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

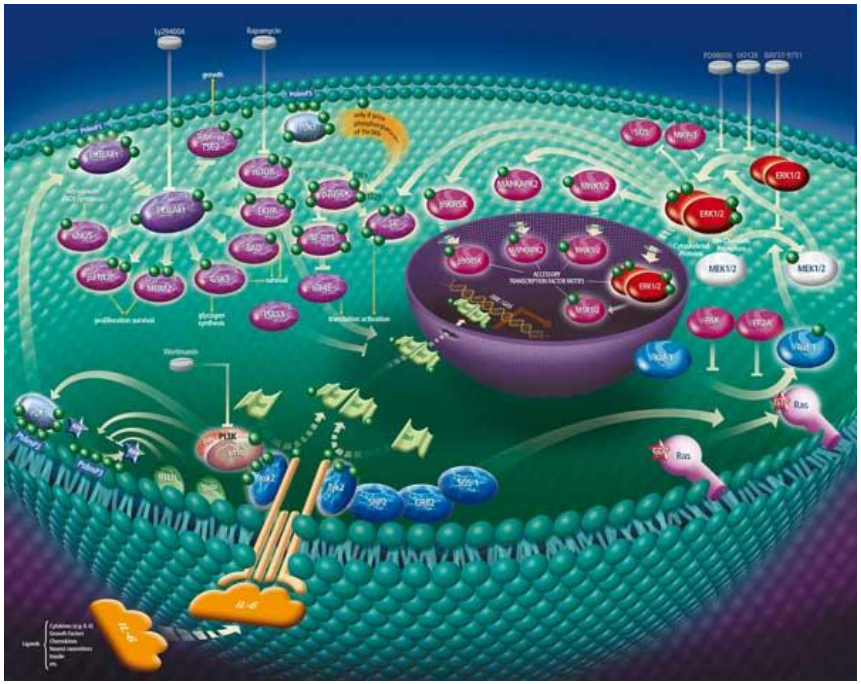
Protein-ligand complexes

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
  - …

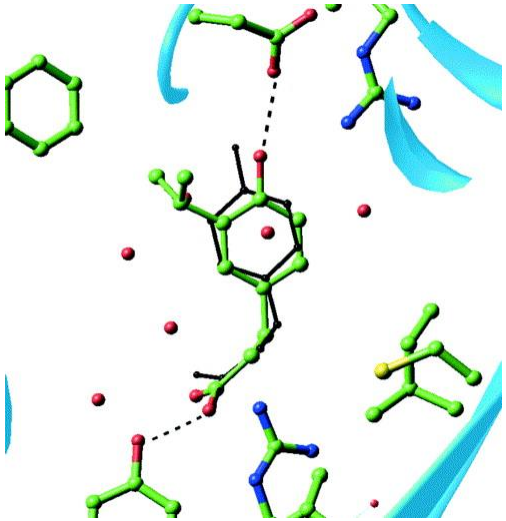


Biological relevance

- **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)
    - …

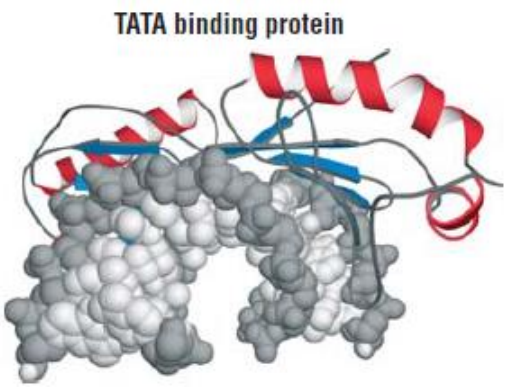
Biological relevance

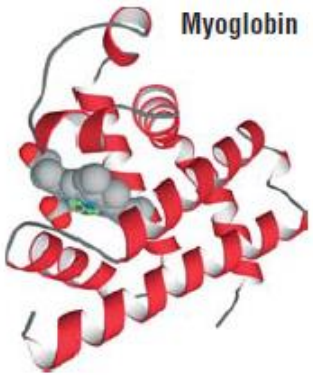
- **Drug discovery**
  - Binding of small molecules to macromolecules
    - Identification of targets (enzymes, receptors, ...)
    - Identification of potential target inhibitors/activators
    - Optimization of target modulators
    - Repurposing of drugs – finding new receptors
    - Adverse side-effects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

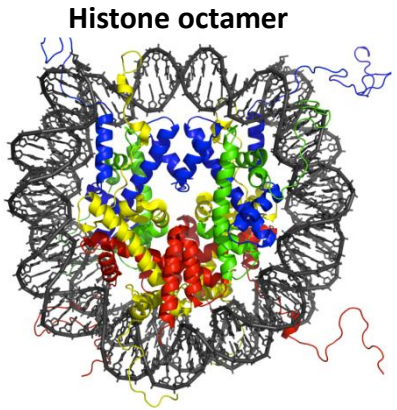


Biophysical aspects

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



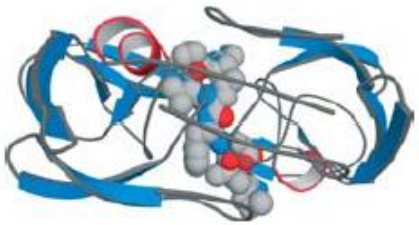




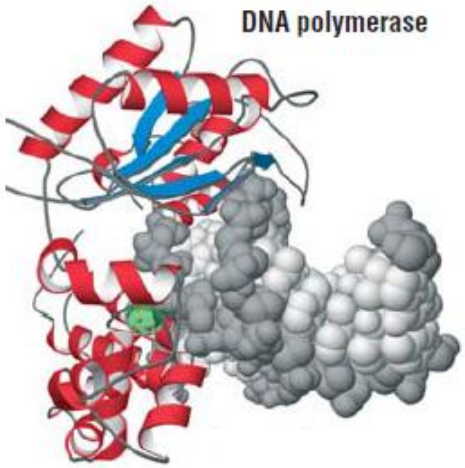
Biophysical aspects

- **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

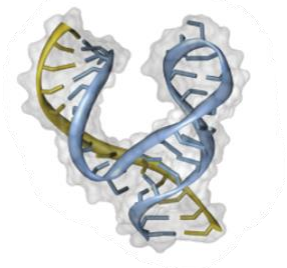
HIV protease



DNA polymerase

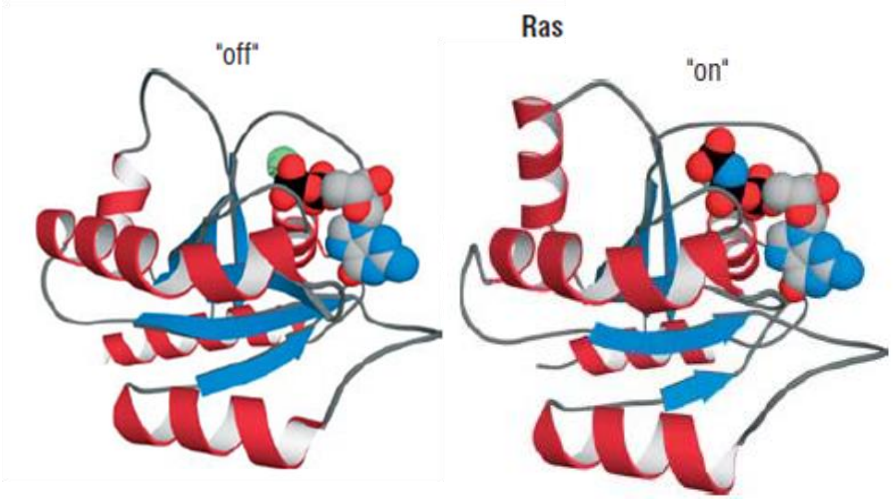


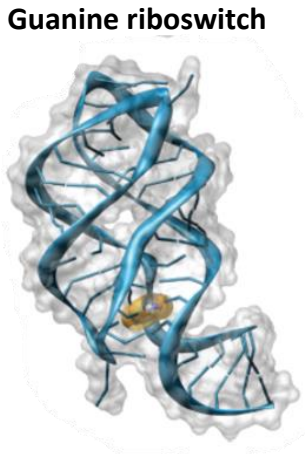
Hammerhead ribozyme



Biophysical aspects

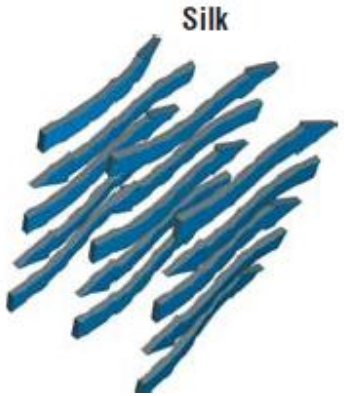
- **Signaling**
  - Conformational changes in response to
    - Ligand binding
    - Properties of surrounding environment (pH, forces…)
  - Different conformations recognized by different proteins in signaling pathways → control of cellular processes

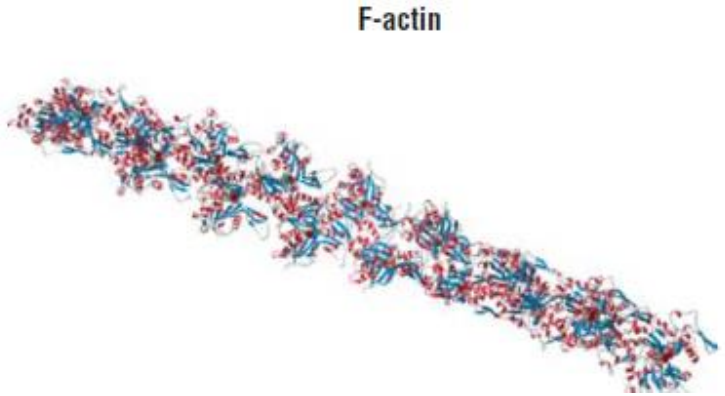




Biophysical aspects

- **Formation of complex structures**
  - Structural elements of complex systems
  - Governed by specific association of protein subunits
    - With themselves
    - Other proteins, carbohydrates, lipids, …



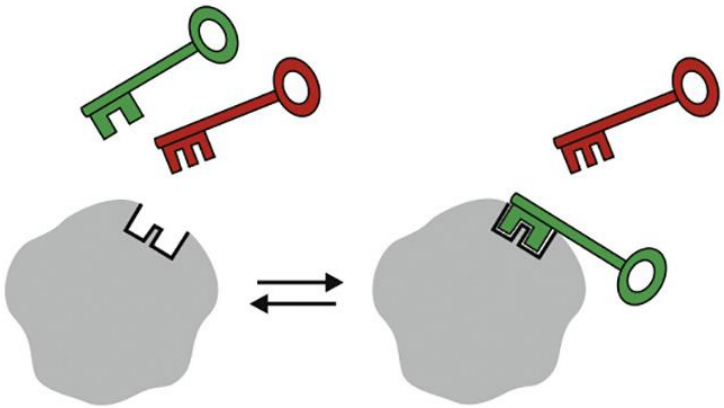


Molecular recognition

- 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

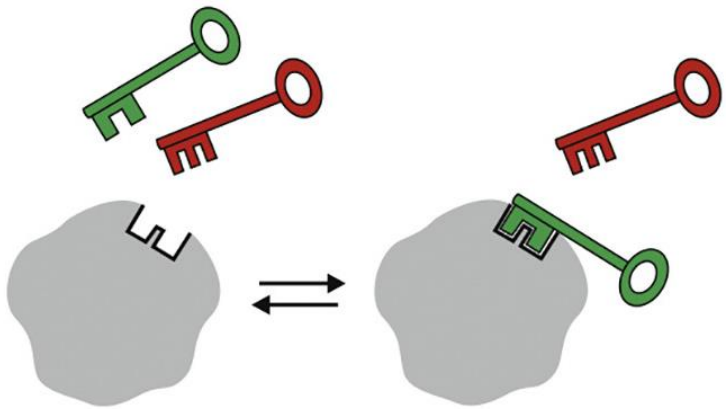
Lock-and-key model

- **E. Fisher – 1894**



Lock-and-key model

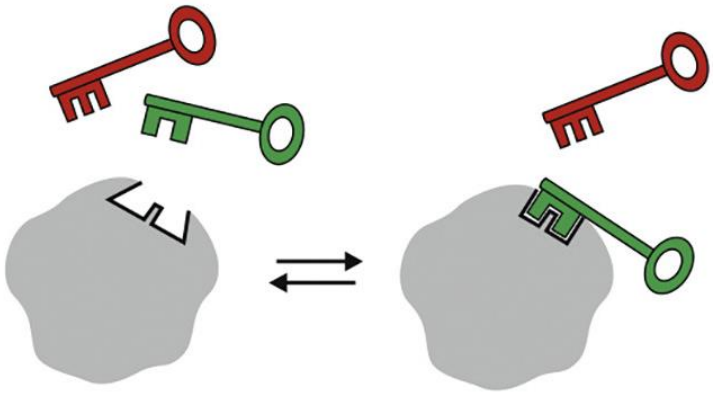
- **E. Fisher – 1894**
- Complementarity between receptor's binding site and the ligand
  - Size & shape
  - Physicochemical properties



- Both ligand and receptor are considered rigid
  - Not sufficient to explain allostery, non-competitive inhibition, or catalysis
  - → Model dismissed, only used for educational purposes

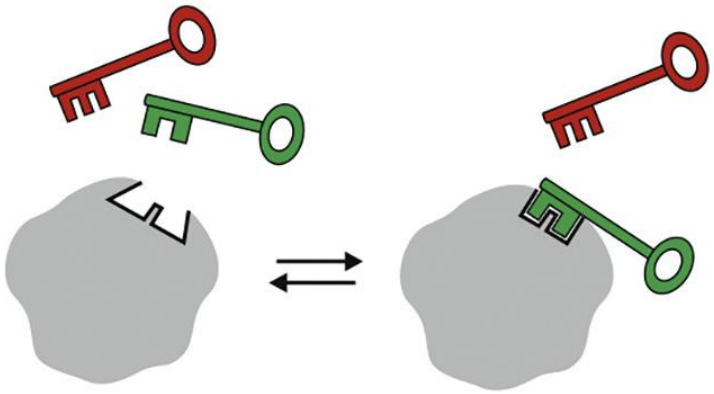
Induced-fit model

- **D. E. Koshland – 1956**



Induced-fit model

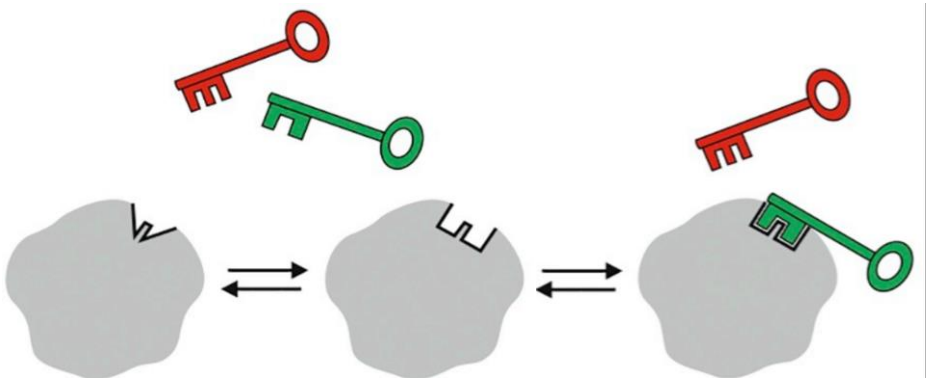
- **D. E. Koshland – 1956**
- Only partial complementarity necessary



- 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

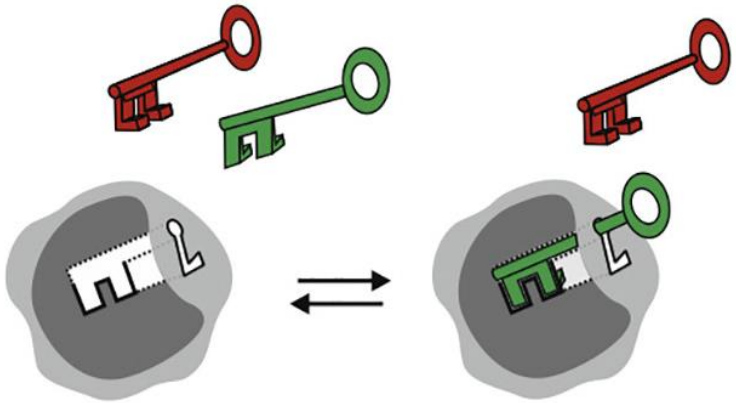
- **B. F. Straub – 1964**
  - This model is also called: *conformational selection*, *fluctuation-fit* or *population selection*
- Receptor and ligand flexible → considered as ensembles



- 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

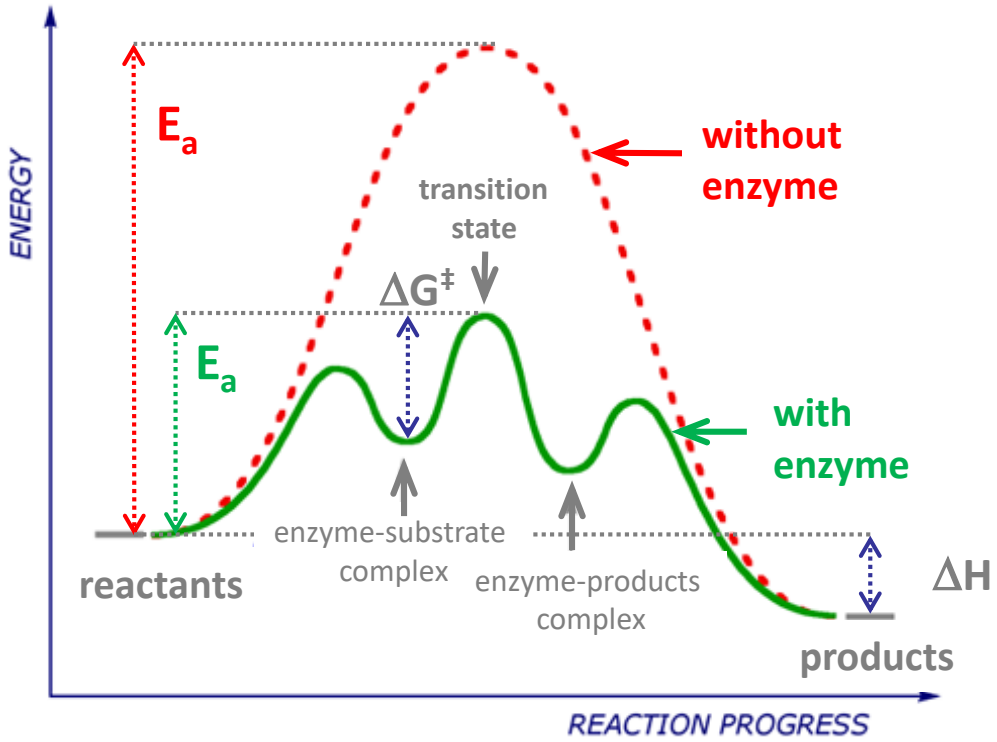
- **Z. Prokop – 2012**
- When the receptor has a buried active site and tunnels



- 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:

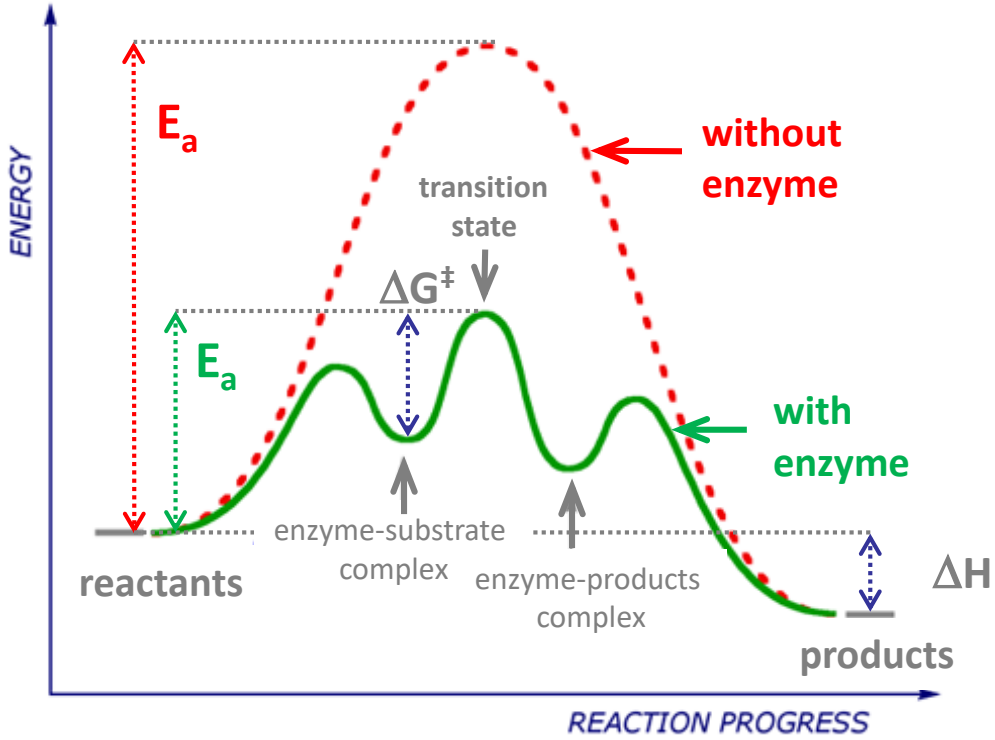
$$k = Ae^{\frac{-E_a}{RT}}$$

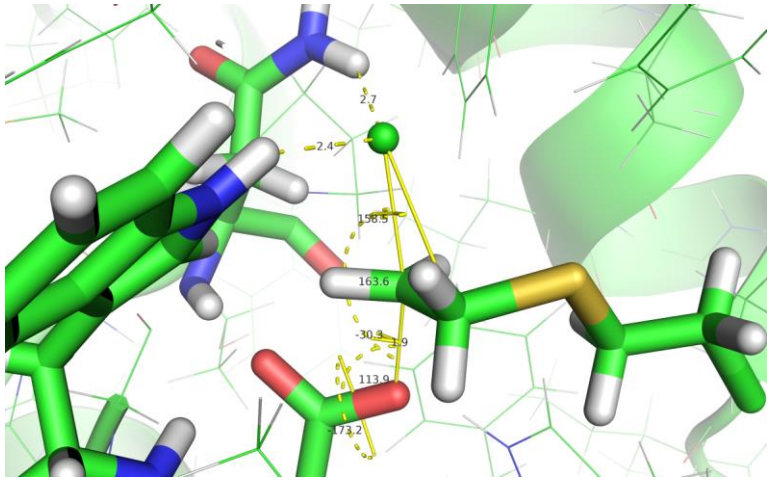
(Arrhenius equation)

- Lower E_a → 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)



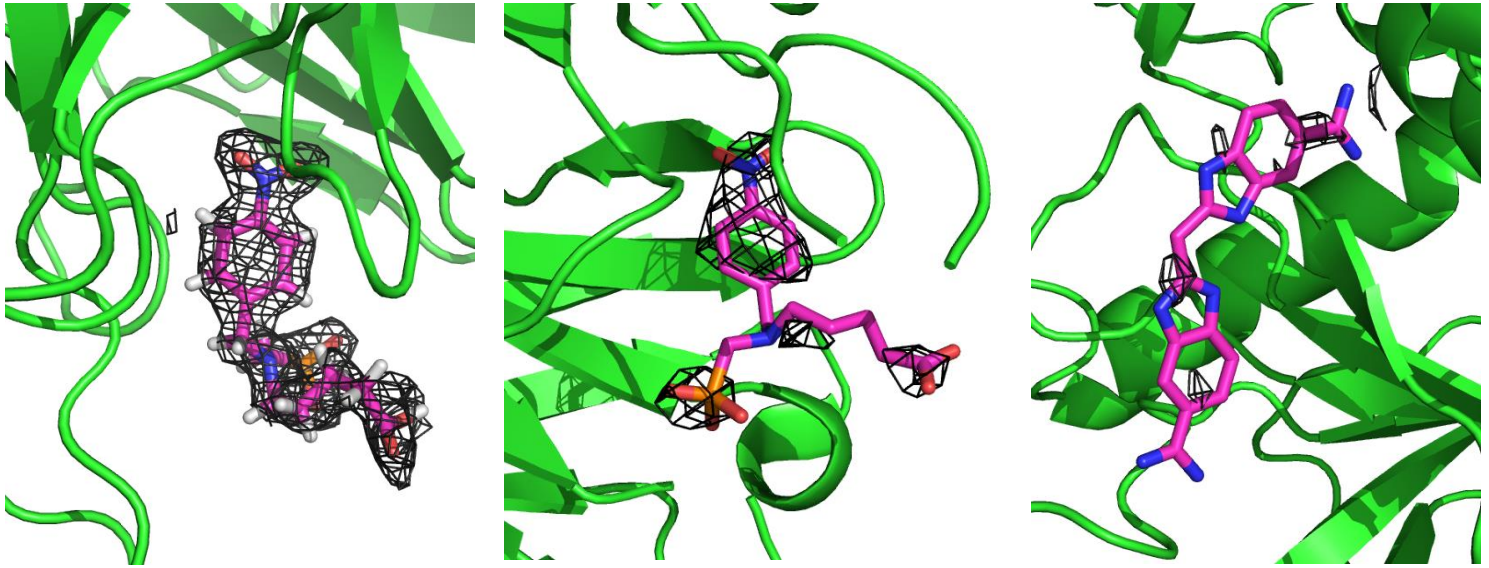


Structure of complexes

- 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

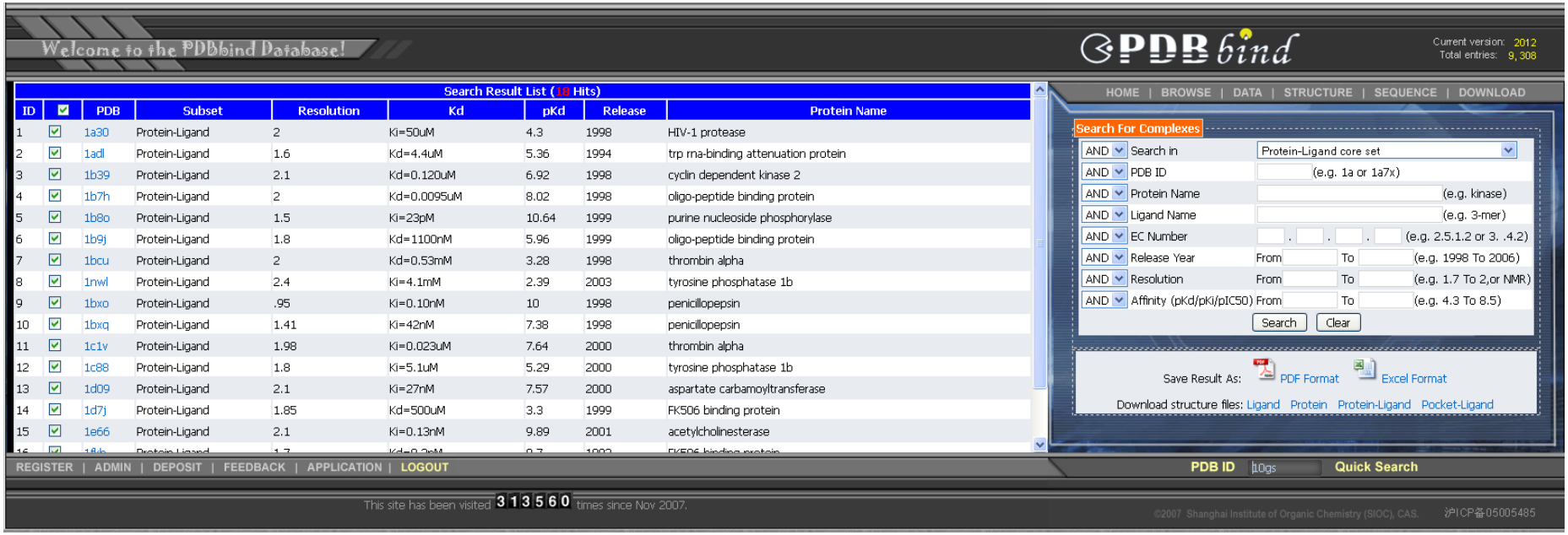


Databases of complexes

- **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

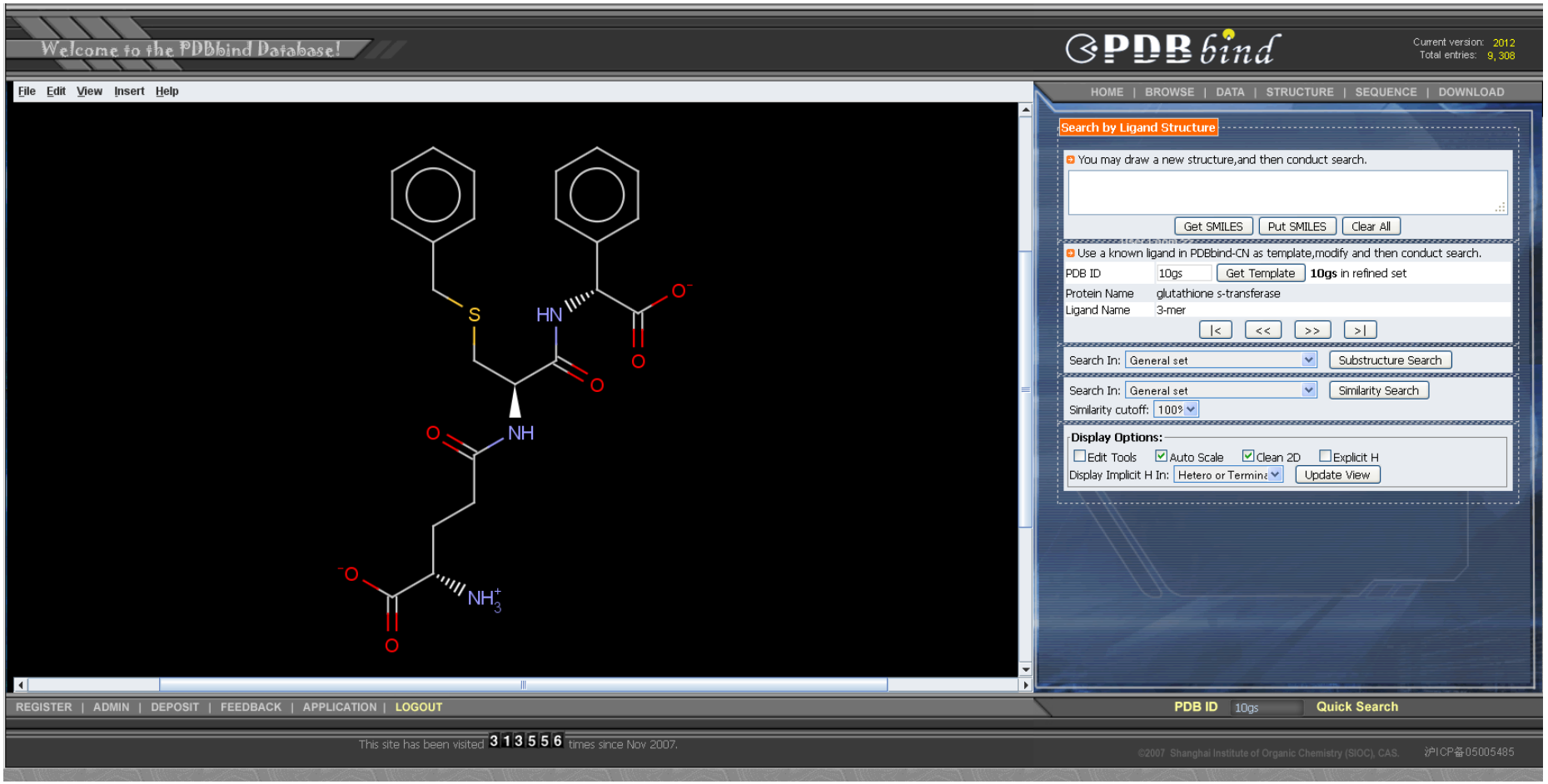
Databases of complexes

- **PDBbind**



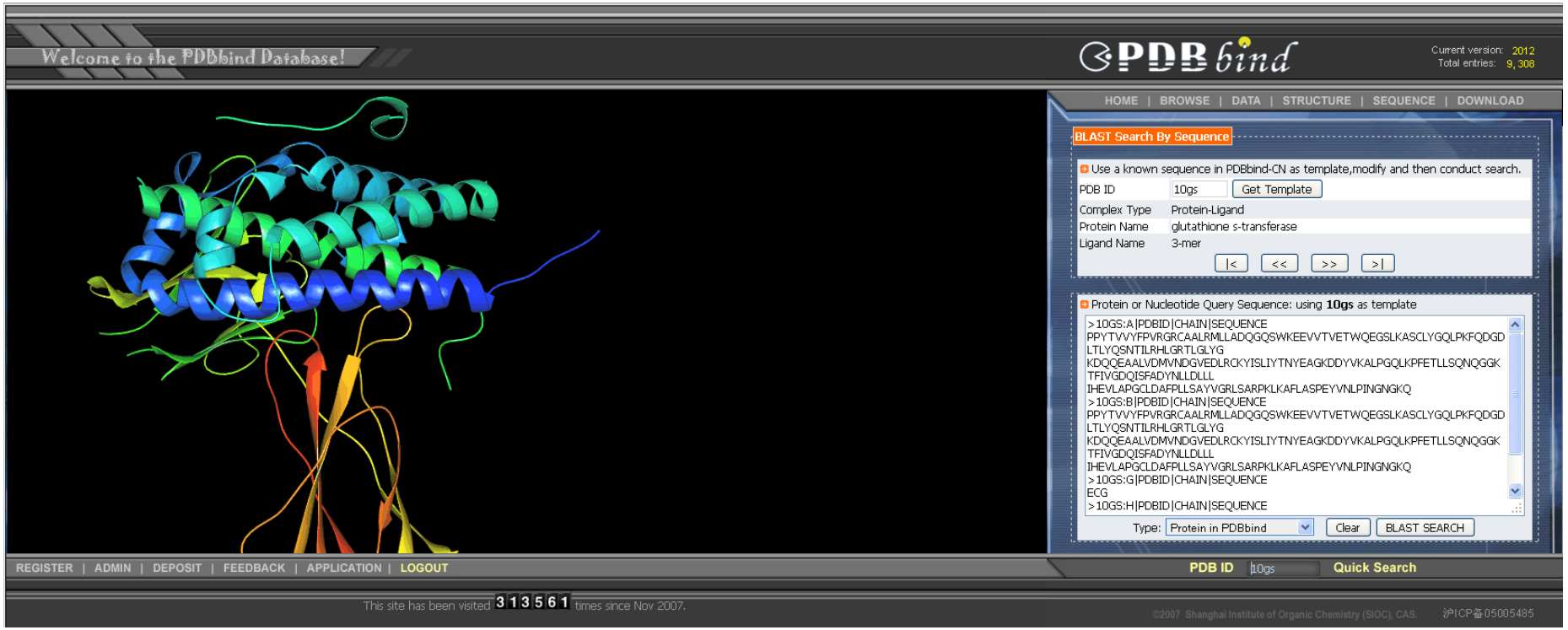
Databases of complexes

- **PDBbind**



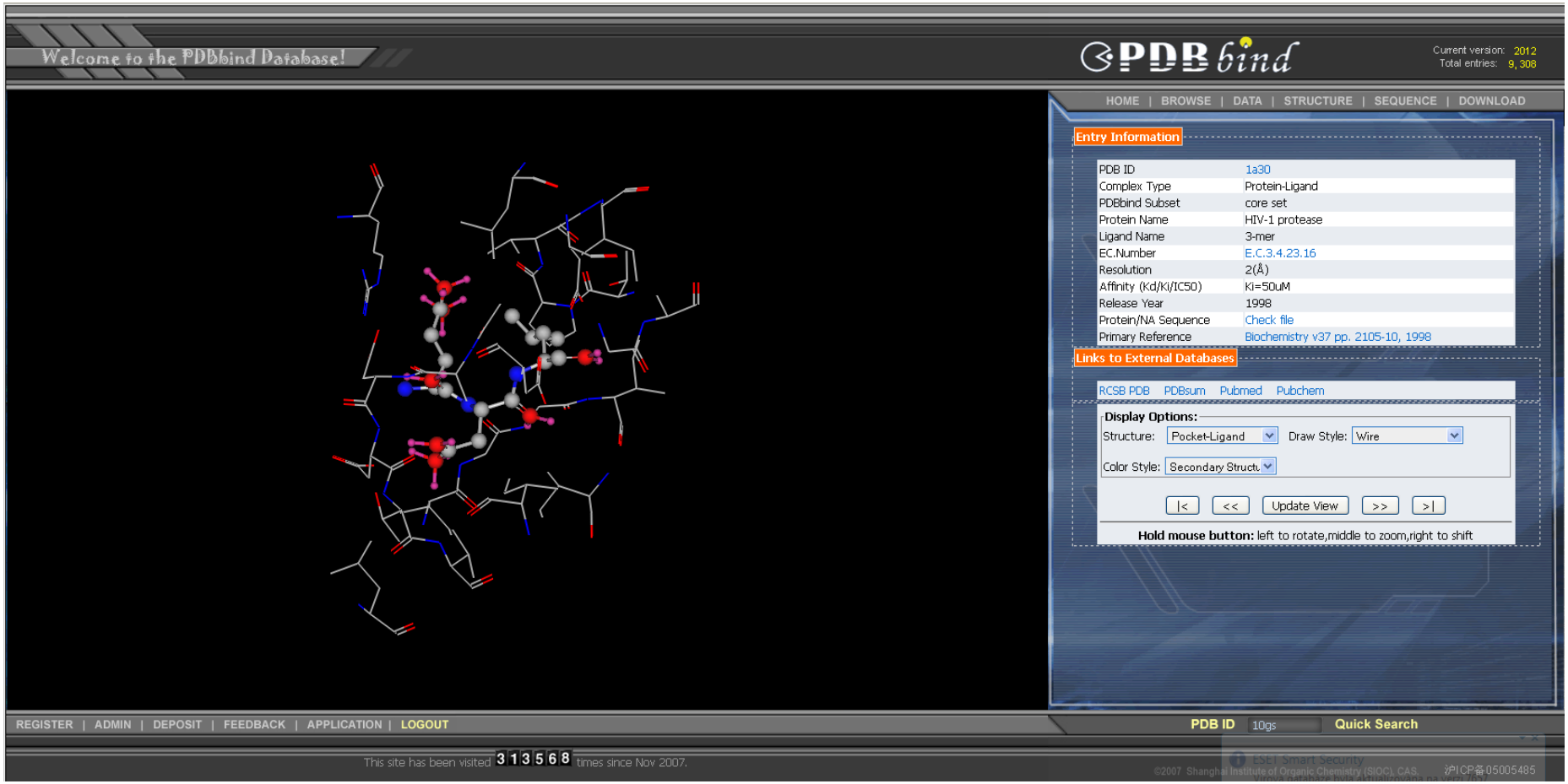
Databases of complexes

- **PDBbind**



Databases of complexes

- **PDBbind**

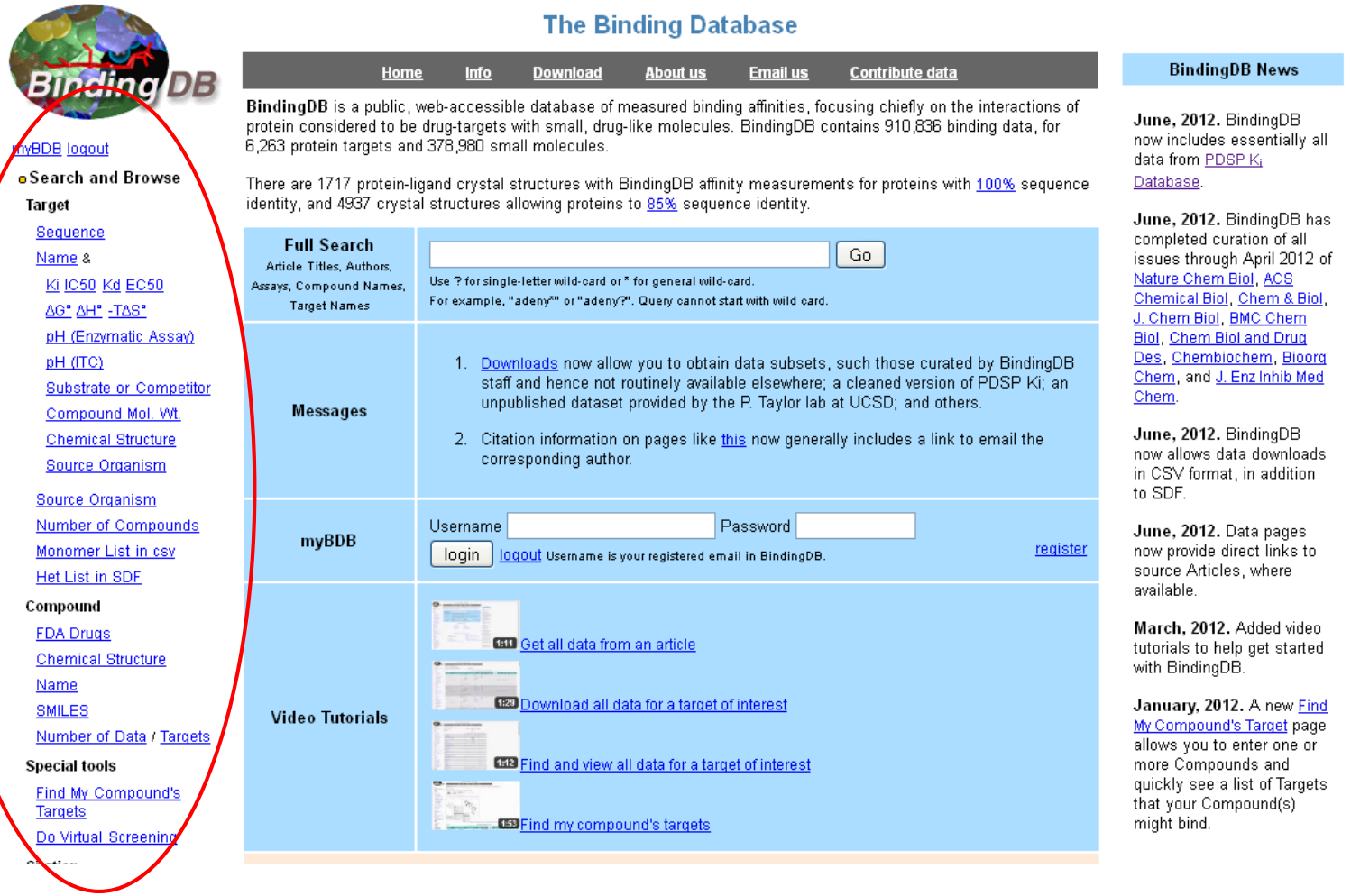


Databases of complexes

- **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

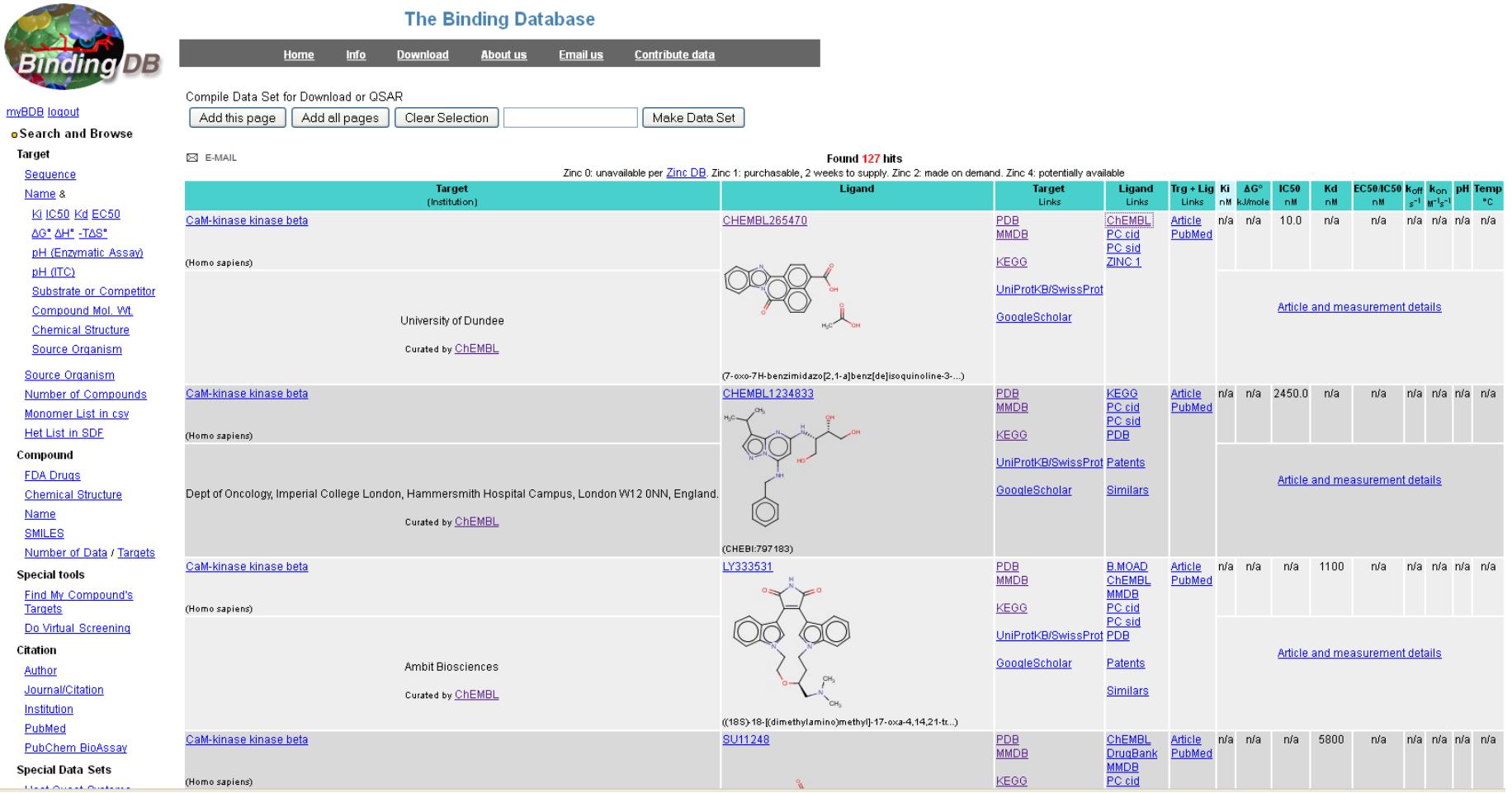
Databases of complexes

BindingDB



Databases of complexes

BindingDB

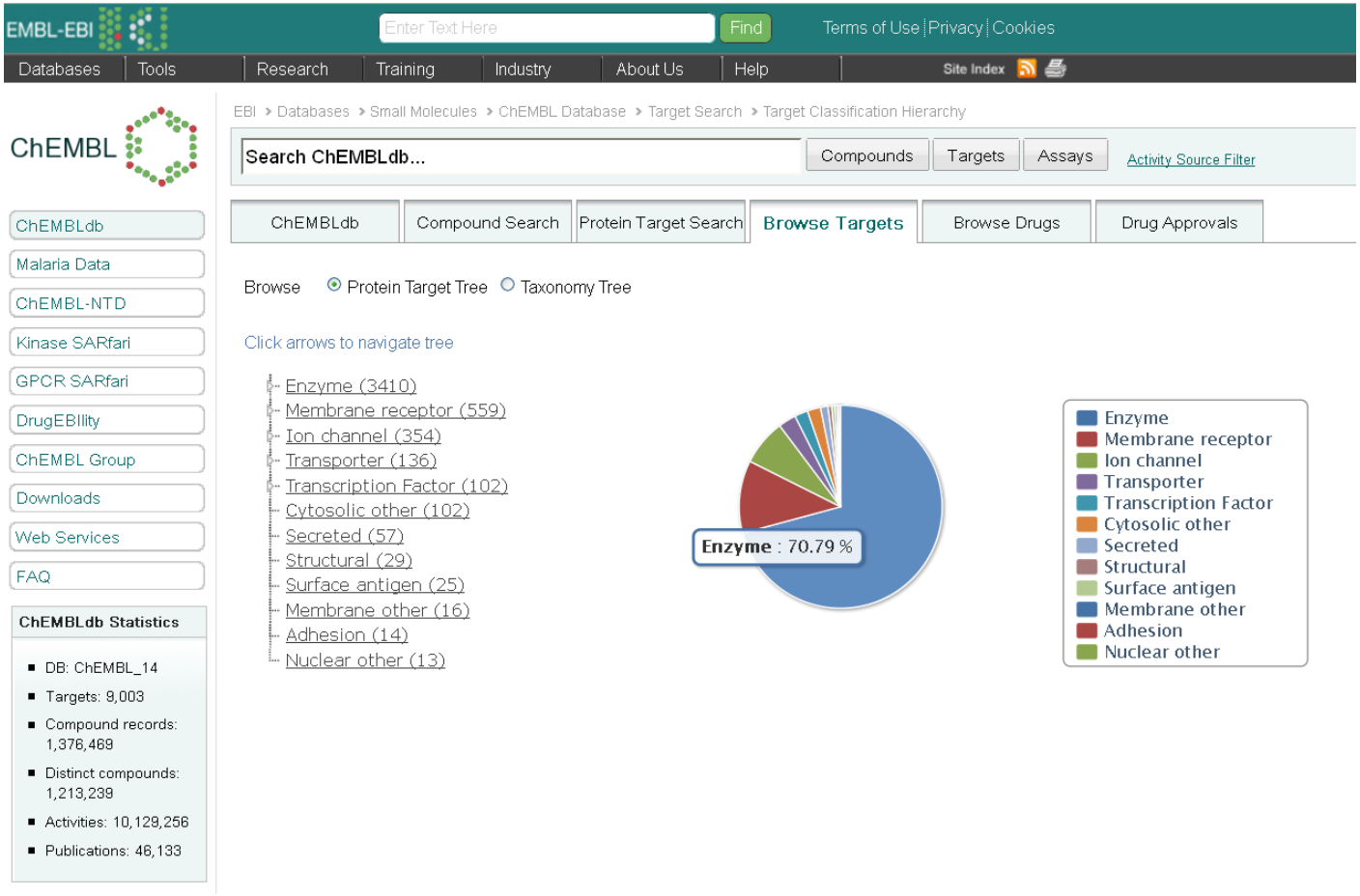


Databases of complexes

- **ChEMBL**
  - https://www.ebi.ac.uk/chembldb/
  - 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

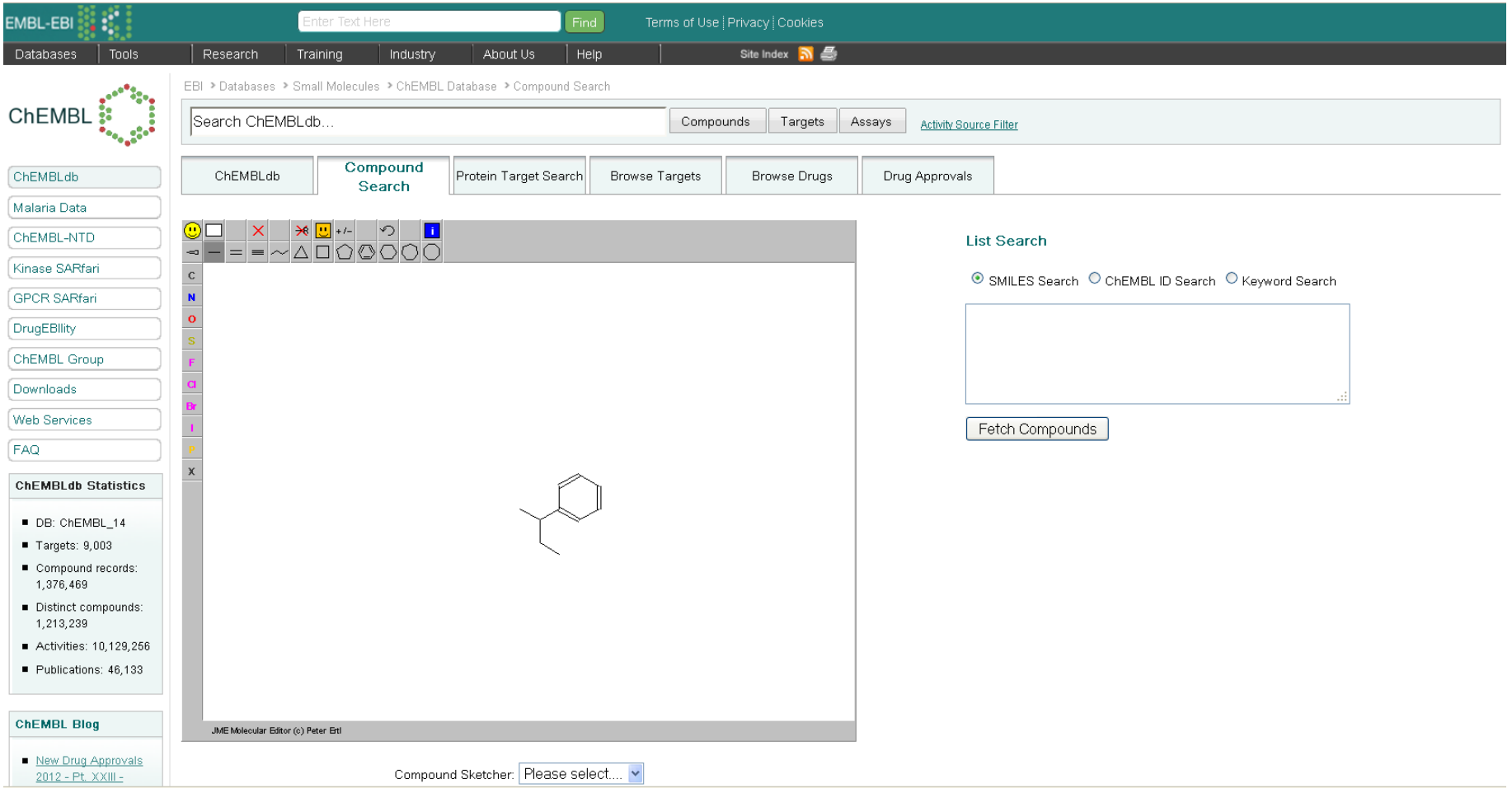
Databases of complexes

- **ChEMBL**



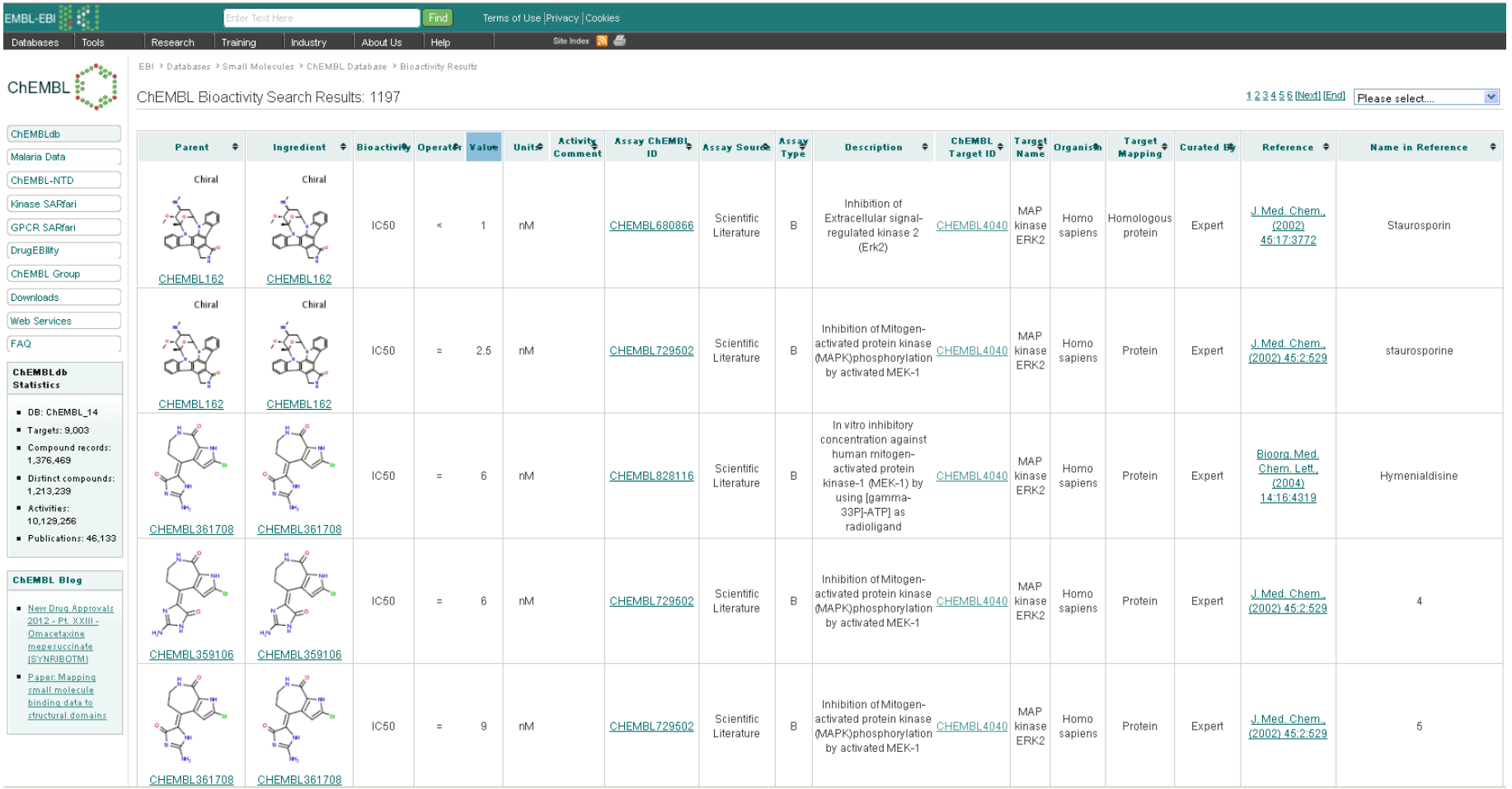
Databases of complexes

- **ChEMBL**



Databases of complexes

- **ChEMBL**



EMBL-EBI | Enter Text Here | Find | Terms of Use | Privacy | Cookies

Databases | Tools | Research | Training | Industry | About Us | Help | Site Index

ChEMBL

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

ChEMBLdb | Malaria Data | ChEMBL-NTD | Kinase SARfari | GPCR SARfari | DrugEBIlity | 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 (SYNRIBOTM)
- Paper: Mapping small molecule binding data to structural domains

ChEMBL Bioactivity Search Results: 1197

1 2 3 4 5 6 [Next] [End] Please select...

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		CHEMBL828116	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

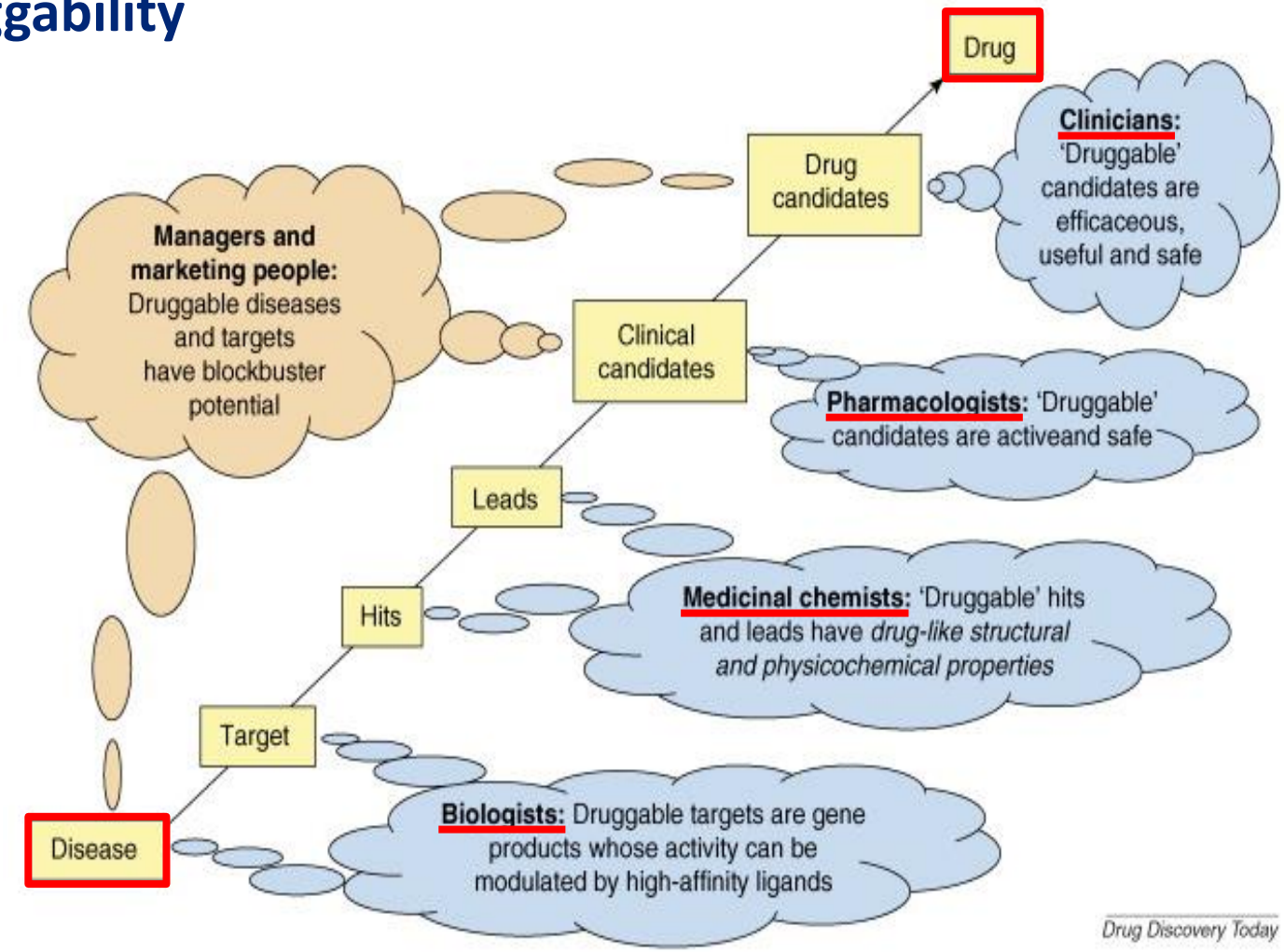
Protein druggability

- **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

Protein druggability

- **Druggability**



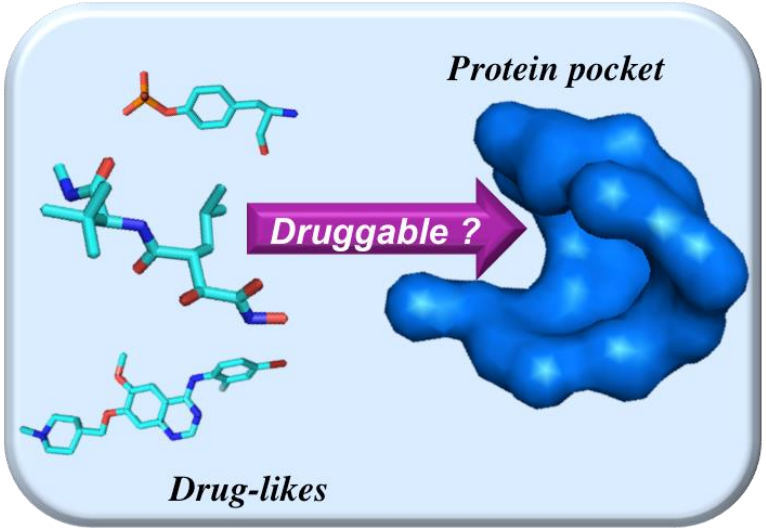
Protein druggability

- **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

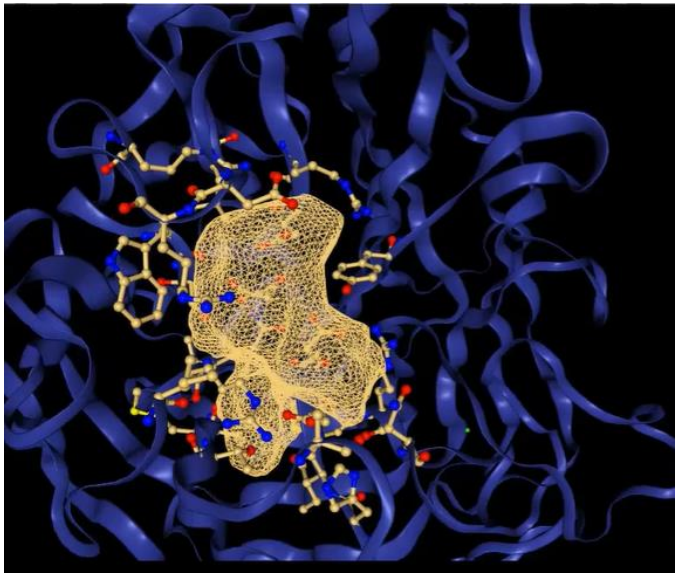
Protein druggability server

- **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)

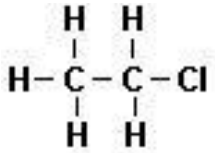


Small molecules

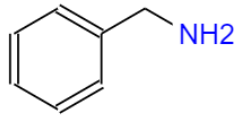
- Representation of small molecules
- Databases of small molecule
  - Cambridge Structural Database
  - PUBCHEM database
  - ZINC database
- Preparation of small molecule structure

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)

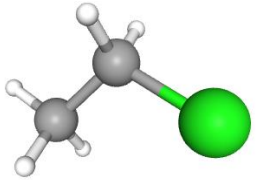


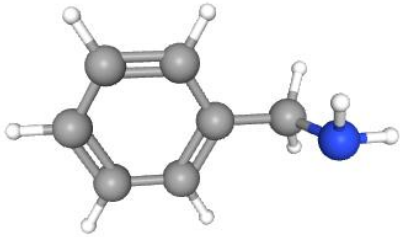
→ CCCl



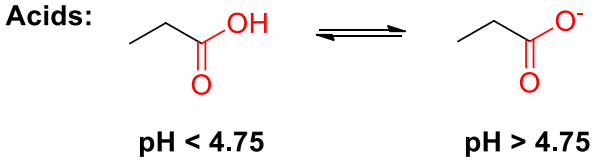
→ C1=CC=C(C=C1)CN

- **3D – atomic coordinates**
  - Usually: PDB, SDF or MOL2 files





- Beware: may have different protonation states



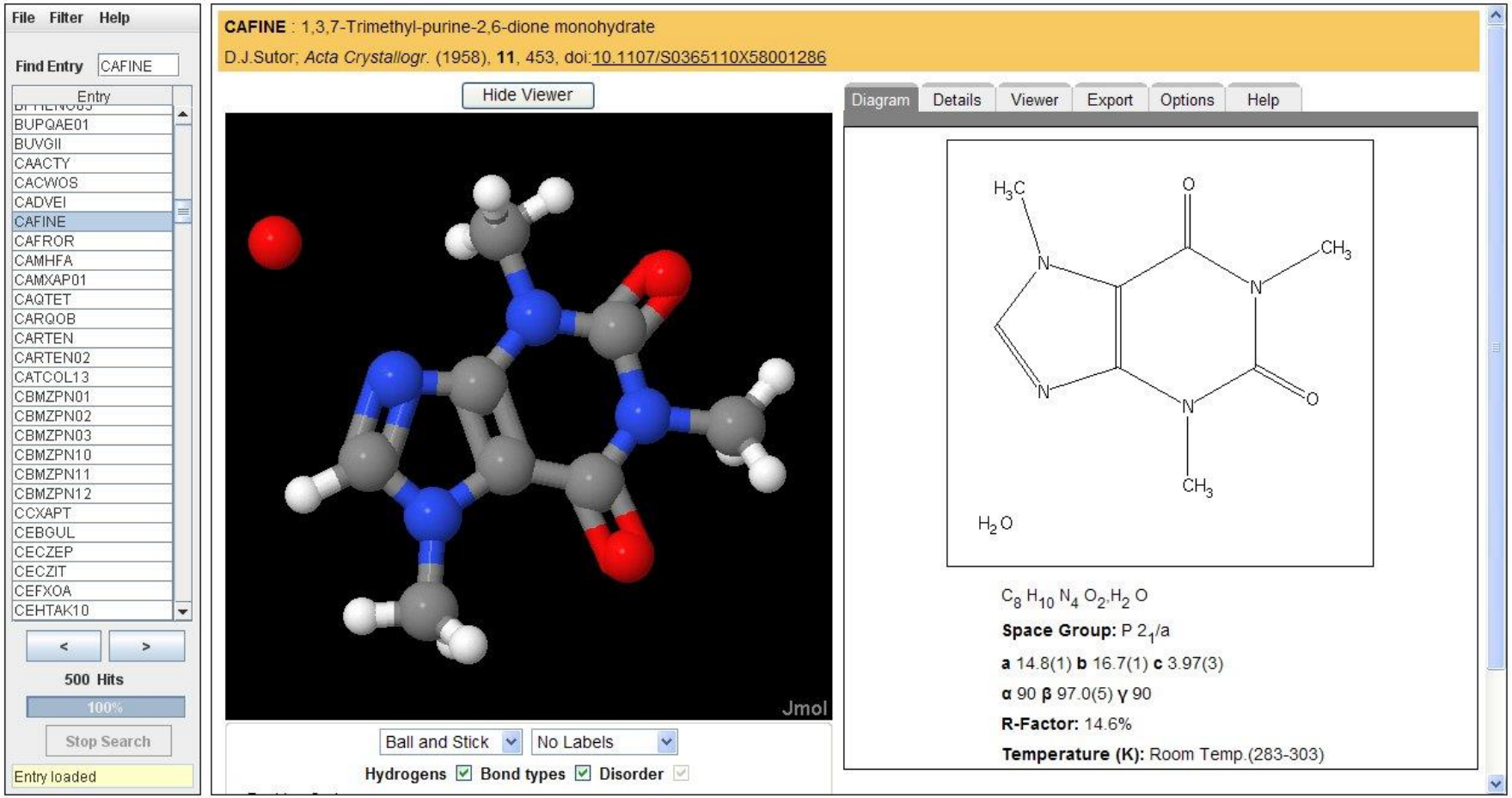
Databases of small molecule

- **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**



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

Databases of small molecule

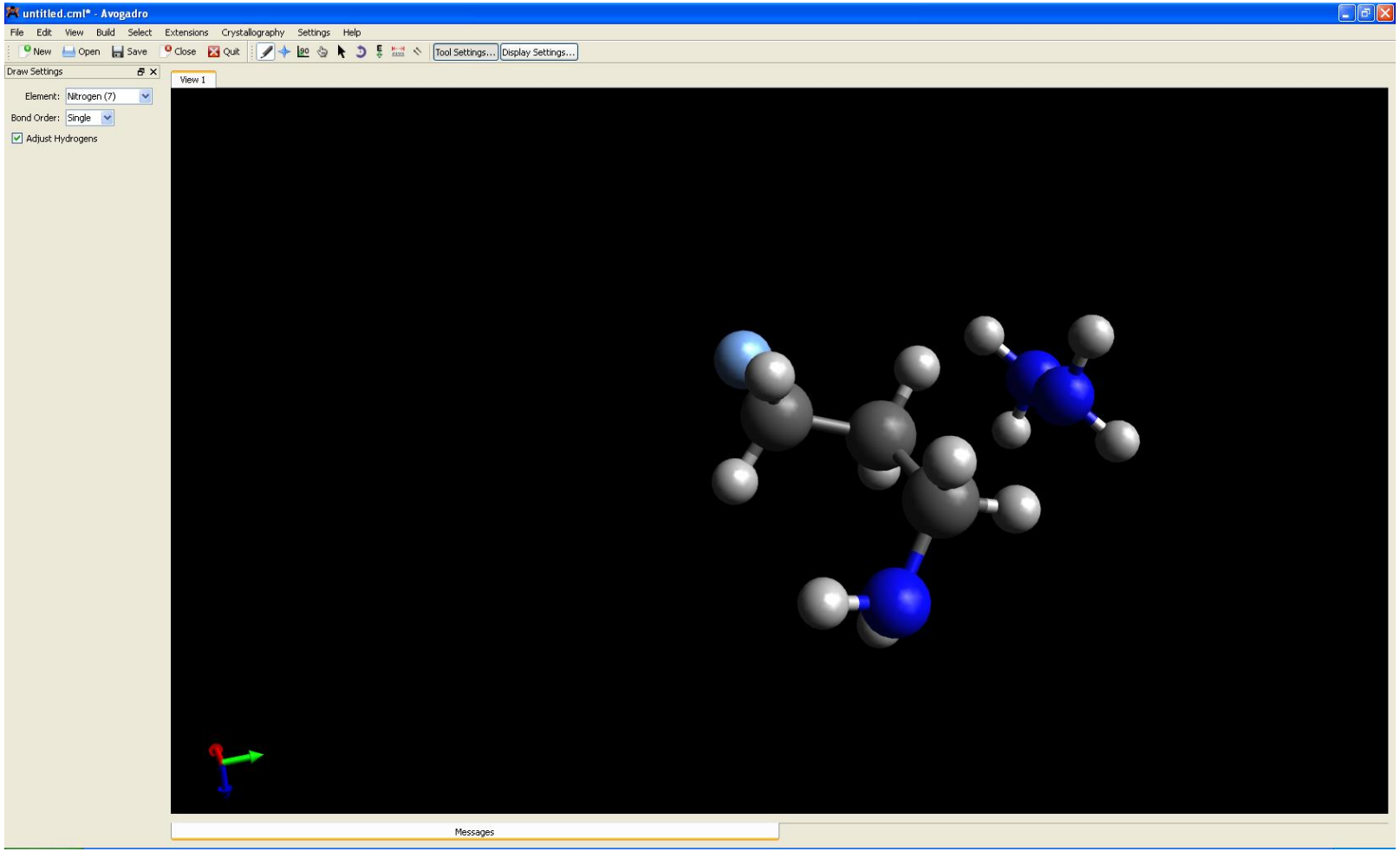
- **ZINC database**
  - http://zinc.docking.org/
  - Free public resource for ligand discovery
  - 3D coordinates in ready-to-dock formats (ex: added hydrogens, partial atomic charges, …)
  - Molecules in biologically relevant protonation and tautomeric forms
  - About 37 billion unique molecules grouped 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

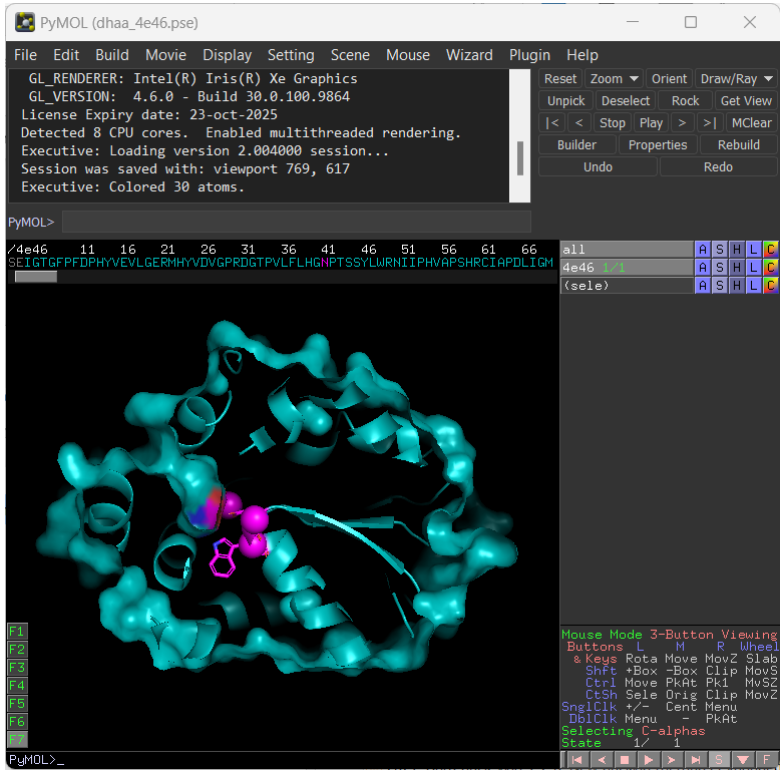
Preparation of small molecule structure

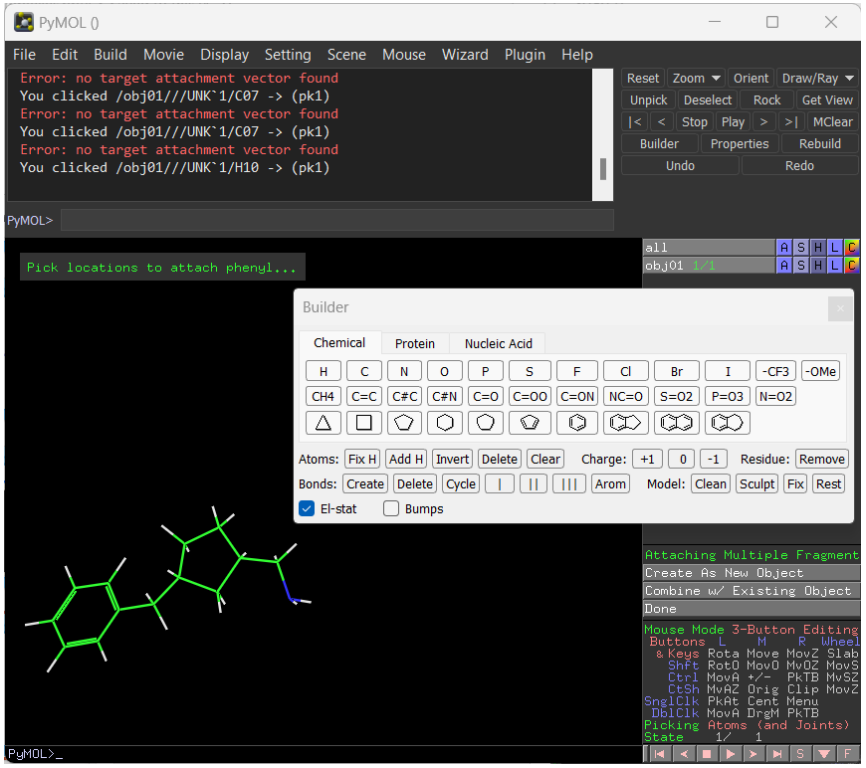
- **AVOGARO**



Preparation of small molecule structure

- **PyMOL**
  - https://pymol.org/
  - Powerful molecular visualizer and editor

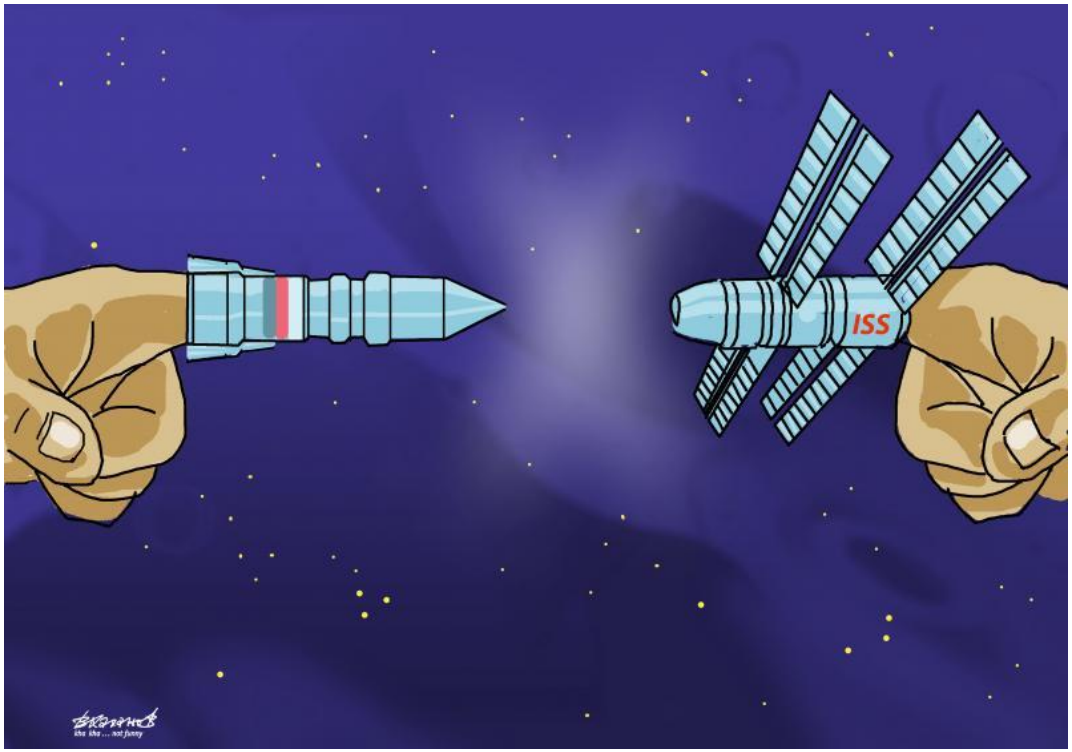




Preparation of small molecule structure

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

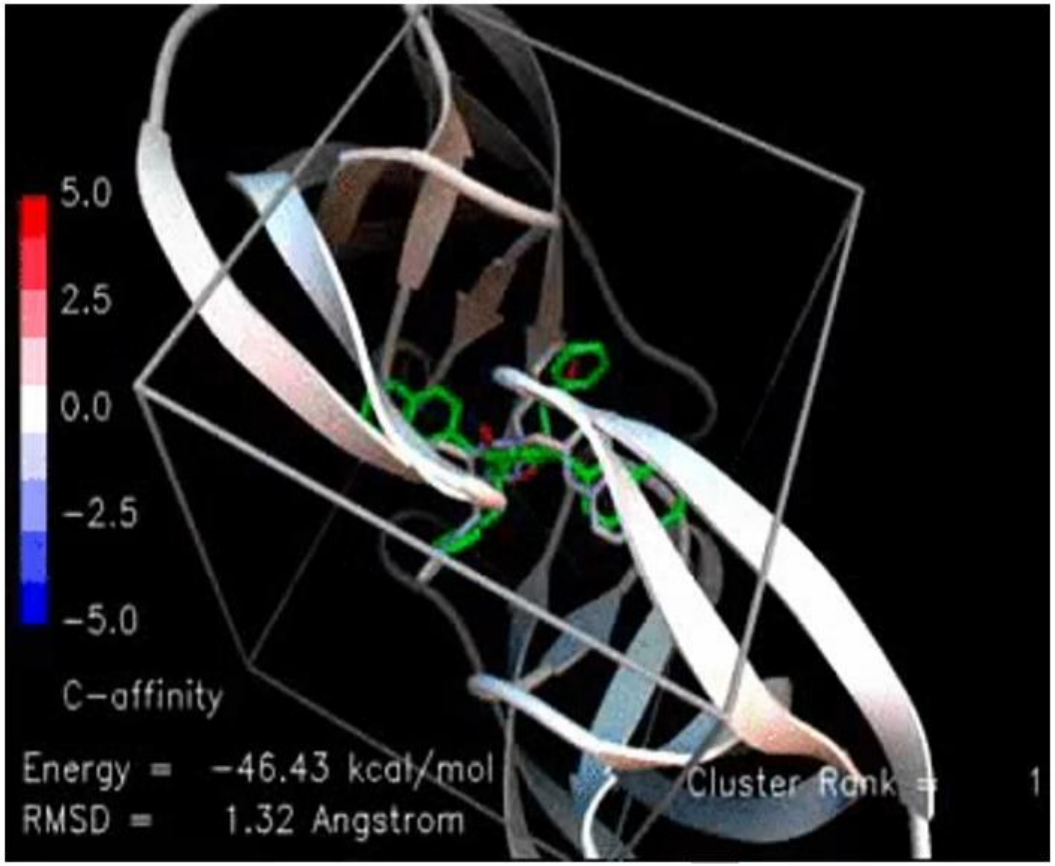
Molecular docking



Molecular docking

- Useful when experimental data is not available

  or for virtual screening



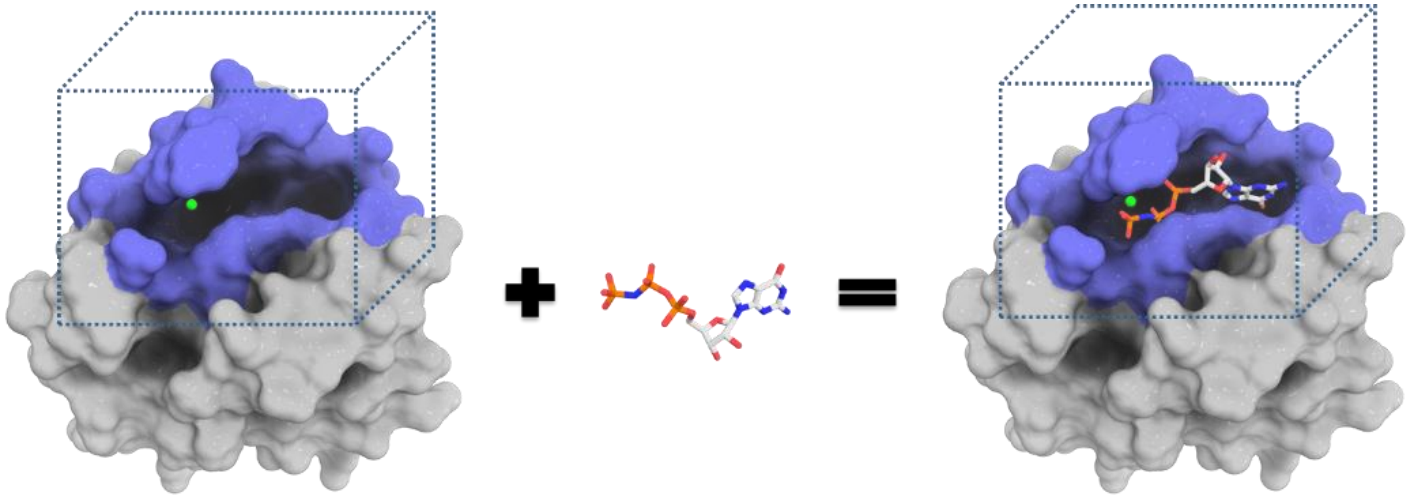
Score →
RMSD →

Crystal (experimental)

Docking attempts

Molecular docking

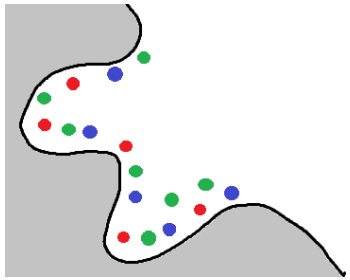
- Several components/steps
  - Receptor representation
  - Ligand representation
  - Search of binding modes
  - Scoring



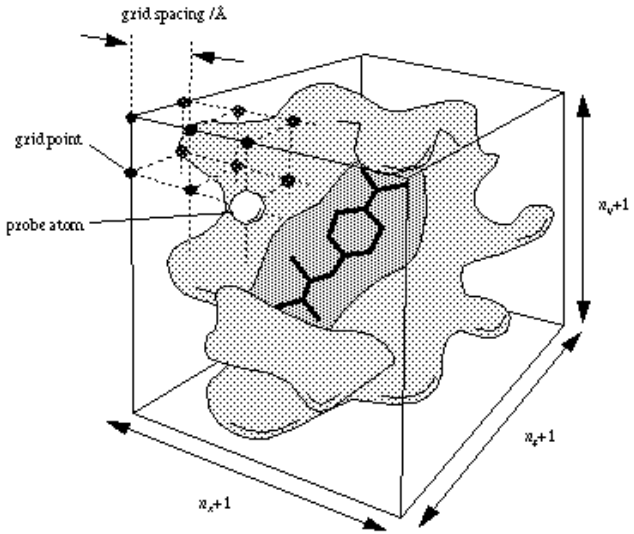
Receptor Ligand Complex

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, …)

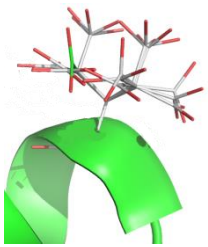


  - Grid representation – entire searched region is covered by orthogonal equidistant points carrying information about interactions of probe atom at this point with receptor atoms

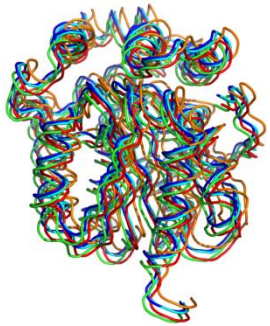


Receptor representation

- **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

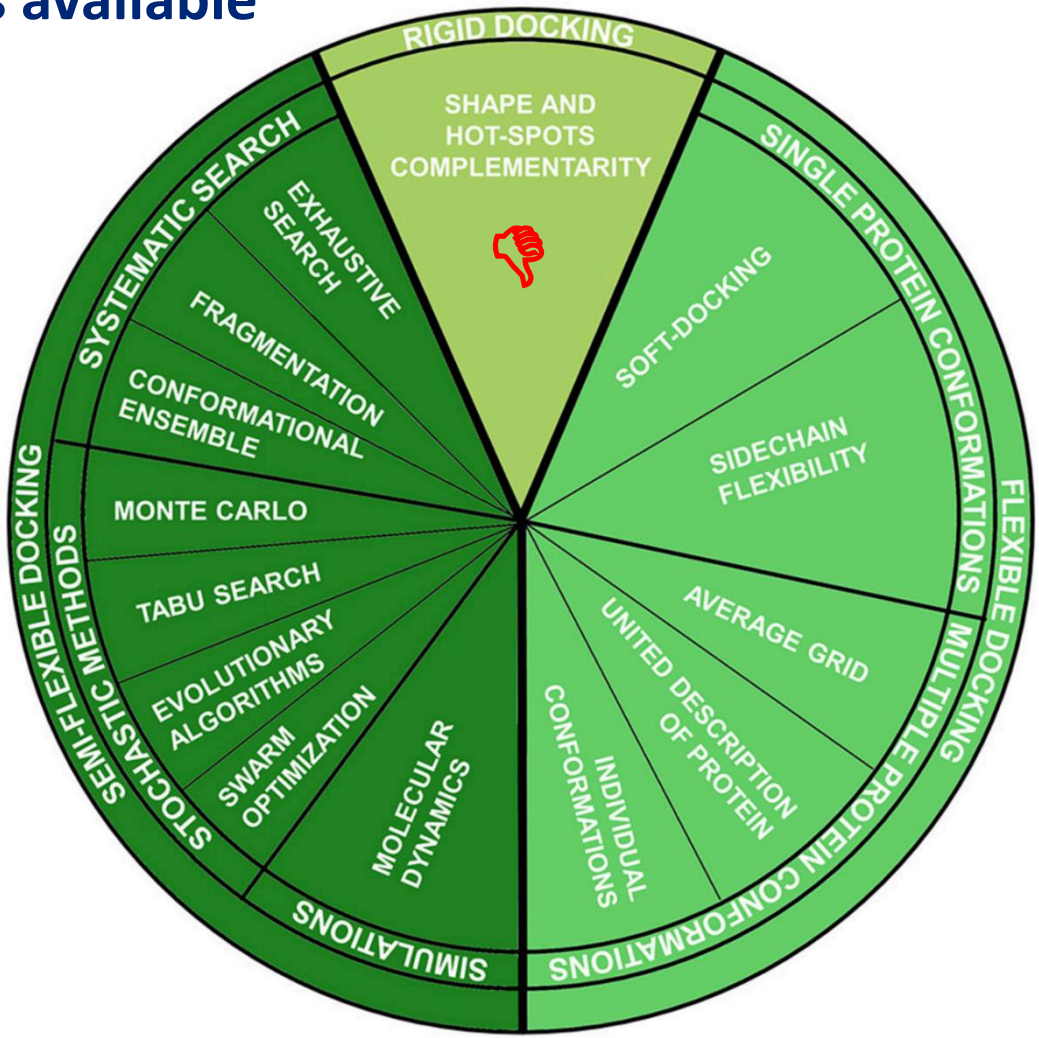


Ligand representation

- **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)



Molecular docking – search

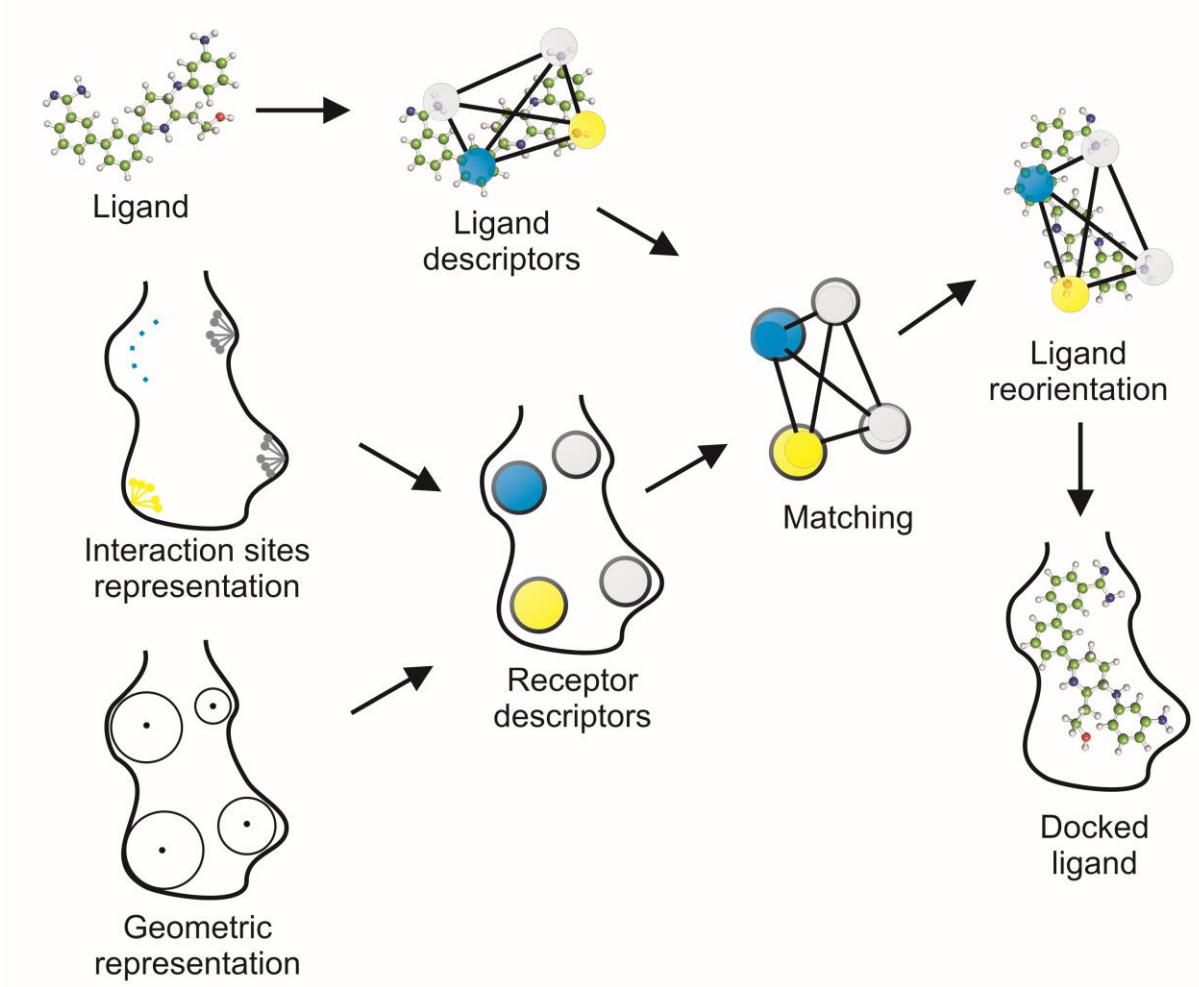
- **Geometry-based and combinatorial algorithms**
  - Assumes that binding is governed by shape and/or physicochemical complementarity between the ligand and the receptor
  - Assumes that the degree of complementarity is proportional to the binding energy which is not always true especially for more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

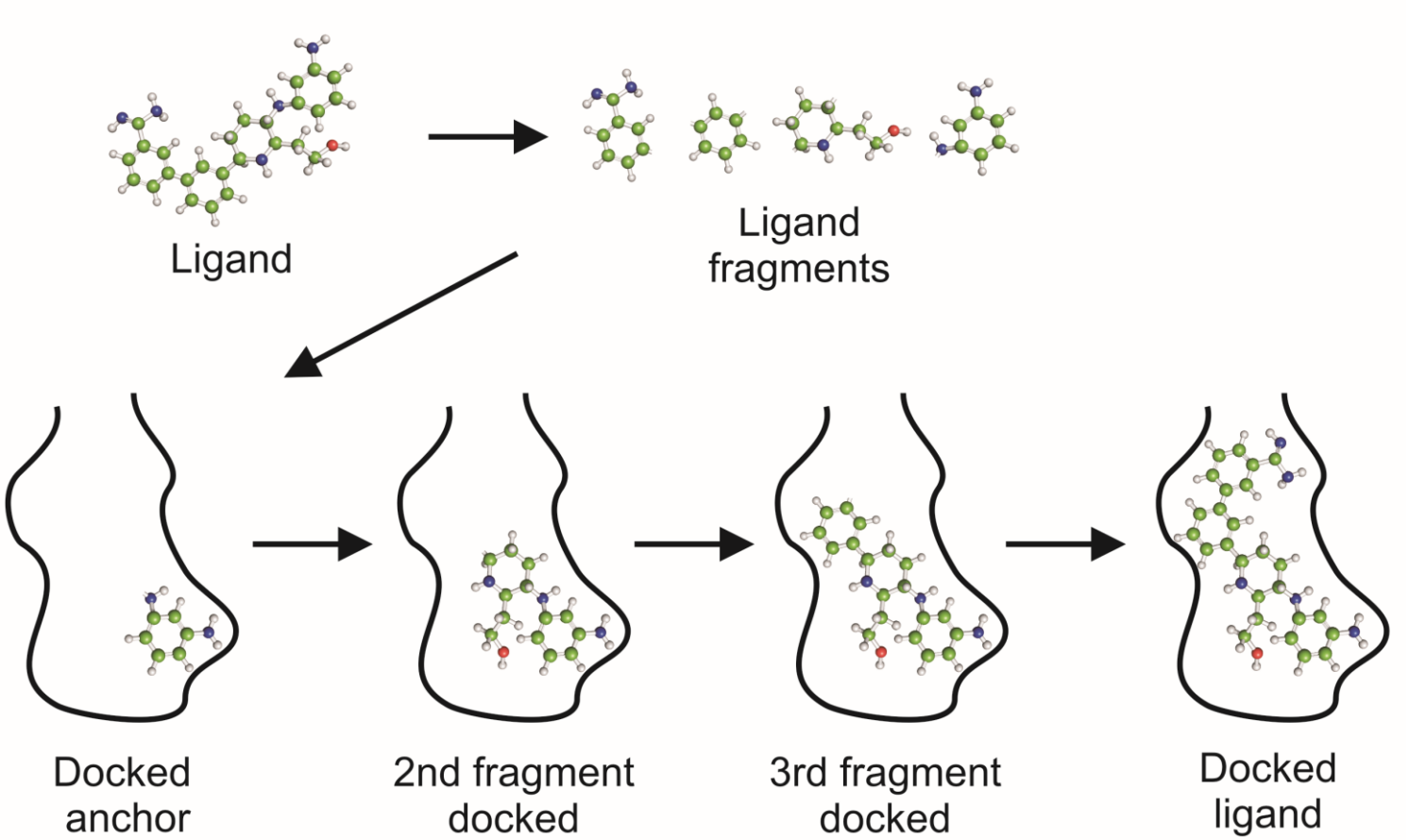


Geometry-based algorithms

- **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**

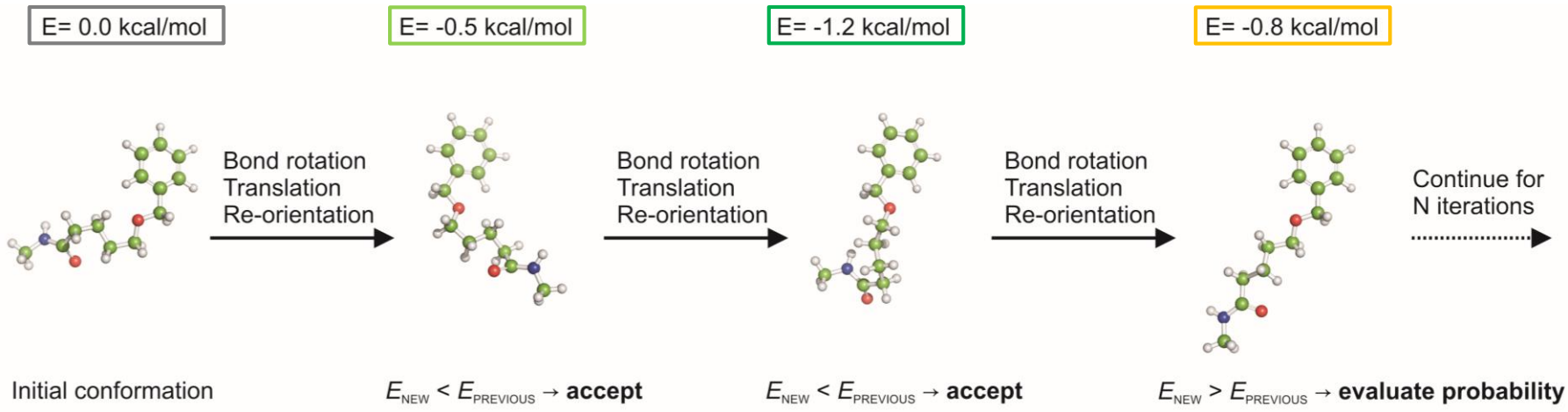


Stochastic energy-driven algorithms

- **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 then 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**

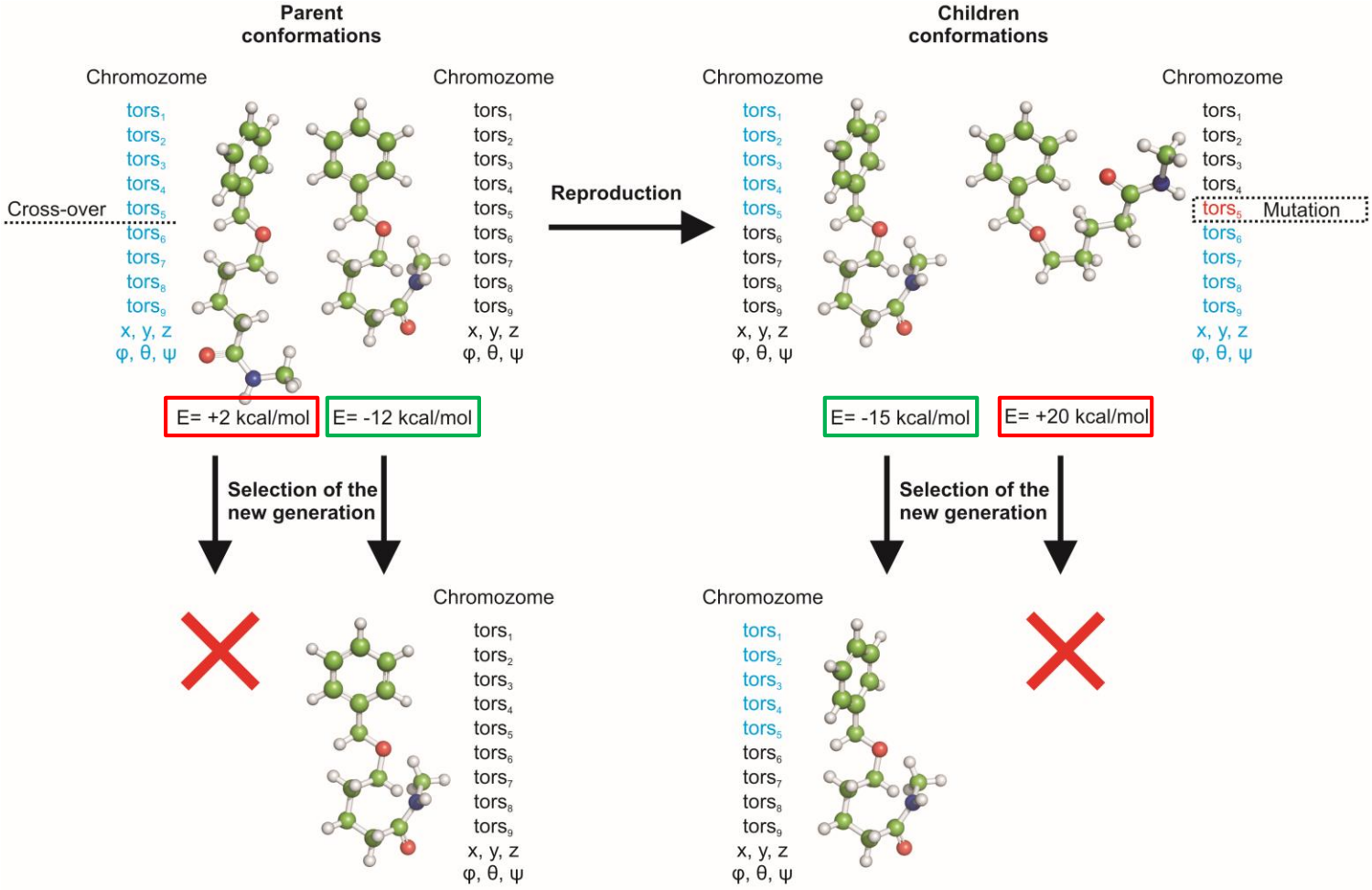


Stochastic energy-driven algorithms

- **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



Molecular docking – scoring

- **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

Molecular docking – scoring

- **Categories of scoring functions**
  - Empirical
  - Knowledge-based
  - Force field-based
  - Machine learning

Molecular docking – scoring

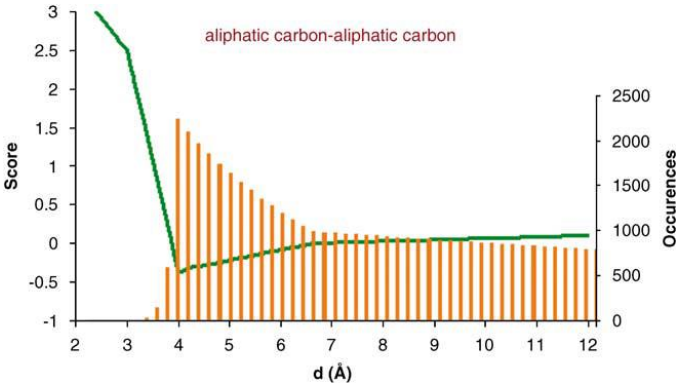
- **Categories of scoring functions**
  - **Empirical**
    - Derived by fitting of following equation to experimental binding affinities of known protein-ligand complexes

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

    - 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, …)

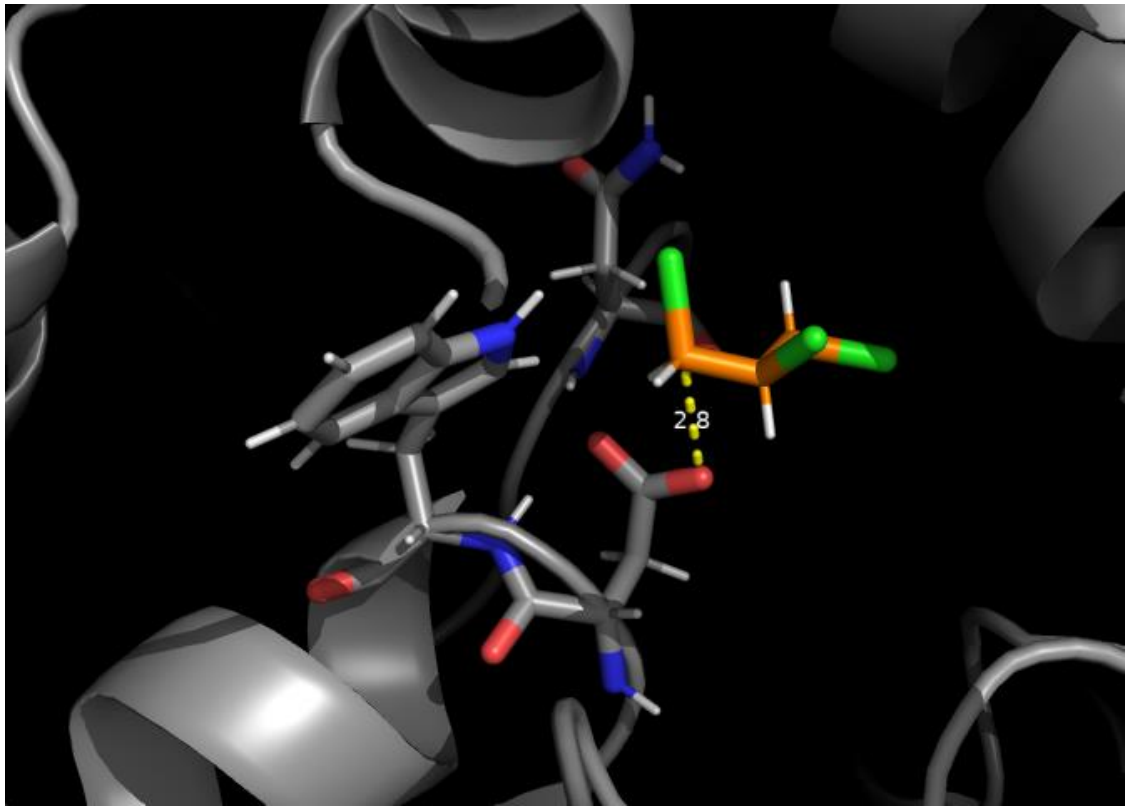
Molecular docking – scoring

- **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

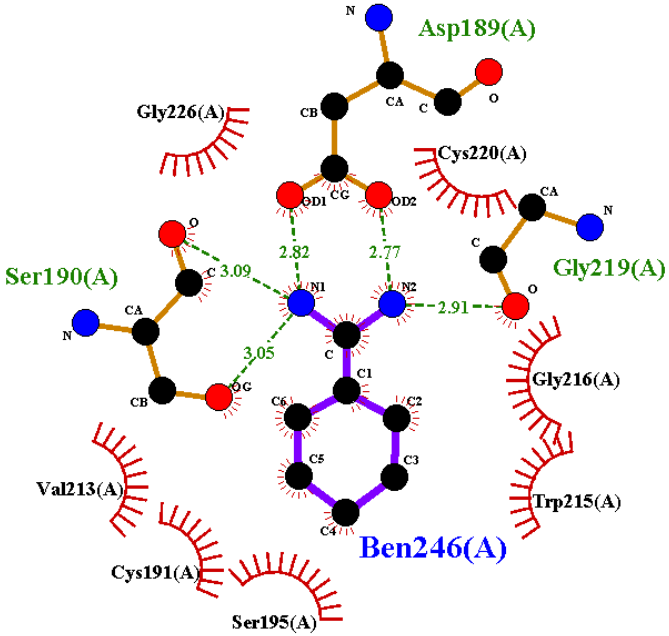


Intermolecular interactions

- **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

Transport of small molecules

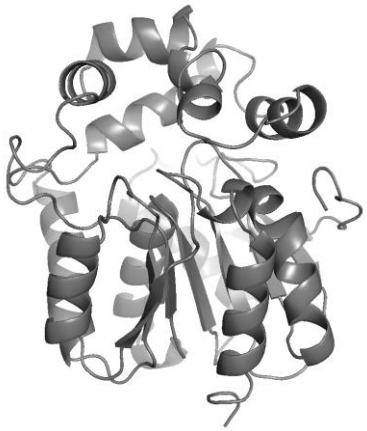
- Describe trajectory of ligands through tunnels
- Based on geometry or molecular docking
  - Fast but low accuracy
  - Good for screening purposes
  - CaverDock, MoMA-LigPath, SLITHER
- Based on force field
  - Run multiple MD simulations
  - Accurate but computationally demanding
  - Metadynamics,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

Transport of small molecules

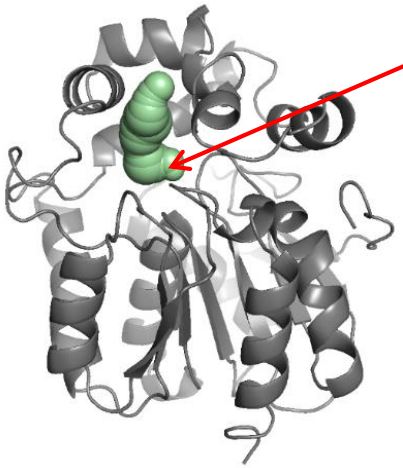
- CaverDock
  - https://loschmidt.chemi.muni.cz/caverdock/
  - Analysis of tunnels by Caver
  - Discretization of identified tunnel into discs
  - Molecular docking by AutoDock Vina to every disc

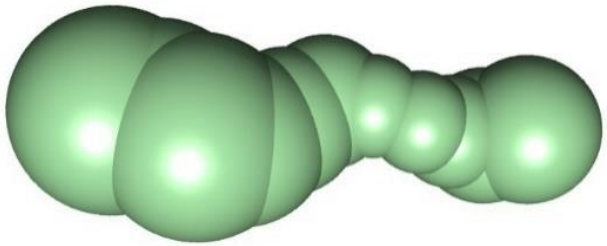
- Caver Web
  - https://loschmidt.chemi.muni.cz/caverweb/
  - Web interface for Caver and CaverDock

CaverDock

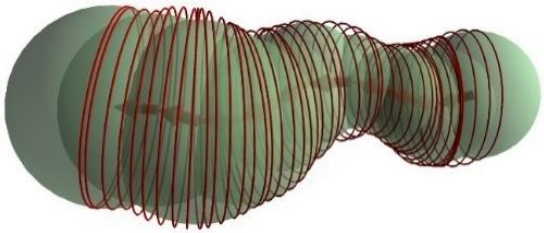


CAVER

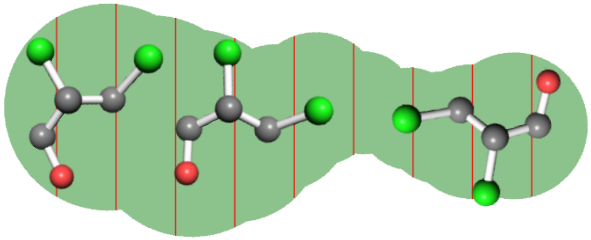




Discretization

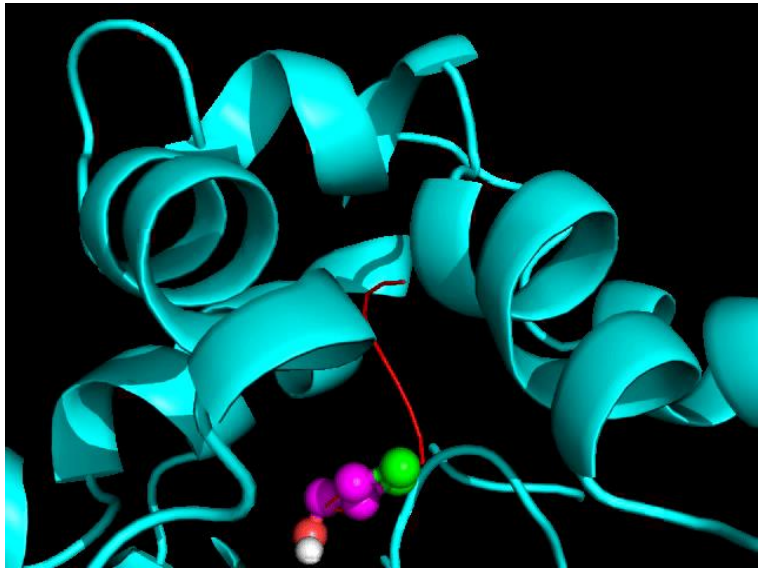


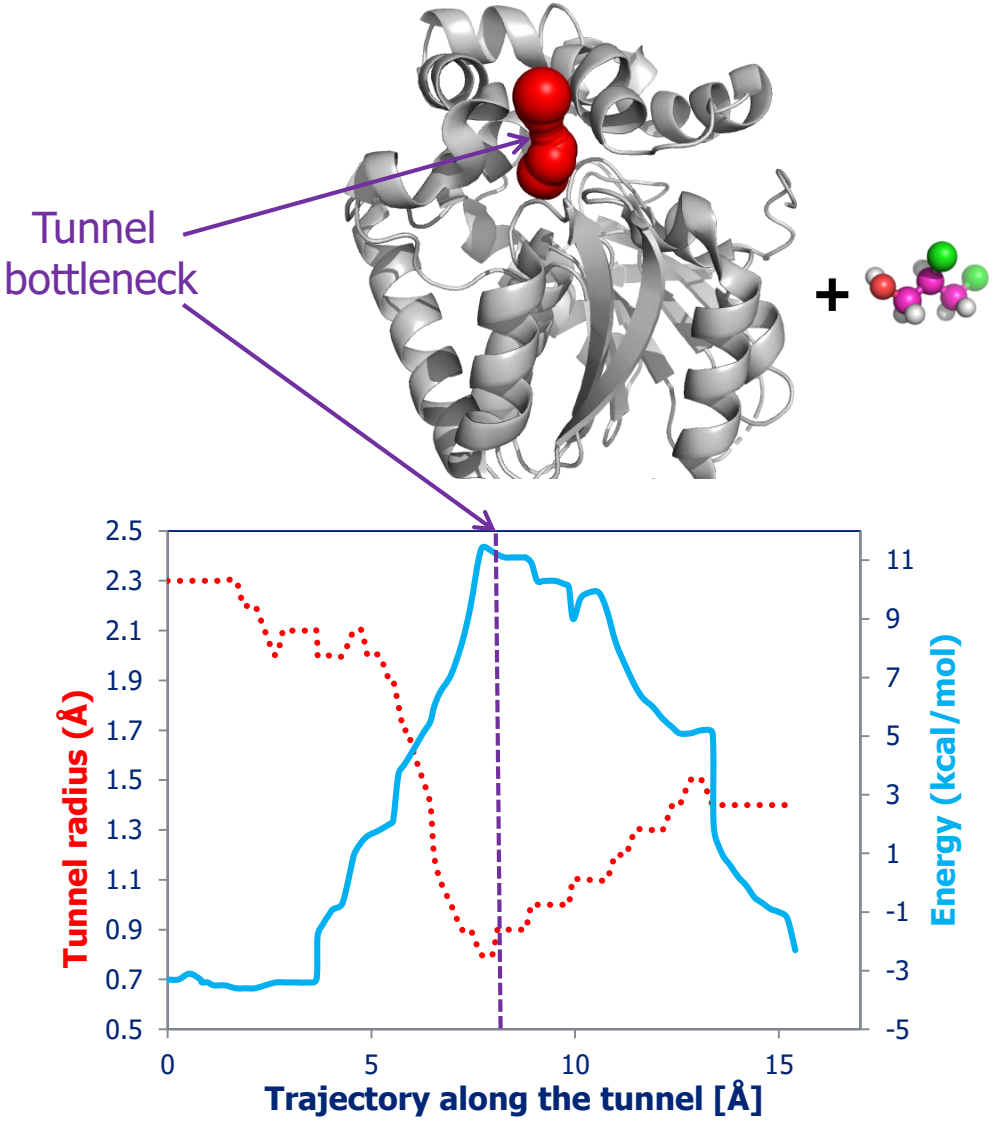
CaverDock



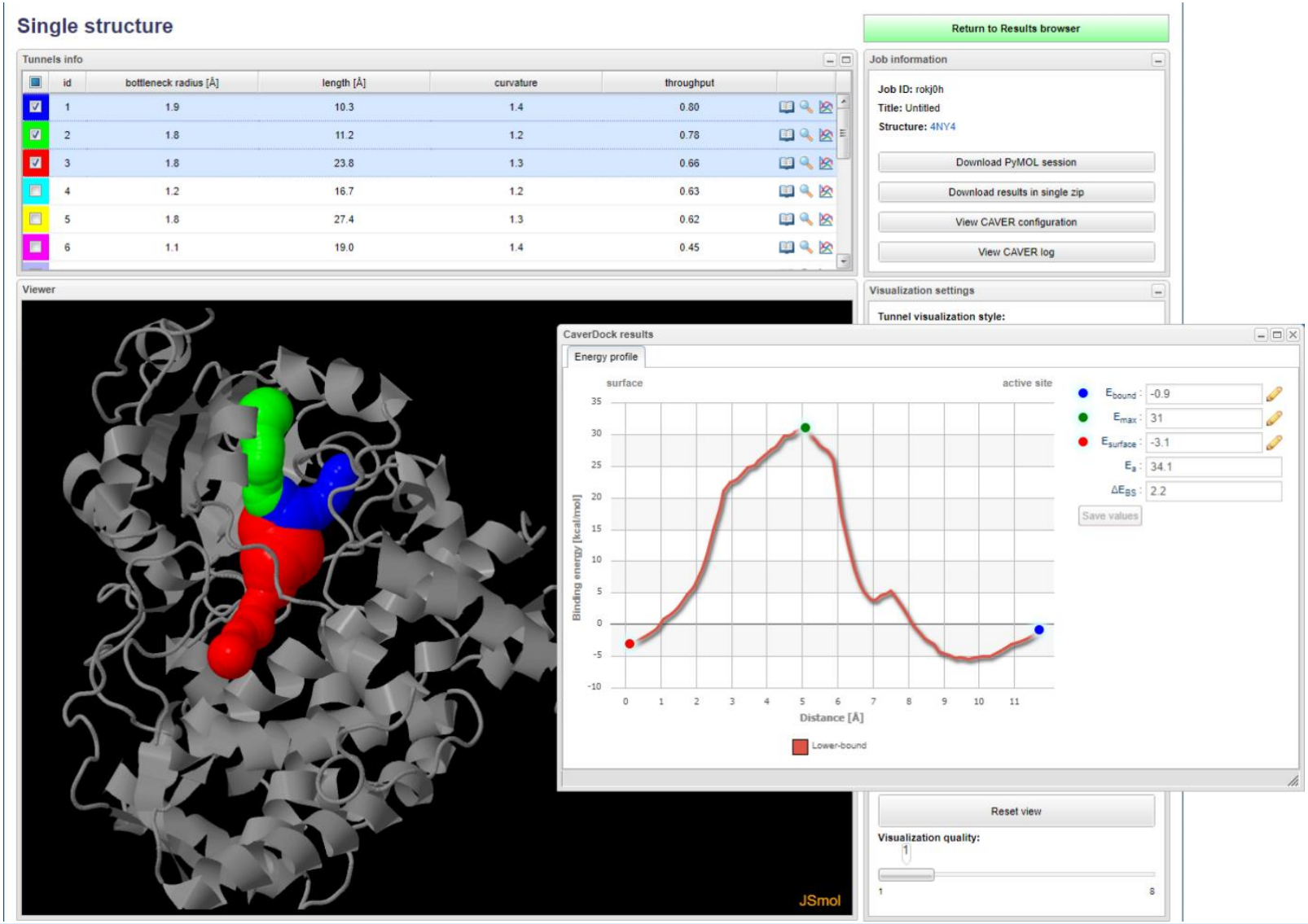
CaverDock

- Results provided:
  - Ligand trajectory
  - Energy profile





CaverDock over Caver Web



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