Секция «Биоинформатика»

Classification of families of DNA-recognizing protein domains based on structural features of DNA-protein complexes

Научный руководитель – Спирин Сергей Александрович

Панова Вера Викторовна
Student (specialist)
Московский государственный университет имени М.В.Ломоносова, Факультет биоинженерии и биоинформатики, Москва, Россия
E-mail: nooroka17@gmail.com

DNA-protein interactions play a central role, for example, in such cellular processes as DNA replication, transcription, and repair [2]. Prediction of the interactions between DNA and protein is a significant but not trivial task [1]. Currently, 6048 DNA-protein structures are available in open databases. A classification of DNA-protein makes it possible to effectively study the patterns of DNA-protein recognition.

This work aims to correct errors in the existing version of NPIDB ([3], http://npidb.belozersky.msu.ru/), to create a classification of the structures of DNA-protein complexes using the current corrected data from NPIDB, to describe structural features and to make an internal classification of the most popular families.

The classification is based on the principles from [5], but instead of domains allocated in protein chains according to the data of the SCOP databank whose support was discontinued in 2009, the domains allocated according to the Pfam databank [4] are classified. The classification is based on contacts between DNA and protein molecules. Hydrogen bonds, hydrophobic interactions, and water bridges are considered. The contact type is a pair of contacting structural elements (one element from the protein is a helix, beta-sheet, or turn/unstructured segment, one from DNA is the major groove, the minor groove, or the sugar-phosphate backbone). The interaction mode for the domain and domain structure is defined as a list of contact types, and the interaction class for a family of domains is defined as the intersection of domain interaction modes.

A set of Python programs determining interaction and interaction classes was written. Only protein structures in complexes with a double DNA helix (at least 6 complementary pairs), solved by X-ray analysis with a resolution < 3 Å were considered. The data were statistically processed, the number of structures and domains with each interaction mode and the number of families with each interaction class were obtained. Several particular families have been described.

I would like to thank Sergey Spirin for supervising and Eugene Baulin for his help in work.

References

1) Gao M, Skolnick J. From Nonspecific DNA–Protein Encounter Complexes to the Prediction of DNA–Protein Interactions // PLoS Comp. Biol. 2009; 5. doi: 10.1371/journal.pcbi.1000341

2) Kulandaisamy, A. et al. (2017). Dissecting and analyzing key residues in protein-DNA complexes // J. of Molecular Recognition, 31(4). doi: 10.1002/jmr.2692

3) Kirsanov D.D. et al. NPIDB: nucleic acid-protein interaction database // Nucleic Acid Research 2013, 41 (D1):D517–D523, doi: 10.1093/nar/gks1199

4) Mistry, J., et al. (2020). Pfam: The protein families database in 2021 // Nucleic Acids Research, 49(D1). doi: 10.1093/nar/gkaa913

5) Zanegina O.N. et al. An updated version of NPIDB includes new classifications of DNA-protein complexes and their families // Nucleic Acid Research 2016, 45 (D1):D144–D153, doi: 10.1093/nar/gkv1339