

Предсказание кластеров биосинтеза нерибосомальных пептидов

Научный руководитель – Конанов Дмитрий Николаевич

Кривонос Данил Вадимович

Студент (магистр)

Российский химико-технологический университет имени Д.И. Менделеева, Факультет биотехнологии и промышленной экологии (БПЭ), Москва, Россия

E-mail: danil01060106@gmail.com

Нерибосомальные пептиды (NRP) - группа структурно разнообразных вторичных метаболитов, синтезируемых крупным ферментом NRP-синтетазой (NRPS)[Oide, 2020, P. 317]. Любой модуль NRPS обязательно включает в себя 3 домена: домен аденилирования (А-домен), домен конденсации (С-домен) и домен переносчика (РСР-домен). Считается, что последовательность мономеров в конечном продукте зависит в большей степени от А-доменов, который способен осуществлять захват аминокислот из окружающей среды[Bloudoff, 2017, P. 1587].

Наша задача состояла в создании инструмента, позволяющего по химической структуре соединения определять его возможных продуцентов. Современные методы предсказания субстратов отдельных модулей основаны на специфичности А-доменов. Однако, описанные в литературе механизмы конденсации позволяют предположить субстратную специфичность и для С-доменов.

На основе MiBIG нами была собрана база данных, включающая более 2000 последовательностей для С- и А- доменов из модулей с известным субстратом. Для каждого субстрата независимо были обучены скрытые марковские модели (НММ). Для каждой модели были получены распределения *e-value* для последовательностей с заведомо несоответствующим субстратом, чтобы иметь возможность оценивать достоверность схожести (попадания) с оглядкой на негативный контроль.

Модульность NRPS позволяет оформить данные в формат матрицы с позиционно-специфичными значениями (position-specific score matrix или PSSM), где каждая строка соответствует субстрату, а каждый столбец модулю. Такие матрицы достаточно давно используются в биоинформатике и могут быть использованы как шаблон для выравнивания конечного продукта на предполагаемый кластер синтеза.

Метрика выравнивания была определена как простая сумма по соответствующим позициям матрицы. Для оценки достоверности каждого конкретного выравнивания применялся перестановочный тест.

Для тестирования метода был выбран Leave-One-Object подход. AUC-ROC составила 0.82, что позволяет говорить о том что, во-первых, субстратная специфичность у С-доменов есть, во-вторых, что ее достаточно чтобы с хорошей точностью искать кластеры синтеза заданных соединений. Для более специфичных А-доменов AUC-ROC составила 0.99, что говорит о высокой предсказательной силе модели.

Источники и литература

- 1) Oide S., Turgeon B.G. Natural roles of nonribosomal peptide metabolites in fungi//Mycoscience, 2020, Vol. 61, No, P. 101-110.
- 2) Bloudoff K., Martin T. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity//Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2017, Vol. 1865, No.11, P. 1587-1604.