

НИКИТИНА А.А., ОРЛОВ А.А., ОСОЛОДКИН Д.И., ПАЛЮЛИН В.А., ЗЕФИРОВ Н.С.

МГУ им. М.В. Ломоносова, ИПВЭ им. М.П. Чумакова, Москва, Россия

СОЗДАНИЕ ВЫСОКОКАЧЕСТВЕННО АННОТИРОВАННОЙ БИБЛИОТЕКИ СОЕДИНЕНИЙ, ТЕСТИРОВАННЫХ НА ПРОТИВОВИРУСНУЮ АКТИВНОСТЬ

Цель: получение наиболее полной и качественно аннотированной базы данных соединений, тестируемых на противовирусную активность, на основе общедоступной базы данных биоактивностей ChEMBL_20.

Материалы и методы: в качестве источника данных была использована SQL-версия базы данных ChEMBL_20. Поиск и анализ информации об организме мишени действия соединений осуществлялся на основе значений полей `assays.assay_organism` и `target_dictionary.organism` с применением специально составленных списков допустимых значений. Поиск данных, имеющих некорректную аннотацию в полях `assays.assay_organism` и `target_dictionary.organism`, производился по полю `assays.description` с использованием словаря подстрок, состоящего из систематических наименований и сокращённых названий вирусов.

Результаты: проведен анализ данных для соединений, тестируемых на противовирусную активность, на основе информации, содержащейся в базе данных ChEMBL_20. Оценено количество данных, имеющих полные и корректные аннотации по мишени действия, а также записей, содержащих информацию о противовирусной активности, но аннотированных некорректно. Проведена дополнительная аннотация активностей по организму-мишени с использованием данных, извлечённых из полей `assays.assay_organism`, `target_dictionary.organism` и `assays.description`. Показано, что полученная база содержит большее количество соединений и активностей, относящихся к вирусам, по сравнению с данными, полученными с использованием простого поиска по таксономии через web-интерфейс ChEMBL_20.

Выводы: составленная база данных содержит информацию более высокого качества по сравнению с сырыми данными, содержащимися в ChEMBL_20, и будет использована для исследования химического пространства противовирусных соединений.