

Московский физико-технический институт (научно-исследовательский
университет)
Физтех-школа физики и исследований им. Ландау
ОП Вычислительная физика конденсированного состояния и живых систем
Объединенный институт высоких температур РАН

Диссертация допущена к защите
зав. кафедрой

_____ Норман Г.Э.

«_____» _____ 2021 г.

Выпускная квалификационная работа на соискание степени БАКАЛАВРА

**Тема: Молекулярно-Динамическая модель
кристалла Лизоцима**

Направление: 03.03.01 – Прикладные математика и физика
Магистерская программа: –

Выполнил студент гр. 786 _____ Поляченко Ю. А.

Научный руководитель,
д. ф.-м. н., в. н. с. _____ Стегайлов В.В.

Консультант по ,
к. ф.-м. н., доц. _____ Кондратюк Н.Д.

Москва – 2021

Оглавление

Введение	4
Глава 1. Теория и эксперимент по измерению объемного модуля упругости K	6
1.1. Поверхностная адсорбция	7
1.2. Модель Flory-Huggins	11
1.2.1. Хим. потенциал пара в контакте с жидкостью	11
1.2.2. Вклад упругости в μ	12
1.2.3. Вклад смешивания в μ	13
1.3. Эксперимент	18
Глава 2. Построение МД модели кристалла Лизоцима	21
2.1. Подготовка PDB-белка к расчету	21
2.2. Сходимость уровня гидротации	26
2.3. 2 метода расчета модуля упругости K	29
2.3.1. Флюктуационный метод	29
2.3.2. Метод малых деформаций	30
Глава 3. Результаты моделирования	33
3.1. Верификация прямых методов расчета K	33
3.2. Используемые баростат и термостат	34
3.3. Сходимость K по параметрам термостата и баростата	39
3.4. Расчет K прямыми методами	42
3.5. Анализ мобильности воды	45
3.5.1. Максимальное смещение	46
3.5.2. Коэффициент диффузии	47
3.5.3. Корреляция смещения молекулы со временем движения	51

3.5.4.	Энтропия $r^2(t)$ как мера равномерности	53
3.5.5.	Предположение о природе мобильности	59
3.6.	2-фазное моделирование для учета влажности	62
3.6.1.	Ограничения малостью системы	62
3.6.2.	Расчет относительной влажности	65
3.6.3.	Расчет K из баланса хим. потенциала	69
Заключение		72
Список литературы		73

Введение

Актуальность работы. Всестороннее изучение белков может быть полезно в широком круге задач, стоящих сейчас перед человечеством, т.к. белки - тип биомолекул, выполняющих основную часть функций в клетке. Лучшее понимание устройства белков может способствовать, например, разработке новых методов исправления их неправильной работы, что значило бы излечение многих серьезных болезней. На сегодняшний день существуют методы определения последовательности белков, работающие по большей части в автоматическом режиме и не требующие больших затрат ресурсов. Однако для глубокого понимания устройства белка необходимо знать не только его аминокислотную последовательность, но и пространственную структуру. Метод кристаллографии, считающийся сейчас классическим для определения 3D-структуры белков, уже менее тривиален в реализации. Одна из главных сложностей в нем - необходимость кристаллизации белка для его исследования. Кристаллизация полипептидов часто сопряжена с созданием необычных и при этом строго контролируемых физических условий, что уже говорит о сложности процесса. Иногда же в силу особенностей конкретной молекулы ее кристаллизация вообще не представляется возможной. В таких случаях может быть проведена замена частей белка, мешающих процессу. Это опять же многократно усложняет процесс, т.к. нужно предпринимать попытки кристаллизации многих подобных белков. В связи с вышеперечисленными трудностями методы молекулярного моделирования активно используются для поиска мутаций, способствующих кристаллизации белка. Однако возможная роль молекулярного моделирования не исчерпывается этими вопросами.

Цель работы состояла в том, чтобы довести молекулярно-динамическую модель кристаллического Лизоцима до приемлимого воспроизведения экспериментальных измерений его модуля упругости. Имея такую модель,

можно будет интерпретировать результаты экспериментов с молекулярно-атомистической точки зрения. Такая модель также заметно упростит исследование кристалла лизоцима в различных условиях, т.к. не будет необходимости каждый раз проводить экспериментальные измерения. Наконец, подход создания МД модели, разработанный для кристаллического лизоцима, может быть модифицирован для создания МД моделей более сложных белков, таких как Гемоглобин. Это может помочь найти для таких белков мутации, способствующие кристаллизации, т.к. опять же не будет необходимости экспериментально проверять множество различных малых мутаций белка в различных физических условиях.

Работа состоит из трёх глав и заключения.

В главе 1 излагается теория, на основе которой далее описывается эксперимент по измерению объемного модуля упругости K кристаллического лизоцима, результаты которого в дальнейшем используются как референсные значения.

В главе 2 рассказывается об этапах построения МД модели и о методах расчета K из молекулярно-динамических траекторий.

В главе 3 приведена теория, связанная с атомистическим моделированием, и с ее применением обсуждаются результаты расчета K .

В заключении кратко обсуждаются полученные результаты и указываются возможные направления дальнейшей работы.

Глава 1

Теория и эксперимент по измерению объемного модуля упругости K

В этой главе описывается эксперимент по измерению объемного модуля упругости кристаллического лизоцима. В дальнейшем эти экспериментальные точки будут использованы в качестве референсных значений для проверки корректности создаваемой МД модели.

В разделе 1.1 излагается теория Brunauer-Emmett-Teller (BET) [1], описывающая процесс поверхностной адсорбции.

В разделе 1.2 описана сорбция воды в полимерах с точки зрения теории Flory-Huggins [2, 3] и ее модификаций. Сорбция рассматривается как смешивание воды и молекул белка на 3-мерной сетке.

В разделе 1.3 описан ЯМР-эксперимент и его обработка, результаты которой далее используются как референсные значения модуля упругости белкового кристалла.

Гидратация играет важную роль в свертывании белков, их динамике и функциях [4–7]. Например ферментативная активность лизоцима заметно возрастает при уровнях гидратации выше $h = 0.2$ (в граммах воды на грамм сухого белка) [4, 8]. Взаимодействие белков с водой также было одной из центральных тем при изучении свертывания с того момента как концепция гидрофобного взаимодействия была введена Каузмана в 1959 [4].

Для лизоцима уровень $h = 0.2$ соответствует давлению пара $P/P_0 \sim 0.7$ [9], где P_0 – давление насыщенного пара при данной температуре. При дальнейшем насыщении происходит дополнительное поглощение воды в сравнении с линейным ростом при $P/P_0 < 0.7$. Опубликованы работы, связывающие именно это изменение в поведении лизоцима с его ключевыми функциями

[5, 10]. На данный момент не существует консенсуса относительно механизма этого дополнительного поглощения [2, 5].

Гравиметрический метод, широко применяемый для исследования белков в физиологическом диапазоне температур, может быть неудобен для работы при температурах ниже комнатной и приближающихся к $0\text{ }^{\circ}\text{C}$. Поэтому в рассматриваемой экспериментальной работе уровень гидратации измерялся с помощью ядерного магнитного резонанса водородов 1H в системе [11, 12].

1.1. Поверхностная адсорбция

Прежде чем приводить и как-либо интерпретировать экспериментальные измерения уровней гидратации лизоцима, нужно ознакомиться с самим процессом и тем как его описывают в литературе. Относительно простой моделью, описывающей адсорбцию воды на белковой поверхности при данной влажности, является Brunauer-Emmett-Teller (BET) теория [1]. Водяной пар рассматривается как идеальный газ, что оправдано при наших температурах ¹.

Считая пар идеальным газом, можно сказать, что в установившемся со-

¹ Плотность насыщенного водяного пара ρ_{sat} при интересующих нас температурах составляет $\sim 10\text{g}/\text{m}^3$. Можно рассмотреть 2 критерия идеальности водяного пара:

1. Отношение размера молекулы воды d_{H_2O} к характерной длине свободного пробега $\lambda \sim 1/n\sigma$.

$$\frac{\lambda}{d_{H_2O}} \sim \frac{\rho_{liq}}{\rho_{gas}} \sim \frac{10^3}{10^{-2}} \sim 10^5 \quad (1.1)$$

2. Отношение кулоновской диполь-дипольной энергии U_{coul} при данной плотности к кинетической энергии $k_B T$. $U_{coul} \sim k_C p^2 / r^3$, где k_C - кулоновская константа для системы СИ, p - дипольный момент молекулы воды, r - характерное расстояние между молекулами воды. Оценочно $1/r^3 \sim n = \rho_{gas} N_A / \mu$, где N_A - число Авогадро, μ - молярная масса воды. $p \sim 6 \cdot 10^{-30}$ Кл·м, откуда

$$\frac{k_B T}{U_{coul}} \sim \frac{k_B T}{k_C p^2 n} \sim \frac{\mu R T}{(p N_A)^2 \rho_{gas} k_C} \sim 10^5 \quad (1.2)$$

Как видно, оба критерия дают хорошую идеальность пара при нашей плотности.

стоянии не происходит изменения поверхностного адсорбированного слоя воды. Идея качественно отражена на рисунке

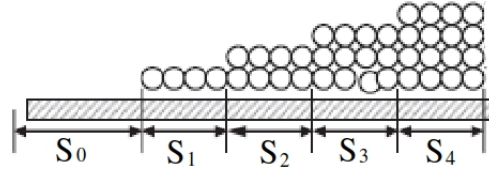


Рис. 1.1. BET-модель поверхностной адсорбции. В статическом состоянии величины S_0 , S_1 , ... не должны изменяться.

Делается предположение, что адсорбция происходит индивидуально на точках адсорбции, т.е. «столбы» молекул, адсорбированных на соседних точках связывания, не влияют друг на друга. Уравнение баланса для сохранения S_0 будет выглядеть как

$$a_1(T)PS_0 = b_1(T)S_1e^{-E_1/RT}, \quad (1.3)$$

где P – давление адсорбируемого пара, S_0 и S_1 – площади поверхностей, покрытых 0-ем и 1-им слоями адсорбированной воды соответственно, E_1 – молярная теплота адсорбции 1-ого слоя, $a_1(T)$ и $b_1(T)$ – остальные константы и не-экспоненциальная зависимость от температуры, R – термодинамическая константа, T – температура. Поясним почему имеет смысл такая запись. Слева стоит скорость уменьшения S_0 -поверхности за счет адсорбции на ней 1-ого слоя воды (что переводит S_0 -поверхность в S_1 -поверхность), т.е. $v_{0 \rightarrow 1}$. Эта скорость пропорциональна давлению пара т.к. плотность пара пропорциональна давлению. Справа стоит скорость испарения с 1-ого слоя, т.е. скорость $v_{1 \rightarrow 0}$ перехода S_1 в S_0 . Есть подавляющий фактор $e^{-E_1/RT}$, т.к. молекулам воды нужно затратить некоторую энергию чтобы вырваться из адсорбированного слоя.

Для всех последующих слоев уравнения неизменности S_k строятся из аналогичных соображений. Скорость адсорбции на k -м слое, переводящей S_k

в $S_{k+1} : v_{k \rightarrow k+1} = PS_k a_{k+1}(T)$. Скорость испарения с k -го слоя $v_{k \rightarrow k-1} = b_k(T)S_k e^{-E_k/RT}$. Адсорбция на $(k-1)$ -м слое $v_{k-1 \rightarrow k} = PS_k a_k(T)$. И наконец испарение с $(k-1)$ -го слоя $v_{k+1 \rightarrow k} = b_{k+1}(T)S_{k+1} e^{-E_{k+1}/RT}$. В итоге

$$\begin{aligned} a_{k+1}(T)PS_k + b_k(T)S_k e^{-E_k/RT} = \\ = b_{k+1}(T)S_{k+1} e^{-E_{k+1}/RT} + a_k(T)PS_{k-1}, \end{aligned} \quad (1.4)$$

$$k = 1 \dots N-1.$$

Подставляя (1.3) в (1.4) при $k = 1$, получаем

$$a_2(T)PS_1 = b_2(T)S_2 e^{-E_2/RT}, \quad (1.5)$$

что аналогично (1.3). Далее по индукции получаем окончательную систему

$$S_k = S_{k-1}P \frac{a_k(T)}{b_k(T)} e^{E_k/RT} = S_{k-1}Px_k(T), \quad k = 1 \dots N. \quad (1.6)$$

Тогда объем адсорбированной воды

$$V = H \sum_{k=0}^N kS_k = S_0 \sum_{k=1}^N kP^k \prod_{j=0}^k x_j(T), \quad (1.7)$$

где H – высота 1 слоя.

Мы нигде явно не использовали, что общая поверхность не меняется. Это можно исправить, нормировав на нее:

$$\frac{V}{V_{mono}} = \frac{\sum_{k=1}^N kP^k \prod_{j=0}^k x_j(T)}{1 + \sum_{k=1}^N P^k \prod_{j=0}^k x_j(T)}, \quad (1.8)$$

где V_{mono} – объем 1 полностью заполненного слоя.

Слои начиная с $k = 2$ находятся в приблизительно одинаковых условиях (если считать, что эффект от подложки быстро спадает), поэтому естественно считать $x_k(T) = x(T)$, $E_k = E_L \quad \forall k > 1$.

Так же, можно пренебречь внешними ограничениями на количество слоев (такие как например размер пор белка), т.е. считать $N \gg 1$. Тогда

$$\frac{V}{V_{mono}} = \frac{cS_0 \sum_{k=1}^{\infty} k(xP)^k}{S_0(1 + c \sum_{k=1}^{\infty} (xP)^k)} = \frac{cxP}{(1 - xP)(1 + xP(c - 1))}, \quad (1.9)$$

где $xP < 1$, $c = x_1/x \sim e^{(E_1 - E_L)/RT}$. Нахождение в пристеночном слое как правило более выгодно [1], поэтому обычно $c > 1$.

Из физических соображений ясно, что при $P \rightarrow P_0(T)$, где $P_0(T)$ – давление насыщенных паров при данной T , должно выполняться $V/V_{mono} \rightarrow \infty$. Если $c > 1$, то достиг этого можно только положив $x(T) = 1/P_0(T)$. Наконец, распространена ситуация когда $E_1 - E_L \gtrsim RT$, что приводит к $c \gg 1$ и позволяет получить совсем простую приближенную версию зависимости

$$\frac{V}{V_{mono}} = \frac{1}{1 - P/P_0}. \quad (1.10)$$

К сожалению, данная зависимость не описывает адсорбцию на поверхности белков. Одним из главных проблем (1.10) и (1.9) является незивисимость (или слабая зависимость в случае (1.9)) от температуры при построении зависимости от относительной влажности. Это не выполняется в лизоциме [9], где количество адсорбированной кристаллом белка воды зависит от относительной влажности по разному при температурах порядка физиологических и при $T \lesssim 10C^\circ$. Плохая применимость (1.10) к белкам объяснима, т.к. при ее получении были сделаны предположения, плохо работающие в белках. Например – пренебрежение внешними ограничениями количества возможных слоев адсорбции. В лизоциме есть пора размером порядка $\sim 1 \text{ \AA}$, что сравнимо с размерами молекул воды, а поэтому переход от (1.8) к (1.9) не корректен в лизоциме. Поэтому для описания экспериментальных точек был использован другой подход.

1.2. Модель Flory-Huggins

Идея данного подхода состоит в том, чтобы рассмотреть сорбцию воды при высоких влажностях как процесс растворения полимера в воде. В природе смесь будет существовать в равновесии с внешней средой, и это равновесие описывается равенством химических потенциалов молекул воды в паре и в «белковом растворе»:

$$\frac{\partial G_{vap}}{\partial \nu_w} = \mu_{vap} = \mu_{Wprot} = \frac{\partial G_{Wprot}}{\partial \nu_w}, \quad (1.11)$$

где G_{vap} – энергия гиббса пара, G_{Wprot} – энергия гиббса системы белок+вода, ν_w – количество молей воды. G_{Wprot} состоит из 2 главных вкладов [3] – упругая энергия разбухания белка G_{el} и энтропийный вклад G_{mix} .

1.2.1. Хим. потенциал пара в контакте с жидкостью

Как было показано выше в (1.1), (1.2), пар при наших условиях можно считать идеальным газом. Поэтому $dG = -SdT + VdP$ при $T = const$ и с учетом аддитивности S и V превращается в

$$G_{vap}(\nu, P, T) = \nu \left(\mu_A(T, P_A) + RT \int_{P_A}^P \frac{dP}{P} \right), \quad (1.12)$$

где P_A – произвольно выбранное референсное давление, $\mu_A(T, P_A)$ – хим. потенциал при соответствующих параметрах.

Далее

$$\mu_{vap} = \frac{\partial G_{vap}}{\partial \nu} = \mu_A(T, P_A) + RT \ln \left(\frac{P}{P_A} \right). \quad (1.13)$$

Физический смысл хим. потенциала – изменение свободной энергии системы при постоянных (P, T) . Система всегда стремится минимизировать

свободную энергию, поэтому если $\mu < 0$, то количество частиц с таким μ будет стремиться увеличиваться, а если $\mu > 0$, то наоборот уменьшаться. Мы знаем, что в случае контакта с жидкой водой испарение будет происходить до тех пор, пока $P < P_0(T)$, где $P_0(T)$ – давление насыщенного пара при данной T . Если же $P > P_0(T)$, то начнется конденсация, т.е. самопроизвольное уменьшение количества молекул пара (за счет их перехода в жидкость). Таким образом $\mu_{var}(P < P_0) < 0$, $\mu_{var}(P > P_0) > 0$, откуда

$$\mu_A(T, P_0) + RT \ln \left(\frac{P_0}{P_A} \right) = \mu_{var}(P = P_0(T)) = 0. \quad (1.14)$$

Подставляя это обратно в (1.13) получаем

$$\mu_{var}(T, P) = RT \ln \left(\frac{P}{P_0(T)} \right). \quad (1.15)$$

1.2.2. Вклад упругости в μ

Изменени энергия гиббса системы вода+белок для (1.11) складывается из энтропийного вклада смешивания [13] и упругой энергии [3].

Разберем сначала упругий вклад. При сорбции воды белок разбухает. Оценить увеличение размеров можно как

$$\frac{1}{v_2} = \frac{V}{V_2} = 1 + \frac{V_1}{V_2} = \left(\frac{l + \Delta l}{l} \right)^3, \quad (1.16)$$

где v_2 – объемная доля атомов белка в «растворе», V_1, V_2 – объемы воды и белка соответственно, $l, \Delta l$ – характерный размер системы и его изменение.

Тогда, используя, что плотность упругой энергии $w = K\epsilon_l^2/2$, в 1 порядке по деформации получим

$$G_{el} \approx V \frac{K(\Delta l/l)^2}{2} = (V_1 + V_2) \frac{K}{2} \left(\left(1 + \frac{V_1}{V_2} \right)^{1/3} - 1 \right)^2, \quad (1.17)$$

где K – объемный модуль упругости белка.

Тогда вклад в хим. потенциал воды

$$\mu_{el} = \frac{\partial G_{el}}{\partial \nu_1} = \frac{\partial V_1}{\partial \nu} \frac{\partial G_{el}}{\partial V_1} = \frac{\partial V_1}{\partial \nu} \frac{K}{2} (v_2^{-1/3} - 1) \left(\frac{5}{3} v_2^{-1/3} - 1 \right) \quad (1.18)$$

1.2.3. Вклад смешивания в μ

Смешение воды с белком имеет как энергетический так и энтропийный вклады, что выражается как

$$\mu_{mix} = \frac{\partial G_{mix}}{\partial \nu_w} = \frac{\partial H_{mix}}{\partial \nu_w} - T \frac{\partial S_{mix}}{\partial \nu_w}. \quad (1.19)$$

Оценим энтропийный вклад. Рассмотрение происходит в рамках модели Flory–Huggins. Идея описания процесса смешивания представлена на рисунке:

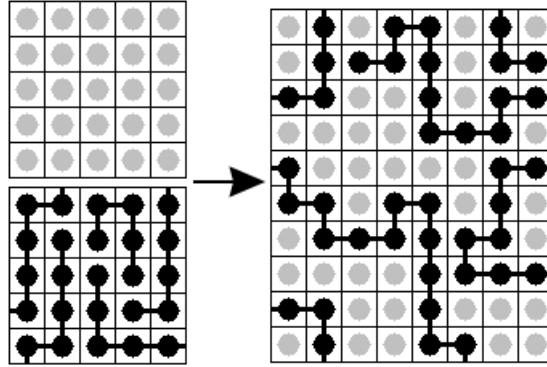


Рис. 1.2. Описание смешивания воды с полимером в рамках модели Flory–Huggins.

Основные приближения данной модели:

1. Смешение происходит на «жесткой» сетке.
2. Звенья полимера неразличимы.

В такой модели можно попытаться оценить количество возможных состояний W и воспользоваться $S = k_B \ln(W)$. Возьмем сетку из n_0 ячеек, заполненных водой, и будем добавлять туда молекулы полимера. Пусть размер

полимера – x ячеек (обозначения взяты из [13]), γ – количество ближайших соседей клетки ($\gamma = 2 \cdot D$, где D – размерность системы). Тогда если в системе уже есть N полимеров, то начальный участок новой молекулы можно расположить $n_0 - xN$ способами. 2-е звено полимера можно положить в еще не занятые клетки. Если на свободной сетке у клетки γ соседей, но уже заполнено xN из n_0 ячеек, то в среднем количество клеток, доступных для постановки 2-го звена полимера, будет $\gamma(1 - xN/n_0)$. Для 3-го и последующих звеньев количество возможных мест расположения будет $\alpha(N) = (\gamma - 1)(1 - xN/n_0)$. Тогда есть всего $\gamma/(\gamma - 1)\alpha^{x-1}(N)$ вариантов расположить полимер не сдвигая начальную точку, которую в свою очередь можно расположить $n_0 - xN$ способами. В итоге, имея в системе N полимеров, можно добавить следующий $(N + 1)$ -й

$$\nu_{N+1} = \frac{n_0 - xN}{2} \frac{\gamma}{\gamma - 1} \alpha(N)^{x-1} \quad (1.20)$$

способами. Множитель $1/2$ отражает, что начало и конец полимера не различимы, но без $1/2$ это вырождение учитывается 2 раза.

Тогда всего для N полимеров есть

$$W = \frac{1}{N!} \prod_{k=1}^N \nu_k, \quad (1.21)$$

что упрощается [13] до

$$W = \left(\frac{n_0}{2N} \left(\frac{\gamma - 1}{e} \right)^{x-1} \right)^N \left(\frac{n_0}{n_0 - xN} \right)^{n_0 - xN}. \quad (1.22)$$

$$S_{mix} = -k \left[n \ln \left(\frac{n}{n + xN} \right) + N \ln \left(\frac{N}{n + xN} \right) \right] + k(x - 1)N[\ln(\gamma - 1) - 1] - kN \ln 2, \quad (1.23)$$

где $n = n_0 - xN$ – количество молекул воды.

Тогда

$$\frac{\partial S_{mix}}{\partial \nu_w} = N_A \frac{\partial S_{mix}}{\partial n} = R(-\ln(1 - v_2) - v_2(1 - 1/x)), \quad (1.24)$$

где $v_2 = xN/n_0 \approx V_2/V$ – доля молекул полимера. По построению x – количество таких «частей» полимера, чтобы работала модель «жесткой» решетки, на которой происходит смешивание. А это значит, что объем белка будет относиться к общему объему так же как количества ячеек, поэтому v_2 здесь имеет тот же смысл что и выше, где это было объемной долей.

Справедливо упомянуть, что такое описание не учитывает возможность самопересечений не-соседних звеньев одного полимера, что значит, что в реальности энтропия будет меньше чем полученная нами.

Теперь рассмотрим энергетический вклад того что 2 чистых вещества смешались [14]. В силу быстрого спадания дисперсионных сил и само-формирующегося экранирования кулоновских сил можно считать, что энергия связи в растворе формируется в основном за счет взаимодействия соседних молекул. Будем считать, что есть N_1 молекул 1-го типа, имеющих s_1 мест-для-связывания, и N_2 молекул 2-го типа с s_2 мест-для связывания. Тогда всего в системе $N_1s_1 + N_2s_2$ мест-связывания, формурующих $N_c = (N_1s_1 + N_2s_2)/2$ контантов.

Формирование контактов можно рассмотреть как выборание 2 экземпляров из группы объектов, содержащей N_1s_1 объектов 1 типа и N_2s_2 2 типа. Тогда для каждого из 2 выборов вероятность выбрать объект 1 типа будет $p_1 = N_1s_1/(N_1s_1 + N_2s_2)$. Тогда вероятность сформировать связь 1–1 $p_{11} = p_1^2$, откуда количество связей 1–1:

$$N_{11} = p_{11}N_c = \left(\frac{N_1s_1}{N_1s_1 + N_2s_2} \right)^2 \frac{N_1s_1 + N_2s_2}{2} = \frac{(N_1s_1)^2}{2(N_1s_1 + N_2s_2)}. \quad (1.25)$$

Аналогично получаем N_{22} и наконец $N_{12} = N_c - N_{11} - N_{22}$.

Тут мы пренебрегли изменением количества доступных для выбора объектов, произошедшем при выборе 1-го объекта. Это допустимо при $(N_1s_1 + N_2s_2) \gg 1$, что верно во всех практически интересных случаях.

Принимая энергии связей 1–1, 2–2, 1–2 за E_{11} , E_{22} , E_{12} соответственно, получаем

$$U = \frac{(N_1s_1)^2 E_{11} + 2N_1s_1N_2s_2 E_{12} + (N_2s_2)^2 E_{22}}{2(N_1s_1 + N_2s_2)}. \quad (1.26)$$

В случае чистых веществ (до смешивания) имеем

$$U_1 = \frac{N_1s_1}{2} E_{11}, \quad U_2 = \frac{N_2s_2}{2} E_{22}, \quad (1.27)$$

поэтому

$$\Delta U_{mix} = U - U_1 - U_2 = \frac{N_1N_2s_1s_2}{N_1s_1 + N_2s_2} (2E_{12} - E_{11} - E_{22}). \quad (1.28)$$

Вспоминая определение x как размер полимера в ячейках, ясно, что $x = s_2/s_1$, откуда (возвращаясь к обозначениям (1.23))

$$\Delta U = \frac{nxN}{n + xN} s_1 (2E_{12} - E_{11} - E_{22}) = k_B T \chi \frac{nxN}{n + xN}. \quad (1.29)$$

Было показано [3, 9, 15], что в интересующем нас диапазоне температур 0–50 $^{\circ}\text{C}$ параметр χ меняется от температура не более чем на $\sim 5\%$ по закону $\chi = \alpha + \beta/RT$.

Исходя из модели смешивания, объем системы не изменяется (т.е. смешивание идеальное). Давление также постоянно. Поэтому $\Delta(PV)_{mix} = 0$, откуда

$$\Delta H_{mix} = \Delta(U + PV)_{mix} = \Delta U_{mix} \quad (1.30)$$

и наконец

$$\frac{\partial \Delta H_{mix}}{\partial \nu_w} = N_A \frac{\partial \Delta H_{mix}}{\partial n} = RT \chi \frac{(xN)^2}{(n + xN)^2} = \chi RT v_2^2. \quad (1.31)$$

В итоге, подставляя (1.31) и (1.24) в (1.19), получаем

$$\mu_{mix} = RT (\chi v_2^2 + \ln(1 - v_2) + v_2(1 - 1/x)) \quad (1.32)$$

Тогда объединяя (1.15), (1.32) и (1.18) получаем уравнение на баланс хим. потенциалов воды в паре и в смеси с белком

$$RT \ln \left(\frac{P}{P_0} \right) = N_A \bar{V}_1 \frac{K}{2} (v_2^{-1/3} - 1) \left(\frac{5}{3} v_2^{-1/3} - 1 \right) + RT (\chi v_2^2 + \ln(1 - v_2) + v_2(1 - 1/x)), \quad (1.33)$$

где \bar{V}_1 – объем обной ячейки на Рис. 1.2.

Мы не знаем точное значение x , но ясно что $x \gg 1$. Деля на $RT v_2^2$ получаем окончательный вид, позволяющий находить значения K по линейному участку:

$$\frac{\ln(P/P_0) - \eta \ln(1 - v_2) - \eta v_2}{v_2^2} = \chi(T) + \frac{K \bar{V}_1}{2k_B T} \left(\frac{1}{v_2^{1/3}} - 1 \right) \left(\frac{5}{3v_2^{1/3}} - 1 \right) \frac{1}{v_2^2}, \quad (1.34)$$

где $0 < \eta < 1$ – неизвестный параметр, учитывающий то, что атомы такого компактного глобулярного белка как лизоцим могут не все в полной мере участвовать в процессе смешивания, что отразится на H_{mix} и S_{mix} .

Для нас важно, что (1.34) представляется в виде

$$f(P, T) = \chi(T) + K(T) \cdot g(P, T), \quad (1.35)$$

что позволяет провести измерения при различных P и фиксированной T , и по наклону зависимости определить K . Если считать что K не зависит от P , что в случае лизоцима верно для $P/P_0 \gtrsim 0.7$ [9], то

$$K(T) = \left. \frac{\partial f(P, T)/\partial P}{\partial g(P, T)/\partial P} \right|_{P/P_0 > 0.7} \quad (1.36)$$

1.3. Эксперимент

В этом разделе описывается получение и анализ экспериментальных данных [9], используемых далее как референсные значения. Использовался кристаллизованный очищенный высушенный лизоцим куриного яичного белка. Он помещался в ЯМР-камеру с контролируемой температурой и давлением водяного пара. Однократный импульс ($\sim 5 \mu s$) возбуждал ядра 1H импульсом в 0.8 Тл, что соответствует резонансной частоте 1H в 34 МГц. Спектр исследуемого порошка лизоцима состоял из 2 основных частей. Широкий пик с шириной на полувысоте ~ 40 kHz и узкий пик с ~ 2.4 kHz. Пример на Рис. 1.3:

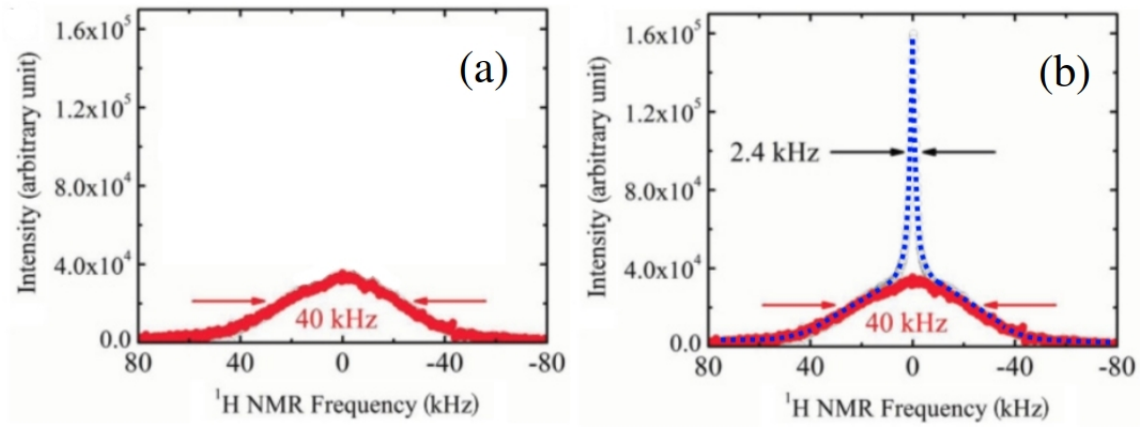


Рис. 1.3. 1H ЯМР-спектры порошка лизоцима в осушенном (а, красный профиль) и влажном (б, синий профиль) состояниях.

При откачке камеры до 10^{-3} Торг узкий пик рассасывается за ~ 24 часа, в то время как широкий пик не изменяется. Из этого делается вывод [16], что широкий пик порождается протонами 1H в составе самого лизоцима, в то время как узкий – протонами из сорбированной воды. Задержка измерения уста-

новки $\sim 8 \mu s$, что сравнимо с обратной шириной красного пика $1/\Delta\nu \sim 25 \mu s$. Поэтому при измерении количества водорода по этому пику надо учитывать затухание сигнала. Зная, что в данном случае затухание происходит по гауссовому закону [9], можно восстановить изначальную амплитуду сигнала [17]. Это восстановление – основной источник ошибки, составляющей $\sim 5\%$ [9]. Эта ошибка влияет на абсолютные значения h , но не на относительные, поэтому форма получаемых изотерм сорбции не изменяется.

Зная плотности воды и белка $\rho_w = \rho_1 = 1 \text{ g/cm}^3$, $\rho_{prot} = \rho_2 = 1.38 \text{ g/cm}^3$ [18], можно найти

$$v_2 = \frac{V_2}{V_1 + V_2} = \frac{1}{1 + V_1/V_2} = \frac{1}{1 + \frac{m_1 \rho_2}{m_2 \rho_1}} = \frac{1}{1 + h \frac{\rho_2}{\rho_1}} = \frac{1}{1 + 1.38h}, \quad (1.37)$$

что уже можно подставлять в (1.34).

На Рис. 1.4 построены изотермы в виде (1.34) при $\eta = 1$. Линейный участок $[\ln(P/P_0) - \ln(1 - v_2) - v_2]/v_2^2 > 0.6$ отвечает $P/P_0 > 0.7$ [9]. Объем \bar{V}_1 ячейки на 1 молекулу воды можно оценить как $\bar{V}_1 = (\mu_w/\rho_w)/N_A \sim 0.03 \text{ nm}^3$. Если принять, что χ и K не зависит от P , что в рассматриваемом случае верно [9], то их значения находятся из аппроксимаций линейных участков на Рис. 1.4.

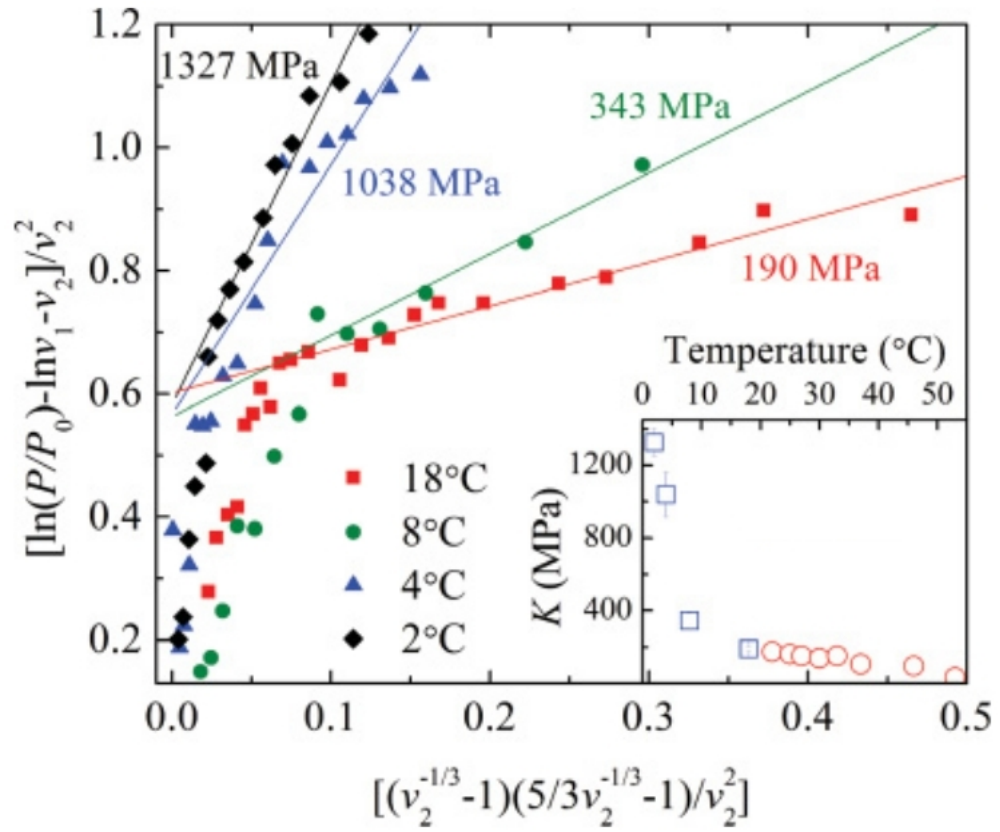


Рис. 1.4. Изотермы вида (1.34). Кривые линейны при $[\ln(P/P_0) - \ln(1 - v_2) - v_2]/v_2^2 > 0.6$. Использовалось значение $\eta = 1$. Малый график: Полученные значения упругости влажного кристалла в зависимости от температуры. Нанесены результаты, полученные в [9] (синие точки) и в [19] (красные точки)

Глава 2

Построение МД модели кристалла Лизоцима

В предыдущей главе были приведены все необходимые теоретические сведения для дальнейшего численного решения краевых задач модели Томаса-Ферми с поправками. В этой главе будет показано, как реализован алгоритм расчёта.

В предыдущей главе был описан метод, которым были получены референсные значения укрутости K , согласовавшиеся с ранее опубликованными значениями.

В разделе 2.1 описана подготовка кристалла лизоцима из базы PDB к симуляции. Далее в 2.2 продемонстрирована сходимость количества добавляемой в белок воды по некоторым базовым параметрам. Наконец, в 2.3 описаны 2 используемых метода расчета модуля упругости по имеющимся МД траекториям.

2.1. Подготовка PDB-белка к расчету

В качестве начальных данных используется модель белка лизоцима 1lee [20, 21]. PDB-файл загружается из онлайн базы данных rcsb.org/structure/1LEE. Данная структура получена рентгеноструктурным анализом. Изначально в файле явно хранится лишь 1 из 8 цепочек белка. Также в файле есть информация о молекулах воды, зарегистрированных при рентгеноструктурном анализе. И наконец, записаны операции симметрии, позволяющие имея структуру 1 цепочки восстановить положение всех 8-ми цепочек, составляющих одну периодическую ячейку кристалла.

При данной операции возникает проблема. Некоторые молекулы воды,

размноженные и перенесенные по тем же операциям симметрии, что и атомы цепочки белка, начинают перекрываться с другими молекулами воды. Видимо это происходит потому, что сложно точно отнести молекулы воды, полученные на рентгене, к той или иной цепочке белка. Если же такого соотношения воды с белковыми цепями нет, то ожидаемо, что часть молекул воды, явно хранимых в оригинальном файле, перекроются с другой водой, т.к. одна из молекул воды в такой перекрывшейся паре будет относиться не к той цепочки белка, которая хранится в файле явно. А поэтому такая воды не должна была участвовать в отражениях по симметрии в данной позиции. Для решения подобных перекрытий используется пакет UCSF Chimera [22]. В нем расстояния между атомами кислорода сортируются. Четко выделяется группа контактов с расстоянием < 0.01 нм, т.к. более длинные контакны уже имеют длину > 0.2 нм. Выбираются все атомы, участвующие в таких наложившихся парах. Из них выбирается минимальный набор атомов, удаление которого решит все перекрытия, и стирается из файла.

Далее происходит протонация белка. Оригинальные рентгеноструктурные данные не содержат протонов ^1H , т.к. последние слабо влияют на электронную плотность, которая на самом деле и рассеивает рентген. А нечувствительность электронной плотности к нахождению протонов затрудняет вписывание известной последовательности белка в измеренную 3-мерную структуру плотности. Процесс протонации соответствует алгоритму PDB2PQR [23, 24]. Опишем его подробнее.

Первым этапом вычисляются pK_a аминокислотных остатков белка. Это значения, такие, что когда pH среды переходит pK_a данного остатка, то меняется протонирование этого остатка. Строго говоря этот подход не совсем точен, т.к. бывает, что pH среды близко к pK_a остатка. Тогда в реальности доля протонированных по разному остатков будет сравнима. Мы не можем воспроизвести данный эффект в расчете, т.к. в модели сильно ограничено ко-

личество оригинальных белков, а все периодические образы протонированы одинаково по построению. Покажем, что это не сильное отклонение модели от реальности. По определению

$$pKa = \log_{10} \left(\frac{[AH] \cdot mol/L}{[A^-][H^+]} \right), \quad pH = -\log_{10} \left(\frac{[H^+]}{mol/L} \right) \Rightarrow \frac{[A^-]}{[HA]} = 10^{pH-pKa}, \quad (2.1)$$

где $[\dots]$ – концентрации соответствующих компонент в единицах моль/литр.

Видно, что при отличии pH среды от pKa всего на 1 концентрации остатков с разной протонацией будут отличаться уже в 10 раз, что уже почти эквивалентно однородной протонации в нашей модели, т.к. учитываются ~ 10 оригинальных цепей белка.

Определение pH также позволяет оценить, нужно ли пытаться включать в модель явные ионы диссоциировавшей воды

$$\begin{aligned} N_{H^+} &= 10^{-pH} N_A \frac{V}{[L]} \sim 0.6 \cdot 10^{-pH} \frac{V}{[nm^3]}, \\ N_{OH^-} &= 10^{pH-14} N_A \frac{V}{[L]} \sim 0.6 \cdot 10^{pH-14} \frac{V}{[nm^3]}, \end{aligned} \quad (2.2)$$

где V – объем ячейки, N_{H^+} , N_{OH^-} – количества ионов соответствующего типа в данном объеме. В используемой ячейке $V \sim 400 \text{ nm}^3$. В эксперименте $pH = 4.5$ [21], что дает $N_{H^+} \sim 10^{-2} \ll 1$. Далее, $pH < 7$, поэтому $N_{OH^-} < N_{H^+} \ll 1$. А итоге можно сказать, что явный учет диссоциации воды в ячейке не возможен при данном размере расчетов области.

Вообще говоря, окружение аминокислотного остатка влияет на его pKa . Процесс расчета описан в [25]. Изложим его кратко. Значения pKa в окружении полимерами складывается из основного pKa^{water} для свободного остатка, и из поправки за счет окружения ΔpKa . Референсные значения pKa^{wat} известны с хорошей точностью [25]. Поправка складывается из нескольких вкладов.

$$\Delta pKa = \Delta pKa^{QQ} + \Delta pKa^{desolv} + \Delta pKa^{HB} + \Delta pKa^{RE} \quad (2.3)$$

- ΔpKa_i^{QQ} – сумма вкладов в pKa_i от всех зарядов системы.
- ΔpKa_i^{desolv} – гидрофобные остатки эффективно отталкивают воду. Расталкивать чистую воду может быть сложнее или легче чем воду, окруженную белком, что и учитывает данное слагаемое.
- ΔpKa^{HB} – вклад водородных связей в энергию
- ΔpKa^{RE} – вклад прочих диполь-зарядных взаимодействий

Вычислив pKa всех остатков, можно на основе экспериментального pH определить для остатков количество необходимых дополнительных протонов на основе (2.1).

Для некоторых остатков, например HIS, ASN, TYR, существуют несколько состояний протонации с одинаковым зарядом. Для них нужно дополнительно проводить оптимизацию положения отдельных протонов и целых боковых цепей остатка.

Наконец, в процедуру протонации можно включить оптимизацию положения протонов в молекулах воды, что может уменьшить релаксационный период.

Имея протонированный белок, его нужно нейтрализовать добавлением ионов Na^+ и Cl^- . Так же, некоторое количество этих ионов отражены в рентгеноструктурных данных, и они тоже требуют нейтрализации. Недостающие ионы добавляются в вообще говоря случайные точки с тем условием, чтобы это было не слишком энергетически неоптимально. Это разумно, т.к. те ионы, которые устойчивы в системе, отражены на рентгеновских данных, а остальные ионы мобильны, поэтому не так важно где конкретно помещать их в начальный момент времени.

Далее идет процесс сольватации системы. В нем в систему добавляется определенное количество воды. Вопрос какое именно обсуждается далее.

В силу большого количества случайных факторов, участвовавших в создании системы до данного момента, ее итоговое состояние может быть энергетически крайне невыгодно из-за перекрытия атомов или неадекватной взаимной геометрии боковых цепей остатков. Это решается путем симуляции в режиме минимизации энергии. В данном случае используется «метод наискорейшего спуска». Формальное описание алгоритма:

1. Выбрать начальный размер шага, например $h_0 = 0.01$ нм.

2.

$$\mathbf{r}_{n+1} = \mathbf{r}_n + h_n \frac{\mathbf{F}_n}{\max(|\mathbf{F}_n|)}, \quad (2.4)$$

где r_n – $3N$ -мерный вектор координат всех атомов, F_n – $3N$ -мерный вектор всех сил на атомы.

3. Считаем новую энергию системы V_{n+1}

Если $V_{n+1} < V_n$, то $h_{n+1} = 1.2h_n$.

Если $V_{n+1} > V_n$, то $h_{n+1} = 0.2h_n$.

4. Повторяем пп. 2-3 пока не закончится заданное пользователем количество итераций, либо пока максимальная сила на атом не станет меньше заданного пользователем порога. Значение физической пороговой силы можно оценить как $\omega\sqrt{mk_BT}$, где ω – характерная частота в системе.

По сути такая «динамика» эквивалентна движению атомов в абсолютно вязкой жидкости с ограничением максимальной скорости. В такой модели из-за вязкости $\sim v$ мгновенно устанавливается режим $v \sim F$, что превращает классическую схему эйлера решения уравнений движения в (2.4).

Наконец после данного этапа система достаточно хорошо описывает реальность, чтобы при запуске динамики с физически-корректными законами движения она не развалилась.

2.2. Сходимость уровня гидротации

В предыдущем разделе 2.1 оставался один свободный параметр – количество воды, добавляемое в систему при сольвации. Имея возможность симулировать систему, можно рассчитывать в ней уже физические параметры, например давление. Нам известно, что в эксперименте давление было порядка атмосферного. Можно попробовать использовать эту информацию чтобы подобрать количество воды, приводящее к экспериментальному давлению. Для этого при каждой интересовавшей нас температуре были проведены несколько симуляций с разным количеством добавляемой воды. В каждой симуляции рассчитывалось давление, и далее интерполяцией находилось количество воды, приводящее к целевому давлению. Идея представлена на Рис. 2.1.

Таким образом возможно найти необходимое количество воды для заданных температуры и размера системы. Количество воды в системе – физически-измеримая величина, поэтому она не должна зависеть от таких модельных параметров как например размер ПГУ-ячейки или время усреднения. Сходимость по этим параметрам была достигнута, что показано на Рис. 2.2, 2.3. На Рис. 2.3 линии для размеров ПГУ-ячейки начиная с 2 лежат заметно кучнее чем линия для размера 1. Из этого можно заключить, что при усреднении в 1 нс. размер ПГУ-ячейки в 2 элементарных ячейки достаточен. Было проверено, что увеличение времени усреднения не приводит к расхождению подобных зависимостей для размеров расчетной области больше 2. На Рис. 2.2 видно, что данного времени усреднения не хватает чтобы различить зависимости для разных размеров ячейки. Подобная картина пропадает на

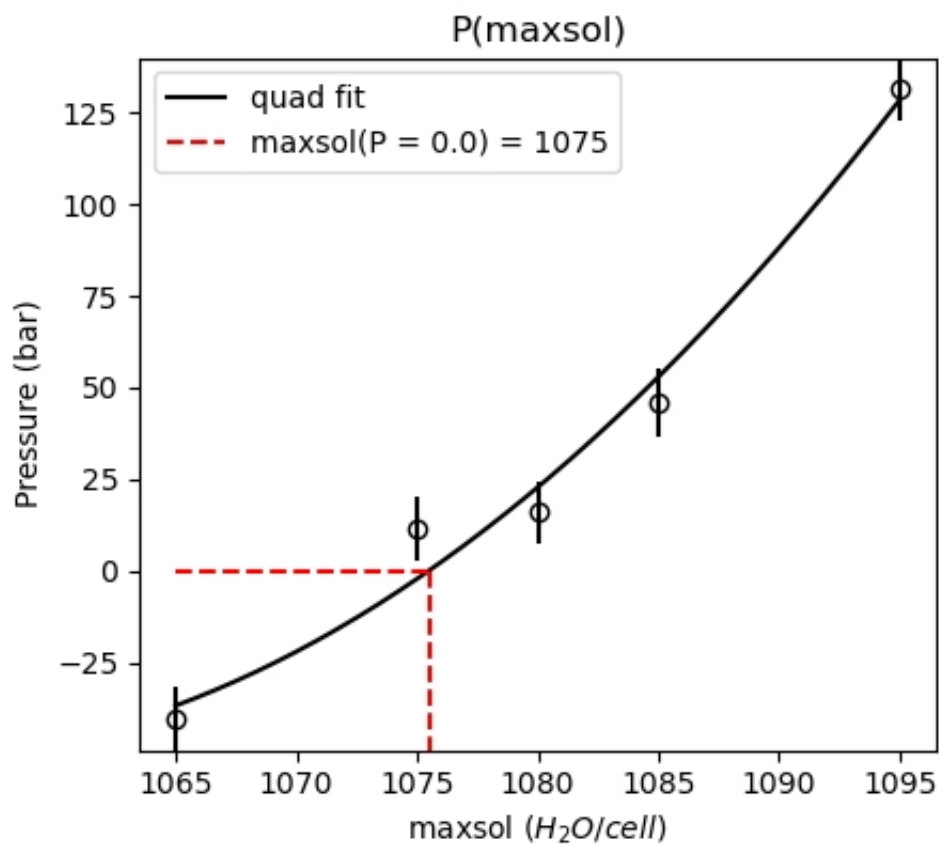


Рис. 2.1. Пример определения количества воды, приводящего к экспериментальному (фактически нулевому) давлению для заданной температуре. Данная кривая получена для системы размером в 2-е элементарные ячейки при $T = 15^\circ\text{C}$.

времени усреднения 1 нс., почему оно и было выбрано для «успешного» Рис. 2.3.

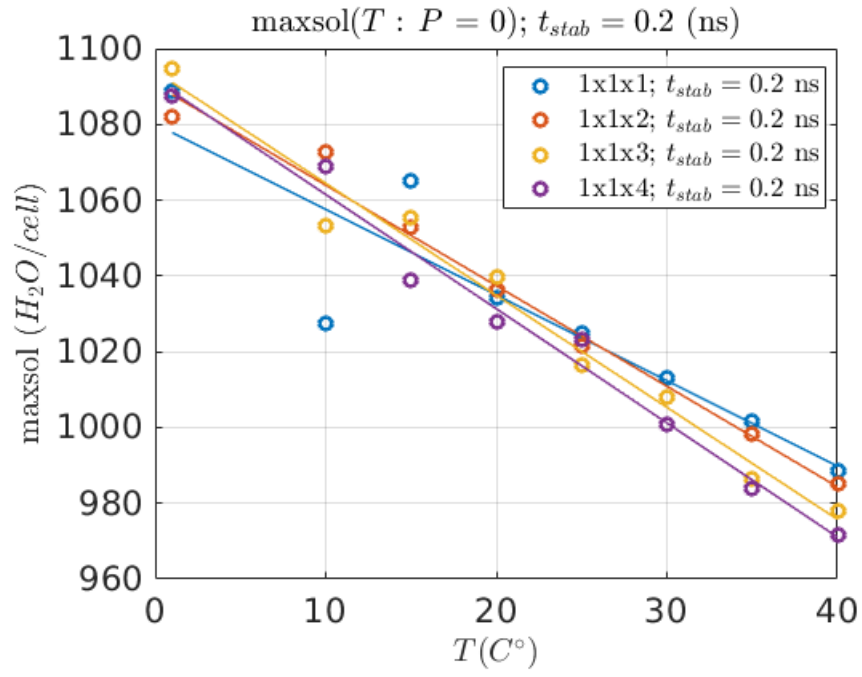


Рис. 2.2. Зависимость $maxsol$, рассчитанного аналогично Рис. 2.1, от температуры при различных временах усреднения и размерах расчетной области. Время усреднения $t_{stab} = 0.2$ ns.

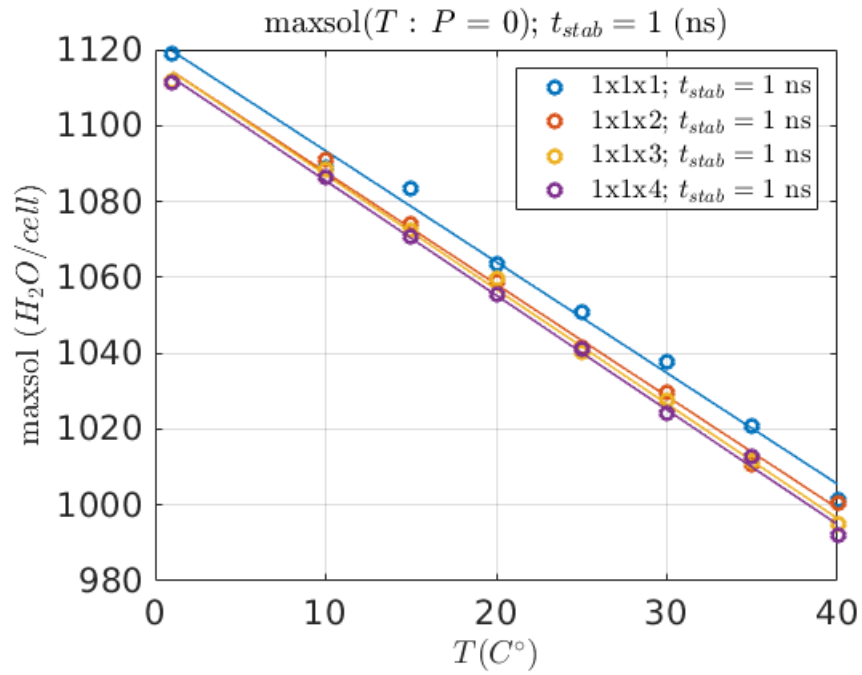


Рис. 2.3. Зависимость $maxsol$, рассчитанного аналогично Рис. 2.1, от температуры при различных временах усреднения и размерах расчетной области. Время усреднения $t_{stab} = 1.0$ ns.

2.3. 2 метода расчета модуля упругости K

Проверив сходимость физических параметров системы по нефизичным, можно переходить к расчету интересующей нас физической величины - объемного изотермического модуля упругости.

$$K = -V \left(\frac{\partial P}{\partial V} \right)_T \quad (2.5)$$

Помимо метода (1.36), применявшегося в обработке опроного эксперимента, рассмотрено еще 2 более прямых метода расчета K :

2.3.1. Флюктуационный метод

Можно показать [26], что при отклонениях от равновесия изменение энтропии системы + термостата ($k_B = 1$)

$$\Delta S_\Sigma = -\frac{\delta^2 E}{T}. \quad (2.6)$$

Используя

$$\begin{aligned} dE &= TdS - PdV + \mu dN \\ \delta &= \delta T \frac{\partial}{\partial T} + \delta V \frac{\partial}{\partial V} + \delta N \frac{\partial}{\partial N} \\ \frac{\partial(T, S)}{\partial(-P, V)} &= \frac{\partial(T, S)}{\partial(\mu, N)} = -1 \end{aligned} \quad (2.7)$$

можно получить

$$\delta^2 E = \delta T^2 \frac{\partial S}{\partial T} - \delta V^2 \frac{\partial P}{\partial V} + \delta N^2 \frac{\partial \mu}{\partial N} + 2\delta V \delta N \frac{\partial \mu}{\partial V}. \quad (2.8)$$

Подставим (2.8) в (2.6). Учитывая $\delta N = 0$ получаем

$$w \sim e^{\Delta S_\Sigma} = \exp \left(-\frac{\delta^2 E}{T} \right) = \exp \left(-\frac{\delta V^2 K}{VT} - \frac{\delta T^2}{T} \frac{\partial S}{\partial T} \right), \quad (2.9)$$

откуда переходя обратно в СИ

$$K_{\sigma^2} = \frac{k_B T V}{\langle \Delta V^2 \rangle} \quad (2.10)$$

Данные для расчета по (2.10) нужно собирать в NPT ансамбле, т.к. именно он воспроизводит естественные флуктуации объема и сохраняет количество частиц, что необходимо для корректного применения (2.9).

2.3.2. Метод малых деформаций

Идея метода представлена на Рис. 2.4:

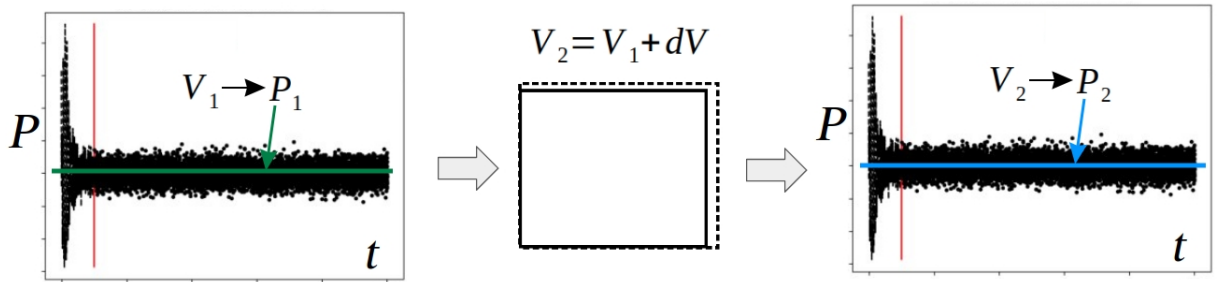


Рис. 2.4. Идея расчета K по определению. Сначала система выводится на равновесие в NVT ансамбле, затем ячейка немного увеличивается в размере, после чего система стабилизируется на новом значении давления.

По равновесным P_1, V_1, P_2, V_2 рассчитывается

$$K_{dV} = -\frac{V_1 + V_2}{2} \frac{P_2 - P_1}{V_2 - V_1} \quad (2.11)$$

В данной схеме есть малый параметр dV , малость которого не совсем очевидна. Расширение ячейки происходит мгновенно, при нем в моделируемой псевдо-бесконечной системе образуется сетка из вакуума, которая затем схлопывается на этапе стабилизации. Поскольку деформация происходит неравномерно по объему, то использовать критерий малости $dV/V \ll 1$ некорректно. Вместо этого, деформация сильно искажает химические связи, пересекающие периодическую границу, и не затрагивает остальные. Поэтому

предложено использовать критерий, основанный на малости энергии деформации химических связей на периодической границе. Возмущение, вносимое деформацией, будет малым, если оно искажает систему не сильно больше чем ее естественные колебания, т.к. естественные колебания точно соответствуют всем физическим ограничениям в системе. Естественные колебания порождаются температурой. Следовательно, имеет смысл сравнивать энергии температурного движения атомов с дополнительной энергией деформации, вносимой искусственным расширением ячейки: $\Delta Q \sim k_B T$. Химические связи часто описываются гармоническим потенциалом, что имело место и в используемом нами

$$\frac{\Delta Q}{k_B T} = \frac{\alpha \delta L^2}{2k_B T} \lesssim 1, \quad (2.12)$$

где α – жесткость химической связи, определяемая используемой параметризацией.

Обычно α указывается в молярных единицах, поэтому оценка малости деформации

$$\delta L \lesssim \sqrt{\frac{2RT}{\alpha}}. \quad (2.13)$$

Для типичных значений жесткостей это упрощается до

$$\delta L \lesssim [pm] \sqrt{\frac{T/[K]}{20}} \sim 4pm \quad (T \sim 300K) \quad (2.14)$$

С другой стороны, при расчете по (2.11) мы вычитаем близкие значения давлений, которые при этом имеют погрешность. Можно оценить ее исходя из закона $\sim 1/\sqrt{N}$:

$$\Delta P \sim \frac{C_1}{\sqrt{tN}} \sim \frac{C}{\sqrt{tV}}. \quad (2.15)$$

Такая запись разумна т.к. количество статистики, используемое для получения P , линейно растет с размером системы V и с временем усреднения t . Объем системы задается точно, поэтому не вносит погрешности, а поэтому можно записать

$$K = V \frac{\delta P}{\delta V}. \quad (2.16)$$

Далее использованы обозначения ΔX – погрешность расчета X , δX – разность X при различных объемах.

Тогда

$$\frac{\Delta K}{K} = \frac{V \delta V}{V \delta P \delta V} \frac{C \sqrt{2}}{\sqrt{tV}} = \frac{C \sqrt{2}}{\delta P} \frac{1}{\sqrt{tV}}. \quad (2.17)$$

Поскольку локальной характеристикой вещества является K , то логично представить $\delta P \sim K \delta V / V$. Также, объем в первом порядке увеличивается как $\delta V \sim 3V^{2/3} \delta L$. Тогда еще одно ограничение на амплитуду смещения

$$\frac{\Delta K}{K} = \frac{C \sqrt{2}}{K} V^{-1/6} \frac{1}{3\sqrt{t} \delta L} \ll 1. \quad (2.18)$$

Объединяя это с (2.13), получаем ограничение на время усреднения

$$t \gg V^{-1/3} \frac{\alpha}{RT} \left(\frac{C}{3K} \right)^2. \quad (2.19)$$

Значение C можно оценить из одного расчета. $K = 100$ МПа для оценки было взято из [19] – оно соответствует $T \sim 40C^\circ$ и дает верхнюю оценку на (2.19) для более высоких T , т.к. K растет с уменьшением T . В итоге $t > 200$ нс. для системы размером $1 \times 1 \times 2$ элементарных ячеек. Это достаточно большое время для одной симуляции, поэтому для практического использования был проведен переборный поиск оптимальных значений малости деформации и времени усреднения, описанный ниже.

Глава 3

Результаты моделирования

В этой главе приведены результаты и анализ расчётов.

В разделе 3.1 предложенные выше методы расчета K проверяются на работоспособность на простой системе чистой воды.

Далее в разделе 3.2 обсуждается теория баро- и термо-статирования, результаты применения которых затем проверяются на сходимость в разделе 3.3.

Значения K рассчитаны в разделе 3.4. Обсуждаются возможные причины разхождения с экспериментом. В разделе 3.6 делается попытка исправить одну из возможных причин расхождения – неверный процесс релаксации системы перед расчетами – при помощи проведения 2-фазного моделирования для нахождения равновесных параметров системы.

3.1. Верификация прямых методов расчета K

Прежде чем применять методы (2.11) и (2.10) для «сложной» системы воды+белок, надо проверить их на более простом примере. Например для воды хорошо известно экспериментальное значение K . К тому же, вода будет участвовать и в целевом расчете, поэтому проверить ее адекватное воспроизведение хотя бы в чистом виде все равно необходимо. Тут, как и далее, используется модель воды tip4p/2005 [27]. Выбор в пользу более сложной модели воды вместо часто используемых в симуляциях белков SPC/E [28–30] и tip3p [31, 32] сделан потому, что эффект роста K при уменьшении температуры наблюдается на масштабах $\sim 10K$, в то время как даже такие менее тонкие эффекты как например температура кристаллизации воды занижает-

ся моделью SPC/E на $\sim 50K$.

На Рис. 3.1 приведено сравнение экспериментального K с рассчитанными по (2.11) и (2.10). Флюктуационный метод совпадает с экспериментом в пределах погрешности, но эта погрешность достаточно велика $\sim 10\%$. Метод малых смещений дает меньшую погрешность, и дает отклонение от эксперимента на $\sim 5\%$.

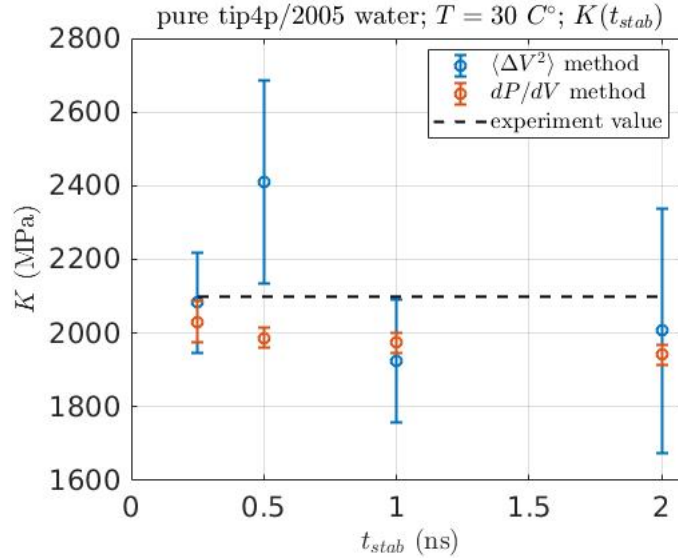


Рис. 3.1. Сравнение K , полученных двумя прямыми методами (2.11) и (2.10), с экспериментальными значениями. t_{stab} – время в начале моделирования, отбрасываемое при анализе. Как видно, система из чистой воды выходит на равновесие уже после 0.25 нс. эволюции.

3.2. Используемые баростат и термостат

Численное решение неизменных уравнений Ньютона приводит к симуляции NVE ансамбля, однако часто бывает интересно поведение системы в других условиях, таких как NVT и NPT . Их воспроизведения в расчете не так тривиально. Например, существуют достаточно идейно-простые методы термостатирования системы, такие как термостат Берендсена. Физически он заключается в насильном выполнении

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}, \quad (3.1)$$

где T – текущая температура системы, T_0 – целевая температура, t – время, τ – характерное время релаксации системы к целевой температуре.

При дискретизации (3.1) аппроксимируется [33] масштабированием скоростей частиц на фактор

$$\lambda = \sqrt{1 + \left(\frac{T_0}{T} - 1\right) \frac{\Delta t}{\tau_T}}, \quad \tau_T = \tau \frac{3N_{df}k_B/2}{C_V}, \quad (3.2)$$

где C_V – теплоемкость системы при постоянном объеме, N_{df} – количество степеней свободы в системе.

Проблема таких «наивных» подходов заключается в том, что они как правило искажают распределение частиц в фазовом пространстве таким образом, что средние значения термодинамических величин не затрагиваются, но флуктуации и тем более высшие моменты начинают расходиться с естественным поведением моделируемых ансамблей. В случае термостатирования эта проблема решается термостатом Нозе-Гувера [34, 35]. Его идея состоит в модификации уравнений движения путем введения дополнительной степени свободы s «резервуара», с которым связана динамика всех реальных частиц системы:

$$\begin{aligned} \frac{d^2 \mathbf{r}_i}{dt^2} &= \frac{\mathbf{F}_i}{m_i} - \frac{p_s}{Q} \frac{d\mathbf{r}_i}{dt} \\ \frac{dp_s}{dt} &= \frac{N_{df}k_B(T - T_0)}{Q} \end{aligned} \quad (3.3)$$

Как видно, при $T > T_0$ коэффициент трения положителен и энергия диссипирует, поэтому T уменьшается. При $T < T_0$ же наоборот происходит разгон частиц т.к. «трение» отрицательно. Это порождает затухающие осцилляции температуры около целевого значения. Характерный период осцилляций

$$\tau_T \sim 2\pi \sqrt{\frac{Q}{2N_{df}k_B T}} \quad (3.4)$$

Уравнения (3.3) не гамильтоновы, но можно показать [34], что гамильтониан

$$\begin{aligned} \mathcal{H}_0(\mathbf{p}, \mathbf{r}) &= \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \\ \mathcal{H}(\mathbf{p}, \mathbf{r}, p_s, s) &= \mathcal{H}_0\left(\frac{\mathbf{p}}{s}, \mathbf{r}\right) + \frac{p_s^2}{2Q} + N_{df}k_B T \ln(s) \end{aligned} \quad (3.5)$$

приводит к системе, аналогичной (3.3) в переменных $(q, p/s)$.

Наличие гамильновой формулировки позволяет применить формализм расложения Троттера для оператора эволюции системы, что в свою очередь позволяет без затруднений выписывать численные схемы, аналогичные схемам Verlet, для сложных ансамблей [36].

При этом симуляция (3.5) в NVE (точнее $NV\mathcal{H}$) ансамбле будет соответствовать симуляции невозмущенной системы \mathcal{H}_0 в NVT . Это происходит потому, что $NV\mathcal{H}$ -статсумма системы

$$Z_{N-H} = \frac{1}{N!} \int dp_s \int ds \int d\mathbf{p} \int d\mathbf{r} \delta[\mathcal{H}(\mathbf{p}, \mathbf{r}, p_s, s) - E] \quad (3.6)$$

представляется [34] как

$$\begin{aligned} Z_{N-H} &= \frac{1}{N_{df} + 1} \left(\frac{2\pi Q}{kT_0} \right)^{1/2} \exp\left(\frac{E}{k_B T_0} \right) Z_{NVT}\left(\frac{\mathbf{p}}{s}, \mathbf{r}\right), \\ Z_{NVT}(\mathbf{p}, \mathbf{r}) &= \frac{1}{N!} \int d\mathbf{p} \int d\mathbf{r} \exp\left[-\frac{\mathcal{H}_0(\mathbf{p}, \mathbf{r})}{k_B T_0} \right] \end{aligned} \quad (3.7)$$

что при усреднении любой термодинамической величины, зависящей от $(\mathbf{p}/s, \mathbf{r})$, эффективно приведет к усреднению лишь по $Z_{NVT}(\mathbf{p}/s, \mathbf{r})$, что как раз эквивалентно симуляции NVT ансамбля. Выражаясь более строго,

$$\left\langle A \left(\frac{\mathbf{p}}{s}, \mathbf{r} \right) \right\rangle_{N-H} = \langle A(\mathbf{p}, \mathbf{r}) \rangle_{NVT} \quad (3.8)$$

Следует уточнить, что реально используемая схема принципиально следует изложенной идее, но не точно ее повторяет. Дело в том, что уравнения (3.3) могут быть не эргодичны, и есть примеры, когда это действительно так [37]. Для решения этой проблемы предложено использовать цепи Нозе-Гувера [38]. Добавление даже нескольких звеньев уже существенно улучшает эргодичность, хотя и не делает ее математически точной. Доказано, что при стремлении количества цепей к бесконечности динамика системы будет эргодичной. В расчетах используется цепь из 10 звеньев, как достаточная для подавляющего большинства расчетов. Внимательно рассматривать значения этого параметра может быть разумно при проведении термодинамического интегрирования, чего в нашем случае не происходит, поэтому используется стандартное значение.

Для приложения внешних напряжений так же существуют как более интуитивные методы, которые однако не воспроизводят динамику системы со всей полнотой, так и более сложные методы, которые доказано моделируют NPT ансамбль.

Например есть аналог термостата Берендсена [39], реализующий теперь идею

$$\frac{d\hat{P}}{dt} = \frac{\hat{P}_0 - \hat{P}}{\tau_p}, \quad (3.9)$$

которая при дискретизации реализуется масштабированием ячейки и координат всех атомов в ней с помощью матрицы

$$\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\tau_p} \beta_{ij} \{P_{0ij} - P_{ij}(t)\}, \quad (3.10)$$

где β_{ij} – тензор изотермической сжимаемости, определяемый как [40]

$$\beta_{ij}\Delta P = \varepsilon_{ij} = \frac{1}{2} \left[(\hat{h}\hat{h}_0^{-1})^T \hat{h}\hat{h}_0^{-1} - 1 \right], \quad (3.11)$$

где $\hat{h} = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ – матрица, составленная из базисных векторов периодической расчетной ячейки, \hat{h}_0 – равновесное положение ячейки.

Качественно понять определение (3.11) можно так. Если \hat{h} и \hat{h}_0 диагональны, то можно перейти к рассмотрению деформации по одной произвольно выбранной оси, а по остальным рассуждения будут аналогичны. При проекции на 1 ось диагональные \hat{h} и \hat{h}_0 переходят в $L = L_0 + \Delta L$ и L_0 соответственно, что дает при выборе например оси X :

$$\beta_{xx}\Delta P = \frac{1}{2} \left[\left(\frac{L_0 + \Delta L}{L_0} \right)^2 - 1 \right] \approx \frac{\Delta L}{L_0} \Rightarrow \beta_{xx} = \frac{1}{L_0} \frac{\Delta L}{\Delta P} \sim \left(\frac{1}{L} \frac{\partial L}{\partial P} \right) \Big|_{L=L_0} \quad (3.12)$$

Как и в случае в термостатом Берендсена, данный алгоритм относительно прост, однако воспроизводимая им динамика вообще говоря отличается от естественной, хотя и не изменяет средние значения термодинамических величин.

Парринело и Рахман [40] предложили подход, аналогичный подходу Нозе-Гувера. Уравнения движения системы модифицируются как

$$\begin{aligned} \frac{d^2 \mathbf{r}_i}{dt^2} &= \frac{\mathbf{F}_i}{m_i} - \left[\hat{h}^{-1} \frac{d\hat{h}}{dt} + \left(\hat{h}^{-1} \frac{d\hat{h}}{dt} \right)^T \right] \frac{d\mathbf{r}_i}{dt}, \\ \frac{d\hat{h}^2}{dt^2} &= V\hat{W}^{-1}(\hat{h}^T)^{-1} (\hat{P} - \hat{P}_0) \end{aligned} \quad (3.13)$$

где V – объем системы, \hat{W} – параметр силы баростата, аналог массы для степеней свободы размеров периодической ячейки.

Эти уравнения сохраняют гамильтониан

$$\mathcal{H}_{P-R} = \sum \frac{m_i \mathbf{v}_i^2}{2} + U(\mathbf{r}_1, \dots, \mathbf{r}_N) + \frac{\text{Tr}(\dot{\hat{h}}^T \hat{W} \dot{\hat{h}})}{2} + \frac{\text{Tr}(\hat{P})}{3} \det(\hat{h}), \quad (3.14)$$

где \mathbf{v}_i учитывает только движение частицы внутри ячейки, но не изменение самой ячейки. Т.е. если представить $\mathbf{r} = \hat{h}\mathbf{s}$, то $\mathbf{v} = \dot{\hat{h}}\mathbf{s}$, но не $\dot{\hat{h}}\mathbf{s} + \hat{h}\dot{\mathbf{s}}$.

Гамильтониан (3.14) равен энтальпии системы H с точностью $\sim 1/N$, что как раз приводит $NP\mathcal{H}_{P-R}$ ансамбль к $NP\mathcal{H}$ ансамблю.

Параметр \hat{W} не очень интуитивен, поэтому полезно указать его связь с более явными величинами:

$$\left(\hat{W}^{-1}\right)_{ij} = \frac{4\pi^2 \beta_{ij}}{3\tau_p^2 \max(\{L_k\})}, \quad (3.15)$$

где L_k – размеры ячейки.

3.3. Сходимость K по параметрам термостата и баростата

Как упоминалось выше, расчет термодинамических величин, таких как K , при активном термо- или баро-статировании может сказаться на получаемых результатах. Поэтому прежде чем проводить вычислительные измерения K , нам нужно найти такие параметры термостата и баростата τ_T , τ_P (являющиеся параметрами, отсутствующими в реальности), чтобы получаемые значения K уже не зависели от них. Также должна быть достигнута сходимость K по времени усреднения.

Такой анализ был проведен, и его результаты приведены на Рис. 3.2:

Видно, что времени усреднения 20 нс. не хватает чтобы получать стабильные K при любой силе внешнего воздействия. Времени же в 40 нс. уже достаточно. На том же графике для 40 нс. видна граница силы воздействия,

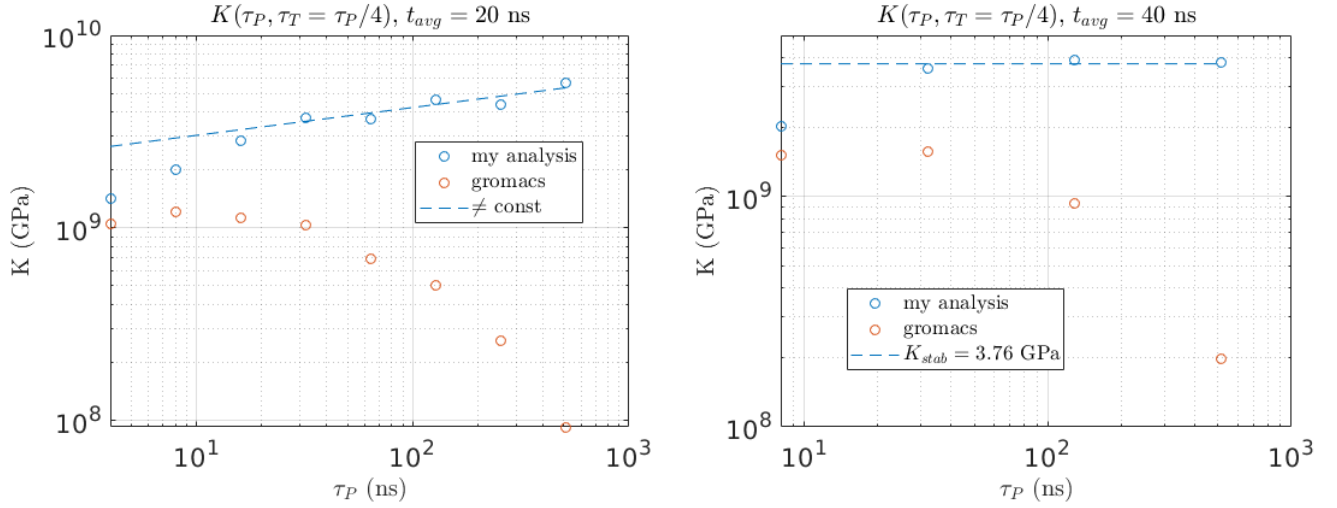


Рис. 3.2. Проверка сходимости K по силе баростата и термостата для флуктуационного метода расчета (2.10). Для уменьшения количества пробных расчетов одновременно менялись силы баростата и термостата, что отражено как $\tau_T = \tau_P/4$. Погрешности точек размеров с символы точек. Синие точки - анализ расчетных данных, проведенный в данной работе. Красные точки - значения, автоматически рассчитываемые Громаксом. Отличие обсуждается ниже.

после которой зависимость от силы термостата пропадает. Но делать теостат слишком слабым тоже не оптимально, т.к. это скажется на скорости уменьшения погрешностей точек, что приведет к необходимости заметно увеличивать время симуляции. В итоге было выбрано значение $\tau_P = 128$ нс. и соответствующее значение τ_T .

На Рис. 3.2 так же присутствуют красные точки, отвечающие значением K , автоматически вычисленным пакетом GROMACS [41]. В нем используется так же формула (2.10), но отличие состоит в том, что Громакс не учитывает наличие начального этапа релаксации. Причина наглядно пояснена на Рис. 3.3.

Для метода малых смещений (2.11) нужно провести аналогичную проверку по сходимости параметров метода. В данном случае это малость изменения объема и сила термостата, выражаемая τ_T . Результаты тестов представлены на Рис. 3.4

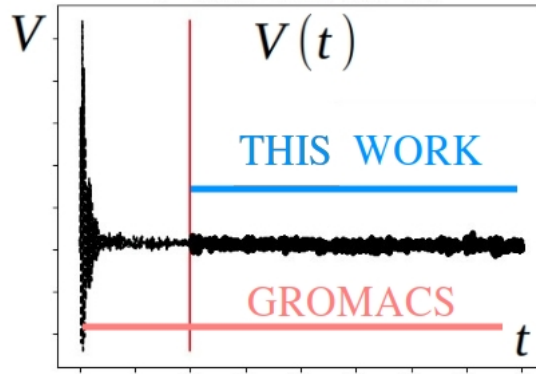


Рис. 3.3. Сравнение реализаций расчета K по (2.10), использованных в данной работе и автоматически проводимых Громаксом. Громакс использует для расчета $\langle \Delta V^2 \rangle$ вся траекторию, не учитывая начальный этап релаксации. Это приводит к завышению $\langle \Delta V^2 \rangle$, что наглядно видно на рисунке – во время релаксации колебания объема значительно больше чем в равновесном состоянии. И чем слабее баростат, тем дольше занимает этап релаксации, и соответственно тем больше завышается значение $\langle \Delta V^2 \rangle$, что все больше занижает значения K . Этот эффект явно виден на Рис. (3.2)

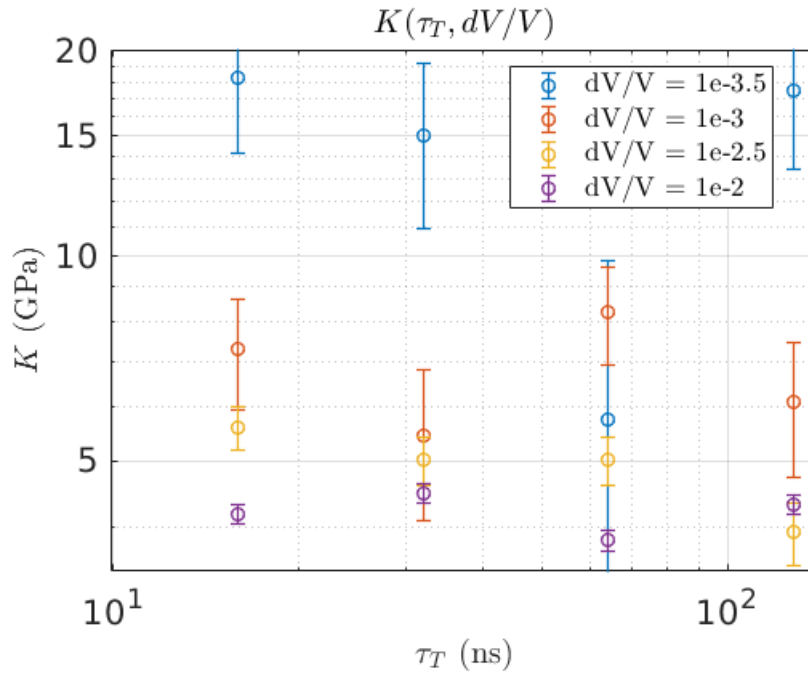


Рис. 3.4. Проверка сходимости метода расчета K малыми смещениями (2.11) по малости изменения объема и по силе термостата. Для согласованности было использовано время усреднения 40 нс., т.к. выше было найдено, что это время оптимально для расчетов по флуктуационному методу. Видно, что в области $\tau_T > 32$ нс., $dV/V > 10^{-2.5}$ значения K стабильны в пределах погрешности.

Для флюктуационного метода выше было выбрано $\tau_T = 32$ нс., поэтому здесь для согласованности мы взяли то же значение, т.к. оно входит в область плато, указанную на Рис. 3.4. Значение dV было выбрано минимально возможным еще входящим в плато из соображений наименьшего искусственного возмущения системы. Можно было бы пытаться увеличивать время усреднения, и таким образом получать плато на еще меньших значениях dV , т.к. ошибка возникает из-за того что при малых dV ожидаемое изменение dP также мало, поэтому хуже регистрируется при малом усреднении. Но чтобы уменьшить допустимое dV на пол порядка, нужно увеличить время усреднения в 10 раз. Поэтому было решено оставить время усреднения и выбрать dV достаточно большое, чтобы этого времени хватало для различения dP , т.е. в данном случае $dV/V = 10^{-2.5}$.

3.4. Расчет K прямыми методами

Получив сходимость по нефизичным параметрам методов, мы можем наконец применить их для расчета K белка при различных температурах. Используются ранее выбранные параметры методов:

- NVT: $dV/V \sim 10^{-2.5}$, $\tau_T = 32$ нс., $t_{avg} = 40$ нс.
- NPT: $\tau_P = 128$ нс., $\tau_T = 32$ нс., $t_{avg} = 40$ нс.

Так же используется метод определения количества воды в системе, описанный в разделе 2.2, основанный на подборе количества воды для достижения экспериментального давления.

Результаты расчетов представлены на Рис. ??:

Из-за самосогласованности результатов, полученных двумя независимыми методами, можно говорить о физичности рассчитываемого K . Однако, по-

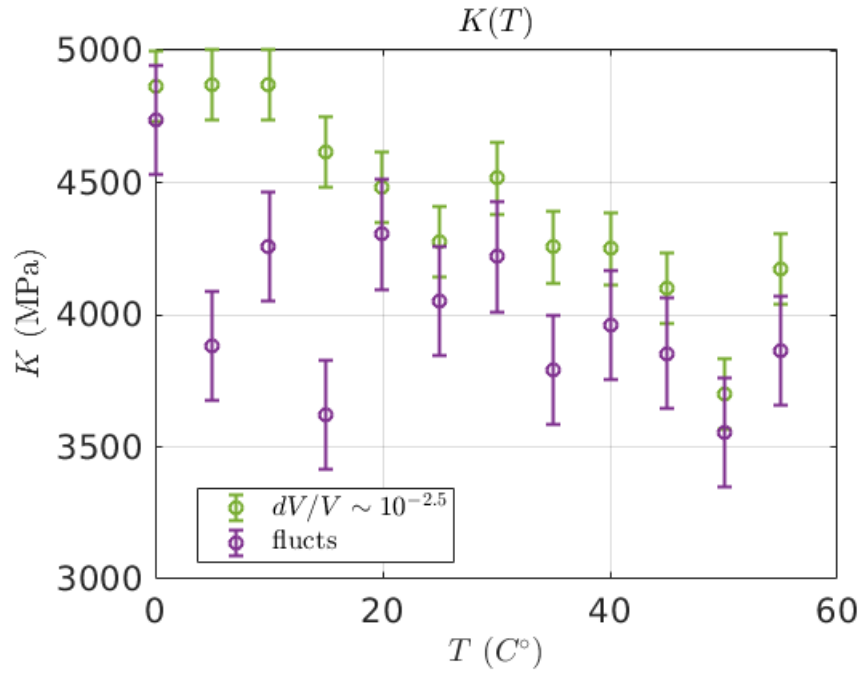


Рис. 3.5. Результаты расчета K прямыми методами (2.11) и (2.10) при гидратации, описанной в разделе 2.2. Методы дают согласованные значения, что говорит в пользу корректности полученных результатов.

лученные значения заметно отличаются от экспериментальных, что демонстрируется на Рис. 3.6:

У данного расхождения может быть множество причин, и их исследование составляет нижеизложенную часть представленной работы и имеет перспективу для дальнейшего развития.

Одной из возможных причин может быть то, что используемая параметризация белка не воспроизводит именно K . В расчетах использовалась параметризация amber03 [42] с моделью воды tip4p/2005 [27]. Как упоминалось в разделе 3.1, вода предположительно [9] играет важную роль в эффекте излома зависимости $K(T)$, в то время как стандартные модели воды SCP/E и tip3p воспроизводят даже такой параметр как температуру кристаллизации с недопустимой в данном случае ошибкой. Поэтому было решено использовать воду tip4p, которая воспроизводит по крайней мере фазовую диаграмму с приемлимой в данном случае точностью. Выбор данной модели воды же

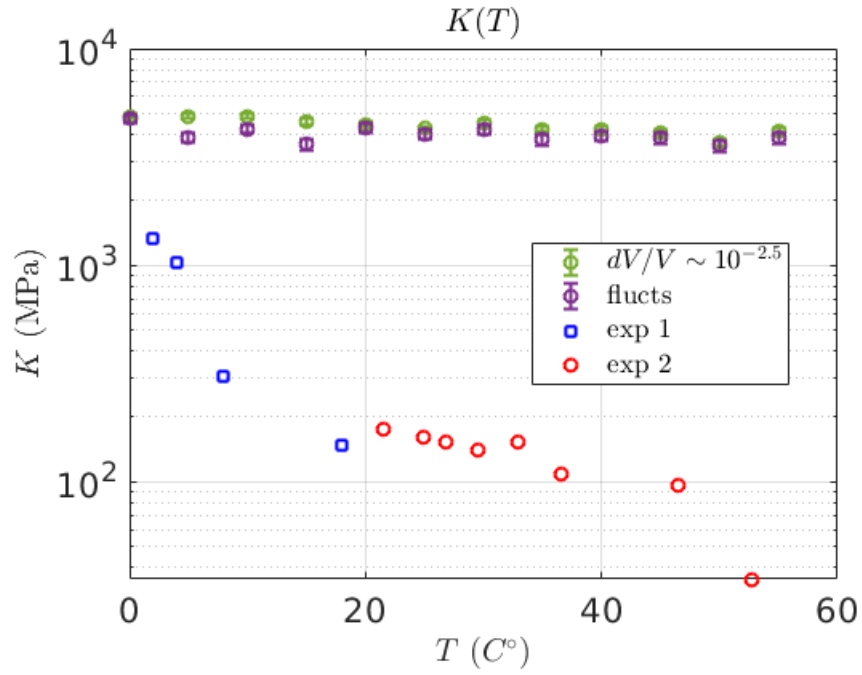


Рис. 3.6. Сравнение значений K , рассчитанных двумя методами, с экспериментальными значениями. Кресты (exp 1) – работа [9], Круги (exp 2) – [19]. Возможные причины расхождения обсуждаются далее.

повлек использование параметризации amber03, т.к. именно она проверенно согласованна с водой tip4p/2005.

Параметры потенциала amber03 были получены из квантовомеханических расчетов. Верификация же проводилась сравнением карт Рамачандрана углов полипептидной связи. Авторам не известно о тестах потенциала amber03, подтверждающих воспроизведение им упругих свойств Лизоцима, поэтому это может не иметь места с достаточной точностью. Можно привести кафельные рассуждения, почему воспроизведение карт Рамачандрана не обязано свидетельствовать об ожидаемых упругих свойствах кристалла. Дело в том, что карты Рамачандрана характеризуют равновесное состояние белка, т.е. распределение средних значений углов полипептидной связи. Упругие же свойства при малых деформациях будут зависеть не от равновесных значений углов, а от резкости потенциальных ям, в которых находятся эти средние значения. Об этом говорит возможность расчета K из флуктуаций объема, т.е.

из динамической характеристики. А поэтому карты рамачандрана зависят не совсем от тех же сочетаний параметров потенциала, от которых зависит K , что значит, что их воспроизведение не так жестко связано с воспроизведением K . Стоит однако заметить, что заметная корреляция должна быть, поскольку большое количество совпавших с экспериментом точек на карте рамачандрана накладывает много ограничений на возможные значения параметров потенциала.

Другой возможной причиной мы считаем неверную работу с водой в системе. Дело в том, что в эксперименте порошок лизоцима находился в равновесии с водяными парами. Поэтому если мы воспроизведем контакт сред пар-белок в расчете, то система сама может прийти к правильному количеству воды в белке, вытеснив лишнюю воду в соответствующий объем изначально пустого пространства.

3.5. Анализ мобильности воды

Интересным исследованием, проведенным в процессе работы, было изучение мобильности молекул воды внутри кристалла белка. Данный вопрос может быть актуален в контексте расчета K , т.к. наличие большого количества жестко связанной с белком воды может сильно влиять на жесткость всей системы в целом. Поэтому возникает задача классификации молекул воды на «мобильные» с «связанные». Подобная классификация может быть полезна еще и тем, что позволит проследить динамику распределения молекул воды по мобильности. Из физических соображения ясно, что в системе, вышедшей на равновесие, подобное распределение не должно изменяться со временем, хотя конкретные молекулы конечно будут могут менять свою мобильность в процессе эволюции.

3.5.1. Максимальное смещение

Для построения распределения воды по мобильности нужно определить, что именно мы называем мобильностью молекулы. Простейшим вариантом является максимальное смещение на время симуляции, т.е. $m_i = \max_t((\mathbf{r}_i(t) - \mathbf{r}_i(0))^2) = \max_t(\Delta r_i(t)^2)$. Типичное распределение по данной характеристике приведено на Рис. 3.7:

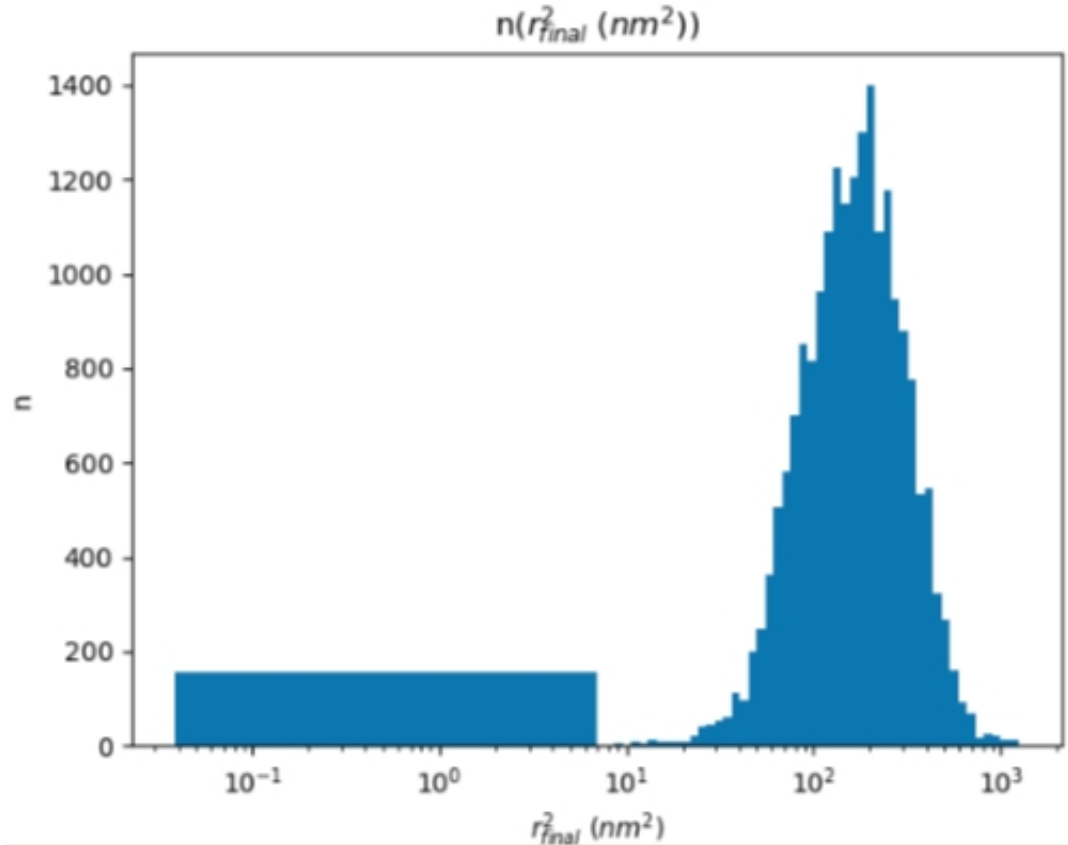


Рис. 3.7. Распределение молекул воды по максимальному за время симуляции смещению.

Видно, что распределение на Рис. 3.7 представляет собой один главный пик и небольшое количество сильно связанных молекул. Эта небольшая группа сильно отличается по свойствам от остальной воды и будет легко обнаружима на всех дальнейших методах анализа. Основная же группа имеет куполообразную форму с единственным пиком, поэтому плохо классифицируется. Дело в том, что качественно зависимости $\Delta r(t)^2$ делятся на 2 типа (Рис. 3.8):

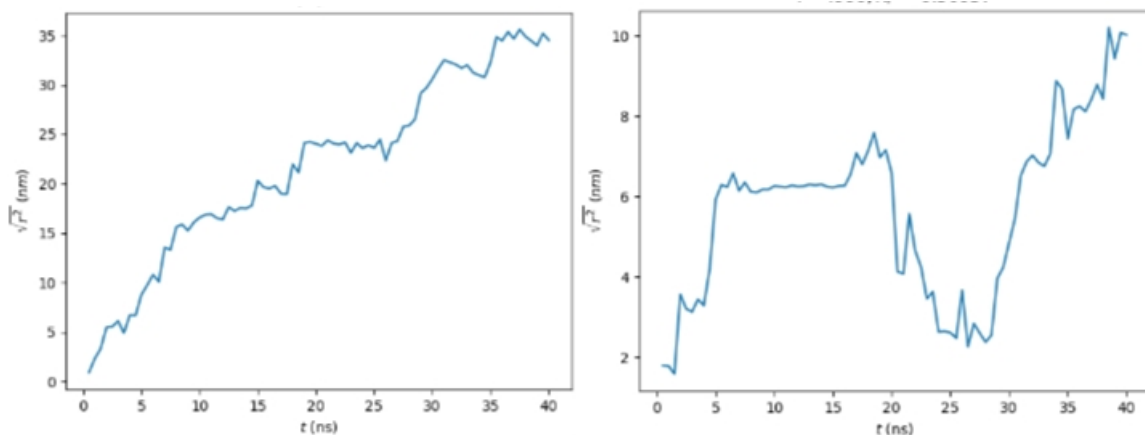


Рис. 3.8. Типичные виды зависимостей $\sqrt{\Delta r(t)^2}$.

Как видно на Рис. 3.8, есть молекулы, просто почти свободно движущиеся в системе (рисунок слева), а есть такие, которые часть времени находятся в практически связанном состоянии (плато на графике), а в другие моменты времени быстро перемещаются. Оба режима перемещения могут давать схожее максимальное смещение, однако из физических соображений ясно, что классифицировать молекулы из этих 2 групп как похожие – не очень корректно.

3.5.2. Коэффициент диффузии

Следующим испытанным вариантом метрики мобильности было значение коэффициента диффузии молекулы D , т.е. фактически наклон графика $\Delta r^2(t)$.

Проблема D как метрики мобильности опять же заключается в наличии единственного пика, что видно на Рис. 3.9. Это не позволяет разумно разбить молекулы на 2 группы. Были предприняты также попытки вводить определенный граничный D_{mob} , и считать мобильными все молекулы с $D > D_{mob}$. Однако недостаток такого подхода в том, что характерные D в системе сильно зависят от температуры, поэтому для воспроизводимости классификации при различных температурах нужно вводить температурную зависимость и в

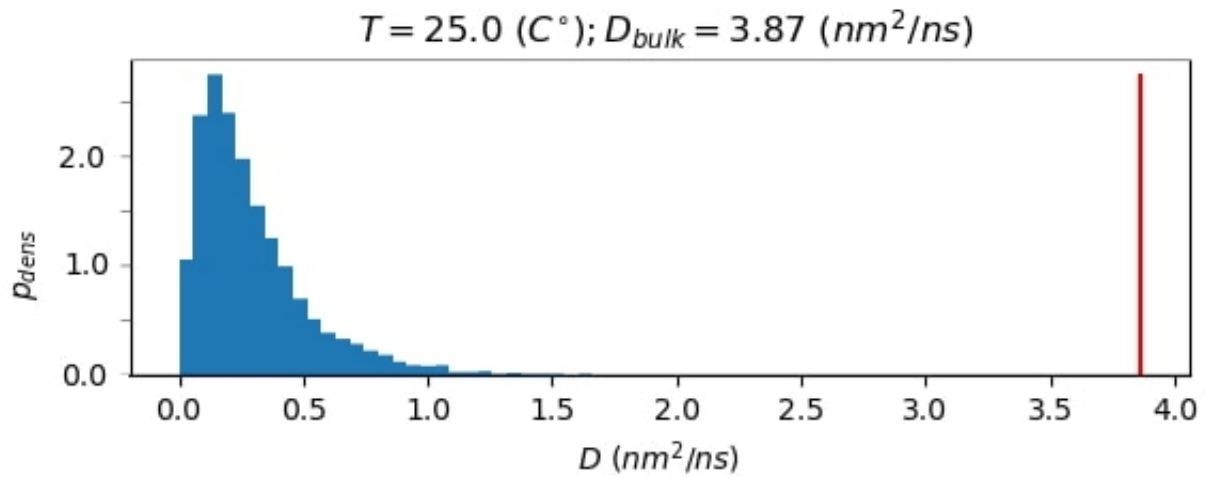


Рис. 3.9. Распределение молекул воды по их коэффициенту диффузии

$D_{mob}(T)$. Это можно сделать, измерив или рассчитав D_{pureW} в чистой воде при различных T , и положив $D_{mob}(T) = \alpha D_{pureW}(T)$, $\alpha < 1$. $D_{mob}(T)$ можно также получать, анализируя распределения, подобные Рис. 3.9, результаты чего представлены на Рис. 3.10:

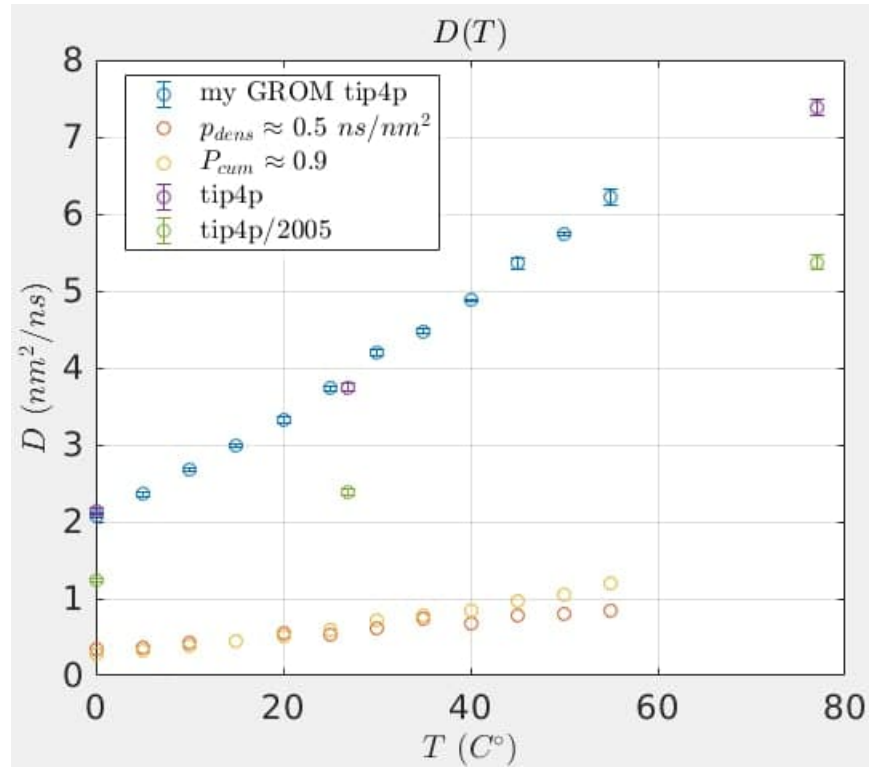


Рис. 3.10. Результаты расчета $D_{mob}(T)$ по двум критериям. 1 – отсечение по пороговому значению плотности вероятности. Такой критерий имеет смысл т.к. по Рис. 3.9 видно, что распределение D молекул воды начинает монотонно падать практически сразу после 0. Если принять монотонность спада $p(D)$, то более строго $D_{dens}(T) = \sup(\{D : p(D(T)) > p_0(T)\})$. На рисунке $p_0 = 0.5 \text{ ns/nm}^2$. 2 – отсечение по кумулятивной вероятности, т.е. взятие за $D_{mob}(T)$ значения D молекулы, обладающей D большим, чем у заданной доли от всех молекул. Более строго $D_{cum}(T) = \sup(\{D : (\sum_{i=1}^N D_i < D) / N < P_{cum}\})$. На рисунке $P_{cum} = 0.9$.

Анализируя два предложенных метода нахождения $D_{mob}(T)$, можно заметить, что в методе отсечения по пороговой плотности вероятности пороговый параметр имеет размерность. Это приводит к тому, что для получения физически-эквивалентных значений $D_{mob}(T)$ необходимо будет изменять значение пороговой плотности вероятности $p_0(T)$. Данная особенность уменьшает ценность данного метода, т.к. цель в данном случае была в формализации и унификации процесса нахождения $D_{mob}(T)$ путем исключения человеческого фактора из алгоритма, в то время как выбор $p_0(T)$ должен делать человек исходя не из четко сформулированных критериев, а из здравого смысла и

вида профиля $p(D)$.

Метод же, основанный на пороговой кумулятивной вероятности, имеет безразмерный пороговый параметр, что является преимуществом. Пользуясь этим методом, можно получить «степень свободы» воды в белке в зависимости от температуры. Разумной мерой «свободности» может являться отношение $D_{mob}/D_{bulk} < 1$. Чем ближе оно к 1, тем свободнее вода. На рисунке Рис. 3.11 представлена зависимости «свободности» от температуры:

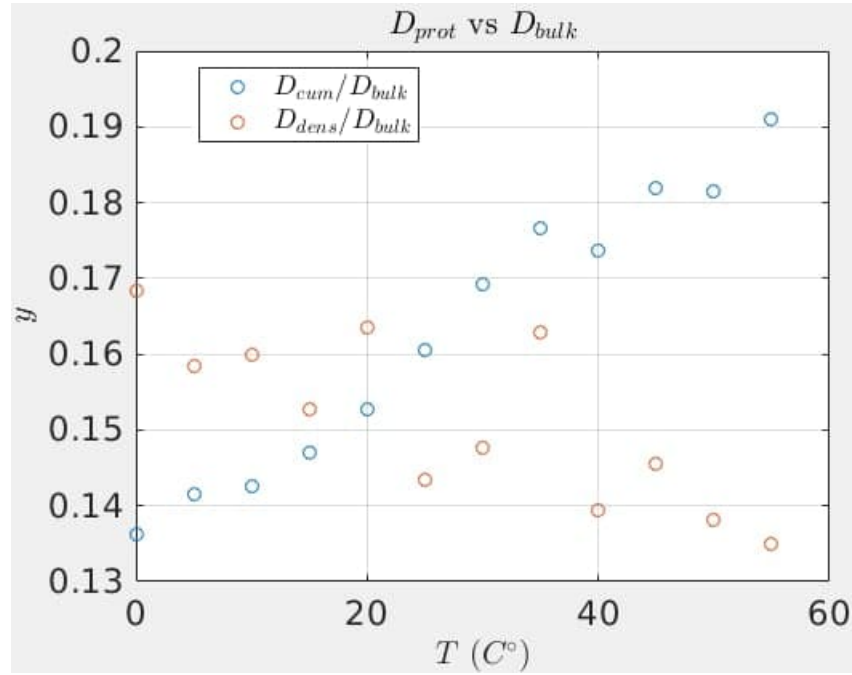


Рис. 3.11. Сравнение средней мобильности воды в системе с белком и в чистом виде. Для нахождения D_{mob} были использованы 2 описанных выше метода с $p_0 = 0.5$ (nm^2/ns) и $P_{cum} = 0.9$.

Значения, основанные на p_0 , падают с ростом T . Это ожидаемо, т.к. отсечка происходит по плотности вероятности, которая должна удовлетворять условию нормировки $\int_0^{D_{bulk}} p(D)dD = 1$. Это условие приводит к тому, что движение точки $D_{dens}(T) = \sup(\{D : p(D(T)) > p_0(T)\})$ вправо (т.е. рост D_{dens}) происходит медленнее чем рост D_{bulk} . Рост же D_{cum}/D_{bulk} говорит о том, что с ростом T вода приближается к свободному состоянию, что опять же физично и ожидаемо.

Анализ динамики распределения $p(D, t)$ также позволяет провести дополнительную верификацию того, что система действительно вышла на равновесие.

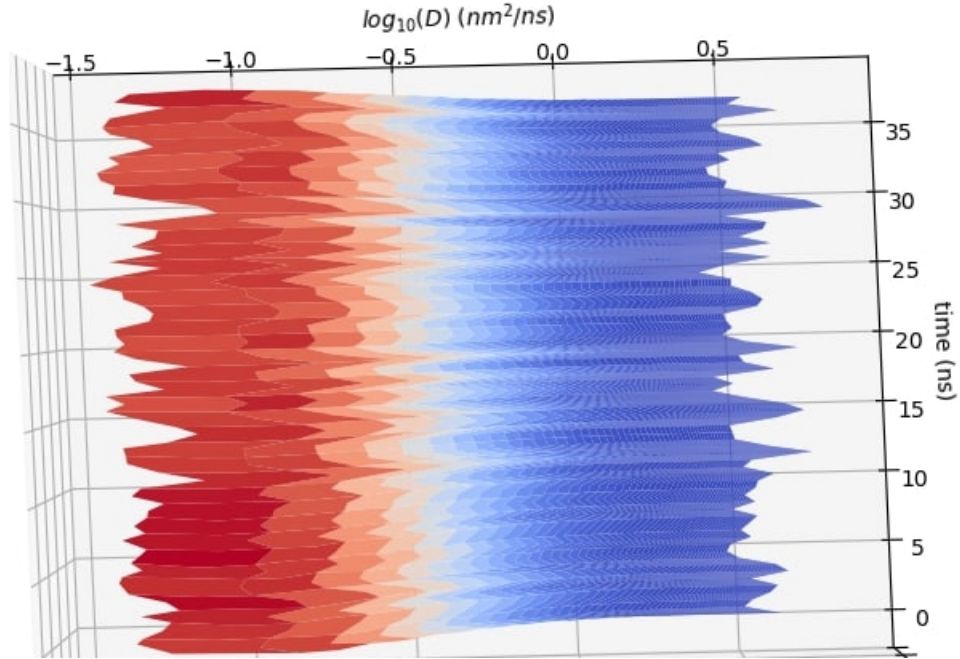


Рис. 3.12. Динамика распределения молекул воды по их коэффициенту диффузии. Видно, что распределение пришло в равновесие.

На Рис. видно, что распределение не изменяется со временем, поэтому можно считать это аргументом в пользу стабильности системы.

3.5.3. Корреляция смещения молекулы со временем движения

Чтобы решить проблему размерности, присутствующую при анализе мобильности через D , было предложено обезразмерить измеряемый наклон $r^2(t)$, т.е. перейти от $D \sim \Delta r^2 / \Delta t$ к коэффициенту корреляции между смещением и временем этого смещения $R(r^2(t), t)$. Это изначально безразмерная величина, поэтому любые отчески по ней так же не будут содержать размерных эффектов, таких как рост D_{bulk} с температурой. Для идеальной диффузии должно выполняться $R \approx 1$, или точнее $0 < 1 - R \ll 1$. Типичное распределение молекул воды по R представлено на Рис. 3.13:

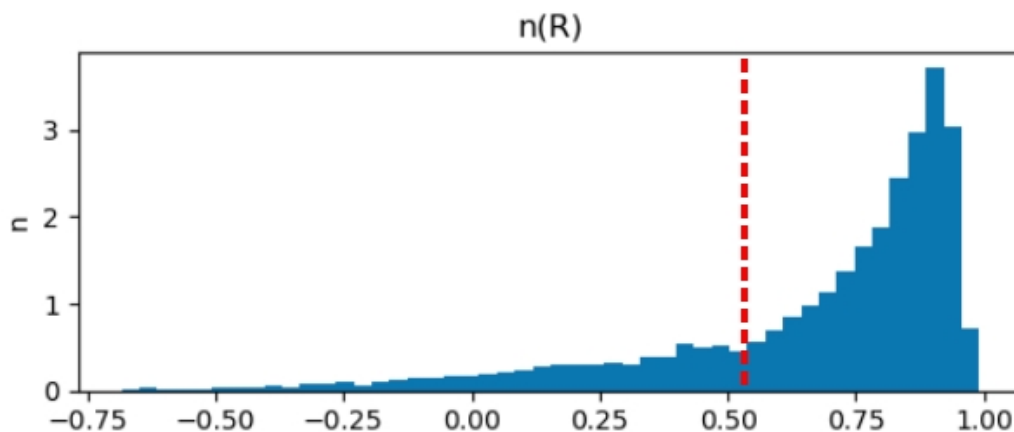


Рис. 3.13. Распределение молекул воды по их корреляции Пирсона $R(r^2(t), t)$. Красным пунктиром отмечено то где человеческий глаз находит изменение характера зависимости распределения.

Возможное наличие точки качественного изменения характера распределения при критическом R_0 , отмеченном на Рис. 3.13 красным пунктиром, дает надежду, что это R_0 может быть использовано для классификации молекул, не зависящей от размерных эффектов. Это может быть сделано, однако использование R как единственной характеристики, отражающей мобильность молекулы, не совсем корректно. Дело в том, что использование R хотя и улучшает ситуацию с проблемой монотонности движения, описанной в 3.5.1, однако не решает ее в достаточной мере, что продемонстрировано на примерах на Рис. 3.14:

На Рис. 3.14 видно, что динамика частиц явно отличается. На графике слева молекула движется относительно равномерно, в то время как справа молекула совершает явные перескоки из одной потенциальной ямы в другую. Однако несмотря на явно различный характер движения данных молекул, они обладают очень близкими $R(r^2(t), t)$. Для решения этой проблемы было предложено перейти к одновременному учету 2 характеристик зависимости $r^2(t)$, одна из которых будет характеризовать активность смещения молекулы, а другая – равномерность движения.

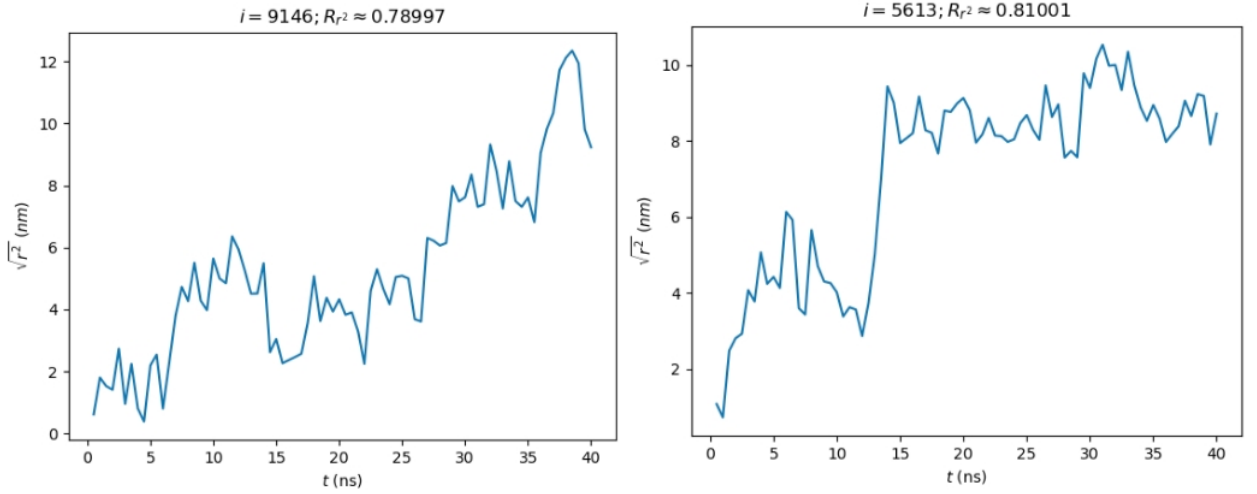


Рис. 3.14. Примеры зависимостей $\sqrt{\Delta r(t)^2}$, демонстрирующие некорректность использования $R(r^2(t), t)$ как единственного критерия мобильности.

3.5.4. Энтропия $r^2(t)$ как мера равномерности

В качестве меры равномерности смещения было предложено использовать энтропию распределения точек $\{r^2(t_i)\}$, определяемую как

$$S_d = - \sum_i P_i \ln(P_i) \quad (3.16)$$

в дискретном случае, и как

$$S_c = - \int_X p(x) \ln[p(x)] dx \quad (3.17)$$

в случае непрерывных распределений.

В данном случае сложность заключалась в выборе между (3.16) и (3.17) применительно к нашему случаю. Дело в том, что использование (3.16) осложнено тем, что физически в каждый момент времени t есть непрерывное распределение молекул воды $p(t, S)$. Но «физически» значит «в бесконечной системе». У нас же количество молекул воды ограничено, поэтому мы лишь оцениваем истинное распределение $p(S)$. Для оценки распределения по конечному количеству измерения необходимо вводить разбиение по оси аргументов,

чтобы построить гистограмму, предположительно аппроксимирующую целевое распределение. Но не ясно, какое разбиение делать для гистограммы, чтобы найти оптимум между количеством точек в каждом интервале и подробностью разрешения оси. Поэтому от выбора разбиения будет зависеть результат, рассчитанный по (3.16).

Использование же (3.17) осложнено 2 моментами. Во-первых эта формула не совсем безразмерна, т.к. мы берем логарифм от размерной величины (плотность вероятности $p(x)$ имеет размерность $1/[x]$, т.е. в нашем случае $1/nm^2$). Во-вторых, у нас опять же имеется лишь конечное количество молекул воды, что приводит к необходимости как-то оценивать распределение. Было предложено оценивать его, накидывая на каждую точку $r^2(t_i)$ гауссовый купол определенной ширины. Данный метод дает разумную оценку распределения, но имеет свободный параметр в виде ширины «гауссового пакета». Его выбор делает

$$\sigma = [\max(\{r^2(t_i)\}) - \min(\{r^2(t_i)\})]/N, \quad (3.18)$$

где $N = |\{r^2(t_i)\}|$ – количество точек по времени. Для поманя такого определения можно обратиться например к случаю свободной диффузии, в котором будет

$$\sigma \sim DT/N = DN\delta t/N = D\delta t, \quad (3.19)$$

где δt – шаг печати кадров по времени.

Если расписать (3.17) в такой постановке, то можно увидеть, что проблема размерности как и проблема выбора σ решаются:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=1}^N \exp\left(-\frac{(x-x_k)^2}{2\sigma^2}\right),$$

$$S = - \int_{\mathcal{R}} p(x) \ln[p(x)] dx \quad (3.20)$$

что упрощается до

$$S = N \ln[\sigma\sqrt{2\pi}] + \sum_{k=1}^N \int_{\mathcal{R}} du \exp\left(-\frac{(u-y_k)^2}{2}\right) \ln \left[\sum_{j=1}^N \exp\left(-\frac{(u-y_j)^2}{2}\right) \right], \quad (3.21)$$

где $y_k = x_k/\sigma$.

В данных формулах x – значения вдоль оси r^2 , т.к. мы рассматриваем плотность распределения $r^2(t)$, спроецированного на ось Y . Тогда $x_k = r^2(t_k)$.

Как видно из (3.21), выражение для S представляется в виде

$$S = S_0(\sigma, N) + S_{rel}(\sigma, N, \{x_k\}). \quad (3.22)$$

В каждом расчете мы знаем σ и N заранее, поэтому можем исключить влияние $S_0(\sigma, N)$. В итоге за энтропию распределения мы будем считать S_{rel} из (3.22). Можно предположить, что эффект размерности не будет давать системного вклада, т.к. для например предельного случая свободной диффузии $x_k \sim kD\delta t/(D\delta t) = k$, что позволяет явно вычислить значения S_{rel} для различных N . Результаты вычисления показаны на Рис. 3.15:

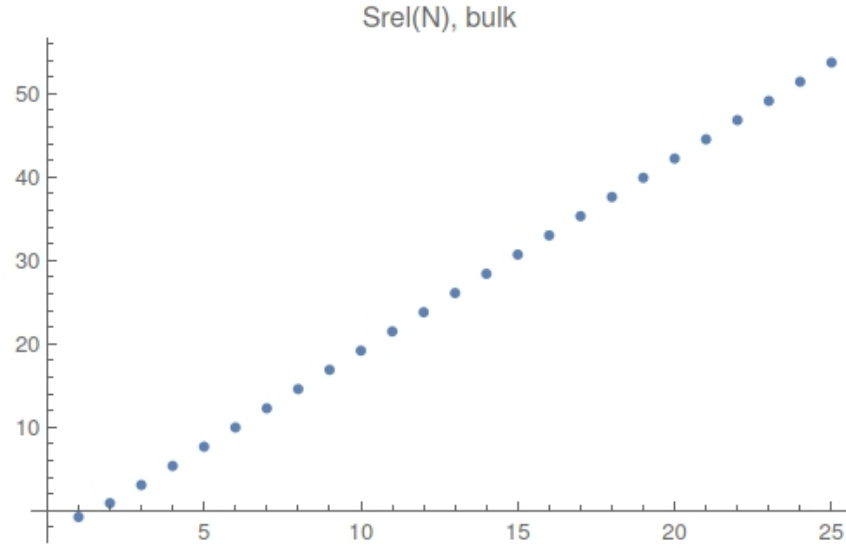


Рис. 3.15. Зависимость $S_{rel,bulk}(N)$ в предположении идеальной свободной диффузии, т.е. для $r^2(t) = 6Dt$.

Видно, что с хорошей точностью $S_{rel,bulk} \approx 2N$, что дает $S_{bulk} \approx N[2 + \ln(D\delta t\sqrt{2\pi})]$. Поэтому в данной работе за итоговую энтропию бралось $S_c = (S - S_0)/N$. Тогда для чистой воды это значения было бы вычислимо и $S_{c,bulk} \sim 2$.

Типичное распределение молекул в пространстве (S_c, R) показано на Рис. 3.16:

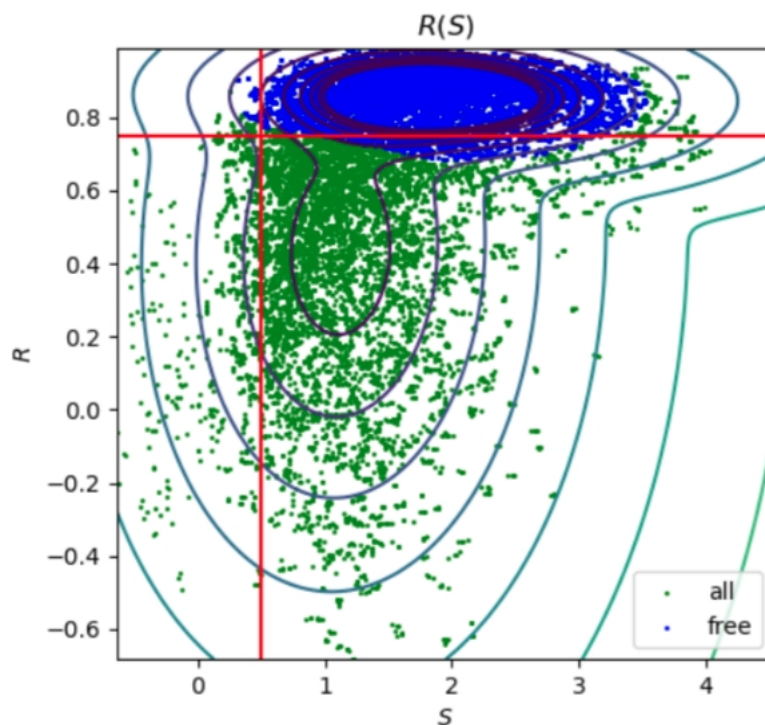


Рис. 3.16. Распределение молекул воды в пространстве S_c, R .

Точки в верхней части распределения смещаются как правило дальше чем остальные, а чем правее точка, тем равномернее происходит это смещение. Исходя из вида 3.16 было предложено 2 способа классификации молекул воды как мобильных. Первый – введение 2 граничных значения R_0, S_0 , и пометка всех молекул с $R > R_0, S > S_0$ как мобильных. Это отражено красными линиями. Второй способ – провели гауссовую кластеризацию и взять все молекулы, относящиеся в верхней правой группе. По виду зависимости понятно, что изначально нужно искать 2 кластера. Этот метод отражен цветом точек.

Далее мы исследуем 2 предложенных метода кластеризации. Для начала, можно посмотреть на динамику доли «свободных» молекул с температурой, что показано на Рис. 3.17:

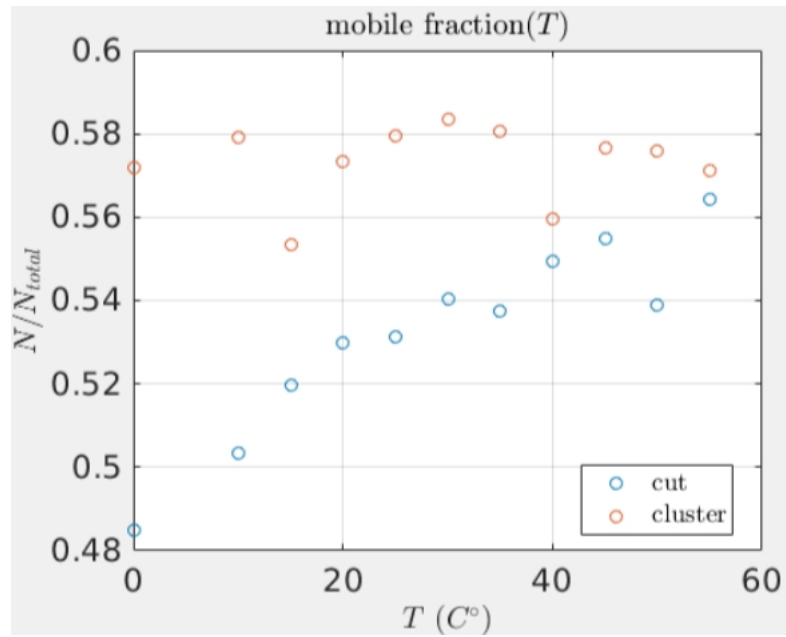


Рис. 3.17. Изменение доли молекул, определяемых как мобильные методами кластеризации и обрезки.

По Рис. 3.17 видно, что размер верхнего кластера не зависит от температуры, в то время как с ростом T все больше и больше молекул переходят границу ($> R_0, > S_0$). Также интересно пронаблюдать динамику движения центра кластера мобильных молекул, показанную на Рис. 3.18:

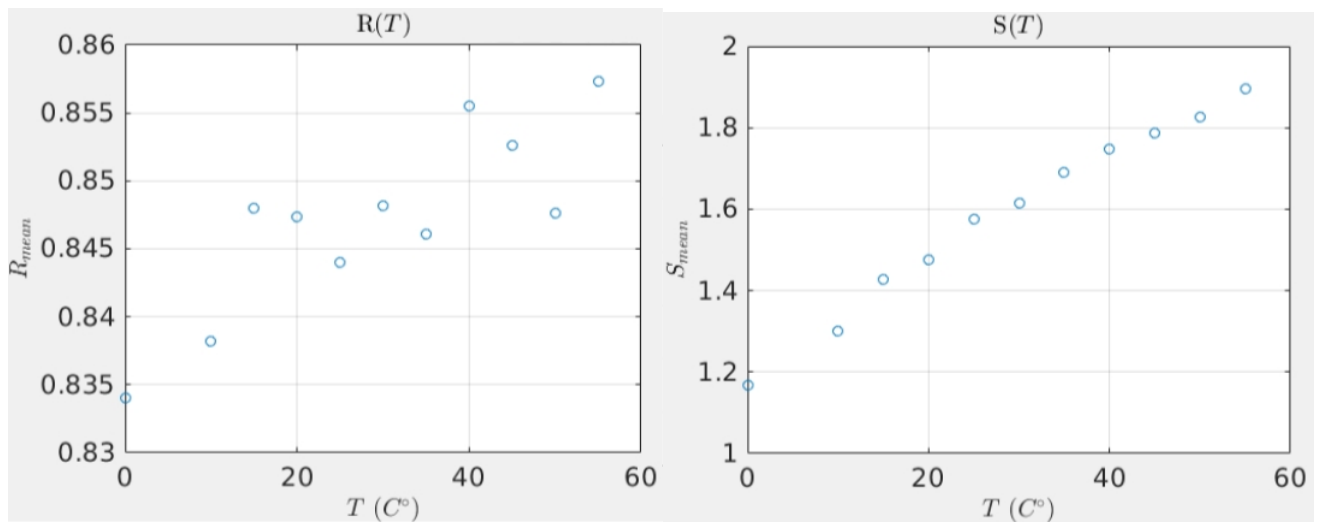


Рис. 3.18. Движение центра кластера мобильных молекул с изменением температуры.

Соединяя данные с Рис. 3.18 и 3.17 можно заключить, что существует кластер молекул, не изменяющийся в размере (Рис. 3.17), но становящийся-

ся более мобильным с ростом температуры (Рис. 3.18). Рост мобильности отражается в основном в стабилизации движения свободных молекул, что ожидаемо, т.к. с ростом температуры потенциальные ямы, мешающие равномерности движения, становятся менее заметны для частиц. Рост же количества мобильных молекул, определяемых методом R_0, S_0 объясняется тем, что при росте T хвосты мобильного кластера несимметрично входят в область ($> R_0, > S_0$), и слева в эту область начинают заползать уже не столь мобильные молекулы.

Поэтому в итоге было принято решение использовать метод гауссовой кластеризации для определения мобильности молекул.

3.5.5. Предположение о природе мобильности

В разделах выше было найдено 2 кластера молекул воды. Молекулы из одного из которых движутся более свободно чем молекулы из другого. Можно предположить, что попадание молекулы в тот или другой кластер может зависеть от степени ее привязанности к белку. Важным механизмом связи воды и белка выступают водородные связи, поэтому они могут играть здесь важную роль. Для оценки доли их вклада в связывание воды можно построить распределение расстояний от молекул воды до белка, что показано на Рис. 3.19:

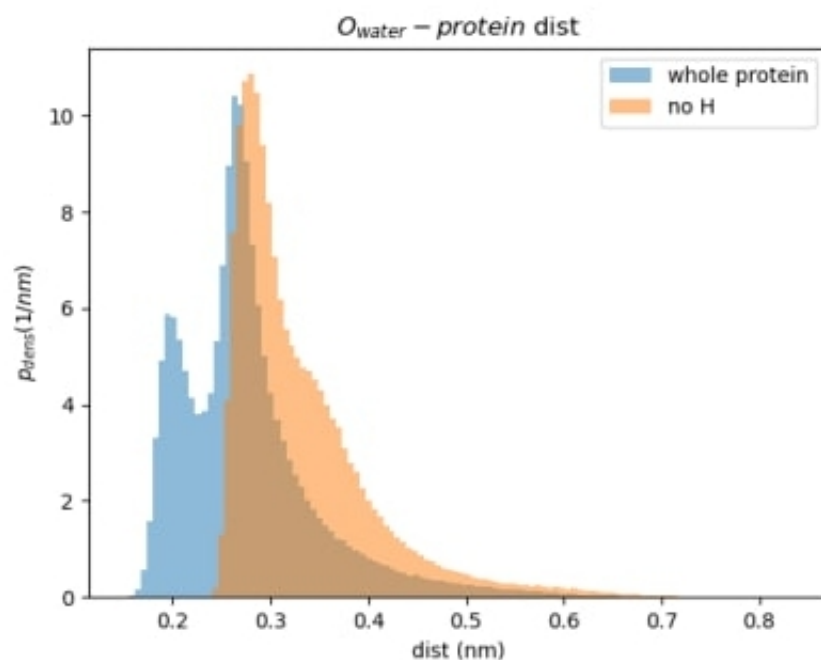


Рис. 3.19. Распределения расстояний от центра атома кислорода в молекуле воды до центра ближайшего атома белка. Оранжевый – ближайшие соседи ищутся среди всех атомов белка кроме водородов. Синий цвет – среди всех атомов белка.

На Рис. 3.19 видно, что при включении водородов в состав белка появляется второй пик в распределении на уровне $\sim 2 \text{ \AA}$, что хорошо согласуется с типичной известной длиной водородной связи.

Преобладание водорода также видно и на распределениях ближайших расстояний по тем элементам, с которыми вода оказалась ближайшим соседом. Они представлены на Рис. 3.20:

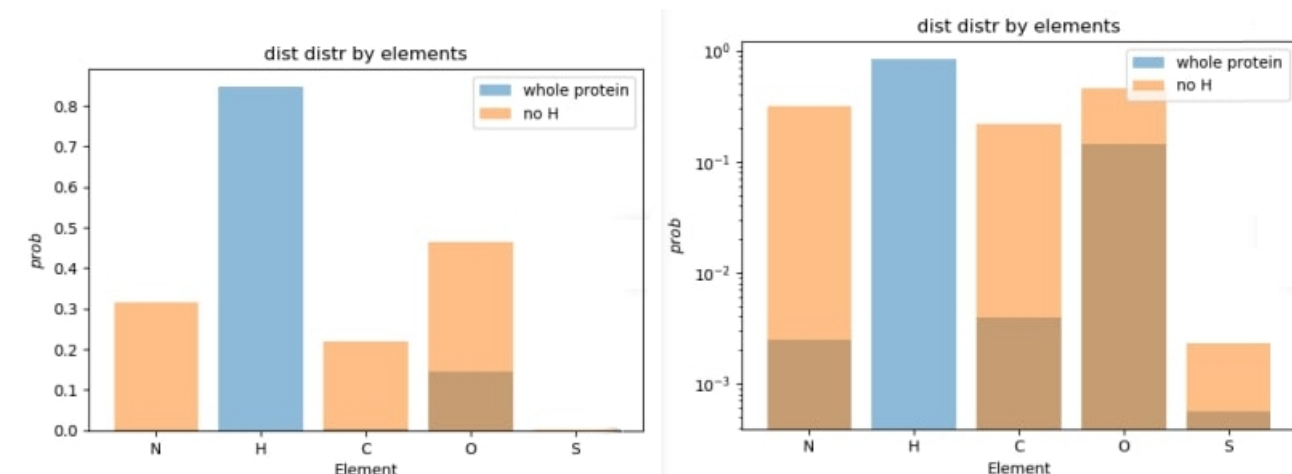


Рис. 3.20. Распределения ближайших расстояний элементам, с которыми эти расстояния созданы. На левом рисунке видно, что после исключения водорода уже не наблюдается такого явного преобладания соседства с каким-либо одним элементом, что объясняет отсутствие других (кроме длины водородной связи) явных пиков на Рис. 3.19.

Наконец Рис. 3.19 как мы считаем позволяет объяснить наличие 2 кластеров на Рис. 3.16 и также поясняет смену режимов динамики молекул воды, показанную еще на Рис. 3.8.

Итак, можно предположить, что вода, находящаяся в левом пике на расстоянии $\sim 2 \text{ \AA}$ от некоторого водорода в белке будет связана водородной связью и не будет свободно диффундировать. Это порождает участки плато на $r^2(t)$ на Рис. 3.8. Молекулы же, не связанные водородной связью, будут двигаться более свободно, что отвечает более равномерным участкам $r^2(t)$ на Рис. 3.8. О менее связанном состоянии свидетельствует также и отсутствие явного преобладания соседства с каким-либо одним элементом после исключения водорода, т.к. если молекула воды просто случайно движется в пространстве, то она будет относительно случайно оказываться ближайшей к разным атомам белка в соответствии с их долей на поверхности. Сравнение Рис. 3.20 с долей атомов на поверхности белка является одной из возможных будущих микро-задач.

Наличие же 2 кластеров на Рис. 3.16 можно попытаться объяснить тем,

что некоторые молекулы чаще оказываются связаны водородной связью, что мешает их подвижности. Для проверки данного предположения в будущем можно наложить распределение молекул воды по 2-м пикам на Рис. 3.19 на распределение Рис. 3.20.

3.6. 2-фазное моделирование для учета влажности

Для воспроизведения контакта белок-пар предложено провести 2-фазное моделирование, аналогичное тому что используется для определения фазового равновесия. В одном из направлений периодическая ячейка увеличивается в размере на расстояние, достаточное, чтобы белковые слои, которые по факту все еще находятся в одной системе, взаимодействовали пренебрежимо мало. По оставшимся 2 осям система не изменяется. В итоге моделируется квази-бесконечный тонкий слой белкового кристалла. Вообще говоря, в рассматриваемом случае действительно будет достигаться фазовое равновесие между водой в белке и в газообразном состоянии. Подобные расчеты позволят более прямо воспроизвести эксперимент [9] и использовать (1.34) и (1.36) для определения K . Они также позволят найти параметры решетки белкового кристалла, естественные для данной влажности.

3.6.1. Ограничения малостью системы

Симуляция в полном NPT ансамбле физически правильна в данной ситуации, т.к. в пределе большой системы (т.е. малых флуктуаций) и большого времени усреднения заранее заданное количество воды в системы бы перераспределилось по следующей схеме:

1. Заданная температура определяет давление насыщенного пара для данной системы, т.е. для контакта влажный белок – пар при данной соле-

ности воды.

2. Заданное количество воды, изначально находящееся только в белке, определит объем, занимаемый паром:

2.1. Если целевое давление выше чем давление насыщенного пара, то баростат сожмет всю воду в жидкое состояние, что полностью убережет пар из системы. Лишняя вода будет просто выпадать жидким слоем на поверхностях белкового слоя.

2.2. Если целевое давление ниже чем давление насыщенного пара, то пар, постепенно испаряясь при данной температуре, начнет преодолевать давление баростата. Это приведет к увеличению объема ячейки, что в свою очередь уменьшит давление пара. Процесс будет повторяться пока в белке не останется количество воды, естественное для данной влажности.

К сожалению, на практике такая схема не реализуется. При полном *NPT*-ансамбле размер ячейки, в котором происходит контакт сред, просто схлопывается, пока не останется упрется в белок, что видно на Рис. 3.21:

Из-за эффекта схлопывания вакуума, показанного на Рис. 3.21, было предложено отключить баростат по оси раздела сред. Физически это может привести к тому, что при добавлении слишком большого количества воды она при выходе из белка будет конденсироваться на его поверхности. В реальности данный эффект отсутствует, а в расчете он может быть нивелирован подбором изначального количества воды в системе, которой должно быть недостаточно для заполнения всего вакуума насыщенным паром + насыщение белка водой при данной температуре.

Еще один эффект, тоже возможно порожденный малостью системы, выражается в отсутствии единых стабильных значений постоянных кристалли-

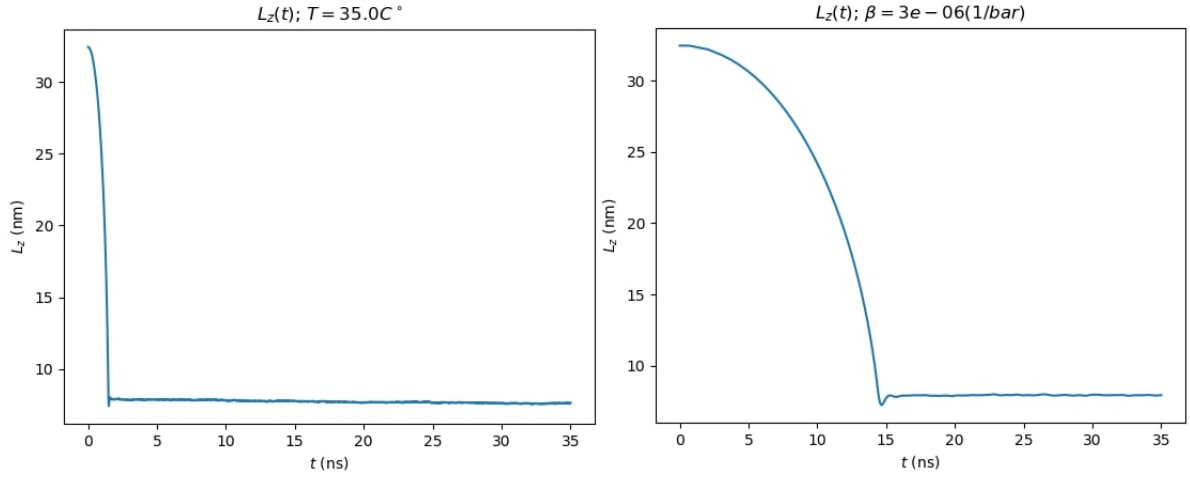


Рис. 3.21. Эволюция L_z – размера ячейки по оси, нормальной к болковому слою – при различных параметрах термостата. На правом рисунке, при получении которого баростат был искусственно ослаблен по оси Z , видно, что даже при достаточном для испарения времени баростат все равно схлопывает вакуум.

ческой решетки. При проведении симуляций, отличающихся только значениями инициализатора случайных чисел, используемого например при генерации начальных скоростей, система может прийти к разным итоговым состояниям, что продемонстрировано на Рис. 3.22:

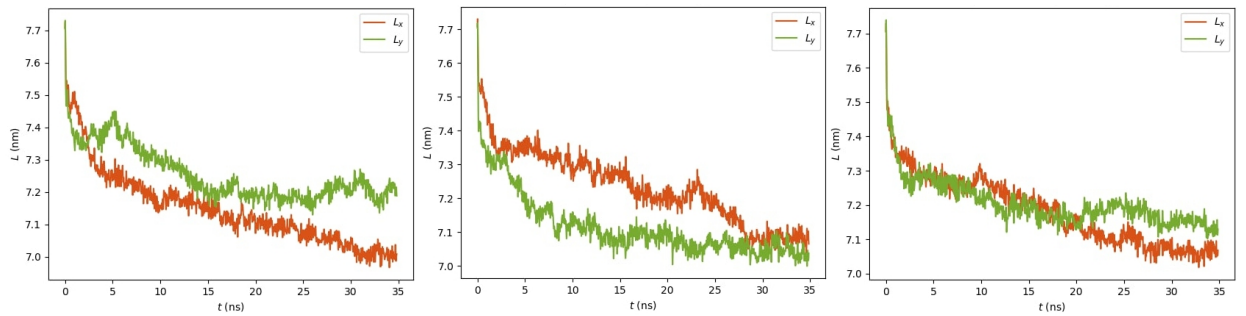


Рис. 3.22. Эволюция размеров ячейки при одинаковых физических параметрах но различных начальных инициализациях генератора случайных чисел. Оранжевый – L_x , Зеленый – L_y . Размеры могут как эволюционировать близко, а потом разойтись (рисунок слева), так и наоборот различаться в начале, но потом сойтись (центральный рисунок), так и все время эволюционировать близко (рисунок справа).

Скорее всего неединственность итоговой конформации связана с тем,

что при малом размере системы флуктуации структуры происходят эффективно одновременно по всему объему (или в данном случае по всей площади) белкового слоя. А такие глобальные флуктуации могут оказаться более устойчивыми из-за крупномасштабного самодействия, отсутствующего в реальной системе. Из-за данной неединственности затруднительно говорить о зависимости равновесных значений объема или отдельных размера ячейки от температуры. Предположительно при увеличении расчетной ячейки на большее количество элементарных ячеек белка эффект должен исчезнуть, ли по крайней мере должна появиться какая-то одна явно преобладающая конформация. Это является одним из направлений дальнейшего исследования. Данный эффект может быть важен непосредственно для определения равновесных размеров ячейки при заданной влажности.

3.6.2. Расчет относительной влажности

Размер ячейки играет ключевую роль в обоих прямых методах расчета K , т.к. динамический метод опирается на флуктуации объема, а в статическом происходит одномоментное изменение объема. Для расчетов же по (1.34) точное значение размеров ячейки не играет очевидно-важной роли, т.к. в (1.34) явно участвуют лишь относительная влажность и массовая доля воды в системе. К тому же данный метод использовался в эксперименте [9].

Опишем алгоритм построения изотерм, аналогичных Рис. 1.4. Начальными данными является температура системы, изначальное количество воды, содержащейся на старте симуляции только внутри белка, и толщина слоя вакуума. Был выбран слой вакуума в 25 нм., т.к. обычно белки заметно влияют друг на друга если находятся ближе ~ 10 нм..

По молекулярно-динамической траектории строится эволюция профиля плотности воды по оси Z , т.е. по нормали к границе раздела сред. Типичный

профиль в установившемся состоянии изображен на Рис. 3.23:

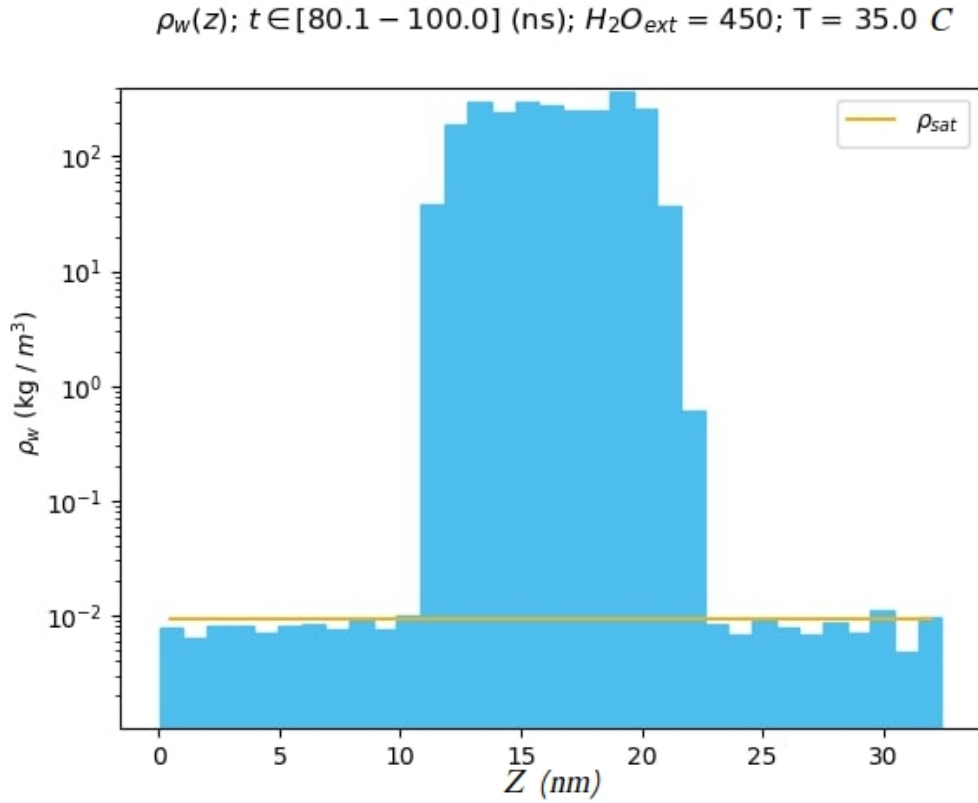


Рис. 3.23. Типичный профиль плотности молекул воды в установившемся состоянии. Усреднение проводилось по отрезку траектории в 20 нс. В систему добавлялось по 450 молекул воды на элементарную ячейку, содержащую 8 цепей лизоцима, к 1656 молекулам на ячейку, уже имеющимся на рентгеновских данных.

Отметим, что на подобных профилях видно формирование слоя воды, которая вышла из белка но уже не смогла поднимать плотность пара выше плотности насыщенного при данной температуре, что продемонстрировано на Рис. 3.24.

Плотность насыщенного пара при данной температуре, отмеченная на профилях плотности выше и используемая далее для построения изотерм, рассчитана фитированием функции вида

$$\rho_{sat}(T) = \exp \left(a - \frac{b}{T} \right) T^{-c}, \quad (3.23)$$

где a , b , c – подбираемые параметры, на данных зависимости плотно-

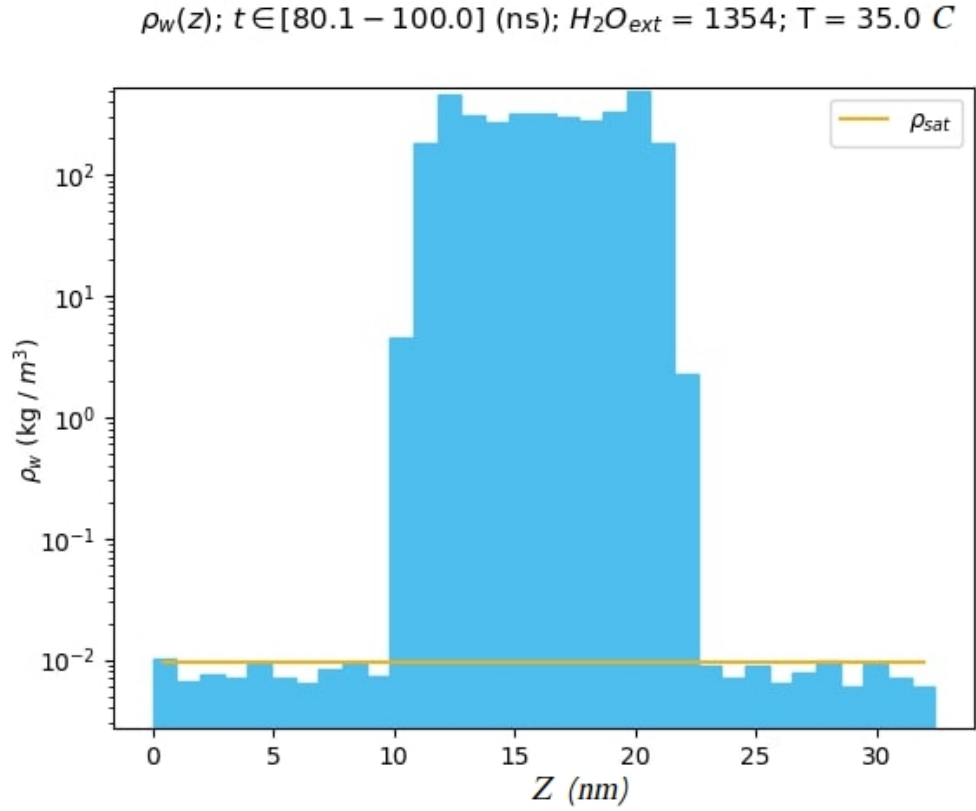


Рис. 3.24. Профиль аналогичный Рис. 3.23, но добавляется не 450 молекул воды на ячейку, а 1354. Можно видеть увеличения плотности воды по краям белкового слоя, которые свидетельствуют о конденсации воды, не влезшей ни в белок ни в вакуум в газообразном состоянии.

сти насыщенного пара воды tip4p/2005 в контакте с чистой водой [43]. Вид зависимости (3.23) получается из уравнения Клапейрона-Клаузиуса при пренебрежении удельным объемом жидкости и принятии модели идеального газа для пара и постоянства теплоемкостей газа и жидкости на исследуемом диапазоне температур.

Общий вид уравнения Клапейрона-Клаузиуса записывается как

$$\frac{dP}{dT} = \frac{q}{T(v_1 - v_2)}, \quad (3.24)$$

где v_1, v_2 – удельные (по массе) объемы фаз, q – удельная теплота перехода. В силу $\rho_{\text{vap}} \ll \rho_{\text{liq}}$ можно сказать что $v_{\text{vap}}/v_{\text{liq}} = \rho_{\text{liq}}/\rho_{\text{vap}} \gg 1$. Далее, принимая модель идеального газа для пара, получаем $v_{\text{vap}} = V_{\text{vap}}/m_{\text{vap}} =$

$\nu RT/P\nu\mu = RT/P\mu$, получаем

$$\frac{dP}{dT} = \frac{Pq\mu}{RT^2}, \quad (3.25)$$

где μ – молярная масса вещества, претерпевающего переход, q – удельная теплота перехода. Она складывается из внутренней энергии и работы по расширению пара при нагревании:

$$q = u_{vap} - u_{liq} + P(v_{vap} - v_{liq}). \quad (3.26)$$

Принимая $u = u_1 + \int_{T_1}^T c_v dT$, $c_v = c_P - R/\mu$ получаем

$$\frac{dP}{dT} = \frac{P\mu[q_1 + (c_{P,vap} - c_{liq})(T - T_1)]}{RT^2}, \quad (3.27)$$

откуда интегрированием по T получается (3.23).

На профилях плотности видно, что можно считать плотностью пара среднее значение плотности по области $z \in [0; 9] \cup [25; 32]$ нм. Таким образом были получены зависимости плотности пара от времени для 2-фазной системы. Было обнаружено, что вычисление изотерм, аналогичных Рис. 1.4, может требовать заметно больших времен релаксации чем вычисление K прямым методом. Дело в том, что в системе происходят процессы релаксации с разными характерными временами. Обычно для стабилизации плотности пара при тощине слоя вакуума в 25 нм. требовалось около 30 нс., что продемонстрировано на Рис. 3.25:

Однако оказывается, что в системе происходят и заметно более долгие процессы, т.к. при увеличении времени симуляции на порядок мы видим изменение плотности пара, что изображено на Рис. 3.26:

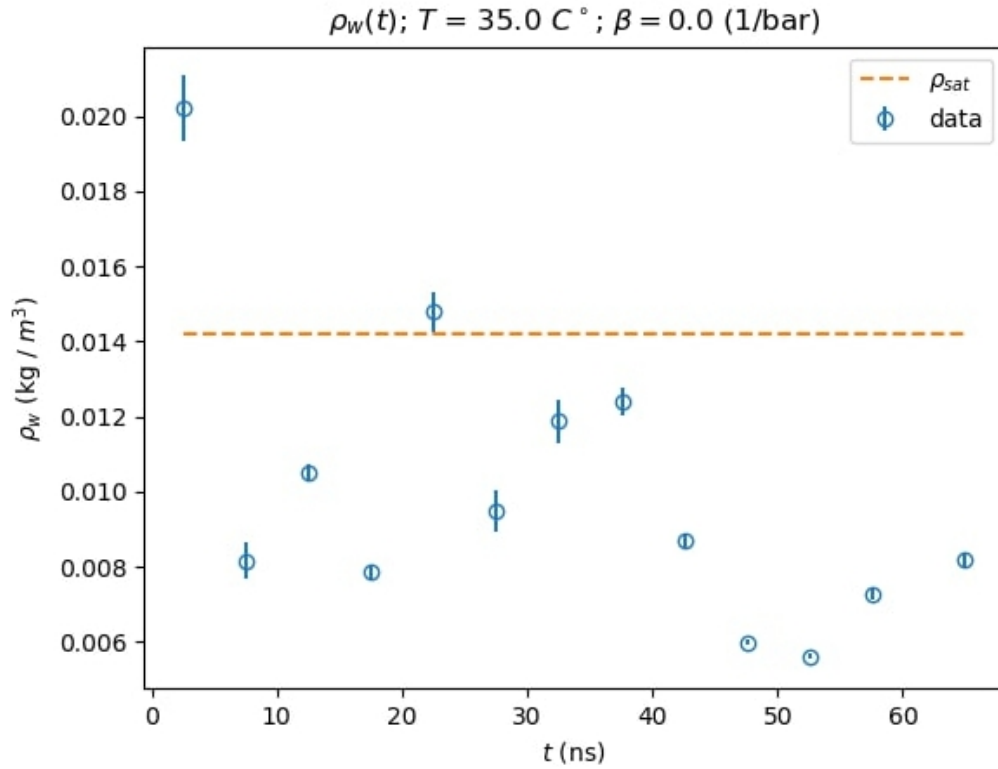


Рис. 3.25. Зависимость плотности пара от времени для $T = 35\text{ }^{\circ}\text{C}$. Время усреднения для каждой точки – 5 нс. Видно, что примерно 30 нс. достаточно для стабилизации зависимости.

3.6.3. Расчет K из баланса хим. потенциала

Учитывая особенности вычисления относительной влажности, описанные в предыдущем разделе, мы наконец можем построить изотермы вида (1.34) и найти по ним значения K белка.

Вычисление K по (1.34) требует проведения нескольких расчетов длительностью по несколько сотен нс. каждый, поэтому на данный момент результатов, полученных этим методом не много. Результаты расчетов для $T = 18\text{ }^{\circ}\text{C}$ представлены на Рис. (3.27).

С одной стороны расхождение расчетных значений с экспериментальными все еще в несколько раз. Но с другой стороны результат расчетов по данному методу ошибается заметно меньше чем оба испытанных прямых метода – ошибка в 5 раз против 20 раз. Тут стоит упомянуть, что в самом экспери-

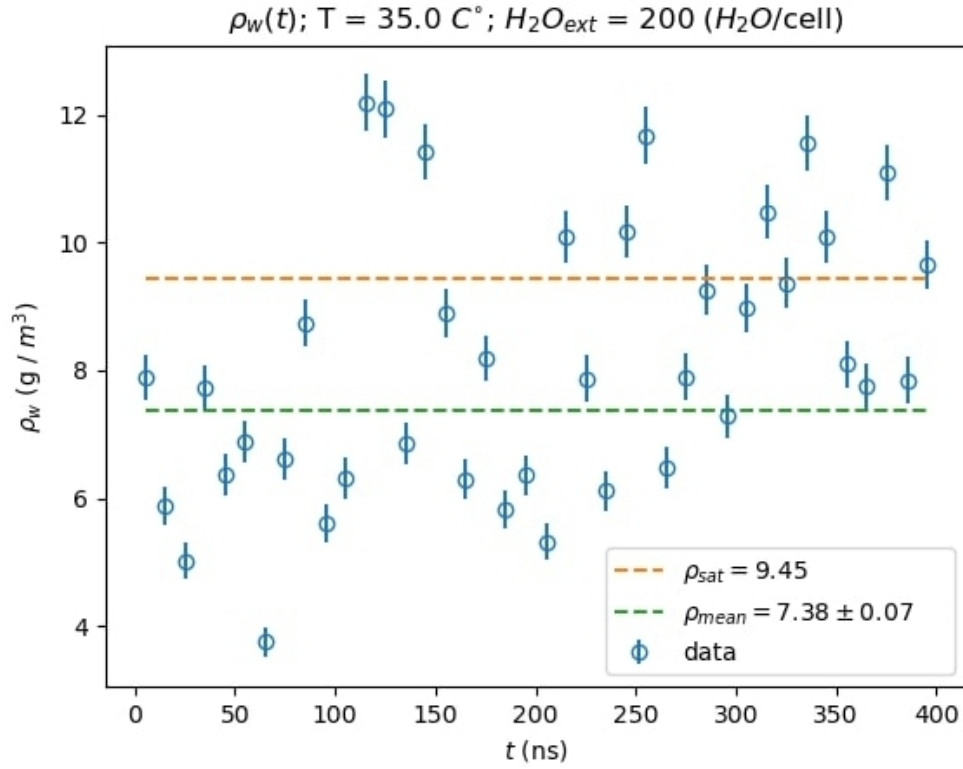


Рис. 3.26. Зависимость плотности пара от времени для $T = 35\text{ }^{\circ}\text{C}$. Время усреднения для каждой точки – 10 нс. Видно, что примерно на моменте 250 нс. начинается качественное изменение в системе.

менте [9] авторы указывают на неизвестный параметр – долю атомов белка, участвующих в смешивании с водой. Вариация этого параметра в пределах, выбранных авторами эксперимента, приводит к изменению получаемого K до 2 раз.

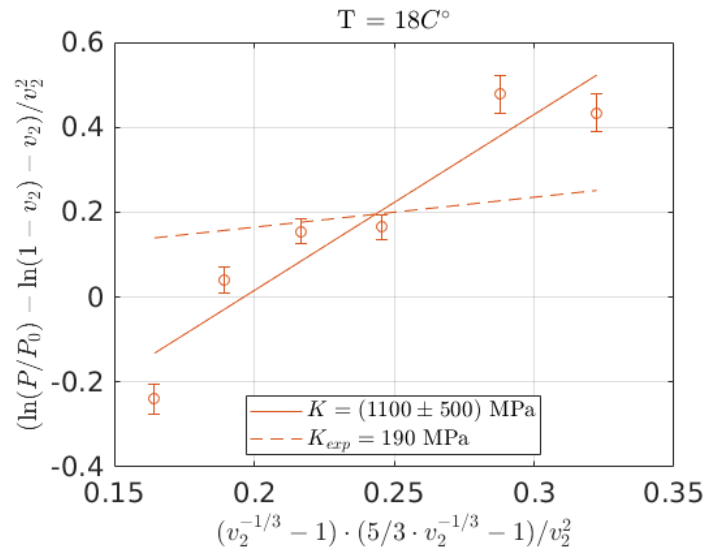


Рис. 3.27. Изотерма вида (1.34), рассчитанная для $T = 18 C^\circ$. Линейная аппроксимация на всем интервале делает потому, что интервал изначально подбирался так, чтобы на экспериментальной зависимости Рис. 1.4 при данных массовых долях воды зависимость (1.34) действительно была линейной. Для наглядности нанесена также линия с наклоном, соответствующем экспериментальному значению K при данной температуре [9].

Заключение

В ходе выполнения данной работы:

- Сформулирован алгоритм подготовки кристалла к MD-расчету.
- Проверены сходимости по параметрам системы.
- Реализовано 2 независимых прямых метода расчета K системы кристалл+вода. K , рассчитанные по этим 2 методам, согласуются между собой, но плохо согласуются с экспериментом.
- Вода кластеризована по мобильности на 2 группы.
- Дано возможное объяснение наличия 2 различных по мобильности групп.
- Протестировано влияние различных моделей воды на мобильность воды и остальные параметры системы.
- Лизоцим стабилизирован при различных влажностях с использованием 2-фазного моделирования.
- Найден приближенный диапазон влажности, оптимальный для расчета модуля упругости.

В дальнейшем планируется провести расчет K при различных влажностях. Увеличение расчетной ячейки так же может улучшить совпадение с экспериментом. Возможно проведение расчетов с другими параметризациями белка для улучшения совпадения с экспериментом.

Список литературы

1. Brunauer S., Emmett P. H., Teller E. Adsorption of Gases in Multimolecular Layers // [Journal of the American Chemical Society](#). 1938. — feb. Vol. 60, no. 2. P. 309–319. URL: <https://doi.org/10.1021/ja01269a023>.
2. Kuntz I., Kauzmann W. [Hydration of Proteins and Polypeptides](#) // [Advances in Protein Chemistry](#). Elsevier, 1974. P. 239–345. URL: [https://doi.org/10.1016/s0065-3233\(08\)60232-6](https://doi.org/10.1016/s0065-3233(08)60232-6).
3. Rowen J. W., Simha R. Interaction of Polymers and Vapors. // [The Journal of Physical and Colloid Chemistry](#). 1949. — jun. Vol. 53, no. 6. P. 921–930. URL: <https://doi.org/10.1021/j150471a019>.
4. Kauzmann W. [Some Factors in the Interpretation of Protein Denaturation](#) // [Advances in Protein Chemistry](#). Elsevier, 1959. P. 1–63. URL: [https://doi.org/10.1016/s0065-3233\(08\)60608-7](https://doi.org/10.1016/s0065-3233(08)60608-7).
5. Rupley J. A., Careri G. [Protein Hydration and Function](#) // [Advances in Protein Chemistry](#). Elsevier, 1991. P. 37–172. URL: [https://doi.org/10.1016/s0065-3233\(08\)60197-7](https://doi.org/10.1016/s0065-3233(08)60197-7).
6. Levy Y., Onuchic J. N. WATER MEDIATION IN PROTEIN FOLDING AND MOLECULAR RECOGNITION // [Annual Review of Biophysics and Biomolecular Structure](#). 2006. — jun. Vol. 35, no. 1. P. 389–415. URL: <https://doi.org/10.1146/annurev.biophys.35.040405.102134>.
7. Frauenfelder H., Chen G., Berendzen J. et al. A unified model of protein dynamics // [Proceedings of the National Academy of Sciences](#). 2009. — feb. Vol. 106, no. 13. P. 5129–5134. URL: <https://doi.org/10.1073/pnas.0900336106>.
8. Daniel R., Dunn R., Finney J., Smith J. The Role of Dynamics in Enzyme Activity // [Annual Review of Biophysics and Biomolecular Structure](#). 2003. — jun. Vol. 32, no. 1. P. 69–92. URL: <https://doi.org/10.1146/annurev>.

- biophys.32.110601.142445.
9. Wang H.-J., Kleinhammes A., Tang P. et al. Temperature dependence of lysozyme hydration and the role of elastic energy // *Physical Review E*. 2011. — mar. Vol. 83, no. 3. URL: <https://doi.org/10.1103/physreve.83.031924>.
 10. Careri G., Gratton E., Yang P.-H., Rupley J. A. Correlation of IR spectroscopic, heat capacity, diamagnetic susceptibility and enzymatic measurements on lysozyme powder // *Nature*. 1980. — apr. Vol. 284, no. 5756. P. 572–573. URL: <https://doi.org/10.1038/284572a0>.
 11. Mao S., Kleinhammes A., Wu Y. NMR study of water adsorption in single-walled carbon nanotubes // *Chemical Physics Letters*. 2006. — apr. Vol. 421, no. 4-6. P. 513–517. URL: <https://doi.org/10.1016/j.cplett.2006.02.011>.
 12. Wang H.-J., Xi X.-K., Kleinhammes A., Wu Y. Temperature-Induced Hydrophobic-Hydrophilic Transition Observed by Water Adsorption // *Science*. 2008. — oct. Vol. 322, no. 5898. P. 80–83. URL: <https://doi.org/10.1126/science.1162412>.
 13. Flory P. J. Thermodynamics of High Polymer Solutions // *The Journal of Chemical Physics*. 1942. — jan. Vol. 10, no. 1. P. 51–61. URL: <https://doi.org/10.1063/1.1723621>.
 14. Staverman A. J. The cohesive energy of liquid mixtures I // *Recueil des Travaux Chimiques des Pays-Bas*. 1937. Vol. 56, no. 9. P. 885–890. URL: <https://doi.org/10.1002/rec1.19370560908>.
 15. Flory P. *Principles of Polymer Chemistry*. Baker lectures 1948. Cornell University Press, 1953. ISBN: 9780801401343. URL: <https://books.google.com/books?id=CQ0EbEkT5ROC>.
 16. Diakova G., Goddard Y. A., Korb J.-P., Bryant R. G. Changes in protein structure and dynamics as a function of hydration from ^1H second moments //

- Journal of Magnetic Resonance*. 2007. — dec. Vol. 189, no. 2. P. 166–172. URL: <https://doi.org/10.1016/j.jmr.2007.09.005>.
17. Abragam A. *The Principles of Nuclear Magnetism*. International Series of Monogr. Clarendon Press, 1983. ISBN: 9780198520146. URL: https://books.google.ru/books?id=9M8U_JK7K54C.
 18. Millero F. J., Ward G. K., Chetirkin P. Partial specific volume, expansibility, compressibility, and heat capacity of aqueous lysozyme solutions. // *Journal of Biological Chemistry*. 1976. — jul. Vol. 251, no. 13. P. 4001–4004. URL: [https://doi.org/10.1016/s0021-9258\(17\)33347-1](https://doi.org/10.1016/s0021-9258(17)33347-1).
 19. Gorelov A., Morozov V. Mechanical denaturation of globular protein in the solid state // *Biophysical Chemistry*. 1987. — dec. Vol. 28, no. 3. P. 199–205. URL: [https://doi.org/10.1016/0301-4622\(87\)80090-x](https://doi.org/10.1016/0301-4622(87)80090-x).
 20. Sauter C., Otalora F., Gavira J.-A. et al. STRUCTURE OF TETRAGONAL HEN EGG WHITE LYSOZYME AT 0.94 Å FROM CRYSTALS GROWN BY THE COUNTER-DIFFUSION METHOD. 2001. — aug. URL: <https://doi.org/10.2210/pdb1iee/pdb>.
 21. Sauter C., Otálora F., Gavira J.-A. et al. Structure of tetragonal hen egg-white lysozyme at 0.94 Å from crystals grown by the counter-diffusion method // *Acta Crystallographica Section D Biological Crystallography*. 2001. — jul. Vol. 57, no. 8. P. 1119–1126. URL: <https://doi.org/10.1107/s0907444901008873>.
 22. Pettersen E. F., Goddard T. D., Huang C. C. et al. UCSF Chimera?A visualization system for exploratory research and analysis // *Journal of Computational Chemistry*. 2004. Vol. 25, no. 13. P. 1605–1612. URL: <https://doi.org/10.1002/jcc.20084>.
 23. Dolinsky T. J., Nielsen J. E., McCammon J. A., Baker N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations // *Nucleic Acids Research*. 2004. — jul. Vol. 32, no. Web Server.

- P. W665–W667. URL: <https://doi.org/10.1093/nar/gkh381>.
24. Dolinsky T. J., Czodrowski P., Li H. et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations // [Nucleic Acids Research](#). 2007. — may. Vol. 35, no. Web Server. P. W522–W525. URL: <https://doi.org/10.1093/nar/gkm276>.
 25. Olsson M. H. M., Søndergaard C. R., Rostkowski M., Jensen J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions // [Journal of Chemical Theory and Computation](#). 2011. — jan. Vol. 7, no. 2. P. 525–537. URL: <https://doi.org/10.1021/ct100578z>.
 26. Landau L., Lifshitz E. [Statistical Physics: Volume 5](#). No. v. 5. Elsevier Science, 2013. ISBN: 9780080570464. URL: <https://books.google.ru/books?id=VzgJN-XPTRsC>.
 27. Abascal J. L. F., Vega C. A general purpose model for the condensed phases of water: TIP4P/2005 // [The Journal of Chemical Physics](#). 2005. — dec. Vol. 123, no. 23. P. 234505. URL: <https://doi.org/10.1063/1.2121687>.
 28. Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., Hermans J. [Interaction Models for Water in Relation to Protein Hydration](#) // The Jerusalem Symposia on Quantum Chemistry and Biochemistry. Springer Netherlands, 1981. P. 331–342. URL: https://doi.org/10.1007/978-94-015-7658-1_21.
 29. Berweger C. D., van Gunsteren W. F., Müller-Plathe F. Force field parametrization by weak coupling. Re-engineering SPC water // [Chemical Physics Letters](#). 1995. — jan. Vol. 232, no. 5-6. P. 429–436. URL: [https://doi.org/10.1016/0009-2614\(94\)01391-8](https://doi.org/10.1016/0009-2614(94)01391-8).
 30. Berendsen H. J. C., Grigera J. R., Straatsma T. P. The missing term in effective pair potentials // [The Journal of Physical Chemistry](#). 1987. — nov. Vol. 91, no. 24. P. 6269–6271. URL: <https://doi.org/10.1021/j100308a038>.

31. Jorgensen W. L., Chandrasekhar J., Madura J. D. et al. Comparison of simple potential functions for simulating liquid water // [The Journal of Chemical Physics](#). 1983. — jul. Vol. 79, no. 2. P. 926–935. URL: <https://doi.org/10.1063/1.445869>.
32. Neria E., Fischer S., Karplus M. Simulation of activation free energies in molecular systems // [The Journal of Chemical Physics](#). 1996. — aug. Vol. 105, no. 5. P. 1902–1921. URL: <https://doi.org/10.1063/1.472061>.
33. Berendsen H. J. C. [Transport Properties Computed by Linear Response through Weak Coupling to a Bath](#) // [Computer Simulation in Materials Science](#). Springer Netherlands, 1991. P. 139–155. URL: https://doi.org/10.1007/978-94-011-3546-7_7.
34. Nosé S. A molecular dynamics method for simulations in the canonical ensemble // [Molecular Physics](#). 1984. — jun. Vol. 52, no. 2. P. 255–268. URL: <https://doi.org/10.1080/00268978400101201>.
35. Hoover W. G. Canonical dynamics: Equilibrium phase-space distributions // [Physical Review A](#). 1985. — mar. Vol. 31, no. 3. P. 1695–1697. URL: <https://doi.org/10.1103/physreva.31.1695>.
36. Martyna G. J., Tuckerman M. E., Tobias D. J., Klein M. L. Explicit reversible integrators for extended systems dynamics // [Molecular Physics](#). 1996. — apr. Vol. 87, no. 5. P. 1117–1157. URL: <https://doi.org/10.1080/00268979600100761>.
37. Cooke B., Schmidler S. C. Preserving the Boltzmann ensemble in replica-exchange molecular dynamics // [The Journal of Chemical Physics](#). 2008. — oct. Vol. 129, no. 16. P. 164112. URL: <https://doi.org/10.1063/1.2989802>.
38. Martyna G. J., Klein M. L., Tuckerman M. Nosé–Hoover chains: The canonical ensemble via continuous dynamics // [The Journal of Chemical Physics](#). 1992. — aug. Vol. 97, no. 4. P. 2635–2643. URL: <https://doi.org/10.1063/1.463940>.

39. Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F. et al. Molecular dynamics with coupling to an external bath // [The Journal of Chemical Physics](#). 1984. — oct. Vol. 81, no. 8. P. 3684–3690. URL: <https://doi.org/10.1063/1.448118>.
40. Parrinello M., Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method // [Journal of Applied Physics](#). 1981. — dec. Vol. 52, no. 12. P. 7182–7190. URL: <https://doi.org/10.1063/1.328693>.
41. Abraham M. J., Murtola T., Schulz R. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers // [SoftwareX](#). 2015. — sep. Vol. 1-2. P. 19–25. URL: <https://doi.org/10.1016/j.softx.2015.06.001>.
42. Duan Y., Wu C., Chowdhury S. et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations // [Journal of Computational Chemistry](#). 2003. — oct. Vol. 24, no. 16. P. 1999–2012. URL: <https://doi.org/10.1002/jcc.10349>.
43. Bauer B. A., Patel S. Properties of water along the liquid-vapor coexistence curve via molecular dynamics simulations using the polarizable TIP4P-QDP-LJ water model // [The Journal of Chemical Physics](#). 2009. — aug. Vol. 131, no. 8. P. 084709. URL: <https://doi.org/10.1063/1.3200869>.