

Лекция 13: Свойства лигандов, построение лигандов, QSAR

Курс: Молекулярное моделирование в применении к биомолекулам

Головин А.В. ¹

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

Москва, 2020

Содержание

Свойства лигандов

QSAR

Фрагментарное построение лиганда



Описание молекул

- **Molecular descriptors.** Параметры молекул это некие свойства в количественном описании, которые можно легко рассчитать зная только формулу молекулы.

Примеры:

- Молекулярная масса
- Количество атомов
- Распределение вода/октан
- Электротопологические индексы
- Молярная поляризуемость
- Топологические двугранные углы



Описание молекул

- 1D, брутто формула
 - Молекулярная масса
 - Количество атомов
- 2D, структура
 - Распределение вода/октан
 - Fingerprints
 - Топологические двугранные углы
- 3D, пространственная структура
 - Дипольный момент
 - Объем и поверхность молекулы
 - Частичные заряды на атомах



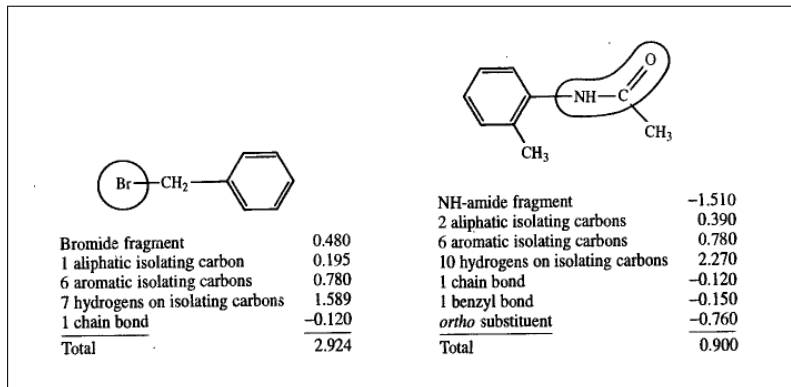
Распределение октанол/вода

- $\log P$ где P степень распределения вещества в системе октанол/вода.
- Для некоторых веществ трудно определить экспериментально.
- Можно использовать метод пертурбации свободной энергии. Есть ряд неудобств.
- Используют фрагментарный подход. Распределение для молекулы равно сумме распределения компонентов с коэффициентами.



Реализация: CLOGP

- Предложено разбивать молекулу по одинарным связям на фрагменты.



Молярная рефрактивность

Молярная рефрактивность отражает способность газа преломлять свет, что так же отражает поляризуемость молекулы.

$$MR = \frac{(n^2 - 1)MW}{(n^2 + 1)d}$$

d это плотность

n это некий индекс не сильно изменяющийся для органических соединений.

- На основании этого значения часто судят о размере молекулы и собственно плотности вещества
- Применяют разбиение на фрагменты, как в CLOGP



Топологические индексы.

- Хол и Кир предложили множество индексов позволяющих судить о молекуле в одной цифре.

$$v = \frac{p_i - h_i}{Z_i - p_i - 1}$$

p : количество s и p электронов ; h : количество атомов водорода при атоме; $d = p - h$






Chi — молекулярные соединения:

$${}^0\chi = \sum_{atoms} (d_i)^{-1/2}$$

$${}^0\chi^v = \sum_{atoms} (v_i)^{-1/2}$$

$${}^1\chi = \sum_{atoms} (d_i d_j)^{-1/2}$$

$${}^1\chi^v = \sum_{atoms} (v_i v_j)^{-1/2}$$

	4	3	2	1	4.828	2.914	1.707
	5	4	1	0	4.992	2.808	1.922
	5	3	2	0	4.992	2.770	2.183
	6	4	0	0	5.155	2.643	2.488
	7	3	0	0	5.207	2.561	2.914

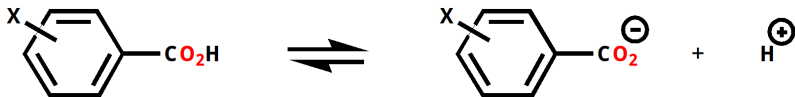
Leach, 1996



Электронный эффект

Константа δ отражает влияние заместителя на смещение электронной плотности.

Эффект можно увидеть при отрыве протона



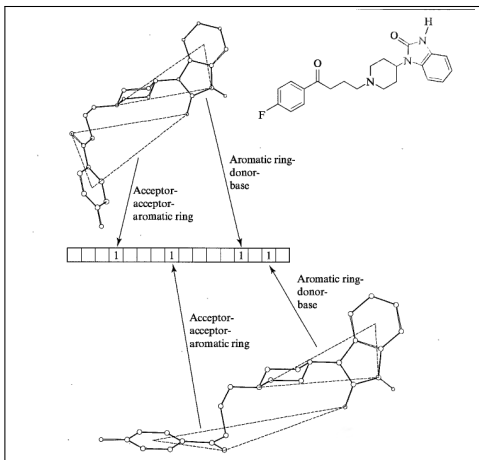
X=H

$$K_h = K_{diss} = \frac{[PhCO_2^-]}{[PhCO_2H]}$$



Фармакофорные индексы

- Как правило выделяют три точки.
- Определяют расстояния между точками для разных конформеров.



Leach, 1996

Создание выборки

- Для поиска заготовки для лекарства важно создать хорошую стартовую выборку соединений.
- Используют рассчитанные параметры молекул для оценки качества выборки.
- Надо добиться нормального распределения параметров.
- Избегают высокой корреляции для избежания пере представленности выборки.



Оценка подобия соединений

- Dice:

$$S_{a,b} = \frac{2 \sum_i^N x_{ia} x_{ib}}{\sum_i^N x_{ia}^2 + \sum_i^N x_{ib}^2}$$

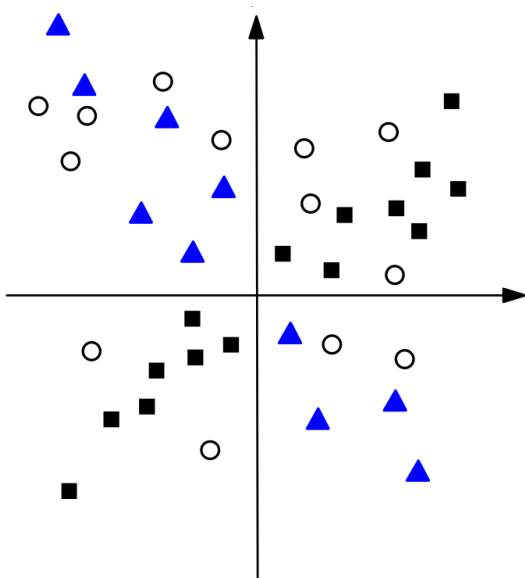
или

$$S_{a,b} = 2 \frac{|x_a \cap x_b|}{|x_a| + |x_b|}$$

- Тanimoto коэффициент.
- Cosine коэффициент.
- Евклидово расстояние.



Redundancy



QSAR, количественные соотношения структура/активность

- QSAR — построение математической модели для описания соотношений структура-активность
Задача предсказать свойства молекулы *in vivo*
В целом такое соотношение можно записать как:

$$v = f(p)$$

где v : активность, а p : свойства из структуры

$$\log\left(\frac{1}{C}\right) = k_1 \log P - k_2 (\log P)^2 + k_3 \sigma + k_4$$

$$\pi = \log\left(\frac{P_X}{P_H}\right); \quad \log\left(\frac{1}{C}\right) = k_1 \pi - k_2 \pi^2 + k_3 \sigma + k_4$$

- На сегодняшний день существует множество модификаций этой простой формулы, которые учитывают множество параметров молекулы



QSAR, выбор соединений

- Для поиска коэффициентов надо синтезировать набор соединений с разными P .
- Желательно равномерное распределение выбранных соединений в шкале P .
- Возможна также вариация соединений при разных pH и температуре.



Получение уравнения

- Простейший подход это определение зависимости активности от параметра как линейной регрессии. Тогда определение коэффициентов это просто метод наименьших квадратов.
- Линейная регрессия расширяется до множественной линейной регрессии, где более чем одна независимая переменная.
- Считается, что для статистически значимого определения необходимо не менее 5 соединений на каждый параметр.
- Существуют генетические алгоритмы для поиска коэффициентов регрессии.



Кросс-валидация

Распространённый способ проверки это выборочное удаление данных из выборки и сравнение результатов.



Использование QSAR

- Главная задача это предсказать оптимальную структуру
- Часто бывает, что QSAR хорошо работает при интерполяции, а не при экстраполяции.
- Бывает необходимо использовать не линейную зависимость, применяют параболическую.
- Существует билинейная модель:

$$\log(1/C) = k_1 \log P - k_2 (\log(\beta P + 1)) + k_3$$

- Отсутствие корреляций для параметра скорее всего означает, что этот параметр не задействован в механизме.



ADME

- Absorption, Distribution, Metabolism, Elimination
- „Fail early, fail fast, fail cheap“

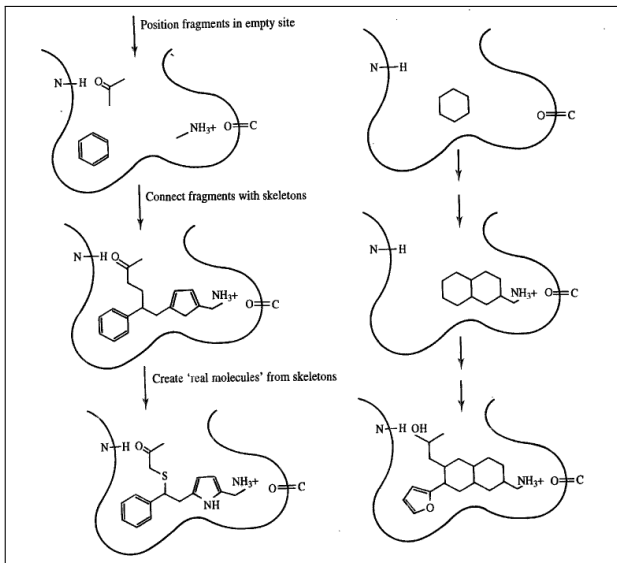


Фрагментарное построение лиганда

- Сканирование по базам данных удобно, так как позволяет тут же проверить молекулу.
- Сканирование не предполагает создание лиганда *de novo*.
- Суть фрагментарного построения лиганда состоит в поиске мест в активном центре белка, где связываются небольшие фрагменты молекул
- Соединяя фрагменты связями при сохранении места связывания, добиваются высоких констант.



Фрагментарное построение лиганда



Фрагментарное построение лиганда, реализация

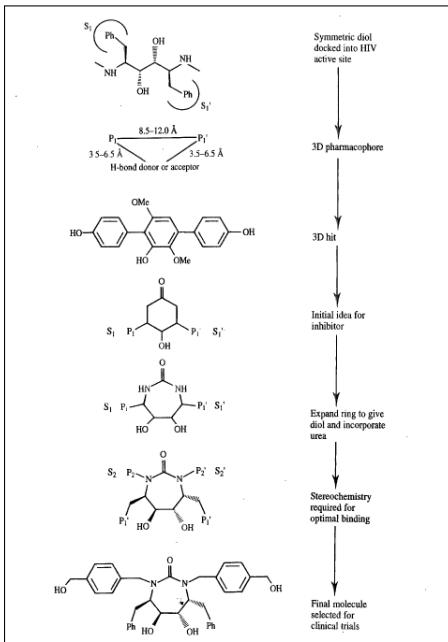
- GRID аналог докинга.
- MCSS сайт наполняется фрагментами и с помощью EM вычленяется место где фрагмент наиболее предпочтителен. Взаимодействия между фрагментами не учитываются.
- LUDI использует информацию из банка PDB для задания фрагментов образующих водородные связи и т.д.
- Необязательно всё моделировать, можно использовать PCA и YMP для определения места связывания фрагмента.



Реализация ковалентного связывания фрагментов

- Если у Вас есть два и более фрагмента, то можно искать способ их соединения по базам данных.
- Реализовано в CAVEAT.
- Можно строить автоматически строить скелеты. Главный критерий это сохранение взаимного положения фрагментов.
- Переход от скелета к молекуле сложен, так как надо реализовать возможность синтеза молекулы.





Leach, 1996