

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра биомедицинской информатики

ТЕВ
Никита Михайлович

**ПОСТРОЕНИЕ МОЛЕКУЛЯРНО-ДИНАМИЧЕСКОЙ МОДЕЛИ
ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ ТЕЛОМЕРАЗЫ ЧЕЛОВЕКА –
ПЕРСПЕКТИВНОЙ МИШЕНИ ДЛЯ КОНСТРУИРОВАНИЯ
НОВЫХ ПРОТИВООПУХОЛЕВЫХ ПРЕПАРАТОВ**

Дипломная работа

Научный руководитель:
доктор химических наук,
профессор А.М. Андрианов

Допущена к защите

« ___ » _____ 2021 г.

Зав. кафедрой биомедицинской информатики
кандидат физико-математических наук, доцент Ю.Л. Орлович

Минск, 2021

РЕФЕРАТ

Дипломная работа, 45 страниц, 13 рисунков, 2 таблицы, 24 источника.

Ключевые слова: ТЕЛОМЕРАЗА, ГОМОЛОГИЧНОЕ МОДЕЛИРОВАНИЕ МОЛЕКУЛЯРНАЯ ДИНАМИКА, КЛАСТЕРИЗАЦИЯ, КВАНТОВАЯ ХИМИЯ, AMBER, MORAS.

Объект исследования: теломераза человека, ее отдельные структурные домены, а также родственные соединения.

Цель работы: построение молекулярно-динамической модели пространственной структуры теломеразы человека на основе структурных данных родственных шаблонов.

Результат: рассмотрена литература и научные статьи, связанные с изучением холоферментного комплекса теломеразы. Изучены теоретические основы гомологичного моделирования, молекулярной динамики и квантово-химических расчетов. С использованием полученных знаний проведено гомологичное моделирование пространственной структуры фрагмента (TERT) теломеразы человека, осуществляющего обратную транскрипцию. Построена и проанализирована молекулярно-динамическая траектория, моделирующая поведение данной структуры во времени. На основе построенной траектории проведена дополнительная оптимизация модели с помощью методов квантовой химии.

Область применения: полученная в результате работы модель пространственной структуры теломеразы может быть использована в других исследованиях как исходная точка для дальнейшего анализа, а также с целью поиска ингибиторов теломеразы, как потенциальных лекарственных препаратов с широким спектром противоопухолевого действия.

РЭФЕРАТ

Дыпломная праца, 45 старонак, 13 малюнкаў, 2 табліцы, 24 крыніцы.

Ключавыя словы: ЦЕЛАМЕРАЗА, ГАМАЛАГІЧНАЕ МАДЭЛІРАВАННЕ, МАЛЕКУЛЯРНАЯ ДЫНАМІКА, КЛАСТЭРЫЗАЦЫЯ, КВАНТАВАЯ ХІМІЯ, АМВЕР, МОРАС.

Аб'ект даследавання: целамераза чалавека, яе асобныя структурныя дамены, а таксама роднасныя злучэнні.

Мэта работы: пабудова малекулярна-дынамічнай мадэлі прасторавай структуры целамеразы чалавека на аснове структурных дадзеных роднасных шаблонаў.

Вынік: разгледжана літаратура і навуковыя артыкулы, звязаныя з вывучэннем холаферментнага комплексу целамеразы. Вывучаны тэарэтычныя асновы гамалагічнага мадэліравання, малекулярнай дынамікі і квантава-хімічных разлікаў. З выкарыстаннем атрыманых ведаў праведзена гамалагічнае мадэліраванне прасторавай структуры фрагмента (TERT) целамеразы чалавека, ажыццяўляючага зваротную транскрыпцыю. Пабудавана і прааналізавана малекулярна-дынамічная траекторыя, якая мадэліруе паводзіны дадзенай структуры з цягам часу. На аснове пабудаванай траекторыі праведзена дадатковая аптымізацыя мадэлі з дапамогай метадаў квантавай хіміі.

Вобласць ужывання: атрыманая ў выніку працы мадэль прасторавай структуры целамеразы можа быць выкарыстана ў іншых даследаваннях як зыходная кропка для далейшага аналізу, а таксама з мэтай пошуку інгібітараў целамеразы, як патэнцыйных лекавых прэпаратаў з шырокім спектрам проціпухліннага дзеяння.

ABSTRACT

Diploma thesis, 45 pages, 13 pictures, 2 tables, 24 sources.

Keywords: TELOMERASE, HOMOLGY MODELING, MOLECULAR DYNAMICS, CLUSTERING, QUANTUM CHEMISTRY AMBER, MOPAC.

Object of research: human telomerase, its individual structural domains and related compounds.

Objective: building of a molecular dynamic model of the spatial structure of human telomerase based on the structural data of related templates.

The result: literature and scientific articles related to the study of the telomerase holoenzyme complex were reviewed. The theoretical foundations of homology modeling, molecular dynamics and quantum-chemical calculations have been studied. Using the knowledge gained, a homology modeling of the fragment (TERT) of human telomerase carrying out reverse transcription was carried out. A molecular dynamic trajectory which simulates the behavior of this structure in time has been constructed and analyzed. On the basis of the constructed trajectory, an additional optimization of the model was carried out using the methods of quantum chemistry.

The scope: the resulting model of the spatial structure of telomerase can be used in other studies as a starting point for further analysis, as well as for the search for telomerase inhibitors as potential drugs with a wide spectrum of antitumor action.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	6
ГЛАВА 1. ПОСТРОЕНИЕ МОДЕЛИ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ ТЕЛОМЕРАЗЫ ЧЕЛОВЕКА С ИСПОЛЬЗОВАНИЕМ ГОМОЛОГИЧНОГО МОДЕЛИРОВАНИЯ.....	7
1.1 ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ	7
1.1.1 Перспективы использования теломеразы в качестве мишени для разработки лекарственных препаратов	7
1.1.2 Исследования структуры теломеразы в различных организмах	8
1.1.3 Гомологичные белки и моделирование белков по гомологии.....	9
1.2 ПОСТРОЕНИЕ МОДЕЛИ КАТАЛИТИЧЕСКОЙ СУБЪЕДИНИЦЫ ТЕЛОМЕРАЗЫ ЧЕЛОВЕКА.....	11
1.2.1 Моделирование каталитической субъединицы теломеразы человека по гомологии	11
1.2.2 Минимизация и проверка полученной структуры.....	15
ГЛАВА 2. ПОСТРОЕНИЕ МОЛЕКУЛЯРНО-ДИНАМИЧЕСКОЙ МОДЕЛИ КАТАЛИТИЧЕСКОЙ СУБЪЕДИНИЦЫ ТЕЛОМЕРАЗЫ ЧЕЛОВЕКА	17
2.1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ МОЛЕКУЛЯРНО-ДИНАМИЧЕСКОГО МОДЕЛИРОВАНИЯ	17
2.1.1 Базовые принципы молекулярной динамики	17
2.1.2 Методология подготовки системы для молекулярной динамики	18
2.2 МОЛЕКУЛЯРНО-ДИНАМИЧЕСКОЕ МОДЕЛИРОВАНИЕ КАТАЛИТИЧЕСКОЙ СУБЪЕДИНИЦЫ ТЕЛОМЕРАЗЫ ЧЕЛОВЕКА	20
2.2.1 Создание файлов с топологией системы и координатами атомов	20
2.2.2 Задание начального состояния системы и ее уравнивание	22
2.2.3 Расчет траекторий системы методами молекулярной динамики	24
ГЛАВА 3. КОНФОРМАЦИОННЫЙ ПОИСК И КВАНТОВО-ХИМИЧЕСКОЕ МОДЕЛИРОВАНИЕ	28
3.1 ЧИСЛЕННЫЕ МЕТОДЫ ОПТИМИЗАЦИИ	28
3.1.1 Численные методы оптимизации функций и их классификация	28
3.1.2 Метод наискорейшего спуска.....	29
3.1.3 Метод сопряженных градиентов	30
3.1.4 Алгоритм L-BFGS.....	31
3.2 ОПТИМИЗАЦИЯ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ МОДЕЛИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ КВАНТОВОЙ ХИМИИ.....	32
3.2.1 Базовые принципы квантовой химии.....	32
3.2.2 Выделение конформаций белка на основе траектории молекулярной динамики	34
3.2.3 Оптимизация конформаций белка с использованием методов квантовой химии	37
3.2.4 Анализ результатов.....	40
ЗАКЛЮЧЕНИЕ	43
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	44

ВВЕДЕНИЕ

Теломеры – это концевые некодирующие участки ДНК, состоящие из повторяющихся нуклеотидных последовательностей TTAGGG. Концы теломер покрыты шелтериновым гексабелковым комплексом [1], который выполняет защитную функцию, не допуская деградации и слияния концов хромосом [2]. В процессе деления клетки ДНК-полимераза не способна обеспечить полную репликацию концов хромосом, поэтому при каждом делении длина теломер сокращается. В конечном итоге длина теломер становится меньше некоторой определенной критической величины, что приводит к старению клетки и её гибели [3].

Для поддержания длины теломер служит специальный фермент – теломераза. Теломераза включает в себя особую обратную транскриптазу, работающую в комплексе с теломеразной РНК. Также в состав фермента входит ряд вспомогательных компонент, необходимых для функционирования в организме [4].

На сегодняшний день не существует точной модели пространственной структуры теломеразы человека. Сложность изучения теломеразы обусловлена ее чрезвычайно низким содержанием в клетке и трудностью получения отдельных компонентов в растворимой форме в достаточном количестве. При этом задача построения модели пространственной структуры теломеразы человека имеет большое значение в связи с тем, что данный фермент является крайне перспективной мишенью для противоопухолевой терапии, а также для разработки лекарств против ряда других тяжелых заболеваний.

Целью данной дипломной работы является построение модели пространственной структуры теломеразы человека на основе структурных данных родственных шаблонов. В частности, в данном случае в качестве исходных данных для гомологичного моделирования была использована пространственная структура каталитического ядра фермента теломеразы красного мучного жука *Tribolium castaneum*.

ГЛАВА 1

ПОСТРОЕНИЕ МОДЕЛИ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ ТЕЛОМЕРАЗЫ ЧЕЛОВЕКА С ИСПОЛЬЗОВАНИЕМ ГОМОЛОГИЧНОГО МОДЕЛИРОВАНИЯ

1.1 ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ

1.1.1 Перспективы использования теломеразы в качестве мишени для разработки лекарственных препаратов

Теломераза присутствует во всех клетках человеческого организма, однако в большинстве соматических клеток ее активность подавляется, поэтому их пролиферативный потенциал строго ограничен. В клетках раковых опухолей ситуация прямо противоположная – они могут делиться практически бесконечно. Подобный эффект достигается за счет того, что в них теломераза всегда активна и постоянно достраивает теломеры на концах хромосом и тем самым блокирует естественный механизм клеточного старения [5].

В ходе исследований были обнаружены отдельные виды раковых опухолей, в которых не наблюдалось высокой активности теломеразы, и выяснилось, что для подобных случаев характерна спонтанная ремиссия заболевания [6]. Этот факт говорит о том, что активная теломераза не является непосредственной причиной возникновения рака, однако является необходимым условием для роста и развития раковых опухолей.

Стратегия использования теломеразы при противоопухолевой терапии базируется на том, что теломеры в раковых клетках значительно короче, чем в других клетках организма, поэтому при подавлении активности теломеразы опухолевые клетки погибают быстрее остальных. Таким образом, образуется терапевтическое окно, позволяющее уничтожать раковые клетки, не нанося существенного вреда остальному организму [7].

Помимо ингибирования теломеразы, в терапевтических целях может использоваться и ее активация. Вторичным эффектом многих дегенеративных заболеваний является уменьшение длины теломер, и активация теломеразы в соответствующих клетках позволила бы эффективно бороться с такими болезнями [10]. Кроме того, есть данные о том, что в процессе старения утратившие способность к делению клетки долгое время могут оставаться жизнеспособными, выделяя при этом вещества, влияющие на возрастные заболевания [8]. Предполагается, что временное повышение активности теломеразы в соответствующих клетках могло бы повлиять на негативные

аспекты, связанные со старением организма.

Таким образом, исследование теломеразы человека является перспективной задачей с высокой практической значимостью, над которой трудятся ученые по всему миру. Однако, несмотря на большое количество исследований и научных статей, механизм действия теломеразы все еще изучен мало – доподлинно неизвестно ни точное строение фермента, ни факторы, влияющие на его активность, даже по поводу его реальной роли в организме ведутся споры. Построение качественной модели теломеразы дало бы мощный стимул для разработки новых лекарственных препаратов и позволило бы сильно продвинуться в борьбе со многими заболеваниями, связанными с активностью теломеразы, для которых на сегодняшний день не существует эффективного лечения.

1.1.2 Исследования структуры теломеразы в различных организмах

Данный раздел посвящен рассмотрению наиболее значительных результатов, полученных в ходе исследования пространственного строения теломеразы в различных организмах.

В 2007г. была получена структура РНК-связывающего домена (TRBD-домен) теломеразной обратной транскриптазы (TERT) для инфузории *Tetrahymena thermophila*. Данный домен является необходимым для работы фермента, и на то время структурных гомологов для него не существовало [9].

В 2008г. с помощью рентгеноструктурного анализа была получена полная структура каталитической субъединицы красного мучного жука *Tribolium castaneum*, имеющая кольцевидную форму. Сравнение TRBD-домена *Tribolium castaneum* с аналогичным доменом *Tetrahymena thermophila*, показало, что между ними присутствует определенное сходство [10].

Первая структура теломеразы в комплексе с ингибитором была определена в 2015г. Это была каталитическая субъединица *Tribolium castaneum* в комплексе с ингибитором VIBR1532. Данное исследование показало, что лиганд присоединяется к теломеразе вне активного центра фермента, что шло вразрез с имеющимися предположениями об ингибировании теломеразы [11]. Также новая структура имела более высокое разрешение по сравнению с предыдущей моделью каталитической субъединицы *Tribolium castaneum* – 2.3Å против 2.7Å, соответственно.

К сожалению, несмотря на весьма существенные результаты, полученные для других организмов, исследования структуры человеческой теломеразы пока что не позволили напрямую получить ее модель. В 2011г. была предпринята попытка построения пространственной структуры полного комплекса теломеразы человека при помощи гомологичного моделирования. Полученный результат позволил сделать важные выводы о расположении и взаимодействии некоторых структурных доменов фермента [12]. Также в 2018г. была продемонстрирована построенная методом криоэлектронной микроскопии модель холоферментного комплекса теломеразы человека, связанной с субстратом, позволяющая оценить архитектуру данного комплекса [13]. Однако, компоненты этого комплекса были представлены частично моделями, полученными на основе гомологичного моделирования с использованием шаблонов белков разных организмов, другие фрагменты комплекса были разрешены с помощью крио-ЭМ. Структура имеет довольно низкое разрешение в 10.2\AA , что затрудняет ее эффективное использование в качестве мишени с целью разработки к ней эффективных ингибиторов.

1.1.3 Гомологичные белки и моделирование белков по гомологии

Мутационный механизм молекулярной эволюции приводит к появлению семейств родственных белков, имеющих похожие аминокислотные последовательности и трехмерные структуры. Такие белки называются гомологичными. Следует отметить, что гомологичные белки чаще всего выполняют одну и ту же функцию в организме – именно этим фактом и объясняется сходство в их пространственном строении. При этом из-за большого количества мутаций, произошедших в процессе эволюции, первичная и вторичная структуры у таких белков могут довольно сильно различаться. Степень подобных расхождений обычно пропорциональна филогенетическому различию между соответствующими видами – чем более эволюционно далеки два вида, тем сильнее будут различия в аминокислотной последовательности их белков.

Определение аминокислотной последовательности белка на сегодняшний день уже является тривиальной задачей, чего нельзя сказать об определении его трехмерной структуры. В связи с тем, что экспериментальное установление трехмерной структуры – занятие долгое, сложное и дорогостоящее, для этой цели зачастую, а также ввиду отсутствия экспериментальных данных о структуре, используются различные теоретические методы. Одним из таких

методов является моделирование трехмерной структуры белков по гомологии. Данный метод базируется как раз на том факте, что родственные белки имеют схожее строение.

В своей традиционной форме моделирование по гомологии включает следующие этапы [14]:

1. Поиск родственных белков.
2. Выделение консервативных и переменных областей.
3. Выравнивание последовательностей исследуемого и шаблонного белков в консервативных областях.
4. Конструирование пространственной структуры консервативных областей с использованием координат структуры шаблонного белка.
5. Конструирование переменных областей.
6. Моделирование боковых цепей.
7. Оптимизация полученной структуры при помощи минимизации энергии или методов молекулярной динамики.

На первом шаге данного метода аминокислотная последовательность исследуемого белка сравнивается с последовательностями всех белков, для которых известно их пространственное строение. Если в ходе сравнения обнаруживается сходство с одним из белков более чем на 30%, то его структура в дальнейшем может быть использована в качестве шаблона для моделирования. Впрочем, если информация о семействе белка известна заранее, как часто и бывает, то этот этап можно пропустить.

В трехмерной структуре родственных белков часть областей имеет практически идентичную структуру, а часть может отличаться. Такие области называют, соответственно, консервативными и переменными. Для дальнейшего моделирования необходимо выделить консервативные и переменные области. Это может делаться различными способами – но чем больше известно структур родственных белков, тем точнее и проще становится этот процесс.

После того, как консервативные области выделены, в них можно выполнить выравнивание аминокислотных последовательностей и таким образом перенести последовательность исследуемого белка на трехмерную структуру белка-шаблона. Однако необходимо смоделировать еще и структуру

вариабельных областей. Обычно для этого применяется либо поиск схожих сегментов в других белках, либо моделирование *de novo*.

После выполнения всех описанных выше этапов уже имеется модель основной цепи белка. Следующим шагом является добавление боковых цепей. Это также довольно нетривиальная задача, поскольку многие боковые цепи имеют хотя бы одну степень свободы и могут находиться в нескольких возможных конформациях. Одним из вариантов решения является подход, основанный на том факте, что боковые цепи обычно принимают лишь некоторые из множества допустимых конформаций, и можно использовать наиболее вероятные из них. Однако при этом необходимо учитывать, что реальная конформация боковой цепи зависит также от локального окружения аминокислоты, вторичной структуры, контактов с другими остатками и еще ряда факторов.

Окончательная модель, полученная в результате гомологичного моделирования, с большой долей вероятности содержит зоны стерических напряжений и неоптимально расположенные боковые цепи. Поэтому такая модель обычно нуждается в уточнении и дополнительной оптимизации структуры. Для этого применяется минимизация энергии и молекулярная динамика. Минимизация энергии позволяет сравнительно быстро обнаружить конформацию, соответствующую локальному энергетическому минимуму, однако данный метод не способен преодолевать энергетические барьеры, разделяющие локальные минимумы на поверхности потенциальной энергии молекулы. Молекулярная динамика по сравнению с минимизацией требует гораздо более серьезных вычислительных мощностей, но дает возможность найти оптимальную геометрию системы за счет более полного исследования конформационного пространства.

1.2 ПОСТРОЕНИЕ МОДЕЛИ КАТАЛИТИЧЕСКОЙ СУБЪЕДИНИЦЫ ТЕЛОМЕРАЗЫ ЧЕЛОВЕКА

1.2.1 Моделирование каталитической субъединицы теломеразы человека по гомологии

Для постановки начальной задачи моделирования необходимо понимать, какие структурные единицы входят в состав теломеразы. Основу теломеразы образует так называемый коровый фермент. Коровый фермент обеспечивает условия для функционирования теломеразы *in vitro*, то есть в лабораторных

условиях. Для обеспечения активности теломеразы в живом организме (*in vivo*) кроме корового фермента необходимо так же присутствие рибонуклеопротеидной структуры, состоящей из ряда специальных белков и взаимодействующих с ними факторов. Данная структура обеспечивает непосредственное связывание теломеразы с теломерной последовательностью хромосомы.

В свою очередь коровый фермент состоит из теломеразной РНК (TER) и теломеразной обратной транскриптазы (TERT). Теломеразная РНК содержит матричный участок, а также несколько других функционально важных элементов вторичной структуры. Теломеразная обратная транскриптаза представляет собой белок, содержащий каталитически важный домен, поэтому ее также часто называют каталитической субъединицей теломеразы. Кроме того, в TERT также входит ряд других характерных доменов, принимающих участие в связывании TER и ДНК-субстрата [15].

Таким образом, теломераза представляет собой сложный комплекс, и известные методы не позволяют провести моделирование подобной структуры целиком. При этом следует отметить, что наиболее важной функциональной частью теломеразы является каталитическая субъединица. Поэтому вместо моделирования полной структуры теломеразы человека было решено ограничиться построением модели ее каталитической субъединицы.

В качестве родственного белка-шаблона для моделирования использовалась структура каталитической субъединицы теломеразы мучного жука *Tribolium castaneum* (PDB 5CQG). Выбор этой структуры обусловлен тем, что она имеет самое высокое разрешение из всех доступных на данный момент моделей теломеразы. Аминокислотная последовательность TERT человека была взята из открытой базы данных белковых последовательностей UniProt.

Аминокислотные последовательности каталитических субъединиц у жука и человека сильно различаются (по данным сервиса SWISS-MODEL подобие составляет всего 22,2%), поэтому было принято решение при осуществлении гомологичного моделирования использовать не всю последовательность целиком, а только ее часть. Для моделирования был оставлен фрагмент последовательности TERT человека, соответствующий структурному домену STE (thumb-domain).

Процесс моделирования выполнялся при помощи веб-сервиса SWISS-MODEL [16]. В результате была получена модель, по пространственной структуре точно повторяющая каталитическую субъединицу *Tribolium*

castaneum, однако аминокислотная последовательность домена СТЕ в ней уже соответствовала человеческой.

Далее было необходимо провести моделирование для остальных структурных доменов каталитической субъединицы. Основным препятствием на этом этапе стало то, что для аминокислотной последовательности TERT человека на рассматриваемых участках не удалось найти хорошего гомолога, поэтому была велика вероятность получить на выходе очень далекую от реальности структуру. Чтобы избежать этого, было принято решение в полученной на предыдущем этапе модели заменить все остатки теломеразы жука на остатки теломеразы человека вручную. Для осуществления подобной замены необходимо построить выравнивание последовательностей, и подобная работа уже была проделана ранее в другом исследовании, поэтому мы воспользовались готовым результатом (рисунок 1.1) [13].

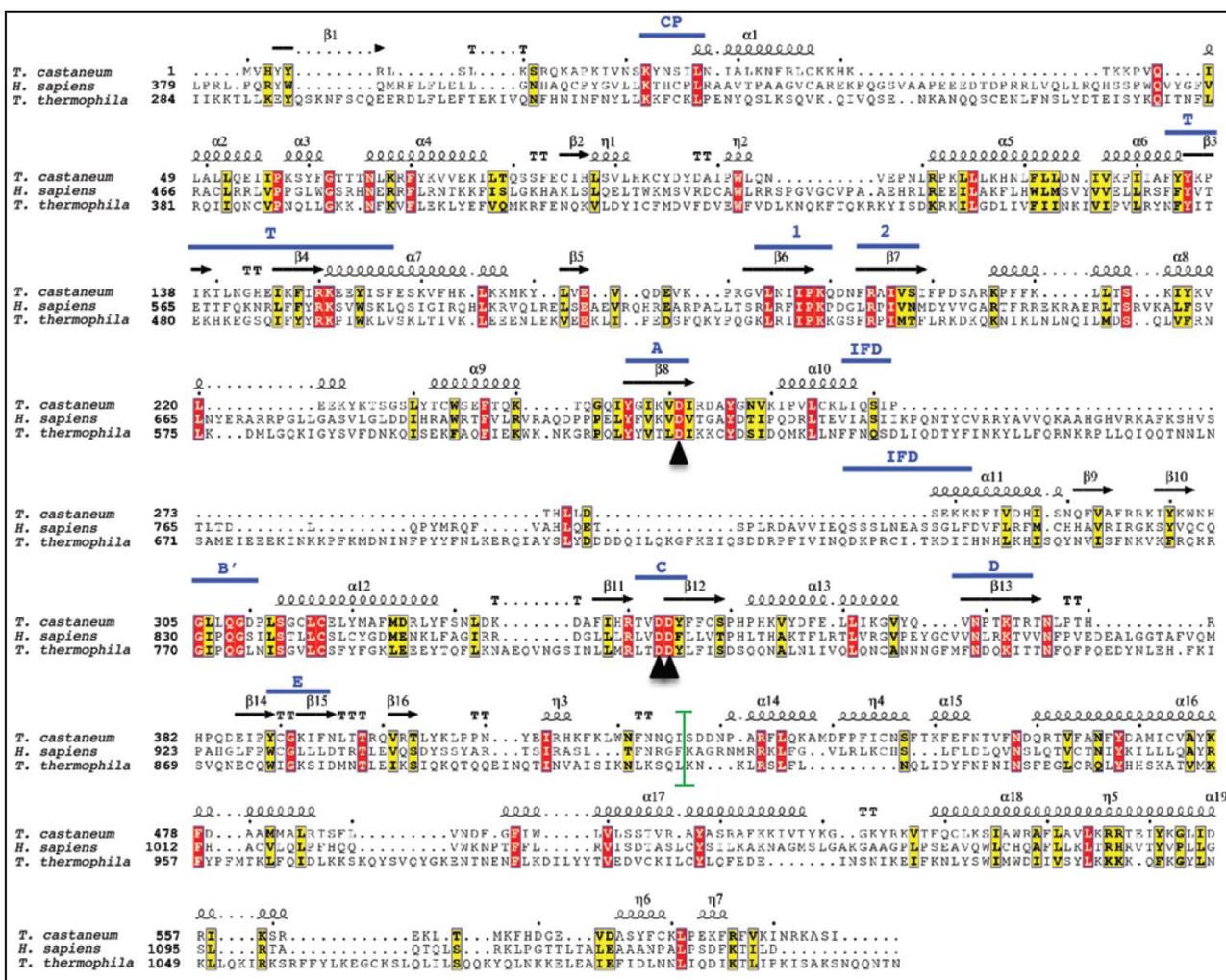


Рисунок 1.1 - Выравнивание последовательностей TERT человека, *Tribolium castaneum* и *Tetrahymena thermophila*. Зеленой чертой отмечено начало СТЕ домена

Отдельно следует упомянуть остатки, для которых не нашлось пары в выравнивании. На основании известных данных о структуре теломеразы можно предположить, что большинство таких остатков либо является случайными мутациями, либо входит в отдельные известные мотивы – например, у человека есть отсутствующий у *Tribolium castaneum* участок IFD (рисунок 1.2), который

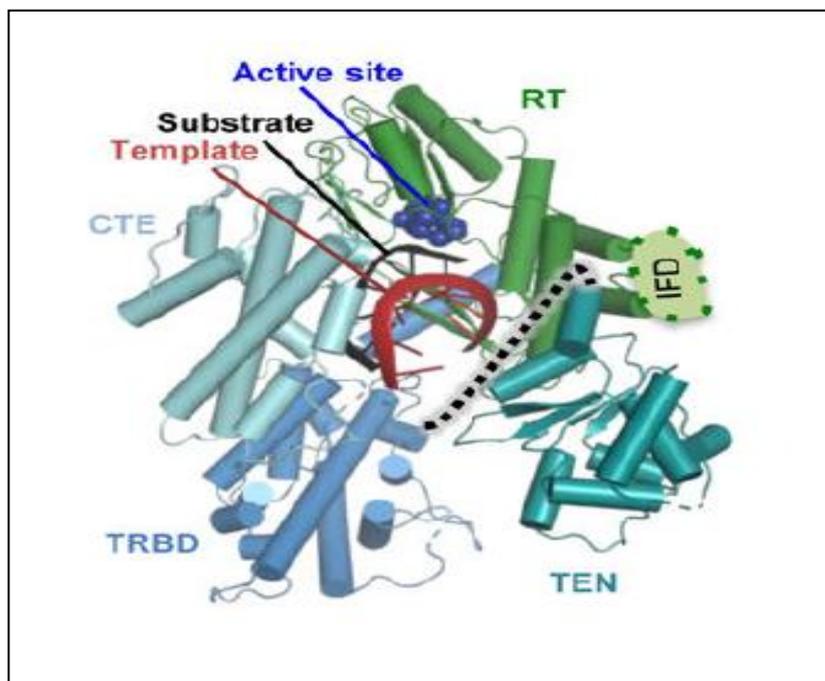


Рисунок 1.2 – Структура связанной с субстратом каталитической субъединицы *Tribolium castaneum* с выделенными структурными доменами.

расположен между спиральями $\alpha 10$ и $\alpha 11$ (рисунок 1.1). Такие участки обычно располагаются близко к поверхности белка, и не являются критически важным для работы фермента. В то же время моделирование строения подобных участков затруднительно из-за недостатка данных. Поэтому данные участки были исключены из рассмотрения для получения более надежной структуры.

Итоговая структура, полученная после замены остатков, полностью повторяет пространственное строение каталитической субъединицы *Tribolium castaneum*, однако ее аминокислотная последовательность максимально возможно приближена к последовательности теломеразы человека.

После того, как была произведена замена всех остатков, необходимо было изменить их нумерацию, чтобы она соответствовала нумерации соответствующих остатков в теломеразе человека. Для этой цели был написан скрипт на языке Python, который принимает на вход файл с белком в формате PDB и данные об изменении номеров остатков в заданном формате, после чего на выходе получается файл PDB с уже измененной нумерацией.

1.2.2 Минимизация и проверка полученной структуры

Для того чтобы полученная в результате моделирования структура имела более правильное пространственное строение, ее необходимо минимизировать. Для минимизации использовался встроенный функционал программы Chimera. Сначала реализованный в программе метод добавляет к структуре водороды и рассчитывает заряды атомов. После этого сначала выполняется минимизация энергии методом наискорейшего спуска, что позволяет устранить наиболее нежелательные положения элементов, а затем – минимизация методом сопряженного градиента, которая работает значительно дольше, но позволяет получить более точный результат. Для обоих методов было выполнено по 500 шагов, начальный размер шага был установлен на $0,02\text{\AA}$. Если сравнить структуры до и после минимизации (рисунок 1.3), то можно заметить, что изменения локализованы преимущественно в областях петель и незначительны.

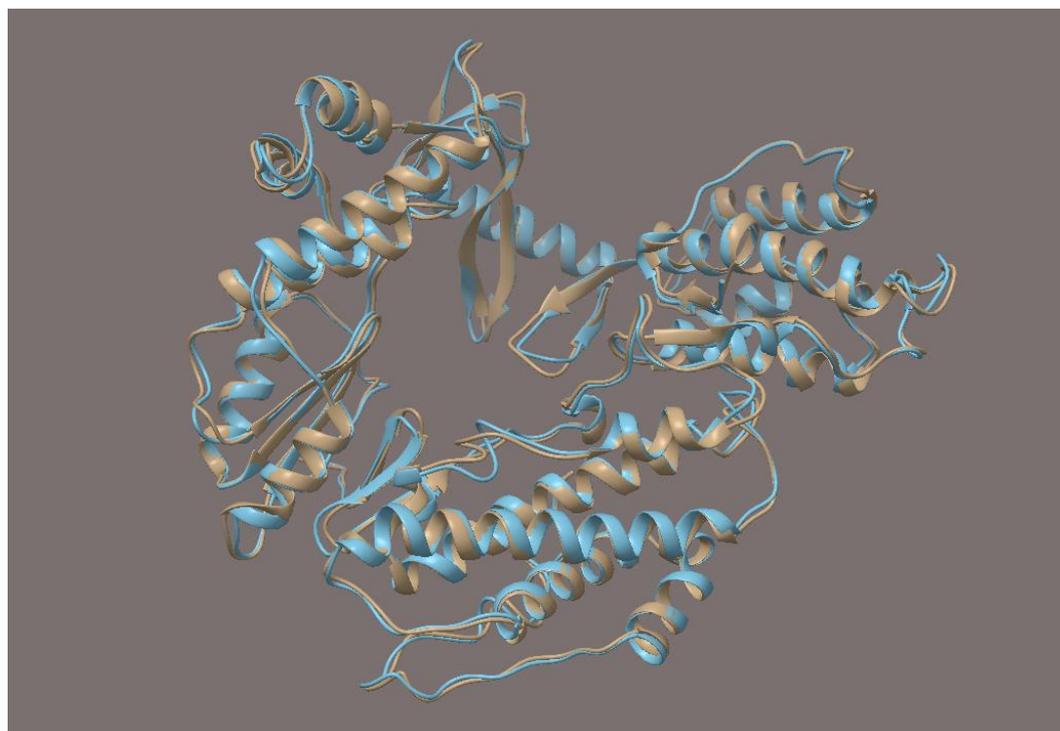


Рисунок 1.3 – Сравнение структур до минимизации энергии (синий цвет) и после минимизации (песочный цвет)

Последним действием на данном этапе является оценка качества полученной структуры. Это необходимо для того, чтобы убедиться, что модель правдоподобна с химической точки зрения и не имеет явных дефектов. Для этой цели использовался программный пакет PROCHECK [17]. Данный пакет обеспечивает детальную проверку стереохимии белковой структуры. В частности, с его помощью можно оценить смоделированную структуру с

помощью карты Рамачандрана – визуальное представление торсионных углов белковой молекулы. Карта Рамачандрана позволяет оценить количество аминокислотных остатков, находящихся в стерически запрещенных положениях, и на основе этих данных сделать выводы о качестве модели. В нашем случае в оптимальные зоны карты попало почти 79% остатков (рисунок 1.4). Оценка этого результата производится с учетом разрешения структуры – 2,3Å. В среднем по данным PROCHECK в качественных структурах подобного разрешения около 81% остатков находятся в оптимальных зонах карты Рамачандрана, поэтому можно считать, что построенная модель не имеет значительных дефектов с точки зрения геометрии молекулы.

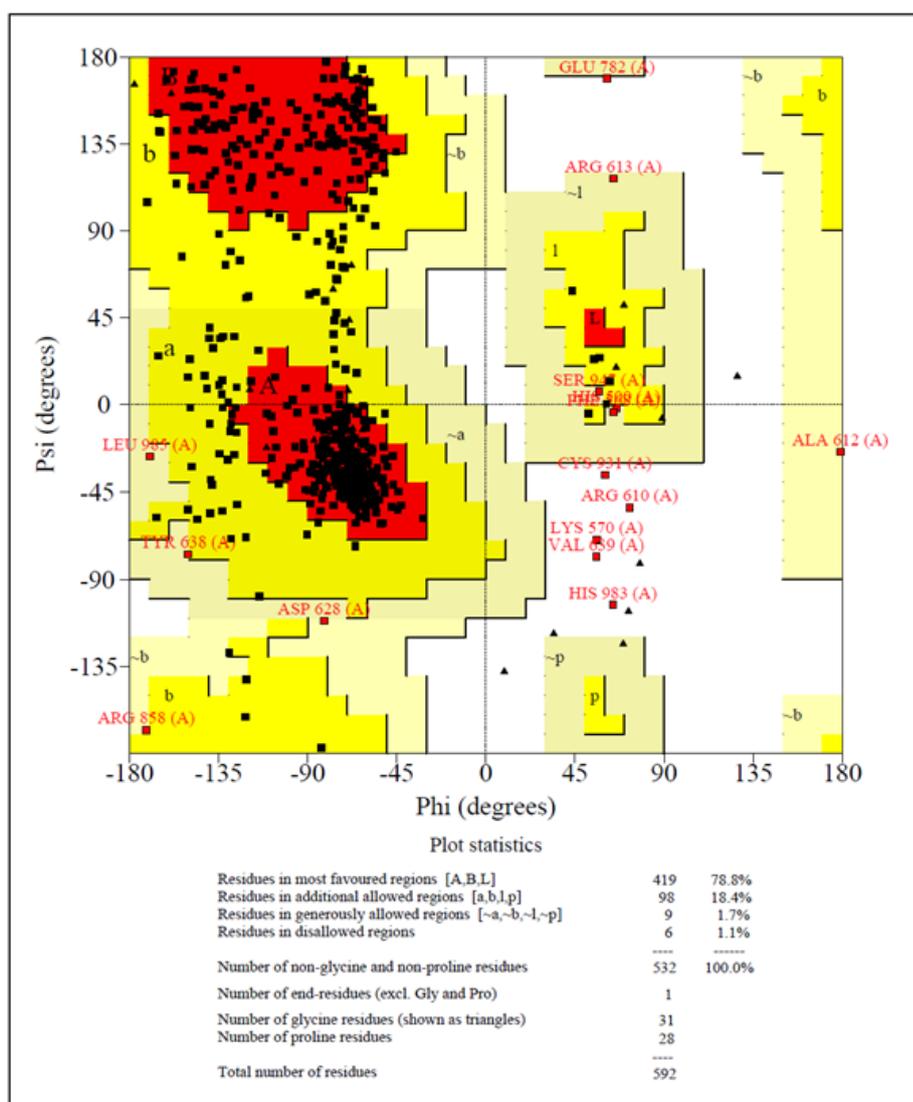


Рисунок 1.4 – Карта Рамачандрана для построенной модели

ГЛАВА 2

ПОСТРОЕНИЕ МОЛЕКУЛЯРНО-ДИНАМИЧЕСКОЙ МОДЕЛИ КАТАЛИТИЧЕСКОЙ СУБЪЕДИНИЦЫ ТЕЛОМЕРАЗЫ ЧЕЛОВЕКА

2.1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ МОЛЕКУЛЯРНО- ДИНАМИЧЕСКОГО МОДЕЛИРОВАНИЯ

2.1.1 Базовые принципы молекулярной динамики

Белки и их комплексы допускают множество конформаций, с которыми связаны их биологические функции. Несмотря на то, что в какой-то мере их структуры могут быть исследованы методами рентгеноструктурного анализа, ЯМР-спектроскопии или криоэлектронной микроскопии, эти методы предоставляют только информацию о статическом состоянии белка, но при этом не способны описать конформационные переходы и динамику. Колебательные движения в белках представлены довольно широким временным диапазоном: от 10^{-14} секунд (атомы связи C – H) до 10^{-9} секунд (конформационный переход). Для всестороннего анализа таких процессов были разработаны методы молекулярной динамики. С помощью вычислительных мощностей, предоставляемых современной техникой, молекулярная динамика позволяет с высокой точностью описывать и предсказывать динамику любых систем.

Основная идея, лежащая в основе методов молекулярной динамики – предсказание поведения каждого отдельного атома системы как функции от времени. При этом молекула может быть представлена как множество точек, имеющих электрический заряд (атомы) и соединенных пружинами (межатомные связи). Чтобы предсказать поведение длин и углов связей, а также взаимодействия между атомами, используются силовые поля. Силовое поле представляет собой набор математических уравнений, описывающих потенциальную энергию взаимодействия атомов.

Для моделирования динамики системы в течение определенного промежутка времени виртуально создаются условия, в которых модель белка помещается в растворитель (обычно в роли растворителя выступает вода). Затем на основе силовых полей рассчитываются скорости и направления движения атомов системы, и атомы сдвигаются в рассчитанных направлениях в течение очень короткого промежутка времени (порядка 10-15 пс) [18]. Затем те же действия повторяются уже с учетом нового положения атомов.

Совокупность координат атомов системы на каждом шаге молекулярной динамики формирует так называемую траекторию системы, непосредственно описывающую поведение системы на заданном временном промежутке. Состояние системы на отдельном шаге называется фреймом траектории.

При этом нельзя также отрицать того факта, что методы молекулярной динамики обладают рядом недостатков:

- моделирование базируется на законах Ньютона, то есть для описания функции потенциальной энергии используется не квантовая механика, а классическая;
- молекулярная динамика не учитывает электронную структуру молекулы, атомы представляются единичными неделимыми частицами;
- удаленные взаимодействия обрезаются, что тоже вносит некоторые погрешности;
- граничные условия между ячейками недостаточно реалистичны.

2.1.2 Методология подготовки системы для молекулярной динамики

Общая схема подготовки системы для моделирования молекулярной динамики выглядит следующим образом:

- построение топологии молекулы;
- выбор формы и размера ячейки для моделирования;
- добавление в ячейку растворителя и ионов;
- минимизация энергии системы;
- задание начальных параметров системы (температура, давление и т.д.) и уравнивание системы с учетом этих параметров).

Рассмотрим каждый из этих пунктов подробнее. Сначала необходимо построить топологию молекулы. Топология включает в себя информацию о типах атомов, связях между ними, параметрах силовых полей и т.д. Многие пакеты для молекулярно-динамического моделирования имеют в своем составе специальные программы, которые позволяют осуществлять подобную подготовку. В частности силовые поля пакета AMBER представляют потенциал взаимодействия атомов на основе формулы (2.1):

$$\begin{aligned}
 U = & \sum_{banded} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\phi [\cos(n\phi + \delta) + 1] \\
 & + \sum_{non-banded} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} \right] \quad (2.1)
 \end{aligned}$$

Здесь первая сумма выражает энергию деформации длины ковалентной связи, вторая – энергию изменения валентных углов, третья – энергию вращения двугранных углов между связями, четвертая – энергию взаимодействия атомов, не связанных ковалентной связью (состоит из энергии электростатического взаимодействия и потенциала Леннарда-Джонса).

Выбор формы и размера ячейки для моделирования осуществляется с таким расчетом, чтобы она включала в себя исследуемую молекулу, и при этом оставалось достаточно пространства для молекул растворителя. Нужно понимать, что слишком большой размер ячейки приведет к значительному объему ненужных вычислений и замедлит работу. При этом слишком маленький размер ячейки может негативно повлиять на качество итогового результата.

Кроме того, чтобы повысить качество моделирования, но при этом избежать лишних вычислений, часто используется подход, называемый периодическими граничными условиями [19]. Такой подход позволяет избежать различных неточностей, связанных с поведением атомов у границы ячейки. Идея заключается в том, что ячейка повторяется многократно во всех направлениях, формируя многомерную сеть (рисунок 2.1).

При таком подходе каждая частица представлена в единственном экземпляре, но расчеты производятся с учетом образов, а любые изменения, связанные с основной частицей, отразятся и на ее образах. Если частица покидает основную ячейку, то считается, что она просто заново входит в нее с противоположной стороны, за счет чего количество частиц в каждой ячейке остается постоянным.

Растворитель в ячейке может быть задан двумя основными способами – явным и неявным. В случае явного задания в систему явным образом вносятся молекулы растворителя, принимающие участие во взаимодействиях. При неявном способе задания растворитель представляется в виде непрерывной среды, обладающей некоторым набором усредненных параметров, моделирующих свойства реальной воды. Кроме растворителя, в ячейку также

добавляют положительно и отрицательно заряженные ионы, чтобы сделать общий заряд системы нейтральным.

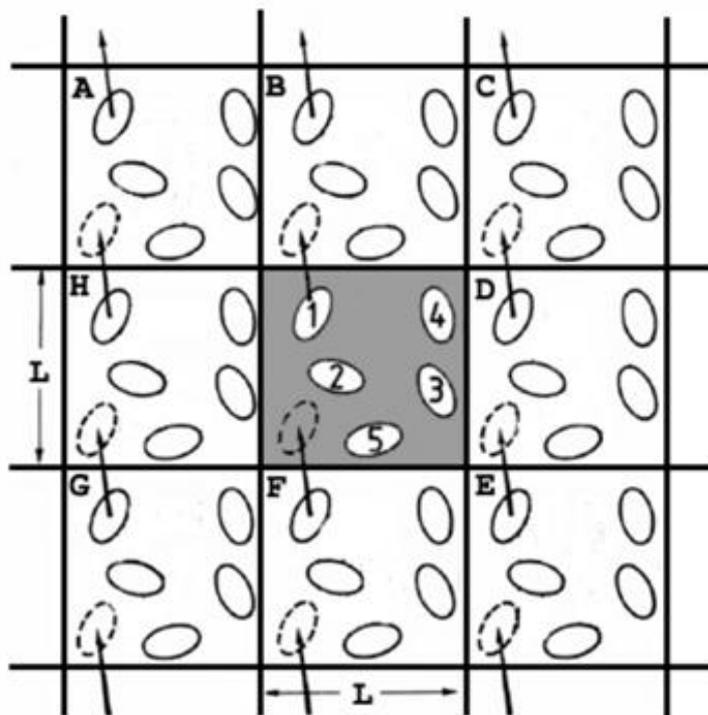


Рисунок 2.1 – Схема периодических граничных условий

После этого необходимо задать начальные параметры системы, и дать ей прийти в равновесное состояние с учетом этих условий. Обычно молекулярно-динамические расчеты проводятся в изобарно-изотермических условиях, так как такие условия наиболее близки к реальным условиям протекания исследуемых процессов. Для контроля температуры и давления используются термостаты и баростаты, на самом деле обычно представляющие собой соответствующие математические модели. Этот этап, по сути, представляет собой симуляцию молекулярной динамики, просто на более коротком промежутке времени и в специфичных условиях.

2.2 МОЛЕКУЛЯРНО-ДИНАМИЧЕСКОЕ МОДЕЛИРОВАНИЕ КАТАЛИТИЧЕСКОЙ СУБЪЕДИНИЦЫ ТЕЛОМЕРАЗЫ ЧЕЛОВЕКА

2.2.1 Создание файлов с топологией системы и координатами атомов

Необходимым условием для осуществления моделирования с помощью

пакета AMBER является наличие файлов, описывающих топологию и координаты атомов исследуемой системы. Под топологией здесь подразумевается информация о типах атомов, связях между ними, параметрах силовых полей и т.д. Параметры, описывающие топологию системы, в ходе симуляции молекулярной динамики остаются постоянными.

Для создания этих файлов была использована входящая в пакет AMBER программа LEaP. Данная программа заменяет собой несколько более старых модулей, ранее являвшихся частью AMBER, и на данный момент является базовым решением при подготовке входных данных для моделирования молекулярной динамики. LEaP содержит в себе набор команд, который позволяет читать и записывать файлы в разных форматах, производить различные операции с атомами и аминокислотными остатками, конструировать силовые поля.

Перед запуском LEaP из начальной модели теломеразы необходимо удалить все атомы водорода. Дело в том, что LEaP автоматически добавляет атомы водорода к системе, что может привести к ошибке, если данные атомы уже присутствуют во входных файлах.

После того, как атомы водорода удалены из начального файла, производится запуск программы LEaP с набором команд, определяющим необходимые преобразования системы.

На начальном этапе необходимо загрузить файлы с информацией о силовом поле, которое будет использоваться для создания топологии системы. Для моделирования было выбрано силовое поле ff14SB, представляющее собой продукт развития силового поля ff99SB. В ff14SB было пересмотрено использование многих параметров, связанных с торсионными углами главной цепи, а также были сбалансированы параметры основных элементов вторичной структуры белка [21]. Эти изменения позволили добиться значительно более точных результатов при дальнейшем моделировании.

Следующим этапом предварительной подготовки системы является создание ячейки с растворителем и ионами, в которую помещается целевая структура. Ячейка формируется следующим образом – сначала определяется минимальная оболочка в форме прямоугольного параллелепипеда, содержащая растворяемую структуру. Затем к полученным размерам со всех сторон добавляется величина отступа от границы (значение этого параметра было выбрано равным 10Å). После этого в ячейку добавляются молекулы растворителя. В качестве растворителя была выбрана модель воды TIP3P, в

которой молекула воды представлена одним атомом кислорода и двумя атомами водорода с соответствующими зарядами и потенциалами Леннарда-Джонса. В ходе этого процесса итоговый размер ячейки может немного увеличиться, так как растворитель должен эффективно заполнить весь предоставленный объем, не оставляя пустых участков. Кроме того, в ячейку также добавляются ионы Na^+ и Cl^- для того, чтобы уравновесить суммарный заряд системы и сделать его нейтральным. Итоговый размер ячейки с теломеразой, водой и ионами получился равным $112,5\text{Å} \times 92,1\text{Å} \times 99,2\text{Å}$.

После выполнения всех вышеперечисленных этапов программа LEaP завершает свою работу. В качестве итогового результата генерируются два файла с расширениями *.prmtop* и *.inpcrd*, содержащие информацию о топологии системы и координаты входящих в нее атомов, соответственно.

2.2.2 Задание начального состояния системы и ее уравнивание

Входные файлы, описывающие систему, являются необходимым условием для моделирования молекулярной динамики. Однако чтобы получить качественный результат при моделировании, необходимо дополнительно осуществить предварительную подготовку системы. Здесь под предварительной подготовкой подразумевается минимизация системы и задание ее начального состояния (температура, плотность и т.д.). Для этой цели была использована входящая в пакет AMBER программа *sander*, предоставляющая широкий набор возможностей по минимизации энергии и симуляции молекулярной динамики.

Первым шагом на этапе предварительной подготовки является минимизация энергии системы. Минимизация позволяет избежать ситуаций, когда начальная конформация содержит перекрывающиеся атомы и группы атомов а также зоны сильных стерических напряжений. В этих случаях начальная энергия системы может быть очень велика, что приводит к ошибкам в ходе моделирования.

Минимизация энергии позволяет сравнительно быстро обнаружить конформацию, соответствующую локальному энергетическому минимуму, однако данный метод не способен преодолевать энергетические барьеры, разделяющие различные конформации белка. В ходе эксперимента минимизация проводилась сначала с помощью метода наискорейшего спуска, а затем с помощью метода сопряженного градиента. В методе наискорейшего

спуска шаги делаются в направлении градиента функции потенциальной энергии, предыдущие шаги при этом не учитываются. Этот метод хорош при оптимизации вдали от минимума, так как размер шага рассчитывается исходя из величины градиента, однако вблизи локального минимума скорость сходимости метода наискорейшего спуска сильно снижается. Метод сопряженных градиентов учитывает также и предыдущие шаги, что позволяет эффективно использовать его в окрестности локальных минимумов. Таким образом, эти два метода хорошо дополняют друг друга – сначала метод наискорейшего спуска быстро приближается к точке локального минимума энергии, а затем положение минимума уточняется методом сопряженных градиентов.

Рассмотрим параметры, которые использовались при минимизации. Для обоих методов количество шагов алгоритма было установлено равным 500. Кроме этого, были добавлены условия, исключаяющие водородные связи из подсчета энергии, а также ограничивающие изменение длин водородных связей ($n_{tf} = n_{tc} = 2$). Чтобы избежать артефактов, связанных с границами ячейки, были использованы периодические граничные условия ($n_{tb} = 1$). Кроме того, были добавлены позиционные ограничения на атомы главной цепи белка (Ca, C, O, N).

Следующим этапом после выполнения минимизации является постепенный нагрев системы до температуры, при которой будет выполняться моделирование. Нагрев производился от 0К до 300К. При этом был использован тот же набор ограничений, что и при минимизации системы.

Для того чтобы контролировать температуру системы в рамках молекулярно-динамического подхода используется абстрактная концепция, получившая название термостата. Существует несколько различных вариантов термостатов, но в данном исследовании использовался термостат Ланжевена [21]. Динамика Ланжевена принимает в расчет трение, создаваемое молекулами растворителя, а также возмущения системы, возникающие при столкновении частиц. В частности, в ее основе лежит уравнение Ланжевена (2.2):

$$MX'' = -\nabla U(x) - \gamma MX' + \sqrt{2M\gamma k_B T} R(t) \quad (2.2)$$

Здесь M – масса частицы; X – координата частицы (соответственно X' – скорость и X'' – ускорение); $-\nabla U(x)$ – слагаемое, соответствующее силе электростатического взаимодействия; γ – постоянная демпфирования, имеющая размерность, обратную времени (параметр модели, в нашем случае было выбрано значение $2c-1$); k_B – постоянная Больцмана; T – температура; $R(t)$ –

стационарный гауссовский случайный процесс с нулевым математическим ожиданием. Таким образом, все последнее слагаемое отвечает за учет трения и случайных взаимодействий, порождаемых столкновениями частиц.

Третий этап подготовки системы – стабилизация плотности растворителя с целью создать постоянное давление и избежать наличия “полостей” в растворе из-за неоднородного распределения частиц. Параметры и ограничения на данном этапе были выбраны аналогично предыдущим. Наибольшее значение тут имеет факт использования периодических граничных условий, что позволяет объему ячейки изменяться, тем самым стабилизируя давление и плотность. Таким образом, моделируемая система представляет собой так называемый NPT ансамбль, имеющий константное количество частиц, давление и температуру.

Этапы стабилизации температуры и плотности моделировались на протяжении 50пс каждый. После этого следует последний этап уравнивания длиной 500пс – его задачей является гарантия того, что система действительно находится в стабильном состоянии.

2.2.3 Расчет траекторий системы методами молекулярной динамики

После выполнения всех предыдущих этапов мы имеем начальную конформацию системы, которая является стартовой точкой уже для непосредственного проведения расчетов методом молекулярной динамики.

Так как моделирование траектории проводится уже на значительно более длинном участке по сравнению с моделированием в ходе уравнивания системы, для того, чтобы расчеты можно было завершить за разумное время, нужны более высокие вычислительные мощности. Поэтому на этом этапе вычисления проводились на видеокарте Nvidia GeForce RTX 2080 Ti.

Длина моделируемой траектории составила 100нс, параметры моделирования остались такими же, как и при уравнивании системы. Из полученной траектории была взята последняя конформация, соответствующая времени в 100нс, для сравнения с исходной структурой (рисунок 2.2). Можно заметить, что они сохранили сильное сходство, что подтверждает сделанное изначально предположение о схожести пространственных структур теломеразы мучного жука и человека. Наблюдается сдвиг одного из структурных доменов (верхний правый угол на рисунке 2.2). С одной стороны, это может быть

обычной сменой конформации, но с другой стороны возможно также и то, что это указывает на начало распада кольцеобразной структуры каталитической субъединицы теломеразы.

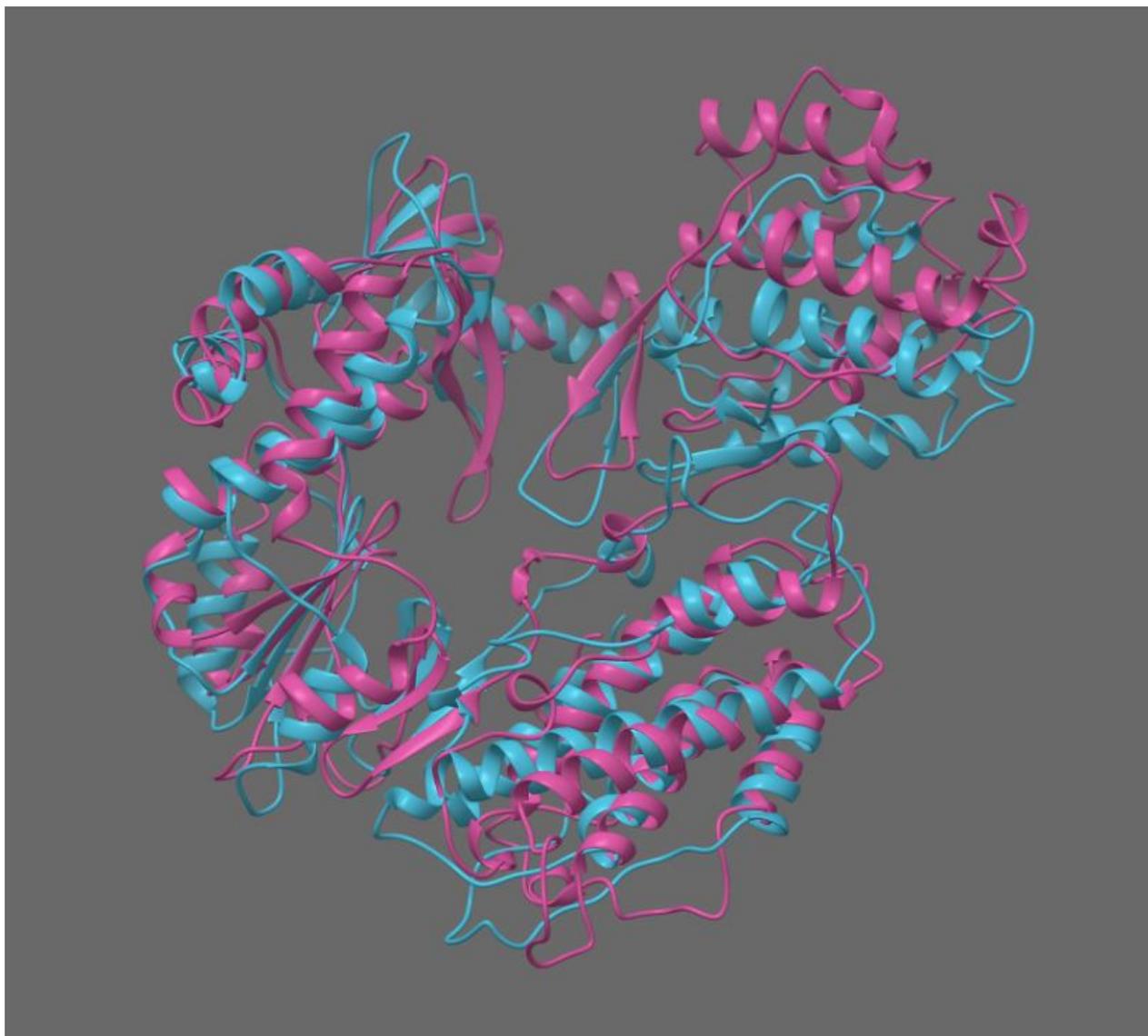


Рисунок 2.2 – Сравнение финальной конформации на 100 нс (розовый цвет) и исходной структуры (синий цвет)

Действительно, использование гомологичного моделирования не может гарантировать устойчивость полученной структуры. Поэтому для того, чтобы проверить, что структура теломеразы стабильна, было принято решение повторить моделирование, увеличив длину траектории со 100нс до 200нс, для того, чтобы данные о поведении системы были более полными.

Для анализа полученной в результате моделирования траектории был построен график среднеквадратичного отклонения от исходной структуры (рисунок 2.3). Предварительно из файла с траекторией были удалены молекулы растворителя и нейтрализующие ионы, а также выполнено выравнивание каждого из фреймов на исходную структуру, чтобы исключить из расчетов поступательный сдвиг структуры в пространстве с течением времени.

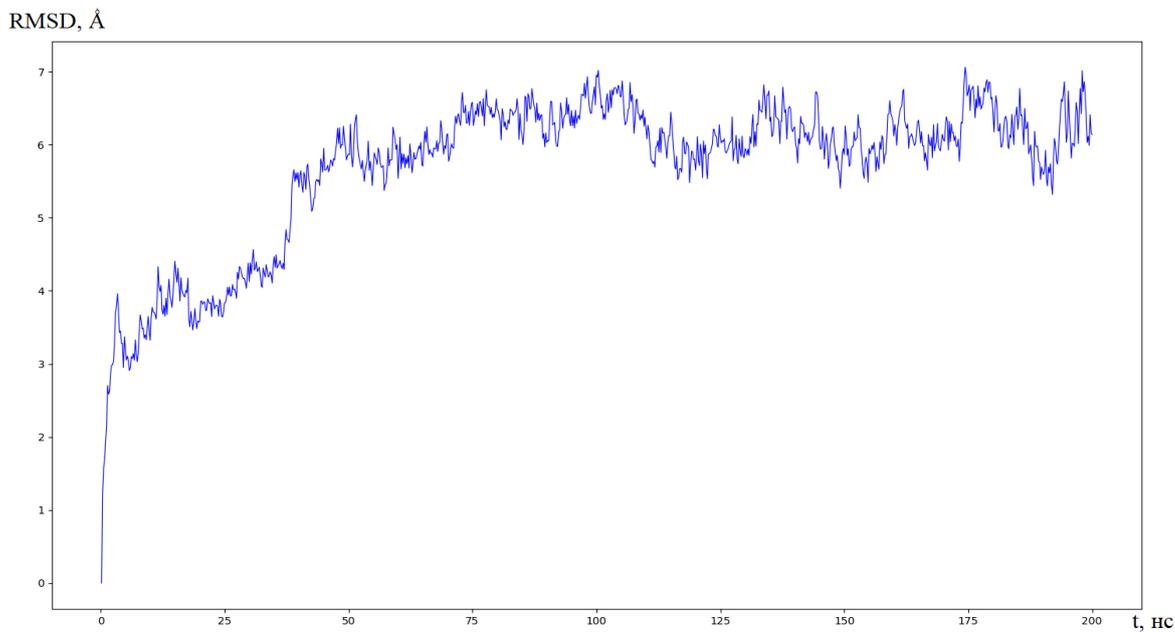


Рисунок 2.3 – График среднеквадратичного отклонения траектории от исходной структуры

Видно, что сначала состояние системы остается относительно стабильным примерно до 40 нс. Затем происходит резкий скачок, и система вновь занимает стабильное положение, на что указывает отсутствие дальнейших перепадов в графике. Такое поведение исследуемой метрики свидетельствует в пользу гипотезы о том, что наблюдаемый сдвиг домена представляет собой обычную смену конформации белка, а сама молекула при этом сохраняет структурную целостность.

Для того чтобы отдельно проанализировать колебания каждого из аминокислотных остатков используется метрика среднеквадратичного колебания (root mean square fluctuation, RMSF), вычисляемая по формуле (2.3):

$$RMSF_i = \left[\frac{1}{T} \sum_{t_j=1}^T |r_i(t_j) - r_i^{ref}|^2 \right]^{1/2} \quad (2.3)$$

где t_j – моменты времени, соответствующие фреймам траектории, $r_i(t_j)$ –

позиция частицы i в момент времени t_j , r_i^{ref} – начальная позиция. Полученный график можно видеть на рисунке 2.4.

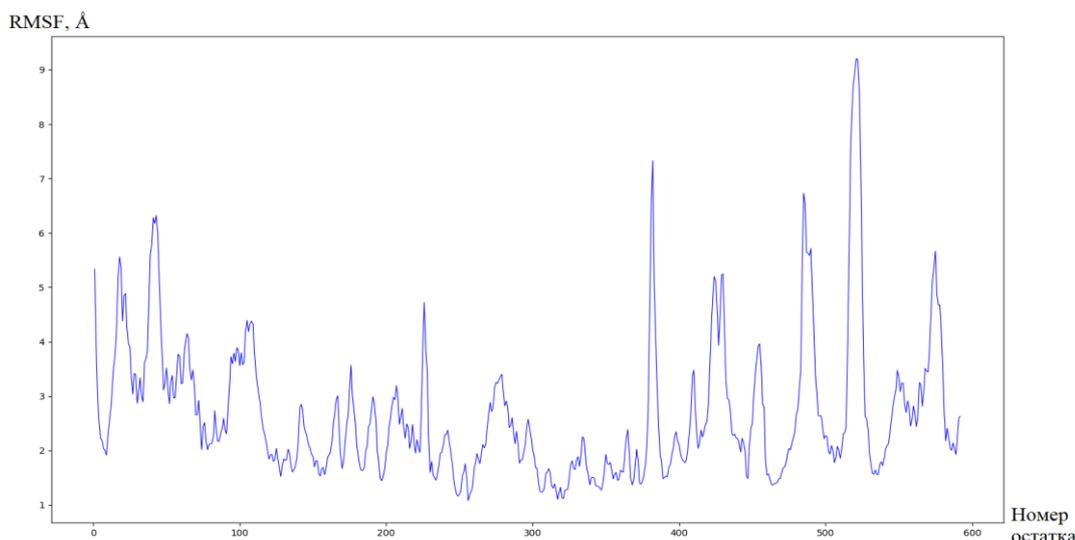


Рисунок 2.4 – График среднеквадратичного колебания аминокислотных остатков

Также был построен график зависимости радиуса инерции (radius of gyration, ROG) молекулы от времени (рисунок 2.5). Данная характеристика отражает компактность белковой структуры и вычисляется по формуле (2.4):

$$R_g = \left(\frac{\sum_i |r_i|^2 m_i}{\sum_i m_i} \right)^2 \quad (2.4)$$

где m_i – масса атома i , а r_i – расстояние от атома i до центра масс молекулы.

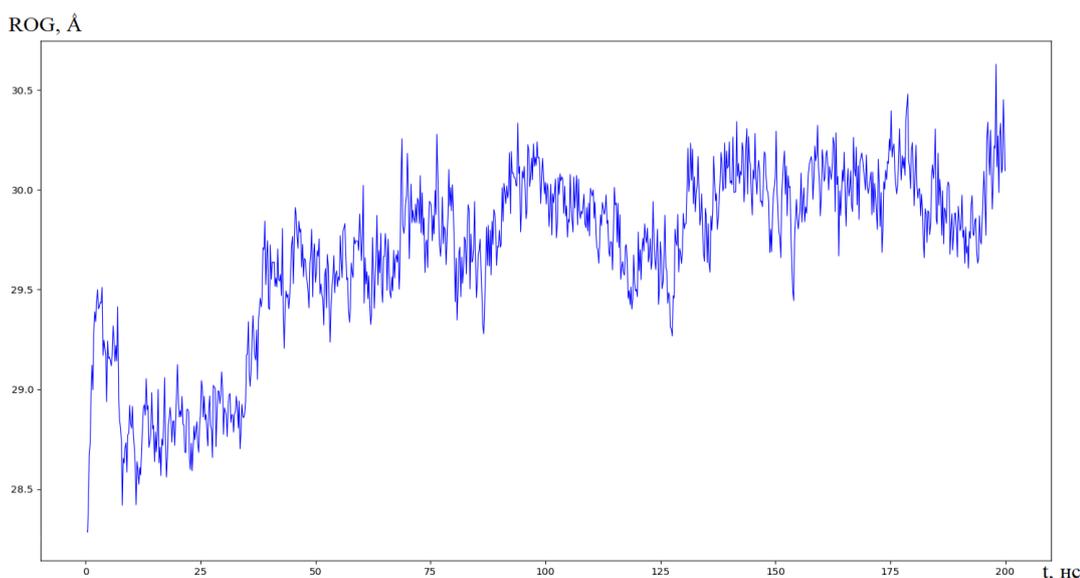


Рисунок 2.5 – График зависимости радиуса инерции от времени моделирования

ГЛАВА 3

КОНФОРМАЦИОННЫЙ ПОИСК И КВАНТОВО-ХИМИЧЕСКОЕ МОДЕЛИРОВАНИЕ

3.1 ЧИСЛЕННЫЕ МЕТОДЫ ОПТИМИЗАЦИИ

3.1.1 Численные методы оптимизации функций и их классификация

Задача нахождения оптимального значения (максимума или минимума) функции имеет множество практических приложений. Однако на практике в большинстве случаев целевая функция является слишком сложной или ее точный вид и вовсе невозможно установить. В такой ситуации решить задачу, аналитическими или геометрическими методами не представляется возможным, поэтому прибегают к численным методам оптимизации.

Рассмотрим задачу минимизации некоторой функции $f(x) \rightarrow \min$. Большинство численных методов являются итерационными. Их работа заключается в построении последовательности точек x^k , сходящихся к точке локального минимума x^* . В качестве начального приближения выбирается некоторая точка x^0 , а затем на каждой итерации $k \geq 0$ происходит смещение в направлении d_k на шаг t_k к следующей точке x^{k+1} . Таким образом, точка x^{k+1} вычисляется по следующей формуле:

$$x^{k+1} = x^k + t_k d_k.$$

При этом направление и длина шага на каждой итерации вычисляются исходя из численных характеристик оптимизируемой функции в одной или нескольких предыдущих точках – обычно в роли таких характеристик выступают значения функции, а также ее производных различных порядков. В зависимости от того какие именно характеристики функции используются для вычисления точек последовательности, численные методы делятся на три класса:

- Методы нулевого порядка используют только информацию о значениях оптимизируемой функции.
- Методы первого порядка в дополнение к информации о значениях самой функции используют также информацию о значениях вектора первых производных (градиента) этой функции.
- Методы второго порядка в дополнение к информации о значениях самой

функции и ее градиента, используют также информацию о значениях матрицы вторых производных функции (гессиан).

Кроме того, методы второго порядка, в зависимости от того, используют ли они явное вычисление гессиана или же аппроксимируют его на основе истории изменений градиента, делятся соответственно на ньютоновские и квазиньютоновские.

3.1.2 Метод наискорейшего спуска

Идея метода наискорейшего спуска заключается в том, чтобы на каждой итерации выбирать направление, которое обеспечивало бы самое быстрое убывание целевой функции при фиксированной длине шага. Способ выбора такого направления зависит от того, какая норма используется при решении задачи. В случае L2-нормированного пространства известно, что направление наискорейшего убывания функции совпадает с вектором ее антиградиента, поэтому в этих случаях метод наискорейшего спуска называют еще методом градиентного спуска. Каждая следующая точка в этом методе вычисляется по формуле:

$$x^{k+1} = x^k - t_k \nabla f(x^k),$$

где $\nabla f(x)$ – вектор градиента функции f в точке x .

Длина шага t_k может выбираться различными способами. Самый простой вариант – все время использовать шаг одинаковой длины, однако в таком случае существует вероятность того, метод не будет сходиться, поэтому чаще используется шаг переменной длины. Доказано, что если функция f выпукла, а ее производная непрерывна, то выбор длины шага по формуле (3.1) гарантирует сходимость метода к точке локального минимума.

$$t_k = \frac{|(x^k - x^{k-1})^T [\nabla f(x^k) - \nabla f(x^{k-1})]|}{\|\nabla f(x^k) - \nabla f(x^{k-1})\|^2} \quad (3.1)$$

Наглядную демонстрацию работы метода наискорейшего спуска можно видеть на рисунке 3.1. Следует обратить внимание, что векторы градиента функции в любой точке направлены перпендикулярно линиям уровня.

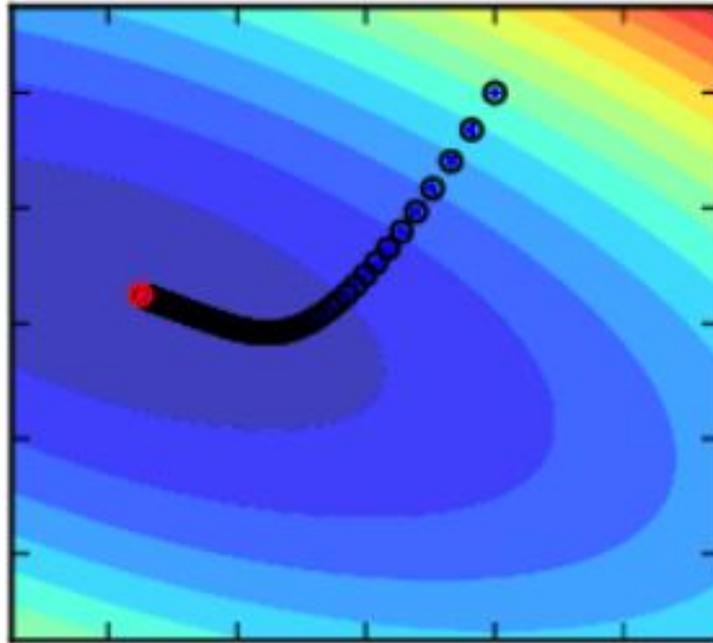


Рисунок 3.1 – иллюстрация работы метода наискорейшего спуска

3.1.3 Метод сопряженных градиентов

Одной из модификаций метода градиентного спуска является метод сопряженных градиентов. Доказано что если целевая функция является квадратичной, то метод сопряженных градиентов сходится за конечное число шагов, не превышающее количества переменных функции. Несмотря на то, что метод сопряженных градиентов лучше всего подходит для оптимизации квадратичных функций, существует также весьма эффективная его вариация, обобщенная на случай произвольной целевой функции – метод Полака-Райбера.

Стратегия построения точек в методе Полака-Райбера похожа на стратегию, использующуюся в методе градиентного спуска. Направление шага на каждой итерации вычисляется по следующей формуле:

$$d_0 = -\nabla f(x^0),$$

$$d_k = -\nabla f(x^k) + b_{k-1}d_{k-1}$$

Накопление погрешностей в ходе итераций может приводить к тому, что векторы d_k перестают указывать направление убывания целевой функции. Чтобы компенсировать этот эффект, каждые n шагов коэффициент b_{k-1} приравнивают к нулю. На практике значение параметра n часто принимают

равным размерности рассматриваемого пространства. Во всех же остальных случаях b_{k-1} вычисляется по формуле (3.2):

$$b_{k-1} = \frac{(\nabla f(x^k), \nabla f(x^k) - \nabla f(x^k))}{(\nabla f(x^k), \nabla f(x^k))} \quad (3.2)$$

В редких случаях метод Полака-Райбера может расходиться, однако его сходимость может быть гарантирована выбором $b_{k-1} = \max \{b_{k-1}; 0\}$.

3.1.4 Алгоритм L-BFGS

Методы наискорейшего спуска и сопряженных градиентов используют при работе только информацию о значениях градиента оптимизируемой функции. Алгоритм Бroyдена-Флетчера-Гольдфарба-Шанно (BFGS) является методом второго порядка – кроме градиента он использует еще и гессиан функции. Вместо того чтобы вычислять гессиан явным образом, как это делается в ньютоновских методах, BFGS аппроксимирует гессиан, используя для этого историю вычислений градиента по всем итерациям.

Общая схема метода схожа с двумя рассмотренными ранее. Обозначим гессиан функции в точке x^k как H_k . Тогда шаг d_k на каждой итерации вычисляется по формуле (3.3).

$$d_k = -H_k^{-1} \nabla f(x^k) \quad (3.3)$$

Длина шага выбирается таким образом, чтобы минимизировать значение функции по заданному направлению, то есть находится такое значение t_k , при котором значение $f(x^k + t_k d_k)$ будет минимально.

После этого определяются векторы $s_k = x^{k+1} - x^k$ и $y_k = \nabla f(x^{k+1}) - \nabla f(x^k)$. Данные векторы используются для того, чтобы по формуле (3.4) обновить приближение гессиана.

$$H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T \quad (3.4)$$

где I – единичная матрица, а $\rho_k = 1/y_k^T s_k$.

Из-за такой схемы вычислений алгоритм BFGS плохо подходит для

решения нелинейных задач с большим количеством неизвестных. Для обновления гессиана он требует хранить в оперативной памяти полную матрицу гессиана с предыдущей итерации, поэтому количество необходимой для вычислений памяти растет квадратично относительно размерности задачи. Специально для таких случаев была разработана модификация алгоритма под названием L-BFGS, оптимизирующая использование памяти. На каждой итерации метод L-BFGS заново рассчитывает приближение гессиана, используя несколько (обычно от 5 до 20) предыдущих значений градиента. При этом он построен таким образом, что не требует явного хранения матрицы гессиана ни на каком из этапов вычислений. Таким образом, в случае L-BFGS, потребляемая память зависит от размерности задачи линейно.

3.2 ОПТИМИЗАЦИЯ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ МОДЕЛИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ КВАНТОВОЙ ХИМИИ

3.2.1 Базовые принципы квантовой химии

Квантовая химия представляет собой один из раздел химии, в котором строение и свойства химических соединений, их взаимодействия и превращения в химических реакциях рассматриваются с помощью методов квантовой механики, а также на основе экспериментально установленных закономерностей.

Модели, описывающие химические системы, должны учитывать множество различных параметров – начиная от обычного взаимодействия частиц и заканчивая такими сложными явлениями, как квантовые эффекты неопределенности положения. Для того, чтобы облегчить решение соответствующих уравнений, используют различные уровни приближений, упрощающие, или и вовсе исключаящие из рассмотрения те или иные эффекты.

Одним из основных объектов, лежащих в основе теории квантовой механики, является волновая функция (3.5), которая описывает чистое состояние рассматриваемой системы:

$$\Psi(R_1, R_2, \dots, R_M, \vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t) \quad (3.5)$$

где R_i – координата ядра с индексом i , \vec{r}_i – координата электрона с индексом i , t – время.

Квантовая химия в качестве приближения использует предположение о том, что система всегда находится в стационарном состоянии, то есть ее энергия не зависит от времени [22]. Тогда, проведя некоторые преобразования с использованием аппарата квантовой механики, можно показать, что квадрат модуля волновой функции не зависит от времени (3.6):

$$|\Psi|^2 = \Phi(R_1, R_2, \dots, R_M, \vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) \quad (3.6)$$

Более того, величина $|\Psi|^2$ является плотностью вероятности обнаружения частицы в заданной точке пространства. Таким образом, вероятность обнаружить частицу в заданной точке не меняется со временем.

Далее большинство методов квантовой химии сводится к решению одной из вариаций уравнения Шредингера, которое позволяет найти волновую функцию системы и ее энергетический спектр. Энергетический спектр представляет собой множество всех возможных энергий квантовых состояний системы, и в случае, когда система находится в ограниченном объеме пространства, является дискретной величиной. Используя энергетический спектр и волновую функцию, можно вычислить энергию системы [23].

Для решения уравнения Шредингера на сегодняшний день существует два класса методов – неэмпирические и полуэмпирические.

Неэмпирические методы квантовой химии (методы *ab initio*) — методы приближенного решения уравнения Шредингера, не использующие для работы экспериментальные данные. Неэмпирические методы описывают состояние системы, исходя только из количества электронов в ней, а также количества и зарядов ядер.

Неэмпирические методы являются наиболее точными расчетными методами квантовой химии, однако их применение требует крайне трудоемких расчетов, из-за чего исследование систем, состоящих даже из нескольких десятков атомов, уже требует использования суперкомпьютеров.

Полуэмпирические методы являются попыткой решить описанную выше проблему за счет некоторых приближений, таких как приближение нулевого дифференциального перекрытия, предполагающее, что никакие две атомные орбитали не имеют общих точек. Кроме того, полуэмпирические методы также используют для вычислений экспериментальные данные, например, известные энергии ионизации атомов и дипольные моменты молекул. Все это позволяет параметризовать многие интегралы либо вообще не вычислять их, и тем

самым значительно ускорить вычисления [24].

3.2.2 Выделение конформаций белка на основе траектории молекулярной динамики

Потенциальная энергия молекулы представляет собой функцию степеней ее свободы (то есть длин связей, валентных и двугранных углов). Если построить график зависимости энергии молекулы от ее геометрических параметров, то получится некоторая поверхность в многомерном пространстве. Конформации, соответствующие локальным энергетическим минимумам на этой поверхности представляют особый интерес, потому что являются наиболее стабильными состояниями молекулы. Такие конформации еще называют конформерами или конформационными изомерами.

Для поиска конформеров существует большое количество методов, называемых методами конформационного поиска. Типичными примерами являются методы случайного поиска с использованием модельного отжига, метод геометрии межатомных расстояний, метод следования низкочастотным модам, методы сглаживания поверхности потенциальной энергии и др. Эти методы довольно сложны в применении, а для больших молекул, таких как молекулы белка, еще и требуют большого объема вычислений.

В данной работе мы решили отойти от классических методов конформационного поиска, и использовать полученную ранее траекторию молекулярной динамики для получения потенциальных конформеров. Как уже было сказано ранее, метод молекулярной динамики в отличие от обычной минимизации энергии обеспечивает более широкий охват конформационного пространства и позволяет преодолевать энергетические барьеры, разделяющие различные конформеры. Таким образом, полученная в результате молекулярно-динамического моделирования траектория может содержать одновременно несколько конформаций молекулы, лежащих в областях, соответствующих различным локальным энергетическим минимумам.

Каждый фрейм траектории соответствует какой-то случайной конформации исследуемого белка. В условиях неограниченных вычислительных ресурсов оптимальным решением было бы для каждой такой конформации провести минимизацию энергии, а из полученных структур выбрать ту, для которой итоговое значение энергии будет наименьшим из всех. Однако даже в нашем эксперименте траектория состоит из тысячи фреймов, что

делает решение данной задачи невозможным за разумное время. Поэтому необходимо каким-то образом выбрать из траектории небольшое количество конформаций, и проводить минимизацию только для них.

При выборе конформаций следует учитывать, что многие из них мало отличаются друг от друга. При минимизации энергии для таких конформаций получатся с большой вероятностью мы будем спускаться к одному и тому же локальному минимуму, что не несет никакой ценности в контексте задачи поиска конформеров. Поэтому для выбора конформаций было решено провести кластеризацию всех фреймов траектории.

Кластеризацией называют задачу разделения множества объектов на несколько подмножеств, называемых кластерами. При этом на множестве водится метрика, характеризующая расстояние между двумя объектами. Считается, что чем меньше расстояние между объектами, тем более они похожи друг на друга. Разбиение на кластеры производится таким образом, чтобы объекты внутри каждого кластера были более похожи друг на друга, чем на объекты из других кластеров. На рисунке 3.2 приведен пример кластеризации множества точек на плоскости, где мерой подобия является евклидово расстояние.

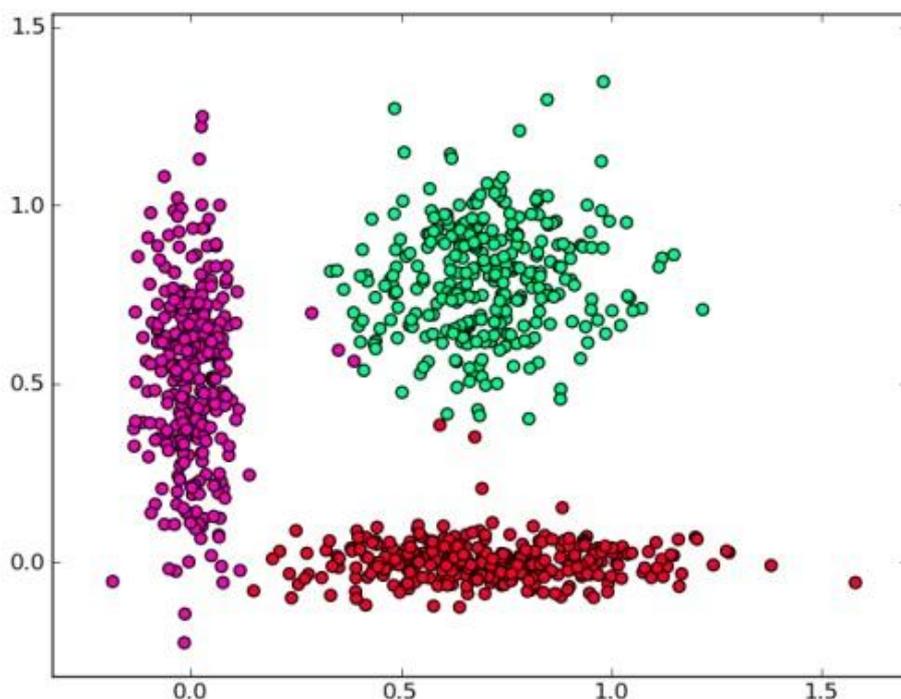


Рисунок 3.2 – График среднеквадратичного колебания аминокислотных остатков

В нашем случае кластеризация позволяет выделить кластеры конформаций, которые схожи друг с другом, но при этом в той или иной

степени отличаются от конформаций из других кластеров. В качестве метрики при этом было решено выбрать RMSD, так как эта величина хорошо характеризует степень структурного подобия конформаций.

В качестве алгоритма кластеризации был использован метод *k*-средних (*k*-means). Суть этого метода заключается в следующем: сначала задается количество кластеров *k*, на которые будет производиться разбиение (в нашем случае $k = 30$). Затем из всего множества случайным образом выбирается *k* объектов – они называются центроидами и исполняют роль представителей кластеров. Дальнейшие действия представляют собой итеративный алгоритм. Для каждого из остальных объектов находится самая близкая к нему центроида, и данный объект попадает в соответствующий ей кластер. После этого для каждого кластера вычисляется новая центроида – среднее по всем объектам внутри кластера. На этом итерация заканчивается, и описанные действия повторяются заново, начиная с момента разбиения объектов по кластерам.

В таблице 1 представлены результаты кластеризации. Для дальнейшего анализа были отобраны конформации, являющиеся центроидами кластеров на последней итерации алгоритма.

Таблица 1 – результаты кластеризации траектории

Номер кластера	Количество конформаций в кластере	Номер фрейма-представителя	Время на траектории, соответствующее фрейму-представителю, нс	Среднее RMSD внутри кластера, Å	Среднее RMSD относительно других кластеров, Å
29	6	3	1	1.731	5.272
25	16	13	3	2.098	5.000
26	15	26	5	1.864	4.660
21	23	49	10	2.049	4.632
19	27	73	15	1.895	4.608
6	45	107	21	1.781	4.189
2	54	165	33	1.952	3.842
9	36	193	39	1.844	4.102
13	33	244	49	2.002	4.196
20	25	263	53	1.846	4.013
22	23	289	58	1.753	3.730
14	32	324	65	1.683	3.828
17	30	346	69	1.889	3.590
18	28	378	76	1.751	3.885
4	52	412	82	1.910	3.790
0	72	495	99	2.074	3.597
10	35	515	103	1.865	3.919
15	31	565	113	1.779	3.442
1	70	593	119	2.032	3.474

Номер кластера	Количество конформаций в кластере	Номер фрейма-представителя	Время на траектории, соответствующее фрейму-представителю, нс	Среднее RMSD внутри кластера, Å	Среднее RMSD относительно других кластеров, Å
5	48	681	136	2.053	3.728
7	45	733	147	1.834	3.402
23	19	758	152	1.884	3.677
28	13	794	159	1.685	3.611
8	36	822	164	1.749	3.719
3	54	849	170	1.855	3.346
11	35	877	175	1.943	4.093
12	34	920	184	1.962	3.567
16	31	950	190	1.842	3.425
27	14	973	195	1.643	4.290
24	18	987	197	2.018	4.221

3.2.3 Оптимизация конформаций белка с использованием методов квантовой химии

В ходе предыдущего этапа работы нами было построено тридцать кластеров, центроиды которых соответствуют различным вариантам пространственной структуры теломеразы. Теперь данные конформации должны быть подвергнуты процедуре минимизации энергии с целью получения соответствующих оптимальных устойчивых конформаций.

Для минимизации энергии было принято решение использовать методы квантовой химии, чтобы достигнуть максимально возможной точности. Доступ к программному обеспечению, позволяющему проводить подобные расчеты, в полной мере обеспечивает пакет MORAC (Molecular Orbital PACkage). Данный программный пакет специально был разработан для решения задач квантовой химии и содержит в себе реализации таких полуэмпирических методов как MNDO, AM1, PM3, RM1, PM6, PM7 и др.

Чтобы избежать путаницы, следует отдельно упомянуть, что неэмпирические методы оперируют метрикой потенциальной энергии, а полуэмпирические – энтальпией образования молекулы. Так как в данной работе мы рассматриваем конформации одной и той же молекулы, набор атомов остается неизменным, и потенциальная энергия отличается от энтальпии образования на константную величину.

Так как квантово-химические расчеты весьма трудоемки, для вычислений использовались ресурсы национальной грид-сети РБ, а именно – UNICORE-

сайта SKIF_GEO. Данный кластер оснащен большим количеством узлов с объемом оперативной памяти 192 Гб, что позволяет сравнительно эффективно выполнять даже очень объемные расчеты.

Для того чтобы полученные результаты были корректными и точными, MORAS требует проведения предварительной подготовки структуры. В случае с оптимизацией белковой молекулы необходимо сначала добавить к ней атомы водорода. Несмотря на то, что исходные структуры уже содержали атомы водорода, было решено эти атомы удалить, а затем добавить заново. Благодаря этому правилу, по которым добавлялись атомы, в точности соответствовали логике, которую MORAS использует в ходе вычислений. Кроме того, ранее при преобразовании траектории из нее были удалены нейтрализующие ионы, а обновление водородов позволяет решить эту проблему за счет протонирования и депротонирования боковых цепей белка. Нейтрализация структуры необходима, потому что в противном случае при наличии рядом нескольких аминокислотных остатков с одинаковым зарядом, может возникнуть высокая концентрация зарядов, что приведет к конформационным изменениям и нереалистичной итоговой структуре.

Добавление атомов водорода к структуре в MORAS производится с помощью ключевого слова ADD-H, при этом большинство атомов молекулы автоматически нейтрализуется. Исключениями являются атомы металлов, которые обычно существуют как ионы (Na^+ , Ca^+), не участвующие в образовании ковалентных связей атомы галогенов (Cl , F), а также некоторые органические участки, для которых ионизированное состояние является естественным. Такой подход к нейтрализации приводит к исчезновению всех солевых мостиков, которые образуются между положительно и отрицательно заряженными боковыми радикалами аминокислотных остатков белка. Чтобы наличие солевых мостиков учитывалось при добавлении атомов водорода, используется ключевое слово SITE=(SALT).

После того, как атомы водорода добавлены к структуре, выполняется их первичная оптимизация. Этот промежуточный этап важен, потому что по сравнению с основной оптимизацией выполняется он довольно быстро, но при этом позволяет значительно приблизить состояние структуры к локальному энергетическому минимуму. Для запуска соответствующего алгоритма с помощью ключевого слова NOOPT отключается оптимизация для всех типов атомов, а затем с помощью слова OPT-H включается оптимизация только для атомов водорода. Также важными ключевыми словами на этом этапе являются MOZYME, EPS=78.4 и GNORM=20, однако их значение будет рассмотрено

немного позднее.

На этом подготовка системы заканчивается, и можно переходить непосредственно к оптимизации всей структуры белка. Для оптимизации использовался полуэмпирический квантовый метод PM7 как самый современный и точный из доступных. Для запуска алгоритма используются ключевые слова CUTOFF=6, EPS=78.4, MOZYME и GNORM=20. Параметр CUTOFF задает значение в ангстремах, соответствующее максимальному расстоянию, на котором взаимодействие между атомами еще будет учитываться. Таким образом обрезаются слабые связи между атомами, расположенными далеко друг от друга, что позволяет значительно ускорить вычисления.

Ключевое слово EPS обозначает наличие растворителя. По умолчанию MOPAC производит моделирование в газообразной среде, что упрощает вычисления, но не позволяет учитывать различные эффекты, требующие наличия растворителя, например, эффект Гофмейстера. Значение EPS=78.4 моделирует неявный водный растворитель с диэлектрической проницаемостью равной 78,4 на основе модели COSMO (Conductor-like Screening Model). Модель COSMO представляет растворитель как континуум с заданной величиной диэлектрической проницаемости ϵ . При этом вокруг молекулы создается “оболочка”, за которую растворитель не проникает. Данная оболочка обычно представляет собой результат пересечения сфер с центрами в атомах и радиусами, приблизительно на 20% превышающими Ван-дер-Ваальсовы радиусы соответствующих атомов. Распределение электрических зарядов в молекуле известно, а значит можно рассчитать и заряды \check{q} для сегментов оболочки. Для расчета соответствующих зарядов растворителя применяется специальная понижающая функция, вычисляемая по формуле (3.7):

$$f(\epsilon) = \frac{\epsilon - 1}{\epsilon + \alpha} \quad (3.7)$$

Здесь ϵ – диэлектрическая проницаемость растворителя, α – константа, обычно принимающая значения от 0 до 0,5. Исходя из этого, заряды растворителя вычисляются как:

$$q = f(\epsilon) * \check{q} \quad (3.8)$$

На основании зарядов растворителя и известного распределения зарядов молекулы может быть вычислена энергия взаимодействия молекулы с растворителем для использования в дальнейших расчетах.

В квантовой химии классические полуэмпирические методы основаны на матричной алгебре. Одним из недостатков подобного подхода, ограничивающим его практическое применение, является то, что сложность вычислений возрастает как кубический полином относительно количества атомов в исследуемой системе, что делает моделирование больших систем затруднительным. Ключевое слово MOZYME позволяет вместо стандартной методики использовать метод локализованных молекулярных орбиталей (LMO) для вычисления энергии системы. Данный метод был специально разработан для того, чтобы облегчить вычисления в случае больших органических соединений, таких как белки. Метод локализованных молекулярных орбиталей позволяет решать уравнения самосогласованных полей (SCF equations) за время, прямо пропорциональное размеру системы.

Для оптимизации энергии системы в AMBER используется алгоритм L-BFGS. Так как он является итерационным, и сходимость к точке локального минимума, где градиент равен нулю, может занять очень много времени, используется критерий останова, задаваемый параметром GNORM. В этой работе использовалось значение $GNORM = 20$. Это означает, что как только на одной из итераций значение градиента функции станет меньше 20, будет принято решение о том, что текущая точка с большой вероятностью находится в окрестности локального минимума, и алгоритм завершит свою работу.

3.2.4 Анализ результатов

Результатом работы алгоритма является оптимизированная структура в PDB формате. Для оценки итогового значения энтальпии образования и потенциальной энергии структуры используется дополнительный расчет с помощью MORAS. Это необходимо, потому что оценка, производимая методом LMO в ходе оптимизации, имеет склонность к накоплению погрешности, что в перспективе большого числа итераций сказывается на точности финального результата.

В таблице 2 представлены результаты оптимизации для восьми конформаций, показавших лучшие итоговые значения энтальпии образования. Так как начальными данными для оптимизации являлись центроиды кластеров, для идентификации начального состояния используется соответствующий центроиду номер фрейма траектории, а также время от начала моделирования молекулярной динамики, которое соответствует данному фрейму.

Таблица 2 – результаты кластеризации траектории

Номер кластера	Номер фрейма-представителя	Время на траектории, соответствующее фрейму-представителю, нс	Энтальпия образования, ккал/моль
17	346	69	-48133
18	378	76	-47930
11	877	175	-47807
16	950	190	-47774
7	733	147	-47754
28	794	159	-47747
3	849	170	-47683
22	289	58	-47659

Также для более наглядной демонстрации результатов был построен график, где каждому фрейму траектории поставлено в соответствие значение энтальпии образования ΔH для центроиды соответствующего ему кластера (рисунок 3.3). Для того чтобы можно было лучше соотнести эти результаты с анализом траектории, который проводился после молекулярно-динамического моделирования, на график энтальпии по оси времени был наложен график RMSD (рисунок 3.4).

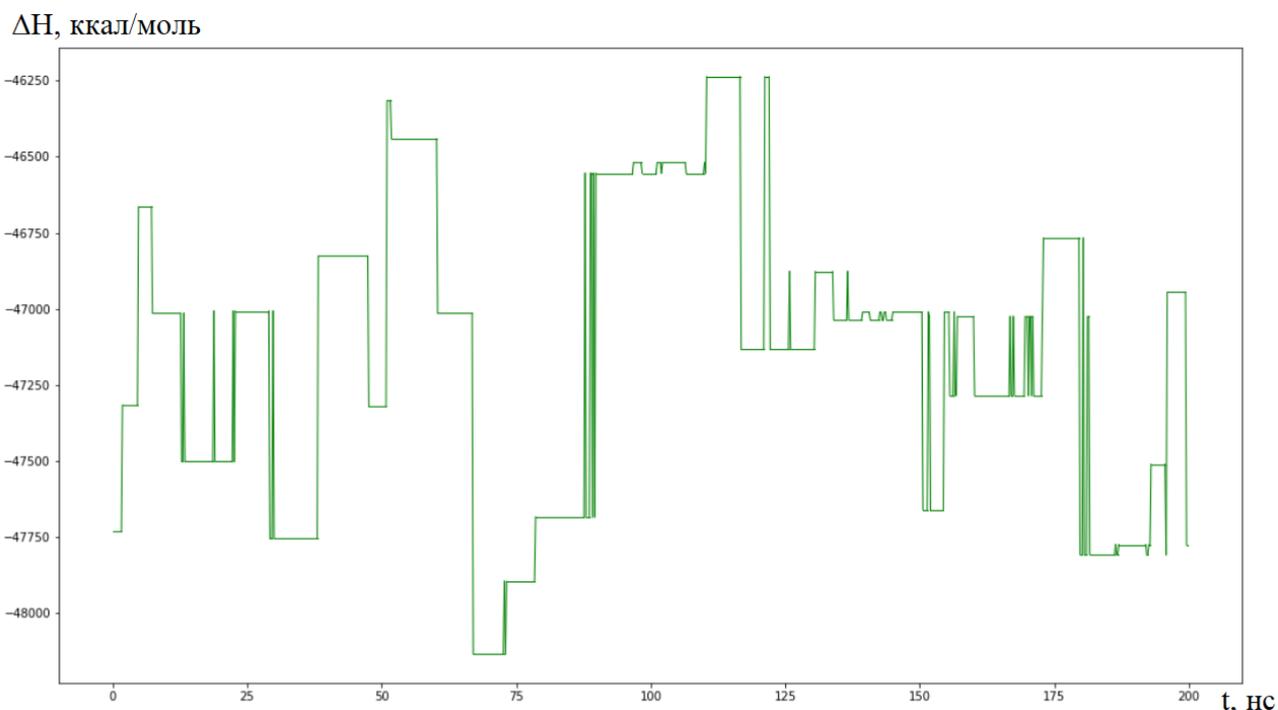


Рисунок 3.3 – График зависимости энтальпии образования от времени моделирования

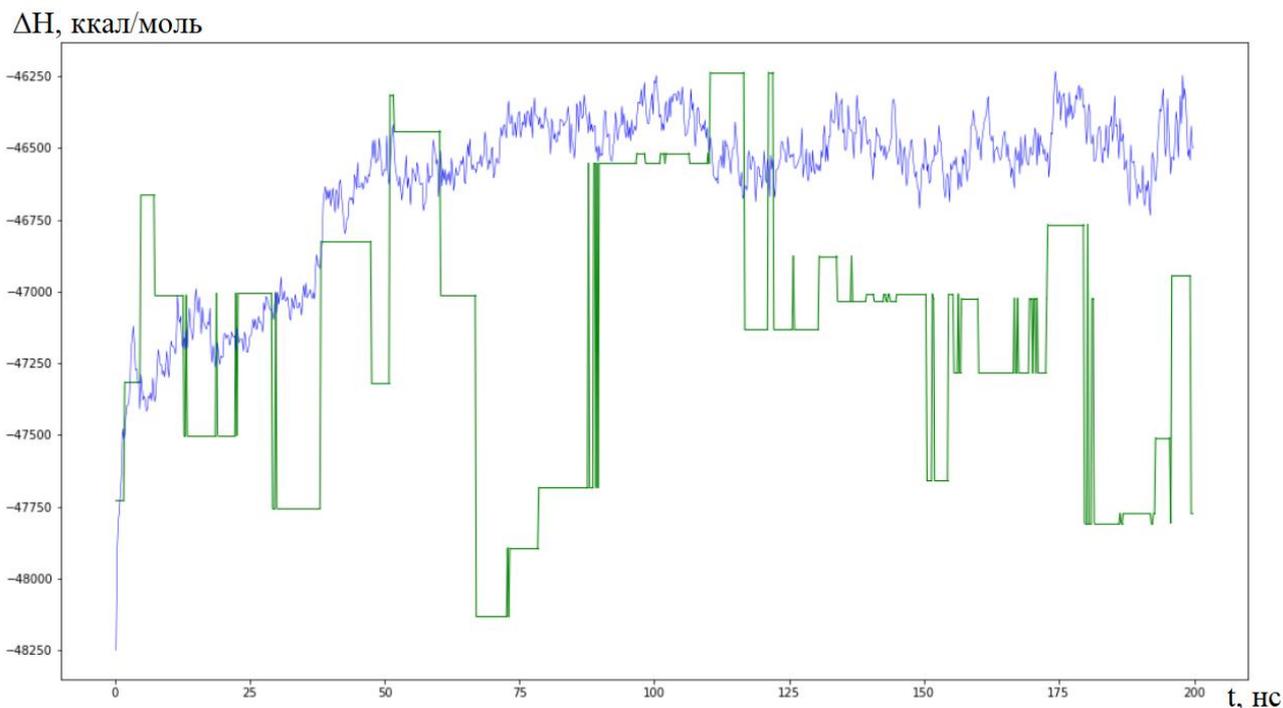


Рисунок 3.4 – График зависимости энтальпии образования от времени моделирования с наложенным по оси времени графиком среднеквадратичного отклонения от исходной структуры

Для сравнения результатов аналогичная процедура оптимизации энергии также была проведена для теломеразы жука, которая использовалась нами в качестве шаблона для моделирования. Энтальпия образования этой структуры была оценена МОРАС в -51830 ккал/моль, что не так сильно отличается от значений, полученных для модели теломеразы человека, и говорит о хорошем качестве структуры модели.

Итоговым результатом этой работы можно считать лучшую конформацию из полученных в результате оптимизации. Полученная модель пространственной структуры теломеразы может быть использована в других исследованиях как исходная точка для дальнейшего анализа, а также на практике с целью поиска ингибиторов теломеразы, как потенциальных лекарственных препаратов с широким спектром противоопухолевого действия.

ЗАКЛЮЧЕНИЕ

В ходе выполнения дипломной работы были изучены теоретические основы гомологичного моделирования, молекулярной динамики и квантово-химического моделирования, а также отработаны практические навыки применения этих методов для решения реальных задач. Помимо этого, была рассмотрена литература и научные статьи, связанные с изучением фермента теломеразы.

С использованием полученных в ходе подготовки знаний и навыков было проведено гомологичное моделирование пространственной структуры фрагмента теломеразы человека. Затем была построена и проанализирована молекулярно-динамическая траектория, моделирующая поведение данной структуры во времени. На основе построенной траектории была проведена дополнительная оптимизация модели с помощью методов квантовой химии.

Полученная в результате работы модель пространственной структуры теломеразы может быть использована в других исследованиях как исходная точка для дальнейшего анализа, а также на практике с целью поиска ингибиторов теломеразы, как потенциальных лекарственных препаратов с широким спектром противоопухолевого действия.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Ozturk M.B. Current insights to regulation and role of telomerase in human diseases // *Antioxidants* (Basel). – 2017. – Vol. 6, №1. – P. 1–13.
2. Blackburn E.H. Structure and function of telomeres // *Nature*. – 1991. – Vol. 350. – P. 569–573.
3. Harley C.B., Futcher A.B., Greider C.W. Telomeres shorter during aging of human fibroblast // *Nature*. – 1990. – Vol. 345. – P. 458–460.
4. Baginski M., Serbakowska K. In silico design of telomerase inhibitors // *Drug Discovery Today*. – 2020. – Vol. 25, №7. – P. 1213–1222.
5. Kim, N.W. Specific association of human telomerase activity with immortal cells and cancer // *Science*. – 1994. – Vol. 266. – P. 2011–2015.
6. Shay, J. Role of Telomeres and Telomerase in Aging and Cancer // *Cancer Discovery*. – 2016. – Vol. 6, №6. – P. 584–593.
7. Schmitt C.A. Cellular senescence and cancer treatment // *Biochim Biophys Acta*. – 2007. – Vol. 1775, №1. – P. 5–20.
8. Tchkonina T., Zhu Y., van Deursen J., Campisi J., Kirkland J.L. Cellular senescence and the senescent secretory phenotype: therapeutic opportunities // *J Clin Invest*. – 2013. – Vol. 123, №3. – P. 966–972.
9. Rouda S., Skordalakes E., Structure of the RNA-Binding Domain of Telomerase: implications for RNA recognition and binding // *Structure*. – 2007. – Vol. 15, №11. – P. 1403–1412.
10. Gillis A.J., et al. (2008) Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT // *Nature*. – 2008. – Vol. 455. – P. 633–637.
11. Mitchell M., Gillis A., Futahashi M., Fujiwara H., Skordalakes E. Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA // *Nat Struct Mol Biol*. – 2010. – Vol. 17, №4. – P. 513–518.
12. Bryan C. Structural basis of telomerase inhibition by the highly specific BIBR1532 // *Structure*. – 2015. – Vol. 23, №10. – P. 1934–1942.
13. Nguyen T.H.D. Cryo-EM structure of substrate-bound human telomerase holoenzyme // *Nature*. – 2018. – Vol. 557. – P. 190–195.
14. Молекулярное моделирование: теория и практика / Х.-Д. Хельтье [и др.]; пер. с англ. – 2-е изд. – Москва: БИНОМ, 2013. – 319 с.

15. Зверева М.Э., Щербакова Д.М., Донцова О.А. Теломераза: структура, функции и пути регуляции активности // Успехи биологической химии. – 2010. – Vol. 50. – P. 155–202.
16. Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G. SWISS-MODEL: homology modelling of protein structures and complexes // Nucleic Acids Res. – 2018. – Vol. 46. – P. 296–303.
17. Laskowski R.A., MacArthur M.W., Moss D.S., Thornton J.M. PROCHECK: a program to check the stereochemical quality of protein structures // J. Appl. Cryst. – 1993. – Vol. 46. – P. 283-291.
18. Singh, D.B. Computer-Aided Drug Design. / D.B Singh – Springer Singapore, 2020. – 306с.
19. Hollingsworth S.A., Dror R.O. Molecular dynamics simulation for all // Neuron. – 2018. – Vol. 99, №6. – P. 1129–1143.
20. Tian C., Kasavajhala K., Belfon K.A.A. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution // J. Chem. Theory Comput. – 2019. – Vol. 16. – P. 528–552.
21. The Art of Molecular Dynamics Simulation / Rapaport D.C. // Cambridge University Press, 2004. – 286с.
22. Modern Quantum Chemistry / Sinanoglu O. // Sterling Chemistry Laboratory Yale University New Haven, Connecticut, 1965. – 47с.
23. Квантовая механика и квантовая химия / Степанов Н.Ф. // Изд. Московского ун-та, 2001. - 42с.
24. Полуэмпирические расчетные методы квантовой химии / Блатов В.А., Шевченко А.П., Пересыпкина Е.В. // Самара: Универс-групп, 2005. - 56с.