

УТОЧНЕНИЕ СТРУКТУРЫ НЕУПОРЯДОЧЕННОГО КОМПОНЕНТА РАСТВОРА БЕЛКОВЫХ КРИСТАЛЛОВ С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

Мустафин Х.С., Гуцин И.Ю

Московский физико-технический институт,
Россия, 141700, г. Долгопрудный, Институтский пер., 9.,
E-mail: khalid.mustafin@phystech.edu

Количество разрешенных структур биологических молекул в PDB в настоящее время превышает 190 000; большинство из них определяется методами кристаллографии. Многие компьютерные программы упростили и ускорили процесс определения структуры [1]. Однако некоторые задачи требуют осторожного подхода в интерпретации результатов кристаллографии.

Ярким примером является задача моделирования компонент раствора белковых кристаллов. Среда, в которой выращивается кристалл белка, неизбежно содержит воду, ионы, нативные лиганды и т. д. В среднем около половины всего объема кристалла состоит из неупорядоченного раствора [2]. Неточная интерпретация электронной плотности компонент растворителя, особенно вблизи поверхности белка, может привести к неверным выводам о его физико-химических свойствах. Даже небольшие компоненты, такие как ионы хлора, могут играть важную роль в функционировании белка и быть неотъемлемой частью его структуры. В то же время ионы хлора имеют низкий аномальный сигнал и их легко спутать с молекулами воды [3].

В данной работе мы исследовали возможность классификации молекул воды и ионов хлора в кристаллографических структурах белков с использованием моделей машинного обучения. Мы представляем автоматизированную схему для анализа и классификации ионов хлора и молекул воды. Также нами были обнаружены кристаллографические структуры, в которых некорректно моделируются ионы хлора.

Литература.

1. Wlodawer A. et al. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination // FEBS J. 2013. Vol. 280, № 22. P. 5705–5736.
2. Weichenberger C.X. et al. The solvent component of macromolecular crystals // Acta Crystallogr. Sect. D Biol. Crystallogr. 2015. Vol. 71, № 5. P. 1023–1038.
3. Skitchenko R.K. et al. Census of halide-binding sites in protein structures // Bioinformatics / ed. Elofsson A. 2020. Vol. 36, № 10. P. 3064–3071.