

# ПРЕДИКТИВНАЯ МОДЕЛЬ ПО ОЦЕНКЕ АФФИННОСТИ БЕЛКОВЫХ И ПЕПТИД-БЕЛКОВЫХ КОМПЛЕКСОВ В ШИРОКОМ ДИАПАЗОНЕ КОНСТАНТ ВЗАИМОДЕЙСТВИЙ

Тычинин Д.И, Шайтан К.В, Новоселецкий В.Н.

Биологический факультет МГУ им. М.В. Ломоносова, кафедра Биоинженерии

Проблемы вычислительного определения аффинности белковых комплексов до сих пор остаются актуальными. В настоящее время популярностью пользуются модели машинного обучения и нейросети, в которых аффинность комплекса определяется по набору признаков аминокислотных последовательностей (sequence-based), значительно реже учитываются конформационные особенности белок-белковых взаимодействий.

Объективно sequence-based модели успешно работают для оценки аффинности коротких пептидов в комплексе с малоподвижным участком белка: ярким примером послужит успех моделирования взаимодействия пептидов с главным комплексом гистосовместимости человека, например модели: MHCflurry2.0 [1] и многие другие. Однако, не существует универсального протокола для оценки аффинности систем, обладающих высокой подвижностью и относительно большими поверхностями активных сайтов связывания. В настоящей работе представлен вариант реализации модели машинного обучения для определения аффинности белок-белковых взаимодействий, основанный на структурных данных.

Модель для вычислительной оценки энергии взаимодействия основана на градиентном бустинге с использованием программного пакета LGBM [2]. Обучение модели было выполнено на 494 кристаллографических структурах белковых комплексов отобранных из PDBbind [3]. Валидационная выборка составила 50 комплексов (10% от общего набора данных). Комплексы разделены на классы аффинности по показателям констант взаимодействий. Вычислительно белок-белковое взаимодействие в каждом комплексе представлялось нами в виде матрицы контактов атомов 9 групп различных типов. Расчет расстояний производился с помощью программного пакета MDAnalysis [4]. AUC полученной модели составил 0.841, что превосходит результаты других моделей [5].

1. Timothy J., et.al MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing, Cell Systems, 21 October 2020, Pages 418-419

2. Guolin Ke, et.al. LightGBM: a highly efficient gradient boosting decision tree, NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems December 2017 Pages 3149–3157

3. PDBBind, URL: [pdbind.org.cn](http://pdbind.org.cn)

4. Richard J. Gowers, et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference, pages 98–105, 2016.

5. Wajid et al, Learning protein binding affinity using privileged information, BMC Bioinformatics volume 19, Article number: 425 (2018)