

Лекция 10

Хемоинформатика и виртуальный скрининг

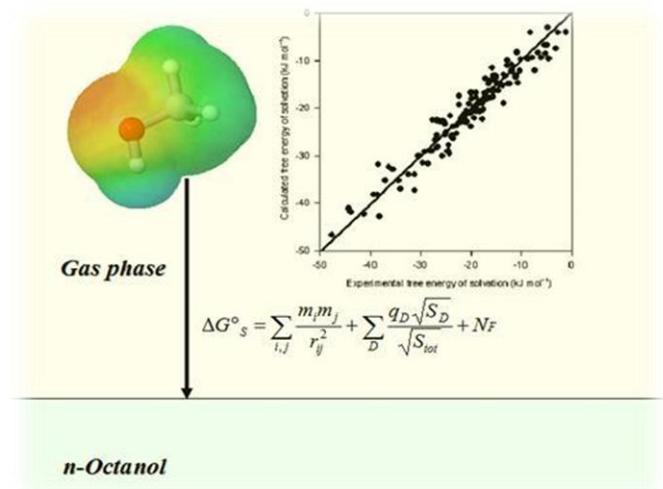
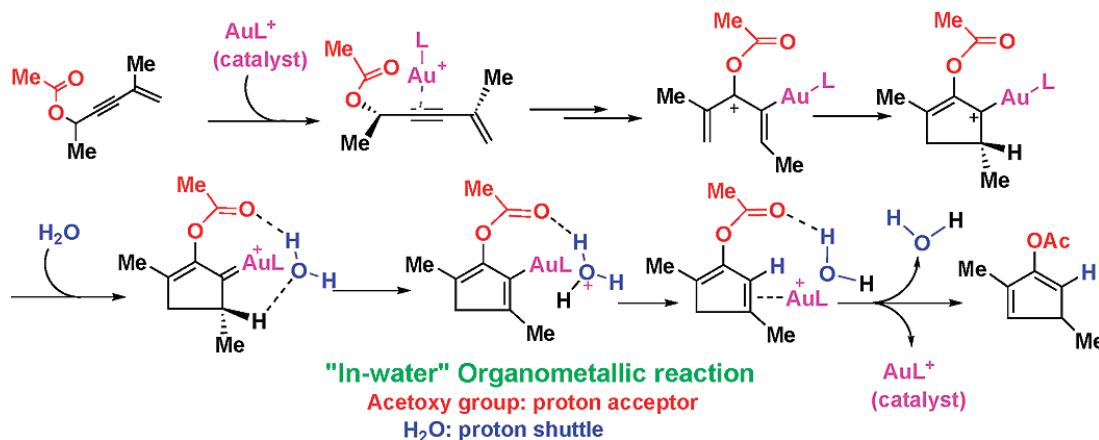
апрель 2018

Хемоинформатика

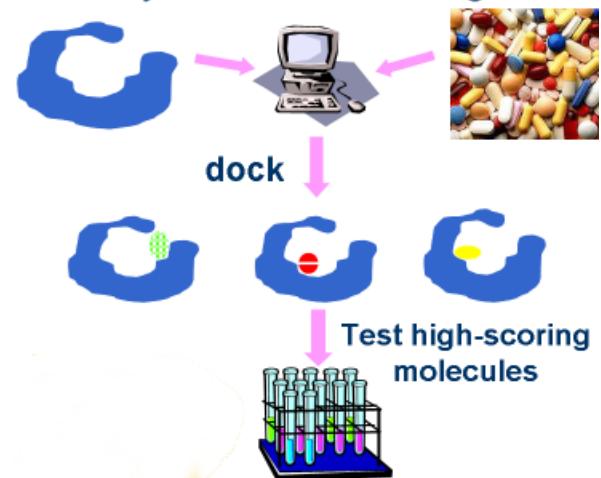
- это применение методов информатики для решения химических задач.

Область применения:

- Предсказание свойств химических соединений (QSPR)
- Поиск по химическому подобию, фармакофорный поиск, виртуальный скрининг
- Компьютерный синтез



Screening for Novel Inhibitors
by Molecular Docking



Хемоинформатика и все-все-все...



Представление структуры молекул

Молекулярный граф – связный неориентированный граф, находящийся во взаимно-однозначном соответствии со структурной формулой химического соединения таким образом, что вершинам графа соответствуют атомы молекулы, а рёбрам графа — химические связи между этими атомами.

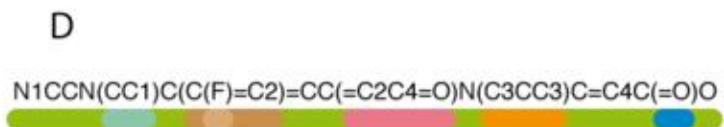
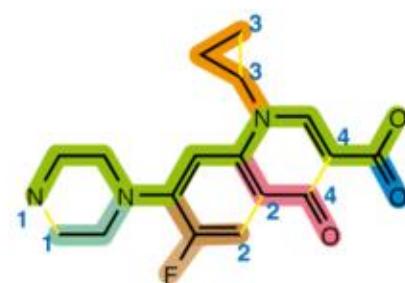
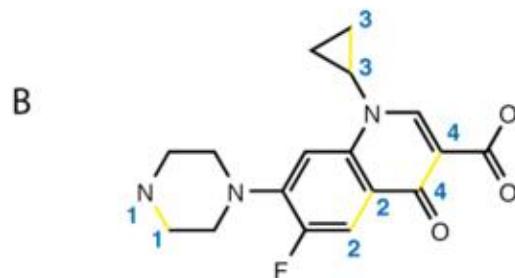
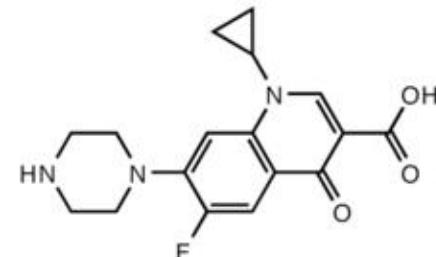
Способы записи:

- Линейные нотации (SMILES, SMARTS, SLN, InChI)
- Матрица смежности
- Структурные файлы (общие – MOL, SDF,... специальные - MOL2, HIN,...)
- Chemical Markup Language

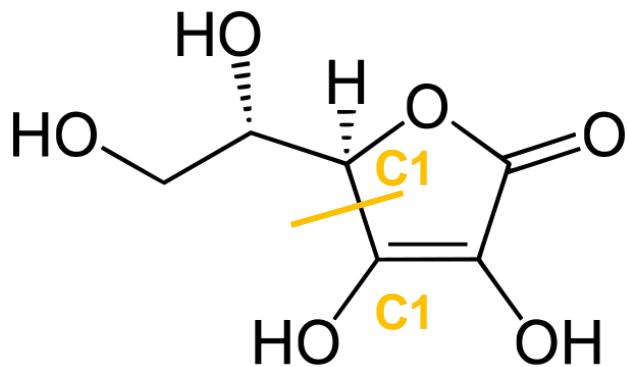
Линейные нотации. SMILES

SMILES (англ. *Simplified Molecular Input Line Entry Specification*, спецификация упрощенного представления молекул в строке ввода) — система правил (спецификация) однозначного описания состава и структуры молекулы химического вещества с использованием строки символов ASCII.

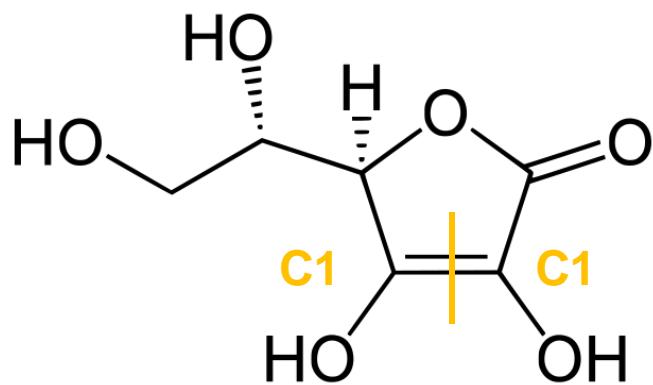
| | |
|-------------------|----------|
| Вода | O |
| Этанол | CCO |
| Углекислый газ | O=C=O |
| Синильная кислота | C#N |
| Циклогексан | C1CCCCC1 |
| Бензол | c1ccccc1 |



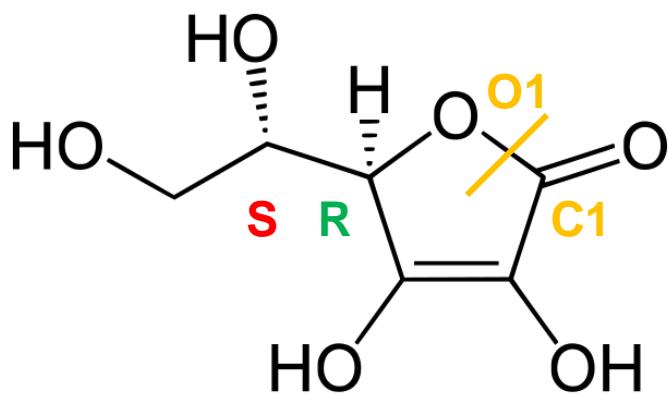
Линейные нотации. SMILES



OCC(O)C1OC(=O)C(O)=C1(O)



OC=1C(OC(=O)C=1O)[C@H](O)CO



O=C1C(O)=C(O)[C@H](O1)[C@@H](O)CO

Линейные нотации. InChI

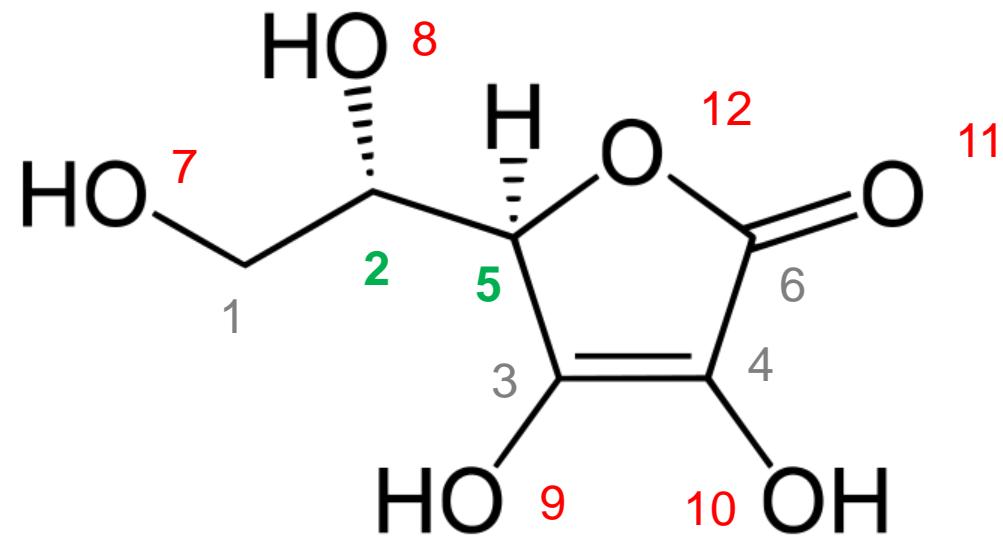
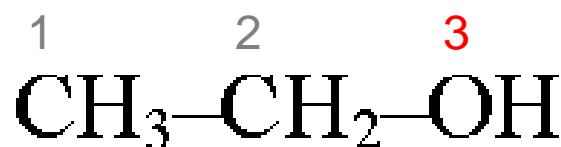
InChI (International Chemical Identifier) — текстовый идентификатор химического соединения для стандартизации кодирования молекулярной информации и представления её в читаемом виде.

Этанол

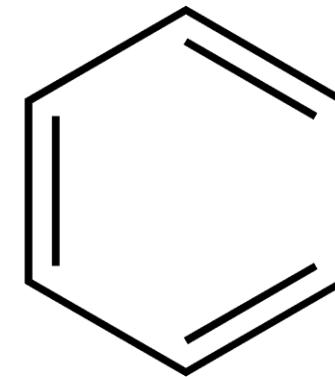
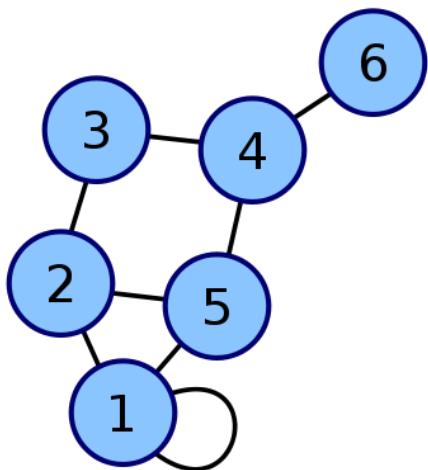
InChI=1/C2H6O/c1-2-3/h3H,2H2,1H3

Аскорбиновая
кислота

InChI=1/C6H8O6/c7-1-2(8)5-3(9)4(10)6(11)12-5/h2,5,7-10H,1H2/t2-,5+/m0/s1



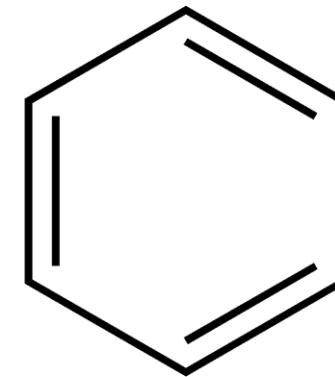
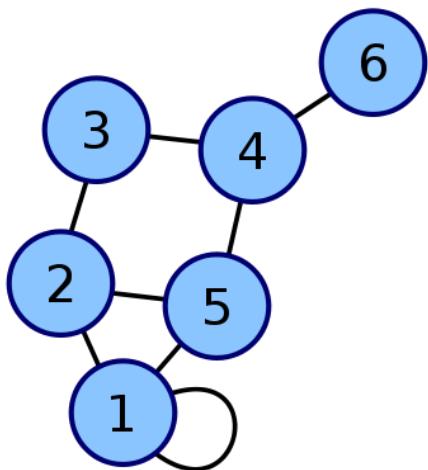
Матрица смежности



$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

?

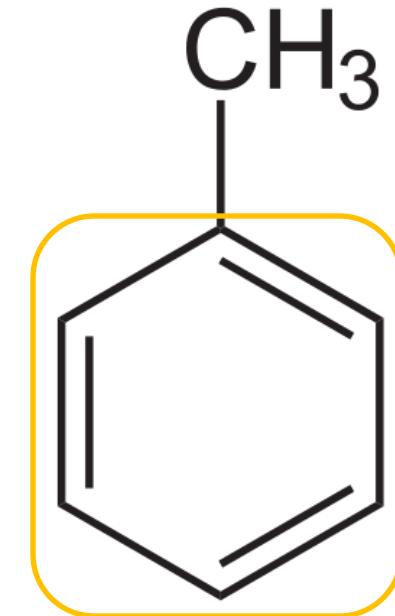
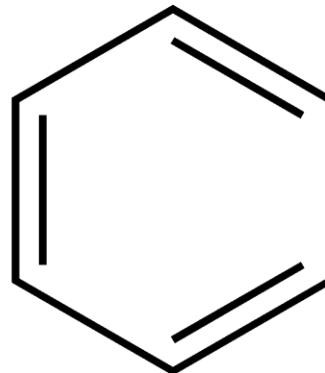
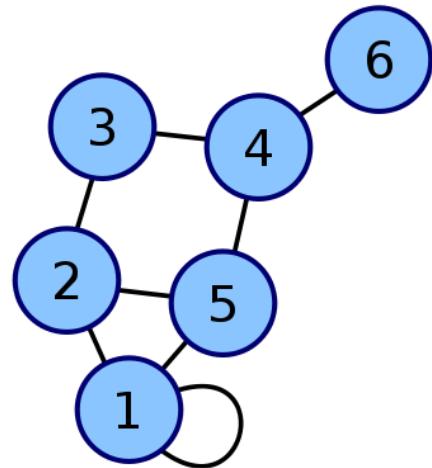
Матрица смежности



$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Матрица смежности



$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Структурные файлы. MOL

benzene

ACD/Labs0812062058

6 6 0 0 0 0 0 0 0 1 v2000

1.9050 -0.7932 0.0000 C 0

1.9050 -2.1232 0.0000 C 0

0.7531 -0.1282 0.0000 C 0

0.7531 -2.7882 0.0000 C 0

-0.3987 -0.7932 0.0000 C 0

-0.3987 -2.1232 0.0000 C 0

2 1 1 0 0 0 0

3 1 2 0 0 0 0

4 2 2 0 0 0 0

5 3 1 0 0 0 0

6 4 1 0 0 0 0

6 5 2 0 0 0 0

M END

Контроль посещаемости сегодня

Заполнить форму через QR-код



3 мин

2 мин

1 мин

Всё!



либо записаться в список на перерыве

Предсказание свойств соединений

Поиск количественных соотношений структура-свойство — процедура построения моделей, позволяющих по структурам химических соединений предсказывать их разнообразные свойства.

Основная гипотеза – сходные соединения имеют сходные свойства.

QSAR - *Quantitative Structure-Activity Relationship* – биологические свойства

QSPR - *Quantitative Structure-Property Relationship* – физические и физико-химические свойства

Примеры:

Температуры плавления и кипения

Вязкость

Давление насыщенных паров

Плотность

Химические сдвиги в спектрах ^1H ЯМР

Растворимость

...

Предсказание свойств соединений

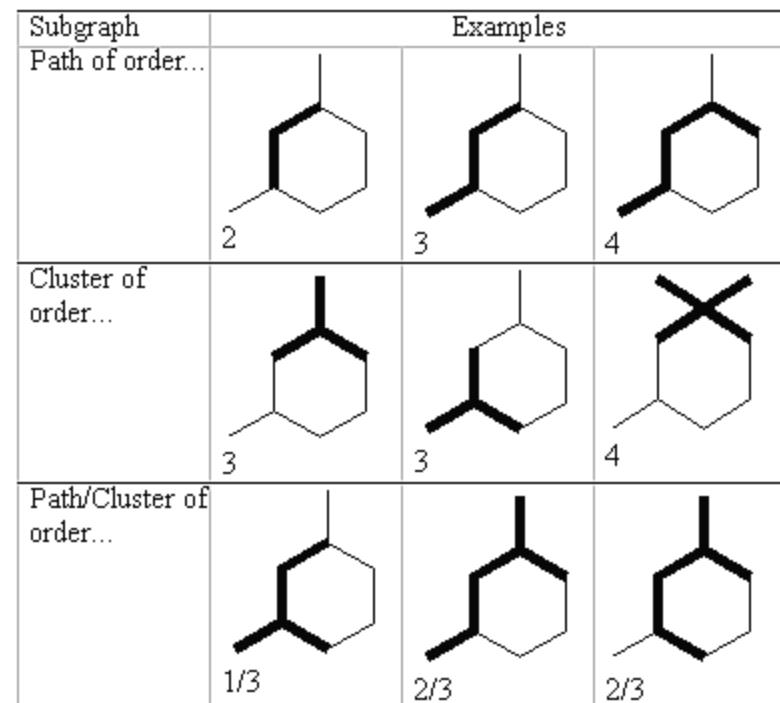
Молекулярные дескрипторы:

"The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment." (Todeschini and Consonni, 2000)

Молекулярные дескрипторы:

- Теоретические (число кратных связей, наличие молекулярных фрагментов,...)
- Экспериментальные (гидрофобность, поляризумость, показатель преломления,...)

Дескрипторы инвариантны, т.е. не зависят от положения молекулы в пространстве.



Молекулярные дескрипторы

Фрагментные дескрипторы – отражают факт наличия фрагмента в молекулярном графе (бинарные) или число вхождений фрагмента (целочисленные)

Физико-химические дескрипторы – соответствуют измеряемым физ-хим величинам (липофильность (LogP), молярная рефракция (MR), молекулярный вес (MW), молекулярные объемы и площади поверхностей,...)

Квантово-химические дескрипторы — величины, получаемые в результате квантово-химических расчетов (энергии граничных орбиталей, частичные заряды на атомах, порядки связей,...)

Дескрипторы молекулярных полей — величины, аппроксимирующие значения молекулярных полей путем вычисления энергии взаимодействия пробного атома, помещенного в узел решетки, с рассматриваемой молекулой

...

Предсказание свойств соединений

В самой общей форме: значение свойства – это некая функция от некого набора дескрипторов.

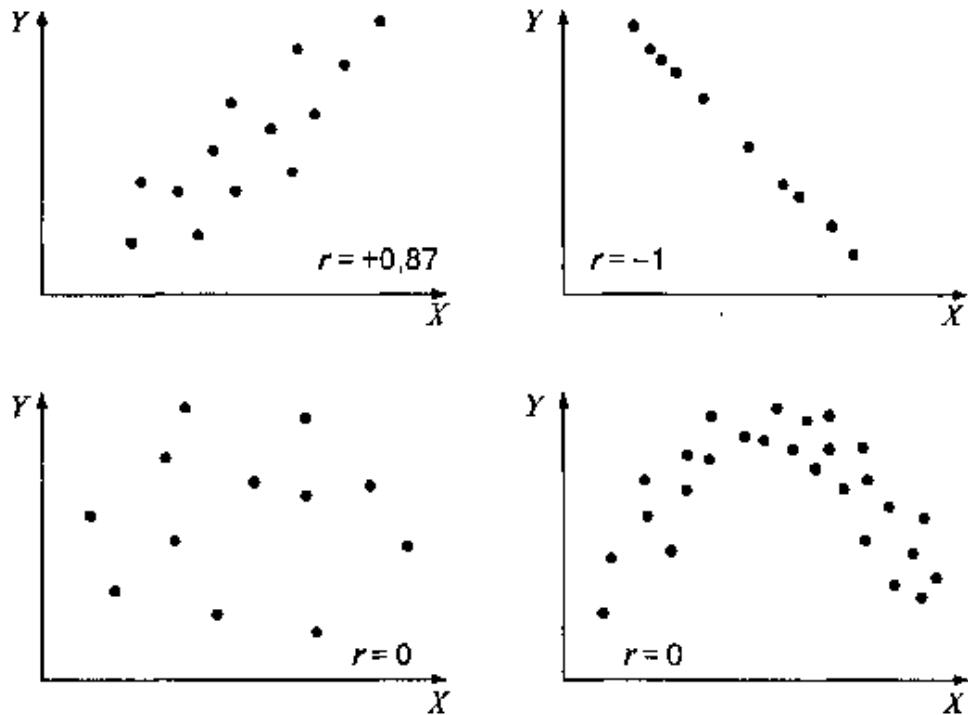
Цель: найти оптимальную функцию и оптимальный набор.

Выявленная связь должна быть проверена =>

- Сравнение модели с экспериментом (коэффициент корреляции)
- Обучающая выборка, тестовая выборка
- Перекрестная проверка (cross-validation) (для маленьких выборок)
- Рандомизация (для больших выборок)

Успех QSAR-модели зависит от точности исходных данных, выбора подходящих дескрипторов и статистических методов и полноценной проверки модели.

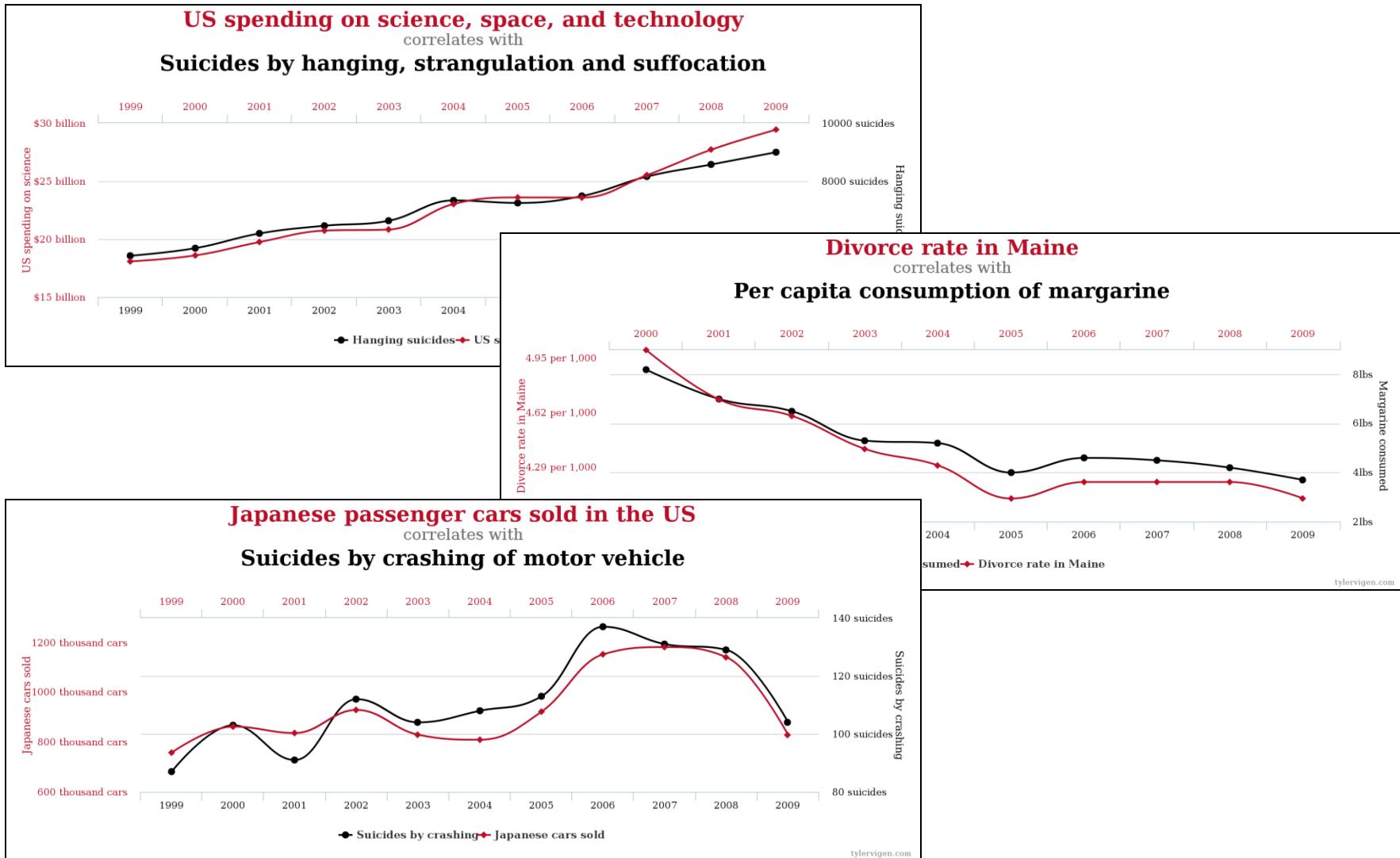
Сравнение модели с экспериментом. Коэффициент корреляции



$$r_{XY} = \frac{\text{cov}_{XY}}{\sigma_X \sigma_Y} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}.$$

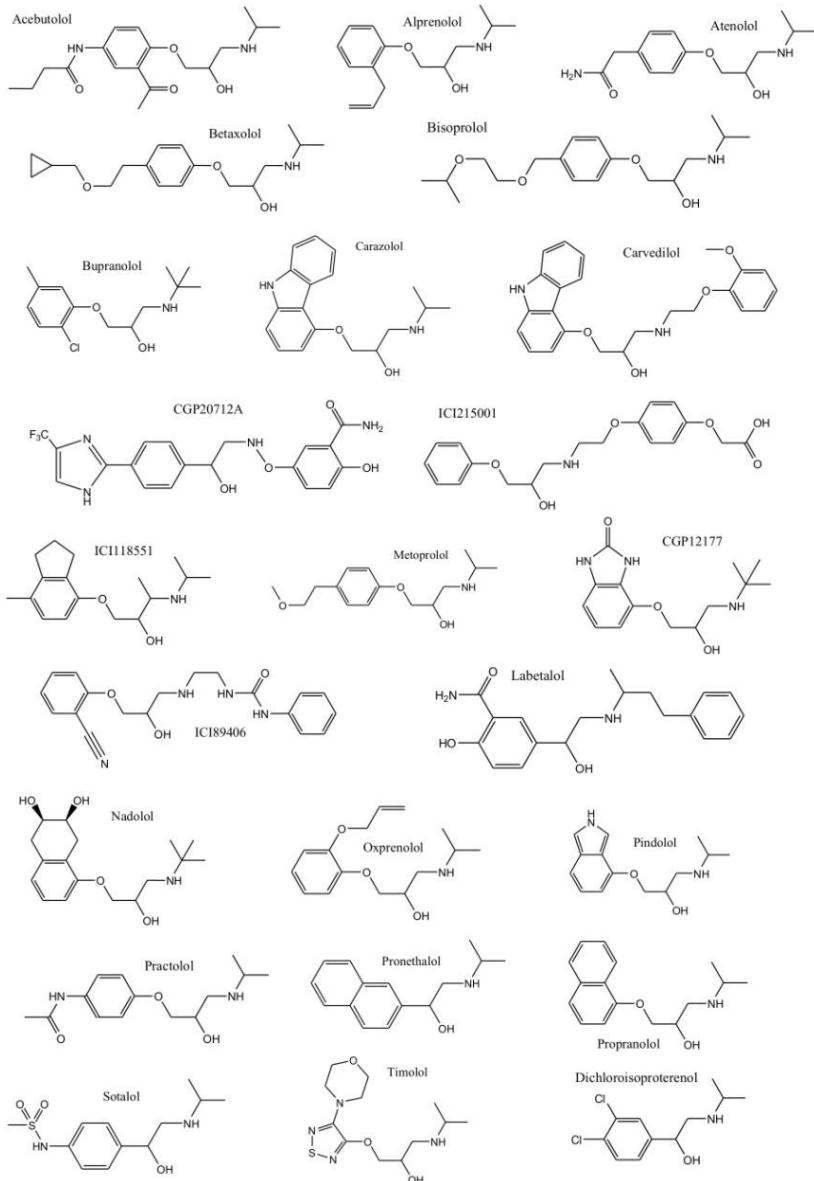
Отсутствие корреляции означает неадекватность выбранной модели

Неожиданные корреляции

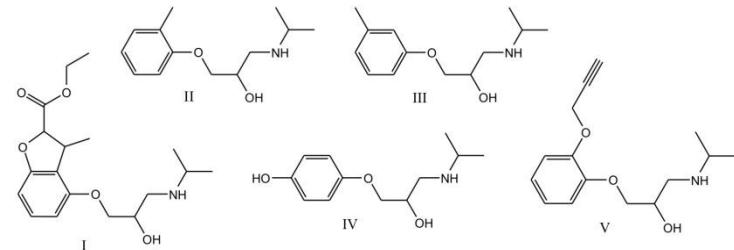


Наличие корреляции не всегда что-то означает. Хотя... 😊

Обучающая и тестовая выборки

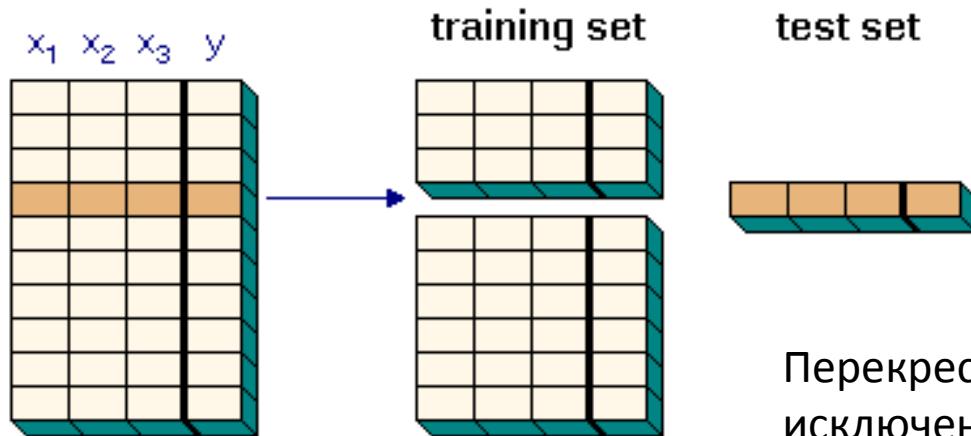


Лиганды бета2-адренергического рецептора



Выборки должны быть репрезентативными,
т.е. отражать генеральную совокупность со
всей возможной полнотой

Перекрестная проверка и рандомизация



Перекрестная проверка с одним исключенным – leave-one-out cross-validation

Для изучения свойств деревьев необязательно рассматривать каждое дерево в лесу.

Аналогично в клинических исследованиях.



Какие свойства можно предсказывать?

Физические свойства индивидуальных низкомолекулярных соединений

Температура кипения (BP)

Вязкость

Плотность

Показатель преломления

Температура плавления (MP)

Константы ионизации (кислотности или основности)

...

Спектроскопические свойства

Положение длинноволновой полосы поглощения симметричных цианиновых красителей

Химические сдвиги в спектрах ^1H ЯМР

...

Физические свойства, обусловленные межмолекулярными взаимодействиями молекул разного типа

Растворимость в воде ($\text{Log}S_w$)

Коэффициент распределения *n*-октанол/вода ($\text{Log}P$)

...

Физические и физико-химические свойства полимеров

Показатель преломления полимеров

Коэффициент проницаемости через полиэтилен низкой плотности

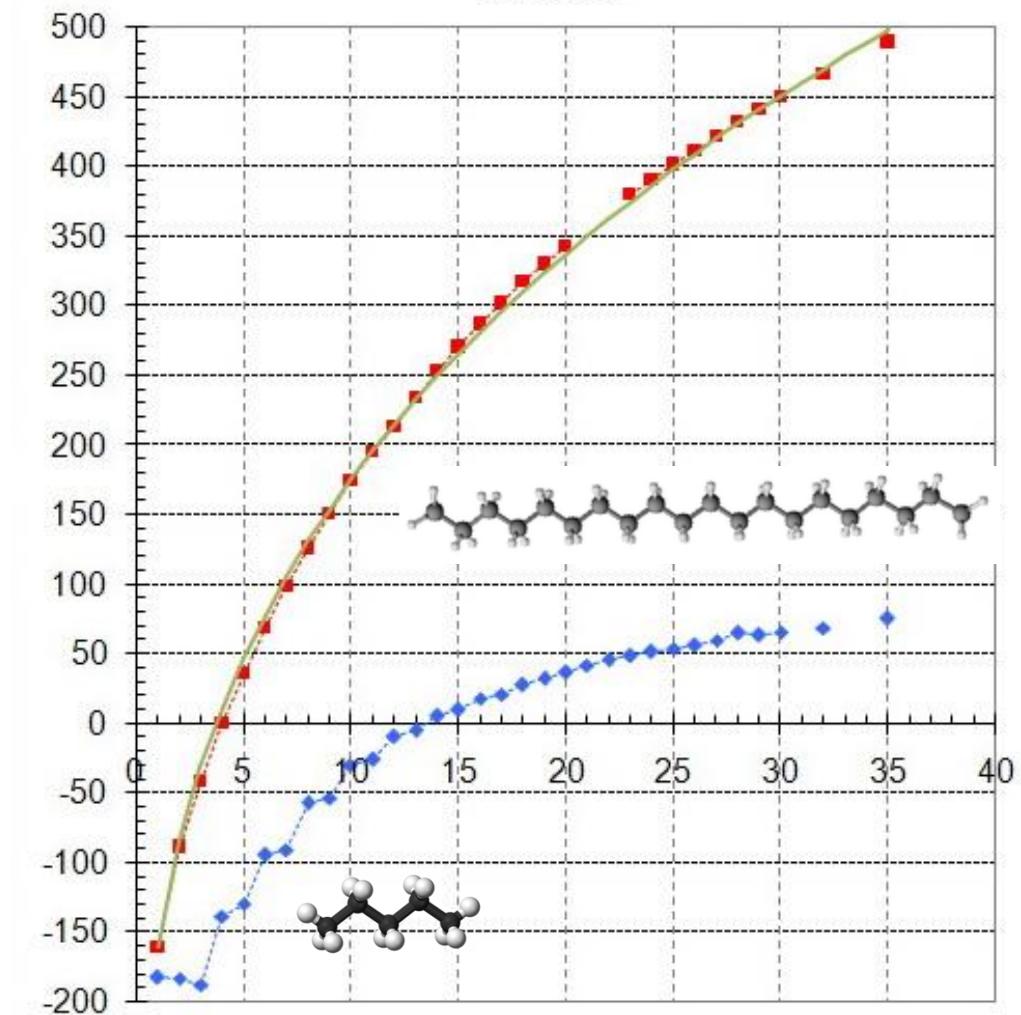
...

Предсказание температур кипения

Предсказание температуры кипения линейных алканов



Алканы



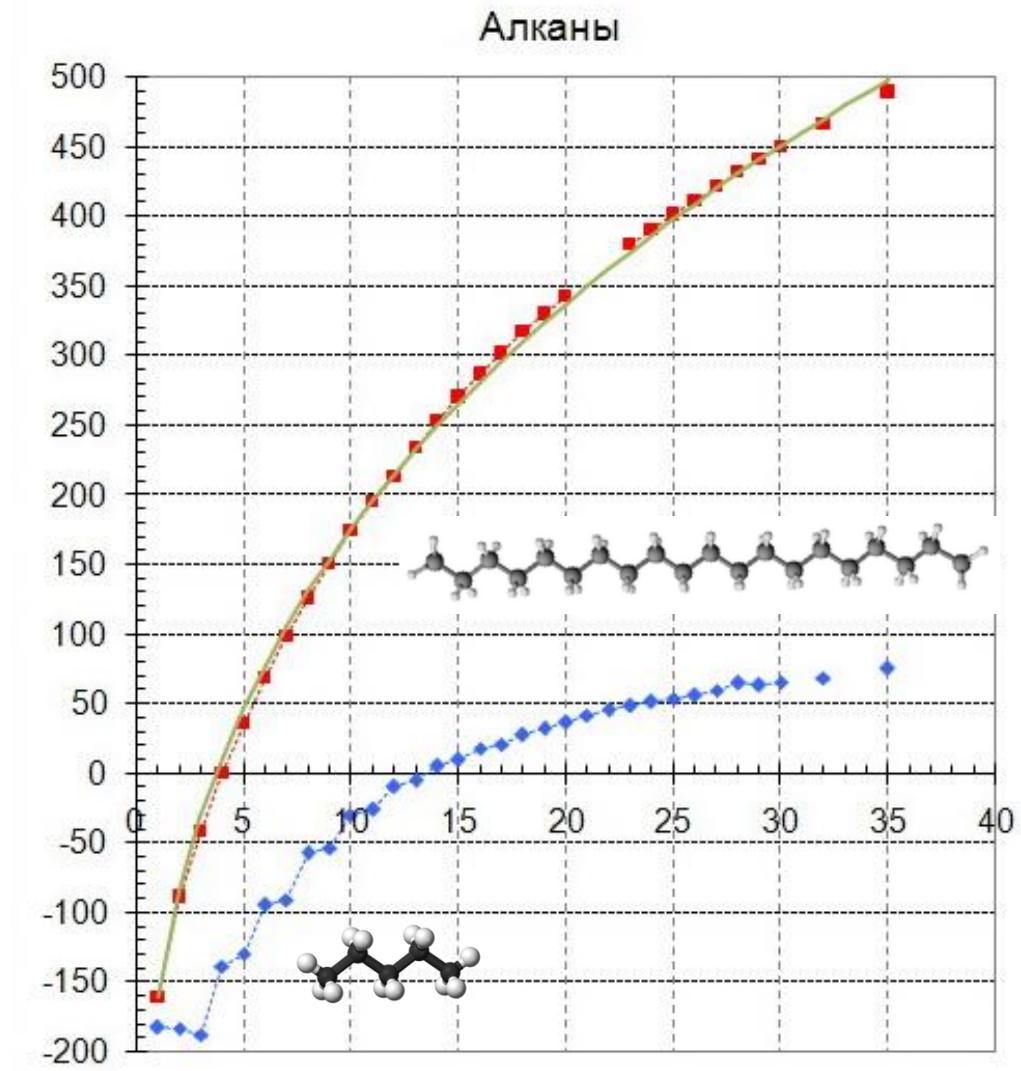
Предсказание температур кипения

Предсказание температуры кипения линейных алканов



$$T(n) = 295 \cdot n^{0.33} - 455$$

$$R^2 = 0.999645$$



Предсказание коэффициента гидрофобности

(Ghose et al., 1988)

$$\log P_{ow} = \sum_i n_i \alpha_i$$

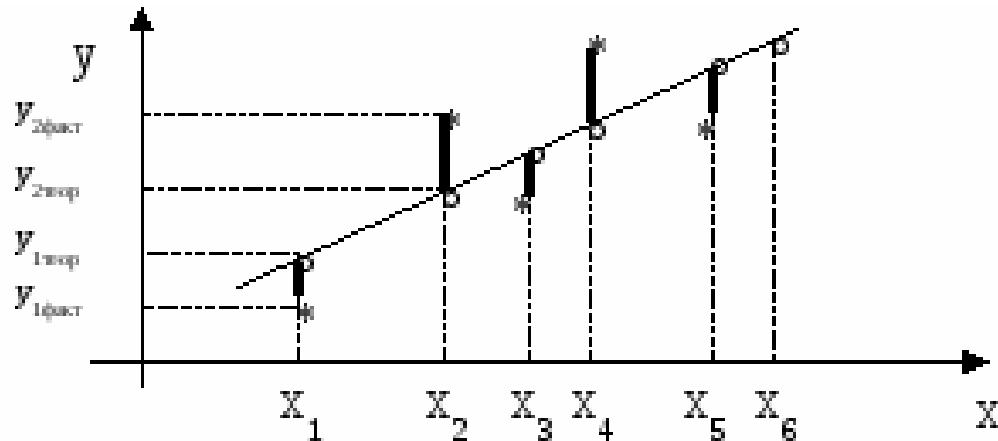
n_i – число атомов типа *i*
α_i – вклад атома типа *i*

Table I. Classification of atoms, and their contributions to octanol-water partition coefficient which is a measure of hydrophobicity.

| Type | Description ^a | Hydrophobic Contribution ^b | No. of Compounds | Frequency of Use | T-test | Molar Refraction ^c |
|-------|-------------------------------------|---------------------------------------|------------------|------------------|--------|-------------------------------|
| C in: | | | | | | |
| 1 | :CH ₃ R, CH ₄ | -0.6037 | 360 | 548 | 100.00 | 2.3000 |
| 2 | :CH ₂ R ₂ | -0.4295 | 216 | 454 | 100.00 | 2.3071 |
| 3 | :CHR ₃ | -0.3426 | 45 | 50 | 100.00 | 2.4926 |
| 4 | :CR ₄ | -0.1155 | 24 | 24 | 74.32 | 2.3000 |
| 5 | :CH ₃ X | -1.0578 | 157 | 224 | 100.00 | 3.4006 |
| 6 | :CH ₂ RX | -0.8188 | 257 | 402 | 100.00 | 3.2624 |
| 7 | :CH ₂ X ₂ | -0.1540 | 5 | 5 | 51.00 | 3.6770 |
| 8 | :CHR ₂ X | -0.5995 | 73 | 118 | 100.00 | 3.0137 |
| 9 | :CHR ₂ X ₂ | 0.0095 | 27 | 27 | 7.85 | 3.225 |
| 10 | :CHX ₃ | 0.5134 | 4 | 4 | 96.02 | 3.2401 |
| 11 | :CR ₃ X | -0.4807 | 14 | 14 | 99.97 | 2.6140 |
| 12 | :CR ₂ X ₂ | 0.2853 | 2 | 2 | 58.14 | 3.1488 |
| 13 | :CRX ₃ | 0.5335 | 34 | 36 | 100.00 | 2.3010 |
| 14 | :CX ₄ | 1.1114 | 6 | 6 | 100.00 | 3.3559 |
| 15 | :—CH ₃ | -0.1654 | 25 | 21 | 97.01 | 2.5071 |

Линейная регрессия

Метод наименьших квадратов (**Гаусс, 1795; Лежандр, 1805**)



$$x = \{x_i\}, i = 1, 2, 3, \dots$$

$$\sum_i (y_i - f(x_i))^2 \rightarrow \min_x$$

$$y_i = a + bx_i + cz_i$$



$$y_i = a + bx_i + cz_i + dx_i z_i$$



$$y_i = a + bx_i + cz_i + bcx_i z_i$$



Функция может быть любой, но линейной по коэффициентам !!!

Какие свойства можно предсказывать?

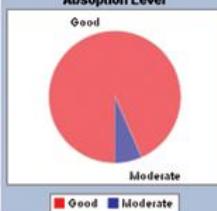
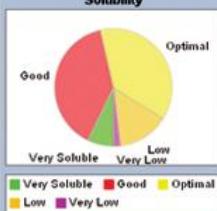
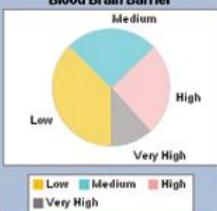
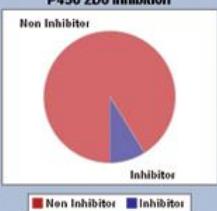
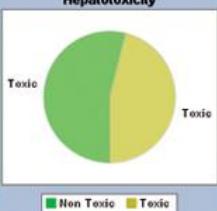
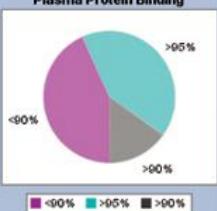
ADMET – Absorption, Distribution, Metabolism, Excretion, and Toxicity

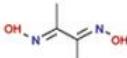
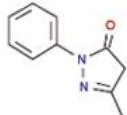
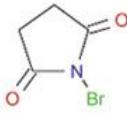
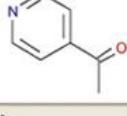
C:\Program Files\SciTegec45\public\users\rbrown\ADMET_report.html - Microsoft Internet Explorer provided by SciTegec

File Edit View Favorites Tools Help

Back Search Favorites Favorites Home Mail Print Find Stop Stop

ADMET Report for Library AB-123

| Absorption Level | Solubility | Blood Brain Barrier | P450 2D6 inhibition | Hepatotoxicity | Plasma Protein Binding |
|---|---|---|--|---|---|
|  |  |  |  |  |  |

| Molecule | Absorption Level | Solubility | Blood Brain Barrier | CYP450 2D6 | Hepatotoxicity | Plasma Protein Binding |
|---|------------------|------------|---------------------|---------------|----------------|------------------------|
|  | Good | 0.285 | -1.188 | Non Inhibitor | Non Toxic | Less Than 90% |
|  | Good | -2.158 | -0.313 | Non Inhibitor | Non Toxic | Greater Than 95% |
|  | Good | -1.097 | -0.626 | Non Inhibitor | Non Toxic | Less Than 90% |
|  | Good | -1.034 | -0.477 | Non Inhibitor | Toxic | Greater Than 95% |

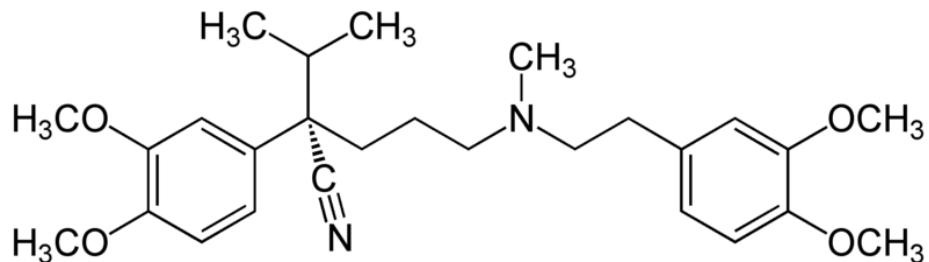
Done My Computer

26

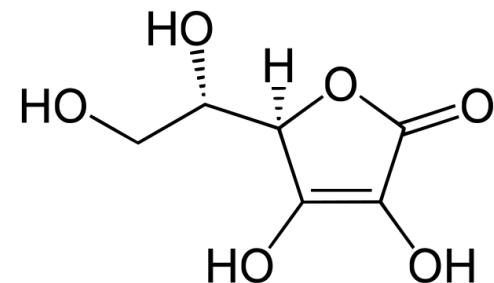
Предсказание биологической активности

«Правило пяти» (Lipinski, 1997) (Rule of thumb):

- Не более 5 доноров водородных связей
- Не более 10 акцепторов водородных связей
- Относительная молярная масса не более 500
- LogP не более 5



Верапамил
Mw=454 LogP=3,79



Аскорбиновая кислота
Mw=176 LogP=-1,9

О мерах сходства выборок

Коэффициент Жаккара («коэффициент флористической общности»):

$$K_J = \frac{c}{a + b - c}$$

Первый предложенный коэффициент сходства!
(P. Jaccard, 1901)

a — количество видов на первой пробной площадке,

b — количество видов на второй пробной площадке,

c — количество видов, общих для 1-ой и 2-ой площадок.



О мерах сходства выборок

Коэффициент Танимото (для битовых строк X_i и Y_i) (1960):

$$S_T = \frac{\sum_i (X_i \wedge Y_i)}{\sum_i (X_i \vee Y_i)}$$

\wedge – логическое И
 \vee – логическое ИЛИ

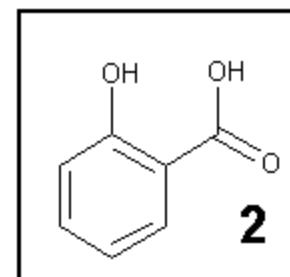
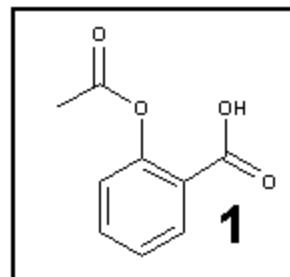
| | | | | | | | | |
|--------------|---|---|---|---|---|---|---|--------------------|
| A | <table border="1"><tr><td>1</td><td>0</td><td>1</td><td>1</td><td>0</td><td>1</td></tr></table> | 1 | 0 | 1 | 1 | 0 | 1 | $ A = 4$ |
| 1 | 0 | 1 | 1 | 0 | 1 | | | |
| B | <table border="1"><tr><td>1</td><td>1</td><td>0</td><td>1</td><td>0</td><td>0</td></tr></table> | 1 | 1 | 0 | 1 | 0 | 0 | $ B = 3$ |
| 1 | 1 | 0 | 1 | 0 | 0 | | | |
| $A \wedge B$ | <table border="1"><tr><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td></tr></table> | 1 | 0 | 0 | 1 | 0 | 0 | $ A \wedge B = 2$ |
| 1 | 0 | 0 | 1 | 0 | 0 | | | |
| $A \vee B$ | <table border="1"><tr><td>1</td><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td></tr></table> | 1 | 1 | 1 | 1 | 0 | 1 | $ A \vee B = 5$ |
| 1 | 1 | 1 | 1 | 0 | 1 | | | |

$$S_T(A, B) = \frac{2}{5}$$



О мерах сходства выборок

Similarity Searching



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| 2 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| | | | | | | | | |

A = Number of bits set in both = 3

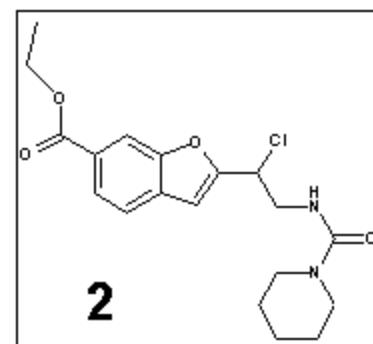
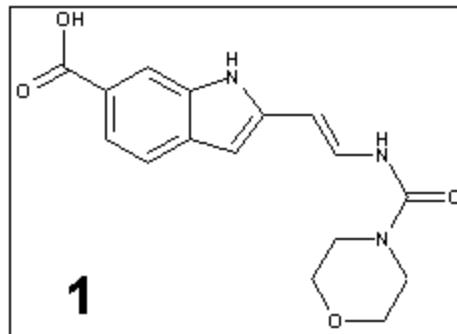
B = Number of bits set in (1), but not in (2) = 2

C = Number of bits set in (2), but not in (1) = 0

$$\text{TANIMOTO COEFFICIENT} = A / (A + B + C)$$
$$= 3 / (3 + 2 + 0) = 0.6 \text{ or } 60\%$$

О мерах сходства выборок

Similarity Searching: Problem 1



| | | | | | | | |
|---|--|--|--|--|--|--|--|
| 1 | | | | | | | |
| 2 | | | | | | | |
| | | | | | | | |
| 1 | | | | | | | |
| 2 | | | | | | | |
| | | | | | | | |

Предсказание биологической активности

Better solutions for your research and development

PASS online

It is easy to use

[GO for prediction »](#)



Methods

Multilevel Neighborhoods of Atoms (MNA) structure descriptors of a molecule are generated on the basis of connection table and table of atoms types presented the compound...

» [read more](#)



Applications

PASS predicts simultaneously 3678 kinds of activity with mean accuracy of prediction about 95% (leave-one-out cross validation) on the basis of the compound's structural formula. ...

» [read more](#)



Publications

Current and past publications, including statistical reports, surveys, press releases, circulars and legislation, are available in electronic format from this section.

» [read more](#)



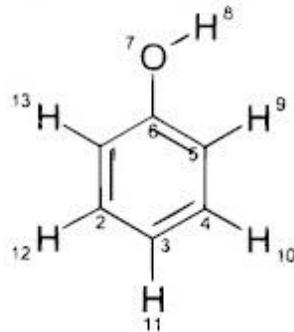
Downloads

You can download and use locally PASS demo version, which predicts 50 biological activities, and test it on your computer. It possible make a batch with several compounds at the same time.

» [click for it](#)

Предсказание биологической активности

Table 2. Representation of Phenol by MNA Descriptors of Zero, First, and Second Levels (MNA/0, MNA/1, MNA/2)^a



MNA – multilevel neighborhoods of atoms

(Filimonov , Poroikov et al., 1999)

| atom | MNA/0 | MNA/1 | MNA/2 |
|------|-------|---------|--------------------------|
| 1 | C | C(CC-H) | C(C(CC-H)C(CC-O)-H(C)) |
| 2 | C | C(CC-H) | C(C(CC-H)C(CC-H)-H(C)) |
| 3 | C | C(CC-H) | C(C(CC-H)C(CC-H)-H(C)) |
| 4 | C | C(CC-H) | C(C(CC-H)C(CC-H)-H(C)) |
| 5 | C | C(CC-H) | C(C(CC-H)C(CC-O)-H(C)) |
| 6 | C | C(CC-O) | C(C(CC-H)C(CC-H)-O(C-H)) |
| 7 | -O | -O(C-H) | -O(C(CC-O)-H(-O)) |
| 8 | -H | -H(-O) | -H(-O(C-H)) |
| 9 | -H | -H(C) | -H(C(CC-H)) |
| 10 | -H | -H(C) | -H(C(CC-H)) |
| 11 | -H | -H(C) | -H(C(CC-H)) |
| 12 | -H | -H(C) | -H(C(CC-H)) |
| 13 | -H | -H(C) | -H(C(CC-H)) |

^a Hyphen (-) is the chain marker for the atoms in the chains.

Предсказание биологической активности

Мера сходства – модифицированный коэффициент Танимото

Calculation of Similarity. We have modified the Tanimoto coefficient to take into account the different frequencies of descriptors. The similarity between two molecules, A and B, is given by

$$\text{sim}(A, B) = \frac{\sum_{i=1}^M \min[A(i), B(i)]}{\sum_{i=1}^M A(i) + \sum_{i=1}^M B(i) - \sum_{i=1}^M \min[A(i), B(i)]} \quad (1)$$

where $A(i)$ and $B(i)$ are the counts of descriptor i in the molecules A and B, respectively; M is the total number of various descriptors in the dictionary.

Предсказание биологической активности

Ограничения предсказательной способности

PASS cannot predict the activity spectrum for essentially new compound if all its descriptors are new and so they don't occur in the training set. If a compound has more than 2 new descriptors it is rather new and prediction results may be considered as pilot.

In some cases PASS predicts both agonist's and antagonist's (blocker and stimulator) actions simultaneously. Thus, only experiments can clarify the biological activity of a compound, but it has an affinity to appropriate receptor (enzyme).

PASS does not predict if the compound will become a drug, but helps to select the most prospective leads.

Базы данных химических соединений

NCBI PubChem Compound

Search PubChem Compound for epinephrine Go Clear Save

Advanced Search Preview/Index History Clipboard Details

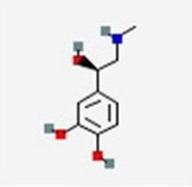
Display Summary Show 20 Sort By Send to

Tools: Links: Related Structures, BioAssays, BioSystems, Literature, Other Links

All: 120 Rule of 5: 62

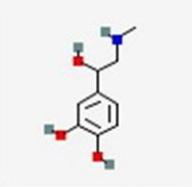
Items 1 - 20 of 120 Page | 1 of 6 Next

1: CID: 5816 Related Structures, BioAssays, BioSystems, Literature, Other Links



epinephrine; Adrenalin; adrenaline ...
IUPAC: 4-[(1R)-1-hydroxy-2-(methylamino)ethyl]benzene-1,2-diol
MW: 183.204420 g/mol | MF: C₉H₁₃NO₃
Tested in BioAssays: All: 274, Active: 22; BioActivity Analysis
Vasoconstrictor Agents... more

2: CID: 838 Related Structures, BioAssays, Literature, Other Links



epinephrine; DL-Adrenaline; Racetepinefrine ...
IUPAC: 4-[1-hydroxy-2-(methylamino)ethyl]benzene-1,2-diol
MW: 183.204420 g/mol | MF: C₉H₁₃NO₃
Tested in BioAssays: All: 17, Active: 1; BioActivity Analysis
Vasoconstrictor Agents... more

Compounds: 95,276,293
Substances: 249,470,154
BioAssays: 1,252,883
Tested Compounds: 2,570,179
Tested Substances: 4,157,676
RNAi BioAssays: 170
BioActivities: 235,470,936
Protein Targets: 10,857
Gene Targets: 22,106

Selected Compounds Compound Count

BioActivity Experiments

| | Compound Count |
|-----------------------------------|----------------|
| BioAssays, Active | 39 |
| BioAssays, Tested | 60 |
| Protein 3D Structures | 6 |
| Crystal Structure Of Dipeptide... | 1 |

BioMedical Annotation

| | Compound Count |
|-------------------------|----------------|
| Pharmacological Actions | 69 |
| Sympathomimetics | 37 |
| BioSystems | 16 |

Depositor Category

| | Compound Count |
|-------------------------|----------------|
| Biological Properties | 120 |
| Chemical Vendors | 55 |
| Journal Publishers | 47 |
| NIH Molecular Libraries | 52 |

Recent activity

Turn Off Clear

Базы данных химических соединений



University of California, San Francisco

About UCSF

Search UCSF

UCSF Medical Center

Shoichet
Laboratory

docking.org

ZINC 12

Not Authenticated – sign in

Active cart: Temporary Cart (0 items)

About Search Subsets Help Social G+

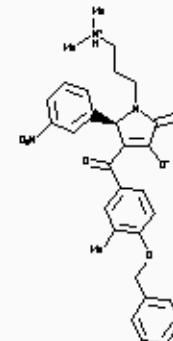
Quick Search Bar...

Go

Please consider switching to [ZINC15](#), which is superior to ZINC12 in most ways. If you prefer ZINC12 after trying ZINC15, we would like to know why @chem4biology so that we can get you to make the switch.

Molecule of the Minute

[8817205](#)



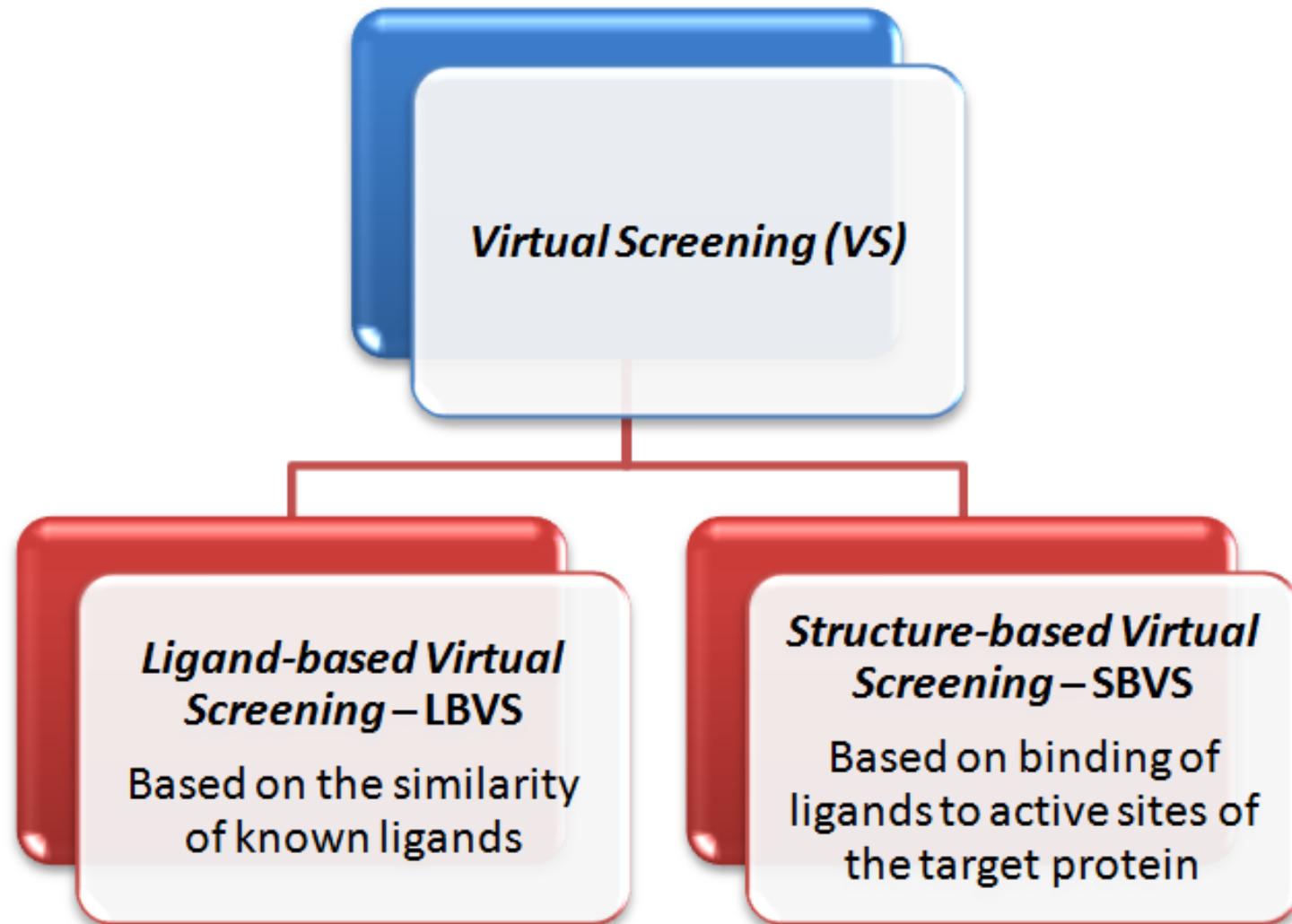
Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 35 million purchasable compounds in ready-to-dock, 3D formats. ZINC is provided by the [Irwin](#) and [Shoichet](#) Laboratories in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF). To cite ZINC, please reference: Irwin, Sterling, Mysinger, Bolstad and Coleman, *J. Chem. Inf. Model.* 2012 DOI: [10.1021/ci3001277](https://doi.org/10.1021/ci3001277). The original publication is Irwin and Shoichet, *J. Chem. Inf. Model.* 2005;45(1):177-82 [PDF](#), [DOI](#). We thank [NIGMS](#) for financial support (GM71896).

ZINC ID, Drug Name, SMILES, Catalog, Vendor Code, Target & r

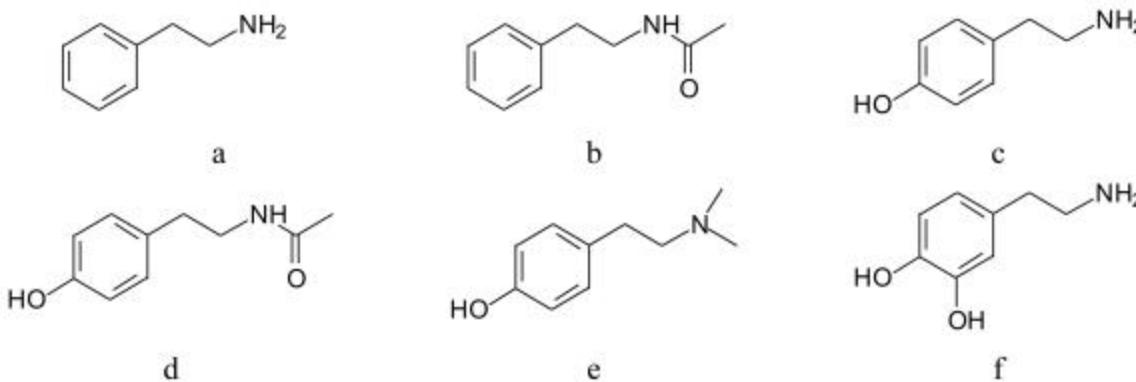
Go

Structure/Draw Physical Properties Catalogs & Vendors ZINC IDs Targets Rings Combination

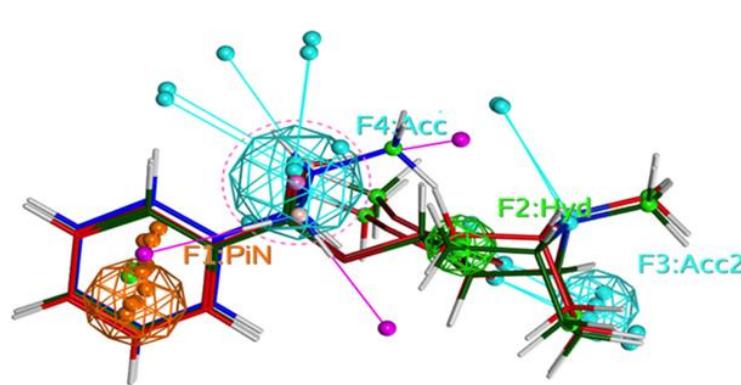
Виртуальный скрининг



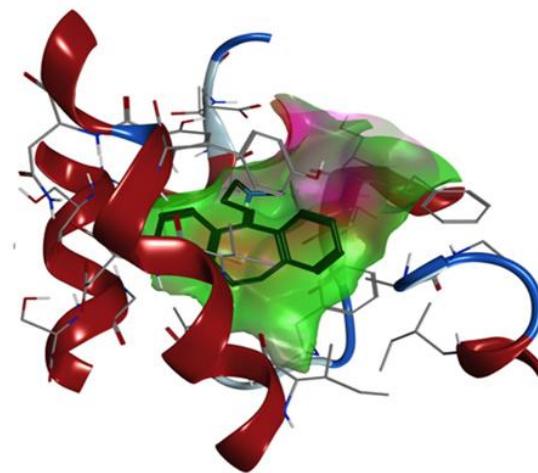
Виртуальный скрининг



Отбор по формальным признакам



Фармакофорный поиск



Молекулярный докинг

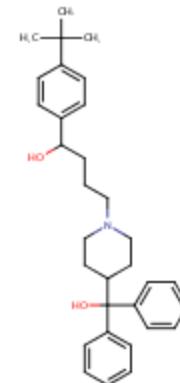
Отбор по формальным признакам. SwissSimilarity

Run parameters

Library screened ZincDrugLike
Screening method Combined
Date Fri Nov 10 11:13:52 2017

If you publish these results, please, cite the following paper: Zoete, V., Daina, A., Bovigny, C., & Michielin, O. SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening., *J. Chem. Inf. Model.*, 2016, 56(8), 1399-1404.

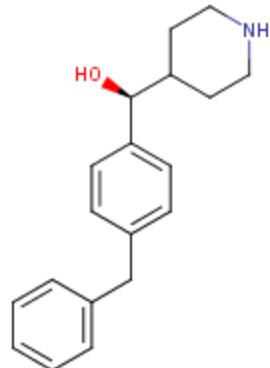
Query Molecule



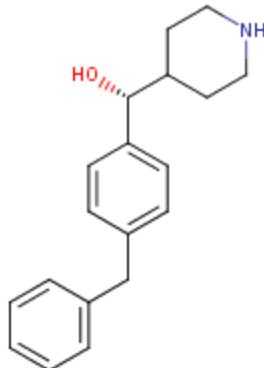
Results

Retrieve data: [CSV](#) [JSON](#) [PDF](#) [PPT](#) [Email](#)

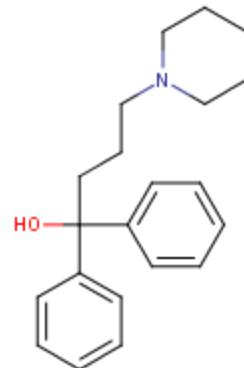
ZINC80181216
Score : 0.983



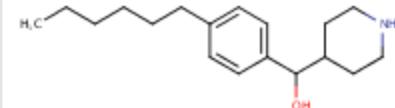
ZINC80181212
Score : 0.983



ZINC00968266
Score : 0.972



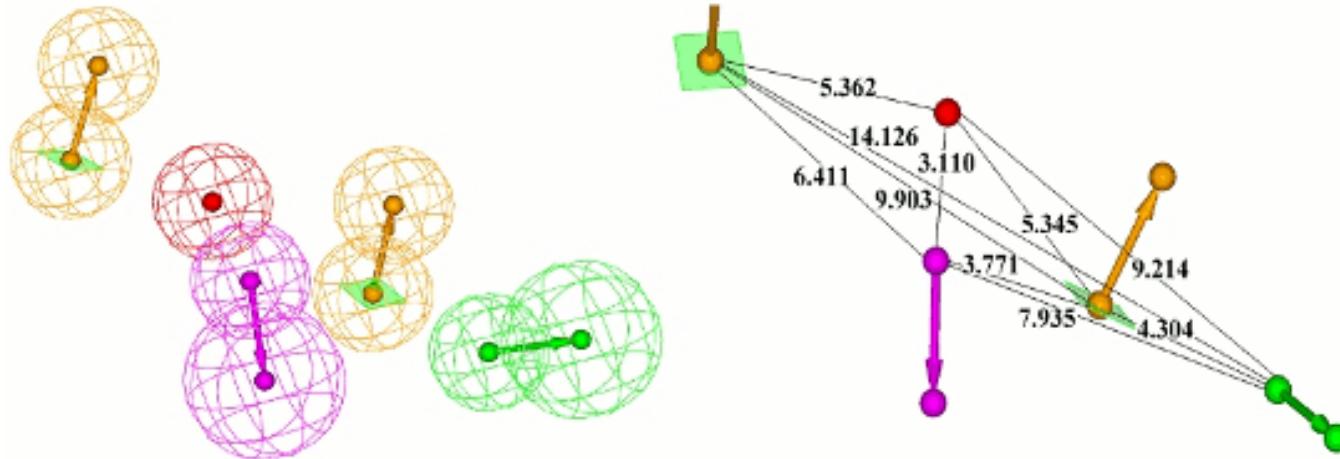
ZINC74369400
Score : 0.969



Фармакофорный поиск

Pharmacophore – “proposed receptor pattern” (Kier, 1971)

Фармакофор – набор пространственных и электронных признаков, необходимых для обеспечения оптимальных супрамолекулярных взаимодействий со специфической биологической мишенью, которые могут вызывать (или блокировать) ее биологический ответ (ИЮПАК).

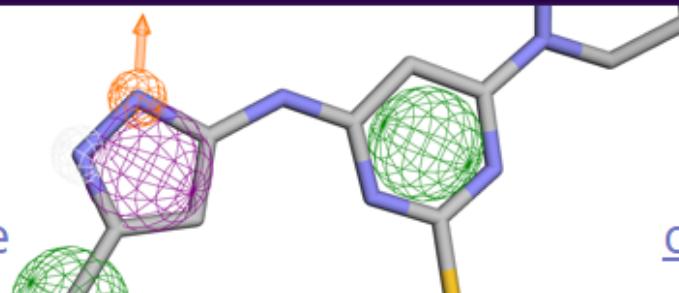


Pharmacophore model for β_2 -adrenoreceptor agonists generated by HipHop.
Features are portrayed as mashed spheres, color-coded as follows: green, hydrogen-bond acceptor, magenta, hydrogen-bond donor, orange, aromatic ring, red, positive ionizable feature.

Фармакофорный поиск. PharmIt

pharmit

interactive exploration of chemical space



[code](#)
[help](#)

[contact](#)

search

virtual screening in your browser

[enter pharmit search](#)

start from PDB: ligand ▾

binding site waters: ignore ▾

[submit](#)

[examples](#)

create

submit your own chemical libraries

[log in to manage libraries](#)

email:

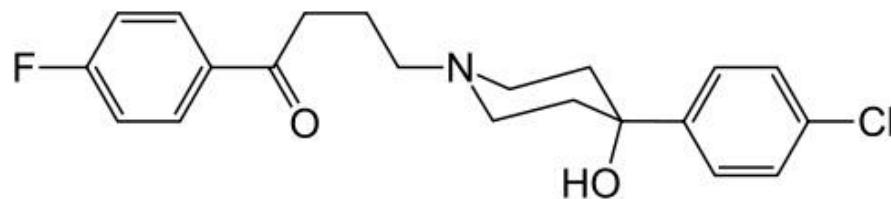
password:

[log in](#)

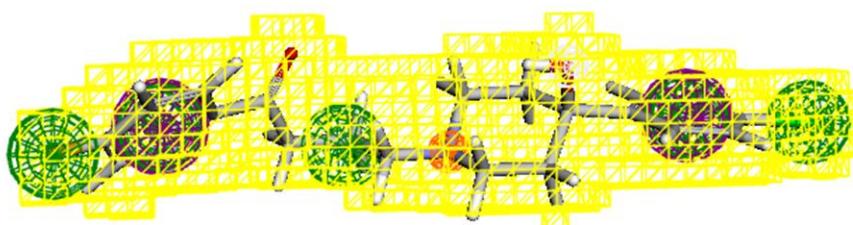
[register new account](#)

[log in as guest](#)

Фармакофорный поиск. PharmIt



Галоперидол



Search PubChem ▾

Pharmacophore Search -> Shape Filter

Load Receptor... Load Features...

Pharmacophore

- Aromatic (6.08,0.16,-0.02) Radius 1.1
- Aromatic (-6.31,0.28,-0.05) Radius 1.1
- HydrogenDonor (2.94,1.31,-1.29) Radius 0.5
- HydrogenAcceptor (0.47,-0.38,0.73) Radius 0.5
- HydrogenAcceptor (2.94,1.31,-1.29) Radius 0.5
- HydrogenAcceptor (-3.19,-0.63,-1.63) Radius 0.5
- Hydrophobic (6.08,0.16,-0.02) Radius 1.0
- Hydrophobic (-6.31,0.28,-0.05) Radius 1.0
- Hydrophobic (9.18,-0.06,0.22) Radius 1.0
- Hydrophobic (-8.77,1.37,0.43) Radius 1.0
- Hydrophobic (-1.89,-1.26,0.6) Radius 1.0

Load Session... Save Session...

| Pharmacophore Results | | | |
|-----------------------|-------|------|-------|
| Name | RMSD | Mass | RBnds |
| PubChem-118753530 | 0.390 | 377 | 6 |
| PubChem-8957387 | 0.486 | 396 | 9 |
| PubChem-7773328 | 0.492 | 379 | 9 |
| PubChem-7773305 | 0.519 | 397 | 9 |
| PubChem-122716498 | 0.556 | 381 | 9 |
| PubChem-119265 | 0.581 | 378 | 6 |
| PubChem-8957383 | 0.606 | 373 | 10 |
| PubChem-10200496 | 0.624 | 392 | 6 |
| PubChem-11257780 | 0.645 | 390 | 6 |
| PubChem-2795977 | 0.779 | 389 | 10 |

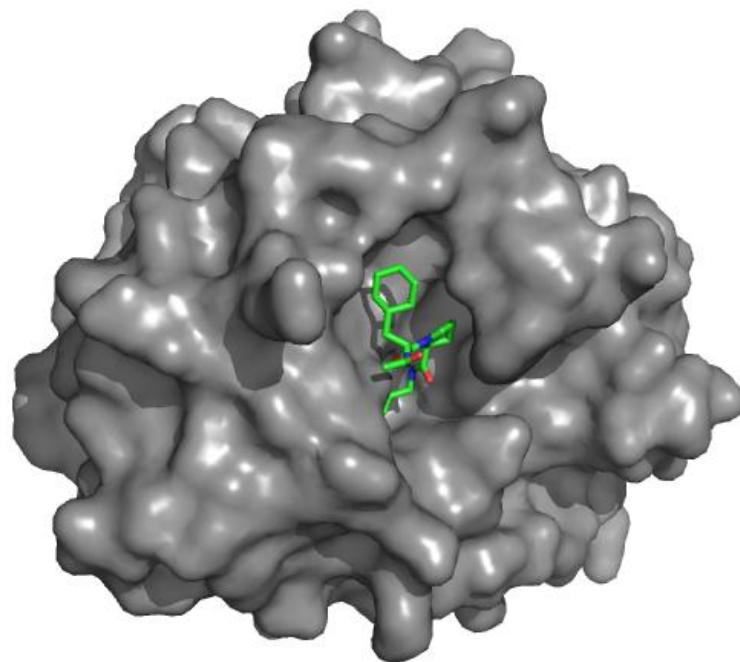
Showing 1 to 10 of 10 hits

Previous 1 Next

Minimize

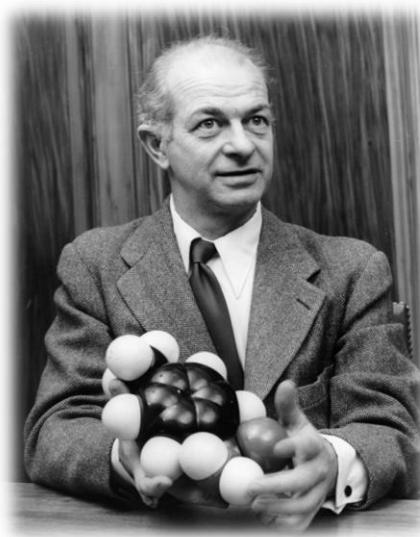
Save..43

Молекулярный докинг

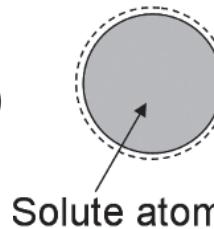


Молекулярные поверхности

(Corey & Pauling, 1953)



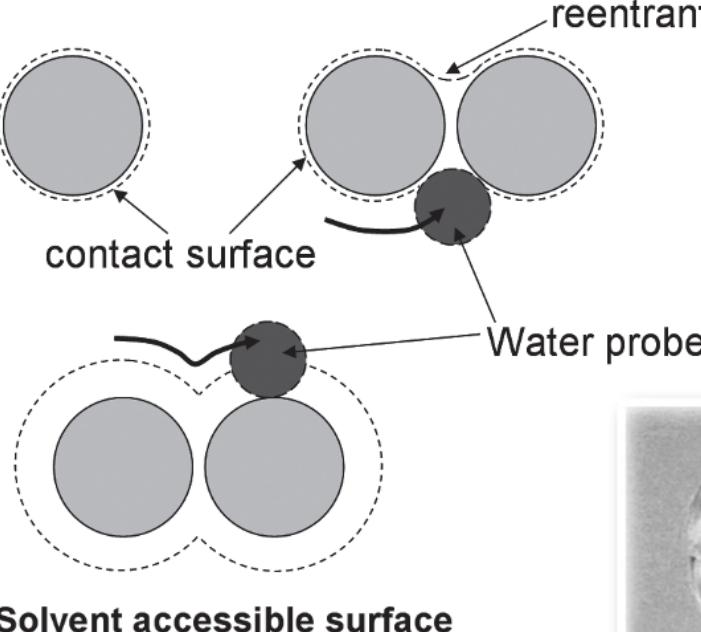
Van der Waals surface



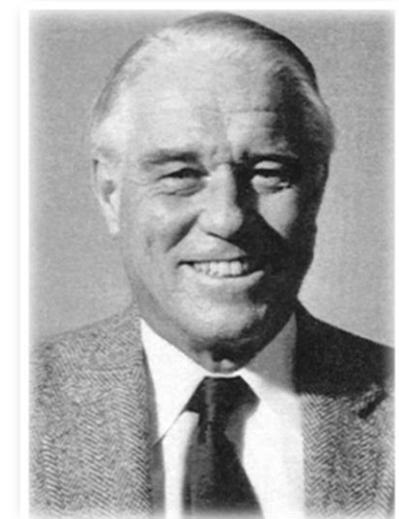
contact surface

Molecular surface

reentrant surface



(Richards, 1977)

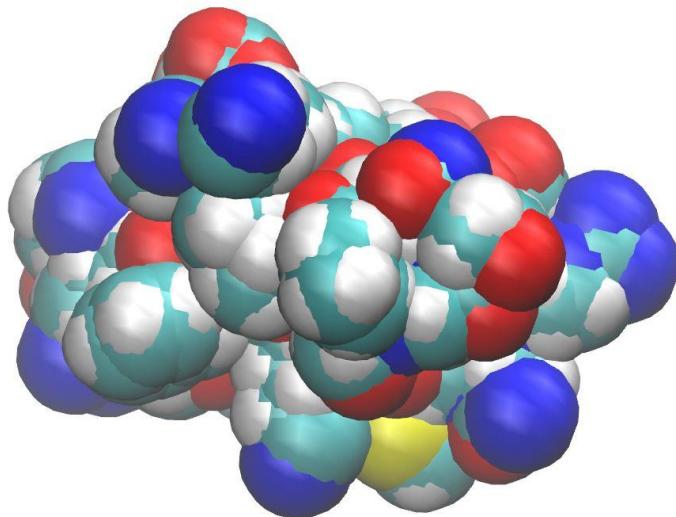
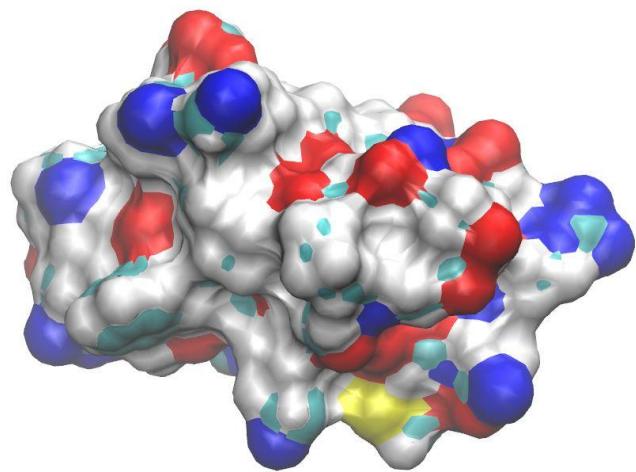
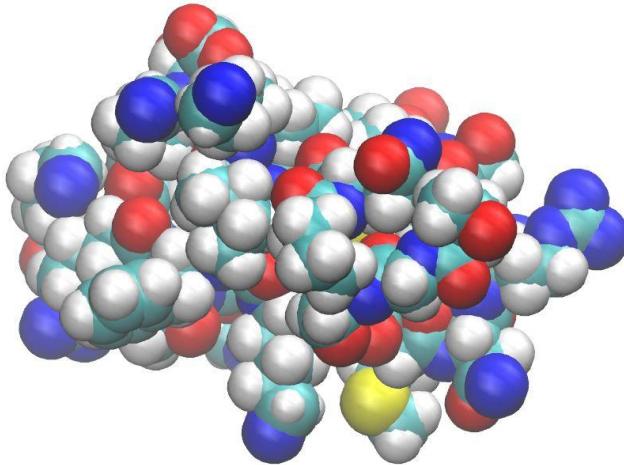
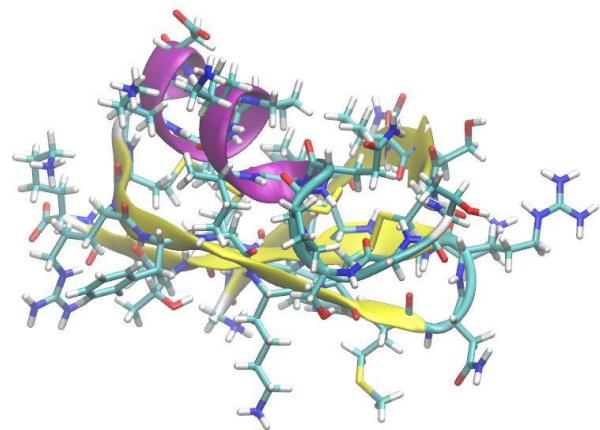


Solvent accessible surface

Linus Pauling
(1901 - 1994)

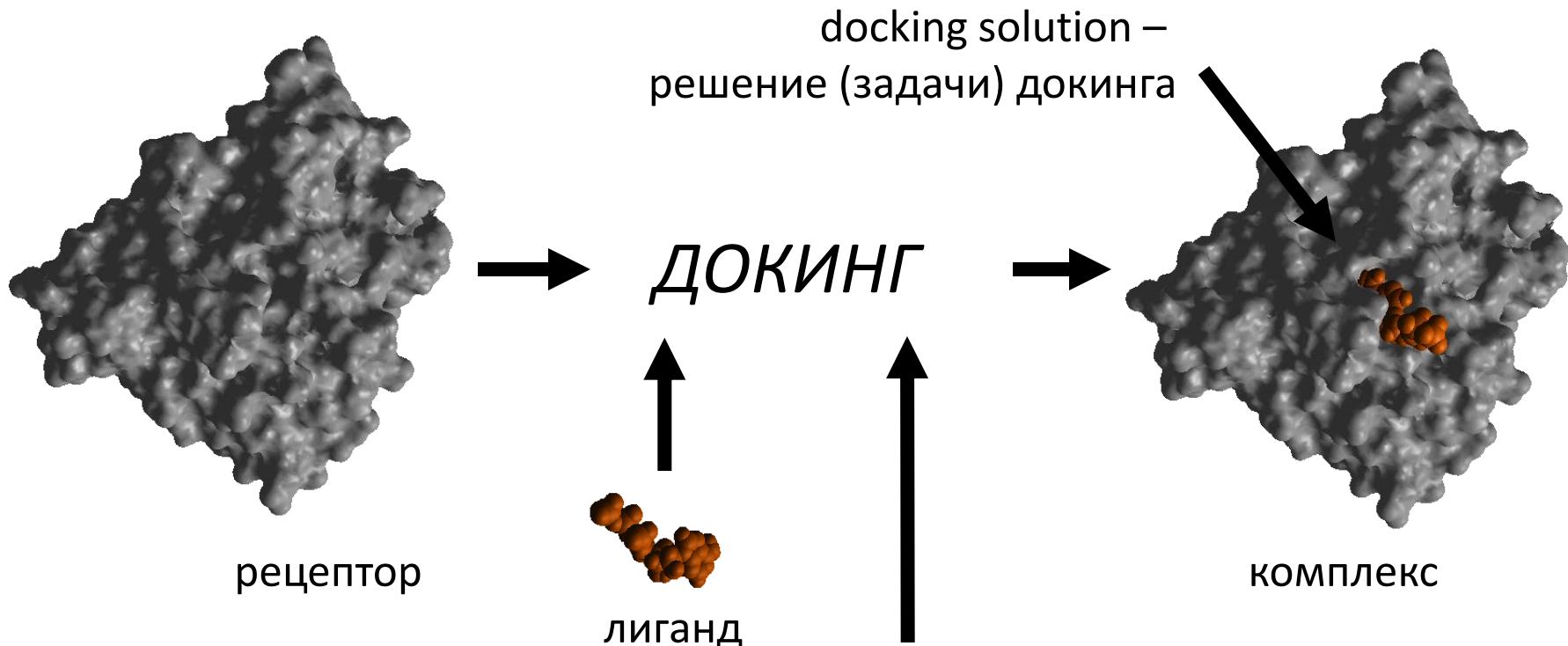
Frederic Richards
(1925 - 2009)

Молекулярные поверхности



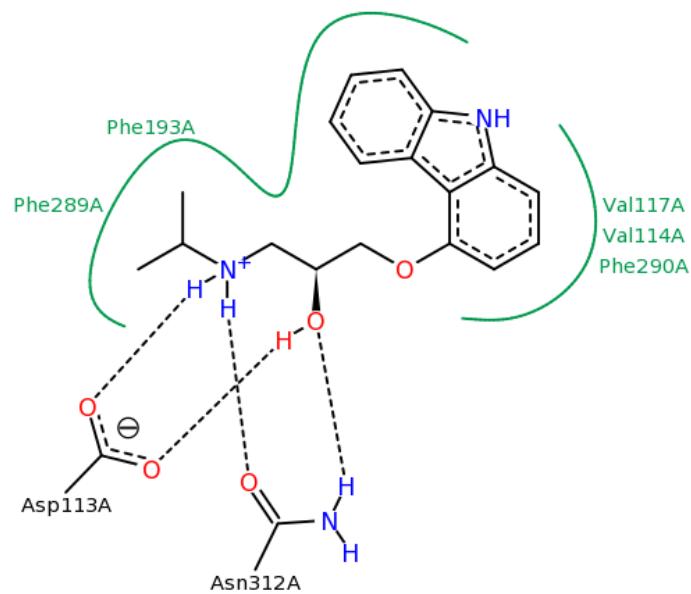
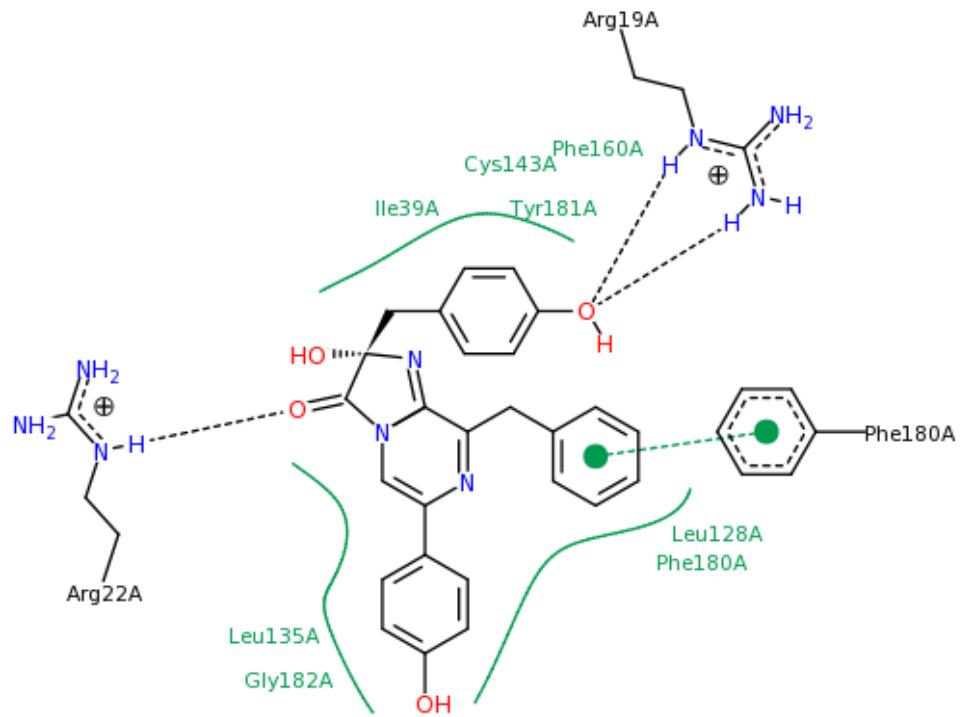
Структура аджитоксина (pdb-код 1agt) в ленточном и стержневом представлении, в виде сфер ван-дер-Ваальса, а также ее молекулярная поверхность и поверхность, доступная растворителю (SAS)

Общая постановка задачи



оценочная функция:
*водородные связи, гидрофобные взаимодействия,
стэкинг-взаимодействия, ...*

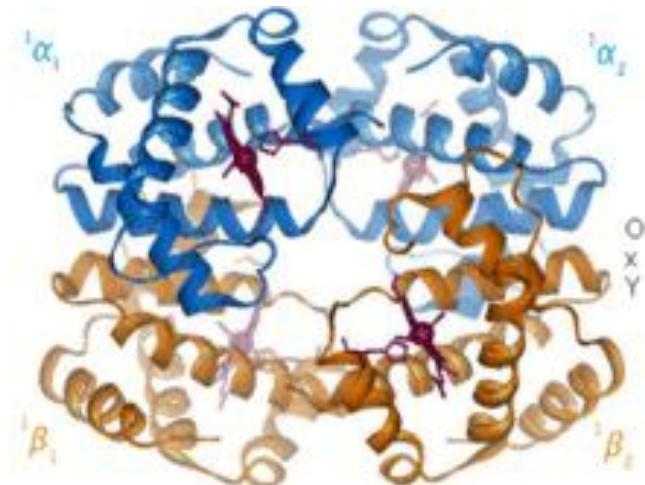
Взаимодействия



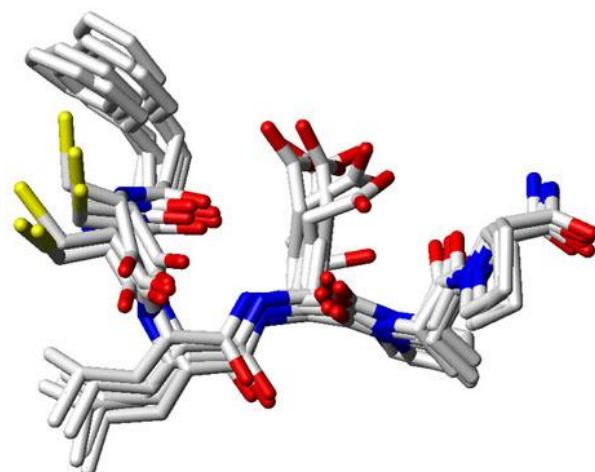
Представление белков и лигандов

Гибкость белка:

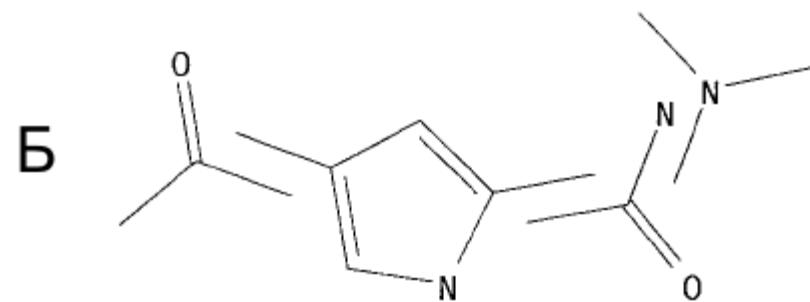
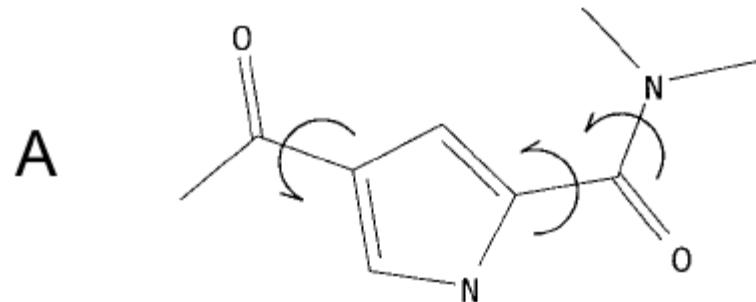
- быстрые движения малого масштаба – движение боковых цепей и петель
- медленные движения крупного масштаба – движение доменов
- ренатурация частично развернутых белков



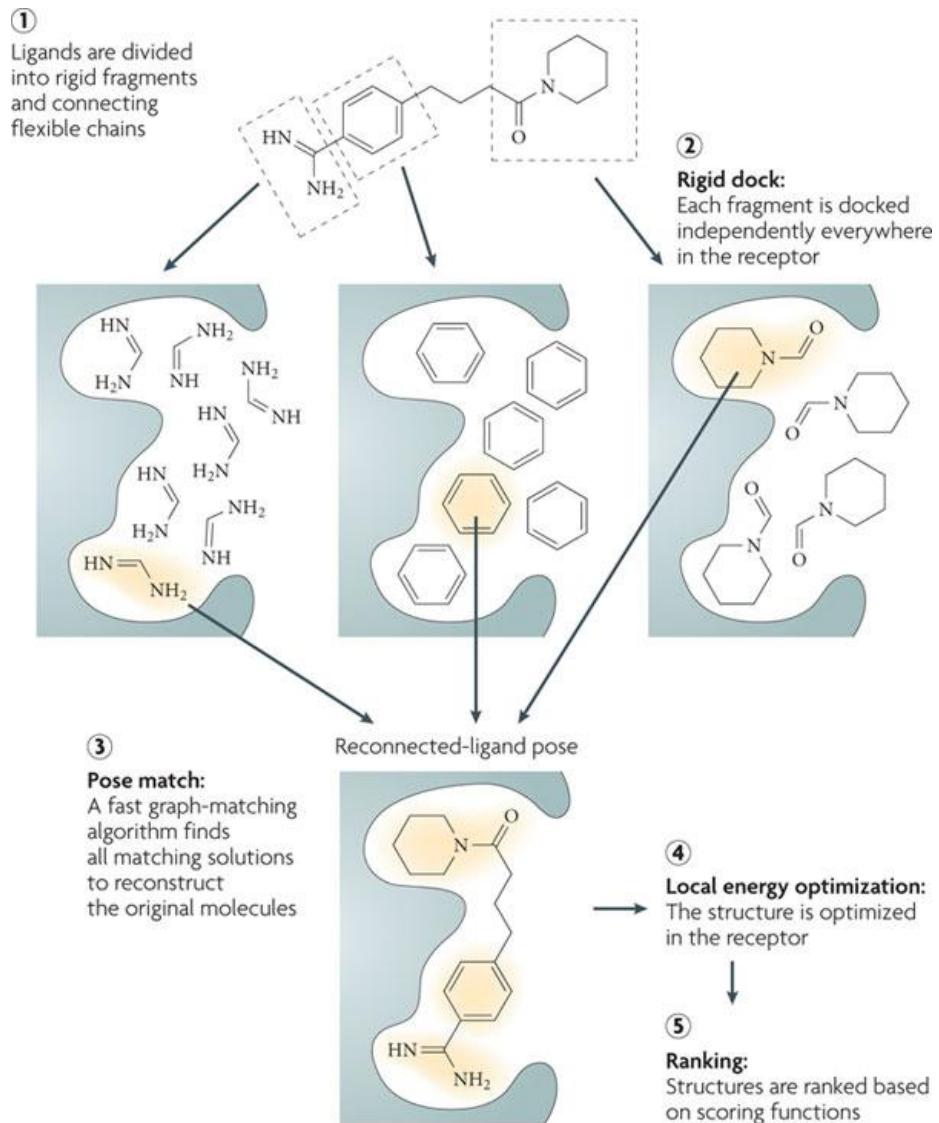
Гибкость лиганда



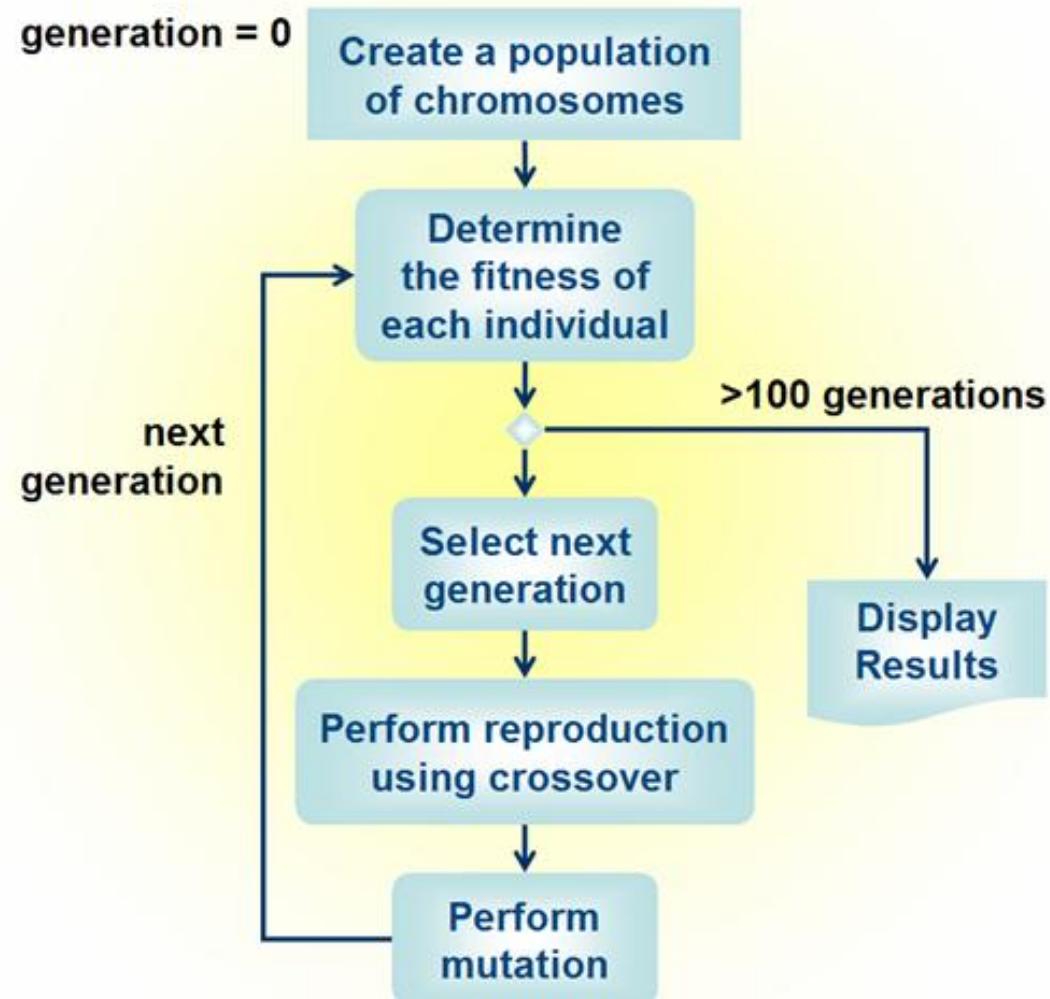
Алгоритмы. Последовательная сборка



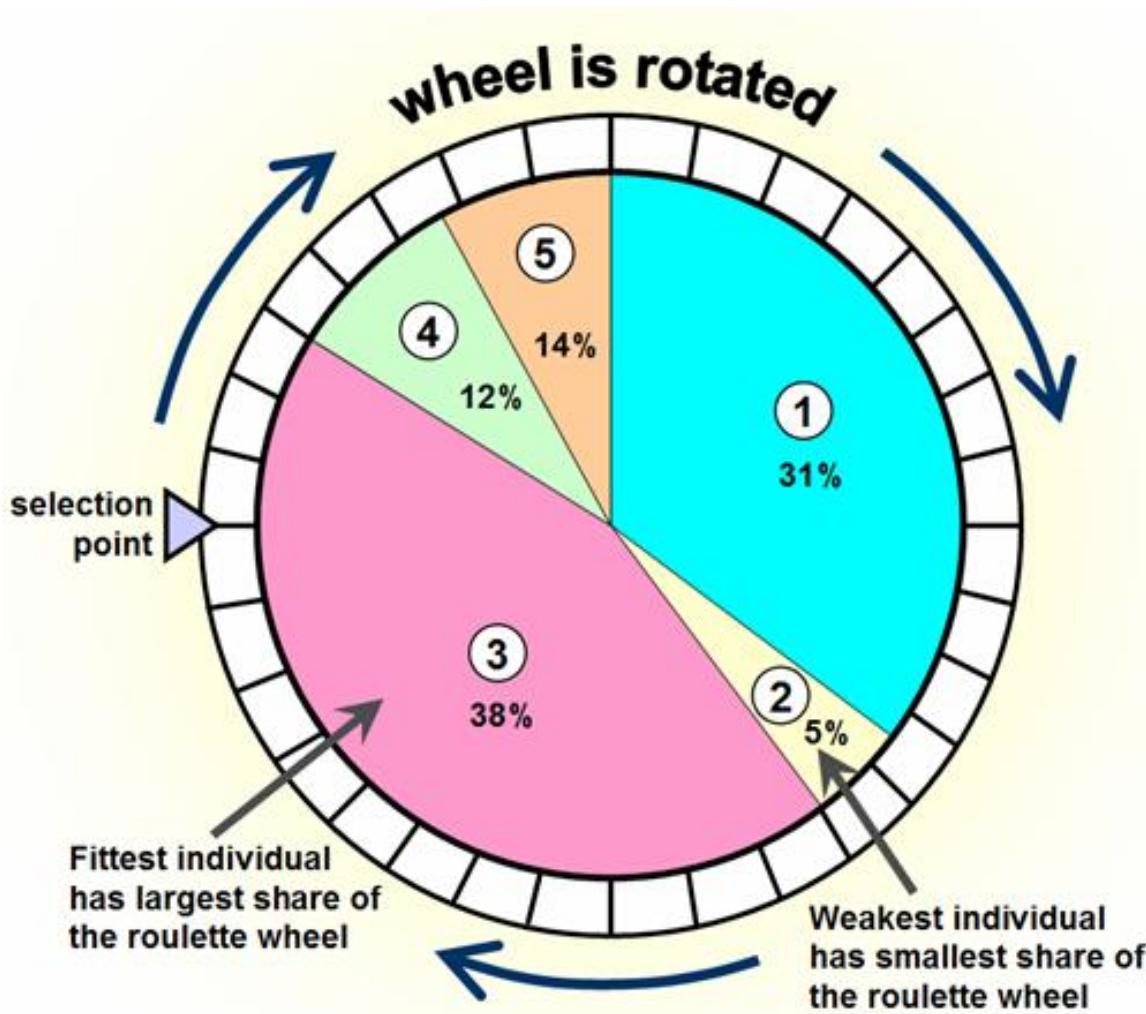
Алгоритмы. Сборка из фрагментов



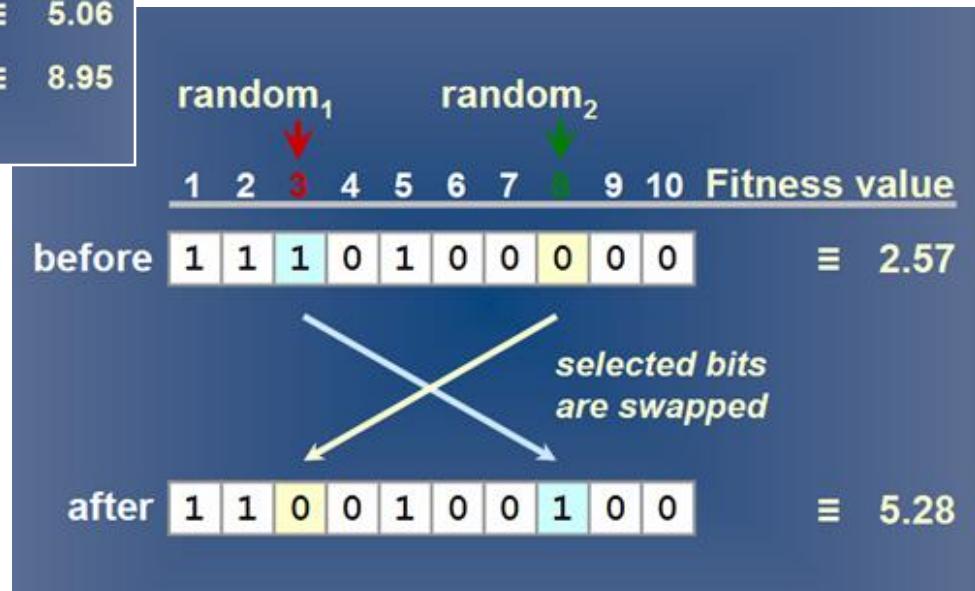
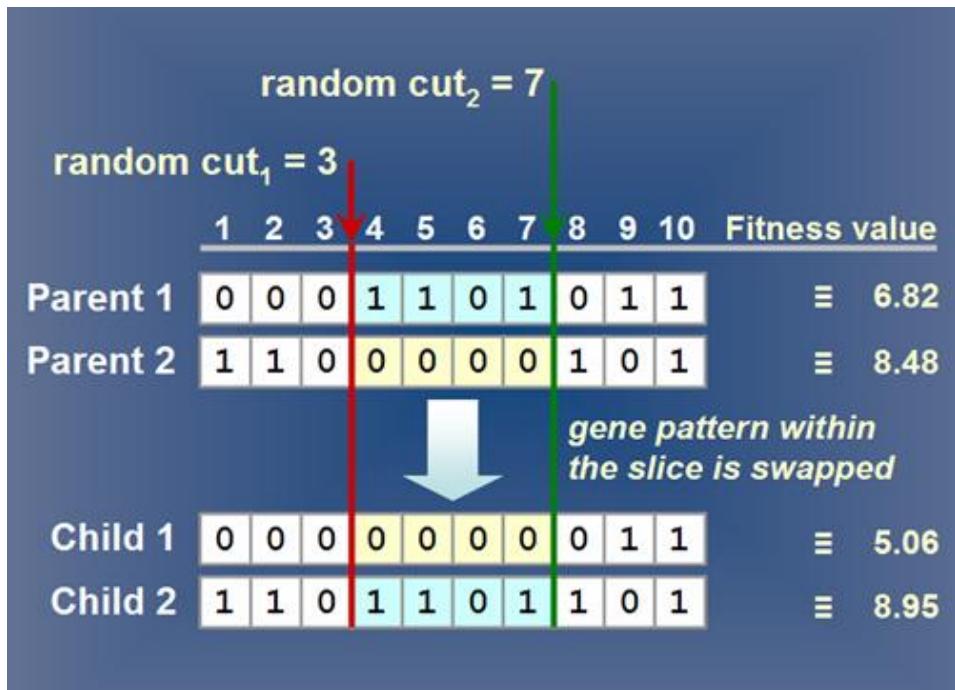
Алгоритмы. Генетический алгоритм



Алгоритмы. Генетический алгоритм



Алгоритмы. Генетический алгоритм

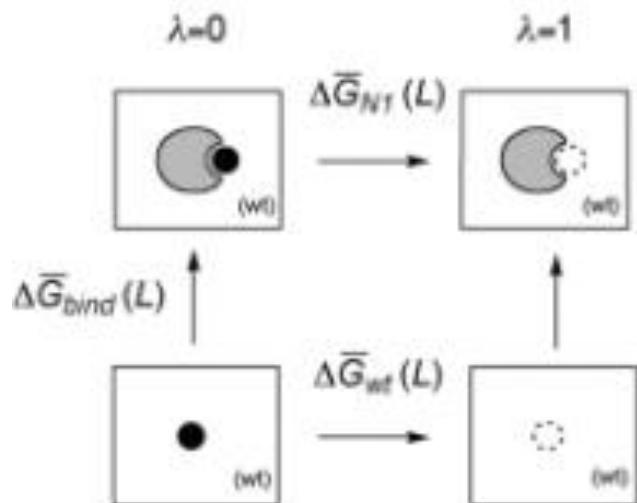


Расчет энергии связывания

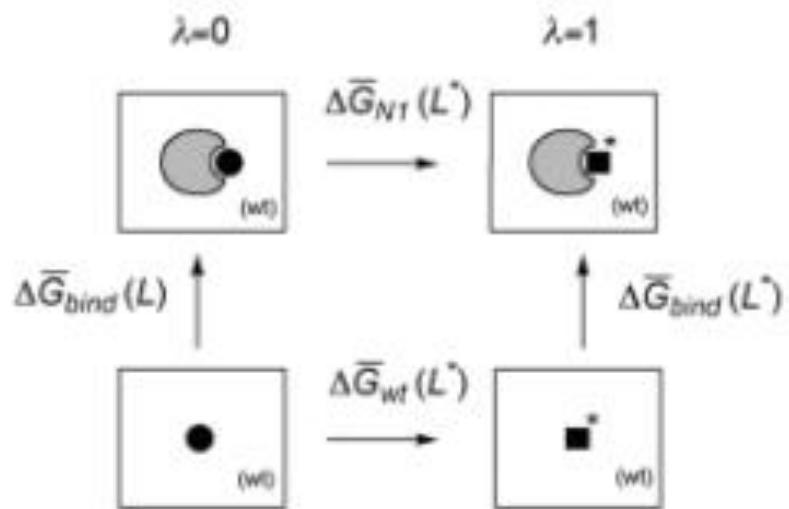
Термодинамическое интегрирование

$$\Delta G_{TI}^0 = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

a Annihilation Perturbation



b Modification Perturbation



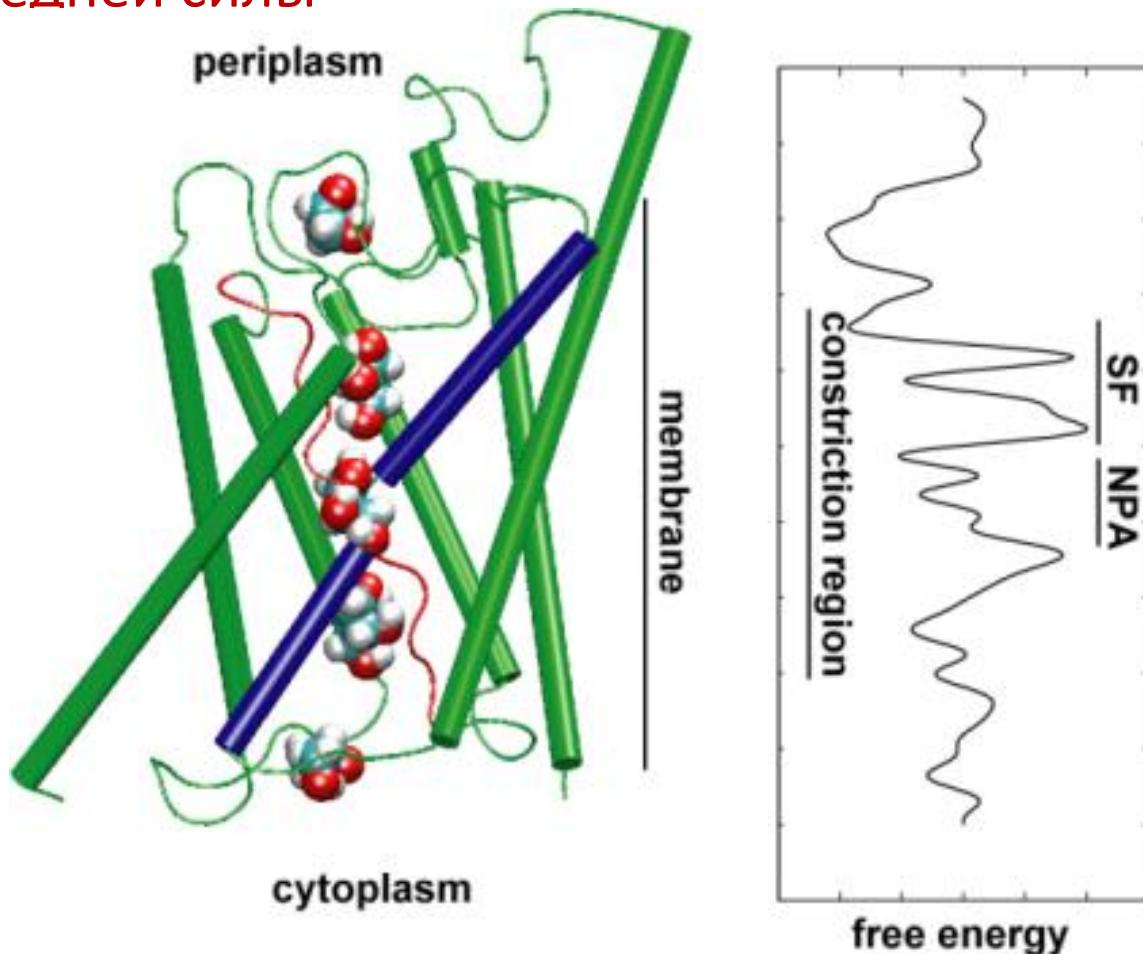
$$\Delta \bar{G}_{bind}(L) = \Delta \bar{G}_{wt}(L) - \Delta \bar{G}_{NT}(L)$$

$$\Delta \bar{G}_{bind} = \Delta \bar{G}_{bind}(L') - \Delta \bar{G}_{bind}(L) = \Delta \bar{G}_{wt}(L') - \Delta \bar{G}_{NT}(L')$$

Не используется в методах молекулярного докинга

Расчет энергии связывания

Потенциал средней силы



Не используется в методах молекулярного докинга

Оценочные функции

– Forcefield-based

- Based on terms from molecular mechanics forcefields
- GoldScore, DOCK, AutoDock

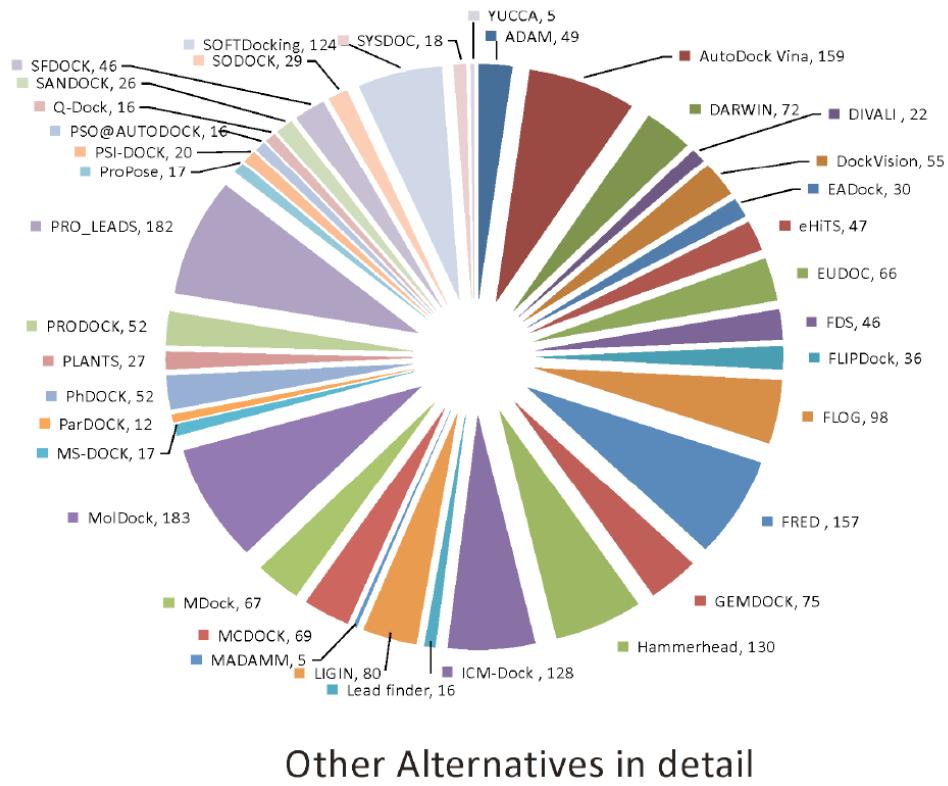
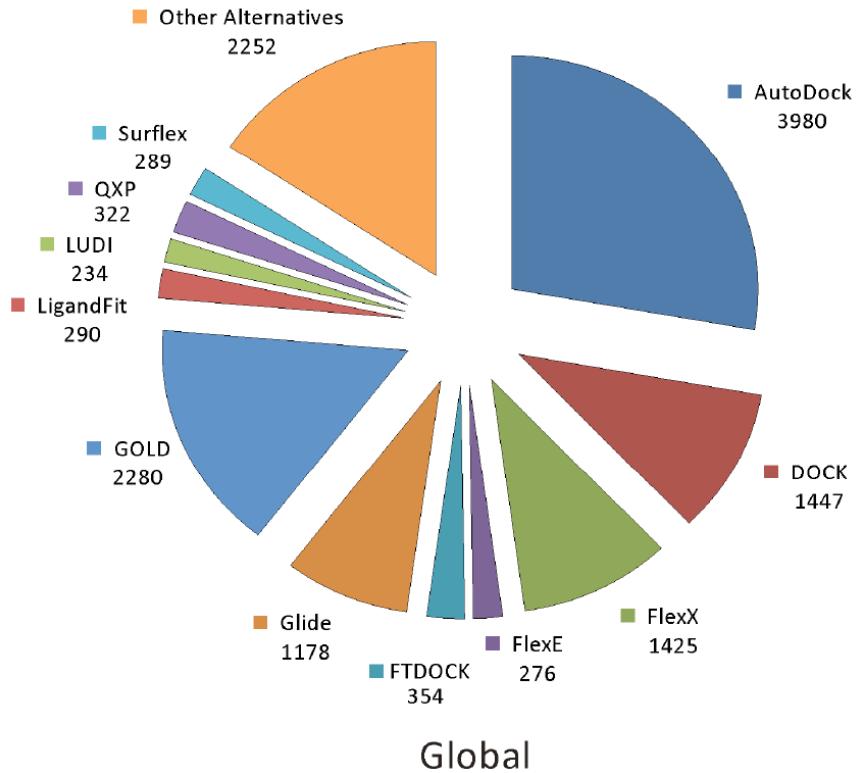
– Empirical

- Parameterised against experimental binding affinities
- ChemScore, PLP, Glide SP/XP

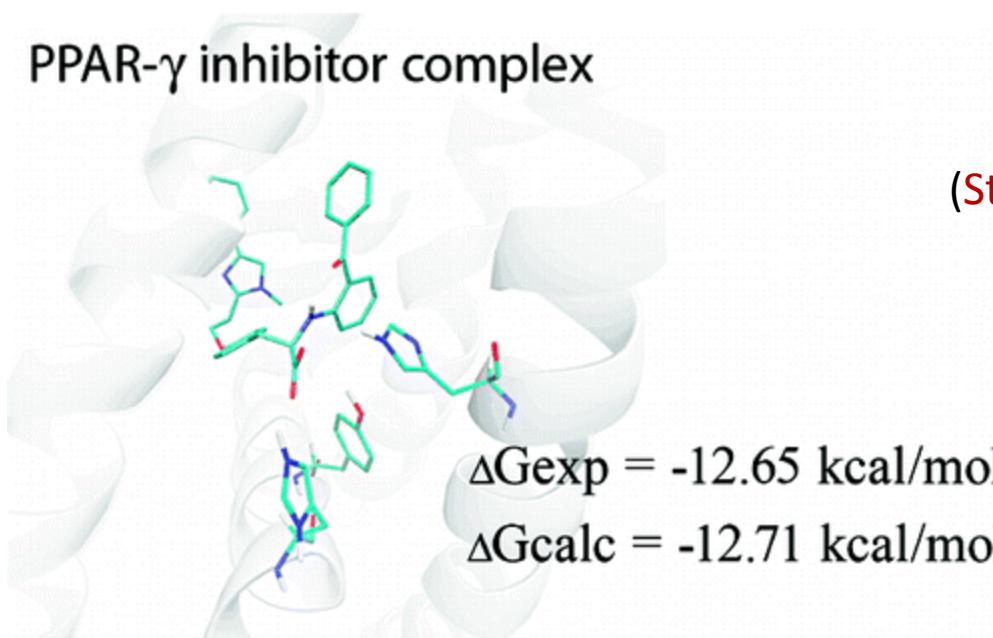
– Knowledge-based potentials

- Based on statistical analysis of observed pairwise distributions
- PMF, DrugScore, ASP

Программы для докинга



Программы для докинга. LeadFinder



Программы для докинга. SwissDock



Swiss Institute of
Bioinformatics



Molecular
Modelling Group

SwissDock

[Home](#)[Target Database](#)[Submit Docking](#)[Command Line Access](#)[Help Forum](#)[Contact](#)

Target selection

Search for targets:

ie. [PDB code](#), [protein name](#), [sequence](#), or [URL](#)
or [upload file \(max 5MB\)](#)

Ligand selection

Search for ligands:

ie. [ZINC AC](#), [ligand name](#) or category (like [scaffolds](#) or [sidechains](#)), or [URL](#)
or [upload file \(max 5MB\)](#)

Description

Job name (required):

Help

Search for a ligand

A success rate >80% can be achieved with drug-like ligand with less than 15 free dihedral angles.

You can search for ligands using a ZINC accession number (AC), its name, or its category.

ZINC AC and names will be looked for in the ZINC database.

Names and categories (scaffolds or sidechains) will be searched for in our database of 58 compounds consisting of 27 scaffolds and 31 sidechains. See [here](#) and [here](#) for further details.

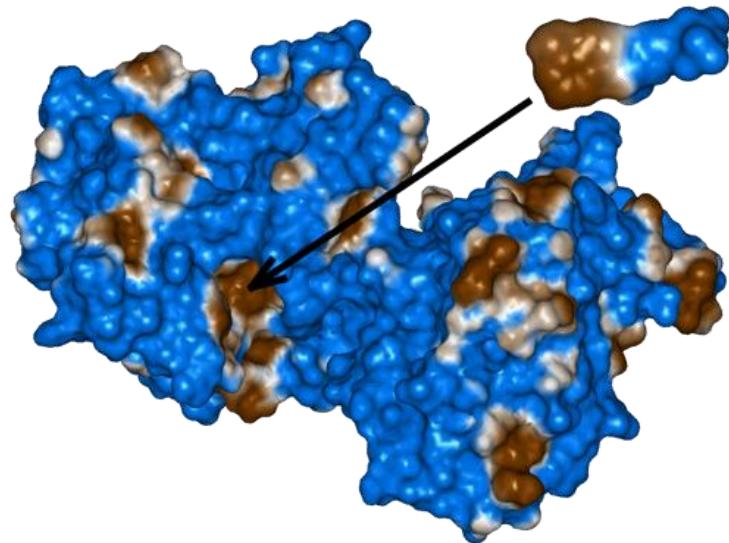
Load a ligand from a URL

You can also load a file from a URL, provided that it is either:

- a MOL2 file with all hydrogens and 3D coordinates. Check atom [chirality](#), and adjust protonation states according to your needs (e.g. carboxylate groups are usually deprotonated at physiological pH), and make sure that it has a correct topology (we recommend [UCSF Chimera](#), [OpenBabel](#), [MarvinSketch](#), [XDrawChem](#), [ChemDraw](#)).
- a ZIP file containing files in the [CHARMM format](#) (PDB/RTF/PAR).

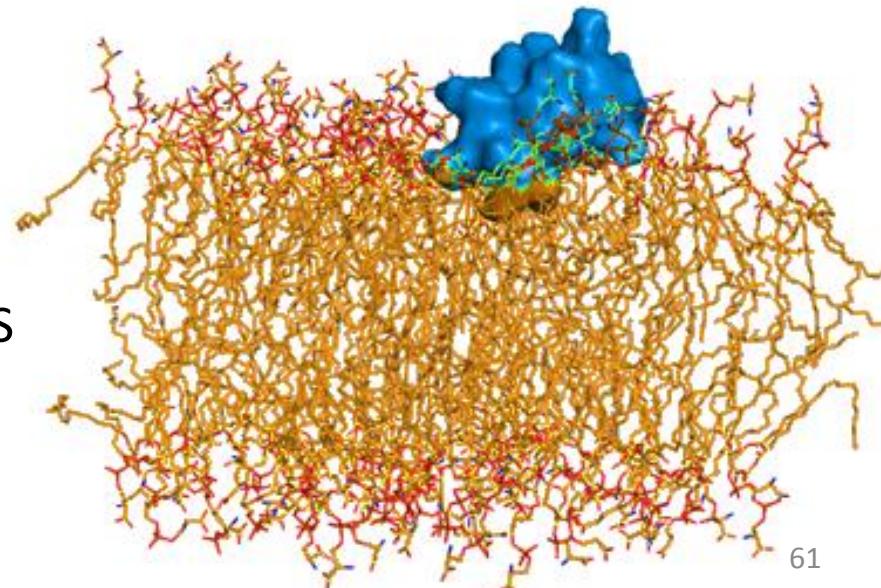
Before moving on, make sure that the protonation states are reasonable, since they have a big impact on the docking outcome.

Гидрофобные взаимодействия в биомолекулярных системах



Комплекс АТФ – Са-АТФаза
(Toyoshima et al., 2004)
■ - гидрофобный
■ - гидрофильный

Комплекс пептида pAntp с бислоем DOPS
(Polyansky et al., 2009)



Protein-Ligand ATtractions Investigation NUMerically

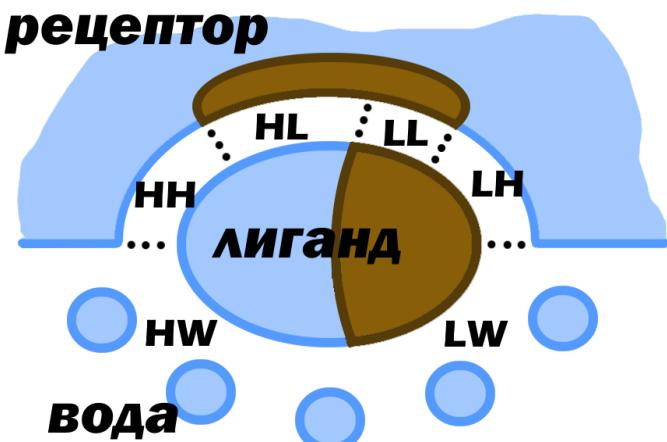
Results

<http://model.nmr.ru/platinum/>

Receptor name: 1VFP_prt.mol2

Reference ligand: 1VFP_atp.pdb.

| Ligand name | r.m.s.d. | IFp | # H-bonds | S _{L/L} | S _{H/H} | S _{buried} | S _{total} | Match ¹ | Match ² | # Stack. | # Stack. Gua-π |
|--------------------|----------|-------|-----------|------------------|------------------|---------------------|--------------------|--------------------|--------------------|----------|----------------|
| 1VFP_atp | 0.00 | 1 | 5.36 | 0.95 | 248.23 | 290.06 | 319.68 | 0.7794 | 0.0424 | 1.933 | 0 |
| gold_soln_atp_m1_2 | 1.94 | 0.714 | 5.55 | 5.25 | 248.80 | 289.11 | 322.97 | 0.7866 | 0.208 | 0.804 | 0 |
| gold_soln_atp_m1_5 | 8.16 | 0.208 | 2.04 | 0.00 | 234.87 | 287.97 | 335.25 | 0.7006 | 0 | 0 | 0.121 |
| gold_soln_atp_m1_4 | 1.19 | 0.632 | 4.27 | 0.25 | 244.18 | 280.63 | 313.54 | 0.7796 | 0.0128 | 1.891 | 0 |
| gold_soln_atp_m1_1 | 1.21 | 0.684 | 4.16 | 0.06 | 242.53 | 279.18 | 309.37 | 0.7842 | 0.0034 | 1.742 | 0 |
| gold_soln_atp_m1_3 | 3.89 | 0.55 | 1.87 | 1.77 | 229.30 | 252.59 | 311.52 | 0.7418 | 0.1083 | 1.659 | 0 |



PLATINUM

(Pyrkov, Efremov et al., 2009)

Обучающие наборы

Положение лигандов и константы связывания уже известны

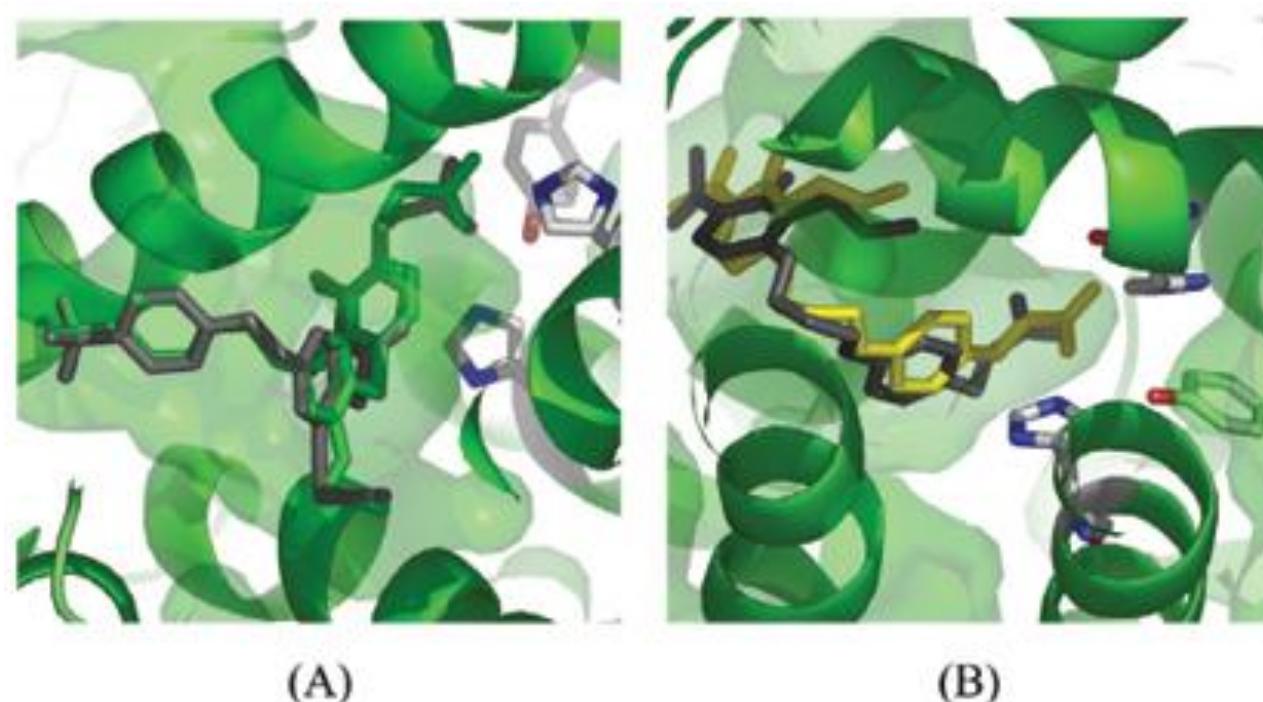
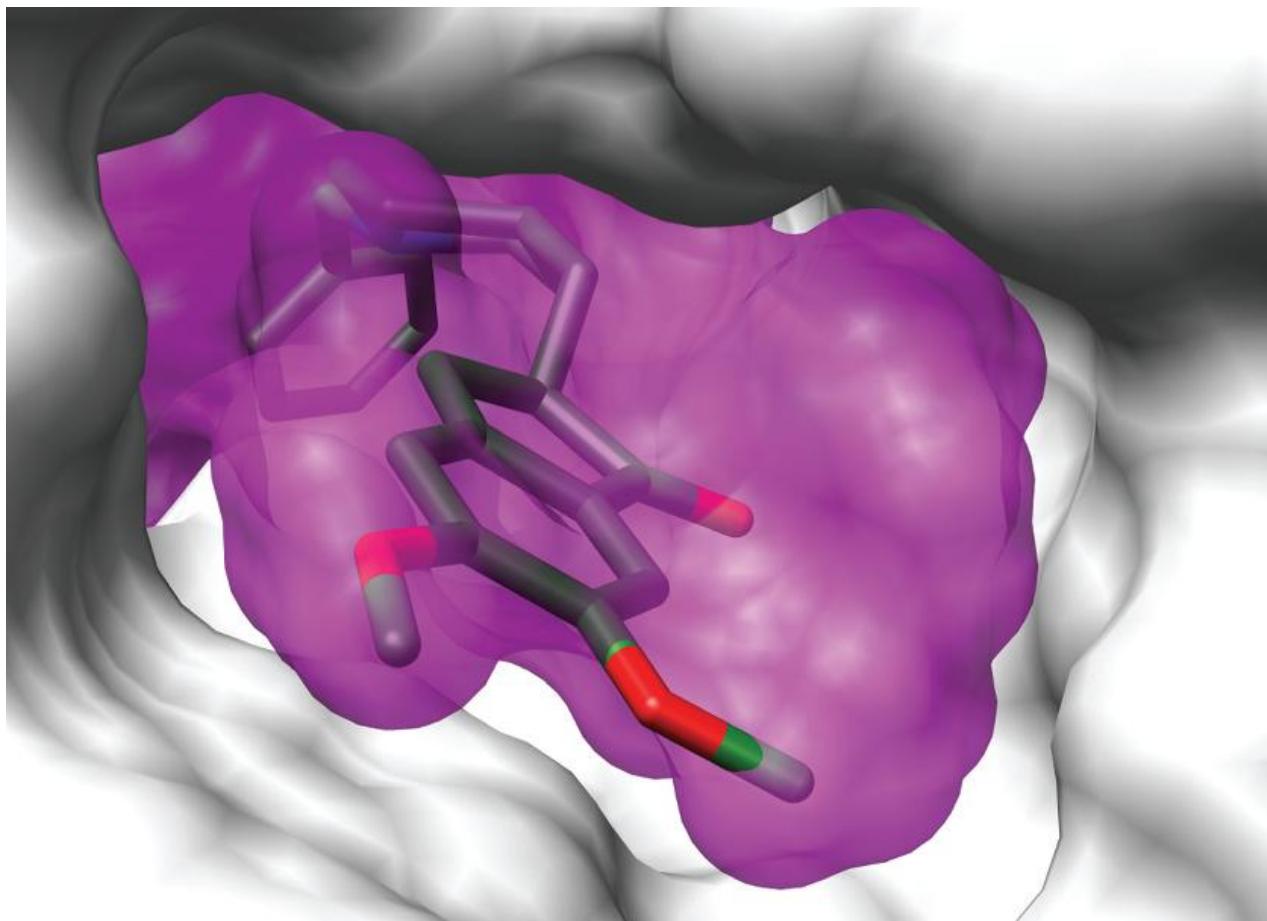


Figure 4. (A) Redocking and (B) crossdocking results. Both crystallographic ligands are shown in gray and the docked ligands are in green (D32) and yellow (L41), respectively. See online for color image.

Аннотация функции



Активный центр на поверхности молекулы холинэстеразы (pdb-код 1EVE) с расположенной в нем молекулой арисепта.

Докинг макромолекул

