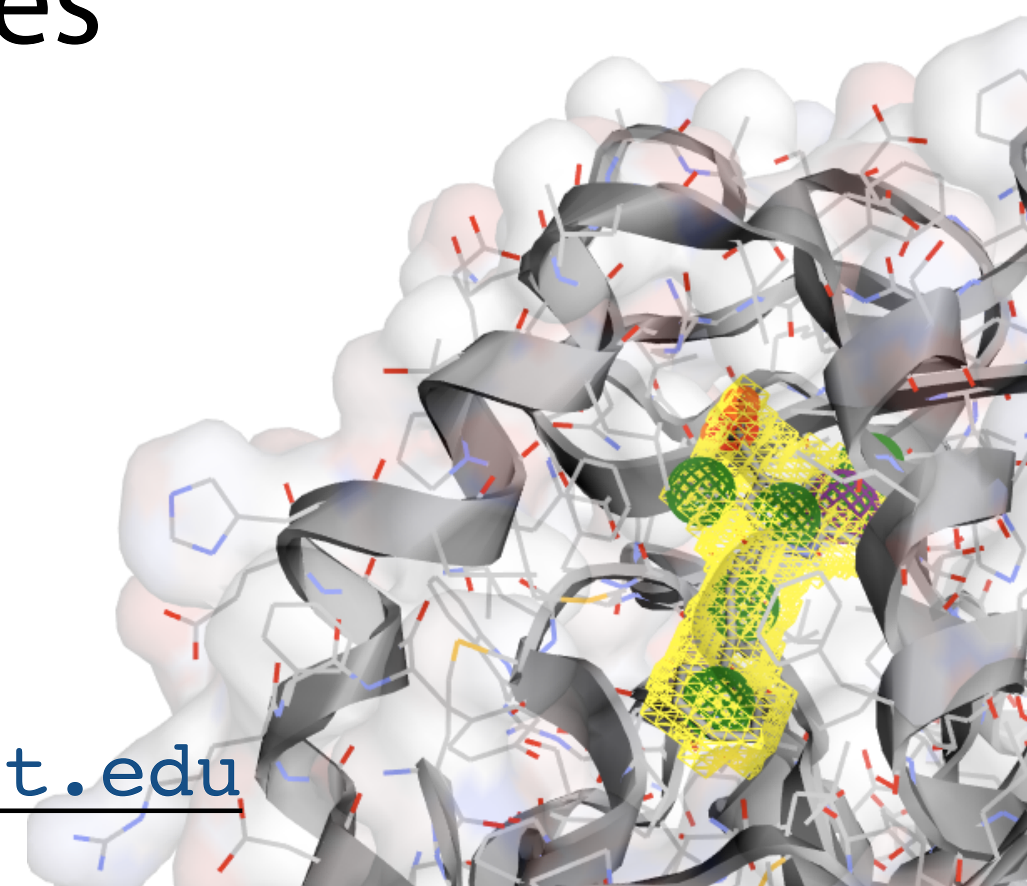
University of Pittsburgh
Department of Computational & Systems Biology

Computational Drug Discovery

David Ryan Koes

11/6/2018



http://bits.csb.pitt.edu

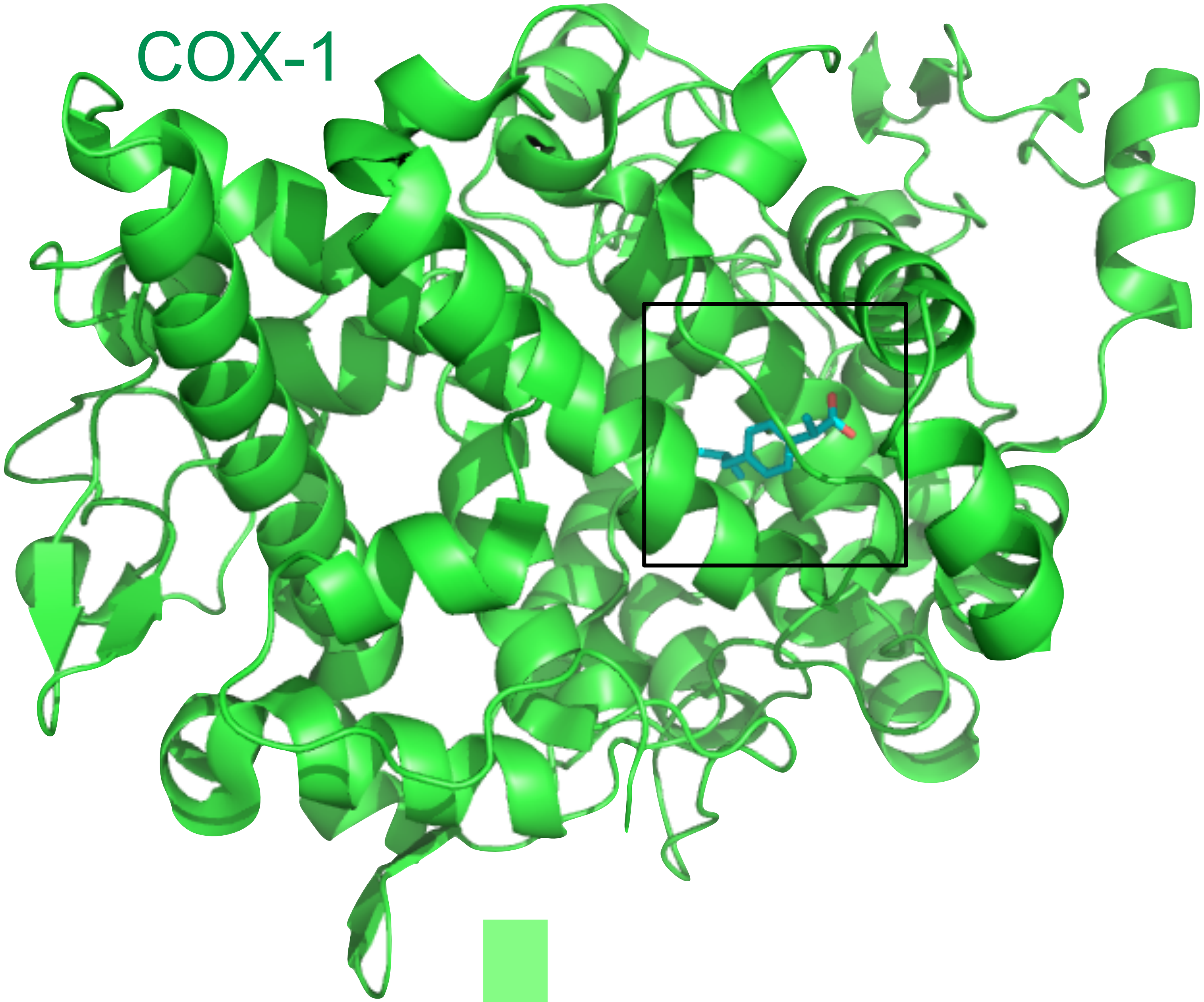
What is a drug?

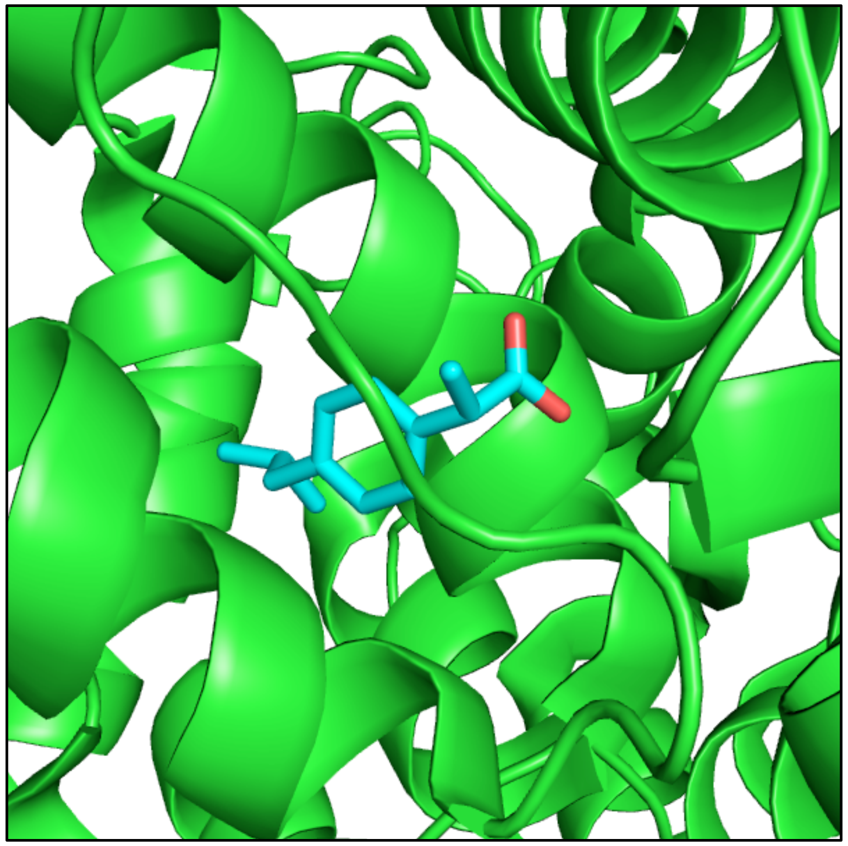
According to the Food, Drug, and Cosmetic Act (1) **:** a substance recognized in an official pharmacopoeia or formulary (2) **:** a substance intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease (3) **:** **a substance** other than food **intended to affect the structure 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

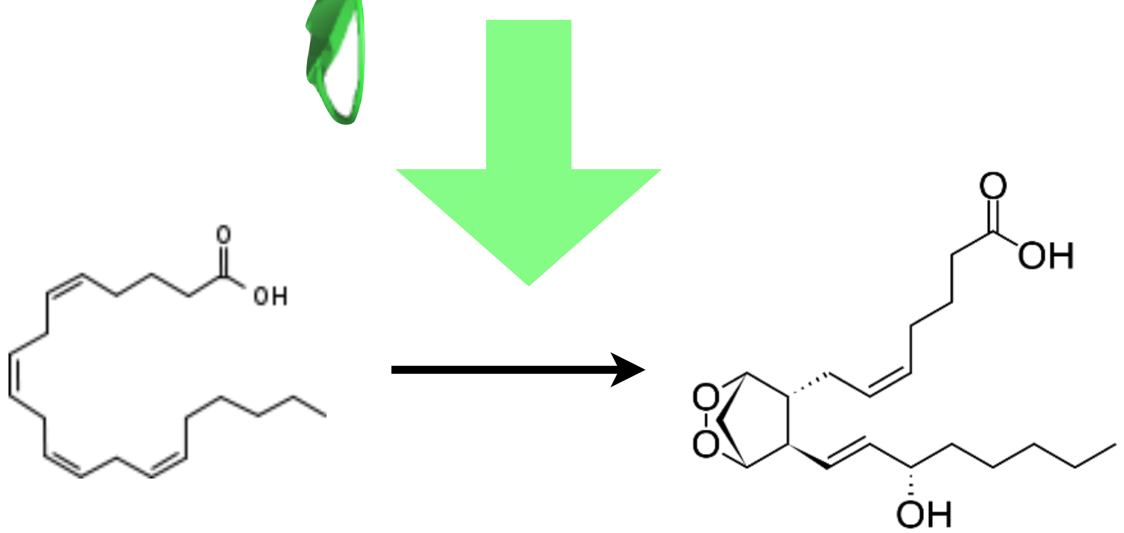
What is a drug?

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

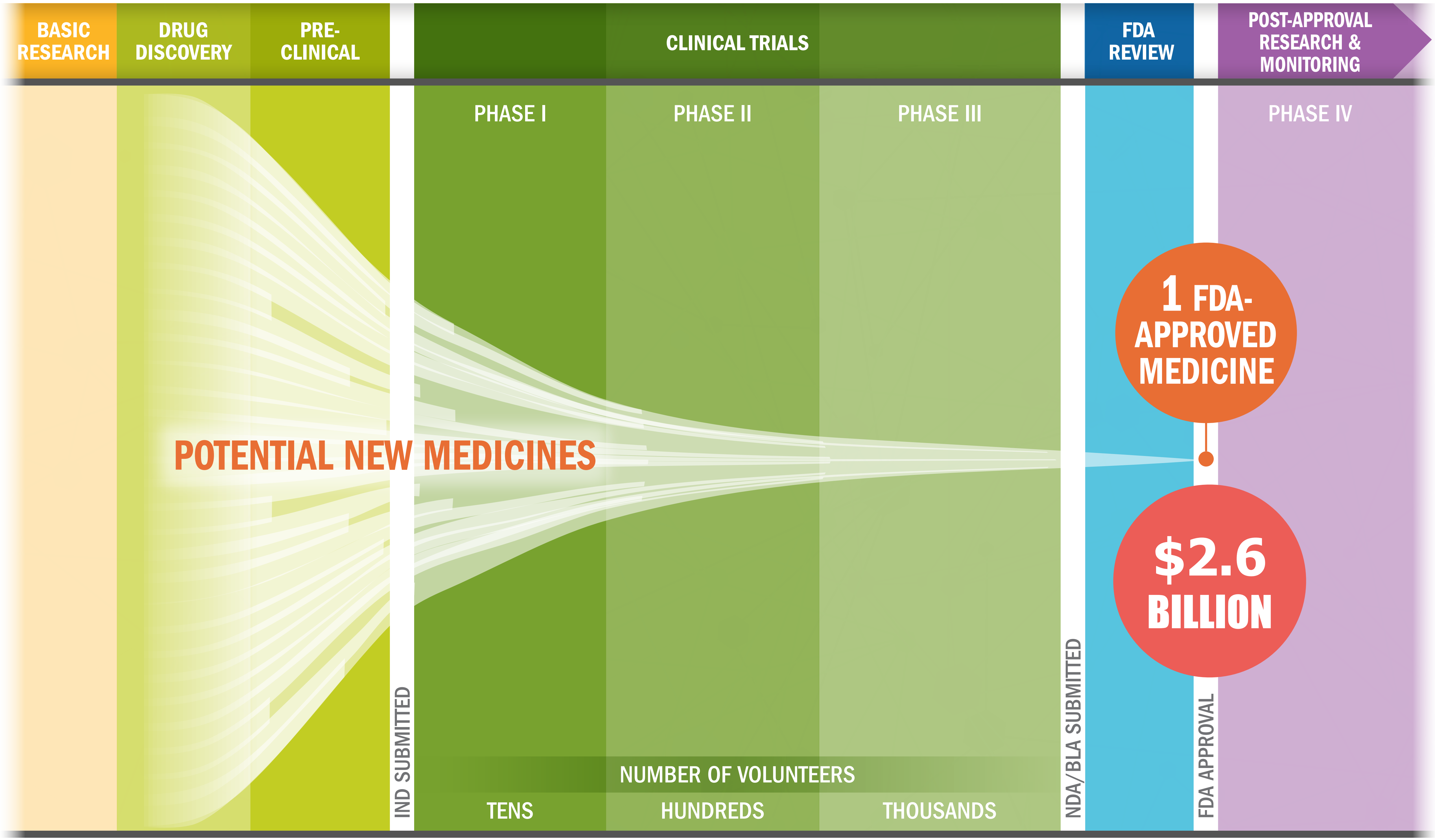
A small molecule intended to affect the structure/function of macromolecules


COX-1



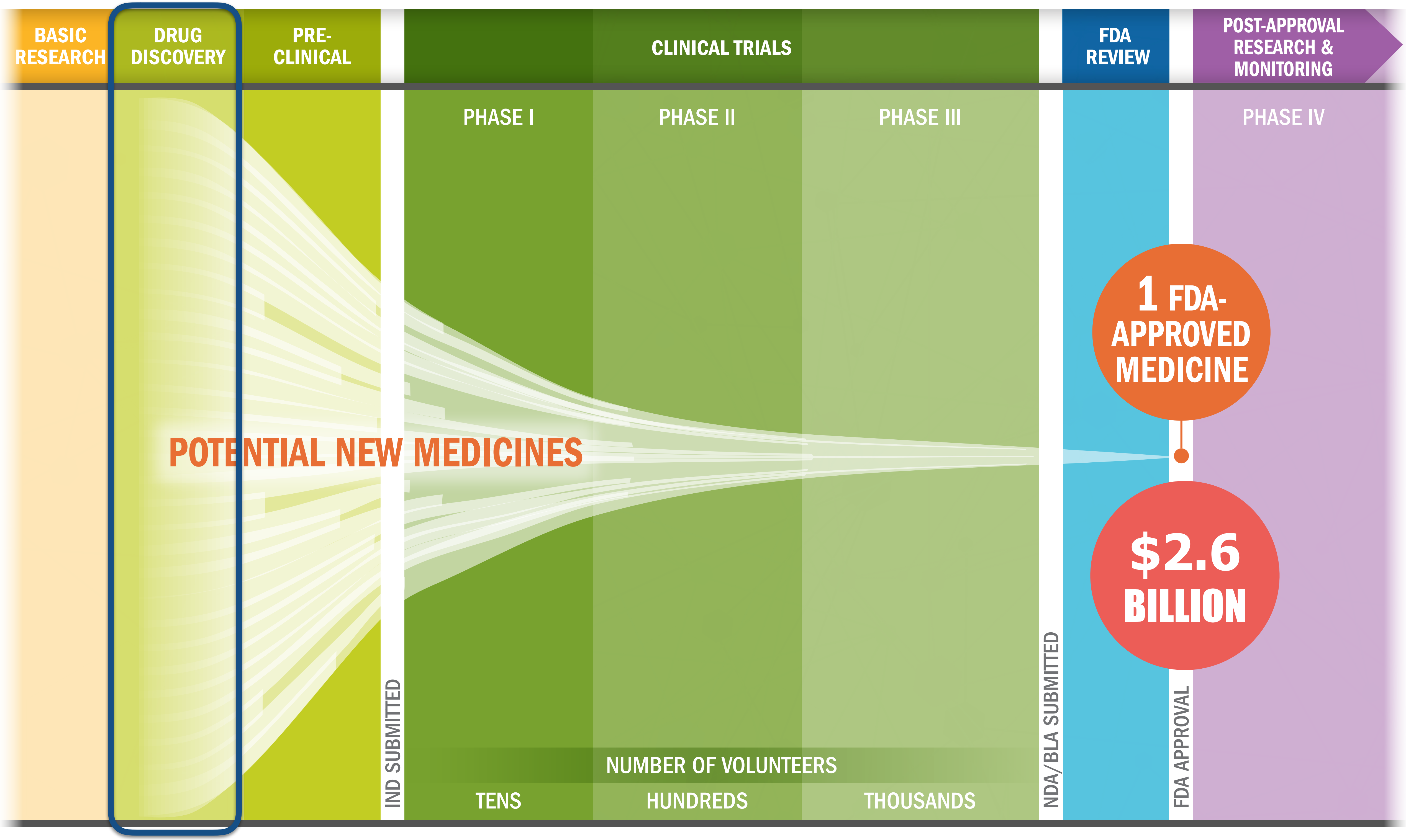

O
OH
O
OH
O
O
OH

THE BIOPHARMACEUTICAL RESEARCH AND DEVELOPMENT PROCESS



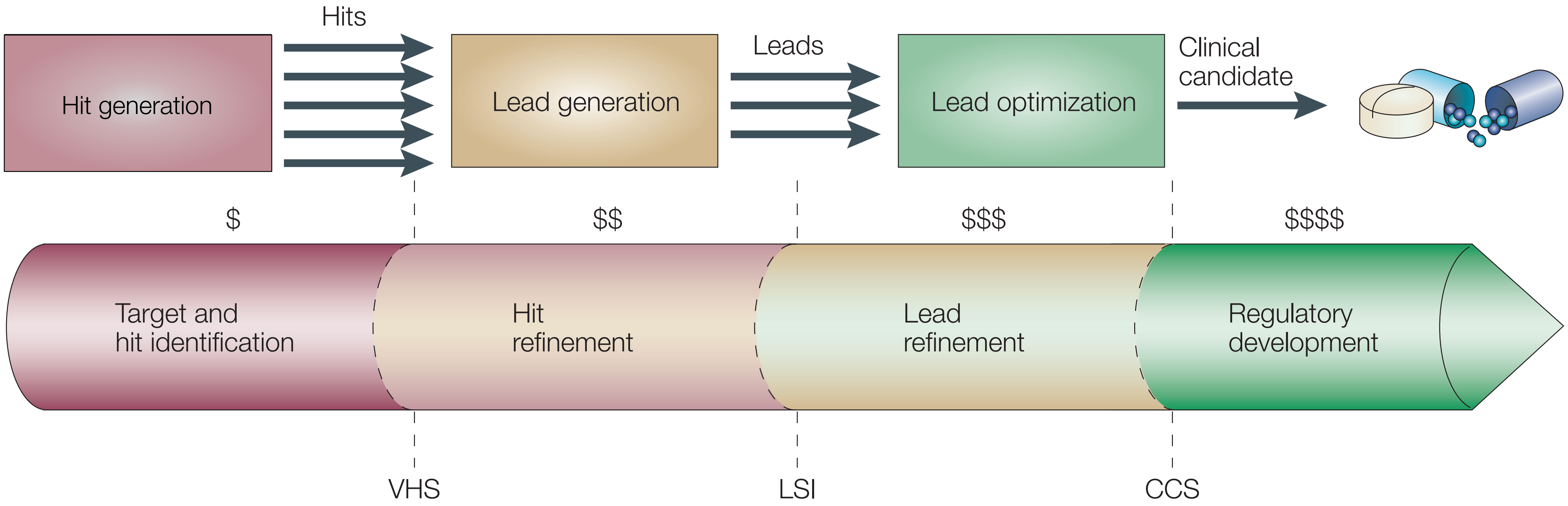
Source: Pharmaceutical Research and Manufacturers of America (http://phrma.org)

THE BIOPHARMACEUTICAL RESEARCH AND DEVELOPMENT PROCESS



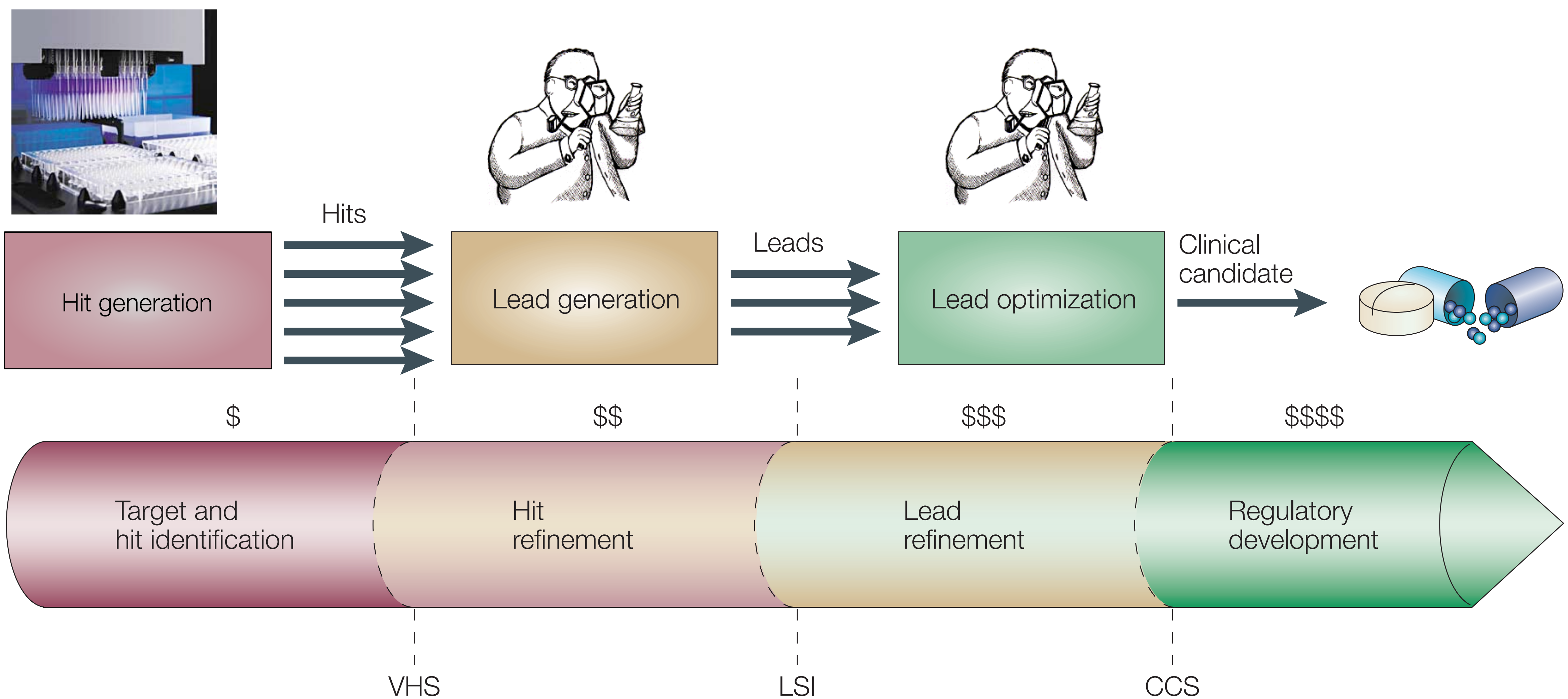
Source: Pharmaceutical Research and Manufacturers of America (http://phrma.org)

Drug Discovery



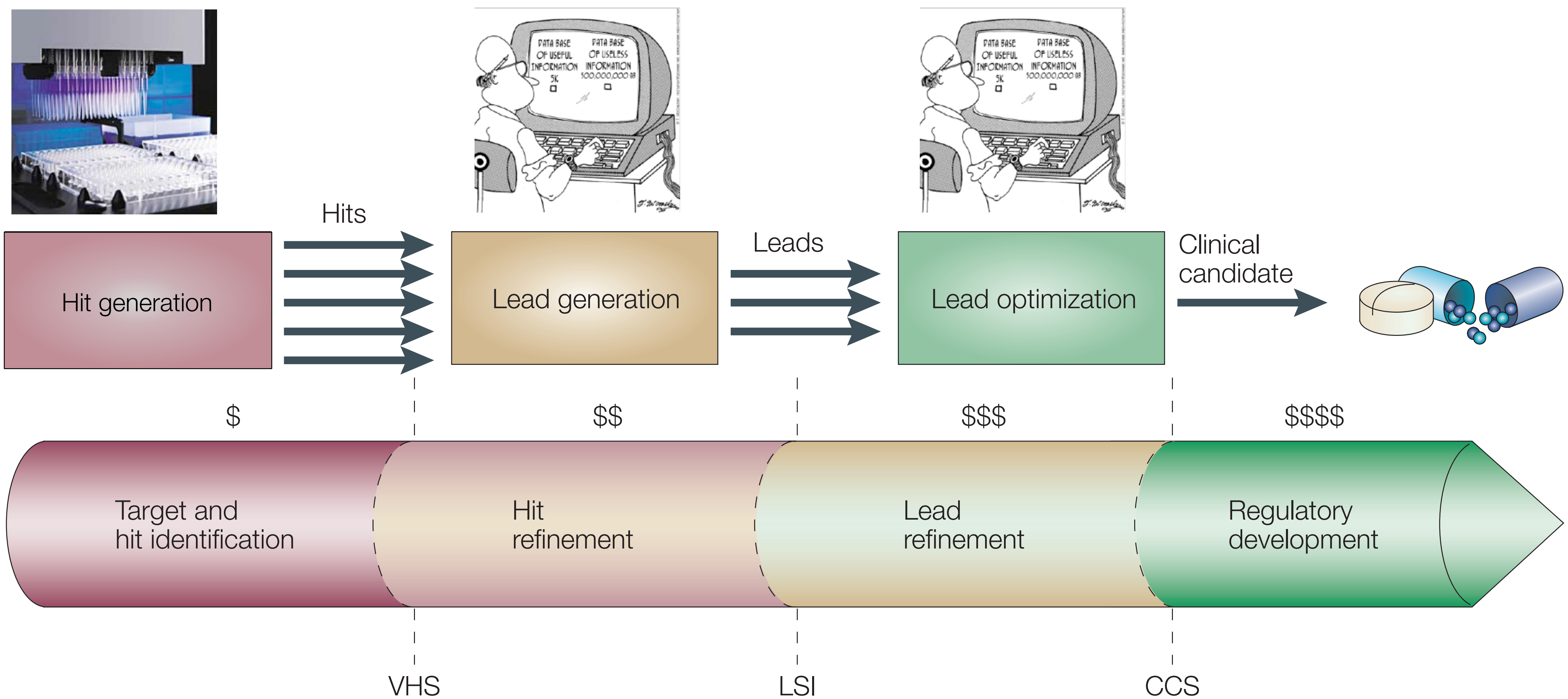
Drug Discovery

High Throughput Screening



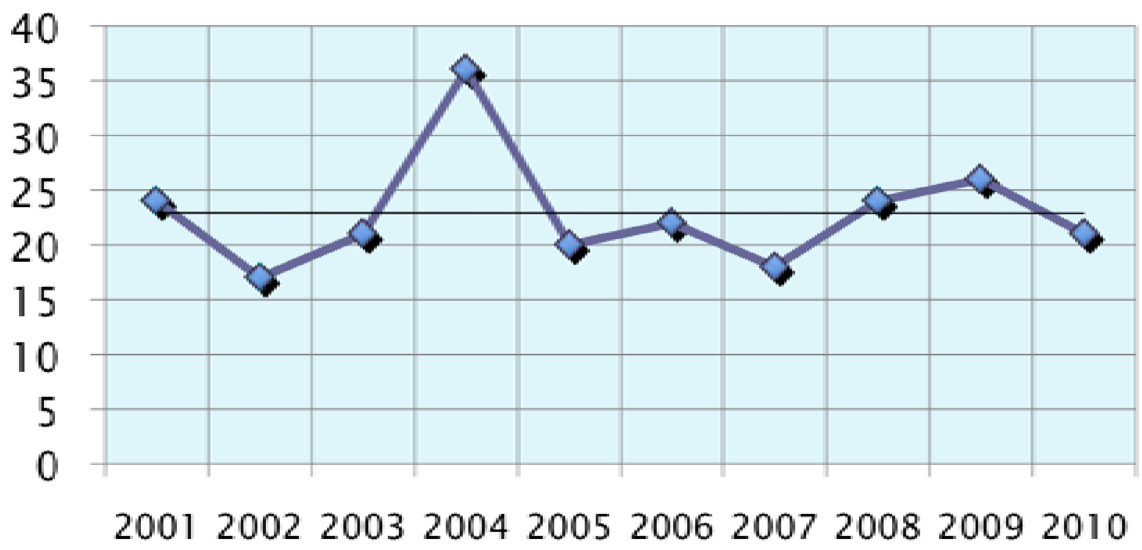
Drug Discovery

High Throughput Screening

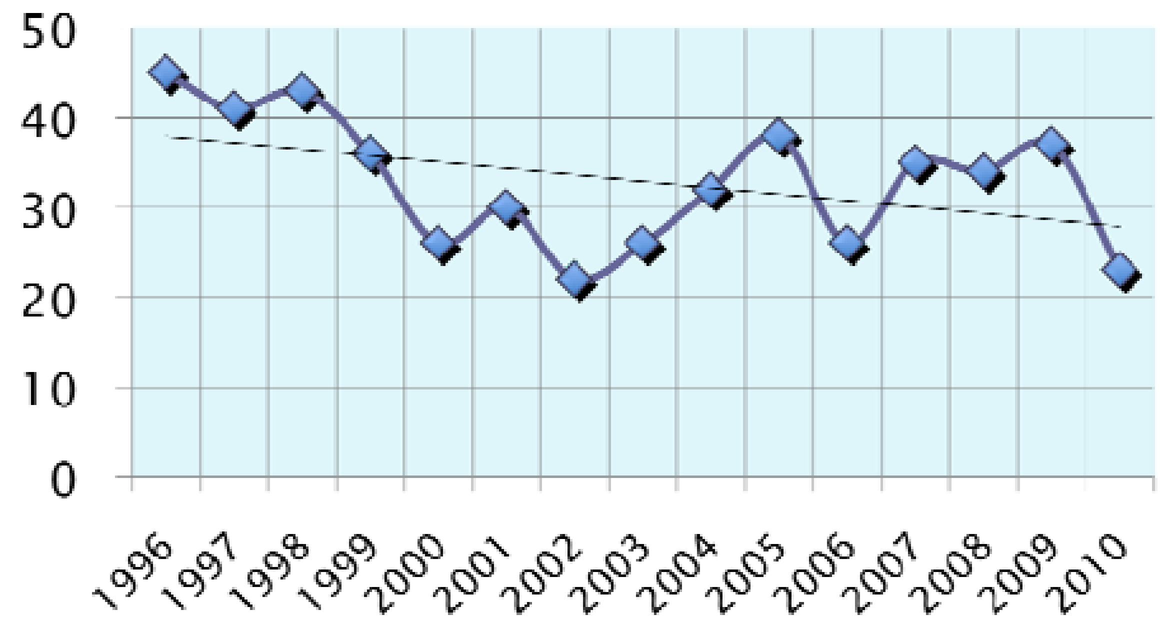


The State of Drug Development

New Drugs Approved



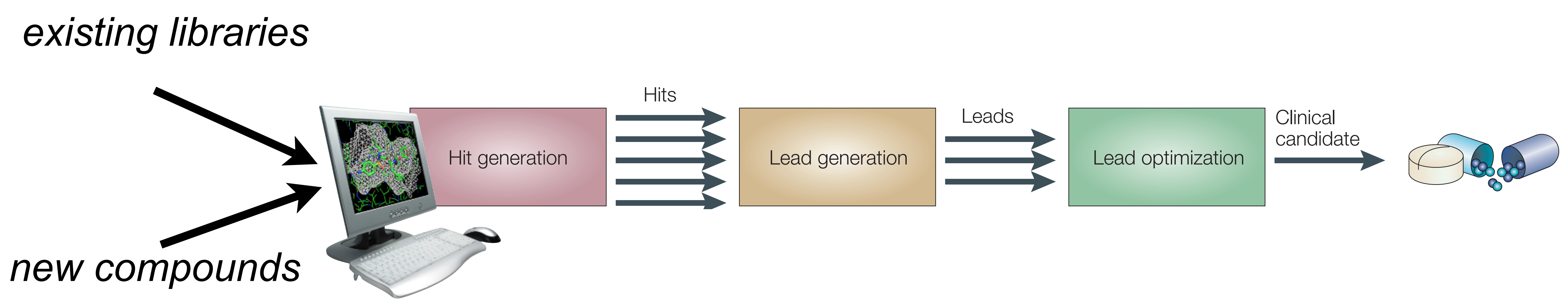
New Drug Applications



http://www.fda.gov/downloads/AboutFDA/Transparency/Basics/UCM247465.pdf

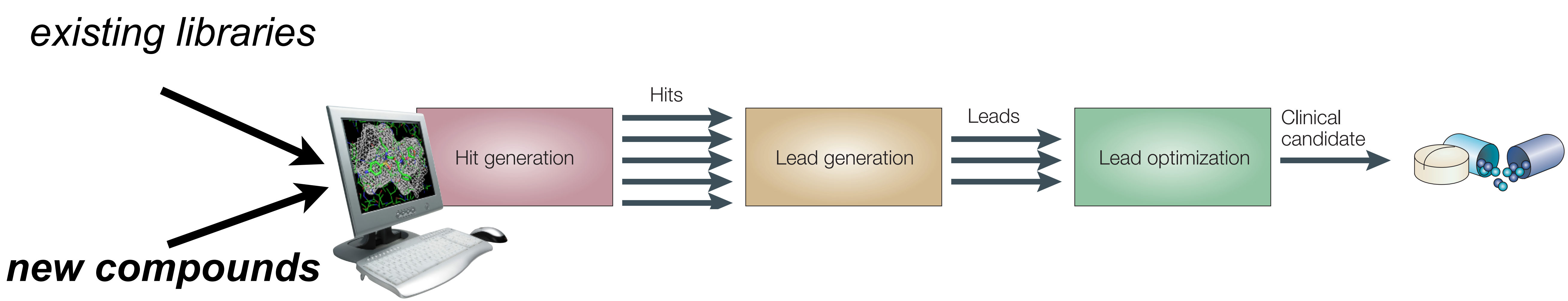
Computational Drug Discovery

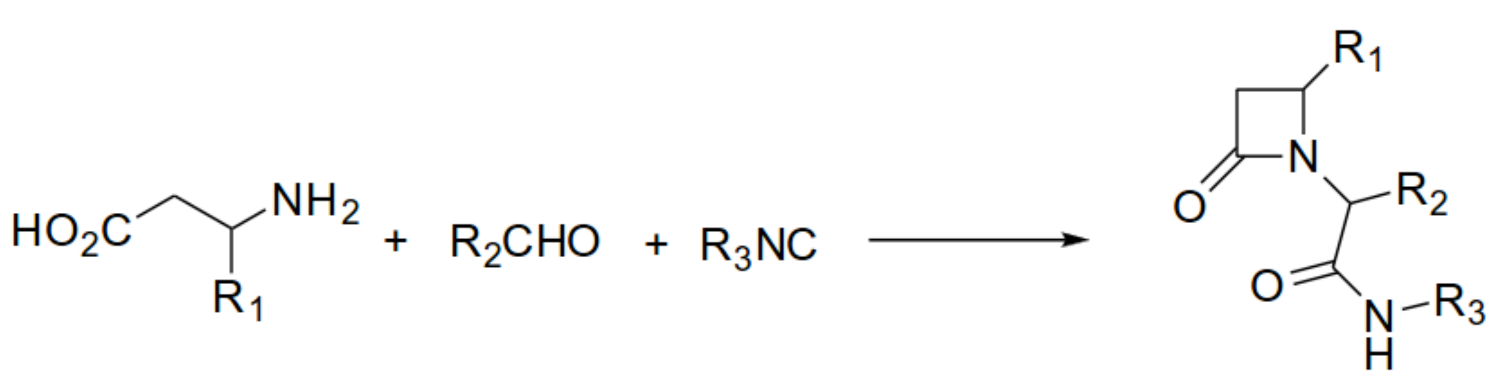
Virtual Screening



Computational Drug Discovery

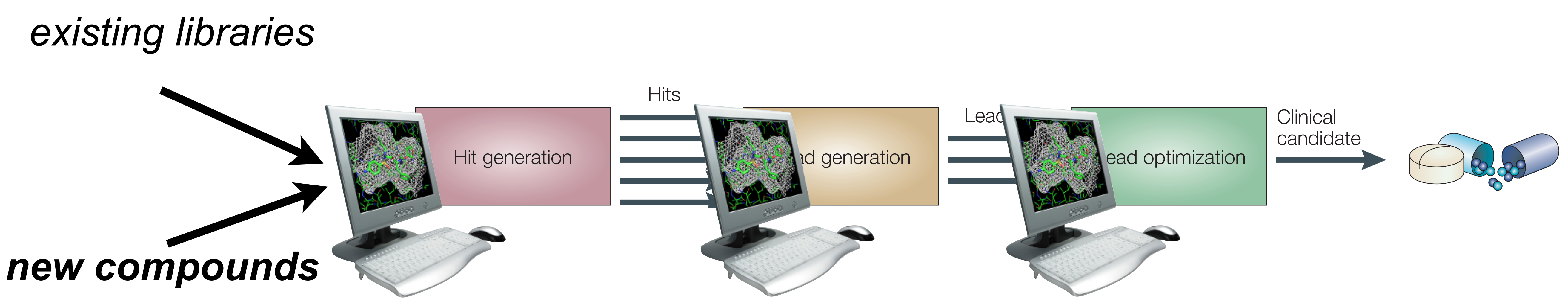
Virtual Screening

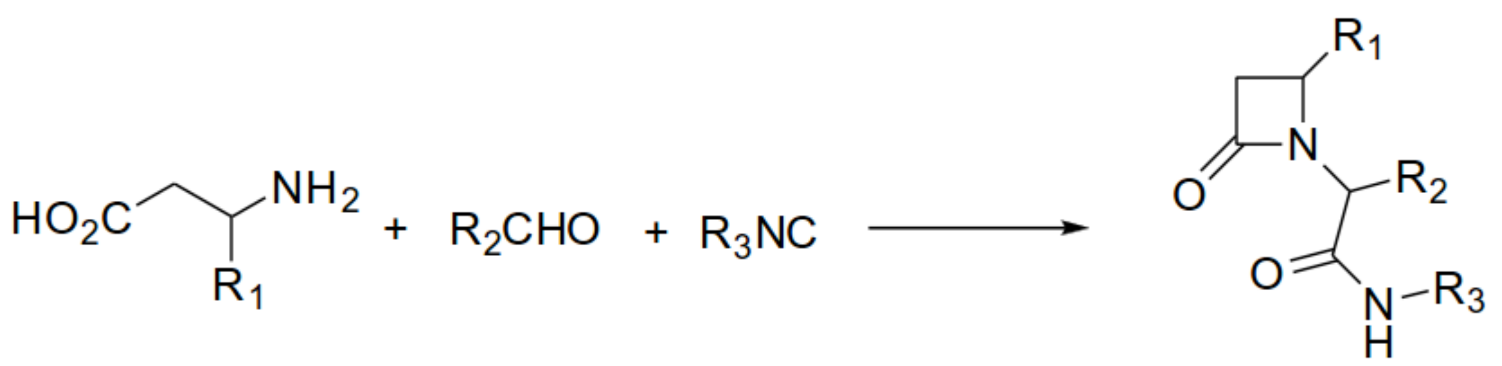




Computational Drug Discovery

Virtual Screening





Kinds of Virtual Screening

ADMET

Ligand Based

- similarity to known binder
- QSAR
- pharmacophore

Receptor Based

- dock and score
- simulation

  MM/GBSA, MM/PBSA, thermodynamic integration, free energy perturbation, Jarzynski, umbrella sampling, Monte Carlo, weighted ensemble, metadynamics…

Kinds of Virtual Screening

ADMET

Ligand Based

- similarity to known binder
- QSAR
- pharmacophore

Receptor Based

- dock and score
- simulation

MM/GBSA, MM/PBSA, thermodynamic integration, free energy perturbation, Jarzynski, umbrella sampling, Monte Carlo, weighted ensemble, metadynamics…

Not going to cover today

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

Ligand Based: Similarity

Fingerprint Methods

- map molecules to a descriptor space:

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

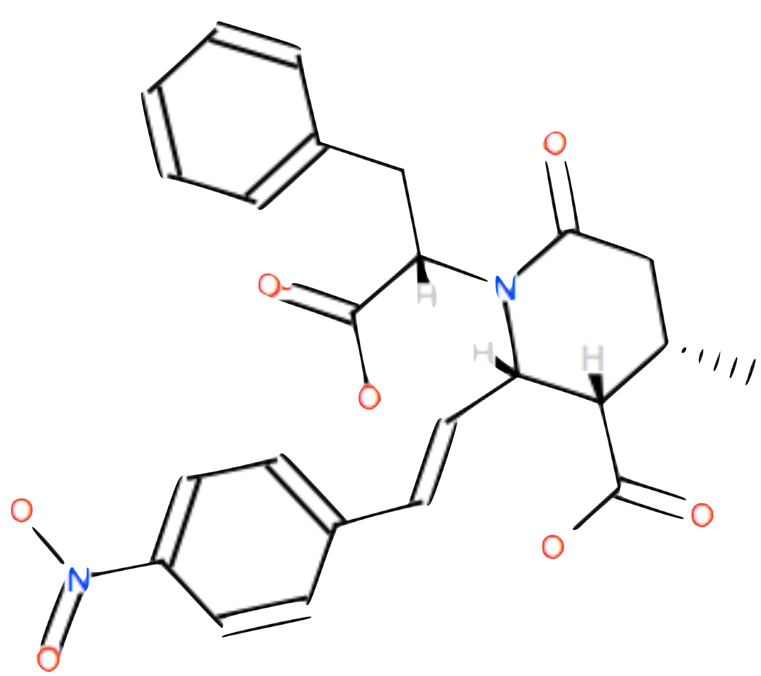
- similarity is "distance" between descriptors
- for bit vectors, Tanimoto distance used

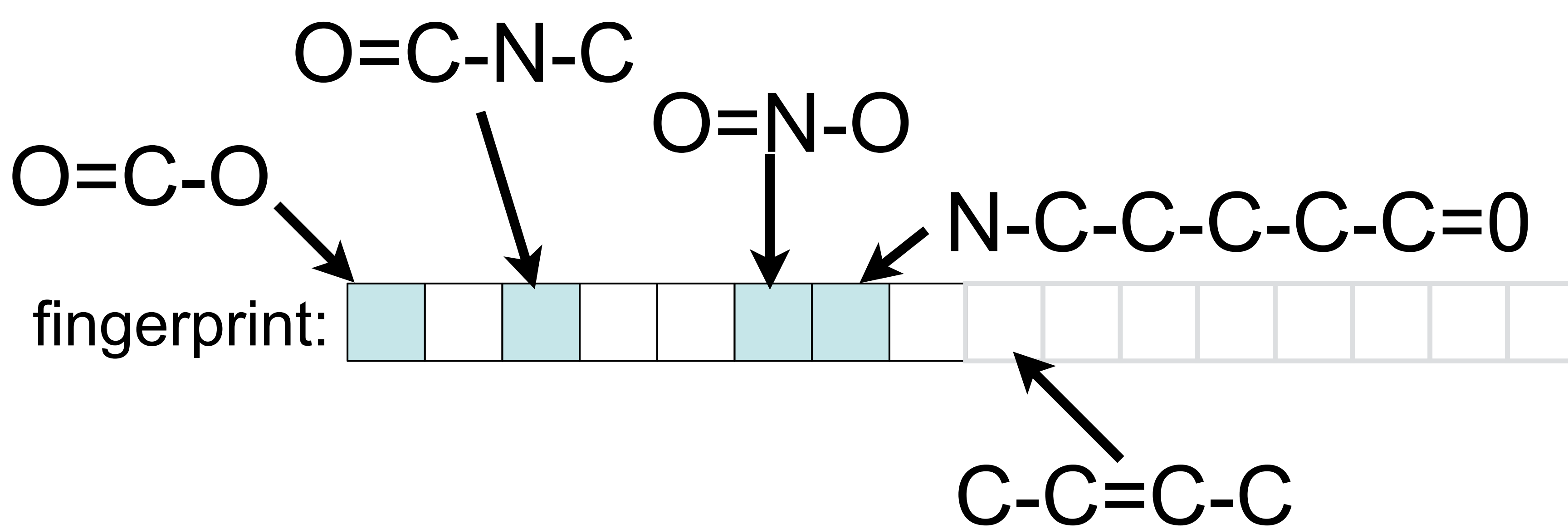
$$T(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Topological Fingerprints

Daylight/FP2 Fingerprints

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

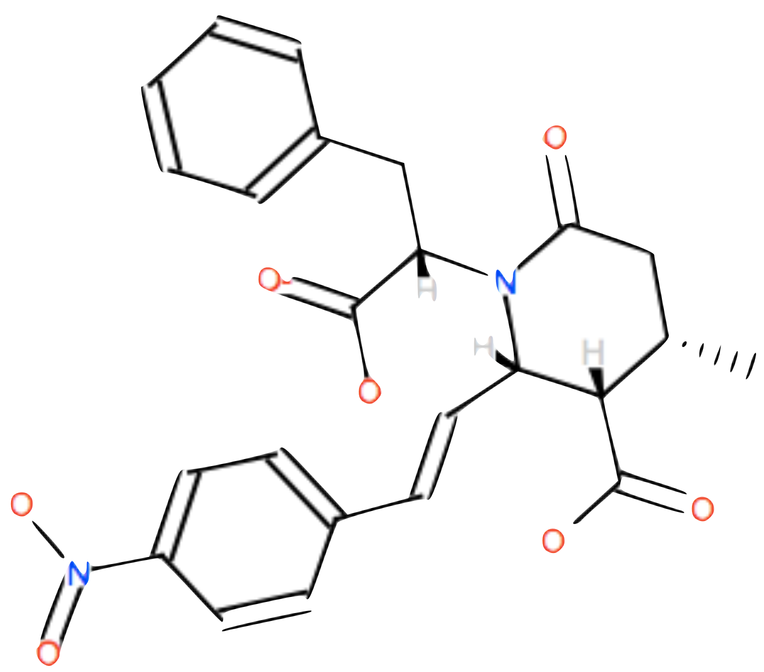


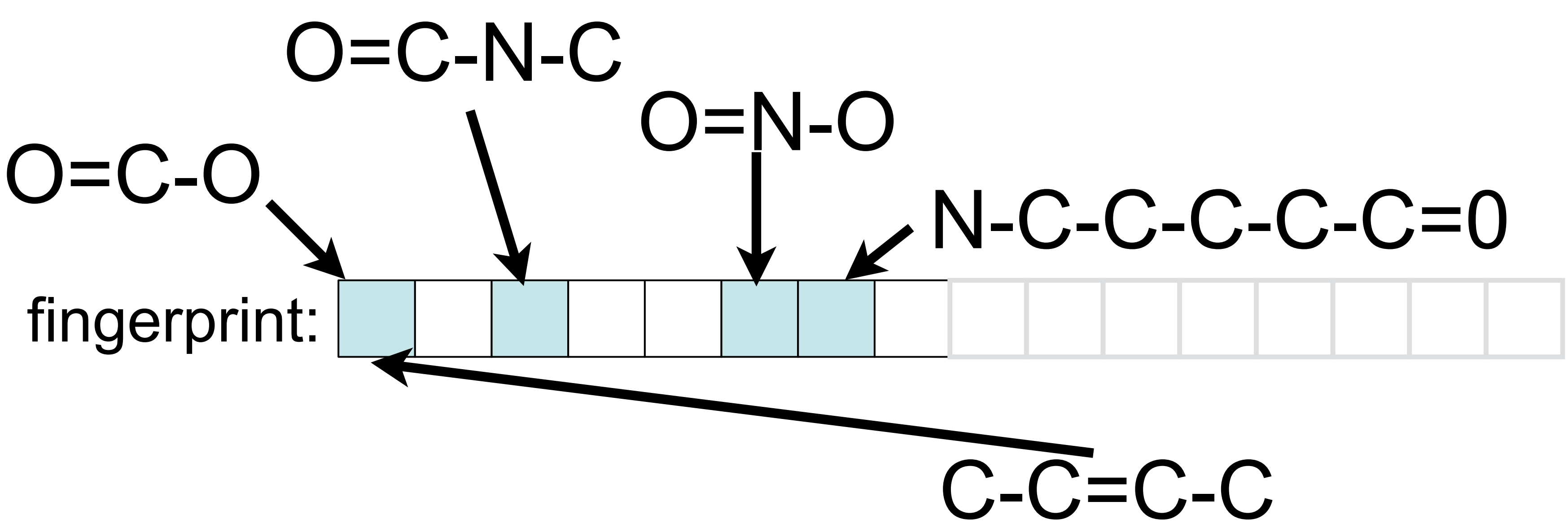


Topological Fingerprints

Daylight/FP2 Fingerprints

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

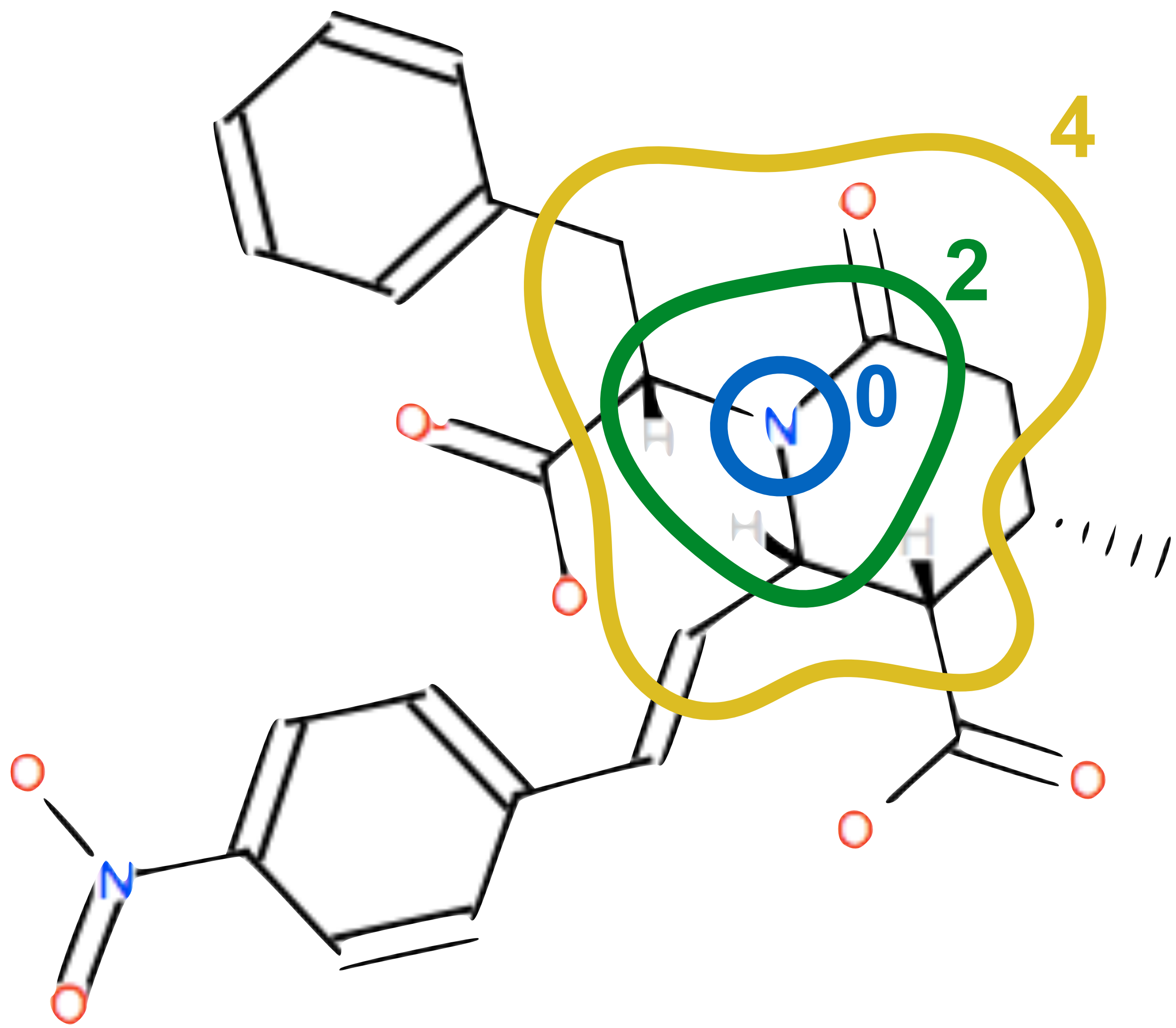




Topological Fingerprints

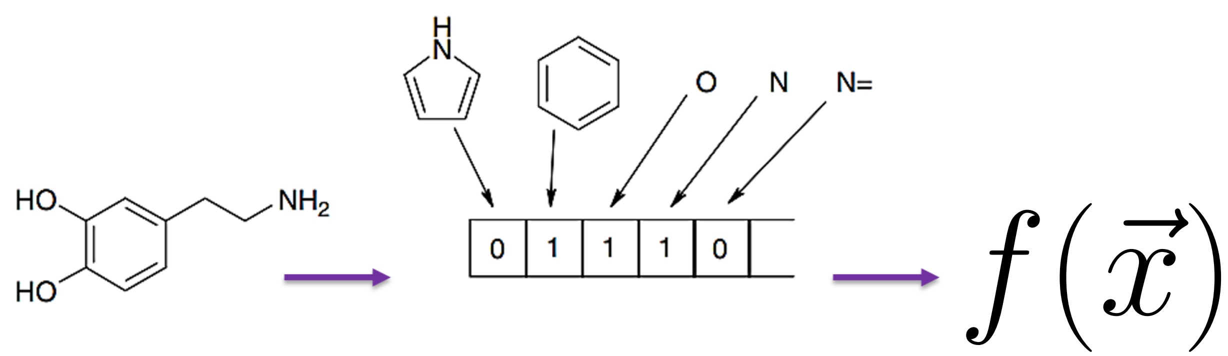
ECFP4

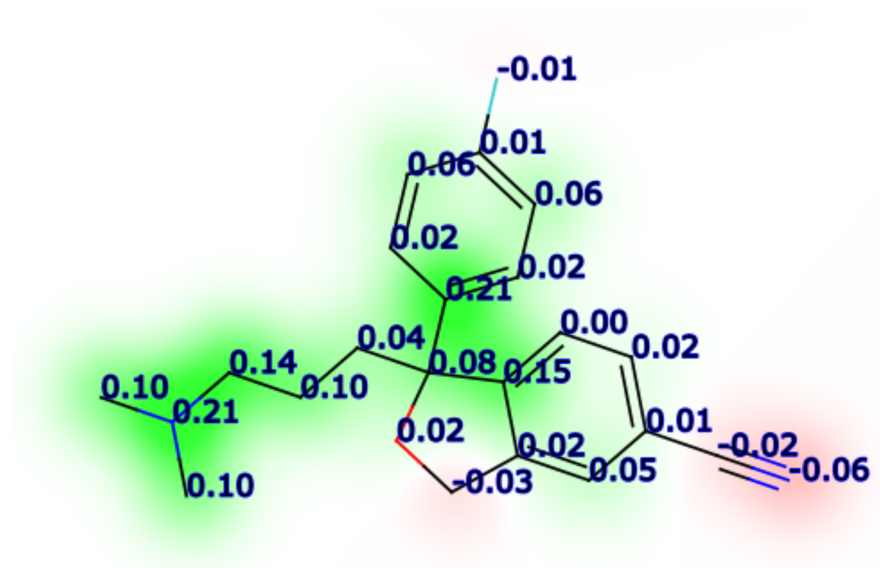
- all substructures with diameter 4 around every atom



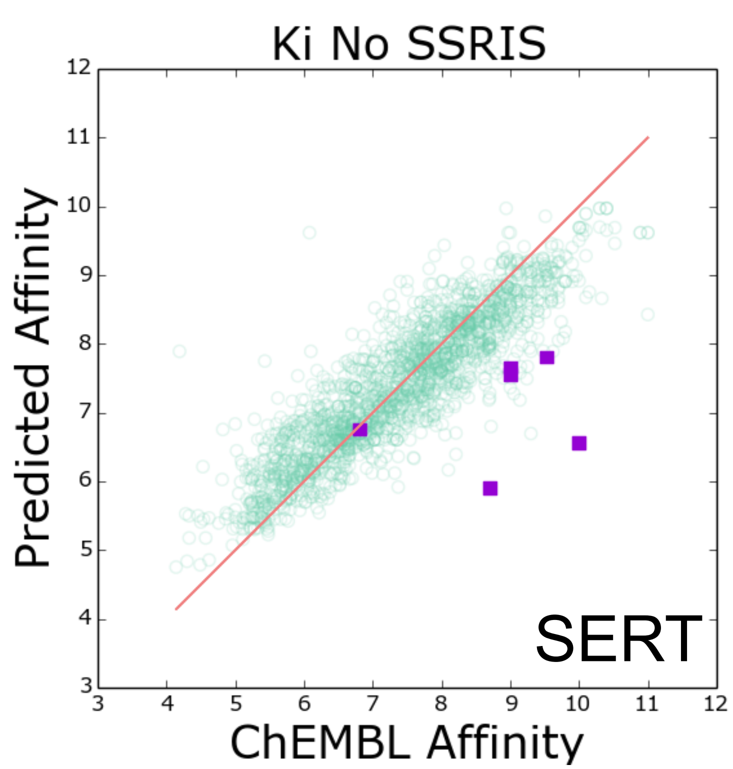
Ligand Based: QSAR

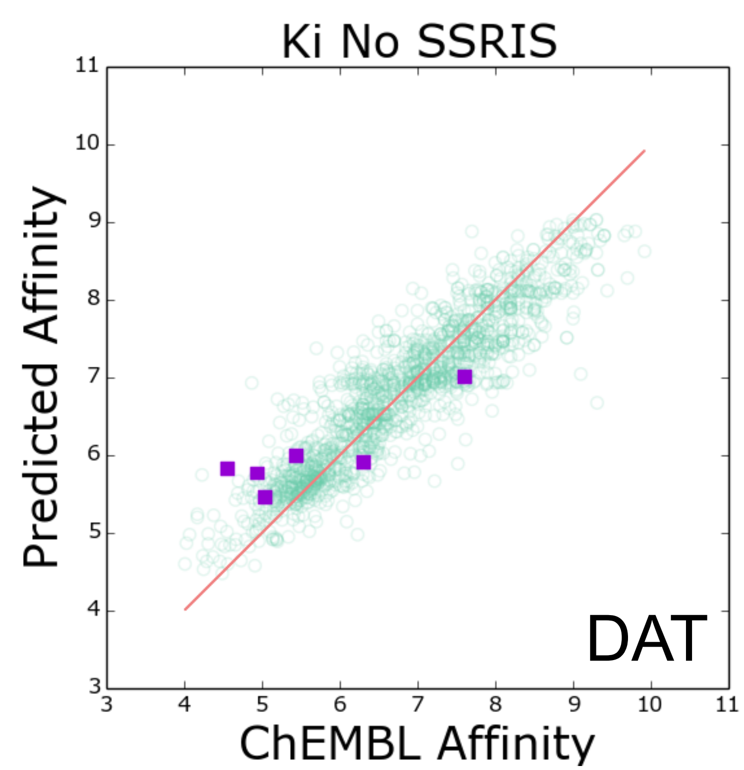
Quantitative Structure/Activity Relationships





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

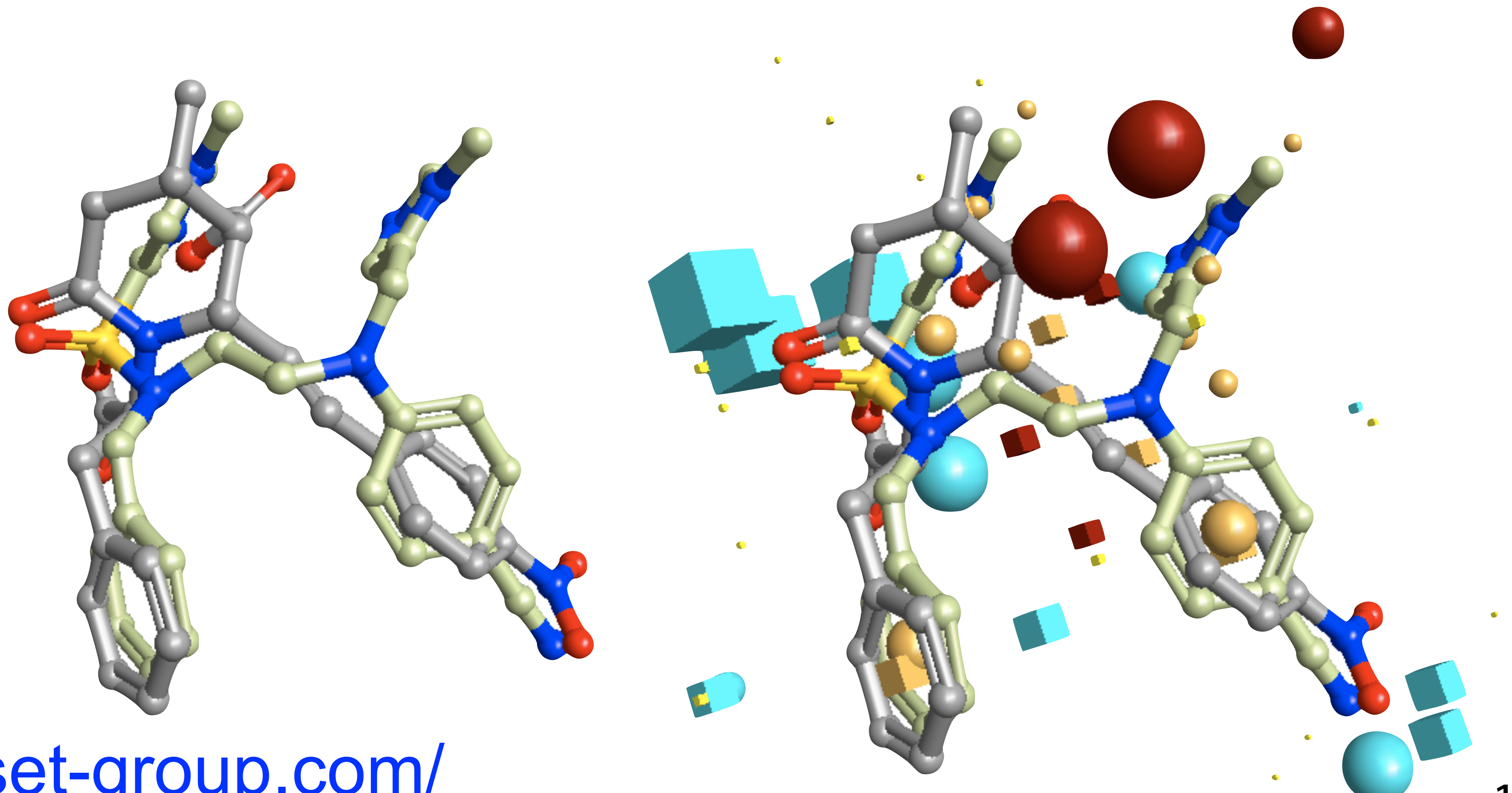




Ligand Based: Similarity

Superposition Methods

- compute "overlap" between molecules
- consider shape, electrostatics, **pharmacophores**

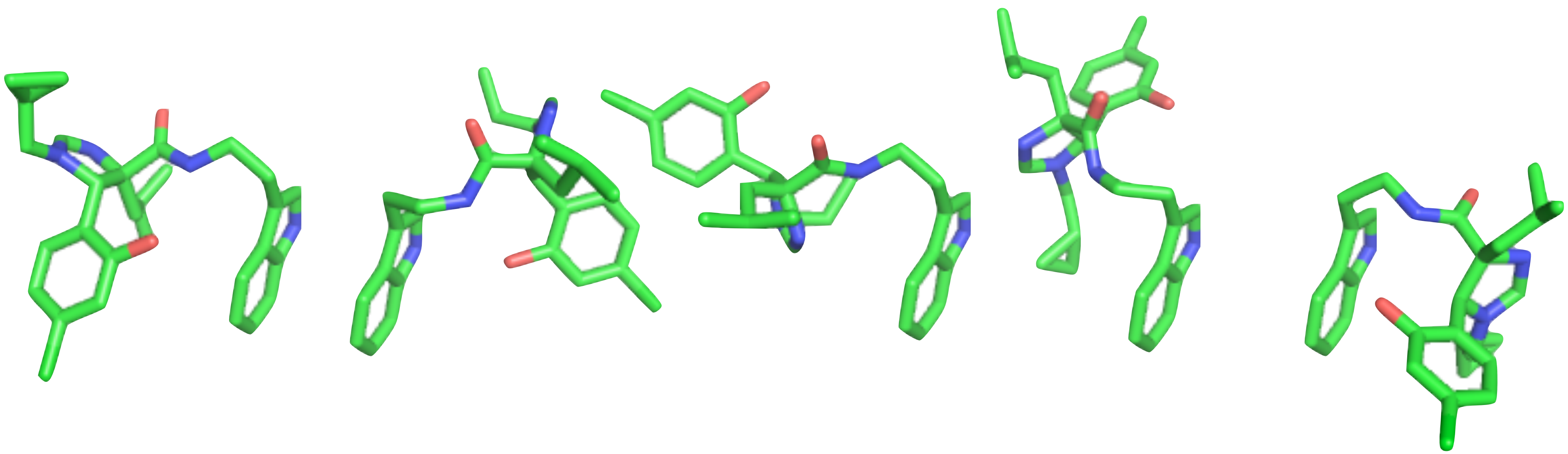


http://www.cresset-group.com/

Representing Compounds

Conformations

A single compound has many different shapes



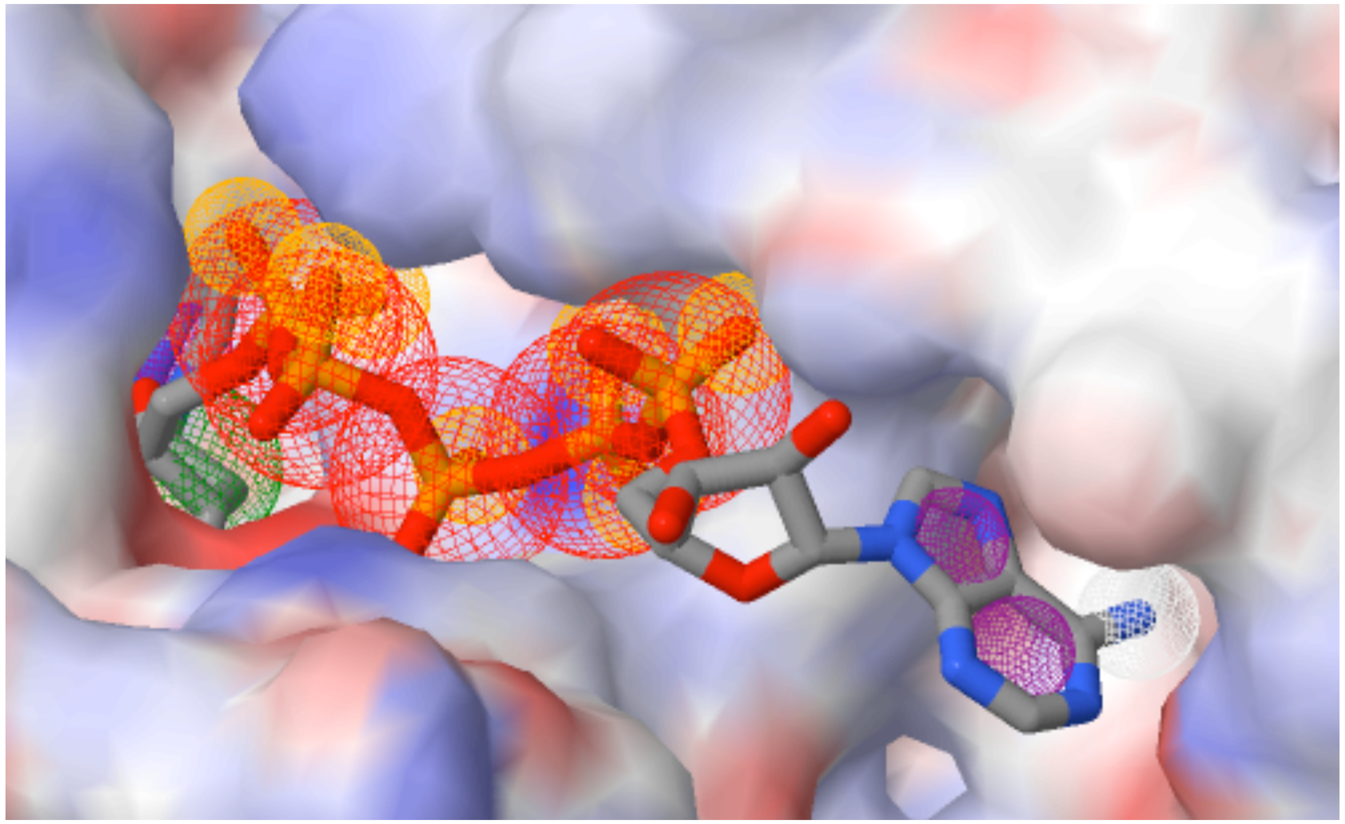
Choices: Store sampling of explicit conformations, search for a good conformation, ignore conformations (2D only)

Ligand Based: Pharmacophore

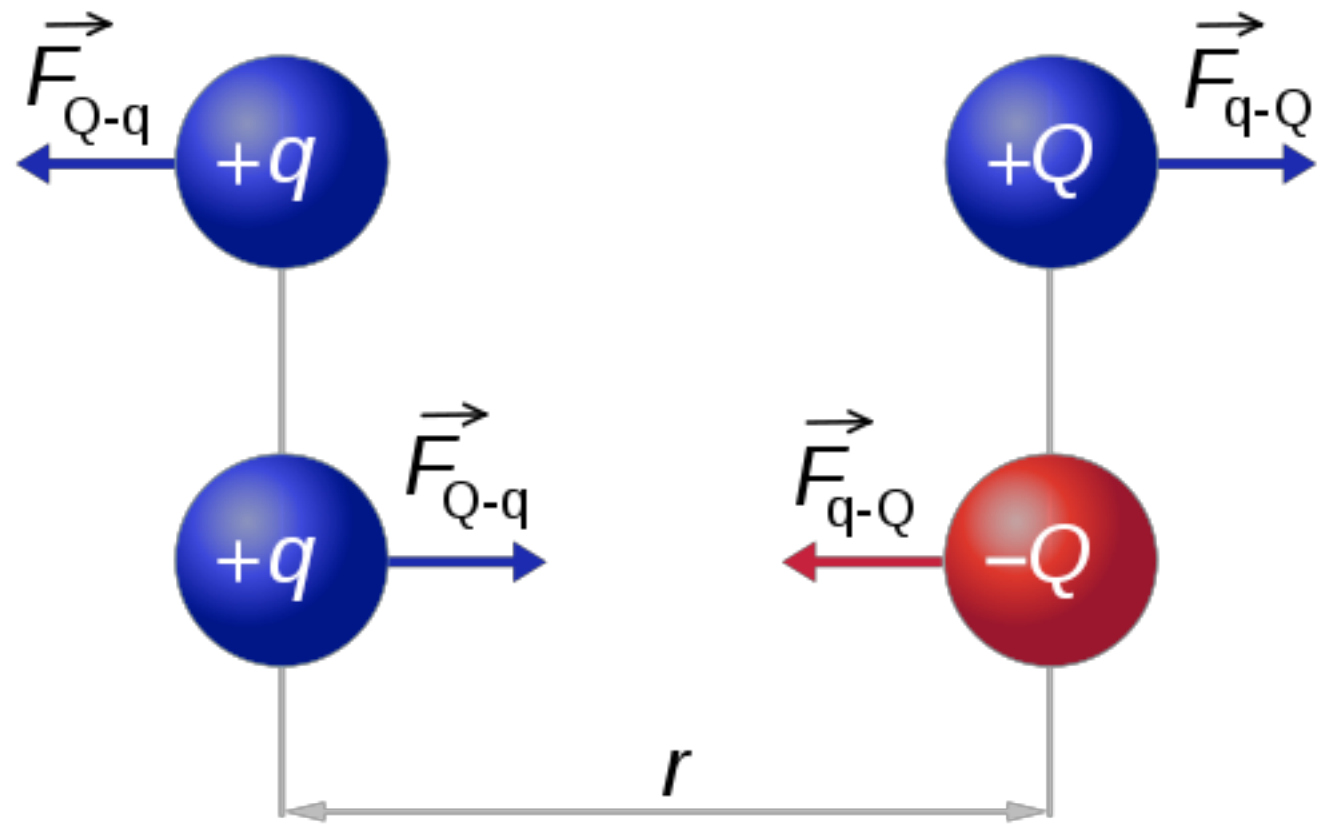
Pharmacophore:

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.

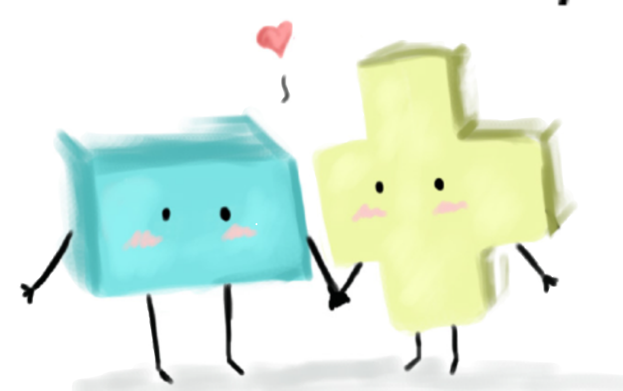
Common Features:
aromatic ring
hydrophobic area
positive ionizable
negative ionizable
hydrogen bond donor
hydrogen bond acceptor



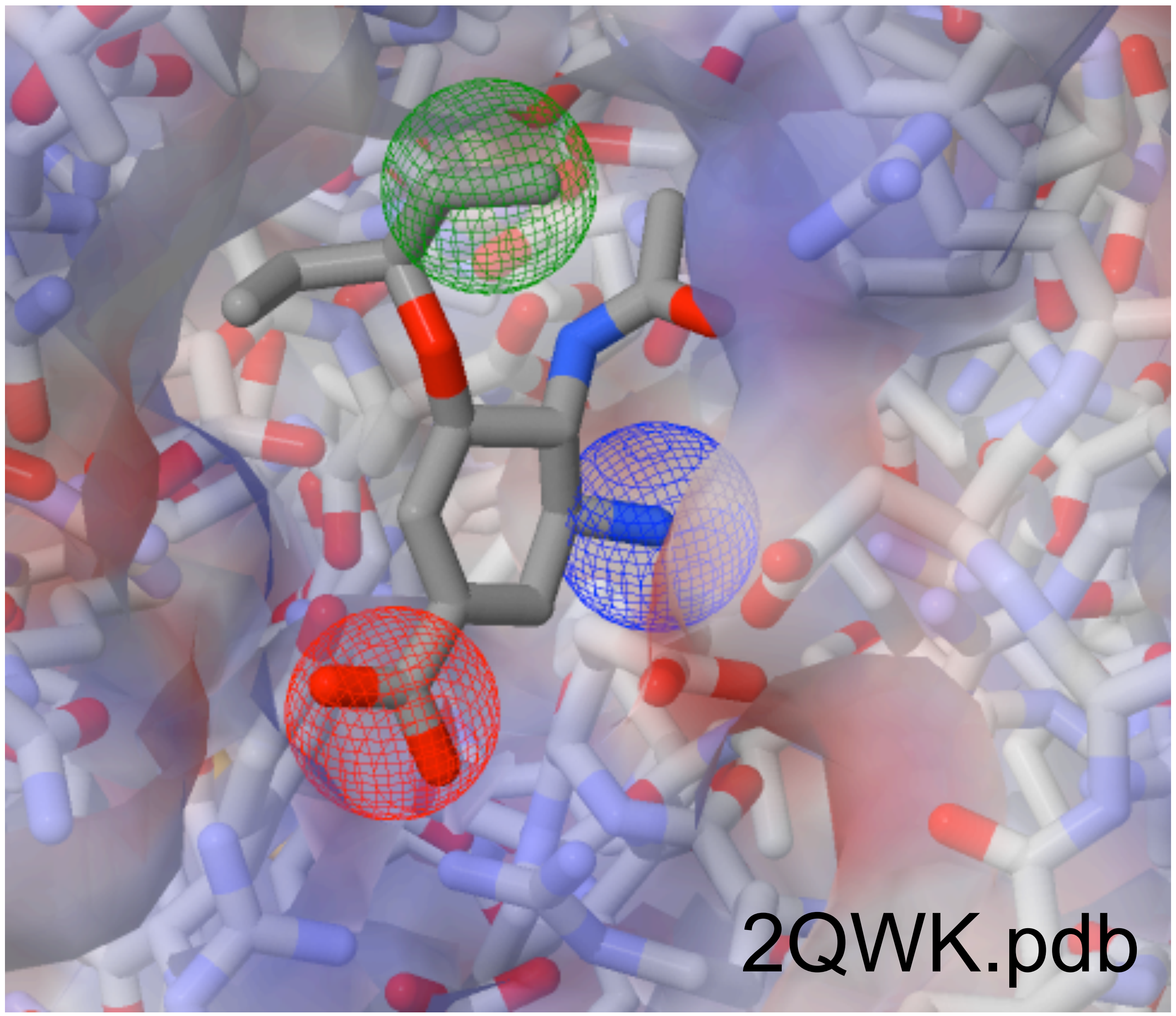
Charge-Charge

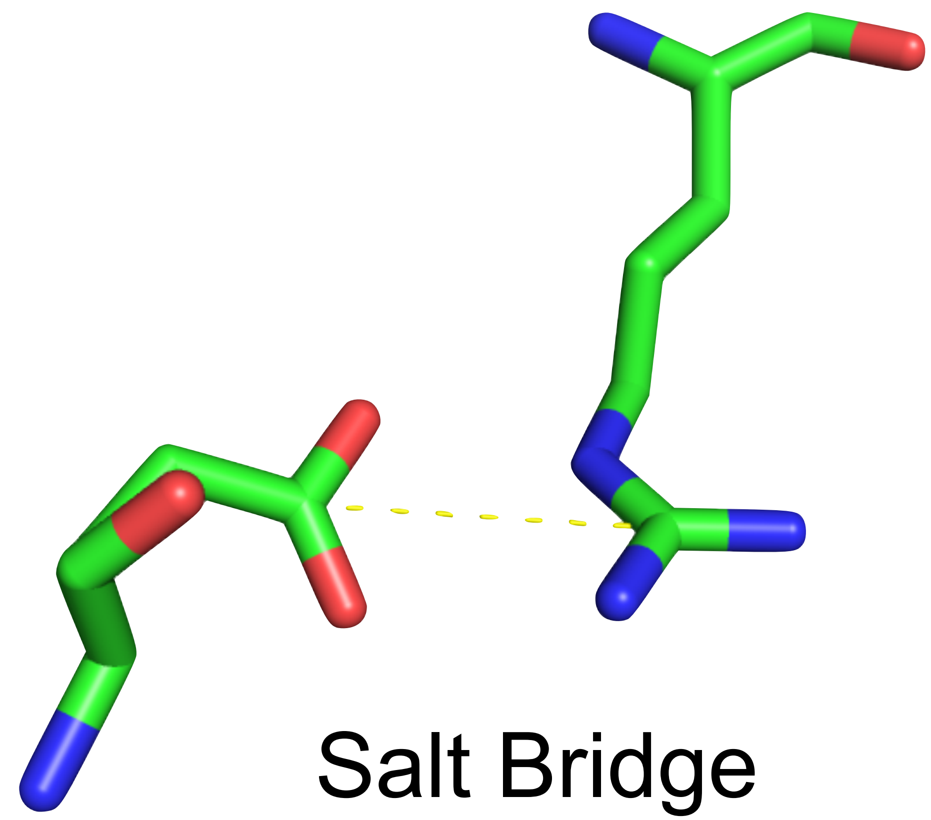


$$|\vec{F}_{Q\text{-}q}| = |\vec{F}_{q\text{-}Q}| = k\,\frac{|q \times Q|}{r^2}$$

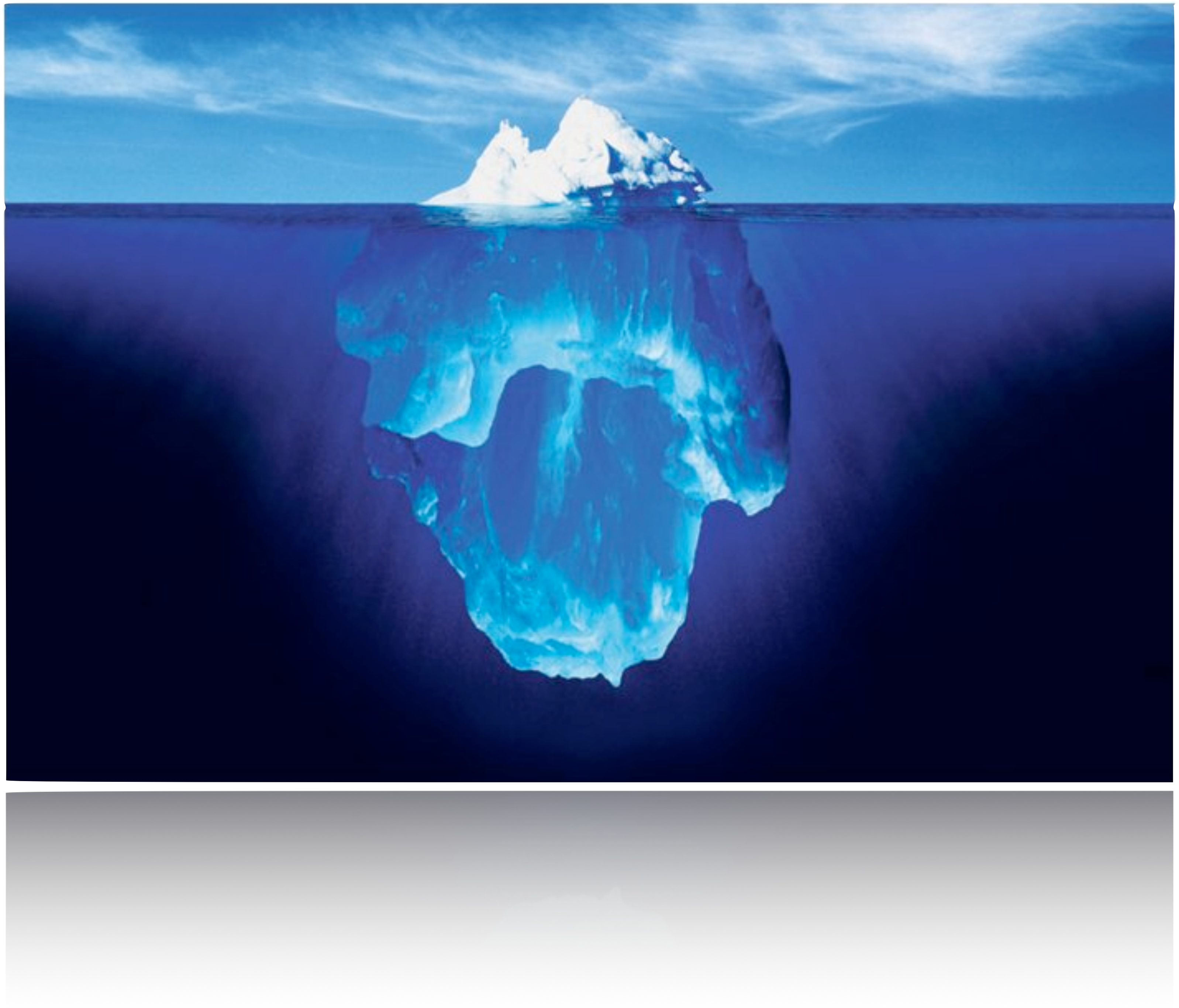


Charge-Charge

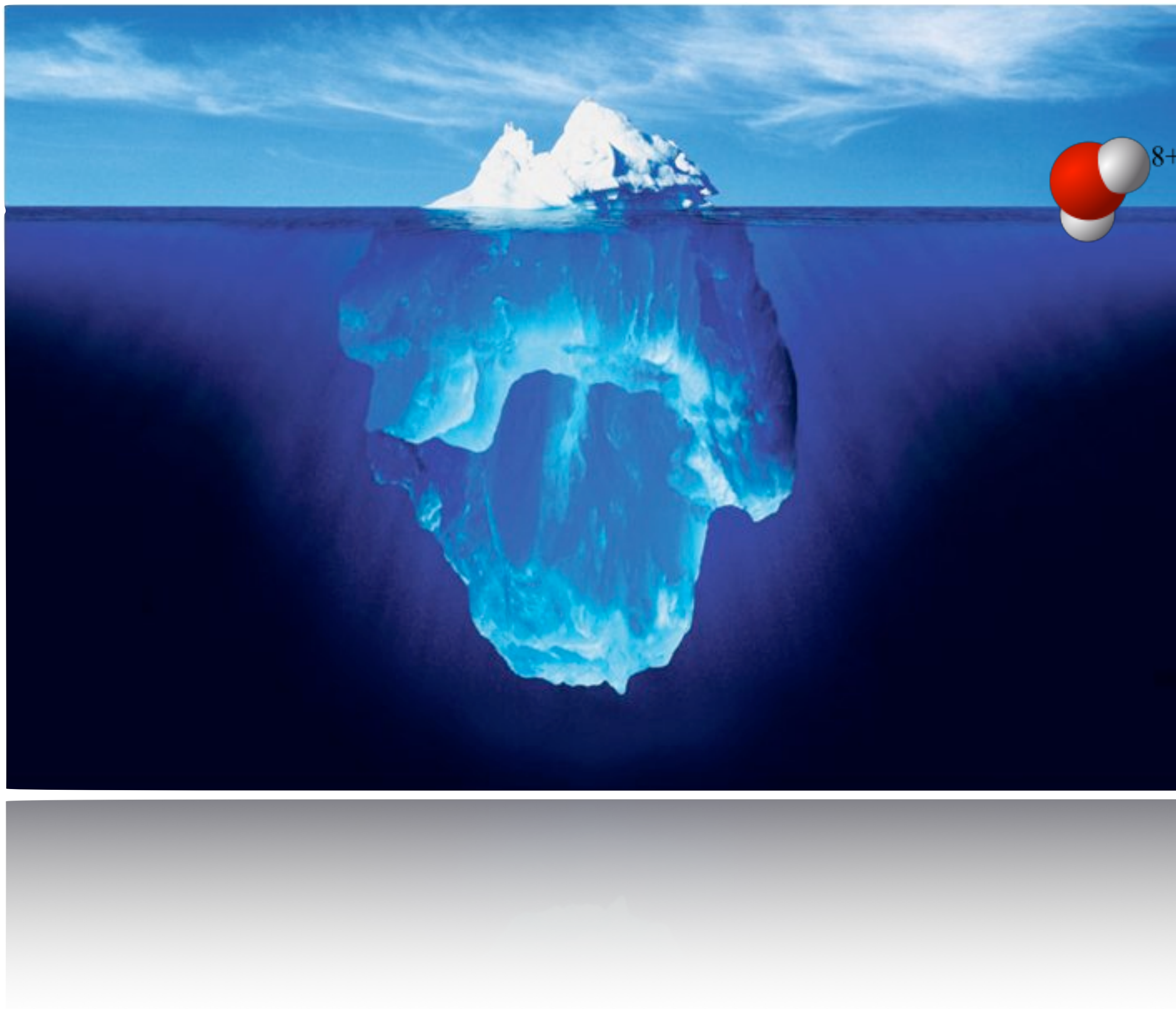


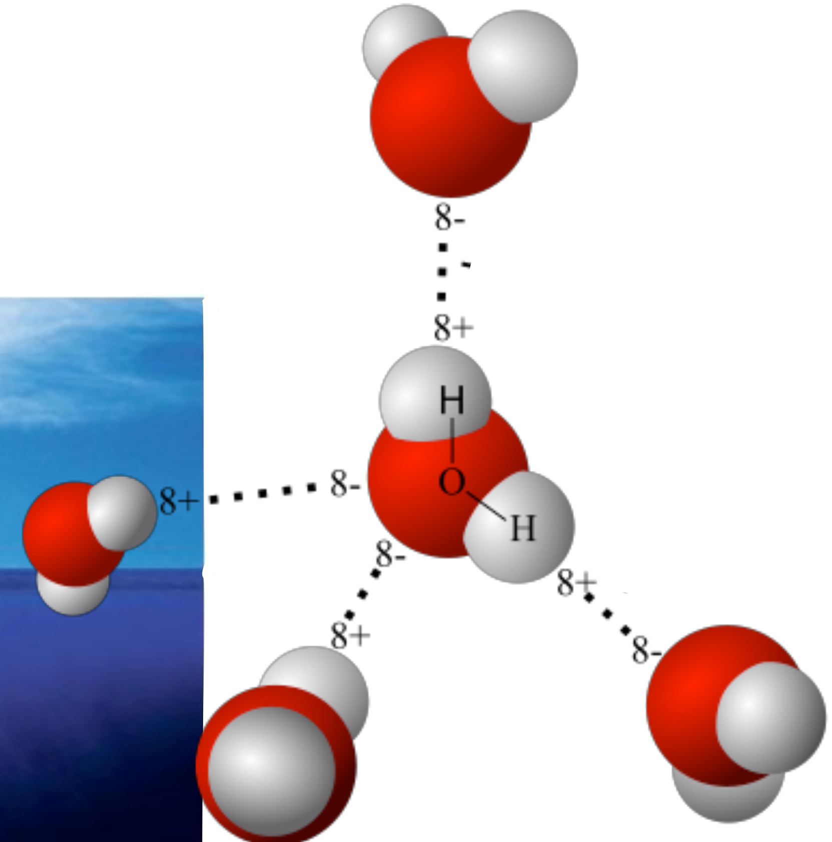


Hydrogen Bond

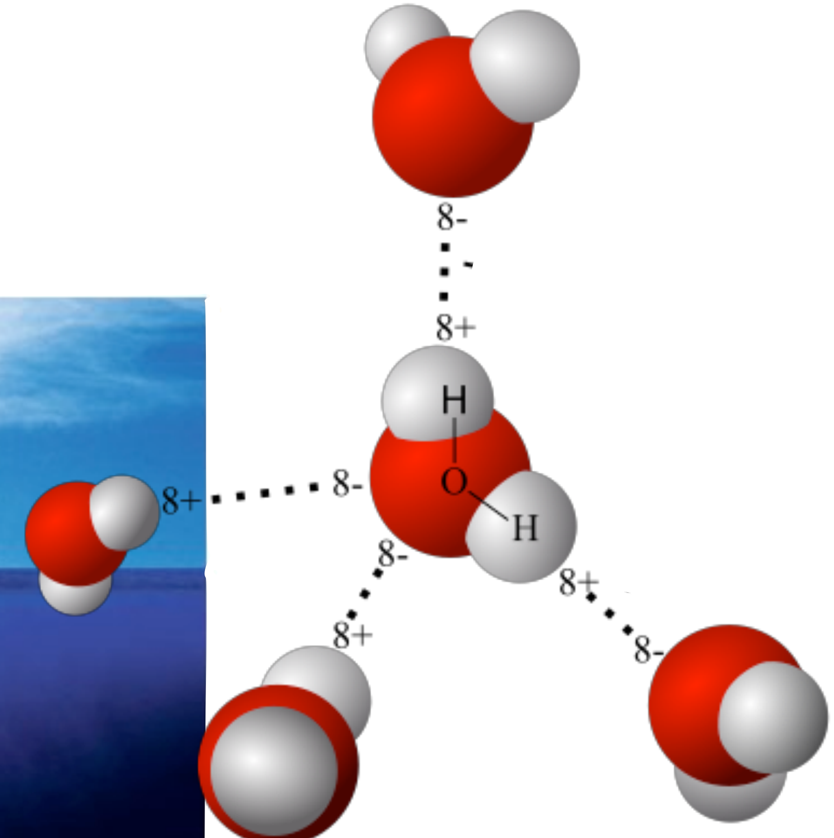


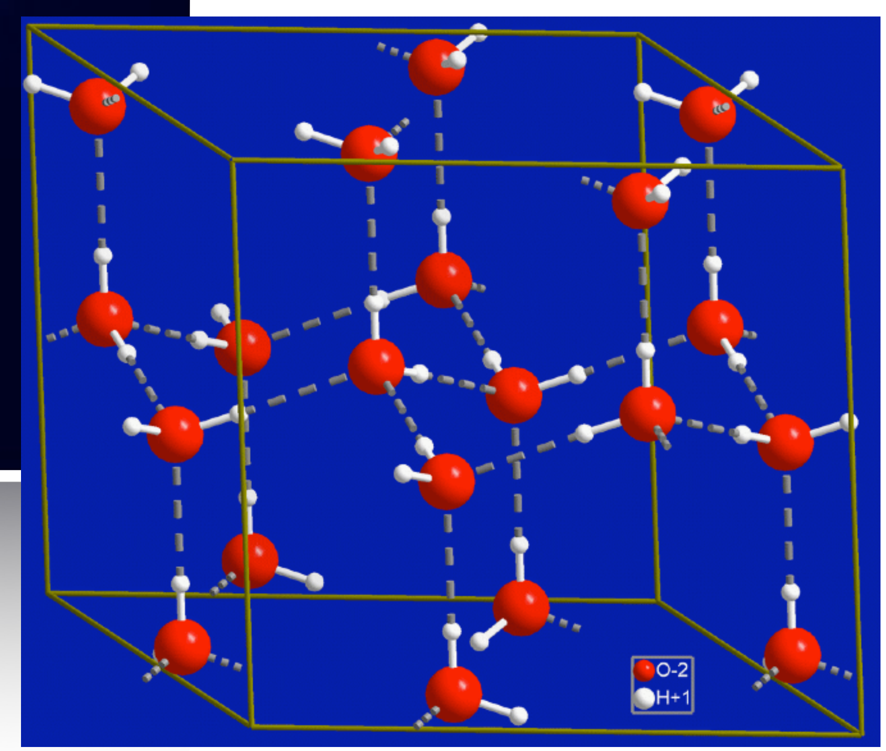
Hydrogen Bond



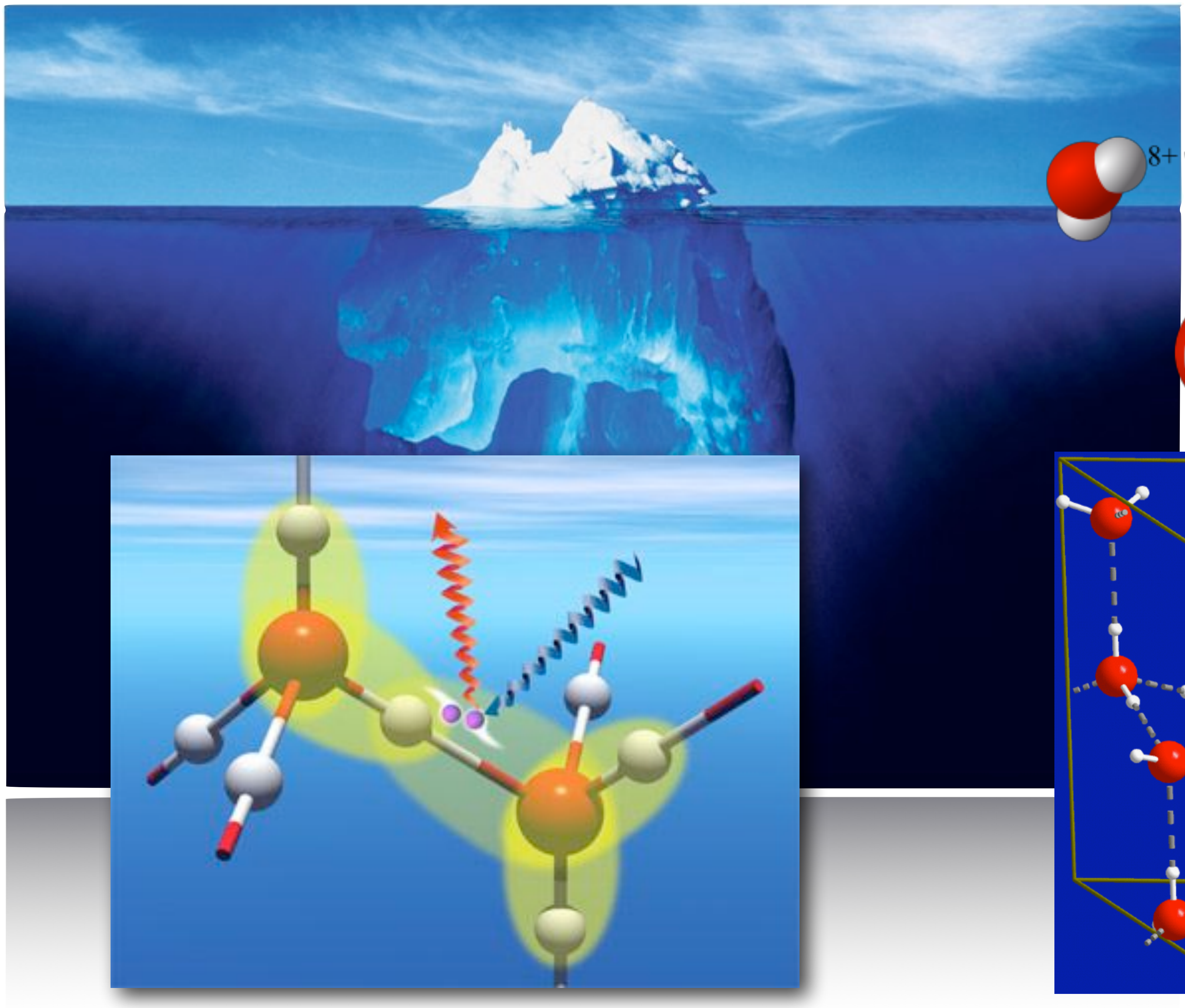


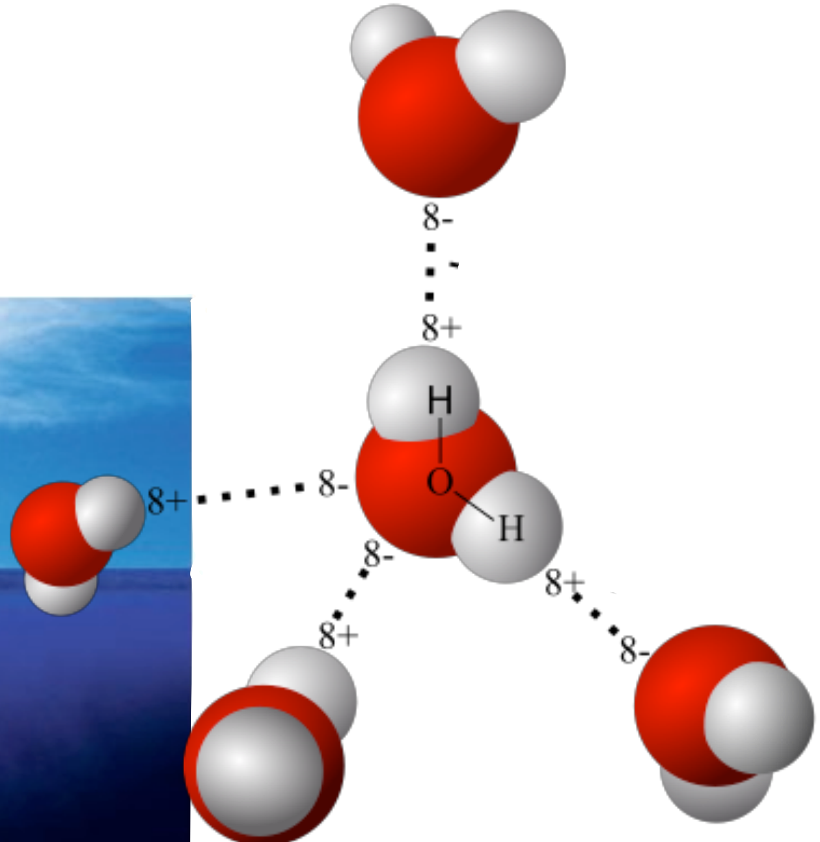
Hydrogen Bond

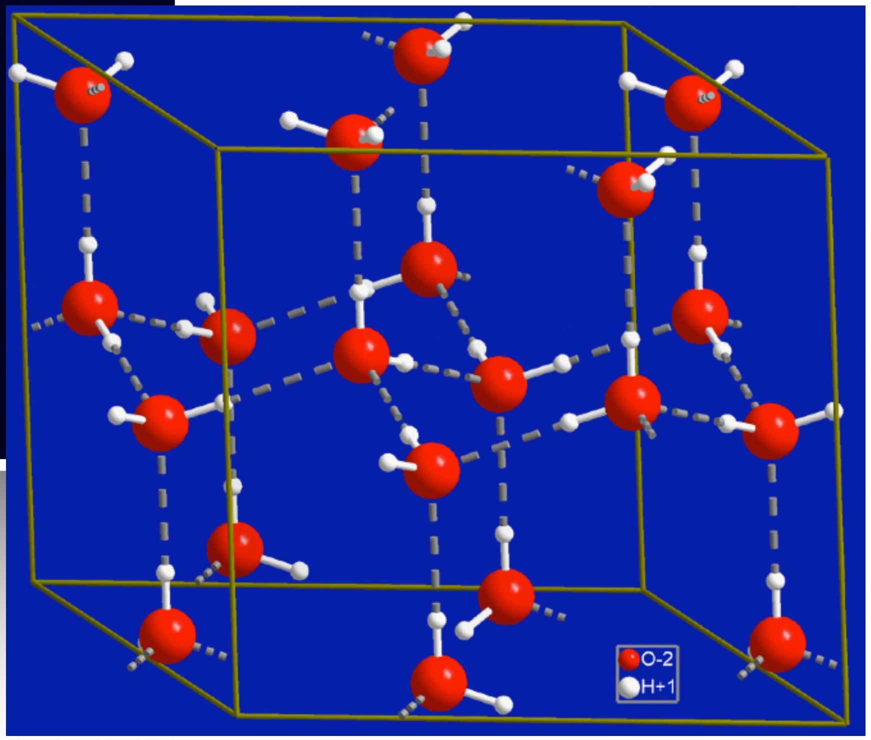


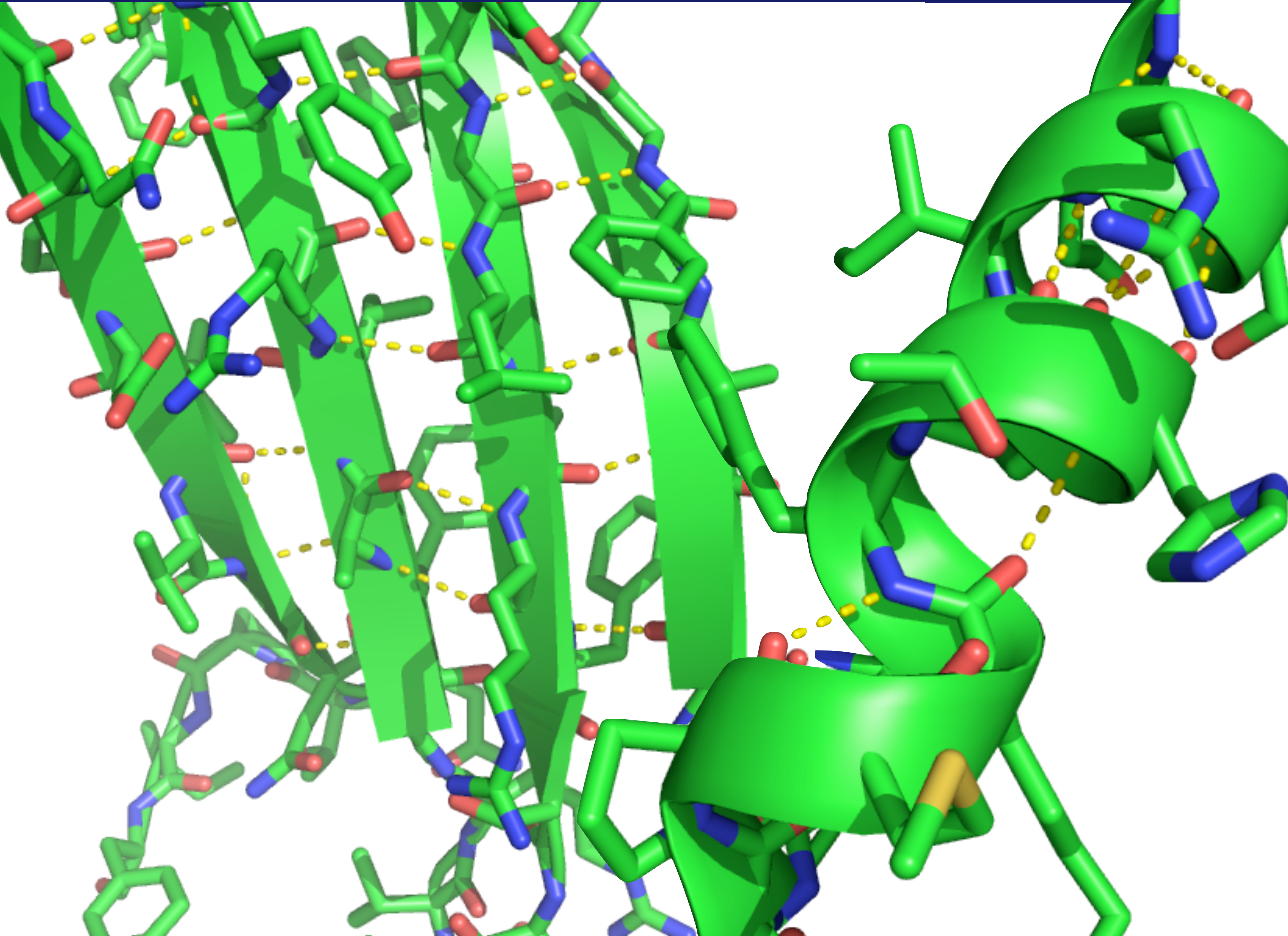


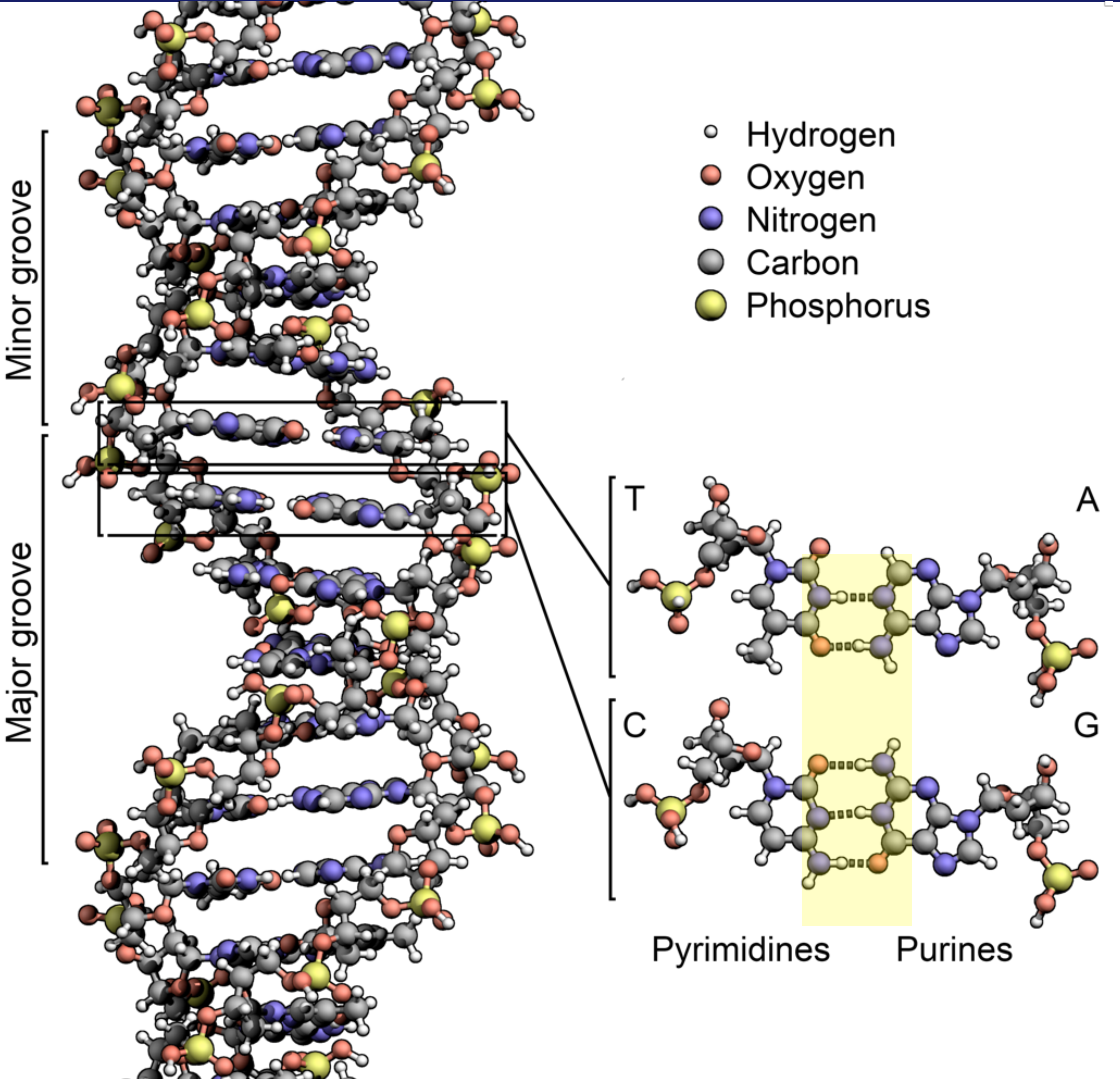
Hydrogen Bond



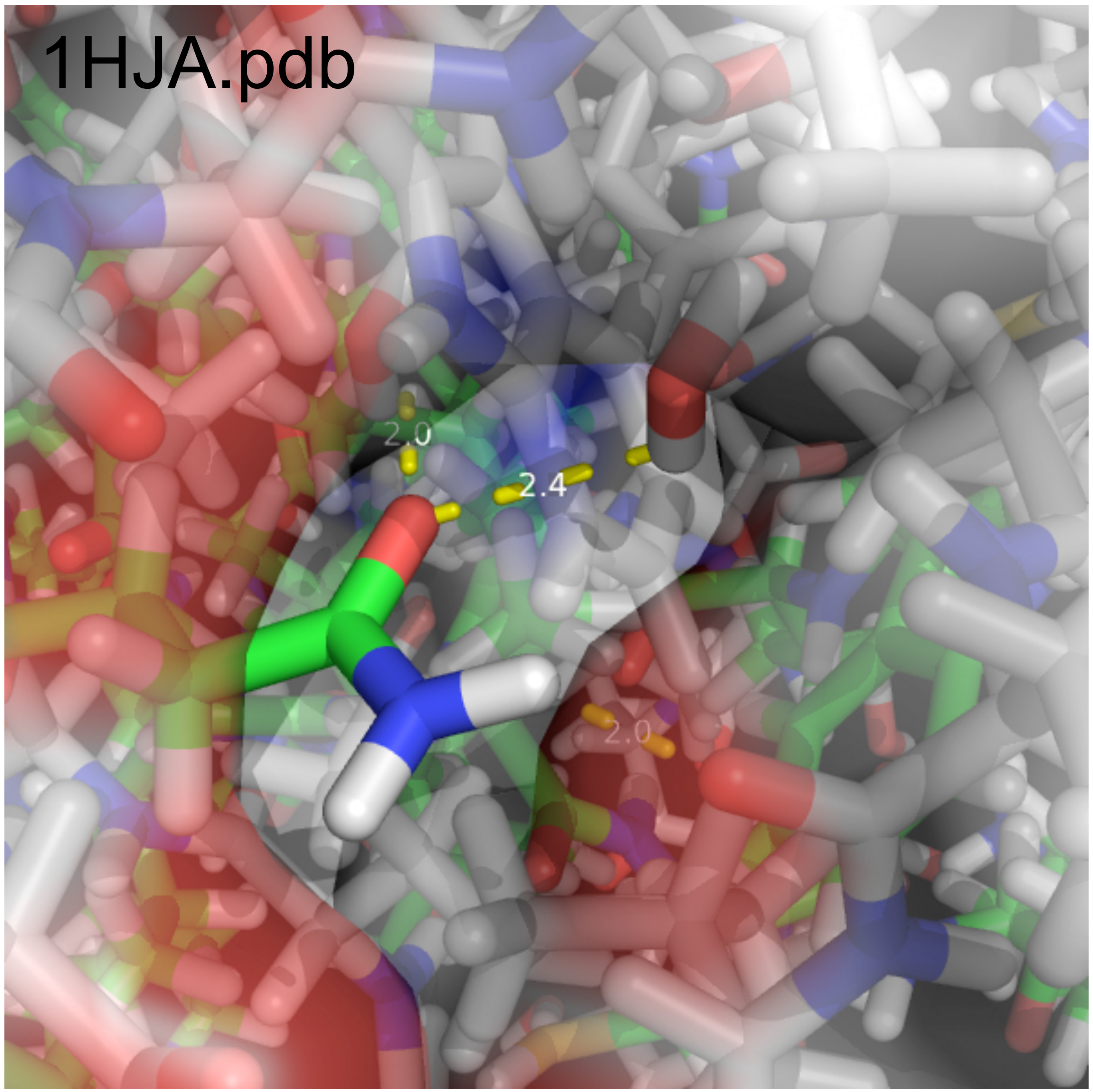







Minor groove
Major groove
Hydrogen
Oxygen
Nitrogen
Carbon
Phosphorus
T
A
C
G
Pyrimidines
Purines

Hydrogen Bond



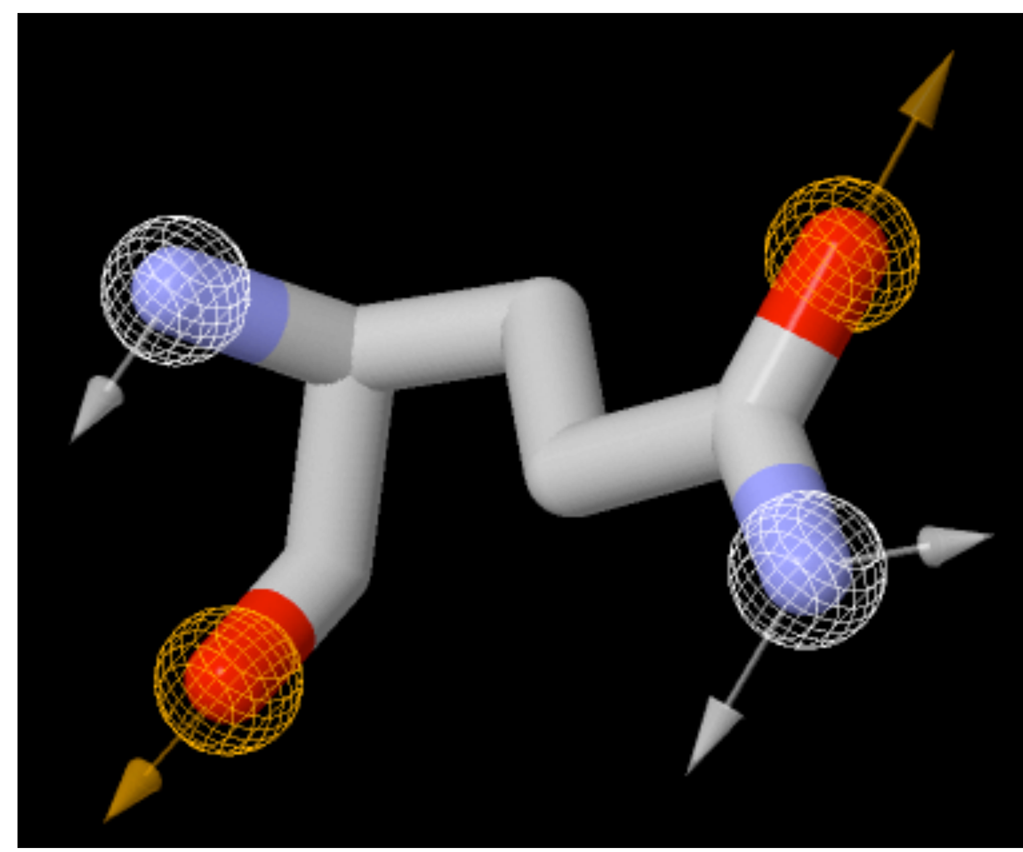
Distance:

D-A: 2.5Å – 3.5Å (4.0Å?)

H-A: 1.5Å – 2.5Å

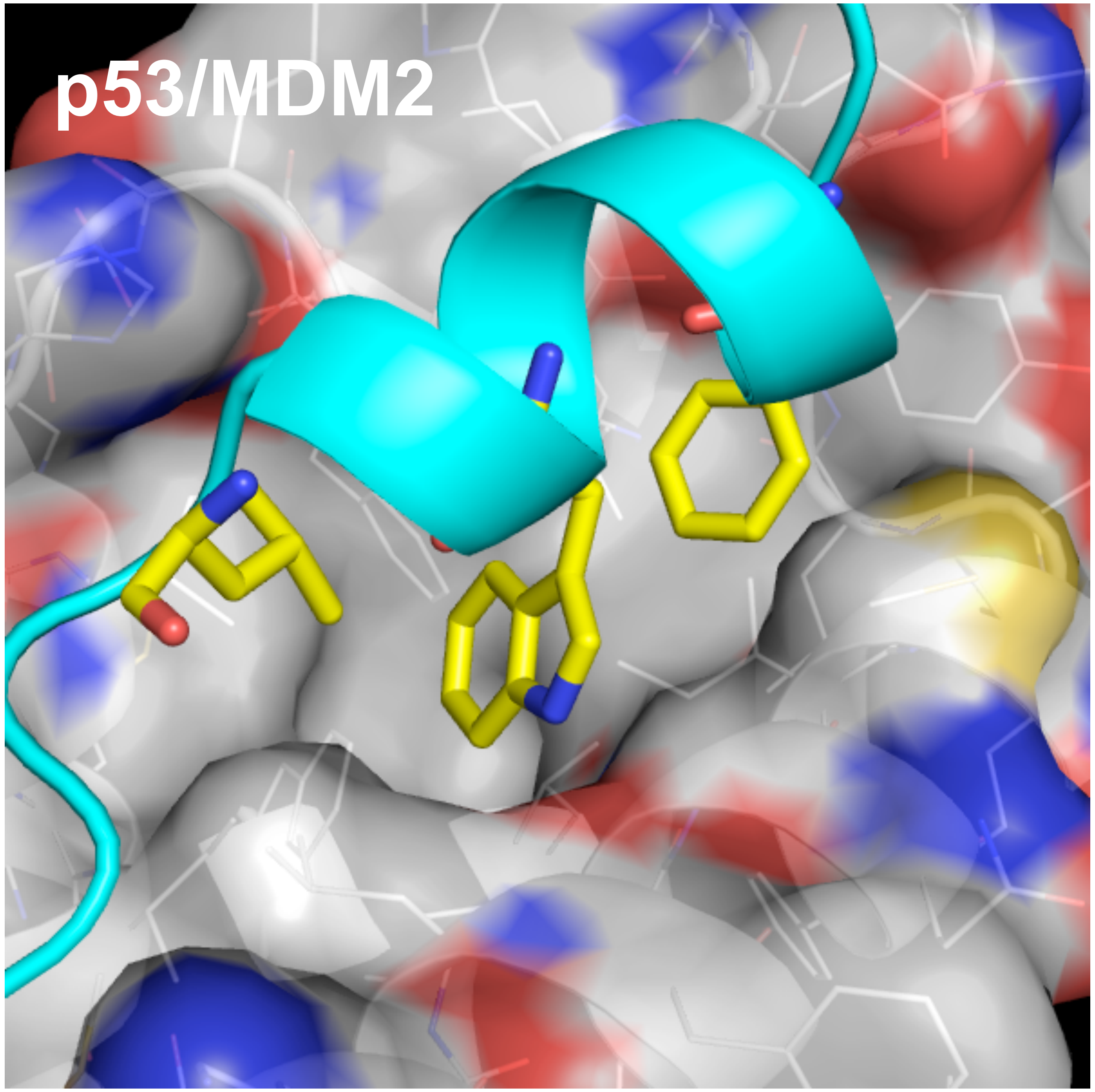
Angle:

Depends on context

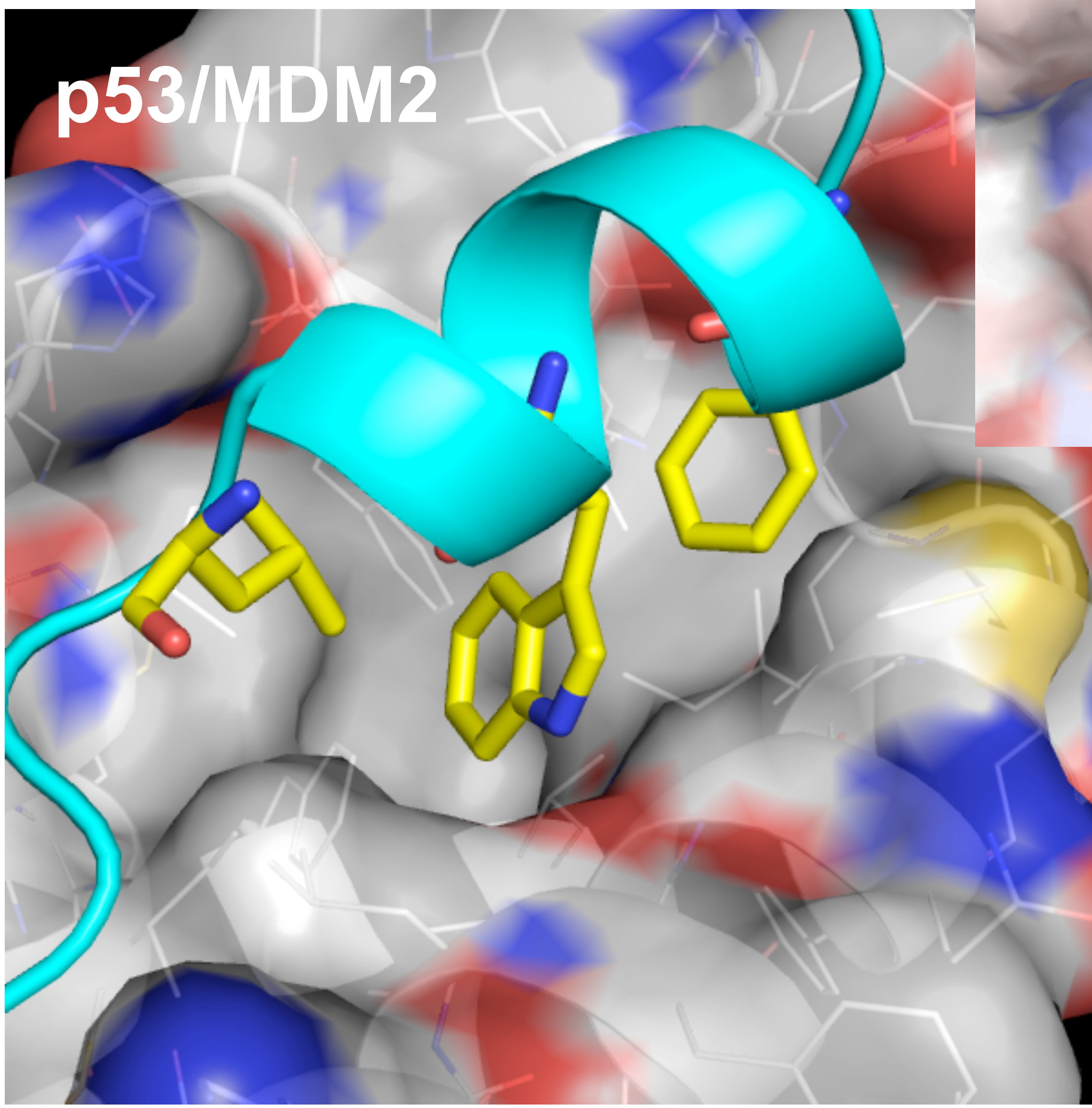


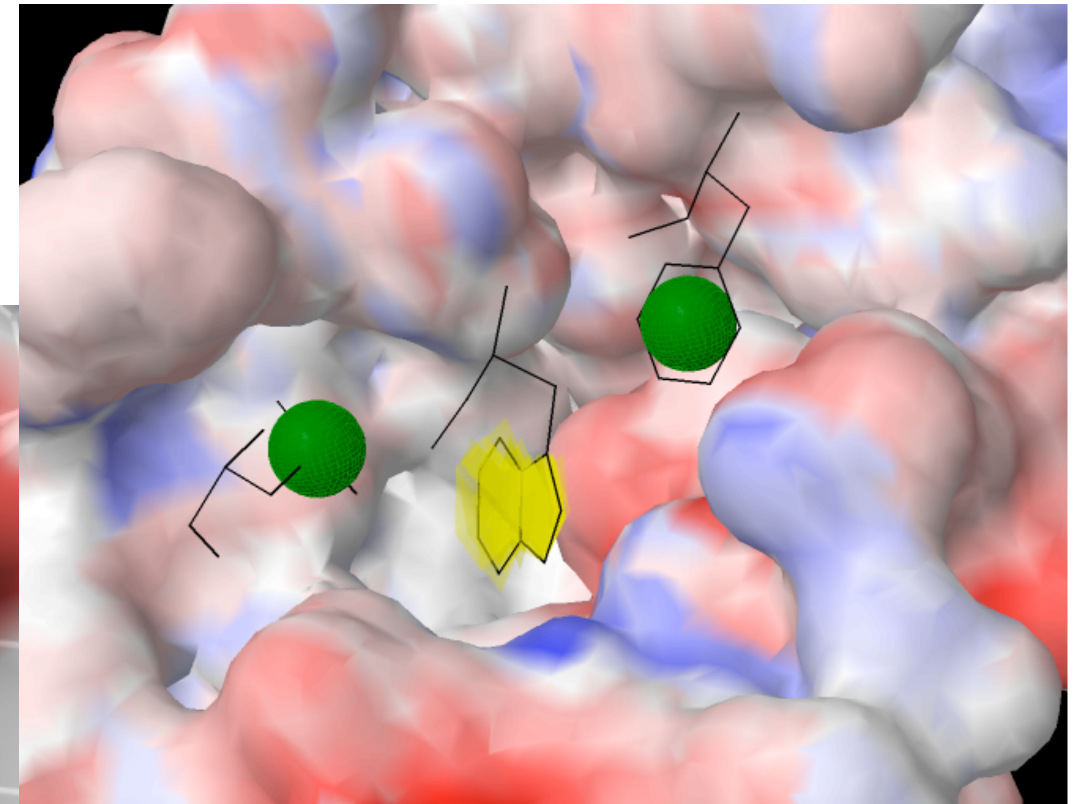
Hydrophobic

Hydrophobic

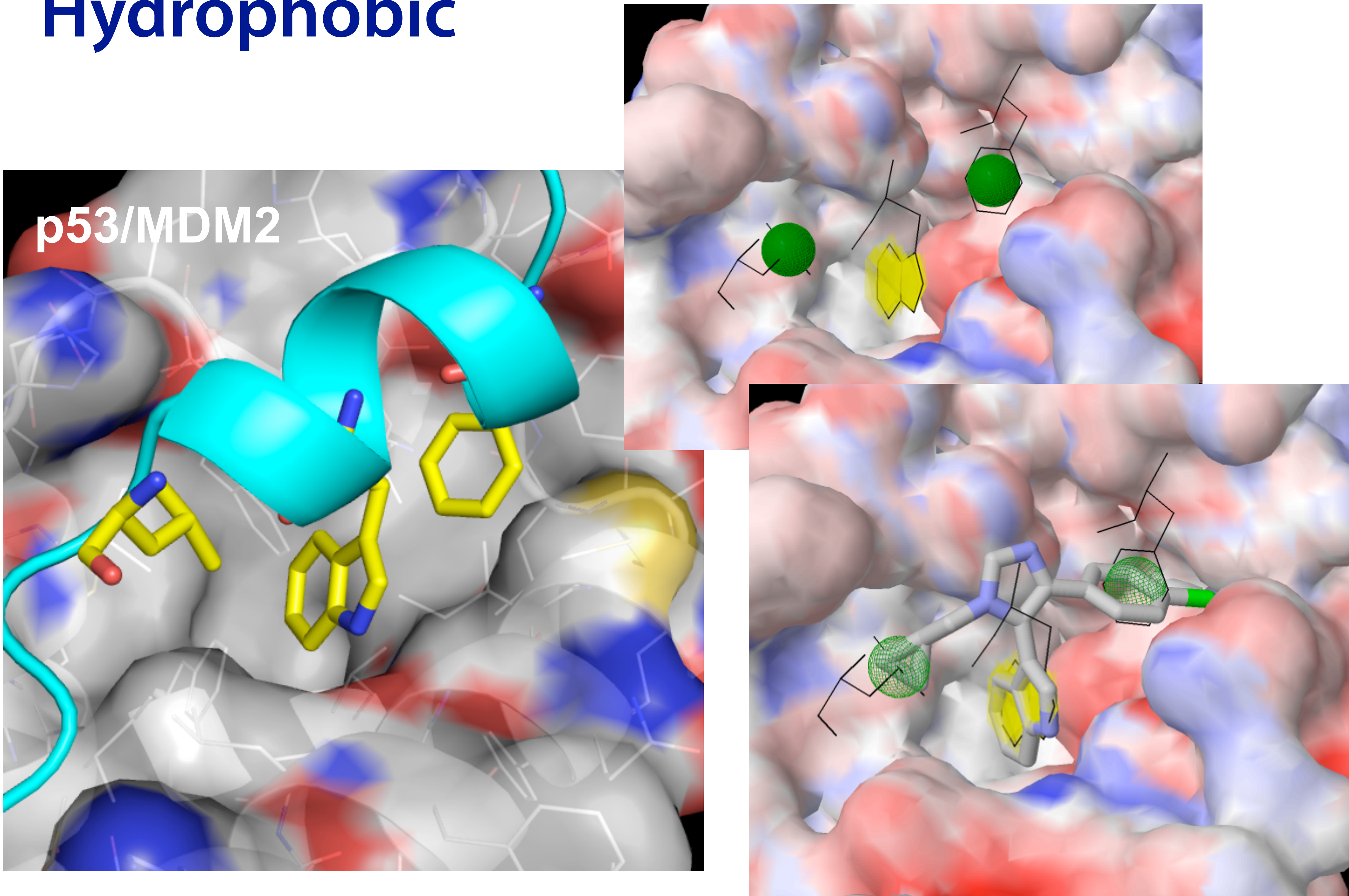


Hydrophobic

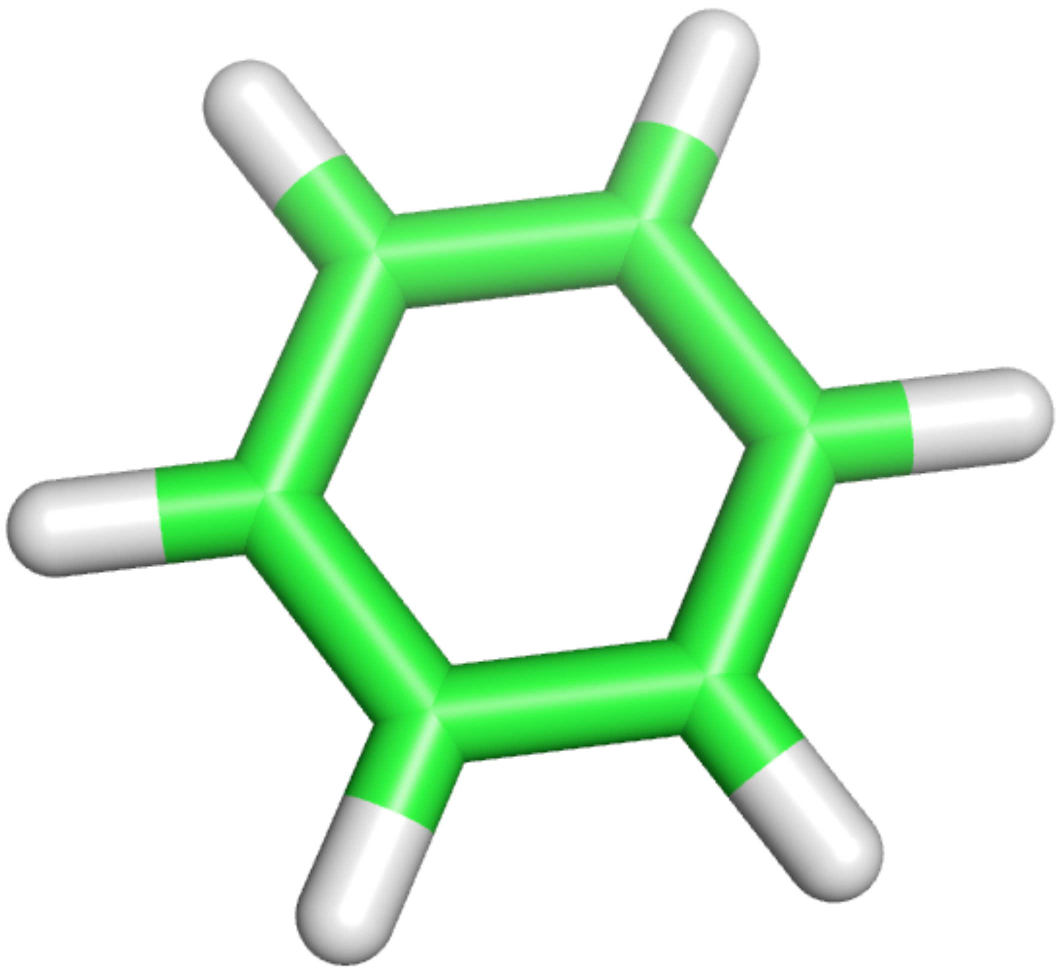




Hydrophobic

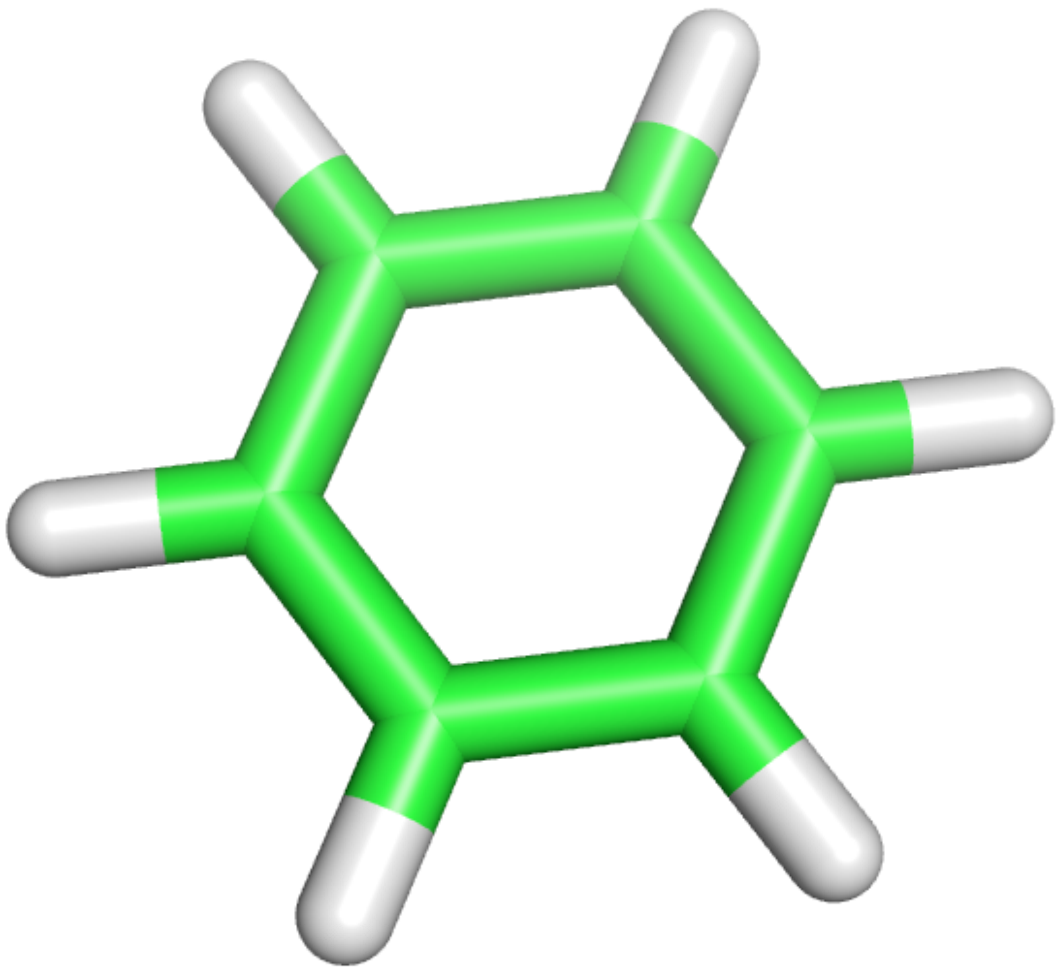


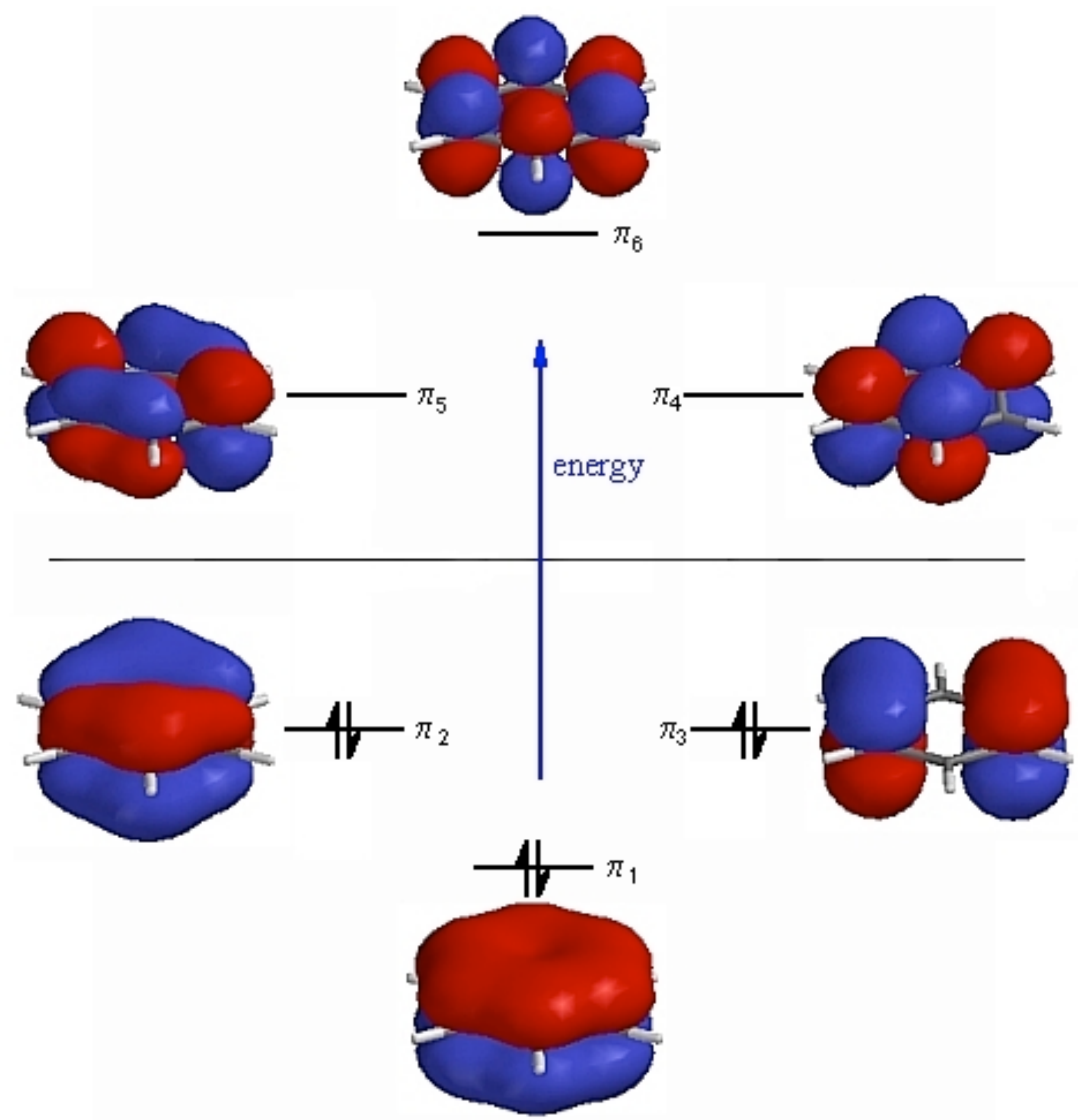
Aromatic



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

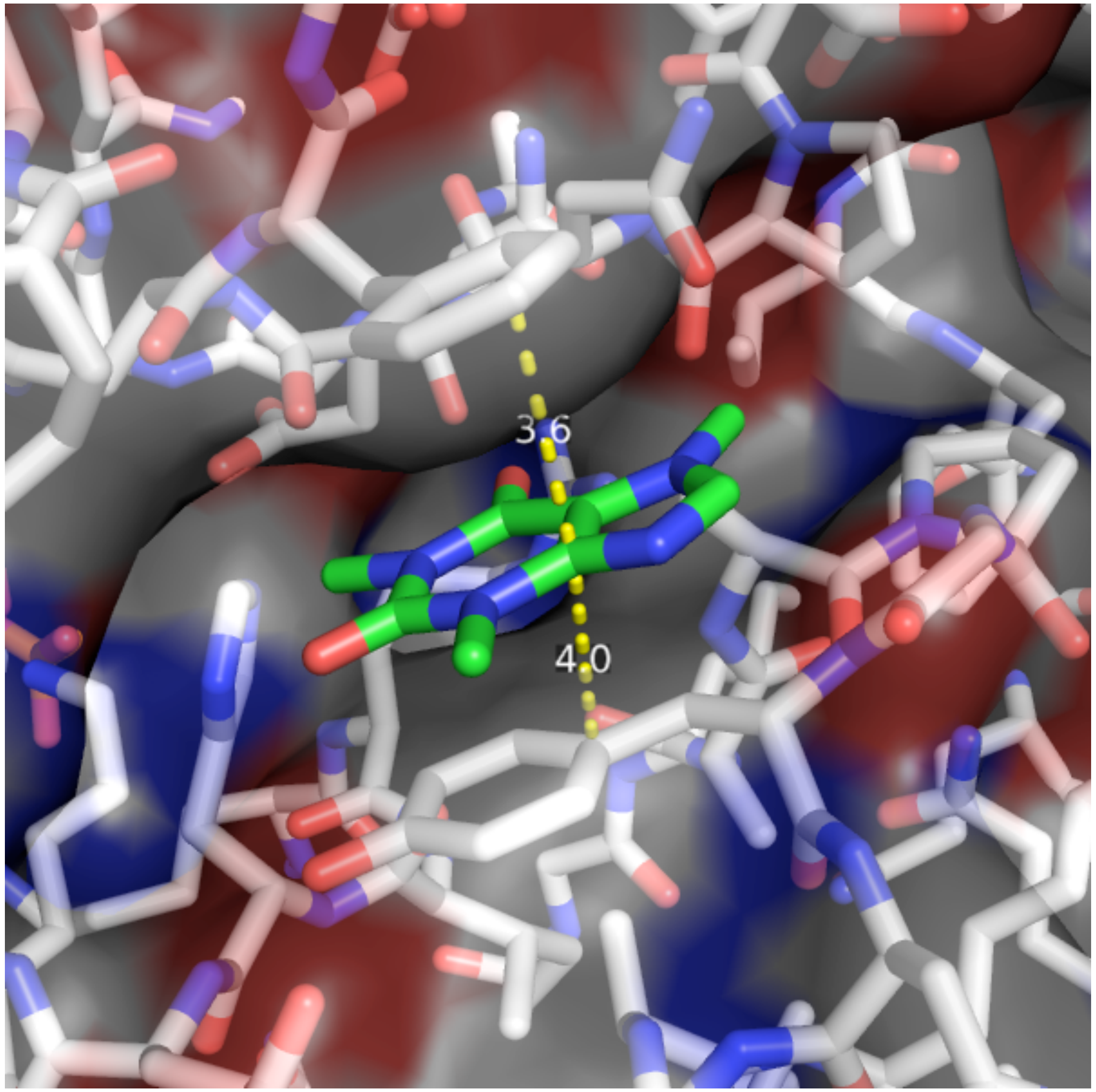
Aromatic



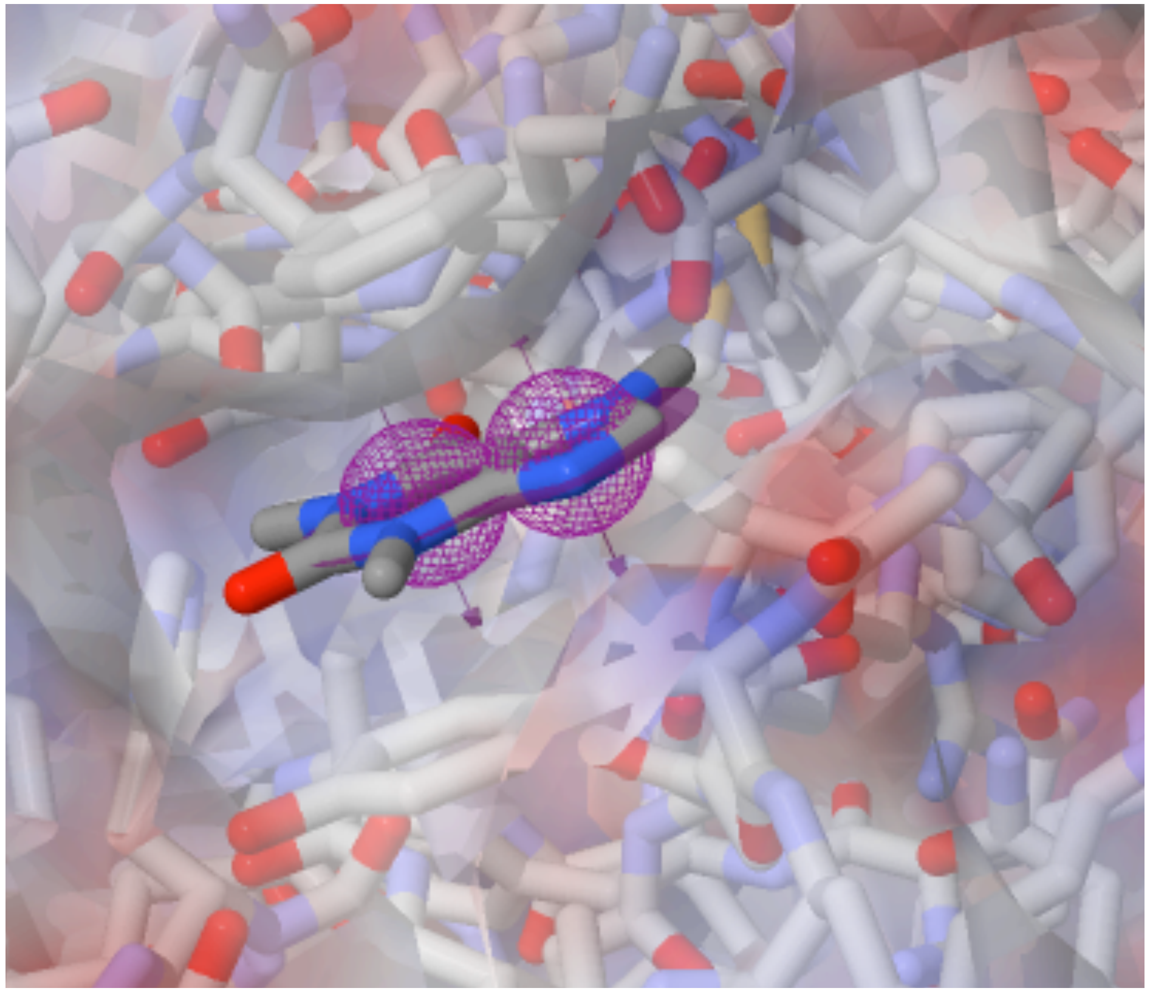


http://www.chem.ucalgary.ca/courses/351/Carey5th/Ch11/benzene-mo.jpg

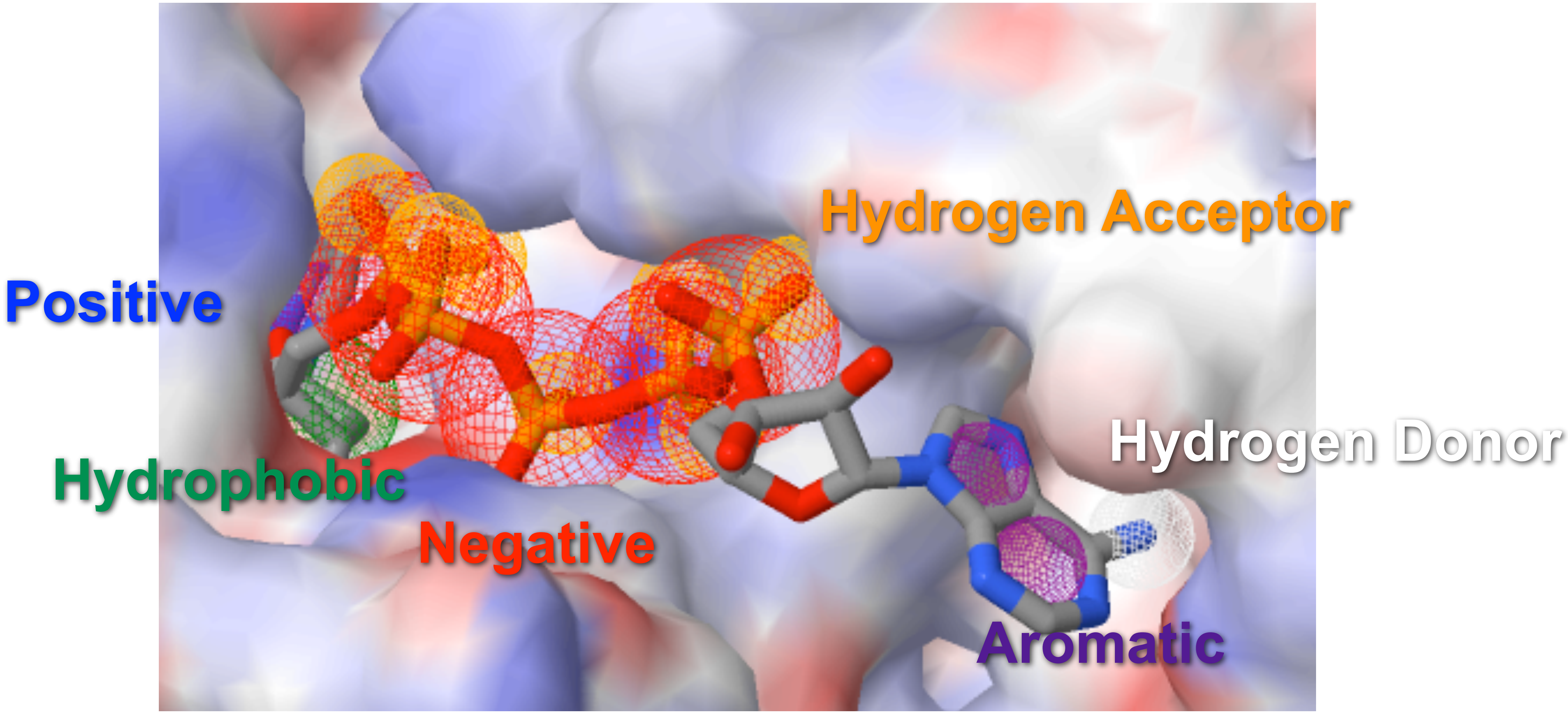
Aromatic



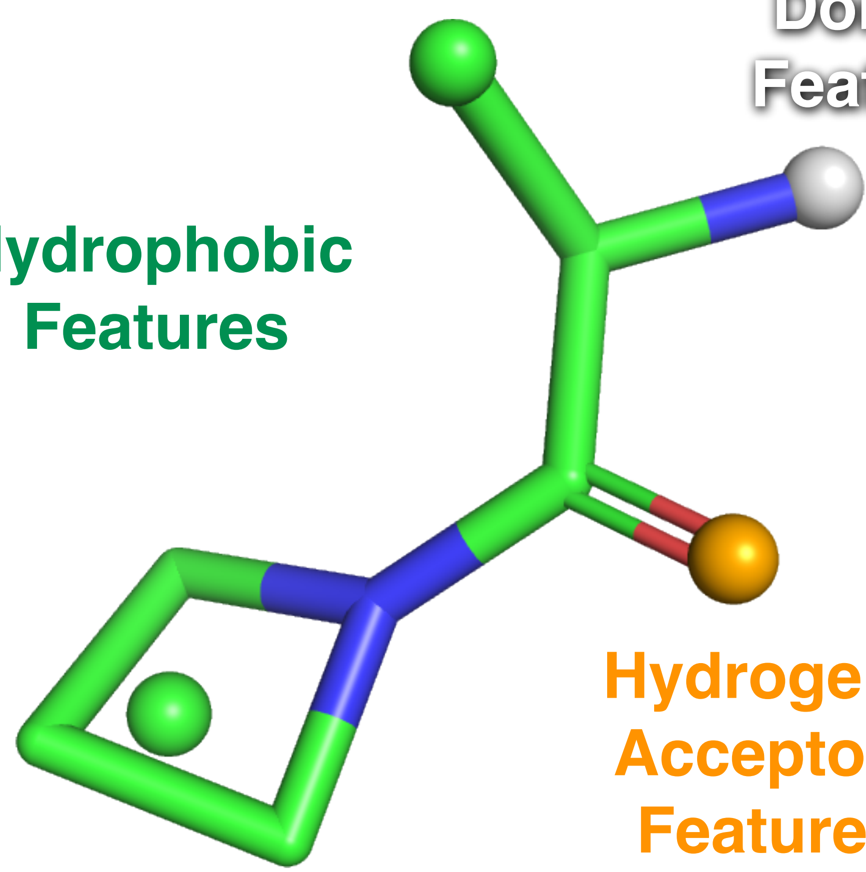
Rings offset
Interplanar distance: 3.3-3.8Å



Pharmacophore Features



Pharmer Efficient and Exact Pharmacophore Search

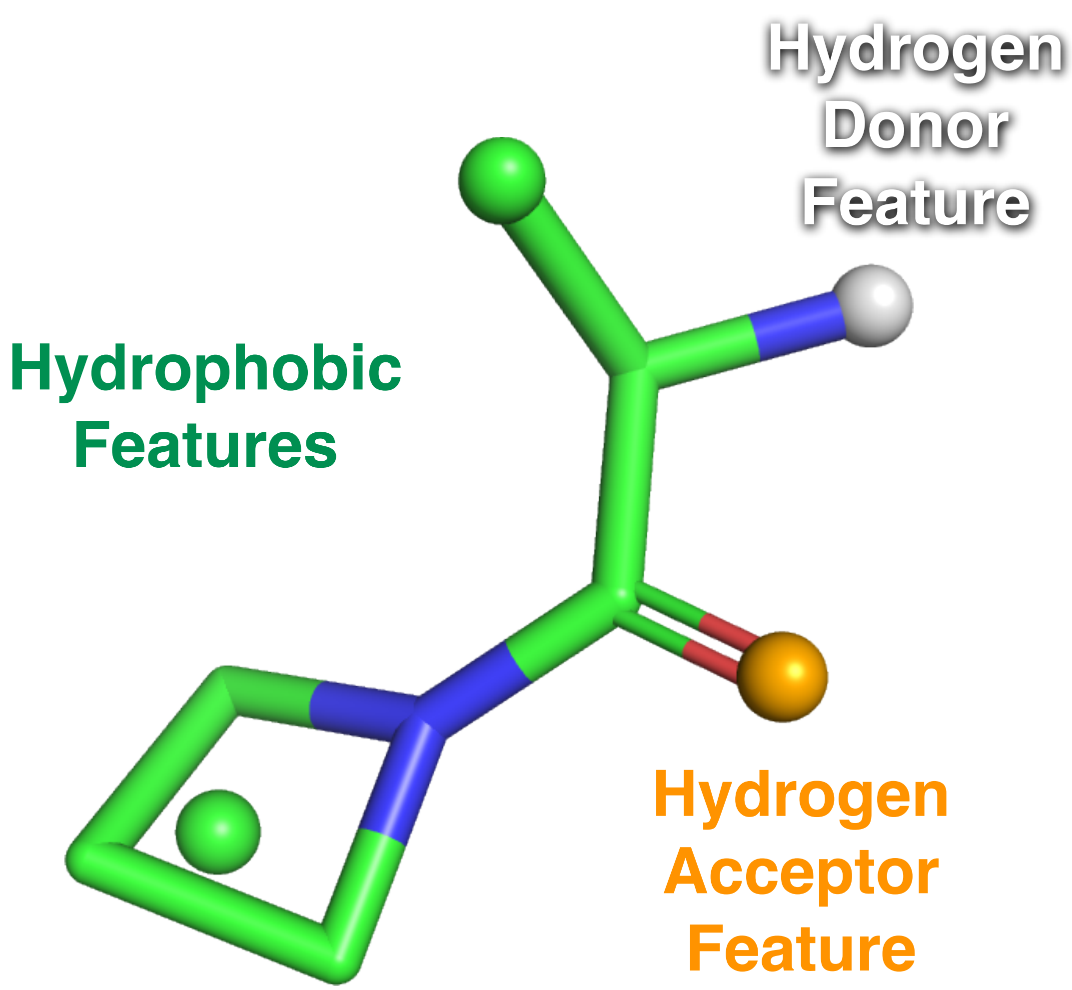


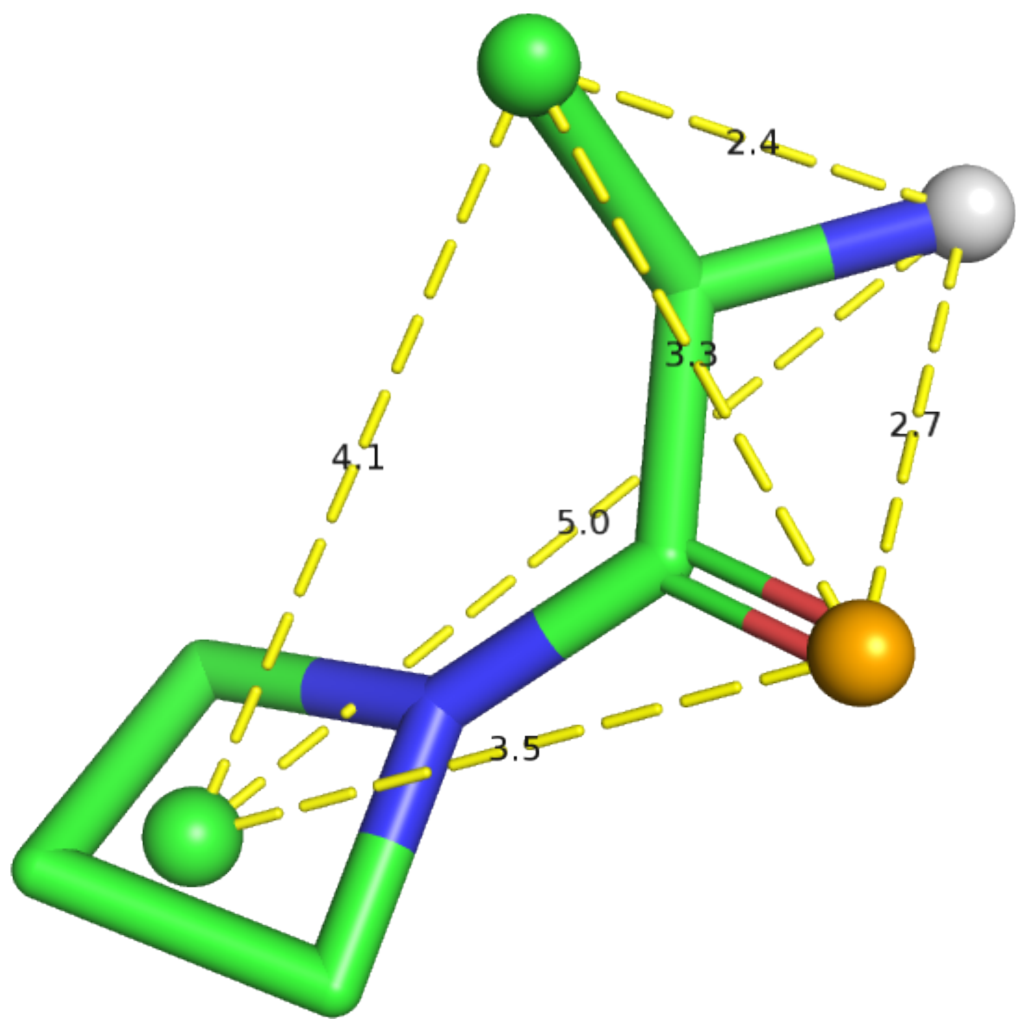
Pharmacophore

A spatial arrangement of molecular features essential for biological activity

Koes, D. R., & Camacho, C. J. (2011). Pharmer: efficient and exact pharmacophore search. *Journal of Chemical Information and Modeling, 51*(6), 1307-1314. doi:10.1021/ci200097m
Koes, D. R., & Camacho, C. J. (2012). ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic acids research, 40*(Web Server issue), W409-414. doi:10.1093/nar/gks378

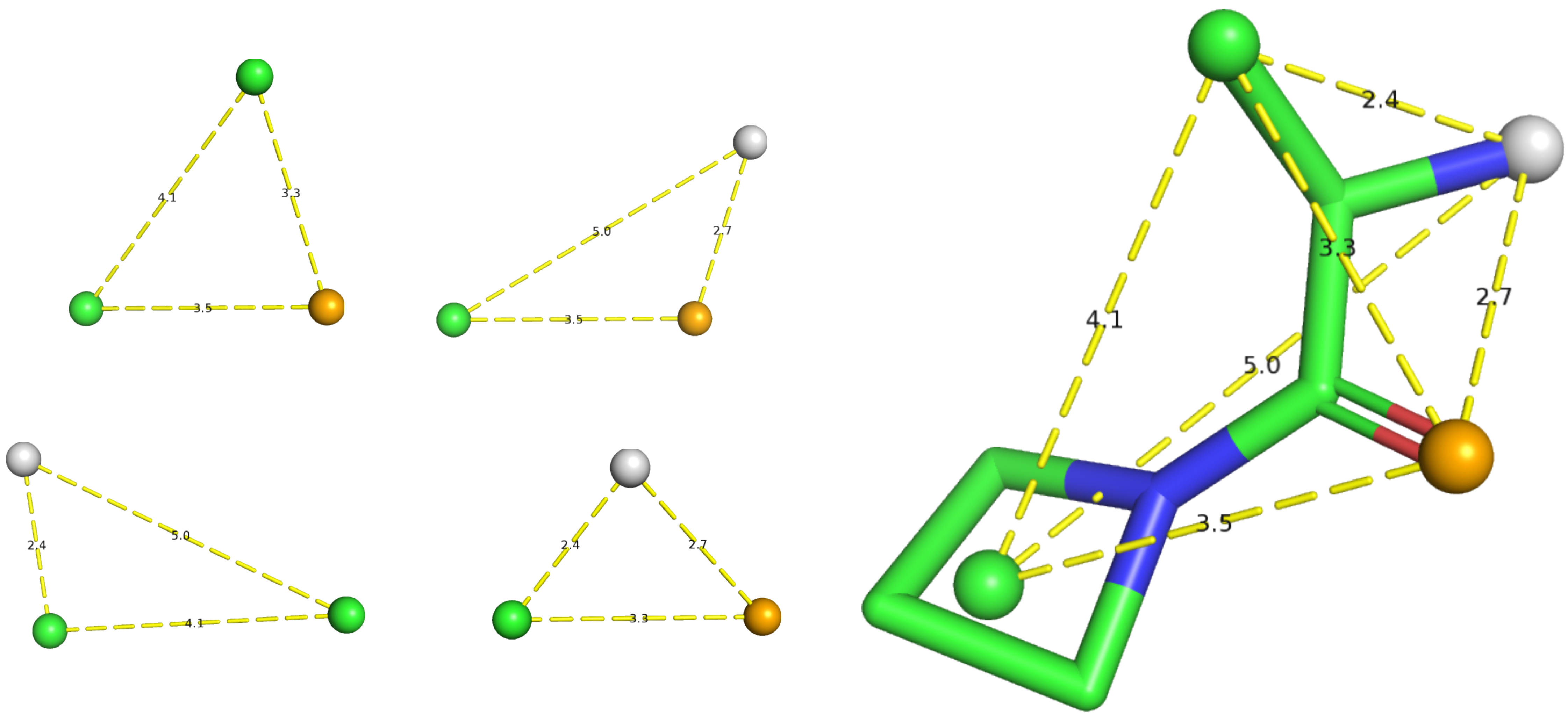
Pharmer Efficient and Exact Pharmacophore Search





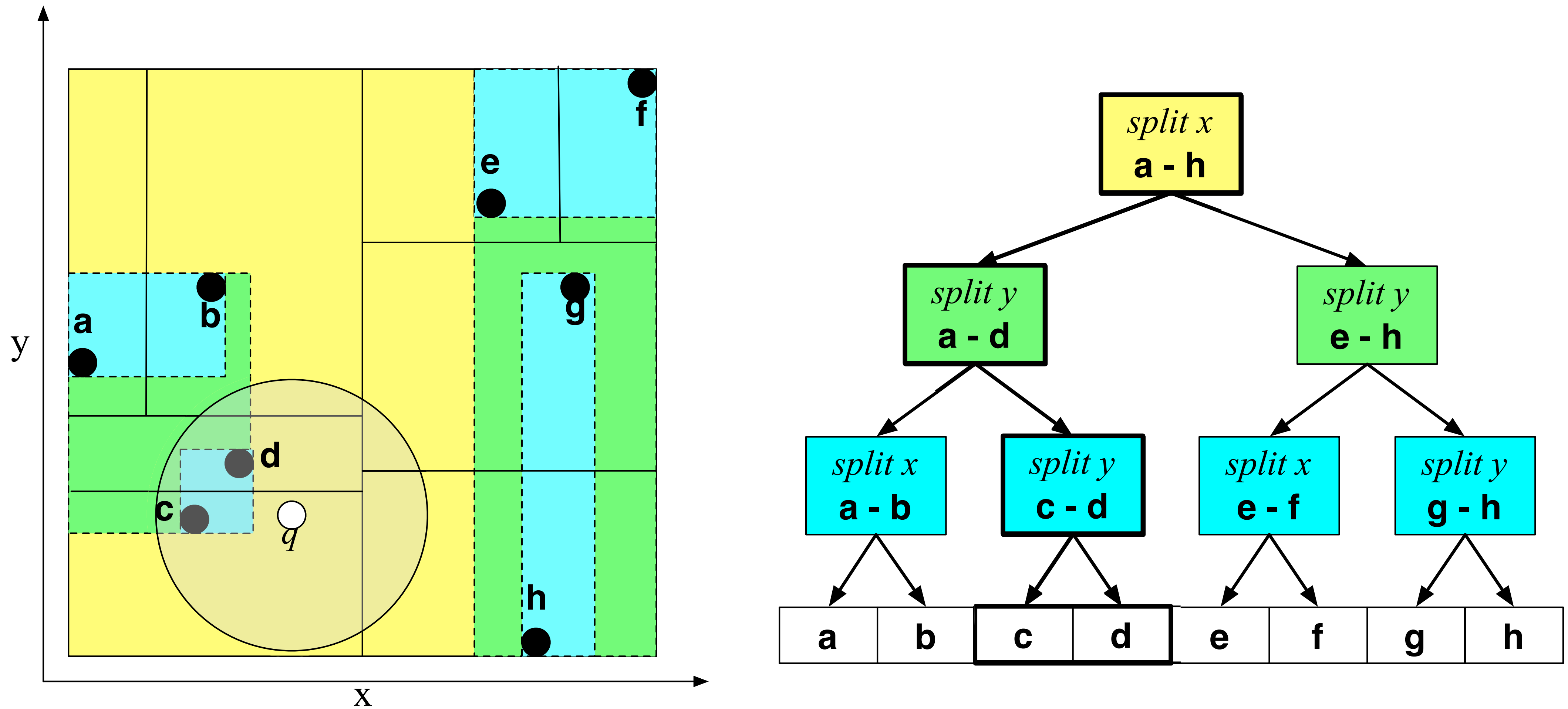
Pharmer

Efficient and Exact Pharmacophore Search



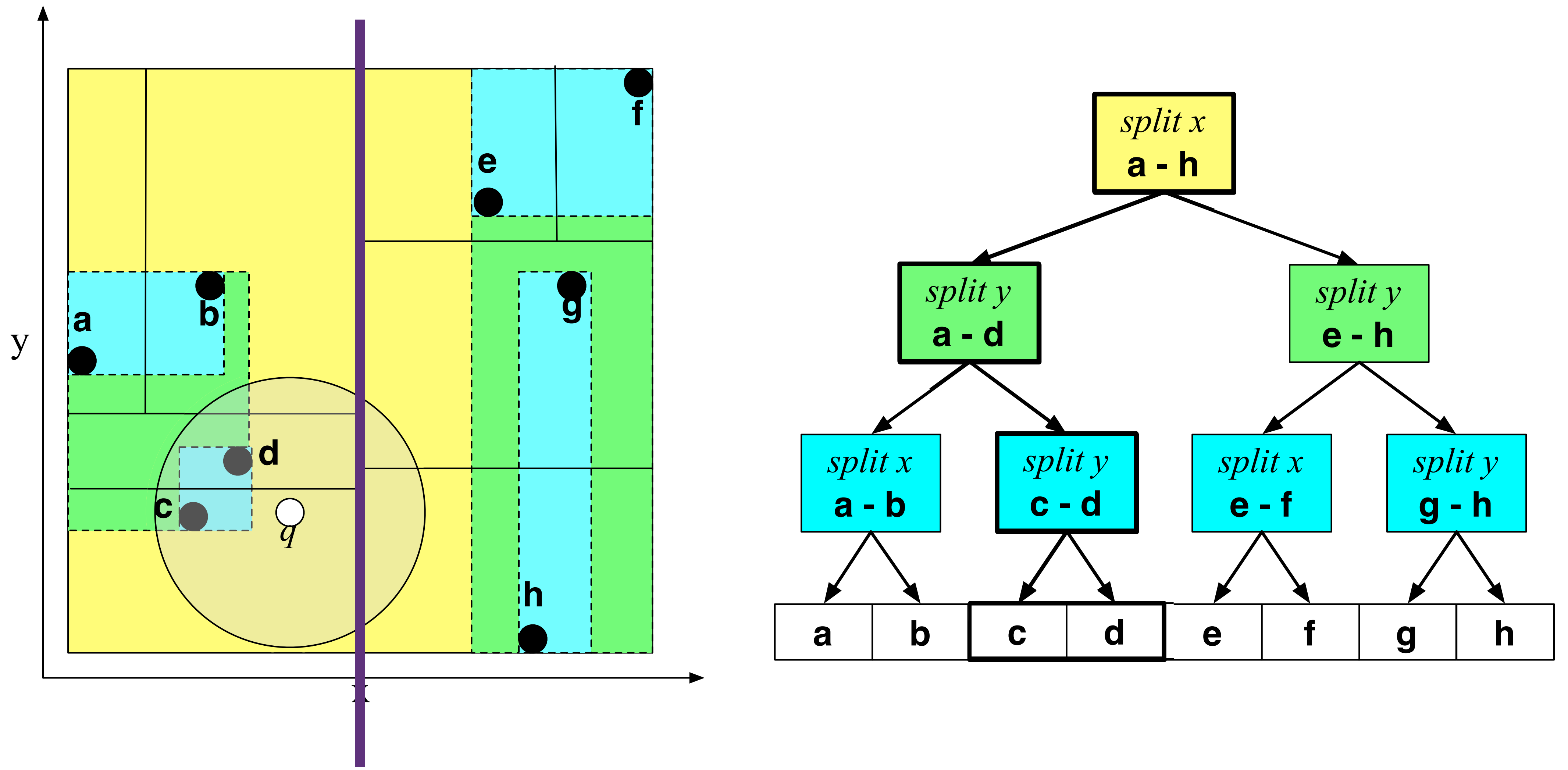
Pharmer

Efficient and Exact Pharmacophore Search



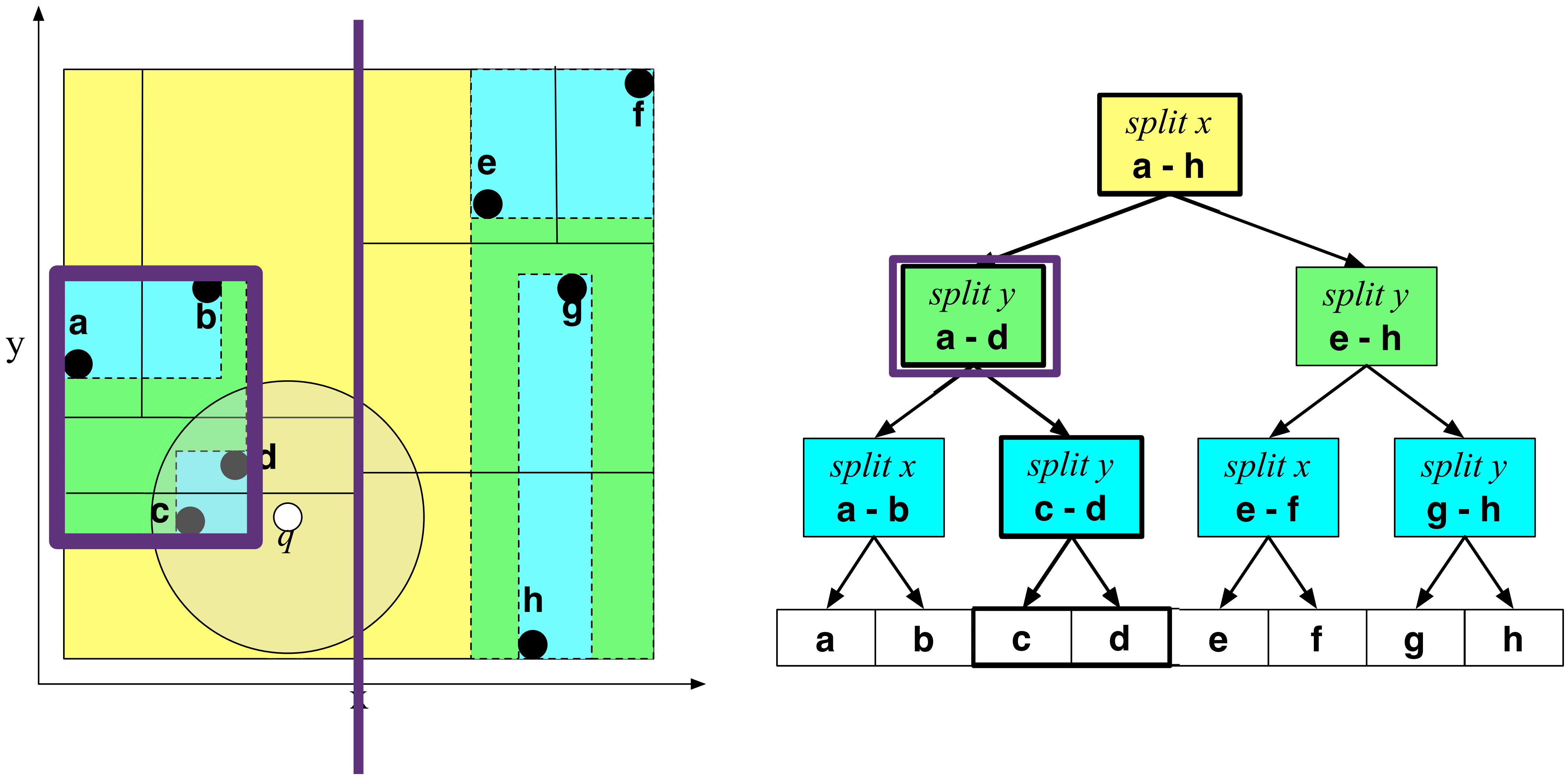
Pharmer

Efficient and Exact Pharmacophore Search

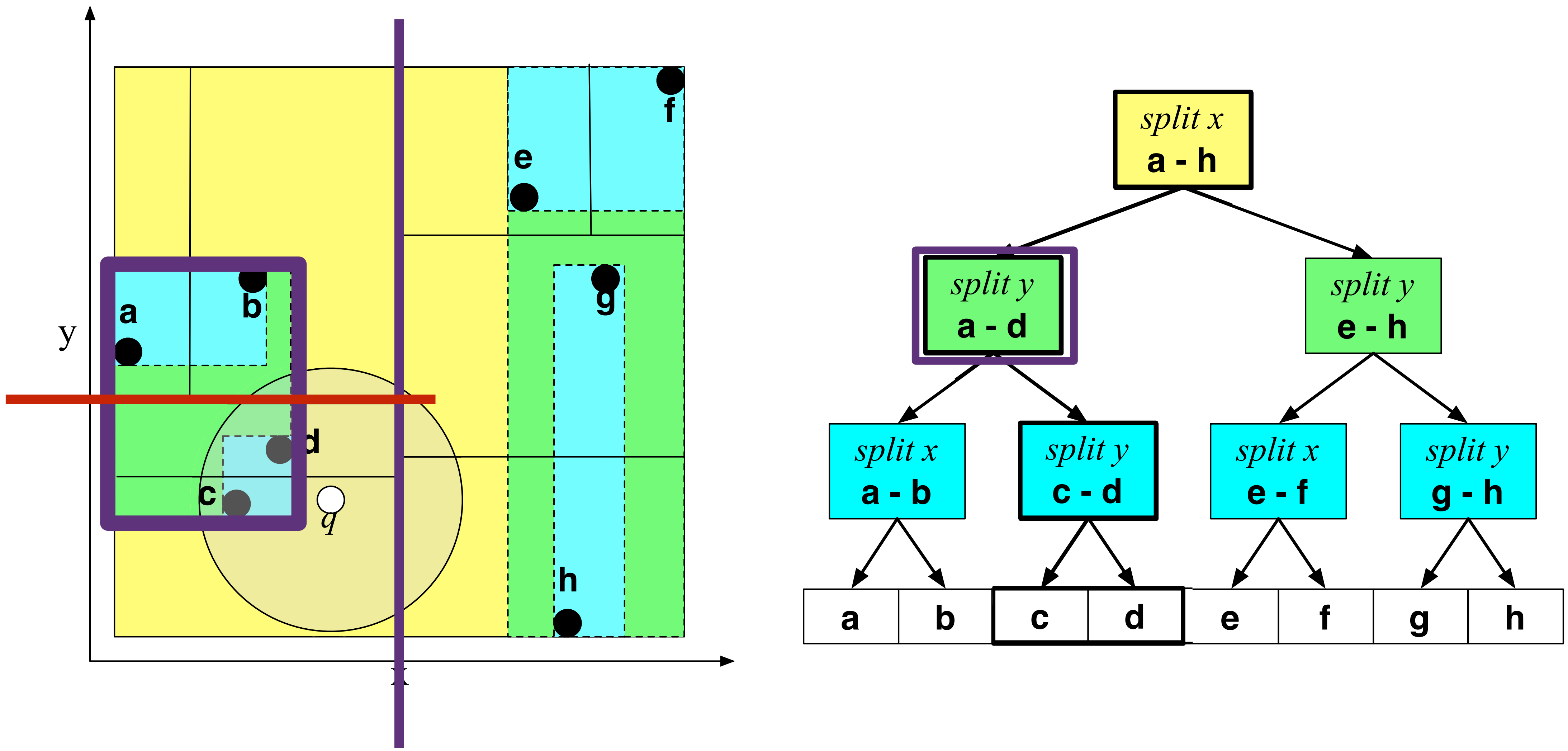


Pharmer

Efficient and Exact Pharmacophore Search

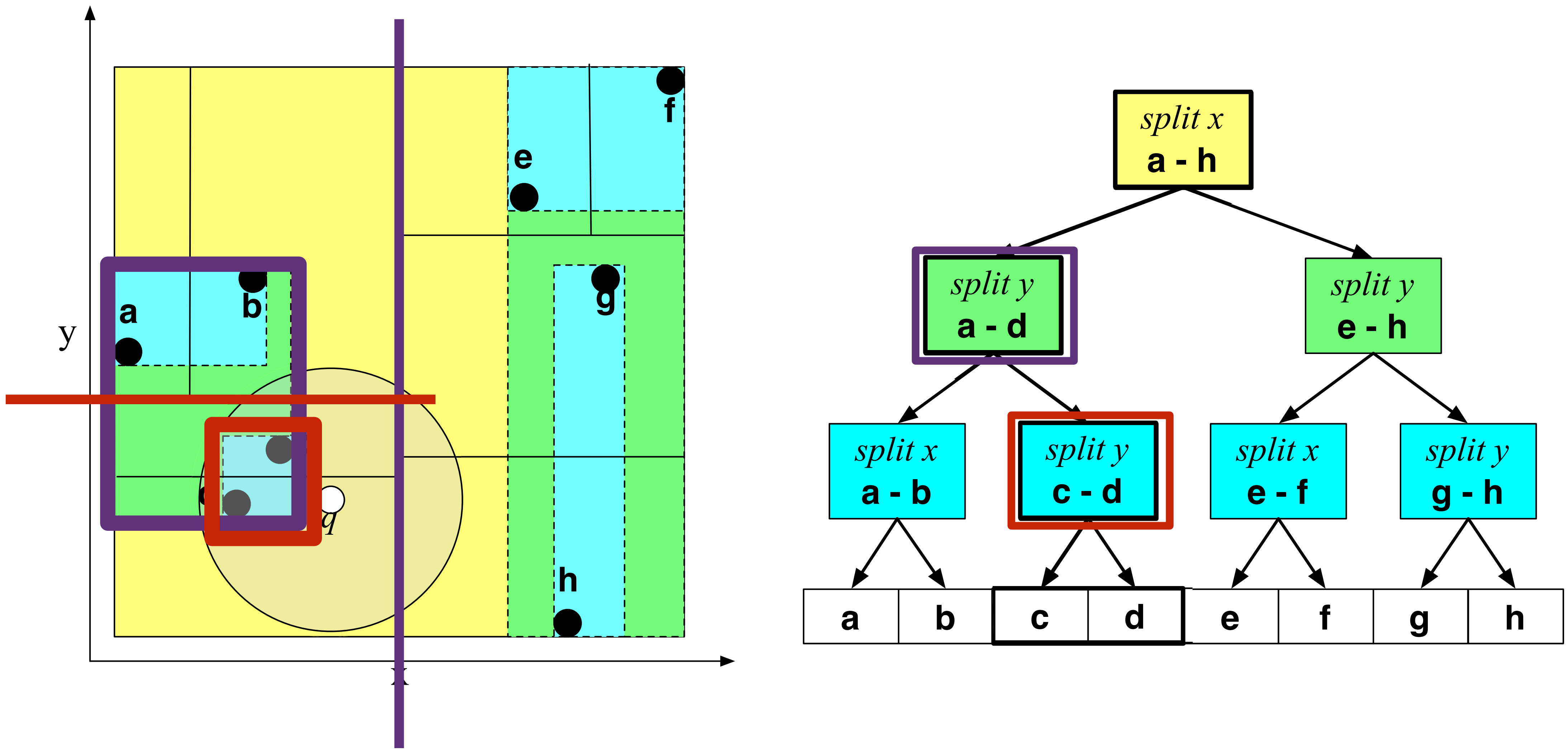


Pharmer Efficient and Exact Pharmacophore Search



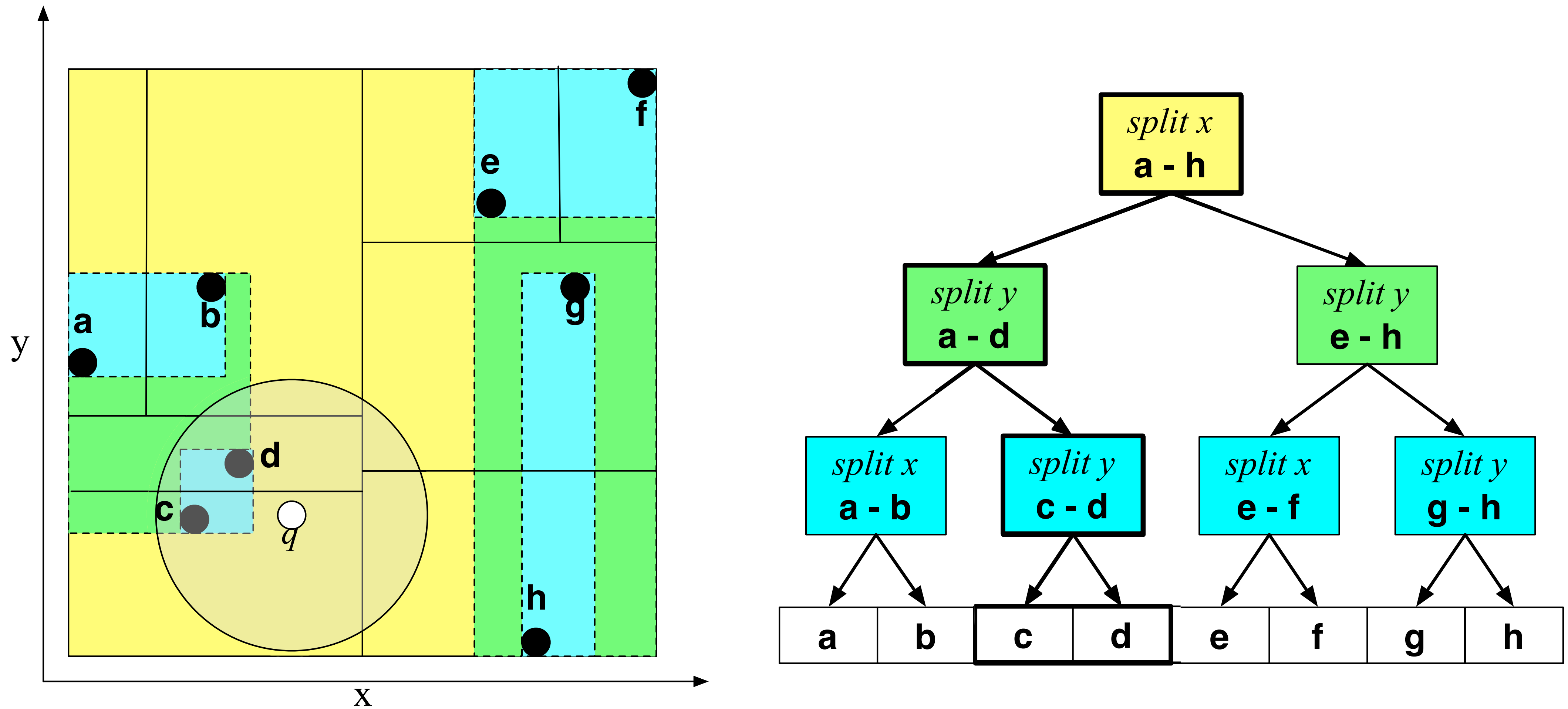
Pharmer

Efficient and Exact Pharmacophore Search



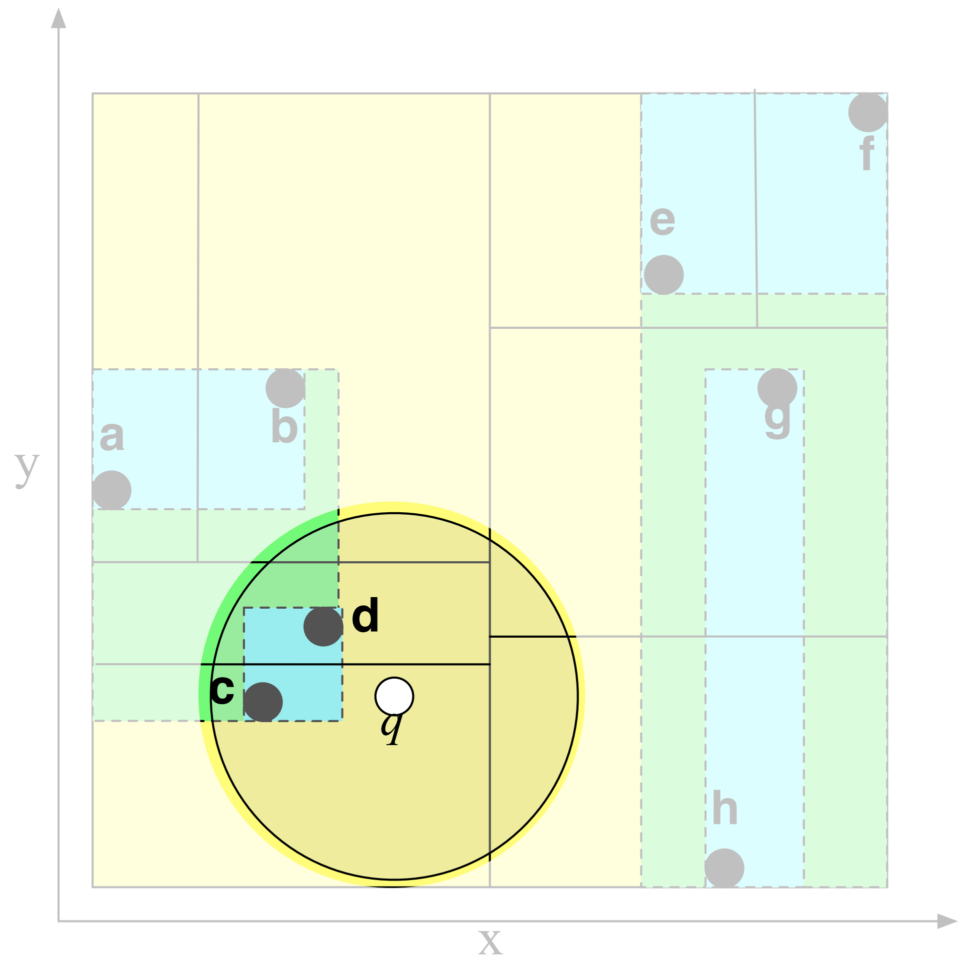
Pharmer

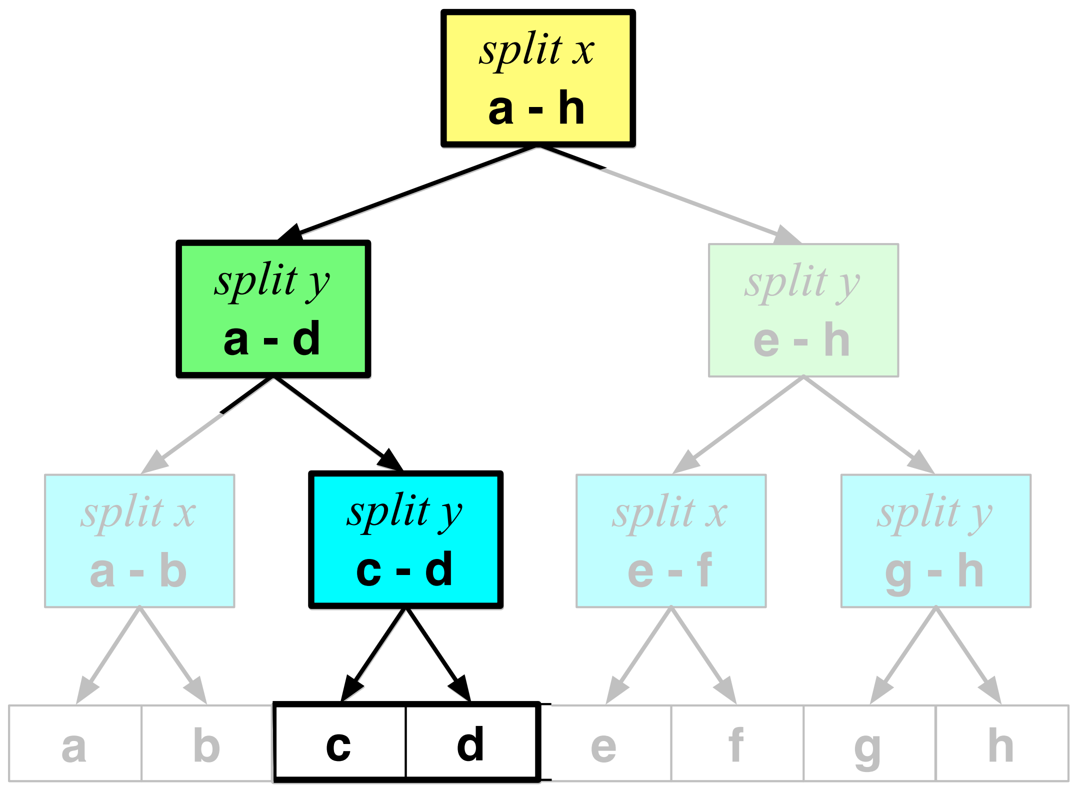
Efficient and Exact Pharmacophore Search



Pharmer

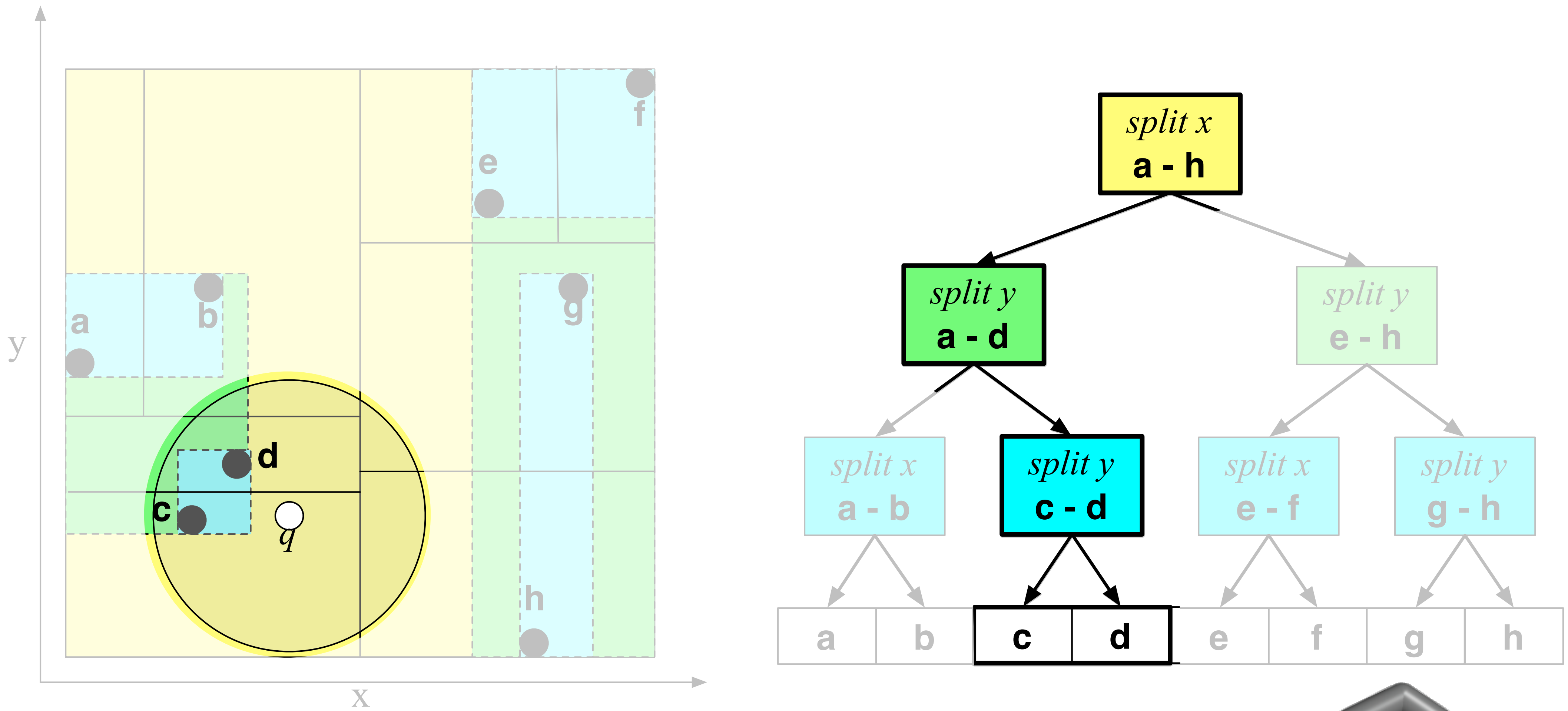
Efficient and Exact Pharmacophore Search

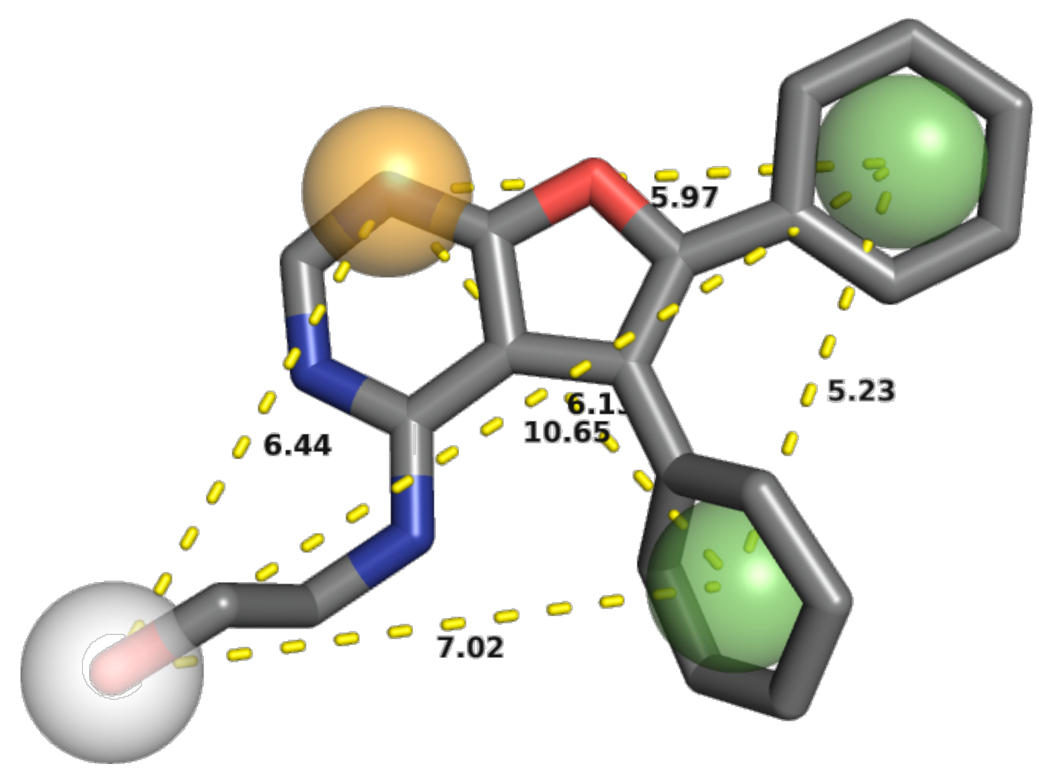




Pharmer

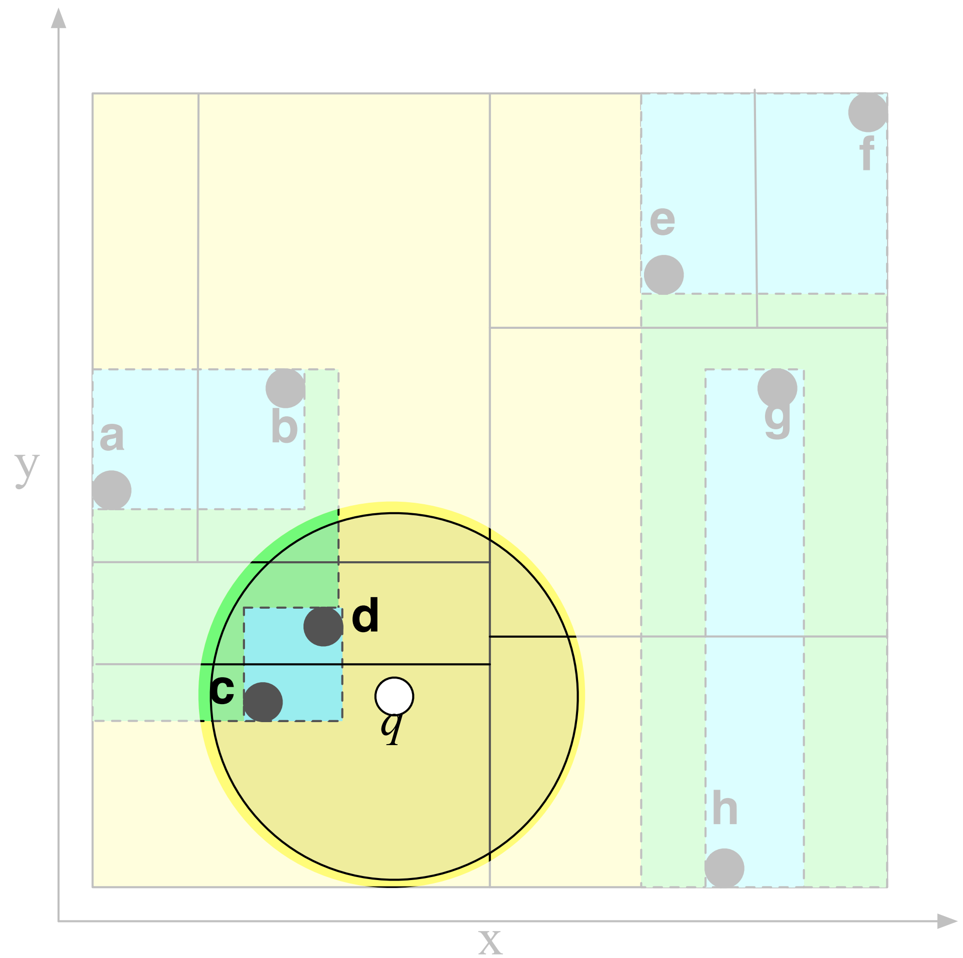
Efficient and Exact Pharmacophore Search

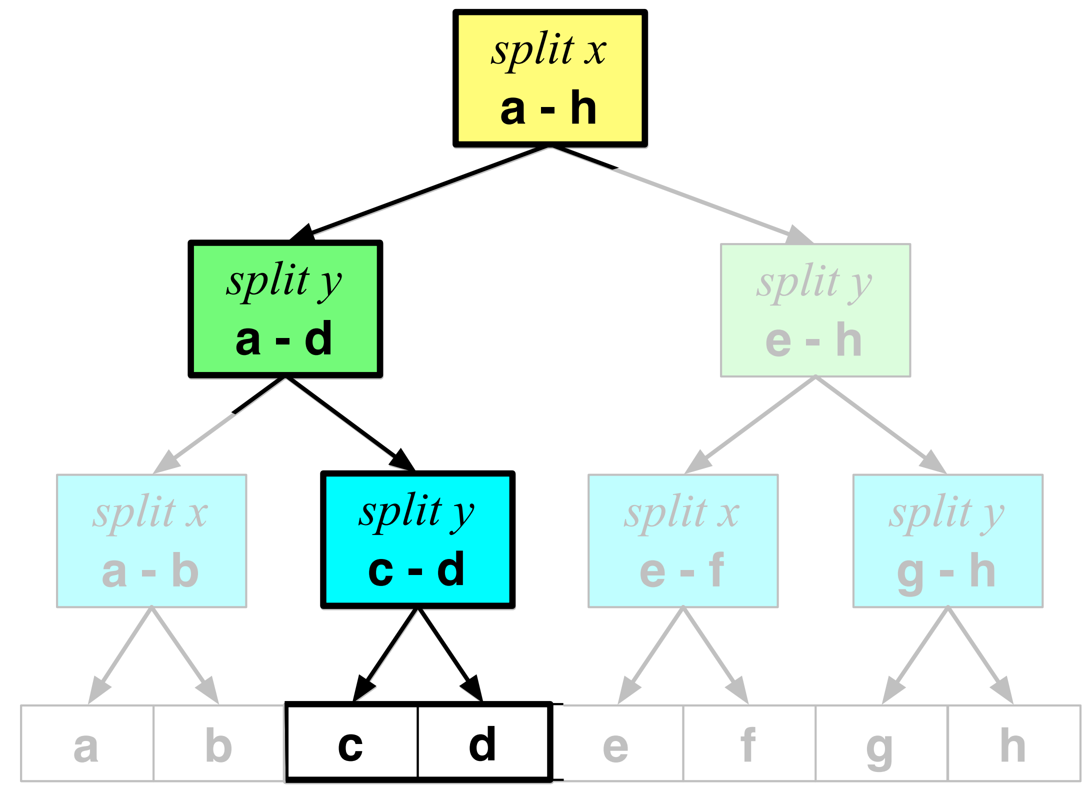


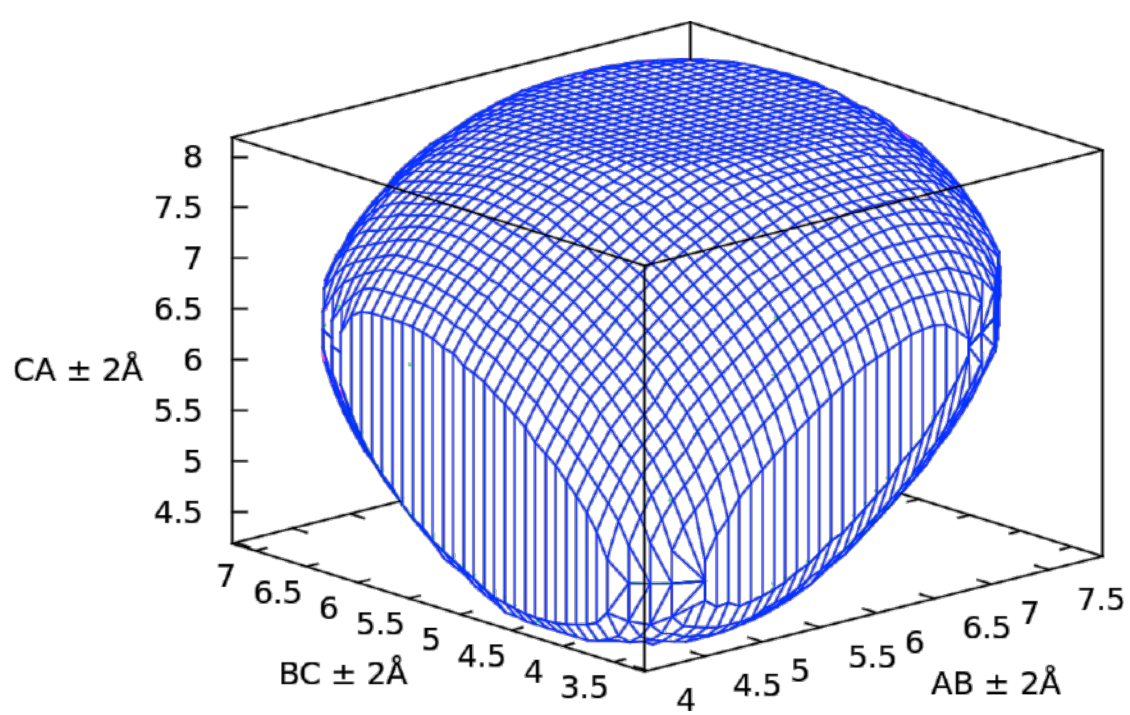


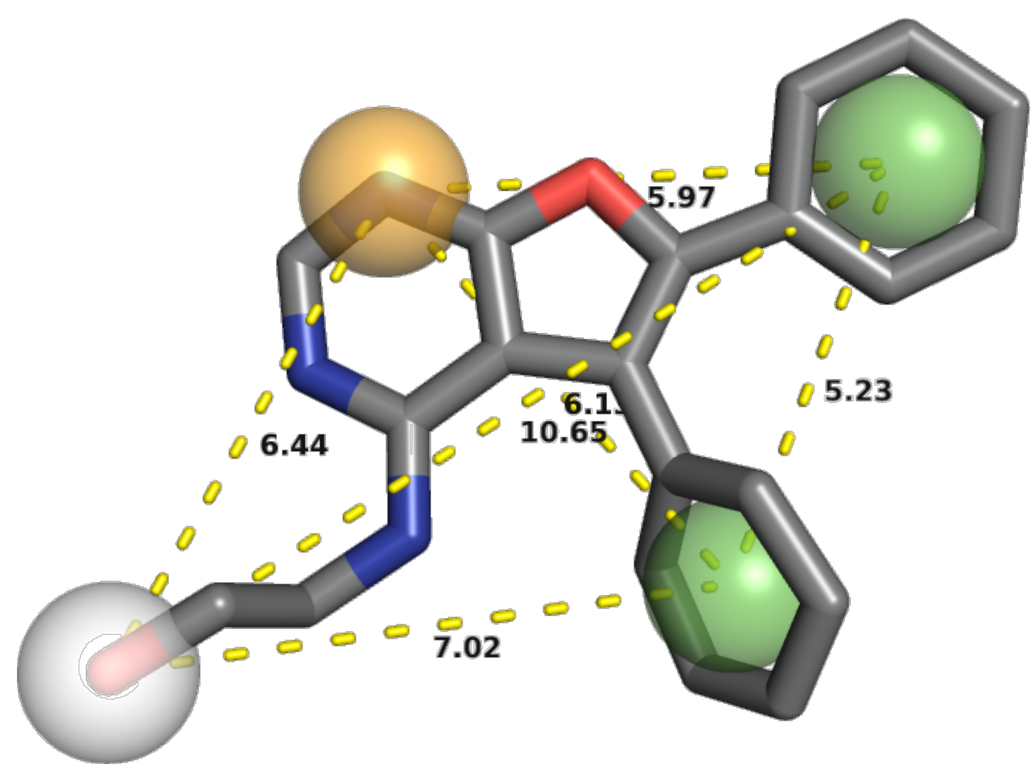
Pharmer

Efficient and Exact Pharmacophore Search

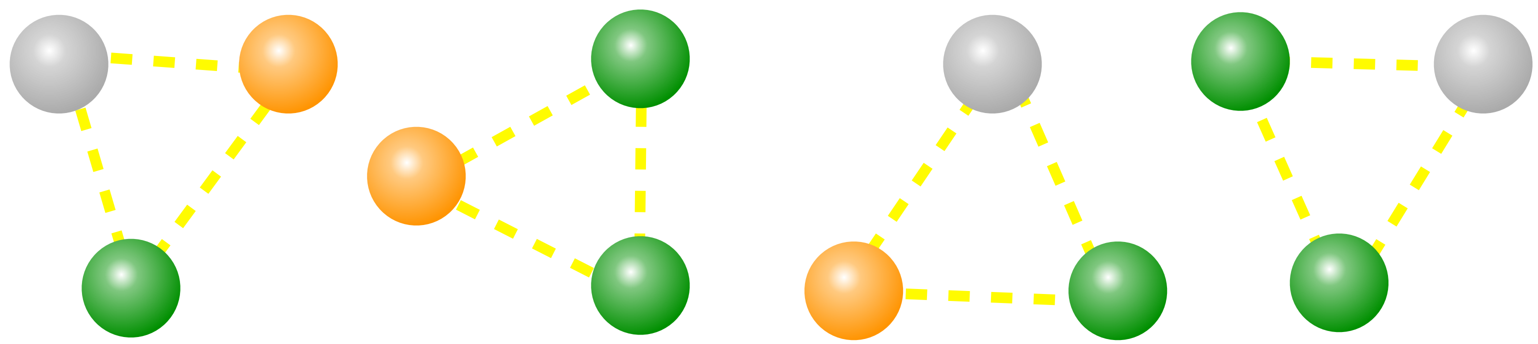


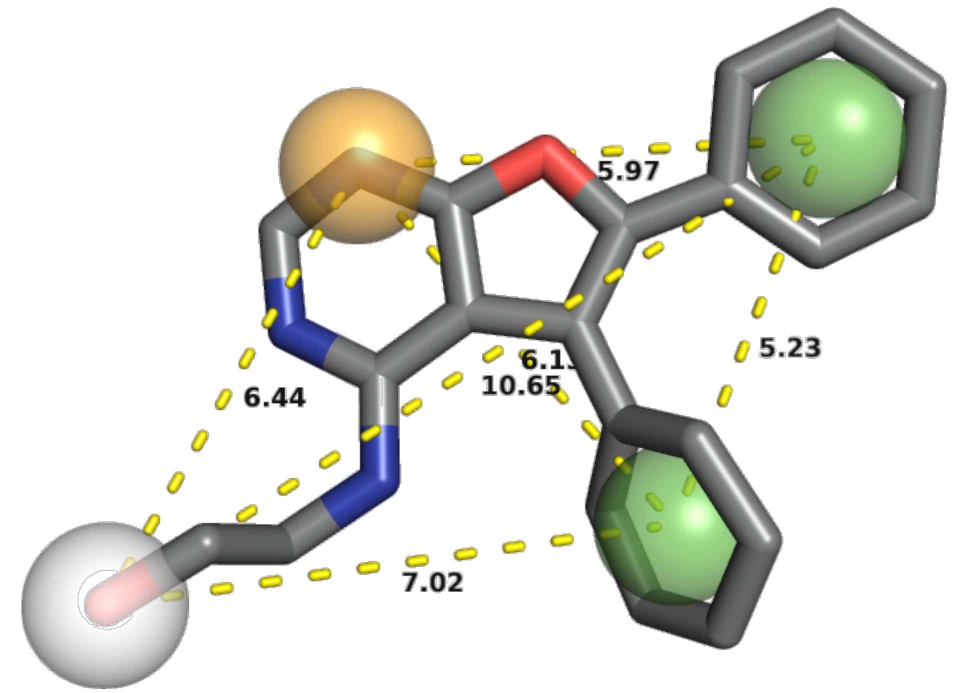






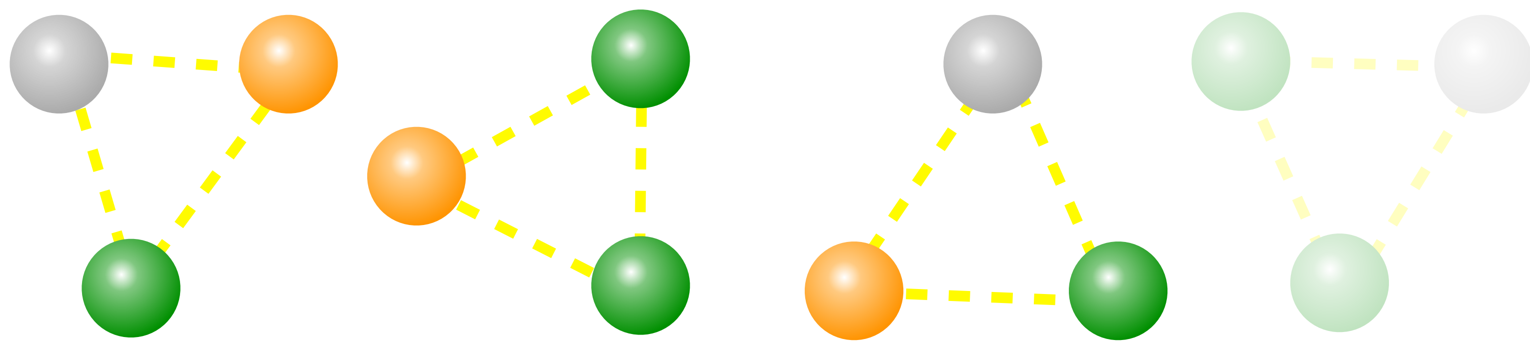
Pharmer Efficient and Exact Pharmacophore Search

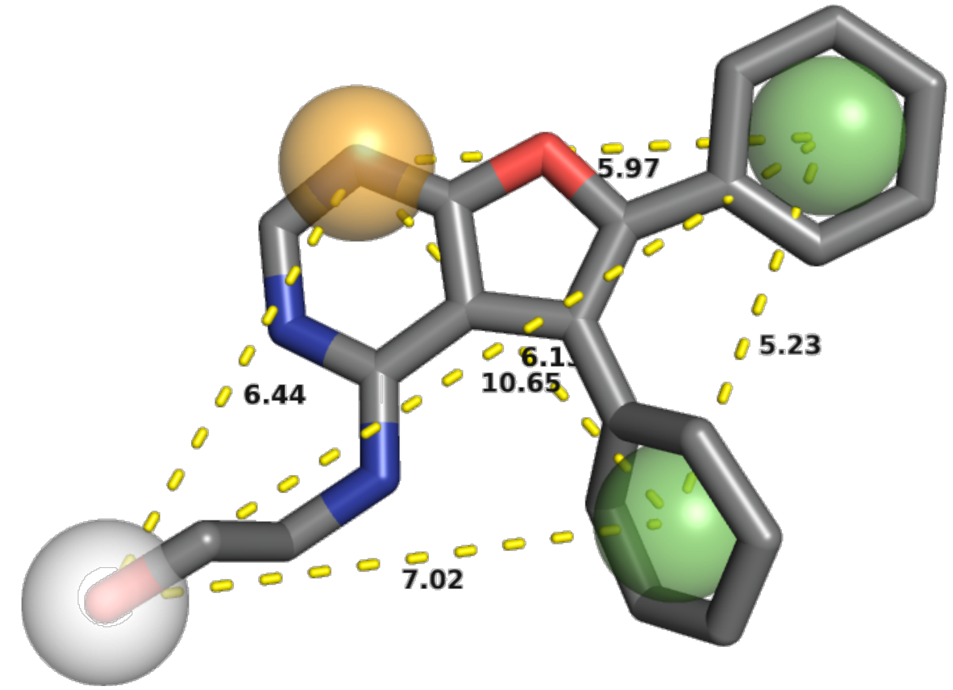




Pharmer

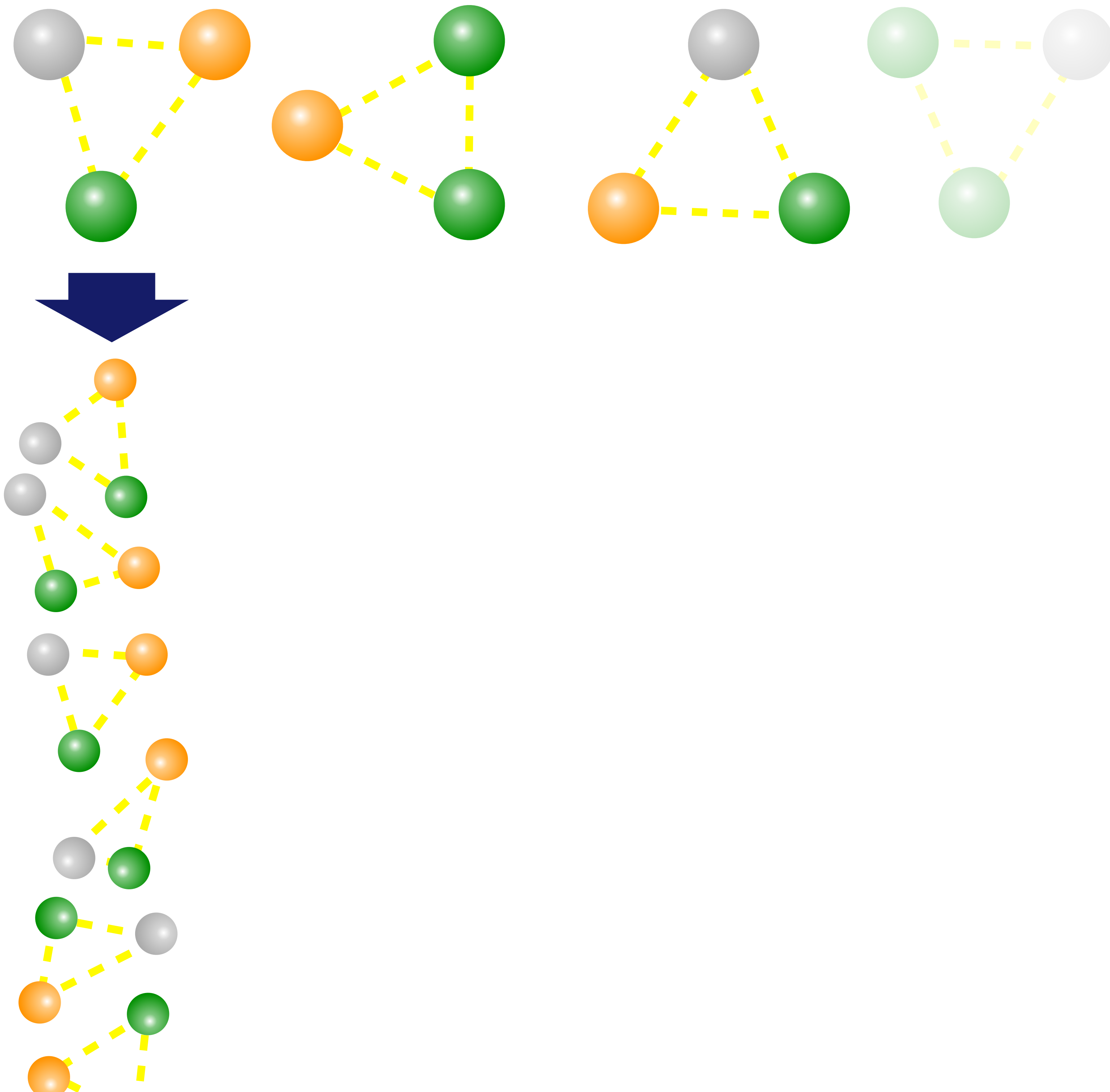
Efficient and Exact Pharmacophore Search

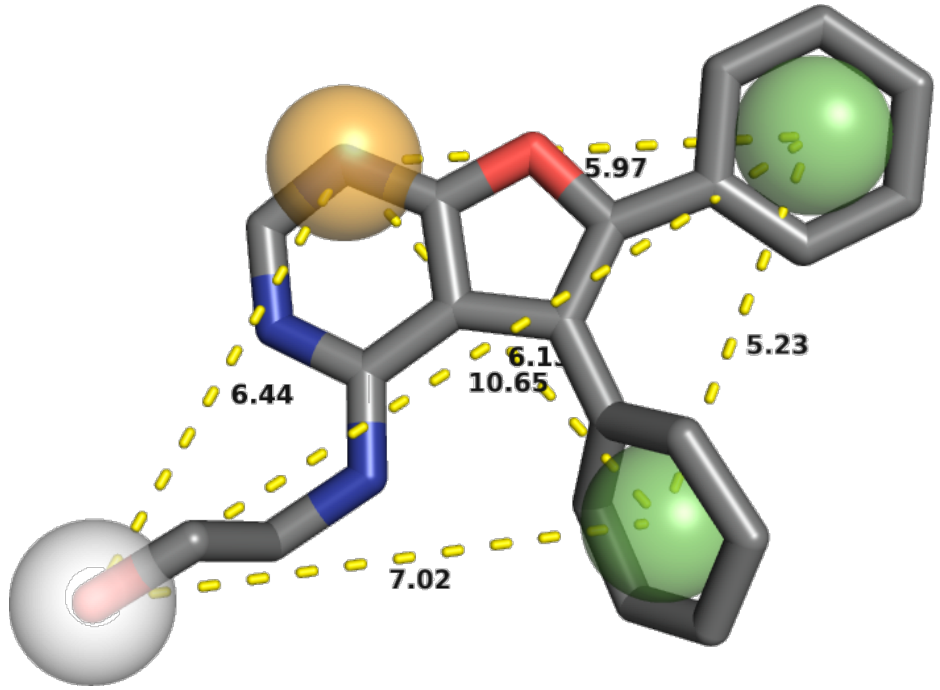




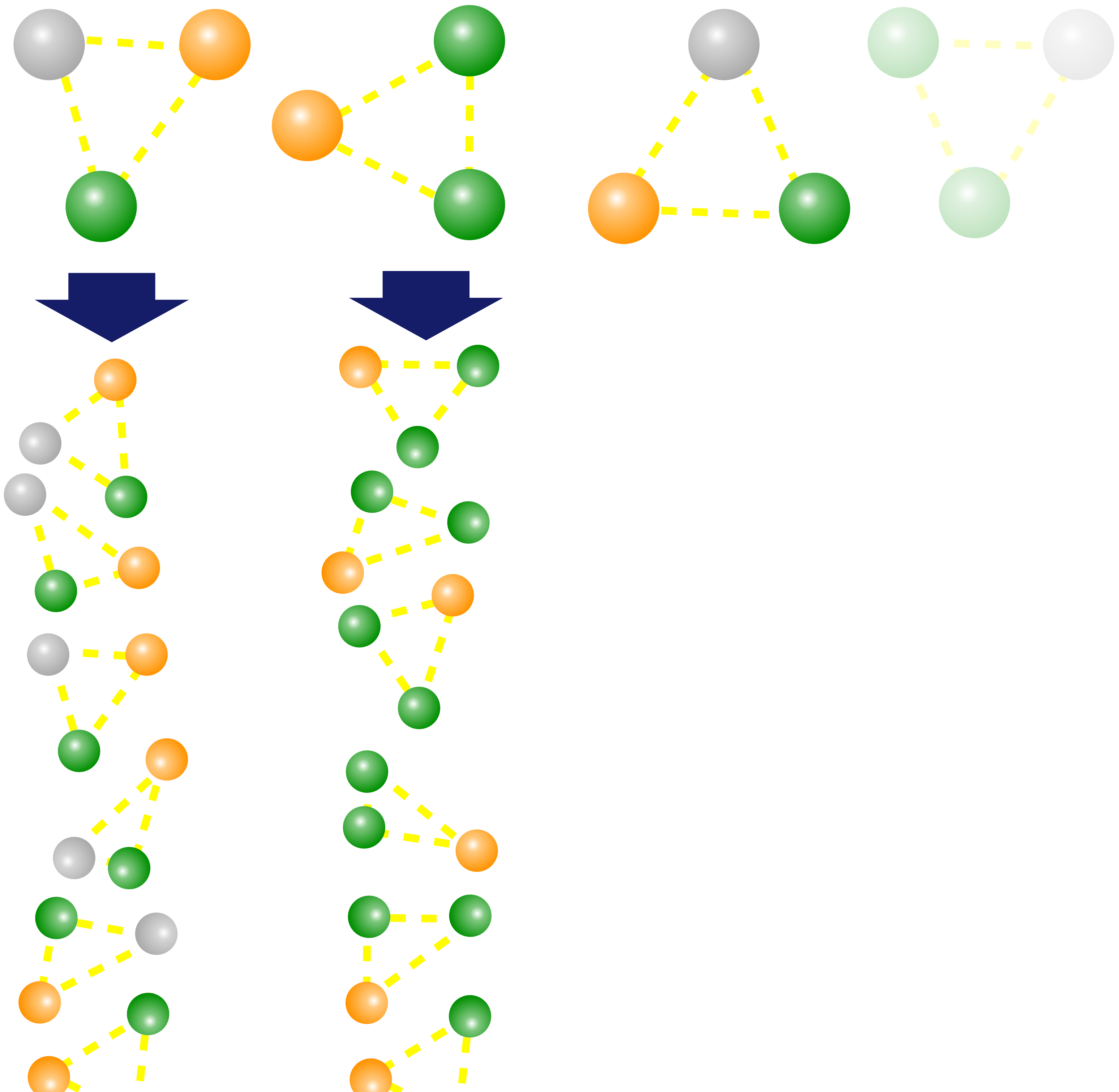
Pharmer

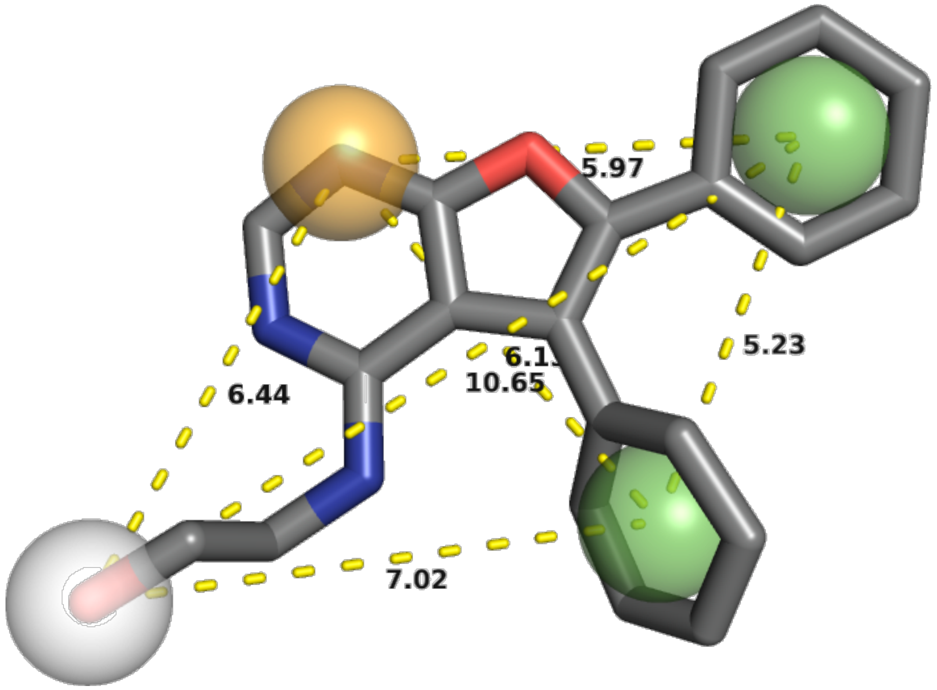
Efficient and Exact Pharmacophore Search





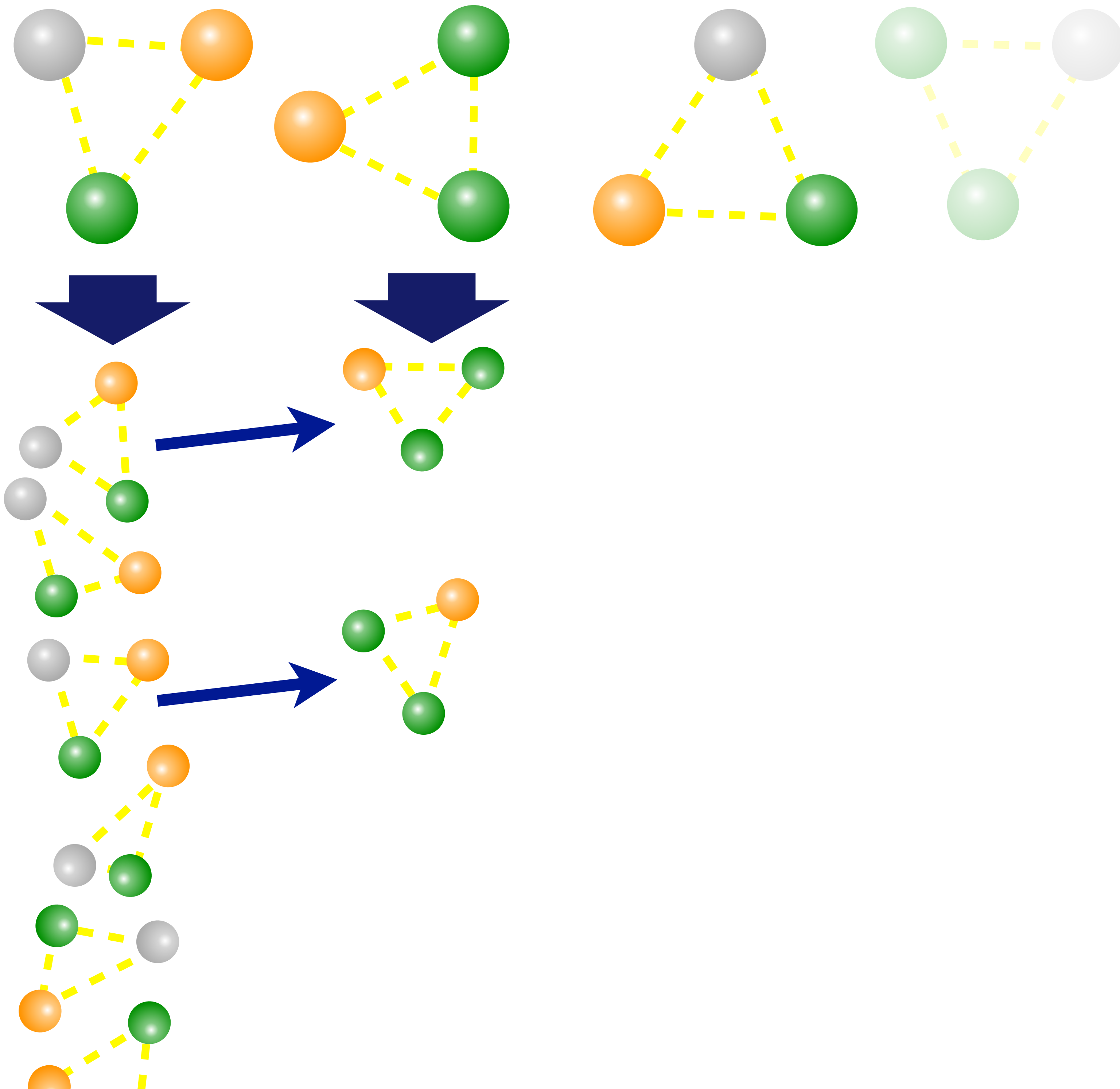
Pharmer Efficient and Exact Pharmacophore Search

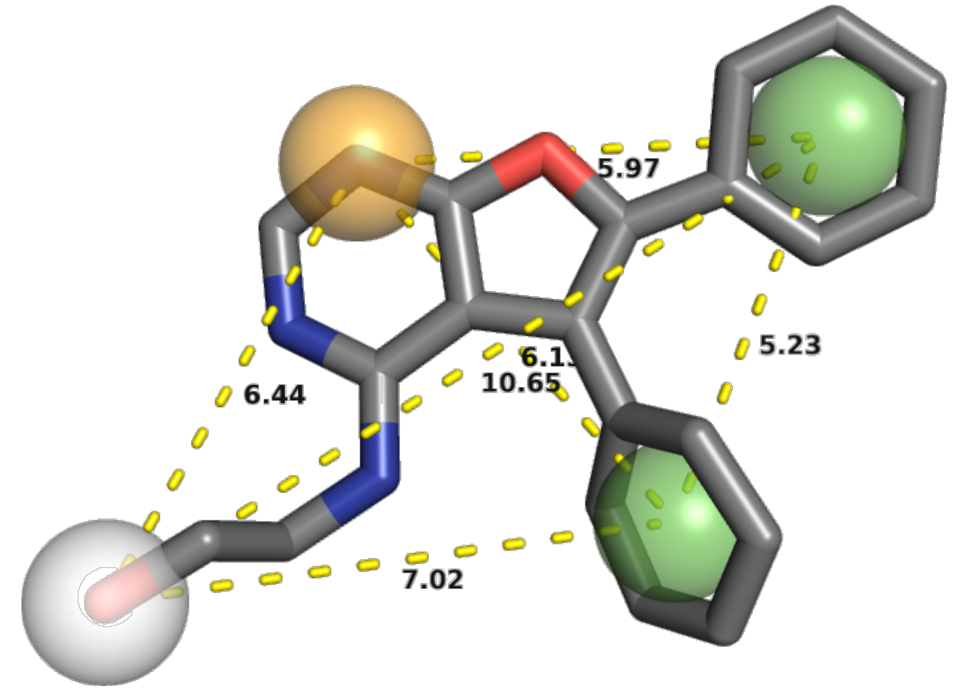




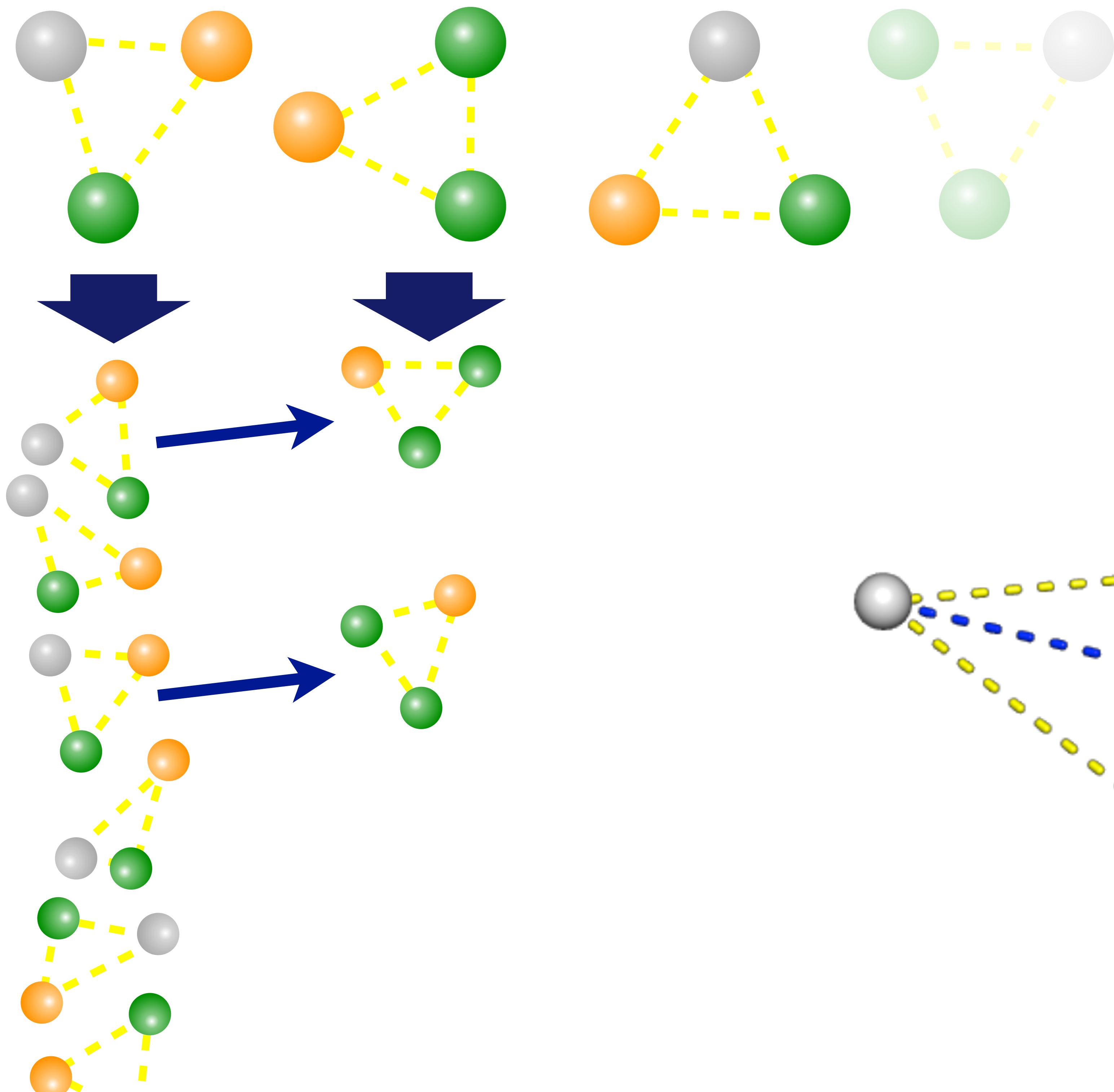
Pharmer

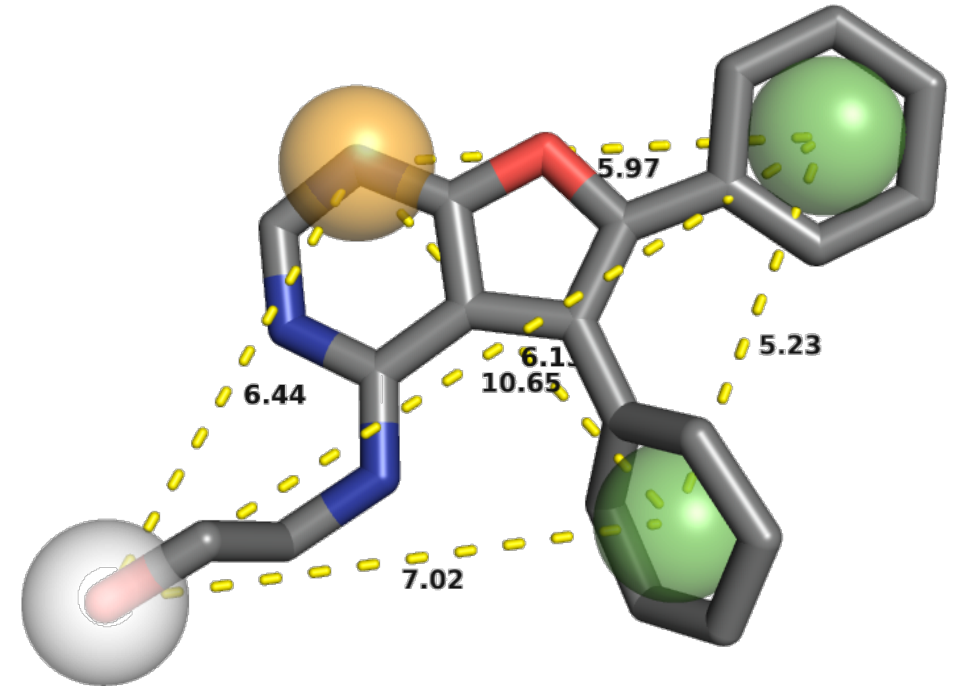
Efficient and Exact Pharmacophore Search

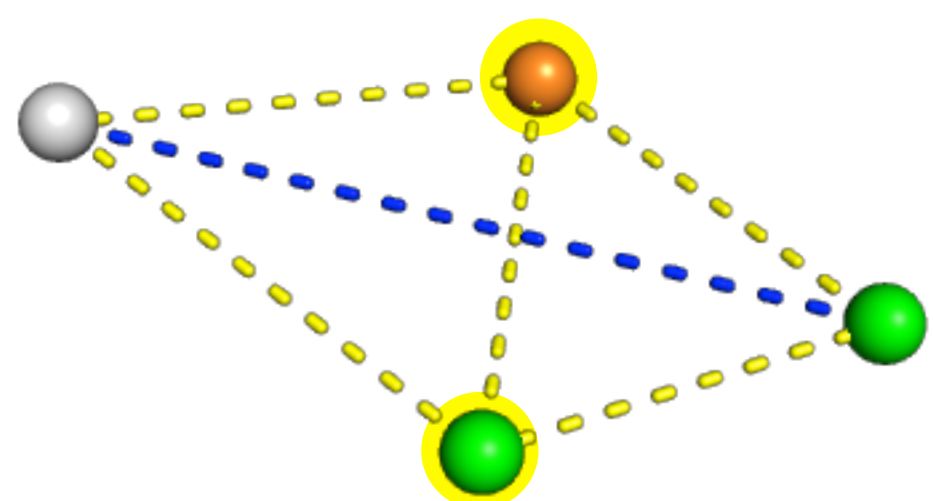




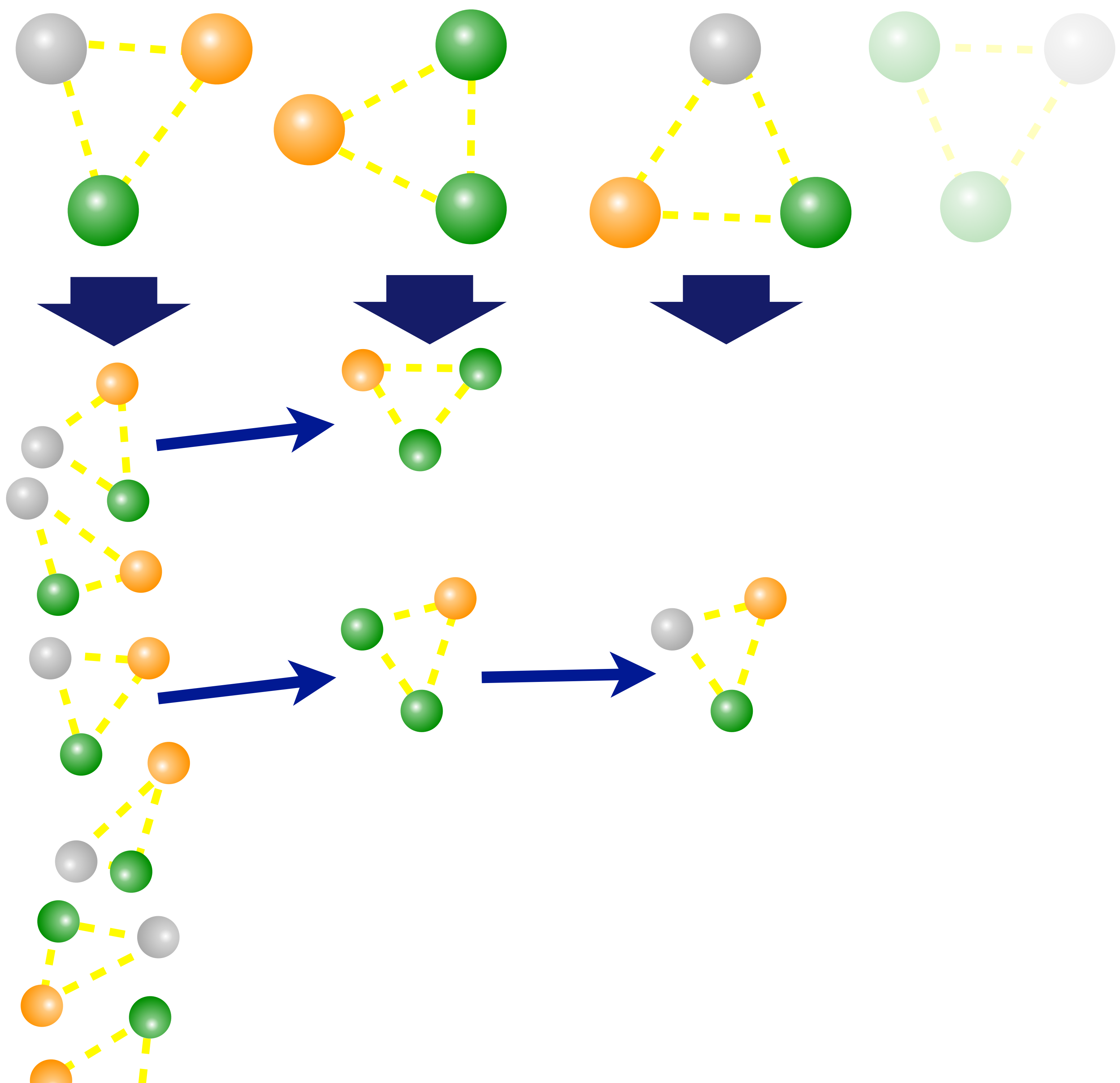
Pharmer Efficient and Exact Pharmacophore Search

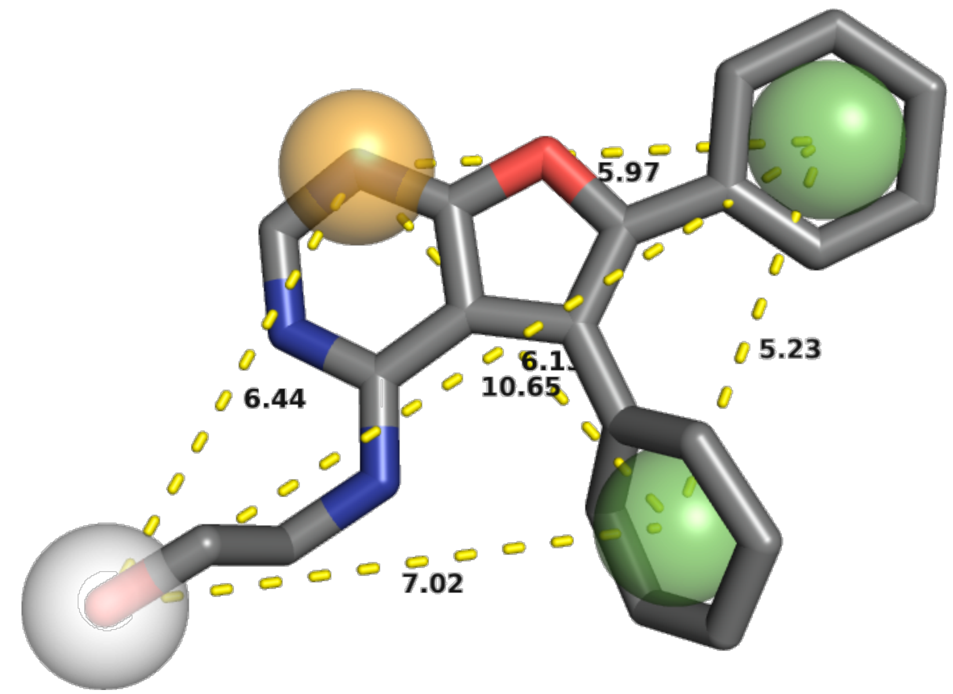




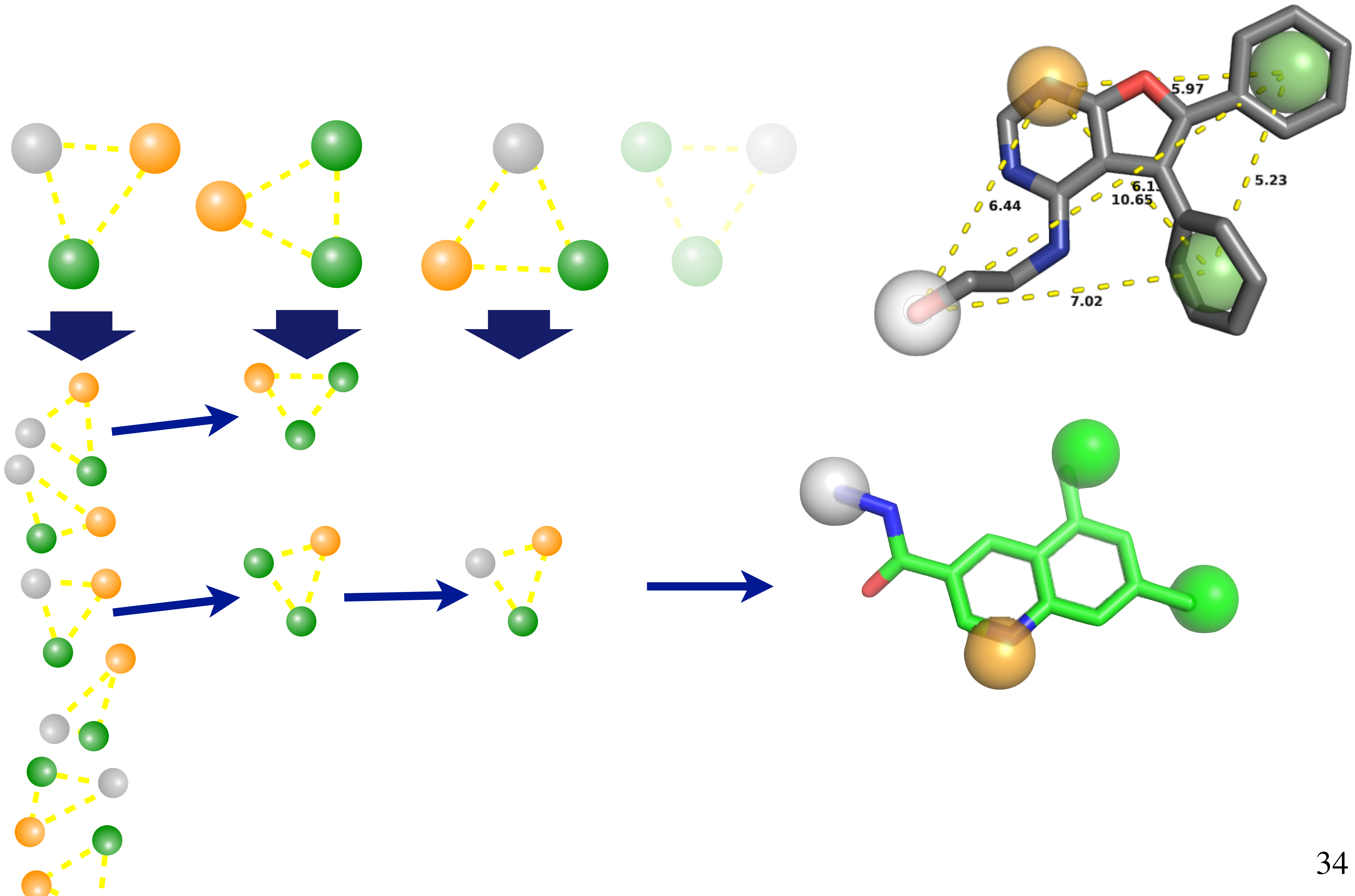


Pharmer Efficient and Exact Pharmacophore Search



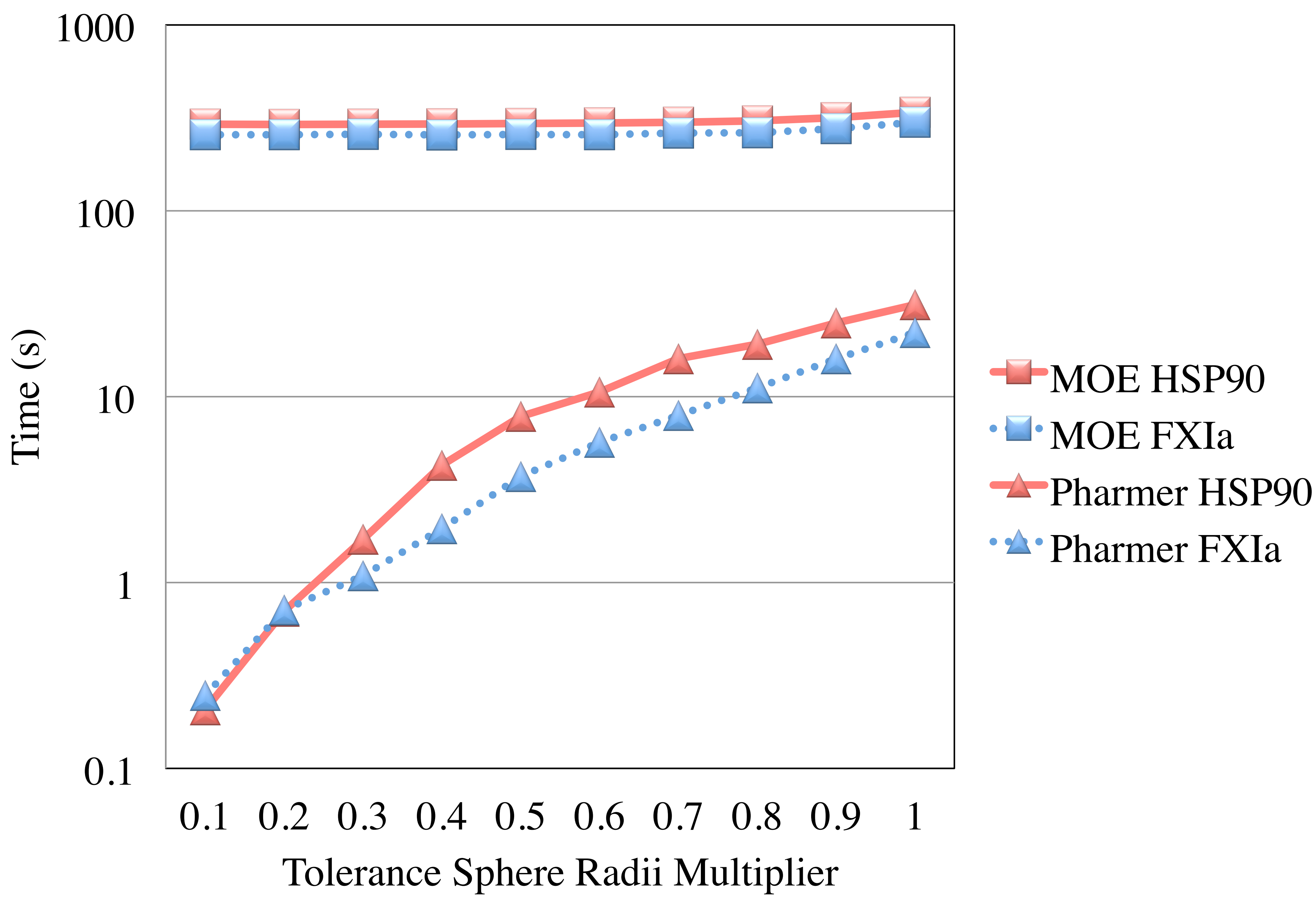


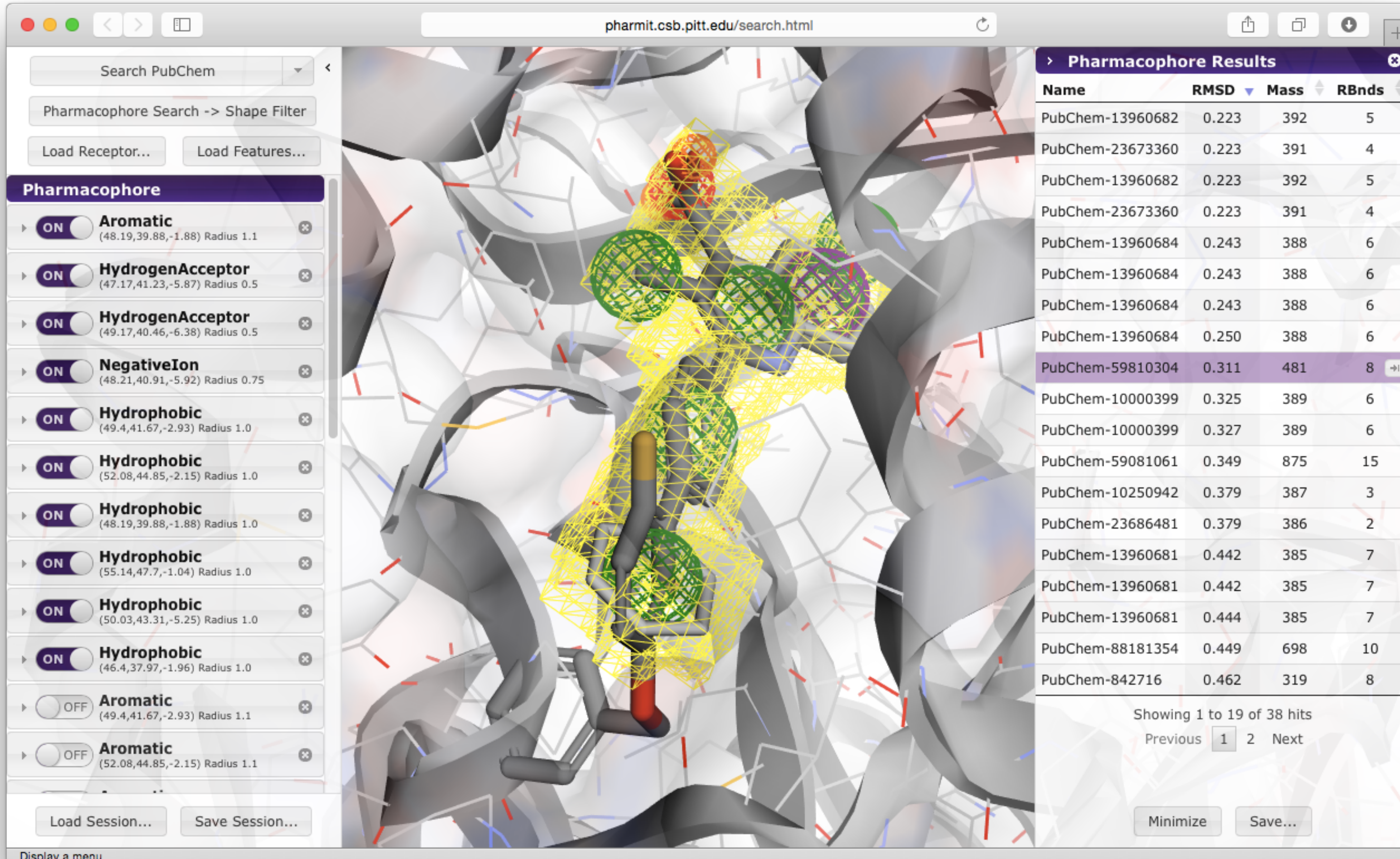
Pharmer Efficient and Exact Pharmacophore Search



Pharmer

Efficient and Exact Pharmacophore Search





http://pharmit.csb.pitt.edu

Kinds of Virtual Screening

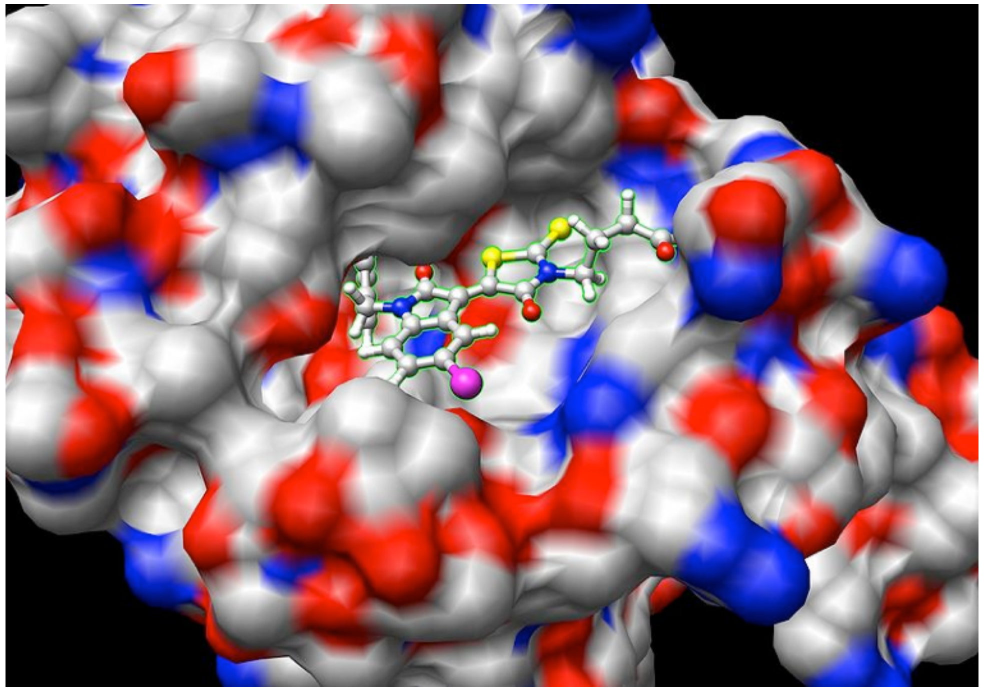
ADMET

Ligand Based

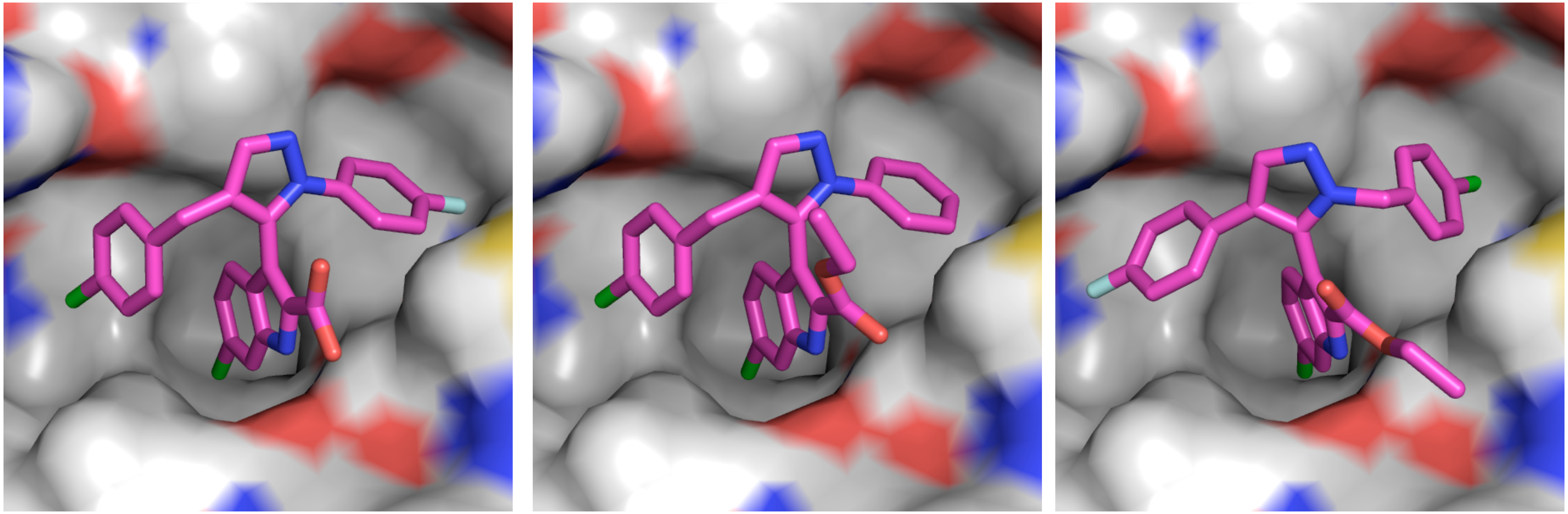
- similarity to known binder
- QSAR
- pharmacophore

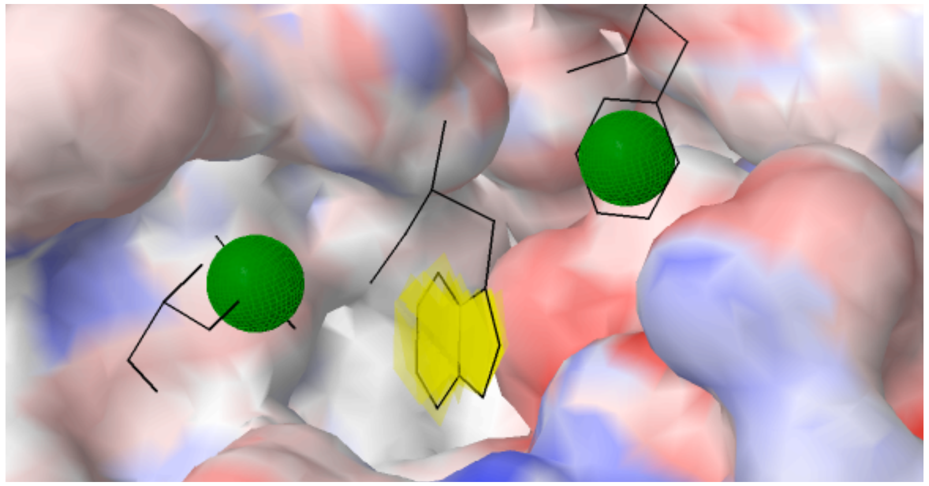
Receptor Based

- **dock and score**

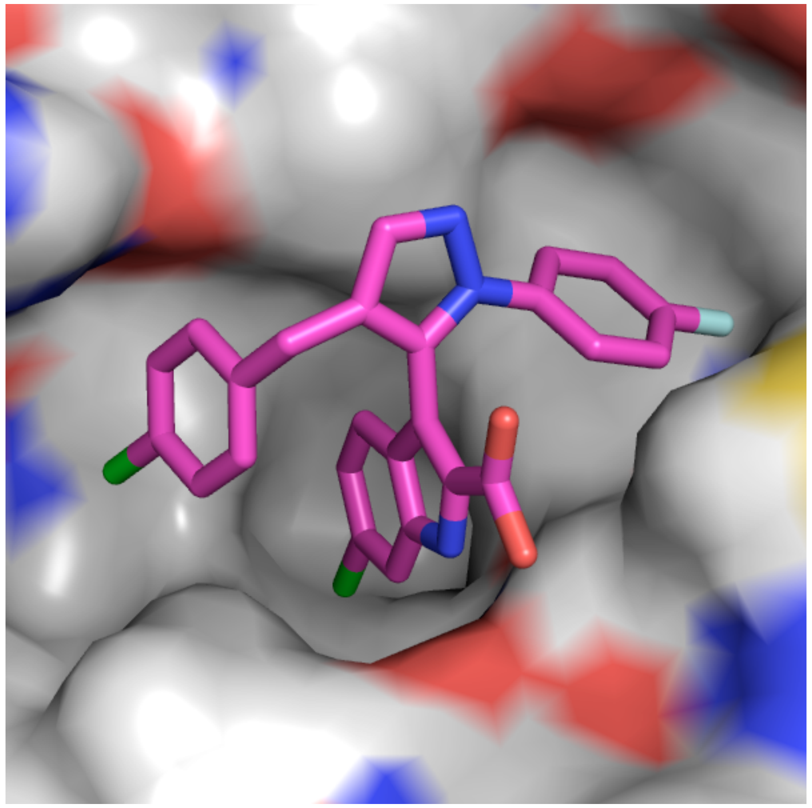


Pharmacophores Aren't Enough

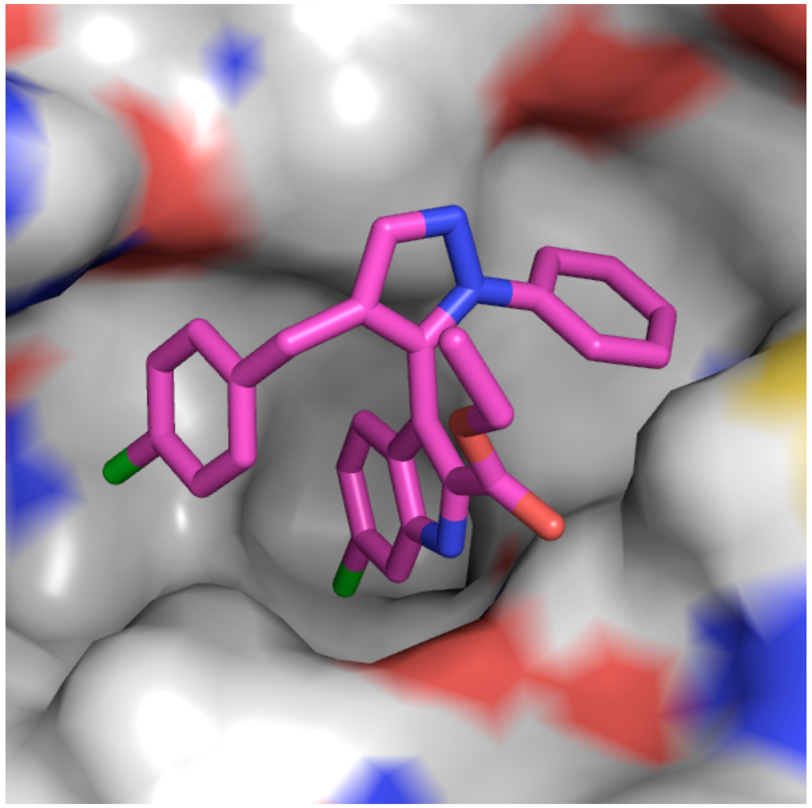




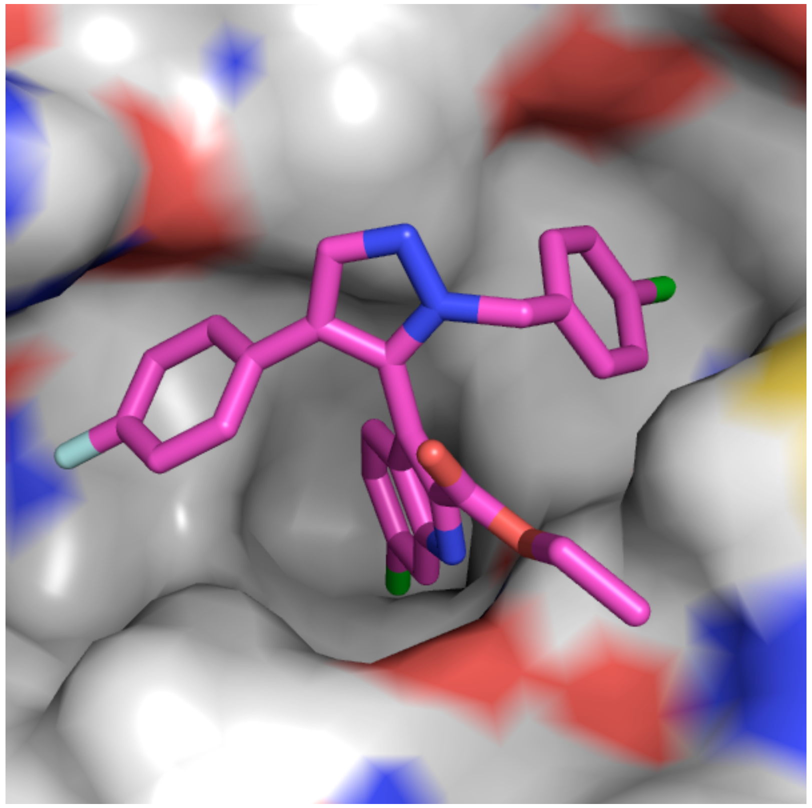
Pharmacophores Aren't Enough



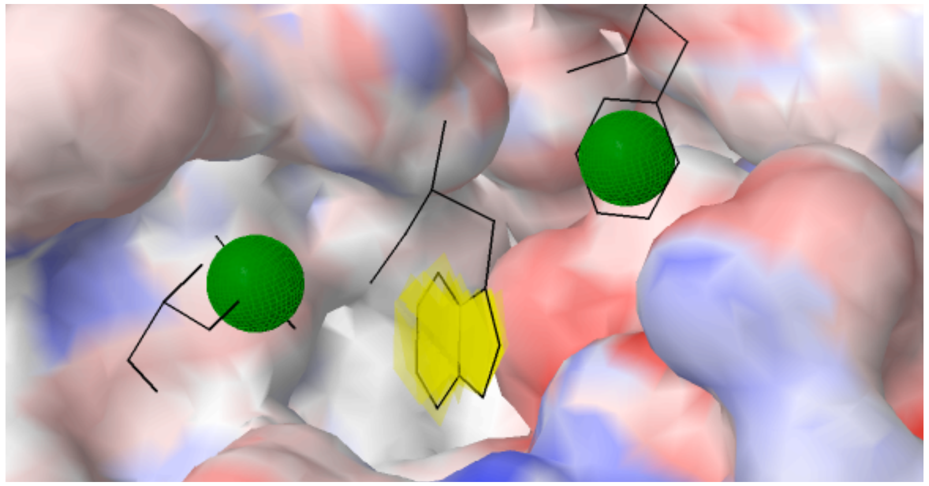
.2µM



50µM



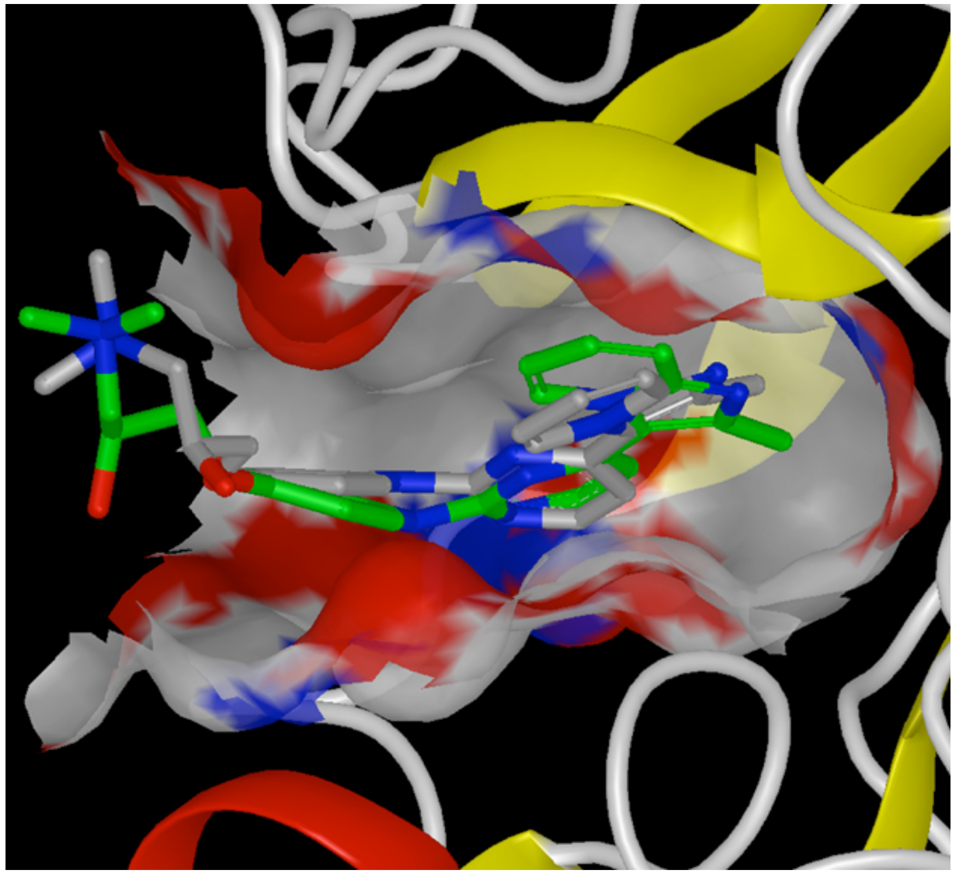
n.i.



Docking

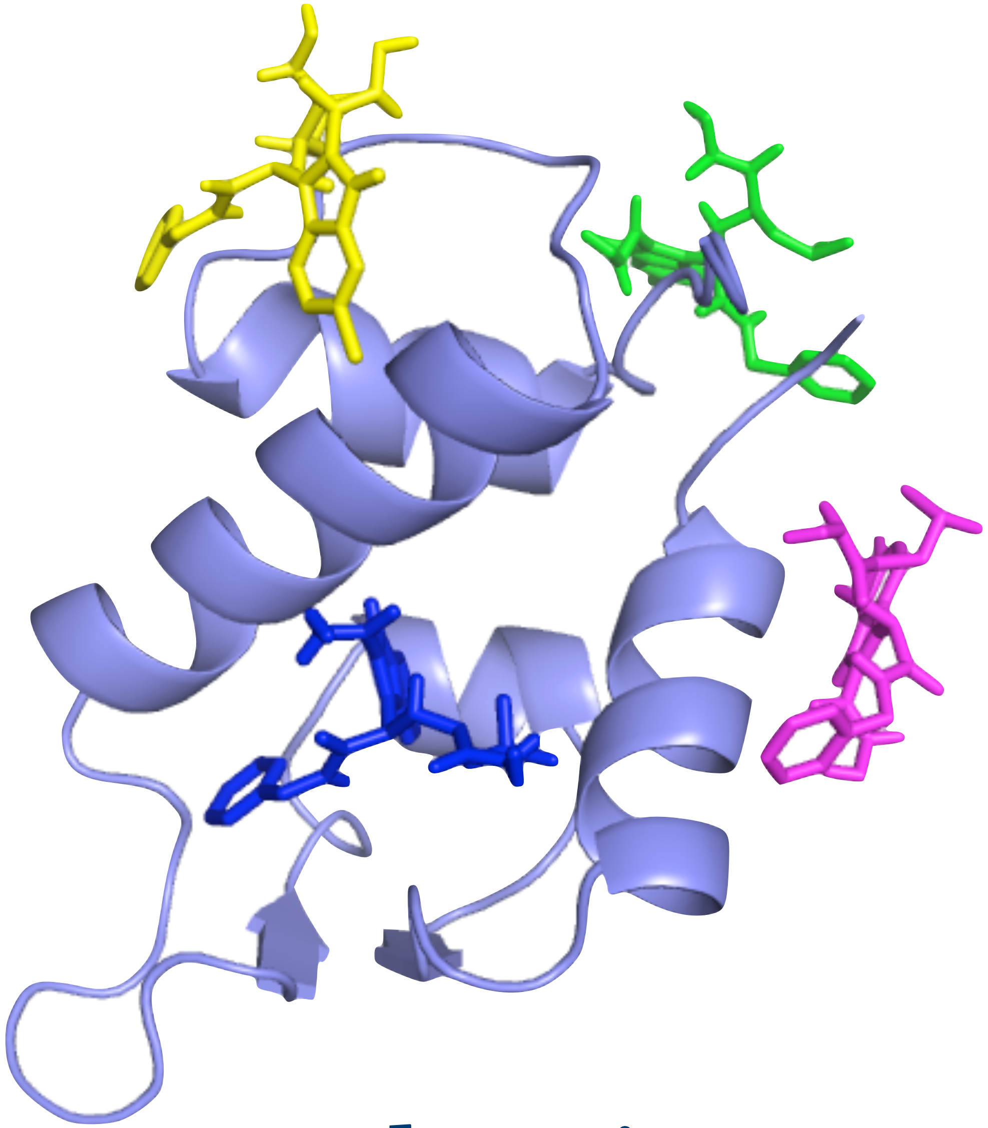
Determine the **conformation** and **pose** of a ligand at a docking site

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



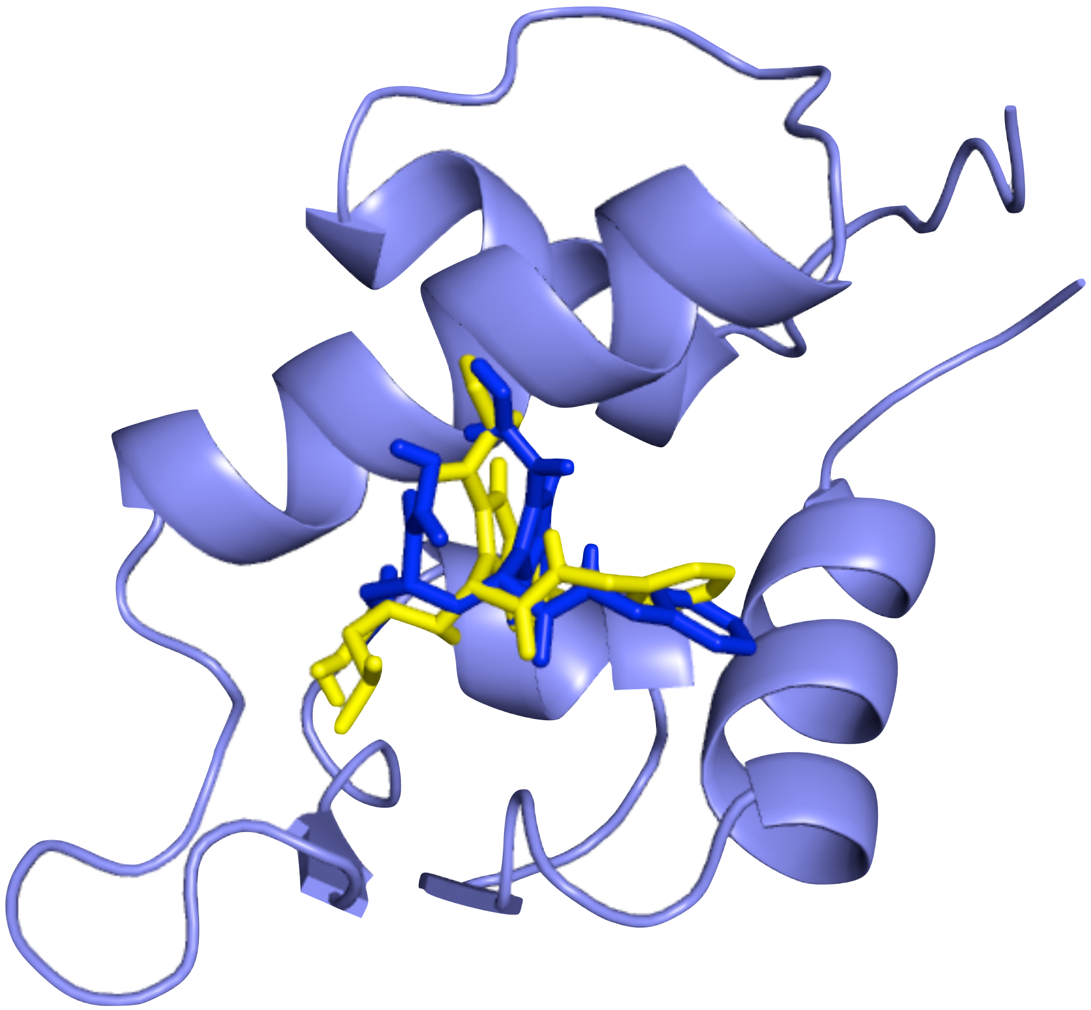
Two Phase Docking

1. Global Pose Estimation



Stochastic

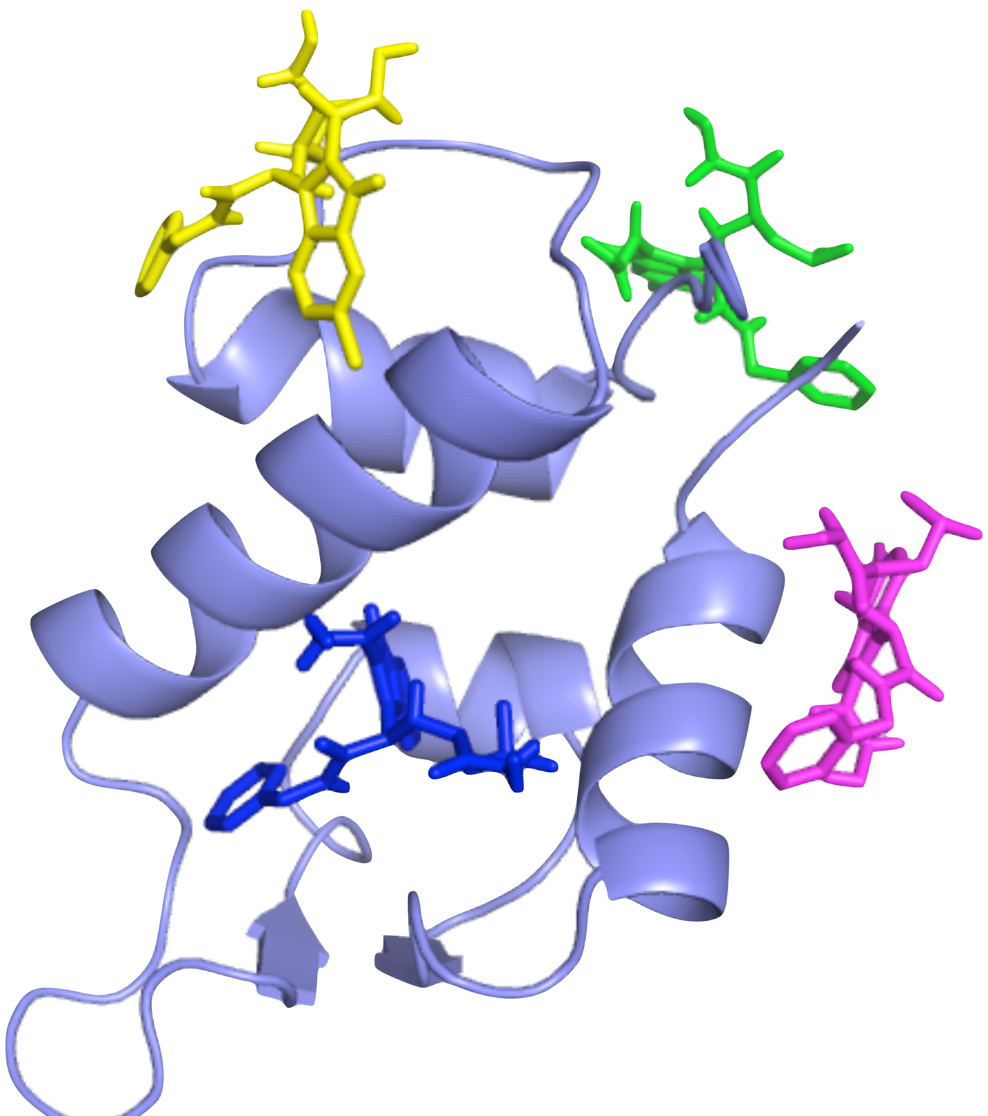
2. Local Refinement



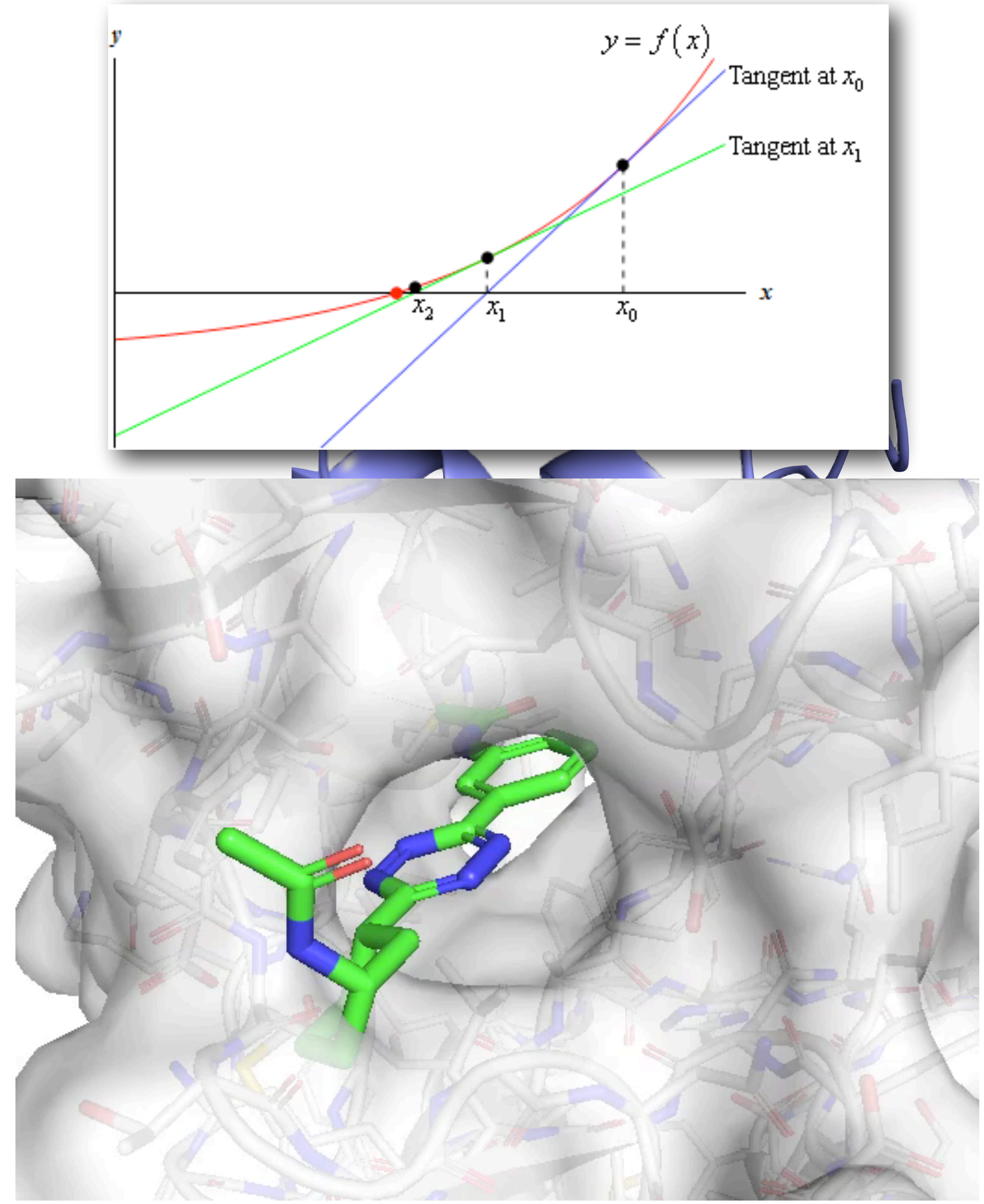
Minimization

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?

Scoring Goals

Affinity Prediction

-how well does it bind?

Inactive/Active Discrimination

-does it bind?

Pose Prediction

-how does it bind?

Speed

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 known structure, not physical principles

Consensus

Force Field Scoring

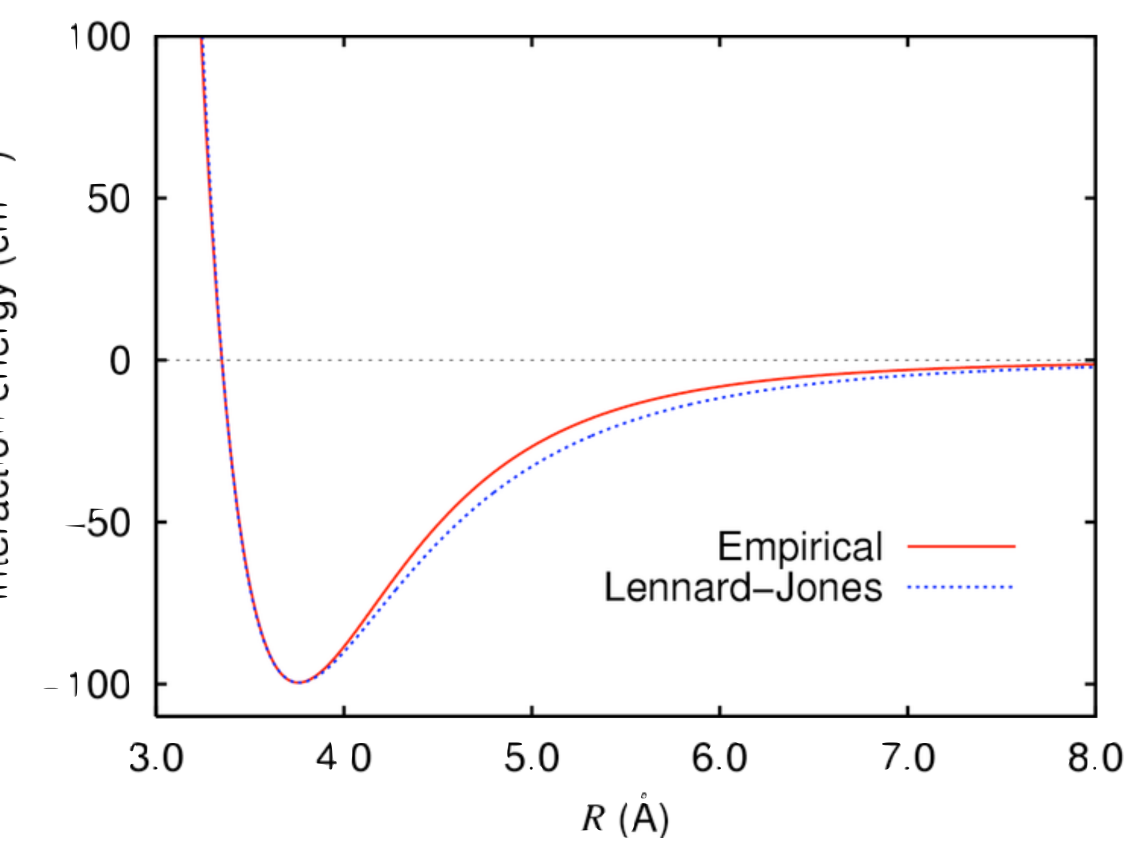
	Protein–ligand	Internal ligand
G-Score	$E_{vdW} + E_{H-bond} = \sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{8}} - \frac{B_{ij}}{d_{ij}^{4}}\right) + (E_{da} + E_{ww}) - (E_{dw} + E_{aw})\right]$	$E_{vdw} + E_{torsion} = \sum_{lig}\left(\frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{6}}\right) + \sum_{lig}\frac{1}{2}V\left[1 + \frac{n}{\lvert n\rvert}\cos(\lvert n\rvert\omega)\right]$
D-Score	$E_{vdW} + E_{electrostatic} = \sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{12}} + \frac{B_{ij}}{d_{ij}^{6}}\right) + 332.0\frac{q_i q_j}{\in(d_{ij})d_{ij}}\right]$	
Gold	$E_{vdW} + E_{electrostatic} = \sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{a}} + \frac{B_{ij}}{d_{ij}^{b}}\right) + 332.0\frac{q_i q_j}{\in(d_{ij})d_{ij}}\right]$	$E_{vdW} + E_{electrostatic} = \sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{a}} + \frac{B_{ij}}{d_{ij}^{b}}\right) + 332.0\frac{q_i q_j}{\in(d_{ij})d_{ij}}\right]$ $+ \text{optional } E_{H-bond}$
AutoDock	$E_{vdW} + E_{H-bond} + E_{electrostatic} = \sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^{6}}\right) + E(t)\times\left(\frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}}\right) + 332.0\frac{q_i q_j}{\in(d_{ij})d_{ij}}\right]$ $E(t) = \text{angular weight factor}$	$E_{vdW} + E_{H-bond} + E_{electrostatic} = \sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^{6}}\right) + E(t)\left(\frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}}\right) + 332.0\frac{q_i q_j}{4(d_{ij})d_{ij}}\right]$ $E(t) = \text{angular weight factor}$
DOCK (v4.0)	$E_{vdW} + E_{electrostatic} = \sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{a}} + \frac{B_{ij}}{d_{ij}^{b}}\right) + 332.0\frac{q_i q_j}{\in(d_{ij})d_{ij}}\right]$	

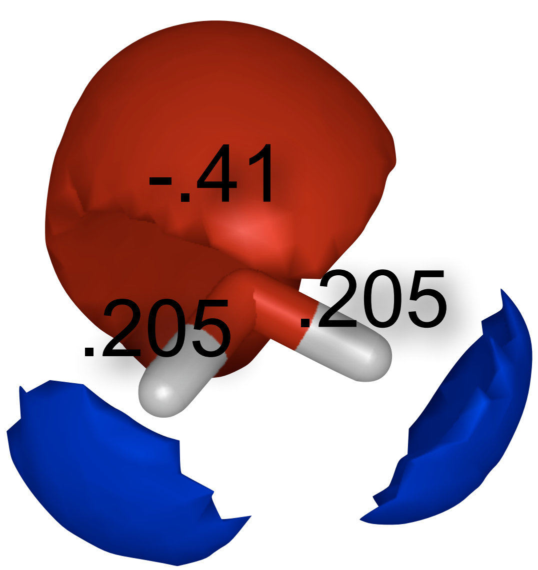
Dock 4.0

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 Scoring

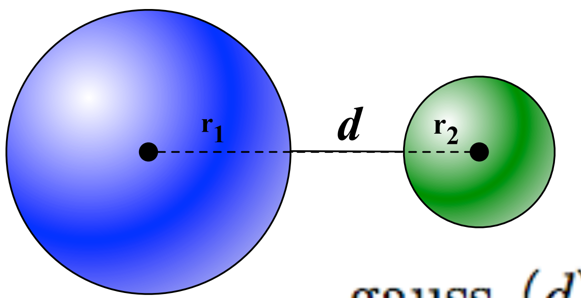
	Functional form
LUDI	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta\alpha) + \Delta G_{hydrophobic} \sum_{hydrophobic} \lvert A_{hydrophobic} \rvert + \Delta G_{rotor} N_{rotor} + \Delta G_0$ $A_{hydrophobic}$ = molecular surface area
F-Score	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta\alpha) + \Delta G_{aromatic} \sum_{aromatic} f(\Delta R, \Delta\alpha) + \Delta G_{contact} \sum_{contact} f(\Delta R, \Delta\alpha) + \Delta G_{rotor} N_{rotor} + \Delta G_0$
Chem-Score	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{metal} \sum_{metal} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} \sum_{lipo} f(\Delta R) + \Delta G_{rotor} \sum_{rotor} f(P_{nl}, P'_{nl}) + \Delta G_0$

Empirical Scoring

	Functional form
LUDI	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta\alpha) + \Delta G_{hydrophobic} \sum_{hydrophobic} \lvert A_{hydrophobic} \rvert + \Delta G_{rotor} N_{rotor} + \Delta G_0$ $A_{hydrophobic} = \text{molecular surface area}$
F-Score	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta\alpha) + \Delta G_{aromatic} \sum_{aromatic} f(\Delta R, \Delta\alpha) + \Delta G_{contact} \sum_{contact} f(\Delta R, \Delta\alpha) + \Delta G_{rotor} N_{rotor} + \Delta G_0$
Chem-Score	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{metal} \sum_{metal} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} \sum_{lipo} f(\Delta R) + \Delta G_{rotor} \sum_{rotor} f(P_{nl}, P'_{nl}) + \Delta G_0$

regression coefficient

AutoDock Vina



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

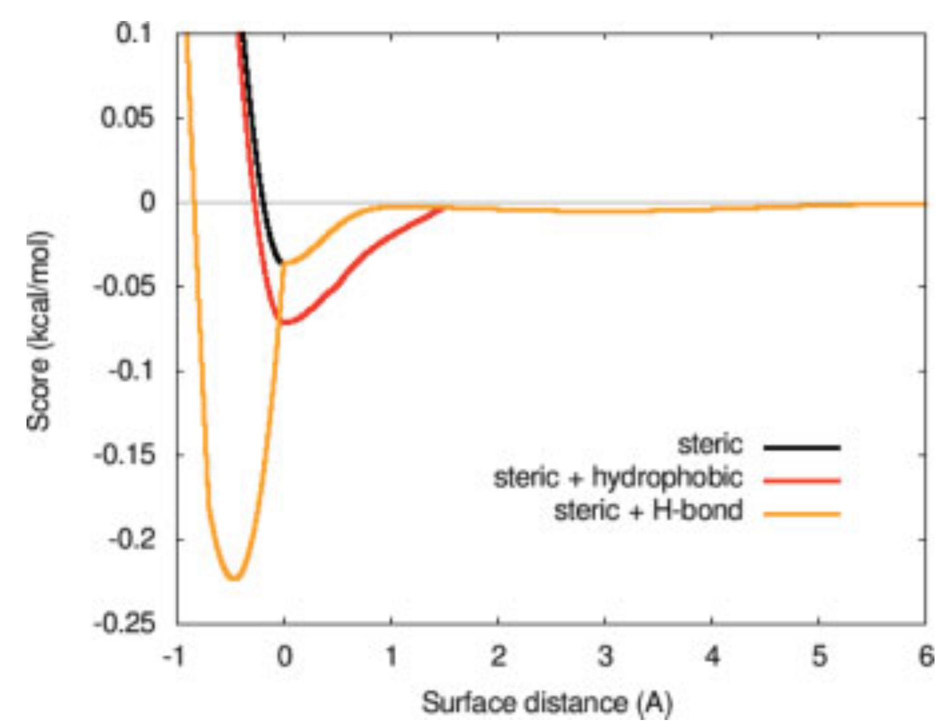
$$\mathrm{gauss}_2(d) = w_{\mathrm{guass}_2} e^{-((d-3)/2)^2}$$

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

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

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

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

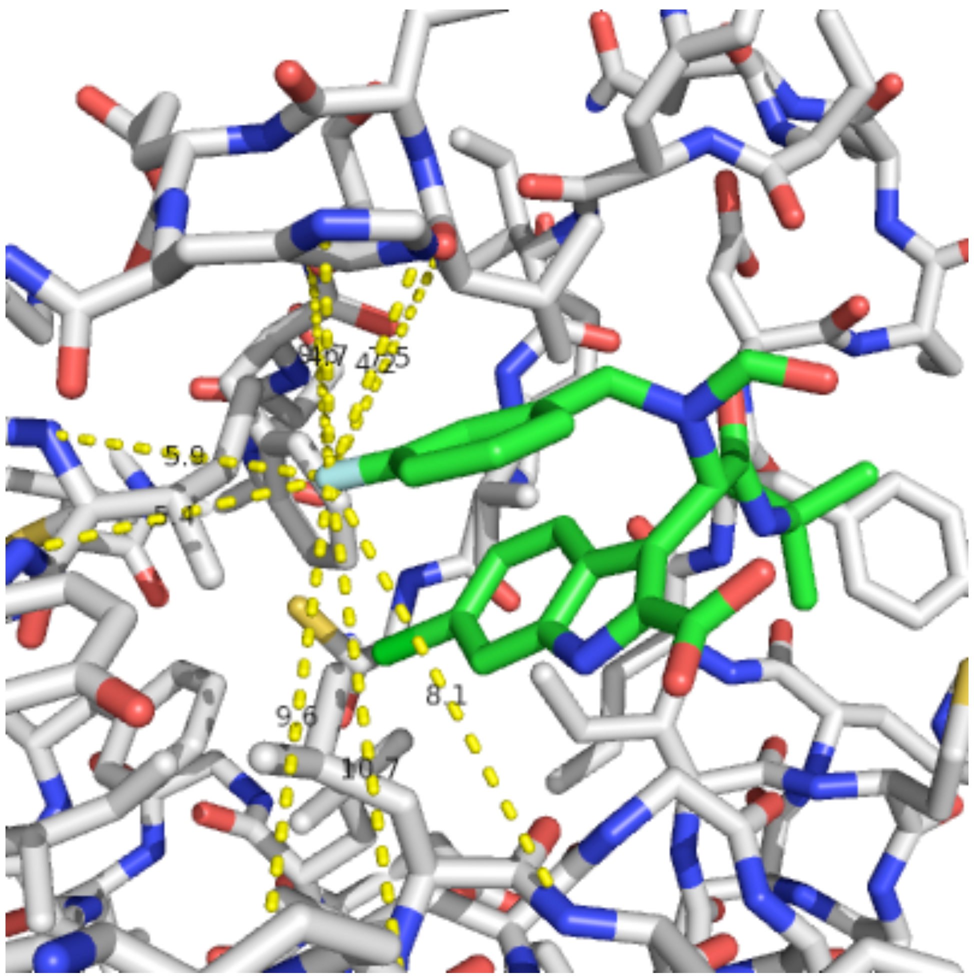


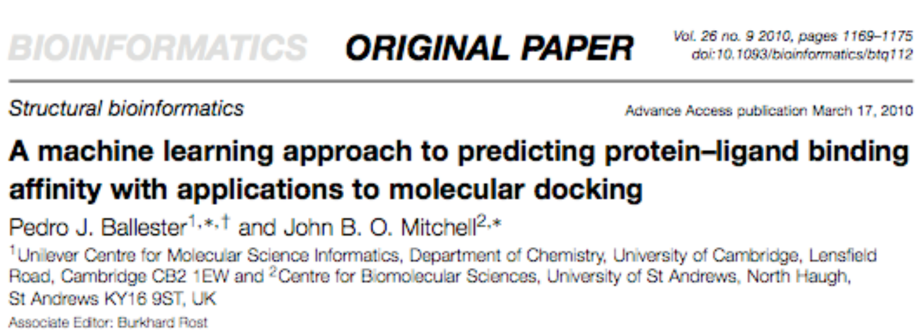
Knowledge Based

	Functional form
PMF	Parametrized pairwise potential PMF score: $PMF = \sum_{prot}\sum_{lig} A_{ij}(d_{ij}) \quad A_{ij}(d_{ij}) = -k_B T \ln\left[f_{Vol\_corr}^{j}(r) \frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}} \right]$ where k_B is the Boltzmann constant, $f_{Vol\_corr}^{j}(r)$ is a ligand volume correction factor and $\frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}}$ indicates a radial distribution function for a protein atom i and a ligand atom j.
DrugScore (v1.2)	$\Delta W = \gamma \sum_{prot}\sum_{lig} \Delta W_{ij}(r) + (1-\gamma) \times \left[\sum_{lig} \Delta W_i(SAS, SAS_0) + \sum_{prot} \Delta W_j(SAS, SAS_0) \right]$ SAS = Solvent accessible surface area terms, W_{ij} = distance dependent pairwise potential
SMoG	$G = \sum_{ij} g_{ij}\Delta_{ij}; \qquad \Delta_{ij} = \begin{cases} 0 & \text{(i, j more than 5 Å)} \\ 1 & \text{(i, j within 5 Å)} \end{cases}; \qquad g_{ij} = -kT \log\left[\frac{p_{ij}}{\overline{p}}\right];$ p_{ij} an $\overline{p}$ are interatomic and averaged interactomic interactions

RF-Score

Pairwise Distance Counts (<12Å)




BIOINFORMATICS ORIGINAL PAPER Vol. 26 no. 9 2010, pages 1169–1175
doi:10.1093/bioinformatics/btq112

Structural bioinformatics — Advance Access publication March 17, 2010

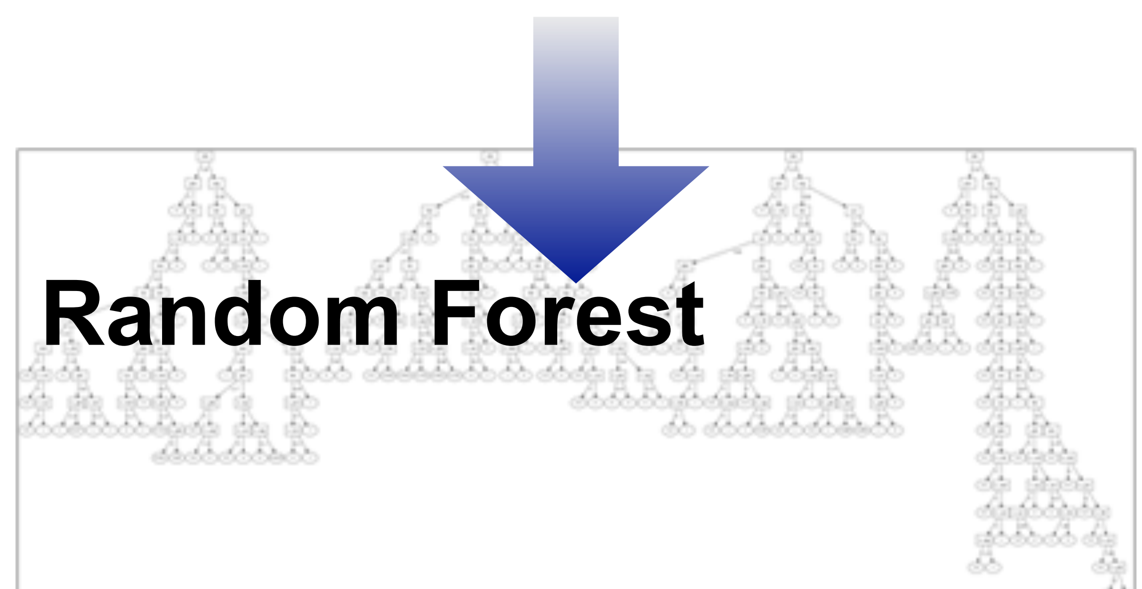
A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking

Pedro J. Ballester[1,*,†] and John B. O. Mitchell[2,*]
[1]Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW and [2]Centre for Biomolecular Sciences, University of St Andrews, North Haugh, St Andrews KY16 9ST, UK
Associate Editor: Burkhard Rost

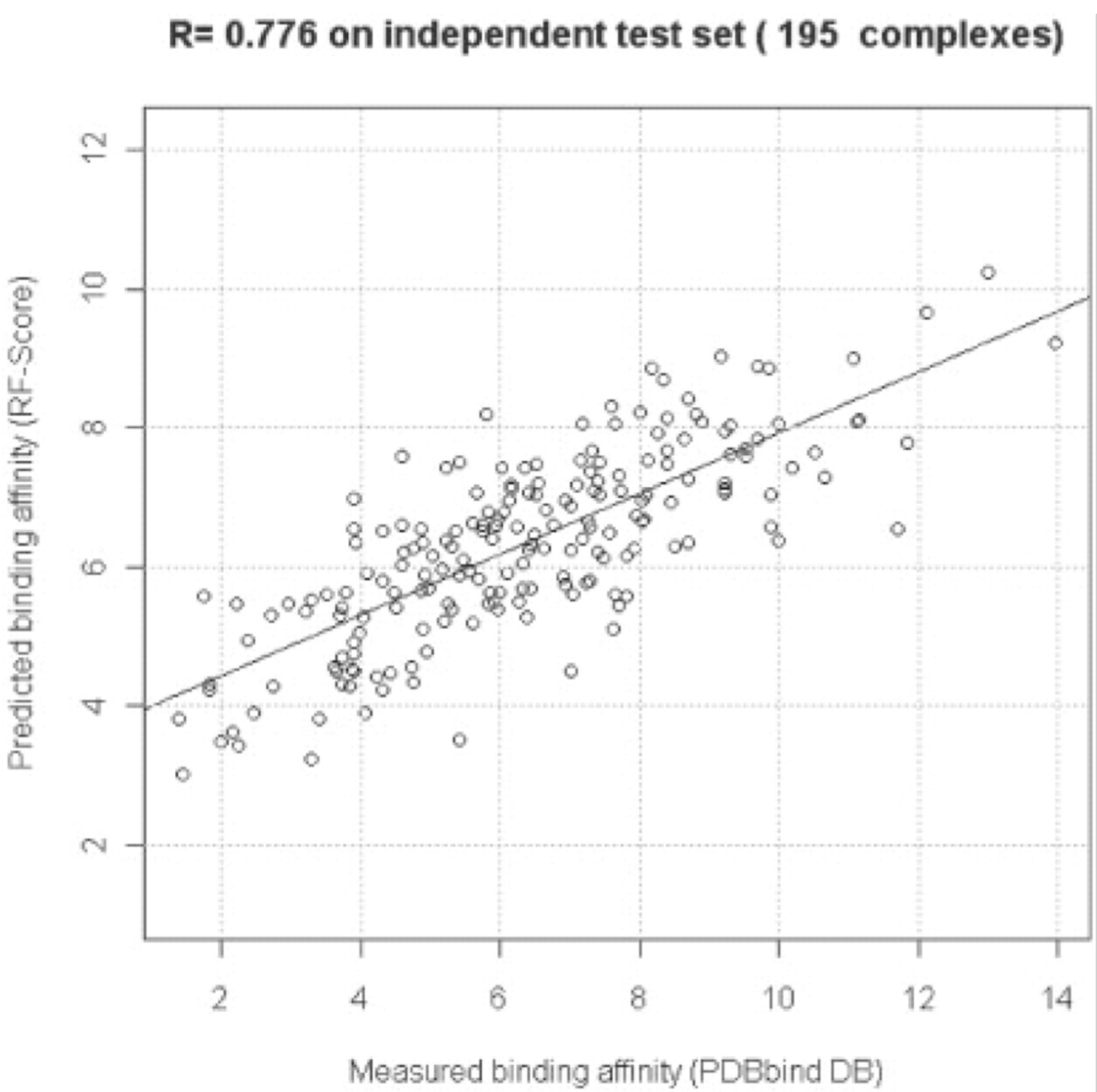
Protein

Ligand

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



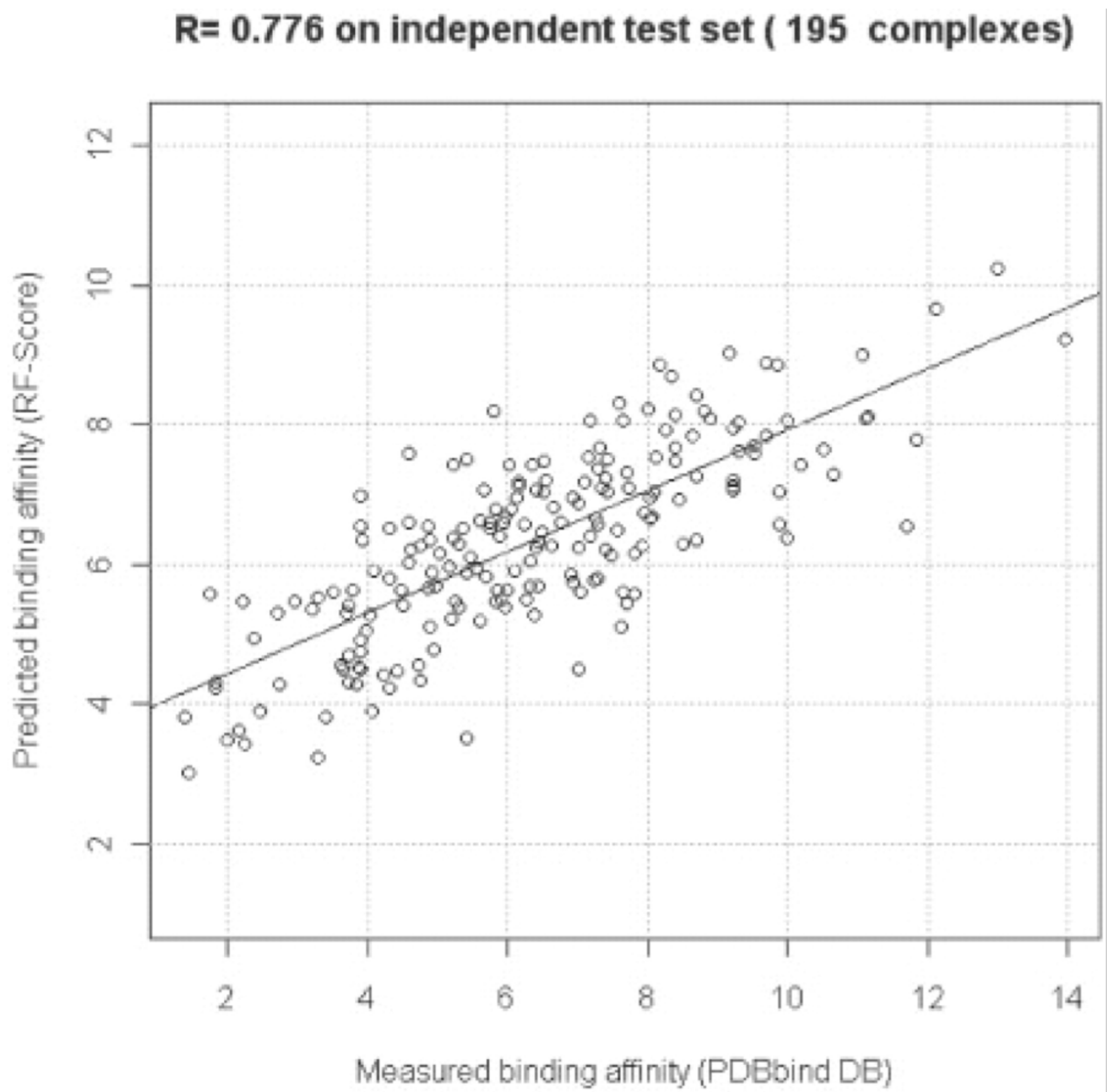
RF-Score Output



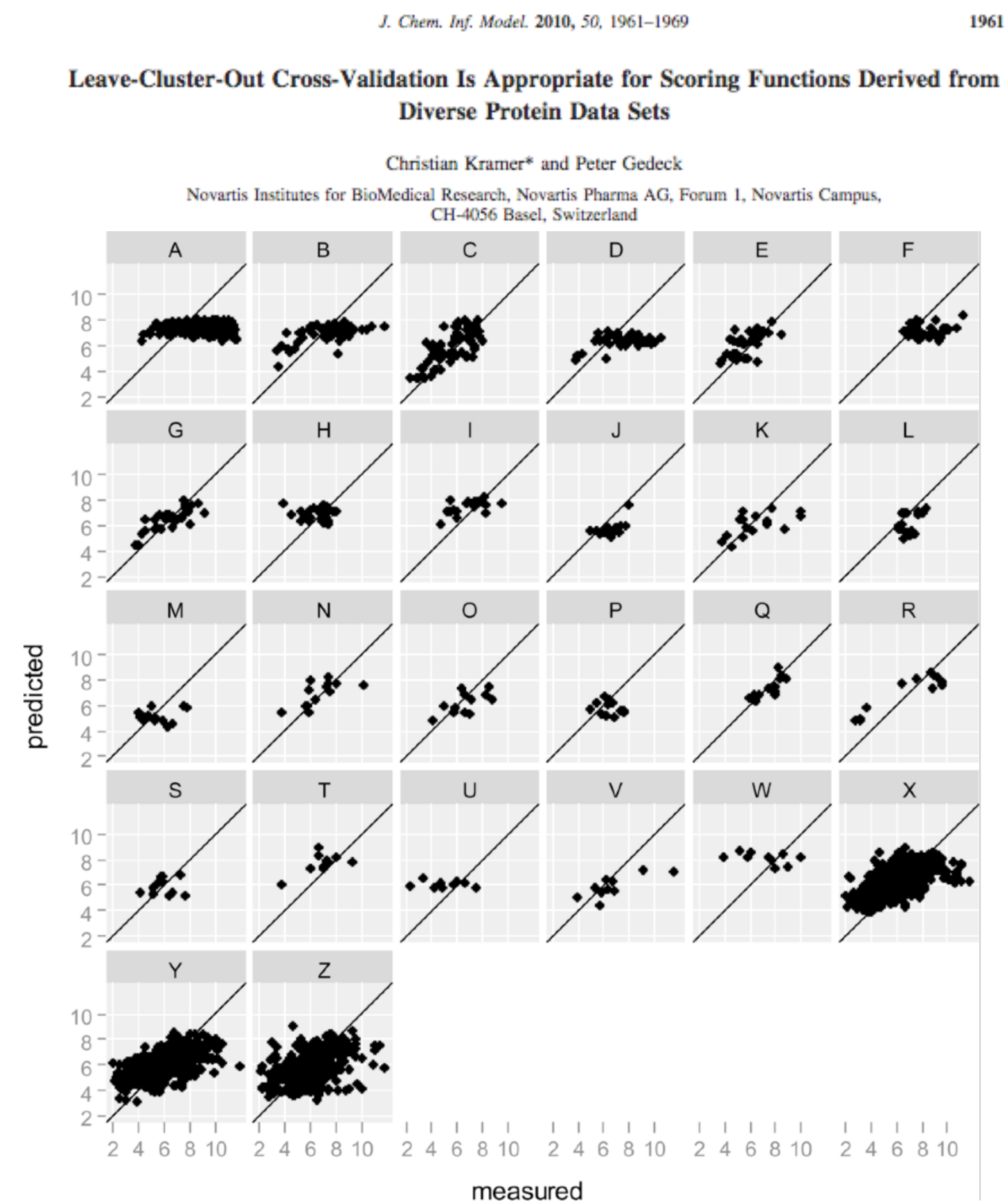
RMSE = 1.58

Scoring function	*R*	*Rs*	RMSE
RF-Score	**0.776**	**0.762**	**1.58**
X-Score::HMScore	0.644	0.705	1.83
DrugScore^{CSD}	0.569	0.627	1.96
SYBYL::ChemScore	0.555	0.585	1.98
DS::PLP1	0.545	0.588	2
GOLD::ASP	0.534	0.577	2.02
SYBYL::G-Score	0.492	0.536	2.08
DS::LUDI3	0.487	0.478	2.09
DS::LigScore2	0.464	0.507	2.12
GlideScore-XP	0.457	0.435	2.14
DS::PMF	0.445	0.448	2.14
GOLD::ChemScore	0.441	0.452	2.15
SYBYL::D-Score	0.392	0.447	2.19
DS::Jain	0.316	0.346	2.24
GOLD::GoldScore	0.295	0.322	2.29
SYBYL::PMF-Score	0.268	0.273	2.29
SYBYL::F-Score	0.216	0.243	2.35

RF-Score Output



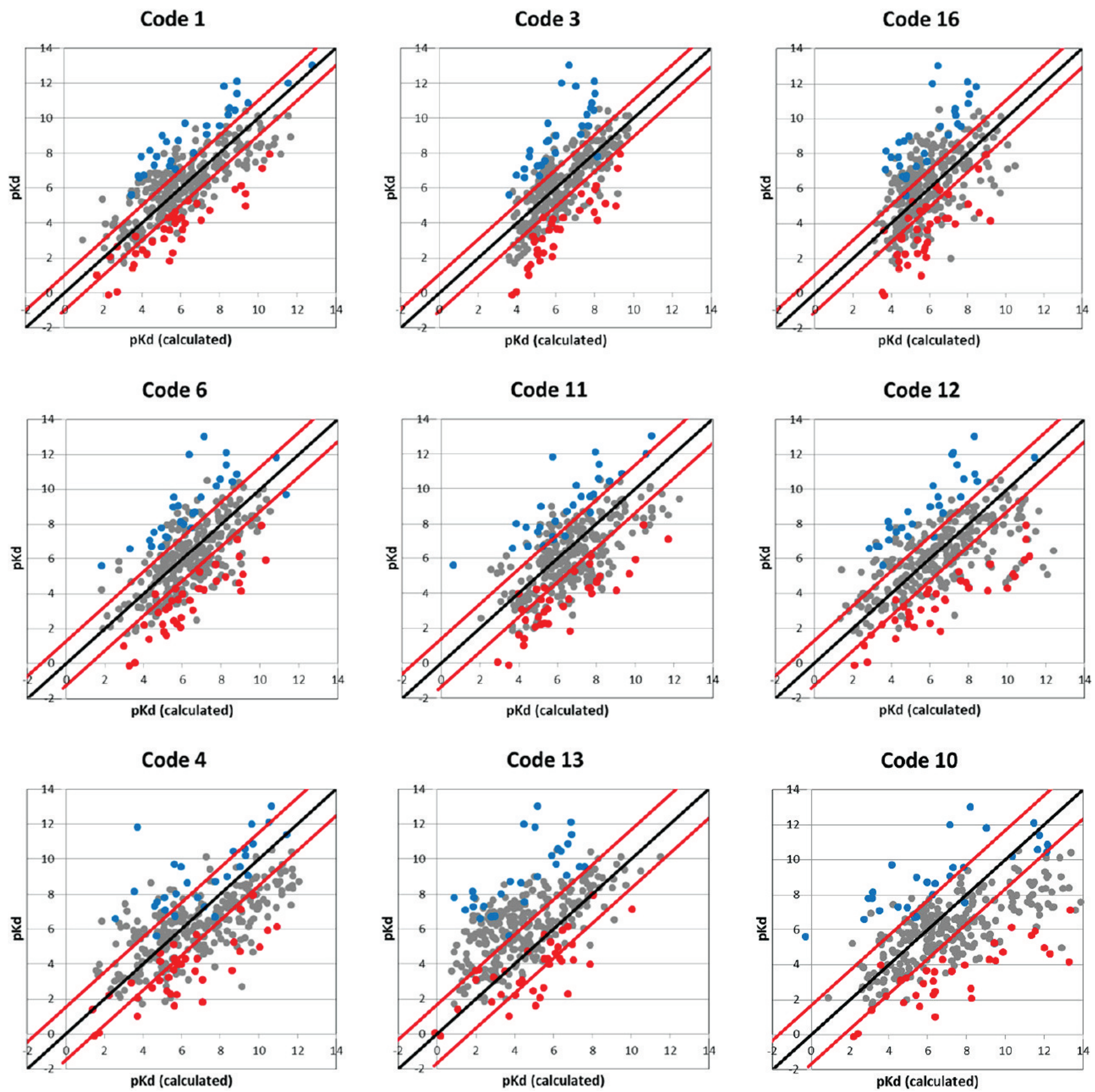
RMSE = 1.58



R = 0.46; RMSE = 1.6

Scoring

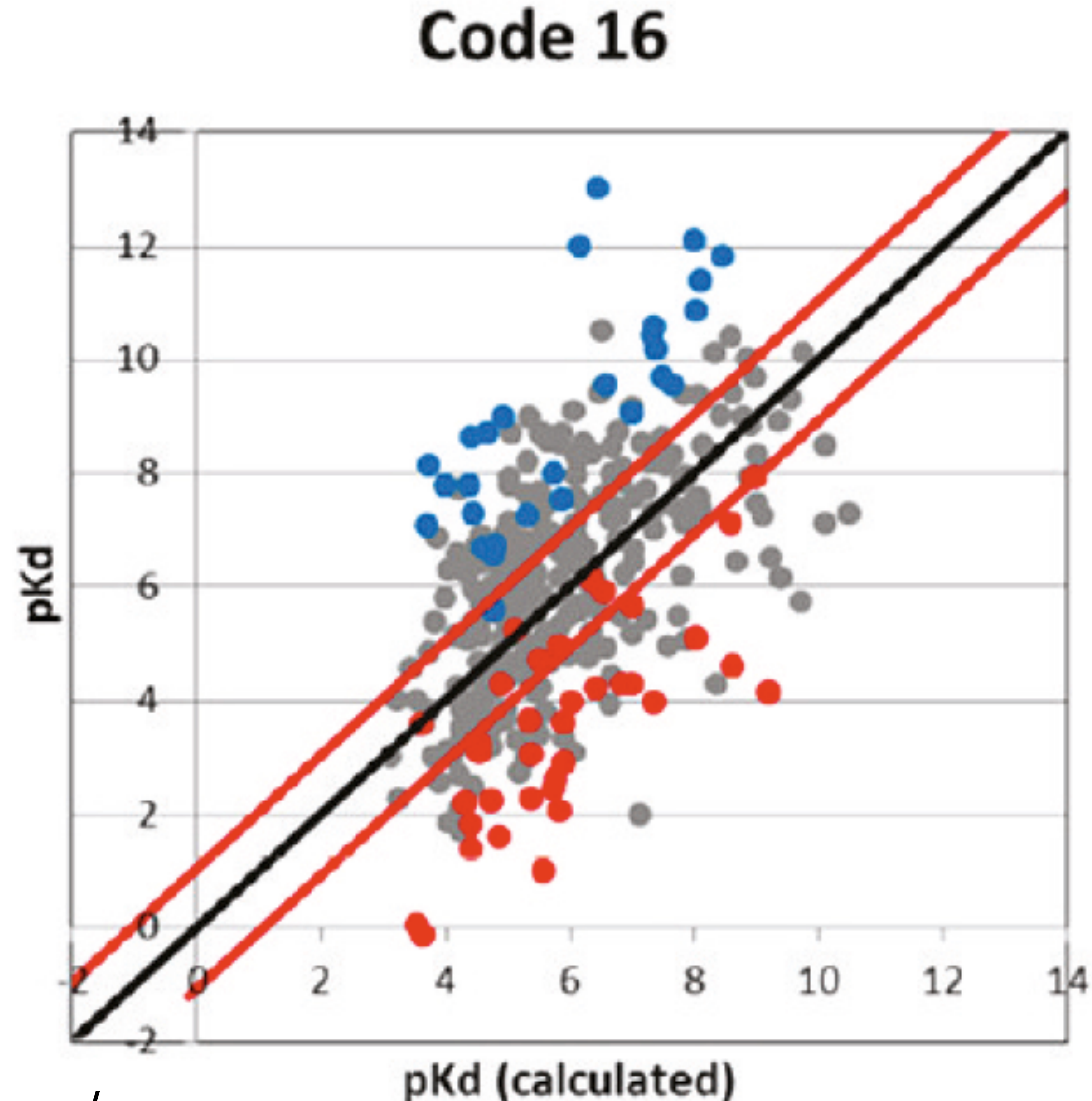
Ideally, score would equal affinity – but this is an unsolved problem.



http://www.csardock.org/

Scoring

Ideally, score would equal affinity – but this is an unsolved problem.

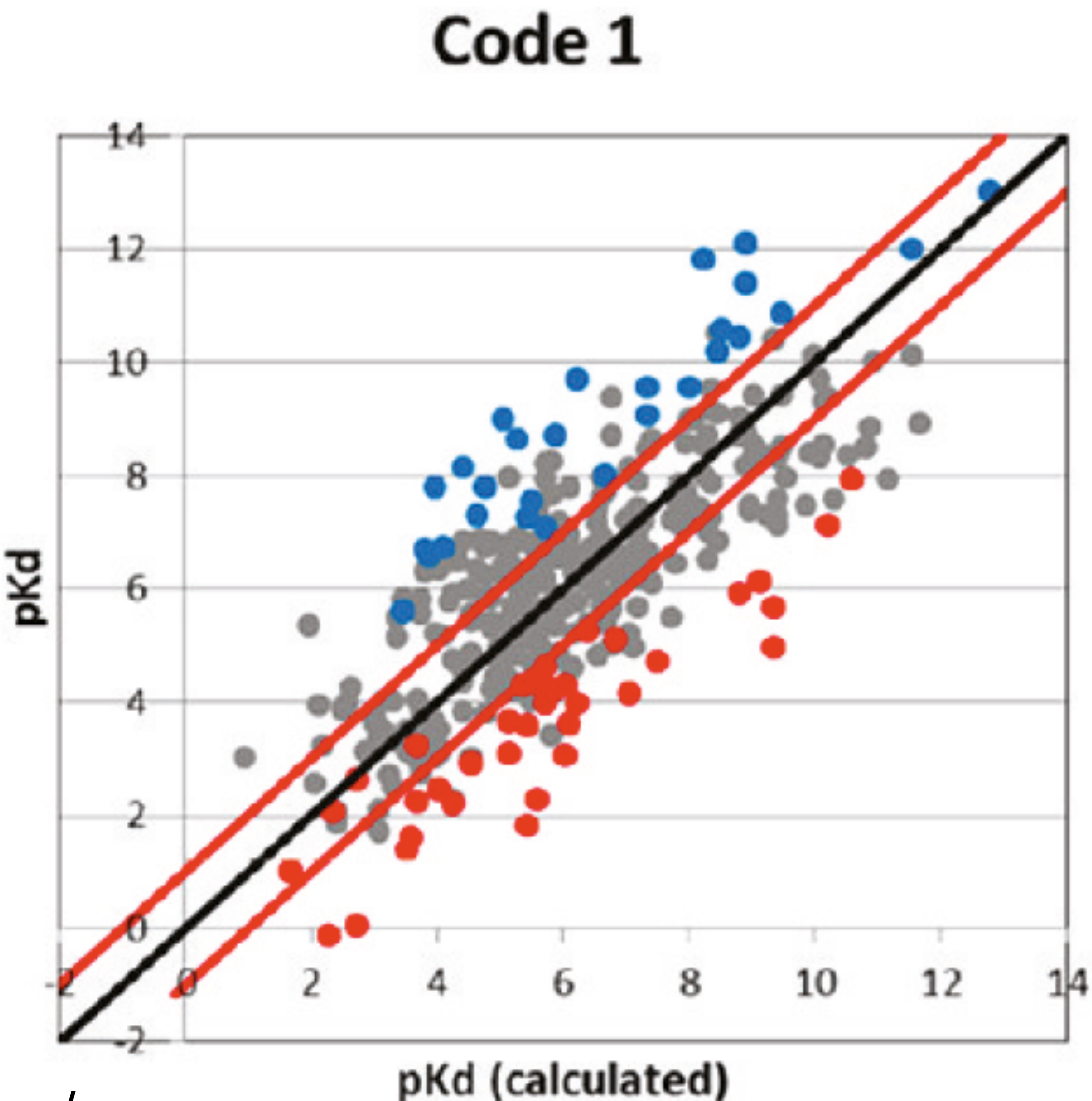


$R^2 = 0.28$
RMSE = 1.9

http://www.csardock.org/

Scoring

Ideally, score would equal affinity – but this is an unsolved problem.

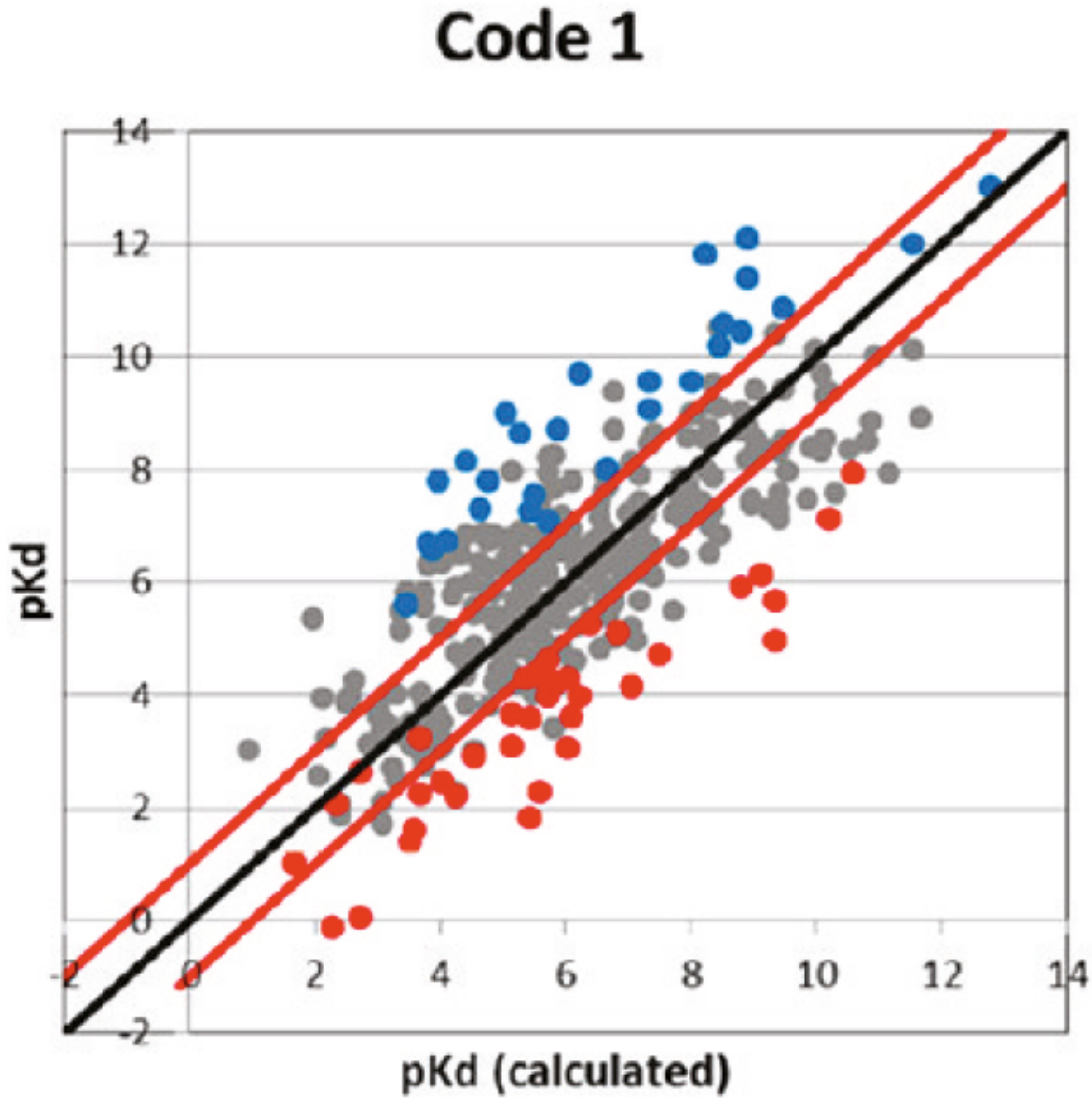


$R^2 = 0.58$
RMSE = 1.51

http://www.csardock.org/

Scoring

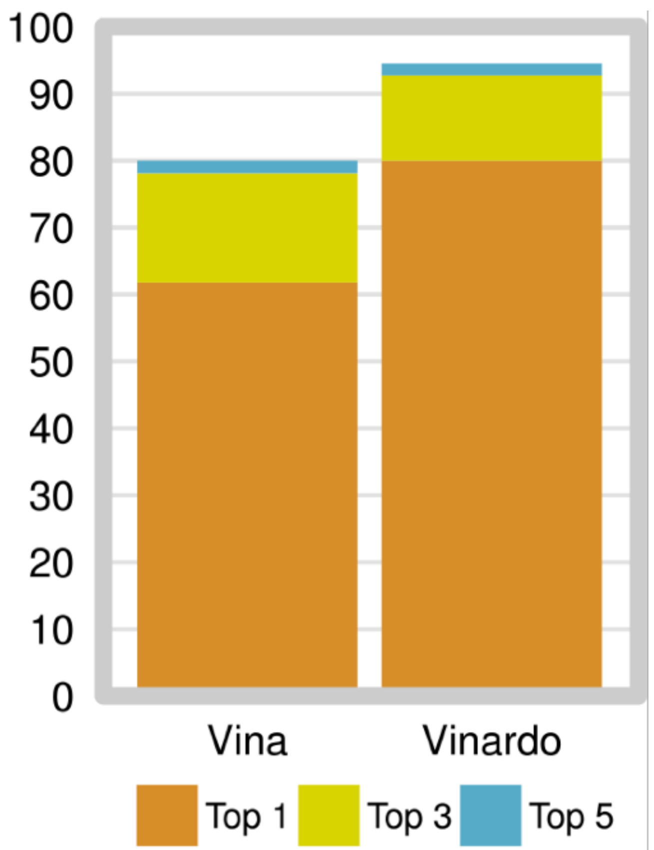
Ideally, score would equal affinity – but this is an unsolved problem.

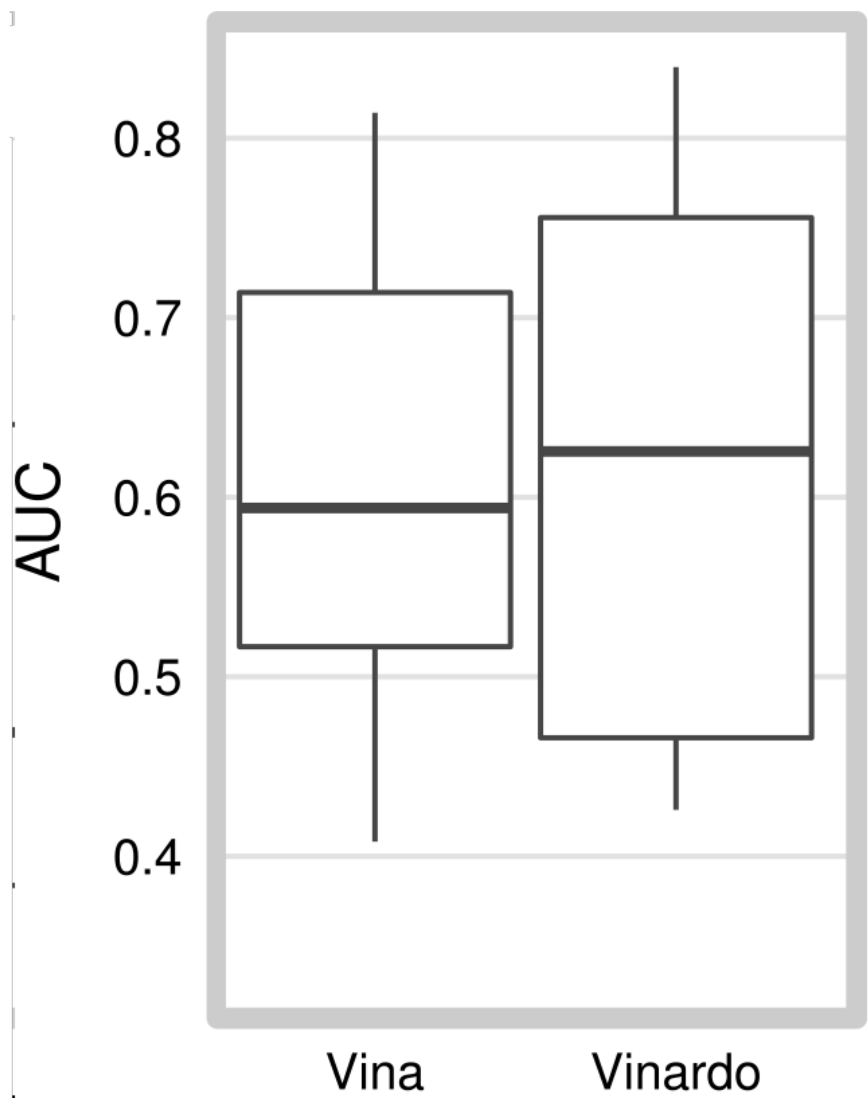


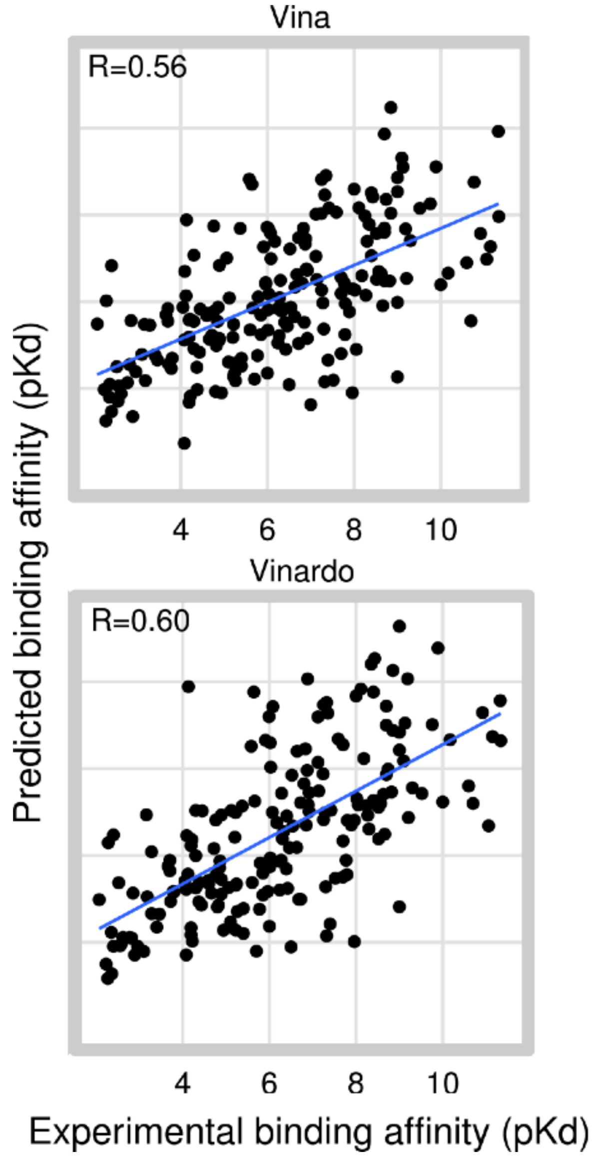
$R^2 = 0.58$
RMSE = 1.51

http://www.csardock.org/

Scoring State of the Art







Pose Prediction | Binding Discrimination | Affinity Prediction

Quiroga R, Villarreal MA (2016) Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLoS ONE* 11(5): e0155183. doi:10.1371/journal.pone.0155183

Can we do better?

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

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 PubCher

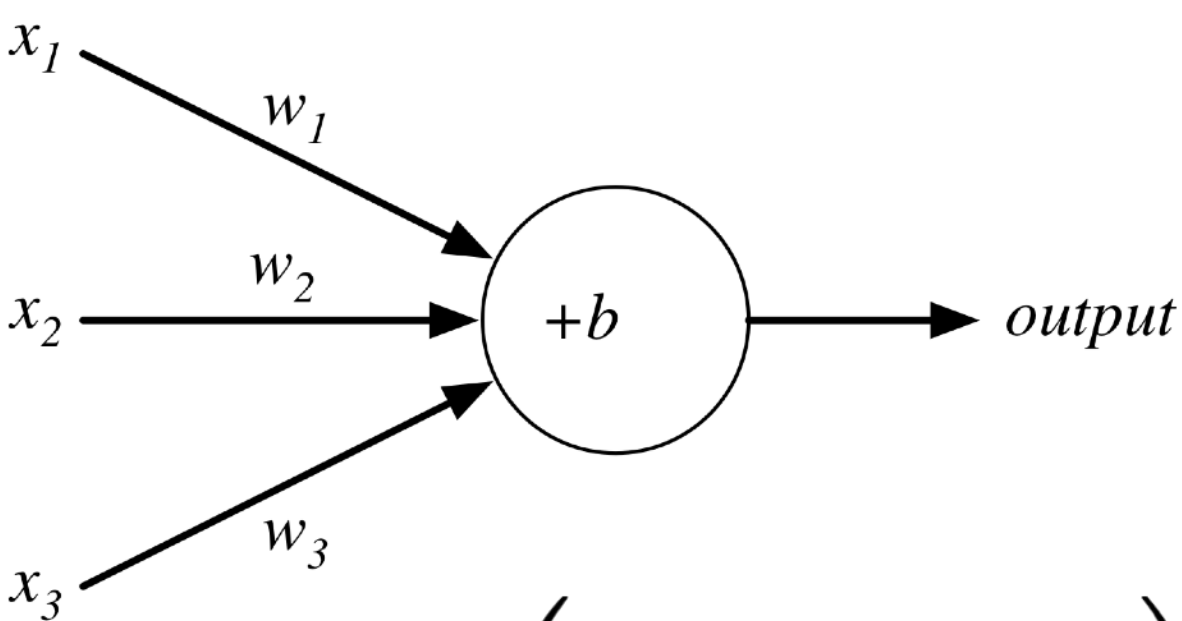
125,526 structures in the PDB

16,179 annotated complexes in PDBbind

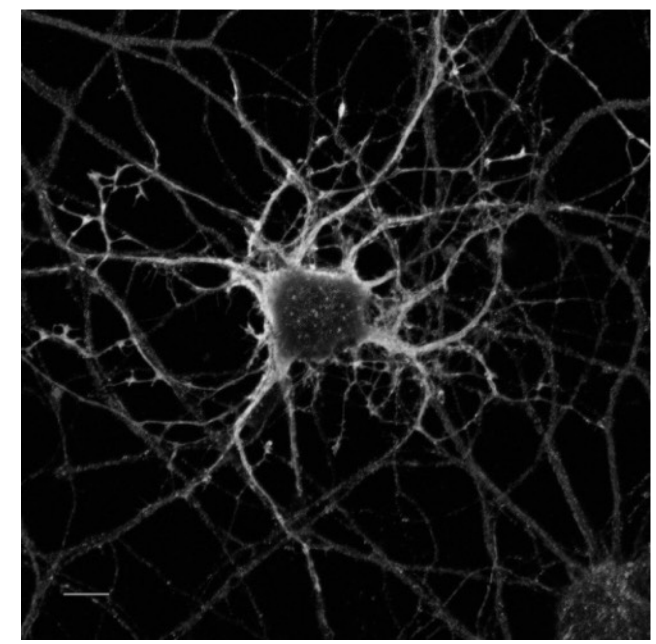
Machine Learning

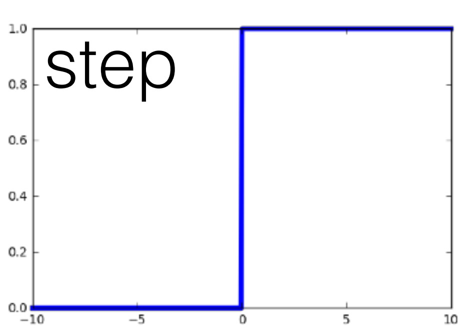
Features $X \rightarrow$ **Model** $\rightarrow y$ Prediction

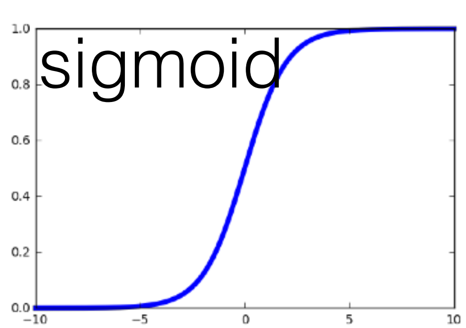
Neural Networks

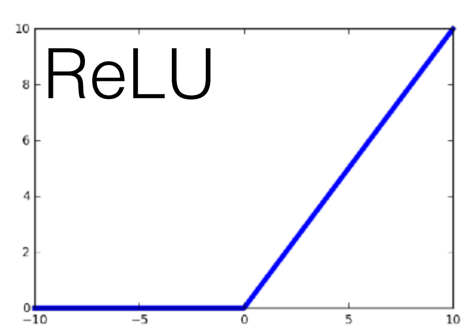


$$output = \sigma\left(\sum_i w_i x_i + b\right)$$

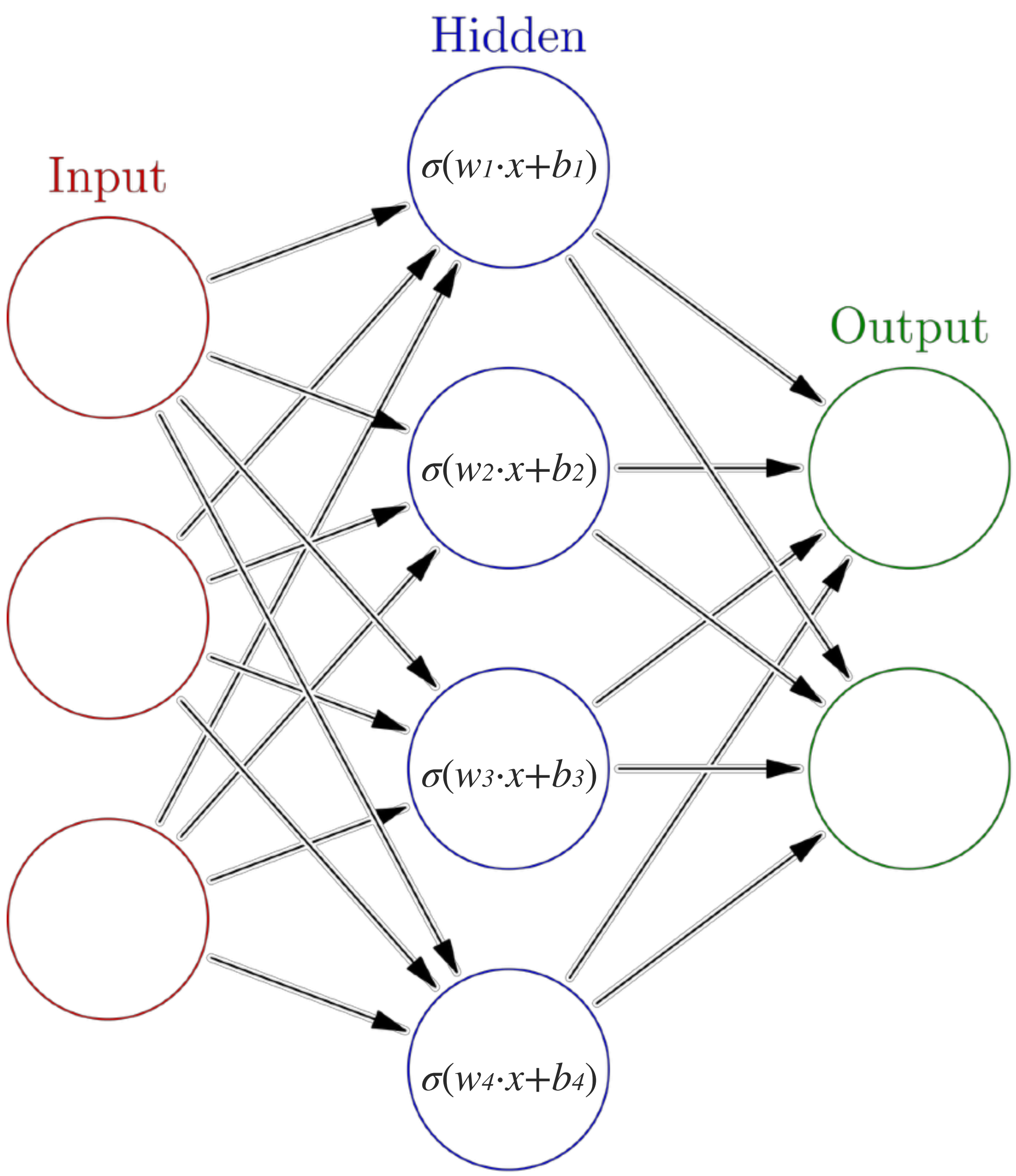


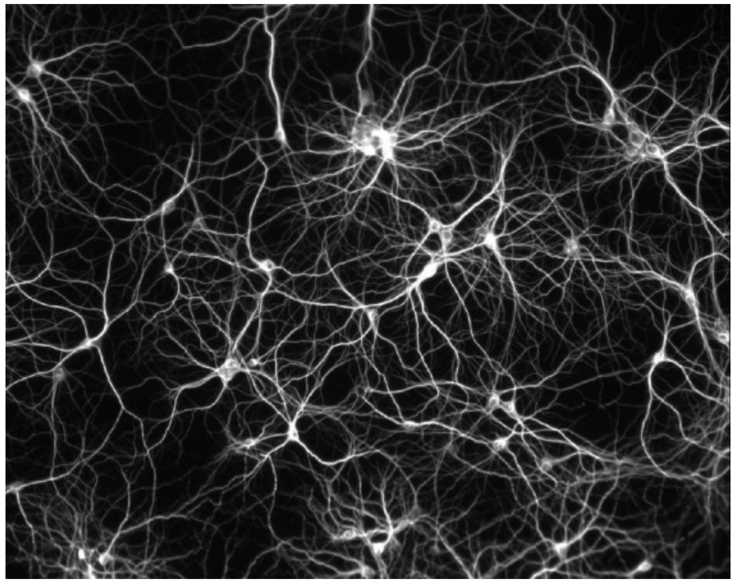






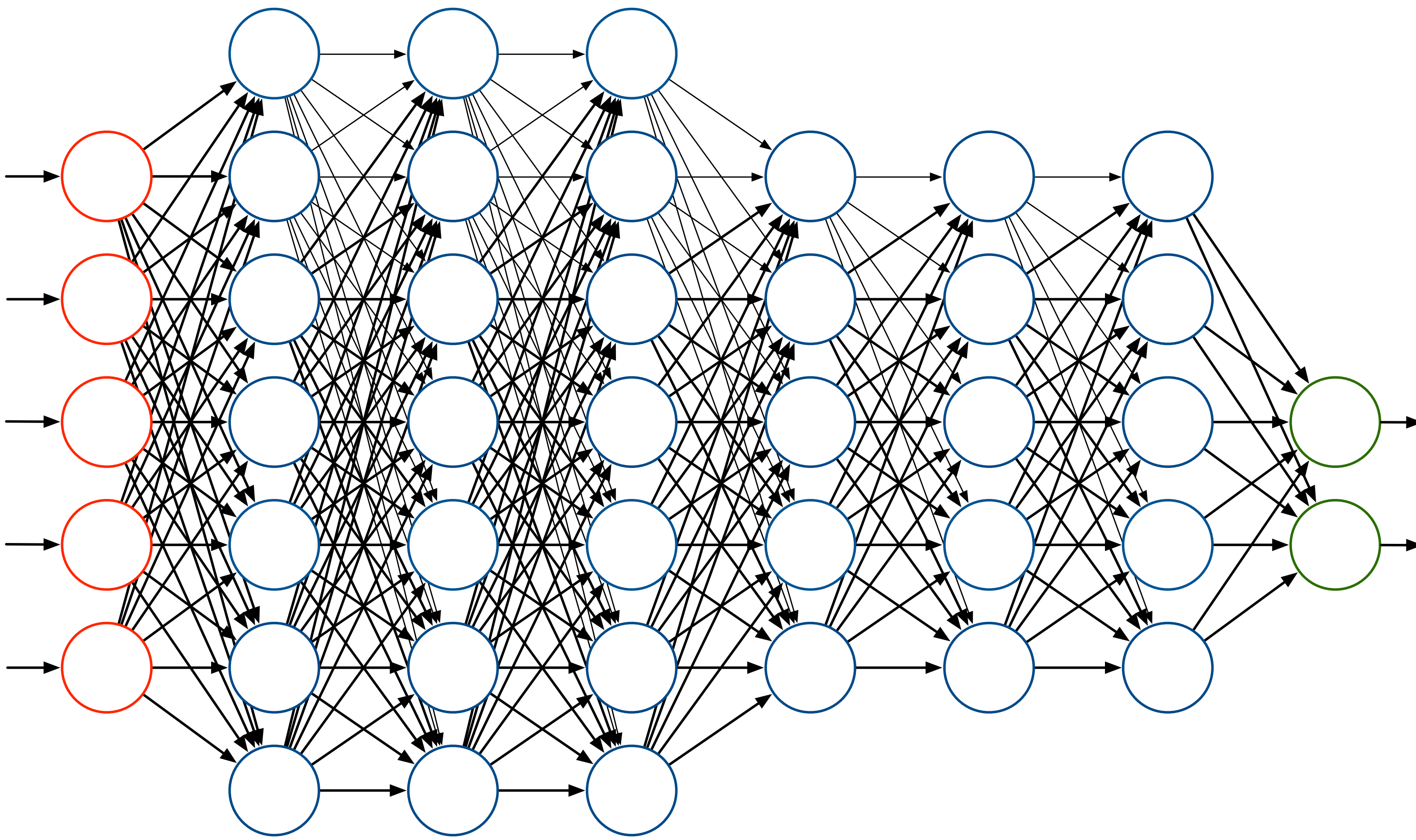
Neural Networks



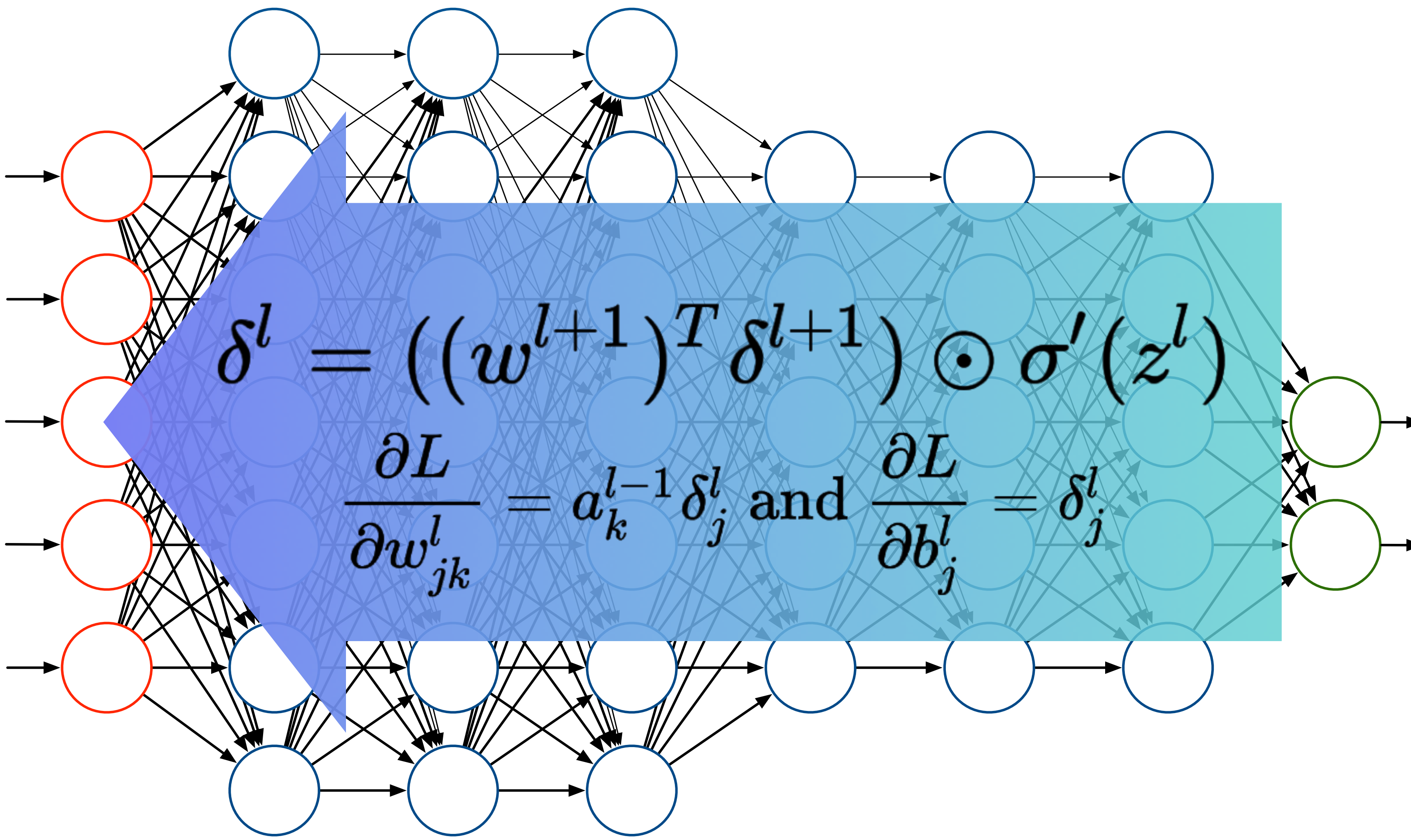


The **universal approximation theorem** states that, under reasonable assumptions, a feedforward **neural network** with a finite number of nodes **can approximate any continuous** function to within a given error over a bounded input domain.

Deep Learning

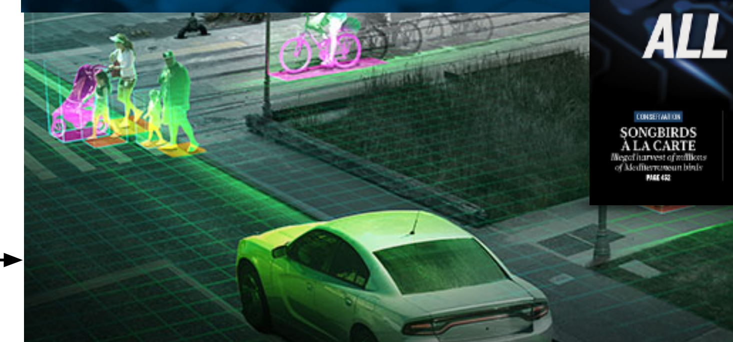


Deep Learning



$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

$$\frac{\partial L}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j \text{ and } \frac{\partial L}{\partial b^l_j} = \delta^l_j$$



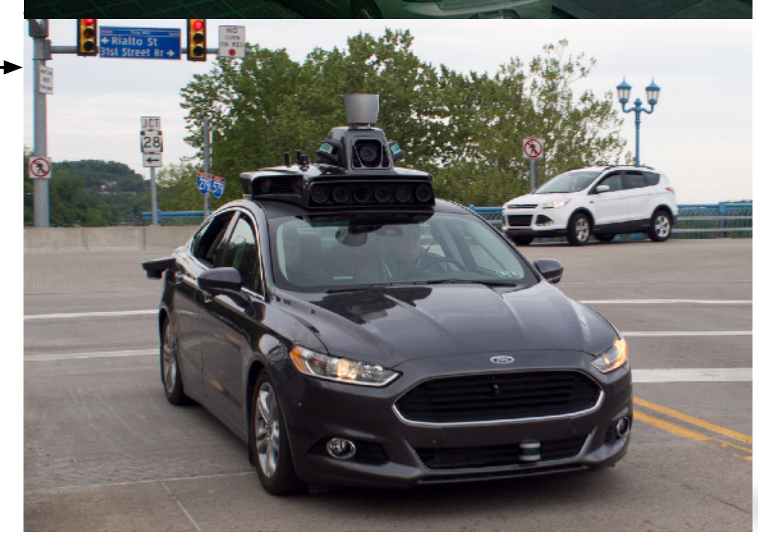
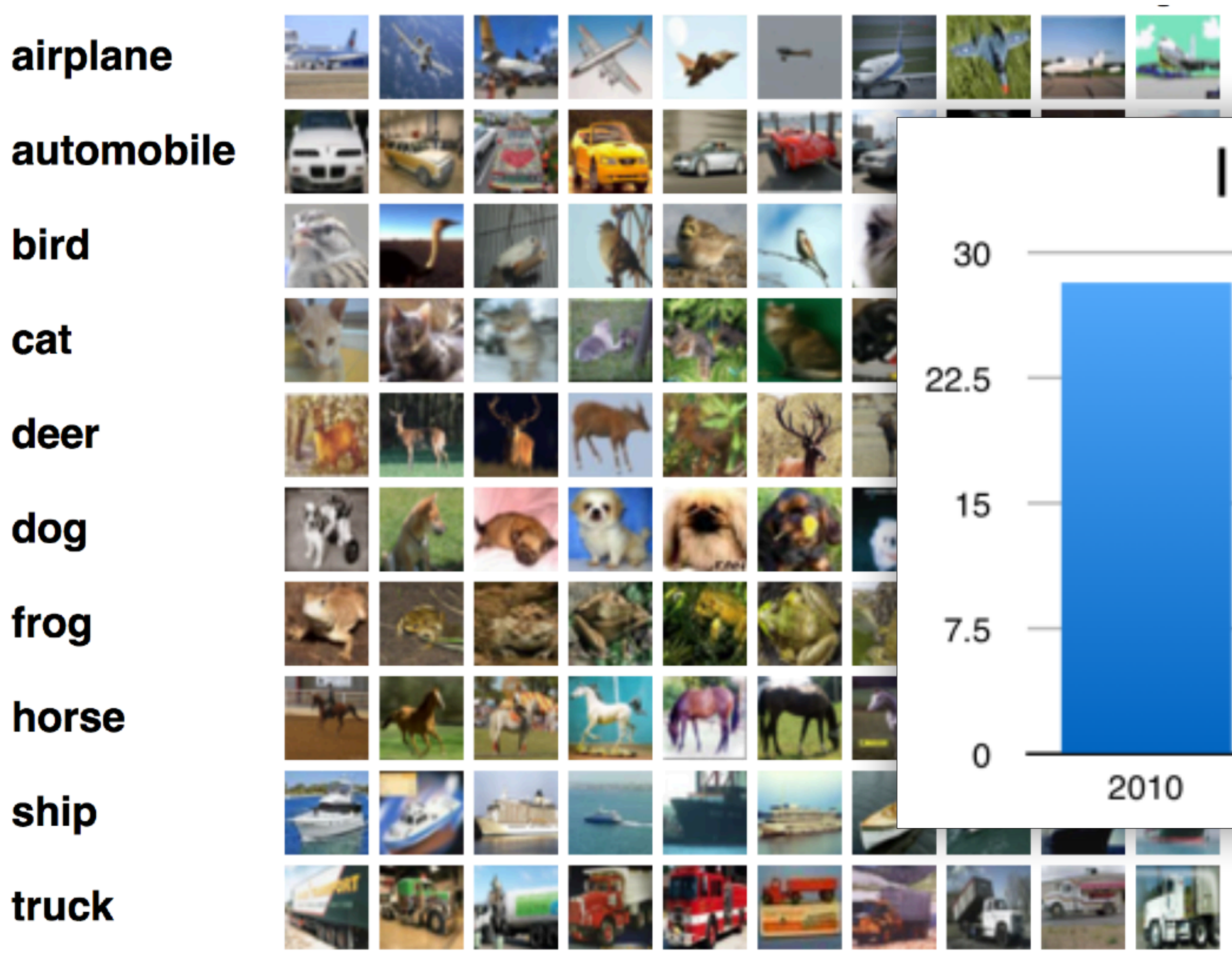
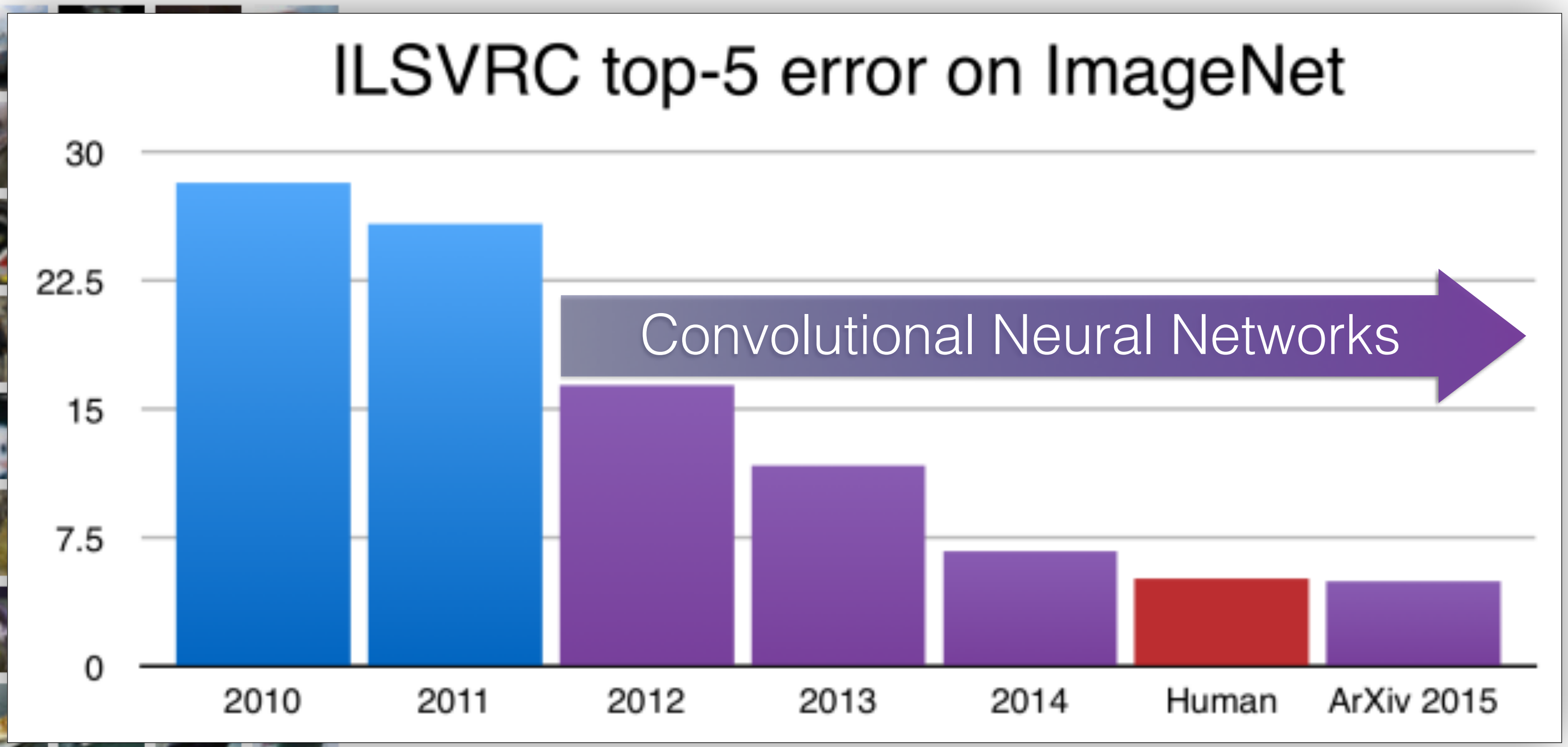


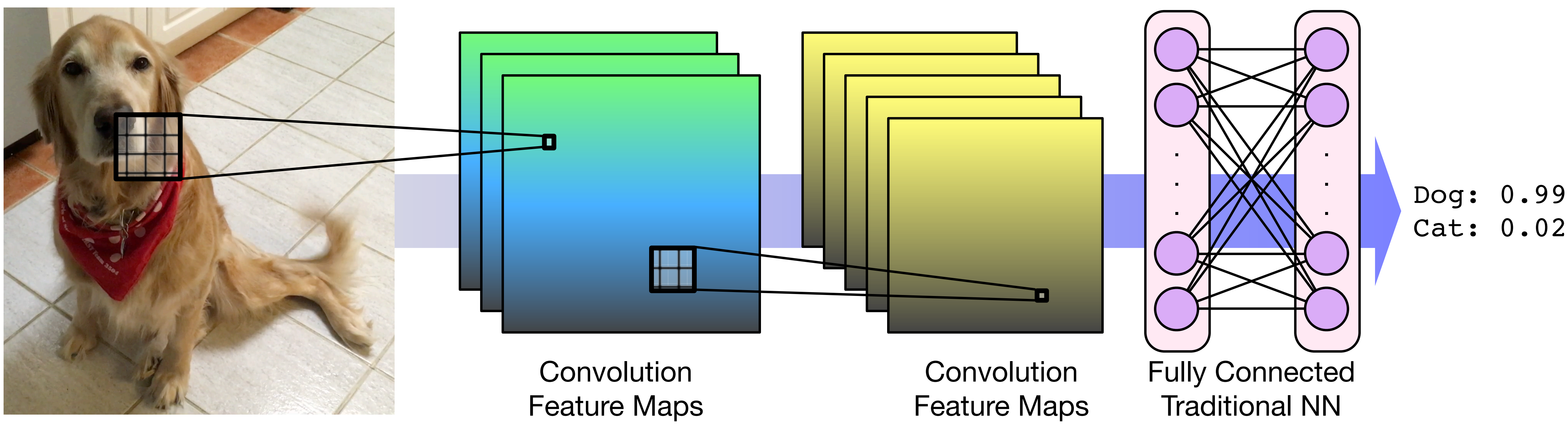
Image Recognition

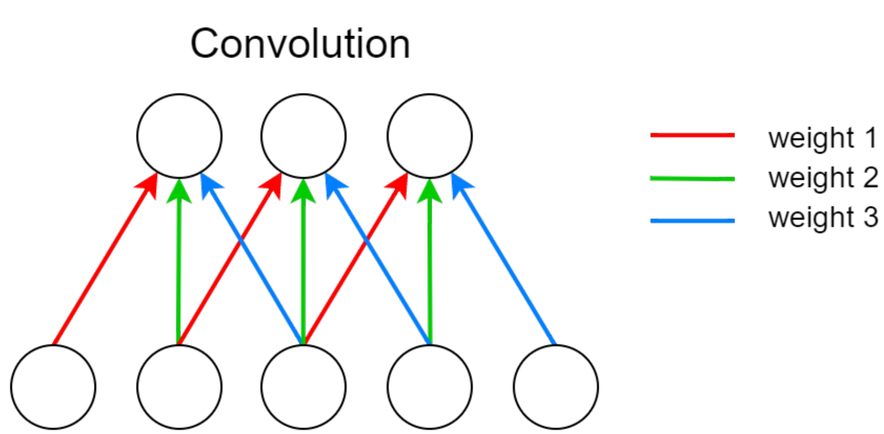


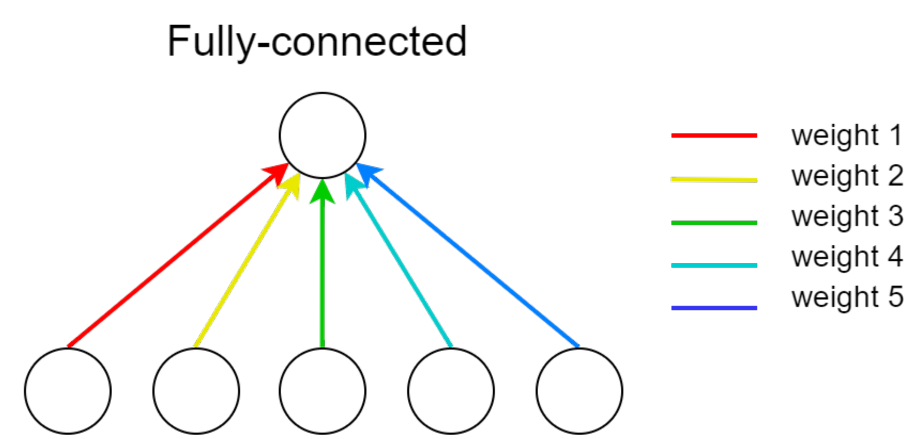


https://devblogs.nvidia.com

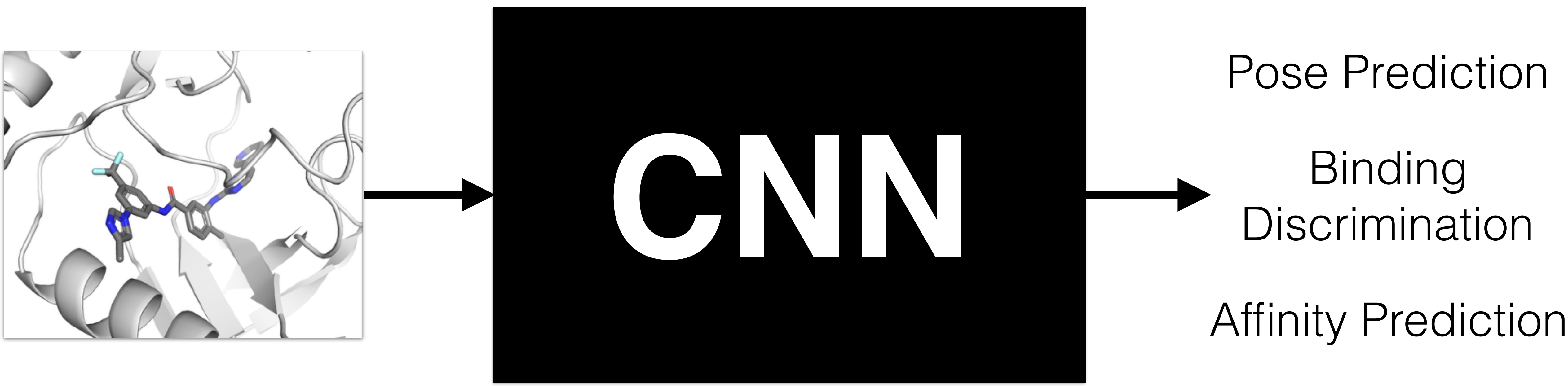
Convolutional Neural Networks



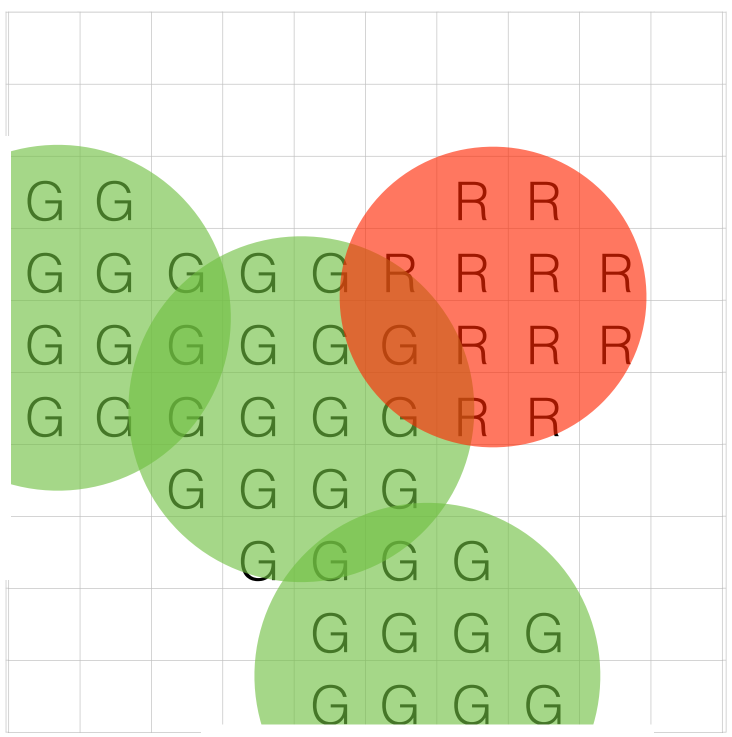




CNNs for Protein-Ligand Scoring

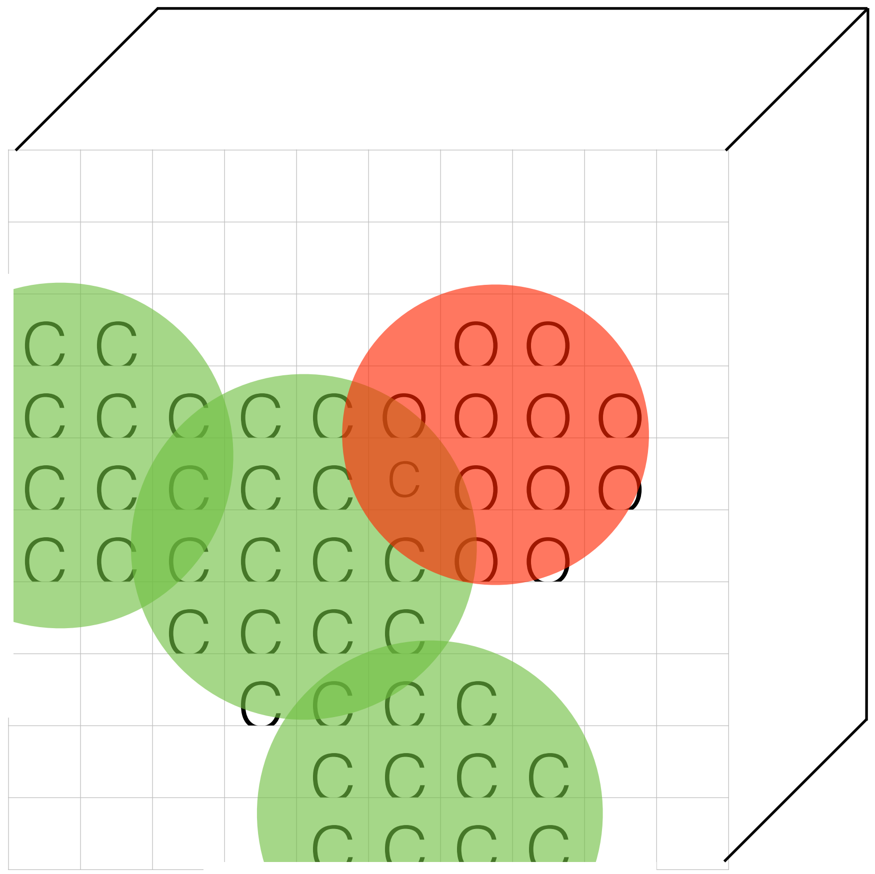


Protein-Ligand Representation



(R,G,B) pixel

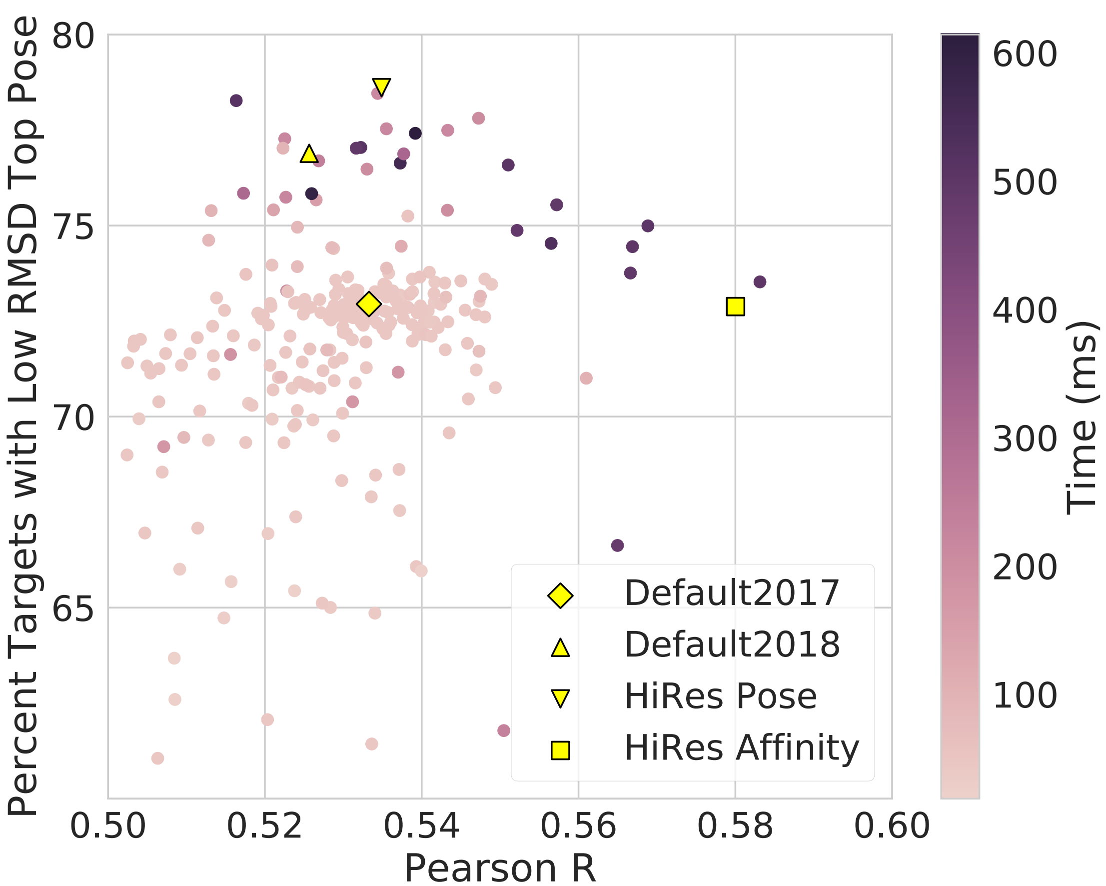
Protein-Ligand Representation

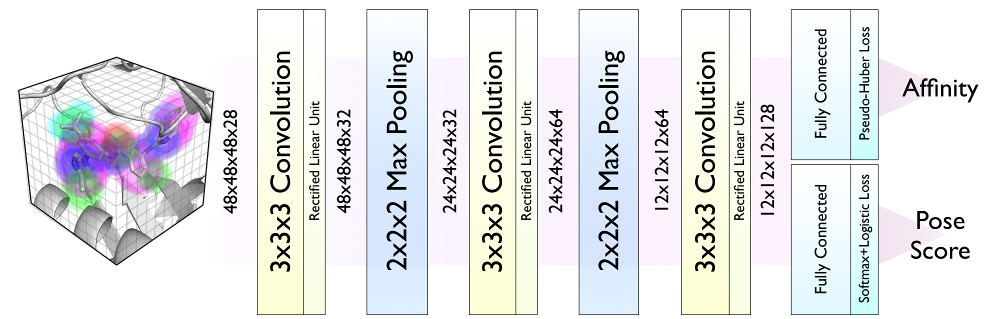


(R,G,B) pixel → (Carbon, Nitrogen, Oxygen,…) **voxe**

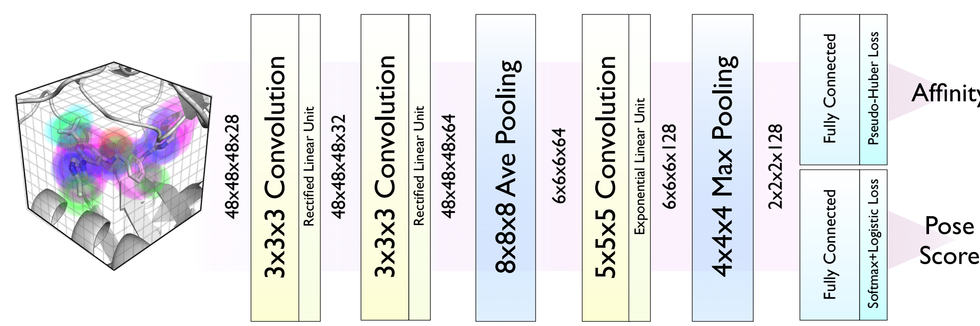
The only parameters for this representation are the choice of **grid resolution**, **atom density**, and **atom types**.

Optimized Models



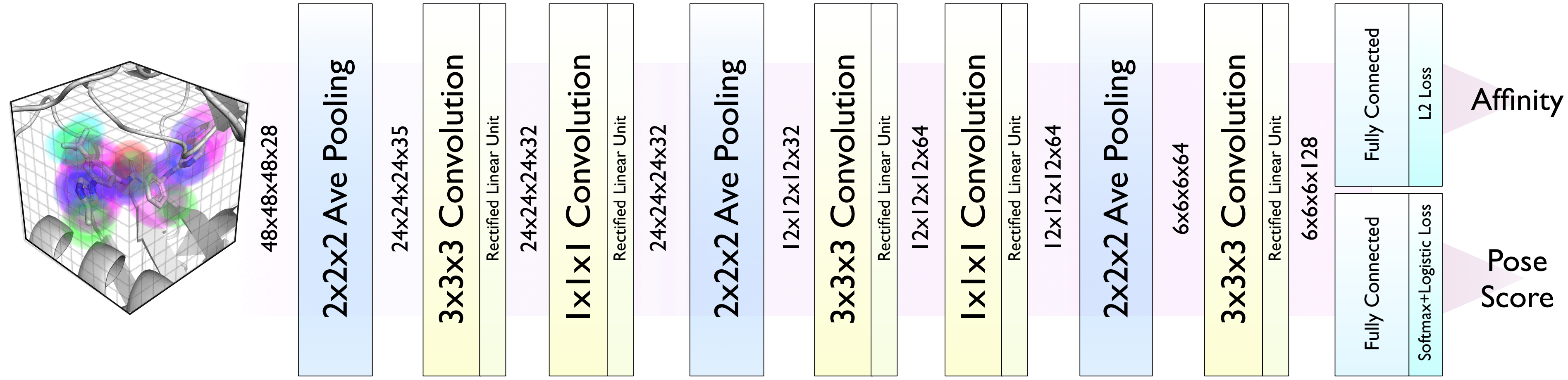


HiRes Pose

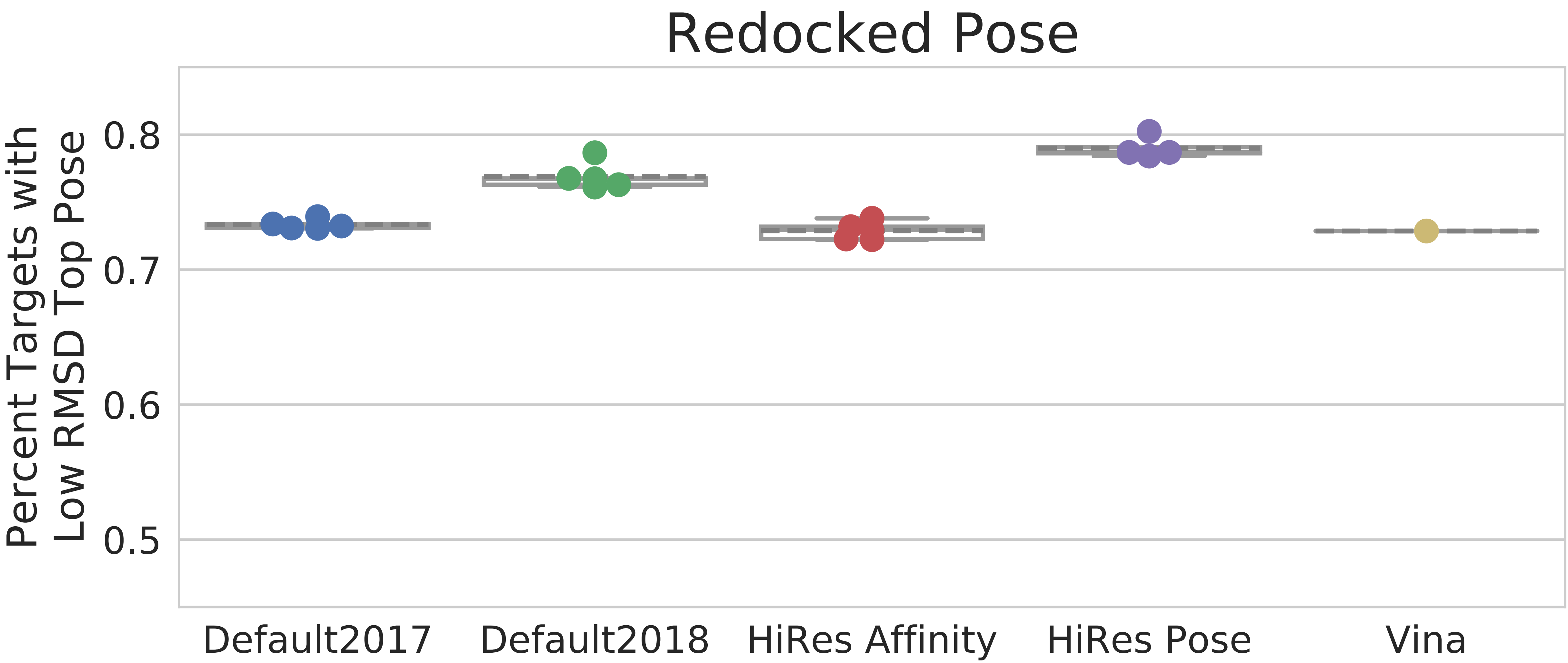


HiRes Affinity

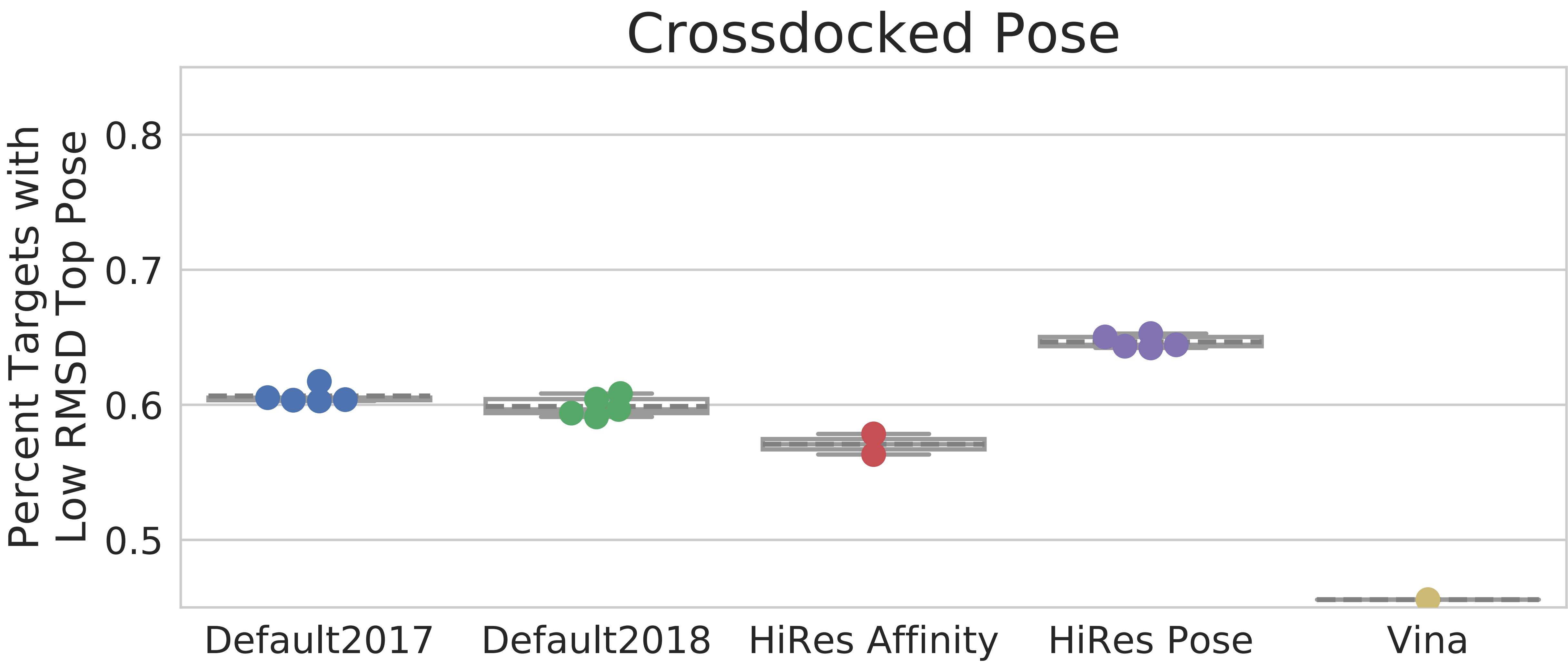
Default2018



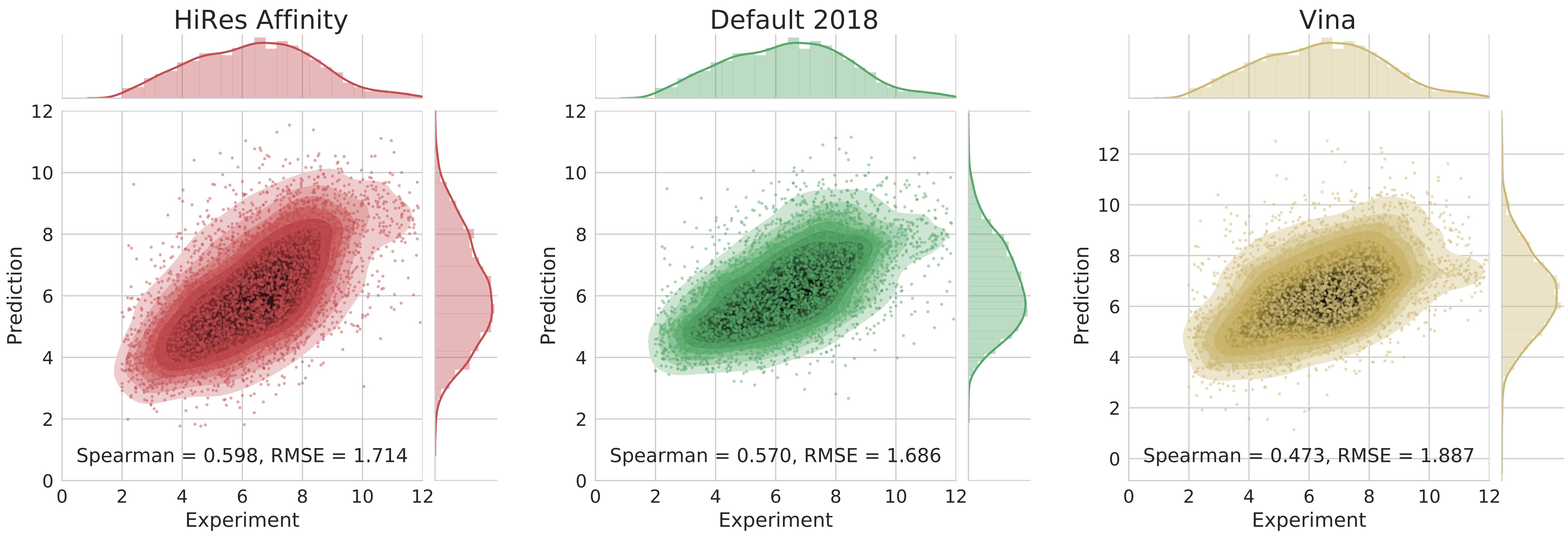
Pose Results



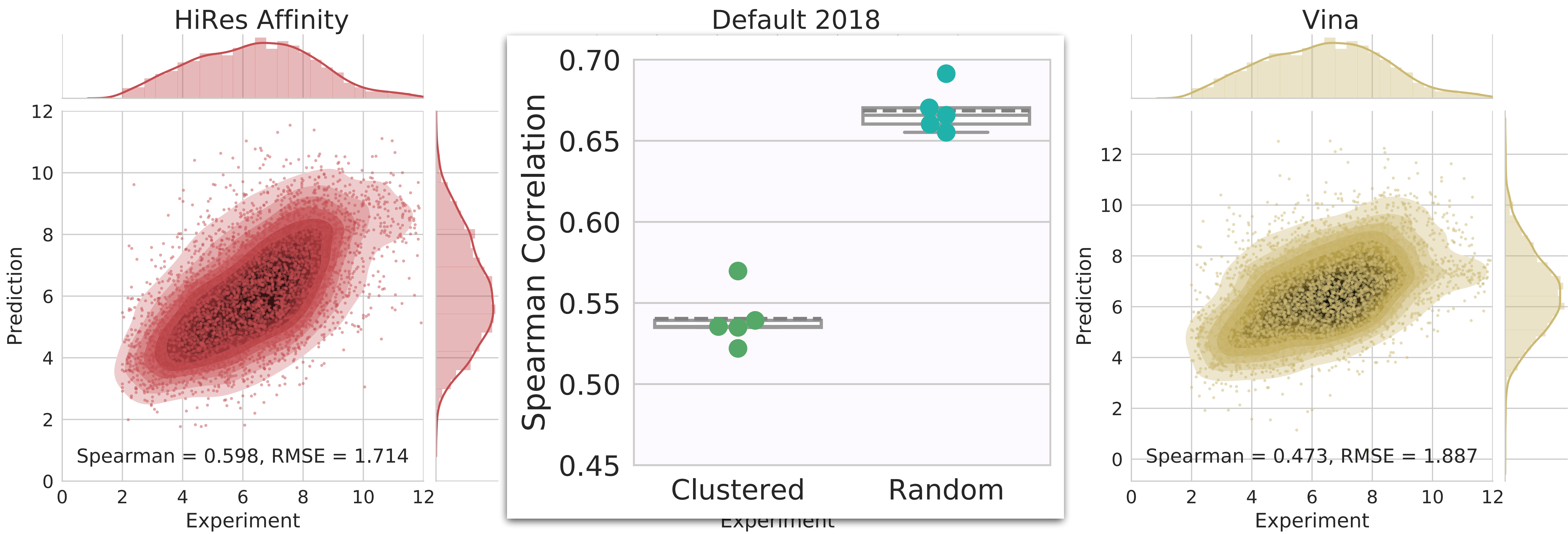
Pose Results



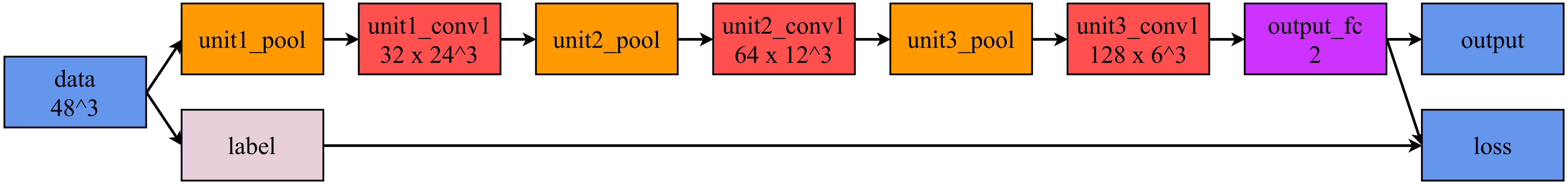
Affinity Results



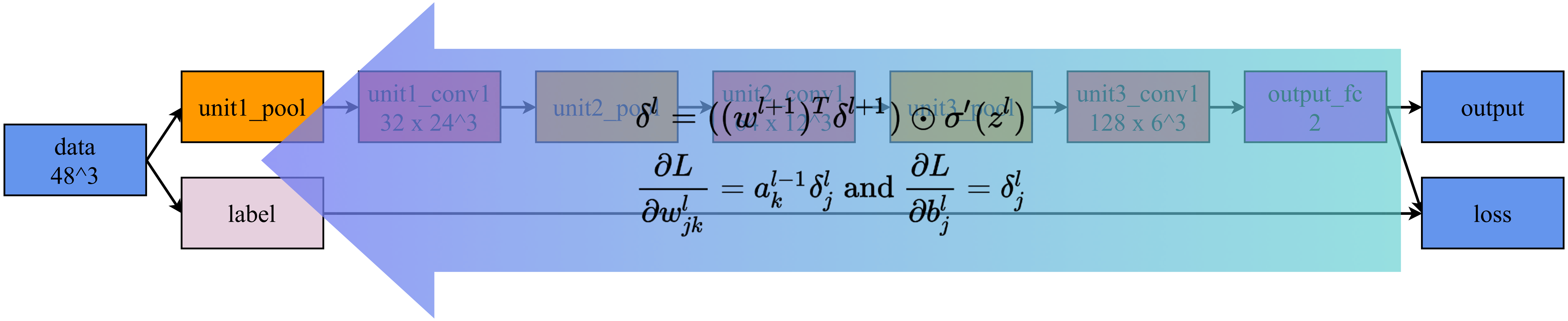
Affinity Results



Beyond Scoring



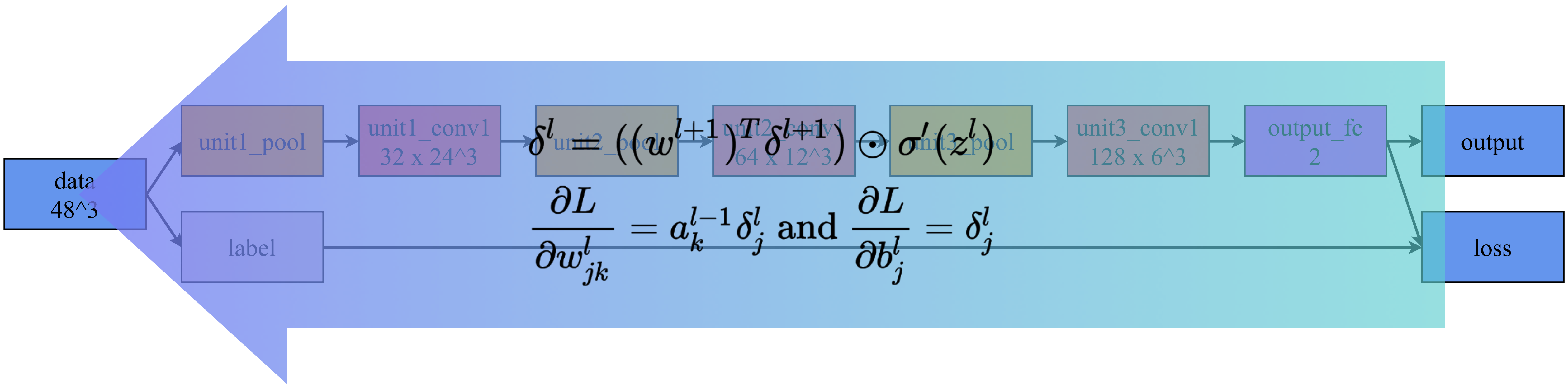
Beyond Scoring



$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

$$\frac{\partial L}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j \text{ and } \frac{\partial L}{\partial b^l_j} = \delta^l_j$$

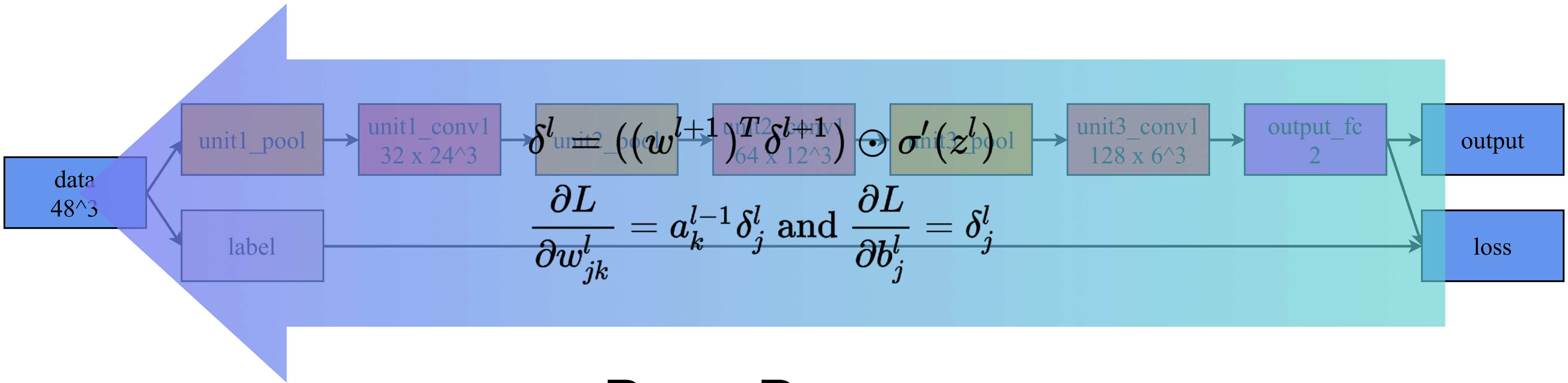
Beyond Scoring



$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

$$\frac{\partial L}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j \text{ and } \frac{\partial L}{\partial b^l_j} = \delta^l_j$$

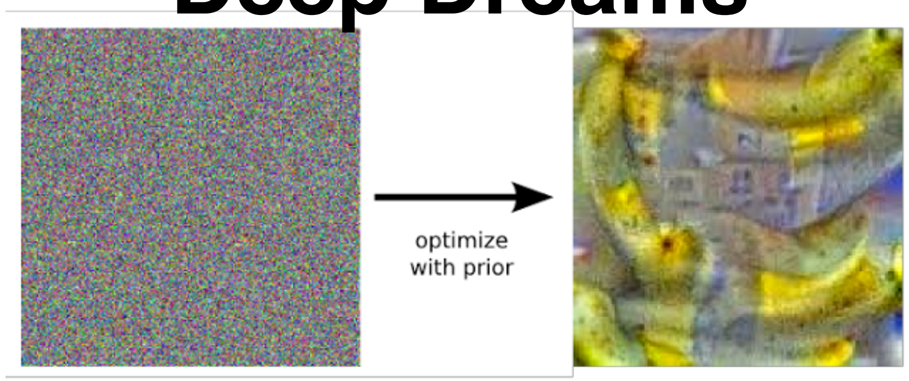
Beyond Scoring



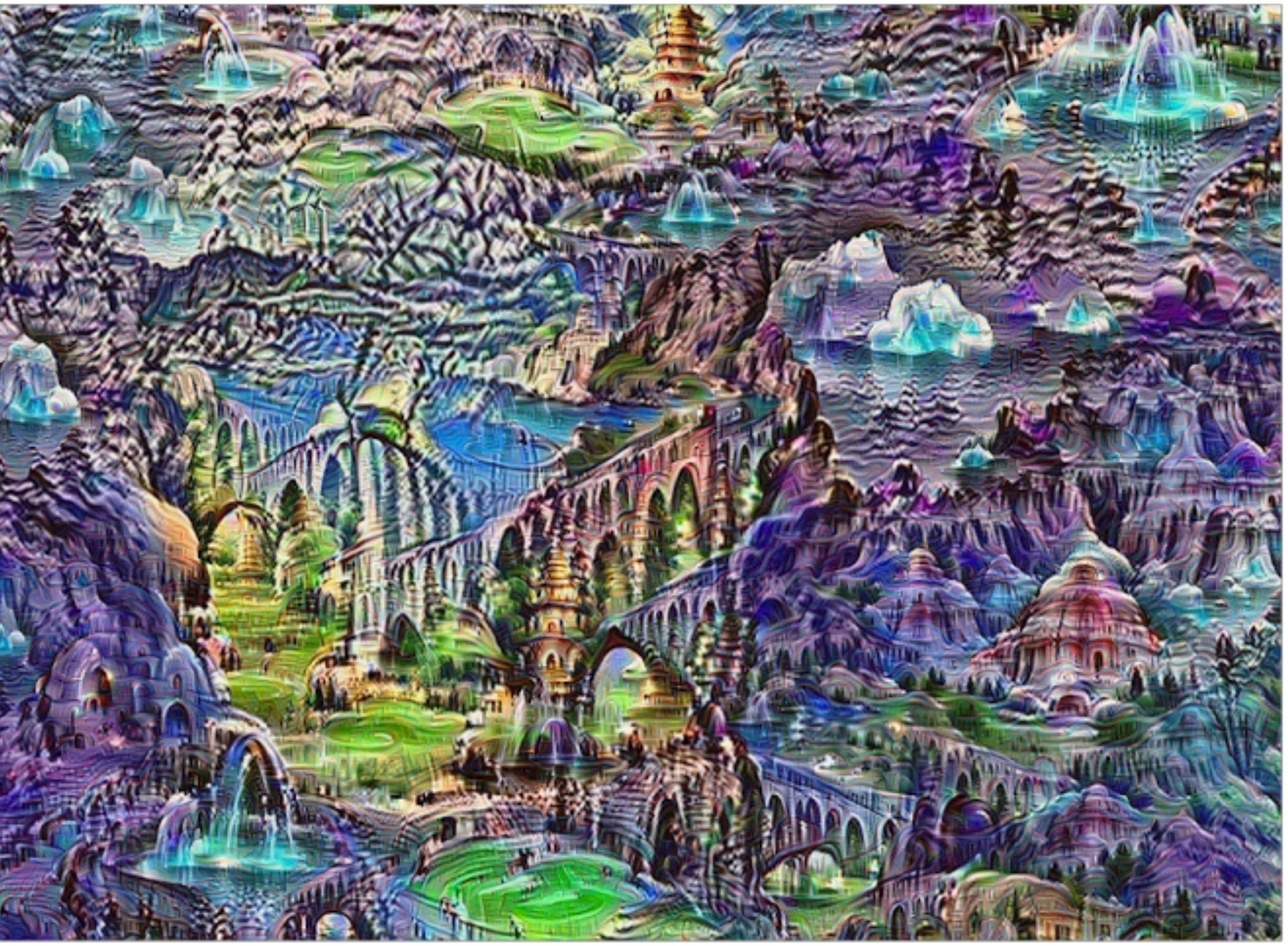
$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

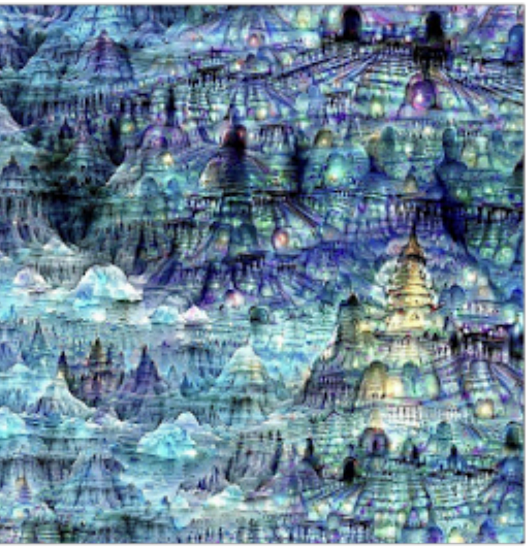
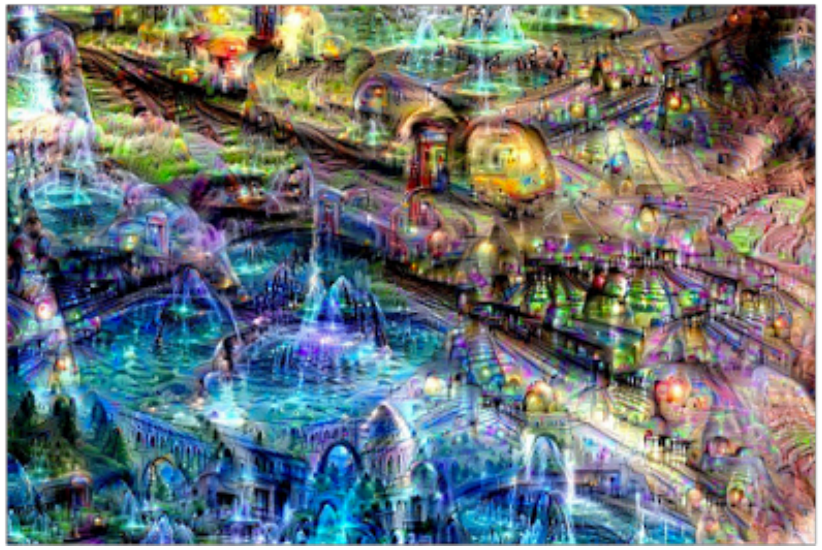
$$\frac{\partial L}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j \text{ and } \frac{\partial L}{\partial b^l_j} = \delta^l_j$$

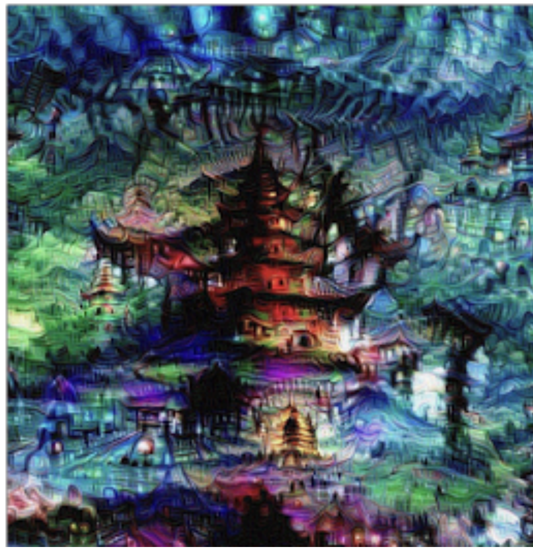
Deep Dreams



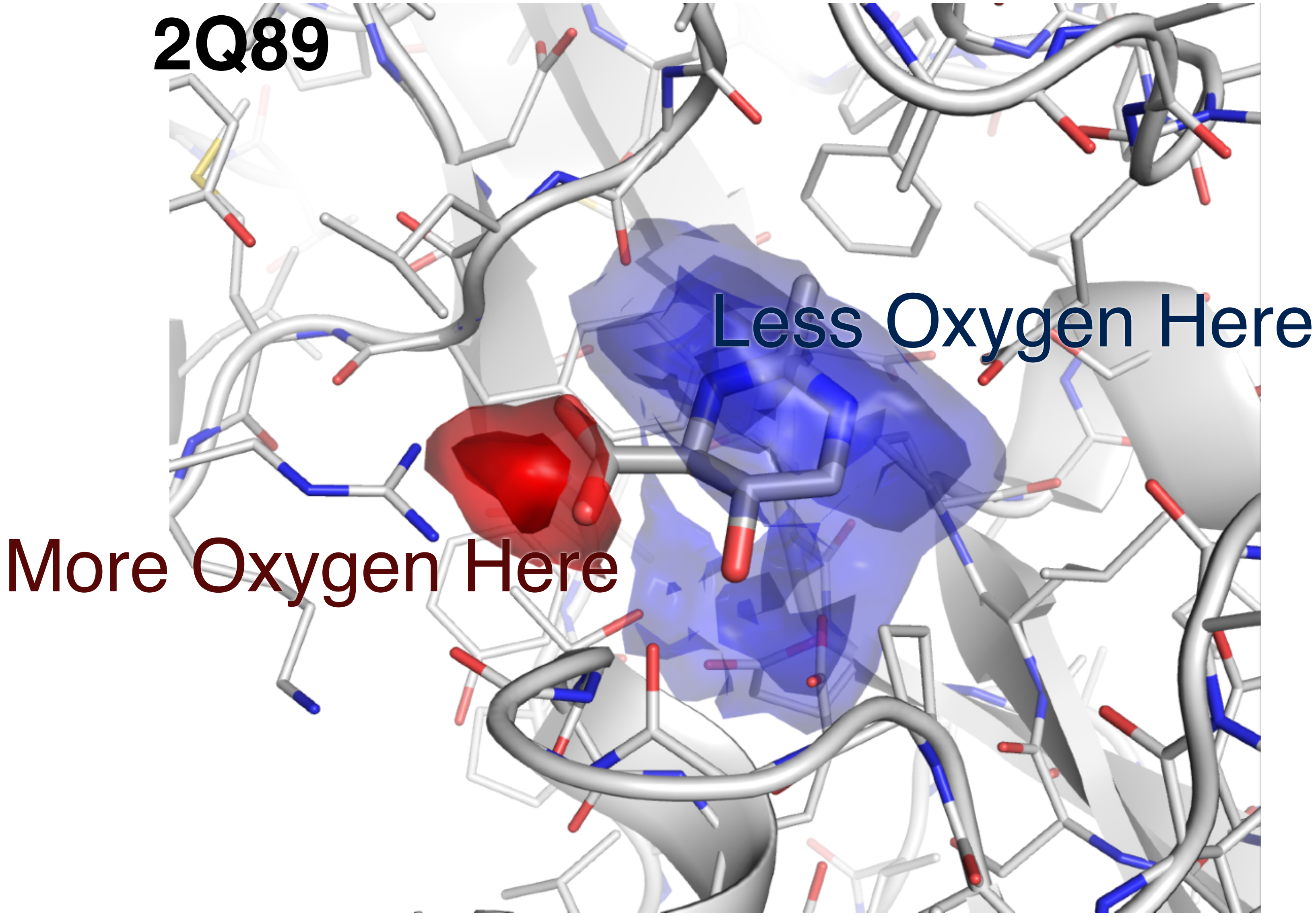
https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html



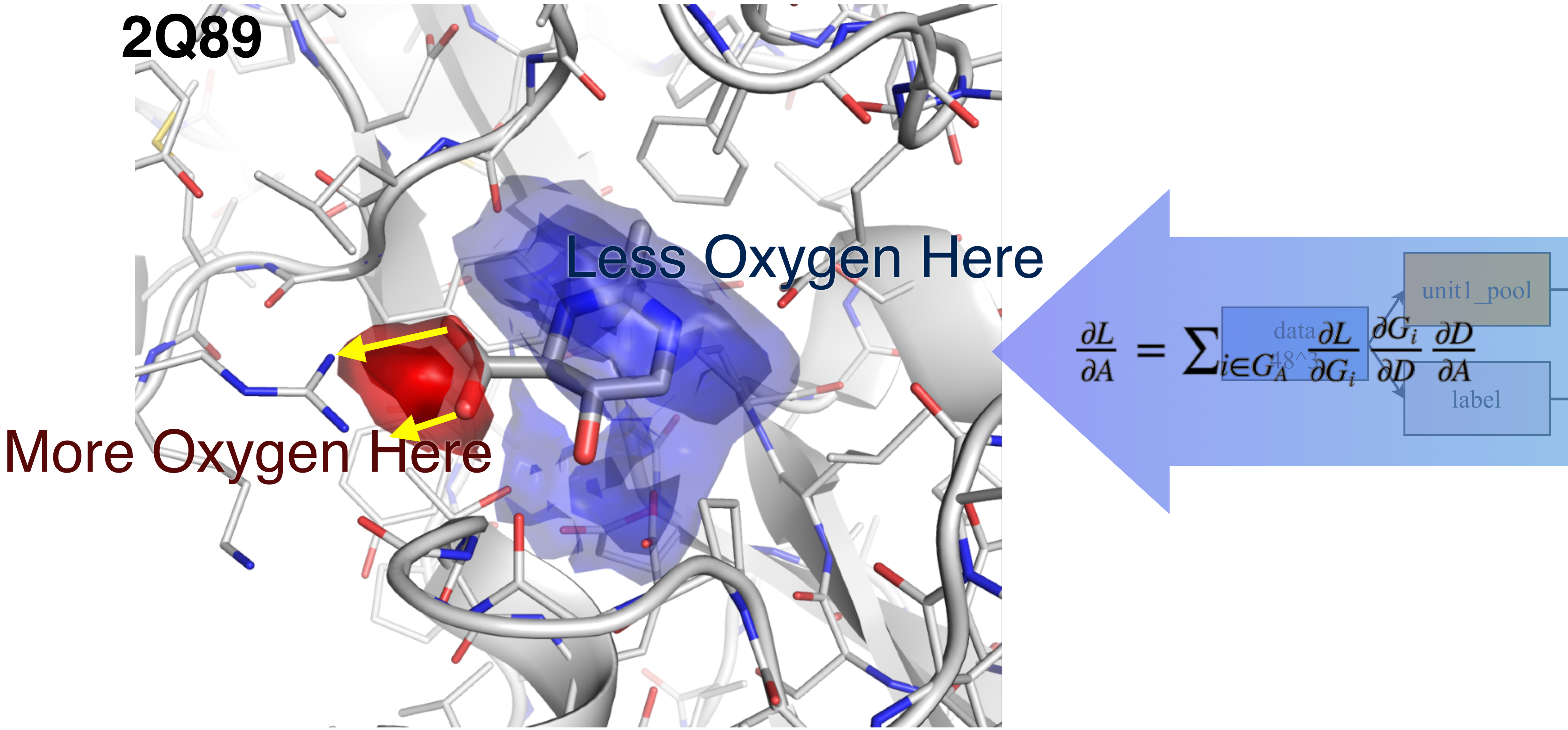



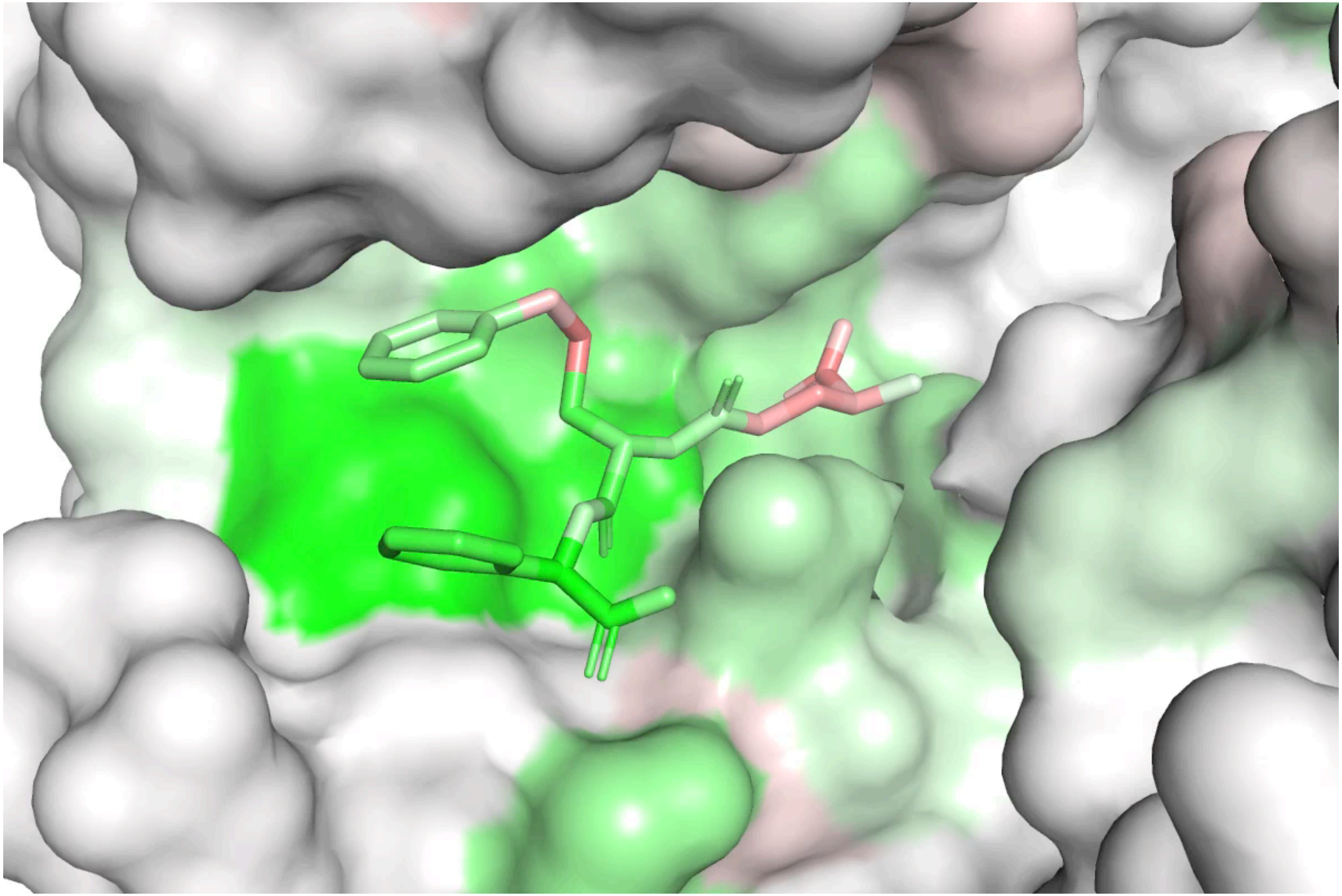


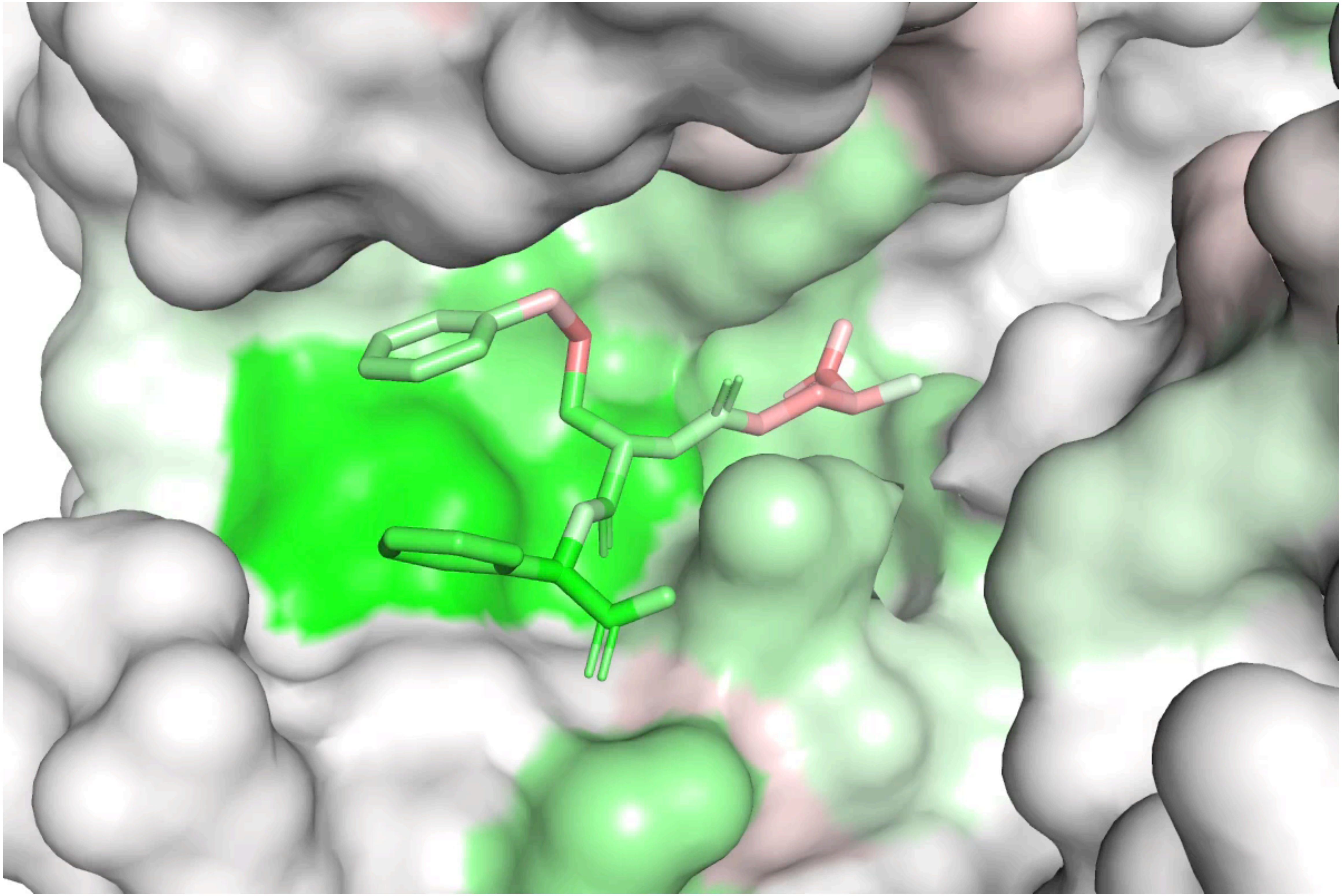
Beyond Scoring

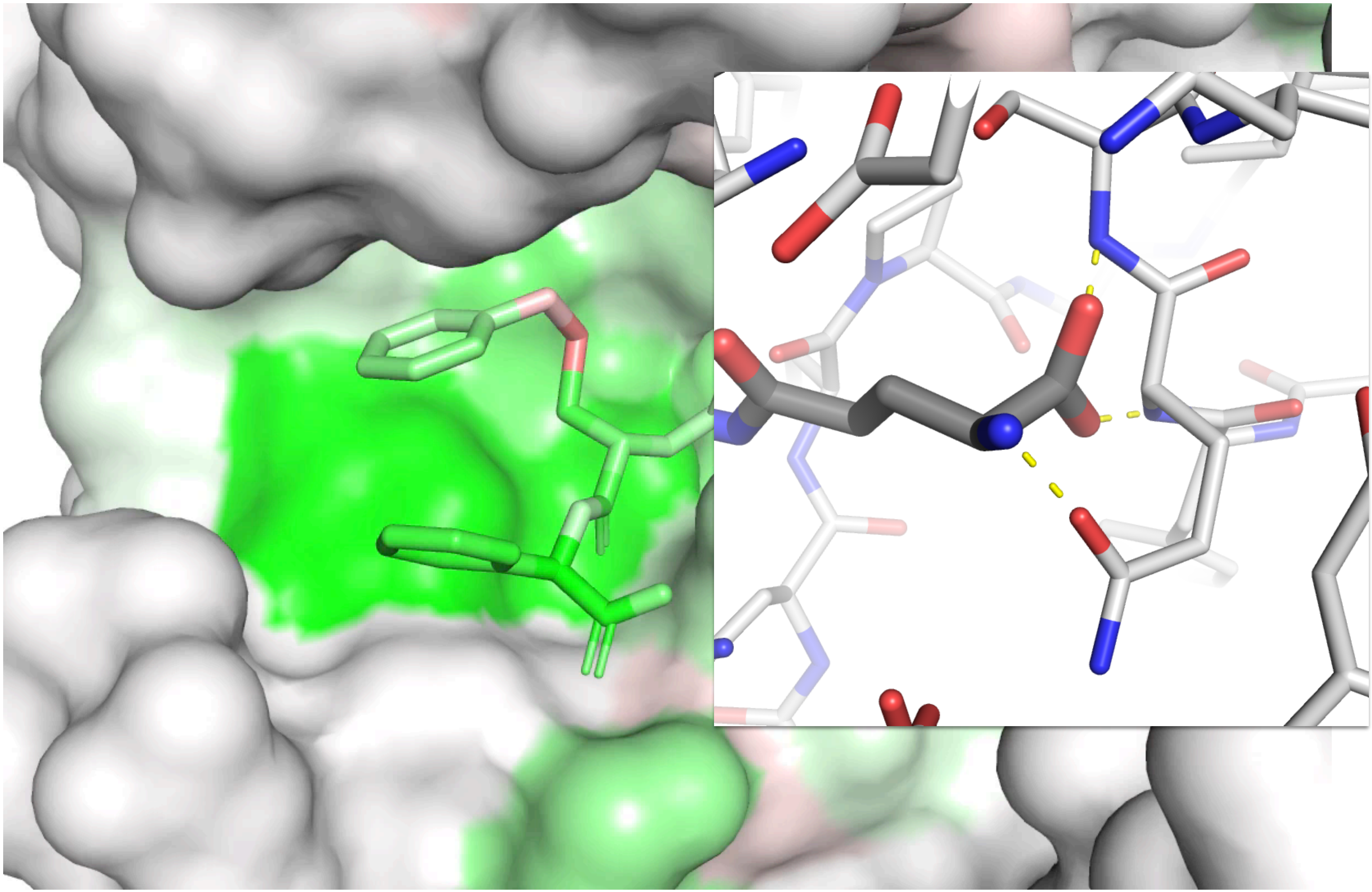


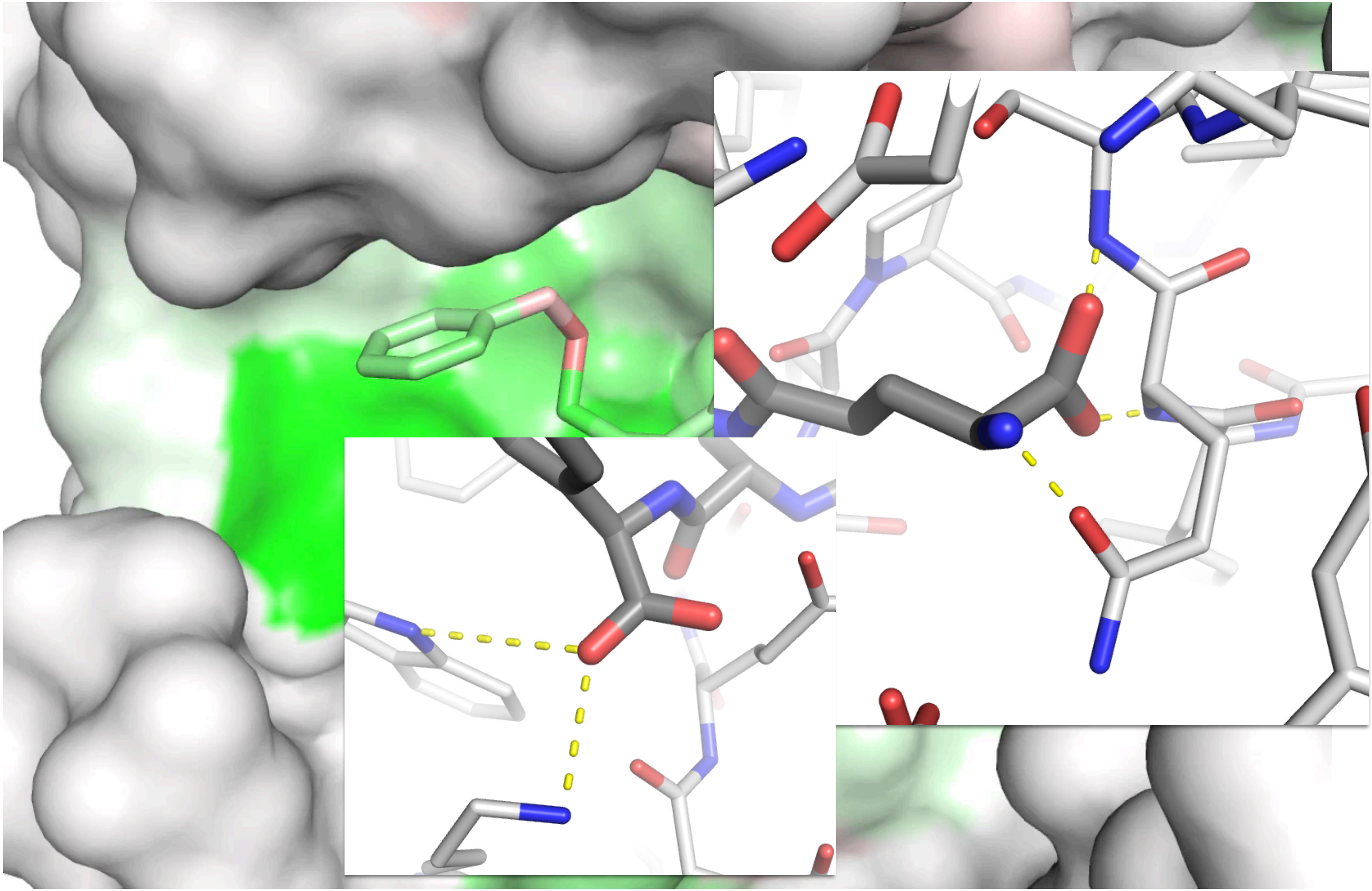
Beyond Scoring



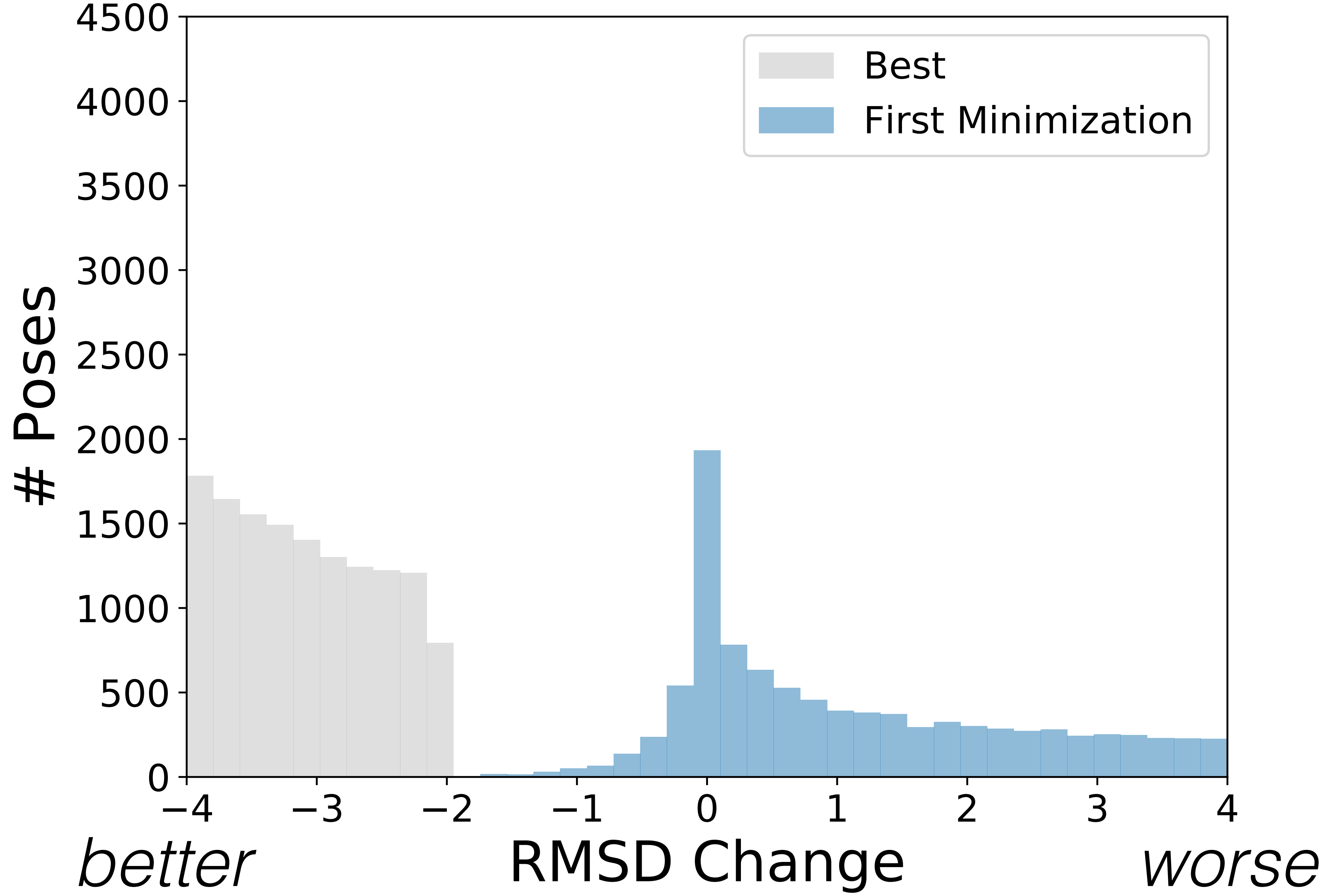


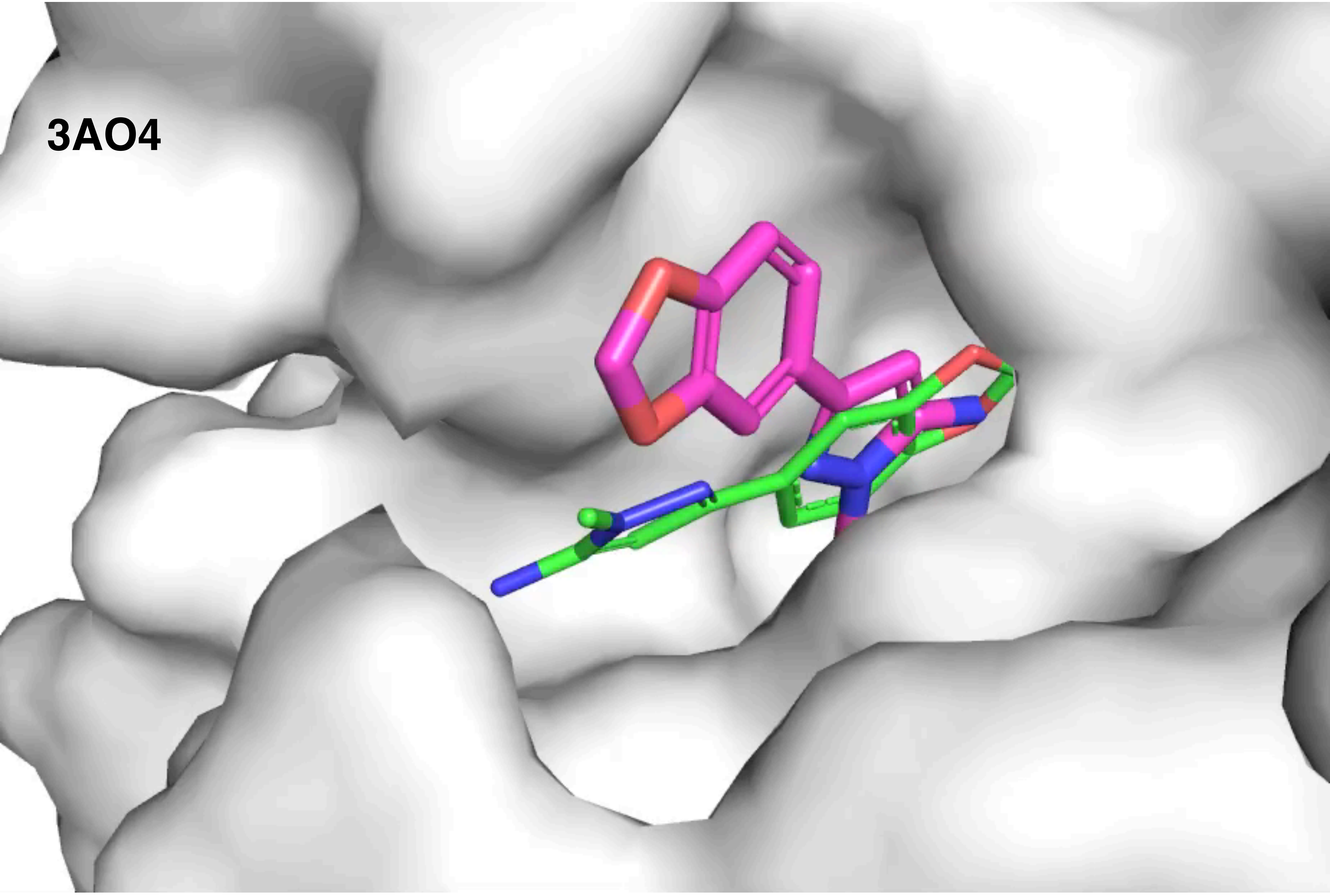


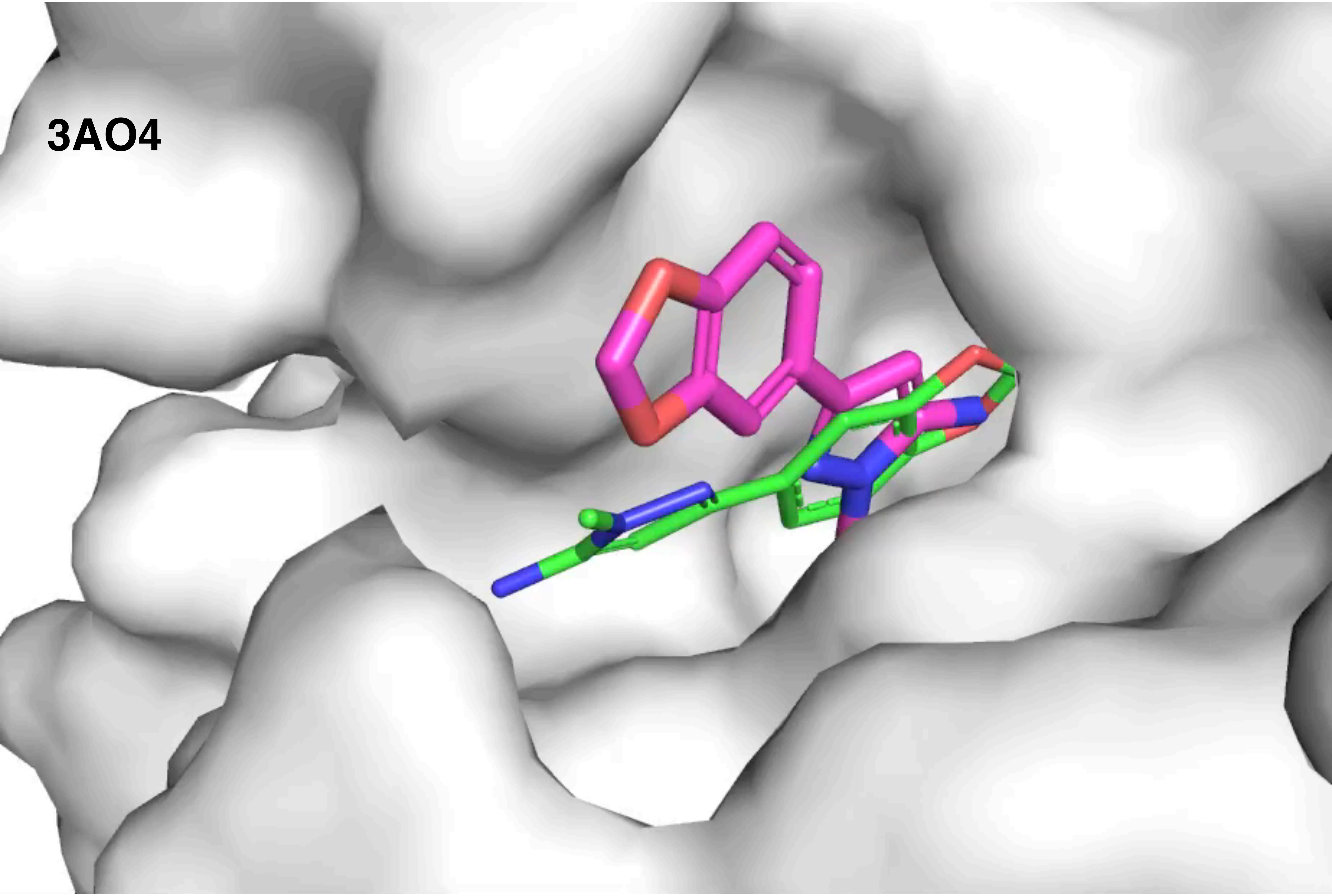




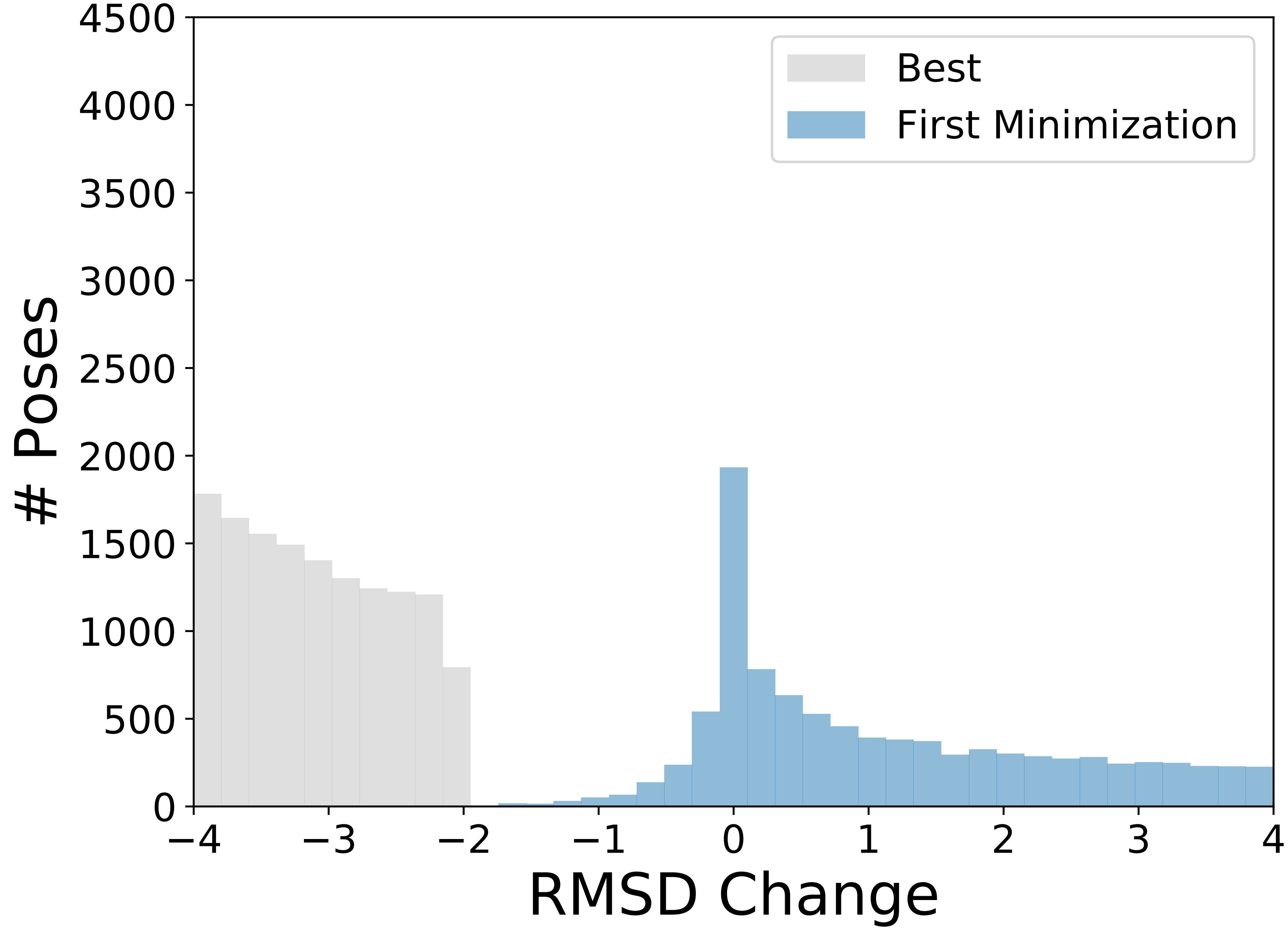
Minimizing Low RMSD Poses



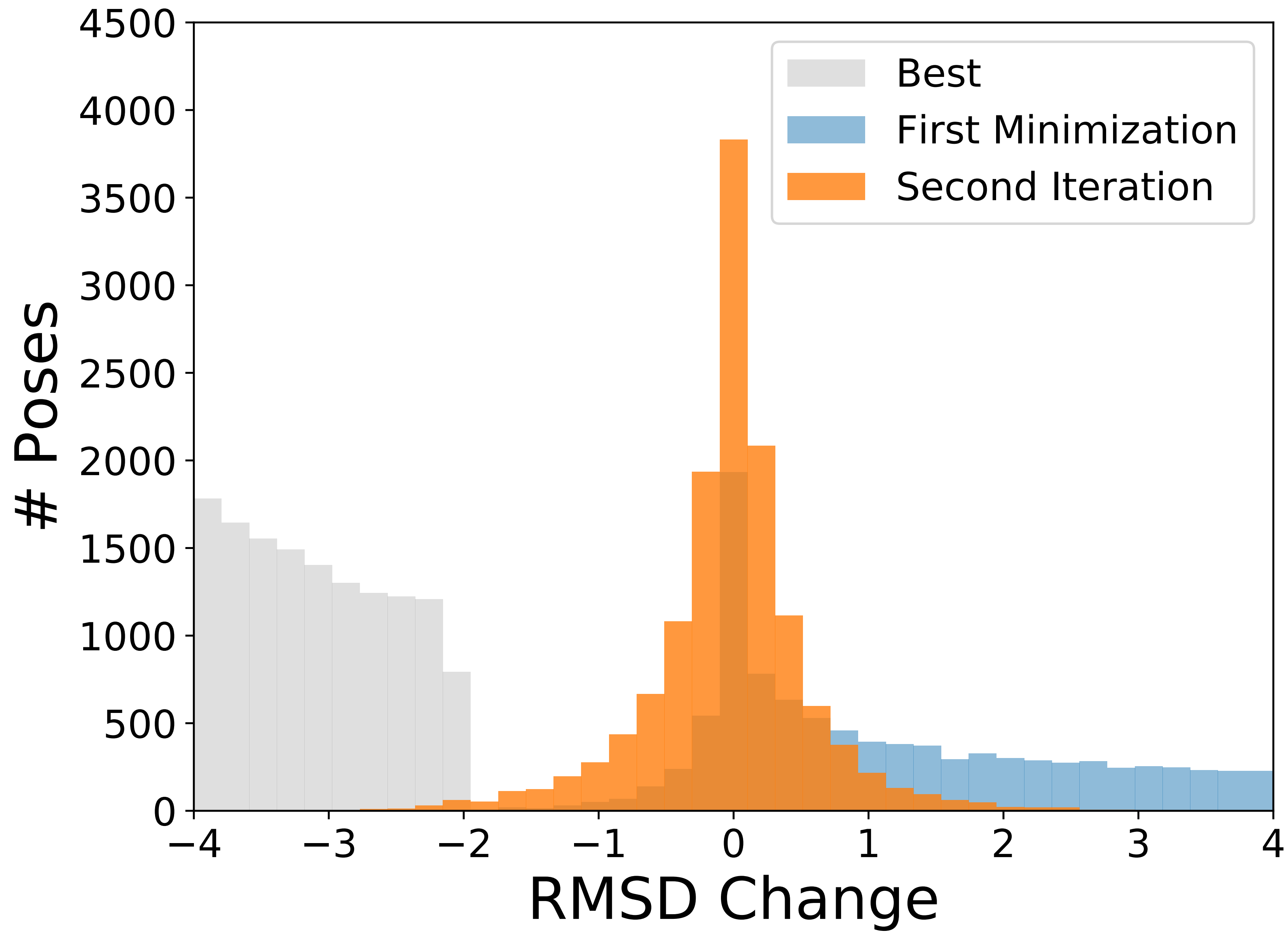

3AO4


3AO4

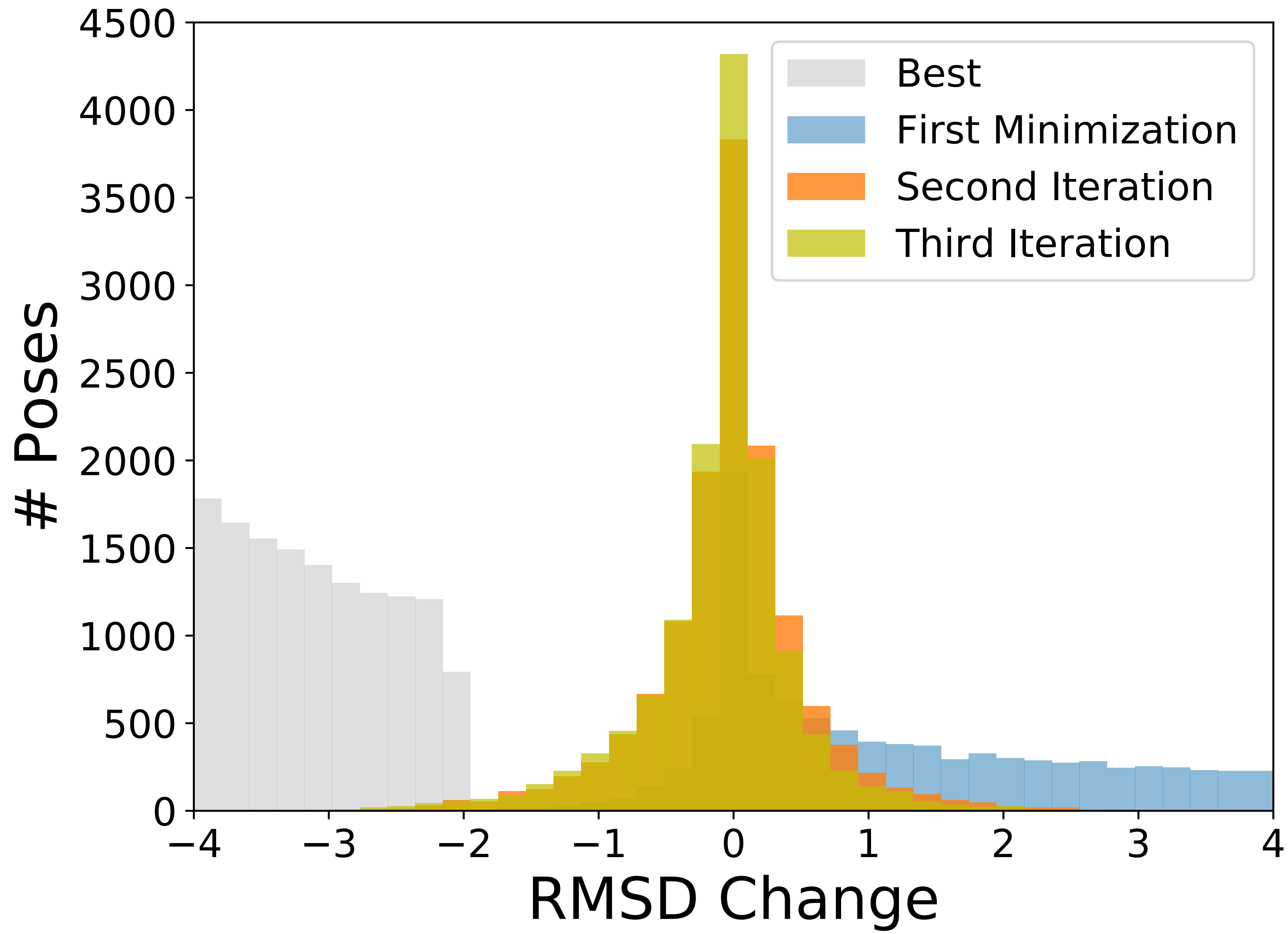
Iterative Refinement

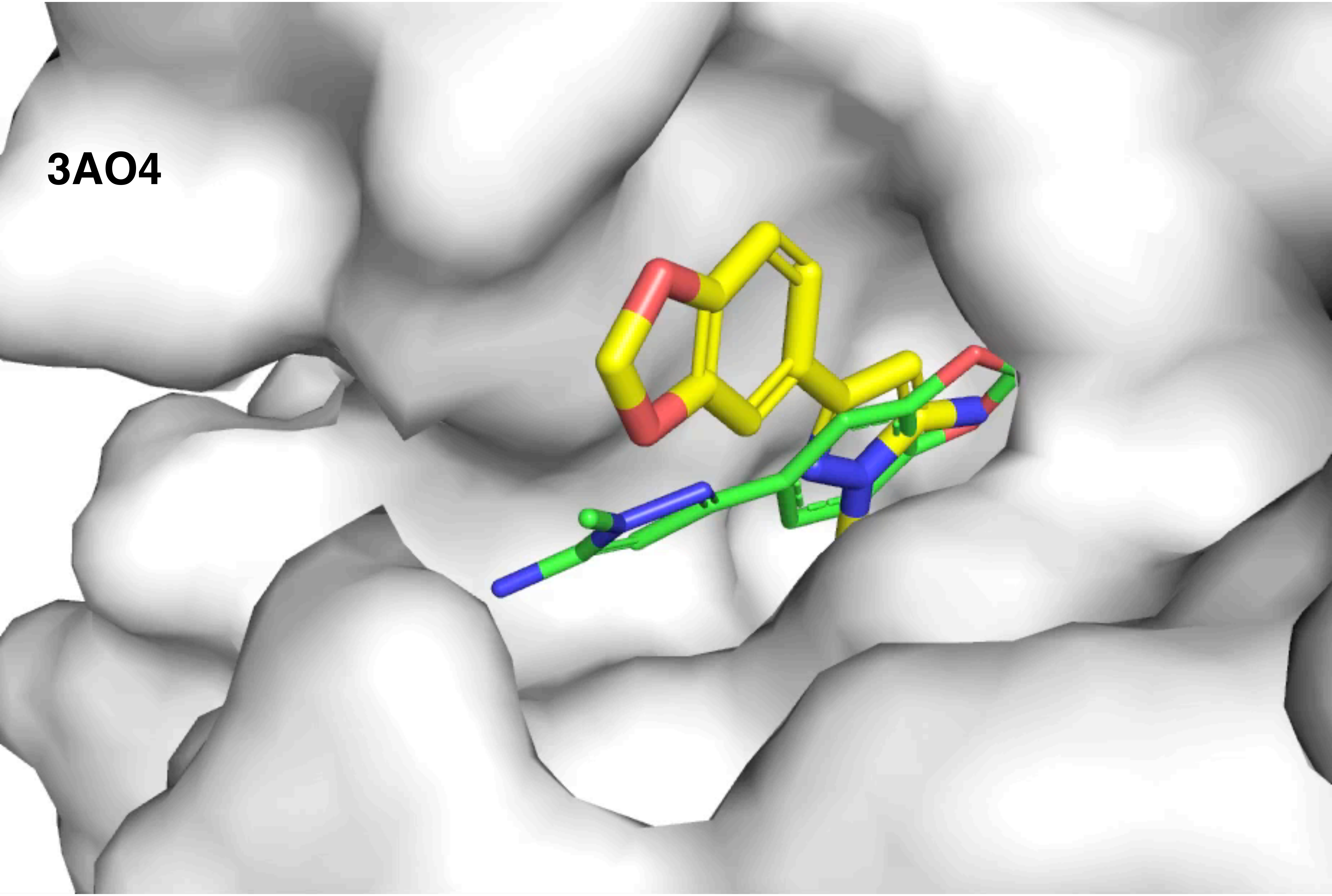


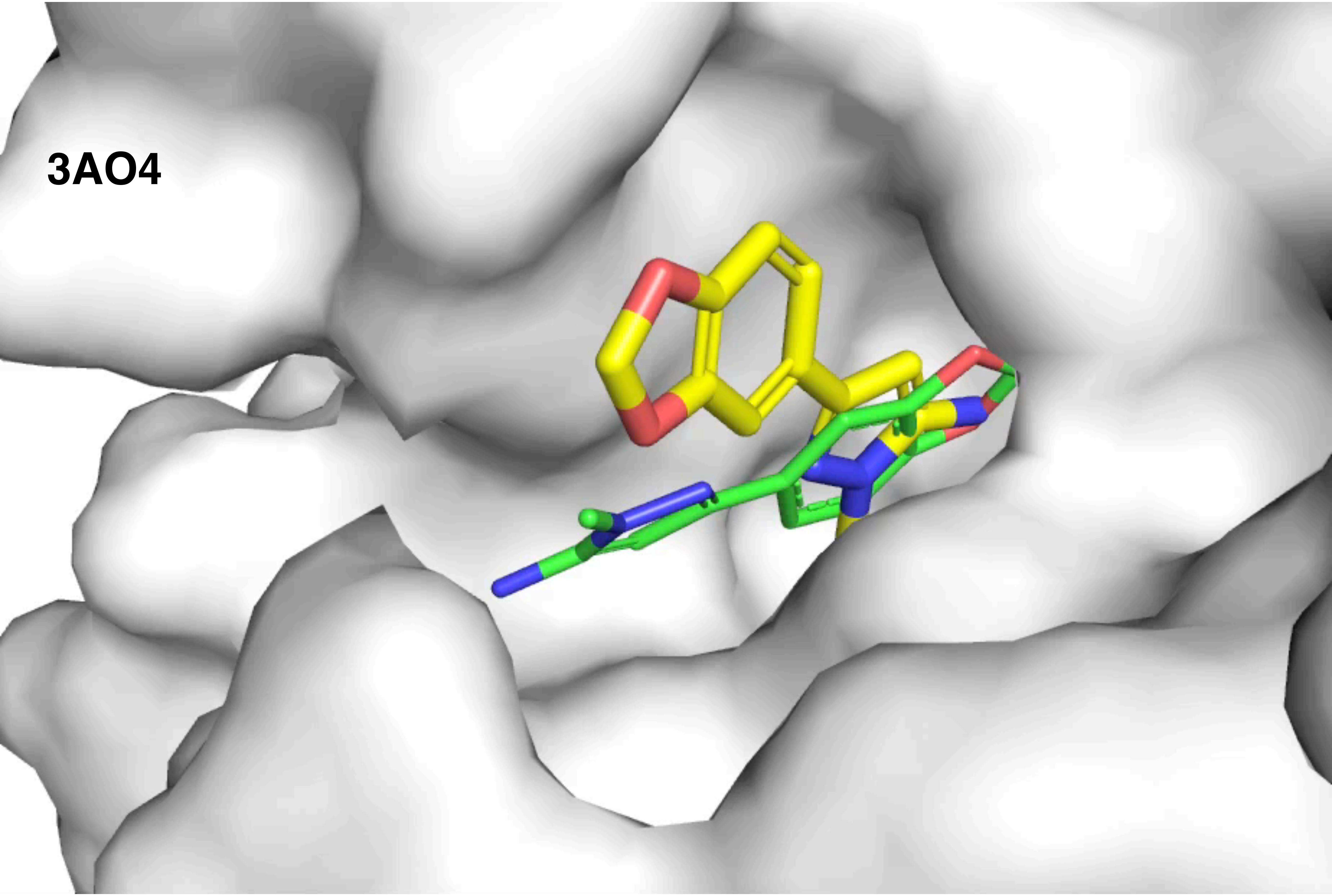
Iterative Refinement



Iterative Refinement




3AO4


3AO4

Related Work

MolecuLeNet: A continuous-filter convolutional neural network for modeling quantum interactions

Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, Klaus-Robert Müller

(Submitted on 26 Jun 2017)

Automatic chemical design using a data-driven continuous representation of molecules

Rafael Gómez-Bombarelli, David Duvenaud, José Miguel Hernández-Lobato, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, Alán Aspuru-Guzik

(Submitted on 7 Oct 2016 (v1), last revised 6 Jan 2017 (this version, v2))

AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery

Izhar Wallach, Michael Dzamba, Abraham Heifets

(Submitted on 10 Oct 2015)

ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost

Justin S. Smith, Olexandr Isayev, Adrian E. Roitberg

(Submitted on 27 Oct 2016 (v1), last revised 6 Feb 2017 (this version, v4))

Convolutional Networks on Graphs for Learning Molecular Fingerprints

David Duvenaud, Dougal Maclaurin, Jorge Aguilera-Iparraguirre, Rafael Gómez-Bombarelli, Timothy Hirzel, Alán Aspuru-Guzik, Ryan P. Adams

(Submitted on 30 Sep 2015 (v1), last revised 3 Nov 2015 (this version, v2))

Atomic Convolutional Networks for Predicting Protein-Ligand Binding Affinity

Joseph Gomes, Bharath Ramsundar, Evan N. Feinberg, Vijay S. Pande

(Submitted on 30 Mar 2017)

Deep Architectures and Deep Learning in Chemoinformatics: The Prediction of Aqueous Solubility for Drug-Like Molecules

Alessandro Lusci*†, Gianluca Pollastri†, and Pierre Baldi*‡
† School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland
‡ Department of Computer Science, University of California, Irvine, Irvine, California 92697, United States

J. Chem. Inf. Model., **2013**, *53* (7), pp 1563–1575
DOI: 10.1021/ci400187y
Publication Date (Web): June 24, 2013

Low Data Drug Discovery with One-shot Learning

Han Altae-Tran, Bharath Ramsundar, Aneesh S. Pappu, Vijay Pande

(Submitted on 10 Nov 2016)

Massively Multitask Networks for Drug Discovery

Bharath Ramsundar, Steven Kearnes, Patrick Riley, Dale Webster, David Konerding, Vijay Pande

(Submitted on 6 Feb 2015)

Protein–Ligand Scoring with Convolutional Neural Networks

Matthew Ragoza†‡, Joshua Hochuli‡¶, Elisa Idrobo§, Jocelyn Sunseriǀ, and David Ryan Koes*ǀ
†Department of Neuroscience, ‡Department of Computer Science, ¶Department of Biological Sciences, and ǀDepartment of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States
§ Department of Computer Science, The College of New Jersey, Ewing, New Jersey 08628, United States

J. Chem. Inf. Model., **2017**, *57* (4), pp 942–957
DOI: 10.1021/acs.jcim.6b00740
Publication Date (Web): April 3, 2017

@david_koes

github.com/gnina

http://bits.csb.pitt.edu