

**LOSCHMIDT
LABORATORIES**



Protein-ligand complexes

Outline

- ❑ Biological relevance
- ❑ Molecular recognition
- ❑ Structure of complexes
- ❑ Protein druggability
- ❑ Small molecules
- ❑ Molecular docking
- ❑ Evaluation of complexes
- ❑ Transport of small molecules

Protein-ligand complexes



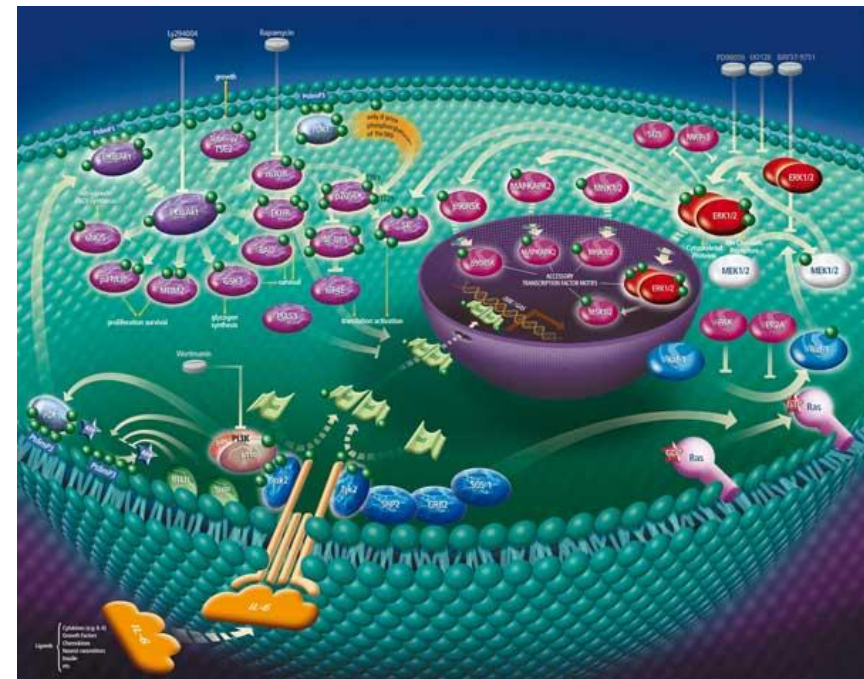
Why do we care?

Examples?

Biological relevance

□ Cell signaling & regulation

- Binding of small molecules to receptors
 - Molecular function of ligands/receptors
 - Selectivity of receptors
- Signaling pathways
- Transport mechanisms
- Homeostasis of the cell
- ...

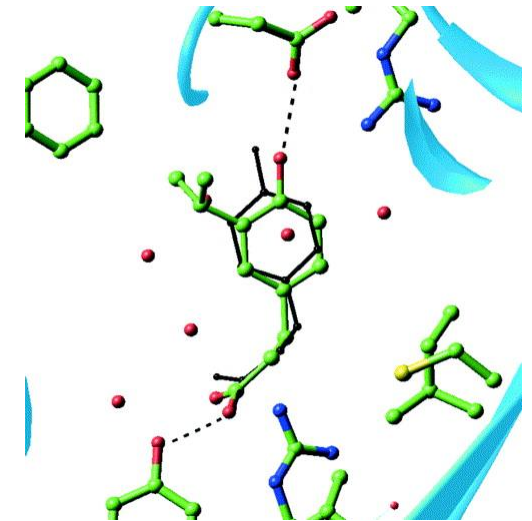


□ **Metabolism**

- Binding of small molecules to enzymes
 - Molecular function of enzymes
 - Activation of enzymes and molecular pathways
 - Bioactivation and clearance of drugs and xenobiotics (P450s,...)
 - Enzymatic cascades
 - Metabolic interferences (competing pathways)
 - ...

□ Drug discovery

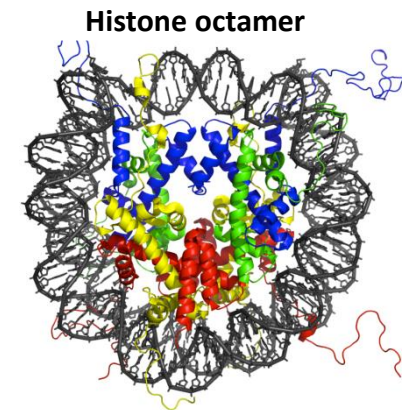
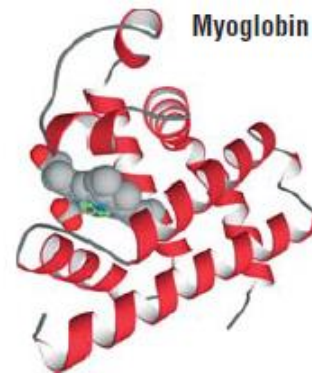
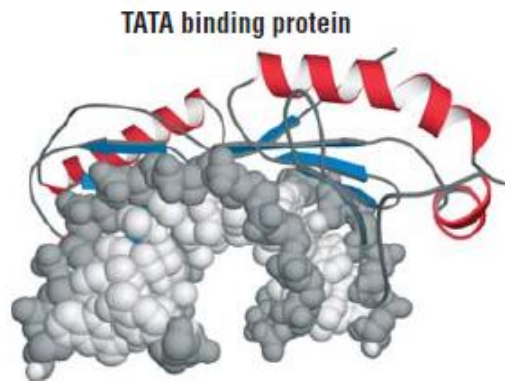
- Binding of small molecules to macromolecules
 - Identification of targets (enzymes, receptors, ...)
 - Identification of potential target inhibitors/activators
 - Optimization of target modulators
 - Repurposing of drugs – finding new receptors
 - Adverse side-effects due to binding to off-targets
 - ...



Biophysical aspects

□ Binding

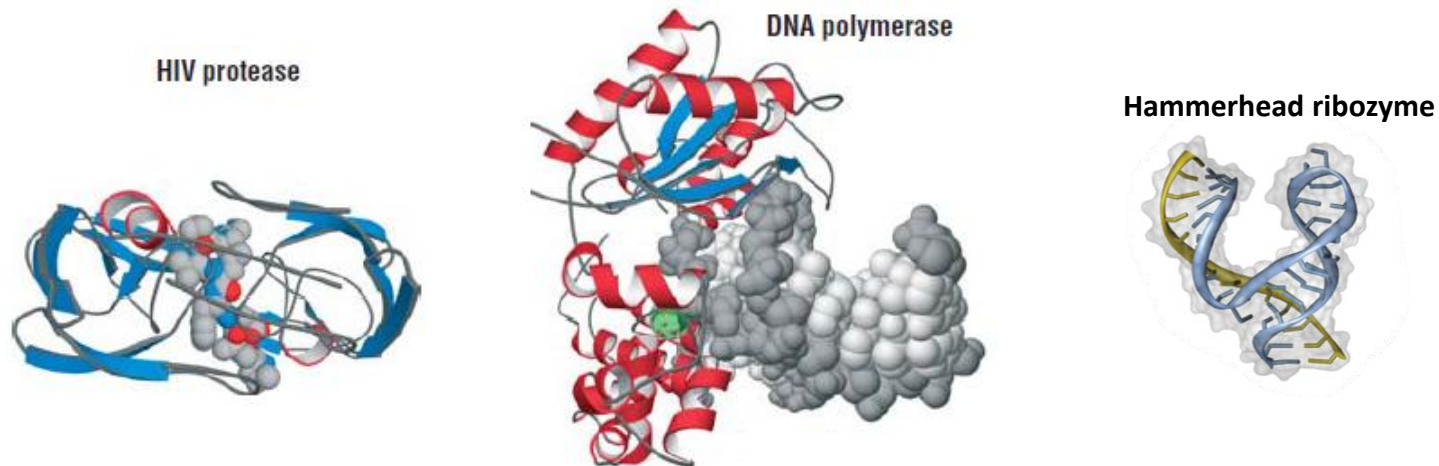
- Specific binding governed by complementarity
 - Geometry and shape
 - Physicochemical properties (interactions)



Biophysical aspects

□ Catalysis

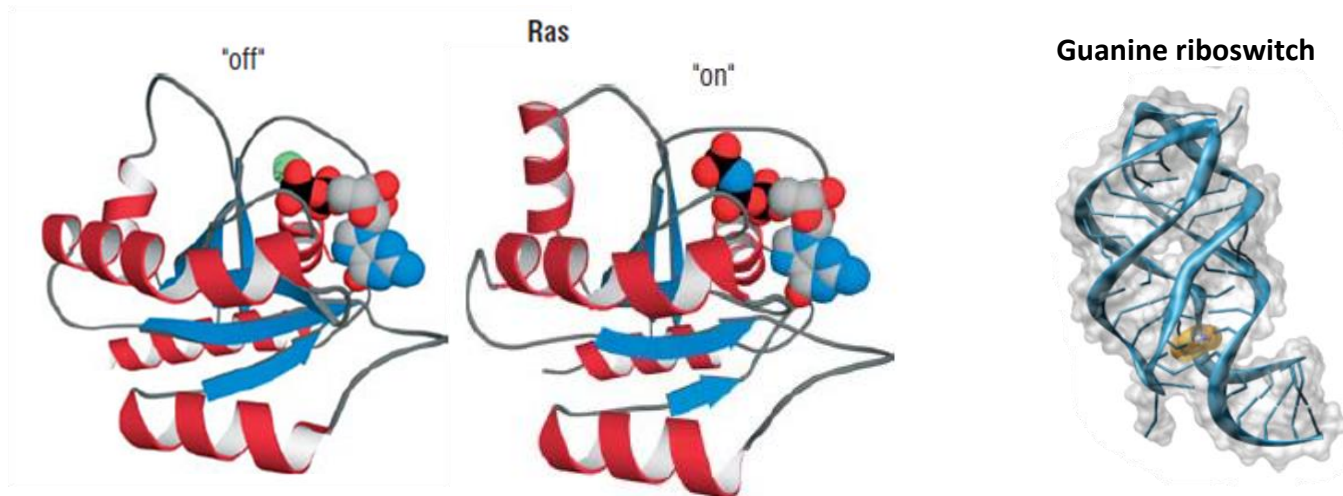
- Chemical reactions can be accelerated up to 17 orders of magnitude
- Binding to active site decreases the energy barrier of the reaction
- Stabilization of Transition State



Biophysical aspects

□ Signaling

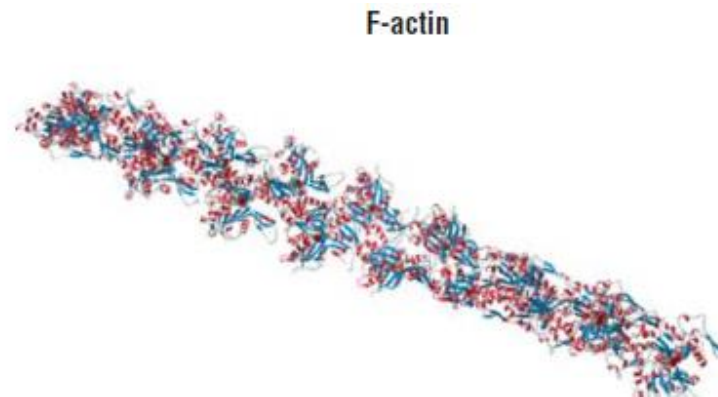
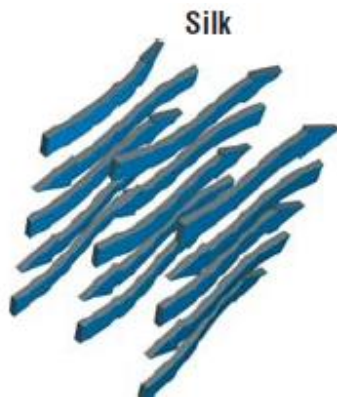
- Conformational changes in response to
 - Ligand binding
 - Properties of surrounding environment (pH, forces...)
- Different conformations recognized by different proteins in signaling pathways → control of cellular processes



Biophysical aspects

□ Formation of complex structures

- Structural elements of complex systems
- Governed by specific association of protein subunits
 - With themselves
 - Other proteins, carbohydrates, lipids, ...



Molecular recognition

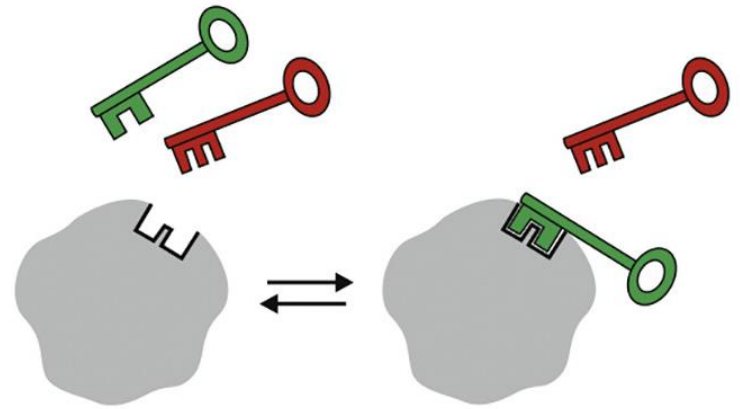


- ❑ Molecular recognition refers to the **specific interactions between two or more molecules** through non-covalent bonding
- ❑ Different biological roles
 - ❑ Specific binding
 - ❑ Catalysis
 - ❑ Signaling
- ❑ Several models to explain molecular recognition

Lock-and-key model



□ E. Fisher – 1894



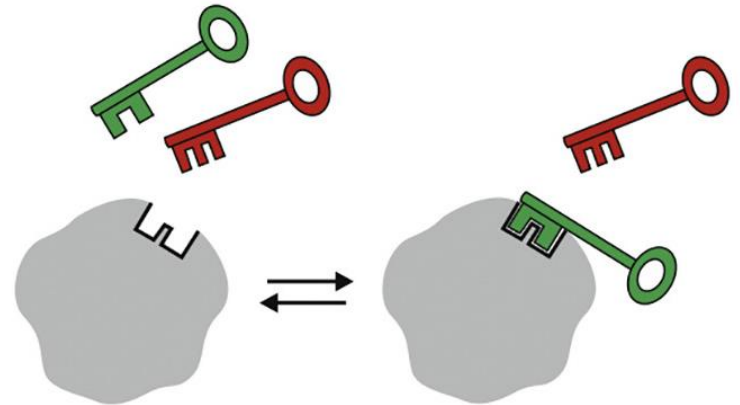
Lock-and-key model



- **E. Fisher – 1894**

- Complementarity between receptor's binding site and the ligand

- Size & shape
- Physicochemical properties



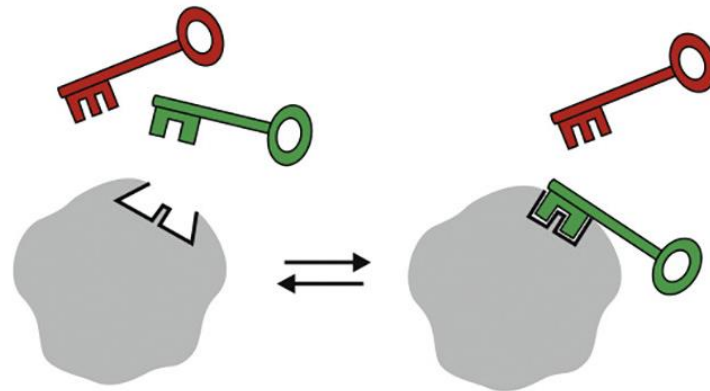
- **Both ligand and receptor are considered rigid**

- Not sufficient to explain allostery, non-competitive inhibition, or catalysis
- → **Model dismissed, only used for educational purposes**

Induced-fit model



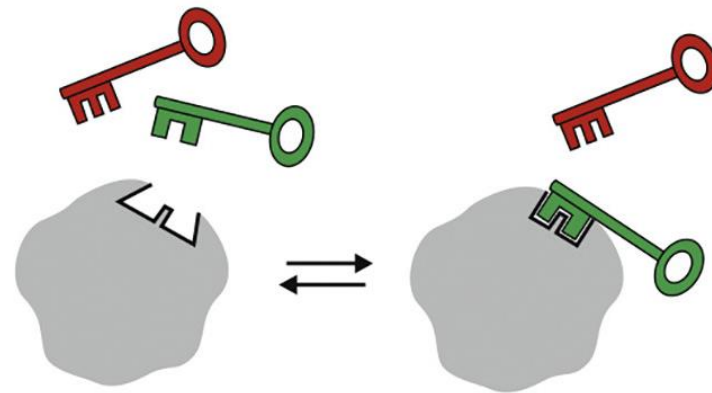
- D. E. Koshland – 1956



Induced-fit model



- D. E. Koshland – 1956
- Only partial complementarity necessary



- Both ligand and receptor can undergo conformational adjustments **upon** complexation
 - Conformation of the bound receptor does not exist in its free state

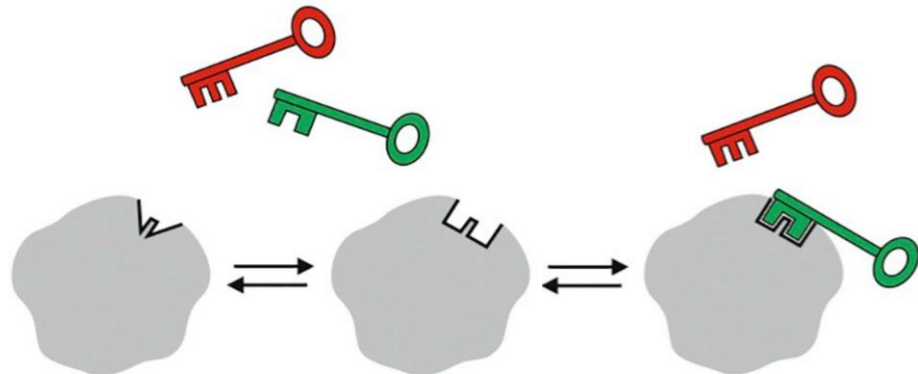
Selected-fit model



□ B. F. Straub – 1964

- This model is also called: *conformational selection, fluctuation-fit or population selection*

□ Receptor and ligand flexible → considered as ensembles



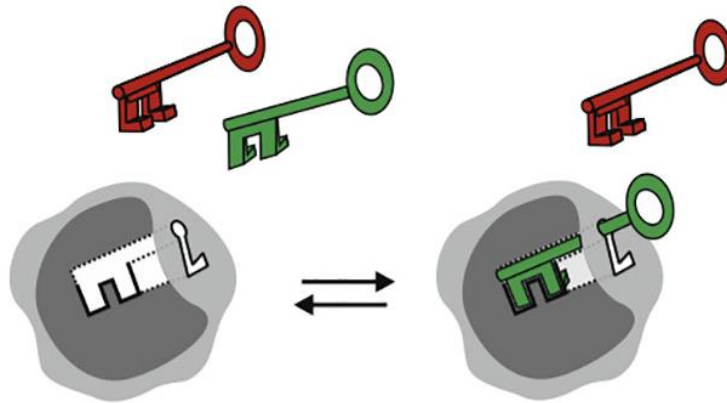
□ Complex is formed in a lock-and-key fashion when two complementary configurations occur

- Conformation of the bound receptor exists also in its free state

Keyhole-lock-key model



- Z. Prokop – 2012
- When the receptor has a buried active site and tunnels

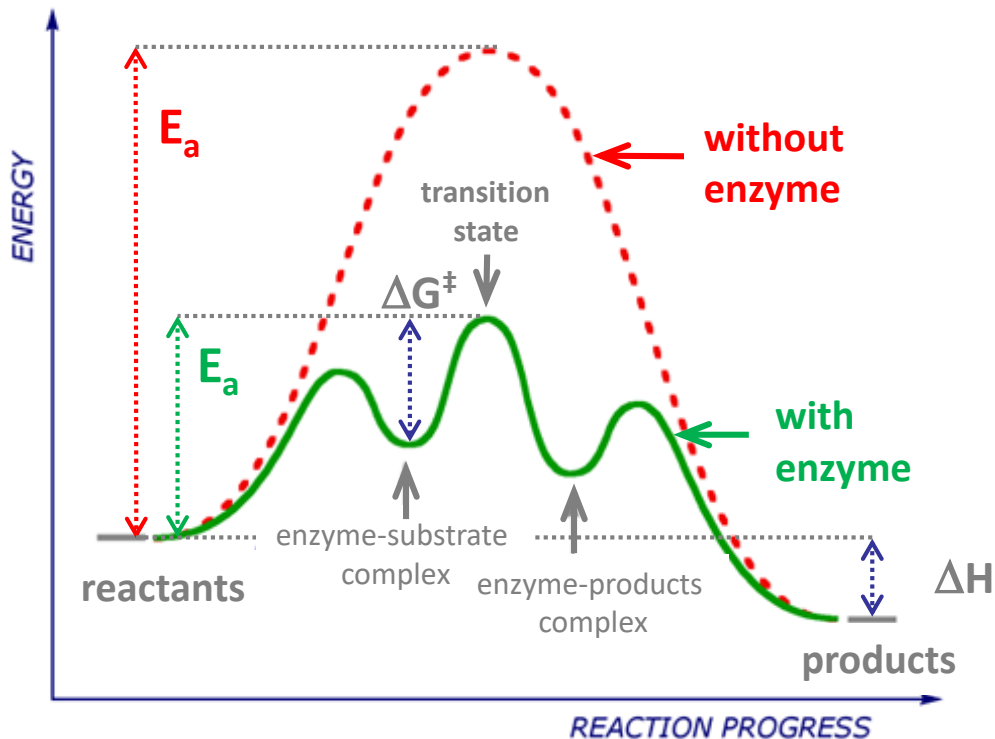


- Complementarity with the ligand is needed **for both** the **active site** and **tunnel**
- Explains the extra selectivity filter provided by the tunnel

Biocatalysis



- Enzymes increase the speed of chemical reactions by decreasing the activation barrier



- Kinetic rate:

$$k = Ae^{\frac{-E_a}{RT}}$$

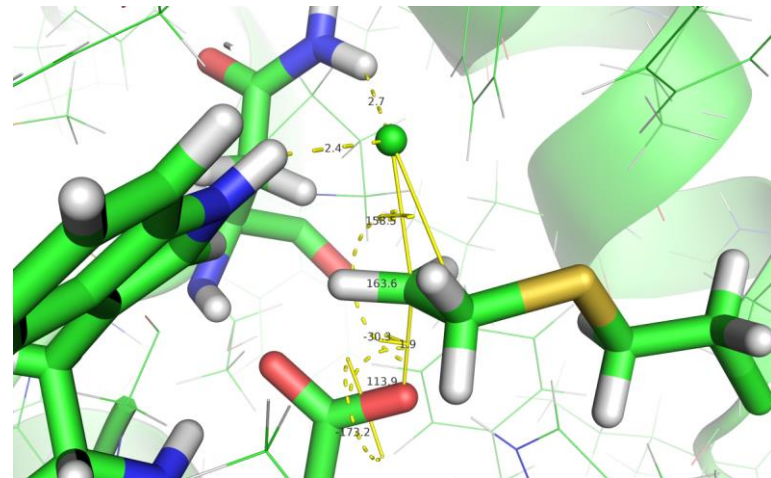
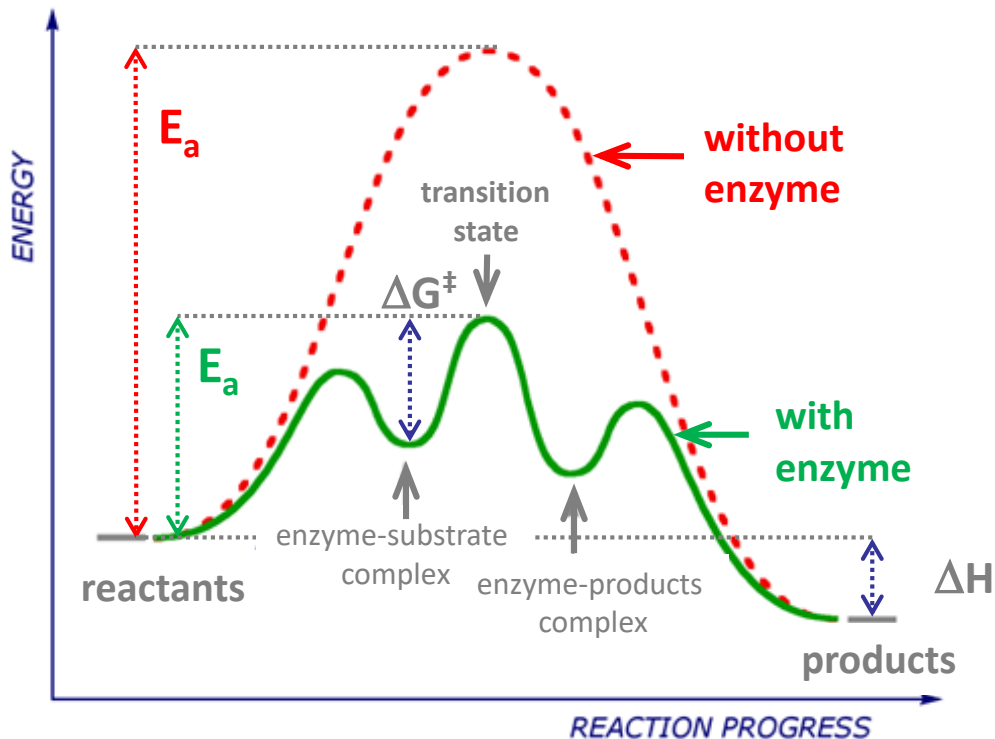
(Arrhenius equation)

- Lower $E_a \rightarrow$ higher k
(faster reaction)

Biocatalysis



- Enzymes increase the speed of chemical reactions by decreasing the activation barrier
- Provide environments that stabilize the transition state(s)



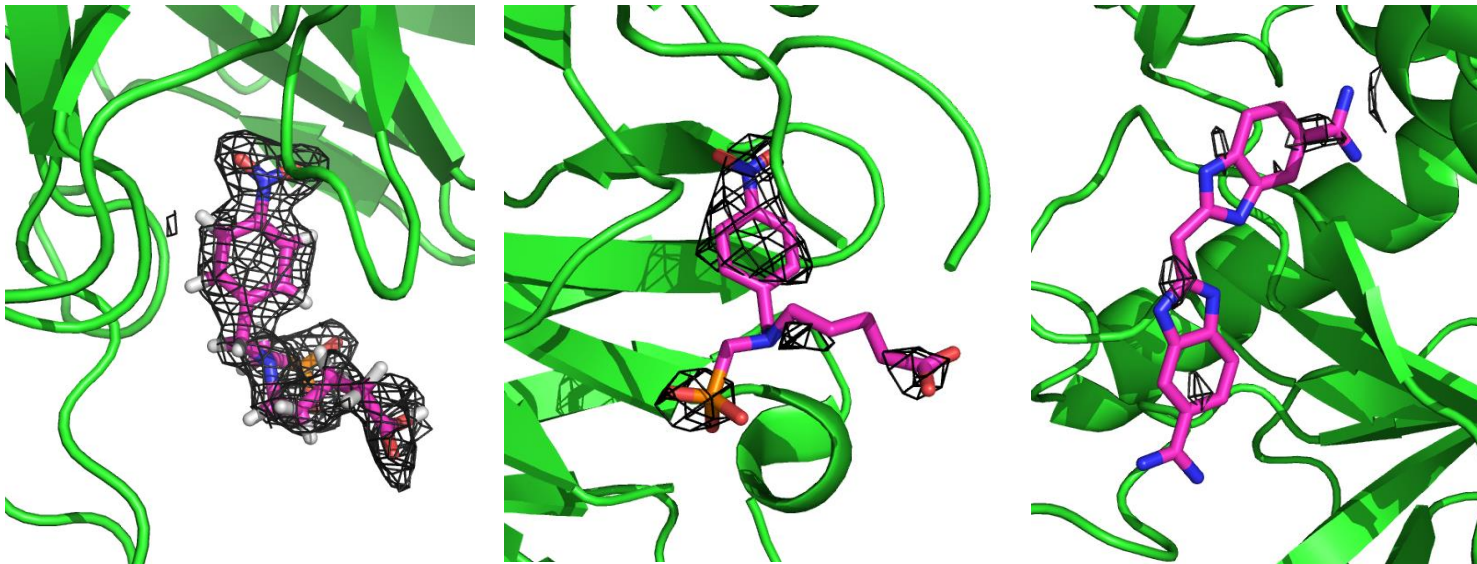
Structure of complexes



- ❑ Complexes in RSCB PDB
- ❑ Databases of complexes
 - PDBbind
 - BindingDB
 - ChEMBL
 - ...
- ❑ Experimentally determined complexes!

Complexes in RSCB PDB

- ❑ **Limited number of available complexes**
 - >180,000 protein structures
 - >101,000 structures with ligands
- ❑ **Limited information on conformation of bound ligand**
 - Ligands often quite mobile -> uncertainties -> **need to be verified**



Databases of complexes

□ PDBbind

- <http://www.pdbbind.org.cn>
- Binding **affinity** data and **structural** information on >16,500 complexes
 - >13,500 protein-ligand
 - >120 nucleic acid-ligand
 - >800 protein-nucleic acid
 - >2,000 protein-protein complexes
- Data collected from >29,000 original references
- Provides also a "refined set" and "core set" compiled as high-quality data sets of protein-ligand complexes for docking/scoring studies

Databases of complexes

□ PDBbind

Welcome to the PDBbind Database!

Current version: 2012
Total entries: 9,308

HOME | BROWSE | DATA | STRUCTURE | SEQUENCE | DOWNLOAD

Search For Complexes

AND Search in Protein-Ligand core set

AND PDB ID (e.g. 1a or 1a7x)

AND Protein Name (e.g. kinase)

AND Ligand Name (e.g. 3-mer)

AND EC Number . . . (e.g. 2.5.1.2 or 3. .4.2)

AND Release Year From To (e.g. 1998 To 2006)

AND Resolution From To (e.g. 1.7 To 2, or NMR)

AND Affinity (pkd/pki/pIC50) From To (e.g. 4.3 To 8.5)

Search Clear

Save Result As: PDF Format Excel Format

Download structure files: Ligand Protein Protein-Ligand Pocket-Ligand

REGISTER | ADMIN | DEPOSIT | FEEDBACK | APPLICATION | LOGOUT

PDB ID | Quick Search

This site has been visited 313560 times since Nov 2007.

©2007, Shanghai Institute of Organic Chemistry (SIOC), CAS. IPI CP 备 05005485

| ID | ✓ | PDB | Subset | Resolution | Kd | pKd | Release | Protein Name |
|----|---|------|----------------|------------|-------------|-------|---------|-------------------------------------|
| 1 | ✓ | 1a30 | Protein-Ligand | 2 | Ki=50uM | 4.3 | 1998 | HIV-1 protease |
| 2 | ✓ | 1adl | Protein-Ligand | 1.6 | Kd=4.4uM | 5.36 | 1994 | trp rna-binding attenuation protein |
| 3 | ✓ | 1b39 | Protein-Ligand | 2.1 | Kd=0.120uM | 6.92 | 1998 | cyclin dependent kinase 2 |
| 4 | ✓ | 1b7h | Protein-Ligand | 2 | Kd=0.0095uM | 8.02 | 1998 | oligo-peptide binding protein |
| 5 | ✓ | 1b8o | Protein-Ligand | 1.5 | Ki=23pM | 10.64 | 1999 | purine nucleoside phosphorylase |
| 6 | ✓ | 1b9j | Protein-Ligand | 1.8 | Kd=1100nM | 5.96 | 1999 | oligo-peptide binding protein |
| 7 | ✓ | 1bcu | Protein-Ligand | 2 | Kd=0.53mM | 3.28 | 1998 | thrombin alpha |
| 8 | ✓ | 1nwl | Protein-Ligand | 2.4 | Ki=4.1mM | 2.39 | 2003 | tyrosine phosphatase 1b |
| 9 | ✓ | 1bxo | Protein-Ligand | .95 | Ki=0.10nM | 10 | 1998 | penicillopepsin |
| 10 | ✓ | 1bxq | Protein-Ligand | 1.41 | Ki=42nM | 7.38 | 1998 | penicillopepsin |
| 11 | ✓ | 1c1v | Protein-Ligand | 1.98 | Ki=0.023uM | 7.64 | 2000 | thrombin alpha |
| 12 | ✓ | 1c88 | Protein-Ligand | 1.8 | Ki=5.1uM | 5.29 | 2000 | tyrosine phosphatase 1b |
| 13 | ✓ | 1d09 | Protein-Ligand | 2.1 | Ki=27nM | 7.57 | 2000 | aspartate carbamoyltransferase |
| 14 | ✓ | 1d7j | Protein-Ligand | 1.85 | Kd=500uM | 3.3 | 1999 | FK506 binding protein |
| 15 | ✓ | 1e66 | Protein-Ligand | 2.1 | Ki=0.13nM | 9.89 | 2001 | acetylcholinesterase |
| 16 | ✓ | 1f4k | Protein-Ligand | 1.7 | Kd=0.2uM | 9.7 | 1999 | FK506 binding protein |

Databases of complexes

□ PDBbind

The screenshot displays the PDBbind website interface. At the top left, it says "Welcome to the PDBbind Database!". The top right corner shows the PDBbind logo, the current version (2012), and the total number of entries (9,308). Below the header is a navigation menu with options: HOME, BROWSE, DATA, STRUCTURE, SEQUENCE, and DOWNLOAD. The main content area is divided into two sections. On the left, a chemical structure is displayed, showing a complex molecule with a central chiral center, a sulfur atom, and various functional groups including a carboxylate group and a protonated amine group. On the right, there is a search interface titled "Search by Ligand Structure". It includes a text input field for drawing a new structure, buttons for "Get SMILES", "Put SMILES", and "Clear All". Below this, there is a section for using a known ligand as a template, with fields for "PDB ID" (10gs) and "Get Template" (10gs in refined set). It also shows the "Protein Name" (glutathione s-transferase) and "Ligand Name" (3-mer). There are navigation buttons for the refined set. Further down, there are search options for "Search In:" (General set) and "Similarity Search", and a "Similarity cutoff:" (100%). At the bottom of the search interface, there are "Display Options:" including checkboxes for "Edit Tools", "Auto Scale", "Clean 2D", and "Explicit H", and a dropdown for "Display Implicit H In:" (Hetero or Termin). The bottom of the page features a footer with navigation links (REGISTER, ADMIN, DEPOSIT, FEEDBACK, APPLICATION, LOGOUT), a visitor counter (313556), a copyright notice (©2007 Shanghai Institute of Organic Chemistry (SIOC), CAS), and a license number (沪ICP备05005485).

Databases of complexes

□ PDBbind



Welcome to the PDBbind Database!

Current version: 2012
Total entries: 9,308

HOME | BROWSE | DATA | STRUCTURE | SEQUENCE | DOWNLOAD

BLAST Search By Sequence

Use a known sequence in PDBbind-CN as template, modify and then conduct search.

PDB ID: 10gs

Complex Type: Protein-Ligand

Protein Name: glutathione s-transferase

Ligand Name: 3-mer

Protein or Nucleotide Query Sequence: using 10gs as template

```
>10GS:A|PDBID|CHAIN|SEQUENCE
PPYTVVYFPVVRGRCAALRMLLADQGQSWKEEVVTVETWQEGSLKASCLYGQLPKFQDGD
LTLYQSNILRHLGRITLGLYG
KDQQEAAALVDMVNDGVEDLRCKYISLIYTNYEAGDDYVKALPGQLKPFETLLSQNGGK
TFIVGDQISFADYNLLDLL
IHEVLAPGCLDAFPLLSAYVGRLSARPKLKAFASPEYVNLPIGNGKQ
>10GS:B|PDBID|CHAIN|SEQUENCE
PPYTVVYFPVVRGRCAALRMLLADQGQSWKEEVVTVETWQEGSLKASCLYGQLPKFQDGD
LTLYQSNILRHLGRITLGLYG
KDQQEAAALVDMVNDGVEDLRCKYISLIYTNYEAGDDYVKALPGQLKPFETLLSQNGGK
TFIVGDQISFADYNLLDLL
IHEVLAPGCLDAFPLLSAYVGRLSARPKLKAFASPEYVNLPIGNGKQ
>10GS:G|PDBID|CHAIN|SEQUENCE
ECG
>10GS:H|PDBID|CHAIN|SEQUENCE
```

Type: Protein in PDBbind

REGISTER | ADMIN | DEPOSIT | FEEDBACK | APPLICATION | LOGOUT

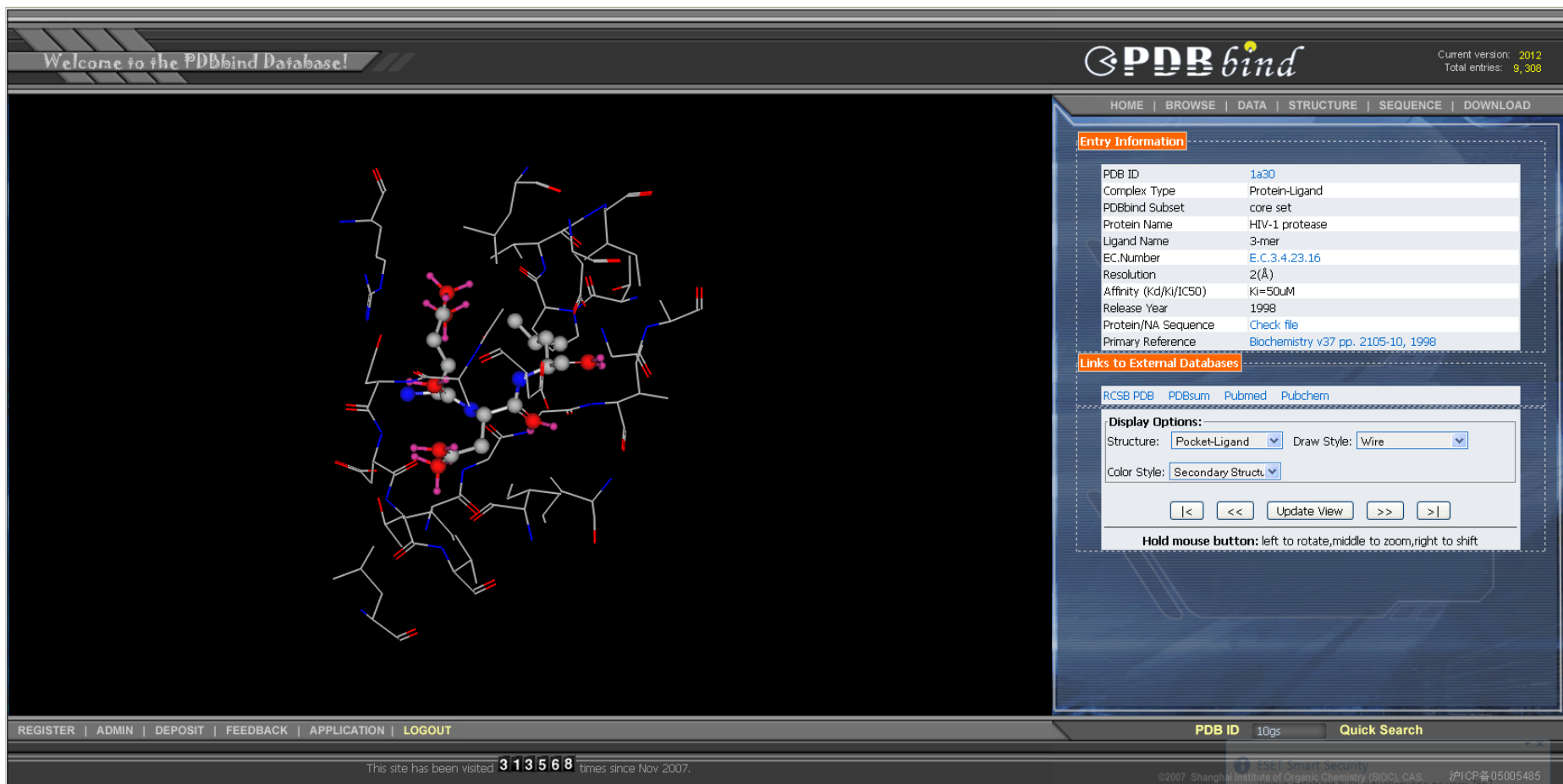
PDB ID: 10gs

This site has been visited 313561 times since Nov. 2007.

©2007, Shanghai Institute of Organic Chemistry (SIOC), CAS. 沪ICP备05005485

Databases of complexes

□ PDBbind



Welcome to the PDBbind Database!

Current version: 2012
Total entries: 9,308

HOME | BROWSE | DATA | STRUCTURE | SEQUENCE | DOWNLOAD

Entry Information

| | |
|-----------------------|--|
| PDB ID | 1a30 |
| Complex Type | Protein-Ligand |
| PDBbind Subset | core set |
| Protein Name | HIV-1 protease |
| Ligand Name | 3-mer |
| EC Number | E.C.3.4.23.16 |
| Resolution | 2(Å) |
| Affinity (Kd/Ki/IC50) | Ki=50uM |
| Release Year | 1998 |
| Protein/NA Sequence | Check file |
| Primary Reference | Biochemistry v37 pp. 2105-10, 1998 |

Links to External Databases

[RCSB PDB](#) [PDBsum](#) [Pubmed](#) [Pubchem](#)

Display Options:

Structure: Draw Style:

Color Style:

Hold mouse button: left to rotate, middle to zoom, right to shift

REGISTER | ADMIN | DEPOSIT | FEEDBACK | APPLICATION | LOGOUT

PDB ID: 10gs Quick Search

This site has been visited **313568** times since Nov 2007.

©2007 Shanghai Institute of Organic Chemistry (SIOC), CAS. ICP备05005485

Databases of complexes

□ BindingDB

- www.bindingdb.org
- Focus on the interactions of proteins considered to be **drug-targets** with **drug-like** molecules
- Contains about 1,500,000 entries of binding data
 - >7,000 protein targets
 - >650,000 small molecules
- Crystal structures of complexes with measured affinity
 - >2,500 – for proteins with 100% sequence identity
 - >6,000 – for proteins up to 85% sequence identity

Databases of complexes

□ BindingDB



[myBDB](#) [logout](#)

Search and Browse

Target

[Sequence](#)

[Name &](#)

[K_i IC50](#) [K_d EC50](#)

[ΔG°](#) [ΔH°](#) [-TΔS°](#)

[pH \(Enzymatic Assay\)](#)

[pH \(ITC\)](#)

[Substrate or Competitor](#)

[Compound Mol. Wt.](#)

[Chemical Structure](#)

[Source Organism](#)

[Source Organism](#)

[Number of Compounds](#)

[Monomer List in csv](#)

[Het List in SDF](#)

Compound

[FDA Drugs](#)

[Chemical Structure](#)

[Name](#)

[SMILES](#)

[Number of Data / Targets](#)

Special tools

[Find My Compound's](#)

[Targets](#)

[Do Virtual Screening](#)

The Binding Database

[Home](#)

[Info](#)

[Download](#)





[About us](#)

[Email us](#)

[Contribute data](#)

BindingDB is a public, web-accessible database of measured binding affinities, focusing chiefly on the interactions of protein considered to be drug-targets with small, drug-like molecules. BindingDB contains 910,836 binding data, for 6,263 protein targets and 378,980 small molecules.

There are 1717 protein-ligand crystal structures with BindingDB affinity measurements for proteins with [100%](#) sequence identity, and 4937 crystal structures allowing proteins to [85%](#) sequence identity.

| | |
|---|--|
| Full Search Article Titles, Authors, Assays, Compound Names, Target Names | <input type="text"/> <input type="button" value="Go"/> Use ? for single-letter wild-card or * for general wild-card. For example, "adeny*" or "adeny?". Query cannot start with wild card. |
| Messages | <ol style="list-style-type: none">Downloads now allow you to obtain data subsets, such those curated by BindingDB staff and hence not routinely available elsewhere; a cleaned version of PDSP K_i; an unpublished dataset provided by the P. Taylor lab at UCSD; and others.Citation information on pages like this now generally includes a link to email the corresponding author. |
| myBDB | Username <input type="text"/> Password <input type="text"/> <input type="button" value="login"/> logout Username is your registered email in BindingDB. register |
| Video Tutorials |  1:11 Get all data from an article  1:29 Download all data for a target of interest  1:12 Find and view all data for a target of interest  1:53 Find my compound's targets |

BindingDB News

June, 2012. BindingDB now includes essentially all data from [PDSP K_i Database](#).

June, 2012. BindingDB has completed curation of all issues through April 2012 of [Nature Chem Biol](#), [ACS Chemical Biol](#), [Chem & Biol](#), [J. Chem Biol](#), [BMC Chem Biol](#), [Chem Biol and Drug Des](#), [Chembiochem](#), [Bioorg Chem](#), and [J. Enz Inhib Med Chem](#).

June, 2012. BindingDB now allows data downloads in CSV format, in addition to SDF.

June, 2012. Data pages now provide direct links to source Articles, where available.

March, 2012. Added video tutorials to help get started with BindingDB.

January, 2012. A new [Find My Compound's Target](#) page allows you to enter one or more Compounds and quickly see a list of Targets that your Compound(s) might bind.

Databases of complexes

❑ BindingDB



The Binding Database

Home Info Download About us Email us Contribute data

Compile Data Set for Download or QSAR

Add this page Add all pages Clear Selection Make Data Set

myBDB logout

Search and Browse

Target

Sequence

Name &

Ki IC50 Kd EC50

ΔG° ΔH° ΔTAS°

pH (Enzymatic Assay)

pH (ITC)

Substrate or Competitor

Compound Mol. Wt.

Chemical Structure

Source Organism

Source Organism

Number of Compounds

Monomer List in csv

Het List in SDF

Compound

FDA Drugs

Chemical Structure

Name

SMILES

Number of Data / Targets

Special tools

Find My Compound's

Targets

Do Virtual Screening

Citation

Author

Journal/Citation

Institution

PubMed

PubChem BioAssay

Special Data Sets

Host Guest Complexes

E-MAIL

Found 127 hits

Zinc 0: unavailable per [Zinc DB](#). Zinc 1: purchasable, 2 weeks to supply. Zinc 2: made on demand. Zinc 4: potentially available

| Target (Institution) | Ligand | Target Links | Ligand Links | Trg + Lig Links | Ki nM | ΔG° kJ/mole | IC50 nM | Kd nM | EC50/IC50 nM | k_{off} s^{-1} | k_{on} $M^{-1}s^{-1}$ | pH | Temp $^{\circ}C$ |
|---|--|---|---|---|----------|-------------------------------|------------|----------|-----------------|-----------------------|----------------------------|-----|---------------------|
| CaM-kinase kinase beta (Homo sapiens) University of Dundee Curated by ChEMBL | CHEMBL265470 (7-oxo-7H-benzimidazo[2,1-a]benz[de]isoquinoline-3-ylidene)acetic acid | PDB MMDB KEGG UniProtKB/SwissProt GoogleScholar | ChEMBL PC cid PC sid ZINC 1 | Article PubMed | n/a | n/a | 10.0 | n/a | n/a | n/a | n/a | n/a | n/a |
| CaM-kinase kinase beta (Homo sapiens) Dept of Oncology, Imperial College London, Hammersmith Hospital Campus, London W12 0NN, England. Curated by ChEMBL | CHEMBL1234833 (CHEBI:797183) | PDB MMDB KEGG UniProtKB/SwissProt GoogleScholar | KEGG PC cid PC sid PDB | Article PubMed | n/a | n/a | 2450.0 | n/a | n/a | n/a | n/a | n/a | n/a |
| CaM-kinase kinase beta (Homo sapiens) Ambit Biosciences Curated by ChEMBL | LY333531 ([(18S)-18-[(dimethylamino)methyl]-17-oxa-4,14,21-tri...) | PDB MMDB KEGG UniProtKB/SwissProt GoogleScholar | B.MOAD ChEMBL MMDB PC cid PC sid PDB | Article PubMed | n/a | n/a | n/a | 1100 | n/a | n/a | n/a | n/a | n/a |
| CaM-kinase kinase beta (Homo sapiens) | SUI1248 | PDB MMDB KEGG | ChEMBL DrugBank MMDB PC cid | Article PubMed | n/a | n/a | n/a | 5800 | n/a | n/a | n/a | n/a | n/a |

Databases of complexes

□ ChEMBL

- <https://www.ebi.ac.uk/chembl/db/>
- Is a manually curated database of bioactive molecules with **drug-like properties**
- Database of binding, functional and ADME (Absorption, Distribution, Metabolism, and Excretion) and toxic. information
- Contains >15,000,000 **activity** data
 - >12,000 protein targets
 - >1,700,000 distinct small molecules
- Data collected from >67,000 original publications
- Smart clustering of relevant information

Databases of complexes

ChEMBL

EMBL-EBI [Terms of Use](#) | [Privacy](#) | [Cookies](#)

Databases Tools Research Training Industry About Us Help [Site Index](#)

EBI > Databases > Small Molecules > ChEMBL Database > Target Search > Target Classification Hierarchy

Search ChEMBLdb... [Activity Source Filter](#)

Browse Protein Target Tree Taxonomy Tree

Click arrows to navigate tree

- Enzyme (3410)
- Membrane receptor (559)
- Ion channel (354)
- Transporter (136)
- Transcription Factor (102)
- Cytosolic other (102)
- Secreted (57)
- Structural (29)
- Surface antigen (25)
- Membrane other (16)
- Adhesion (14)
- Nuclear other (13)

| Target Type | Count |
|----------------------|-------|
| Enzyme | 3410 |
| Membrane receptor | 559 |
| Ion channel | 354 |
| Transporter | 136 |
| Transcription Factor | 102 |
| Cytosolic other | 102 |
| Secreted | 57 |
| Structural | 29 |
| Surface antigen | 25 |
| Membrane other | 16 |
| Adhesion | 14 |
| Nuclear other | 13 |

- Enzyme
- Membrane receptor
- Ion channel
- Transporter
- Transcription Factor
- Cytosolic other
- Secreted
- Structural
- Surface antigen
- Membrane other
- Adhesion
- Nuclear other



ChEMBLdb Statistics


- DB: ChEMBL_14
- Targets: 9,003
- Compound records: 1,376,469
- Distinct compounds: 1,213,239
- Activities: 10,129,256
- Publications: 46,133

Databases of complexes

ChEMBL


EMBL-EBI [Terms of Use](#) | [Privacy](#) | [Cookies](#)

Databases | Tools | Research | Training | Industry | About Us | Help | [Site Index](#)  

ChEMBL 

EBI > Databases > Small Molecules > ChEMBL Database > Compound Search

[Activity Source Filter](#)



ChEMBLdb Statistics

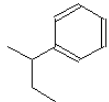
- DB: ChEMBL_14
- Targets: 9,003
- Compound records: 1,376,469
- Distinct compounds: 1,213,239
- Activities: 10,129,256
- Publications: 46,133

ChEMBL Blog

- [New Drug Approvals 2012 - Pt. XXIII -](#)

List Search

SMILES Search ChEMBL ID Search Keyword Search



JME Molecular Editor (c) Peter Ertl

Compound Sketcher:

Databases of complexes

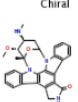
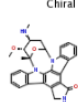
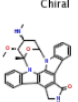
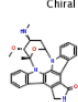
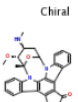
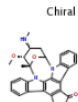
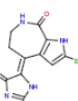
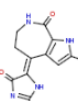
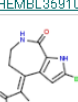
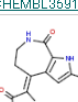
ChEMBL

EMBL-EBI [Terms of Use](#) | [Privacy](#) | [Cookies](#)

Databases | Tools | Research | Training | Industry | About Us | Help | [Site Index](#)

EBI > Databases > Small Molecules > ChEMBL Database > Bioactivity Results

ChEMBL Bioactivity Search Results: 1197 1 2 3 4 5 6 (Next) (End)

| Parent | Ingredient | Bioactivity | Operator | Value | Units | Activity Comment | Assay ChEMBL ID | Assay Source | Assay Type | Description | ChEMBL Target ID | Target Name | Organism | Target Mapping | Curated By | Reference | Name in Reference |
|--|--|-------------|----------|-------|-------|------------------|------------------------------|-----------------------|------------|--|----------------------------|-----------------|--------------|--------------------|------------|--|-------------------|
| Chiral  CHEMBL162 | Chiral  CHEMBL162 | IC50 | < | 1 | nM | | CHEMBL680866 | Scientific Literature | B | Inhibition of Extracellular signal-regulated kinase 2 (Erk2) | CHEMBL4040 | MAP kinase ERK2 | Homo sapiens | Homologous protein | Expert | J. Med. Chem. (2002) 45:17-3772 | Staurosporin |
| Chiral  CHEMBL162 | Chiral  CHEMBL162 | IC50 | = | 2.5 | nM | | CHEMBL729502 | Scientific Literature | B | Inhibition of Mitogen-activated protein kinase (MAPK) phosphorylation by activated MEK-1 | CHEMBL4040 | MAP kinase ERK2 | Homo sapiens | Protein | Expert | J. Med. Chem. (2002) 45:2-529 | staurosporine |
|  CHEMBL361708 |  CHEMBL361708 | IC50 | = | 6 | nM | | CHEMBL628116 | Scientific Literature | B | In vitro inhibitory concentration against human mitogen-activated protein kinase-1 (MEK-1) by using [gamma-33P]-ATP as radioligand | CHEMBL4040 | MAP kinase ERK2 | Homo sapiens | Protein | Expert | Bioorg. Med. Chem. Lett. (2004) 14:16:4319 | Hymenialdisine |
|  CHEMBL359106 |  CHEMBL359106 | IC50 | = | 6 | nM | | CHEMBL729502 | Scientific Literature | B | Inhibition of Mitogen-activated protein kinase (MAPK) phosphorylation by activated MEK-1 | CHEMBL4040 | MAP kinase ERK2 | Homo sapiens | Protein | Expert | J. Med. Chem. (2002) 45:2-529 | 4 |
|  CHEMBL361708 |  CHEMBL361708 | IC50 | = | 9 | nM | | CHEMBL729502 | Scientific Literature | B | Inhibition of Mitogen-activated protein kinase (MAPK) phosphorylation by activated MEK-1 | CHEMBL4040 | MAP kinase ERK2 | Homo sapiens | Protein | Expert | J. Med. Chem. (2002) 45:2-529 | 5 |

ChEMBLdb

Malaria Data

ChEMBL-NTD

Kinase SARfari

GPCR SARfari

DrugEBLity

ChEMBL Group

Downloads

Web Services

FAQ

ChEMBLdb Statistics

- DB: ChEMBL_14
- Targets: 9,003
- Compound records: 1,376,469
- Distinct compounds: 1,213,239
- Activities: 10,129,256
- Publications: 46,133

ChEMBL Blog

- [New Drug Approvals 2012 - Pt. XXIII - Omacetaxine mepesuccinate \(SYNRIEOTM\)](#)
- [Paper: Mapping small molecule binding data to structural domains](#)



❑ Druggability

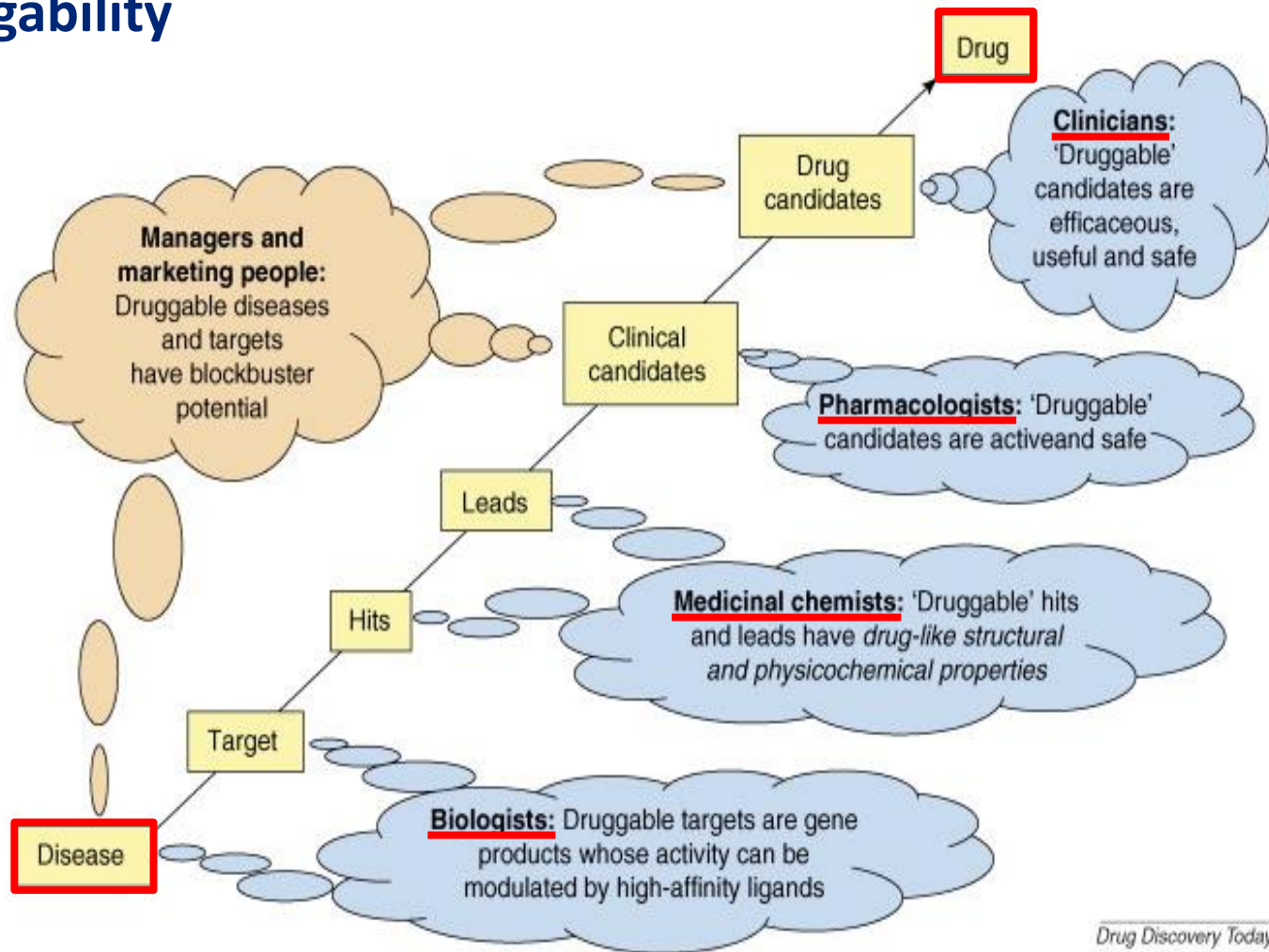
- ❑ Likelihood of a **particular protein** to be modulated or targeted by a **drug-like molecule** in a way that leads to a **therapeutic effect**
- ❑ It means bind with high affinity to selective, bioavailable, low-molecular weight molecules

❑ Lipinski's rule of 5 (for orally-active drugs)

- ❑ MW < 500 Da
- ❑ < 5 H-bond donors (NH, OH); < 10 H-bond acceptors (F, O, N)
- ❑ Partition coefficient ($\log P_{o/w}$) < 5
- ❑ Usually 1 violation is acceptable



□ Druggability





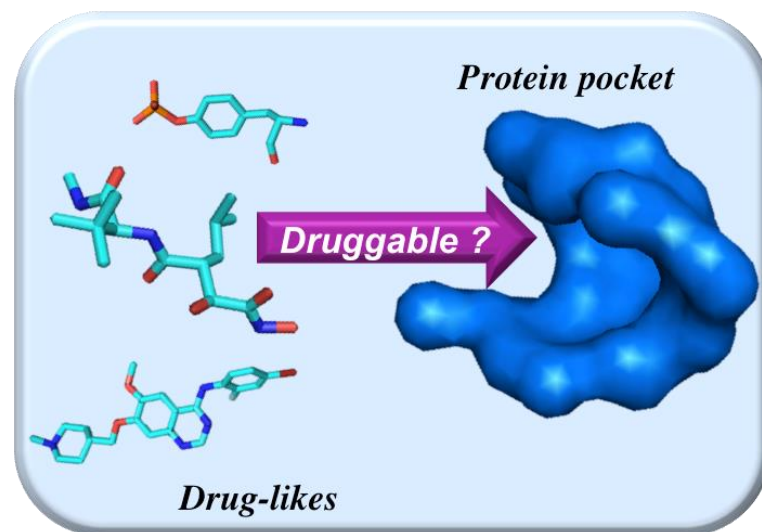
- ❑ **Prediction of protein druggability**
 - By similarity to known target
 - Sequence of binding domain
 - Structural features of binding sites
 - From databases of known targets
 - Predictive tools: PockDrug Server, DoGSiteScorer, ...

- ❑ **Important in target identification phase of drug discovery**
- ❑ Unfortunately, many resources are only private or commercial

Protein druggability server

□ PockDrug-Server

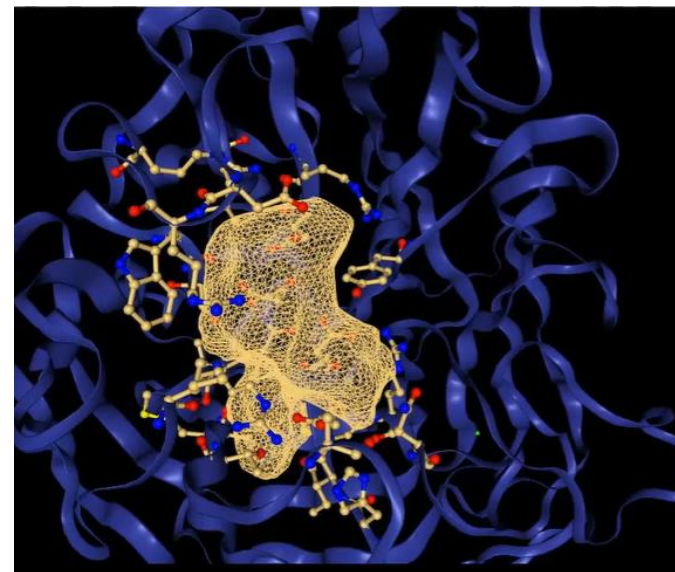
- <http://pockdrug.rpbs.univ-paris-diderot.fr/>
- Automatic tool combining pocket detection, characterization and druggability prediction
- Based on:
 - Physicochemical features
 - Geometry, volume, shape
- Druggability probability for one pocket or to compare two pockets



Protein-ligand interactions server

□ Proteins *Plus*

- <https://proteins.plus/>
- Meta-server providing global support for the initial steps in analysing protein structures
- Structure search, quality assessment, protein pocket detection, protein-ligand and protein-protein interactions
- Predicts binding sites and estimates their druggability (using [DoGSiteScorer](#))





- ❑ Representation of small molecules
- ❑ Databases of small molecule
 - Cambridge Structural Database
 - PUBCHEM database
 - ZINC database
- ❑ Preparation of small molecule structure

Representation of small molecules

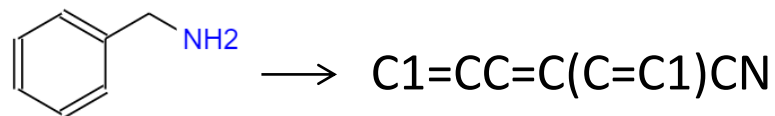
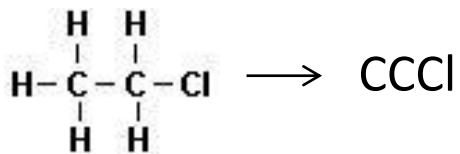


□ 1D – atom based (empirical formula)

- C_2H_5Cl

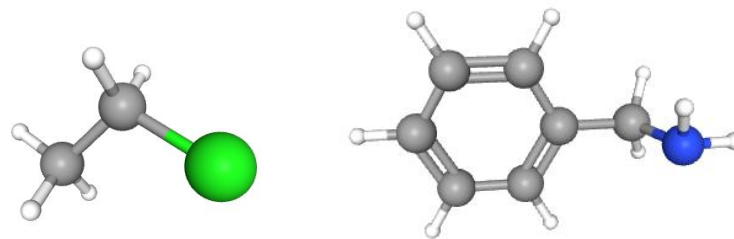
□ 2D – chemical structure diagram -> connection

- Topology or SMILES (Simplified Molecular Line Entry System)

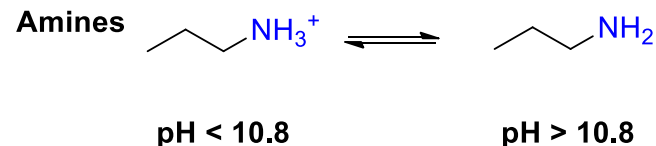
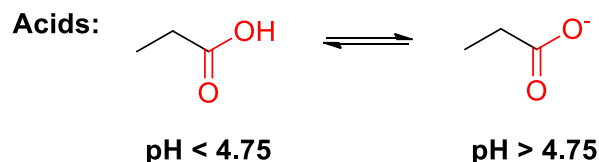


□ 3D – atomic coordinates

- Usually: PDB, SDF or MOL2 files



□ Beware: may have different protonation states



□ Cambridge Structural Database

- <http://www.ccdc.cam.ac.uk/products/csd/>
- The world largest **repository of crystal structures** of small molecules
- >900,000 structures with 3D coordinates available
- CSD is distributed **commercially**
- Free interactive demo for educational purposes
(**only ~750 structures**)
- <https://www.ccdc.cam.ac.uk/Community/educationalresources/teaching-database/>

Databases of small molecule

Cambridge Structural Database

File Filter Help

Find Entry

Entry

- DIFFNO03
- BUQAE01
- BUVGII
- CAACTY
- CACWOS
- CADVEI
- CAFINE
- CAFROR
- CAMHFA
- CAMXAP01
- CAGTET
- CARQOB
- CARTEN
- CARTEN02
- CATCOL13
- CBMZPN01
- CBMZPN02
- CBMZPN03
- CBMZPN10
- CBMZPN11
- CBMZPN12
- CCXAPT
- CEBGUL
- CECZEP
- CECZIT
- CEFXOA
- CEHTAK10

< >

500 Hits

100%

Stop Search

Entry loaded

CAFINE : 1,3,7-Trimethyl-purine-2,6-dione monohydrate
D.J.Sutor; *Acta Crystallogr.* (1958), **11**, 453, doi:10.1107/S0365110X58001286

Hide Viewer

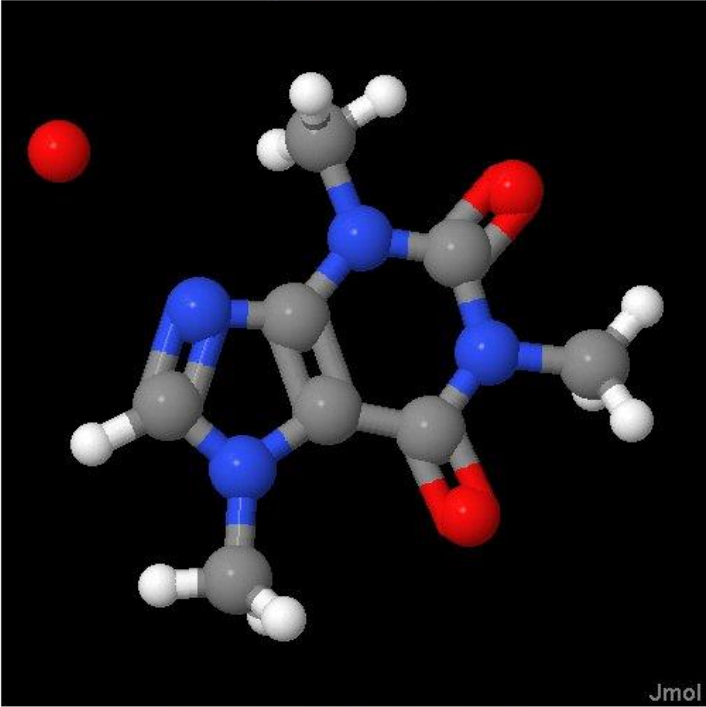
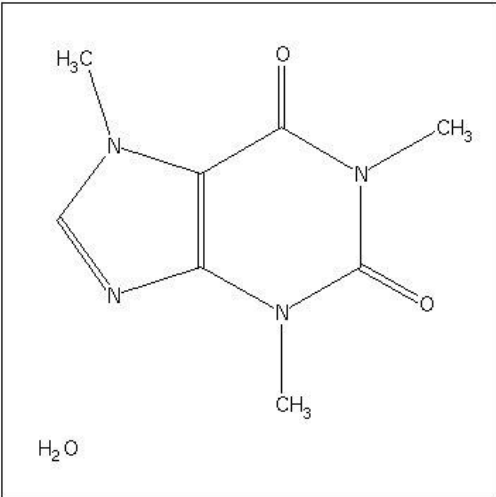


Diagram Details Viewer Export Options Help



$C_8H_{10}N_4O_2 \cdot H_2O$
Space Group: P 2₁/a
a 14.8(1) **b** 16.7(1) **c** 3.97(3)
α 90 **β** 97.0(5) **γ** 90
R-Factor: 14.6%
Temperature (K): Room Temp. (283-303)

Ball and Stick No Labels

Hydrogens Bond types Disorder

Jmol

Databases of small molecule



□ PubChem

- <http://pubchem.ncbi.nlm.nih.gov/>
- **World largest open repository** of experimental data identifying the biological activities of small molecules
- **Substances:** >270 M chemical entities
- **Compound:** >**111 M unique chemical structures**. Compounds may be searched by chemical properties and are pre-clustered by structure comparison into identity and similarity groups
- **BioAssays:** >1.4 M biological experiments
- **Bioactivities:** >300 M biological activity data points

□ ZINC database

- <http://zinc.docking.org/>
- **Free public resource** for ligand discovery
- 3D coordinates in **ready-to-dock formats** (ex: added hydrogens, partial atomic charges, ...)
- Molecules in **biologically relevant** protonation and tautomeric **forms**
- About **37 billion** unique molecules grouped by classes
 - >750,000,000 – commercially available molecules
 - >10,000,000 – drug-like molecules
 - > 5,000 – FDA-approved drugs
 - ...

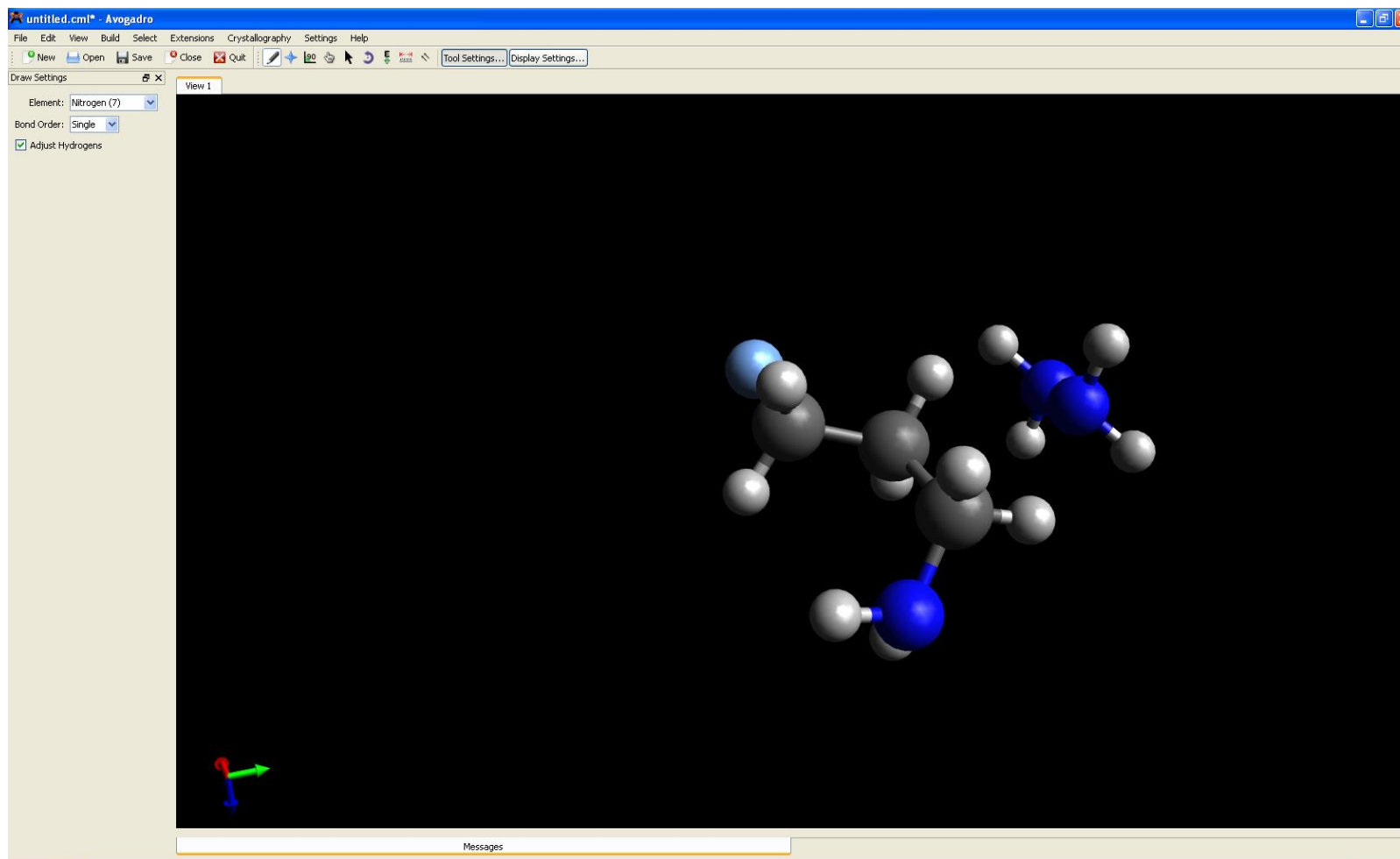
Preparation of small molecule structure

□ AVOGARO

- <https://avogadro.cc/>
- Free, open-source molecule editor and visualizer
- Intuitive & easy to use
- Useful to convert file formats
- Embedded molecular minimization and molecular mechanics
- Interface to quantum chemistry packages

Preparation of small molecule structure

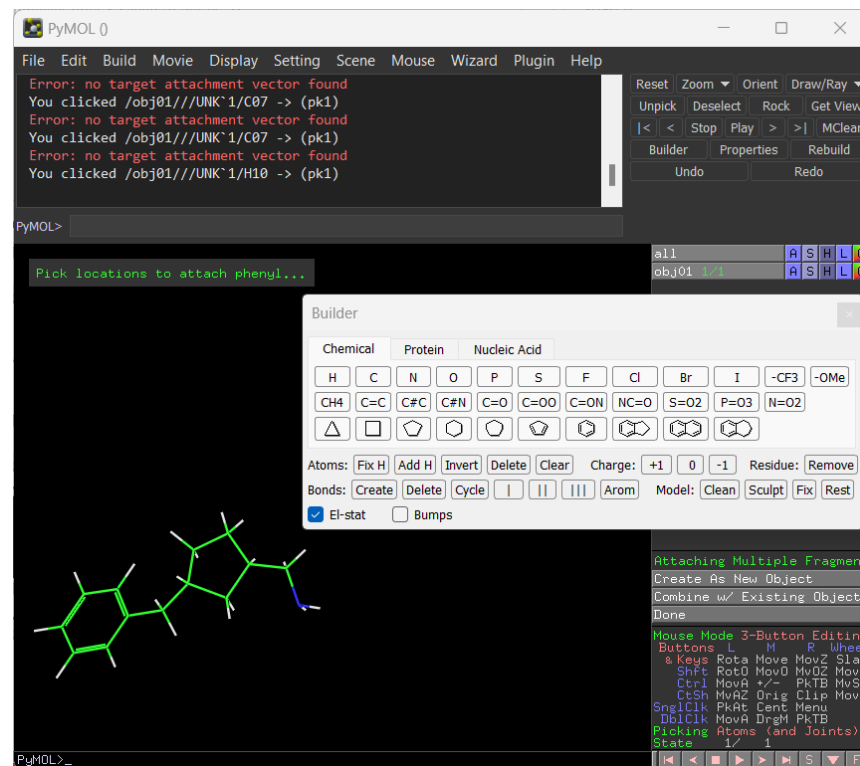
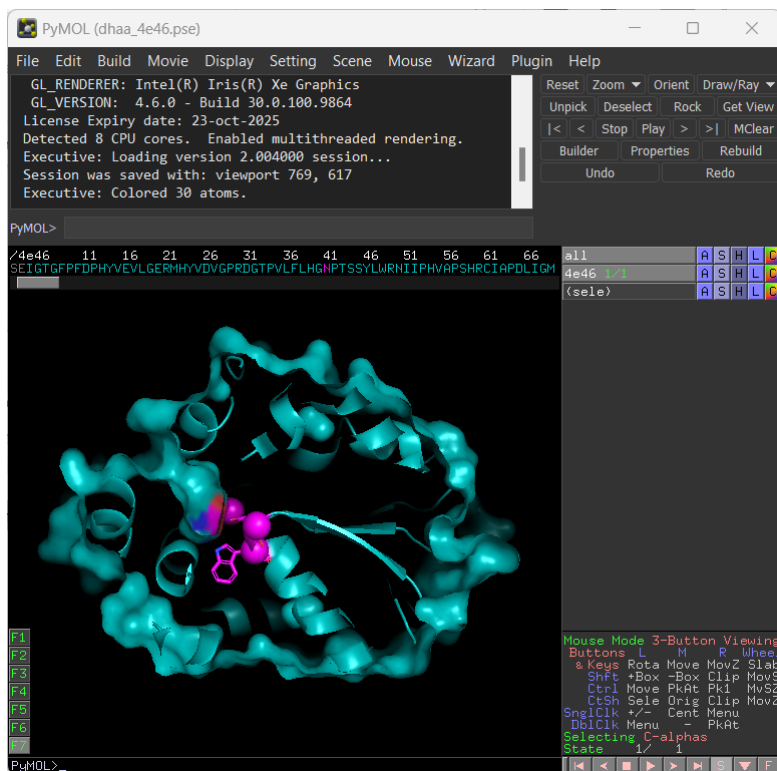
□ AVOGARO



Preparation of small molecule structure

PyMOL

- <https://pymol.org/>
- Powerful molecular visualizer and editor



Preparation of small molecule structure

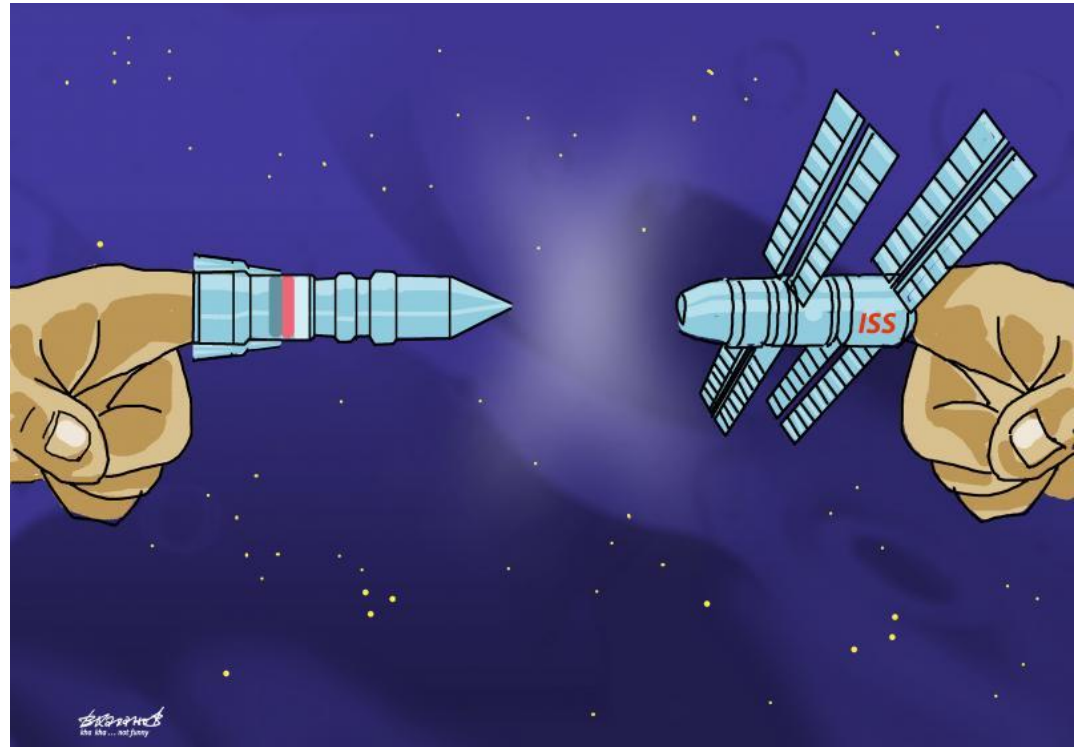
□ Open Babel

- <https://openbabel.org/>
- Free, open-source
- Widely used molecule format converter
- Command line and graphical interface

Molecular docking



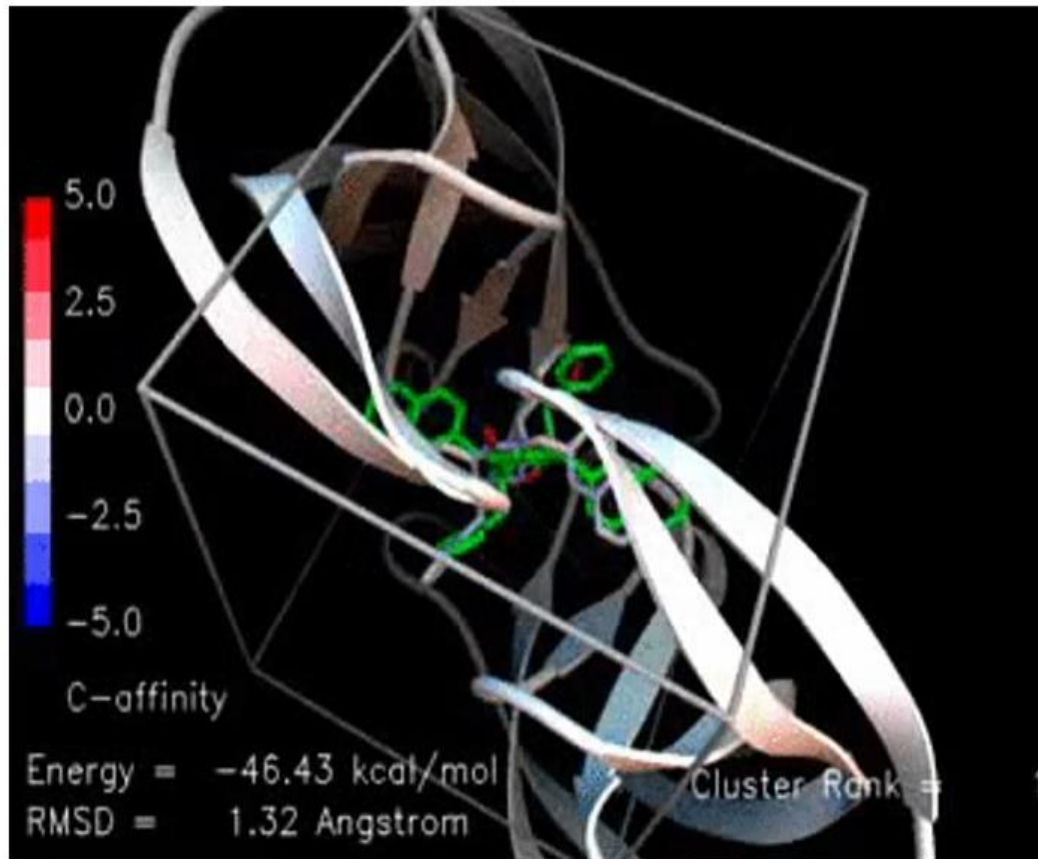
What is it?



Molecular docking



- Useful **when experimental data is not available**
or for virtual screening

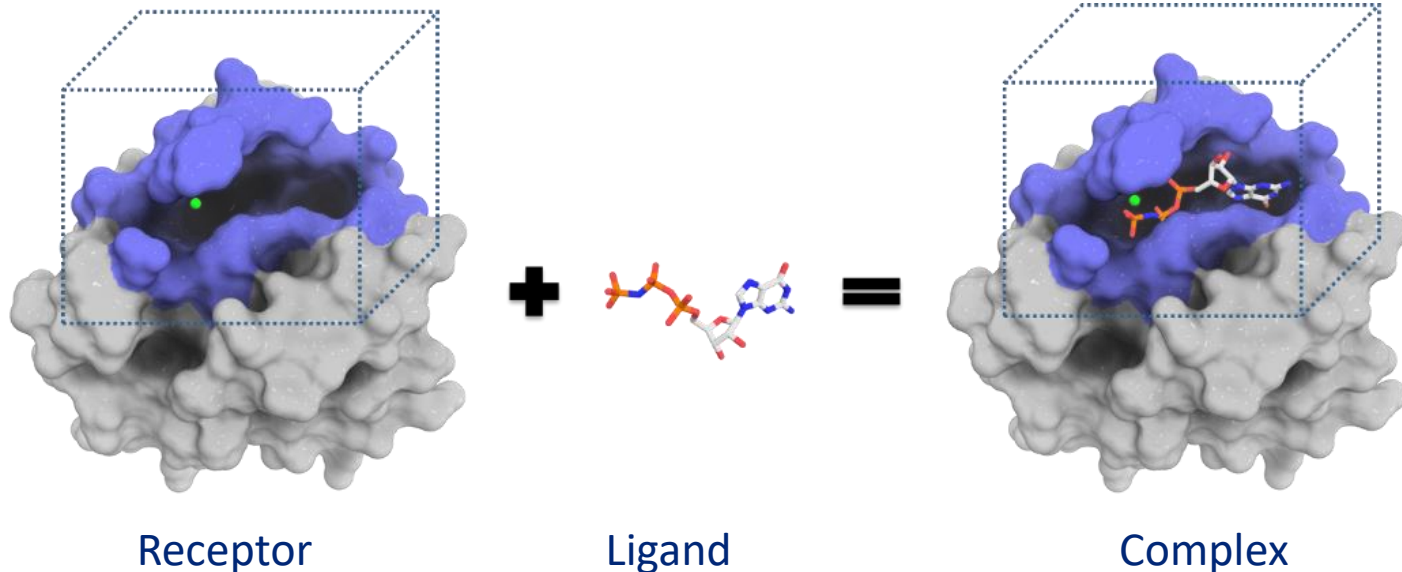


- Crystal (experimental)
- Docking attempts

Molecular docking



- ❑ Several components/steps
 - ❑ Receptor representation
 - ❑ Ligand representation
 - ❑ Search of binding modes
 - ❑ Scoring

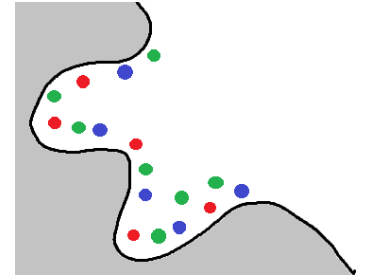


Receptor representation

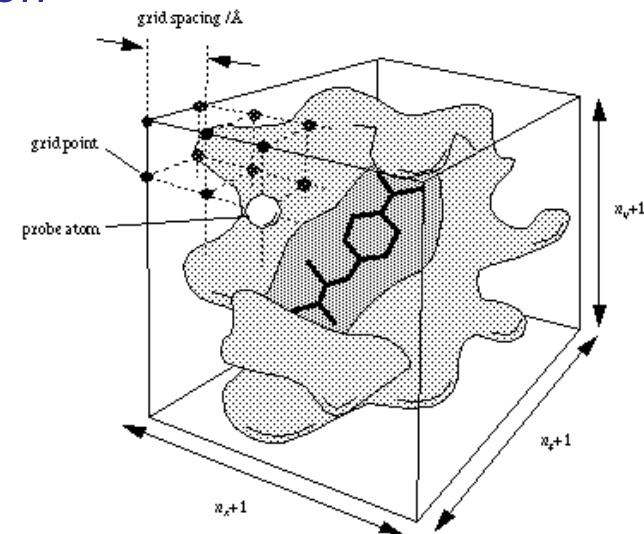


□ Receptor represented only by relevant binding site

- **Descriptor representation** – derived from geometry and interaction abilities of binding site (H-bond donor/acceptor, hydrophobic contacts, ...)



- **Grid representation** – entire searched region is covered by orthogonal equidistant points carrying information about interactions of probe atom at this point with receptor atoms

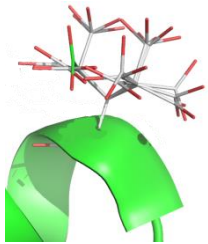


Receptor representation



□ Receptor flexibility

- **Fully rigid approximation**
- **Soft docking** – employs tolerant “soft” scoring functions to simulate plasticity of otherwise rigid receptor
- **Explicit side-chain flexibility** – optimization of residues by rotating part of their structure or rotation of whole side-chains using predefined rotamer libraries
- **Docking to molecular ensemble of protein structure** – obtained from multiple crystal structures, from NMR structure determination or from a trajectory produced by MD simulation





- **Ligands represented by all atoms or just some**
 - Non-polar hydrogens can be united with their respective parent carbon atoms to reduce number of atoms in calculation

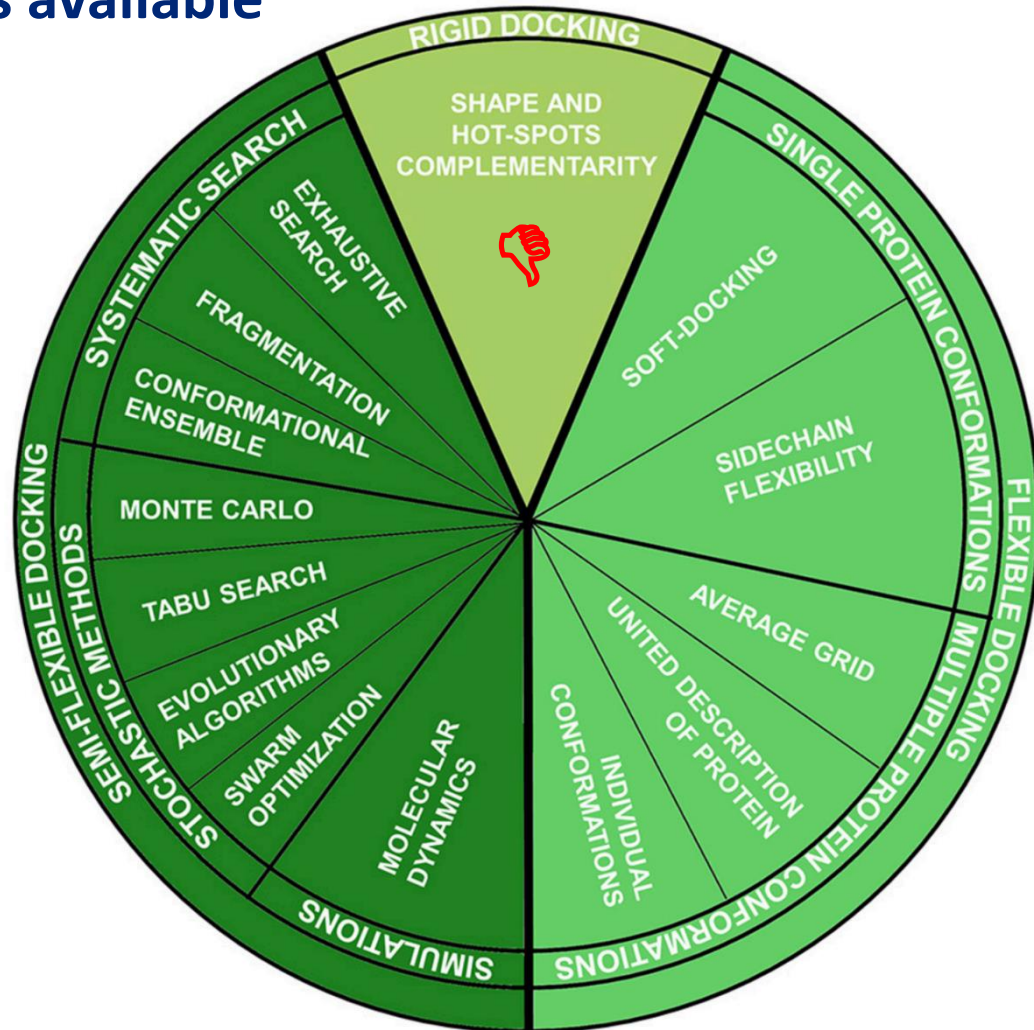
- **Ligand flexibility**
 - Only rotation about single bonds
 - **Docking of a library of pre-generated ligand conformations** – applicable only to quite rigid ligands due to exponential increase in number of possible conformers with number of rotatable bonds
 - **Direct sampling of ligand conformational space during searching**
 - **Fragment-based techniques** – ligand is cut into several fragments and rigidly docked into binding site

Molecular docking – search



□ Many search algorithms available

- Rigid docking 🙄
- **Semi-flexible** 👍
- Fully flexible 👍 👍
(but demanding)





□ **Geometry-based and combinatorial algorithms**

- Assumes that binding is governed by shape and/or physicochemical complementarity between the ligand and the receptor
- **Assumes that the degree of complementarity is proportional to the binding energy** which is not always true especially for more polar ligands

□ **Energy-driven and stochastic algorithms**

- Tries to locate directly the global minimum of the binding free energy corresponding to the experimental structure
- **Random basis** of these methods **requires multiple independent runs** of docking calculations to achieve consistent results

Geometry-based algorithms

□ Matching algorithms

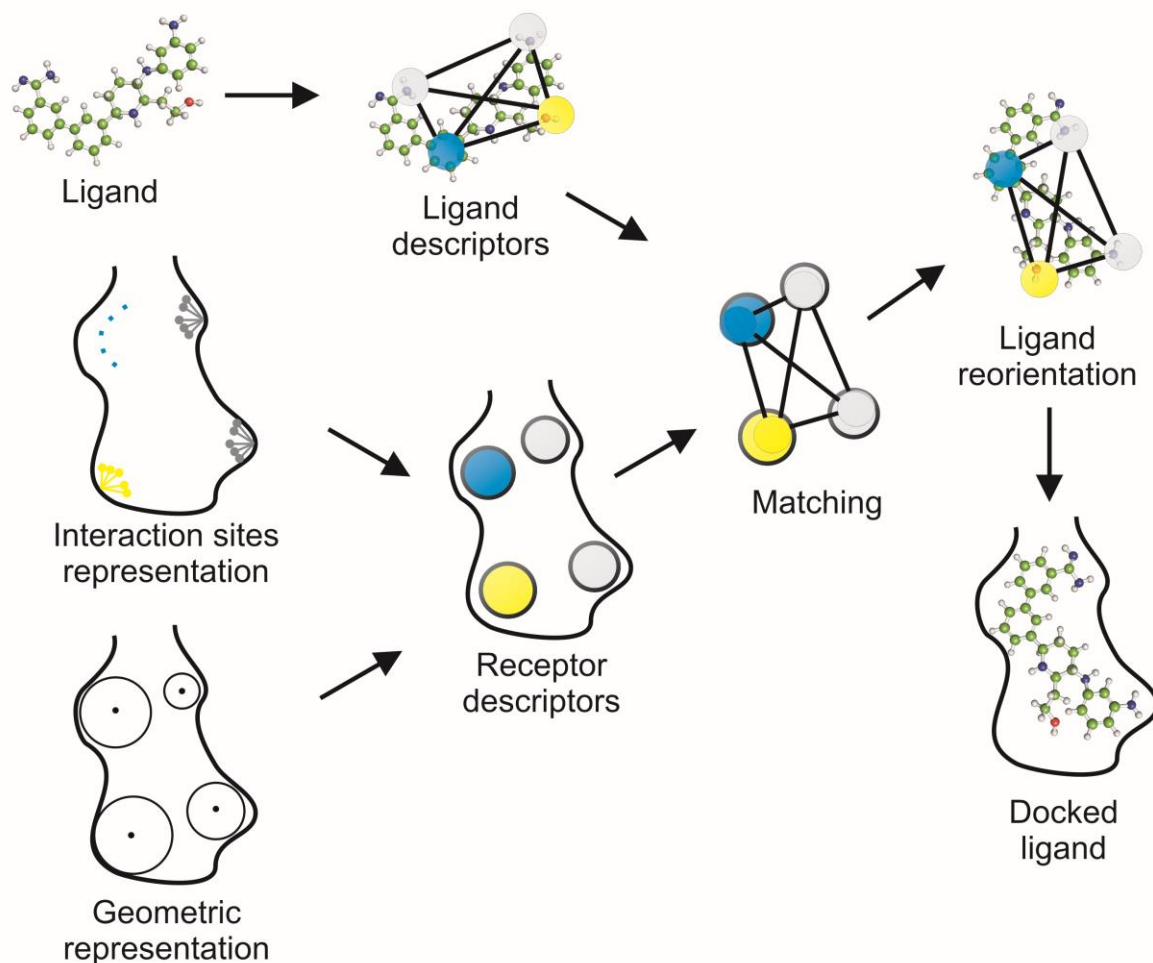
- Represent a ligand and a receptor binding site by **descriptors** derived from their geometry and/or presence of particular interaction sites
- Try to **align/match complementary parts** of ligand and binding site and in this way predict the ligand binding mode

□ SW packages

- DOCK – <http://dock.compbio.ucsf.edu/>
- SLIDE – <http://www.kuhnlab.bmb.msu.edu/software/slide/>
- ...

Geometry-based algorithms

□ Matching algorithms



Geometry-based algorithms

□ **Fragment-based algorithms**

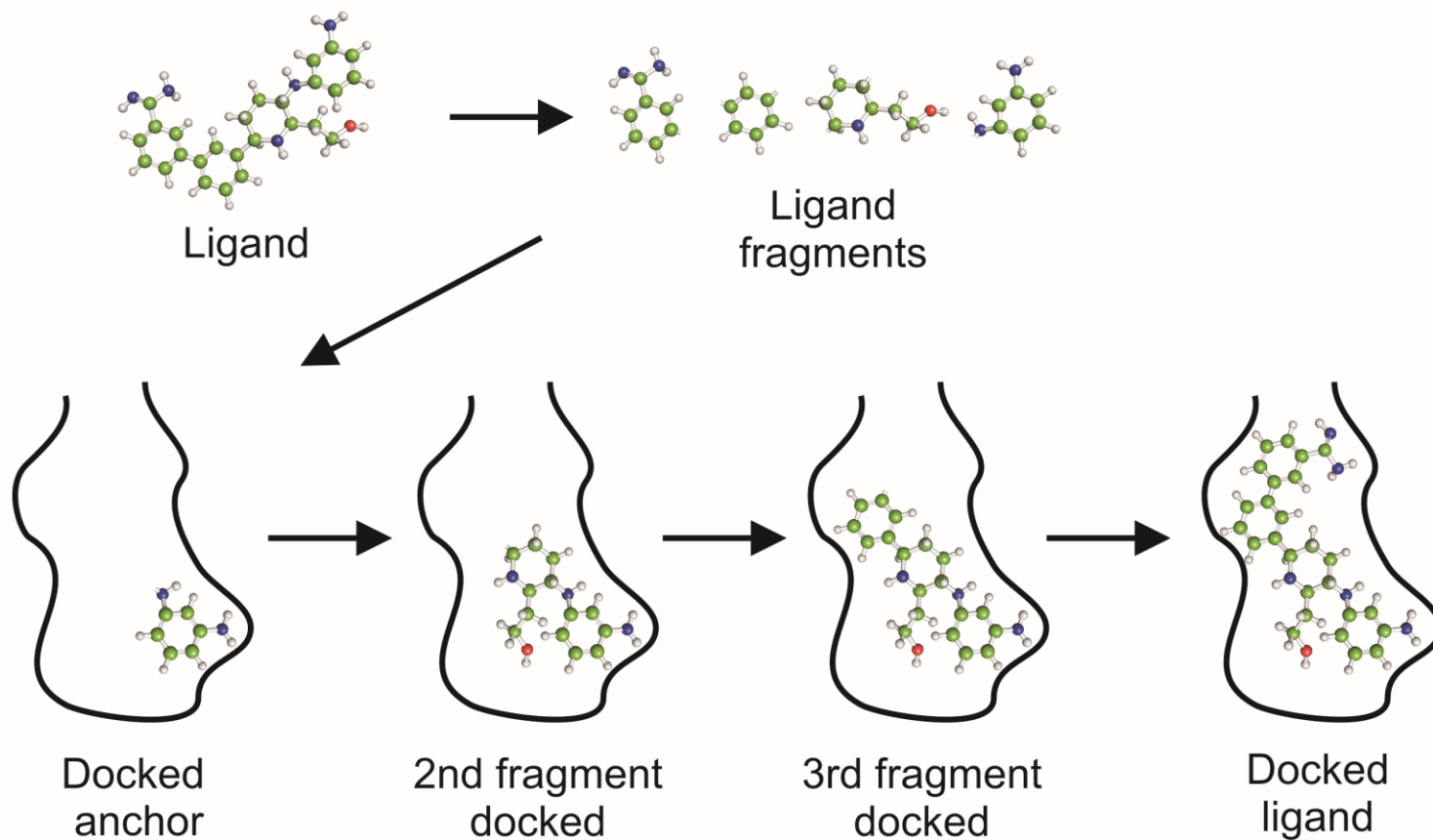
- Ligand is initially fragmented into rigid parts
- Two approaches to obtain whole docked molecule
 - **Incremental construction** – fragments are incrementally docked into the receptor until whole ligand is constructed
 - **Fragment-placing and linking** – all fragments are docked simultaneously and then joined together

□ **SW packages**

- FlexX – <http://www.biosolveit.de/FlexX/>
- eHITS – <http://www.simbiosys.ca/ehits/>
- ...

Geometry-based algorithms

□ Fragment-based algorithms



Stochastic energy-driven algorithms

□ Monte Carlo algorithms

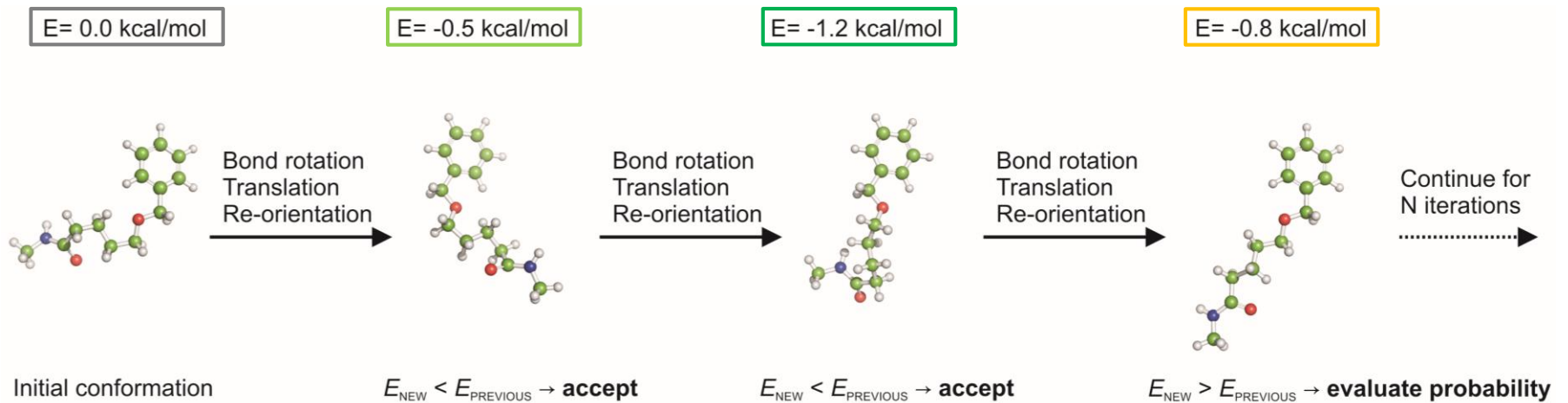
- Explore protein-ligand interactions space by **iteratively introducing random changes** into a position, orientation or conformation of the ligand and evaluating new configuration using **acceptance criterion**
- New configuration is **always accepted** if its energy is more favorable than the energy of previous configuration or **accepted with some probability** reflecting energy difference to previous configuration

□ SW packages

- Autodock Vina – <http://vina.scripps.edu>
- Glide – <http://www.schrodinger.com/Glide>
- ...

Stochastic energy-driven algorithms

□ Monte Carlo algorithms



Stochastic energy-driven algorithms

□ Genetic algorithms

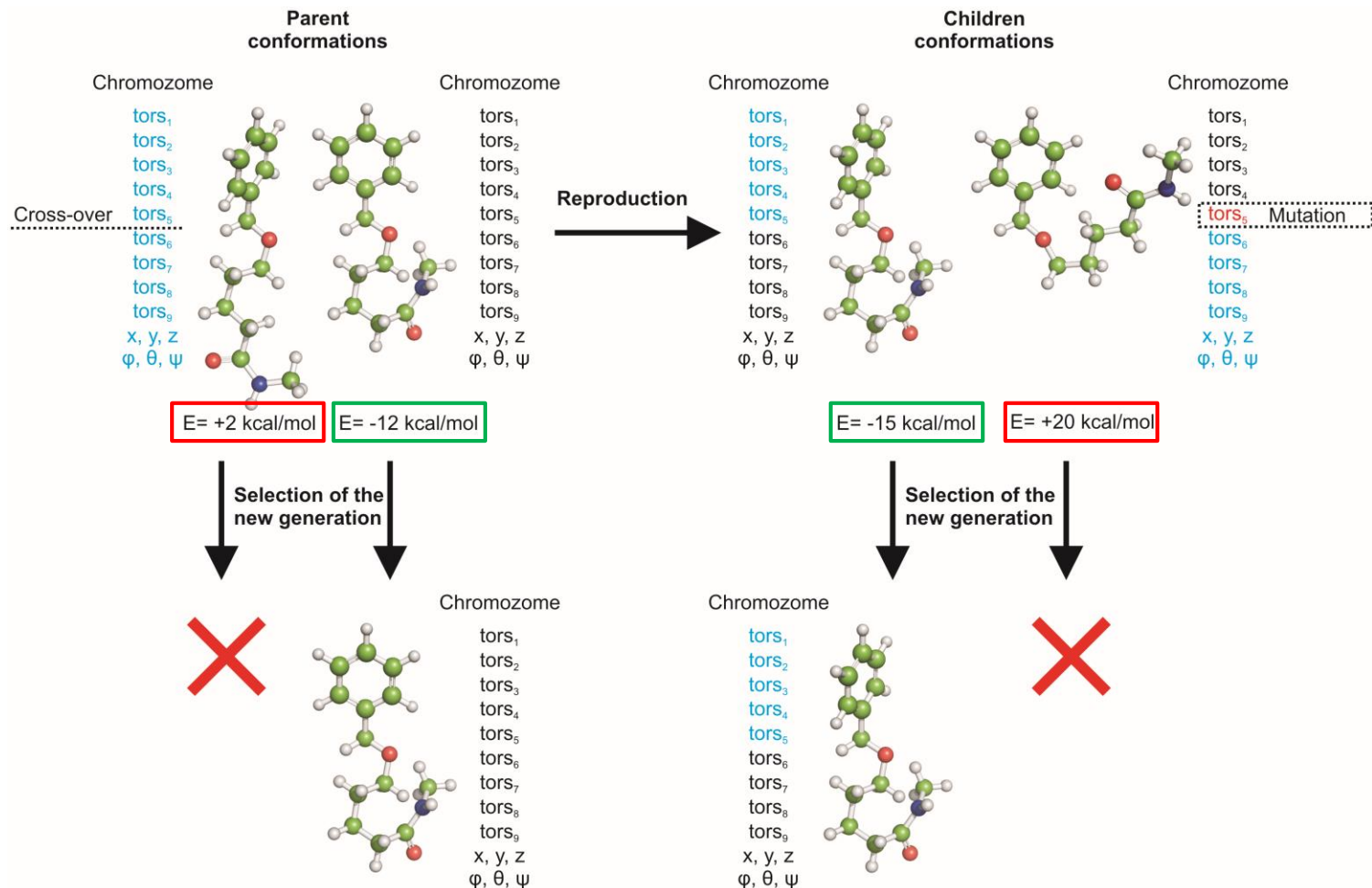
- Configurations of the ligand from randomly generated initial population are encoded in their “genes” which are subject of random genetic modification (single point mutation, crossover, ...)
- Individuals with better fitness (binding energy) have higher chance to survive and reproduce to next generation
- Overall fitness of population is increasing with each new generation

□ SW packages

- AutoDock – <http://autodock.scripps.edu>
- GOLD – http://www.ccdc.cam.ac.uk/products/life_sciences/gold/
- ...

Stochastic energy-driven algorithms

Genetic algorithms





□ Scoring function

- Evaluate all the binding modes from the searching algorithms
- Must be computationally efficient and provide accurate description of protein-ligand interactions

□ Application of scoring functions to rank

- Several configurations of one ligand bound to one protein – essential for prediction of the best binding mode
- Different ligands bound to one protein – determination of substrate or inhibitor specificity
- One ligand bound to several different proteins – functional annotation of proteins and study of drug selectivity



□ Categories of scoring functions

- Empirical
- Knowledge-based
- Force field-based
- Machine learning

Molecular docking – scoring

□ Categories of scoring functions

▪ Empirical

- Derived by fitting of following equation to experimental binding affinities of known protein-ligand complexes

$$\Delta G_{bind} = \alpha.\Delta G_{hb} + \beta.\Delta G_{lipo} + \gamma.\Delta G_{el} + \delta.\Delta G_{rot} + \dots$$

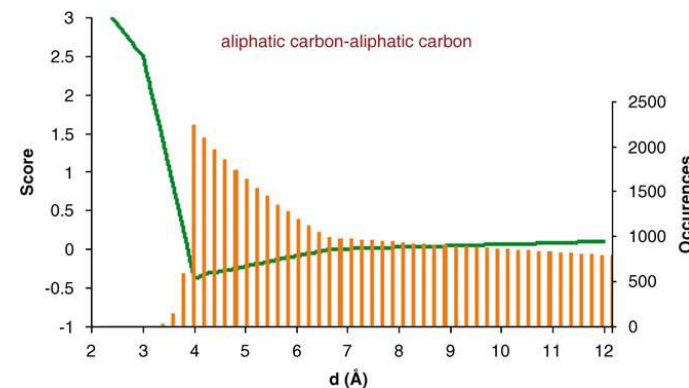
- Rapid evaluations
- Arbitrary selection of terms included in the equation → failure when binding is governed by any excluded type of interaction
- Weights are dependent on the chosen training set

Molecular docking – scoring

□ Categories of scoring function

▪ Knowledge-based

- Capture the **knowledge** about protein-ligand binding that is implicitly **stored in structural data** by **statistical analysis**
- **Atom-pair potentials** derived from distances found for such pair in training structural data
- **Rapid evaluations**
- Describe **all types of interactions** without any preselection
- **Problem** when structural data **do not** contain **sufficient information** on specific atom-pairs (ex. halogens, metals, ...)



Molecular docking – scoring

□ Categories of scoring function

▪ Force field-based

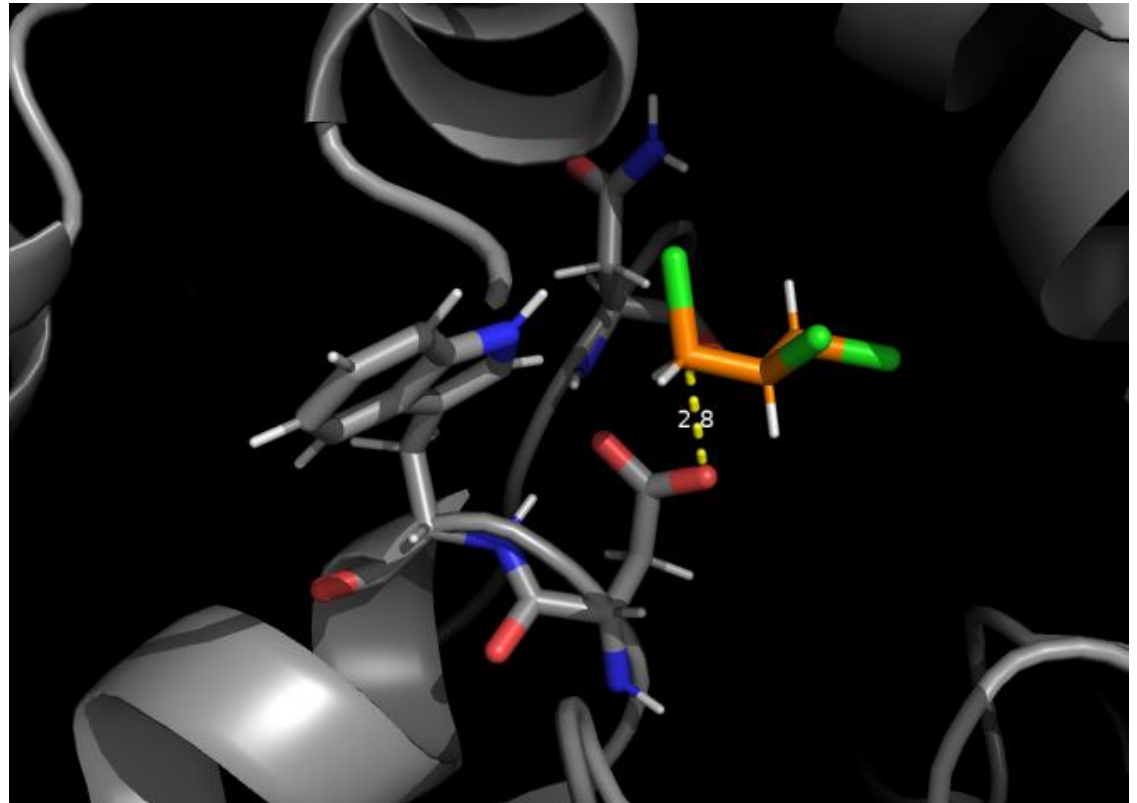
- Use the non-bonded terms of **well-established force fields**
- Provide **precise affinities**
- **Computationally demanding** → employed for rescoring selected binding modes (**not during searching**)

$$E_{total} = \sum_{\text{bonds}} K_r (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2$$
$$+ \sum_{\text{torsions}} \frac{V_n}{2} [1 \pm \cos(n\phi - \gamma)]$$
$$+ \sum_{\text{non-bonded}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}} \right]$$

Evaluation of complexes



- ❑ Intermolecular interactions
- ❑ Binding energies



Intermolecular interactions



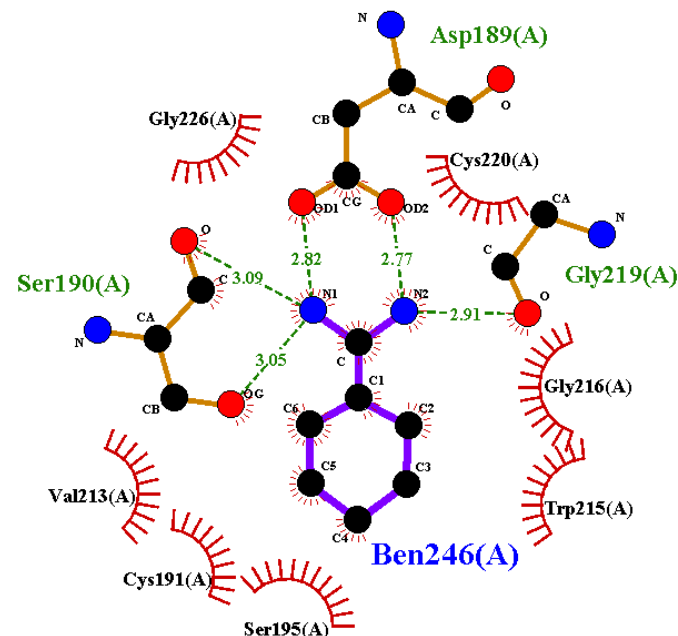
□ **Most common types**

- Hydrogen bonds
- Hydrophobic
- Aromatic
- Ionic bonds

Intermolecular interactions

□ Visualization

- Schematic diagrams showing hydrogen bonds and hydrophobic contacts



□ Tools

- LigPlot⁺
 - Stand alone application
 - <http://www.ebi.ac.uk/thornton-srv/software/LigPlus/>
 - Pre-calculated for protein-ligand complexes in PDBsum (pictorial database of PDB structures)

Binding energies

□ Binding Affinity Prediction of Protein-Ligand (BAPPL) server

- <http://www.scfbio-iitd.res.in/software/drugdesign/bappl.jsp>

- Calculates binding free energy

of a protein-ligand complex

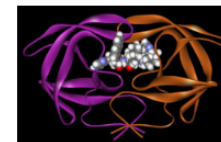
using all-atom-energy-based

empirical scoring function

- Only for non-metallo protein-ligand complexes



BAPPL server



Welcome to the BAPPL server

Binding Affinity Prediction of Protein-Ligand (BAPPL) server computes the binding free energy of a non-metallo protein-ligand complex using an all atom energy based empirical scoring function [1] & [2].

BAPPL server provides two methods as options:

Method 1 : Input should be an energy minimized protein-ligand complex with hydrogens added, protonation states, partial atomic charges and van der Waals parameters (R^m and S) assigned for each atom. The server directly computes the binding affinity of the complex using the assigned parameters. For format specifications on the input, please refer to the [README](#) file.

Method 2 : Input should be an energy minimized protein-ligand complex with hydrogens added and protonation states assigned. The net charge on the ligand should be specified. The server derives the partial atomic charges of the ligand using the AM1-BCC procedure [3] and GAFF [5] force field for van der Waals parameters. Cornell et al. force field [4] is used to assign partial atomic charges and van der Waals parameters for the proteins. For format specifications on the input, please refer to the [README](#) file.

For the purpose of validation of the empirical scoring function [1] a dataset of 161 non-metallo protein-ligand complexes has been prepared. [Click here](#) to access the Protein-Ligand Complex Dataset.

Transport of small molecules



- ❑ Describe trajectory of ligands through tunnels
- ❑ Based on geometry or molecular docking
 - Fast but low accuracy
 - Good for screening purposes
 - CaverDock, MoMA-LigPath, SLITHER
- ❑ Based on force field
 - Run multiple MD simulations
 - Accurate but computationally demanding
 - Metadynamics, steered MD, adaptive sampling, etc.

Transport of small molecules



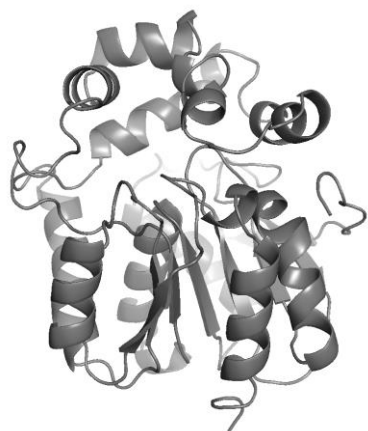
□ CaverDock

- <https://loschmidt.chemi.muni.cz/caverdock/>
- Analysis of tunnels by Caver
- Discretization of identified tunnel into discs
- Molecular docking by AutoDock Vina to every disc

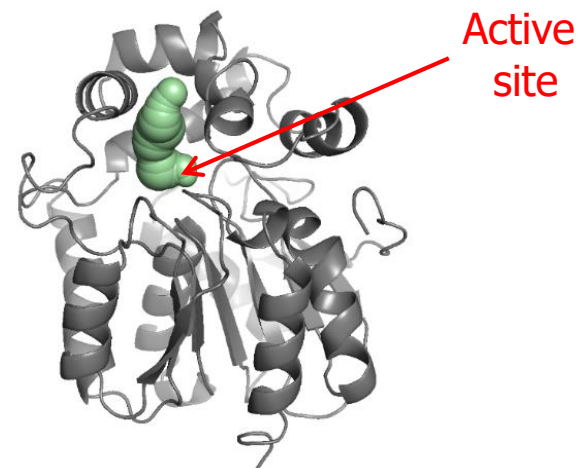
□ Caver Web

- <https://loschmidt.chemi.muni.cz/caverweb/>
- Web interface for Caver and CaverDock

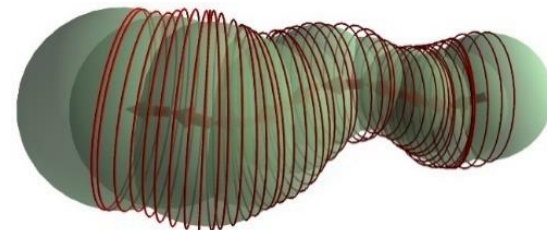
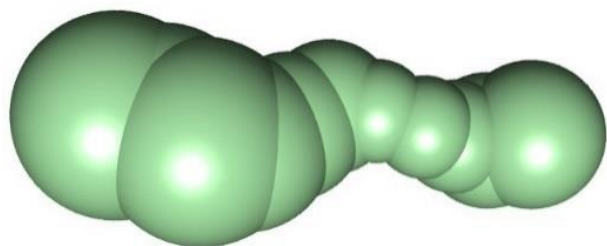
CaverDock



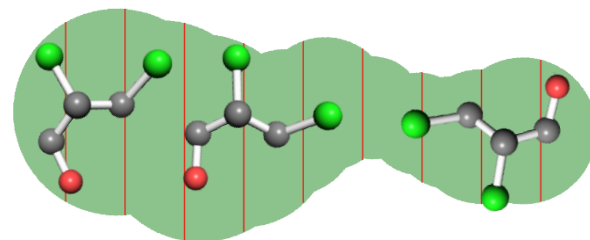
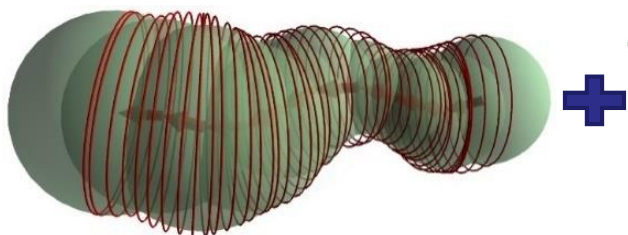
CAVER



Discretization

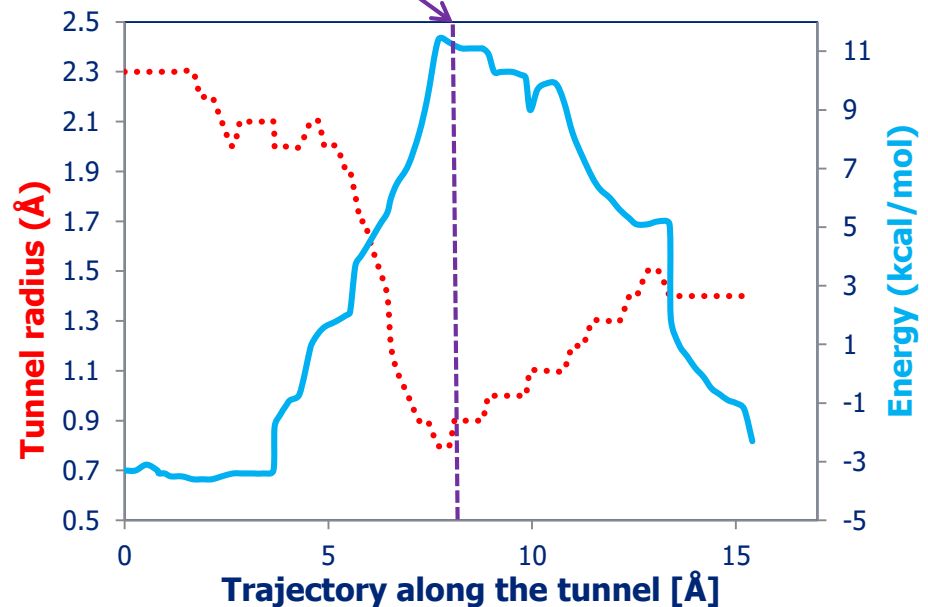
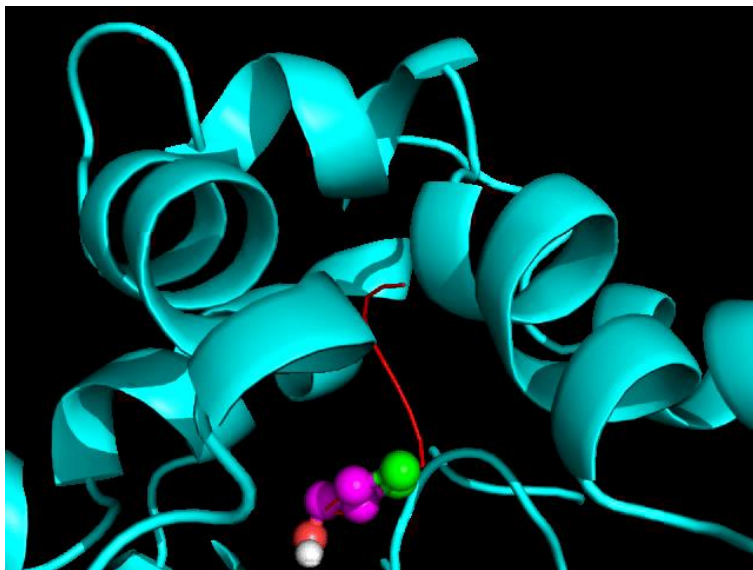
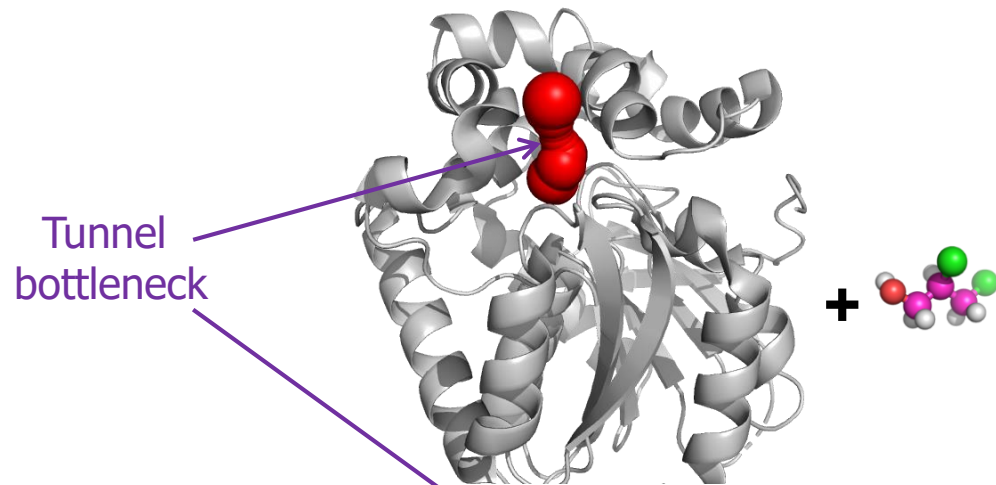


CaverDock



CaverDock

- Results provided:
 - Ligand trajectory
 - Energy profile



CaverDock over Caver Web

Single structure

Tunnels info

| id | bottleneck radius [Å] | length [Å] | curvature | throughput | |
|-------------------------------------|-----------------------|------------|-----------|------------|------|
| <input checked="" type="checkbox"/> | 1 | 1.9 | 10.3 | 1.4 | 0.80 |
| <input checked="" type="checkbox"/> | 2 | 1.8 | 11.2 | 1.2 | 0.78 |
| <input checked="" type="checkbox"/> | 3 | 1.8 | 23.8 | 1.3 | 0.66 |
| <input type="checkbox"/> | 4 | 1.2 | 16.7 | 1.2 | 0.63 |
| <input type="checkbox"/> | 5 | 1.8 | 27.4 | 1.3 | 0.62 |
| <input type="checkbox"/> | 6 | 1.1 | 19.0 | 1.4 | 0.45 |

Return to Results browser

Job information

Job ID: rokj0h
Title: Untitled
Structure: 4NY4

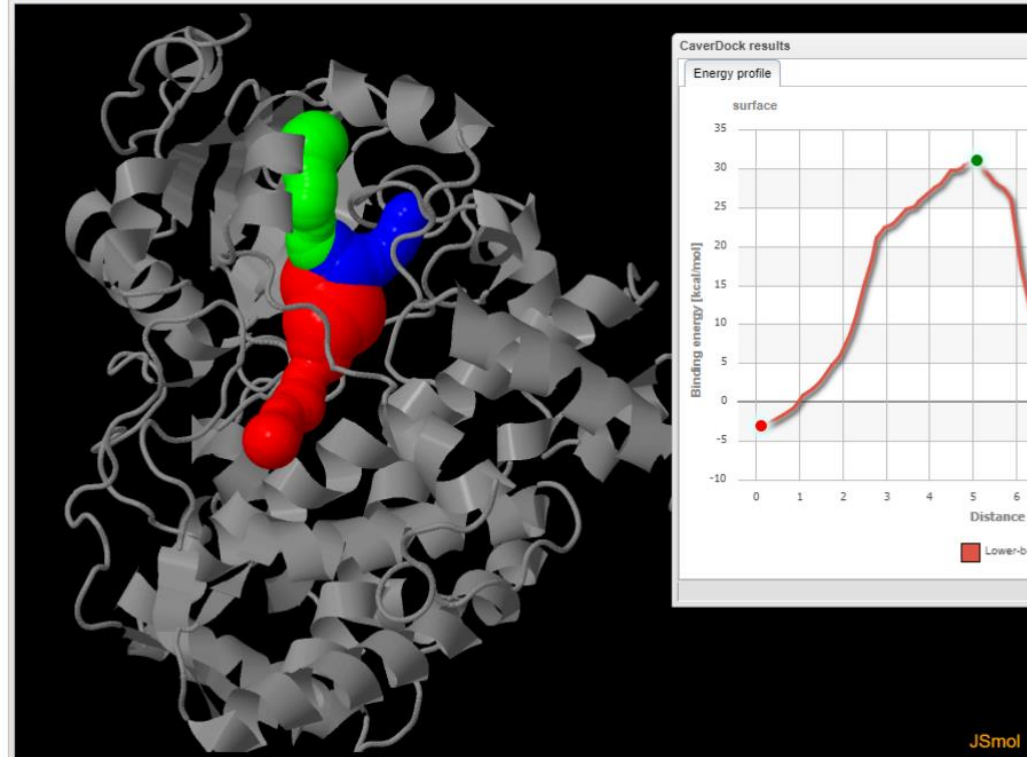
Download PyMOL session

Download results in single zip

View CAVER configuration

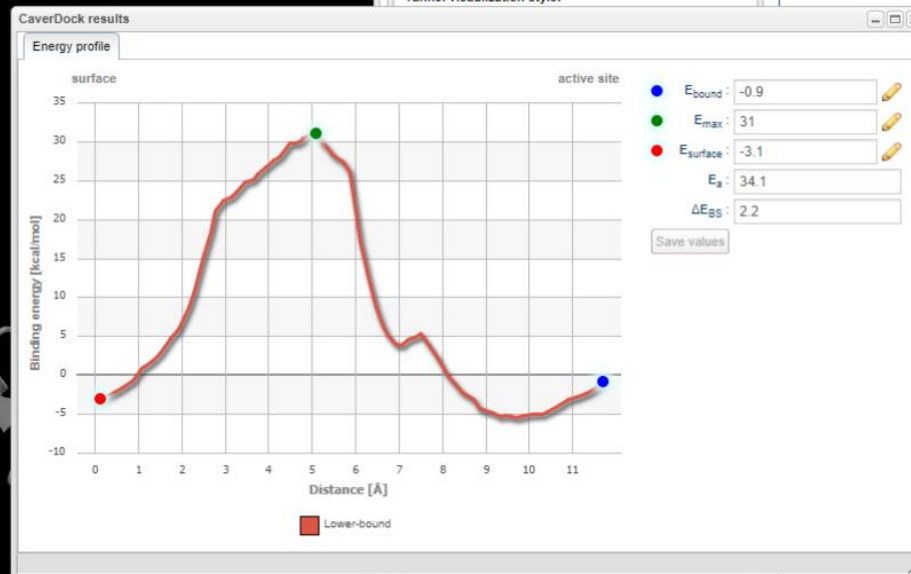
View CAVER log

Viewer



Visualization settings

Tunnel visualization style:



Reset view

Visualization quality:



References I

- ❑ Gu, J. & Bourne, P. E. (2009). **Structural Bioinformatics, 2nd Edition**, Wiley-Blackwell, Hoboken.
- ❑ Pérot, S. *et al.* (2010). Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discovery Today* **15**: 656-667.
- ❑ Moitessier, N. *et al.* (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British Journal of Pharmacology* **153**: S7-S26.

References II

- ❑ Bolton, E. E. *et al.* (2008). PubChem: Integrated platform of small molecules and biological activities. *Annual Reports in Computational Chemistry* **4**: 217-241.
- ❑ Gaulton, A. *et al.* (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* **40**: D1100-D1107.
- ❑ Irwin, J. J. *et al.* (2012). ZINC: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling* **52**: 1757-1768.
- ❑ Santos, R. *et al.* (2017). A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery*. **16**: 19-34