Структурная Биоинформатика

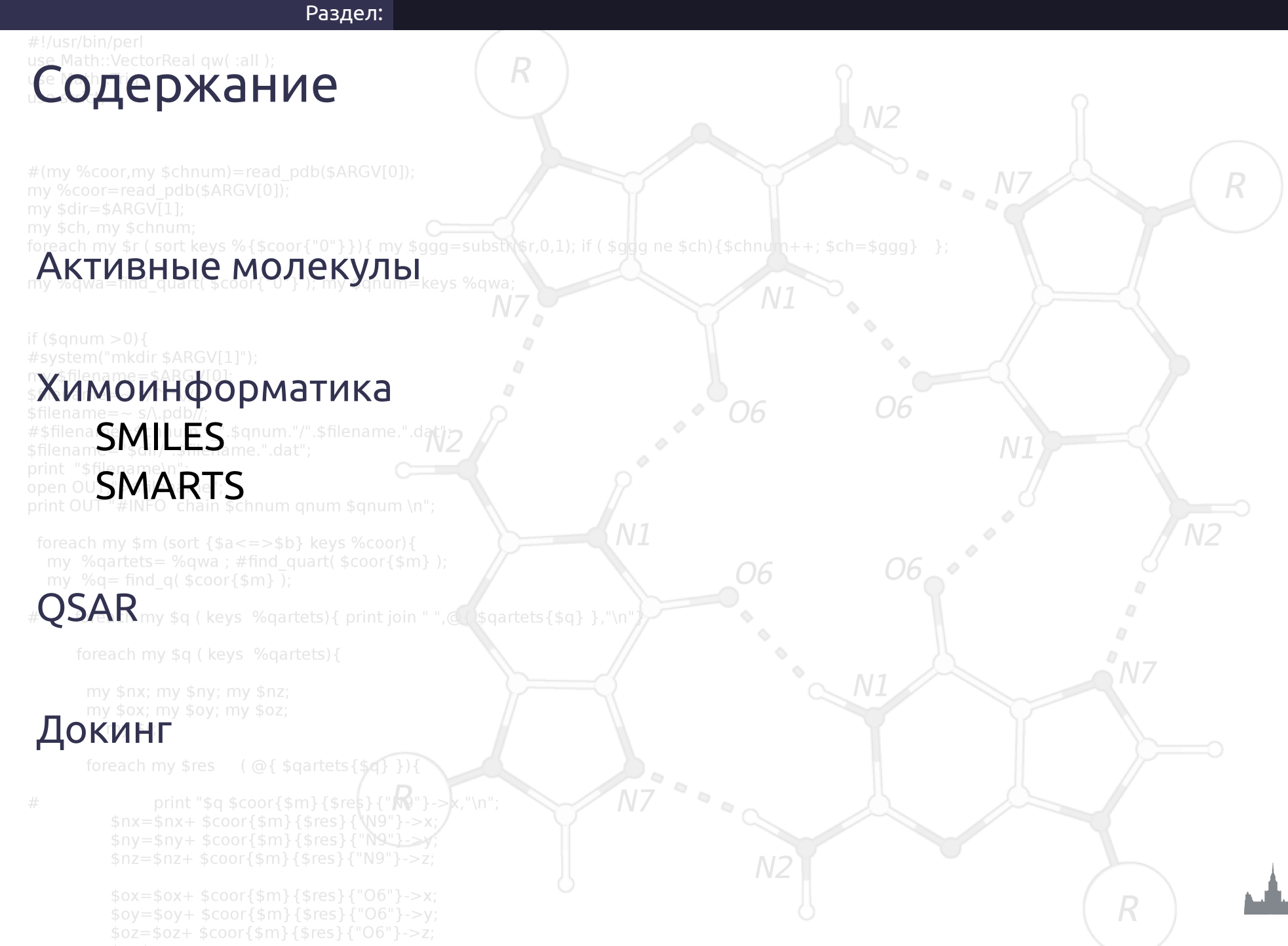
Лекция 6. Поиск новых биоактивных молекул и химоинформатика.

Головин А.В. [1]

[1] МГУ им М.В. Ломоносова, Факультет Биоинженерии и Биоинформатики

Москва, 2012

Содержание



Активные молекулы

- В основном биологически активные молекулы взаимодействуют нековалентно с биополимерами
- Агонисты связываются как нативные лиганды и дают тот же эффект
- Антагонисты конкурируют или препятствуют связыванию нативного лиганда
- Обратные агонисты связываются и оказывают эффект, обратный эффекту нативного лиганда
- Хорошие молекулы показывают высокую комплементарность поверхности биополимера

Свойства лекарства

- Лекарством обычно являются не только те молекулы, которые хорошо связываются с биополимером.
- Лекарство должно иметь приемлемую растворимость
- Часто бывает, что лекарству надо проникнуть сквозь мембрану.
- Хорошо когда лекарство в итоге метаболизируется, а не накапливается в тканях.

Как искать активные молекулы?

- Можно пытаться искать вещества в биоматериалах.
- Можно проводить роботизированное сканирование библиотеки соединений на активность в разных тестах.
- Недостаток сканирования: не все тесты можно адаптировать под робота.
- Возможен высокий уровень шума из-за неспецифических взаимодействий
- Можно применить фильтрацию по подобию соединений, для этого нужны ИТ.

Компьютерное представление молекул

- Хранение в компьютере молекулы как изображения имеет малую ценность
- Большинство современных баз данных представляет молекулу как граф, с узлами и рёбрами
- Графы представляются как таблицы связей.

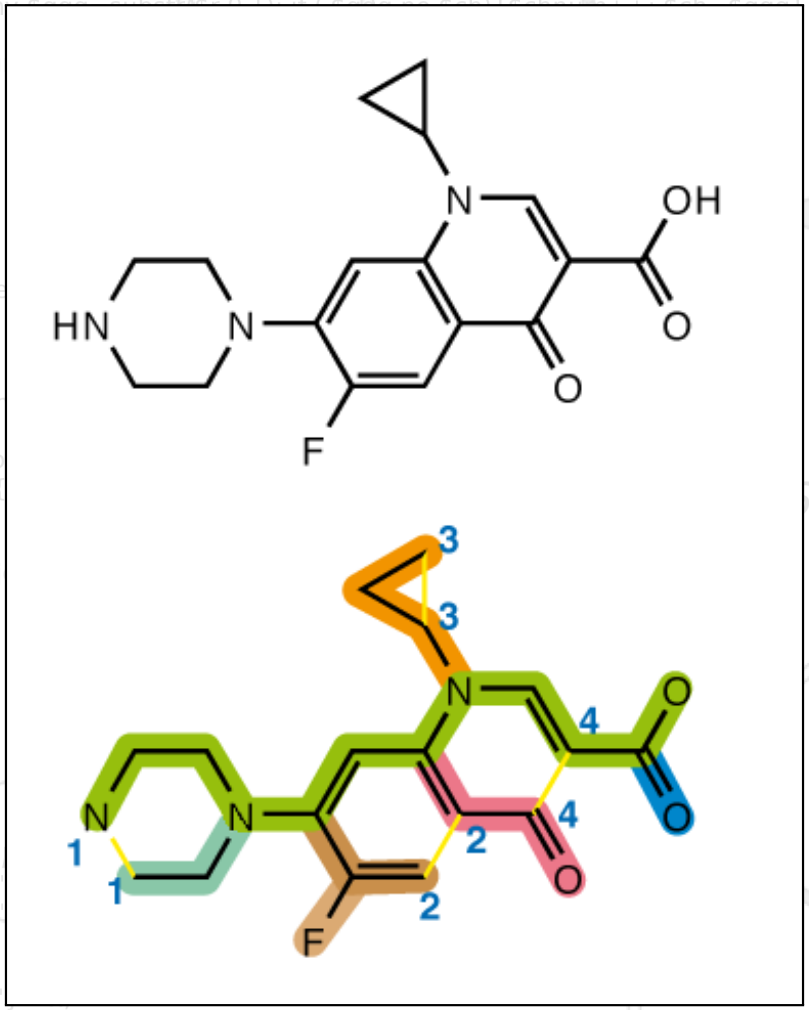
```
 Marvin  04200617372D

   4  3  0  0  0  0               999 V2000
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.7145   -0.4125    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -0.7145   -0.4125    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.8250    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
  1  4  2  0  0  0  0
  2  1  1  0  0  0  0
  3  1  1  0  0  0  0
 M  END
```

Содержание

Линейное представление молекул, SMILES

Молекула представляется в виде диаграммы и каждый атом проходится только один раз



Линейное представление молекул, SMILES

Примеры:

CC ethane	[OH3+] hydronium ion
O=C=O carbon dioxide	[2H]]O[[2H] deuterium oxide
C#Nhydrogen cyanide	[235U] uranium-235
CCN(CC)CC triethylamine	F/C=C/F E-difluoroethene
CC(=O)Oacetic acid	F/C=C/F Z-difluoroethene
C1CCCCC1 cyclohexane	N[C@@H](C)C(=O)O L-alanine
c1ccccc1 benzene	N[C@H](C)C(=O)O D-alanine

Реакции в виде SMILES

```
[I-].[Na+].C=CCBr>>[Na+].[Br-].C=CCI
(C(=O)O).(OCC)>>(C(=O)OCC).(O)
```

Стандартизация SMILES

- Очевидно, что одну молекулу можно описать разными способами.
- Морган в 1965 году предложил рассматривать каждый атом по свойству его окружения.
- Стандартные SMILES называют Unique.

Input SMILES	Unique SMILES
OCC	CCO
[CH3][CH2][OH]	CCO
C-C-O	CCO
C(O)C	CCO
OC(=O)C(Br)(Cl)N	NC(Cl)(Br)C(=O)O
ClC(Br)(N)C(=O)O	NC(Cl)(Br)C(=O)O
O=C(O)C(N)(Br)Cl	NC(Cl)(Br)C(=O)O

Описание SMILES: атомы

- Однобуквенные атомы, а именно : B, C, N, O, P, S, F, Cl, Br, I записываются как есть, как один символ.
- Все остальные атомы записываются в квадратных скобках [Pt]
- Так как атомы водорода обычно не указываются, то “валентность” атомов определятся как наименьшая из ближайших Т.е. B (3), C (4), N (3,5), O (2), P (3,5), S (2,4,6).
- “Валентности”, отличные от “нормальных”, указывают в скобках [S], [H+], [Fe+2], [OH-], [Fe++], [OH3+], [NH4+]

Описание SMILES: связи

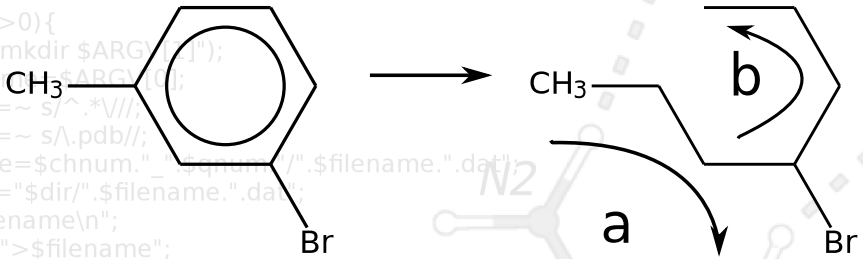
CC	этан
C=C	этилен
O=C=O	CO2
C#N	HCN
CCO	этанол
[H][H]	водород

Ветвление цепи отображается в скобках ()

Пример: CCC(CC)COO

Описание SMILES: циклы

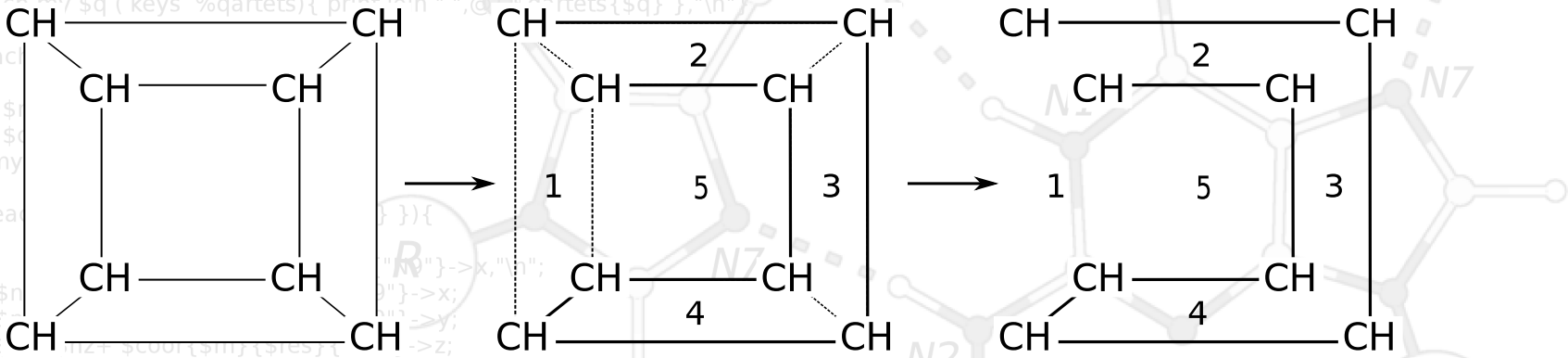
- C1CCCCC1 циклогексан
- Или более сложный пример:



a) CC1=CC(=CC=C1)Br

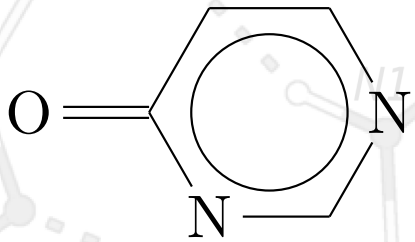
b) CC1=CC(Br)=CC=C1

- Самый сложный пример: C12C3C4C1C5C4C3C25



Описание SMILES: ароматика

- SMILES для определения ароматичности использует расширенный алгоритм Хюккеля.
- c1ccccc1 eq C1=CC=CC=C1 тут все атомы находятся в sp2-гибридизации
- c1cccc1 eq C1=CC=CC1 , последний атом в гибридизации sp3.
- Ароматичными могут быть атомы: C, N, O, P, S, As, Se, и * .
- Пример: c1cnc[nH]c(=O)1



Содержание

Активные молекулы

Химоинформатика

SMILES

SMARTS

QSAR

Докинг

SMARTS: паттерны для SMILES

В принципе, SMARTS это SMILES + операторы логики и варианты в позициях.

Примеры для атомов:

C	алифатический углерод
c	ароматический углерод
a	любой ароматический атом
[#6]	любой атом углерода
[++]	атом с зарядом +2
[R]	атом в кольце
[D3]	атом с тремя связями (не с водородами)
[X3]	атом с тремя связями, включая водороды
[v3]	атом с валентностью 3.

SMARTS: логические операторы и примеры

Логика:

!e1	not e1
e1&e2	a1 and e2
e1,e2	e1 or e2
e1;e2	a1 and e2

Примеры:

[!C;R]	не алифатический C в кольце
[n;H1], [n&H1], [nH1]	H в пирроле
[c,n&H1]	C или H в пирроле
[X3&H0]	Атом с тремя связями не с H
[c,n;H1]	N или C в связи с одним H1

Поиск по 3D-базам данным

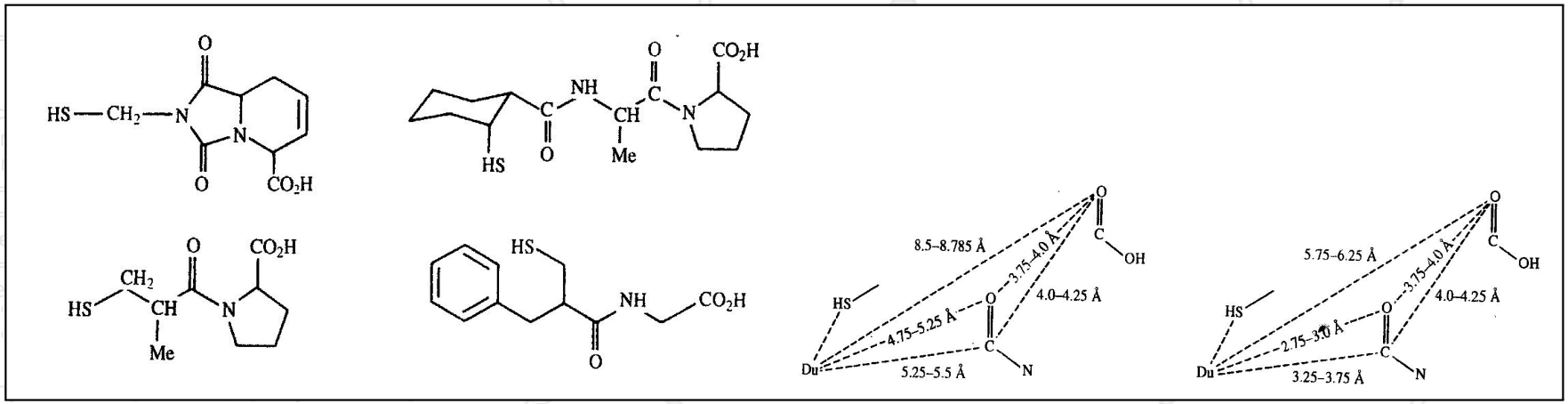
- Поиск в 2D-пространстве хорош для поиска подобных молекул, но биологически активные молекулы действуют благодаря специфической 3D-структуре.
- Взаимодействие с биополимером может происходить благодаря нужному расположению в пространстве некоторых групп. При этом различие в 2D-структуре может быть весьма существенным.
- Фармакофор — это набор свойств, которые являются общими для некоторой группы активных молекул.
- Пример: Антигистаминный 3D-фармакофор

Проблемы с фармакофорами

- Если молекулы более или менее подвижны, то это накладывает дополнительные требования на учёт конформационных превращений.
- Для определения фармакофора надо определить, какой набор групп располагается в биополимере идентично.
- Надо быть уверенным, что выбранный набор молекул связывается с белком в одном и том же месте. Однозначное указание на это можно получить только экспериментально.

Систематический поиск

- Есть проблема:



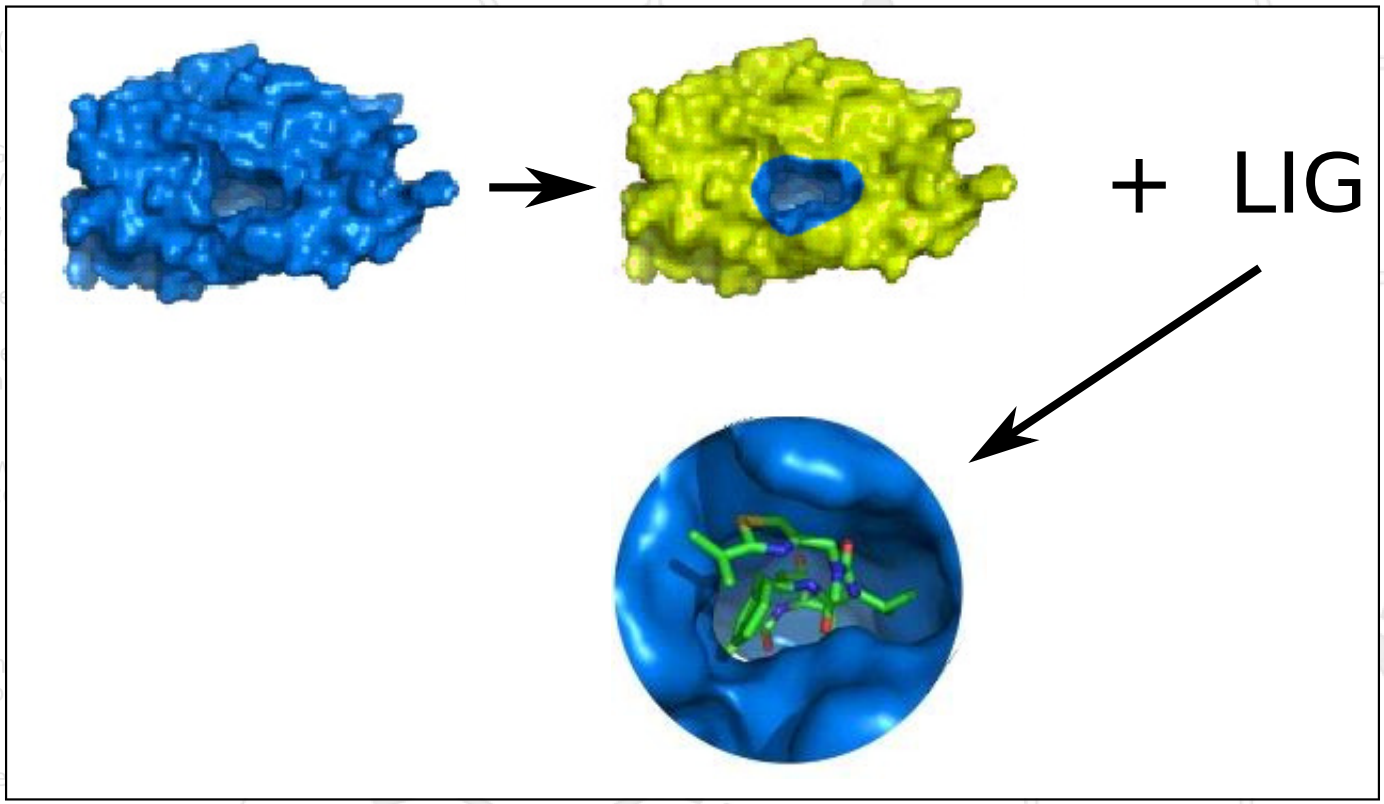
- Выбирают точки, которые по мнению исследователей определяют активность. Делают конформационный поиск для всех молекул. Если находят пересечения по геометрии, то на основе этих точек и геометрии пересечения формулируют фармакофор.

Алгоритмы ансамблей геометрических ограничений и МД.

- Конформационный поиск можно делать по-разному:
- Оптимизировать молекулы, вводя дистанционные ограничения
- Проводить молекулярную динамику с дистанционными ограничениями.
- Этим алгоритмам, естественно, надо знать, какие точки использовать.
- Генетические алгоритмы могут сами искать точки фармакофора.

Докинг белок-лиганд

Компьютерный метод, воспроизводящий связывание лиганда с белком



В результате мы можем узнать положение лиганда в комплексе с белком и оценить константу связывания

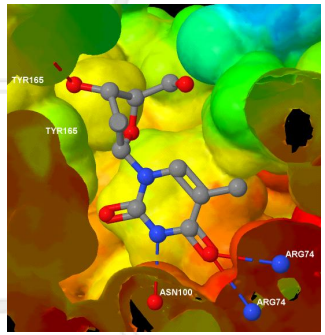
Положение в сайте связывания

- Сайт связывания — место связывания лиганда
- Геометрия связывания — место связывания, ориентация и конформация лиганда

Использование докинга

Основные цели докинга:

- Виртуальный поиск лигандов
- Определение геометрии связывания лиганда



Если мы знаем, как связывается лиганд, то:

- Мы можем
  узнать, какие части важны для связывания
- Можно предложить изменения для улучшения
  константы связывания
- Можем избежать ошибок

Два основных компонента программ для докинга

- Алгоритм поиска
  - Установление места связывания
  - Установление геометрии связывания
- Алгоритм расчёта константы связывания областей с низкой энергией.

Реализация

Сегодня существует много программ для докинга

- AutoDock, DOCK, e-Hits, FlexX, FRED, Glide, GOLD, LigandFit, QXP, Surflex-Dock…и т.д.
- разные алгоритмы оценки аффинности и разные алгоритмы поиска
- Важно не путать лиганд-белок докинг и белок-белок докинг

Практические аспекты

- Часто PDB-структура содержит молекулы воды, почти всегда их надо убрать.
- Надо добавлять протоны к структуре; His?
- Часто в PDB неточно определена ориентация некоторых групп, что сказывается на паттерне водородных связей.
- Протонирование лиганда и его таутомерные формы.

Rigid|Flexible докинг

- Rigid: лиганд не имеет внутренних степеней свободы, т.е. вращение вокруг связей запрещено.
- Flexible: предполагает учёт вращения вокруг связей лиганда.
- Часто белок рассматривается как жёсткое тело