

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

Работа 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

Выполнил студент гр. 786 _____ Поляченко Юрий Анатольевич

Научный руководитель,

д. ф.-м. н., в. н. с.

_____ Стегайлов В.В.

Москва – 2021

Оглавление

Введение	3
Глава 1. Описание эксперимента по измерению объемного модуля упругости K	5
1.1. Поверхностная адсорбция	6
1.2. Модель Flory-Huggins	10
1.2.1. Хим. потенциал пара в контакте с жидкостью	10
1.2.2. Вклад упругости в μ	11
1.2.3. Вклад смешивания в μ	12
1.3. Эксперимент	17
Глава 2. Построение МД модели кристалла Лизоцима	20
2.1. Подготовка PDB-белка к расчету	20
2.2. Сходимость уровня гидротации	25
2.3. 2 метода расчета модуля упругости K по траекториям	26
Глава 3. Результаты моделирования	27
3.1. Сравнение расчетных и экспериментальных K	27
3.2. Анализ мобильности молекул воды	27
3.3. 2-фазное моделирование для учета влажности	27
Заключение	28
Список литературы	29

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. Но для глубокого понимания устройства белка необходимо знать не только его аминокислотную последовательность, но и трехмерную 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 белков 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 кристаллизация может быть проведена с заменой частей белка, мешающих ей. Это опять же многократно усложняет процесс, т.к. нужно предпринимать попытки кристаллизации многих подобных белков. 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-атомарной точки зрения. Также такая модель заметно упростит исследование 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 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

В главе 3 приведены основные результаты численного моделирования.

В заключении кратко обсуждаются полученные результаты и делаются соответствующие выводы. Изложены также перспективы развития работы.

Глава 1

Описание эксперимента 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 белка с того момента как концепция гидрофобного взаимодействия была введена Каузмана в 1959 [4].

Для лизоцима уровень $h = 0.2$ соответствует давлению пара $P/P_0 \sim 0.7$, 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, 9]. 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 C°. Поэтому в данной экспериментальной работе уровень гидратации измерялся с помощью ядерного магнитного резонанса водородов ^1H в системе [10, 11].

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 в 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 $d_{\text{H}_2\text{O}}$ к характерной длине свободного пробега $\lambda \sim 1/n\sigma$.

$$\frac{\lambda}{d_{\text{H}_2\text{O}}} \sim \frac{\rho_{liq}}{\rho_{gas}} \sim \frac{10^3}{10^{-2}} \sim 10^5 \quad (1.1)$$

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

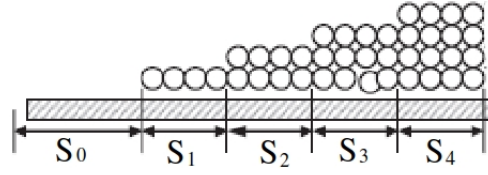


Рис. 1.1. BET-модель поверхностной адсорбции. В статическом состоянии величины S_0 , S_1 , ... не должны изменяться.

Делается предположение, что адсорбция происходит индивидуально на точках адсорбции и «столбы» молекул, адсорбированных на соседних точках, 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

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

Тогда объем адсорбированной воды

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

где H – высота 1 слоя.

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

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

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 $P_0(T)$ – давлении 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 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 выполнено в лизоциме [12], где количество адсорбированной кристаллом белка воды зависит от относительной влажности по разному при температурах порядка физиологических и при $T < 10C^\circ$. Плохая применимость (1.10) с белкам объяснима, т.к. при выводе были сделаны предположения, не выполняемые в белке. Например – пренебрежение внешними ограничениями количества возможных слоев адсорбции. В лизоциме есть пора размером порядка $\sim 1 \text{ \AA}$, что сравнимо с размерами молекул воды и поэтому делает переход от (1.8) к (1.9) не корректным. 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. 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 помтоянных (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. 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 $\mu_{vap}(P < P_0) < 0$, $\mu_{vap}(P > P_0) > 0$, откуда

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

Подставляя это обратно в (1.13) получаем

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

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:

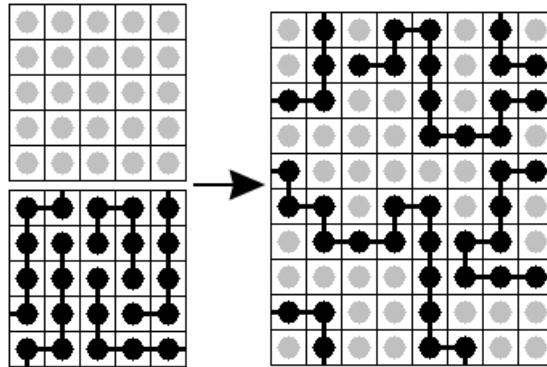


Рис. 1.2. Описание смешивания воды с полимером в рамках модели Flory–Huggins.

Основные приближения данной модели:

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

В такой модели можно попытаться оценить количество возможных состояний 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, 12, 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 так же 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 $P/P_0 \gtrsim 0.7$ [12], 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 [12], 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:

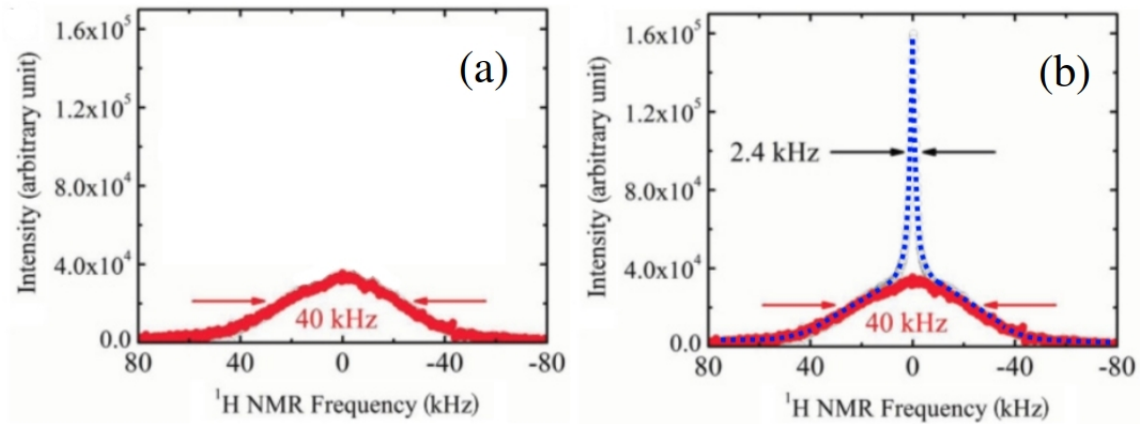


Рис. 1.3. 1H ЯМР-спектры порошка лизоцима в осушенном (а, красный профиль) и влажном (б, синий профиль) состояниях.

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 [12], можно восстановить изначальную амплитуду сигнала [17]. Это восстановление – основной источник ошибки, составляющей $\sim 5\%$ [12]. 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 $P/P_0 > 0.7$ [12]. 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 [12], то их значения находятся из аппроксимаций линейных участков на Рис. 1.4.

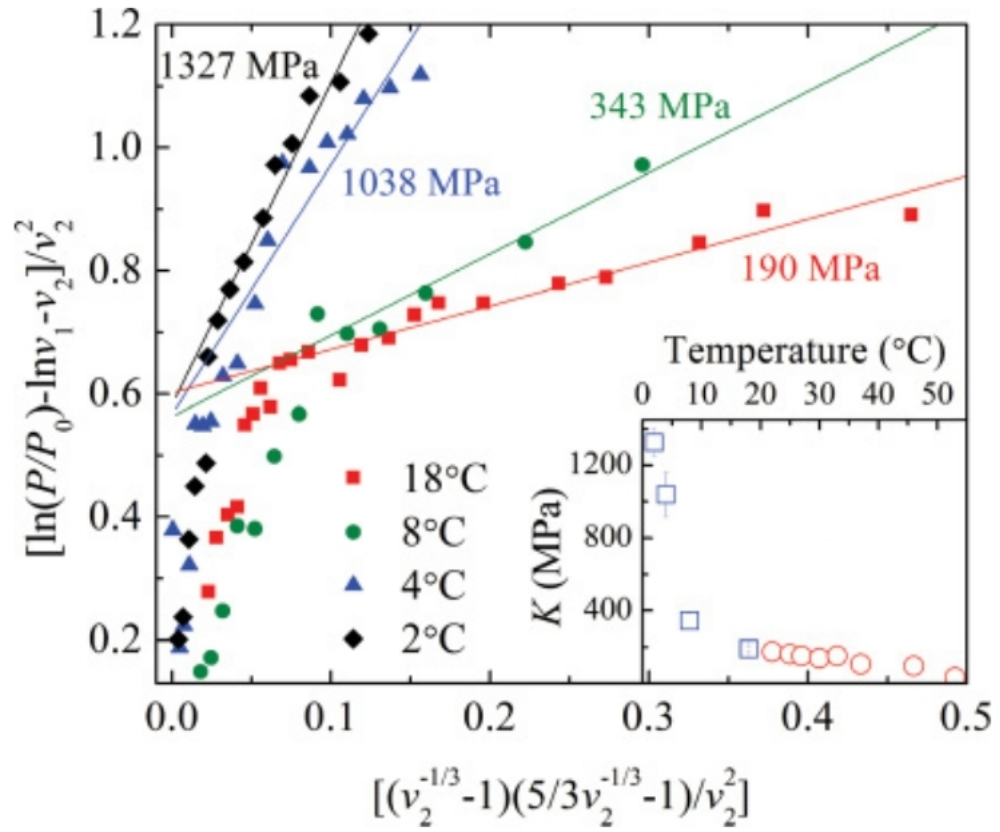


Рис. 1.4. Изотермы вида (1.34). Кривые линейны при $[\ln(P/P_0) - \ln(1 - v_2) - v_2]/v_2^2 > 0.6$. Использовалось значение $\eta = 1$. Малый график: Полученные значения упругости влажного кристалла в зависимости от температуры. Нанесены результаты, полученные в [12] (синие точки) и в [19] (красные точки)

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. Считаю 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

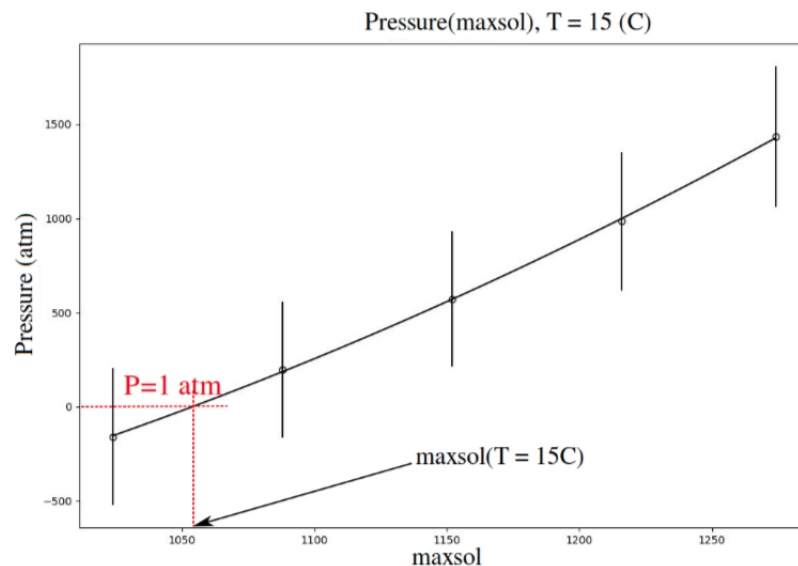


Рис. 2.1. Пример определения количество воды, приводящего к экспериментальному (фактически нулевому) давлению для заданной температуре.

2.3. 2 метода расчета модуля упругости K по траекториям

3

Глава 3

Результаты моделирования

В этой главе приведены результаты и анализ расчётов.

В разделе 3.1 сравниваются значения K , полученные двумя методами, описанными выше, и экспериментальные значения. Обсуждаются возможные причины расхождения.

Далее в разделе 3.2 исследуется мобильность молекул воды, т.к. изменения именно в ней могут провоцировать изменение характера поведения K с температурой.

Затем в 3.3 получена зависимость плотности пара от количества воды, добавляемой в систему. По ней можно определить количество воды, необходимое для воспроизведения эксперимента в заданной влажности

3.1. Сравнение расчетных и экспериментальных K

1

3.2. Анализ мобильности молекул воды

2

3.3. 2-фазное моделирование для учета влажности

3

Заключение

В ходе выполнения данной работы:

- Сформулирован алгоритм подготовки кристалла к MD-расчету.
- Проверены сходимости по параметрам системы.
- Реализовано 2 независимых метода расчета K кристалла. K , рассчитанные по этим 2 методам, согласуются между собой, но плохо согласуются с экспериментом.
- Проанализирована модильность молекул воды в белке.
- Проведено сравнение результатов при различных моделях воды.
- 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. Так же 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. 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>.
 10. 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>.
 11. 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>.
 12. 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>.
 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>.