

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

Лебеденко Ольга Олеговна

**Расчёты измеряемых параметров ЯМР на основе  
данных МД моделирования биомолекулярных систем:  
новые методы и приложения**

Научная специальность

1.3.8. Физика конденсированного состояния

Диссертация на соискание ученой степени  
кандидата физико-математических наук

Научный руководитель:

Ph.D.

Николай Русланович Скрынников

Санкт-Петербург — 2024

## Оглавление

	Стр.
Введение . . . . .	5
<b>Глава 1. Диффузия неупорядоченных белков. Исследование N-концевого фрагмента гистона H4 методами МД моделирования и ЯМР спектроскопии . . . . .</b>	<b>23</b>
1.1 Введение . . . . .	23
1.2 Результаты и обсуждение . . . . .	29
1.2.1 Модели воды для моделирования неупорядоченных белков . . . . .	29
1.2.2 МД моделирование N-концевого фрагмента гистона H4	36
1.2.3 Расчет коэффициентов диффузии белков по данным МД моделирования . . . . .	39
1.2.4 Сравнение экспериментальных значений коэффициентов диффузии белков с данными МД моделирования . . . . .	42
1.2.5 Эмпирические методы оценки коэффициента диффузии белков . . . . .	50
1.2.6 Ускоренный протокол МД моделирования диффузионных свойств белков . . . . .	56
1.3 Материалы и методы . . . . .	59
1.3.1 МД моделирование . . . . .	59
1.3.2 Обработка МД траекторий . . . . .	65
1.3.3 ЯМР эксперимент . . . . .	72
1.4 Заключение . . . . .	74
1.5 Дополнительная информация . . . . .	78
<b>Глава 2. Парамагнитное усиление ядерной спиновой релаксации: исследование хвоста гистона H4 в составе нуклеосомной частицы методами МД моделирования и ЯМР спектроскопии . . . . .</b>	<b>89</b>
2.1 Введение . . . . .	89
2.2 Результаты и обсуждение . . . . .	93
2.2.1 Спин-меченые образцы нуклеосомы . . . . .	93

2.2.2	Измерения PRE в спин-меченых образцах нуклеосомы	96
2.2.3	Расчет эффекта PRE из МД траекторий . . . . .	98
2.2.4	Сравнение экспериментальных и расчетных значений PRE . . . . .	100
2.3	Материалы и методы . . . . .	104
2.3.1	Приготовление образцов . . . . .	104
2.3.2	ЯМР эксперименты . . . . .	106
2.3.3	МД моделирование . . . . .	107
2.4	Заключение . . . . .	114
2.5	Дополнительная информация . . . . .	116
<b>Глава 3. Использование ЯМР и МД моделирования для изучения переворотов ароматических колец в кристаллах убиквитина . . . . .</b>		
		<b>128</b>
3.1	Введение . . . . .	128
3.2	Результаты и обсуждение . . . . .	133
3.2.1	Селективно изотопно-меченые остатки Phe в трех кристаллических формах убиквитина . . . . .	133
3.2.2	Перевороты ароматического кольца Phe4 происходят в субмиллисекундном временном интервале . . . . .	136
3.2.3	Перевороты ароматического кольца Phe4 происходят на временных масштабах 10-20 нс . . . . .	140
3.2.4	МД моделирование переворотов ароматических колец .	144
3.3	Материалы и методы . . . . .	147
3.3.1	Получение и очистка белков . . . . .	147
3.3.2	Кристаллизация белков . . . . .	148
3.3.3	ЯМР эксперимент . . . . .	149
3.3.4	МД моделирование . . . . .	153
3.4	Заключение . . . . .	154
3.5	Дополнительная информация . . . . .	160
3.5.1	Расчет скорости релаксации $R_{1\rho}$ . . . . .	160
3.5.2	Учет других форм движения, отличных от переворотов ароматических колец . . . . .	161
<b>Заключение . . . . .</b>		<b>172</b>

Список сокращений . . . . .	177
Благодарности . . . . .	179
Список литературы . . . . .	180

## Введение

### Актуальность

После получения первой кристаллографической структуры белка миоглобина [1], белковые молекулы рассматривались как полностью структурированные, статичные объекты. Потребовалось значительное время для того, чтобы получить экспериментальные свидетельства конформационной подвижности в белках. В числе первых Вютрих и Вагнер показали, что боковые цепи в гидрофобном ядре глобулярного белка совершают скачкообразные переходы между ротамерными состояниями [2]. Таким образом, несмотря на наличие однозначно определяемой структуры, гидрофобное ядро белка до некоторой степени обладает свойствами полимерного расплава. Спустя некоторое время анализ кристаллографических данных заставил предположить, что индивидуальные домены в составе многодоменных белков способны смещаться по отношению друг к другу и что такого рода подвижность играет важную роль в процессе энзиматического катализа [3]. Тогда же было показано, что ряд белковых молекул и, в частности, белки-гистоны, содержат в своем составе неупорядоченные участки, которые, несмотря на отсутствие структуры, играют важную функциональную роль [4].

В настоящее время динамика белковых молекул является предметом активных исследований. В той или иной форме, динамика оказывает влияние на сворачивание и стабильность белков, связывание белков с лигандами, аллостерические эффекты при передаче сигнала и другие функциональные свойства белковых молекул. Внутренне неупорядоченные белки и неупорядоченные домены в составе структурированных белков играют важную роль в архитектуре и механизме действия молекулярных машин, таких как рибосома [5], в организации и функционировании хроматина [6] и ядерных поровых комплексов [7], при сборке клеточного цитоскелета [8] и формировании фазово-разделенных органелл [9]. При этом неупорядоченные пептидные цепи в высокой степени подвержены протеолитической деградации, аномальным посттрансляционным модификациям и, в конечном счете, переходу в патологические формы

(включая прионы и амилоидные фибриллы). Благодаря этим своим особенностям, неупорядоченные белки играют ключевую роль в патогенезе нейродегенеративных заболеваний [10].

Исследования неупорядоченных белков и, в более широком смысле, динамики белковых молекул требуют особого экспериментального инструментария и новых подходов к моделированию белков, отличающихся от тех, которые используются при изучении белковой структуры. Основой структурных исследований являются данные рентгеновской дифрактометрии и, с недавних пор, криоэлектронной микроскопии. Эти данные лишь опосредованно и в ограниченной мере способны пролить свет на динамику белковых молекул. Например, отсутствие электронной плотности при построении модели белковой молекулы по данным кристаллографии или криоэлектронной микроскопии свидетельствует о высокой подвижности соответствующего участка пептидной цепи. Для получения более детальной картины необходимо использовать специальный арсенал экспериментальных методов, среди которых основное место принадлежит спектроскопии ядерного магнитного резонанса (ЯМР). Эксперименты ЯМР позволяют наблюдать сигналы от индивидуальных атомов в составе неупорядоченной белковой цепи и получать информацию о динамических параметрах, относящихся как к непосредственному окружению данного атома, так и к его более дальнему окружению. В основе большинства измерений лежат многомерные гетероядерные эксперименты, позволяющие использовать в качестве зондов ядра  $^{15}\text{N}$  и  $^{13}\text{C}$  основной цепи белка или его боковых цепей. Использование ядер  $^{15}\text{N}$  и  $^{13}\text{C}$  вместо протонов существенно упрощает анализ информации, поскольку интерпретация результатов, как правило, сводится к рассмотрению небольшой спиновой системы, включающей в себя от двух до четырех спинов.

Среди многочисленных параметров, измеряемых с помощью экспериментов ЯМР, следует выделить химические сдвиги и скорости спиновой релаксации. Химический сдвиг отражает особенности окружения отдельно взятого атома. В частности, химические сдвиги чувствительны к так называемой «остаточной структуре» (англ. residual structure), то есть статистической предрасположенности неупорядоченной последовательности образовывать короткоживущие  $\alpha$ -спирали,  $\beta$ -шпильки или другие, более сложные, структурные элементы. Помимо этого анализ

спектров ЯМР позволяет идентифицировать эффект так называемого «обменного уширения», возникающий в результате относительно медленных (в диапазоне от микросекунд до секунд) переходов между состояниями с различными химическими сдвигами. Существует множество специализированных экспериментов, позволяющих охарактеризовать параметры таких обменных процессов. Скорости спиновой релаксации гетероядер (в первую очередь, спинов  $^{15}\text{N}$  и  $^{13}\text{C}$ ) отражают амплитуды и характерные временные константы определенных форм динамики, модулирующих спиновые взаимодействия с участием данных ядер. Как правило, речь идет об изменении ориентации вектора, соединяющего атом азота или углерода и непосредственно связанный с ним протон. Наиболее широко распространенными релаксационными измерениями являются измерения скорости продольной ( $R_1$ ) и поперечной ( $R_2$ ) спиновой релаксации. Однако помимо этого современные экспериментальные методы позволяют также измерять десятки других релаксационных параметров для различных корреляций в многоспиновой системе. Особое место среди релаксационных измерений занимают измерения так называемого эффекта парамагнитного усиления релаксации (англ. paramagnetic relaxation enhancement, PRE). Для измерения эффекта PRE в состав системы вводится парамагнитная метка. В большинстве случаев парамагнитная метка представляет собой нитроксильный радикал, который вступает в реакцию присоединения с боковой цепью цистеина посредством малеимидной химии. Специфика эффекта PRE заключается в том, что этот эффект отражает дальнедействующее дипольное взаимодействие между парамагнитным центром и индивидуальным ядерным спином, так что парамагнитная метка выступает в качестве зонда для исследования стохастической динамики в рассматриваемой системе.

Репертуар методов ЯМР, используемых для исследования неупорядоченных белков, не ограничивается анализом химических сдвигов и скоростей релаксации. Ценную информацию можно получить также из экспериментов ЯМР с использованием импульсных градиентов магнитного поля, позволяющих измерять коэффициент трансляционной диффузии неупорядоченных белков. В свою очередь, коэффициент трансляционной диффузии представляет собой важный параметр, характеризующий компактность конформационного ансамбля неупорядоченного белка.

Аналогичную информацию можно также извлечь из данных малоуглового рентгеновского рассеяния, однако эта методика обладает низкой чувствительностью по отношению к небольшим белкам и пептидам.

Кроме того, среди экспериментальных методов, используемых для исследования неупорядоченных белков, следует отметить такие методы, как круговой дихроизм, аналитическое ультрацентрифугирование и протеолитическая фрагментация. Отдельная категория экспериментов связана с использованием ковалентных меток: помимо упомянутых выше измерений PRE, сюда можно отнести спектроскопию электронного парамагнитного резонанса (ЭПР), Ферстеровский перенос энергии (FRET) и его разновидности (например, BRET), перекрестное сшивание (например, за счет образования дисульфидных связей) и другие подходы. Однако эти методы относительно малоинформативны по сравнению со спектроскопией ЯМР, а использование громоздких меток способно приводить к изменениям в деликатном конформационном равновесии неупорядоченного белка.

Очевидно, что исследование неупорядоченных белков также требует применения альтернативных моделей, отличающихся от статических белковых структур. В качестве таких моделей первоначально использовались структурные ансамбли, т.е. наборы из десятков или сотен структур, воспроизводящих конформационное многообразие белка. В таких моделях каждому конформеру приписывался статистический вес, отражающий предполагаемую частоту его встречаемости в рассматриваемой системе. Для построения таких моделей был разработан ряд специальных программ, наибольшую известность среди которых получили программы ENSEMBLE [11] и ASTEROIDS [12]. В общем виде алгоритм можно описать следующим образом. В качестве отправной точки используется набор из большого числа конформеров, сгенерированных случайным образом (без должного учета конформационных предпочтений данной пептидной последовательности). Затем статистические веса, присвоенные каждому конформеру, оптимизируются таким образом, чтобы воспроизвести набор экспериментальных данных, имеющийся в наличии для данного белка. Одним из недостатков подобных статических моделей является отсутствие в них временной шкалы, т.е. информации о характерных скоростях перехода из одного конформационного состояния в другое. Также следует отметить, что качество подобных моделей целиком



зависит от используемых при их конструировании экспериментальных данных; сравнительно небольшой объем данных, сложность в их интерпретации и невысокая информативность ведут к тому, что создание подобных моделей требует больших усилий, а результаты оказываются плохо определены.

В качестве альтернативного решения Шу и Скрынников предложили использовать траектории Молекулярной Динамики (МД) в качестве модели неупорядоченных белков [13]. В дальнейшем этот подход получил большое развитие и использовался для анализа широкого круга неупорядоченных систем [14; 15]. В принципе, такая модель позволяет исчерпывающим образом охарактеризовать поведение внутренне неупорядоченного белка, включая скорости перехода между различными конформационными состояниями. Также нужно отметить, что в основе модели лежит априорное знание в виде силового поля, которое достаточно реалистичным образом отражает взаимодействия между различными группами пептидной цепи, а также взаимодействие каждой из этих групп с водой. Таким образом, есть основания предполагать, что достоверная модель может быть получена из первых принципов, без привлечения экспериментальных данных.

Однако на этом пути встает ряд препятствий технического характера. Во-первых, такого рода МД моделирование требует исключительно длинных траекторий с тем, чтобы получить приемлемое покрытие обширного фазового пространства, доступного для неупорядоченного белка. Внедрение в практику компьютеров, оснащенных графическими ускорителями, а также ряд усовершенствований в алгоритмах МД моделирования (например, использование методов ускоренной молекулярной динамики) позволяют в значительной мере решить данную проблему, по крайней мере для неупорядоченных фрагментов небольшого размера. Во-вторых, силовые поля традиционно создавались, оптимизировались и тестировались для структурированных, глобулярных белков. Для такого рода систем накоплен большой опыт исследований, позволяющих оценивать сильные и слабые стороны существующих силовых полей и вносить соответствующие коррективы при их обновлении. Первые попытки использования стандартных силовых полей для моделирования неупорядоченных белков привели к систематическим отклонениям от экспериментальных данных. Как выяснилось, классические силовые

поля имеют тенденцию переоценивать склонность к образованию (нестабильных)  $\alpha$ -спиралей в неупорядоченных белках и их склонность к коалесценции, т.е. образованию компактных форм по типу «расплавленной глобулы» [16]. Дальнейшие исследования показали, что эти артефакты явились следствием переоценки взаимодействий белок-белок в сравнении с взаимодействиями белок-вода. Для коррекции этой проблемы были разработаны новые варианты силовых полей, включая, в частности, новые модели воды [17]. Однако, в дальнейшем выяснилось, что в некоторых случаях новые модели воды оказываются недостаточно хорошо сбалансированными и приводят к частичной дестабилизации глобулярных белков [18].

Таким образом, к настоящему времени сложилась ситуация, когда достоверность МД моделей неупорядоченных белков, а также белков, содержащих в своем составе неупорядоченные элементы, находится под вопросом. Имеющийся опыт экспериментальной валидации носит разрозненный и ограниченный характер. Количество исследуемых систем достигает нескольких десятков. Также и количество различных схем молекулярной динамики, используемых при их моделировании, исчисляется десятками. Для валидации моделей применяются достаточно разнородные и при этом ограниченные наборы экспериментальных данных. В ходе валидации расчеты измеряемых параметров зачастую осуществляются по упрощенным и даже грубым вычислительным схемам. В итоге многочисленные публикации в этой области рисуют достаточно пеструю картину, где разные научные коллективы доказывают преимущества различных схем моделирования, причем их выводы подчас вступают в противоречие друг с другом. Такое не слишком удовлетворительное положение вещей послужило мотивацией для написания настоящей диссертационной работы. Настоящее исследование посвящено разработке и применению различных схем МД моделирования, настроенных оптимальным образом для последующих расчетов параметров ЯМР. В работе использованы новые и усовершенствованные алгоритмы для расчета параметров ЯМР на основе данных МД моделирования. Наконец, в работе проводится сравнение расчетных параметров с результатами экспериментальных измерений. Такого рода анализ позволяет более надежным и достоверным образом осуществлять валидацию

рассматриваемых моделей МД. В свою очередь, успешная валидация дает возможность для использования модели в качестве уникального источника информации о динамике неупорядоченных белков и, в более широком смысле, о динамике белковых молекул в целом.

## Цели работы

Первая глава посвящена разработке методологии для расчета диффузионных параметров неупорядоченного белка по данным МД моделирования. Исследование проводилось на примере N-концевого участка гистона H4 с привлечением соответствующих экспериментальных данных. В рамках настоящей работы был поставлен вопрос о том, как обеспечить получение большого объема данных МД, необходимых для решения этой задачи, путем эффективного использования доступных вычислительных ресурсов. Отдельное внимание было уделено апробации различных моделей воды в контексте данной задачи. Для того, чтобы придать работе общий характер, дополнительно было проведено исследование трансляционной и вращательной диффузии для небольшого глобулярного белка убиквитина. Основным вопросом, поставленным в ходе настоящей работы, можно сформулировать следующим образом: возможно ли на основании анализа диффузионных свойств неупорядоченного белка сделать содержательный вывод о мере компактности соответствующего конформационного ансамбля? Также была поставлена цель сравнить строгие результаты, полученные в ходе нашего исследования, с результатами различных эмпирических схем для расчета диффузионных параметров по данным МД моделирования.

Вторая глава посвящена расчетам и интерпретации скоростей парамагнитной релаксации (PRE) в сложной многокомпонентной системе, включающей в себя как структурированные, так и неупорядоченные элементы. В качестве такой системы была исследована нуклеосомная частица, где нитроксильная метка прикреплялась к различным остаткам на поверхности гистона H3 (в его структурированной части), а для наблюдения были выбраны спектральные сигналы, соответствующие подвижному N-концевому фрагменту гистона H4. Исходно было понятно, что данные PRE, полученные в таких образцах, несут информацию о локализации

N-концевого хвоста гистона H4 по отношению к телу нуклеосомной частицы. Однако предстояло ответить на следующие вопросы. Возможно ли содержательным образом интерпретировать данные PRE на основе данных МД моделирования? Необходимо ли при этом реализовать строгую схему для расчета скоростей PRE, учитывающую трансляционную динамику хвоста H4, или для этих целей можно ограничиться использованием упрощенной формулы Соломона-Бломбергена? Удастся ли добиться желаемой сходимости расчетов PRE в траекториях с продолжительностью в несколько микросекунд?

В третьей главе обсуждается использование данных МД для интерпретации экспериментальных результатов, относящихся к скачкообразным движениям боковых цепей фенилаланина в кристалле небольшого глобулярного белка убиквитина. Основная задача состояла в изучении того, может ли моделирование МД использоваться в качестве дополнительного источника информации, позволяющего уточнить результаты анализа данных ЯМР – а именно, идентифицировать происхождение расщепления в некоторых из наблюдаемых спектральных резонансов, объяснить отсутствие некоторых ожидаемых сигналов, и прояснить характер конформационных переходов в боковых цепях фенилаланина. Помимо этого, была поставлена задача выявить структурные факторы, влияющие на скорость переворотов ароматических колец.

### **Научная и практическая значимость**

1. Как описано выше, исследования неупорядоченных белков сталкиваются с дефицитом экспериментальной информации. В этих условиях измерения коэффициента трансляционной диффузии посредством эксперимента ЯМР с использованием импульсных градиентов магнитного поля представляют собой потенциально важный элемент исследования, позволяющий охарактеризовать конформационный ансамбль белка по признаку его компактности. В настоящей работе представлена схема для МД моделирования неупорядоченного белка (пептида), позволяющая максимально строгим образом рассчитать его диффузионные параметры. В качестве первого шага записывается траектория белка в

ячейке малого размера. Затем отдельные кадры из этой траектории используются в качестве стартовых точек для того, чтобы записать набор из коротких траекторий в ячейках среднего и большого размера. Экстраполяция полученных результатов к пределу бесконечно большой ячейки позволяет получить искомое значение коэффициента диффузии. Такой подход позволяет максимально эффективно использовать имеющиеся вычислительные мощности, поскольку для записи коротких траекторий может быть задействован кластер из вычислительных серверов под контролем программы для управления очередью. Помимо этого данный подход также создает благоприятные условия для моделирования с использованием термодинамического ансамбля NVE (что обеспечивает максимально строгий уровень моделирования), поскольку при записи коротких траекторий отсутствует риск дрейфа температуры. В настоящей работе было также показано, что при моделировании диффузии неупорядоченного белка можно с успехом использовать не только стандартный шаг интегрирования 1-2 фс, но и удлиненный шаг 4 фс. Необходимым условием для этого является приложение схемы перераспределения массы водорода (англ. Hydrogen Mass Repartition scheme, HMR) не только к молекуле белка, но и к молекулам воды в объеме ячейки. Данный результат открывает возможности для дальнейшего ускорения вычислений.

Моделирование и расчеты были проведены для неупорядоченного N-концевого фрагмента гистона H4, состоящего из 25 аминокислотных остатков (N-H4). В ходе исследования было показано, что на результаты существенным образом влияет вязкость МД воды. В связи с этим наблюдением в рамках настоящей диссертационной работы была систематическим образом исследована вязкость воды для моделей TIP4P-Ew, TIP4P-D и OPC. После коррекции на вязкость воды было обнаружено, что расчетное значение для коэффициента диффузии пептида N-H4 в воде TIP4P-Ew значительно завышено по сравнению с данными эксперимента. Это является следствием чрезмерной компактности пептида в траектории с использованием TIP4P-Ew. Напротив, диффузионные параметры полученные в воде TIP4P-D и OPC согласуются с экспериментальными данными. Таким образом, можно заключить, что модель воды TIP4P-Ew не подходит для исследования неупорядоченных белков, в то время

как две оставшиеся модели могут быть с успехом использованы в подобных исследованиях. Данный вывод подтверждается также анализом релаксационных данных для атомов  $^{15}\text{N}$  в основной цепи пептида. Тем самым в настоящей работе показан пример успешной валидации различных моделей молекулярной динамики на основе диффузионных данных и данных спиновой релаксации. При выборе из двух успешных моделей воды, TIP4P-D и OPC, можно отдать предпочтение последней, т.к. она требует лишь минимальной коррекции в отношении вязкости и дает несколько более точные результаты при моделировании диффузионных параметров малого глобулярного белка убиквитина.

В современной литературе можно найти десятки примеров работ, где коэффициент трансляционной диффузии используется для валидации различных МД моделей неупорядоченных белков (пептидов). Однако вместо строго расчета коэффициента трансляционной диффузии, как это сделано в настоящей диссертационной работе, авторы используют различные эмпирические подходы – например, программы для гидродинамических вычислений HYDROPRO и HullRadSAS, а также формулы Nygaard и Kirkwood-Riseman. Сравнение результатов, получаемых с помощью вышеназванных эмпирических методов, с результатами полноценного анализа, проведенного в настоящей работе, показывает, что эмпирические методы терпят неудачу. В частности, применение эмпирических методов может приводить к полностью ошибочным выводам – к примеру, полученные результаты указывают на то, что вода TIP4P-Ew является оптимальным выбором для моделирования неупорядоченного пептида. Основная причина данной неудачи заключается в том, что программы HYDROPRO и HullRadSAS и основанные на их применении эмпирические формулы предназначены для статичных (жестких) белковых структур и, соответственно, не учитывают роли сегментальной динамики в трансляционной диффузии неупорядоченного белка.

2. Нуклеосомная коровая частица (англ. nucleosome core particle, NCP) является фундаментальной структурной единицей хранения генетической информации. NCP представляет собой сборку из восьми белков-гистонов, вокруг которой обернута в 1.7 оборота двухнитевая ДНК. При этом концевые участки гистонов (хвосты) имеют вид протяженных неупорядоченных цепочек, направленных в растворитель.

С функциональной точки зрения гистоновые хвосты играют важнейшую роль, обеспечивая взаимодействия нуклеосомы с транскрипционными факторами и различными хроматин-ассоциированными белками, а также оказывая влияние на структуру хроматина. Взаимодействие гистоновых хвостов с различными лигандами регулируется посредством своеобразного кода, задаваемого ковалентными модификациями боковых цепей. Ввиду своего неупорядоченного характера, гистоновые хвосты не поддаются исследованию стандартными методами структурного анализа и могут быть исследованы почти исключительно с помощью ЯМР спектроскопии. В числе возможных спектроскопических экспериментов особое место занимают измерения эффекта PRE, которые позволяют охарактеризовать локализацию гистоновых хвостов по отношению к телу нуклеосомы. При проведении таких экспериментов парамагнитные метки прикрепляются к структурированным элементам NCP (в настоящем исследовании для этой цели были избраны остатки 36, 65, 79 и 125 в составе гистона H3), а сигналы регистрируются для подвижных остатков гистонового хвоста (в данном случае N-концевой участок гистона H4, обогащенного изотопом  $^{15}\text{N}$ ). При этом до настоящего времени не существовало теоретической модели, позволяющей корректным образом интерпретировать данные PRE в подобной системе, которая включает в себя как структурированные, так и неупорядоченные элементы.

В рамках настоящей диссертационной работы была исследована возможность интерпретации данных PRE в нуклеосомной частице на основе модели в виде конформационного ансамбля с использованием хорошо известного уравнения Соломона-Бломбергена. Полученные результаты показывают бессодержательный характер такой интерпретации. К примеру, полученные данные PRE можно воспроизвести используя модель из двух конформеров, что очевидно несовместимо с базовыми представлениями о поведении гистоновых хвостов. Аналогичным образом, данные можно воспроизвести, используя ансамбль из 42 конформеров, где значения PRE полностью определяются двумя из них, а остальные фактически соответствуют случайным конформациям гистонового хвоста. Причиной такого неудовлетворительного положения дел является крайне высокая чувствительность скоростей PRE к расстояниям между парамагнитным центром и ядерными спинами  $^{1}\text{H}^{\text{N}}$ , выступающими в роли зондов.

В качестве альтернативного решения автором настоящего исследования была предложена интерпретация, основанная на прямом расчете скоростей PRE из данных молекулярной динамики. При этом строгий расчет принимает во внимание не только изменение ориентации вектора, соединяющего парамагнитный спин с ядерным спином, но и модуляцию межспинового расстояния за счет трансляционного движения мобильного гистонового хвоста. Результаты, полученные из траектории NCP в воде TIP4P-D, записанной автором данного исследования, а также набора траекторий в воде OPC, записанного Пенгом с коллегами, показывают хорошее качественное согласие с данными эксперимента. Таким образом в данной работе удалось успешно валидировать модели МД для нуклеосомной частицы на основе экспериментальных данных PRE. В качестве следующего шага прошедшие такого рода проверку модели МД были использованы для получения информации о пространственном положении гистонового хвоста H4. Проведенный анализ показал, что N-концевой участок гистона H4 локализуется преимущественно в окрестности нуклеосомной ДНК, что обусловлено в первую очередь электростатическим взаимодействием между положительно заряженным хвостом и отрицательно заряженной ДНК. При этом хвост H4 сохраняет высокую подвижность, взаимодействуя с ДНК по механизму «нечеткого взаимодействия» (англ. fuzzy interaction). Локализация хвоста в окрестности ДНК ограничивает его доступность для связывания с хроматин-ассоциированными белками, включая гистон-модифицирующие ферменты и транскрипционные факторы. Для дальнейшего исследования подвижных гистоновых хвостов в составе нуклеосомной частицы желательно улучшить сходимость расчетов PRE на основе данных МД моделирования. Как показал проведенный в настоящей работе анализ, даже траектории длиной в десятки микросекунд не обеспечивают в этом отношении удовлетворительной сходимости. Основной причиной для этого является уже упомянутая выше чувствительность PRE к межспиновым расстояниям. Для улучшения сходимости можно рекомендовать применение распределенных вычислительных ресурсов, ускоренные схемы моделирования (например, с использованием шага интегрирования 4 фс) и другие подобные решения.



3. Скачкообразные переходы боковых цепей в глобулярном белке можно рассматривать как индикатор пластичности белковой структуры и присутствия в ней значительных динамических флуктуаций. Нашими соавторами из исследовательской группы П. Шанды была исследована динамика боковых цепей фенилаланина в трех различных кристаллических формах убиквитина. Для проведения измерений методом ЯМР спектроскопии с вращением под магическим углом (англ. magic angle spinning, MAS) были изготовлены образцы с новой схемой изотопного обогащения, позволяющей селективно ввести спины  $^1\text{H}$  и  $^{13}\text{C}$  в позициях  $\epsilon_1$ ,  $\epsilon_2$  ароматического кольца. Оба остатка фенилаланина в молекуле убиквитина, Phe4 и Phe45, расположены на поверхности белковой молекулы и, таким образом, потенциально чувствительны к влиянию кристаллических контактов. В настоящем диссертационном исследовании показано, как анализ данных МД моделирования дает возможность дополнить и уточнить интерпретацию, основанную на экспериментальных данных ЯМР.

В частности, в диссертационной работе показано, что расщепление сигнала от Phe4 в кубической форме кристалла связано с упаковкой двух остатков Phe4, относящихся к белковым молекулам типа В в составе асимметричной ячейки, на кристаллическом интерфейсе. С помощью программы SHIFTX2 на основе данных МД моделирования были рассчитаны химические сдвиги спинов  $^1\text{H}^\epsilon$  и  $^{13}\text{C}^\epsilon$ , включающие в себя вклад от кольцевого тока в ароматическом кольце, подтвердившие отнесение наблюдаемых сигналов в данной кристаллической форме. Анализ траектории МД также показал, что динамика боковых цепей фенилаланина сводится к скачкам торсионного угла  $\chi_2$  и, следовательно, может быть описана как перевороты (англ. flips) ароматического кольца, тогда как угол  $\chi_1$  остается неизменным. Данные МД моделирования свидетельствуют о том, что перевороты ароматического кольца в остатке Phe4 происходят в масштабе времени наносекунд, тогда как в остатке Phe45 характерный временной масштаб составляет десятки и сотни микросекунд, причем данное соотношение выполняется независимо от исследуемой кристаллической формы. Этот результат позволяет объяснить отсутствие экспериментально наблюдаемых сигналов от остатка Phe45, что можно приписать эффекту обменного уширения, связанному с микросекундной

динамикой. В ходе исследования МД была также предпринята попытка объяснить различия в характерных временах переворотов ароматических колец Phe4 и Phe45. В частности были рассмотрены гипотезы о том, что скорость переворотов определяется конформацией боковой цепи, доступным для вращения объемом (который можно охарактеризовать через доступную для растворителя площадь поверхности остатка) и кристаллическими контактами. Ни одна из этих гипотез не нашла прямого подтверждения, заставляя предположить, что скорость переворотов определяется более тонкими кооперативными эффектами, имеющими место в структуре убиквитина. Таким образом, было продемонстрировано, что данные МД моделирования белковых кристаллов являются ценным источником информации, взаимодополняющим результаты экспериментального исследования методом твердотельного ЯМР.

### **Аппробация**

Материалы диссертации были представлены на следующих конференциях:

1. «65-th Biophysical Society Annual Meeting», онлайн, 22 февраля – 26 февраля 2021.
2. Международная конференция «Современное развитие методов магнитного резонанса 2021», Казань, Россия, 1 ноября – 5 ноября 2021.
3. VI ежегодная конференция ИТБМ СПбГУ «Актуальные проблемы трансляционной биомедицины – 2022», Санкт-Петербург, Россия, 25 июля – 25 июля 2022 г.
4. «Chinese Biophysics Congress 2024», Ланьчжоу, Китай, 25 июля – 28 июля 2024.

### **Личный вклад**

Глава 1. Планирование исследования, получение, обработка и анализ данных МД моделирования, интерпретация полученных результатов и подготовка материалов для публикации осуществлены лично автором. В частности, автору принадлежит идея и реализация специальной

схемы МД моделирования для исследования трансляционной диффузии разупорядоченного белка. Используемые в работе данные ЯМР в своей основной части были получены В.А. Саликовым под руководством и при участии И.С. Подкорытова в Лаборатории био-ЯМР СПбГУ. При обработке и анализе данных МД были использованы программные скрипты, созданные С.А. Измайловым.

Глава 2. Планирование и реализация исследования в части, касающейся МД моделирования, обработка и анализ данных моделирования, интерпретация полученных результатов и подготовка материалов для публикации осуществлены лично автором. В частности, автором были реализованы в виде программного кода ряд теоретических схем для расчёта скоростей парамагнитной релаксации (PRE) на основе данных МД. Подготовка образцов нуклеосомы и измерения данных PRE методом ЯМР спектроскопии были выполнены В. Саном в лаборатории К. Джароньяка.

Глава 3. Планирование и реализация исследования в части, касающейся МД моделирования, обработка и анализ данных моделирования, интерпретация полученных результатов и подготовка материалов для публикации осуществлены лично автором. Подготовка образцов кристаллического убиквитина и экспериментальное исследование внутренней динамики в этих образцах методом ЯМР спектроскопии были выполнены Д. Гауто и Л. Беккер в лаборатории П. Шанды.

На всех этапах исследования постановка задач и интерпретация результатов осуществлялась под научным руководством Н.Р. Скрынникова.

## **Публикации**

Основные результаты по теме диссертации изложены в 3 печатных работах в журналах, включенных в систему цитирования Scopus и Web of Science:

1. Using NMR diffusion data to validate MD models of disordered proteins: Test case of N-terminal tail of histone H4 [Text] / O. O. Lebedenko [et al.] // Biophysical Journal. — 2024. — Jan. — Vol. 123, no. 1. — P. 80—100.

2. Conformational and Interaction Landscape of Histone H4 Tails in Nucleosomes Probed by Paramagnetic NMR Spectroscopy [Text] / W. Sun [et

al.] // Journal of the American Chemical Society. — 2023. — Sept. — Vol. 145, no. 46. — P. 25478—25485.

3. Aromatic ring flips in differently packed ubiquitin protein crystals from MAS NMR and MD [Text] / D. F. Gauto [et al.] // Journal of Structural Biology: X. — 2023. — Dec. — Vol. 7. — P. 1—12.

### **Объем и структура работы**

Диссертационная работа состоит из введения, трёх глав с изложением основных результатов исследования, заключения и списка литературы. Каждая из глав в свою очередь включает в себя введение, описание результатов и их обсуждение, описание использованных в ходе исследования материалов и методов, выводы, и дополнительные материалы. Работа изложена на 209 страницах, содержит 37 рисунков и 9 таблиц. Библиографический аппарат состоит из 310 ссылок.

### **Основные научные результаты**

1. Разработана высокоэффективная схема МД моделирования для предсказания диффузионных параметров разупорядоченных белков [19].
2. Показано, что получаемые в результате моделирования параметры трансляционной диффузии отражают, во-первых, вязкость моделируемой воды и, во-вторых, степень компактности конформационного ансамбля моделируемого белка [19].
3. Показано, что используемые в литературе эмпирические подходы для расчёта диффузионных параметров разупорядоченных белков на основе данных МД моделирования обладают низкой точностью и зачастую могут приводить к качественно неверным заключениям [19].
4. Было обнаружено, что для ускорения МД моделирования трансляционной диффузии может быть использован шаг интегрирования 4 фс, однако для сохранения высокой точности результатов необходимо прикладывать схему перераспределения

массы водорода (NMR) не только к молекуле белка, но и к молекулам воды [19].

5. Реализована на высоком уровне строгости схема расчёта скоростей парамагнитной релаксации (PRE) на основе данных МД моделирования для систем, содержащих в своём составе разупорядоченные участки пептидной цепи (нуклеосомная частица) [20].
6. Показано, что данные МД моделирования для нуклеосомной частицы находятся в качественном согласии с экспериментальными данными PRE и могут быть использованы для реконструкции пространственного распределения подвижных гистоновых хвостов [20].
7. Посредством анализа набора из сверх-длинных МД траекторий нуклеосомной частицы, показано, что задача о расчёте PRE ставит очень высокие требования к сходимости МД модели, а попытки использования данных PRE для формирования модели в виде статичного конформационного ансамбля ведут к плохо определённым результатам [20].
8. Показано, что данные МД моделирования белковых кристаллов способны дополнить и уточнить результаты анализа белковой динамики на основе экспериментальных данных ЯМР и, в частности, позволяют относить некоторые из наблюдаемых сигналов ЯМР и объяснять отсутствие других сигналов [21].

### **Положения выносимые на защиту**

1. Новая схема для МД моделирования разупорядоченных белков, представленная в данной диссертационной работе, позволяет с высокой эффективностью получать данные, необходимые для расчёта скорости трансляционной диффузии белка.
2. Сравнительный анализ измерений методом диффузионного ЯМР и расчётов на основе данных МД моделирования позволяет сделать заключение о достоверности МД модели в части, касающейся степени компактности моделируемого разупорядоченного белка. Для небольших пептидов получаемая информация

носит уникальный характер, поскольку альтернативный метод, малоугловое рентгеновское рассеяние, характеризуется в этом случае значительной экспериментальной погрешностью.

3. Представленная методика для расчёта диффузионных параметров на основе данных МД моделирования позволяет установить, что широко используемые для этой цели эмпирические алгоритмы обладают низкой точностью и зачастую ведут к качественно неверным выводам в отношении характеристик моделируемого конформационного ансамбля.
4. В диссертационном исследовании представлена новая схема для расчёта скоростей парамагнитной релаксации (PRE) на основе данных МД моделирования, где строгим образом учитывается не только вращательная динамика исследуемой системы, но также и сегментальная динамика её разупорядоченных элементов (на примере подвижных гистоновых хвостов в составе нуклеосомной частицы).
5. Анализ экспериментальных данных PRE на основе МД модели нуклеосомной частицы показывает, что положительно заряженные гистоновые хвосты локализуются в окрестности отрицательно заряженной нуклеосомной ДНК, взаимодействуя с ней по механизму нечёткого взаимодействия (англ. *fuzzy interaction*) и сохраняя при этом относительно высокую степень конформационной подвижности.
6. Анализ данных PRE в рамках МД модели нуклеосомной частицы продемонстрировал, что для сходимости расчётов PRE необходимо использовать траектории совокупной длиной в сотни микросекунд. Аномальная чувствительность данных PRE к межспиновым расстояниям не позволяет строить на их основе достоверные структурные модели, где конформационное многообразие гистоновых хвостов отражено в виде статичного конформационного ансамбля.
7. При исследовании кристаллического убиквитина методом твердотельного ЯМР данные МД моделирования помогают осуществить отнесение спектров и интерпретацию результатов в отношении скачкообразной динамики боковых цепей белка.

# Глава 1. Диффузия неупорядоченных белков. Исследование N-концевого фрагмента гистона H4 методами МД моделирования и ЯМР спектроскопии

## 1.1 Введение

В последние несколько десятилетий возрос интерес к внутренне неупорядоченным белкам (от англ. Intrinsically Disordered Proteins, IDPs) и белкам, содержащим внутренне неупорядоченные домены (от англ. Intrinsically Disordered Protein Regions, IDPR). С эволюционной точки зрения внутренне неупорядоченные белки относятся к важным компонентам современного клеточного механизма — доля IDPs от общего числа белков увеличивается в ряду организмов археи – бактерии – эукариоты [22]. Известно, что IDPs и IDRs входят в состав таких молекулярных машин, как рибосома [23], хроматин [6], ядерный поровый комплекс [7]. Кроме того, IDPs и IDRs являются ключевыми участниками таких молекулярных процессов, как сборка компонентов цитоскелета [8] и формирование немембранных органелл путем фазового разделения [9]. Главное преимущество IDPs над упорядоченными белками заключается в их более высокой эффективности в процессе передачи клеточных сигналов. Сигнальные пути, в которых задействованы IDPs, характеризуются более высокой степенью сопряженности и имеют более широкий диапазон клеточного ответа [24]. Однако увеличение эффективности имеет свою цену: неупорядоченные белки чувствительны к протеолитическому расщеплению [25], подвержены случайным модификациям [26], склонны к переходам в прионную форму [27] и образованию амилоидных фибрилл [28]. Указанные уязвимости определяют печально известную роль неупорядоченных белков в развитии нейродегенеративных расстройств и онкологических заболеваний [29].

В настоящее время ведутся активные поиски лекарственных препаратов, воздействующих на IDPs [30]. Трудности возникают как при разработке низкомолекулярных соединений, взаимодействующих с внутриклеточными IDPs мишенями, так и при создании терапевтических

антител, нейтрализующих внеклеточные IDPs мишени ( $A\beta$ -пептид и тау-белок) [31; 32].

Со структурной точки зрения IDPs не имеют выделенной конформации в пространстве и существуют в виде динамического ансамбля конформеров. Упрощенно, в свободном состоянии IDPs можно рассматривать как «случайный клубок» (от англ. random coil). При этом в большинстве случаев при формировании комплекса с упорядоченным белком-мишенью происходит переход IDPs из неупорядоченной формы в структурированную, так называемое «сворачивание при связывании» (от англ. folding upon binding) [33]. Однако возможен сценарий, когда IDPs сохраняют неупорядоченную форму даже после образования комплекса с упорядоченным белком, так называемое «нечеткое взаимодействие» (от англ. fuzzy interaction) [34; 35]. Более того, известны случаи, когда IDP остается полностью неупорядоченным в связанном состоянии с другим IDP-партнером [36]. Долгое время считалось, что в свободном состоянии «случайного клубка» IDPs не имеют каких-либо структурных особенностей [37]. Однако впоследствии выяснилось, что для IDPs характерно наличие «остаточной структуры» (от англ. residual structure) [38; 39], которая влияет на сродство IDP к белку-мишени (эффект, как правило, невелик, но биологически значим). Например, склонность к образованию структурных элементов в форме спирали в трансактивационном домене белка р53 влияет на его сродство к белку Mdm2 [40]. Этот и многие другие подобные примеры усиливают интерес к структурным исследованиям неупорядоченных белков.

Классические методы структурной биологии вроде криоэлектронной микроскопии, рентгеноструктурного анализа, и спектроскопии ЯМР, основанной на эффекте Оверхаузера (от англ. Nuclear Overhauser effect, NOE), сообщают лишь косвенную информацию о структурном состоянии IDPs и IDRs и в общем случае оказываются малоинформативными для исчерпывающего описания динамического ансамбля неупорядоченных белков. Ценная структурная информация может быть получена из данных ЯМР экспериментов, таких как измерение химических сдвигов, скоростей ядерной спиновой релаксации, эффекта парамагнитного усиления ядерной спиновой релаксации (от англ. paramagnetic relaxation enhancement, PRE), а также измерение диффузионных свойств в ЯМР



экспериментах с импульсным градиентом магнитного поля, ЯМР с ИГМП, (от англ. pulsed field gradient, PFG). В частности, диффузионные эксперименты дают информацию о компактности конформационного ансамбля IDP, для которого пространственная конфигурация отдельно взятого конформера может варьироваться от полностью вытянутого состояния «случайного клубка» до достаточно компактного состояния «расплавленной глобулы». ЯМР с ИГМП эксперименты имеют особое практическое значение по отношению к коротким IDP (неупорядоченным пептидам) с молекулярной массой несколько кДа, т.к. альтернативный метод малоуглового рентгеновского рассеяния, МУРР, (от англ. small-angle X-ray scattering, SAXS) оказывается неэффективным по отношению к небольшим системам. Дело в том, что при небольшой массе снижается чувствительность [41], а усиление сигнала с помощью синхротронных источников высокой интенсивности ведет к радиационному разрушению образца [42]. В свою очередь, диффузионные методы спектроскопии ЯМР могут быть приложены к IDPs произвольного размера.

Экспериментальные данные для свернутых белков представляют в виде пространственной модели, сведения о которой помещают в базу данных белковых структур PDB (от англ. Protein Data Bank). Возникает вопрос: что должна представлять собой эквивалентная модель для неупорядоченных белков? В классическом подходе пространственную модель IDPs определяют, как конечный структурный ансамбль, в котором вероятность появления каждого из конформеров кодируется своим статистическим весом. Наиболее известными подходами для построения статистических ансамблей IDPs являются алгоритмы ENSEMBLE [11] и ASTEROIDS [12]. Относительно недавно было высказано предположение о том, что траектории молекулярной динамики (МД) могут выступать в качестве полноценной структурной модели для неупорядоченных белков. Преимущество в том, что валидация МД моделей может быть осуществлена напрямую при помощи параметров ЯМР и других экспериментальных данных [13]. Действительно, МД траектория уже содержит всю необходимую информацию о вероятностях перехода между отдельными конформационными состояниями, что позволяет напрямую вычислять параметры ЯМР, такие как значения скоростей релаксации и величина эффекта PRE.

На первый взгляд, использование МД траекторий в качестве «структурно-динамической модели» IDPs кажется естественным решением, однако на практике эта идея сталкивается с рядом сложностей. Во-первых, классические силовые поля (и модели воды) оказываются непригодными для моделирования неупорядоченных белков. Во-вторых, требуются достаточно длинные траектории МД, чтобы адекватным образом покрыть обширное фазовое пространство для конформационного многообразия неупорядоченного белка. Важно отметить, что за прошедшее десятилетие был достигнут существенный прогресс в решении упомянутых трудностей. Так, был разработан ряд новых специализированных силовых полей и моделей воды для моделирования неупорядоченных белков (а также для моделирования свернутых белков) [18; 43–45]. В то же время, благодаря техническому прогрессу улучшилась ситуация со статистикой при моделировании IDPs, а именно: появились ускорители математических вычислений на графических процессорах (от англ. graphics processing unit, GPU) и специальные схемы ускоренной молекулярной динамики [15; 46–48]. Благодаря прогрессу последних лет стало возможным записывать длинные статистически самосогласованные траектории МД для IDPs небольшого размера (т.е. для длинного пептида) и использовать полученную модель напрямую для сравнения с экспериментальными данными [49; 50]. Для IDPs большего размера Блэкледж и др. разработали алгоритм ABSURD, суть которого заключается в представлении структурной модели IDP, в виде набора коротких МД траекторий с регулируемым статистическим весом [51].

В настоящей работе был изучен вопрос: могут ли быть использованы данные диффузионных методов ЯМР с импульсным градиентом магнитного поля для валидации МД модели неупорядоченного белка? Как уже было отмечено, коэффициент трансляционной диффузии  $D_{tr}$  является одним из ключевых параметров, сообщающих информацию о компактности конформационного ансамбля IDPs. Однако расчет  $D_{tr}$  по данным МД моделирования (с целью дальнейшего сравнения с результатами эксперимента) представляет собой нетривиальную задачу.

На практике хорошо известно, что расчетные значения  $D_{tr}$  зависят от размера ячейки моделирования. Данный артефакт был отмечен для простых жидкостей, полимеров и белков, см. [52–58]. Происхождение

эффекта обязано ограничениям, накладываемым в МД моделировании, а именно: равенству нулю полного импульса системы. Это требование приводит к ситуации, при которой из полной динамики системы вычитается некоторая её часть. Эффект становится существенным для систем небольшого размера, что в конечном итоге ведет к замедлению диффузионного движения в сравнении с ожидаемой [59]. Для того чтобы обойти данную трудность, Йех и Хаммер предложили проводить моделирование в ячейках увеличивающегося размера и экстраполировать значение  $D_{tr}$  к пределу бесконечно большой ячейки [59]. Несмотря на то, что данный подход является общепринятым решением, он довольно сложен с вычислительной точки зрения. В частности, требуется запись длинных МД траекторий для IDP в ячейках, содержащих большое количество молекул воды.

Кроме того, возникают сложности с определением  $D_{tr}$  по данным моделирования в ансамбле NPT (характеризуется фиксированным числом частиц  $N$ , постоянным давлением  $P$  и температурой  $T$ ) с использованием термостата Ланжевена. Дело в том, что термостат Ланжевена искажает естественные гидродинамические потоки в системе [60]. Так, при наиболее часто используемых значениях коэффициента трения  $\gamma$  (отвечает за регулировку температуры) происходит существенное увеличение вязкости растворителя и, как следствие, замедление общей динамики, т.е. занижение эффективного значения  $D_{tr}$  [61]. В принципе, для того чтобы учесть эффект замедления динамики, можно ввести эмпирическую схему коррекции. Относительно недавно, Хикс и его коллеги сформулировали алгоритм вычисления поправок для коэффициента вращательной диффузии  $D_{rot}$  [62]. Тем не менее предложенный алгоритм не является универсальным, т.е. зависит от размеров моделируемого белка и предполагает громоздкие вычисления. Очевидно, что в отношении IDPs подобный эмпирический подход столкнется с дополнительными трудностями. Наилучшим решением является использование другого термостата, такого как термостат масштабирования скоростей или термостат Бусси-Парринелло (от англ. Bussi-Parrinello velocity rescaling thermostat) [63].

Настоящая работа посвящена исследованию неупорядоченного N-концевого фрагмента гистона H4 человека (аминокислотные остатки с 1 по 25). N-концевой фрагмент гистона H4 (далее обозначается как N-H4)

играет важную роль в транскрипции генов, сборке и ремоделировании хроматина [64]. В дополнение к пептиду N-H4, в качестве отправной точки была изучена трансляционная диффузия известного упорядоченного белка убиквитина (далее обозначается как Ub). И наконец, систематическим образом была исследована самодиффузия молекул воды с целью сопоставления расчетных значений  $D_{tr}$  для N-H4 и Ub с вязкостью конкретной модели воды [65].

Моделирование проводилось на базе пакета программ Амбер20 [66] с использованием силового поля ff14SB [67] в трех различных моделях воды: TIP4P-Ew [68], TIP4P-D [18] и OPC [69]. Как для белка Ub, так и для пептида N-H4 было показано, что модель воды TIP4P-Ew ведет к значительному завышению расчетных значений коэффициентов диффузии  $D_{tr}$ . В свою очередь, модель воды TIP4P-D несколько занижает  $D_{tr}$ , в то время как модель воды OPC совсем незначительно завышает  $D_{tr}$ . Наблюдаемые расхождения в значительной мере обусловлены различием вязкости для трех моделей воды. В отношении неупорядоченного пептида N-H4 полученный результат содержит некоторую информацию о компактности конформационного ансамбля. В частности, полученные данные свидетельствует о том, что использование модели воды TIP4P-Ew ведет к чрезмерно компактному конформационному ансамблю пептида N-H4. Дополнительно, результаты МД моделирования были валидированы при помощи экспериментальных данных по измерению ЯМР релаксации ядер  $^{15}\text{N}$ . В пределах допустимого уровня погрешности результаты моделирования с использованием как модели воды OPC, так и в модели воды TIP4P-D согласуются с данными эксперимента.

Интересно отметить, что противоположный вывод следует из анализа данных МД моделирования при помощи пакета программ HYDRPRO [70] и метода, основанного на параметризации HYDRPRO [71]. Оба подхода говорят, что модель воды TIP4P-Ew обеспечивает лучшее согласие с экспериментом для пептида N-H4. Такой ошибочный вывод предупреждает о том, что необходимо с осторожностью относиться к результатам расчетов при помощи алгоритма HYDRPRO, который в момент создания не задумывался, как инструмент для работы с неупорядоченными белками. Аналогичным образом следует проявлять аккуратность в отношении других эмпирических схем для предсказания  $D_{tr}$  на основе статистических

конформационных ансамблей IDPs [72; 73]. Как показано в настоящей работе, расчет  $D_{tr}$  напрямую из данных МД моделирования может стать полезным инструментом для дальнейших исследований структурных особенностей IDPs.

## 1.2 Результаты и обсуждение

### 1.2.1 Модели воды для моделирования неупорядоченных белков

МД моделирование биологических макромолекул опирается в основном на трехчастичные и четырехчастичные жесткие модели воды с точечными зарядами [74]. Притяжение или отталкивание между молекулами воды задается потенциалом, который включает кулоновские и леннард-джонсовские взаимодействия. Геометрические параметры, величины точечных зарядов и константы леннард-джонсовские взаимодействия подбираются таким образом, чтобы свойства модели соответствовали экспериментально измеряемым характеристикам воды. Например, часто используемая трехчастичная модель TIP3P хорошо соответствует геометрии молекулы воды в газовой фазе, а также дает хорошее приближение для плотности и теплоты испарения жидкости [75]. Для четырехчастичной модели воды TIP4P точка расположения отрицательного заряда смещена от центра атома кислорода. Введение такого фиктивного заряда улучшает согласие между моделью TIP4P и экспериментальными данными нейтронного рассеяния для радиальных функций распределения кислород-кислород и кислород-водород. Позднее параметры модели TIP4P были ре-оптимизированы для метода суммирования Эвальда, который заменил простой подход, основанный на подсчете кулоновских взаимодействий в окрестности ограниченного радиуса [68]. За последние два десятилетия модель воды TIP4P-Ew стала одной из наиболее популярных в задачах моделирования биологических макромолекул наряду с классическими моделями TIP3P и SPC/E [76].

Около пятнадцати лет назад растущий интерес к неупорядоченным белкам и появление ускорителей компьютерных вычислений на графических процессорах стали отправной точкой в моделировании гибких пептидов. Достаточно быстро стало очевидно, что моделирование IDPs терпит неудачу из-за неточных параметров для описания процесса сольватации, т.е. процесса взаимодействия белка с растворителем [16]. В частности, предсказываемый выигрыш энергии при переносе молекулы пептида из вакуума в растворитель (значение свободной энергии сольватации) оказывался меньше измеренного экспериментально. Для того чтобы устранить это расхождение, было предпринято несколько попыток тонкой настройки параметров, описывающих взаимодействия Ван-дер-Ваальса [77; 78]. Усилия в этом направлении привели к разработке усовершенствованной версии силового поля Amber ff99SB [17; 79], которая позднее получила название ff99SB-UCB. Впоследствии для силового поля ff99SB-UCB была подготовлена модифицированная версия модели воды TIP4P-Ew. У этой модели воды параметры для атома кислорода были подобраны таким образом, чтобы улучшить согласие с экспериментальными значениями свободной энергии сольватации для набора коротких пептидов. К сожалению, упомянутая модификация модели воды TIP4P-Ew не получила какого-либо уникального названия.

Вскоре после этого был представлен ещё один вариант модели воды под названием TIP4P-D [18]. Модель TIP4P-D была сконструирована на основе модели TIP4P/2005 [80] путем редактирования параметров Ван-дер-Ваальса для атома кислорода с тем условием, чтобы процесс растворения стал более выигрышным с точки зрения энтальпийного слагаемого в выражении для свободной энергии сольватации. Кроме того, для воды TIP4P-D были скорректированы заряды на атомах, так что было улучшено согласие с экспериментальными значениями теплот испарения и зависимостью плотности от температуры. Сконструированная модель была протестирована в сочетании с несколькими силовыми полями, среди которых силовое поле Amber ff99SB-ILDN [81]. Модель воды TIP4P-D продемонстрировала хорошие результаты в отношении моделирования неупорядоченных белков, вместе с тем в отношении свернутых белков была отмечена тенденция к дестабилизации пространственной структуры.

Два года спустя была представлена комбинация модифицированного варианта модели воды TIP3P и силового поля CHARMM36m, заявленная как универсальная для моделирования как свернутых, так и неупорядоченных белков [43]. При создании новой модели воды были переопределены параметры Ван-дер-Ваальса. Для уточненного варианта модели воды TIP3P авторы не дали нового названия.

Аналогичные усилия были предприняты в области разработки новых вариантов силовых полей общего назначения. Среди них можно выделить силовое поле a99SB-disp, которое подходит для моделирования как свернутых, так и неупорядоченных белков [44]. Силовое поле a99SB-disp используется в совокупности с несколько измененной версией модели воды TIP4P-D. Новая комбинация силового поля и модели воды была тщательным образом протестирована на наборе из 21 белка.

Примерно в то же время, отталкиваясь от абсолютно иных принципов, была сконструирована модель воды OPC. Поиск оптимального расположения точечных зарядов осуществлялся без учета геометрии ковалентных связей, а параметры Ван-дер-Ваальса для атома кислорода были подобраны таким образом, чтобы наилучшим образом соответствовать основным экспериментальным характеристикам воды [69]. Позднее выяснилось, что модель воды OPC позволяет достичь хороших результатов и при моделировании IDPs [45]. В частности в МД траектории для N-концевого участка гистона H4 (26 аминокислотных остатков), записанной с использованием силового поля Amber99SB и модели воды OPC, конформационный ансамбль содержал гораздо больше вытянутых конформаций в сравнении с траекторией в стандартной модели воды TIP3P.

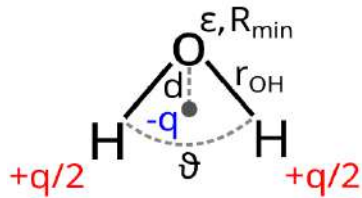
Имеется ряд работ, где проведен сравнительный анализ различных комбинаций силовых полей и моделей воды [14; 49; 82–89]. Эффективность конкретной комбинации оценивалась по соответствию между моделируемыми параметрами и экспериментальными данными, такими как радиус инерции (измерения МУРР), химические сдвиги и константы косвенного диполь-дипольного взаимодействия  $^3J(\text{H}^N, \text{H}^\alpha)$  (ЯМР измерения). Сравнительный анализ показал, что недавно разработанные модели воды в большей мере подходят для моделирования неупорядоченных белков, чем модели воды предыдущего поколения, такие как TIP3P. Сделать более конкретные выводы достаточно сложно, поскольку

выяснилось, что на данный момент отсутствует универсальный набор параметров наилучший для моделирования всех систем, рассмотренных в перечисленных работах. Стоит добавить, что из недавно предложенных вариантов только модель воды TIP4P-D была тщательно исследована в нескольких работах. Для более определенных выводов требуется расширение тестового набора IDPs, включение новых экспериментальных параметров с целью валидации МД моделей и запись более длинных МД траекторий. При этом поиск оптимальных параметров для модели воды необходимо рассматривать как неотъемлемый элемент задачи поиска оптимальных параметров силового поля.

В настоящей работе было исследовано три модели воды, TIP4P-Ew, TIP4P-D и OPC в контексте задачи оценки коэффициентов диффузии по данным МД моделирования для упорядоченных и неупорядоченных белков. Модель воды TIP4P-Ew — это классическая модель, которая была выбрана в качестве отправной точки. Модель воды TIP4P-D — это наиболее популярная специализированная модель, предназначенная для моделирования неупорядоченных белков. И наконец, модель воды OPC — это модель воды последнего поколения, которая подходит как для моделирования свернутых, так и для моделирования неупорядоченных белков. Каждая из этих моделей относится к четырехчастичным моделям воды с фиксированным зарядом на атомах, что упрощает процедуру сравнения (т.к. моделирование требует одинаковых вычислительных затрат). Для удобства параметризация используемых моделей воды показана на Рисунке 1.1.

Для корректного объяснения результатов оценки коэффициента диффузии по данным МД траектории, требуется знание вязкости рассматриваемой модели воды. В нашем исследовании мы оценили коэффициенты самодиффузии  $D_{tr}^{H_2O}$  для моделей воды TIP4P-Ew, TIP4P-D и OPC. Дело в том, что приведенные в литературе коэффициенты самодиффузии для этих моделей воды получены по данным моделирования относительно небольшого количества молекул воды без учета зависимости  $D_{tr}$  от размера ячейки моделирования [65]. Кроме того, опубликованные данные  $D_{tr}^{H_2O}$  доступны лишь для ограниченного набора значений температуры моделирования. Поэтому было принято решение оценить





$$V_{LJ}^{water-water}(r) = \epsilon \left[ \left( \frac{R_{min}}{r} \right)^{12} - 2 \left( \frac{R_{min}}{r} \right)^6 \right]$$

TIP4P-EW		TIP4P-D		OPC	
q	1.0484	q	1.1600	q	1.3583
r <sub>OH</sub>	0.9572 Å	r <sub>OH</sub>	0.9572 Å	r <sub>OH</sub>	0.8724 Å
θ	104.52	θ	104.52	θ	103.60
d	0.1250 Å	d	0.1546 Å	d	0.1594 Å
ε	0.1628 kcal/mol	ε	0.2238 kcal/mol	ε	0.2128 kcal/mol
R <sub>min</sub> /2	1.7759 Å	R <sub>min</sub> /2	1.7763 Å	R <sub>min</sub> /2	1.7771 Å

Рисунок 1.1 — Модели воды TIP4P-Ew, TIP4P-D и OPC. Параметры Ван-дер-Ваальса относятся к атому кислорода, а параметр  $r$  обозначает расстояние между атомами кислорода.

$D_{tr}^{H_2O}$  самостоятельно, используя те же параметры моделирования, что и в исследовании диффузии белковых молекул.

Для каждой модели воды TIP4P-Ew, TIP4P-D и OPC было записано три траектории длиной 150 нс в ячейках моделирования малого, среднего и большого размера. Размер ячейки моделирования был таким же, как в соответствующих траекториях белка Ub (например, для модели воды OPC ячейка большого размера содержала 65243 молекулы воды). Моделирование было выполнено для двух температур 298 и 303 К (в соответствии с МД моделированием N-H4 и Ub). После обработки МД траекторий по протоколу, описанному в разделе 1.3, для каждой модели воды при каждой температуре было получено по три значения  $D_{tr}^{H_2O}$ . На Рисунке 1.2 показана зависимость  $D_{tr}^{H_2O}$  от обратного линейного размера ячейки моделирования,  $1/L$  (цветовая кодировка указана в подписи к рисунку). Линейная экстраполяция  $D_{tr}^{H_2O}$  к пределу ячейки бесконечно большого размера позволяет получить окончательные значения для коэффициента самодиффузии воды, предсказываемые МД. Экстраполированные значения  $D_{(tr,0)}^{H_2O}$  соответствуют точкам пересечения пунктирных линий с осью  $y$ . Эти экстраполированные значения  $D_{(tr,0)}^{H_2O}$  предназначены для сравнения экспериментальными данными (взяты из [90]), которые показаны на рисунке красными горизонтальными линиями. Основные результаты собраны в Таблице 1.

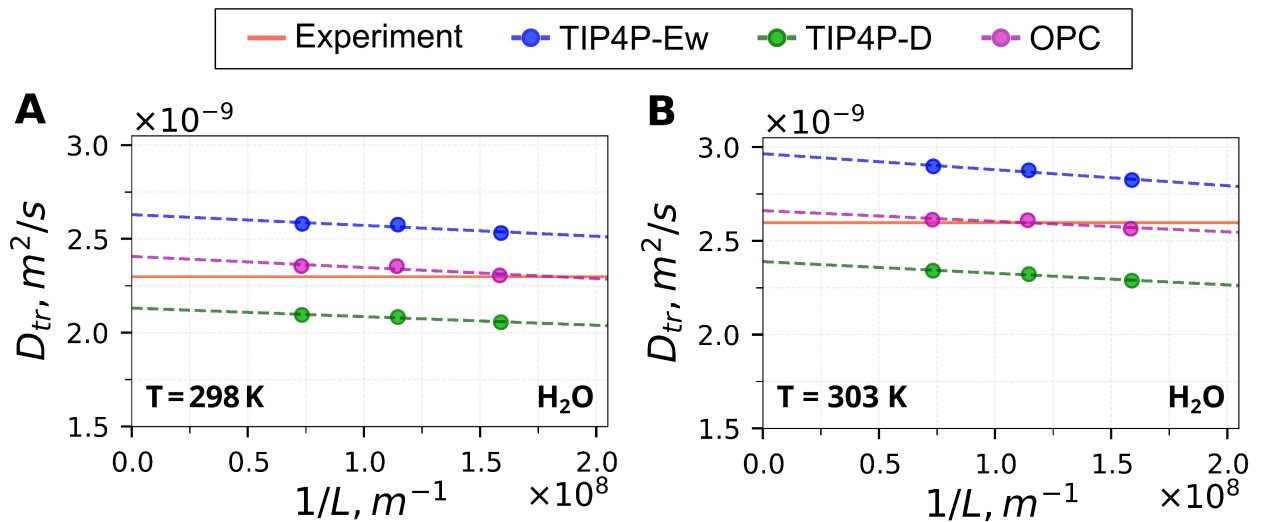


Рисунок 1.2 — Коэффициенты самодиффузии воды,  $D_{tr}^{H_2O}$ , рассчитанные по данным МД моделирования с использованием моделей воды TIP4P-Ew, TIP4P-D и OPC при температуре (А) 298 К и (В) 303 К. Результаты вычислений  $D_{tr}^{H_2O}$  представлены как функция обратного линейного размера ячейки моделирования,  $1/L$ . Методом линейной регрессии осуществлена экстраполяция зависимости к пределу ячейки моделирования бесконечно большого размера (пунктирная линия на рисунке). Точка пересечения регрессионной прямой с осью  $y$  соответствует истинному значению коэффициента самодиффузии,  $D_{(tr,0)}^{H_2O}$ , для соответствующей модели воды (см. Таблицу 1). Экспериментальные данные [90] показаны горизонтальными красными линиями.

Если обратиться к Рисунку 1.2, можно увидеть, что существует хоть и небольшая, но хорошо заметная зависимость значений  $D_{tr}^{H_2O}$  от размера ячейки моделирования. Поэтому если основной целью моделирования является сравнение с результатами эксперимента, то следует использовать именно экстраполированные значения,  $D_{(tr,0)}^{H_2O}$ . Это же утверждение о необходимости использовать  $D_{(tr,0)}^{H_2O}$  справедливо и для обсуждения свойств воды, как растворителя при моделировании белковых молекул. Данные представленные на Рисунке 1.2 и в Таблице 1 свидетельствуют о том, что коэффициент самодиффузии  $D_{(tr,0)}^{H_2O}$  для модели воды TIP4P-Ew сильно завышен, для модели воды TIP4P-D слегка занижен, а для модели воды OPC имеет минимальные отклонения в сторону больших значений по отношению к результатам эксперимента.

Таблица 1 — Коэффициенты самодиффузии воды,  $D_{(tr,0)}^{H_2O}$ , известные из эксперимента и рассчитанные по данным МД моделирования с использованием моделей воды TIP4P-Ew, TIP4P-D и OPC при температуре 298 К или 303 К.

Температура (К)	$D_{tr}^{H_2O}$ ( $10^{-9}$ м <sup>2</sup> /с)			
	Эксперимент	МД моделирование		
		TIP4P-Ew	TIP4P-D	OPC
298	2.30	2.63	2.13	2.41
303	2.60	2.96	2.39	2.60

Показательным оказывается сравнение полученных результатов с литературными данными. По литературным данным для модели воды TIP4P-Ew коэффициент самодиффузии при температуре 298 К равен  $2.4 \cdot 10^{-9}$  м<sup>2</sup>/с [68], что значительно ниже значения полученного в настоящей работе  $2.63 \cdot 10^{-9}$  м<sup>2</sup>/с. Причина расхождения в том, что в опубликованные ранее значения были получены из данных моделирования небольшого количества молекул воды (512 молекул) с использованием термостата Андерсена [91], который не подходит для моделирования диффузионных характеристик системы [61]. Схожее наблюдение в отношении модели воды TIP4P-Ew было сделано исследовательской группой Экономоу, которая сообщила, что скорректированное на размер ячейки значение коэффициента диффузии составляет  $2.7 \cdot 10^{-9}$  м<sup>2</sup>/с [65].

Для модели воды OPC коэффициент самодиффузии при 298 К, как это следует из литературы, равен  $2.3 \cdot 10^{-9}$  м<sup>2</sup>/с [69], что немного ниже значения, полученного нами  $2.41 \cdot 10^{-9}$  м<sup>2</sup>/с. Здесь необходимо иметь в виду, что в своем моделировании авторы использовали небольшое количество молекул воды (804 молекулы) и алгоритм термостатирования Ланжевена, который приводит к искажению динамических характеристик системы (подробно обсуждается во введении).

И наконец, для модели воды TIP4P-D из литературных данных известен коэффициент самодиффузии при температуре 300 К, равный  $2.1 \cdot 10^{-9}$  м<sup>2</sup>/с. Это значение немного ниже значений  $2.13 \cdot 10^{-9}$  и  $2.39 \cdot 10^{-9}$  м<sup>2</sup>/с, которые были получены в рамках описываемого исследования при температуре 298 и 303 К соответственно. Следует отметить, что

авторы использовали 3054 молекулы воды и алгоритм термостатирования Ноэ-Хувера [91; 92], который не приводит к искажению динамических характеристик системы.

### 1.2.2 МД моделирование N-концевого фрагмента гистона H4

Для N-концевого фрагмента гистона H4 (N-H4) был записан большой набор МД траекторий суммарной длиной 75 мкс (см. подробности в Таблице 4). Пептид N-H4 условно может быть разделен на два фрагмента, а именно: фрагмент, содержащий большое количество остатков глицина (аминокислотные остатки 1-15), и фрагмент, содержащий положительно заряженные остатки (аминокислотные остатки 16-25). В первых работах, где исследовался этот пептид, сообщалось о склонности N-H4, в особенности его ацетилированной формы, к образованию  $\alpha$ -спиралей и формированию  $\beta$ -листов [93–96]. Однако в более поздних экспериментальных и вычислительных работах присутствие остаточной структуры не подтвердилось как в случае изолированного пептида N-H4, так и в случае N-концевого хвоста гистона H4 в составе нуклеосомной частицы (от англ. nucleosome core particle, NCP) [49; 97–102]. В целом, эксперименты ЯМР выявили некоторые любопытные аспекты поведения пептида N-H4 в составе нуклеосомы, а именно: динамическое («нечеткое») взаимодействие хвоста гистона H4 с нуклеосомной ДНК, а также конкурентное связывание хвостом гистона H3 и хвостом гистона H4 участков на поверхности нуклеосомной ДНК [50; 100–102].

В МД траекториях, записанных в каждой из моделей воды TIP4P-Ew, TIP4P-D и OPC, динамика пептида N-H4 связана с конформационными переходами между большим числом случайных состояний. Тем не менее, для моделирования в воде TIP4P-Ew можно заметить склонность пептида N-H4 к образованию структур, похожих на шпильки. В этих траекториях наблюдаются стабильные структурные состояния, для которых пространственная конформация, хоть и претерпевает постоянные изменения, но остается узнаваемой в течение нескольких сотен наносекунд. Иногда пептидная цепь формирует так называемый «мини-фолд», который

не имеет вторичной структуры, но также сохраняет свою топологию в течение несколько сотен наносекунд. Подобного рода подвижные структурные мотивы удерживаются небольшим числом водородных связей между основной и боковыми цепями белка по типу: основная-основная, основная-боковая и боковая-боковая. С другой стороны, в МД траекториях в воде TIP4P-D пептид N-H4 стремится принять вытянутую конформацию. Формирование шпилек происходит реже и носит более локальный характер. Среди таких событий заслуживает внимание образование небольшой шпильки на С-концевой части пептида (см. работу [49]). И наконец, структурное состояние пептида N-H4 в МД траектории в воде OPC можно охарактеризовать, как промежуточное между поведением пептида в воде TIP4P-Ew и в воде TIP4P-D, поскольку конформационный ансамбль представляет собой смесь частично свернутых и вытянутых конформаций.

Коэффициент трансляционной диффузии чувствителен к виду конформационного ансамбля, который различается для трех моделей воды TIP4P-D, TIP4P-Ew и OPC. Компактность конформационного ансамбля можно выразить при помощи функции распределения по радиусу инерции,  $P(R_g)$ . Соответствующие распределения  $P(R_g)$  по результатам моделирования пептида N-H4 в моделях воды TIP4P-Ew, TIP4P-D и OPC показаны на Рисунке 1.3 (синяя, зеленая и пурпурная гистограммы). Все три распределения бимодальные. Для воды TIP4P-Ew в распределении преобладает узкий пик чуть ниже отметки в  $10 \text{ \AA}$ , а второй более широкий пик находится на отметке в  $13 \text{ \AA}$ . Для воды TIP4P-D основной широкий пик сосредоточен примерно на  $16 \text{ \AA}$ , в то время как второстепенная компонента проглядывается на  $11 \text{ \AA}$ . И наконец, для воды OPC заметен широкий пик с большей площадью под кривой на отметке около  $15 \text{ \AA}$ , в то время как второй более узкий пик расположен примерно на отметке  $11 \text{ \AA}$ . Средние значения радиуса инерции  $R_g$  составляют  $11.5$ ,  $13.8$  и  $15.3 \text{ \AA}$  для траекторий TIP4P-Ew, OPC и TIP4P-D соответственно. В совокупности эти данные указывают на то, что результаты моделирования в воде TIP4P-Ew и воде TIP4P-D сильно отличаются друг от друга, а результаты моделирования в воде OPC несколько ближе к воде TIP4P-D, чем к воде TIP4P-Ew.

Результаты, показанные на Рисунке 1.3, имеют вполне ожидаемый характер. Как было отмечено ранее, при моделировании в воде TIP4P-Ew энергетически более выгодными оказываются взаимодействия белок-белок

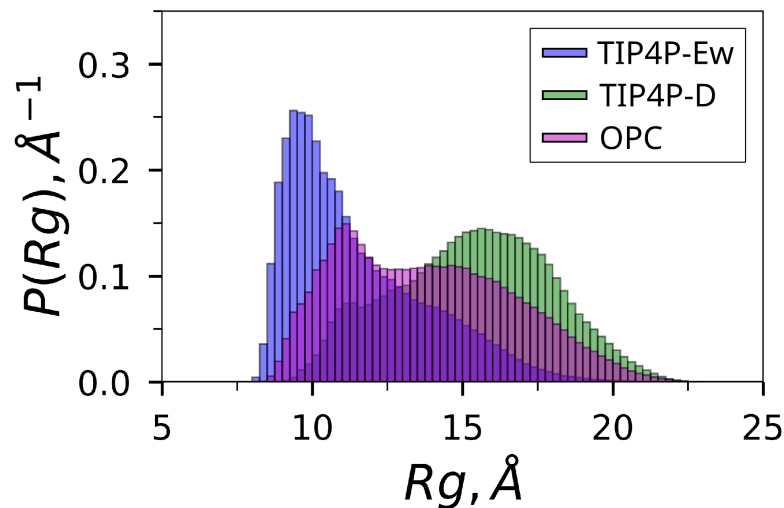


Рисунок 1.3 — Распределения по радиусам инерции  $P(R_g)$  для пептида N-H4 по данным МД траекторий, записанным с использованием моделей воды TIP4P-Ew, TIP4P-D и OPC (моделирование в ячейке малого размера в ансамбле NPT). Соответствующие распределения для МД траекторий в ячейках среднего и большого размеров почти не отличаются от представленных на настоящем рисунке (см. Рисунок 1.12 (B, C)). Подобный результат ожидаем ввиду того, что МД моделирование в ячейках среднего и большого размера было осуществлено в виде набора коротких траекторий длиной 10 нс, начальные конформации для которых были взяты из кадров МД траектории в ячейке малого размера (см. Рисунок 1.8). Распределения  $P(R_g)$  в ансамблях NVE и NPT для МД траекторий в модели воды TIP4P-D практически неотличимы, в то время как для модели воды TIP4P-Ew существуют небольшие, но тем не менее заметные различия (см. Рисунок 1.12 (D-F)). Подобное поведение объясняется неидеальной статистической сходимостью, а именно: для модели воды TIP4P-D характерны более вытянутые конформации для гибких пептидов, а для модели воды TIP4P-Ew заметна склонность к формированию частично упорядоченных структурных мотивов в форму шпильки.

в сравнении с взаимодействиями белок-вода, что ведет к увеличению вероятности появления более компактных конформаций пептида N-H4. В то же время, при моделировании в воде TIP4P-D баланс смещен в противоположную сторону: более благоприятными становятся вытянутые конформации с высоким уровнем сольватации. В контексте данной работы примечательно, что конформационные ансамбли пептида N-H4 оказываются различными при моделировании в трех моделях воды

TIP4P-Ew, TIP4P-D и OPC. Ключевой вопрос заключается в том, можно ли с помощью экспериментальных данных для коэффициента трансляционной диффузии  $D_{tr}$  отличить три конформационных ансамбля. Ответ на этот вопрос будет рассмотрен в оставшейся части данной главы.

В качестве завершения данного раздела, будут приведены результаты моделирования белка Ub (краткая информация о записанных МД траекториях дана в Таблице 5). Теоретически можно представить, что молекула Ub становится чуть более компактной в воде TIP4P-Ew и чуть более увеличенной в воде TIP4P-D. На практике подобное явление может произойти вследствие «залипания» боковых цепей аминокислотных остатков на поверхности белка в воде TIP4P-Ew, и обращение боковых цепей наружу к растворителю в воде TIP4P-D. В действительности предполагаемый эффект оказался незначительным. Так, средние значения радиусов инерции  $R_g$  для убиквитина практически совпадают 11.69, 11.70 и 11.81 Å при моделировании в моделях воды TIP4P-Ew, OPC и TIP4P-D (данные относятся к моделированию в ансамбле NPT с использованием ячейки малого размера). Таким образом, можно утверждать, что структурное состояние убиквитина не зависит от выбора модели воды.

### 1.2.3 Расчет коэффициентов диффузии белков по данным МД моделирования

Согласно протоколу, описанному в разделе 1.3, из МД траекторий для пептида N-H4 были вычислены значения коэффициентов трансляционной диффузии,  $D_{tr}^{N-H4}$ , в зависимости от размера ячейки моделирования для трех моделей воды TIP4P-Ew, TIP4P-D и OPC. Полученную зависимость экстраполировали к пределу бесконечно большой ячейки моделирования. Полученный таким образом коэффициент диффузии,  $D_{(tr,0)}^{(N-H4)}$ , считается истинным значением для конкретной модели воды. На Рисунке 1.4 А проиллюстрирована процедура экстраполяции в случае МД траекторий для пептида N-H4 (суммарная длина траекторий в NPT ансамбле 45 мкс). Значения  $D_{tr}^{N-H4}$  для каждой ячейки моделирования (цвет кружков на графике соответствует определенной модели воды) хорошо

ложатся на прямую линию, что указывает на хорошую статистическую сходимость и служит подтверждением корректности выбранной схемы для устранения зависимости коэффициента диффузии от размера ячейки моделирования [59]. При дальнейшем рассмотрении результатов можно заметить, что коэффициент диффузии  $D_{(tr,0)}^{(N-H4)}$  в воде TIP4P-Ew в значительной степени завышен относительно экспериментального значения (см. на положение точки пересечения пунктирной линии с осью  $y$  и на сплошную красную линию на Рисунке 1.4 (А)). Вместе с тем, коэффициент диффузии  $D_{(tr,0)}^{(N-H4)}$  умеренно занижен в воде TIP4P-D (зеленый цвет) и слегка завышен воде OPC (пурпурный цвет) в сравнении с экспериментальными данными. Наиболее близкое к экспериментальному значение наблюдается для модели воды OPC, см. Таблицу 2.

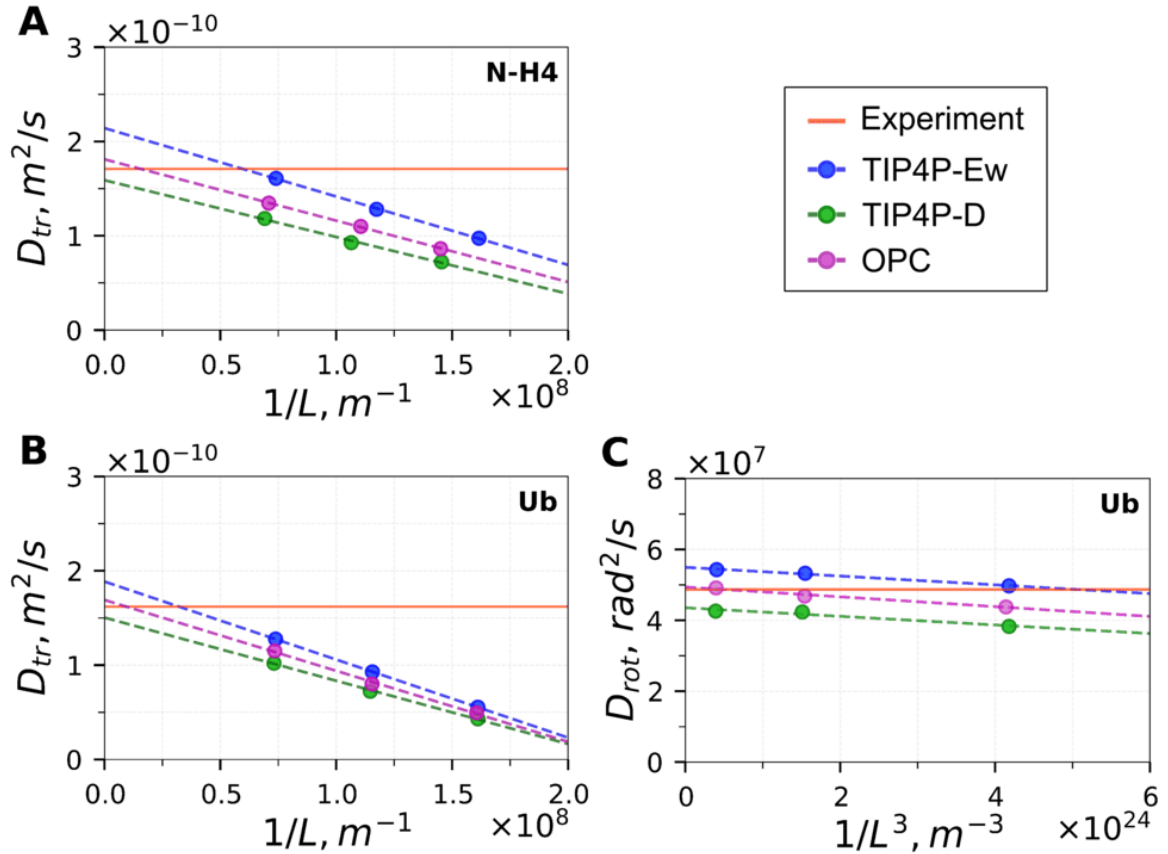


Рисунок 1.4 — Коэффициенты диффузии (А) пептида N-H4 и (В, С) белка Ub, рассчитанные по данным МД моделирования с использованием моделей воды TIP4P-Ew, TIP4P-D и OPC. Моделирование проводилось в ансамбле NPT при температуре 298 К (для N-H4) или 303 К (для Ub). Результаты вычислений  $D_{tr}$  представлены как функция обратного линейного размера ячейки моделирования,  $1/L$ . Методом линейной регрессии осуществлена экстраполяция зависимости к пределу ячейки



моделирования бесконечно большого размера (пунктирная линия на рисунке). Точка пересечения регрессионной прямой с осью  $y$  соответствует истинному значению коэффициента трансляционной диффузии,  $D_{(tr,0)}^{(N-H4)}$  или  $D_{(tr,0)}^{Ub}$  для конкретной модели воды. Результаты расчетов  $D_{rot}$  представлены, как функция меры обратного объема ячейки моделирования,  $1/L^3$ . Аналогичным образом методом линейной регрессии осуществлена экстраполяция зависимости к пределу ячейки моделирования бесконечно большого размера (пунктирная линия на рисунке). Точка пересечения регрессионной прямой с осью  $y$  соответствует истинному значению коэффициента вращательной диффузии,  $D_{(rot,0)}^{Ub}$  для конкретной модели воды. Можно заметить, что точка пересечения с осью  $y$  для функции  $D_{tr}$  от  $1/L$  определяется более однозначным образом, чем для функции  $D_{rot}$  от  $1/L^3$ . В частности, наблюдается небольшое отклонение значений  $D_{rot}$  от линейной зависимости в ячейке моделирования среднего размера (средняя точка на панели (С)). Более подробный анализ, основанный на сравнении результатов моделирования в ансамблях NPT и NVE, а также оценку погрешности по методу «складного ножа» (от англ. jackknife), не выявил значимых отличий в точности для расчетных значений  $D_{tr}$  и  $D_{rot}$ . Экспериментальные данные показаны горизонтальными красными линиями.

Аналогичным образом по данным МД моделирования для белка Ub были вычислены значения коэффициентов трансляционной,  $D_{tr}^{Ub}$ , и вращательной диффузии,  $D_{rot}^{Ub}$  (суммарная длина траекторий в NPT ансамбле составила 18 мкс). В целом результаты, показанные на Рисунке 1.4 (В, С), имеют ту же направленность, что и для пептида N-H4, а именно: полученная оценка для коэффициентов диффузии слегка завышена для воды TIP4P-Ew, слегка занижена для воды TIP4P-D и наиболее близка к экспериментальному значению для воды OPC. Интерпретация полученных результатов сопряжена с необходимостью учитывать действие следующих факторов. Во-первых, существует различие в вязкости для разных моделей воды, см. Рисунок 1.2 и Таблицу 1. Во-вторых, существует различие в структуре конформационного ансамбля пептида N-H4 для разных моделей воды, см. Рисунок 1.3. Влияние этих факторов на результирующие значения  $D_{tr}$  и  $D_{rot}$  будет проанализировано в следующем разделе.

В дополнение к NPT ансамблю, также была записана серия траекторий в ансамбле NVE для двух моделей воды TIP4P-Ew и TIP4P-D с суммарной длиной 30 мкс для пептида N-H4 и 12 мкс для белка Ub. Этапы обработки и анализа данных для моделирования в ансамбле NVE не отличались от описанных выше для ансамбля NPT. Набор независимых данных для  $D_{tr}$  и  $D_{rot}$  позволяет оценить статистическую сходимость для выбранного протокола расчета.

Сравнение результатов моделирования в ансамбле NPT и NVE представлено на Рисунке 1.13 и в Таблице 6. Хороший уровень согласия подтверждает тот факт, что термостат Бусси-Парринелло является оправданным выбором в задачах моделирования диффузионных свойств белковых молекул. Расхождение между результатами моделирования в ансамблях NPT и NVE не превышает 2.0 % по модулю, что сопоставимо с уровнем экспериментальной погрешности. Так, для коэффициента трансляционной диффузии белка Ub экспериментальная погрешность составляет 2.5 % (см. Таблицу 6). Таким образом, точность расчетов находится на уровне, достаточном для адекватного сравнения с данными эксперимента. В частности, можно считать, что различия в несколько процентов между расчетом и экспериментом не являются значимыми.

#### 1.2.4 Сравнение экспериментальных значений коэффициентов диффузии белков с данными МД моделирования

На Рисунке 1.5 в процентном отношении показано расхождение между экспериментальными и расчетными значениями коэффициентов диффузии для воды, убиквитина и пептида N-H4. Удобнее начать обсуждение результатов с упорядоченного белка убиквитина, для которого конформационное состояние строго определено, а скорость диффузии зависит только от вязкости растворителя. Данные о вязкости растворителя в виде  $D_{tr}^{H_2O}$  при температуре 303 К представлены на Рисунке 1.5 (А). Мы сопоставим их с данными для коэффициента трансляционной диффузии убиквитина,  $D_{tr}^{Ub}$  (см. Рисунок 1.5 (В)).

Таблица 2 — Экспериментальные и расчетные значения коэффициентов диффузии пептида N-H4 и белка Ub. Расчетные значения из данных МД моделирования были получены путем анализа кривой MSD( $\tau$ ) с учетом зависимости от размера ячейки моделирования см. Рисунок 1.4. В таблице также содержатся результаты для четырех эмпирических схем расчета: HYDROPRO, Nygaard, Kirkwood-Riseman и HullRadSAS. Экспериментальное значение  $D_{tr}^{N-H4}$  и  $D_{tr}^{Ub}$  получены в рамках описываемого исследования, значение  $D_{rot}^{Ub}$  взято из литературы [103]. Экспериментальные значения скорректированы к значениям вязкости и температуры в МД моделировании.

Коэффициент диффузии	Эксперимент	МД моделирование														
		TIP4P-Ew				TIP4P-D				OPC						
		МД	HYDROPRO	Nygaard	Kirkwood-Riseman	HullRadSAS	МД	HYDROPRO	Nygaard	Kirkwood-Riseman	HullRadSAS	МД	HYDROPRO	Nygaard	Kirkwood-Riseman	HullRadSAS
$D_{tr}^{N-H4}$ ( $10^{-10}$ m <sup>2</sup> /c)	1.71	2.14	1.75	1.72	2.34	2.05	1.59	1.56	1.62	1.96	1.8	1.81	1.63	1.66	2.11	1.9
$D_{tr}^{Ub}$ ( $10^{-10}$ m <sup>2</sup> /c)	1.62	1.89	1.61	1.66	—	1.75	1.5	1.6	1.65	—	1.74	1.69	1.61	1.66	—	1.75
$D_{rot}^{Ub}$ ( $10^7$ rad <sup>2</sup> /c)	4.87	5.50	4	—	—	4.14	4.35	3.87	—	—	4.05	4.94	3.98	—	—	4.13

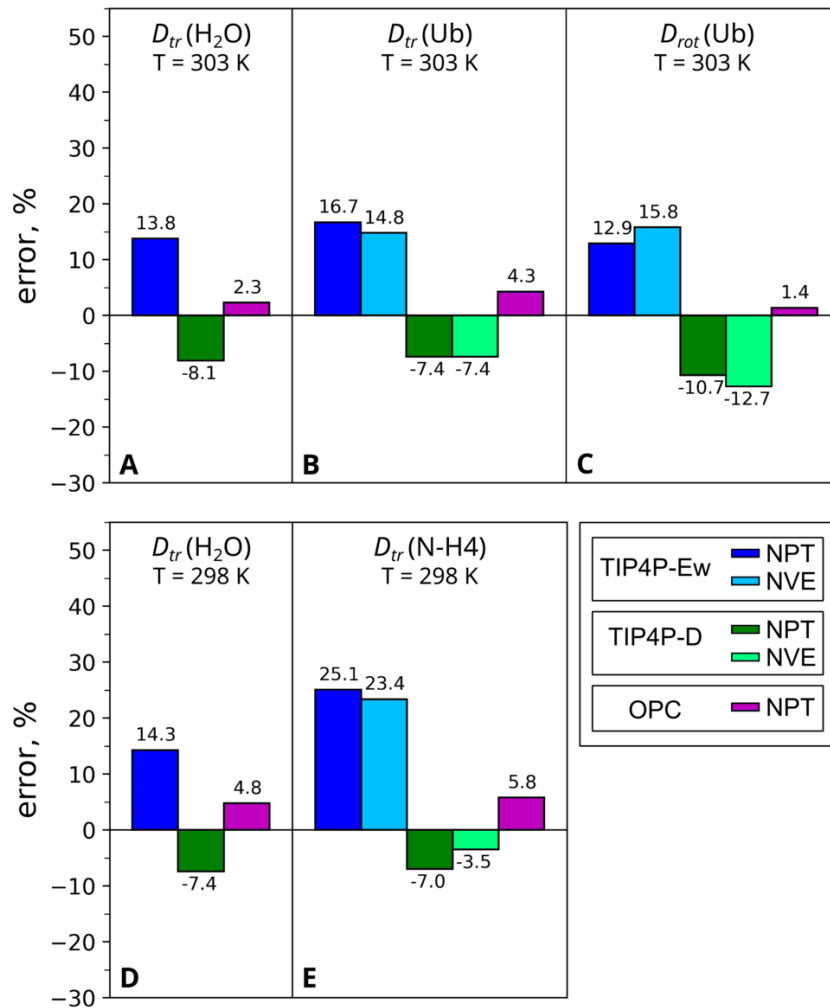


Рисунок 1.5 — Расхождение в процентном отношении между экспериментальными и расчетными значениями коэффициентов диффузии для воды, белка убиквитина (Ub) и пептида N-H4 (на основе данных из Таблиц 1, 2 и 6). Основные параметры в эксперименте и МД моделировании указаны непосредственно на рисунке. Необходимо отметить, что в процентном отношении различия между экспериментом и данными МД моделирования для коэффициентов диффузии находятся приблизительно на одном и том же уровне для температурных условий 298 К и 303 К (см. панели (A, D)).

Модель воды TIP4P-Ew переоценивает значение экспериментального коэффициента самодиффузии воды  $D_{tr}^{\text{H}_2\text{O}}$  на 13.8 % (синий столбец на Рисунке 1.5 (A)). Коэффициент трансляционной диффузии  $D_{tr}^{\text{Ub}}$  в модели воды TIP4P-Ew также завышен относительно экспериментальных значений на 16.7 % или 14.8 % для МД моделирования в ансамбле NPT или NVE (синий или голубой столбец на Рисунке 1.5 (B)).

Модель воды TIP4P-D недооценивает значение экспериментального коэффициента самодиффузии воды  $D_{tr}^{H2O}$  на 8.1 % (зеленый столбец на Рисунке 1.5 (A)). Коэффициент трансляционной диффузии  $D_{tr}^{Ub}$  в модели воды TIP4P-D также занижен относительно экспериментальных значений на 7.4 % для МД моделирования в ансамблях NPT и NVE (темно-зеленый и светло-зеленый столбец на Рисунке 1.5 (B)).

Модель воды OPC переоценивает значение экспериментального коэффициента самодиффузии воды  $D_{tr}^{H2O}$  на 2.3 % (пурпурный столбец на Рисунке 1.5 (A)). Коэффициент трансляционной диффузии  $D_{tr}^{Ub}$  в модели воды OPC завышен относительно экспериментальных значений на 4.3 % для МД моделирования в ансамбле NPT (пурпурный столбец на Рисунке 1.5 (B)).

Таким образом, сравнивая результаты на Рисунках 1.5 (A, B), можно увидеть, что расхождения находятся в пределах нескольких процентов. Это означает, что различия в расчетных значениях коэффициентов трансляционной диффузии для белка Ub в разных моделях воды объясняются различием в вязкости растворителя. Отсюда следует общая рекомендация, что для успешного моделирования трансляционной диффузии свернутого белка достаточно выбрать модель воды с корректным значением вязкости (например, модель воды OPC) или в случае других моделей воды осуществить коррекцию экспериментальных значений к истинному значению вязкости (требуется для моделей воды TIP4P-Ew и TIP4P-D).

Сделанные выводы подтверждают результаты для коэффициентов вращательной диффузии убиквитина  $D_{rot}^{Ub}$  см. Рисунок 1.5 (C). Расхождение экспериментальных и расчетных значений  $D_{rot}^{Ub}$  может быть объяснено различием в вязкости растворителя. В частности, модель воды OPC несколько переоценивает значения экспериментального коэффициента вращательной диффузии  $D_{rot}^{H2O}$  на 1.4 % (пурпурный столбец на Рисунке 1.5 (C)). С другой стороны, для моделей воды TIP4P-Ew или TIP4P-D достаточно применить простую коррекцию в отношении вязкости, чтобы получить точный результат для  $D_{rot}^{Ub}$ .

Кроме того, на Рисунке 1.5 высота столбцов на панели (A) практически совпадает с высотой столбцов на панелях (B, C). Незначительные отклонения объясняются погрешностью эксперимента

и погрешностью МД расчетов (см. зеленые и светло-зеленые столбцы для результатов моделирования в NPT и NVE ансамбле на Рисунке 1.5 (С)). Как уже было отмечено (см. предыдущий раздел), отклонение на несколько процентов экспериментальных и расчетных значений можно считать несущественным.

Далее перейдем к обсуждению результатов для пептида N-H4, который представляет основной интерес в контексте данной работы. Поскольку пептид N-H4 неупорядочен в растворе, точность оценок коэффициента диффузии по данным МД моделирования будет зависеть от двух факторов: от вязкости моделируемого растворителя (данный фактор уже обсуждался на примере убиквитина) и непосредственно от структуры конформационного ансамбля (см. Рисунок 1.3).

На Рисунке 1.5 в процентном отношении показано расхождение между экспериментальными и расчетными значениями коэффициентов диффузии для воды и пептида N-H4 при температуре 298 К (панели (D, E) соответственно). Сравнивая значения  $D_{tr}^{H_2O}$  и  $D_{tr}^{N-H4}$ , «невооруженным глазом» можно заметить существенное отклонение для модели воды TIP4P-Ew. С одной стороны, модель воды TIP4P-Ew переоценивает значения экспериментального коэффициента самодиффузии воды  $D_{tr}^{H_2O}$  на 14.3 % (синий столбец на Рисунке 1.5 (D)). С другой стороны, для этой модели воды отклонение коэффициента трансляционной диффузии  $D_{tr}^{N-H4}$  от экспериментального значения существенно выше: 25.1 % или 23.4 % в случае моделирования в ансамбле NPT или NVE (синий или голубой столбец на Рисунке 1.5 (E)). Зафиксированное расхождение слишком велико, чтобы его можно было объяснить исключительно ошибками эксперимента или погрешностью в МД моделировании. Причина расхождений заключается в структуре конформационного ансамбля N-H4 пептида.

Обратимся вновь к Рисунку 1.3. Из соответствующего распределения по радиусу инерции можно заметить, что пептид N-H4 в модели воды TIP4P-Ew имеет склонность принимать компактные конформации. Такие компактные частицы перемещаются быстрее, что и приводит к более высокому, чем это ожидалось, значению коэффициента диффузии. Иными словами, наш анализ  $D_{tr}$  выявил непригодность модели воды TIP4P-Ew для моделирования конформационного ансамбля N-H4 пептида.

Дополнительные доказательства этого факта приведены в обсуждении данных по скоростям спиновой релаксации.

Возвращаясь к обсуждению Рисунков 1.5 (D, E), следует отметить, что для модели воды TIP4P-D и модели воды OPC данные по коэффициентам самодиффузии воды совместимы с данными для коэффициентов диффузии N-H4 пептида. Модель воды TIP4P-D недооценивает значение экспериментального коэффициента самодиффузии воды  $D_{tr}^{H_2O}$  на 7.4 % (зеленый столбец на Рисунке 1.5 (E)). Коэффициент трансляционной диффузии  $D_{tr}^{N-H_4}$  при моделировании в воде TIP4P-D занижен относительно экспериментальных значений на 7.0 % или 3.5 % в случае моделирования в ансамбле NPT или ансамбле NVE (темно-зеленый и светло-зеленый столбец на Рисунке 1.5 (E)). Наконец, модель воды OPC переоценивает значения реального коэффициента самодиффузии воды  $D_{tr}^{H_2O}$  на 4.8 % (пурпурный столбец на Рисунке 1.5 (D)). А коэффициент трансляционной диффузии  $D_{tr}^{N-H_4}$  в модели воды OPC завышен относительно экспериментальных значений на 5.8 % в случае моделирования в ансамбле NPT (пурпурный столбец на Рисунке 1.5 (E)).

По-видимому, обе модели воды TIP4P-D и OPC достаточно успешно справляются с воспроизведением конформационного ансамбля пептида N-H4. Можно сделать предположение, что истинное распределение по радиусу инерции  $P(R_g)$  для пептида N-H4 находится где-то посередине между оценками TIP4P-D и OPC (см. Рисунок 1.3). Различия в предсказании коэффициента диффузии пептида N-H4 моделями TIP4P-D и OPC можно объяснить различиями в вязкости растворителя.

Для того чтобы усилить сделанные наблюдения, обратимся к данным по  $^{15}\text{N}$  скоростям релаксации ЯМР [49]. В ЯМР эксперименте для каждого аминокислотного остатка пептида N-H4 были измерены скорости продольной релаксации,  $R_1$ , и скорости поперечной кросс-корреляционной релаксации,  $\eta$  (между дипольным механизмом и механизмом через анизотропию химического сдвига). Кроме того, значения  $R_1$  и  $\eta$  для каждого аминокислотного остатка N-H4 пептида были рассчитаны по данным из набора записанных МД траекторий (см. 1.3). Сравнение экспериментальных и расчетных значений для  $R_1$  и  $\eta$  показано на Рисунке 1.6.

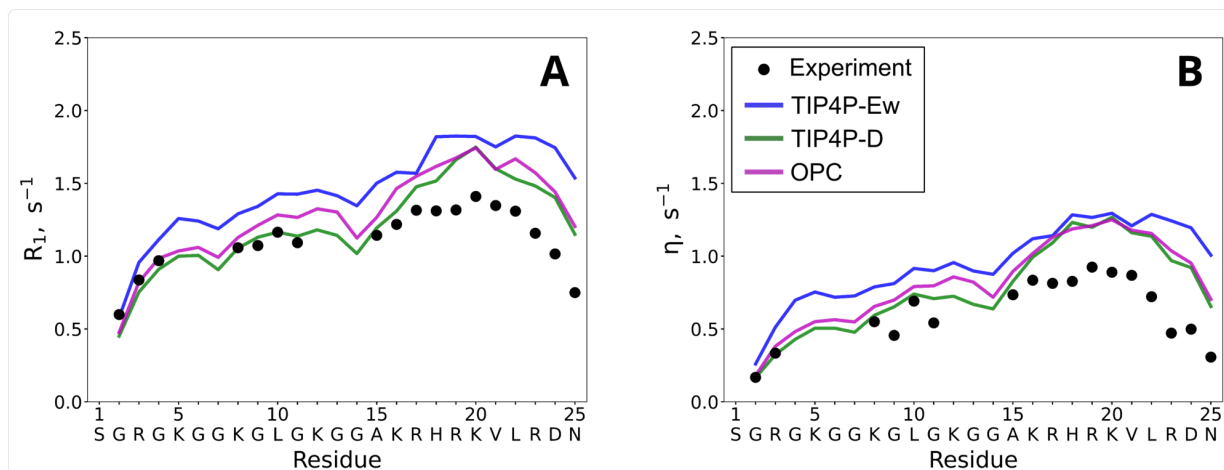


Рисунок 1.6 — Экспериментальные и расчетные значения  $^{15}\text{N}$  скоростей релаксации для пептида N-H4: (A)  $R_1$  и (B)  $\eta$ . Обозначения приведены в подписи к рисунку. Расчетные значения основаны на данных МД моделирования в NPT ансамбле в ячейке малого размера. Анализ данных показал, что расчетные значения скоростей релаксации практически не зависят от размера ячейки моделирования (см. Рисунок 1.14), поскольку спиновая релаксация в пептиде N-H4 в большей мере объясняется интенсивной конформационной динамикой, чем общим вращательным движением.

Особый интерес представляют скорости поперечной кросс-релаксации,  $\eta$ , см. Рисунок 1.6 (B). В отличие от продольной релаксации, поперечная кросс-релаксация имеет более наглядную интерпретацию, т.к. демонстрирует более простую зависимость от характерного времени движения (замедление динамики белка приводит к увеличению  $\eta$ ) [104]. Обсуждение результатов для  $^{15}\text{N}$  скоростей релаксации удобнее начать с модели воды TIP4P-Ew. Ранее было показано, что вязкость воды TIP4P-Ew существенно ниже, чем следует из эксперимента, см. Таблицу 1. Отсюда можно ожидать, что локальная динамика пептида N-H4 в воде TIP4P-Ew окажется быстрее, чем в реальности. А это означает, что полученные из данных МД моделирования значения скоростей кросс-релаксации  $\eta$  ожидаются заниженными относительно эксперимента. Однако эти предположения не подтверждаются фактическими результатами — в действительности, расчетные значения  $\eta$  существенно выше экспериментальных значений в случае модели воды



TIP4P-Ew (синяя кривая для модели воды TIP4P-Ew и черные кружки для экспериментальных значений на Рисунке 1.6 (B)).

По-видимому, результаты, представленные на Рисунке 1.6 (B), зависят от тонкой структуры конформационного ансамбля N-H4. Как было отмечено ранее, для МД моделирования в воде TIP4P-Ew наблюдается тенденция к образованию компактных структурных состояний (удерживаются за счет образования водородных связей), см. раздел 1.3.2). Медленная конформационная динамика для таких компактных структурных состояний приводит к более высоким значениям скоростей релаксации. Эффект оказывается достаточно сильным даже на фоне низкой вязкости модели воды TIP4P-Ew, в результате чего расчетные значения превышают экспериментальные значения скоростей релаксации в 1.5 раза (см. Рисунок 1.6 (B)). Таким образом, данные по скоростям релаксации совместно с данными о коэффициенте диффузии указывают на то, что конформационный ансамбль пептида N-H4 оказывается слишком компактным при моделировании в воде TIP4P-Ew. Полученный результат согласуется с выводами предыдущих работ о том, что классические модели воды, как правило, приводят к излишне компактным моделям неупорядоченных белков [16].

Возвращаясь к обсуждению Рисунка 1.6, следует подчеркнуть, что как моделирование в воде TIP4P-D, так и моделирование в воде OPC обеспечивает более хорошее согласие с экспериментальными значениями скоростей релаксации для пептида N-H4 в сравнении с моделью воды TIP4P-Ew. Строго говоря, обе модели достаточно точно оценивают значения скоростей релаксации для более неупорядоченной N-концевой части пептида N-H4, но переоценивают скорости релаксации для C-концевой части, см. Рисунок 1.6 (зеленый профиль — для TIP4P-D, синий профиль — для TIP4P-Ew, черные кружки — экспериментальные значения). Небольшая разница в оценке скоростей релаксации для моделей воды TIP4P-D и OPC не позволяет отдать предпочтение той или иной модели воды (т.е. невозможно указать на модель, которая обеспечивает более реалистичное описание конформационного ансамбля N-H4, см. Рисунок 1.3). И хотя анализ скоростей релаксации указывает на проблемы с моделированием в воде TIP4P-Ew, в отношении моделей воды TIP4P-D и

ОРС сделать однозначный выбор не представляется возможным, требуются дополнительные исследования.

### 1.2.5 Эмпирические методы оценки коэффициента диффузии белков

В настоящей работе был предложен протокол для вычисления коэффициента трансляционной диффузии напрямую из данных МД моделирования белковых молекул. Более простой алгоритм реализован в программе HYDROPRO [70], в которой расчет гидродинамических свойств макромолекул осуществляется на основе координат их атомов. Существует ряд работ, в которых программа HYDROPRO была использована для оценки гидродинамического радиуса,  $R_h$ , и коэффициента трансляционной диффузии,  $D_{tr}$ , неупорядоченных белков. В частности был предложен подход, в котором алгоритм HYDROPRO применяется к индивидуальным кадрам из МД траектории неупорядоченного белка, после чего результаты усредняются для получения итоговых значений  $R_h$  и  $D_{tr}$  [82; 105–108]. Описанный протокол также был опробован в отношении конформационных ансамблей IDPs, сконструированных при помощи алгоритмов, отличных от молекулярной динамики [109–111]. Кроме того, недавно Нюгором (от англ. Nygaard) был предложен ещё один вариант эмпирической параметризации для выражения, устанавливающего взаимосвязь между средним радиусом инерции  $R_g$  и гидродинамическим радиусом  $R_h$  для неупорядоченных белков [71]. В нескольких работах параметризация Nygaard была протестирована в контексте задачи предсказания коэффициента трансляционной диффузии IDPs белков по данным МД моделирования [88; 89].

Строго говоря, оценка коэффициента  $D_{tr}$  неупорядоченного белка при помощи HYDROPRO не подкреплена какой-либо прочной теоретической базой. В руководстве пользователя HYDROPRO указано, что метод предназначен для расчета гидродинамических характеристик жестких макромолекул [70]; расчеты гидродинамических характеристик гибких макромолекул требуют привлечения других подходов [112; 113].

По сути, в алгоритме HYDROPRO заложено допущение о том, что все конформеры в ансамбле IDPs, когда диффундируют в растворе, находятся в «замороженном состоянии». В реальности диффузия гибкого пептида сопряжена с непрерывным переходом из одной конформации в другую. Можно предположить, что диффузия неподвижных «замороженных» конформеров должна быть в среднем медленнее, чем для гибкого пептида. Следует отметить, что эта гипотеза была проверена и оказалась справедливой при моделировании цепочек простых полимеров [114; 115].

В настоящей работе для проверки этого предположения, была разработана специальная схема МД моделирования пептида N-H4. Для этого была записана серия из пятисот коротких траекторий для пептида N-H4 длиной 10 нс в ячейке среднего размера с использованием модели воды OPC. В каждой короткой МД траектории конформация пептида N-H4 была зафиксирована с помощью большого количества слабых ограничений, накладываемых на расстояния между всеми доступными парами атомов (см. раздел 1.3). Полученный в результате набор коротких траекторий полноценным образом отражает конформационный ансамбль для пептида N-H4 с той лишь разницей, что все конформеры диффундируют, как жесткие («замороженные») молекулы. Подобная модель ведет к значению  $D_{tr}^{N-H4}$  ниже основного результата (МД моделирование без ограничений) на 12 %. МД моделирование по той же схеме, но уже в ячейке малого размера, приводит к значению  $D_{tr}^{N-H4}$ , которое ниже на 8 % относительно основного результата. В совокупности наблюдения подтверждают гипотезу о том, что ансамблю IDPs, состоящему из «замороженных» конформеров, соответствует более низкое значение коэффициента диффузии.

Отталкиваясь от приведенного выше обсуждения, можно ожидать, что применение алгоритма HYDROPRO к отдельным кадрам из оригинальных МД траекторий (без ограничений), должно привести к заниженным значениям коэффициентов диффузии. Для проверки этого утверждения мы сделали следующее. С помощью алгоритма HYDROPRO был проведен расчет коэффициентов диффузии для пептида N-H4 по 5000 кадрам из МД траекторий в ячейке малого размера, записанных с использованием моделей воды TIP4P-Ew, TIP4P-D и OPC. В этих расчетах были использованы экспериментальные значения вязкости и плотности воды при температуре 298 К, а все остальные параметры были установлены

по умолчанию. На Рисунке 1.7 (А) показано сравнение экспериментальных данных и результатов оценки коэффициента трансляционной диффузии,  $D_{tr}$ , при помощи алгоритма HYDROPRO.

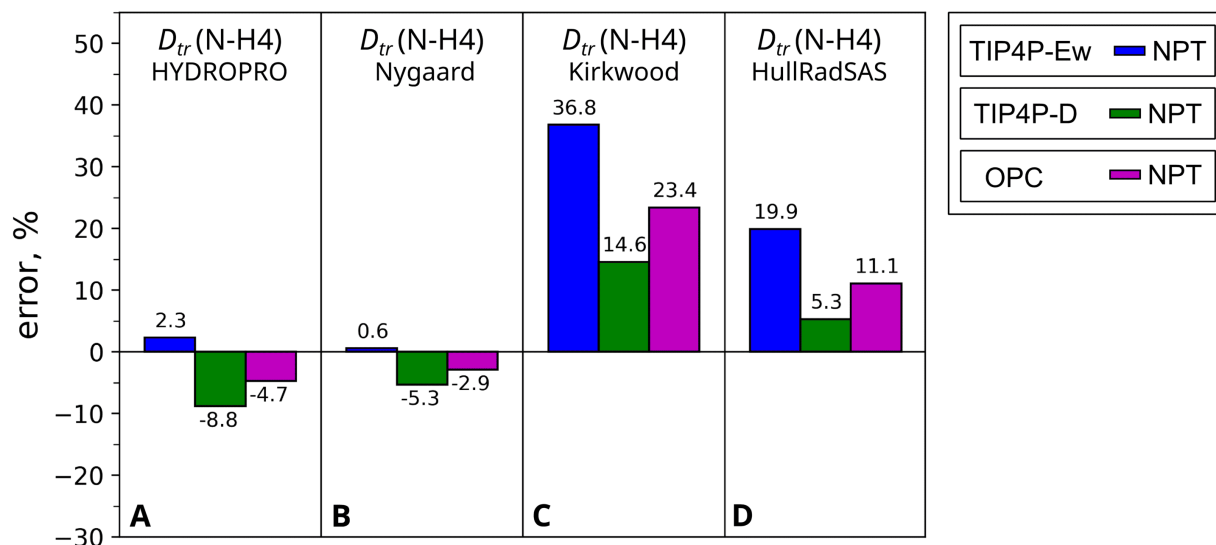


Рисунок 1.7 — Оценка расхождений между экспериментальными и расчетными значениями коэффициентов диффузии для пептида N-H4. Расчеты были проведены для трех ансамблей конформеров, построенных на основе 5000 кадров из МД траекторий в ячейке меньшего размера, записанных с использованием моделей воды TIP4P-Ew, TIP4P-D и OPC и ансамбля NPT (синие, зеленые и пурпурные столбцы соответственно). Для МД моделирования в ячейках среднего и большего размера результаты почти идентичны, что неудивительно, ввиду схожести конформационных ансамблей, см. Рисунок 1.12. Показанные на рисунке значения  $D_{tr}$  получены с помощью следующих инструментов: А) программа HYDROPRO; (В) эмпирическая формула Nygaard, основанная на данных HYDROPRO [71]; (С) модель Kirkwood-Riseman, формула (1.5); (D) алгоритм HullRadSAS. Для методов (В) и (С) полученные значения  $R_h$  были преобразованы в  $D_{tr}$  по формуле Эйнштейна-Стокса (1.3). Альтернативный метод (см. формулу (1.4)), отталкивающийся от данных для коэффициента диффузии эталонной молекулы, сообщает схожие результаты (не представлены).

Если обратиться к результатам алгоритма HYDROPRO (см. Рисунок 1.7 (А)), то можно заметить что наилучшее соответствие с экспериментальным значением коэффициента трансляционной диффузии для пептида N-H4 наблюдается в случае модели воды TIP4P-Ew.

Отклонение от данных эксперимента для этой модели воды составляет всего 2.3 %, что существенно ниже по сравнению с моделью воды TIP4P-D и моделью воды OPC. Этот результат противоречит выводам на основе прямого расчета по данным МД моделирования (см. раздел 1.3.2), согласно которым модель воды TIP4P-Ew приводит к слишком компактному конформационному ансамблю пептида N-H4 и обладает низкой вязкостью. Мы считаем, что результаты HYDRPRO ошибочны, что определяется следующими двумя факторами. Во-первых, в алгоритме HYDRPRO отдельные конформеры в структурном ансамбле неупорядоченного белка рассматриваются как жесткие молекулы. Этот фактор смещает оценку для коэффициента трансляционной диффузии в сторону меньших значений. Во-вторых, расчет при помощи алгоритма HYDRPRO проведен на кадрах из МД траектории TIP4P-Ew, где неупорядоченный пептид имеет склонность к образованию более компактных структур. Этот фактор смещает оценку реального коэффициента трансляционной диффузии в сторону больших значений. В совокупности комбинация этих факторов приводит к компенсации ошибок и ложному выводу о том, что модель воды TIP4P-Ew является наиболее подходящей для моделирования неупорядоченного пептида. В этой связи следует также упомянуть еще об одном факторе: в расчетах HYDRPRO значение вязкости растворителя задается пользователем (обычно соответствует экспериментальным значениям) и, следовательно, в такой ситуации не учитывается вязкость моделируемого растворителя, т.е. вязкость воды TIP4P-Ew.

Последний фактор имеет как положительное, так и отрицательное влияние. С одной стороны, результаты не чувствительны к некорректной вязкости растворителя, как это происходит при расчете коэффициента диффузии напрямую из траектории молекулярной динамики (см. раздел 1.3.2). С другой стороны, тот факт, что в подходе HYDRPRO не учитывается вязкость моделируемого растворителя, может привести к неправильным выводам при интерпретации данных для МД ансамбля конформеров (что в дальнейшем скажется на других расчетах, к примеру, на расчете скоростей спиновой релаксации).

В завершении обсуждения алгоритма HYDRPRO, важно отметить, что этот подход демонстрирует очень хорошую точность при оценке  $D_{tr}$  для убиквитина (см. Рисунок 1.15 (A)). Результаты практически не

зависят от используемой модели воды. Это объясняется тем, что структура свернутого белка практически нечувствительна к выбору модели воды (см. раздел 1.3.2). Одновременно с этим, HYDROPRO довольно плохо справляется с задачей оценки вращательного коэффициента диффузии,  $D_{rot}$ , демонстрируя уровень ошибки вплоть до 20 % (см. Рисунок 1.15 (D)). Похожие результаты для  $D_{rot}$  были получены авторами одной из предыдущих работ [116]. В совокупности полученные данные указывают на существующие ограничения гидродинамических моделей, опирающихся на алгоритм HYDROPRO.

Следует отметить, что эмпирическая параметризация Nygaard, устанавливающая взаимосвязь между радиусом инерции,  $R_g$ , и гидродинамическим радиусом,  $R_h$ , рассчитанным при помощи HYDROPRO, хорошо применима к рассматриваемым белкам (Рисунок 1.7 (A) в сравнении с Рисунок 1.7 (B) и Рисунок 1.7 (A) в сравнении с Рисунок 1.7 (B)). Однако это лишь означает, что критика подхода HYDROPRO для оценки коэффициента трансляционной диффузии IDPs, остается справедливой и в отношении подхода Nygaard.

Таким образом, можно дать общую рекомендацию не использовать подход HYDROPRO в приложении к конформационным ансамблям IDPs, поскольку алгоритм не предназначен для работы с неупорядоченными белками и может приводить к ложным результатам.

Относительно недавно группой Lindorff-Larsen были предложены альтернативные схемы для расчета коэффициента трансляционной диффузии для статистического ансамбля неупорядоченных белков. Для набора IDPs различной длины авторы сконструировали ансамбль при помощи известной программы Flexible-Meccano [117] и при помощи моделирования методом динамики Ланжевена в крупнозернистом силовом поле CALVADOS [118]. Далее, индивидуальным конформерам назначались веса таким образом, чтобы добиться наилучшего согласия с данными экспериментов малоуглового рассеяния. Затем авторы попытались установить взаимосвязь между конформационным ансамблем и данными о гидродинамическом радиусе  $R_h$  ( $D_{tr}$ ) из диффузионных ЯМР экспериментов.

Изначально в качестве такой модели авторы рассмотрели обобщенную модель Kirkwood-Riseman, описывающую гидродинамические свойства

гибких полимеров [119]. Позднее выяснилось, что эта модель дает достаточно точные значения  $R_h$  для конформационных ансамблей с некоторыми ограничениями на размер белка [72]. Однако вскоре экспериментальные данные были уточнены, и оказалось, что модель Kirkwood-Riseman имеет тенденцию к завышению коэффициентов трансляционной диффузии для IDPs [73].

Относительно недавно теми же авторами был предложен алгоритм HullRadSAS [120]. Алгоритм HullRadSAS в приложении к конформационным ансамблям дает довольно точные оценки для  $R_h$  [73]. Однако на уровне базовых принципов подход HullRadSAS имеет сходство с подходом HYDROPRO. Аналогично HYDROPRO, подход HullRadSAS не задумывался изначально как инструмент для работы с неупорядоченными белками. Алгоритм HullRadSAS эмпирическим путем устанавливает взаимосвязь между набором моделей и данными  $R_h$  ( $D_{tr}$ ).

В настоящей работе были протестированы модель Kirkwood-Riseman и подход HullRadSAS (см. раздел 1.3). Результаты для модели Kirkwood-Riseman выглядят крайне неудовлетворительно (см. Рисунок 1.7 (C)), что согласуется с недавними наблюдениями [73] о том, что модель Kirkwood-Riseman склонна переоценивать экспериментальное значение коэффициента диффузии. Так, модель Kirkwood-Riseman в приложении к кадрам из МД траектории приводит к заметным ошибкам в 14.6 % и 23.4 % относительно ожидаемого значения  $D_{tr}$  для моделирования в воде TIP4P-D и в воде OPC. В то же самое время прямой расчет  $D_{tr}$  из данных МД моделирования позволяет говорить о практически количественной точности в случае использования модели воды TIP4P-D или модели воды OPC.

Для алгоритма HullRadSAS (см. Рисунок 1.7 (D)) результаты чуть лучше согласуются с данными прямого расчета по траекториям молекулярной динамики. Однако количественное согласие неидеально. Ранее по результатам прямого расчета был сделан вывод, что если для модели воды TIP4P-D и модели воды OPC скорректировать значение вязкости, то для диффузионных характеристик пептида N-H4 будет получено хорошее согласие с экспериментом (см. Рисунок 1.5 (E, D)). Напротив, алгоритм HullRadSAS сообщает завышенные оценки экспериментального значения  $D_{tr}$  на 5.3 % и 11.1 % для моделирования в воде TIP4P-D и в воде OPC. Таким образом, используя алгоритм

HullRadSAS можно прийти к неверному выводу о том, что модель воды OPC плохо подходит для моделирования неупорядоченных белков.

Алгоритм HullRadSAS, как и алгоритм HYDROPRO не был задуман для работы с неупорядоченными белками, т.е. в отношении HullRadSAS остается справедливой рекомендация об осторожном его использовании в отношении IDPs. В том случае, когда основной задачей является оценка качества МД модели неупорядоченного пептида, наиболее надежный подход — извлечь коэффициент диффузии напрямую из данных МД моделирования (как это продемонстрировано в настоящей работе), а не полагаться на эмпирические инструменты для оценки  $D_{tr}$  по отдельным кадрам МД траектории.

### 1.2.6 Ускоренный протокол МД моделирования диффузионных свойств белков

Расчет коэффициента диффузии  $D_{tr}$  напрямую из МД траекторий на основе анализа зависимости  $MSD(\tau)$  представляется более надежным подходом по сравнению с эмпирическими схемами оценки  $D_{tr}$  на основе координат атомов для ансамбля конформеров. Главным препятствием в использовании прямого подхода является то, что МД моделирование отнимает много времени. В настоящей работе было показано, что временные затраты можно сократить в несколько раз за счет использования специального протокола моделирования, который предполагает параллельную запись набора коротких траекторий в ячейках среднего и большого размеров. В настоящем разделе на примере белка убиквитина (Ub) будет рассмотрена дополнительная возможность для ускорения расчетов за счет увеличения шага интегрирования уравнений движения.

С этой целью было записано три дополнительных серии траекторий белка Ub с увеличенным шагом интегрирования, а именно:  $\Delta t = 2$  фс,  $\Delta t = 4$  фс (в которой в соответствии с рекомендациями разработчиков Amber схема перераспределения массы водорода, HMR, прикладывалась к Ub и не применялась к растворителю — воде) и  $\Delta t = 4$  фс (где схема HMR



прикладывалась как к Ub, так и растворителю — воде). Каждая серия состоит из трех МД траекторий в ячейках малого, среднего и большого размеров и имеет суммарную длительность 6 мкс. Исключением является серия МД траекторий с использованием NMR, для которой протокол моделирования не отличается от описанного в разделе 1.3. Моделирование осуществлялось в воде OPC, как одной из наиболее успешно проявившихся себя в стандартной схеме моделирования. Результаты представлены в Таблице 3.

Анализируя данные Таблицы 3, можно заметить, что переход от  $\Delta t = 1$  фс к  $\Delta t = 2$  фс ведет к увеличению коэффициента диффузии  $D_{tr}^{Ub}$  и  $D_{rot}^{Ub}$  для Ub на несколько процентов. С другой стороны, переход к большему шагу интегрирования,  $\Delta t = 4$  фс, в сочетании со схемой NMR на Ub приводит к более значительному увеличению  $\sim 15-20\%$ . При этом, использование схемы NMR, приложенной как к белку Ub, так и к растворителю (воде), позволяет привести результаты моделирования с шагом  $\Delta t = 4$  фс в соответствии с результатами для  $\Delta t = 1$  фс.

В руководстве пользователя пакета Амбер в качестве выбора по умолчанию предлагается схема  $\Delta t = 4$  фс с коррекцией NMR, приложенной только к белковой молекуле, но не к растворителю [121]. Дело в том, что в МД моделировании молекулы воды по умолчанию рассматриваются как жесткая система, внутренние движения которой ограничены действием алгоритма SETTLE [122]. Тем не менее в предыдущих работах отмечается, что алгоритм SETTLE не всегда устраняет все высокочастотные колебания в молекуле воды [123]. Причиной таких высокочастотных движений могут служить короткодействующие нековалентные взаимодействия между соседними молекулами воды [123]. Мы предполагаем, что применение схемы NMR ко всей системе (белок и растворитель) позволяет подавить эти высокочастотные движения и, таким образом, повысить точность МД моделирования с увеличенным шагом интегрирования  $\Delta t = 4$  фс.

Полученные результаты носят предварительный характер, тем не менее, полученные для Ub данные не теряют своей значимости в случае растворителя (молекул воды) и неупорядоченного N-H4 пептида. По-видимому, для этих систем моделирование с использованием шага интегрирования  $\Delta t = 4$  фс дает точные параметры коэффициентов диффузии тогда и только тогда, когда схема NMR применяется ко всей

Таблица 3 — Коэффициент диффузии убиквитина (Ub) для дополнительной серии МД траекторий с увеличенным шагом интегрирования уравнений движения и частичным или полным действием схемы коррекции HMR. Остальные параметры моделирования не отличались от приведенных в разделе 1.3. В моделировании были использованы модель воды ОРС и ансамбль NPT. Экстраполяция значения коэффициента диффузии к пределу ячейки бесконечного размера показана на Рисунке 1.16

Белок	Коэффициент диффузии	Эксперимент	Simulations			
			$\Delta t = 1$ фс без HMR	$\Delta t = 2$ фс без HMR	$\Delta t = 4$ фс HMR (белок)	$\Delta t = 4$ фс HMR (белок + H <sub>2</sub> O)
Ub	$10^{-10} \text{ m}^2/\text{с}$	1.62	1.69	1.8	1.93	1.68
	$10^7 \text{ rad}^2/\text{с}$	4.87	4.94	4.97	5.91	4.72

системе (белок и растворитель). Если сделанный вывод подтвердится, то оценки коэффициента диффузии белка из МД траекторий можно будет осуществлять в четыре раза быстрее, чем при использовании классического протокола.

### 1.3 Материалы и методы

#### 1.3.1 МД моделирование

Начальные координаты для пептида N-H4 (аминокислотная последовательность SGRGKGGKGLGKGGAKRHRKVLRDN) были получены по протоколу, описанному ранее [49]. Во-первых, генерировался набор из 2000 случайных конформеров пептида N-H4 при помощи программы Scwrl4 [124] и программного обеспечения, распространяемого через сервер <http://unfolded.uchicago.edu> [125]. Затем энергия каждой конформации минимизировалась, после чего структуры были упорядочены в соответствии с их энергиями в растворителе GBneck2 [126]. И наконец, из центральной части распределения по энергии случайным образом была выбрана одна структура, которая в дальнейшем использовалась как стартовая во всех МД траекториях.

Подготовленная начальная структура протонировалась при помощи программы PROPKA [127] (рН 4.0 в соответствии с условиями эксперимента). В эксперименте низкое значение рН позволяет свести к минимуму обмен протонами между амидными группами и водой. Следует отметить, что протонирование пептида N-H4 при рН 4.0 не отличается от протонирования при физиологическом значении рН 7.2, за исключением единственного остатка гистидина [49]. Моделирование осуществлялось при помощи программного пакета Amber 20 в силовом поле ff14SB с использованием трех различных моделей воды: TIP4P-Ew, TIP4P-D и OPC. Для каждой выбранной модели воды были записаны траектории в трех ячейках моделирования увеличивающегося размера. Сначала будет описан

протокол моделирования в ячейке малого размера (схему моделирования в ячейках среднего и большого размера мы опишем позже).

Ячейка моделирования в виде усеченного октаэдра была сконструирована с использованием программы LEaP при помощи команды *SolvateOct*. При этом параметр, задающий минимальное расстояние от атомов пептида до границ ячейки, был установлен равным 12 Å для модели воды TIP4P-Ew и 15 Å для моделей воды TIP4P-D и OPC. Нейтральный заряд ячейки моделирования обеспечивался добавлением девяти ионов Cl<sup>-</sup> [128]. После чего была проведена минимизация энергии с гармоническими ограничениями, наложенными на положения тяжелых атомов пептида N-H4. Далее, система была нагрета до целевой температуры 298 К. В качестве последнего этапа подготовки система была уравновешена в ансамбле NVE в течение 1 нс.

Температура в ходе моделирования поддерживалась при помощи алгоритма термостатирования Бусси [63], который не искажает естественную динамику системы. Термостат Бусси был запрограммирован для платформы Amber 20 (опция *ntt* = 11). Контроль давления осуществлялся при помощи баростата Берендсена [129] с временной константой 1 пс. Для нековалентных взаимодействий был использован радиус обрезки 11 Å. Дальнодействующие взаимодействия рассчитывались по методу суммирования Эвальда с параметрами по умолчанию для шага сетки и сплайн-интерполяции. Все связи с участием атомов водорода были ограничены алгоритмом SHAKE [130]. Шаг интегрирования составлял 1 фс (для согласия с дополнительной серией МД траекторий в ансамбле NVE, см. ниже). В ходе моделирования координаты системы сохранялись на диск через каждую 1 пс траектории. Моделирование проводилось на вычислительном кластере, оснащенный графическими ускорителями NVIDIA GeForce GTX 1080, RTX 2080 Ti и RTX 3080. Длительность траектории пептида N-H4 в ячейке моделирования малого размера для каждой из трех моделей воды равнялась 5 мкс.

Необходимо отметить, что приведенное выше описание протокола относится к МД моделированию неупорядоченного пептида в ячейке ограниченного размера с наложенными периодическими граничными условиями. Возникает вопрос: влияют ли граничные условия на конформационное многообразие. Ответ зависит от количества столкновений

пептида N-H4 со своими периодическими образами за время моделирования. Поэтому в записанных МД траекториях мы сосчитали долю кадров, в которых атомы основной копии пептида N-H4 сближаются с атомами периодического образа на расстояние  $5 \text{ \AA}$  и менее. Оказалось, что количество таких контактов составляет долю процента от общего числа кадров в МД траектории, а именно 0.06 %, 0.36 % и 0.17 % для моделирования с использованием модели воды TIP4P-Ew, TIP4P-D и ОРС соответственно. Следовательно, можно сделать вывод о том, что эффект граничных условий может быть проигнорирован без какого-либо ущерба для точности расчетов.

Как уже было отмечено, для корректного извлечения  $D_{tr}$  и  $D_{rot}$  из траекторий молекулярной динамики требуется провести моделирование в ячейках увеличивающегося размера. В случае ячейки малого размера МД траектория записывается достаточно быстро, в то время как МД моделирование в ячейках большого размера требует значительных временных затрат. Для того чтобы уменьшить время моделирования, была разработана специальная схема записи МД траекторий в ячейках среднего и большого размера, см. Рисунок 1.8. Идея состоит в следующем, из МД траектории в ячейке малого размера извлекаются кадры (координаты атомов пептида N-H4) в моменты времени 0, 10, 20, ..., 4990 нс, которые затем используются в качестве стартовых конформаций для моделирования в ячейках среднего и большого размера. Ячейка среднего размера была сконструирована при помощи команды SolvateOst с минимальной толщиной водного слоя вокруг пептида равной  $24 \text{ \AA}$ , ячейка большого размера — с толщиной водного слоя равной  $48 \text{ \AA}$ . Далее, происходило уравнивание сконструированных ячеек (по протоколу, описанному выше для ячейки малого размера). И наконец, осуществлялся запуск коротких траекторий длиной  $x$  нс (см. Рисунок 1.8).

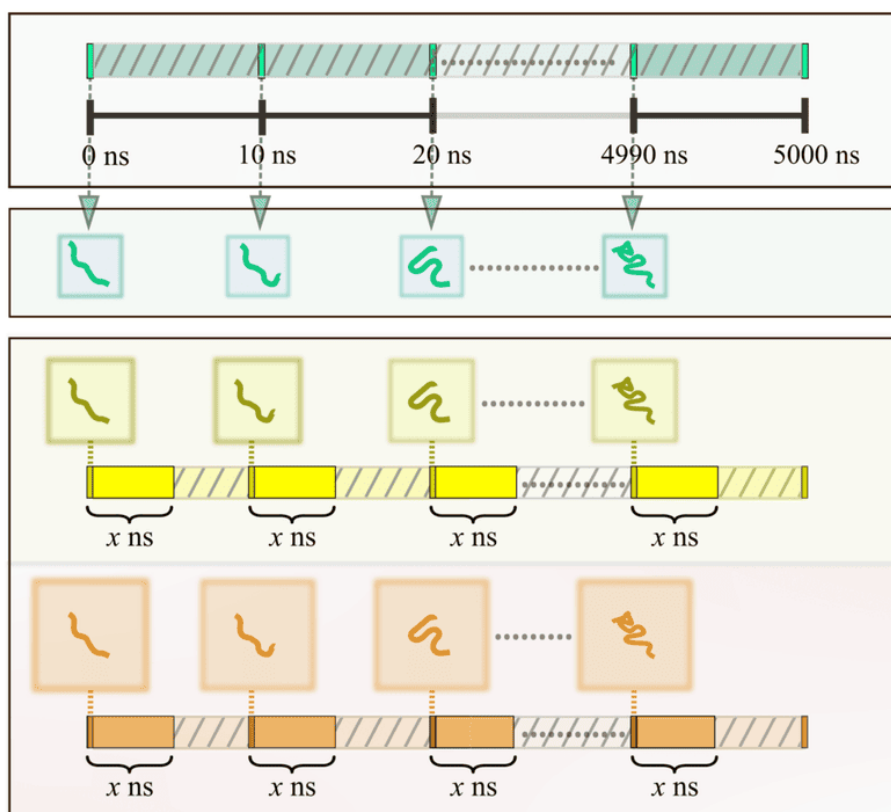


Рисунок 1.8 — Схема, иллюстрирующая протокол моделирования N-H4 пептида в ячейках малого, среднего и большого размера. Первоначально была записана серия коротких траекторий длиной  $x = 5$  нс, которая затем была расширена до  $x = 10$  нс

В данной работе была записана серия коротких траекторий, длина каждой из которых составила  $x = 5$  нс. Предложенная схема обеспечивает существенную экономию вычислительного времени. Во-первых, благодаря уменьшению суммарной длины МД траекторий достигается двукратный выигрыш во времени моделирования, см. Рисунок 1.8. Во-вторых, благодаря использованию высокопроизводительно вычислительного кластера, оснащенного графическими ускорителями и поддерживающего параллельные вычисления коротких МД траекторий (в настоящей работе расчеты проводились на кластере с системой очередей SLURM), удастся в несколько раз увеличить общую производительность. Одновременно с этим, схема  $x = 5$  нс обеспечивает эффективное семплирование конформационного пространства в ячейках среднего и большого размеров, так как в качестве стартовых конфигураций выбираются кадры траектории пептида N-H4 длиной 5 мкс в ячейке малого размера. Для того чтобы продемонстрировать

надежность полученных результатов, записанная серия с  $x = 5$  нс была расширена до  $x = 10$  нс.

В дополнение к серии МД траекторий в ансамбле NPT с термостатом Бусси, была записана серия МД траекторий в ансамбле NVE для двух моделей воды TIP4P-Ew и TIP4P-D. Хорошо известно, что моделирование в ансамбле NVE соответствует «естественной» динамике системы, свободой от потенциального искажения алгоритмами термостатирования. Однако дрейф температуры при моделировании в ансамбле NVE делает его малоприменимым для повсеместного использования [131; 132]. Мы провели моделирование в ансамбле NVE с целью валидации результатов, полученных для ансамбля NPT с термостатом Бусси. Протокол записи траекторий в ансамбле NVE не отличался от описанного выше для ансамбля NPT, включая шаг интегрирования равный 1 фс. Малое значение шага интегрирования способствует уменьшению дрейфа температуры. Для МД траекторий пептида N-H4 в ячейке малого размера с использованием ансамбля NVE наблюдался небольшой дрейф температуры за время моделирования в 5 мкс, а именно: увеличение температуры на 1.2 К для траектории в воде TIP4P-Ew и увеличение на 0.5 К для траектории в воде TIP4P-D. Таким образом, существующий дрейф температуры невелик и позволяет использовать МД траектории в ансамбле NVE для валидации МД траекторий в ансамбле NPT.

Полное описание всех МД траекторий для пептида N-H4, полученных в рамках данной работы, приведено в Таблице 4. Суммарная длина всех траекторий для пептида N-H4 составляет 75 мкс.

Отдельно, для изучения влияния локальной подвижности пептида N-H4 на его диффузионные свойства, было проведено моделирование пептида N-H4 в условиях ограниченной динамики (что соответствует условиям при расчете  $D_{tr}$  с помощью алгоритма HYDROPRO). Для модели воды OPC был записан набор из пятисот траекторий длиной 10-нс в ячейке среднего размера. Каждая из траекторий была записана в условиях мягких ограничений, наложенных на расстояния между всеми парами атомов пептида N-H4 (всего 73 536 ограничений). Соответствующий потенциал был определен, как потенциальная яма с параболическим дном  $k(d - d_0)^2$  заданным от  $0.99 d_0$  до  $1.01 d_0$ , где  $d_0$  — межатомное расстояние в исходном кадре, а значение  $k$  было установлено равным

1 ккал/моль · Å<sup>2</sup>. За пределами параболической потенциальной ямы действовал линейный потенциал [66]. Смысл действия приложенных ограничений — поддержание начальной конформации пептида для каждой 10-нс траектории в неизменном виде. Для записанных таким образом траекторий среднеквадратичное отклонение координат атомов от начальной конформации пептида N-H4 не превышало 0.15 Å. За время моделирования в условиях ограниченной динамики температура, давление и объем оставались близкими к установленным значениям, а дополнительная энергия, необходимая для поддержания начальной конформации в неизменном виде, составляла не более 0.11 % от полной энергии. Аналогичным образом было осуществлено моделирование в ячейке малого размера, а именно: непрерывная траектория в модели воды OPC была разбита на пятьсот 10-нс фрагментов, после чего начальная конформация из каждого фрагмента была использована, как стартовая для записи траектории длиной 10-нс в условиях ограниченной динамики.

Дополнительно были записаны МД траектории для известного модельного белка убиквитина (Ub). Начальные координаты системы соответствовали структуре 1UBQ [133]. Протонирование было проведено с помощью программы tleap (согласно оценкам PROPKA состояние протонирования для белка Ub не отличается в диапазоне от pH 4.5 до 6.5). Моделирование проводилось при температуре 303 К; остальные параметры аналогичны используемым в моделировании пептида N-H4. Схема МД моделирования также не отличалась от описанной выше, см. Рисунок 1.8, т.е. МД траектория в ячейке малого размера была записана как непрерывная траектория, тогда как моделирование в ячейках среднего и большого размеров представляло собой совокупность коротких траекторий по схеме  $x = 5$  нс (в дальнейшем каждая траектория была продлена до  $x = 10$  нс). Поскольку убиквитин — упорядоченный белок, для полноценного исследования его конформационного пространства не требуется такая длинная траектория, как в случае неупорядоченного N-H4 пептида. Поэтому время моделирования было сокращено с 5 до 2 мкс. Аналогичный подход был использован при моделировании белка Ub в ансамбле NVE для двух моделей воды TIP4P-Ew и TIP4P-D. В случае белка Ub температурный дрейф для траекторий в ансамбле NVE в ячейке малого размера оказался незначительным (0.4 К и 0.2 К для моделирования в воде



TIP4P-Ew и в воде TIP4P-D), что позволило осуществлять моделирование по схеме, показанной на Рисунке 1.8. Список всех МД траекторий белка Ub, полученных в рамках настоящей работы, приведен в Таблице 5. Суммарная длина всех траекторий Ub составляет 30 мкс.

Для того чтобы оценить вязкость моделей воды TIP4P-Ew, TIP4P-D и OPC, был записан набор траекторий в ячейках моделирования увеличивающегося размера, которые содержали исключительно молекулы воды. Геометрические параметры для ячейки малого, среднего и большого размеров были установлены равными геометрическим параметрам для соответствующих ячеек моделирования белка Ub. Общая длина всех траекторий для моделирования молекул воды составила 1.35 мкс (длина каждой индивидуальной траектории 0.14 мкс).

И наконец, были протестированы схемы МД моделирования с различной величиной шага интегрирования уравнений движения, а именно: было проведено моделирование белка Ub в воде OPC с использованием шага интегрирования в 2 фс и 4 фс с двумя опциями с применением и без применения алгоритма перераспределения массы водорода (от англ. Hydrogen Mass Repartition scheme, HMR) ([134]).

### 1.3.2 Обработка МД траекторий

К МД траекториям, записанным в периодических граничных условиях, был применен алгоритм разворачивания к непрерывному виду (от англ. unwrapping). При этом особое внимание уделялось флуктуациям геометрических размеров ячейки моделирования в ансамбле NPT [135]. Для этого был использован собственный скрипт, написанный на языке python, на основе функционала библиотеки ruhmolpp2.

Коэффициенты трансляционной диффузии  $D_{tr}$  были определены по результатам анализа среднего квадрата смещения центра масс белка (от англ. mean square displacement, MSD) [136]. Для набора из  $N$  траекторий, где длина каждой траектории равна  $\chi$  нс, средний квадрат смещения центра масс был рассчитан по формуле

$$MSD(\tau) = \frac{1}{N \cdot n} \sum_{i=1}^N \sum_{j=1}^n |\vec{r}_i(t_j + \tau) - \vec{r}_i(t_j)|^2 \quad (1.1)$$

где  $\vec{r}_i$  определяет положение центра масс белка в  $i$ -й траектории,  $t_j$  и  $t_j + \tau$  пара точек, разделенных промежутком времени  $\tau$ ,  $n$  количество таких пар для траектории длиной  $x$ . В случае непрерывной траектории формула (1.1) сводится к известному выражению для MSD [136].

Расчетное значение  $D_{tr}$  было получено в результате аппроксимации зависимости  $MSD(\tau)$  при помощи линейной функции. Для корректного извлечения  $D_{tr}$  важно аккуратным образом выбрать временной интервал  $\tau$ , на котором осуществляется аппроксимация. С одной стороны, по мере увеличения  $\tau$  возрастает погрешность соответствующих значений  $MSD(\tau)$ , при этом ошибки оказываются коррелированы [137; 138]. С другой стороны, небольшие значения  $\tau$  также следует исключить из рассмотрения, в особенности при наличии локальной динамики в системе (подробное обсуждение приведено ниже).

Конформационная динамика неупорядоченного белка ведет к (ограниченному) смещению центра масс белка. Это явление можно рассматривать, как ограниченную диффузию. Для пептида N-H4 в МД траектории было зафиксировано два динамических режима: быстрый режим с характерным временем около  $\sim 100$  пс и медленный режим с характерным временем около  $\sim 1$  нс [49]. Таким образом, вклад локальной динамики пептида N-H4 в зависимость  $MSD(\tau)$  проявляется для значений  $\tau$ , принадлежащих субнаносекундному диапазону.

Полученные данные МД моделирования указывают на справедливость приведенных выше рассуждений (см. Рисунок 1.9 (E)). В то же время, экспериментальные значения  $D_{tr}$  не содержат информации о внутренней динамике, поскольку измерения соответствуют значениям  $\tau$  от десятков до сотен миллисекунд. Поэтому аппроксимацию кривой  $MSD(\tau)$  необходимо реализовать таким образом, чтобы минимизировать вклад внутренней динамики.

Согласно сформулированным критериям, для аппроксимации зависимости  $MSD(\tau)$  был выбран временной интервал [100 пс, 1 нс] (см. Рисунок 1.9). Введенная схема аппроксимации является устойчивой, а именно: между различными вариантами временных интервалов для

подгонки  $D_{tr}$  не было обнаружено значимых отличий (см. Рисунок 1.9). Дополнительно была проведена оценка расхождений для значений  $D_{tr}$ , рассчитанных на основе двух наборов траекторий, соответствующих схемам моделирования  $x = 5$  нс и  $x = 10$  нс. Как показано на Рисунке 1.10 (А), для указанных схем моделирования полученные значения  $D_{tr}$  хорошо согласуются друг с другом. Тем не менее, схема моделирования  $x = 10$  нс позволяет определить  $D_{tr}$  с максимальной точностью. Поэтому во всех расчетах в настоящей главе была использована схема моделирования с  $x = 10$  нс.

Схема расчета  $D_{tr}$  для белка Ub не отличается от описанного выше протокола для пептида N-H4. Схожую структуру имеет и процедура расчета  $D_{tr}$ , коэффициента самодиффузии воды. Но более быстрая внутренняя динамика молекул воды (фемтосекунды) позволяет использовать более короткий временной интервал [1 пс, 100 пс] для подгонки кривой MSD, чтобы уменьшить статистическую неопределенность [137; 138]. Значения  $D_{tr}$  были рассчитаны для 100 молекул воды, выбранных случайным образом из ячейки моделирования, после чего результат усреднялся для получения итогового значения коэффициента самодиффузии.

Далее, была построена зависимость  $D_{tr}$  от обратного линейного размера ячейки моделирования. В принципе, любая характеристика линейного размера ячейки моделирования может быть выбрана для этой цели (например, длина ребра  $a$  для ячейки моделирования в форме усеченного октаэдра). В настоящей работе в качестве соответствующей характеристики был выбран диаметр сферы, вписанной в усеченный октаэдр,  $L = \sqrt[3]{a}$ . Удобным оказывается то, что значение  $L$  указано в последней строке файла координат rst7.

Для набора МД траекторий пептида N-H4 в ячейках моделирования среднего и большого размеров наблюдалась небольшая вариация в геометрических параметрах ячейки (в зависимости от начальной конформации пептида). Например, в случае модели воды ОРС вариация составила  $L = 142 \pm 9$  Å для 500 коротких траекторий в ячейке моделирования большого размера. Вариация обратного линейного размера ячейки  $1/L$  при этом составляет  $0.0071 \pm 0.0004$  Å<sup>-1</sup>. При экстраполяции  $D_{tr}$  к пределу бесконечно большой ячейки было использовано значение  $0.0071$  Å<sup>-1</sup> (см. Рисунок 1.4). Как альтернативный подход, можно было бы

выбрать ячейку моделирования универсального размера (к примеру, самую большую ячейку моделирования из всех начальных конфигураций).

Сравнение результатов моделирования в ансамблях NVE и NPT позволяет косвенным образом получить представление о погрешности процедуры расчета  $D_{tr}$  из МД траекторий. Для того чтобы напрямую оценить погрешность значений  $D_{tr}$ , рассчитанных по данным МД траекторий, был реализован метод «складного ножа». МД траектория в ячейке малого размера была разделена на фрагменты длиной 10 нс, и каждому такому фрагменту была сопоставлена короткая траектория в ячейках моделирования среднего и большого размеров. После чего из полученного набора случайным образом исключалось 20 % фрагментов длиной 10-нс, и на основе этого отредактированного набора проводился расчет коэффициента трансляционной диффузии. Описанную процедуру повторяли 1000 раз. Стандартное отклонение для полученного распределения  $D_{tr}$  было принято за искомую погрешность значения  $D_{tr}$  (см. Таблицу 6).

Коэффициент вращательной диффузии,  $D_{rot}$ , может быть определен только для упорядоченного белка Ub, в отличие от коэффициента трансляционной диффузии,  $D_{tr}$ , который может быть рассчитан как для упорядоченного белка Ub, так и для неупорядоченного пептида N-H4. Расчет  $D_{rot}$  по данным МД траектории предполагает процедуру, включающую несколько этапов. Во-первых, совмещение всех кадров из МД траектории Ub с первым кадром путем наложения  $C\alpha$  атомов, принадлежащих участкам вторичной структуры. Во-вторых, параметризация всех операций наложения при помощи матриц вращения. В-третьих, введение в рассмотрение псевдомолекулы, сферы с единичным радиусом, на поверхности которой равномерно распределены 64 вектора (имитируют N-H<sup>N</sup> вектора) [139]. В-четвертых, применение матриц вращения к псевдомолекуле (в результате получается траектория для псевдомолекулы, в которой закодировано вращательное движение белка). В-пятых, вычисление корреляционных функций  $g(\tau)$  для всех 64 векторов по формуле

$$g(\tau) = \frac{1}{N \cdot n} \sum_{i=1}^N \sum_{j=1}^n P_2(\cos \chi_i(t_j, t_j + \tau)) \quad (1.2)$$

где  $P_2(x)$  полином Лежандра второго порядка,  $\chi_i(t_j, t_j + \tau)$  угол между направлениями вектора в моменты времени  $t_j$  и  $t_j + \tau$  для  $i$ -й траектории; усреднение осуществляется как по  $n$  парам временных точек, разделенных интервалом  $\tau$ , так и по  $N$  траекториям длиной  $x$  нс [140]. Внутренняя сумма в уравнении (1.2) была вычислена с использованием быстрого преобразования Фурье [141].

В-шестых, вычисление корреляционной функции  $G(\tau)$ , как результата сложения корреляционных функций для 64 векторов с соответствующими весовыми коэффициентами [139]. В-седьмых, аппроксимация  $G(\tau)$  при помощи моноэкспоненциальной функции, определение  $\tau_{rot}$  и последующий расчет  $D_{rot} = 1/((6\tau_{rot}))$ . Как и в случае с трансляционной диффузией, важно выбрать подходящий временной интервал для подгонки  $G(\tau)$ . Этот вопрос достаточно подробным образом был изучен на примере длинных траекторий молекулярной и броуновской динамики для Ub (данные не показаны). В результате мы пришли к выводу, что наиболее точные оценки получаются при использовании интервала  $[0, 2 \text{ нс}]$ . Примеры функций  $G(\tau)$  и варианты моноэкспоненциальной аппроксимации показаны на Рисунке 1.11.

Описанная выше процедура извлечения коэффициента вращательной диффузии позволяет получать значения  $D_{rot}$ , которые могут быть напрямую сопоставлены с результатами измерений ЯМР. Существуют также альтернативные схемы, которые можно использовать для извлечения  $D_{rot}$  из траекторий молекулярной динамики. Например, по МД траектории можно рассчитать значения скоростей релаксации  $^{15}\text{N}$  (см. обсуждение ниже), затем определить анизотропный тензор вращательной диффузии убиквитина [142], и наконец, перейти к изотропному коэффициенту вращательной диффузии  $D_{rot}$ . В настоящей работе был использован наиболее простой и наиболее общий подход [140].

Как и в случае коэффициента трансляционной диффузии, результаты расчетов коэффициента вращательной диффузии зависят от размера ячейки моделирования и требуют экстраполяции к пределу бесконечно большой ячейки моделирования. Отличие в том, для  $D_{rot}$  характерна зависимость от величины обратного объема ячейки, тогда как значения  $D_{tr}$  зависят от обратного линейного размера [143]. При экстраполяции можно использовать любую величину, пропорциональную объёму ячейки моделирования, удобным оказывается параметр  $L^3$ . Результаты

экстраполяции показаны на Рисунке 1.10 (В). Оценка погрешности определения  $D_{rot}$  по данным МД траекторий была осуществлена при помощи метода «складного ножа», идея которого была описана выше для  $D_{tr}$ .

Для проведения расчетов  $D_{tr}$  и  $D_{rot}$  был разработан собственный программный пакет, который находится в открытом доступе в репозитории по ссылке <https://github.com/bionmr-spbu-projects/2023-UBQ-NH4-DIFFUSION>. В репозитории содержатся все необходимые инструкции по установке и запуску, а также пример анализа короткой траектории Ub для демонстрации работы скриптов. Анализ траекторий МД опирается на функционал библиотеки `рухmolpp2`, написанной на языке `python` и доступной в репозитории по ссылке <https://github.com/bionmr-spbu/рухmolpp2>.

Не смотря на то, что концепция вращательной диффузии не применима по отношению к неупорядоченным белкам, сопоставимая информация о динамике может быть получена из данных гетероядерной релаксации. В настоящей работе для каждого аминокислотного остатка пептида N-H4 из МД траектории были рассчитаны значения скоростей продольной ядерной релаксации,  $R_1$ , и поперечной кросс-корреляционной релаксации,  $\eta$  (между дипольным механизмом релаксации и механизмом релаксации, обусловленным анизотропией химического сдвига (от англ. *chemical shift anisotropy, CSA*)) [49]. Соответствующие корреляционные функции были рассчитаны по формуле (1.2). Для кросс-корреляционной функции  $\chi$  соответствует углу между вектором N-H<sup>N</sup> и главной осью аксиально-симметричного тензора <sup>15</sup>N CSA в момент времени  $t + \tau$ . При работе с массивом коротких траекторий в ячейках среднего и большого размеров расчеты проводились для схемы моделирования  $x = 10$  нс (также как для  $D_{tr}$  и  $D_{rot}$ ).

Аппроксимация корреляционных функций осуществлялась в виде суммы четырех экспонент  $\sum_{k=1}^4 c_k \exp(-\tau/\tau_k)$ . При этом набор допустимых значений для четырех характерных времен  $t_k$  был ограничен временными интервалами [1 пс, 10 пс], [10 пс, 100 пс], [100 пс, 1 нс] и [1 нс, 10 нс], а весовые коэффициенты  $c_k$  были отнормированы таким образом, чтобы их сумма равнялась 1.0. Поиск оптимальных параметров  $t_k$  и  $c_k$  осуществлялся по алгоритму Левенберга-Марквардта из библиотеки `SciPy`

(функция `curve_fit`) [144]. Найденные значения  $t_k$  и  $c_k$  были использованы для расчета функции спектральной плотности, которая, в свою очередь, входит в выражение для вычисления  $R_1$  и  $\eta$  [145]. При расчете скоростей релаксации были использованы стандартные значения для длины связи N-H (1.02 Å), для анизотропии химического сдвига  $^{15}\text{N}$  (-170 ppm) и для угла между главной осью тензора анизотропии химического сдвига  $^{15}\text{N}$  и вектора N-H<sup>N</sup> (20°) [146].

Алгоритм HYDROPRO был запущен на отдельных кадрах, взятых с шагом равным 1 нс из МД траекторий в ячейке моделирования малого размера. Запуск алгоритма осуществлялся с экспериментальными значениями для вязкости и плотности воды при температуре 298 К или 303 К (для N-H4 и Ub соответственно). Остальные параметры были инициализированы значениями по умолчанию. Алгоритм HullRadSAS был запущен аналогичным образом [120]. Кроме того, был осуществлен расчет по эмпирической схеме, устанавливающей взаимосвязь между средним радиусом инерции в МД траектории,  $R_g$ , и гидродинамическим радиусом,  $R_h$ , известным из результатов запуска алгоритма HYDROPRO [71]. Согласно рекомендациями авторов перечисленных подходов, полученные значения для гидродинамического радиуса  $R_h$  были преобразованы в соответствующие значения коэффициентов трансляционной диффузии  $D_{tr}$  по формуле Эйнштейна-Стокса.

$$D_{tr} = \frac{k_b T}{6\pi\eta_w R_h} \quad (1.3)$$

где  $k_b$  константа Больцмана,  $T$  температура, а  $\eta_w$  вязкость воды. Другое определение  $D_{tr}$  через коэффициент диффузии референсной молекулы, 1,4-диоксана по формуле

$$D_{tr} = D_{tr}^{dioxane} \frac{R_h^{dioxane}}{R_h} \quad (1.4)$$

где гидродинамический радиус диоксана предполагается равным 2.27 Å [73], а коэффициент диффузии диоксана в воде  $D_{tr}^{dioxane}$  при температуре 298 К равным  $1.10 \times 10^{-9} \text{ m}^2/\text{c}$  [147]. Формула Kirkwood-Riseman была использована в следующем виде [148]

$$R_h = 1 \left\langle \frac{1}{N^2} \sum_{i \neq j} \frac{1}{r_{ij}} \right\rangle \quad (1.5)$$

где  $r_{ij}$  — расстояние между  $C\alpha$  атомами для  $i$  и  $j$  аминокислотных остатков,  $N$  — количество остатков в последовательности неупорядоченного белка, а угловые скобки обозначают усреднение по всем кадрам из МД траектории. Преобразование  $R_h$  в  $D_{tr}$  было выполнено согласно формуле (1.3) или формуле (1.4)

### 1.3.3 ЯМР эксперимент

Пептид N-H4 был синтезирован компанией Permic Co. Ltd (Сучжоу, Китай). Образец с концентрацией пептида N-H4 равной 1 mM был приготовлен в 20 mM буфере NaAc-d3, содержащем 5 % D<sub>2</sub>O (pH 4.0). Белок Ub экспрессирован рекомбинантным образом и очищен по ранее описанному протоколу [149]. Далее, для приготовления образца белка Ub с концентрацией 0.23 mM лиофильно высушенный убиквитин растворяли в 20 mM буфере NaAc-d3, содержащем 5 % D<sub>2</sub>O (pH 6.0). Низкая концентрация белка Ub была выбрана для того, чтобы уменьшить долю димеров в образце [150]. Все диффузионные эксперименты с пептидом N-H4 и белком Ub проводились в ампулах Шигеми объемом 250 мл [151].

Все измерения коэффициента трансляционной диффузии были проведены на спектрометре Bruker Avance III 500 МГц, оборудованном датчиком BBI (для ампул диаметром 5 мм) с катушками для создания импульсных градиентов магнитного поля по оси Z. Эксперименты проводились при 298 К для пептида N-H4 и 303 К для белка Ub. Коэффициент трансляционной диффузии был измерен в экспериментах с двойным стимулированным эхом (от англ. double-stimulated echo, DSTe) для компенсации тепловой конвекции [152] с применением последовательности 3-9-19 WATERGATE [153] для подавления сигнала растворителя (код импульсной последовательности основан на стандартных импульсных программах фирмы Bruker dstebpgp3s и stebpgp1s19). Все градиентные импульсы имели сглаженную прямоугольную форму Bruker SMSQ10.100. Длительность каждой компоненты биполярного кодирующего



или декодирующего импульсов была установлена равной 2.7 мс. Всего было записано двадцать спектров с амплитудами импульсных градиентов от 2.4 до 50 Г/см. Диффузионная задержка  $\Delta$  между градиентными импульсами составляла 0.1 с, длительность регистрируемого сигнала спада свободной индукции равнялась 3.6 с, релаксационная задержка составляла 5 с.

В качестве области интегрирования для получения коэффициентов диффузии была выбрана область, соответствующая сигналам от алифатической части белка в отдалении от резонансного сигнала воды. Для пептида N-H4 несколько спектральных линий в области химических сдвигов от 0.93 до 0.73 м.д., а для белка Ub одна спектральная линия с химическим сдвигом -0.17 м.д. Оценка интегральной интенсивности сигналов была осуществлена с помощью нового алгоритма, разработанного в нашей лаборатории [154], реализация которого доступна через веб-сервер DDfit (Diffusion Data fit), <https://ddfit.bio-nmr.spbu.ru>. Далее, значения  $D_{tr}$  были рассчитаны по формуле Стейскала-Таннера [155] с поправками Ершова-Мюллера [152].

Экспериментальные значения были скорректированы с учетом присутствия в образце небольшого количества  $D_2O$  и ацетата натрия, пользуясь тем фактом, что коэффициент  $D_{tr}$  обратно пропорционален вязкости растворителя (по уравнению Эйнштейна-Стокса). Кроме того, для убиквитина результат был скорректирован на присутствие димерных форм в образце (вклад димеров в ЯМР сигнал составил 8 % по отношению к сигналу разбавленного образца) [150]. Поправка составила 1.7 % от исходного значения  $D_{tr}$ , в этом смысле выбранные экспериментальные условия обеспечивают практически полное отсутствие чувствительности к небольшой доле димеров в образце. Таким образом, скорректированные значения  $D_{tr}$  соответствуют мономерной форме белка, растворенного в чистой воде, и тем самым являются пригодными для сравнения с результатами МД расчетов.

Значение  $D_{rot}$  для белка Ub было взято из ранее опубликованной работы, в которой был собран уникальный набор данных для  $^{15}N$  скоростей релаксации в нескольких магнитных полях для образцов с различными концентрациями белка, включая сильно разбавленный образец [103]. Для выбранных условий эксперимента (рН 4.5) авторы не обнаружили признаков димеризации белка Ub, однако при более высокой концентрации

зафиксировали неспецифическое взаимодействие молекул белка Ub между собой. Для того чтобы минимизировать влияние эффекта димеризации, было выбрано значение  $D_{rot}$ , полученное для низкой концентрации белка 0.2 мМ. Данный результат был скорректирован по уравнению Эйнштейна-Стокса с учетом экспериментальных условий: температура 296.6 К и присутствие D<sub>2</sub>O. Данные <sup>15</sup>N скоростей релаксации были взяты из ранее опубликованной работы [49].

## 1.4 Заключение

На сегодняшний день, метод молекулярной динамики позволяет получить полноценную модель для длинных пептидов и небольших неупорядоченных белков. Однако подобные вычислительные модели требуют тщательной экспериментальной валидации. Полезным представляется выбрать модельную систему для изучения совместными усилиями нескольких исследовательских групп. На эту роль могли бы подойти RS-пептид [43; 72; 82; 156] или исследованный в настоящей работе пептид N-H4 [49; 99]. Описанная тщательным образом система с внутренней неупорядоченностью послужила бы аналогом хорошо изученных модельных систем глобулярных белков таких, как убиквитин и лизоцим.

Коэффициент трансляционной диффузии  $D_{tr}$ , известный из экспериментов ЯМР с импульсным градиентом магнитного поля (ИГМП), является значимым параметром для валидации моделей неупорядоченных белков. Однако при более детальном рассмотрении можно обнаружить ряд сложностей. Стоит отметить, что коэффициент диффузии  $D_{tr}$  достаточно грубый параметр. Коэффициенты диффузии мономерной и гомодимерной форм белка отличаются на фактор  $2^{1/3}$ , т.е. всего в  $\approx 1.26$  раза. В сущности, коэффициент диффузии оказывается нечувствительным к небольшим вариациям в структуре конформационного ансамбля неупорядоченного белка, т.е. к предпочтению пептидом более компактных или более развернутых форм.

Другая трудность — технически сложная постановка диффузионного эксперимента ЯМР с ИГМП из-за необходимости учитывать влияние

конвекции и искажение базовой линии остаточными сигналами воды [157]. Одновременно с этим, расчет  $D_{tr}$  по данным МД моделирования представляет собой нетривиальную задачу. В настоящей работе показано, для корректной оценки  $D_{tr}$  требуется аккуратный выбор алгоритма термостатирования и использование специальной схемы моделирования в ячейках увеличивающегося размера. И наконец, необходимо знать точные значения вязкости для используемой модели воды.

В рамках описанного исследования на пути к преодолению указанных трудностей, первым делом был реализован и внедрен в пакет программ Амбер алгоритм термостатирования Бусси на основе масштабирования скоростей. Одновременно с этим, была предложена специальная схема фрагментации МД траектории в ячейке малого размера, которая позволяет эффективным образом семплировать конформационное пространство пептида при дальнейшем моделировании в ячейках среднего и большого размера. Как это следует из предварительных данных, время моделирования может быть также сокращено за счет перехода к увеличенному шагу интегрирования уравнений молекулярной динамики равному 4 фс с применением коррекции NMR ко всей моделируемой системе (белок и растворитель). Отдельно в настоящей работе была исследована вязкость моделей воды TIP4P-Ew, TIP4P-D и OPC. Разработанный протокол моделирования был откалиброван на примере МД модели небольшого упорядоченного белка убиквитина.

Центральная темой данной работы стал анализ диффузионных свойств неупорядоченного пептида N-H4 при использовании набора МД траекторий в ансамблях NPT и NVE общей продолжительностью 75 мкс. В результате было установлено, что МД моделирование с использованием классической модели воды TIP4P-Ew ведет к формированию слишком компактного конформационного ансамбля для пептида N-H4. Сделанный вывод также подтверждается сравнением расчетных и экспериментальных значений  $^{15}\text{N}$  скоростей релаксации. Действительно, в случае МД моделирования в модели воды TIP4P-Ew пептид N-H4 проявляет склонность к формированию устойчивых конформаций (так называемый «мини-фолд»), топология которых благодаря случайному набору водородных связей остается неизменной в течение нескольких сотен наносекунд.

С другой стороны, моделирование как с использованием модели воды TIP4P-D, так и с использованием модели воды OPC, ведет к конформационному ансамблю пептида N-H4, который хорошо согласуется с экспериментальными данными для коэффициента трансляционной диффузии  $D_{tr}$ . Дело в том, что параметр  $D_{tr}$  оказывается недостаточно чувствительным к небольшим отличиям между двумя конформационными ансамблями (средние значения радиуса инерции для моделей воды TIP4P-D и OPC равны 15.3 и 13.8 Å). Анализ скоростей  $^{15}\text{N}$  релаксации также не позволяет выявить лучшую модель, а именно: обе траектории в воде TIP4P-D и OPC демонстрируют хорошее согласие с экспериментальными данными. Таким образом, на основе имеющейся информации невозможно установить, какую модель воды TIP4P-D или OPC следует предпочесть для моделирования неупорядоченных белков. Для ответа на этот вопрос требуются дополнительные данные.

В качестве альтернативы строгой процедуре расчета  $D_{tr}$  по данным МД моделирования на практике также используются различные эмпирические методы. Распространенным подходом является применение алгоритма HYDROPRO для вычисления  $D_{tr}$  по набору кадров из МД траектории. В настоящей работе было показано, что подобные эмпирические методы могут приводить к неверным выводам. Например, согласно HYDROPRO модель воды TIP4P-Ew является наиболее подходящей для моделирования пептида N-H4, что противоречит результатам строгого расчета. Основная причина в том, что алгоритм HYDROPRO был сконструирован для оценки диффузионных свойств жестких биологических молекул и не был задуман, как инструмент для изучения белков с внутренней неупорядоченностью.

Совсем недавно были предложены эмпирические подходы для оценки коэффициента трансляционной диффузии неупорядоченных белков по модели конформационного ансамбля. Соответствующие алгоритмы были откалиброваны на наборах экспериментальных данных малоуглового рентгеновского рассеяния и данных диффузионных экспериментов ЯМР. В настоящей работе было установлено, что новые алгоритмы также склонны давать неточные оценки, вероятно, из-за погрешности в экспериментальных данных, которая может быть связана с использованием 1,4-диоксана в качестве эталонной молекулы [73]. В этом смысле, коэффициент

трансляционной диффузии, рассчитанный напрямую из данных МД моделирования, является важным параметром для валидации и уточнения моделей конформационных ансамблей неупорядоченных белков.

## 1.5 Дополнительная информация

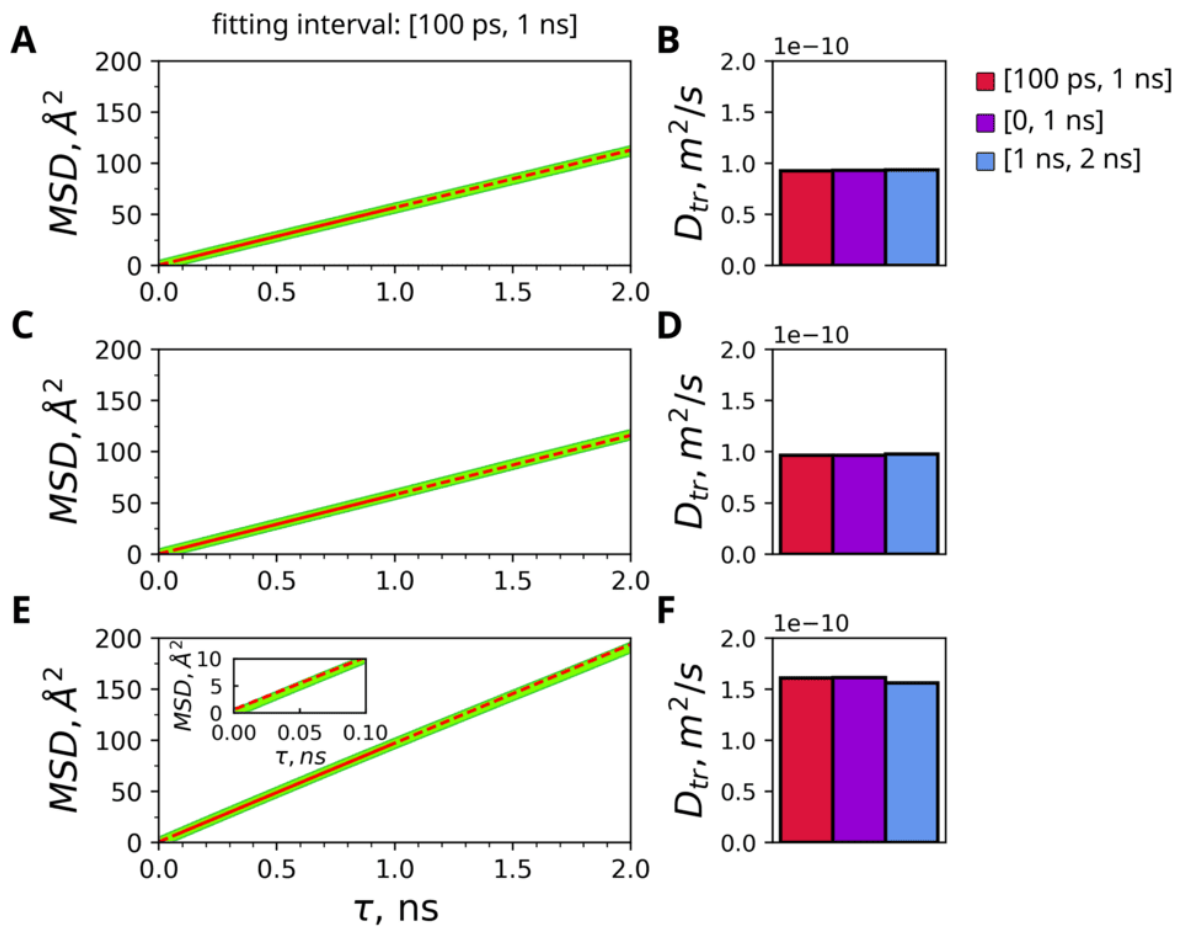


Рисунок 1.9 — Коэффициент трансляционной диффузии пептида N-H4, рассчитанный по данным МД моделирования путем анализа зависимости  $MSD(\tau)$ . Аппроксимация кривых  $MSD(\tau)$  (широкие зеленые линии), вычисленных по формуле (1.1), была осуществлена с помощью линейной функции  $D_{tr} \cdot (\tau + b)$ . Наилучший вариант функции подгонки на заданном интервале  $\tau$  показан сплошной красной линией (за пределами интервала  $\tau$  пунктирной линией). В левой колонке приведены примеры поведения функции подгонки на интервале [100 пс, 1 нс]. В правой колонке приведены сведения о значениях  $D_{tr}$ , полученных с использованием различных интервалов для подгонки (расшифровка цветовой кодировки содержится в условных обозначениях на рисунке). (А, В) МД траектория, записанная с использованием модели воды TIP4P-D в ансамбле NPT для ячейки моделирования среднего размера. (С, D) МД траектория, записанная с использованием модели воды TIP4P-D в ансамбле NVE для ячейки моделирования среднего размера. (Е, F) МД траектория, записанная с

использованием модели воды TIP4P-Ew в ансамбле NPT для ячейки моделирования большого размера. Специальная вставка на панели (Е) иллюстрирует начальный участок кривой  $MSD(\tau)$  совместно с наилучшим вариантом функции подгонки. Для целей построения данного рисунка были выбраны МД траектории, которые демонстрируют наиболее сильную зависимость от выбранного интервала для подгонки.

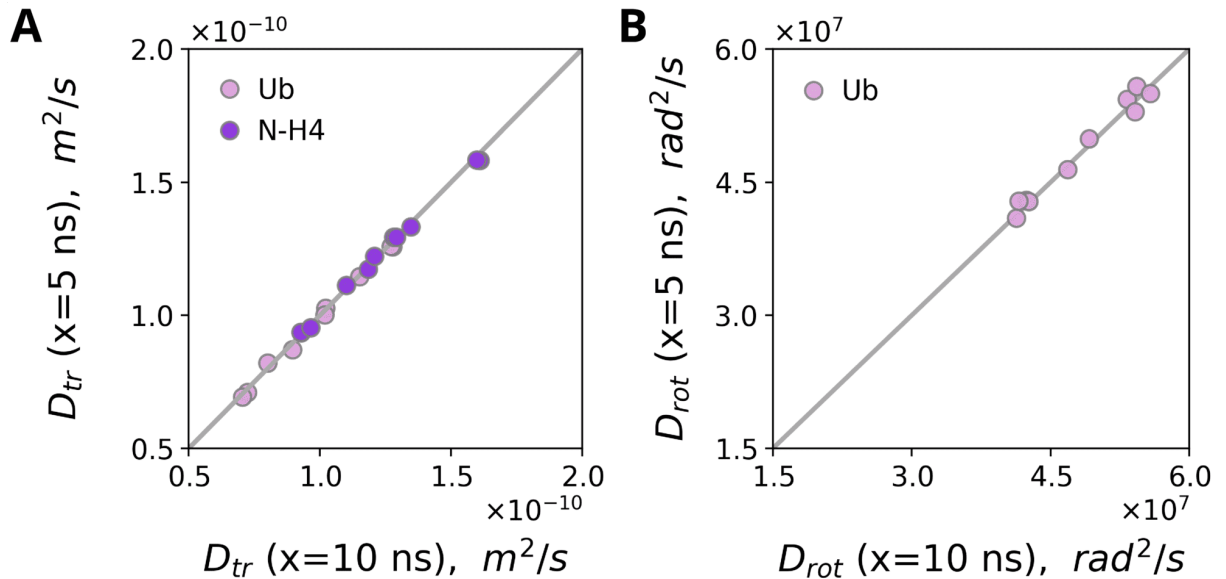


Рисунок 1.10 — (А) Корреляция между значениями  $D_{tr}$ , полученными в результате обработки набора МД траекторий по схеме МД моделирования  $x = 5$  нс и  $x = 10$  нс. Точки на графике относятся к результатам на наборах коротких траекторий в случае пептида N-H4 и белка Ub (фиолетовые и розовые символы соответственно) в ячейках моделирования среднего и большого размеров с использованием ансамблей NPT и NVE. (В) Корреляция между значениями  $D_{rot}$ , полученными в результате обработки набора МД траекторий по схеме моделирования  $x = 5$  нс и  $x = 10$  нс. Точки на графике относятся к результатам на наборах коротких траекторий в случае пептида N-H4 и белка Ub (фиолетовые и розовые символы соответственно) в ячейках моделирования среднего и большого размеров с использованием ансамблей NPT и NVE.

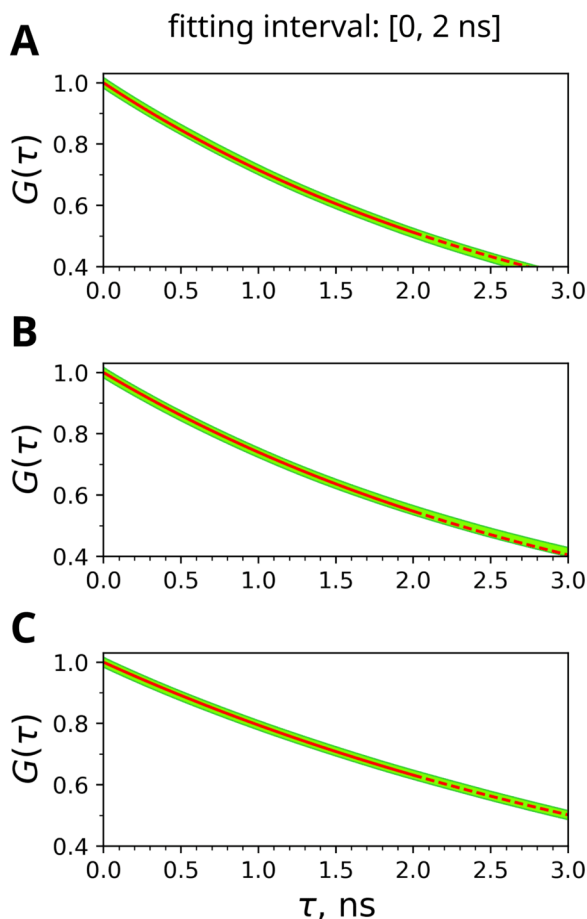


Рисунок 1.11 — Коэффициент вращательной диффузии  $D_{rot}$  белка Ub, рассчитанный по данным МД моделирования путем анализа корреляционных функций  $G(\tau)$  (см. в разделе Материалы и методы). Аппроксимация корреляционных функций  $G(\tau)$  (широкие зеленые линии) была осуществлена с помощью моноэкспоненциальной функции подгонки  $\exp(-6D_{rot}\tau)$ . Наилучший вариант функции подгонки на заданном интервале  $\tau$  [0,2 нс] показан сплошной красной линией (за пределами интервала  $\tau$  пунктирной линией). (А) Фрагментированная МД траектория ( $x = 10$  нс), записанная с использованием модели воды TIP4P-Ew в ансамбле NPT для ячейки моделирования большого размера. (В) непрерывная МД траектория, записанная с использованием модели воды TIP4P-Ew в ансамбле NVE для ячейки моделирования малого размера. (С) непрерывная МД траектория, записанная с использованием модели воды TIP4P-D в ансамбле NPT для ячейки моделирования малого размера.



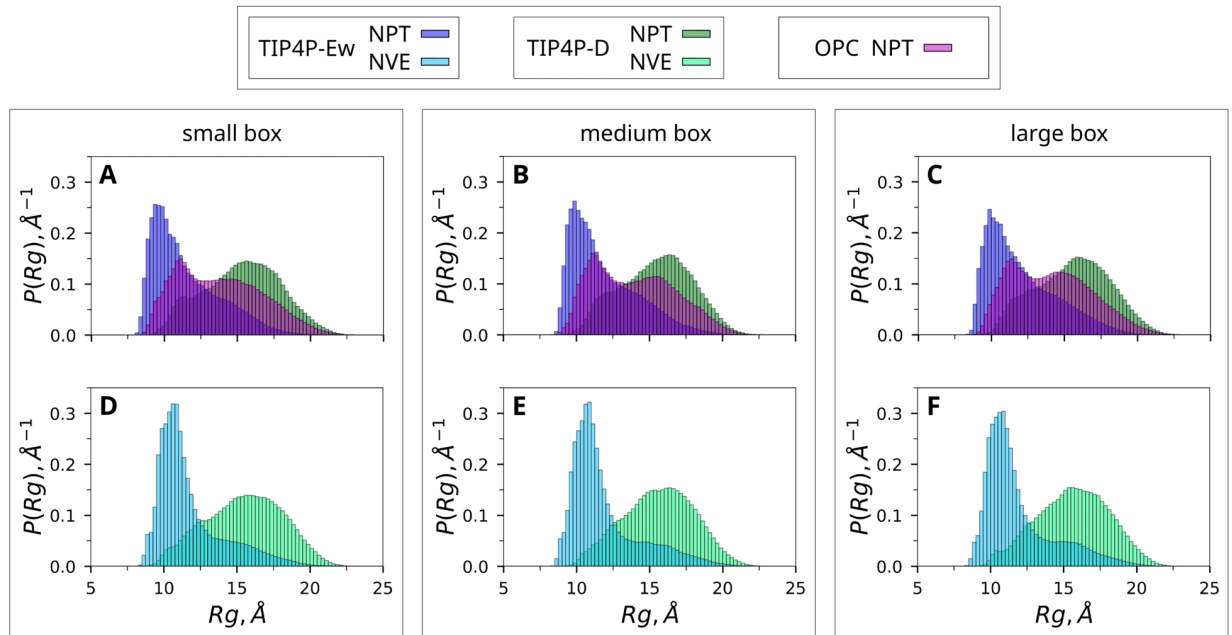


Рисунок 1.12 — Распределения по радиусам инерции  $P(R_g)$  для пептида N-H4, по данным МД траекторий, записанных в моделях воды TIP4P-Ew, TIP4P-D и OPC с использованием ансамбля NPT совместно с термостатом Бусси: (A) МД моделирование в ячейках малого размера (соответствует Рисунку 1.3); (B) МД моделирование в ячейках среднего размера; (C) МД моделирование в ячейках большого размера. Кроме того, на рисунке представлены распределения по радиусам инерции  $P(R_g)$  для пептида N-H4, по данным МД траекторий, записанных в моделях воды TIP4P-Ew и TIP4P-D с использованием ансамбля NVE: (A) МД моделирование в ячейках малого размера; (B) МД моделирование в ячейках среднего размера; (C) МД моделирование в ячейках большого размера. Расшифровка цветовой кодировки содержится в условных обозначениях на рисунке.

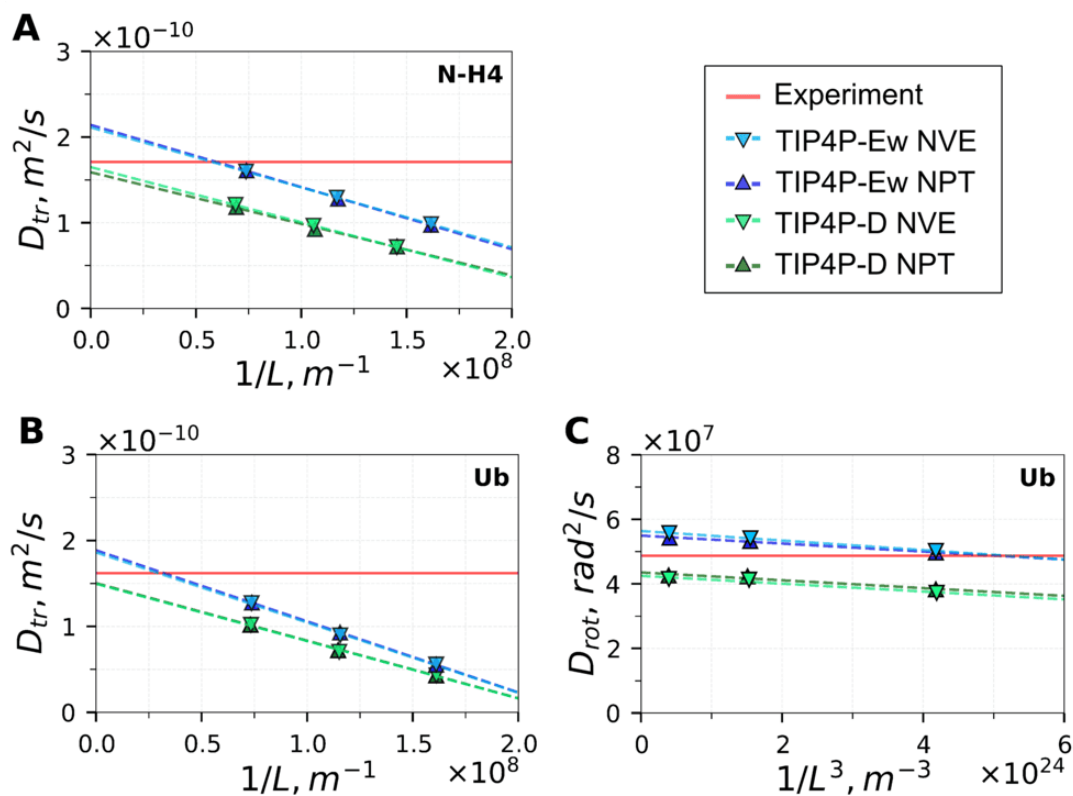


Рисунок 1.13 — Коэффициенты диффузии (А) пептида N-H4 и (В, С) белка Ub, следующие из анализа МД траекторий, записанных в моделях воды TIP4P-Ew и TIP4P-D. Результаты представлены для моделирования в ансамбле NPT (темно-синие и темно-зеленые символы; соответствует 1.4) и ансамбле NVE (светло-синие и светло-зеленые символы). Протокол обработки данных не отличается от использованного при подготовке Рисунка 1.4.

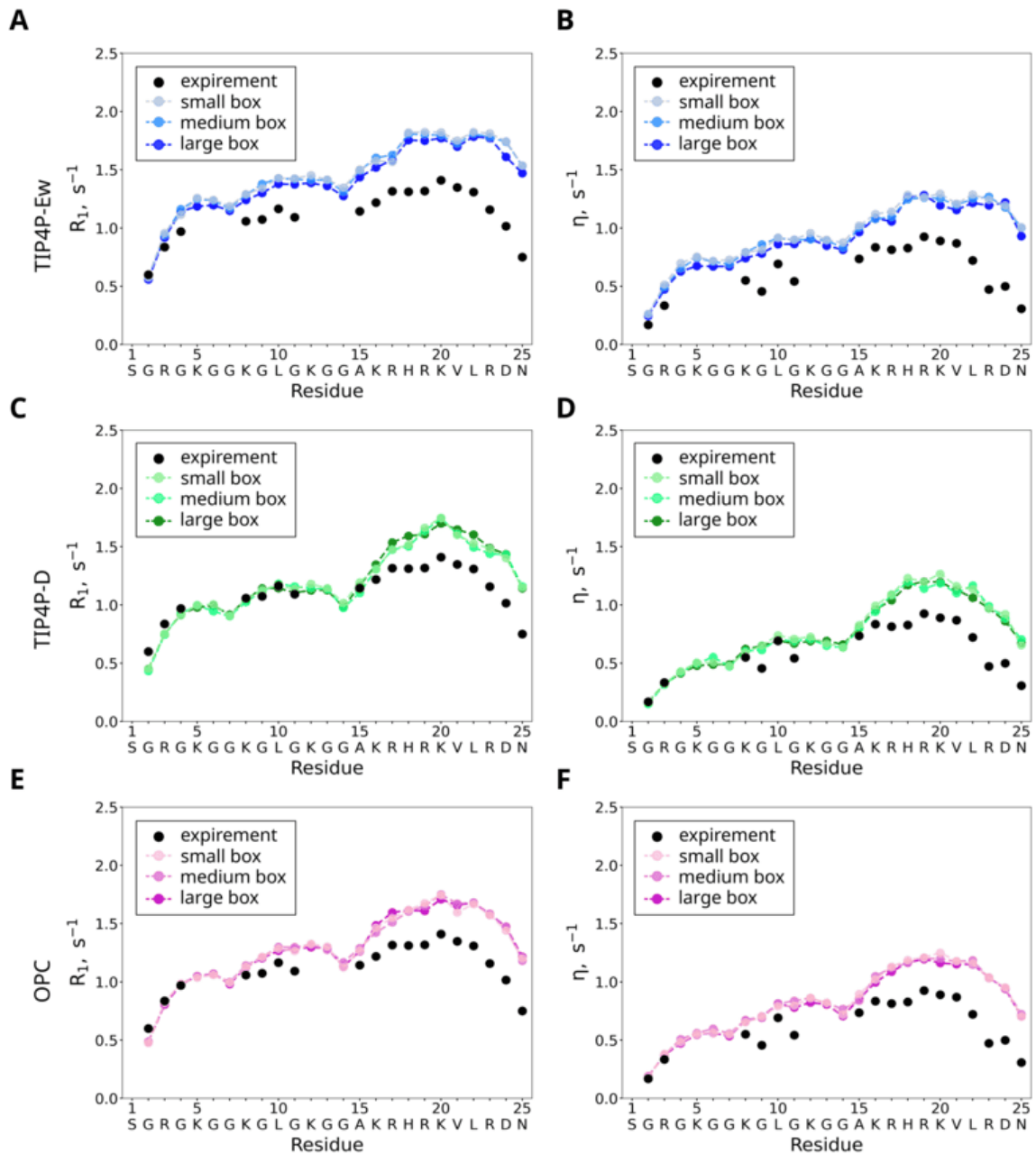


Рисунок 1.14 — Экспериментальные и расчетные скорости  $^{15}\text{N}$  спиновой релаксации  $R_1$  и  $\eta$  в пептиде N-H4, следующие из анализа МД траекторий, записанных в моделях воды: (A, B) TIP4P-Ew, (C, D) TIP4P-D и (E, F) OPC. Результаты моделирования в ячейках увеличивающегося размера для ансамбля NPT проиллюстрированы различными оттенками одного и того же цвета, как указано в условных обозначениях на рисунке.

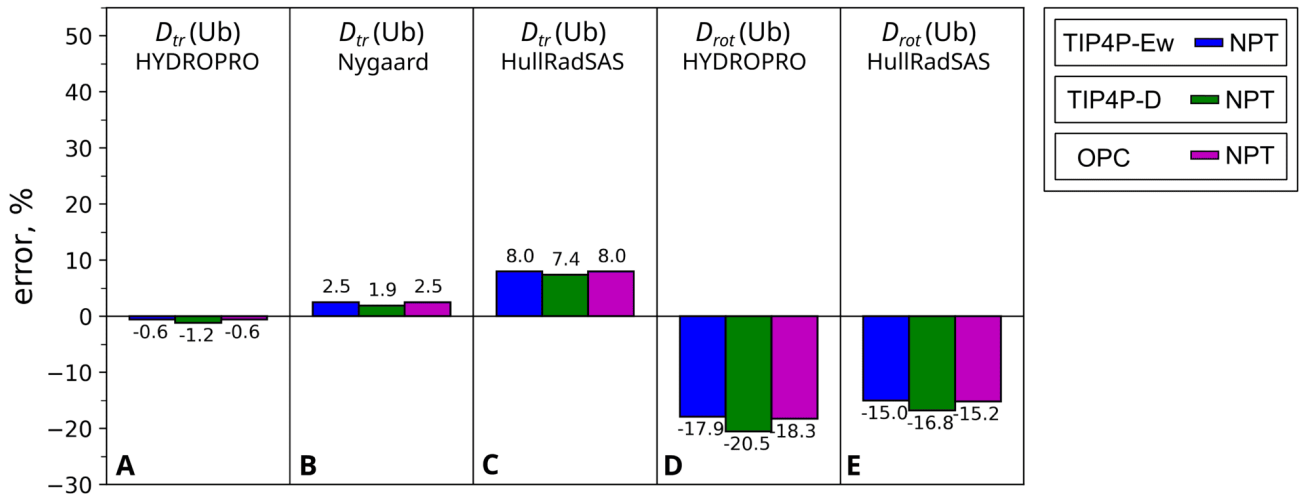


Рисунок 1.15 — Оценка расхождений между экспериментальными и расчетными значениями коэффициентов диффузии для белка Ub. Расчеты проведены для трех ансамблей конформеров, построенных на основе 2000 кадров, взятых из МД траектории в ячейке малого размера, записанных с использованием моделей воды TIP4P-Ew, TIP4-D и OPC и ансамбля NPT (синие, зеленые и пурпурные столбцы соответственно). Показанные на рисунке значения  $D_{tr}$  получены с помощью следующих инструментов: А) программа HYDROPRO; (В) эмпирическая формула Nygaard, опирающаяся на результаты HYDROPRO (С) программа HullRadSAS. Для методов (В) и (С) полученные значения  $R_h$  были преобразованы в  $D_{tr}$  по формуле Эйнштейна-Стокса (1.3). Альтернативный метод (см. формулу (1.4)), отталкивающийся от данных для коэффициента диффузии эталонной молекулы, дает схожие результаты (не представлены). Показанные на рисунке значения  $D_{rot}$  получены с помощью следующих вычислительных инструментов: (D) программы HYDROPRO и (E) программы HullRadSAS.

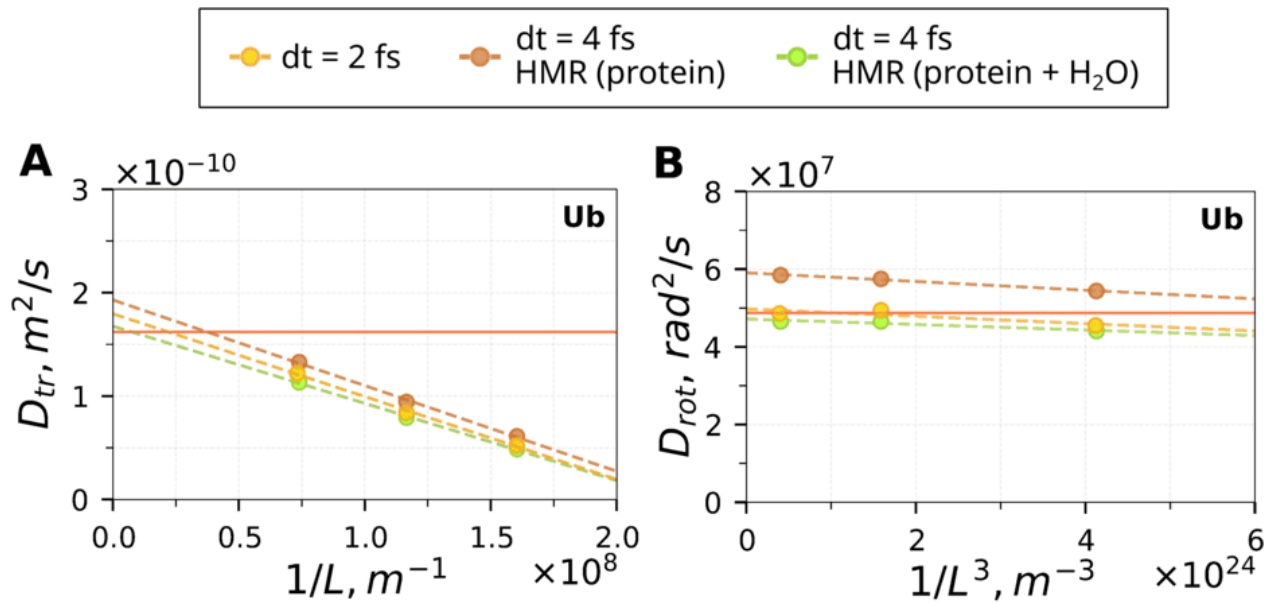


Рисунок 1.16 — (А) Коэффициент трансляционной и (В) вращательной диффузии белка Ub, следующие из анализа МД траекторий, записанных в модели воды ОРС для увеличенного шага интегрирования уравнений молекулярной динамики:  $\Delta t = 2$  фс (оранжевые символы),  $\Delta t = 4$  фс со схемой HMR, приложенной только к белку (коричневые символы), и  $\Delta t = 4$  фс со схемой HMR, приложенной ко всей системе (белок Ub и растворитель; светло-зеленые символы).

Таблица 4 — Список МД траекторий для пептида N-H4

Белок	Модель воды	Ансамбль	Ячейка моделирования	Толщина водного слоя	МД протокол
N-H4	TIP4P-Ew	NPT	малая	12 Å	1 × 5000 нс
			средняя	24 Å	500 × 10 нс
			большая	48 Å	500 × 10 нс
		NVE	малая	12 Å	1 × 5000 нс
			средняя	24 Å	500 × 10 нс
			большая	48 Å	500 × 10 нс
	TIP4P-D	NPT	малая	15 Å	1 × 5000 нс
			средняя	24 Å	500 × 10 нс
			большая	48 Å	500 × 10 нс
		NVE	малая	15 Å	1 × 5000 нс
			средняя	24 Å	500 × 10 нс
			большая	48 Å	500 × 10 нс
	OPC	NPT	малая	15 Å	1 × 5000 нс
			средняя	24 Å	500 × 10 нс
			большая	48 Å	500 × 10 нс

Таблица 5 — Список МД траекторий для белка Ub

Белок	Модель воды	Ансамбль	Ячейка моделирования	Толщина водного слоя	МД протокол
Ub	TIP4P-Ew	NPT	малая	12 Å	1 × 2000 нс
			средняя	24 Å	200 × 10 нс
			большая	48 Å	200 × 10 нс
		NVE	малая	12 Å	1 × 2000 нс
			средняя	24 Å	200 × 10 нс
			большая	48 Å	200 × 10 нс
	TIP4P-D	NPT	малая	12 Å	1 × 2000 нс
			средняя	24 Å	200 × 10 нс
			большая	48 Å	200 × 10 нс
		NVE	малая	12 Å	1 × 2000 нс
			средняя	24 Å	200 × 10 нс
			большая	48 Å	200 × 10 нс
	OPC	NPT	малая	12 Å	1 × 2000 нс
			средняя	24 Å	200 × 10 нс
			большая	48 Å	200 × 10 нс

Таблица 6 — Коэффициенты диффузии пептида N-H4 и белка Ub, известные из эксперимента и рассчитанные по данным МД моделирования. Расчет коэффициентов трансляционной диффузии осуществлялся напрямую по данным МД моделирования путем анализа  $MSD(\tau)$  с учетом зависимости полученных значений от размера ячейки моделирования (см. Рисунок 1.4 для моделирования в ансамбле NPT и Рисунок 1.13 для моделирования в ансамбле NVE). Погрешность экспериментальных данных  $D_{tr}$  определялась по значениям ковариационной матрицы [158], рассчитанной с помощью функции `curve_fit` в библиотеке `ScuPy` [144] на базе сервера `DDfit`. Погрешность экспериментальных данных  $D_{rot}$  была взята из ранее опубликованной работы [103]. Погрешность расчетных значений  $D_{tr}$  и  $D_{rot}$  была получена при помощи метода «складного ножа» (от англ. `jackknife`), в котором из МД расчетов случайным образом исключается 20 % выборки (см. раздел 1.3).

Белок	Коэффициент диффузии	Эксперимент	МД моделирование					
			TIP4P-Ew		TIP4P-D		OPC	
			NPT	NVE	NPT	NVE	NPT	NPT
N-H4	$D_{tr}$ ( $10^{-10}$ m <sup>2</sup> /c)	$1.71 \pm 0.01$	$2.14 \pm 0.02$	$2.11 \pm 0.02$	$1.59 \pm 0.01$	$1.65 \pm 0.03$	$1.81 \pm 0.02$	
Ub	$D_{tr}$ ( $10^{-10}$ m <sup>2</sup> /c)	$1.62 \pm 0.04$	$1.89 \pm 0.02$	$1.86 \pm 0.02$	$1.50 \pm 0.02$	$1.50 \pm 0.02$	$1.69 \pm 0.02$	
	$D_{rot}$ ( $10^7$ rad <sup>2</sup> /c)	$4.87 \pm 0.02$	$5.50 \pm 0.04$	$5.64 \pm 0.05$	$4.35 \pm 0.04$	$4.25 \pm 0.04$	$4.94 \pm 0.05$	



## Глава 2. Парамагнитное усиление ядерной спиновой релаксации: исследование хвоста гистона H4 в составе нуклеосомной частицы методами МД моделирования и ЯМР спектроскопии

### 2.1 Введение

Хроматин — весьма динамичная структура, представленная в клетках эукариот в виде нуклеопротеинового комплекса, который обеспечивает хранение, реализацию и передачу генетической информации. Базовой повторяющейся единицей хроматина является нуклеосомная коровая частица (от англ. nucleosome core particle, NCP), которая состоит из 147 пар нуклеотидов двойной спирали ДНК, закрученных  $\sim 1.7$  раз вокруг октамера, образованного двумя копиями каждого из гистонов H2A, H2B, H3 и H4. В свою очередь, каждый гистон содержит ядро — структурированный глобулярный домен, и гистоновый хвост — положительно заряженный N-концевой домен, длиной от 15 до 35 аминокислотных остатков. В ранних экспериментальных работах, основанных на расщеплении пептидных связей и использовании метода ядерного магнитного резонанса (ЯМР), были получены первые доказательства того, что N-концевые гистоновые хвосты выступают с поверхности нуклеосомы и являются чрезвычайно подвижными [159; 160]. Справедливость сделанных предположений стала очевидной после получения методом рентгеноструктурного анализа первой структуры высокого разрешения для нуклеосомной частицы [161]. Дело в том, что при исследовании нуклеосомной частицы методом рентгеноструктурного анализа (а также методом криоэлектронной микроскопии) для самых крайних N-концевых аминокислотных остатков электронная плотность, как правило, отсутствует, а для аминокислотных остатков, расположенных чуть дальше в полипептидной цепочке (ближе к гистоновому ядру), электронная плотность крайне слабая и носит разрозненный характер. В некоторых случаях при взаимодействии гистоновых хвостов с соседними нуклеосомами в кристаллической решетке, отдельные фрагменты могут быть частично разрешены, что до некоторой

степени позволяет восстановить положение гистоновых хвостов (см. Рисунок 2.1 (С)) [161; 162].

Тот факт, что гистоновые хвосты не обладают выделенной конформацией в пространстве, чрезвычайно подвижны и выступают с поверхности нуклеосомы в сторону растворителя, определяет их функциональную значимость. Гистоновые хвосты являются основным местом для прикрепления сотни различных хроматин-ассоциированных белков (от англ. chromatin-associated proteins, CAPs), которые участвуют в сборке и ремоделировании хроматина, активации и подавлении транскрипции [163]. Кроме того, гистон-модифицирующие ферменты (от англ. histone-modifying enzymes, HMEs) могут как вносить, так и убирать различные посттрансляционные модификации (от англ. post-translational modifications, PTMs) в гистоновые хвосты. К числу наиболее распространенных посттрансляционных модификаций относится метилирование остатков лизина и аргинина, а также ацетилирование остатка лизина (добавляет дополнительный отрицательный заряд) [164]. Совокупность посттрансляционных модификаций получила название «гистонового кода» [165], через который реализуется тонкая настройка сродства гистоновых хвостов к белкам CAPs, а также регуляция взаимодействий с соседними нуклеосомами в хроматине [166]. Обобщая, можно сказать, что гистоновые хвосты находятся в центре чрезвычайно обширной и сложной динамической сети взаимодействий, детальное представление о которой (на уровне отдельных атомов), на сегодняшний день, носит довольно ограниченный характер.

В рамках описываемого исследования детальным образом был охарактеризован конформационный ансамбль подвижного неупорядоченного N-концевого фрагмента гистона H4 в составе нуклеосомной частицы. В предыдущих исследованиях методами ЯМР в жидкости и твердом теле было показано, что первые ~15-20 N-концевых остатков гистона H4 в значительной мере неупорядочены и обладают высокой степенью подвижности как в составе отдельных нуклеосом [97], так и в составе объединения нуклеосом в структуры более высокого порядка от незначительного уровня компактизации [167] до высококонденсированного состояния [98]. Благодаря тому, что N-концевой фрагмент гистона H4 положительно заряжен, можно ожидать, что он будет находиться вблизи

отрицательно заряженной нуклеосомной ДНК (нДНК) и взаимодействовать с ней. Однако комбинированный подход, включающий как проведение ЯМР экспериментов по измерению скоростей ядерной спиновой релаксации, так и моделирование методом молекулярной динамики (МД) длительностью несколько микросекунд, показал, что остатки N-концевого фрагмента гистона H4 весьма подвижны и образуют лишь динамические контакты с нДНК (электростатика, образование водородных связей) даже в условиях низкой ионной силы раствора [50]. Подобные динамические взаимодействия получили название «нечетких взаимодействий» (от англ. «fuzzy interaction»), а образующийся при этом комплекс — «нечеткого комплекса» (от англ. «fuzzy complex») [50; 168]. И хотя в предыдущих работах была получена ценная информация о характере взаимодействий хвоста гистона H4 с нДНК, использованные ранее экспериментальные методы сообщали лишь косвенную информацию о положении хвоста гистона H4 относительно поверхности нуклеосомы и нДНК. Локализация хвоста гистона H4 в составе нуклеосомной частицы по большей части остается малоисследованной, хотя важна в контексте понимания доступности хвоста гистона H4 для CAPs (и HMEs) белков, способности формировать контакты с соседними нуклеосомами [166] и вступать в конкуренцию за сайты связывания с хвостом гистона H3 [101; 169].

Как уже было отмечено, структурные методы высокого разрешения, такие как рентгеноструктурный анализ и криоэлектронная микроскопия, оказываются малоэффективными по отношению к гибким гистоновым хвостам. Ценная информация может быть получена при помощи методов ЯМР спектроскопии [50; 97; 98; 100; 101; 167; 168; 170–180]. В частности, полезными оказываются данные по измерению эффекта парамагнитного ускорения ядерной спиновой релаксации (от англ. paramagnetic relaxation enhancement, PRE). Значения PREs могут быть преобразованы в межатомные расстояния (вплоть до 20–30 Å) от амидного протона индивидуального остатка основной цепи до парамагнитной метки, присоединенной ковалентным образом к определенному сайту на поверхности белковой молекулы. Для упорядоченных белковых молекул ограничения на межатомные расстояния, следующие из данных PRE, могут быть использованы для построения и уточнения структурной модели [181–183]. В частности, известны структурные

исследования нуклеосомной частицы (в том числе с CAPs белками), опирающиеся на результаты измерения эффекта PRE [184–186]. В перечисленных работах эффект PRE рассматривался в рамках хорошо известной модели Соломона-Бломбергена (от англ. Solomon-Bloembergen model) [187]. В модели Соломона-Бломбергена эффект PRE определяется произведением двух компонент, зависящих от величины межатомного расстояния и локальной динамики в системе. В свою очередь, параметры, характеризующие локальную динамику, могут быть определены из ЯМР экспериментов по измерению гетероядерных скоростей релаксации [188]. Наряду с моделью Соломона-Бломбергена, были предложены и более сложные модели, в которых дополнительным образом учитывается конформационная динамика гибкого линкера парамагнитной метки [183; 189]. В отличие от свернутых белков, для неупорядоченных белковых доменов, таких как N-концевые фрагменты гистонов в составе нуклеосомы, две компоненты, характеризующие межатомное расстояние и внутреннюю динамику, оказываются коррелированы. Для подобных систем количественный анализ PRE эксперимента зачастую опирается на эмпирические модели, в которых неупорядоченный белок рассматривается, как «случайный клубок» (от англ. random coil) [190–192]. Другой возможный подход предполагает введение допущений при описании сложной динамики гибкой полипептидной цепи [193–195]. В итоге полученная из данных PRE информация может быть использована при построении конформационного ансамбля неупорядоченного белка, состоящего из конечного набора пространственных конфигураций [196–198]. Кроме того, были предложены подходы, позволяющие разделить структурный и динамический вклад в величину эффекта PRE [199; 200].

В более ранних работах по изучению неупорядоченных белков было показано, что значения PREs могут быть вычислены напрямую из траекторий молекулярной динамики по строгим теоретическим выражениям [13; 201]. Однако по историческим причинам использование МД моделирования для изучения неупорядоченных белков было осложнено двумя факторами, а именно: короткие МД траектории и низкое качество силовых полей в отношении моделирования неупорядоченных полипептидов. На настоящий момент, оба фактора перестали быть ограничивающими. С появлением графических ускорителей стало

возможным на регулярной основе записывать длинные МД траектории для частично и полностью неупорядоченных белков в ячейках моделирования большого размера [202]. Одновременно с этим было разработано новое поколение силовых полей (в частности, новые модели воды), предназначенные для моделирования как упорядоченных, так и неупорядоченных белков [203; 204]. В рамках описываемого исследования, были использованы последние достижения в области МД моделирования, чтобы продемонстрировать, что длинные МД траектории N-концевого фрагмента гистона H4 в составе нуклеосомы с нитроксильной парамагнитной меткой, введенной в различные позиции гистона H3, могут быть успешно использованы для анализа PRE эксперимента. В результате была получена релевантная модель конформационного ансамбля и динамических контактов неупорядоченного хвоста гистона H4 в составе нуклеосомы.

## 2.2 Результаты и обсуждение

### 2.2.1 Спин-меченые образцы нуклеосомы

В рамках описываемого исследования было исследовано четыре спин-меченых образца нуклеосомной частицы, каждый из которых был восстановлен из гистонов, полученных рекомбинантным путем в *Xenopus laevis*, и фрагмента ДНК Widom 601 длиной 147 пар оснований нуклеотидов [205]. В последовательности гистона H3 единственный остаток цистеина был заменен на аланин в позиции S110A. Как известно из литературы: данная мутация не влияет на нативную структуру гистона H3 [97; 98]. Далее, на основе полученной конструкции было подготовлено четыре точечных мутанта гистона H3 с цистеином, введенным в один из четырех сайтов: K36C, L65C, K79C и Q125C (см. Рисунок 2.1 (A), более подробное описание приведено в разделе 2.3). Каждый из точечных мутантов гистона H3 совместно с  $^{15}\text{N}$  спин-меченым гистонам H4 и нативными варианты гистонов H2A и H2B были использован для сборки

четырёх вариантов гистонового октамера. Далее к полученным образцам добавлялся избыток предшественника нитроксильной спиновой метки MTSL, вступающий в реакцию присоединения с остатком цистеина в уникальном сайте гистона H3 [206; 207]. Модифицированный в результате реакции присоединения аминокислотный остаток называют R1 в соответствии с установившейся практикой (см. Рисунок 2.1 (B)). Таким образом, было подготовлено четыре варианта октамера гистонов с парамагнитной меткой, введенной в уникальный сайт гистона H3. И наконец, структура нуклеосомной частицы была восстановлена путем смешивания ДНК Widom 601 с конкретным вариантом октамера гистонов. Полученные образцы нуклеосомной частицы были использованы при проведении ЯМР экспериментов. Для удобства введены краткие обозначения образцов: 36R1, 65R1, 79R1 и 125R1.

Важно отметить, что все сайты в последовательности гистона H3, выбранные для введения точечных мутаций, находятся на поверхности нуклеосомной частицы вблизи нДНК. В их число входят: сайт K36 в неупорядоченном регионе, предшествующем  $\alpha$ N спирали, два сайта L65 и K79 на конце  $\alpha$ 1 спирали, и сайт Q125 в центре спирали  $\alpha$ 3. Потенциальные возмущения в структуре нуклеосомной частицы были сведены к минимуму, благодаря использованию протокола, который предполагает присоединение парамагнитной метки к восстановленному октамеру гистонов, а не к индивидуальной молекуле гистона H3. Точность процедуры мечения была подтверждена при помощи масс-спектрометрии и гель-электрофореза (см. Рисунок 2.3 и Рисунок 2.4). Весьма важно, что перечисленные сайты находятся вблизи структурированных участков в гистоне H3 и тем самым представляют собой набор референсных точек в структуре нуклеосомной частицы, относительно которых удобно исследовать конформационную динамику хвоста гистона H4 (см. Рисунок 2.1 (C)). Выбранные референсные точки находятся на расстоянии 20-60 Å друг от друга и тем самым обеспечивают получение независимых наборов данных о пространственной локализации хвоста гистона H4.

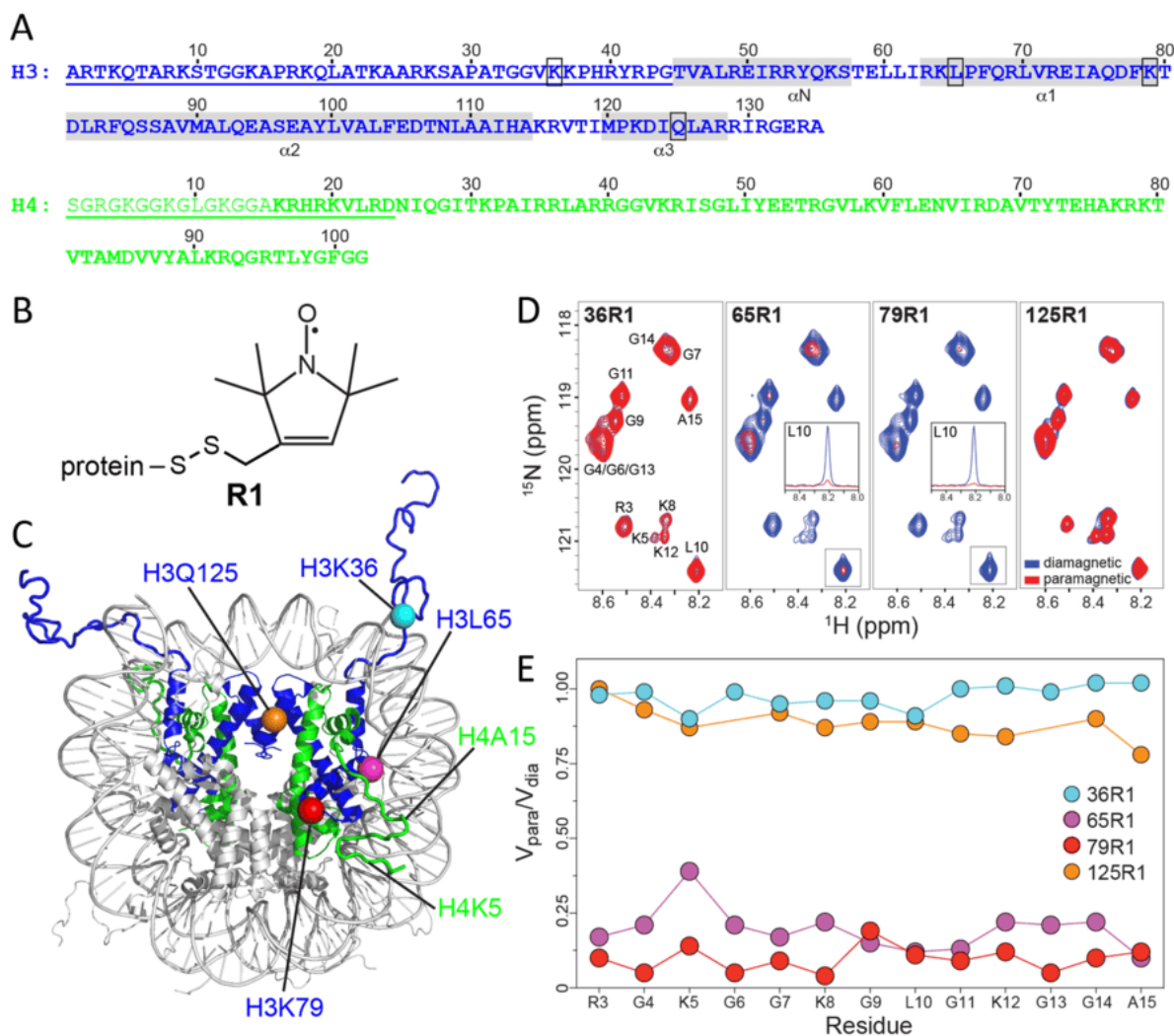


Рисунок 2.1 — Измерение эффекта PRE для аминокислотных остатков хвоста гистона Н4 в нуклеосоме. (А) Аминокислотная последовательность для гистонов Н3(C110А) и Н4 взята из *Xenopus laevis*. На рисунке подчеркнуты остатки, принадлежащие неупорядоченным фрагментам, для которых по данным рентгеноструктурного анализа наблюдается низкая электронная плотность и высокие значения В-факторов (структура 1КХ5 из базы данных PDB) [162]. Серыми прямоугольниками обозначены спиральные участки гистона Н3. Черной рамкой обведены остатки, предназначенные для введения парамагнитной метки. Мелким шрифтом выделены остатки, принадлежащие гибкой части гистона Н4 вплоть до остатка А15. (В) Структурная формула парамагнитной спиновой метки (MTSL), вступающей в реакцию присоединения с остатком цистеина белковой молекулы. Модифицированная боковая цепь в соответствии с общепринятой практикой обозначается как R1 [206; 207]. (С) Кристаллическая структура нуклеосомы (PDB ID: 1КХ5). Гистон Н3 окрашен в синий цвет, гистон Н4 в зеленый цвет. Сайты

введения парамагнитной метки для одной из копий гистона Н3 обозначены пронумерованными сферами ( $C^\alpha$  атомы соответствующих аминокислотных остатков). (D)  $^{15}N$ - $^1H^N$  HSQC ЯМР спектры хвоста гистона Н4 в составе нуклеосомы, содержащей MTSL-меченый гистон Н3 и  $^{15}N$ -меченый гистон Н4. Также показано наложение спектра парамагнитного образца (красный цвет) на референсный спектр диамагнитного образца (синий цвет); выбран одинаковый уровень отсечки для всех образцов нуклеосомы (см. верхнюю часть каждой панели). Диамагнитные образцы были получены путем добавления аскорбата натрия к парамагнитному образцу, меченному нитроксильной спиновой меткой. Во всех ЯМР спектрах пик A15 и пики глицинов записаны отраженными вдоль оси  $^{15}N$  относительно верхнего края спектра. На вставке показан одномерный спектр для остатка L10 в образцах 65R1 и 79R1. (E) Отношение объемов пиков парамагнитного образца к диамагнитному  $V_{para}/V_{dia}$ , которое определяет значение PRE (см. Таблицу 7). Благодаря высокому отношению сигнал / шум, погрешность измерений не превышает 0.01, так что доверительные интервалы не выходят за границы символов на рисунке. Цветовая кодировка соответствует панели (C).

### 2.2.2 Измерения PRE в спин-меченых образцах нуклеосомы

Измерение эффекта PRE для остатков спин-меченого хвоста гистона Н4 под влиянием парамагнитного центра нитроксильной метки, присоединенной к гистону Н3, было проведено с помощью метода гетероядерной одноквантовой корреляционной спектроскопии (от англ. heteronuclear single quantum coherence, HSQC). Для этого был записан HSQC спектр для парамагнитного образца, меченного нитроксильной спиновой меткой, и HSQC спектр для диамагнитного образца. Диамагнитный образец получали путем добавления аскорбата натрия к парамагнитному образцу, в результате чего происходила нейтрализация нитроксильной группы до гидроксиламина (см. раздел 2.3).



В HSQC спектре  $^{15}\text{N}$  меченого гистона Н4 в составе нуклеосомной частицы (см. Рисунок 2.1 (D)) наблюдается ограниченное количество пиков, а именно: все наблюдаемые сигналы соответствуют наиболее подвижным остаткам хвоста гистона Н4 (аминокислотные остатки  $\sim 1-15$ ; важно отметить, что для двух последних N-концевых остатков не регистрируется сигнал ЯМР из-за обмена протона амидной группы с растворителем [97]). При этом в HSQC спектре регистрируется единственный набор пиков для двух N-концевых фрагментов гистона Н4 в составе нуклеосомной частицы, что указывает на отсутствие существенных структурных и динамических различий между двумя копиями хвостов. Динамика остальных аминокислотных остатков (ближе к гистоновому ядру и в самом гистоновом ядре) определяется медленным вращением нуклеосомной частицы ( $\sim 160$  нс [50]), что препятствует наблюдению сигнала от таких остатков в спектре ЯМР (соответствующие пики сильно уширены).

Из наложения HSQC спектра для парамагнитного образца на референсный HSQC спектр для диамагнитного образца нуклеосомной частицы (см. Рисунок 2.1 (D)) можно заметить, что внедрение парамагнитной метки в позиции 65 и 79 гистона Н3 приводит к резкому уменьшению интенсивности сигнала для большинства аминокислотных остатков хвоста гистона Н4. С другой стороны, внедрение парамагнитной метки в позиции 36 и 125 гистона Н3 способствует лишь незначительному снижению интенсивности сигнала для аминокислотных остатков хвоста Н4. Данный результат легко понять на качественном уровне, если принять во внимание тот факт, что сайты 65 и 79 в гистоне Н3 расположены в непосредственной близости от точки прикрепления N-концевого фрагмента хвоста гистона Н4 к гистоновому ядру. В то время как сайты 36 и 125 в гистоне Н3 удалены от хвоста гистона Н4 (аминокислотный остаток 36 гистона Н3 отделен от хвоста гистона Н4 витком нуклеосомной ДНК). Для каждого остатка хвоста гистона Н4 в каждом исследуемом образце значения PREs определялись из отношения объёмов пиков,  $V_{para}/V_{dia}$ , по ранее описанному протоколу [201] (см. Рисунок 2.1Е, а также раздел 2.3 и Таблицу 7).

### 2.2.3 Расчет эффекта PRE из МД траекторий

Экспериментальные данные PRE были использованы для валидации двух МД моделей нуклеосомной частицы. Первая модель представляет ранее записанную нами траекторию длиной 2 мкс в силовом поле Amber ff14SB [67] с использованием модели воды TIP4P-D [18]. В предыдущем исследовании было показано, что скорости  $^{15}\text{N}$  релаксации для аминокислотных остатков N-концевого фрагмента гистона H4, рассчитанные по данной траектории, хорошо согласуются с экспериментальными значениями [50]. Вторая модель представляет собой набор МД траекторий нуклеосомной частицы суммарной длиной в 41 мкс [102], который был записан в лаборатории Панченко с использованием того же силового поля Amber ff14SB совместно с моделью воды OPC [208]. Модель воды TIP4P-D является специализированной моделью, предназначенной для моделирования неупорядоченных пептидов (таких как хвост гистона H4). В противоположность этому, модель воды OPC изначально создавалась как модель общего назначения и лишь позднее зарекомендовала себя в задачах моделирования систем с внутренней неупорядоченностью [45]. Процедура расчета значений PRE по данным МД моделирования подробно описана в разделе 2.3. Следует отметить, что выполненные в настоящей работе вычисления опираются на строгие выражения теории Редфилда, в которых явным образом учитывается динамика ядерных спинов  $^1\text{H}^{\text{N}}$  (расположены на хвосте гистона H4) относительно парамагнитного центра метки MTSL (расположен в ядре гистона H3). При этом основной вклад в динамику вносит движение, связанное с конформационными переходами неупорядоченного хвоста гистона H4, которое можно рассматривать как ограниченную диффузию. Такое движение изменяет как пространственную ориентацию, так и длину вектора между протоном  $^1\text{H}^{\text{N}}$  и неспаренным электроном MTSL. Характерные для подобного движения корреляционные функции приведены на Рисунке 2.5.

Прежде чем перейти к сравнению экспериментальных данных и результатов МД моделирования, следует отметить, что для исследуемых образцов нуклеосомной частицы с введенной парамагнитной меткой

наблюдается относительно небольшая вариация значений PRE от одного аминокислотного остатка к другому в N-концевом фрагменте гистона H4 (см. Рисунок 2.1 (E) и Таблицу 7). Из-за недостаточно хорошей сходимости (причины более подробно рассмотрены ниже) процедура расчета эффекта PRE, как правило, не улавливает небольшие вариации в значениях PRE от остатка к остатку. В настоящей работе, для сравнения экспериментальных и расчетных данных для каждого образца нуклеосомной частицы было принято использовать значение PRE, усредненное по всем аминокислотным остаткам хвоста гистона H4.

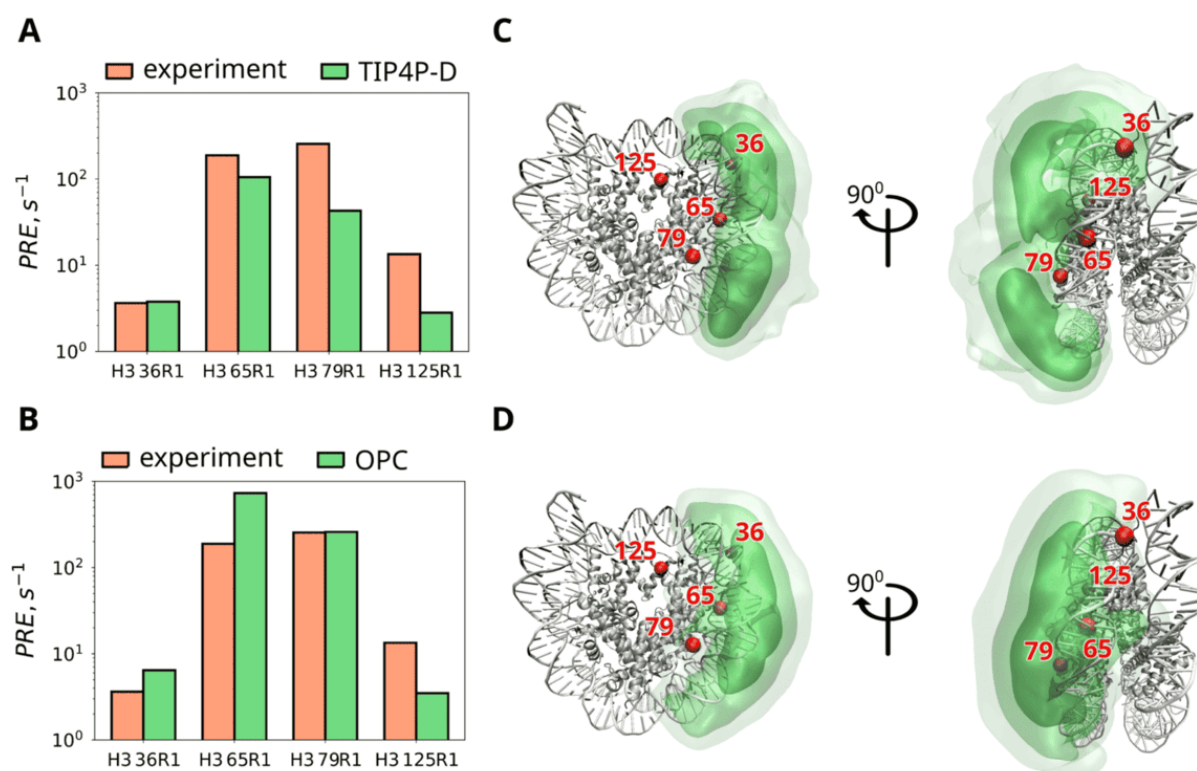


Рисунок 2.2 — Сравнение значений PRE, следующих из эксперимента и расчетов методом молекулярной динамики. (А, В) Экспериментальные (коралловые столбцы) и соответствующие расчетные (зеленые столбцы) значения PRE, полученные в результате усреднения сигнала для аминокислотных остатков с 3 по 15 по двум копиям хвоста гистона H4 в составе нуклеосомы. Расчетные значения получены на основе данных МД траекторий, записанных в моделях воды TIP4P-D и OPC (панели А и В соответственно). (С, D) Пространственная локализация хвоста гистона H4 (аминокислотные остатки с 3 по 15) в МД траекториях, записанных в моделях воды TIP4P-D и OPC (панели С и D). Рисунок был сгенерирован на основе координат структуры 3LZ0 [209]. Данные от

двух копий хвостов гистона H4 были объединены на одном рисунке путем вращения нуклеосомной частицы вокруг оси псевдосимметрии второго порядка. Совокупность координат  $C^\alpha$  атомов для аминокислотных остатков с 1 по 15 хвоста гистона H4 показана в виде изоповерхностей, заключающих в себе 50 %, 90 % и 99 % интегральной плотности, что соответствует темному, умеренному и светлому оттенкам зеленого цвета. Сайты расположения парамагнитных меток показаны в виде пронумерованных красных сфер (см. Рисунок 2.1 (С)).

#### 2.2.4 Сравнение экспериментальных и расчетных значений PRE

На Рисунке 2.2 (А, В) приведено сопоставление экспериментальных и расчетных значений PREs. Коралловые столбцы соответствуют экспериментальным результатам, а зеленые столбцы — данным, полученным из расчетов методом молекулярной динамики с использованием модели воды TIP4P-D и OPC. Расчетные значения PREs опираются на строгие теоретические выражения, которые не содержат каких-либо настраиваемых (эмпирических) параметров. Можно заметить, что оба набора расчетных значений PREs на качественном уровне демонстрируют хорошее согласие с экспериментом как в случае модели воды TIP4P-D, так и в случае модели воды OPC, а именно: для образцов H3 79R1 и 65R1 значения PREs более высокие в сравнении с образцами H3 125R1 и 36R1. И хотя МД моделирование не достигает строго количественного согласия с экспериментом, примечательно, что экспериментальные данные содержатся в интервале, границами которого являются результаты МД моделирования в моделях воды TIP4P-D и OPC. К примеру, для модели воды TIP4P-D наблюдается хорошее согласие с экспериментом в случае образца 36R1, а для модели воды OPC соответствующее значение несколько завышено. Одновременно с этим, модель воды TIP4P-D слегка недооценивает, а модель воды OPC существенно переоценивает экспериментальный результат для образца 65R1. Кроме того, для модели воды TIP4P-D среднее значение PRE гораздо ниже результатов эксперимента для образца 79R1, в то

время как для модели воды OPC наблюдается хорошее согласие. На основе этих наблюдений, можно сделать вывод, что в совокупности обе МД траектории соответствуют релевантной модели эффекта PRE для спин-меченого образца нуклеосомной частицы.

Далее будут рассмотрены причины отсутствия количественного согласия между экспериментальными и расчетными значениями PRE, представленными на Рисунке 2.2. Наиболее вероятно, что основным фактором является недостаточная сходимость МД расчетов значений PRE для неупорядоченного хвоста гистона H4 длиной 25 аминокислотных остатков, конформационная динамика которого охватывает достаточно большой регион вблизи поверхности нуклеосомной частицы. Действительно, существует сильная зависимость величины эффекта PRE от расстояния между взаимодействующими спинами (см. формулу (2.3)). Даже кратковременное сближение хвоста гистона H4 с парамагнитной меткой в МД траектории может существенным образом повлиять на наблюдаемое значение PRE. Другими словами, извлечение корректных значений PRE для каждого аминокислотного остатка из данных МД моделирования осложняется недостаточным уровнем сходимости [198]. Данное факт проиллюстрирован на Рисунке 2.6, где представлено сравнение четырех наборов МД траекторий в модели воды OPC.

Тем не менее, между данными МД моделирования и результатами эксперимента существует полуколичественное согласие (см. Рисунок 2.2 (А, В)). Следовательно, полученные МД модели могут быть использованы для извлечения информации о локализации хвоста гистона H4 в нуклеосомной частице. На Рисунке 2.2 (С, D) в виде изоповерхностей показана совокупность координат в МД траектории для аминокислотных остатков с 1 по 15, принадлежащих хвосту гистона H4. Каждая изоповерхность включает в себе различный уровень интегральной плотности, закодированный оттенками зеленого цвета. Можно заметить, что хвост гистона H4 в значительной мере делокализован в пространстве, что является вполне ожидаемым поведением для неупорядоченного пептида. Одновременно с этим, наблюдается отчетливая тенденция на сближение хвоста гистона H4 с нДНК. Последнее утверждение согласуется с отмеченными выше наблюдениями других исследователей о том, что хвосты гистонов вступают в «нечеткое взаимодействие» с нДНК [50; 168].

Обратимся вновь к Рисунку 2.2 (С, D). Можно видеть, что положение парамагнитных меток H3 65R1 и 79R1 совпадает с центром распределения для координат аминокислотных остатков хвоста гистона H4. Вместе с тем, парамагнитные метки H3 36R1 и 125R1 находятся ближе к периферии на некотором расстоянии от хвоста гистона H4, что хорошо видно из фронтальной и боковой проекции нуклеосомной частицы. Интересно отметить, что различные наборы МД траекторий, записанные с использованием модели воды OPC, демонстрируют некоторую вариативность относительно формы распределения для координат аминокислотных остатков хвоста гистона H4 (см. Рисунок 2.6). Из общего набора можно выделить МД траектории, для которых распределение несколько смещено по направлению к сайту расположения парамагнитной метки 125R1, что определяет более высокие значениям PRE для этих траекторий и лучшее согласие с экспериментом. Из этого наблюдения следует, что более детальная характеристика положения хвоста гистона H4 в нуклеосомной частице требует более длинных МД траекторий и дополнительных экспериментальных данных для валидации.

Не смотря на ряд технических трудностей, связанных с недостаточной сходимостью МД моделирования для количественной оценки эффекта PRE на уровне отдельного аминокислотного остатка, представленный в настоящей работе анализ PRE на основе данных МД представляет собой самую общую и эффективную стратегию. Можно ожидать, что в ближайшее время качество результатов существенно улучшится, благодаря увеличению продолжительности моделирования и совершенствованию параметров для моделей воды. В этом отношении важно отметить, что традиционные подходы, опирающиеся на статистические ансамбли конформеров, сталкиваются с рядом сложностей в приложении к задаче оценки величины эффекта PRE, см. обсуждение ниже [198].

Во-первых, подходы, основанные на статистических конформационных ансамблях, игнорируют наличие внутренней динамики в системе. В частности, по конформационному ансамблю нельзя определить времена корреляции,  $\tau_c$ , характерные для выделенного аминокислотного остатка, знание которых необходимо для расчета значений PRE. Для хвоста гистона H4 в составе нуклеосомы МД моделирование показывает, что значения  $\tau_c$  изменяются от одного аминокислотного остатка к другому

и могут быть значительно короче, чем время вращения нуклеосомной частицы,  $\tau_{rot}$ . Кроме того, значения  $\tau_c$  не могут быть надежным способом определены из экспериментальных данных  $^{15}\text{N}$  релаксации. Следовательно, при построении конформационного ансамбля необходимо делать предположения о значениях  $\tau_c$ , либо вводить  $\tau_c$  в качестве дополнительного настраиваемого параметра.

Во-вторых, в исследовании Гангули и Чена [198] был сделан вывод о том, что в большинстве случаев одного PRE эксперимента недостаточно для того, чтобы однозначным образом сконструировать конформационный ансамбль неупорядоченного белка. В частности, из-за сильной зависимости значений PRE от межъядерного расстояния  $r^{-6}$  оказывается, что для наилучшего соответствия с экспериментальными значениями PRE достаточно определить один или несколько конформеров, в то время как остальные конформеры могут быть выбраны произвольным образом. Сконструированный таким образом конформационный ансамбль может не соответствовать реальному состоянию неупорядоченного белка.

В данной работе также были получены подтверждения для сделанных ранее наблюдений (см. Рисунок 2.7 и Рисунок 2.8). На Рисунке 2.7 представлен конформационный ансамбль минимального размера, состоящий всего из двух конформеров хвоста гистона H4, который очевидным образом не соответствует реальному многообразию пространственных конфигураций, но, тем не менее, приводит к значениям PRE, которые хорошо согласуются с результатами эксперимента. На Рисунке 2.8 показан расширенный вариант конформационного ансамбля, состоящий из 42 конформеров хвоста гистона H4, для которого значения PRE определяются главным образом двумя конформационными состояниями с низкой заселенностью, в то время как оставшиеся 40 конформеров выбраны произвольным образом.

И наконец, валидация полученных МД моделей для нуклеосомной частицы не ограничивалась экспериментальными данными по измерению эффекта PRE. Так, обе МД траектории, записанные с использованием моделей воды TIP4P-D и OPC, продемонстрировали хороший уровень согласия с экспериментом при оценке химических сдвигов для аминокислотных остатков с 1 по 15, принадлежащих хвосту гистона H4, правильно предсказав неупорядоченное состояние для фрагмента,

богатого остатками глицина. МД траектория в модели воды TIP4P-D, также оказалась успешной в воспроизведении экспериментальных значений  $^{15}\text{N}$  скоростей релаксации для аминокислотных остатков хвоста гистона H4 [50]. Напротив, МД траектория в модели воды OPC переоценивает склонность гистоновых хвостов к формированию контактов с нДНК и, как следствие, предсказывает неточные значения для  $^{15}\text{N}$  скоростей релаксации. В тоже время для МД траектории в воде TIP4P-D наблюдается слабая тенденция к раскручиванию нДНК, которая затрагивает оба конца сверхспирализованной нДНК. Подобное поведение, вероятно, отражает известные проблемы со стабильностью системы в модели воды TIP4P-D [18]. В противоположность этому, МД траектория в модели воды OPC отвечает стабильному состоянию системы с небольшими динамическими флуктуациями в положении нДНК (см. Рисунок 2.9).

## 2.3 Материалы и методы

### 2.3.1 Приготовление образцов

Клетки штамма *E. coli* DH5 $\alpha$  трансформировали плазмидой pJ201, содержащей 32 копии последовательности ДНК Widom 601 длиной 147 пар нуклеотидов, затем клетки растили на среде LB. Амплифицированную плазмиду выделяли с помощью набора QIAGEN Plasmid Giga. Для выделения ДНК плазмиду обрабатывали рестриктазой EcoRV-HF (New England Biolabs). Затем осуществляли очистку ДНК путем осаждения полиэтиленгликолем (PEG), как было описано ранее [210]. А именно, 4 М NaCl и 40 % PEG 6000 были добавлены к плазмиде, обработанной EcoRV, в соотношении 0.192:0.346:1 v/v/v. Полученную смесь инкубировали на льду в течение 1 ч, а затем центрифугировали (27000 об/мин, 4 °C, 20 мин). Супернатант, содержащий целевую последовательность ДНК Widom 601, собирали и добавляли к нему холодный этанол в соотношении 1:2.5 v/v. Осажденную ДНК центрифугировали (27000 об/мин, 4 °C, 40 мин),



супернатант сливали, осадок высушивали на воздухе в течение 10 мин и затем растворяли в 0.5 кратном буфере (5 мМ Трис, 0,5 мМ ЭДТА, рН 7.5).

Экспрессия гистонов *Xenopus laevis* Н2А, Н2В, Н3(С110А), содержащих точечную мутацию в одном из сайтов: К36С, L65С, К79С или Q125С, и гистона Н4 осуществлялась в клетках штамма *E. coli* BL21 (DE3) рLysS. Очистка полученных гистонов проводилась, как это было описано ранее, с использованием гель-фильтрационной и ионообменной хроматографии в 7 М растворе мочевины с последующим диализом в растворе 2 мМ β-меркаптоэтанола и завершающим этапом лиофилизации. Одновременно с этим <sup>15</sup>N-меченый вариант гистона Н4 был получен с использованием минимальной среды с <sup>15</sup>NH<sub>4</sub>Cl (1 г/л) в качестве единственного источника азота. Точечные мутанты цистеина гистона Н3, предназначенные для введения парамагнитной метки, были получены путем сайт-направленного мутагенеза плазмиды Н3 дикого типа с использованием ДНК-полимеразы Herculase II (Agilent).

Сборка октамера гистонов осуществлялась путем смешивания четырех развернутых форм гистонов (Н2А, Н2В, один из точечных мутантов гистона Н3 (К36С, L65С, К79С или Q125С) и <sup>15</sup>N-меченый гистон Н4), приготовленных в концентрациях 10 мг/мл, в денатурирующих условиях (7 М гуанидин гидрохлорид, 20 мМ Трис, 10 мМ дитиотреитол, рН 7.5) в молярном соотношении Н2А:Н2В:Н3:Н4, равном 1.2:1.2:1:1. После чего проводился двукратный диализ для переноса образца в рефолдинг-буфер (1 кратный ТЕ, 2 М NaCl, 5 мМ ВМЕ, рН 8.0). По завершению диализа осуществлялось концентрирование образца до 5 мл при помощи центрифужного фильтра Amicon (30 кДа, Millipore). Чтобы предотвратить образование неправильных дисульфидных связей между цистеинами в гистоне Н3, к раствору, содержащему восстановленный октамер гистонов, добавляли дитиотреитол до конечной концентрации 10 мМ. Раствор выдерживали на льду в течение 30 минут, а затем очищали методом гель-фильтрационной хроматографии в рефолдинг-буфере (без ВМЕ и дитиотреитола), как описано ранее [211]. Элюированные фракции, содержащие очищенный октамер, объединяли и инкубировали в течение ночи в темноте при температуре 4 °С с 20-кратным молярным избытком парамагнитной метки (1-оксил-2,2,5,5-тетраметил-дельта-3-пирролин-3-метил)-метантиосульфата

(MTSL; Toronto Research Chemicals, Inc.). Для удаления избытка MTSL, октамер переносили в 0.5-кратный буфер TE и концентрировали до конечной концентрации 15 мг/мл с помощью центробежного фильтра Amicon. Факт успешного введения метки MTSL подтверждался с помощью масс-спектрометрии MALDI-TOF (см. Рисунок 2.3). Процедура мечения оказалась эффективной на 80-90 %, как это следует из анализа масс-спектров до и после введения метки.

ДНК и октамер гистонов смешивали в молярном соотношении ДНК:октамер, равном 1:0.65, в водном буфере 0.5 x TE, 2 М NaCl и 1 mM бензамидин гидрохлорид гидрата. NaCl удаляли путем двукратного диализа с использованием 0.5 кратный TE, 1 mM буфера бензамидин гидрохлорид гидрата при температуре 4 °C. Для того чтобы отделить восстановленные нуклеосомные частицы от свободной ДНК, проводилось центрифугирование в градиенте сахарозы с использованием 5-30 % градиента сахарозы в 0.5-кратном буфере. Фракции, содержащие нуклеосомные частицы, объединяли и удаляли сахарозу путем замены буфера на буфер 0.5 x TE с использованием центробежного фильтра Amicon. Подтверждение факта образования и проверка чистоты образцов нуклеосомных частиц осуществлялась при помощи анализа электрофоретической подвижности по результатам нативного электрофореза в полиакриламидном геле (см. Рисунок 2.4).

### 2.3.2 ЯМР эксперименты

ЯМР эксперименты на образцах нуклеосом, восстановленных из  $^{15}\text{N}$ -меченого гистона H4 и MTSL-меченого гистона H3, проводились при температуре 25 °C на ЯМР спектрометре Bruker 850 МГц Avance III HD с криодатчиком TCI. Образцы нуклеосом были приготовлены в концентрации ~40-45 мкМ в буферном растворе, содержащем 5 mM Трис, 0.5 mM ЭДТА, 100 mM NaCl, 0.1 mM MgCl<sub>2</sub> и 5 % D<sub>2</sub>O при pH 7.0. После ЯМР измерений к каждому образцу, содержащему парамагнитную метку, добавлялся аскорбат натрия для того, чтобы перевести систему в диамагнитное состояние, а именно: аскорбат натрия добавлялся в 20-кратном избытке с

последующей инкубацией в течение 4-7 часов при температуре 4 °С. После чего проводили ЯМР измерения для диамагнитной системы. Значения PRE для каждого аминокислотного остатка хвоста гистона H4 определялись в стандартном  $^{15}\text{N}$ - $^1\text{H}$  HSQC эксперименте с подавлением сигнала воды из отношения объемов пиков для парамагнитного и диамагнитного образца, согласно следующему выражению  $V_{para}/V_{dia} = \exp(-PRE \cdot 2\tau_{inept})$ , как это было описано ранее [201]. Длительность импульсной последовательности INEPT в HSQC эксперименте составляла 4.56 мс, а период восстановления сигнала был выбран равным 2 с, что обеспечивало полное восстановление сигнала как для парамагнитных, так и для диамагнитных образцов.

### 2.3.3 МД моделирование

Ранее были записаны МД траектории нуклеосомной частицы в модели воды TIP4P-D (одна траектория длиной в 2 микросекунды) и в модели воды OPC (двадцать две траектории общей длиной 41 мкс) [50; 102]. Оба набора траекторий были записаны в одном и том же силовом поле Amber ff14SB [67] с использованием близких параметров за исключением модели воды. Обсудим важные параметры, которые были использованы при записи этих траекторий.

Во-первых, МД моделирование с использованием модели воды TIP4P-D отталкивается от начальной структуры 3LZ0 из базы данных PDB. В то время как, МД моделирование, сделанное Пенгом и соавторами в модели воды OPC, опирается на другую начальную структуру 1KX5 из базы данных PDB [162], у которой аминокислотная последовательность гистонов соответствует *Xenopus laevis*, а последовательность ДНК отличается. Оригинальная последовательность ДНК была заменена на последовательность гена KRAS из организма человека и расширена с обоих концов линкерными фрагментами длиной 20 пар нуклеотидов. Существует вероятность, что подобное вмешательство может негативным образом отразиться на качестве МД моделей нуклеосомной частицы. Так, в некоторых траекториях внешний виток нуклеосомной ДНК

демонстрирует высокую подвижность, тем не менее, признаков постепенного раскручивания ДНК обнаружено не было (см. Рисунок 2.9).

Во-вторых, МД моделирование с использованием модели воды TIP4P-D проводилось при температуре 25 °С с добавлением 100 мМ NaCl, что соответствует экспериментальным условиям в данной работе. В то время как, МД моделирование с использованием модели воды OPC проводилось при температуре 36 °С с добавлением 150 мМ NaCl. И хотя более высокая температура и концентрация соли должны способствовать увеличению подвижности гистоновых хвостов, напротив, наблюдалось замедление динамики при моделировании в модели воды OPC относительно моделирования в модели воды TIP4P-D.

В-третьих, МД траектории в модели OPC, опубликованные Пенгом и соавторами [102] были сохранены с шагом в 1 нс. Для того чтобы исследовать этот аспект, мы обработали нашу траекторию в модели воды TIP4P-D в двух вариантах: с шагом в 1 нс и с оригинальным шагом в 1 пс. Как оказалось, значения PRE и поперечной скорости релаксации  $R_2$  ядер  $^{15}\text{N}$  практически нечувствительны к шагу, с которым записываются кадры в МД траектории. Подобный результат ожидаем, если учесть, что продольная скорость релаксации  $R_1$  ядер  $^{15}\text{N}$  более чувствительна к быстрым движениям, в отличие от скорости  $R_2$ , которая в основном определяется медленными движениями в системе. В целях расчета скоростей релаксации  $R_1$  можно рекомендовать шаг записи кадров МД траектории равный 10 пс.

В-четвертых, как при моделировании с использованием модели воды TIP4P-D, так и при моделировании с использованием модели воды OPC был использован термостат Ланжевена. Мы рекомендуем выбирать термостат Бусси-Парринелло [63], так как данный термостат лучше подходит для исследования динамики системы, в частности для расчета скоростей релаксации.

Обработка МД траекторий. Для каждого аминокислотного остатка парамагнитный вклад в ядерную спиновую релаксацию  $^1\text{H}^{\text{N}}$  спинов определяется следующими выражениями

$$\Gamma_2 = \frac{1}{20} D^2 (J(0) + 4(J(\omega_H))), \quad (2.1)$$

$$J(\omega) = \int_0^{\infty} C(\tau) \cos(\omega\tau) d\tau, \quad (2.2)$$

$$C(\tau) = \left\langle \frac{P_2(\cos \theta(t, t + \tau))}{r^3(t)r^3(t + \tau)} \right\rangle \quad (2.3)$$

где  $D$  — константа диполь-дипольного взаимодействия, равная  $D = 4.9764 \cdot 10^{-22} \text{ м}^3 \cdot \text{с}^{-1}$  для метки MTSL,  $J(\omega)$  — функция спектральной плотности,  $\omega_H$  — ларморовская частота прецессии протона,  $C(\tau)$  — корреляционная функция диполь-дипольного взаимодействия,  $P_2(x)$  — многочлен Лежандра второго порядка,  $\theta(t, t + \tau)$  — угол между векторами  $\vec{r}(t)$  и  $\vec{r}(t + \tau)$ , соединяющий парамагнитный центр с соответствующим ядром  $^1\text{H}^N$  в моменты времени  $t$  и  $t + \tau$ ,  $r(t)$  и  $r(t + \tau)$  — длины векторов, а угловые скобки обозначают усреднение по времени и по ансамблю (состоящему в данном случае из двух копий гистона Н4).

Учитывая тот факт, что нуклеосомная частица обладает осью псевдосимметрии второго порядка, релаксация для каждого спина  $^1\text{H}^N$  хвоста гистона Н4 объясняется взаимодействием с парамагнитным центром двух меток MTSL, которые соответствуют двум копиям гистона НЗ и условно обозначаются, как «ближняя» (от англ. proximal) и «дальняя» метка (от англ. remote)

$$PRE = \Gamma_2^{proximal} + \Gamma_2^{remote} \quad (2.4)$$

Расчет значений PRE из МД траекторий проводился следующим образом. Сначала осуществляли наложение всех кадров из МД траектории на референсную структуру 3LZ0 путем наложения  $\text{C}^\alpha$  атомов, принадлежащих участкам вторичной структуры гистонного ядра. Затем вычисляли корреляционные функции следующим образом

$$c(\tau) = \frac{1}{N_{total}} \sum_{i=1}^{N_{traj}} \sum_{j=1}^2 \sum_{k=1}^{N_{timepointpairs}^{(i)}} \frac{P_2(\cos \theta(t, t + \tau))}{r^3(t)r^3(t + \tau)} \quad (2.5)$$

где первая операция суммирования выполняется по всем записанным траекториям для рассматриваемой системы, вторая операция суммирования выполняется по двум хвостам гистона Н4, а третья операция суммирования

выполняется по всем парам временных интервалов  $t, t + \tau$  в соответствующей траектории (при фиксированном  $\tau$ ). Полученный результат нормируется на  $N_{total}$  общее количество слагаемых в тройной сумме в уравнении (2.5). При вычислении  $\Gamma_2^{proximal}$  вторая операция суммирования подразумевает, что задействованы два вектора  $\vec{r}(t)$ : (1) между первой копией хвоста гистона H4 и его ближней парамагнитной меткой и (2) между второй копией хвоста гистона H4 и его ближней парамагнитной меткой. Аналогичным образом рассчитывается  $\Gamma_2^{remote}$ . При этом важно отметить, что отличие между ближней и дальней меткой не сводятся лишь к различию в расстоянии. Между ближней и дальней меткой существуют отличия, основанные на фундаментальных соображениях о симметрии в системе. В частности, поворот на 180 градусов вокруг диадной оси приводит к перестановке двух ближних меток (а также двух дальних меток).

Стоит подчеркнуть, что уравнение (2.5) идеальным образом описывает ситуацию, при которой необходимо учесть данные от двух (предположительно эквивалентных) хвостов гистона H4 в моделируемой нуклеосомной частице, а также принять во внимание данные, поступающие из нескольких МД траекторий. В более широком смысле, в этом уравнении для рассматриваемой системы осуществляется усреднение, как по ансамблю, так и по времени моделирования. Подход, основанный на уравнении (2.5), является более предпочтительным в сравнении с вариантом, где расчет PRE проводится отдельно для каждой из двух копий гистоновых хвостов и каждой МД траектории из существующего набора, после чего полученные результаты усредняются для получения итогового ответа.

Для учета вращения нуклеосомной частицы корреляционные функции  $c(\tau)$ , рассчитанные по уравнению (2.5), умножаются на  $\exp(-t/\tau_{rot})$  (вращательное движение было исключено на этапе наложения кадров МД траектории на референсную структуру)

$$C(\tau) = c(\tau) \exp(-t/\tau_{rot}) \quad (2.6)$$

Значение  $\tau_{rot}$ , равное 163 нс, получено из расчетов HYDROPRO, как сообщалось ранее [50]. На следующем этапе аппроксимация рассчитанной корреляционной функции  $C(\tau)$  осуществляется в виде суммы нескольких

экспонент [50]. Далее вычисляется значение спектральной плотности по уравнению (2.2), и наконец, соответствующее значение PRE по уравнению (2.4).

В настоящей работе в МД траекториях не использовалось явное представление для координат метки MTSL. В принципе, моделирование четырех образцов нуклеосомы с меткой MTSL, прикрепленной к различным сайтам, является возможной стратегией [212], однако требует больших временных и вычислительных затрат на запись четырех длинных МД траекторий. Вместо этого в качестве положения парамагнитного центра метки MTSL были использованы координаты тяжелых атомов на конце боковой цепи соответствующего аминокислотного остатка: атом  $N^{\zeta}$  для K36 и K79,  $C^{\delta 1}$  для L65 и  $N^{\epsilon 2}$  для Q125 в гистоне H3. Можно считать, что динамика нативных боковых цепей слабо отличается от движения метки MTSL. Более сильным аргументом является то, что вклад динамики метки MTSL в эффект наблюдаемый PRE является вторичным по сравнению с динамикой неупорядоченного хвоста гистона H4. В целом, эффект замены явного представления координат атомов метки MTSL на координаты тяжелых атомов боковой цепи находится на уровне статистической погрешности МД моделирования эффекта PRE.

Далее следует отметить, что уравнение (2.5) зависит не только от флуктуаций ориентации вектора  $\vec{r}(t)$ , соединяющего пару взаимодействующих спинов, но и от флуктуаций расстояния между взаимодействующими спинами  $r(t)$ , т.е. длины вектора  $\vec{r}(t)$ . Главным образом, изменение вектора  $\vec{r}(t)$  вызвано движением гибкого хвоста гистона H4. Ситуация напоминает взаимную трансляционную диффузию двух частиц. Так, довольно выражено влияние ограниченной трансляционной диффузии хвоста гистона на корреляционные функции  $C(\tau)$ . Для атома  $^1H^N$ , расположенного на конце гибкого гистонового хвоста H4 и взаимодействующего с ближней парамагнитной меткой, корреляционная функция затухает за десятки наносекунд, что намного быстрее, чем затухание из-за вращения нуклеосомной частицы (см. Рисунок 2.5).

Для того чтобы визуализировать локализацию хвостов гистона H4 в составе нуклеосомной частицы (см. Рисунок 2.2 и Рисунок 2.6), были собраны данные от двух копий гистоновых хвостов. В качестве первого шага, была подготовлена повернутая копия референсной структуры 3LZ0.

Операция поворота представляла собой наложения первой копии каждого гистона H2A, H2B, H3 и H4, на вторую копию того же гистона и наоборот (наложение проводилось по атомам C<sup>α</sup> вторичной структуры). Данное преобразование соответствует повороту на 180 градусов вокруг диадной оси. Далее все кадры МД траектории были совмещены с повернутой структурой 3LZ0, что дало повернутую копию исходной траектории. И наконец, из повернутой траектории были извлечены координаты атомов C<sup>α</sup> и добавлены в исходную траекторию. Описанным путем была сгенерирована редуцированная версия траектории, включающая координаты референсной структуры 3LZ0 и координаты атомов C<sup>α</sup> для двух (совместно локализованных) хвостов гистона H4 из МД траектории. Координаты атомов C<sup>α</sup> были использованы для построения трехмерной плотности распределения с помощью модуля VolMap программного пакета VMD v.1.9.1 [213] со следующими параметрами: разрешение (от англ. resolution) 1 Å, сглаживающий фактор по Гауссу (от англ. Gaussian smoothing factor) 3, веса (от англ. weights) не устанавливались. Визуализация карты трехмерной плотности была осуществлена в программе VMD с использованием представления в виде изоповерхностей (от англ. Isosurface representation) [213].

Рисунок 2.7 был получен следующим образом. Сначала был сформирован набор конформеров, в который вошли кадры нуклеосомной частицы из траектории в модели воды TIP4P-D, выбранные с шагом в 1 нс. При этом каждая копия хвоста гистона H4 рассматривалась как индивидуальный конформер с отличающимся пространственным положением относительно сайтов введения парамагнитных меток. Для каждого такого конформера были рассчитаны значения PRE по формуле (2.1) и (2.4), но с заменой уравнения (2.2) на более простое выражение [193]

$$J(\omega) = \frac{1}{r^6} \frac{\tau_c}{1 + \omega^2 \tau_c^2} \quad (2.7)$$

где  $r$  — расстояние между выбранным протоном амидной группы в хвосте гистона H4 и парамагнитным центром для конкретного конформера, а  $\tau_c$  — характерное время корреляции.



Как обсуждалось выше, при использовании классической процедуры построения конформационного ансамбля ограниченного экспериментальными данными PRE необходимо сделать предположения о значении  $\tau_c$ . При отсутствии дополнительной информации о системе естественным решением является установить  $\tau_c$  равным  $\tau_{rot} = 163$  нс. На физическом уровне подобный выбор означает, что перемещение хвоста гистона H4 по поверхности нуклеосомной частицы предполагается относительно малым на временном интервале соответствующем вращению нуклеосомной частицы  $\tau_{rot}$  (что является достаточно произвольным выбором). Здесь необходимо учесть, что наблюдение сигналов в спектре ЯМР для хвоста гистона H4 обусловлено тем фактом, что хвост гистона H4 остается в высокой степени подвижным, не смотря на образование слабых контактов с поверхностью нуклеосомной частицы («нечеткое взаимодействие»). После этого был осуществлен поиск ансамбля из двух конформеров, для которого расчетные значения PRE наилучшим образом согласуются с данными эксперимента (согласно Рисунку 2.2 (А)). Для поиска такой пары был исследован набор из четырех тысяч конформеров; для каждой пары конформеров были оптимизированы относительные веса, определяющие вклад в итоговое значение PRE.

Рисунок 2.8 был получен следующим образом. Сначала был сформирован набор конформеров, в который вошли кадры из набора траекторий нуклеосомной частицы в модели воды OPC [102] (кадры были выбраны с шагом 10 нс). Описанным выше способом было выявлено два конформера, которые наиболее точным образом воспроизводят экспериментальные значения PRE, но при этом мы предполагали, что заселенность обоих конформационных состояний равна 10 %. В дополнение, были отобраны 40 конформеров, которые вносят несущественный вклад в итоговое значение PRE; каждому конформеру был присвоен вес равный 2 %.

Конформационные ансамбли, представленные на Рисунках 2.7 и 2.8 опираются на набор кадров из МД траектории, однако сконструировать подобный набор конформеров можно и другим способом. Тем не менее стоит отметить, что применение стандартных инструментов, вроде Flexible-messano [117] или IDPConformerGenerator [214] оказывается недостаточным. После того, как сгенерирована случайная конформация хвоста гистона H4 её необходимо "пришить" к ядру нуклеосомной частицы;

после чего необходимо произвести проверку модели на наличие стерических затруднений и в соответствии с результатами проверки либо сохранить случайную конформацию, либо исправить, либо отбросить. И даже в этом случае, конформационные ансамбли неупорядоченных белков, опирающиеся на данные PRE, не всегда являются репрезентативными, на что впервые указали Гангули и Чен [198]. В настоящей работе несостоятельность подхода проиллюстрирована на Рисунках 2.7 и 2.8. Результат можно улучшить, если увеличить количество образцов с парамагнитной меткой, введенной в различные позиции, или же принять во внимание данные других экспериментов (например, химические сдвиги, скорости релаксации и т.д.).

## 2.4 Заключение

Изучение конформационного многообразия гистоновых хвостов необходимо для понимания их функций и сети взаимодействий, в которые вовлечены гистоновые хвосты как в случае индивидуальных нуклеосом, так и в случае объединения нуклеосом в структуры более высокого порядка. При этом особый интерес представляет то, каким образом изменяется пространственная конфигурация и набор контактов гистоновых хвостов в ответ на различные посттрансляционные модификации и образование комплекса с хроматин-ассоциированными белками. В этом отношении уникальным источником информации являются методы ЯМР спектроскопии в жидкости и твердом теле. В этом исследовании мы показали, что измерения PRE в образцах нуклеосом с использованием нескольких парамагнитных меток должны быть одним из центральных элементов таких многоэкспериментальных наборов данных ЯМР. В частности, было продемонстрировано, что комбинированный подход, предполагающий одновременное использование экспериментальных данных PRE и МД моделирование, опирающееся на длинные траектории, позволяет извлекать ценную информацию о локализации гистоновых хвостов. Можно предположить, что в будущем разработанный подход может быть применен к задаче изучения более высокого уровня компактизации нуклеосом, а

также комплексов нуклеосом с хроматин-ассоциированными белками и исследования других биомолекулярных комплексов, включающих функционально-активные гибкие сегменты. И наконец, следует отметить, что данные PRE, как и другие экспериментальные параметры ЯМР, могут быть использованы для валидации МД моделей и совершенствования существующих силовых полей. В совокупности полученные результаты способствуют лучшему пониманию функции хроматина.

## 2.5 Дополнительная информация

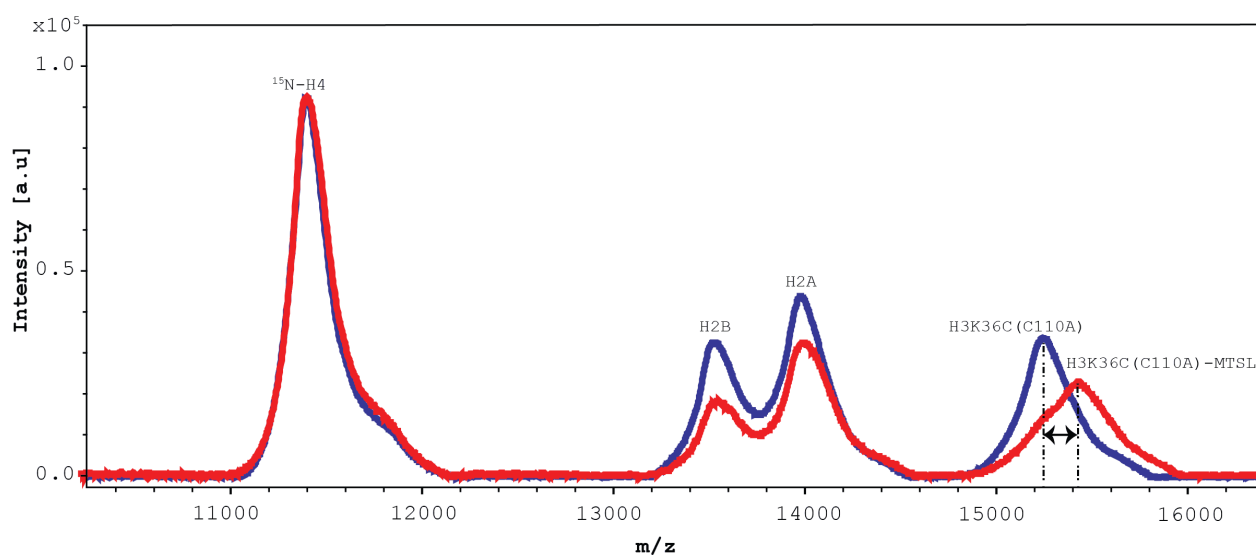


Рисунок 2.3 — Репрезентативный масс-спектр MALDI-TOF до (синий цвет) и после (красный цвет) введения парамагнитной метки MTSL для октамера гистона, который содержит  $^{15}\text{N}$ -labeled H4 и H3(C110A) K36C. Наблюдаемые значения для отношения массы к заряду  $m/z$  согласуются с ожидаемым увеличением молекулярной массы гистона H3(C110A) K36C на 184 Да в результате присоединения метки MTSL. Вертикальная шкала соответствует интенсивности пика для  $^{15}\text{N}$ -меченого хвоста гистона H4.

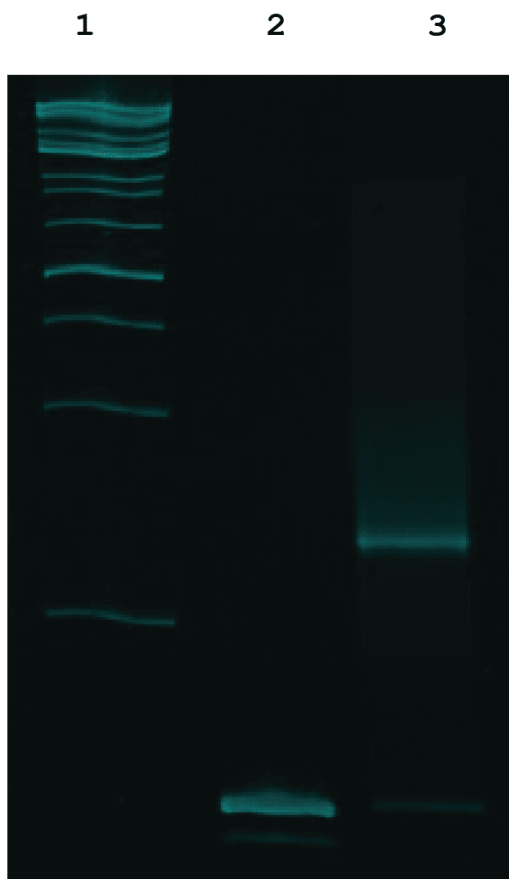


Рисунок 2.4 — Нативный полиакриламидный гель (5 % полиакриламид, 0.3 кратный TBE буфер: 27 мМ Трис-борат, 0.6 мМ EDTA), окрашенный бромистым этидием. Дорожка номер 1: ДНК-маркер (New England Biolabs). Дорожка номер 2: последовательность ДНК Widom 601 длиной 147 пар нуклеотидов. Дорожка 3: нуклеосомная коровая частица после очистки в градиенте сахарозы.

$$\frac{C(\tau)^{PRE}}{C(0)^{PRE}} \text{ vs. } C(\tau)^{tumbling} \text{ vs. } C(\tau)^{NH}$$

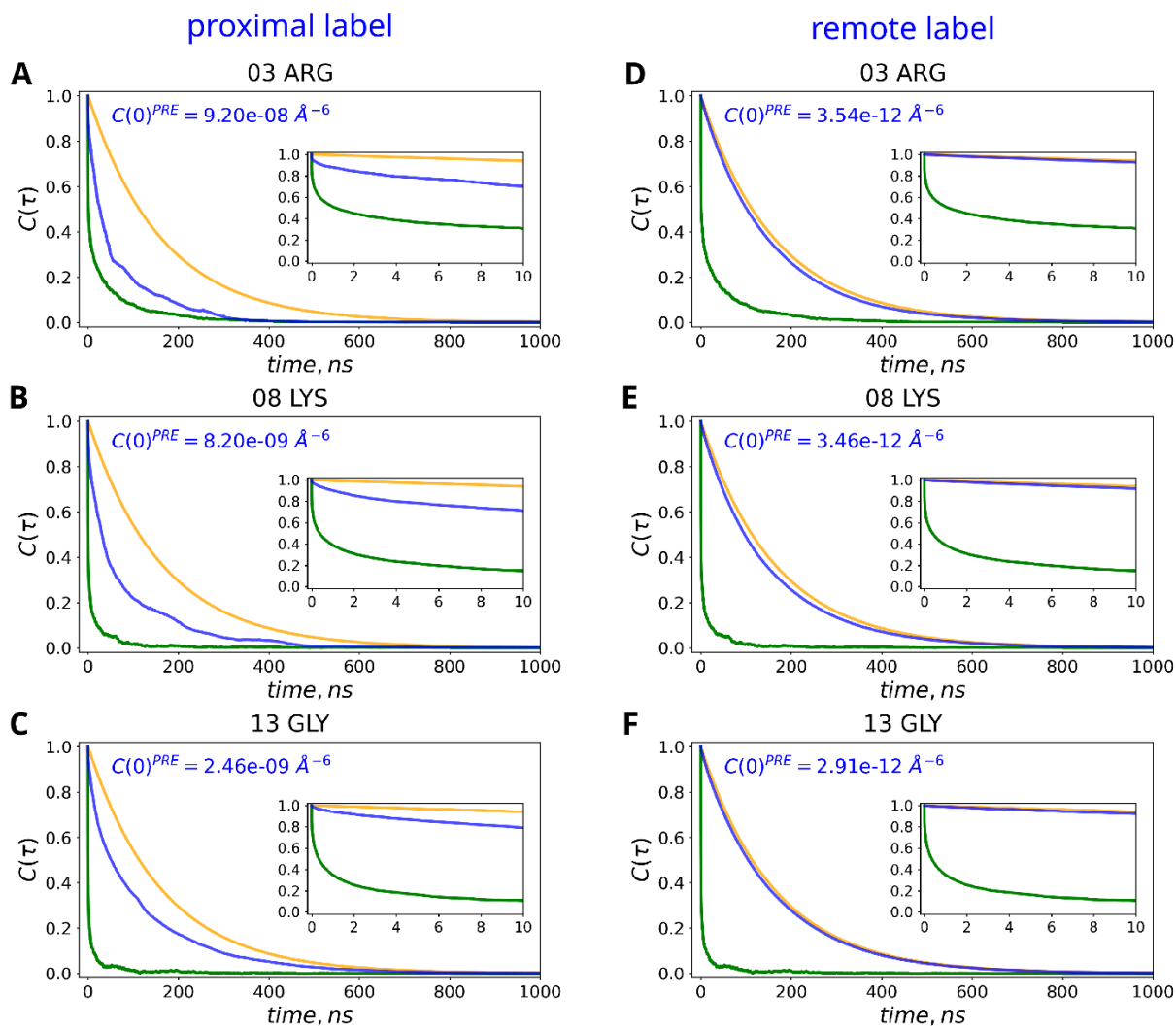


Рисунок 2.5 — Примеры корреляционных функций  $C(\tau)$ , рассчитанных по данным МД траектории нуклеосомной частицы в модели воды TIP4P-D длиной 2 мкс (синие кривые). Результаты приведены для трех аминокислотных остатков R3, K8 и G13 хвоста гистона H4 в образце нуклеосомной частицы, которая содержит спин-меченый в позиции L65S гистон H3 (условные обозначения указаны на рисунке). Для удобства визуализации корреляционные функции отнормированы на единицу (соответствующее  $C(0)$  значение указано на каждой панели рисунка). Кроме того, на рисунке показана корреляционная функция для вращения нуклеосомной части с характерным временем  $\tau_{rot}$  равным 163 нс [50] (оранжевая кривая), и корреляционная функция дипольного взаимодействия  $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$  для трех выбранных амидных сайтов (зеленые кривые). Важно отметить, что функции корреляции  $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$  существенным

образом отличаются от функций корреляции, которые возникают при расчете эффекта PRE. К примеру, рассмотрим следующую ситуацию: остаток R3 хвоста гистона H4 образует слабые контакты с нДНК, так что оставшаяся часть хвоста с 3 по 15 аминокислотные остатки формирует большую гибкую петлю. В этом случае движение гистонового хвоста относительно парамагнитной метки может происходить очень медленно, в то время как конформационная динамика петлевого сегмента может быть очень быстрой. Показанный рисунок иллюстрирует тот факт, что данные азотной релаксации не могут достоверным образом описать динамику в системе в контексте эффекта PRE. В целом, корреляционные функции, рассчитанные из МД траектории в модели воды OPC (не показаны), не отличаются от полученных в результате МД моделирования в модели воды TIP4P-D. Тем не менее существует некоторая тенденция более медленного убывания корреляционной функции, что согласуется с нашим наблюдением о том, что хвосты гистона H4 менее подвижны в модели воды OPC

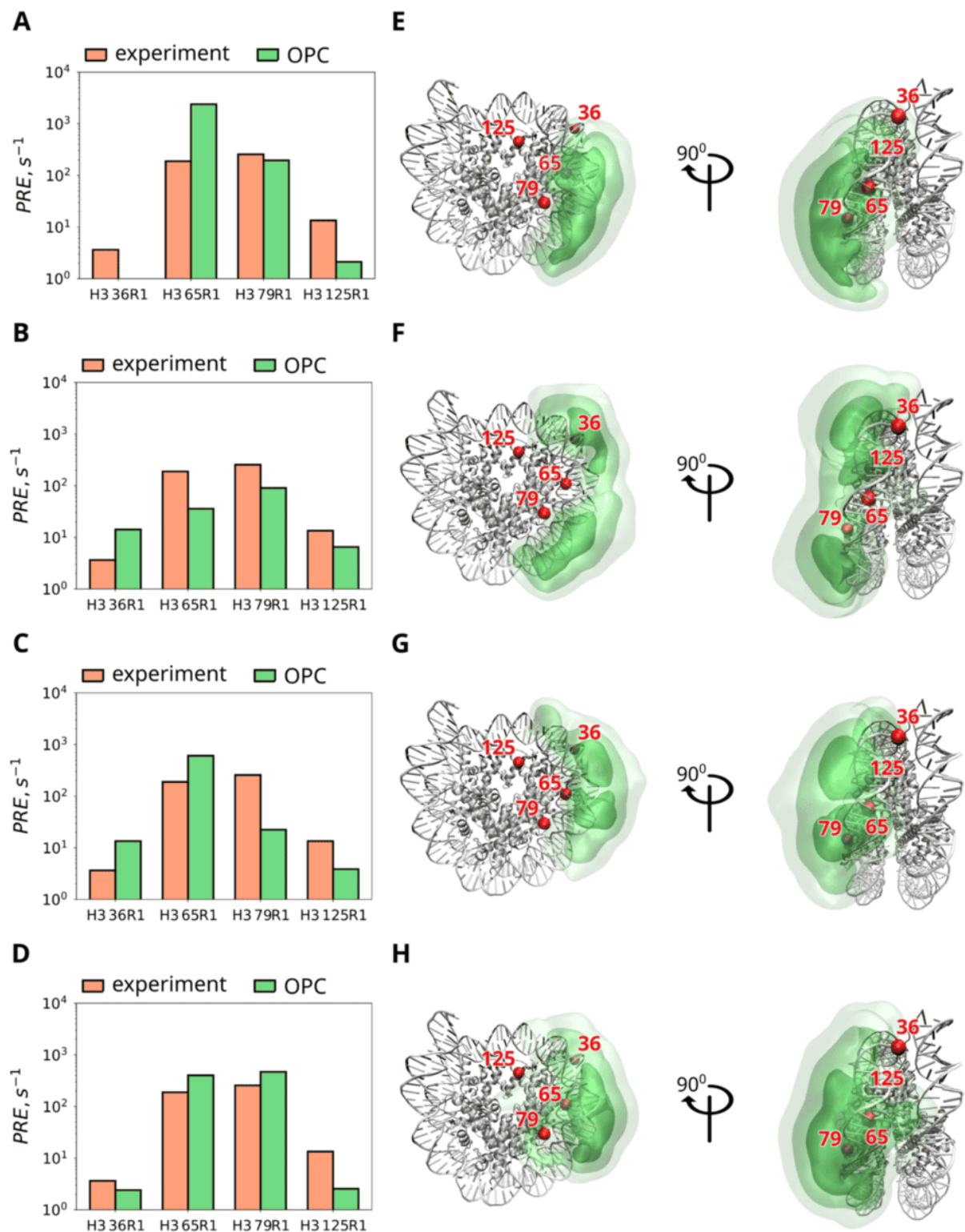


Рисунок 2.6 — (A-D) Экспериментальные значения PRE (коралловые столбцы), усредненные по остаткам с 3 по 15, принадлежащие хвосту гистона H4, и соответствующие расчетные значения из данных МД моделирования (зеленые столбцы) для четырех схем моделирования в воде OPC. Авторы обозначили четыре схемы моделирования, как модели A-D (краткое описание приведено ниже, дополнительные



подробности см. по ссылке [102]). (Е-Н) Локализация хвоста гистона Н4 в четырех вариантах МД моделирования в воде ОРС. Способ генерации данного рисунка совпадает с использованным при подготовке Рисунка 2.2. Введенное авторами название для моделей А-Д соответствует протоколу моделирования, использованному при подготовке начальной модели нуклеосомной частицы [102]. Модель А основана на структуре PDB:1KX5, в которой последовательность ДНК была заменена на расширенную последовательность человеческого онкогена KRAS. Начальные координаты гистоновых хвостов были взяты из PDB:1KX5, а также частично перенесены из PDB:1AOI. Недостающие остатки на концах гистоновых хвостов были достроены путем линейного наращивания с сохранением конформации для остатков, имеющих в экспериментальной модели. Модель В аналогична модели А, за исключением того, что начальные координаты гистоновых хвостов были взяты из PDB:1KX5 и PDB:1EQZ. В модели С частично разрешенные гистоновые хвосты в PDB:1KX5 были обрезаны, а затем линейно восстановлены с геометрией, соответствующей ориентации двух последних остатков в сайтах обрезки хвостов. Возникающие стерические затруднения были разрешены при помощи программы Modeler [215]. Модель D аналогична модели С, но не предполагает выполнения требования на симметрию положения гистоновых хвостов относительно диадной оси.

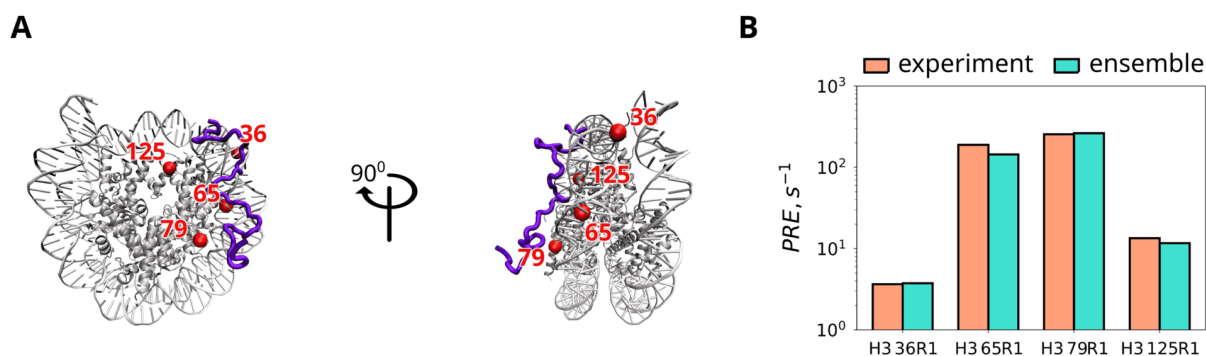


Рисунок 2.7 — (А) Ансамбль из двух конформеров нуклеосомной частицы, который наилучшим образом согласуется с экспериментальными значениями PRE. Ансамбль сконструирован на основе МД траектории в модели воды TIP4P-D, по протоколу, описанному в разделе 2.3. Конформеры хвостов гистона H4 показаны фиолетовым цветом, а сайты расположения парамагнитных меток показаны в виде красных сфер. Статистический вес конформационных состояний составляет 93 % и 7 % соответственно. (В) Сравнение экспериментальных значений PRE и значений PRE, рассчитанных по ансамблю конформеров (коралловые и бирюзовые столбцы соответственно). Расчетные значения PRE для каждого индивидуального аминокислотного остатка были получены на основе формулы Соломона-Бломбергена для спектральной плотности (2.7). Полученные значения PRE усреднялись по аминокислотным остаткам с 3 по 15, принадлежащим хвосту гистона H4, и по двум конформерам в ансамбле с учетом их статистического веса. Ансамбль из двух конформеров успешно воспроизводит экспериментальные значения PRE, хотя никоим образом не отражает реальное многообразие конформационных состояний хвоста гистона H4 в составе нуклеосомы.

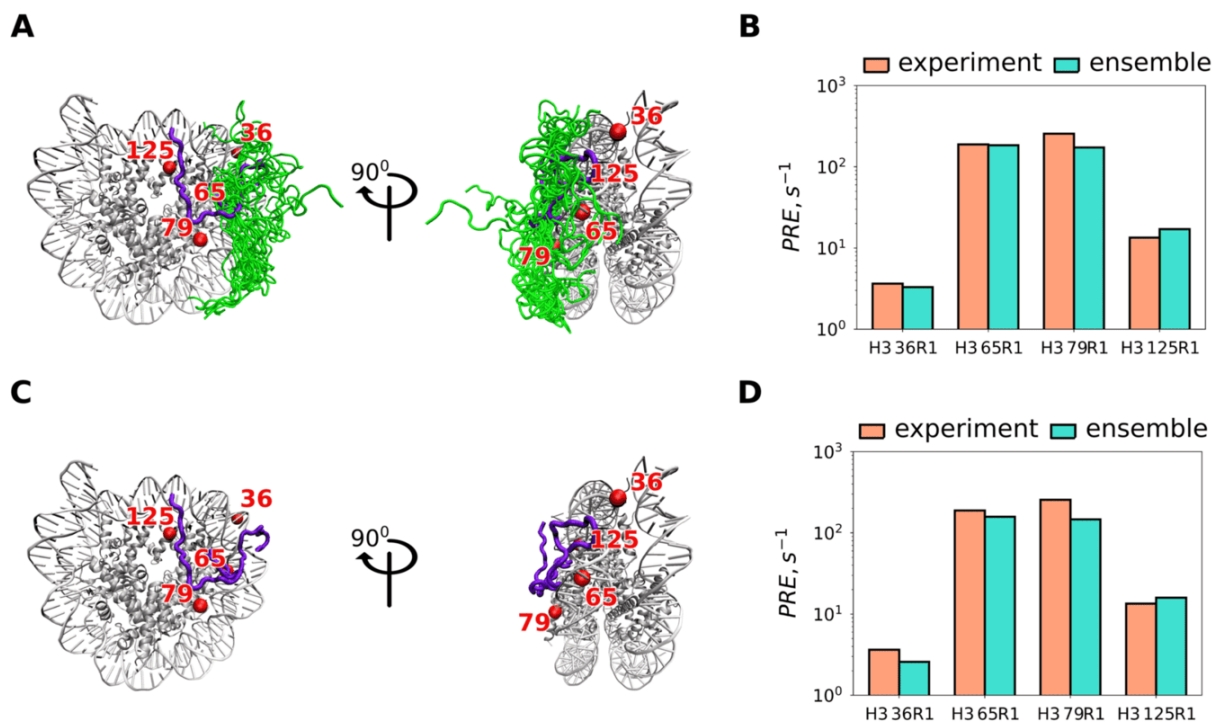


Рисунок 2.8 — (А) Ансамбль из 42 конформеров нуклеосомной частицы, который наилучшим образом согласуется с экспериментальными значениями PRE. Ансамбль сконструирован на основе МД траектории в модели воды OPC по протоколу, описанному в разделе 2.3. Конформеры хвостов гистона H4 показаны фиолетовым и зеленым цветом, а сайты расположения парамагнитных меток показаны в виде красных сфер. Статистический вес каждого конформационного состояния, показанного фиолетовым цветом, составляет 10 %. Тем не менее, данные конформеры вносят основной вклад в расчетные значения PRE. Статистический вес каждого из 40 конформационных состояний, показанных зеленым цветом, составляет по 2 %. Данные состояния вносят несущественный вклад в расчетные значения PRE. (В) Сравнение экспериментальных значений PRE и значений PRE, рассчитанных по ансамблю конформеров (коралловые и бирюзовые столбцы соответственно). Расчетные значения PRE для каждого индивидуального аминокислотного остатка были получены на основе формулы Соломона-Бломбергена для спектральной плотности (2.7). Полученные значения PRE усреднялись по аминокислотным остаткам с 3 по 15, принадлежащим хвосту гистона H4, и по двум конформерам в ансамбле с учетом их статистического веса. (С) Редуцированный ансамбль конформеров нуклеосомной частицы, содержащий два конформера, дающих наибольший вклад в значения PRE (показаны те же конформеры,

что и на панели (A)), при этом статистический вес каждого из них равен 10 %. Необходимо отметить, что проиллюстрированные здесь конформеры полностью отличаются от тех, что показаны на Рисунке 2.7. (D) Сравнение экспериментальных значений PRE и значений PRE, рассчитанных по редуцированному (относительно ансамбля, показанного на панели (C)) ансамблю конформеров (коралловые и бирюзовые столбцы соответственно). И хотя формально ансамбль из 42 конформеров успешно воспроизводит экспериментальные значения PRE, данный ансамбль не может рассматриваться как релевантная модель для хвоста гистона H4 в виду того, что экспериментальные ограничения лишь в малой степени определяют его вид.

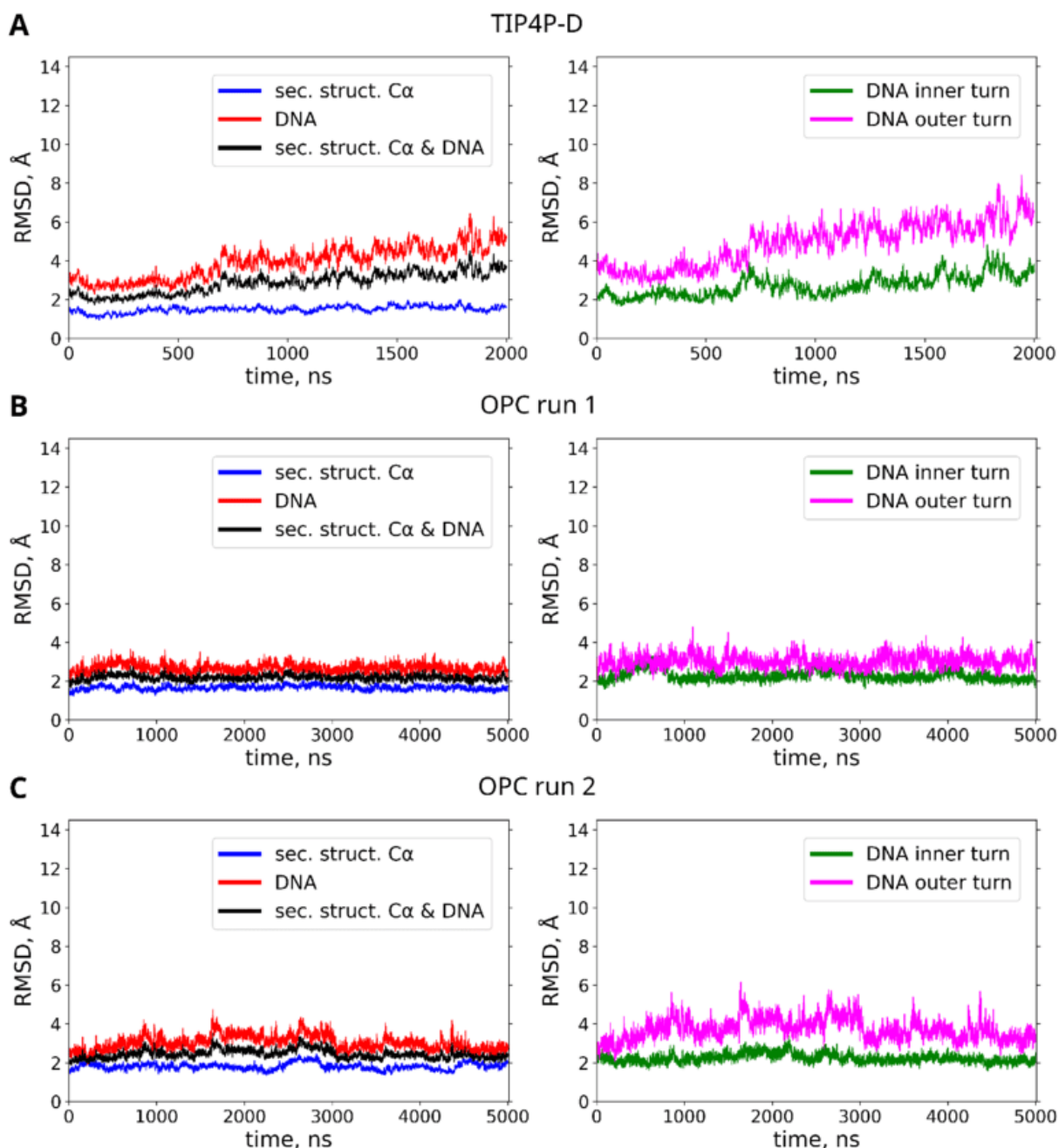


Рисунок 2.9 — (А) Профили RMSD, рассчитанные по МД траектории нуклеосомной частицы в модели воды TIP4P-D длиной 2 мкс. Все кадры МД траектории были совмещены с референсной структурой 3LZ0 путем наложения  $C^\alpha$  атомов гистонового кодра (т.е. были использованы  $C^\alpha$  атомы, которые находятся в  $\alpha$ -спиральных участках). Далее значения RMSD были рассчитаны с использованием различных наборов атомов: (1)  $C^\alpha$  атомы, которые были использованы при выравнивании (синяя кривая); (2) атомы N1 и N9 из нуклеиновых оснований нДНК (красная кривая); (3) комбинация из наборов атомов (1) и (2) (черная кривая); (4) атомы N1 и N9 из внутреннего витка нДНК, т.е. нуклеотидные основания с -38 до 38 (зеленая

кривая); (5) атомы N1 и N9 из внешнего витка нДНК, т.е. нуклеотидные основания с -72 до -39 и от 39 до 72 (пурпурная кривая). Профили RMSD рассчитывались с временным интервалом в 1 нс. (В, С) Профили RMSD для двух репрезентативных МД траекторий нуклеосомной частицы длиной 5 мкс в модели воды OPC, принадлежащих к схемам моделирования В и D в работе [102]. Важно отметить, что определение внутреннего и внешнего витков нДНК совпадает с введенным ранее (без учета линкерных сегментов).

Таблица 7 — Экспериментальные значения PRE для амидных протонов хвоста гистона H4 в четырех образцах нуклеосомной частицы, содержащих гистон H3 с введенной парамагнитной меткой 36R1, 65R1, 79R1 и 125R1. Значение PRE для каждого аминокислотного остатка рассчитывалось из отношения объемов пиков в  $^{15}\text{N}$ - $^1\text{H}$  ЯМР HSQC спектре для парамагнитного образца к диамагнитному образцу. Переход в диамагнитное состояние инициировался путем добавления избытка аскорбата натрия (см., Рисунок 2.1 и раздел 2.3). Для трех перекрывающихся сигналов от остатков глицина G4, G6 и G13 приведено усредненное значение. В случае образца H3 36R1 для трех остатков глицина зафиксированы небольшие отрицательные значения PRE, которые были округлены до нуля (немодифицированные значения указаны в квадратных скобках). В двух последних строках таблицы приведено среднее значение и стандартное отклонение для эффекта PRE, которые были рассчитаны по всем остаткам хвоста гистона H4.

остаток H4	PRE( $\text{c}^{-1}$ )			
	H3 36R1	H3 65R1	H3 79R1	H3 125R1
R3	2.5	194.1	255.3	0.3
G4/G6/G13	0.7	172.1	327.1	7.6
K5	11.3	105.0	216.4	14.9
G7	5.5	194.8	265.6	9.3
K8	4.9	165.3	350.5	15.4
G9	4.8	210.4	180.0	12.2
L10	10.8	229.0	239.0	13.2
G11	0.1	227.7	268.6	17.5
K12	0.0 [-1.6]	165.2	229.2	19.4
G14	0.0 [-2.2]	164.9	248.0	11.7
A15	0.0 [-2.4]	250.1	235.0	26.8
Среднее	3.7	189.0	255.9	13.5
Стандартное отклонение	4.2	40.5	48.1	6.8

## Глава 3. Использование ЯМР и МД моделирования для изучения переворотов ароматических колец в кристаллах убиквитина

### 3.1 Введение

Ароматические аминокислотные остатки выполняют важные функции в белках. Известно, что ароматические остатки обеспечивают стабилизацию структуры белка, формируя плотно упакованный гидрофобный кластер — «гидрофобное ядро». Кроме того, ароматические остатки играют важную роль в молекулярном узнавании и связывании [216], зачастую располагаясь в активном центре ферментов и на интерфейсе взаимодействия белок-белок и белок-малая молекула. На протяжении последних четырех десятилетий динамика боковых цепей ароматических остатков была предметом особенного интереса исследователей [217–220]. Причина такого интереса в том, что боковые цепи ароматических остатков достаточно «громоздкие» и для изменения их положения в пространстве, особенно для переворотов (от англ. flips) ароматических колец, требуется значительный свободный объем. В свою очередь, для того чтобы возник свободный объем и переворот ароматических колец стал возможным, должно произойти согласованное изменение конформации окружающих остатков. Взаимосвязь с конформационными перестройками делает изучение переворотов ароматических колец особенно привлекательным в контексте исследования функциональной активности ферментов и ионных каналов [221; 222].

По определению переворот ароматических колец остатков Phe и Tyr — это поворот боковой цепи аминокислотного остатка вокруг двугранного угла  $\chi_2$  на  $180^\circ$ . Важно отметить, что при перевороте ароматических колец происходит переход между двумя структурно неразличимыми состояниями, поэтому классические методы, такие как рентгеноструктурный анализ, не могут зафиксировать такое событие. Незаменимой тут оказывается спектроскопия ядерного магнитного резонанса (ЯМР). Она позволяет детальным образом охарактеризовать локальную динамику боковых цепей ароматических остатков, включая характерное время переворотов, а также



амплитуду и частоту покачиваний относительно оси ароматического кольца. Спектроскопия ЯМР в растворе является информативным инструментом для описания переворотов ароматических колец, поскольку две симметрично эквивалентных пары спинов находятся в разном химическом окружении (*ortho*-СН или *meta*-СН в позициях  $H^e-C^e$  или в позициях  $H^d-C^d$  соответственно). Вид спектра ЯМР меняется по мере увеличения частоты переворотов: при малой частоте спектр состоит из двух наборов хорошо различимых пиков, увеличение частоты приводит к уширению спектральных линий, и, наконец, при больших частотах наблюдается один набор пиков, усредненный по двум состояниям. Растущий арсенал экспериментальных подходов позволяет количественным образом охарактеризовать подобного рода обменные процессы [223—226]. Изучение зависимости ЯМР спектра от температуры и давления, позволяет извлечь ценную информацию о переходном состоянии и свободном объеме, которые возникают в ходе переворота ароматических колец [227—229]. Так, в одной из недавних работ исследователям удалось стабилизировать переходное состояние и изучить его при помощи структурных методов высокого разрешения [230]. Эксперименты по измерению ядерной спиновой релаксации в растворе являются другим источником ценной информации о динамике ароматических колец в структуре белка [229; 231].

Вопрос о том, каким образом упаковка молекул в кристалле белка влияет на динамику ароматических остатков, остается недостаточно изученным ввиду того, что методами рентгеноструктурной кристаллографии невозможно зарегистрировать события переворота ароматических колец. Метод ЯМР в твердом теле с вращением под магическим углом, ВМУ (от англ. magic-angle spinning, MAS), позволяет получить детальную информацию о различных вариантах упаковки белковых молекул, в том числе в кристаллах. Наряду с этим, ЯМР с ВМУ может быть применен для изучения динамики ароматических остатков [220; 232—234] и динамики белковой молекулы в целом (см. обзорные работы: [235—242]). Недавно была предложена методика ЯМР с ВМУ для количественной оценки подвижности ароматических остатков в широком диапазоне характеристических времен, основанная на селективном изотопном мечении образца и детекцией сигнала на протонах [243]. В этом подходе были использованы образцы белка с высокой степенью дейтерирования, в

которых спиновая метка  $^1\text{H}$ - $^{13}\text{C}$  вводилась в сайты  $\text{C}^\zeta$  (*para*-CH) или  $\text{C}^\epsilon$  (*meta*-CH), или  $\text{C}^\delta$  (*ortho*-CH). Современные ЯМР датчики с частотой ВМУ 40-50 кГц позволяют получать  $^1\text{H}$ - $^{13}\text{C}$  корреляционные спектры с высоким разрешением. Благодаря тому, что пара спинов  $^1\text{H}$ - $^{13}\text{C}$  хорошо изолирована, количественная информация о динамике системы не подвержена влиянию скалярных и дипольных взаимодействий со спинами, удаленными от исследуемой пары. Арсенал методов ЯМР для изучения динамики молекул в твердом теле более обширный, чем в жидкости. В дополнение к химическому сдвигу и скоростям релаксации, измерение которых доступно и в жидкости, в твердом теле появляется возможность измерить константу дипольного взаимодействия и охарактеризовать локальную динамику в микро- и миллисекундном диапазоне времен в экспериментах с одновременным использованием зависимости спектров от частоты ВМУ и амплитуды радиочастотных импульсов; к таким методикам относится измерение дисперсии релаксации вблизи вращательного резонанса (от англ. NEar-Rotary resonance Relaxation Dispersion, NERRD) [244; 245]. В константе дипольного взаимодействия содержится информация об амплитуде движения: под действием динамики, существующей в системе, происходит усреднение тензора дипольного взаимодействия по всему конформационному пространству, которое доступно межзатомному вектору. Используемые в описываемом исследовании методы, позволяют увидеть даже анизотропию основного движения, которая вызвана, например, двухпозиционными переворотами колец. Кроме того, эксперименты ЯМР с ВМУ по измерению ядерной спиновой релаксации позволяют исследовать динамику системы в широком диапазоне времен [236–238]. И хотя чувствительность измерений не одинакова для различных временных интервалов, измерение скоростей релаксации в экспериментах ЯМР с ВМУ можно осуществить для широкого диапазона времен практически без «слепых зон» [246; 247]. В частности, из экспериментов NERRD можно извлечь информацию о движениях, которые происходят в микросекундном (мкс) или наносекундном (нс) временных интервалах [247].

В описываемом исследовании методами спектроскопии ЯМР с ВМУ была исследована динамика ароматических колец остатков фенилаланина в трех кристаллических формах белка убиквитина. В экспериментах были использованы в высокой степени дейтерированные образцы с двухспиновой

меткой  $^1\text{H}$ - $^{13}\text{C}$ , введенной селективным образом. Три типа исследованных кристаллов далее будут обозначаться как MPD-ub, cubic-PEG-ub и rod-PEG-ub (в соответствии с формой кристаллов и использованным кристаллизационным агентом — метилпентандиол или полиэтиленгликоль). Упаковка молекул убиквитина в кристаллах и, в частности, упаковка боковых цепей остатков Phe, показана на Рисунке 3.1.

Три кристаллические формы отличаются не только количеством молекул убиквитина и их взаимной ориентацией внутри элементарной ячейки, но также и содержанием воды, которое равно 58 % для cubic-PEG-ub, 49 % для MPD-ub и 40 % для rod-PEG-ub. В предыдущих исследованиях для этих трех кристаллических форм была выявлена конформационная динамика основной цепи с характерным временем в микросекундном диапазоне, но отличающейся амплитудой. В наименее плотно упакованном кристалле cubic-PEG-ub было обнаружено качательное движение с амплитудой в несколько градусов (колебание молекул белка в пределах кристаллической решетки, англ. *rocking motion*) [244; 248]. Интересно, что ориентация боковых цепей для каждого из двух остатков Phe отличается в трех кристаллических формах, несмотря на то, что оба остатка расположены на поверхности белка, а не погружены внутрь гидрофобного ядра. Для структуры убиквитина высокого разрешения в растворе PDB ID: 1D3Z, полученной методом ЯМР, площадь поверхности доступная растворителю (от англ. *solvent accessible surface area*, SASA) для остатка Phe4 составляет  $59 \text{ \AA}^2$  (33 %), в то время как для остатка Phe45 соответствующее значение равно  $42 \text{ \AA}^2$  (23 %). Доступность остатка для растворителя зависит от упаковки молекул убиквитина в конкретной кристаллической форме (Рисунок 3.1 (D-I)), что, в свою очередь, определяет количество контактов атомов боковой цепи остатка Phe с атомами соседних молекул убиквитина в кристалле (Рисунок 3.1 (J)).

Для того чтобы изучить, каким образом упаковка молекул в кристаллах убиквитина влияет на динамику фенильных колец, в описываемом исследовании был использован комбинированный подход, основанный на экспериментах ЯМР в твердом теле с ВМУ и моделировании методом молекулярной динамики (МД). Методами спектроскопии ЯМР не удалось задетектировать сигнал от остатка Phe45 из-за сильного обменного уширения вследствие медленных переворотов фенильных колец

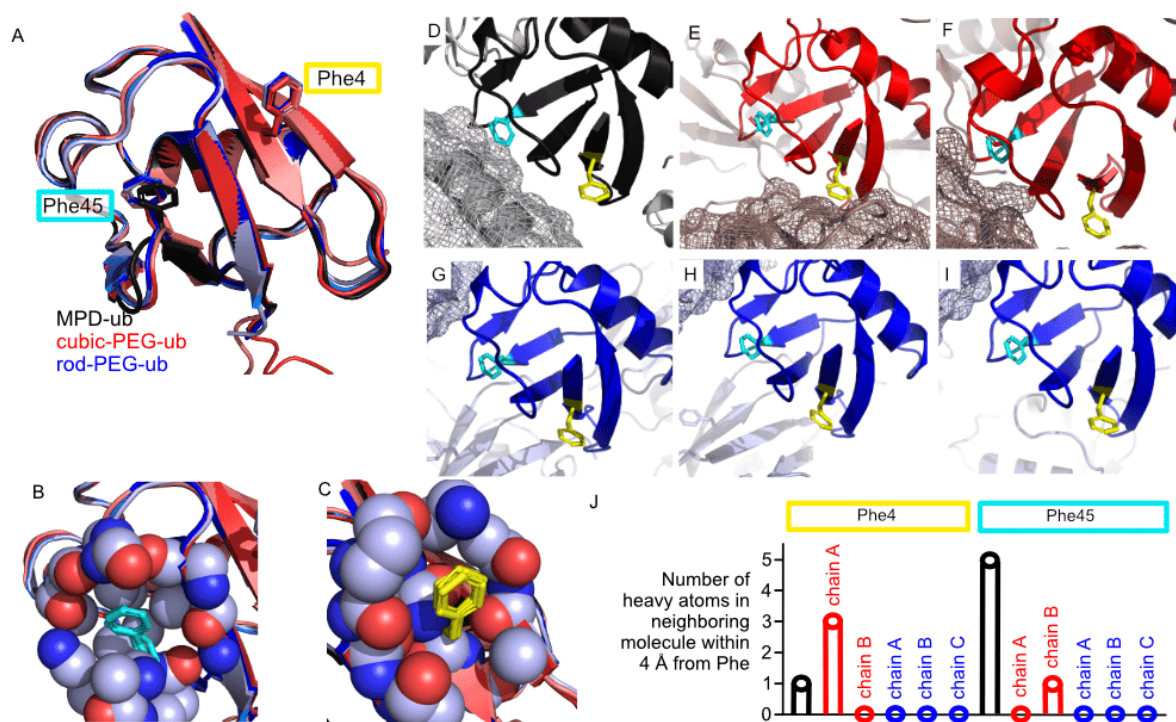


Рисунок 3.1 — Пространственная структура для трех кристаллических форм убиквитина. Согласно введенным обозначениям: "MPD-ub" (черный цвет), "cubic-PEG-ub" (красный цвет) и "rod-PEG-ub" (синий цвет). Три кристаллические формы соответствуют структурам 3ONS, 3N30 и 3ENV из базы данных PDB (от англ. Protein Data Bank). Количество неэквивалентных молекул в элементарной ячейке равно: 1 (MPD-ub), 2 (cubic-PEG-ub) и 3 (rod-PEG-ub). Панель (A) Наложение атомов основной цепи белка для шести неэквивалентных молекул (1+2+3) в трех кристаллических формах. На панелях (B) и (C) в увеличенном масштабе показаны остатки Phe45 и Phe4. Сферами обозначены атомы, находящиеся в окрестности 5 Å от боковой цепи каждого из ароматических остатков (иллюстрация на примере одной из молекул кристалла rod-PEG-ub; темно-синим цветом показан атом азота, красным цветом показан атом кислорода и светло-голубым цветом атом углерода). Соседние молекулы, находящиеся в прямом контакте с центральной молекулой, отображены в виде поверхности. (J) Количество контактов для остатков Phe4 (левая часть панели (J)) или остатка Phe45 (правая часть панели (J)) в окрестности 4 Å с тяжелыми атомами соседних молекул убиквитина.

с характерным временем в микросекундном диапазоне. Одновременно с этим, для остатка Phe4 положение пиков в спектре ЯМР совпадает

для всех кристаллических форм, несмотря на разный состав буфера, значения pH и упаковку молекул в кристалле. Интересно, что в одной из кристаллических форм существует отличие в молекулярном окружении для двух неэквивалентных молекул убиквитина, которое приводит к разным значениям химических сдвигов и скоростей спиновой релаксации. В данной работе показано, что перевороты ароматических колец в большей степени определяются внутримолекулярными взаимодействиями, нежели межмолекулярными контактами для рассматриваемых трех кристаллических форм белка убиквитина. Различие в скоростях переворота фенильных колец для остатков Phe4 и Phe45 подтверждается данными ЯМР экспериментов и МД моделирования. При этом оказывается, что наблюдаемое различие напрямую не связано ни с площадью поверхности доступной растворителю, ни с конкретным ротамерным состоянием и, вероятно, определяется другими внутримолекулярными факторами.

## 3.2 Результаты и обсуждение

### 3.2.1 Селективно изотопно-меченые остатки Phe в трех кристаллических формах убиквитина

Общепринятым методом для получения спектров ЯМР с ВМУ высокого разрешения с детектированием по протонам является использование образца высокой степени дейтерирования и быстрого вращения под магическим углом [249; 250]. Для регистрации сигнала от сайтов, участвующих в химическом обмене (в том числе для амидных протонов основной цепи), наиболее простым решением является полное дейтерирование образца с последующим обратным обменом на протоны растворителя  $^1\text{H}_2\text{O}$ . Для наблюдения атомов боковой цепи (сайты, не участвующие в химическом обмене) можно воспользоваться несовершенством процедуры дейтерирования образца, т.е. присутствием небольшого количества  $^1\text{H}$  в дейтерированном образце [251; 252], или же осуществить селективное мечение при помощи

соединения-прекурсора, в котором целевая химическая группа практически полностью протонирована. Последний подход широко используется для регистрации метильных групп при проведении экспериментов ЯМР в жидкости [253; 254] и в твердом теле [255; 256]. Подобным образом можно получить сигнал ЯМР с высоким разрешением и для других боковых цепей. Селективное мечение позволяет добиться более узких линий в спектре ЯМР в сравнении с полностью протонированным образцом измеренном даже при самых высоких доступных на сегодняшний день частотах ВМУ (100 кГц) [243].

В описываемом исследовании белок убиквитин был экспрессирован рекомбинантным образом, при этом все сайты, участвующие в химическом обмене, были дейтерированы, а в два сайта  $\text{H}^\epsilon\text{-C}^\epsilon$  каждого из двух остатков фенилаланина была введена двухспиновая метка  $^1\text{H}\text{-}^{13}\text{C}$ . Такое мечение мы обозначим как  $\alpha\text{-}[^2\text{H}, ^{15}\text{N}], \text{Phe}(\epsilon 1, \epsilon 2)\text{-}^1\text{H}, ^{13}\text{C}$ . Селективность процедуры введения спиновой метки достигалась путем добавления соединения-прекурсора к бактериальной культуре перед индукцией (35 мг прекурсора-кетокислоты на литр культуры) [257], см. подробное описание в разделе 3.3, а также см. Рисунок 3.2 (А).

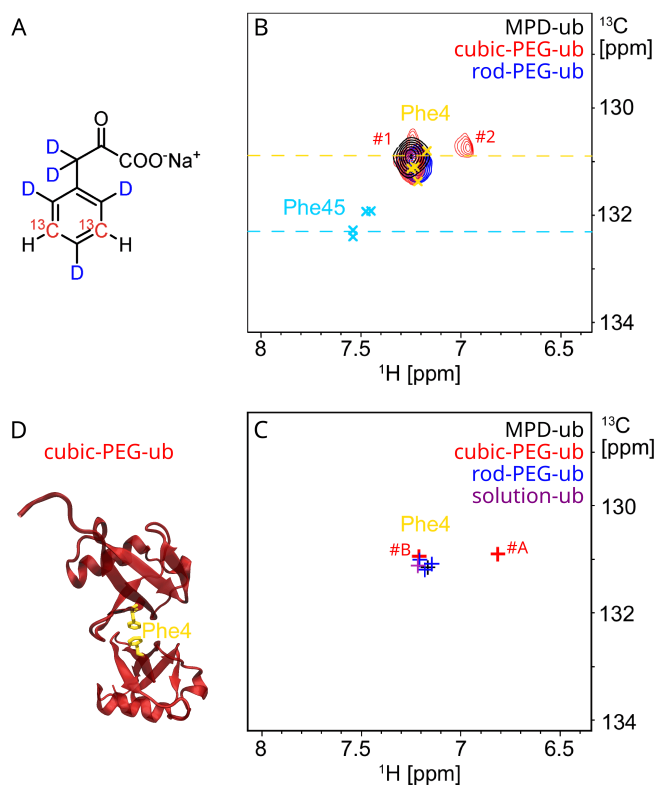


Рисунок 3.2 — Использование метода ЯМР в твердом теле с вращением под магическим углом (ВМУ) для исследования динамики остатков Phe, которые содержат двойную изотопную метку. (А) Предшественник

кетокислоты, содержащий пару спинов  $^1\text{H}$ - $^{13}\text{C}$  в двух положениях; используется для введения изотопной метки в остаток Phe. (B) Наложение корреляционных спектров  $^1\text{H}^\epsilon$ - $^{13}\text{C}^\epsilon$  для трех кристаллических форм белка убиквитина согласно данным экспериментов по кросс-поляризации (от англ. cross-polarization, CP). Для получения соответствующих ЯМР спектров была использована импульсная последовательность для  $^1\text{H}$ - $^{13}\text{C}$  кросс-поляризации и осуществлена детекция сигнала ядер  $^1\text{H}$  [243]. В спектрах кристаллов MPD-ub и rod-PEG-ub зарегистрирован один пик, в то время как для кристалла cubic-PEG-ub наблюдается два пика (на рисунке обозначены как #1 и #2). С помощью крестиков обозначено отнесение сигналов для остатков Phe4 и Phe45 белка убиквитина, известное из ранее опубликованных экспериментов ЯМР в растворе (BMRB 6457, 5387, 16228 и 27356). Горизонтальная линия золотистого цвета (для Phe4) и горизонтальная линия голубого цвета (для Phe45) соответствуют отнесению сигнала для кристаллической формы убиквитина (данные известны только для  $^{13}\text{C}$ ; BMRB 25123). (C) Численная оценка значений химического сдвига для пары спинов  $^1\text{H}^\epsilon$ - $^{13}\text{C}^\epsilon$ , принадлежащих остатку Phe4, при помощи программы SHIFTX2. Цветовая кодировка содержится в условных обозначениях на рисунке. Стоит отметить, что химический сдвиг протонов заметным образом отличается для двух неэквивалентных молекул A и B в кристалле cubic-PEG-ub. Наиболее вероятно, что сдвиг объясняется кольцевыми токами в результате стэкинг-взаимодействия ароматических колец двух остатков Phe4, принадлежащих соседним молекулам типа A в кристаллической решетке.

На Рисунке 3.2 (B) показана ароматическая область  $^1\text{H}$ - $^{13}\text{C}$  в спектрах ЯМР для каждой из кристаллических форм белка убиквитина. Показано также отнесение сигналов, известное по литературным данным из экспериментов ЯМР в жидкости для  $^1\text{H}^\epsilon$ - $^{13}\text{C}^\epsilon$  [258—261] и из экспериментов ЯМР в твердом теле с ВМУ на атомах углерода для  $^{13}\text{C}^\epsilon$  [262]. Интересно, что для всех вариантов проведения эксперимента, включая метод, использованный в описываемом исследовании, положение пиков практически совпадает даже в случае детекции сигнала в клетке и обратных мицеллах.

### 3.2.2 Перевороты ароматического кольца Phe4 происходят в субмиллисекундном временном интервале

В полученном ЯМР спектре наблюдается лишь сигнал от остатка Phe4, а отклик остатка Phe45 отсутствует (причины будут обсуждаться ниже). В случае остатка Phe4 для двух кристаллических форм MPD-ub и rod-PEG-ub в спектре ЯМР наблюдается единственный пик. Для кристалла MPD-ub, элементарная ячейка которого содержит единственную молекулу убиквитина, подобный результат ожидаем. В тоже время, для кристалла rod-PEG-ub, элементарная ячейка которого содержит три неэквивалентные молекулы убиквитина, результат менее очевиден при условии того, что в эксперименте для многих амидных сайтов основной цепи регистрируется три различных сигнала [248]. Из подобного поведения можно заключить, что остаток Phe4 имеет схожее химическое окружение в каждой из трех неэквивалентных молекул убиквитина, что и объясняет появление единственного сигнала в спектре ЯМР. В свою очередь, в кристалле cubic-PEG-ub наблюдается два сигнала, а именно: положение одного из пиков (обозначен как (#1)) совпадает с координатами пиков в других кристаллических формах, в то время как второй пик (обозначен как (#2)) смещен примерно на 0.25 миллионных долей (мд) в высокое поле по оси протонных химических сдвигов (см. Рисунок 3.1). Причиной удвоения сигнала является разное химическое окружение для остатка Phe4 в двух неэквивалентных молекулах убиквитина в кристалле cubic-PEG-ub (см. Рисунок 3.2 (E-J)). Можно предположить, что пик (#1), положение которого близко к сигналу ЯМР в жидкости, относится к молекуле В, поскольку не образует межмолекулярных контактов, а пик (#2) принадлежит молекуле А, которая вступает во взаимодействие с соседней молекулой в кристалле cubic-PEG-ub.

Для проверки этой гипотезы при помощи программы SHIFTX2 был осуществлен расчет химических сдвигов [263]  $^1\text{H}^\epsilon\text{-}^{13}\text{C}^\epsilon$  для молекулы убиквитина в растворе и в трех кристаллических формах. Заранее стоит отметить, что существующие вычислительные инструменты для оценки химических сдвигов в боковых цепях аминокислот имеют ограниченную точность. Дело в том, что, на сегодняшний день, объем доступных



экспериментальных данных оказывается недостаточным для построения релевантной модели. Тем не менее, используя численные методы, удалось воспроизвести результаты для остатка Phe4. В частности, для молекулы А в кристалле cubic-PEG-ub подтверждается сдвиг сигнала вверх по полю вдоль протонной оси на 0.39 мд относительно сигнала от молекулы В. Сдвиг сигнала обусловлен кольцевыми токами при стэкинг-взаимодействии пары ароматических колец остатка Phe4, принадлежащих двум соседним молекулам в кристалле cubic-PEG-ub (см. Рисунок 3.2 (D)). При этом если провести расчет при помощи SHIFTX2 для каждой из молекул в отдельности, а затем совместить полученные результаты, то окажется, что два сигнала сливаются в один пик. Эффект влияния кольцевых токов на экранирование протонов в сопряженных ароматических системах достаточно хорошо изучен и есть все основания полагать, что удвоение сигнала для остатка Phe4 объясняется именно межмолекулярным стэкингом ароматических колец [264; 265].

Несовершенство численных методов оценки химических сдвигов становится очевидным, если обратиться к результатам для остатка Phe45. Алгоритм SHIFTX2 правильным образом предсказывает относительное положение пиков (химические сдвиги для остатка Phe45 не отличаются для убиквитина в растворе и трех кристаллических формах), однако абсолютные значения плохо согласуются с данными эксперимента (см. Рисунок 3.7).

Ввиду того что для остатка Phe4, который содержит две изотопные метки в еположениях, наблюдается единственный пик, можно предположить, что в результате субмиллисекундной динамики, связанной с переворотами фенильного кольца, происходит усреднение изотропной компоненты химического сдвига. Изучение процесса усреднения дипольного взаимодействия при помощи ЯМР с ВМУ позволяет извлечь более подробную информацию о динамике ароматических колец. Из литературы известно, что тензор дипольного взаимодействия для пары  $^1\text{H}$ - $^{13}\text{C}$  усредняется в результате быстрых движений с характерным временем меньше чем 10-100 мкс (Рисунок 28 в источнике [237]). При перевороте ароматических колец полного усреднения не происходит, анизотропная часть дипольного тензора уменьшается до теоретического значения 62.5 % (параметр порядка  $S=0.625$ ; англ. order parameter). Более того, в

пределе жесткой системы (от англ. rigid-limit) аксиально-симметричный тензор становится биаксиальным (от англ. biaxial). Термин биаксиальный соответствует используемому термину асимметричный тензор с параметром асимметрии  $\eta$ , где  $\eta=0$  для аксиально-симметричного тензора. Поскольку термин биаксиальный более точно отражает форму тензора и зачастую применяется в других областях физики [266], в описываемом исследовании используется именно этот термин. Можно показать, что повороты ароматических колец не только уменьшают анизотропную часть дипольного тензора до значения  $S=0.625$ , но также приводят к увеличению параметра биаксиальности  $\eta=0.6$  [243]. Адаптированная версия эксперимента REDOR (от англ. Rotational Echo DOuble Resonance, REDOR) [267] позволяет определять параметр порядка и параметр биаксиальности тензора [256]. На Рисунке 3.3 (А) показаны кривые REDOR для трех кристаллических форм убиквитина. Для всех кристаллических форм полученные экспериментальные параметры хорошо согласуются с теоретической моделью для переворота ароматического кольца, т.е. в пределах погрешности параметр порядка соответствует ожидаемому значению 0.625, а доверительный интервал для параметра биаксиальности включает ожидаемое значение 0.6 (см. Рисунок 3.3 (В)). Для некоторых кристаллов параметр порядка и биаксиальности несколько ниже, чем для других, что, вероятно, связано с присутствием дополнительных движений малой амплитуды. Тем не менее очевидно, что в пределах допустимого уровня погрешности, экспериментальные наблюдения могут быть объяснены исключительно переворотами ароматических колец.

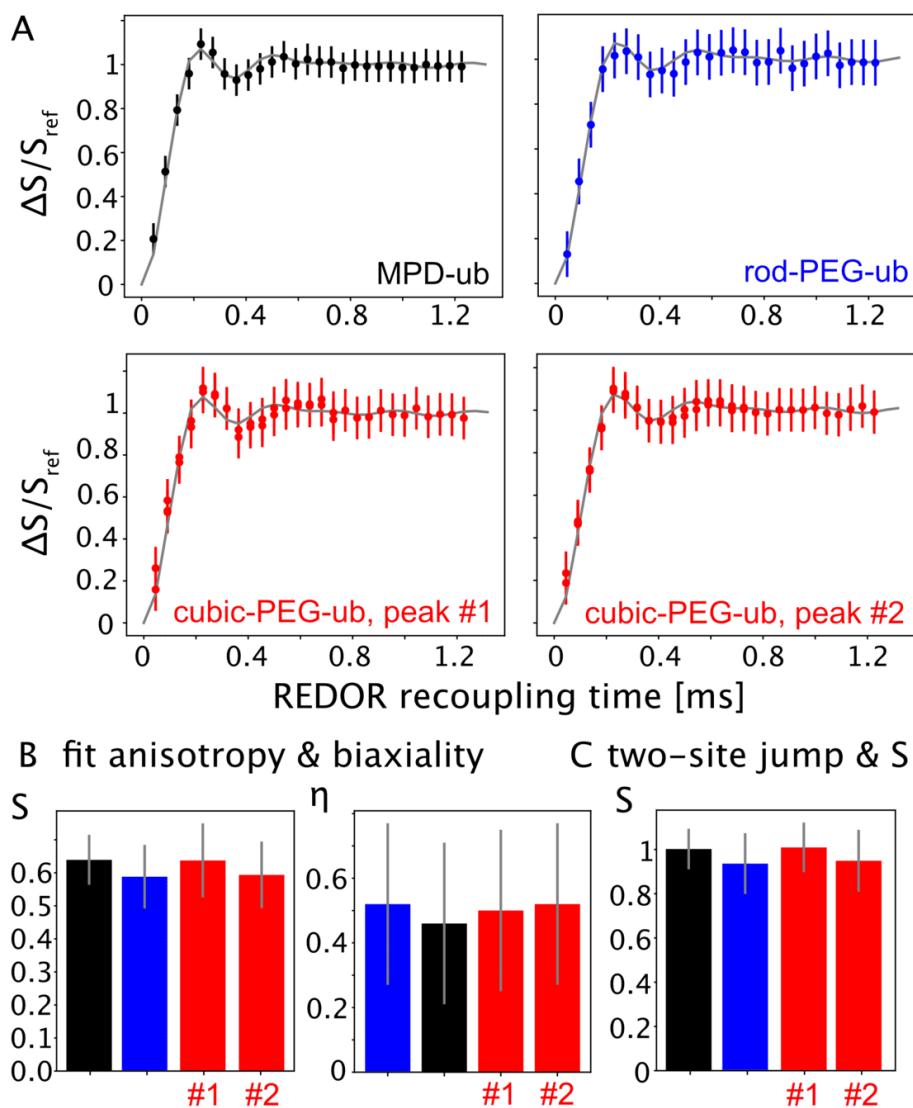


Рисунок 3.3 — Измерение параметров тензора дипольного взаимодействия для пары спинов  $^1\text{H}$ - $^{13}\text{C}$ , принадлежащих остатку Phe4. (A) Эксперимент REDOR. Рекаплинг-кривые для различных кристаллических форм. (B) Оптимизация параметров тензора дипольного взаимодействия путем поиска по сетке значений (параметр порядка  $S$  (параметр анизотропии) и параметр биаксиальности (асимметрии)  $\eta$ ). (C) Оптимальное значение параметра порядка для модели прыжков между двумя эквивалентными состояниями ( $120^\circ$ ), выявленное путем поиска по сетке значений (параметр порядка, тензор анизотропии  $\delta_D$ ). Найденное таким образом значение параметра порядка ( $S = \delta_{D,\text{fitted}}/\delta_{D,\text{rigid}}$ ) близко к 1. Этот факт указывает на то, что значения параметров тензора дипольного взаимодействия в большей степени определяются динамикой, связанной со скачкообразными прыжками между двумя состояниями, в то время как другие движения в системе не вносят заметного вклада.

### 3.2.3 Перевороты ароматического кольца Phe4 происходят на временных масштабах 10-20 нс

Скорости ядерной спиновой релаксации чувствительны к амплитуде и характерным временам для движений, присутствующих в системе. В описываемом исследовании были измерены скорости продольной спиновой релаксации  $R_1$  и скорости спиновой релаксации во вращающейся системе координат  $R_{1\rho}$  ядер  $^{13}\text{C}$ , а также скорости спиновой релаксации во вращающейся системе координат ( $R_{1\rho}$ ) ядер  $^1\text{H}$ . Полученные значения были использованы для определения скорости переворота ароматических колец. С этой целью было осуществлено сравнение экспериментальных  $^{13}\text{C}$  скоростей релаксации (см. Рисунок 3.4 (А, В), а также Таблицу 8) и расчетных значений скоростей релаксации по модели, в которой предполагается, что в системе доминирует движение, связанное с переворотом ароматических колец (см. Рисунок 3.4 (D)).

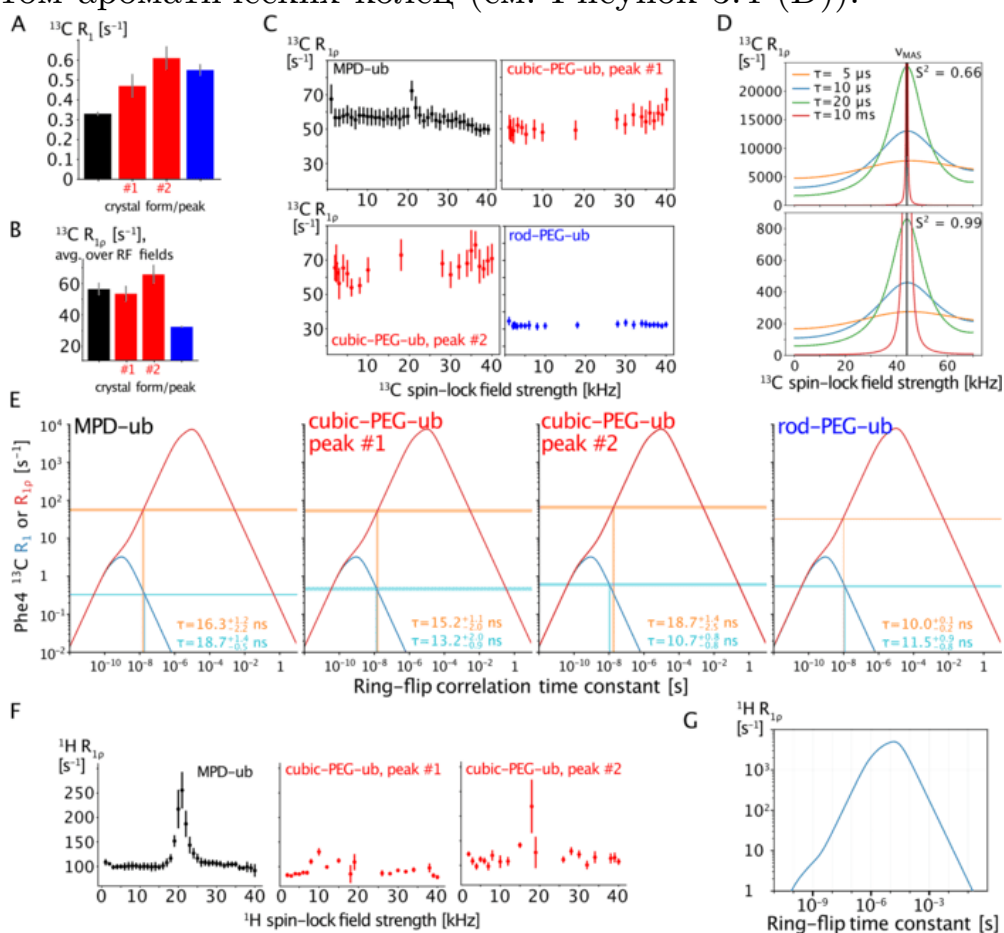


Рисунок 3.4 — Результаты ЯМР экспериментов с ВМУ для остатка Phe4 в трех кристаллических формах. Данные были получены на

частоте ВМУ 44.053 кГц. Цветовая кодировка соответствует трем кристаллическим формам (черный, красный и синий цвет). (А) Значения  $^{13}\text{C}$  скорости продольной релаксации ( $R_1$ ) и (В) скорости релаксации во вращающейся системе координат ( $R_{1\rho}$ ). (С) Измерение  $^{13}\text{C}$   $R_{1\rho}$  как функции мощности импульса спин-лока (дисперсия релаксации). При приближении к условию  $\nu_{\text{RF}} = \nu_{\text{MAS}}$ , профили скоростей релаксации остаются практически неизменными (см. раздел 3.5.1), что указывает на отсутствие динамики в микросекундном диапазоне времен. (D) Расчетные профили  $^{13}\text{C}$  NERRD для варьируемых значений характерных времен и зафиксированного значения параметра порядка, которое было установлено равным либо  $S=0.66$  (см. верхняя панель; движение, связанное с переворотом ароматических колец), либо равным  $S=0.99$  (см. нижняя панель; покачивания — движения с амплитудой меньшей, чем при перевороте ароматического кольца). Вращательный резонанс на частоте ВМУ равной 44.053 кГц показан черной вертикальной прямой. (E) Определение корреляционного времени для переворота ароматического кольца остатка Phe4 по данным  $^{13}\text{C}$  скоростей релаксации. Профили расчетных значений скоростей релаксации показаны в виде кривых  $\Lambda$ -формы, рассчитанных как функция временного масштаба событий переворота ароматических колец ( $R_{1\rho}$  оранжевый цвет;  $R_1$  синий цвет). Горизонтальные линии соответствуют экспериментальным значениям скоростей релаксации, а вертикальные пунктирные линии показывают отсечку для быстрого динамического режима. Экспериментальные значения содержатся в Таблице 8. Наблюдается хорошее согласие между скоростями переворота, извлеченными из данных  $R_{1\rho}$  и  $R_1$ . (F) Профили дисперсии релаксации  $^1\text{H}$   $R_{1\rho}$ . Для кристаллической формы MPD-ub заметен рост значений  $R_{1\rho}$  при условии  $R_{1\rho}$  (условия HORROR-[268]), благодаря восстановлению дипольных гомоядерных взаимодействий  $^1\text{H}$ - $^1\text{H}$ . Эффект наблюдается в узком диапазоне частот (несколько кГц) вблизи половины частоты вращения. (G) Расчетные значения скоростей релаксации  $^1\text{H}$   $R_{1\rho}$  (см. раздел 3.5.1).

При расчетах скоростей релаксации параметр порядка для вектора  $^1\text{H}$ - $^{13}\text{C}$  был зафиксирован в соответствии с ожидаемым теоретическим значением, варьировалось только характерное корреляционное время. В

диапазоне характерных времен 10-20 нс расчетные скорости релаксации совпадают с экспериментальными значениями для всех кристаллических форм, в том числе для двух неэквивалентных сайтов в cubic-PEG-ub. Значения времен корреляции, извлекаемые из данных  $R_{1\rho}$  and  $R_1$  хорошо согласуются друг с другом. Этот факт подтверждает предположение о том, что релаксационные процессы в системе объясняются единственным доминирующим движением, связанным переворотом ароматического кольца. Кроме того, был оценен вклад в наблюдаемые значения скоростей релаксации движений, отличных от переворота ароматического кольца. С этой целью была рассмотрена расширенная модель, в которой предполагалось существование дополнительных динамических режимов. Значения соответствующих параметров были определены на основе данных МД моделирования (см. Рисунок 3.8 и раздел 3.5.2). Анализ скоростей релаксации  $R_1$  и  $R_{1\rho}$  ядер  $^{13}\text{C}^\epsilon$ , проведенный с использованием расширенной модели, показал, что дополнительные движение не существенны для извлекаемых значений констант скорости переворота ароматических колец (см. Рисунок 3.9.)

И хотя полученные данные однозначно указывают на то, что перевороты ароматических колец происходят в наносекундном диапазоне, при помощи экспериментов  $^{13}\text{C}$   $R_{1\rho}$  NERRD было исследовано наличие у остатка Phe4 медленной динамики с малой амплитудой [244; 245]. В NERRD экспериментах скорость релаксации  $R_{1\rho}$  рассматривается как функция мощности радиочастотного импульса спин-лока. При наличии движений в микросекундном диапазоне значения скоростей релаксации  $R_{1\rho}$  резко увеличиваются, когда частота спин-лока приближается к частоте ВМУ. Кроме того, в качестве дополнительной проверки на присутствие в системе микросекундных движений была использована модификация эксперимента  $^1\text{H}$   $R_{1\rho}$  NERRD (несколько более сложная для получения количественных оценок) [269; 270]. На Рисунке 3.4 (С, F) показаны  $^{13}\text{C}$  и  $^1\text{H}$  экспериментальные профили NERRD. В целом, для всех образцов профили NERRD плоские, что указывает на отсутствие динамики микросекундном диапазоне. Наблюдаемый рост значений  $^1\text{H}$  NERRD для  $\nu_{\text{RF}} = \frac{1}{2}\nu_{\text{MAS}}$  обусловлен восстановлением дипольных гомоядерных взаимодействий при условии HORROR [268]. Этот эффект распространяется на довольно узкий диапазон частот, намного меньше

ожидаемого при наличии микросекундной динамики [270]. Важно отметить, для остатка Phe4 экспериментальное значение  $^1\text{H}$   $R_{1\rho}$  составляет 50-120  $\text{с}^{-1}$ , что оказывается выше значения, ожидаемого для движения с характерным временем в интервале 10-20 нс (соответствует 30-50  $\text{с}^{-1}$ ; см. Рисунок 3.4G)). Это позволяет предположить, что потеря когерентности фаз, вызванная дипольным взаимодействием, ответственна более чем за половину ожидаемой скорости затухания сигнала [271; 272].

Отталкиваясь от результатов расчета, показанных на Рисунке 3.4, можно дать объяснение отсутствию сигнала для остатка Phe45. Если предположить, что переворот ароматического кольца для остатка Phe45 происходит на 1-2 порядка медленнее (молекулярная динамика дает фактор замедления около 50; см. ниже), чем для остатка Phe4, то времена релаксации  $^1\text{H}$  и  $^{13}\text{C}$  для остатка Phe45 будут составлять около 1 мс (а дополнительная дипольная расфазировка сократила бы время жизни  $^1\text{H}$  еще больше). Такая быстрая релаксация ослабит передачу когерентности в ходе эксперимента и уширит сигналы настолько, что они станут ненаблюдаемы. Следовательно, намагниченность Phe45 будет быстро затухать во время последовательности импульсов и во время детекции, уничтожая спектральный сигнал.

Интересно отметить, что в спектре кристаллической формы cubic-PEG-ub для пика #2 (относится к молекуле А) значения скоростей релаксации  $^1\text{H}$  и  $^{13}\text{C}$   $R_{1\rho}$  выше, чем для пика #1 (относится к молекуле В). Вероятно, этот факт является причиной того, что интенсивность для пика #2 ниже. Ввиду того, что скорости переворота ароматического кольца довольно схожи, причина более быстрой релаксации в случае пика #2 недостаточно ясна. Возможным объяснением является более близкое расположение протонов для молекулы А по отношению к соседней молекуле в кристалле (см. Рисунок 3.1 J).

Наконец, следует упомянуть, что молекулы в кристалле cubic-PEG-ub вовлечены в качательное движение, в то время как в более плотно упакованных кристаллах MPD-ub и rod-PEG-ub подобная форма движения не обнаруживается [244; 248]. Характерное время для подобных покачиваний составляет десятки микросекунд, как это следует из профилей  $^{15}\text{N}$  NERRD. В этом отношении интересно, что профили  $^{13}\text{C}$  NERRD для кристаллической формы cubic-PEG-ub (но не для двух других

кристаллических форм) показывают небольшую тенденцию к росту значений при соблюдении условия  $\nu_{RF} = \nu_{MAS} = 44$  кГц. По всей вероятности, слабый тренд в профилях NERFD объясняется тем, что динамика боковых цепей чувствительна к общему качательному движению молекул в кристалле. Однако на количественном уровне эффект оказывается ниже, чем можно было бы ожидать исходя из параметров характеризующих качательное движение. Данное наблюдение позволяет сделать вывод о том, что динамика боковой цепи практически не коррелирует с динамикой основной цепи. Кроме того, обнаружить качательное движение на фоне внутренней динамики для пары  $^1\text{H}-^{13}\text{C}$  в боковой цепи ароматических остатков намного сложнее, чем в случае амидных групп основной цепи, так как абсолютные значения константы, характеризующей скорость динамики, в 10 раз выше для  $^1\text{H}-^{13}\text{C}$ , чем для амидной группы  $^{15}\text{N}$ .

В совокупности, данные ЯМР экспериментов указывают на то, что для всех кристаллических форм характерное время переворота ароматического кольца для остатка Phe4 составляет около 10-20 нс. Для остатка Phe45 перевороты ароматического кольца происходят реже, что препятствует обнаружению сигнала в спектре ЯМР. В следующей секции приведено подтверждение этих выводов при помощи данных моделирования методом молекулярной динамики.

### 3.2.4 МД моделирование переворотов ароматических колец

Метод молекулярной динамики (МД) представляет собой дополнительный источник ценной информации о событиях переворота ароматических колец. В настоящей работе для молекулы убиквитина в растворе и в трех типах кристаллической решетки были записаны МД траектории длиной в несколько микросекунд. Данные МД моделирования были использованы для того, чтобы объяснить результаты эксперимента, а именно: различие в частоте событий переворота ароматических колец для остатков Phe4 и Phe45, а также тот факт, что сигнал от остатка Phe45 не наблюдаем в эксперименте. На Рисунке 3.5 показана зависимость значений двугранных углов  $\chi_1$  и  $\chi_2$  от времени моделирования для боковых цепей



двух остатков Phe4 и Phe45 в растворе и в трех кристаллических формах. При этом общее время моделирования достаточно велико, поскольку длина траектории для индивидуальной молекулы убиквитина составляет две микросекунды, а общее количество молекул в кристаллической ячейке от 24 до 48 штук.

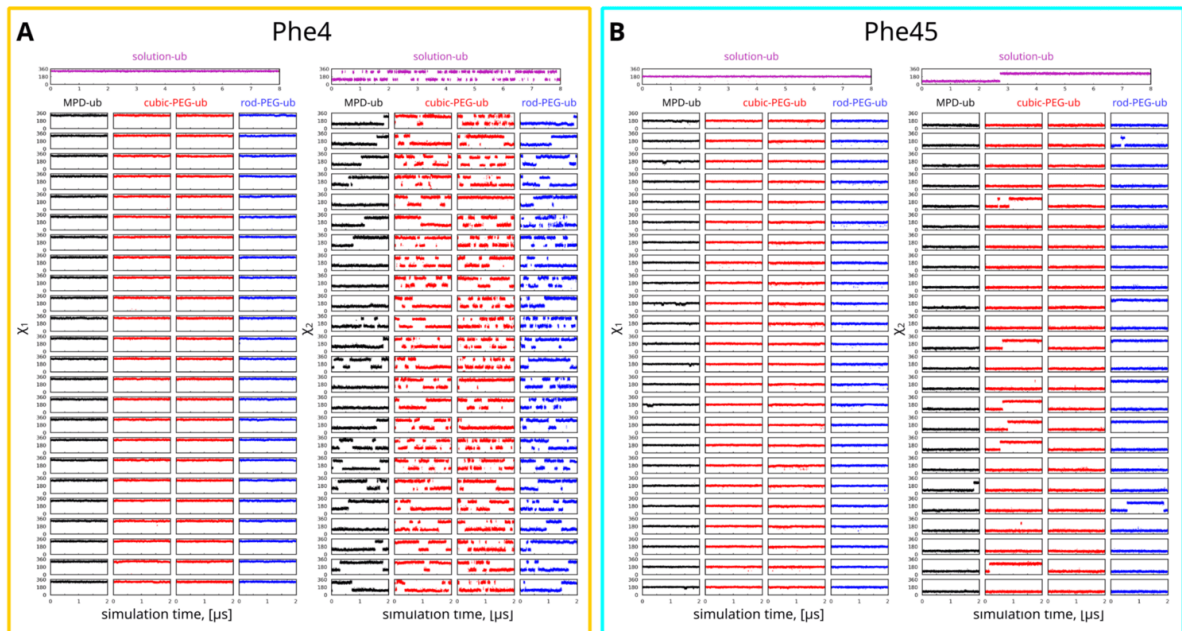


Рисунок 3.5 — Зависимость значений торсионных углов  $\chi_1$  и  $\chi_2$ , принадлежащих боковой цепи остатков (А) Phe4 и (В) Phe45, от времени моделирования белка убиквитина в растворе и в трех кристаллических формах. Для более наглядной визуализации ротамерных прыжков вместо стандартного диапазона  $[-180-180^\circ]$  был использован диапазон значений  $[0-360^\circ]$ . Цветовая кодировка согласуется с введенной ранее: черный цвет (кристалл MPD-ub), красный цвет (кристалл cubic-PEG-ub), синий цвет (кристалл rod-PEG-ub) и пурпурный цвет (моделирование в жидкости). Ячейка моделирования для cubic-PEG-ub кристалла содержит 48 молекул убиквитина, распределенных поровну между двумя неэквивалентными молекулами типа А (молекулы 1–24) и типа В (молекулы 25–48). Ячейка моделирования для rod-PEG-ub кристалла содержит 24 молекулы убиквитина, распределенных поровну между тремя неэквивалентными молекулами типа А (молекулы 1–8), типа В (молекулы 9–16) and типа С (молекулы 17–24). Более детально протокол моделирования убиквитина в трех кристаллических формах обсуждается в разделе 3.3

Двугранный угол  $\chi_1$ , изменение которого соответствует повороту вокруг оси  $C^\beta-C^\gamma$ , фактически остается постоянным в течение времени моделирования для каждой рассматриваемой системы. Более интересная ситуация наблюдается в отношении двугранного угла  $\chi_2$ , изменение которого соответствует движению ароматического кольца (переворот и покачивания с малой амплитудой относительно равновесного значения). Для каждой из рассматриваемых систем зарегистрированы случаи переворота ароматического кольца, принадлежащего остатку Phe4. Для кристаллических форм cubic-PEG-ub и rod-PEG-ub перевороты ароматических колец происходят достаточно часто за время моделирования в 2 мкс. В случае кристаллической формы MPD-ub перевороты ароматических колец наблюдаются чуть реже, чем в случае PEG-ub кристаллов, хотя согласно экспериментальным данным для трех кристаллических форма не наблюдается значимых различий. В совокупности, по данным всех МД траекторий в растворе и в кристаллической форме характерное время переворота ароматического кольца остатка Phe4 составляет  $\tau_{div}=204$  нс (характерные времена для индивидуальных траекторий см. в Таблице 9;  $\tau_{div}$  — общее время моделирования, разделенное на общее количество событий переворота). Результат, полученный по данным МД моделирования, на порядок превышает значение в 10-20 нс, известное из экспериментов по измерению скоростей релаксации. Расхождение соответствует переоценке высоты барьера на 1.5 ккал/моль, что находится на уровне погрешности самых современных силовых полей [273].

В случае остатка Phe45 перевороты ароматического кольца происходят гораздо реже, чем в случае остатка Phe4, что справедливо для всех МД траекторий в растворе и в трех кристаллических формах. В частности, данные МД моделирования указывают на то, что соответствующая константа скорости переворотов ароматического кольца для остатка Phe45 в 47 раз ниже, чем для остатка Phe4. Полученные значения константы скорости для остатка Phe45 можно считать лишь приблизительной оценкой ввиду плохой статистики (небольшого количества событий переворота в МД траекториях). Тем не менее данные МД моделирования однозначным образом указывают на более медленную локальную динамику боковой цепи остатка Phe45, в сравнении с остатком

Phe4. Более медленная динамика для остатка Phe45 препятствует его обнаружению в спектрах ЯМР с ВМУ из-за сильного уширения спектральной линии.

### 3.3 Материалы и методы

#### 3.3.1 Получение и очистка белков

Ниже приведен протокол бактериальной экспрессии, который был использован для получения дейтерированного образца убиквитина, в котором остатки фенилаланина содержат  $^1\text{H}$ - $^{13}\text{C}$  изотопные метки в  $\epsilon$ -положениях. Сначала была осуществлена трансформация плазмиды pET21b, несущей ген убиквитина человека, в клетки штамма *Escherichia coli* BL21(DE3). После чего трансформированные клетки росли в течение 48 часов на среде M9/D<sub>2</sub>O, содержащей в качестве единственных источников азота и углерода 1 г/л  $^{15}\text{ND}_4\text{Cl}$ , 2 г/л D-глюкозы-d<sub>7</sub>. Далее бактериальная культура росла при температуре 37 °C. Когда оптическая плотность на длине волны 600 нм (OD<sub>600</sub>) достигала значения 0.65, на 1 литр культуры добавляли 35 мг прекурсора-кетокислоты (см. Рисунок 3.2 (А)). Через час, при постоянном перемешивании на шейкере при 37 °C, когда значение OD<sub>600</sub> достигло 0.95, к культуре добавлялся индуктор ИПТГ (Изопропил- $\beta$ -D-1-тиогалактопиранозид) в концентрации 1 мМ. Индукция проводилась в течение 3 часов при 37 °C. Финальное значение OD<sub>600</sub> составило 2.2.

По завершению этапа индукции, клетки ресуспендировали в 20 мл буфера 50 мМ Трис-НСl (рН 8.0) в присутствии ингибиторов пептидаз: 2 мкг/мл лейпептина и 2 мкг/мл пепстатина. Разрушение клеток осуществлялось при помощи ультразвука. Полученный лизат центрифугировали при 46000 об./мин в течение 30 минут с использованием ротора JA25-50. После чего супернатант подвергался диализу в 300 мл буфера 50 мМ Трис-НСl (рН 8.0). Затем образец вновь центрифугировали при 46000 об./мин в течение 30 минут и далее наносили на анионообменную хроматографическую на колонку Q-Sepharose объемом

40 мл. Фракции, содержащие убиквитин, концентрировали и наносили на гель-фильтрационную колонку Superdex 200 10/300 G, предварительно уравновешенную 50 mM Tris-HCl (pH 8.0). Далее образец подвергали диализу против воды Milli-Q Ultra до полного удаления буфера. И наконец, образец белка был лиофильно высушен в течение 24 часов.

### 3.3.2 Кристаллизация белков

В описываемом исследовании были использованы те же кристаллы убиквитина, что и в предыдущем исследовании, отличие состояло в том, что изотопные метки вводились в другие сайты [248]. Кристаллы были получены по методу сидящей капли (от англ. sitting-drop) в буферных растворах, состав которых описан ниже. Во всех случаях кристаллы были получены с использованием планшетов для кристаллизации (Hampton research Cryschem plate, HR3-159) для этого смешивали 40 мкл образца с 450 мкл маточного раствора (раствор с осадителем, против которого кристаллизуется белок, англ. reservoir buffer).

Для получения MPD-ub кристаллов убиквитин помещался в буфер А с концентрацией 20 мг/мл (20 mM ацетата аммония, pH 4.3). Далее буфер В (50 mM цитрата, pH 4.1) смешивали с метилпентадиолом (MPD) в соотношении 40:60 (буфер В:MPD) и 450 мкл этой смеси добавляли к маточному буферу в лунки на планшете. В "сидячей капле" 37 мкл раствора убиквитин / буфер А смешивали с 10 мкл раствора буфер В:MPD. Для обеспечения герметичности планшет покрывали клейкой лентой CrystalClear и держали при 4 °C. Через 1-2 недели вырастали кристаллы игольчатой формы (от англ. sea-urchin like).

Для получения cubic-PEG-ub кристаллов маточный буфер был приготовлен из 450 мкл буфера С (100 mM 2-этансульфоновой кислоты (pH 6.3), 20 % (по массе) PEG 3350 и 100 mM ацетата цинка). Раствор белка (20 мг/мл убиквитина) в буфере D (20 mM ацетата аммония, pH 4.8) смешивали с буфером С в соотношении 1:1 и помещали 45 мкл этого раствора в лунку на планшете для сидящей капли. В течение одной недели при температуре около 23 °C были получены cubic-PEG-ub кристаллы.

Для получения rod-PEG-ub кристаллов маточный буфер был приготовлен из буфера E (50 mM HEPES) (pH 7.0), 25 % PEG 1500 и 25 mM ацетата цинка). Белок помещали в буфер D в соотношении 20 мг/мл и затем смешивали с маточным буфером 1:1 по аналогии с протоколом для cubic-PEG-ub.

Сформировавшиеся кристаллы белка переносили в специально изготовленную центрифугу (в специальную воронку, установленную поверх ВМУ ротора Bruker, диаметром 1.3 мм; размеры воронки соответствовали размерам ротора Beckman SW32). Кристаллы каждого типа (около 3 мг белка) центрифугировали в отдельные ВМУ роторы, затем крышки ВМУ роторов приклеивали двухкомпонентным эпоксидным клеем, чтобы избежать потерь воды.

### 3.3.3 ЯМР эксперимент

Все эксперименты были проведены на спектрометре Bruker Avance III с рабочей частотой на ядрах  $^1\text{H}$  600 МГц (14.1 Тл), оборудованном датчиком ВМУ (вращение под магическим углом) с диаметром ротора 1.3 мм. Спектрометр оснащен основной радиочастотной (РЧ) катушкой, настроенной на частоты ядер  $^1\text{H}$ ,  $^{13}\text{C}$  и  $^{15}\text{N}$ , а также дополнительной катушкой на частоте  $^2\text{H}$ . Частота вращения ротора была установлена на 40-50 кГц (указана на рисунках и ниже) и поддерживалась постоянной с точностью до 10 Гц. Эффективная температура образца составляла около 28 °С. Температуру определяли по сигналу от кристалла MPD (2-метил-2,4-пентандиол) (не чувствителен к изменению температуры) и по сигналу растворителя (воды), используя уравнение  $T[^\circ\text{C}] = 455 - 90 \cdot \delta_{\text{H}_2\text{O}}$ , где  $\delta_{\text{H}_2\text{O}}$  — сдвиг сигнала воды, выраженный в миллионных долях (мд).

Импульсная последовательность для корреляционных экспериментов (от англ. correlation experiments, hCN) представлена на Рисунке 3.8 в работе [243]. Импульсная последовательность включает в себя возбуждение ядер  $^1\text{H}$  с последующим переносом намагниченности путем кросс-поляризации на ядра  $^{13}\text{C}$  для регистрации химического сдвига (с развязкой порядка 10 кГц  $^1\text{H}$  WALTZ-16). Далее следует

преобразование когерентности  $^{13}\text{C}$  в когерентность  $^{13}\text{C}_z$  с последующим подавлением сигнала воды в течение 40 мс. Для подавления сигналов воды применялась серия импульсов на частоте резонанса ядер  $^1\text{H}$  с амплитудой 18 кГц, длительностью 820 мкс и модуляцией фазы ( $\pm 15^\circ$ ). В непрямом измерении сигнал ядер  $^{13}\text{C}$  регистрировался в течение 12-15 мс с шириной спектрального окна 15 мд. (2250 Гц); измерение сигнала ядер  $^1\text{H}$  проводилось за время около 50 мс.

При переносе намагниченности путем кросс-поляризации (от Н до С и от С до Н) были использованы РЧ импульсы длительностью 400 мкс и мощностью 85 кГц и 35 кГц при частоте ВМУ равной 50 кГц; для возбуждения ядер  $^1\text{H}$  подавались импульсы с постепенным увеличением мощности от 90 % до 100 % при этом указанная амплитуда РЧ импульса соответствует значению в средней точке. Длительность жестких рч импульсов составляла от 2.5 до 2.6 мкс для  $^1\text{H}$ , от 3.4 до 3.5 мкс для  $^{15}\text{N}$  и 3.2 мкс для  $^{13}\text{C}$ .

Эксперименты REDOR со сдвигом по времени,  $^{13}\text{C} R_1$ ,  $^{13}\text{C} R_{1\rho}$  и  $^1\text{H} R_{1\rho}$  опирались на стандартный метод hCN корреляции, импульсная последовательность для которого представлена на Рисунке 3.8 в ранее опубликованной работе [243]. В эксперименте по измерению  $^1\text{H} R_{1\rho}$  спин-лок применялся непосредственно после возбуждения ядер  $^1\text{H}$ . За исключением эксперимента по измерению релаксации ядер  $^1\text{H}$  все последовательности были составлены при помощи NMRlib [274].

Эксперимент REDOR проводился при частоте ВМУ равной 4.053 кГц (период вращения ротора составлял 22.7 мкс). В последовательности рекаплинг-импульсов (от англ. recoupling — восстановление взаимодействия) в эксперименте REDOR длительности  $^1\text{H} \pi$  и  $^{13}\text{C} \pi$  импульсов составляли 5 мкс и 6 мкс соответственно. Эксперимент REDOR был реализован со сдвигом половины импульсов  $^1\text{H}$  от центра периода вращения ротора, как описано ранее [256; 275]. Смещение было таковым, что самый короткий промежуток времени между двумя  $^1\text{H}$  импульсами составлял 0.5 мкс, т.е. центры двух последовательно приложенных импульсов были разделены промежутком времени равным 5.5 мкс. Таким образом, регистрировалось 17 временных точек в диапазоне от 2 до 36 периодов вращения ротора с шагом в 2 периода (по одному с каждой стороны от центрального импульса  $^{13}\text{C}$ ). Полученные данные REDOR записывались как серия одномерных спектров.

Для измерения  $^{13}\text{C}_{1\rho}$  было записано десять одномерных экспериментов с отличающейся длительностью импульса спин-лока от 1 до 45 мс, которые были реализованы для 20 значений амплитуды спин-лока в диапазоне от 2 до 40 кГц. В середине релаксационного периода подавался  $\pi$  импульс на ядра  $^1\text{H}$  для подавления эффектов кросс-корреляционной релаксации [276]. В ходе эксперимента по измерению  $^1\text{H}$   $R_{1\rho}$  было записано десять одномерных спектров с действием спин-лока во временных интервалах от 1 до 25 мс, которые были реализованы для 25 вариантов значений напряженности радиочастотного поля в диапазоне от 1 кГц (для кристаллической формы MPD-ub) или от 2 кГц (для кристаллической формы cubic-PEG-ub) до 40 кГц.

Для анализа данных интегрирование пиков в спектрах ЯМР осуществлялось при помощи собственных скриптов, написанных на языке python. Профили затухания релаксационных кривых были аппроксимированы при помощи простой экспоненциальной функции. Для интерпретации данных REDOR была проведена серия расчетов рекаплинга при помощи пакета программы GAMMA [277]. Моделирование проводилось для различных параметров анизотропии и асимметрии, в результате чего была получена двумерная таблица расчетных значений временных профилей. Как экспериментальные, так и расчетные данные были аппроксимированы с помощью функции  $\Delta S/S_{\text{ref}} = (S_{\text{ref}} - S_{\text{rec}})/S_{\text{ref}}$ , где  $S_{\text{rec}}$  и  $S_{\text{ref}}$  интенсивности сигнала в эксперименте с рекаплингом и в контрольном эксперименте (т.е. с применением и без применения импульсов к ядрам  $^1\text{H}$ ). Далее оценивалось отклонение расчетной кривой от экспериментальной кривой по параметру хи-квадрат, т.е. как сумма квадратов отклонения между экспериментальной и расчетной интенсивностями сигнала  $\Delta S/S_{\text{ref}}$ , разделенная на квадрат экспериментальной ошибки. В результате, по рассчитанной двумерной таблице были найдены оптимальные значения параметров анизотропии и асимметрии, для которых значение хи-квадрат минимально. В качестве альтернативного подхода к аппроксимации экспериментальных данных была рассмотрена модель, в которой перевороты ароматических колец учитываются явным образом, т.е. принимался во внимание тот факт, что при перевороте ароматического кольца происходит изменение ориентации вектора  $^1\text{H}$ - $^{13}\text{C}$  в пространстве на  $120^\circ$ . В итоге была составлена одномерная

таблица для результатов моделирования, в которой использовались различные значения анизотропии тензора дипольного взаимодействия.

Оценка доверительного интервала была осуществлена по методу Монте-Карло. Опираясь на экспериментальное значение погрешности (для интенсивности или для значений  $\Delta S/S_{\text{ref}}$ ) и лучший вариант кривой подгонки (для релаксационных или для REDOR данных), был реализован алгоритм внесения шума в данные посредством случайного выбора точек из нормального распределения в окрестности точки, принадлежащей лучшей кривой подгонки. Таким образом была получена тысяча искусственно зашумленных кривых. К этому набору зашумленных данных была применена процедура обработки, описанная выше. Действуя подобным образом, получали значения стандартного отклонения для интересующих параметров.

Расчет скоростей релаксации (см. Рисунок 3.4 (D, E, G)) опирался на строгие аналитические выражения, следующие из теории Рэдфилда, как это описано в разделе 3.5.1. Соответствующая схема вычисления  $R_{1\rho}$  описана в работе [278]. В данных расчетах значение параметра порядка было установлено равным  $S = 0.661$  (если не указано иное значение), что соответствует модели прыжков между двумя равновероятными состояниями, т.е.  $S^2 = (3\cos^2\varphi + 1)/4 = 0.437$ , где предполагается что прыжок сопровождается изменением ориентации в пространстве вектора  $^1\text{H}-^{13}\text{C}$  на  $120^\circ$  [279]. Необходимо отметить, что значение  $S = 0.661 = \sqrt{0.437}$  несколько отличается от значения  $S = 0.625$ , которое используется для масштабирования анизотропии тензора дипольного взаимодействия при анализе данных REDOR. Данный факт следует из теоретического описания, в котором учитывается параметр асимметрии. Во всех расчетах было использовано значение длины связи равное  $1.09 \text{ \AA}$  и значение анизотропии тензора дипольного взаимодействия равное  $23327 \text{ Гц}$  [231]. Тензор анизотропии химического сдвига рассматривался как аксиально-симметричный с  $\Delta\delta = -159 \text{ мд.}$  [280]. Важно отметить, что применимость теории Рэдфилда для анализа медленной динамики в системе является предметом дискуссии [237; 281]. Тем не менее, для рассматриваемого диапазона значений скоростей релаксации теория Рэдфилда остается справедливой [282]. Так, были выявлены некоторые расхождения между теоретическим расчетом по теории Рэдфилда и численным моделированием;



по-видимому, в этих случаях несостоятельность теоретических оценок связана с тем фактом, что спад сигнала является многоэкспоненциальным. Другими словами, при расчете скоростей релаксации следует соблюдать аккуратность [283; 284].

### 3.3.4 МД моделирование

Для получения информации о динамике боковой цепи остатков фенилаланина было проанализировано четыре МД траектории: три траектории в кристаллической форме MPD-ub (2 мкс), cubic-PEG-ub (2 мкс) и rod-PEG-ub (2 мкс), а также одна траектория убиквитина в растворе (8 мкс). В траектории для кристаллической формы MPD-ub моделируется суперячейка, состоящая из 4 элементарных ячеек, построенных на основе кристаллической структуры 3ONS [285]. В общей сложности, ячейка моделирования содержит 24 молекулы убиквитина и 8 772 молекулы воды [286]. Размеры ячейки моделирования были масштабированы на фактор 1.016, с целью воспроизвести расширение кристалла при комнатной температуре. Для кристаллической формы cubic-PEG-ub ячейка моделирования содержит одну элементарную ячейку, содержащую 48 молекул убиквитина, а для кристаллической формы rod-PEG-ub — две элементарные ячейки, содержащие 24 молекулы убиквитина.

В соответствии с общепринятой схемой моделирование кристаллов белка [287] было реализовано посредством использования периодических граничных условий, которые прикладываются к границам ячейки моделирования. Все МД траектории для кристаллических форм белка убиквитина были записаны при помощи пакета программ Amber 16 [288] с использованием силового поля ff14SB [273]. МД траектория белка убиквитина в растворе была записана при помощи пакета программ Amber 11 [289] с использованием силового поля ff99SB [290; 291]. Более детальные подробности, касающиеся протокола МД моделирования, можно найти в ранее опубликованных работах [244; 248]. Анализ траекторий МД опирался на функционал библиотеки ruхmolpp2, написанной на языке python; доступна в репозитории по ссылке <https://github.com/>

bionmr-spbu/pyxmolpp2. В частности, в библиотеке реализован алгоритм для вычисления площади поверхности доступной растворителю.

Расчет химических сдвигов был проведен для кристаллических решеток, сконструированных на основе кристаллографических структур ZONS [285], ZN30 [292] и ZENV [293]. Для вычисления химических сдвигов из кристаллографической решетки были вырезаны фрагменты, соответствующие интересующей молекуле убиквитина и её ближайшим соседям. На выделенных фрагментах запускали алгоритм SHIFTX2 [263] для предсказания химических сдвигов  $^1\text{H}^\epsilon$ ,  $^{13}\text{C}^\epsilon$  в боковой цепи остатков фенилаланина. Расчетные значения химических сдвигов усредняли по парам атомов  $\epsilon_1$ ,  $\epsilon_2$ . Кроме того, были оценены химические сдвиги убиквитина в растворе, для этих целей использовалась высокоточная ЯМР-структура 1D3Z.

Для расчета количества переворотов ароматических колец, была реализована следующая схема. Переворот считался случившимся, если произошел переход между двумя состояниями,  $\chi_2=[60\div 180^\circ]$  и  $\chi_2=[240\div 360^\circ]$  (см. Рисунок 3.5). Конформации соответствующие интервалу между двумя промежутками встречаются достаточно редко и могут быть проигнорированы. Описанным образом был рассчитано количество переворотов для остатков Phe4 и Phe45 для моделирования убиквитина в жидкости и в трех кристаллических формах. МД расчеты показывают, что частота ротамерных прыжков для остатка Phe45 в среднем в 47 раз ниже, чем частота прыжков для остатка Phe4.

### 3.4 Заключение

В описываемом исследовании было показано, что подход, сочетающий введение изотопных меток селективным образом и эксперименты ЯМР в твердом теле с ВМУ, позволяет получить представление о динамике аминокислотных остатков фенилаланина в кристаллах убиквитина. Благодаря хорошему разрешению в спектре ЯМР  $^1\text{H}$  удалось наблюдать два различных химических окружения для остатка Phe4, которые соответствуют двум молекулам убиквитина в кристалле cubic-PEG-ub.

Стэкинг-взаимодействие ароматических колец двух остатков Phe4, принадлежащих соседним молекулам убиквитина в кристалле, приводит к заметному смещению сигнала протонов одной из молекул в сильное поле и, как результат, к появлению двух хорошо различимых пиков для этих двух молекул в кристалле cubic-PEG-ub. В тоже время для кристалла rod-PEG-ub, асимметричная ячейка которого содержит три неэквивалентных молекулы, наблюдается единственный пик, поскольку во всех молекулах боковая цепь остатка Phe4 направлена в сторону растворителя и не вступает во взаимодействия с соседними молекулами. Ни для одной из трех кристаллических форм не удалось обнаружить сигнал от остатка Phe45 из-за уширения линии, связанного с медленными переворотами ароматических колец. Данное предположение было подтверждено при помощи метода МД моделирования.

Для остатка Phe45 характерное время переворотов ароматического кольца можно оценить равным 1 мкс, используя известный из эксперимента временной диапазон 10-20 нс (см. Рисунок 3.4 (E)), характерный для событий переворота ароматического кольца для Phe4, и коэффициент замедления динамики боковой цепи остатка Phe45 по отношению к остатку Phe4, известный из МД моделирования.

Используя известный из эксперимента диапазон 10-20 нс (см. Рисунок 3.4 (E)) для характерного времени переворота ароматического кольца Phe4, и коэффициент замедления динамики боковой цепи остатка Phe45 по отношению к остатку Phe4 около 50, известный из МД моделирования, характерное время переворотов ароматического кольца остатка Phe45 можно оценить в 1 мкс.

Для того чтобы оценить время распада намагниченности была рассчитана скорость поперечной релаксации  $R_{1\rho}$  для протонов (была использована система 4 спинов, см. подробности в разделе 3.5) и скорость поперечной релаксации  $R_{1\rho}$  для углерода, как функция характерного времени для переворота ароматического кольца. Полученные результаты демонстрируют, что время жизни когерентности при действии спин-лока составляет доли миллисекунд (см. Рисунок 3.4 (E, G)). В соответствии с ожиданиями, столь быстрое затухание сигнала будет означать уширение и исчезновение когерентности в hCN эксперименте ещё до момента регистрации сигнала. В одномерном протонном спектре обнаружить отклик

остатка Phe45 также не удается, т.к. сигналы ароматических остатков перекрываются с сигналами амидных групп. Полное дейтерирование всех амидных протонов, т.е. использование 100 % дейтерированного буфера не представляется возможным, поскольку только осаждающий агент MPD вносит около 20 %  $^1\text{H}$  в растворитель.

В одной из работ сигнал от остатка Phe45 всё же был обнаружен в кристалле MPD-ub при помощи спектроскопии ЯМР в твердом теле с ВМУ в условиях более низкой температуры (273 К) [262]. Можно предположить, что перевороты ароматического кольца остатка Phe45 в кристалле MPD-ub происходят намного медленнее, чем в других кристаллах, т.е. на масштабах времен порядка сотни мкс. Эта гипотеза находит подтверждение в данных МД моделирования кристалла MPD-ub, для которого было зафиксировано всего одно событие переворота ароматического кольца Phe45 (см. Рисунок 3.5 (В)). При условии справедливости данного предположения, понижение температуры до 273 К может дополнительным образом замедлить перевороты ароматического кольца Phe45, вплоть до миллисекундных значений [294]. В этих условиях сигнал от остатка Phe45 стал бы наблюдаемым (см. Рисунок 3.4 (Е, G)). В описываемом исследовании самая низкая температура, которую удалось достичь при условии быстрого вращения образца под магическим углом, составила около 15-20 °С, при этом для остатка Phe45 сигнал не был обнаружен (данные не показаны).

Интересно отметить, что значения В-факторов для атомов углерода  $\delta$  и  $\epsilon$ , принадлежащих остатку Phe4 или остатку Phe45, оказываются сопоставимыми для всех кристаллических форм (см. Рисунок 3.10). В целом, неудивительно, что рентгеновская кристаллография не способна зафиксировать расхождения, которые детектируются методом спектроскопии ЯМР. Дело в том, что при криогенных температурах динамика боковых цепей крайне ограничена, в результате чего пропадает различие в скоростях переворота ароматического кольца между остатками Phe4 и Phe45.

Дальнейшая работа была направлена на выявление причин существенного отличия в частоте переворотов ароматического кольца для остатков Phe4 и Phe45. С одной стороны, оба остатка Phe4 и Phe45 расположены на поверхности молекулы убиквитина и решающим

обстоятельством, определяющим различие, может оказаться разная площадь поверхности, доступной растворителю (от англ. solvent accessible surface area, SASA). Фактически параметр SASA соотносится со свободным объемом, доступным для ароматического кольца фенилаланина (точнее в данном случае с объемом, заполненным водой). Другим фактором, который может оказывать воздействие на частоту переворотов ароматического кольца, является расположение боковой цепи остатка фенилаланина относительно основной цепи белка, т.е. ротамерное состояние.

На Рисунке 3.6 представлена структурная информация о ротамерном состоянии и показателе SASA для двух остатков Phe4 и Phe45 в МД траекториях белка убиквитина. По тепловым картам, отображающим распределение по конформациям для боковых цепей, можно заметить, что остатки Phe4 и Phe45 находятся в двух разных ротамерных состояниях, что возможно объясняет отличия в частоте переворотов ароматических колец. Для того чтобы проверить последнее предположение, были рассмотрены конформационные состояния остатков Phe, располагающихся на поверхности белковой глобулы, в МД моделировании нескольких негомологичных белков. В результате было обнаружено, что встречаются остатки фенилаланина, которые принадлежат к тому же ротамерному состоянию, что и Phe4, но при этом не вовлеченные в динамику, связанную с переворотами ароматического кольца. И наоборот, были остатки фенилаланина, которые принадлежат к тому же ротамерному состоянию, что и Phe45, но при этом подвержены частым событиям переворота ароматических колец (см. Рисунок 3.11). Это наблюдение делает несостоятельным предположение о том, что отличия в ротамерном состоянии могут объяснить разную скорость переворота ароматических колец. На Рисунке 3.6 также показаны распределения показателей SASA по данным МД моделирования, а также значения SASA для экспериментальных структур в растворе и кристаллической форме (кроме того, значения SASA для каждой неэквивалентной молекулы в асимметричной ячейке кристалла продемонстрированы на Рисунке 3.12). Для показателей SASA также не выявлено четкой корреляции с частотой переворотов ароматических колец. Обобщая полученные данные, можно сказать, что частота переворотов ароматических колец контролируется комбинацией структурных и динамических факторов, которые охватывают

как состояние самих остатков фенилаланина, так и их непосредственное окружение. Однако набор этих факторов ещё предстоит полностью идентифицировать.

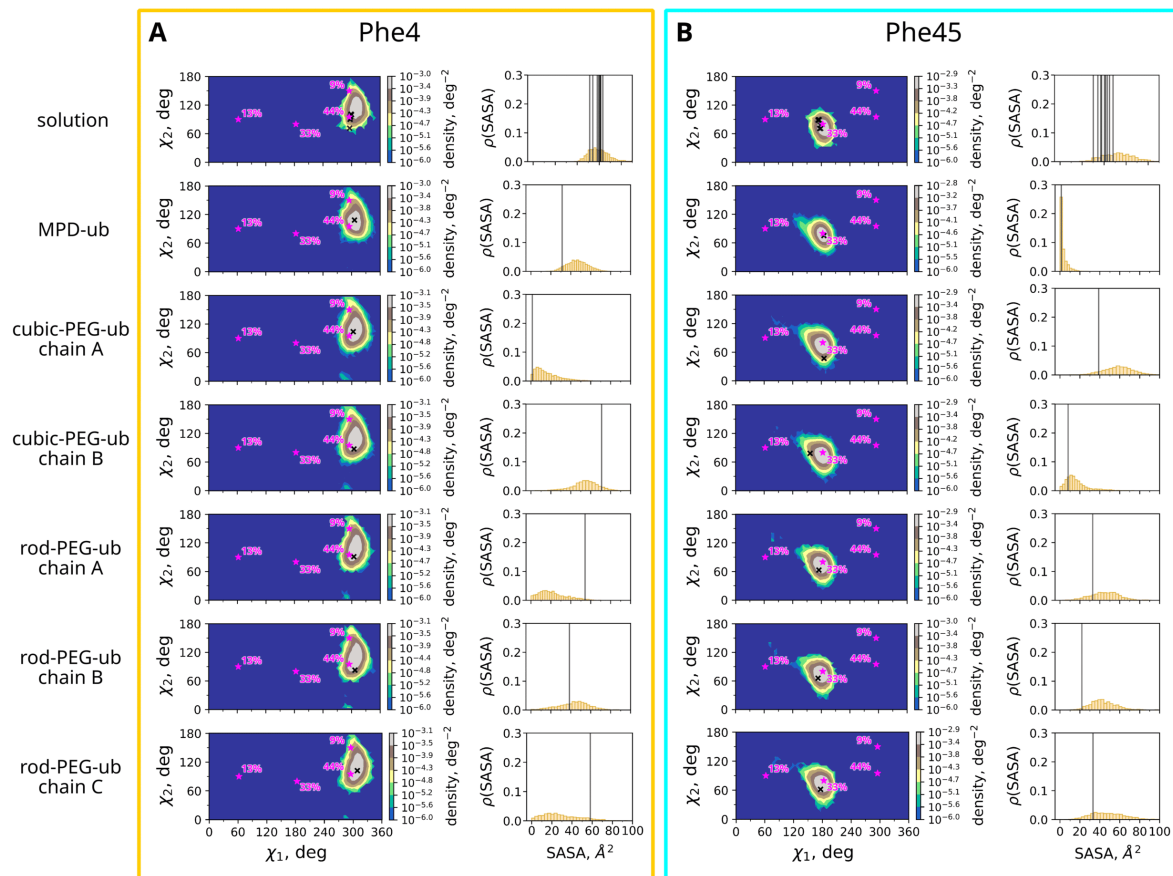


Рисунок 3.6 — Ротамерное состояние и площадь поверхности, доступной растворителю, для остатков (А) Phe4 и (В) Phe45 по структурным данным и данным МД моделирования. Условные обозначения для соответствующих кристаллических форм и образца убиквитина в жидкости указаны в левой части графика. Первая колонка: тепловая карта ( $\chi_1, \chi_2$ ), демонстрирующая вероятность распределения для конформаций боковой цепи остатков Phe4 в МД моделировании. С помощью черных крестиков показаны значения  $\chi_1, \chi_2$ , характерные для структуры убиквитина в растворе (1D3Z, 10 conformers) и для трех кристаллических форм (3ONS, 3N30, 3ENV). В дополнение, пурпурными звездами показаны канонические ротамерные состояния для остатка фенилаланина, взятые из [295]. Диапазон возможных значений для  $\chi_1$  и  $\chi_2$  был выбран таким же  $[0 \div 360^\circ]$ , как и для Рисунка 3.5. Вторая колонка: гистограммы, показывающие распределение показателя SASA для остатка Phe4, полученные по данным МД моделирования. Кроме того, вертикальной черной линией показаны значения SASA для остатка Phe4 в случае структуры убиквитина в растворе и трех кристаллических

формах. В третьей и четвертой колонке проиллюстрированы аналогичные данные для остатка Phe45.

В предыдущей работе было обнаружено качательное движение молекул убиквитина в кристалле cubic-PEG-ub, которое приводит к сильным эффектам в экспериментальных профилях NERRD и, в целом, соответствует более высоким значениям скоростей релаксации  $R_{1\rho}$ . Однако в рамках настоящей работы выяснилось, что качательное движение не оказывает значимого влияния на скорости углеродной релаксации в остатке фенилаланина, а также не приводит к заметными изменениям экспериментальных профилей NERRD (см. Рисунок 3.4 (C)). Вероятно, что динамика, связанная с переворотом ароматических колец, маскирует качательное движение молекул в кристалле.

Обобщая полученные результаты, можно сделать вывод о том, что упаковка молекул в кристаллах убиквитина оказывает несущественное влияние на динамику ароматических колец. В случае остатка Phe4 частота переворотов ароматического кольца практически полностью совпадает для всех кристаллических форм, а для остатка Phe45 переворот ароматического кольца происходит существенно реже, что является причиной отсутствия сигнала от остатка Phe45 в спектрах ЯМР. Интересным представляется исследование влияния упаковки молекул в кристалле на скорость переворотов ароматических колец для остатков, расположенных в гидрофобном ядре белка. В этом случае для того, чтобы переворот кольца стал возможен, должно произойти согласованное изменение конформации окружающих остатков (получившее название «дыхание белка» (от англ. «breathing motion»)). Можно предположить, что кристаллические контакты выступают мягкими ограничениями и не влияют на подобное движение, связанное с «дыханием» белка. Предложенная в описываемом исследовании стратегия, сочетающая селективное изотопное мечение и серию ЯМР экспериментов в твердом теле с ВМУ является хорошим инструментом для изучения подобных вопросов.

### 3.5 Дополнительная информация

#### 3.5.1 Расчет скорости релаксации $R_{1\rho}$

Скорость релаксации  $^1\text{H}$   $R_{1\rho}$  определяется суммой вкладов от гетероядерного дипольного взаимодействия  $^1\text{H}$ - $^{13}\text{C}$  ( $R_{1\rho}^{\text{HC}}$ ), гомоядерного дипольного взаимодействия  $^1\text{H}$ - $^1\text{H}$  ( $R_{1\rho}^{\text{HH}}$ ) и вкладом релаксации за счёт анизотропии химического сдвига (chemical shift anisotropy, CSA)  $^1\text{H}$  CSA ( $R_{1\rho}^{\text{CSA}}$ ) [296]

$$R_{1\rho}^{\text{HC}} = \frac{1}{4}d_{\text{HC}}^2 \left\{ \frac{1}{3} \sin^2(\beta_{\text{eff}}) \left[ J(\omega_e - 2\omega_r) + 2J(\omega_e - \omega_r) + 2J(\omega_e + \omega_r) + J(\omega_e + 2\omega_r) + 9J(\omega_C) \right] + \frac{1}{4}(3 + \cos(2\beta_{\text{eff}})) \left[ 3J(\omega_{\text{H}}) + 6J(\omega_{\text{H}} + \omega_C) + J(\omega_{\text{H}} - \omega_C) \right] \right\} \quad (3.1)$$

$$R_{1\rho}^{\text{HH}} = \frac{3}{4}d_{\text{HH}}^2 \left\{ \frac{1}{8} \sin^2(2\beta_{\text{eff}}) \left[ J(\omega_r - \omega_e) + J(\omega_r + \omega_e) \right] + \frac{1}{2} \sin^4(\beta_{\text{eff}}) \left[ J(\omega_r - 2\omega_e) + J(\omega_r + 2\omega_e) \right] + \frac{1}{16} \sin^2(2\beta_{\text{eff}}) \left[ J(2\omega_r - \omega_e) + J(2\omega_r + \omega_e) \right] + \frac{1}{4} \sin^4(\beta_{\text{eff}}) \left[ J(2\omega_r - 2\omega_e) + J(2\omega_r + 2\omega_e) \right] + \frac{1}{4}(7 - 3 \cos(2\beta_{\text{eff}})) \left[ J(\omega_{\text{H}}) \right] + \frac{1}{2}(5 + 3 \cos(2\beta_{\text{eff}})) \left[ J(2\omega_{\text{H}}) \right] \right\} \quad (3.2)$$

$$R_{1\rho}^{\text{CSA}} = \frac{1}{3}\omega_{\text{H}}^2\Delta\sigma^2 \left\{ \frac{1}{9} \sin^2(\beta_{\text{eff}}) \left[ J(\omega_e - 2\omega_r) + 2J(\omega_e - \omega_r) + 2J(\omega_e + \omega_r) + J(\omega_e + 2\omega_r) \right] + \frac{1}{4}(3 + \cos(2\beta_{\text{eff}})) \left[ J(\omega_{\text{H}}) \right] \right\} \quad (3.3)$$

где  $d_{\text{HC}}$  и  $d_{\text{HH}}$  гетеро- и гомоядерная константы дипольного взаимодействия, определяемые как  $d_{12} = -\frac{\mu_0 \gamma_1 \gamma_2 \hbar}{4\pi r_{12}^3}$ ;  $\mu_0$  магнитная



проницаемость вакуума,  $\gamma_i$  гиромагнитное отношение,  $\hbar$  постоянная Планка,  $r_{12}$  межъядерное расстояние.  $\Delta\delta = \frac{3}{2}(\delta_{zz} - \delta_{iso})$  анизотропия химического сдвига,  $\omega_e$  амплитуда эффективного поля,  $\beta_{eff}$  угол наклона эффективного поля,  $\omega_r$  частота вращения под магическим углом,  $\omega_H$  и  $\omega_C$  ларморовские частоты ядер  $^1H$  и  $^{13}C$ . Спектральная плотность  $J(\omega)$  определена как  $J(\omega) = \frac{2}{5}(1 - S^2) \frac{\tau_c}{1 + (\omega\tau_c)^2}$ , с параметром порядка  $S$  и корреляционным временем  $\tau_c$ .

В случае гомоядерной дипольной релаксация протонов  $H^e$  (см. Рисунок 3.4 (G)) рассматривалась система из четырех спинов, имитирующая окружение остатка Phe45 в дейтерированном селективно-меченом  $^1H, ^{13}C$  образце. Система из четырех спинов содержит представляющий основной интерес протон  $H^e$ , который непосредственно связан с  $C^e$ , другим протоном  $H^e$  и находится вблизи от амидного протона  $H^N$  основной цепи (остаток Thr66). Для межъядерного расстояния использовались следующие значения 1.09 Å для  $C^e-H^e$ , 4.3 Å для  $H^e-H^e$  и 3.0 Å for  $C^e-H^N$ . В дополнение к дипольным взаимодействиям в системе четырех спинов была также учтена анизотропия химического сдвига протона  $H^e$ . Значение анизотропии химического сдвига  $\Delta\delta$  было установлено равным 9 мд [297]. Скорости релаксации были рассчитаны для магнитного поля, соответствующего ларморовской частоте прецессии равной протонов 600 МГц, при частоте ВМУ равной 44.053 кГц, эффективном поле 80 кГц и эффективном угле наклона  $90^\circ$ . Параметр порядка  $S$  был установлен равным 0.661 для гетероядерной релаксации  $^{13}C^e-^1H^e$ , 1.0 для гомоядерной релаксации  $^1H^e-^1H^e$  и 0.5 для гомоядерной релаксации  $^1H^e-^1H^N$  [279].

### 3.5.2 Учет других форм движения, отличных от переворотов ароматических колец

Кроме переворотов ароматических колец вклад в спиновую релаксацию могут также вносить и другие типы движений в системе. Например, флуктуации значений торсионных углов  $\chi_1$  и  $\chi_2$ . Для того чтобы оценить вклад подобных движений, отличных от переворота ароматических колец, был проведен анализ по следующей схеме (см.

Рисунок 3.4 (Е)). Из МД траектории MPD-ub кристалла было выбрано шесть молекул убиквитина, для которых за время моделирования не было зафиксировано ни одного переворота кольца Rhe4. По этим шести молекулам была рассчитана усредненная  $^1\text{H}^\epsilon\text{-}^{13}\text{C}^\epsilon$  дипольная корреляционная функция  $g(\tau)$ , которая показана черной кривой на Рисунке 3.8. Кривая постепенно убывает и выходит на плато со значением 0.84. Похожее поведение наблюдается для корреляционной функции, характеризующей подвижность основной цепи, динамика которой также ограничивается небольшими флуктуациями [298]. В общем случае, спад  $g(\tau)$  имеет многоэкспоненциальный характер [299]. Соответственно аппроксимация корреляционной функции может быть осуществлена про помощи комбинации нескольких экспонент, как это предполагается в расширенной модели Липари-Сабо (от англ. Lipari-Szabo model) [300]. В настоящей работе было использовано четыре экспоненты для подгонки корреляционной функции  $g(\tau)$  в рамках принятых предположений, что четыре типа движения независимы и характеризуются соответствующими значениями параметра порядка  $S_i^2$  и корреляционного времени  $\tau_i$ . Полная корреляционная функция была представлена, как произведение четырех корреляционных функций [301]

$$g_{fit}(\tau) = ((1 - S_1^2)\exp(-\tau/\tau_1) + S_1^2) \times ((1 - S_2^2)\exp(-\tau/\tau_2) + S_2^2) \times \\ \times ((1 - S_3^2)\exp(-\tau/\tau_3) + S_3^2) \times ((1 - S_4^2)\exp(-\tau/\tau_4) + S_4^2) \quad (3.4)$$

где  $g(\tau)$  корреляционная функция, показанная на Рисунке 3.8 (пунктирная зеленая кривая). Характерные времена  $\tau_1 = 0.98$  пс,  $\tau_2 = 32$  пс,  $\tau_3 = 0.59$  нс,  $\tau_4 = 9.2$  нс отличаются на порядок, что подтверждает статистическую независимость соответствующих режимов движения. Характерное время  $\tau_1$  чрезвычайно короткое и, по сути, определяется шагом, с которым записана МД траектория. Очевидно, вклад подобного движения в релаксацию пренебрежительно мал, следовательно выражение для  $g_{fit}(\tau)$  можно упростить

$$g_{red}(\tau) = S_1^2 \times ((1 - S_2^2)\exp(-\tau/\tau_2) + S_2^2) \times ((1 - S_3^2)\exp(-\tau/\tau_3) + S_3^2) \times \\ \times ((1 - S_4^2)\exp(-\tau/\tau_4) + S_4^2) \quad (3.5)$$

Более того, в параметре порядка  $S_1^2$  содержится информация о субмикросекундных колебаниях и либрациях, которая уже учтена в константе дипольного взаимодействия и анизотропии химического сдвига (от англ. chemical shift anisotropy, CSA). Таким образом, выражение для корреляционной функции принимает вид

$$g_{red}(\tau) = ((1 - S_2^2)\exp(-\tau/\tau_2) + S_2^2) \times ((1 - S_3^2)\exp(-\tau/\tau_3) + S_3^2) \times ((1 - S_4^2)\exp(-\tau/\tau_4) + S_4^2) \quad (3.6)$$

На основе полученного выражения для корреляционной функции было сформулировано более общее выражение, в котором учитываются перевороты ароматических колец

$$G(\tau) = ((1 - S_2^2)\exp(-\tau/\tau_2) + S_2^2) \times ((1 - S_3^2)\exp(-\tau/\tau_3) + S_3^2) \times ((1 - S_4^2)\exp(-\tau/\tau_4) + S_4^2) \times ((1 - S_{flip}^2)\exp(-\tau/\tau_{flip}) + S_{flip}^2) \quad (3.7)$$

где  $S_{flip}^2 = 0.437$  [279] параметр порядка для переворотов ароматического кольца, а значения  $S_2^2, \tau_2$ ,  $S_3^2, \tau_3$  и  $S_4^2, \tau_4$  зафиксированы в соответствии с данными молекулярной динамики (см. Рисунок 3.8), а  $\tau_{flip}$  единственная варьируемая величина. Можно заметить, что  $\tau_2 \ll \tau_3 \ll \tau_4, \tau_{flip}$  что позволяет дополнительным образом упростить выражение

$$J(\omega) = \frac{2}{5} \left[ (1 - S_2^2) \frac{\tau_2}{1 + (\omega\tau_2)^2} + S_2^2 (1 - S_3^2) \frac{\tau_3}{1 + (\omega\tau_3)^2} + S_2^2 S_3^2 \xi \right] \quad (3.8)$$

$$\begin{aligned} \xi = & (1 - S_4^2)(1 - S_{flip}^2) \frac{\tau'}{1 + (\omega\tau')^2} + (1 - S_4^2) S_{flip}^2 \frac{\tau_4}{1 + (\omega\tau_4)^2} + \\ & + S_4^2 (1 - S_{flip}^2) \frac{\tau_{flip}}{1 + (\omega\tau_{flip})^2} \end{aligned} \quad (3.9)$$

$$\tau' = \left( \frac{1}{\tau_4} + \frac{1}{\tau_{flip}} \right)^{-1} \quad (3.10)$$

Важно отметить, что сделанное упрощение справедливо только в случае, когда  $\tau_2$ ,  $\tau_3$  и  $(\tau_4, \tau_{flip})$  отличаются, по крайней мере, на порядок величины, см. Рисунок 3.8 и Таблицу 8.

Для интерпретации экспериментальных результатов, представленных на Рисунке 3.9 были использованы формулы (3.8) - (3.10). Сравнивая полученный результат с показанным на Рисунке 3.4 (Е) можно оценить вклад движений, отличный от переворотов ароматических колец и количественно оценить, на сколько изменится время  $\tau_{flip}$ .

В заключение необходимо отметить, что МД моделирование не всегда дает точные параметры, характеризующие локальную динамику. Более того, различные типы движений, определяющие вид корреляционной функции, могут быть частично коррелированы. Кроме того, используемые константные значения, например,  $r_{CH} = 1.09 \text{ \AA}$ , известны не с абсолютной точностью и являются предметом дискуссий, посвященных эффекту усреднения колебаний [302; 303]. Поэтому данные на Рисунке 3.9 следует рассматривать лишь как полуколичественную оценку.

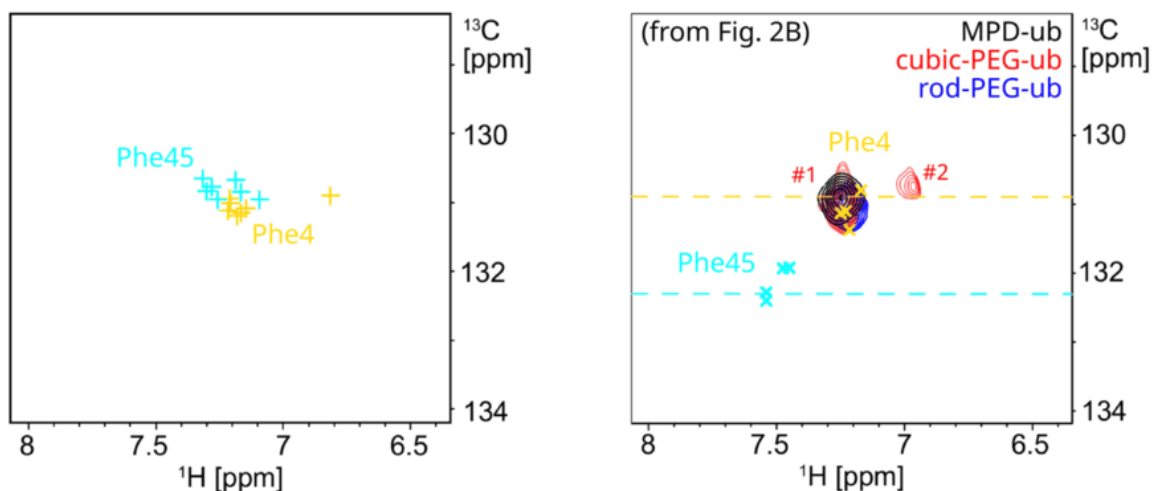


Рисунок 3.7 — Результаты использования программы SHIFTX2 для оценки химического сдвига в системе спинов  $^1\text{H}^\epsilon$ - $^{13}\text{C}^\epsilon$ , принадлежащих остатку Phe4 (желтый цвет) и остатку Phe45 (голубой цвет). Программа SHIFTX2 дает схожие оценки в случае остатка Phe45 для образца убиквитина в растворе и в составе трех кристаллических форм. Более того, SHIFTX2 приводит к корректной оценке для химических сдвигов в случае остатка Phe4, включая сдвиг сигнала для молекулы типа А в кристаллической форме cubic-PEG-ub, который вызван стэкинг-взаимодействием ароматических колец двух остатков Phe4, принадлежащих соседним молекулам в кристалле (см. Рисунок 3.2С). Однако эта программа ошибается с абсолютным значением величин химического сдвига для остатка Phe45 (в сравнении с экспериментом, см. Рисунок 3.2В).

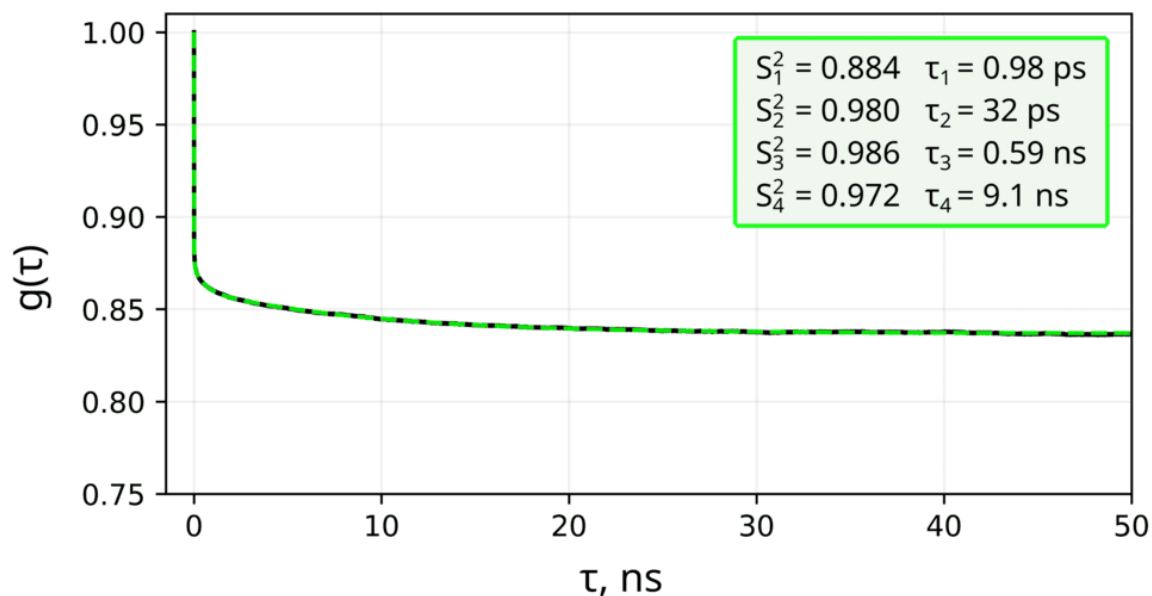


Рисунок 3.8 —  $^1\text{H}^\varepsilon\text{-}^{13}\text{C}^\varepsilon$  dipolar correlation function reflecting phenylalanine motions other than ring flips. The data are from six Phe4 residues that do not experience any flips in our MPD-ub simulation (see Fig. 3.5A). The time correlation function  $P_2(\cos\theta)$  (black curve) has been fitted using the 4-exponential decay function parameterized in the spirit of model-free model, Eq. (3.4) (dashed green curve). The best-fit order parameters and correlation times are listed in the inset. The effect of these non-flip modes on our analyses of spin relaxation rates is illustrated in Fig. 3.9.

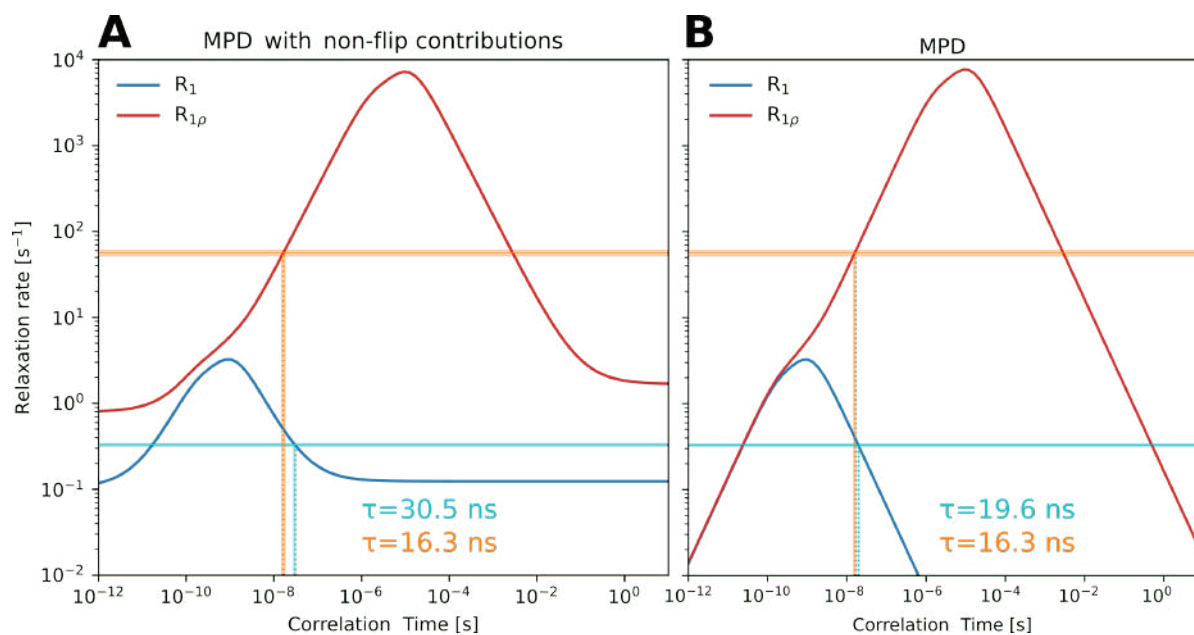


Рисунок 3.9 — Корреляционная функция дипольного взаимодействия  $^1\text{H}^\varepsilon$ - $^{13}\text{C}^\varepsilon$ , отражающая локальную динамику остатков фенилаланина, отличную от переворотов ароматических колец. Для иллюстрации были отобраны шесть остатков Phe4 из МД траектории MPD-ub кристалла, для которых не зафиксировано ни одного события переворота ароматического кольца (см. Рисунок 3.5А). Аппроксимация корреляционной функции  $P_2(\cos\theta)$  (черная кривая) осуществлялась при помощи суммы четырех экспонент (зеленая пунктирная кривая) (по мотивам англ. model-free model), см. формулу (3.4). Найденные оптимальные значения для параметра порядка и времен корреляции показаны на вставке непосредственно на самом рисунке. Эффект подобных движений, не связанных с переворотом ароматических колец, на значения скоростей релаксации продемонстрирован на Рисунке 3.9.

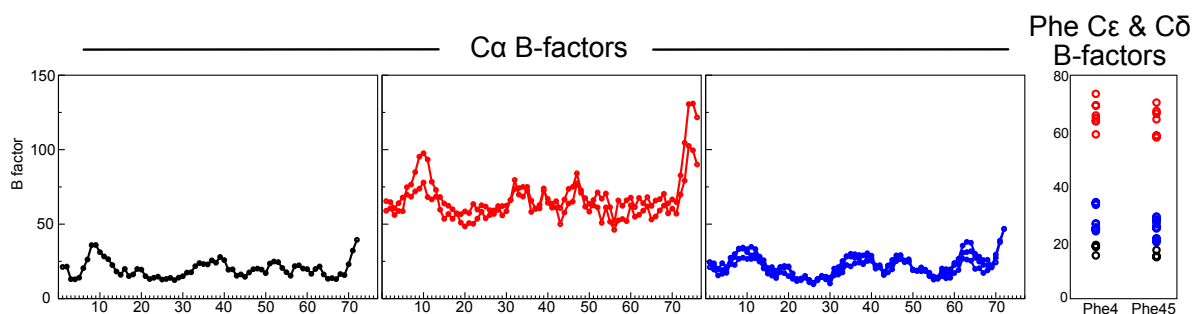


Рисунок 3.10 — B-факторы для трех кристаллических форм MPD-ub (черный цвет, PDB 3ONS), cubic-PEG-ub (красный цвет, PDB 3N30) и rod-PEG-ub (синий цвет, PDB 3ENV). В левой части рисунка показаны B-факторы для атомов C $\alpha$ , а в правой части рисунка B-факторы для атомов C $\delta$  и C $\epsilon$ , принадлежащих ароматическому кольцу остатков фенилаланина. Важно отметить, что большой сдвиг для значений B-факторов в случае кристалла cubic-PEG-ub вероятно связан с качательным движением основной цепи в этой кристаллической форме [248].

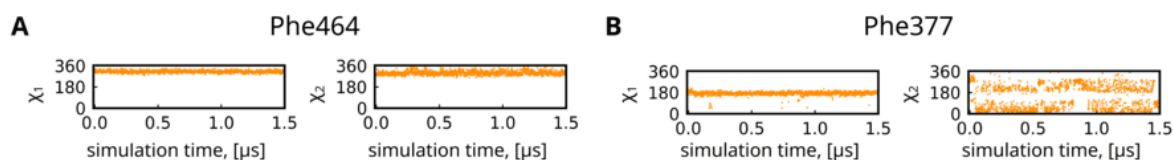


Рисунок 3.11 — Временная зависимость торсионных углов  $\chi_1$  и  $\chi_2$  для боковых цепей остатков, расположенных на поверхности белковой глобулы, (А) Phe464 и (В) Phe377 по данным МД моделирования рецептор-связывающего домена шиповидного белка вируса SARS-CoV-2 в комплексе с мини-белком LCB1 (PDB ID: 7JZU [304]). МД траектория длиной 1.5 мкс была записана в силовом поле Amber ff14SB с использованием модели воды TIP4P-Ew. Ротамерное состояние для остатка Phe464 не отличается от ротамерного состояния для остатка Phe4 в убиквитине, состояние  $m-85^\circ$  согласно работе Ловелл [295]. Остаток Phe464, как и остаток Phe4, в значительной мере экспонирован в раствор, а именно: показатель SASA равен  $72 \text{ \AA}^2$  (значение получено из МД траектории). В то же время в отличие от остатка Phe4 в убиквитине, для остатка Phe464 не было зафиксировано ни одного события переворота ароматического кольца за время моделирования,  $\tau_{div} \geq$  равно 1500 нс. Одновременно с этим, остаток Phe377 находится в том же ротамерном состоянии, что и остаток Phe45 в убиквитине, состояние  $t80^\circ$ . Кроме того, площадь поверхности доступной растворителю для остатка Phe377 сопоставима с соответствующим значением для остатка Phe45 и равна  $57 \text{ \AA}^2$ . Несмотря на имеющиеся сходства, в отличие от остатка Phe45 в убиквитине, остаток Phe377 демонстрирует быструю динамику, связанную с переворотом ароматического кольца,  $\tau_{div} = 10$  нс. Два приведенных примера демонстрируют, что ротамерное состояние боковой цепи остатка фенилаланина не является единственным фактором, определяющим скорость переворота ароматического кольца.



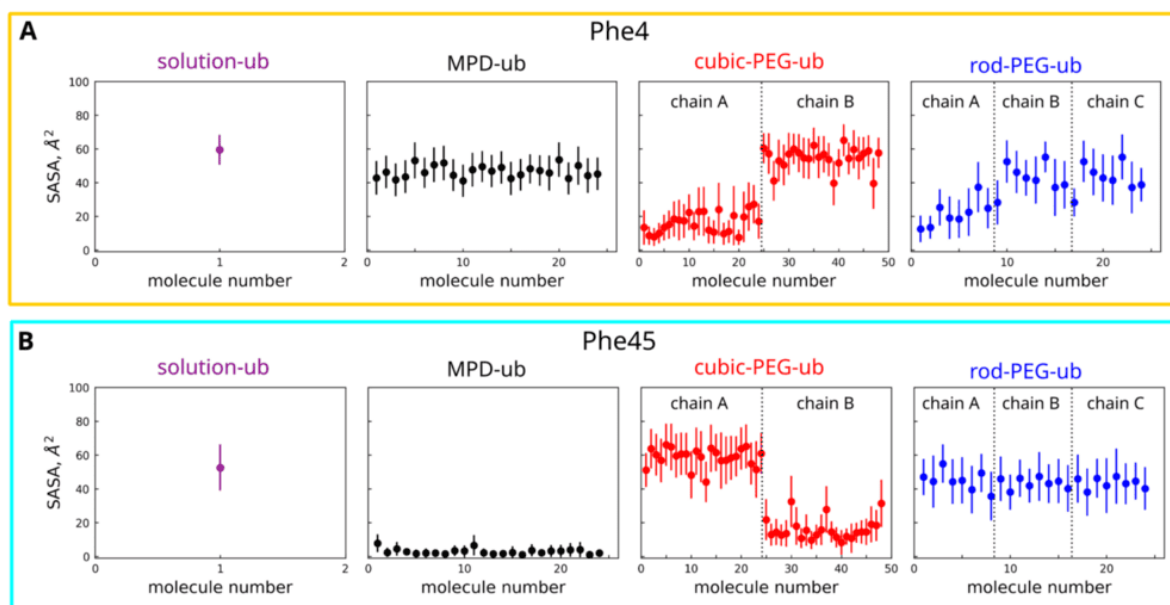


Рисунок 3.12 — Значения для площади поверхности, доступной растворителю (от англ. solvent-accessible surface areas, SASA) для двух остатков фенилаланина, полученные в результате усреднения по данным МД траекторий длиной 2 мкс для трех кристаллических форм и по данным МД траектории длиной 8 мкс для убиквитина в растворе. Можно заметить, что для значений показателя SASA отсутствует явная корреляция со скоростью переворота ароматического кольца для остатков Phe4 и Phe45 (см. Таблицу 8). Цветовая кодировка совпадает с введенной ранее на предыдущих рисунках: черный цвет (MPD-ub), красный цвет (cubic-PEG-ub), синий цвет (rod-PEG-ub) и пурпурный цвет (образец убиквитина в растворе). Ячейка моделирования в случае кристалла MPD-ub содержит 24 эквивалентных молекулы убиквитина типа А. Ячейка моделирования в случае кристалла cubic-PEG-ub содержит 48 молекул убиквитина, распределенных поровну между молекулами типа А (молекулы 1-24) и молекулами типа В (молекулы 25-48). Ячейка моделирования в случае кристалла rod-PEG-ub содержит 24 молекулы убиквитина, распределенных поровну между молекулами типа А (молекулы 1-8), молекулами типа В (молекулы 9-16) и молекулами типа С (молекулы 17-24).

Таблица 8 — Экспериментальные значения скоростей релаксации  $R_1$  и  $R_{1\rho}$  для остатка Phe4 в различных типах кристаллических форм убиквитина и соответствующие расчетные значения характерных времен корреляции  $\tau$ . Значения скоростей релаксации  $R_{1\rho}$  были получены в экспериментах с амплитудой спин-лока равной 30 кГц.

Crystal	$R_1$ [ $s^{-1}$ ]	$\tau_{R_1}$ [ns]	$R_{1\rho}$ [ $s^{-1}$ ]	$\tau_{R_{1\rho}}$ [ns]
<b>MPD-ub</b>	$0.33 \pm 0.01$	$18.7^{+1.4}_{-0.2}$	$56.5 \pm 4.1$	$16.3^{+1.2}_{-2.2}$
<b>cubic-PEG-ub, peak (#1)</b>	$0.47 \pm 0.05$	$13.2^{+2.0}_{-0.9}$	$53.5 \pm 5.1$	$15.2^{+1.1}_{-2.0}$
<b>cubic-PEG-ub, peak (#2)</b>	$0.61 \pm 0.05$	$10.7^{+0.8}_{-0.8}$	$65.8 \pm 6.2$	$18.7^{+1.4}_{-2.5}$
<b>rod-PEG-ub</b>	$0.55 \pm 0.03$	$11.5^{+0.9}_{-0.8}$	$32.5 \pm 0.7$	$10.0^{+0.1}_{-0.2}$

Таблица 9 — Характерные времена переворотов остатков Phe в убиквитине по данным МД моделирования. Время  $\tau_{div}$  рассчитано, как длина траектории, деленная на количество переворотов. Корреляционное время  $\tau_{corr}$  вычислено при помощи выражений, следующих из теории Редфилда. Для этих целей в МД траектории были идентифицированы все временные точки, на которых происходит переворот фенильного кольца и по этим данным была сконструирована псевдотраектория для дипольного вектора (т.е. для вектора  $N^E-C^E$ ), ориентация которого изменялась на  $120^\circ$  при каждом событии переворота. Полученная таким образом траектория отражает динамику остатков Phe, связанную с переворотом ароматических колец, и пренебрегает другими движениями, такими как быстрые флуктуации относительно оси ароматического кольца. Псевдотраектория была использована для расчета корреляционной функции  $P_2(\cos(\theta))$ , аппроксимация которой осуществлялась с помощью  $g(\tau) = S^2 + (1 - S^2)\exp(-\tau/\tau_{corr})$ , где параметр порядка  $S^2 = 0.437$  согласно модели прыжков между двумя эквивалентными состояниями [279].

	Phe4		Phe45*	
	$\tau_{div}$ (ns)	$\tau_{corr}$ (ns)	$\tau_{div}$ (ns)	$\tau_{corr}$ (ns)
solution	98	66	4000	-
MPD-ub	906	762	48000	-
cubic-PEG-ub chain A	215	170	4000	-
cubic-PEG-ub chain B	118	174	>48000	-
rod-PEG-ub chain A	236	217	4000	-
rod-PEG-ub chain B	140	242	>16000	-
rod-PEG-ub chain C	334	245	8000	-

## Заключение

За последние тридцать лет современная биология пришла к осознанию того, что внутренне неупорядоченные белковые последовательности играют огромную роль в жизнедеятельности клетки как в нормальном, так и в патологическом состоянии. При этом функция неупорядоченных белков, как правило, реализуется посредством взаимодействия с белками-партнерами. Основным сценарием при подобном взаимодействии является так называемое «сворачивание при связывании» (англ. *folding upon binding*). Этот механизм предполагает, что неупорядоченный белок приобретает структуру при взаимодействии со своей структурированной мишенью; при этом мишень играет роль своего рода матрицы, на которой происходит фолдинг изначально неупорядоченного белка и которая стабилизирует сформированную структуру. Однако репертуар различных модальностей связывания с участием неупорядоченных белков не ограничивается сворачиванием при связывании. Например, белковый лиганд может частично сохранять свою неупорядоченную природу после связывания с мишенью, образуя так называемый «нечеткий комплекс» (англ. *fuzzy complex*). Более того, в литературе описаны примеры систем, где два неупорядоченных белка, связываясь друг с другом, остаются при этом в неупорядоченном состоянии. В нашей недавней работе предложена общая классификация для различных вариантов связывания с участием неупорядоченных белков [305].

Выступая в роли адвоката дьявола, можно задаться следующими вопросами: в чем состоит значимость неупорядоченного белка, как такового, в случае, когда его функция реализуется через упорядочивание при связывании? Не следует ли рассматривать неупорядоченный белок как случайную полимерную цепочку, детали поведения которой не представляют самостоятельного интереса, сосредоточившись вместо этого на анализе полностью структурированного комплекса? Для ответа на эти вопросы необходимо вспомнить, что неупорядоченные белки обладают определенными структурными предпочтениями, которые наиболее ярко проявляются в форме остаточной структуры (например, отдельные участки в составе неупорядоченной цепочки нередко образуют короткоживущие

$\alpha$ -спирали). Зачастую именно такие элементы остаточной структуры определяют связывание неупорядоченного белка с его структурированной мишенью; при этом реализуется схема связывания «ключ-замок» (англ. lock and key) в противоположность схеме «индуцированного соответствия» (англ. induced fit) [306]. Из этого следует, что присутствие остаточной структуры в неупорядоченном белке непосредственно влияет на параметры связывания ( $k_{on}$ ,  $K_d$ ). В свою очередь, даже небольшие изменения в  $k_{on}$  и  $K_d$  способны оказывать существенное влияние на жизнедеятельность клетки в том случае, когда речь идет о ключевых сигнальных каскадах. Таким образом, именно остаточная структура заслуживает пристального изучения. Иными словами, в первую очередь нас интересует то, что отличает неупорядоченный белок от случайной цепочки.

Очевидно, что при описании неупорядоченных белков классические методы структурного моделирования требуют замены. В качестве естественной альтернативы выступают модели, получаемые с помощью метода молекулярной динамики, т.е. траектории МД. Однако такой подход сопряжен с рядом сложностей. Сами по себе модели МД не отличаются высокой точностью. Известно, что в ходе длительного моделирования белковых молекул имеет место эффект структурного дрейфа; это означает, что силовое поле МД в большинстве случаев не способно успешно воспроизвести глобальный минимум свободной энергии, соответствующий экспериментально определенной белковой структуре [307]. Для неупорядоченных белков ситуация усугубляется тем, что в данном случае отсутствует даже локальный минимум свободной энергии, соответствующий стартовой структуре белка, а это в свою очередь создает условия для беспрепятственного дрейфа. Следует также принять во внимание, что силовые поля МД традиционно разрабатывались для моделирования глобулярных белков. Практика показывает, что использование таких классических силовых полей для моделирования неупорядоченных белков ведет к систематическим ошибкам – например, неупорядоченный белок в процессе моделирования может превратиться в избыточно плотную глобулу [18].

Таким образом, при использовании метода МД для моделирования неупорядоченных белков остро встает проблема совершенствования силовых полей, а решение этой проблемы в свою очередь требует валидации

МД моделей на основе экспериментальных данных. Такого рода работа в последнее время активно ведется рядом исследовательских групп, однако эти усилия в настоящее время носят достаточно разрозненный характер. Объектом валидации становятся различные силовые поля в сочетании с различными моделями воды, а в качестве тестовых систем используются разнообразные белки и пептиды. При этом привлекаемые для анализа наборы экспериментальных данных, как правило, имеют весьма ограниченный характер и варьируют от одного исследования к другому. В то же время для расчета измеряемых параметров нередко применяются упрощенные и даже грубые формулы, что в значительной мере обесценивает результаты сравнительного анализа. Все это привело к тому, что в настоящее время не существует консенсуса относительно оптимального протокола для МД моделирования частично или полностью неупорядоченных белковых систем, а получаемые результаты подчас носят противоречивый характер.

По нашему мнению для достижения прогресса в этой области необходимо сосредоточиться на исследовании нескольких модельных систем. В качестве одной из таких систем может быть выбран пептид N-H4, обсуждаемый в главе 1 настоящей диссертационной работы. Для каждой из исследуемых систем должен быть собран обширный набор экспериментальных данных, который следует использовать для систематического тестирования различных вариантов силовых полей (включая модели воды). Получаемые таким образом результаты можно использовать для идентификации силовых полей (моделей воды), наилучшим образом подходящих для моделирования неупорядоченных белков, а в перспективе и для активного совершенствования имеющихся силовых полей. В контексте такого рода систематических исследований значительный выигрыш может быть получен от объединения усилий нескольких научных коллективов. Здесь уместно отметить, что у автора настоящей работы имеется опыт организации подобного межлабораторного сотрудничества в целях изучения эффекта посттрансляционных модификаций на динамику гистоновых хвостов (исследование проводится в настоящее время и не получило освещения в данной диссертационной работе).

С точки зрения совершенствования силовых полей, существуют определенные возможности для оптимизации параметров в имеющихся полях. Такая работа ведется в ряде исследовательских групп и приносит значимые результаты [308]. Тем не менее рано или поздно на этом пути развития будет достигнут предел и научное сообщество встанет перед необходимостью использовать для моделирования более реалистичные и сложные алгоритмы. В свою очередь это потребует более значительных вычислительных ресурсов. Однако непрерывное совершенствование вычислительной техники позволяет, по крайней мере частично, удовлетворить этот запрос. В качестве одного из примеров такого подхода можно привести поляризуемые силовые поля. Впрочем, следует констатировать, что на сегодняшний день применение поляризуемых силовых полей не привело к качественному улучшению в моделировании биомолекулярных систем.

При этом следует иметь в виду, что приложения молекулярной динамики к конформационно неупорядоченным системам требует использования максимально длительных МД траекторий. Это является необходимым условием для того, чтобы должным образом воспроизвести конформационное многообразие подобных систем. Особенно остро эта проблема встает при использовании данных МД для расчета таких параметров, как скорости парамагнитной релаксации, поскольку эти расчеты предъявляют крайне высокие требования к сходимости МД модели (см. главу 2 настоящей работы). Во многих случаях для решения данной проблемы можно использовать специализированные схемы МД моделирования. Например, при исследовании динамики боковых цепей в глобулярных белках (см. главу 3) значительную пользу могут принести методы метадинамики.

Наконец, нужно отметить, что дальнейшее развитие исследований в этой области предполагает более богатый и полный молекулярный контекст. Например, это относится к моделированию посттрансляционных модификаций в гистоновых хвостах, о которых уже было сказано выше. Здесь же можно упомянуть моделирование молекулярных сборок с увеличивающимся числом компонентов. К примеру, в рамках исследования нуклеосомной частицы нами были получены предварительные результаты моделирования для тетрануклеосомы в кристаллической форме, что

позволяет получить качественные представления о роли гистоновых хвостов в упаковке хроматина. В качестве еще одного примера исследований, проводимых в настоящее время автором данной работы, можно упомянуть МД моделирование взаимодействия пионерского транскрипционного фактора Cbfl с нуклеосомной частицей, где в роли связующего элемента выступает гистоновый хвост H3.

Пожалуй, наиболее ярким примером *in silico* исследований неупорядоченных белков является разработка терапевтических лигандов, способных связываться с неупорядоченными последовательностями. Попытки использования для этой цели малых молекул до сих пор имели лишь ограниченный успех, хотя и на этом пути можно отметить некоторые многообещающие результаты [309]. Можно предположить, что сама природа неупорядоченных белков не предусматривает эффективных механизмов для связывания с малыми молекулами (в отличие от структурированных белков, которые обладают такой способностью). В то же время в природе неупорядоченных белков заложена способность связываться с физиологическими мишенями в виде структурированных белков. Это открывает возможности для разработки белковых лигандов неупорядоченных белков с помощью методов белковой инженерии. Первый пример такой работы был опубликован летом 2024 года коллективом исследователей под руководством Д. Бейкера [310]. Символично, что в день, когда автор пишет эти строки, было объявлено о присуждении Д. Бейкеру Нобелевской премии за успехи в области белковой инженерии. Можно предвидеть, что компьютерные методы, имеющие целью создание белковых лигандов для неупорядоченных белковых мишеней, станут одним из действенных инструментов медицинской технологии завтрашнего дня. Разработка достоверных компьютерных моделей IDP и их валидация на основе экспериментальных данных, чему посвящена настоящая диссертационная работа, являются важным элементом прилагаемых в этом направлении усилий.



## Список сокращений

<b>BBI</b>	широкополосный инверсный датчик
<b>BRET</b>	резонансный перенос энергии биолюминесценции
<b>CSA</b>	анизотропия химического сдвига
<b>FRET</b>	Ферстеровский резонансный перенос энергии
<b>GPU</b>	графический процессор
<b>IDPs</b>	внутренне неупорядоченные белки
<b>IDPR</b>	внутренне неупорядоченные регионы
<b>HMR</b>	схема перераспределения масс для атомов водорода
<b>HSQC</b>	Гетероядерная                      одноквантовая                      корреляционная спектроскопия
<b>MSD</b>	средний квадрат смещения
<b>NCP</b>	нуклеосомная коровая частица
<b>N-H4</b>	пептид SGRGKGGKGLGKGGAKRHRKVL RDN
<b>NOE</b>	ядерный эффект Оверхаузера
<b>NPT</b>	термодинамический ансамбль с постоянным числом частиц, давлением и температурой
<b>NVE</b>	термодинамический ансамбль с постоянным числом частиц, объёмом и энергией
<b>OPC</b>	модель воды с оптимальными точечными зарядами
<b>PDB</b>	база данных белковых структур
<b>PRE</b>	парамагнитное усиление ядерной спиновой релаксации
<b>RMSD</b>	среднеквадратичное отклонение
<b>SASA</b>	площадь поверхности доступная растворителю
<b>SAXS</b>	малоугловое рентгеновское рассеяние
<b>TIP3P</b>	переносимый трехточечный межмолекулярный потенциал
<b>TIP4P-Ew</b>	переносимый трехточечный межмолекулярный потенциал - Эвальд
<b>TIP4P-D</b>	переносимый трехточечный межмолекулярный потенциал - дисперсия
<b>ДНК</b>	дезоксирибонуклеиновая кислота
<b>ВМУ</b>	вращение под магическим углом
<b>ИГМП</b>	импульсный градиент магнитного поля

<b>МД</b>	молекулярная динамика
<b>МУРР</b>	малоугловое рентгеновское рассеяние
<b>ЭПР</b>	электронный парамагнитный резонанс
<b>ЭДТА</b>	этилендиаминтетрауксусная кислота

## Благодарности

В заключение хочется выразить глубокую признательность и благодарность научному руководителю, Николаю Руслановичу Скрынникову, и коллегам по лаборатории биомолекулярного ЯМР, Ивану Сергеевичу Подкорытову и Измайлову Сергею Александровичу за многолетнюю поддержку и возможность работать в научном коллективе мирового уровня. Отдельно хочется поблагодарить Ивана Сергеевича Подкорытова за советы и содействие при работе над диссертацией. Моей научной работе в значительной мере способствовала атмосфера открытой научной дискуссии и товарищеские отношения, сложившиеся в Лаборатории био-молекулярного ЯМР благодаря её бывшим и нынешним сотрудникам: Елене Бушмановой, Полине Вагановой, Александру Коваленко, Светлане Корбан, Александре Лёвкиной, Дмитрию Лузику, Олегу Михайловскому, Михаилу Половинкину, Ольге Рогачевой, Владиславу Саликову, Ирине Тюряевой и Борису Харькову. Я также глубоко признательна руководителям лабораторий, сотрудничество с которыми внесло большой вклад в успех данной работы: Полю Шанде, Кристоферу Джароняку, Дэвиду Кейсу, Алексею Онуфриеву и Алексею Шайтану.

Автор благодарен ресурсным центрам СПбГУ на базе которых проводилась часть вычислений и экспериментов представленных в диссертации. В их число входят Ресурсные центры «Вычислительный центр», «Магнитно-резонансные методы исследования», «Методы анализа состава вещества», «Развитие молекулярных и клеточных технологий», «Рентгенодифракционные методы исследования», «Термогравиметрические и калориметрические методы исследования» и Криогенный отдел научного парка СПбГУ. Автор также выражает глубокую признательность Физическому факультету СПбГУ и в его составе Кафедре ядерно-физических методов исследования.

## Список литературы

1. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis / J. C. Kendrew [et al.] // *Nature*. — 1958. — Mar. — Vol. 181, no. 4610. — P. 662—666.

2. Wüthrich, K. NMR investigations of the dynamics of the aromatic amino acid residues in the basic pancreatic trypsin inhibitor / K. Wüthrich, G. Wagner // *FEBS Letters*. — 1975. — Feb. — Vol. 50, no. 2. — P. 265—268.

3. Bennett, W. S. Glucose-induced conformational change in yeast hexokinase. / W. S. Bennett, T. A. Steitz // *Proceedings of the National Academy of Sciences*. — 1978. — Sept. — Vol. 75, no. 10. — P. 4848—4852.

4. The Conformation of Histone H5: Isolation and Characterisation of the Globular Segment / F. J. AVILES [et al.] // *European Journal of Biochemistry*. — 1978. — Sept. — Vol. 88, no. 2. — P. 363—371.

5. A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome / Z. Peng [et al.] // *Cellular and Molecular Life Sciences*. — 2013. — Aug. — Vol. 71, no. 8. — P. 1477—1504.

6. Musselman, C. A. Characterization of functional disordered regions within chromatin-associated proteins / C. A. Musselman, T. G. Kutateladze // *iScience*. — 2021. — Feb. — Vol. 24, no. 2. — P. 102070.

7. Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded / D. P. Denning [et al.] // *Proceedings of the National Academy of Sciences*. — 2003. — Feb. — Vol. 100, no. 5. — P. 2450—2455.

8. Intrinsic structural disorder in cytoskeletal proteins / M. Guharoy [et al.] // *Cytoskeleton*. — 2013. — June. — Vol. 70, no. 10. — P. 55—57.

9. Uversky, V. N. Intrinsically disordered proteins in overcrowded milieu: membrane-less organelles, phase separation, and intrinsic disorder / V. N. Uversky // *Current Opinion in Structural Biology*. — 2017. — June. — Vol. 44. — P. 18—30.

10. Uversky Vladimir, N. Intrinsic disorder in proteins associated with neurodegenerative diseases / N. Uversky Vladimir // *Frontiers in Bioscience*. — 2009. — June. — Vol. 14, no. 1. — P. 5188.

11. Choy, W.-Y. Calculation of ensembles of structures representing the unfolded state of an SH3 domain / W.-Y. Choy, J. D. Forman-Kay // *Journal of Molecular Biology*. — 2001. — May. — Vol. 308, no. 5. — P. 1011–1032.

12. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from nmr residual dipolar couplings / G. Nodet [et al.] // *Journal of the American Chemical Society*. — 2009. — Nov. — Vol. 131, no. 49. — P. 17908–17918.

13. Xue, Y. Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements,  $^{15}\text{N}$  relaxation, and molecular dynamics simulations: how fast is segmental diffusion in denatured ubiquitin? / Y. Xue, N. R. Skrynnikov // *Journal of the American Chemical Society*. — 2011. — Aug. — Vol. 133, no. 37. — P. 14614–14628.

14. Henriques, J. Molecular dynamics simulations of intrinsically disordered proteins: on the accuracy of the TIP4P-D water model and the representativeness of protein disorder models / J. Henriques, M. Skepö // *Journal of Chemical Theory and Computation*. — 2016. — June. — Vol. 12, no. 7. — P. 3407–3415.

15. Shrestha, U. R. Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations / U. R. Shrestha, J. C. Smith, L. Petridis // *Communications Biology*. — 2021. — Feb. — Vol. 4, no. 1.

16. Best, R. B. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse / R. B. Best, J. Mittal // *The Journal of Physical Chemistry B*. — 2010. — Nov. — Vol. 114, no. 46. — P. 14916–14923.

17. Nerenberg, P. S. Optimizing protein–solvent force fields to reproduce intrinsic conformational preferences of model peptides / P. S. Nerenberg, T. Head-Gordon // *Journal of Chemical Theory and Computation*. — 2011. — Mar. — Vol. 7, no. 4. — P. 1220–1230.

18. Water dispersion interactions strongly influence simulated structural properties of disordered protein states / S. Piana [et al.] // *The Journal of Physical Chemistry B*. — 2015. — Apr. — Vol. 119, no. 16. — P. 5113–5123.

19. Using NMR diffusion data to validate MD models of disordered proteins: Test case of N-terminal tail of histone H4 / O. O. Lebedenko [et al.] // *Biophysical Journal*. — 2024. — Jan. — Vol. 123, no. 1. — P. 80–100.

20. Conformational and interaction landscape of histone H4 tails in nucleosomes probed by paramagnetic NMR spectroscopy / W. Sun [et al.] // *Journal of the American Chemical Society*. — 2023. — Nov. — Vol. 145, no. 46. — P. 25478—25485.

21. Aromatic ring flips in differently packed ubiquitin protein crystals from MAS NMR and MD / D. F. Gauto [et al.] // *Journal of Structural Biology: X*. — 2023. — Dec. — Vol. 7. — P. 1—12.

22. Xue, B. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life / B. Xue, A. K. Dunker, V. N. Uversky // *Journal of Biomolecular Structure and Dynamics*. — 2012. — June. — Vol. 30, no. 2. — P. 137—149.

23. A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome / Z. Peng [et al.] // *Cellular and Molecular Life Sciences*. — 2013. — Aug. — Vol. 71, no. 8. — P. 1477—1504.

24. Wright, P. E. Intrinsically disordered proteins in cellular signalling and regulation / P. E. Wright, H. J. Dyson // *Nature Reviews Molecular Cell Biology*. — 2014. — Dec. — Vol. 16, no. 1. — P. 18—29.

25. Context-dependent resistance to proteolysis of intrinsically disordered proteins / M. J. Suskiewicz [et al.] // *Protein Science*. — 2011. — June. — Vol. 20, no. 8. — P. 1285—1297.

26. Darling, A. L. Intrinsic disorder and posttranslational modifications: the darker side of the biological dark matter / A. L. Darling, V. N. Uversky // *Frontiers in Genetics*. — 2018. — May. — Vol. 9.

27. Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins / R. Halfmann [et al.] // *Molecular Cell*. — 2011. — July. — Vol. 43, no. 1. — P. 72—84.

28. Intrinsic disorder modulates protein self-assembly and aggregation / A. De Simone [et al.] // *Proceedings of the National Academy of Sciences*. — 2012. — Apr. — Vol. 109, no. 18. — P. 6951—6956.

29. Intrinsic disorder in cell-signaling and cancer-associated Proteins / L. M. Iakoucheva [et al.] // *Journal of Molecular Biology*. — 2002. — Oct. — Vol. 323, no. 3. — P. 573—584.

30. Targeting intrinsically disordered proteins involved in cancer / P. Santofimia-Castaño [et al.] // *Cellular and Molecular Life Sciences*. — 2019. — Sept. — Vol. 77, no. 9. — P. 1695—1707.

31. Ji, C. Current status of clinical trials on tau immunotherapies / C. Ji, E. M. Sigurdsson // *Drugs Journal*. — 2021. — June. — Vol. 81, no. 10. — P. 1135–1152.

32. Immunotherapy for Alzheimer's disease: targeting  $\beta$ -amyloid and beyond / C. Song [et al.] // *Translational Neurodegeneration*. — 2022. — Mar. — Vol. 11, no. 1.

33. Dyson, H. Coupling of folding and binding for unstructured proteins / H. Dyson, P. E. Wright // *Current Opinion in Structural Biology*. — 2002. — Feb. — Vol. 12, no. 1. — P. 54–60.

34. Sigalov, A. B. Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form / A. B. Sigalov, A. V. Zhuravleva, V. Y. Orekhov // *Biochimie*. — 2007. — Mar. — Vol. 89, no. 3. — P. 419–421.

35. Fuxreiter, M. Fuzzy Complexes: polymorphism and structural disorder on protein-protein interactions / M. Fuxreiter, P. Tompa // *Biophysical Journal*. — 2009. — Feb. — Vol. 96, no. 3. — 319a.

36. Extreme disorder in an ultrahigh-affinity protein complex / A. Borgia [et al.] // *Nature*. — 2018. — Feb. — Vol. 555, no. 7694. — P. 61–66.

37. Oldfield, C. J. Intrinsically disordered proteins and intrinsically disordered protein regions / C. J. Oldfield, A. K. Dunker // *Annual Review of Biochemistry*. — 2014. — June. — Vol. 83, no. 1. — P. 553–584.

38. Evidence for Residual Structure in Acid- and Heat-denatured Proteins / K. C. Aune [et al.] // *Journal of Biological Chemistry*. — 1967. — Oct. — Vol. 242, no. 19. — P. 4486–4489.

39. Hughson, F. M. Structural characterization of a partly folded apomyoglobin intermediate / F. M. Hughson, P. E. Wright, R. L. Baldwin // *Science*. — 1990. — Sept. — Vol. 249, no. 4976. — P. 1544–1548.

40. Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells / W. Borchers [et al.] // *Nature Chemical Biology*. — 2014. — Nov. — Vol. 10, no. 12. — P. 1000–1002.

41. Da Vela, S. Methods, development and applications of small-angle X-ray scattering to characterize biological macromolecules in solution / S. Da Vela, D. I. Svergun // *Current Research in Structural Biology*. — 2020. — Vol. 2. — P. 164–170.

42. Smaller capillaries improve the small-angle X-ray scattering signal and sample consumption for biomacromolecular solutions / M. A. Schroer [et al.] // *Journal of Synchrotron Radiation*. — 2018. — June. — Vol. 25, no. 4. — P. 1113–1122.

43. CHARMM36m: an improved force field for folded and intrinsically disordered proteins / J. Huang [et al.] // *Nature Methods*. — 2016. — Nov. — Vol. 14, no. 1. — P. 71–73.

44. Robustelli, P. Developing a molecular dynamics force field for both folded and disordered protein states / P. Robustelli, S. Piana, D. E. Shaw // *Proceedings of the National Academy of Sciences*. — 2018. — May. — Vol. 115, no. 21.

45. Shabane, P. S. General purpose water model can improve atomistic simulations of intrinsically disordered proteins / P. S. Shabane, S. Izadi, A. V. Onufriev // *Journal of Chemical Theory and Computation*. — 2019. — Mar. — Vol. 15, no. 4. — P. 2620–2634.

46. Accelerating the conformational sampling of intrinsically disordered proteins / T. N. Do [et al.] // *Journal of Chemical Theory and Computation*. — 2014. — Sept. — Vol. 10, no. 11. — P. 5081–5094.

47. Stanley, N. Kinetic modulation of a disordered protein domain by phosphorylation / N. Stanley, S. Esteban-Martín, G. De Fabritiis // *Nature Communications*. — 2014. — Oct. — Vol. 5, no. 1.

48. A kinetic ensemble of the Alzheimer's A $\beta$  peptide / T. Löhner [et al.] // *Nature Computational Science*. — 2021. — Jan. — Vol. 1, no. 1. — P. 71–78.

49. What drives  $^{15}\text{N}$  spin relaxation in disordered proteins? combined NMR/MD study of the H4 histone tail / K. Kämpf [et al.] // *Biophysical Journal*. — 2018. — Dec. — Vol. 115, no. 12. — P. 2348–2367.

50. Histone H4 Tails in nucleosomes: a fuzzy interaction with DNA / S. O. Rabdano [et al.] // *Angewandte Chemie International Edition*. — 2021. — Feb. — Vol. 60, no. 12. — P. 6480–6487.

51. Salvi, N. Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation / N. Salvi, A. Abyzov, M. Blackledge // *The Journal of Physical Chemistry Letters*. — 2016. — June. — Vol. 7, no. 13. — P. 2483–2489.



52. Dünweg, B. Molecular dynamics simulation of a polymer chain in solution / B. Dünweg, K. Kremer // *The Journal of Chemical Physics*. — 1993. — Nov. — Vol. 99, no. 9. — P. 6983–6997.

53. Self-diffusion coefficient of the hard-sphere fluid: system size dependence and empirical correlations / D. M. Heyes [et al.] // *The Journal of Physical Chemistry B*. — 2007. — Jan. — Vol. 111, no. 6. — P. 1455–1464.

54. Raabe, G. Molecular dynamics simulation of the effect of bond flexibility on the transport properties of water / G. Raabe, R. J. Sadus // *The Journal of Chemical Physics*. — 2012. — Sept. — Vol. 137, no. 10.

55. Aimoli, C. G. Transport properties of carbon dioxide and methane from molecular dynamics simulations / C. G. Aimoli, E. J. Maginn, C. R. A. Abreu // *The Journal of Chemical Physics*. — 2014. — Oct. — Vol. 141, no. 13.

56. System-size corrections for self-diffusion coefficients calculated from molecular dynamics simulations: The case of CO<sub>2</sub>, n-alkanes, and polyethylene glycol) dimethyl ethers / O. A. Moutos [et al.] // *The Journal of Chemical Physics*. — 2016. — Aug. — Vol. 145, no. 7.

57. Ferrario, V. Simulation of protein diffusion: a sensitive probe of protein–solvent interactions / V. Ferrario, J. Pleiss // *Journal of Biomolecular Structure and Dynamics*. — 2018. — Apr. — Vol. 37, no. 6. — P. 1534–1544.

58. Characterization of long linear and branched alkanes and alcohols for temperatures up to 573.15 K by surface light scattering and molecular dynamics simulations / T. Klein [et al.] // *The Journal of Physical Chemistry B*. — 2020. — Apr. — Vol. 124, no. 20. — P. 4146–4163.

59. Yeh, I.-C. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions / I.-C. Yeh, G. Hummer // *The Journal of Physical Chemistry B*. — 2004. — Sept. — Vol. 108, no. 40. — P. 15873–15879.

60. Dünweg, B. Molecular dynamics algorithms and hydrodynamic screening / B. Dünweg // *The Journal of Chemical Physics*. — 1993. — Nov. — Vol. 99, no. 9. — P. 6977–6982.

61. Basconi, J. E. Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations / J. E. Basconi, M. R. Shirts // *Journal of Chemical Theory and Computation*. — 2013. — June. — Vol. 9, no. 7. — P. 2887–2899.

62. Hicks, A. Removing Thermostat Distortions of Protein Dynamics in Constant-Temperature Molecular Dynamics Simulations / A. Hicks, M. MacAinsh, H.-X. Zhou // *Journal of Chemical Theory and Computation*. — 2021. — Aug. — Vol. 17, no. 9. — P. 5920—5932.

63. Bussi, G. Canonical sampling through velocity rescaling / G. Bussi, D. Donadio, M. Parrinello // *The Journal of Chemical Physics*. — 2007. — Jan. — Vol. 126, no. 1.

64. Hansen, J. C. Structure and function of the core histone N-termini: more than meets the eye / J. C. Hansen, C. Tse, A. P. Wolffe // *Biochemistry journal*. — 1998. — Dec. — Vol. 37, no. 51. — P. 17637—17641.

65. Self-diffusion coefficient of bulk and confined water: a critical review of classical molecular simulation studies / I. N. Tsimpanogiannis [et al.] // *Molecular Simulation*. — 2018. — Sept. — Vol. 45, no. 4/5. — P. 425—453.

66. AMBER 2020 / D. Case [et al.]. — University of California, San Francisco, 2020.

67. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB / J. A. Maier [et al.] // *Journal of Chemical Theory and Computation*. — 2015. — July. — Vol. 11, no. 8. — P. 3696—3713.

68. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew / H. W. Horn [et al.] // *The Journal of Chemical Physics*. — 2004. — May. — Vol. 120, no. 20. — P. 9665—9678.

69. Izadi, S. Building water models: a different approach / S. Izadi, R. Anandakrishnan, A. V. Onufriev // *The Journal of Physical Chemistry Letters*. — 2014. — Sept. — Vol. 5, no. 21. — P. 3863—3871.

70. García de la Torre, J. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure / J. García de la Torre, M. L. Huertas, B. Carrasco // *Biophysical Journal*. — 2000. — Feb. — Vol. 78, no. 2. — P. 719—730.

71. An efficient method for estimating the hydrodynamic radius of disordered protein conformations / M. Nygaard [et al.] // *Biophysical Journal*. — 2017. — Aug. — Vol. 113, no. 3. — P. 550—557.

72. Assessment of models for calculating the hydrodynamic radius of intrinsically disordered proteins / F. Pesce [et al.] // *Biophysical Journal*. — 2023. — Jan. — Vol. 122, no. 2. — P. 310—321.

73. Revisiting the use of dioxane as a reference compound for determination of the hydrodynamic radius of proteins by pulsed field gradient NMR spectroscopy / E. E. Tranchant [et al.] // *bioRxiv*. — 2023. — June.

74. Onufriev, A. V. Water models for biomolecular simulations / A. V. Onufriev, S. Izadi // *WIREs Computational Molecular Science*. — 2017. — СЕНТ. — Т. 8, № 2.

75. Comparison of simple potential functions for simulating liquid water / W. L. Jorgensen [et al.] // *The Journal of Chemical Physics*. — 1983. — July. — Vol. 79, no. 2. — P. 926–935.

76. Berendsen, H. J. C. The missing term in effective pair potentials / H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma // *The Journal of Physical Chemistry*. — 1987. — Nov. — Vol. 91, no. 24. — P. 6269–6271.

77. Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations / D. L. Mobley [et al.] // *Journal of Chemical Theory and Computation*. — 2009. — Jan. — Vol. 5, no. 2. — P. 350–358.

78. Best, R. B. Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association / R. B. Best, W. Zheng, J. Mittal // *Journal of Chemical Theory and Computation*. — 2014. — Oct. — Vol. 10, no. 11. — P. 5113–5124.

79. Optimizing solute–water van der waals interactions to reproduce solvation free energies / P. S. Nerenberg [et al.] // *The Journal of Physical Chemistry B*. — 2012. — Apr. — Vol. 116, no. 15. — P. 4524–4534.

80. Abascal, J. L. F. A general purpose model for the condensed phases of water: TIP4P/2005 / J. L. F. Abascal, C. Vega // *The Journal of Chemical Physics*. — 2005. — Dec. — Vol. 123, no. 23.

81. Improved side-chain torsion potentials for the Amber ff99SB protein force field / K. Lindorff-Larsen [et al.] // *Proteins: Structure, Function, and Bioinformatics*. — 2010. — Apr. — Vol. 78, no. 8. — P. 1950–1958.

82. Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment / S. Rauscher [et al.] // *Biophysical Journal*. — 2016. — Feb. — Vol. 110, no. 3. — 358a.

83. Miller, M. S. Osmotic pressure simulations of amino acids and peptides highlight potential routes to protein force field parameterization / M. S. Miller,

W. K. Lay, A. H. Elcock // *The Journal of Physical Chemistry B.* — 2016. — Apr. — Vol. 120, no. 33. — P. 8217–8229.

84. Choice of force field for proteins containing structured and intrinsically disordered regions / V. Zapletal [et al.] // *Biophysical Journal.* — 2020. — Apr. — Vol. 118, no. 7. — P. 1621–1633.

85. Gil Pineda, L. I. Performance of CHARMM36m with modified water model in simulating intrinsically disordered proteins: a case study / L. I. Gil Pineda, L. N. Milko, Y. He // *Biophysics Reports.* — 2020. — June. — Vol. 6, no. 2/3. — P. 80–87.

86. Conformational preferences of an intrinsically disordered protein domain: a case study for modern force fields / S. M. Gopal [et al.] // *The Journal of Physical Chemistry B.* — 2020. — Dec. — Vol. 125, no. 1. — P. 24–35.

87. Abriata, L. A. Assessment of transferable forcefields for protein simulations attests improved description of disordered states and secondary structure propensities, and hints at multi-protein systems as the next challenge for optimization / L. A. Abriata, M. Dal Peraro // *Computational and Structural Biotechnology Journal.* — 2021. — Vol. 19. — P. 2626–2636.

88. Thermodynamics and kinetics of the amyloid- $\beta$  peptide revealed by Markov state models based on MD data in agreement with experiment / A. Paul [et al.] // *Chemical Science.* — 2021. — Vol. 12, no. 19. — P. 6652–6669.

89. Pedersen, K. B. Predicting molecular properties of  $\alpha$ -synuclein using force fields for intrinsically disordered proteins / K. B. Pedersen, J. C. Flores-Canales, B. Schiøtt // *Proteins: Structure, Function, and Bioinformatics.* — 2022. — Aug. — Vol. 91, no. 1. — P. 47–61.

90. Holz, M. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate  $^1\text{H}$  NMR PFG measurements / M. Holz, S. R. Heil, A. Sacco // *Physical Chemistry Chemical Physics.* — 2000. — Vol. 2, no. 20. — P. 4740–4742.

91. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature / H. C. Andersen // *The Journal of Chemical Physics.* — 1980. —  $\Phi\text{EBP}$ . — T. 72, № 4. — C. 2384–2393.

92. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions / W. G. Hoover // *Physical Review A.* — 1985. — Mar. — Vol. 31, no. 3. — P. 1695–1697.

93. Acetylation Increases the  $\alpha$ -Helical Content of the Histone Tails of the Nucleosome / X. Wang [et al.] // *Journal of Biological Chemistry*. — 2000. — Sept. — Vol. 275, no. 45. — P. 35013–35020.

94. Potoyan, D. A. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics / D. A. Potoyan, G. A. Papoian // *Journal of the American Chemical Society*. — 2011. — Apr. — T. 133, № 19. — C. 7405–7415.

95. Yang, D. Structure and binding of the H4 histone tail and the effects of lysine 16 acetylation / D. Yang, G. Arya // *Phys. Chem. Chem. Phys.* — 2011. — Февр. — T. 13, № 7. — C. 2911–2921.

96. The acetylation landscape of the H4 histone tail: disentangling the interplay between the specific and cumulative effects / D. Winogradoff [et al.] // *Journal of the American Chemical Society*. — 2015. — May. — Vol. 137, no. 19. — P. 6245–6253.

97. Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-Terminal basic patch in the nucleosome / B.-R. Zhou [et al.] // *Journal of Molecular Biology*. — 2012. — Aug. — Vol. 421, no. 1. — P. 30–37.

98. Histone H3 and H4 N-Terminal Tails in nucleosome arrays at cellular concentrations probed by magic angle spinning NMR spectroscopy / M. Gao [et al.] // *Journal of the American Chemical Society*. — 2013. — Oct. — Vol. 135, no. 41. — P. 15278–15281.

99. Shabane, P. S. Significant compaction of H4 histone tail upon charge neutralization by acetylation and its mimics, possible effects on chromatin structure / P. S. Shabane, A. V. Onufriev // *Journal of Molecular Biology*. — 2021. — Mar. — Vol. 433, no. 6. — P. 166683.

100. Correlating histone acetylation with nucleosome core particle dynamics and function / T. H. Kim [et al.] // *Proceedings of the National Academy of Sciences*. — 2023. — Apr. — Vol. 120, no. 15.

101. Acetylated histone H4 tail enhances histone H3 tail acetylation by altering their mutual dynamics in the nucleosome / A. Furukawa [et al.] // *Proceedings of the National Academy of Sciences*. — 2020. — Aug. — Vol. 117, no. 33. — P. 19661–19663.

102. Binding of regulatory proteins to nucleosomes is modulated by dynamic histone tails / Y. Peng [et al.] // *Nature Communications*. — 2021. — Sept. — Vol. 12, no. 1.

103. Nanosecond time scale motions in proteins revealed by high-resolution NMR relaxometry / C. Charlier [et al.] // *Journal of the American Chemical Society*. — 2013. — Nov. — Vol. 135, no. 49. — P. 18665—18672.

104. Roberts, G. K. NMR Approaches To Understanding Protein Specificity / G. K. Roberts, L.-Y. Lian // *Biological NMR Spectroscopy*. — Oxford University Press, 02/1997.

105. Determination of an ensemble of structures representing the denatured state of the bovine Acyl-Coenzyme A binding protein / K. Lindorff-Larsen [et al.] // *Journal of the American Chemical Society*. — 2004. — Feb. — Vol. 126, no. 10. — P. 3291—3299.

106. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins / A. H. Mao [et al.] // *Proceedings of the National Academy of Sciences*. — 2010. — Apr. — Vol. 107, no. 18. — P. 8183—8188.

107. Extended experimental inferential structure determination method in determining the structural ensembles of disordered protein states / J. Lincoff [et al.] // *Communications Chemistry*. — 2020. — June. — Vol. 3, no. 1.

108. Ortega, A. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models / A. Ortega, D. Amorós, J. García de la Torre // *Biophysical Journal*. — 2011. — АВГ. — Т. 101, № 4. — С. 892—898.

109. Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques / W.-Y. Choy [et al.] // *Journal of Molecular Biology*. — 2002. — Feb. — Vol. 316, no. 1. — P. 101—112.

110. Bernadó, P. A self-consistent description of the conformational behavior of chemically denatured proteins from nmr and small angle scattering / P. Bernadó, M. Blackledge // *Biophysical Journal*. — 2009. — СЕНТ. — Т. 97, № 10. — С. 2839—2845.

111. Protein dynamics to define and refine disordered protein ensembles / P. M. Naullage [et al.] // *The Journal of Physical Chemistry B*. — 2022. — Feb. — Vol. 126, no. 9. — P. 1885—1894.

112. Amorós, D. Prediction of hydrodynamic and other solution properties of partially disordered proteins with a simple, coarse-grained model / D. Amorós, A. Ortega, J. García de la Torre // *Journal of Chemical Theory and Computation*. — 2013. — Feb. — Vol. 9, no. 3. — P. 1678—1685.

113. García de la Torre, J. Hydrodynamic properties of biomacromolecules and macromolecular complexes: concepts and methods. a tutorial mini-review / J. García de la Torre, J. Hernández Cifre // *Journal of Molecular Biology*. — 2020. — Apr. — Vol. 432, no. 9. — P. 2930—2948.

114. Braga, C. Computer simulation of the role of torsional flexibility on mass and momentum transport for a series of linear alkanes / C. Braga, K. P. Travis // *The Journal of Chemical Physics*. — 2012. — Aug. — Vol. 137, no. 6.

115. Bulacu, M. Effect of bending and torsion rigidity on self-diffusion in polymer melts: A molecular-dynamics study / M. Bulacu, E. van der Giessen // *The Journal of Chemical Physics*. — 2005. — Sept. — Vol. 123, no. 11.

116. The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule / E. Brookes [et al.] // *European Biophysics Journal*. — 2009. — Feb. — Vol. 39, no. 3. — P. 423—435.

117. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables / V. Ozenne [et al.] // *Bioinformatics*. — 2012. — May. — Vol. 28, no. 11. — P. 1463—1470.

118. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties / G. Tesi [et al.] // *Proceedings of the National Academy of Sciences*. — 2021. — Oct. — Vol. 118, no. 44.

119. Kirkwood, J. G. The intrinsic viscosities and diffusion constants of flexible macromolecules in solution / J. G. Kirkwood, J. Riseman // *The Journal of Chemical Physics*. — 1948. — June. — Vol. 16, no. 6. — P. 565—573.

120. Fleming, P. J. Revisiting macromolecular hydration with HullRadSAS / P. J. Fleming, J. J. Correia, K. G. Fleming // *European Biophysics Journal*. — 2023. — Jan. — Vol. 52, no. 4/5. — P. 215—224.

121. AMBER 2023 / D. Case [et al.]. — University of California, San Francisco, 2023.

122. Miyamoto, S. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models / S. Miyamoto, P. A. Kollman // *Journal of Computational Chemistry*. — 1992. — Oct. — Vol. 13, no. 8. — P. 952—962.

123. Jung, J. Optimal temperature and pressure evaluations in molecular dynamics simulations with a large time step / J. Jung, C. Kobayashi, Y. Sugita // *Biophysical Journal*. — 2019. — Feb. — Vol. 116, no. 3. — 141a.

124. Krivov, G. G. Improved prediction of protein side-chain conformations with SCWRL4 / G. G. Krivov, M. V. Shapovalov, R. L. Dunbrack // *Proteins: Structure, Function, and Bioinformatics*. — 2009. — July. — Vol. 77, no. 4. — P. 778—795.

125. Statistical coil model of the unfolded state: Resolving the reconciliation problem / A. K. Jha [et al.] // *Proceedings of the National Academy of Sciences*. — 2005. — Aug. — Vol. 102, no. 37. — P. 13099—13104.

126. Nguyen, H. Improved generalized born solvent model parameters for protein simulations / H. Nguyen, D. R. Roe, C. Simmerling // *Chemical Theory and computation*. — 2013. — Mar. — Vol. 9, no. 4. — P. 2020—2034.

127. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKaPredictions / M. H. M. Olsson [et al.] // *Journal of Chemical Theory and Computation*. — 2011. — Jan. — Vol. 7, no. 2. — P. 525—537.

128. Joung, I. S. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations / I. S. Joung, T. E. Cheatham // *The Journal of Physical Chemistry B*. — 2008. — July. — Vol. 112, no. 30. — P. 9020—9041.

129. Molecular dynamics with coupling to an external bath / H. J. C. Berendsen [et al.] // *The Journal of Chemical Physics*. — 1984. — Oct. — Vol. 81, no. 8. — P. 3684—3690.

130. Ryckaert, J.-P. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes / J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen // *Journal of Computational Physics*. — 1977. — Mar. — Vol. 23, no. 3. — P. 327—341.

131. Effects of molecular dynamics thermostats on descriptions of chemical nonequilibrium / A. J. Page [et al.] // *Journal of Chemical Theory and Computation*. — 2012. — Oct. — Vol. 8, no. 11. — P. 4019—4028.

132. Frenkel, D. Understanding molecular simulation: from algorithms to applications / D. Frenkel, B. Smit, M. A. Ratner // *Physics Today*. — 1997. — July. — Vol. 50, no. 7. — P. 66—66.



133. Vijay-kumar, S. Structure of ubiquitin refined at 1.8 Å resolution / S. Vijay-kumar, C. E. Bugg, W. J. Cook // *Journal of Molecular Biology*. — 1987. — Apr. — Vol. 194, no. 3. — P. 531–544.

134. Long-time-step molecular dynamics through hydrogen mass repartitioning / C. W. Hopkins [et al.] // *Journal of Chemical Theory and Computation*. — 2015. — Mar. — Vol. 11, no. 4. — P. 1864–1874.

135. Bülow, S. von. Systematic errors in diffusion coefficients from long-time molecular dynamics simulations at constant pressure / S. von Bülow, J. T. Bullerjahn, G. Hummer // *The Journal of Chemical Physics*. — 2020. — July. — Vol. 153, no. 2.

136. Haile, J. M. Molecular dynamics simulation / J. M. Haile. — Nashville, TN: John Wiley Sons, 02/1997. — (Wiley professional paperback series).

137. Qian, H. Single particle tracking. analysis of diffusion and flow in two-dimensional systems / H. Qian, M. Sheetz, E. Elson // *Biophysical Journal*. — 1991. — Oct. — Vol. 60, no. 4. — P. 910–921.

138. Michalet, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium / X. Michalet // *Physical Review E*. — 2010. — Oct. — Vol. 82, no. 4.

139. Fliege, J. The distribution of points on the sphere and corresponding cubature formulae / J. Fliege // *IMA Journal of Numerical Analysis*. — 1999. — Apr. — Vol. 19, no. 2. — P. 317–334.

140. Wong, V. Evaluating rotational diffusion from protein md simulations / V. Wong, D. A. Case // *The Journal of Physical Chemistry B*. — 2007. — Dec. — Vol. 112, no. 19. — P. 6013–6024.

141. Numerical recipes in C: the art of scientific computing / William H. Press [et al.]. — 2nd ed. — Cambridge, England: Cambridge University Press, 10/1992.

142. Rotational diffusion anisotropy of human ubiquitin from  $^{15}\text{N}$  NMR relaxation / N. Tjandra [et al.] // *Journal of the American Chemical Society*. — 1995. — Dec. — Vol. 117, no. 50. — P. 12562–12566.

143. Linke, M. Rotational diffusion depends on box size in molecular dynamics simulations / M. Linke, J. Köfinger, G. Hummer // *The Journal of Physical Chemistry Letters*. — 2018. — May. — Vol. 9, no. 11. — P. 2874–2878.

144. SciPy 1.0: fundamental algorithms for scientific computing in Python / P. Virtanen [et al.] // *Nature Methods*. — 2020. — Feb. — Vol. 17, no. 3. — P. 261—272.

145. Palmer, A. G. I. NMR characterization of the dynamics of biomacromolecules / A. G. I. Palmer // *ChemInform journal*. — 2004. — Oct. — Vol. 35, no. 45.

146. Palmer III, A. G. NMR probes of molecular dynamics: overview and comparison with other techniques / A. G. Palmer III // *Annual Review of Biophysics and Biomolecular Structure*. — 2001. — June. — Vol. 30, no. 1. — P. 129—155.

147. Binary mutual diffusion coefficients of aqueous cyclic ethers at 25 °C. tetrahydrofuran, 1,3-dioxolane, 1,4-dioxane, 1,3-dioxane, tetrahydropyran, and trioxane / D. G. Leaist [et al.] // *Journal of Chemical and Engineering Data*. — 2000. — Sept. — Vol. 45, no. 5. — P. 815—818.

148. Clisby, N. High-precision estimate of the hydrodynamic radius for self-avoiding walks / N. Clisby, B. Dünweg // *Physical Review E*. — 2016. — Nov. — Vol. 94, no. 5.

149. Lazar, G. A. De novo design of the hydrophobic core of ubiquitin / G. A. Lazar, J. R. Desjarlais, T. M. Handel // *Protein Science*. — 1997. — June. — Vol. 6, no. 6. — P. 1167—1178.

150. Noncovalent dimerization of ubiquitin / Z. Liu [et al.] // *Angewandte Chemie*. — 2011. — Nov. — Vol. 124, no. 2. — P. 484—487.

151. Price, W. S. Diffusion and its measurement / W. S. Price. — Cambridge: Cambridge University Press, 07/2009. — P. 1—68.

152. Jerschow, A. Suppression of convection artifacts in stimulated-echo diffusion experiments. double-stimulated-echo experiments / A. Jerschow, N. Müller // *Journal of Magnetic Resonance*. — 1997. — Apr. — Vol. 125, no. 2. — P. 372—375.

153. Gradient-tailored water suppression for  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments optimized to retain full sensitivity / V. Sklenar [et al.] // *Journal of Magnetic Resonance, Series A*. — 1993. — Apr. — Vol. 102, no. 2. — P. 241—245.

154. The role of rotational motion in diffusion nmr experiments on supramolecular assemblies: application to Sup35NM fibrils / B. B. Kharkov [et al.] // *Angewandte Chemie*. — 2021. — June. — Vol. 133, no. 28. — P. 15573—15579.

155. Stejskal, E. O. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient / E. O. Stejskal, J. E. Tanner // *The Journal of Chemical Physics*. — 1965. — Jan. — Vol. 42, no. 1. — P. 288–292.

156. Hermann, M. R. Saxs-restrained ensemble simulations of intrinsically disordered proteins with commitment to the principle of maximum entropy / M. R. Hermann, J. S. Hub // *Journal of Chemical Theory and Computation*. — 2019. — Aug. — Vol. 15, no. 9. — P. 5103–5115.

157. // *NMR studies of translational motion*. — Cambridge University Press, 07/2009. — P. 296–307.

158. Boggs, P. T. The computation and use of the asymptotic covariance matrix for measurement error models / P. T. Boggs, J. R. Donaldson. — 1989.

159. Whitlock, J. Folding of DNA by histones which lack their NH<sub>2</sub>-terminal regions. / J. Whitlock, A. Stein // *Journal of Biological Chemistry*. — 1978. — June. — Vol. 253, no. 11. — P. 3857–3861.

160. Cary Peter D. Moss, T. High-resolution proton-magnetic-resonance studies of chromatin core particles / T. Cary Peter D. Moss, E. M. Bradbury // *European Journal of Biochemistry*. — 1978. — Sept. — Vol. 89, no. 2. — P. 475–482.

161. Crystal structure of the nucleosome core particle at 2.8 Å resolution / K. Luger [et al.] // *Nature journal*. — 1997. — Sept. — Vol. 389, no. 6648. — P. 251–260.

162. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution / C. A. Davey [et al.] // *Journal of Molecular Biology*. — 2002. — June. — Vol. 319, no. 5. — P. 1097–1113.

163. Jenuwein, T. Translating the histone code / T. Jenuwein, C. D. Allis // *Science*. — 2001. — Aug. — Vol. 293, no. 5532. — P. 1074–1080.

164. Bannister, A. J. Regulation of chromatin by histone modifications / A. J. Bannister, T. Kouzarides // *Cell Research*. — 2011. — Feb. — Vol. 21, no. 3. — P. 381–395.

165. Strahl, B. D. The language of covalent histone modifications / B. D. Strahl, C. D. Allis // *Nature*. — 2000. — Jan. — Vol. 403, no. 6765. — P. 41–45.

166. The role of the nucleosome acidic patch in modulating higher order chromatin structure / A. A. Kalashnikova [et al.] // *Journal of The Royal Society Interface*. — 2013. — May. — Vol. 10, no. 82. — P. 20121022.

167. Structure and dynamics in the nucleosome revealed by solid-state NMR / X. Shi [et al.] // *Angewandte Chemie International Edition*. — 2018. — July. — Vol. 57, no. 31. — P. 9734–9738.

168. Ghoneim, M. Histone tail conformations: a fuzzy affair with DNA / M. Ghoneim, H. A. Fuchs, C. A. Musselman // *Trends in Biochemical Sciences*. — 2021. — July. — Vol. 46, no. 7. — P. 564–578.

169. Histone H3 and H4 tails play an important role in nucleosome phase separation / E. F. Hammonds [et al.] // *Biophysical Chemistry*. — 2022. — Apr. — Vol. 283. — P. 106767.

170. Modulations of DNA contacts by linker histones and post-translational modifications determine the mobility and modifiability of nucleosomal H3 tails / A. Stützer [et al.] // *Molecular Cell*. — 2016. — Jan. — Vol. 61, no. 2. — P. 247–259.

171. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome / E. A. Morrison [et al.] // *eLife*. — 2018. — Apr. — Vol. 7.

172. Conformational dynamics of histone H3 tails in chromatin / M. Zandian [et al.] // *The Journal of Physical Chemistry Letters*. — 2021. — June. — Vol. 12, no. 26. — P. 6174–6181.

173. Nucleosome composition regulates the histone H3 tail conformational ensemble and accessibility / E. A. Morrison [et al.] // *Nucleic Acids Research*. — 2021. — Apr. — Vol. 49, no. 8. — P. 4750–4767.

174. Dynamic networks observed in the nucleosome core particles couple the histone globular domains with DNA / X. Shi [et al.] // *Communications Biology*. — 2020. — Oct. — Vol. 3, no. 1.

175. Histone H3 core domain in chromatin with different DNA linker lengths studied by  $^1\text{H}$ -Detected solid-state NMR spectroscopy / S. T. Smrt [et al.] // *Frontiers in Molecular Biosciences*. — 2023. — Jan. — Vol. 9.

176. Site-specific studies of nucleosome interactions by solid-state NMR spectroscopy / S. Xiang [et al.] // *Angewandte Chemie International Edition*. — 2018. — Mar. — Vol. 57, no. 17. — P. 4571–4575.

177. The N-terminal tails of histones H2A and H2B adopt two distinct conformations in the nucleosome with contact and reduced contact to DNA / H. Ohtomo [et al.] // *Journal of Molecular Biology*. — 2021. — July. — Vol. 433, no. 15. — P. 167110.

178. Histone H4 lysine 20 mono-methylation directly facilitates chromatin openness and promotes transcription of housekeeping genes / M. Shoaib [et al.] // *Nature Communications*. — 2021. — Aug. — Vol. 12, no. 1.

179. Jennings, C. E. Arginine anchor points govern H3 tail dynamics / C. E. Jennings, C. J. Zoss, E. A. Morrison // *Frontiers in Molecular Biosciences*. — 2023. — May. — Vol. 10.

180. Musselman, C. A. Visualizing conformational ensembles of the nucleosome by NMR / C. A. Musselman, T. G. Kutateladze // *ACS Chemical Biology*. — 2022. — Feb. — Vol. 17, no. 3. — P. 495–502.

181. Battiste, J. L. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited Nuclear Overhauser Effect data / J. L. Battiste, G. Wagner // *Biochemistry journal*. — 2000. — Apr. — Vol. 39, no. 18. — P. 5355–5365.

182. Protein global fold determination using site-directed spin and isotope labeling / V. Gaponenko [et al.] // *Protein Science*. — 2000. — Jan. — Vol. 9, no. 2. — P. 302–309.

183. Iwahara, J. Ensemble approach for NMR structure refinement against  $^1\text{H}$  Paramagnetic Relaxation Enhancement data arising from a flexible paramagnetic group attached to a macromolecule / J. Iwahara, C. D. Schwieters, G. M. Clore // *Journal of the American Chemical Society*. — 2004. — Apr. — Vol. 126, no. 18. — P. 5879–5896.

184. Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR / H. Kato [et al.] // *Proceedings of the National Academy of Sciences*. — 2011. — July. — Vol. 108, no. 30. — P. 12283–12288.

185. Structural insights into the histone H1-nucleosome complex / B.-R. Zhou [et al.] // *Proceedings of the National Academy of Sciences*. — 2013. — Sept. — Vol. 110, no. 48. — P. 19390–19395.

186. A small number of residues can determine if linker histones are bound on or off dyad in the chromatosome / B.-R. Zhou [et al.] // *Journal of Molecular Biology*. — 2016. — Oct. — Vol. 428, no. 20. — P. 3948–3959.

187. Solomon, I. Nuclear magnetic interactions in the HF molecule / I. Solomon, N. Bloembergen // *The Journal of Chemical Physics*. — 1956. — Aug. — Vol. 25, no. 2. — P. 261–266.

188. Rotational diffusion anisotropy of proteins from simultaneous analysis of  $^{15}\text{N}$  and  $^{13}\text{C}\alpha$  nuclear spin relaxation / L. K. Lee [et al.] // *J. Biomol. NMR.* — 1997. — Apr. — Vol. 9, no. 419. — P. 287–298.

189. Iwahara, J. Structure-independent analysis of the breadth of the positional distribution of disordered groups in macromolecules from order parameters for long, variable-length vectors using NMR Paramagnetic Relaxation Enhancement / J. Iwahara, G. M. Clore // *Journal of the American Chemical Society.* — 2010. — Aug. — Vol. 132, no. 38. — P. 13346–13356.

190. Mapping long-range contacts in a highly unfolded protein / M. A. Lietzow [et al.] // *Journal of Molecular Biology.* — 2002. — Sept. — Vol. 322, no. 4. — P. 655–662.

191. Teilum, K. Transient structure formation in unfolded acyl-coenzyme a-binding protein observed by site-directed spin labelling / K. Teilum, B. B. Kragelund, F. M. Poulsen // *Journal of Molecular Biology.* — 2002. — Nov. — Vol. 324, no. 2. — P. 349–357.

192. Characterization of conformational and dynamic properties of natively unfolded human and mouse  $\alpha$ -synuclein ensembles by NMR: implication for aggregation / K.-P. Wu [et al.] // *Journal of Molecular Biology.* — 2008. — May. — Vol. 378, no. 5. — P. 1104–1115.

193. Gillespie, J. R. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. paramagnetic relaxation enhancement by nitroxide spin labels / J. R. Gillespie, D. Shortle // *Journal of Molecular Biology.* — 1997. — Apr. — Vol. 268, no. 1. — P. 158–169.

194. Huang, J.-r. Ensemble Calculations of Unstructured Proteins Constrained by RDC and PRE Data: A Case Study of Urea-Denatured Ubiquitin / J.-r. Huang, S. Grzesiek // *Journal of the American Chemical Society.* — 2009. — Dec. — Vol. 132, no. 2. — P. 694–705.

195. NMR characterization of long-range order in intrinsically disordered proteins / L. Salmon [et al.] // *Journal of the American Chemical Society.* — 2010. — May. — Vol. 132, no. 24. — P. 8407–8418.

196. Dedmon, M. M. Mapping long-range interactions in  $\alpha$ -synuclein using spin-label nmr and ensemble molecular dynamics simulations / M. M. Dedmon, Lindorff-Larsen, others. // *Journal of the American Chemical Society.* — 2004. — Dec. — Vol. 127, no. 2. — P. 476–477.

197. Determination of the free energy landscape of  $\alpha$ -synuclein using spin label nuclear magnetic resonance measurements / J. R. Allison [et al.] // *Journal of the American Chemical Society*. — 2009. — Dec. — Vol. 131, no. 51. — P. 18314—18326.

198. Ganguly, D. Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states / D. Ganguly, J. Chen // *Journal of Molecular Biology*. — 2009. — July. — Vol. 390, no. 3. — P. 467—477.

199. Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure / J. A. Marsh [et al.] // *Journal of Molecular Biology*. — 2007. — Apr. — Vol. 367, no. 5. — P. 1494—1510.

200. Modulation of correlated segment fluctuations in IDPs upon complex formation as an allosteric regulatory mechanism / A. Beier [et al.] // *Journal of Molecular Biology*. — 2018. — Aug. — Vol. 430, no. 16. — P. 2439—2452.

201. Paramagnetic relaxation enhancements in unfolded proteins: Theory and application to drkN SH3 domain / Y. Xue [et al.] // *Protein Science*. — 2009. — June. — Vol. 18, no. 7. — P. 1401—1424.

202. Accelerating molecular modeling applications with graphics processors / J. E. Stone [et al.] // *Journal of Computational Chemistry*. — 2007. — Sept. — Vol. 28, no. 16. — P. 2618—2640.

203. Huang, J. Force field development and simulations of intrinsically disordered proteins / J. Huang, A. D. MacKerell // *Current Opinion in Structural Biology*. — 2018. — Feb. — Vol. 48. — P. 40—48.

204. Recent Force Field Strategies for intrinsically disordered proteins / J. Mu [et al.] // *Journal of Chemical Information and Modeling*. — 2021. — Feb. — Vol. 61, no. 3. — P. 1037—1047.

205. Lowary, P. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning / P. Lowary, J. Widom // *Journal of Molecular Biology*. — 1998. — Feb. — Vol. 276, no. 1. — P. 19—42.

206. Estimation of inter-residue distances in spin labeled proteins at physiological temperatures: experimental strategies and practical limitations / C. Altenbach [et al.] // *Biochemistry journal*. — 2001. — Nov. — Vol. 40, no. 51. — P. 15471—15482.

207. Recent advances in site-directed spin labeling of proteins / W. L. Hubbell [et al.] // *Current Opinion in Structural Biology*. — 1998. — Oct. — Vol. 8, no. 5. — P. 649—656.

208. Izadi, S. Building water models: a different approach / S. Izadi, R. Anandakrishnan, A. V. Onufriev // *The Journal of Physical Chemistry Letters*. — 2014. — Sept. — Vol. 5, no. 21. — P. 3863—3871.

209. Vasudevan, D. Crystal structures of nucleosome core particles containing the ‘601’ strong positioning sequence / D. Vasudevan, E. Y. Chua, C. A. Davey // *Journal of Molecular Biology*. — 2010. — Oct. — Vol. 403, no. 1. — P. 1—10.

210. Reconstitution of nucleosome core particles from recombinant histones and DNA / P. N. Dyer [et al.] // *Methods in Enzymology*. — Elsevier, 2003. — P. 23—44.

211. Luger, K. Preparation of nucleosome core particle from recombinant histones / K. Luger, T. J. Rechsteiner, T. J. Richmond // *Chromatin*. — Elsevier, 1999. — P. 3—19.

212. Structural and dynamic origins of ESR lineshapes in spin-labeled GB1 domain: the insights from spin dynamics simulations based on long MD trajectories / S. A. Izmailov [et al.] // *Scientific Reports*. — 2020. — Jan. — Vol. 10, no. 1.

213. Humphrey, W. VMD: visual molecular dynamics / W. Humphrey, A. Dalke, K. Schulten // *Journal of Molecular Graphics*. — 1996. — Feb. — Vol. 14, no. 1. — P. 33—38.

214. IDPConformerGenerator: a flexible software suite for sampling the conformational space of disordered protein states / J. M. C. Teixeira [et al.] // *The Journal of Physical Chemistry A*. — 2022. — Aug. — Vol. 126, no. 35. — P. 5985—6003.

215. Webb, B. Comparative protein structure modeling using MODELLER / B. Webb, A. Sali // *Current Protocols in Bioinformatics*. — 2016. — June. — Vol. 54, no. 1.

216. Aromatic clusters in protein-protein and protein-drug complexes / E. Lanzarotti [et al.] // *Journal of Cheminformatics*. — 2020. — Vol. 12, no. 1. — P. 30.

217. Wuthrich, K. Internal motion in globular proteins / K. Wuthrich, G. Wagner // *Trends Biochem. Sci.* — 1977. — Vol. 3, no. 10. — P. 227—230.



218. Campbell, I. D. Proton magnetic resonance studies of the tyrosine residues of hen lysozyme-assignment and detection of conformational mobility / I. D. Campbell, C. M. Dobson, R. J. P. Williams // *Proc. R. Soc. B.* — 1975. — Vol. 189, no. 1097. — P. 503–509.

219. Wagner, G. Dynamics of the aromatic amino acid residues in the globular conformation of the basic pancreatic trypsin inhibitor (BPTI). I.  $^1\text{H}$  NMR studies. / G. Wagner, a. DeMarco, K. Wüthrich // *Biophys. Struct. Mech.* — 1976. — Vol. 2, no. 2. — P. 139–158.

220. Protein dynamics by solid-state NMR: aromatic rings of the coat protein in fd bacteriophage / C. Gall [et al.] // *Proc. Natl. Acad. Sci.* — 1982. — Vol. 79, no. 1. — P. 101–105.

221. Hu, F. Mechanisms of proton conduction and gating in influenza M2 proton channels from solid-state NMR / F. Hu, W. Luo, M. Hong // *Science.* — 2010. — Vol. 330, no. 6003. — P. 505–508.

222. Dynamics of aromatic side chains in the active site of FKBP12 / U. Weininger [et al.] // *Biochemistry journal.* — 2017. — Vol. 56, no. 1. — P. 334–343.

223. Slow aromatic ring flips detected despite near-degenerate NMR frequencies of the exchanging nuclei / U. Weininger [et al.] // *Journal of Physical Chemistry.* — 2013. — Vol. 117, no. 31. — P. 9241–9247.

224. Weininger, U. Ring flips revisited:  $^{13}\text{C}$  relaxation dispersion measurements of aromatic side chain dynamics and activation barriers in basic pancreatic trypsin inhibitor / U. Weininger, K. Modig, M. Akke // *Biochemistry journal.* — 2014. — Vol. 53, no. 28. — P. 4519–4525.

225. Raum, H. N. Conformational exchange of aromatic side chains by  $^1\text{H}$  CPMG relaxation dispersion / H. N. Raum, M. Dreydoppel, U. Weininger // *J. Biomol. NMR.* — 2018. — Vol. 72, no. 1. — P. 105–114.

226. Bremi, T. A protocol for the interpretation of side-chain dynamics based on NMR relaxation: application to phenylalanines in antamanide / T. Bremi, R. Brüschweiler, R. R. Ernst // *Journal of the American Chemical Society.* — 1997. — Vol. 119, no. 18. — P. 4272–4284.

227. Transition-state compressibility and activation volume of transient protein conformational fluctuations / M. Dreydoppel [et al.] // *JACS Au.* — 2021. — Vol. 1, no. 6. — P. 833–842.

228. Infrequent cavity-forming fluctuations in HPr from *Staphylococcus carnosus* revealed by pressure- and temperature-dependent tyrosine ring flips / M. Hattori [et al.] // *Protein Sci.* — 2004. — Vol. 13, no. 12. — P. 3104–3114.

229. A sharp thermal transition of fast aromatic-ring dynamics in ubiquitin. / V. Kasinath [et al.] // *Angew. Chem. Int. Ed.* — 2015. — Vol. 54, no. 1. — P. 102–107.

230. Visualizing protein breathing motions associated with aromatic ring flipping / L. Marino P'erez [et al.] // *Nature.* — 2022. — Vol. 602, no. 7898. — P. 695–700.

231. Kasinath, V. A  $^{13}\text{C}$  labeling strategy reveals a range of aromatic side chain motion in calmodulin. / V. Kasinath, K. G. Valentine, A. J. Wand // *Journal of the American Chemical Society.* — 2013. — Vol. 135, no. 26. — P. 9560–9563.

232. Gall, C. Phenylalanine ring dynamics by solid-state deuterium NMR / C. Gall, J. DiVerdi, S. Opella // *Journal of the American Chemical Society.* — 1981. — Vol. 103, no. 17. — P. 5039–5043.

233. Dynamics of hydrophobic core phenylalanine residues probed by solid-state deuterium NMR / L. Vugmeyster [et al.] // *Journal of Physical Chemistry.* — 2015. — Vol. 119, no. 47. — P. 14892–14904.

234. Solvent-driven dynamical crossover in the phenylalanine side-chain from the hydrophobic core of amyloid fibrils detected by  $^2\text{H}$  NMR relaxation / L. Vugmeyster [et al.] // *Journal of Physical Chemistry.* — 2017. — Vol. 121, no. 30. — P. 7267–7275.

235. Bonaccorsi, M. Protein structural dynamics by magic-angle spinning NMR / M. Bonaccorsi, T. Le Marchand, G. Pintacuda // *Curr. Opin. Struct. Biol.* — 2021. — T. 70. — C. 34–43.

236. Lewandowski, J. R. Advances in solid-state relaxation methodology for probing site-specific protein dynamics / J. R. Lewandowski // *Acc. Chem. Res.* — Department of Chemistry, University of Warwick, Coventry CV4 7AL, U.K., 2013. — Vol. 46, no. 9. — P. 2018–2027.

237. Schanda, P. Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: Principles and applications to biomolecules / P. Schanda, M. Ernst // *Prog. Nucl. Magn. Reson. Spectrosc.* — 2016. — T. 96. — C. 1–46.

238. Krushelnitsky, A. Solid-state NMR and protein dynamics / A. Krushelnitsky, D. Reichert // *Prog. Nucl. Magn. Reson. Spectr.* — 2005. — Vol. 47, no. 1/2. — P. 1–25.

239. Probing structure and dynamics of protein assemblies by magic angle spinning NMR spectroscopy / S. Yan [et al.] // *Acc. Chem. Res.* — 2013. — Vol. 46, no. 9. — P. 2047–2058.

240. Reif, B. Deuterated peptides and proteins: structure and dynamics studies by MAS solid-state NMR / B. Reif // *Protein NMR Techniques.* — Springer, 2012. — C. 279–301.

241. Watt, E. D. Recent advances in solid-state nuclear magnetic resonance techniques to quantify biomolecular dynamics / E. D. Watt, C. M. Rienstra // *Anal. Chem.* — 2014. — Vol. 86, no. 1. — P. 58–64.

242. Krushelnitsky, A. Solid-state NMR approaches to internal dynamics of proteins: from picoseconds to microseconds and seconds / A. Krushelnitsky, D. Reichert, K. Saalwachter // *Acc. Chem. Res.* — 2013. — Vol. 46, no. 9. — P. 2028–2036.

243. Aromatic ring dynamics, thermal activation, and transient conformations of a 468 kDa enzyme by specific  $^1\text{H}$ - $^{13}\text{C}$  labeling and fast magic-angle spinning NMR / D. F. Gauto [et al.] // *Journal of the American Chemical Society.* — 2019. — Vol. 141, no. 28. — P. 11183–11195.

244. Slow conformational exchange and overall rocking motion in ubiquitin protein crystals / V. Kurauskas [et al.] // *Nat. Commun.* — 2017. — Vol. 8, no. 1. — P. 1–12.

245. Probing transient conformational states of proteins by solid-state  $R_{1\rho}$  Relaxation-Dispersion NMR Spectroscopy / P. Ma [et al.] // *Angew. Chem. Int. Ed.* — 2014. — Vol. 53, no. 17. — P. 4312–4317.

246. Smith, A. A. Because the light is better here: correlation-time analysis by NMR spectroscopy / A. A. Smith, M. Ernst, B. H. Meier // *Angew. Chem. Int. Ed.* — 2017. — Vol. 129, no. 44. — P. 13778–13783.

247. Zumpfe, K. Model-free or not? / K. Zumpfe, A. A. Smith // *Front. Mol. Biosci.* — 2021. — T. 8. — C. 727553.

248. Observing the overall rocking motion of a protein in a crystal / P. Ma [et al.] // *Nat. Commun.* — 2015. — Vol. 6, no. 1. — P. 1–10.

249.  $^1\text{H}$ -detected biomolecular NMR under fast magic-angle spinning / T. Le Marchand [et al.] // *Chem. Rev.* — 2022. — Vol. 122. — P. 9943–10018.

250. High-resolution proton-detected NMR of proteins at very fast MAS / L. B. Andreas [et al.] // *J. Magn. Reson.* — 2015. — Vol. 253. — P. 36–49.

251. Optimal degree of protonation for  $^1\text{H}$  detection of aliphatic sites in randomly deuterated proteins as a function of the MAS frequency / S. Asami [et al.] // *J. Biomol. NMR.* — 2012. — Vol. 54, no. 2. — P. 155–168.

252. Agarwal, V. Residual methyl protonation in perdeuterated proteins for multi-dimensional correlation experiments in MAS solid-state NMR spectroscopy / V. Agarwal, B. Reif // *J. Magn. Reson.* — 2008. — Vol. 194, no. 1. — P. 16–24.

253. Ruschak, A. M. Methyl groups as probes of supra-molecular structure, dynamics and function / A. M. Ruschak, L. E. Kay // *J. Biomol. NMR.* — 2010. — Vol. 46, no. 1. — P. 75–87.

254. Methyl-specific isotopic labeling: A molecular tool box for solution NMR studies of large proteins / R. Kerfah [et al.] // *Curr. Opin. Struct. Biol.* — 2015. — Vol. 32. — P. 113–122.

255. Protein side-chain dynamics as observed by solution- and solid-state NMR spectroscopy: A similarity revealed / V. Agarwal [et al.] // *Journal of the American Chemical Society.* — 2008. — Vol. 130, no. 49. — P. 16611–16621.

256. Solid-state NMR Measurements of asymmetric dipolar couplings provide insight into protein Side-Chain Motion / P. Schanda [et al.] // *Angew. Chem. Int. Ed.* — 2011. — Vol. 50, no. 46. — P. 11005–11009.

257.  $\alpha$ -Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. / R. J. Lichtenecker [et al.] // *J. Biomol. NMR.* — Institute of Organic Chemistry, University of Vienna, Währingerstr. 38, 1090, Vienna, Austria, roman.lichtenecker@univie.ac.at., 2013. — Vol. 57, no. 4. — P. 327–31.

258. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase / G. Cornilescu [et al.] // *Journal of the American Chemical Society.* — 1998. — Vol. 120, no. 27. — P. 6836–6837.

259. Babu, C. R. Validation of protein structure from preparations of encapsulated proteins dissolved in low viscosity fluids / C. R. Babu, P. F. Flynn, A. J. Wand // *Journal of the American Chemical Society.* — 2001. — Vol. 123, no. 11. — P. 2691–2692.

260. Automated NMR structure determination of stereo-array isotope labeled ubiquitin from minimal sets of spectra using the SAIL-FLYA system / T. Ikeya [et al.] // *J. Biomol. NMR.* — 2009. — Vol. 44, no. 4. — P. 261–272.

261. High-Resolution Protein 3D Structure Determination in Living Eukaryotic Cells / T. Tanaka [et al.] // *Angew. Chem. Int. Ed.* — 2019. — Vol. 131, no. 22. — P. 7362–7366.

262. Structural heterogeneity in microcrystalline ubiquitin studied by solid-state NMR / H. K. Fasshuber [et al.] // *Prot. Sci.* — 2015. — Vol. 24, no. 5. — P. 592–598.

263. SHIFTX2: significantly improved protein chemical shift prediction / B. Han [et al.] // *J. Biomol. NMR.* — 2011. — Vol. 50, no. 1. — P. 43–57.

264. Haigh, C. Ring current theories in nuclear magnetic resonance / C. Haigh, R. Mallion // *Prog. Nucl. Magn. Reson. Spectrosc.* — 1979. — Vol. 13, no. 4. — P. 303–344.

265. Case, D. A. Calibration of ring-current effects in proteins and nucleic acids / D. A. Case // *Journal of Biomolecular NMR.* — 1995. — Vol. 6, no. 4. — P. 341–346.

266. Hess, S. *Tensors for physics* / S. Hess. — Springer, 2015.

267. Gullion, T. Detection of weak heteronuclear dipolar coupling by rotational-echo double-resonance nuclear magnetic resonance / T. Gullion, J. Schaefer // *Advances in Magnetic and Optical Resonance*. T. 13. — Elsevier, 1989. — C. 57–83.

268. Double-quantum homonuclear rotary resonance - efficient dipolar recovery in magic angle spinning nuclear magnetic resonance / N. C. Nielsen [et al.] // *J. Chem. Phys.* — 1994. — Vol. 101, no. 3. — P. 1805–1812.

269. Mechanistic insights into microsecond time-scale motion of solid proteins using complementary  $^{15}\text{N}$  and  $^1\text{H}$  relaxation dispersion techniques / P. Rovó [et al.] // *Journal of the American Chemical Society.* — 2019. — Vol. 141, no. 2. — P. 858–869.

270. Protein conformational dynamics studied by  $^{15}\text{N}$  and  $^1\text{H}$   $R_{1\rho}$  relaxation dispersion: Application to wild-type and G53A ubiquitin crystals / D. F. Gauto [et al.] // *Solid State Nucl. Magn. Reson.* — 2017. — Oct. — Vol. 87. — P. 86–95.

271. VanderHart, D.  $^{13}\text{C}$  NMR rotating frame relaxation in a solid with strongly coupled protons: Polyethylene / D. VanderHart, A. Garroway // *J. Chem. Phys.* — 1979. — Vol. 71, no. 7. — P. 2773—2787.

272. Spin–spin and spin-lattice contributions to the rotating frame relaxation of  $^{13}\text{C}$  in L-alanine / K. Akasaka [et al.] // *J. Chem. Phys.* — 1983. — Vol. 78, no. 6. — P. 3567—3572.

273. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB / J. A. Maier [et al.] // *J. Chem. Theory Comput.* — 2015. — Vol. 11, no. 8. — P. 3696—3713.

274. ssNMRLib: a comprehensive library and tool box for acquisition of solid-state nuclear magnetic resonance experiments on Bruker spectrometers / A. Vallet [et al.] // *Magn. Reson.* — 2020. — Vol. 1, no. 2. — P. 331—345.

275. Schanda, P. Quantitative analysis of protein backbone dynamics in microcrystalline ubiquitin by solid-state NMR Spectroscopy / P. Schanda, B. H. Meier, M. Ernst // *Journal of the American Chemical Society.* — 2010. — Vol. 132, no. 45. — P. 15957—15967.

276. Cross-correlated relaxation of dipolar coupling and chemical-shift anisotropy in magic-angle spinning  $\text{R}_{1\rho}$  NMR measurements: application to protein backbone dynamics measurements / V. Kurauskas [et al.] // *Journal of Physical Chemistry.* — 2016. — Sept. — Vol. 120, no. 34. — P. 8905—8913.

277. Computer simulations in magnetic resonance. An object-oriented programming approach / S. Smith [et al.] // *J. Magn. Reson.* — 1994. — Vol. 106. — P. 75—105.

278. Kurbanov, R. The nuclear magnetic resonance relaxation data analysis in solids: general  $\text{R}_1/\text{R}_{1\rho}$  equations and the model-free approach / R. Kurbanov, T. Zinkevich, A. Krushelnitsky // *J. Chem. Phys.* — 2011. — Vol. 135, no. 18. — P. 184104.

279. Analysis of the backbone dynamics of interleukin-1b using two-dimensional inverse detected heteronuclear  $^{15}\text{N}$ – $^1\text{H}$  NMR spectroscopy / G. M. Clore [et al.] // *Biochemistry.* — 1990. — Vol. 29. — P. 87—91.

280. Carbon-13 chemical shift anisotropies of solid amino acids / C. Ye [et al.] // *Magnetic Resonance in Chemistry.* — 1993. — Aug. — Vol. 31, no. 8. — P. 699—704.

281. Characterization of fibril dynamics on three timescales by solid-state NMR / A. A. Smith [et al.] // *J. Biomol. NMR.* — 2016. — Vol. 65, no. 3. — P. 171–191.

282. Correction to: characterization of fibril dynamics on three timescales by solid-state NMR / A. A. Smith [et al.] // *J. Biomol. NMR.* — 2018. — Vol. 70, no. 3. — P. 203–203.

283. Slow motions in microcrystalline proteins as observed by MAS-dependent  $^{15}\text{N}$  rotating-frame NMR relaxation / A. Krushelnitsky [et al.] // *J. Magn. Reson.* — 2014. — Nov. — Vol. 248. — P. 8–12.

284. Microsecond motions probed by Near-Rotary-Resonance  $R_{1\rho}$   $^{15}\text{N}$  MAS NMR experiments : the model case of protein overall-rocking in crystals / A. Krushelnitsky [et al.] // *J. Biomol. NMR.* — 2018. — Vol. 71, no. 1. — P. 53–67.

285. The structure of human ubiquitin in 2-methyl-2, 4-pentanediol: A new conformational switch / K. Y. Huang [et al.] // *Prot. Sci.* — 2011. — Vol. 20, no. 3. — P. 630–639.

286. Berendsen, H. The missing term in effective pair potentials / H. Berendsen, J. Grigera, T. Straatsma // *J. Phys. Chem.* — 1987. — Vol. 91, no. 24. — P. 6269–6271.

287. Simulations of a protein crystal with a high resolution X-ray structure: evaluation of force fields and water models / D. S. Cerutti [et al.] // *Journal of Physical Chemistry.* — 2010. — Vol. 114, no. 40. — P. 12811–12824.

288. AMBER 16, University of California, San Francisco / D. Case [et al.]. — University of California, San Francisco, 2015.

289. AMBER 11, University of California, San Francisco / D. Case [et al.]. — University of California, San Francisco, 2010.

290. Comparison of multiple Amber force fields and development of improved protein backbone parameters / V. Hornak [et al.] // *Proteins.* — 2006. — Vol. 65, no. 3. — P. 712–725.

291. Improved side-chain torsion potentials for the Amber ff99SB protein force field / K. Lindorff-Larsen [et al.] // *Proteins.* — 2010. — Vol. 78, no. 8. — P. 1950–1958.

292. Crystallographic analysis of metal-ion binding to human Ubiquitin / F. Arnesano [et al.] // *Chemistry European Journal.* — 2011. — Vol. 17, no. 5. — P. 1569–1578.

293. Structural probing of Zn(ii), Cd(ii) and Hg(ii) binding to human ubiquitin / G. Falini [et al.] // *Chemical Communications*. — 2008. — Vol. 45. — P. 5960—5962.

294. Dreydoppel, M. Slow ring flips in aromatic cluster of GB1 studied by aromatic  $^{13}\text{C}$  relaxation dispersion methods / M. Dreydoppel, H. N. Raum, U. Weininger // *J. Biomol. NMR*. — 2020. — Vol. 74, no. 2. — P. 183—191.

295. The penultimate rotamer library / S. C. Lovell [et al.] // *Proteins*. — 2000. — Vol. 40, no. 3. — P. 389—408.

296. Rovó, P. Recent advances in solid-state relaxation dispersion techniques / P. Rovó // *Solid State Nucl. Magn. Reson.* — 2020. — T. 108. — C. 101665.

297. Zhang, R. Proton chemical shift tensors determined by 3D ultrafast MAS double-quantum NMR spectroscopy / R. Zhang, K. H. Mroue, A. Rammamoorthy // *J. Chem. Phys.* — 2015. — Vol. 143, no. 144201. — P. 1—6.

298. Chen, J. Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation / J. Chen, C. L. Brooks, P. E. Wright // *J. Biomol. NMR*. — 2004. — Vol. 29, no. 3. — P. 243—257.

299. Halle, B. Model-free analysis of stretched relaxation dispersions / B. Halle, H. Jóhannesson // *J. Magn. Reson.* — 1998. — Vol. 135, no. 1. — P. 1—13.

300. Deviations from the simple two-parameter model-free approach to the interpretation of  $^{15}\text{N}$  nuclear magnetic relaxation of proteins / M. G. Clore [et al.] // *Journal of the American Chemical Society*. — 1990. — Vol. 112, no. 12. — P. 4989—4991.

301. Skrynnikov, N. R. Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond time-scale side-chain motions / N. R. Skrynnikov, O. Millet, L. E. Kay // *Journal of the American Chemical Society*. — 2002. — Vol. 124, no. 22. — P. 6449—6460.

302. Kowalewski, J. Dipole–dipole coupling constant for a directly bonded CH pair—A carbon-13 relaxation study / J. Kowalewski, M. Effemey, J. Jokisaari // *J. Magn. Reson.* — 2002. — Vol. 157, no. 2. — P. 171—177.

303. NMR Determination of amide N–H equilibrium bond length from concerted dipolar coupling measurements / L. Yao [et al.] // *Journal of the American Chemical Society*. — 2008. — Vol. 130, no. 49. — P. 518—520.



304. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors / L. Cao [et al.] // *Science*. — 2020. — Vol. 370, no. 6515. — P. 426–431.

305. Lebedenko, O. O. Order/Disorder Transitions Upon Protein Binding: A Unifying Perspective / O. O. Lebedenko, A. Sekhar, N. R. Skrynnikov // *Proteins: Structure, Function, and Bioinformatics*. — 2024. — Aug.

306. Koshland, D. E. The Key–Lock Theory and the Induced Fit Theory / D. E. Koshland // *Angewandte Chemie International Edition in English*. — 1995. — Jan. — Vol. 33, no. 23/24. — P. 2375–2378.

307. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations / A. Raval [et al.] // *Journal of proteins structure, function and bioinformatics*. — 2012. — May. — Vol. 80, no. 8. — P. 2071–2079.

308. Nerenberg, P. S. New developments in force fields for biomolecular simulations / P. S. Nerenberg, T. Head-Gordon // *Current Opinion in Structural Biology*. — 2018. — App. — T. 49. — C. 129–138.

309. Zhu, J. Small molecules targeting the disordered transactivation domain of the androgen receptor induce the formation of collapsed helical states / J. Zhu, X. Salvatella, P. Robustelli // *Nature Communications*. — 2022. — Oct. — Vol. 13, no. 1.

310. Diffusing protein binders to intrinsically disordered proteins / C. Liu [et al.] // *bioRxiv*. — 2024. — July.