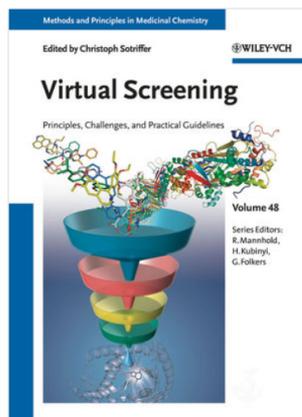


ХИМИЧЕСКАЯ ИНФОРМАТИКА: ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ

Лекции 4-5

Молекулярный докинг

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА



Classification of Current Scoring Functions

J. Liu et al *Journal of Chemical Information and Modeling* 2015 55 (3), 475-482 DOI: 10.1021/ci500731a

Docking and scoring in virtual screening for drug discovery: methods and applications

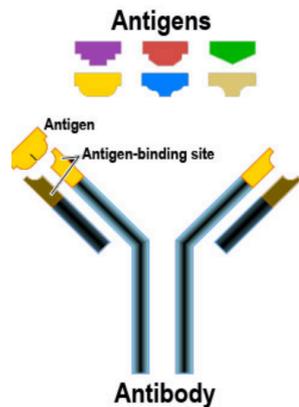
D.B.Kitchen et al *Nature Reviews Drug Discovery* 3, 935-949

Beware of Machine Learning-Based Scoring Functions—On the Danger of Developing Black Boxes

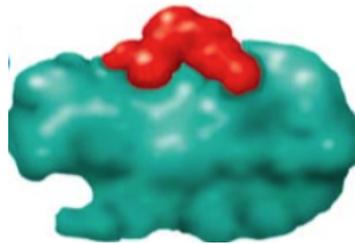
J. Gabel et al *Journal of Chemical Information and Modeling* 2014 54 (10), 2807-2815 DOI: 10.1021/ci500406k

МОЛЕКУЛЯРНЫЙ ДОКИНГ

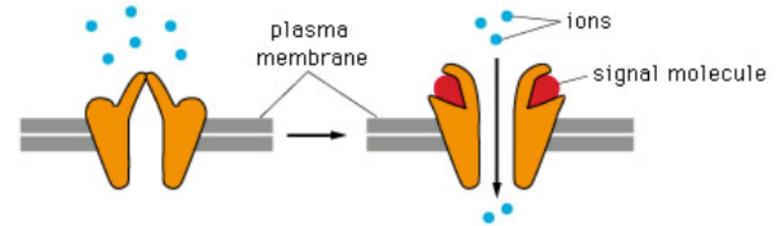
Молекулярный докинг (или молекулярная стыковка) — это метод молекулярного моделирования, который позволяет предсказать наиболее выгодную для образования устойчивого комплекса ориентацию и положение одной молекулы по отношению к другой



антиген-антитело



белок-лиганд



сигнал-рецептор

Прогнозирование

- Ориентации и конформации лиганда внутри активного центра белка
- Значения Оценочной Функции (ОФ или Scoring function), характеризующей энергию связывания

МОЛЕКУЛЯРНЫЙ ДОКИНГ

- **Сайт связывания (или “активный центр”)**

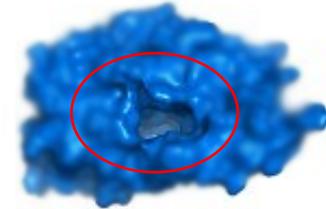
Часть белка, где происходит связывание, в общем случае, полость в структуре белка

- **Положение или способ связывания (“binding mode”)**

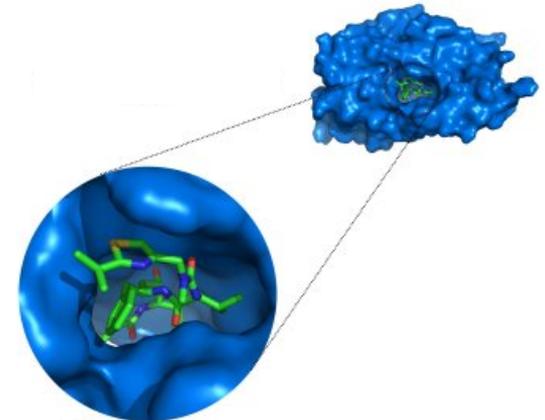
Геометрия лиганда в сайте связывания

Геометрия = **расположение, ориентация и конформация**

Binding site

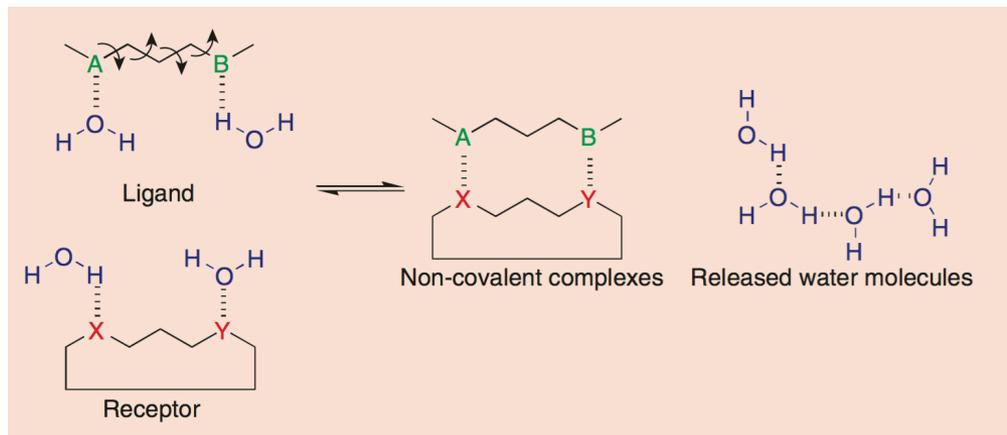


Complex



МОЛЕКУЛЯРНЫЙ ДОКИНГ: ФИЗИКО-ХИМИЧЕСКИЕ ОСНОВЫ МОЛЕКУЛЯРНОГО РАСПОЗНАВАНИЯ

Энергия взаимодействия может быть определена как разность величин свободной энергии в свободном и связанном состояниях или через введение константы связывания:

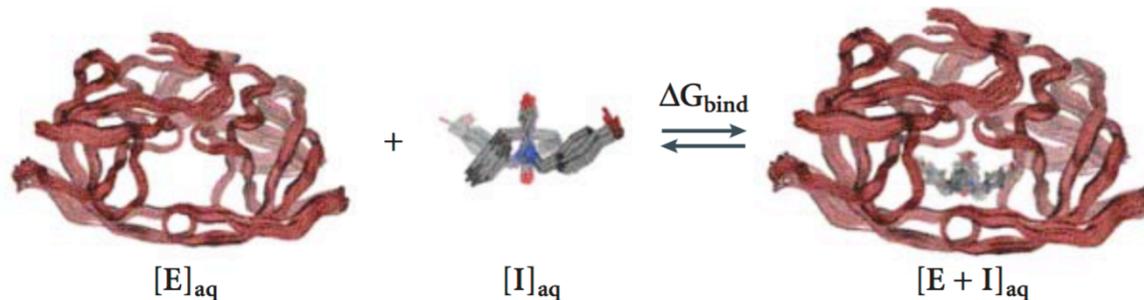


$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ}$$

$$\Delta G^{\circ} = RT \ln K$$

Полярные взаимодействия характеризуются положительным вкладом в энергию связывания
Удаление молекул воды увеличивает энтропию системы

МОЛЕКУЛЯРНЫЙ ДОКИНГ: ФИЗИКО-ХИМИЧЕСКИЕ ОСНОВЫ МОЛЕКУЛЯРНОГО РАСПОЗНАВАНИЯ



Энтальпия

Непосредственный учет взаимодействий между лигандом, растворителем, белком:

- Взаимодействие белок-лиганд
- Взаимодействие лиганд-растворитель
- Взаимодействие растворитель-белок

Энтропия

Конформационная гибкость

Учет гидрофобности

Колебания

Энтропийно-энтальпийная компенсация

Взаимодействие лиганд-мишень должно быть энергетически более выгодным по сравнению с взаимодействием с растворителем

Присутствие у лиганда гидрофобных областей для увеличения положительного сольватационного вклада в энтропию

Как можно более полная стерическая комплементарность с сайтом связывания

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- π interactions
- Arene-arene interactions

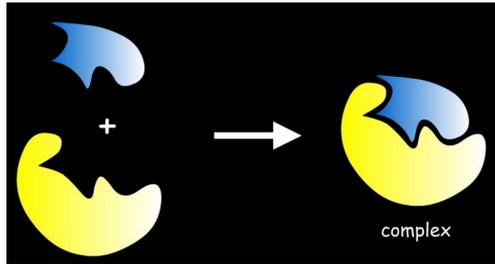
Steric interactions

- 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	ΔG of -20 to -40 kJ/mol $1/r$	
<i>Charge-dipole</i> Depends on orientation of dipole	ΔG of -12 to -20 kJ/mol $1/r^2$	
<i>Dipole-dipole</i> Depends on mutual orientation of dipoles	ΔG of -4 to -12 kJ/mol $1/r^3$	
<i>Charge-induced dipole</i> Depends on polarizability of molecule in which dipole is induced	ΔG of -2 to -10 kJ/mol $1/r^4$	
<i>Dipole-induced dipole</i> Depends on polarizability of molecule in which dipole is induced	ΔG of -2 kJ/mol $1/r^6$	
<i>Dispersion</i> Involves mutual synchronization of fluctuating charges	ΔG of -2 to -4 kJ/mol $1/r^6$	
<i>Hydrogen bond</i> Charge attraction + partial covalent bond	ΔG of -4 to -30 kJ/mol	

МОЛЕКУЛЯРНЫЙ ДОКИНГ: ИСТОРИЯ РАЗВИТИЯ



Принцип "ключ-замок"

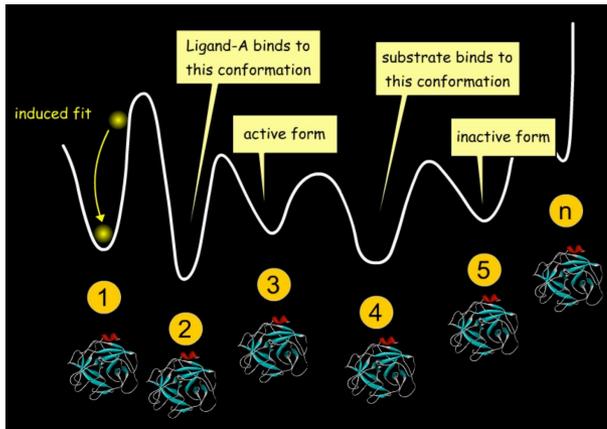
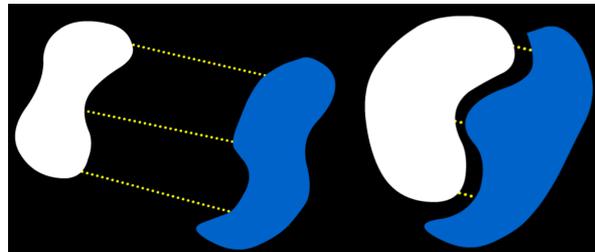
Emil Fischer (1890):
"A substrate fits into the active site of a macromolecule, just like a key fits into a lock"

Модель индуцированного соответствия

Daniel Koshland (1958):
"Both ligand and target mutually adapt to each other through small conformational changes"

Модель конформационного ансамбля

Buyong Ma (2003)
"Proteins are considered as a pre-existing ensemble of conformational states. The plasticity of the protein allows it to switch from one state to another"



МОЛЕКУЛЯРНЫЙ ДОКИНГ: ТИПЫ

В зависимости от наличия информации об активном сайте белка:

«слепой» докинг (blind docking)

Активный сайт белка не известен. Необходимо определить центр связывания лиганда, а затем и его способ. Частным случаем «слепого» докинга является докинг в построенные по гомологии модели белков, когда предполагается что их центр связывания похож на центр связывания шаблонного белка.

прямой (direct docking) докинг.

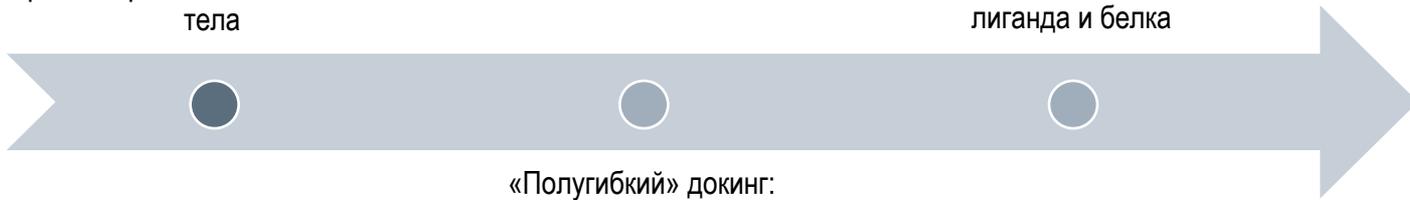
Активный сайт известен. В этом случае интересует прежде всего способ связывания (ориентация) лиганда в центре связывания.

В зависимости от учета гибкости лиганда и белка:

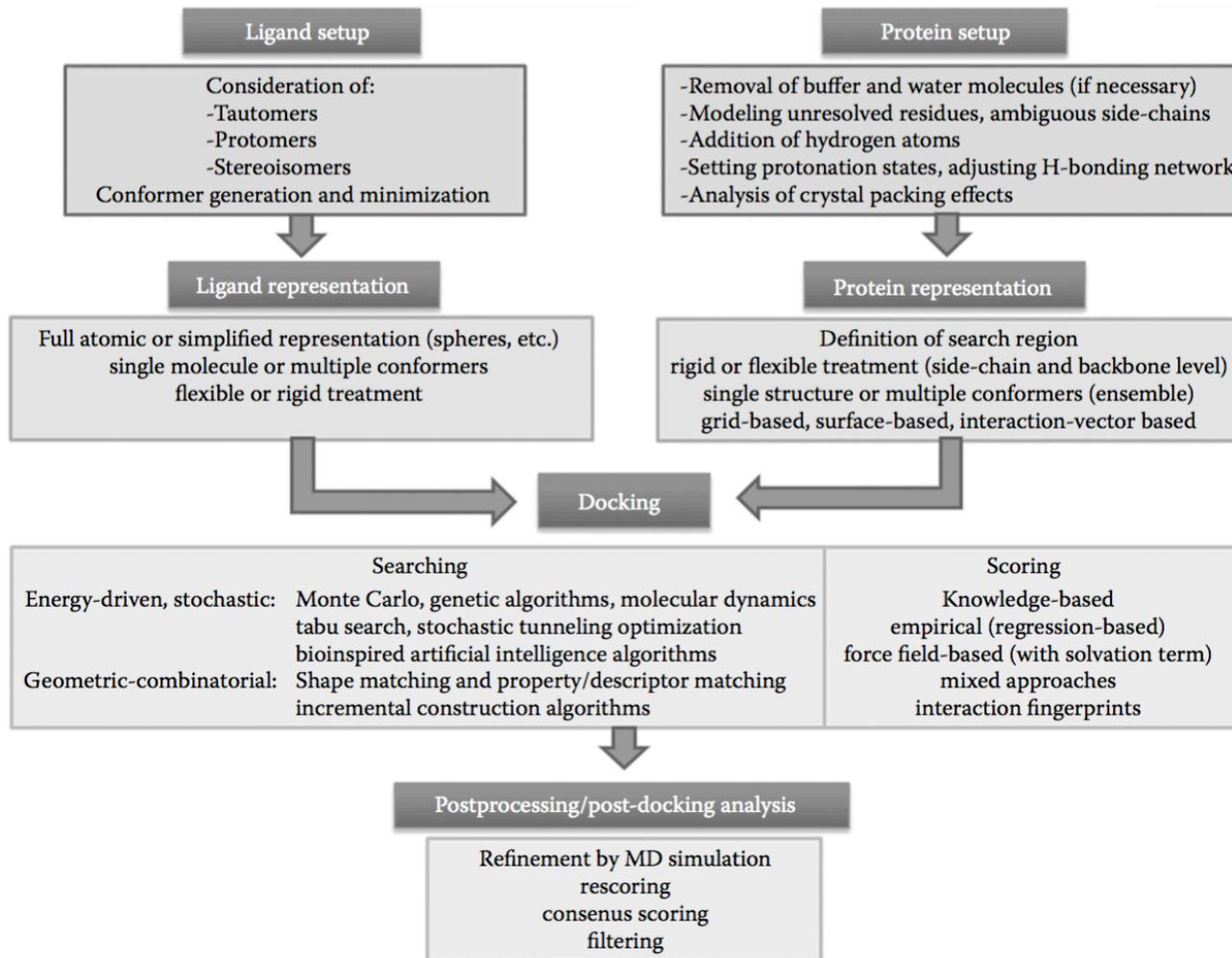
«Жесткий» докинг:
И белок и лиганд
рассматриваются как жесткие
тела

«Гибкий» докинг:
Учитывается гибкость и
лиганда и белка

«Полугибкий» докинг:
Учитывается только
конформационная гибкость
лиганда



МОЛЕКУЛЯРНЫЙ ДОКИНГ: ОБОБЩЕННЫЙ АЛГОРИТМ



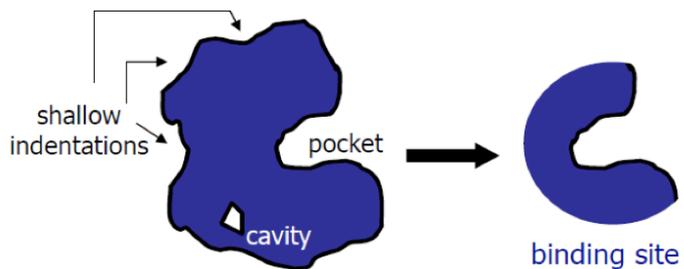
МОЛЕКУЛЯРНЫЙ ДОКИНГ: ПОИСК ЦЕНТРА СВЯЗЫВАНИЯ

Характеристики центра
связывания

Глубина

Объем

Площадь липофильной поверхности



Использование баз данных

sc-PDB

An Annotated Database of Druggable Binding Sites from the Protein DataBank

scPDB release 2013 online: 9283 entries, 3678 proteins and 5608 ligands

Database search

About sc-PDB

Clusters of binding sites

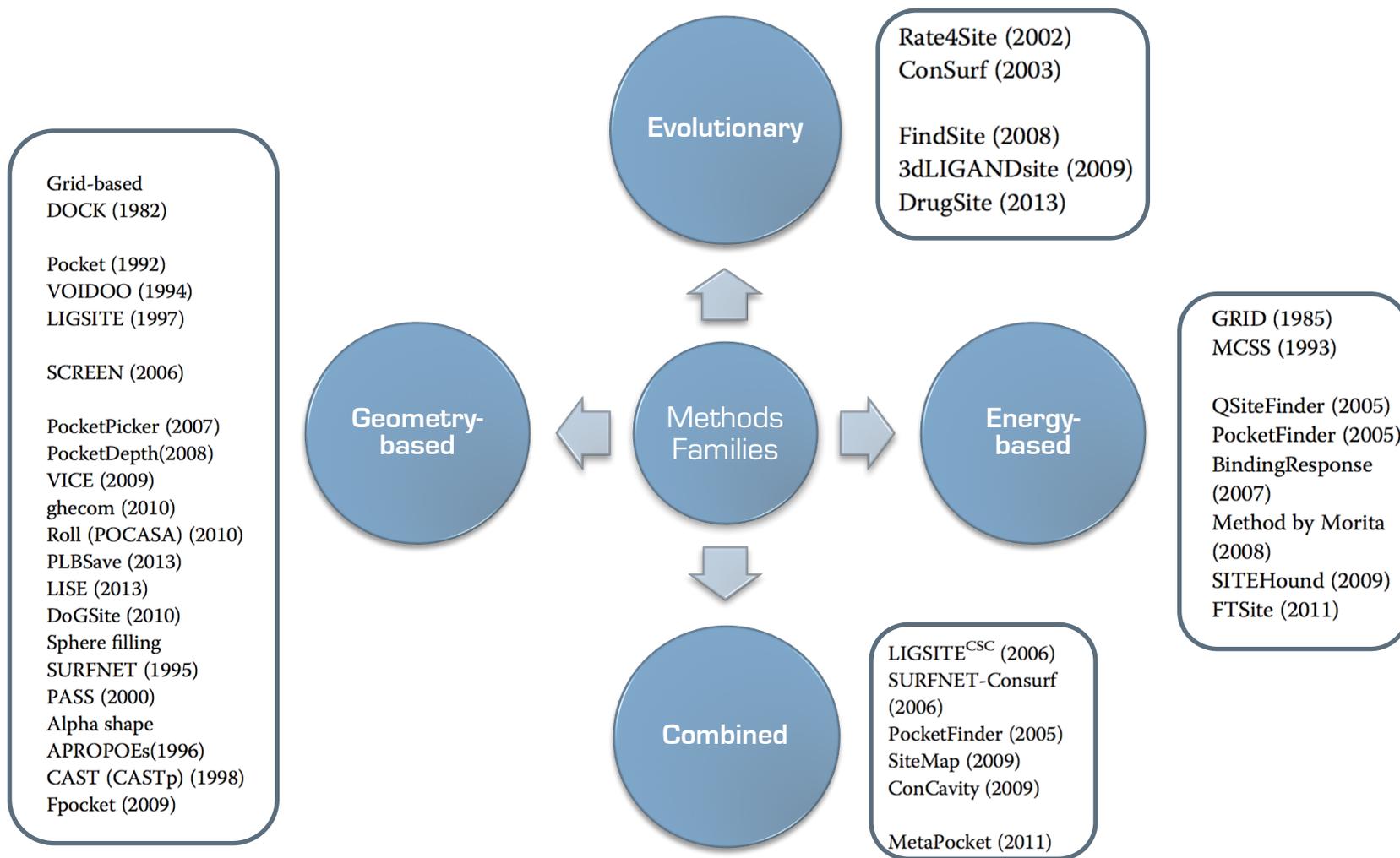
Literature Corner

scPDB Fragments

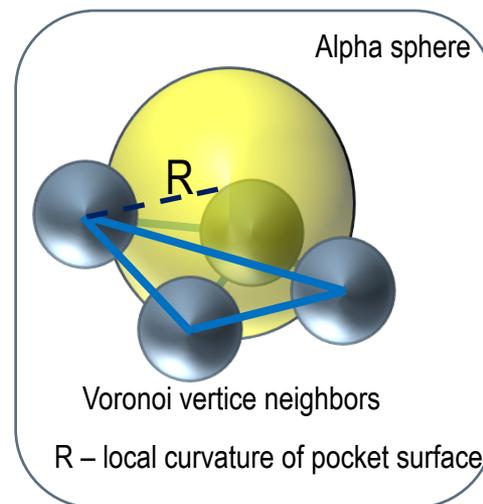
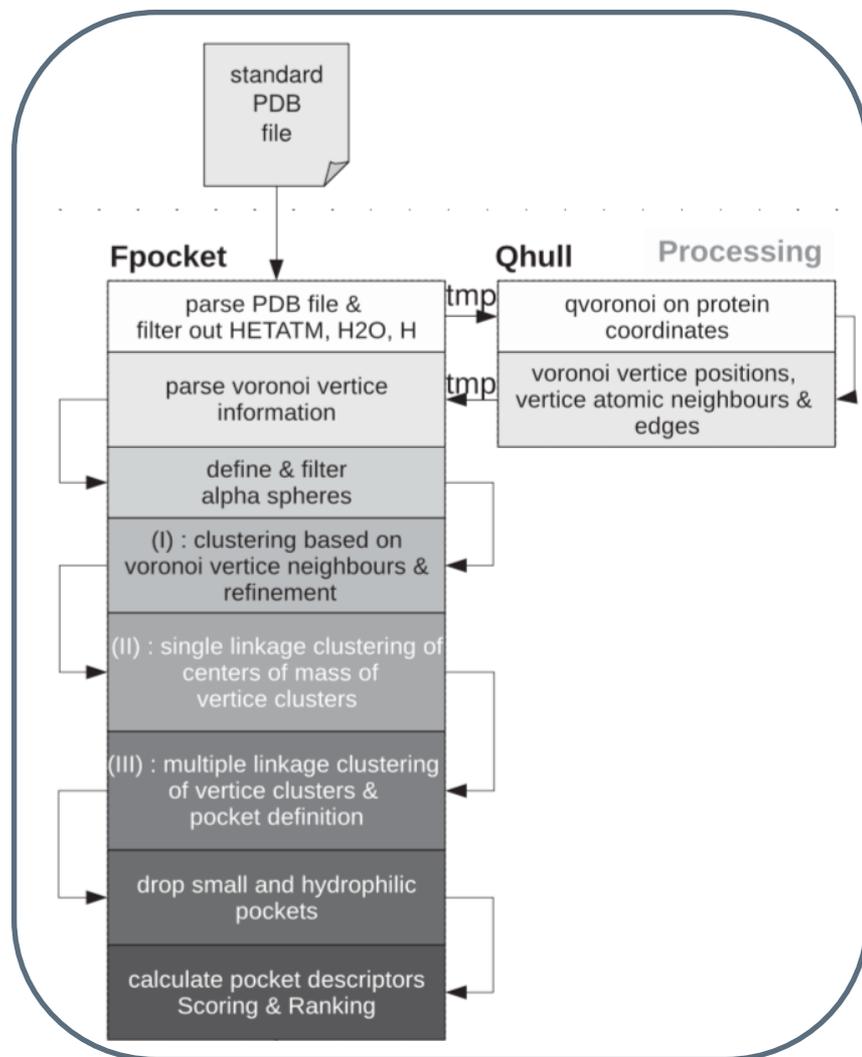
Download sc-PDB

Методы сканирования поверхности

METHODS FOR PROTEIN BINDING SITE DETECTION



GEOMETRY-BASED METHODS: FPOCKET



Descriptors:

- Alpha spheres (number, polarity, density)
- Protein atoms (electronegativity)
- Residues (residue types occurrence, hydrophobicity)
- Volume

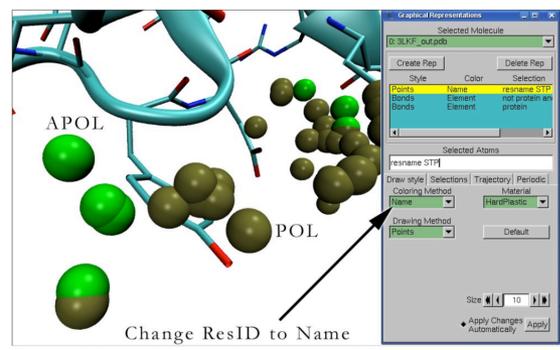
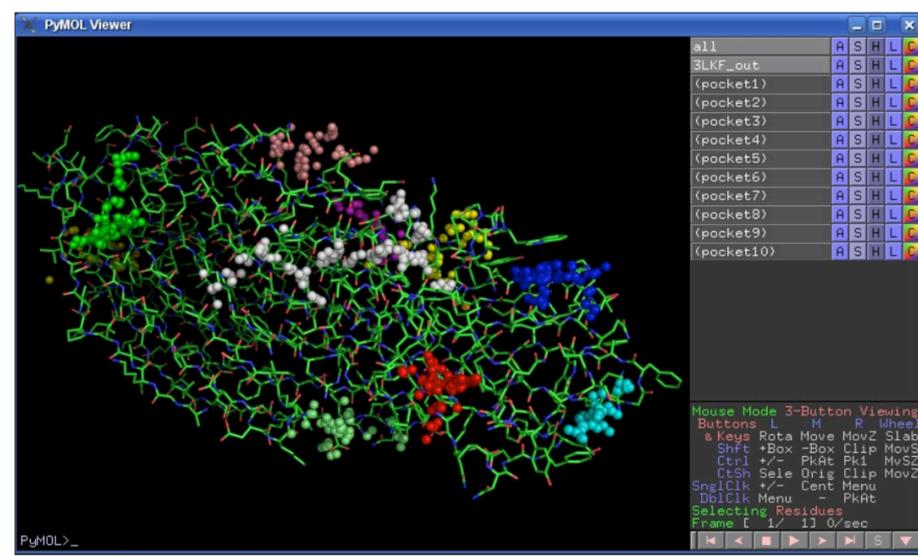
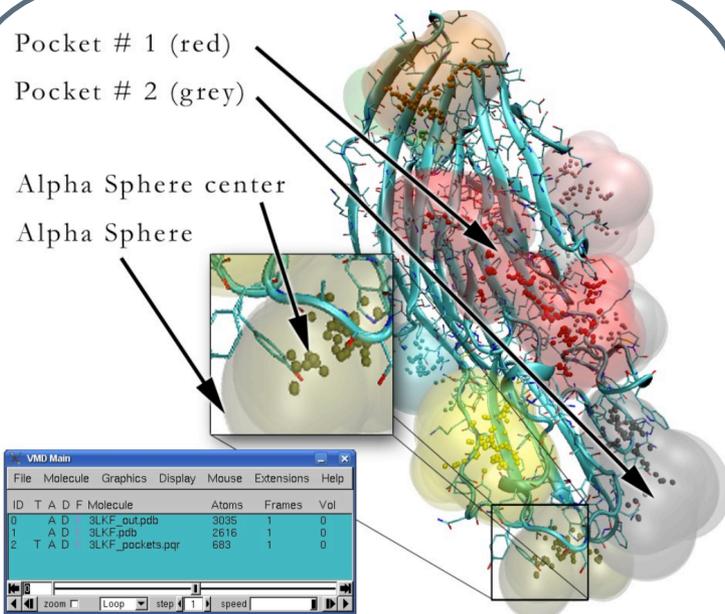
Scoring functions:

- Normalized number of alpha spheres
- Normalized mean local hydrophobic density
- Normalized proportion of apolar alpha spheres
- Polarity score
- Alpha sphere density

GEOMETRY-BASED METHODS: FPOCKET

VMD Visualization

PyMol Visualization



DRUGGABILITY ASSESSMENT

Classification models (whether given protein belongs to druggable family)

Consideration of positive and negative physicochemical contributions

- + Pocket hydrophobicity, sufficient size, buriedness
- Polar contacts, charged residues

Precedent-based or energy-assessing evaluation

ПОДГОТОВКА СТРУКТУРЫ БЕЛКА: БАЗЫ ДАННЫХ БЕЛКОВЫХ СТРУКТУР



RCSB PDB
PROTEIN DATA BANK
since 1998, maintained by
Research Collaboratory for
Structural Bioinformatics



EMBL-EBI
European
Macromolecular
Structure Database



PDBj
Protein Data Bank
Japan

**Data deposit, treatment
and distribution**

Weekly updated, data synchronisation on mirror sites



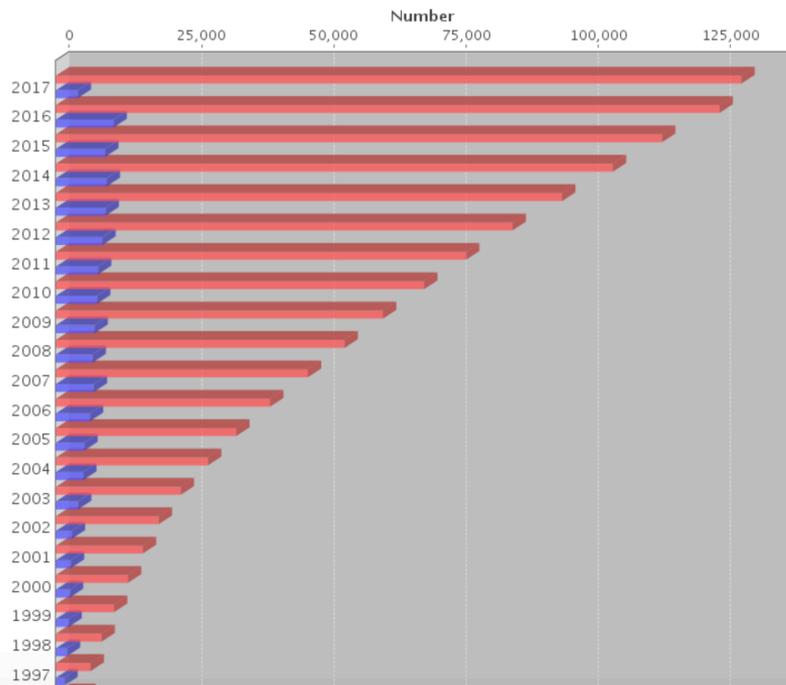
WORLDWIDE PDB
PROTEIN DATA BANK



BMRB
BioMagResBank
A Repository for Data from NMR Spectroscopy
on Proteins, Peptides, and Nucleic Acids
NMR data (since 2006)
Department of Biochemistry
University of Wisconsin-Madison

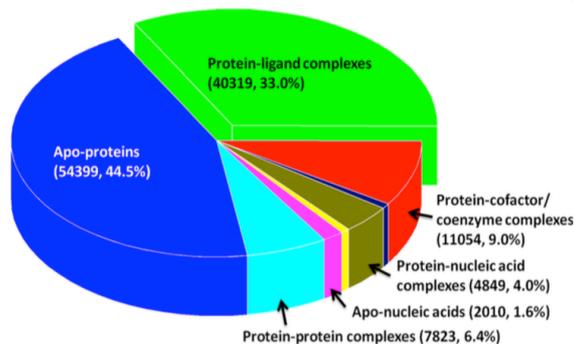
БАЗЫ ДАННЫХ БЕЛКОВЫХ СТРУКТУР: PROTEIN DATA BANK

RCSB
PDB
PROTEIN DATA BANK

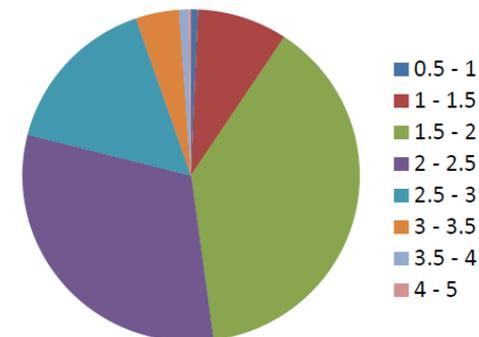


Основные методы определения структуры:

- PCA
- ЯМР



Кол-во структур с различным разрешением



version	entries in PDB ^a	valid complexes ^b	complexes with binding data				total
			protein-ligand	nucleic acid-ligand	protein-protein	protein-nucleic acid	
2007	40 876	11 822	3124	0	0	0	3124
2008	48 092	18 211	3539	40	471	250	4300
2009	55 118	23 284	4277	44	1053	304	5678
2010	62 387	26 434	5075	55	1281	361	6772
2011	70 224	30 259	6051	66	1441	428	7986
2012	78 235	34 180	7121	79	1597	511	9308
2013	87 085	38 918	8302	83	1804	587	10 776
2014	96 592	44 569	10 656	87	1592	660	12 995
2015	105 183	48 821	11 987	109	1807	717	14 260
2016	114 344	53 838	13 308	118	1976	777	16 179

ПОДГОТОВКА СТРУКТУРЫ БЕЛКА: БАЗЫ ДАННЫХ БЕЛКОВЫХ СТРУКТУР

Name	URL	Statistics	Annotation
sc-PDB	http://bioinfo-pharma.u-strasbg.fr/scPDB	9877 pockets 3034 targets 5339 ligands	Atomic coordinates (MOL2) Ligand properties Protein properties Binding site properties Protein–ligand interactions
PDTD	http://www.dddc.ac.cn/pdtd/	1207 pockets 841 targets	Atomic coordinates (PDB, MOL2) Therapeutic area Related disease Protein function
Anonymous	http://www.bcgsc.ca/downloads/yli/	2923 pockets 252 proteins	Atomic coordinates (ICM)
BioDrugScreen	http://www.biodrugscreen.org/	1926 pockets 1549 proteins	Protein–ligand docking scores (9 different scoring functions)
SePreSa	http://sepresa.bio-x.cn/	115 pockets 91 targets	Polymorphism of binding site residues
DAR-CPI	http://cpi.bio-x.cn/drar/	385 pockets 353 proteins	Positive/negative association with 166 drugs

ПОДГОТОВКА СТРУКТУРЫ БЕЛКА

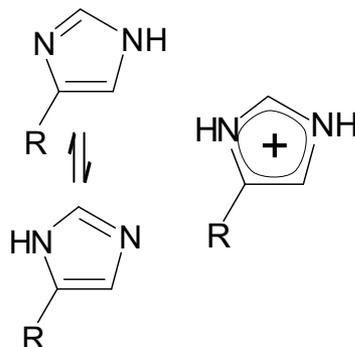
В качестве начальной точки выбрать структуру с хорошим разрешением

- ❖ Когда доступно несколько вариантов структур предпочтение отдается комплексам белок-лиганд перед апоструктурами (свободными белками)
- ❖ Если доступна информация по нескольким комплексам, то выбираются структуры, сокристаллизованные с лигандом в интересующем сайте связывания
- ❖ Если таких структур несколько, то используется их ансамбль, отражающий конформационную гибкость белка

PDB структуры часто содержат молекулы воды. В общем случае, все молекулы воды удаляются за исключением случаев, когда известна их функциональность в координации с лигандом

Добавляются атомы водорода

Корректное обозначение протонирования в активном центре (аспарагин, глутамин, гистидин): таутомерные формы остатков гистидина выбираются с учетом их взаимодействия с известными лигандами в существующих структурах



ПОДГОТОВКА ЛИГАНДА: НЕКОТОРЫЕ БАЗЫ ДАННЫХ

Databases for Virtual Screening

Zinc	http://zinc.docking.org
Zinc15	http://zinc15.docking.org
ZincClick	http://www.symech.it/index.asp?catID=31&lang=en
PubChem	https://pubchem.ncbi.nlm.nih.gov
ChemSpider	http://www.chemspider.com
ChEMBL	https://www.ebi.ac.uk/chembl/
ChemBank	http://chembank.broadinstitute.org
eMolecules	https://www.emolecules.com
DrugBank	https://www.drugbank.ca
BindingDB	https://www.bindingdb.org/bind/index.jsp
SCUBIDOO	http://kolblab.org/scubidoo/view/download.php
DUD	http://dud.docking.org
MUV	https://www.tu-braunschweig.de/pharmchem/forschung/baumann/muv
CoCoCo	http://cococo.isof.cnr.it
Wombat	http://dud.docking.org/wombat/
TCM	http://tcm.cmu.edu.tw
DNP	http://dnp.chemnetbase.com/intro/index.jsp

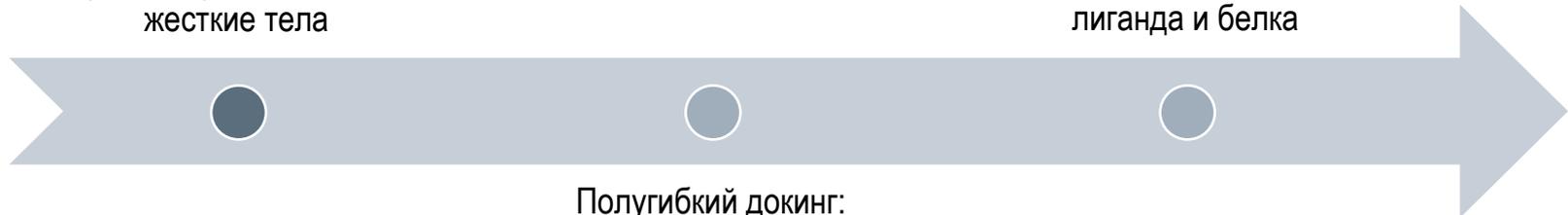
МОЛЕКУЛЯРНЫЙ ДОКИНГ: ГИБКОСТЬ БЕЛКА И ЛИГАНДА

Жесткий докинг:
И белок и лиганд
рассматриваются как
жесткие тела

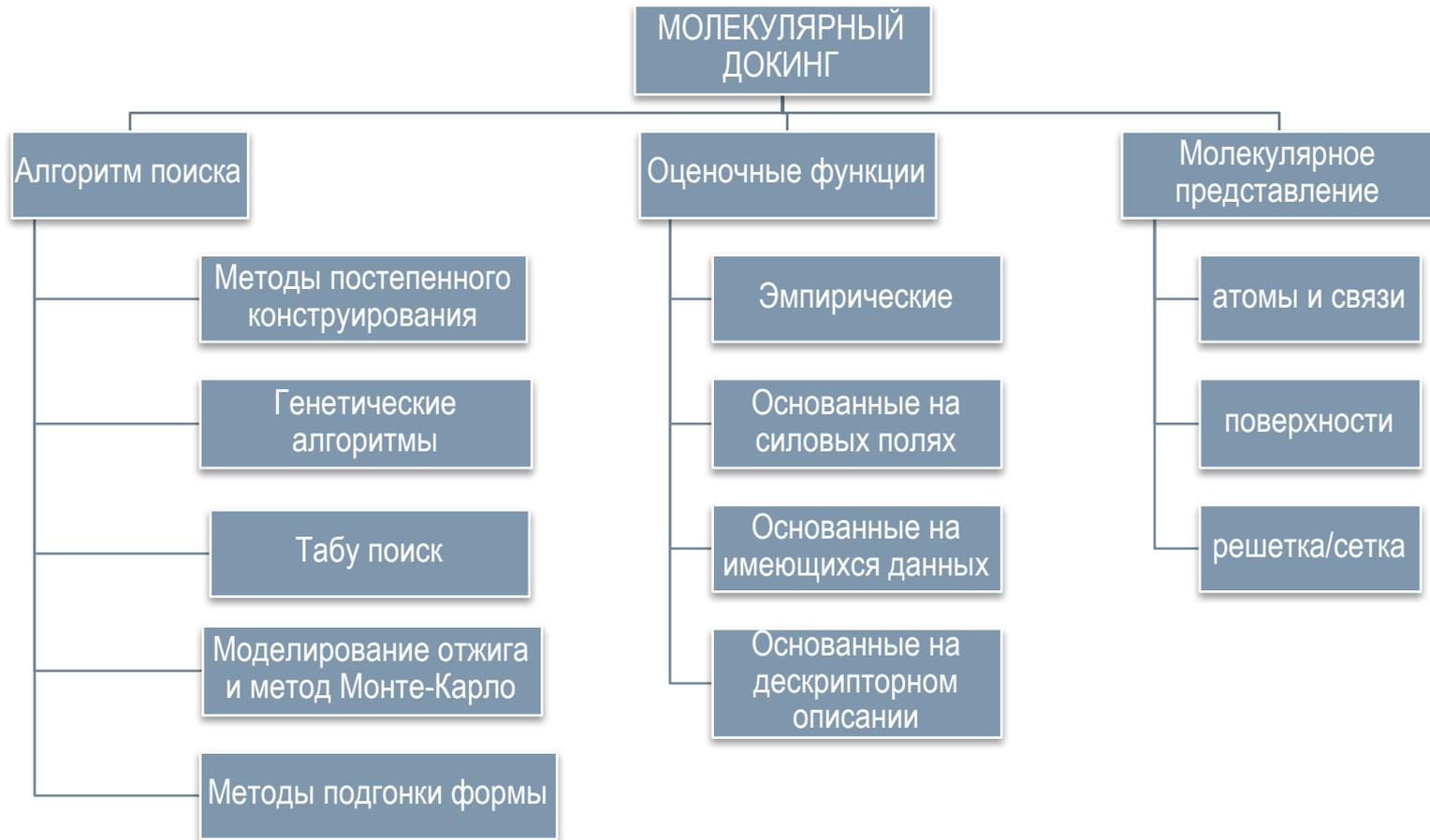
Гибкий докинг:
Учитывается гибкость и
лиганда и белка

Полугибкий докинг:
Только для лиганда
учитывается
конформационная гибкость

Сложность модели



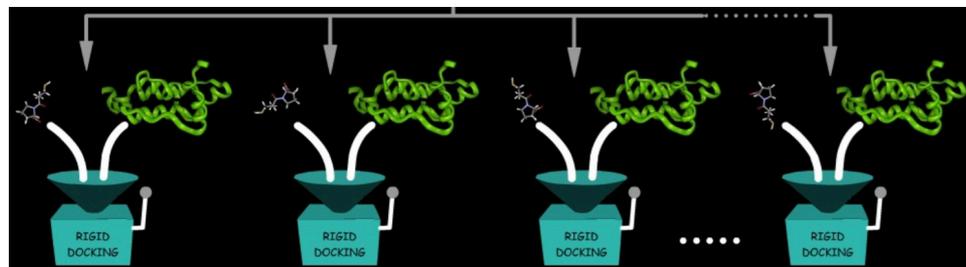
МОЛЕКУЛЯРНЫЙ ДОКИНГ



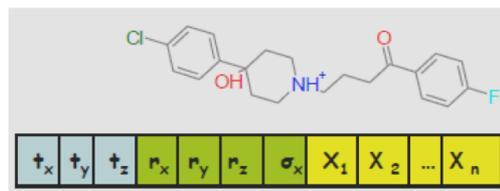
МОЛЕКУЛЯРНЫЙ ДОКИНГ: УЧЕТ СТЕПЕНЕЙ СВОБОДЫ ЛИГАНДА

Учет гибкости лиганда

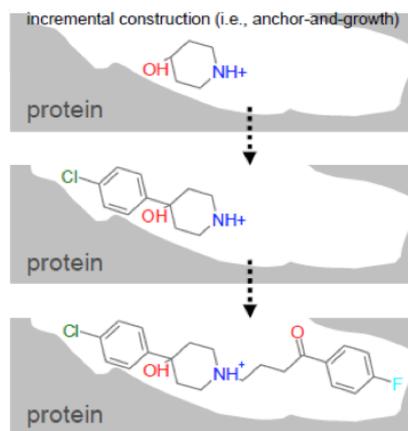
Генерация библиотек конформеров



Стохастические методы



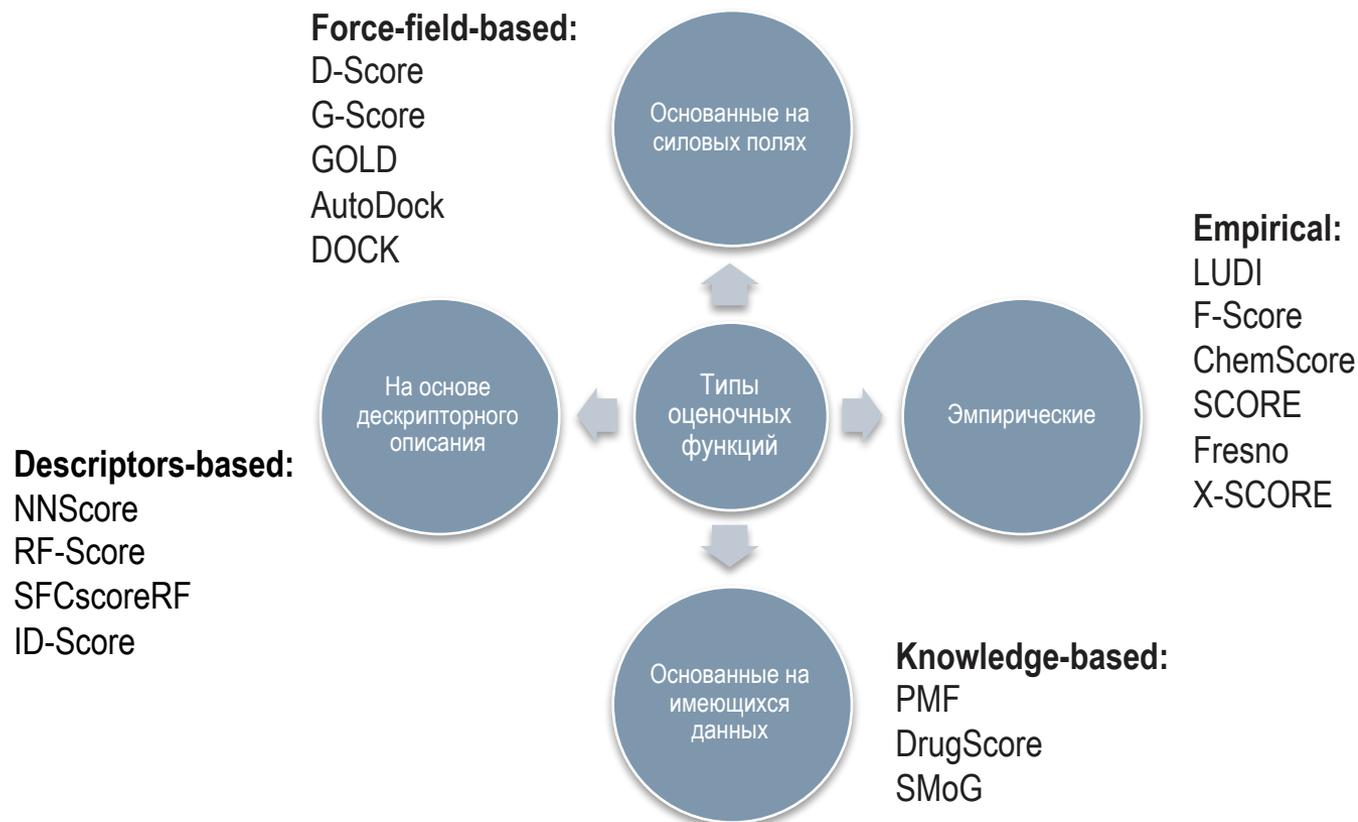
Инкрементный подход



МОЛЕКУЛЯРНЫЙ ДОКИНГ: ОЦЕНОЧНЫЕ ФУНКЦИИ

В процессе докинга служат мерой качества подгонки при оптимизации расположения лиганда

После завершения докинга используются для ранжирования лигандов, для которых было найдено решение задачи докинга



Classification of Current Scoring Functions

J. Liu et al *Journal of Chemical Information and Modeling* **2015** 55 (3), 475-482

DOI: 10.1021/ci500731a

МОЛЕКУЛЯРНЫЙ ДОКИНГ: ОЦЕНОЧНЫЕ ФУНКЦИИ

Оценочные функции, основанные на силовых полях

Для оценки взаимодействий используются классические силовые поля молекулярной механики и делается предположение об аддитивности вкладов в величину энергии взаимодействия.

$$\Delta G_{binding} = \Delta E_{VDW} + \Delta E_{electrostatic} + \Delta E_{H-bond} + \Delta G_{desolvation}$$

Эмпирические оценочные функции

Используют различные зависимости, описывающие свойства, важные для связывания лигандов, для записи так называемого основного уравнения, предсказывающего энергию связывания белка и лиганда.

$$ChemScore = S_{H-bond} + S_{metal} + S_{lipophilic} + P_{rotor} + P_{strain} + P_{clash} + P_{covalent} + P_{constraint}$$

Множественная линейная регрессия используется для оптимизации весовых коэффициентов при вычисленных функциях, с использованием обучающей выборки лиганд-белковых комплексов, для которых известна аффинность и доступна структура.

Эти функции могут описывать полярные взаимодействия (водородные связи и ионные взаимодействия), неполярные взаимодействия (липофильные и ароматические взаимодействия), потерю гибкости лиганда (энтропия) и эффекты десольватации.

МОЛЕКУЛЯРНЫЙ ДОКИНГ: ОЦЕНОЧНЫЕ ФУНКЦИИ

Оценочные функции, основанные на имеющихся данных

Преобразуют структурную информацию, полученную из рентгеноструктурных данных для лиганд-белковых комплексов, в свободные энергии Гельмгольца взаимодействий пар атомов белка и лиганда.

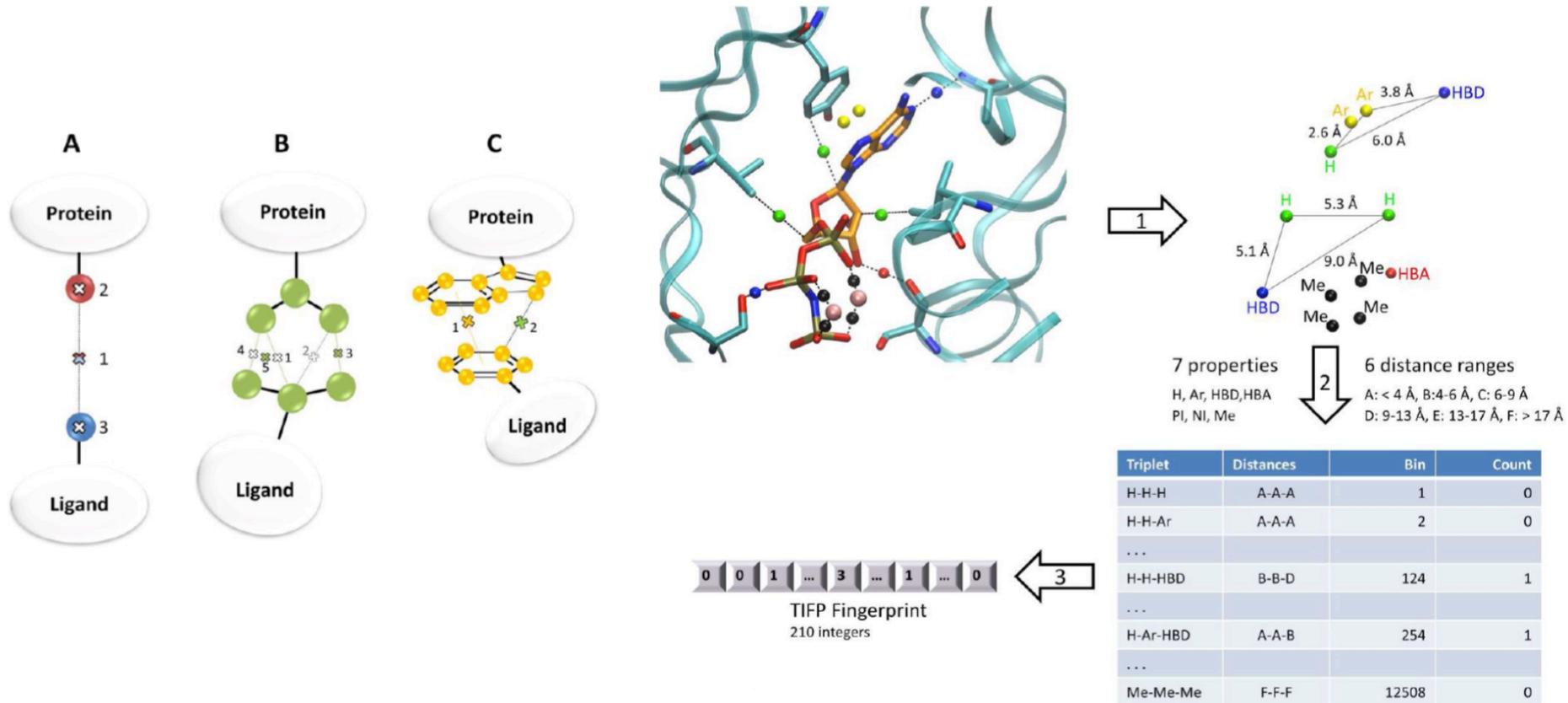
Предполагается, что чем чаще атом белка типа i и атом лиганда типа j находятся на расстоянии r_{ij} , тем выгоднее такое взаимодействие. Каждому типу взаимодействия между атомом белка и атомом лиганда приписывается свободная энергия лиганд-белкового взаимодействия в зависимости от его частоты.

Функция распределения вычисляется по численной плотности вхождений этой пары на расстоянии r в базу лиганд-белковых комплексов. Оценка определяется как сумма по всем межатомным взаимодействиям в комплексе.

$$A = \sum_{lig} \sum_{prot} \omega_{ij}(r) \quad \omega_{ij}(r) = -k_B T \ln[g_{ij}(r)] = -k_B T \ln \left[\frac{\rho_{ij}(r)}{\rho_{ij}^*} \right]$$

МОЛЕКУЛЯРНЫЙ ДОКИНГ: ОЦЕНОЧНЫЕ ФУНКЦИИ

Оценочные функции, основанные на дескрипторном описании



Encoding Protein-Ligand Interaction Patterns in Fingerprints and Graphs

J. Desaphy et al *Journal of Chemical Information and Modeling* 2013 53 (3), 623-637

DOI: 10.1021/ci300566n

ФИЛЬТРОВАНИЕ РЕЗУЛЬТАТОВ ДОКИНГА

Задача: уменьшение числа виртуальных хитов, для которых необходимо провести биологические испытания – докинг нескольких десятков тысяч соединений часто приводит к списку хитов из нескольких тысяч соединений → необходимы продуманные стратегии обработки результатов.

ФИЛЬТРОВАНИЕ РЕЗУЛЬТАТОВ ДОКИНГА

По топологическим свойствам

С помощью консенсусных подходов

С помощью комбинированных вычислительных процедур

По химическому разнообразию

На основании визуальной оценки

ФИЛЬТРОВАНИЕ РЕЗУЛЬТАТОВ ДОКИНГА

Фильтрация по топологическим свойствам

Оценка комплексов с точки зрения стерического соответствия:

- доля объема лиганда, погруженную внутрь полости связывания,
- размер липофильных карманов на поверхности взаимодействия лиганда с белком,
- площадь доступной для растворителя поверхности лиганда,
- число близких контактов между атомами лиганда и белка, не связанными водородными связями.

Фильтрация с помощью консенсусных подходов

Совместное использование различных методов докинга и оценочных функций

Фильтрация с помощью комбинированных вычислительных процедур

Последующее применение для отобранных решений более точных оценочных функций

Фильтрация по химическому разнообразию

Отбор по химическому сходству

УЧЕТ ГИБКОСТИ БЕЛКА

Protein flexibility

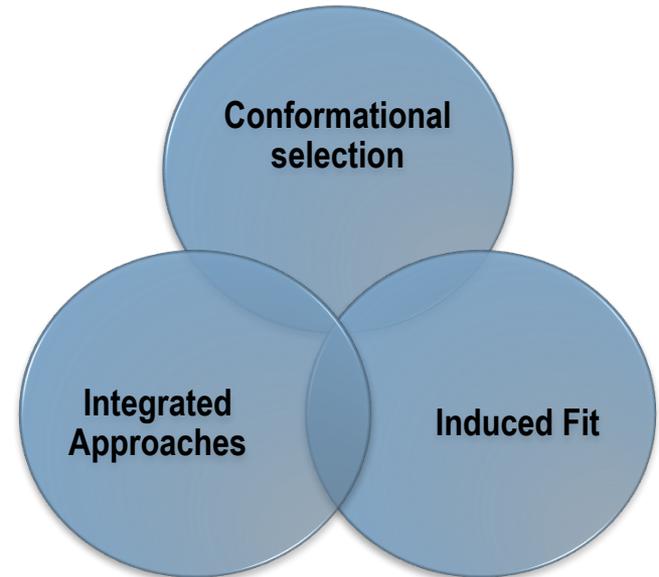
Small atom fluctuations

Side chain flexibility

Backbone flexibility

Domain movements

Protein flexibility analysis approaches



УЧЕТ ГИБКОСТИ БЕЛКА: ВЫБОР ОПТИМАЛЬНЫХ КОНФОРМАЦИЙ

Макромолекула-мишень описывается посредством ансамбля различных конформаций

Получение ансамблей

Разновидности докинга

Экспериментальные методы

Методы молекулярной динамики

Монте-Карло сэмплинг

Анализ гармонических колебаний

На основе ранее имеющейся информации по комплексам белок-лиганд

Последовательный докинг

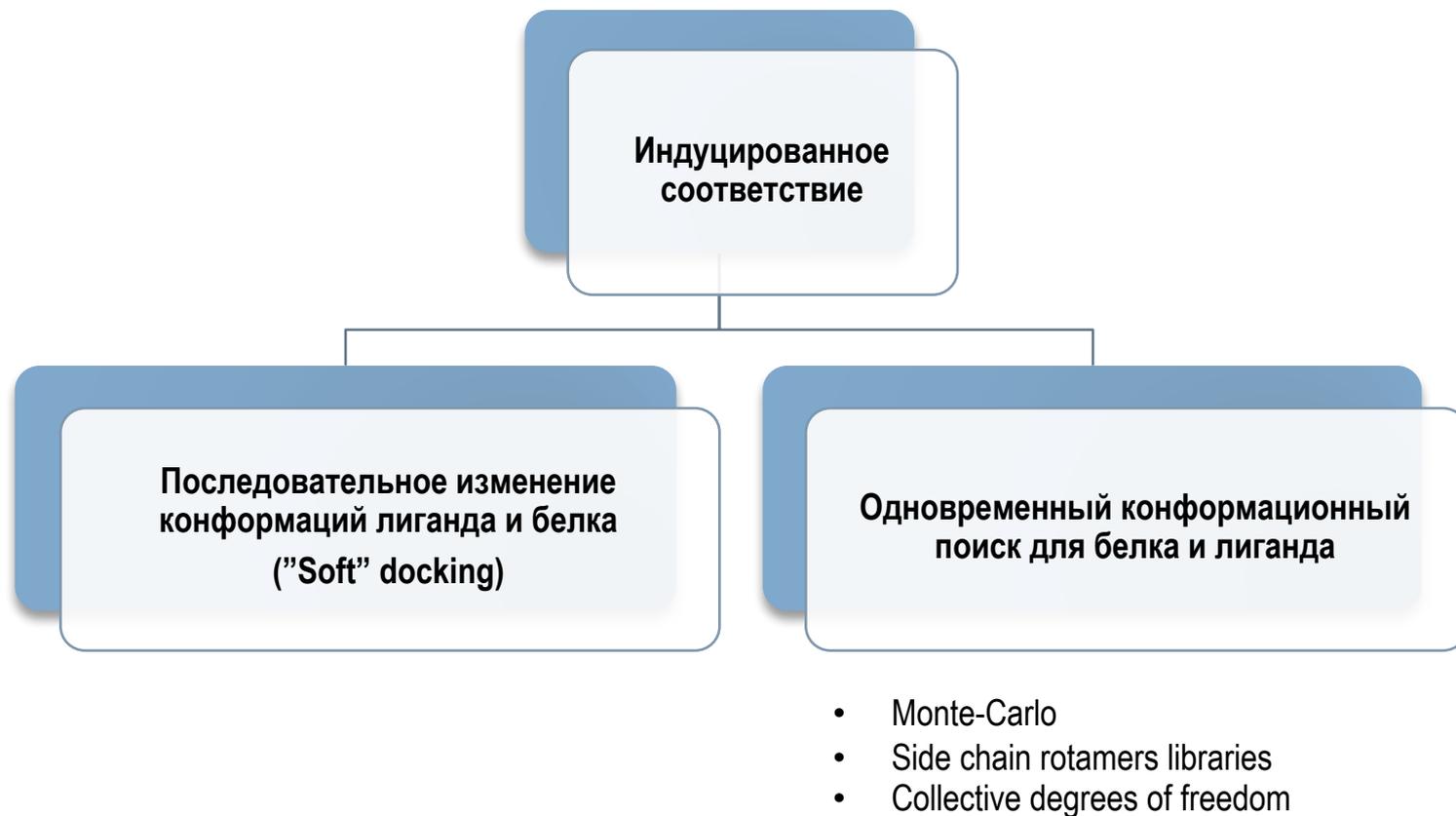
Оценка пространства, занимаемого ансамблем, определение наиболее вероятной конформации

Объединение ансамбля в единую структуру, только несовпадающие подструктуры оптимизируются

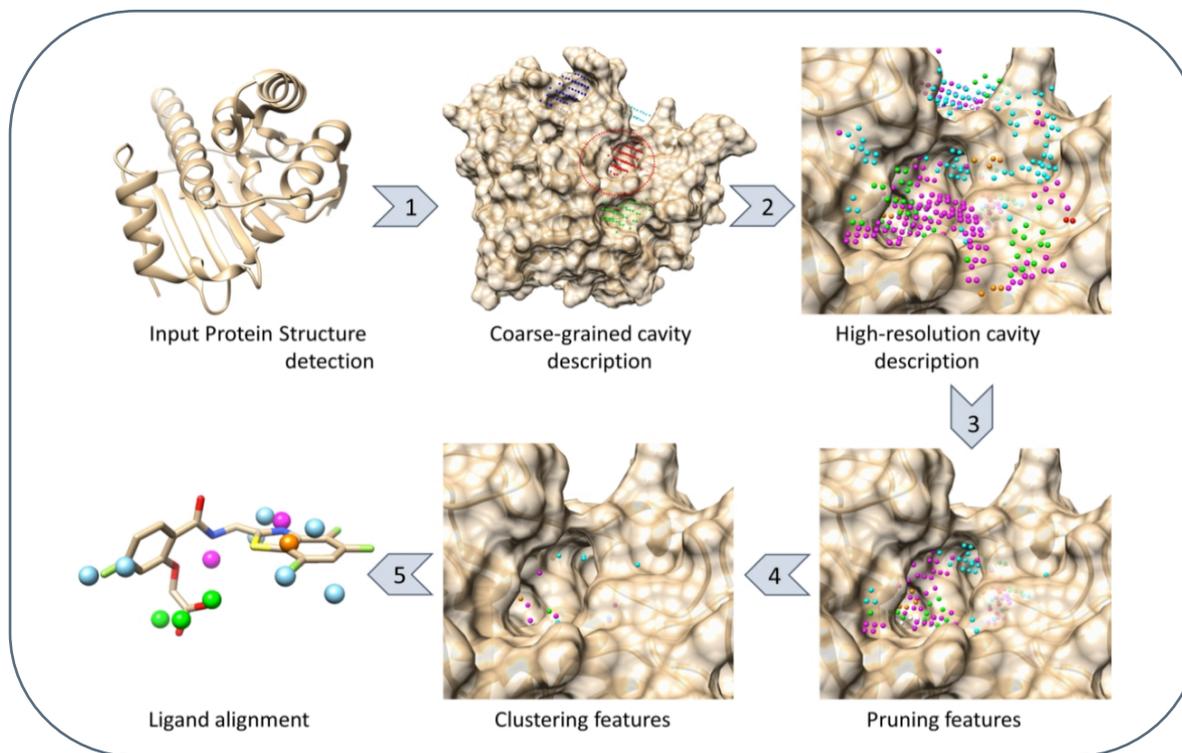
Использование 3D решетки с оценкой потенциальной энергии взаимодействия

УЧЕТ ГИБКОСТИ БЕЛКА: ИНДУЦИРОВАННОЕ СООТВЕТВИЕ

Допускается несоответствие лиганда сайту связывания с возможностью проникновения лиганда из полости в непосредственно область поверхности рецептора



IChem: software to detect, analyze and compare protein-ligand interactions



References:

Desaphy, J., Azdimousa, K., and Rognan, D. (2012) Comparison and druggability prediction of protein-ligand binding pockets from pharmacophore-annotated shapes., *J. Chem. Inf. Model.*, **52**, 2287-2299

Desaphy, J., Ducrot, P., Raimbaud, E. and Rognan, D. (2013) Encoding protein-ligand interaction patterns in fingerprints and graphs. *J. Chem. Inf. Model.*, **53**, 623-637

Desaphy, J. and Rognan, D. (2014) scPDBFrag: a database of protein-ligand interaction patterns for bioisosteric replacements. *J. Chem. Inf. Model.*, **54**, 1908-1918

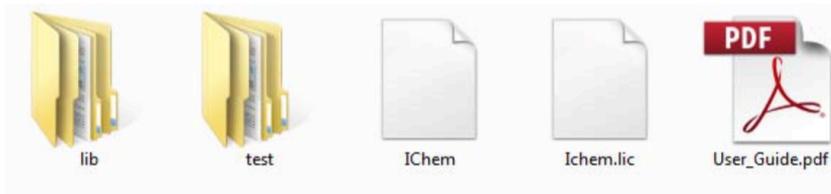
Slynko, I. Da Silva, F., Bret, G. and Rognan, D. (2016) Docking pose selection by interaction pattern graph similarity: application to the D3R grand challenge 2015. *J. Comput.-Aided Mol. Des.*, **30**, 669-683

IChem: software to detect, analyze and compare protein-ligand interactions



Structural Chemogenomics Group

Laboratoire d'Innovation Therapeutique, UMR7200 CNRS/Universite de Strasbourg

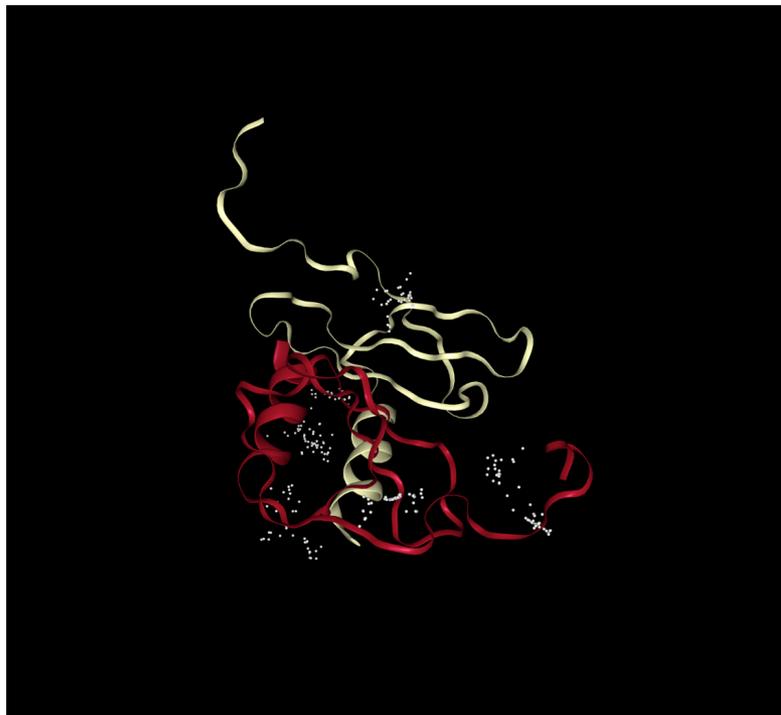


- **lib**: a directory containing necessary template files
- **test**: a directory containing some test input/output files
- **Ichem.lic** a license file
- **IChem**: A 64-bit Linux executable (CentOS 7.1)
- **User_Guide.pdf**: this manual

Tool	Purpose
Binding mode tools	
IFP	Detect and save interactions as fingerprints (binding-site dependent)
ints	Detect and save interaction patterns(binding-site independent)
grim	Detect and save interactions as graphs
Generic tools	
genkey	License file generator (privileged usage)
license	Output license details
pdbconv	Convert PDB into MOL2 files
realign	Rotation/translation of atomic coordinates
sims	Fingerprint comparison
utils	Miscellaneous (buried surface area, ligand fragmentation)
Cavity detection	
Volsite	Cavity detection and druggability/ligandability prediction
Protein-Protein interfaces	
DetectPPI	Cavity detection and druggability/ligandability prediction

Detecting ligand-binding cavities: hydrogen adding with PROTOSS

Protoss is a fully automated hydrogen prediction tool for protein-ligand complexes. It adds missing hydrogen atoms to protein structures (PDB-format) and detects reasonable protonation states, tautomers, and hydrogen coordinates of both protein and ligand molecules.



Representation options:	Surface options:	Viewer options:	More options:
Protein Cartoon	Surface Off	Background Black	Screenshot -Resolution-
Ligand Licorice	Color White	Mouse behavior Default	View Reset
Solvent Off	Opacity 100%	View Perspective View	NGL More info
Clip width Clip distance	<input type="range"/>		

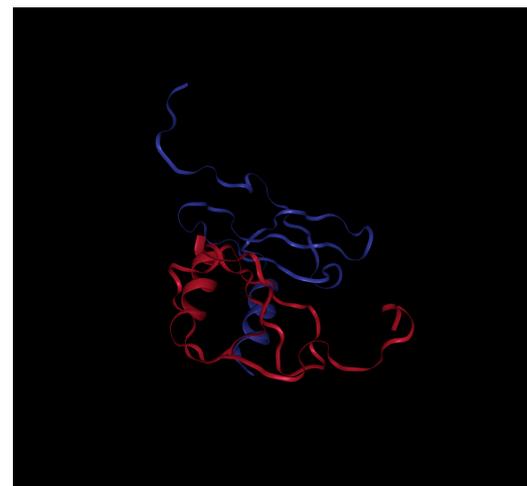


PDB-Code or search term:

① Upload Protein (PDB format): pdb4uai_out.pdb

① Upload Ligand (SDF format): no file selected

[Advanced search](#)



Result

Protoss output info:

PDB entry filter removed 240 atom entries

You can download the full output info by clicking on the 'Download Log' button.

You can see the result of Protoss in the NGL Viewer on the left side. You can switch between the original file and the protoss file by check/uncheck the checkbox modified

Display Protoss file in NGL Viewer instead of original PDB file

ПРОГРАММНЫЙ ПАКЕТ ROSETTA

ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ

Прогнозирование структуры белка

Молекулярный докинг (protein-protein, protein-ligand)

Определение структуры белка из данных физических методов исследования

Дизайн протеинов