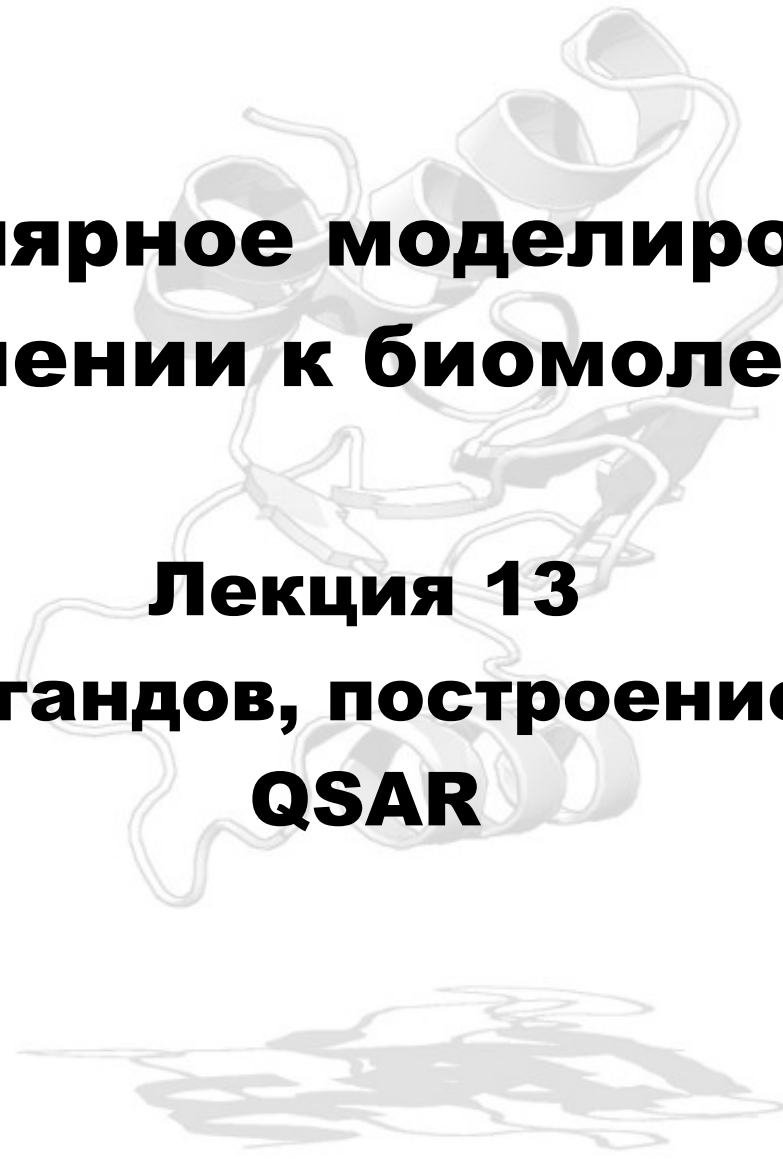


A faint, grayscale background image of a protein structure, likely a ribbon diagram, is centered behind the text. The protein has several alpha-helices and beta-sheets, with a complex, folded shape. It appears to be a monomeric protein, possibly an enzyme or a signaling protein, given its intricate folding.

Молекулярное моделирование в применении к биомолекулам

Лекция 13

**Свойства лигандов, построение лигандов,
QSAR**

Описание молекул

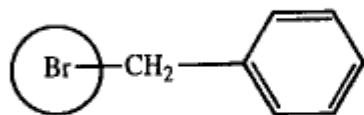
- Molecular descriptors. Параметры молекул это некие свойства в количественном описании, которые можно легко рассчитать зная только формулу молекулы
- Примеры:
 - Молекулярная масса
 - Количество атомов
 - Распределение вода/октан
 - Электротопологические индексы
 - Молярная поляризуемость
 - Топологические двугранные углы

Распределение октанол/вода

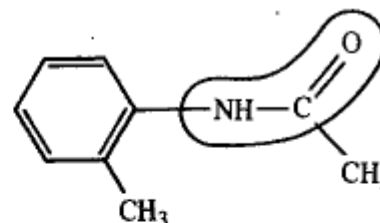
- $\log P$ где P степень распределения вещества в системе октанол/вода.
- Для некоторых веществ трудно определить экспериментально.
- Можно использовать метод пертурбации свободной энергии. Есть ряд неудобств.
- Используют фрагментарный подход. Распределение для молекулы равно сумме распределения компонентов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

Реализация: CLOGP

- Предложено разбивать молекулу по одинарным связям на фрагменты.



Bromide fragment	0.480
1 aliphatic isolating carbon	0.195
6 aromatic isolating carbons	0.780
7 hydrogens on isolating carbons	1.589
1 chain bond	-0.120
<u>Total</u>	<u>2.924</u>



NH-amide fragment	-1.510
2 aliphatic isolating carbons	0.390
6 aromatic isolating carbons	0.780
10 hydrogens on isolating carbons	2.270
1 chain bond	-0.120
1 benzyl bond	-0.150
<i>ortho</i> substituent	-0.760
<u>Total</u>	<u>0.900</u>

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

Топологические индексы.

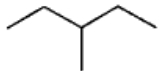
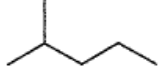
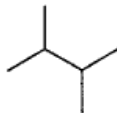
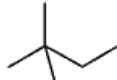
- Хол и Кир предложили множество индексов позволяющих судить о молекуле в одной цифре.
- Основа это дельта значения для тяжёлых атомов
 $\delta_i = \sigma_i - h_i$ $\delta_i^v = Z_i^v - h_i$ $\delta_i^v = (Z_i^v - h_i) / (Z_i - Z_i^v - 1)$ (для $> F$)
 δ : количество сигма электронов
 Z : количество валентных электронов
- Chi — молекулярные соединения

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

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

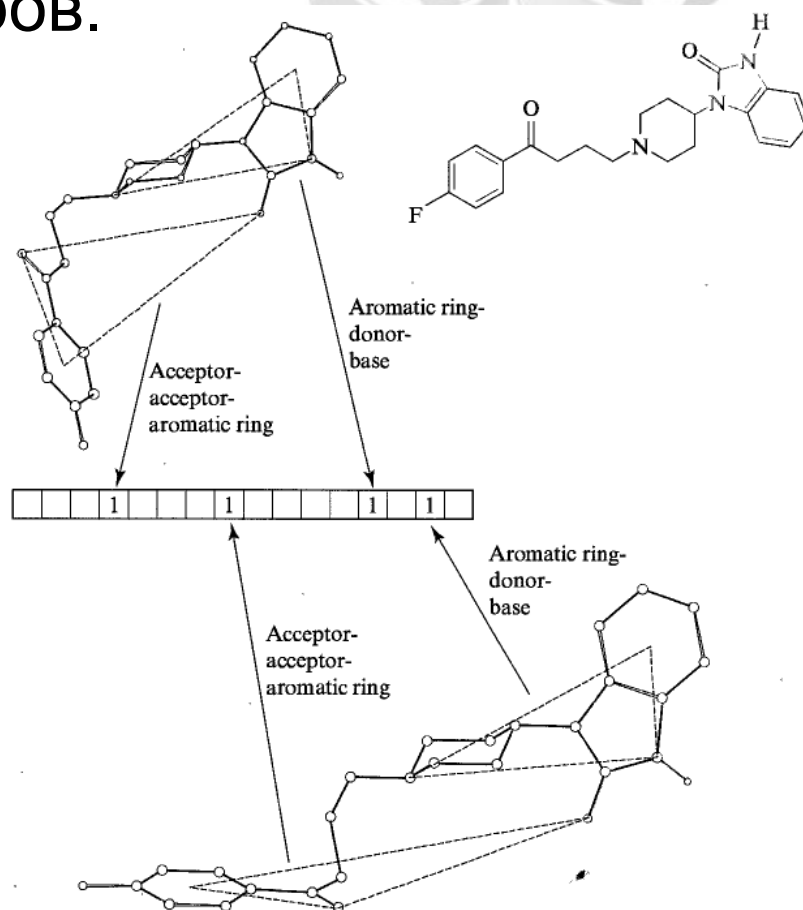
$${}^1\chi = \sum_{atoms} (\delta_i \delta_j)^{-1/2};$$

$${}^1\chi^v = \sum_{atoms} (\delta_i^v \delta_j^v)^{-1/2};$$

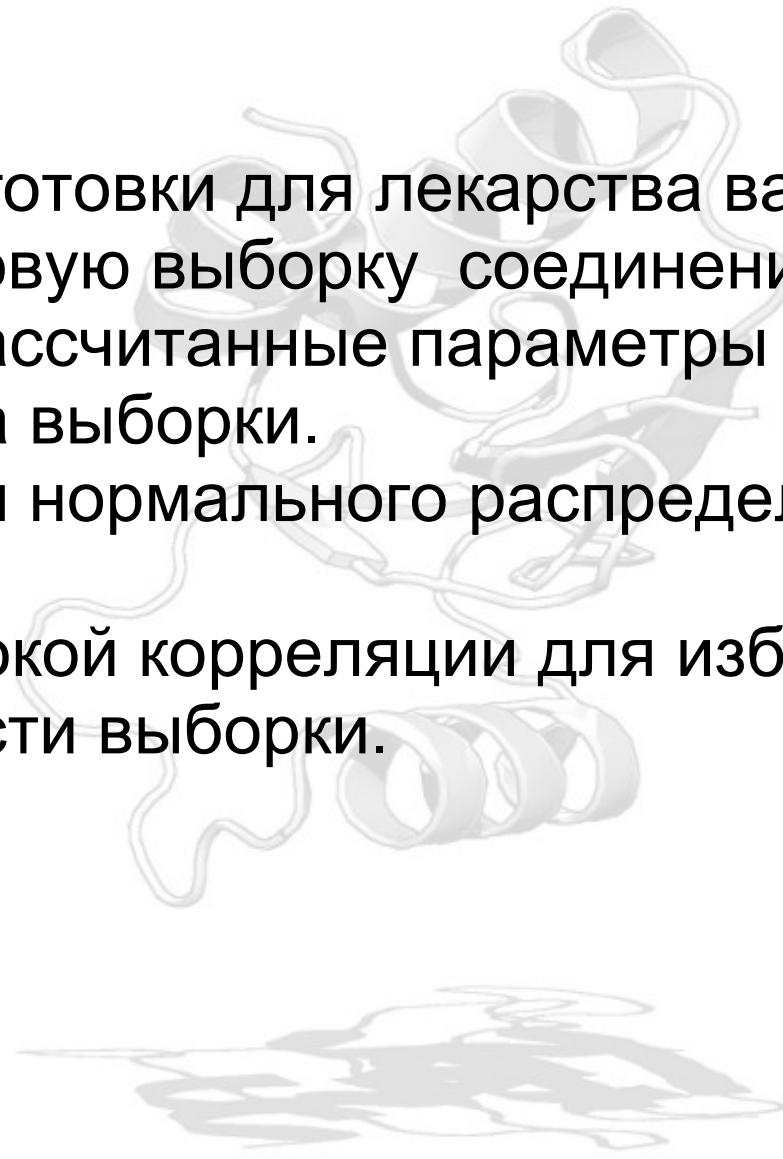
	4	3	2	1	4.828	2.914	1.707
	5	4	1	0	4.992	2.808	1.922
	5	3	2	0	4.992	2.770	2.183
	6	4	0	0	5.155	2.643	2.488
	7	3	0	0	5.207	2.561	2.914

Фармакофорные индексы

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



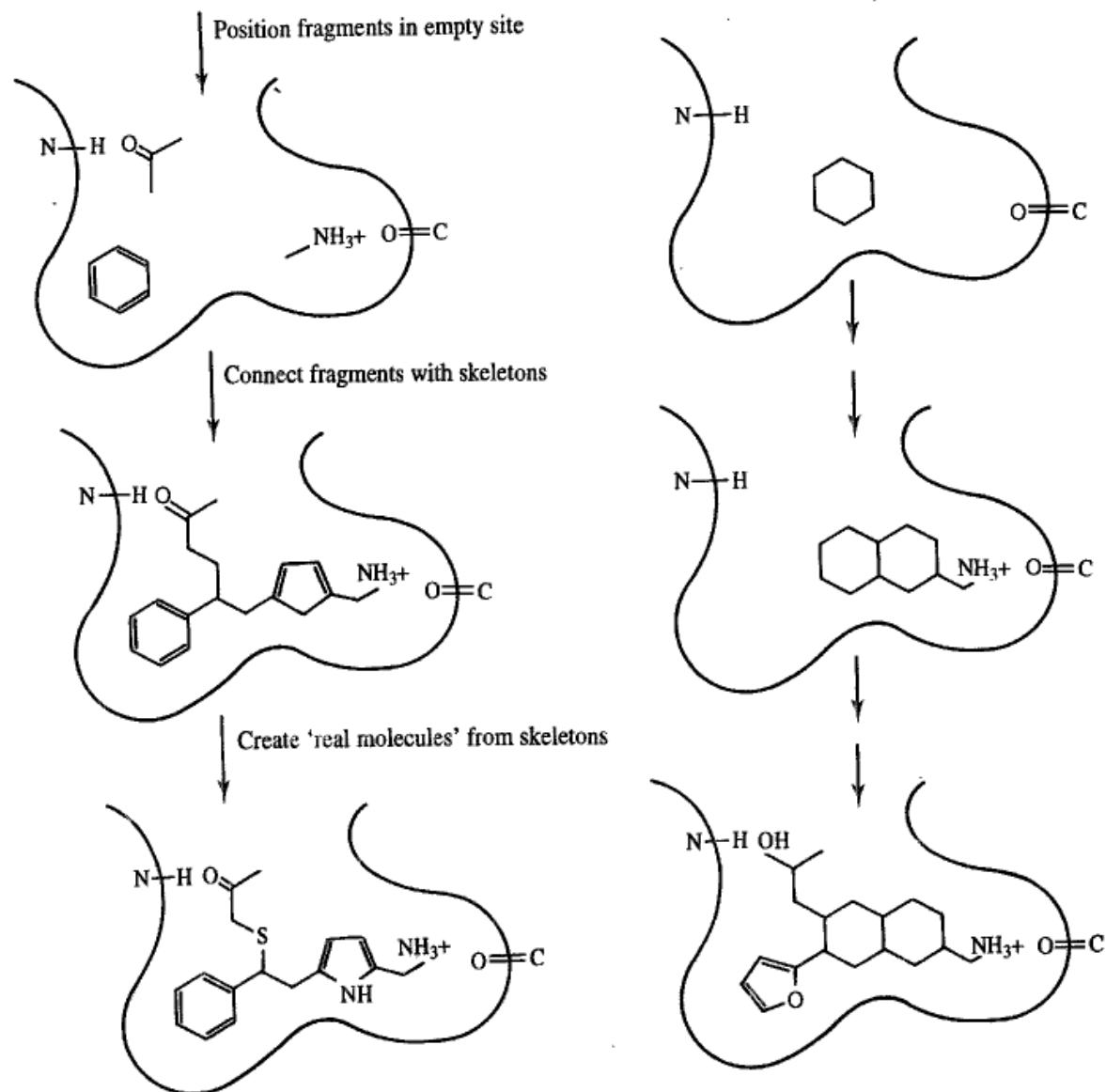
Создание выборки

- 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.
- 

Фрагментарное построение лиганда

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

Фрагментарное построение лиганда

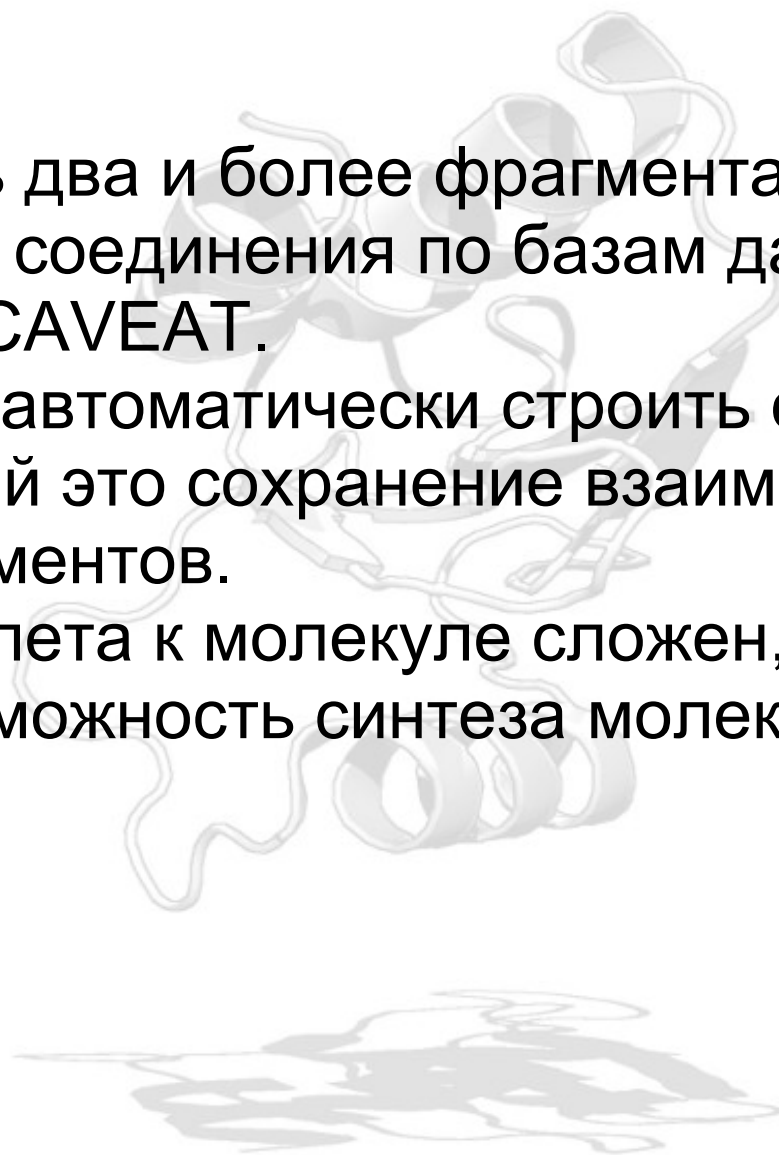


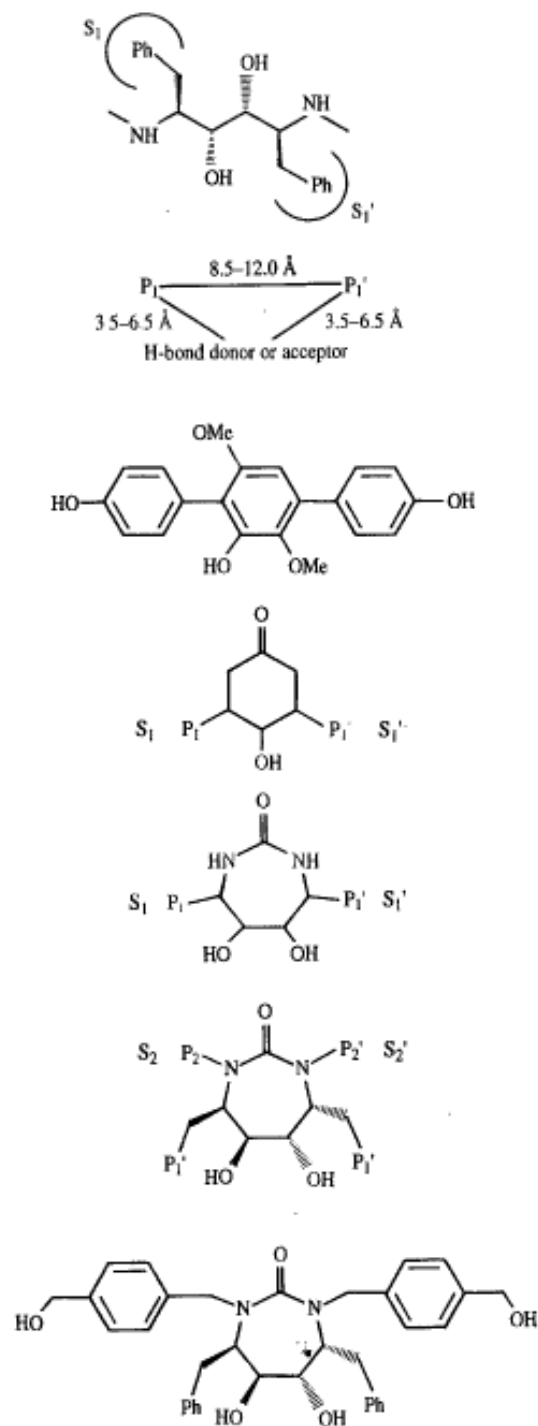
Фрагментарное построение лиганда, реализация

- 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 ЯМР 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.





Symmetric diol
docked into HIV
active site

3D pharmacophore

3D hit

Initial idea for
inhibitor

Expand ring to give
diol and incorporate
urea

Stereochemistry
required for
optimal binding

Final molecule
selected for
clinical trials

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

$$\log(1/C) = k_1 \log P - k_2 (\log P)^2 + k_3 \sigma + k_4$$

$$\pi = \log(P_X / P_H)$$

$$\log(1/C) = k_1 \pi - k_2 \pi^2 + k_3 \sigma + k_4$$

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

QSAR, электронный эффект

Константа δ отражает влияние заместителя на смещение электронной плотности.

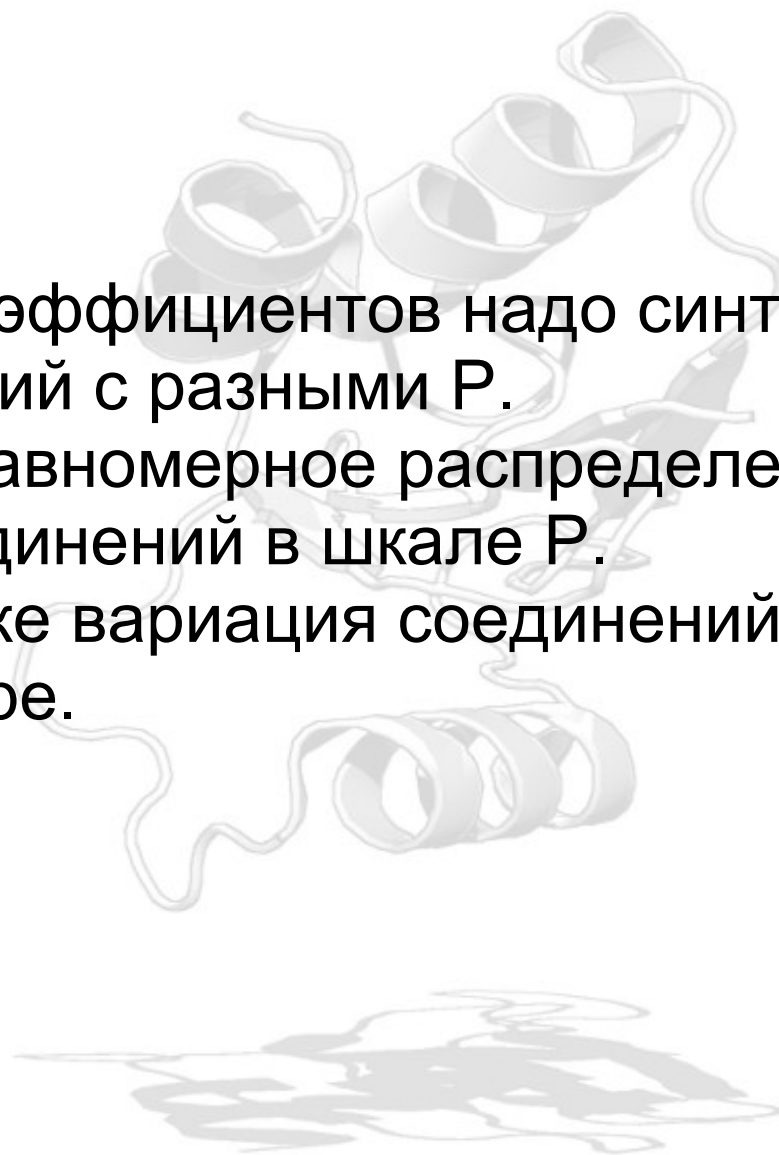
Эффект можно увидеть при отрыве протона



$X=\text{H}$ $K_{\text{H}} = \text{Dissociation constant} = \frac{[\text{PhCO}_2^-]}{[\text{PhCO}_2\text{H}]}$

QSAR, выбор соединений

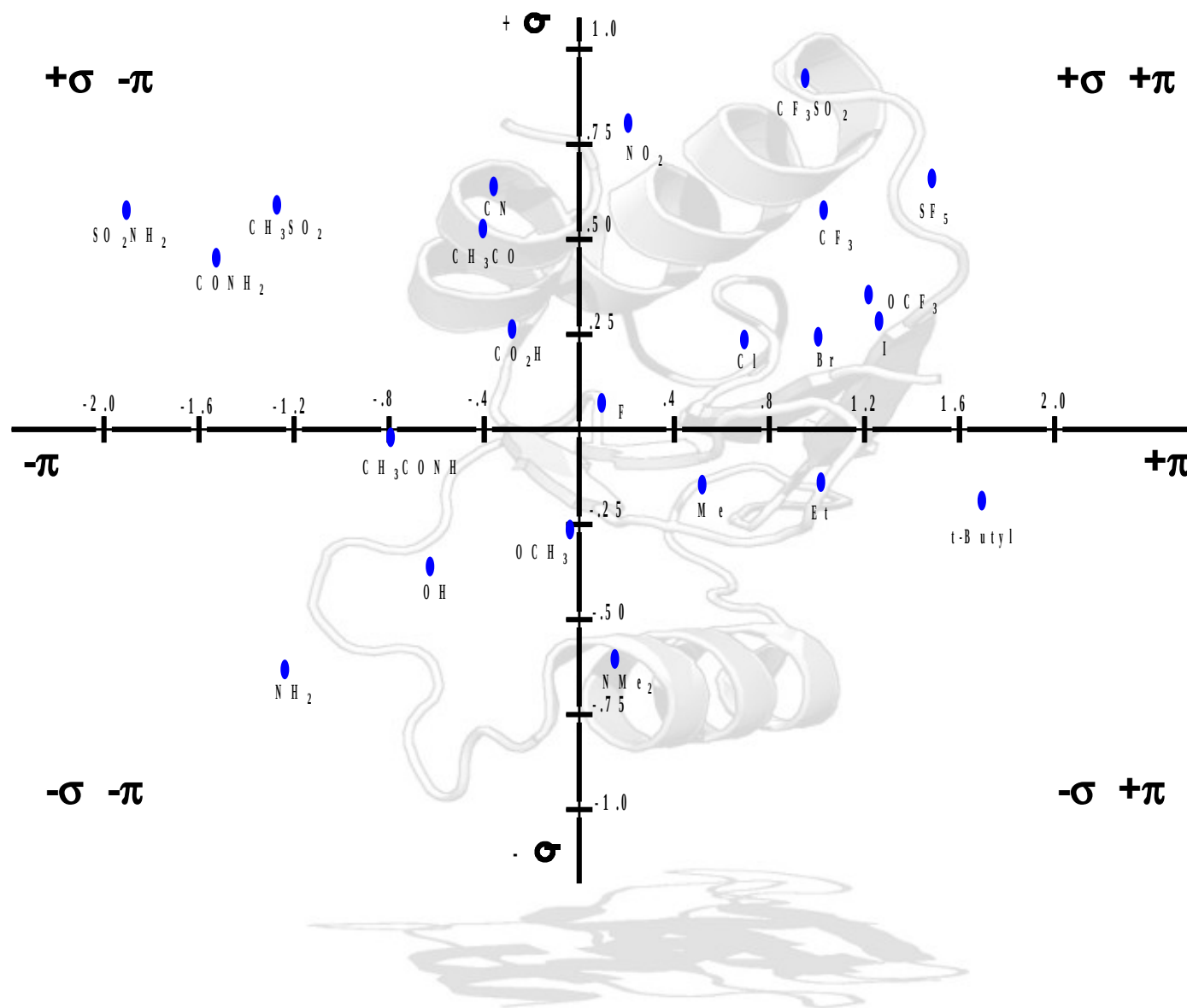
- Для поиска коэффициентов надо синтезировать набор соединений с разными R .
- Желательно равномерное распределение выбранных соединений в шкале R .
- Возможна также вариация соединений при разных pH и температуре.



Получение уравнения

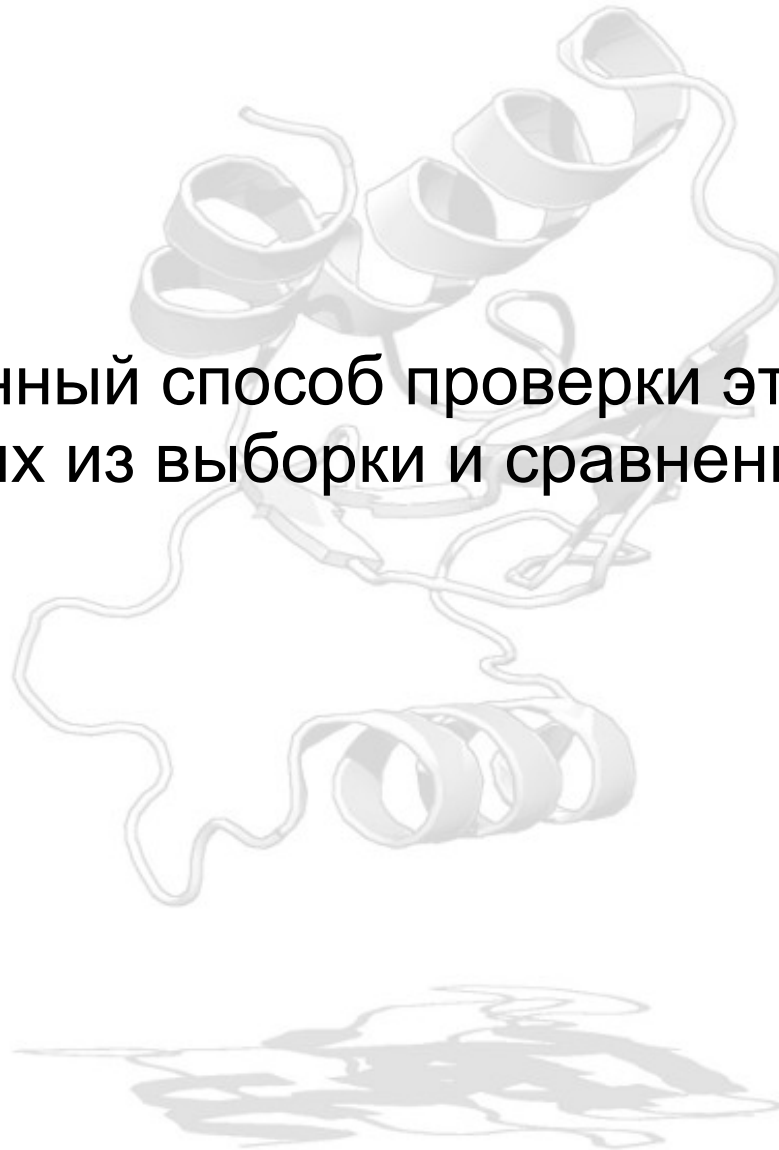
- Простейший подход это определение зависимости активности от параметра как линейной регрессии. Тогда определение коэффициентов это просто метод наименьших квадратов.
- Линейная регрессия расширяется до множественной линейной регрессии, где более чем одна независимая переменная.
- Считается, что для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

Диаграмма Крейга



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.