

**Классификация семейств ДНК-узнающих белковых доменов на основе структурных особенностей ДНК-белковых комплексов**

**Научный руководитель – Спирин Сергей Александрович**

***Панова Вера Викторовна***

*Аспирант*

Московский государственный университет имени М.В.Ломоносова, Факультет биоинженерии и биоинформатики, Москва, Россия

*E-mail: nooroka17@gmail.com*

На настоящий момент в открытых базах данных доступно 6557 ДНК-белковых структур. Разные белки взаимодействуют с двойной спиралью ДНК разными способами. Создание классификации ДНК-белковых взаимодействий и ее интеграция в базу ДНК-белковых комплексов позволяет эффективно исследовать закономерности ДНК-белкового узнавания.

Цель данной работы — исправить ошибки в существующей версии NPIDB ([1], <http://npidb.belozersky.msu.ru/>), а далее, используя исправленные данные из NPIDB, создать классификацию структур ДНК-белковых комплексов на современном материале, а также описать структурные особенности и внутреннюю классификацию наиболее популярных семейств.

Классификация основана на принципах из работ [1,2], но вместо доменов, выделяемых в цепях белка согласно данным банка SCOP (поддержка которого прекращена в 2009 г.), классифицируются домены, выделяемые согласно банку Pfam [3]. Классификация основана на контактах между молекулами ДНК и белка. Рассматриваются водородные связи, гидрофобные взаимодействия и водяные мостики. Тип контакта — это пара контактирующих элементов структуры. В белках мы выделяем три элемента: спираль, бета-лист, поворот или неструктурированный сегмент, в ДНК тоже три элемента: большая бороздка, малая бороздка, сахарофосфатный остов). Способ взаимодействия домена белка определяется как список типов контактов структур этого домена с ДНК (структурой домена мы называем конкретную часть записи PDB с расшифровкой её структуры, т.е. техническую реплику, а доменом — участок белка). Класс взаимодействия для семейства доменов определяется как пересечение способов взаимодействия доменов [2].

Был написан комплекс программ на Python, который определяет способы взаимодействия для доменов и структур доменов, а также классы взаимодействия для семейств. Рассматривались только структуры белка с двойной спиралью ДНК (не менее шести комплементарных пар), решённые методом рентгеноструктурного анализа или криоэлектронной микроскопии, с разрешением менее 3Å. Всего на декабрь 2021 (последнее обновление NPIDB) доступно 4089 структур ДНК-белковых комплексов, удовлетворяющих нашим условиям. В белках этих комплексов выявлено 8525 контактирующих с ДНК структур 856 доменов. Семействам, представленным не менее чем тремя различными доменами с доступными структурами в комплексах с ДНК, приписаны классы взаимодействия, всего таких семейств 100. Всего на данный момент реализуется 37 классов взаимодействия (из 512 теоретически возможных). Самые популярные — (i) спираль с большой бороздкой и петля с большой бороздкой, (ii) спираль с большой бороздкой, (iii) петля с большой бороздкой (по 9 семейств в каждом из этих классов). Описано несколько конкретных семейств, в т.ч. числе шесть семейств, называемых “разнородными”, это семейства, у которых нулевой класс взаимодействия. Показана внутренняя классификация “разнородных” семейств, определены причины, по которым данные семейства являются “разнородными”.

### Источники и литература

- 1) Kirsanov D.D. et al. NPIDB: nucleic acid-protein interaction database. Nucleic Acid Research 2013, 41 (D1):D517–D523, doi: 10.1093/nar/gks1199
- 2) Zanagina 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.
- 3) Mistry, J., et al. (2020). Pfam: The protein families database in 2021. Nucleic Acids Research, 49(D1). doi:10.1093/nar/gkaa913