

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 Заключение

Общее введение, терминология

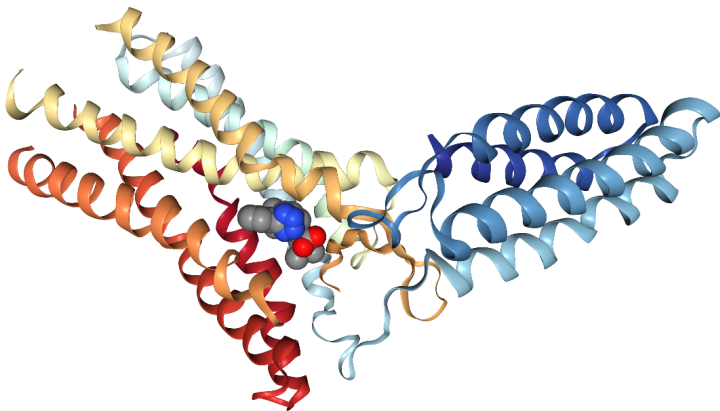


Схема взаимодействия лиганда-лекарства с белком-мишенью на примере структуры GPCR-рецептора AT1R с обратным агонистом олмесартаном из базы PDB.

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.

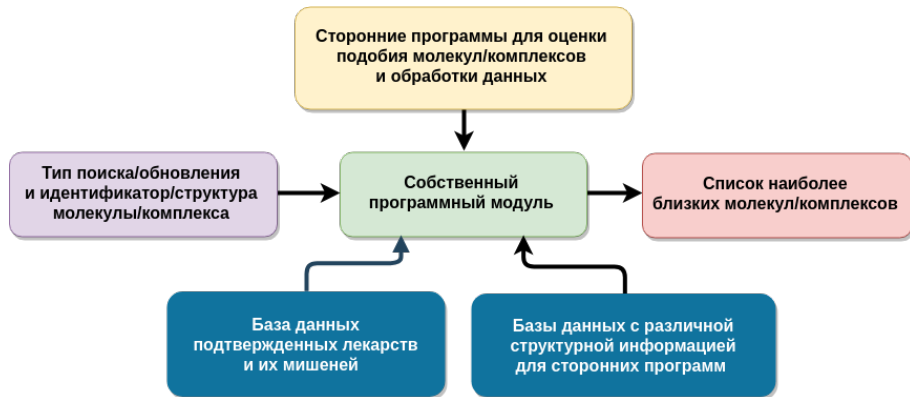
Входные данные и ожидаемые результаты

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)

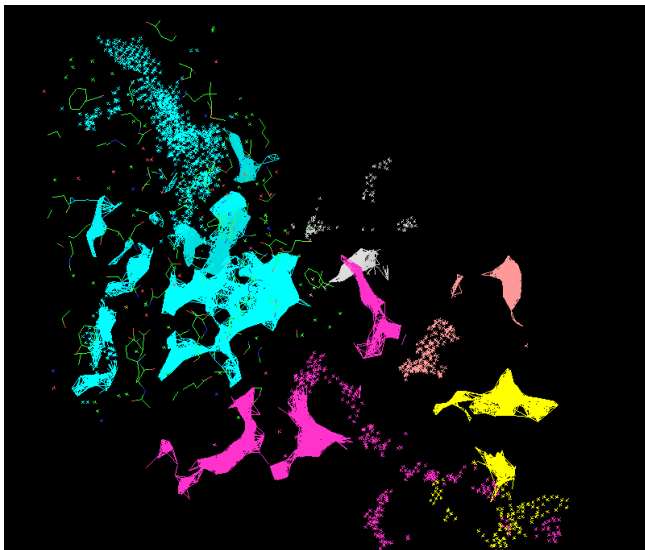
Сайты связывания сравниваются с использованием атомистической структуры полостей в комплексах и распределения физико-химических свойств.

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

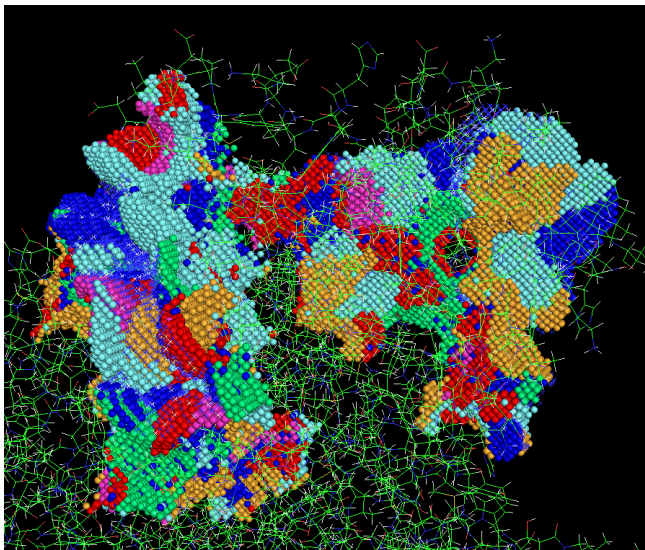
Принципы сравнений

- 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

Сравнение аминокислотных последовательностей мишеней.

```
1 In :
2 get_closest_fastas_from_uniprot('P08100', path_to_data_in_fasta, k=0,
3 align_matrix='blosum62', sim_or_ident=True)
4 Out:
5
6 similarity      identity      name
7 1392      1843    348    |cl|BSEQ0016346|Rhodopsin
8 1931      318    156    |cl|BSEQ0010278|Cholecystokinin_receptor_type_A
9 151       295    147    |cl|BSEQ0016698|Somatostatin_receptor_type_5
10 ...
11 Name = |cl|BSEQ0010278|Cholecystokinin_receptor_type_A
12 Similarity=318.5, identity=156
13 Matrix blosum62, number of alignments = 1
14 MNGTEGPNFYVPFSNATGVVRSPFEYPQYY-LAEP-----WQFS-----
15
16 MLAA YMFL LVLV LGFP INFL TLYV TVQH KKLRTPL NYILLNL AVADLF MVLG GFTSTL YTSL HG YFVFGPTGCNLEGFF
17
18 WSLV VLAIER YVVV CKPM-SNFF GENHAIMG VFTW VMAL ACAAP-PLAG WSR YIP-
19 EGLQC SCGIDYY TLKPEV NNFV IYMFV VHF TIPMI IFFCY GQLVFTV-----KEAAA QQQESATT
20 -----QK-----AEKEV TRMVIIMVIA FLICWVPYA
21
22 .....
23
24 PISFILLLSYTSSCV-
25 NPIIYCFM NKRFR LGMATFP CCPNPG PPGARGE VGE EEEEGGTT GASLSRFSY SHMSASVPPQ
26
27 Score=318.5
28 ...
29 CPU times: user 10min 41s, sys: 1.26 s, total: 10min 42s
30 Wall time: 10min 43s
```

Результаты

Сравнение молекулярных отпечатков лигандов.

По SMILES

```
1 In:
2 get_closest_ligands_from_3d_structure(path_to_structure, path_to_sdf_approved,
   root,
3 fptype='maccs', number_to_print=5)
4 Out:
5 name
6 481          Losartan
7 132          Olmesartan
8 923          Fomepizole
9 1110         Tromethamine
10 947          Gemfibrozil
11
12 smiles          similarity \
13 481 CCCCC1=NC(Cl)=C(CO)N1CC1=CC=C(C=C1)C1=CC=CC=C1... 1.000000
14 132 CCCCC1=NC(=C(N1CC1=CC=C(C=C1)C1=C(C=CC=C1)C1=NN... 0.677469
15 923                      CC1=CCN=C1 0.531250
16 1110                      NC(CO)(CO)CO 0.531250
17 947                      CC1=CC(OCCCC(C)(C)C(O)=O)=C(C)C=C1 0.522267
18
19
20
21
22
23
24
25 CPU times: user 1.87 s, sys: 7.68 ms, total: 1.88 s
26 Wall time: 1.88 s
```

По структурному файлу

```

1 In:
2 smiles = get_smiles_from_name_from_pubchem(name_lig, root)
3 df = get_closest_smiles_names(smiles, root, 5)
4 Out:
5 Name                Tanimoto coeff          Drugbank ID      Fingerprint type
6 0      Losartan                1.000000          DB00678          maccs
7 1    Olmesartan                0.735294          DB00275          maccs
8 2      Valsartan                0.690141          DB00177          maccs
9 3    Irbesartan                0.681159          DB01029          maccs
10 4   Crizotinib                0.657143          DB08865          maccs
11 CPU times: user 2.66 s, sys: 7.99 ms, total: 2.67 s
12 Wall time: 2.66 s

```

Заключение

Построен гибкий протокол, использующий различные методы вычисления сходства лигандов/мишеней/комплексов, и его работа проверена.

В будущем возможно добавлять новые способы, реализовать параллельные версии программ и учитывать известные побочные эффекты лекарств. Предлагается также использовать методы машинного обучения на коэффициентах сходства, что позволит искать неспецифичные взаимодействия более эффективно.