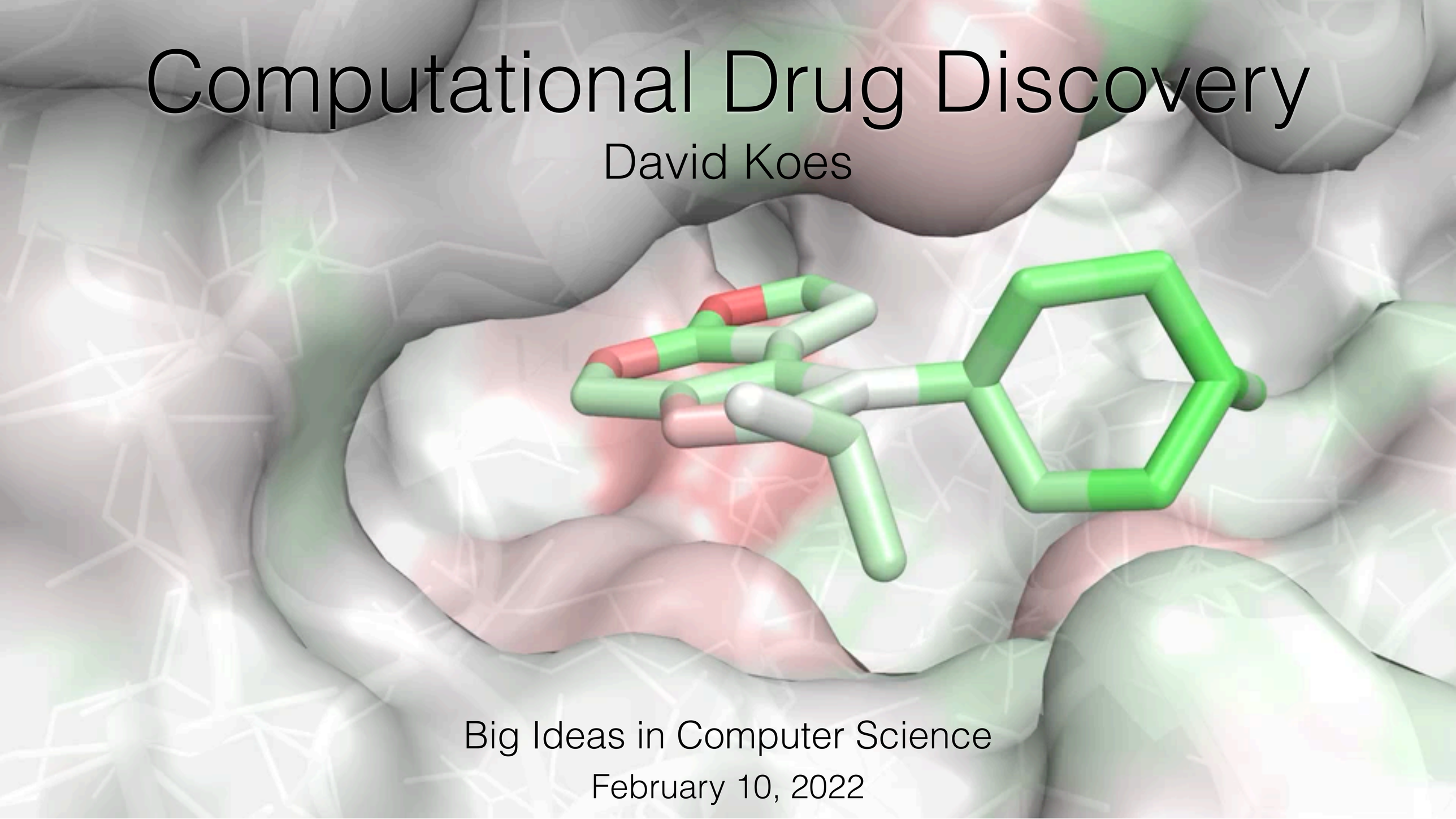
Computational Drug Discovery

David Koes

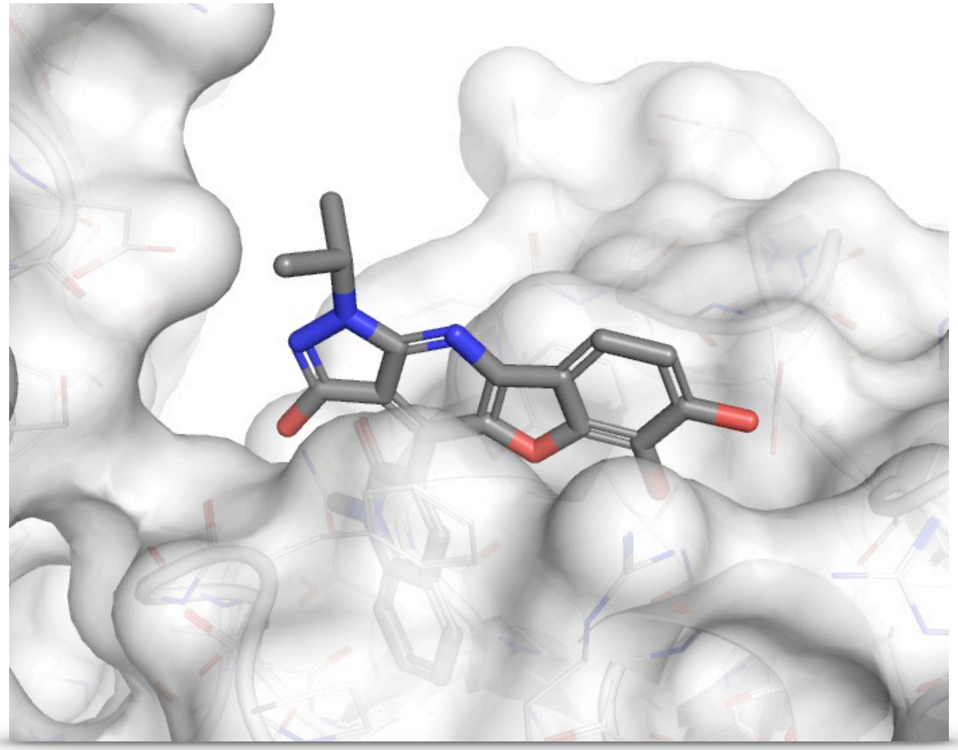


Big Ideas in Computer Science

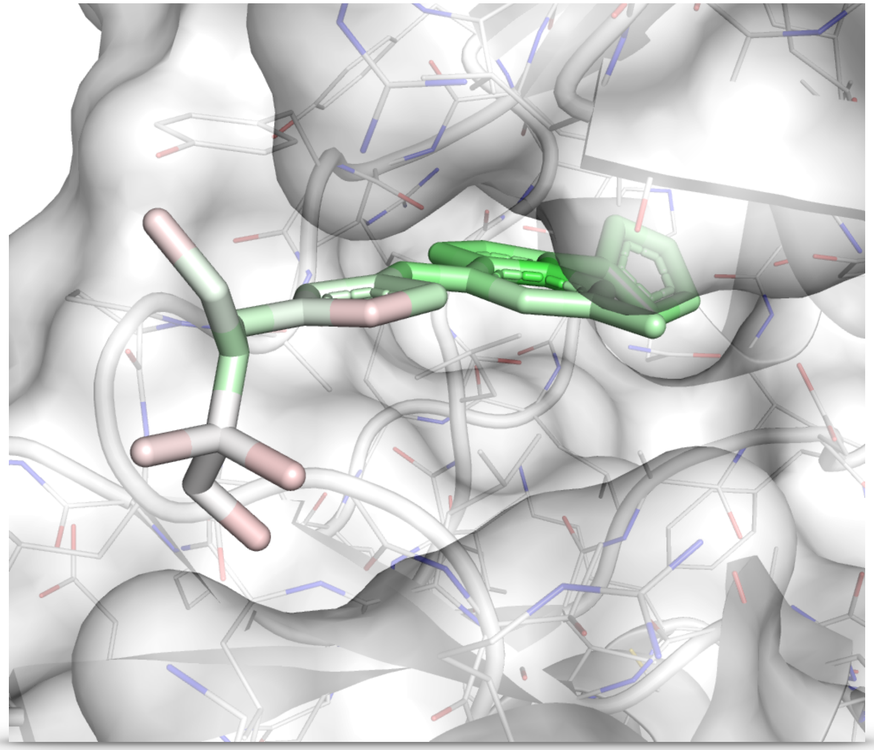
February 10, 2022

Structure Based Drug Design

Pose Prediction Binding Discrimination Affinity Prediction

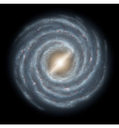


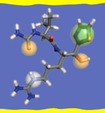
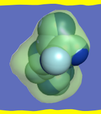
Virtual Screening

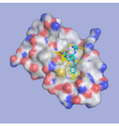
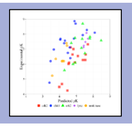


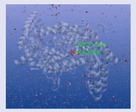
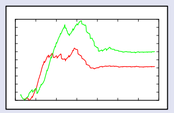
Lead Optimization

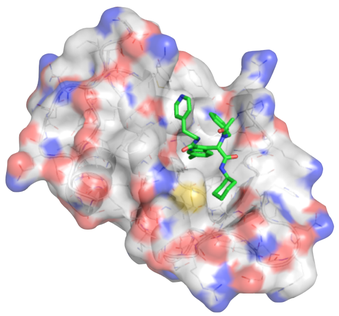
Drug Discovery Funnel

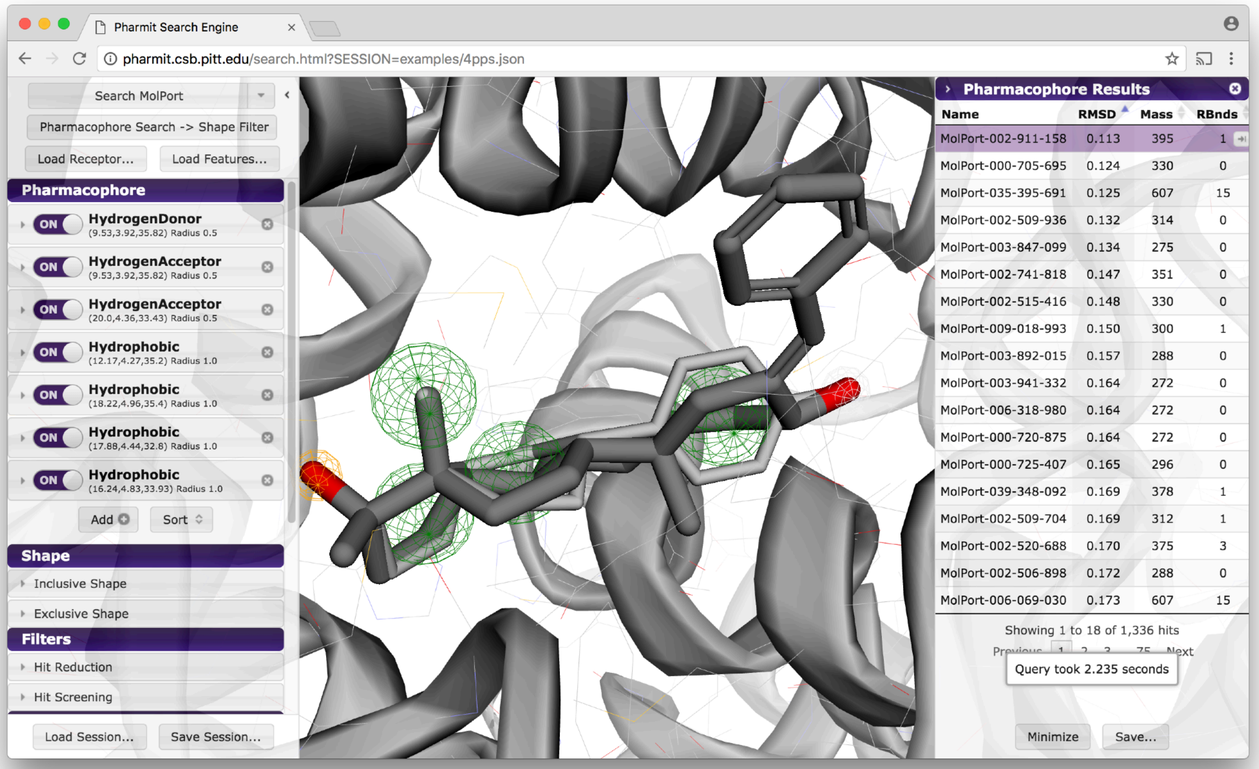
Purchasable  Accessible


Matching



Scoring


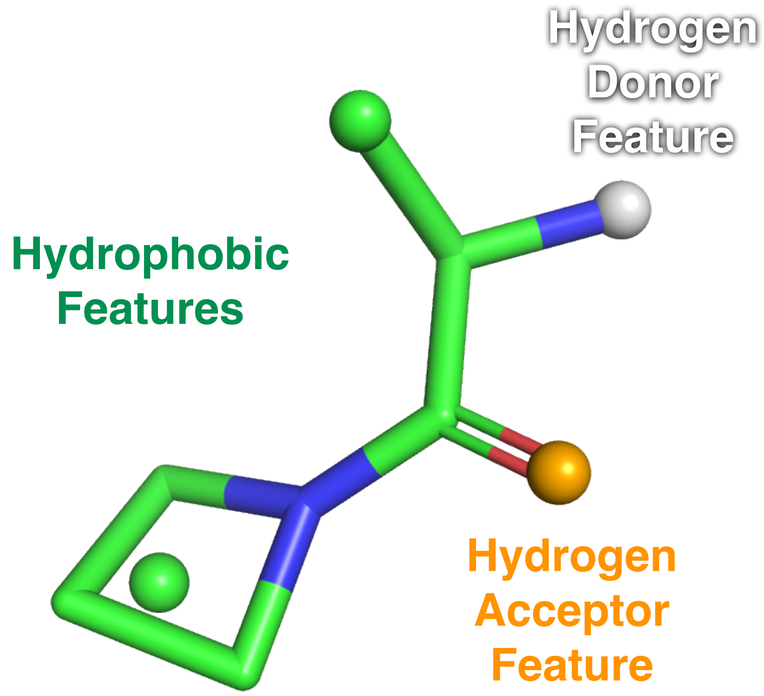

Dynamics






http://pharmit.csb.pitt.edu

Pharmer

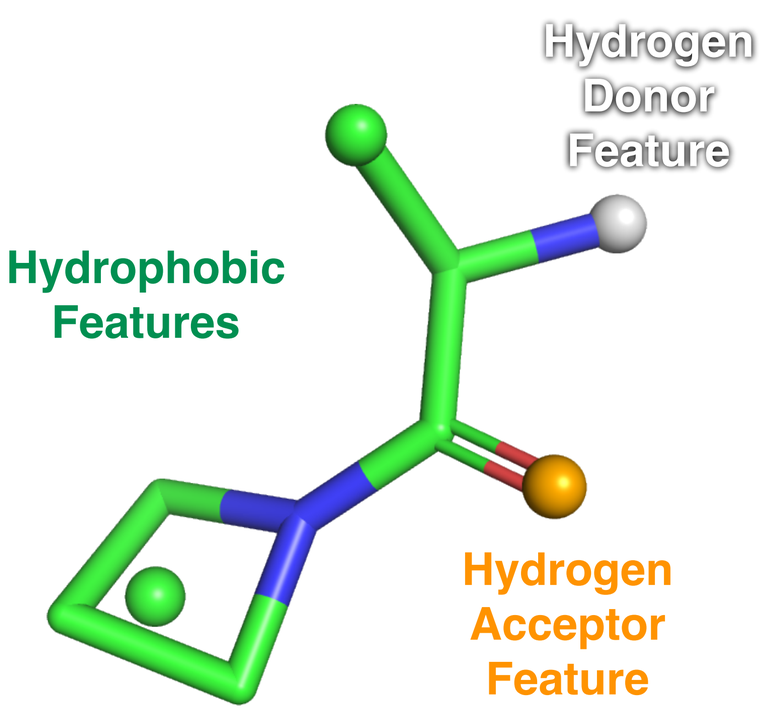


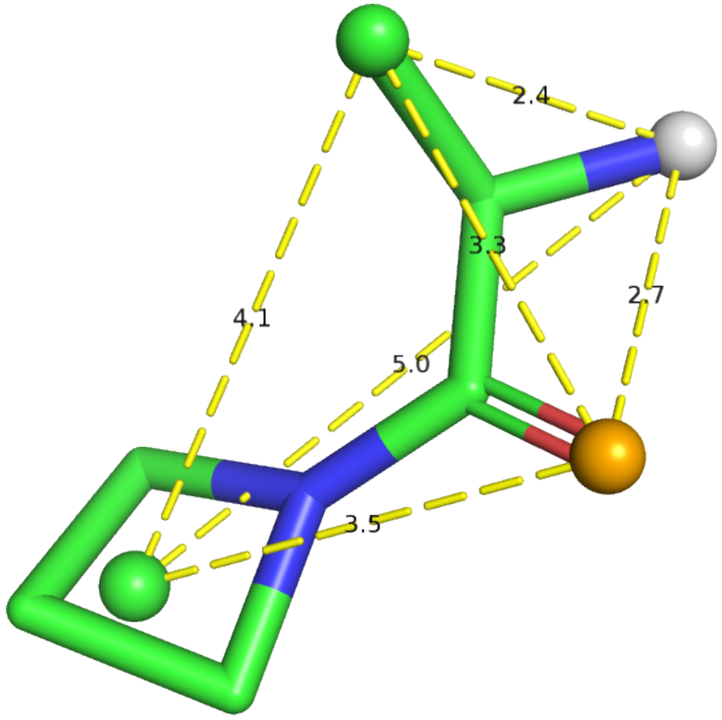
Pharmacophore

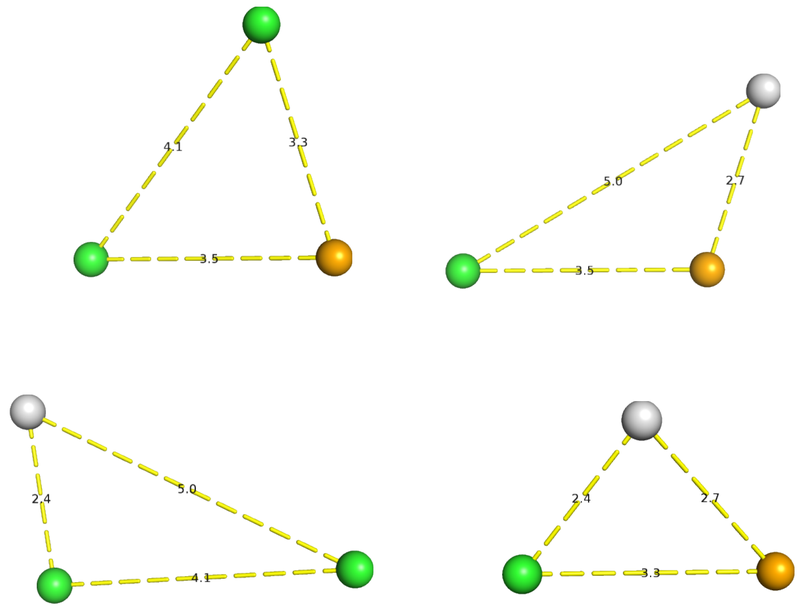
A spatial arrangement of molecular features essential for biological activity

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

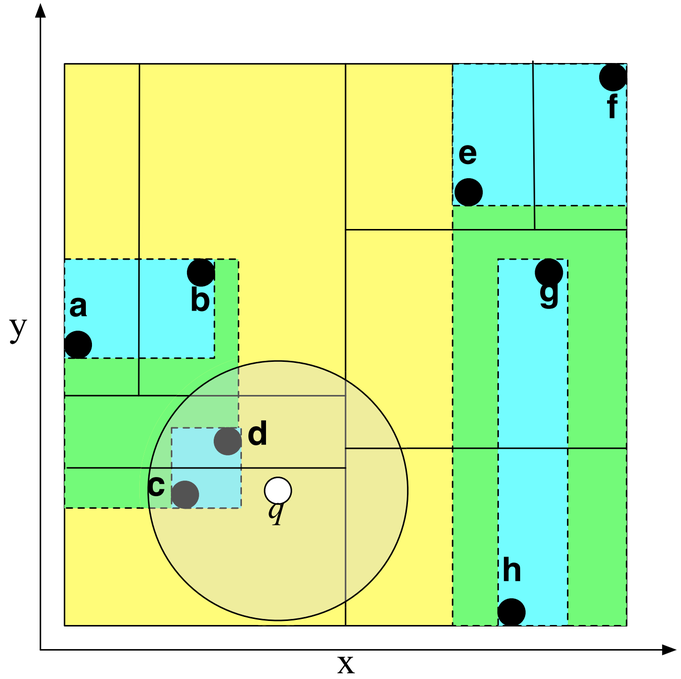
Pharmer

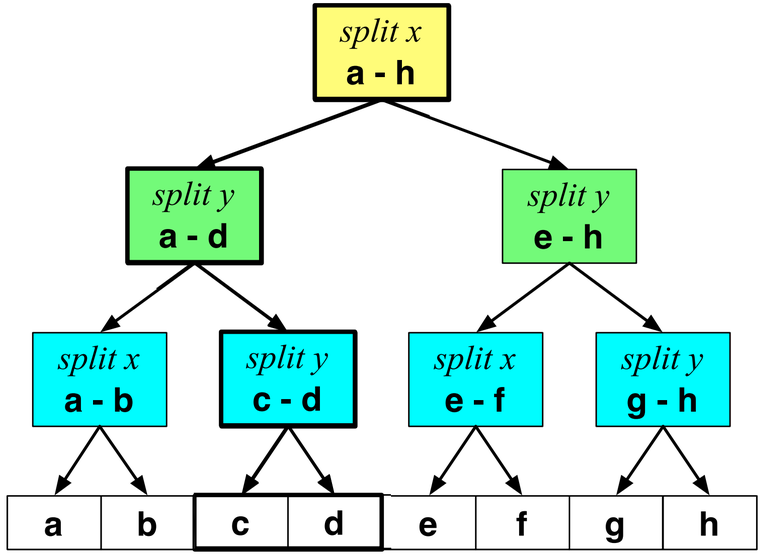

Hydrogen
Donor
Feature
Hydrophobic
Features
Hydrogen
Acceptor
Feature

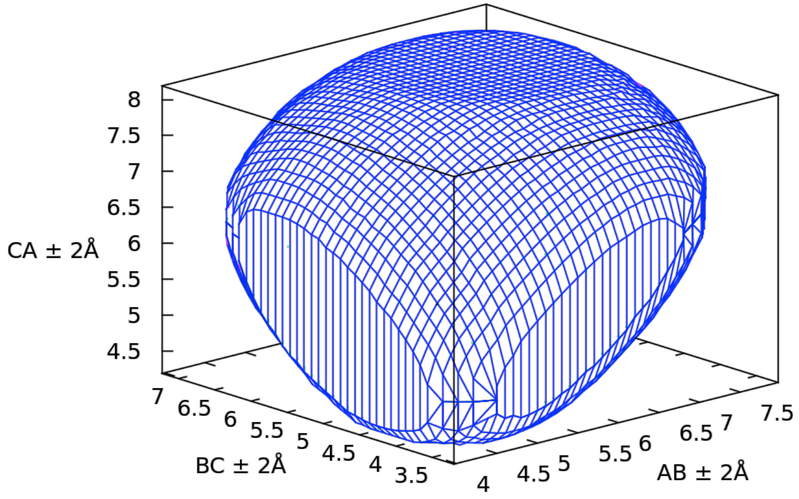

2.4
3.3
2.7
4.1
5.0
3.5

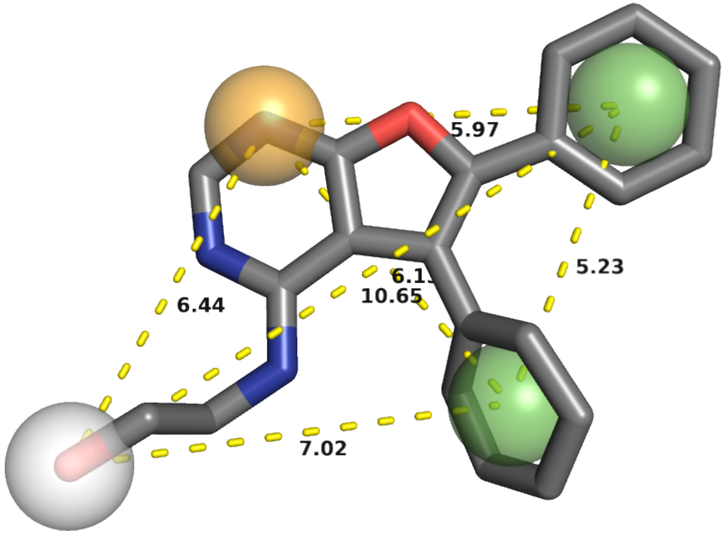

4.1
3.3
3.5
5.0
2.7
3.5
2.4
5.0
4.1
2.4
2.7
3.3

Pharmer

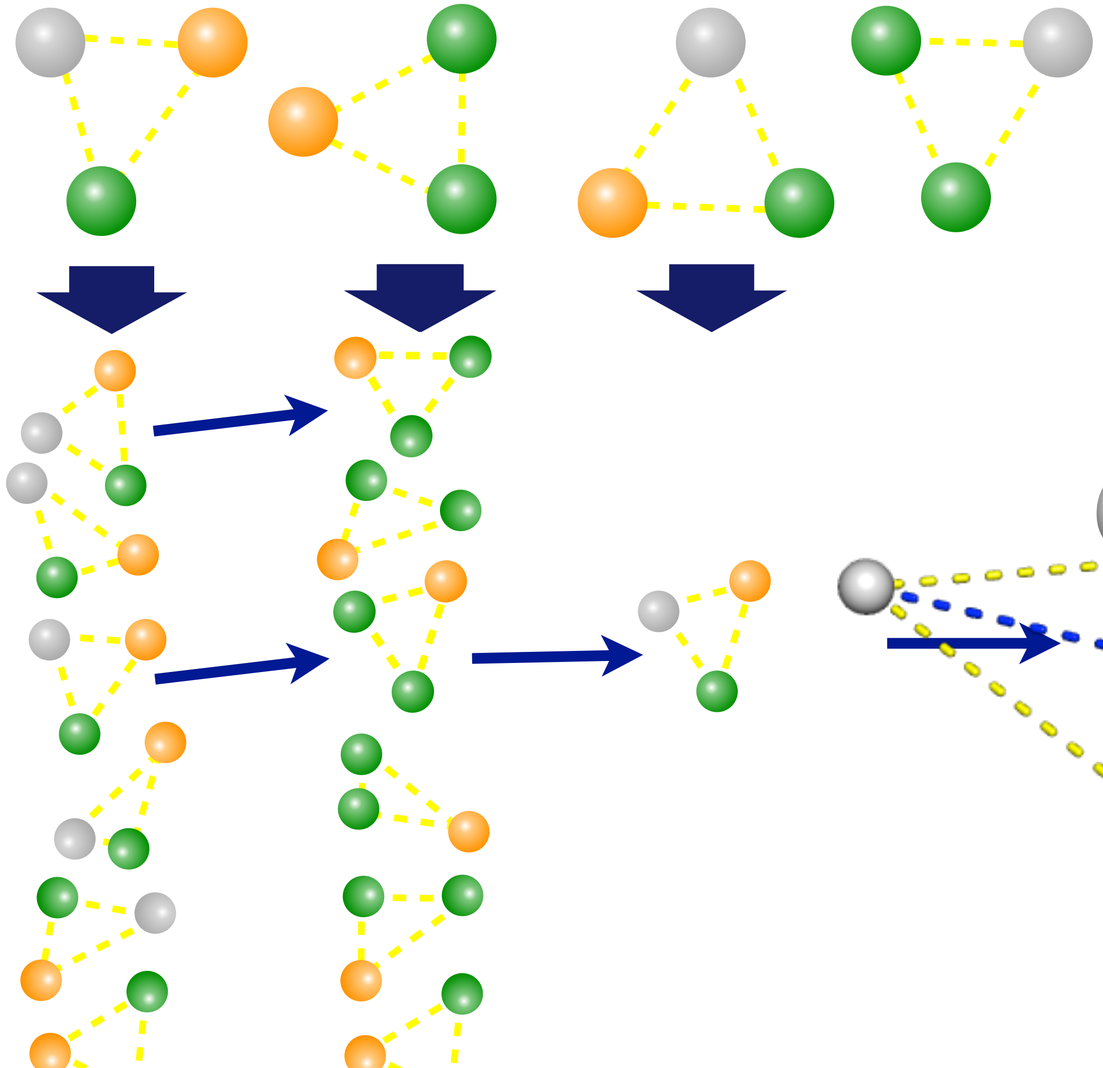

y
x
a
b
c
d
e
f
g
h
q

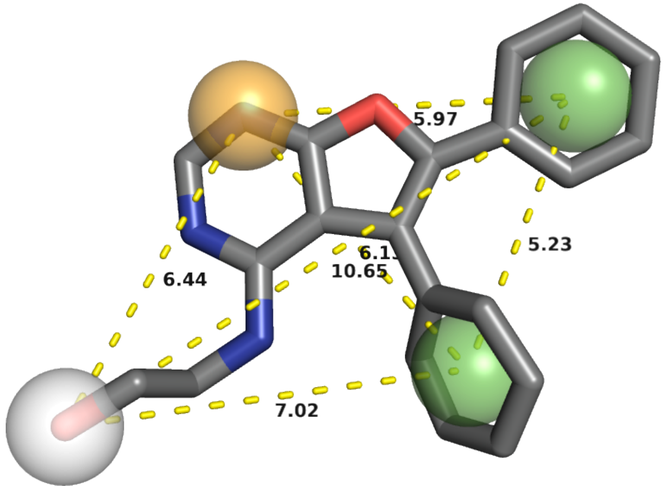

split x
a - h
split y
a - d
split y
e - h
split x
a - b
split y
c - d
split x
e - f
split y
g - h
a
b
c
d
e
f
g
h

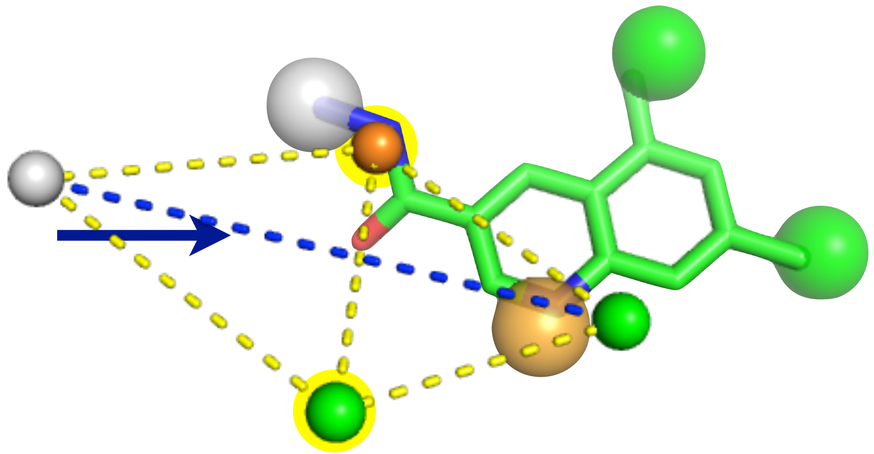

CA ± 2Å
BC ± 2Å
AB ± 2Å


5.97
5.23
6.44
10.65
7.02

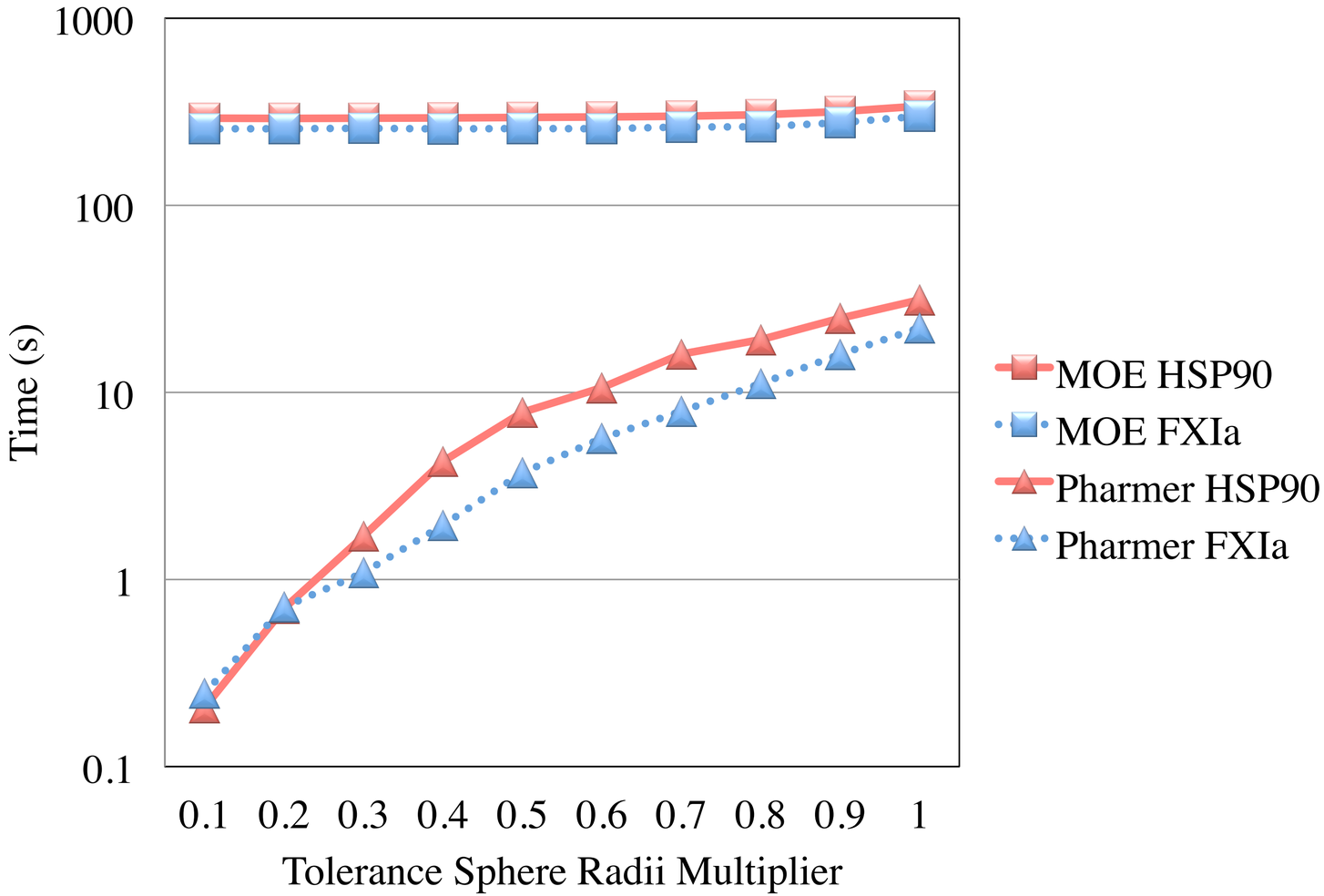
Pharmer

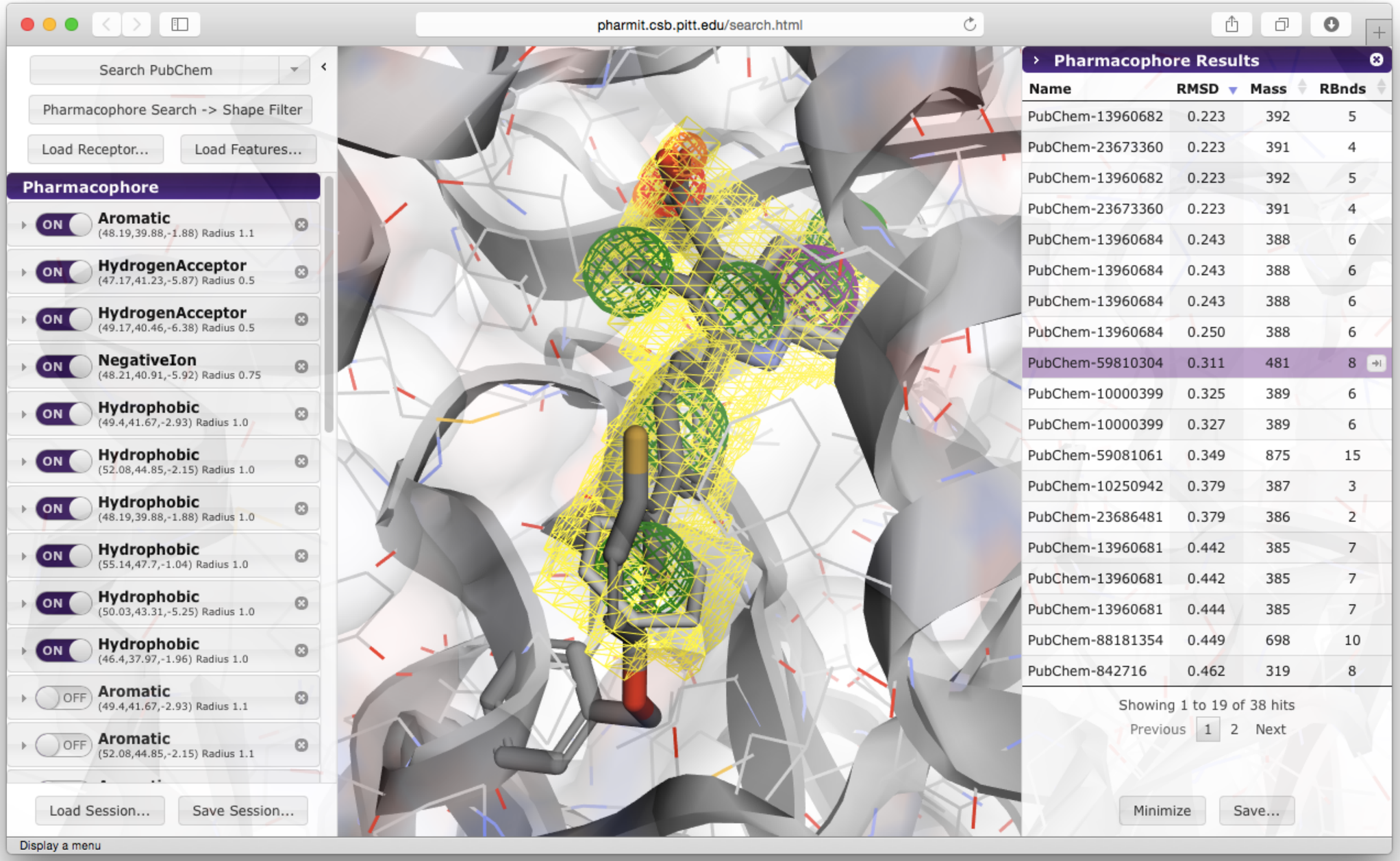






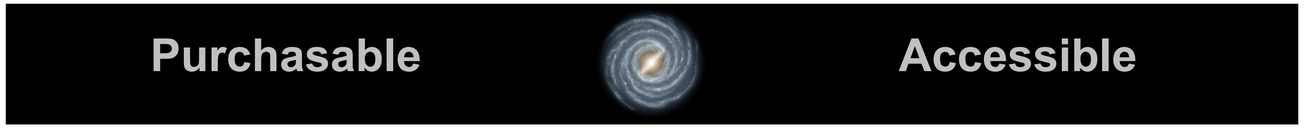
Pharmer

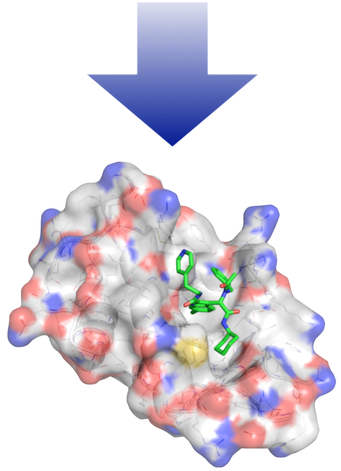

1000
100
10
1
0.1
Time (s)
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Tolerance Sphere Radii Multiplier
MOE HSP90
MOE FXIa
Pharmer HSP90
Pharmer FXIa

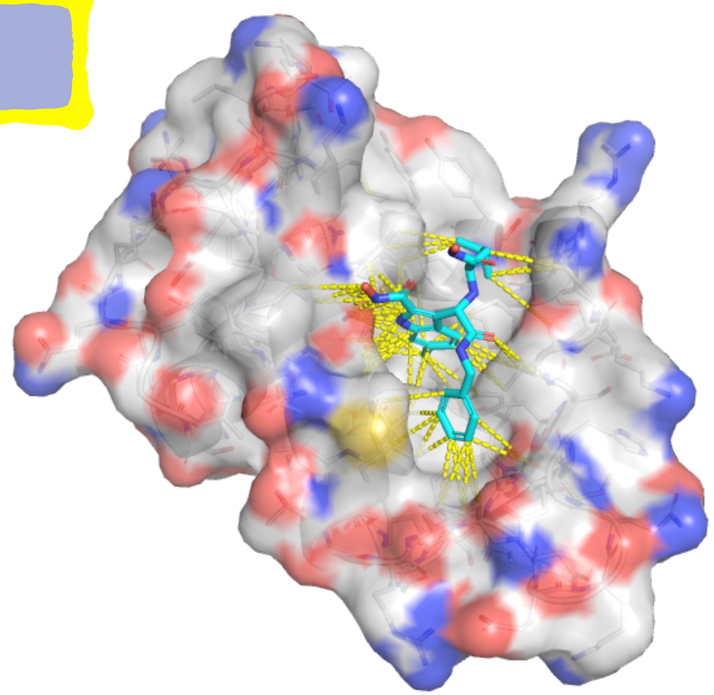


http://pharmit.csb.pitt.edu

Drug Discovery Funnel



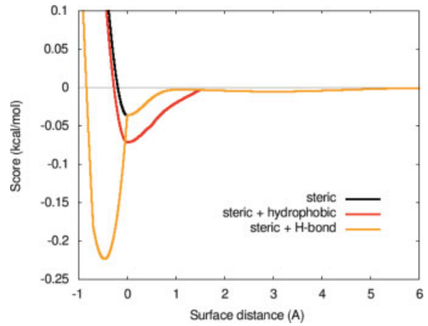




$$\text{gauss}_1(d) = w_{\text{guass}_1} e^{-(d/0.5)^2}$$
$$\text{gauss}_2(d) = w_{\text{guass}_2} e^{-((d-3)/2)^2}$$
$$\text{repulsion}(d) = \begin{cases} w_{\text{repulsion}} d^2 & d < 0 \\ 0 & d \geq 0 \end{cases}$$

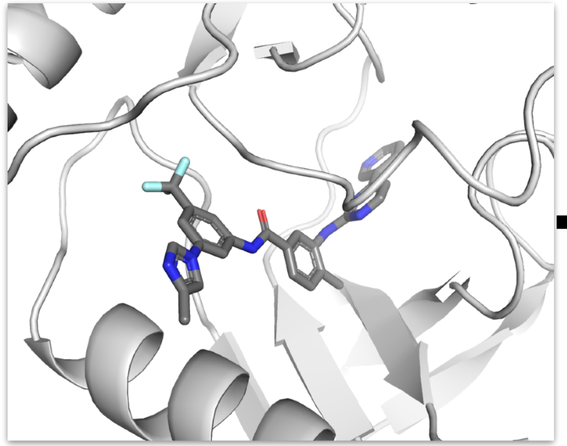
$$\text{hydrophobic}(d) = \begin{cases} w_{\text{hydrophobic}} & d < 0.5 \\ 0 & d > 1.5 \\ w_{\text{hydrophobic}}(1.5 - d) & otherwise \end{cases}$$

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



O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry* 31 (2010) 455-461

Protein-Ligand Scoring



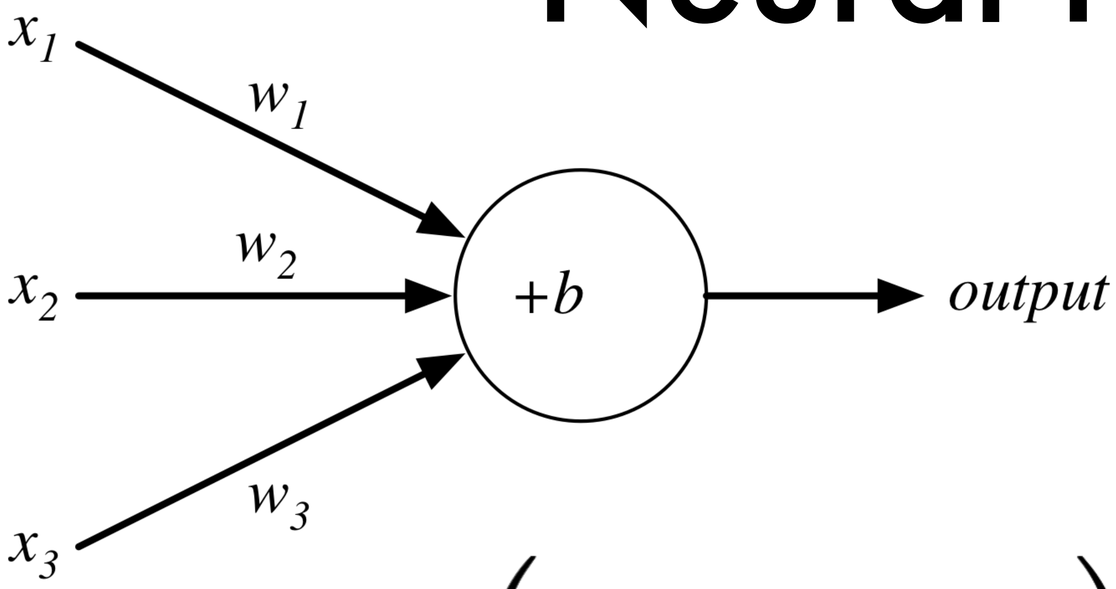
Model

Pose Prediction

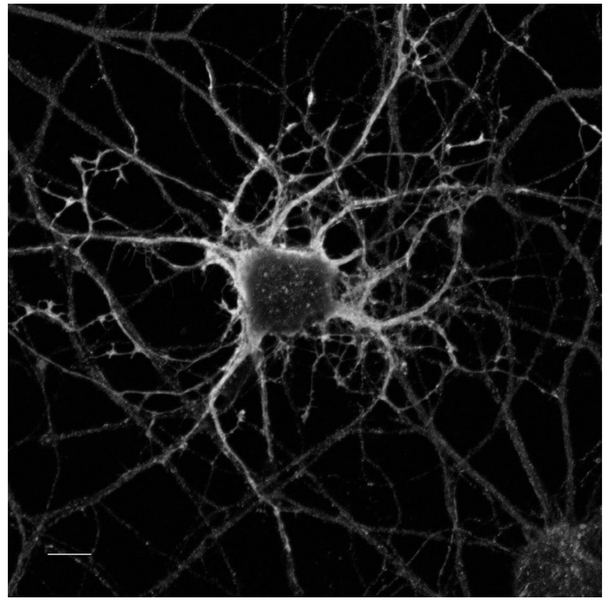
Binding Discrimination

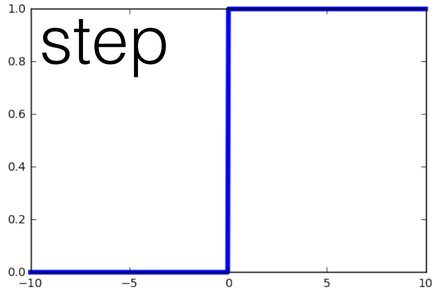
Affinity Prediction

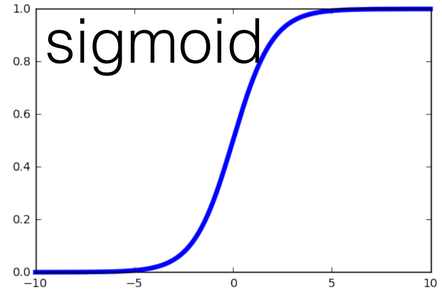
Neural Networks

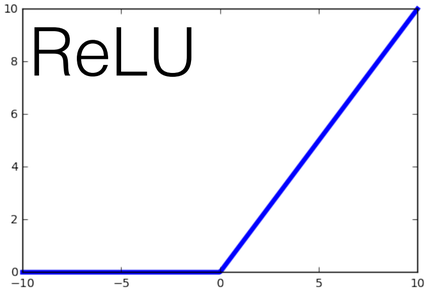


$$output = \sigma\left(\sum_i w_i x_i + b\right)$$

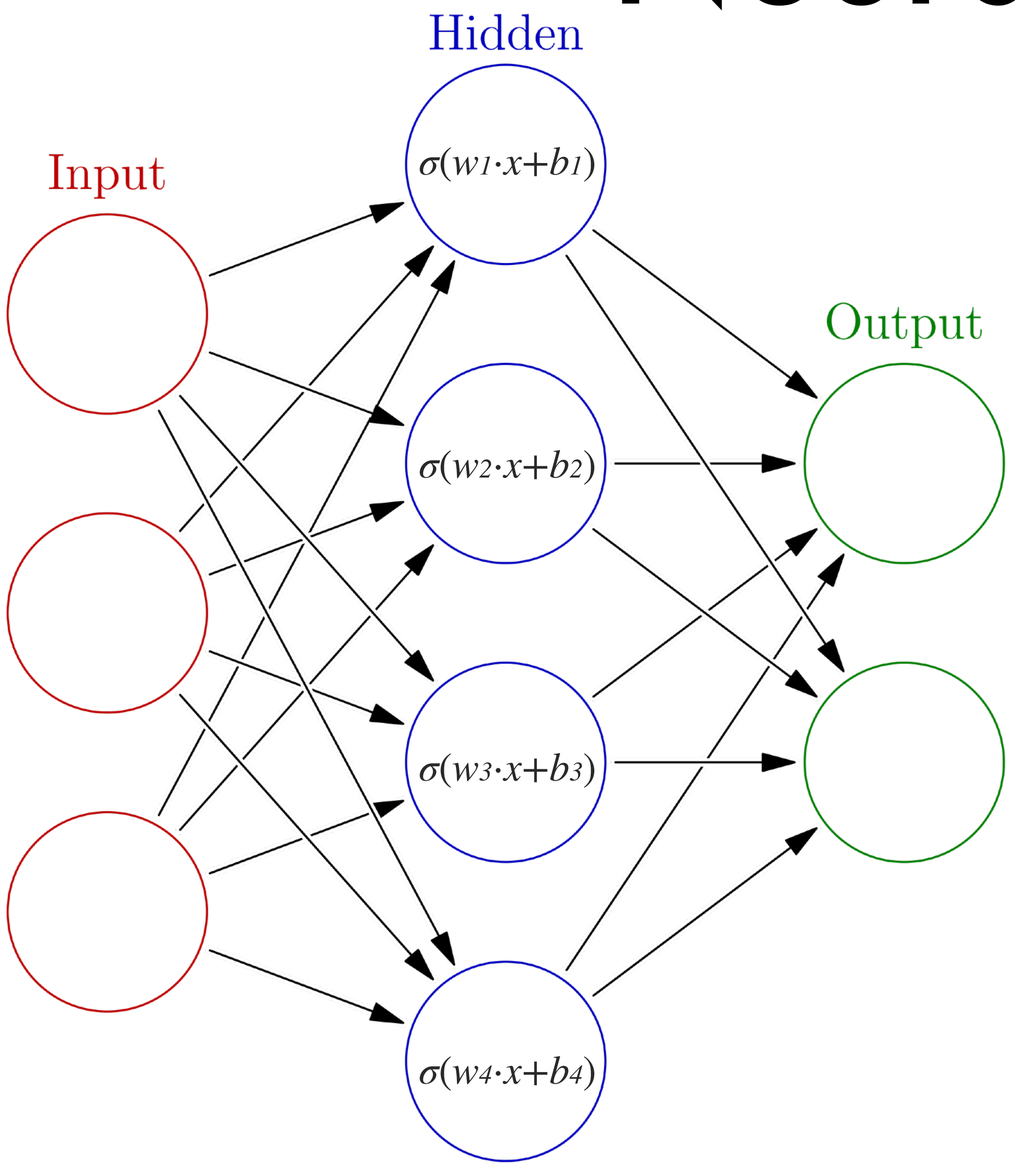


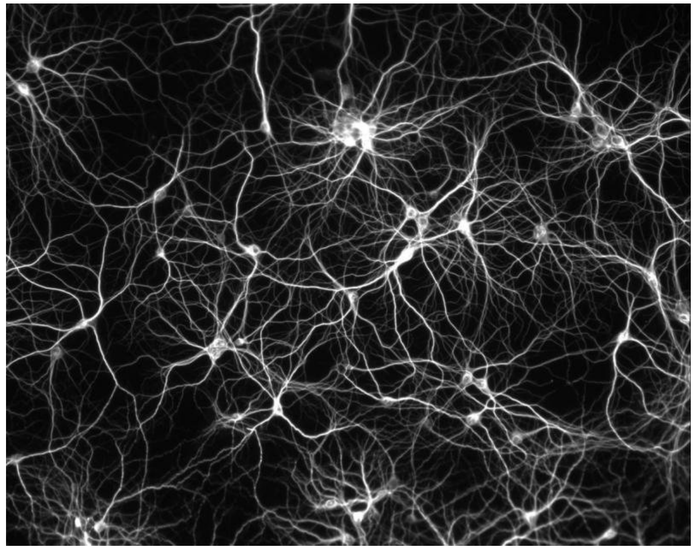






Neural Networks



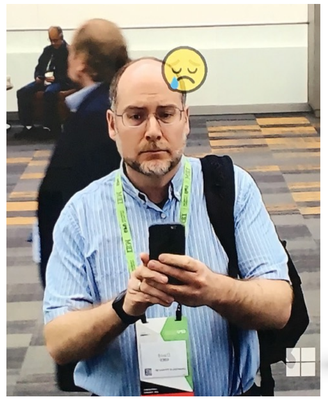


The **universal approximation theorem** states that, under reasonable assumptions, a feedforward **neural network** with a finite number of nodes **can approximate any continuous** function to within a given error over a bounded input domain.

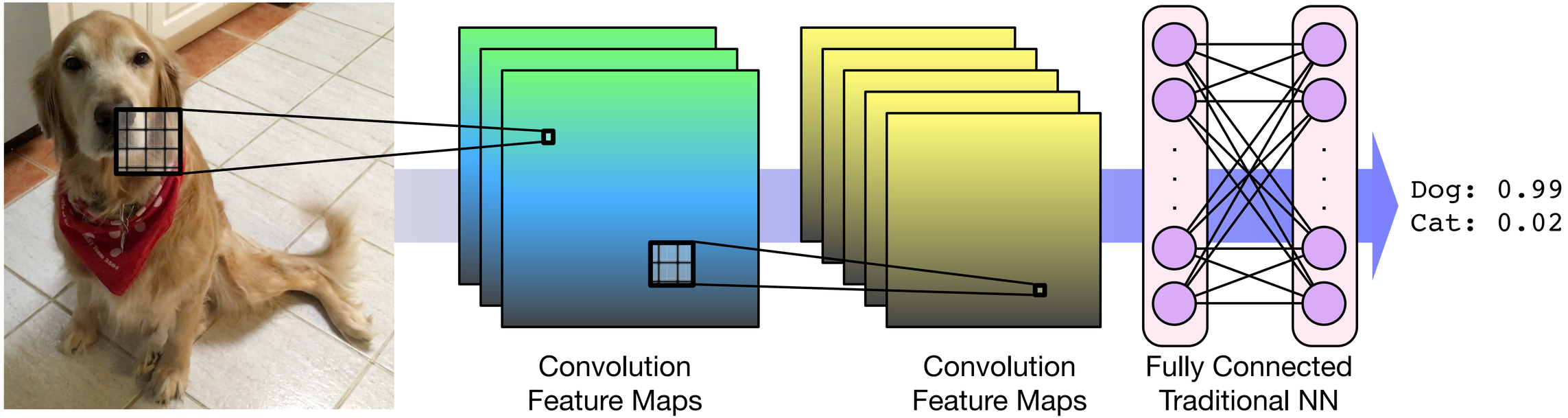
Deep Learning

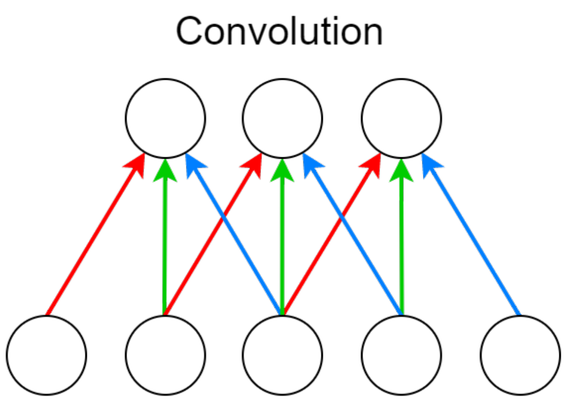
$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

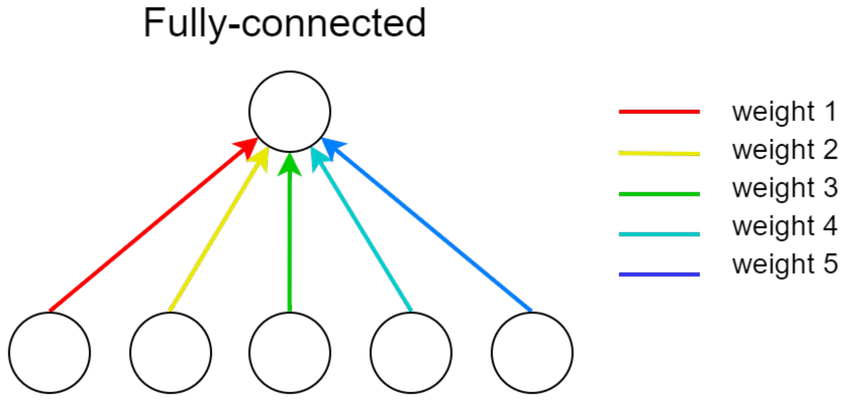
$$\frac{\partial L}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j \text{ and } \frac{\partial L}{\partial b^l_j} = \delta^l_j$$



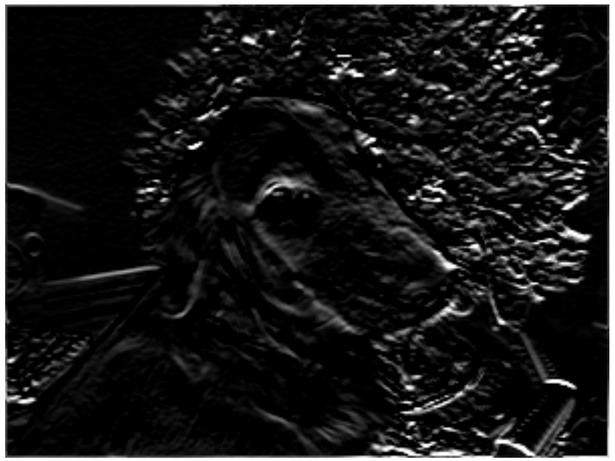
Convolutional Neural Networks



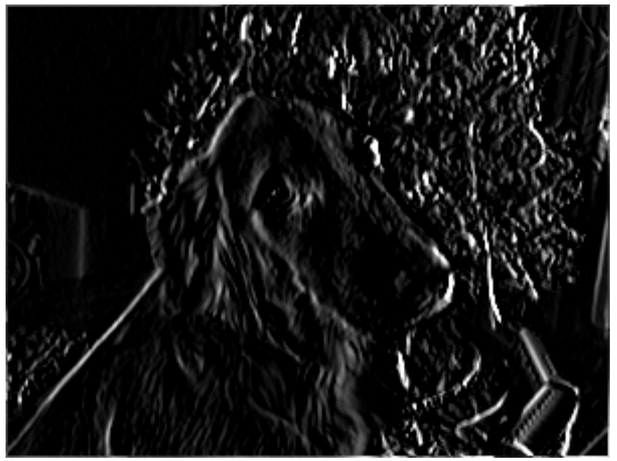




Convolutional Filters



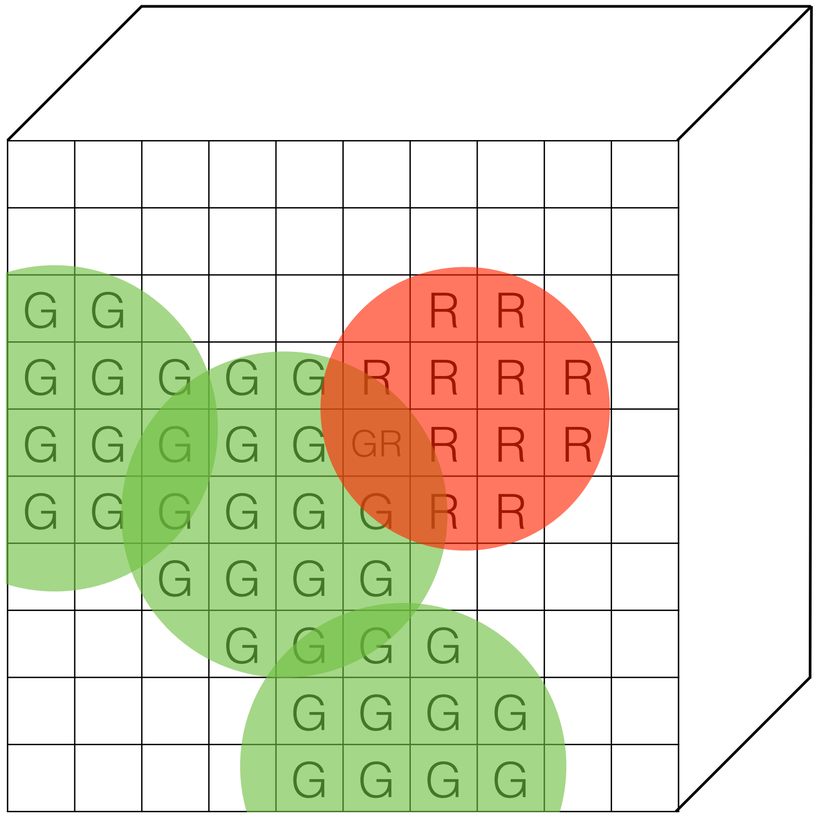
-1	-1	-1
0	0	0
1	1	1



-1	0	1
-1	0	1
-1	0	1

-1	-1	-1
-1	8	-1
-1	-1	-1

Protein-Ligand Representation



(R,G,B) pixel →

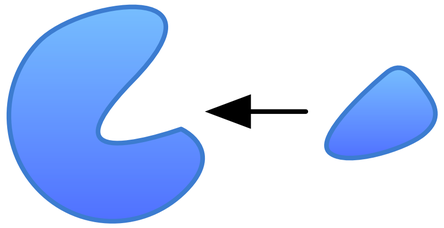
(Carbon, Nitrogen, Oxygen,…) **voxel**

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

Training

PDBbind 2016 refined set

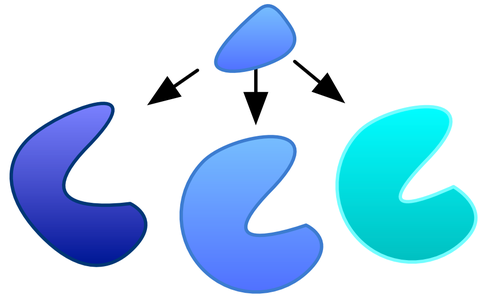
- 4056 protein-ligand complexes
- diverse targets
- wide range of affinities
- generate poses with AutoDock Vina
- include minimized crystal pose



Redocked Training Set

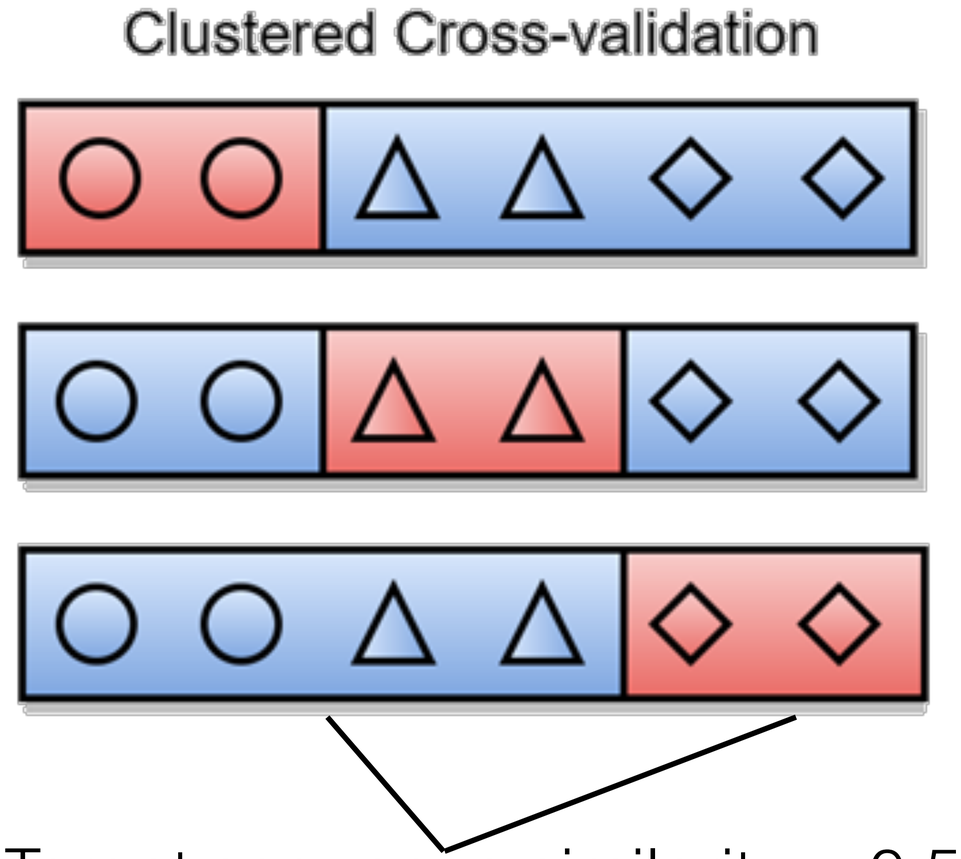
Pocketome

- 2923 distinct pockets
- 27,142 receptor structures
- 4,138,117 non-redundant poses
- generate poses with AutoDock Vina
- include minimized crystal pose



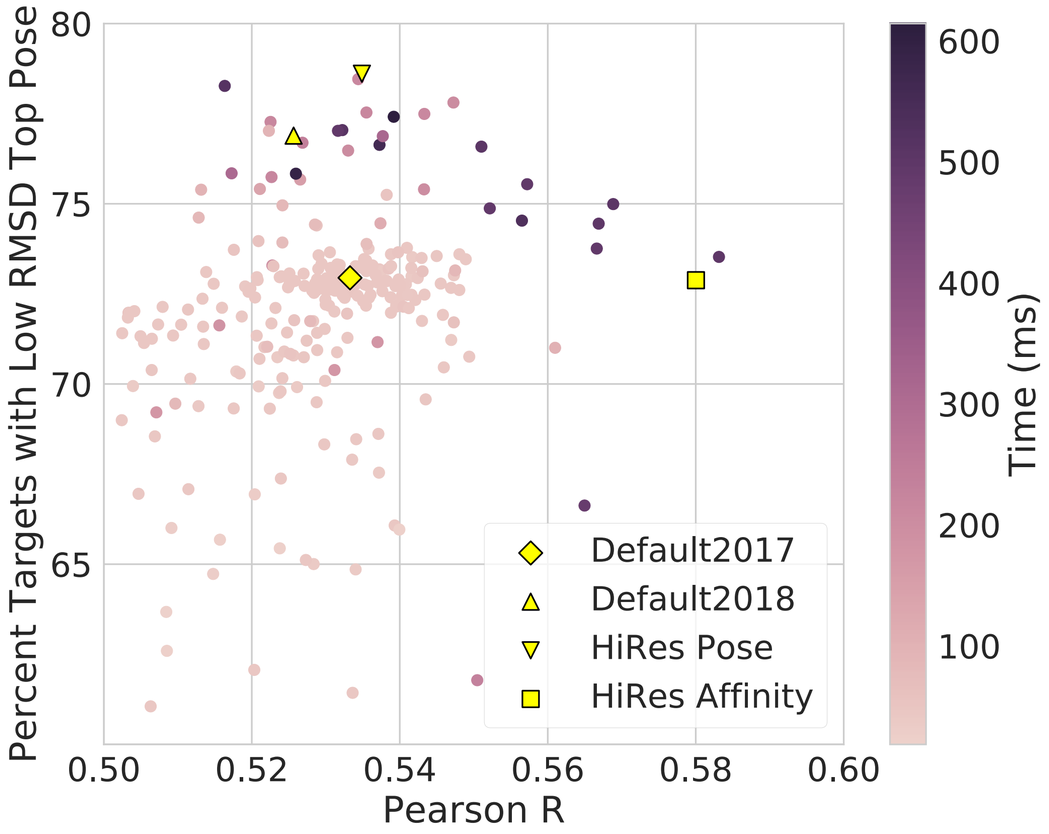
Crossdocked Training Set

Training

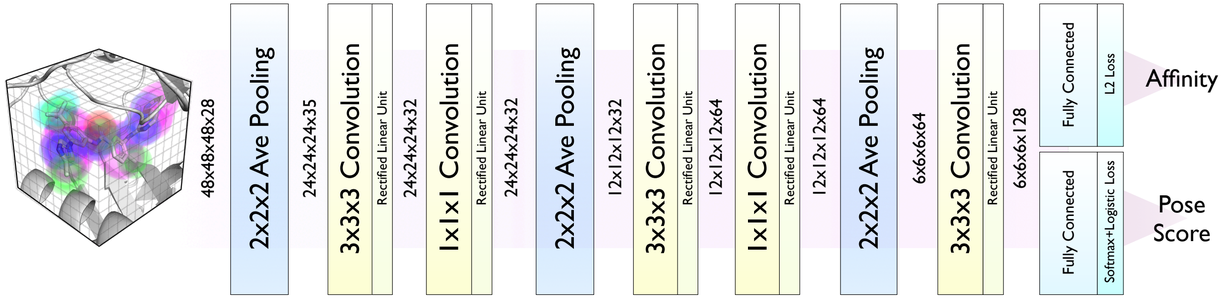


Target sequence similarity < 0.5
AND
Ligand similarity < 0.9

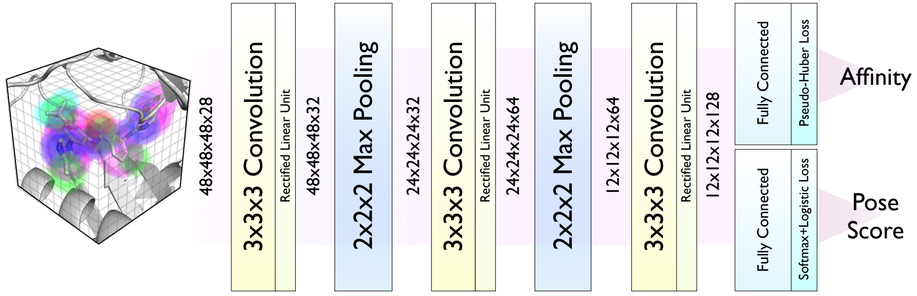
Optimized Models



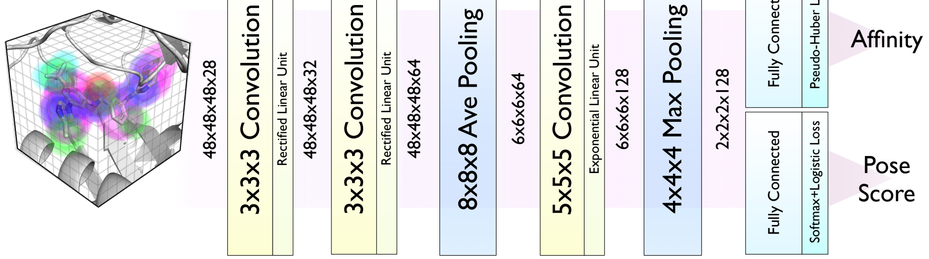
Default2018



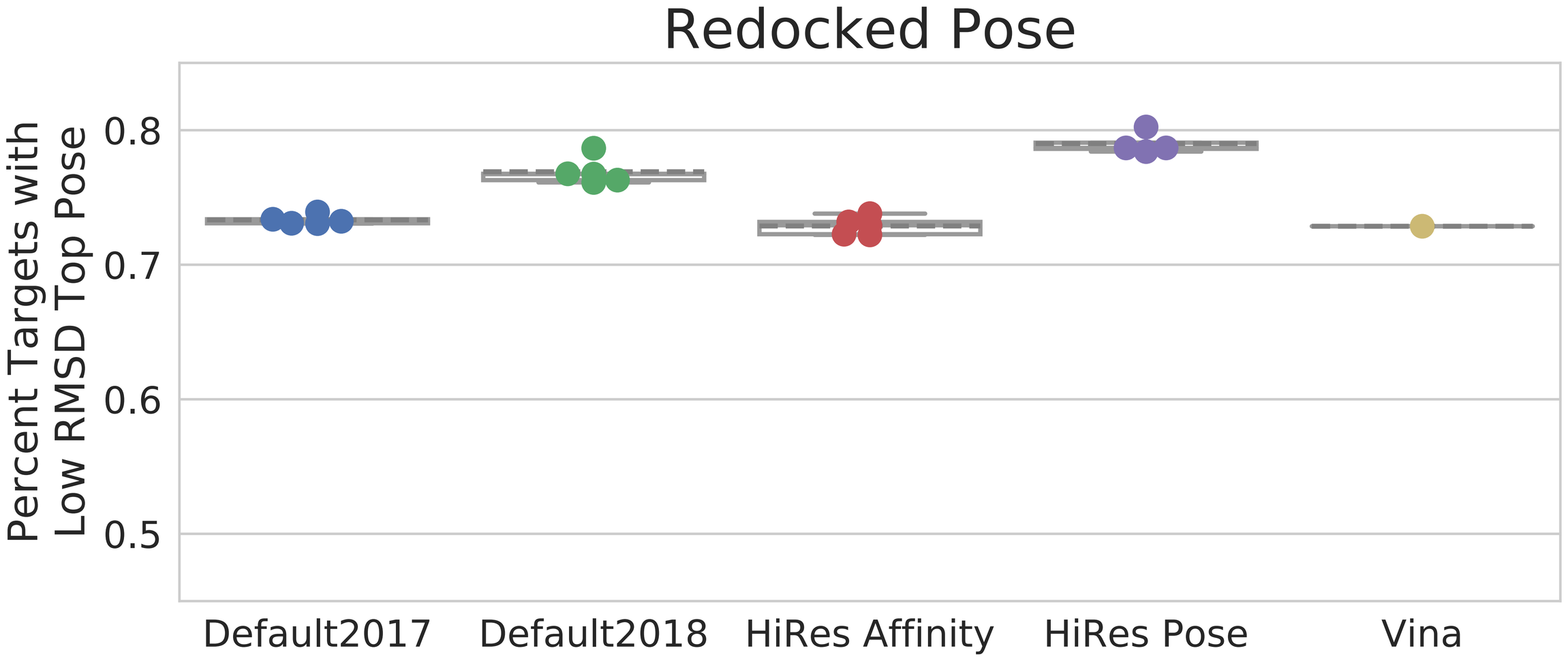
HiRes Pose



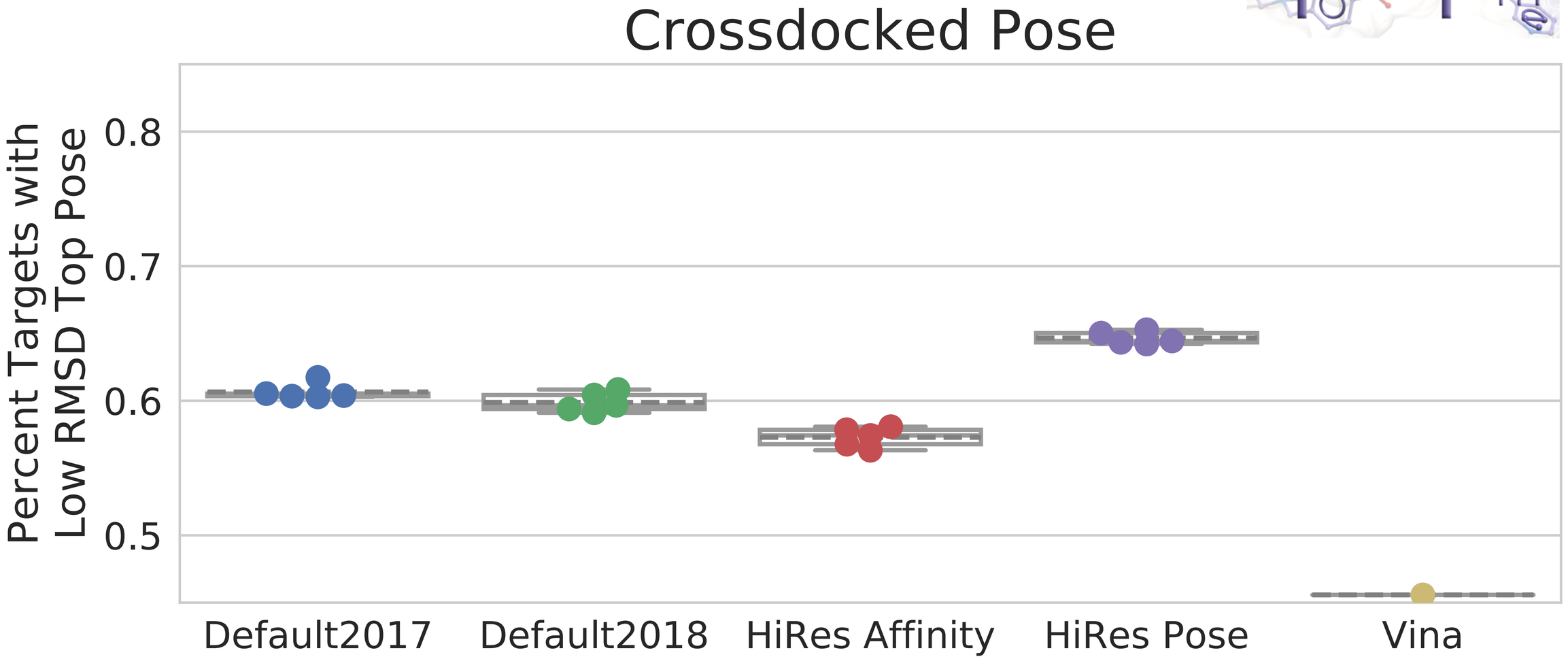
HiRes Affinity



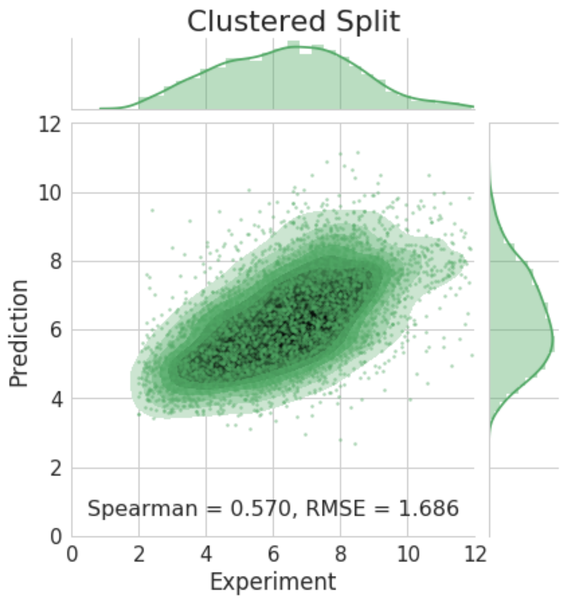
Pose Results

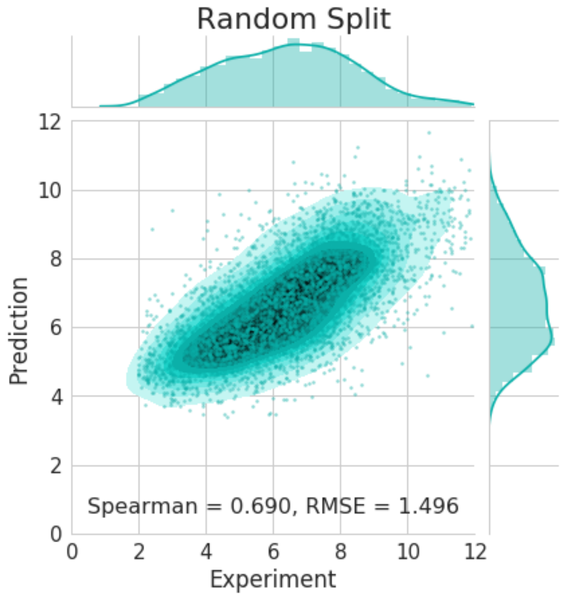


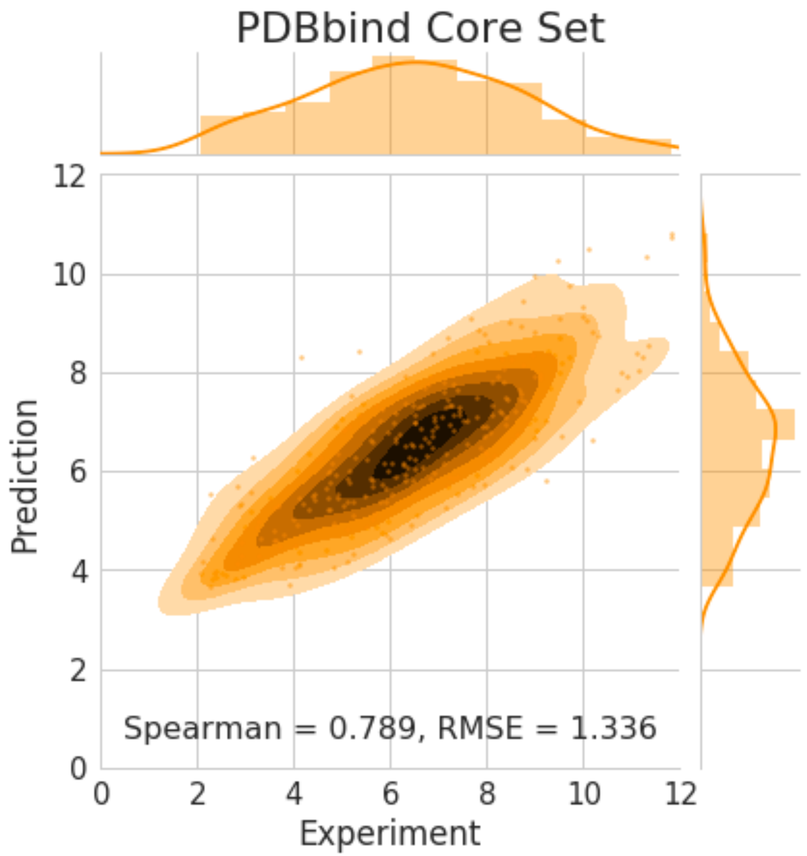
Pose Results

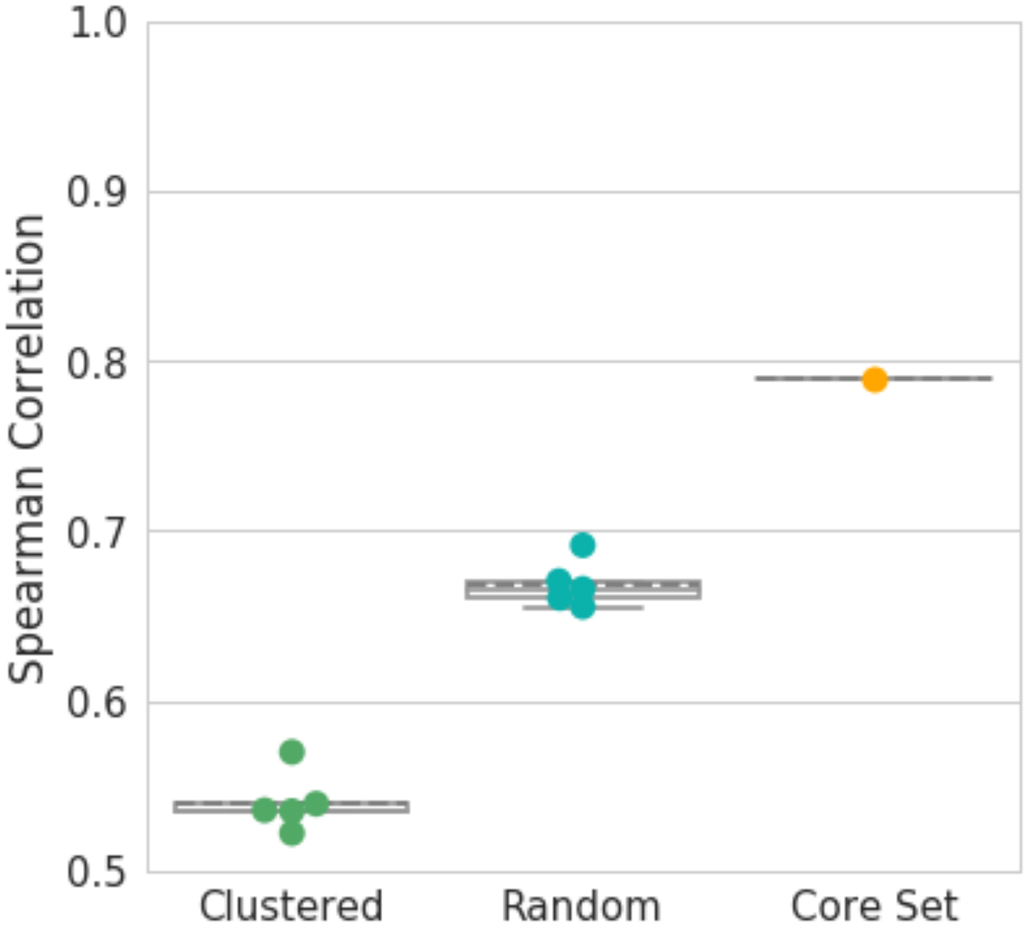


Affinity Results

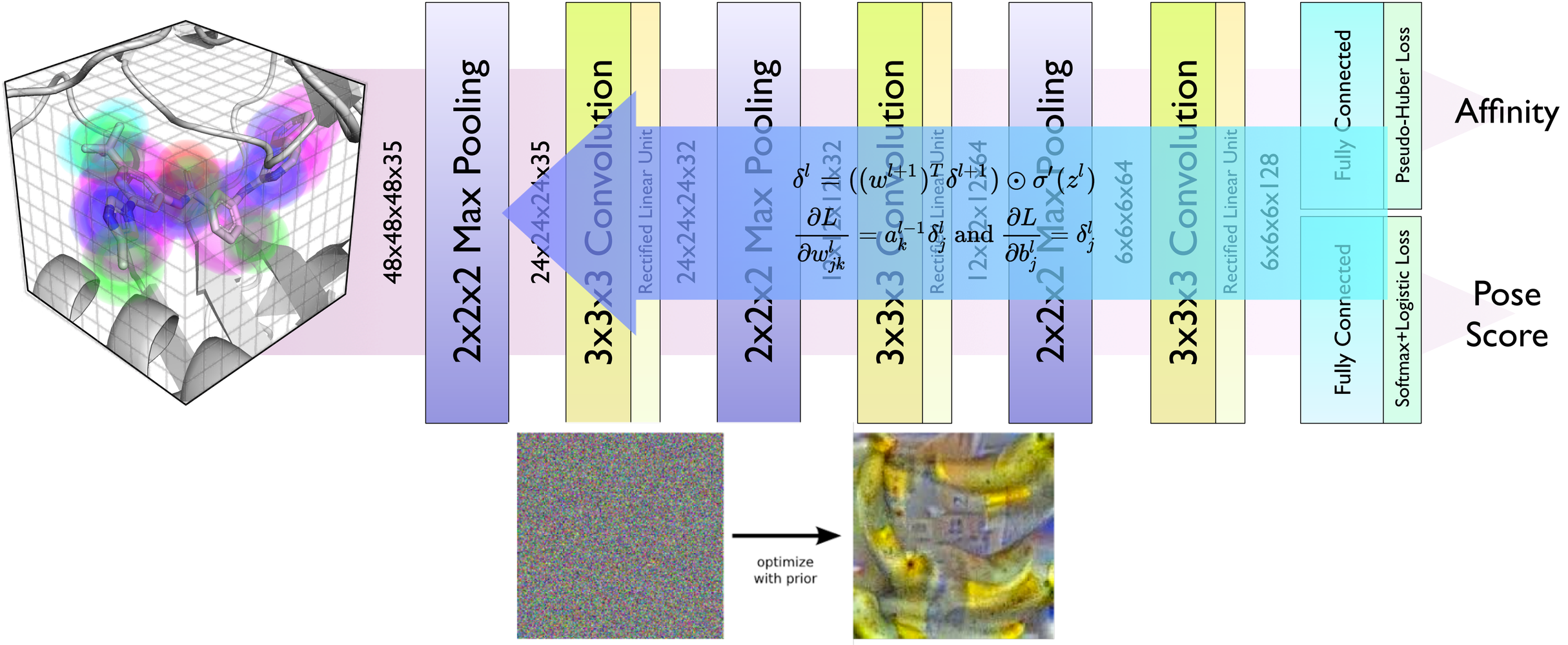




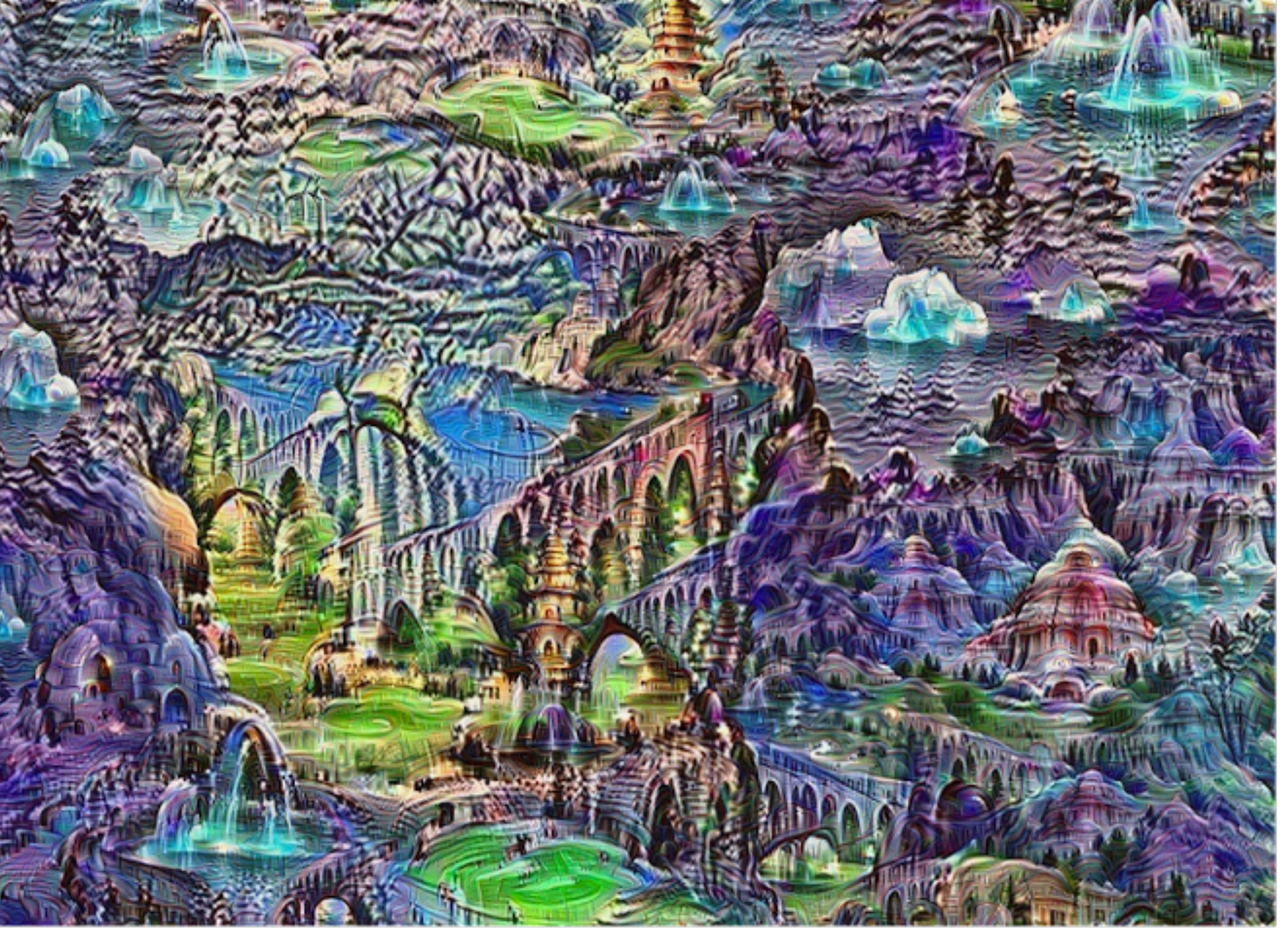
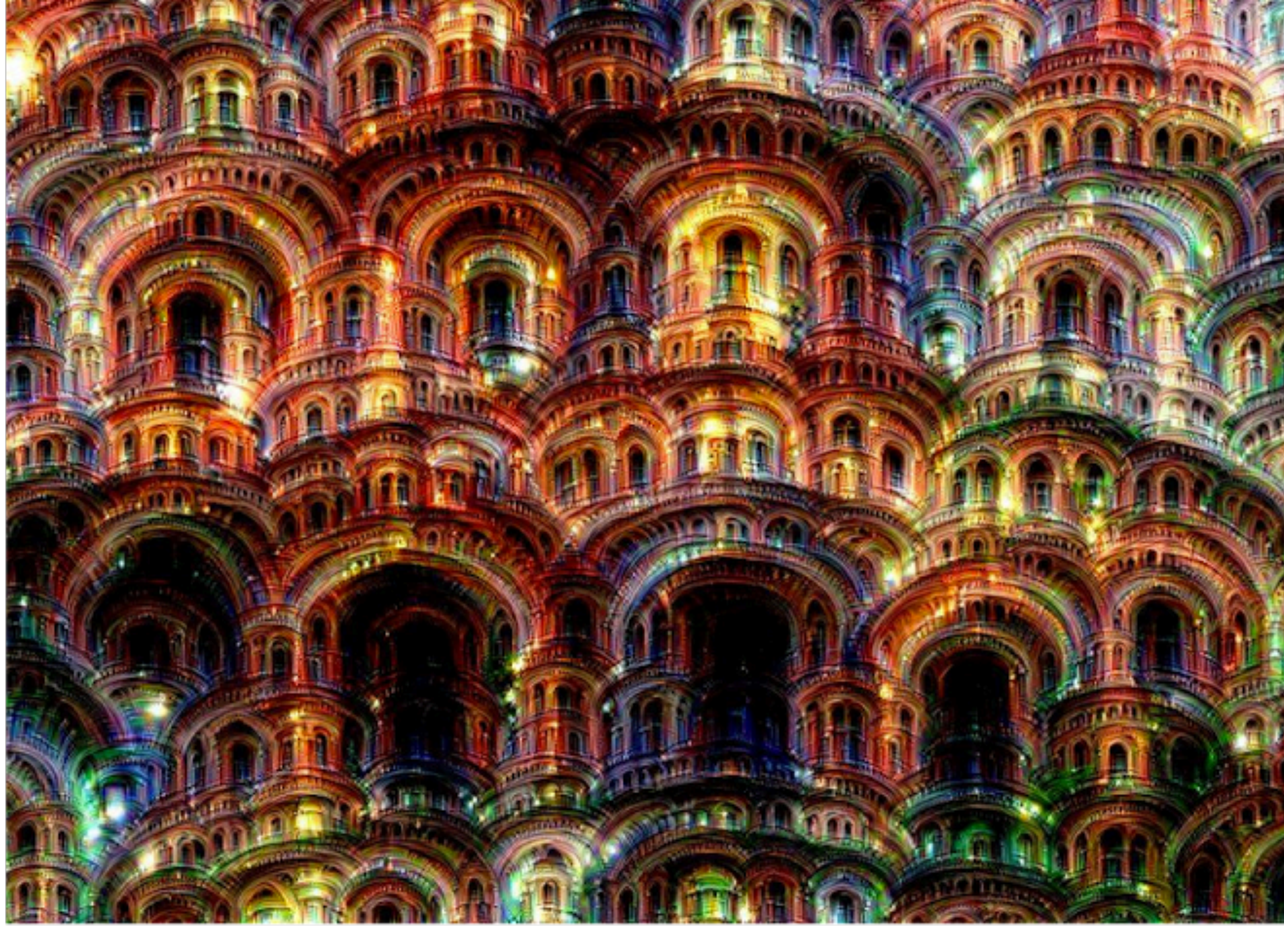
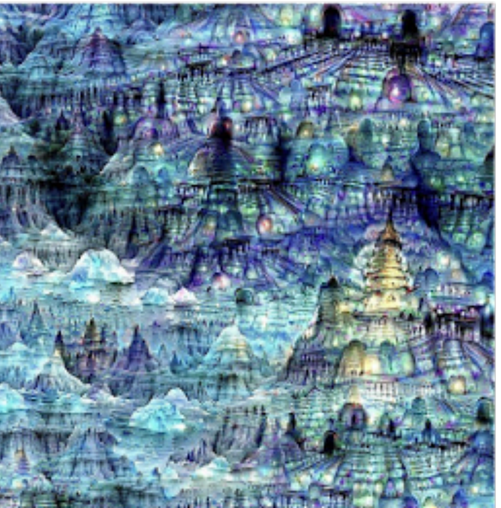
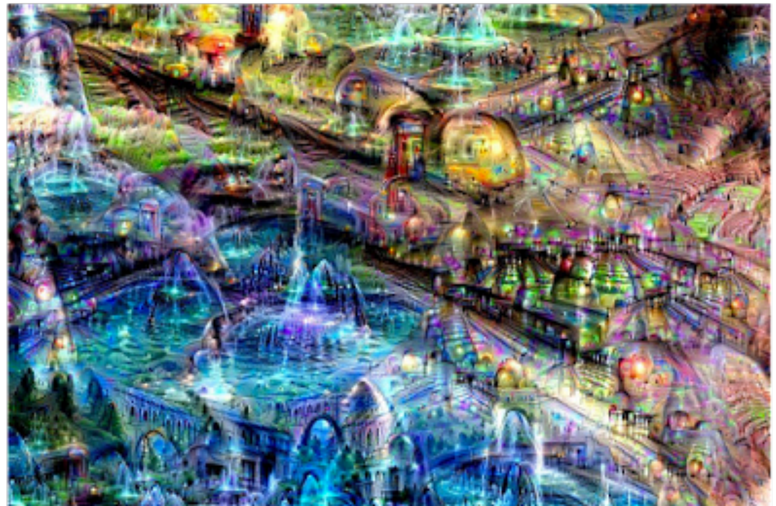


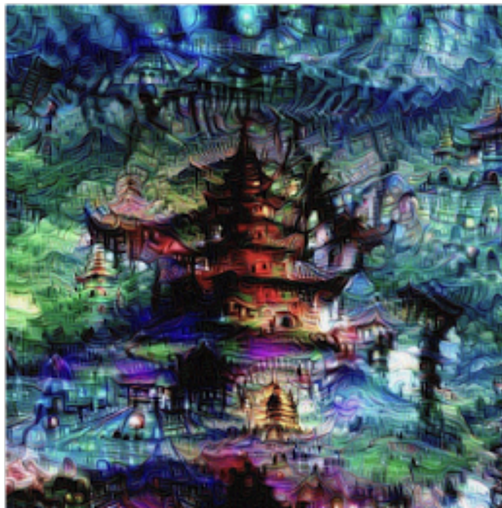


Gradients

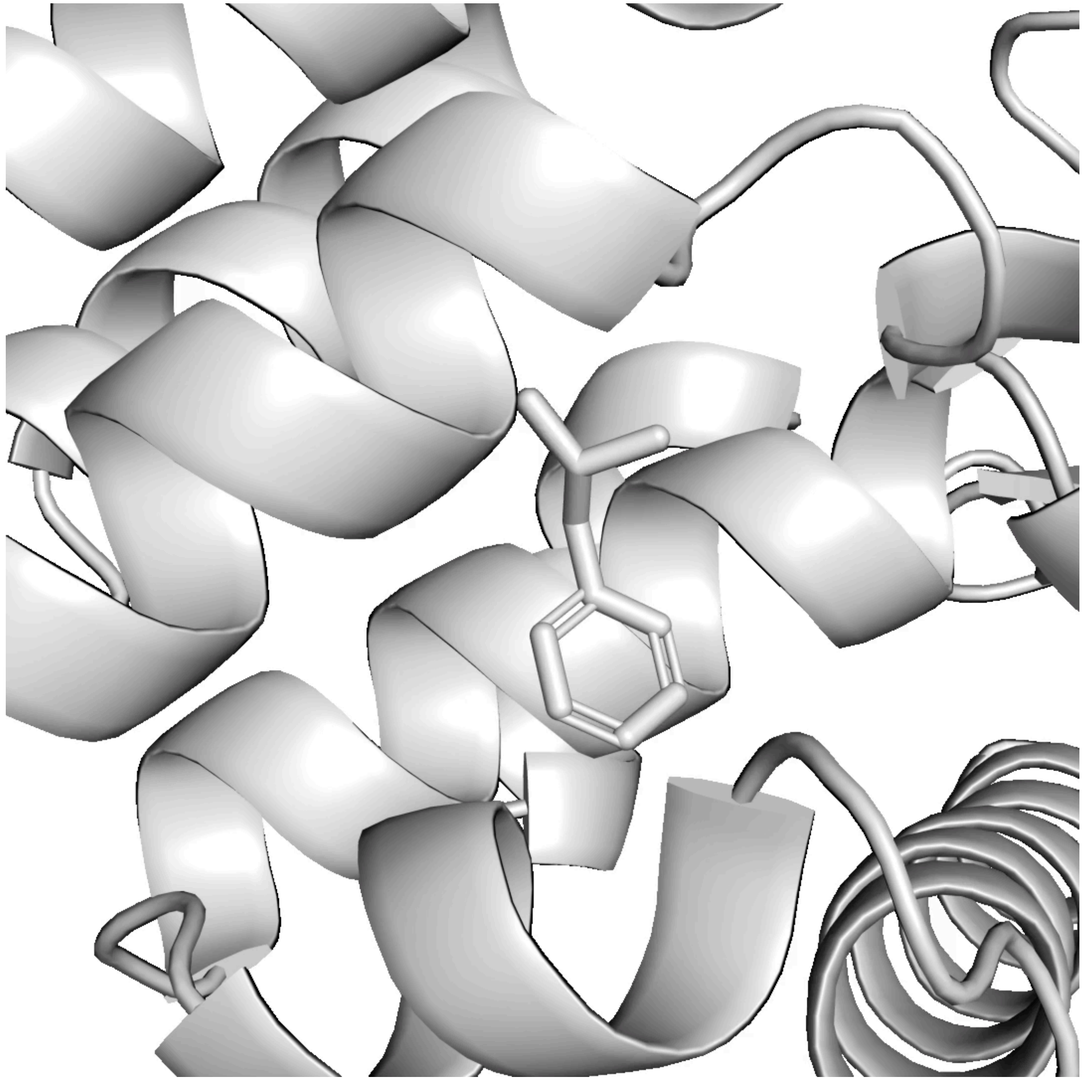


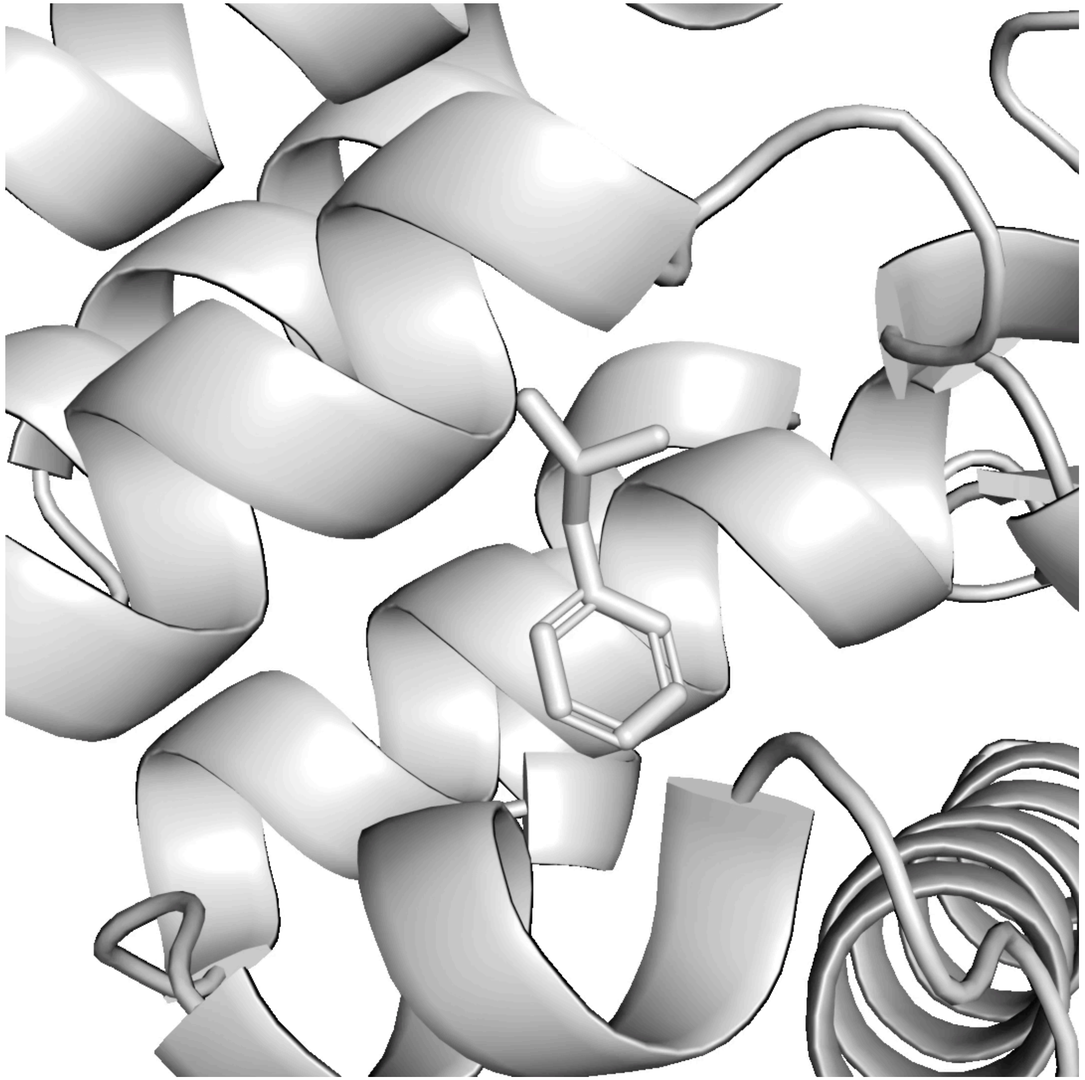
https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html



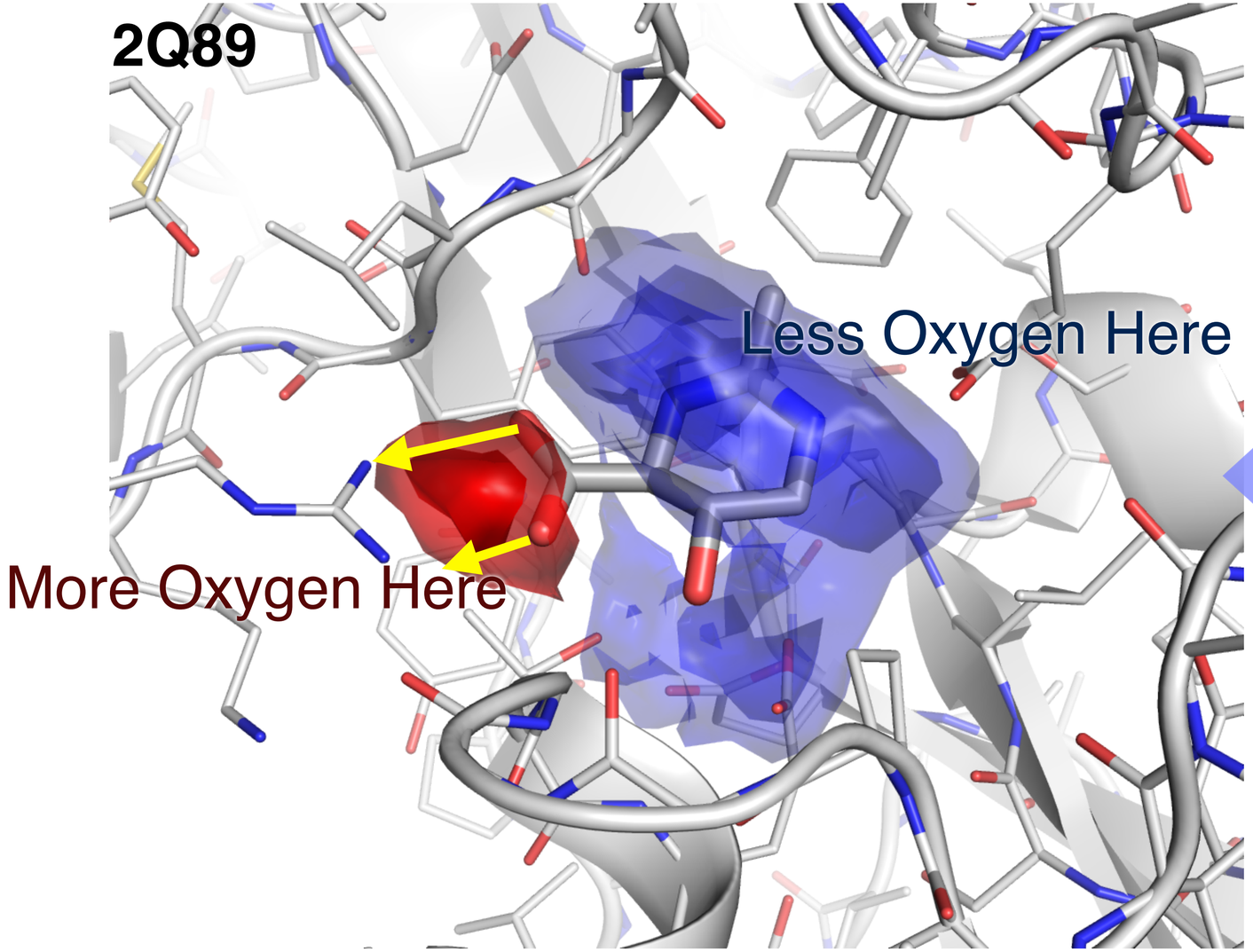
Deep Dreams of Molecules



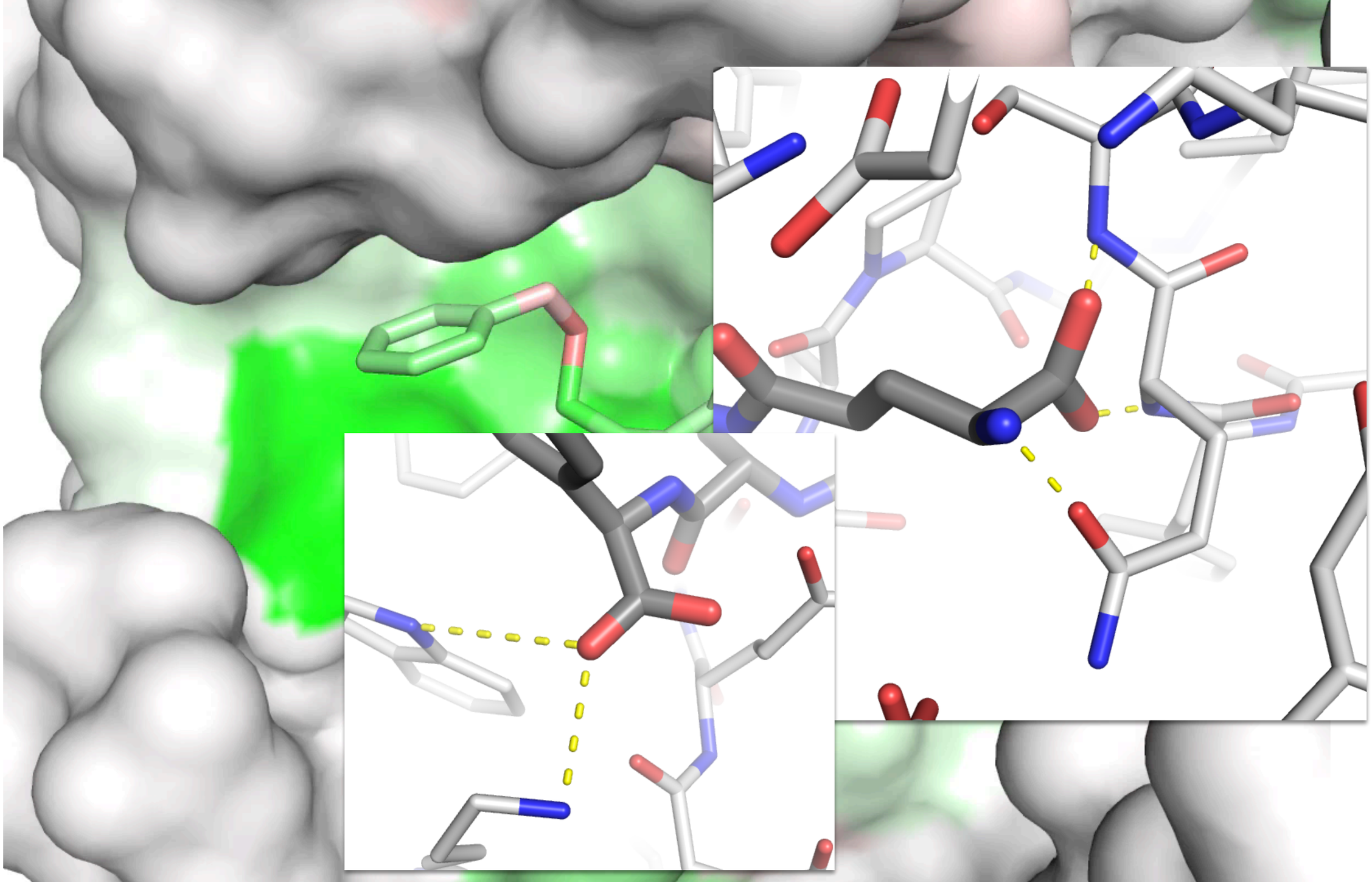


184l

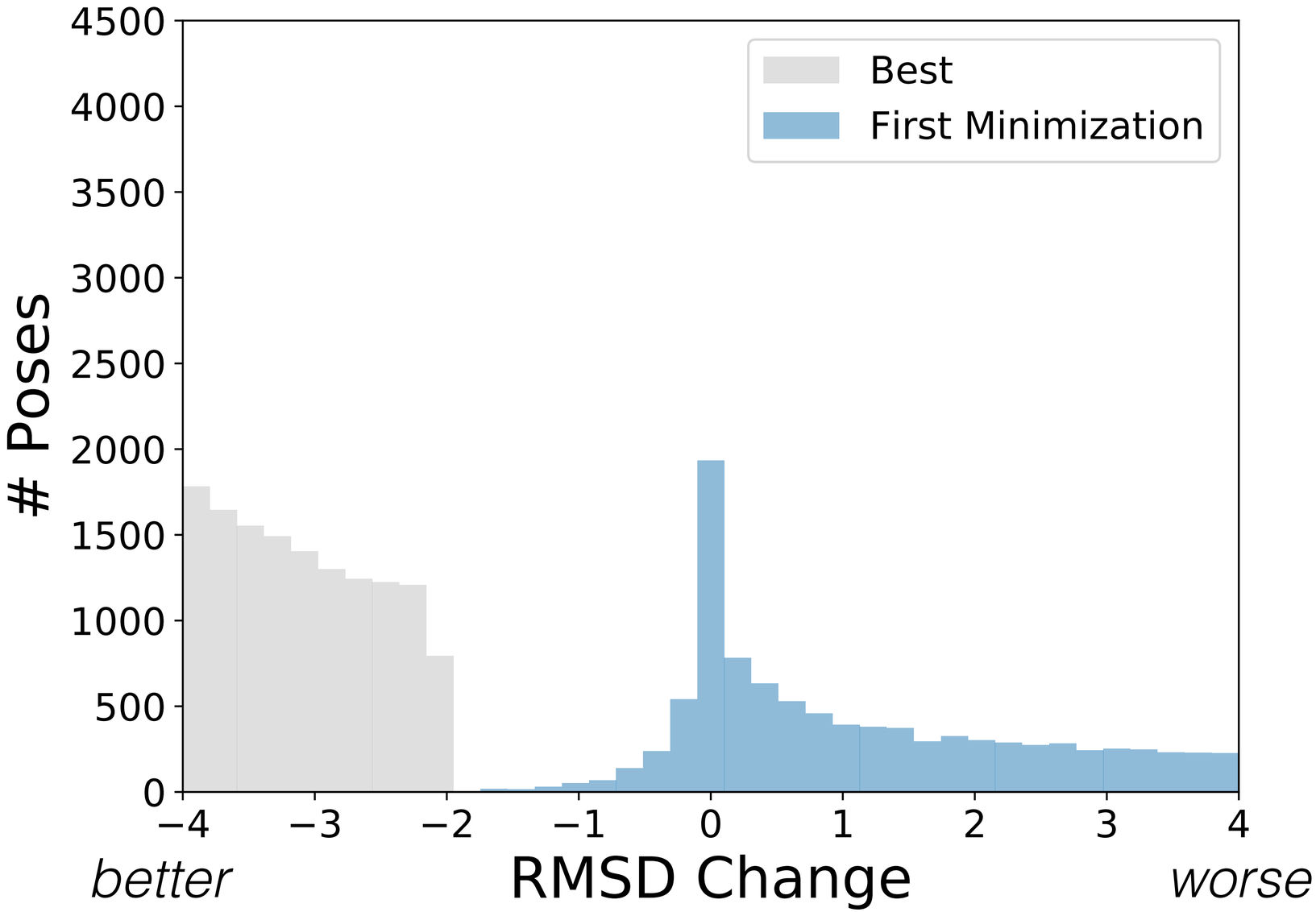
Gradients: Beyond Scoring



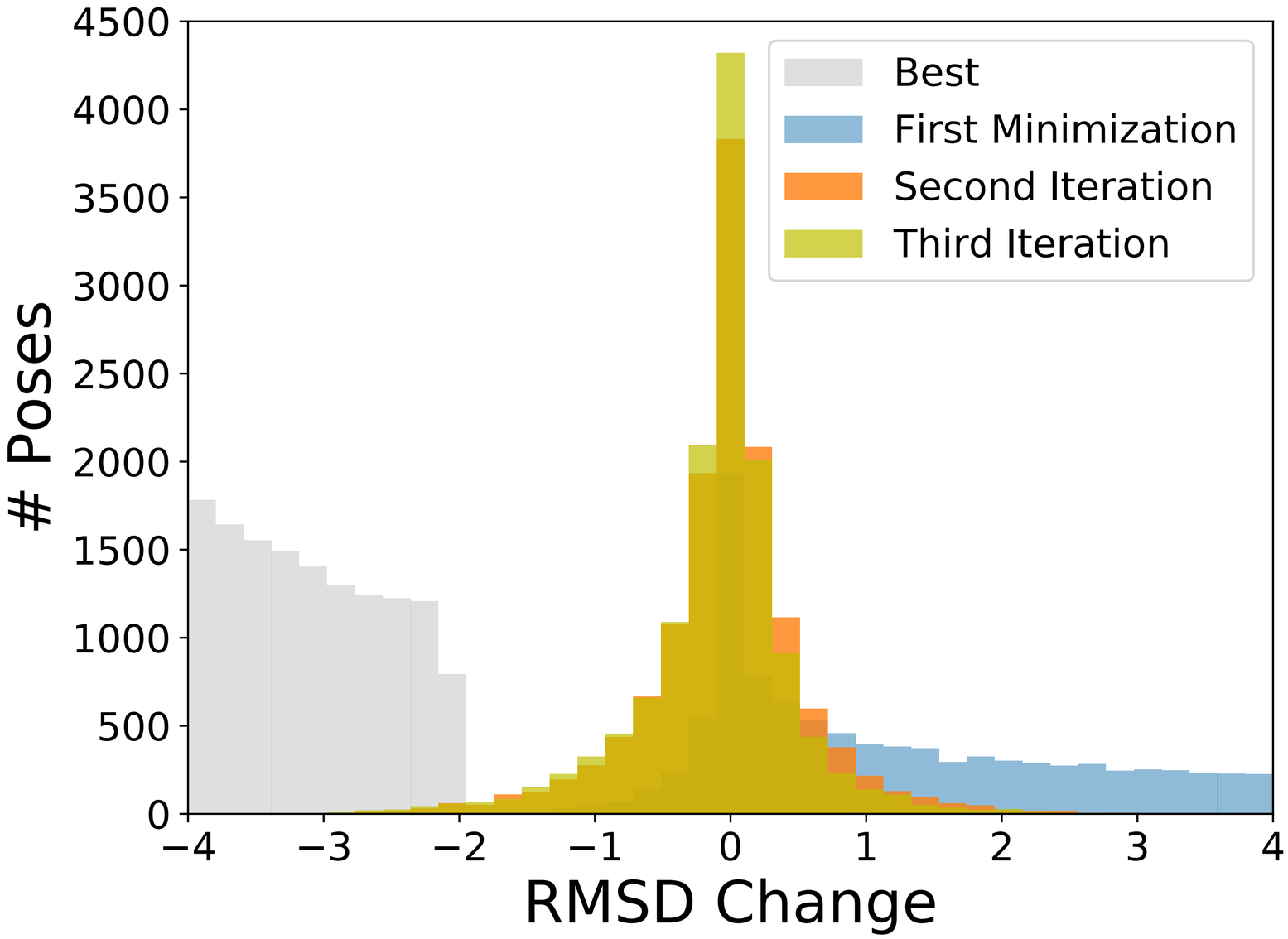
$$\frac{\partial L}{\partial A} = \sum_{i \in G_A} \frac{\partial L}{\partial G_i} \frac{\partial G_i}{\partial D} \frac{\partial D}{\partial A}$$

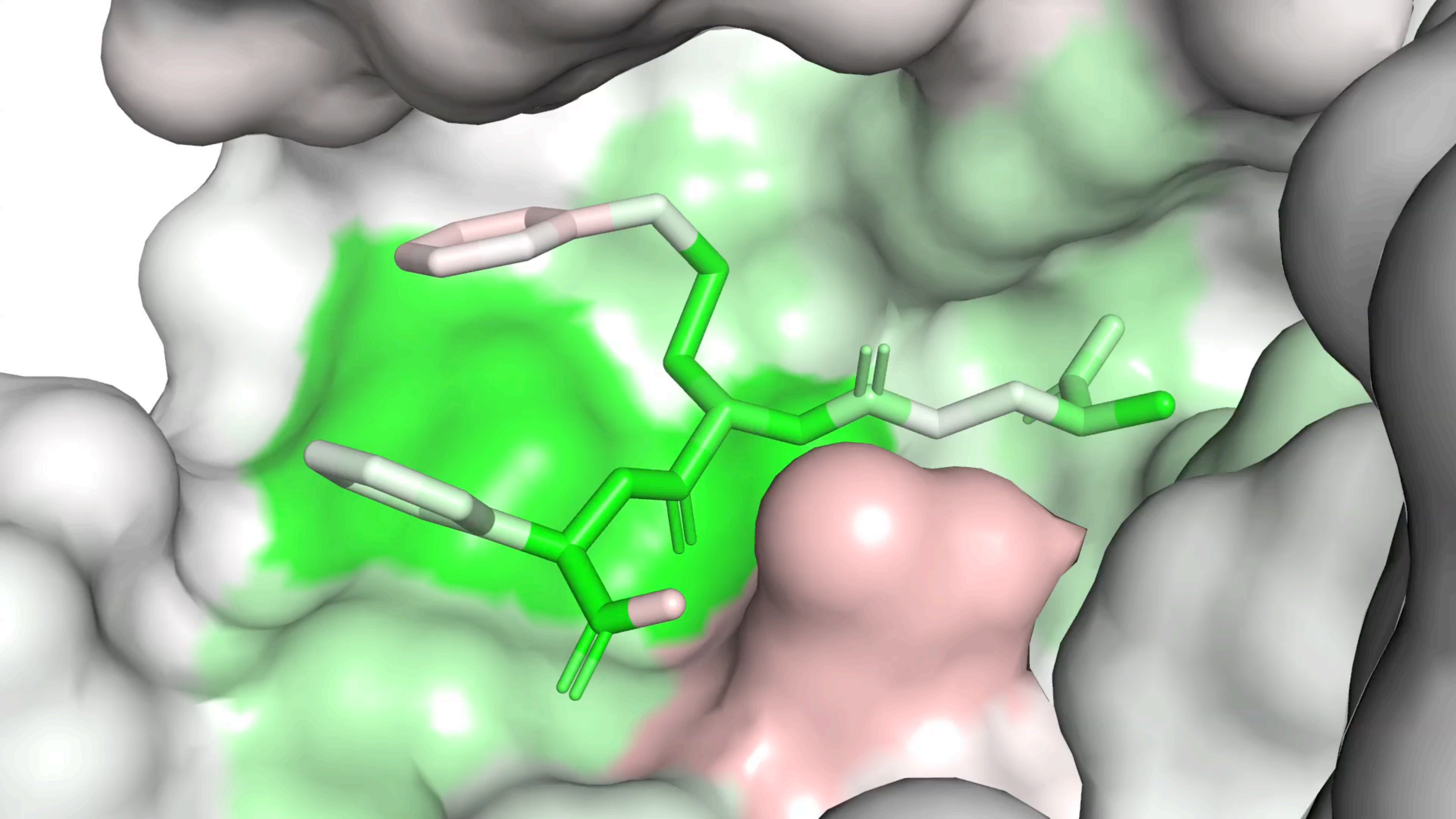


Minimizing Low RMSD Poses



Iterative Refinement





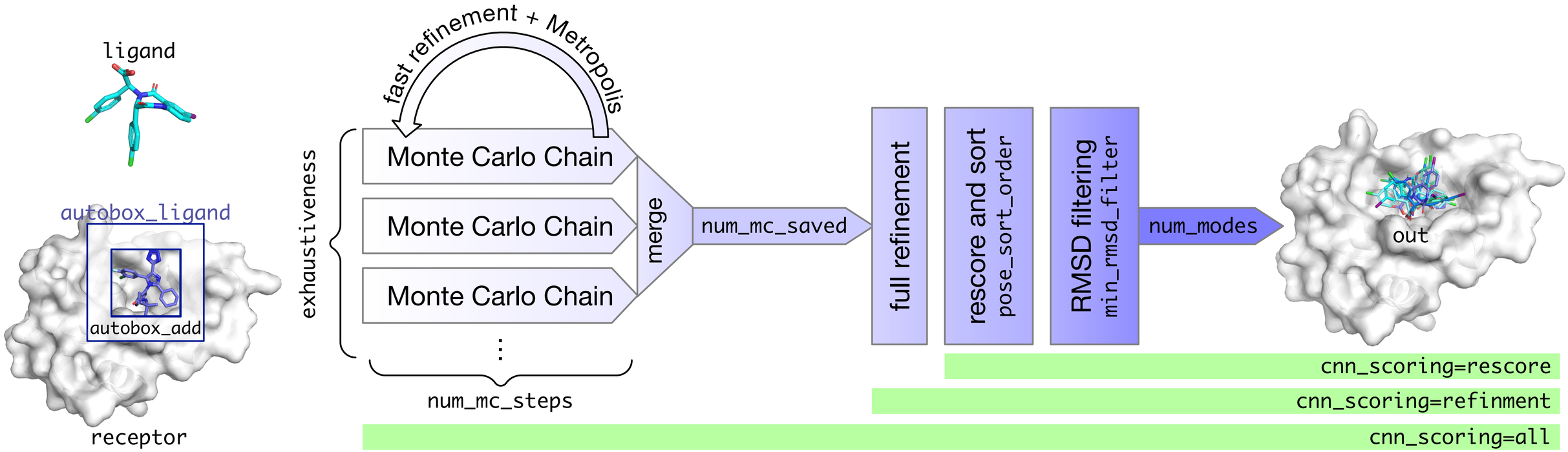
GNINA 1.0

https://github.com/gnina/gnina

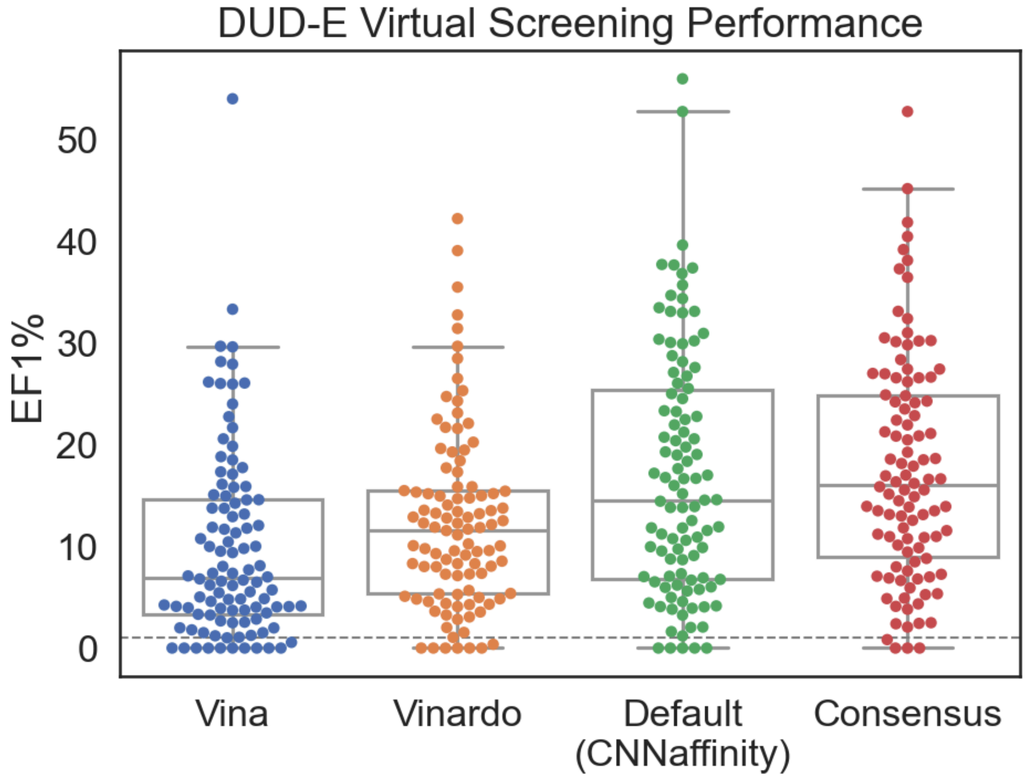
SOFTWARE Open Access

GNINA 1.0: molecular docking with deep learning

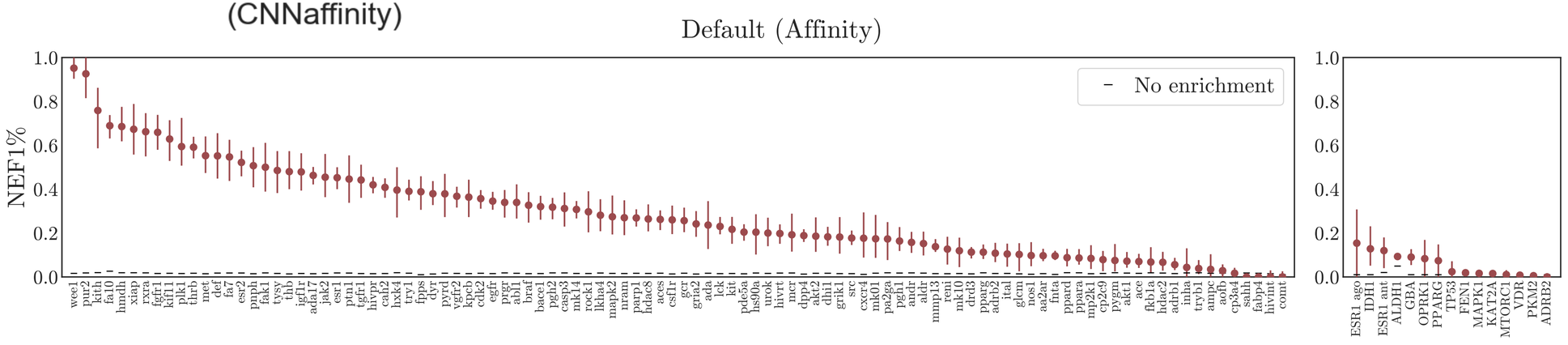
Andrew T. McNutt[1], Paul Francoeur[1], Rishal Aggarwal[2], Tomohide Masuda[1], Rocco Meli[3], Matthew Ragoza[1], Jocelyn Sunseri[1] and David Ryan Koes[1*]



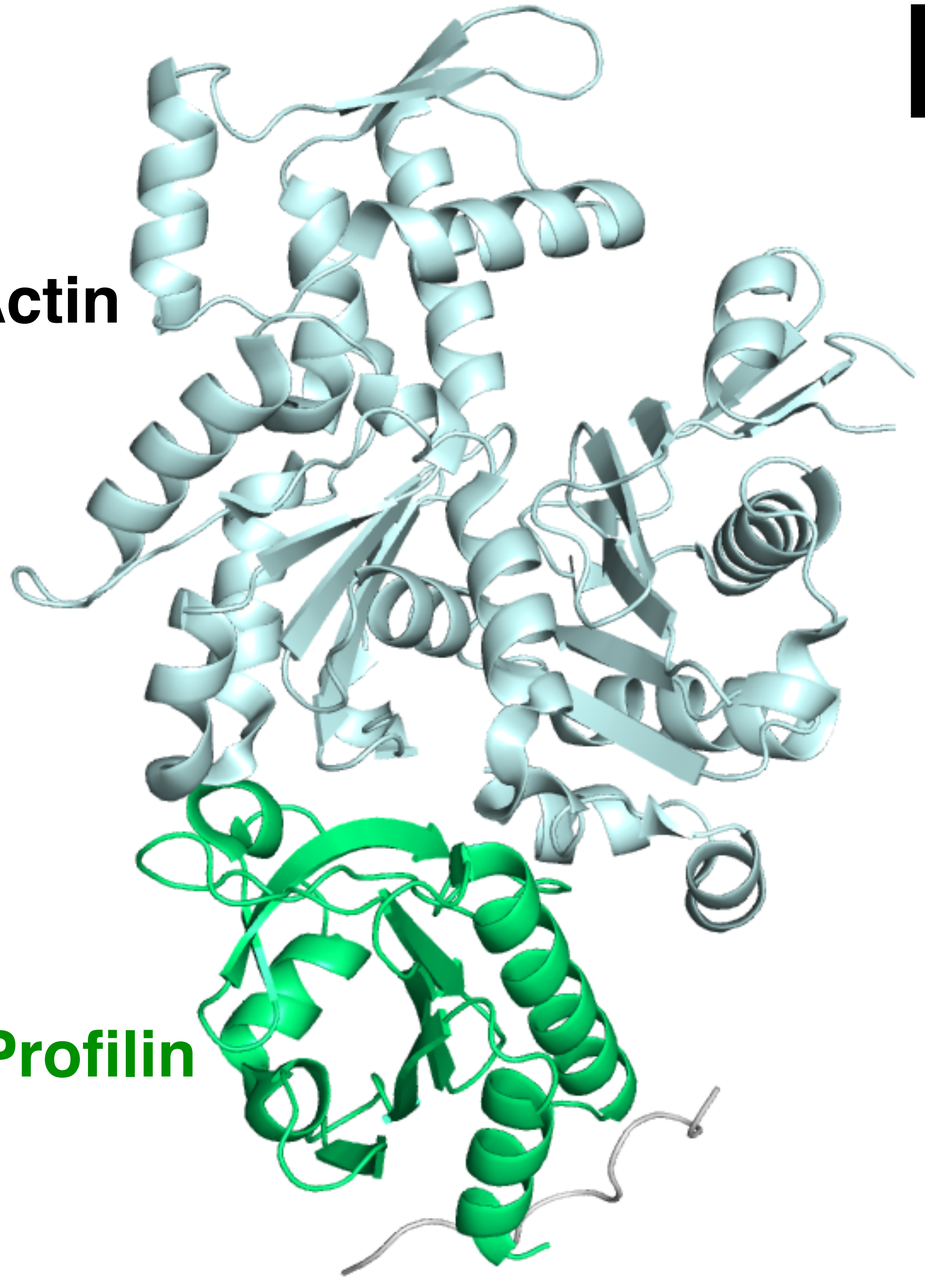
Virtual Screening



Model	DUD-E			LIT-PCBA		
	AUC	NEF1%	EF1%	AUC	NEF1%	EF1%
RFScore-4	0.683	0.0514	3.02	*0.6*	*0.013*	*1.28*
RFScore-VS	0.963	0.857	51.9	*0.542*	0.00733	0.733
Vina	0.745	0.118	7.05	*0.581*	*0.011*	1.1
Vinardo	0.764	0.187	11.4	*0.577*	0.0103	0.99
General (Affinity)	0.756	0.179	11.6	*0.579*	*0.037*	*2.06*
General (Pose)	0.702	0.156	10.3	0.498	*0.0147*	*1.3*
Dense (Affinity)	**0.795**	*0.27*	*17.7*	**0.616**	**0.037**	**2.58**
Dense (Pose)	*0.767*	**0.313**	**20.4**	0.514	*0.0238*	*1.81*
Default (Affinity)	*0.795*	*0.258*	*15.6*	*0.611*	*0.0238*	*1.88*
Default (Pose)	0.744	*0.241*	*15.8*	0.512	*0.0147*	*1.47*



Profilin



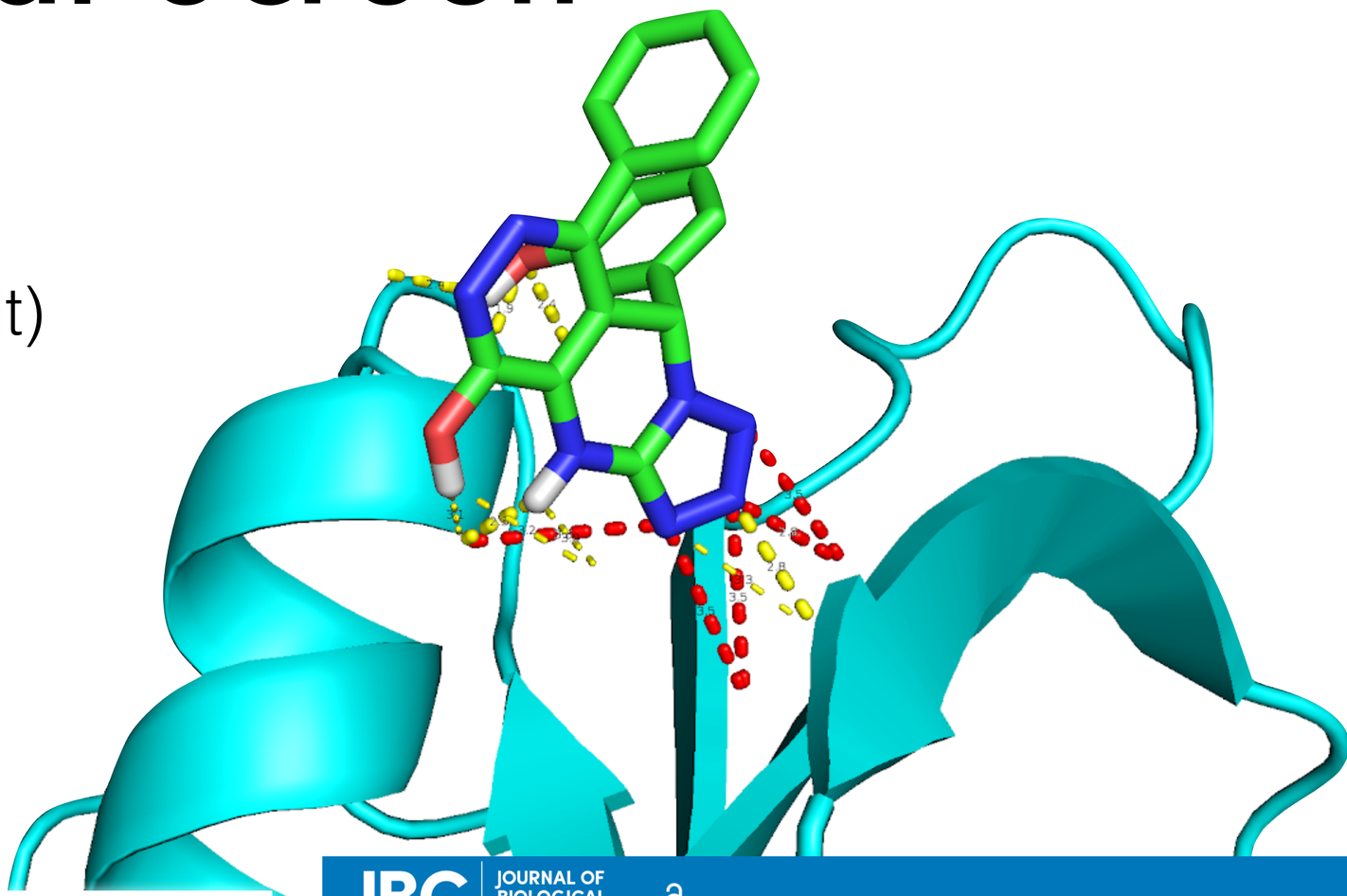
- Actin-binding protein
- Accelerates actin polymerization in presence of proline-rich proteins (e.g. formin, WASP, VASP)
- Sequesters actin otherwise

Dave Gau

Partha Roy

Virtual Screen

- Scaffold Hop Early Hit
- Pharmacophore screen (Pharmit)
- Ranked with Vina and CNN



JBC JOURNAL OF BIOLOGICAL CHEMISTRY

Structure-based virtual screening identifies a small-molecule inhibitor of the profilin 1–actin interaction

David Gau[‡1], Taber Lewis[§], Lee McDermott[¶], Peter Wipf[‡,§], David Koes[¶] and Partha Roy[‡,‖,**2]

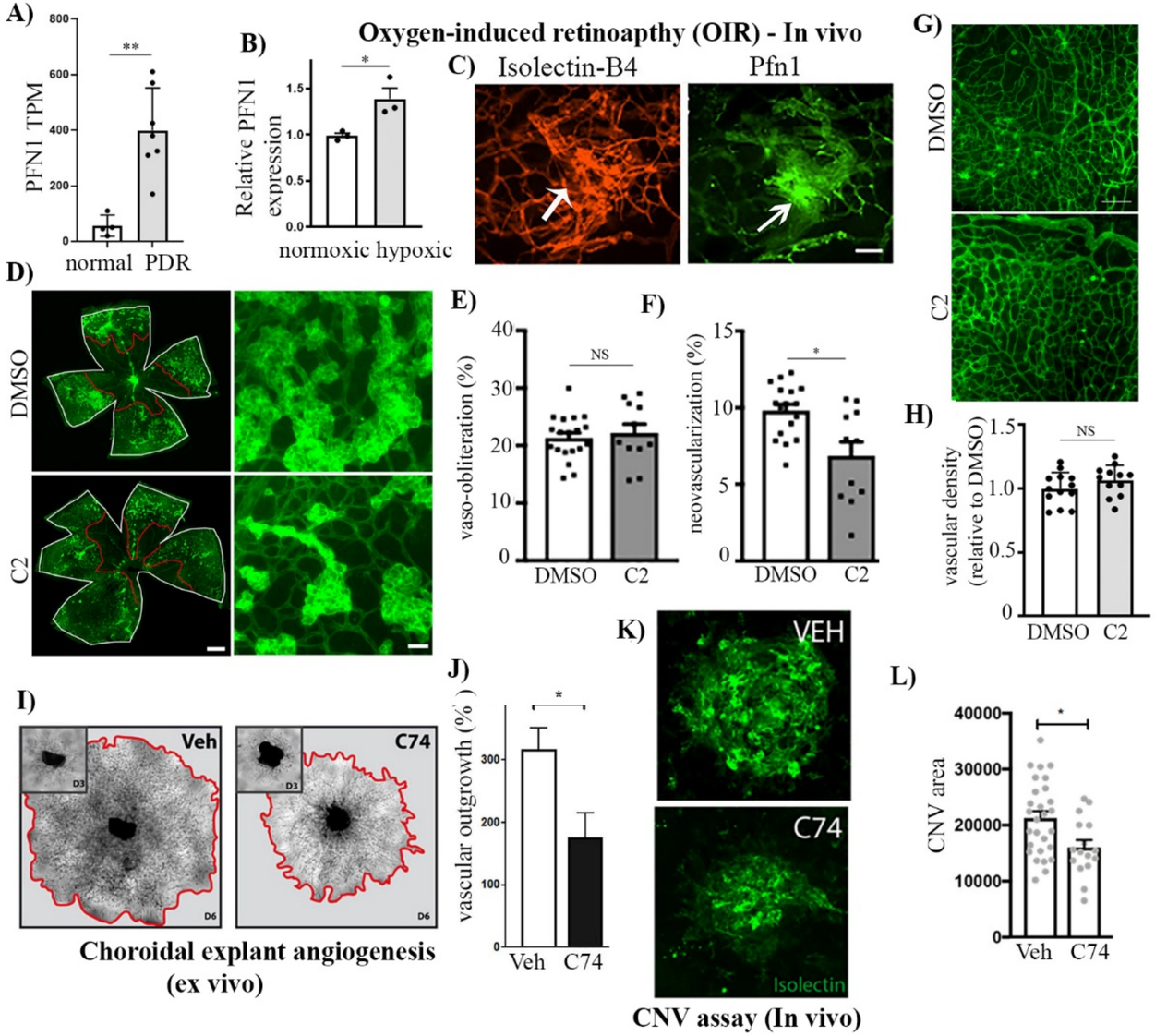
JBC JOURNAL OF BIOLOGICAL CHEMISTRY OPEN ACCESS

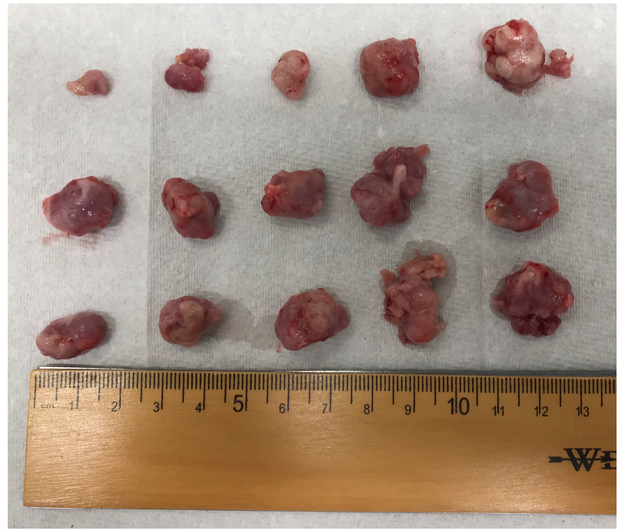
CELL BIOLOGY | VOLUME 295, ISSUE 46, P15636-15649, NOVEMBER 2020

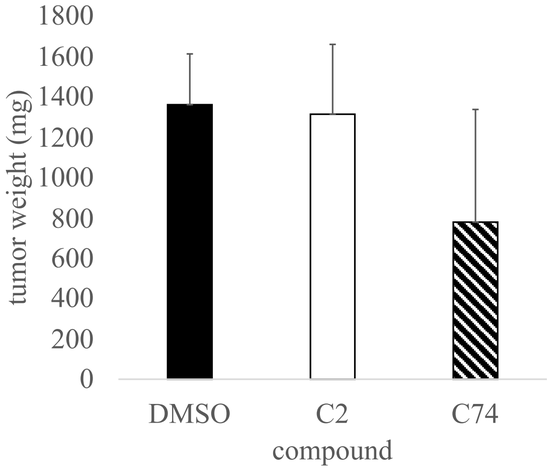
Actin-binding protein profilin1 promotes aggressiveness of clear-cell renal cell carcinoma cells

Abigail Allen • David Gau • Paul Francoeur • ... David Koes • Walter J. Storkus • Partha Roy Show all authors

Results

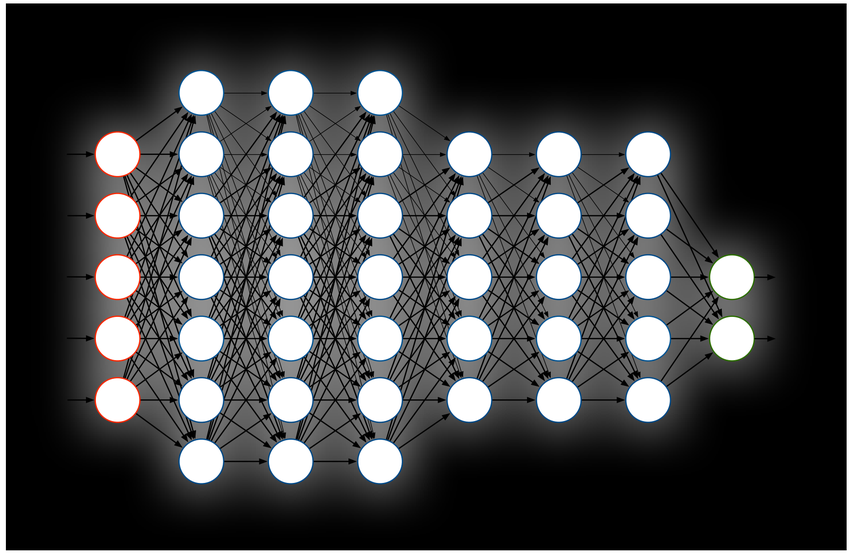

A)
PFN1 TPM
normal PDR
B)
Relative PFN1 expression
normoxic hypoxic
Oxygen-induced retinoapthy (OIR) - In vivo
C) Isolectin-B4
Pfn1
G)
DMSO
C2
D)
DMSO
C2
E)
vaso-obliteration (%)
NS
DMSO C2
F)
neovascularization (%)
*
DMSO C2
H)
vascular density (relative to DMSO)
NS
DMSO C2
I)
Veh
C74
D3
D6
Choroidal explant angiogenesis (ex vivo)
J)
vascular outgrowth (%)
*
Veh C74
K)
VEH
C74
Isolectin
CNV assay (In vivo)
L)
CNV area
*
Veh C74


C74
DMSO
C2

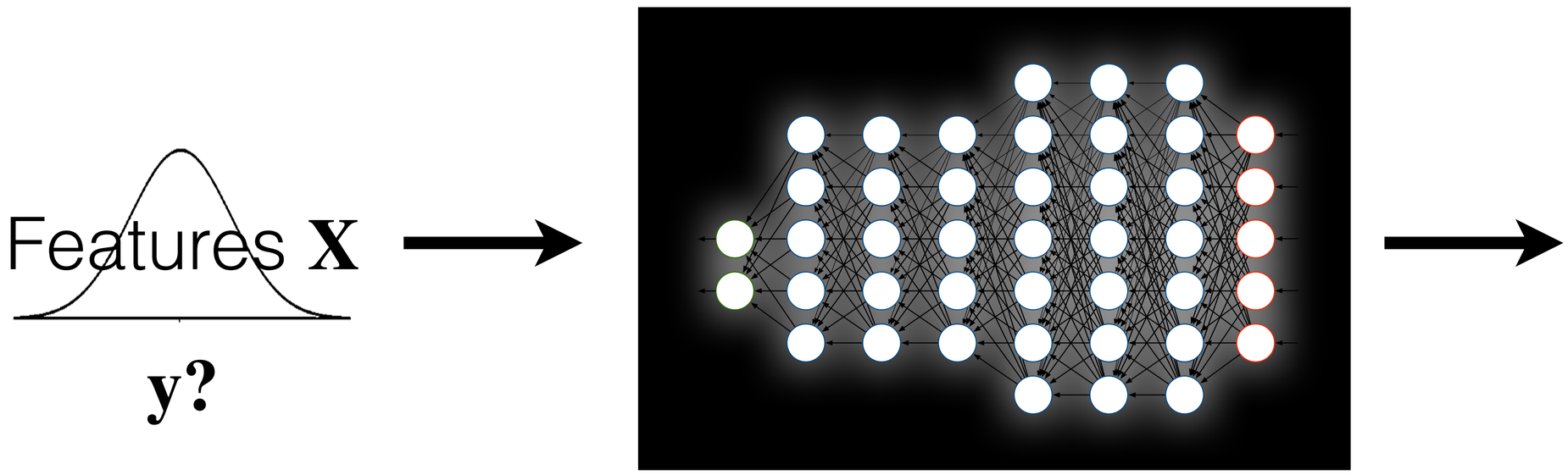

tumor weight (mg)
DMSO
C2
C74
compound

Generative Modeling

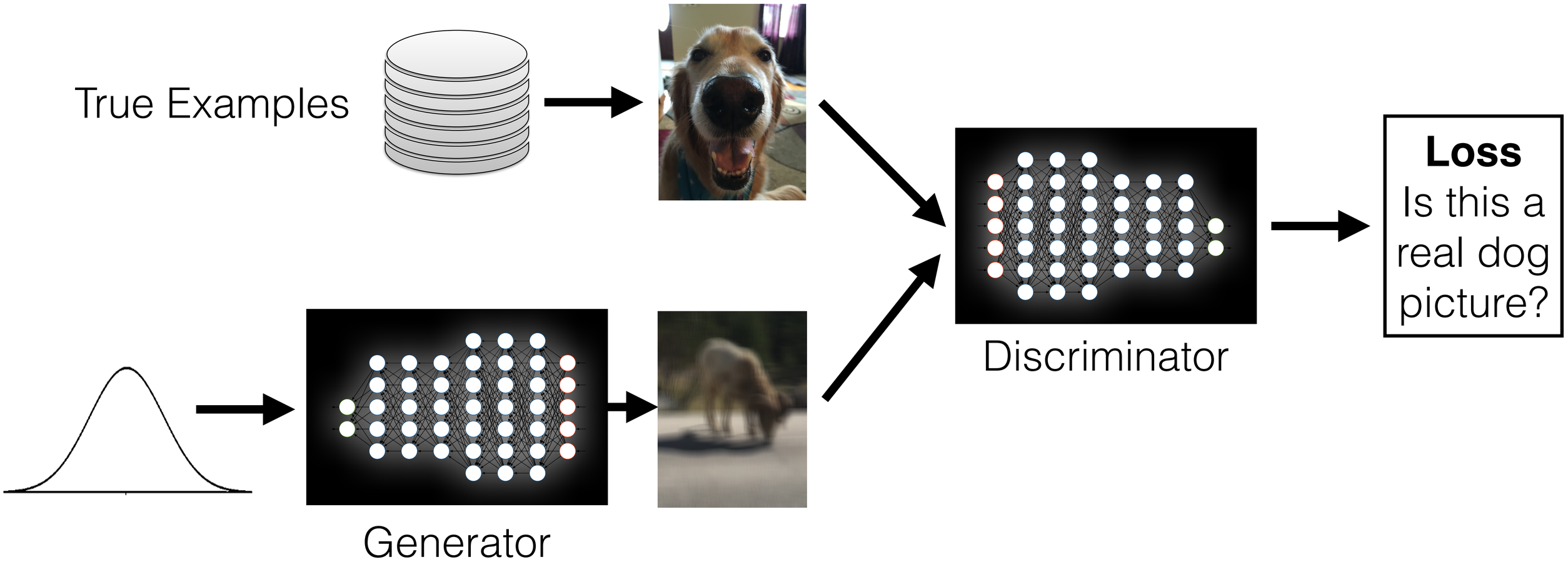
Discriminative Model

Features $\mathbf{X}$  Prediction $\mathbf{y}$

Generative Model



Generative Adversarial Networks



Generative Adversarial Networks

2014

2015

2016

2017

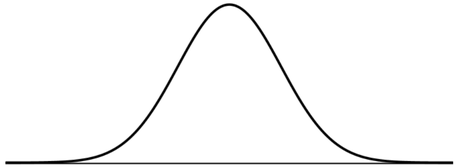
2018

Ian Goodfellow @goodfellow_ian · 2h
4.5 years of GAN progress on face generation. arxiv.org/abs/1406.2661 arxiv.org/abs/1511.06434 arxiv.org/abs/1606.07536 arxiv.org/abs/1710.10196 arxiv.org/abs/1812.04948

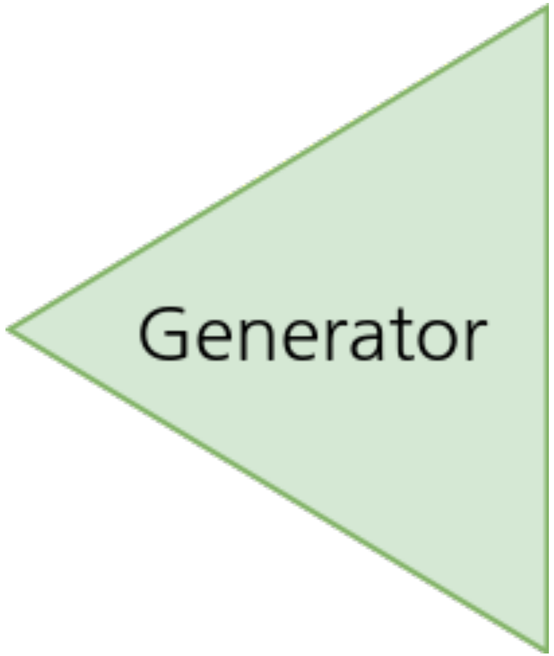
https://thispersondoesnotexist.com

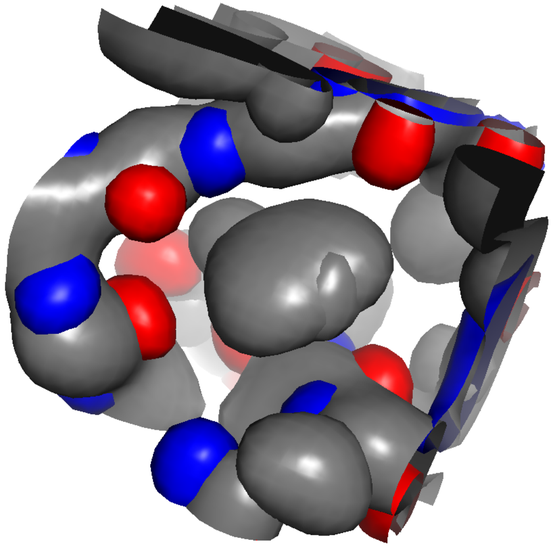
Generative Models

Generative models approximate a data distribution directly. They can map samples from one distribution (noise or input data) to realistic samples from an output distribution of interest.



noise sample

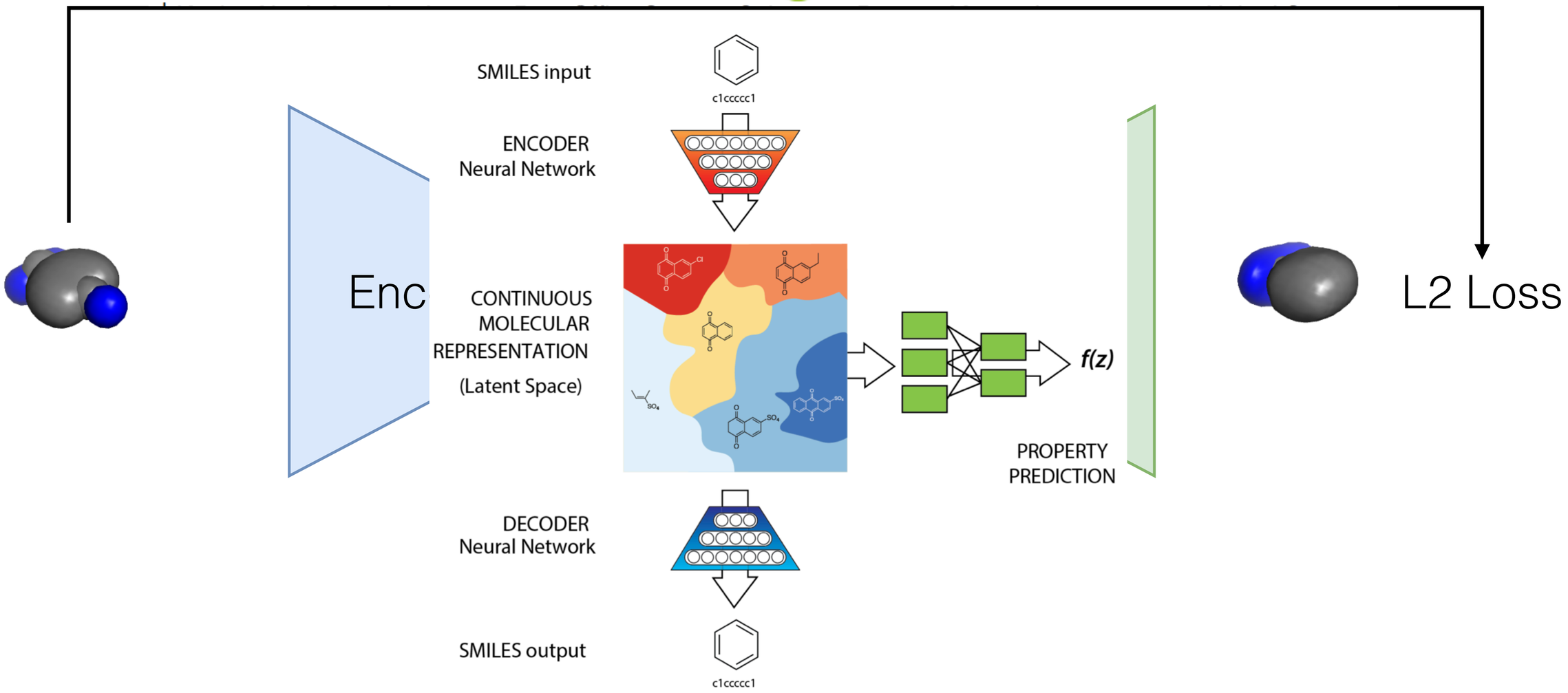




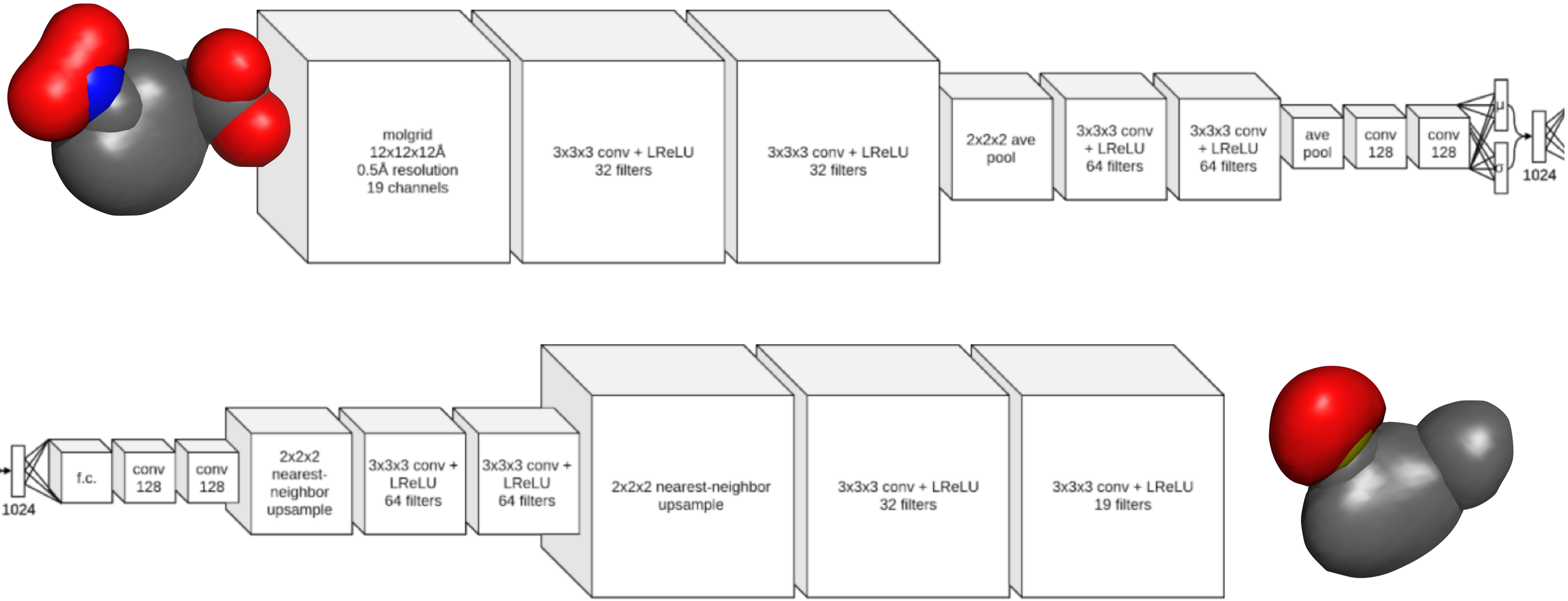
generated receptor & ligand grid

Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules

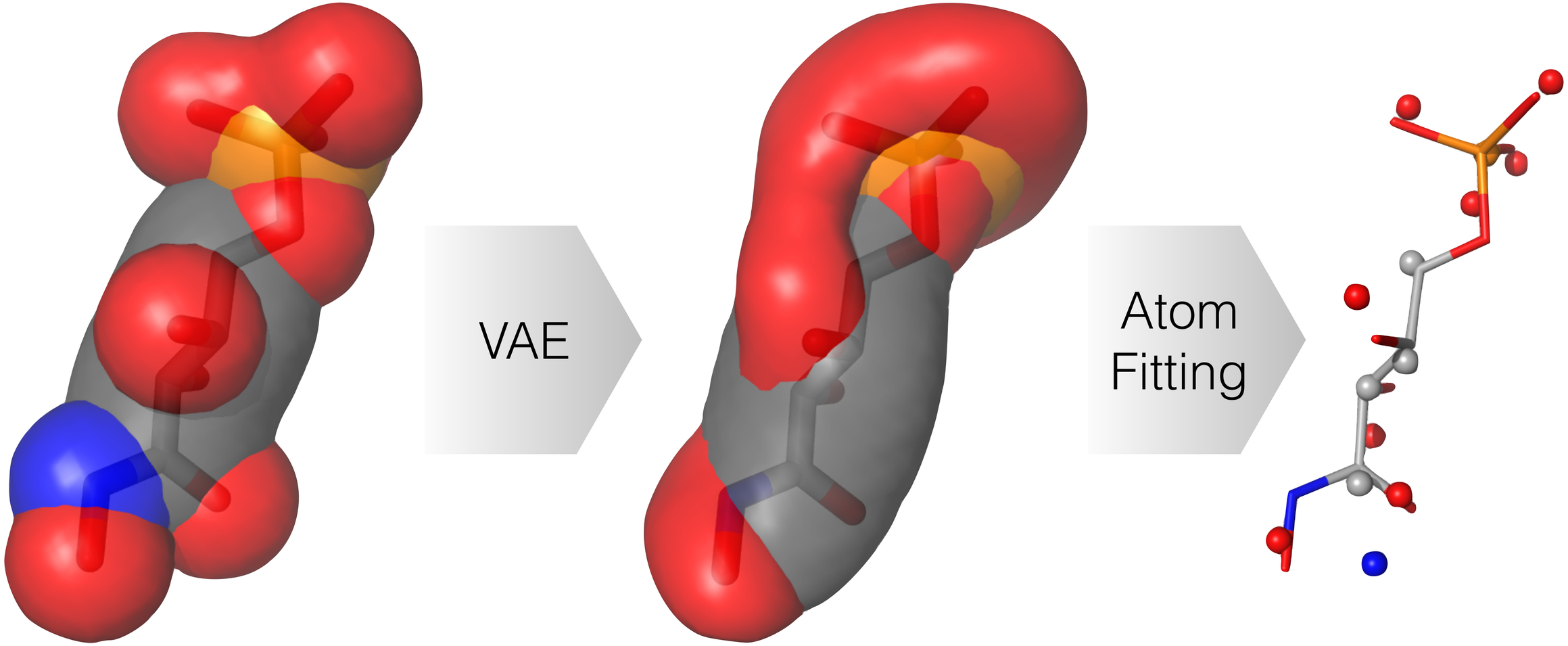
Rafael Gómez-Bombarelli[†#], Jennifer N. Wei[‡#], David Duvenaud[¶#], José Miguel Hernández-Lobato[§#], Benjamín Sánchez-Lengeling[‡], Dennis Sheberla[‡], Jorge Aguilera-Iparraguirre[†], Timothy D. Hirzel[†], Ryan P. Adams[∇⊥], and Alán Aspuru-Guzik[*‡⊥]



Model


molgrid
12x12x12Å
0.5Å resolution
19 channels
3x3x3 conv + LReLU
32 filters
3x3x3 conv + LReLU
32 filters
2x2x2 ave
pool
3x3x3 conv
+ LReLU
64 filters
3x3x3 conv
+ LReLU
64 filters
ave
pool
conv
128
conv
128
μ
σ
1024
1024
f.c.
conv
128
conv
128
2x2x2
nearest-
neighbor
upsample
3x3x3 conv +
LReLU
64 filters
3x3x3 conv +
LReLU
64 filters
2x2x2 nearest-neighbor
upsample
3x3x3 conv + LReLU
32 filters
3x3x3 conv + LReLU
19 filters

Variational Autoencoding Examples

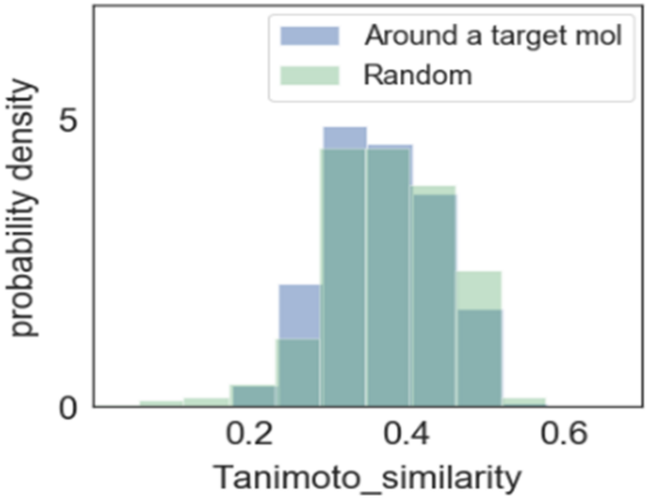


2BES

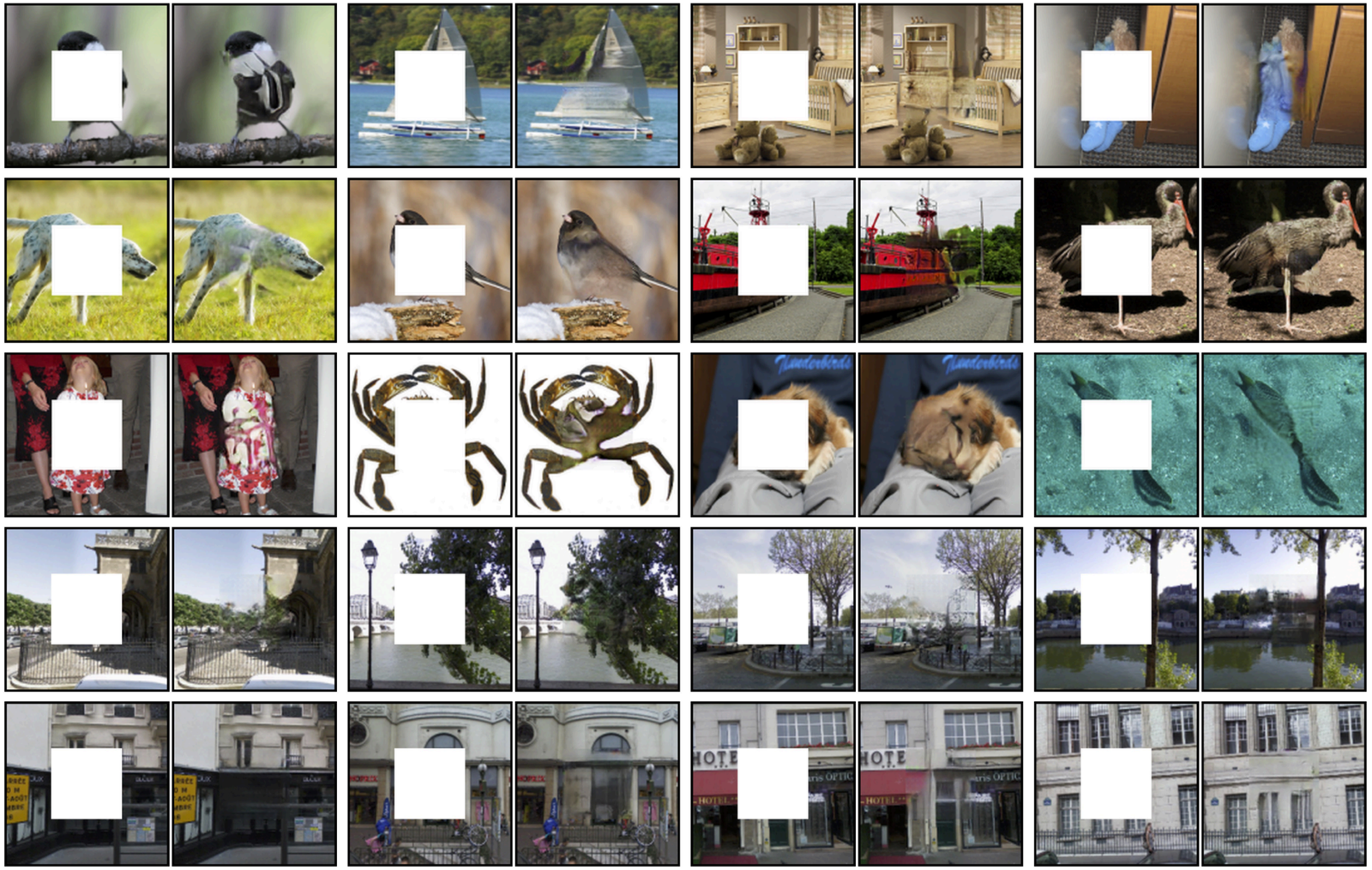
Variational Autoencoding Examples

Around a target molecule (3 samples per each mol)			Random	
mol_1	mol_2	mol_3	mol_1,2,3	mol_4,5,6
(0.45, 0.80, 0.25)	(0.40, 0.82, 0.74)	(0.31, 0.56, 0.87)		
(0.39, 0.80, 0.51)	(0.43, 0.82, 0.48)	(0.35, 0.56, 0.44)		
(0.37, 0.80, 0.34)	(0.33, 0.82, 0.52)	(0.27, 0.56, 0.78)		

pink : true ligand green : generated molecule rigid body alignment

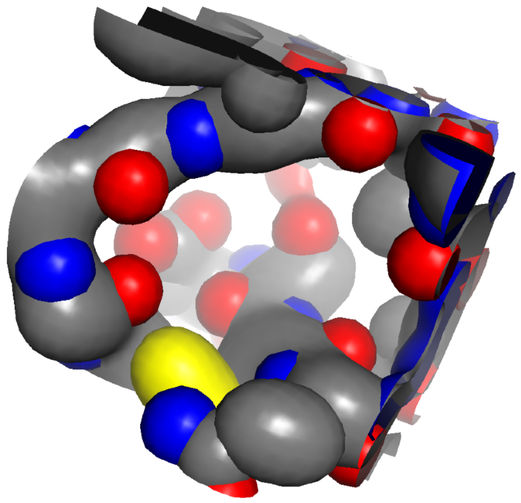


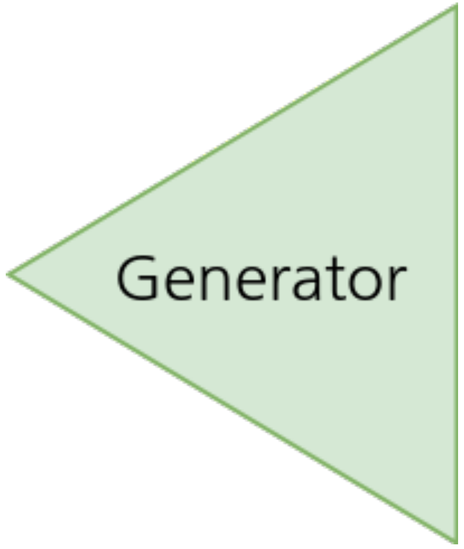
Context Encoding

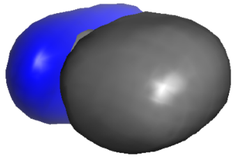


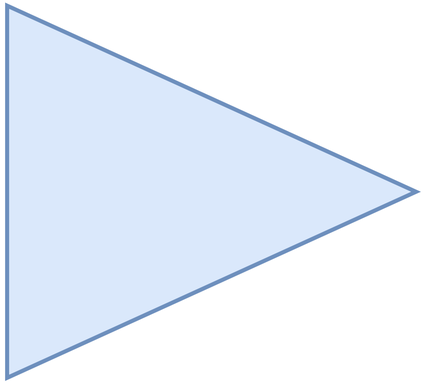
http://people.eecs.berkeley.edu/~pathak/context_encoder/

Context Encoding





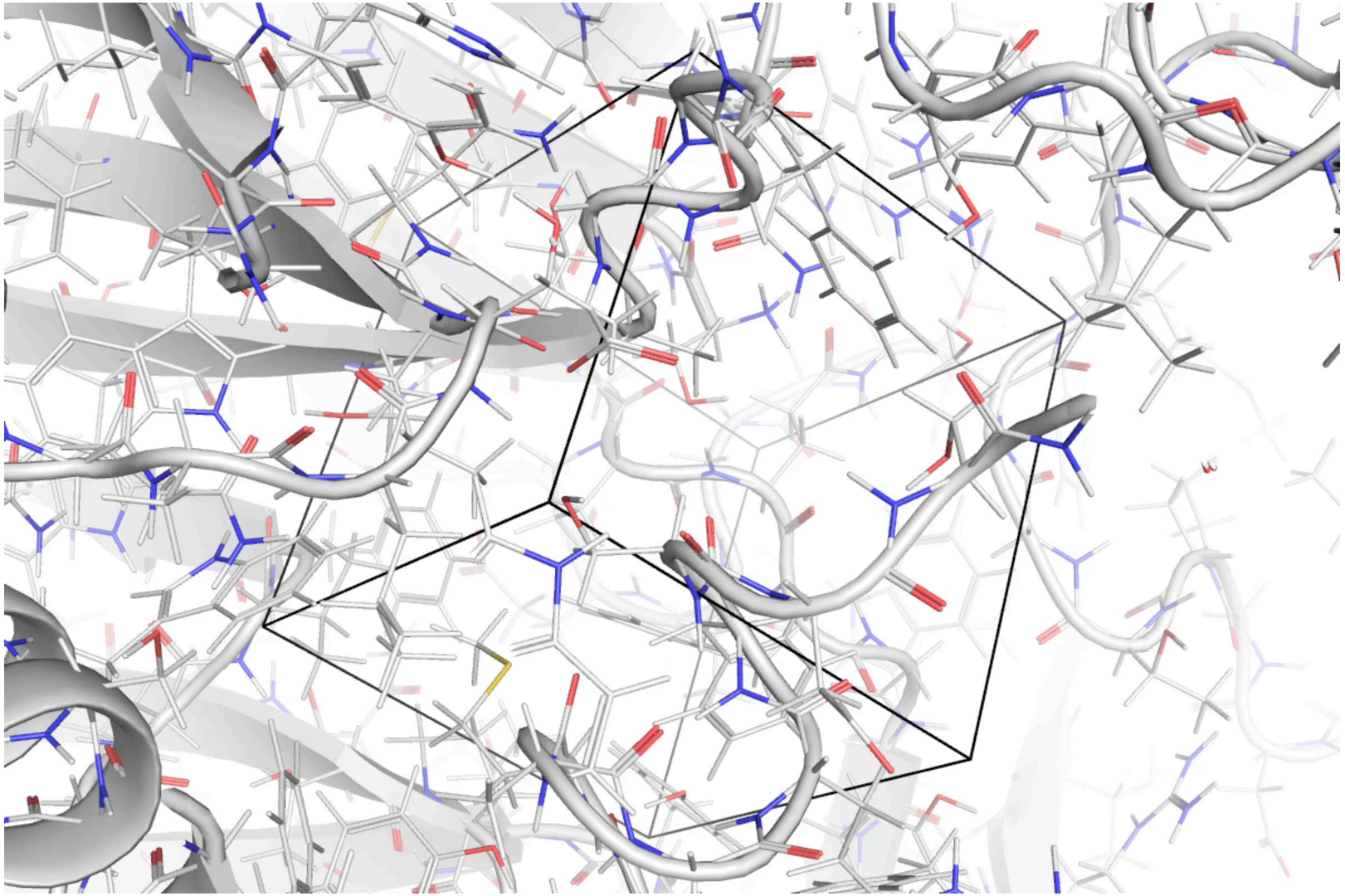




receptor grid

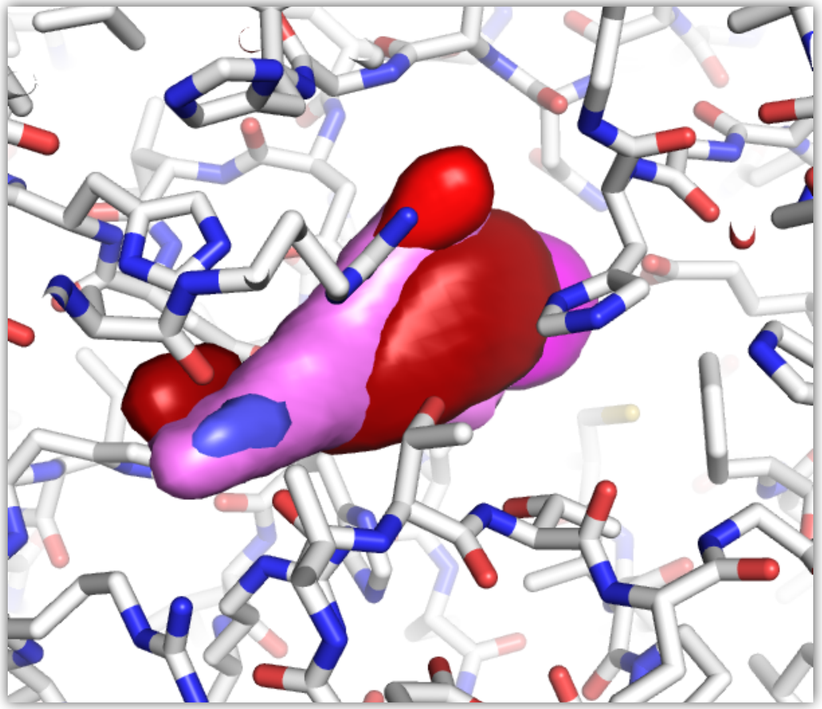
generated ligand grid

Conditioning on the Receptor

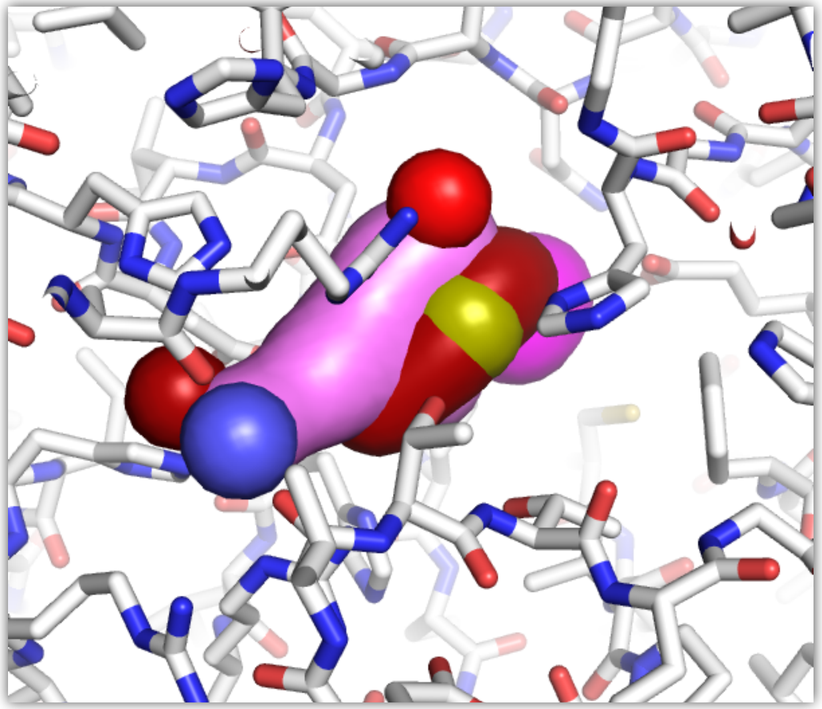


Context Encoding with Fully Convolutional Network

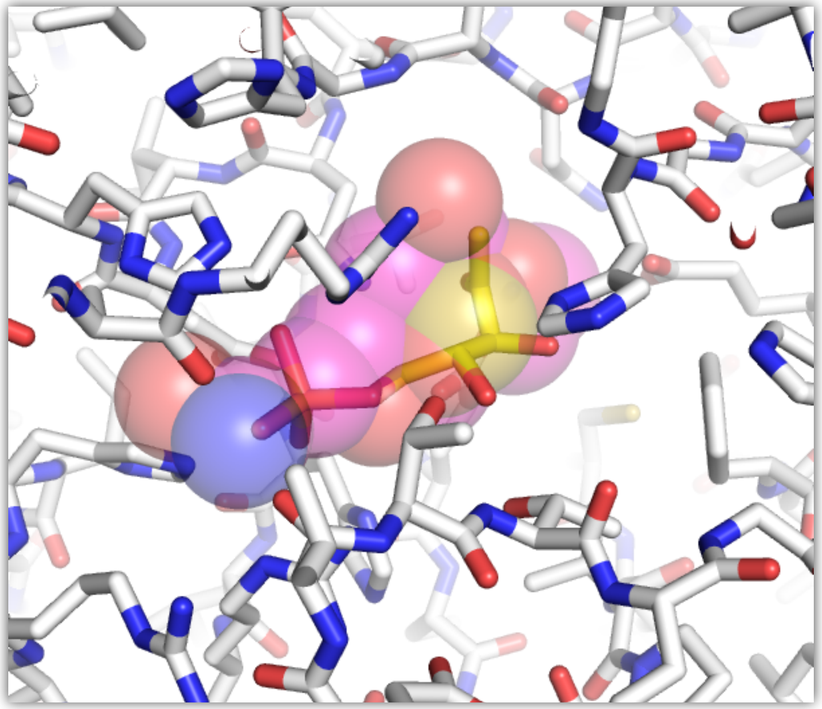
1m5w



Generated



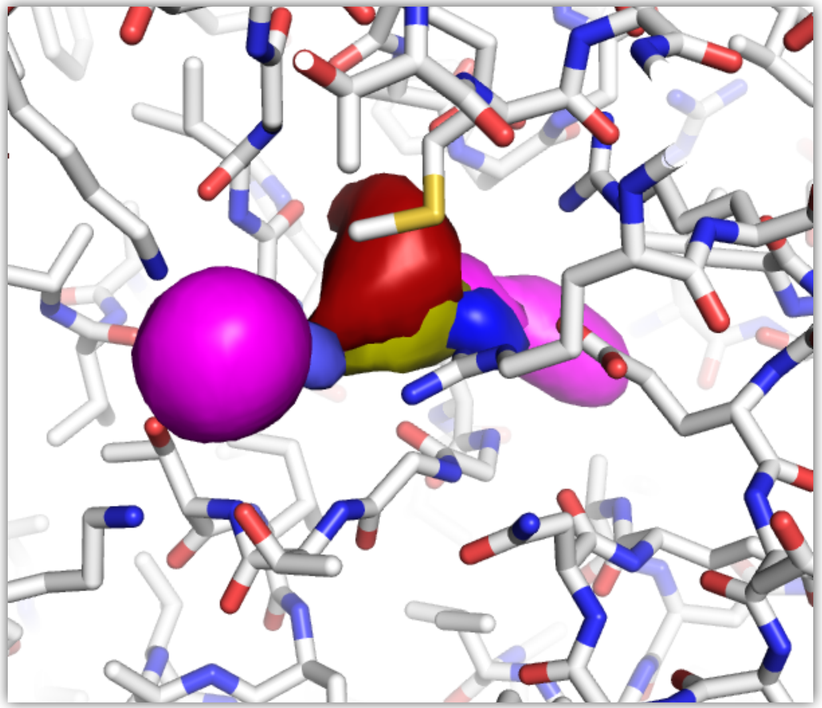
Fit Densities



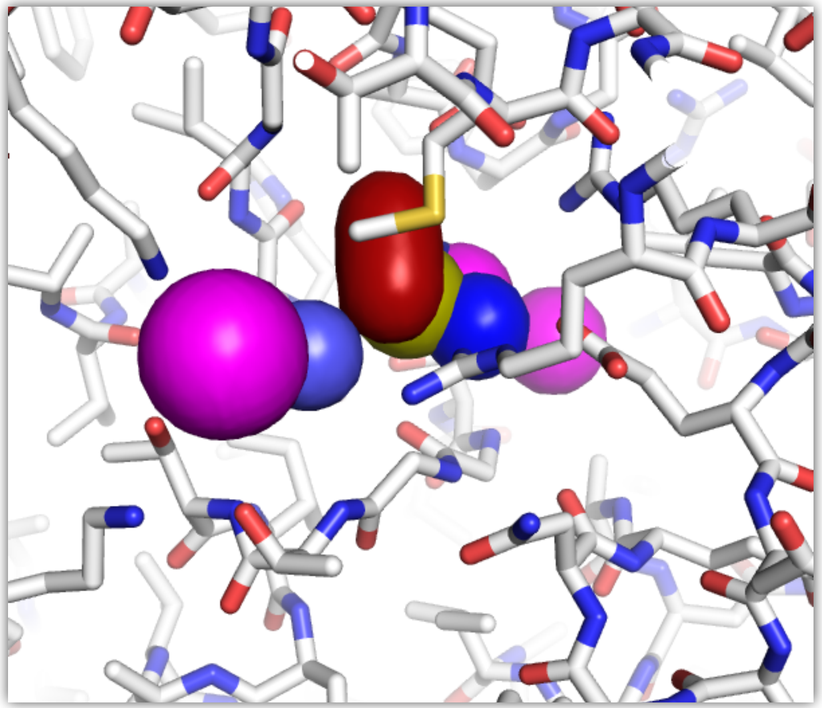
Fit Atoms

Context Encoding with Fully Convolutional Network

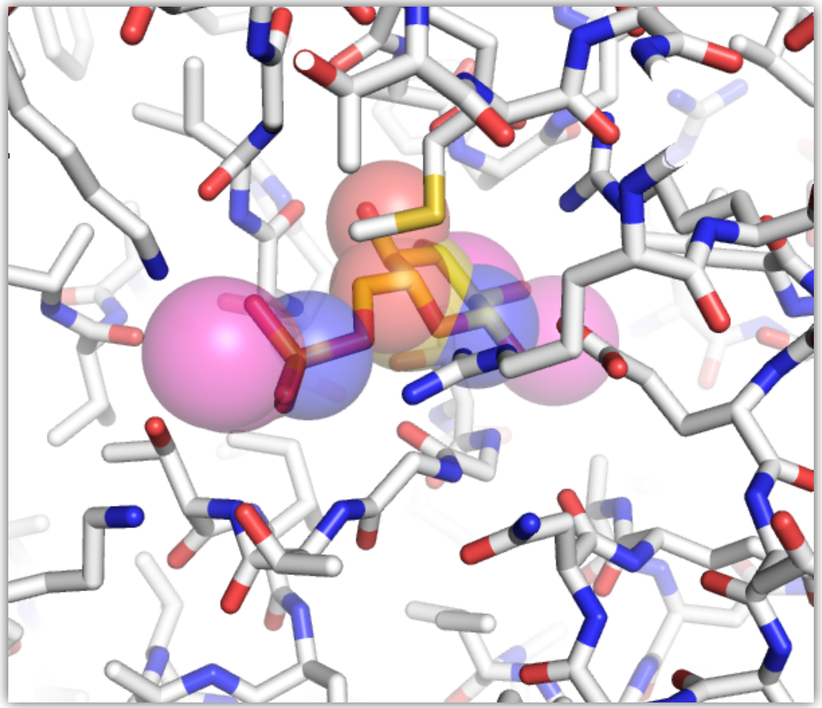
3bxg



Generated



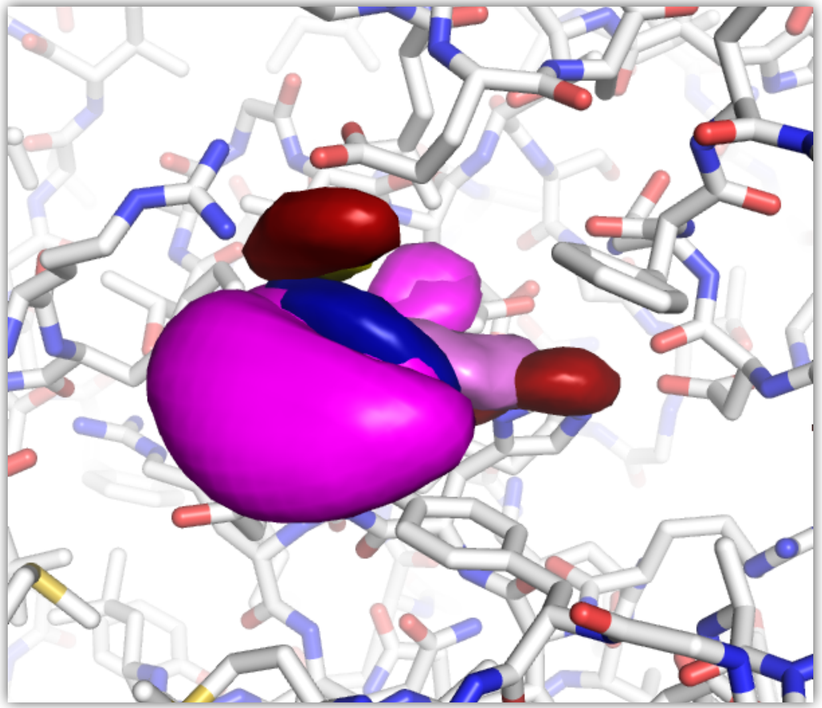
Fit Densities



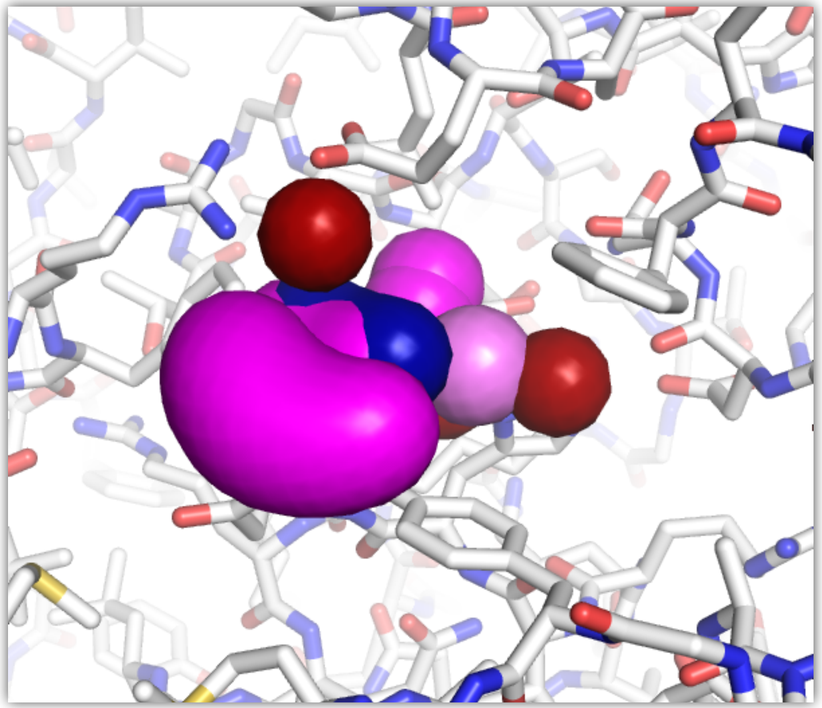
Fit Atoms

Context Encoding with Fully Convolutional Network

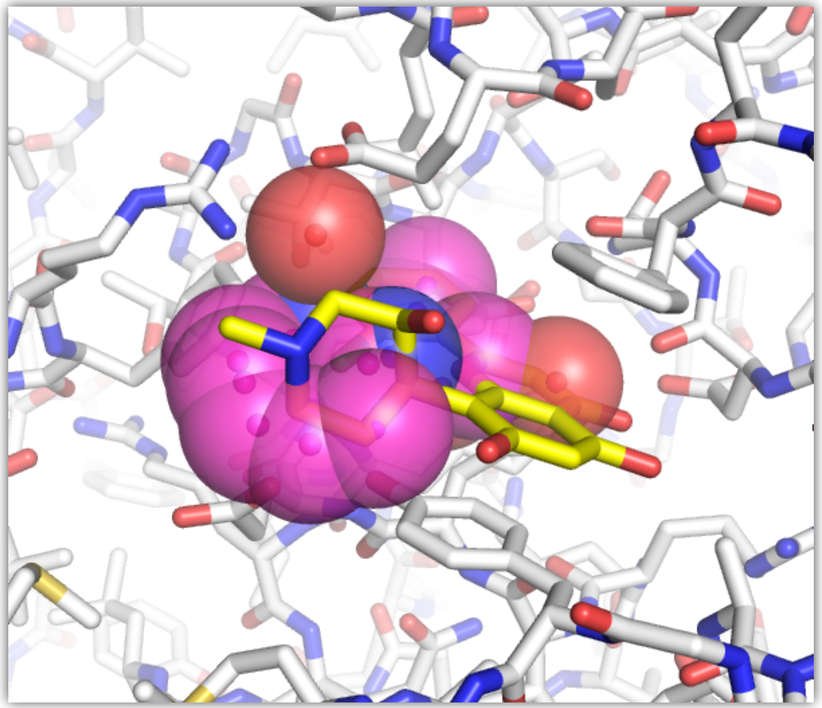
3ebp



Generated



Fit Densities



Fit Atoms

Acknowledgements