

Прогноз противовирусной активности низкомолекулярных соединений на основе протеохеометрики

Карасев Д.А.*, Соболев Б.Н., Филимонов Д.А., Лагунин А.А., Тарасова О.А.,
Поройков В.

Институт биомедицинской химии им. В.Н. Ореховича, Москва, Россия

* d.karasev@ibmc.msk.ru

Ключевые слова: протеохеометрика; низкомолекулярные лекарственноподобные соединения; хемоинформатика; анализ взаимосвязи «структура-активность»; (Q)SAR; противовирусные препараты; вирусные инфекции

Мотивация и цель: Оценка взаимосвязи «структура–активность» ((Q)SAR) на основе представления молекул в виде химических структур широко используется при компьютерном поиске новых лекарственных соединений. Методы протеохеометрики (PCM) позволяют расширить область применимости моделей взаимосвязи «структура-активность», поскольку при построении таких моделей также используются данные о структуре белков-мишеней. Таким образом, становится возможным осуществлять прогноз для белков с неизвестным спектром лигандов. Поскольку белки-мишени, кодируемые генами различных вирусов, могут проявлять значимое структурное сходство [1], использование PCM представляется весьма перспективным при поиске действенных лекарств при новых вирусных угрозах. Цель настоящей работы – разработка подхода PCM и его тестирование при прогнозе активности низкомолекулярных соединений в отношении протеаз коронавируса, энтеровирусов, флавивирусов и вирусов гепатита С.

Методы и алгоритмы: Мы разработали оригинальный метод PCM, основанный на комбинированном описании пар белок-лиганд. Метод был протестирован на данных, представляющих вирусные протеазы из надсемейства химотрипсина и их ингибиторы. Данные о соответствующих взаимодействиях были извлечены с помощью оригинальной процедуры из базы данных ChEMBL. Множественное выравнивание белков было построено на основе попарных трехмерных выравниваний. Это позволило рассчитать дескрипторы белков с учетом областей связывания лигандов [2]. Для описания лигандов были применены дескрипторы атомных окрестностей (MNA-дескрипторы) [3].

Результаты: Тестирование набора лигандов в режиме (Q)SAR показало точность прогноза около 0.95 при валидации методом исключения по одному. Модели PCM, построенные с использованием комбинированных дескрипторов, были тестированы с использованием процедуры, в которой реализовано исключение пар белок-лиганд. Таким образом, было смоделировано предсказание взаимодействия новых лигандов и новых мишеней. Полученные значения точности при этом достигли 0.89.

Выводы: Таким образом, подход PCM показывает некоторое снижение точности по сравнению с (Q)SAR, но обеспечивает эффективное предсказание при отсутствии надежных данных о взаимодействии низкомолекулярных соединений с белками-мишенями соответствующих вирусов.

Финансирование: Исследование выполнено при поддержке программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 гг.) (№ 124050800018-9).

Prediction of antiviral activity of druglike compounds based on proteochemometrics

Karasev D.A.*, Sobolev B.N., Filimonov D.A., Lagunin A.A., Tarasova O.A., Poroikov V.V.

Institute of Biomedical Chemistry, Moscow, Russia

* w.dmitrykarasev@gmail.com

Key words: proteochemometrics; small molecule drug-like compounds; chemoinformatics; structure-activity relationship analysis; (Q)SAR; antiviral drugs; viral infections

Motivation and Aim: An analysis of structure-activity relationships ((Q)SAR) based on the chemical structures of molecules is widely used in the computer search for new drug compounds. Proteochemometrics (PCM) methods expand the applicability area of structure-activity relationship models, as such models can use data on the structures of both protein targets and ligands, which is beneficial for predicting activity spectra for proteins with an unknown spectrum of ligands. Since target proteins encoded by genes of different viruses can exhibit significant structural similarity [1], the use of PCM appears is a promising in the search for effective drugs against new viral threats. The purpose of this work is to develop the PCM approach and test it for predicting the activity of small molecular compounds against proteases of coronaviruses, enteroviruses, flaviviruses and hepatitis C viruses.

Methods and Algorithms: We developed an original PCM method based on the combined description of protein-ligand pairs. The method was tested on data representing viral proteases from the chymotrypsin superfamily and their inhibitors. Data on relevant interactions were extracted using an original procedure from the ChEMBL database. The multiple protein alignment was built based on the pairwise 3D alignments. It enabled the calculation of protein descriptors, taking into account the regions of ligand binding [2]. Atomic neighborhood descriptors (MNA descriptors) were used to describe ligands [3]. **Results:** Testing a set of ligands in (Q)SAR mode with leave-one-out cross validation was characterized by a prediction accuracy of about 0.95. PCM models built using the combination descriptors were tested using a procedure involving the exclusion of protein-ligand pairs. Thus, the prediction of interactions between new ligands and new targets was modeled. The accuracy values obtained in PCM mode reached 0.89.

Conclusion: Thus, the PCM approach shows insignificant decrease in accuracy compared to (Q)SAR, but provides effective prediction in the absence of reliable data on the interaction of small molecules with the target proteins of the corresponding viruses.

Funding: The study was performed in the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030) (No. 124050800018-9).

Список литературы/References

1. Montinen H.A.M., Ravaniti J.J., Poranen M.M. Structural comparison strengthens the higher-order classification of proteases related to chymotrypsin. *PLoS One*. 2019;14(5):e0216659
2. Karasev D.A., Sobolev B.N., Filimonov D.A., Lagunin A.A. Prediction of viral protease inhibitors using proteochemometrics approach. *Comput Biol Chem*. 2024;110:108061
3. Lagunin A., Stepanchikova A., Filimonov D., Poroikov V. PASS: prediction of activity spectra for biologically active substances. *Bioinformatics*. 2000;16(8):747-748