

Протокол для поиска возможных неспецифичных взаимодействий лигандов

Максимов Антон, 527и группа

Научный руководитель
Попов Пётр Анатольевич, PhD

Московский физико-технический институт
(национальный исследовательский университет)

25 июня 2019 г., Долгопрудный

План

- 1 Введение и терминология
- 2 Цель проекта
- 3 Входные данные и ожидаемые результаты
- 4 Структура протокола
- 5 Принципы работы протокола
- 6 Результаты
- 7 Заключение

Общее введение, терминология

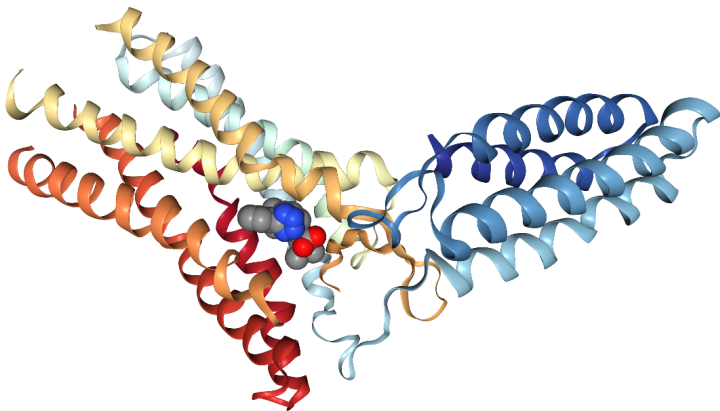


Схема взаимодействия лиганда-лекарства с белком-мишенью на примере структуры GPCR-рецептора AT1R с обратным агонистом олмесартаном из базы PDB.

Общая картина

- Парадигма «болезнь-мишень-лекарство» уступает полифармакологии и лекарственной репозиции.
- *Полифармакология* — одно вещество действует сразу на несколько мишеней, весь метаболический путь.
 - + меньше учет взаимодействий лекарств друг с другом;
 - + просто меньше лекарств → предсказуемость;
 - надо создать такое лекарство.
- *Лекарственная репозиция* – использование уже известных лекарств в новых целях.
 - + уже прошли тесты на людях ⇒ безопасно, экономически выгодно.Примеры: виагра, некоторые лекарства от рака.

Цель проекта

Целью работы являлось построение гибкого программного протокола, способного определять степени подобия лигандов/мишеней/комплексов с референсными данными, основанными на подтвержденных FDA лекарствах и протеомах.

Шаги для достижения цели

- извлечь информацию о подтвержденных FDA лекарствах, лигандах, мишенях, а также референсных протеомах;
- создать функции для конвертации данных одной молекулы в требуемые сторонними программами форматы;
- создать функции для сравнения одного элемента входных данных с одной подтвержденной FDA записью;
- реализовать поиск по референсным данным наиболее подходящих в терминах различных метрик подобия для лигандов/мишеней/комплексов.

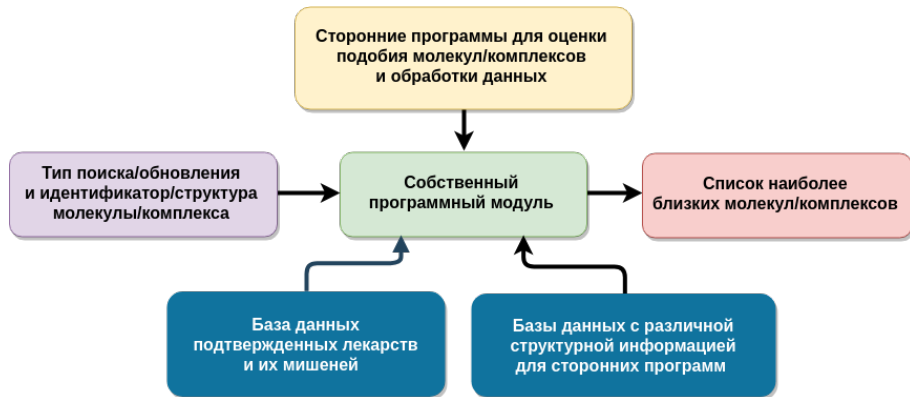
Входные данные и ожидаемые результаты

На вход протокола могут подаваться:

- 1 SMILES лиганда или его структура в файле SDF, MOL2, PDB;
- 2 аминокислотная последовательность мишени, Uniprot ID или PDB структура;
- 3 PDB структура молекулярного комплекса.

Результат — отсортированный по схожести ко входным данным список объектов того же типа с наиболее полной информацией о них и о различиях между входом и выходом.

Общая структура протокола



Принципы работы протокола

Используются различные структурные метрики подобия лигандов/мишеней (последовательность, 2D-, 3D-структуры)

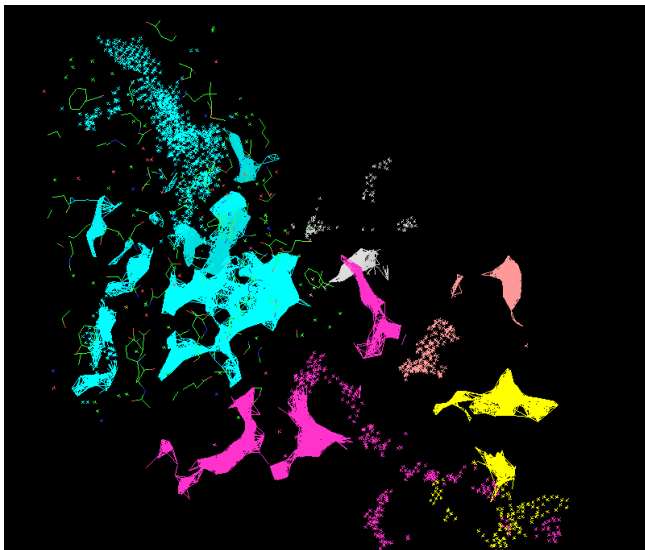
Сайты связывания сравниваются с использованием атомистической структуры полостей в комплексах и распределения физико-химических свойств.

Похожие на подтвержденные молекулы будут иметь повышенную вероятность образовать связь с теми же лигандами/мишенями и быть безвредными для человека.

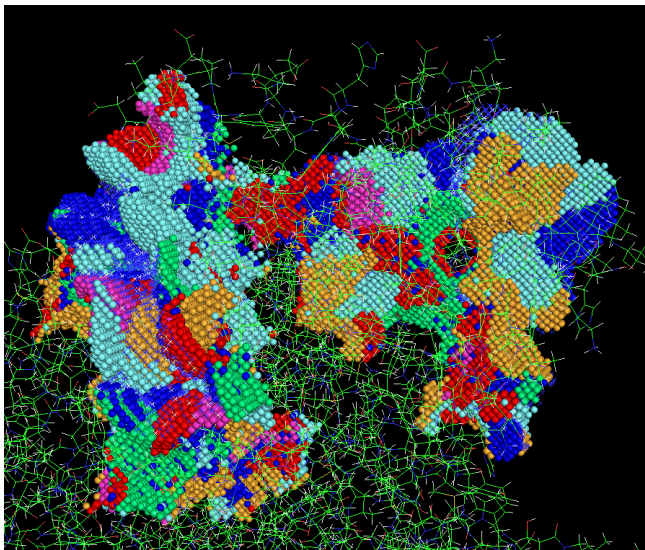
Принципы сравнений

- По аминокислотным последовательностям мишеней — с помощью выравнивания алгоритмом BLAST.
- По структурам мишеней — выравнивание динамическим программированием с последующим вычислением нормированного отклонения структур.
- По структурам лигандов — используя молекулярные отпечатки топологического (по SDF) и текстового (по SMILES).
- По структурам комплексов — используя графовый поиск на размеченных по физико-химическим свойствам полостях (IsoMIF) и по сравнению геометрий топологий структур (TM-align).

Сравнение комплексов: нахождение полостей



Сравнение комплексов: разметка полостей



Результаты

Результаты Сравнение аминокислотных последовательностей мишеней.

```
In :
get_closest_fastas_from_uniprot('P08100', path_to_data_in_fasta, k=0,
align_matrix='blosum62', sim_or_ident=True)
Out:

similarity      identity      name
1392      1843   348   lcl|BSEQ0016346|Rhodopsin
1931      318   156   lcl|BSEQ0010278|Cholecystokinin_receptor_type_A
151      295   147   lcl|BSEQ0016698|Somatostatin_receptor_type_5
...
Name = lcl|BSEQ0010278|Cholecystokinin_receptor_type_A
Similarity=318.5, identity=156
Matrix blosum62, number of alignments = 1
MNGTEGPNFYVPFSNATGVVRSPFEYPQYY-LAEP-----WQFS-----

MLAAYMFLILVLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVLGGFTSTLYTSLHGYFVFGPTGCNLEGFF

WSLVVLAIERYYVVCKPM-SNFFGENHAIMGVAFTWVMALACAA-PLAGWSRYIP-
EGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMIIFFCYGQLVFTV-----KEAAAQQQESATT
-----QK-----AEKEVTRMVIIMVIAFLICWVPYA

.....

PISFILLLSYTSSCV-
NPIIYCFMNKRFRRLGFMATFPCCPNPGPPGARGEVGEEEGGTTGASLSRFSYSHMSASVPPQ
Score=318.5
...
CPU times: user 10min 41s, sys: 1.26 s, total: 10min 42s
Wall time: 10min 43s
```

Заключение