

Хадарович А.Ю., Калиновский А.А., Тузиков А.В.

Предсказание структур димерных белковых комплексов с помощью нейронных сетей глубокого обучения

ГНУ «Объединенный институт проблем информатики НАН Беларуси», Минск, Республика Беларусь

Актуальность и цель. Вступая во взаимодействие, белки образуют белок-белковый комплекс. Задача нахождения трехмерной структуры комплекса, образованного при взаимодействии белков, называется белковым докинггом. Как и в случае индивидуальных белков, экспериментальные методы для определения структуры комплекса могут использоваться в ограниченном количестве случаев и требуют длительного времени, а в массовом порядке и вовсе неприменимы. Поэтому ускоренное развитие получили вычислительные методы, кото-

рые позволяют получать трехмерные структуры комплексов быстро, используя в качестве входных данных трехмерные структуры белков, составляющих комплекс [1,2]. Наибольшее значение при этом имеет область связывания белков, или интерфейс белкового комплекса, поскольку именно в данной области находятся аминокислоты, являющиеся необходимыми для образования комплекса, который, в свою очередь, выполняет соответствующую функцию в организме. Это особенно актуально в исследованиях, посвященных созданию потенциальных лекарственных препаратов, где в сжатые сроки нужно получить множество моделей белковых комплексов и оценить их жизнеспособность и влияние на организм. Поэтому актуальной является задача разработки алгоритма моделирования трехмерной структуры белкового комплекса. В данной работе рассматривались только димерные белковые комплексы.

Материалы и методы. Авторами был предложен алгоритм предсказания белковых интерфейсов гомодимеров (области связывания) на основе глубокого обучения и алгоритм предсказания структуры белкового комплекса по предсказанной карте контактов [3]. Алгоритмы были реализованы на языке программирования Python и библиотеки для машинного обучения pytorch [4] и использованы для предсказания набора белковых гомодимерных и гетеродимерных комплексов из базы данных белков PDB [5].

Результаты и выводы. Алгоритм на основе нейронной сети глубокого обучения позволяет ограничить пространство поиска возможных структур при докинге белковых комплексов, существенно упрощая и ускоряя процесс моделирования. Кроме того, анализ полученных результатов показывает, что разработанный алгоритм моделирования структур димерных белковых комплексов на основе известной матрицы контактов позволяет получить на тестовом множестве белковых комплексов корректные трехмерные модели для 94% гетеродимеров и 96 % гомодимеров. Разработанный алгоритм предсказания структуры белкового комплекса по предсказанной карте контактов может использоваться не только для гомодимеров, но и для гетеродимеров, то есть комплексов, состоящих их двух разных белков, поскольку не зависит от количества аминокислотных остатков в одной и другой белковой цепи. Таким образом, результаты исследования свидетельствуют о том, что данный подход является перспективным для решения задачи моделирования димерных белковых комплексов.

Литература

1. Lecun, Y. Deep learning / Y. Lecun, Y. Bengio, G. Hinton // Nature. – 2015. – Vol. 521, № 7553. – P. 436-444.
2. DT Jones, S.K. High precision in protein contact prediction using fully convolutional

neural networks and minimal sequence features / S.K. DT Jones // *Bioinformatics*. – 2018. – Vol. 34. – P. 3308-3315.

3. А. Ю. Хадарович. Предсказание структуры гомодимерных белковых комплексов на основе глубокой нейронной сети / А. Ю. Хадарович, А. А. Калиновский, А. В. Тузиков // *Информатика*. – 2020. – № 17(2). – С. 44–53.

4. PyTorch: An Imperative Style, High-Performance Deep Learning Library / A. Paszke [et al.] // *Advances in Neural Information Processing Systems 32* / Curran Associates, Inc. ; eds. H. Wallach [et al.]. – 2019. – P. 8024-8035.

5. Berman, H. The Protein Data Bank / H. Berman // *Nucleic Acids Res.* – 2000. – Vol. 28. – P. 235-242.