ВВЕДЕНИЕ В БИОИНФОРМАТИКУ

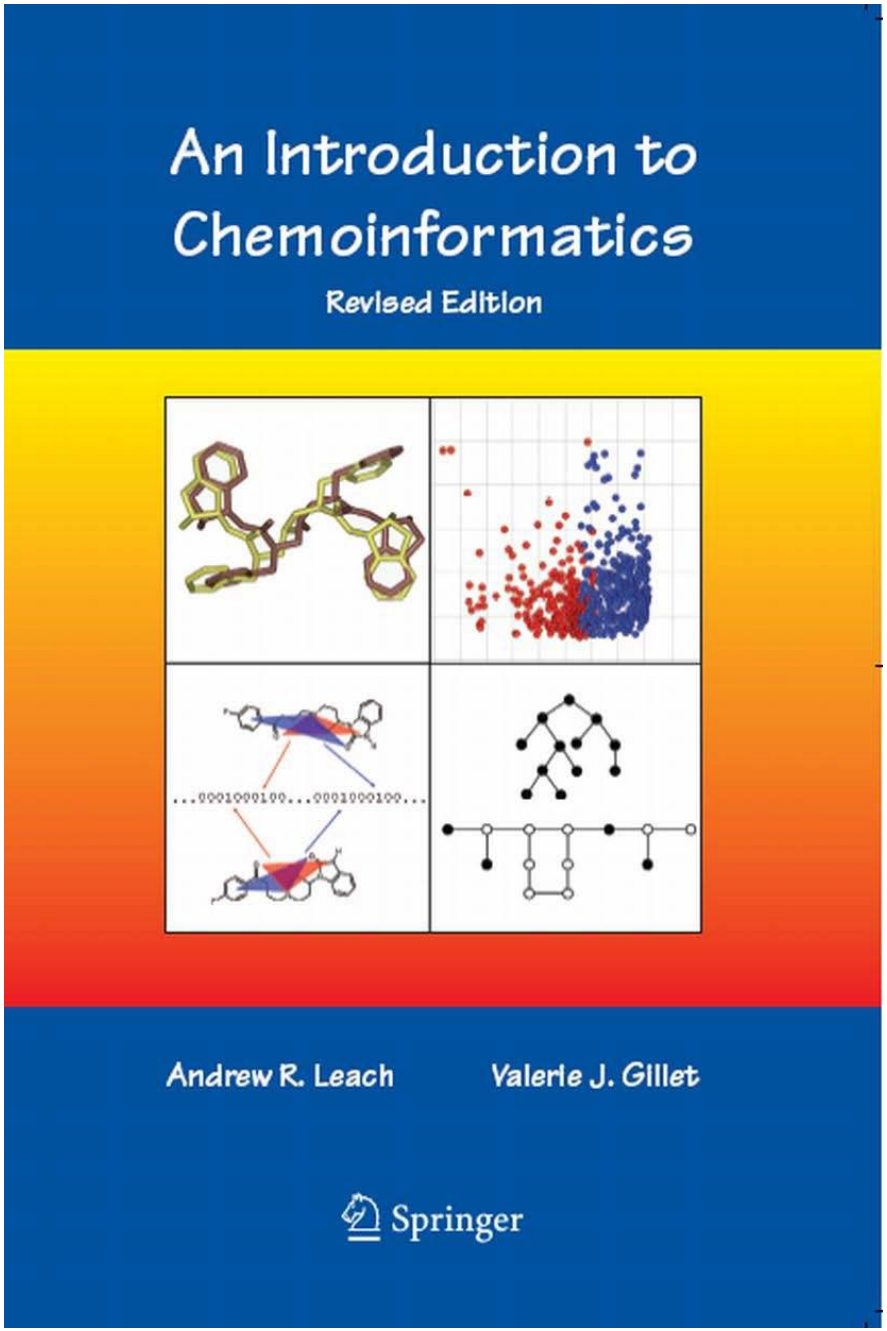
Лекция №21

Хемоинформатика
и
виртуальный скрининг

Новоселецкий Валерий Николаевич
к.ф.-м.н., доц. каф. биоинженерии
valery.novoseletsky@yandex.ru

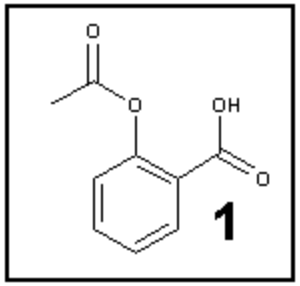
Сайт курса http://intbio.org/bioinf2018-2019

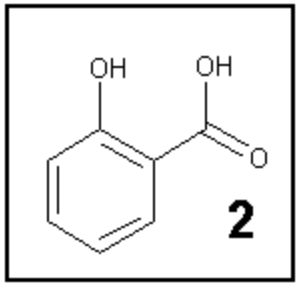
Что почитать?

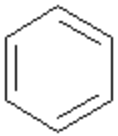
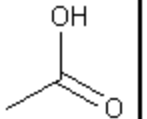
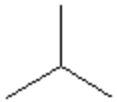
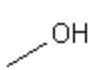
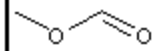
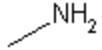
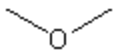
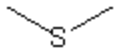


Определение сходства молекул

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

TANIMOTO COEFFICIENT = A / (A + B + C)
= 3 / (3 + 2 + 0) = 0.6 or 60%

О мерах сходства

Коэффициент Танимото (для битовых строк Xi и Yi) (1960):

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

Λ – логическое И
V – логическое ИЛИ

A	1	0	1	1	0	1	$\lvert A\rvert = 4$
B	1	1	0	1	0	0	$\lvert B\rvert = 3$
$A \wedge B$	1	0	0	1	0	0	$\lvert A \wedge B\rvert = 2$
$A \vee B$	1	1	1	1	0	1	$\lvert A \vee B\rvert = 5$

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

Предсказание биологической активности

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

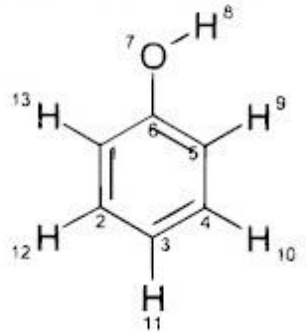
Downloads

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]



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.

MNA – multilevel neighborhoods of atoms

(Filimonov , Poroikov et al., 1999)

PASS

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

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

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.

PASS

Предсказание биологической активности

Ограничения предсказательной способности

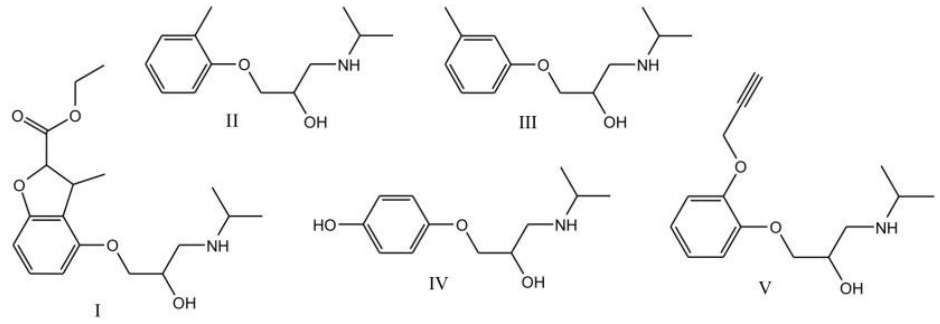
PASS cannot predict the activity spectrum for essentially new compound if all its descriptors are new and so they don't occur in the training set. If a compound has more then 2 new descriptors it is rather new and prediction results may be considered as pilot.

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 then 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).



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 then 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.

Предсказание биологической активности

Swiss Institute of Bioinformatics

SwissTargetPrediction

Home FAQ Help Download Contact Disclaimer

This website allows you to predict the targets of a small molecule. Using a combination of 2D and 3D similarity measures, it compares the query molecule to a library of 280'000 compounds active on more than 2000 targets of 5 different organisms.

The webserver is described in detail in our article: SwissTargetPrediction: a webserver for target prediction of bioactive small molecules, *Nucl. Acids Res.* (2014). For technical information about the prediction algorithm, you can refer to this article: Shaping the interaction landscape of bioactive molecules, *Bioinformatics* (2013) 29:3073-3079.

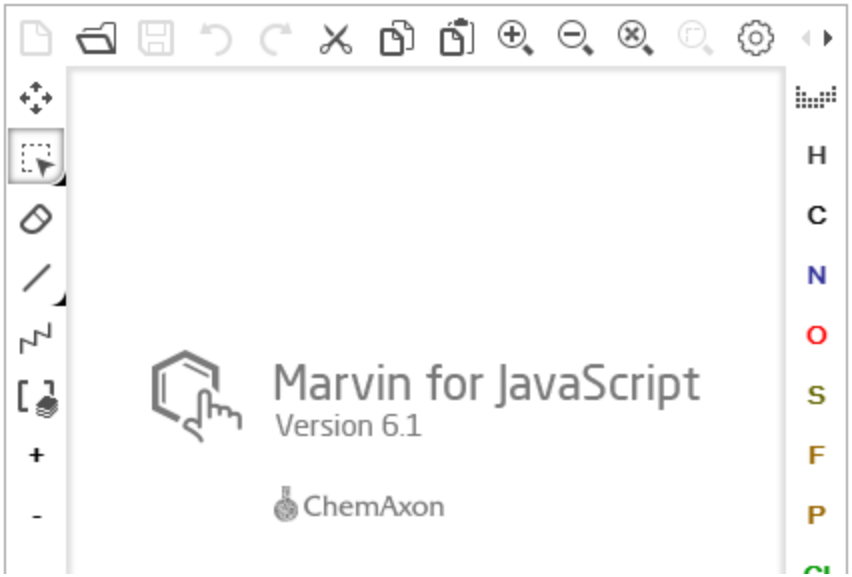
Choose an organism

- (•) Homo sapiens
- ( ) Mus musculus
- ( ) Rattus norvegicus
- ( ) Bos taurus
- ( ) Equus caballus

Paste a SMILES in this box, or draw a molecule

Examples: ▾ Clear

=



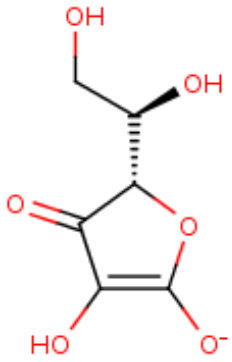
Предсказание биологической активности

List of predicted targets

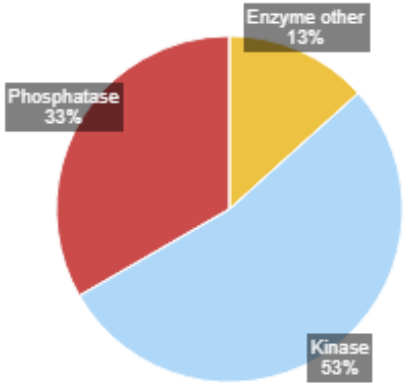
These targets have been predicted using the method described in:

Gfeller D., Michielin O. & Zoete V. Shaping the interaction landscape of bioactive molecules, *Bioinformatics* (2013) 29:3073-3079.

Query Molecule



General Target Classes



Retrieve data:

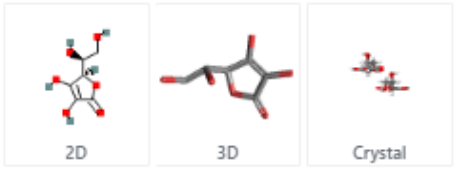
Target	Common name	Uniprot ID	ChEMBL ID	Probability*	# sim. cmpds (3D / 2D)	Target Class
Tyrosyl-DNA phosphodiesterase 1	TDP1	Q9NUW8	CHEMBL1075138		0 / 2	Enzyme
Tubulin--tyrosine ligase	TTL	Q8NG68	CHEMBL5549		0 / 12	Enzyme
Protein kinase C alpha type	PRKCA	P17252	CHEMBL299		0 / 141	Ser_Thr Kinase
Protein kinase C delta type regulatory subunit *(by homology)*	PRKCD	Q05655	CHEMBL2996		0 / 143	Ser_Thr Kinase
Protein kinase C theta type *(by homology)*	PRKCQ	Q04759	CHEMBL3920		0 / 143	Ser_Thr Kinase

Базы данных химических соединений

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

Базы данных химических соединений

Ascorbic acid

PubChem CID:	54670067
Structure:	 2D 3D Crystal
Chemical Safety:	DATASHEET AVAILABLE: Laboratory Chemical Safety Summary (LCSS)
InChI Key:	CIWBSHSKHKDKBQ-JLAZNSOCSA-N
Molecular Formula:	$C_6H_8O_6$ or $HC_6H_7O_6$
UNII:	PQ6CK8PD0R
Depositor-Supplied Synonyms:	l-ascorbic acid ascorbic acid vitamin C 50-81-7 L(+)-Ascorbic acid More...
Molecular Weight:	176.124 g/mol
Dates:	Modify: 2019-03-30 Create: 2011-12-26

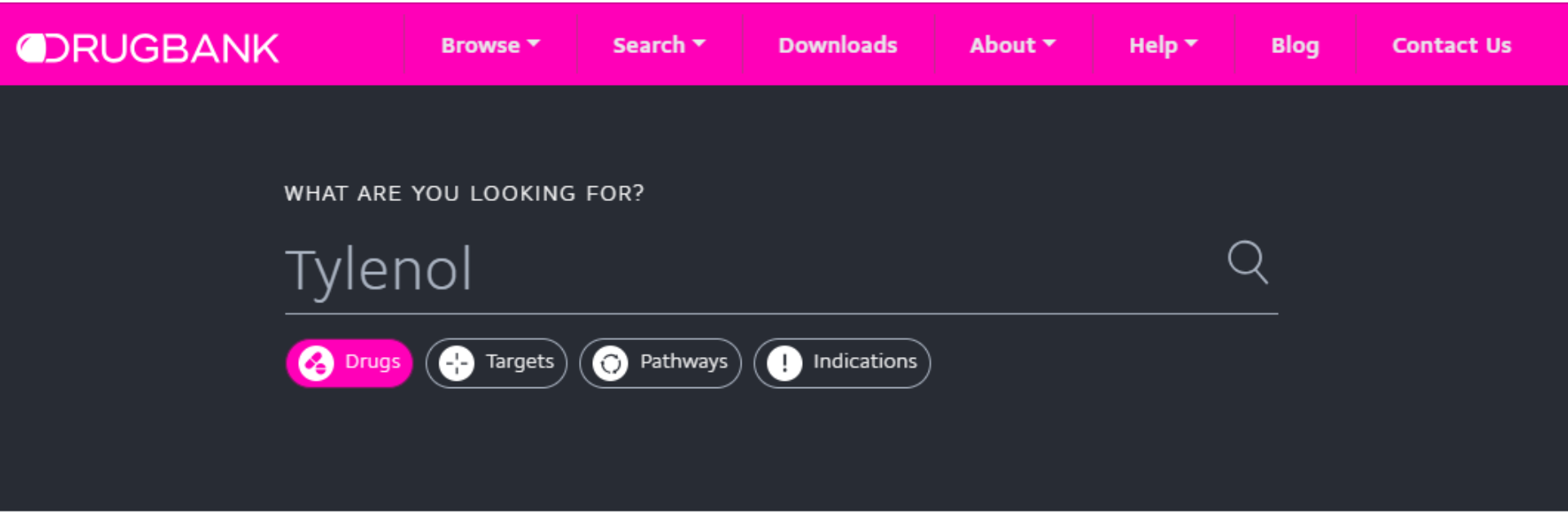
Share Tweet Email

Cite Download

CONTENTS

- Title and Summary
- 1 Structures
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Spectral Information
- 5 Related Records
- 6 Chemical Vendors
- 7 Drug and Medication Information
- 8 Food Additives and Ingredients
- 9 Agrochemical Information
- 10 Pharmacology and Biochemistry
- 11 Use and Manufacturing
- 12 Identification
- 13 Safety and Hazards
- 14 Toxicity
- 15 Literature
- 16 Patents
- 17 Biomolecular Interactions and Pathways

Базы данных химических соединений



DRUGBANK

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information.

Базы данных химических соединений

UCSF University of California, San Francisco | About UCSF | Search UCSF | UCSF Medical Center

Shoichet Laboratory docking.org

ZINC 12

Not Authenticated — sign in

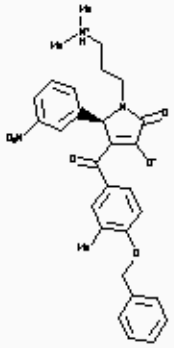
Active cart: Temporary Cart (0 items)

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.

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. 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).

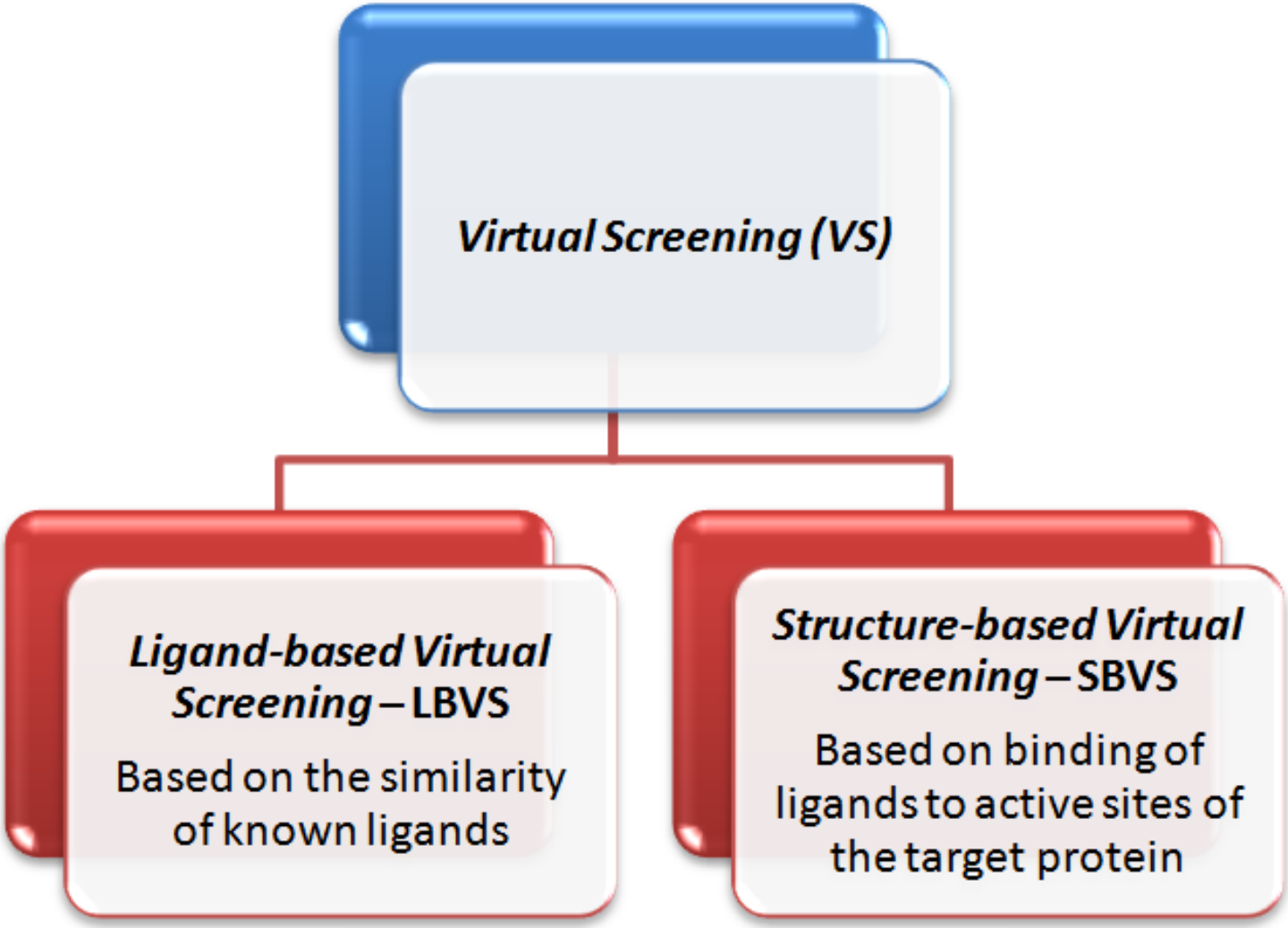
Molecule of the Minute 8817205



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

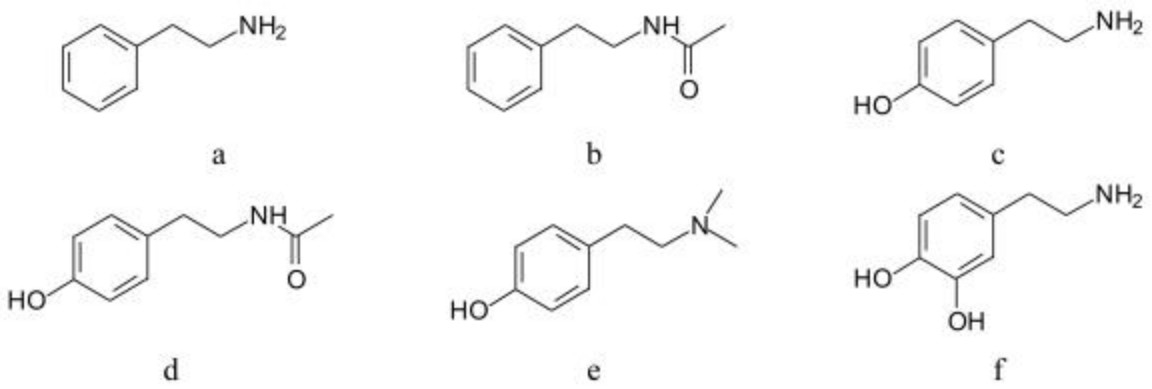
Structure/Draw Physical Properties Catalogs & Vendors ZINC IDS Targets Rings Combination

Виртуальный скрининг

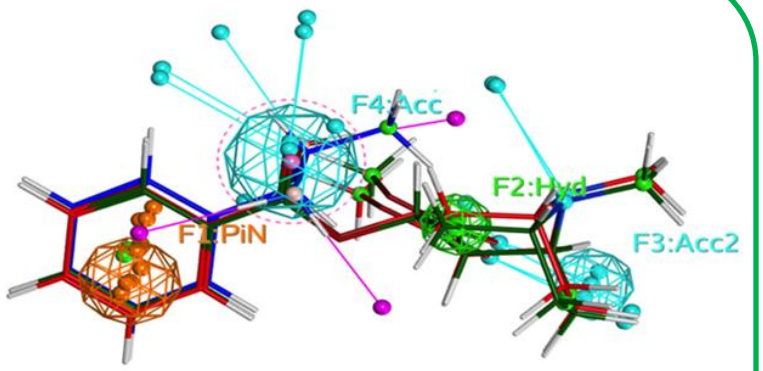


Виртуальный скрининг

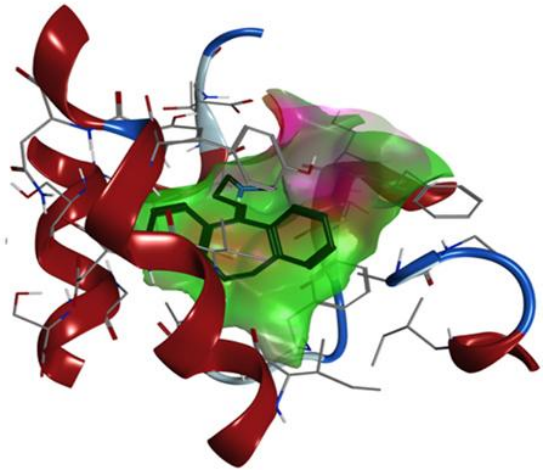
LBVS



Отбор по формальным признакам



Фармакофорный поиск



Молекулярный докинг

SBVS

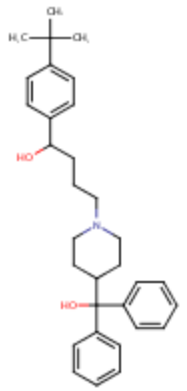
Отбор по формальным признакам. SwissSimilarity

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.

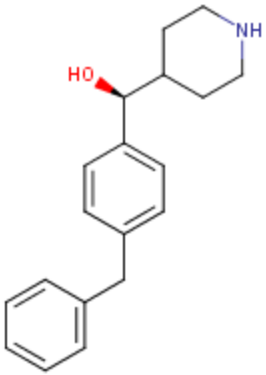
Query Molecule



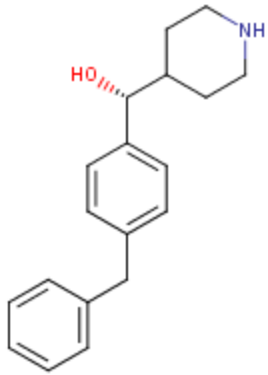
Results

Retrieve data:

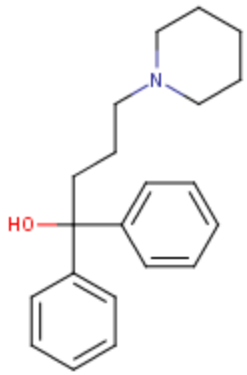
ZINC80181216
Score : 0.983



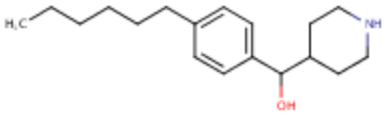
ZINC80181212
Score : 0.983



ZINC00968266
Score : 0.972



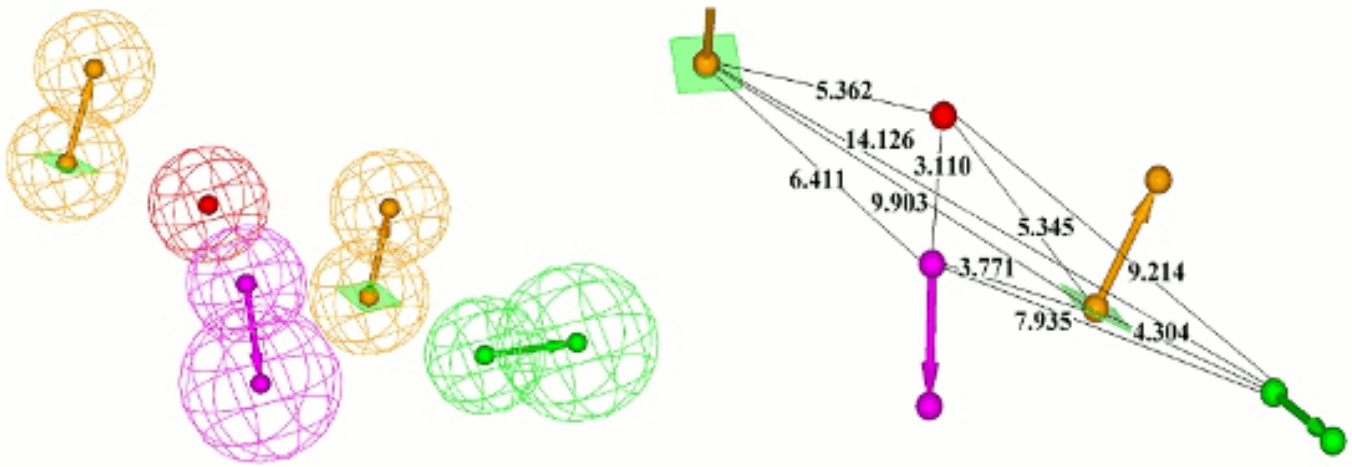
ZINC74369400
Score : 0.969



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).

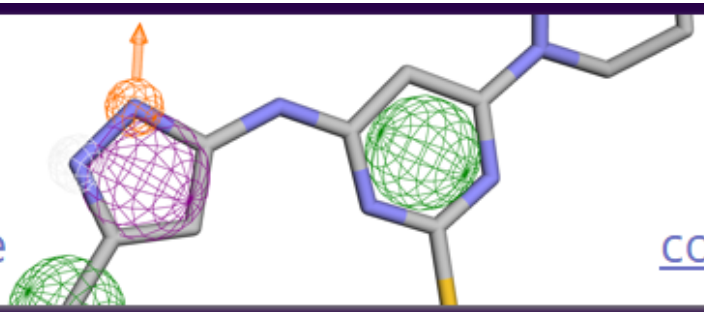


Pharmacophore model for β2-adrenoreceptor agonists generated by HipHop. Features are portrayed as mashed spheres, color-coded as follows: green, hydrogen-bond acceptor, magenta, hydrogen-bond donor, orange, aromatic ring, red, positive ionizable feature.

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:

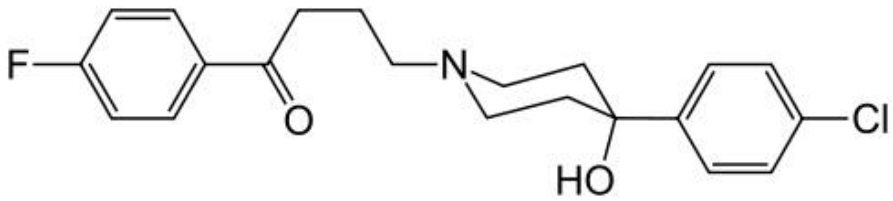
password:

log in

register new account

log in as guest

Фармакофорный поиск. PharmIt



Галоперидол

Search PubChem

Pharmacophore Search -> Shape Filter

Load Receptor... Load Features...

Pharmacophore

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

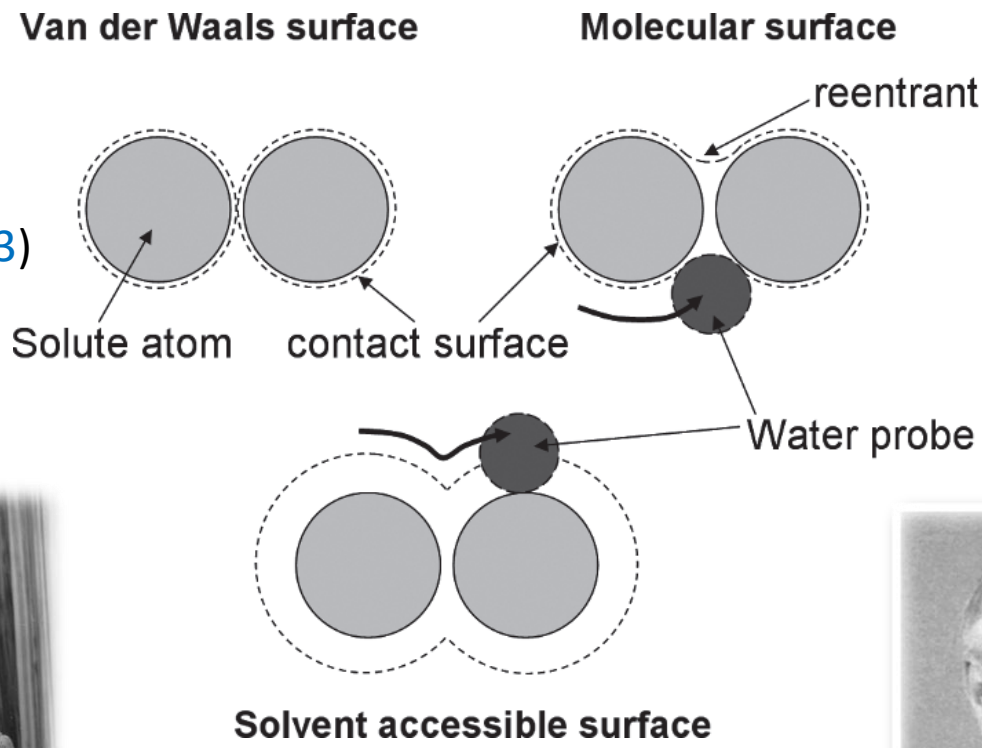
Showing 1 to 10 of 10 hits

Previous 1 Next

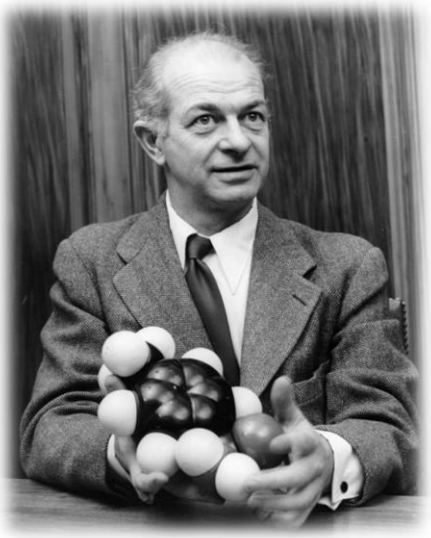
Minimize Save...

Молекулярные поверхности

(Corey & Pauling, 1953)



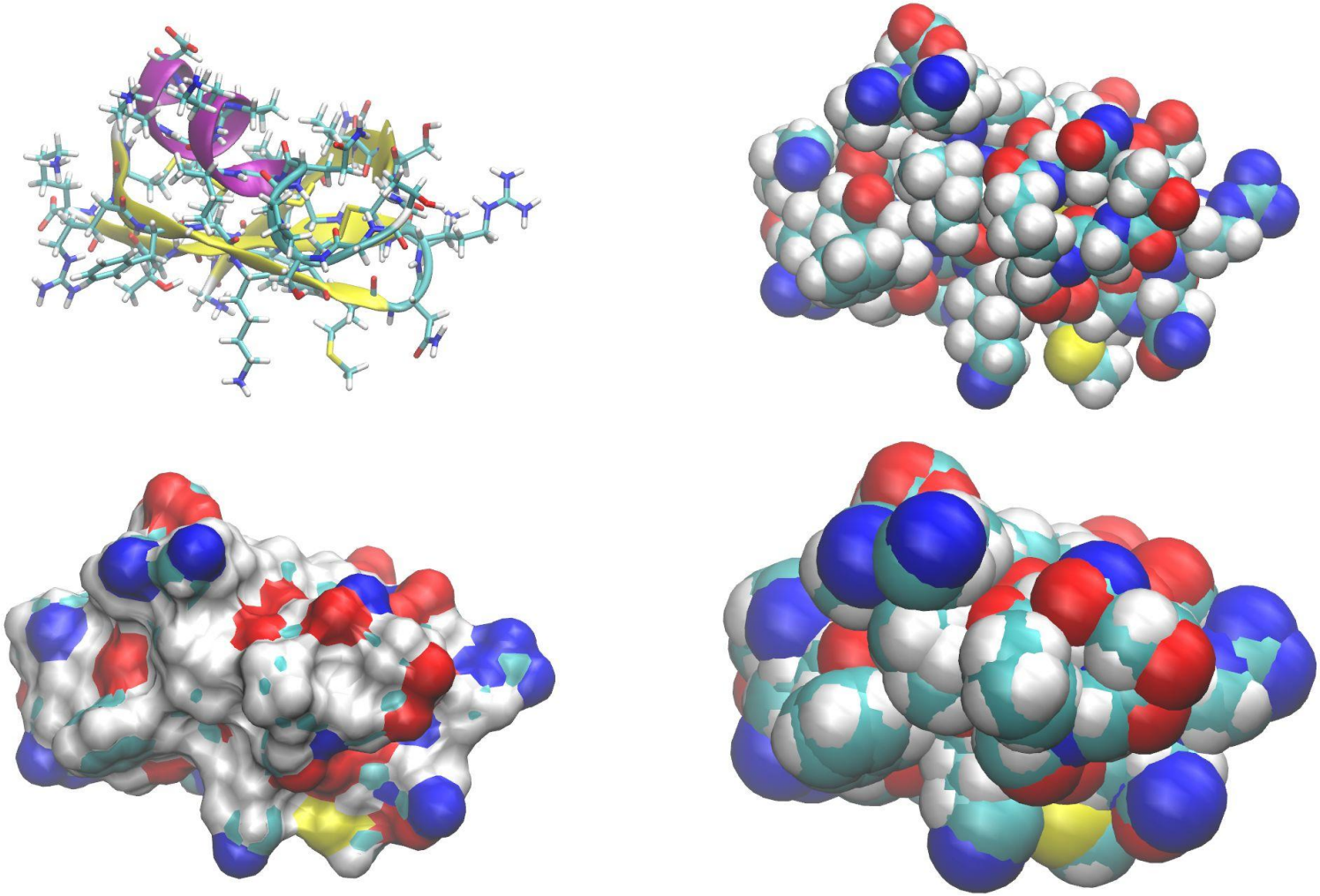
(Richards, 1977)



Linus Pauling
(1901 - 1994)

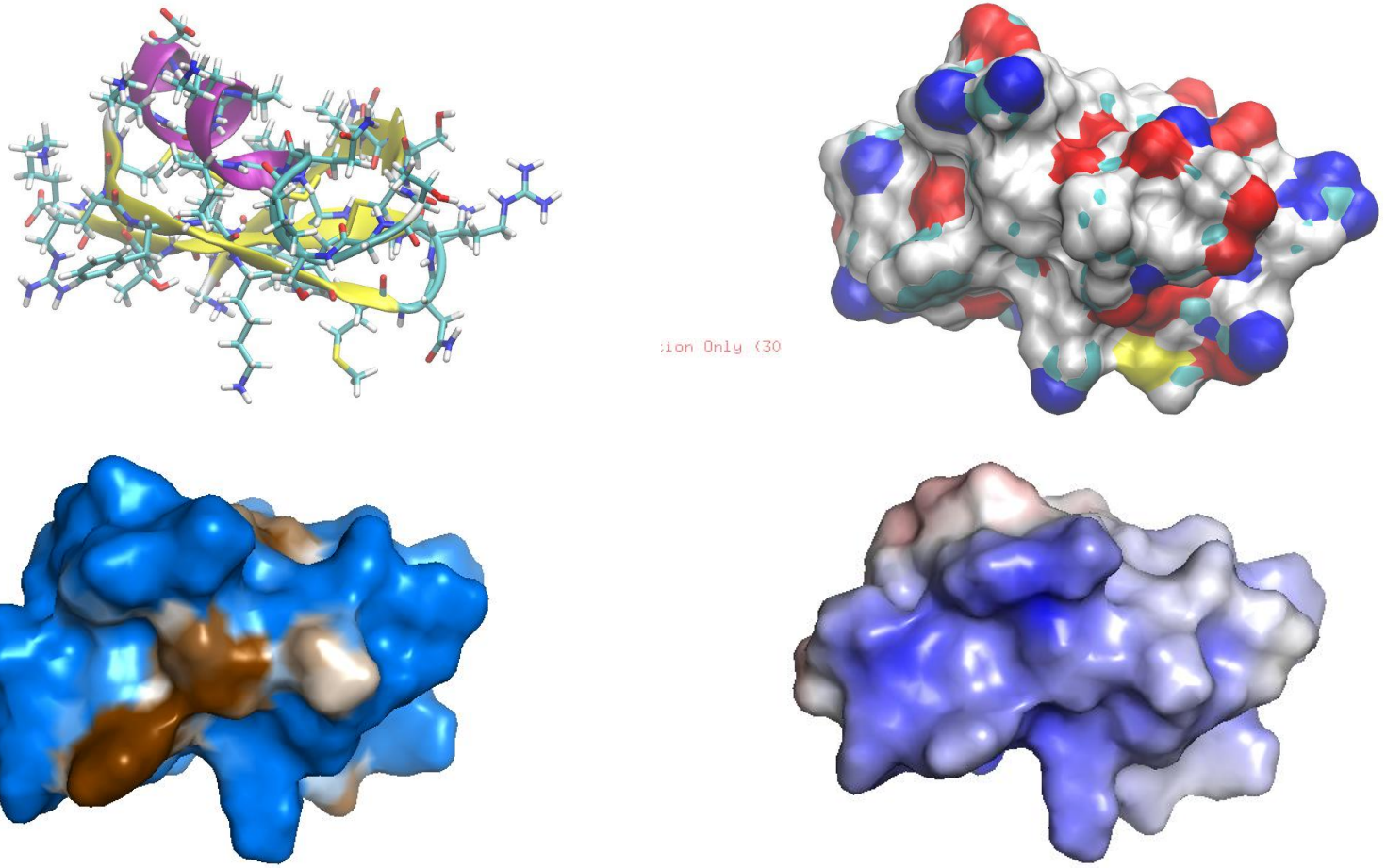
Frederic Richards
(1925 - 2009)

Молекулярные поверхности



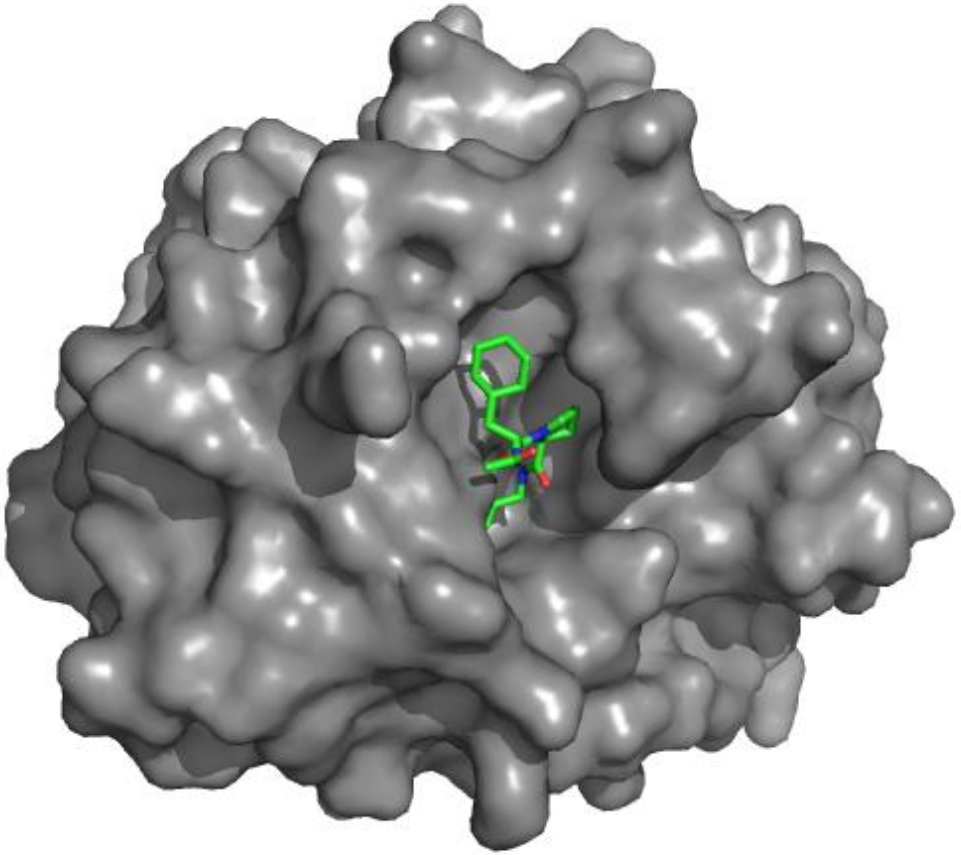
Структура аджитоксина-2 (pdb-код 1agt) в ленточном и стержневом представлении, в виде сфер ван-дер-Ваальса, а также ее молекулярная поверхность и поверхность, доступная растворителю (SAS)

Молекулярные поверхности

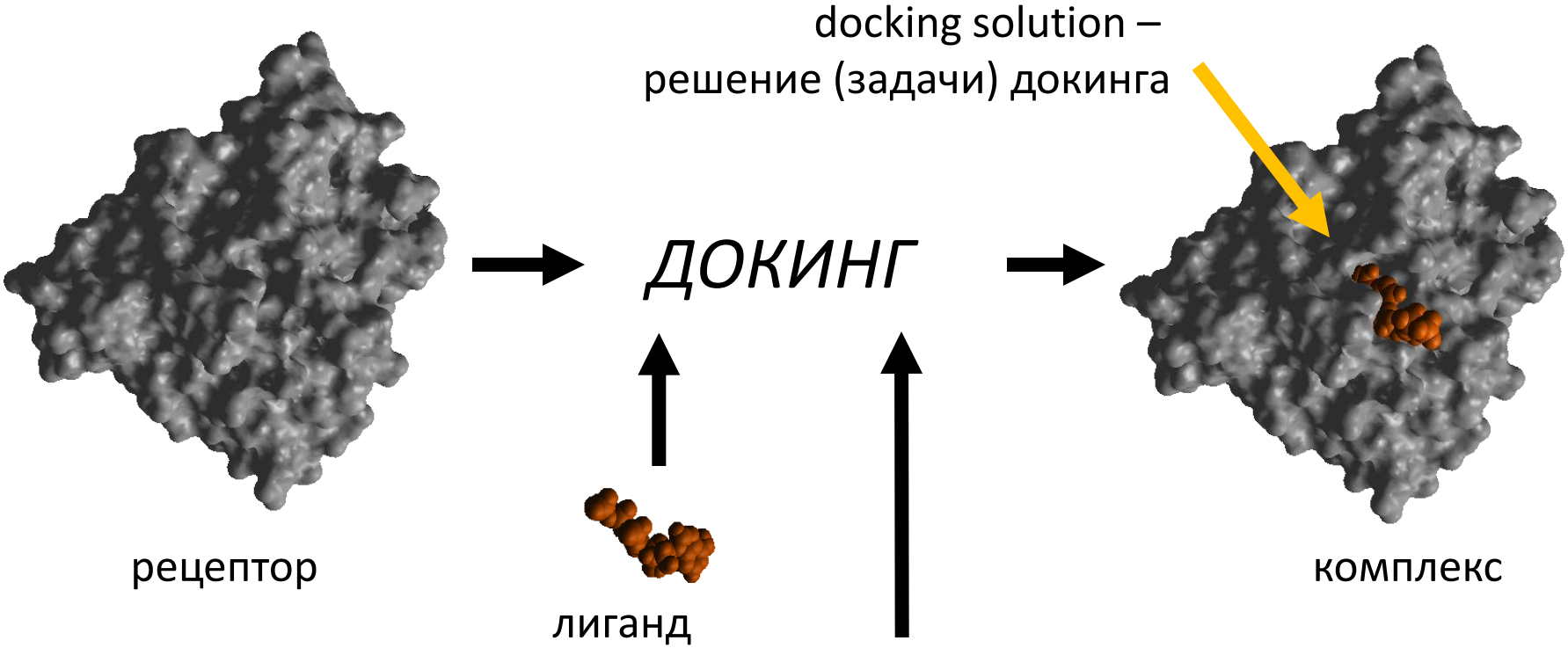


Структура аджитоксина-2 и её молекулярная поверхность, раскрашенная по химическому элементу, гидрофобности и электростатическому потенциалу (синий цвет соответствует положительным значениям, красный отрицательным)

Молекулярный докинг

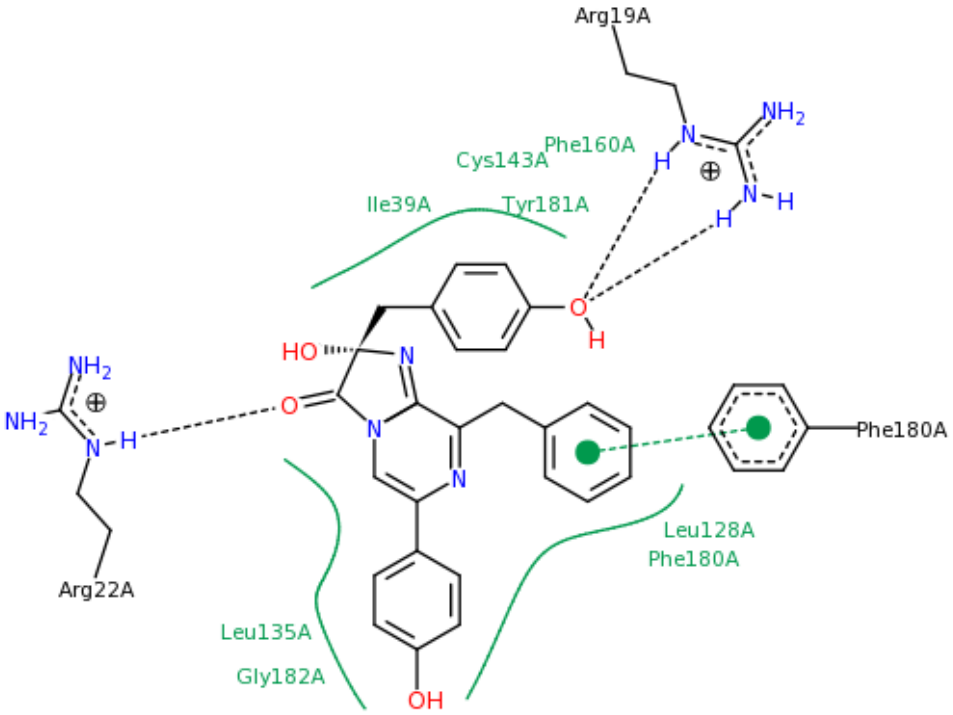


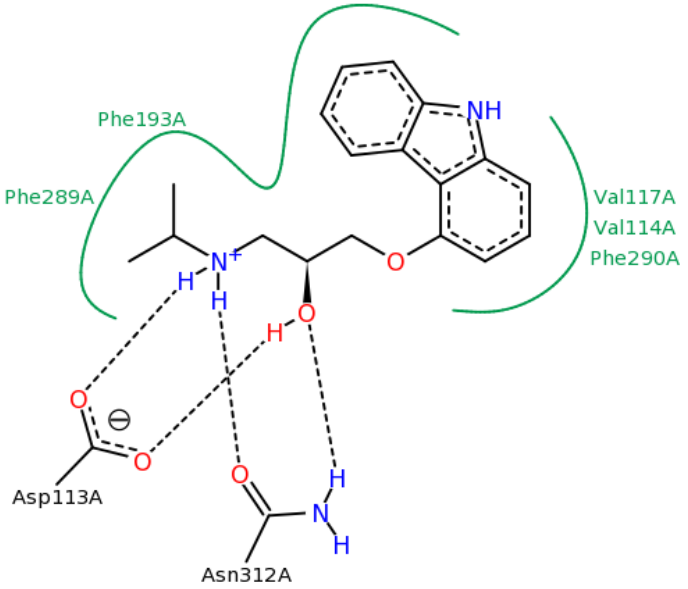
Общая постановка задачи



оценочная функция:
водородные связи, гидрофобные взаимодействия, стэкинг-взаимодействия, …

Взаимодействия

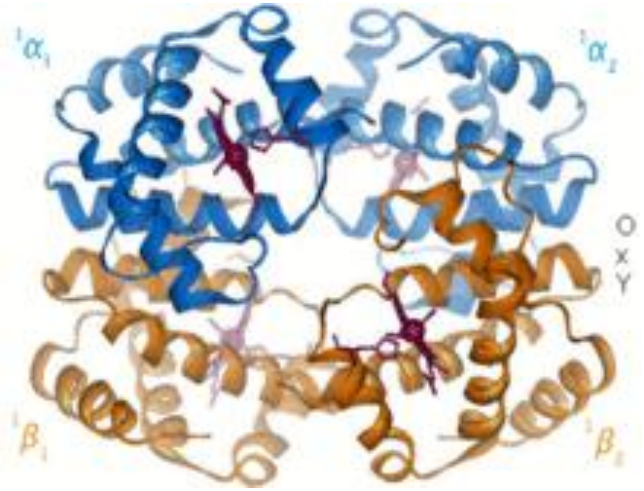




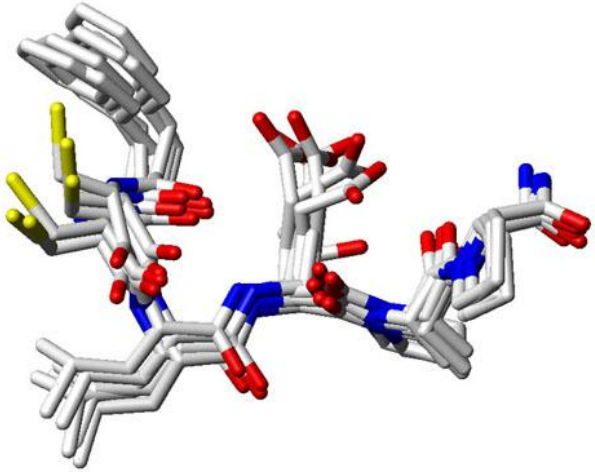
Представление белков и лигандов

Гибкость белка:

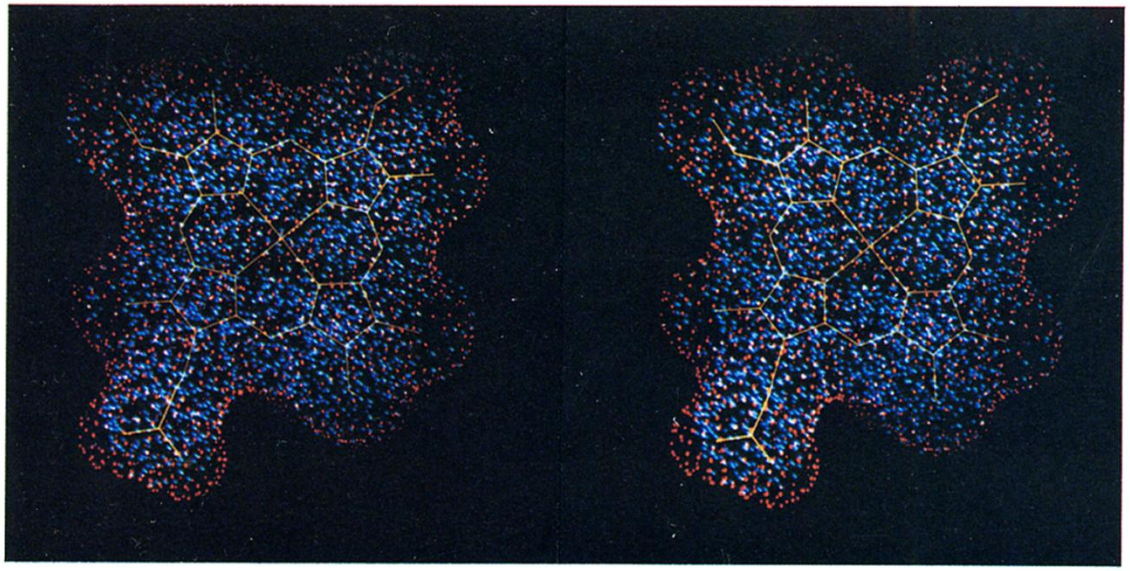
- быстрые движения малого масштаба – движение боковых цепей и петель
- медленные движения крупного масштаба – движение доменов
- ренатурация частично развернутых белков



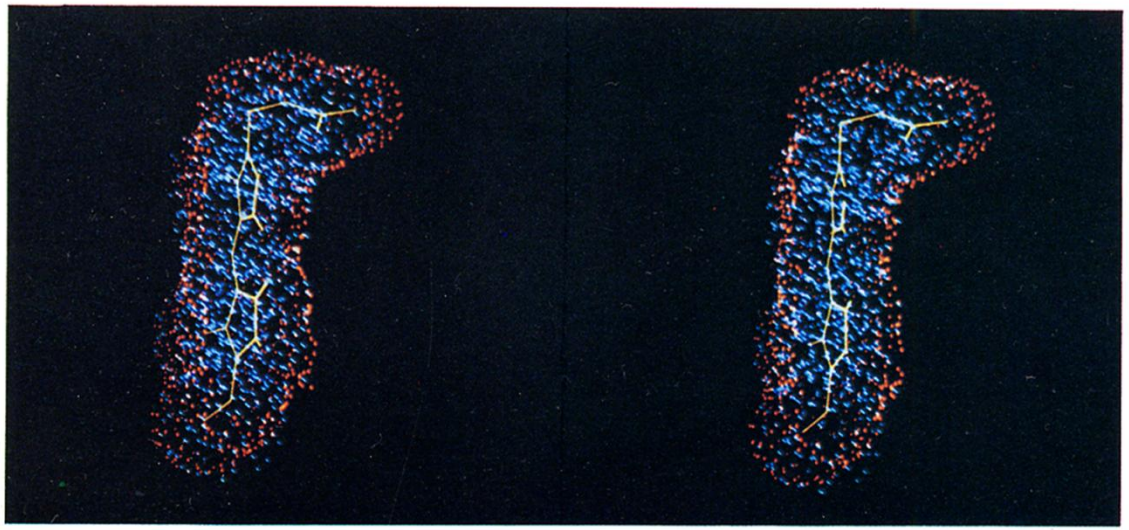
Гибкость лиганда



A Geometric Approach to Macromolecule-Ligand Interactions (1982)

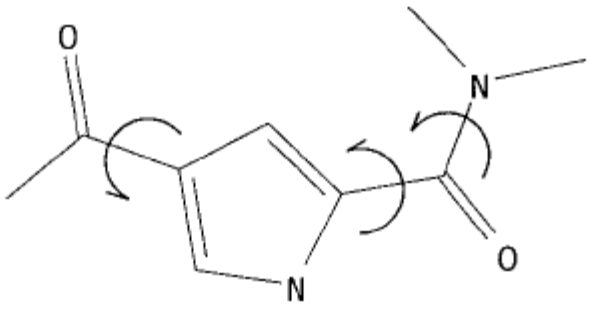


(a)

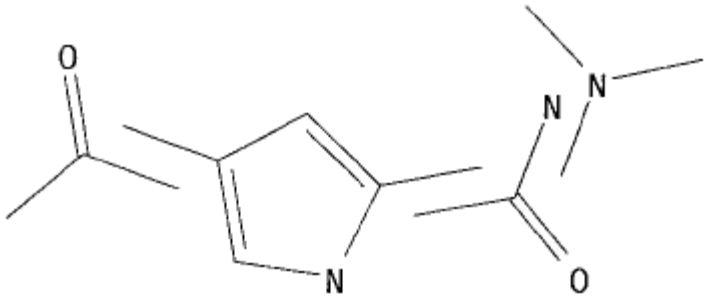


Алгоритмы. Последовательная сборка

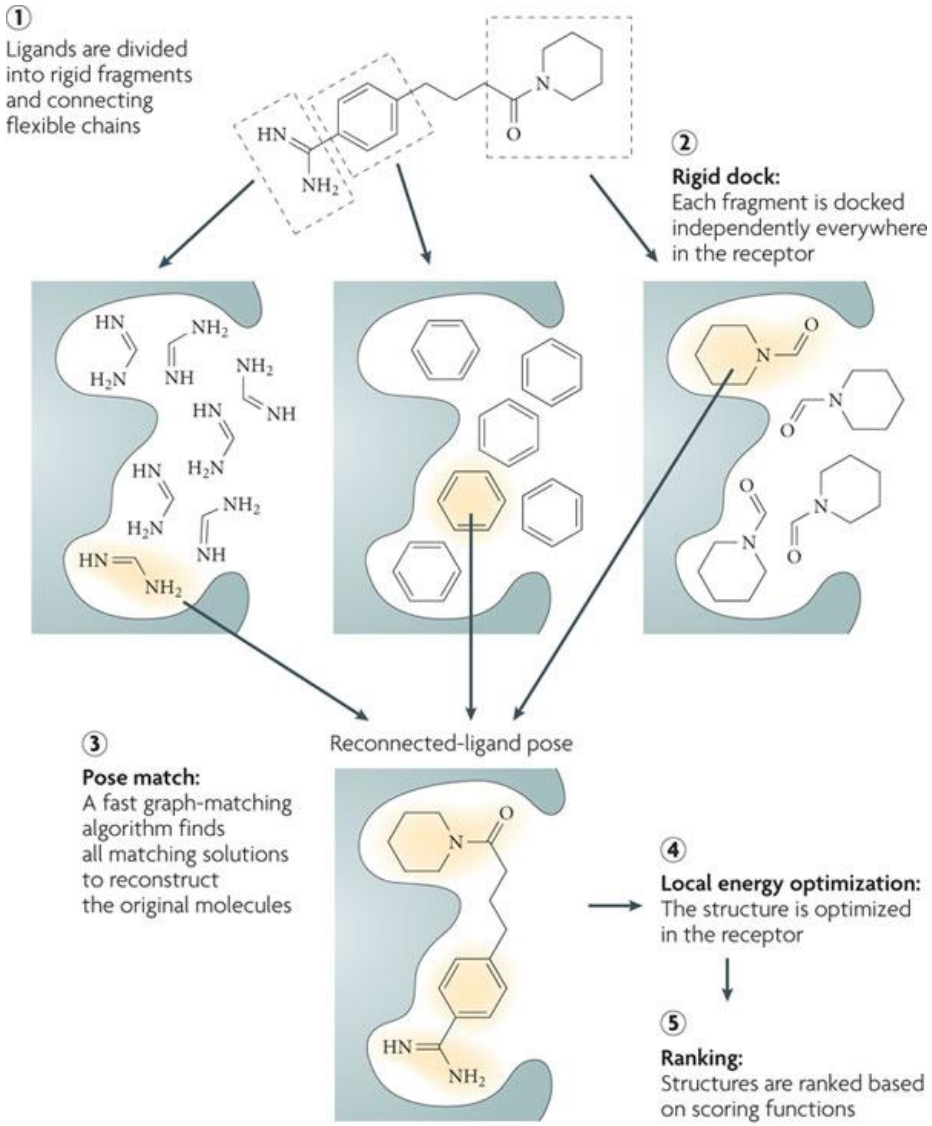
А



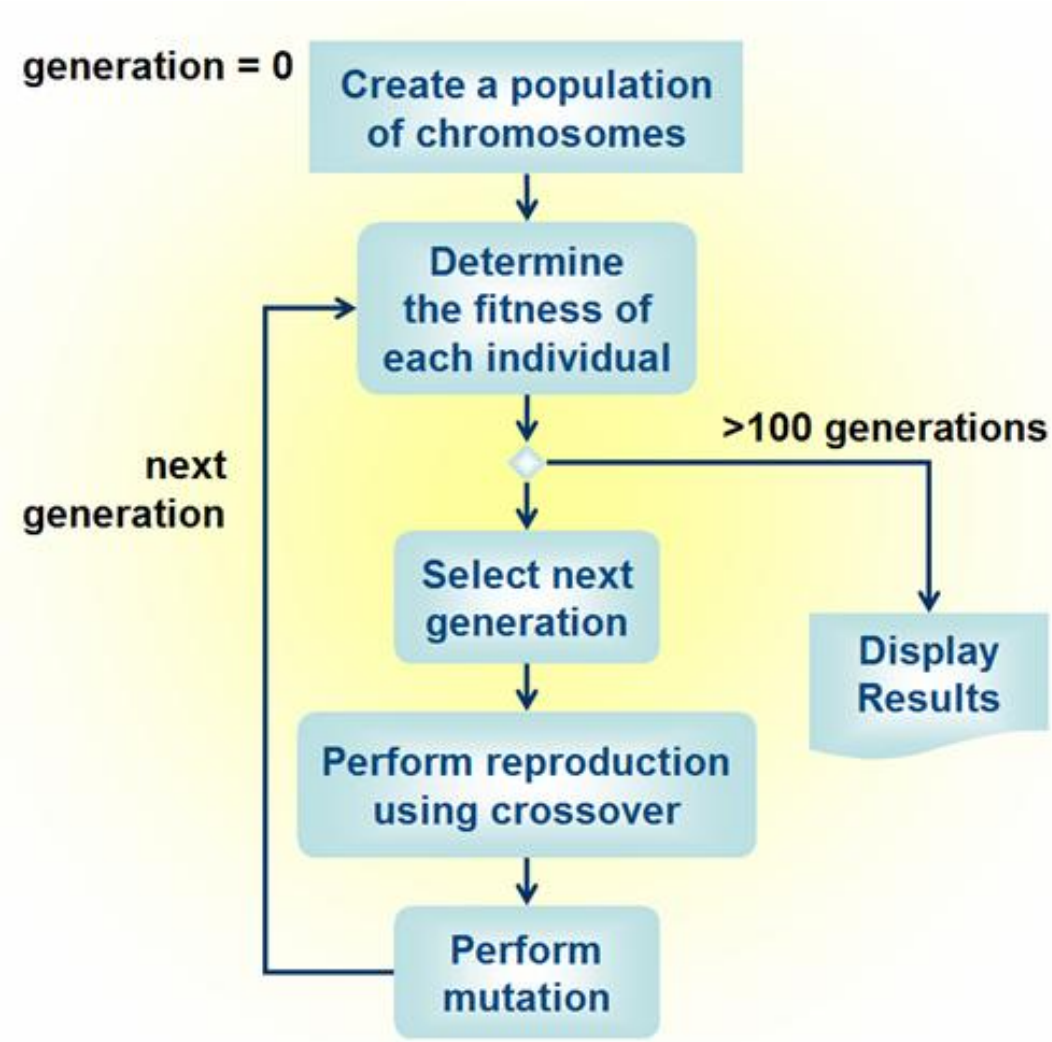
Б



Алгоритмы. Сборка из фрагментов

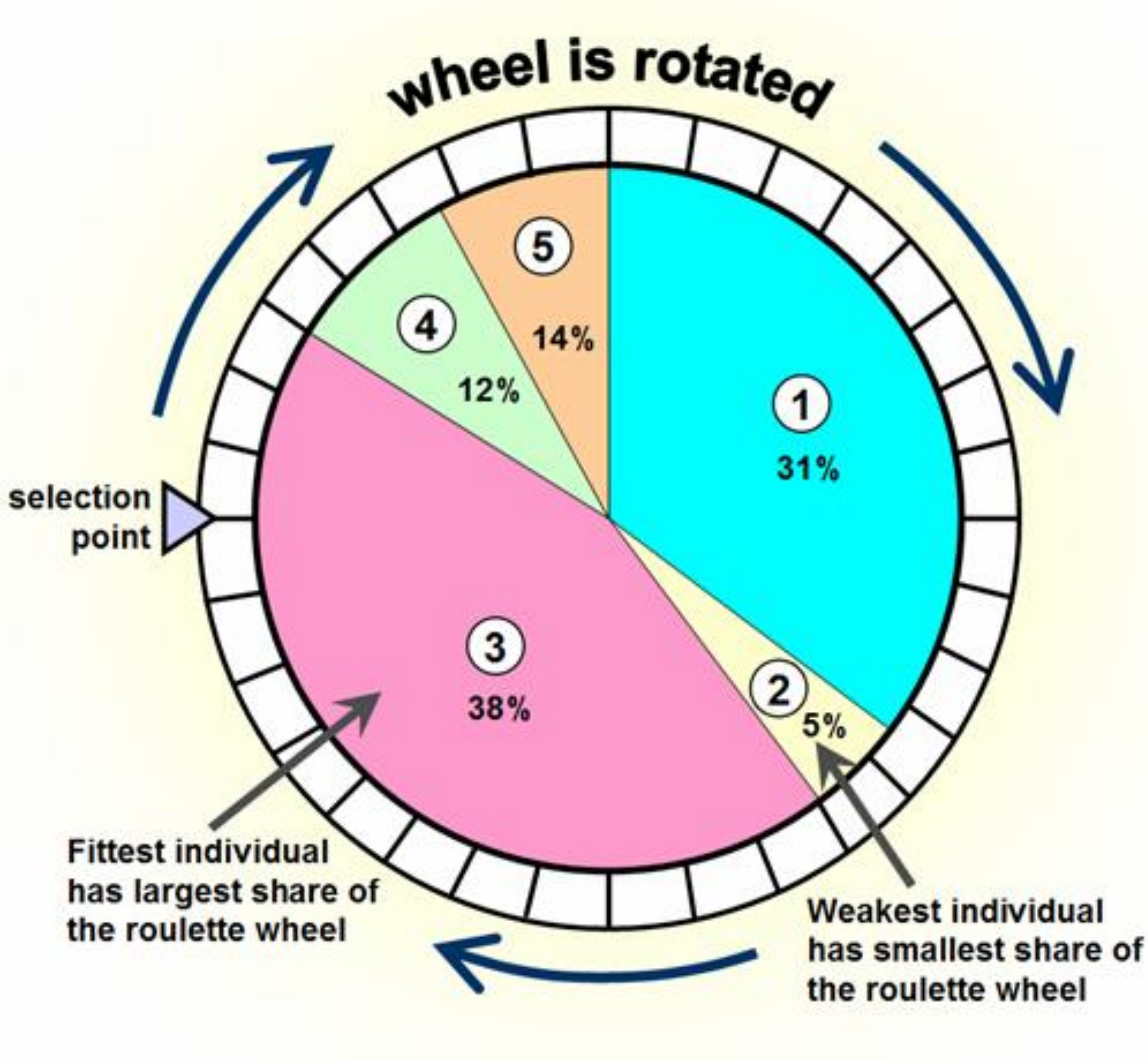


Алгоритмы. Генетический алгоритм

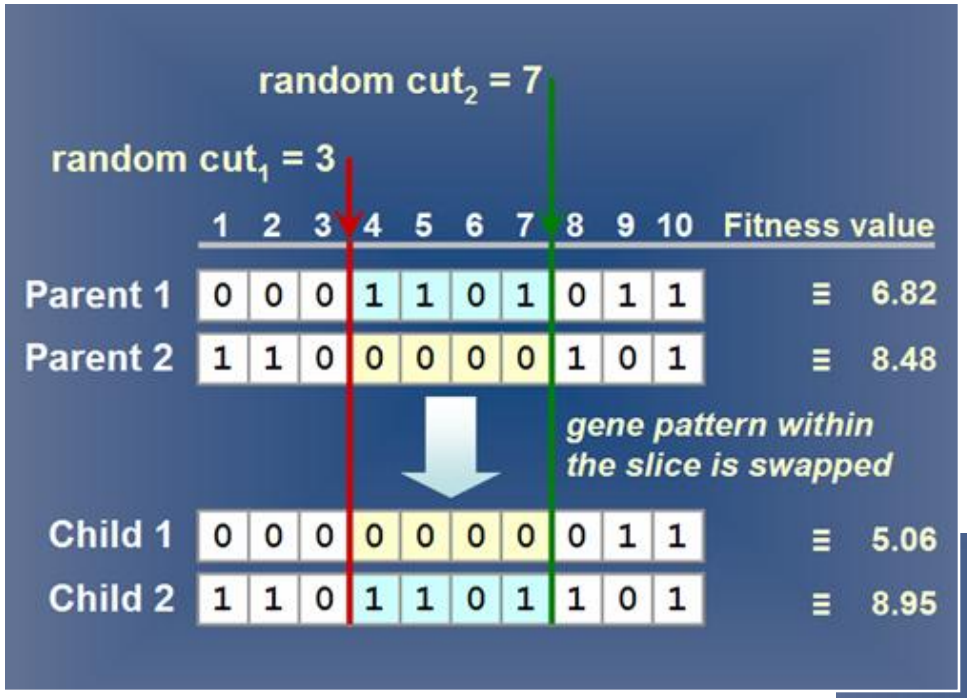


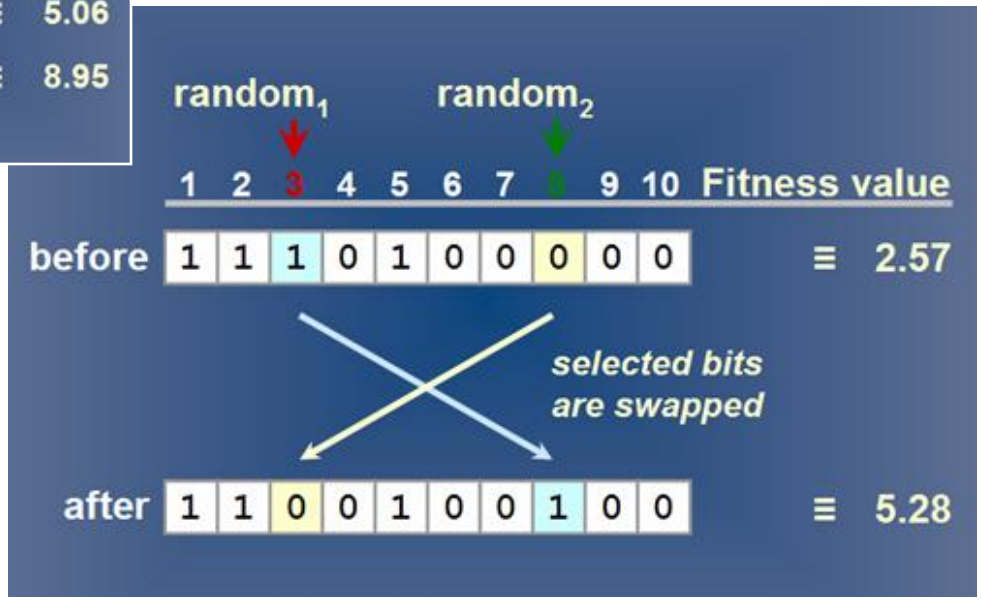
http://www.edc.ncl.ac.uk/highlight/rhjanuary2007.php

Алгоритмы. Генетический алгоритм



Алгоритмы. Генетический алгоритм

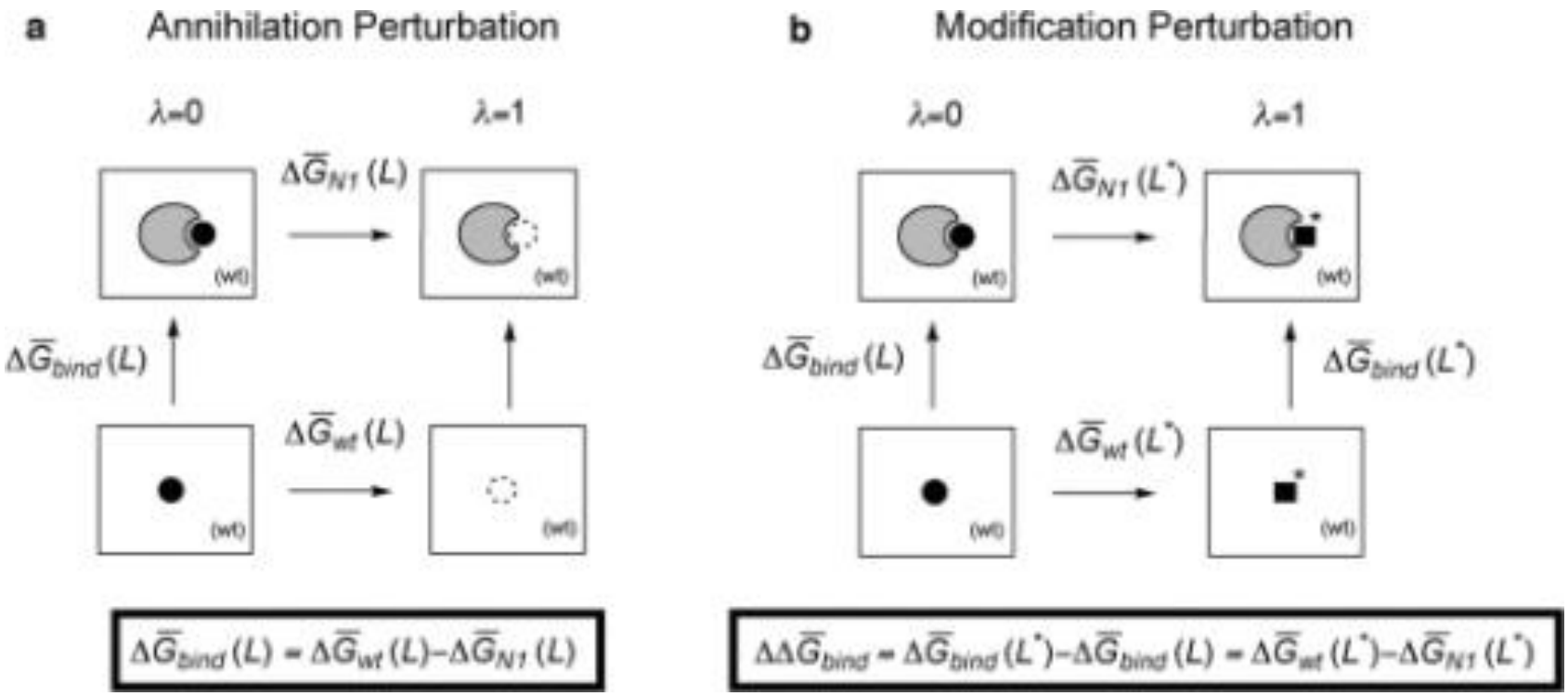




Расчет энергии связывания

Термодинамическое интегрирование

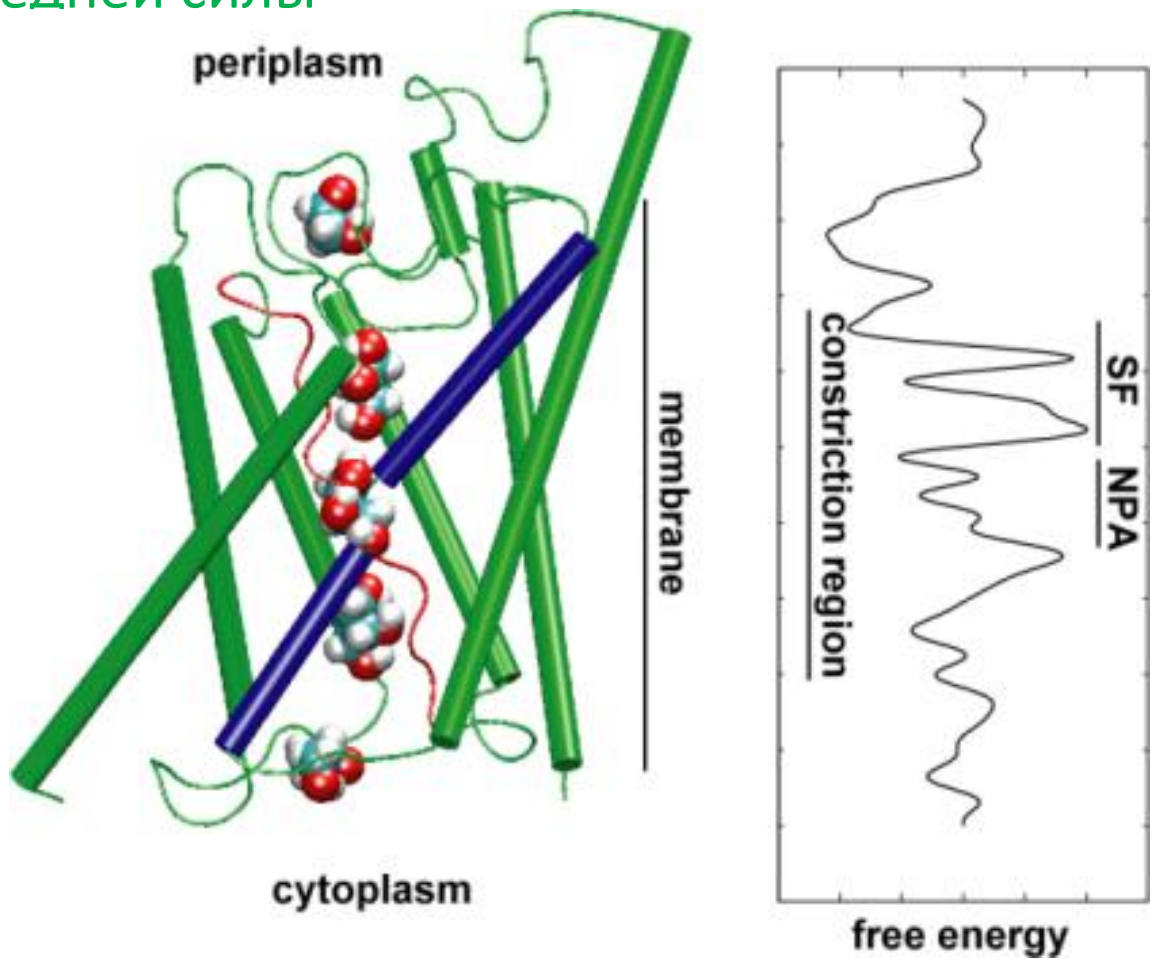
$$\Delta G^{0}_{TI} = \int_{0}^{1} \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$



Не используется в методах молекулярного докинга

Расчет энергии связывания

Потенциал средней силы

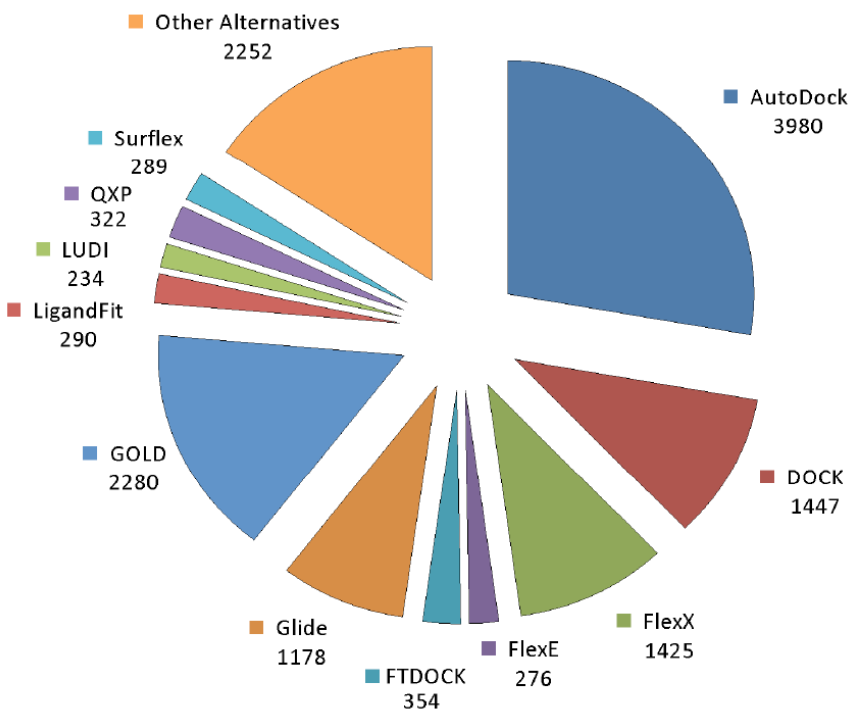


Не используется в методах молекулярного докинга

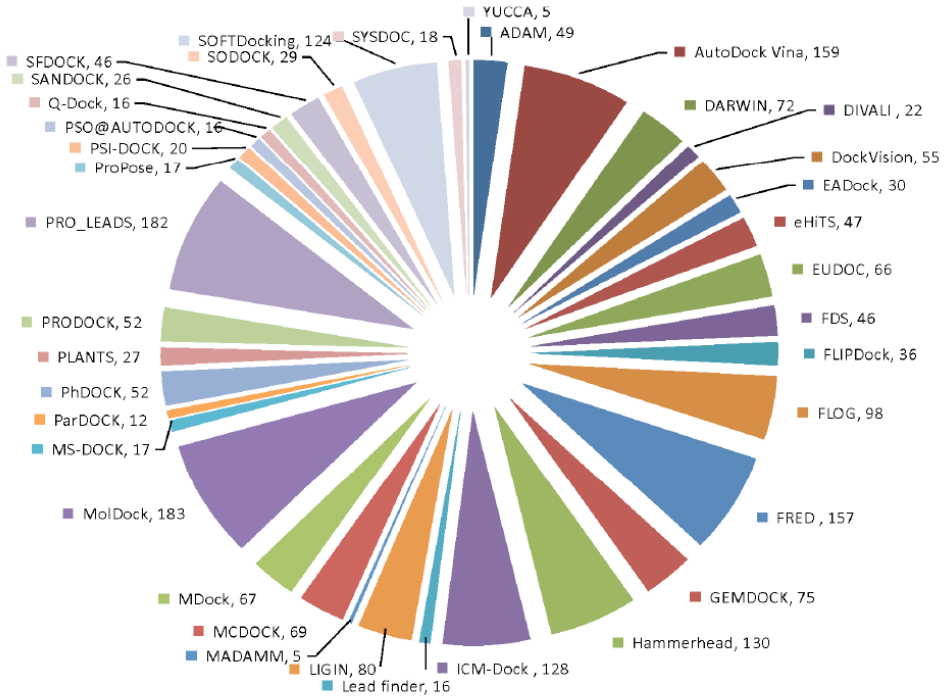
Оценочные функции

- 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



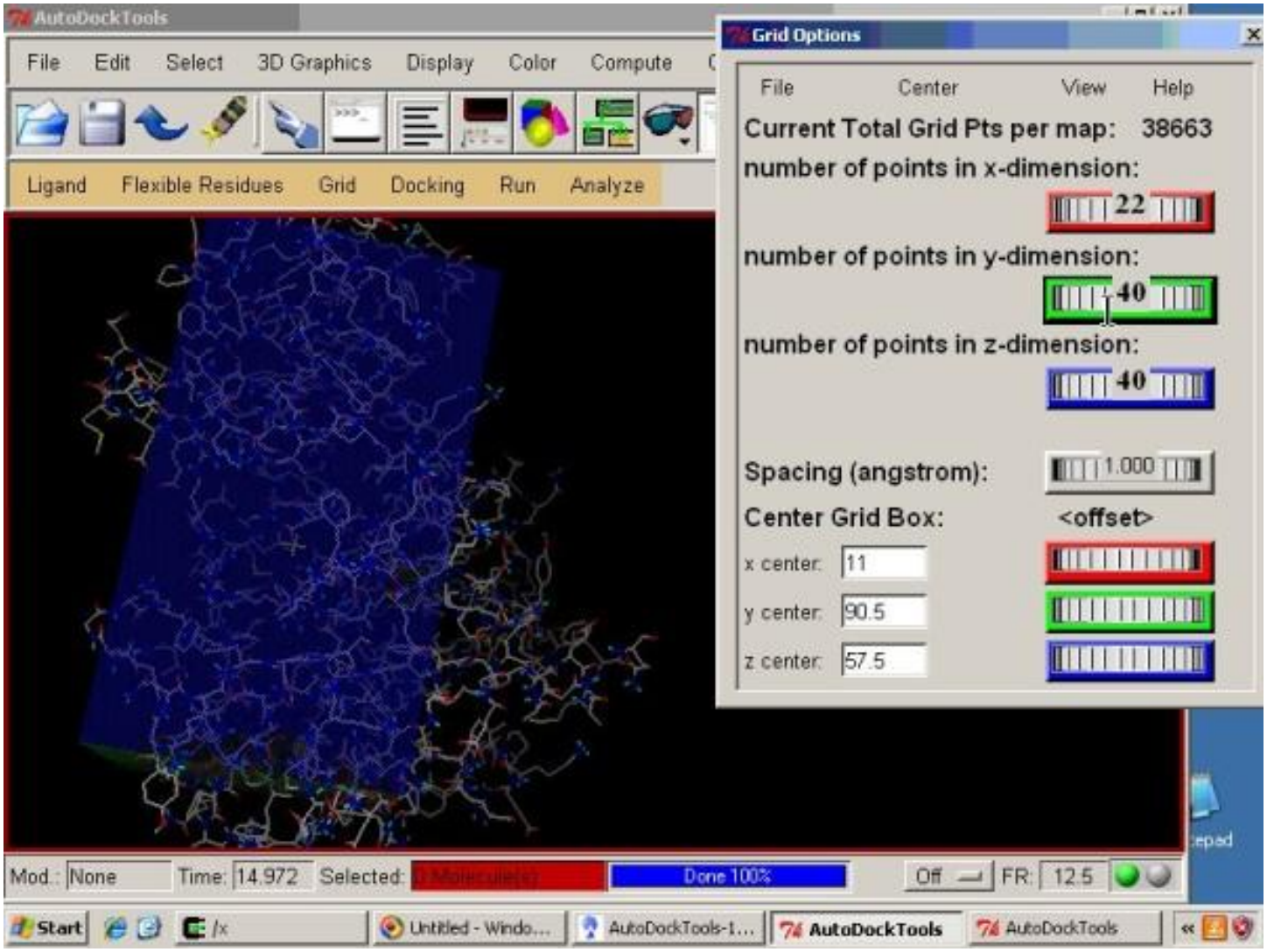
Global



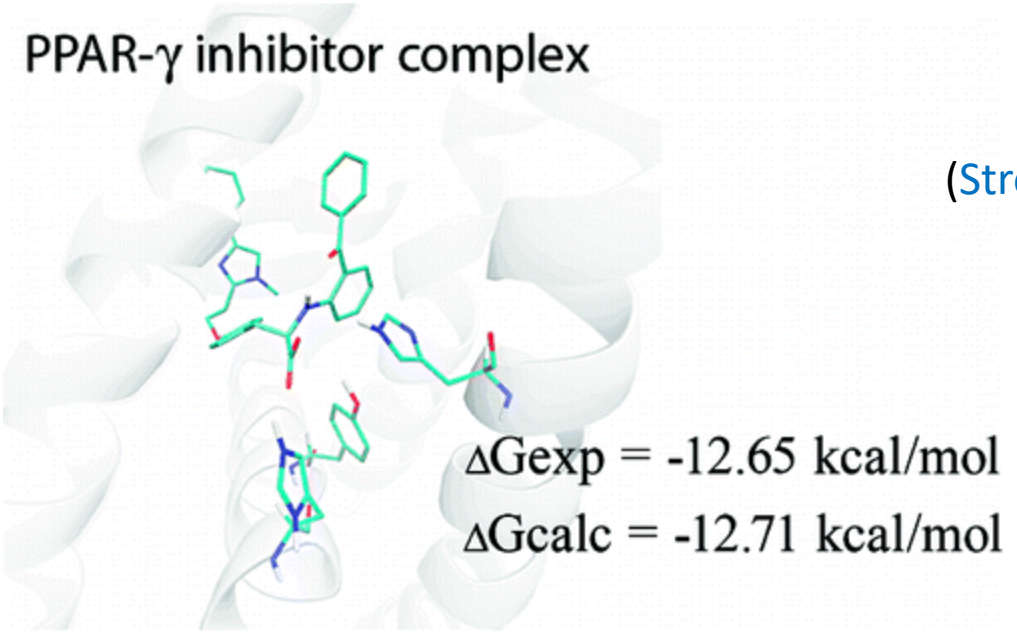
Other Alternatives in detail

Программы для докинга. AutoDock (->AutoDock Vina)

(Goodsell & Olson, 1990; Trott & Olson, 2010)

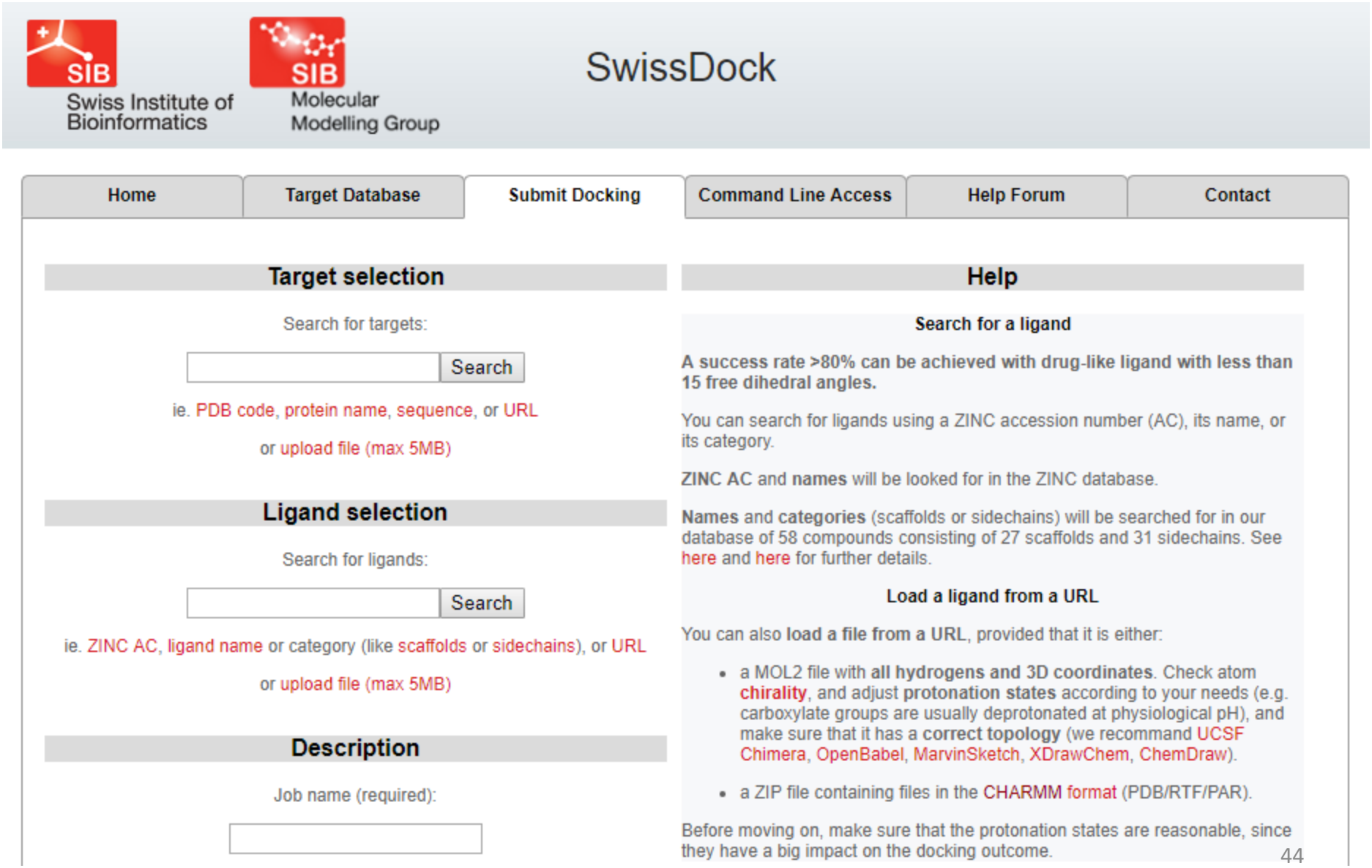


Программы для докинга. LeadFinder

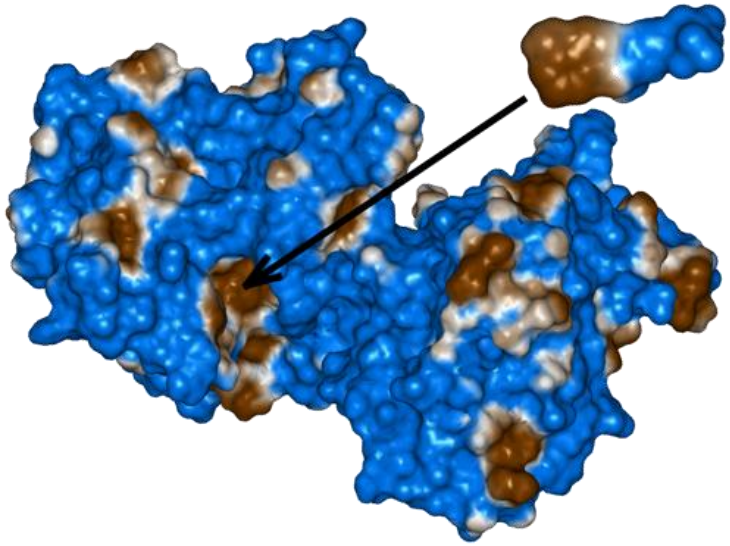


(Stroganov, Chilov et al., 2008)

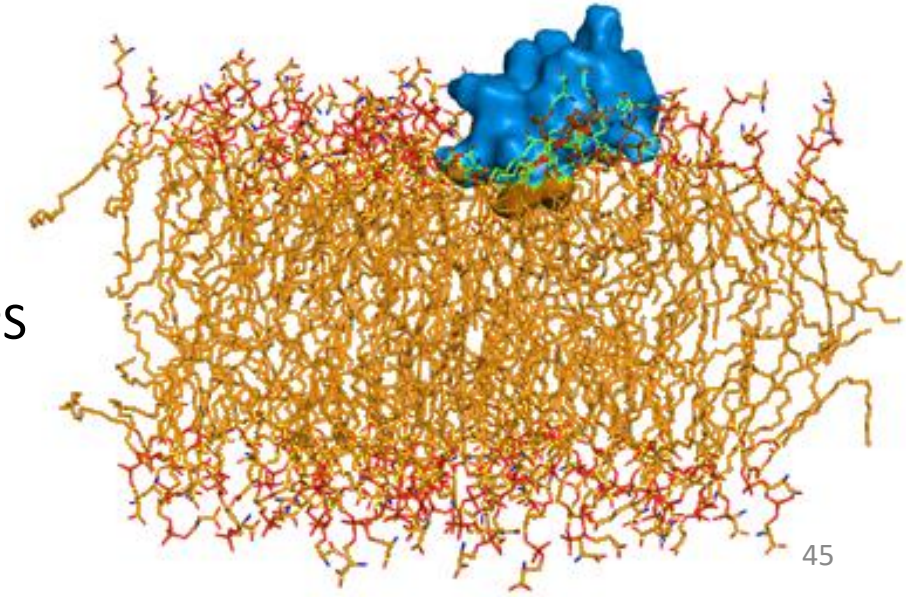
Программы для докинга. SwissDock



Гидрофобные взаимодействия в биомолекулярных системах



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



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

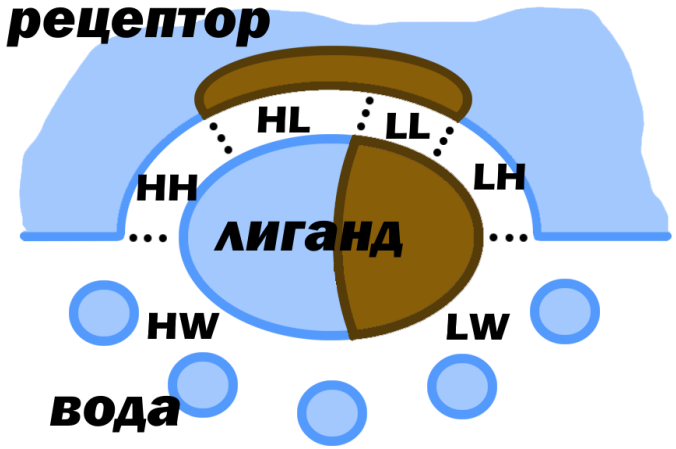
Protein-Ligand ATtractions Investigation NUMerically

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

Results

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

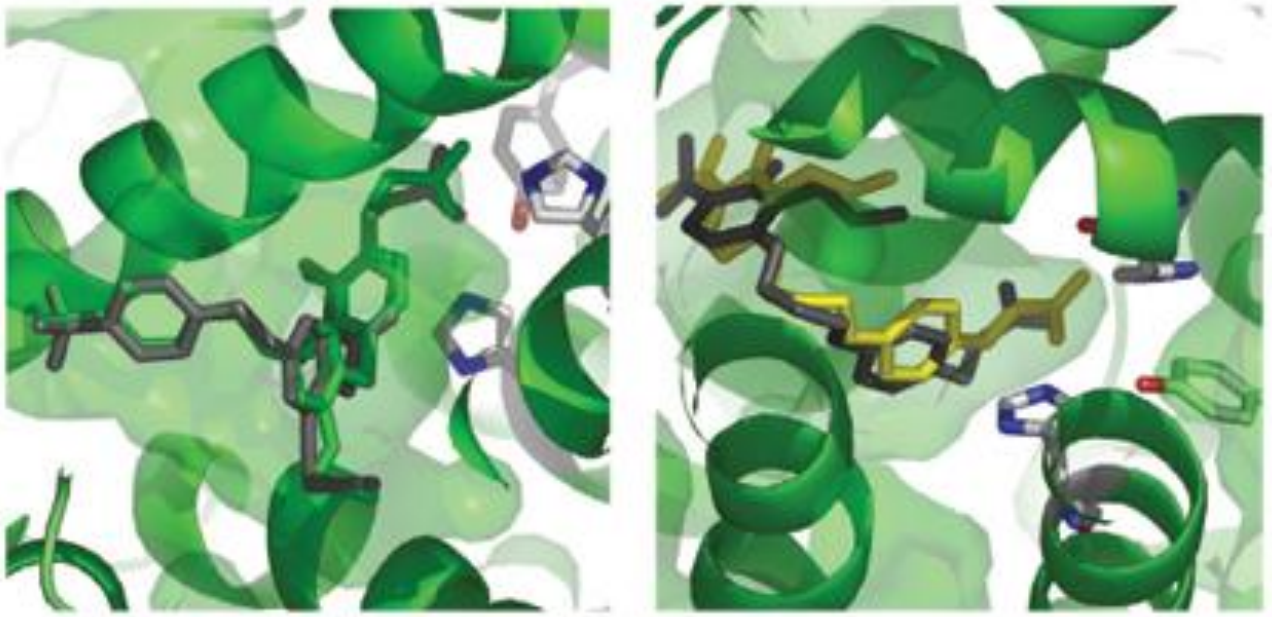
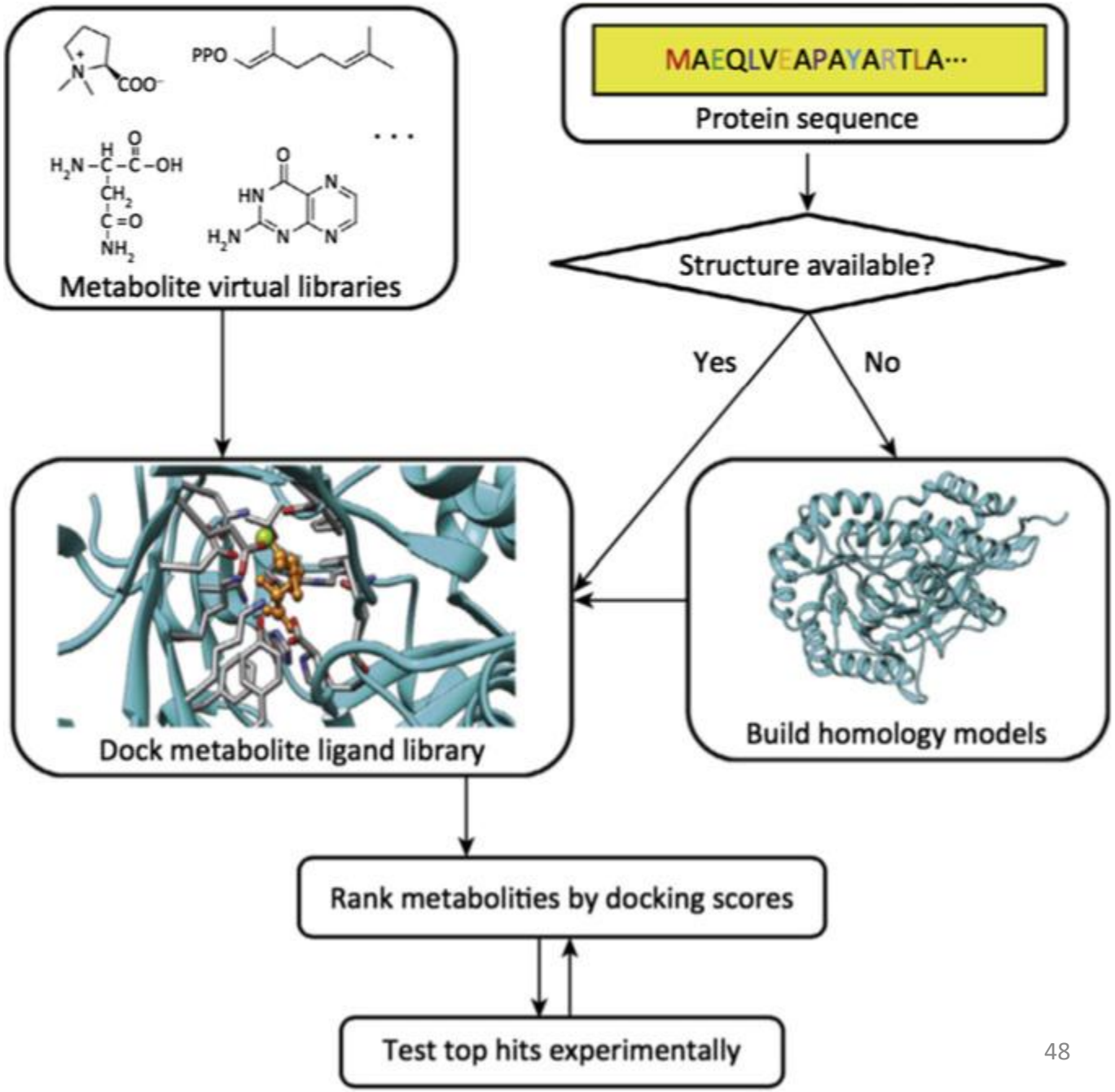


Figure 4. (A) Redocking and (B) crossdocking results. Both crystallographic ligands are shown in gray and the docked ligands are in green (D32) and yellow (L41), respectively. See online for color image.

Аннотация функции



Докинг макромолекул