

# ХИМИЧЕСКАЯ ИНФОРМАТИКА: ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ

---

## Лекция 3

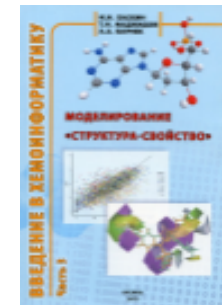
Методы исследования количественных соотношений пространственная  
структура - активность  
(3D QSAR)

# РЕКОМЕНДОВАННАЯ ЛИТЕРАТУРА

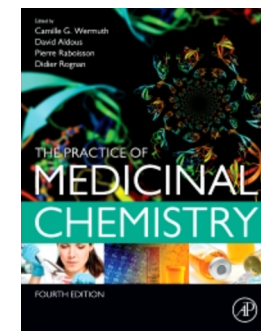
J.-P.Doucet, A.Panaye Three dimensional QSAR: Applications in Pharmacology and Toxicology  
CRC Press (Taylor & Francis Group) (2010), 520 c



Серия учебных пособий «Введение в хемоинформатику», Т.И.Маджидов, И.И.Баскин,  
И.С. Антипин, А.А.Варнек. Ч. 3 Моделирование «структура-свойство»

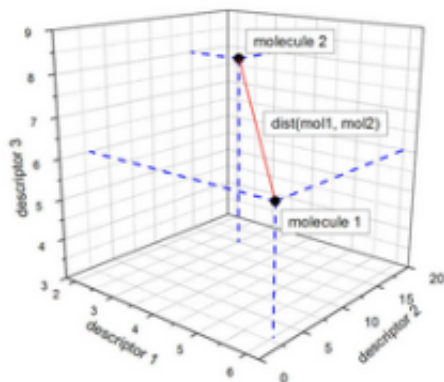
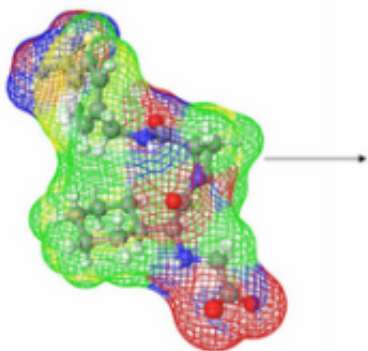


The Practice of Medicinal Chemistry 4th Edition. Edited by David Aldous Pierre Raboisson,  
Didier Rognan, Academic Press (2015), 902 c



# Химическая информатика: взаимосвязь структура - свойство

Поиск количественных соотношений структура-свойство (Quantitative Structure-Activity Relationships - QSAR) — процедура построения моделей, позволяющих по структурам химических соединений предсказывать их свойства.



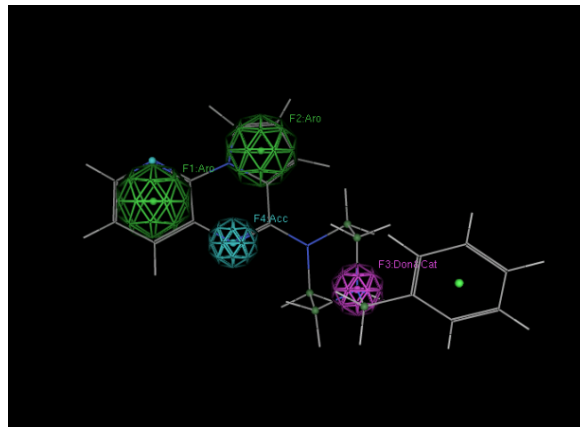
$$\text{Molecule } i : M(i) = \sum_1^j d_j(i)$$

При векторном описании химической структуре ставится в соответствие вектор молекулярных дескрипторов (числовых параметров, характеризующих структуру соединения).

- Quantitative Structure-Activity Relationship (QSAR) - прогноз количественных характеристик биологической активности.
- Quantitative Structure-Property Relationship (QSPR) - прогноз физических и физикохимических свойств органических соединений.
- Structure-Activity Relationship (SAR) - качественное описание соотношений между структурами химических соединений и их биологической активностью

# 3D QSAR

- Установление соответствия между биологической активностью ряда структурно и биологически охарактеризованных соединений и пространственными особенностями различных молекулярных полей, таких как стерические, гидрофобные и электростатические => различия в исследуемом свойстве соотносятся с эквивалентным изменением формы и энергии полей нековалентных взаимодействий
- Поиск оптимального положения фармакофорных фрагментов в пространстве



- доноры водородной связи ( $-OH$ ,  $-NH_2$ )
- акцепторы водородной связи ( $=O$ )
- ароматические группы
- гидрофобные группы

# РОЛЬ ФУНКЦИОНАЛЬНЫХ ГРУПП ВО ВЗАИМОДЕЙСТВИИ ЛИГАНД-РЕЦЕПТОР

$$K_d = \frac{[ligand][receptor]}{[complex]}$$

Free energy of binding:

$$\Delta G^0 = -RT \times \ln(K_d/C^0) = -2.303RT \times \log(K_d/C^0)$$

## Electrostatic Interactions

- Charge-charge interaction or ionic bonds
- Charge-dipole and dipole-dipole interactions
- Inductive interactions
- Hydrogen bonds
- Cation- $\pi$  interactions
- Arene-arene interactions

## Steric interactions

- Dispersion forces
- Short-range repulsive forces
- Conformational energy

## Enthalpy/entropy compensation

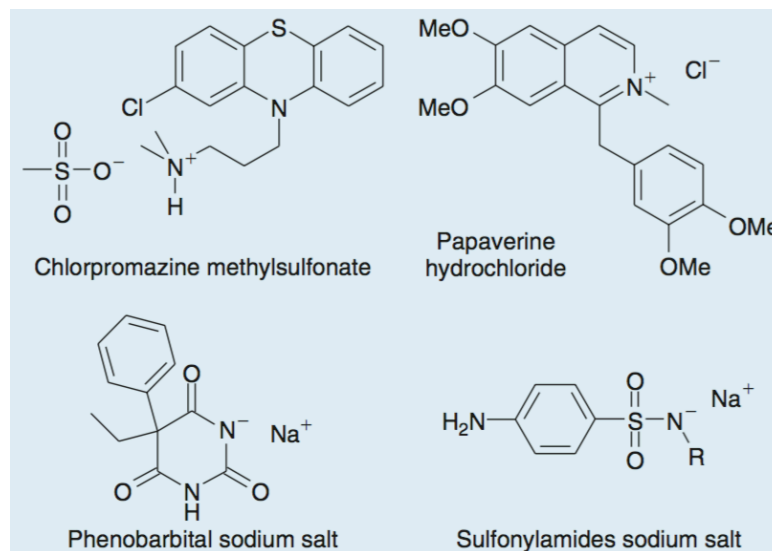
Type of interaction	Energy and dependency on distance	Example
<i>Charge-charge</i> Longest-range force, nondirectional	$\Delta G$ of $-20$ to $-40$ kJ/mol $1/r$	
<i>Charge-dipole</i> Depends on orientation of dipole	$\Delta G$ of $-12$ to $-20$ kJ/mol $1/r^2$	
<i>Dipole-dipole</i> Depends on mutual orientation of dipoles	$\Delta G$ of $-4$ to $-12$ kJ/mol $1/r^3$	
<i>Charge-induced dipole</i> Depends on polarizability of molecule in which dipole is induced	$\Delta G$ of $-2$ to $-10$ kJ/mol $1/r^4$	
<i>Dipole-induced dipole</i> Depends on polarizability of molecule in which dipole is induced	$\Delta G$ of $-2$ kJ/mol $1/r^6$	
<i>Dispersion</i> Involves mutual synchronization of fluctuating charges	$\Delta G$ of $-2$ to $-4$ kJ/mol $1/r^6$	
<i>Hydrogen bond</i> Charge attraction + partial covalent bond	$\Delta G$ of $-4$ to $-30$ kJ/mol	

# РОЛЬ ФУНКЦИОНАЛЬНЫХ ГРУПП ВО ВЗАИМОДЕЙСТВИИ ЛИГАНД-РЕЦЕПТОР

## Main Ionizable Groups in Proteins and Nucleic Acids

Chemical function	Charge	$pK_a$
<i>Fully or almost fully ionized groups at pH 7.4</i>		
Carboxyl $\alpha$ (terminal COOH)	-	1.8–2.4
Carboxyl $\beta$ (Asp)	-	3.7
Carboxyl $\gamma$ (Glu)	-	4.3
Primary phosphoryl	-	0.7–1.0
Secondary phosphoryl	-	5.9–6.0
$\alpha$ -Ammonium (Lys)	+	10.5
Guanidinium	+	12.5
<i>Partially ionized groups at pH 7.4</i>		
Sulfhydryl (Cys)	-	8.2
Imidazolium (His)	+	6.0
$\alpha$ -Ammonium (terminal peptide $NH_2$ )	+	7.5–10.3
N Amidic (Glu)	+	0.1
N Amidic (Asp)	+	8.8
<i>Nonionized groups at pH 7.4</i>		
Phenolic hydroxyl (Tyr)	-	10.0
Heteroaromatic hydroxyl (uracyl, thymine, guanine)	-	9.2–9.8
Osidic hydroxyl	-	12.3–12.6
Amino residue (adenine, guanine, cytosine)	+	3.3–4.6

## Cationic and Anionic drugs



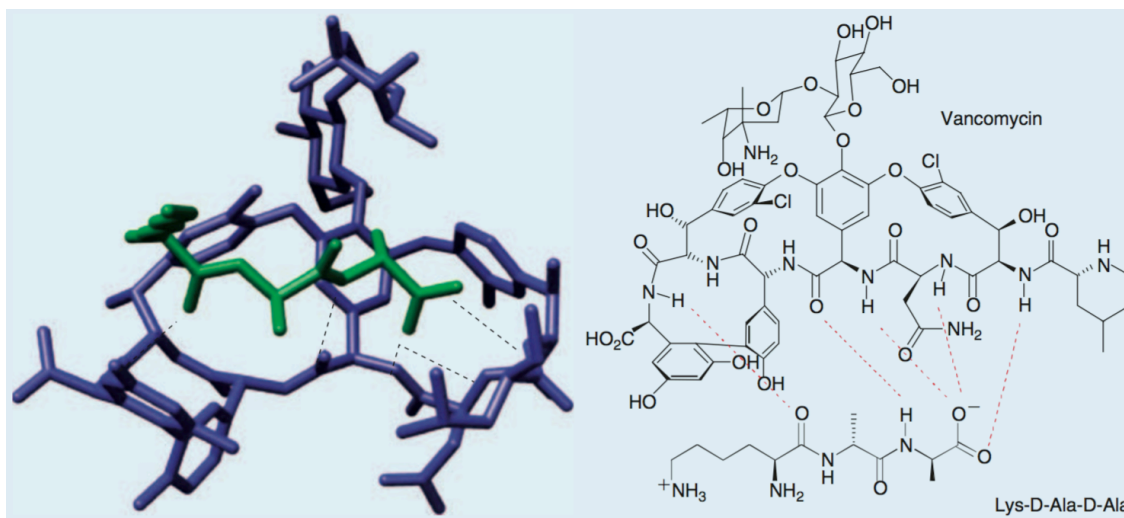
# РОЛЬ ФУНКЦИОНАЛЬНЫХ ГРУПП ВО ВЗАИМОДЕЙСТВИИ ЛИГАНД-РЕЦЕПТОР

Potential Hydrogen Bond Donor and Acceptor Groups Classified According to Their Strength of Interaction

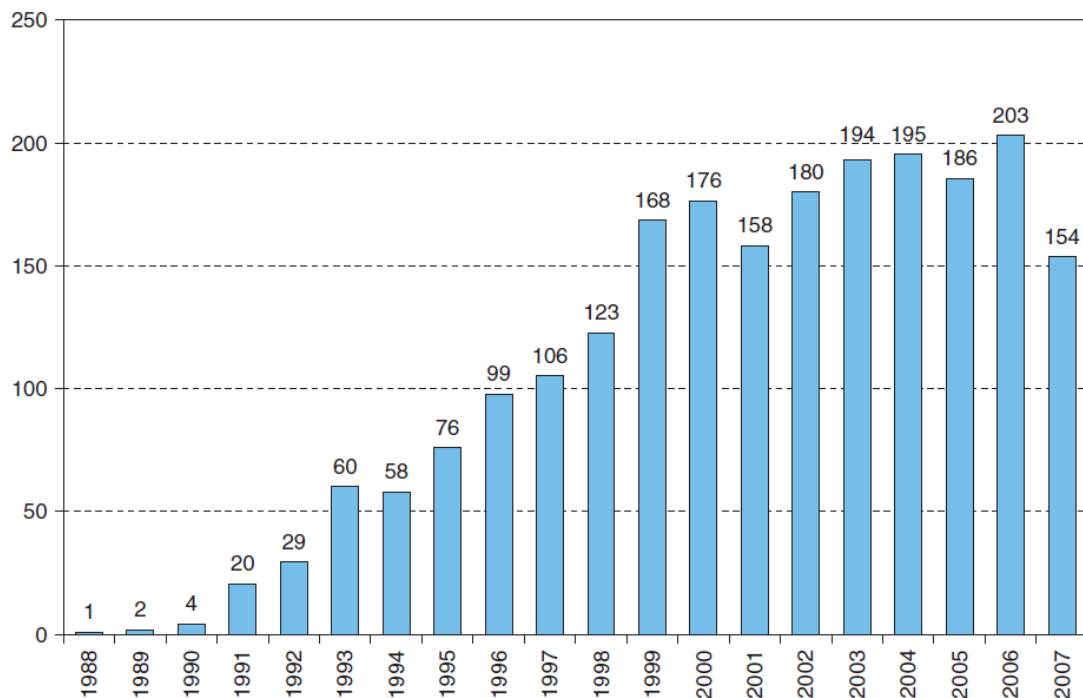
	Donor*	Acceptor
Very strong	$N^+H_3$ , $X^+-H$ , $F-H$	$CO_2^-$ , $O^-$ , $N^-$ , $F^-$
Strong	$O-H$ , $N-H$ , $Hal-H$	$O=C$ , $O-H$ , $N$ , $S=C$ , $F-H$ , $Hal^-$
Weak	$C-H$ , $S-H$ , $P-H$ , $M-H$	$C=C$ , $Hal-C$ , $\pi$ , $S-H$ , $M$ , $Hal-M$ , $Hal-H$ , $Se$

\*X is any atom, Hal is any of the lighter halogens, and M is a transition metal

Crystal structure of a short peptide l-Lys-d-Ala-d-Ala (bacterial cell wall precursor – in green) bound to the antibiotic vancomycin (in blue) through 5 hydrogen bonds.



# 3D QSAR: ПУБЛИКАЦИОННАЯ АКТИВНОСТЬ



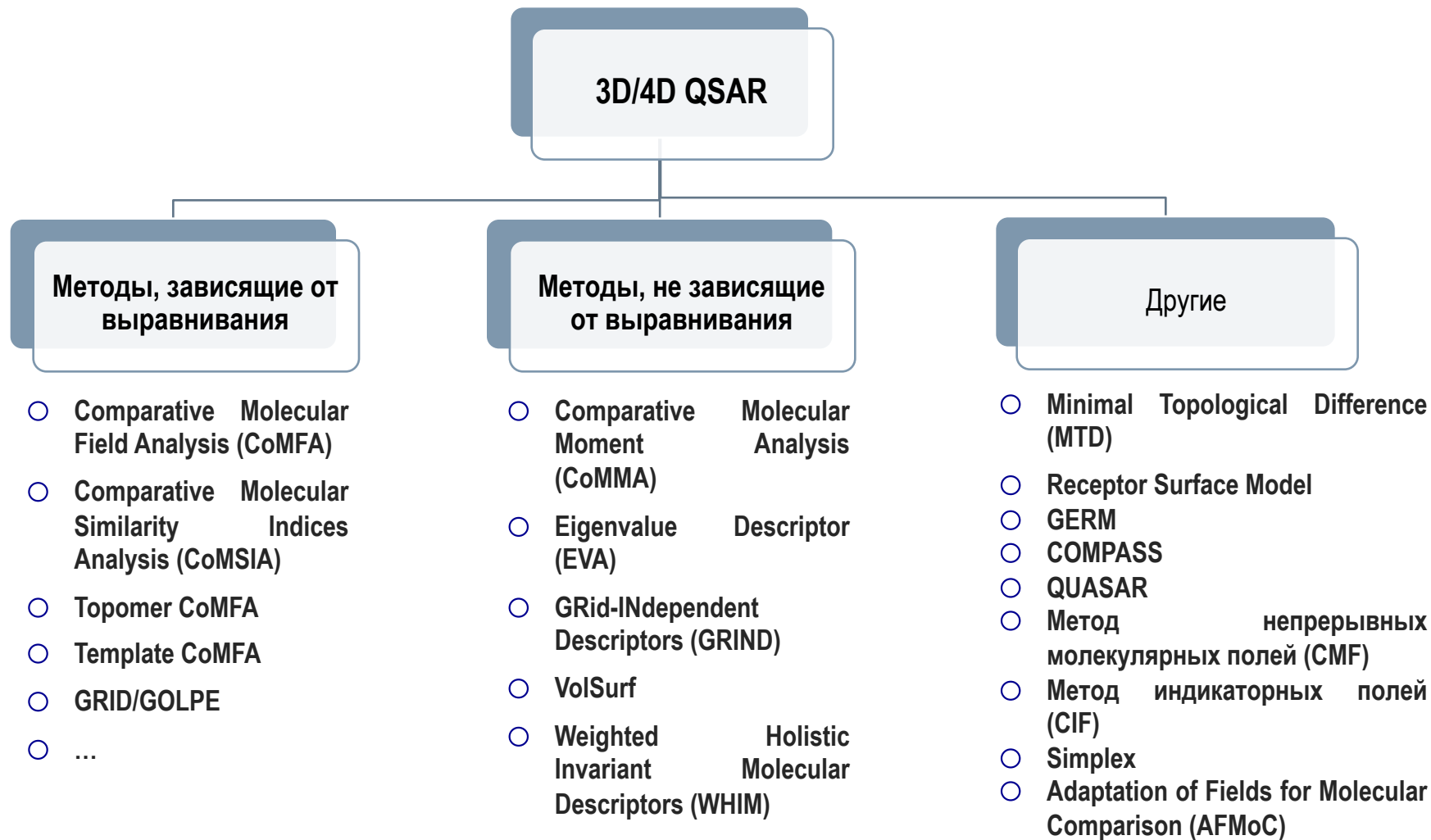
**TABLE 29.1** Scientific Journals Containing Articles about 3D QSAR, as Referred by CAS Online September 2007

Journal	Number of 3D QSAR articles
<i>J. Med. Chem.</i>	381
<i>Bioorg. Med. Chem.</i>	204
<i>J. Comput.-Aided Mol. Des.</i>	130
<i>J. Chem. Inf. Comput. Sci.*</i>	124
<i>Bioorg. Med. Chem. Lett.</i>	64
<i>Quant. Struct.-Act. Relat.</i>	48
<i>Eur. J. Med. Chem.</i>	47
<i>J. Mol. Graph. Model.</i>	45

\*Incl. *J. Chem. Inf. Model.*



# ТАКСОНОМИЯ МЕТОДОВ 3D QSAR

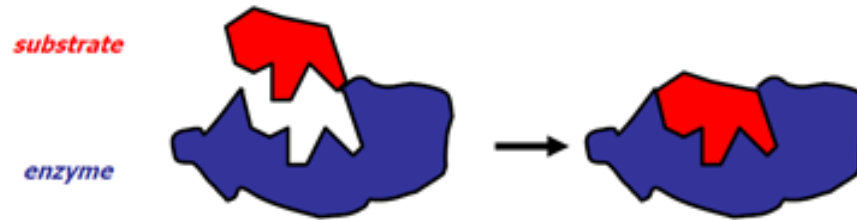


Martin , Y. C. 3D QSAR: current state, scope and limitations *Perspect. Drug Discov. Des.*(1998), 12, 3 – 23.

Norinder , U. Recent progress in CoMFA methodology and related techniques . *Perspect. Drug Discov. Des.* (1998) , 12 , 3 – 23 .

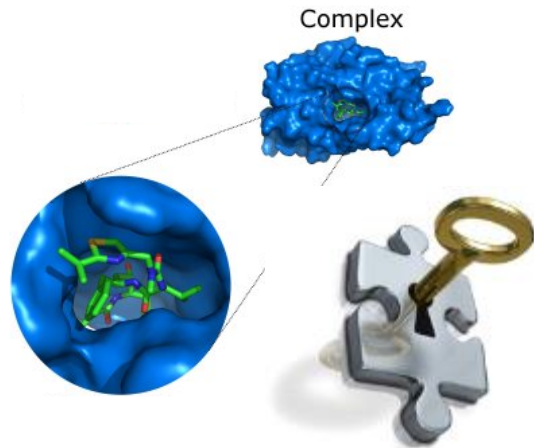
# 3D QSAR

## Принцип «ключ-замок»



Docking

3D QSAR



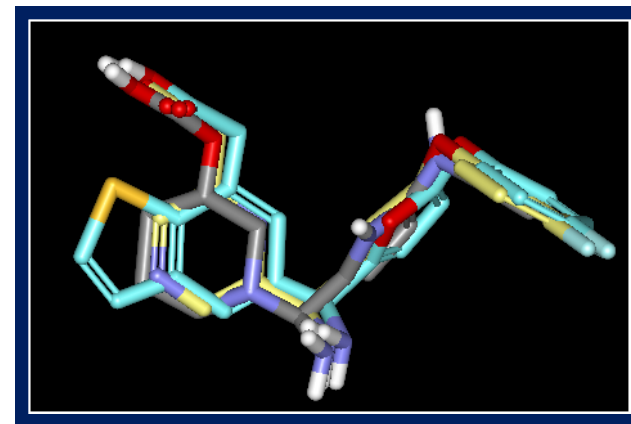
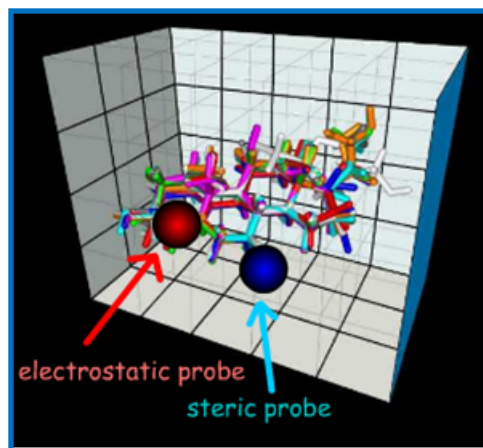
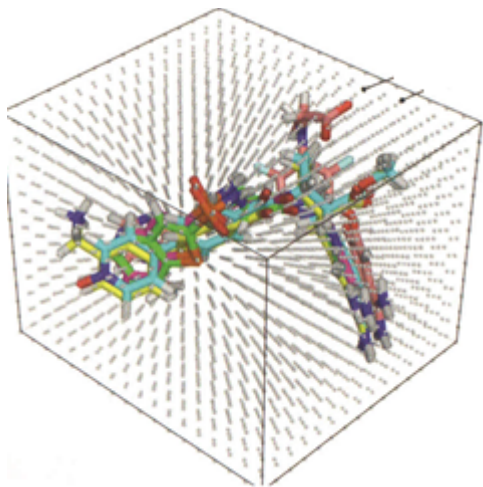
Большая часть алгоритмов  
3D QSAR



3D QSAR, основанный  
на знании строения  
рецептора



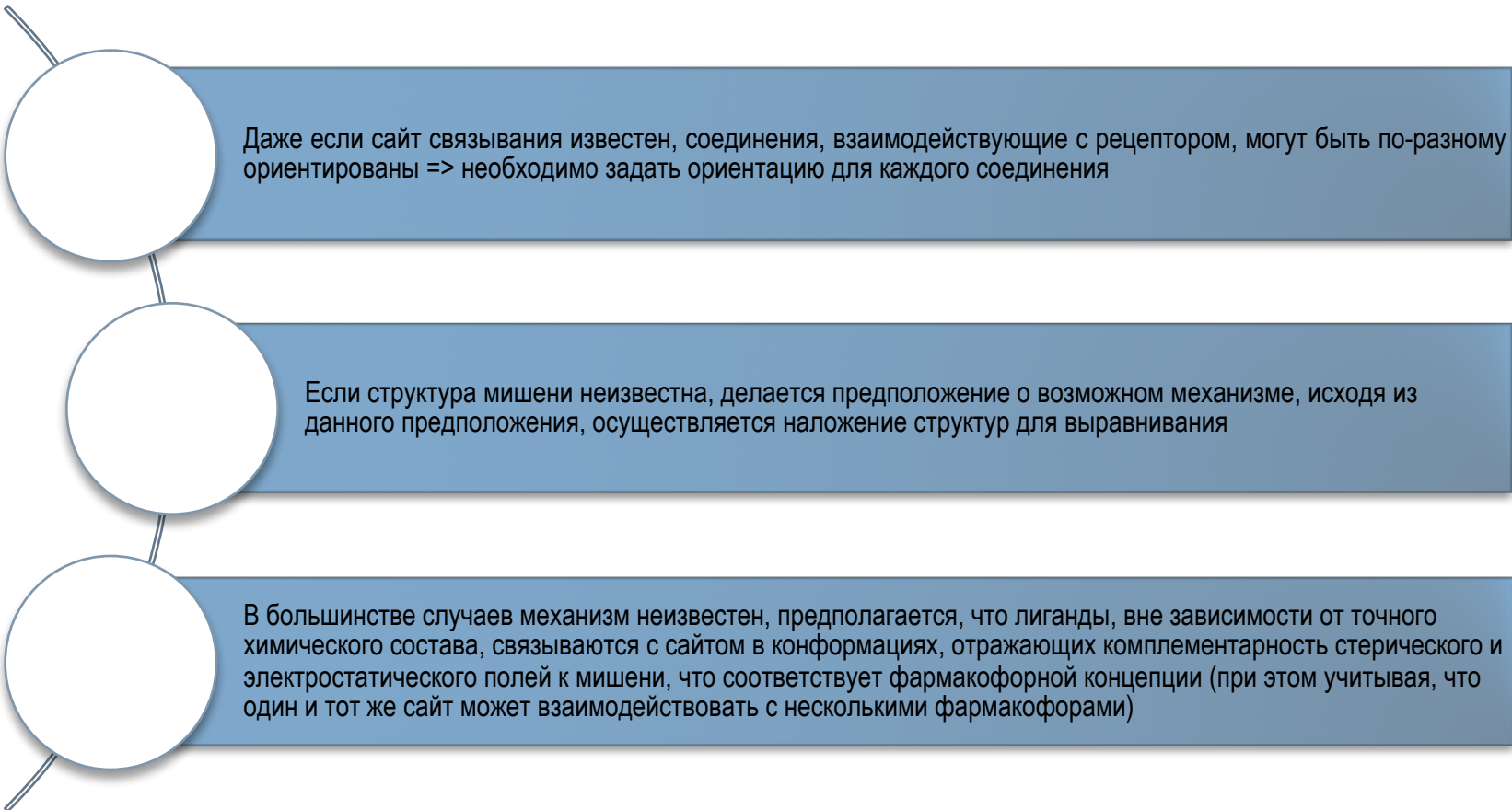
# МЕТОДЫ 3D QSAR, ЗАВИСЯЩИЕ ОТ ВЫРАВНИВАНИЯ



Методы предполагают выравнивание молекул относительно общего темплата

# 3D QSAR

Соединения должны быть выравнены в соответствии с предполагаемым механизмом взаимодействия с мишенью:



Даже если сайт связывания известен, соединения, взаимодействующие с рецептором, могут быть по-разному ориентированы => необходимо задать ориентацию для каждого соединения

Если структура мишени неизвестна, делается предположение о возможном механизме, исходя из данного предположения, осуществляется наложение структур для выравнивания

В большинстве случаев механизм неизвестен, предполагается, что лиганды, вне зависимости от точного химического состава, связываются с сайтом в конформациях, отражающих комплементарность стерического и электростатического полей к мишени, что соответствует фармакофорной концепции (при этом учитывая, что один и тот же сайт может взаимодействовать с несколькими фармакофорами)

# Comparative Molecular Field Analysis (CoMFA): история развития метода

- Основы подхода – Dynamic lattice-oriented molecular modeling system (DYLOMMS) [1983]

(Wise et al Progress in three-dimensional drug design: the use of real-time colour graphics and computer postulation of bioactive molecules in DYLOMMS)

- Начало активного использования метода CoMFA (1984)

(Wold et al The covariance problem in linear regression. The partial least squares approach to generalized inverses J Sci Stat Comp 1984 5 735-743)

- Основная работа по методу CoMFA (1988)

[Cramer et al Comparative molecular field analysis. Effect of shape on binding of steroids to carrier proteins, JACS 1988, 110, p.5959-5976)]

---

# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA): ОСНОВНЫЕ ШАГИ, ИСХОДНЫЕ ПРЕДПОЛОЖЕНИЯ И ОГРАНИЧЕНИЯ МЕТОДА

## Подготовка данных, разделение данных на обучающую и тестовую выборки

- Экспериментальные значения свойства измерены с сопоставимой точностью (в идеальном случае, в одной лаборатории в одинаковых условиях), диапазон величин биологической активности должен охватывать несколько порядков
- Одинаковый механизм действия лигандов (взаимодействие с одинаковым типом рецептора, фермента или ионного канала)
- Метод разрабатывался для описания непосредственно процесса взаимодействия, при возникновении нелинейных явлений, связанных с транспортом и распределением лекарств, ограниченно пригоден (рассматриваются только результаты испытаний *in vitro*)
- При отсутствии точной информации о стереохимии (например, смеси энантиомеров или диастереомеров) соединения не могут быть включены в анализ

## Генерация 3D структур и конформационный анализ

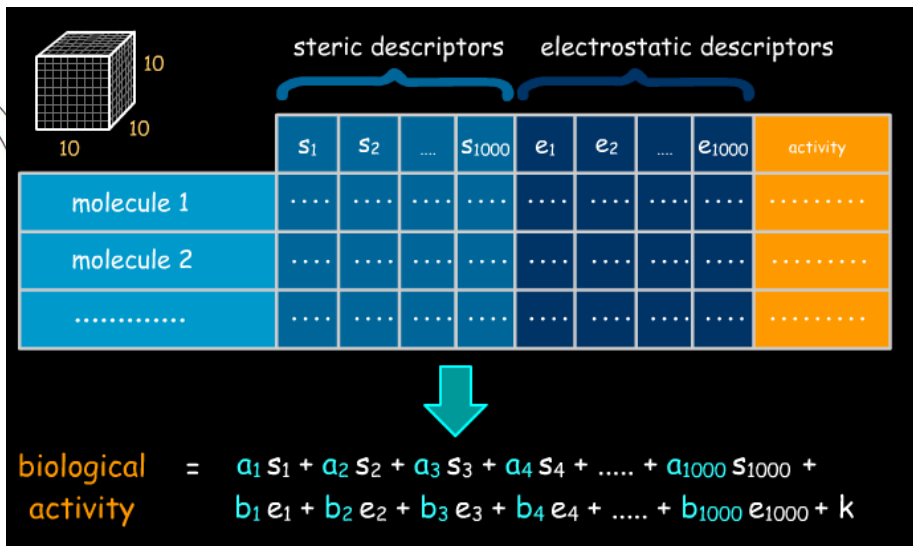
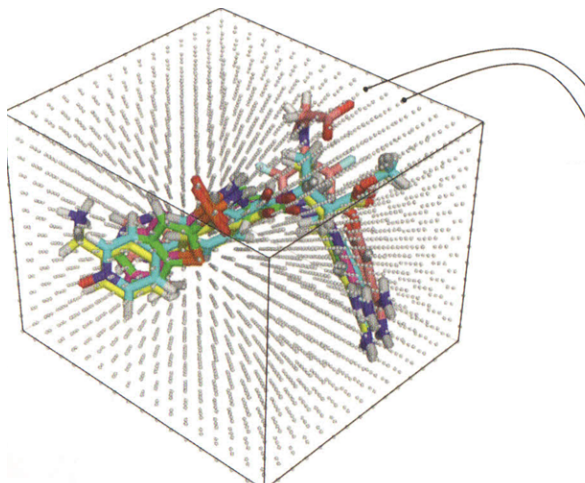
## Выравнивание молекул в соответствии с фармакофорными свойствами и свойствами поверхностей

## Расчет дескрипторов молекулярных полей

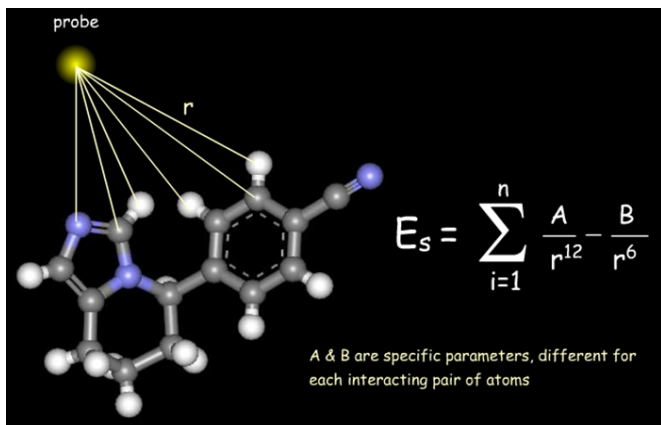
- Предполагается, что стерическое и электростатическое поля должны содержать всю информацию, необходимую для описания биологических свойств набора соединений

## Генерация и валидация моделей, прогноз новых соединений

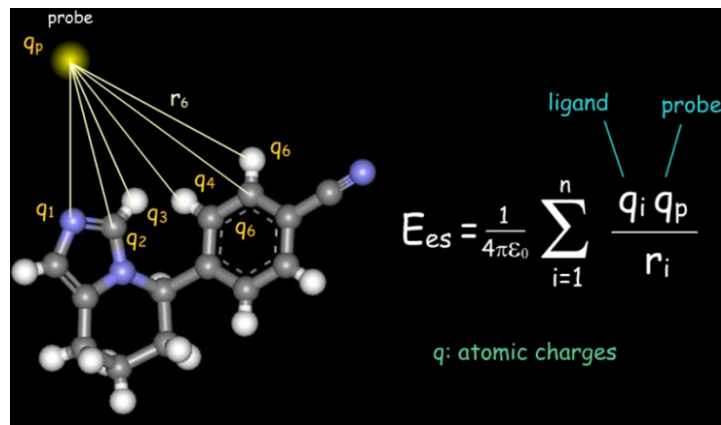
# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA)



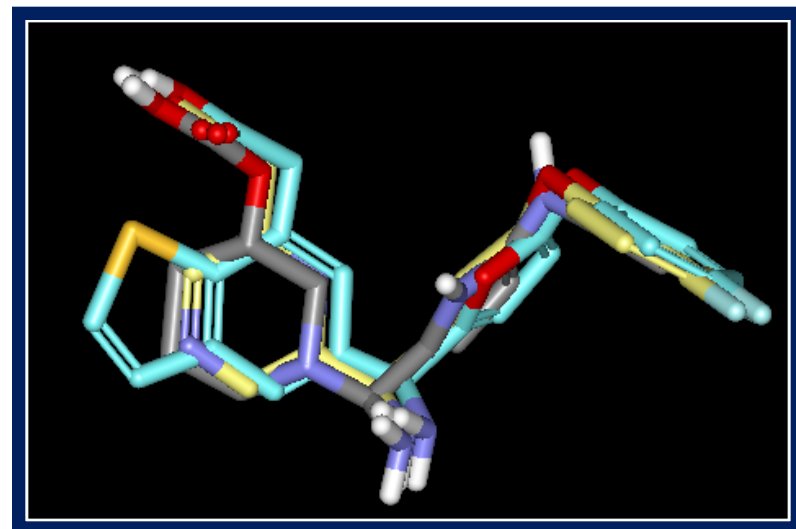
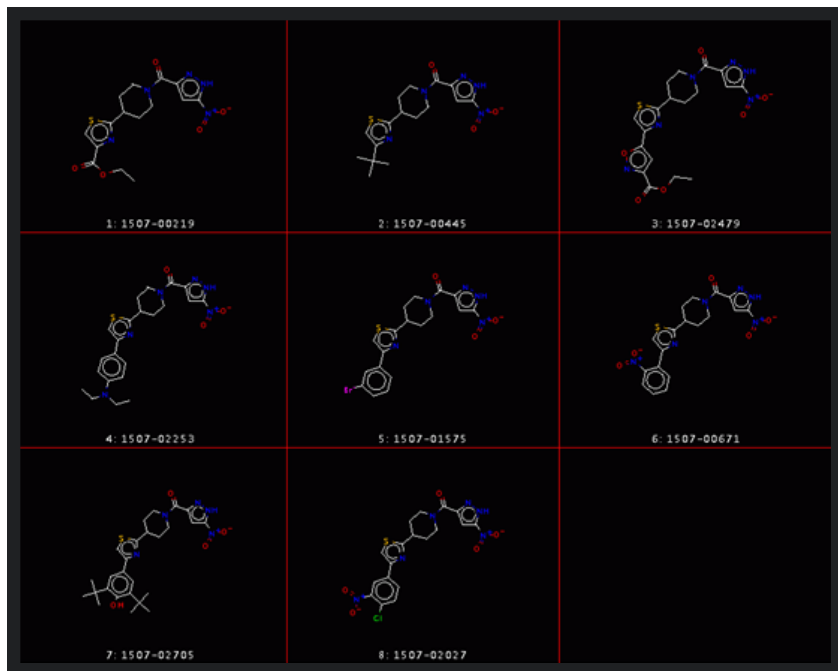
Стерическое поле (потенциал Леннард-Джонса)



Электростатическое поле (Кулоновский потенциал)



# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA): ОСНОВНЫЕ ШАГИ (1)



Установить правила ориентации при наложении молекул.

Выравнивание молекул в соответствии с фармакофорными свойствами и свойствами поверхностей

- Разделить исходный набор данных на обучающую и тестовую выборки сопоставимой разнородности
- Сгенерировать 3D структуры всех соединений набора данных



# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA): ПРОЦЕДУРА ВЫРАВНИВАНИЯ

## Совмещение "атом-на-атом"

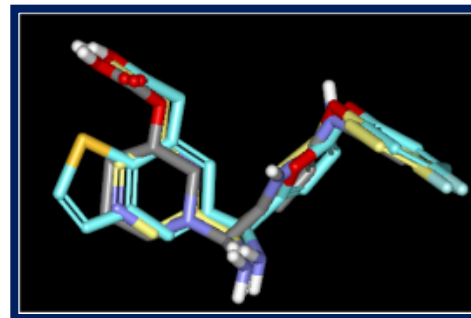
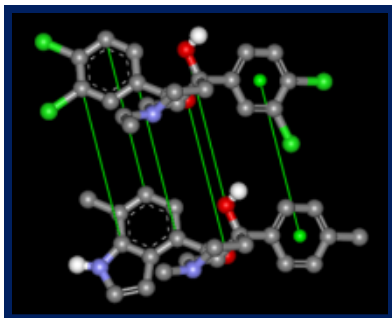
- Стандартная процедура, использующая метод наименьших квадратов
- Метод активного аналога
- SEAL

## Совмещение молекулярных полей

- Максимизация меры подобия для молекулярных поверхностей
- Вычисление молекулярных полей физико-химических свойств с использованием потенциала Леннард-Джонса или замена вычислений с помощью регулярных решеток интегрированием гауссовых функций

# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA): ПРОЦЕДУРА ВЫРАВНИВАНИЯ

## Метод наименьших квадратов



Молекулы совмещают путем минимизации среднеквадратичного отклонения расстояний между парами соответствующих атомов в этих молекулах.

Pros: Возможность обнаружения различий между соединениями, предварительно рассматривающимися как сходные

Cons: Пары совмещаемых атомов должны быть определены заранее. Метод неприменим для тех систем, где предварительно невозможно указать соответствие атомов

# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (СОМФА): ПРОЦЕДУРА ВЫРАВНИВАНИЯ

## Метод активного аналога

- Систематический поиск ансамбля стерически и энергетически разрешенных конформаций
  - Вычисление множества расстояний между фармакофорными группами для каждой из конформаций.
  - Поиск начинают с наиболее жесткой молекулы, на основе полученных результатов проводят исследования конформационного пространства для всех ее более гибких аналогов
-

# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA): ПРОЦЕДУРА ВЫРАВНИВАНИЯ

## Метод SEAL (Steric and Electrostatic ALignment)

Индекс сходства суммируется по всем возможным парам атомов в молекуле:

$$A_F = - \sum_{i=1}^m \sum_{j=1}^n w_{ij} e^{-\alpha r_{ij}^2}.$$

Каждой паре атомов соответствует статистический вес, связанный с относительными расстояниями между этими атомами:

$w$  – коэффициент, регулирующий влияние стерического и электростатического эффектов

$$w_{ij} = w_E q_i q_j + w_S v_i v_j$$

$\alpha$  – коэффициент затухания (обычно 0,2)

$r$  – расстояние между атомами

$q$  – заряд

$v$  – функция атомного радиуса ван дер Ваальса

Так как в процессе сравнения учитываются все возможные пары атомов, при итоговом совмещении учитываются свойства, связанные с итоговой формой молекулы

Выравнивание осуществляется случайным изменением положения одного соединения относительно другого и минимизацией функции выравнивания

# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (СОМФА): ПРОЦЕДУРА ВЫРАВНИВАНИЯ

- **Максимизация меры подобия для молекулярных поверхностей**

M. Manaut et al *Automatic search for maximum similarity between molecular electrostatic fields* JCAMD 5: 371–380, 1991

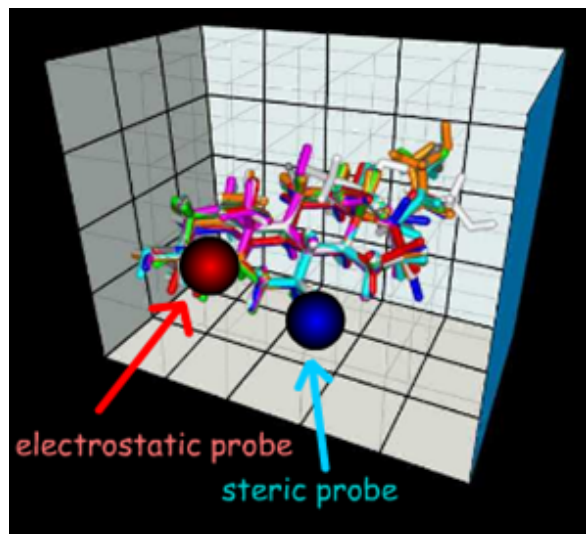
C. Lemmen et al *Computational methods for the structural alignment of molecules* JCAMD, 14: 215–232, 2000

- **Вычисление молекулярных полей физико-химических свойств с использованием потенциала Леннарда-Джонса или замена вычислений с помощью регулярных решеток интегрированием гауссовых функций**

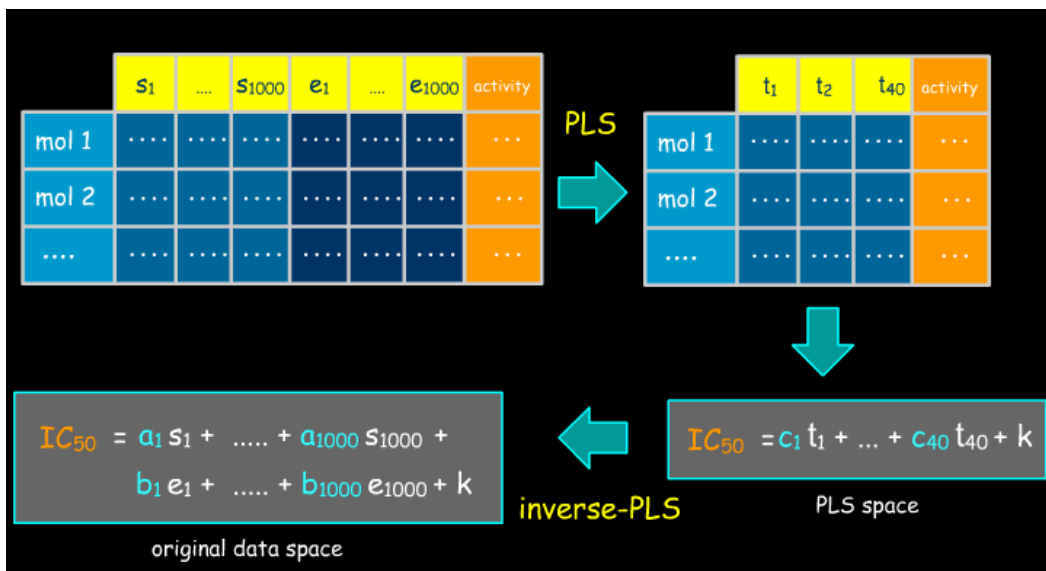
Clark, M. et al *Tetrahedron Comput. Methodol.* 3,47-59 (1990).

Dean, F M. *Molecular recognition: The measurement and search for molecular similarity in ligand-receptor interaction. In: Concepts and Applications of Molecular Similarity. Johnson, M. A., and Maggiora, G. M. (Eds.). Wiley: New York; 211-238 (1990).*

# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA): ОСНОВНЫЕ ШАГИ (2)



- Поместить молекулы в кубическую решетку, которая также обеспечивает достаточный объем вокруг молекул (4Å в каждом направлении от всех молекул, расстояние между узлами решетки 1-2 Å)
- Значения полей рассчитываются для каждого соединения, располагая в каждом узле гипотетической решетки «пробные» атомы.



$$\text{Activity} = c + \sum_{i=1}^N a_i s_i + \sum_{i=1}^N b_i e_i$$

$s_i$  = steric field at the  $i^{\text{th}}$  grid point

$e_i$  = electrostatic field at the  $i^{\text{th}}$  grid point

$a_i$  &  $b_i$  = regression weights

# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA): ОСНОВНЫЕ ШАГИ (3)

## Principal Components Regression (PCR)

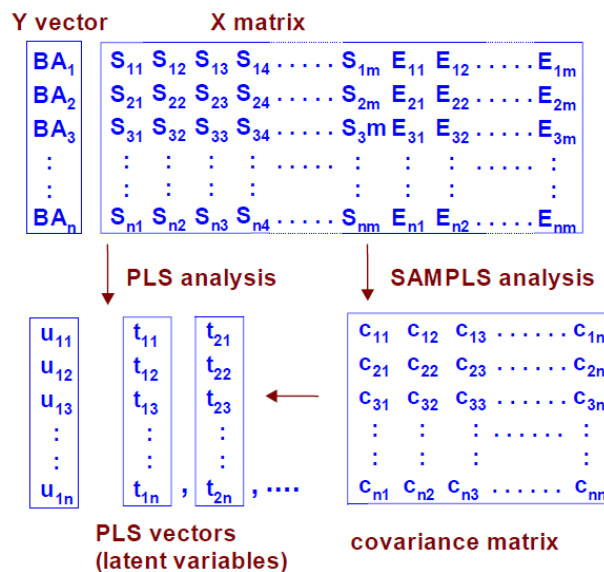
Extract latent variables for **predictors** that explain most of the **predictor variation**  
 $\rightarrow X^T X$  decomposition

## Maximum Redundancy Analysis (MRA, van Wollenberg 1977)

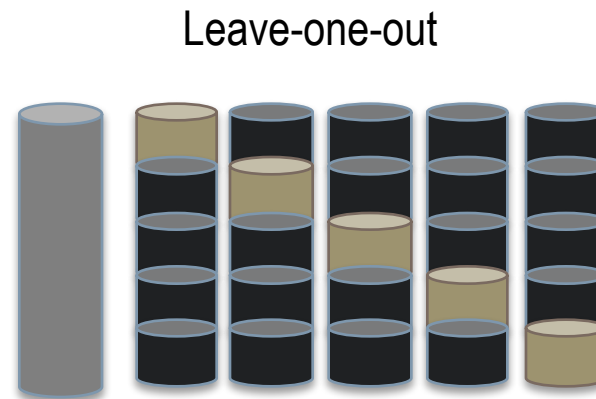
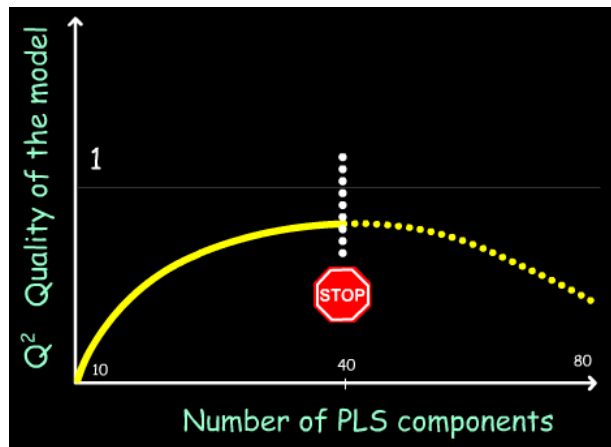
Extract latent variables for **responses** that explain most of the **response variation**  
 $\rightarrow \hat{Y}^T \hat{Y}$  decomposition

## Partial Least Squares/Projection to Latent Structures (PLS, Wold 1960)

$\rightarrow$  „Seek directions in the feature space associated with high variation in the responses but bias them towards accurately predictable directions“  
 $\rightarrow X^T Y$  decomposition



# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA): СТАТИСТИЧЕСКОЕ КАЧЕСТВО МОДЕЛЕЙ



$$\text{Activity} = c + \sum_{i=1}^N a_i s_i + \sum_{i=1}^N b_i e_i$$

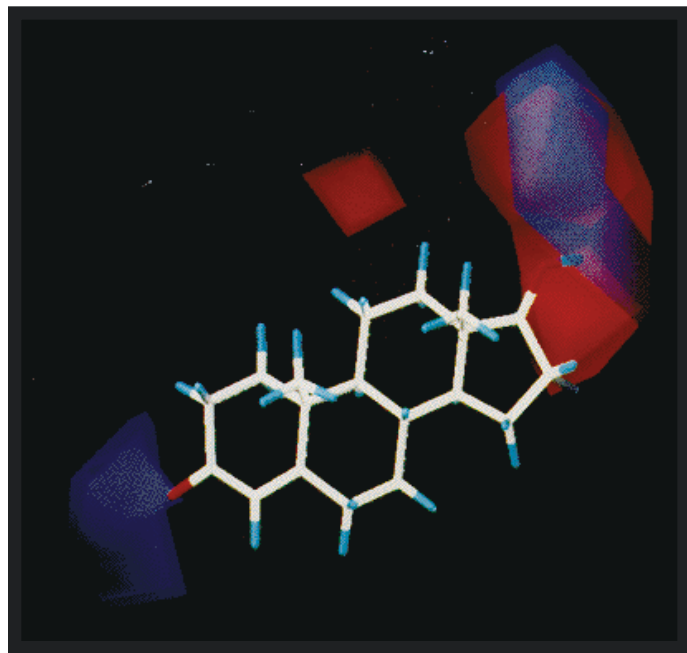
$s_i$  = steric field at the  $i^{\text{th}}$  grid point

$e_i$  = electrostatic field at the  $i^{\text{th}}$  grid point

$a_i$  &  $b_i$  = regression weights

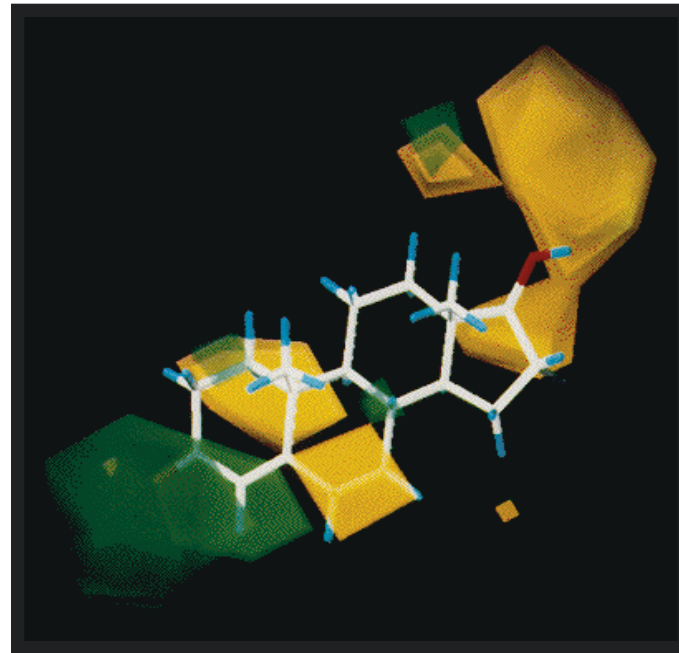


# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA): ГРАФИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ



Electrostatic Contour Map

Blue - favorable electronegative  
Red - the unfavorable regions



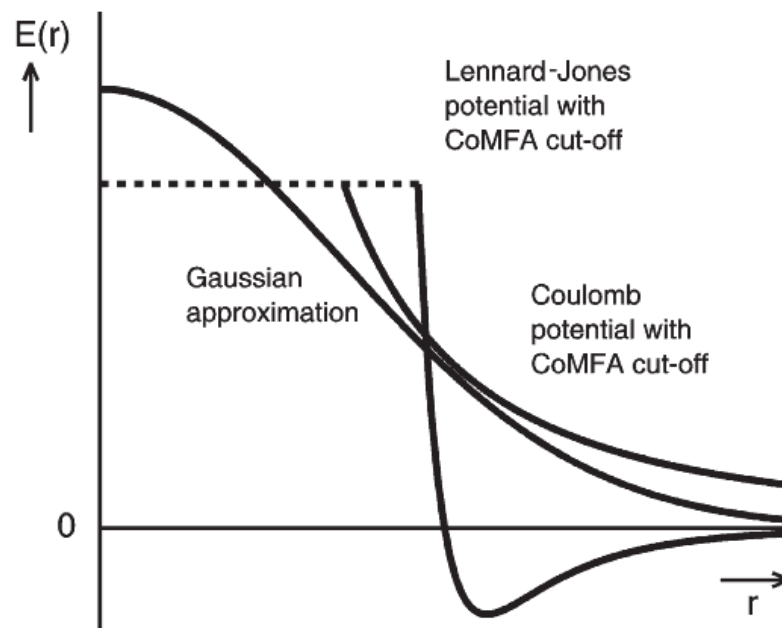
Steric Contour Map

Green - favorable regions  
Yellow – unfavorable regions

# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА МОЛЕКУЛЯРНЫХ ПОЛЕЙ (COMFA): ОГРАНИЧЕНИЯ МЕТОДА

## Ограничения метода:

- Большое количество настраиваемых параметров (ориентация, расположение решетки, размер шага, тип пробного атома и т.д.)
- При использовании отбора переменных фрагментированные контурные карты
- Отсутствие прямого учета гидрофобности
- Необходимость использования искусственно введенных пороговых значений функций
- Недооценка вклада водородных связей
- Статистическая проблема, связанная с большим количеством переменных в модели
- Необходимость использовать данные *in vitro*

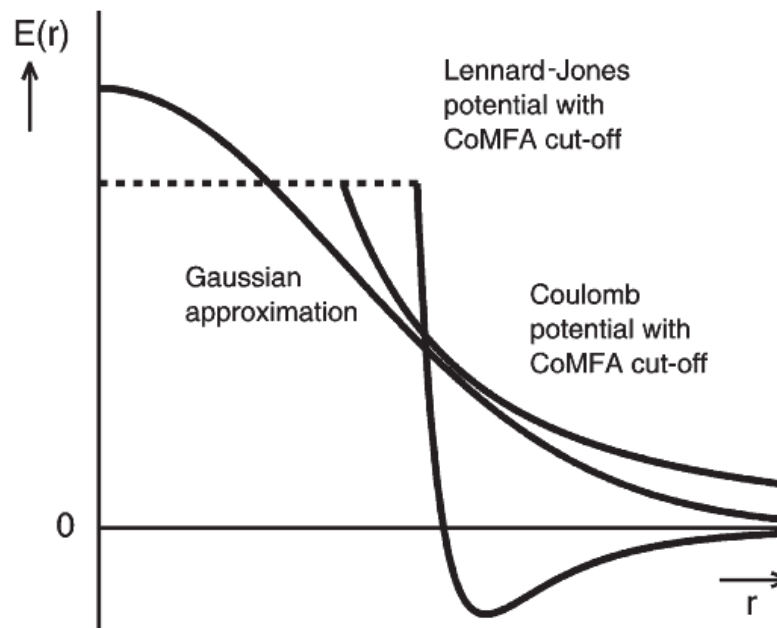


# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА ИНДЕКСОВ МОЛЕКУЛЯРНОГО ПОДОБИЯ (COMSIA)

Вместо энергий взаимодействия рассчитываются индексы сходства:

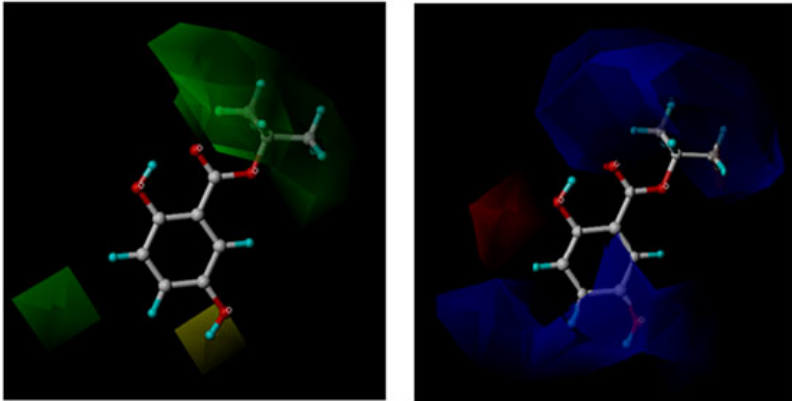
$$A_{F,k}^q(j) = - \sum_i (W_{\text{probe},k} \cdot W_{ik}) \exp(-\alpha r_{iq}^2)$$

$W_{ik}$  значение свойства  $k$  для атома  $i$  (стерический объем, AM1 заряды, эмпирические правила для учета влияния водородных связей, параметры, учитывающие гидрофобность)

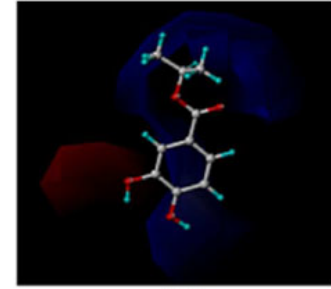


# МЕТОД СРАВНИТЕЛЬНОГО АНАЛИЗА ИНДЕКСОВ МОЛЕКУЛЯРНОГО ПОДОБИЯ (COMSIA)

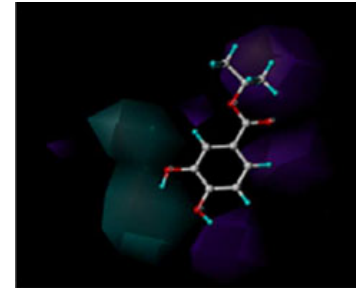
## Conventional CoMFA



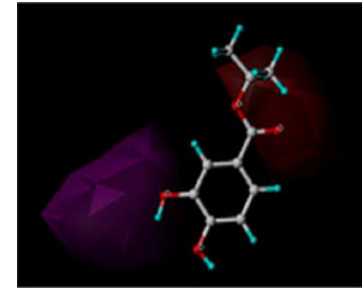
Steric map (green, bulky group desirable; yellow, bulky group undesirable)



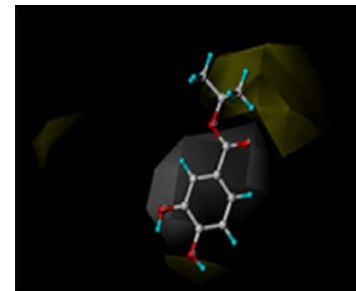
Electrostatic map (blue, electropositive group desirable; red, electronegative group desirable)



Donor map (cyan, donor group desirable; purple, donor group undesirable)



Acceptor map (magenta, acceptor group desirable; red, acceptor group undesirable)



Hydrophobic map (yellow, hydrophobic group desirable; white, hydrophilic group desirable)

# МЕТОД GRID/GOLPE

## GRID

- Большое количество «пробных» атомов: амино (NH<sub>3</sub><sup>+</sup>), карбоксильный и карбонильный кислород, гидроксильная и метильная группы, вода; если пробный атом образован группой, то эта группа рассматривается единым элементом с адаптированным радиусом
- В отличие от CoMFA рассматривается общий потенциал взаимодействия:

$$E_{\text{int}} = \sum E_{\text{lj}} + \sum E_{\text{el}} + \sum E_{\text{hb}}$$

Прародитель методов GRIND и VolSurf

## GOLPE

Процедура, включающая различные методы отбора переменных (метод D-оптимального предварительного отбора переменных, Fractional Factorial Design)

## ДРУГИЕ МЕТОДЫ

TOPOMER CoMFA Метод топомерного сравнительного анализа молекулярных полей  
Topomer CoMFA: A Design Methodology for Rapid Lead Optimization J. Med. Chem. 2003, 46, 374-388

TEMPLATE CoMFA (Метод темплатного сравнительного анализа молекулярных полей)  
Template CoMFA: The 3D-QSAR Grail? R.Cramer et al J. Chem. Inf. Model. 2014, 54, 660-671

FLUFF (FLUFF-BALL)  
FLUFF-BALL, A Template-Based Grid-Independent Superposition and QSAR Technique: Validation Using a Benchmark Steroid Data Set S. Korhonen et al J Chem Inf Comp Sci 2003 43 (6), 1780-1793

SOMFA (Self-Organizing Molecular Field Analysis)  
Self-Organizing Molecular Field Analysis: A Tool for Structure-Activity Studies D. Robinson et al J Med Chem 1999 42 (4), 573-583

SAMFA (Simple Atom Mapping Following Alignment)  
SAMFA: Simplifying Molecular Description for 3D-QSAR J Manchester et al JCIM 2008 48 (6), 1167-1173

---

# МЕТОДЫ 3D QSAR, НЕ ЗАВИСЯЩИЕ ОТ ВЫРАВНИВАНИЯ



# GRID-INDEPENDENT DESCRIPTORS (GRIND): ИСХОДНЫЕ ТРЕБОВАНИЯ

- Независимость дескрипторов от положения и ориентации молекулы в пространстве
  - Возможность химической интерпретации дескрипторов
  - Максимальная информативность
  - Возможность быстрого и автоматизированного расчета
-



# 3D QSAR МЕТОДЫ, НЕ ЗАВИСЯЩИЕ ОТ ВЫРАВНИВАНИЯ: GRID-INDEPENDENT DESCRIPTORS (GRIND)

## Основные этапы метода

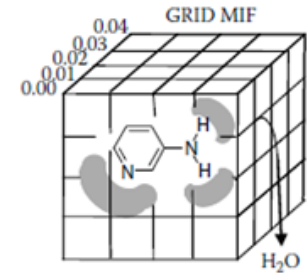
### Расчет молекулярных полей взаимодействия (метод чувствителен к заданной сетке):

Три типа пробных атомов:

DRY (для оценки гидрофобных взаимодействий),

Карбонильный кислород (акцептор водородной связи)

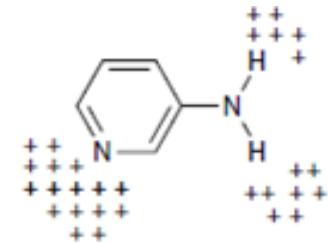
Амидогруппа (донор водородной связи).



### Фильтрация:

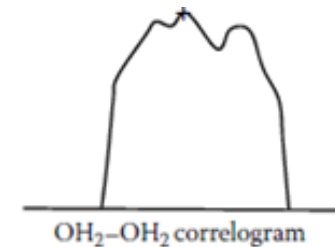
Выделение областей потенциально наиболее выраженного взаимодействия лиганда с сайтом связывания: (i) значения полей должны быть отрицательными (выгодными для взаимодействия) и выраженными и (ii) расстояние между двумя областями должно быть максимально возможным.

Используется алгоритм двухзадачной оптимизации, (учет значения поля в вершине и взаимное расстояние между вершинами).

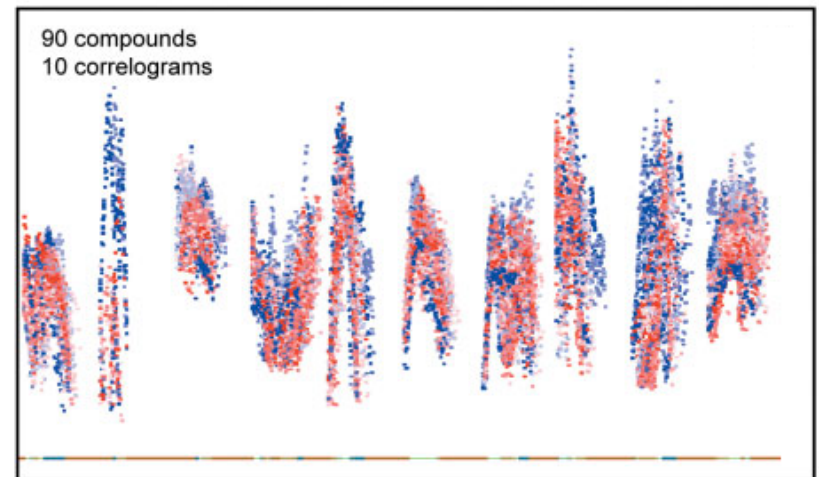
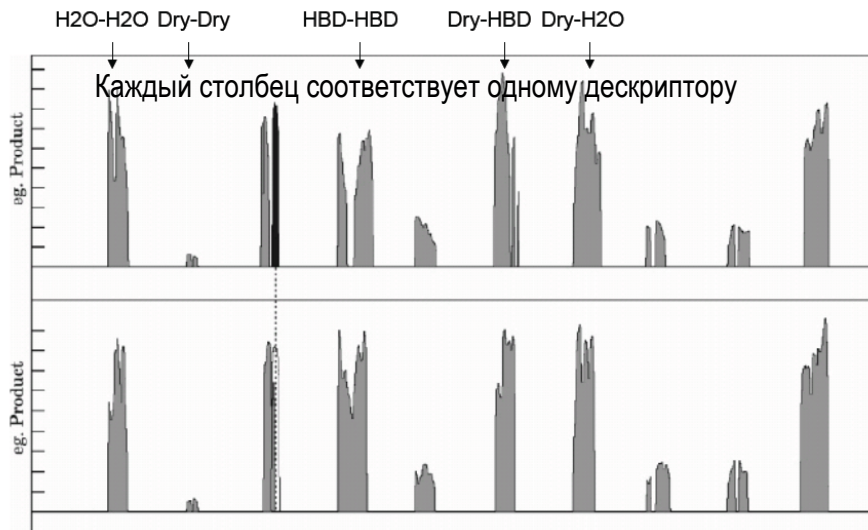
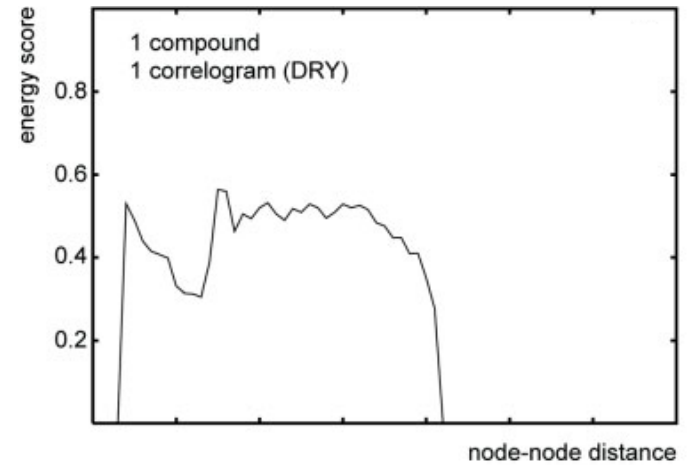
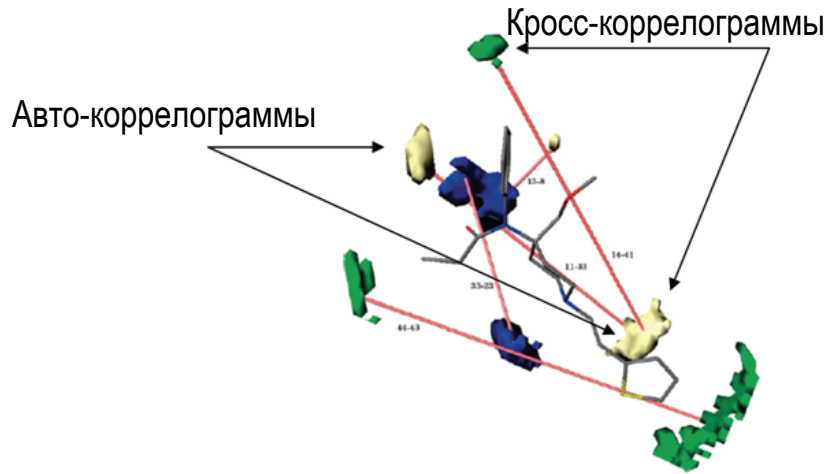


### Кодирование:

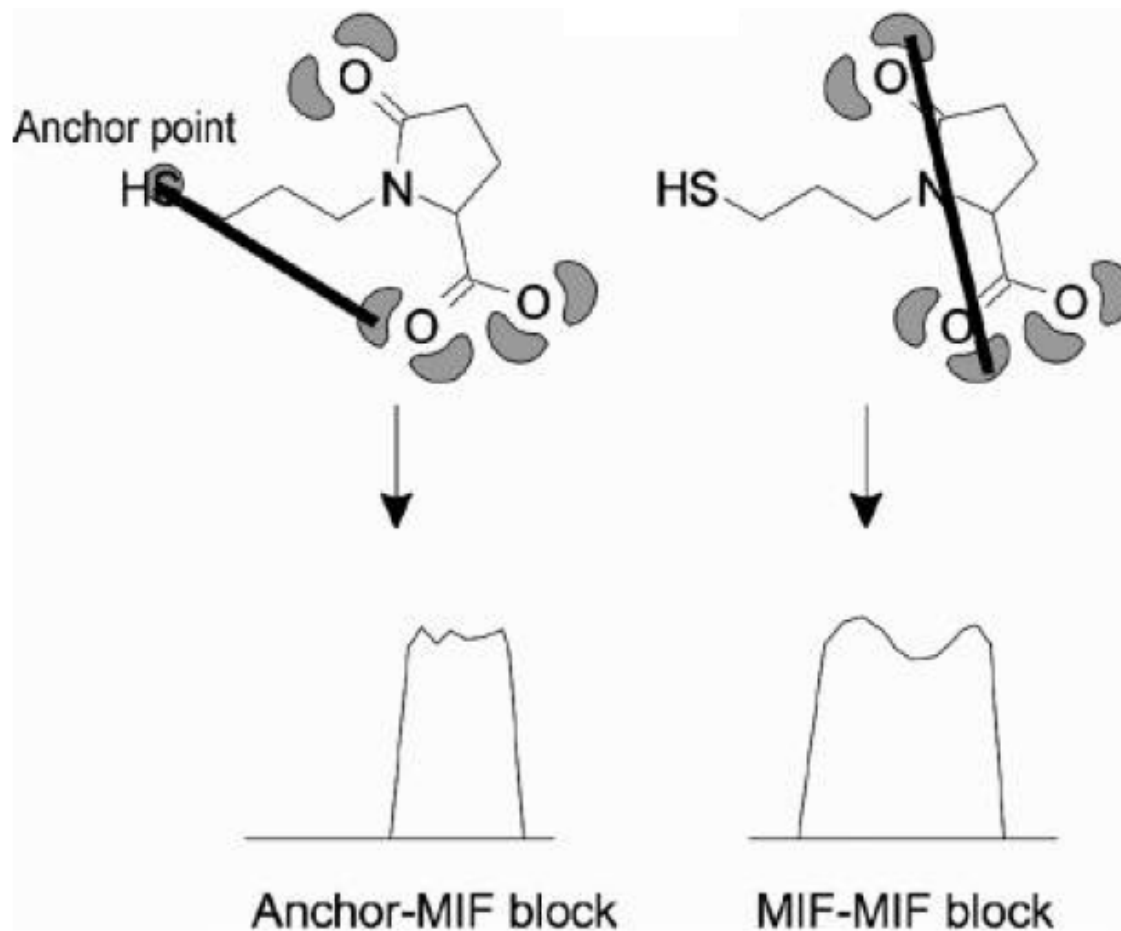
Пространственное расположение описывается при помощи системы, основанной на расстоянии между точками-зондами, соответствующими областям максимально выраженного взаимодействия: определяется дискретный набор фиксированных расстояний («бинов»), каждая пара выделенных значимых областей относится к одному из определенных «бинов», в конце анализа каждый бин представлен парой с максимальным произведением значений молекулярных полей, полученные данные шкалируются на максимальное значение



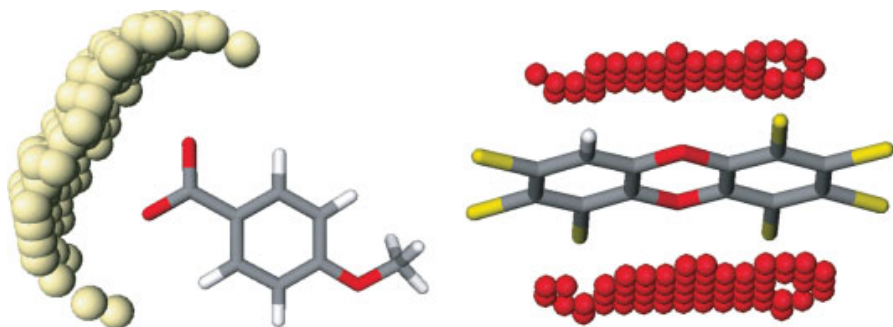
# GRID-INDEPENDENT DESCRIPTORS (GRIND): ГРАФИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ



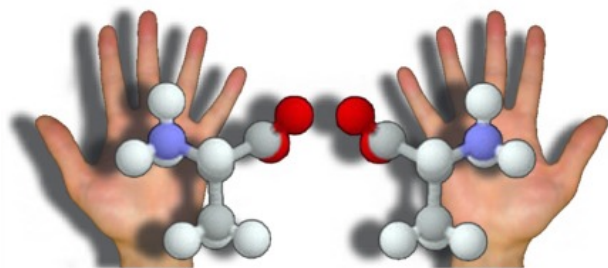
# 3D QSAR МЕТОДЫ, НЕ ЗАВИСЯЩИЕ ОТ ВЫРАВНИВАНИЯ: ANCHOR GRID-INDEPENDENT DESCRIPTORS (GRIND)



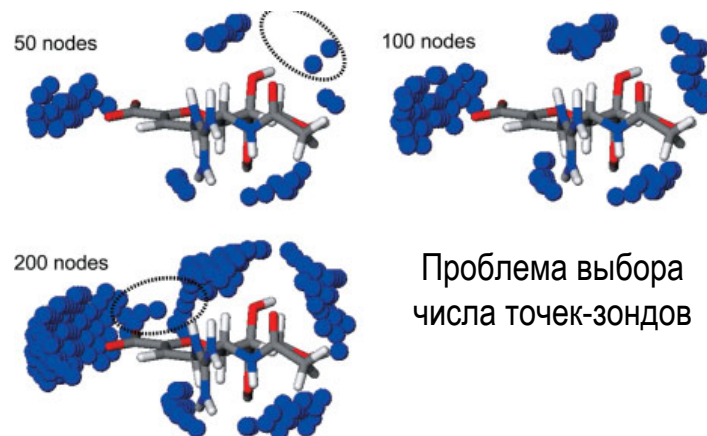
# GRID-INDEPENDENT DESCRIPTORS (GRIND): НЕДОСТАТКИ МЕТОДА



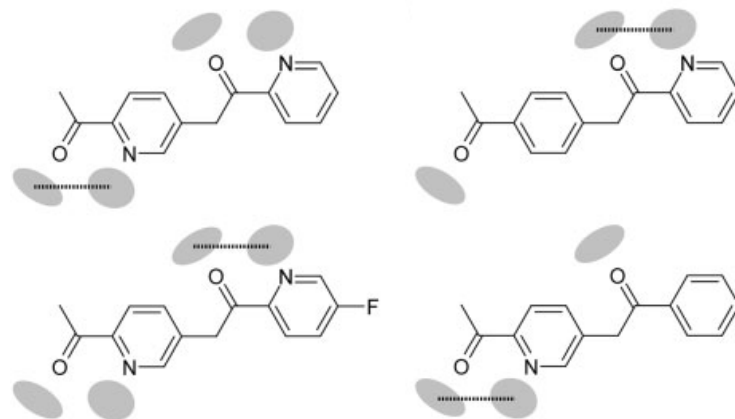
- Заряженные группы могут сосредоточить вокруг себя все точки-зонды
- Присутствие нескольких заместителей с неявно выраженными значениями энергии приводит к малоинформативным результатам



Отсутствие учета хиральности молекул: идентичное описание энантиомеров, диастереомеры могут быть представлены различными коррелограммами



Проблема выбора числа точек-зондов



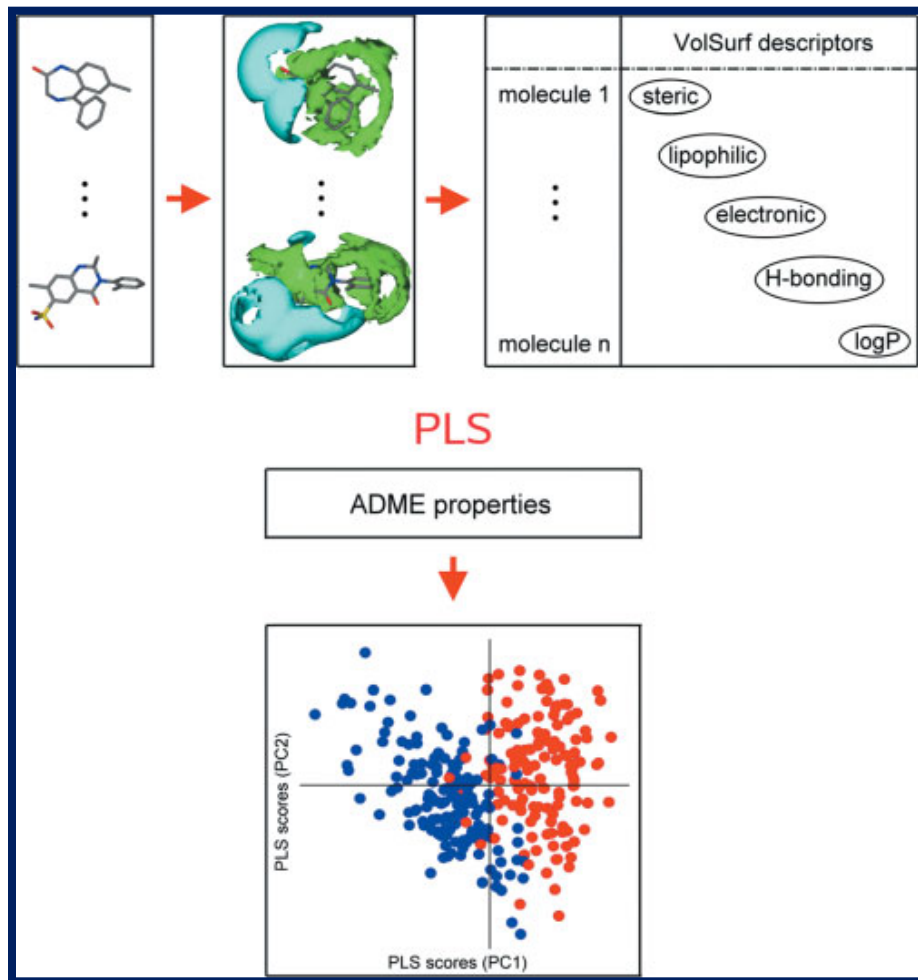
Присутствие нескольких альтернативных сайтов связывания во всех лигандах или части набора данных

# GRID-INDEPENDENT DESCRIPTORS (GRIND)

- ❖ В чем различия между дескрипторами GRIND и основанными на фармакофорном описании данных?
  - ❖ Назовите основные ограничения метода
-

# VOLSURF

Описание формы, полярности, гидрофобности посредством преобразования 3D карт, состоящих из отдельных «вокселей» (соотв. элем. ячейкам), в конкретные числовые значения дескрипторов



## Расчет молекулярных полей взаимодействия

Типы пробных атомов:

- DRY (для оценки гидрофобных взаимодействий),
- Карбонильный кислород (акцептор водородной связи)
- OH2 probe
- ...



## Расчет дескрипторов

# VOLSURF

## 100 дескрипторов

### Дескрипторы размера и формы

- Молекулярный объем
- Молекулярная поверхность
- Отношение объем/поверхность
- Молекулярная сферичность

### Дескрипторы гидрофильных областей

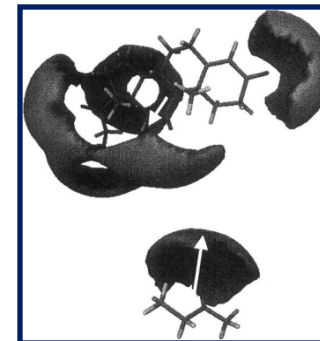
- Гидрофильные дескрипторы
- Факторы вместимости (отношение площади гидрофильных областей к общей площади молекулярной поверхности)

### Дескрипторы гидрофобных областей

### Моменты энергий взаимодействия (Interaction Energy Moments)

отражают дисбаланс между центром масс и барицентром гидрофильных областей

GRID 3D молекулярные поля кортизона (верх) и бутанона (низ) рассчитанные с пробным атомом воды. Выделены гидрофильные области. Стрелки – соответствуют векторам моментов энергий взаимодействий



### Смешанный тип

Локальный минимум энергии взаимодействия  
Расстояния между минимумами энергии  
Гидрофильно-гидрофобный баланс  
Амфифильный момент  
Поляризуемость

# 3D QSAR МЕТОДЫ, НЕ ЗАВИСЯЩИЕ ОТ ВЫРАВНИВАНИЯ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОЛЕКУЛЯРНЫХ МОМЕНТОВ (СОММА)

Может рассматриваться как глобальный подход, который может дать представление, как лиганд отреагирует на приложенный момент или однородное или градиентное электростатическое поле при взаимодействии с рецептором.

В качестве структурных дескрипторов используются молекулярные моменты массы и распределения заряда

Форма характеризуется с помощью главных моментов инерции:  $I_x, I_y, I_z$

$$\mathbf{I} = \sum_i m_i (\mathbf{1} r_i^2 - \vec{r}_i \vec{r}_i)$$

Распределение заряда характеризуется на основе величин дипольного и главного квадрупольного моментов:  $p$  и  $Q$

$$\vec{p} = \int d^3x \vec{x} \rho(\vec{x}), \quad \mathbf{Q} = 1/2 \int d^3x (3\vec{x}\vec{x} - |\vec{x}|^2 \mathbf{1}) \rho(\vec{x})$$

Дескрипторы, устанавливающие взаимосвязь между формой молекулы и распределением заряда:

$p_x, p_y, p_z, d_x, d_y, d_z$  (величины дипольных компонент, величины смещения между центром диполя и центром масс)

$Q_{xx}, Q_{yy}$  (квадрупольные компоненты)



Не зависят от ориентации и положения молекул в пространстве



Для симметричных молекул могут принимать значения  $\infty$

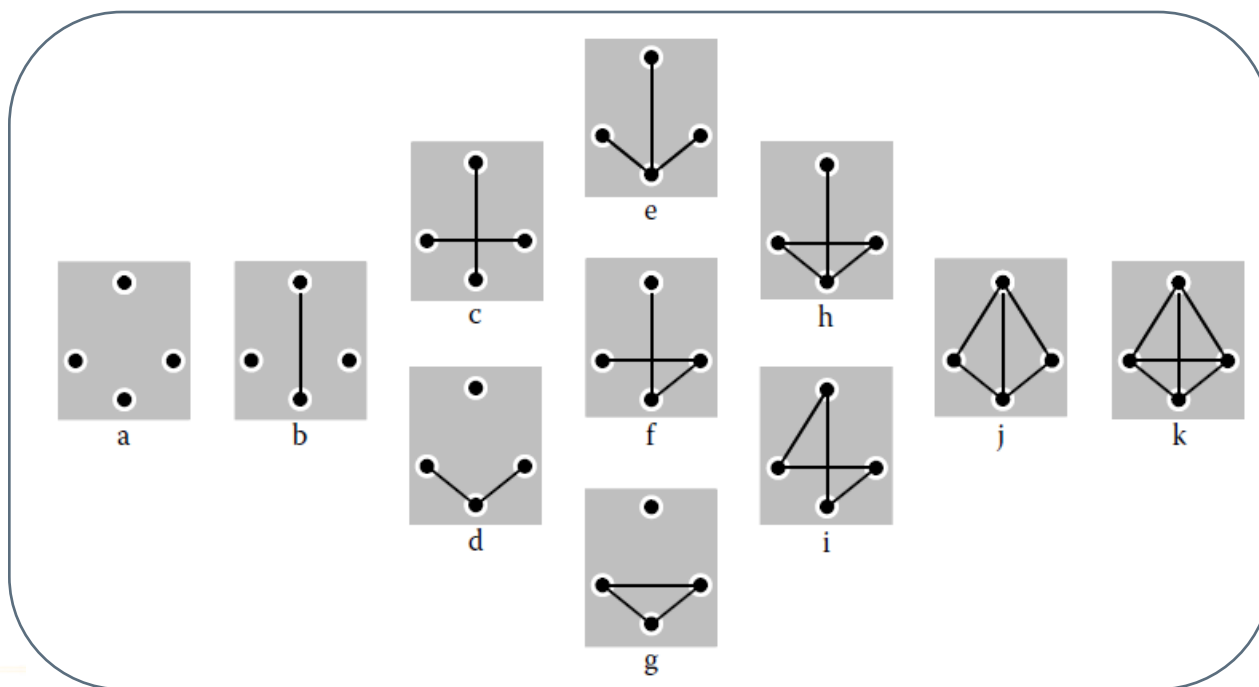


# SIMPLEX (SiRMS) APPROACH

Симплексы - фрагменты (4-х атомные) фиксированного типа, хиральности и симметрии.

Для каждого соединения, конформеры описываются ансамблем симплексов 11 базовых типов. Атомы обозначаются в соответствии с одним из заданных свойств: тип атомов, частичный заряд, липофильность, рефракция, способность атома быть донором и/или акцептором водородной связи

Молекулярные дескрипторы рассчитываются как сумма значений структурных параметров для каждого конформера, взвешенные на его вероятность



# ДРУГИЕ 3D QSAR МЕТОДЫ



# ДРУГИЕ 3D QSAR МЕТОДЫ

Основанные на расчете свойств в дискретных точках пространства, расположенных на общей поверхности активных лигандов или рядом с ней, модели псевдорецепторов

- GERM
- COMPASS
- QUASAR
- Модель поверхности рецептора (Receptor Surface Model)

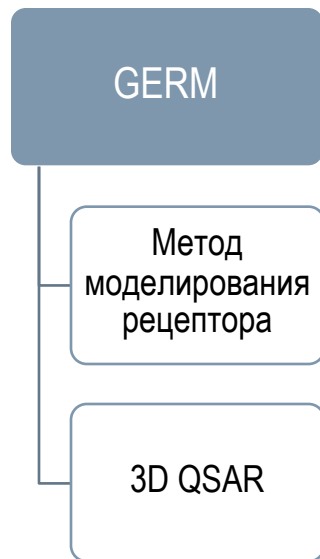
Методы, основанные на информации о рецепторе

- Метод адаптации полей для сравнения молекул (Adaptation of Fields for Molecular Comparison)

Методы, основанные на непрерывных полях

- Метод непрерывных полей (Continuous Molecular Fields)
- Метод индикаторных полей (Continuous Indicator Fields)

# GENETICALLY EVOLVED RECEPTOR MODEL (GERM)



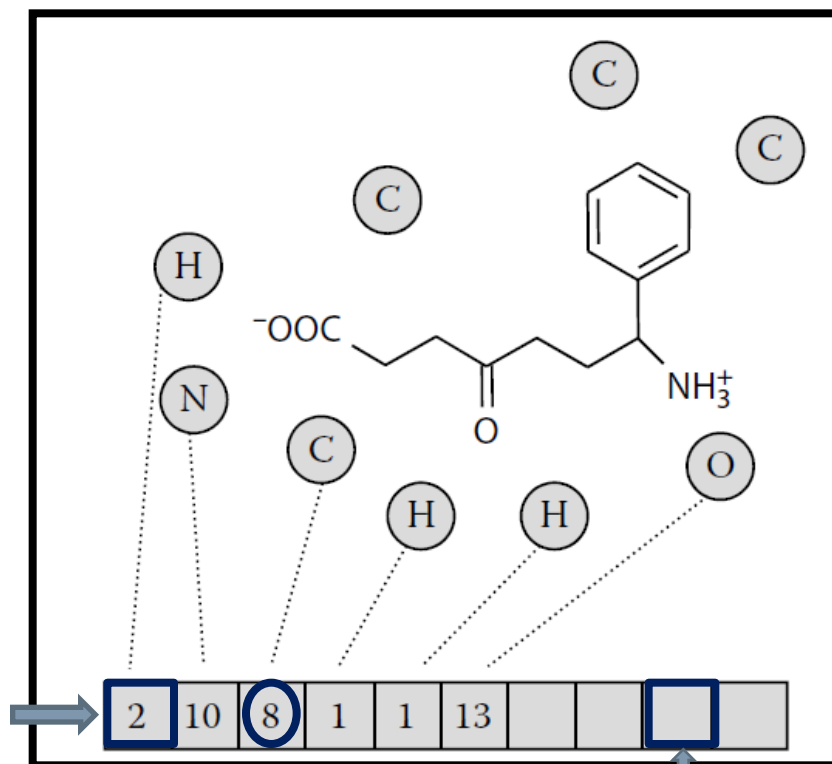
## Исходные требования:

SAR серия, для которой определены «приемлемые» выравнивания «подходящих» конформеров

## Предположения и ограничения:

- один лиганд – один конформер
- Значение биологической активности пропорционально энергии лиганд-рецепторного взаимодействия
- Влияние деформаций лиганда или гибкости рецептора не учитываются

# GENETICALLY EVOLVED RECEPTOR MODEL (GERM)



Атомы - «зонды» равномерно распределяются по поверхности сферы, окружающей лиганд, атом углерода помещается в каждую из точек, позиция корректируется для обеспечения наилучших ван-дер-Ваальсовых взаимодействий с лигандами

15 типов атомов  
40 – 60 атомов- «зондов»



Генетический алгоритм

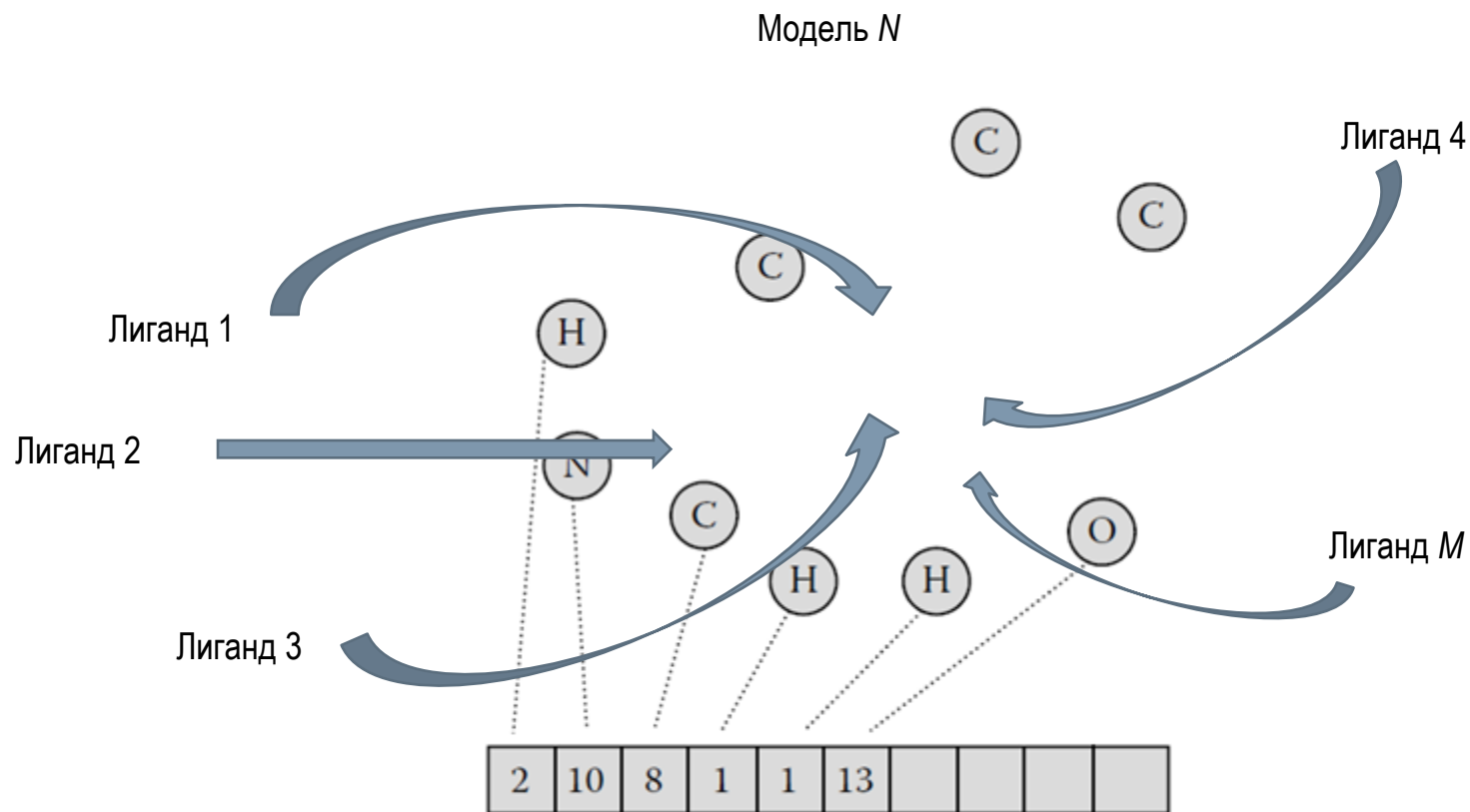
Положение в строке – положение в пространстве

Цифровое значение - тип атома-зонда

«Пустые» позиции ( возможность учета молекул растворителя)

# GENETICALLY EVOLVED RECEPTOR MODEL (GERM)

- ВДВ и электростатические энергии взаимодействия
- Корреляция между  $\log$  (значения свойства) и рассчитанной энергией взаимодействия



# COMPASS

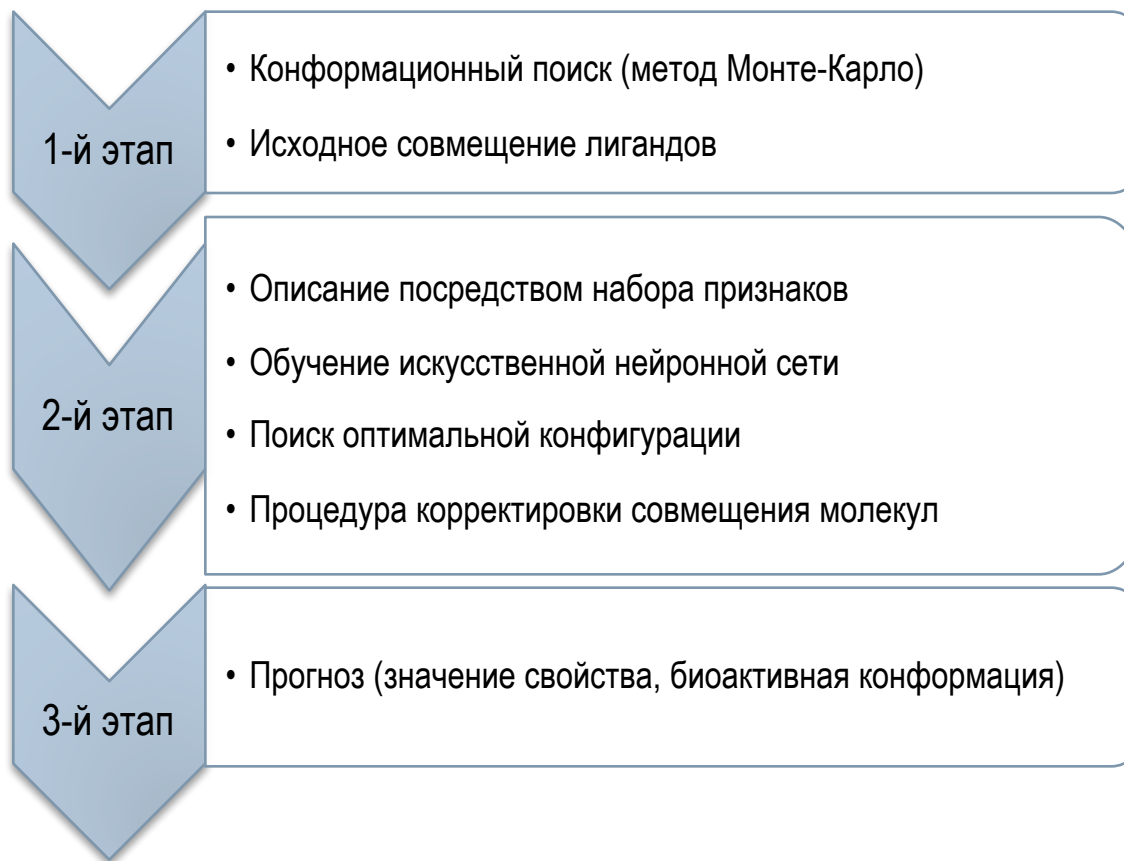
## Отличительные особенности метода:

- Акцент на роли молекулярной поверхности (не исследует «внутренние» зоны)
- Построение моделей при помощи искусственных нейронных сетей
- Автоматический поиск конформаций и процедура совмещения молекул

## Преимущество перед методами, основанными на выравнивании:

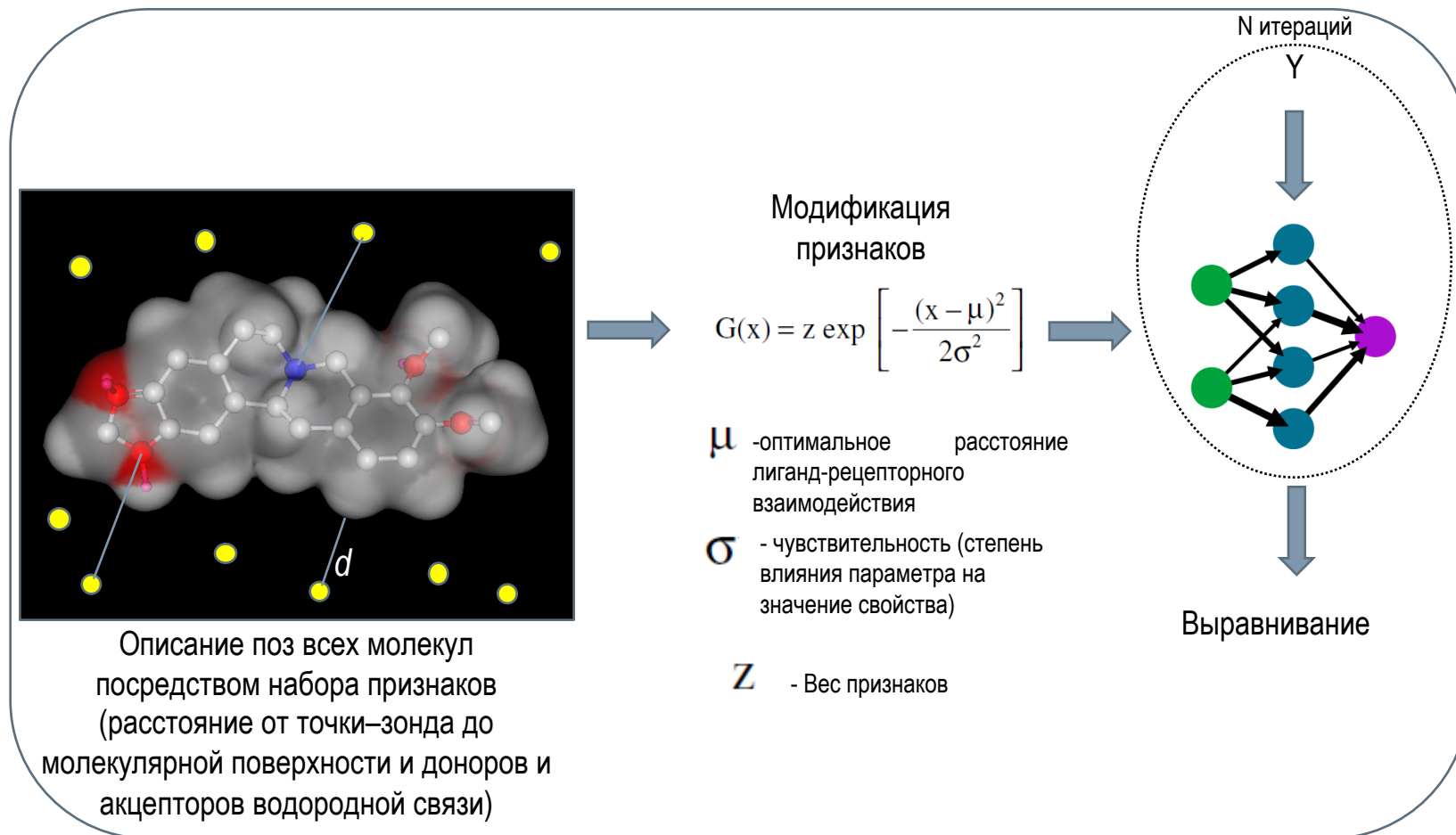
- Более эффективный учет стерического соответствия
- Меньшая вероятность переобучения модели

# COMPASS: основные этапы





# COMPASS: 2-й этап



0-5 раз