

СОВРЕМЕННЫЕ МЕТОДЫ МОДЕЛИРОВАНИЯ И ПРЕДСКАЗАНИЯ СТРУКТУР И ВЗАИМОДЕЙСТВИЯ БИОМОЛЕКУЛ

Федотова Полина Алексеевна, студентка 2 курса Института агrobiотехнологии, ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, polifedou@yandex.ru

Научный руководитель – Шипунова Виктория Олеговна, к.б.н., заведующая лабораторией биохимических исследований канцерогенеза, МФТИ, Физтех, viktoriya.shipunova@phystech.edu

Аннотация. Обзор посвящен современным методам, используемым для моделирования и предсказания 3D-структур белков, что является важной задачей биоинформатики. Рассмотрены два биоинформатических инструмента – AlphaFold, основанный на гомологии, и Rosetta с алгоритмом моделирования *de novo*.

Ключевые слова: структура белка, дизайн белков, взаимодействие биомолекул, машинное обучение, моделирование

В современном мире проблема сворачивания трёхмерной (3D-) структуры белков до сих пор не является полностью решенной.

С одной стороны, именно первичной структурой определяется дальнейший фолдинг, нативная структура белка; есть данные о том, какие именно аминокислоты чаще всего образуют те или иные вторичные структуры (спирали, складки или повороты), но с другой – это не может дать нам ответ на то, как будут взаимодействовать между собой данные вторичные структуры, и мы не можем точно сказать в итоге, как же на самом деле будет выглядеть синтезированный белок. На данные вопросы могут помочь ответить такие биоинформатические инструменты, как программы для дизайна белков, вычисления конформационных подсостояний, программы предсказания 3D-структуры и активных сайтов биомолекул с использованием нейронных сетей, машинного обучения и др.

Существует множество моделей, пытающихся объяснить процесс сворачивания белков, среди них три основных: каркасная модель (*the framework model*), модель гидрофобного коллапса, или расплавленной глобулы (*the hydrophobic collapse model*) и модель доменообразования (*the nucleation model*). Каркасная модель предполагает, что сначала формируются локальные элементы вторичной структуры; затем они диффундируют вместе, сталкиваются и слипаются, образуя правильную третичную структуру. Модель гидрофобного коллапса подразумевает, что белок быстро коллапсирует вокруг своих гидрофобных боковых цепей, а

затем перестраивается из ограниченной конформации этого промежуточного продукта типа «расплавленной глобулы». Модель доменообразования утверждает, что локальные взаимодействия формируют небольшое количество нативной вторичной структуры, которая действует как ядро для распространения наружу дальнейшей нативной структуры. Однако и в случае этих моделей нельзя определенно сказать, каким же образом белок принимает свою нативную конформацию, так как в живых организмах, кроме гидрофобных и электростатических взаимодействий, функционируют шапероны и др. Также существуют проблемы в определении различных взаимодействий биомолекул: белок-белковых взаимодействий, взаимодействий белок-лиганд, белок-рецептор и т.д. Данные проблемы требуют очень большого объема знаний, в большей степени в молекулярной биологии, биофизике и в частности физике белка.

Так как проблемы достаточно обширные и глубокие, решение их с помощью человеческого ресурса является очень трудозатратным и долгим, есть вероятность не учесть какие-то особенности 3D-структуры. На помощь в решении данных проблем приходят компьютерные технологии, позволяющие обрабатывать огромные объемы данных намного быстрее. Раньше такие программы использовали только специалисты из-за их сложной настройки, необходимости знаний вычислительной биологии, программирования, а также из-за недостатка компьютерных вычислительных мощностей. В последнее десятилетие компьютерные мощности значительно возросли, также увеличилось количество разработок, как и потребность в сервисах. На данный момент уже существуют «*user-friendly*» сервисы, которые позволяют не углубляться в сложные физические основы, программирование и другие технические аспекты, при этом позволяют легко пользоваться сложными и очень важными для молекулярной биологии и для биологии в целом биоинформатическими инструментами. Биологи могут сосредоточиться на ключевых исследовательских вопросах, не привлекая штат программистов, что позволяет тратить меньше времени на анализ данных. Данные инструменты могут применяться в самых различных сферах исследований: медицина, фармацевтика, сельское хозяйство – там, где требуется максимально точное предсказание 3D-структуры белка для дальнейшего его использования при разработке медицинских, фармацевтических, ветеринарных и других препаратов.

Самым популярным сервисом предсказания 3D-структур белка по гомологии на данный момент является *AlphaFold* [1]. Он повышает точность предсказания структуры за счет внедрения архитектур нейронных сетей (NN) и процедур обучения, а именно алгоритма глубокого обучения (DL), основанных на эволюционных, физических и геометрических ограничениях белковых структур. Процесс можно

упростить до его работы с двумя входными базами данных: гомологичных последовательностей, обрабатываемых методом множественного выравнивания последовательностей (MSA), и базы возможных трехмерных координат всех тяжелых атомов для данного белка, учитывая физические ограничения. Блок *Evoformer* отражает оценку сети наиболее вероятной структуры в этом блоке. На выходе из данных о корреляции позиций мы получаем распределение вероятностей попарных расстояний. Еще один модуль – структурный, или декодер. Он использует трансформирующую NN, обеспечивает переход от абстрактного представления структуры белка к трехмерным координатам атомов целевых белков. Структурный модуль воспринимает каждый остаток как отдельный объект и прогнозирует вращения и перемещения, необходимые для его размещения. Так, точность данного инструмента может достигать 85 %. Можно отдельно выделить одну из важных сфер применения AlphaFold2: как правило, определение структуры доменов или белков *de novo* с помощью ядерно-магнитного резонанса (NMR), рентгеновской кристаллографии (XRD), электронной криомикроскопии (cryo-EM) требует много временных и финансовых затрат, и благодаря данному инструменту можно относительно заменить данные методы и решить перечисленные проблемы.

Rosetta – программа, разработанная в Институте белкового дизайна под руководством Дэвида Бейкера, она использует алгоритм моделирования *de novo* (с начала) [2]. В отличие от моделирования по гомологии, моделирование *de novo* не зависит от известных структур белка, генерирует трехмерную структуру целевого белка на основе установленных законов физики (квантовой механики). Этот метод проводит поиск конформаций, руководствуясь разработанной энергетической функцией с атомными координатами аминокислот в качестве переменных. В этом процессе образуется множество возможных конформаций, и выбирается та, которая имеет наименьшую энергию. Данный метод зависит от двух факторов: энергетической функции, которая представляет свободную энергию целевого белка по отношению к атомным координатам аминокислот, и эффективного алгоритма конформационного поиска, который может быстро идентифицировать низкоэнергетические состояния.

Таким образом, рассмотренные инструменты упрощают процесс молекулярного дизайна: помогают узнать особенности смоделированного белка до начала эксперимента и правильно подобрать протокол, предсказать особенности активных центров белков, снизить затраты на исследования, повысить точность прогнозирования структур, полученных методами NMR, XRD, cryo-EM, или даже заменить их, предсказать структуру белков, которые сложно поддаются анализу вышеупомянутыми методами, создавать новые современные лекарства для лечения

различных заболеваний. Однако при этом ни один инструмент не является абсолютно точным: невозможно точно сказать, что сконструированный белок будет работать, заменить полноценное исследование *in vitro* и *in vivo*, полностью безошибочно предсказать нативную структуру сложного слитого белка, заменить высококвалифицированного молекулярного биолога и биоинформатика.

Исследование поддержано Минобрнауки РФ, проект FSMG-2023-0015.

Библиографический список

1. AlphaFoldProteinStructureDatabase [Электронный ресурс]. URL: <https://alphafold.ebi.ac.uk/> (дата обращения 21.04.2024).
2. Rosetta Commons [Электронный ресурс]. URL: <https://www.rosettacommons.org/software> (дата обращения 21.04.2024).