

# ПРЕДСКАЗАНИЕ ЛОКАЛЬНОЙ СТРУКТУРЫ БЕЛКА ПО ПОСЛЕДОВАТЕЛЬНОСТИ, ОСНОВАННОЕ НА ФИЗИКО-ХИМИЧЕСКИХ СВОЙСТВАХ АМИНОКИСЛОТ

Мильчевский Ю.В., Никитин А.М., Лукшин С.А., Мильчевская В.Ю.<sup>1</sup>,  
Туманян В.Г.

Институт молекулярной биологии им. В.А. Энгельгардта РАН, Россия, 119991, Москва,  
Вавилова 32, milch@eimb.ru

<sup>1</sup>European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany  
Rastatter strasse 10, Heidelberg 69126, Germany

Последние из опубликованных методов предсказания вторичной структуры белка (Nature 2016) [1] достигли точности предсказания в 84% для трех классов и 72% для восьми классов в соответствии с разметкой DSSP. Несмотря на это, даже по правильно предсказанной вторичной структуре фрагмента восстановить его декартовы координаты невозможно.

Мы представляем новый подход предсказания локальной структуры белка, который позволяет однозначно восстанавливать декартовы координаты предсказанного фрагмента. Для этого мы определили аналог базиса в пространстве локальных структур (далее наз. “обобщённые функции координат”), и обосновали минимальное число элементов, необходимое для представления фрагментов белковой цепи различной длины при различных предположениях о фиксированных/нефиксированных двугранных и/или валентных углах, а, также длинах связей. Далее описан метод предсказания локальной структуры фрагментов из пяти остатков. Путем кластеризации выборки белковых структур из PDB банка мы получили набор локальных структур пента-пептидов, включающий 30 кластеров, т.е. 30 базисных структур. Полученные обобщенные координаты позволяют свести многомерную задачу предсказания пространственной структуры по последовательности к ряду одномерных задач. Так же, как в декартовых координатах положение точки можно задать через расстояния до базисных векторов, в представленных обобщенных координатах структура фрагмента определяется расстояниями до базисных структур. Данные расстояния предсказываются на основе физико-химических свойств последовательности белкового фрагмента [2].

Создан сервер для предсказания локальной структуры белка по последовательности белка: milch.eimb.ru, не имеющий аналогов. Качество предсказания (попадания в свой кластер) для набора из 30 кластеров составляет в среднем 72%

## Литература.

1. Wang, S. et al. Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. Sci. Rep. 6, 18962; doi: 10.1038/srep18962 (2016).
2. <http://www.genome.jp/aaindex/>