Министерство науки и высшего образования Российской Федерации
Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт
Высшая школа теоретической механики и математической физики

Работа допущена к защите
Директор ВШТМиМФ,
д.ф.-м.н., чл.-корр. РАН
А. М. Кривцов
«___»_____2022 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

ФОЛДИНГ БЕЛКА ПРИ ПОМОЩИ ГЛУБОКОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

по направлению подготовки
01.03.03 «Механика и математическое моделирование»
направленность
01.03.03_03 Математическое моделирование процессов нефтегазодобычи

Выполнил:
Студент гр. 5030103/80301 А. Д. Артемьев

Руководитель:
Профессор ВШИИ, д.т.н, профессор В. С. Заборовский

Санкт-Петербург
2022

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО

Физико-механический институт

Высшая школа теоретической механики и математической физики

УТВЕРЖДАЮ

Директор ВШТМиМФ,

А. М. Кривцов

«___»_____2022 г.

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

студенту Артемьеву Алексею Дмитриевичу, гр. 5030103/80301

1. Тема работы: Фолдинг белка при помощи глубокого обучения с подкреплением.

2. Срок сдачи студентом законченной работы: 06.06.2022.

3. Исходные данные по работе: Актуальные научные публикации по теме работы, исследования в области фолдинга белка и применения методов глубокого обучения с подкреплением.

4. Содержание работы (перечень подлежащих разработке вопросов): Постановка задачи квалификационной работы, анализ методов работы с фолдингом белка, реализация методов машинного обучения для решения задачи.

5. Перечень графического материала (с указанием обязательных чертежей): не предусмотрено.

6. Консультанты по работе: отсутствуют.

7. Дата выдачи задания 13.05.2022.

Руководитель ВКР _____ В. С. Заборовский

Задание принял к исполнению 13.05.2022

Студент _____ А. Д. Артемьев

РЕФЕРАТ

На 60 с., 16 рисунков, 1 таблица

ФОЛДИНГ БЕЛКА, ГЛУБОКОЕ ОБУЧЕНИЕ, ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ, МАШИННОЕ ОБУЧЕНИЕ, ЗАДАЧА ОПТИМИЗАЦИИ

В работе предложен метод решения задачи фолдинга белка при помощи глубокого обучения с подкреплением. Создана математическая модель физических процессов при сворачивании гидрофобно-гидрофильного (HP-модель) белка. Получены результаты решения оптимизационной задачи, которые показывают как количественно близкие к оптимальному решению способы сворачивания белка, так и качественно обоснованные биохимической и биофизической теорией.

ABSTRACT

60 pages, 16 figures, 1 table

PROTEIN FOLDING, DEEP LEARNING, REINFORCEMENT LEARNING, MACHINE LEARNING, OPTIMIZATION TASK

The paper proposes a method for solving the protein folding problem using deep reinforcement learning. A mathematical model of physical processes during the folding of hydrophobic-hydrophilic (HP-model) protein has been created. The results of solving the optimization problem are obtained, which show both quantitatively close to the optimal solution methods of protein folding, and qualitatively justified by biochemical and biophysical theory.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ 5
ГЛАВА 1. АКТУАЛЬНОСТЬ ПРОБЛЕМЫ И ПОСТАНОВКА ЗАДАЧ ИССЛЕДОВАНИЯ 8
1.1. Актуальность задачи фолдинга белка 8
1.2. Существующие подходы и актуальность их исследования 11
1.3. Предлагаемые подходы и постановка задачи 14
ГЛАВА 2. МОДЕЛИ И МЕТОДЫ ФОЛДИНГА БЕЛКА 16
2.1. Биохимическое устройство белков 16
2.2. Математическая модель биофизических процессов 21
2.3. Задача фолдинга для HP-модели 28
ГЛАВА 3. АЛГОРИТМЫ И ПРОГРАММЫ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ФОЛДИНГА БЕЛКА 31
3.1. Глубокое обучение с подкреплением 31
3.2. Архитектура среды и модель агента 33
3.3. Дизайн награды 38
ГЛАВА 4. АНАЛИЗ РЕЗУЛЬТАТОВ РЕШЕНИЯ ЗАДАЧ ИССЛЕДОВАНИЯ И ОЦЕНКА ИХ ПРАКТИЧЕСКОЙ ЗНАЧИМОСТИ 40
4.1. Постановка экспериментов 40
4.2. Результаты исследований и графики обучения 42
4.3. Пример фолдинга и верификация результатов 46
ЗАКЛЮЧЕНИЕ 51
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ 54

ВВЕДЕНИЕ

Вопрос о способе сворачивания белков в живых организмах до сих пор является актуальным для исследователей. Проблема фолдинга белков включает в себя более чем 50 лет трудов различных ученых по этой теме, однако полного понимания процессов перехода белка из развернутой первичной структуры в свернутое состояние до сих пор нет.

Сложность понимания этих процессов возникает из-за ограниченных знаний человека о некоторых из биологических процессов. Основной вклад в объяснение модели белка внесли ученые, работавшие в 19–20 веках, поэтому эта модель построена на основных принципах физики и механики в частности. Взаимодействие частиц, например, объясняется силами притяжения и электростатическими силами, также рассматривается энергия связей в цепочке аминокислот. Комбинация этих знаний формирует пограничные науки биохимию и биофизику, цель которых является описание процессов в организмах.

Еще в начале 50-х годов прошлого века Кристиан Бемер Анфинсен – американский ученый в области биохимии – сформулировал закон о том, что последовательность аминокислот в белке определяет его трехмерную структуру. Несмотря на то, что это высказывание обрело известность среди ученых, процесс фолдинга долгое время воспринимался как случайное событие. Только в последние десятилетия исследователи начали понимать, что сворачивание белка – это постоянный процесс в организме человека, требующий сложного взаимодействия большого числа биокомпонентов, нарушение которого приводит к таким заболеваниям, как рак, диабет, болезни Альцгеймера и Паркинсона.

За свои труды в области фолдинга белка Анфинсен получил Нобелевскую премию по химии в 1972 году. Предметом его работы стала небольшая последова-

тельность из 100 аминокислот, называемой рибонуклеаза А, в дальнейшем часто используемая другими учеными для экспериментов.

Примерно в это же время другой американский ученый и молекулярный биолог Сайрус Левинталь рассматривал проблему фолдинга белка с вычислительной стороны. В 1968 году он сформулировал парадокс, заключающийся во временной оценке сворачивания белка в различные конформации. По его предположению, если белку требуется 10^{-13} секунд на проверку одного свернутого состояния, то для цепочки из всего лишь 100 аминокислотных остатков ему требуется перебрать 10^{100} комбинаций, что составило бы около 10^{80} лет. Как известно, это больше на несколько десятков порядков, чем время жизни всей вселенной. Но белок сворачивается за секунды, а значит он не просто перебирает все возможные варианты, а делает это по какому-то определенному, “умному” принципу, неизвестному человеку.

Проблема в решении задачи фолдинга белка также заключается в ограниченности ресурсов для проведения расчетов по получению пространственной структуры белка и временной сложности самой задачи. Даже для упрощенных моделей белка время переборов всех вариантов может быть непозволительно большим, что требует более интеллектуального подхода к формализации задачи и ее решению.

Важность научного обоснования устройства белков и способов их сворачивания сложно переоценить. Решение задачи предсказания структуры белка с использованием методов машинного обучения должно дать ответ на вопрос о том, как белок сворачивается. Если начальная последовательность определяет трехмерную конформацию, а интеллектуальный самообучающийся нейросетевой агент может решить задачу прогнозирования этой структуры, то задача человека состоит в том, чтобы интерпретировать эти результаты. В данном случае объяснимость работы машинного обучения можно сформулировать при помощи прицип топо-

логии для определенной конформации белка, если прибегнуть к области машинного обучения с подкреплением.

ГЛАВА 1. АКТУАЛЬНОСТЬ ПРОБЛЕМЫ И ПОСТАНОВКА ЗАДАЧ ИССЛЕДОВАНИЯ

1.1. Актуальность задачи фолдинга белка

Проблема фолдинга белка является одной из самых интересных нерешенных задач. Неизученность биохимических процессов в организме человека, связанных с синтезом белков, не позволяет понять принципы сворачивания белка для предсказания его нативной структуры. С другой стороны, введение даже упрощенных математических моделей, таких как двумерная HP-модель белка, оставляет задачу NP-полной, что не позволяет разработать эффективные алгоритмы решения. Знание о способе перехода белка от первичной структуры в виде определенной аминокислотной последовательности к более сложной конформации может помочь человеку в борьбе с неизлечимыми болезнями (Альцгеймера [7, 15], Паркинсона [1], Хантингтона), с болезнями сахарного диабета [4], а также в борьбе с раком , причиной которых является мисфолдинг (то есть неправильное сворачивание) белка. Также понимание процесса фолдинг белка откроет возможности в разработке дизайна новых лекарств.

Несмотря на обладание знанием о том, что сохранение гидрофобных связей движет сворачиванием белка, точно предсказывать трехмерную структуру белка человек еще не умеет. Одним из проектов, занимающихся этой задачей, является Rosetta, – они смогли предсказать около 12% трехмерных конформаций для семейства белков, для которых ранее не было результатов. Это стало возможно благодаря методам машинного обучения, имплементации нового алгоритма и распределенному кластеру компьютеров. Их предсказанные значения были встроены в публичную базу данных структур белка, а затем некоторые из них экспериментально подтверждены.

Другим интересным исследовательским проектом является онлайн-головоломка Foldit, в которой пользователю предлагается сворачивать белок, получая награду за более правильную конформацию. Лучшие из решений были проанализированы учеными, и на их основе предполагается найти решения многих научных проблем, связанных, например, с разработкой вакцин. Принцип игры в том, что человеческий мозг способен очень хорошо работать с трехмерными структурами, такими как пазл или конструктор, не тратя времени на перебор всех комбинаций. По сути, белок делает аналогично, – он знает, как правильно сложиться в структуру с минимальной энергией, а не перебирает все возможные варианты. Человек, в свою очередь, не обладает этим знанием и пытается использовать различные методы для построения модели фолдинга.

Основным прорывом за последнее время стал алгоритм компании Google DeepMind под названием AlphaFold, предсказывающий пространственную структуру белка. Методами этого алгоритма стали глубокое обучение в связке с данными из открытых источников по схожим по структуре белкам. Результаты алгоритма представлены как решение задачи соревнования CASP, которое подразумевает предсказание белковых структур. Оно проходит в формате разделения ученых на тех, которые занимаются алгоритмами предсказания структуры для белка, экспериментальные данные по которому еще не проведены, и на тех, кто проводит эксперименты в лаборатории. После решения ученые верифицируют результаты. До 2016 года метрика GDT (Global distance test) составляла не более 40 очков (метрика считается от 0 до 100 очков, где меньше 30 – случайная структура, а 100 – точно предсказанная). В 2021 году алгоритм AlphaFold2 решил задачу с метрикой GDT, равной 92 очкам, что позволяет точно утверждать об актуальности применения методов глубокого обучения в данной задаче.

Также стоит отметить успехи проекта Folding@Home, суть которого заключается в создании одного из самых больших и высокопроизводительных распре-

деленных кластеров на основе персональных компьютеров добровольцев и некоторых из кластеров крупных компаний. В 2020 году его мощность достигла 2.7 экзафлопса, что было больше, чем суммарная производительность всех суперкомпьютеров в мировом рейтинге TOP500. Результаты эксперимента позволили написать более 200 научных работ, вносящих вклад в решение задачи фолдинга белков.

Такие методы, например, позволят смоделировать белок шипов SARS-CoV-2, и при помощи сопоставления с физическими данными можно будет получить точную структуру белка, что даст человеку понять, как вирус взаимодействует с рецепторами дыхательных путей, и найти лекарства от связывания и последующего инфицирования COVID-19.

С вычислительной точки зрения проблема фолдинга белков включает в себя два принципиальных момента, которые необходимо учесть:

1. Первый – выбор модели. Модель белка определяет задачу, которая будет решаться. Интерпретируемость результатов также зависит от выбора типа модели, поэтому критически важно рассматривать проблему как комплексный объект, изученный со всех важных сторон. Для этого требуется изучение и применение обширной теории в молекулярном моделировании, формирующую модель белка на атомарном уровне.

2. Второй важный момент – построение методов решения поставленной оптимизационной задачи на выбранной модели белка. Здесь есть большой простор для исследователя, но для качественно и количественно верного результата следует выбрать подходящие методы. В опубликованных ранее другими исследователями работах можно заметить некоторое медленное продвижение в решении задачи предсказания свернутой структуры белка. Однако, эти методы все равно неспособны полностью решить NP-сложные задачи и низкоэффективны в нахождении глобального оптимума, так как

опираются на анализ локальных данных. Здесь могут выручить методы машинного обучения, которые выделяют общие свойства структуры белка из обучающей выборки в неявной форме. При помощи нейронных сетей и с привлечением глубокого обучения возможно учесть эти закономерности и смоделировать высокоуровневые абстракции. Такой абстракцией может быть граница некоторых областей. Для формализации этих представлений можно обратиться к топологическому анализу данных, который работает с топологическими комплексами с помощью алгебраической топологии.

Сформулировать процесс фолдинга можно с точки зрения топологии. Белки в развернутом виде являются длинной цепочкой аминокислот, закодированной при помощи ДНК, но впоследствии запутываются более сложные структуры. Основные абстракции такого сворачивания является спираль, имеющая вид правозакрученной винтовой линии, и лист (или плоскость). Эти компоненты также складываются в более сложные структуры (третичная структура). В основе этого сворачивания лежит не случайность процесса, а целенаправленное принятие белком вида глобулы, несмотря на многовариантность выбора нативной конформации. Структуру свернутого белка можно описать предсказать, опираясь на объективные физико-химические законы, в частности на распределение поверхностного электрического заряда в глобуле. Можно сказать, что белок – это самособирающаяся машина, которая имеет уникальную форму (то есть топологию), определяющая его функцию в организме. Описать эти процессы возможно при помощи создания математической модели, созданной в данной работе.

1.2. Существующие подходы и актуальность их исследования

Проблема фолдинга белков является популярной оптимизационной задачей, изученной в других работах. Она сформулирована в различных постановках: для разных моделей белка, для разных целей (предсказание какой структуры белка

нужно сделать из какого состояния), при помощи различных методов решения задачи (оптимизационный, глубокое обучение, обучение с подкреплением).

В существующих работах по теме фолдинга белка предложены различные способы решения задачи с использованием HP-модели белка. Все алгоритмы можно условно разделить на три типа:

1. Аппроксимационные алгоритмы оптимизации. Такие методы обеспечивают приближенное решение оптимизационной задачи в рамках заданного времени, причем это строго доказуемо. Это помогает в решении NP-трудных задач, которые скорее всего нельзя решить за полиномиальное время, и к которым относится фолдинг. В работе [27] такой метод позволяет решить задачу для трехмерной сетки с коэффициентом аппроксимации 3/8 от оптимального решения, или с гарантированным коэффициентом 2/5 (работа тех же авторов [28]). В публикации [44] автор достиг результата решения с коэффициентом 2/3 для двумерного случая модели HP.

2. Алгоритмы комбинаторной оптимизации. Комбинаторная оптимизация рассматривает задачу поиска лучшего элемента в дискретном множестве или в тех множествах, которые могут быть сведены к дискретным. В [31] рассмотрены различные методы и теоремы комбинаторной оптимизации, примененных для HP моделей белка. Авторы работы замечают, что упомянутые алгоритмы схожи друг с другом, а максимальный коэффициент аппроксимации достиг 88.1% для некоторых из последовательностей. Основные работы содержат решения задачи фолдинга для трехмерной кубической и гранецентрированной кубической решеток [16, 41] методами программирования в ограничениях. В схожих постановках задача решалась методами линейного программирования [22, 23].

3. Эвристические алгоритмы оптимизации. Такие алгоритмы хоть и не гарантируют нахождение лучшего с точки зрения заданной метрики решения, но в большинстве случаев дают близкое к оптимальному решение, которого часто может быть достаточно, а также позволяют ускорить этот процесс поиска. Примером реализации эвристического алгоритма можно назвать Core-directed chain Growth method (CG), предложенный в работе [21] с использованием трехмерной HP модели белка на последовательностях длиной 20–64 элементов. Методы обучения с подкреплением использованы, например, в [38], где авторы показывают результаты с данным методом и применением верхних доверительных границ для деревьев (R-UCT) для последовательностей из 20–162 элементов. В другой работе [58] также применяются методы обучения с подкреплением и метода Монте-Карло для поиска в дереве (MCTS), а также предлагают некоторые подходы к устройству пространств среды и действий. Схожей работой с использованием самообучающихся агентов в виде процесса обучения с подкреплением является [24, 32, 57]. Также были использованы алгоритмы муравьиного роя [2, 47, 52], имитации отжига [54], квантового отжига [45], генетические [55, 60], методы Монте-Карло [59].

Упомянутые алгоритмы имеют свои достоинства и недостатки, так как решают поставленную задачу в определенных ограничениях: по ресурсам, по необходимому уровню оптимизации, по скорости нахождения оптимума, или комбинации перечисленного. В подходе обучения с подкреплением с использованием нейросетевого агента можно настраивать явно или неявно способность агента выполнять эти критерии. Это позволяет масштабировать возможности оптимизационного алгоритма и комбинирует лучшие из свойств отдельных аппроксимационных, комбинаторных или эвристических алгоритмов.

Наиболее перспективными походами к решению задачи можно назвать методы с использованием глубокого обучения и машинного обучения с подкреплением. Такие подходы зарекомендовали себя в других вычислительно сложных задачах, например, для игры в го [48, 49, 50]. В ней на протяжении большого числа шагов сохраняется высокая неопределенность в правильности тех или иных ходов, с чем искусственный интеллект может хорошо справляться. В задаче фолдинга белков выбор конформации в HP-модели как раз может быть сформулирован с этой точки зрения, что говорит о возможности значительно продвинуться в решении задачи при помощи глубокого обучения с подкреплением.

1.3. Предлагаемые подходы и постановка задачи

Целью работы является решение задачи фолдинга (сворачивания) белка для математической сеточной модели HP методами глубокого обучения с подкреплением.

Задачами работы являются следующие пункты:

1. Сформулировать задачу сворачивания белка с использованием методов машинного обучения с учетом биохимических, физических и математических особенностей.
2. Реализовать библиотеку подпрограмм для проведения экспериментов в соответствие с описанными моделями, алгоритмами и методами.
3. Сформулировать и проверить гипотезы, связанные с обучением нейросетевого агента, и верифицировать результаты.

Выдвигаемой в работе гипотезой является получение обученного агента, который не только решает задачу оптимизации энергии, но и делает это в соответствии с физической и биологической моделями белка.

В работе предлагаются методы решения проблемы фолдинга белка при помощи глубокого обучения с подкреплением. Задача сформулирована как оптимизация свободной энергии свернутого белка в пространстве состояний. Такое пространство учитывает расположение исходной последовательности гидрофобных и гидрофильных остатков аминокислот. Моделью взаимодействия является HP модель белка и представляет собой дискретизацию двумерного пространства на элементы сетки, в каждом из которых возможно разместить такие остатки. Энергия сворачивания представлена дискретной функцией подсчета числа связей между гидрофобными аминокислотными остатками, являющихся смежными при расположении на сетке, но не чередующимися в исходной последовательности, представленной упорядоченным набором двух типов остатков.

С точки зрения машинного обучение предложены две архитектуры нейросетей – в виде набора полносвязных слоев и в виде комбинации сверточных и полносвязных слоев. Для обучения с подкреплением рассмотрены два алгоритма – PPO (Proximal Policy Optimization) и DQN (Deep Q-Learning). Сформулирована и реализована среда для обучения агента сворачиванию в нативную конформацию, а также предложен дизайн награды для правильного обучения. Также предлагаются несколько последовательностей, изученных в других работах, для оптимизации реализованным агентом. Для верификации результатов можно визуализировать действия агента, а также сравнить число связей между гидрофобными аминокислотными остатками в свернутом состоянии.

Предполагается, что агент сможет обучиться по правилам среды действовать так, чтобы в конечном результате получить максимальную награду, то есть найти конформацию белка с минимальной энергией. Особенностью тут является то, что агент не обладает знаниями извне о биологической роли белка, его синтезе в организме, о физическом строении, а также не может использовать полученные человеком известные пары последовательность-структура для изученных белков.

ГЛАВА 2. МОДЕЛИ И МЕТОДЫ ФОЛДИНГА БЕЛКА

2.1. Биохимическое устройство белков

Для постановки задачи фолдинга белка в первую очередь необходимо рассмотреть биологические функции аминокислот и белка, их строение и роль в живом организме. Это позволит понять процессы, происходящие на молекулярном уровне, что поможет при качественной и количественной оценке результатов экспериментов, поставленных в текущей работе.

Поскольку белки являются сложными структурами, следует рассмотреть весь цикл их синтеза в организме, начиная с ДНК (дезоксирибонуклеиновая кислота). ДНК обеспечивает хранение генетической информации и передачу следующим поколениям. При помощи процесса транскрипции из ДНК образуется мРНК (матричная рибонуклеиновая кислота), которая является посредником при передаче информации, закодированной в ДНК к рибосомам, а тРНК переносит аминокислоты для синтеза белка. Далее из тРНК (транспортная РНК) и аминокислот образуются белки – цепочки аминокислот, соединенных между собой. Свойства цепочки аминокислот зависят от свойств ее компонент, а также от того, какую форму имеет образовавшийся белок, как он взаимодействует со средой, что определяет уникальную роль этого белка в организме. Они могут быть частью иммунной системы в виде антител, быть ферментами (то есть катализаторами химических реакций), быть сигнальными гормонами (они контролируют протекание физических процессов, например, инсулин снижает содержание глюкозы в крови, лептин регулирует аппетит), переносить кислород (гемоглобин выполняет эту функцию), формировать мышцы и так далее. Таким образом, белки играют одну из важнейших ролей в организме.

Аминокислоты состоят из центрального атома углерода, карбоксильной группы, аминогруппы и радикала (рис.2.1.). Как правило, в организмах 20 основных аминокислот, которые образуют белки, синтезируемые организмом.

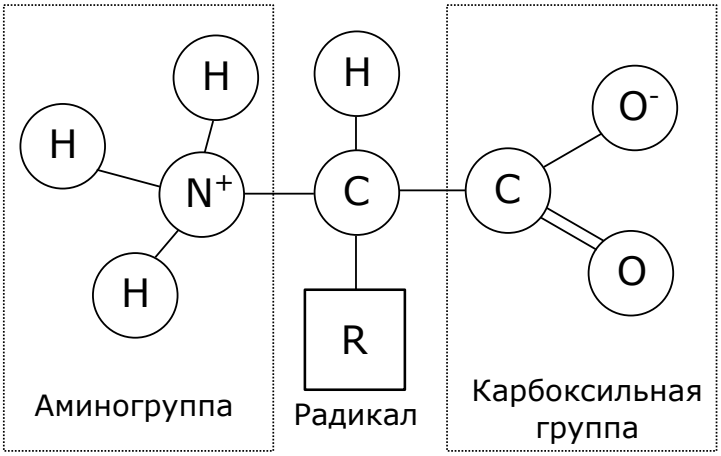

Рисунок 2.1 - Структура аминокислоты

Далее белки имеют несколько уровней структурной организации (рис.2.2.):

1. первичная структура – последовательность ковалентно связанных пептидными связями аминокислот, образующих белок;
2. вторичная структура – соединение между собой разных участков пептидной цепи для ее укладки в более компактную структуру, при котором взаимодействуют пептидные группы и происходит образование между ними водородных связей (примером такой структуры являются параллельный и антипараллельный β-листы и α-спираль);
3. третичная структура – пространственное строение всей молекулы или взаимодействие боковых групп;
4. четвертичная структура – порядок в пространстве, в котором разные полипептиды образуют более сложную структуру.

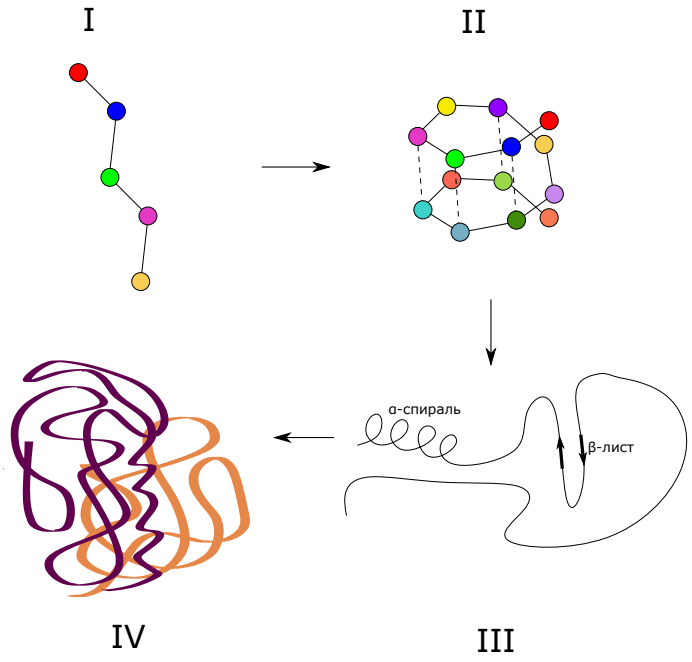

Рисунок 2.2 - Структура белка

Пусть имеется гидроксиаминокислота серин. В составе его боковой группы есть атом кислорода, который сильно электроотрицательный, то есть он отнимает электроны. У атома водорода низкая электроотрицательность относительно атома кислорода, он будет отдавать электроны и на нем будет частично положительный заряд. Такая группа полярна, и эта часть молекулы гидрофильной. Если же рассмотреть валин, у которого боковой радикал углеводородный (в нем нет полярности), то он будет гидрофобным. Боковые группы сильно влияют на структуру белка. Упомянутые аминокислоты могут образовывать цепочку и связываться пептидными связями.

Гидрофобные боковые радикалы стараются избежать контакта с водой, поэтому “прячутся” внутрь молекулы белка, что влияет на ее форму изгиба. Гидрофильные группы в водном растворе находятся с внешней стороны белка и взаимодействуют с водой, и, возможно, образуют водородные связи с другими боковыми группами, что может удерживать структуру белка.

Потеря высокоуровневой структуры при сохранении последовательности аминокислот называется денатурацией белка. Такое может произойти при взаимодействии с определенными химическими компонентами, при изменении температуры окружающей среды или при изменении уровня pH. В такой ситуации связи, обеспечивающие трехмерную структуру белка, разрушаются, и он может принять первичную структуру в виде последовательности аминокислот. Примером данного необратимого процесса является варка или жарка куриного яйца, в котором прозрачный сырой белок становится белым. Однако, некоторые белки способны к ренатурации, то есть возвращать свою исходную природную структуру. Если белки могут самостоятельно переходить от неструктурированного состояния в упакованное, то есть самоорганизовываться, то в аминокислотах содержится вся необходимая для фолдинга информация.

Фолдинг белка – это процесс спонтанного сворачивания полипептидной цепи в уникальную нативную пространственную структуру, то есть переход из первичной структуры в третичную. В живой клетке этот процесс начинается еще при синтезе белка на рибосоме и занимает несколько минут.

Проблемой фолдинга белка называется задача предсказания нативной пространственной структуры белка по данной аминокислотной последовательности. Важность фолдинга белков в организме заключается в том, что неправильное сворачивание белков может привести как к потере определенной функции из-за недостатка белков для ее выполнения, так и к повреждению окружающей клеточной среды. Все это может стать причиной заболеваний, наиболее известные из них – болезнь Альцгеймера, Паркинсона, Хантингтона, Крейтцфельдта-Якоба:

1. Болезнь Альцгеймера – наиболее распространенная форма деменции, симптомами являются расстройство памяти, нарушение речи и когнитивных функций. От этой болезни невозможно излечиться и замедлить ее развитие, существующие меры являются лишь паллиативными.

2. Болезнь Паркинсона – хроническое неврологическое заболевание, вызванное разрушением и гибелью нейронов, вырабатывающих нейромедиатор дофамин, что приводит к серьезным физическим отклонениям. Лекарства от этого заболевания не существует, но предлагаются методы лечения, позволяющие улучшить качество жизни больных и замедлить ход развития болезни.

3. Болезнь Хантингтона (также синдром Гентингтона) – генетическое заболевание нервной системы, характеризующееся психическими расстройствами и непроизвольными движениями в мышцах из-за ошибочных команд головного мозга. Болезнь неизлечима, существующие лекарства позволяют облегчить некоторые из симпотомов болезни.

4. Болезнь Крейтцфельдта-Якоба – дистрофическое заболевание коры большого мозга, базальных ядер и спинного мозга, что приводит к слабоумию, миоклонии (непроизвольные кратковременные сокращения мышц) и другим патологиям центральной нервной системы. Смерть больных наступает в 100% случаев, а лечение может быть только симптоматическим.

Перечисленные болезни, причиной которых является неправильное свертывание белка, неизлечимы и плохо изучены, что является серьезной проблемой в современном мире, для которой нет решения. Таким образом, возможность предсказания структуры белка и динамики сворачивания по его аминокислотной последовательности имеет широкий спектр применения в таких сферах, как профилактика заболеваний и проектирование лекарств.

Чтобы посмотреть на проблему со стороны энергии и процессов, связанных с ней, нужно обратиться к физическому описанию.

2.2. Математическая модель биофизических процессов

При попытке понять физическую основу проблемы фолдинга белков часто обращаются к известному парадоксу Левинталя, который сформулирован американским молекулярным биологом Сайрусом Левинталем в 1969 году [57]. Для этого необходимо рассмотреть белковую цепь. У нее есть большое число возможных состояний – каждый аминокислотный остаток имеет около 10 возможных конформаций, тогда цепочка из 100 остатков будет иметь порядка 10^{100} вариантов. Переход из одной конформации в другую для белка составляет порядка 10^{-13} секунды минимум, тогда перебор всех возможных структур займет около 10^{80} лет. Для сравнения, возраст вселенной по современным оценкам составляет около 14 миллиардов лет, то есть немного больше, чем 10^{10} лет.

Но, как было сказано, фолдинг белка в организмах происходит в течение нескольких секунд или минут, то есть белок не перебирает все возможные варианты конформаций, а имеет что-то, что позволяет это сделать быстро и эффективно, и этим знанием человек еще не обладает. Таким образом, невозможно экспериментально перебрать все возможные нативные структуры белка, так как придется потратить на это на несколько порядков больше времени, чем возраст вселенной. Парадокс в том, что промежуток времени, за который полипептид приобретает нативную пространственную структуру, на много порядков меньше, чем если бы белок просто перебирал все возможные свои конфигурации. Левинталь предположил, что нативная структура белка определяется кинетическими процессами, а не термодинамическими, что значит достижение не глобальному, а просто быстро достижимому минимуму свободной энергии цепочки.

Однако, компьютерные вычислительные эксперименты показывают обратное – сворачивание белка находится под термодинамическим контролем, а не кинетическим, так как цепи аминокислотных последовательностей складываются в

нативную конфигурацию, соответствующую именно глобальному минимуму энергии и являющейся самой стабильной из всех возможных.

На рис.2.3. изображены примеры того, как белок при различных параметрах, описывающих его трехмерную структуру, может иметь различную энергию. Пространство параметров представлено в виде одного параметра по горизонтальной оси графика, однако в действительности таких параметров может быть много. На первом графике (отмечено (а), см. рис.2.3.) представлено описание того, о чем предположил Левинталь – есть лишь единственная конформация белка такая, что энергия минимальна. На втором графике (отмечено (б), см. рис.2.3.) – то, что скорее всего происходит при фолдинге – существует несколько возможных конформаций с локальным минимумом энергии, что соответствует неправильной структуре белка, и существует глобальный минимум энергии взаимодействия молекул белка, что соответствует нативной структуре.

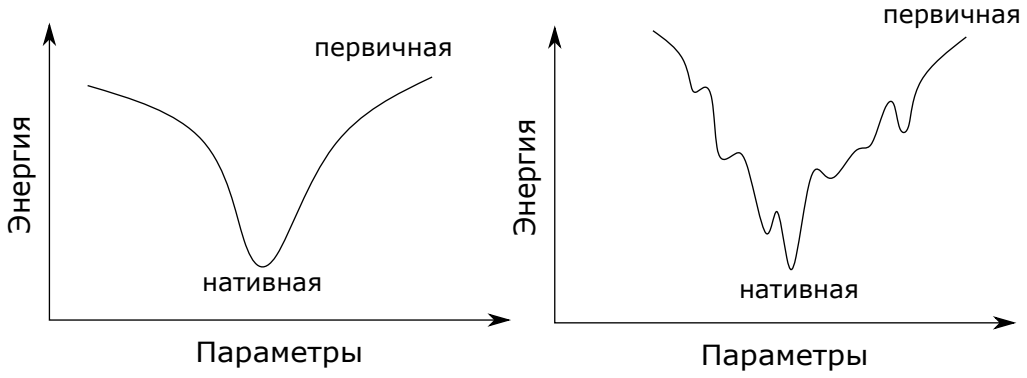

Рисунок 2.3 - Визуализация параметрического пространства с функцией, отображающей значение энергии: (а) соответствует описанию модели с одной единственной возможной конформацией белка в нативном состоянии и соответствующему ей минимуму энергии, (б) – более реалистичное описание состояний в пространстве параметров

Нативная и близкая к ней конформация могут иметь близкие значения энергии, и эта разница определяется щелью между оптимумами. Эта щель имеет ширину во много раз больше, чем $k_B T_{melt}$, где k_B называется постоянной Больцмана (она определяет связь между температурой и энергией $k_B \approx 1.38 * 10^{-23}$ Дж/К), T_{melt} – температура плавления белка. Введение такой щели необходимо для того, чтобы обеспечить процесс сворачивания белка в формате, когда есть разделение высоким барьером свободной энергии нативного и денатурированного состояния.

Для понимания того, как учитывается энергия, можно рассмотреть ячеистую структуру воды на молекулярном уровне. Пусть имеется выражение для свободной энергии:

$$\Delta F = \Delta U - T\Delta S.$$

Здесь ΔF – результирующая свободная энергия, ΔU – изменение энергии, T – температура, ΔS – энтропия.

Ячеистая структура воды стабильна, так как при образовании такой структуры энтропия понижается, значит $\Delta S < 0$, следовательно, $-T\Delta S > 0$, что приводит к увеличению свободной энергии. Эта величина компенсируется тем, что ΔU – энергия взаимодействия, связывающая воду (они обусловлены водородными связями), – большая по абсолютному значению величина и отрицательна, покрывающая второе слагаемое в вышеупомянутой формуле.

Для формирования математической модели фолдинга белка следует рассмотреть основные типы взаимодействий, которые обеспечивают формирование и поддержание структуры белка. Такими взаимодействиями являются [6]:

1. Вандерваальсовы (любые два атома слабо притягиваются, когда находятся на близком расстоянии, что происходит из-за колебаний электрического поля).

2. Водородные связи (взаимодействия между аминокислотными остатками с гидроксильными группами).

3. Электростатические (кулоновские взаимодействия, возникают между противоположно заряженными радикалами).

4. Ковалентные поперечные связи (способность к образованию ковалентной связи).

5. Гидрофобные (некоторые из R-групп аминокислотных остатков стараются расположиться так, чтобы избежать контакта с водой, то есть быть заключенными внутри).

В первую очередь необходимо рассмотреть взаимодействия Ван-дер-Ваальса. Пусть имеются два диполя (рис.2.4.). Они притягиваются друг к другу, так как расстояние между разноименными зарядами, которые притягиваются, меньше, чем расстояние между одноименными зарядами, которые отталкиваются.

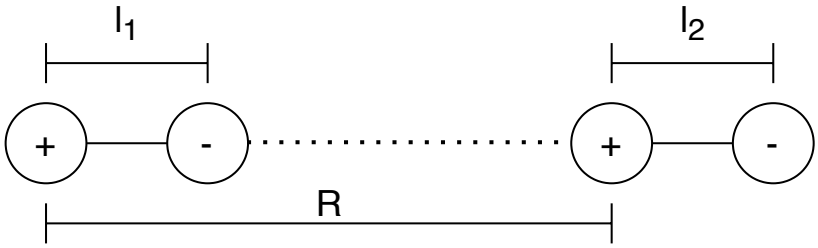

Рисунок 2.4 - Модель двух диполей

Энергия вандерваальсового взаимодействия в данном случае будет иметь вид [9]:

$$u(R) = -\frac{e^2}{R - l_1} - \frac{e^2}{R + l_2} - \frac{e^2}{R + l_2 - l_1} + \frac{e^2}{R},$$

$$R >> l_1, l_2 \Longrightarrow u(R) = -\frac{2e^2 l_1 l_2}{R^3} = -\frac{2P_1 P_2}{R^3}.$$

Если усреднить энергию взаимодействия по всем ориентациям, то получится, что она обратно пропорциональна расстоянию R^6:

$$u(R) = -\frac{2P_1^2P_2^2}{3k_BTR^6} \sim \frac{1}{R^6}.$$

Эта модель необходима при образовании глобул белка, в котором находятся диполи (например, группа NH_2), тогда энергия будет иметь такую зависимость от расстояния.

Хотя в модели белка почти все радикалы нейтральны, в этом случае также возможны взаимодействия, называемые мгновенными наведенными диполями, что соответствует дисперсионным взаимодействиям.

Величина энергии валентных взаимодействий в главной последовательности равна:

$$u_{valence} \sim 20 \div 400 \text{ кДж} \cdot \text{м}^{-1} \sim 5 \div 100 \text{ ккал} \cdot \text{м}^{-1}.$$

Дисперсионные взаимодействия меньше:

$$u_{disp} \sim 1 \div 13 \text{ кДж} \cdot \text{м}^{-1} \sim 1 \div 3 \text{ ккал} \cdot \text{м}^{-1}.$$

Это меньше, чем энергия валентных взаимодействий, но больше, чем энергия тепловых колебаний при температуре около 25 градусов Цельсия:

$$k_BT \sim 2.5 \text{ кДж} \cdot \text{м}^{-1} \sim 0.6 \text{ ккал} \cdot \text{м}^{-1}.$$

Таким образом, вандерваальсовы взаимодействия стабилизируют глобулу, но в среднем они больше, чем энергия тепловых колебаний, что позволяет поддерживать структуру.

Выражение для энергии имеет вид (также называется потенциалом Леннарда-Джонса):

$$u(R) = -\frac{A}{R^6} + \frac{B}{R^{12}},$$

$$A = 1.5 \text{ кДж} \cdot \text{нм}^6 \cdot \text{м}^{-1},$$

$$B = 895 * 10^3 \text{ кДж} \cdot \text{нм}^6 \cdot \text{м}^{-1}.$$

Здесь A – коэффициент вклада отталкивания, B – коэффициент вклада притяжения.

Если нейтральные частицы далеко друг от друга, то энергия их взаимодействия очень небольшая. За это отвечает первое слагаемое в выражении для энергии, оно отрицательное, тогда оно соответствует силе притяжения – атомы будут притягиваться. На рис.2.5. это соответствует области справа от R_0. В изолированном состоянии, когда атомы не взаимодействуют, эта энергия равна нулю.

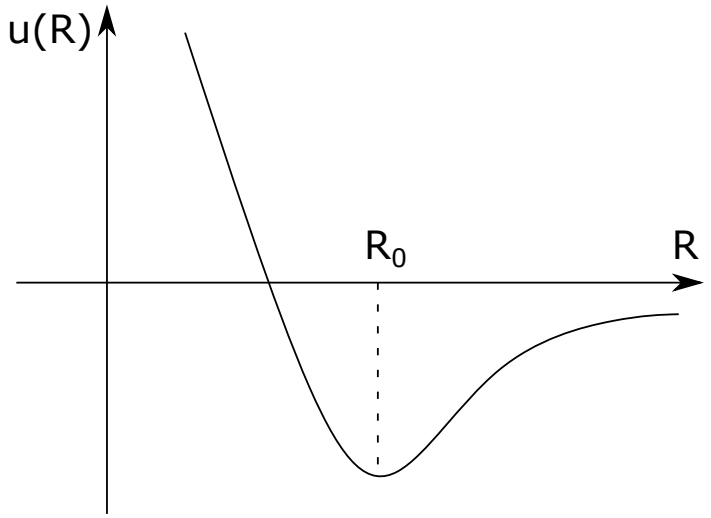

Рисунок 2.5 - Зависимость энергии взаимодействия от расстояния

С другой стороны, если они находятся близко друг к другу, то энергия взаимодействия увеличивается. Это происходит из-за отталкивания электронных обо-

лочек, тогда энергия отталкивания растет как функция, обратно пропорциональная R^{12}, то есть очень быстро. На рис.2.5. это соответствует области слева от R_0.

Выражение для энергии показывает, что есть минимум функции при $R = R_0$. Для некоторых из атомов такие расстояния равны:

$$C \cdots C \Longrightarrow R_0 \approx 0.32\text{нм} = 3.2\text{Å},$$

$$C \cdots O \Longrightarrow R_0 \approx 0.28\text{нм} = 2.8\text{Å},$$

$$H \cdots H \Longrightarrow R_0 \approx 0.2\text{нм} = 2\text{Å}.$$

Эти расстояния порядка двух-трех Å, что соответствует плотному расположению в белке.

Формула полной энергии полипептидной цепочки объемных взаимодействиям имеет вид (ϵ – диэлектрическая проницаемость, q_i – заряды на взаимодействующих атомах, E_ϕ^0 – торсионный барьер (или барьер вращения), ϕ – текущее значение торсионного угла (аналогично для ψ), r – межатомное расстояние):

$$E(\phi,\psi) = \frac{1}{2}E_\phi^0\big(1 - 3\cos(3\phi)\big) + \frac{1}{2}E_\psi^0\big(1 - 3\cos(3\psi)\big) + \\ + \sum_{\text{ij}} 4\epsilon_{ij}\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6} + \frac{q_i q_j}{\epsilon r}.$$

Здесь первые два слагаемых соответствуют торсионному потенциалу (зависит от углов ϕ и ψ), второе слагаемое (сумма) – потенциалу Леннарда-Джонса (вандерваальсовы взаимодействия), а последнее – электростатическим взаимодействиям (если имеются заряды). Это выражение соотносит определенному набору углов поворота одинарных связей структуру и соответствующую ей энергию.

Можно рассмотреть дипептид $CH_3 - CONH - CHR - CONH - CH_3$. В формуле полной энергии для этого дипептида углы принимают значения от -180 до

180 градусов, и функция этих двух переменных должна обладать минимумом, иначе дипептид не будет иметь стабильной структуры. Теоретически показано, что будет иметься два минимума энергии, что соотносится с экспериментами. Переход между стабильными структурами возможен, но маловероятен из-за высокого энергетического барьера. В сложных соединениях, в отличие от дипептидов, таких связей может быть больше.

Экспериментально показано, что в ходе сворачивания белок не перебирает все возможные состояния, а сначала образует α- и β-структуры, которые сохраняются. Поэтому предсказание вторичной структуры можно рассматривать как итеративый процесс, в начале которого находится положение, соответствующее минимуму энергии каких-то выбранных дипептидов из всего соединения, а после этого добавляют третью структуру к ним, и так далее. Причем глобальный оптимум достижим несмотря на выбор начальных дипептидов.

Таким образом, получена комплексная математическая модель описания физических процессов при фолдинге белка. Эта задача является сложной даже на уровне предсказания вторичной структуры.

2.3. Задача фолдинга для HP-модели

Проблема оптимизации представляется как минимизация целевой функции J от некоторого выбранного набора параметров x (в работе это способ укладывания аминокислотных остатков на сетке) при условии выполнения ограничений c:

$$J(x) \to \min_{x \in X} \text{ if } c(x) \leq 0.$$

Целевой функцией здесь будет являться энергия объемных взаимодействий в белке как дискретная функция подсчета связей между остатками по заданным правилам, взятая со знаком минус.

Ограничениями являются:

1. Невозможность расположить более одного элемента цепочки в одну клетку.

2. Заполнение всего замкнутого цепочкой пространства на сетке.

Второе условие естественным образом следует из биохимического и физического описаний: высокоуровневая структура белка содержит наиболее компактно уложенную последовательность аминонокислот, чтобы молекулы воды не могли попасть внутрь белка и разрушить ее.

Гидрофобно-гидрофильная модель (HP, или Hydrophobic-Polar) фолдинга белка представляет собой гидрофобные взаимодействия в белке с помощью обозначения аминокислот как гидрофобные (обозначаются как H, являются неполярными) и гидрофильные (обозначаются как P, являются полярными). Такая модель была широко изучена ранее [25], а задача фолдинга белков, связанная с этой моделью, численно решалась различными оптимизационными алгоритмами, например, муравьиным алгоритмом, стохастические, эволюционные и генетические алгоритмы.

На рис.2.6. приведены два типа модели HP для квадратной сетки: в трехмерном случае и в двумерном случае.

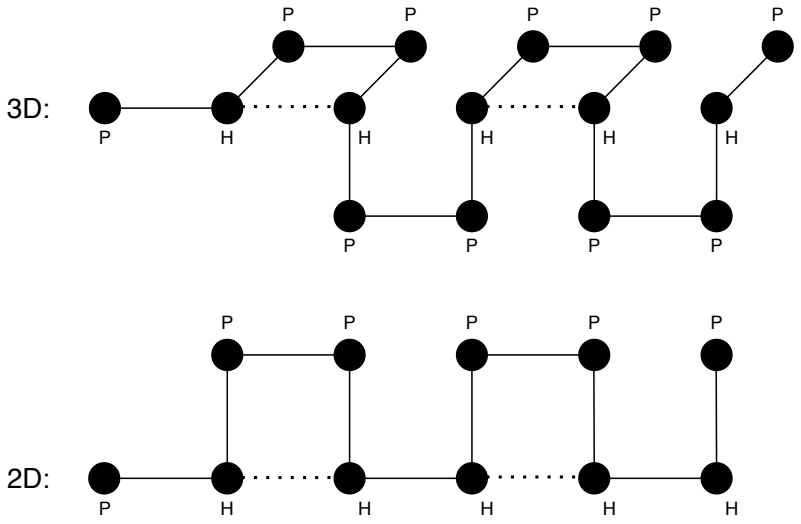

Рисунок 2.6 - HP-модель белка

Задачу, сформулированную в терминах HP-модели, сложно решить перебором даже для небольших последовательностей H и P. С ростом последовательности временная сложность растет экспоненциально, что делает невозможным перебор для больших и более интересных последовательностей.

В работе предложен модуль для расположения элементов на сетке, в котором уже имеются два первых элемента цепочки, так как в любом случае они будут поставлены на сетке, и таким образом можно убрать лишние действия. На каждый следующий шаг можно выбрать одно из четырех действий: поставить элемент слева, снизу, сверху или справа от последнего. Для примера, на использованной конфигурации компьютера алгоритм перебора всех комбинаций обрабатывает порядка 10^6 итераций за час в одном потоке, где каждая итерация является полным расположением последовательности на двумерной сетке и подсчетом энергии сворачивания как числа связей между H-остатками. Чтобы найти нативную структуру белка, описанного всего лишь 18 элементами, нужно $4^{16} \approx 4 * 10^9$ попыток, что займет примерно 4000 часов вычислений. При увеличении цепочки лишь на 1 элемент время на расчеты увеличивается на порядок. Это означает, что даже упрощенная до двумерного случая модель фолдинга белка и при предположении о том, что придется работать только с двумя типами элементов – H и P, перебор не подходит для решения задачи для больших цепочек.

В такой оптимизационной задаче применяют эвристические оптимизационные алгоритмы, позволяющие проводить поиск по всем параметрам системы. Однако, из-за своей сложности, они неэффективны для больших последовательностей аминокислотных остатков и требуют больших вычислительных мощностей в виде суперкомпьютеров, а также время на эти вычисления. В этом случае может помочь решение задачи в терминах обучения с подкреплением, в котором агент, набравшись опыта, сможет научиться действовать наиболее оптимально для заданных правил среды.

ГЛАВА 3. АЛГОРИТМЫ И ПРОГРАММЫ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ФОЛДИНГА БЕЛКА

3.1. Глубокое обучение с подкреплением

Обучение с подкреплением является областью машинного обучения, которая формулирует решение задачи как взаимодействие нейросетевого агента со средой. Обучение агента происходит при получении обратной связи от среды в виде награды или штрафа за те или иные действия, таким образом позволяет настроить агента на решение оптимизационной задачи.

На рис.3.1 представлена условная схема такого процесса обучения.

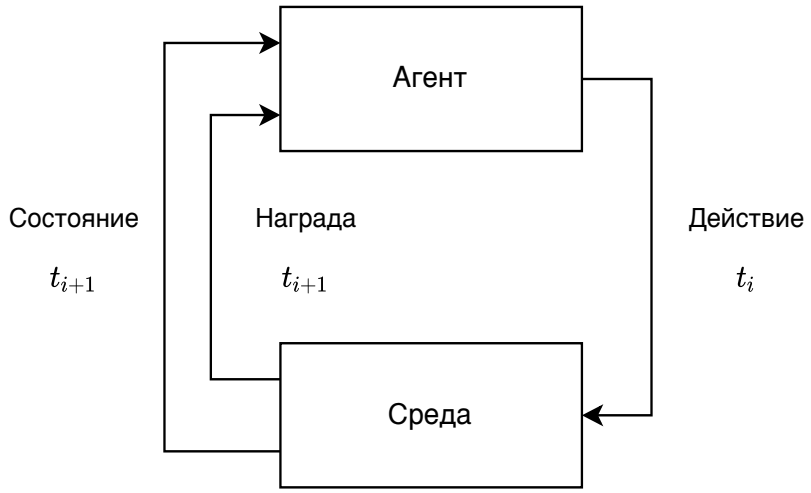

Рисунок 3.1 - Схема обучения с подкреплением

При такой формулировке задачи нет соответствий между входными данными и правильными ответами, как в классическом машинном обучении. Агент в ходе обучения пытается соблюдать баланс между тем, чтобы исследовать новые способы получения награды – это называется исследованием (exploration) – и ис-

пользовать опыт взаимодействия для других действий – это эксплуатация (exploitation).

Для большинства задач глубокого обучения с подкреплением используется алгоритм PPO, так как он наиболее эффективно позволяет обучать агента на практике, имеет относительно простую идею, что позволяет модифицировать его для конкретной задачи в случае необходимости. Также алгоритм реализован в часто используемых фреймворках для обучения с подкреплением, таких как RLLib и Stabel Baselines 3. Основная идея PPO в том, чтобы при помощи имеющихся данных, собранных агентом в ходе исследования действий в среде, сделать максимально большой шаг в обучении политики, но не зайти слишком далеко, чтобы испортить процесс обучения.

Когда говорят о PPO, обычно идет речь о PPO-Clip, хотя есть еще вариант PPO-Penalty. Их отличие заключается во введении ограничений при обновлении политики. В данной работе использован метод PPO-Clip. Алгоритм PPO-Clip можно наблюдать ниже в виде псевдокода [55]:

1. Задать начальные параметры политики θ_0, начальные параметры функции ценности ϕ_0.
2. Для каждого $k = 0,1,2, \ldots$:
    a. Собрать набор траекторий $\mathcal{D}_k = \{\tau_i\}$, используя политику $\pi_k = \pi(\theta_k)$ на действиях в среде.
    b. Вычислить награды $\hat{R}_t$.
    c. Вычислить оценки функций преимуществ $\hat{A}_t$ (используя любой из методов оценки преимущества), основываясь на текущей функции значимости V_{ϕ_k}.

d. Обновить параметры θ_{k+1} политики, максимизируя целевую функцию (обычно при помощи стохастического градиентного подъема).

e. Подогнать параметры ϕ_{k+1} функции ценности, используя среднеквадратичную ошибку (обычно при помощи градиентного спуска).

Таким образом, описанный алгоритм является основным методом обучения агента с подкреплением. Архитектура самого агента и среды рассмотрены далее.

3.2. Архитектура среды и модель агента

Среда представляет собой структуру данных, в которой находится сетка с размещенными на ней аминокислотами на какой-либо шаг. Ее можно представить как марковский процесс принятия решений – МППР, в котором множество состояний и множество действий конечно, то есть этот процесс финитный. Модель обучения с подкреплением состоит из следующих множеств и их элементов:

1. $s_t \in S$ – состояние на момент времени (то есть на какой-то дискретный шаг) t из множества состояний;
2. $a_t \in A(s_t)$ – действие из множества действий;
3. $r_t \in R(s_{t+1})$ – награда за действие a_t, которое приводит среду в состояние s_{t+1}.

На каждом шаге t агент может выбрать одно из четырех действий, которое приведет среду в новое состояние, и получает награду за это действие. В ходе таких взаимодействий со средой агент получит опыт, который должен помочь ему

научиться такой политике действий $\pi: S \to A$, которая максимизирует награду за эпизод $R = \sum_{t=0}^{T} r_i$.

Выбор действия на каждом шаге представлен на рис.3.2. и представляет из себя переход в ячейки, расположенную слева, снизу, сверху или справа от последнего расположенного на сетке элемента.

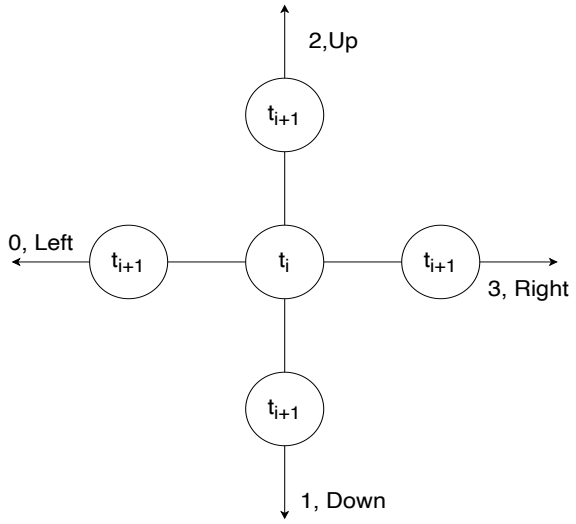

Рисунок 3.2. Варианты действий на каждый шаг в виде выбора одного объекта из множества $\{0,1,2,3\}$

В простейшем случае рассматривается только одна заданная последовательность аминокислотных остатков, что в конечном результате позволяет обучить одного агента для оптимального размещения этой последовательности. Этот способ подходит для проверки главное идеи работы, однако для большей генерализации предлагается такой дизайн среды, что при каждом обновлении будет автоматически генерироваться новая последовательность аминокислотных остатков. Таким образом, агент, предобученный на произвольных случайно сгенерированных последовательностях, сможет делать предсказание структуры белка в свернутом состоянии для последовательности, которой не было в обучающей выборке.

В качестве агента в конечном алгоритме используется полносвязная нейронная сеть со скрытыми нейронными слоями по 512 нейронов каждый, функция активации – ReLU. Такая архитектура имеет под собой основу, схожую по сути с многослойным перцептроном, в котором имеется входной вектор значений $x \in \mathbb{R}^n$ и выходной вектор $y \in \mathbb{R}^m$. Их связывают скрытые слои, в данном случае их число равно 2.

На рис.3.3. приведена условная схема перцептрона с набором входных и выходных параметров и несколькими скрытыми слоями.

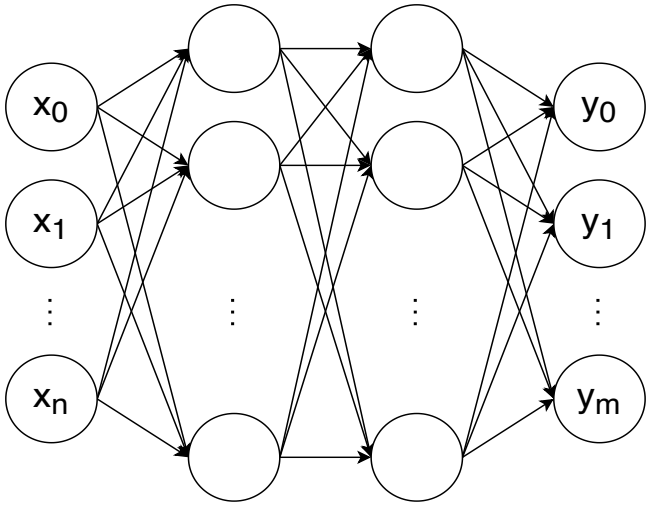

Рисунок 3.3 - Архитектура многослойного перцептрона

Так как наблюдения среды можно представить как несколько двумерных массивов чисел, то для больших последовательностей аминокислот полносвязная нейронная сеть не сможет учесть такое большое входное пространство, а также не будет принимать во внимание те особенности на исходных картах, которые могут быть полезными для агента. Для решения этой проблемы предлагается использовать сверточную нейронную сеть, что доказало свою эффективность в других задачах глубокого обучения.

В случае полносвязной реализации, выходами нейросети являются два набора блоков – набор чисел, описывающих вероятности совершить определенное действие (обозначено на рис.3.4. Policy), и одно число, описывающее состояние (обозначено Value).

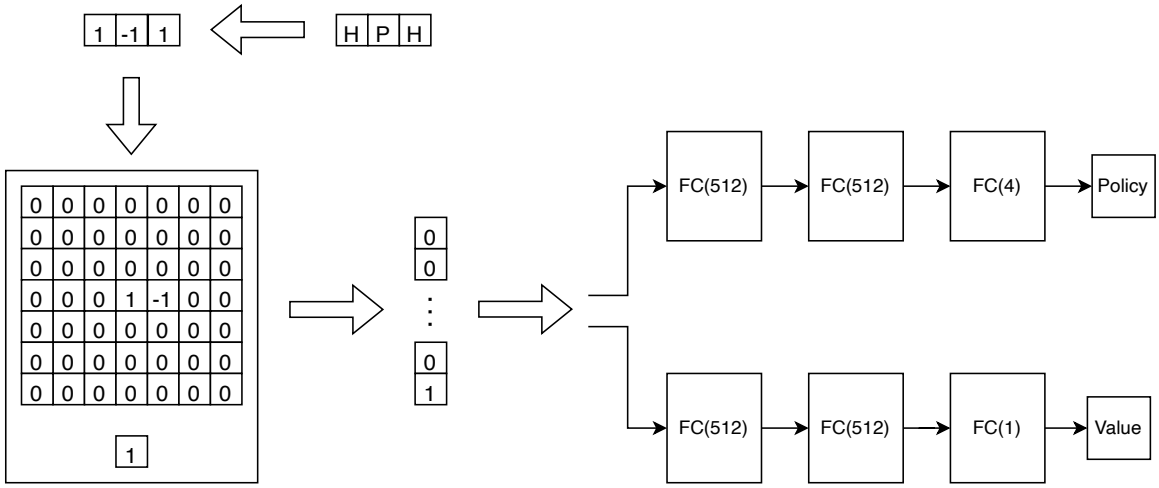

Рисунок 3.4. Архитектура полносвязного агента FC

Здесь и далее: $FC(n)$ означает выходную размерность – n элементов, $Conv(a,(b,b))$ – сверточный слой с a фильтрами и ядром свертки b на b. Функция активации в обоих случаях – ReLU (Rectified Linear Unit).

Сверточная архитектура отличается от полносвязной тем, что в случае перцептрона каждый нейрон следующего слоя связан с нейроном предыдущего слоя, и каждая такая связь имеет собственный коэффициент, называемый весом. В случае сверточного слоя, для разных нейронов используют небольшую матрицу весов, которую двигают с небольшим сдвигом (stride, в данной работе он равен 1) по матрице предыдущего слоя. Таким образом, она нацелена на распознавание образов, что эффективно, например, в классификации изображений и других задачах машинного обучения, и, возможно, в данной задаче. Аналогично FC, CNN нейросеть имеет набор блоков, но уже со свертками (рис.3.5.).

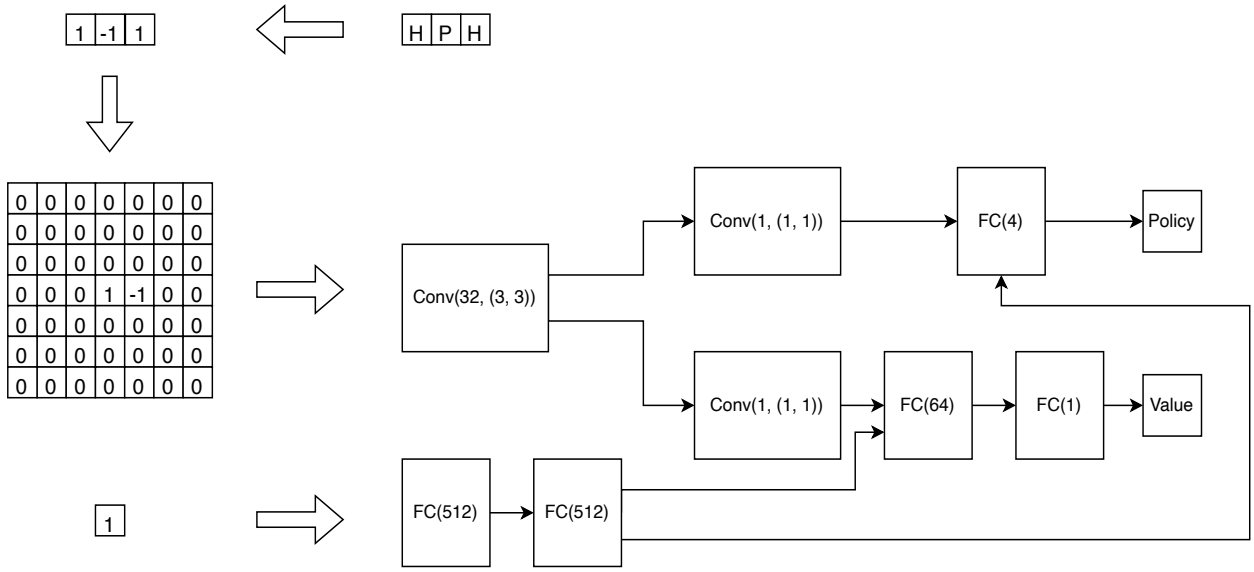

Рисунок 3.5 - Архитектура сверточного агента CNN

В настоящий момент времени стали широко распространены более сложные архитектуры нейросетей, позволяющие более эффективно работать с изображениями, матрицами и графами, относительно полносвязных и сверточных сетей. Поэтому логично сделать предположение о том, что применение графовых нейронных сетей, остаточных сверточных нейронных сетей, или визуальных трансформеров значительно улучшит способность агента решать задачу фолдинга. Кроме того, представление пространства состояния, использованное в данной работе, имеет недостатки, – большое количество значений, неинформативно существующих в матрице среды, – можно заменить на представление в виде графа, что должно улучшить информативность признаков, подаваемых агенту на вход. Также пространство действий в дальнейшем следует заменить, как минимум, на 3 направления вместо 4, потому что нет смысла предсказывать агенту такое действие, сообщающее постановку нового элемента на предыдущий, – в этом случае также можно использовать маскирование действий, широко применяющееся в работах с большим пространством действий.

3.3. Дизайн награды

Для того, чтобы обучить агента правильно располагать последовательности на сетке, нужна обратная связь от среды за действия агента. Так, в случае корректных с точки зрения задачи действий, то есть, например, заполнение свободной ячейки сетки следующим элементом, агент должен либо получать награду, либо не получать штраф, как это сделано в данной работе. За некорректные действия, такие как пересечения уже расположенной цепочки новым элементом, агенту сообщается штраф и происходит окончание эпизода.

В промежуточных шагах агент не получает награды, пока не происходит окончание эпизода. Алгоритм формирования награды приведен далее:

1. Принять: текущий шаг t, максимальное число шагов в эпизоде n, функция подсчета свободной энергии $energy < 0$ от состояния S_t, переменная наличия пересечений $collision$, штраф за пересечение $penalty > 0$.
2. $state_reward \leftarrow 0$.
3. $collision_reward \leftarrow 0$.
4. $reward \leftarrow 0$.
5. Если $t \geq n$, то: $state_reward \leftarrow -energy(S_t)$.
6. Если $collision$, то: $collision_reward \leftarrow -penalty$.
7. $reward \leftarrow state_reward + collision_reward$.
8. Возвратить $reward$.

Функция подсчета энергии устроена так, чтобы сообщить число связей между смежными на сетке H-элементами, но только между теми, которые не стоят ря-

дом в исходной последовательности. Все остальные связи сообщают нулевую энергию. Число таких связей возвращается функцией с обратным знаком, поэтому большее число связей H-H сообщает меньшую энергию. Штраф за пересечение цепочки новым элементом выбран как -1, чтобы агент всегда получал знание о том, что цепочку нельзя пересекать. Награду агент получает только в конце, что усложняет понимание того, как агенту двигаться внутри шагов в среде, однако в итоге позволяет не отвлекаться на получение небольшой награды, чтобы при правильном расположении элементов получить большую награду. Также идейно отказ от промежуточной награды обусловлен тем, что в любом из случаев агент должен расположить всю цепочку белка на сетке, и только в этом случае эпизод будет считаться корректным.

ГЛАВА 4. АНАЛИЗ РЕЗУЛЬТАТОВ РЕШЕНИЯ ЗАДАЧ ИССЛЕДОВАНИЯ И ОЦЕНКА ИХ ПРАКТИЧЕСКОЙ ЗНАЧИМОСТИ

4.1. Постановка экспериментов

Для проведения экспериментов конечное решение представлено в виде библиотеки с набором классов среды, модели агента, файлов конфигураций, скриптов для работы с данными, которые написаны на языке программирования Python. Основным фреймворком для работы с обучением с подкреплением стала библиотека RLLib, для написания моделей используется Torch, а для описания среды взята библиотека Gym. Вся библиотека реализована в модульном формате, чтобы возможно было изменить каждый из компонентов пайплайна, проверять другие гипотезы и усложнять задачу, не меняя методов ее решения или, наоборот, варьируя методы, не меняя правила среды.

В работе предложены следующие эксперименты:

1. Сравнение алгоритмов PPO и DQN обучения с подкреплением. Предполагается, что один из этих методов будет более подходящим в задаче, так как они используют разные способы обучения агента. DQN оптимизирует награду на напрямую, а прогнозирует значения для Q-функции, сохраняет переходы в буфере и случайно обучается на них. PPO напрямую оптимизирует награду, обновляя политику при помощи только самых последних переходов состояний. Сравнить алгоритмы возможно при помощи графиков обучения агента, заметив тенденции более быстрого роста средней награды и достижения более высокого уровня максимальной награды. Выбор этих методов обусловлен их популярностью в применении к другим задачам обучения с подкреплением. Эта популярность вызвана стабильность методов и доказанной и показанной в других работах эффективностью. Данные алго-

ритмы предполагается не имплементировать, а взять готовые решения из фреймворка RLLib, и добавить их в пайплайн для обучения агента.

2. Сравнение нейросетевых реализаций агентов FC и CNN. Для этого необходимо имплементировать сами архитектуры агентов в виде кода, а после этого сравнить графики обучения агентов, по аналогии с предыдущим пунктов сравнения PPO и DQN. Архитектура FC выбрана потому, что FC является базовой моделью, которую можно применить в решении задачи при помощи обучения с подкреплением, в силу своей простоты в имплементации и устройстве, а также наличии многочисленной теории по описанию математической модели сети. Сверточные нейросети часто используют при работе с изображениями и представлениями в виде матриц, что также актуально для текущей постановки задачи. Популярные фреймворки, которые были применены для разработки агентов в данной работе, содержат наборы полносвязный и сверточных слоев, что позволяет собрать из них конечную архитектуру нейронной сети. В работе предложены комбинации этих слоев для реализации соответствующих архитектур, которые интуитивно подходят для решения задачи, но не реализованы в открытом доступе для данной задачи, поэтому итоговая библиотека включает в себя реализацию автором данной работы этих нейросетей.

3. Проведение расчетов на известных последовательностях аминокислотных остатков, являющихся бенчмарками. Здесь предполагается количественно оценить при помощи заранее выбранной метрики насколько хорошо агент находит конформации с минимальным уровнем энергии. Для этого известные последовательности из бенчмарка с открытыми данными подаются на вход отдельному агенту для оптимизации, после чего необходимо сравнить результаты в таблице и оценить средние значения для большего понимания результатов. Также предлагается изучить результаты из других статей с те-

ми же бенчмарками, что позволит сделать выводы о том, получилось ли превзойти решение, предложенное в других работах, или нет.

На основе выполненных экспериментов следует сформулировать выводы, в особенности связанные с физическим и биологическим описанием и математической моделью, полученной в предыдущих главах.

Конфигурация компьютера для расчетов выглядит следующим образом:

1. Процессор 1,4 GHz Quad-Core Intel Core i5 – все вычисления выполнены на CPU без использования GPU.
2. Число потоков для одновременного обучения агентов – 7.

В такой конфигурации эксперимент по обучению агента на одной небольшой последовательности (в итоговой таблице соответствует 18 элементам) занимает около 15 минут для FC-модели и около часа для CNN-модели. При увеличении числа элементов в два раза, то есть до 36 элементов, время обучения увеличивается в 2–4 раза.

4.2. Результаты исследований и графики обучения

В качестве результатов предложены графики обучения агента для случая сравнения алгоритмов PPO и DQN, для случая сравнения архитектур агента FC и CNN, а также пример обучения для одного агента.

На рис.4.1. представлены графики обучения агентов при помощи алгоритмов PPO и DQN. Можно заметить, что хоть DQN в самом начале обучения позволил агенту быстрее найти способ получить большую награду за меньшее число шагов, график обучения получился очень нестабильным и не достиг того уровня

награды, который получилось достичь методом PPO. Также PPO показал большую устойчивость относительно DQN.

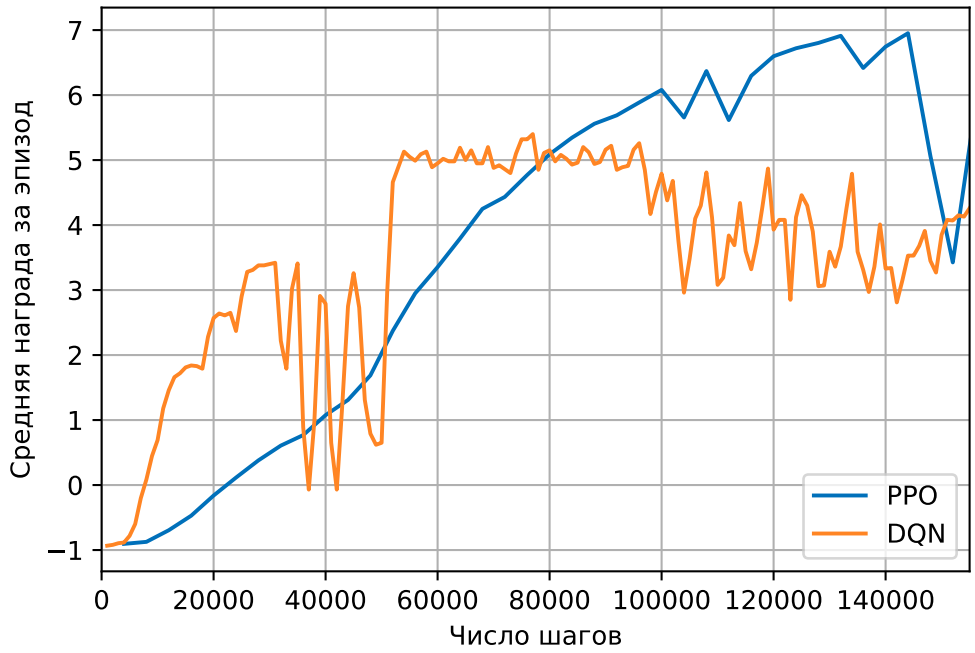

Рисунок 4.1 - Сравнение алгоритмов обучения: PPO и DQN

Далее приведены сравнительные результаты по использованию нейросетевых агентов FC и CNN. Из рис.4.2. видно, что обучение сверточной архитектуры занимает меньше шагов, сделанных агентом в среде, чем в случае полносвязных скрытых слоев, что играет роль при обучении по двум причинам. Первая причина заключается в том, что проведение расчета награды в среде за эпизод хоть и нетребовательный процесс относительно обучения нейросети, например, но все равно занимает какое-то время. Таким образом, чем меньше сделано шагов в среде, тем быстрее агент будет обучаться при прочих равных условиях. Вторая причина в том, что если имеется GPU (graphics processing unit), то обучение сверточной сети занимает значительно меньше времени, чем на CPU (central processing unit), что также уменьшает время обучения агента.

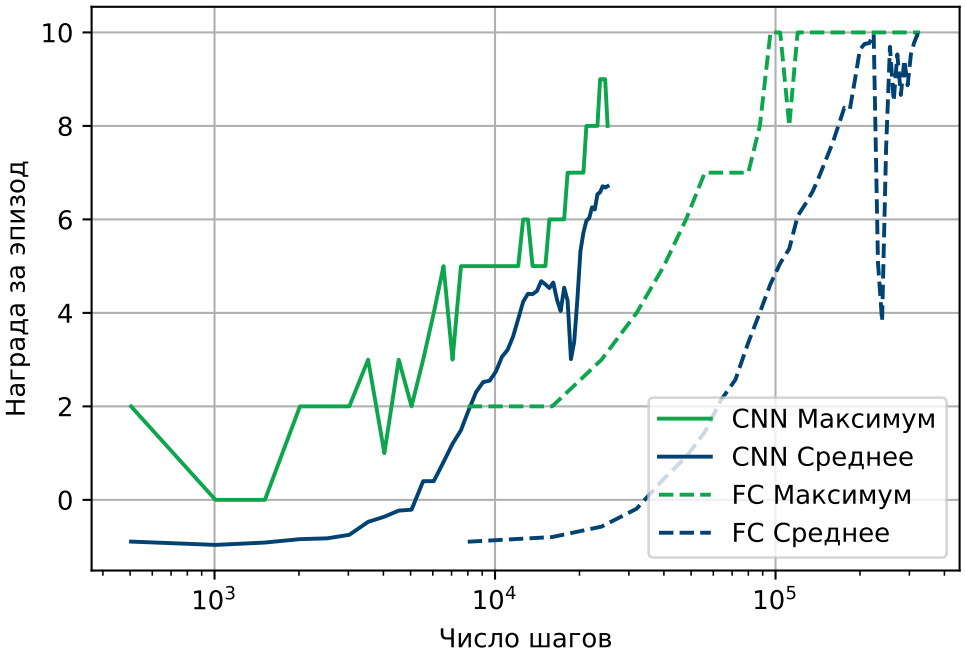

Рисунок 4.2 - Сравнение обучения нейросетевых агентов: CNN и FC

Итоговые результаты обучения агента можно наблюдать на рис.4.3., который показывает сходимость минимального, среднего и максимального значений наград за эпизод к одному значению. На практике такой график позволяет понять обучился ли агент, – если все три функции сходятся к одному значению, то можно говорить об успешном обучении.

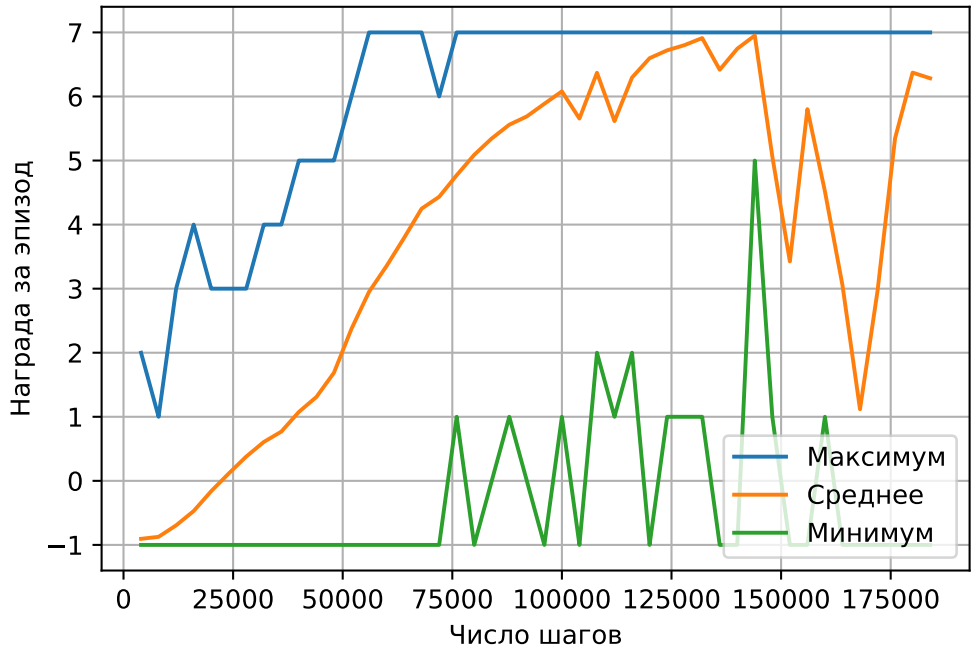

Рисунок 4.3 - Изменение награды по шагам в процессе обучения агента

Полученный график важен с точки зрения обучения нейронных сетей, так как показывает увеличение средней награды агента за его действия. Это является следствием того, что агент делает не просто случайные действия, которые он совершал в самом начале своего обучения, но осмысленно выбирает пути, по которым строить цепочку. Он делает это таким образом, чтобы избежать штрафа за пересечение уже поставленных элементов и максимизировать награду, что означает минимизацию энергии взаимодействия.

Итак, оптимальным для проверки гипотез с точки зрения обучения в данной постановке задачи и экспериментов является использование FC агента с алгоритмом обучения PPO. Эта комбинация позволила получить конформации белков с самым низким уровнем энергии с возможность проведения экспериментов на персональном компьютере небольшой мощности за приемлемое время (не более часа на эксперимент).

4.3. Пример фолдинга и верификация результатов

Для верификации результатов используется набор бенчмарков, предложенных для двумерной квадратной сетки Университетом Брауна. Этот набор включает в себя последовательности аминокислотных остатков, представленных буквами H и P, длиной от 18 до 100 элементов с числом связей от 4 до 50. В данной работе использованы только первые 8 из них из-за высокой вычислительной сложности даже самых коротких.

Полный набор проведенных исследований приведен далее в таблице. Для проверки того, что агент совершает адекватную сборку белка, далее представлен фолдинг для двух начальных последовательностей, которые агент расположил после обучения.

В первом случае, изображенном на рис.4.4., агент пытался оптимизировать последовательность из 18 элементов с энергией -8, соответствующей 8 связям – на рисунке отмечены толстой связью. Видно, что оптимальная с точки зрения награды конформация выглядит как расположение гидрофобных остатков в центр структуры, а гидрофильных – снаружи. Это совпадает с представлениями о том, как белок выглядит в действительности. Описать это представление будет нагляднее при помощи следующего примера фолдинга.

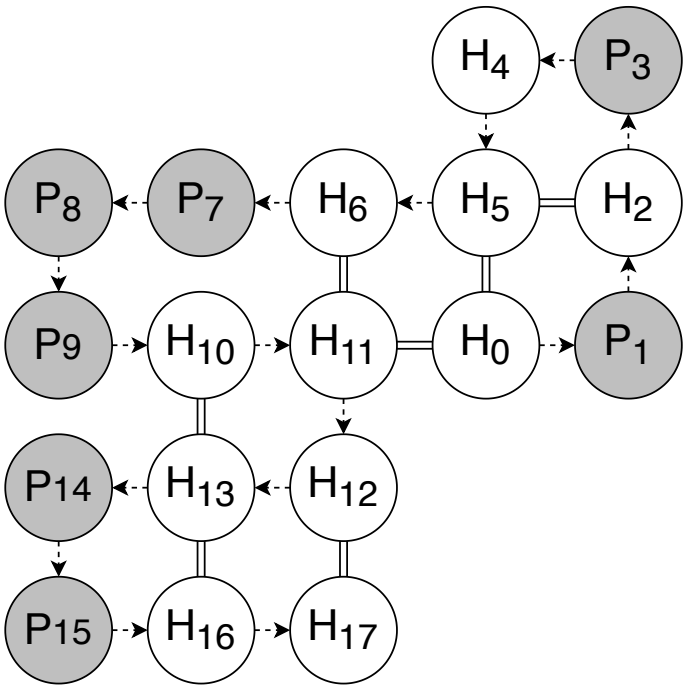

Рисунок 4.4 - Пример свернутого белка с 7 связями для последовательности №2 из таблицы, структура которого найдена агентом

На рис.4.5. наблюдается более сложная последовательность, содержащая уже 20 элементов и имеющая нативную конформацию с уровнем энергии -10, что соответствует 10 связям между элементами H. Агент смог понять какие действия предпринять, чтобы образовать такую конформацию. Таким образом, агент решил задачу для данной последовательности за меньшее число шагов, чем просто перебор всех возможных действий. Объяснить это можно тем, что агент использовал в ходе своего обучения знания о том, какая награда следует за какие действия, то есть, используя этот опыт, стал действовать наиболее оптимально с точки зрения награды.

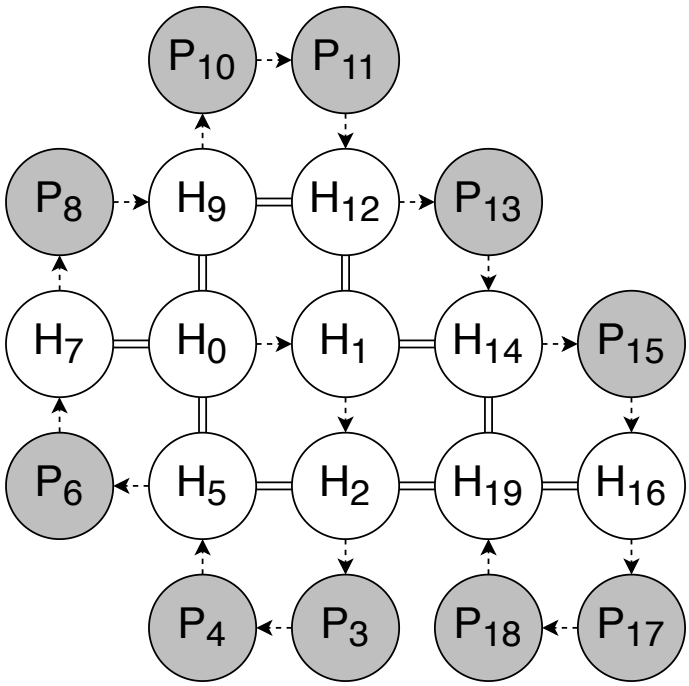

Рисунок 4.5 - Пример свернутого белка с 10 связями для последовательности №5 из таблицы, структура которого найдена агентом

С другой стороны, следует отметить структуру белка, образованную на рис.4.5. Можно наблюдать глобулу, в которой гидрофобные аминокислотные остатки образовали кластер внутри глобулы, а гидрофильные – снаружи. Это совпадает со строением глобулярных белков в организме (например, гемоглобин). Для того, чтобы полимер имел минимальную энергию взаимодействия (а это в свою очередь необходимо для поддержания структуры), внутри глобулы должны находиться гидрофобные элементы, которые не взаимодействуют с водой, иначе структура была бы нестабильной. А снаружи должен образоваться слой гидрофильных остатков, то есть тех, которые могут взаимодействовать с молекулами воды. При этом важно, что внутри такой структуры нет пропусков, как и в реальном белке, – такие свободные пространства в плотно набитой глобуле привели бы к возможности разрушить архитектуру энергетических связей.

Также полученную структуру белка на рис.4.5. можно объяснить с топологической точки зрения. На нем наблюдается наличие внутренней и внешней границы структуры, которую агент собирает из начальной последовательности. С точки зрения биологии такая структура, разделяющая свойства среды и объекта, обеспечивает устойчивость к внешней среде.

Далее стоит перейти к расчетам на остальных начальных последовательностях. Приведенные в таблице далее результаты выполнены с использованием нейросетевого агента с полносвязными скрытыми слоями. Обучение с использованием сверточных слоев дает схожие результаты, поэтому они не приведены в работе.

Примеры последовательностей, число связей для них и коэффициент аппроксимации. Результаты получены с использованием только FC модели

№	Длина	Последовательность	$n_{optimal}$	n_{found}	k
1	18	HHPPPPPHHPPPHPPPHP	4	3	0.75
2	18	HPHPHHHPPPHHHHPPHH	8	7	0.88
3	18	PHPPHPHHHPHHPHHHHH	9	8	0.89
4	20	HPHPPHHPHPPHPHHPPHPH	9	7	0.78
5	20	HHHPPHPHPHPPHPHPHPPH	10	10	1.00
6	24	HHPPHPPHPPHPPHPPHPPHPPHH	9	7	0.78
7	25	PPHPPHHPPPPHHPPPPHHPPPPHH	8	5	0.63
8	36	PPPHHPPHHPPPPPHHHHHHH PPHHPPPPHHPPHPP	14	7	0.50

Для оценки результатов таблицы стоит ввести коэффициент аппроксимации (approximation ratio) k, который равен отношению результата, полученного алго-

ритмом, к оптимальному результату. В задаче фолдинга этим результатом может служить как энергия взаимодействия, так и число связей. Пусть для определенности этот коэффициент равен следующему выражению:

$$k = \frac{n_{found}}{n_{optimal}}.$$

Тогда из таблицы можно сделать выводы о том, что агент находит близкие к оптимальным значениям энергии конформации белка. В одной из последовательности (под номером 5 в таблице) агент смог решить задачу с коэффициентом аппроксимации $k = 1$ относительно бенчмарка, несмотря на большую длину цепочки. Для более длинных цепочек агент также научился строить последовательности с некоторым числом связей между гидрофобными остатками, хоть и не так близко подошел к оптимальному решению. В среднем коэффициент равен $k = 0.84$ для последовательностей 1-6, что больше примерно на 15%, чем показано в другой работе по решению задачи фолдинга белка с использованием глубокого обучения с подкреплением – в [58] средний коэффициент для тех же последовательностей составил $k = 0.69$ (имеются в виду только результаты с использованием глубокого обучения с подкреплением). Если обратиться к последовательностям по всей таблице, приведенной в текущей работе, то результат упомянутой метрики уменьшится до $k = 0.77$, но все равно остается высоким результатом относительно упомянутых при обзоре литературы аналогичных работ.

ЗАКЛЮЧЕНИЕ

В работе рассмотрена и решена задача фолдинга белков с использованием машинного обучения с подкреплением. Задача сформулирована с биологической, физической и оптимизационной точек зрения, создана математическая модель сворачивания белка.

Автором реализован агент и среда, с помощью которой происходит обучение агента, а также выбран алгоритм, реализация которого взята из фреймворка RLLib с открытым исходным кодом, для обучения с подкреплением нейросетевого агента. Для агента придуманы и проверены две архитектуры нейронных сетей – FC и CNN, а также имплементированы в виде кода в библиотеке. Оптимизационная задача сформулирована как минимизация энергии как функции конформации белка. Обученный агент позволяет для конкретной последовательности аминокислотных остатков предсказать третичную структуру. Среда реализована с использованием двумерной квадратной HP-модели белка, а награда, возвращаемая агенту за его действия, позволяет оценить энергию связей как число смежных на сетке гидрофобных остатков, но не идущих последовательно в исходной цепочке. Дизайн среды и награды также выбран и реализован в виде модулей библиотеки. Написанный код в виде библиотеки позволил провести эксперименты, в ходе которого был выбран алгоритм PPO, агент с архитектурой FC, а также приведено обоснование выбора.

Также в ходе работы были изучены релевантные исследования с различных сторон процесса фолдинга, наиболее важные факты из них упомянуты, проведено сравнение их с результатами работы.

Полученные результаты экспериментов говорят о том, что:

1. Для данной постановки задачи агент находит близкие к оптимальным или оптимальные решения, что следует из сравнения коэффициентов аппроксимации, приведенных в таблице.

2. Алгоритм PPO в сравнении с DQN показывает большую устойчивость и превосходит по максимальной награде, которую агент получает в конце эпизода.

3. Средняя за эпизод награда для агента с использованием сверточных слоев сходится к максимальной награде за меньшее число шагов в среде, чем агент, реализованный с полносвязными слоями.

4. Полученные результаты на бенчмарках показывают успешное решение агентом задачи оптимизации с точки зрения введенного коэффициента аппроксимации – на главных последовательностях коэффициент аппроксимации равен $k = 0.84$, а в некоторых случаях этот коэффициент приближается к единице или достигает ее.

5. Качественная оценка действий агента говорит о верном подходе к действиям агента с точки зрения биологического строения белка и физических особенностей, связанных с энергией.

Дальнейшие шаги в исследовании темы фолдинга белка и улучшении результатов могут включать в себя:

- Работу над средой и представлении признаков структуры белка в более сжатом формате.
- Изменение дизайна агента на более сложные и современные архитектуры нейронных сетей для того, чтобы учесть большее число свойств структуры белка.

- Увеличение размерности пространства HP-модели белка до 3 измерений для более правильной с физической точки зрения постановки задачи.
- Проведение экспериментов с большими вычислительными ресурсами на более длинных последовательностях аминокислотных остатков.
- Изменение модели HP белка на непрерывную и, соответственно, более сложную.

Следует выделить качественные результаты по анализу структуры полученного решения с топологической точки зрения. Так как интерпретация действий нейросетевого агента критически важна при решении задачи, следует особенно внимательно подойти к анализу результатов. Фолдинги, полученные в работе, подтверждают гипотезу о том, что агент находит оптимальное решение, обоснованное с точки зрения биологии и физики. Он делает это несмотря на то, что в явном виде не обладает знаниями о том, что правильно, а лишь действует исходя из награды, которую он получает от среды.

Таким образом, поставленные задачи выполнены, а начальная гипотеза о том, что полученный обученный агент не только решает задачу оптимизации энергии, но и делает это в соответствии с биофизической и биохимической моделями белка, проверена и подтверждена.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Абдуллаева Н.Ф. Влияние болезни Паркинсона на фолдинг белка. Инновационные научные исследования, 2022. 2-2(16), 11–17.
2. Бекмуратов Т.Ф. Реализация муравьиного алгоритма фолдинга белков методами программных агентов в распределенных системах / Т. Ф. Бекмуратов, Р. К. Базаров, Д. К. Базаров // Проблемы вычислительной и прикладной математики, 2017. – № 2(8). – С. 103-113. – EDN YPASKR.
3. Галиев Ш.И. Математическая логика и теория алгоритмов. - Казань: Издательство КГТУ им. А. Н. Туполева, 2002. - 270 с.
4. Кайгородова Е.В., Кайгородова Е.В., Завьялова М.В., Завьялова М.В., Богатюк М.В., Перельмутер В.М., Перельмутер В.М. Особенности внутриклеточной локализации молекулярного шаперона Hsp27 в опухолевых клетках рака молочной железы // Международный журнал прикладных и фундаментальных исследований, 2014. – № 4. – С. 201-202;
5. Калмыкова З.А., Кононенко И.В., Смирнова О.М., Шестакова М.В. Сигнальные пути гибели β-клеток при сахарном диабете 2 типа: роль врожденного иммунитета. Сахарный диабет, 2020. 23(2):174-184.
6. Макеев А.В. Основы биологии. - М.: Мир, 1997. - 235 с.
7. Огнивенко В.М., Плетень А.П. Взаимосвязь фолдинга белков с возникновением патологий» нервной системы (обзор), Материалы международной научно - практической конференции «Новая наука: опыт, традиции, инновации», 2016. - с. 26 - 31.
8. Рубин А.Б. Биофизика. В двух книгах. 2015.
9. Рубин А.Б. Биофизика. Видеолекции. 2019.

10. Финкельштейн А.В., Птицын О.Б. Физика белка. Курс лекций с цветными и стереоскопическими иллюстрациями и задачами. — 4-е издание, исправленное и дополненное. — Москва: Книжный дом Университет, 2012. — 524 с. — 500 экз. — ISBN 978-5-98227-834-0.

11. Хельтье Х., Зиппль В., Роньян Д., Фолькерс Г. Молекулярное моделирование. Теория и практика, 2010.

12. Чачанидзе Е.Р. Сравнительный анализ алгоритмов Proximal Policy Optimization и Soft-Actor-Critic // E-Scio, 2020. №5 (44).

13. Achiam J., Held, D., Tamar A., Abbeel P. Constrained Policy Optimization, 2017.

14. Achiam J., Joshua P. Spinning Up in Deep Reinforcement Learning, 2018.

15. Ashraf G.M., Greig N.H., Khan T.A., Hassan I., Tabrez S., Shakil S., Sheikh I.A., Zaidi S.K., Akram M., Jabir N.R., Firoz C.K., Naeem A., Alhazza I.M., Damanhouri G.A., Kamal M.A. Protein misfolding and aggregation in Alzheimer's disease and type 2 diabetes mellitus. CNS Neurol Disord Drug Targets, 2014.

16. Backofen B., Will S. A constraint-based approach to fast and exact structure prediction in three- dimensional protein models. Constraints, 11(1):5–30, 2006.

17. Bahi J., Côté N., Guyeux C., Salomon M. Protein Folding in the 2D Hydrophobic-Hydrophilic (HP) Square Lattice Model is Chaotic. Cognitive Computation. 4. 98-114, 2012.

18. Bao G. Protein Mechanics: A New Frontier in Biomechanics. Exp Mech, 2009.

19. Bergasa-Caceres F., Rabitz H.A. Interdiction of Protein Folding for Therapeutic Drug Development in SARS CoV-2. The Journal of Physical Chemistry B, 2020.

20. Berger B., Leighton T. Protein folding in the hydrophobic-hydrophilic (HP) is NP-complete. In Proceedings of the second annual international conference on

Computational molecular biology (RECOMB ’98). Association for Computing Machinery, New York, NY, USA, 30–39, 1998.

21. Beutler T.C., Dill K.A. A fast conformational search strategy for finding low energy structures of model proteins. Protein Science, 5(10):2037–2043, 1996.

22. Carr R., Hart W.E., Newman A. Bounding a protein’s free energy in lattice models via linear programming. Poster at RECOMB04, 2004.

23. Clote P. Protein structure prediction on the face centered cubic lattice by local search. In Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, pages 241–246, 2008.

24. Czibula G., Bocicor M.-I., Czibula I.-G. A reinforcement learning model for solving the folding problem. International Journal of Computer Technology and Applications, 2:171–182, 2011.

25. Dill K.A. Theory for the folding and stability of globular proteins. Biochemistry, 1985.

26. Finkelstein A.V. 50+ Years of Protein Folding. Biochemistry (Mosc), 2018.

27. Hart W., Istrail S. Fast Protein Folding in the Hydrophobic–Hydrophilic Model within Three-Eighths of Optimal. Journal of computational biology: a journal of computational molecular cell biology, 1996.

28. Hart W., Istrail S. Lattice and off-lattice side chain models of protein folding: linear time structure prediction better than 86% of optimal. Journal of Computational Biology, 4(3):241–259, 1997.

29. Höltje H., Folkers G., Reviewer A.L. Molecular Modeling, Basic Principles and Applications. Computers in Physics, 12, 41-41, 1996.

30. Hoque M., Chetty M., Dooley L. Fast Computation of the Fitness Function for Protein Folding Prediction in a 2D Hydrophobic-Hydrophilic Model. International Journal of Simulation: Systems, Science and Technology, 2005.

31. Istrail S., Lam F. Combinatorial Algorithms for Protein Folding in Lattice Models: A Survey of Mathematical Results. Commun. Inf. Syst. 9 (4) 303 - 346, 2009.

32. Jafari R., Javidi M.M. Solving the protein folding problem in hydrophobic-polar model using deep reinforcement learning. SN Applied Sciences, 2, 1-13, 2020.

33. Javadi Y., Fernandez J.M., Perez-Jimenez R. Protein Folding Under Mechanical Forces: A Physiological View. Physiology, 2013.

34. Jumper J., Evans R., Pritzel A. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589, 2021.

35. Karplus M. The Levinthal paradox: yesterday and today. Fold Des. 2(4):S69-75, 1997.

36. Lau K.F., Dill K.A. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. Macromolecules, 22(10), 3986–3997, 1989.

37. Levinthal C. How to Fold Graciously. Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois. J. T. P. DeBrunner and E. Munck eds., University of Illinois Press Pages 22-24, 1969.

38. Li Y., Keng H., Ye K., Yin S., Li X. FoldingZero: Protein Folding from Scratch in Hydrophobic-Polar Model, 2018.

39. Lin C., Su S. Protein 3D HP Model Folding Simulation Using a Hybrid of Genetic Algorithm and Particle Swarm Optimization. International Journal of Fuzzy Systems. 13. 140-147, 2011.

40. Lyngsø R.B., Pedersen C.N.S. Protein Folding in the 2D HP Model. BRICS Report Series, 6(16), 1999.

41. Mann M., Will S., Backofen R. Cpsp-tools–exact and complete algorithms for high-throughput 3d lattice protein studies. BMC bioinformatics, 9(1):230, 2008.

42. Mnih V., Kavukcuoglu K., Silver D., Graves A., Antonoglou I., Wierstra D., Riedmiller M.A. Playing atari with deep reinforcement learning. CoRR abs/1312.5602, 2013.

43. Nauta M. From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI. arXiv e-prints, 2022.

44. Newman A. A new algorithm for protein folding in the hp model. In Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms, pages 876–884. Society for Industrial and Applied Mathematics, 2002.

45. Perdomo-Ortiz A., Dickson N., Drew-Brook M., Rose G., Aspuru-Guzik A. Finding low-energy conformations of lattice protein models by quantum annealing. Scientific reports, 2:571, 2012.

46. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal Policy Optimization Algorithms. 2017.

47. Shmygelska A., Hoos H.H. An ant colony optimisation algorithm for the 2D and 3D hydrophobic polar protein folding problem. BMC Bioinformatics 6, 30, 2005.

48. Silver D., Huang A., Maddison C. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489, 2016.

49. Silver D., Schrittwieser J., Simonyan K. Mastering the game of Go without human knowledge. Nature 550, 354–359, 2017.

50. Silver D., Hubert T., Schrittwieser J., Antonoglou I., Lai M., Guez A., Lanctot M., Sifre L., Kumaran D., Graepel T., Lillicrap T., Simonyan K., Hassabis D. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science. Dec 7;362(6419):1140-1144, 2018.

51. Sutton R.S., Barto A.G. Reinforcement learning: An introduction, second ed., The MIT Press, 2018.

52. Thalheim T., Merkle D., Middendorf M. Protein folding in the hp-model solved with a hybrid population based aco algorithm. IAENG International Journal of Computer Science, 35(3):291–300, 2008.

53. Tunyasuvunakool K., Adler J., Wu Z. Highly accurate protein structure prediction for the human proteome. Nature 596, 590–596, 2021.

54. Ullah A.D., Steinhöfel K. A hybrid approach to protein folding problem integrating constraint programming with local search. BMC bioinformatics, 11(1):S39, 2010.

55. Unger R., Moult J. Genetic algorithms for protein folding simulations. Journal of molecular biology, 231(1):75–81, 1993.

56. Wang Z., Freitas N., Lanctot M. Dueling network architectures for deep reinforcement learning, CoRR abs/1511.06581, 2015.

57. Wu, H., Yang, R., Fu, Q. Research on predicting 2D-HP protein folding using reinforcement learning with full state space. BMC Bioinformatics 20, 685, 2019.

58. Yu H. Deep Reinforcement Learning for Protein Folding in the Hydrophobic-Polar Model with Pull Moves. NeurIPS, 2020.

59. Zhang J., Kou S.C., Liu J.S. Biopolymer structure simulation and optimization via fragment regrowth monte carlo. The Journal of chemical physics, 126(22):06B605, 2007.

60. Zhang X., Wang T., Luo H., Yang J.Y., Deng Y., Tang J., Yang M.Q. 3D protein structure prediction with genetic tabu search algorithm. BMC Syst Biol. May 28;4 Suppl 1(Suppl 1):S6, 2010.