

Анализ связи структура-активность для ингибиторов белкового синтеза на основе машинного обучения**Academic supervisor – Столбов Леонид Алексеевич****Зуев Давид Александрович***Student (specialist)*

Lomonosov Moscow State University, Факультет биоинженерии и биоинформатики,
Moscow, Россия
E-mail: d.gll@bk.ru

Биосинтез белка (трансляция) представляет собой многостадийный процесс, нарушение которого является основой действия многих противоопухолевых и антипаразитарных препаратов. Актуальной задачей является разработка подходов, позволяющих по химической структуре соединения прогнозировать его потенциальную активность до этапа экспериментального тестирования. Целью настоящей работы является создание моделей на основе машинного обучения, способных прогнозировать ингибирующую активность соединений в отношении конкретных стадий трансляции и молекулярных мишеней.

База данных EuPSIC [2] содержит сведения об ингибиторах эукариотической трансляции. В ходе исследования проведена стандартизация структур для 485 соединений из EuPSIC. Поскольку в базе представлены только активные соединения, для бинарной классификации отрицательные примеры были получены методом взаимного противопоставления, где ингибиторы одного процесса считались неактивными для другого. Химическое пространство было охарактеризовано дескрипторами QNA [3], на основе которых с применением метода самосогласованной экстремальной классификации (SCEC) [1] построены классификационные модели для прогноза ингибирования различных стадий трансляции (инициация, элонгация, терминация), групп мишеней (рибосомные сайты, факторы трансляции, сигнальные пути), а также для индивидуальных мишеней (фактор инициации eIF4A, A- и E-сайты рибосомы). SCEC предназначен для работы с гетерогенными и несбалансированными данными: метод автоматически балансирует вклад конкретных химических структур за счёт адаптивных весов и отбирает наиболее значимые структурные особенности, что критически важно при несбалансированных классах или при использовании структур из других процессов в качестве неактивных.

Результаты пятикратной кросс-валидации показывают высокую прогностическую способность моделей, наибольшие значения AUC достигнуты для моделей прогноза ингибирования стадий инициации (0,93) и элонгации (0,90), а также для групп «Факторы трансляции» (0,91) и «Рибосомные сайты» (0,90). Значения AUC для моделей индивидуальных мишеней составили от 0,87 до 0,98. Построенные модели позволяют осуществлять поиск соединений с заданным механизмом ингибирования трансляции и отбирать среди них кандидатов для создания новых лекарственных препаратов.

Источники и литература

- 1) Столбов Леонид Алексеевич. Разработка подходов к виртуальному скринингу антивирусных соединений с учетом гетерогенности информации: автореферат дис. ... кандидата биологических наук: 1.5.8. ФГБНУ «Научно-исследовательский институт биомедицинской химии имени В. Н. Ореховича», 2023.
- 2) Dmitriev S. E., Vladimirov D. O., Lashkevich K. A. A Quick Guide to Small-Molecule Inhibitors of Eukaryotic Protein Synthesis // Biochemistry (Moscow). 2020. Vol. 85, No. 11. P. 1389–1421.

- 3) Filimonov D. A., Zakharov A. V., Lagunin A. A., Poroikov V. V. QNA-Based Track“ QSAR Approach // SAR and QSAR in Environmental Research. 2009. Vol. 20, No. 7–8. P. 679–709.