измерений. Была промоделирована пустая капсидная часть, а также целый вирус с искусственной РНК, построенной с предположением, что ее структура состоит из 30 двухцепочечных цепей (основа которых была разрешена из кристаллографических данных) и произвольно упакованных одноцепочечных областей, соединяющих разрешенные двухцепочечные [158], последователь-

ность РНК была искусственной, а ее длина (949 нуклеотидов) была аналогична нативному геному (1058). Это ядро РНК также было промоделировано отдельно, без капсида.

Результаты моделирования показали, что вода может проходить через стенку капсида. Авторы также рассчитали константу диффузии ионов хлора в системе. Более интересные результаты для биологических приложений касались стабильности системы. Капсид, содержащий РНК, и РНК по отдельности оказались стабильными в течение всего времени моделирования (13 нс и 10 нс). Однако пустой капсид оказался нестабильным, поскольку белковые блоки сместились к центру капсида, что привело к его общей деформации. Факт нестабильности пустого капсида согласуется с экспериментом. Эти результаты имеют значение для механизма сборки вируса, подчеркивая роль РНК в этом процессе.

Для стабильного целого вириона было обнаружено интересное наблюдение в динамике белковых единиц. Белковые единицы вокруг экваториального «пояса» демонстрировали коррелированное движение, в то время как единицы на противоположных концах демонстрировали антикоррелированное движение. Это один из результатов, который удалось получить только в полноатомной MD целого вируса.

Результаты публикации позволили сделать вывод, что правильная сборка вируса происходит в присутствии РНК, при этом пустые капсиды не образуются. Наши результаты также показали подобное поведение деформации капсида для другого вируса PCV2 [7], когда геном отсутствует внутри капсида. Такая нестабильность вызвана неправильным распределением ионов на поверхностях капсида (внутри и снаружи). Однако в отличие от STMV экспериментально известно, что пустые капсиды PCV2 существуют [100].

Спустя три года после первой публикации была промоделирована система из 4,5 миллионов атомов [159], с акцентом на механические свойства капсида вируса мозаики южной фасоли. В данной работе моделирование Force Probe MD было применено для получения распределения механических свойств (упругой константы и сил текучести) вирусного капсида на атомистическом уровне. Подход позволил проанализировать свойства вирусной оболочки при изменении скорости сферического наконечника зонда и ее размера. Авторы наблюдали три режима деформации под нагрузкой. Они также не обнаружили потока воды через стенку капсида даже при вдавливании кончика сферы, что привело бы к существенному изменению внутреннего объема.

Обычно предполагается, что удаление ионов кальция из структуры капсида вируса мозаики южной фасоли приводит к его набуханию, что является начальной стадией инфицирования клетки, во время которой происходит высвобождение генома [160]. В своей последующей публикации авторы [8] продолжили исследование механических свойств капсида вируса с помощью моделирования Force Probe MD при атомистическом разрешении с целью изучения того, как изменяются механические свойства капсида при удалении кальция. Они также исследовали распределение упругой постоянной и сил текучести на внешней поверхности, а также на внутренней поверхности капсида, что невозможно сделать экспериментально для внутренней поверхности. В работе отмечено, что после удаления кальция механические свойства изменяются раньше, чем изменяется структура. Обнаруженное ослабление вдоль осей пятого порядка капсида породило гипотезу о том, что пентамеры, образующие структуру капсида, служат возможными портами для высвобождения РНК.

Для некоторых вирусов наблюдалось увеличение их диаметра или «набухание» при удалении ионов металлов (например, Ca^{2+}) из структуры капсида [161; 162]. Это свойство играет важную роль в процессе заражения. Моделирование MD позволяет изучать динамику и структурные особенности капсида, связанные с набуханием, как это было сделано в [102] для другого вируса-сателлита, сателлита вируса некроза табака (STNV). Моделирование проводилось для капсидов, содержащих структурные ионы Ca^{2+}, и капсидов после удаления ионов в молекулярных водных растворах в течение 1 микросекунды каждый, РНК в моделирование не включалась.

Авторы проанализировали как изменяется водопроницаемость капсида после удаления ионов кальция и обнаружили, что участки связывания кальция капсида играют разные роли в структуре STNV: некоторые отвечают за набухание после удаления ионов кальция, в то время как другие играют структурную роль и не принимают участия в набухании.

В [163] авторы провели MD моделирование нативного и набухшего состояний вируса хлоротической крапчатости коровьего гороха (CCMV) в вакууме и показали, что изучение структурных переходов в набухшем капсиде требует длительного времени моделирования. Они применили свой подход многомасштабной молекулярной динамики/экстраполяции параметров порядка (MD/OPX) [164] для длительного (200 нс) моделирования набухшего капсида вируса CCMV в вакууме, что позволило исследовать механизмы его структурных переходов.

В своей следующей работе [165] авторы продемонстрировали применение своей методологии MD/OPX для исследования набухания капсида вируса CCMV в среде, имитирующей раствор. Результаты моделирования показали, что набухание капсида, а также заметное при моделировании в вакууме сжатие сопровождались потерей симметрии сначала локально, а затем и во всем капсиде.

Конформацию неструктурированных, динамических фрагментов капсида невозможно измерить экспериментально, но их можно реконструировать, а их структуру и динамику можно изучить в моделировании, как это было сделано в [166] на примере полиовируса. Капсид полиовируса включает четыре белка VP1, VP2, VP3, VP4, которые образуют единую структурную единицу (блок), 60 копий этой единицы составляют весь капсид. Рентгеноструктурный анализ не позволил определить структуру первых 20 аминокислот N-концевого домена VP1, который участвует в процессе инфекцирования. Моделирование проводилось для четырех систем: целый вирус с РНК, пустой капсид, пентамер и изолированный N-концевой домен. Размеры моделируемых систем достигали 4 миллионов атомов. Моделирование проводилось в течение 10 нс для каждой системы в растворе NaCl концентрацией 0,154 mM. Также в системе с РНК для стабилизации структуры были добавлены ионы Mg^{2+}. Было замечено, что в отдельном домене структура α-спирали сохраняется на протяжении всех 10 нс моделирования, этот домен более гибок в пустом капсиде, чем в полном вирусе. Было исследовано поведение ионов в полном вирионе и обнаружено, что ионы Na^{+} и Cl^{-} диффундируют через стенку капсида через небольшую трещину в области оси второго порядка. Na^{+} диффундирует внутрь, Cl^{-} диффундирует наружу, в то время как Mg^{2+} остается связанным с РНК.

Свойства этого вируса также исследовались в 2014 году на более длительных интервалах времени, до 200 нс [103]. В своей работе авторы использовали натрий-фосфатный буфер с pH 7.4 для воспроизведения экспериментальных условий. Общий размер системы составил 6.5 мил-

лионов атомов. После длительного уравновешивания в течение 100 нс наблюдалась диффузия молекул воды через стенку капсида. В состоянии равновесия время полной замены воды внутри капсида внешней водой составило 25 μс. Этот результат объясняет экспериментальные данные о способности полиовируса выдерживать высокое давление и удары давления [167] поскольку вода диффундирует достаточно быстро, чтобы выровнять давление, не разрушая капсид. Также были получены результаты о возможности испарения воды изнутри капсида. Эти результаты важны для разработки методов дезактивации вирусов. Несмотря на диффузию воды, ионы не прошли через капсид за все 200 нс времени моделирования. Это классифицирует капсид как полупроницаемую мембрану. Внутри пустого капсида наблюдалось отрицательное давление, что объяснялось кулоновским взаимодействием раствора внутри капсида с избыточными зарядами капсида.

Аналогичные результаты по функционированию капсида как полупроницаемой мембраны были получены в [22] для цирковируса свиней типа 2 (PCV2). В работе проведено полноатомное моделирование пустого капсида PCV2 в среде, близкой к физиологической, с молекулярной моделью воды. Капсид моделировался двумя методами: классическим MD и гибридным подходом MD/гидродинамики. Размер системы составил 1,9 миллиона атомов, а время моделирования — 10 нс. Было обнаружено, что капсид стабилен через 10 нс. Анализ потоков воды и ионов показал, что вода может пересекать стенку капсида в обоих направлениях, а ионы — нет.

В [168] авторы промоделировали пустой капсид вируса ВИЧ-1 и исследовали его физико-химические свойства. Моделируемое время составило 1.2 μс, а размер системы достигал 64 миллионов атомов. Моделирование проводилось в физиологических условиях с использованием 150mM водного раствора NaCl. Авторы исследовали устойчивость капсида как в целом, так и локально, рассчитали скорость обмена воды между внутренней и внешней частью капсида, определили расположение ионов на поверхности капсида. Были проанализированы электростатические свойства, а также акустические свойства и акустические моды.

Было обнаружено, что капсид способен перемещать ионы с высокой специфичностью, что может играть определенную роль во время инфекцирования, помогая фильтровать определенные молекулы. Ионы натрия переносились через стенку капсида по специальным каналам. Установлено, что ионы хлора образуют слой на внутренней поверхности капсида, а ионы натрия — на внешней поверхности. Аналогичные слои были обнаружены на поверхности PCV2 [7].

Большая часть анализа была посвящена акустическим модам на поверхности капсида. Было обнаружено колебательное движение, распространяющееся через капсид, аналогично капиллярным волнам, характерным для мембран и жидкостей. Интересно, что частоты этих волн оказались в ультразвуковом диапазоне (от 2.38 до 11.9 МГц). Более того, было обнаружено, что волны разных частот локализуются в разных областях капсида (например, самая высокая частота связана с кончиком капсида и т. д.). Была выдвинута гипотеза, что такое поведение играет важную роль в процессе заражения, поскольку способствует передаче информации на очень большие расстояния.

Моделирование MD на основе атомов было проведено для капсида вируса гепатита B [169]. Моделирование проводилось в течение 1.1 μс в молекулярной воде и 150 mM NaCl, таким образом, размер системы составил 6 миллионов атомов. Результаты демонстрируют высокую гибкость капсида и наличие асимметричного искажения, которое может привести к проблемам в структурных измерениях с помощью крио-ЭМ. Капсид HBV также проницаем для молекул воды, ионы

натрия могут перемещаться через стенку капсида быстрее, чем ионы хлорида. Авторы предположили, что эта медленная транслокация ионов хлора может играть важную роль в постепенности структурного искажения капсида. Ионы физиологического раствора могут играть важную роль в посредничестве специфических взаимодействий.

В настоящее время полноатомное моделирование вирусов с использованием MD находится в начале активного развития, о чем свидетельствует небольшое количество публикаций в этой области. Эти работы демонстрируют успешное применение MD к целым вирусам, анализ их стабильности в различных условиях, взаимодействия с окружающей средой и, в некоторых редких случаях, с геномом. Разработанные модели позволяют исследовать макроскопические свойства вирусов на некоторых стадиях их жизненного цикла, исходя из фундаментальной микроскопической динамики. Факт существования стабильных пустых капсидов важен, поскольку он позволяет предположить, что такие капсиды могут существовать в природе при соответствующих условиях и, возможно, играть роль в самосборке вируса. Однако, поскольку моделирование столь крупных систем является новой областью исследований, важно уделять пристальное внимание правильной подготовке системы.

Одним из самых интригующих вопросов в этой области как для экспериментаторов, так и для разработчиков моделей является структура генома внутри капсида и его взаимодействие с капсидом. На сегодняшний день попыток реконструировать геном крайне мало, поскольку экспериментальные данные о его структуре в лучшем случае фрагментарны. Однако быстрый прогресс в определении структуры дает основания полагать, что полноатомное моделирование генома вирусов станет одним из наиболее активно развивающихся направлений исследований в ближайшие годы.

Вопросы вирусной самосборки [170; 171] остаются в значительной степени открытыми. Они связаны с анализом специфических взаимодействий между единицами капсида, а также между геномом и капсидом. Подобные взаимодействия имеют ключевое значение для процесса самосборки и упаковки генома. Моделирование MD с включением всех атомов станет инструментальной методологией в изучении деталей этих взаимодействий, таким образом, косвенно, внеся значительный вклад в реконструкцию биологических механизмов самосборки вирусов. Исследовательская деятельность в этом направлении, вероятно, станет тенденцией в области полноатомной MD вирусов.

В целом, полноатомное моделирование целых вирусов в явной воде доказало свою полезность, но в настоящее время оно находится в начале своего активного использования. Тенденции в экспериментальных методах, компьютерных разработках и вопросах, которые ставит современная биология, таковы, что эти моделирования, вероятно, получат активное развитие в ближайшем будущем.

3.2 Изучение капсида бактериофага MS2 с использованием полноатомной молекулярной динамики: распределение зарядов на капсиде и динамика ионов из окружающего раствора

Современные компьютеры могут моделировать очень большие биомолекулярные системы размером в миллионы атомов (органеллы, клеточные мембраны, целые вирусы и т. д.) с атомистическим разрешением [78; 103; 168]. Необходимая начальная молекулярная структура самой биомолекулы, трехмерные координаты всех атомов в системе, получают с помощью экспериментальных методов высокого разрешения, таких как рентгеновская кристаллография и криоэлектронная микроскопия (крио-ЭМ). Компьютерные модели молекулярной динамики (MD) на основе атомов могут предоставить информацию и уникальные знания о сложных биомолекулярных системах, которые невозможно получить никакими другими способами. Например, реконструкция частей молекулярной системы, таких как гибкие части, которые невозможно измерить экспериментально, обеспечивает понимание динамики при физиологической температуре. Важно отметить, что до недавнего времени рентгеновская кристаллография была основным источником молекулярных структур, а в настоящее время крио-ЭМ измерения предоставляют информацию сопоставимого разрешения, но с решающим преимуществом, заключающимся в возможности измерения асимметричных структур.

Моделирование вирусов является наиболее реалистичным, поскольку, в отличие от других крупных биомолекулярных систем, вирусы являются автономными биологическими единицами, которые существуют изолированно от остального организма (хотя они не могут размножаться изолированно) [23—26]. Например, компьютерные модели клеточных органелл с необходимостью рассматривают только часть системы, которая взаимодействует с остальными частями клетки посредством сложных, плохо изученных механизмов, что затрудняет сравнение с экспериментальными результатами. Для вирусов взаимодействие с водным раствором является единственной внешней силой, определяющей структуру и динамику, а моделирование воды хорошо развито в MD. Несмотря на значительные вычислительные ресурсы, необходимые для такого моделирования, число вирусов, моделируемых с разрешением на уровне атомов, растет (см. обзоры всех известных на сегодняшний день симуляций [172; 27]).

Недавно была создана MD модель капсида вируса PCV2, который является самым маленьким известным несателлитным вирусом (ему не нужен другой вирус для размножения) [7; 22; 25]. В данном разделе описано MD моделирование известного бактериофага, который инфицирует *E. coli*, (+)ssRNA фаг MS2. Этот вирус – идеальный кандидат для такого полностью атомистического моделирования, поскольку он небольшой (его геном короткий, 3569 нуклеотидов), его икосаэдрический капсид лишь незначительно больше, чем у PCV2 (27 нм против 23 нм), но он гораздо более подробно изучен с использованием различных методов молекулярной биологии и биофизики [173—175; 28—30].

Молекулярная структура капсида MS2 была недавно измерена с атомистическим разрешением, что позволило получить не только структуру симметричных частей капсида, но и несколько

белков, которые отличаются от остального капсида, что делает его асимметричным [176]. В последней публикации даже представлены части генома, разрешенные на атомистическом уровне, и большая часть генома с разрешением, позволяющим получить его остов [177]. Таким образом, мы находимся в уникальном положении, поскольку имеем возможность построить очень реалистичную MD-модель капсида вируса, включая его асимметричные части, которые, как известно, играют важную биологическую роль в инфицировании бактерий.

Капсид MS2 в основном состоит из множественных копий одного белка длиной 129 аминокислотных остатков. Белок принимает три различные конформации, обозначаемые как A, B и C, которые образуют два типа димеров: асимметричный AB и симметричный CC. Капсид состоит из 59 AB и 30 CC димеров. Наконец, в стенку капсида встроен белок созревания из 393 остатков [178]. Капсид образует полупроницаемую стенку с порами, через которые проходит вода и ионы. Эта информация была использована для разработки MD модели капсида и анализа ряда его структурных и динамических свойств.

3.2.1 Детали моделирования

Сборка модели капсида

Наша модель капсида основана на крио-ЭМ структуре 5TC1 из Protein Data Bank [179]. Она содержит i) три мономера белка оболочки, образующие блок треугольной формы; ii) белок созревания, последовательность которого отличается от последовательности остальных белков капсида; и iii) 5 мономеров, расположенных рядом с белком созревания, структура которых деформирована белком созревания. Среди этих мономеров в одном из них имеется пробел в информации о структуре (обозначен как цепь G в файле PDB), то есть координаты остатков 68–77 отсутствуют, вероятно, из-за их гибкой или неупорядоченной структуры. Кроме того, в белке созревания также имеется 4 пробела (всего 63 остатка).

Подробная информация о процедуре сборки представлена в приложении Д. Для всех задач использовался пакет моделирования GROMACS 5.1 [180]. Окончательная ячейка моделирования изображена на рисунке 3.2.

Протокол моделирования

Моделирование проводилось с использованием силового поля AMBER03 и модели воды TIP3P. Общее число атомов в системе составило 3528811, включая 1057725 молекул воды.

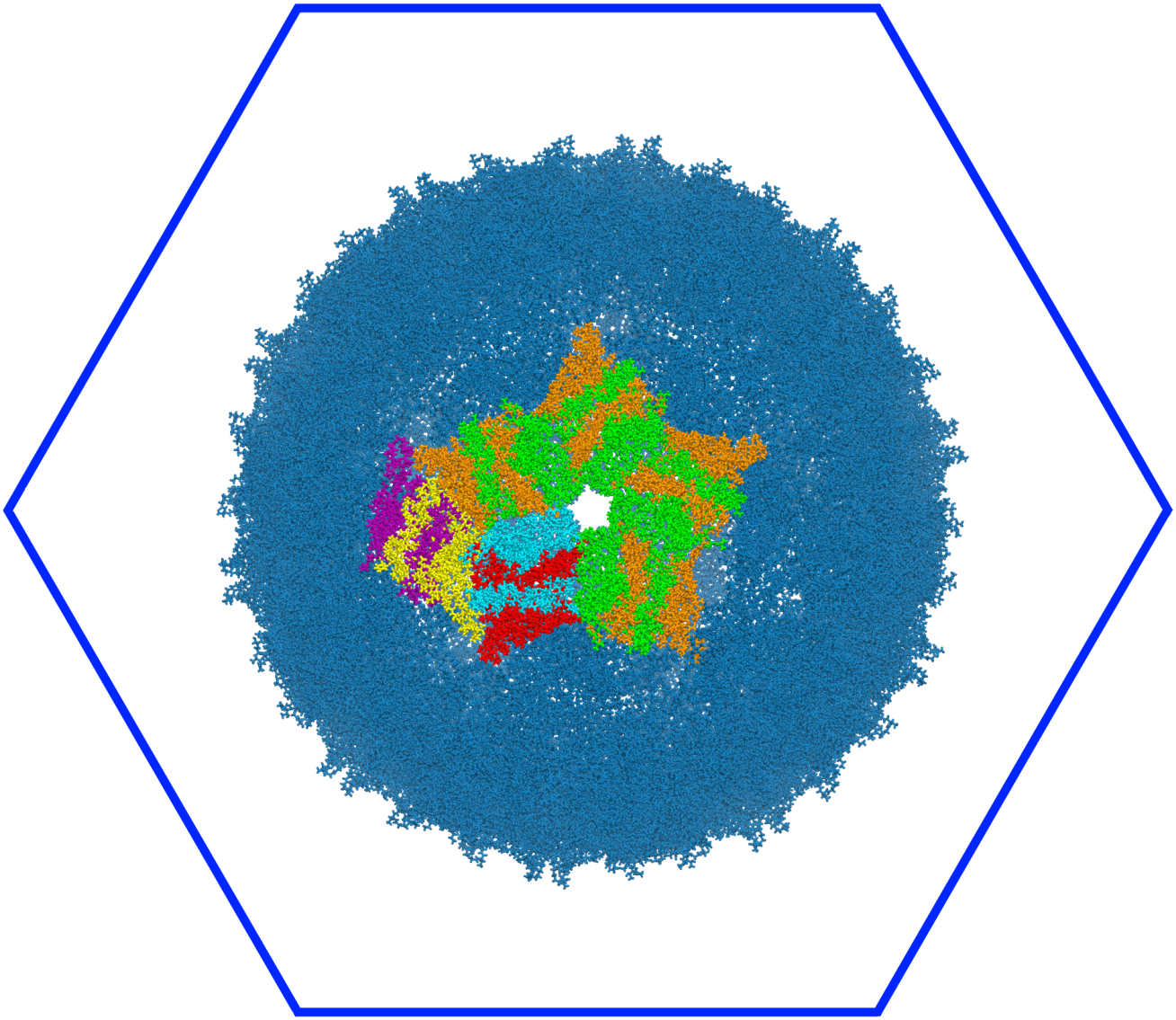


Рисунок 3.2 — Модель капсида в коробке в форме ромбического додекаэдра; несколько белков оболочки окрашены, выделяя пентамер, окружающий пору (зеленые, оранжевые, синие и красные молекулы), и блок тримера (зеленые, оранжевые, синие, красные, фиолетовые и желтые молекулы); ионы и молекулы воды не показаны.

Сначала была выполнена минимизация энергии собранной системы с использованием алгоритма наискорейшего спуска до тех пор, пока наибольшая остаточная сила не стала меньше 5000 кДж нм$^{-1}$. Затем был проведен цикл уравновешивания при постоянном объеме длительносью в 5 нс. Тяжелые атомы капсида были зафиксированны в своих исходных положениях, за исключением реконструированной недостающей последовательности, которой было разрешено принять оптимальную структуру. Начальная температура была установлена на 100 K и в течение первых 200 пс она постепенно повышалась до 298 K; это значение сохранялось до конца эксперимента (использовались параметры имитации отжига GROMACS). Среднее давление в ячейке составило 120 бар, что близко к стандартному давлению. На самом деле, согласно сжимаемости воды при 300 K ($4.5 \cdot 10^{-5}$ bar^{-1} [180]), это дополнительное давление соответствует изменению длины ячей-

ки на 0,07 нм из 37 нм, что незначительно. Таким образом, было решено не использовать баростат для основных прогонов, чтобы избежать внесения нежелательных колебаний объема в систему.

После уравновешивания был проведен продуктивный прогон продолжительностью 50 нс без ограничений. Объем ячейки поддерживался постоянным. Среднее давление во время прогона составило -85 бар.

Во всех прогонах накладывались трехмерные периодические граничные условия, температура контролировалась термостатом с изменением масштаба скорости с постоянной времени 1 пс, временной шаг составлял 2 фс, все связи ограничивались алгоритмом LINCS, электростатические взаимодействия вычислялись методом PME (отсечка при 1 нм), в то время как для взаимодействий Ван-дер-Ваальса отсечка составляла 1 нм.

3.2.2 Результаты и обсуждение

Структура

Основным показателем устойчивости структуры является изменение во времени ее среднеквадратического отклонения (СКО) от исходной конфигурации. Графики для всего капсида и только его остова показаны на рисунке 3.3. Оба значения RMSD увеличились на 20 нс, а затем достигли плато при небольшом значении (~0.27 нм), что свидетельствует об отсутствии существенных смещений. То же самое показывает и изменение радиуса инерции капсида во времени (Рисунок 3.3): он перестает расти после 20 нс, показывая увеличение на ~0.15 нм в конце моделирования по сравнению со структурой крио-ЭМ.

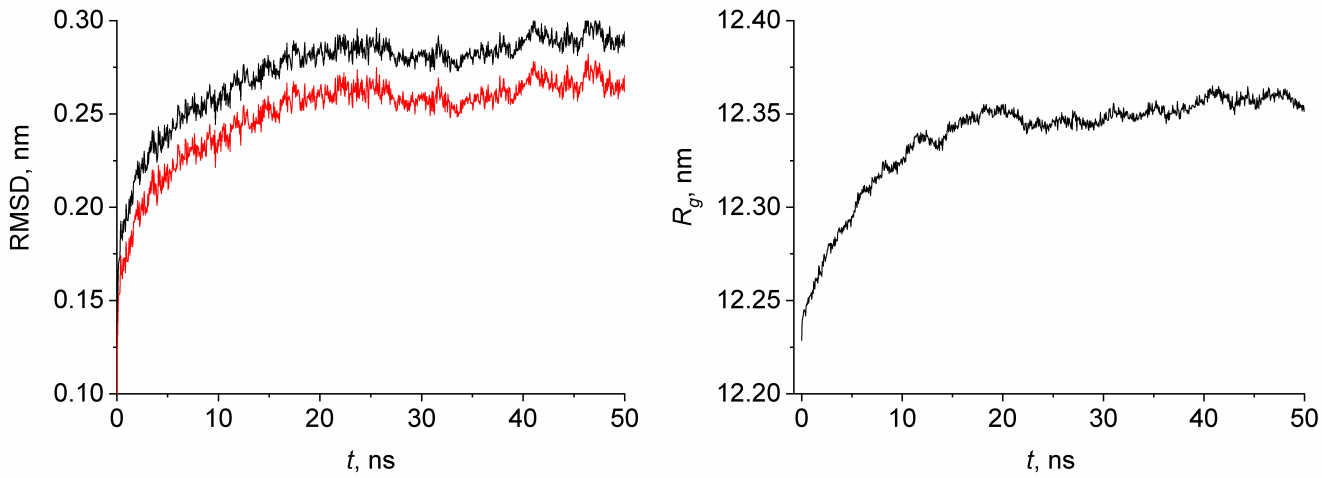


Рисунок 3.3 — Слева: RMSD всего капсида (черный) и его остова (красный) от начальной структуры; справа: радиус вращения капсида.

Интересно рассмотреть среднее смещение отдельных аминокислотных остатков из их исходных положений, Рисунок 3.4. Все остатки имеют схожее низкое среднее смещение (в терминах

RMSD), за исключением остатков 13-15 со значениями ~0.8 нм. Эти остатки располагаются на внешней поверхности вдали от пор и подвергаются воздействию внешнего раствора, что делает их относительно более подвижными (Рисунок 3.5).

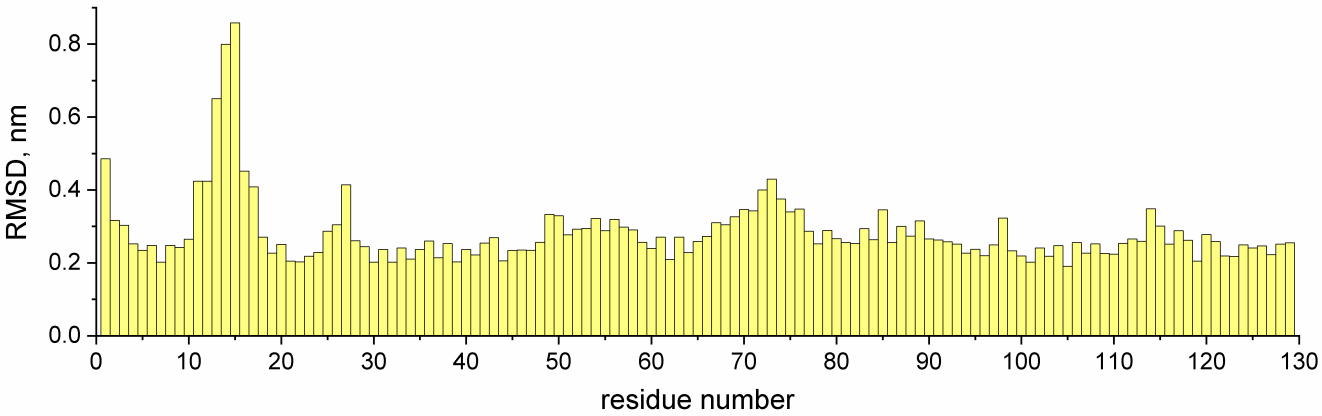


Рисунок 3.4 — RMSD отдельных остатков, усредненных по 89 димерам белков оболочки.

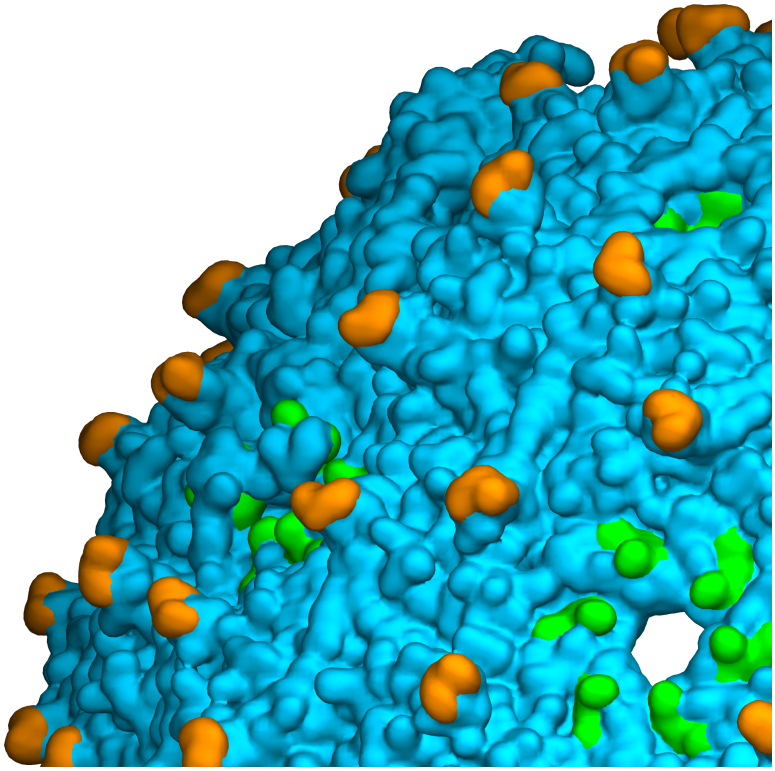


Рисунок 3.5 — Первоначальное расположение остатков 13–15, имеющих наибольшее смещение от структуры крио-ЭМ (оранжевый) и 75–78, образующих петлю FG (зеленый).

С биологической точки зрения важна вторичная структура (ВС) белка оболочки. Моделирование MD позволяет проверить, сохраняется ли ВС из условиий крио-ЭМ при комнатной температуре. Для этого было использовано программное обеспечение DSSP 2.0.4 [180].

Были вычислены «средние» вторичные структуры мономеров белков оболочки типов A, B и C по всем копиям в i) конфигурации крио-ЭМ, ii) конфигурации после ограниченного прогона (то есть начальной конформации для продуктивного прогона), iii) серии промежуточных конфигураций в интервале 5–45 нс и iv) конечной конфигурации, Рисунок 3.6. В частности, для каждого

аминокислотного остатка 1-129 была определена ВС, которая наиболее распространена в мономерах; а последовательность из 129 одноостаточных ВС сформировала «среднюю» ВС. В общей сложности в конечной конфигурации 16% (мономеры типа А) и 13% (мономеры типа В и С) аминокислотных остатков изменили свой средний ВС после 50 нс моделирования по сравнению с конформацией крио-ЭМ.

Ключевой областью белка оболочки является так называемая петля FG (остатки 75-78, Рисунок 3.6), которая играет ключевую роль в самосборке капсида [181]. Ее вторичная структура в мономерах типа А и С идентична и отличается от таковой в мономерах типа В. Проверка показывает, что ВС этого интервала в мономерах типа В остается неизменным в течение всего моделирования, тогда как в мономерах типа А и типа С происходит обратимая модификация в остатках 74–75. Однако соседний интервал 79–81 или 79–82 стабильно реструктурировался в мономеры типа А и В соответственно.

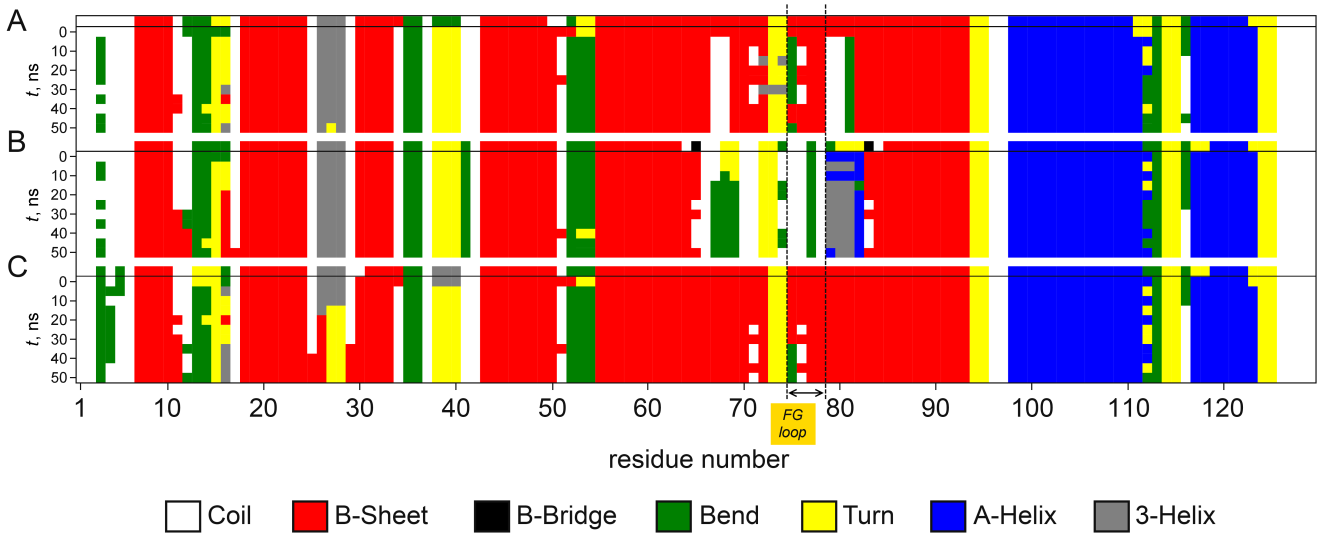


Рисунок 3.6 — Усредненные по капсиду вторичные структуры мономеров белков оболочки типов А, В и С в различные моменты времени моделирования; для каждого вида мономера первая подчеркнутая строка соответствует конфигурации крио-ЭМ; показан контур FG.

Распределение ионов

Капсид MS2 представляет собой сильно заряженную частицу. Поэтому ожидается образование двойного электрического слоя в солевом растворе. Были вычислены распределения ионов вокруг капсида с помощью функций радиального распределения (RDF) относительно центра масс капсида (ЦМ), Рисунок 3.7. RDF на интервалах 10–30 нс и 30–50 нс практически идентичны, что свидетельствует о достижении равновесного распределения ионов.

По-видимому, ионы Cl^- образуют отчетливый слой вблизи внутренней поверхности со средней плотностью в 2,1 раза выше, чем во внутреннем растворе. Удивительно, но ионы Na^+ имеют

очень низкую концентрацию вблизи отрицательной внешней поверхности (~1.3 раза). Это отличается от PCV2, где наблюдается их двукратное увеличение [25]. Небольшой пик при 12.6 нм образован ионами, расположенными в порах, впитанных стенкой или расположенных вблизи стенки в вогнутых областях, в то время как более крупный пик при 14.2 нм образован ионами вблизи выпуклых областей и ионами в растворе, находящемся на расстоянии нескольких слоев воды от поверхности. Таким образом, вся вирусоподобная частица (VLP) будет вести себя как отрицательно заряженная, что соответствует экспериментальным данным, показывающим отрицательные значения ζ-потенциала VLP MS2 [181—183].

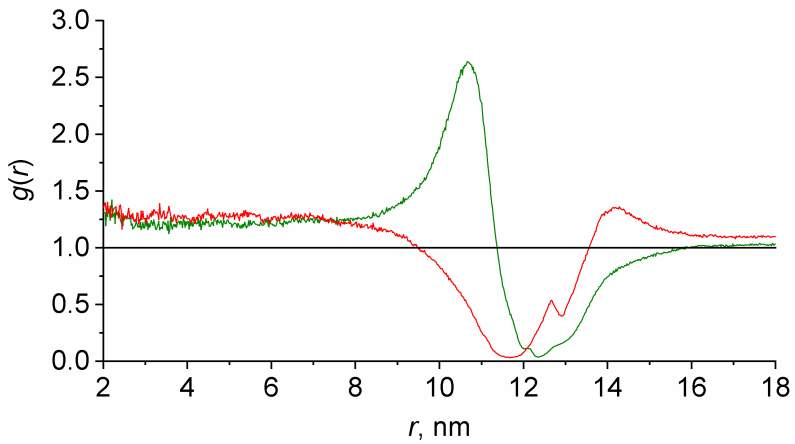


Рисунок 3.7 — Функции радиального распределения ионов Na^+ (красный) и Cl^- (зеленый) относительно ЦМ капсида, рассчитанные в ходе продуктивного прогона.

Потоки воды и ионов через капсид

Способность пропускать или блокировать поток молекул воды и ионов является важным свойством биомолекулярной стенки. Поэтому интересно рассмотреть стенку капсида в этом отношении. MD обеспечивает прямой способ количественной оценки движения раствора в стенке.

Был использован алгоритм, аналогичный нашей работе [22], который основан на работе Андо и др. [103]. Поскольку форма капсида MS2 гораздо ближе к сферическому слою, чем форма капсида PCV2, границу между внутренним и внешним раствором можно определить просто как определенное расстояние от ЦМ.

Атом считался находящимся внутри, если его расстояние до ЦМ капсида было меньше 11.7 нм, и он находился во внешнем растворе, если расстояние было больше 12.7 нм. Эти числа были выведены в результате визуального осмотра капсида в пакете VMD и соответствуют средним радиусам внутренней и внешней поверхностей. Радиусы не были функциями времени, поскольку размер капсида оставался практически постоянным после 10 нс (Рисунок 3.3, R_g изменился на <0.03 нм). Буферная зона толщиной 1 нм была введена для того, чтобы исключить молекулы, остающиеся в порах, и подсчитать только молекулы, фактически прошедшие через стенку капсида. Как будет показано ниже, величина расчетных потоков существенно не меняется в зависимости

от толщины буферной зоны. С использованием этого критерия были сформированы списки индексов молекул воды и ионов, находящихся внутри. Считалось, что частицы, находящиеся внутри (снаружи) капсида в один момент времени и снаружи (внутри) в другой момент времени, покинули (вошли) капсид; а сумма покинувших и вошедших частиц давала чистый поток через стенку за промежуток времени между рассматриваемыми моментами.

Расчет транспортируемых частиц проводился независимо на пяти 10-нс интервалах траектории (0–10 нс, 10–20 нс и т.д.). Для полученных чисел были рассчитаны среднее значение и стандартное отклонение; результаты для интервала 0–10 нс оказались несколько выделяющимися (вероятно, из-за остаточного уравновешивания размера капсида, Рисунок 3.3b), поэтому они были исключены из выборки. Графики для последнего интервала (40–50 нс) показаны на рисунке 3.8; графики для других интервалов очень похожи.

Через стенку идет интенсивный поток молекул воды, достигающий (5800±85) молекул на нс. Величина значения не зависит существенно от принятой толщины буфера: поток равен (5970±80) или (5630±85) молекул на нс, если буферная зона расположена на расстоянии 12–12,5 нм или 11,5–13 нм соответственно. В первые 15 нс изнутри капсида вышло 6800 молекул воды, после чего входящие и исходящие потоки стали взаимно компенсирующими.

Суммарный поток ионов также значителен и достигает (4.5±0.5) Na^+ и (4.9±0.4) Cl^- на нс. Значения существенно не меняются при увеличении или уменьшении толщины буфера. Для Cl^- и, в меньшей степени, Na^+, выходящий поток несколько сильнее входящего: через 50 нс содержание ионов внутри изменилось на -37 для Cl^- и -12 для Na^+.

Ситуация полностью противоположна капсиду PCV2, где транспорт воды был в 70 раз ниже, а транспорт ионов был незначительным [22]. Это объясняется большим диаметром пор в капсиде MS2, достигающим приблизительно 15 нм.

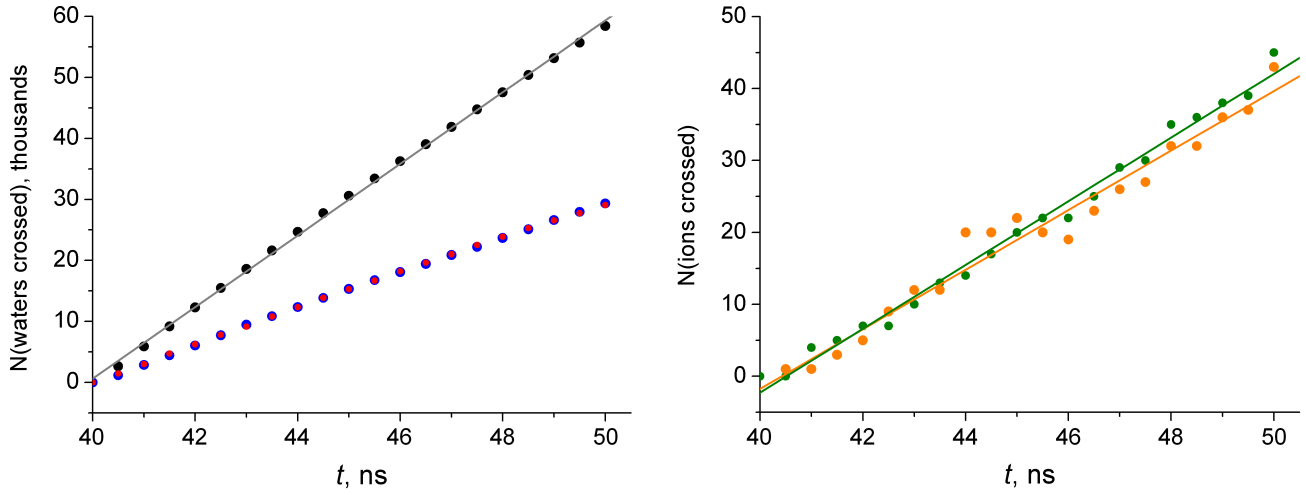


Рисунок 3.8 — Слева: число молекул воды, вошедших (красный) или покинувших (синий) внутреннюю часть капсида, по сравнению со структурой через 40 нс, суммой вошедших и покинувших (черный) и ее линейной аппроксимацией (серый); справа: число ионов Na^+ (оранжевый) или Cl^- (зеленый), пересекших стенку капсида в обоих направлениях, по сравнению с исходной структурой, и ее линейная аппроксимация.

3.2.3 Выводы

Была создана MD-модель капсида MS2, включая его асимметричные части (белок созревания и несколько деформированных белков вокруг) и одну подпоследовательность, отсутствующую в экспериментальных данных. После первоначальной корректировки атомистическая структура капсида при комнатной температуре и в физиологическом растворе оказалась стабильной и в значительной степени совпадала с измеренной крио-ЭМ структурой, за небольшим исключением трех остатков, расположенных на внешней поверхности.

Был проанализирован ряд структурных и динамических свойств капсида, которые невозможно получить экспериментально. Распределение ионов внутри и снаружи капсида выявило выраженный слой ионов хлора на внутренней поверхности, в то время как ионы натрия образовали неожиданно диффузный слой на внешней поверхности капсида. Наконец, поры в стенке капсида обеспечивают существенный транспорт воды и ионов в обоих направлениях.

3.3 Моделирование атомной молекулярной динамики целого капсида вируса PCV2 выявило критически важную роль правильного распределения ионов в стабильности капсида

Детальное знание структуры и функций вирусов на молекулярном уровне необходимо для разработки вакцин и лекарственных препаратов для лечения заболеваний, вызываемых этими вирусами. В последнее время получение такой информации становится возможным благодаря тому, что эксперименты позволяют измерять структуры вирусных частиц на атомистическом уровне разрешения методами рентгеновской кристаллографии и криоэлектронной микроскопии [74—77; 184]. Однако, поскольку измерения проводятся в нефизиологических условиях (например, низкая температура, кристаллическая упаковка образца), полученная структура может отличаться от клеточной среды. Кроме того, эксперименты не позволяют выделить некоторые части вирусных частиц (обычно гибкие или неупорядоченные), которые могут иметь важное значение для структуры и динамики вируса [74; 184].

Моделирование целых вирусных частиц с использованием молекулярной динамики является активно развивающейся областью исследований [8; 78; 101—103; 165; 185—188]. Они i) могут предоставить подробную информацию о структуре и динамике системы в растворе на атомистическом и молекулярном уровнях и ii) становятся осуществимыми для относительно длительных периодов моделирования за счет использования высокопроизводительных вычислений, включая специализированные компьютерные системы, такие как ANTON [189] и MDGRAPE-4 [190] и новые алгоритмы [2; 191].

В этом разделе описаны результаты моделирования вируса свиного цирковируса типа 2 (PCV2). Он относится к семейству Circoviridae и имеет кольцевую одноцепочечную ДНК, поме-

щенную внутри безоболочечного икосаэдрического капсида. Вирус PCV2 является патогенным и вызывает заболевания, связанные с цирковирусом свиней [74; 192; 193]. Этот вирус представляет собой удобный объект для моделирования поскольку он является одним из самых маленьких вирусов диаметром около 20 нм, а его атомистическая структура была определена экспериментально с разрешением 2.3 Ангстрема методом рентгеновской кристаллографии [74]. Капсид состоит из 60 копий одного белка (элементарных субъединиц), структура которых доступна в Protein Data Bank (PDB ID 3R0R). Важно отметить, что гибкие N-концы белков (33 остатка) не удалось разрешить в структуре и, следовательно, они отсутствуют в PDB. Кроме того, кристаллографическая структура не отражает полной картины взаимодействия белков, входящих в состав капсида, и растворителя при комнатной температуре [187; 194] где молекулы воды и ионы играют ключевую роль в динамической корректировке конформаций белка. Это связано с тем, что, как принято считать, гидратационная вода является основной «движущей силой» динамики белков, регулирующей как локальные, так и крупномасштабные движения в белках [1].

Белки капсида PCV2 способны спонтанно собираться в пустой капсид (также называемый вирусоподобными частицами), который стабилен при отсутствии ДНК внутри, как показано в эксперименте в физиологических условиях [99; 100] хотя и с довольно низким разрешением (была видна только общая форма пустого капсида). Однако причина такой стабильности пустого капсида без ДНК внутри до сих пор неясна.

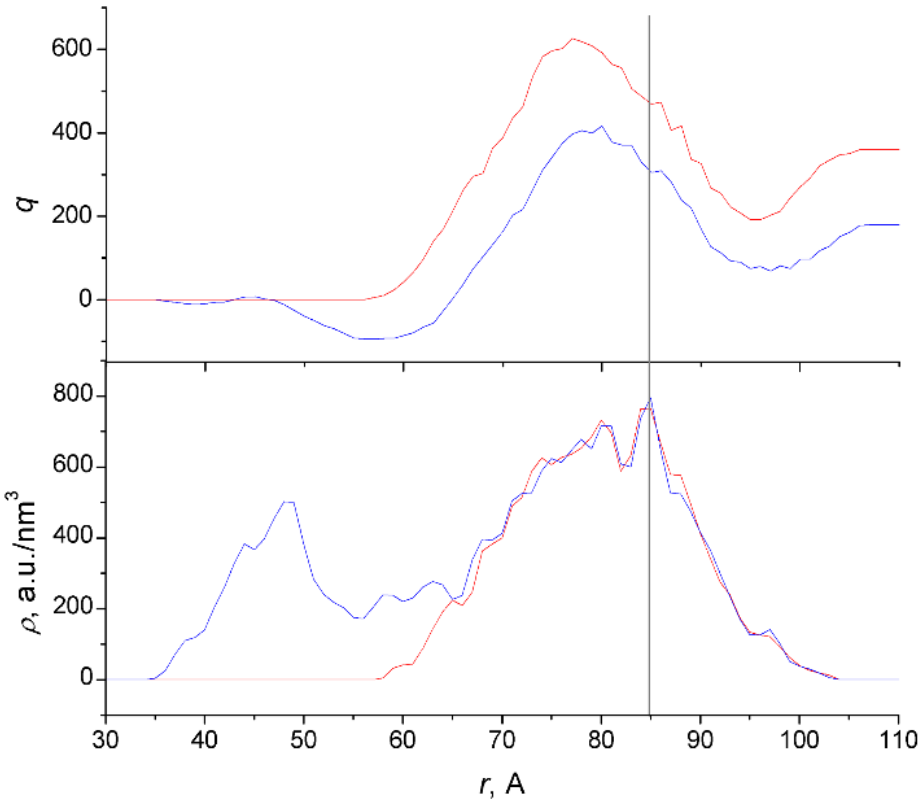


Рисунок 3.9 — Зависимости общего заряда (вверху) и плотности (внизу) атомов капсида от расстояния от центра масс капсида; черные кривые соответствуют капсиду без хвостов, красные кривые соответствуют капсиду с хвостами; вертикальная линия соответствует середине стенки капсида.

Выполнив моделирование в физиологических условиях (pH, соленость и температура) с использованием кристаллографических координат, была выдвинута гипотеза, что одна из возмож-

ных причин стабильности пустого капсида связана с зарядами, переносимыми ионами в растворе, в частности хлорид-ионами, которые могут имитировать геном в его отсутствии и заменять его функции, связанные с поддержанием структуры капсида. Имитация генома отрицательными зарядами уже была показана для сателлитного вируса табачной мозаики и сателлитного вируса некроза табака, где ионы хлора расположены в тех же областях, что и геном [186]. Поэтому для проверки этой гипотезы рассмотрим случаи с различным количеством ионов.

Было установлено что, хотя кристаллографическая структура достаточно хорошо согласуется со структурами MD, при моделировании капсидов с низким и высоким числом ионов возникает заметное различие, что подтверждает нашу гипотезу.

3.3.1 Моделируемая система

Капсид PCV2 состоит из 60 копий структурных субъединиц, расположенных в икосаэдрической симметричной структуре. Субъединицы собраны без зазоров, за исключением небольших пор на осях 5-го порядка, потенциально достаточно больших для транспорта воды и ионов.

Было проведено MD-моделирование всего вирусного капсида с атомистическим разрешением в атомитической воде. Первоначальная структура элементарной субъединицы была взята из рентгеновских измерений, за исключением первых 33 аминокислот N-конца (обозначаемые «хвостами»), которые расположены внутри капсида [74]. Последовательность отсутствующего N-конца известна из последовательности гена, используемого для экспрессии элементарной субъединицы. В данной работе была использована структура вирусоподобной частицы консенсусной последовательности PCV2, которая была выбрана для рентгеновской кристаллографии как наиболее удобная для экспрессии. Структуру N-конца определить не удалось, скорее всего, из-за его неупорядоченной конформации [184]. Обратите внимание, что используемые N-концы имеют слегка отрицательный заряд в отличие от настоящего вируса, где они обладают значительным положительным зарядом. Это облегчает их взаимодействие с отрицательно заряженной ДНК. Однако в нашем случае ДНК отсутствует, и, таким образом, заряд N-конца не имеет решающего значения для самосборки капсида.

При получении исходных капсидов без хвостов элементарная субъединица без N-конца была скопирована 60 раз с учетом симметрии капсида и собрана в полный икосаэдрический капсид с использованием утилит VIPERdb [105]. Та же процедура была применена для сборки капсида, включая хвосты, при этом структура хвостов была реконструирована на основе известной последовательности с использованием метода моделирования гомологии Modeller [104] и добавление его к элементарной субъединице из PDB. Затем были проведены структурные уточнения хвостов внутри всего капсида с использованием молекулярной операционной среды (MOE) [195] (Chemical Computing Group Inc.).

Капсид несет большой общий положительный заряд в 360 e (без хвостов) или 180 e (с хвостами, которые заряжены отрицательно). Заряд распределен неравномерно между внутренней и внешней поверхностями. Для того чтобы корректно нейтрализовать этот избыточный заряд, бы-

ли построены зависимости интегрального заряда атомов капсида от расстояния от центра масс капсида (ЦМ), рисунок 3.9. Для удаления шума с графиков было использовано усреднение.

Таблица 8 — Первоначальный состав моделируемой системы

система	капсид без хвостов				капсид с хвостами			
	NI		LI		NI		LI	
размер ячейки (нм)	24.68		24.68		24.65		24.65	
	внутри	снаружи	внутри	снаружи	внутри	снаружи	внутри	снаружи
$N_{compens}$	606 Cl^-	246 Na^+	300 Cl^-	60 Cl^-	407 Cl^-	227 Na^+	200 Cl^-	20 Na^+
N_{Na+}	1966		1720		1947		1740	
N_{Cl-}	2326		2080		2127		1920	
N_{water}	568707		569199		559368		559782	
N_{atoms}	1898573		1899557		1897998		1898826	

NI — Система с нейтрализующим числом ионов; LI — Система с малым числом ионов;
$N_{compens}$ — число ионов, добавляемых к капсидам для компенсации заряда.
N_{Na+} – общее количество ионов натрия; N_{Cl-} – общее количество ионов хлора;
N_{water} – общее количество молекул воды; N_{atoms} – общее количество атомов в системе

Для сравнения также построен график плотности атомов капсида в зависимости от расстояния от ЦМ капсида. Отчетливый провал на графике для капсида с хвостами соответствует хвостам, которые вытянуты к центру капсида. Максимумы на графиках плотности расположены при ~8.5 нм и соответствует середине стенки капсида. Однако максимумы зарядовых кривых расположены на более коротком расстоянии ~7.8 нм. Это позволило сделать вывод, что положительный заряд сосредоточен на внутренней поверхности капсида и имеет величину, равную значению в максимуме графика заряда (606 для капсида без хвостов и 407 для капсида с ними). Внешняя поверхность несет отрицательный заряд меньшей величины (246 для капсида без хвостов и 227 для капсида с ними). В целом капсид имеет положительный заряд. Поскольку капсид оказался высокополярным, было решено по отдельности нейтрализовать заряд каждой поверхности.

Положительный заряд внутренней поверхности нейтрализовали ионами Cl^-, помещенными внутрь капсида (606 ионов для капсида без хвостов и 407 ионов для капсида с ними). Отрицательный заряд внешней поверхности нейтрализовали ионами Na^+, помещенными в объем раствора (246 для капсида без хвостов и 227 для капсида с ними). Системы с таким числом ионов далее в тексте будем называть «нейтрализующими». Наконец, для моделирования физиологических условий были добавлены 1720 ионов Na^+ и Cl^-, равномерно распределенных в системе, что соответствует 0.9 массовых процента буфера NaCl.

Также для обоих капсидов было проведено моделирование с уменьшенным количеством ионов, поместив половину необходимого количества Cl^- внутрь капсидов (300 в капсид без хвостов и 200 в капсид с хвостами). Для поддержания электронейтральности всей системы также были добавлены ионы снаружи капсида (60 Cl^- для капсида без хвостов и 20 Na^+ для капсида с

хвостами). Системы с таким числом ионов будем называть «низкими», чтобы отличать их от «нейтрализующих» систем. Также был добавлен NaCl в физиологической концентрации (1720 Na^+ и Cl^-). Подробный алгоритм добавления ионов представлен в приложении Ж. Важно, что все моделируемые системы в целом были электронейтральными.

Для каждой симуляции использовались два разных силовых поля: GROMOS53a6 [116] и AMBER03 [196], чтобы исследовать зависимость результатов от деталей силового поля межатомных взаимодействий.

Для разных силовых полей также были выбраны разные модели воды: SPC для GROMOS и TIP3P для AMBER. Всего было изучено восемь систем: четыре капсида с нейтрализующим числом ионов (два капсида без хвостов и два капсида с хвостами) и четыре таких же системы с дефицитом ионов внутри. Таблица 8 обобщает начальные условия систем в силовом поле AMBER.

Начальные структуры были приведены в равновесие до начала моделирования MD посредством минимизации энергии и последующего прогона MD с ограничениями положения капсида, увеличивая температуру с 200 K до 300 K (см. приложение Ж). Для моделирования MD использовалась обработка кулоновских взаимодействий методом PME. После первоначального уравновешивания каждая система был промоделирована в течение 10 нс и эти интервалы времени были использованы для анализа.

В конце 10-нс моделирования все капсиды с нейтрализующим числом ионов сохранили симметрию, тогда как моделирования с меньшим числом ионов потеряли симметрию капсида во время моделирования и показали явные признаки распада.

3.3.2 Результаты

Стабильность, структура и распределение заряда

Для количественной демонстрации устойчивости конструкции были проанализированы изменения среднеквадратичного отклонения (RMSD) и радиуса инерции (R_g) для атомов остова всего капсида со временем в силовом поле AMBER, рисунок 3.10 (графики силового поля GROMOS представлены в приложении). Расчеты RMSD и R_g были выполнены для одного и того же набора аминокислот для всех систем и не включали хвосты.

Сначала рассмотрим результаты, полученные для капсидов с нейтрализующим числом ионов.

Графики 3.10a,c показывают RMSD и R_g для капсидов с нейтрализующим числом ионов без хвостов (черная кривая) и с хвостами (красная кривая). Стабильное значение RMSD для капсида без хвостов при 0.2 нм указывает на стабильное состояние вирусной частицы и небольшое отклонение от исходной структуры, что является обычным консенсусом в литературе [101; 166] и вместе с постоянными значениями R_g позволяют сделать вывод, что капсид без хвостов стабилен в тече-

ние 10 нс моделирования. Несмотря на наличие медленного роста RMSD для капсида с хвостами можно сделать вывод, что капсид находится в стабильном состоянии, поскольку R_g остается постоянным. Рост RMSD в случае с хвостами можно объяснить локальными перестройками в белках или некоторыми перемещениями частей белка, которые связаны с гибкими хвостами, не влияющими на общую форму капсида. Такое поведение совпадает с экспериментальными данными.

Отдельные результаты получены для случая, когда меньшее количество ионов помещено внутрь капсида, рисунки 3.10c,d. Обе структуры демонстрируют непрерывный рост RMSD. Значение для капсида без хвостов достигает почти ~0.4 нм, а значение для структуры с хвостами - почти ~0.3 нм. Кроме того, непрерывное увеличение R_g для обоих капсидов предполагает нестабильность начальной конформации каждого капсида с меньшим числом ионов.

Отмечено, что для всех случаев с малым количеством ионов капсид демонстрирует резкие изменения и теряет симметрию, что говорит о том, что исходная структура нестабильна в данных условиях. Потеря симметрии для капсида без хвостов и с хвостами происходит с разной скоростью. Сделан вывод, что капсид с хвостами имеет дополнительные взаимодействия между хвостами и внутренней поверхностью капсида, которые играют роль якорей, препятствующих быстрому разрушению структуры, однако этого недостаточно для поддержания симметричной структуры.

Графики для тех же систем в силовом поле GROMOS показывают похожее поведение (приложение Ж).

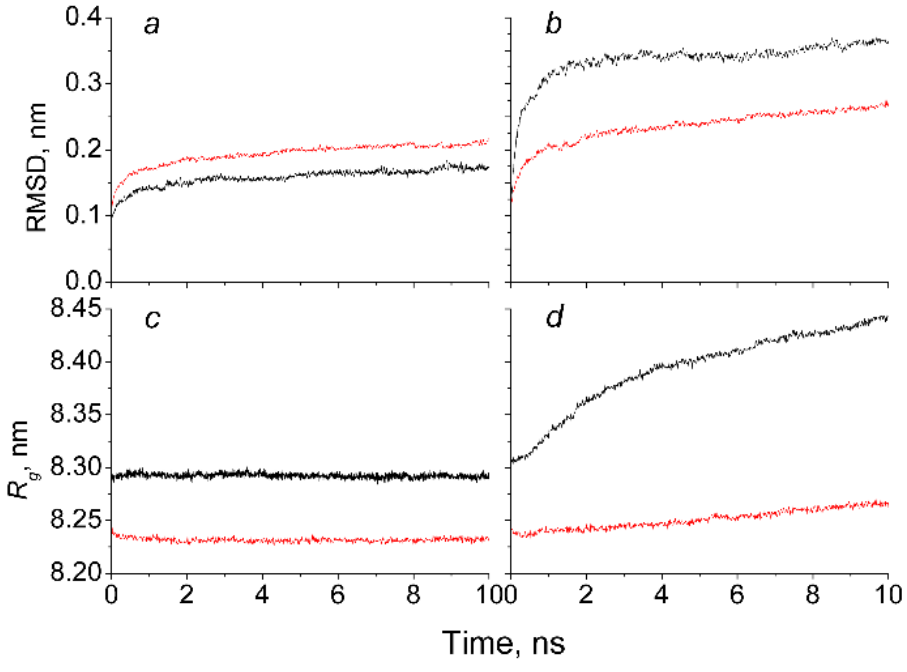


Рисунок 3.10 — Изменение во времени среднеквадратичного отклонения, RMSD, (верхние графики) и радиуса инерции, R_g, (нижние графики) атомов основной цепи относительно их первоначальной структуры как эталонной структуры в силовом поле AMBER; расчеты RMSD и R_g не включали хвостовую последовательность для всех структур и состояли из одинакового числа аминокислот; красный цвет обозначает капсид с хвостами, черный — капсид без хвостов; a, c – капсиды с нейтрализующим числом ионов, b, d – капсид с малым числом ионов.

Совпадающие значения R_g относительно осей x, y и z неизменны во времени, рисунок 3.11e,f и демонстрируют сохранение симметрии для системы с нейтрализующим числом ионов, рисунок 3.11e, и потеря симметрии для системы с меньшим числом ионов который показывает, что капсид расширяется неравномерно, образуя большие трещины в оболочке. Это поведение не имеет тенденции к изменению со временем, даже после дополнительных 10 нс моделирования, рисунок 3.11f.

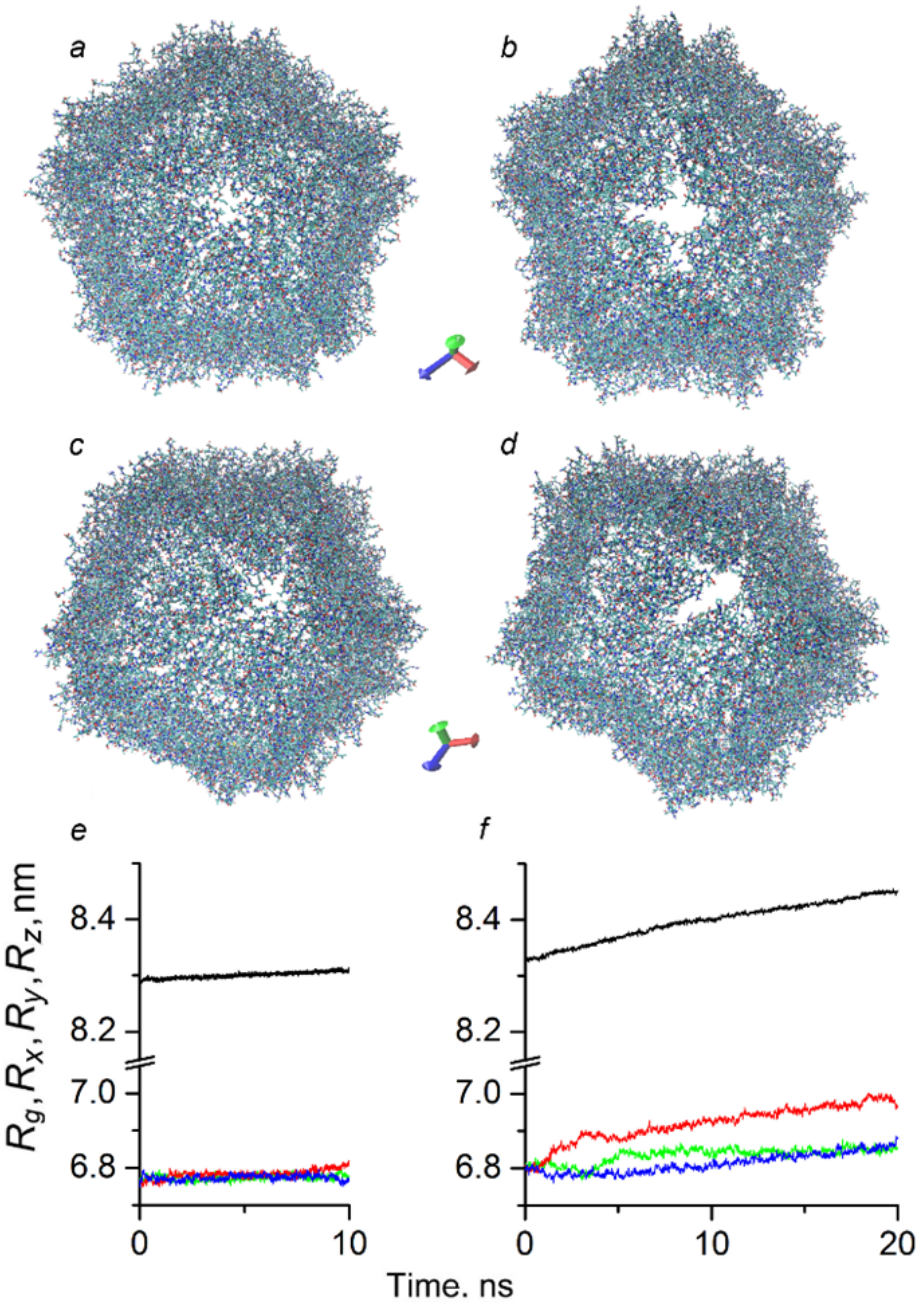


Рисунок 3.11 — Атомистические структуры капсида PCV2 без хвостов после моделирования: (a,c) атомистические структуры капсида PCV2 с нейтрализующим числом ионов, (e) радиус инерции для разных осей (зеленый — ось X, красный — ось Y, синий — ось Z); (b,d) — структуры капсида с меньшим числом ионов и (f) радиус инерции для разных осей.

Подводя итог, можно сказать, что результаты, полученные для восьми систем после 10 нс моделирования MD, показывают, что количество ионов внутри играет ключевую роль в стабильности капсида и поддержании его симметрии.

Как классический MD, так и гибридный методы позволили получить стабильный капсид с небольшим среднеквадратичным отклонением относительно кристаллической структуры. Это означает, что экспериментально измеренная атомистическая структура при низкой температуре в нефизиологических условиях сохраняется в растворе при комнатной температуре с клеточноподобной средой.

Структура капсида, промоделированная с использованием нашего гибридного метода, показывает небольшое среднеквадратичное отклонение по сравнению с кристаллической структурой, Рисунок 3.12. Это показывает, что наша методология сохраняет атомистические структурные свойства в атомистической зоне системы.

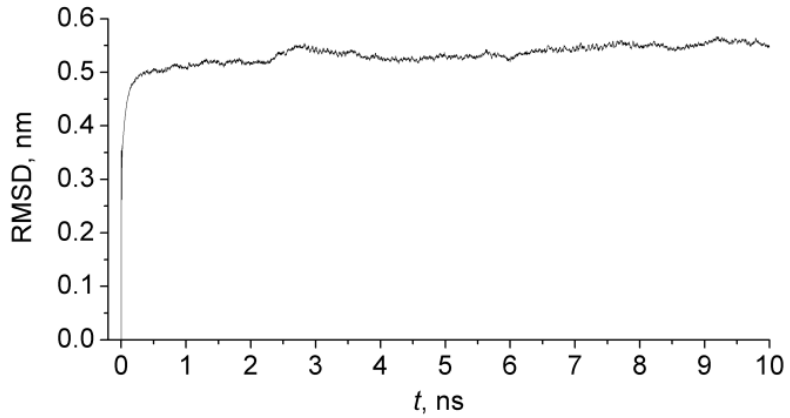


Рисунок 3.12 — Среднеквадратичное смещение атомов основной цепи капсида от начальной конфигурации во время гибридного расчета MD/гидродинамики.

Для оценки изменения по сравнению с исходной структурой были дополнительно построены зависимости интегрального заряда атомов капсида от расстояния до центра масс капсида для конечной структуры. Графики даны в приложении Ж. Графики очень похожи, высоты максимумов остались неизменными.

Наблюдаемая потеря симметрии капсида и образование трещин в определенных областях позволяет предположить наличие качественного различия в распределении ионов внутри капсидов. Рисунок 3.13d показывает, что рядом с положительной внутренней стенкой капсида образуется плотный слой ионов хлора, если количество ионов правильное. Однако в случае малого количества ионов, рисунок 3.13b, ионы хлора неравномерно накапливаются на внутренней поверхности капсида, образуя отчетливые пустоты вблизи пор. Показан случай распределения ионов в капсиде без хвостов, поскольку наличие хвостов усложняет картину.

Предложено два возможных механизма, объясняющих потерю симметрии капсида.

1) Когда внутри капсида недостаточно ионов хлора, его положительная внутренняя стенка покрывается отрицательными ионами неравномерно, некоторые участки стенки имеют дефицит зарядов отрицательных ионов (эти участки в основном находятся вблизи пор капсида), рисунок 3.13b. Из-за этого несбалансированного распределения заряда ненейтрализованные положительные части внутренней стенки притягиваются к отрицательному облаку ионов хлора внутри

капсида, в результате чего эта часть стенки схлопывается внутрь и капсид в этом месте приобретает приблизительно сферическую форму. Это коррелирует с выводами [186], где авторы наблюдали накопление ионов хлора внутри пор симметрии 5-го порядка.

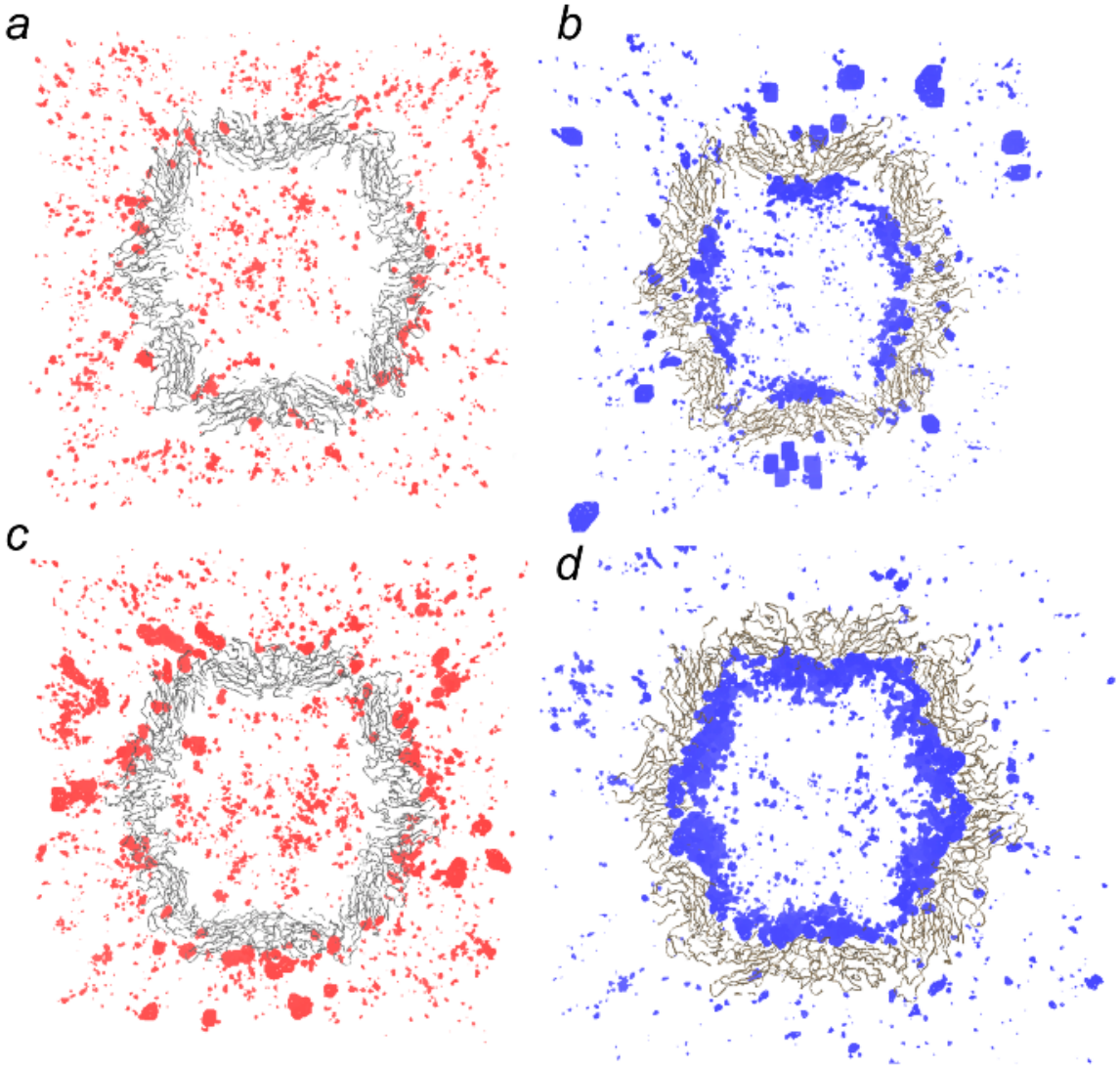


Рисунок 3.13 — Распределение ионов после первой наносекунды моделирования MD для малого числа ионов (вверху) и для нейтрализующего числа ионов (внизу); красная плотность (a,c) представляет собой плотность ионов натрия, а синяя плотность (b,d) — плотность ионов хлора, усредненную за одну наносекунду моделирования.

2) Как было видно в ходе моделирования, недостаток ионов хлора внутри и вокруг пор позволяет ионам натрия проходить через поры, увеличивая и без того избыточный не нейтрализованный положительный заряд внутренней стенки вблизи поры, что приводит к ослаблению взаимодействия между субъединицами и в конечном итоге к появлению широких трещин в капсиде в этих местах. В случае нейтрализации количества ионов хлора внутри капсида ионы хлора образуют плотный слой отрицательного заряда, равномерно распределенного по внутренней поверхности. Этому слою соответствует аналогичный положительный слой ионов натрия на внешней поверхности капсида, рисунок 3.13c,d. Эти два слоя больших противоположных зарядов сильно притягиваются друг к другу, стабилизируя всю структуру капсида.

3.3.3 Поток ионов и воды через стенку капсида PCV2

Был рассчитан транспорт ионов натрия, ионов хлора и молекул воды через стенку капсида в течение 10 нс. Рисунок 3.14a,c,e показывают результаты для системы A, а Рисунок 3.14b,d,f - для системы Б (как описано в разделе 1.3.6). Важно отметить, что диапазоны осей Y на соответствующих графиках равны для облегчения сравнения.

Рисунок 3.14a,b демонстрирует изменения динамики ионов хлора (синие точки) и натрия (красные точки) для систем A и B. Количество ионов натрия и ионов хлора практически одинаково в течение всех 10 нс. Отметим, что количество ионов, проходящих через стенку капсида, ничтожно мало, и не существует взаимокомпенсирующих потоков ионов в капсид и из него.

Более интересные результаты наблюдаются для потока молекул воды. Рисунок 3.14c,d демонстрирует зависимость количества молекул воды внутри капсида от времени. Это число колеблется вокруг постоянного значения. Колебания имеют больший период в системе A (нет хвостов), Рисунок 3.14c, и меньше в системе B (с хвостами), Рисунок3.14d. Тем не менее, поток молекул воды можно обнаружить в течение моделируемых 10 нс.

Рисунок 3.14e,f показывают общее количество молекул воды, которые пересекли капсид между начальной конфигурацией и конфигурацией в момент времени t. Он вычисляется как сумма (i) числа молекул, которые остаются внутри капсида в начальной конфигурации и находятся снаружи него в момент времени t, и (ii) наоборот. Средний поток, полученный как наклон линейной аппроксимации графиков, составляет 73±3 молекулы на нс для системы A и 60±2 для системы B. Две первые точки не использовались для подгонки из-за остаточного уравновешивания, происходящего после снятия ограничений с атомов капсида.

В гибридном случае поток ионов также отсутствует, что согласуется с результатами чистого MD (Рисунок 3.15a). Напротив, поток воды составляет 66±2 молекулы на нс, что несколько (на 10%) ниже, чем при чистом MD-моделировании, Рисунок 3.15b. Это показывает, что гидродинамическая часть моделирования по-прежнему искажает динамику атомистического движения, но эффект имеет ограниченную величину. Заметным эффектом является большая разница между величинами входящих и исходящих потоков.

3.3.4 Выводы

В заключение следует отметить, что наше моделирование демонстрирует, что нейтрализующее количество ионов (особенно ионов хлора) необходимо для стабильности капсида, ионы хлора покрывают положительную внутреннюю стенку капсида, создавая плотный слой отрицательного заряда.

Наконец, сравнивая два использованных силовых поля, приходим к выводу, что AMBER дает результат в виде стабильного капсида, который ближе соответствует кристаллографической

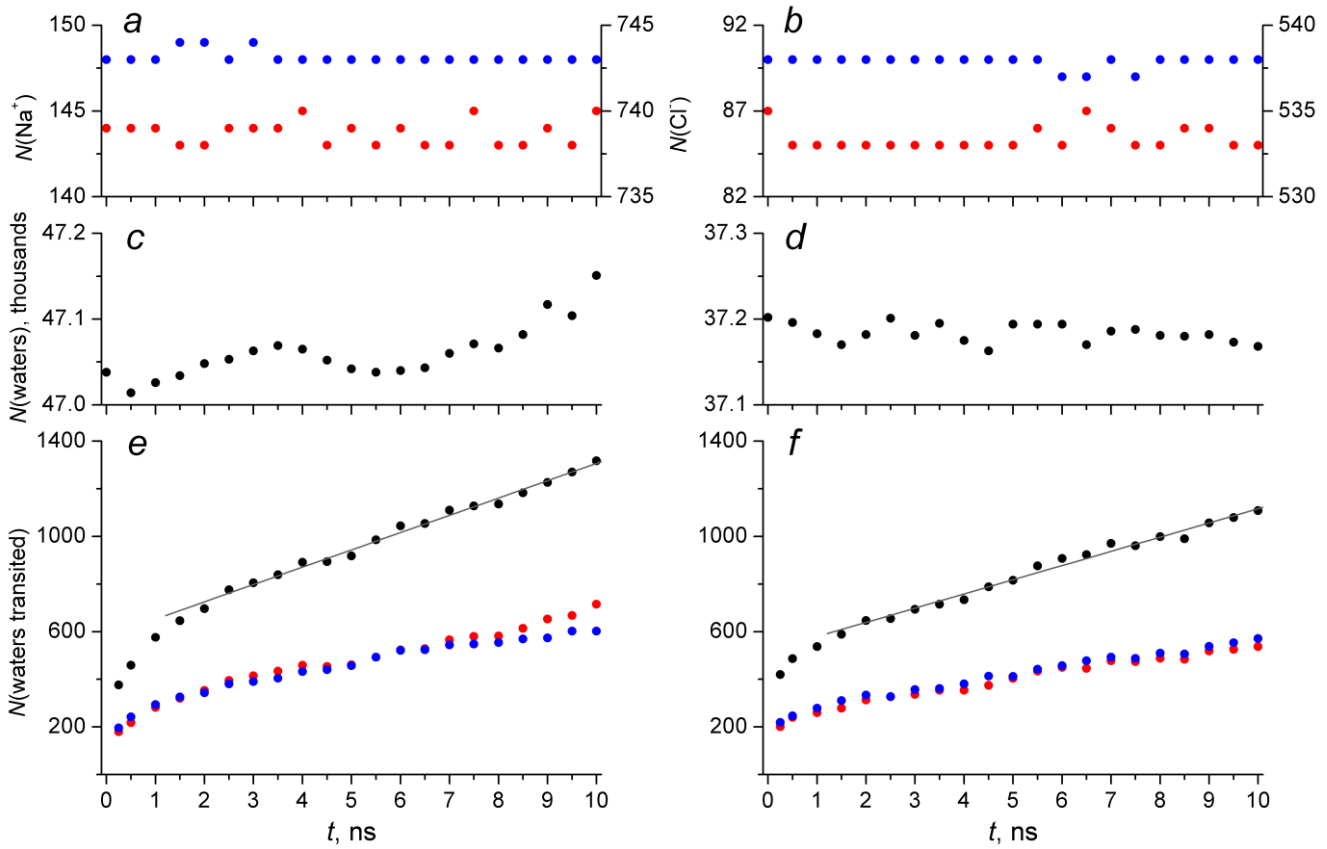


Рисунок 3.14 — Поток воды и ионов в системе A (a,c,e) и системе B (b,d,f) в течение 10 нс моделирования; a,b) количество ионов Na^+ (красный) и Cl^- (синий) внутри капсида как функция времени, c,d) количество молекул воды внутри капсида как функция времени, e,f) общее количество молекул воды, которые пересекли стенку капсида в обоих направлениях по сравнению с исходной структурой (черный) и ее линейной аппроксимацией (серый), количество вошедших (красный) и вышедших (синий) молекул воды.

структуре. Неясно, означает ли это, что силовое поле AMBER лучше подходит для моделирования PCV2, или GROMOS также представит стабильный результат; для получения окончательных результатов требуется больше времени моделирования.

Кроме того, наши результаты показывают, что среднеквадратичное отклонение между структурой с нейтрализующим числом ионов с N-концами и без них в растворе при 300 K и рентгеновской структурой в среде кристаллической упаковки при 100 K не является высоким. Это показывает, что кристаллографическая структура отражает структуру in vivo.

Был проведен анализ потоков молекул воды и ионов через стенку вирусного капсида. Анализ был выполнен для структур вирусного капсида, полученных с помощью рентгеновского анализа и промоделированных с использованием полноатомной молекулярной динамики с атомистической моделью растворителя в течение 10 нс. Анализ показывает, что в течение 10 нс ионы из раствора практически не пересекают оболочку капсида. Это контрастирует с молекулами воды, которые обладают способностью проходить через стенку капсида в обоих направлениях. Моделирование полиовируса методом MD [103] демонстрирует аналогичное поведение молекулярной проницаемости капсида полиовируса. Авторы пришли к выводу, что капсид полиовируса может действовать как полупроницаемая мембрана. Здесь наблюдается похожее поведение. Эта функ-

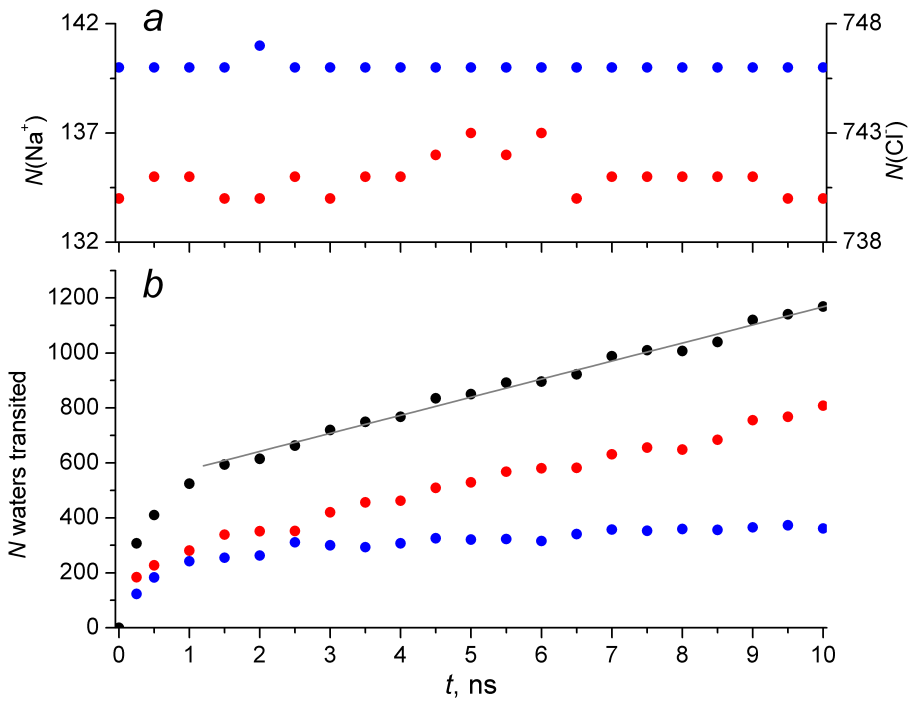


Рисунок 3.15 — а) Количество ионов Na^+ (красный) и Cl^- (синий) внутри капсида как функция времени; б) общее количество молекул воды, которые пересекли стенку капсида в обоих направлениях по сравнению с исходной структурой (черный) и ее линейной аппроксимацией (серый), количество вошедших (красный) и вышедших (синий) молекул воды во время гибридного расчета MD/гидродинамики.

ция капсида может играть важную механическую роль в выдерживании высокого давления или изменений давления. Моделирование той же системы в гибридном подходе MD/гидродинамики позволило получить структуру вируса, а также величины потоков ионов и воды, близкие к полученным при чистом моделировании MD, что подтверждает правильность нашего подхода.

3.4 Подробности распределения заряда в стабильном вирусном капсиде PCV2

На сегодняшний день одной из самых сложных проблем в экспериментальном и вычислительном плане в моделировании вирусов является локализация генома внутри капсида и его взаимодействие с капсидом. Экспериментально асимметричная упаковка цепи ДНК или РНК не позволяет кристаллографам получить ее структуру, в то время как разрешение одночастичной крио-ЭМ часто недостаточно для определения атомистической структуры генома. Отсутствие надежной экспериментальной информации делает практически невозможным построение разумной начальной структуры для моделирования MD.

Интересную корреляцию между расположением ионов хлора в растворе и молекулой генома обнаружили авторы [186]. Они наблюдали, что при отсутствии генома ионы хлора занимают те же места рядом с внутренней стенкой капсида, что и фосфатные группы генома. Наше полноатом-

ное моделирование MD [7] пустого капсида цирковируса свиней (PCV2) показало, что количество ионов внутри капсида имеет решающее значение для его стабильности при отсутствии генома внутри. Экспериментально известно, что капсид PCV2 может быть стабильным без генома [74; 197] и его атомистическая структура была измерена с помощью рентгеновского анализа [74]. Наши результаты показывают, что дефицит ионов хлора вблизи внутренней поверхности капсида дестабилизирует его структуру.

Целью этого раздела является изучение молекулярных деталей распределения ионов и определение того, какие области внутренней поверхности должны быть покрыты ионами хлора для сохранения целостности структуры капсида. Основная идея состоит в сравнении результатов MD для стабильного капсида и для ситуации с нестабильным капсидом. Системы готовятся таким образом, что единственное различие между ними заключается в количестве ионов; все остальные свойства, включая структуру самого капсида, сохраняются идентичными. Был проведен детальный анализ распределения ионов и выявлены критические места, вызывающие дестабилизацию капсида.

3.4.1 Детали моделирования

Подготовка системы и протоколы моделирования MD приведены в наших предыдущих публикациях [7; 22]. Субъединица белка капсида PCV2 взята из Protein Data Bank (PDB) [198] и полный пустой капсид реконструирован с помощью программы VIPERdb [105].

Моделирование MD проводилось с использованием GROMACS [126; 199], программа VMD [200] использовался для визуализации. Моделирование и анализ проводились в силовом поле AMBER с использованием модели воды TIP3P.

Капсид имеет положительный заряд $+360$ e, и заряд распределен непропорционально между внутренней и внешней поверхностями [7]. С целью нейтрализации этого заряда было проведено две отдельные процедуры добавления ионов хлора внутрь капсида и ионов натрия снаружи. Число ионов составили 606 Cl^- и 246 Na^+, поскольку было обнаружено, что для полной нейтрализации положительного заряда внутренней поверхности необходим заряд -606. Наконец, на третьем этапе были добавлены 1720 иона Na^+ и 1720 иона Cl^-, случайным образом распределённых по ячейке, чтобы имитировать физиологический раствор с концентрацией 0.15 M. Эту систему будем называть естественной системой (EC). Для приготовления искусственной системы (ИС) была добавлена половина необходимого количества ионов хлора внутрь капсида, т.е. 300 Cl^-. Поскольку заряд капсида составляет $+360$ e, во внешний раствор вместо Na^+ было помещено 60 Cl^-. Также был добавлен тот же буфер хлорида натрия. Обе системы в целом были электронейтральными.

Была проведена минимизация энергии, после чего проведено 1 нс MD моделирование систем с фиксированными положениями всех атомов капсида. В ходе подготовки к продуктивному прогону температура постепенно повышалась с 200K до 300K [7; 22]. Наконец, было проведено продуктивное моделирование MD в течение 10 нс при 300K. Для расчета электростатических

взаимодействий использовался метод PME, а для ван-дер-ваальсовых взаимодействий — метод отсечения с радиусом 1 нм.

В естественной структуре капсида имеются поры, которые соединяют внутреннюю и внешнюю части вируса. В нашей работе [22] было обнаружено, что ионы натрия и хлорида не могут проходить через поры. Однако молекулы воды могут диффундировать в обоих направлениях. Это позволило нам сделать вывод, что стенка капсида PCV2 функционирует как полупроницаемая мембрана, аналогично тому, что было показано для капсида полиовируса [103].

Был проведен анализ функции радиального распределения (RDF) ионов в растворе в зависимости от расстояния от центра капсида для естественных и искусственных систем. Анализ проводился с использованием последнего интервала в 200 пс из 1 нс прогона с ограничениями. Продуктивный прогон, когда ограничения были сняты, не использовался, поскольку анализировались причины, которые впоследствии привели к деформации капсида, произошедшей когда капсид не был зафиксирован. В течение 800 пс ограниченного MD-прогона ионы могли свободно перемещаться и достигать равновесного распределения.

3.4.2 Результаты и обсуждение

На рисунке 3.16 показана типичная деформация капсида в искусственной системе с неоптимальным количеством ионов внутри. Количественно показано, что эта деформация не имеет тенденции к прекращению и в конечном итоге приводит к образованию больших трещин в стенке капсида, которые ставят под угрозу общую целостность вируса [7]. Также показано, что это вызвано отсутствием отрицательного заряда в слое раствора, прилегающем к внутренней поверхности капсида, который компенсирует положительный заряд поверхности.

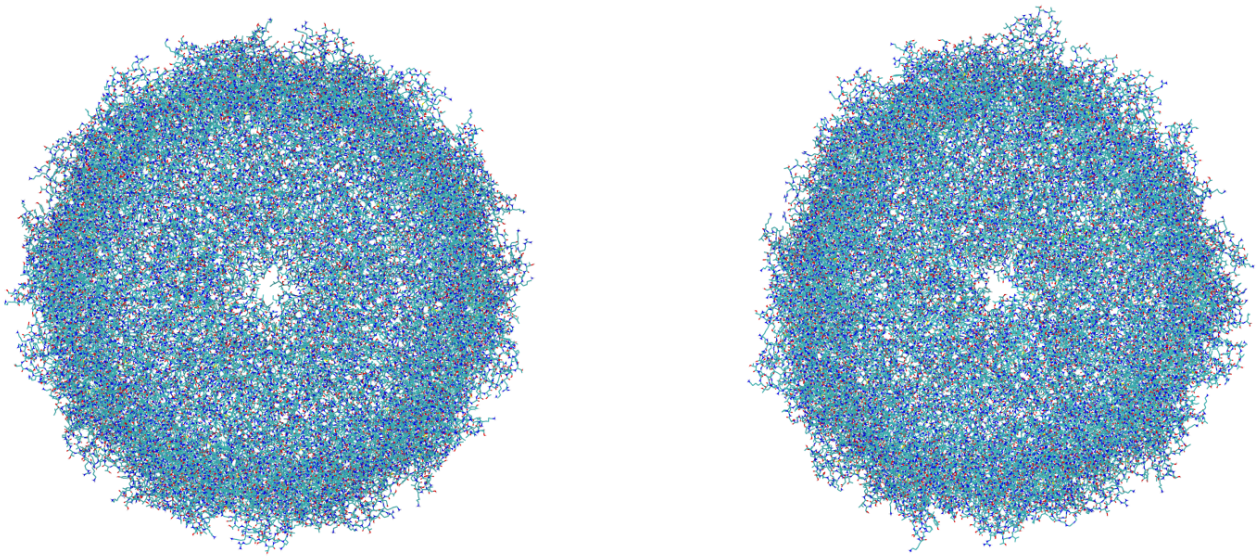


Рисунок 3.16 — Стабильные (слева) и деформированные (справа) капсиды; деформация прогрессирует и приводит к разрушению капсида.

На снимке распределения ионов в стабильной системе видно довольно плотное скопление ионов хлора вблизи внутренней поверхности, Рисунок 3.17. Это, конечно, ожидаемо; однако количественные данные раскрывают интересные особенности этого распределения, в частности, по сравнению с нестабильной системой. На рисунке 3.18 показаны функции распределения ионов в зависимости от расстояния от центра масс капсида (ЦМ).

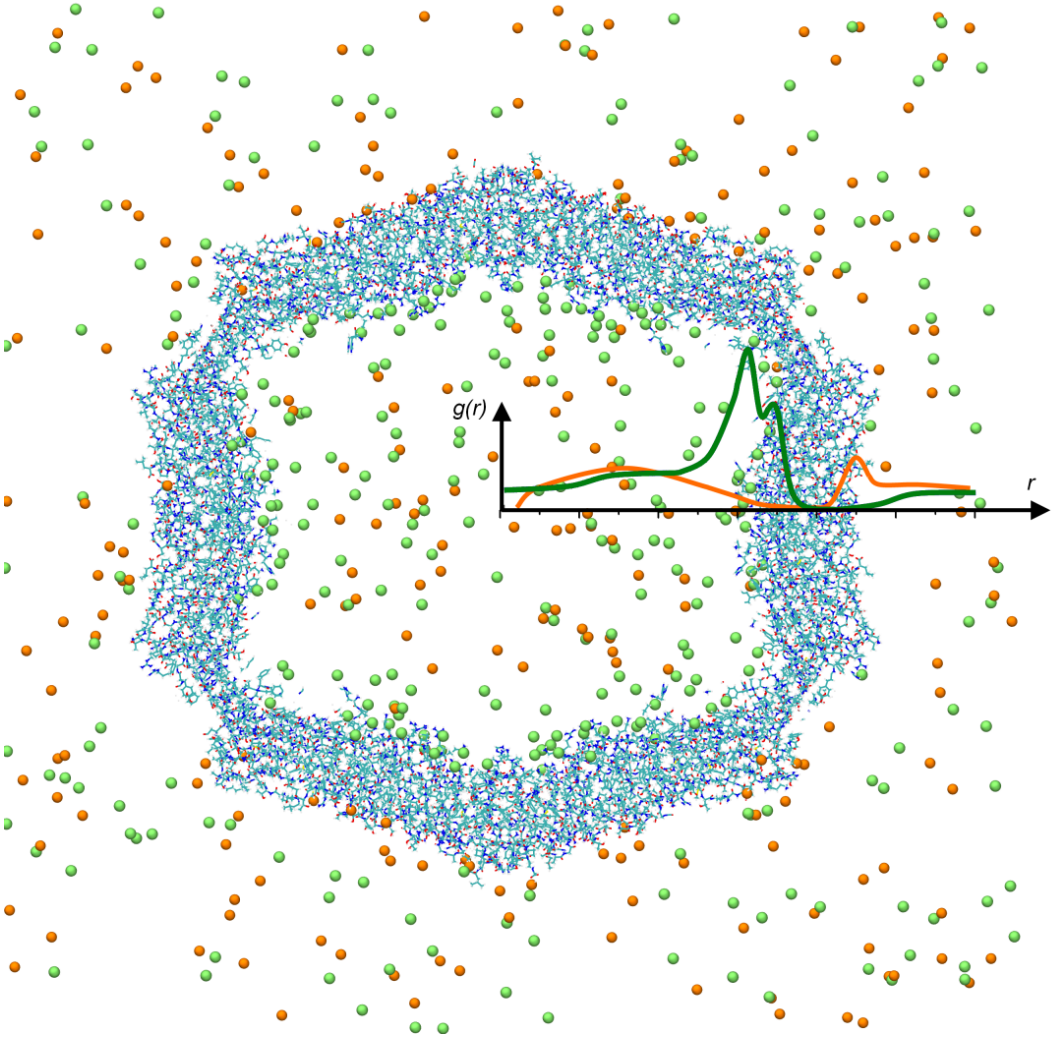


Рисунок 3.17 — Схематическое изображение распределения ионов Cl^- (зеленый) и Na^+ (оранжевый) в системе и RDF, количественно определяющая их концентрацию в зависимости от расстояния от центра масс капсида.

Видно, что в капсиде с нейтрализующим числом ионов наблюдается расщепленный пик при ~7 нм, тогда как в капсиде с малым количеством ионов этот пик отсутствует. Рисунок 3.16 также дает подсказку для объяснения этой особенности: хотя форма капсида близка к сферической, он имеет выраженные внутренние «полости» в окрестностях пор. Они образованы икосаэдрической формой капсида, отчетливо видимой, если выделить атомы, находящиеся на расстоянии более 7.5 нм от ЦМ (желтые области на рисунке 3.20). Число Cl^- в этих «углах» достигает 190, и в нестабильной системе они практически полностью отсутствуют. Пик на ~6 нм также ниже для нестабильного капсида, в котором отсутствует примерно 100 ионов.

Ионы натрия в обоих случаях распределены внутри капсида одинаково. Однако в естественной системе вокруг внешней поверхности существует дополнительный слой ионов Na^+. Это

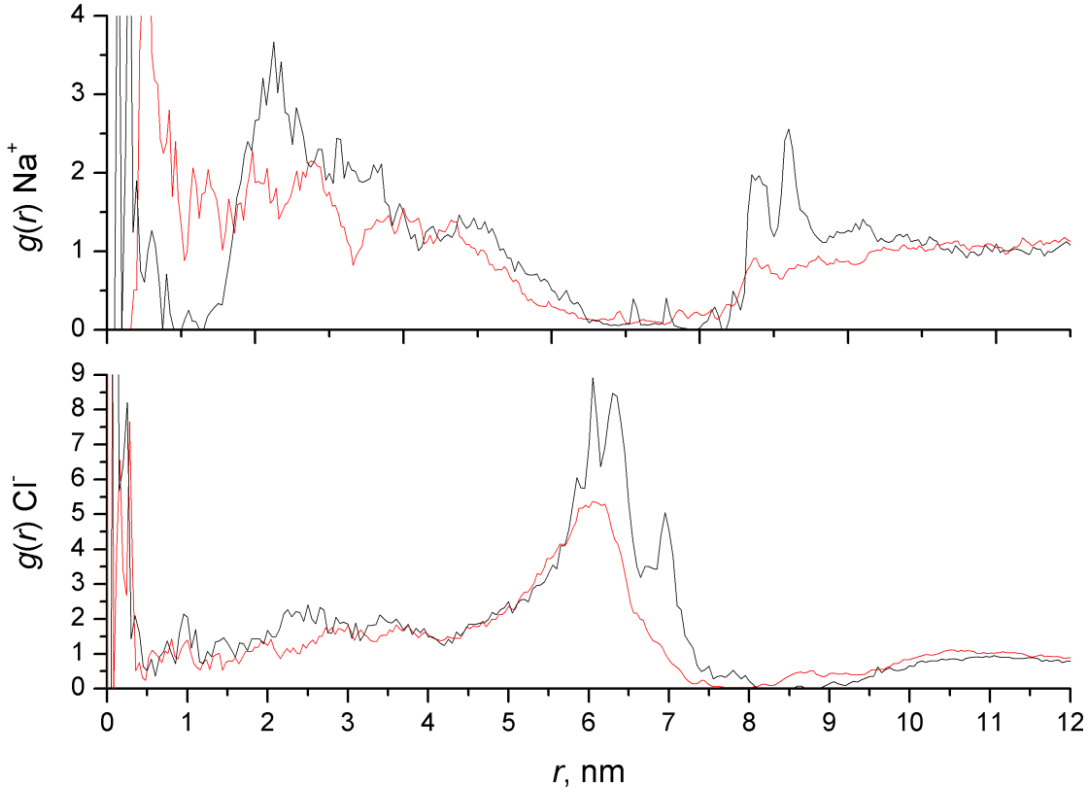


Рисунок 3.18 — Функции радиального распределения ионов Na^+ (вверху) и Cl^- (внизу) относительно центра масс капсида; черные линии соответствуют стабильному капсиду, красные линии соответствуют деформированному капсиду.

может быть следствием разницы в общем количестве Na^+ в объеме раствора (в ЕС оно на 226 больше, чем в ИС) и их первоначальном распределении, но можно предположить, что это не так. Общее количество Na^+ снаружи велико в обоих случаях (~1600 в ИС), этого должно быть достаточно для формирования слоя, если это необходимо. Более того, при необходимости в ИС должен был появиться хотя бы небольшой тонкий слой, но никаких доказательств этому нет.

RDF количественно определяют распределение ионов между сферическими слоями, но не показывают, как ионы распределяются внутри каждого слоя, в частности в важных слоях, прилегающих к внутренней поверхности капсида. Для анализа деталей накопления ионов в определенных местах внутренней поверхности для каждой аминокислоты было рассчитано среднее количество остатков этой аминокислоты, имеющих по крайней мере один ион Cl^- поблизости (на расстоянии 5A). Очевидно, это число изменяется от 0 (например, аминокислота заряжена отрицательно и отталкивает Cl^-) до 60 (этот остаток в каждом из 60 белков контактирует с ионом Cl^-). Распределения для обеих систем (естественной и искусственной) показаны на рисунке 3.19, и в приложении Е, Рисунок Е.1.

Для стабильного капсида несколько аминокислот имеют поразительно высокие значения этой занятости Cl^-: для остатков 172(Asn), 174(Lys) и 175(Arg) почти все белки имеют Cl^- вокруг этой аминокислоты ($\geqslant$ 55 из 60). Далее следуют остатки 94(Arg), 95(Lys), 173(Asn) и 176(Asn) со

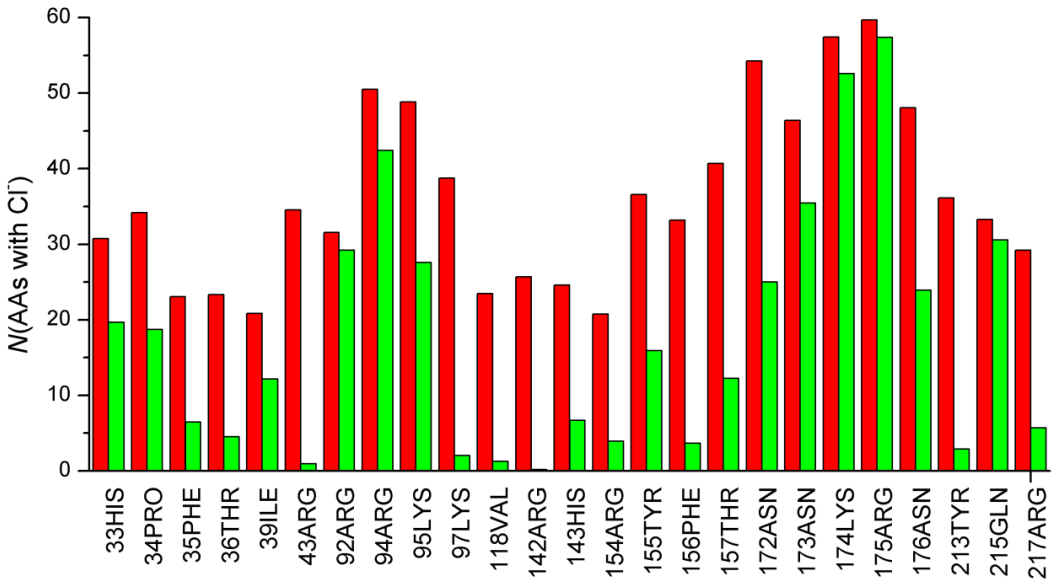


Рисунок 3.19 — Количество остатков, имеющих ион Cl^- ближе, чем 5A; показаны только аминокислоты, имеющие более 20 остатков, контактирующих с Cl^- в стабильном капсиде; красный — естественная система, зеленый — искусственная система.

значениями $\geqslant$ 47 из 60. Важно отметить, что количество 174(Lys) и 175(Arg) практически одинаковы в системах EC и ИС. Эти наблюдения свидетельствуют о том, что описанные регионы чрезвычайно привлекательны для анионов. Поэтому они могут легко служить местом прикрепления вирусного генома, упакованного внутри капсида.

Белки капсида образуют тримеры как геометрические единицы, из них построен капсид. Один из таких тримеров показан на рисунке 3.20. Интересно, что аминокислоты, богатые Cl^- из трех белковых мономеров, образующих тример, расположены очень близко друг к другу, несмотря на высокий положительный заряд этих аминокислот. Они не распределены равномерно по внутренней поверхности капсида, как можно было бы ожидать, чтобы минимизировать взаимное электростатическое отталкивание. Таким образом, наличие отрицательного заряда вблизи этой группы остатков, вероятно, имеет решающее значение при образовании тримера. Это позволяет предположить, что образованию тримеров из мономеров существенно способствовало бы присутствие в растворе ДНК, выступающей в качестве носителя, обеспечивая необходимый отрицательный заряд. Эти богатые хлоридом области внутренней поверхности показаны на рисунке 3.20.

Еще одно наблюдение из рисунка 3.19 заключается в том, что в ИС имеется ряд остатков, которые теряют значительно больше ионов хлора, чем другие, по сравнению с EC. Это остатки 43(Arg), 97(Lys), 118(Val), 142(Arg), 156(Phe), 213(Tyr) которые также имеют высокую занятость в EC, но теряют почти все ионы хлора при переходе от стабильного капсида к нестабильному. Они разделены на две группы (выделены зеленым и синим цветом на рисунке 3.20) и расположены на «ободке» полостей, содержащих поры.

Трудно выделить одну конкретную область капсида, где начинается деформация и которая вызывает его коллапс. По-видимому, большая часть оболочки, состоящая из нескольких белков, в какой-то момент складывается внутрь. Однако бывают случаи, когда рядом с порой образует-

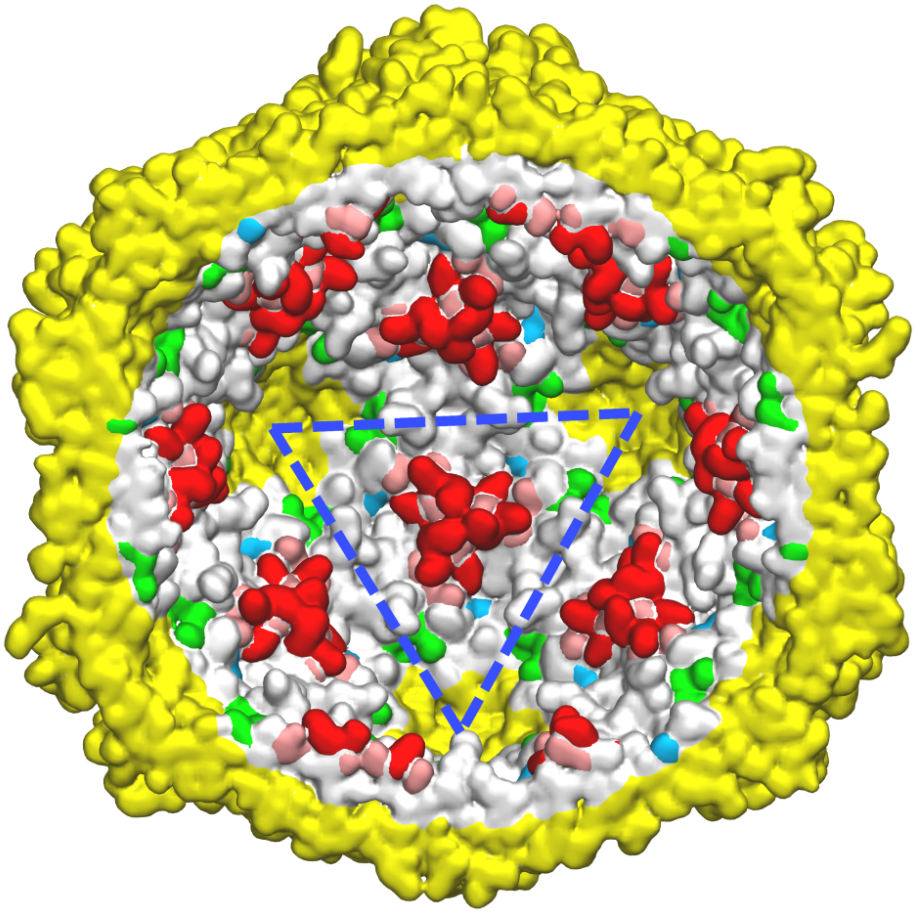


Рисунок 3.20 — Половина капсида, показывающая расположение богатых хлоридом аминокислот 172(Asn), 174(Lys), 175(Arg) (красные) и 94(Arg), 95(Lys), 173(Asn), 176 (Asn) (розовые) и теряющих хлорид аминокислот 43(Arg), 97(Lys), 142(Arg), 213(Tyr) (зеленые) и 118(Val), 156(Phe) (синие); остатки на расстоянии более 75A от центра масс капсида показаны желтым цветом; один тример обозначен пунктирной линией.

ся заметная трещина. Было проанализировано относительное расположение трещины и остатков, которые были определены как наиболее пострадавшие в нестабильной системе. Результат показан на рисунке 3.21. Аминокислоты 43(Arg), 97(Lys), 142(Arg), 213(Tyr) (выделены зеленым на рисунке 3.20) расположены в области трещины, то есть отсутствие Cl^- вблизи них может быть критическим фактором, инициирующим дестабилизацию всего капсида.

Наконец, была рассчитана карта распределения электростатического потенциала внутри недеформированного капсида с помощью программы APBS [201]. Были использованы настройки по умолчанию, подвижные ионы не добавлялись. Молекулы и ионы воды были исключены из расчетов, поскольку их присутствие экранировало бы потенциал капсида. Полученная карта представлена на рисунке 3.22 в нескольких вариантах. Поскольку голый капсид заряжен сильно положительно, значения потенциала варьируются от -2 В до +25 В. Карта была построена с использованием нескольких масштабов, и только масштаб, который был очень сильно смещен к верхнему пределу, позволил примерно распознать места, богатые хлоридами. Были проанализированы две поверхности: первая имеет толщину по умолчанию (определяется атомными радиусами атомов белка), а вторая находится на расстоянии примерно 5 А от атомов и, таким

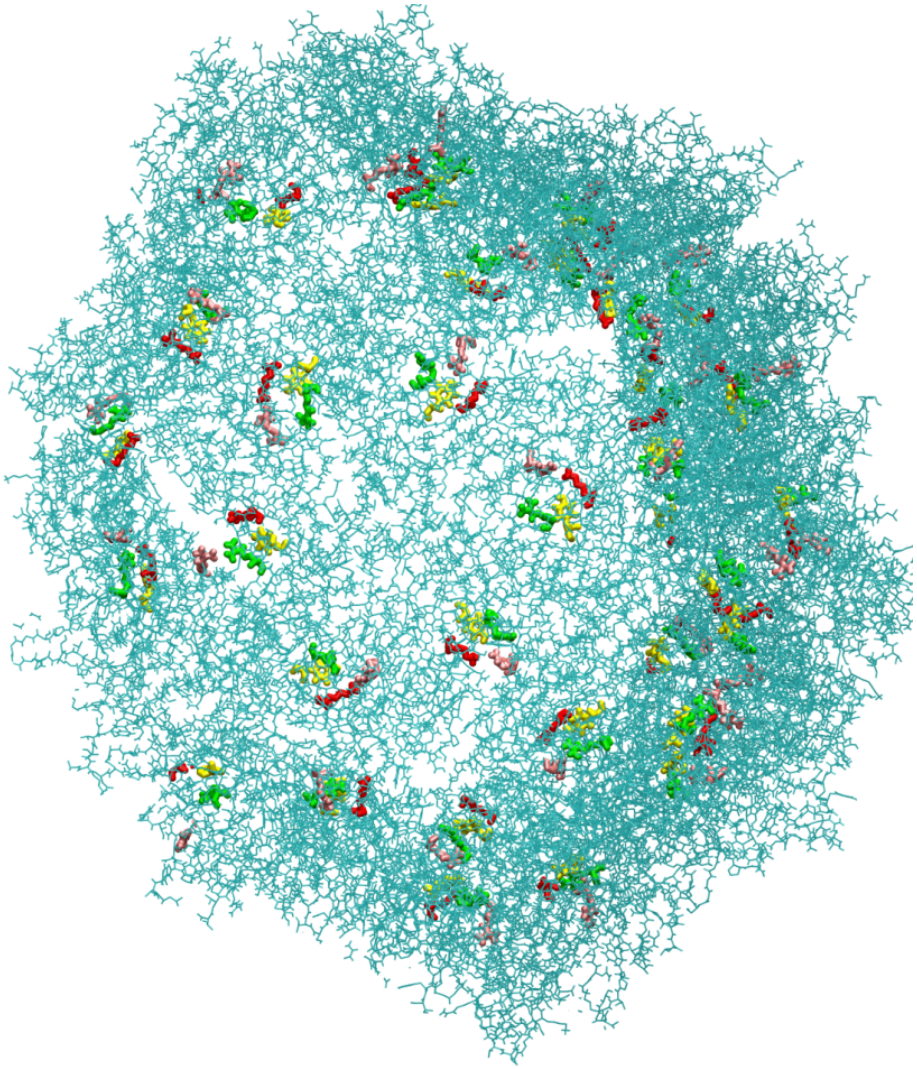


Рисунок 3.21 — Половина деформирующегося капсида после 10 нс моделирования MD; аминокислоты 43(Arg), 97(Lys), 142(Arg) и 213(Tyr) выделены зеленым, красным, розовым и желтым цветами соответственно.

образом, примерно соответствует поверхности, на которой расположены ионы. Незначительное изменение электростатического потенциала означает, что наблюдаемое распределение ионов трудно объяснить только зарядами остатков. Играет роль сложное взаимодействие между притяжением зарядов аминокислот и стерическими эффектами, включающими воду и противоионы.

3.4.3 Выводы

Распределение ионов хлора может имитировать локализацию фосфатных групп ДНК/РНК, как это было показано в [186] для другого вируса. Это также может иметь биологические последствия, поскольку стабильность капсида может определяться определенным расположением цепи ДНК.

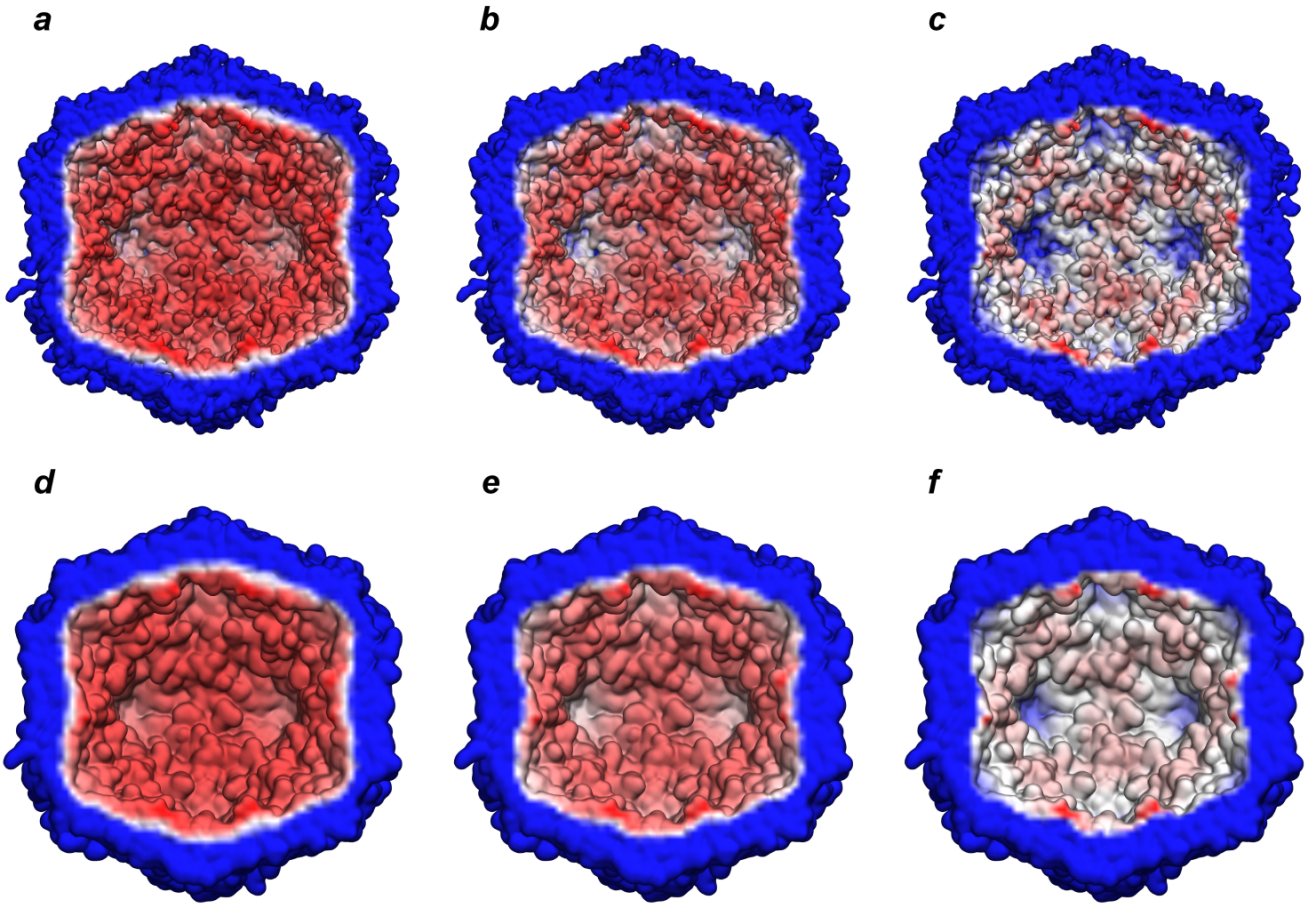


Рисунок 3.22 — Карты электростатического потенциала внутренней части капсида; верхний ряд: поверхность капсида, нижний ряд: поверхность приблизительно на расстоянии 5 А от капсида; шкала варьируется в пределах 13,7 – 23 В (a, d); 17 – 23 В (b, e); 20 – 23 В (c, f), где синий цвет соответствует нижнему пределу, а красный – верхнему пределу шкалы.

Было проведено сравнение двух систем, отличающихся только количеством ионов внутри капсида, и детальный анализ локального распределения ионов на внутренней поверхности стенки капсида с целью выявления того, как распределение ионов влияет на стабильность капсида.

Было обнаружено несколько аминокислот, которые очень привлекательны для ионов хлора в капсиде, и они одинаковы для обеих систем. Интересно, что распределение электростатического потенциала, создаваемого атомами капсида по внутренней поверхности капсида, явно не указывает на предпочтительные места связывания ионов Cl^-.

Были обнаружены аминокислоты, которые заняты хлорид-ионами в ЕС, однако в ИС они содержат гораздо меньше (практически ноль) хлорид-ионов. Поскольку эти аминокислоты располагаются в трещинах ИС, можно предположить, что дефицит ионов хлора вблизи этих аминокислот может вызвать появление трещин в капсиде и дальнейшую деформацию структуры капсида, что может привести к его разрушению. Поэтому можно выдвинуть гипотезу, что эти аминокислоты должны взаимодействовать с отрицательным зарядом хлорид-ионов или фосфатных групп генома.

3.5 Молекулярная динамика капсидного белка PCV2 при постоянном значении pH раскрывает механизм сборки капсида

Свиные цирковирусы (PCV) типов 1-3 представляют собой икосаэдрические вирионы размером 20 нм, симметрии T=1, которые имеют кольцевые одноцепочечные ДНК-геномы и принадлежат к роду *Circovirus* , семье *Circoviridae* [202; 203]. Эти вирусы морфологически схожи, но генетически, антигенно и патогенно различны. PCV1 является апатогенным и впервые был обнаружен в клетках почек свиньи (PK-15). PCV2 является высокопатогенным, поражает практически все органы хозяина и вызывает заболевание, ассоциированное с цирковирусом свиней (PCVAD), или заболевание, вызванное цирковирусом свиней (PCVD), которое приводит к иммуносупрессии хозяина и смерти от вторичных инфекций [204—208]. Естественными обьектами заражения животных PCV2 в дикой природе являются грызуны, телята, лисы, норки, козы, енотовидные собаки, кабаны и свиньи [209—214]. Контролируемые инфекции PCV2 в лабораторных условиях включают мышей BALB/c, мышей Kumming и несколько линий человеческих клеток [215—219]. Инфекции у телят и грызунов приводят к иммунодепрессии хозяина [197; 209; 210; 213; 214]. Недавно было показано, что PCV3 ассоциируется с многочисленными заболеваниями у свиней, включая репродуктивную недостаточность, дерматит и синдром нефропатии [220—222]. Свинофермы — это многомиллиардные отрасли в странах, где разводят свиней, и вспышки PCV2 могут иметь разрушительные экономические последствия [217]. Способность PCV2 преодолевать видовой барьер, индуцировать свой патогенный фенотип и инфицировать клетки человека в культуре позволяет предположить, что этот патоген находится на грани превращения в зоонозный и потенциально представляет угрозу для здоровья человека. Кроме того, PCV2 является сдерживающим фактором для ксенотрансплантации, поскольку свиньи считаются жизнеспособным источником органов для людей [223—225]. Несмотря на то, что программа вакцинации против PCV2 уже разработана, высокая частота мутаций PCV2 ($1{,}2 \times 10^{-6}$ замен положения$^{-1}$ год$^{-1}$) и рекомбинации являются причиной того, что недавние генотипы демонстрируют устойчивость к вакцинации [226; 227]. Следовательно, необходимы альтернативные методы лечения, которые нарушают одну или несколько стадий жизненного цикла PCV2. Перспективной стратегией является нарушение самосборки вирусов внутри клетки; однако для этого необходимо знание критических факторов, необходимых для успешной сборки инфекционного вириона. Исследования, описывающие сборку капсида PCV2 немногочисленны [197; 228; 229]. Недавние исследования показали, что сборка капсида происходит в ядре клеток, экспрессирующих капсидный белок (CP) [230]. Знание механизма сборки PCV2 расширит наши знания о сборке капсида вируса и откроет путь к разработке терапевтических средств подавляющих самосборку PCV2.

Белок PCV2 CP (27.8 кДа) имеет N-конец из сорока двух аминокислот, из которых почти половина — аргинины. Далее будем называть этот N-конец мотивом богатым аргинином (ARM). ARM обладают множественными сигналами ядерной локализации (NLS), которые отвечают за транспортировку вируса в ядро во время процесса проникновения и/или упаковки генома [231]. Другие функции, связанные с ARM, включают разрушение липосомальной мембраны для про-

никновения в клетку и взаимодействие с отрицательно заряженными фосфатами генома [232]. Экспрессия CP в клетках *Trichoplusia ni* (насекомое) или HEK293 (человек) приводит к образованию нуклеиновой кислоты, содержащей капсиды внутри клеток [74; 230; 233]. Структуры капсида в этих системах экспрессии практически неотличимы и демонстрируют, что CP принимает каноническую форму вирусного желеобразного рулета, распространенную во многих вирусных капсидных белках. Однако ARM не были промоделированы ни в одном из этих исследований [74; 230; 234—236]. Вероятнее всего, это связано с тем, что ARM либо не принимают икосаэдрическую симметрию, налагаемую при определении структуры, либо ARM неструктурированы.

Капсиды PCV2 собираются внутри клетки, чтобы заключить в себя нуклеиновую кислоту; однако варианты PCV2, разработанные для устранения этого взаимодействия, не способны собираться в капсиды внутри клетки [230]. Другие авторы показали, что капсиды могут собираться in vitro, заключая в себе нуклеиновую кислоту, что предполагает, что взаимодействие между CP и нуклеиновой кислотой является необходимым условием для сборки капсида [237]. Нам удалось разработать протокол in vitro для сборки капсидов PCV2 в отсутствии нуклеиновой кислоты. Это дает нам уникальную возможность изучать сборку капсида в отсутствие нуклеиновой кислоты. Объединив наши биохимические исследования с исследованиями *in silico*, было показано, что ARM не является необходимым для сборки капсида *in vitro* и протонирование CP отвечает за регуляцию сборки капсида. В частности, CP собираются в капсиды при pH 5, а не при pH 3 или 7. Это говорит о том, что pH 5 каким-то образом имитирует клеточные условия, необходимые для самосборки капсида, и CP при этом pH, вероятно, будет напоминать структуру, способную к сборке до сборки клеточного капсида. Одновременно с этим, подход *in silico* демонстрирует структурные изменения при pH 3 и 7, ответственные за ингибирование сборки. Подчеркнем, что определение атомной структуры мономерного CP при различных значениях pH с использованием экспериментальных подходов не является тривиальной задачей. Было показано, что моделирование методом полноатомной молекулярной динамики всего капсида PCV2 точно отражает его структуру и динамику и, таким образом, может использоваться в качестве надежного источника информации, дополняющего и объясняющего экспериментальные данные [7; 22; 25; 27].

В этом разделе была использована MD для изучения конформационного ландшафта CP при различных значениях pH. Использование традиционной MD для изучения влияния pH на структуру проблематично, поскольку этот подход сохраняет единственное состояние протонирования для каждого остатка на протяжении всего моделирования. Таким образом, теряется зависимость между изменениями в состояниях протонирования и изменениями в структуре, и отбираются потенциально неточные конформации. В вычислительной части этой работы интерпретированы биохимические результаты, чтобы определить, что происходит с мономерной структурой CP при этих значениях pH, используя недавно разработанное моделирование молекулярной динамики при постоянном pH (CpHMD) с методами обмена репликами (RE) в pH-измерении (CpH REMD) [238].

В отличие от классического подхода MD, моделирование CpHMD позволяет изменять состояния протонирования аминокислот в процессе моделирования. Метод RE обеспечивает улучшенную выборку. Сборка капсида происходит в результате сборки мономерных CP в более крупные олигомеры; таким образом, для понимания сборки капсида необходимы атомистические

знания о мономере CP. Были изучены структурные особенности CP, зависящие от pH, с помощью CpH REMD.

Было показано, что CpH REMD обеспечивает интерпретацию наблюдаемых явлений обусловленных pH. Конформация CP, подходящая для сборки капсида, зависит от положения петли GH, расположенной на трех осях икосаэдра симметрии (*iaos*), и это положение регулируется процессом протонирования/депротонирования с участием двух пар кислых аминокислот, которые мы называем диадами. Наши результаты также могут объяснить регуляцию самосборки PCV2 внутри клетки. Недавно было показано, что петля GH взаимодействует с нуклеиновой кислотой в оси третьего порядка нуклеиновой кислоты, содержащей VLP [230]. Можно предположить, что сборка цитоплазматического капсида запрещена электростатическим отталкиванием этих диад, и что взаимодействие петли GH с нуклеиновой кислотой внутри ядра может быть движущим фактором преодоления этого отталкивания. Насколько нам известно, это первое описание того, как PCV2 регулирует сборку своего капсида.

3.5.1 Экспериментальная часть

Очистка капсидного белка PCV2

Подготовка белка CP$*$ (рис 3.23) была проведена как описано ранее [74]. CP, в котором отсутствует ARM (CP$^{\Delta ARM}$) был создан путем генетического слияния небольшого модификатора юбиквитина His6 (SUMO) на N-конце CP (аминокислоты 42–233) (плазмида pET28a, Novagen) с использованием сервисов GenScript. Оба белка были синтезированны в *Escherichia coli*, штамм Rosetta DE3 (Novagen). Клетки выращивали в среде Terrific Broth (TB) при температуре 37°C до середины логарифмической фазы в присутствии соответствующего антибиотика. Экспрессия белка была индуцирована 0,25 мМ изопропилового спирта β-D-1-тиогалактопиранозид (ИПТГ) в течение 16 ч при 20°C. Клетки собирали центрифугированием (30 мин, 4000 об/мин, 4°C). Осадок лизировали в 50 мМ N-циклогексил-3-аминопропансульфоновой кислоте (CAPS) pH 10,5, 50 мМ хлориде натрия (NaCl), 0,2 мМ фенилметансульфонилфториде (PMSF), 0,5 мМ трис (2-карбоксиэтил)фосфине (TCEP) и 25 единицах бензоназы (MilliporeSigma). Лизат центрифугировали (30 мин, 16 000 об/мин, 4 °C) и супернатант подвергали гравитационной хроматографии Ni-NTA. Белок элюировали лизирующим буфером, дополненным 1 М имидазолом при pH 10.5. Элюат обрабатывали 25 единицами бензоназы при диализе в лизирующем буфере в течение ночи. Белок был дополнительно очищен с помощью тандемной катионно-анионообменной хроматографии (колонки HiTrap SP HP и HiTrap Q FF, GE healthcare). Белок связывался с колонкой HiTrapQ FF и элюировался примерно 200 мМ NaCl. К элюенту добавляли 250 мМ L-Arg и концентрировали белок до 20 мг/мл$^{-1}$. Концентрацию белка определяли методом Гилла и фон Хиппеля, аликвотируя по 50µl, мгновенно замораживали в $N_2(l)$, и хранили при температуре -80°C.

SUMO-специфическая экспрессия и очистка протеазы

Лаборатория доктора Патрика Дж. Лолла предоставила конструкт экспрессии для SUMO-специфической протеазы (dtuD1). Белок был экспрессирован и очищен в соответствии с протоколом опубликованном в [239].

Сборка капсида in vitro

CP$^{\Delta ARM}$ белок инкубировали с dtuD1 для высвобождения свободного CP$^{\Delta ARM}$. Состав полученного материала был подтвержден с помощью анализа SDS-PAGE. Для определения условий сборки капсида 3 мг/мл капсидного белка смешивали с различными буферами сборки в соотношении 1:1 и инкубировали при 4 °C в течение 48 часов. Первая группа буфера для сборки включала 12% полиэтиленгликоля (ПЭГ) 3350, 5% изопропанола, 0.6 M цитрата аммиака (pH 3.0, pH 4.0, pH 4.5, pH 5.0, pH 6.0, pH 7.0 и pH 8.0). Вторая группа буфера для сборки включала 0.6 M цитрата аммиака (pH 5.0), 5% изопропанола, 3%, 6%, 9%, 12% и 20% ПЭГ 3350.

Негативно окрашенные сетки были приготовлены путем нанесения 5 µл образца при концентрации 0,1 мг мл$^{-1}$ на сетку с углеродным покрытием 400 меш в течение 3 секунд. Образец был извлечен и окрашен 2% уранилформиатом (УФ). Для каждого условия были подготовлены и отображены три эквивалентные сетки. Для получения изображений образца с увеличением 60 000 раз использовался электронный микроскоп JEOL 1230 в Нью-Йоркском центре структурной биологии (NYSBC). Визуализация проводилась с использованием пакета Leginon [240].

Моделирование молекулярной динамикой

Все моделирования проводились с использованием пакета AMBER 18 [241]. Начальная структура для моделирования была получена из кристаллической структуры CP$*$ субъединицы капсида PCV2 (запись PDB 3R0R) [74]. Неразрешенные первые 32 аминокислоты структуры были реконструированы с помощью топологического моделирования в MODELLER [104] и добавлены в структуру CP. Первоначальная сольватированная система была подготовлена с помощью модуля Leap AmberTools. Белок был сольватирован с использованием модели воды TIP3P в кубической ячейке размером, обеспечивающим 10 Å между белком и стенками. Для нейтрализации системы и имитации ионной силы 0.15 M было добавлено 128 ионов Cl^- и 115 ионов Na^+. Общее число атомов в системе составило 129762. Параметры силового поля AMBER FF10 [241] был выбраны для белка. Алгоритм SHAKE был применен к связям, включающим водородные связи [242]. Дальнодействующие кулоновские взаимодействия были описаны с помощью метода сетки ча-

стиц Эвальда (PME) [59] и расстояние отсечки было установлено равным 8 Å. Взаимодействия Ван-дер-Ваальса рассчитывались с использованием отсечки в 8 Å.

Минимизация энергии сольватированной системы выполнялась методом наискорейшего спуска с последующим применением алгоритма сопряженных градиентов для 5000 шагов. Затем система с минимизированной энергией нагревалась в течение 400 пс от 10 K до 300 K с использованием динамики Ланжевена (частота столкновений 5 $пс^{-1}$). Затем было проведено 4-нс моделирование системы в ансамбле NPT. Температура и давление поддерживались постоянными на уровне 300 K и 1 атм с помощью термостата Ланжевена и баростата Берендсена соответственно. Кроме того, было проведено дополнительное 4-нс моделирование в ансамбле NVT с термостатом Ланжевена, поддерживающим температуру 300 K.

Для моделирования MD при постоянном pH была использована дискретная модель состояния протонирования, предложенная Монганом и разработанная Суэйлсом для MD в атомистическом растворителе [238]. После каждых 200 фемтосекунд моделирования MD, выполненного в атомистическом растворителе, была предпринята попытка изменения протонирования для всех титруемых остатков в неявном растворителе [243], время релаксации растворителя было установлено на уровне 200 фс. При моделировании CpHMD были титрованы все остатки Asp, Glu и His (всего 32 ионизируемых остатка). Было выполнено 10 реплик в pH-измерении, охватывающем диапазон pH 3.0 - 7.5 с интервалом 0.5 единицы pH, каждая реплика была промоделирована в течение 1μc в ансамбле NVT, в результате чего получили в общей сложности 10μсимуляции. Попытки обмена между репликами были установлены каждые 200 фс моделирования. Моделирование CpHMD проводилось с использованием PMEMD с ускорением на GPU [244—248]. Графики обмена для каждой реплики представлены в приложении.

Анализ данных моделирования

Реконструкции траектории для анализа данных выполнялись для pH 3, 5 и 7 с использованием программы Cpptraj в AMBERTools. Структурный анализ при каждом значении pH проводился на структурах, извлеченных из минимумов свободной энергии при каждом значении pH. Для кластеризации извлеченных конформаций был применен иерархический агломеративный подход, все атомы против всех Cα, с отсечкой в 2 Å [249]. Обработка состояний протонирования проводилась с помощью cphstats в AMBER [244—248] и значения pKa рассчитывались с использованием уравнения Хендерсона–Хассельбаха на основе соотношения временных точек с протонированной и депротонированной аминокислотой с помощью собственной программы [243]. Cpptraj в AMBERTools использовался для расчета расстояния между указанными остатками центра масс, площади поверхности, доступной растворителю (SASA), радиуса инерции (Rg) и среднеквадратичного отклонения C_α [249]. Значения площади закрытой поверхности (BSA) были рассчитаны с использованием сервера PDBePISA [250]. Контурные диаграммы для карт плотности, гистограм-

мы расстояний и ленточные диаграммы были созданы в Origin 2018b (OriginLab), MOE (Chemical Computing Group, Inc) и UCSF Chimera [251].

3.5.2 Результаты

Сборка капсида in vitro обусловлена pH

PCV2 CP (GenBank: AAO39759.1) может быть экспрессирован и очищен из *E. coli* до однородности мономерных субъединиц [74; 232]. В этих образцах не содержатся нуклеиновые кислоты, что определяется по соотношению 260 нм/280 нм. Инкубация CP в условиях низкого pH в течение нескольких недель приводит к образованию капсидов; однако удаление или модификация ARM приводит к надежной и быстрой сборке капсида [232]. Это может быть связано с тем, что нуклеиновая кислота необходима для нейтрализации положительного заряда связанного с ARM. Поэтому было решено заменить и исключить ARM из наших исследований. Рассматриваются два варианта: 1) CP*, который заменяет ARM на аффинную метку из гексагистидина, за которой следует линкер из двадцати шести аминокислот (His^6-линкер); и 2) $CP^{\Delta ARM}$, который исключает первые сорок две аминокислоты (Рисунок 3.23). Буфер сборки капсида был определен путем оптимизации состава буфера, используемого для получения кристаллов капсидов PCV2 для рентгеновской кристаллографии [74]. Снижение pH буфера сборки с 8.0 до 3.0 приводит к образованию наибольшего количества капсидов при pH 5 (Рисунок 3.23b). Собранные капсиды не наблюдаются при значениях pH выше 6,0, несмотря на увеличение концентрации CP или PEG. Это говорит о том, что протонирование одного или нескольких титруемых остатков (Asp, Glu или His) может отвечать за регуляцию сборки капсида. Для дальнейшего изучения этой возможности была использована молекулярная динамика для изучения конформаций CP* при различных значениях pH. Отметим, что способность $CP^{\Delta ARM}$ собираться с профилем pH аналогичным CP*, говорит о том, что последовательность N-конца в CP* не влияет на сборку капсида (Рисунок 3.23).

Конформационная выборка CP с моделированием CpH-REMD

Чтобы понять связь между сборкой капсида и pH, были исследованы репрезентативные конформации CP* при pH 3, 5 и 7 используя моделирование CpHMD в сочетании с pH-REMD для улучшения конформационной выборки. Было использовано 10 реплик, каждая в 1μc, охватывая диапазон pH от 3.0 до 7.5 с шагом 0.5 единицы pH. В последовательности CP имеется 32 аминокислоты (Asp, Glu, His), которые титруются в этом диапазоне pH. Эти остатки были целью титрования во время моделирования MD.

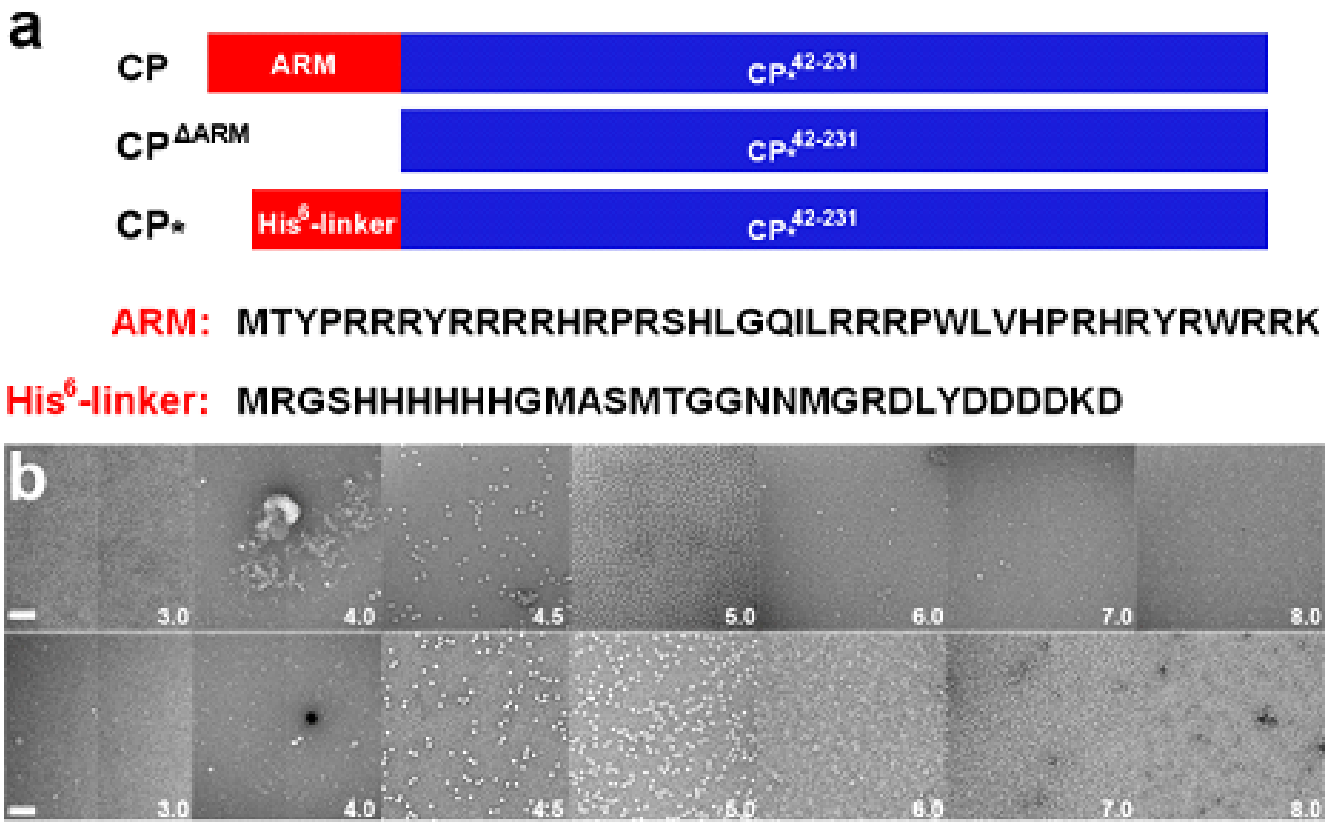


Рисунок 3.23 — Сборка капсида PCV2 in vitro; (а) схематические изображения полноразмерного CP, CP$^{\Delta ARM}$ и CP$*$ описывающий различия между белками. Аминокислотные последовательности N-концов приведены ниже; (б) микрофотографии, полученные с помощью трансмиссионного электронного микроскопа, отрицательно окрашенного in vitro собранного капсида PCV2 вверху, влияние pH (белый текст) на сборку CP$*$; внизу, влияние pH (белый текст) на сборку CP$^{\Delta ARM}$.

Координаты для N-конца не присутствуют ни в одной из экспериментальных структур PCV2 [74; 230; 232; 233; 237]; поэтому они были промоделированы с помощью MODELLER [104]. Сходимость моделирования контролировалась путем наблюдения за эволюцией поверхностей свободной энергии (FES) как функции SASA и RMSD как функции Rg (рисунки 3.1 — 3.11 в приложении 3). Сравнение Rg и SASA FES показывает, что CP принимает более компактную форму при pH 7 с меньшими Rg и SASA. RMSD рассчитывали на основе выравнивания каждой конформации со структурой капсидного CP (запись PDB 3R0R). Первоначальные выравнивания дали большие (>20 Å) Значения RMSD и Rg (<29 Å). По нашему предположению, гибкий N-конец может быть ответственен за большие значения Rg, SASA и RMSD. Исключение His^6-линкера значительно уменьшило значения RMSD (<7.5 Å) и Pr (<22 Å) (рисунки 3.4 — 3.6 в приложении 3). Анализ FES показал, что наше моделирование CpH-REMD сходится в пределах 1μc. Отметим, что поскольку наши биохимические эксперименты демонстрируют сопоставимые профили сборки для CP$*$ и CP$^{\Delta ARM}$ можно сделать вывод, что зависимые от pH конформационные изменения, регулирующие самосборку, должны включать сегмент CP$*^{42-231}$. Далее обсуждается последовательность без His^6-линкера (CP$*^{42-231}$).

Поверхности свободной энергии при разных значениях pH демонстрируют разные конформации CP

На рисунке 3.24 показываны конформационные ландшафты для каждого pH с помощью поверхностей свободной энергии в координатах Rg и RMSD. В течение 1 μc были идентифицированы густонаселенные области (синие), соответствующие минимальной энергии конформаций CP. Распределение этих областей уникально для каждого значения pH. Минимум наблюдается при значении Rg 19.7 Å и значении RMSD 4.6 Å для pH 3. Минимум для pH 5 наблюдается при Rg 20.0 Å и RMSD 4.0 Å. При pH 7 с Rg 20.0 наблюдаются два минимума при Rg 20.0 Å и RMSD 4.4 Å (минимум 1) и Rg 19.5 Å RMSD 5.0 Å (минимум 2).

Наличие двух минимумов предполагает наличие по крайней мере двух преобладающих конформаций при pH 7. Примечательно, что значение RMSD минимума при pH 5 ниже, чем при pH 3 или 7, что означает, что конформация $CP*^{42-231}$ при pH 5 может быть более похожа на конформацию, наблюдаемую в собранном капсиде.

pH изменяет конформацию петель CD и GH

Репрезентативные конформации CP для каждого pH были идентифицированы с использованием структурной кластеризации извлеченных структур из минимумов при каждом pH. Чтобы определить влияние pH на структуру CP были сравнены репрезентативные конформации каждого кластера (центроид) со структурой CP капсида (Рисунок 3.25a) [74; 236]. Четыре области значительно отклоняются от структуры CP капсида: N-конец (His^6-линкер), CD-петля (остатки 80-92), GH-петля (остатки 166-194) и C-конец (остатки 225-231) (Рисунок 3.25a,b и 3.26). На N-конце имеется тринадцать титруемых остатков (7 His и 6 Asp). Наши расчеты pKa показали, что остатки His и Asp на N-конце полностью и частично протонированы при pH 3 соответственно (Таблица 9). Увеличение pH от 3 приводит к частичной депротонизации остатков His и полной депротонизации остатков Asp при pH 5 и полной депротонизации обоих при pH 7. В собранном капсиде N-конец помогает определить интерфейс CP:CP в оси 2-го порядка (Рисунок 3.25a, 3.26a). При pH 3 и 5 N-конец расположен вблизи оси 2-го порядка (Рисунок 3.26b, c). С другой стороны, при pH 7 он располагается вблизи поверхности, определяющей внутреннюю поверхность капсида, из-за ионных взаимодействий между отрицательно заряженным Asp на N-конце и положительно заряженными Arg и Lys на внутренней поверхности капсида (Рисунок 3.26d, e).

В собранном капсиде петля CD контактирует с соответствующей петлей в соседней субъединице (Рисунок 3.25a). Наши результаты показывают, что петля CD принимает различные конформации при каждом значении pH (Рисунок 3.26); однако отсутствие титруемых остатков в петле предполагает, что ей может быть присуща наблюдаемая изменчивость. Биохимические исследования показали, что замена и усечение петли CD не отменяют сборку капсида, что сви-

детельствует о том, что петля CD, вероятно, не является решающим фактором в сборке капсида [228]. Положения GH-петли более интересны: петля занимает то же положение, что и в собранном капсиде при pH 5, однако петля находится ближе к EF-петле при pH 3 или закручивается и перемещается ближе к CD-петле при pH 7 (Рисунок 3.26f-j). Анализ экспериментальной структуры капсида показал, что на петлю GH приходится до трети (1000 Å^2) поверхности CP, которая погружается при сборке капсида [252]. Ниже рассматривается значение петли GH более подробно. В собранном капсиде C-конец также помогает определить интерфейс CP:CP в оси 2-го порядка. Наши результаты показывают, что C-конец занимает различные положения при каждом значении pH (Рисунок 3.26a-e и рисунки 3.16, 3.18 — 3.21 в приложении 3). На C-конце имеется единственный титруемый остаток (Asp228), и расчеты pKa показывают, что он становится депротонированным при значениях pH выше 5 (см. ниже). Чжан и др. недавно показали что только атомы основной цепи аминокислот 225-231 необходимы для сборки капсида, так как удаление, но не замена этого региона отменяет сборку капсида [229]. Вероятно, C-конец принимает необходимое положение и конформацию, когда две субъединицы взаимодействуют по оси 2-го порядка.

Две аспарагиновые диады, расположенные в петлях EF и GH, регулируют сборку капсида

В собранном капсиде два аминокислотных остатка в EF-петле (Asp126 и Asp127) и два в GH-петле (Asp168 и Asp172) расположены в непосредственной близости друг от друга (Рисунок 3.26f-j). Их состояние протонирования может быть ответственно за наблюдаемую конформационную изменчивость петли GH, поэтому они были титрованы в ходе моделирования. Чтобы количественно оценить положение петли GH относительно петли EF при трех значениях pH, были рассчитаны расстояния между аминокислотами этих двух петель для извлеченных конформаций (Рисунок 3.24). На рисунке 3.27 представлено распределение расстояний между центрами масс для остатков 168 и 126 (далее обозначаемых как 168:126) (Рисунок 3.27a, c, e, g) и остатками 172:126 (Рисунок 3.27b, d, f, h).

Было проанализировано наблюдаемое распределение расстояний при каждом значении pH с расстояниями в экспериментально определенной структуре капсида. При pH 3 мономодальное распределение между остатками 168:126 (10-11.2 Å) было похоже на расстояние, наблюдаемое в структуре капсида (10.7 Å); однако расстояния между остатками 172:126 были распределены мономодально (8.1-8.7 Å) и были значительно больше, чем у капсидной структуры (5.1 Å) (Рисунок 3.27a, b). При pH 5 мультимодальное распределение между остатками 168:126 (8.5-10.5 Å и 10.9-14.5 Å) и 172:126 (5.1-6.7 Å и 8.1-9.3 Å) было похоже на расстояния, наблюдаемые в структуре капсида (Рисунок 3.27c, d). При pH 7 минимум 1, мультимодальное распределение между остатками 168:126 (10.3-13.1 Å и 16.0-20.0 Å) было похоже на расстояние, наблюдаемое в структуре капсида; однако мультимодальное распределение между остатками 172:126 (8.2-14.4 Å и 18.7-21.3 Å) значительно больше, чем наблюдается в структуре капсида (Рисунок 3.27e, f). Эти

значения отражают наблюдаемое скручивание петли GH на Рисунок 3.27d, где Asp168 занимает то же положение, что и в собранном капсиде, тогда как Asp172 отталкивается от EF-петли. При pH 7 минимум 2 бимодальное распределение между аминокислотами 168:126 и 172:126 значительно больше, чем в структуре капсида (Рисунок 3.27g, h). Увеличение расстояния между этими диадами при pH 7 по сравнению с 5 показывает, что петля GH отдаляется от петли EF при нейтральном pH.

Наиболее важным фактором, ответственным за это движение, могут быть силы отталкивания между двумя депротонированными диадами. Чтобы рассмотреть эту возможность, были рассчитаны значения pKa титруемых аминокислот из конформаций, извлеченных из каждого минимума (Рисунок 3.24, Таблица 9). На основании рассчитанных значений pKa, показанных в таблице 9, Asp126 и Asp127 (петля EF) преимущественно протонированы при pH 3, частично депротонированы при pH 5 и преимущественно депротонированы при pH 7 для обоих минимумов. Asp168 (петля GH) преимущественно протонируется при pH 3 и депротонируется при pH 5 и pH 7 (для обоих минимумов). Asp172 протонируется при pH 3, протонируется при pH 5 и депротонируется при pH 7 (для обоих минимумов). Протонирование четырех аспарагиновых кислот (две диады) при pH 3 позволяет удерживать петли EF и GH в непосредственной близости (Рисунок 3.26g). При pH 5 Asp126, Asp127 и Asp168 депротонируются, а Asp172 протонируется. Депротонирование этих трех аспарагиновых кислот приводит к электростатическому отталкиванию, отталкивающему петлю GH от петли EF (Рисунок 3.26h) в положение, наблюдаемое в структуре капсида (Рисунок 3.26f). При pH 7 все четыре аспарагиновые кислоты депротонированы, и дополнительное электростатическое отталкивание отодвигает GH-петлю дальше от EF-петли (Рисунок 3.26i, j). Почти для всех представительных конформаций в двух минимумах при pH 7 наблюдалось скручивание петли GH (Рисунок 3.26i, j).

3.5.3 Обсуждение результатов

Способность вирионов регулировать сборку своих CP для создания инфекционного вириона может обеспечить им большую эволюционную приспособленность по сравнению с вирионами, CP которых неконтролируемо собираются в агрегаты, пустые или заполненные клеточной нуклеиновой кислотой капсиды внутри инфицированной клетки. Некоторые из этих регуляторных механизмов могут быть присущи вирусному CP или капсиду. Понимание механизмов, ответственных за эту регуляцию, необходимо для понимания жизненных циклов вирусов и разработки терапевтических средств, позволяющих нарушить формирование инфекционных вирионов. Было показано, что экспрессия CP PCV2 в клетках млекопитающих приводит к сборке капсида исключительно внутри ядра [230]. Это говорит о том, что PCV2 способен регулировать образование капсида, ингибируя сборку в цитоплазме и запуская ее в ядре, где происходит репликация генома PCV2.

Внутри клетки CP собирается при pH, близком к нейтральному, чтобы заключить в себя нуклеиновую кислоту. In vitro CP собирается при pH, близком к нейтральному, только в присутствии нуклеиновой кислоты; таким образом, взаимодействие между CP и нуклеиновой кислотой явля-

ется необходимым условием для сборки капсида [237]. Взаимодействие CP:нуклеиновая кислота, вероятно, обусловлено высоким содержанием Arg в ARM. Чтобы уменьшить взаимодействие CP:нуклеиновая кислота и сосредоточиться на взаимодействии CP:CP во время сборки капсида, было решено заменить и исключить ARM из наших исследований сборки. Как оказалось, CP PCV2, лишенный нуклеиновой кислоты, собирается in vitro только в кислых условиях (pH 5, Рисунок 3.23). Более ранние структурные исследования показали, что эти капсиды неотличимы от капсидов, собранных в клетках насекомых и млекопитающих [74; 230; 232; 236]. Следовательно, при отсутствии нуклеиновой кислоты pH 5 может выступать в качестве движущего фактора для сборки CP. Чтобы понять, как pH 5 влияет на сборку капсида, были промоделированы CP при различных значениях pH, используя моделирование MD при постоянном pH с техникой обмена репликами в pH-измерении. Моделирование CpH-REMD показало, что петля GH занимает положение, наблюдаемое в капсиде, только при pH 5 (Рисунок 3.26f, h). Расчет значений pKa с использованием конформаций CP, полученных с помощью CpH-REMD, продемонстрировал, что петля GH занимает правильное положение при pH 5, поскольку Asp126/Asp127 (петля EF) и Asp168 (петля GH) депротонированы, а Asp172 (петля GH) протонирован (Рисунок 3.26h). Однако эти аспарагиновые кислоты депротонируются при pH 7, и электростатическое отталкивание между ними отталкивает петлю GH от петли EF, что подавляет сборку капсида (Рисунок 3.26i, j). Выравнивание всех последовательностей CP PCV2 из GenBank показало, что по крайней мере три кислые аминокислоты (Asp/Glu) присутствуют в четырех позициях и что Asp172 абсолютно консервативен [230]. Это подтверждает нашу гипотезу о том, что PCV2 использует Asp172 для регуляции сборки капсида. В контексте инфекции электростатическое отталкивание, описанное в исследовании CpH-REMD, еще предстоит изучить, поскольку предполагается, что pH цитоплазмы и ядра близок к нейтральному.

Карта электростатического потенциала капсида PCV2 показала, что его внутренняя поверхность заряжена положительно [22]. Аминокислоты, ответственные за эти заряды, вероятно, взаимодействуют с отрицательно заряженными фосфатами нуклеиновой кислоты, упакованной в капсид [25]. Действительно, было показано, что замены аминокислот, которые мешают взаимодействию фосфата нуклеиновой кислоты с CP, нарушают сборку капсида [230]. Учитывая, что Khayat и др. продемонстрировали, что петля GH (Lys179) взаимодействует с фосфатами остова нуклеиновой кислоты и что эта петля важна для сборки капсида (как показали эти исследования), можно предположить, что взаимодействие между петлей GH и нуклеиновой кислотой может отвечать за правильное позиционирование петли GH для сборки капсида [230]. В то время как другие авторы продемонстрировали сборку капсида *E. coli* экспрессированный CP при pH 7, остается определить, содержал ли их материал нуклеиновую кислоту, которая могла бы облегчить процесс сборки [229; 253]. Наши результаты также показывают, что присутствие CP ARM не является необходимым для сборки капсида при pH 5 *in vitro*.

Сообщения о неудачах вакцины PCV2 и изменениях генотипа указывают на необходимость разработки противовирусных терапевтических средств [74; 197; 254]. Наши результаты показывают, что определенное положение петли GH важно для сборки капсида PCV2, и отклонение от этого положения может препятствовать сборке. Таким образом, соединения, способные контролировать положение петли GH, могут служить потенциальными кандидатами для ингиби-

рования сборки капсида PCV2. Можно предположить, что эту задачу могут выполнить два класса соединений. Один класс расположил бы петлю GH таким образом, что сборка стала бы невозможной. Альтернативный класс может способствовать формированию конформации CP, способной собираться в капсиды внутри цитоплазмы, где геном PCV2 отсутствует. Такие аггрегаты представляют собой неинфекционные частицы. Эти соединения могут быть эффективными ингибиторами, способствуя сборке капсида в отсутствие нуклеиновой кислоты. Таким образом, наблюдаемые конформации CP имеют физиологическое значение.

В заключение, наши результаты имеют фундаментальное и практическое значение, поскольку они обеспечивают понимание того, как вирус регулирует свою самосборку, и открывают новые возможности для разработки противовирусных терапевтических средств. Насколько нам известно, это первое объяснение механизма сборки PCV2.

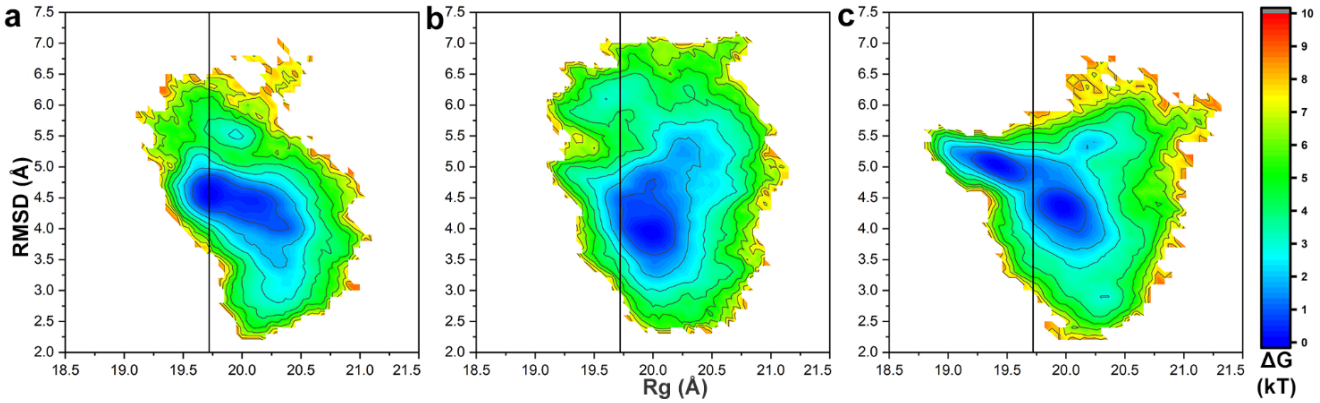


Рисунок 3.24 — Конформационный ландшафт, полученный с помощью капсидного белка CP$*^{42-231}$; поверхность свободной энергии как функция параметров RMSD (CP$*^{42-231}$ Cα атомов) как функция Rg (CP$*^{42-231}$Cα атомов) из моделирования при (a) pH 3, (b) pH 5, (c) pH 7; RMSD рассчитывали путем сравнения конформаций, полученных из 1µc моделирования к эталонному CP в собранном капсиде (запись PDB 3R0R); вертикальные черные линии обозначают значение Rg записи PDB 3R0R.

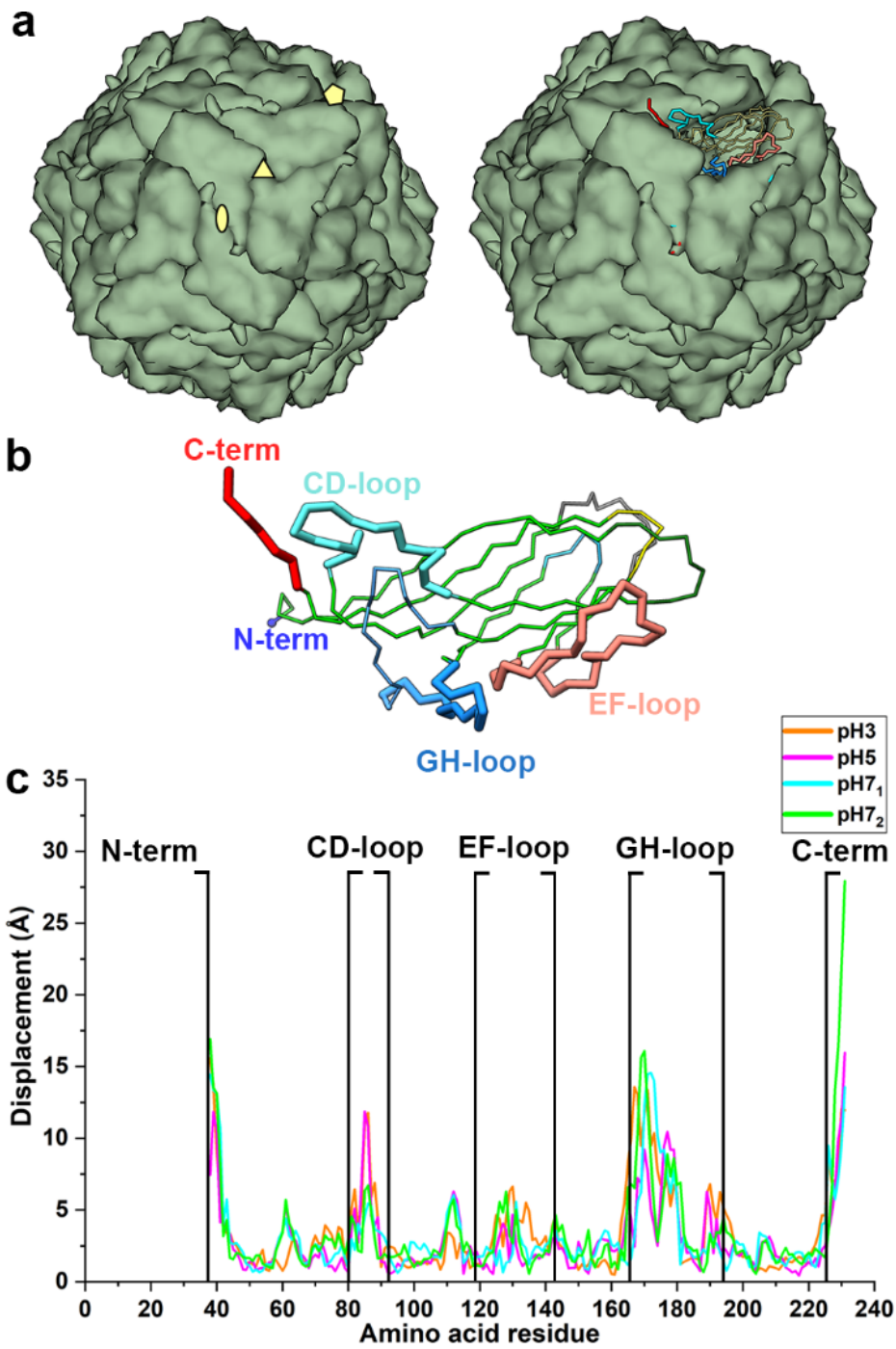


Рисунок 3.25 — Конформационная изменчивость CP при pH 3, 5 и 7; (a) слева: капсид PCV2 с CP, представленным в виде зеленых точек; оси симметрии икосаэдра показаны как пятиугольник (5-го порядка), треугольник (3-го порядка) и эллипс (2-го рядка); справа: одна из субъединиц показана как Cα представление трубки с важными петлями (N-конец, CD и GH в виде толстых трубок); (б) PCV2 CP Cα представление трубки с маркировкой важных петель (толстых трубок); (c) расстояния между эквивалентными Cα атомами репрезентативных конформаций при pH 3, 5, 7 после выравнивания со структурой капсида (запись PDB 3R0R); панели a и b были сделаны с использованием структуры капсида; в контексте собранного капсида N-конец находится внутри капсида около оси 2-го порядка: CD-петля расположена на поверхности капсида вблизи оси 2-го порядка, GH-петля скрыта около оси 3-го порядка, а C-конец расположен на поверхности капсида вблизи оси 2-го порядка.

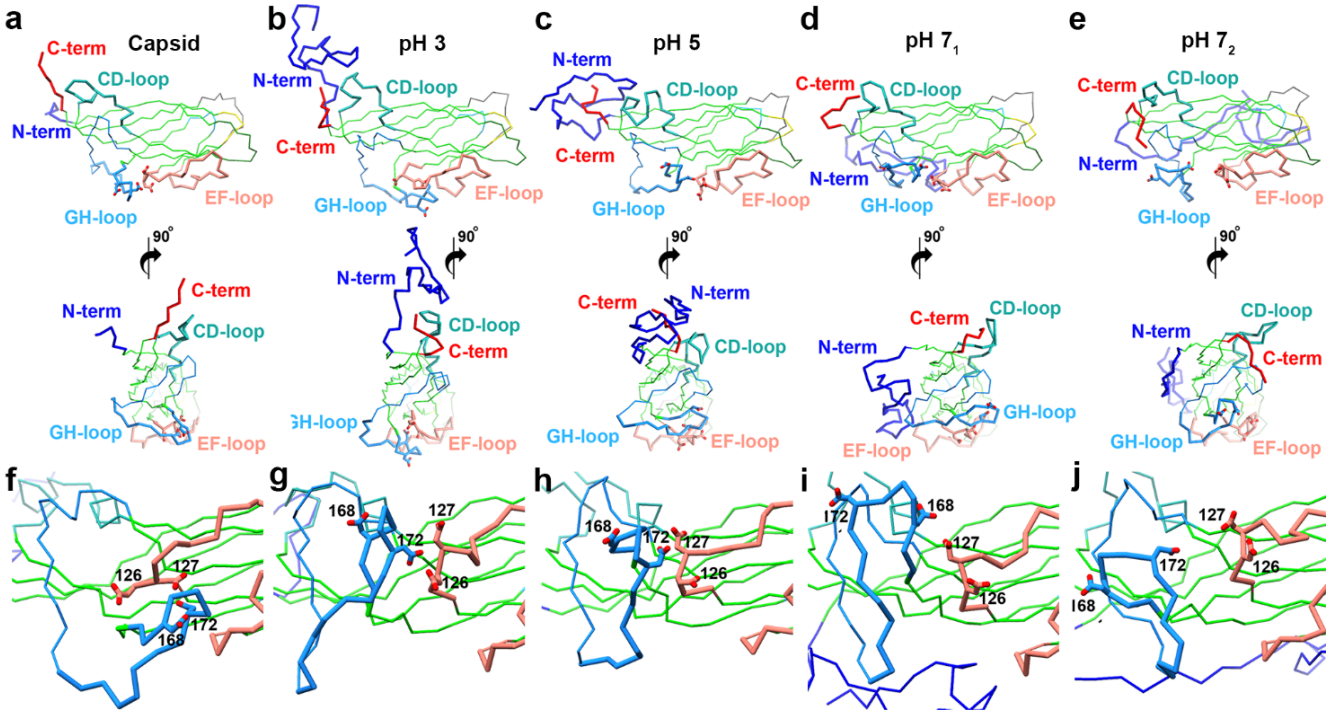


Рисунок 3.26 — Репрезентативные конформации из первого кластера при каждом значении pH; области со значительным отклонением от структуры капсида выделены в соответствии с рисунком 3.25; a) CP, извлеченный из собранных капсидов, b) представитель наиболее заселенного кластера при pH 3, c) представитель наиболее заселенного кластера при pH 5, d) представитель наиболее заселенного кластера из первого пика при pH 7, e) представитель наиболее заселенного кластера из второго пика при pH 7, f-j) крупный план EF- (розовый) и GH-петель (синий), показывающий их кислотные диады f) CP, извлеченный из собранных капсидов, g) представитель наиболее заселенного кластера при pH 3, h) представитель наиболее заселенного кластера при pH 5, i) представитель наиболее заселенного кластера из первого пика при pH 7, j) представитель наиболее заселенного кластера из второго пика при pH 7; значения RMSD для репрезентативных структур: (b)–4.6Å, (c)–3.9Å, (d)–4.3Å и (e)–5.1Å.

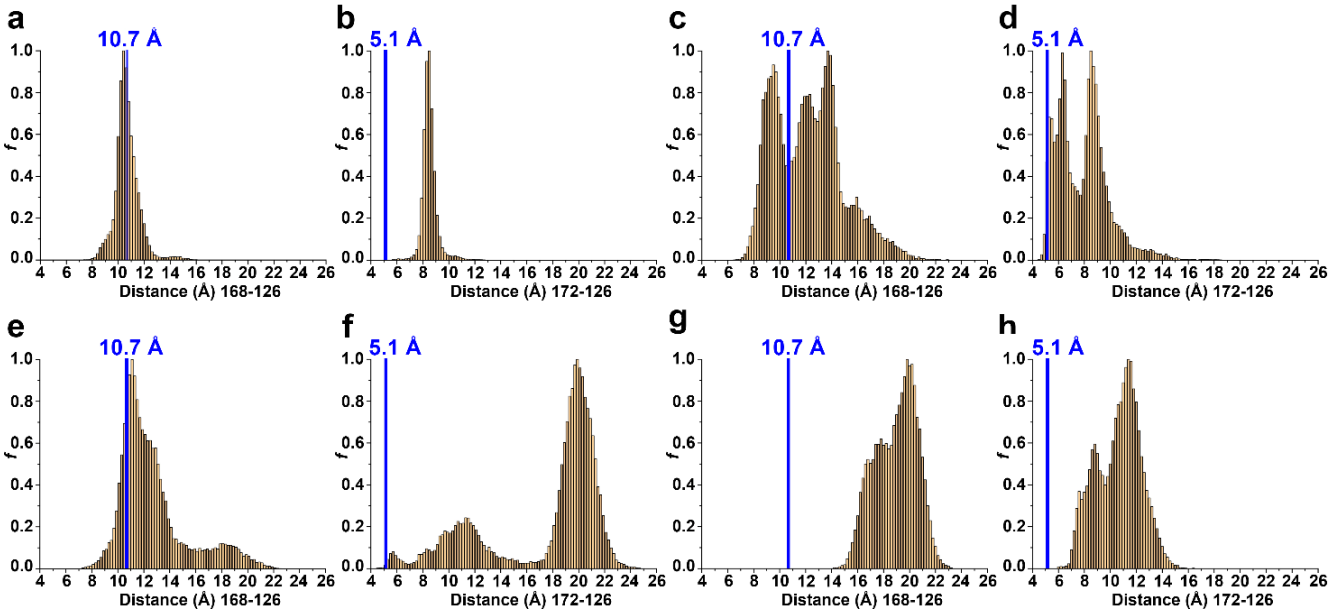


Рисунок 3.27 — Распределение расстояний между центрами масс для остатков 126 и 168 (a, c, e, g) и остатков 126 и 172 (b, d, f, h); при (a, b) pH 3; (c, d) pH5; (e, f) pH 7 (минимум 1); и (g, h) pH7 (минимум 2); синяя вертикальная линия на каждом графике представляет расстояние между этими остатками в структуре капсида (10.7 Å между 168 и 126; 5.1 Å между 172 и 126); распределение было рассчитано для всех структур, извлеченных из минимумов свободной энергии поверхности при каждом pH.

Таблица 9 — Рассчитанные значения pKa титруемых аминокислот по сравнению со значениями pKa которые были вычисленны с использованием конформаций, извлеченных из минимумов pH 3, 5 и 7. «p» и «d» обозначают протонированную и депротонированную фазы соответственно. Аминокислоты Asp126, Asp127, Asp168 и Asp172 выделены жирным шрифтом

	pH 3	**pH 5**	**pH 7_1**	**pH 7_2**			**pH 3**	**pH 5**	**pH 7_1**	**pH 7_2**
HIS 10	5.1 p	5.6 p	6.0 d	5.5 d		GLU 104	3.3 p	3.0 d	3.3 d	3.4 d
HIS 11	5.3 p	5.5 p	6.1 d	5.5 d		ASP 115	3.9 p	4.3 d	5.3 d	4.8 d
HIS 12	4.7 p	5.4 p	5.8 d	5.5 d		**ASP 126**	**3.5** p	**4.0** d	**4.6** d	**4.2** d
HIS 13	5.1 p	5.5 p	5.9 d	5.7 d		**ASP 127**	**3.6** p	**4.4** d	**5.0** d	**5.1** d
HIS 14	4.9 p	5.5 p	6.3 d	6.4 d		ASP 139	3.6 p	4.4 d	4.9 d	4.1 d
HIS 15	5.0 p	5.7 p	6.0 d	5.8 d		HIS 148	4.9 p	5.2 p	5.4 d	5.2 d
ASP 29	3.0	3.9 d	3.9 d	3.8 d		HIS 157	6.6 p	6.3 p	6.9 d	7.4 p
ASP 32	2.8 d	3.9 d	4.5 d	5.2 d		**ASP 168**	**3.7** p	**3.9** d	**4.6** d	**4.5** d
ASP 33	2.9 d	3.6 d	3.9 d	4.3 d		**ASP 172**	**6.1** p	**5.8** p	**5.6** d	**4.2** d
ASP 34	3.0	4.1 d	5.0 d	3.4 d		ASP 194	2.0 d	2.8 d	3.1 d	3.3 d
ASP 35	3.5 p	4.1 d	4.6 d	2.6 d		HIS 195	6.3 p	6.5 p	6.8 d	7.3 p
ASP 37	3.3 p	4.0 d	4.3 d	3.6 d		GLU 203	4.3 p	5.7 p	5.9 d	6.1 d
HIS 38	5.6 p	6.1 p	6.3 d	6.2 d		ASP 208	2.8 d	3.4 d	3.5 d	3.9 d
ASP 70	4.9 p	5.1 p	5.2 d	5.6 d		GLU 210	4.6 p	5.1 p	5.1 d	5.0 d
ASP 78	2.3 d	3.2 d	3.8 d	3.8 d		GLU 223	4.3 p	4.9 d	5.3 d	5.3 d
GLU 94	2.9 d	3.7 d	4.4 d	4.0 d		ASP 228	4.4 p	5.0	5.3 d	4.9 d

Глава 4. Белки и взаимодействие белок-лиганд

4.1 Немарковская модель динамики пептидов

Движение и функция белка определяются элементарными конформационными переходами, описываемыми изменениями двугранных углов остова биополимера. Обычно углы колеблются вокруг усредненного значения, приписываемого «метастабильным» конформационным состояниям, и иногда существенно изменяют своё значение, что соответствует конформационным переходам [31—33].

Динамика таких метастабильных состояний часто описывается с помощью все более популярной модели марковских состояний (MSM) [255—261]. Суть модели заключается в определении дискретных конформаций белка таким образом, чтобы переходы между ними в дискретные моменты времени представлялись марковским процессом. Поскольку математический аппарат марковских процессов хорошо развит, модель предоставляет полезную информацию о временной эволюции белковой системы.

Одной из активных тем в исследованиях MSM является валидация определения состояний. Не существует строгого описания того, как разделить непрерывное пространство двугранных углов белка на области, обозначающие марковские состояния. Важным требованием MSM является марковское свойство состояний, то есть независимость переходов от предыдущих шагов. Это накладывает концептуальное ограничение на временной масштаб модели: временной шаг должен быть не меньше памяти системы.

На практике состояния либо приписываются областям, примерно соответствующим областям притяжения на поверхности потенциальной или свободной энергии белка, либо строятся из «микросостояний». Последние определяются с помощью мелкой сетки, покрывающей все доступное двугранное пространство системы. Микросостояния группируются в «макросостояния» на основе времени, проведенного в них системой [262]. Макросостояния представляют собой совокупности микросостояний, для которых переходы внутри макросостояния более вероятны, чем между макросостояниями. Поскольку микросостояния должны быть «кратковременными», то есть переходов внутри микросостояния следует избегать, для корректной реконструкции макросостояний их требуется очень много.

Недостатков MSM можно избежать, если разрешить состояниям быть немарковскими. Естественным развитием теории марковских процессов является теория скрытых марковских процессов. Здесь переходы могут зависеть от предыдущих шагов, и эти зависящие от времени состояния группируются в последовательности состояний таким образом, что сами последовательности становятся марковскими.

В данном разделе развит вариант скрытого марковского описания конформационных переходов. Использована методология, называемая «Вычислительная механика» (КМ), разработанная Кратчфилдом *и др.* [130; 263; 264]. Показано, что применение этого подхода к динамике бел-

ков (которую далее называем немарковской моделью состояния, nMSM) позволяет обнаружить кратковременные «переходные» состояния, сократив временной шаг на порядок по сравнению с обычной MSM. Описывая переходы в таком мелком масштабе времени можно выяснить подробные механизмы переходов. В частности, становится возможным анализ элементарных конформационных переходов, являющихся строительными блоками всех конформационных движений белков. Вычислительная механика была использована в изучении молекулярных систем в нескольких контекстах [17; 20; 34]. В этом разделе, она применена для прямого описания конформационных движений пептида.

Поскольку механизмы переходов теперь могут быть раскрыты, продолжительность самих переходов (не время между переходами, а сам процесс конформационных перестроек) может быть количественно определена. Этого обычно невозможно достичь в MSM, поскольку время переходов обычно меньше, чем временной шаг MSM. Кроме того, можно идентифицировать и изучать «повторные пересечения» когда система демонстрирует несколько быстрых пересечений границы разделяющей конформационные состояния – важная проблема, давно признанная в этой области [35].

4.1.1 Молекулярная система

Был изучен цвиттер-ион L-аланил-L-аланин, Рисунок 4.4. Это очень удобная модель, потому что i) конформация молекулы полностью определяется двумя двугранными углами ψ и φ, ii) в воде конформация $\psi \approx 2.5$, $\varphi \approx -2{,}2$ радиан преобладает, однако существуют очень редкие переходы в две другие метастабильные конформации ($\psi \approx -1$, $\varphi \approx -2.2$ и $\psi \approx 2.5$, $\varphi \approx 1$) и iii) переходы происходят только в воде, поскольку в вакууме отрицательно заряженная группа COO^- сильно взаимодействует с положительно заряженной группой NH_3^+, исключая все конформации, кроме той, в которой группы находятся на минимальном расстоянии друг от друга. Таким образом, молекула является хорошей моделью конформационных изменений белка и в то же время технически удобна: переходы четко определены и хорошо разделены во времени.

Моделирование проводилось с использованием программного пакета GROMACS [126]. Пептид был сольватирован 396 молекулами воды. Размеры ячейки составляли 25x22x22Å. Силовое поле было либо Gromos 53a6 [116; 265; 266] (Рисунок 4.4) либо OPLS (оптимизированные потенциалы для моделирования жидкостей) (Рисунок 4.5) [267]. Поля оптимизированы для биомолекулярных систем, взаимодействующих с водой. Электростатические взаимодействия обрабатывались методом суммирования Particle Mesh Ewald (PME). Использовались периодические граничные условия. Температура поддерживалась на уровне 300K с помощью термостата Берендсена [268]. Алгоритм интегрирования был типа Верле, а шаг интегрирования составлял 0.002 пс. Перед отбором траекторий для анализа система была тщательно уравновешена. Общее время моделирования составило $2\mu s$ для каждой системы.

4.1.2 Методология моделирования

Три метастабильных состояния, отчетливо видимые на плотности состояний, Рисунок 4.4, позволяют ввести простую естественную дискретизацию конформационных состояний. Также дискретизируя время с шагом Δt непрерывная траектория MD может быть преобразована в строку символов $\{s_i\}, i = 0 \dots N$, где s_i равно «A», «B» или «C» в зависимости от того, куда попадает точка траектории в момент времени t_i, N — количество временных шагов.

Анализ $\{s_i\}$ можно сделать, построив на их основе MSM или, как в этом разделе, применив скрытую марковскую модель.

Вычислительная механика строит статистику на основе бесконечно длинных историй («прошлого») символов s_i, представляющих состояние системы в моменты времени t_i, $\overleftarrow{s}_i \equiv \{\dots s_{i-2}s_{i-1}s_i\}$, анализируя «будущие» $\overrightarrow{s}_i \equiv \{s_{i+1}s_{i+2}\dots\}$ после каждого прошлого. Фокус смещается с физических конформационных состояний s_i на их последовательности $\overleftarrow{s}_i$ и $\overrightarrow{s}_i$. Таким образом, в картину включаются естественные взаимозависимости конформационных состояний, при этом сохраняется хорошо разработанный математический аппарат марковских процессов.

Алгоритм группирует прошлые события в классы, называемые «обусловливающими состояниями» ε_j. Критерием группировки является эквивалентность вероятностей будущих событий, то есть два прошлых события $\overleftarrow{s}_i$ и $\overleftarrow{s}_j$ относятся к одному и тому же обуславливающему состоянию, если распределения их будущих значений одинаковы:

$$P(\overrightarrow{s}|\overleftarrow{s}_i) = P(\overrightarrow{s}|\overleftarrow{s}_j), \tag{4.1}$$

где $P(X|Y)$ — вероятность X при условии Y. Таким образом, вместо переходов между состояниями s_i динамика системы описывается вероятностными переходами между обуславливающими состояниями ε_i. Важно отметить, что последовательность ε_i это по определению *Марковский* процесс независимо от свойств исходного процесса s_i. Совокупность обуславливающих состояний вместе с вероятностями перехода между ними называется «ε-машина». Подробное определение ε-машины дано в приложении Г. ε-машины могут быть реконструированы на основе наблюдаемых данных $\{s_i\}$ с использованием алгоритма `CSSR`, описаного и реализованого в [144]. Вычислительная механика успешно применяется во многих реальных физических системах. Одним из наиболее важных примеров является использование ε-машины для анализа структуры сложных неупорядоченных материалов [269; 270].

Теоретически ВМ формулируется с использованием предположения о бесконечно длинных прошлых и будущих. На практике необходимо выбрать конечную длину истории l, и это один из настраиваемых параметров алгоритма `CSSR` (длина будущего всегда выбирается равной 1). Число возможных историй растет экспоненциально с длиной истории. Поэтому для длинных историй также необходимо экспоненциальное увеличение числа точек данных.

Второй параметр алгоритма `CSSR` – уровень значимости σ используемый при сравнении распределений $P(\overrightarrow{s}|\overleftarrow{s}_i)$ для группировки историй в обуславливающие состояния (используется критерий Колмогорова-Смирнова). Слишком большие значения σ (слишком строгий порог для

того, чтобы два распределения считались эквивалентными) приводят к искусственному появлению слишком большого количества обуславливающих состояний. Это эквивалентно недостаточно большой выборке историй. Та же ситуация имеет место при слишком большой длине истории, поскольку число возможных историй слишком велико, а для умеренно длинных экспериментальных временных рядов распределения $P(\overrightarrow{s}|\overleftarrow{s}_i)$ становятся статистически незначимыми.

Поэтому для получения надежных результатов представляется необходимым провести анализ ε-машины как функции этих двух параметров. Слишком длинная история или слишком большое значение σ приводит к статистически неверным результатам. Как рекомендуют авторы CSSR, величину σ следует выбирать таким образом, чтобы существовало «плато» в числе обуславливающих состояний как функции l. Если таких значений l несколько, то следует выбрать наименьшее (согласно принципу минимальности ВМ). Это постоянное значение l является «истинным» значением длины истории для стабильной архитектуры ε-машины.

ε-машина состоит из последовательностей конформационных состояний. Поэтому, строго говоря, неверно отождествлять состояния машины ε_i с конформационными состояниями s_i. Однако если система находится длительное время в одном из конформационных состояний, то состояния ε-машины в эти времена в основном состоят из последовательности повторяющихся конформационных состояний (например, для состояния «A» это $\overleftarrow{s}_i \equiv \ldots AAAA$). Таким образом, возможно приблизительное соответствие между метастабильными состояниями MSM и nMSM.

Самые интересные моменты времени когда ε-машина генерирует кратковременные переходные состояния. Они описывают механизмы перехода между метастабильными конформационными состояниями.

4.1.3 Результаты

Для исследуемого пептида минимальный временной шаг, при котором конформационные состояния являются марковскими, составляет $\Delta t = 6ps$. Динамика представлена обычной матрицей перехода, а переходом является момент времени, когда символ «A» на временном шаге i изменяется на символ «B» на временном шаге $i+1$ (Рисунок 4.4). Матрица дает вероятности переходов:

	s_{i+1}	A	B	C
s_i				
A		0.996	0.003	0.001
B		0.261	0.737	0.002
C		0.097	0.002	0.901

В этом представлении переход $A \to B$ происходит в среднем один раз в 3.1 нс.

Для временного шага $\Delta t = 5$ динамика становится немарковской, что и показывает ε-машина. Следующая матрица показывает количество реконструированных обуславливающих

состояний ε_i как функция длины истории l и уровня значимости σ (см. раздел 4.1.2):

l	1	2	3	4	5	6	7	8	9
σ									
0.1000	3	6	6	6	6	6	6	6	25
0.0656	3	7	7	8	8	8	8	8	8
0.0430	3	4	11	13	14	14	18	24	57
0.0282	3	4	4	10	11	13	20	24	41
0.0185	3	4	4	4	4	4	4	4	4
0.0122	3	4	4	4	4	4	4	4	4
0.0080	3	4	4	4	4	4	4	4	4
0.0052	3	4	4	4	4	4	4	4	4
0.0034	3	4	4	4	4	4	4	4	4
0.0023	3	4	4	4	4	4	4	4	4
0.0015	3	4	4	4	4	4	4	4	4
0.0010	3	4	4	4	4	4	4	4	4

Присутствует четкое «плато» с минимальным количеством шагов два, необходимых для реконструкции ε-машины.

Поскольку каждому моменту времени присваивается обуславливающее состояние (совокупность одной или нескольких последовательностей $\overleftarrow{s}_i \equiv \{\ldots s_{i-2}s_{i-1}s_i\}$), можно собрать и проанализировать конформации пептида, соответствующие каждому обуславливающему состоянию. Распределение конформаций для каждого состояния вместе с ε-машиной показано на рисунке 4.1.

Состояние «0» на рисунке 4.1 соответствует конформационному состоянию «A», поскольку состоит из последовательностей «AA». Распределение конформаций пептида (Рисунок 4.1, слева) для состояния «0» совпадает с распределением, расположенным в области «A» на рисунке 4.4. Аналогично, состояния «3» и «1» соответствуют конформационным состояниям «B» и «C» соответственно. Дополнительное четвертое состояние, обозначенное «2», описывает переходы между «A» и «B» (последовательности «AB» и «BA»). При данном временном разрешении (5пс) точность времени недостаточна для различения прямого и обратного переходов, поэтому состояние «2» включает в себя как процессы $A \to B$ так и $B \to A$. Длина $l = 2$ временных шагов также определяет временной масштаб, в котором внутренний механизм переходов становится важным: 5пс.

При дальнейшем уменьшении временного шага раскрывается более подробная картина. При $\Delta t = 2{,}5$ три шага становятся коррелированными и минимальная ε-машина состоит из 7 обуслав-

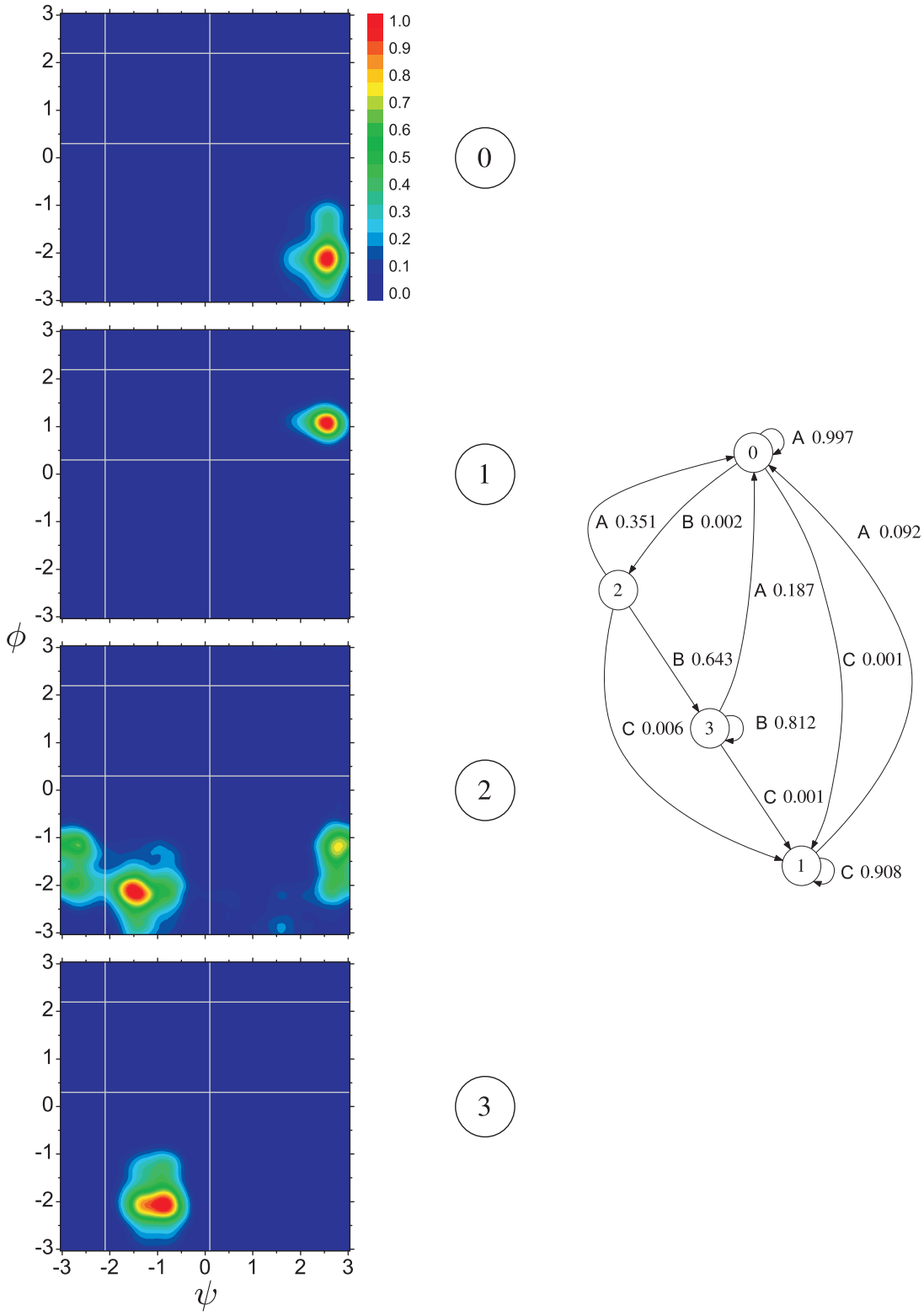


Рисунок 4.1 — Справа: ε-машина для временного шага 5 пс, длина последовательностей $l = 2$, состояние «0» соответствует исходному состоянию «A» (в основном состоит из последовательностей AA), аналогично, состояние «3» соответствует «B», состояние «2» описывает переходы между A и B (состоит из последовательностей AB и BA); слева: распределение конформаций, соответствующих каждому обуславливающему состоянию; для визуализации выбраны моменты времени на последнем символе историй.

ливающих состояний:

l	1	2	3	4	5	6	7	8	9
σ									
0.1000	4	11	11	12	12	12	12	12	54
0.0656	4	11	11	12	12	12	12	12	40
0.0430	4	11	11	12	12	12	12	12	40
0.0282	4	9	9	9	9	9	9	9	9
0.0185	3	9	9	9	9	9	9	9	9
0.0122	3	9	9	9	9	9	9	9	9
0.0080	3	8	7	7	7	7	7	7	7
0.0052	3	8	7	7	7	7	7	7	7
0.0034	3	8	7	7	7	7	7	7	7
0.0023	3	8	7	7	7	7	7	7	7
0.0015	3	6	8	7	7	16	23	34	29
0.0010	3	6	9	8	8	17	24	35	30

Теперь все переходы между парами конформационных состояний AB и AC разрешены (прямые переходы между B и C практически никогда не реализуются). Присутствует та же шкала времени 5 пс, однако для описания деталей необходим один дополнительный шаг в середине процесса.

Наиболее подробную информацию можно получить при $\Delta t = 0.3$. Это минимальный временной шаг, достижимый для доступного объема данных, то есть при наличии переходов, осуществляемых в течение 1μ. Для этого необходимо не менее 4 временных шагов, а число реконструируемых обуславливающих состояний равно 14:

l	1	2	3	4	5	6	7	8	9
σ									
0.1000	4	9	17	21	26	33	41	80	151
0.0656	4	9	16	17	22	27	28	76	134
0.0430	4	9	15	19	20	31	34	84	157
0.0282	4	9	16	19	19	24	31	87	172
0.0185	4	9	15	17	19	20	20	30	139
0.0122	4	9	15	14	14	14	14	31	143
0.0080	4	9	15	14	14	14	14	31	143
0.0052	4	9	15	14	14	14	14	31	143
0.0034	4	9	16	14	14	14	14	31	143
0.0023	4	9	17	18	19	21	29	53	139
0.0015	4	9	17	17	17	17	17	17	140
0.0010	4	9	13	13	13	13	13	13	69

Несмотря на множество обнаруженных состояний обуславливающие состояния, соответствующие метастабильным конформациям, могут быть идентифицированы аналогичным образом. Состояние «0» соответствует «A», тогда как состояние «9» описывает метастабильную

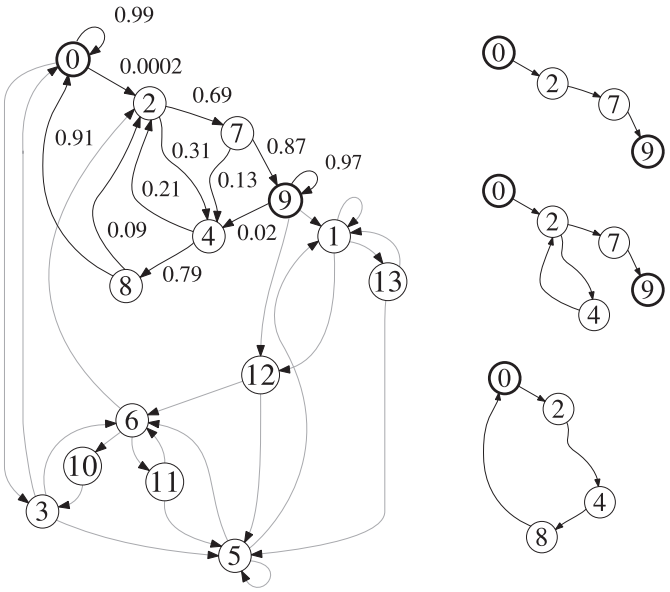


Рисунок 4.2 — Слева: ε-машина для временного шага 0.3 пс, длина последовательностей $l = 4$, состояние «0» соответствует исходному состоянию «A» (в основном состоит из последовательностей $AAAA$), аналогично, состояние «9» соответствует «B», состояние «9» может быть достигнуто только из состояния «0» через состояния «2», «7», «4», «8», которые описывают механизм (пути) перехода $A \to B$; справа: три типичных случая во время перехода: прямой переход (вверху, вероятность 0.56), переход с несколькими повторными пересечениями (в середине, вероятность 0.03), неудачная попытка перехода (внизу, вероятность 0.27), вероятности этих случаев даны в предположении, что вероятность перехода из состояния «0» или в состояние «9» любым возможным путем равна 1.

конформацию «B». ε-машина очень удачно описывает детали перехода $A \to B$: это обуславливающие состояния «2», «7», «4» и «8», Рисунок 4.2.

Система входит в переход $A \to B$ при переходе из обуславливающего состояния «0» в обуславливающее состояние «2» (переход в состояние «3» из «0» соответствует переходу $A \to C$). Из состояния «2» есть две возможности: одна ведет к переходу в состояние «9», соответствующее фактическому переходу $A \to B$, другой в конечном итоге возвращается в состояние «0» и представляет собой неудавшийся конформационный переход. Предполагая, что переход от «0» к «9» или обратно к «0» имеет вероятность 1, вероятности различных маршрутов системы суммируются в таблице 10.

Число возвратов весьма значительно: только 63 процента попыток оказываются успешными. Кроме того, даже если преобладающие последовательности являются прямыми (например, 0279), повторения состояний также не являются незначительными (например, 024279). Они показывают, что повторные переходы очень важны в динамике.

Для этой модели переход $A \to B$ (это левый столбец таблицы 10) происходит в среднем один раз в 2.4 нс, что существенно меньше значения, предоставляемого MSM (3.1 нс).

Можно визуализировать как ε-машина дает очень подробную картину переходов, потока двугранных углов $\psi - \varphi$. Два маршрута, соответствующие левому и правому столбцам таблицы 10 показаны на рисунке 4.3. Малый шаг по времени позволяет достаточно точно идентифици-

Таблица 10 — Последовательности обуславливающих состояний, описывающие два пути, реализуемые при выходе из состояния «0», и их вероятности, см. Рисунок 4.2

$A \to B$		$A \to A$	
последовательность	вероятность	последовательность	вероятность
0279	0.56	02480	0.27
024279	0.03	027480	0.08
0248279	0.02	0242480	0.01
0274279	0.009	02482480	0.007
02748279	0.006	02742480	0.004
02482424279	0.001	02427480	0.003
		0242482480	0.001
всего:	0.63	всего:	0.37

ровать момент перехода. Таким образом, можно показать различные моменты, предшествующие переходу. Интересно, что динамика, по-видимому, очень похожа для обоих маршрутов вплоть до ≈ 3 пс до перехода. Затем она становится другой, что приводит либо к переходу к конформации B, либо к возврату в A.

В целом механизм показывает, что прямой переход очень быстрый, он занимает всего ≈ 4 временных шага, то есть 1.2пс. Значимость этого факта в том, что это время существенно меньше минимально допустимого масштаба времени для стандартного MSM. Поэтому важные детали могут быть упущены, по крайней мере для небольших пептидов, таких как изучаемый. Причина относительно больших требуемых временных шагов в MSM заключается в множественных повторных пересечениях, которые маскируются большими временными шагами и, таким образом, делают динамику марковской. Это является еще одним доказательством полезности nMSM в изучении динамики пептидов.

4.1.4 Выводы

Представлено развитие метода марковских состояний, которое учитывает немарковское поведение конформационных состояний белка. Метод позволяет избежать ряда недостатков, присущих MSM. В частности, для определения корректных метастабильных состояний не требуется множества микросостояний. Динамика системы автоматически находит необходимые состояния.

Важно, что методология дает возможность сократить временной шаг, тем самым увеличивая временное разрешение модели. В принципе, разрешение ограничивается только объемом доступных для анализа данных, тогда как в стандартном MSM ограничение разрешения по времени является концептуальным.

Наш метод дает подробное описание механизмов переходов. Он определяет важность множественных повторных переходов и количественно оценивает их вероятности. Поскольку ме-

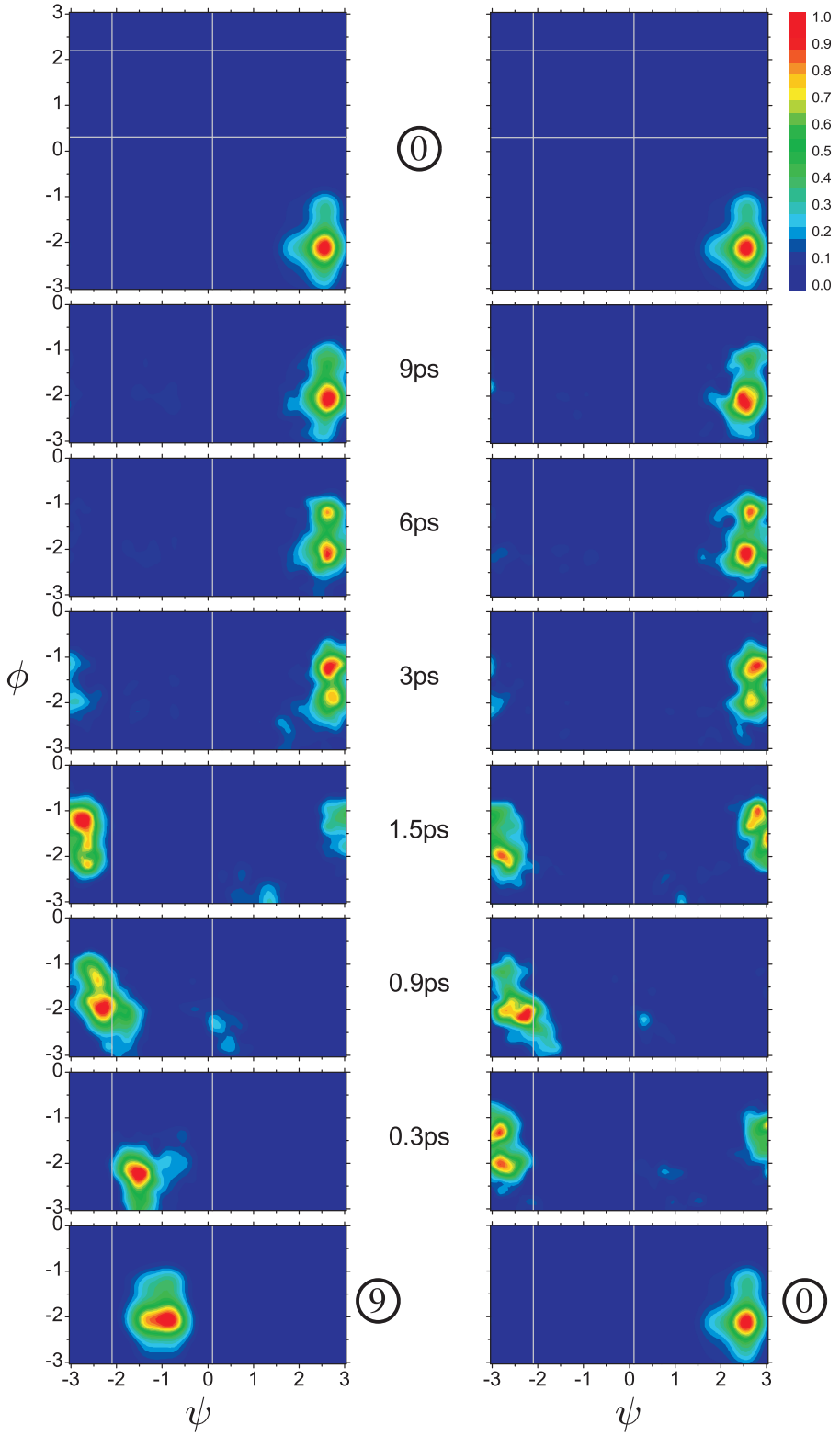


Рисунок 4.3 — Поток двугранных углов, описывающий две различные возможности перехода из состояния «0»: слева, переход $A \to B$, справа, неудавшийся переход и возврат в состояние A; числа показывают время до завершения перехода (то есть, когда система достигает либо состояния «0», либо состояния «9»); для визуализации выбраны моменты времени в последнем символе историй.

ханизмы переходов теперь можно выяснить с высоким временным разрешением, можно получить подробную молекулярную картину конформационных движений.

Механизмы переходов для этого конкретного пептида показывают, что, вероятно, существуют и другие важные степени свободы, которые необходимо включить в картину. Можно предположить, что такие степени свободы должны включать молекулы воды, и это является предметом наших последующих исследований.

4.2 Вода управляет конформационными переходами пептидов

Недавние исследования динамики белков показывают, что вода играет главную роль в движении белков. Действительно, существует большой объем экспериментальных и компьютерных свидетельств [271—274] показывающи тесную связь между динамикой воды и конформациями белков. Фрауэнфельдер и коллеги экспериментально показали, что доминирующие конформационные движения белка подчиняются динамике гидратной оболочки и объемного растворителя [3; 36—39], в то время как сама молекула белка обеспечивает «активную матрицу», необходимую для направления динамики воды в сторону биологически значимых конформационных изменений. Изменение динамики воды в оболочке размером почти в десяток диаметров молекулы воды вокруг белков обнаружено в [275]. Исследование нейтронного рассеяния продемонстрировало, что граничная (гидратационная) вода является основной «движущей силой» динамики белков, регулирующей как локальные, так и крупномасштабные движения в белках [17; 276; 40—42]. Наконец, была продемонстрирована критическая роль сольватирующей воды в важной прикладной области связывания белка с лигандом [277].

Несмотря на обширные исследования динамики белков, исследования *элементарных конформационных движений* редки [43—45]. С другой стороны, знание конкретных молекулярных механизмов, включая роль молекул воды, управляющих конформационными движениями, крайне востребовано, поскольку эти элементарные конформационные изменения в конечном итоге определяют все перестройки белков в целом.

В данном разделе проанализирован пептид, смоделированный с помощью молекулярной динамики, уделяя особое внимание моментам элементарных конформационных изменений, включая атомистические молекулы воды. Показано, что вода действительно является движущей силой изменений, и продемонстрированы конкретные механизмы этого явления.

4.2.1 Анализируемые величины: конформационные переходы в диаланине

Были изучены две различные модели системы. Одна — объединённое атомное силовое поле GROMOS [278], другая — OPLS (оптимизированные потенциалы для моделирования жидкостей) [267]. Это одни из самых популярных моделей MD для пептидов и белков. Они обе дают оди-

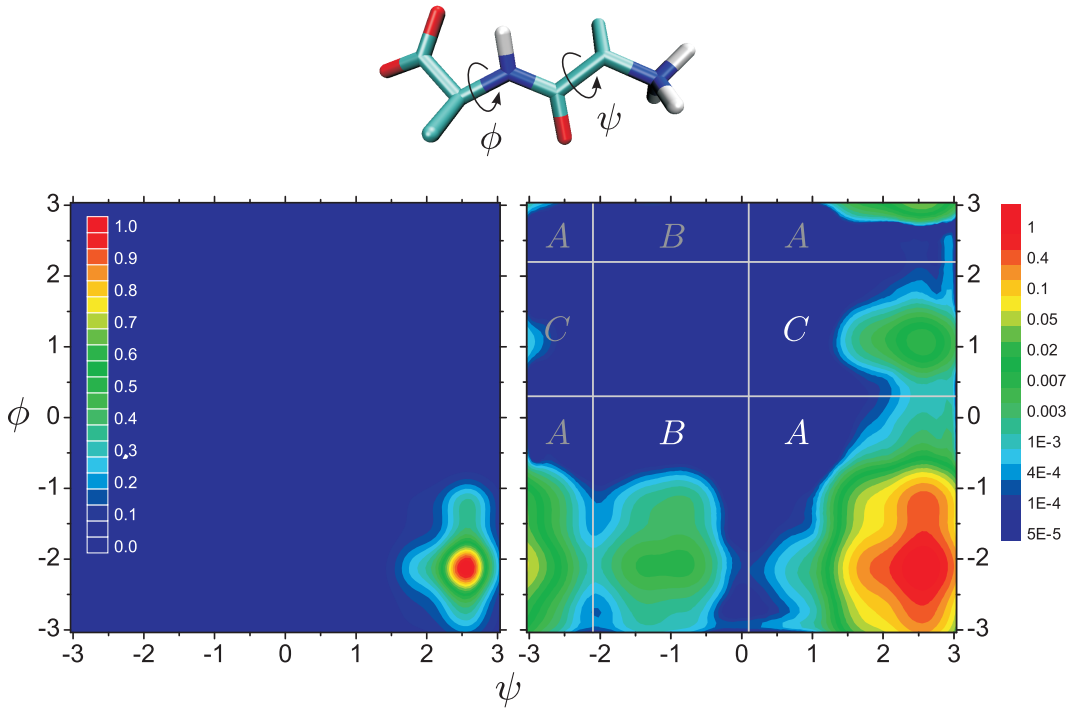


Рисунок 4.4 — Вверху: цвиттер-ион L-аланил-L-аланина (силовое поле GROMOS, см. приложение); слева: Нормализованные вероятности конформаций (график Рамачандрана), образованные 2μs траекториями; справа: те же вероятности, подчеркивающие наличие двух второстепенных конформаций и разбиение для символизации.

наковые результаты в наших исследованиях, несмотря на существенно различное представление атомов и их взаимодействий для молекул пептида и воды.

Важным концептуальным моментом является определение конформационных состояний и моментов переходов между ними. Поскольку нас интересуют динамические свойства, определяем состояния как динамически метастабильные состояния, то есть конформации, в которых молекула проводит значительно больше времени по сравнению со временем, которое она тратит на переходы между конформациями. Это отражено в вероятностях конформаций молекулы, рассчитанных как вероятность нахождения точки траектории MD с определенными значениями ψ и φ за все время моделирования MD. Имеются три хорошо разделенных метастабильных состояния, отчетливо видимых в пространстве вероятностей конформации, Рисунок 4.4. Они позволяют ввести простую естественную дискретизацию конформаций, которые обозначаем как «A», «B» или «C».

Теперь можно определить моменты переходов. Описанная дискретизация определяет границы конформационных состояний. Таким образом, можно было бы определить моменты переходов между состояниями, когда траектория пересекает границы. Однако границы очерчивают вероятности конформаций, усредненные по всей траектории. Отдельные участки траектории не переходят напрямую из одного состояния в другое, а извиваются сложным образом, часто пересекая границы много раз, прежде чем принять новую конформацию. Это приводит к «мерцающим» состояниям, когда они образуют последовательности очень кратковременных чередующихся состояний, таких как «$\ldots ABABAB \ldots$», которые, очевидно, не удовлетворяют требуемому свойству метастабильности. Иногда после нескольких таких переходов траектория возвращается в исходное состояние,

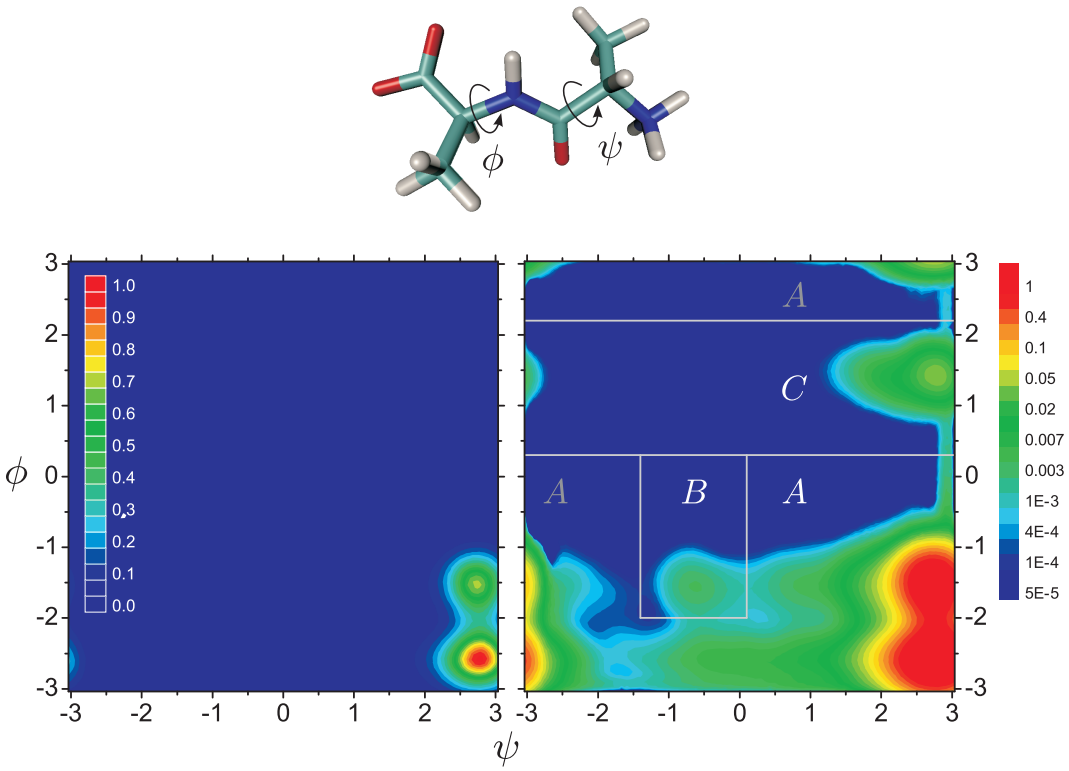


Рисунок 4.5 — Вверху: цвиттер-ион L-аланил-L-аланина (силовое поле OPLS); слева: Нормализованные вероятности конформаций (график Рамачандрана), образованных 2μs траектория; справа: те же вероятности, подчеркивающие наличие двух второстепенных конформаций и разбиение для символизации.

не задерживаясь в новом состоянии достаточно долго (это создает хорошо известную проблему «повторяющихся переходов» в химической кинетике [279—281]).

Наш подход к решению этой проблемы с использованием скрытой Марковской модели в реализации ε-машины описан в разделе 4.1.

4.2.2 Поведение воды

Предметом нашего исследования было поведение воды. Было рассчитано распределение атомов кислорода (водорода) в пространстве *путем усреднения по выбранным временным интервалам*. Вычисляемое поле $f(x,t)$ дает вероятность обнаружения атома кислорода (водорода) в небольшом объеме вокруг местоположения x в момент времени t *в преддверии перехода*. Эта вероятность равна числу атомов в малом объеме, деленному на общее число атомов в системе. Если бы атомы были распределены равномерно, это число было бы равно плотности кислорода (водорода), деленной на атомную массу. Преимущество такого вероятностного представления состоит в том, что фокус смещается от физической плотности. Рассчитанное количество *не* равно плотности, поскольку последняя получается путем деления числа атомов (умноженного на их массу) на объем. Структура воды определяется сетью водородных связей, что подразумевает при-

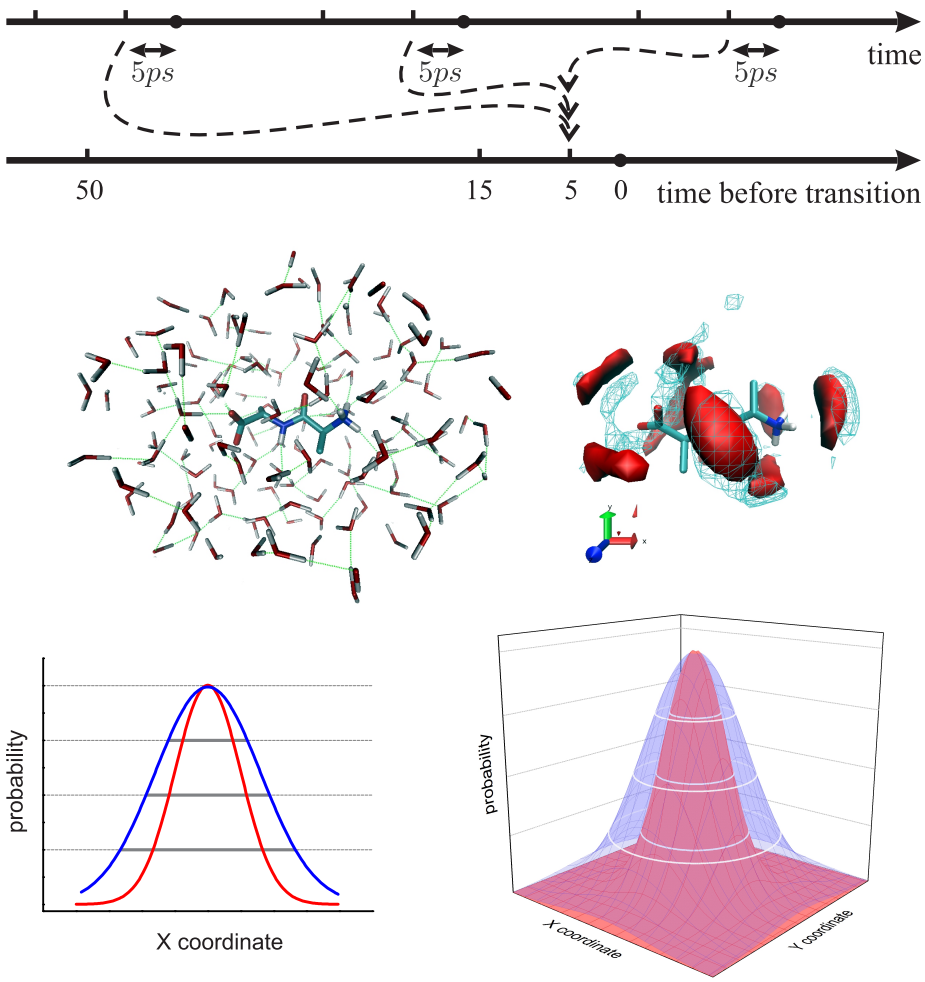


Рисунок 4.6 — Вверху: сбор временных точек для статистики «время до перехода»; точки на осях «время» — это моменты перехода; в середине слева: снимок молекулярной системы, показывающий ≈ 2-3 слоя молекул воды, связанных водородными связями друг с другом и пептидом; в середине, справа: распределение атомов кислорода (красный, сплошной) и водорода (синий, сетчатый) усреднено по всей симуляции (показаны изоповерхности при значении вероятности 0.047 (значение в предположении полностью равномерного распределения равно 0.031) для модели GROMOS); внизу: одномерное и двумерное распределение вероятностей, иллюстрирующее сужение пика для высокой вероятности атомов; красное, более узкое распределение описывает ситуацию, когда атомы имеют более определенные местоположения, а не более равномерно распределены, как характерно для синего распределения.

мерно одинаковое расстояние между молекулами воды повсюду (и, следовательно, одинаковую локальную плотность), Рисунок 4.6, в центре слева. Однако вероятности могут существенно различаться в разных положениях, указывая на предпочтительные положения атомов, Рисунок 4.6, в центре справа.

Переход из конформации A в конформацию B соответствует повороту группы NH_3 на $\approx 180^\circ$ (правая сторона молекулы на Рисунок 4.6, в середине), в то время как во время перехода $A \to C$ группа COO вращается аналогичным образом. Оба конца молекулы несут заряды: отрицательный на участках COO и положительный на участках NH_3, что приводит к относительно прочному присоединению молекул воды к этим концам. Водородно-связанные молекулы воды

с атомами кислорода и водорода образуют области высокой вероятности атомов воды, Рисунок 4.6, в середине, что указывает на то, что в среднем в этих местах более вероятно обнаружение атомов воды.

Это интуитивно понятный результат. Более интересен момент непосредственно перед (1-10пс) переходами. В это время вода перестраивается вокруг пептида, однако хотелось бы подчеркнуть, что перестройка носит коллективный, относительно медленный (по сравнению с движением отдельных атомов) характер. Вода не «уступает» место атомам пептида, а вся структура окружающей воды изменяется одновременно с изменением средней конформации пептида. Вокруг пептида нет «плотной» воды. Области высокой вероятности показывают, что вода имеет тенденцию оставаться в определенных местах вокруг пептида, а не то, что локальная плотность (количество атомов на единицу объема) выше. Эти области указывают на пики вероятности. Чем шире пики, тем равномернее распределение, и наоборот, узкие пики указывают на предпочтение определенных мест.

Для количественной оценки этого эффекта полезна аналогия с одномерными и двумерными распределениями (Рисунок 4.6, нижний). Для одномерного распределения ширина пика на разных уровнях вероятности дает меру тенденции атомов находиться в местах с высокой вероятностью. Для двумерного случая ширина становится площадью двумерных областей, для которых значение вероятности выше выбранного уровня, Рисунок 4.6, нижний. Наконец, трехмерное распределение, как в нашем случае с молекулами, можно охарактеризовать объемами, охватываемыми изоповерхностями, соответствующими различным уровням вероятности.

Было обнаружено, что эти объемы значительно уменьшаются перед переходом на всех уровнях вероятности, достаточно высоких, чтобы различить пики вероятности вокруг пептида, Рисунок 4.7, левый. Результат для двух изученных моделей и двух переходов $A \to B$ и $A \to C$ показаны на рисунке 4.7, справа. В период времени когда вращение группы NH_3 (COO) наиболее существенно, от ≈ 10 до ≈ 1, размер областей с высокой вероятностью появления воды значительно сокращается. Это указывает на то, что в эти периоды вода имеет тенденцию оставаться в более определенных местах вокруг пептида. Другими словами, поток траекторий MD во время переходов концентрируется в более узких «каналах».

Наиболее информативный временной интервал перед переходом, 1–10 пс, имеет тот же порядок, что и наименьший возможный временной шаг для обычного MSM (6 пс). Таким образом, использование нашей скрытой марковской модели имеет решающее значение для анализа деталей переходов. Рисунок 4.7, справа, показывает, что эффект очень похож для двух изученных конформационных переходов и обеих моделях MD.

Общие изменения структуры воды происходят одновременно с изменением двугранного угла, отражая конформационный переход. Можно количественно оценить степень зависимости между областями воды с высокой вероятностью и двугранными углами пептида, вычислив динамические корреляции между ними. Это делается с помощью метода линейной стохастической оценки (LSE), который широко применяется в динамике жидкостей: от визуализации когерентных структур в турбулентных потоках [282] до идентификации источников шума в турбулентных струях [283].

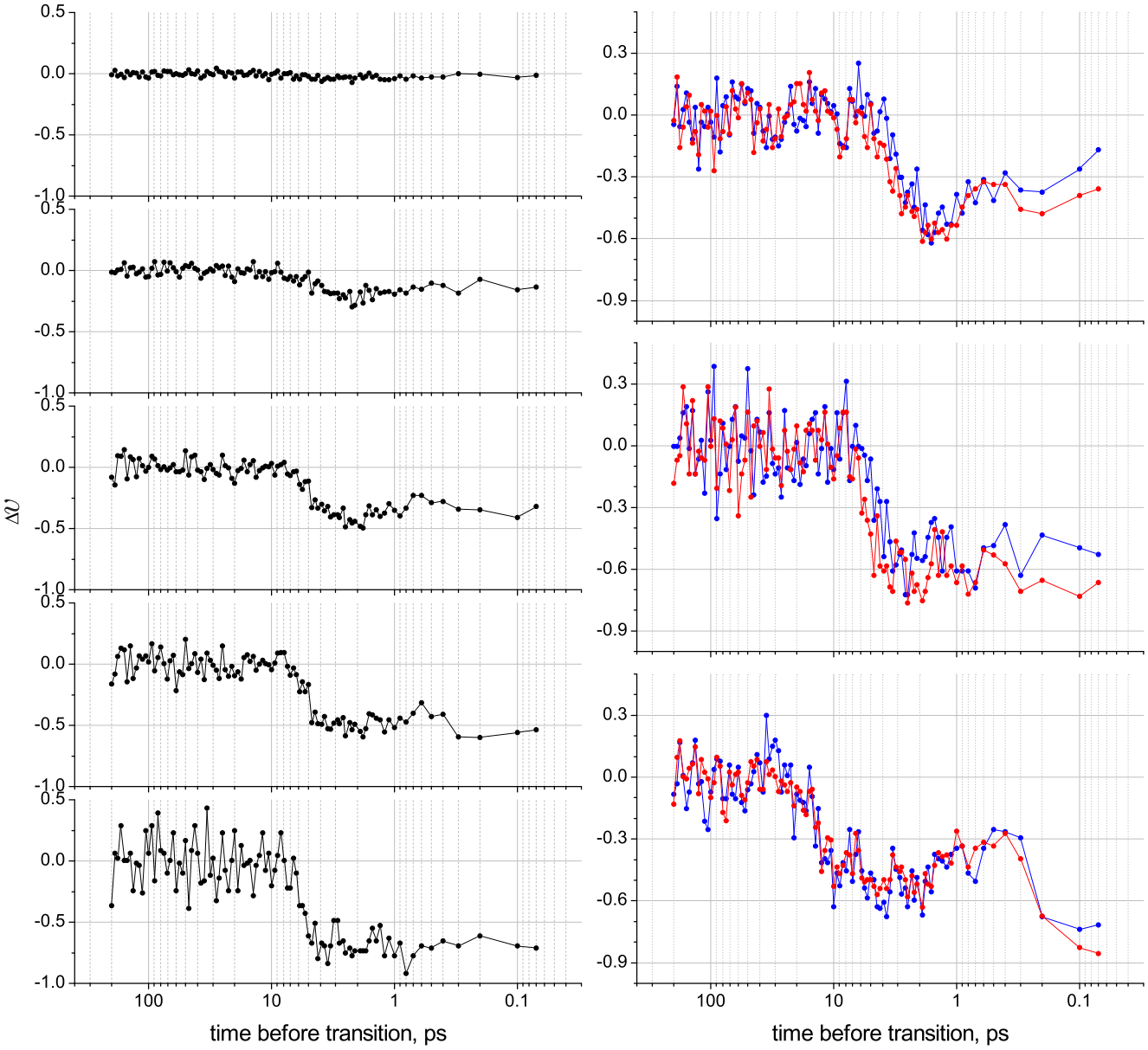


Рисунок 4.7 — Слева: зависимость объема областей высокой вероятности от «времени до перехода», нормализованная к средним значениям, $\Delta v = \frac{v - \langle v \rangle}{\langle v \rangle}$, где v — абсолютный объем (см. текст), а $\langle v \rangle$ — средний объем за все время моделирования, показаны площади изоповерхностей на уровнях вероятности (сверху вниз) 0.013, 0.041, 0.046, 0.051 и 0.059; справа: то же самое для кислорода (красный) и водорода (синий) для перехода из состояния A в состояние B, модель GROMOS (вверху), переход $A \to C$, модель GROMOS (в середине) и переход $A \to C$, модель OPLS (внизу), показаны площади изоповерхностей на уровне вероятности 0.055.

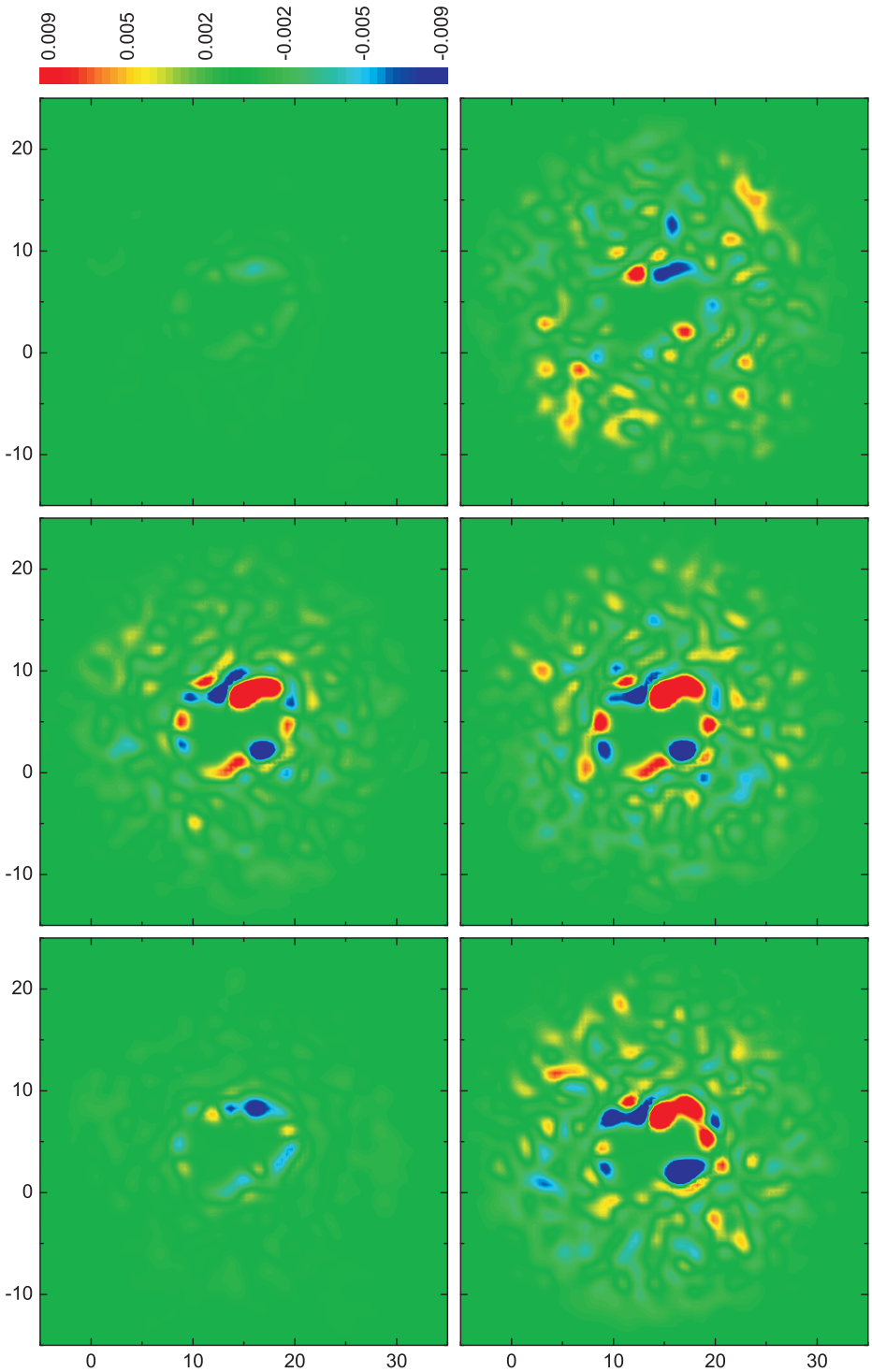


Рисунок 4.8 — Сечение xy (значение координаты z выбрано таким образом, что поле флуктуации вероятности $f'(x,t)$ (справа), и его часть коррелируемой с конформацией пептида $f'_c(x,t)$ (слева) для кислорода; время до перехода составляет, сверху вниз, $200ps$, $0.8ps$ и $0ps$; шкала ограничена интервалом $-0.009 \ldots 0{,}009$ для наглядности, максимумы пиков достигают значений 0.045 и -0.041 (приведены результаты для модели GROMOS).

В рамках LSE поле вероятностей $f(x,t)$ раскладывается на линейную комбинацию двух частей: коррелированной и некоррелированной с двугранными углами пептида. Для извлечения коррелированного поля используется функция $f(x,t)$ сначала преобразуемая во временные колебания путем вычитания среднего значения по времени: $f'(x,t) = f(x,t) - \langle f(x,t)\rangle_t$. Затем рассматривая эти колебания воды как стохастический сигнал по времени, можно получить линейную стохастическую аппроксимацию $f'_c(x,t) = \alpha(x)\varphi(t) + \beta(x)\psi(t)$ в $f'(x,t)$, где $\varphi(t)$, $\psi(t)$ – средние углы в момент времени t (то есть стохастические переменные) и α, β — константы. Стохастическое соответствие находится путем минимизации статистической ошибки $\langle f' - f'_c\rangle_t$ [284]. $f'_c(x,t)$ представляет собой колебания воды, коррелирующие с углами пептида и α, β — коэффициенты корреляции.

Для нахождения коэффициентов для каждой точки объема воды решается система линейных уравнений относительно α, β. Эти линейные уравнения, полученые из $\langle f'(x,t)\varphi(t)\rangle = \langle\{\alpha\varphi(t) + \beta\psi(t)\}\varphi(t)\rangle$ (и аналогично для $\psi(t)$), есть

$$\begin{aligned}\alpha\langle\varphi(t)\varphi(t)\rangle + \beta\langle\psi(t)\varphi(t)\rangle &= \langle f'(x,t)\varphi(t)\rangle \\ \alpha\langle\varphi(t)\psi(t)\rangle + \beta\langle\psi(t)\psi(t)\rangle &= \langle f'(x,t)\psi(t)\rangle.\end{aligned}$$

Теория линейной стохастической оценки и результаты для модели OPLS.

В рамках линейной стохастической оценки переменная поля воды $f(x,t)$ понимается как опорный стохастический сигнал $s(t)$, время t как стохастическая переменная, а углы $\varphi(t)$, $\psi(t)$ – как пространство управляющих (стохастических) процессов $x_1(t),\ldots,x_n(t); n = 1,2$. Для извлечения коррелированной части стохастического сигнала относительно процессов управления предполагается, что шумовая часть сигнала ортогональна пространству процессов управления. Следовательно, полезная коррелированная часть сигнала $\hat{s}$ получается путем проецирования исходного сигнала на процессы управления, т.е. как условное среднее $\langle s(t)|x_1(t),\ldots,x_n(t)\rangle \rightarrow \hat{s}$ [284]. Проекция $\hat{s}$ находится как линейная комбинация процессов управления, $\hat{s} = \sum a_i x_i$. Умножая обе части последнего уравнения на процесс управления, интегрируя результат по времени и используя тот факт, что шумовая часть ортогональна процессам управления, $\langle\hat{s},x_1(t)\rangle = \langle\sum a_i(x_i,x_1(t))\rangle = \langle s(t),x_1(t)\rangle$, получаем линейное уравнение для коэффициентов a_i. Дополнительные $n-1$ уравнений получаются путем повторения вышеописанной процедуры для всех остальных оставшихся процессов управления $x_2(t),\ldots,x_n(t)$. Затем неизвестные коэффициенты a_i определяются путем решения соответствующей линейной системы. Видно, что решение линейных уравнений соответствует минимуму математического ожидания остатка $\langle s(t) - \hat{s}\rangle$, остаток которого представляет собой шум, который необходимо было отделить от полезного сигнала [284]. Согласно теории обработки сигналов [284], отфильтрованный сигнал $\hat{s}$ также соответствует глобальному минимуму ошибки, т.е. стохастическая подгонка соответствует глобальному оптимуму, в случае, когда все рассматриваемые процессы имеют совместное нормальное распределение с нулевым математическим ожиданием. В нашем случае, чтобы убедиться, что распределение управляющих процессов является гауссовым, исходное распределение углов пептидов соответствующим образом перенормируется.

Поле флуктуаций вероятности $f'(x,t)$ и его часть, коррелирующая с конформацией пептида $f'_c(x,t)$ показаны на рисунке 4.8 для нескольких репрезентативных моментов времени для модели GROMOS.

Колебания для воды (правая колонка) значительно сильнее непосредственно перед переходом (0.8 пс) по сравнению со стабильным периодом ($\approx$ 200 пс). Удивительно, что условно усредненные колебания воды (левая колонка) практически не коррелируют с пептидом во все времена, за исключением короткого периода непосредственно перед переходом (некоторые корреляции также заметны уже в ≈ 5 пс до перехода, что согласуется с началом уменьшения вероятности на рисунке 4.7). Интересно, что в 0 пс, когда процесс перехода завершен, вероятность воды становится некоррелированной с пептидом, как и в стабильные периоды. Однако колебания его остаются сильными, лишь немного слабее, чем на уровне 0.8 пс. Можно объяснить этот эффект большой инерцией водной оболочки, которой требуется сравнительно много времени для затухания колебаний. Тот факт, что эти колебания воды после перехода отделены от пептида, подчеркивает обнаруженное явление сильных взаимодействий воды и пептида именно во время процесса перехода.

Для модели OPLS эти результаты очень похожи для перехода $A \to C$. Переход $A \to B$ плохо разрешен для этой модели, и поэтому прямое сравнение с моделью GROMOS невозможно.

Колебания плотности воды и их часть, коррелированная с двугранными углами для модели OPLS, показаны на рисунке 4.9.

4.2.3 Выводы

Подводя итог, было обнаружено, что (i) от 10 до 1 пс до перехода, когда двугранные углы изменяются больше всего, молекулы воды имеют тенденцию располагаться в более определенных положениях вокруг пептида по сравнению с более равномерным распределением в другие моменты времени; (ii) во время перехода динамика распределения воды становится сильно коррелированной с динамикой двугранных углов; и (iii) эти корреляции полностью отсутствуют в периоды стабильной конформации.

Сделан вывод, что вода и пептид ведут себя как единая динамическая система. Во время конформационного перехода пептид и окружающая его вода претерпевают переходы вместе. Это контрастирует с метастабильными периодами, когда их динамика по существу разъединена. Переход характеризуется более конкретно определенной сетью водородных связей воды. Таким образом, вполне вероятно, что вода является движущей силой всего процесса конформационных переходов.

4.3 Самосборка молекул трегалозы на поверхности лизоцима: гипотеза разбитого стекла

Растворы сахаров помогают биомолекулам сохранить свою структуру в жестких условиях, включая обезвоживание и высокие температуры. Среди природных дисахаридов трегалоза, по-видимому, является одним из наиболее эффективных стабилизирующих агентов [285—291]. Трегалоза, также называемая микозой или грибным сахаром [285], представляет собой невосстанавливающий гомодисахарид, в котором две единицы D-глюкопиранозы связаны вместе $\alpha -$ 1,1-гликозидной связью.

Несмотря на многочисленные экспериментальные и теоретические исследования взаимодействия трегалозы с белком [292—300] мало что известно о молекулярном механизме стабилизирующего действия трегалозы на биомолекулярную структуру поскольку экспериментальные результаты трудно интерпретировать на атомистическом уровне. Разумно предположить, что стабилизация является результатом особых взаимодействий молекул сахара с белком, которые приводят к образованию нетривиальных, стабилизирующих молекулярных структур. Существуют три основные гипотезы, описывающие такие структуры [285]:

- *Гипотеза механического захвата (стеклования):* захват биомолекул в стекловидной матрице трегалозы, образованной в среде *концентрированного* раствора трегалозы с высокой вязкостью. Это должно защитить естественную конформацию биомолекул подобно насекомым, пойманным в янтаре.
- *Гипотеза замены воды:* защита биомолекул посредством прямого взаимодействия трегалозы с поверхностными группами биомолекул через водородные связей. Эта гипотеза предполагает, что большая часть молекул воды в первой гидратной оболочке биомолекулы должна быть заменена трегалозой.
- *Гипотеза захвата воды:* улавливание молекул воды в промежуточном слое между сахарами и биомолекулярной поверхностью.

Отметим, что первая гипотеза применима только к растворам с высокой концентрацией трегалозы ($\gg$ 1M) и низким содержанием воды. Трегалоза, по-видимому, обладает особенно благоприятными свойствами в этих условиях [301]. Однако это не объясняет почему трегалоза также проявляет хорошие стабилизирующие свойства в низко- и среднеконцентрированных растворах ($\approx$0.05-1.0 M), естественных для живых организмов [302—304].

Обе гипотезы, *замена воды* и *захват воды*, подтверждаются различными наборами экспериментальных и теоретических данных. Эксперименты по инфракрасной спектроскопии, описанные в [305] показывают, что между трегалозой и лизоцимом образуется большое количество прямых водородных связей, что свидетельствует о *гипотезе замены воды*. Однако есть также много экспериментальных [292—295] и теоретических [296; 297; 299; 300] данных показывающих, что взаимодействие белка и сахара лучше описывается в терминах гипотезы *захвата воды*.

Целью данного исследования является пересмотр вопроса о природе взаимодействия трегалозы с белками и выявление деталей молекулярного уровня связывания трегалозы и воды с

белками в среднеконцентрированных растворах с помощью долговременного моделирования молекулярной динамики лизоцима в водном растворе трегалозы (0.75 M).

Исследуя конкретные молекулярные *структуры*, образованные путем взаимодействия молекул трегалозы с белком в воде, были также обнаружены отличия таких структур в *динамике* белка. Обнаружена четкая корреляция между гибкостью белка и кластерами молекул сахара что может быть связано со структурной стабильностью белка.

Сворачивание и разворачивание лизоцима занимает гораздо больше времени чем это в настоящее время возможно смоделировать для полной системы вместе с необходимым количеством (несколько тысяч) молекул воды и трегалозы в водном растворе [306; 307]. Однако можно проанализировать гибкость атомов остова лизоцима и это должно быть связано со стабильностью третичной структуры белка. Эта связь была продемонстрирована для термофильных белков которые биохимически активны при высокой температуре. Экспериментальные исследования показали, что эта термостабильность напрямую связана с гибкостью этих белков [308] и что термофильные белки столь же гибки при высокой температуре как мезофильные при комнатной температуре [309]. Поэтому, можно ожидать, что трегалоза влияет на гибкость лизоцима. Если гипотеза механического захвата верна то снижение гибкости должно произойти для всего белка, тогда как остальные альтернативные гипотезы должны привести к более локализованным эффектам [46].

4.3.1 Детали моделирования

Было проведено 30-нс моделирование методом атомистической молекулярной динамики при комнатной температуре (300 K) водного раствора лизоцима куриного яйца в присутствии и в отсутствие трегалозы. Белок был помещен в центр кубической ячейки, которая также содержала (i) $\sim 2.2 \cdot 10^4$ молекул воды для основного водного раствора; (ii) $\sim 1.4 \cdot 10^4$ молекул воды и 256 молекул трегалозы для водного раствора сахара. Моделирование начиналось с конформации белка предоставленной данными ЯМР в Protein Data Bank (код PDB 1e8l [310]; была взята 26-я конформация из 50 конформаций, предоставленных этой записью PDB) с помощью GROMACS 3.3. В начале моделирования молекулы трегалозы были случайным образом распределены по всей системе что дало концентрацию 0.75 M. Первоначальное распределение молекул трегалозы по ячейке не показало никаких кластеров. Перед началом прогона каждая система была уравновешена в течение 4 нс с фиксированными положениями атомов белка. В ходе этого уравновешивания первоначальное случайное распределение молекул трегалозы изменилось образовав кластеры на поверхности белка которые сохранялись на протяжении всего прогона. Раствор нейтрализовали путем добавления в ячейку моделирования нескольких противоионов. Было использовано силовое поле GROMOS 53a6 [116] для белка, ионов и трегалозы вместе с моделью воды SPC/E [311]. Были использованы GROMOS 53a*-совместимые оптимизированные заряды для трегалозы, взятые из работы [312]. Силовое поле было выбрано из-за его адекватного описания белков и олигоса-

харидов [116; 266; 312]. Временной шаг интегрирования MD составил 2 фс, электростатические взаимодействия были обработаны с помощью метода коррекции поля реакции [313].

Для анализа плотности сахара и воды в системе поверхности изоплотности были рассчитаны как число атомов, занимающих трехмерные узлы сетки, и усреднены за 30 нс моделирования (узел сетки считается занятым, если он лежит внутри сферы атомного радиуса с центром на любом атоме любой молекулы сахара или воды). На рисунках 4.10,4.11 представлена плотность трегалозы по отношению к среднему числу молекул в растворе. На рисунке показан объем, в котором содержание сахара в 6.4 раза выше, чем в остальной части раствора. Это значение было выбрано потому, что оно дает четкое представление о высокой плотности сахара вокруг некоторых частей белка. Также существуют некоторые обласи, где содержание воды в воде превышает средний показатель более чем в 1.5 раза, поскольку в этих местах содержание сахара ниже, и это также показано на рисунке 4.10.

4.3.2 Результаты

Для атомов основной цепи белка были рассчитаны среднеквадратичные флуктуации (RMSF). Распределение RMSF по положению вдоль остова показано на рисунке 4.12.

Анализ имеющихся экспериментальных данных ЯМР [310] ясно показывает, что модель гибкости пептида хорошо воспроизводится при моделировании, Рисунок 4.13. Ожидались более высокие значения RMSF при моделировании поскольку структуры, полученные с помощью ЯМР, были ограничены экспериментальными измерениями.

При сравнении гибкости пептида в чистой воде и растворах сахара отчетливо видно уменьшение подвижности атомов основной цепи белка в растворе трегалозы, особенно в аминокислотах С-конца (остатки 123-129), которые обладают высокой подвижностью в объеме водного раствора, Рисунок 4.12. Такое неравномерное снижение подвижности белка нельзя объяснить только повышением вязкости раствора сахара, поскольку при данной концентрации сахара она должна быть всего лишь в два раза выше, чем в растворе чистой воды [314].

Отметим, что аминокислоты 30 - 40 и 48 - 50 становятся более подвижными при добавлении трегалозы к водному раствору. Это увеличение подвижности остова при использовании трегалозы происходит вокруг положения двух аминокислот активного центра GLU35 и ASP52, как показано на рисунке 4.10, это увеличение гибкости связано с активным центром фермента. Лизоцимы специфически связываются с пептидогликанами и олигосахаридами, обнаруженными в клеточных стенках бактерий [315—319]. Ферменты гидролизуют гликозидные связи в этих молекулах, искажая кольцо в конформацию полукресла. В этом напряженном состоянии гликозидная связь легко разрывается [316; 317; 319]. Таким образом, аминокислоты активного центра лизоцима куриного яйца должны взаимодействовать с сахарами, но это взаимодействие будет оптимизировано для процесса гидролиза, а не прочного связывания. Интересно отметить, что активный центр соответствует единственным положительным пикам на рисунке 4.12.

Для выяснения механизма влияния трегалозы был построен график плотности сахара по отношению к поверхности белка, Рисунок 4.10, 4.11. Из этих рисунков можно сразу сделать два важных вывода: (i) молекулы сахара образуют кластеры *только в непосредственной близости от белка* (Рисунок 4.11) и (ii) расположение кластеров *соответствуют местоположению пониженной гибкости*. Кроме того, на поверхности лизоцима в растворе трегалозы имеются также небольшие кластеры воды высокой плотности, и это согласуется с *гипотезой захвата воды*. Отметим, что увеличение плотности воды вокруг лизоцима наблюдалось экспериментально с помощью рентгеновского и нейтронного рассеяния в работе [320]. Размер кластеров сахара по отношению к отдельным молекулам трегалозы показан на рисунке 4.14 где некоторые молекулы сахара явно изображены для случайно выбранного кадра моделирования.

Было подсчитано количество водородных связей белок-сахар, и результаты показаны на рисунке 4.15 с соответствующими структурными типами, Рисунок 4.16. Аминокислоты, образующие большинство водородных связей с сахаром либо обладают пониженной подвижностью либо располагаются поблизости от менее подвижных аминокислот. Таким образом, наша гипотеза заключается в том, что образование водородных связей (и, следовательно, иммобилизующий эффект) в белке способствует образованию долгоживущих сахарных кластеров, что, в свою очередь, снижает гибкость остова белка. Это согласуется с *гипотезой замены воды*.

Чтобы понять общие тенденции в механизме связывания белков с водой и трегалозой, было рассчитано среднее число внутренних белок-белковых водородных связей как для водных, так и для трегалозных растворов и обнаружено, что в среднем оно на 10% меньше для раствора трегалозы, чем для раствора чистой воды. В таблице 11 показано, сколько водородных связей образуется при взаимодействии белка с трегалозой и белка с водой. Результаты показаны только для тех аминокислот, боковые цепи которых способны образовывать водородные связи (заряженные, полярные и некоторые ароматические аминокислоты). При концентрации сахара использованной в исследовании (0.75 M) трегалоза имеет лишь десятую часть водородных связей. Однако трегалоза образует более 20% от общего числа водородных связей с большинством боковых цепей. Более того, взаимодействия трегалозы с кислыми аминокислотами (GLU и ASP) еще более благоприятны — они образуют около 40 % от общего числа водородных связей с боковыми цепями этих аминокислот. Это является явным указанием на преимущественное взаимодействие молекул трегалозы с полярными и, особенно, кислыми аминокислотами.

4.3.3 Обсуждение

Гибкость белкового остова связана со стабильностью третичной структуры, но детали молекулярного механизма этого и даже знак эффекта не ясны [321]. Однако имеются данные демонстрирующие корреляцию между гибкостью белков и их структурной стабильностью.

Эксперименты по релаксации спина ЯМР продемонстрировали, что химический денатурант усиливает колебания белка, что в конечном итоге приводит к его разворачиванию [322].

Таблица 11 — Водородные связи (ВС) боковых цепей белка с трегалозой и водой в % от общего числа ВС аминокислота-растворитель

Аминокислота	Белок-трегалоза	Белок-вода
LYS	22	78
ARG	22	78
GLU	39	61
ASP	44	56
GLN	20	80
ASN	24	76
SER	26	74
THR	21	79
HIS	14	86
TYR	11	89
TRP	24	76

Интересно, что агенты, стабилизирующие белки, такие как триметиламин N-оксид (TMAO), использованные в исследовании, также снижают гибкость белка.

В той же публикации [322] сделан вывод, что TMAO не связывается специфически с белком, что свидетельствует о единообразном стабилизирующем эффекте. Аналогично, исследования молекулярной динамики приводят авторов [323] к идее представить эффект стабилизации лизоцима трегалозой как образование «стеклоподобного» вещества. Рамановское рассеяние также выявило стабилизирующий эффект трегалозы на лизоцим, сохраняющий свернутый вид при более высоких температурах по сравнению с растворами чистой воды [324]. Эффект сахара становится очевидным начиная с низких концентраций сахара $\approx$20%. Молекулярная динамика [299] и эксперимент [325] демонстрируют снижение колебаний лизоцима в растворах трегалозы.

Несмотря на то что в настоящее время невозможно сделать вывод о том, что меньшая гибкость приводит к большей стабильности, наши выводы добавляют еще один пример в поддержку этой гипотезы.

Наши результаты показывают что действие трегалозы неоднородно а взаимодействия специфичны для областей между вторичными структурными элементами структуры белка. Есть одна область где белок становится более гибким при добавлении трегалозы – активный центр фермента. Функция лизоцима заключается в гидролизе гликозидных связей а структура трегалозы связана с ее природными субстратами. Молекулярно динамическое моделирование не позволяет анализировать химические реакции, однако специфическое взаимодействие трегалозы с активным центром, повышающее гибкость, позволяет предположить каталитическую функцию фермента.

4.3.4 Выводы

Приходим к выводу, что распределение трегалозы по поверхности лизоцима неравномерно и трегалоза образует кластеры на поверхности. Поэтому выводы экспериментальных и теоретических исследований, предполагающие равномерное распределение, могут быть ошибочными. Результаты можно интерпретировать как подтверждающие как гипотезу захвата воды, так и гипотезу замещения воды; структура участков трегалозы согласуется с последней, тогда как наличие кластеров воды рядом с кластерами сахара согласуется с первой. Неравномерное распределение трегалозы означает, что обе гипотезы верны, но для разных частей структуры. Из-за различной химической природы структурных элементов белков некоторые из них (около 30 %) предпочитают напрямую взаимодействовать с сахарами. Однако, несмотря на большое количество молекул трегалозы вблизи поверхности белка, для молекул воды также остается достаточно места и большая часть поверхности белка (около 70 %) остается гидратированной.

Кластеры трегалозы значительно снижают подвижность соседних аминокислот лизоцима за исключением нескольких аминокислот, расположенных близко к активному центру. При этом большая часть кластеров сосредоточена вокруг петель и менее структурированных элементов белка. Следовательно, это можно интерпретировать как то, что трегалоза оказывает наибольшее влияние на стабильность *третичной* структуры белка, а не на *вторичные* структурные элементы.

4.4 Диффузия лиганда на поверхности белка, наблюдаемая при моделировании молекулярной динамики

Малые молекулы (лиганды) взаимодействуют (связываются) с белками и изменяют их биологическую функцию иногда с желательными с медицинской точки зрения последствиями. Взаимодействие обычно описывается как химическая реакция между лигандом и белком, образующими комплекс лиганд-белок. Сила взаимодействия измеряется константой сродства обратной равновесной константе диссоциации. Последняя представляет собой отношение констант скорости диссоциации (числа комплексов диссоциирующих на компоненты за единицу времени) к скорости ассоциации (обратного процесса). Таким образом, эффективность лиганда как кандидата в лекарственное средство определяется константами процессов связывания и диссоциации. Обе константы скорости можно измерить экспериментально, однако конкретные молекулярные механизмы лучше всего выявляются с помощью компьютерного моделирования. Современное оборудование и современные методы моделирования позволяют проводить реалистичное моделирование малых белков в масштабах времени от сотен наносекунд до микросекунд, в течение которых происходят элементарные события ассоциации.

Значительные усилия были предприняты для расчета констант связывания с помощью моделирования молекулярной динамикой. Одной из самых сложных задач в этом исследо-

вании является получение кинетики связывания непосредственно из моделирования, избегая использования термодинамических приближений, которые могли бы маскировать механизмы определяющие скорости процессов [47—49].

Муавад *и др.* провели исследование путей миграции CO в цитохроме $P540_{cam}$ [326]. 37 траекторий длительностью всего 1 нс каждая оказались достаточными для идентификации нескольких промежуточных сайтов связывания лиганда при его диффузии в растворителе после диссоциации в результате флэш-фотолиза. Статистические данные были недостаточными для расчета скоростей, но достаточно хорошими для выявления метастабильных состояний и сложных перегруппировок воды внутри белка. Есть еще несколько публикаций, в которых предпринимаются попытки рассчитать скорость связывания лигандов на основе молекулярной динамики. Кондрашов *и др.* изучали высвобождение NO из нитрофорина 4 (гемового белка) [327]. Однако многочисленные допущения и короткое время моделирования не дали возможности провести осмысленное сравнение с экспериментальными константами скорости. Выход лиганда из миоглобина был промоделирован Шей [328]. Была использована сложная и довольно специализированная модель процесса, хотя рассматривались только прямые статистические данные по траекториям. В той же парадигме Камачо *и др.* сообщили о моделировании [329] выполненом с использованием броуновской динамики. Ограниченное представление атомистических движений не позволило раскрыть атомистические детали механизма связывания. Моделирование миграции CO между растворителем и местом связывания в миоглобине с атомистическими деталями, представленно Ruscio *и др.* [330]. Эти авторы оценили скорость по 90 траекториям связывания / диссоциации общей длиной 7 μс. Согласие с экспериментом было достаточно хорошим. Недавно в двух публикациях были представлены оценки констант связывания, рассчитанные непосредственно по результатам моделирования [331; 332]. Другими словами, в дополнение к моделированию начальных (несвязанных) и конечных (связанных) состояний, важно проследить полный путь лиганда между состояниями, поскольку он может содержать этапы, ограничивающие скорость процесса.

Один из таких механизмов изменения скорости связывания связан с «диффузией» лиганда по поверхности белка. Хорошо известно, что диффузия лиганда на поверхности белка ускоряет поиск лигандом места связывания [333] и, таким образом, облегчает связывание (см. обзор [334] и ссылки в нем). Доказательства, подтверждающие этот механизм, были получены для связывания белка с ДНК и связывания рецепторов с поверхностью клетки [335; 336]. Однако механизм поверхностной диффузии не обсуждался в контексте связывания белков с *малыми лигандами*. Этот процесс находится в той же концептуальной парадигме, а именно, повышение скорости ассоциации за счет диффузионного поиска в пространстве меньшей размерности (поверхность вместо объема), и как таковой может обеспечить химически важный механизм связывания. Насколько нам известно, решающий вопрос «диффундируют ли малые лиганды по поверхности при связывании с белками?» остается без ответа. Ответ не очевиден, поскольку взаимодействие небольшой, но многоатомной молекулы лиганда со сложной структурой поверхности белка, опосредованное атомистически представленными молекулами воды, является весьма запутанным. Он создает тонкий баланс сил, который может привести или не привести к эффекту притяжения лиганда к белку.

Поскольку лиганд мал, силы принципиально отличаются от уже исследованных случаев связывания белка с ДНК и клеткой.

Здесь приведены результаты *прямого* моделирования систем DHFR-TMP и DHFR-BrWR (*Дигидрофолатредуктаза микобактерии туберкулеза* в комплексе с ее кофактором NADPH связывается с триметопримом (TMP) или Br-WR9920 (BrWR), аналогом противомалярийного средства WR99210) в течение почти микросекунды с использованием полноатомной молекулярной динамики с атомистической водой. Этот белок является популярной целью для исследований антибактериальных агентов [337; 338].

Было обнаружено, что лиганды притягиваются к поверхности белка. Однако притяжение недостаточно сильное, чтобы обездвижить лиганд, поэтому лиганд может перемещаться по поверхности прежде чем отсоединится. Таким образом, моделирование наглядно демонстрирует существование давно обсуждаемого канала реакции – неспецифического связывания с последующей поверхностной диффузией к месту связывания [333; 334]. Это – главный результат нашей работы, подразумевающий облегчение процесса связывания. Кроме того, была дана грубая оценка прочности неспецифического связывания, удерживающего лиганд на поверхности. Была использована модель Берга и Перселла [333] описывающая поиск места связывания на сфере с учетом диффузии на поверхности сферы. Наши оценки в рамках этой модели показывают, что связывание действительно достаточно сильное, чтобы существенно изменить скорость связывания [50].

4.4.1 Моделируемая система

Структура несвязанного (исходного) состояния была построена на основе рентгеноструктурного анализа тройных комплексов бактерий DHFR-NADP-TMP (PDB ID: 1DG5) и DHFR-NADP-BrWR (PDB ID: 1DG7) [339]. Сообщалось [339], что наличие или отсутствие лиганда на участке связывания фолата мало влияет на общую структуру DHFR M.tuberculosis (в отличие от DHFR E.coli, для которого наблюдаются перемещения субдоменов при связывании субстрата [340]). Таким образом, молекул лиганда была помещена в 25 от кристаллографического положения в исходной конформации и NADP был заменен на NADPH. Этот комплекс был сольватирован в прямоугольной ячейке содержащей молекулы воды TIP3P [341]. Размеры ячейки были выбраны таким образом, чтобы минимальное расстояние любого атома растворенного вещества от стенок составляло 15 Å. Хотелось бы подчеркнуть, что это минимальное расстояние, тогда как в среднем лиганды могут находиться на расстоянии до $\approx$ 50 Åот поверхности белка. Перед сбором данных для анализа исходные структуры были тщательно уравновешены.

Десять отдельных симуляций различались начальным распределением импульсов всех атомов в системе. Время моделирования варьировалось от 20 нс до 100 нс, а общее время моделирования составило приблизительно 800 нс. Все моделирования проводились с использованием пакета AMBER 8.0 [342] модифицировано для MDGRAPE-3 [343; 344]. Алгоритм SHAKE [242] был применен к связям с участием атомов водорода с использованием шага времени интегрирования 2.0 фс. Дальнодействующие кулоновские взаимодействия были рассчитаны с помощью

метода Эвальда (PME) [345]. Вещественная пространственная компонента метода PME была рассчитана с использованием MDGRAPE-3, в то время как пространственная компонента волнового числа для этого метода и связанные взаимодействия были рассчитаны компьютерами-хозяевами. Для оптимизации баланса между временем расчета этих компонентов расстояние отсечки составляло 14 Å. Каждая система постепенно нагревалась до 310 K в течение первых 50 пс. Методы контроля температуры и давления Берендсена [268] использовались для поддержания постоянной температуры и давления на уровне 310 K и 1 атм. Для молекулы DHFR использовалось силовое поле AMBER ff03. Поскольку NADPH, TMP и BrWR не были включены в стандартное силовое поле ff03, их параметры силового поля были определены с использованием модуля версии 1.27 AMBER 8.0 [342] с использованием общего атомного силового поля (GAFF) [346] и атомные заряды были определены с помощью квантово-химических расчетов с использованием метода Хатри-Фока/базисного набора 6-31G**.

4.4.2 Результаты

Было обнаружено, что через несколько наносекунд после начала моделирования лиганд находит поверхность белка и слабо связывается с ней. На рисунке 4.17 показано несколько репрезентативных графиков расстояния между центрами масс лиганда и ближайшим остатком белка. Это соответствовало образованию нескольких водородных связей TMP или BrWR с поверхностью белка и отсутствию воды между лигандом и белком. Очевидно, большую часть времени лиганд проводит на поверхности белка. Он также перемещается вдоль поверхности (медленно диффундирует по ней), несколько типичных мест расположения TMP показаны на рисунке 4.17 вместе с траекторией диффузии на поверхности. Наконец, лиганд время от времени отсоединяется от белка и свободно перемещается в воде, прежде чем прикрепиться к поверхности в каком-то другом месте.

Хотелось бы подчеркнуть, что во всех десяти экспериментах места в которых лиганд связывался были разными. Таким образом, связывание было неспецифическим, в отличие от нескольких случаев связывания лиганда с экспериментально известным местом связывания, где он в конечном итоге был захвачен.

Упрощенная модель описанного взаимодействия лиганда с поверхностью может помочь оценить влияние поверхностной диффузии на скорость связывания. Предположим, что взаимодействие осуществляется через прямоугольную потенциальную яму глубиной ΔU и шириной l так, что на расстояниях от поверхности, меньших l, лиганд притягивается к поверхности [347], Рисунок 4.18. Тогда соотношение τ времени, проведенного лигандом в слое толщиной l вокруг белка, t_s, и в объеме, t_b, составляет $\tau = \frac{t_s}{t_b}$. Если система находится в равновесии, то это отношение должно быть равно отношению соответствующих объемов $\beta = \frac{sl}{V-sl}$ умноженное на фактор Больцмана: $\tau = \beta^{-\frac{\Delta U}{kT}}$, где s — площадь поверхности, а V — объем системы без учета объема

белка. Отсюда можно найти глубину потенциальной ямы:

$$\Delta U = -kT \ln \frac{\tau}{\beta}. \tag{4.2}$$

Время нахождения лиганда на поверхности зависит от l. Это происходит всегда, когда расстояние, показанное на рисунке 4.17 меньше чем l. Значения $V = 367640$ Å^3 и $s = 9020$ Å^2 привело к увеличению ΔU с $1.2kT$ до $2.1kT$, при изменении l с 7 до 10 Å. Были выбраны эти значения l несколько произвольно, основываясь на значениях d на рисунке 4.17 в моменты времени, когда расстояние d приблизительно постоянно, что соответствует расположению лиганда непосредственно на поверхности белка (состояние «поверхностной диффузии»).

Эти значения ΔU можно сравнить с теоретической моделью поиска места связывания по тому же механизму поверхностной диффузии. Эта модель разработана в классическом труде Берга и Перселла [333]. Авторы оценили поток лигандов к месту связывания, представленному небольшим поглощающим участком радиуса r_s на сферическом белке радиусом r_p. Лиганд мог искать место связывания, либо диффундируя в воде с коэффициентом диффузии D и отскакивая от поверхности, либо, кроме того, слабо прикрепляясь к поверхности и диффундируя по ней в течение некоторого времени с коэффициентом диффузии D'. Именно этот последний механизм здесь и обсуждается. Берг и Перселл показали [333] что для того, чтобы механизм поверхностной диффузии был преобладающим, потенциальная яма ΔU должна быть глубже чем

$$\Delta U > kT \ln \left(\frac{D}{D'} \frac{r_s}{\pi l} \ln \frac{r_p^2}{r_s^2} \right). \tag{4.3}$$

Наш белок можно приблизительно представить сферой радиусом $r_p = 15$ имеющий место связывания радиусом $r_s = 6.3$. Можно предположить, что соотношение $\frac{D}{D'}$ имеет порядок 10. Для этих значений и l, находящейся в том же интервале $7 - 10$, ΔU варьируется от $2.0kT$ до $1.6kT$. Эти значения фактически совпадают с нашими оценками, приведенными выше. Таким образом, это показывает, что описанный механизм (поверхностная диффузия) действительно важен для кинетики связывания.

Полученные значения ΔU используя как формулу (4.2) так и формулу (4.3) следует понимать как очень грубые оценки, в пределах допуска по порядку величины. Например, изменение в ΔU из ур. (4.2) может быть в пределах $\approx 2 - 5kT$, если площадь поверхности s изменяется на ≈ 3000 Å^2. Аналогично, тот же допуск $\approx 2 - 5kT$ получается из уравнения (4.3) если отношение констант диффузии $\frac{D}{D'}$ изменяется в диапазоне $1 - 1000$. Поэтому, даже если приведенные оценки дают верный порядок величин, их не следует воспринимать слишком буквально.

Стоит также отметить, что было обнаружено похожее поведение для другой системы. Взаимодействие трегалозы с лизоцимом демонстрирует преимущественное расположение молекул трегалозы рядом с поверхностью белка [46]. Целью исследования было изучение стабильности структуры белка, а динамика лиганда на поверхности не исследовалась.

4.4.3 Выводы

Подводя итог, было продемонстрировано существование механизма, облегчающего поиск лигандом места связывания. Механизм заключается в диффузии лиганда на поверхности белка и уже обсуждался в литературе, но не был явно продемонстрирован в реалистичных молекулярных системах. Была также приблизительно оценена прочность этого неспецифического связывания и оно было признано существенным для кинетики связывания.

4.5 Энергия связывания серии новых веществ с альбумином

Разработка новых противоопухолевых препаратов является актуальной задачей медицинской химии. Перспективным классом являются производные цианиновых красителей с кватернизованным атомом азота, проявляющие широкий спектр биологической активности [348—350]. Их цитотоксические свойства реализуются через взаимодействие с внутриклеточными мишенями [351—353]. Как и антрациклины, синтезированные производные тиазолопиримидина проявляют способность к интеркаляции в ДНК и ингибированию топоизомеразы II. Дополнительным механизмом действия является накопление в митохондриях с нарушением их мембранного потенциала [351; 352].

Важным аспектом фармакокинетики является транспорт в кровотоке. Человеческий сывороточный альбумин (ЧСА) играет ключевую роль в переносе веществ. Связывание с альбумином влияет на распределение, период полувыведения и биодоступность лекарства [354]. Поэтому, изучение афинности новых кандидатов в лекарственные молекулы к альбумину является важным этапом разработки лекарств.

В данном разделе описываны новые препараты, синтезированные нашими коллегами из Первого Санкт-Петербургского государственного медицинского университета (вещества STR1-STR4, Рисунок 4.19) с точки зрения их афинности к альбумину. Показано, что расчетные методы дают результаты, количественно воспроизводящие экспериментальные значения. В качестве основного метода моделирования были использованы расчеты абсолютной свободной энергии связывания (ABFE) с использованием подхода BAT2 демонстрирующего высокую прогностическую точность с среднеквадратичным отклонением 1.0-1.5 ккал/моль при валидации на тестовых системах включающих BRD4(2) и 2-метил-5-(метиламино)-6-фенилпиридазин-3(2H)-он, HIV-1 протеазу и N-альфа-(метоксикарбонил)-N-(3-[(2R)-морфолин-2-ил]пропил]-бета-фенил-L-фенилаланинамид [355].

Систематические названия синтезированных веществ STR1-STR4, соответственно, следующие (Рисунок 4.19):

STR1: 2-амино-5-(4-(диметиламино)стирил)-7-метил[1,3,4]тиадиазоло[3,2-a]пиримидин-4-ия перхлорат;

STR2: 2,7-диметил-5-(4-(диметиламино)стирил)-[1,3,4]тиадиазоло[3,2-a]пиримидин-4-ия перхлорат;

STR3: 5-(4-(диметиламино)стирил)-7-метил-3-фенилтиазоло[3,2-a]пиримидин-4-ия перхлорат;

STR4: 5-(4-(диметиламино)стирил)-7-метил-3-фенилтиазоло[3,2-a]пиримидин-4-ия перхлорат.

Результаты моделированния сравнивали с экспериментальными данными, полученными в той же лаборатории методом калориметрического титрования.

4.5.1 Калориметрическое титрование

Взаимодействие исследуемых соединений STR1-STR4 с человеческим сывороточным альбумином изучали в строго контролируемых изотермических условиях при температуре 298.15 K, используя высокоточный метод изотермического калориметрического титрования. Эксперименты проводились на современной микрокалориметрической системе Nano ITC 2G производства компании TA Instruments. Особенностью данного прибора является использование прецизионной золотой измерительной ячейки объемом 1 мл, что обеспечивает высокую чувствительность и точность регистрации тепловых эффектов, возникающих при молекулярном взаимодействии лигандов с белковой мишенью [356]. Параметры взаимодействия веществ были рассчитаны с использованием термодинамической модели независимого связывания.

4.5.2 Подготовка белковых структур BAT2

Известно шесть сайтов связывания ЧСА с различными лигандами (IA, IIA, IIIA, IB, IIB, IIIB), Рисунок 4.20. Механизм взаимодействия лиганда с ЧСА может включать конформационные изменения в белке. Этот факт усложняет расчет свободной энергии связывания как меру аффинности лиганда, приводя к существенным ошибкам при использовании методов, предполагающих жесткую структуру белка (к коротым относятся, например, большинство методов докинга). Этот же факт предопределил наш выбор метода BAT2, основанного на стандартных МД расчетах, т.е. естественным образом учитывающим конформационную гибкость мишени.

В данном исследовании было изучено связывание лигандов с главным сайтом связывания IIA. Этот сайт связывания был выбран по нескольким причинам: (i) Сайт IIA характеризуется оптимальным сочетанием гидрофобных и полярных остатков, что обеспечивает возможность формирования разнообразных типов взаимодействий (π-катионные, ван-дер-ваальсовы, водородные связи); (ii) многочисленные исследования подтверждают, что сайт IIA альбумина является основным сайтом связывания для многих классов лекарственных соединений, что позволяет проводить корректное сравнение с литературными данными; (iii) проведение расчетов молекулярной дина-

мики для всех потенциальных сайтов связывания потребовало бы непропорционально больших вычислительных ресурсов без существенного прироста информативности.

Было вручную выбираны три атома в ЧСА (обозначаемые P1, P2, P3) используя программный пакет BAT2 [355], которые служили точкой привязки для определения положения и ориентации лиганда в сайте связывания с помощью Boresch-style restraints. Основными критериями выбора в нашем исследовании являются следующие:

1. атомы должны находиться в жёстких областях белка, предпочтительно в альфа-спиралях или бета-листах, чтобы минимизировать нежелательные конформационные изменения;
2. атомы должны быть частью белкового остова;
3. необходимо избегать геометрических вырожденностей: расстояния между атомами не должны быть слишком короткими, а углы между векторами P1-P2 и P1-P3 не должны быть близки к 0° или 180°.

Далее задавалась сферическая зона поиска лиганда вокруг точки, координаты которой задаются относительно атома P1 в декартовой системе отсчёта. Эта зона используется для проверки, остался ли лиганд в сайте связывания после этапа уравновешивания системы. Для корректной работы алгоритма автоматического выбора якорных атомов лиганда и определения зоны поиска, BAT2 требует, чтобы белок был ориентирован в пространстве единообразно для всех расчётов. Для этого поданная на вход структура белка автоматически выравнивается алгоритмом BAT2 на оптимизированную структуру с помощью утилиты lovoalign. Это гарантирует, что декартовы координаты зоны поиска и якорных атомов остаются последовательными для всех лигандов, нацеленных на один и тот же сайт связывания. BAT2 использует стандартные шаблоны протонирования для остатков белка из выбранного силового поля. Протонирование групп обычно задаётся исходя из стандартных значений pKa для pH = 7.0. Затем было выбрано силовое поле для белка OpenMM и модель воды TIP3P через входной файл BAT2. BAT2 с помощью утилиты tleap из пакета AMBERTools помещает ЧСА в систему с водой. Был выбран буферный зазор по осям XYZ после чего в систему автоматически добавлялись ионы для достижения электронейтральности.

4.5.3 Подготовка структур лиганда

Лиганды STR1, STR2, STR3, STR4 использовались с предварительно добавленными атомами водорода. Требовалось задать специфическое состояние ионизации, которое автоматические алгоритмы могут определить неверно. В этом случае было необходимо вручную указать общий заряд лиганда во входном файле BAT2. Было сгенерировали параметры лиганда с помощью инструментов Antechamber (из пакета AMBER) и предоставлены BAT2 готовые файлы с частичными зарядами и с параметрами связей, углов и двугранных углов. После того как система белок-лиганд была выровнена по оптимизированной структуре, BAT2 автоматически выбирал три атома лиганда (L1, L2, L3), которые были использованы для применения трансляционно-ротационных (TR) ограничений относительно якорных атомов белка. Сначала программа выбирала атом (L1) лиганда, ближайший к заранее заданному центру сайта связывания. Затем определяла атом (L2), для

которого угол P1-L1-L2 был максимально близок к 90°, а расстояние L1-L2 находилось в заданном пользователем диапазоне. Третий атом (L3) выбирался аналогично L2, с учётом угла L1-L2-L3 и расстояния L2-L3.

4.5.4 Прогоны молекулярной динамики (МД) для расчёта ABFE в BAT2

Сначала лиганды автоматически подгонялись в сайт связывания белка с помощью докинга, с использованием программы AutoDock Vina. Это генерировало несколько (от 1 до 10) различных начальных поз. Затем система была уравновешена, чтобы позволить сгенерированной оболочке и боковым цепям белка адаптироваться к присутствию лиганда, а также проверить, стабильна ли начальная поза лиганда.

После сольватации и добавления ионов, лигандов и белка BAT2 применяет к системе три типа ограничений:

1. COM-ограничения (Center of Mass): удерживают центр масс белка в центре системы, позволяя ему свободно вращаться;
2. конформационные ограничения белка: могут быть применены к двугранным углам остова в жёстких регионах белка для ускорения сходимости;
3. TR-ограничения (Translational/Rotational) и конформационные ограничения лиганда: жёстко фиксируют положение и ориентацию лиганда относительно якорных атомов белка, а также его внутреннюю конформацию.

Далее система подвергается энергетической минимизации для устранения стерических перекрываний. Система постепенно нагревается от низкой до целевой температуры в течение 100 пс с использованием термостата Ланжевена. Короткие прогоны по 15 пс проводятся для поднятия давления системы до 1 атм с использованием баростата Монте-Карло. Проводится серия коротких МД-прогонов, в ходе которых ограничения на лиганд (как TR, так и конформационные) постепенно и последовательно ослабляются до нуля. Это позволяет системе плавно релаксировать вокруг лиганда. После полного снятия ограничений с лиганда проводится более длительная симуляция (длительность задаётся пользователем), в течение которой лиганд может найти устойчивое положение в сайте связывания. По окончании этапа BAT2 проверяет, находится ли хотя бы один атом лиганда внутри заданной сферической зоны поиска сайта связывания. Если лиганд покинул сайт, эта поза помечается как нестабильная, и расчёт свободной энергии для неё не проводится. Если поза признана стабильной после уравновешивания, система перестраивается на основе уравновешенной конфигурации. При этом заново определяются якорные атомы лиганда и значения для ограничений, чтобы они соответствовали новой конформации.

Для каждого из выбранных компонентов [355] составляющих ΔG^0, запускается свой набор независимых МД-прогонов (окон). Каждое окно соответствует определённому значению параметра связи. Для каждого значения параметра связи (λ) проводится короткий уравновешивающий прогон для адаптации системы. Затем проводится основной прогон, в ходе которго траектория сохраняется для последующего анализа. После завершения всех прогонов BAT2 автоматически

анализирует данные с помощью выбранного метода MBAR (Multistate Bennett Acceptance Ratio), вычисляет вклад каждого компонента и их сумму (ΔG^0).

После этого было использовано уравнение (4.4) которое представляет собой наиболее общее выражение для расчёта свободной энергии связывания, учитывающее возможность существования нескольких стабильных поз лиганда в сайте белка [357]:

$$\Delta G^0_{bind} = -RT \ln \sum_{i}^{N_{poses}} e^{-\beta \Delta G^0_i}, \tag{4.4}$$

где ΔG^0_{bind} - стандартная абсолютная свободная энергия связывания, R - универсальная газовая постоянная, T - абсолютная температура, $\beta = 1/RT$, N_{poses} - количество всех стабильных, неперекрывающихся поз лиганда, ΔG^0_i - стандартная свободная энергия связывания, рассчитанная для конкретной позы i.

4.5.5 Результаты

На рисунке 4.20 представлена схема расположения веществ STR1-STR4 в сайтах связывания ЧСА по результатам начального докинга. Анализ результатов определения аффинности показал, что все вещества связываются с главным сайтом связывания (IIA) в белке. Это можно объяснить определённым расположением молекулы, что требует значительной деформации белка для десорбции данных соединений, а также наличием водородных связей между компонентами. Молекулы демонстрируют схожее поведение, за исключением соединения 4, где не образовалось достаточно связей для удержания молекулы в сайте, что также подтвердилось отсутствием экспериментального значения для этого соединения. Результаты моделирования, начатого с исходных конфигураций с расположением соединений вблизи поверхности ЧСА, показывают важность водородных связи при связывании веществ STR1-STR4. В таблице 12 представленны результаты энергии связывания полученные в программе BAT2.

Экспериментальные термодинамические параметры взаимодействия веществ STR1-STR4 с ЧСА были рассчитаны с использованием термодинамической модели независимого связывания и представленны в таблице 13. Сравнение этих значений со значениями, полученными в результаты моделирования показывает, что они согласуются с расхождением в 1-6 кДж/моль, что говорит о достаточной точности результатов [354].

4.5.6 Выводы

Моделирование для определения аффинности с расчетом абсолютной свободной энергии связывания (ABFE) с использованием пакета BAT2 подтвердило стабильность комплексов веществ STR1-STR3 с основным сайтом IIA ЧСА в течение 20 нс. Расчетные значения свободной

Таблица 12 — Результаты расчетов энергии связывания лигандов с ЧСА в сайте IIA

Соединение	Номер конформации	Энергия связывания BAT2 в различных конформациях, кДж/моль (ΔG_i^0)	Энергия связывания полученная калориметрическим титрованием, кДж/моль	Энергия связывания полученная по распределению Больцмана, кДж/моль (ΔG_{bind}^0)
STR1	1	-31	-38.97	-38.409
	2	-19,96		
	3	-37,32		
	4	-24,31		
	5	-19,71		
	6	-24,23		
	7	-16,40		
	8	-35,40		
	9	-17,99		
STR2	1	-34,35	-37.75	-43.698
	2	-43,64		
STR3	1	-18,24	-40.21	-39.114
	2	-38,95		
	3	-25,86		
	4	-27,57		
	5	-20,25		
	6	-17,70		
	7	-31,63		
STR4	1	0,59	-	-24.316

энергии связывания продемонстрировали хорошую корреляцию с экспериментальными данными, с расхождением в пределах 1–6 кДж/моль, что свидетельствует о высокой прогностической точности примененного вычислительного подхода.

Экспериментальные исследования методом калориметрческого титрования (ITC) количественно подтвердили высокое сродство веществ STR1-STR3 к ЧСА с константами связывания (K_a) порядка $10^7 - 10^8$ M^{-1}. Термодинамический анализ показал, что связывание для веществ 1 и 2 является энтальпийно выгодным, тогда как для вещества 3 движущей силой является благоприятное энтропийное изменение. Для вещества 4 значимое связывание с ЧСА экспериментально не обнаружено, что согласуется с результатами докинга.

Таким образом, установлено, что синтезированные производные тиазолопиримидина STR1-STR3 обладают высокой афинностью к ЧСА, что указывает на их значительный потенциал в качестве кандидатов для дальнейшей фармакокинетической оптимизации и доклинических исследований в качестве перспективных противоопухолевых агентов. Комплексный подход, со-

Таблица 13 — Термодинамические параметры связывания веществ STR1-STR4 с ЧСА. ΔH, ΔS, ΔG – изменения энтальпии, энтропии и энергии Гиббса, K_a – константа связывания, n – стехиометрический коэффициент связывания

Термодинамические параметры связывания	STR1	STR2	STR3	STR4
K_a, моль	$0.71 \cdot 10^7$	$0.41 \cdot 10^7$	$0.1 \cdot 10^8$	-
n	1.39	1.2	1.03	-
ΔH, кДж/моль	-33.83	-42.93	-23.28	-
ΔG, кДж/моль	-38.97	-37.75	-40.21	-
ΔS, кДж/моль/К	17.24	17.39	56.74	-

четающий вычислительные и экспериментальные методы, доказал свою эффективность для прогнозирования и анализа белок-лигандных взаимодействий.

4.6 Константа скорости диссоциации белок-лиганд, полученная с помощью полноатомного моделирования

Сродство соединения к связыванию, количественно определяемое константой диссоциации K_D, является ключевым свойством молекулы соединения для разработки лекарственных препаратов. K_D определяется как отношение констант скорости процессов диссоциации и ассоциации в системе белок-лиганд, $K_D = \frac{k_{\text{off}}}{k_{\text{on}}}$, где $k_{\text{off(on)}}$ — константа скорости диссоциации (ассоциации), а $\tau_{\text{off(on)}} = 1/k_{\text{off(on)}}$ — время диссоциации (ассоциации).

Расчет скоростей ассоциации и диссоциации с использованием молекулярного моделирования является активной областью исследований, см. последние обзоры в [358—360]. Более того, показано, что кинетические свойства, а не K_D, лучше коррелируют с экспериментальной эффективностью препарата [359—361].

Моделирование полноатомной молекулярной динамики в большинстве случаев не позволяет напрямую рассчитать кинетику ассоциации и диссоциации белок-лиганд, поскольку экспериментальные значения находятся в диапазоне времен порядка секунд, что на много порядков больше, чем доступные в настоящее время времена моделирования прямой MD. Это особенно актуально для времени диссоциации, поскольку оно намного больше времени ассоциации для потенциальных лекарств (что делает их хорошими кандидатами). Поэтому применяется ряд методов оценки скоростей диссоциации и выяснения механизмов диссоциации с использованием нано- и микросекундного моделирования [362]. Несмотря на недавние успехи, рассчитанные скорости диссоциации воспроизводят экспериментальные значения «в пределах множителя 2-20» [363],

они “все еще намного меньше экспериментального значения” (на 2 порядка величины) [364], или даже они соответствуют экспериментальным значениям с погрешностью “до 4 порядков” [365].

Это несоответствие в последнее время [360] является одной из основных трудностей, с которыми сталкивается подход MD к связыванию белка с лигандом, подчеркивая, что основным фактором, влияющим на столь большое отклонение, является «экстраполяция методами моделирования для вывода долговременных характеристик».

В данной работе был использован метод случайного ускорения MD (RAMD) в его варианте, называемом τ-RAMD [366; 367] для получения времен диссоциации лиганда изониазида, диссоциирующего от белка каталазы. Изониазид является основным препаратом для лечения туберкулеза, который воздействует на каталазу, жизненно важный белок для функционирования микобактерии туберкулеза [368]. Идея метода заключается в приложении к лиганду небольшой силы, сохраняя величину силы постоянной, но периодически и случайным образом изменяя ее направление с целью ускорения процесса отсоединения. Моделирование останавливается, когда лиганд достигает заданного расстояния от активного центра, в этот момент он считается диссоциированным и регистрируется время диссоциации. Этот подход в определенной степени аналогичен сопряжению пространственного разделения и скорости химической ассоциации/диссоциации, обсуждаемому в [369] но обеспечивает более простую реализацию, необходимую для полностью атомистического подхода MD. В результате τ-RAMD предоставляет набор времен диссоциации в зависимости от величины приложенной силы.

Используя последние результаты стохастической теории скоростей реакций, показано, что моделированные данные могут быть использованы для оценки времени диссоциации, которое количественно соответствует экспериментальному значению $\tau_{\text{off}}^{exp} = 50$ секунд [51].

4.6.1 Теория

Компьютерный эксперимент, реализованный с использованием τ-RAMD, генерирует данные в виде количества лигандов, которые остаются связанными с белком в момент времени t: $N(t)$. Нормализовав к 1 в нулевой момент времени, получаем вероятность выживания $\frac{N(t)}{N(0)}$ лиганда, связанного с белком в момент времени t. Поэтому первым вопросом для теории будет “Как определить время диссоциации τ_{off} исходя из вероятности выживания $\frac{N(t)}{N(0)}$ для осмысленного сравнения с экспериментально измеренными τ_{off}^{exp}?”. Поскольку это значение τ_{off}, полученное при моделировании, зависит от искусственно приложенной силы f, вторым вопросом будет “как экстраполировать моделируемое время диссоциации на нулевую силу для сравнения с реальным экспериментом?”. Как показывают наши данные, ответы на оба вопроса требуют нетривиальных физических подходов.

Определение τ_{off}

Возникновение стохастической теории скоростей реакций, включая процессы диссоциации, восходит к основополагающей работе Х.А. Крамерса [370] который рассматривал микроскопическое происхождение макроскопических процессов химической кинетики как случайное движение термически активированных частиц пересекающих потенциальный барьер. Дальнейшее развитие этой теории можно найти в подробных обзорах [371—374].

В простейшем случае классической кинетики Крамерса плотность вероятности пространственно-временного распределения случайных частиц удовлетворяет уравнению Фоккера-Планка. Интегриривания его решения в пределах барьера дает желаемую временную эволюцию вероятности выживания $\frac{N(t)}{N(0)}$ спадающую как экспоненциальная функция; соответственно, τ_{off} определяется как величина обратная экспоненциальному коэффициенту.

Однако существует также вероятность аномальной кинетики, которая приводит к дробному уравнению Фоккера-Планка с неэкспоненциальным поведением релаксации [375]. Следует отметить, что такие неэкспоненциальные, так называемые «неспектральные», моды могут проявляться даже в случае классического уравнения Фоккера-Планка, когда начальные условия берутся из широких, сильно нестационарных начальных плотностей вероятности [376]. В этом случае можно выявить двухэтапный процесс релаксации, когда в ходе временной эволюции динамики системы ведущий степенной режим сменяется обычным спектральным (экспоненциальным) режимом релаксации [377].

Подводя итог, можно сказать, что динамика вероятности диссоциации лиганда от белка может иметь два режима:

(I) неэкспоненциальный на малых временах, вызванный неравновесными начальными условиями, возникающими из-за сложных межмолекулярных взаимодействий в системе под влиянием приложенной внешние силы τ-RAMD,

(II) классическая экспоненциальная релаксация в течение более длительного времени, когда указанные выше начальные условия уравновешиваются.

Начальный неэкспоненциальный режим является кратковременным и, таким образом, не обнаруживается экспериментом. Поэтому можно предположить, что τ_{off} определяется вторым, гораздо более продолжительным режимом и равен обратной величине своего экспоненциального коэффициента. Далее для краткости будем использовать τ_{off} и τ как синонимы.

Зависимость времени диссоциации τ от приложенной силы

Существует несколько работ по исследоавнию снижения потенциального барьера диссоциации либо под влиянием дополнительных приложенных сил, либо за счет активации скорости частиц в контексте диссоциации или времени первого прохождения из потенциальной ямы

[378—381]. Однако эти модели довольно абстрактны и имеют дело с искусственным численным моделированием, а не с реальными биофизическими системами.

Насколько нам известно, первая попытка учесть влияние внешней силы f на связь рецептор-лиганд была предложена Г.И. Беллом [382], который рассматривал характерное время жизни ассоциированного состояния τ в простейшей форме

$$\tau = \nu_0 \exp\left[\left(E_0 - \gamma f\right)/k_B T\right], \tag{4.5}$$

где ν_0 — функция собственной частоты колебаний системы в связанном состоянии, что соответствует стандартной теории Крамерса. Соответственно, когда $f \to 0$, значение $\tau(f = 0)$ сводится к обратной константе диссоциации Крамерса для невозмущенной системы. E_0 — энергия связи, γ — некоторый феноменологический параметр, а k_B и T — постоянная Больцмана и температура системы.

Обратите внимание, что уравнение (4.5) можно формально рассматривать как решение обыкновенного дифференциального уравнения

$$\frac{d\tau}{df} = -\frac{\gamma}{k_B T}\tau. \tag{4.6}$$

С простейшей точки зрения размерного анализа параметр γ имеет смысл некоторой характерной длины, которую частица должна преодолеть под действием силы, которую можно рассматривать как работу, уменьшающую начальную свободную энергию барьера/связи. Очевидно, что эта работа и, следовательно, γ зависят от силы f. Мы предлагаем модель, предполагающую, что эта характерная «длина» уменьшается с силой тем же больцмановским образом:

$$\gamma = \gamma_0 \exp\left[\left(-\gamma' f\right)/k_B T\right]. \tag{4.7}$$

Аналогичная линия рассуждений использовалась для описания так называемых «связей-зацепок», прочность которых увеличивается с ростом приложенной силы [383; 384]. Таким образом, можно предположить, что данное поведение носит универсальный характер для сложных биологических молекул.

Подставляя уравнение (4.7) в уравнение (4.6), получаем дифференциальное уравнение

$$\frac{d\tau}{df} = -\frac{\gamma_0}{k_B T} e^{-\frac{\gamma' f}{k_B T}}\tau, \tag{4.8}$$

которое легко решается методом разделения переменных:

$$\tau = \tau_0 e^{\frac{\gamma_0}{\gamma'}\left(e^{-\frac{\gamma' f}{k_B T}} - 1\right)}, \tag{4.9}$$

где $\tau_0 = \nu_0 \exp\left[-E_0/k_B T\right]$ — время диссоциации для невозмущенной системы. Легко видеть, что τ_0 равно решению (4.9) при $f = 0$. В направлении больших сил решение (4.9) асимптотически стремится к $\tau_\infty = \tau_0 \exp\left(-\gamma_0/\gamma'\right)$. Она имеет конечное значение, что согласуется со стохастическим характером модели, поскольку даже если приложенная сила полностью разрушает барьер, частице требуется некоторое время, чтобы покинуть окрестность своего начального положения посредством случайного блуждания. В то же время, $\tau_\infty << \tau_0$ в нескольких порядках величины,

т.е. $\gamma' << \gamma_0$. Отметим также, что для слабых сил возмущения $\gamma' f / k_B T << 1$, решение (4.9) сводится к выражению Белла (4.5) с $\gamma = \gamma_0$.

Уравнение (4.9) можно линеаризовать как

$$\ln\left(\ln\left(\frac{\tau}{\tau_0} e^{\frac{\gamma_0}{\gamma'}}\right)\right) = \ln\left(\frac{\gamma_0}{\gamma'}\right) - \frac{\gamma'}{kT} f. \tag{4.10}$$

Это выражение содержит безразмерные и строго положительные аргументы логарифмов, но они содержат неизвестные параметры, недоступные при прямых измерениях или моделировании. Отсюда, ур. (4.10) играет роль качественного аргумента, демонстрирующего возможное происхождение функциональной зависимости в виде дважды логарифмической зависимости времени выхода от приложенной силы. Так как ур. (4.10) содержит комбинацию феноменологических параметров, удобнее применить некоторое перемасштабирование, чтобы получить более простое выражение для дальнейшего анализа моделируемых данных.

Изменив масштаб времени τ с помощью константы $\frac{\gamma_0/\gamma'}{\tau_0}$, $\tilde{\tau} = \tau \frac{\gamma_0/\gamma'}{\tau_0}$ получаем $\tilde{\tau}_0 = \tau_0 \frac{\gamma_0/\gamma'}{\tau_0} = e^{\gamma_0/\gamma'}$ из которого $\gamma_0/\gamma' = \ln(\tilde{\tau}_0)$, $\ln\left(\frac{\gamma_0}{\gamma'}\right) = \ln\left(\ln\left(\tilde{\tau}_0\right)\right)$ и уравнение (4.10) становится

$$\ln\left(\ln\left(\tilde{\tau}\right)\right) = \ln\left(\ln\left(\tilde{\tau}_0\right)\right) - \frac{\gamma'}{kT} f, \tag{4.11}$$

обеспечивая линейную зависимость между двойным логарифмом перемасштабированного времени диссоциации и приложенной силой. Очевидно, что при $f = 0$ время диссоциации τ равно τ_0 для немасштабированного времени диссоциации.

4.6.2 Молекулярная модель и детали моделирования

Каталаза из микобактерии туберкулеза (*Mt*KatG) является мишенью для изониазида. Однако экспериментальных атомистических данных, позволяющих установить начальную структуру комплекса для MD, нет. К счастью, каталаза микобактерии туберкулеза (MtKatG) и каталаза burkholderia pseudomallei (BpKatG) имеют очень похожую атомистическую структуру и активность против изониазида [385]. Поскольку экспериментальная структура комплекса MtKatG-изониазид не имеется в банке данных белков [386], атомные координаты были получены путем наложения кристаллических структур MtKatG (PDB: 1sj2) и комплекса BpKatG-INH (PDB: 5syi) с использованием UCSF Chimera [387]. Подробная информация о моделировании молекулярной динамики представлена в приложении И.1.

Несколько расчетов τRAMD проводилось путем приложения к лиганду различных сил: 550, 500, 450, 400, 350, 300 и 250 $\frac{\text{kJ}}{\text{mol·nm}}$. Если расстояние между центрами масс лиганда и белка изменялось менее чем на 0.025 нм, направление силы изменялось случайным образом. Максимальное расстояние ЦОМ, на котором лиганд гарантированно покидал поверхность белка, было установлено равным 5 нм. При каждом значении силы было выполнено несколько запусков, $N(0)$ (до 200), с использованием идентичных начальных координат и скоростей, при этом единственным

отличающимся параметром было случайное начальное число для генерации случайной силы. Поскольку изониазид является небольшим лигандом, а его конформация и положение в активном центре практически не меняются со временем, варьирование связанного состояния (т.е. получение нескольких исходных структур) не была необходимой и не влияла на конечный результат (время диссоциации).

Обработка данных

Для каждого значения силы набор времен диссоциации был пересчитан в зависимости вероятности выживания от времени моделирования. Для момента времени t количество $N(t)$ рассчитывалось как общее число комплексов, которые не диссоциировали к этому моменту времени. Это равно количеству τ-RAMD выполненных в наборе (для заданного значения f), длительность которого больше, чем t. Чтобы получить вероятность выживания, $N(t)$ затем делили на общее количество запусков $N(0)$ в наборе, t варьировалось от нуля до продолжительности самого длинного запуска в наборе, а $N(t)/N(0)$ изменялось от единицы при $t = 0$ до нуля при последнем значении t, Рисунок 4.21.

Получение времени диссоциации

Очевидно, что графики вероятностей выживания $N(t)/N(0)$ для каждой силы демонстрируют два режима, Рисунок 4.21. На первом этапе распад следует колоколообразной кривой, которую можно описать как $\ln[-\ln(N(t)/N(0))] = p\ln(t) + p\ln(t_0)$, где p - показатель степени, а t_0 - некоторое характерное время, приводящее к выявленной временной зависимости $N(t) = N_0\exp(-(t/t_0)^p)$ показано черной пунктирной кривой на рисунке 4.21. Для этих двух примеров p равны 1.4 и 1.6.

Однако через некоторое время τ_{frac} вероятность выживания демонстрирует резкое изменение динамики, начиная следовать линейной зависимости $\ln(N(t)/N_0)$ от t, что соответствует обычному процессу релаксации $\frac{dN(t)}{dt} = -\lambda N(t)$ со скоростью распада λ определяющей время диссоциации τ. К концу экспоненциального распада оставшиеся долгоживущие комплексы образуют «полки» с постоянными значениями N, которые искажают наклон аппроксимирующей прямой. Эти «полки» были образованы очень небольшим числом недиссоциирующих комплексов со скачкообразными изменениями, далекими от непрерывной зависимости модели подгонки данных. Для определения порога статистически значимых данных и получения надежного соответствия значения в конце отсекались одно за другим до тех пор, пока наклон не переставал меняться (см. Рисунок И.1 в приложении для репрезентативного примера). В некоторых случаях

(как на рисунке 4.21 (a)) все значения были сохранены для подгонки, поскольку отсечение конечных точек не изменяло наклон.

Время диссоциации обратно пропорционально скорости, $\tau = \frac{1}{\lambda}$, для всех сил, перечисленных в таблице 14 а также времена перехода между двумя режимами. Обратите внимание, что значения τ и τ_{frac} близки друг к другу, что подтверждает интерпретацию начального режима как существенно неравновесных переходных процессов, происходящих за времена, меньшие характерного времени релаксации системы. Поэтому он был исключен из дальнейшего анализа. Подогнанные кривые для всех значений силы f приведены в приложении И.2.

Экстраполяция к нулевой силе

Зависимость полученных значений времени диссоциации τ от приложенной силы f на моль была приведена к линеаризованной форме путем последовательного дважды логарифмического преобразования, как показано на рисунке 4.22. Очевидная линейность в зависимости от f подтверждает теоретическую модель (4.11). Линейная аппроксимация этих значений

$$\ln\left(\ln(\tilde{\tau})\right) = \ln\left(\ln(\tilde{\tau}_0)\right) - \kappa f \tag{4.12}$$

была выполнена с использованием стандартного набора инструментов Curve Fitting Toolbox пакета MATLAB, который использует алгоритм QR-факторизации. Обратите внимание, что были использованы размерные времена (пс), полученные в результате процедуры обработки данных для подгонки, чтобы избежать ненужных осложнений с множественными параметрами, введенными при рассмотрении возможной теоретической модели, что приводит к такой двойной логарифмической функциональной форме. Поскольку нас интересует только величина τ_0, этот вид подгонки напрямую дает желаемый параметр так как исходный τ_0 совпадает с масштабированным $\tilde{\tau}_0$.

4.6.3 Результаты и обсуждение

Подгонка данных из таблицы 14 используя уравнение (4.12) приводит к $R^2 = 0.978$ и $RMSE = 0.073$ для выбранных масштабированных единиц.

Эта процедура подгонки дает среднее значение наклона, равное $\kappa = 0.0041$ с доверительными интервалами от 0.0038 до 0.0044 на уровне стандартного отклонения. Второй параметр подогнанной прямой линии (4.12) имеет среднее значение $\ln\left(\ln(\tau_0)\right) = 3.441$ с доверительным интервалом от 3.315 до 3.563 на уровне стандартного отклонения. Числовые значения соответствуют пикосекундам как размерности времени.

Рассчитанный $\ln\left(\ln(\tau_0)\right)$ предполагает нормальное распределение вокруг найденного среднего значения. Однако, возводя его в степень дважды для полученя τ_0 существенно меняет тип

Таблица 14 — Время диссоциации τ и моменты перехода от дробно-экспоненциального к классическому режиму релаксации τ_{frac}

f,кДж/(нм · мол)	τ, пс	τ_{frac}, пс
250	78397	70000
300	11044	10000
350	1319	1300
400	259	300
450	98	106
500	87	150
550	22	24

распределения вероятностей и требует более сложной процедуры определения τ_0 и его неопределенности. Они были оценены с помощью машины неопределенности NIST [388] с алгоритмом Монте-Карло, моделирующим ансамбль из 10^6 реализаций. После преобразования распределение вероятностей становится сильно «длиннохвостым» и перекошенным степенным законом, что может привести к расходящимся статистическим моментам (см. Рисунок И.3 в приложении).

Для этого типа распределений надежной статистической мерой наиболее вероятного значения является медиана, которая в нашем случае была равна $M(\tau_0) = 36{,}1$ секунд. Статистически значимые отклонения от этого значения количественно определяются медианой абсолютного значения отклонений $M(|\tau_0 - (\tau_0)|)$, что в нашем случае равно 35.9 секундам, а статистически значимые значения распределены между 36.1-35.9=0.2 и 36,1+35,9=72.0 секундами.

Подводя итог, найденное экстраполированное значение τ_0 составляет 36.1 секунды, со статистически значимыми границами 0.2 и 72.0 с, что соответствует экспериментальному значению $50 \pm 8\text{s}$ количественно в пределах неопределенностей экстраполяции и эксперимента.

4.6.4 Выводы

Была применена методология τ-RAMD для получения вероятностей диссоциации лиганда от белка. Также была предложена теория для этих вероятностей, которая описывает их временную эволюцию в соответствии с двумя режимами: неэкспоненциальным для малых времен и стандартным экспоненциальным для больших времен. Были выявлены эти два режима в данных, полученных в результате моделирования. Наконец, была предложена модель, позволяющая экстраполировать полученные времена диссоциации на значение нулевой силы, количественно соответствующее экспериментально измеренному значению 50 секунд. Это контрастирует с оригинальным подходом τ-RAMD, в котором не предпринималось попыток такой экстраполяции. Важно отметить, что экстраполяция была выполнена на девять порядков величины τ, от наносекунд до секунд. Тем не менее, экстраполированное значение количественно воспроизводит экспериментальное, в отличие от большинства современных методов, описанных в литературе.

Наша модель может быть использована в других условиях, например, в иммунологии для альтернативного расчета «времени диссоциации» как отношения времени «иммунологического синапса» [389—391].

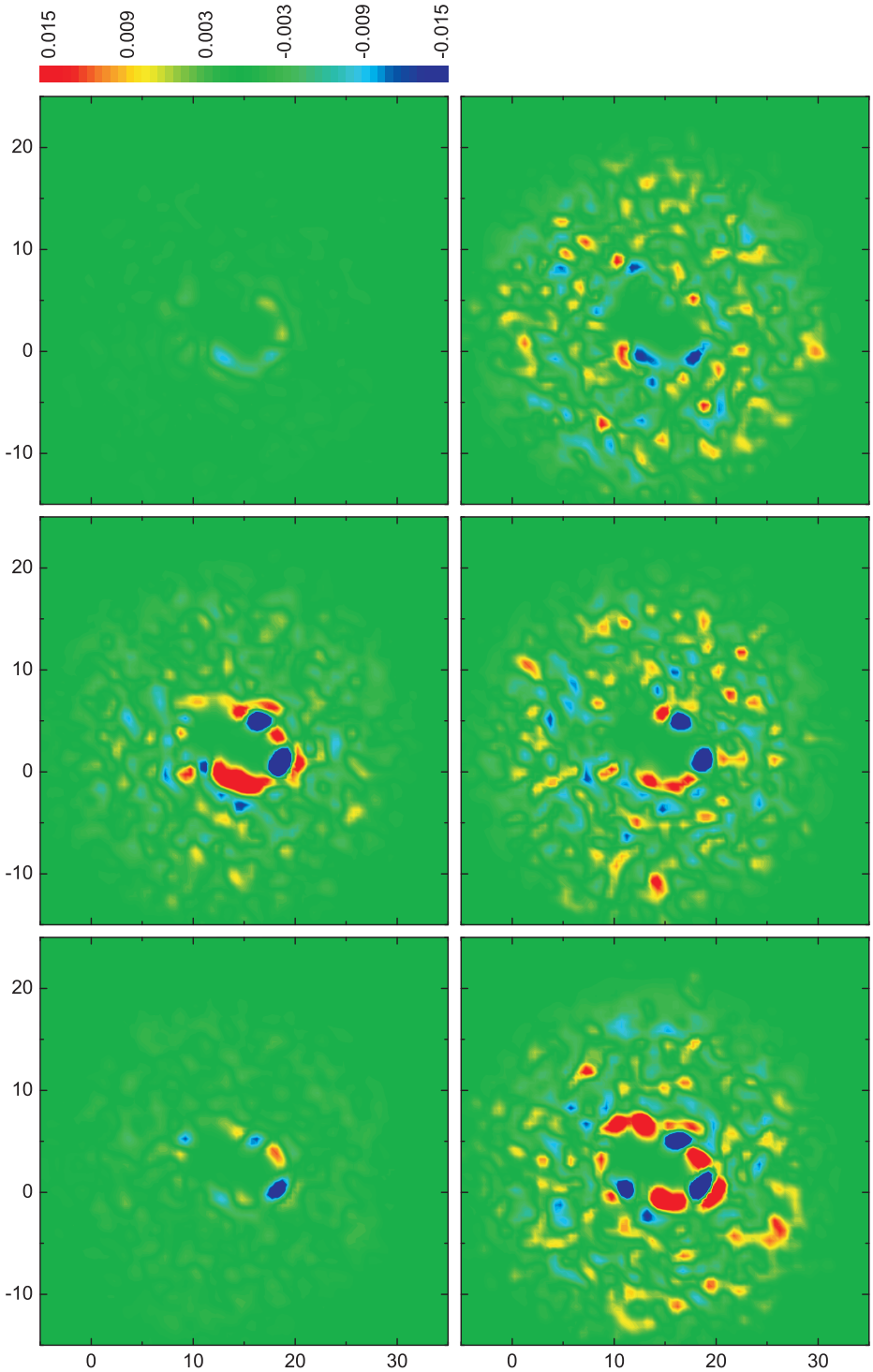


Рисунок 4.9 — Сечение xy (значение координаты z выбрано таким образом, чтобы секущая плоскость проходила через центр масс пептида; другие значения координаты z показывают качественно те же результаты) поля флуктуации вероятности $f'(x,t)$ (справа), и его часть коррелирующая с конформацией пептида $f'_c(x,t)$ (слева) для водорода; время до перехода составляет, сверху вниз, $200ps$, $0.4ps$ и $0ps$; шкала ограничена интервалом $-0.015 \ldots 0{,}015$ для наглядности, максимумы пиков достигают значений 0.04 и -0.055 (показаны результаты для модели OPLS).

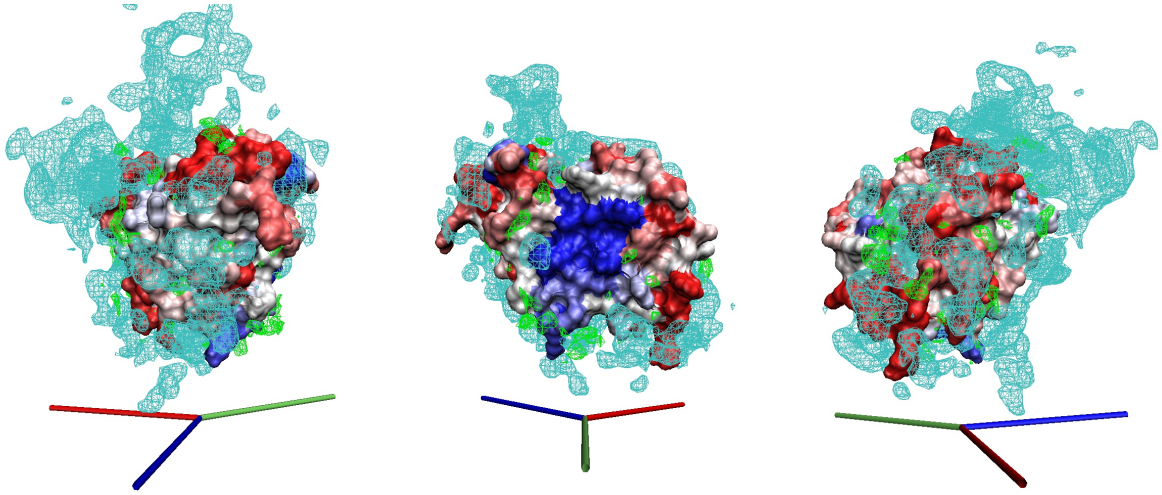


Рисунок 4.10 — Поверхность белка окрашена в соответствии с разницей в гибкости остова, вызванной трегалозой (цвета соответствуют графику разницы, показанному на рисунке 4.12); голубой: плотность трегалозы (в 6.4 раза выше, чем в основной массе); зеленый: плотность воды (в 1.5 раза выше, чем у объемной воды), показана молекулярная структура с трех разных точек зрения

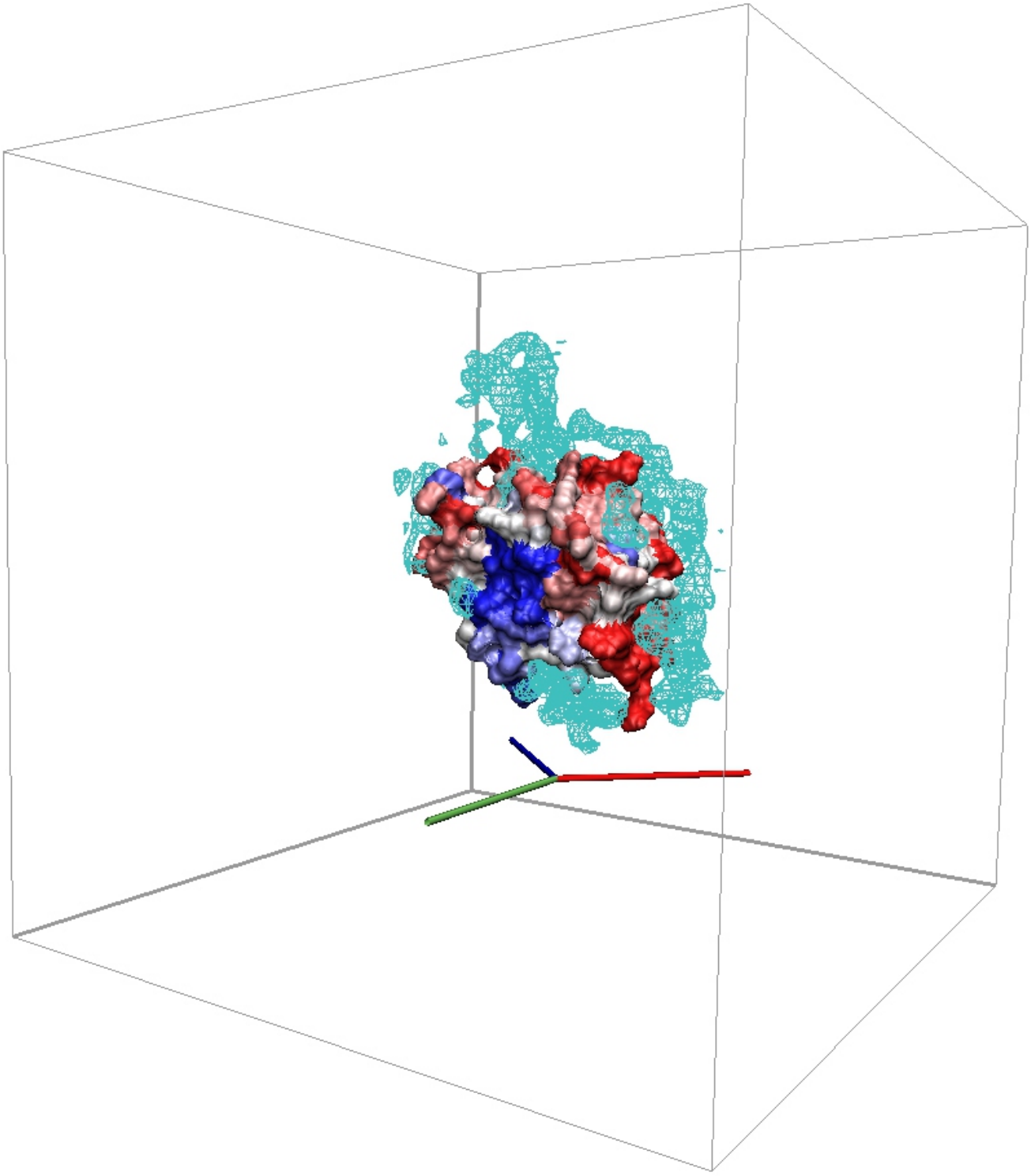


Рисунок 4.11 — Поверхность белка окрашена в соответствии с разницей в гибкости остова, вызванной трегалозой; голубой: плотность трегалозы (в 6.4 раза выше, чем в основной массе); показан коробка моделирования, подчеркивающий отсутствие сахара везде, за исключением области, прилегающей к белку.

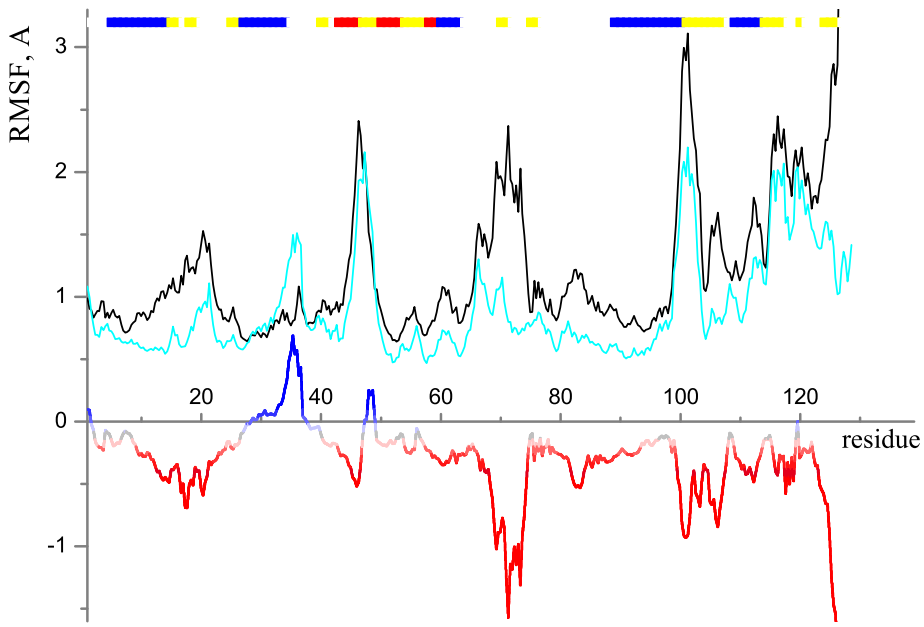


Рисунок 4.12 — RMSF атомов основной цепи белка; черный: 30 нс моделирование в растворе чистой воды; голубой: 30 нс моделирование в растворе трегалозы; многоцветная линия: разница в RMSF между растворами воды и трегалозы (голубая и черная линии), цвета используются для отображения разницы на поверхности белка, Рисунок 4.10, 4.11 и структур белка, Рисунок 4.16; структурные мотивы для каждой аминокислоты показаны в виде цветной полосы над кривыми: красная - β-лист, синий - α-спираль, желтая - петля.

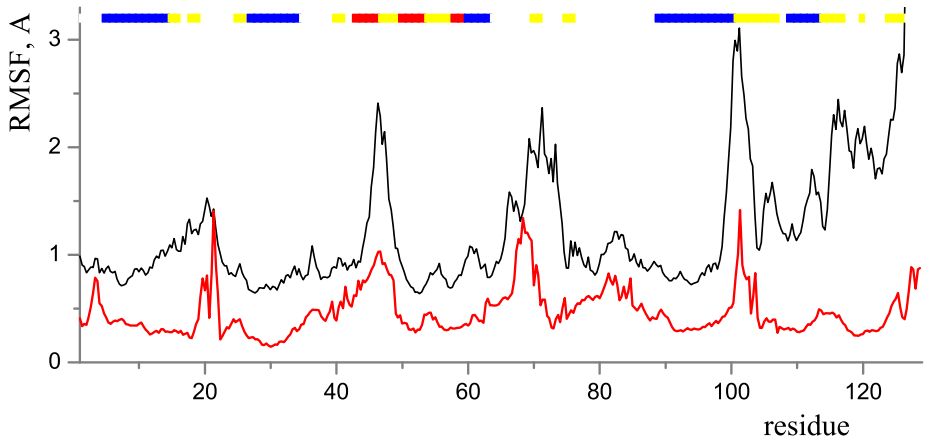


Рисунок 4.13 — RMSF атомов основной цепи белка; черный: 30 нс моделирование в чистом водном растворе; красный: 50 ЯМР-структур из [310]; структурные мотиве для каждой аминокислоты показаны в виде цветной полосы над кривыми: красная - β-лист, синий - α-спираль, желтая - петля.

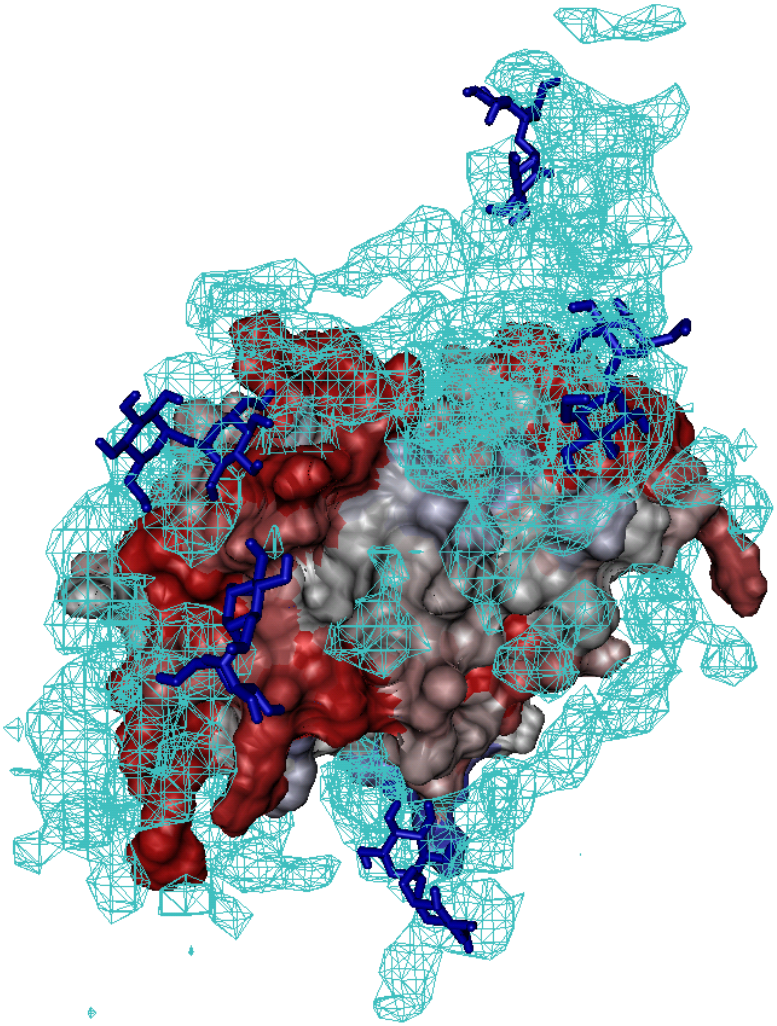


Рисунок 4.14 — Несколько случайно выбранных молекул сахара показаны для сравнения размера участков сахара высокой плотности с размером молекул сахара.

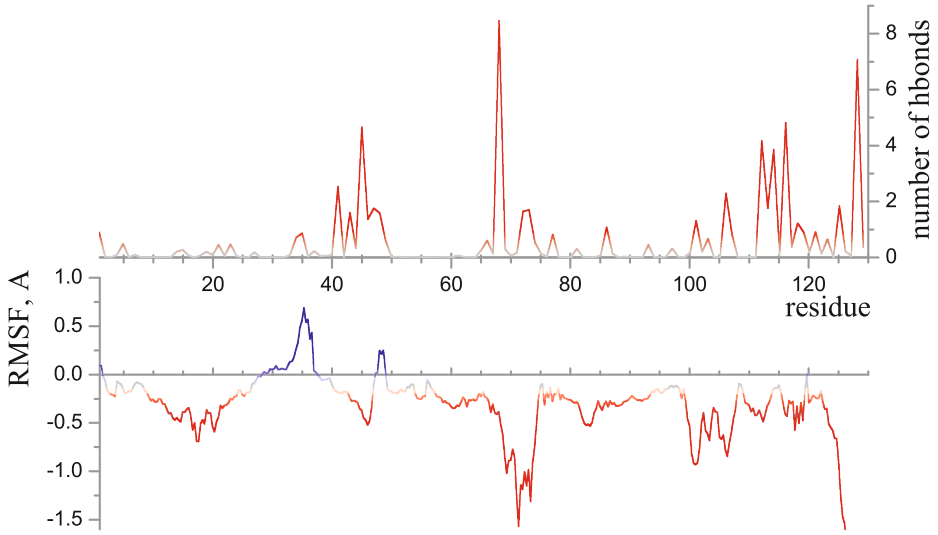


Рисунок 4.15 — Среднее количество водородных связей лизоцима и трегалозы на аминокислоту; также показана разница в гибкости аминокислот (такая же, как на рисунке 4.12).

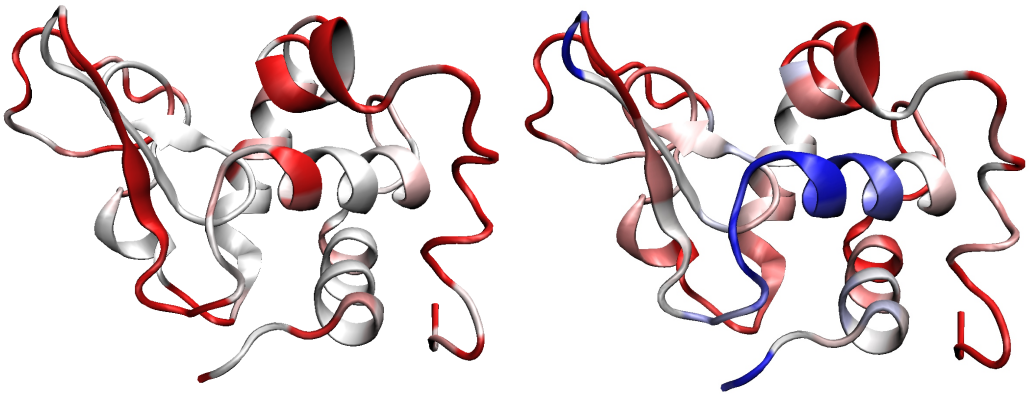


Рисунок 4.16 — Слева: среднее число водородных связей на аминокислоту; справа: гибкость аминокислот (окраска соответствует рисунку 4.15).

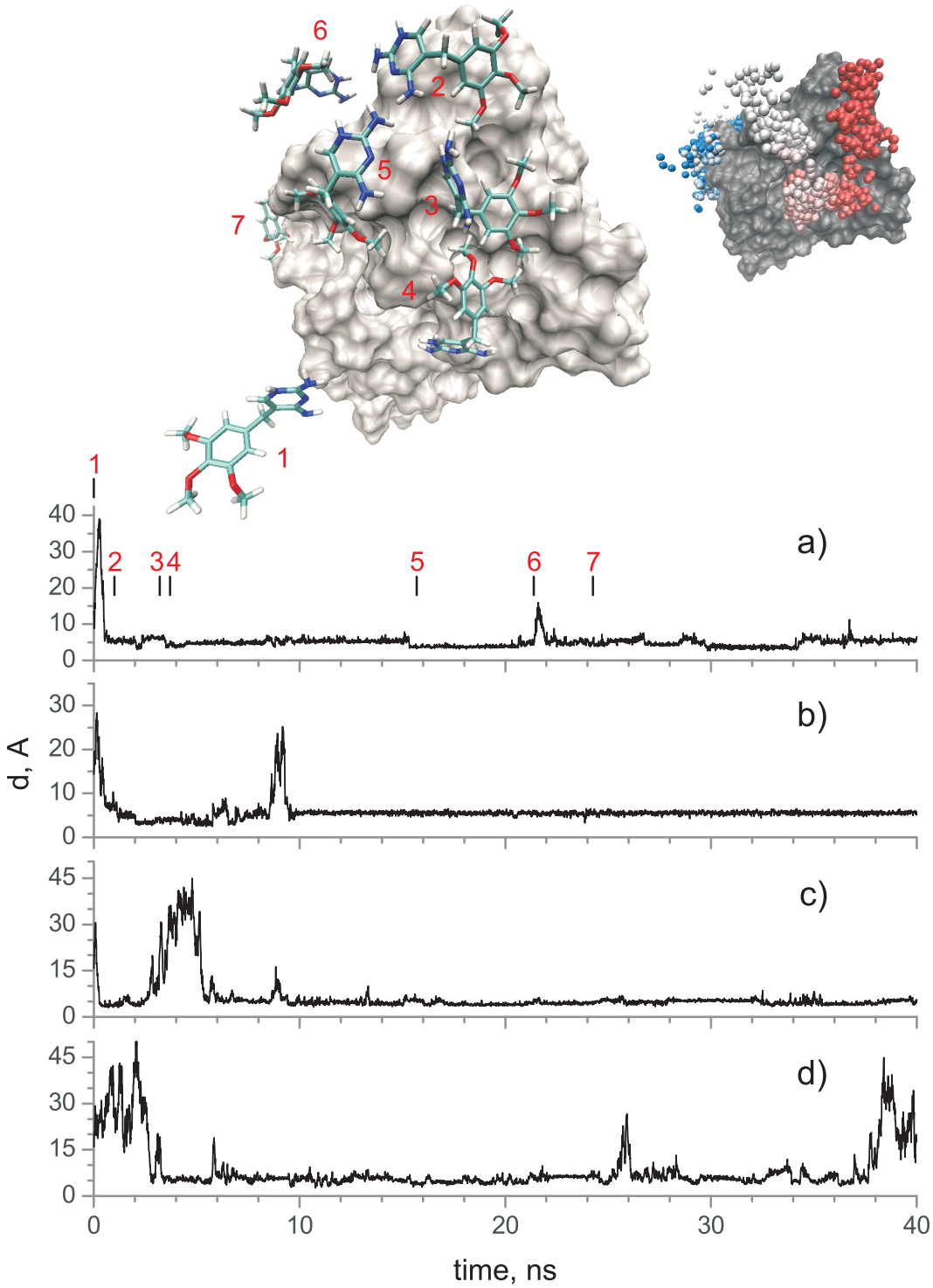


Рисунок 4.17 — Расстояние d между центрами масс лиганда TMP (a, b, c) или BrWR (d) и ближайшим остатком белка; низкие, постоянные значения d соответствуют интервалам времени, когда лиганды находятся на поверхности (и между лигандом и поверхностью нет молекул воды), а значения выше ≈ 7 показывают диффузию лигандов в объеме; несколько репрезентативных положений лиганда TMP на поверхности (и, иногда, вне нее) показаны вверху, числа на графике времени (a), от 1 до 7, соответствуют положениям лиганда на поверхности; верхний правый рисунок показывает диффузию TMP на поверхности из положения 2 в положение 7, лиганд представлен в виде одного шара, постепенно меняющего свой цвет вдоль траектории от красного (положение 2) через белый (приблизительно положение 5) к синему (положение 7).

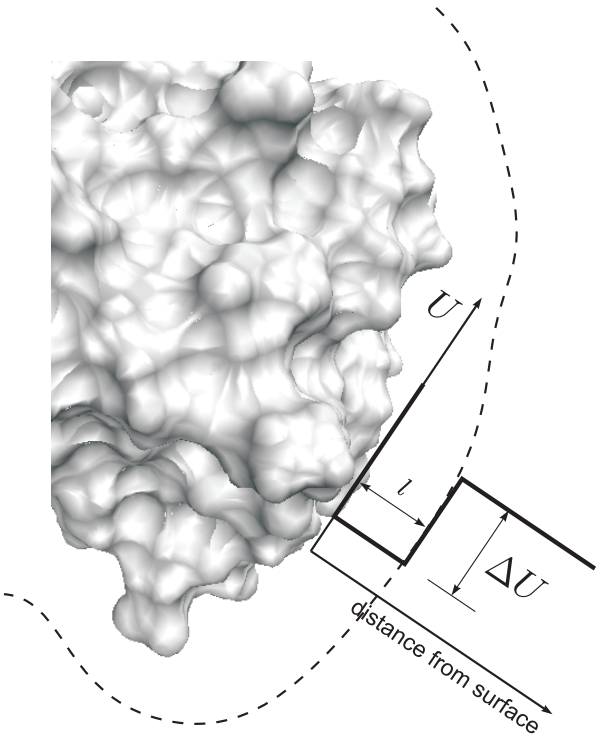


Рисунок 4.18 — Модель потенциала прямоугольной ямы для взаимодействия лиганда с поверхностью.

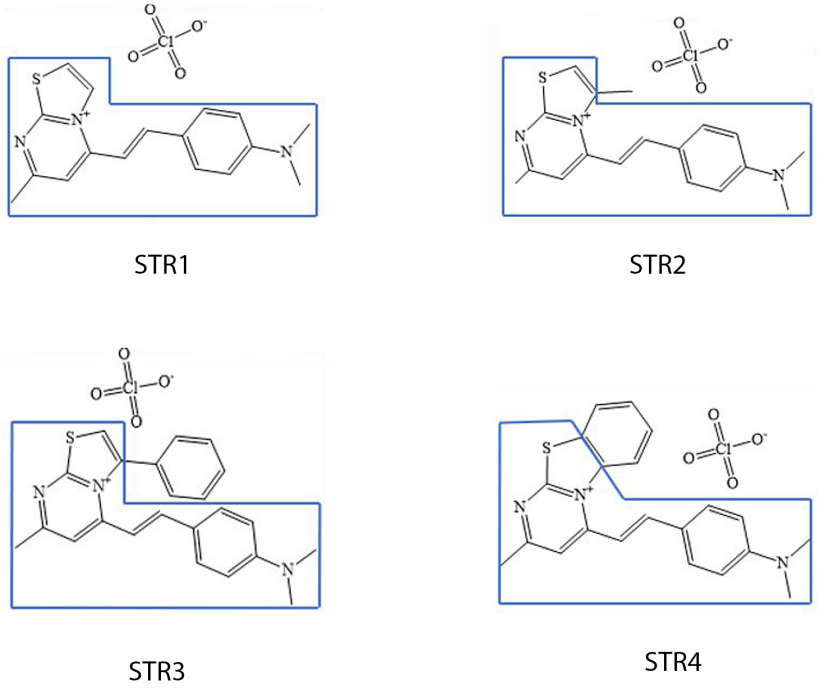


Рисунок 4.19 — Структурные формулы веществ STR1-STR4, использованных в исследовании взаимодействия с альбумином.

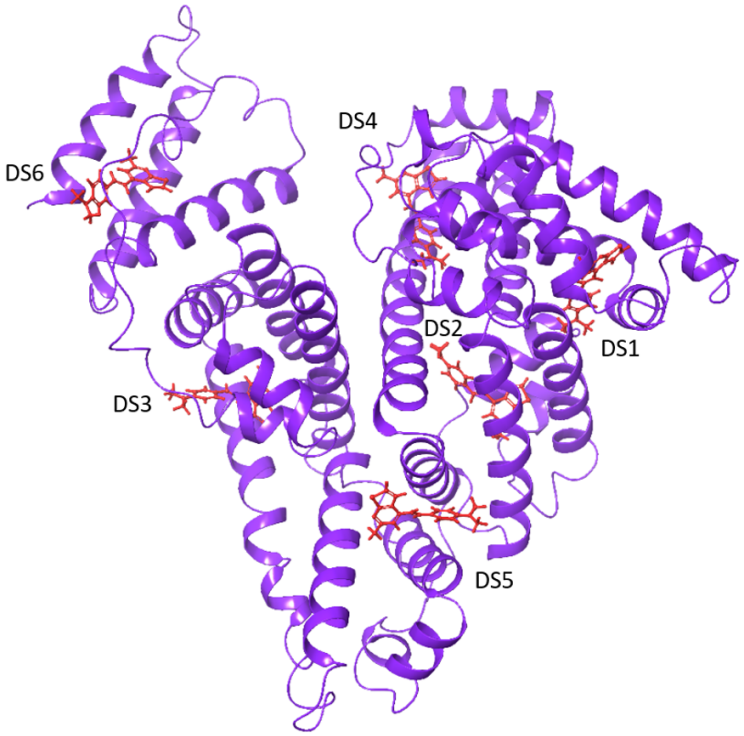


Рисунок 4.20 — Расположение лигандов в сайтах связывания ЧСА, DS1 – IA, DS2 – IIA, DS3 – IIIA, DS4 – IB, DS5 – IIB, DS6 – IIIB.

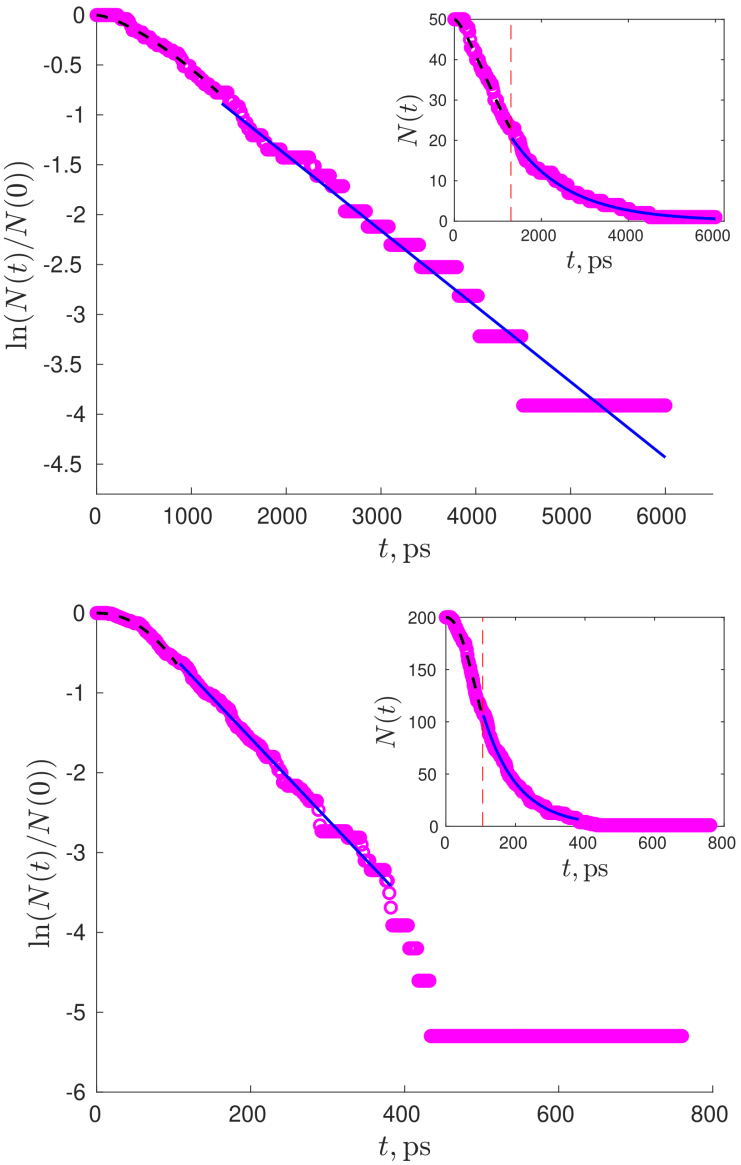


Рисунок 4.21 — Подгонка вероятности того, что лиганд останется связанным с белком, с использованием моделей для двух режимов; результаты для внешней силы, равной 350 кДж/(нм · mol) (а) и 500 кДж/(нм · mol) (б); черная штриховая линия — неэкспоненциальная модель, синяя линия — экспоненциальная модель, красная штрихпунктирная линия — момент переключения между моделями; подобранные значения параметров приведены в таблице 14.

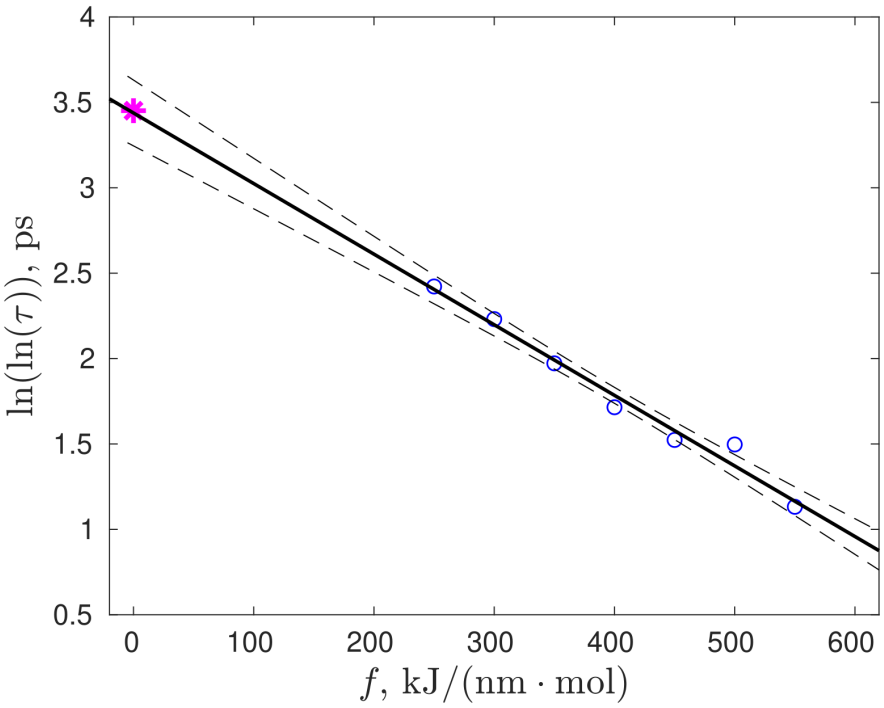


Рисунок 4.22 — Последовательность времен диссоциации, определенная с помощью моделирования MD, линеаризованная с помощью преобразования координат как функции приложенных сил на моль (точки) и их линейной подгонки (сплошная линия); пунктирные кривые обозначают границы предсказания с уровнем достоверности, равным стандартному отклонению; звёздочкой отмечено экспериментальное значение.

Заключение

Основные результаты работы заключаются в следующем.

1. Разработан теоретический подход для представления жидкого вещества с точки зрения атомистической классической молекулярной динамики и континуальной гидродинамики *одновременно,* обеспечивающий плавный переход одного представления в другое с изменением пространственного и временного масштабов, при этом соблюдаются фундаментальные физические законы сохранения на всех промежуточных масштабах.
2. Численное решение полученных уравнений движения реализовано в общеупотребительном пакете GROMACS с настройкой для вычислений на параллельных высокопроизводительных суперкомпьютерах. Разработанный код протестирован и валидирован с использованием наиболее производительных суперкомпьютеров, таких как K-computer, Sulis и другие.
3. В рамках разработанного подхода проведено моделирование ряда жидких систем, по мере усложнения: жидкий аргон, жидкая вода, водные растворы пептидов, белков и, наконец, целый вирусный капсид в физиологическом растворе. Рассчитаны физико-химические и биохимические характеристики систем и показано их соответствие как эталонным расчётным моделям, так и экспериментальным данным.
4. Проведено молекулярно-динамическое моделирование полноатомных моделей целых капсидов бактериофага MS2 и вируса PCV2. Исследованы биохимические и биофизические свойства этих систем, как то: распределение зарядов на капсиде и в растворе, показана ключевая роль растворённых ионов в стабильности вирусного капсида; динамика потоков ионов между капсидом и внешним раствором; биохимические механизмы самосборки вирусного капсида при определенном pH.
5. Изучена динамика молекулярных систем с точки зрения нелинейной динамики и теории сложности. Показаны особенности этой хаотической динамики и ее отличие от стохастического сигнала, наличие островов с “липкими” границами в фазовом пространстве соотнесено с квазипериодическими движениями молекул воды.
6. Применён подход “Computational Mechanics” к молекулярным траекториям и показан его потенциал в выявлении долговременных корреляций, значительно превышающих статистические корреляции рассчитанные стандартными методами. Выдвинуто предположение о связи этих долговременных корреляций с конформационными переходами в белках, потенциально приводящих к сворачиванию белка. Та же методология “Computational Mechanics” применена для выявления немарковского характера динамики конформационных переходов пептида в водном растворе, показана роль молекул воды в этом процессе.
7. Проведено моделирование нескольких систем белок-лиганд, где в качестве лиганда использовались малые молекул-модели лекарств. Рассчитанные биохимические и физико-химические характеристики систем, включая константы ассоциации и диссоциации максимально близко совпали с экспериментальными значениями.

В заключение хотелось бы обозначить направления дальнейших исследований по теме диссертации.

Продолжается работа над разработкой гибридных моделей связывающих молекулярную динамику и флуктуационную гидродинамику. Новый подход основан на использовании ограничений на механическую систему, описываемую уравнениями движения частиц МД. В качестве таких ограничений выступают гидродинамические поля, динамика которых должна воспроизводится усреднёнными по атомам величинами плотности, скорости и внутренней энергии (температуры). Первые результаты недавно опубликованы в следующих работах [52; 53].

Также продолжается работа в построении полноатомных моделей целых вирусов. На сегодняшний день нам удалось первыми в мире создать модель целого бактериофага MS2 *включая строение реального генома внутри капсида* [54]. Эта работа вызвала существенный интерес в экспериментальных исследованиях криоэлектронной микроскопии, что нашло отражение в нашем докладе на симпозиуме “Faraday Discussions” [55].

Наконец, хотелось бы упомянуть нашу последнюю работу в области моделирования вирусов: изучение деталей взаимодействия так называемых флагеллотропных бактериофагов с бактериальными жгутиками. Особенность этого класса бактериофагов состоит в том, что они используют свои хвостовые фибриллы для охватывания бактериального жгутика, после чего, в результате вращения жгутика, бактериофаг “накручивается” на жгутик как гайка на болт. Было проведено гидродинамичекое моделирование этого процесса и показано совпадение скорости накручивания с экпериментальными оценками [56]. В настоящее время мы с коллегами из университета Осаки проводим крио-электронные измерения для определения молекулярной структуры комплекса бактериофаг-жгутик. Имея такую структуру мы сможем приступить к полноатомному и гибридному моделированию процесса.

В завершении работы хочется ещё раз отметить большой потенциал описанных методов компьютерного моделирования. Гибридность позволяет преодолеть теоретический барьер невозможности атомистического моделирования макрообъектов при сохранении всех ключевых параметров системы. У нас нет никаких сомнений что данное направление принесёт много важных и интересных открытий как в познании фундаментальных законов природы вещества так и в прикладных направлениях.

Словарь терминов

MD : Молекулярная Динамика
CG : Coarse Grained model
DOF : степени свободы
MARTINI : “гранулированное” силовое поле
LL-FH : флуктуационная гидродинамика Ландау-Лифшица
HC : уравнения гидродинамики Навье-Стокса
FH : флуктуационная гидродинамика
MB : модель воды Мерседес-Бенц
VACF : корреляционная функция скорости
MS2 : бактериофаг MS2
PCV2 : вирус PCV2
EOS : уравнение состояния
RDF : радиальная функция распределения
STD : стандартное отклонение
RMSD : среднеквадратичное отклонение
RMSF : среднеквадратичные флуктуации

Список литературы

1. Multiscale Modeling and Simulation of Composite Materials and Structures / под ред. W. Young, H. David, R. Talreja. — 6-е изд. — Chichester : Springer, 2007.
2. *Vlachos, D.* A Review of Multiscale Analysis: Examples from Systems Biology, Materials Engineering, and Other Fluid Surface Interacting Systems / D. Vlachos // Advances in Chemical Engineering. — 2005. — Т. 30, № 5. — 1 61. — DOI: 10.1021/ja0276321.
3. *Frauenfelder, H.* A Unified Model of Protein Dynamics / H. Frauenfelder, G. Chen, J. Berendzen [и др.] // Proceedings of the National Academy of Sciences. — 2009. — Март. — Т. 106, № 13. — С. 5129—5134. — DOI: 10.1073/pnas.0900336106.
4. *West, B.* Membrane-Protein Interactions in a Generic Coarse-Grained Model for Lipid Bilayers / B. West, F. L. Brown, F. Schmid // Biophysical Journal. — 2009. — Янв. — Т. 96, № 1. — С. 101—115. — DOI: 10.1529/biophysj.108.138677.
5. *Lonsdale, R.* A Multiscale Approach to Modelling Drug Metabolism by Membrane-Bound Cytochrome P450 Enzymes / R. Lonsdale, S. L. Rouse, M. S. P. Sansom [и др.] // PLoS Computational Biology / под ред. M. Punta. — 2014. — Июль. — Т. 10, № 7. — e1003714. — DOI: 10.1371/journal.pcbi.1003714.
6. *Kasson, P. M.* ”Cross-graining”: efficient multi-scale simulation via Markov state models. / P. M. Kasson, V. S. Pande // Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing. — 2010. — С. 260—268. — ppublish.
7. *Larsson, D. S. D.* Virus Capsid Dissolution Studied by Microsecond Molecular Dynamics Simulations / D. S. D. Larsson, L. Liljas, D. van der Spoel // PLoS Computational Biology / под ред. M. Nilges. — 2012. — Май. — Т. 8, № 5. — e1002502. — DOI: 10.1371/journal.pcbi.1002502.
8. *Zink, M.* Primary Changes of the Mechanical Properties of Southern Bean Mosaic Virus upon Calcium Removal / M. Zink, H. Grubmüller // Biophysical Journal. — 2010. — Февр. — Т. 98, № 4. — С. 687—695. — DOI: 10.1016/j.bpj.2009.10.047.
9. *Flekkoy, E. G.* Hybrid model for combined particle and continuum dynamics / E. G. Flekkoy, G. Wagner, J. Feder // Europhys. Lett. — 2000. — Т. 52. — С. 271—276.
10. *Asproulis, A.* A hybrid molecular continuum method using point wise coupling / A. Asproulis, M. Kalweit, D. Drikakis // Advances in Engineering Software. — 2012. — Т. 46. — С. 85—92.
11. *O’Connell, S.* Molecular dynamics-continuum hybrid computations: a tool for studying complex fluid flows / S. O’Connell, P. Thompson // Phys. Rev. E. — 1995. — Т. 52. — С. 5792.
12. *Kotsalis, E. M.* Control of density fluctuations in atomistic-continuum simulations of dense liquids / E. M. Kotsalis, J. H. Walther, P. Koumoutsakos // Physical Review E. — 2007. — Июль. — Т. 76, № 1. — С. 016709. — DOI: 10.1103/physreve.76.016709.

13. *Kotsalis, E. M.* Control algorithm for multiscale flow simulations of water / E. M. Kotsalis, J. H. Walther, E. Kaxiras [и др.] // Physical Review E. — 2009. — Апр. — Т. 79, № 4. — С. 045701. — DOI: 10.1103/physreve.79.045701.

14. *Fabritiis, G. D.* Multiscale Modeling of Liquids with Molecular Specificity / G. D. Fabritiis, R. Delgado-Buscalioni, P. V. Coveney // Phys. Rev. Lett. — 2006. — Т. 97, № 13. — С. 134501. — DOI: 10.1103/PhysRevLett.97.134501.

15. *Nie, X. B.* A continuum and molecular dynamics hybrid method for micro- and nano-fluid flow / X. B. Nie, S. Y. Chen, W. N. E [и др.] // J. Fluid Mech. — 2004. — Т. 500. — С. 55—64.

16. *Batchelor, G. K.* An introduction to fluid dynamics / G. K. Batchelor. — Cambridge university press, 2000.

17. *Golse, F.* Chapter 3 The Boltzmann Equation and Its Hydrodynamic Limits / F. Golse //. Т. 2 / под ред. C. Dafermos, E. Feireisl. — North-Holland, 2006. — С. 159—301. — (Handbook of Differential Equations: Evolutionary Equations). — DOI: http://dx.doi.org/10.1016/S1874-5717(06)80006-X.

18. *Pasquale, N.* Mixing atoms and coarse-grained beads in modelling polymer melts / N. Pasquale, D. Marchisio, P. Carbone // Journal of Chemical Physics. — 2012. — Т. 137. — С. 164111—164119.

19. *Noid, W. G.* The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models / W. G. Noid, J.-W. Chu, G. S. Ayton [и др.] // The Journal of Chemical Physics. — 2008. — Июнь. — Т. 128, № 24. — С. 244114. — DOI: 10.1063/1.2938860.

20. *Müller-Plathe, F.* Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back / F. Müller-Plathe // ChemPhysChem. — 2002. — Т. 3, № 9. — С. 754—769. — DOI: https://doi.org/10.1002/1439-7641(20020916)3:9<754::AID-CPHC754>3.0.CO;2-U.

21. *Tang, Y.-H.* Accelerating dissipative particle dynamics simulations on GPUs: Algorithms, numerics and applications / Y.-H. Tang, G. E. Karniadakis // Computer Physics Communications. — 2014. — Нояб. — Т. 185, № 11. — С. 2809—2822. — DOI: 10.1016/j.cpc.2014.06.015.

22. *Hoogerbrugge, P. J.* Simulating Microscopic Hydrodynamic Phenomena with Dissipative Particle Dynamics / P. J. Hoogerbrugge, J. M. V. A. Koelman // Europhysics Letters (EPL). — 1992. — Июнь. — Т. 19, № 3. — С. 155—160. — DOI: 10.1209/0295-5075/19/3/001.

23. *Español, P.* Dissipative particle dynamics with energy conservation / P. Español // Europhysics Letters (EPL). — 1997. — Дек. — Т. 40, № 6. — С. 631—636. — DOI: 10.1209/epl/i1997-00515-8.

24. *J., H. W.* / H. W. J., Xu, S. Zhang [и др.] // IEIT J. Adapt. Dyn. Comput. — 2011. — Т. 2. — С. 26—32.

25. *Berendsen, H.* GROMACS: A message-passing parallel molecular dynamics implementation / H. Berendsen, D. van der Spoel, R. van Drunen // Computer Physics Communications. — 1995. — Сент. — Т. 91, № 1—3. — С. 43—56. — DOI: 10.1016/0010-4655(95)00042-e.

26. *Goga, N.* / N. Goga, S. Costashe, S. Marrink // Mater. Plast. — 2009. — Т. 46, № 1. — С. 53.

27. *Goga, N.* Multiscaling algorithms for molecular dynamics simulations with GROMACS / N. Goga, S. Marrink, S. V. Costache [и др.] // 2009 3rd Annual IEEE Systems Conference. — IEEE, 03.2009. — DOI: 10.1109/systems.2009.4815825.

28. *Marrink, S. J.* The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations / S. J. Marrink, H. J. Risselada, S. Yefimov [и др.] // The Journal of Physical Chemistry B. — 2007. — Июнь. — Т. 111, № 27. — С. 7812—7824. — DOI: 10.1021/jp071097f.

29. *Buti, F.* BioShape: a spatial shape-based scale-independent simulation environment for biological systems / F. Buti, D. Cacciagrano, F. Corradini [и др.] // Procedia Computer Science. — 2010. — Май. — Т. 1, № 1. — С. 827—835. — DOI: 10.1016/j.procs.2010.04.090.

30. *Noid, W. G.* Multiscale Coarse-Graining and Structural Correlations: Connections to Liquid-State Theory / W. G. Noid, J.-W. Chu, G. S. Ayton [и др.] // The Journal of Physical Chemistry B. — 2007. — Март. — Т. 111, № 16. — С. 4116—4127. — DOI: 10.1021/jp068549t.

31. *Praprotnik, M.* A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation / M. Praprotnik, L. D. Site, K. Kremer // The Journal of Chemical Physics. — 2007. — Апр. — Т. 126, № 13. — С. 134902. — DOI: 10.1063/1.2714540.

32. *Potestio, R.* Hamiltonian Adaptive Resolution Simulation for Molecular Liquids / R. Potestio, S. Fritsch, P. Español [и др.] // Physical Review Letters. — 2013. — Март. — Т. 110, № 10. — С. 108301. — DOI: 10.1103/physrevlett.110.108301.

33. *Español, P.* Statistical mechanics of Hamiltonian adaptive resolution simulations / P. Español, R. Delgado-Buscalioni, R. Everaers [и др.] // The Journal of Chemical Physics. — 2015. — Февр. — Т. 142, № 6. — С. 064115. — DOI: 10.1063/1.4907006.

34. *Lighthill, M. J.* On sound generated aerodynamically I. General theory / M. J. Lighthill // Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences. — 1952. — Март. — Т. 211, № 1107. — С. 564—587. — DOI: 10.1098/rspa.1952.0060.

35. *GOLDSTEIN, M. E.* A generalized acoustic analogy / M. E. GOLDSTEIN // Journal of Fluid Mechanics. — 2003. — Июль. — Т. 488. — С. 315—333. — DOI: 10.1017/s0022112003004890.

36. *GOLDSTEIN, M. E.* The aeroacoustics of slowly diverging supersonic jets / M. E. GOLDSTEIN, S. J. LEIB // Journal of Fluid Mechanics. — 2008. — Март. — Т. 600. — С. 291—337. — DOI: 10.1017/s0022112008000311.

37. *Karabasov, S. A.* Understanding jet noise / S. A. Karabasov // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. — 2010. — Авг. — Т. 368, № 1924. — С. 3593—3608. — DOI: 10.1098/rsta.2010.0086.

38. *Buckley, S.* Mechanism of Fluid Displacement in Sands / S. Buckley, M. Leverett // Transactions of the AIME. — 1942. — Дек. — Т. 146, № 01. — С. 107—116. — DOI: 10.2118/942107-g.

39. *Ben-Naim, A.* Statistical Mechanics of “Waterlike” Particles in Two Dimensions. I. Physical Model and Application of the Percus–Yevick Equation / A. Ben-Naim // The Journal of Chemical Physics. — 1971. — Май. — Т. 54, № 9. — С. 3682—3695. — DOI: 10.1063/1.1675414.

40. *Landau, L. D.* Statistical Physics, Part 1 (Third Edition) / L. D. Landau, E. M. Lifshitz, L. E. Reichl // Physics Today. — 1981. — Янв. — Т. 34, № 1. — С. 74—74. — DOI: 10.1063/1.2889978.

41. *De Zarate, J. M. O.* Hydrodynamic fluctuations in fluids and fluid mixtures / J. M. O. De Zarate, J. V. Sengers. — Elsevier, 2006.

42. *Bell, J. B.* Numerical methods for the stochastic Landau-Lifshitz Navier-Stokes equations / J. B. Bell, A. L. Garcia, S. A. Williams // Physical Review E. — 2007. — Июль. — Т. 76, № 1. — С. 016708. — DOI: 10.1103/physreve.76.016708.

43. *Donev, A.* On the accuracy of finite-volume schemes for fluctuating hydrodynamics / A. Donev, E. Vanden-Eijnden, A. Garcia [и др.] // Communications in Applied Mathematics and Computational Science. — 2010. — Июнь. — Т. 5, № 2. — С. 149—197. — DOI: 10.2140/camcos.2010.5.149.

44. *Markesteijn, A.* A new non-linear two-time-level Central Leapfrog scheme in staggered conservation-flux variables for fluctuating hydrodynamics equations with GPU implementation / A. Markesteijn, S. Karabasov, V. Glotov [и др.] // Computer Methods in Applied Mechanics and Engineering. — 2014. — Нояб. — Т. 281. — С. 29—53. — DOI: 10.1016/j.cma.2014.07.027.

45. *Fabritiis, G. D.* Fluctuating hydrodynamic modeling of fluids at the nanoscale / G. D. Fabritiis, M. Serrano, R. Delgado-Buscalioni [и др.] // Physical Review E (Statistical, Nonlinear, and Soft Matter Physics). — 2007. — Т. 75, № 2. — С. 026307. — DOI: 10.1103/PhysRevE.75.026307.

46. *Voulgarakis, N. K.* Bridging fluctuating hydrodynamics and molecular dynamics simulations of fluids / N. K. Voulgarakis, J.-W. Chu // The Journal of Chemical Physics. — 2009. — Апр. — Т. 130, № 13. — С. 134111. — DOI: 10.1063/1.3106717.

47. *Ladd, A. J. C.* Short-time motion of colloidal particles: Numerical simulation via a fluctuating lattice-Boltzmann equation / A. J. C. Ladd // Physical Review Letters. — 1993. — Март. — Т. 70, № 9. — С. 1339—1342. — DOI: 10.1103/physrevlett.70.1339.

48. *Markesteijn, A. P.* Flow injection of polymers into nanopores / A. P. Markesteijn, O. B. Usta, I. Ali [и др.] // Soft Matter. — 2009. — Т. 5, № 22. — С. 4575. — DOI: 10.1039/b909208k.

49. *Brody, J. P.* Biotechnology at low Reynolds numbers / J. P. Brody, P. Yager, R. E. Goldstein [и др.] // Biophysical journal. — 1996. — Т. 71, № 6. — С. 3430—3441.

50. *Shapiro, E.* Interface capturing in dual-flow microfluidics / E. Shapiro, D. Drikakis, J. Gargiuli [и др.] // Journal of Computational and Theoretical Nanoscience. — 2007. — Т. 4, № 4. — С. 802—806.

51. *Ivorra, B.* Two-and three-dimensional modeling and optimization applied to the design of a fast hydrodynamic focusing microfluidic mixer for protein folding / B. Ivorra, J. L. Redondo, J. G. Santiago [и др.] // Physics of fluids. — 2013. — Т. 25, № 3. — С. 032001.

52. *Koplik, J.* Continuum deductions from molecular hydrodynamics / J. Koplik // APS March Meeting Abstracts. — 1998. — U1—02.

53. *Mansur, E. A.* A state-of-the-art review of mixing in microfluidic mixers / E. A. Mansur, Y. Mingxing, W. Yundong [и др.] // Chinese Journal of Chemical Engineering. — 2008. — Т. 16, № 4. — С. 503—516.

54. *Kumar, P.* Molecular dynamics study of orientational cooperativity in water / P. Kumar, G. Franzese, S. V. Buldyrev [и др.] // Physical Review E. — 2006. — Т. 73, № 4. — С. 041505.

55. *Umezawa, K.* Solvent flow patterns fluctuating largely around a protein and correlation with solvent density fluctuations: a molecular dynamics study / K. Umezawa, R. Morikawa, H. Nakamura [и др.] // The Journal of chemical physics. — 2010. — Т. 132, № 15. — 04B608.

56. *Frauenfelder, H.* A unified model of protein dynamics / H. Frauenfelder, G. Chen, J. Berendzen [и др.] // Proceedings of the National Academy of Sciences. — 2009. — Т. 106, № 13. — С. 5129—5134.

57. *Naim, A. B.* Statistical Mechanics of 'Waterlike' Particles in Two Dimensions. I. Physical Model and Application of the Percus–Yevick Equation / A. B. Naim // J. Chem. Phys. — 1971. — Т. 54. — С. 3682.

58. *Allen, M. P.* Computer Simulation of Liquids / M. P. Allen, D. J. Tildesley. — USA : Oxford University Press, 2000.

59. *Darden, T.* Particle mesh Ewald: An N log(N) method for Ewald sums in large systems / T. Darden, D. York, L. Pedersen // J. Chem. Phys. — 1993. — Т. 98. — С. 10089.

60. *Silverstein, K. A. T.* A Simple Model of Water and the Hydrophobic Effect / K. A. T. Silverstein, A. D. J. Haymet, K. A. Dill // J. Am. Chem. Soc. — 1998. — Т. 120. — С. 3166—3175.

61. *Kataoka, Y.* Studies of Liquid Water by Computer Simulations. III. Dynamical Properties of a 2D Model / Y. Kataoka // Bull. Chem. Soc. Jpn. — 1983. — Т. 57. — С. 1522—1527.

62. *Kataoka, Y.* A Molecular Dynamical Study of the Mutual Diffusion Coefficient and Cooperative Motion in a 2-Dimensional Aquaeous Solution / Y. Kataoka // Bull. Chem. Soc. Jpn. — 1985. — Т. 59. — С. 3341—3346.

63. *Okazaki, K.* Study of liquid water by computer simulations. I. Static properties of a 2D model / K. Okazaki, S. Nosé, Y. Kataoka [и др.] // The Journal of Chemical Physics. — 1981. — Т. 75, № 12. — С. 5864—5874. — DOI: 10.1063/1.442036.

64. *Hirata, F.* A realization of 'V structure' in liquid water / F. Hirata, P. J. Rossky // J. Chem. Phys. — 1981. — Т. 74. — С. 6867—6874.

65. *Dias, C. L.* Microscopic Mechanism for Cold Denaturation / C. L. Dias, T. Ala-Nissala, M. Karttunen [и др.] // Physical Review Letters. — 2008. — Т. 100. — С. 118101.

66. *Dias, C. L.* Using Microscopic Mechanism for Pressure and Cold Denaturations of Proteins / C. L. Dias // Physical Review Letters. — 2012. — Т. 109. — С. 048104.

67. *Kell, G. S.* Density, Thermal Expansivity, and Compressibility of Liquid Water from 0° to 150° C. : Correlations and Tables for Atmospheric Pressure and Saturation Reviewed and Expressed on 1968 Temperature Scale / G. S. Kell // Journal Of Chemical and Engineering Data. — 1975. — Т. 20. — С. 97—105.

68. *Urbič, T.* A two-dimensional model of water: Solvation of nonpolar solutes / T. Urbič, V. Vlachy, Y. V. Kalyuzhnyi [и др.] // J. Chem. Phys. — 2002. — Т. 116. — С. 723.

69. *Urbič, T.* Orientation-dependent integral equation theory for a two-dimensional model of water / T. Urbič, V. Vlachy, Y. V. Kalyuzhnyi [и др.] // J. Chem. Phys. — 2003. — Т. 118. — С. 5516.

70. *Urbič, T.* Theory for the solvation of nonpolar solutes in water / T. Urbič, V. Vlachy, Y. V. Kalyuzhnyi [и др.] // J. Chem. Phys. — 2007. — Т. 127. — С. 174505.

71. *Urbič, T.* An improved thermodynamic perturbation theory for Mercedes-Benz water / T. Urbič, V. Vlachy, Y. V. Kalyuzhnyi [и др.] // J. Chem. Phys. — 2007. — Т. 127. — С. 174511.

72. *Urbič, T.* Mercedes–Benz water molecules near hydrophobic wall: Integral equation theories vs Monte Carlo simulations / T. Urbič, M. F. Holovkol // J. Chem. Phys. — 2011. — Т. 135. — С. 134706.

73. *Dias, C. L.* The hydrophobic effect and its role in cold denaturation / C. L. Dias, T. A.-N. J. Wong-ekkabut, I. Vattulainen [и др.] // Cryobiology. — 2010. — Т. 60. — С. 91—99.

74. *Khayat, R.* The 2.3-Angstrom Structure of Porcine Circovirus 2 / R. Khayat, N. Brunn, J. A. Speir [и др.] // Journal of Virology. — 2011. — Т. 85, № 15. — С. 7856—7862. — DOI: 10.1128/JVI.00737-11.

75. *Lane, S. W.* Construction and crystal structure of recombinant STNV capsids / S. W. Lane, C. A. Dennis, C. L. Lane [и др.] // Journal of molecular biology. — 2011. — Т. 413, № 1. — С. 41—50.

76. *Duquerroy, S.* The picobirnavirus crystal structure provides functional insights into virion assembly and cell entry / S. Duquerroy, B. Da Costa, C. Henry [и др.] // The EMBO journal. — 2009. — Т. 28, № 11. — С. 1655—1665.

77. *Wang, X.* Hepatitis A virus and the origins of picornaviruses / X. Wang, J. Ren, Q. Gao [и др.] // Nature. — 2015. — Т. 517, № 7532. — С. 85—88.

78. *Zhao, G.* Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics / G. Zhao, J. R. Perilla, E. L. Yufenyuy [и др.] // Nature. — 2013. — Т. 497, № 7451. — С. 643—646.

79. *Yang, B.* Porcine circovirus (PCV) removal by Q sepharose fast flow chromatography / B. Yang, H. Wang, C. Ho [и др.] // Biotechnology progress. — 2013. — Т. 29, № 6. — С. 1464—1471.

80. *De Fabritiis, G.* Fluctuating hydrodynamic modeling of fluids at the nanoscale / G. De Fabritiis, M. Serrano, R. Delgado-Buscalioni [и др.] // Physical Review E. — 2007. — Т. 75, № 2. — С. 026307.

81. *Landau, L.* Statistical Physics / L. Landau, E. Lifshitz. — Oxford: Pergamon, 1980.

82. *Landau, L.* Fluid Mechanics / L. Landau, E. Lifshitz. — Pergamon press, 1966.

83. *Markesteijn, A.* Time asynchronous relative dimension in space method for multi-scale problems in fluid dynamics / A. Markesteijn, S. Karabasov // Journal of Computational Physics. — 2014. — Т. 258, № 10. — С. 137—164. — DOI: 10.1021/ja00089a029.

84. *Omelchenko, Y. A.* Self-adaptive time integration of flux-conservative equations with sources / Y. A. Omelchenko, H. Karimabadi // Journal of Computational Physics. — 2006. — Т. 216, № 1. — С. 179—194.

85. *Crighton, D. G.* Modern methods in analytical acoustics lecture notes / D. G. Crighton, A. P. Dowling, J. Ffowcs-Williams [и др.]. — 1992.

86. *Drew, D.* Theory of Multicomponent Fluids / D. Drew, S.L.Passman. — New York : Springers, 1999.

87. *Hynninen, T.* A molecular dynamics implementation of the 3D Mercedes-Benz water model / T. Hynninen, C. Dias, A. Mkrtchyand [и др.] // Computer Physics Communication. — 2011.

88. *Naim, A. B.* Molecular Theory of Water and Aqueous Solutions / A. B. Naim. — Singapore : World Scientific, 2010.

89. *Pearson, E. M.* Laplace-transform technique for deriving thermodynamic equations from the classical microcanonical ensemble / E. M. Pearson // Phys. Rev. A. — 1985. — Т. 32. — С. 3030—3039.

90. *Eisenberg, D.* The structure and properties of water / D. Eisenberg, W. Kauzmann. — London : Oxford University Press, 1969.

91. *Mark, P.* Structure and Dynamics of the TIP3P, SPC, and SPC /E Water Models at 298 K / P. Mark, L. Nilsson // J. Phys. Chem. A. — 2001. — Т. 105 (43). — С. 9954—9960.

92. *Tuckerman, M. E.* Statistical Mechanics: Theory and Molecular Simulation / M. E. Tuckerman. — Oxford : Oxford University Press, 2012.

93. *Nosé, S.* A molecular dynamics method for simulations in the canonical ensemble / S. Nosé // Molecular Physics. — 1984. — Июнь. — Т. 52, № 2. — С. 255—268. — DOI: 10 . 1080 / 00268978400101201.

94. *Hoover, W. G.* Canonical dynamics: Equilibrium phase-space distributions / W. G. Hoover // Phys. Rev. A. — 1985. — Март. — Т. 31, № 3. — С. 1695—1697. — DOI: 10.1103/PhysRevA.31.1695.

95. *Hess, B.* LINCS: A linear constraint solver for molecular simulations / B. Hess, H. Bekker, H. J. C. Berendsen [и др.] // Journal of Computational Chemistry. — 1997. — Т. 18, № 12. — С. 1463—1472. — DOI: https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

96. *Cino-Ozuna, A. G.* Characterization of a new disease syndrome associated with porcine circovirus type 2 in previously vaccinated herds / A. G. Cino-Ozuna, S. Henry, R. Hesse [и др.] // Journal of clinical microbiology. — 2011. — Т. 49, № 5. — С. 2012—2016.

97. *Segalés, J.* The natural history of porcine circovirus type 2: from an inoffensive virus to a devastating swine disease? / J. Segalés, T. Kekarainen, M. Cortey // Veterinary microbiology. — 2013. — Т. 165, № 1/2. — С. 13—20.

98. *Yin, S.* Self-assembly of virus-like particles of porcine circovirus type 2 capsid protein expressed from Escherichia coli / S. Yin, S. Sun, S. Yang [и др.] // Virology journal. — 2010. — Т. 7, № 1. — С. 1—5.

99. *Nawagitgul, P.* Open reading frame 2 of porcine circovirus type 2 encodes a major capsid protein / P. Nawagitgul, I. Morozov, S. R. Bolin [и др.] // Journal of General Virology. — 2000. — Т. 81, № 9. — С. 2281—2287.

100. *Fan, H.* Immunogenicity of empty capsids of porcine circovius type 2 produced in insect cells / H. Fan, C. Ju, T. Tong [и др.] // Veterinary research communications. — 2007. — Т. 31, № 4. — С. 487—496.

101. *Freddolino, P. L.* Molecular dynamics simulations of the complete satellite tobacco mosaic virus / P. L. Freddolino, A. S. Arkhipov, S. B. Larson [и др.] // Structure. — 2006. — Т. 14, № 3. — С. 437—449.

102. *Larsson, D. S.* Virus capsid dissolution studied by microsecond molecular dynamics simulations / D. S. Larsson, L. Liljas, D. van der Spoel // PLoS computational biology. — 2012. — Т. 8, № 5. — e1002502.

103. *Andoh, Y.* All-atom molecular dynamics calculation study of entire poliovirus empty capsids in solution / Y. Andoh, N. Yoshii, A. Yamada [и др.] // The Journal of chemical physics. — 2014. — Т. 141, № 16. — 10B611_1.

104. *Šali, A.* Comparative protein modelling by satisfaction of spatial restraints / A. Šali, T. L. Blundell // Journal of molecular biology. — 1993. — Т. 234, № 3. — С. 779—815.

105. *Carrillo-Tripp, M.* VIPERdb2: an enhanced and web API enabled relational database for structural virology / M. Carrillo-Tripp, C. M. Shepherd, I. A. Borelli [и др.] // Nucleic acids research. — 2009. — Т. 37, suppl_1. — С. D436—D442.

106. *Markesteijn, A.* A New Non-Linear Two-Time-Level Central Leapfrog Scheme in Staggered Conservation-Flux Variables for Fluctuating Hydrodynamics Equations with GPU Implementation / A. Markesteijn, S. Karabasov, V. Glotov [и др.] // Computer Methods in Applied Mechanics and Engineering. — 2014. — Т. under revision.

107. *Glotov, V.* New Two Level Leapfrog Scheme for Modeling the Stochastic Landau–Lifshitz Equations / V. Glotov, V. Goloviznin, S. Karabasov [и др.] // Computational Mathematics and Mathematical Physics. — 2014. — Т. 54 (2). — С. 315—334.

108. *Balboa, F.* Staggered Schemes for Fluctuating Hydrodynamics / F. Balboa, J. B. Bell, R. Delgado-Buscalioni [и др.] // Multiscale Modeling and Simulation. — 2012. — Янв. — Т. 10, № 4. — С. 1369—1408. — DOI: 10.1137/120864520.

109. *Irving, J. H.* The Statistical Mechanical Theory of Transport Processes. IV. The Equations of Hydrodynamics / J. H. Irving, J. G. Kirkwood // The Journal of Chemical Physics. — 1950. — Июнь. — Т. 18, № 6. — С. 817—829. — DOI: 10.1063/1.1747782.

110. *Markesteijn, A. P.* A comparison of the value of viscosity for several water models using Poiseuille flow in a nano-channel / A. P. Markesteijn, R. Hartkamp, S. Luding [и др.] // The Journal of Chemical Physics. — 2012. — Апр. — Т. 136, № 13. — С. 134104. — DOI: 10.1063/1.3697977.

111. *Sunda, A. P.* Parametric dependence on shear viscosity of SPC/E water from equilibrium and non-equilibrium molecular dynamics simulations / A. P. Sunda, A. Venkatnathan // Molecular Simulation. — 2013. — Авг. — Т. 39, № 9. — С. 728—733. — DOI: 10.1080/08927022.2012.762098.

112. *GUO, G.-J.* Equilibrium molecular dynamics calculation of the bulk viscosity of liquid water / G.-J. GUO, Y.-G. ZHANG // Molecular Physics. — 2001. — Февр. — Т. 99, № 4. — С. 283—289. — DOI: 10.1080/00268970010011762.

113. *Fanourgakis, G. S.* Determining the Bulk Viscosity of Rigid Water Models / G. S. Fanourgakis, J. S. Medina, R. Prosmiti // The Journal of Physical Chemistry A. — 2012. — Март. — Т. 116, № 10. — С. 2564—2570. — DOI: 10.1021/jp211952y.

114. *González, M. A.* The shear viscosity of rigid water models / M. A. González, J. L. F. Abascal // The Journal of Chemical Physics. — 2010. — Март. — Т. 132, № 9. — С. 096101. — DOI: 10.1063/1.3330544.

115. *Shang, B. Z.* Fluctuating hydrodynamics for multiscale modeling and simulation: Energy and heat transfer in molecular fluids / B. Z. Shang, N. K. Voulgarakis, J.-W. Chu // The Journal of Chemical Physics. — 2012. — Июль. — Т. 137, № 4. — С. 044117. — DOI: 10.1063/1.4738763.

116. *Oostenbrink, C.* A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6 / C. Oostenbrink, A. Villa, A. E. Mark [и др.] // Journal of Computational Chemistry. — 2004. — Окт. — Т. 25, № 13. — С. 1656—1676.

117. *Chu, J.-W.* Coarse-grained modeling of the actin filament derived from atomistic-scale simulations / J.-W. Chu, G. A. Voth // Biophysical journal. — 2006. — Т. 90, № 5. — С. 1572—1582.

118. *Makino, J.* Scientific Simulations with Special-Purpose Computers / J. Makino, M. Taiji. — John Wiley & Sons, 1997.

119. *Shaw, D. E.* Anton, a Special-purpose Machine for Molecular Dynamics Simulation / D. E. Shaw, M. M. Deneroff, R. O. Dror [и др.] // Commun. ACM. — New York, NY, USA, 2008. — Июль. — Т. 51, № 7. — С. 91—97. — DOI: 10.1145/1364782.1364802.

120. *Voulgarakis, N. K.* Bridging fluctuating hydrodynamics and molecular dynamics simulations of fluids / N. K. Voulgarakis, J.-W. Chu // J. Chem. Phys. — 2009. — Т. 130.

121. *Baig, T.* In-silico studies of oncogene protein with anti-cancer drugs / T. Baig, H. Sheikh, P. K. Tripathi // Global Journal of Biotechnology & Biochemistry. — 2014. — Т. 9, № 3. — С. 65—75.

122. *Berezhkovskii, A. M.* Time and length scales for diffusion in liquids / A. M. Berezhkovskii, G. Sutmann // Phys. Rev. E. — 2002. — Т. 65. — С. 060201.

123. *Sasai, M.* Long time fluctuation of liquid water: 1/f spectrum of energy fluctuation in hydrogen bond network rearrangement dynamics / M. Sasai, I. Ohmine, R. Ramaswamy // The Journal of Chemical Physics. — 1992. — Т. 96, № 4. — С. 3045—3053. — DOI: 10.1063/1.461950.

124. *Shudo, A.* Origin of slow relaxation in liquid-water dynamics: A possible scenario for the presence of bottleneck in phase space / A. Shudo, K. Ichiki, S. Saito // EPL (Europhysics Letters). — 2006. — Т. 73, № 6. — С. 826.

125. *Berendsen, H. J. C.* Interaction Model for Water in Relation to Protein Hydration / H. J. C. Berendsen, J. Postma, W. van Gunsteren [и др.] // Intermolecular Forces / под ред. B. Pullman. — D. Reidel Publishing Company, Dordrecht, 1981. — С. 331—342.

126. *Van der Spoel, D.* GROMACS: Fast, flexible, and free / D. Van der Spoel, E. Lindahl, B. Hess [и др.] // Journal of Computational Chemistry. — 2005. — Т. 26, № 16. — С. 1701—1718.

127. *Berendsen, H. J. C.* Transport properties computed by linear response through weak coupling to a bath / H. J. C. Berendsen // Computer Simulations in Material Science / под ред. M. Meyer, V. Pontikis. — Kluwer, 1991. — 139 155.

128. *Christiansen, F.* Symbolic encoding in symplectic maps / F. Christiansen, A. Politi // Nonlinearity. — 1996. — Т. 9, № 6. — С. 1623.

129. *Buhl, M.* Statistically relaxing to generating partitions for observed time-series data / M. Buhl, M. B. Kennel // Physical Review E. — 2005. — Т. 71. — С. 046213.

130. *Crutchfield, J. P.* Inferring Statistical Complexity / J. P. Crutchfield, K. Young // Phys. Rev. Lett. — 1989. — Июль. — Т. 63, № 2. — С. 105—108.

131. *Bollt, E. M.* What symbolic dynamics do we get with a misplaced partition? : On the validity of threshold crossings analysis of chaotic time-series / E. M. Bollt, T. Stanford, Y.-C. Lai [и др.] // Physica D: Nonlinear Phenomena. — 2001. — Т. 154, № 3/4. — С. 259—286. — DOI: DOI: 10.1016/S0167-2789(01)00242-1.

132. *Theiler, J.* Testing for nonlinearity in time series: the method of surrogate data / J. Theiler, S. Eubank, A. Longtin [и др.] // Physica D: Nonlinear Phenomena. — 1992. — Т. 58, № 1—4. — С. 77—94. — DOI: DOI: 10.1016/0167-2789(92)90102-S.

133. *Zaslavsky, G. M.* Hamiltonian Chaos and Fractional Dynamics / G. M. Zaslavsky. — Oxford University Press, 2005.

134. *Zaslavsky, G. M.* Maxwell's demon as a dynamical model / G. M. Zaslavsky, M. Edelman // Phys. Rev. E. — 1997. — Нояб. — Т. 56, № 5. — С. 5310—5320. — DOI: 10.1103/PhysRevE.56.5310.

135. *Shojiguchi, A.* Fractional behavior in multidimensional Hamiltonian systems describing reactions / A. Shojiguchi, C.-B. Li, T. Komatsuzaki [и др.] // Phys. Rev. E. — 2007. — Нояб. — Т. 76, № 5. — С. 056205. — DOI: 10.1103/PhysRevE.76.056205.

136. *Afraimovich, V.* / V. Afraimovich, G. Zaslavsky // Chaos. — 2003. — Т. 13. — С. 519.

137. *Wackerbauer, R.* A comparative classification of complexity measures / R. Wackerbauer, A. Witt, H. Atmanspacher [и др.] // Chaos, Solitons & Fractals. — 1994. — Т. 4. — С. 133—173.

138. *Eckmann, J.-P.* Ergodic theory of chaos and strange attractors / J.-P. Eckmann, D. Ruelle // Rev. Mod. Phys. — 1985. — Т. 57. — С. 617—656.

139. *Zaslavsky, G.* Fractional Kinetics, and Anomalous Transport / G. Zaslavsky // Phys. Repts. — 2001. — T. 371. — C. 461—580.

140. *Afraimovich, V.* Fractal and miltifractal properties of exit times and Poincare recurrences / V. Afraimovich, G. Zaslavsky // Phys. Rev. E. — 1997. — T. 55. — C. 5418—5426.

141. *Lind, D.* An introduction to symbolic dynamics and coding / D. Lind, B. Marcus. — New York, NY, USA : Cambridge University Press, 1995.

142. *Shalizi, C. R.* What Is a Macrostate? From Subjective Measurements to Objective Dynamics / C. R. Shalizi, C. Moore // Studies in History and Philosophy of Modern Physics. — 2003. — T. submitted.

143. *Takens, F.* Detecting strange attractors in turbulence / F. Takens //. T. 898. — Springer Berlin / Heidelberg, 1981. — Гл. Detecting strange attractors in turbulence. C. 366—381. — (Lecture Notes in Mathematics). — DOI: 10.1007/BFb0091903.

144. *Shalizi, C. R.* Blind Construction of Optimal Nonlinear Recursive Predictors for Discrete Sequences / C. R. Shalizi, K. L. Shalizi // Uncertainty in Artificial Intelligence: Proceedings of the Twentieth Conference / под ред. M. Chickering, J. Halpern. — Arlington, Virginia : AUAI Press, 2004. — C. 504—511.

145. *Chirikov, B. V.* Asymptotic Statistics of Poincaré Recurrences in Hamiltonian Systems with Divided Phase Space / B. V. Chirikov, D. L. Shepelyansky // Phys. Rev. Lett. — 1999. — Янв. — T. 82, № 3. — C. 528—531. — DOI: 10.1103/PhysRevLett.82.528.

146. *Chirikov, B. V.* A Universal Instability of Many-Dimensional Oscillator Systems. / B. V. Chirikov // Phys. Rep. — 1979. — T. 52. — C. 264—379.

147. *Feldman, D. P.* The organization of intrinsic computation: Complexity-entropy diagrams and the diversity of natural information processing / D. P. Feldman, C. S. McTague, J. P. Crutchfield // Chaos: An Interdisciplinary Journal of Nonlinear Science. — 2008. — T. 18, № 4. — C. 043106. — DOI: 10.1063/1.2991106.

148. *Amigo, J. M.* Combinatorial detection of determinism in noisy time series / J. M. Amigo, S. Zambrano, M. A. F. Sanjuan // EPL (Europhysics Letters). — 2008. — T. 83, № 6. — C. 60005.

149. *Zaslavsky, G.* Chaos, fractional kinetics, and anomalous transport / G. Zaslavsky // Phys. Rep. — 2002. — T. 371. — C. 461—580.

150. *Misguich, J. H.* Motion in a stochastic layer described by symbolic dynamics / J. H. Misguich, J.-D. Reuss, Y. Elskens [и др.] // Chaos: An Interdisciplinary Journal of Nonlinear Science. — 1998. — T. 8, № 1. — C. 248—256. — DOI: 10.1063/1.166303.

151. *Baldovin, F.* Nonstandard entropy production in the standard map / F. Baldovin, C. Tsallis, B. Schulze // Physica A: Statistical Mechanics and its Applications. — 2003. — T. 320. — C. 184—192. — DOI: 10.1016/S0378-4371(02)01584-4.

152. *Boffetta, G.* Predictability: a way to characterize complexity / G. Boffetta, M. Cencini, M. Falcioni [и др.] // Physics Reports. — 2002. — T. 356, № 6. — C. 367—474. — DOI: 10.1016/S0370-1573(01)00025-4.

153. *Faber, N. T.* On time-dependent orbital complexity in gravitational N-body simulations / N. T. Faber, C. M. Boily, S. Portegies Zwart // Monthly Notices of the Royal Astronomical Society. — 2008. — Т. 386, № 1. — С. 425—439. — DOI: 10.1111/j.1365-2966.2008.13041.x.

154. *Kandrup, H. E.* Chaos, complexity, and short time Lyapunov exponents: two alternative characterisations of chaotic orbit segments / H. E. Kandrup, B. L. Eckstein, B. O. Bradley // Astron. Astrophys. — 1997. — Т. 320. — С. 65—73.

155. *Sideris, I. V.* Chaos and the continuum limit in the gravitational *N* -body problem. II. Nonintegrable potentials / I. V. Sideris, H. E. Kandrup // Phys. Rev. E. — 2002. — Июнь. — Т. 65, вып. 6. — С. 066203. — DOI: 10.1103/PhysRevE.65.066203.

156. *Shevchenko, I.* Isentropic perturbations of a chaotic domain / I. Shevchenko // Physics Letters A. — 2004. — Т. 333, № 5/6. — С. 408—414. — DOI: 10.1016/j.physleta.2004.10.066.

157. *Yuan, S.* Cryo-EM structure of a herpesvirus capsid at 3.1 Å / S. Yuan, J. Wang, D. Zhu [и др.] // Science. — 2018. — Т. 360, № 6384. — eaao7283.

158. *Larson, S. B.* Satellite tobacco mosaic virus RNA: structure and implications for assembly / S. B. Larson, A. McPherson // Current opinion in structural biology. — 2001. — Т. 11, № 1. — С. 59—65.

159. *Zink, M.* Mechanical properties of the icosahedral shell of southern bean mosaic virus: a molecular dynamics study / M. Zink, H. Grubmüller // Biophysical journal. — 2009. — Т. 96, № 4. — С. 1350—1363.

160. *Hsu, C.* Stabilizing effect of divalent metal ions on virions of southern bean mosaic virus / C. Hsu, O. Sehgal, E. Pickett // Virology. — 1976. — Т. 69, № 2. — С. 587—595.

161. *Speir, J. A.* Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy / J. A. Speir, S. Munshi, G. Wang [и др.] // Structure. — 1995. — Т. 3, № 1. — С. 63—78.

162. *Jones, T. A.* Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution / T. A. Jones, L. Liljas // Journal of molecular biology. — 1984. — Т. 177, № 4. — С. 735—767.

163. *Miao, Y.* Viral structural transition mechanisms revealed by multiscale molecular dynamics/order parameter extrapolation simulation / Y. Miao, P. J. Ortoleva // Biopolymers: Original Research on Biomolecules. — 2010. — Т. 93, № 1. — С. 61—73.

164. *Miao, Y.* Molecular dynamics/order parameter extrapolation for bionanosystem simulations / Y. Miao, P. J. Ortoleva // Journal of computational chemistry. — 2009. — Т. 30, № 3. — С. 423—437.

165. *Miao, Y.* All-atom multiscale simulation of cowpea chlorotic mottle virus capsid swelling / Y. Miao, J. E. Johnson, P. J. Ortoleva // The Journal of Physical Chemistry B. — 2010. — Т. 114, № 34. — С. 11181—11195.

166. *Roberts, J. A.* Investigation of a predicted N-terminal amphipathic α-helix using atomistic molecular dynamics simulation of a complete prototype poliovirus virion / J. A. Roberts, M. J. Kuiper, B. R. Thorley [и др.] // Journal of Molecular Graphics and Modelling. — 2012. — Т. 38. — С. 165—173.

167. *Kingsley, D. H.* Inactivation of hepatitis A virus and a calicivirus by high hydrostatic pressure / D. H. Kingsley, D. G. Hoover, E. Papafragkou [и др.] // Journal of food protection. — 2002. — Т. 65, № 10. — С. 1605—1609.

168. *Perilla, J. R.* Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations / J. R. Perilla, K. Schulten // Nature communications. — 2017. — Т. 8, № 1. — С. 15959.

169. *Hadden, J. A.* All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits / J. A. Hadden, J. R. Perilla, C. J. Schlicksup [и др.] // Elife. — 2018. — Т. 7. — e32478.

170. *Jiang, J.* Early stage P22 viral capsid self-assembly mediated by scaffolding protein: Atom-resolved model and molecular dynamics simulation / J. Jiang, J. Yang, Y. V. Sereda [и др.] // The Journal of Physical Chemistry B. — 2015. — Т. 119, № 16. — С. 5156—5162.

171. *Moerman, P.* Kinetics versus thermodynamics in virus capsid polymorphism / P. Moerman, P. Van Der Schoot, W. Kegel // The Journal of Physical Chemistry B. — 2016. — Т. 120, № 26. — С. 6003—6009.

172. *Hadden, J. A.* All-atom virus simulations / J. A. Hadden, J. R. Perilla // Current opinion in virology. — 2018. — Т. 31. — С. 82—91.

173. *Kovacs, E. W.* Dual-surface-modified bacteriophage MS2 as an ideal scaffold for a viral capsid-based drug delivery system / E. W. Kovacs, J. M. Hooker, D. W. Romanini [и др.] // Bioconjugate chemistry. — 2007. — Т. 18, № 4. — С. 1140—1147.

174. *Licis, N.* Long-range translational coupling in single-stranded RNA bacteriophages: an evolutionary analysis / N. Licis, J. van Duin, Z. Balklava [и др.] // Nucleic acids research. — 1998. — Т. 26, № 13. — С. 3242—3246.

175. *Licis, N.* Forced retroevolution of an RNA bacteriophage / N. Licis, Z. Balklava, J. Van Duin // Virology. — 2000. — Т. 271, № 2. — С. 298—306.

176. *Koning, R. I.* Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ / R. I. Koning, J. Gomez-Blanco, I. Akopjana [и др.] // Nature communications. — 2016. — Т. 7, № 1. — С. 12524.

177. *Dai, X.* In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus / X. Dai, Z. Li, M. Lai [и др.] // Nature. — 2017. — Т. 541, № 7635. — С. 112—116.

178. *Golmohammadi, R.* The refined structure of bacteriophage MS2 at 2· 8 Å resolution / R. Golmohammadi, K. Valegård, K. Fridborg [и др.] // Journal of molecular biology. — 1993. — Т. 234, № 3. — С. 620—639.

179. *Abraham, M. J.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers / M. J. Abraham, T. Murtola, R. Schulz [и др.] // SoftwareX. — 2015. — Т. 1. — С. 19—25.

180. *Kabsch, W.* Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features / W. Kabsch, C. Sander // Biopolymers: Original Research on Biomolecules. — 1983. — Т. 22, № 12. — С. 2577—2637.

181. *Perkett, M. R.* The allosteric switching mechanism in bacteriophage MS2 / M. R. Perkett, D. T. Mirijanian, M. F. Hagan // The Journal of chemical physics. — 2016. — Т. 145, № 3. — С. 035101.

182. *Armanious, A.* Viruses at solid–water interfaces: a systematic assessment of interactions driving adsorption / A. Armanious, M. Aeppli, R. Jacak [и др.] // Environmental science & technology. — 2016. — Т. 50, № 2. — С. 732—743.

183. *Shao, F.* Toxicity of silver nanoparticles on virus / F. Shao // J. Mater. Environ. Sci. — 2014. — Т. 5. — С. 587—590.

184. *Perilla, J. R.* All-atom molecular dynamics of virus capsids as drug targets / J. R. Perilla, J. A. Hadden, B. C. Goh [и др.] // The journal of physical chemistry letters. — 2016. — Т. 7, № 10. — С. 1836—1844.

185. *Goh, B. C.* Atomic modeling of an immature retroviral lattice using molecular dynamics and mutagenesis / B. C. Goh, J. R. Perilla, M. R. England [и др.] // Structure. — 2015. — Т. 23, № 8. — С. 1414—1425.

186. *Larsson, D. S.* Screening for the location of RNA using the chloride ion distribution in simulations of virus capsids / D. S. Larsson, D. van der Spoel // Journal of Chemical Theory and Computation. — 2012. — Т. 8, № 7. — С. 2474—2483.

187. *Wieder, M.* Comparing pharmacophore models derived from crystal structures and from molecular dynamics simulations / M. Wieder, U. Perricone, T. Seidel [и др.] // Monatshefte für Chemie-Chemical Monthly. — 2016. — Т. 147. — С. 553—563.

188. *Perilla, J. R.* Molecular dynamics simulations of large macromolecular complexes / J. R. Perilla, B. C. Goh, C. K. Cassidy [и др.] // Current opinion in structural biology. — 2015. — Т. 31. — С. 64—74.

189. *Shaw, D. E.* Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer / D. E. Shaw, J. Grossman, J. A. Bank [и др.] // SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. — IEEE. 2014. — С. 41—53.

190. *Ohmura, I.* MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations / I. Ohmura, G. Morimoto, Y. Ohno [и др.] // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. — 2014. — Т. 372, № 2021. — С. 20130387.

191. *De Vivo, M.* Role of molecular dynamics and related methods in drug discovery / M. De Vivo, M. Masetti, G. Bottegoni [и др.] // Journal of medicinal chemistry. — 2016. — Т. 59, № 9. — С. 4035—4061.

192. *Rose, N.* Epidemiology and transmission of porcine circovirus type 2 (PCV2) / N. Rose, T. Opriessnig, B. Grasland [и др.] // Virus research. — 2012. — Т. 164, № 1/2. — С. 78—89.

193. *Segalés, J.* Porcine circovirus type 2 (PCV2) infections: clinical signs, pathology and laboratory diagnosis / J. Segalés // Virus research. — 2012. — Т. 164, № 1/2. — С. 10—19.

194. *Terada, T.* Comparative molecular dynamics simulation study of crystal environment effect on protein structure / T. Terada, A. Kidera // The Journal of Physical Chemistry B. — 2012. — Т. 116, № 23. — С. 6810—6818.

195. *Ccgi, M.* Molecular operating environment (MOE), 2013.08 / M. Ccgi // Chemical Computing Group Inc., Montreal. — 2016. — Т. 354.

196. *Duan, Y.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations / Y. Duan, C. Wu, S. Chowdhury [и др.] // Journal of computational chemistry. — 2003. — Т. 24, № 16. — С. 1999—2012.

197. *Wang, N.* In silico analysis of surface structure variation of PCV2 capsid resulting from loop mutations of its capsid protein (Cap) / N. Wang, Y. Zhan, A. Wang [и др.] // The Journal of general virology. — 2016. — Т. 97, № 12. — С. 3331.

198. *Berman, H. M.* The protein data bank, 1999– / H. M. Berman, J. Westbrook, Z. Feng [и др.]. — 2006.

199. *Hess, B.* GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation / B. Hess, C. Kutzner, D. Van Der Spoel [и др.] // Journal of chemical theory and computation. — 2008. — Т. 4, № 3. — С. 435—447.

200. *Humphrey, W.* VMD: visual molecular dynamics / W. Humphrey, A. Dalke, K. Schulten // Journal of molecular graphics. — 1996. — Т. 14, № 1. — С. 33—38.

201. *Baker, N. A.* Electrostatics of nanosystems: application to microtubules and the ribosome / N. A. Baker, D. Sept, S. Joseph [и др.] // Proceedings of the National Academy of Sciences. — 2001. — Т. 98, № 18. — С. 10037—10041.

202. *Todd, D.* Circoviruses: immunosuppressive threats to avian species: a review / D. Todd // Avian Pathology. — 2000. — Т. 29, № 5. — С. 373—394.

203. *Rosario, K.* Revisiting the taxonomy of the family Circoviridae: establishment of the genus Cyclovirus and removal of the genus Gyrovirus / K. Rosario, M. Breitbart, B. Harrach [и др.] // Archives of Virology. — 2017. — Т. 162. — С. 1447—1463.

204. *Mankertz, A.* Identification of a protein essential for replication of porcine circovirus / A. Mankertz, J. Mankertz, K. Wolf [и др.] // Journal of General Virology. — 1998. — Т. 79, № 2. — С. 381—384.

205. *Opriessnig, T.* Porcine circovirus type 2–associated disease: update on current terminology, clinical manifestations, pathogenesis, diagnosis, and intervention strategies / T. Opriessnig, X.-J. Meng, P. G. Halbur // Journal of Veterinary Diagnostic Investigation. — 2007. — Т. 19, № 6. — С. 591—615.

206. *Ramamoorthy, S.* Porcine circoviruses: a minuscule yet mammoth paradox / S. Ramamoorthy, X.-J. Meng // Animal Health Research Reviews. — 2009. — Т. 10, № 1. — С. 1—20.

207. *Segalés, J.* Porcine circovirus type 2 (PCV2) infections: clinical signs, pathology and laboratory diagnosis / J. Segalés // Virus research. — 2012. — Т. 164, № 1/2. — С. 10—19.

208. *Deng, Z.* Transmission of porcine circovirus type 2b (PCV2b) in Kunming mice / Z. Deng, A. Yuan, W. Luo [и др.] // Acta Veterinaria Hungarica. — 2013. — Т. 61, № 2. — С. 234—243.

209. *Li, L.* Possible cross-species transmission of circoviruses and cycloviruses among farm animals / L. Li, T. Shan, O. B. Soji [и др.] // Journal of General Virology. — 2011. — Т. 92, № 4. — С. 768—772.

210. *Halami, M. Y.* Susceptibility of calves to porcine circovirus-2 (PCV2) / M. Y. Halami, M. Freick, A. A. Shehata [и др.] // Veterinary microbiology. — 2014. — Т. 173, № 1/2. — С. 125—131.

211. *Wang, Z.* Fusion between fluid liposomes and intact bacteria: study of driving parameters and in vitro bactericidal efficacy / Z. Wang, Y. Ma, H. Khalil [и др.] // International journal of nanomedicine. — 2016. — Т. 11. — С. 4025.

212. *Wang, X.* Phylogenetic analysis of two goat-origin PCV2 isolates in China / X. Wang, W. Li, X. Xu [и др.] // Gene. — 2018. — Т. 651. — С. 57—61.

213. *Song, T.* First detection and phylogenetic analysis of porcine circovirus type 2 in raccoon dogs / T. Song, J. Hao, R. Zhang [и др.] // BMC veterinary research. — 2019. — Т. 15, № 1. — С. 1—7.

214. *Song, T.* First detection and genetic analysis of fox-origin porcine circovirus type 2 / T. Song, S. Zhang, J. Hao [и др.] // Transboundary and emerging diseases. — 2019. — Т. 66, № 1. — С. 1—6.

215. *Kiupel, M.* Viral replication and lesions in BALB/c mice experimentally inoculated with porcine circovirus isolated from a pig with postweaning multisystemic wasting disease / M. Kiupel, G. Stevenson, J. Choi [и др.] // Veterinary Pathology. — 2001. — Т. 38, № 1. — С. 74—82.

216. *Hattermann, K.* Infection studies on human cell lines with porcine circovirus type 1 and porcine circovirus type 2 / K. Hattermann, C. Roedner, C. Schmitt [и др.] // Xenotransplantation. — 2004. — Т. 11, № 3. — С. 284—294.

217. *Pinheiro, A. L. B. C.* Verification of natural infection of peridomestic rodents by PCV2 on commercial swine farms / A. L. B. C. Pinheiro, L. H. S. Bulos, T. S. Onofre [и др.] // Research in veterinary science. — 2013. — Т. 94, № 3. — С. 764—768.

218. *Liu, X.* Immunogenicity evaluation of inactivated virus and purified proteins of porcine circovirus type 2 in mice / X. Liu, T. Ouyang, T. Ma [и др.] // BMC veterinary research. — 2018. — Т. 14. — С. 1—7.

219. *Liu, X.* Human cells are permissive for the productive infection of porcine circovirus type 2 in vitro / X. Liu, T. Ouyang, H. Ouyang [и др.] // Scientific Reports. — 2019. — Т. 9, № 1. — С. 5638.

220. *Tischer, I.* Characterization of papovavirus and picornavirus-like particles in permanent pig kidney cell lines. / I. Tischer, R. Rasch, G. Tochtermann [и др.] // Zenibl. Bukt. — 1974. — Т. 226, № 2. — С. 153—67.

221. *Tischer, I.* A very small porcine virus with circular single-stranded DNA / I. Tischer, H. Gelderblom, W. Vettermann [и др.] // Nature. — 1982. — Т. 295, № 5844. — С. 64—66.

222. *Palinski, R.* A novel porcine circovirus distantly related to known circoviruses is associated with porcine dermatitis and nephropathy syndrome and reproductive failure / R. Palinski, P. Piñeyro, P. Shang [и др.] // Journal of virology. — 2017. — Т. 91, № 1. — e01879—16.

223. *Ekser, B.* Xenotransplantation of solid organs in the pig-to-primate model / B. Ekser, P. Rigotti, B. Gridelli [и др.] // Transplant immunology. — 2009. — Т. 21, № 2. — С. 87—92.

224. *Pierson III, R. N.* Current status of xenotransplantation and prospects for clinical application / R. N. Pierson III, A. Dorling, D. Ayares [и др.] // Xenotransplantation. — 2009. — Т. 16, № 5. — С. 263—280.

225. *Denner, J.* Porcine circoviruses and xenotransplantation / J. Denner, A. Mankertz // Viruses. — 2017. — Т. 9, № 4. — С. 83.

226. *Duffy, S.* Rates of evolutionary change in viruses: patterns and determinants / S. Duffy, L. A. Shackelton, E. C. Holmes // Nature Reviews Genetics. — 2008. — Т. 9, № 4. — С. 267—276.

227. *Xiao, C.-T.* Global molecular genetic analysis of porcine circovirus type 2 (PCV2) sequences confirms the presence of four main PCV2 genotypes and reveals a rapid increase of PCV2d / C.-T. Xiao, P. G. Halbur, T. Opriessnig // Journal of General Virology. — 2015. — Т. 96, Pt_7. — С. 1830—1841.

228. *Wang, D.* Structure-based design of porcine circovirus type 2 chimeric VLPs (cVLPs) displays foreign peptides on the capsid surface / D. Wang, S. Zhang, Y. Zou [и др.] // Frontiers in cellular and infection microbiology. — 2018. — Т. 8. — С. 232.

229. *Zhan, Y.* The carboxyl terminus of the porcine circovirus type 2 capsid protein is critical to virus-like particle assembly, cell entry, and propagation / Y. Zhan, W. Yu, X. Cai [и др.] // Journal of Virology. — 2020. — Т. 94, № 9. — e00042—20.

230. *Khayat, R.* Structural characterization of the PCV2d virus-like particle at 3.3 Å resolution reveals differences to PCV2a and PCV2b capsids, a tetranucleotide, and an N-terminus near the icosahedral 3-fold axes / R. Khayat, K. Wen, A. Alimova [и др.] // Virology. — 2019. — Т. 537. — С. 186—197.

231. *Shuai, J.* Mapping of the nuclear localization signals in open reading frame 2 protein from porcine circovirus type 1 / J. Shuai, W. Wei, L. Jiang [и др.] // Acta Biochimica et Biophysica Sinica. — 2008. — Т. 40, № 1. — С. 71—77.

232. *Dhindwal, S.* The arginines in the n-terminus of the porcine circovirus 2 virus-like particles are responsible for disrupting the membranes at neutral and acidic pH / S. Dhindwal, S. Feng, R. Khayat // Journal of molecular biology. — 2019. — Т. 431, № 17. — С. 3261—3274.

233. *Liu, Z.* 2.9 Å resolution cryo-EM 3D reconstruction of close-packed virus particles / Z. Liu, F. Guo, F. Wang [и др.] // Structure. — 2016. — Т. 24, № 2. — С. 319—328.

234. *Harrison, S.* Tomato bushy stunt virus at 2.9 Å resolution / S. Harrison, A. Olson, C. Schutt [и др.] // Nature. — 1978. — Т. 276. — С. 368—373.

235. *Crowther, R.* Comparison of the structures of three circoviruses: chicken anemia virus, porcine circovirus type 2, and beak and feather disease virus / R. Crowther, J. Berriman, W. Curran [и др.] // Journal of virology. — 2003. — Т. 77, № 24. — С. 13036—13041.

236. *Dhindwal, S.* Porcine circovirus 2 uses a multitude of weak binding sites to interact with heparan sulfate, and the interactions do not follow the symmetry of the capsid / S. Dhindwal, B. Avila, S. Feng [и др.] // Journal of virology. — 2019. — Т. 93, № 6. — e02222—18.

237. *Mo, X.* Structural roles of PCV2 capsid protein N-terminus in PCV2 particle assembly and identification of PCV2 type-specific neutralizing epitope / X. Mo, X. Li, B. Yin [и др.] // PLoS pathogens. — 2019. — Т. 15, № 3. — e1007562.

238. *Swails, J. M.* Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: implementation, testing, and validation / J. M. Swails, D. M. York, A. E. Roitberg // Journal of chemical theory and computation. — 2014. — Т. 10, № 3. — С. 1341—1352.

239. *Weeks, S. D.* Ligation independent cloning vectors for expression of SUMO fusions / S. D. Weeks, M. Drinker, P. J. Loll // Protein expression and purification. — 2007. — Т. 53, № 1. — С. 40—50.

240. *Carragher, B.* Leginon: an automated system for acquisition of images from vitreous ice specimens / B. Carragher, N. Kisseberth, D. Kriegman [и др.] // Journal of structural biology. — 2000. — Т. 132, № 1. — С. 33—45.

241. *Hornak, V.* Comparison of multiple Amber force fields and development of improved protein backbone parameters / V. Hornak, R. Abel, A. Okur [и др.] // Proteins: Structure, Function, and Bioinformatics. — 2006. — Т. 65, № 3. — С. 712—725.

242. *Ryckaert, J.* Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes / J. Ryckaert, G. Ciccotti, H. Berendsen // Journal of Computational Physics. — 1977. — Т. 23. — С. 327—341.

243. *Mongan, J.* Constant pH molecular dynamics in generalized Born implicit solvent / J. Mongan, D. A. Case, J. A. McCammon // Journal of computational chemistry. — 2004. — Т. 25, № 16. — С. 2038—2048.

244. *Gotz, A. W.* Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born / A. W. Gotz, M. J. Williamson, D. Xu [и др.] // Journal of chemical theory and computation. — 2012. — Т. 8, № 5. — С. 1542—1555.

245. *Le Grand, S.* SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations / S. Le Grand, A. W. Götz, R. C. Walker // Computer Physics Communications. — 2013. — Т. 184, № 2. — С. 374—380.

246. *Salomon-Ferrer, R.* Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald / R. Salomon-Ferrer, A. W. Gotz, D. Poole [и др.] // Journal of chemical theory and computation. — 2013. — Т. 9, № 9. — С. 3878—3888.

247. *Case, D.* AMBER 2018; 2018 / D. Case, I. Ben-Shalom, S. Brozell [и др.] // University of California, San Francisco. — 2018.

248. *Mermelstein, D. J.* pH-dependent conformational dynamics of beta-secretase 1: A molecular dynamics study / D. J. Mermelstein, J. A. McCammon, R. C. Walker // Journal of Molecular Recognition. — 2019. — Т. 32, № 3. — e2765.

249. *Roe, D. R.* PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data / D. R. Roe, T. E. Cheatham III // Journal of chemical theory and computation. — 2013. — Т. 9, № 7. — С. 3084—3095.

250. *Krissinel, E.* Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions / E. Krissinel, K. Henrick // Acta Crystallographica Section D: Biological Crystallography. — 2004. — Т. 60, № 12. — С. 2256—2268.

251. *Pettersen, E. F.* UCSF Chimera—a visualization system for exploratory research and analysis / E. F. Pettersen, T. D. Goddard, C. C. Huang [и др.] // Journal of computational chemistry. — 2004. — Т. 25, № 13. — С. 1605—1612.

252. *Krissinel, E.* Inference of macromolecular assemblies from crystalline state / E. Krissinel, K. Henrick // Journal of molecular biology. — 2007. — Т. 372, № 3. — С. 774—797.

253. *Wu, P.-C.* Characterization of porcine circovirus type 2 (PCV2) capsid particle assembly and its application to virus-like particle vaccine development / P.-C. Wu, W.-L. Lin, C.-M. Wu [и др.] // Applied microbiology and biotechnology. — 2012. — Т. 95. — С. 1501—1507.

254. *Franzo, G.* Phylodynamic analysis of porcine circovirus type 2 reveals global waves of emerging genotypes and the circulation of recombinant forms / G. Franzo, M. Cortey, J. Segalés [и др.] // Molecular phylogenetics and evolution. — 2016. — Т. 100. — С. 269—280.

255. *Schuette, C.* A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo / C. Schuette, A. Fischer, W. Huisinga [и др.] // J. Comput. Phys. — 1999. — Т. 151, № 1. — С. 146—168. — DOI: DOI: 10.1006/jcph.1999.6231.

256. *Swope, W. C.* Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory / W. C. Swope, J. W. Pitera, F. Suits // The Journal of Physical Chemistry B. — 2004. — Т. 108, № 21. — С. 6571—6581. — DOI: 10.1021/jp037421y.

257. *Swope, W. C.* Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and a β-Hairpin Peptide / W. C. Swope, J. W. Pitera, F. Suits [и др.] // The Journal of Physical Chemistry B. — 2004. — Т. 108, № 21. — С. 6582—6594. — DOI: 10.1021/jp037422q.

258. *Noe, F.* Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states / F. Noe, I. Horenko, C. Schuette [и др.] // The Journal of Chemical Physics. — 2007. — Т. 126, № 15. — С. 155102. — DOI: 10.1063/1.2714539.

259. *Chodera, J. D.* Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics / J. D. Chodera, N. Singhal, V. S. Pande [и др.] // The Journal of Chemical Physics. — 2007. — Т. 126, № 15. — С. 155101. — DOI: 10.1063/1.2714538.

260. *Park, S.* Validation of Markov state models using Shannon's entropy / S. Park, V. S. Pande // The Journal of Chemical Physics. — 2006. — Т. 124, № 5. — С. 054118. — DOI: 10.1063/1.2166393.

261. *Chodera, J. D.* Long-Time Protein Folding Dynamics from Short-Time Molecular Dynamics Simulations / J. D. Chodera, W. C. Swope, J. W. Pitera [и др.] // Multiscale Modeling & Simulation. — 2006. — Т. 5, № 4. — С. 1214—1226. — DOI: 10.1137/06065146X.

262. *Prinz, J.-H.* Markov models of molecular kinetics: Generation and validation / J.-H. Prinz, H. Wu, M. Sarich [и др.] // The Journal of Chemical Physics. — 2011. — Т. 134, № 17. — С. 174105. — DOI: 10.1063/1.3565032.

263. *Crutchfield, J. P.* Computation at the Onset of Chaos / J. P. Crutchfield, K. Young // Entropy, Complexity, and Physics of Information, SFI Studies in the Sciences of Complexity, VIII / под ред. by W. Zurek. — Addison-Wesley, Reading, Massachusetts, 1990.

264. *Crutchfield, J. P.* The calculi of emergence: computation, dynamics and induction / J. P. Crutchfield // Physica D. — 1994. — Авг. — Т. 75, № 1—3. — С. 11—54.

265. *Hess, B.* Hydration thermodynamic properties of amino acid analogues: A systematic comparison of biomolecular force fields and water models / B. Hess, N. F. A. van der Vegt // Journal of Physical Chemistry B. — 2006. — Т. 110, № 35. — С. 17616—17626.

266. *Oostenbrink, C.* Validation of the 53A6 GROMOS force field / C. Oostenbrink, T. A. Soares, N. F. A. van der Vegt [и др.] // European Biophysics Journal with Biophysics Letters. — 2005. — Т. 34, № 4. — С. 273—284.

267. *Jorgensen, W. L.* The Opls [optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin / W. L. Jorgensen, J. Tirado-Rives // J. Am. Chem. Soc. — 1988. — Т. 110, № 6. — С. 1657—1666. — DOI: 10.1021/ja00214a001.

268. *Berendsen, H. J. C.* Molecular dynamics with coupling to an external bath / H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren [и др.] // The Journal of Chemical Physics. — 1984. — Т. 81, № 8. — С. 3684—3690.

269. *Varn, D. P.* Discovering planar disorder in close-packed structures from x-ray diffraction: Beyond the fault model / D. P. Varn, G. S. Canright, J. P. Crutchfield // Phys. Rev. B. — 2002. — Нояб. — Т. 66, вып. 17. — С. 174110. — DOI: 10.1103/PhysRevB.66.174110.

270. *Varn, D. P.* Inferring planar disorder in close-packed structures *via* ε-machine spectral reconstruction theory: structure and intrinsic computation in zinc sulfide / D. P. Varn, G. S. Canright, J. P. Crutchfield // Acta Crystallographica Section B. — 2007. — Апр. — Т. 63, № 2. — С. 169—182. — DOI: 10.1107/S0108768106043084.

271. *Born, B.* Solvation Dynamics of Model Peptides Probed by Terahertz Spectroscopy. Observation of the Onset of Collective Network Motions / B. Born, H. Weingartner, E. Brandermann [и др.] // J. Am. Chem. Soc. — 2009. — Т. 131, № 10. — С. 3752—3755. — DOI: 10.1021/ja808997y. — PMID: 19275262.

272. *Pagnotta, S. E.* The Dynamical Behavior of Hydrated Glutathione: a Model for Protein-Water Interactions / S. E. Pagnotta, S. Cerveny, A. Alegria [и др.] // Phys. Chem. Chem. Phys. — 2010. — Т. 12, вып. 35. — С. 10512—10517. — DOI: 10.1039/C003493B.

273. *Johnson, Q.* Water's Contribution to the Energetic Roughness from Peptide Dynamics / Q. Johnson, U. Doshi, T. Shen [и др.] // J. Chem. Theory Comput. — 2010. — Т. 6, № 9. — С. 2591—2597. — DOI: 10.1021/ct100183s.

274. *Zhang, L.* Protein Hydration Dynamics and Molecular Mechanism of Coupled Water−Protein Fluctuations / L. Zhang, Y. Yang, Y.-T. Kao [и др.] // J. Am. Chem. Soc. — 2009. — Т. 131, № 30. — С. 10677—10691. — DOI: 10.1021/ja902918p.

275. *Ebbinghaus, S.* An Extended Dynamical Hydration Shell Around Proteins / S. Ebbinghaus, S. J. Kim, M. Heyden [и др.] // Proc. Natl. Acad. Sci. USA. — 2007. — Т. 104, № 52. — С. 20749—20752. — DOI: 10.1073/pnas.0709207104.

276. *Combet, S.* Further Evidence That Interfacial Water is The Main "driving Force" of Protein Dynamics: a Neutron Scattering Study on Perdeuterated C-phycocyanin / S. Combet, J.-M. Zanotti // Phys. Chem. Chem. Phys. — 2012. — Т. 14, вып. 14. — С. 4927—4934. — DOI: 10.1039/C2CP23725C.

277. *Limongelli, V.* Sampling Protein Motion and Solvent Effect During Ligand Binding / V. Limongelli, L. Marinelli, S. Cosconati [и др.] // Proc. Natl. Acad. Sci. USA. — 2012. — Т. 109, № 5. — С. 1467—1472. — DOI: 10.1073/pnas.1112181108.

278. *Gunsteren, W. F. van*. Biomolecular Simulation: The GROMOS96 Manual and User Guide / W. F. van Gunsteren, S. R. Billeter, A. A. Eising [и др.]. — Vdf Hochschulverlag AG an der ETH Z rich, Z rich, Switzerland, 1996.

279. *MacKay, R. S.* Transport In 3d Volume-Preserving Flows / R. S. MacKay // J. Nonlinear Scie. — 1994. — Т. 4, № 1. — С. 329—354. — DOI: 10.1007/BF02430637.

280. *Pechukas, P.* Statistical approximations in collision theory, in: Dynamics of molecular collisions Pt B / P. Pechukas ; под ред. M. WH. — Plenum, 1976.

281. *Chandler, D.* Statistical Mechanics of Isomerization Dynamics in Liquids and the Transition State Approximation / D. Chandler // J. Chem. Phys. — 1978. — Т. 68, № 6. — С. 2959—2970. — DOI: 10.1063/1.436049.

282. *Adrian, R. J.* Stochastic Estimation of Conditional Structure: a Review / R. J. Adrian // Appl. Sci. Res. — 1994. — Т. 53, вып. 3. — С. 291—303.

283. *Kerherve, F.* Jet turbulence characteristics associated with downstream and sideline sound emission / F. Kerherve, P. Jordan, J. Delville [и др.] // Proceeding of the 16th AIAA/CEAS Aeroacoustics Conference, AIAA, Reston, Virginia, USA. — 2010. — AIAA 2010—3965.

284. *Papoulis, A.* Probability, Random Variables and Stochastic Processes / A. Papoulis, S. Pillai. — McGraw Hill, 2002.

285. *Jain, N. K.* Effect of trehalose on protein structure / N. K. Jain, I. Roy // Protein Science. — 2009. — Т. 18. — С. 24—36. — DOI: 10.1002/pro.3.

286. *Haines, A. H.* Non-equivalence of D- and L-trehalose in stabilising alkaline phosphatase against freeze-drying and thermal stress. Is chiral recognition involved? / A. H. Haines // Org. Biomol. Chem. — 2006. — Т. 4. — С. 702—706. — DOI: 10.1039/b514473f.

287. *Heikal, A.* The stabilisation of purified, reconstituted P-glycoprotein by freeze drying with disaccharides / A. Heikal, K. Box, A. Rothnie [и др.] // Cryobiology. — 2009. — Т. 58, № 1. — С. 37—44. — DOI: DOI: 10.1016/j.cryobiol.2008.10.125.

288. *Hedoux, A.* Thermostabilization Mechanism of Bovine Serum Albumin by Trehalose / A. Hedoux, J. F. Willart, L. Paccou [и др.] // Journal Of Physical Chemistry B. — 2009. — Апр. — Т. 113, № 17. — С. 6119—6126.

289. *Liao, Y.-H.* Investigation of the stabilisation of freeze-dried lysozyme and the physical properties of the formulations / Y.-H. Liao, M. B. Brown, G. P. Martin // European Journal of Pharmaceutics and Biopharmaceutics. — 2004. — Т. 58, № 1. — С. 15—24. — DOI: DOI: 10.1016/j.ejpb.2004.03.020.

290. *Crowe, J. H.* Preservation of Membranes in Anhydrobiotic Organisms: The Role of Trehalose / J. H. Crowe, L. M. Crowe, D. Chapman // Science. — 1984. — Т. 223, № 4637. — С. 701—703. — DOI: 10.1126/science.223.4637.701.

291. *Crowe, L.* Is trehalose special for preserving dry biomaterials? / L. Crowe, D. Reid, J. Crowe // Biophys. J. — 1996. — Т. 71, № 4. — С. 2087—2093. — DOI: doi:10.1016/S0006-3495(96)79407-9.

292. *Cordone, L.* Harmonic behavior of trehalose-coated carbon-monoxy-myoglobin at high temperature / L. Cordone, M. Ferrand, E. Vitrano [и др.] // Biophysical Journal. — 1999. — Т. 76, № 2. — С. 1043—1047.

293. *Cordone, L.* Internal dynamics and protein-matrix coupling in trehalose-coated proteins / L. Cordone, G. Cottone, S. Giuffrida [и др.] // Biochimica et Biophysica Acta-Proteins and Proteomics. — 2005. — Июнь. — Т. 1749, № 2. — С. 252—281.

294. *Cordone, L.* Role of residual water hydrogen bonding in sugar/water/biomolecule systems: a possible explanation for trehalose peculiarity / L. Cordone, G. Cottone, S. Giuffrida // Journal of Physics-Condensed Matter. — 2007. — Т. 19, № 20. — С. 205110.

295. *Lerbret, A.* Low-frequency vibrational properties of lysozyme in sugar aqueous solutions: A Raman scattering and molecular dynamics simulation study / A. Lerbret, F. Affouard, P. Bordat [и др.] // Journal Of Chemical Physics. — 2009. — Дек. — Т. 131, № 24. — С. 245103.

296. *Cottone, G.* Molecular dynamics simulation of sucrose- and trehalose-coated carboxy-myoglobin / G. Cottone, S. Giuffrida, G. Ciccotti [и др.] // Proteins-Structure Function and Bioinformatics. — 2005. — Май. — Т. 59, № 2. — С. 291—302.

297. *Lins, R. D.* Trehalose-protein interaction in aqueous solution / R. D. Lins, C. S. Pereira, P. H. Hunenberger // Proteins-Structure Function and Bioinformatics. — 2004. — Апр. — Т. 55, № 1. — С. 177—186.

298. *Pereira, C. S.* Interaction of the disaccharide trehalose with a phospholipid bilayer: A molecular dynamics study / C. S. Pereira, R. D. Lins, I. Chandrasekhar [и др.] // Biophysical Journal. — 2004. — Апр. — Т. 86, № 4. — С. 2273—2285.

299. *Lerbret, A.* How Do Trehalose, Maltose, and Sucrose Influence Some Structural and Dynamical Properties of Lysozyme? Insight from Molecular Dynamics Simulations / A. Lerbret, P. Bordat, F. Affouard [и др.] // The Journal of Physical Chemistry B. — 2007. — Т. 111, № 31. — С. 9410—9420. — DOI: 10.1021/jp071946z.

300. *Lerbret, A.* Molecular dynamics simulations of lysozyme in water/sugar solutions / A. Lerbret, F. Affouard, P. Bordat [и др.] // Chemical Physics. — 2008. — Апр. — Т. 345, № 2/3. — С. 267—274.

301. *Sola-Penna, M.* Stabilization against thermal inactivation promoted by sugars on enzyme structure and function: Why is trehalose more effective than other sugars? / M. Sola-Penna, J. R. Meyer-Fernandes // Archives Of Biochemistry And Biophysics. — 1998. — Т. 360, № 1. — С. 10—14.

302. *Hill, D. R.* Extracellular polysaccharide of Nostoc commune (cyanobacteria) inhibits fusion of membrane vesicles during desiccation / D. R. Hill, T. W. Keenan, R. F. Helm [и др.] // Journal of Applied Phycology. — 1997. — Т. 9, № 3. — С. 237—248.

303. *Eroglu, A.* Intracellular trehalose improves the survival of cryopreserved mammalian cells / A. Eroglu, M. Russo, R. Bieganski [и др.] // Nature Biotechnology. — 2000. — Февр. — Т. 18, № 2. — С. 163—167.

304. *Robinson, C. H.* Cold adaptation in Arctic and Antarctic fungi / C. H. Robinson // New Phytologist. — 2001. — Т. 151, № 2. — С. 341—353.

305. *Allison, S. D.* Hydrogen bonding between sugar and protein is responsible for inhibition of dehydration-induced protein unfolding / S. D. Allison, B. Chang, T. W. Randolph [и др.] // Archives of Biochemistry and Biophysics. — 1999. — Т. 365, № 2. — С. 289—298.

306. *Miranker, A.* Detection of transient protein folding populations by mass spectrometry / A. Miranker, C. V. Robinson, S. E. Radford [и др.] // Science. — 1993. — Т. 262, № 5135. — С. 896—900. — DOI: 10.1126/science.8235611.

307. *Tsui, V.* Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate / V. Tsui, C. Garcia, S. Cavagnero [и др.] // Protein Science. — 1999. — Т. 8. — С. 45—49. — DOI: 10.1110/ps.8.1.45.

308. *Tehei, M.* Adaptation to high temperatures through macromolecular dynamics by neutron scattering / M. Tehei, G. Zaccai // FEBS Journal. — 2007. — Т. 274, № 16. — С. 4034—4043. — DOI: 10.1111/j.1742-4658.2007.05953.x.

309. *Zavodszky, P.* Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins / P. Zavodszky, J. Kardos, A. Svingor [и др.] // Proceedings of the National Academy of Sciences of the United States of America. — 1998. — Т. 95, № 13. — С. 7406—7411.

310. *Schwalbe, H.* A refined solution structure of hen lysozyme determined using residual dipolar coupling data / H. Schwalbe, S. B. Grimshaw, A. Spencer [и др.] // Protein Science. — 2001. — Т. 10, № 4. — С. 677—688. — DOI: 10.1110/ps.43301.

311. *Nezbeda, I.* A family of primitive models of water: three-, four and five-site models / I. Nezbeda, J. Slovak // Molecular Physics. — 1997. — Т. 90, № 3. — С. 353—372. — DOI: 10 . 1080 / 002689797172471.

312. *Lins, R. D.* A new GROMOS force field for hexopyranose-based carbohydrates / R. D. Lins, P. H. Hunenberger // Journal Of Computational Chemistry. — 2005. — Т. 26, № 13. — С. 1400—1412.

313. *Warshel, A.* Modeling electrostatic effects in proteins / A. Warshel, P. K. Sharma, M. Kato [и др.] // Biochimica et Biophysica Acta-Proteins and Proteomics. — 2006. — Нояб. — Т. 1764, № 11. — С. 1647—1676.

314. *Sampedro, J. G.* Trehalose-enzyme interactions result in structure stabilization and activity inhibition. The role of viscosity / J. G. Sampedro, S. Uribe // Molecular And Cellular Biochemistry. — 2004. — Т. 256, № 1/2. — С. 319—327.

315. *Johnson, L. N.* Structure Of Some Crystalline Lysozyme-Inhibitor Complexes Determined By X-Ray Analysis At 6 A Resolution / L. N. Johnson, D. C. Phillips // Nature. — 1965. — Т. 206, № 4986. — С. 761—&.

316. *Blake, C. C. F.* Structure Of Hen Egg-White Lysozyme - A 3-Dimensional Fourier Synthesis At 2a Resolution / C. C. F. Blake, D. F. Koenig, G. A. Mair [и др.] // Nature. — 1965. — Т. 206, № 4986. — С. 757—&.

317. *Kelly, J. A.* X-Ray Crystallography Of The Binding Of The Bacterial-Cell Wall Trisaccharide Nam-Nag-Nam To Lysozyme / J. A. Kelly, A. R. Sielecki, B. D. Sykes [и др.] // Nature. — 1979. — Т. 282, № 5741. — С. 875—878.

318. *Artymiuk, P. J.* Crystallographic Studies Of The Dynamic Properties Of Lysozyme / P. J. Artymiuk, C. C. F. Blake, D. E. P. Grace [и др.] // Nature. — 1979. — Т. 280, № 5723. — С. 563—568.

319. *Vocadlo, D. J.* Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate / D. J. Vocadlo, G. J. Davies, R. Laine [и др.] // Nature. — 2001. — Авг. — Т. 412, № 6849. — С. 835—838.

320. *Svergun, D. I.* Protein hydration in solution: Experimental observation by x-ray and neutron scattering / D. I. Svergun, S. Richard, M. H. J. Koch [и др.] // Proceedings Of The National Academy Of Sciences Of The United States Of America. — 1998. — Март. — Т. 95, № 5. — С. 2267—2272.

321. *Kamerzell, T. J.* The complex inter-relationships between protein flexibility and stability / T. J. Kamerzell, C. R. Middaugh // Journal of Pharmaceutical Sciences. — 2008. — Т. 97, № 9. — С. 3494—3517. — DOI: 10.1002/jps.21269.

322. *Doan-Nguyen, V.* The effects of cosolutes on protein dynamics: The reversal of denaturant-induced protein fluctuations by trimethylamine N-oxide / V. Doan-Nguyen, J. P. Loria // Protein Science. — 2007. — Т. 16, № 1. — С. 20—29. — DOI: 10.1110/ps.062393707.

323. *Dirama, T. E.* Coupling between lysozyme and trehalose dynamics: Microscopic insights from molecular-dynamics simulations / T. E. Dirama, J. E. Curtis, G. A. Carri [и др.] // The Journal of Chemical Physics. — 2006. — Т. 124, № 3. — С. 034901. — DOI: 10.1063/1.2159471.

324. *Ionov, R.* Sugar bioprotective effects on thermal denaturation of lysozyme: Insights from Raman scattering experiments and molecular dynamics simulation / R. Ionov, A. Hedoux, Y. Guinet [и др.] // Journal of Non-Crystalline Solids. — 2006. — Т. 352, № 42—49. — С. 4430—4436. — DOI: DOI : 10 . 1016 / j . jnoncrysol . 2006 . 01 . 105. — Proceedings of the 5th International Discussion Meeting on Relaxations in Complex Systems, 5th International Discussion Meeting on Relaxations in Complex Systems.

325. *Hedoux, A.* Analysis of Sugar Bioprotective Mechanisms on the Thermal Denaturation of Lysozyme from Raman Scattering and Differential Scanning Calorimetry Investigations / A. Hedoux, J.-F. Willart, R. Ionov [и др.] // The Journal of Physical Chemistry B. — 2006. — Т. 110, № 45. — С. 22886—22893. — DOI: 10.1021/jp061568i.

326. *Mouawad, L.* CO migration pathways in cytochrome P450cam studied by molecular dynamics simulations / L. Mouawad, C. Tetreau, S. Abdel-Azeim [и др.] // Protein Science. — 2007. — Т. 16, № 5. — С. 781—794. — DOI: 10.1110/ps.062374707.

327. *Kondrashov, D. A.* Nonequilibrium Dynamics Simulations of Nitric Oxide Release: Comparative Study of Nitrophorin and Myoglobin / D. A. Kondrashov, W. R. Montfort // The Journal of Physical Chemistry B. — 2007. — Т. 111, № 31. — С. 9244—9252. — DOI: 10.1021/jp071136n.

328. *Sheu, S.-Y.* Selectivity principle of the ligand escape process from a two-gate tunnel in myoglobin: Molecular dynamics simulation / S.-Y. Sheu // The Journal of Chemical Physics. — 2006. — Т. 124, № 15. — С. 154711. — DOI: 10.1063/1.2185626.

329. *Camacho, C. J.* Kinetics of Desolvation-Mediated Protein-Protein Binding / C. J. Camacho, S. Kimura, C. DeLisi [и др.] // Biophysical Journal. — 2000. — Т. 78, № 3. — С. 1094—1105. — DOI: 10.1016/S0006-3495(00)76668-9.

330. *Ruscio, J. Z.* Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin / J. Z. Ruscio, D. Kumar, M. Shukla [и др.] // Proceedings of the National Academy of Sciences. — 2008. — Т. 105, № 27. — С. 9204—9209. — DOI: 10.1073/pnas.0710825105.

331. *Shan, Y.* How Does a Drug Molecule Find Its Target Binding Site? / Y. Shan, E. T. Kim, M. P. Eastwood [и др.] // Journal of the American Chemical Society. — 2011. — Т. 133, № 24. — С. 9181—9183. — DOI: 10.1021/ja202726y.

332. *Buch, I.* Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations / I. Buch, T. Giorgino, G. De Fabritiis // Proceedings of the National Academy of Sciences. — 2011. — Т. 108, № 25. — С. 10184—10189. — DOI: 10.1073/pnas.1103547108.

333. *Berg, H.* Physics of chemoreception / H. Berg, E. Purcell // Biophysical journal. — 1977. — Т. 20, № 2. — С. 193—219. — DOI: doi:10.1016/S0006-3495(77)85544-6.

334. *Zhou, H.-X.* Rate theories for biologists / H.-X. Zhou // Quarterly Reviews of Biophysics. — 2010. — Т. 43, № 02. — С. 219—293. — DOI: 10.1017/S0033583510000120.

335. *Alsallaq, R.* Protein association with circular DNA: Rate enhancement by nonspecific binding / R. Alsallaq, H.-X. Zhou // JOURNAL OF CHEMICAL PHYSICS. — CIRCULATION & FULFILLMENT DIV, 2 HUNTINGTON QUADRANGLE, STE 1 N O 1, MELVILLE, NY 11747-4501 USA, 2008. — Март. — Т. 128, № 11. — DOI: {10.1063/1.2888996}.

336. *Holyst, R.* Reduction of dimensionality in a diffusion search process and kinetics of gene expression / R. Holyst, M. Blazejczyk, K. Burdzy [и др.] // PHYSICA A. — PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS, 2000. — Март. — Т. 277, № 1/2. — С. 71—82.

337. *Hawser, S.* Dihydrofolate reductase inhibitors as antibacterial agents / S. Hawser, S. Lociuro, K. Islam // Biochemical Pharmacology. — 2006. — Т. 71, № 7. — С. 941—948. — DOI: 10.1016/j.bcp.2005.10.052.

338. *Argyrou, A.* Mycobacterium tuberculosis dihydrofolate reductase is a target for isoniazid / A. Argyrou, M. W. Vetting, B. Aladegbami [и др.] // Nat Struct Mol Biol. — 2006. — Май. — Т. 13, № 5. — С. 408—413.

339. *Li, R.* Three-dimensional structure of M. tuberculosis dihydrofolate reductase reveals opportunities for the design of novel tuberculosis drugs / R. Li, R. Sirawaraporn, P. Chitnumsub [и др.] // Journal of Molecular Biology. — 2000. — Т. 295, № 2. — С. 307—323. — DOI: DOI: 10.1006/jmbi.1999.3328.

340. *Sawaya, M. R.* Loop and Subdomain Movements in the Mechanism of Escherichia coli Dihydrofolate Reductase: Crystallographic Evidence / M. R. Sawaya, J. Kraut // Biochemistry. — 1997. — Т. 36, № 3. — С. 586—603. — DOI: 10.1021/bi962337c.

341. *Jorgensen, W. L.* Comparison of simple potential functions for simulating liquid water / W. L. Jorgensen, J. Chandrasekhar, J. D. Madura [и др.] // The Journal of Chemical Physics. — 1983. — Т. 79, № 2. — С. 926—935. — DOI: 10.1063/1.445869.

342. *Case, D.* AMBER 8 / D. Case, T. Darden, T. Cheatham [и др.]. — University of California San Francisco, 2004.

343. *Narumi, T.* A 55 TFLOPS simulation of amyloid-forming peptides from yeast prion Sup35 with the special-purpose computer system MDGRAPE-3 / T. Narumi, Y. Ohno, N. Okimoto [и др.] // Proceedings of the 2006 ACM/IEEE conference on Supercomputing. — Tampa, Florida : ACM, 2006. — (SC '06). — DOI: 10.1145/1188455.1188506.

344. *Taiji, M.* MDGRAPE-3 chip: a 165 Gflops application specific LSI for molecular dynamics simulations / M. Taiji. — IEEE Computer Society. pp. in CD-ROM, 2004.

345. *Darden, T.* Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems / T. Darden, D. York, L. Pedersen // The Journal of Chemical Physics. — 1993. — Т. 98, № 12. — С. 10089—10092. — DOI: 10.1063/1.464397.

346. *Wang, J.* Development and testing of a general amber force field / J. Wang, R. M. Wolf, J. W. Caldwell [и др.] // Journal of Computational Chemistry. — 2004. — Т. 25, № 9. — С. 1157—1174. — DOI: 10.1002/jcc.20035.

347. *Zhou, H.-X.* Enhancement of Association Rates by Nonspecific Binding to DNA and Cell Membranes / H.-X. Zhou, A. Szabo // Phys. Rev. Lett. — 2004. — Окт. — Т. 93, вып. 17. — С. 178101. — DOI: 10.1103/PhysRevLett.93.178101.

348. *Arshad, N.* Investigations on Anticancer Potentials by DNA Binding and Cytotoxicity Studies for Newly Synthesized and Characterized Imidazolidine and Thiazolidine-Based Isatin Derivatives / N. Arshad, M. I. Mir, F. Perveen [и др.] // Molecules. — 2022. — Янв. — Т. 27, № 2. — С. 354. — DOI: 10.3390/molecules27020354.

349. *Minotti, G.* Anthracyclines: Molecular Advances and Pharmacologic Developments in Antitumor Activity and Cardiotoxicity / G. Minotti, P. Menna, E. Salvatorelli [и др.] // Pharmacological Reviews. — 2004. — Июнь. — Т. 56, № 2. — С. 185—229. — DOI: 10.1124/pr.56.2.6.

350. *Delgado, J. L.* Topoisomerases as anticancer targets / J. L. Delgado, C.-M. Hsieh, N.-L. Chan [и др.] // Biochemical Journal. — 2018. — Янв. — Т. 475, № 2. — С. 373—398. — DOI: 10.1042/bcj20160583.

351. *Sahoo, D.* Quest for Mode of Binding of 2-(4-(Dimethylamino)styryl)-1-methylpyridinium Iodide with Calf Thymus DNA / D. Sahoo, P. Bhattacharya, S. Chakravorti // The Journal of Physical Chemistry B. — 2010. — Янв. — Т. 114, № 5. — С. 2044—2050. — DOI: 10.1021/jp910766q.

352. *Bahner, C. T.* Quaternary Salts of Styryl Pyridines and Quinolines / C. T. Bahner, E. S. Pace, R. Prevost // Journal of the American Chemical Society. — 1951. — Июль. — Т. 73, № 7. — С. 3407—3408. — DOI: 10.1021/ja01151a120.

353. *Paponov, B. V.* Synthesis and cytotoxic activity investigations of DNA-tropic 4-dimethylaminostyryl derivatives of N-methyl-pyrimidobenzimidazolium iodides / B. V. Paponov, A. N. Sokolova, P. S. Zdereva [и др.] // The Scientific Notes of the Pavlov University. — 2024. — Авг. — Т. 31, № 3. — С. 79—88. — DOI: 10.24884/1607-4181-2024-31-3-79-88.

354. *Ross, G. A.* The maximal and current accuracy of rigorous protein-ligand binding free energy calculations / G. A. Ross, C. Lu, G. Scarabelli [и др.] // Communications Chemistry. — 2023. — Окт. — Т. 6, № 1. — DOI: 10.1038/s42004-023-01019-9.

355. *Heinzelmann, G.* BAT2: an Open-Source Tool for Flexible, Automated, and Low Cost Absolute Binding Free Energy Calculations / G. Heinzelmann, D. J. Huggins, M. K. Gilson // Journal of Chemical Theory and Computation. — 2024. — Авг. — Т. 20, № 15. — С. 6518—6530. — DOI: 10.1021/acs.jctc.4c00205.

356. *Mikolaichuk, O. V.* Biocompatibility and bioactivity study of a cytostatic drug belonging to the group of alkylating agents of the triazine derivative class / O. V. Mikolaichuk, V. V. Sharoyko, E. A. Popova [и др.] // Journal of Molecular Liquids. — 2021. — Дек. — Т. 343. — С. 117630. — DOI: 10.1016/j.molliq.2021.117630.

357. *Kaus, J. W.* How To Deal with Multiple Binding Poses in Alchemical Relative Protein–Ligand Binding Free Energy Calculations / J. W. Kaus, E. Harder, T. Lin [и др.] // Journal of Chemical Theory and Computation. — 2015. — Май. — Т. 11, № 6. — С. 2670—2679. — DOI: 10.1021/acs.jctc.5b00214.

358. *Decherchi, S.* Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation / S. Decherchi, A. Cavalli // Chem. Rev. — 2020. — Т. 120, № 23. — С. 12788—12833. — DOI: 10.1021/acs.chemrev.0c00534.

359. *Nunes-Alves, A.* Recent progress in molecular simulation methods for drug binding kinetics / A. Nunes-Alves, D. B. Kokh, R. C. Wade // Curr. Opin. Struct. Biol. — 2020. — Т. 64. — С. 126—133. — DOI: 10.1016/j.sbi.2020.06.022.

360. *Ahalawat, N.* An appraisal of computer simulation approaches in elucidating biomolecular recognition pathways / N. Ahalawat, J. Mondal // J. Phys. Chem. Lett. — 2021. — Т. 12. — С. 633—641. — DOI: 10.1021/acs.jpclett.0c02785.

361. *Bruce, N. J.* New approaches for computing ligand–receptor binding kinetics / N. J. Bruce, G. K. Ganotra, D. B. Kokh [и др.] // Curr. Opin. Struct. Biol. — 2018. — Апр. — Т. 49. — С. 1—10. — DOI: 10.1016/j.sbi.2017.10.001.

362. *Mapplebeck, S.* Simulation of protein pulling dynamics on second time scale with boxed molecular dynamics / S. Mapplebeck, J. Booth, D. Shalashilin // J. Chem. Phys. — 2021. — Т. 155, № 8. — С. 085101. — DOI: 10.1063/5.0059321.

363. *Wolf, S.* Multisecond ligand dissociation dynamics from atomistic simulations / S. Wolf, B. Lickert, S. Bray [и др.] // Nat. Commun. — 2020. — Июнь. — Т. 11, № 1. — DOI: 10.1038/s41467-020-16655-1.

364. *Capelli, R.* Accuracy of molecular simulation-based predictions of k_off values: a metadynamics study / R. Capelli, W. Lyu, V. Bolnykh [и др.] // J. Phys. Chem. Lett. — 2020. — Т. 11, № 15. — С. 6373—6381. — DOI: 10.1021/acs.jpclett.0c00999.

365. *Ribeiro, J. M. L.* Kinetics of Ligand–Protein Dissociation from All-Atom Simulations: Are We There Yet? / J. M. L. Ribeiro, S.-T. Tsai, D. Pramanik [и др.] // Biochemistry. — 2018. — Дек. — Т. 58, № 3. — С. 156—165. — DOI: 10.1021/acs.biochem.8b00977.

366. *Kokh, D. B.* A workflow for exploring ligand dissociation from a macromolecule: Efficient random acceleration molecular dynamics simulation and interaction fingerprint analysis of ligand trajectories / D. B. Kokh, B. Doser, S. Richter [и др.] // J. Chem. Phys. — 2020. — Сент. — Т. 153, № 12. — С. 125102. — DOI: 10.1063/5.0019088.

367. *Kokh, D. B.* Estimation of Drug-Target Residence Times by τ-Random Acceleration Molecular Dynamics Simulations / D. B. Kokh, M. Amaral, J. Bomke [и др.] // J. Chem. Theory Comput. — 2018. — Май. — Т. 14, № 7. — С. 3859—3869. — DOI: 10.1021/acs.jctc.8b00230.

368. *Unissa, A. N.* Overview on mechanisms of isoniazid action and resistance in Mycobacterium tuberculosis / A. N. Unissa, S. Subbian, L. E. Hanna [и др.] // Infect. Genet. Evol. — 2016. — Нояб. — Т. 45. — С. 474—492. — DOI: 10.1016/j.meegid.2016.09.004.

369. *Jagger, B. R.* Quantitative ranking of ligand binding kinetics with a multiscale milestoning simulation approach / B. R. Jagger, C. T. Lee, R. E. Amaro // J. Phys. Chem. Lett. — 2018. — Т. 9. — С. 4941—4948. — DOI: 10.1021/acs.jpclett.8b02047.

370. *Kramers, H. A.* Brownian motion in a field of force and the diffusion model of chemical reactions / H. A. Kramers // Physica. — 1940. — Т. 7, № 4. — С. 284—304. — DOI: DOI: 10.1016/S0031-8914(40)90098-2.

371. *Hänggi, P.* Reaction-rate theory: fifty years after Kramers / P. Hänggi, P. Talkner, M. Borkovec // Rev. Mod. Phys. — 1990. — Т. 62. — С. 251. — DOI: 10.1103/RevModPhys.62.251.

372. *Ebeling, W.* Stochastic dynamics of reacting biomolecules / W. Ebeling, L. Schimansky-Geier, Y. M. Romanovsky. — World Scientific (Singapore), 2002.

373. *Pollak, E.* Reaction rate theory: What it was, where is it today, and where is it going? / E. Pollak, P. Talkner // Chaos. — 2005. — Т. 15. — С. 026116. — DOI: 10.1063/1.1858782.

374. *Bernetti, M.* Kinetics of drug binding and residence time / M. Bernetti, M. Masetti, W. Rocchia [и др.] // Ann. Phys. Chem. — 2019. — Т. 70. — С. 143—171. — DOI: 10.1146/annurev-physchem-042018-052340.

375. *Dybiec, B.* Estimation of the smallest eigenvalue in fractional escape problems: Semi-analytics and fits / B. Dybiec, I. M. Sokolov // Comput. Phys. Commun. — 2015. — Февр. — Т. 187. — С. 29—37. — DOI: 10.1016/j.cpc.2014.10.007.

376. *Toenjes, R.* Nonspectral Relaxation in One Dimensional Ornstein-Uhlenbeck Processes / R. Toenjes, I. M. Sokolov, E. B. Postnikov // Phys. Rev. Lett. — 2013. — Апр. — Т. 110, № 15. — С. 150602. — DOI: 10.1103/physrevlett.110.150602.

377. *Thiel, F.* Nonspectral modes and how to find them in the Ornstein-Uhlenbeck process with white μ-stable noise / F. Thiel, I. M. Sokolov, E. B. Postnikov // Phys. Rev. E. — 2016. — Май. — Т. 93, № 5. — С. 052104. — DOI: 10.1103/physreve.93.052104.

378. *Lin, H.-J.* Bell's expression and the generalized Garg form for forced dissociation of a biomolecular complex / H.-J. Lin, H.-Y. Chen, Y.-J. Sheng [и др.] // Phys. Rev. Lett. — 2007. — Т. 98. — С. 088304. — DOI: 10.1103/PhysRevLett.98.088304.

379. *Friddle, R. W.* Unified model of dynamic forced barrier crossing in single molecules / R. W. Friddle // Phys. Rev. Lett. — 2008. — Т. 100. — С. 138302. — DOI: 10.1103/PhysRevLett.100.138302.

380. *Abkenar, M.* Dissociation rates from single-molecule pulling experiments under large thermal fluctuations or large applied force / M. Abkenar, T. H. Gray, A. Zaccone // Phys. Rev. E. — 2017. — Т. 95, № 4. — С. 042413. — DOI: 10.1103/PhysRevE.95.042413.

381. *Scacchi, A.* Escape rate of transiently active brownian particle in one dimension / A. Scacchi, J. M. Brader, A. Sharma // Phys. Rev. E. — 2019. — Т. 100. — С. 012601. — DOI: 10.1103/PhysRevE.100.012601.

382. *Bell, G. I.* Models for the specific adhesion of cells to cells / G. I. Bell // Science. — 1978. — Т. 200. — С. 618—627. — DOI: 10.1126/science.347575.

383. *Pereverzev, Y. V.* Force-induced deformations and stability of biological bonds / Y. V. Pereverzev, O. V. Prezhdo // Phys. Rev. E. — 2006. — Т. 73. — С. 050902. — DOI: 10.1103/PhysRevE.73.050902.

384. *Prezhdo, O. V.* Theoretical aspects of the biological catch bond / O. V. Prezhdo, Y. V. Pereverzev // Acc. Chem. Res. — 2009. — Т. 42. — С. 693—703. — DOI: 10.1021/ar800202z.

385. *Singh, R.* Comparative study of catalase-peroxidases (KatGs) / R. Singh, B. Wiseman, T. Deemagarn [и др.] // Arch. Biochem. Biophys. — 2008. — Т. 471, № 2. — С. 207—214. — DOI: 10.1016/j.abb.2007.12.008.

386. Protein Data Bank: the single global archive for 3D macromolecular structure data // Nucleic Acids Res. — 2019. — Т. 47, № D1. — С. D520—D528. — DOI: 10.1093/nar/gky949.

387. *Pettersen, E. F.* UCSF Chimera—a visualization system for exploratory research and analysis / E. F. Pettersen, T. D. Goddard, C. C. Huang [и др.] // J. Comput. Chem. — 2004. — Т. 25, № 13. — С. 1605—1612. — DOI: 10.1002/jcc.20084.

388. *Lafarge, T.* The NIST uncertainty machine / T. Lafarge, A. Possolo // NCSLI Measure. — 2015. — Т. 10. — С. 20—27. — DOI: 10.1080/19315775.2015.11721732.

389. *Chattopadhyay, A. K.* Close contact fluctuations: The seeding of signalling domains in the immunological synapse / A. K. Chattopadhyay, N. J. Burroughs // Europhys. Lett. — 2007. — Февр. — Т. 77, № 4. — С. 48003. — DOI: 10.1209/0295-5075/77/48003.

390. *Bush, D. R.* Contact time periods in immunological synapse / D. R. Bush, A. K. Chattopadhyay // Phys. Rev. E. — 2014. — Окт. — Т. 90, вып. 4. — С. 042706. — DOI: 10.1103/PhysRevE.90.042706.

391. *Bush, D. R.* Temporal dynamics in an immunological synapse: Role of thermal fluctuations in signaling / D. R. Bush, A. K. Chattopadhyay // Phys. Rev. E. — 2015. — Июль. — Т. 92, вып. 1. — С. 012706. — DOI: 10.1103/PhysRevE.92.012706.

392. *Shalizi, C.* Quantifying self-organization with optimal predictors / C. Shalizi, K. Shalizi, R. Haslinger // Physical Review Letters. — 2004. — Т. 93, № 11. — С. 118701-1 —118701-4.

393. *MacKerell Jr, A. D.* All-atom empirical potential for molecular modeling and dynamics studies of proteins / A. D. MacKerell Jr, D. Bashford, M. Bellott [и др.] // J. Phys. Chem. B. — 1998. — Т. 102, № 18. — С. 3586—3616. — DOI: 10.1021/jp973084f.

394. *Zoete, V.* SwissParam: a fast force field generation tool for small organic molecules / V. Zoete, M. A. Cuendet, A. Grosdidier [и др.] // J. Comput. Chem. — 2011. — Т. 32, № 11. — С. 2359—2368. — DOI: 10.1002/jcc.21816.

Публикации по теме диссертации

1. *Nerukh, D.* Water-Peptide Dynamics during Conformational Transitions / D. Nerukh, S. Karabasov // The Journal of Physical Chemistry Letters. — 2013. — Т. 4, № 5. — С. 815—819. — DOI: 10.1021/jz400051p (IF 6,7).

2. *Markesteijn, A.* Concurrent multiscale modelling of atomistic and hydrodynamic processes in liquids / A. Markesteijn, S. Karabasov, A. Scukins [и др.] // Phil. Trans. R. Soc. — 2014. — Т. A 372, № 2021. — С. 20130379. — DOI: 10.1098/rsta.2013.0379 (IF 2,77).

3. *Pavlov, E.* Visualising and controlling the flows in biomolecular systems at and between multiple scales: from atoms to hydrodynamics at different locations in time and space / E. Pavlov, M. Taiji, A. Scukins [и др.] // Faraday Discussions. — 2014. — Т. 169, № 1. — С. 285—302. — DOI: 10.1039/c3fd00159h (IF 3,46).

4. *Korotkin, I.* Two-phase flow analogy as an effective boundary condition for modelling liquids at atomistic resolution / I. Korotkin, D. Nerukh, E. Tarasova [и др.] // Journal of Computational Science. — 2016. — Т. 17, Part 2. — С. 446—456. — DOI: 10.1016/j.jocs.2016.03.012 (IF 1,67).

5. *Scukins, A.* Multiscale Molecular Dynamics/Hydrodynamics implementation of two dimensional ‘Mercedes Benz’ water model / A. Scukins, D. Nerukh, E. Pavlov [и др.] // European Physical Journal E. — 2015. — Т. 224, № 12. — С. 2217—2238. — DOI: 10.1140/epjst/e2015-02409-8 (IF 1,99).

6. *Scukins, A.* Molecular Dynamics implementation of BN2D or ‘Mercedes Benz’ water model / A. Scukins, V. Bardik, E. Pavlov [и др.] // Computer Physics Communications. — 2015. — Т. 190. — С. 129—138. — DOI: 10.1016/j.cpc.2014.12.019 (IF 3,09).

7. *Tarasova, E.* All-Atom Molecular Dynamics Simulations of Entire Virus Capsid Reveal the Role of Ion Distribution in Capsid’s Stability / E. Tarasova, V. Farafonov, R. Khayat [и др.] // The Journal of Physical Chemistry Letters. — 2017. — Т. 8, № 4. — С. 779—784. — DOI: 10.1021/acs.jpclett.6b02759 (IF 8,71).

8. *Korotkin, I.* A Hybrid Molecular Dynamics/Fluctuating Hydrodynamics Method for Modelling Liquids at Multiple Scales in Space and Time / I. Korotkin, S. Karabasov, D. Nerukh [и др.] // The Journal of Chemical Physics. — 2015. — Т. 143. — С. 014110. — DOI: 10.1063/1.4923011 (IF 2, 96).

9. *Bardik, V.* Equation of state for water in the small compressibility region / V. Bardik, D. Nerukh, E. Pavlov [и др.] // Ukrainian Journal of Physics. — 2012. — Т. 57, № 1. — С. 49—53.

10. *Bardik, V.* Free energy functional expansion as the generalized approach to the equation of state of dense fluids / V. Bardik, D. Nerukh, E. Pavlov [и др.] // Ukrainian Journal of Physics. — 2012. — Т. 57, № 6. — С. 612—618. — DOI: 10.15407/ujpe57.6.612 (IF 0,44).

11. *Karabasov, S.* Multiscale modelling: approaches and challenges / S. Karabasov, D. Nerukh, A. Hoekstra [и др.] // Phil. Trans. R. Soc. — 2014. — Т. A 372. — С. 20130390. — DOI: 10.1098/rsta.2013.0390 (IF 2,77).

12. *Nerukh, D.* Quantifying the complexity of chaos in multibasin multidimensional dynamics of molecular systems / D. Nerukh, G. Karvounis, R. Glen // Complexity. — 2004. — Т. 10, вып. 2, № 2. — С. 40—46. — DOI: 10.1002/cplx.20055 (IF 1,17).

13. *Luzanov, A.* Complexity and chirality indices for molecular informatics: differential geometry approach / A. Luzanov, D. Nerukh // Functional materials. — 2005. — Т. 12, № 1. — С. 55—64.

14. *Luzanov, A.* Simple One-electron Invariants of Molecular Chirality / A. Luzanov, D. Nerukh // Journal of Mathematical Chemistry. — 2007. — Т. 41, № 4. — С. 417—435. — DOI: 10.1007/s10910-006-9087-8 (IF 1,25).

15. *Nerukh, D.* Computational mechanics reveals nanosecond time correlations in molecular dynamics of liquid systems / D. Nerukh // Chemical Physics Letters. — 2008. — Т. 457, № 4—6. — С. 439—443. — DOI: 10.1016/j.cplett.2008.04.043 (IF 2,16).

16. *Nerukh, D.* Complex Temporal Patterns in Molecular Dynamics: a Direct Measure of the Phase Space Exploration by the Trajectory at Macroscopic Time Scales / D. Nerukh, V. Ryabov, R. C. Glen // Phys. Rev. E. — 2008. — Т. 77. — С. 036225. — DOI: 10.1103/PhysRevE.77.036225 (IF 2,56).

17. *Nerukh, D.* Molecular phase space transport in water: Non-stationary random walk model / D. Nerukh, V. Ryabov, M. Taiji // Physica A: Statistical Mechanics and its Applications. — 2009. — Т. 388, № 22. — С. 4719—4726. — DOI: 10.1016/j.physa.2009.07.041 (IF 1,55).

18. *Ryabov, V.* Quantifying long time memory in phase space trajectories of molecular liquids / V. Ryabov, D. Nerukh // Journal of Molecular Liquids. — 2011. — Т. 159, № 1. — С. 99—104. — DOI: 0.1016/j.molliq.2010.11.016 (IF 2,14).

19. *Nerukh, D.* Computational Mechanics of Molecular Systems / D. Nerukh, V. Ryabov // Computational Mechanics Research Trends / под ред. H. P. Berger. — Nova Science Publishers, 2010. — С. 219—247. — (Computer Science, Technology and Applications).

20. *Ryabov, V.* Computational mechanics of molecular systems: Quantifying high-dimensional dynamics by distribution of Poincare recurrence times / V. Ryabov, D. Nerukh // Chaos: An Interdisciplinary Journal of Nonlinear Science. — 2011. — Т. 21, № 3. — С. 037113. — DOI: 10.1063/1.3608125 (IF 2,05).

21. *Ryabov, V.* Statistical complexity of low and high dimensional systems / V. Ryabov, D. Nerukh // Journal of Atomic, Molecular, and Optical Physics. — 2012. — Т. 2012. — С. 589651. — DOI: doi:10.1155/2012/589651.

22. *Tarasova, E.* Complete virus capsid at all-atom resolution: Simulations using molecular dynamics and hybrid molecular dynamics/hydrodynamics methods reveal semipermeable membrane function / E. Tarasova, I. Korotkin, V. Farafonov [и др.] // Journal of Molecular Liquids. — 2017. — Т. 245, Supplement C. — С. 109—114. — DOI: 10.1016/j.molliq.2017.06.124 (IF 4,51).

23. *Zhuravlova, Z. Y.* On stress state of the PCV2 virus / Z. Y. Zhuravlova, D. A. Nerukh, V. V. Reut [и др.] // Bulletin of Taras Shevchenko National University of Kyiv Series: Physics & Mathematics. — 2017. — Т. 3. — С. 59—61.

24. *Zhuravlova, Z.* Investigation of idealized virus capsid model with the dynamic elasticity apparatus / Z. Zhuravlova, D. Nerukh, V. Reut [и др.] // Journal of Mathematical Sciences. — 2017. — Т. 60, № 2. — С. 92—104.

25. *Tarasova, E.* Details of charge distribution in stable viral capsid / E. Tarasova, V. Farafonov, M. Taiji [и др.] // Journal of Molecular Liquids. — 2018. — Т. 265. — С. 585—591. — DOI: 0.1016/j.molliq.2018.06.019 (IF 4,51).

26. *Tarasova, E.* Molecular dynamics study of the role of ions in maintaining virus capsid stability / E. Tarasova, V. Farafonov, D. Nerukh // Ensemble. — 2019. — Т. 21, № 1. — С. 50—54. — DOI: https://doi.org/10.11436/mssj.21.50.

27. *Tarasova, E.* All-Atom Molecular Dynamics Simulations of Whole Viruses / E. Tarasova, D. Nerukh // The Journal of Physical Chemistry Letters. — 2018. — Т. 9. — С. 5805—5809. — DOI: 10.1021/acs.jpclett.8b02298 (IF 7,33).

28. *Farafonov, V. S.* Estimation of Nanoparticle's Surface Electrostatic Potential in Solution Using Acid-Base Molecular Probes I: In Silico Implementation for Surfactant Micelles / V. S. Farafonov, A. V. Lebed, D. A. Nerukh [и др.] // The Journal of Physical Chemistry B. — 2023. — Т. 126, № 4. — С. 1022—1030. — DOI: 10.1021/acs.jpcb.2c07012 (IF 2,77).

29. *Vodolazkaya, N.* Estimation of Nanoparticle's Surface Electrostatic Potential in Solution Using Acid-Base Molecular Probes. III. Experimental Hydrophobicity/Hydrophilicity and Charge Distribution of MS2 Virus Surface / N. Vodolazkaya, M. Nikolskaya, A. Laguta [и др.] // The Journal of Physical Chemistry B. — 2022. — Т. 126, № 41. — С. 8166—8176. — DOI: 10.1021/acs.jpcb.2c04491 (IF 2,77).

30. *Vodolazkaya, N.* Influence of various colloidal surfactants on the stability of MS2 bacteriophage suspension. The charge distribution on the PCV2 virus surface / N. Vodolazkaya, A. Laguta, V. Farafonov [и др.] // Journal of Molecular Liquids. — 2023. — Т. 387. — С. 122644. — DOI: 0.1016/j.molliq.2023.122644 (IF 5,69).

31. *Nerukh, D.* Complexity of classical dynamics of molecular systems. I. Methodology / D. Nerukh, G. Karvounis, R. C. Glen // The Journal of Chemical Physics. — 2002. — Т. 117, № 21. — С. 9611—9617. — DOI: 10.1063/1.1518010 (IF 3,08).

32. *Nerukh, D.* Complexity of classical dynamics of molecular systems. II. Finite statistical complexity of a water–Na+ system / D. Nerukh, G. Karvounis, R. C. Glen // The Journal of Chemical Physics. — 2002. — Т. 117, № 21. — С. 9618—9622. — DOI: 10.1063/1.1518011 (IF 3,08).

33. *Nerukh, D.* Why are MD simulated protein folding times wrong? / D. Nerukh // Advances in Computational Biology. Т. 680 / под ред. H. R. Arabnia. — Springer, 2010. — (Advances in Experimental Medicine and Biology).

34. *Nerukh, D.* Identifying and correcting non-Markov states in peptide conformational dynamics / D. Nerukh, C. H. Jensen, R. C. Glen // The Journal of Chemical Physics. — 2010. — Т. 132, № 8. — С. 084104. — DOI: 10.1063/1.3328781 (IF 3,09).

35. *Nerukh, D.* Non-Markov state model of peptide dynamics / D. Nerukh // Journal of Molecular Liquids. — 2012. — Т. 176. — С. 65—70. — DOI: 10.1016/j.molliq.2012.06.011 (IF 2,14).

36. *Karvounis, G.* Water network dynamics at the critical moment of a peptide's beta-turn formation: A molecular dynamics study / G. Karvounis, D. Nerukh, R. C. Glen // The Journal of Chemical Physics. — 2004. — Т. 121, № 10. — С. 4925—4935. — DOI: 10.1063/1.1780152 (IF 3,04).

37. *Nerukh, D.* Dynamic Complexity of Chaotic Transitions in High-Dimensional Classical Dynamics: Leu-Enkephalin Folding / D. Nerukh, G. Karvounis, R. C. Glen // Computational Life Sciences II. Т. 4216 / под ред. M. R. Berthold, R. C. Glen, I. Fischer. — Springer Berlin Heidelberg, 2006. — С. 129—140. — (Lecture Notes in Computer Science). — DOI: 10.1007/11875741_13.

38. *Jensen, C. H.* Sensitivity of peptide conformational dynamics on clustering of a classical molecular dynamics trajectory / C. H. Jensen, D. Nerukh, R. C. Glen // The Journal of Chemical Physics. — 2008. — Т. 128, № 11. — С. 115107. — DOI: 10.1063/1.2838980 (IF 3,15).

39. *Jensen, C. H.* Controlling protein molecular dynamics: How to accelerate folding while preserving the native state / C. H. Jensen, D. Nerukh, R. C. Glen // The Journal of Chemical Physics. — 2008. — Т. 129, № 22. — С. 225102. — DOI: 10.1063/1.3025888 (IF 3,15).

40. *Bardik, V.* Light scattering study of human serum albumin in pre-denaturation: Relation to dynamic transition in water at 42C / V. Bardik, V. Gotsulskii, E. Pavlov [и др.] // Journal of Molecular Liquids. — 2012. — Т. 176. — С. 60—64. — DOI: 10.1016/j.molliq.2012.09.012 (IF 2,14).

41. *Atamas, N.* The effect of water dynamics on conformation changes of albumin in pre-denaturation state: photon correlation spectroscopy and simulation / N. Atamas, V. Bardik, A. Bannikova [и др.] // Journal of Molecular Liquids. — 2017. — Т. 235. — С. 17—23. — DOI: 0.1016/j.molliq.2017.01.017 (IF 4,51).

42. *Atamas, N.* Water dynamics and stability of major blood proteins at pre-denaturation stage / N. Atamas, V. Bardik, O. Grishina [и др.] // Atti della Accademia Peloritana dei Pericolanti – Classe di Scienze Fisiche, Matematiche e Naturali. — 2019. — Т. 97, S2. — A16. — DOI: 10.1478/AAPP.97S2A16.

43. *Nerukh, D.* 21-residue peptide's dynamics at and between elementary structural transitions / D. Nerukh, M. Taiji // Journal of Molecular Liquids. — 2009. — Т. 147, № 1/2. — С. 124—127. — DOI: 10.1016/j.molliq.2008.10.015 (IF 3,52).

44. *Nerukh, D.* Dynamical frustration of protein's environment at the nanoseconds time scale / D. Nerukh // Journal of Molecular Liquids. — 2009. — Т. 145, № 3. — С. 139—144. — DOI: 0.1016/j.molliq.2008.06.012 (IF 1,44).

45. *Ruzhytska, S.* Identification of metastable states in peptide's dynamics / S. Ruzhytska, M. N. Jacobi, C. H. Jensen [и др.] // The Journal of Chemical Physics. — 2010. — Т. 133, № 16. — С. 164102. — DOI: 10.1063/1.3493333 (IF 3,09).

46. *Fedorov, M. V.* Self-assembly of trehalose molecules on a lysozyme surface: the broken glass hypothesis / M. V. Fedorov, J. M. Goodman, D. Nerukh [и др.] // Physical Chemistry Chemical Physics. — 2011. — Т. 13, вып. 6. — С. 2294—2299. — DOI: 10.1039/C0CP01705A (IF 3,57).

47. *Qaiser, H.* Structural insight into TNF-α inhibitors through combining pharmacophore-based virtual screening and molecular dynamic simulation / H. Qaiser, M. Saeed, D. Nerukh [и др.] // Journal of Biomolecular Structure and Dynamics. — 2020. — Т. 39, № 16. — С. 5920—5939. — DOI: 10.1080/07391102.2020.1796794 (IF 3,31).

48. *Tariq, S. S.* Impact of mutations in SARS-CoV-2 recombinant sub-variant XBB.1.16 on the binding affinity with Human ACE2 receptor / S. S. Tariq, K. Zia, M. Nur-e-Alam [и др.] // Journal of Molecular Graphics and Modelling. — 2024. — С. 108813. — DOI: 10.1016/j.jmgm.2024.108813 (IF 3,15).

49. *Abdjan, M. I.* Exploring the mechanism of action of spirooxindoles as a class of CDK2 inhibitors: a structure-based computational approach / M. I. Abdjan, M. Shafiq, D. Nerukh [и др.] // Phys. Chem. Chem. Phys. — 2024. — advance article. — DOI: 10.1039/D4CP00844H (IF 2,87).

50. *Nerukh, D.* Ligand Diffusion on Protein Surface Observed in Molecular Dynamics Simulation / D. Nerukh, N. Okimoto, A. Suenaga [и др.] // The Journal of Physical Chemistry Letters. — 2012. — Т. 3, № 23. — С. 3476—3479. — DOI: 10.1021/jz301635h (IF 6,48).

51. *Maximova, E.* Protein-Ligand Dissociation Rate Constant from All-Atom Simulation / E. Maximova, E. B. Postnikov, A. I. Lavrova [и др.] // The Journal of Physical Chemistry Letters. — 2021. — Т. 12, № 43. — С. 10631—10636. — DOI: 10.1021/acs.jpclett.1c02952 (IF 6,48).

52. *Bakumenko, M.* The Multiscale Hybrid Method with a Localized Constraint. I. A Modified Control Volume Function for the Hybridized Mass and Momentum Equations / M. Bakumenko, V. Bardik, D. Nerukh // Ukrainian Journal of Physics. — 2023. — Т. 68, № 8. — С. 517. — DOI: 10.15407/ujpe68.8.517 (IF 0,59).

53. *Bakumenko, M.* The Multiscale Hybrid Method with a Localized Constraint. II. Hybrid Equations of Motion Based on Variational Principles / M. Bakumenko, V. Bardik, V. Farafonov [и др.] // Ukrainian Journal of Physics. — 2024. — Т. 69, № 4. — С. 269. — DOI: 10.15407/ujpe69.4.269 (IF 0,8).

54. *Farafonov, V. S.* Complete Virion Simulated: All-Atom Model of an MS2 Bacteriophage with Native Genome / V. S. Farafonov, M. Stich, D. A. Nerukh // Journal of Chemical Theory and Computation. — 2023. — Т. 19, № 21. — С. 7924—7933. — DOI: 10.1021/acs.jctc.3c00846 (IF 5,77).

55. *Farafonov, V. S.* Reconstruction and validation of entire virus model with complete genome from mixed resolution cryo-EM density / V. S. Farafonov, M. Stich, D. Nerukh // Faraday Discuss. — 2022. — Т. 240, вып. 0. — С. 152—167. — DOI: 10.1039/D2FD00053A (IF 3,46).

56. *Karabasov, S. A.* The nut-and-bolt motion of a bacteriophage sliding along a bacterial flagellum: a complete hydrodynamics model / S. A. Karabasov, M. A. Zaitsev, D. A. Nerukh // Scientific Reports. — 2023. — Т. 13, № 1. — С. 9077. — DOI: 10.1038/s41598-023-36186-1 (IF 4,6).

57. *Tarasova, E.* Constant pH molecular dynamics of porcine circovirus 2 capsid protein reveals a mechanism for capsid assembly / E. Tarasova, N. Okimoto, S. Feng [и др.] // Physical Chemistry Chemical Physics. — 2021. — Т. 23, № 43. — С. 24617—24626. — DOI: 10.1039/d1cp02874j (IF 3,45).

58. *Nerukh, D.* Molecular phase space transport in water: Non-stationary random walk model / D. Nerukh, V. Ryabov, M. Taiji // Physica A: Statistical Mechanics and its Applications. — 2009. — Т. 388, № 22. — С. 4719—4726. — DOI: 10.1016/j.physa.2009.07.041 (IF 1,55).

59. *Farafonov, V. S.* MS2 bacteriophage capsid studied using all-atom molecular dynamics / V. S. Farafonov, D. Nerukh // Interface Focus. — 2019. — Т. 9, № 3. — С. 20180081. — DOI: 10.1098/rsfs.2018.0081 (IF 2,58).

Список рисунков

1.1 Временные и пространственные масштабы биомолекулярных процессов и пределы возможностей MD моделирования. 16
1.2 Схематическое представление гибридной модели; форма функции s (внизу) определяет чисто MD частицы (справа), которые постепенно трансформируются в пассивные метки, следующие за гидродинамическим потоком (слева). 19
1.3 Результаты моделирования с флуктуирующей гидродинамикой, сравнение с теорией; CS2 соответствует мелкомасштабной части домена, а CS1 соответствует крупным масштабам. 21
1.4 Результаты моделирования тепловых флуктуаций в воде при равновесии для ряда контрольных объемов (ячейки 1 и 256 соответствуют наибольшим контрольным объемам, а ячейка 128 соответствует наименьшей): мгновенная плотность (а) и x составляющая скорости (б) флуктуаций. 22
1.5 Результаты моделирования тепловых флуктуаций в воде при равновесии для диапазона контрольных объемов: стандартное отклонение флуктуаций плотности в зависимости от размера системы. 23
1.6 В плоскости xy область 2D моделирования показана черными точками, представляющими частицы MD; фаза FH отображается в виде цветного поля; параметр связи s в этом примере имеет круговую симметрию, где в центре $s = 0$ а по краям $s = 1$, и его проекция на плоскость xy показана цветом. 23
1.7 Переменный параметр s и MD-сфера внутри вычислительной ячейки. Красная зона — это чистая область MD ($s = 0$), а синяя — область флуктуирующей гидродинамики ($s = 1$) для сферически симметричного случая s-функции; молекулы воды схематически показаны для примера (не в масштабе). 24
1.8 Модель воды «Mercedes Benz» [57]. 34
1.9 Флуктуации производной первого порядка по площади для разных размеров системы (указано количество молекул), $T^* = 0.21$. 44
1.10 Сходимость производной первого порядка потенциала MB по площади и изотермической сжимаемости, $T^* = 0.21$. 44
1.11 Автокорреляционные функции поступательной и вращательной скоростей модели MB. 45
1.12 Мгновенный снимок структуры воды MB, $T^* = 0.15$. 45
1.13 Функция радиального распределения для воды MB. 45
1.14 Стандартное отклонение скорости вращения с временным шагом Δt^*. 46
1.15 Уравнение состояния для модели воды MB. 46
1.16 а) Поступательная и б) вращательная автокорреляционные функции в $T^* = 0.195$ для моделей MB и SPC. 47
1.17 Функция радиального распределения при $T^* = 0.16$. Центральная молекула показана зеленым цветом. «Междоузельная» вода выделена пурпурным цветом. 47

1.18 Распределение функции плотности вероятности для массы (слева) и x-компоненты импульса (справа), теоретические распределения в каждом случае показаны сплошными черными линиями. 50
1.19 Исходное строение капсида: слева – общий вид с границами области моделирования и разделение системы на области: 1 — чистая MD, 2 — гибридная область, 3 — чистая гидродинамика; пунктирная линия показывает подразделение гибридной области для термостатирования; справа – срез, показывающий «хвосты» (система Б); отдельные белки и хвосты показаны разным цветом. 52
1.20 Зависимость интегрального заряда капсидных слоев от радиуса наибольшего слоя; сплошная кривая соответствует системе А, штриховая кривая соответствует системе Б; вертикальные линии обозначают примерные границы стенки капсида. 53
1.21 Диаланин в гибридной MD/FH воде; цвет частиц постепенно меняется с синего (область FH) на белый (область смешанной MD-FH) и на красный (чистая область MD). 59
1.22 Вычисленная функция $MSD(t)$ для задачи гибридного моделирования LL-FH/MD диффузии пептида в воде (сплошная линия) и в чистом MD (штриховая линия). 60
1.23 Капсид PCV2 в гибридной воде LL-FH/MD; капсид находится внутри области чистой MD (белая), которая постепенно изменяется через промежуточную LL-FH/MD область (серый) в область FH (синий); мелкие клетки имеют размер 1 нм что примерно соответствует размеру ячейки FH. 61
1.24 Масштабирование с количеством вычислительных узлов для моделирования MD с помощью K-компьютера в RIKEN. 63
1.25 Разработка ускорителей семейства GRAPE (Gordon Bell Призы, выигранные в 1995, 1996, 1999, 2000 (двойной), 2001, 2003 и 2006). 64
1.26 Системная плата MDGRAPE-4 и система на кристалле. На плате расположено 8 SoC и 48 оптических передатчиков/приемников для работы в сети. SoC имеет 8 модулей расчета нековалетных сил (PP), 8 модулей процессоров общего назначения (GP), каждый из которых имеет 8 ядер, память (GM) и сетевые интерфейсы (NIF). 64
1.27 a) Количество времени, которое занимает один цикл GPU как функция количества ячеек (N); b) отношение времени, необходимого для выполнения одного цикл на процессоре по сравнению со временем на графическом процессоре как функция количества ячеек в одном измерении кубической области. 65
1.28 Стандартные отклонения средних скоростей (компонента x) в чистом моделировании MB, моделировании LL-FH и связанных системах. 66
1.29 Разность скоростей в смеси и в отдельных фазах a) $\tilde{u}_x - \sum_p u_{xp}$и b) $\tilde{u}_x - u_x$ для $s = 0.1$ и $s = 0.8$. 67
1.30 Профили компонент скорости x фазы MD $\sum_p u_{xp}$ и фазы FH u_x; a) $s = 0.1$, b) $s = 0.8$. 67
1.31 Функция радиального распределения для различных значений s. 68
1.32 Функция автокорреляции поступательной скорости для различных значений s, рассчитанных из скорости частиц. 68
1.33 Функция радиального распределения MB частицы в центре гибридной области параметра s, где $s = 0$. 69

1.34 Автокорреляционная функция поступательной скорости частиц MB в центре гибридной области s, где $s = 0$. 70
1.35 Стандартное отклонение флуктуаций скорости для аргона для четырех разных случаев; чистое решение MD ($s = 0$) без BLOB фильтра, MD часть решения для $s = 0.1$ и $s = 0.8$ и решение LL-FH. 71
1.36 Стандартные отклонения флуктуаций плотности аргона для пяти разных случаев; чистое решение MD ($s = 0$) без и с BLOB фильтром, часть решения MD для $s = 0.1$ и $s = 0.8$ и решение LL-FH. 72
1.37 Функции радиального распределения для различных значений s для аргона. 73
1.38 Автокорреляционные функции скорости для различных постоянных s значений для аргона. 73
1.39 Автокорреляционные функции скорости в случае переменной s для аргона. 74
1.40 Изменение средней по времени плотности и импульса в гибридной зоне $0 < s < 1$ показана вертикальными пунктирными линиями (слева $s = 0$ и справа $s = 1$) для аргона. 75
1.41 Изменение стандартного отклонения плотности и импульса в гибридной зоне $0 < s < 1$, показано вертикальными пунктирными линиями (слева, $s = 0$ и справа, $s = 1$) для аргона. 76
1.42 Автокорреляции (а) плотности и (б) x-скорости компонент для аргона для положения x_0 в центре гибридной зоны MD/LL-FH. 76
1.43 Сигнал флуктуирующей плотности, полученный в MD части решения по сравнению с аналитическим решением (слева), включая исходный MD сигнала без фазового и пространственного усреднения, (справа) увеличенный вид с включением эталонного фазового и усредненное по пространству решение без учета акустической волны. . . . 77
1.44 Стандартные отклонения пульсаций скорости для SPC/E воды для четырех различных случаев; чистый MD ($s = 0$), два варианта MD/LL-FH с $s = 0.1$ и $s = 0.8$ и чистый LL-FH ($s = 1$). 78
1.45 Стандартные отклонения флуктуаций плотности для SPC/E воды для пяти разных случаев: чистый MD ($s = 0$) без BLOB фильтра, часть решения MD для $s = 0{,}1$ и $s = 0{,}8$ и решение LL-FH. 79
1.46 Автокорреляционные функции скорости для различных постоянных значений s для SPC/E воды. 80
1.47 Разрез, показывающий SPC/E воду в случае переменной s; белые и красные молекулы - молекул в MD-области, синие сферы — капли воды в области с преобладанием гидродинамики. 81
1.48 Функции радиального распределения кислород-кислород для разных значений s для SPC/E воды. 82
1.49 Функции радиального распределения кислород-водород для разных значений s для SPC/E воды. 83
1.50 Автокорреляционные функции скорости в случае переменной s для SPC/E воды. . . . 84

1.51 Переменный параметр s и MD-сферас небольшой пептидной молекулой, цвиттерионной формой диаланина, внутри в центре сферы. Красная сфера — область чистой MD ($s = 0$), синий — область флуктуационной гидродинамики ($s =$1). 84
1.52 График MSD(t)) для воды SPC/E и белка для чистого MD. 85
1.53 Графики MSD(t) для воды (гибридные MD), рассчитанные с разными радиусами обрезания. График чистог MD приведен для сравнения. 85
1.54 Графики MSD(t) для белка (гибридная MD), рассчитанные с разными радиусами обрезания; график для чистого MD приведен для сравнения. 86
1.55 Начальная конформация капсида PCV-2 по экспериментальным рентгеновским данным. 88
1.56 Капсид PCV2 нестабилен в чистом MD. 88
1.57 Капсид PCV-2 стабилен в гибридном моделировании LL-FH/MD: исходная форма вируса после уравновешивания, соответствующая экспериментальная структура; B — смоделированная структура в гибридном LL-FH/MD; C это сравнение A и B (синий - моделирование, красный - исходный вирус после уравновешивания). 89
1.58 Скользящие средние значения среднеквадратичного отклонения (RMSD) атомов основной цепи от их исходной структуры для расчета чистой MD (вверху) и гибридной модели LL-FH (внизу). 90

2.1 Процесс преобразования непрерывного сигнала скорости атомов $\mathbf{v}$ в символьную последовательность; справа показана символизация с 2, 3, 4 и 5 символами; 3-символьный алфавит использовался во всех последующих расчетах. 94
2.2 Сходимость условной вероятности $p(0) = p(0|\{010200100\})$ с длиной временного ряда для произвольно выбранного символьного слова ‘010200100’; прямая показывает линейную аппроксимацию результата расчета значения $p(0)$ для различной длины символьной последовательности; коэффициент $\overline{p}(0)$ подгонки определяет наклон линии, тогда как среднеквадратичное отклонение точек от линейной подгонки обозначается как $\sigma_{p(0)}$. 95
2.3 Стандартное отображение; символизация с помощью трехсимвольного алфавита (слева); два периодических острова встроены в хаотичное море; один из них в увеличенном виде показан справа. 97
2.4 Условные вероятности ‘0’, $p(0|\{\})$, и ‘1’, $p(1|\{\})$ (слева) и их тренд $\overline{p}(0)$ (в середине) и дисперсия $\sigma_{p(0)}$ (справа) как функция вхождения слова $p(s_i)$, вычисленного для символьной строки длиной $3 \cdot 10^7$ для чисто случайного процесса (троичное разложение числа π). 97
2.5 То же, что и предыдущая картинка, но рассчитано для суррогатного временного ряда. 98
2.6 То же, что и на предыдущих двух рисунках, но для сигнала от воды. 98
2.7 То же, что и предыдущие три рисунка, но для стандартного отображения. 99
2.8 То же, что и предыдущий рисунок, но для другого разбиения стандартного отображения 100

2.9 Иллюстрация вырожденности макронаблюдаемой проекции траектории полномерного фазового пространства. Та же последовательность наблюдаемой (скорости) $\{\mathbf{v}_t\mathbf{v}_{t+1}\mathbf{v}_{t+2}\mathbf{v}_{t+3}\}$ генерируется двумя различными частями траектории фазового пространства $\{\mathbf{q}_t\mathbf{q}_{t+1}\mathbf{q}_{t+2}\mathbf{q}_{t+3}\}$ и $\{\mathbf{q}'_t\mathbf{q}'_{t+1}\mathbf{q}'_{t+2}\mathbf{q}'_{t+3}\}$. 103

2.10 Автокорреляционные функции $C(\tau) \equiv \frac{1}{T}\sum_t^T \mathbf{v}_t \cdot \mathbf{v}_{t+\tau}$ для исходной скорости водорода в объемной воде (черный) и сигнала, состоящего из 27 символов (красный). . 104

2.11 Схематическое изображение последовательностей, используемых для определения формулы (2.5); точки фазового пространства $\{\mathbf{q}\}$ и $\{\mathbf{q}'\}$ двух частей траектории образуют последовательности Маркова соответствующие последовательности наблюдений A и A' не являются марковскими, так как одно и то же значение A_t приводит к разным A_{t+1} и A'_{t+1} в зависимости от предыдущих значений A_{t-1} и A'_{t-1}; однако, если обе истории $\{\ldots A_{t-2}A_{t-1}A_t\}$ и $\{\ldots A'_{t-2}A'_{t-1}A_t\}$ принадлежат к одному и тому же обусловливающему состоянию ε_t, следующее обусловливающее состояние ε_{t+1} определяется без знания ε_{t-1}, таким образом делая $\{\varepsilon\}$ последовательностью Маркова. 106

2.12 Гистограмма случайной величины φ иллюстрирующая равномерное симметричное распределение точек в плоскости поперечного сечения $v_z = 0$ траектории скорости водорода. 109

2.13 Логарифм числа обусловливающих состояний $\log_2 n_{st}$ в ε-машине как функция длины истории l, погрешности σ и длительности временного ряда; слева: суррогатный временной ряд Фурье [132], справа: молекулярный сигнал; вверху: длительность временного ряда 60 нс, посередине: 450 нс, внизу: 1μs. 111

2.14 Схема функции распределения вероятностей для времени возврата Пуанкаре τ. Экспоненциальный спад при малых τ изменяется на, по-видимому, более медленный закон спада на больших временах. 112

2.15 Распределение времен возврата Пуанкаре, усредненное по всем обусловливающим состояниям; a) Стандартное отображение; b) временной ряд воды, соответствующий скорости атома водорода. 114

2.16 Кластеризация обусловливающих состояний для временного ряда скорости атома водорода (a) и стандартного отображения (b) на «периодические» (ромбы) и «хаотические» (треугольники) классы; параметр D построен в виде графика в зависимости от частоты возникновения обуславливающих состояний. 115

2.17 Анализ Фурье временных рядов воды; спектры мощности (a, d) и гистограммы времен возврата (b, c, e, f) для типичных обуславливающих состояний, относящихся к разным типам: «периодическому» состоянию (a-c) и «хаотическому» состоянию (d-f); гистограммы на рисунках (c, f) представляют собой увеличенные и сглаженные фрагменты гистограмм, показанных на рисунках (b, e); спектры в (a, d) являются функциями обратной частоты. 116

2.18 Статистическая сложность в зависимости от (логарифма) длины интервала анализа для временного ряда скорости водорода (верхняя кривая) и четырех суррогатных временных рядов (нижние кривые): три независимые реализации алгоритма фазовой перестановки (красный, зеленый и синий) и один временной ряд белого шума, пропущенного через линейный фильтр нижних частот (черный) (из [19]); обратите внимание, что значение статистической сложности для данных, полученных с помощью простого генератора случайных чисел, близко к нулю и не зависит от длины временного ряда при достаточно большом значении последней (не показано). . 118
2.19 Значения C_μ и количество обуславливающих состояний для различных длин истории (снизу вверх $l = 2 \dots 9$) для стандартной траектории отображения как функции длины данных. 121
2.20 То же, что на рисунке 2.19 но для случайного суррогата. 122
2.21 Зависимость статистической сложности от параметра K (количество точек данных 10^8). 122
2.22 Условные распределения следующего символа для всех историй при длине истории $l = 8$. Каждая точка представляет собой историю. Общее количество точек (историй) составляет 3^8. 123
2.23 Условное распределение вероятностей для истории с наибольшим отклонением от марковского свойства (черный треугольник); добавление символа в историю кардинально меняет положение точки на диаграмме; три круга соответствуют добавлению ‘0’, ‘1’ или ‘2’ к истории $l = 8$ (крестики); красный треугольник представляет собой случайно выбранную историю с явным марковским свойством, т.е. распределение вероятностей не зависит от добавленного символа. 124
2.24 Участки хаотической траектории, соответствующие истории с большим отклонением от свойства цепи Маркова; отображаются только точки, соответствующие трем центральным символам в истории; по-видимому, история включает в себя один из двух периодических островов, показанных на рисунке 2.3. . . 125
2.25 Участки хаотической траектории, соответствующие истории без отклонений от свойства цепи Маркова; отображаются только точки, соответствующие трем центральным символам в истории; по-видимому, история не включает ни один из двух периодических островов, показанных на рисунке 2.3. 126

3.1 Разрез полноатомной модели капсида PCV2 без генома, погруженного в ячейку, заполненную атомистической водой [7]. 128
3.2 Модель капсида в коробке в форме ромбического додекаэдра; несколько белков оболочки окрашены, выделяя пентамер, окружающий пору (зеленые, оранжевые, синие и красные молекулы), и блок тримера (зеленые, оранжевые, синие, красные, фиолетовые и желтые молекулы); ионы и молекулы воды не показаны. 135
3.3 Слева: RMSD всего капсида (черный) и его остова (красный) от начальной структуры; справа: радиус вращения капсида. 136
3.4 RMSD отдельных остатков, усредненных по 89 димерам белков оболочки. 137

3.5 Первоначальное расположение остатков 13–15, имеющих наибольшее смещение от структуры крио-ЭМ (оранжевый) и 75–78, образующих петлю FG (зеленый). 137
3.6 Усредненные по капсиду вторичные структуры мономеров белков оболочки типов A, B и C в различные моменты времени моделирования; для каждого вида мономера первая подчеркнутая строка соответствует конфигурации крио-ЭМ; показан контур FG. 138
3.7 Функции радиального распределения ионов Na^+ (красный) и Cl^- (зеленый) относительно ЦМ капсида, рассчитанные в ходе продуктивного прогона. 139
3.8 Слева: число молекул воды, вошедших (красный) или покинувших (синий) внутреннюю часть капсида, по сравнению со структурой через 40 нс, суммой вошедших и покинувших (черный) и ее линейной аппроксимацией (серый); справа: число ионов Na^+ (оранжевый) или Cl^- (зеленый), пересекших стенку капсида в обоих направлениях, по сравнению с исходной структурой, и ее линейная аппроксимация. . . 140
3.9 Зависимости общего заряда (вверху) и плотности (внизу) атомов капсида от расстояния от центра масс капсида; черные кривые соответствуют капсиду без хвостов, красные кривые соответствуют капсиду с хвостами; вертикальная линия соответствует середине стенки капсида. 142
3.10 Изменение во времени среднеквадратичного отклонения, RMSD, (верхние графики) и радиуса инерции, R_g, (нижние графики) атомов основной цепи относительно их первоначальной структуры как эталонной структуры в силовом поле AMBER; расчеты RMSD и R_g не включали хвостовую последовательность для всех структур и состояли из одинакового числа аминокислот; красный цвет обозначает капсид с хвостами, черный — капсид без хвостов; a, c – капсиды с нейтрализующим числом ионов, b, d – капсид с малым числом ионов. 146
3.11 Атомистические структуры капсида PCV2 без хвостов после моделирования: (a,c) атомистические структуры капсида PCV2 с нейтрализующим числом ионов, (e) радиус инерции для разных осей (зеленый — ось X, красный — ось Y, синий — ось Z); (b,d) — структуры капсида с меньшим числом ионов и (f) радиус инерции для разных осей. 147
3.12 Среднеквадратичное смещение атомов основной цепи капсида от начальной конфигурации во время гибридного расчета MD/гидродинамики. 148
3.13 Распределение ионов после первой наносекунды моделирования MD для малого числа ионов (вверху) и для нейтрализующего числа ионов (внизу); красная плотность (a,c) представляет собой плотность ионов натрия, а синяя плотность (b,d) — плотность ионов хлора, усредненную за одну наносекунду моделирования. 149
3.14 Поток воды и ионов в системе A (a,c,e) и системе B (b,d,f) в течение 10 нс моделирования; a,b) количество ионов Na^+ (красный) и Cl^- (синий) внутри капсида как функция времени, c,d) количество молекул воды внутри капсида как функция времени, e,f) общее количество молекул воды, которые пересекли стенку капсида в обоих направлениях по сравнению с исходной структурой (черный) и ее линейной аппроксимацией (серый), количество вошедших (красный) и вышедших (синий) молекул воды. 151

3.15 а) Количество ионов Na^+ (красный) и Cl^- (синий) внутри капсида как функция времени; б) общее количество молекул воды, которые пересекли стенку капсида в обоих направлениях по сравнению с исходной структурой (черный) и ее линейной аппроксимацией (серый), количество вошедших (красный) и вышедших (синий) молекул воды во время гибридного расчета MD/гидродинамики. 152
3.16 Стабильные (слева) и деформированные (справа) капсиды; деформация прогрессирует и приводит к разрушению капсида. 154
3.17 Схематическое изображение распределения ионов Cl^- (зеленый) и Na^+ (оранжевый) в системе и RDF, количественно определяющая их концентрацию в зависимости от расстояния от центра масс капсида. 155
3.18 Функции радиального распределения ионов Na^+ (вверху) и Cl^- (внизу) относительно центра масс капсида; черные линии соответствуют стабильному капсиду, красные линии соответствуют деформированному капсиду. 156
3.19 Количество остатков, имеющих ион Cl^- ближе, чем 5A; показаны только аминокислоты, имеющие более 20 остатков, контактирующих с Cl^- в стабильном капсиде; красный — естественная система, зеленый — искусственная система. 157
3.20 Половина капсида, показывающая расположение богатых хлоридом аминокислот 172(Asn), 174(Lys), 175(Arg) (красные) и 94(Arg), 95(Lys), 173(Asn), 176 (Asn) (розовые) и теряющих хлорид аминокислот 43(Arg), 97(Lys), 142(Arg), 213(Tyr) (зеленые) и 118(Val), 156(Phe) (синие); остатки на расстоянии более 75A от центра масс капсида показаны желтым цветом; один тример обозначен пунктирной линией. . 158
3.21 Половина деформирующегося капсида после 10 нс моделирования MD; аминокислоты 43(Arg), 97(Lys), 142(Arg) и 213(Tyr) выделены зеленым, красным, розовым и желтым цветами соответственно. 159
3.22 Карты электростатического потенциала внутренней части капсида; верхний ряд: поверхность капсида, нижний ряд: поверхность приблизительно на расстоянии 5 A от капсида; шкала варьируется в пределах 13,7 – 23 B (a, d); 17 – 23 B (b, e); 20 – 23 B (c, f), где синий цвет соответствует нижнему пределу, а красный – верхнему пределу шкалы. 160
3.23 Сборка капсида PCV2 in vitro; (a) схематические изображения полноразмерного CP, $CP^{\Delta ARM}$ и CP$*$ описывающий различия между белками. Аминокислотные последовательности N-концов приведены ниже; (б) микрофотографии, полученные с помощью трансмиссионного электронного микроскопа, отрицательно окрашенного in vitro собранного капсида PCV2 вверху, влияние pH (белый текст) на сборку CP$*$; внизу, влияние pH (белый текст) на сборку $CP^{\Delta ARM}$. 167
3.24 Конформационный ландшафт, полученный с помощью капсидного белка $CP*^{42-231}$; поверхность свободной энергии как функция параметров RMSD ($CP*^{42-231}$ Cα атомов) как функция Rg ($CP*^{42-231}$Cα атомов) из моделирования при (a) pH 3, (b) pH 5, (c) pH 7; RMSD рассчитывали путем сравнения конформаций, полученных из 1μc моделирования к эталонному CP в собранном капсиде (запись PDB 3R0R); вертикальные черные линии обозначают значение Rg записи PDB 3R0R. 172

3.25 Конформационная изменчивость CP при pH 3, 5 и 7; (a) слева: капсид PCV2 с CP, представленным в виде зеленых точек; оси симметрии икосаэдра показаны как пятиугольник (5-го порядка), треугольник (3-го порядка) и эллипс (2-го рядка); справа: одна из субъединиц показана как Cα представление трубки с важными петлями (N-конец, CD и GH в виде толстых трубок); (б) PCV2 CP Cα представление трубки с маркировкой важных петель (толстых трубок); (c) расстояния между эквивалентными Cα атомами репрезентативных конформаций при pH 3, 5, 7 после выравнивания со структурой капсида (запись PDB 3R0R); панели a и b были сделаны с использованием структуры капсида; в контексте собранного капсида N-конец находится внутри капсида около оси 2-го порядка: CD-петля расположена на поверхности капсида вблизи оси 2-го порядка, GH-петля скрыта около оси 3-го порядка, а C-конец расположен на поверхности капсида вблизи оси 2-го порядка. . . . 173

3.26 Репрезентативные конформации из первого кластера при каждом значении pH; области со значительным отклонением от структуры капсида выделены в соответствии с рисунком 3.25; a) CP, извлеченный из собранных капсидов, b) представитель наиболее заселенного кластера при pH 3, c) представитель наиболее заселенного кластера при pH 5, d) представитель наиболее заселенного кластера из первого пика при pH 7, e) представитель наиболее заселенного кластера из второго пика при pH 7, f-j) крупный план EF- (розовый) и GH-петель (синий), показывающий их кислотные диады f) CP, извлеченный из собранных капсидов, g) представитель наиболее заселенного кластера при pH 3, h) представитель наиболее заселенного кластера при pH 5, i) представитель наиболее заселенного кластера из первого пика при pH 7, j) представитель наиболее заселенного кластера из второго пика при pH 7; значения RMSD для репрезентативных структур: (b)–4.6Å, (c)–3.9Å, (d)–4.3Å и (e)–5.1Å. 174

3.27 Распределение расстояний между центрами масс для остатков 126 и 168 (a, c, e, g) и остатков 126 и 172 (b, d, f, h); при (a, b) pH 3; (c, d) pH5; (e, f) pH 7 (минимум 1); и (g, h) pH7 (минимум 2); синяя вертикальная линия на каждом графике представляет расстояние между этими остатками в структуре капсида (10.7 Å между 168 и 126; 5.1 Å между 172 и 126); распределение было рассчитано для всех структур, извлеченных из минимумов свободной энергии поверхности при каждом pH. 175

4.1 Справа: ε-машина для временного шага 5 пс, длина последовательностей $l = 2$, состояние «0» соответствует исходному состоянию «A» (в основном состоит из последовательностей AA), аналогично, состояние «3» соответствует «B», состояние «2» описывает переходы между A и B (состоит из последовательностей AB и BA); слева: распределение конформаций, соответствующих каждому обуславливающему состоянию; для визуализации выбраны моменты времени на последнем символе историй. 182

4.2 Слева: ε-машина для временного шага 0.3 пс, длина последовательностей $l = 4$, состояние «0» соответствует исходному состоянию «A» (в основном состоит из последовательностей $AAAA$), аналогично, состояние «9» соответствует «B», состояние «9» может быть достигнуто только из состояния «0» через состояния «2», «7», «4», «8», которые описывают механизм (пути) перехода $A \to B$; справа: три типичных случая во время перехода: прямой переход (вверху, вероятность 0.56), переход с несколькими повторными пересечениями (в середине, вероятность 0.03), неудачная попытка перехода (внизу, вероятность 0.27), вероятности этих случаев даны в предположении, что вероятность перехода из состояния «0» или в состояние «9» любым возможным путем равна 1. 184

4.3 Поток двугранных углов, описывающий две различные возможности перехода из состояния «0»: слева, переход $A \to B$, справа, неудавшийся переход и возврат в состояние A; числа показывают время до завершения перехода (то есть, когда система достигает либо состояния «0», либо состояния «9»); для визуализации выбраны моменты времени в последнем символе историй. 186

4.4 Вверху: цвиттер-ион L-аланил-L-аланина (силовое поле GROMOS, см. приложение); слева: Нормализованные вероятности конформаций (график Рамачандрана), образованные 2μs траекториями; справа: те же вероятности, подчеркивающие наличие двух второстепенных конформаций и разбиение для символизации. 188

4.5 Вверху: цвиттер-ион L-аланил-L-аланина (силовое поле OPLS); слева: Нормализованные вероятности конформаций (график Рамачандрана), образованных 2μs траектория; справа: те же вероятности, подчеркивающие наличие двух второстепенных конформаций и разбиение для символизации. 189

4.6 Вверху: сбор временных точек для статистики «время до перехода»; точки на осях «время» — это моменты перехода; в середине слева: снимок молекулярной системы, показывающий ≈ 2-3 слоя молекул воды, связанных водородными связями друг с другом и пептидом; в середине, справа: распределение атомов кислорода (красный, сплошной) и водорода (синий, сетчатый) усреднено по всей симуляции (показаны изоповерхности при значении вероятности 0.047 (значение в предположении полностью равномерного распределения равно 0.031) для модели GROMOS); внизу: одномерное и двумерное распределение вероятностей, иллюстрирующее сужение пика для высокой вероятности атомов; красное, более узкое распределение описывает ситуацию, когда атомы имеют более определенные местоположения, а не более равномерно распределены, как характерно для синего распределения. 190

4.7 Слева: зависимость объема областей высокой вероятности от «времени до перехода», нормализованная к средним значениям, $\Delta v = \frac{v-\langle v\rangle}{\langle v\rangle}$, где v — абсолютный объем (см. текст), а $\langle v\rangle$ — средний объем за все время моделирования, показаны площади изоповерхностей на уровнях вероятности (сверху вниз) 0.013, 0.041, 0.046, 0.051 и 0.059; справа: то же самое для кислорода (красный) и водорода (синий) для перехода из состояния A в состояние B, модель GROMOS (вверху), переход $A \to C$, модель GROMOS (в середине) и переход $A \to C$, модель OPLS (внизу), показаны площади изоповерхностей на уровне вероятности 0.055. 192

4.8 Сечение xy (значение координаты z выбрано таким образом, что поле флуктуации вероятности $f'(x,t)$ (справа), и его часть коррелируемой с конформацией пептида $f'_c(x,t)$ (слева) для кислорода; время до перехода составляет, сверху вниз, $200ps$, $0.8ps$ и $0ps$; шкала ограничена интервалом $-0.009 \ldots 0{,}009$ для наглядности, максимумы пиков достигают значений 0.045 и -0.041 (приведены результаты для модели GROMOS). 193

4.9 Сечение xy (значение координаты z выбрано таким образом, чтобы секущая плоскость проходила через центр масс пептида; другие значения координаты z показывают качественно те же результаты) поля флуктуации вероятности $f'(x,t)$ (справа), и его часть коррелирующая с конформацией пептида $f'_c(x,t)$ (слева) для водорода; время до перехода составляет, сверху вниз, $200ps$, $0.4ps$ и $0ps$; шкала ограничена интервалом $-0.015 \ldots 0{,}015$ для наглядности, максимумы пиков достигают значений 0.04 и -0.055 (показаны результаты для модели OPLS). 221

4.10 Поверхность белка окрашена в соответствии с разницей в гибкости остова, вызванной трегалозой (цвета соответствуют графику разницы, показанному на рисунке 4.12); голубой: плотность трегалозы (в 6.4 раза выше, чем в основной массе); зеленый: плотность воды (в 1.5 раза выше, чем у объемной воды), показана молекулярная структура с трех разных точек зрения 222

4.11 Поверхность белка окрашена в соответствии с разницей в гибкости остова, вызванной трегалозой; голубой: плотность трегалозы (в 6.4 раза выше, чем в основной массе); показан коробка моделирования, подчеркивающий отсутствие сахара везде, за исключением области, прилегающей к белку. 223

4.12 RMSF атомов основной цепи белка; черный: 30 нс моделирование в растворе чистой воды; голубой: 30 нс моделирование в растворе трегалозы; многоцветная линия: разница в RMSF между растворами воды и трегалозы (голубая и черная линии), цвета используются для отображения разницы на поверхности белка, Рисунок 4.10, 4.11 и структур белка, Рисунок 4.16; структурные мотивы для каждой аминокислоты показаны в виде цветной полосы над кривыми: красная - β-лист, синий - α-спираль, желтая - петля. 224

4.13 RMSF атомов основной цепи белка; черный: 30 нс моделирование в чистом водном растворе; красный: 50 ЯМР-структур из [310]; структурные мотиве для каждой аминокислоты показаны в виде цветной полосы над кривыми: красная - β-лист, синий - α-спираль, желтая - петля. 224

4.14 Несколько случайно выбранных молекул сахара показаны для сравнения размера участков сахара высокой плотности с размером молекул сахара. 225
4.15 Среднее количество водородных связей лизоцима и трегалозы на аминокислоту; также показана разница в гибкости аминокислот (такая же, как на рисунке 4.12). 225
4.16 Слева: среднее число водородных связей на аминокислоту; справа: гибкость аминокислот (окраска соответствует рисунку 4.15). 226
4.17 Расстояние d между центрами масс лиганда TMP (a, b, c) или BrWR (d) и ближайшим остатком белка; низкие, постоянные значения d соответствуют интервалам времени, когда лиганды находятся на поверхности (и между лигандом и поверхностью нет молекул воды), а значения выше ≈ 7 показывают диффузию лигандов в объеме; несколько репрезентативных положений лиганда TMP на поверхности (и, иногда, вне нее) показаны вверху, числа на графике времени (a), от 1 до 7, соответствуют положениям лиганда на поверхности; верхний правый рисунок показывает диффузию TMP на поверхности из положения 2 в положение 7, лиганд представлен в виде одного шара, постепенно меняющего свой цвет вдоль траектории от красного (положение 2) через белый (приблизительно положение 5) к синему (положение 7). . . 227
4.18 Модель потенциала прямоугольной ямы для взаимодействия лиганда с поверхностью. 228
4.19 Структурные формулы веществ STR1-STR4, использованных в исследовании взаимодействия с альбумином. 228
4.20 Расположение лигандов в сайтах связывания ЧСА, DS1 – IA, DS2 – IIA, DS3 – IIIA, DS4 – IB, DS5 – IIB, DS6 – IIIB. 229
4.21 Подгонка вероятности того, что лиганд останется связанным с белком, с использованием моделей для двух режимов; результаты для внешней силы, равной 350 кДж/(нм · mol) (а) и 500 кДж/(нм · mol) (б); черная штриховая линия — неэкспоненциальная модель, синяя линия — экспоненциальная модель, красная штрихпунктирная линия — момент переключения между моделями; подобранные значения параметров приведены в таблице 14. 230
4.22 Последовательность времен диссоциации, определенная с помощью моделирования MD, линеаризованная с помощью преобразования координат как функции приложенных сил на моль (точки) и их линейной подгонки (сплошная линия); пунктирные кривые обозначают границы предсказания с уровнем достоверности, равным стандартному отклонению; звёздочкой отмечено экспериментальное значение. 231

Д.1 Суммарный заряд атомов капсида в сфере радиуса r с центром в центре масс капсида (красная кривая); синяя кривая — сглаженный исходный заряд; пунктирная серая кривая показывает радиальную функцию распределения атомов капсида (масштабирована по вертикали для наглядности); вертикальные линии обозначают границы, охватывающие 90% атомов капсида 298

Е.1 Количество остатков, имеющих ион Cl^- ближе, чем 5; красный - естественная система, зеленый - искусственная система 299

Ж.1 Изменение во времени среднеквадратичного отклонения, RMSD, (верхние графики) и радиуса инерции, R_g, (нижние графики) атомов основной цепи относительно их первоначальной структуры как эталонной структуры в силовом поле AMBER; расчеты RMSD и Rg не включали хвостовую последовательность для всех структур и состояли из одинакового числа аминокислот красный цвет обозначает капсид с хвостами, черный — капсид без хвостов; a,c – капсиды с нейтрализующим числом ионов, b,d – капсид с малым числом ионов 301

Ж.2 Изменение во времени среднеквадратичного отклонения, $RMSD$, (верхние графики) и радиуса инерции, R_g, (нижние графики) атомов основной цепи относительно их первоначальной структуры как эталонной структуры в силовом поле Gromos; расчеты RMSD и Rg не включали хвостовую последовательность для всех структур и состояли из одинакового числа аминокислот; красный цвет обозначает капсид с хвостами, черный — капсид без хвостов. a,c – капсиды с нейтрализующим числом ионов, b,d – капсид с малым числом ионов 302

Ж.3 Зависимости интегрального заряда атомов капсида от расстояния от центра масс капсида; синие кривые соответствуют капсиду с хвостами, красные кривые соответствуют капсиду без хвостов; сплошные кривые соответствуют конечным конфигурациям, пунктирные — начальным; вертикальная линия соответствует середине стенки капсида 302

3.1 Конформационный ландшафт, полученный с помощью CP* во время моделирования при pH 3; поверхности свободной энергии как функция параметров SASA всех остатков в зависимости от Rg всех остатков из 1000 нс моделирования; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231 303

3.2 Конформационный ландшафт, полученный с помощью CP* во время моделирования при pH 5; поверхности свободной энергии как функция параметров SASA всех остатков в зависимости от Rg всех остатков из 1000 нс моделирования; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231 304

3.3 Конформационный ландшафт, полученный с помощью белка CP* во время моделирования при pH 7; поверхности свободной энергии как функция параметров SASA всех остатков в зависимости от Rg всех остатков из 1000 нс моделирования; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231 304

3.4 Конформационный ландшафт, отобранный CP* в течение 1000 нс моделирования при pH 3; поверхности свободной энергии как функции параметров а) RMSD всех атомов Cα по сравнению с Rg всех остатков; б) RMSD 38-231 атомов Cα по сравнению с Rg 38-231 остатков из моделирования; среднеквадратическое отклонение рассчитывалось относительно эталонной рентгеновской структуры 305

3.5 Конформационный ландшафт, отобранный CP* в течение 1000 нс моделирования при pH 5; поверхности свободной энергии как функции параметров a) RMSD всех атомов Cα по сравнению с Rg всех остатков; б) RMSD 38-231 атомов Cα по сравнению с Rg 38-231 остатков; среднеквадратическое отклонение рассчитывалось относительно эталонной рентгеновской структуры 306
3.6 Конформационный ландшафт, отобранный CP* в течение 1000 нс моделирования при pH 7; поверхности свободной энергии как функции параметров a) RMSD всех атомов Cα по сравнению с Rg всех остатков; б) RMSD 38-231 атомов Cα по сравнению с Rg 38-231 остатков из моделирования; среднеквадратическое отклонение рассчитывалось относительно эталонной рентгеновской структуры 307
3.7 Диаграмма свободной энергии как функции параметра SASA всех остатков при pH 3, 5 и 7; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231 308
3.8 Диаграмма свободной энергии как функции параметра Rg всех остатков при pH 3, 5 и 7; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231 308
3.9 Диаграмма свободной энергии как функция параметра RMSD всех остатков при pH 3, 5 и 7; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231 308
3.10 Диаграмма свободной энергии как функция параметра RMSD 38-231 атомов Cα при pH 3, 5 и 7; среднеквадратичное отклонение рассчитывалось относительно эталонной рентгеновской структуры 309
3.11 Диаграмма свободной энергии как функции параметра Rg остатков 38-231 при pH 3, 5 и 7 309
3.12 Совокупно рассчитанные значения pKa для всех остатков Asp/Glu/His из симуляций при pH 3 (a), pH 5 (b), pH 7 (c) как функция времени симуляции 309
3.13 Обмен участками для каждой реплики; зависимость обменов между значениями pH от времени моделирования для отдельных реплик; коэффициент принятия составляет 21% 310
3.14 Гистограмма, показывающая долю извлеченных структур на кластер для 10 лучших кластеров; количество кластеров, начиная с 0 310
3.15 Выравнивание репрезентативных структур из наиболее заселенных кластеров при pH 3, 5, 7 на кристаллической структуре: N-концы окрашены, а остальная часть структуры — серого цвета 311
3.16 Выравнивание репрезентативных структур из наиболее заселенных кластеров при pH 3, 5, 7 на кристаллической структуре: C-концы окрашены, а остальная часть структуры — серого цвета 311

3.17 Конформационная изменчивость первых кластеров при pH 3 (a), pH 5 (b), pH 71 (c) и pH 72 (d); кластеризация проводилась с использованием аминокислот 38–231; однако изображения создаются с использованием всех аминокислот; на этих изображениях показана оболочка (тонкие линии) всех кадров в кластере; тонкие огибающие демонстрируют малые среднеквадратичные отклонения между кадрами, тогда как широкие огибающие демонстрируют большие среднеквадратичные отклонения 312
3.18 pH-зависимые области CP; (a) позиционное смещение репрезентативных структур из первых трех заселенных кластеров при pH 3 по сравнению с рентгеновской структурой; (b) C_α след репрезентативных структур из первых трех кластеров при pH 3 313
3.19 pH-зависимые области CP; (a) позиционное смещение репрезентативных структур из первых трех заселенных кластеров при pH 5 по сравнению с рентгеновской структурой; (b) C_α след репрезентативных структур из первых трех кластеров при pH 5 314
3.20 pH-зависимые области CP; (a) позиционное смещение репрезентативных структур из первых трех заселенных кластеров при pH 7_1 по сравнению с рентгеновской структурой; (b) C_α след репрезентативных структур из первых трех кластеров при pH 7_1 315
3.21 pH-зависимые области CP; (a) позиционное смещение репрезентативных структур из первых трех заселенных кластеров при pH 7_2 по сравнению с рентгеновской структурой; (b) C_α след репрезентативных структур из первых трех кластеров при pH 7_2 316
3.22 Соотношение выделенных структур распределеных в соответствии с расстояниями между остатками 168 и 127 a,b,c,d и остатками 172 и 127 e,f,g,h; при pH 3 – a,e; pH5 – b,f; pH 7 (первый минимум) – c,g; pH7 (2 минимум) – d,h; синяя вертикальная линия представляет расстояния между остатками в структуре капсида 317

И.1 Иллюстрация того, как наклон экспоненциального режима (τ^{-1}) в полулогарифмическом масштабе может меняться в зависимости от количества отсеченных с конца значений (N_{cut}) для силы $F = 550$ кДж/(нм · мол); последняя точка в данных, когда $N(t) = 1$, образует длинное горизонтальное плато, которое значительно увеличивает τ; после $N_{cut} = 2$, τ стабилизирует, а обрезка 2 точек обеспечивает надежность подгонки 319
И.2 Подгонка вероятностей выживания комплекса белок-лиганд с двумя режимами при различных внешних силах: черная пунктирная линия — неэкспоненциальный режим, синяя линия — экспоненциальный режим, черная штрихпунктирная линия — момент переключения между режимами 321
И.3 Распределения подобранного параметра, смоделированные методом Монте-Карло $\ln(\ln(\tau_0))$ (слева) и трансформированный τ_0 (справа) (длинный хвост при больших значениях обрезан, чтобы показать статистически значимую часть) 321

Список таблиц

1 Переменные, использованные в моделировании MD. r_{HB} — длина водородной связи, m_{H_2O} — масса молекулы воды, ε_{HB} — энергия водородной связи, ε_{LJ} — энергия Леннард-Джонса, σ^*_{LJ} — $0.7\, r^*_{HB}$. 42

2 Параметры моделирования, используемые в GROMACS для аргона и SPC/E воды и значения вязкости, используемые в коде LL-FH. 49

3 Приведенные единицы для жидкого аргона. 49

4 Количество ионов, добавленных к капсиду для нейтрализации заряда. 53

5 Параметры моделирования MD: L – размер кубической ячейки, t – время работы, Δt – временной шаг интегрирования, τ определяет постоянную времени для температурной связи, а T - целевая температура для температурной связи, R_{vdw} и R_{coul} расстояния обрезания (для электростатики Леннарда-Джонса и Particle-Mesh Ewald для кулоновских соответственно). 62

6 Коэффициент диффузии D получен в гибридном моделировании. 83

7 Окончательные значения D (погрешность для результатов чистого MD пренебрежимо мала) 86

8 Первоначальный состав моделируемой системы 144

9 Рассчитанные значения pKa титруемых аминокислот по сравнению со значениями pKa которые были вычисленны с использованием конформаций, извлеченных из минимумов pH 3, 5 и 7. «p» и «d» обозначают протонированную и депротонированную фазы соответственно. Аминокислоты Asp126, Asp127, Asp168 и Asp172 выделены жирным шрифтом 176

10 Последовательности обуславливающих состояний, описывающие два пути, реализуемые при выходе из состояния «0», и их вероятности, см. Рисунок 4.2 185

11 Водородные связи (ВС) боковых цепей белка с трегалозой и водой в % от общего числа ВС аминокислота-растворитель 200

12 Результаты расчетов энергии связывания лигандов с ЧСА в сайте IIA 211

13 Термодинамические параметры связывания веществ STR1-STR4 с ЧСА. ΔH, ΔS, ΔG – изменения энтальпии, энтропии и энергии Гиббса, K_a – константа связывания, n – стехиометрический коэффициент связывания 212

14 Время диссоциации τ и моменты перехода от дробно-экспоненциального к классическому режиму релаксации τ_{frac} 219

Приложение А

Предел возможностей метода Молекулярной Динамики

Сколько времени нужно на расчет одного шага MD по следующему алгоритму?

```
do step=1, nstep
  p = p + 0.5*dt*f
  r = r + dt*p/m
  f = force(r)
  p = p + 0.5*dt*f
end do
```

где `r` и `p` – координата и импульс атома, `f` – сила, действующая на атом и `dt` – временной шаг. Ответ: минимум $4 \cdot 3 = 12$ операций на шаг для одного атома.

За “операцию” будем упрощённо считать обмен данными между CPU и памятью. Допустим, что обмен происходит со скоростью света $3 \cdot 10^8$ м/с и расстояние между CPU и памятью 1 см. Тогда на одну операцию будет уходить $\frac{0.01}{3 \cdot 10^8} = \frac{1}{3} \cdot 10^{-10}$ секунды. А на один шаг MD – $\frac{12}{3} \cdot 10^{-10} = 4 \cdot 10^{-10}$ секунды.

Обычный шаг по времени в MD равен 2 фемтосекунды. Т.е. чтобы расчитать движение атома в течение 1 секунды надо $\frac{1}{2} \cdot 10^{15}$ шагов или $\frac{1}{2} \cdot 10^{15} \cdot 4 \cdot 10^{-10} = 2 \cdot 10^5$ секунд.

Сколько атомов в живом существе? Вот несколько примеров:

– самая маленькая живая клетка: $3 \cdot 10^8$,

– клетка человека: $1 \cdot 10^{14}$,

– условная “кошка”: 10^{10} клеток или $3 \cdot 10^8 \cdot 10^{10} = 3 \cdot 10^{18}$ атомов.

Естественно, будем вести расчёты параллельно на многих процессорах. Один из самых быстрых в мире суперкомпьютеров Fugaku состоит из 7 630 848 “ядер”. Допустим, что у нас в распоряжении 100 миллионов ($1 \cdot 10^8$) ядер. Тогда для расчёта жизни в течение 1 секунды нужно

– самая маленькая живая клетка: $\frac{3 \cdot 10^8}{10^8} = 3$ цикла или $3 \cdot 2 \cdot 10^5$ секунд = **7 дней**,

– клетка человека: $\frac{10^{14}}{10^8} = 10^6$ циклов или $10^6 \cdot 2 \cdot 10^5$ секунд = **6342 года**,

– условная “кошка”: $\frac{3 \cdot 10^{18}}{10^8} = 3 \cdot 10^{10}$ циклов или **190 258 752 года**.

Другими словами, чтобы промоделировать 1 секунду жизни условной “кошки” “идеальному” компьютеру современной архитектуры понадобилось бы *несколько сотен миллионов лет расчетов*. Т.е. моделирование организма с атомистическим разрешением *принципиально невозможно*.

Приложение Б

Вывод уравнений движения с учетом сохранения массы и импульса

Б.1 Сохранение массы

Смесь фаз FH и MD имеет плотность

$$\tilde{\rho} = s\rho + (1-s)\sum_p \rho_p \tag{Б.1}$$

и импульс

$$\tilde{u}_j\tilde{\rho} = su_j\rho + (1-s)\sum_p \rho_p u_{jp}. \tag{Б.2}$$

Обратите внимание, что $\sum_p$ берется по ячейке и включает в себя $N(t)$ частиц в ячейке.

Нам требуется сохранение массы и импульса смеси «жидкостей» FH и MD при всех значениях концентрации s.

Законы сохранения массы каждой из фаз имеют вид:

$$\frac{\partial}{\partial t}s\rho + \frac{\partial}{\partial x_i}\tilde{u}_i s\rho = J^\rho, \tag{Б.3}$$

$$\frac{\partial}{\partial t}(1-s)\sum_p \rho_p + \frac{\partial}{\partial x_i}(1-s)\sum_p u_{ip}\rho_p = -J^\rho, \tag{Б.4}$$

где J^ρ — сток/источник, который переносит массу между фазами, а скорость переноса в (Б.3) — сохраняющаяся скорость $\tilde{u}_i$, то есть средняя скорость смеси.

Для выполнения закона сохранения массы введем следующий динамический закон:

$$\frac{D}{Dt_0}\left(\tilde{\rho} - \sum_p \rho_p\right) = L^\rho \cdot \left(\tilde{\rho} - \sum_p \rho_p\right), \tag{Б.5}$$

где $\frac{D}{Dt_0} = \frac{\partial}{\partial t}\cdot + \frac{\partial}{\partial x_i}\tilde{u}_i\cdot$ и L^ρ — оператор, который приводит соответствующее отклонение к заданному значению в пределах зоны $0 < s < 1$ и возвращает 0 при $s = 0$ и $s = 1$.

Для текущей реализации была использована

$$L^\rho \cdot \left(\tilde{\rho} - \sum_p \rho_p\right) = \frac{\partial}{\partial x_i}\left[s\,(1-s)\,\alpha\frac{\partial}{\partial x_i}\left(\tilde{\rho} - \sum_p \rho_p\right)\right], \tag{Б.6}$$

управляемой параметром $\alpha > 0$.

Ур. (Б.5) можно переписать как

$$\frac{\partial}{\partial t}\left(\tilde{\rho}-\sum_p \rho_p\right)+\frac{\partial}{\partial x_i}\tilde{u}_i\left(\tilde{\rho}-\sum_p \rho_p\right)=\frac{\partial}{\partial x_i}\left[s\left(1-s\right)\alpha\frac{\partial}{\partial x_i}\left(\tilde{\rho}-\sum_p \rho_p\right)\right], \tag{Б.7}$$

и подставив $\tilde{\rho}$ в уравнении выше получаем разницу плотностей между фазами FH и MD:

$$\frac{\partial}{\partial t}s\left(\rho-\sum_p \rho_p\right)+\frac{\partial}{\partial x_i}\tilde{u}_i s\left(\rho-\sum_p \rho_p\right)=\frac{\partial}{\partial x_i}\left[s\left(1-s\right)\alpha\frac{\partial}{\partial x_i}\left(\tilde{\rho}-\sum_p \rho_p\right)\right]. \tag{Б.8}$$

В ур. (Б.8) собираем члены для выражения J^{ρ} в соответствии с уравнением (Б.3)

$$J^{\rho}=\frac{\partial}{\partial t}s\sum_p \rho_p+\frac{\partial}{\partial x_i}\tilde{u}_i s\sum_p \rho_p+\frac{\partial}{\partial x_i}\left[s\left(1-s\right)\alpha\frac{\partial}{\partial x_i}\left(\tilde{\rho}-\sum_p \rho_p\right)\right] \tag{Б.9}$$

Замена J^{ρ} в уравнении (Б.4) дает

$$\begin{aligned}\frac{\partial}{\partial t}(1-s)\sum_p \rho_p+\frac{\partial}{\partial x_i}(1-s)\sum_p \rho_p u_{ip}=-\frac{\partial}{\partial t}s\sum_p \rho_p-\frac{\partial}{\partial x_i}\tilde{u}_i s\sum_p \rho_p-\\ -\frac{\partial}{\partial x_i}\left[s\left(1-s\right)\alpha\frac{\partial}{\partial x_i}\left(\tilde{\rho}-\sum_p \rho_p\right)\right],\end{aligned} \tag{Б.10}$$

что можно упростить до

$$\begin{aligned}\frac{\partial}{\partial t}\sum_p \rho_p+\frac{\partial}{\partial x_i}s\tilde{u}_i\sum_p \rho_p+\frac{\partial}{\partial x_i}(1-s)\sum_p \rho_p u_{ip}=\\ =-\frac{\partial}{\partial x_i}\left[s\left(1-s\right)\alpha\frac{\partial}{\partial x_i}\left(\tilde{\rho}-\sum_p \rho_p\right)\right].\end{aligned} \tag{Б.11}$$

Уравнения сохранения массы в виде

$$\frac{\partial}{\partial t}\sum_p \rho_p+\frac{\partial}{\partial x_i}\sum_p \frac{dx_{ip}}{dt}\rho_p=0 \tag{Б.12}$$

дает измененные скорости $\frac{dx_{ip}}{dt}$ частиц MD, которые включают источник J^{ρ}.

Объединяя ур. (Б.12) с ур. (Б.11) дает

$$\begin{aligned}-\frac{\partial}{\partial x_i}\sum_p \frac{dx_{ip}}{dt}\rho_p+\frac{\partial}{\partial x_i}\tilde{u}_i\sum_p \rho_p+\frac{\partial}{\partial x_i}(1-s)\sum_p \rho_p u_{ip}=\\ =-\frac{\partial}{\partial x_i}\left[s\left(1-s\right)\alpha\frac{\partial}{\partial x_i}\left(\tilde{\rho}-\sum_p \rho_p\right)\right].\end{aligned} \tag{Б.13}$$

Интегрируя уравнение выше, получаем

$$\sum_p \frac{dx_{ip}}{dt}\rho_p=\tilde{u}_i\sum_p \rho_p+(1-s)\sum_p \rho_p u_{ip}+\left[s\left(1-s\right)\alpha\frac{\partial}{\partial x_i}\left(\tilde{\rho}-\sum_p \rho_p\right)\right], \tag{Б.14}$$

для каждой частицы в ячейке имеем

$$\frac{dx_{ip}}{dt} = u_{ip} + s(\tilde{u}_i - u_{ip}) + s(1-s)\alpha(x)\frac{1}{\rho_p N(t)}\frac{\partial}{\partial x_i}\left(\tilde{\rho} - \sum_p \rho_p\right), \tag{Б.15}$$

где $N(t)$ — количество частиц в ячейке.

Б.2 Сохранение импульса

Процедура аналогичная процедуре сохранения массы применена к сохранению импульса. Для «жидкости» FH используем следующее уравнение сохранения импульса:

$$\frac{\partial}{\partial t} s\rho u_j + \frac{\partial}{\partial x_i}\tilde{u}_i u_j s\rho = sF_j + J^u, \tag{Б.16}$$

и для MD «жидкости»

$$\frac{\partial}{\partial t}(1-s)\sum_p \rho_p u_{jp} + \frac{\partial}{\partial x_i}(1-s)\sum_p u_{ip}u_{jp}\rho_p = \frac{(1-s)}{V_{cell}}\sum_p F_{jp} - J^u. \tag{Б.17}$$

Для выполнения закона сохранения импульса введем следующий динамический закон

$$\frac{D}{Dt_0}\left(\tilde{u}_j\tilde{\rho} - \sum_p \rho_p u_{jp}\right) = L^u \cdot \left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp}\right) + sF_j, \tag{Б.18}$$

где оператор внешней силы равен

$$L^u \cdot \left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp}\right) = \frac{\partial}{\partial x_j}\left[s\,(1-s)\,\beta\frac{\partial}{\partial x_j}\left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp}\right)\right], \tag{Б.19}$$

и $\beta > 0$.

Расписывая ур. (Б.18) получаем:

$$\begin{aligned}\frac{\partial}{\partial t}\left(\tilde{u}_j\tilde{\rho} - \sum_p \rho_p u_{jp}\right) + \frac{\partial}{\partial x_i}\tilde{u}_i\left(\tilde{u}_j\tilde{\rho} - \sum_p \rho_p u_{jp}\right) = sF_j + \\ + \frac{\partial}{\partial x_j}\left[s\,(1-s)\,\beta\frac{\partial}{\partial x_j}\left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp}\right)\right],\end{aligned} \tag{Б.20}$$

где $\tilde{u}_i\tilde{\rho}$ можно заменить на (Б.2) дающее

$$\begin{aligned}\frac{\partial}{\partial t}s\left(u_j\rho - \sum_p \rho_p u_{jp}\right) + \frac{\partial}{\partial x_i}s\tilde{u}_i\left(u_j\rho - \sum_p \rho_p u_{jp}\right) = sF_j + \\ + \frac{\partial}{\partial x_j}\left[s\,(1-s)\,\beta\frac{\partial}{\partial x_j}\left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp}\right)\right].\end{aligned} \tag{Б.21}$$

В ур. (Б.21) собираем члены для явного выражения J^u в соответствии с уравнением (Б.16)

$$J^u = \frac{\partial}{\partial t} s \sum_p \rho_p u_{jp} + \frac{\partial}{\partial x_i} s\tilde{u}_i \sum_p \rho_p u_{jp} + \frac{\partial}{\partial x_j}\left[s\,(1-s)\,\beta \frac{\partial}{\partial x_j}\left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp} \right) \right] \tag{Б.22}$$

Если известна скорость обмена импульсом J^u, ее можно подставить в уравнение сохранения импульса для MD (Б.17)

$$\frac{\partial}{\partial t}\sum_p \rho_p u_{jp} + \frac{\partial}{\partial x_i}(1-s)\sum_p u_{ip}u_{jp}\rho_p = \frac{(1-s)}{V_{cell}}\sum_p F_{jp} - \frac{\partial}{\partial x_i} s\tilde{u}_i \sum_p \rho_p u_{jp} - \tag{Б.23}$$
$$-\frac{\partial}{\partial x_j}\left[s\,(1-s)\,\beta \frac{\partial}{\partial x_j}\left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp} \right) \right].$$

Для уравнения сохранения импульса в виде

$$\frac{\partial}{\partial t}\sum_p \rho_p u_{jp} + \frac{\partial}{\partial x_i}\sum_p \frac{dx_{ip}}{dt} u_{jp}\rho_p = \sum_p \rho_p \frac{du_{jp}^N}{dt}, \tag{Б.24}$$

что приводит к измененным силам MD ‘жидкости’ $\frac{du_{jp}^N}{dt}$, который включает в себя скорость обменна импульса J^u.

Ур. (Б.24) и ур. (Б.23) дает

$$\frac{\partial}{\partial x_i}\sum_p \frac{dx_{ip}}{dt} u_{jp}\rho_p = \frac{\partial}{\partial x_i} s\tilde{u}_i \sum_p u_{jp}\rho_p + \tag{Б.25}$$
$$+\frac{\partial}{\partial x_i}(1-s)\sum_p u_{jp}\rho_p u_{ip} + \frac{\partial}{\partial x_i}\left[s\,(1-s)\,\alpha \frac{\partial}{\partial x_i}\left(\tilde{\rho} - \sum_p \rho_p \right) \right] \frac{\sum_p u_{jp}}{N(t)},$$

измененная сила/ускорение, действующая на частицы в ячейке, равна:

$$\sum_p \rho_p \frac{du_{jp}^N}{dt} = \frac{(1-s)}{V_{cell}}\sum_p F_{jp} + \tag{Б.26}$$
$$+\frac{\partial}{\partial x_i}\left[s\,(1-s)\,\alpha \frac{\partial}{\partial x_i}\left(\tilde{\rho} - \sum_p \rho_p \right) \right] \frac{\sum_p u_{jp}}{N(t)} - \frac{\partial}{\partial x_j}\left[s\,(1-s)\,\beta \frac{\partial}{\partial x_j}\left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp} \right) \right].$$

Для отдельных частиц получаем

$$\frac{du_{jp}^N}{dt} = \frac{(1-s)}{\rho_p V_{cell}} F_{jp} + \frac{1}{\rho_p N(t)}\frac{\partial}{\partial x_i}\left[s(1-s)\alpha(x)\frac{\partial}{\partial x_i}\left(\tilde{\rho} - \sum_p \rho_p \right) \frac{\sum_p u_{jp}}{N(t)} \right] - \tag{Б.27}$$
$$-\frac{1}{\rho_p N(t)}\frac{\partial}{\partial x_i}\left[s(1-s)\beta(x)\frac{\partial}{\partial x_i}\left(\tilde{u_j}\tilde{\rho} - \sum_p \rho_p u_{jp} \right) \right].$$

Б.3 Гибридная вода модели MB

Вводя новые переменные

$$\rho' = \tilde{\rho} - \sum_p \rho_p, \tag{Б.28}$$

$$u'\rho' = \tilde{u}_j\tilde{\rho} - \sum_p \rho_p u_{jp} \tag{Б.29}$$

уравнения LL-FH можно записать как:

$$\frac{\partial}{\partial t}\rho' + \frac{\partial}{\partial x_i}\tilde{u}_i\rho' = Q_\rho, \tag{Б.30}$$

где

$$Q_\rho = \frac{\partial}{\partial x_i}\left[s\left(1-s\right)\alpha\frac{\partial}{\partial x_i}\left(\rho'\right)\right] \tag{Б.31}$$

и

$$\frac{\partial}{\partial t}u_j'\rho' + \frac{\partial}{\partial x_i}\tilde{u}_i u_j'\rho' = sF_j + Q_u, \tag{Б.32}$$

где

$$Q_u = \frac{\partial}{\partial x_j}\left[s\left(1-s\right)\beta\frac{\partial}{\partial x_j}\left(u_j'\rho'\right)\right]. \tag{Б.33}$$

Чистая сила FH используется из классических уравнений LL-FH

$$F_j = \frac{\partial}{\partial x_i}\left(\Pi_{ij} + \tilde{\Pi}_{ij}\right), \tag{Б.34}$$

где тензор напряжений рассчитывается как

$$\Pi_{ij} = -\left(p - \xi\nabla\cdot\vec{u}\right)\delta_{ij} + \eta\left(\partial_i u_j + \partial_j u_i - 2D^{-1}\nabla\vec{u}\cdot\delta_{ij}\right), \tag{Б.35}$$

и тензор стохастического напряжения как

$$\tilde{\Pi}_{ij} = \sqrt{\frac{2k_BT}{\delta t\cdot V_{cell}}}\left(\sqrt{2}\sqrt{\eta}\cdot G_{ij}^s + \sqrt{D}\sqrt{\xi}\frac{tr[G]}{D}E_{ij}\right) \tag{Б.36}$$

где $\partial_i = \frac{\partial}{\partial x_i}$ и G — гауссова матрица, $G_{ij}^s = \frac{G_{ij}+G_{ij}^T}{2} - \frac{tr[G]}{D}E_{ij}$, E_{ij} — единичная матрица, а $tr[]$ обозначает след матрицы.

Уравнения LL-FH решаются с помощью центрированной конечно-разностной схемы второго порядка, основанной на характеристическом разложении.

Приложение В

Вывод гибридных уравнений движения в дифференциальной форме

Подставляя выражения для $\frac{d\mathbf{x}_p}{dt}$ и $\frac{du_{ip}}{dt}$ из (1.35) в (1.34) дает:

$$\delta_t\left(\sum_{p=1,N(t)} m_p\right)+\sum_{\gamma=1,6}\left(\sum_{p=1,N_\gamma(t)}\mathbf{u}_p\rho_p\right)d\mathbf{n}^\gamma\cdot\delta t=-\sum_{\gamma=1,6}\left(\sum_{p=1,N_\gamma(t)}s(\bar{\mathbf{u}}-\mathbf{u}_p)\rho_p\right)d\mathbf{n}^\gamma\cdot\delta t$$
$$-\sum_{\lambda=1,6}\left(s(1-s)\cdot\alpha\cdot\tfrac{1}{V}\sum_{\gamma=1,6}\left(\bar{\rho}-\sum_{q=1,N_\gamma(t)}\rho_q\right)d\mathbf{n}^\gamma\right)d\mathbf{n}^\lambda\cdot\delta t \tag{B.1}$$

и

$$\delta_t\left(\sum_{p=1,N(t)} m_p u_{ip}\right)+\sum_{\gamma=1,6}\left(\sum_{p=1,N_\gamma(t)}\mathbf{u}_p\rho_p u_{ip}\right)d\mathbf{n}^\gamma\cdot\delta t=\sum_{p=1,N(t)}(1-s)F_{ip}\cdot\delta t$$
$$+\sum_{k=1,3}\sum_{\gamma=1,6}\left(s(1-s)\beta\tfrac{1}{V}\sum_{\lambda=1,6}\left(\bar{\rho}\cdot\bar{u}_i-\sum_{q=1,N_\lambda(t)}\rho_q u_{iq}\right)dn_k^\lambda\right)dn_k^\gamma\,\delta t- \tag{B.2}$$
$$\sum_{\gamma=1,6}\left(\sum_{p=1,N_\gamma(t)}s(\bar{\mathbf{u}}-\mathbf{u}_p)\rho_p u_{ip}\right)d\mathbf{n}^\gamma\delta t$$

соответственно.

Вычитая из уравнений следующие истинные тождества для плотности и импульса (B.1) и (B.2), соответственно,

$$\begin{gathered}\delta_t\left(s\sum_{p=1,N(t)} m_p\right)+\sum_{\gamma=1,6}s\left(\sum_{p=1,N_\gamma(t)}\mathbf{u}_p\rho_p\right)d\mathbf{n}^\gamma\delta t=\\ \delta_t\left(s\sum_{p=1,N(t)} m_p\right)+\sum_{\gamma=1,6}s\left(\sum_{p=1,N_\gamma(t)}\mathbf{u}_p\rho_p\right)d\mathbf{n}^\gamma\delta t\end{gathered} \tag{B.3}$$

и

$$\begin{gathered}\delta_t\left(s\sum_{p=1,N(t)} m_p u_{ip}\right)+\sum_{\gamma=1,6}s\left(\sum_{p=1,N_\gamma(t)}\mathbf{u}_p\rho_p u_{ip}\right)d\mathbf{n}^\gamma\delta t=\\ =\delta_t\left(s\sum_{p=1,N(t)} m_p u_{ip}\right)+\sum_{\gamma=1,6}s\left(\sum_{p=1,N_\gamma(t)}\mathbf{u}_p\rho_p u_{ip}\right)d\mathbf{n}^\gamma\delta t,\end{gathered} \tag{B.4}$$

получаются следующие уравнения:

$$\delta_t\left((1-s)\sum_{p=1,N(t)} m_p\right)+\sum_{\gamma=1,6}(1-s)\left(\sum_{p=1,N_\gamma(t)}\mathbf{u}_p\rho_p\right)d\mathbf{n}^\gamma\delta t=-\delta_t\left(s\sum_{p=1,N(t)} m_p\right)-$$
$$-\sum_{\gamma=1,6}\left(\sum_{p=1,N_\gamma(t)}s\bar{\mathbf{u}}\rho_p\right)d\mathbf{n}^\gamma\delta t-\sum_{\lambda=1,6}\left(s(1-s)\cdot\alpha\cdot\tfrac{1}{V}\sum_{\gamma=1,6}\left(\bar{\rho}-\sum_{q=1,N_\gamma(t)}\rho_q\right)d\mathbf{n}^\gamma\right)d\mathbf{n}^\lambda\delta t \tag{B.5}$$

и

$$\delta_t\left((1-s)\sum_{p=1,N(t)} m_p u_{ip}\right)+\sum_{\gamma=1,6}(1-s)\left(\sum_{p=1,N_\gamma(t)}\mathbf{u}_p\rho_p u_{ip}\right)d\mathbf{n}^\gamma\delta t-\sum_{p=1,N(t)}(1-s)F_{ip}\,\delta t=$$
$$-\delta_t\left(s\sum_{p=1,N(t)} m_p u_{ip}\right)-\sum_{\gamma=1,6}\left(\sum_{p=1,N_\gamma(t)} s\bar{\mathbf{u}}\rho_p u_{ip}\right)d\mathbf{n}^\gamma\delta t+$$
$$+\sum_{k=1,3}\sum_{\gamma=1,6}\left(s(1-s)\cdot\beta\cdot\tfrac{1}{V}\sum_{\lambda=1,6}\left(\bar{\rho}\cdot\bar{u}_i-\sum_{q=1,N_\gamma(t)}\rho_q u_{iq}\right)dn_k^\lambda\right)dn_k^\gamma\,\delta t. \tag{B.6}$$

Сравнение (B.5) и (B.5) с (1.27) и (1.29) соответственно дает:

$$\delta_t J^{(\rho)}=\delta_t\left(s\sum_{p=1,N(t)} m_p\right)+\sum_{\gamma=1,6}\left(\sum_{p=1,N_\gamma(t)} s\bar{\mathbf{u}}\rho_p\right)d\mathbf{n}^\gamma\cdot\delta t+$$
$$+\sum_{\lambda=1,6}\left(s(1-s)\cdot\alpha\cdot\tfrac{1}{V}\sum_{\gamma=1,6}\left(\bar{\rho}-\sum_{q=1,N_\gamma(t)}\rho_q\right)d\mathbf{n}^\gamma\right)d\mathbf{n}^\lambda\delta t \tag{B.7}$$

и

$$\delta_t J_i^{(\mathbf{u})}=\delta_t\left(s\sum_{p=1,N(t)} m_p u_{ip}\right)+\sum_{\gamma=1,6}\left(\sum_{p=1,N_\gamma(t)} s\bar{\mathbf{u}}\rho_p u_{ip}\right)d\mathbf{n}^\gamma\delta t-$$
$$-\sum_{k=1,3}\sum_{\gamma=1,6}\left(s(1-s)\cdot\beta\cdot\tfrac{1}{V}\sum_{\lambda=1,6}\left(\bar{\rho}\cdot\bar{u}_i-\sum_{q=1,N_\gamma(t)}\rho_q u_{iq}\right)dn_k^\lambda\right)dn_k^\gamma\,\delta t. \tag{B.8}$$

Подставляя приведенные выше выражения (B.7) и (B.8) в уравнения «фазы» континуума (1.26) и (1.28), в итоге получаем по массе

$$\delta_t(sm)+\sum_{\gamma=1,6}(s\rho\bar{\mathbf{u}})\,d\mathbf{n}^\gamma\delta t=\delta_t\left(s\sum_{p=1,N(t)} m_p\right)+\sum_{\gamma=1,6}\left(\sum_{p=1,N_\gamma(t)} s\bar{\mathbf{u}}\rho_p\right)d\mathbf{n}^\gamma\delta t+$$
$$+\sum_{\lambda=1,6}\left(s(1-s)\cdot\alpha\cdot\tfrac{1}{V}\sum_{\gamma=1,6}\left(\bar{\rho}-\sum_{q=1,N_\gamma(t)}\rho_q\right)d\mathbf{n}^\gamma\right)d\mathbf{n}^\lambda\delta t \tag{B.9}$$

и импульсу

$$\delta_t(smu_i)+\sum_{\gamma=1,6}(s\rho u_i\bar{\mathbf{u}})\,d\mathbf{n}^\gamma\delta t=\delta_t\left(s\sum_{p=1,N(t)} m_p u_{ip}\right)+\sum_{\gamma=1,6}\left(\sum_{p=1,N_\gamma(t)} s\bar{\mathbf{u}}\rho_p u_{ip}\right)d\mathbf{n}^\gamma\delta t+$$
$$+s\sum_{j=1,3}\sum_{\gamma=1,6}\left(\Pi_{ij}+\tilde{\Pi}_{ij}\right)dn_j^\gamma\,\delta t-$$
$$-\sum_{k=1,3}\sum_{\gamma=1,6}\left(s(1-s)\cdot\beta\cdot\tfrac{1}{V}\sum_{\lambda=1,6}\left(\bar{\rho}\cdot\bar{u}_i-\sum_{q=1,N_\gamma(t)}\rho_q u_{iq}\right)dn_k^\lambda\right)dn_k^\gamma\,\delta t \tag{B.10}$$

со следующими истинными равенствами

$$\delta_t\left((1-s)\sum_{p=1,N(t)} m_p\right)+\sum_{\gamma=1,6}(1-s)\left(\sum_{p=1,N_\gamma(t)}\bar{\mathbf{u}}\rho_p\right)d\mathbf{n}^\gamma\delta t=$$
$$=\delta_t\left((1-s)\sum_{p=1,N(t)} m_p\right)+\sum_{\gamma=1,6}(1-s)\left(\sum_{p=1,N_\gamma(t)}\bar{\mathbf{u}}\rho_p\right)d\mathbf{n}^\gamma\delta t \tag{B.11}$$

и

$$\begin{aligned}&\delta_t\left((1-s)\sum_{p=1,N(t)}m_p u_{ip}\right)+\sum_{\gamma=1,6}(1-s)\left(\sum_{p=1,N_\lambda(t)}s\bar{\mathbf{u}}\rho_p u_{ip}\right)d\mathbf{n}^\gamma\delta t=\\&=\delta_t\left((1-s)\sum_{p=1,N(t)}m_p u_{ip}\right)+\sum_{\gamma=1,6}(1-s)\left(\sum_{p=1,N_\gamma(t)}s\bar{\mathbf{u}}\rho_p u_{ip}\right)d\mathbf{n}^\gamma\delta t\end{aligned}\tag{B.12}$$

соответственно, и используя переменные плотности и импульса смеси, $\bar{\rho}=s\rho+(1-s)\sum\limits_{p=1,N(t)}\rho_p$ и $\bar{m}\bar{u}_i=smu_i+(1-s)\sum\limits_{p=1,N(t)}m_p u_{ip}$, соответственно, в конечном итоге дает:

$$\begin{aligned}&\delta_t\left(\bar{m}-\sum_{p=1,N(t)}m_p\right)+\sum_{\gamma=1,6}\bar{\mathbf{u}}\left(\bar{\rho}-\sum_{q=1,N_\gamma(t)}\rho_q\right)d\mathbf{n}^\gamma\delta t=\\&-\sum_{\lambda=1,6}\left(s(1-s)\cdot\alpha\cdot\tfrac{1}{V}\sum_{\gamma=1,6}\left(\bar{\rho}-\sum_{q=1,N_\gamma(t)}\rho_q\right)d\mathbf{n}^\gamma\right)d\mathbf{n}^\lambda\delta t\end{aligned}\tag{B.13}$$

и

$$\begin{aligned}&\delta_t(\bar{m}\cdot\bar{u}_i-\sum_{p=1,N(t)}m_p u_{ip})+\sum_{\gamma=1,6}s\bar{\mathbf{u}}\left(\bar{\rho}\cdot\bar{u}_i-\sum_{p=1,N_\gamma(t)}\rho_p u_{ip}\right)d\mathbf{n}^\gamma\delta t=\\&=s\sum_{j=1,3}\sum_{\gamma=1,6}\left(\Pi_{ij}+\tilde{\Pi}_{ij}\right)dn_j^\gamma\,\delta t-\\&-\sum_{k=1,3}\sum_{\gamma=1,6}\left(s(1-s)\cdot\beta\cdot\tfrac{1}{V}\sum_{\lambda=1,6}\left(\bar{\rho}\cdot\bar{u}_i-\sum_{q=1,N_\lambda(t)}\rho_q u_{iq}\right)dn_k^\lambda\right)dn_k^\gamma\,\delta t.\end{aligned}\tag{B.14}$$

Полученные уравнения (B.13) и (B.14) идентичны (1.34).

Приложение Г

Вычислительная механика

Рассматриваются все прошлые s_i^- и будущие s_i^+ половины бесконечных в обе стороны символьных последовательностей с центрами в точках времени i. Два прошлых события s_1^- и s_2^- определяются как эквивалентные, если условные распределения по их будущим событиям $P(s^+|s_1^-)$ и $P(s^+|s_2^-)$ равны. *Обуславливающее состояние* $\varepsilon(s_i^-)$ — это набор всех прошлых эквивалентных s_i^-: $\varepsilon_i \equiv \varepsilon(s_i^-) = \{\lambda : P(s^+|\lambda) = P(s^+|s_i^-)\}$. В каждый момент система находится в одном из обусловливающих состояний и переходит в следующее с вероятностью, заданной матрицей перехода $T_{ij}\lambda P(\varepsilon_j|\varepsilon_i)$. Матрица перехода определяет асимптотические вероятности обусловливающих состояний как ее левый собственный вектор $P(\varepsilon_i)T = P(\varepsilon_i)$, где $\Sigma_i P(\varepsilon_i) = 1$. Совокупность обусловливающих состояний вместе с вероятностями перехода определяет ε-*машину*.

Доказано [392], что ε-машина есть

- *достаточная* статистика, то есть содержащая полную статистическую информацию о данных;
- *минимальная достаточная* статистика, поэтому обуславливающие состояния не могут быть разделены на более мелкие состояния;
- *уникальная минимальная достаточная* статистика, любая другая просто переименовывает те же самые состояния.

Приложение Д

Моделирование вируса MS2

Д.1 Детали моделирования вируса MS2

Для построения начальных координат для моделирования сначала расположили 60 копий тримерного блока в соответствии с икосаэдрической симметрией, что образовало замкнутую структуру. Затем белок созревания и соседние деформированные мономеры были введены путем замены 2-х и 5-ти мономеров соответственно.

Наконец, последовательность, содержащая остатки 68-77, была скопирована из мономера, обозначенного как цепь A в файле PDB, и вручную выровнена для заполнения пробела в мономере, обозначенном как цепь G. Чтобы хорошо подогнать ее и устранить стерические столкновения, весь мономер был подвергнут минимизации энергии, при которой все атомы, кроме реконструированной последовательности, были заморожены. Здесь и далее для этих целей использовался алгоритм наискорейшего спуска; минимизация проводилась до тех пор, пока наибольшая остаточная сила не становилась меньше 5000 кДж нм^{1}, что гарантировало отсутствие сбоев при последующем моделировании MD. Полученную структуру скопировали в капсид. На этом процесс сборки был завершен, рисунок 3.2. Мы не проводили реконструкцию недостающих последовательностей в белке созревания, поскольку они расположены далеко от стенки капсида и, таким образом, не влияют на исследуемые свойства.

Атомы водорода были добавлены к аминокислотам, а состояния протонирования были установлены с помощью инструмента *gmx pdb2gmx* с настройками по умолчанию.

Д.2 Подготовка ячейки моделирования

Поскольку капсид имеет почти сферическую форму, для коробки моделирования была выбрана форма ромбического додекаэдра, чтобы минимизировать ее объем. Длина стороны была установлена в 21,4 нм, что обеспечивает запас 2×5,5 нм вокруг капсида, поскольку применялись периодические граничные условия. В триклинном представлении ячейки размеры составляют 37×37×37 нм с углами $\alpha = \beta = 60°$, $\gamma = 90°$. Молекулы воды были добавлены в коробку с помощью *gmx solvate*. Проведена минимизация энергии.

Капсид разделяет систему на две части: внутреннюю полость и внешний раствор. Капсид несет высокий суммарный заряд (+192 e), который распределен неравномерно: его внутренняя поверхность заряжена положительно, а внешняя — отрицательно. Поэтому нейтрализацию заряда следует проводить для каждой части отдельно. Мы следовали процедуре, которую использовали

в нашей работе по PCV2 [7; 22; 25]. Был вычислен суммарный заряд атомов капсида в сфере радиусом r с центром в центре масс капсида (ЦМ), рисунок Д.1. Полученная кривая была сглажена путем скользящего усреднения с окном 0.3 нм для удаления шума.

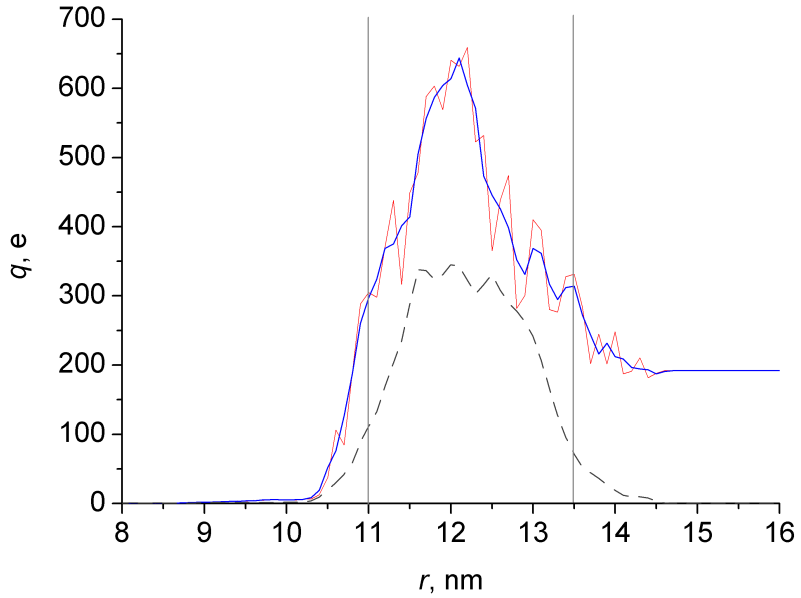


Рисунок Д.1 — Суммарный заряд атомов капсида в сфере радиуса r с центром в центре масс капсида (красная кривая); синяя кривая — сглаженный исходный заряд; пунктирная серая кривая показывает радиальную функцию распределения атомов капсида (масштабирована по вертикали для наглядности); вертикальные линии обозначают границы, охватывающие 90% атомов капсида

Интегральный заряд увеличивается с расстоянием до +644 e при 12.1 нм, а затем уменьшается до +192 e. Положение максимума примерно соответствует середине стенки капсида. Эти факты указывают на то, что капсид сильно поляризован, внутренняя поверхность имеет заряд +644 e, а внешняя поверхность заряжена –452 e. Согласно этой информации, 644 иона Cl^- были помещены в полость капсида, а 452 иона Na^+ были распределены во внешнем растворе. Для этого использовались утилиты *gmx genion* и скрипты написанные нами. Наконец, фоновый электролит в количестве 2971 ионов Na^+ и Cl^- был случайным образом распределен по всей ячейке, что сделало общую концентрацию соли соответствующей концентрации физиологического раствора (0.9% по массе NaCl). Это число рассчитывалось на основе общего числа молекул воды в клетке и количественного соотношения 1:358 между молекулами NaCl и воды в физиологическом растворе.

На этом сборка первоначальной модели капсида MS2 в физиологическом растворе была завершена.

Приложение Е

Детали распределения ионов на внутренней поверхности капсида PCV2

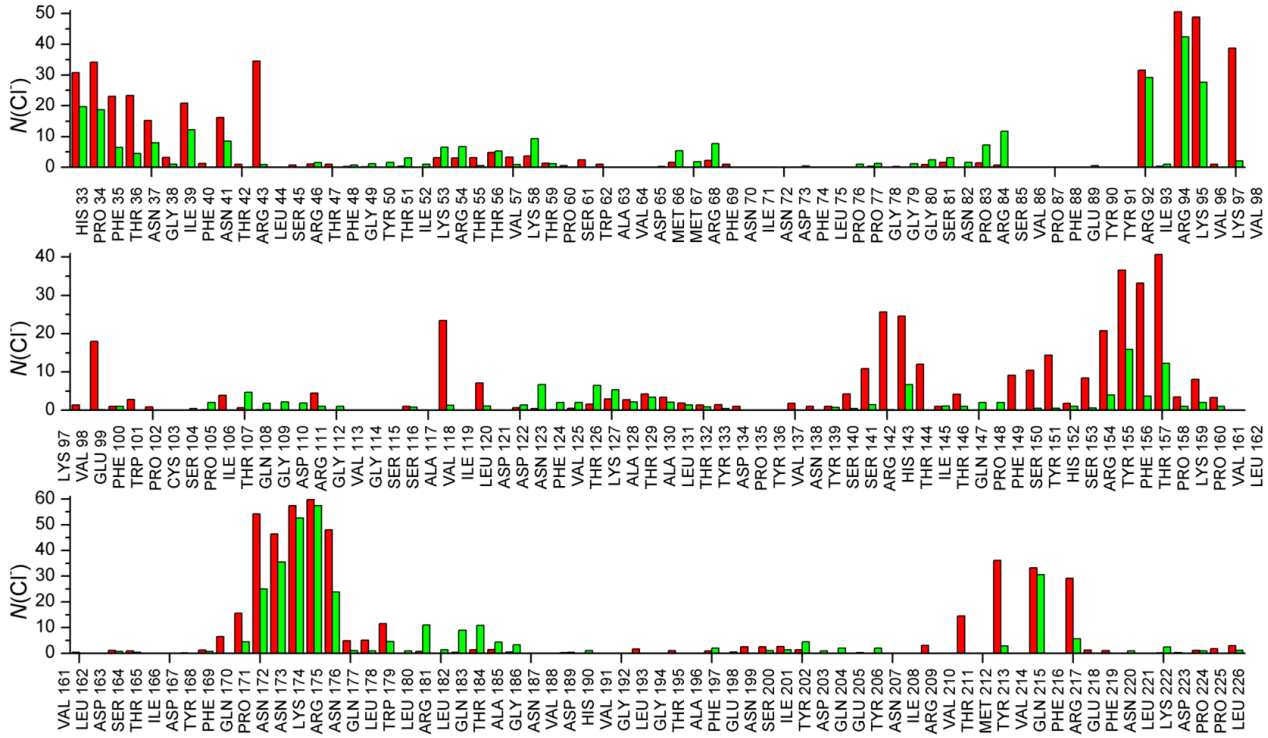


Рисунок Е.1 — Количество остатков, имеющих ион Cl^- ближе, чем 5; красный - естественная система, зеленый - искусственная система

Приложение Ж

Подробности моделирования PCV2

Подробное описание подготовки конструкций для моделирования, процесса моделирования, графики *RMSD*(t) и $R_g(t)$ для силового поля GROMOS, распределение заряда в конечных конфигурациях.

Было выполнено моделирование MD с использованием GROMACS. Для моделирования MD всех моделей в GROMOS96 применялась модель воды SPC силового поля 53a6, а в Amber03 использовалась модель воды TIP3P. Наши расчеты не включали ДНК.

Моделирование MD для всех случаев проводилось с периодическими граничными условиями в кубической коробке со стороной около 27 нм, начальные конфигурации с ионами были подготовлены следующим образом. Сначала кристаллографическая конфигурация была сольватирована в воде с использованием утилиты genbox GROMACS. Коробка представляла собой куб с ребром 26.6 нм. Затем с помощью нашего программного обеспечения из системы была извлечена сфера радиусом 3 нм (для капсида с хвостами) или 4.5 нм (для капсида без хвостов), расположенная в центре масс капсида. Радиусы были выбраны таким образом, чтобы охватывать только молекулы воды без атомов капсида. Молекулы сохранялись целыми, т.е. все атомы, связанные с охватываемыми сферой, включались в последнюю. Затем необходимое количество случайно выбранных молекул воды в извлеченной сфере (407 для капсида с хвостами и 606 для капсида без хвостов) было заменено ионами Cl^- с помощью утилиты genion GROMACS. Полученное решение было помещено обратно в основную систему путем объединения их pdb-файлов в текстовом редакторе. После этого из системы с помощью собственного программного обеспечения был извлечен сферический слой радиусом 11 нм и толщиной 0.5 нм. Как и прежде, в нем была только вода. Случайно выбранные молекулы воды в слое были заменены 227 (для капсида с хвостами) или 246 (для капсида без хвостов) ионами Cl^- с помощью утилиты genion. Полученный раствор был помещен обратно в основную систему в текстовом редакторе. Наконец, путем замены молекул воды с помощью утилиты genion были добавлены ионы Na^+ и Cl^- в концентрации, соответствующей физиологическому раствору (0.9% по массе, т.е. 1720 ионов каждого знака). Ионы были равномерно распределены по всей системе, как внутри, так и снаружи капсида.

Перед началом расчета MD был использован процесс минимизации энергии, после чего расчет MD был выполнен с ограничениями положения капсида при T $=$ 200K и при 300K. После 1 нс процесса уравновешивания было проведено MD-моделирование без позиционных ограничений и при T = 300 K в течение 10 нс.

Для моделирования был использован термостат v-rescale и схему отсечки Верле. Расстояние отсечки для списка ближних соседей, электростатики Эвальда и взаимодействия Леннарда-Джонса составило 1.0 нм для каждого. Все параметры моделирования для всех систем были одинаковыми.

Результаты моделирования были проанализированы с помощью среднеквадратичного отклонения (*RMSD*) и радиуса инерции (R_g). *RMSD* и R_g были рассчитаны в GROMACS как функция времени

$$RMSD = \sqrt{\frac{1}{M}\sum_{i=1}^{N} m_i \left\| r_{i0} - r_i(t) \right\|^2}$$

$R_g = \sqrt{\frac{\sum_i \|r_i(t)-r_c\|^2 m_i}{\sum_i m_i}}$, где M — масса капсида, r_i(t) – положение атома i в момент времени t, r_{i0}-положение атома i в кристаллической структуре, m_i $-$ $-$ масса атома i и r_c - центр масс капсида.

Только для атомов остова, имеющих в качестве референтной структуры их исходную структуру. Расчеты *RMSD*и R_g не включают хвостовую последовательность для всех структур и состоят из одинакового числа аминокислот.

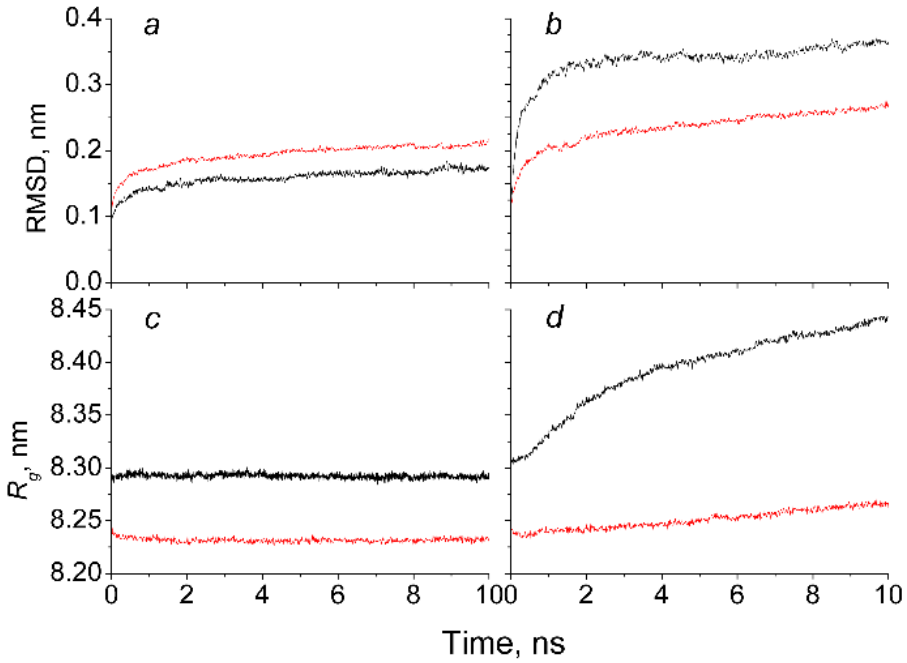


Рисунок Ж.1 — Изменение во времени среднеквадратичного отклонения, RMSD, (верхние графики) и радиуса инерции, R_g, (нижние графики) атомов основной цепи относительно их первоначальной структуры как эталонной структуры в силовом поле AMBER; расчеты RMSD и Rg не включали хвостовую последовательность для всех структур и состояли из одинакового числа аминокислот красный цвет обозначает капсид с хвостами, черный — капсид без хвостов; a,c – капсиды с нейтрализующим числом ионов, b,d – капсид с малым числом ионов

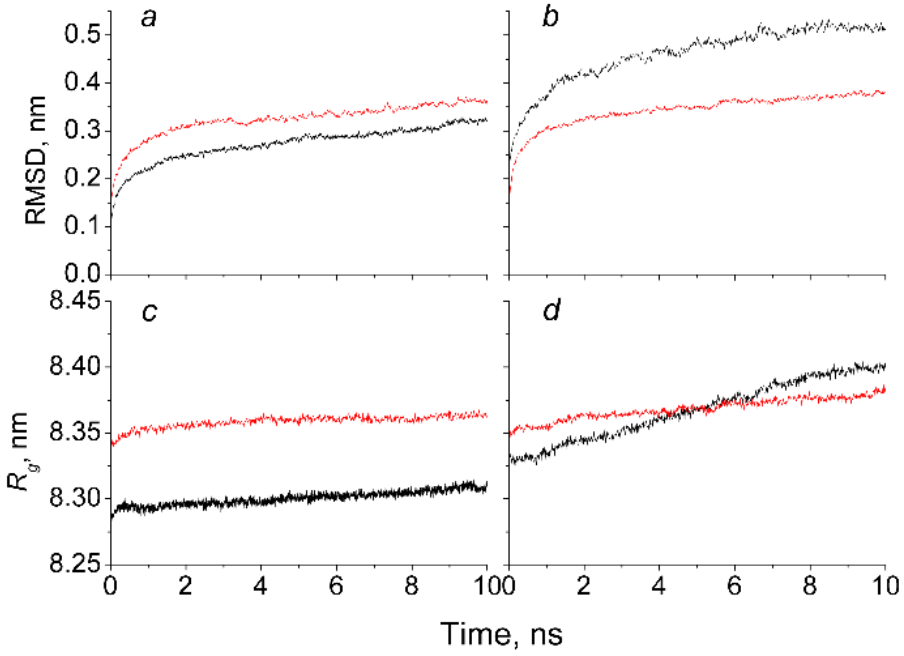


Рисунок Ж.2 — Изменение во времени среднеквадратичного отклонения, $RMSD$, (верхние графики) и радиуса инерции, R_g, (нижние графики) атомов основной цепи относительно их первоначальной структуры как эталонной структуры в силовом поле Gromos; расчеты RMSD и Rg не включали хвостовую последовательность для всех структур и состояли из одинакового числа аминокислот; красный цвет обозначает капсид с хвостами, черный — капсид без хвостов. a,c – капсиды с нейтрализующим числом ионов, b,d – капсид с малым числом ионов

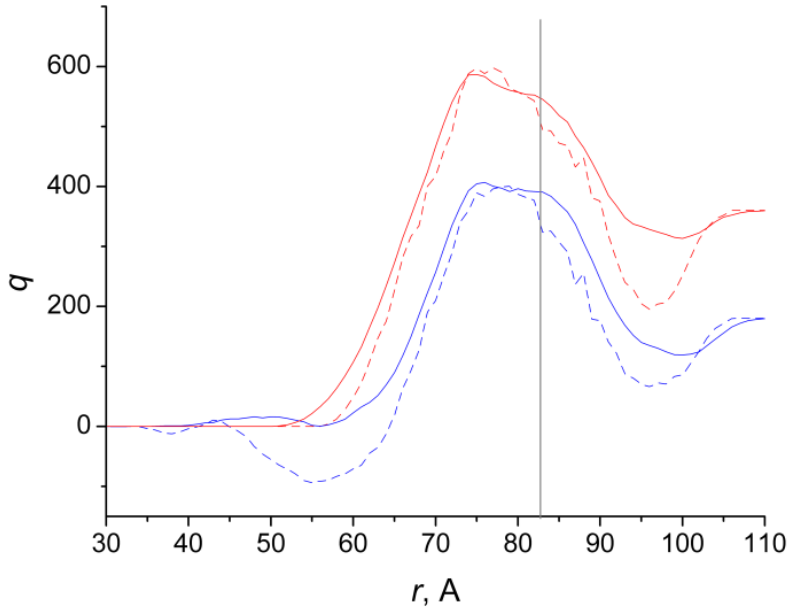


Рисунок Ж.3 — Зависимости интегрального заряда атомов капсида от расстояния от центра масс капсида; синие кривые соответствуют капсиду с хвостами, красные кривые соответствуют капсиду без хвостов; сплошные кривые соответствуют конечным конфигурациям, пунктирные — начальным; вертикальная линия соответствует середине стенки капсида

Приложение 3

Моделирование белков PCV2 про постоянном pH

3.1 Сходимость результатов

Рис. 3.1 — 3.3 показывают конформационный ландшафт, отобранный по капсидному белку с использованием двумерного распределения параметров SASA против Rg.

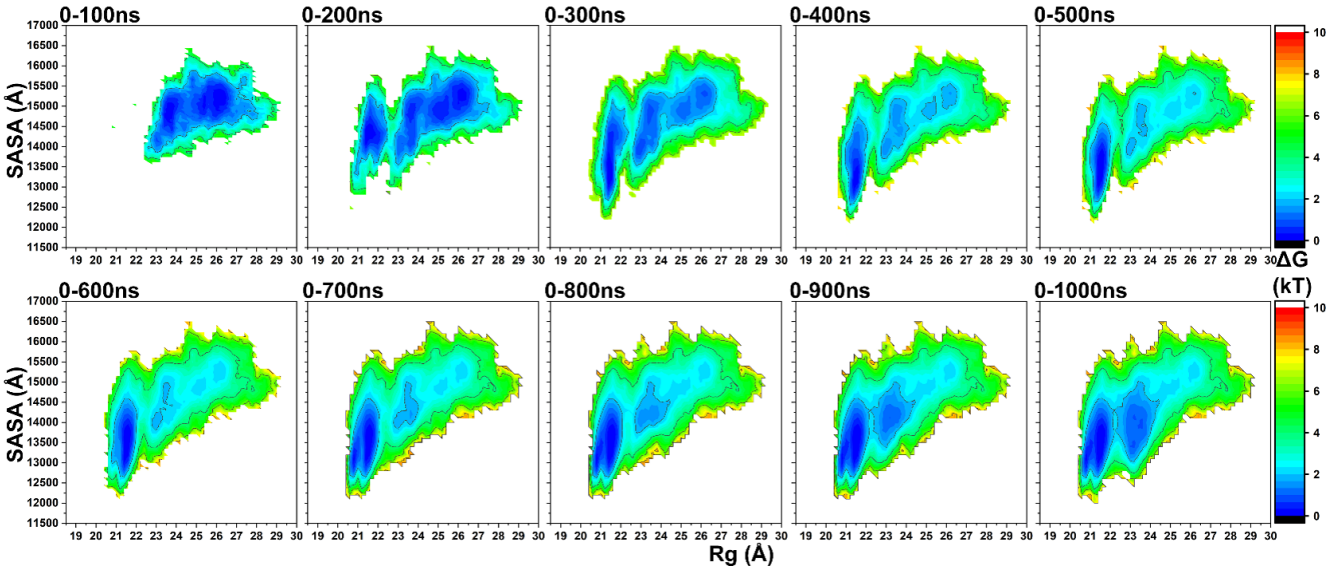


Рисунок 3.1 — Конформационный ландшафт, полученный с помощью CP* во время моделирования при pH 3; поверхности свободной энергии как функция параметров SASA всех остатков в зависимости от Rg всех остатков из 1000 нс моделирования; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231

Рис. 3.4 — 3.6 показывают конформационный ландшафт, отобранный по капсидному белку с использованием двумерного распределения параметров RMSD против Rg.

3.2 Сходимость значений pKa

Рис. 3.7 — 3.13 показывает конформационный ландшафт, выбранный по капсидному белку с использованием одномерного распределения параметров SASA, RMSD и Rg.

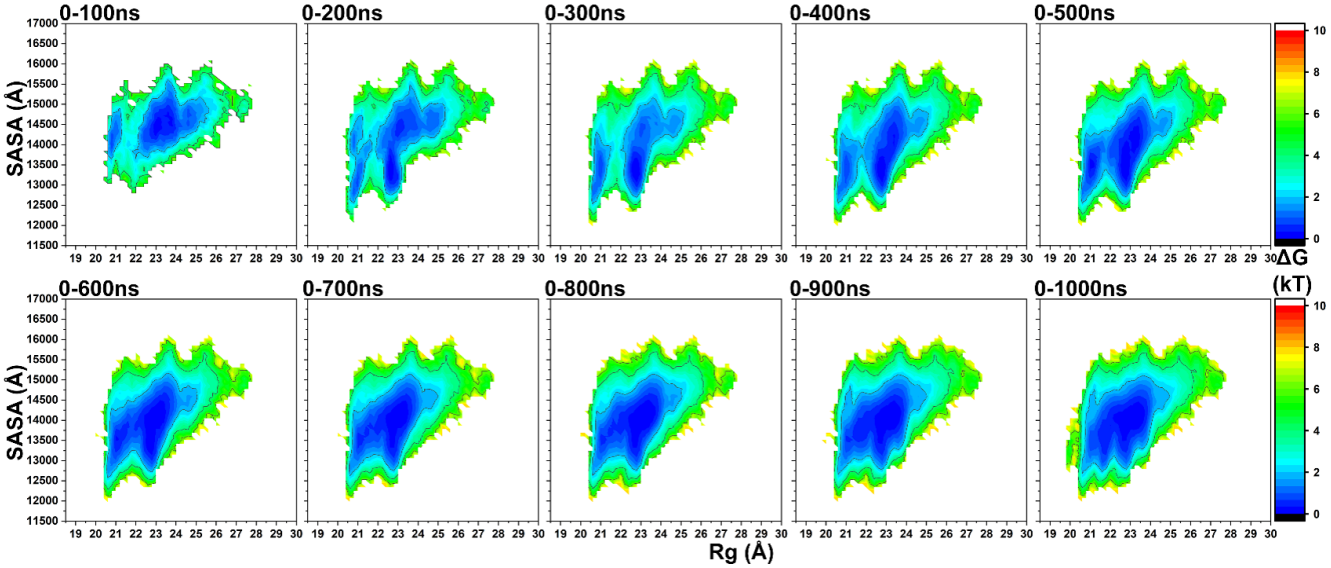


Рисунок 3.2 — Конформационный ландшафт, полученный с помощью CP* во время моделирования при pH 5; поверхности свободной энергии как функция параметров SASA всех остатков в зависимости от Rg всех остатков из 1000 нс моделирования; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231

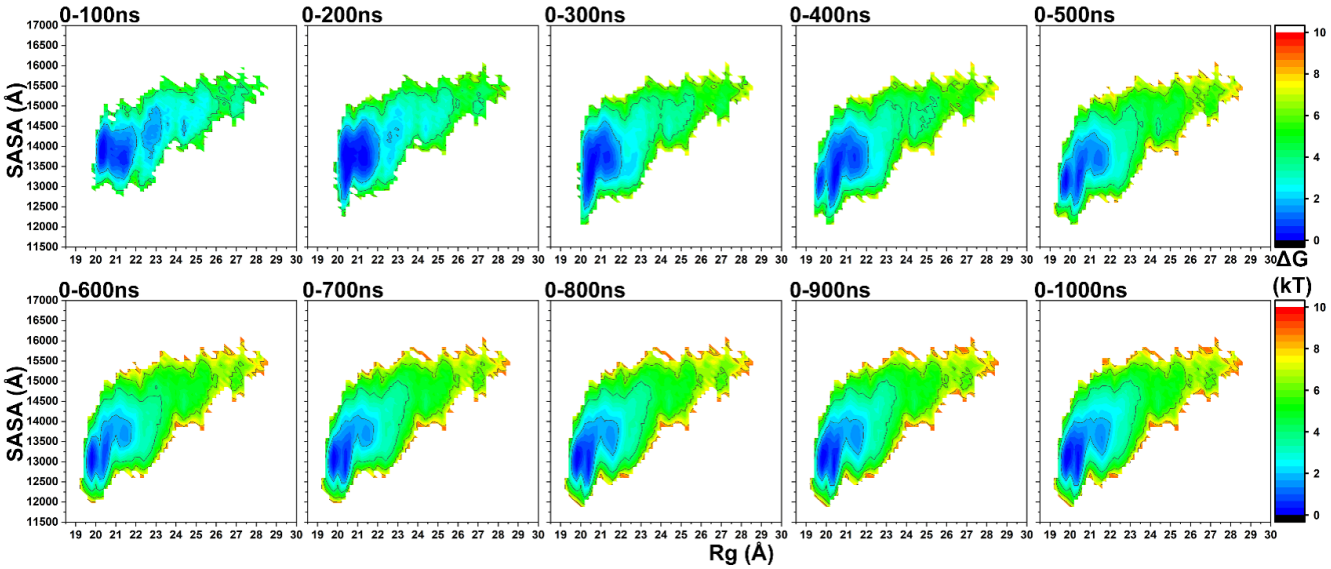


Рисунок 3.3 — Конформационный ландшафт, полученный с помощью белка CP* во время моделирования при pH 7; поверхности свободной энергии как функция параметров SASA всех остатков в зависимости от Rg всех остатков из 1000 нс моделирования; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231

3.3 Репрезентативные структуры

Кластерный анализ показывает, что CP более конформационно изменчив при pH 5, как показано на рис. 3.14.

При pH 3 первый кластер описывает 75% конформаций, второй кластер описывает 10% конформаций, а третий кластер описывает 4% конформаций. При pH 5 первый кластер описывает

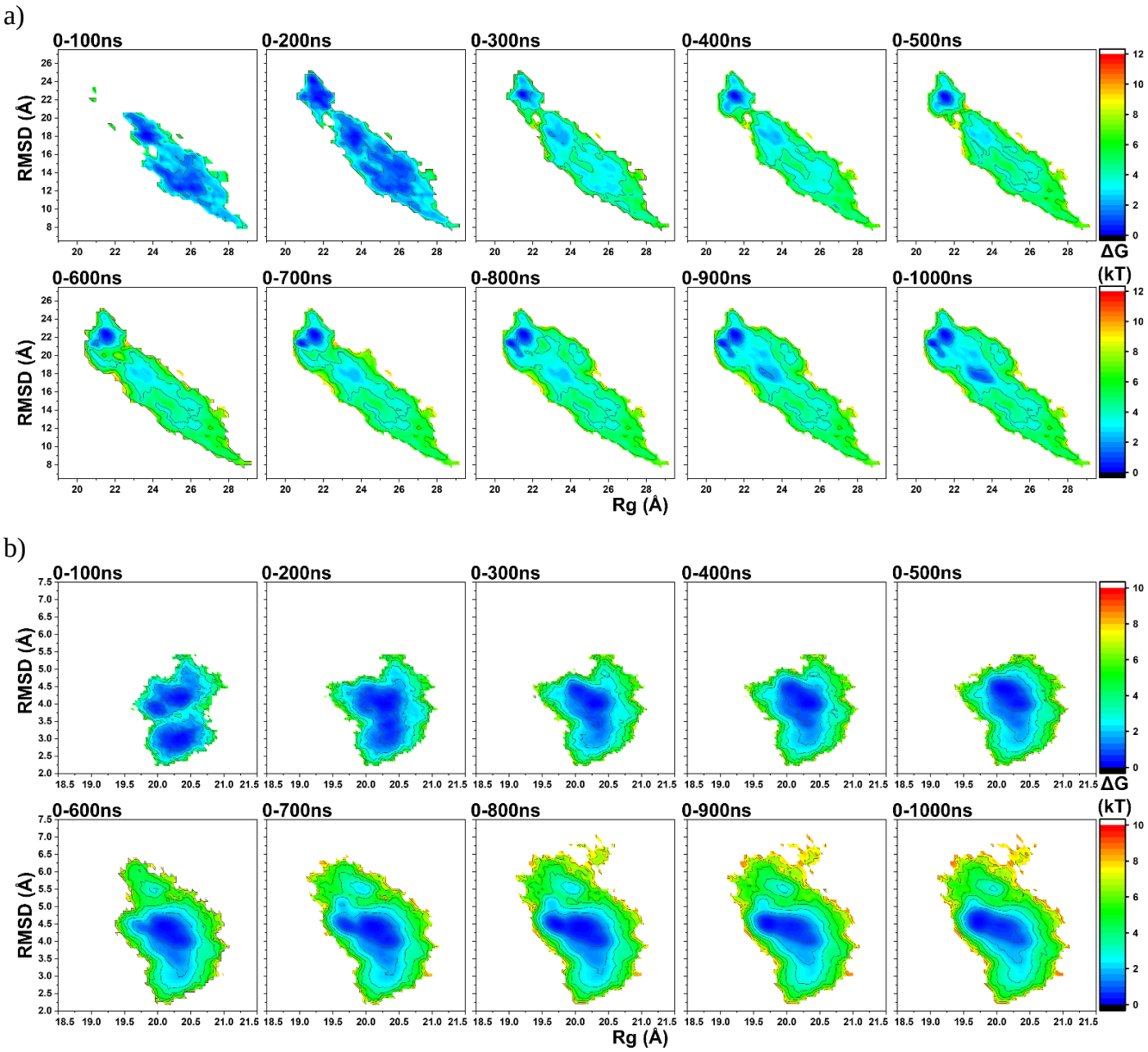


Рисунок 3.4 — Конформационный ландшафт, отобранный CP* в течение 1000 нс моделирования при pH 3; поверхности свободной энергии как функции параметров а) RMSD всех атомов Cα по сравнению с Rg всех остатков; б) RMSD 38-231 атомов Cα по сравнению с Rg 38-231 остатков из моделирования; среднеквадратическое отклонение рассчитывалось относительно эталонной рентгеновской структуры

14% конформаций, второй и третий кластеры описывают чуть менее 8% конформаций каждый. Для первого минимума при pH 7 первый кластер описывает 59% конформаций, второй кластер описывает 6% конформаций, а третий кластер описывает 3% конформаций. Преобладающее большинство конформаций при pH 3 и 7 описывается первым кластером, что позволяет предположить, что CP принимает одну конформацию в пределах каждого из этих значений pH. Однако сопоставимое распределение конформаций в первых трех кластерах при pH 5 позволяет предположить, что CP может быть более конформационно изменчивым при этом pH.

Сравнение репрезентативных структур при различных значениях pH показано на рис. 3.15 и 3.16

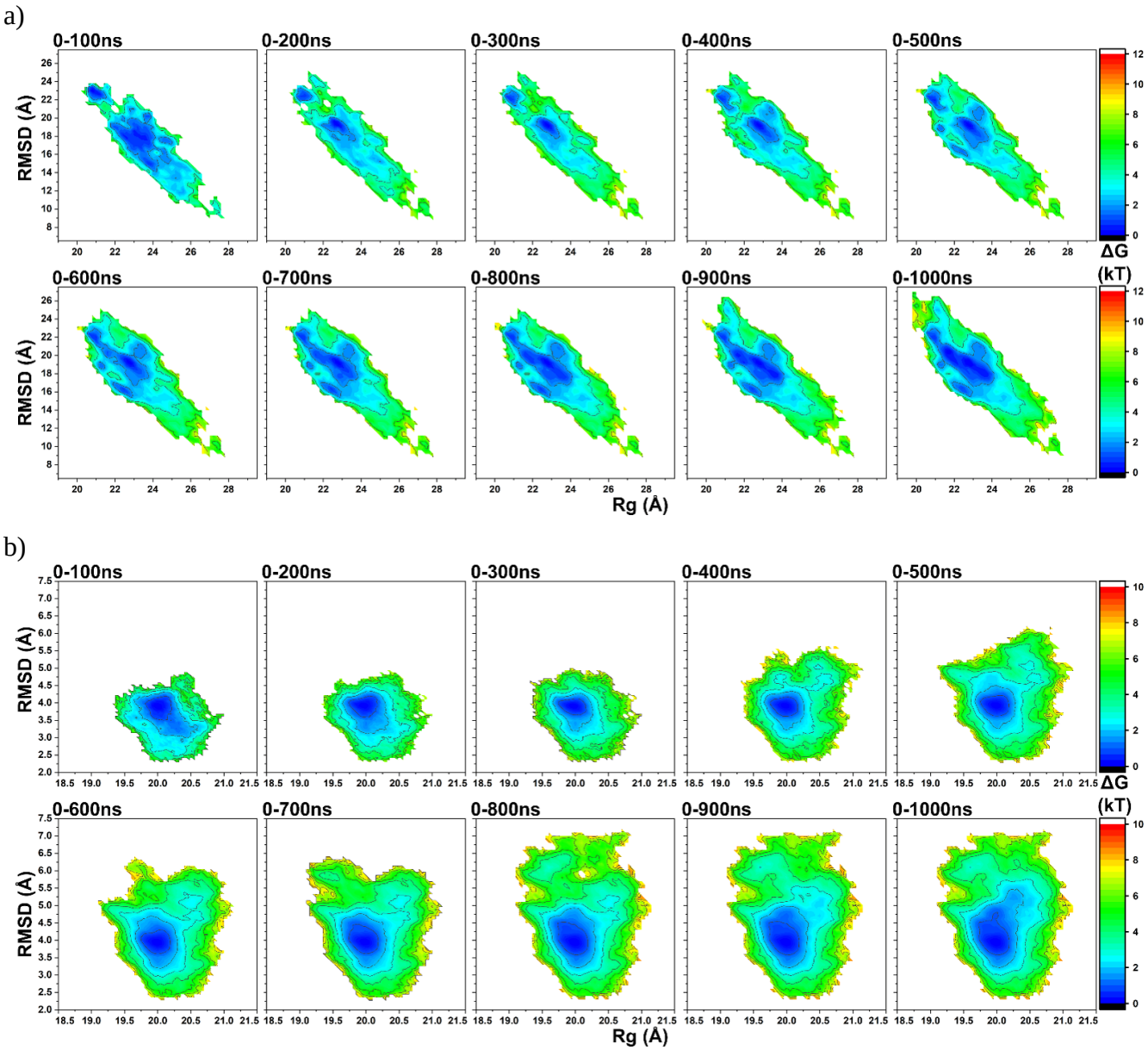


Рисунок 3.5 — Конформационный ландшафт, отобранный CP* в течение 1000 нс моделирования при pH 5; поверхности свободной энергии как функции параметров а) RMSD всех атомов Cα по сравнению с Rg всех остатков; б) RMSD 38-231 атомов Cα по сравнению с Rg 38-231 остатков; среднеквадратическое отклонение рассчитывалось относительно эталонной рентгеновской структуры

Конформационная изменчивость показана на рис. 3.17.

Для визуализации конформационной изменчивости внутри каждого кластера было создано составное изображение всех членов кластера, наложив каждый на центроид. Несмотря на то, что эти изображения являются качественными, они передают как величину, так и направление изменчивости. Это имеет преимущество перед традиционными количественными методами, такими как метод среднеквадратичного отклонения, которые фиксируют величину изменения, но игнорируют направление изменения. Как и ожидалось, N-концы демонстрируют конформационную изменчивость; однако N-концы каждого кластера располагаются в сопоставимых положениях по отношению к остальной части CP. Это неожиданно, поскольку N-концы не были включены в

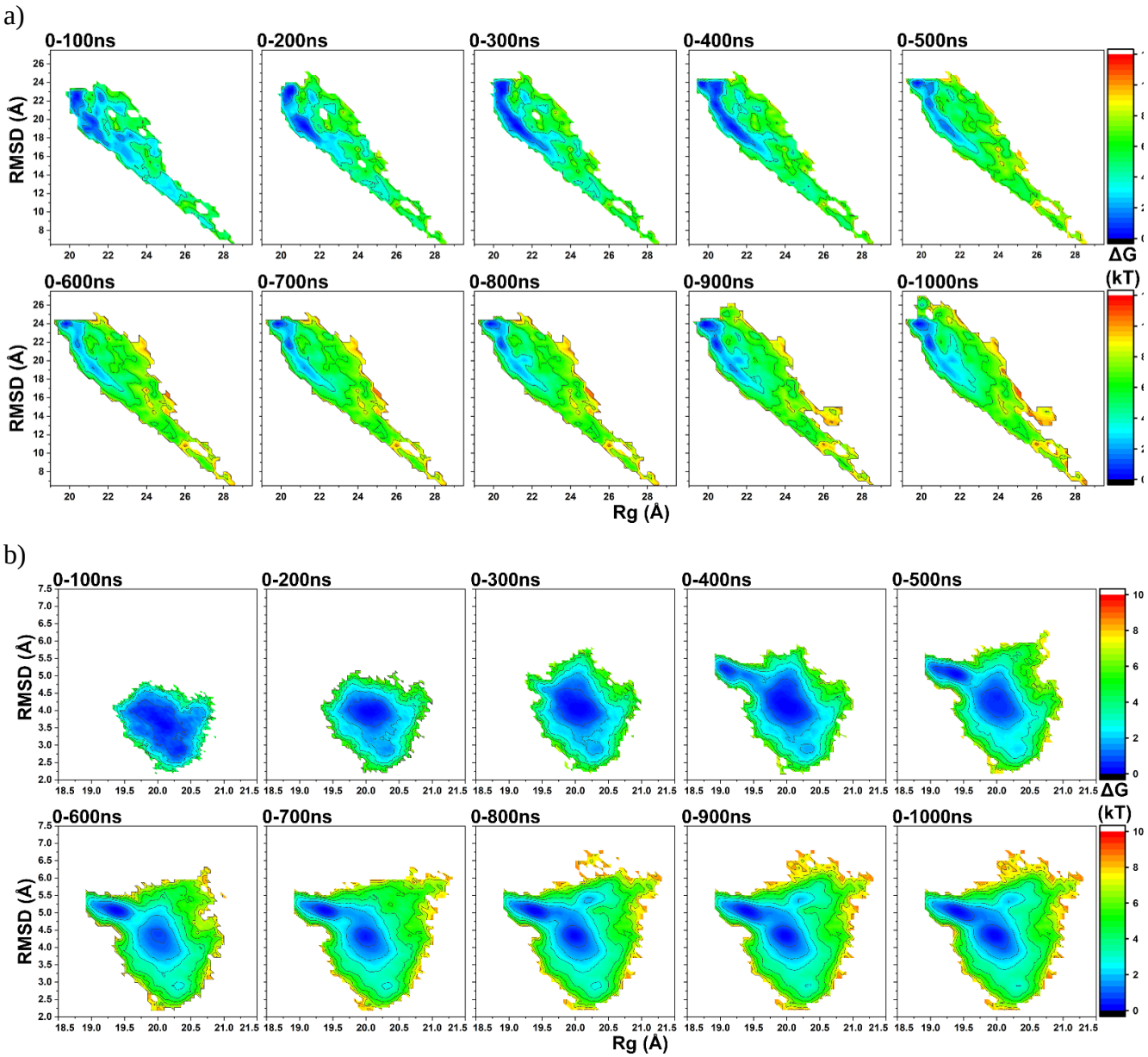


Рисунок 3.6 — Конформационный ландшафт, отобранный CP* в течение 1000 нс моделирования при pH 7; поверхности свободной энергии как функции параметров а) RMSD всех атомов Cα по сравнению с Rg всех остатков; б) RMSD 38-231 атомов Cα по сравнению с Rg 38-231 остатков из моделирования; среднеквадратическое отклонение рассчитывалось относительно эталонной рентгеновской структуры

кластерный анализ. Первый кластер pH 3 содержит 76537 кадров. N-конец занимает большое пространство, противопоставленное единственной стороне CP. Первый кластер pH 5 содержит 5724 кадра. N-конец занимает меньшее пространство по сравнению с pH 3, а его положение и ориентация соответствуют остальной части CP. Первый кластер pH 7 содержит 19513 кадра. N-конец занимает большее пространство; однако эта оболочка имеет ту же общую форму и занимает сопоставимые положения внутри кластера — вдоль внутренней стороны CP. Первый кластер pH 7 (минимум 2) содержит 25076 структур. Конформационная изменчивость N-конца ограничена и локализуется на поверхности CP, которая определяет внутреннюю поверхность капсида.

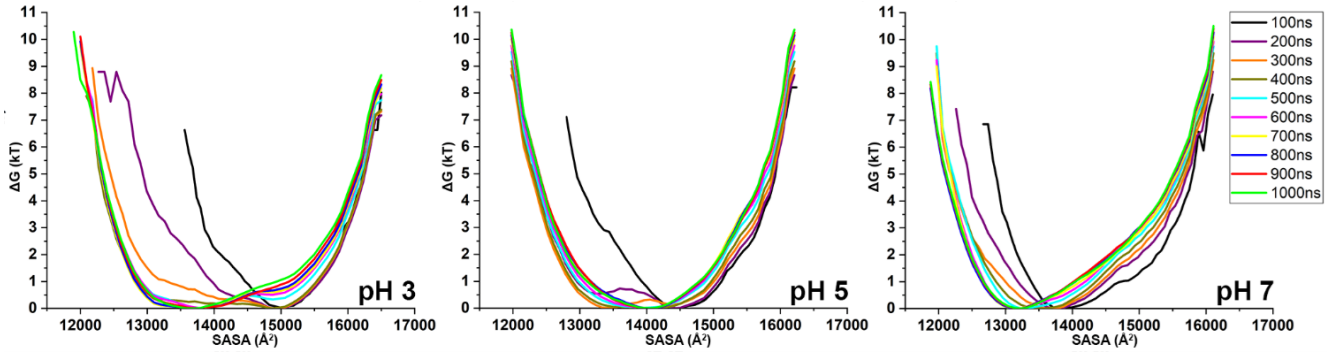


Рисунок 3.7 — Диаграмма свободной энергии как функции параметра SASA всех остатков при pH 3, 5 и 7; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231

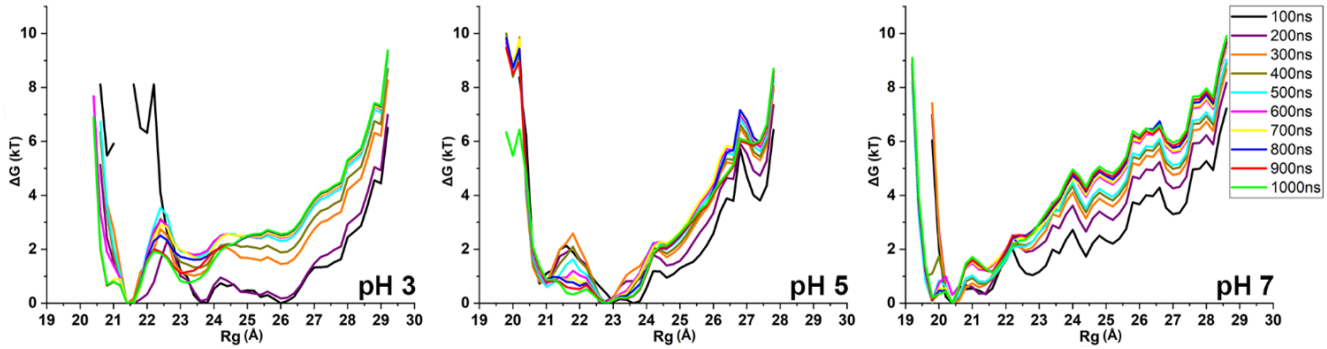


Рисунок 3.8 — Диаграмма свободной энергии как функции параметра Rg всех остатков при pH 3, 5 и 7; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231

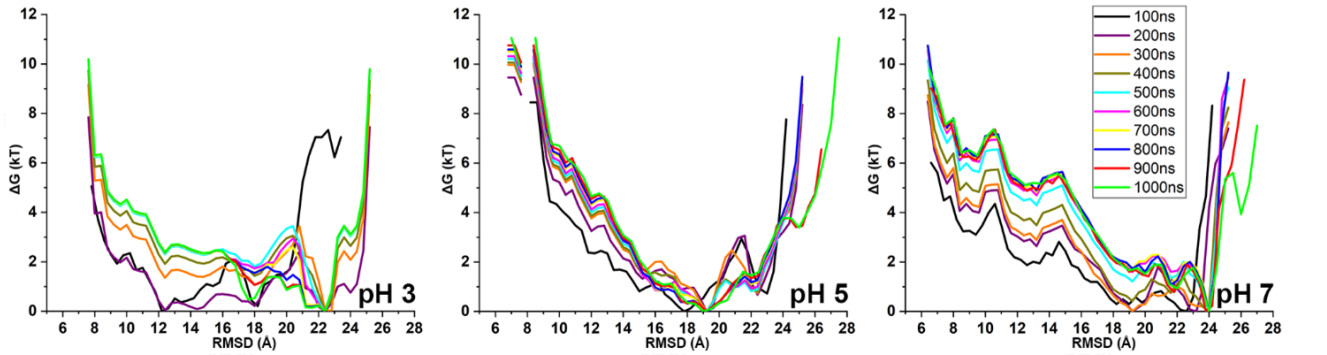


Рисунок 3.9 — Диаграмма свободной энергии как функция параметра RMSD всех остатков при pH 3, 5 и 7; анализ был проведен для всех остатков в субъединице, включая гибкую N-концевую область - аминокислоты 1-231

Сравнение центроидов первых трех кластеров при каждом значении pH показано на рис. 3.18 — 3.22.

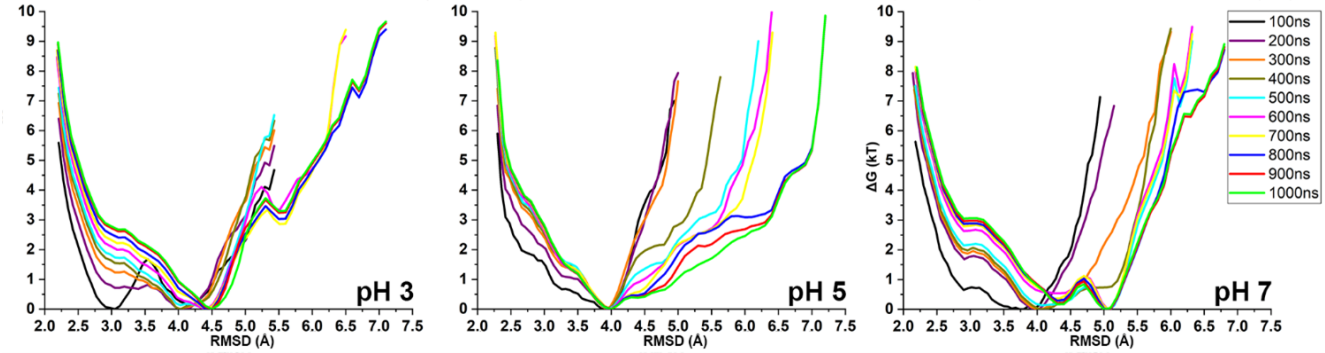


Рисунок 3.10 — Диаграмма свободной энергии как функция параметра RMSD 38-231 атомов Cα при pH 3, 5 и 7; среднеквадратичное отклонение рассчитывалось относительно эталонной рентгеновской структуры

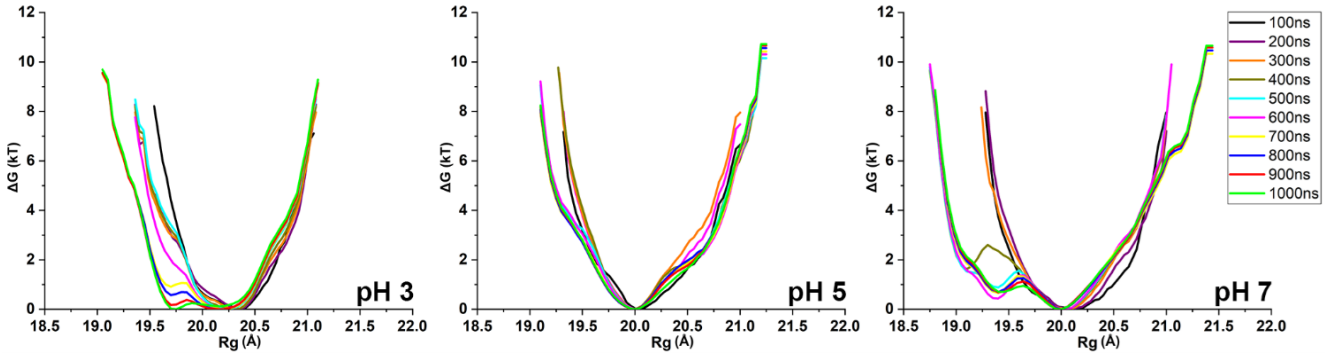


Рисунок 3.11 — Диаграмма свободной энергии как функции параметра Rg остатков 38-231 при pH 3, 5 и 7

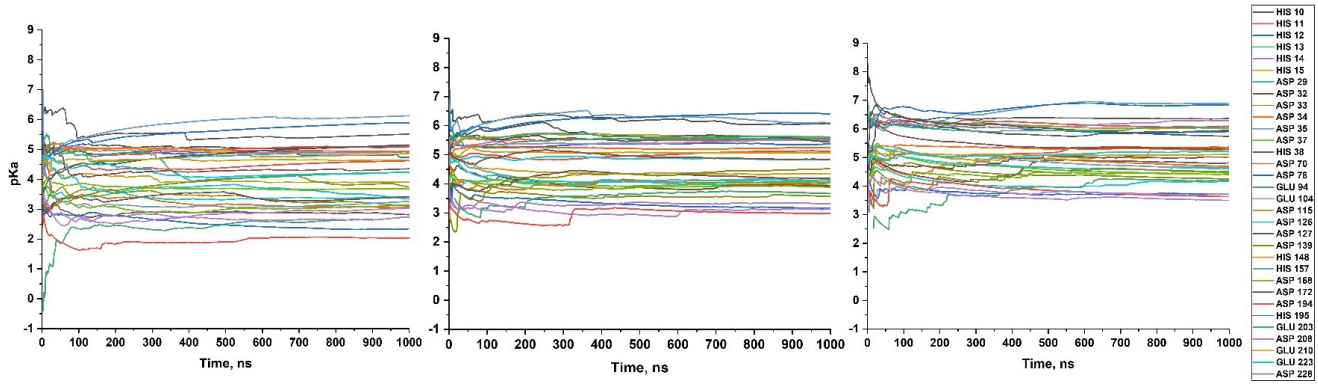


Рисунок 3.12 — Совокупно рассчитанные значения pKa для всех остатков Asp/Glu/His из симуляций при pH 3 (a), pH 5 (b), pH 7 (c) как функция времени симуляции

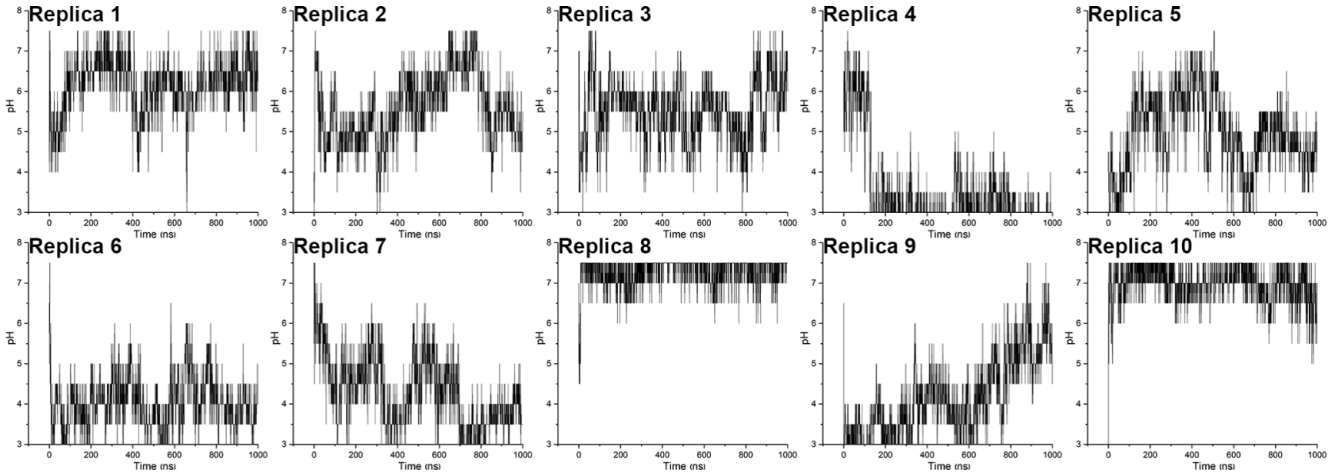


Рисунок 3.13 — Обмен участками для каждой реплики; зависимость обменов между значениями pH от времени моделирования для отдельных реплик; коэффициент принятия составляет 21%

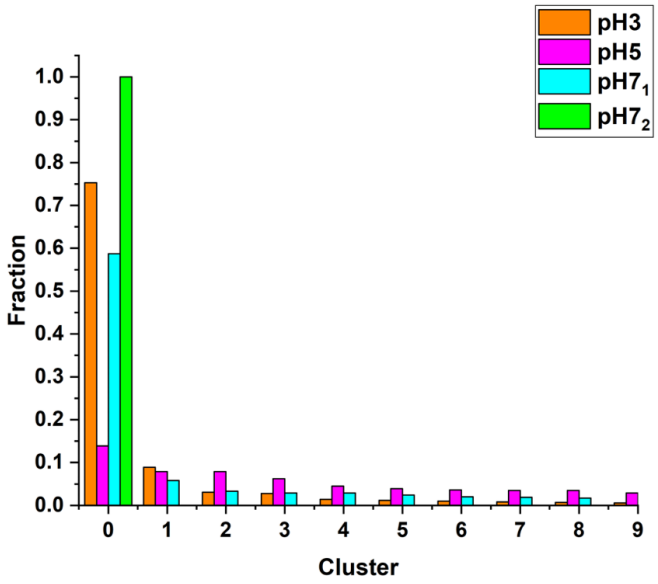


Рисунок 3.14 — Гистограмма, показывающая долю извлеченных структур на кластер для 10 лучших кластеров; количество кластеров, начиная с 0

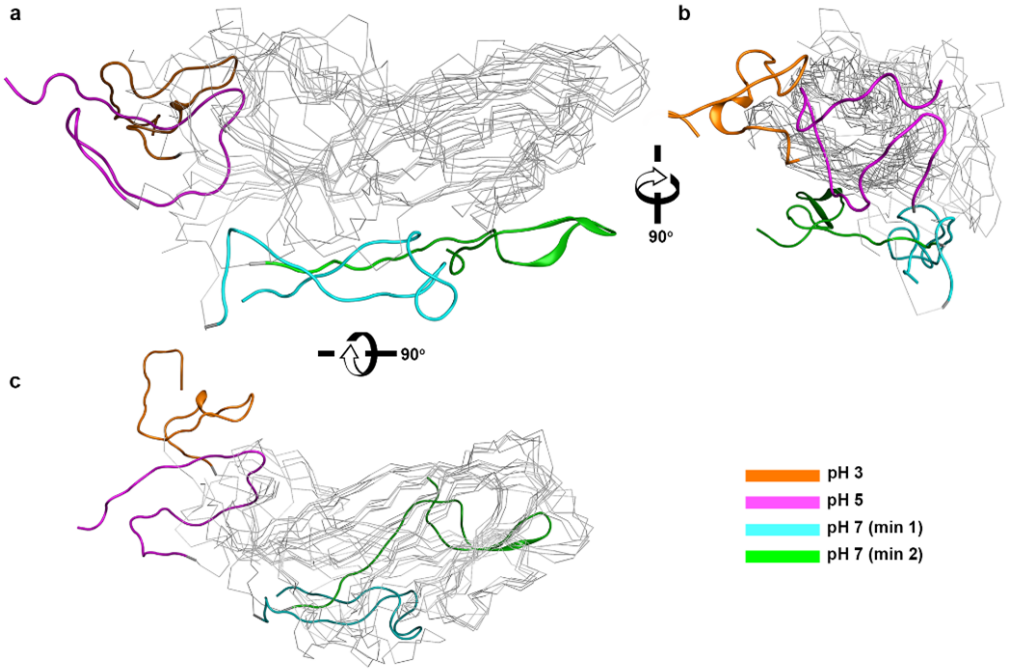


Рисунок 3.15 — Выравнивание репрезентативных структур из наиболее заселенных кластеров при pH 3, 5, 7 на кристаллической структуре: N-концы окрашены, а остальная часть структуры — серого цвета

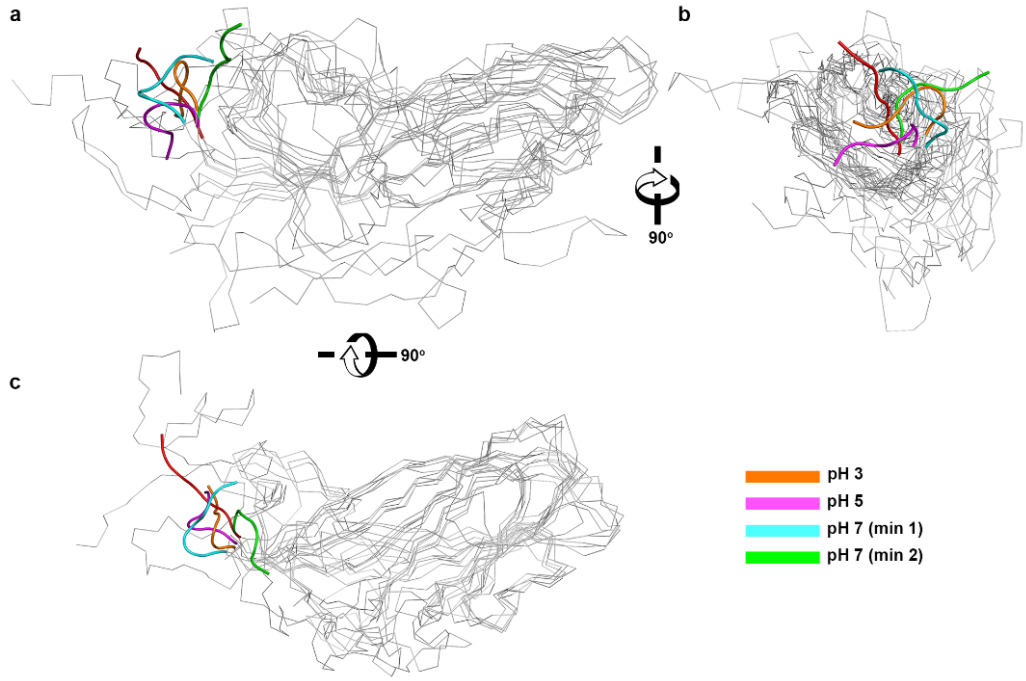


Рисунок 3.16 — Выравнивание репрезентативных структур из наиболее заселенных кластеров при pH 3, 5, 7 на кристаллической структуре: C-концы окрашены, а остальная часть структуры — серого цвета

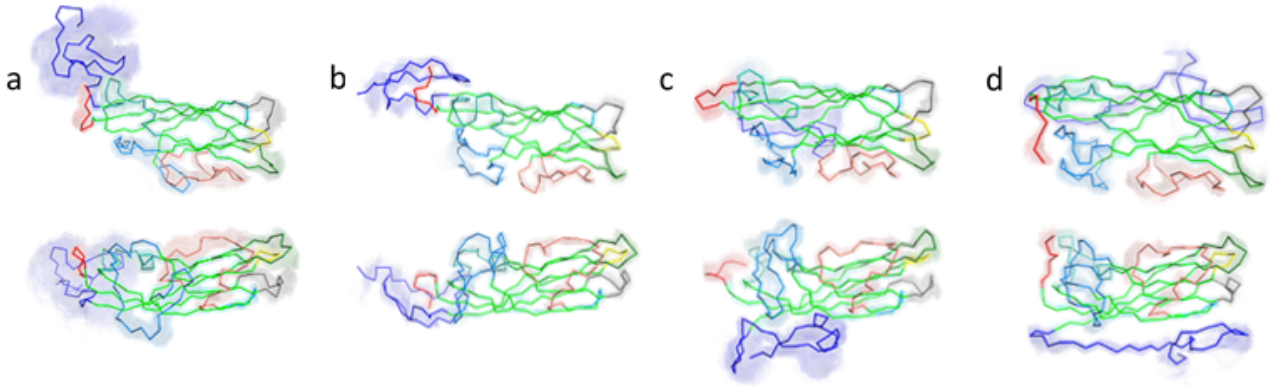
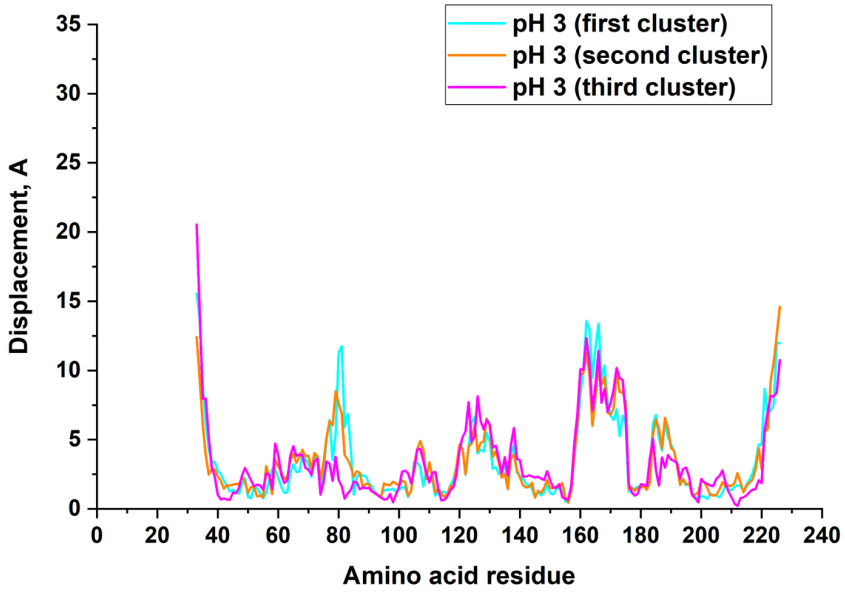


Рисунок 3.17 — Конформационная изменчивость первых кластеров при pH 3 (a), pH 5 (b), pH 71 (c) и pH 72 (d); кластеризация проводилась с использованием аминокислот 38–231; однако изображения создаются с использованием всех аминокислот; на этих изображениях показана оболочка (тонкие линии) всех кадров в кластере; тонкие огибающие демонстрируют малые среднеквадратичные отклонения между кадрами, тогда как широкие огибающие демонстрируют большие среднеквадратичные отклонения

a)



b)

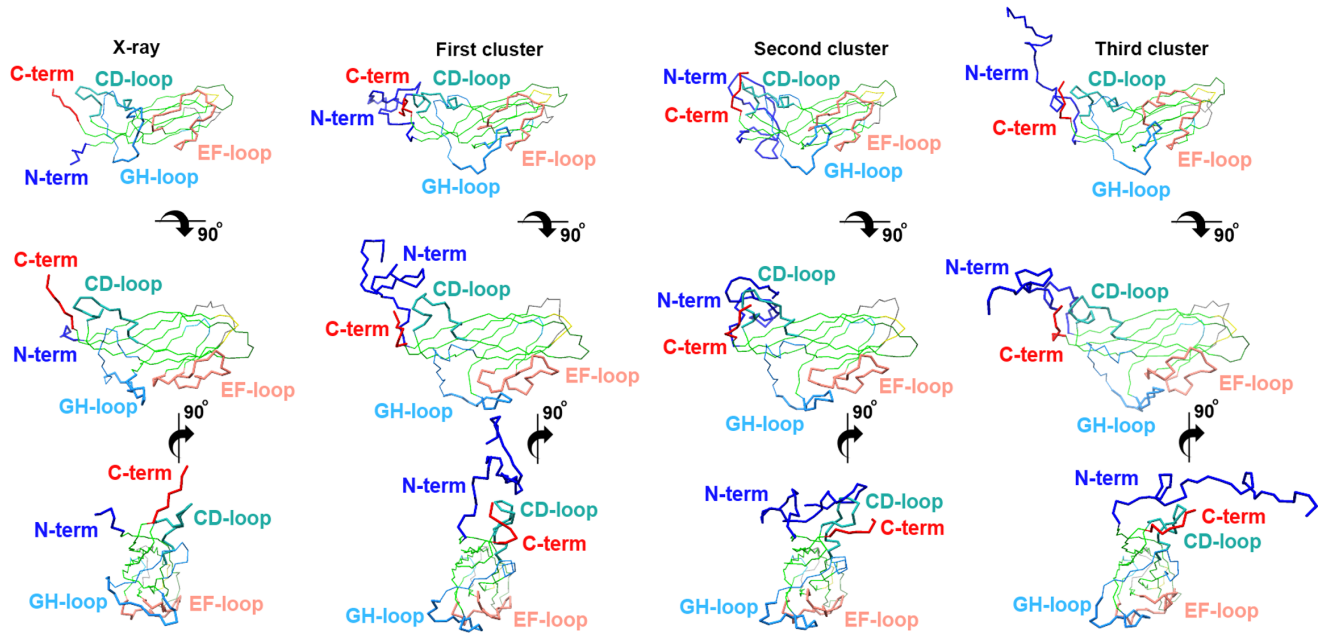
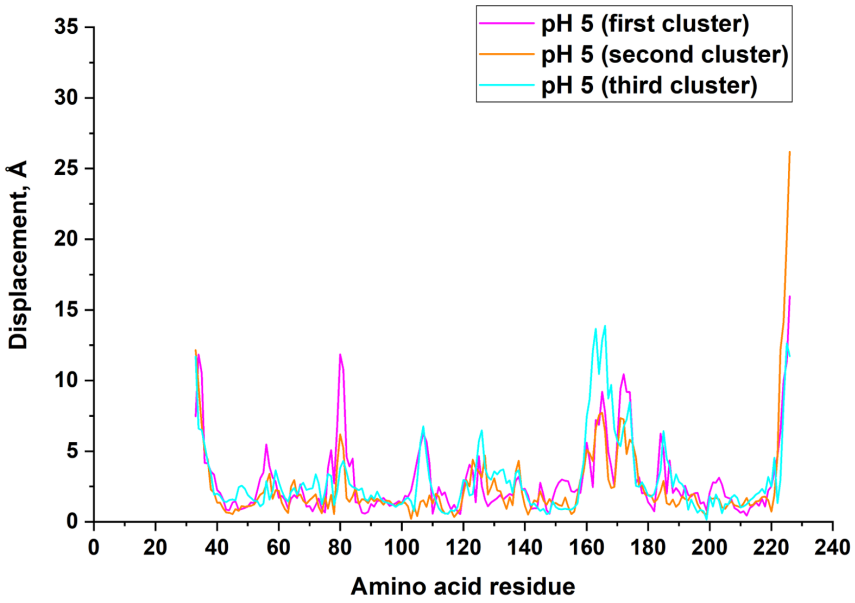


Рисунок 3.18 — pH-зависимые области CP; (a) позиционное смещение репрезентативных структур из первых трех заселенных кластеров при pH 3 по сравнению с рентгеновской структурой; (b) C_α след репрезентативных структур из первых трех кластеров при pH 3

a)



b)

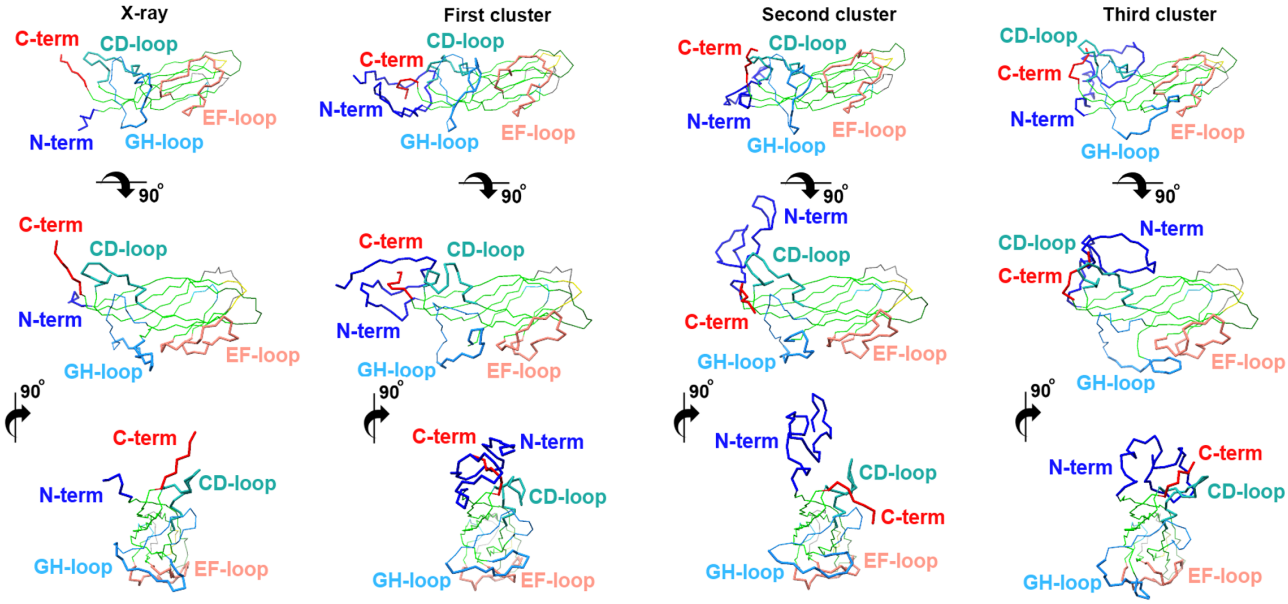
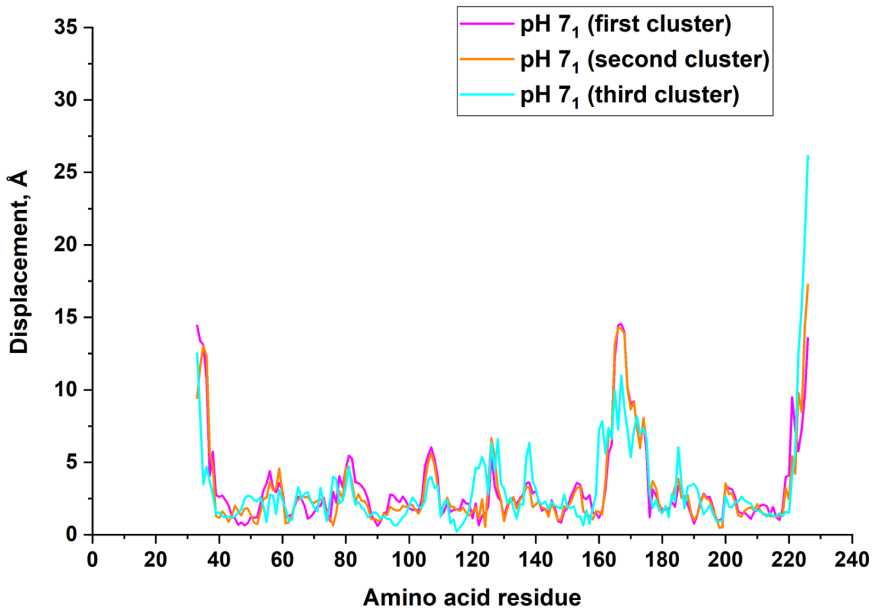


Рисунок 3.19 — pH-зависимые области CP; (a) позиционное смещение репрезентативных структур из первых трех заселенных кластеров при pH 5 по сравнению с рентгеновской структурой; (b) C_α след репрезентативных структур из первых трех кластеров при pH 5

a)



b)

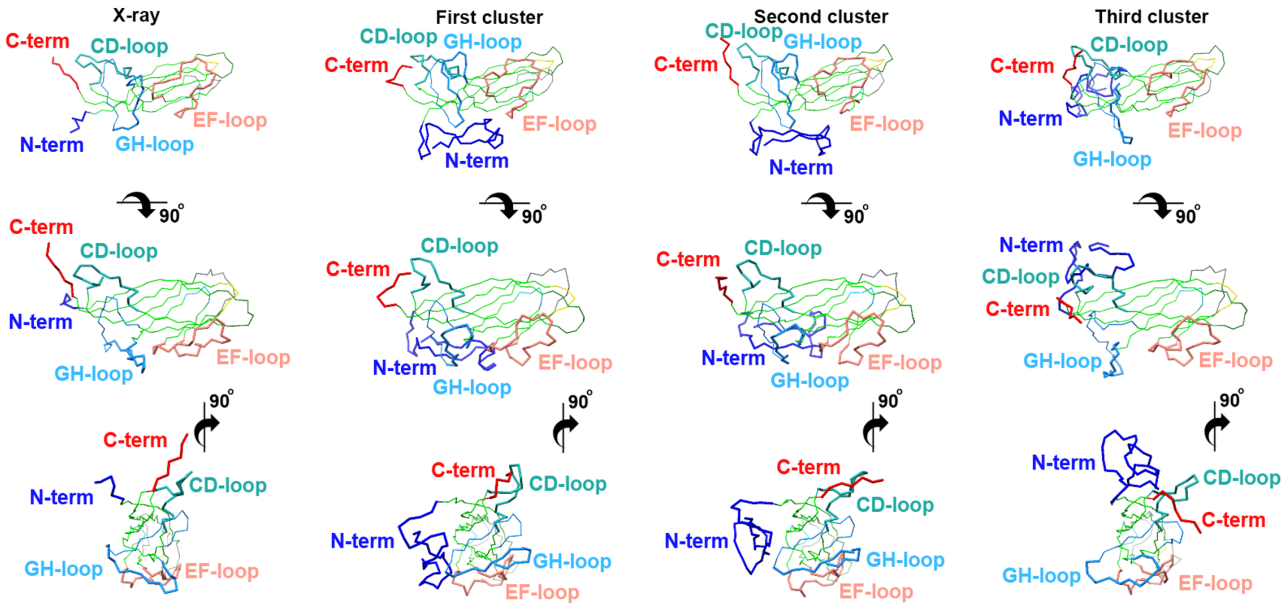
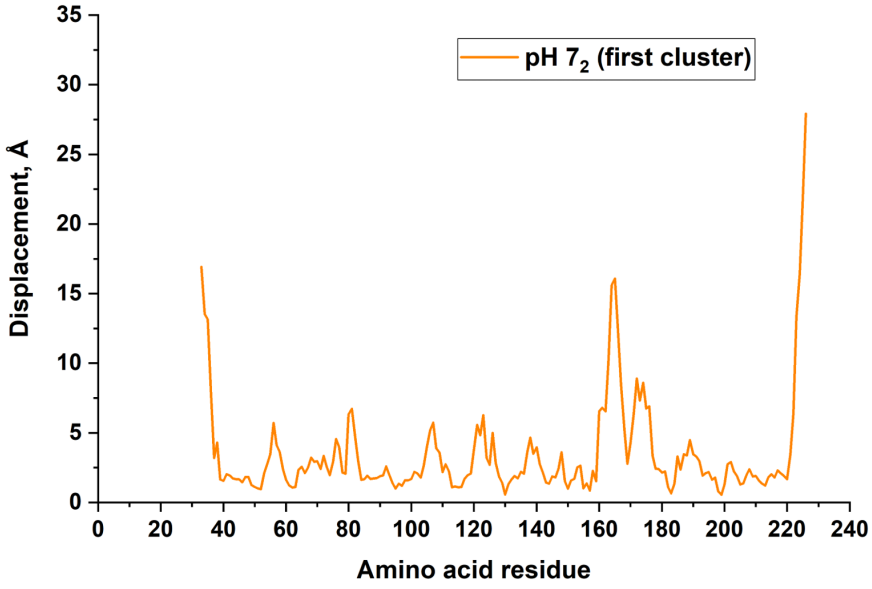


Рисунок 3.20 — pH-зависимые области CP; (a) позиционное смещение репрезентативных структур из первых трех заселенных кластеров при pH 7$_1$ по сравнению с рентгеновской структурой; (b) C$_\alpha$ след репрезентативных структур из первых трех кластеров при pH 7$_1$

a)



b)

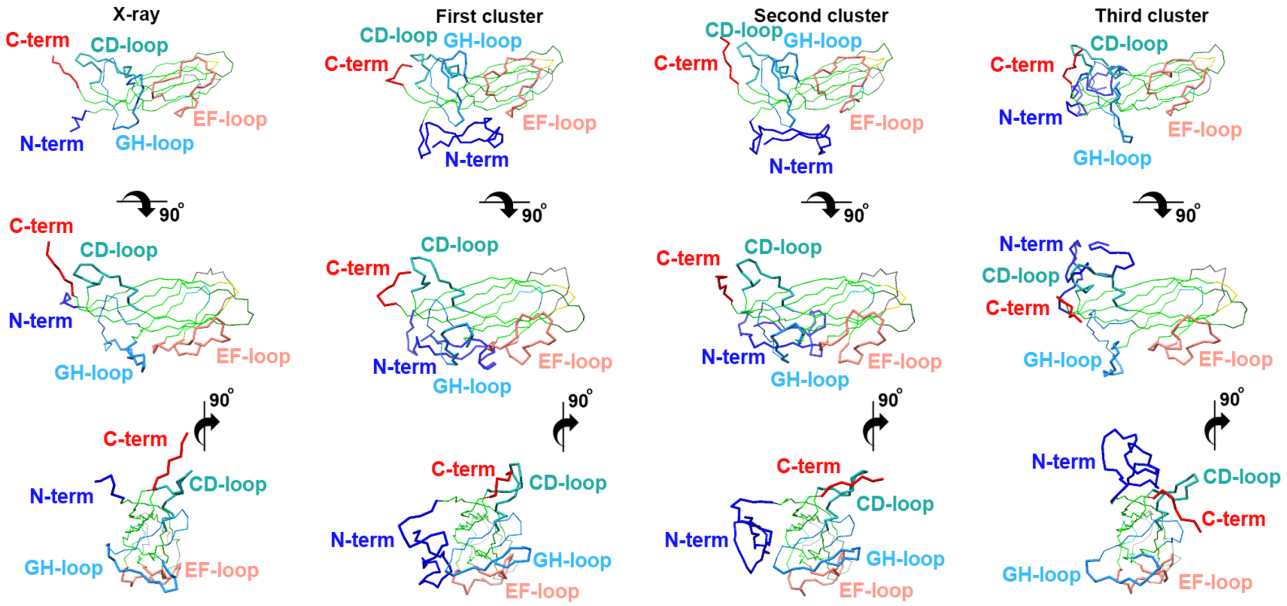


Рисунок 3.21 — pH-зависимые области CP; (a) позиционное смещение репрезентативных структур из первых трех заселенных кластеров при pH 7_2 по сравнению с рентгеновской структурой; (b) C_α след репрезентативных структур из первых трех кластеров при pH 7_2

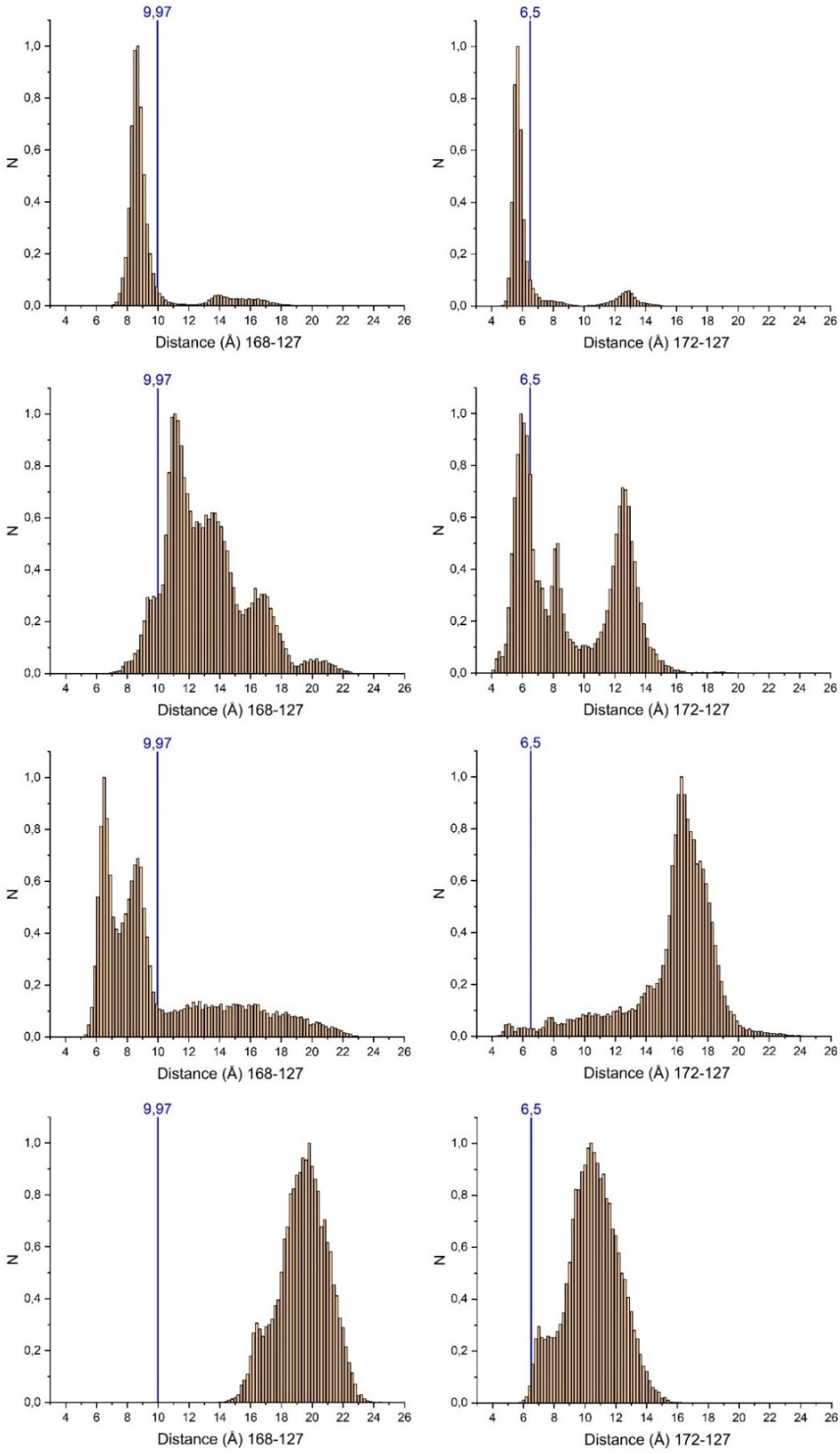


Рисунок 3.22 — Соотношение выделенных структур распределеных в соответствии с расстояниями между остатками 168 и 127 a,b,c,d и остатками 172 и 127 e,f,g,h; при pH 3 – a,e; pH5 – b,f; pH 7 (первый минимум) – c,g; pH7 (2 минимум) – d,h; синяя вертикальная линия представляет расстояния между остатками в структуре капсида

Приложение И

Константа скорости диссоциации белок-лиганд, полученная с помощью полноатомного моделирования

И.1 Подробности моделирования молекулярной динамики

Моделирование молекулярной динамики проводилось с использованием пакета GROMACS 2020.1 с реализованным методом τ-RAMD [366]. Было выбрано силовое поле [393] CHARMM36 для белков и SwissParam [394] был использован для создания файла топологии для лиганда. Система была помещена в додекаэдрическую коробку на расстоянии 1 нм от краев, сольватированную молекулами воды TIP3P. Ионы Na^+ и Cl^- добавляли в физиологической концентрации 150 mM. После этого энергия системы была минимизирована с использованием метода наискорейшего спуска, и комплекс был последовательно уравновешен в ансамблях NVT и NPT в течение 100 пс при 300 K и 1 бар с термостатом масштабирования скорости и баростатом Берендсена. Затем система была подвергнута 230 нс моделированию MD с термостатом масштабирования скорости и баростатом Парринелло-Рахмана с целью достижения стабильной начальной структуры (RMSD достигло значения 0.25 нм) для дальнейшего расчета τ-RAMD. Временной шаг был равен 2 fc.

И.2 Вероятность выживания

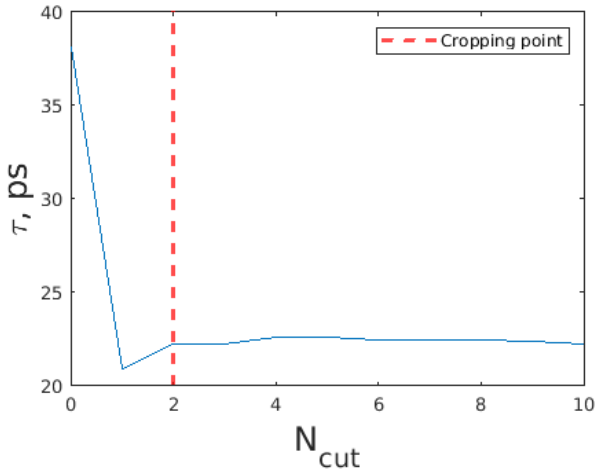
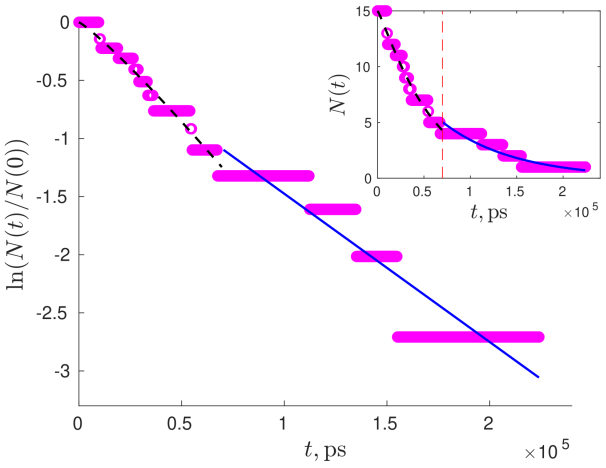
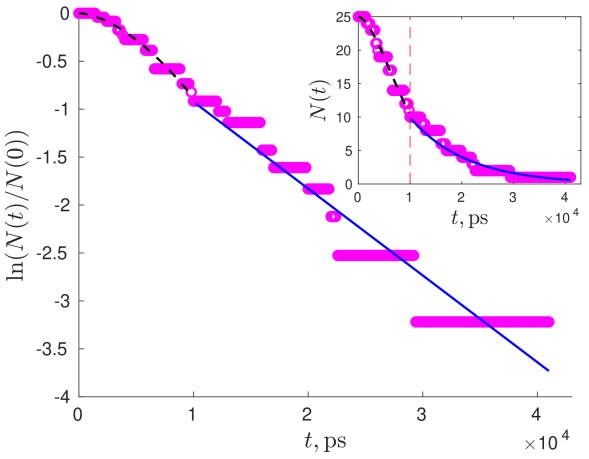


Рисунок И.1 Иллюстрация того, как наклон экспоненциального режима (τ^{-1}) в полулогарифмическом масштабе может меняться в зависимости от количества отсеченных с конца значений (N_{cut}) для силы $F = 550$ кДж/(нм · мол); последняя точка в данных, когда $N(t) = 1$, образует длинное горизонтальное плато, которое значительно увеличивает τ; после $N_{cut} = 2$, τ стабилизирует, а обрезка 2 точек обеспечивает надежность подгонки

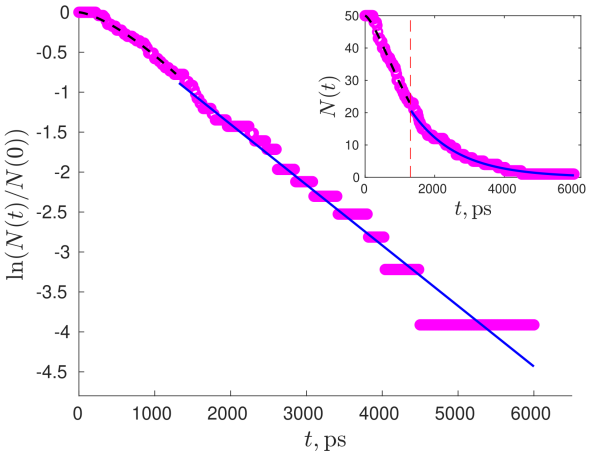
а) $F = 250$ kJ/(nm · mol)



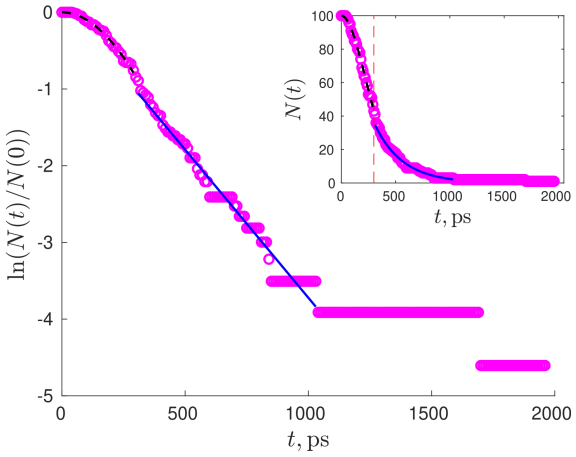
б) $F = 300$ kJ/(nm · mol)



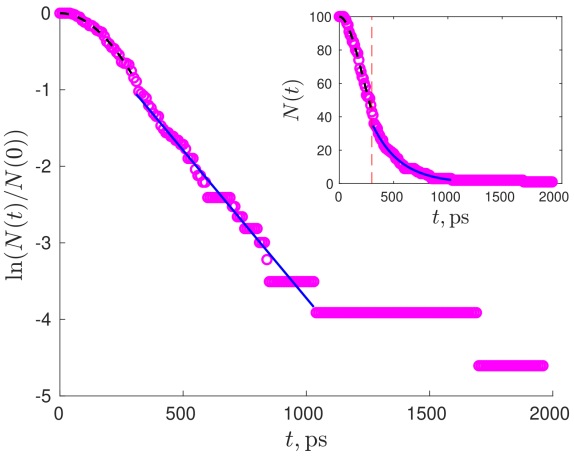
в) $F = 350$ kJ/(nm · mol)



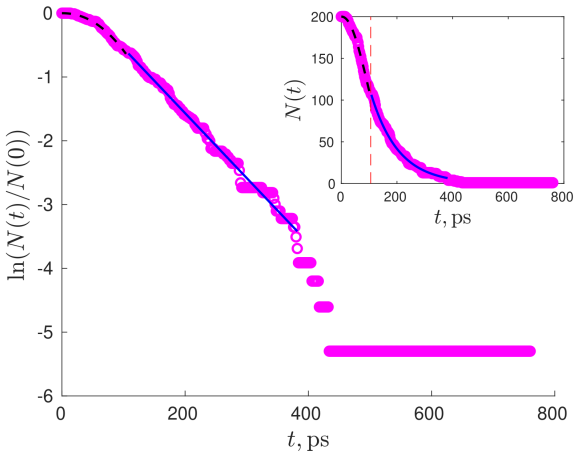
г) $F = 400$ kJ/(nm · mol)



д) $F = 450$ kJ/(nm · mol)



е) $F = 500$ kJ/(nm · mol)



ж) $F = 550$ kJ/(nm · mol)

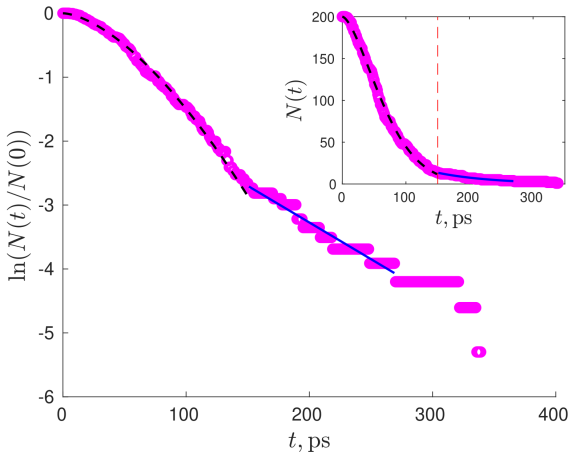


Рисунок И.2 Подгонка вероятностей выживания комплекса белок-лиганд с двумя режимами при различных внешних силах: черная пунктирная линия — неэкспоненциальный режим, синяя линия — экспоненциальный режим, черная штрихпунктирная линия — момент переключения между режимами

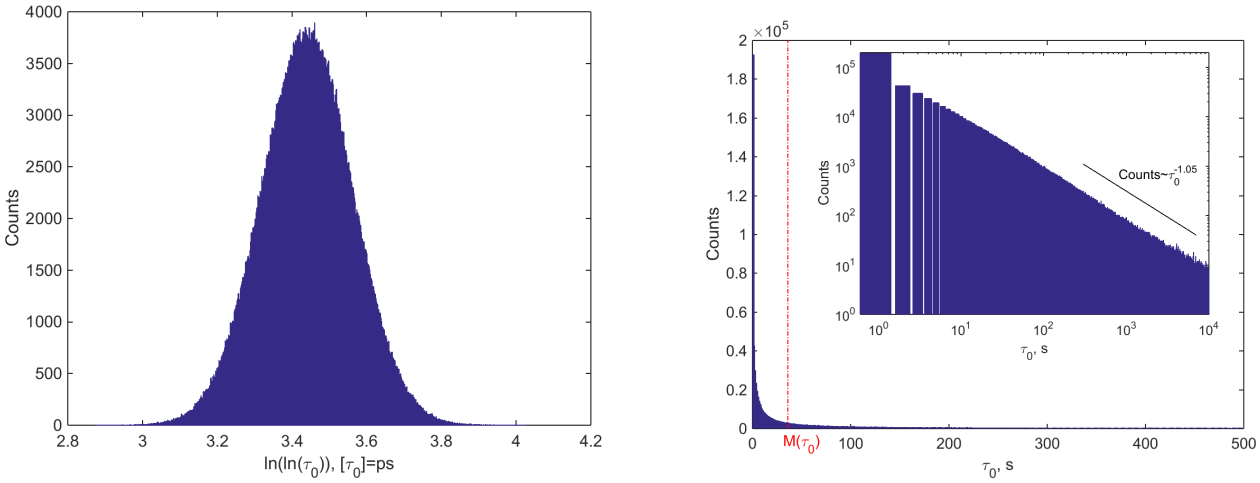


Рисунок И.3 Распределения подобранного параметра, смоделированные методом Монте-Карло $\ln(\ln(\tau_0))$ (слева) и трансформированный τ_0 (справа) (длинный хвост при больших значениях обрезан, чтобы показать статистически значимую часть)