CLARIFYING THE STRUCTURE OF THE SOLVENT COMPONENT OF PROTEIN CRYSTALS USING MACHINE LEARNING MODELS

K. Mustafin, I. Gushchin

Moscow Institute of Physics and Technology
9 Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russian Federation
E-mail: khalid.mustafin@phystech.edu

The number of resolved structures of biological molecules in the PDB currently exceeds 190,000; most of them are determined by crystallographic methods. Many computer programs have made the process of structure determination easier and faster [1]. However, some tasks require a careful approach to interpreting the results of crystallography.

A prime example is the problem of modeling the solvent component of protein crystals. The environment in which a protein crystal is grown inevitably contains water, ions, native ligands, etc. On average, about half of the entire crystal volume consists of disordered solvent [2]. Inaccurate interpretation of the electron densities of the solvent, especially near the surface of the protein, can lead to incorrect conclusions about its physical and chemical properties. Even small solvent components such as chloride ions can play an important role in the functioning of a protein and be an integral part of its structure. At the same time, chloride ions have a low anomalous signal and can easily be confused with water molecules [3].

In this work, we investigated the possibility of classifying water molecules and chloride ions in the crystallographic structures of proteins using machine-learning models. We present a complete end-to-end scheme for the analysis and classification of chloride ions and water molecules. We also found crystallographic structures, in which chloride atoms are modeled incorrectly.

References.
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.