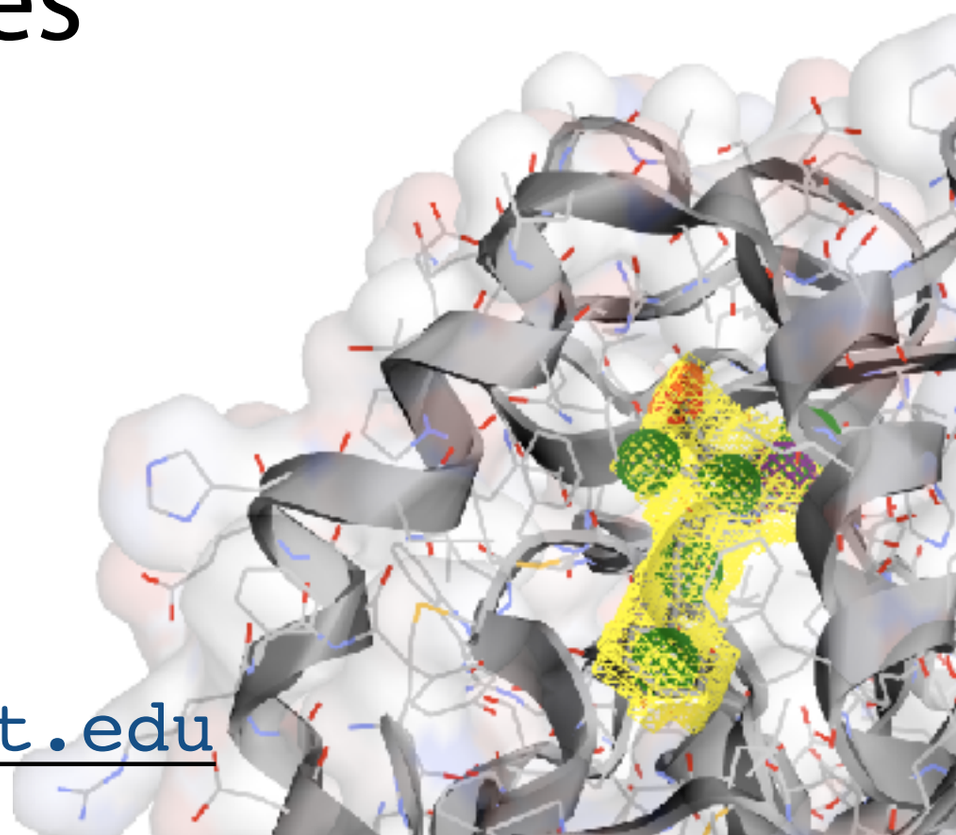
Computational Drug Discovery

David Ryan Koes

7/9/2018

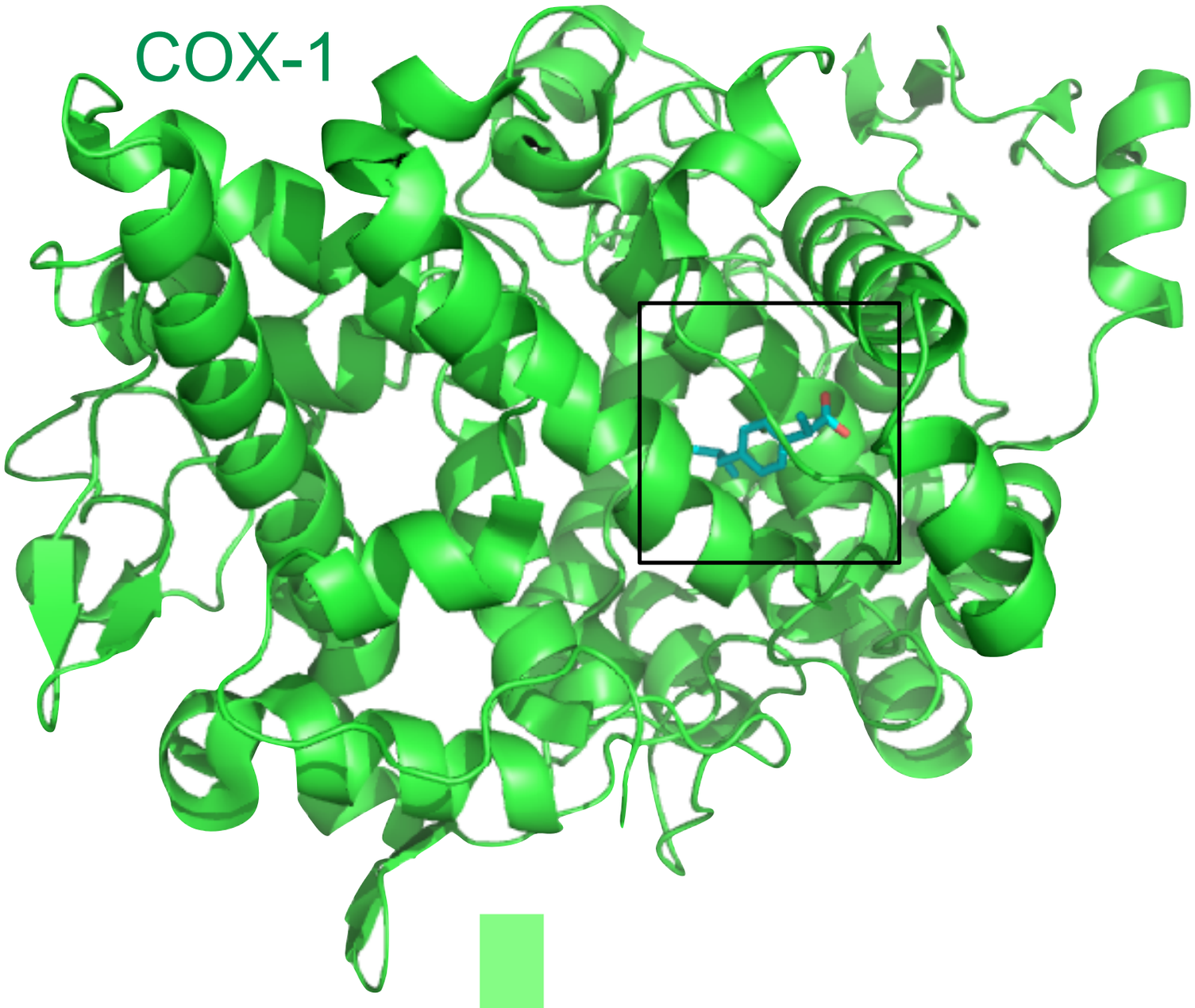


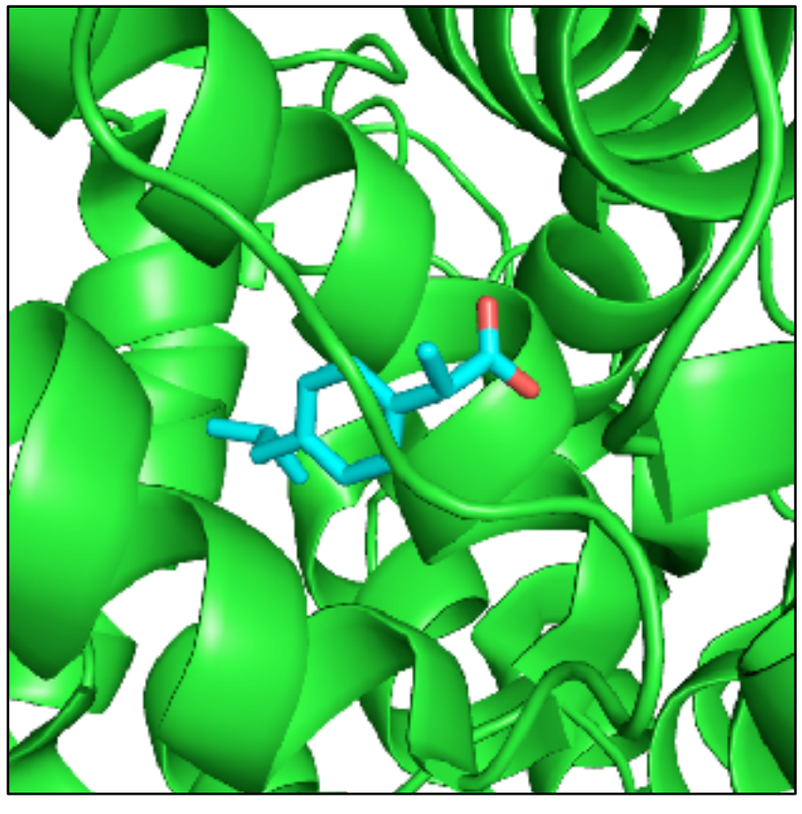
http://bits.csb.pitt.edu

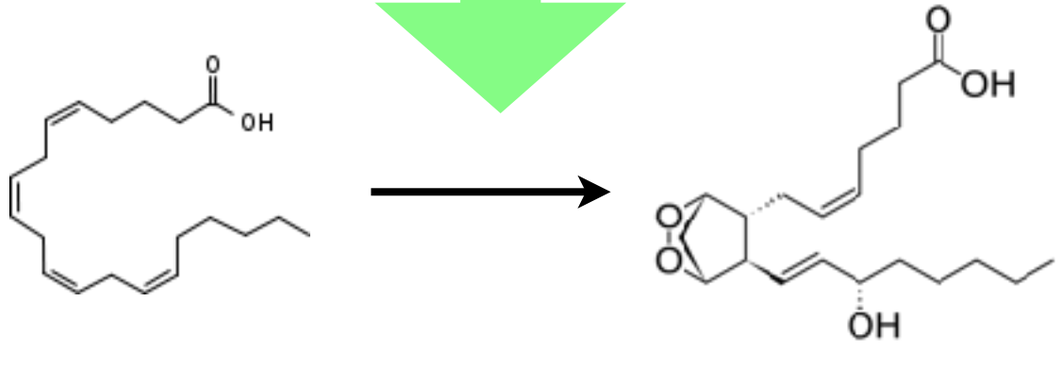
What is a drug?

According to the Food, Drug, and Cosmetic Act (1) **:** a substance recognized in an official pharmacopoeia or formulary (2) **:** a substance intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease (3) **:** **a substance** other than food **intended to affect the structure or function** of the body (4) **:** a substance intended for use as a component of a medicine but not a device or a component, part, or accessory of a device
http://www.merriam-webster.com/dictionary/drug

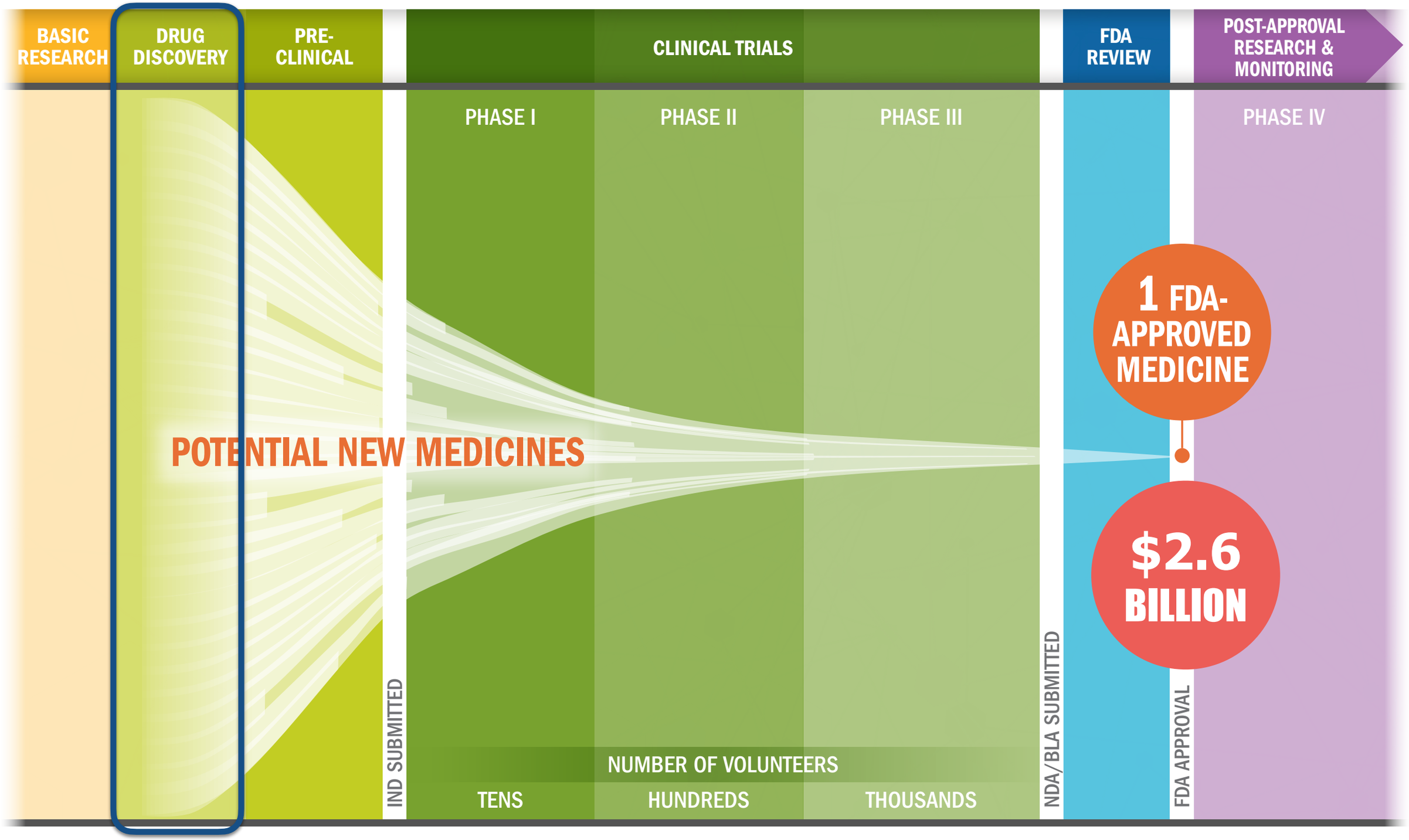
A small molecule intended to affect the structure/function of macromolecules


COX-1




OH
O
OH
O
O
OH

THE BIOPHARMACEUTICAL RESEARCH AND DEVELOPMENT PROCESS

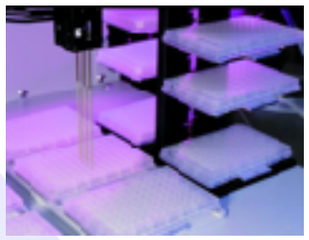


Source: Pharmaceutical Research and Manufacturers of America (http://phrma.org)

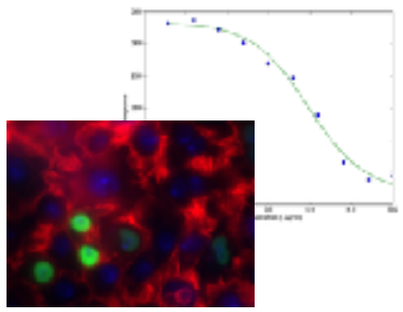
1. Does the compound do what you want it to?
2. Does the compound **not** do what you **don't** want it to?
3. Is what you want it to do the right thing?

Drug Discovery

Target Identification



Screening



Lead Identification

Lead Optimization

Compounds **Hits** **Leads** **Clinical Candidates**

Cost

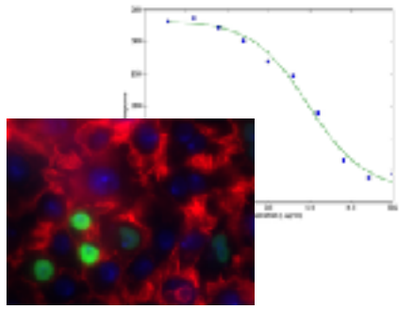
Computational Drug Discovery

Omics

Target Identification

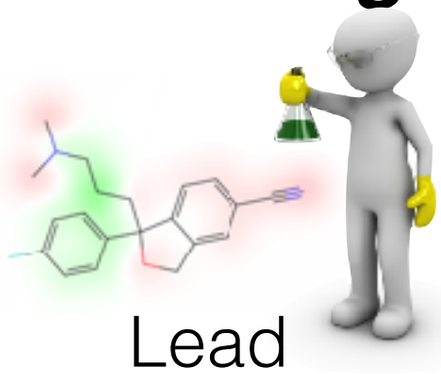
Virtual

Screening



Lead Identification

Modeling



Lead Optimization

Compounds **Hits** **Leads** **Clinical Candidates**

Cost

Kinds of Virtual Screening

ADMET

Ligand Based

- similarity to known binder
- QSAR
- pharmacophore

Receptor Based

- dock and score
- simulation

ADMET

Absorption
Distribution
Metabolism
Excretion
Toxicity

Will this be a usable drug?

Screening for ADMET:
Cytochrome P450 interaction
Lipinksi's Rule of Five
QSPR: Quantitative Structure Property Relationship

Kinds of Virtual Screening

ADMET

Ligand Based

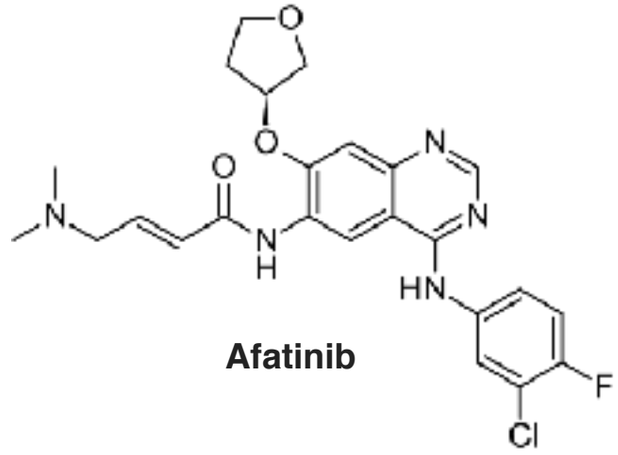
- similarity to known binder

- QSAR

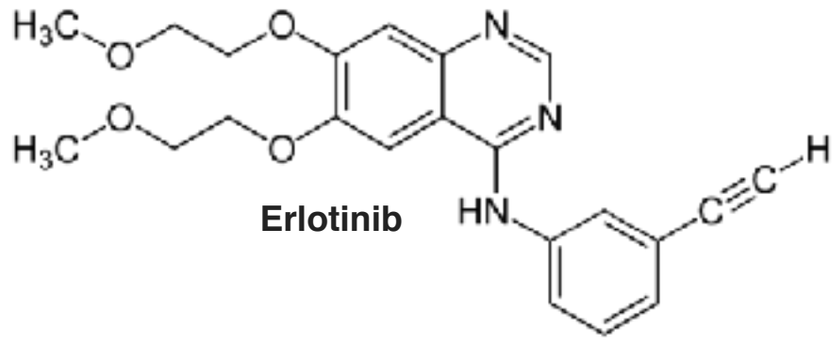
- pharmacophore

Receptor Based

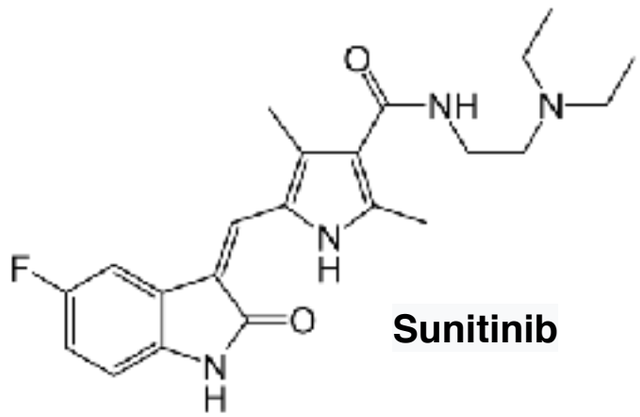
- dock and score



Afatinib



Erlotinib



Sunitinib

Ligand Based: Similarity

Fingerprint Methods

- map molecules to a descriptor space:

1D: molecule weight, #h-bonds, etc.
2D: paths, bond distances between atom-pairs

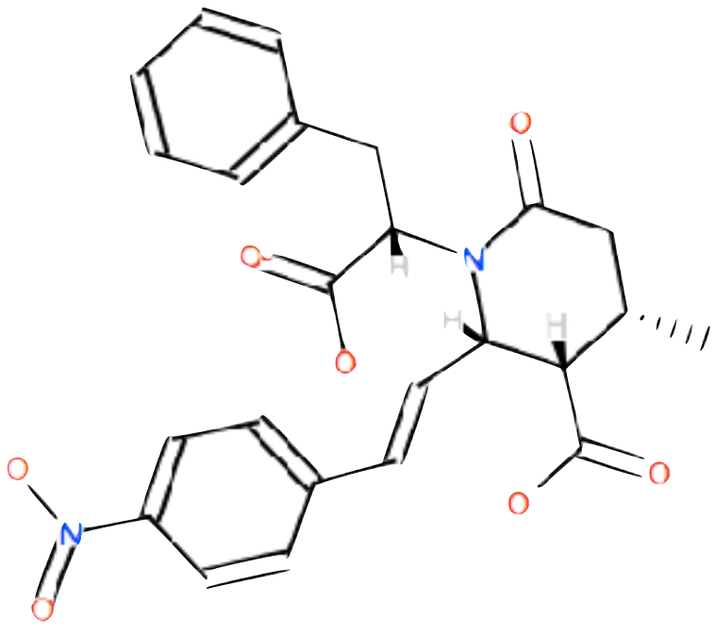
- similarity is “distance” between descriptors
- for bit vectors, Tanimoto distance used

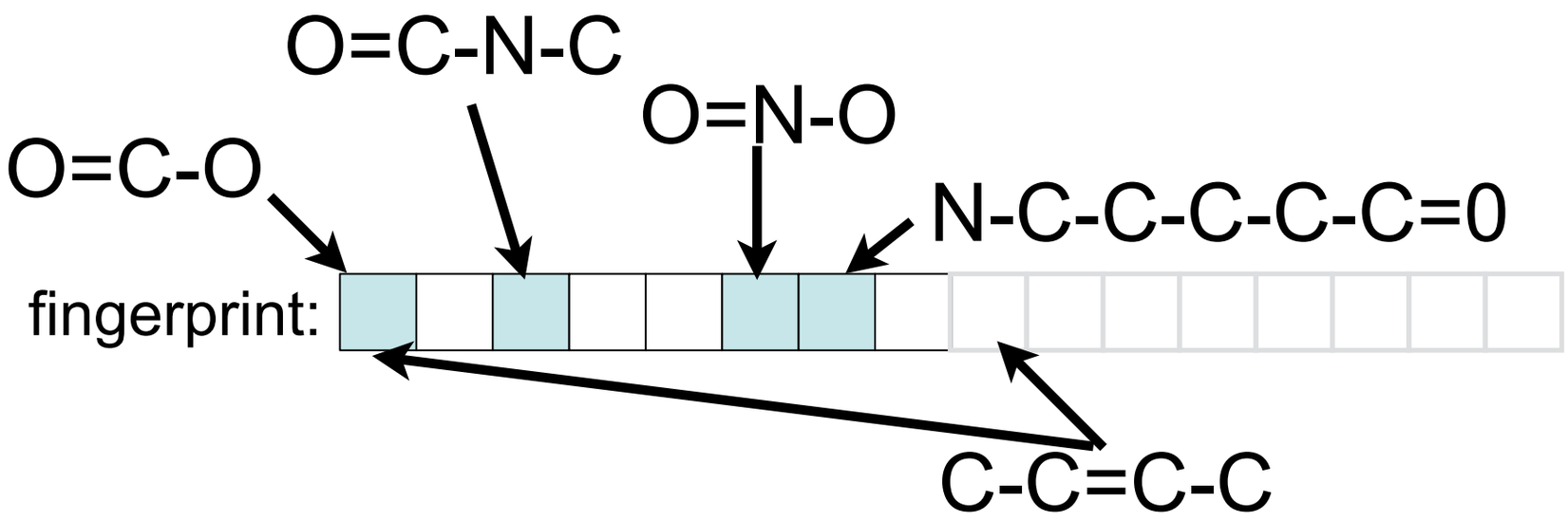
$$T(A,B) = \frac{\left|A \cap B\right|}{\left|A \cup B\right|}$$

Topological Fingerprints

Daylight/FP2 Fingerprints

- all paths up to 7 bonds long
- each path corresponds to bit position (**hashing**)
- fast similarity checking (Tanimoto)

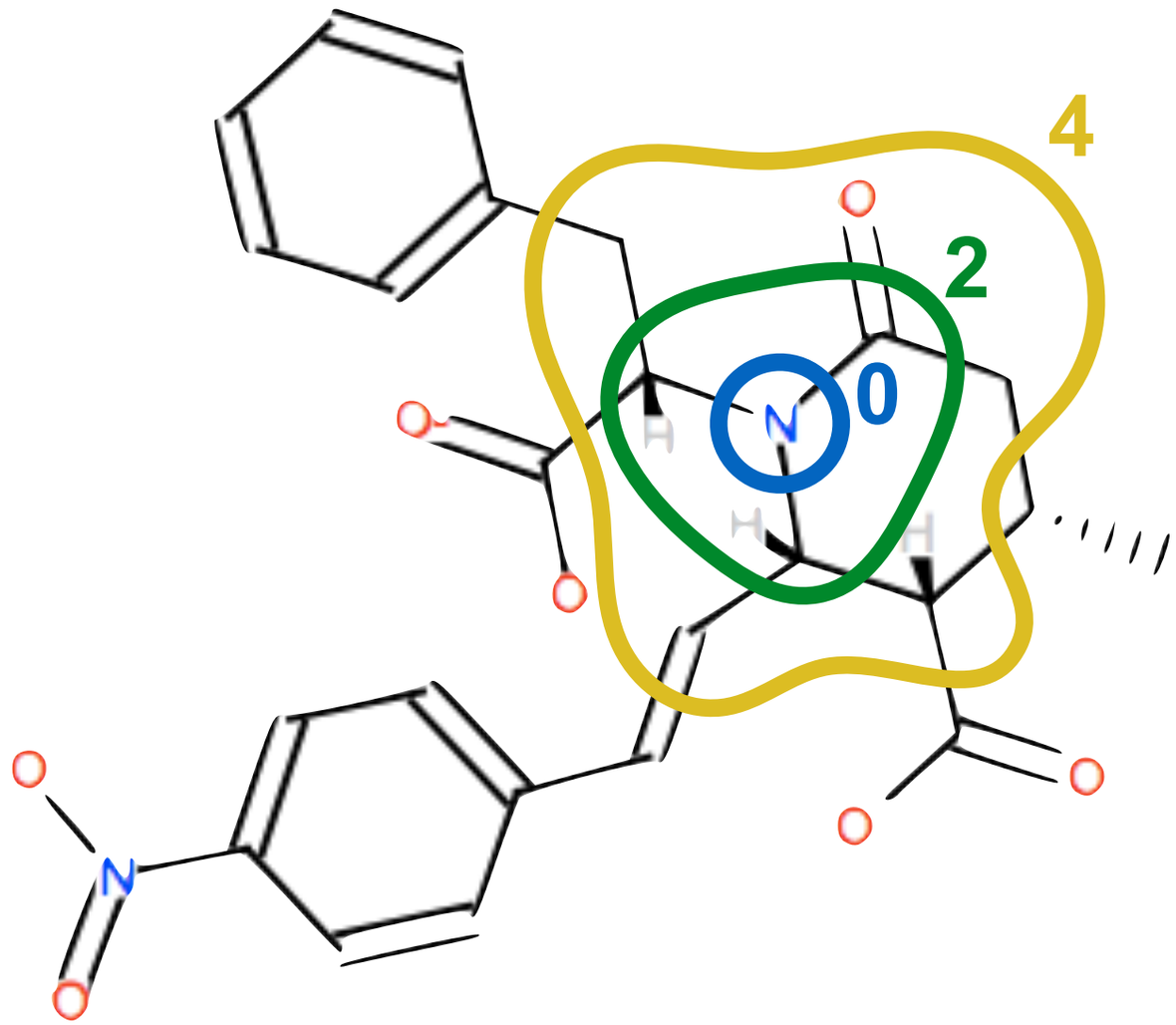




Topological Fingerprints

ECFP4

- all substructures with diameter 4 around every atom

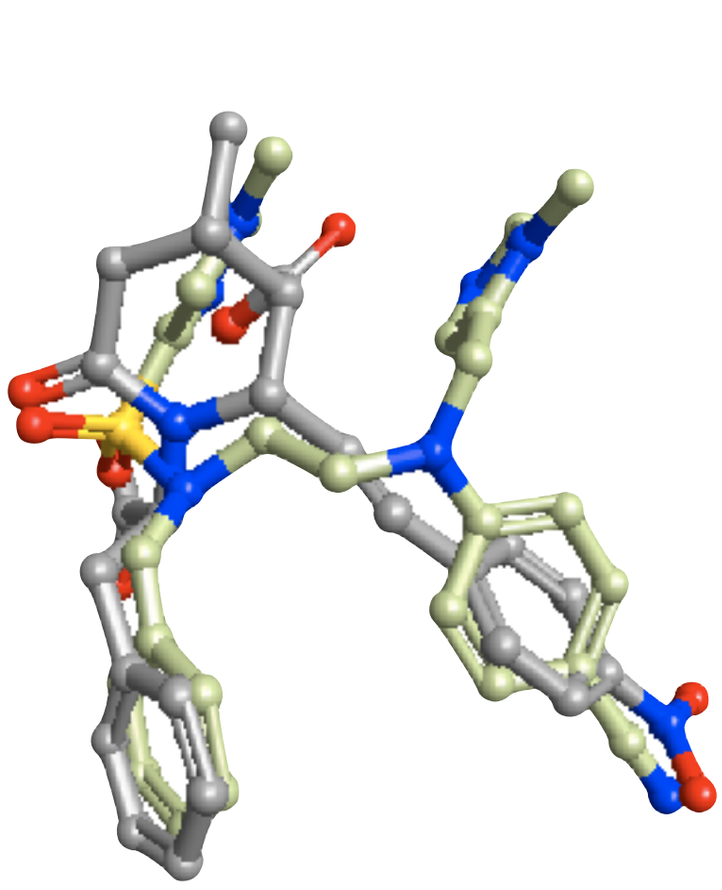


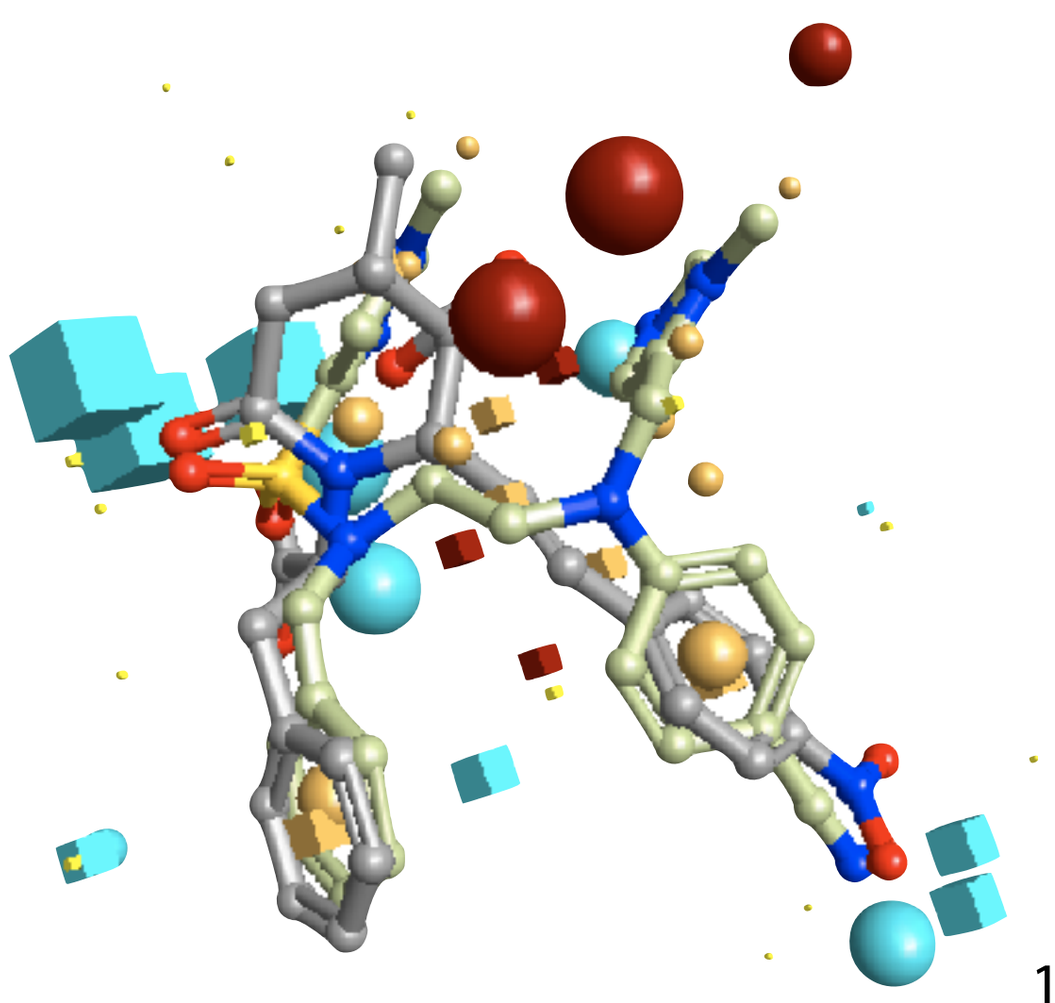
Ligand Based: Similarity

Superposition Methods

- compute "overlap" between molecules

- consider shape, electrostatics, **pharmacophores**





http://www.cresset-group.com/

Ligand Based: QSAR

Quantitative Structure/Activity Relationships

Properties

Compounds

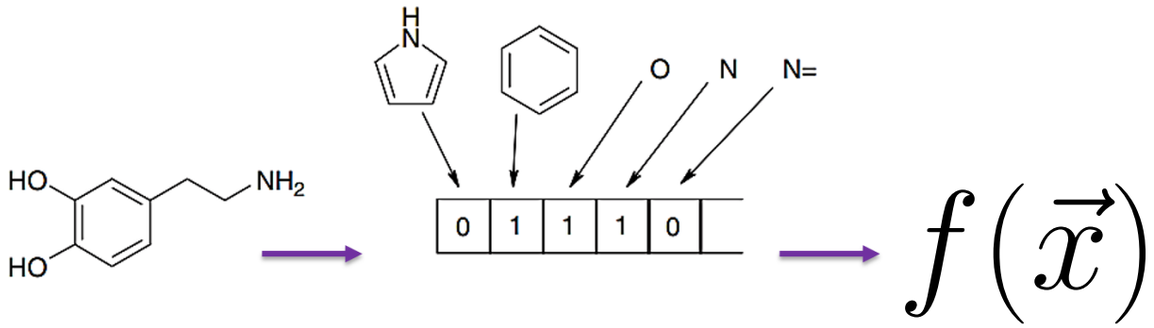
Cmpd Number	Cmpd Name	X	Log EC_{50}	π	Calculated Log EC_{50}	Residual
1	6a	H	1.07	0	0.79	0.28
2	6b	Cl	0.09	0.71	0.21	-0.12
3	6d	NO_2	0.66	-0.28	1.02	-0.36
4	6e	CN	1.42	-0.57	1.26	0.16
5	6f	C_6H_5	-0.62	1.96	-0.81	0.19
6	6g	$N(CH_3)_2$	0.64	0.18	0.65	-0.01
7	6h	I	-0.46	1.12	-0.12	-0.34

Biological Activity = Learned ~~linear~~ function of properties

3D-QSAR: includes geometric/structural properties

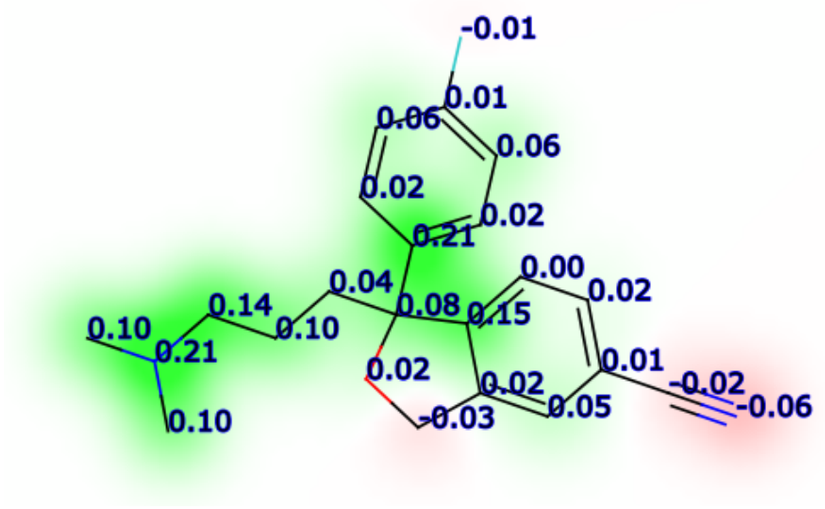
Ligand Based: QSAR

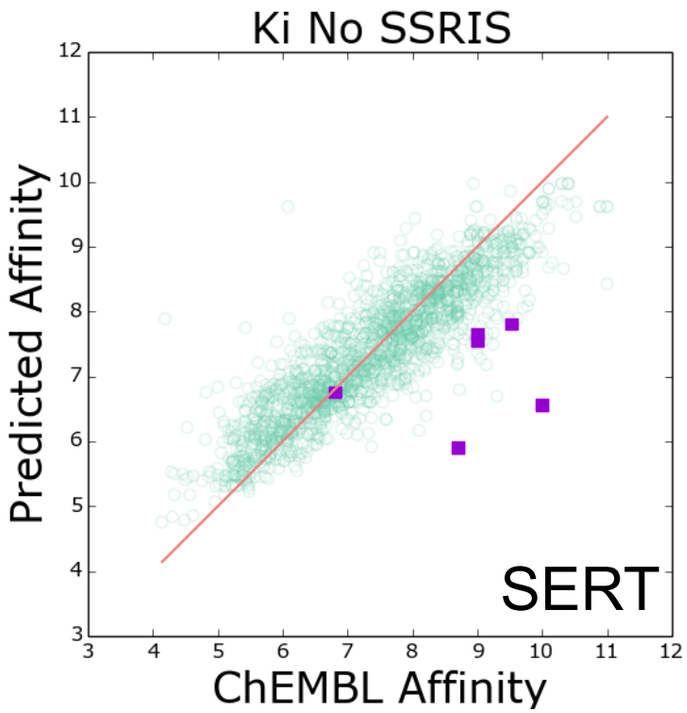
Quantitative Structure/Activity Relationships

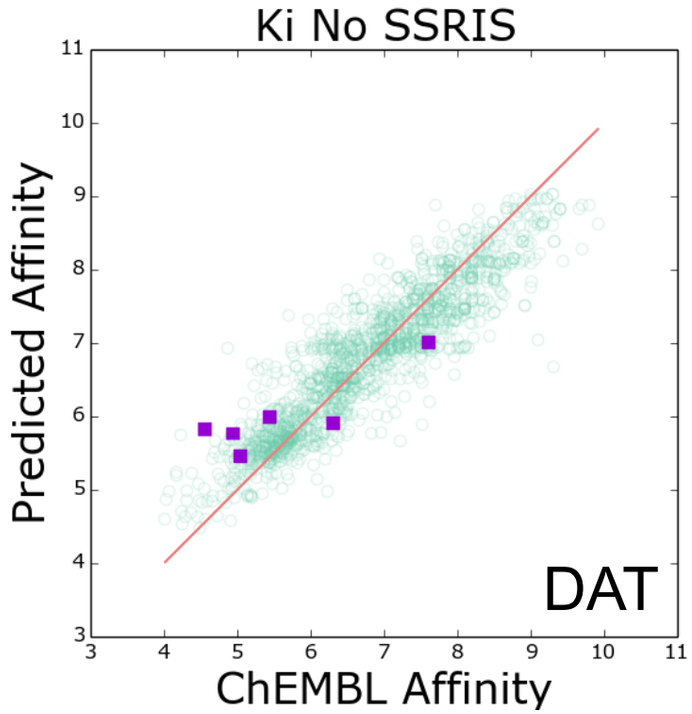


$$f(\vec{x})$$

$$f(\vec{x}) = w_1\vec{x}_1 + w_2\vec{x}_2 + w_3\vec{x}_3 + ... + b$$





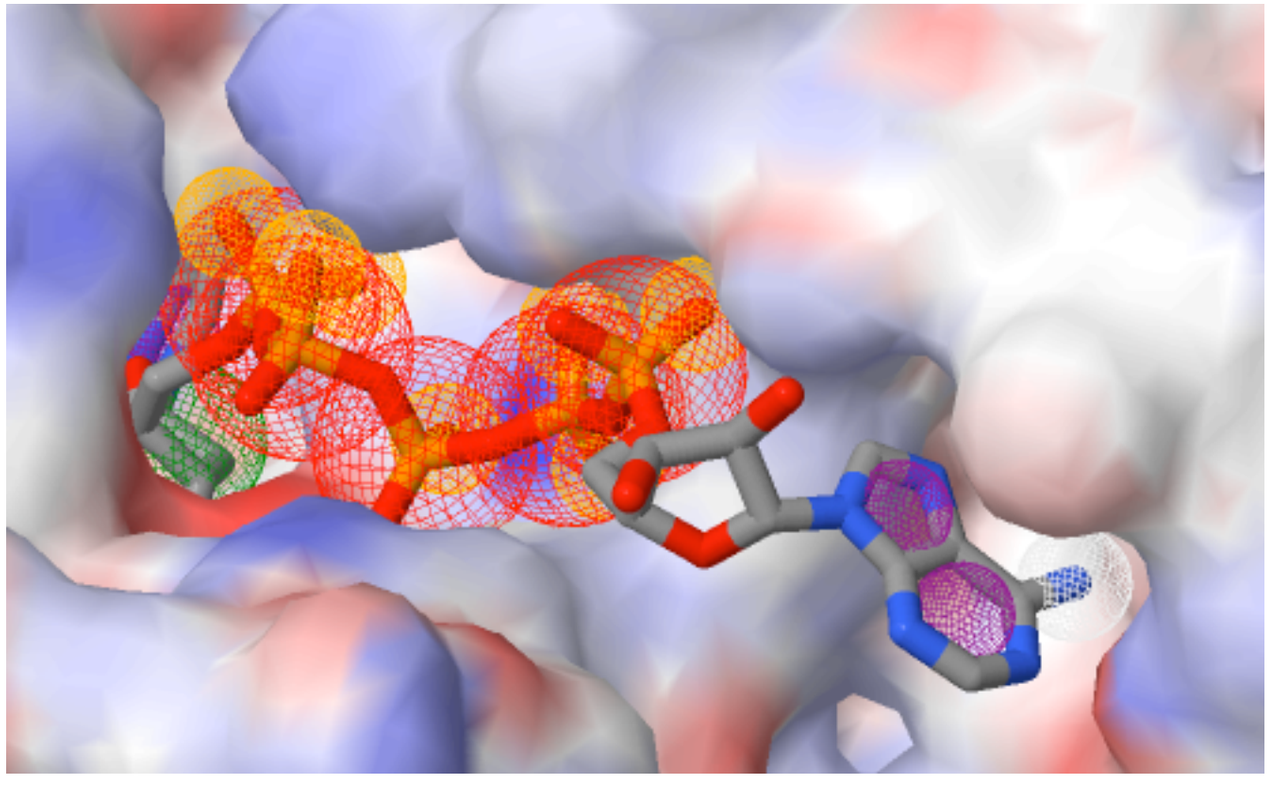


Ligand/Receptor Based: Pharmacophore

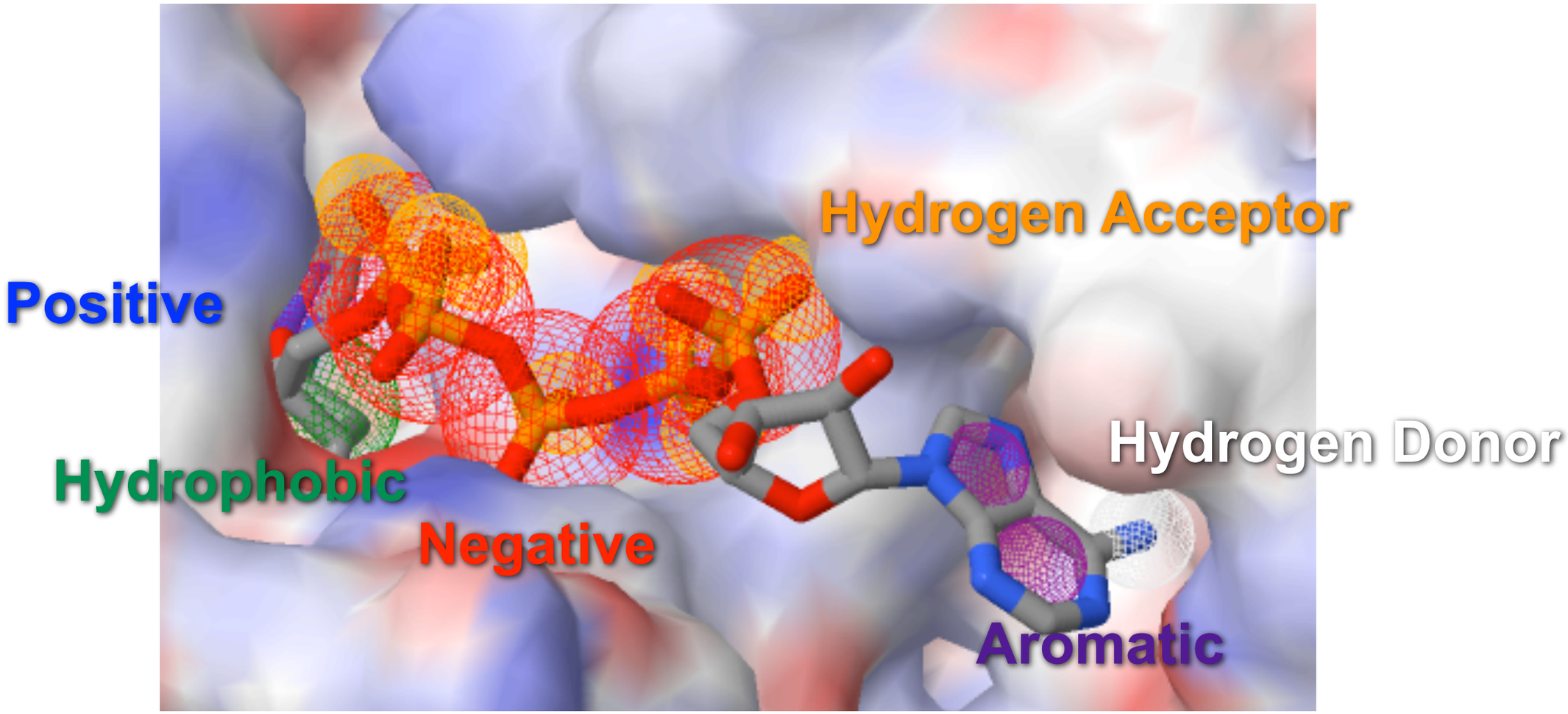
Pharmacophore:

IUPAC: The ensemble of steric and electronic features that is necessary to ensure the optimal supra-molecular interactions with a specific biological target structure and to trigger (or to block) its biological response.

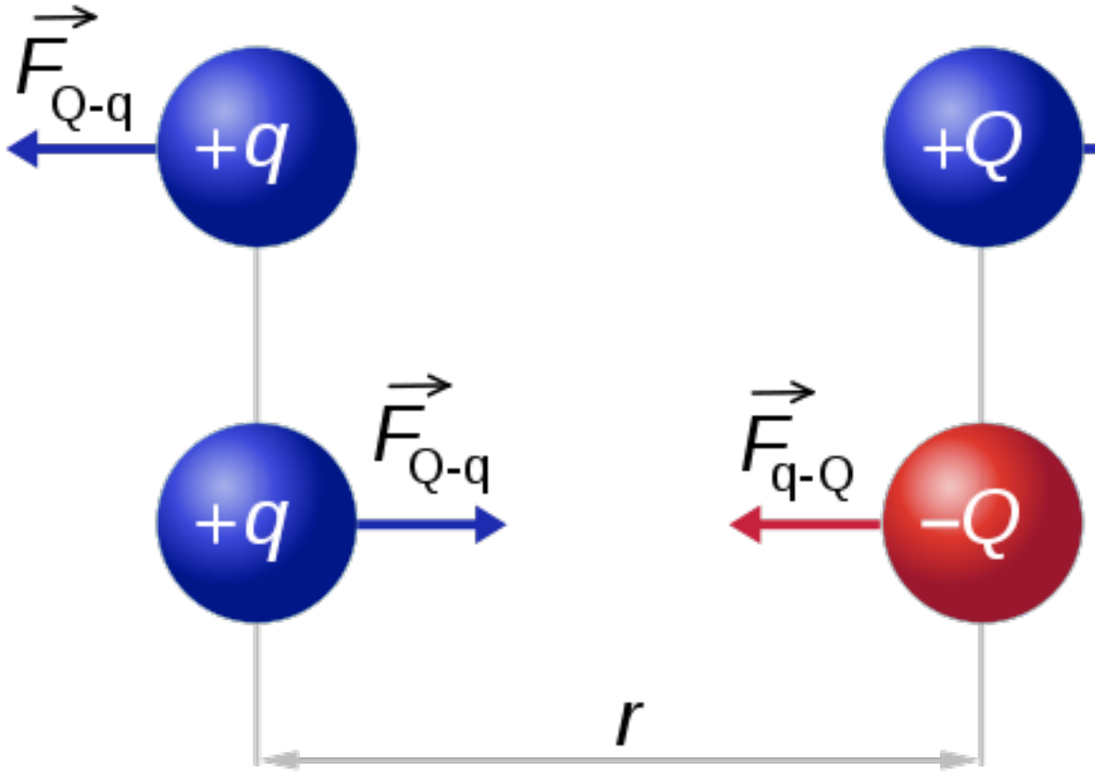
Common Features:
aromatic ring
hydrophobic area
positive ionizable
negative ionizable
hydrogen bond donor
hydrogen bond acceptor



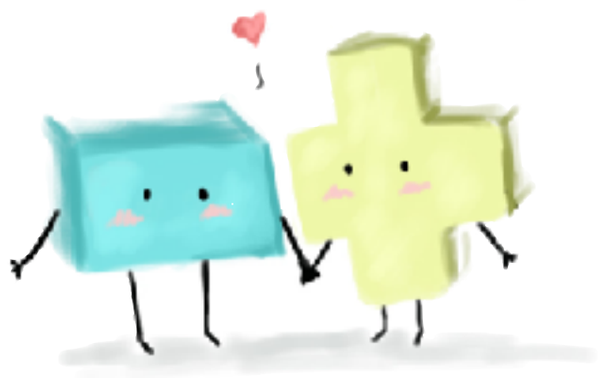
Pharmacophore Features



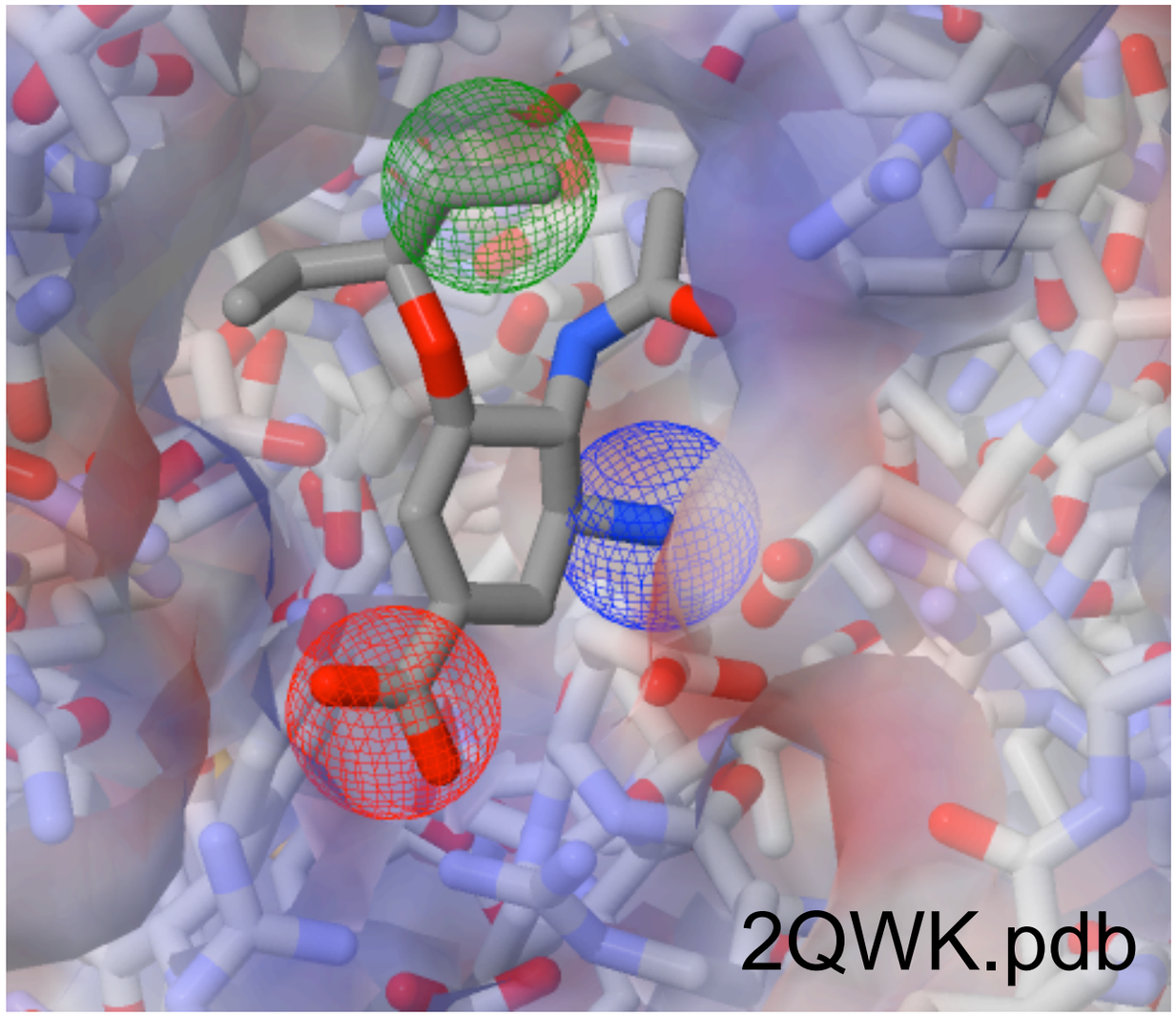
Charge-Charge

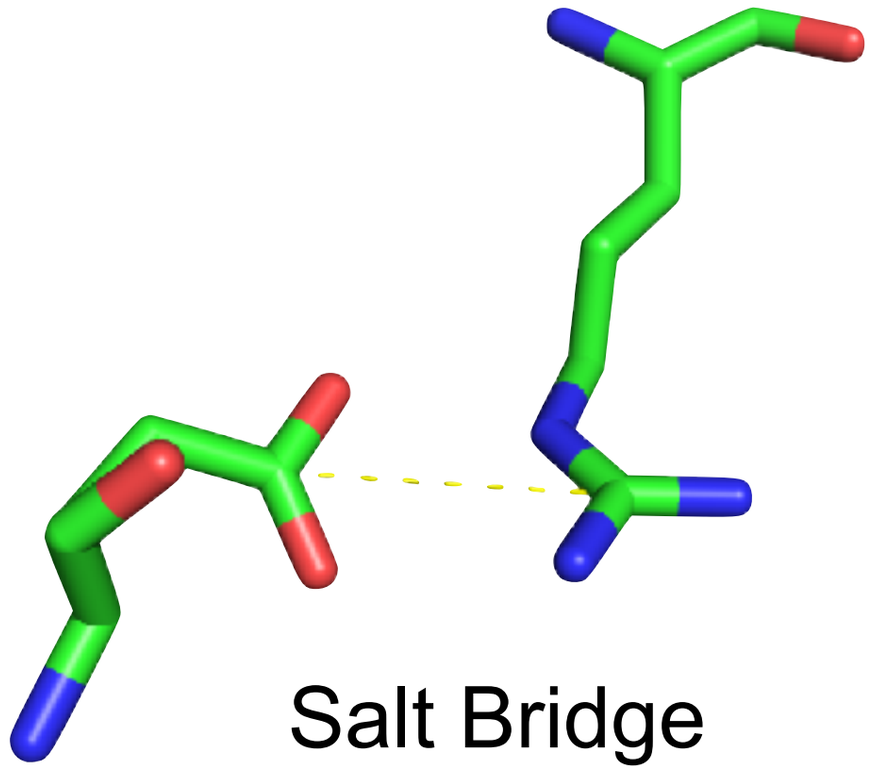


$$|\vec{F}_{Q\text{-}q}| = |\vec{F}_{q\text{-}Q}| = k\,\frac{|q| \times |Q|}{r^2}$$



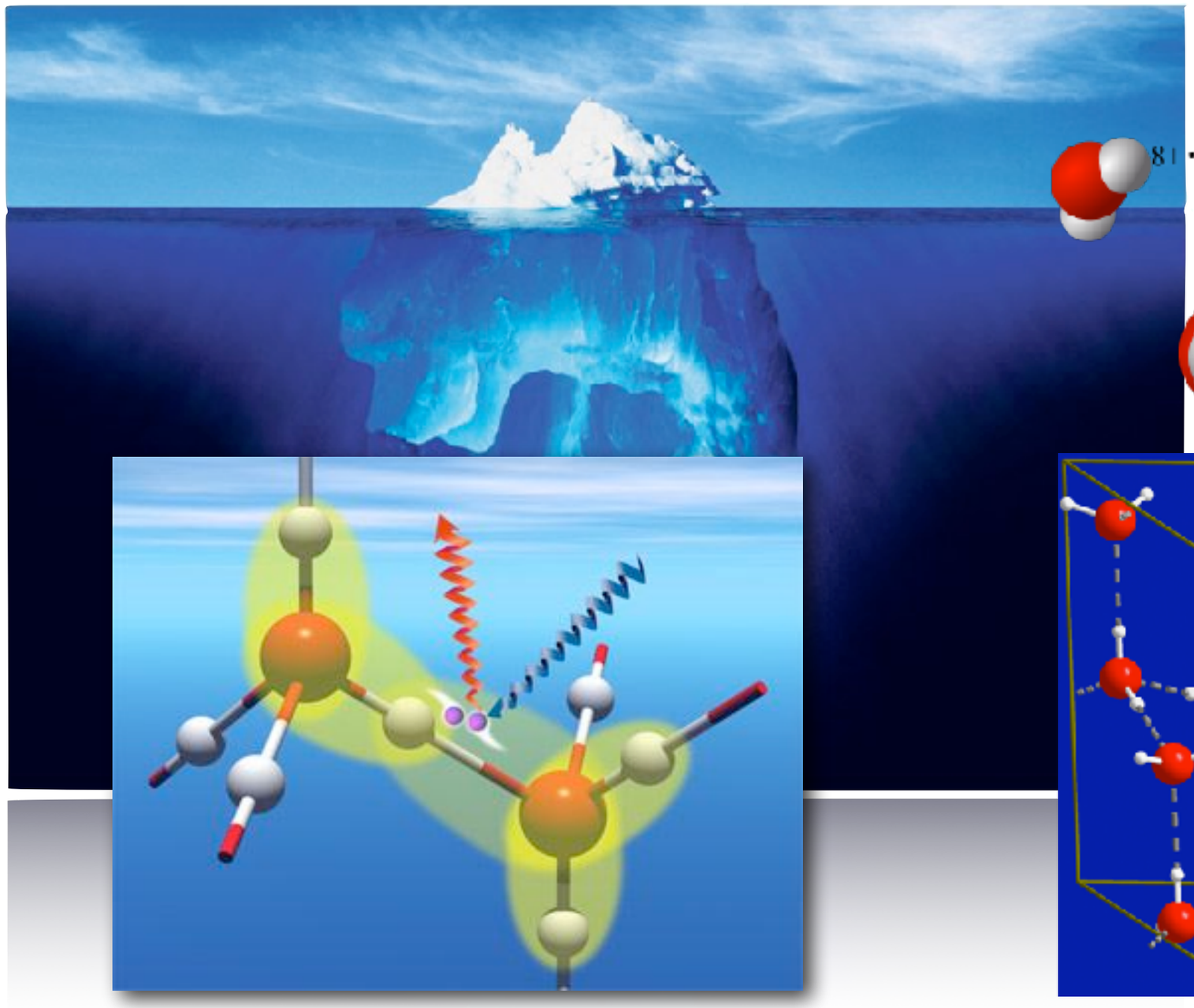
Charge-Charge

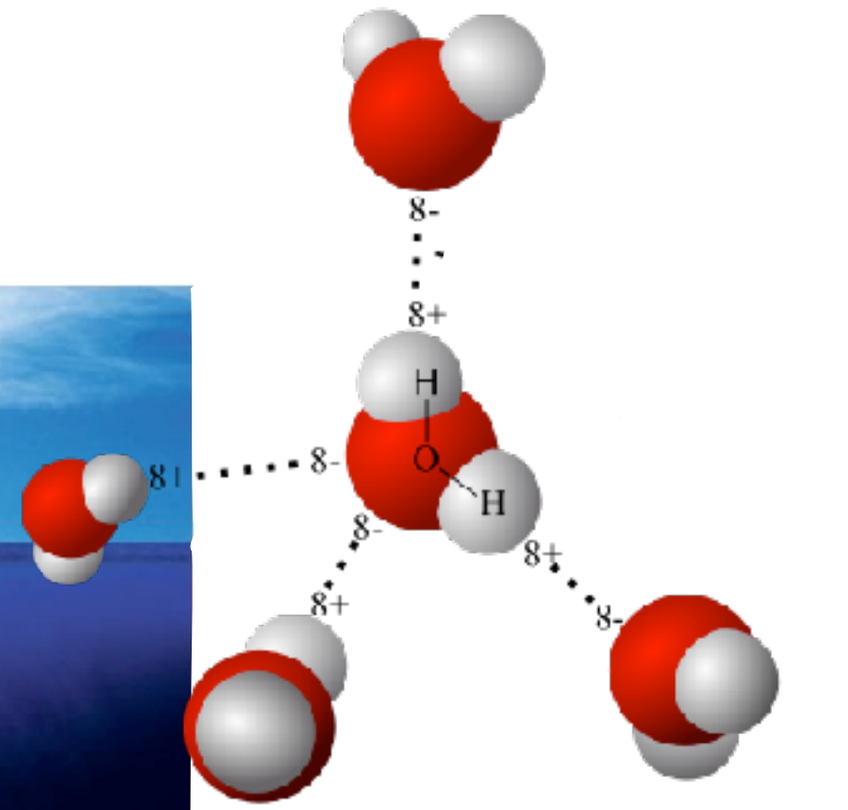


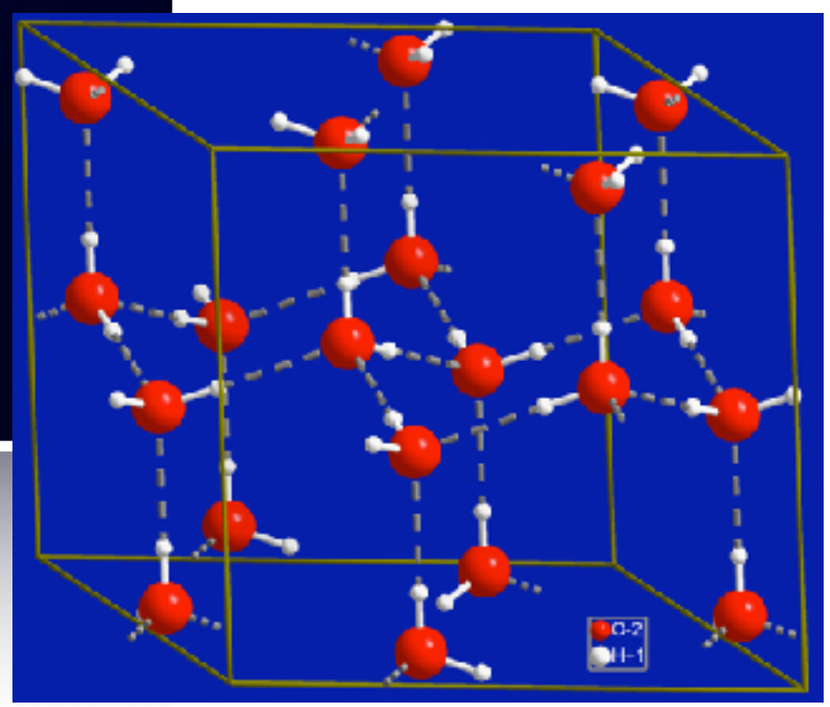


Inhibitor of the influenza virus neuraminidase (antiviral agent)

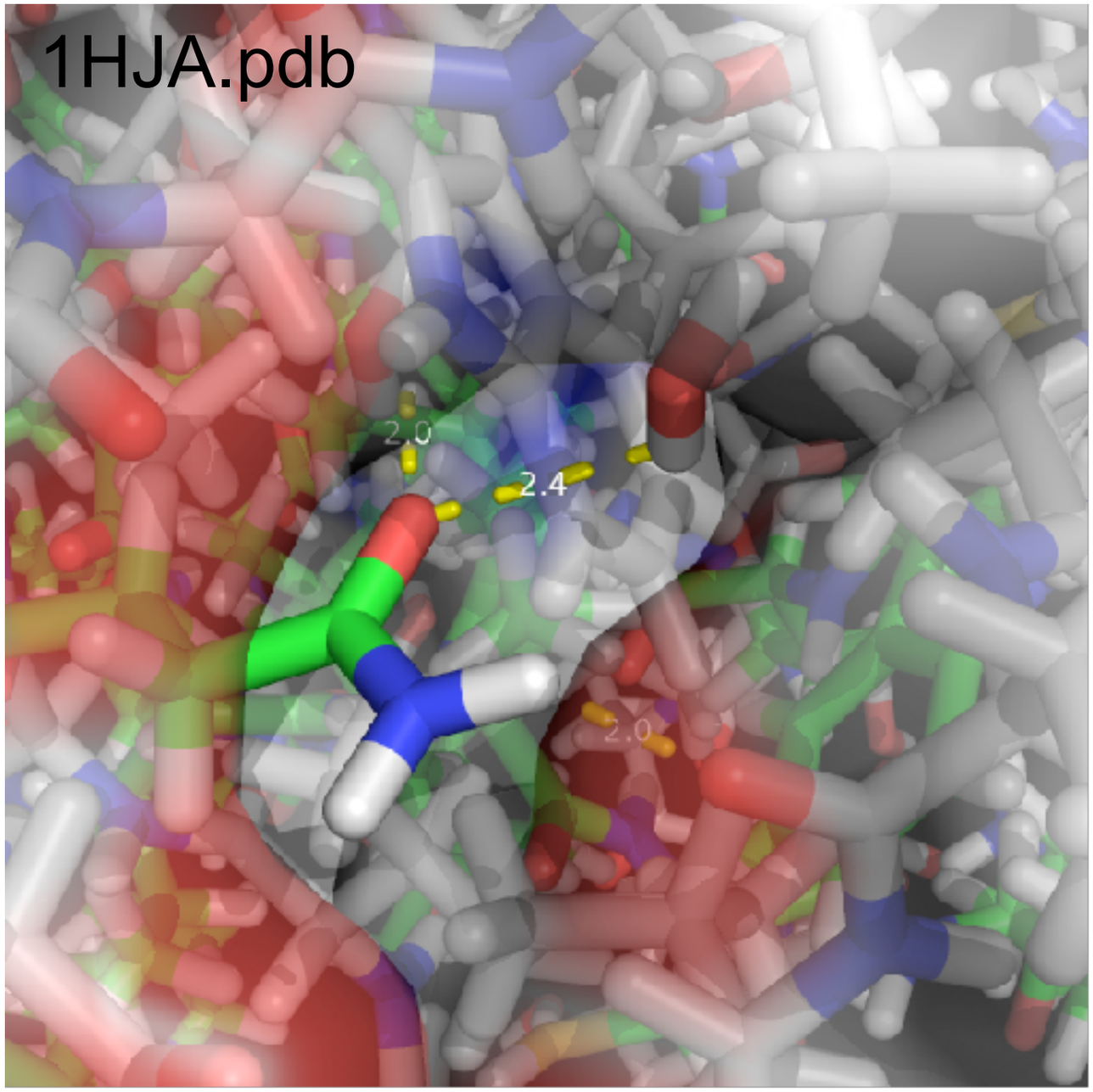
Hydrogen Bond



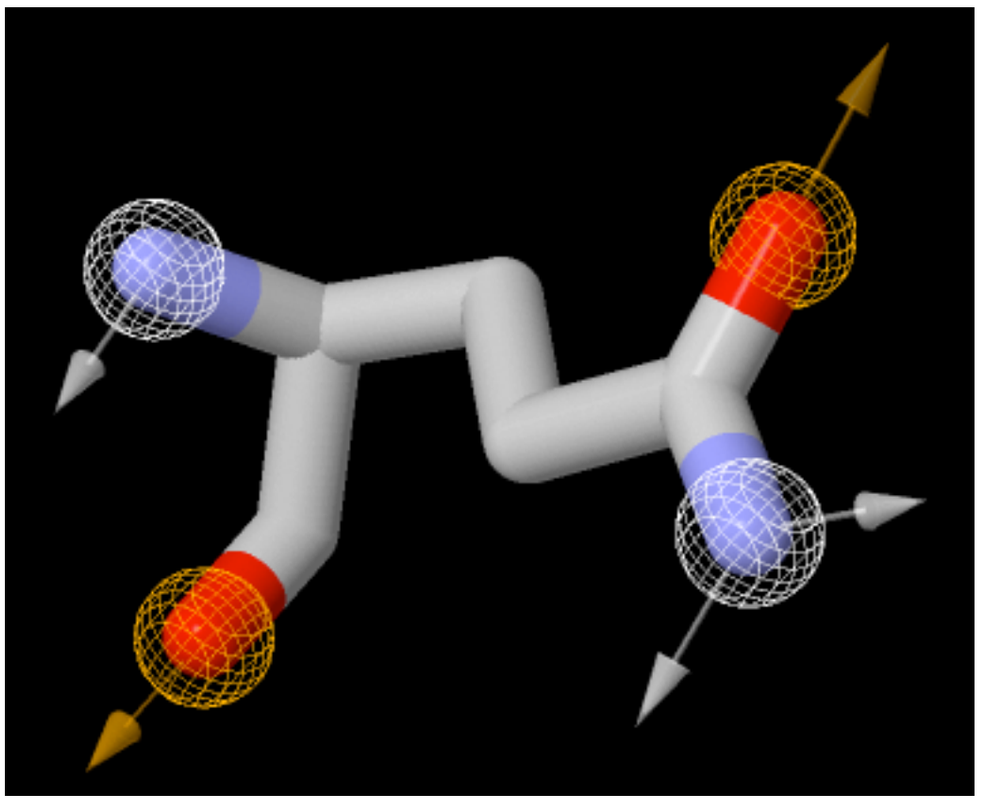




Hydrogen Bond



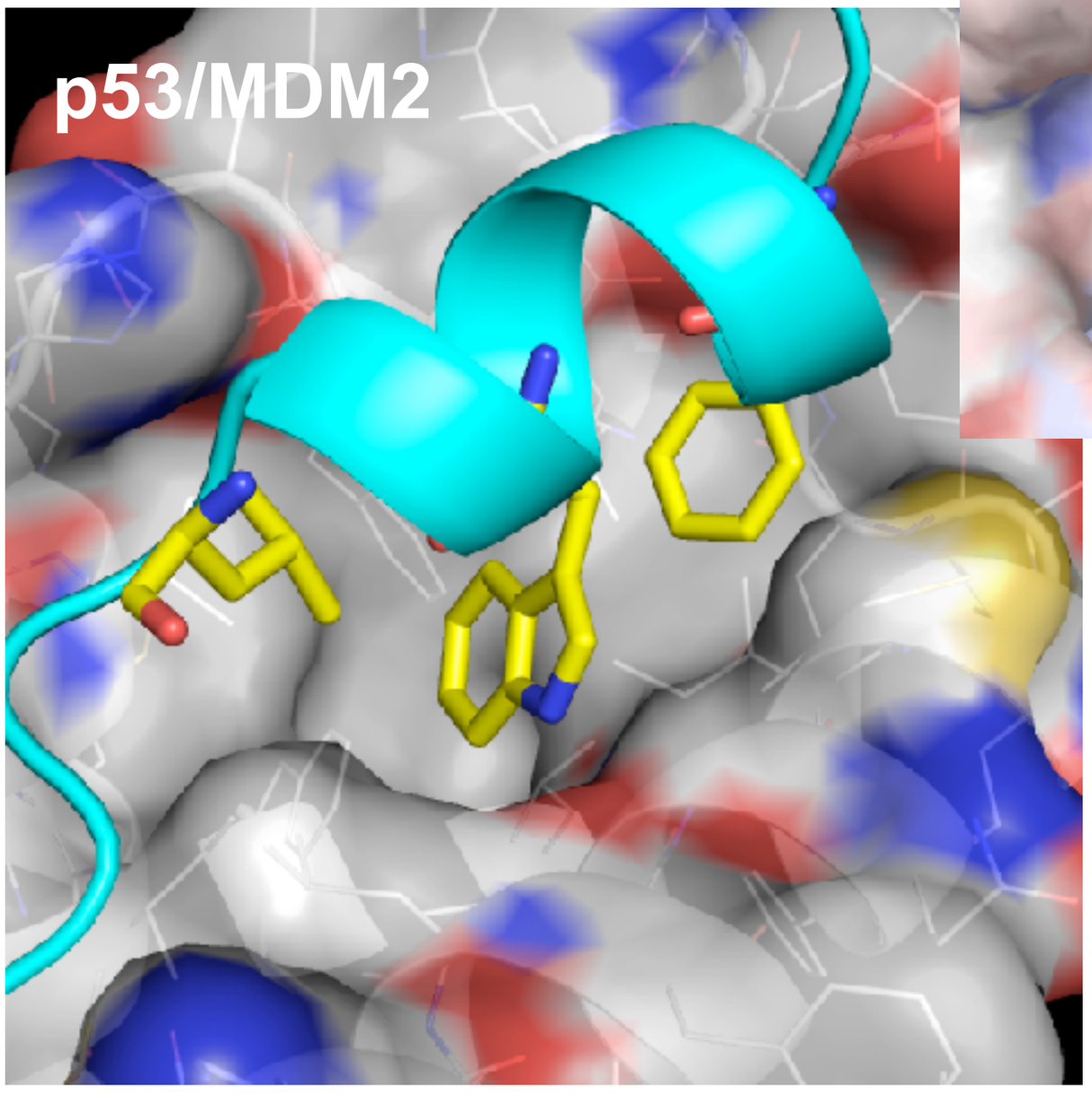
Distance:
D-A: 2.5Å – 3.5Å (4.0Å?)
H-A: 1.5Å – 2.5Å
Angle:
Depends on context

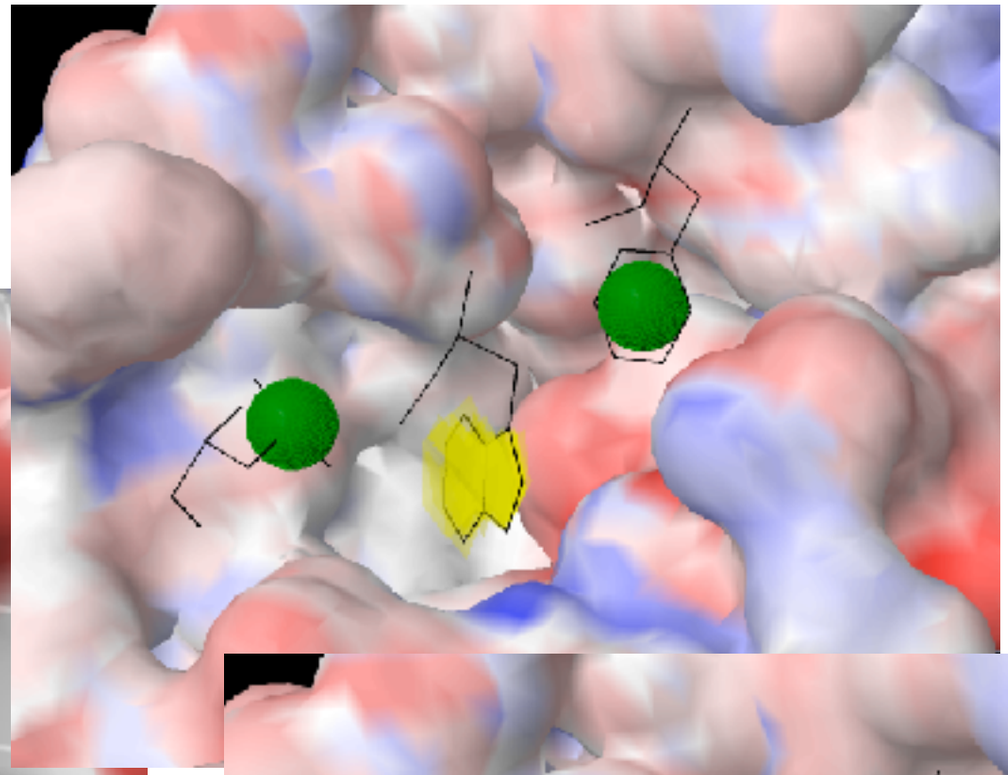


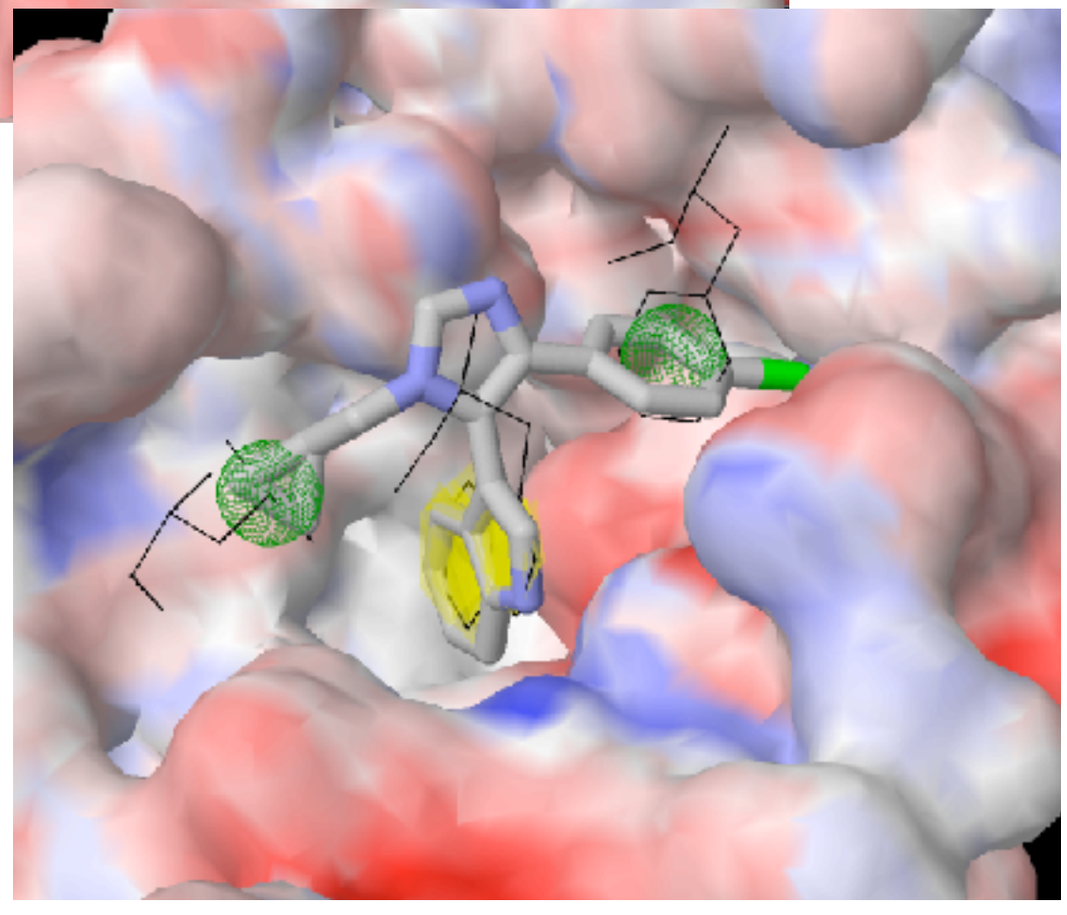
Turkey Ovomucoid Inhibitor

Hydrophobic

Hydrophobic

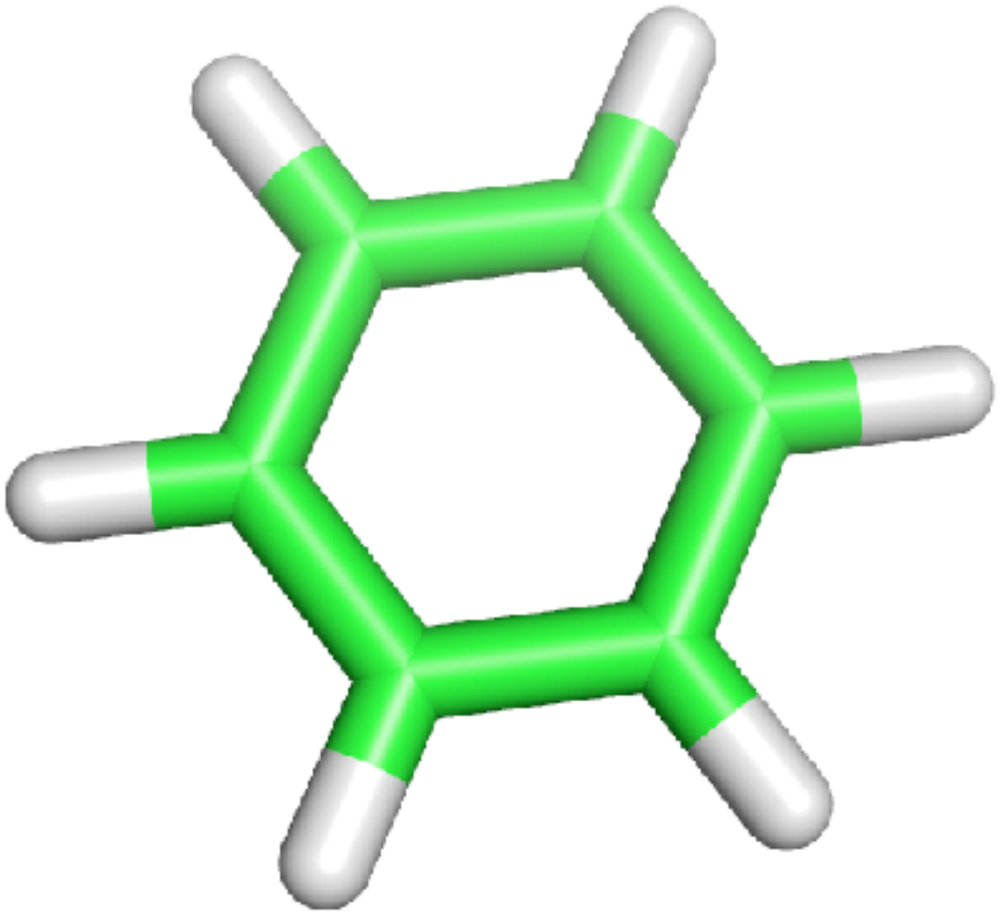


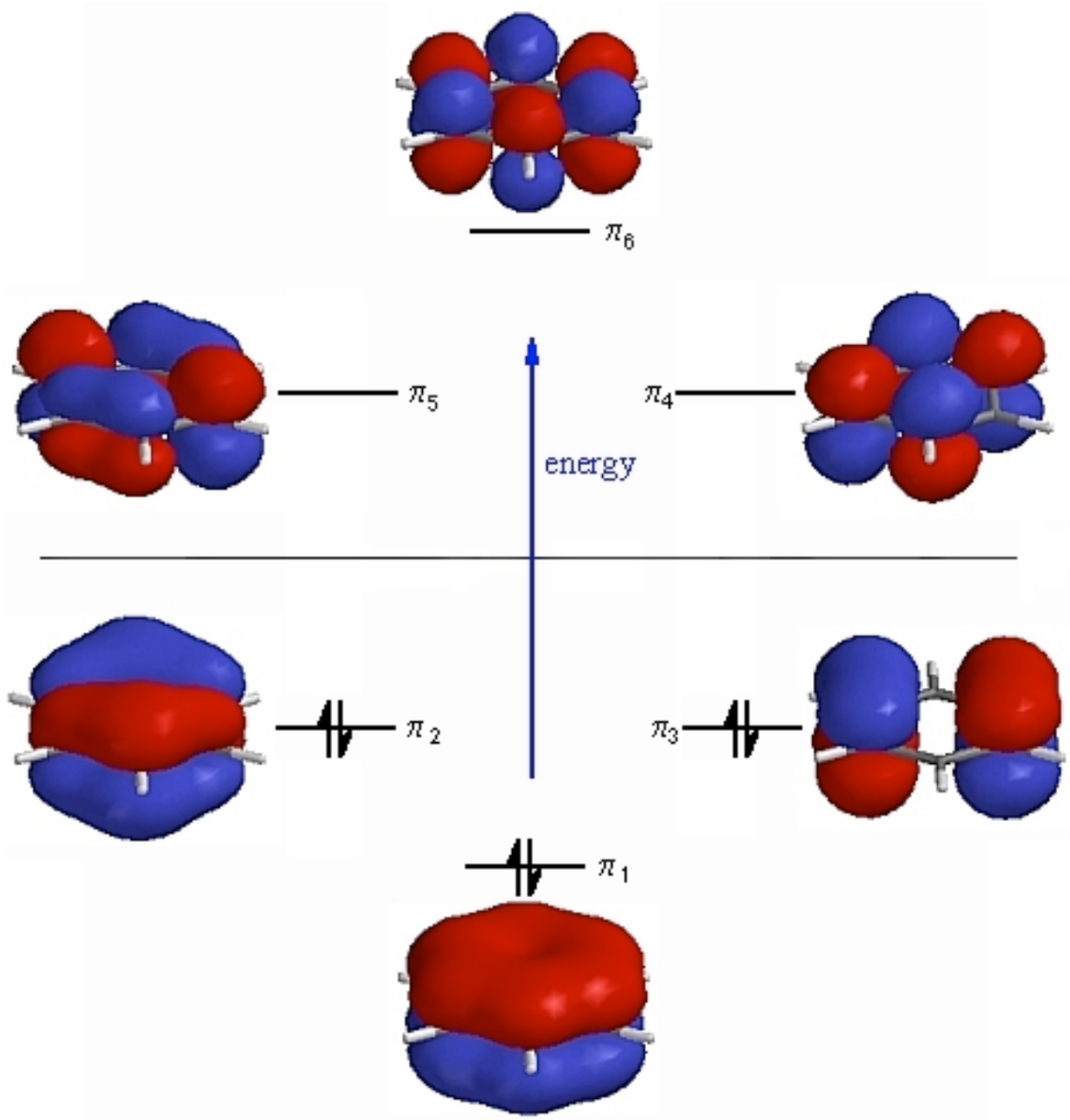




MDM2 (over expressed in >50% of cancers) down-regulates p53 (guardian of the genome)

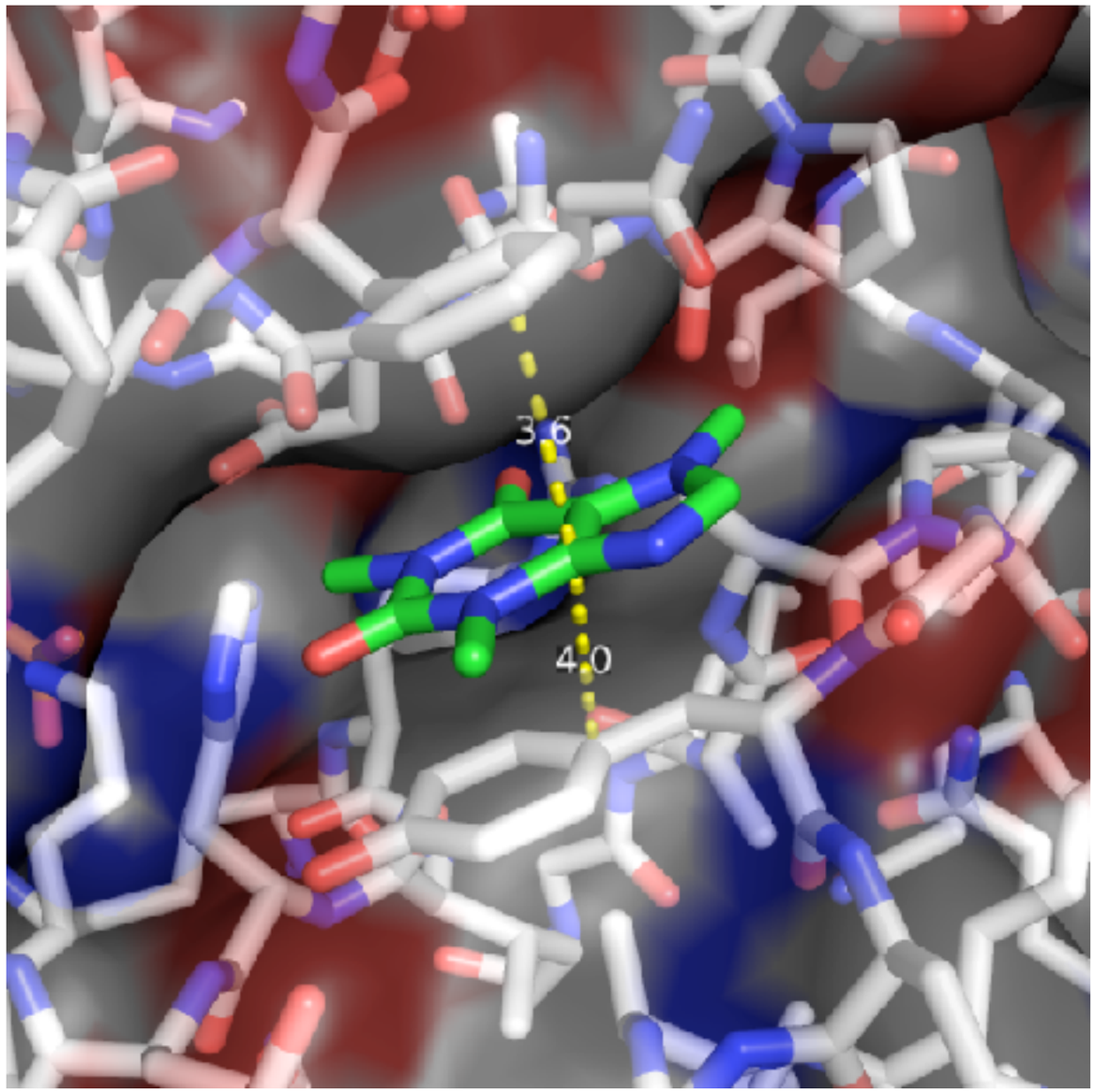
Aromatic



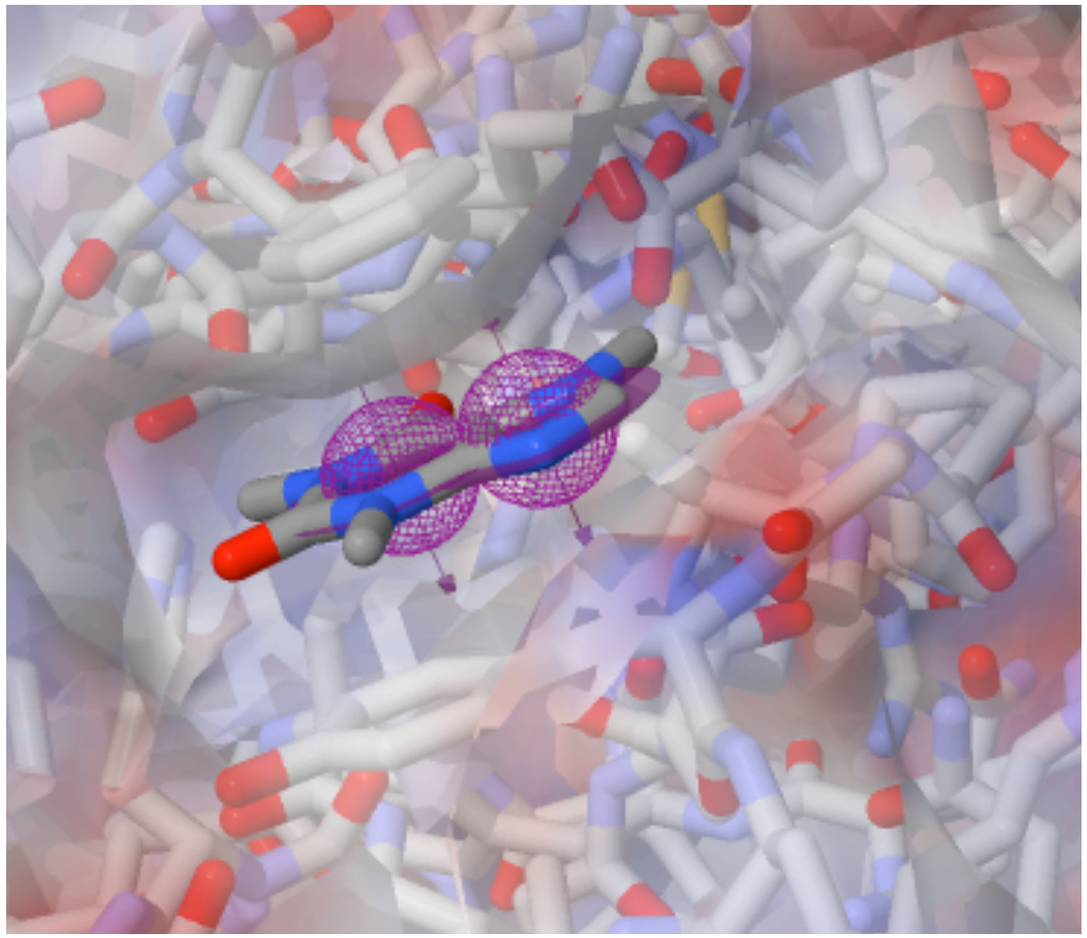


http://www.chem.ucalgary.ca/courses/351/Carey5th/Ch11/benzene-mo.jpg

Aromatic



Rings offset
Interplanar distance: 3.3-3.8Å



Human liver glycogen phosphorylase a complexed with caffeine

Kinds of Virtual Screening

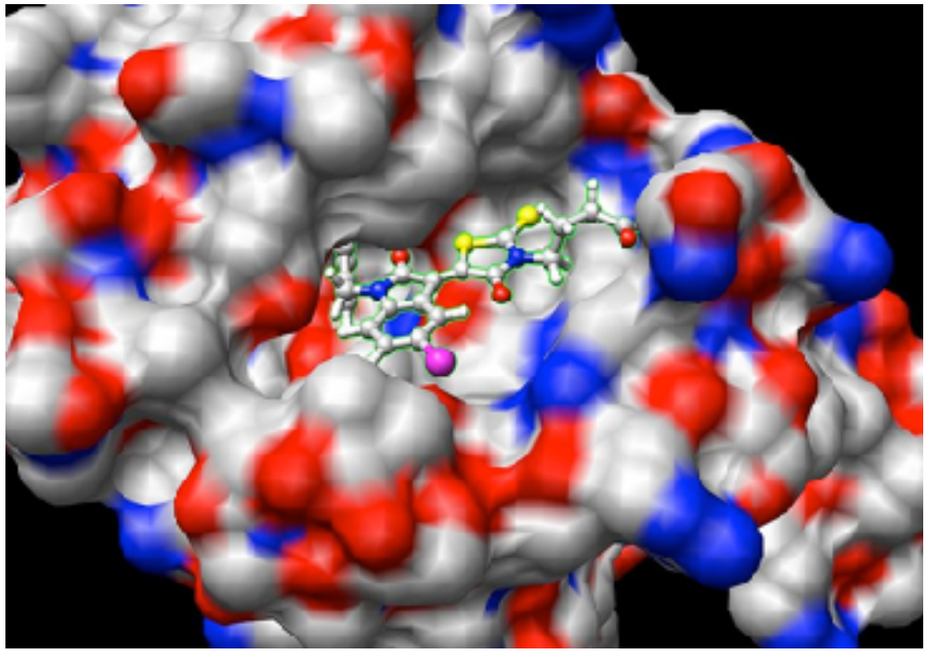
ADMET

Ligand Based

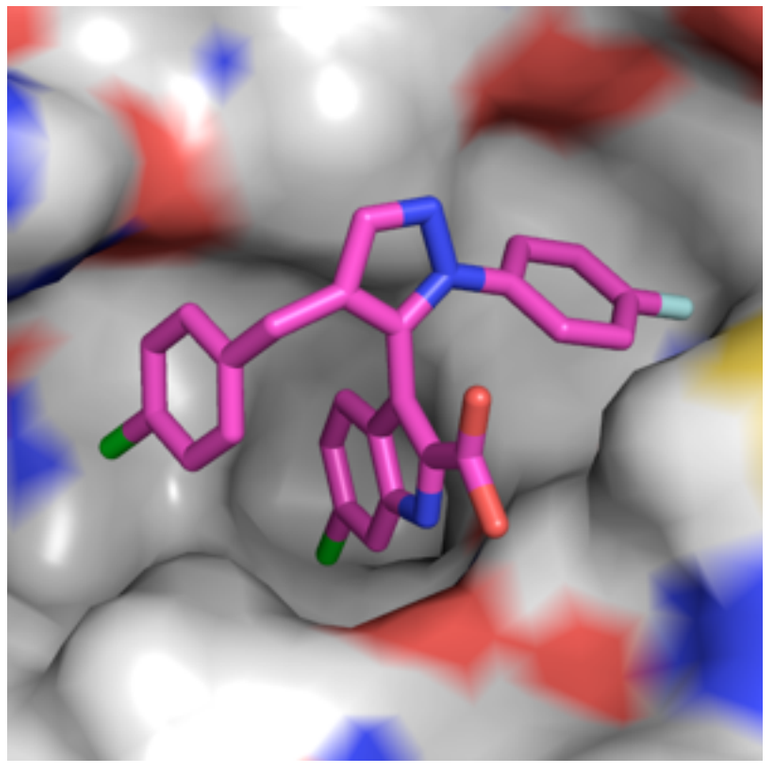
- similarity to known binder
- QSAR
- pharmacophore

Receptor Based

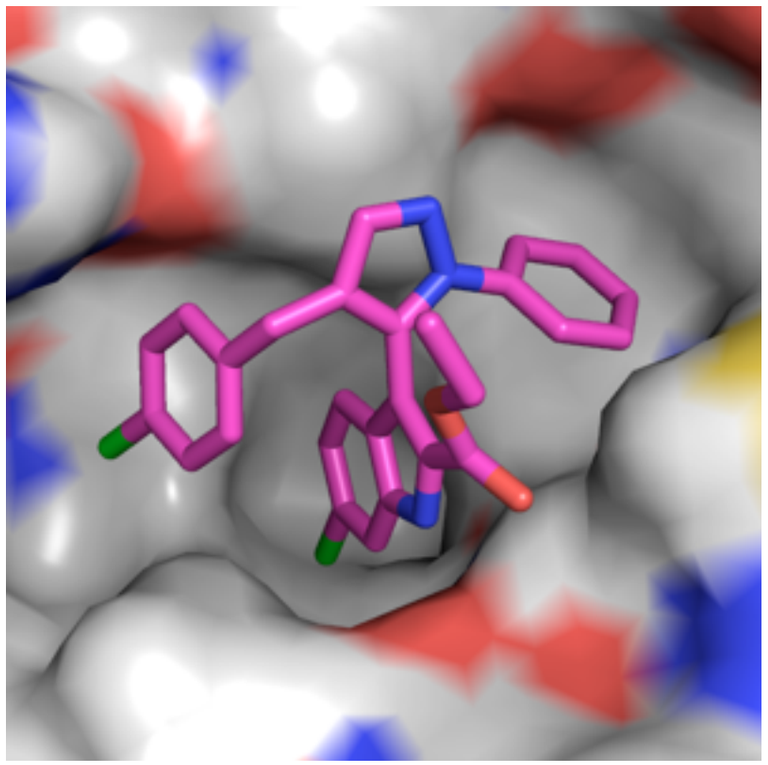
- dock and score



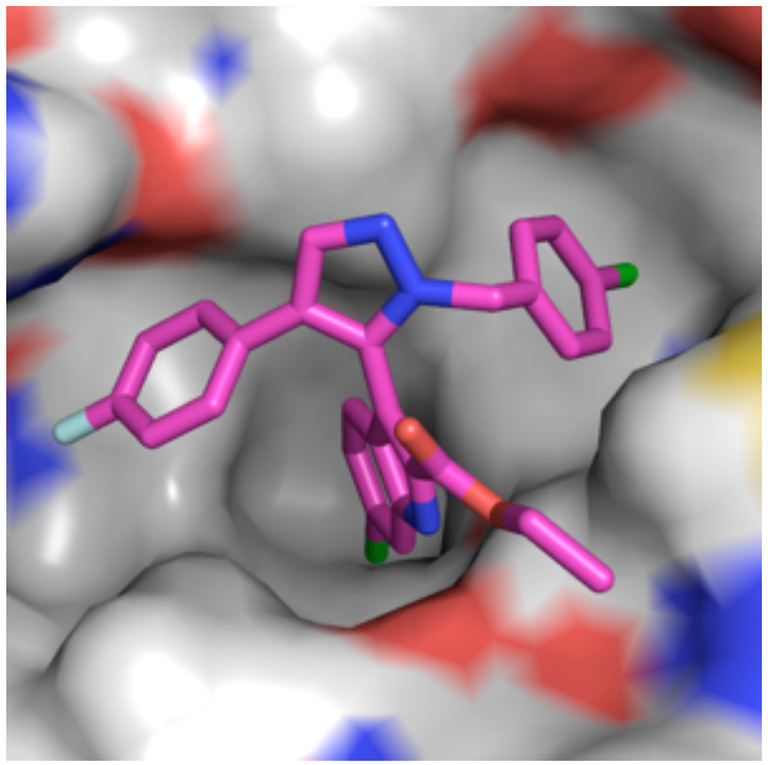
Pharmacophores Aren't Enough



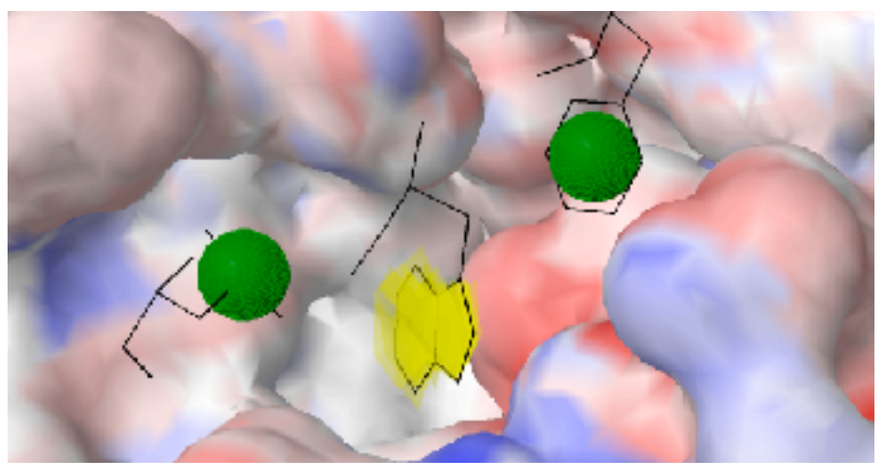
.2μM



50μM



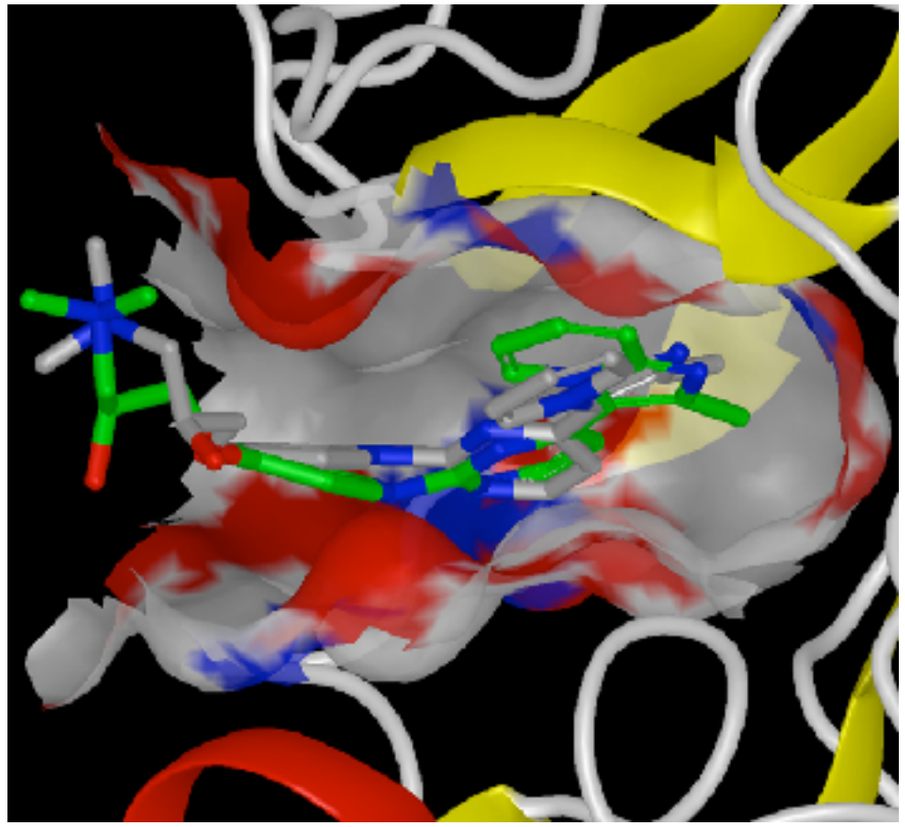
n.i.



Docking

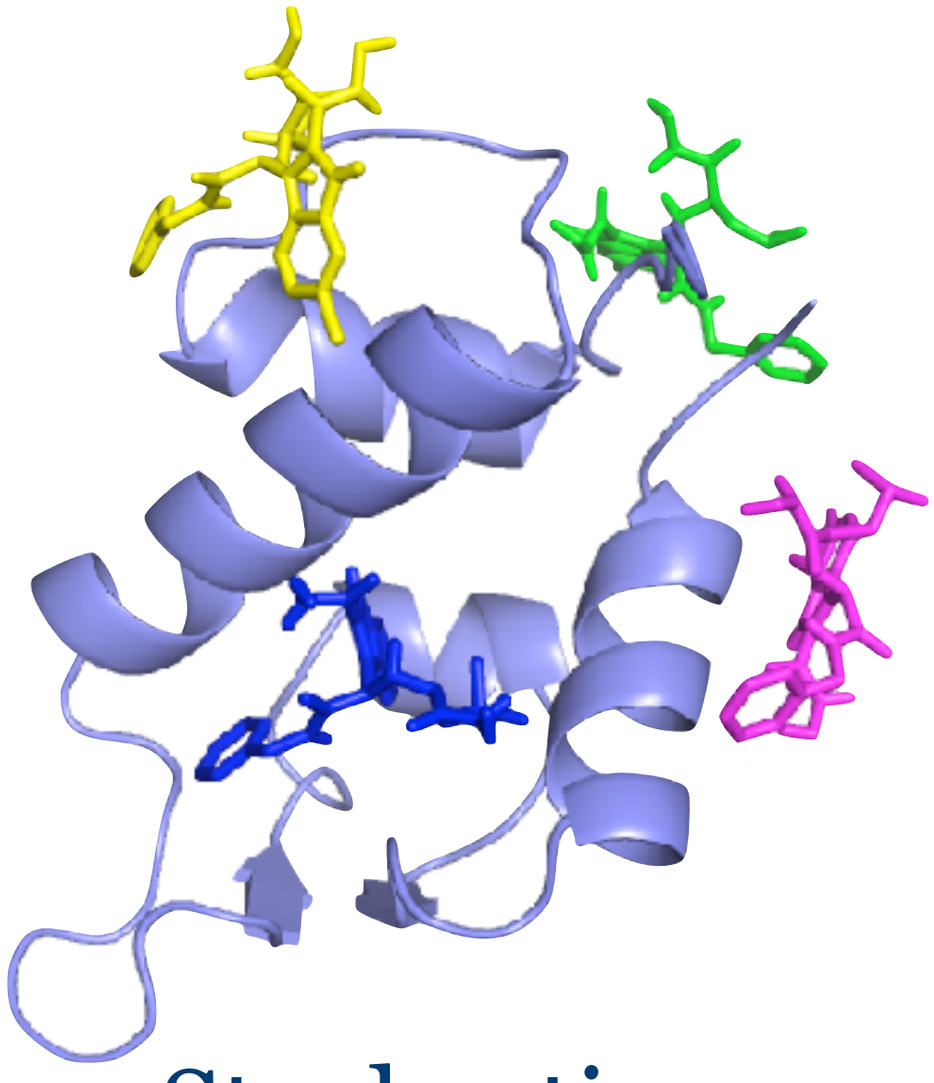
Determine the **conformation** and **pose** of a ligand at a docking site

Challenge is to find conformation and pose with the best **score**

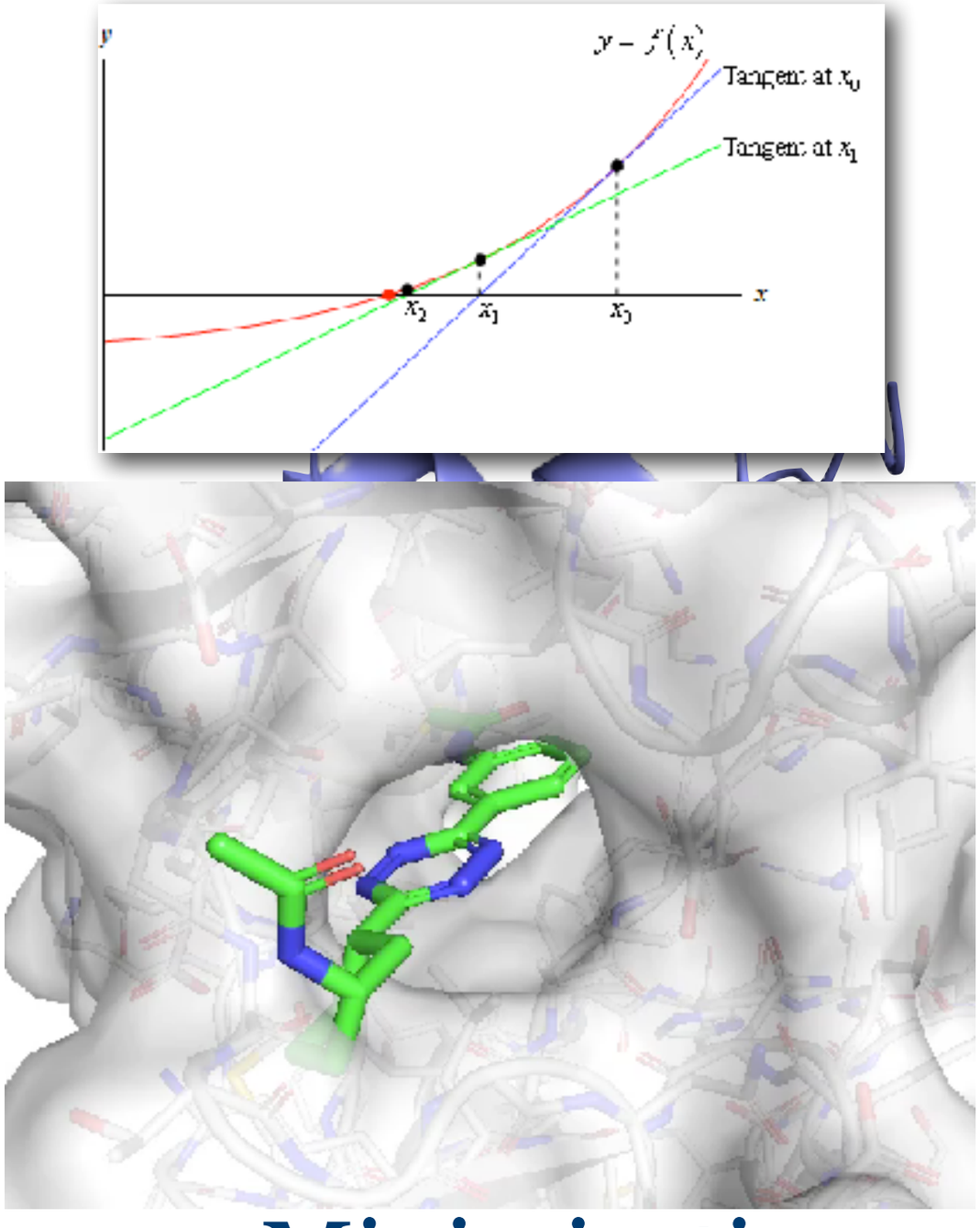


Two Phase Docking

1. Global Pose Estimation



Stochastic



Minimization

Scoring Goals

Affinity Prediction

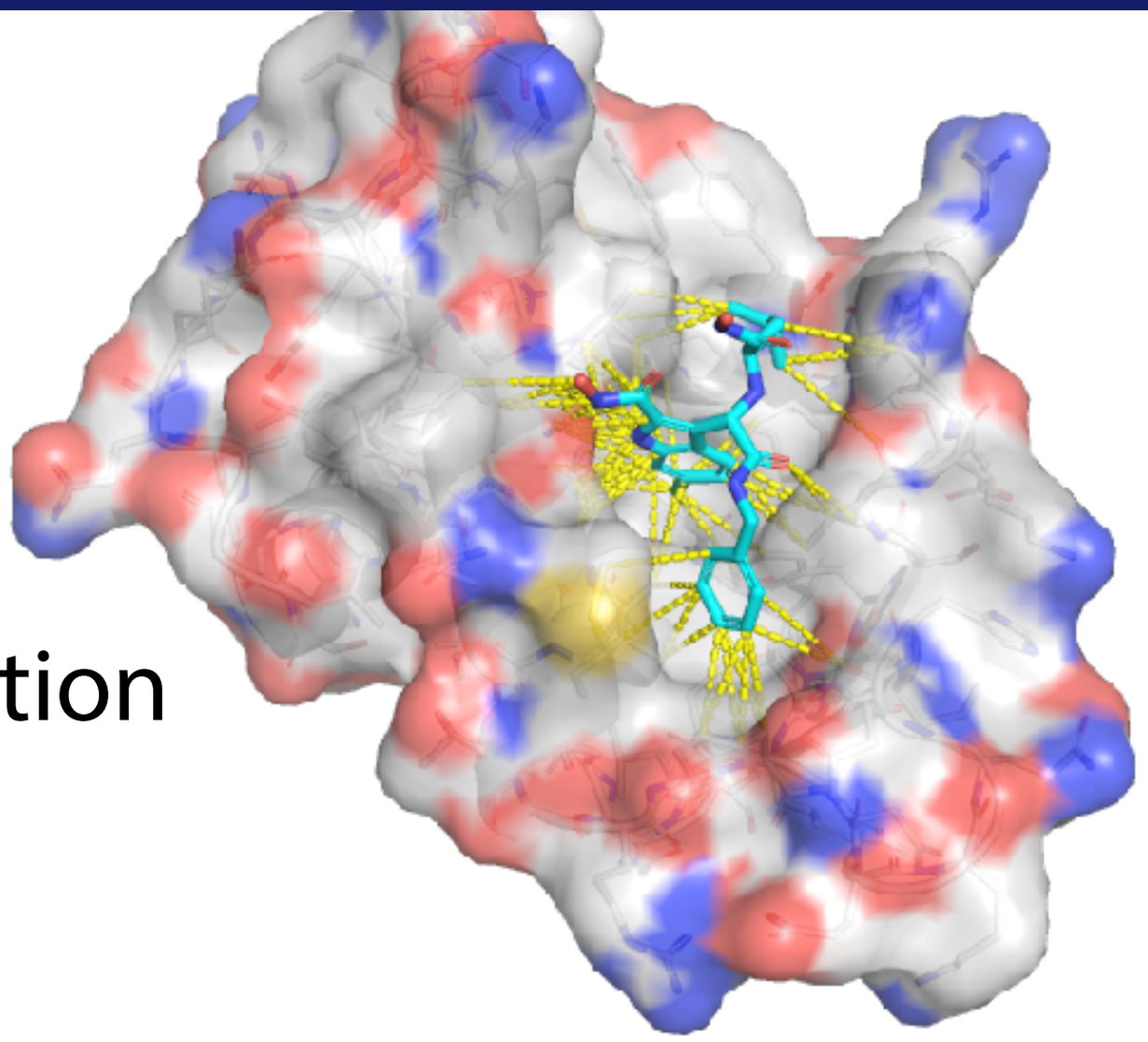
-how well does it bind?

Inactive/Active Discrimination

-does it bind?

Pose Prediction

-how does it bind?



Speed

Approximations:
Rigid or semi-rigid receptor
Implicit water model

Scoring Types

Force-field based

inter- and intra- molecular forces
van der Waals, electrostatic, torsional

Empirical

parameterized function is fit to binding energy data

Knowledge based

scoring function based on data, not physical principles

Consensus

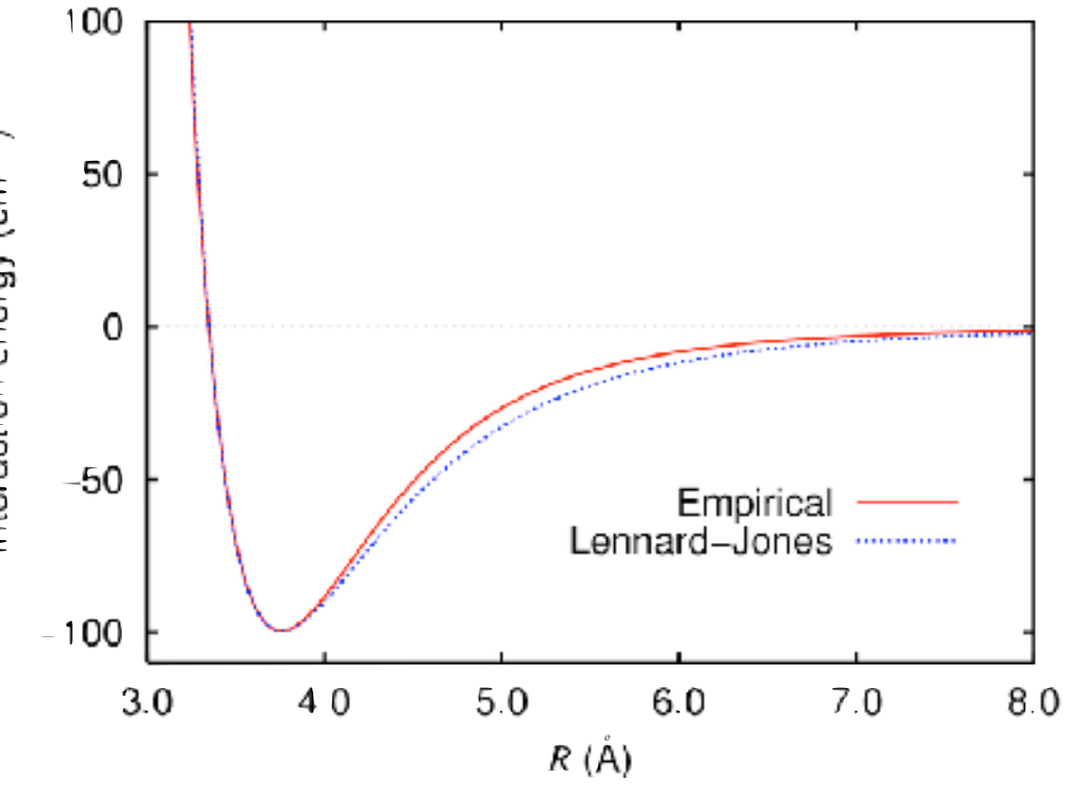
combine multiple scoring functions

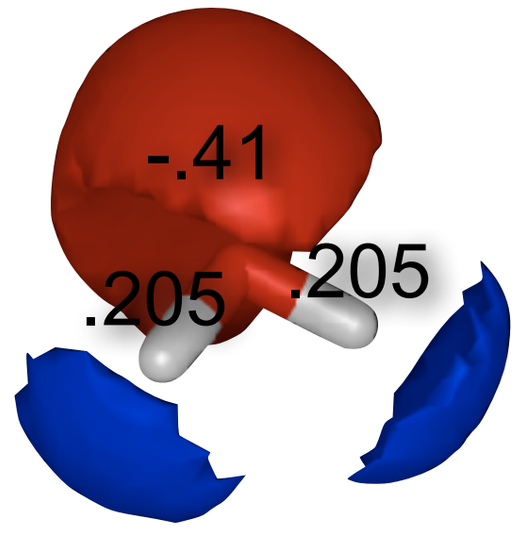
Force Field: Dock 4.0

Coulomb's Law
q: partial charges
D: dielectrict constant

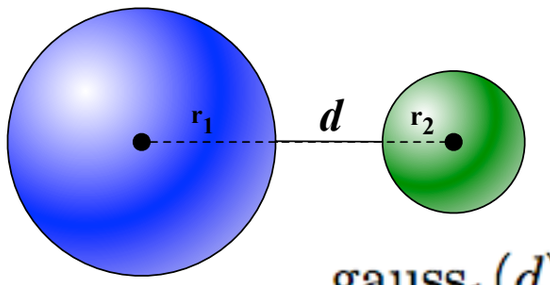
$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^{a}} - \frac{B_{ij}}{r_{ij}^{b}} + 332 \frac{q_i q_j}{D r_{ij}} \right)$$

van der Waals
a = 12, b = 6
Lennard-Jones potential





Empirical: AutoDock Vina



$$\text{gauss}_1(d) = w_{\text{guass}_1} e^{-(d/0.5)^2}$$

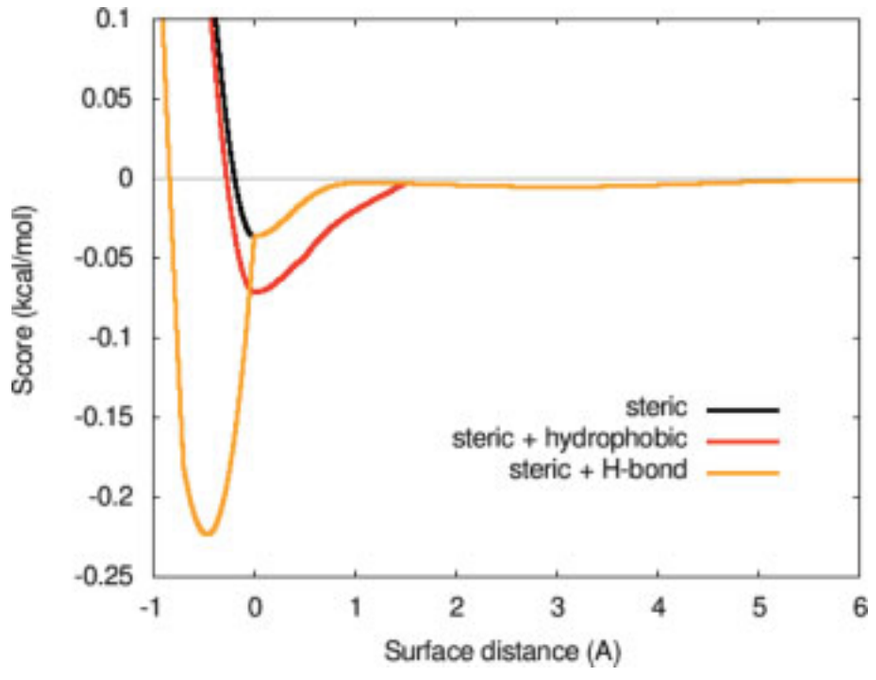
$$\text{gauss}_2(d) = w_{\text{guass}_2} e^{-((d-3)/2)^2}$$

$$\text{repulsion}(d) = \begin{cases} w_{\text{repulsion}} d^2 & d < 0 \\ 0 & d \geq 0 \end{cases}$$

$$\text{hydrophobic}(d) = \begin{cases} w_{\text{hydrophobic}} & d < 0.5 \\ 0 & d > 1.5 \\ w_{\text{hydrophobic}}(1.5 - d) & otherwise \end{cases}$$

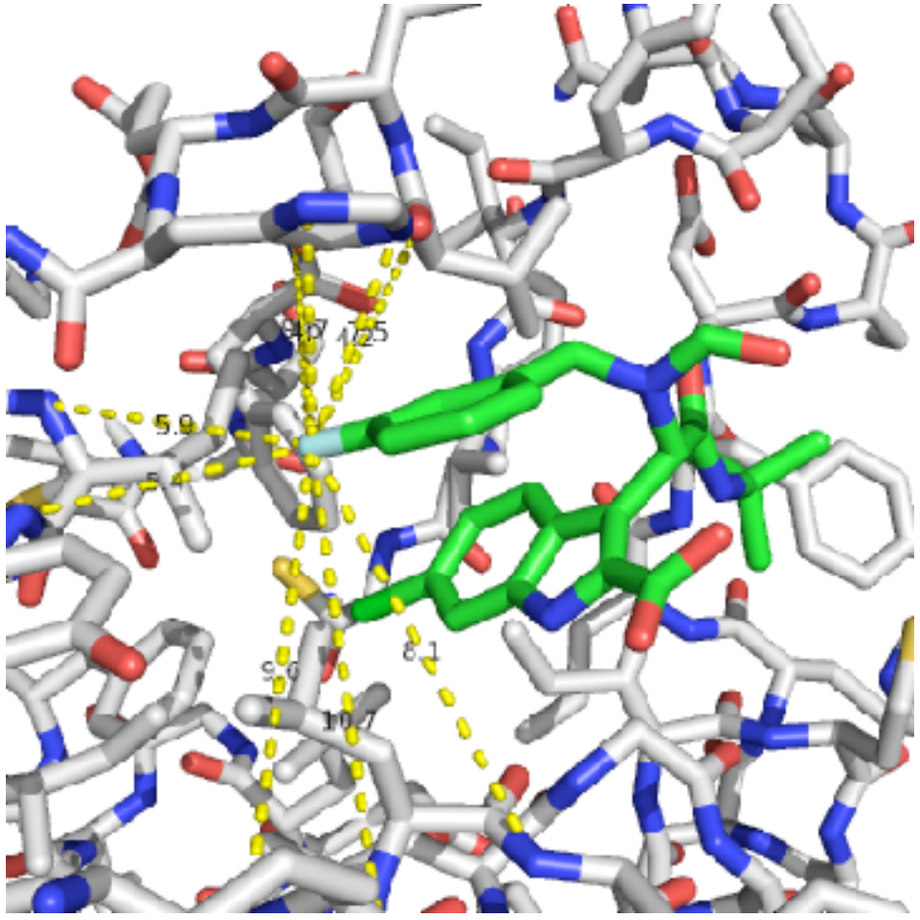
$$\text{hbond}(d) = \begin{cases} w_{\text{hbond}} & d < -0.7 \\ 0 & d > 0 \\ w_{\text{hbond}}(-\frac{10}{7} d) & otherwise \end{cases}$$

Weight	Term
−0.0356	gauss_1
−0.00516	gauss_2
0.840	Repulsion
−0.0351	Hydrophobic
−0.587	Hydrogen bonding
0.0585	N_{rot}



Knowledge Based: RF-Score

Pairwise Distance Counts (<12Å)



Protein

Ligand

	C	N	O	S
C				
N				
O				
S				
P				
F		9		
Cl				
Br				
I				

Random Forest

BIOINFORMATICS **ORIGINAL PAPER**

Structural bioinformatics

A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking

Pedro J. Ballester and John B. O. Mitchell

Can we do better?

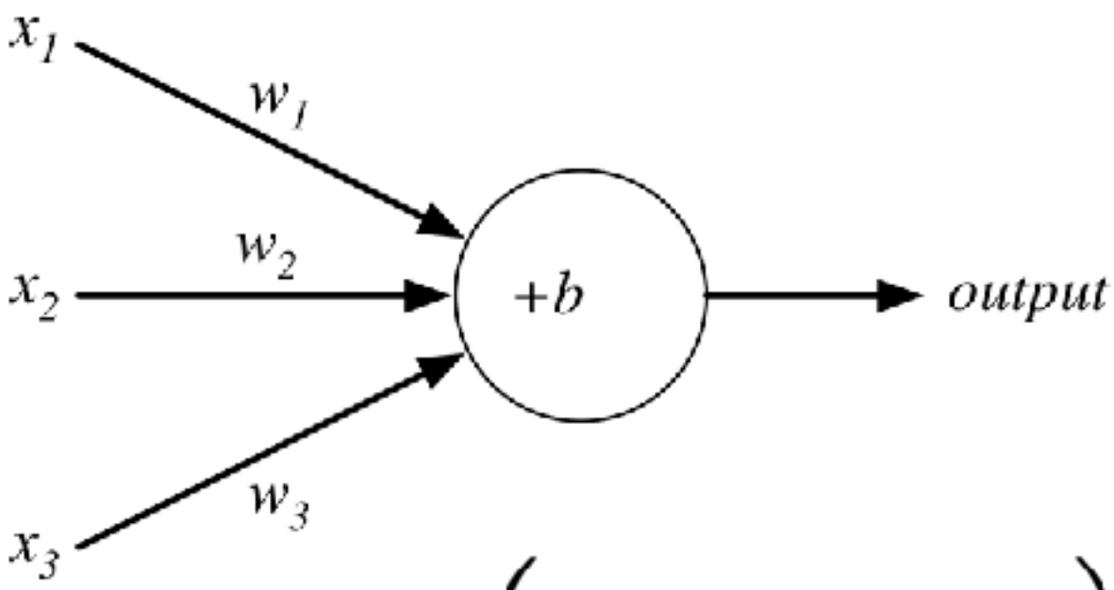
Accurate pose prediction, binding discrimination, **and** affinity prediction without sacrificing performance?

Key Idea: Leverage "big data"

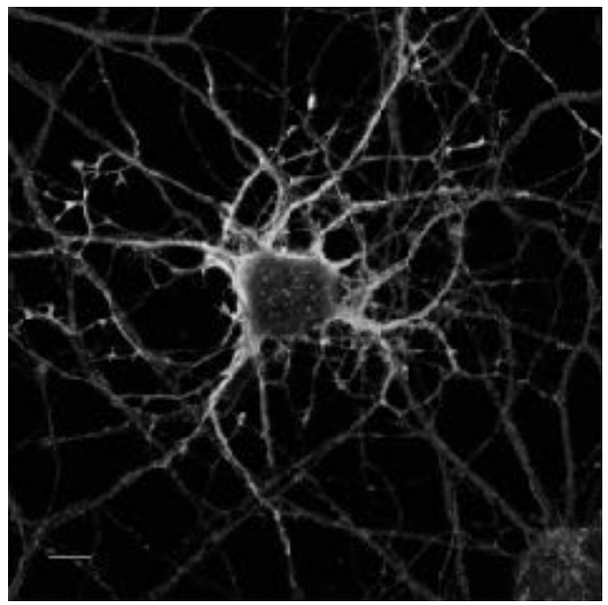
- 231,655,275 bioactivities in PubChem
- 125,526 structures in the PDB
- 16,179 annotated complexes in PDBbind

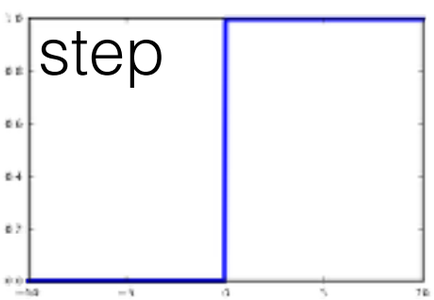
Machine Learning

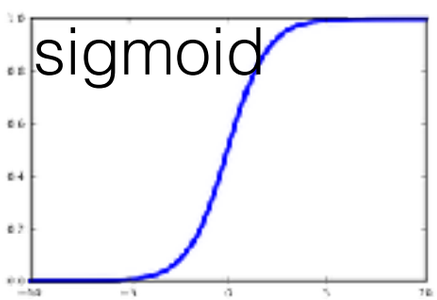
Neural Networks

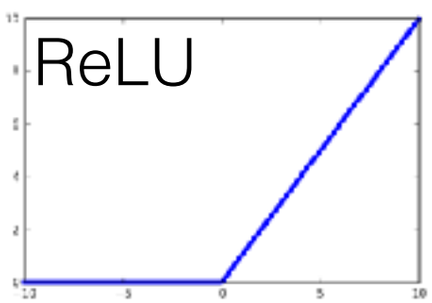


$$output = \sigma\left(\sum_i w_i x_i + b\right)$$

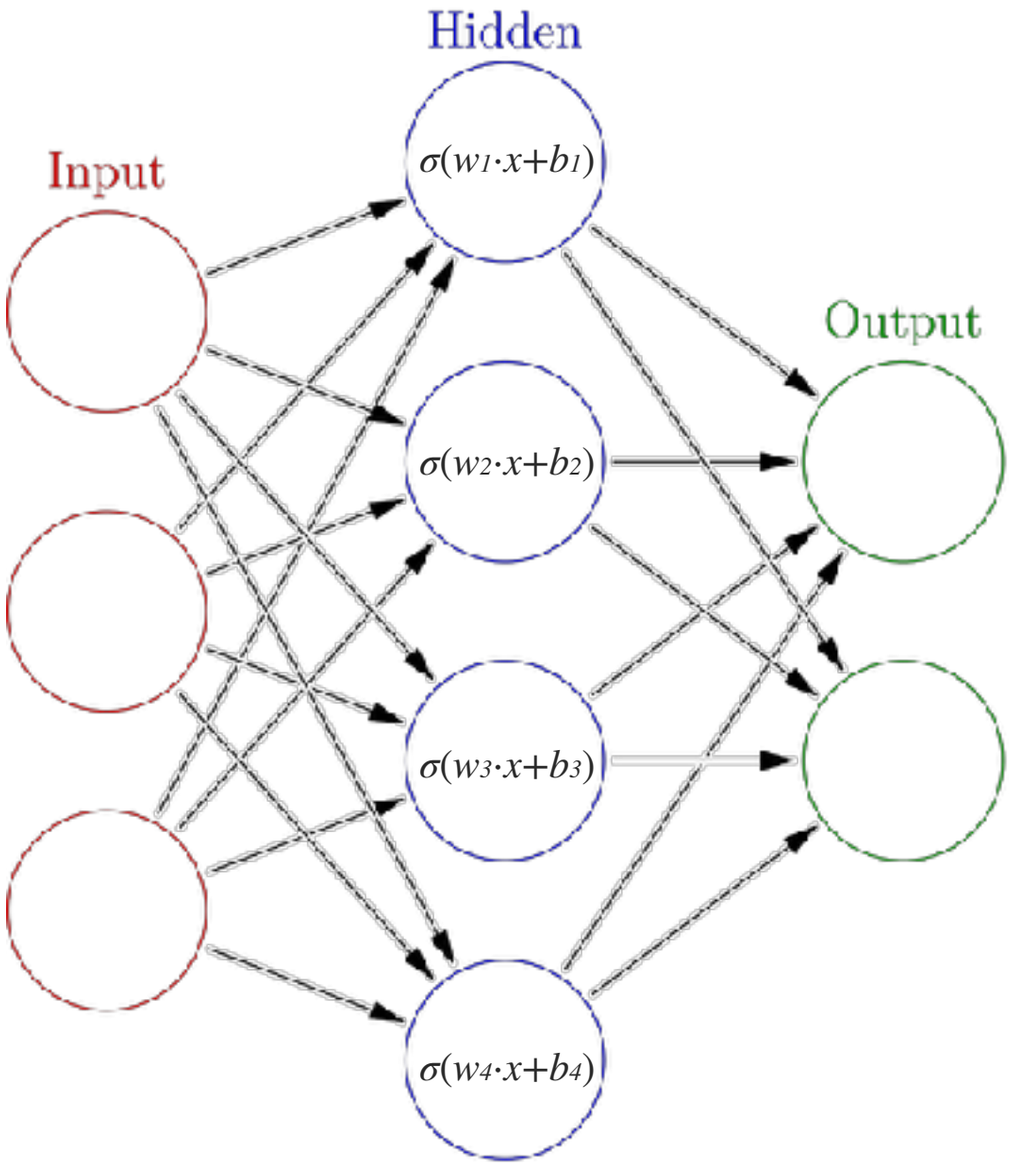


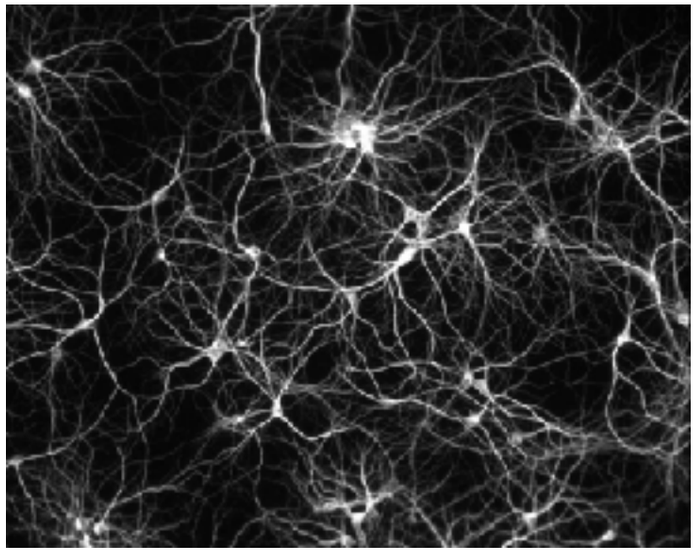






Neural Networks





The **universal approximation theorem** states that, under reasonable assumptions, a feedforward **neural network** with a finite number of nodes **can approximate any continuous** function to within a given error over a bounded input domain.

Deep Learning

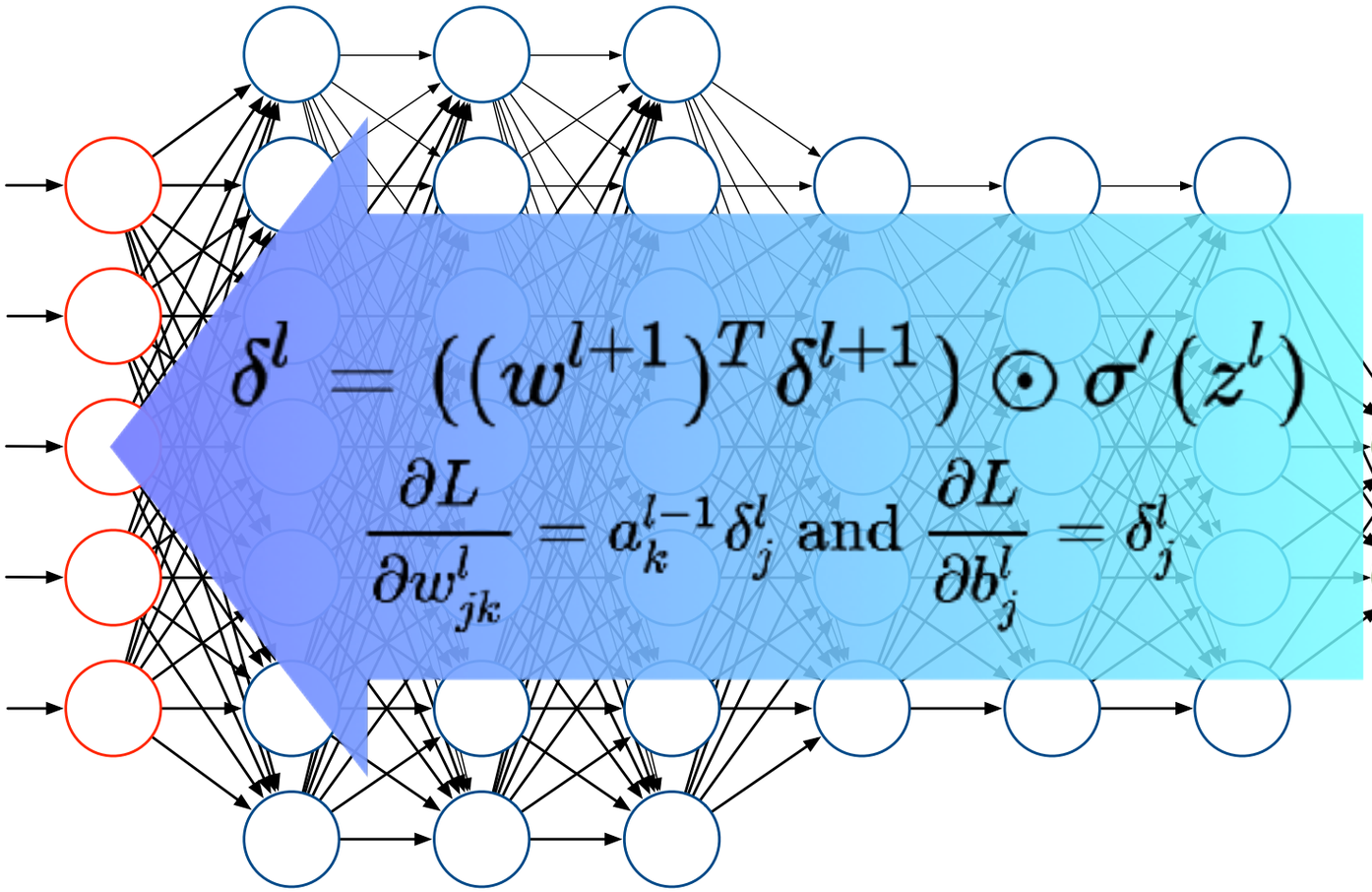
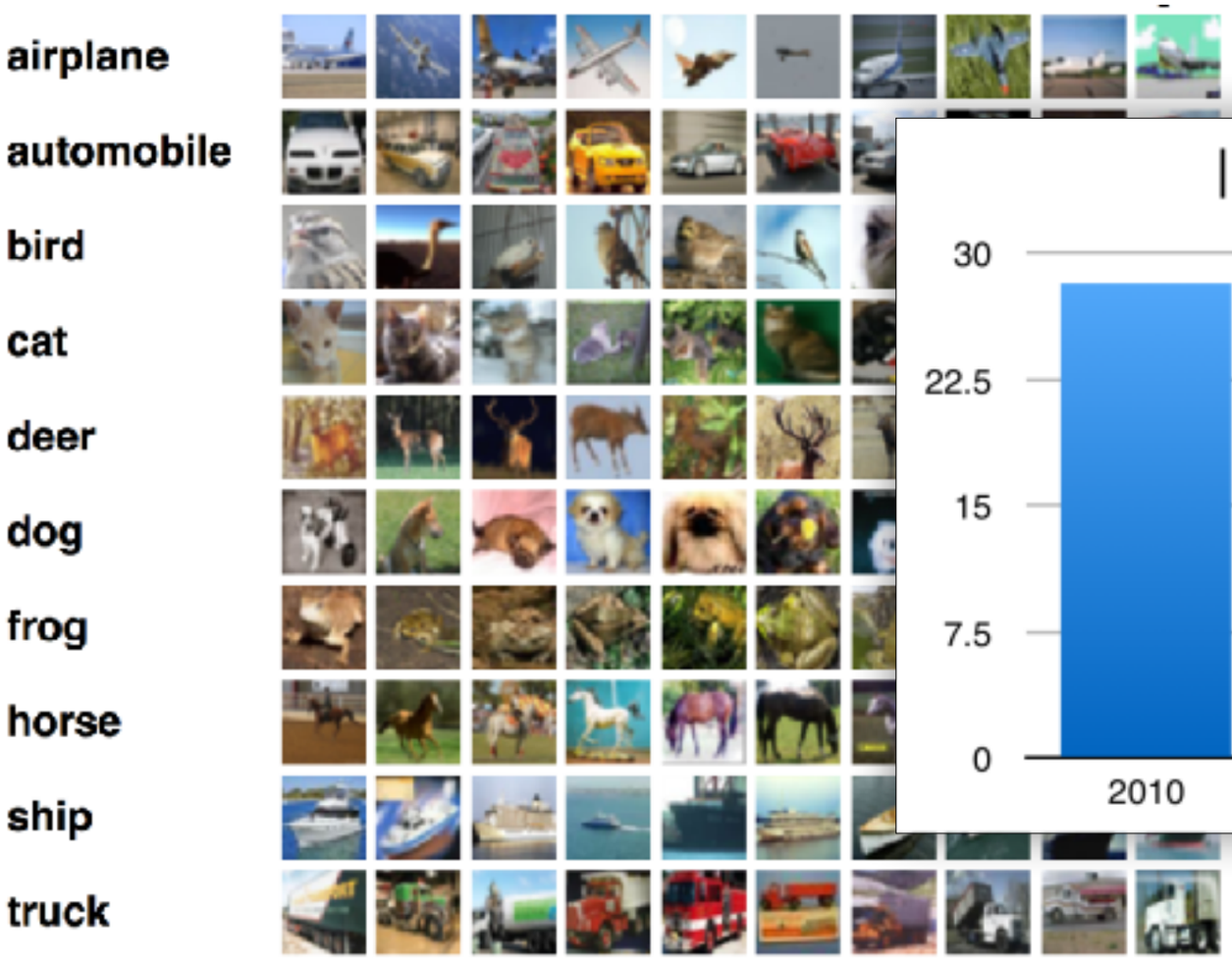
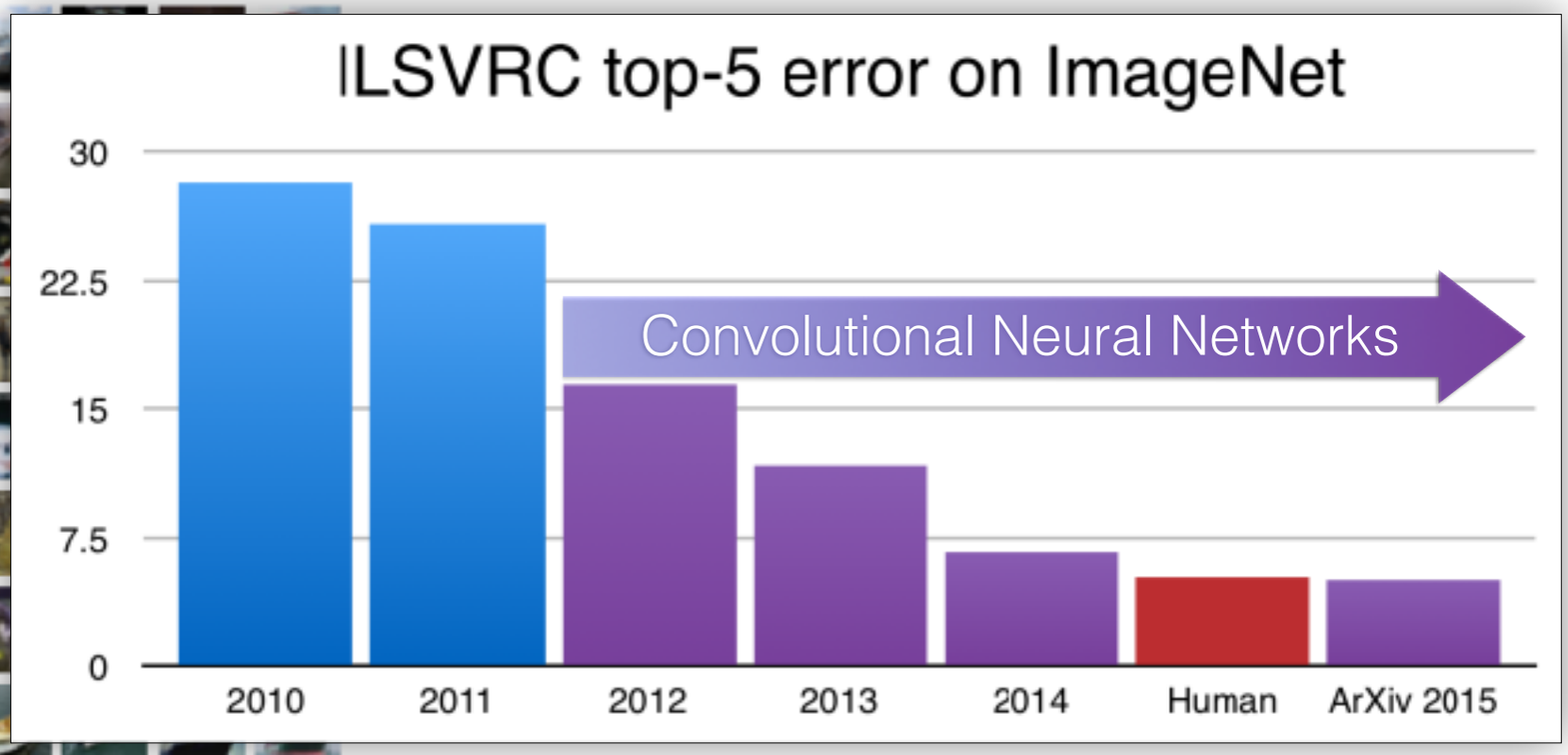


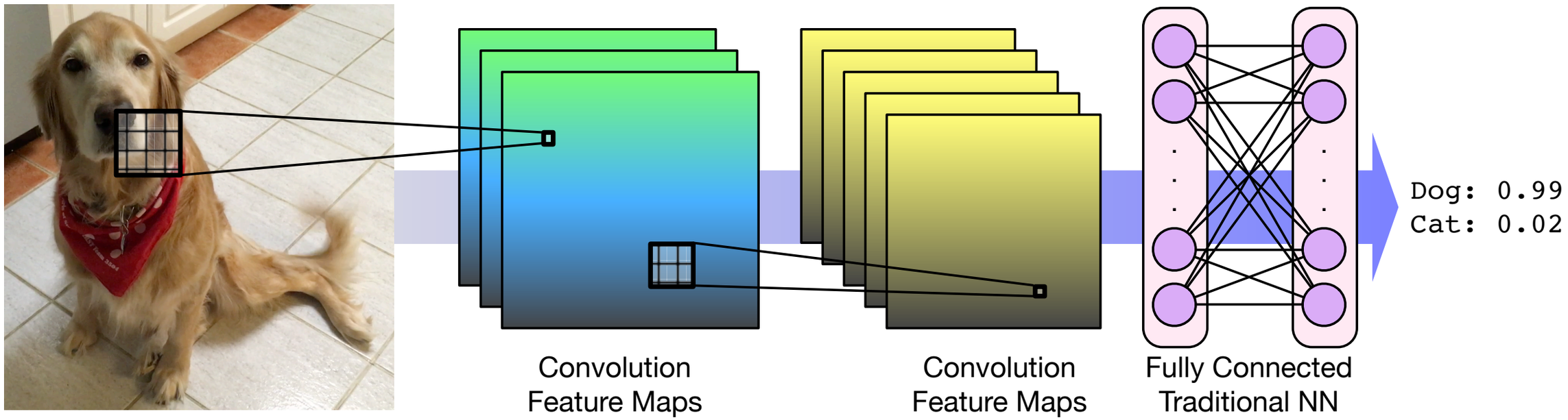
Image Recognition

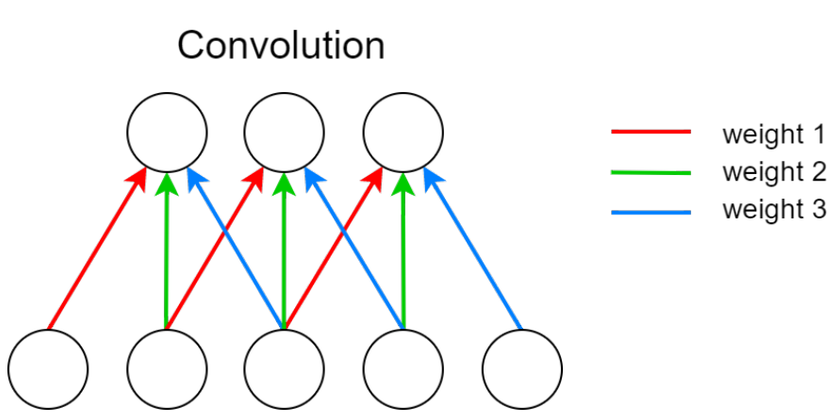


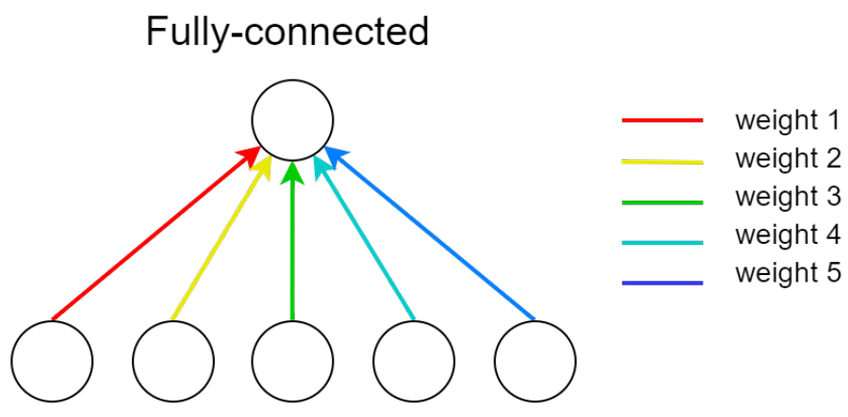


https://devblogs.nvidia.com

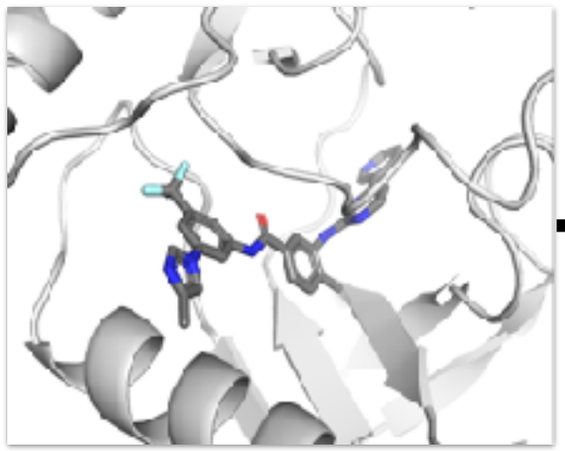
Convolutional Neural Networks







CNNs for Protein-Ligand Scoring



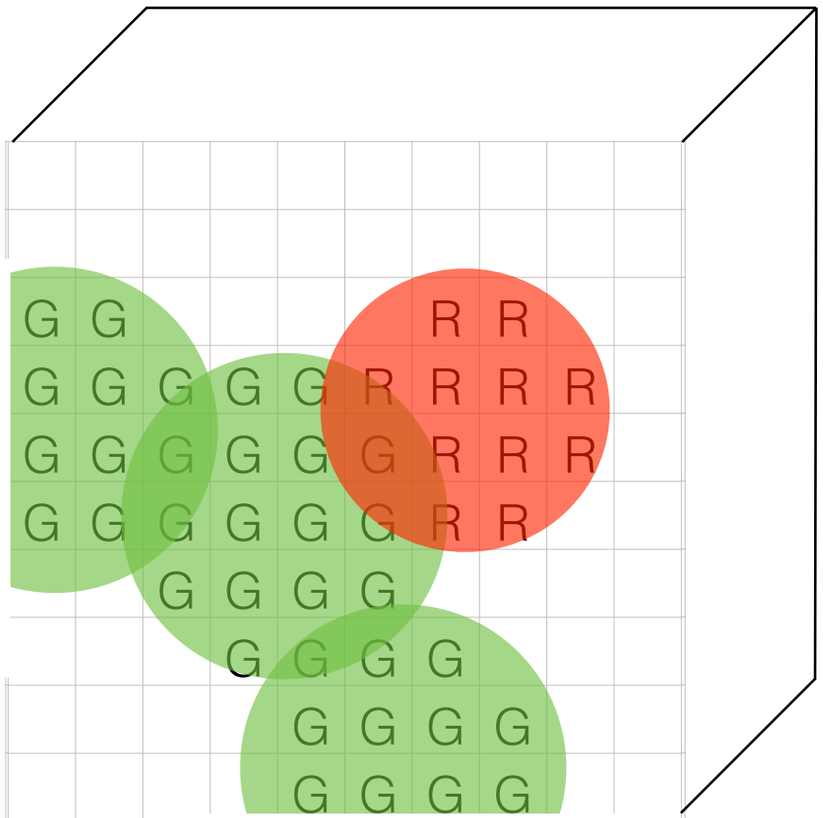
CNN

Pose Prediction

Binding Discrimination

Affinity Prediction

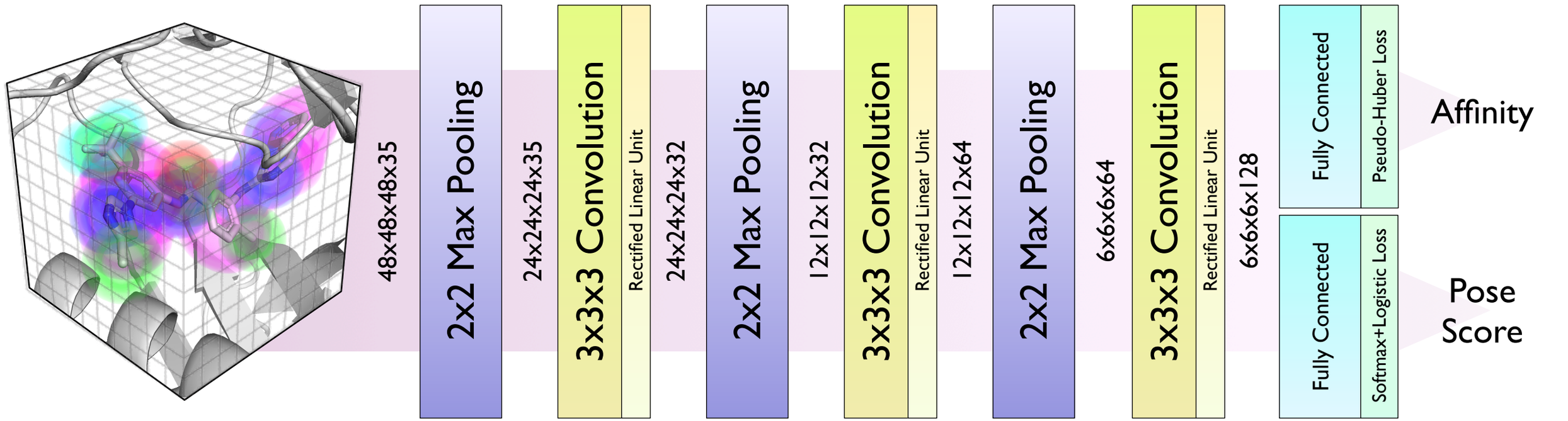
Protein-Ligand Representation



(R,G,B) pixel → (Carbon, Nitrogen, Oxygen,…) **voxe**

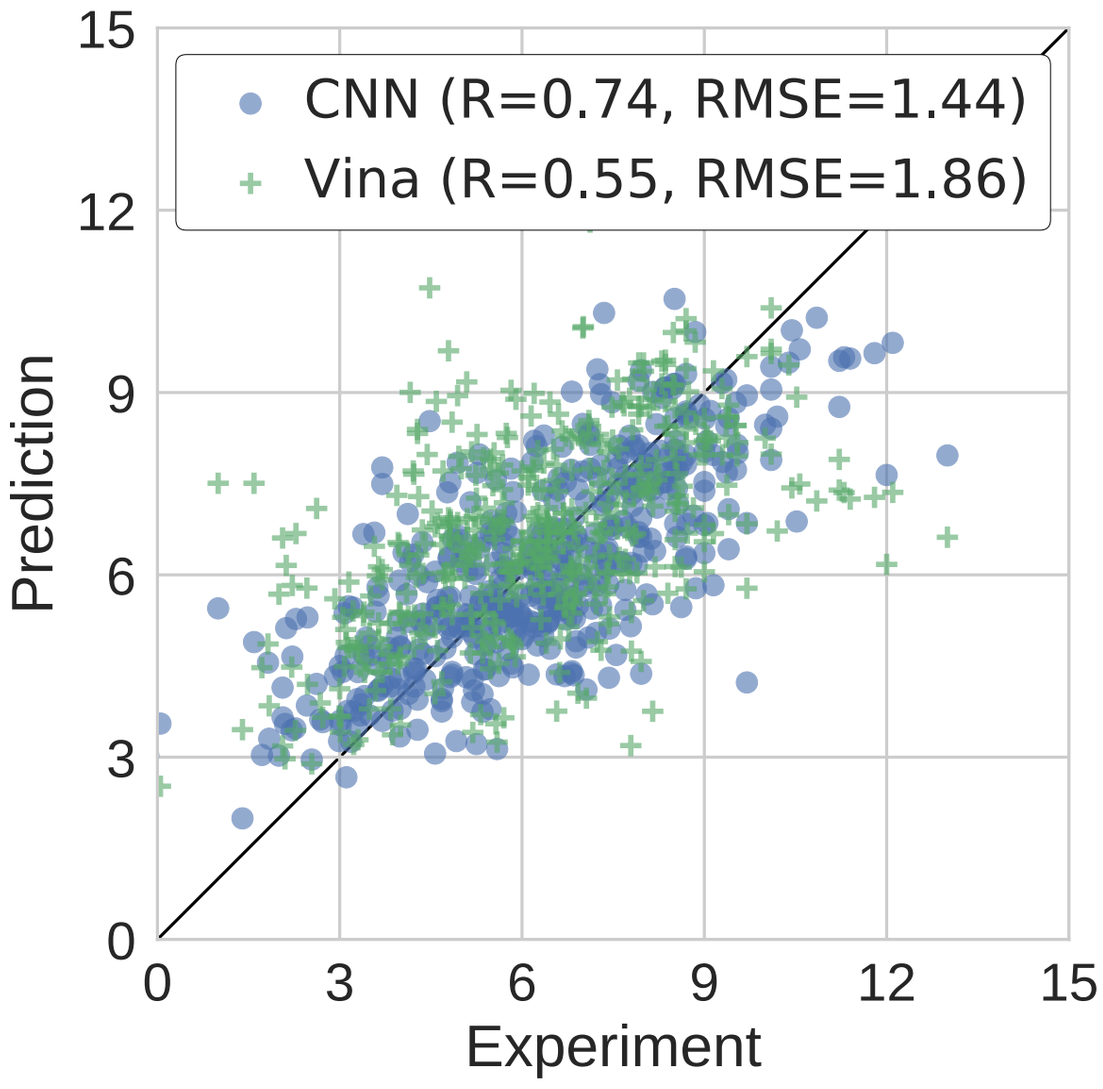
The only parameters for this representation are the choice of **grid resolution**, **atom density**, and **atom types**.

Model

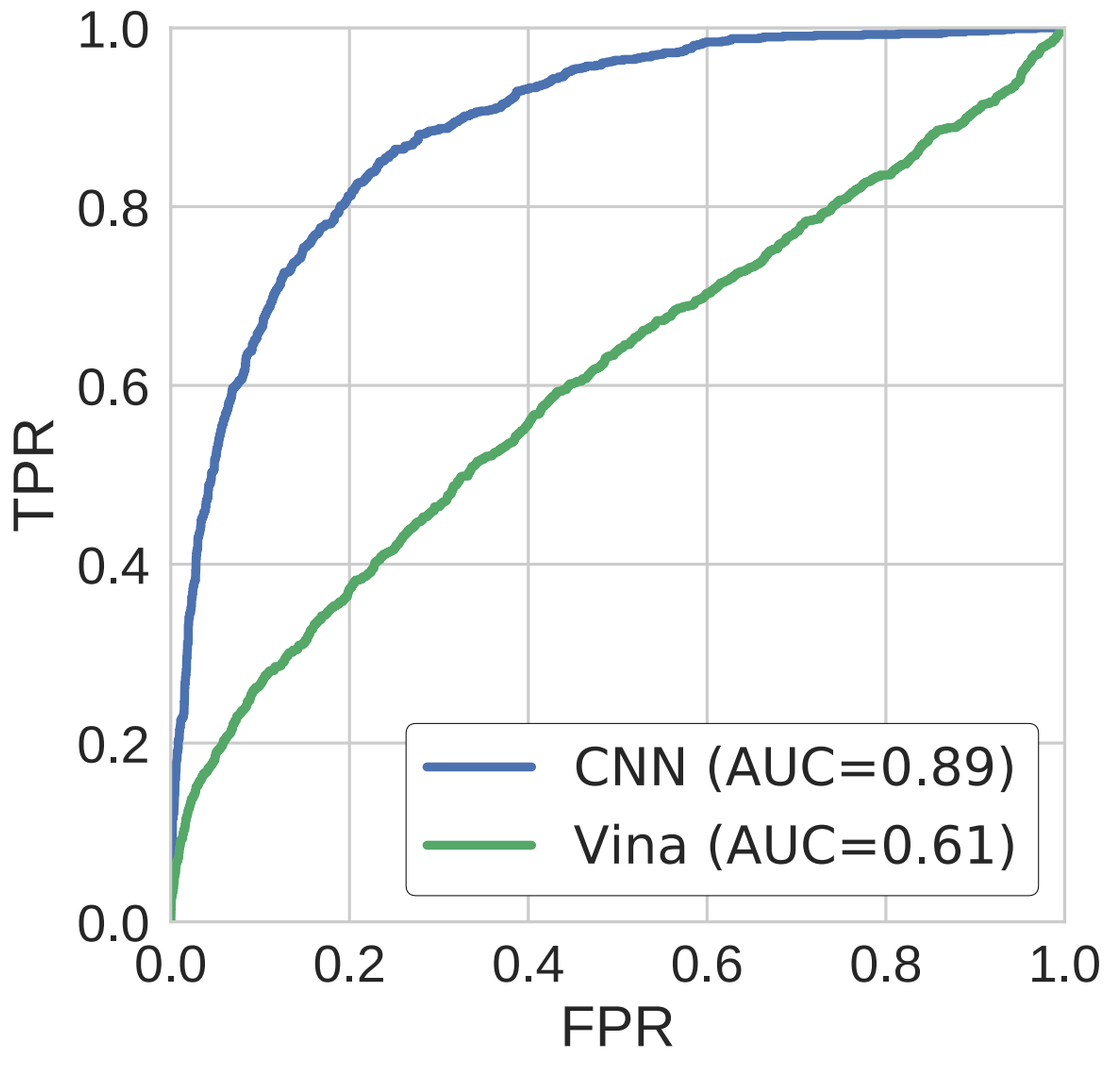

48x48x48x35
2x2 Max Pooling
24x24x24x35
3x3x3 Convolution
Rectified Linear Unit
24x24x24x32
2x2 Max Pooling
12x12x12x32
3x3x3 Convolution
Rectified Linear Unit
12x12x12x64
2x2 Max Pooling
6x6x6x64
3x3x3 Convolution
Rectified Linear Unit
6x6x6x128
Fully Connected
Pseudo-Huber Loss
Affinity
Fully Connected
Softmax+Logistic Loss
Pose Score

Results

Affinity Prediction

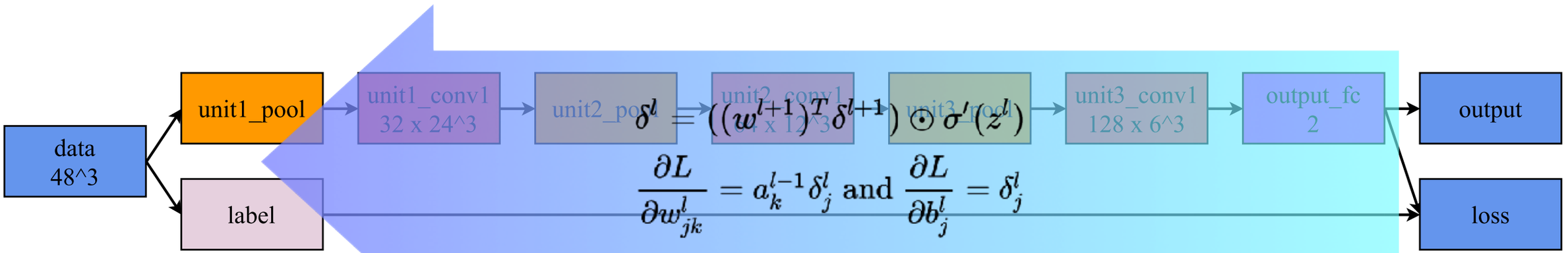


Pose Prediction

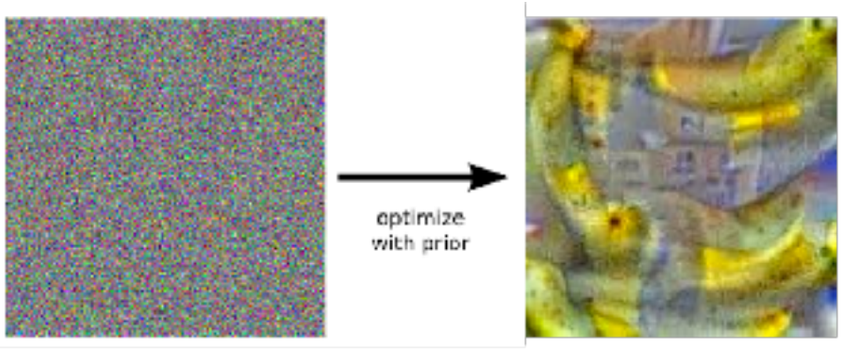


Trained on PDBbind refined; tested on CSAR 🤔

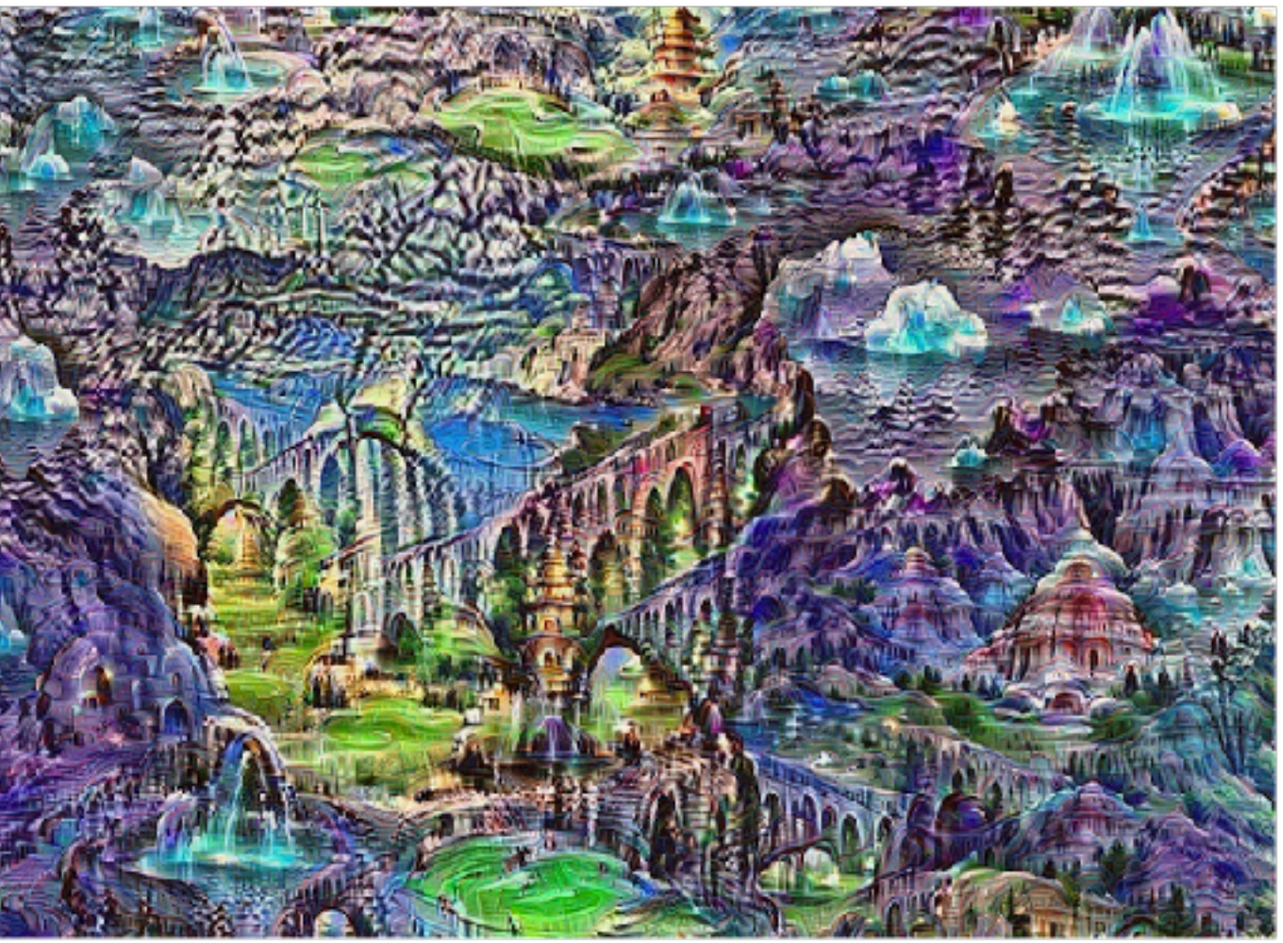
Beyond Scoring

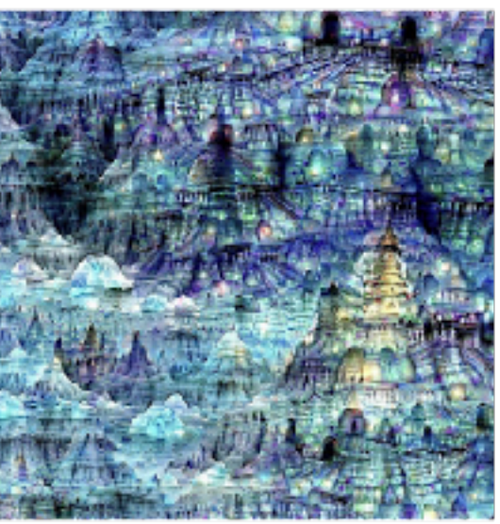


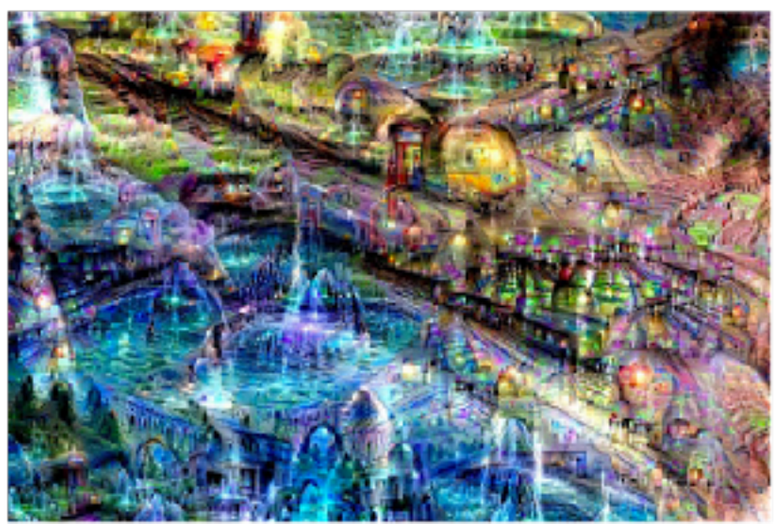
Deep Dreams

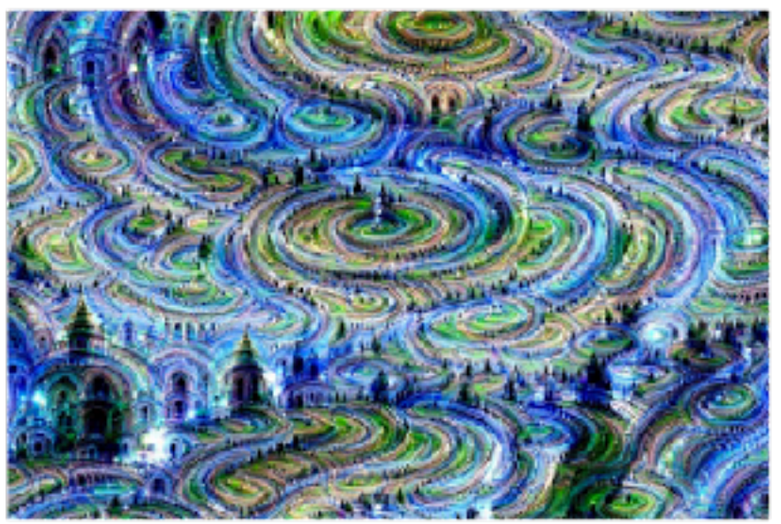


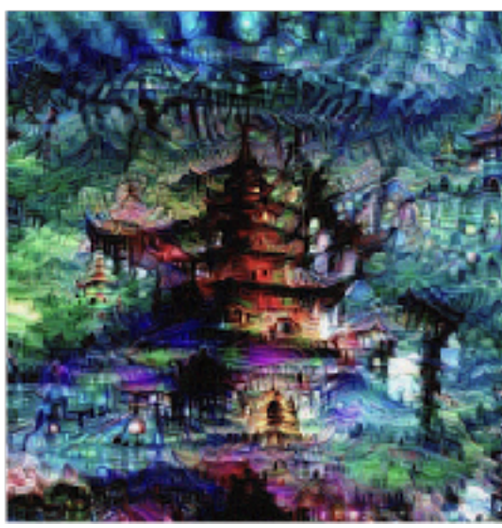
https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html





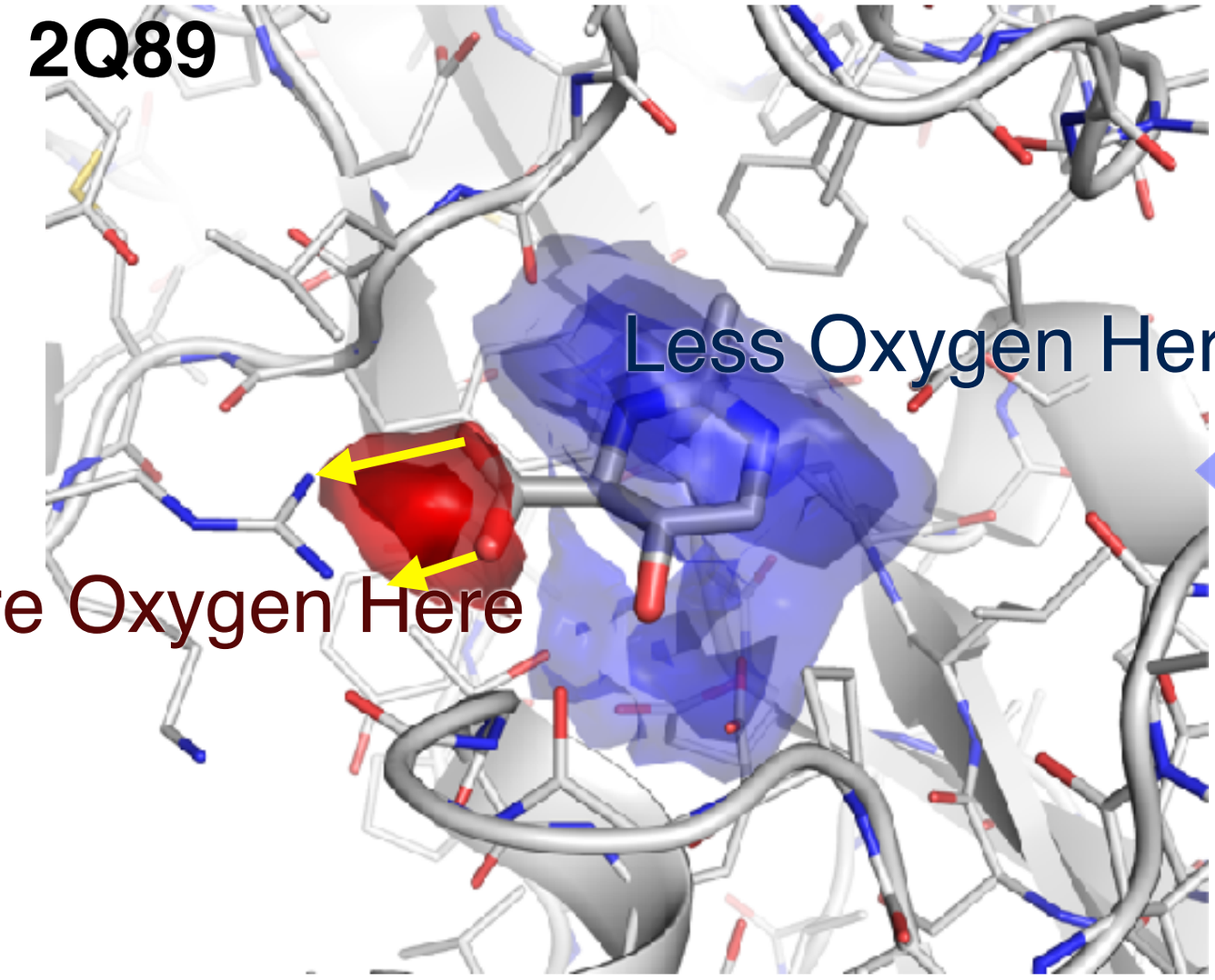






https://deepdreamgenerator.com/#gallery

Beyond Scoring

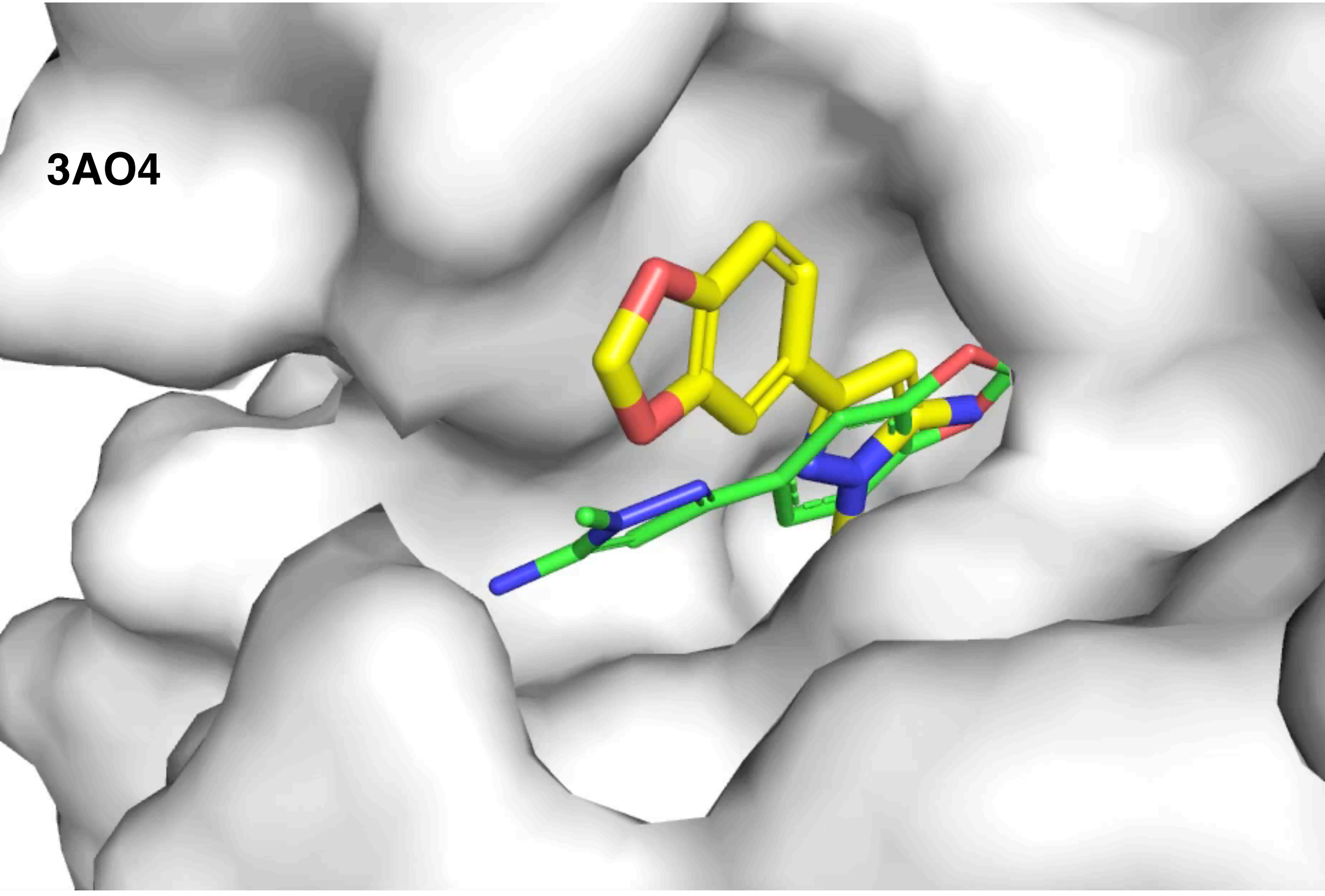


$$\frac{\partial L}{\partial A} = \sum_{i \in G_A^{48^3}} \frac{\partial L}{\partial G_i} \frac{\partial G_i}{\partial D} \frac{\partial D}{\partial A}$$

data

unit1_pool

label


3AO4

Key Concepts

Ligand-Based Virtual Screening

Identifying new active compounds based on similarity to known active compounds; fingerprint is a bit vector representation of a molecule

Pharmacophore

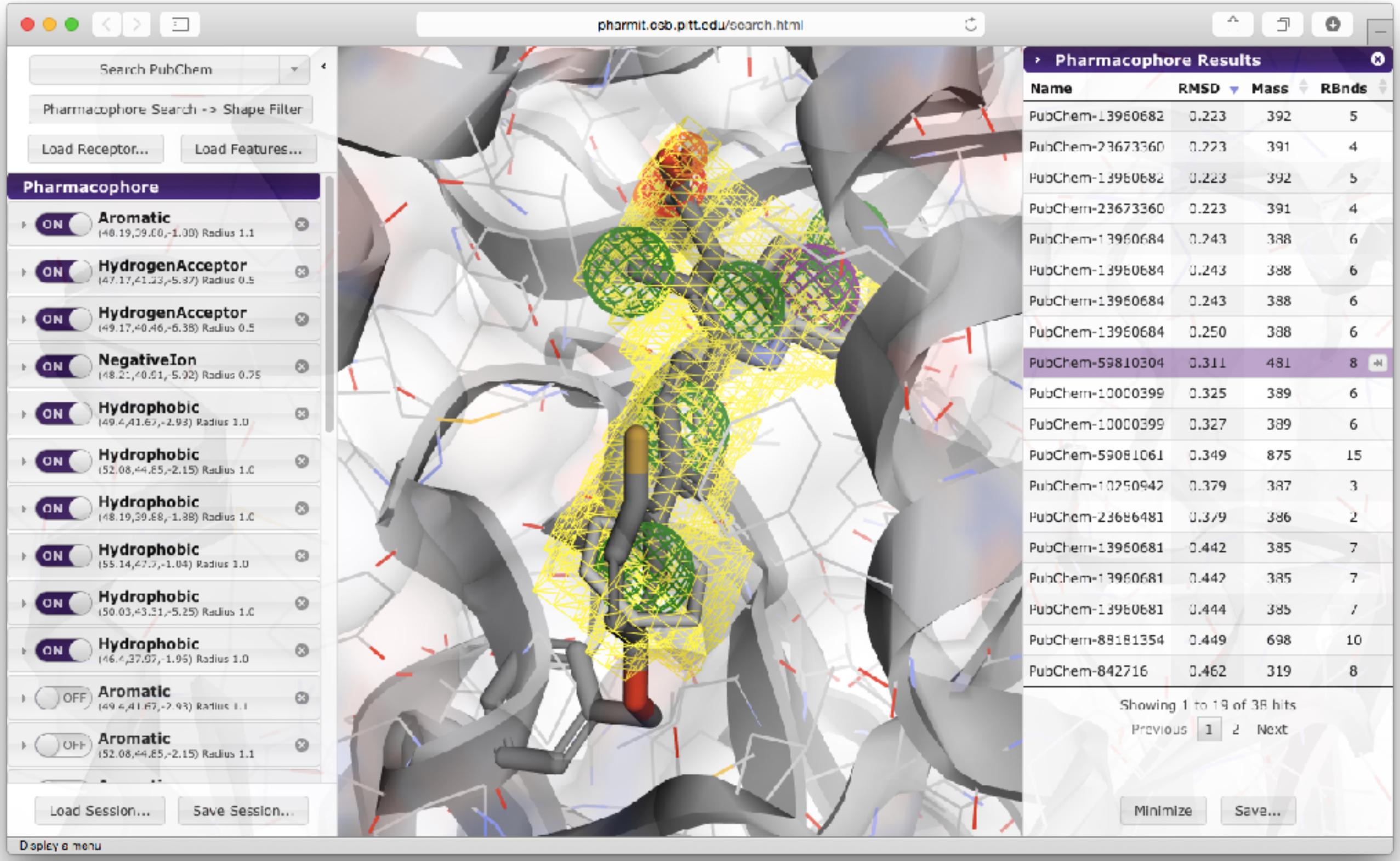
A spatial arrangement of molecular features essential for biological activity - hydrogen bonding, hydrophobic, charged, etc.

Docking

Predict the position, pose and affinity of a molecule using the receptor structure

Scoring

force field … empirical … knowledge based



http://pharmit.csb.pitt.edu