

---

---

# Связь стабильности $\alpha$ -спирали с ее аминокислотным составом

Я. А. Орлов<sup>1</sup> и Н. К. Победин<sup>1</sup>

<sup>1</sup>Физтех-школа физики и исследований им. Ландау.

Московский физико-технический институт (национальный исследовательский университет)

Получено: 14 мая 2024 г.

Keywords and phrases:  $\alpha$ -спираль, ?

## 1. Введение

?какой-то текст про цели задачи туда сюда?

## 2. Теоретические сведения

### 2.1. Понятие о структуре белков

Белки — это и молекулярные машины, и строительные блоки, и оружие живой клетки. Важнейшая и почти монополярная функция белков - ферментативный катализ химических превращений в клетке и вокруг нее.

При всем разнообразии, работа белков всегда базируется на высоко специфическом — как у ключа с замком (точнее: как у гибкого ключа с гибким замком) - взаимодействии белка с обрабатываемой им молекулой. Для этого взаимодействия необходима достаточно «твердая» (во всяком случае, у «работающего белка) пространственная структура. Поэтому биологическая функция белков (как и других важнейших для жизни макромолекул — ДНК и РНК) тесно связана с определенностью их трехмерных структур. Не только разрушение - даже небольшие изменения этих структур часто ведут к утере или резкому изменению активности белков. Знание молекулярной трехмерной структуры белка необходимо для понимания функционирования белковой молекулы.

Нековалентные взаимодействия, поддерживающие пространственное строение белка, значительно слабее химических связей, фиксирующих последовательность мономеров — аминокислот в белковой цепи. Эта последовательность — она называется «первичной структурой белка» (рис. 1) — создается в ходе матричного биохимического синтеза согласно «инструкции», записанной в гене.

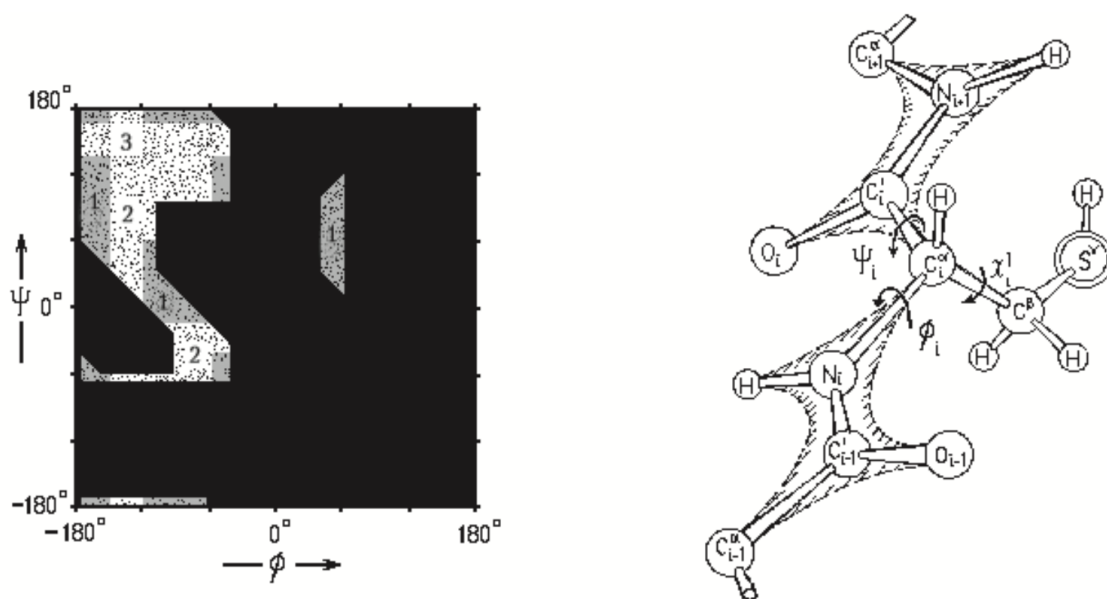
Архитектуры белков сложны и разнообразны — в противоположность универсальности двойной спирали ДНК. Однако и в белках прослеживается набор «стандартных» структур. В своей работе мы будем исследовать  $\alpha$ -спирали, которые часто изображаются спиральными лентами или цилиндрами (рис. 1)



Рис. 1.  $\alpha$ -спираль и  $\beta$ -структура

## 2.2. Карты Рамачандрана

Аминокислоты связаны между собой пептидными связями C' и N атомов



На рисунке 2 справа изображен полипептид: главная цепь и боковая группа (Ser) на ней. Пептидные группы заштрихованы. Показаны углы внутреннего вращения в главной ( $\phi, \psi, \omega$ ) и боковой ( $\chi^1$ ) цепях. Индексы  $i-1, i, i+1$  показывают последовательность аминокислотных остатков в цепи. Стрелки указывают направление вращения ближней к нам части цепи относительно более отдаленной ее части, ведущее к росту угла поворота.

Слева изображена его карта Рамачандрана - изображенные в координатах ( $\phi, \psi$ ) запрещенные и разрешенные конформации (то есть положения при повороте на на углы  $\phi$  и  $\psi$ ) остатка.

Рис. 3-5. Карта запрещенных (■) и разрешенных (□, I - точки на белом фоне, II - точки на сером фоне) конформаций крупных остатков при вращении по углам  $\phi, \psi$  в белковой цепи. В области □ разрешены все три ротамера  $C^\gamma$  атома по  $\chi^1$ , в области II - 2 ротамера, в области I - 1 ротамер.

### 2.3. Свойства аминокислотных остатков

Список 20 «стандартных», т. е. кодируемых ДНК-аминокислотных остатков дан в таблице ниже, там же дан их молекулярный вес и встречаемость в белках. Структуры этих аминокислотных остатков представлены на рис. 2.

А.к. ост.	Наличие		Число $\gamma$ б.гр.	Диполь/заряд	pK	Яркая тенденция быть:							
	NH	C <sup><math>\beta</math></sup>				до- $\alpha_N$	В спирали $\alpha_N$	за- $\alpha$	$\alpha_C$	$\alpha_C$	$\beta$	в петлях	в ядре
Gly	+	–						–			–	+	
Ala	+	+						+				–	
Pro	–	+	1				+	–	–	–	–	+	
Glu	+	+	1	COOH $\Rightarrow$ CO <sub>2</sub> <sup>–</sup>	4,3	+	+		–	–	–		–
Asp	+	+	1	COOH $\Rightarrow$ CO <sub>2</sub> <sup>–</sup>	3,9	+	+	–	–	–	–	+	–
Gln	+	+	1	OCNH <sub>2</sub>									–
Asn	+	+	1	OCNH <sub>2</sub> <sup>2</sup>		+		–		+	–	+	–
Ser	+	+	1	OH		+						+	–
His	+	+	1	NH; и N $\Rightarrow$ NH <sup>+</sup>	6,5		–		+	+			
Lys	+	+	1	NH <sub>2</sub> $\Rightarrow$ NH <sub>3</sub> <sup>+</sup>	10,5	–	–		+	+	–		–
Arg	+	+	1	HNC (NH <sub>2</sub> ) <sub>2</sub> <sup>3</sup> +	12,5	–	–		+	+	–	+	–
Thr	+	+	2	OH		+					+		
Ile	+	+	2								+	–	+
Val	+	+	2								+	–	+
Leu	+	+	1					+			+	–	+
Met	+	+	1					+			+	–	+
Phe	+	+	1								+	–	+
Tyr	+	+	1	OH $\Rightarrow$ O <sup>–</sup>	10,1			–			+		+
Cys	+	+	1	SH $\Rightarrow$ S <sup>–</sup>	9,2			–			+		+
Trp	+	+	1	NH							+		+

Попробуем понять основные закономерности этой таблицы, исходя из того, что мы уже изучили. При этом мы будем использовать следующую логику: так как белок в целом стабилен, он должен в основном состоять из стабильных элементов, т.е. именно они должны наблюдаться в его структуре чаще всего, а нестабильные должны наблюдаться редко.

Почему пролин не любит вторичной структуры? Потому, что у него уменьшена возможность завязывать водородные связи, а именно на них и держится вторичная структура. Почему он, тем не менее, любит N-конец спирали? Потому, что здесь, в водородные связи, и здесь пролину нечего терять... С другой стороны, угол  $\varphi$  в пролине фиксирован его кольцом примерно при  $-60^\circ$ , т. е. его конформация уже почти «готова» для  $\alpha$ -спирали (рис. 10-2а).

Почему глицин не любит вторичной структуры и предпочитает нерегулярные участки («клубок»)? Потому, что для него допустима очень широкая область углов ( $\varphi\psi$ ) на карте Рамачандрана (рис. 10-2б), ему легко принимать самые разнообразные конформации, лежащие вне вторичной структуры.

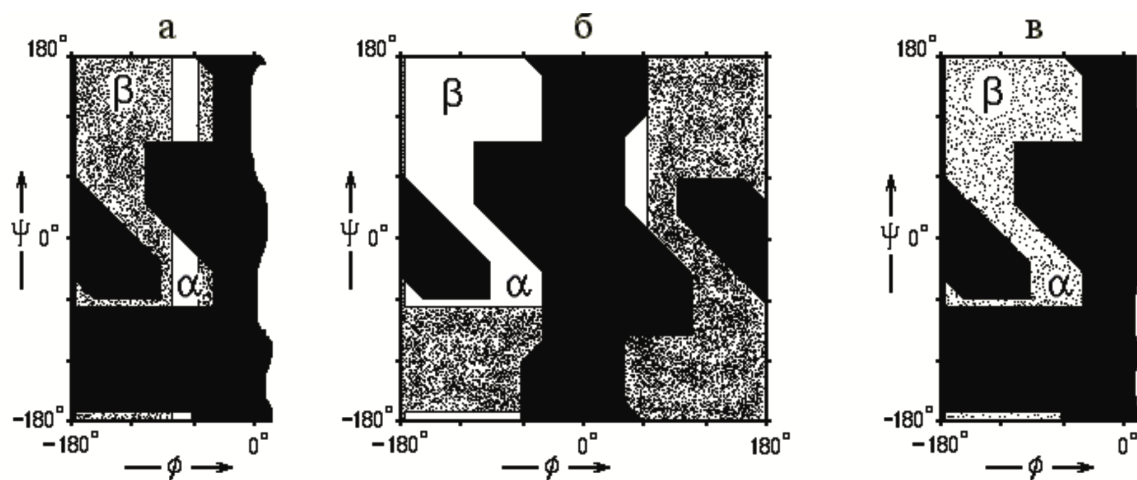


Рис. 2. Запрещенные и разрешенные конформации различных аминокислотных остатков и - на их фоне - конформации  $\alpha$  - и  $\beta$ -структуры. (а) Разрешенные для пролина конформации (□) на фоне конформаций, разрешенных для аланина (I - точки на белом фоне); ■- конформации, запрещенные для них обоих. (б) Разрешенные конформации аланина (□) на фоне конформаций I, разрешенных лишь для глицина; ■ — области, запрещенные для всех остатков. (в) Карта запрещенных (■) и разрешенных (□, I) конформации боковой группы по углу  $\chi^1$ , в области I часть углов  $\chi^1$  запрещена

Наоборот, аланин - с более узкой, но включающей и  $\alpha$ -, и  $\beta$ -конформацию разрешенной областью на карте Рамачандрана (рис. 10-26) - предпочитает нерегулярным конформациям  $\alpha$ -спираль (и отчасти  $\beta$ -структуру).

Остальные гидрофобные остатки (т.е. остатки без зарядов и диполей в боковой цепи) предпочитают, как правило,  $\beta$ -структуру. Почему? Потому, что их крупные  $\gamma$ -атомы могут там располагаться более свободно (рис. 10-2в). Особенно это важно для боковых групп с двумя крупными  $\gamma$ -атомами.

А вот аминокислоты с полярными группами в боковых цепях предпочитают поверхностные нерегулярные участки («клубок»), где эти полярные группы могут завязать водородные связи как с полипептидной цепью, так и с водой. Особенно заметна эта тенденция для наиболее полярных, заряженных при «нормальном» pH7 остатков, и для самых коротких (см. рис. 10-1) полярных боковых цепей, с наиболее приближенными к главной цепи полярными группами. Кстати, по той же причине, поскольку у них там есть возможность завязать дополнительную водородную связь, короткие полярные боковые группы любят места у обоих концов спирали.

Некое исключение среди аминокислот с диполями в боковой цепи составляют триптофан и тирозин, имеющие маленький диполь на фоне большой гидрофобной части, и цистеин, у которого (т.е. у SH-группы которого) водородные связи совсем слабые. Они ведут себя, в общем, так же, как гидрофобные остатки.

Мы видим также, что отрицательно заряженные боковые группы предпочитают N-конец спирали (точнее: N-концевой виток и один-два остатка перед ним) и не любят C-концевой виток (и пару остатков за ним), а положительно заряженные - предпочитают C-конец спирали и не любят ее N-онец. Почему? Потому, что на N-конце из спирали торчат NH-группы и на нем образуется заметный положительный заряд, и «минусы» боковых цепей притягиваются к нему, а «плюсы» отталкиваются от него. А C-конец спирали заряжен, наоборот, отрицательно, и там эффект противоположен: около C-конца любят собираться «плюсы» боковых цепей, а «минусы» его избегают.

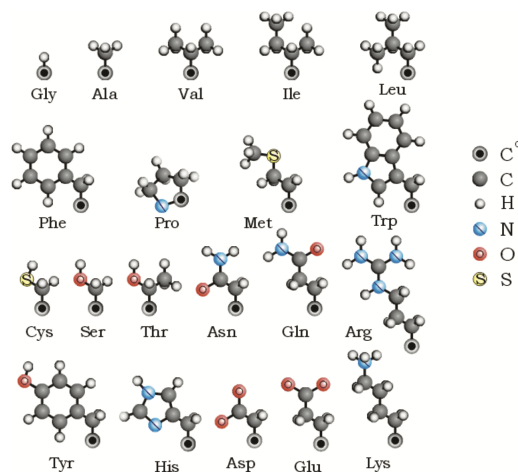


Рис. 3.

Что касается расположения остатков внутри белка или на его поверхности, здесь общая тенденция заключается в том, что полярные (гидрофильные) боковые группы находятся снаружи, где они могут контактировать с полярной же водой («подобное растворяется в подобном»!). Отрываться от воды полярным группам плохо - теряются водородные связи. Особенно плохо отрываться заряженным группам: переход из среды с высокой диэлектрической проницаемостью (из воды) в среду с низкой (ядро белка) ведет к большому повышению свободной энергии. И действительно, ионизированных групп внутри белка практически нет (а почти все исключения связаны либо с координационными связями с ионом металла, либо с активными центрами, ради которых, собственно, белок и создан...).

Наоборот, большинство гидрофобных боковых групп находятся внутри белка - они-то и создают здесь гидрофобное ядро (опять: «подобное растворяется в подобном»!). Мы уже говорили, что гидрофобность группы тем больше, чем больше ее неполярная поверхность: именно ее нужно упрятать от воды. Для чисто неполярных групп гидрофобный эффект прямо пропорционален их поверхности, а для групп с полярными вкраплениями - их поверхности за вычетом поверхности этих вкраплений.

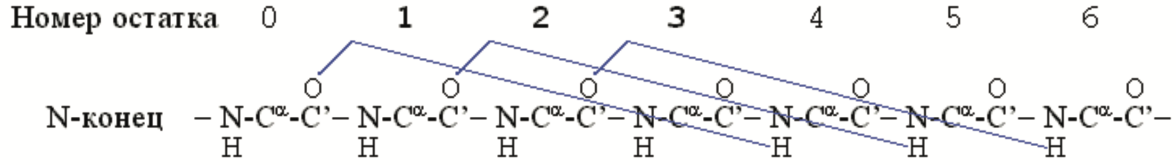
Слипание гидрофобных групп - главная движущая сила образования белковой глобулы. Главная, но не единственная - еще есть образование водородных связей во вторичной структуре (о чем мы уже говорили) и образование плотной, квазикристаллической упаковки внутри белка (о чем мы еще поговорим в свое время).

Для создания гидрофобного ядра белковой цепью она должна входить в него с уже насыщенными водородными связями - ведь иначе ее полярным пептидным группам от воды придется оторваться, а разрыв водородной связи дорог. Поэтому в гидрофобное ядро вовлекается цепь, уже образовавшая (или образующая при этом) вторичную структуру и тем самым насытившая водородные связи пептидных групп в главной цепи. Однако при этом в ядро должны увлекаться только гидрофобные остатки вторичной структуры, а входящие в нее полярные остатки должны остаться вне ядра, потому и на  $\alpha$ -спиралях, и на  $\beta$ -структурных участках выделяются гидрофобные и гидрофильные поверхности; для их создания необходимо определенное чередование соответствующих групп в белковой цепи.

Все закономерности, о которых мы сейчас говорили, используются как для конструирования искусственных белков, так и для предсказания - по аминокислотным последовательностям - вторичной структуры белков, а также для предсказаний тех участков их цепи, что глубоко погружены в белок, или, наоборот, тех участков, что лежат на поверхности белка.

#### 2.4. Стабильность $\alpha$ -спирали

Первая водородная связь в  $\alpha$ -спирали  $(CO)_0 - (HN)_4$  фиксирует конформации трех остатков - 1, 2, 3; следующая водородная связь  $(CO)_1 - (HN)_5$  дополнительно фиксирует конформацию только одного остатка - остатка 4; связь  $(CO)_2 - (HN)_6$  дополнительно фиксирует остаток 5, и т. д.



Значит, если в спирали фиксировано  $n$  остатков, то их фиксирует  $n - 2$  водородные связи. Рассмотрим свободную энергию образования такой спирали из клубка в водном окружении («клубок» - это полимер без фиксированной структуры и без взаимодействия дальних по цепи звеньев). Эту свободную энергию можно записать как

$$\Delta F_{\alpha} = F_{\alpha} - F_{\text{клуб.}} = (n - 2)f_H - nTS_{\alpha} = -2f_H + n(f_H - TS_{\alpha}). \quad (1)$$

Здесь  $f_H$  - свободная энергия образования водородной связи и сопутствующих ей взаимодействий в  $\alpha$ -спирали (вы помните, что  $f_H$  - не просто энергия, как было бы в вакууме: в нее входит и энергия, и энтропия перестроек водородных связей в водном окружении), а  $S_{\alpha}$  - потеря энтропии при фиксации одного остатка в спирали.

Вы видите, что в  $\Delta F_{\alpha}$  есть два члена. Один ( $-2f_H$ ) не зависит от длины спирали; величина

$$f_{\text{init}} = -2f_H \quad (2)$$

традиционно называется свободной энергией инициации спирали (на самом деле,  $f_{\text{init}}$  - суммарная свободная энергия обеих границ спирали с клубком; она учитывает и инициацию, и терминацию спирали). Другой член,  $n(f_H - TS_{\alpha})$ , прямо пропорционален длине спирали; величина

$$f_{\text{el}} = (f_H - TS_{\alpha}) \quad (3)$$

называется свободной энергией элонгации спирали на один остаток. В общем виде

$$\Delta F_{\alpha} = f_{\text{init}} + nf_{\text{el}}. \quad (4)$$

При этом отношение вероятности чисто спирального состояния цепи из  $n$  остатков к ее же чисто клубковому (и начисто лишенному спиральных примесей) состоянию, равно

$$\exp(-\Delta F_{\alpha}/kT) = \exp(-f_{\text{init}}/kT) [\exp(-f_{\text{el}}/kT)]^n = \sigma s^n. \quad (5)$$

Здесь используются общепринятые обозначения:

- фактор элонгации спирали:  $s = \exp(-f_{\text{el}}/kT)$ ;
- фактор инициации спирали:  $\sigma = \exp(-f_{\text{init}}/kT)$ .

Ясно, что  $\sigma \ll 1$ , так как  $\sigma = \exp(-f_{\text{init}}/kT) = \exp(+2f_H/kT)$ ; а свободная энергия водородной связи - большая отрицательная величина, порядка нескольких  $kT$ .

Величина  $\exp(-\Delta F_{\alpha}/kT) = \sigma s^n$  - это просто константа равновесия  $n$  звеньев.

Выясним вопрос о том, как образуется спираль при изменении условий среды (температуры, растворителя и т. д.) - фазовым переходом или постепенно? Поначалу кажется, что такая отличная от клубка структура, как спираль, должна «вымораживаться» из него путем фазового перехода - как лед из воды... Однако на этот счет есть теорема Ландау, которая гласит, что в системе, где обе фазы одномерны, - фазовый переход первого рода невозможен.

"Одномерность" означает, что свободная энергия (или по сути размер границы раздела) не зависит размера кусков фаз. Соответственно в трехмерной системе сосуществование двух разных фаз энергетически не выгодно, а в одномерной выгодно. Поэтому в одномерной системе фазы перемешиваются.

Найдем характерную длину  $n_0$  спирального участка в середине перехода спираль-клубок

Рассмотрим цепь из  $N$  звеньев при температуре «середины перехода», где спираль и клубок имеют равную свободную энергию, т. е.  $f_{el} = 0$ . При этом свободная энергия элонгации спирали (а равно и клубка) - ноль, ее инициации -  $f_{init}$ , а число возможных положений спирали в цепи из  $N$  звеньев - порядка  $N^2/2$  (она может начинаться и кончаться в любом месте при единственном условии, что ее длина - не менее трех остатков); и ни расположение спирали в цепи, ни ее длина (при  $f_{el} = 0$ ) не влияют на ее свободную энергию. Для получения качественной оценки пренебрежем мелочами (цифрами) по сравнению с главным (буквами). Итак: размещений спирали - порядка  $N^2$ , т. е. их энтропия -  $k \cdot 2 \ln(N)$ , а полная свободная энергия внедрения куска новой фазы (спирали с флуктуирующими концами) в цепь длины  $N$  - примерно  $f_{init} - 2kT \ln(N)$ . Если она, эта свободная энергия, больше нуля - новая фаза не внедрится; если она меньше нуля - новая фаза может внедриться, и даже неоднократно. Значит, смешение клубковой и спиральной фаз начинается в кусках длины  $N \sim n_0$ , а величина  $n_0$  получается из уравнения  $f_{init} - 2kT \ln(n_0) = 0$ . Итак: характерная длина кусков спирали и клубка в середине перехода

$$n_0 = \exp(+f_{init}/2kT) = \sigma^{-1/2} \quad (6)$$

Наконец, зная  $n_0$ , можно вычислить  $f_{init}$  и  $\sigma$ . Для большинства аминокислот  $n_0 \approx 30$ ,  $f_{init} \approx 4$  ккал/моль, и  $\sigma \approx 0,001$ .

Теперь мы можем найти свободную энергию образования водородной связи (вкуче со всеми сопутствующими образованию Н-связи в  $\alpha$ -спирали взаимодействиями): согласно (2),  $f_H = -f_{init}/2 \approx -2$  ккал/моль. Можно найти и конформационную энтропию, теряемую при фиксации одного звена в  $\alpha$ -спирали: согласно формуле (3), при  $f_{el} = 0$ ,  $TS_\alpha = f_H \approx -2$  ккал/моль.

Оба параметра стабильности спирали - и  $f_{el}^\alpha$ , и  $f_{init}$  - зависят от температуры, но по-настоящему сильно влияет на стабильность спирали именно отклонение величины  $f_{el}$  от 0. Дело в том, что в спирали, состоящей из  $\sim n_0$  звеньев, это отклонение умножается на большое число  $n_0$  и уже в таком виде входит в свободную энергию спирали. Когда величина  $f_{el} n_0/kT$  составляет порядка +1 (более точная оценка:  $f_{el} n_0/kT = +2$ ), спиральность практически исчезает, а когда  $f_{el} n_0/kT = -2$  - практически исчезает клубок.

Стабильность  $\alpha$ -спирали обычно падает с ростом температуры и с добавлением полярных денатурантов и растет с добавлением слабополярных растворителей.

Для измерения влияния отдельных аминокислотных остатков на стабильность спиралей сейчас чаще всего используются короткие (длиной  $\sim n_0$  или менее) полипептиды. В них может образоваться только одна спираль, и оценить влияние каждой аминокислотной замены на спиральность здесь наиболее просто.

### 3. Методы

В работе рассмотрены  $\alpha$ -спирали, состоящие из 12 аминокислот. Рассмотрение их всех не представляется возможным ( $20^{12} = 4 \cdot 10^{15}$  последовательностей), поэтому были взяты только выборочные последовательности.

Трехмерные структуры  $\alpha$ -спиралей генерировались с помощью Molefactory, расширения пакета VMD; со встроенными параметрами спирали  $\phi = -57^\circ$ ,  $\psi = -47^\circ$ .

Молекулярная динамика моделировалась в пакете Gromacs. Использовалось поле сил CHARMM27 и модель воды SPC/E. Далее использовались периодические граничные условия с ячейкой моделирования в виде ромбического додекаэдра (экономит объем моделируемой воды по сравнению с кубической), при этом в процессе моделирования соблюдалось соглашение о том, чтобы расстояние между белковой структурой и ее периодическим изображением превышало расстояние, на котором обнуляется сила.

Система была нейтрализована (белок в нейтральных условиях имеет заряд) с помощью ионов натрия.

Перед началом моделирования система проходила несколько этапов подготовки, первый из них минимизация энергии. На этом этапе система изменялась в направлении действующих в ней сил. Это позволяет понизить начальную потенциальную энергию и уменьшить напряжения в молекуле (которые в действительности возникали из-за неестественной геометрии  $\alpha$ -спиралей для некоторых последовательностей), чтобы при начале моделирования эта потенциальная энергия не высвободилась мгновенно в виде температуры, и система не коллапсировала. Минимизация проводилась градиентным спуском, пока сила не опускалась до порогового значения в  $500 \text{ кДж}/(\text{моль} \cdot \text{нм})$ . При этом энергия опускалась как правило до  $10^{-5} \text{ кДж}/(\text{моль} \cdot \text{нм})$ .

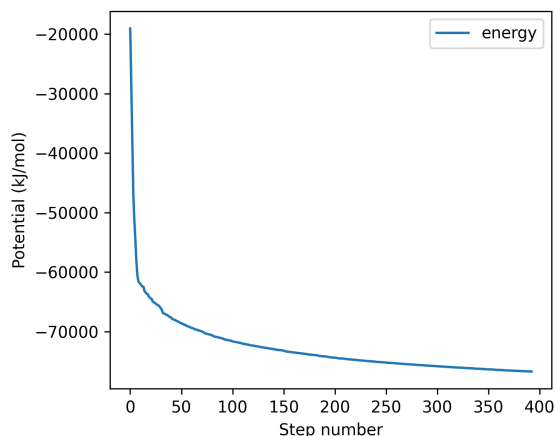


Рис. 5. Потенциальная энергия в процессе ее минимизации

Далее проводилось уравнивание с NVT-ансамблем (постоянное число частиц, объем и температура), при этом частицы белка удерживались неподвижными до некоторого порога силы действующей на него. Это позволяло системе прийти в равновесие при заданной температуре. Температура была выбрана 300K. Генерация исходных скоростей частиц производилась в соответствии с распределением Максвелла при этой температуре. Уравнивание заключалось в моделировании поведения системы с термостатом. Моделирование велось с помощью схемы Leap-frog с шагом 2фс на 50000 шагов (100пс). В качестве термостата был взят термостат Berendsen'a. При этом было достигнуто состояние, при котором температура колебалась не более, чем на 10K, в зависимости от размеров системы.

Далее было проведено уравнивание с NPT-ансамблем (постоянное число частиц, давление и температура) для того, чтобы подобрать подходящий для системы объем. Аналогично, моделирование длилось 100пс с шагом 2фс. При этом также использовалась схема Leap-frog, а баростат Паринелло-Рахмана. В итоге плотность в системе колебалась на  $20 \text{ г}/\text{см}^3$ . Давление при этом было очень нестабильно и колебалось на сотни атмосфер в связи с малым размером системы.

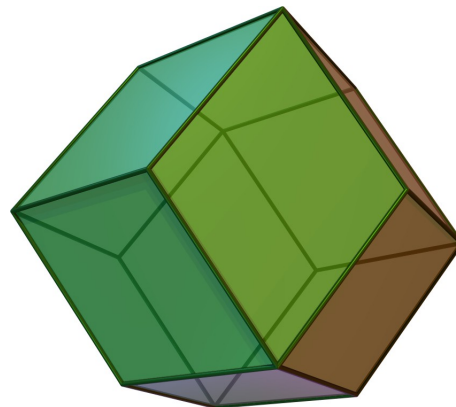


Рис. 4. Ромбический додекаэдр



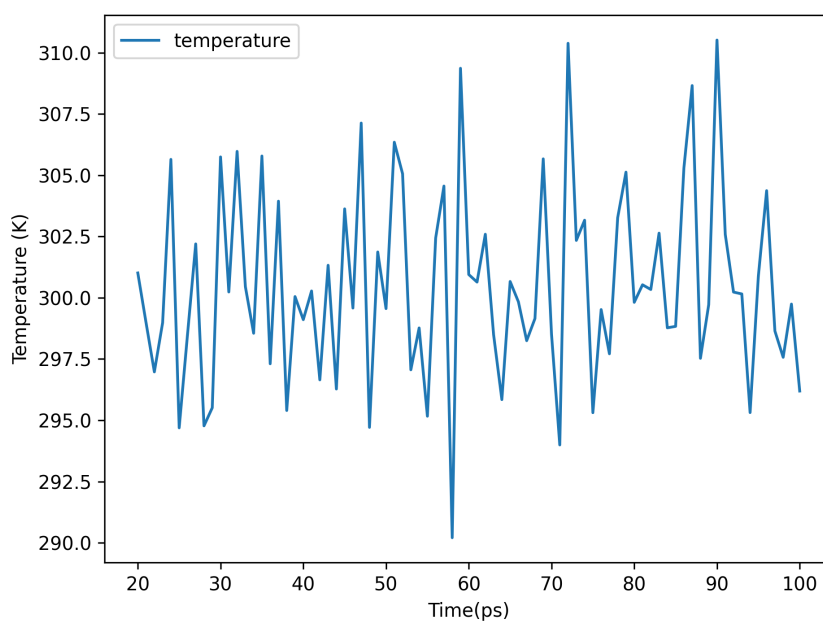


Рис. 6. Температура во время NVT-уравновешивания

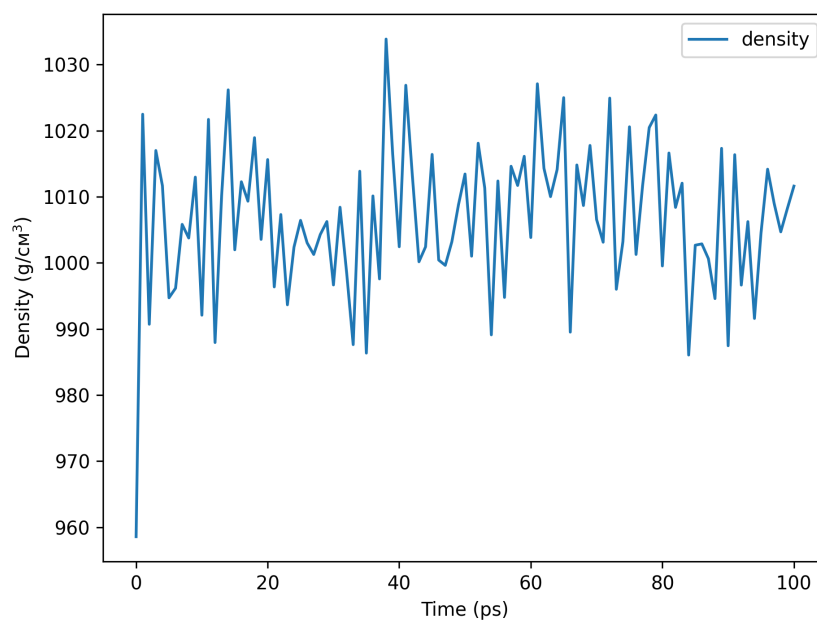


Рис. 7. Плотность во время NPT-уравновешивания

И наконец после этого можно было проводить моделирование. Оно также велось с помощью схемы Leap-frog с шагом 2фс на 4нс. Такое большое время моделирования было доступно благодаря очень небольшому размеру белка.

Траектория частиц рассматривалась с помощью VMD. Анализ макропараметров проводился с помощью встроенных инструментов Gromacs, таких как rms для нахождения средне-квадратичного отклонения атомов от исходной “идеальной” структуры  $\alpha$ -спирали, gyrate для определения радиуса гирации и hbond для подсчета водородных связей белка.

## Список литературы

1. Финкельштейн А. В., Птицын О. Б. Физика белка: (Addison-Wesley, London, 1984).