Docking Deeply: Molecular Docking with Deep Learning Potentials.

David Koes

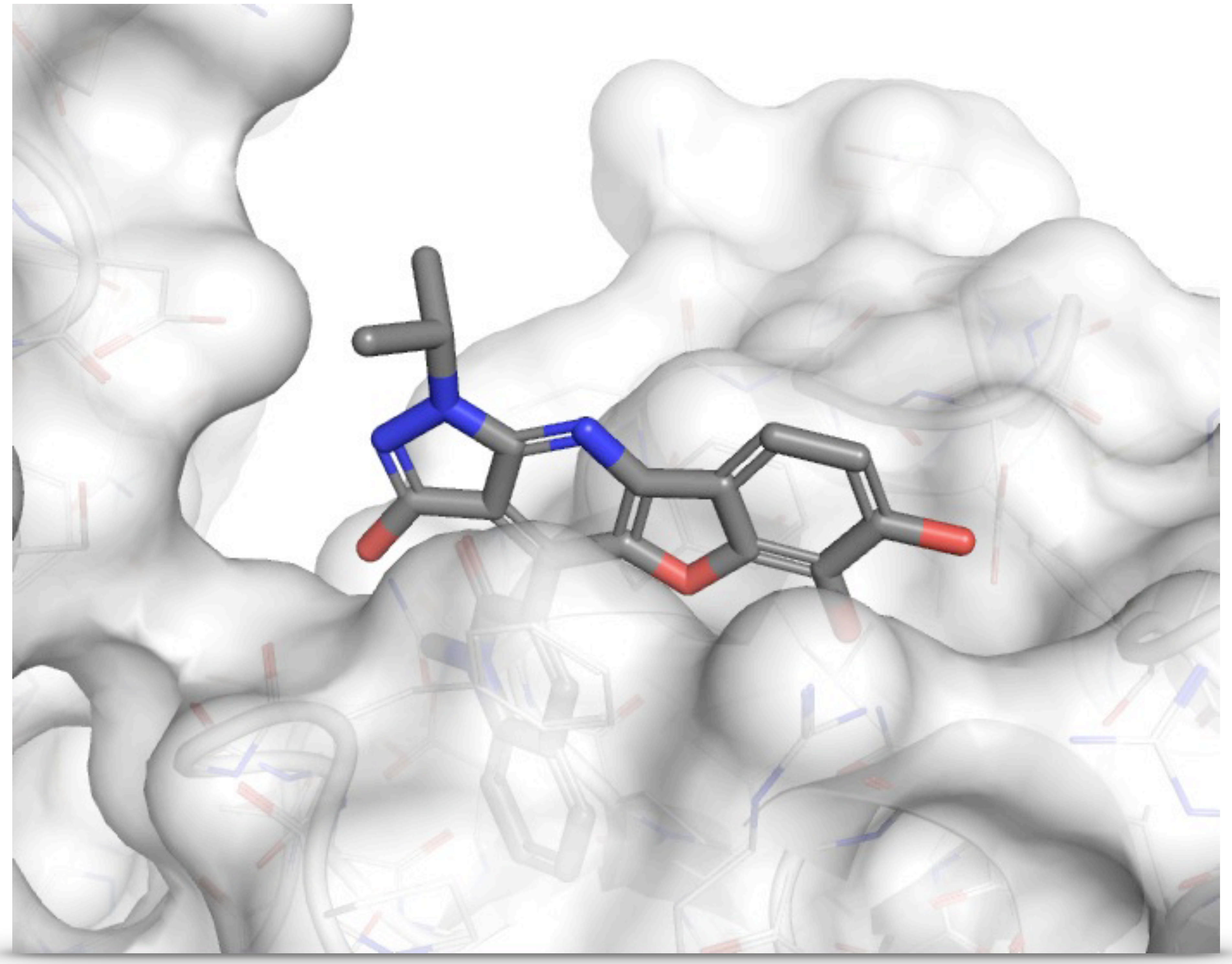
@david_koes

2018 Workshop on Free Energy Methods, Kinetics and Markov State Models in Drug Design
Cambridge, MA
May 16, 2018

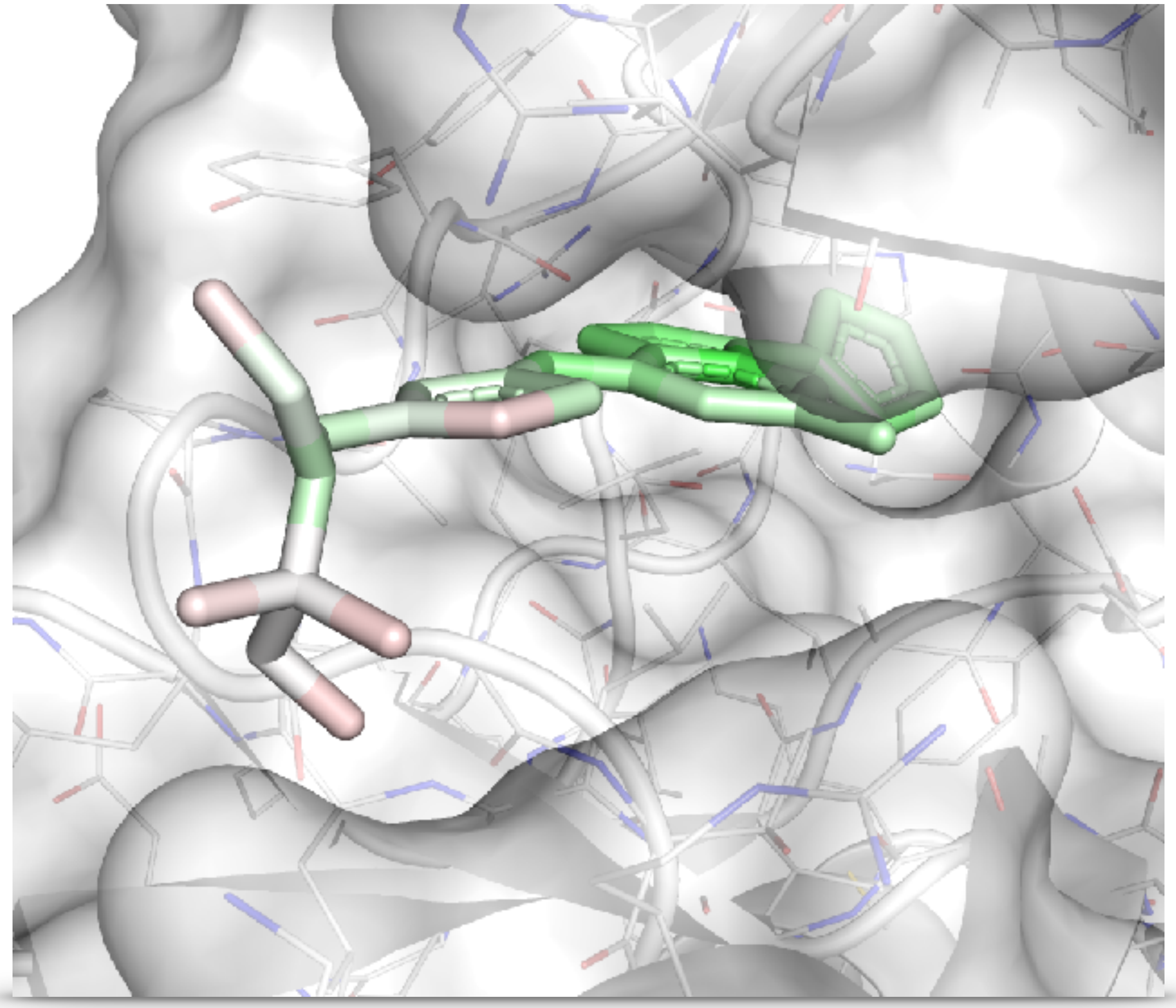
UNIVERSITY OF PITTSBURGH
VERITAS VIRTUS 1787

Structure Based Drug Design

Virtual Screening



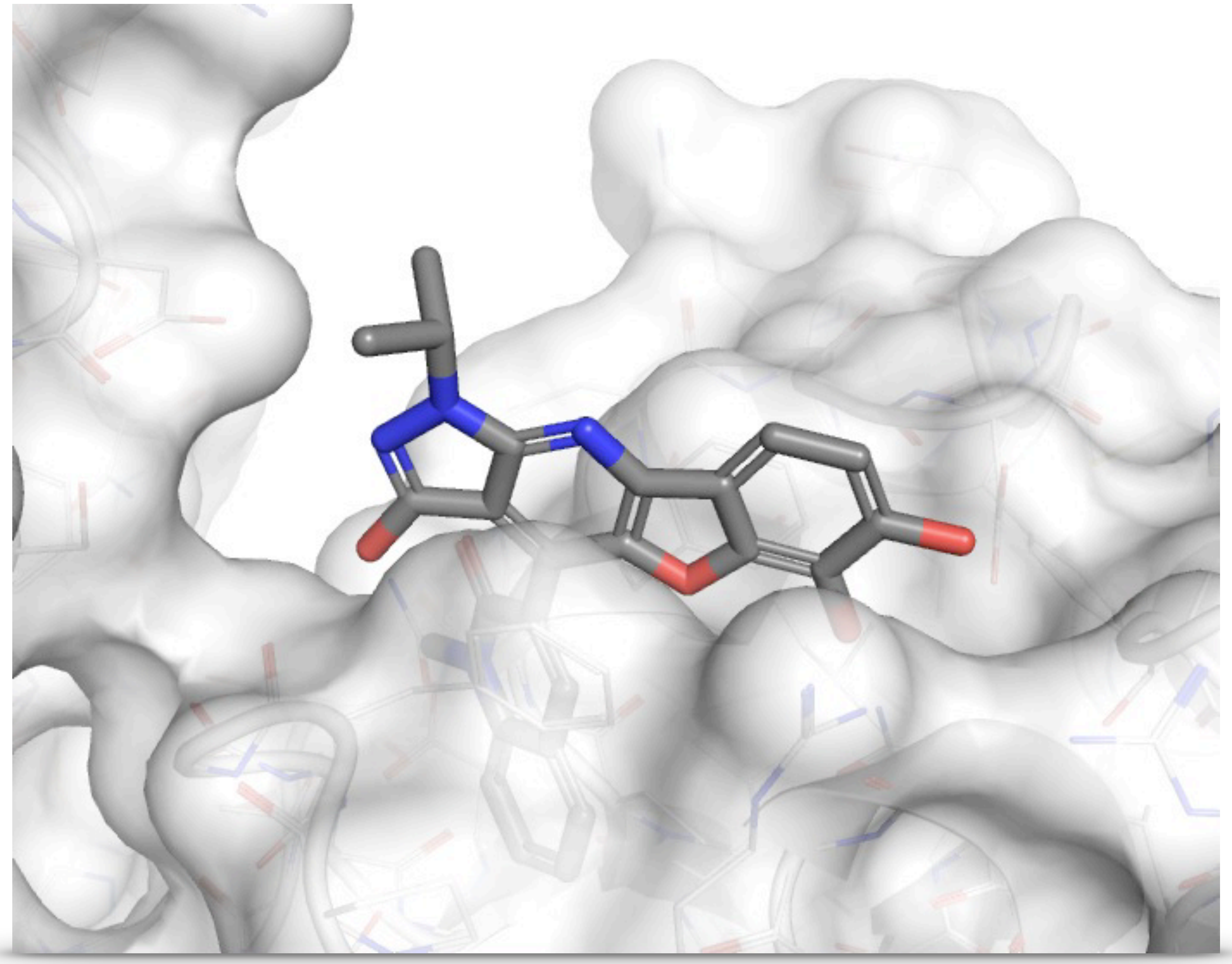
Lead Optimization



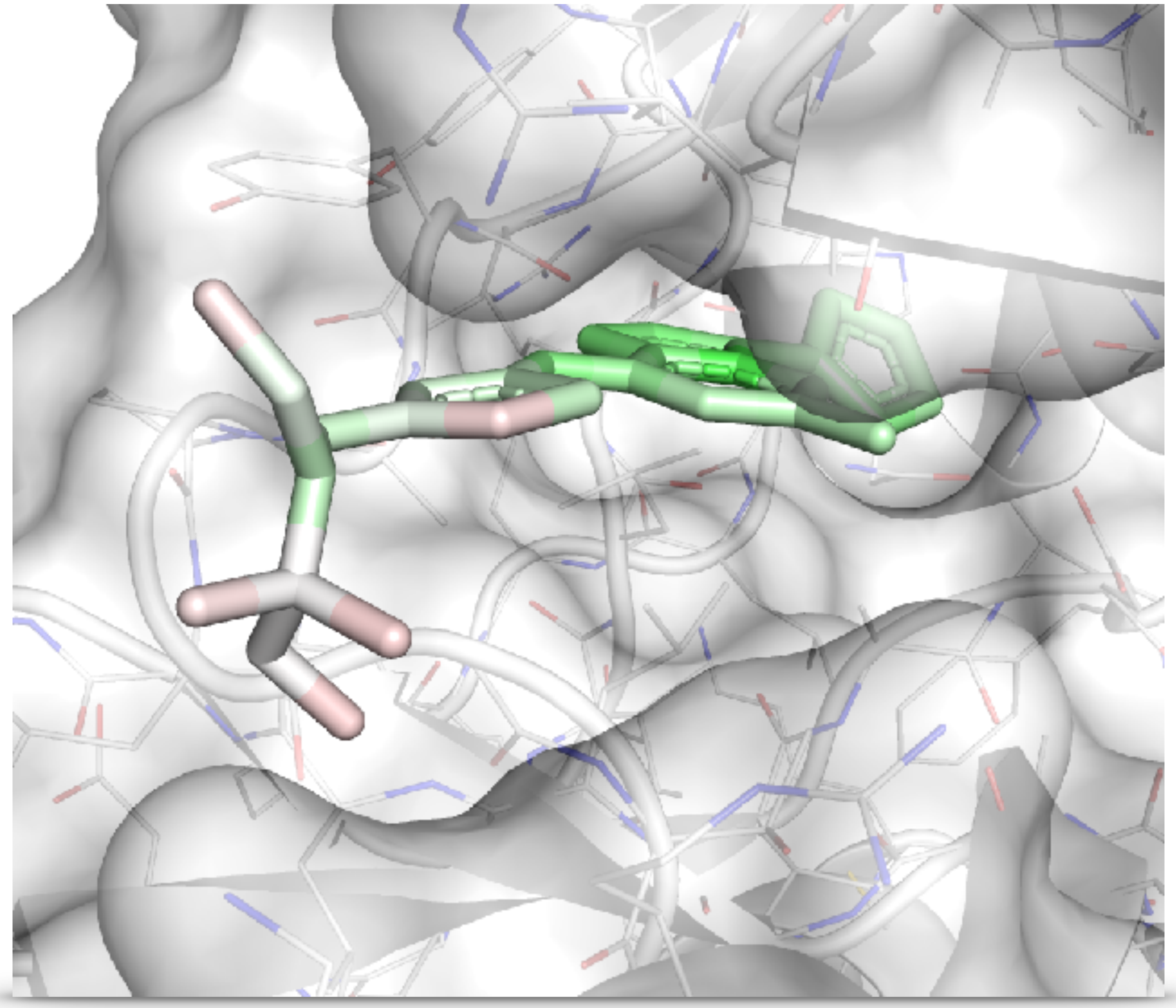
Pose Prediction Binding Discrimination **Affinity Prediction**

Structure Based Drug Design

Virtual Screening

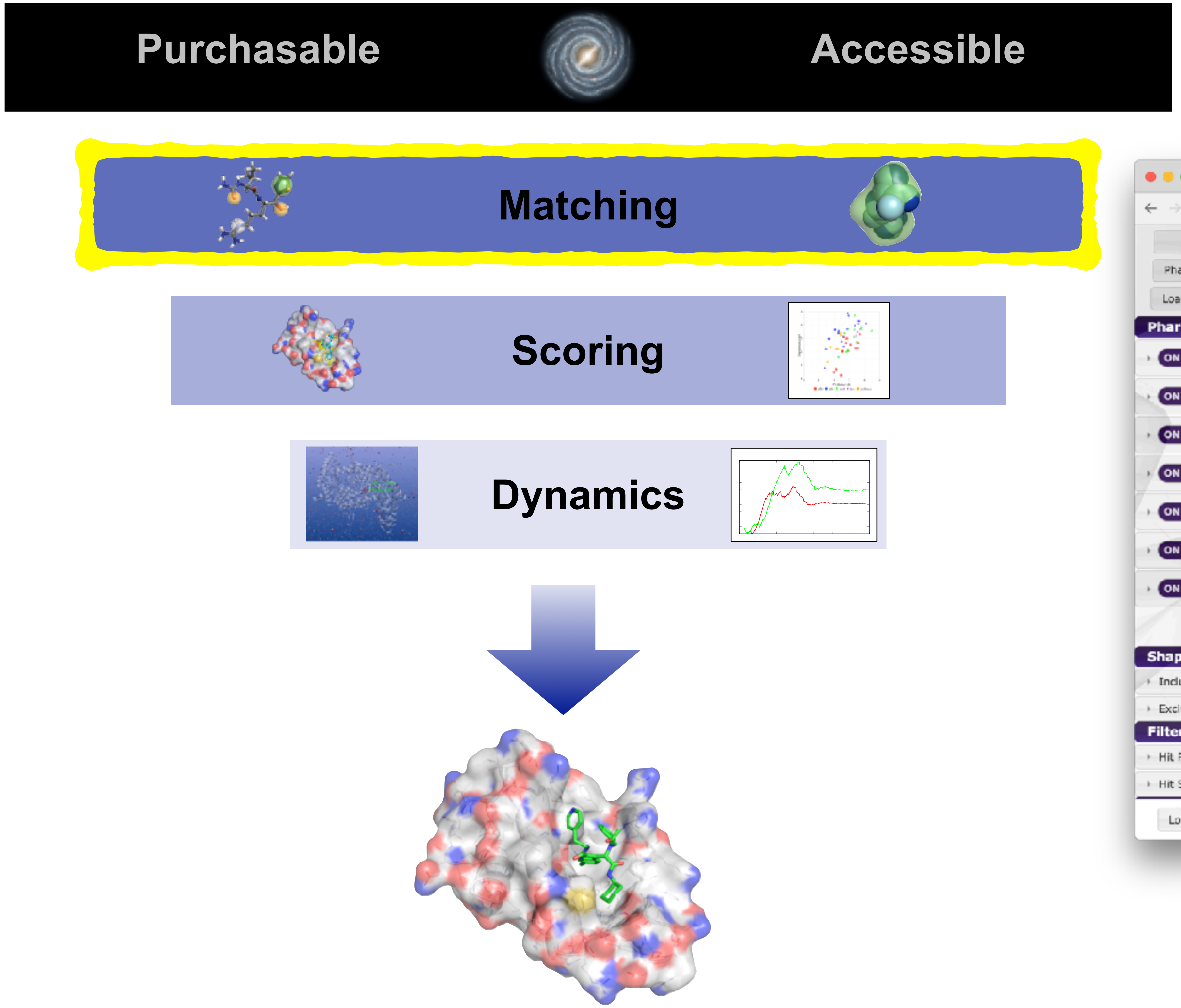


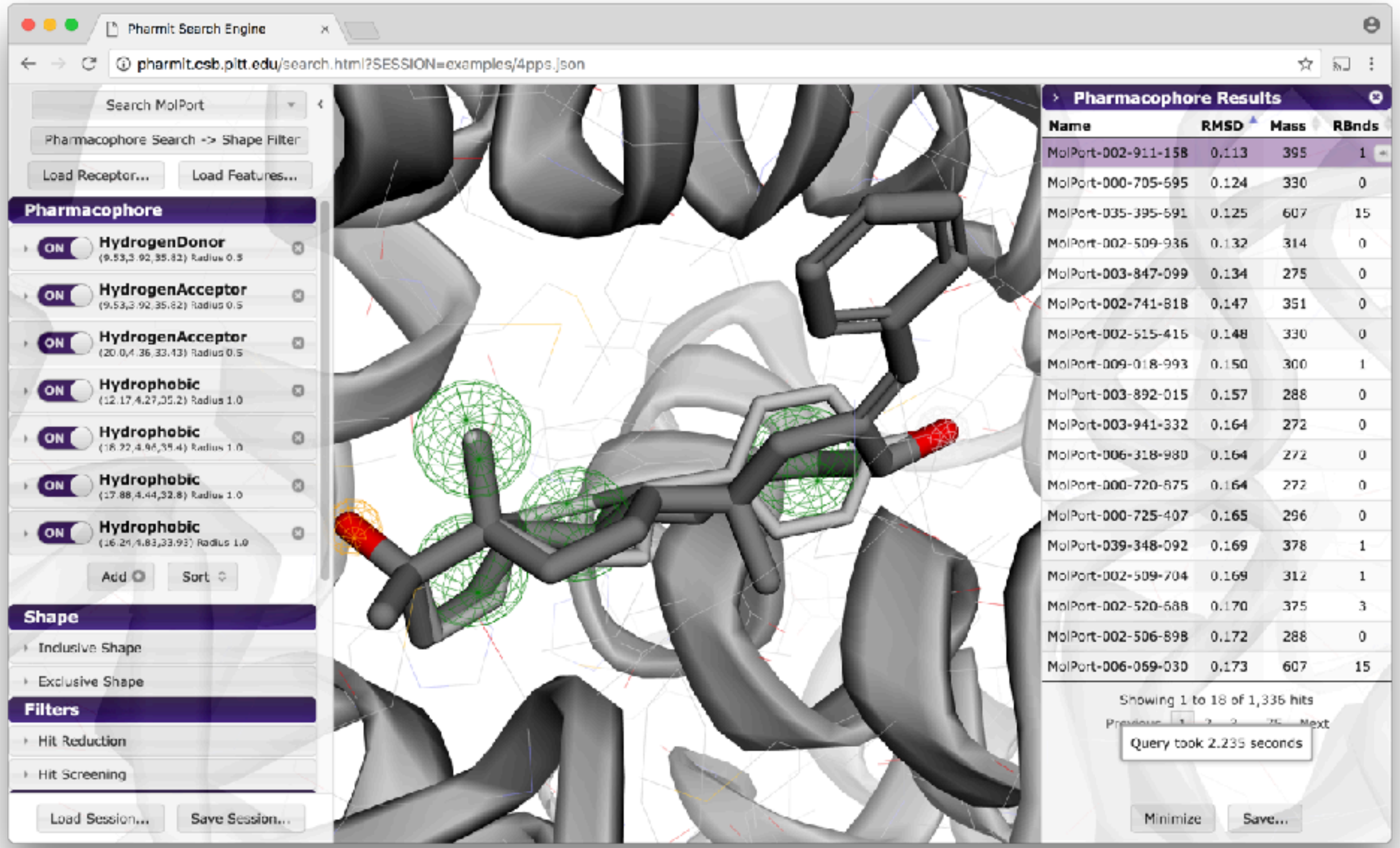
Lead Optimization



Pose Prediction Binding Discrimination **Affinity Prediction**

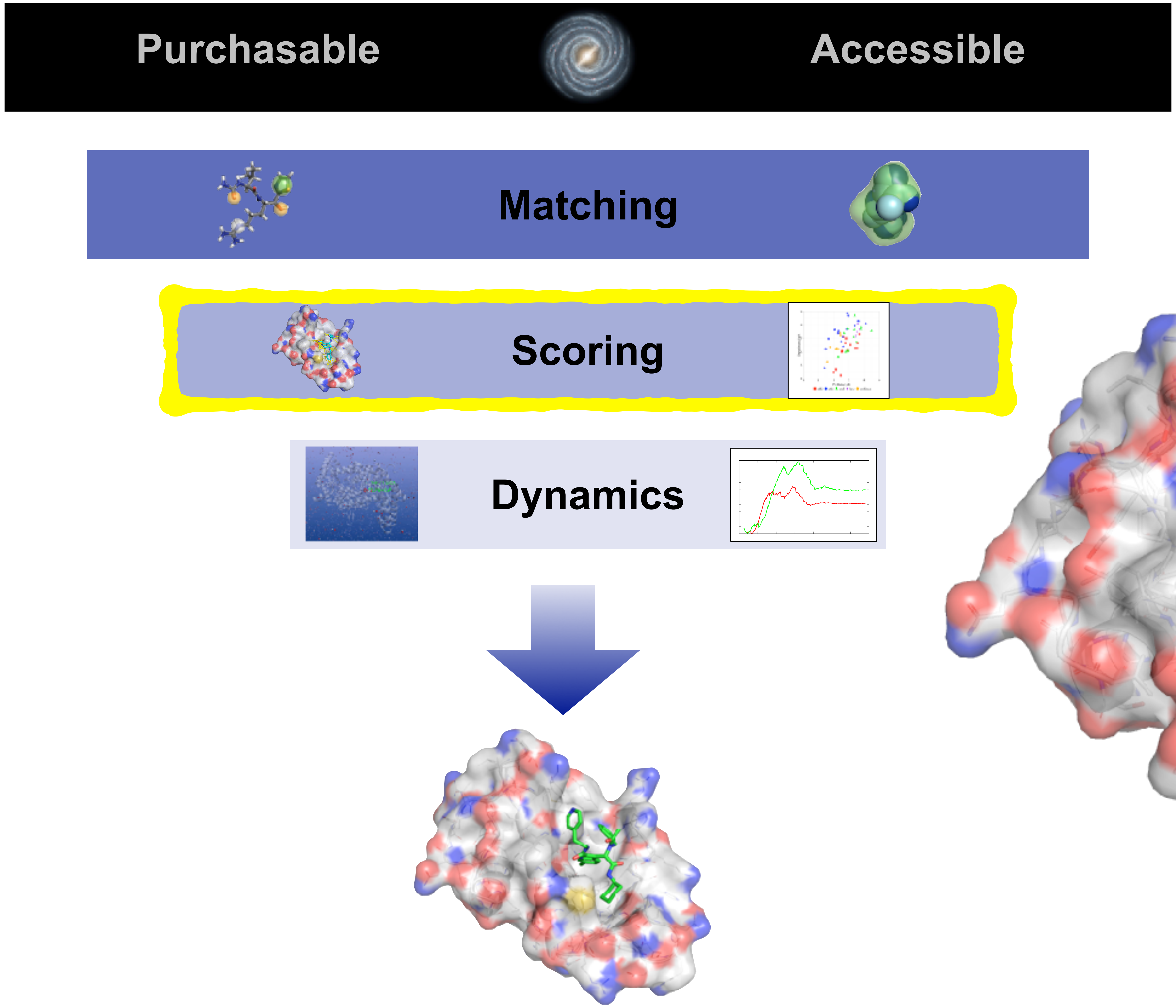
Drug Discovery Funnel





http://pharmit.csb.pitt.edu

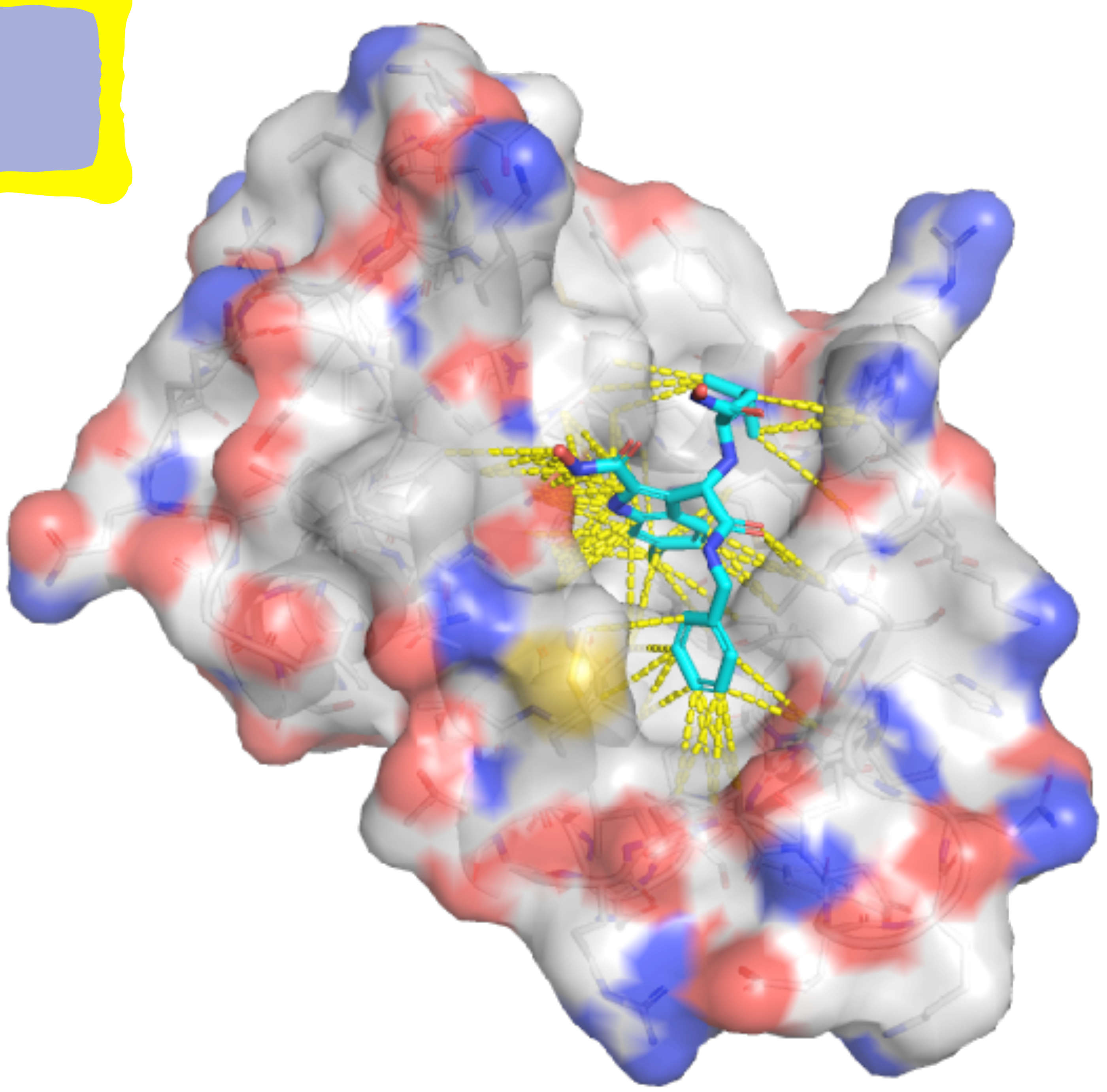
Drug Discovery Funnel

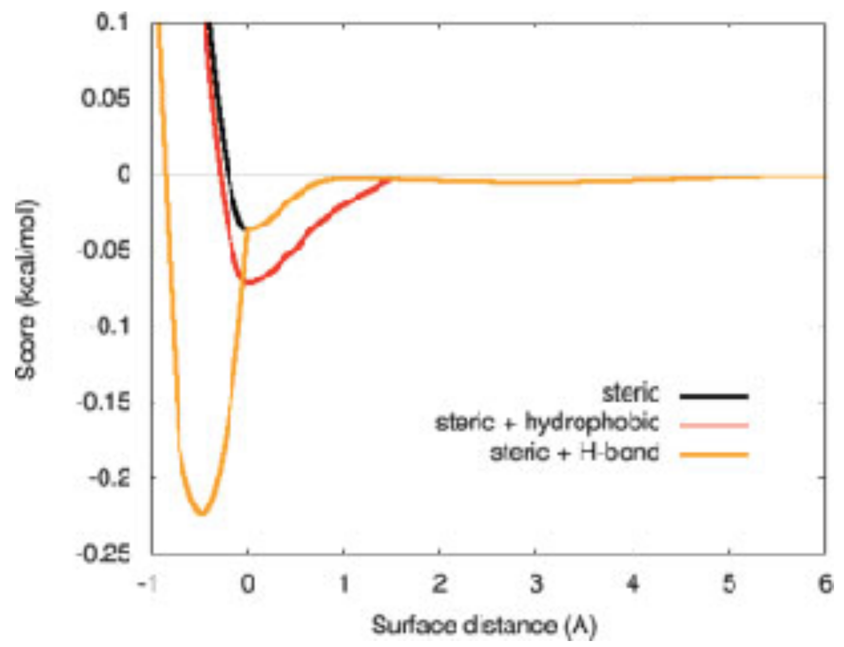


$$
\begin{aligned}
\text{gauss}_1(d) &= w_{\text{guass}_1} e^{-(d/0.5)^2} \\
\text{gauss}_2(d) &= w_{\text{guass}_2} e^{-((d-3)/2)^2} \\
\text{repulsion}(d) &= \begin{cases} w_{\text{repulsion}} d^2 & d < 0 \\ 0 & d \geq 0 \end{cases}
\end{aligned}
$$

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

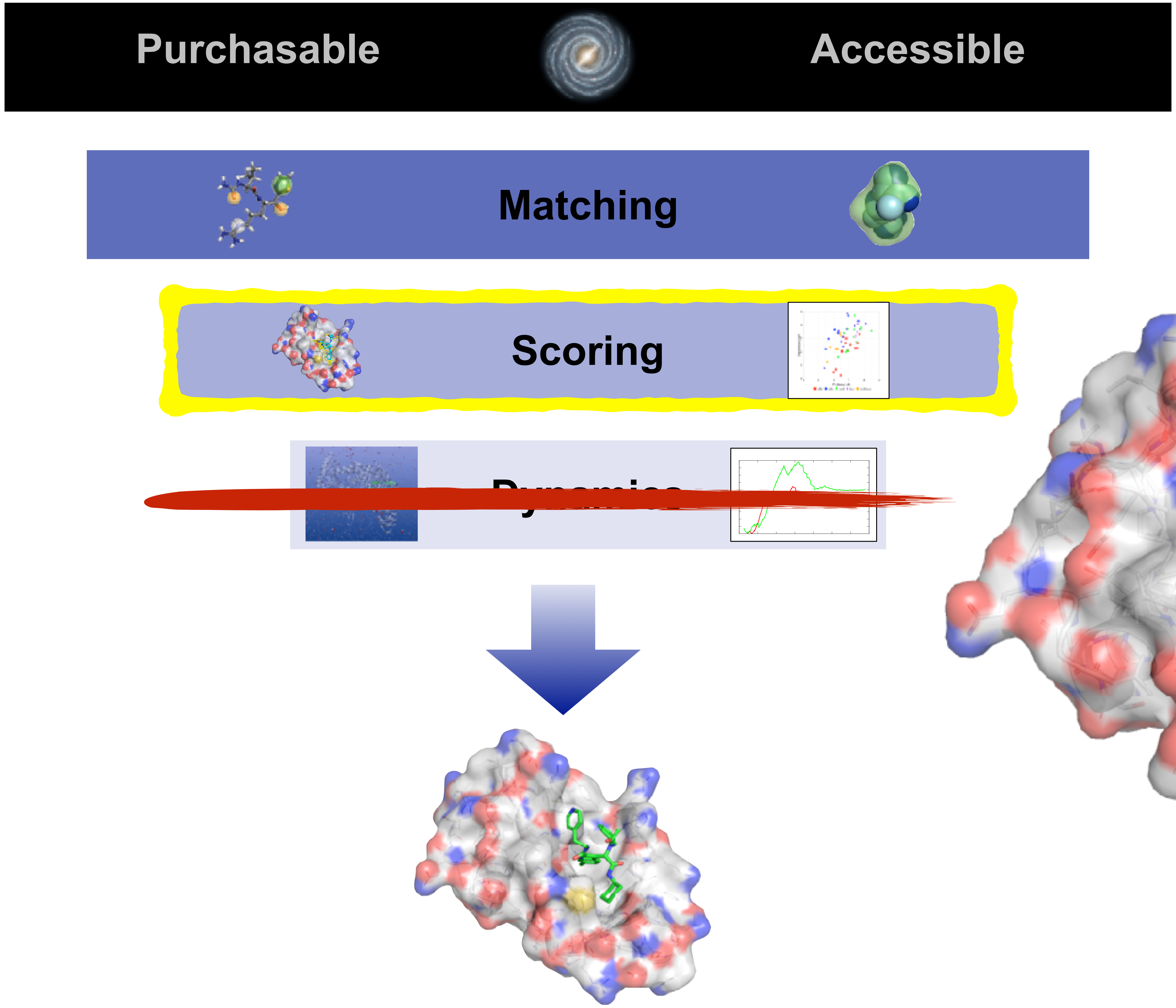
$$
\text{hbond}(d) = \begin{cases} w_{\text{hbond}} & d < -0.7 \\ 0 & d > 0 \\ w_{\text{hbond}}(-\frac{10}{7}d) & \textit{otherwise} \end{cases}
$$

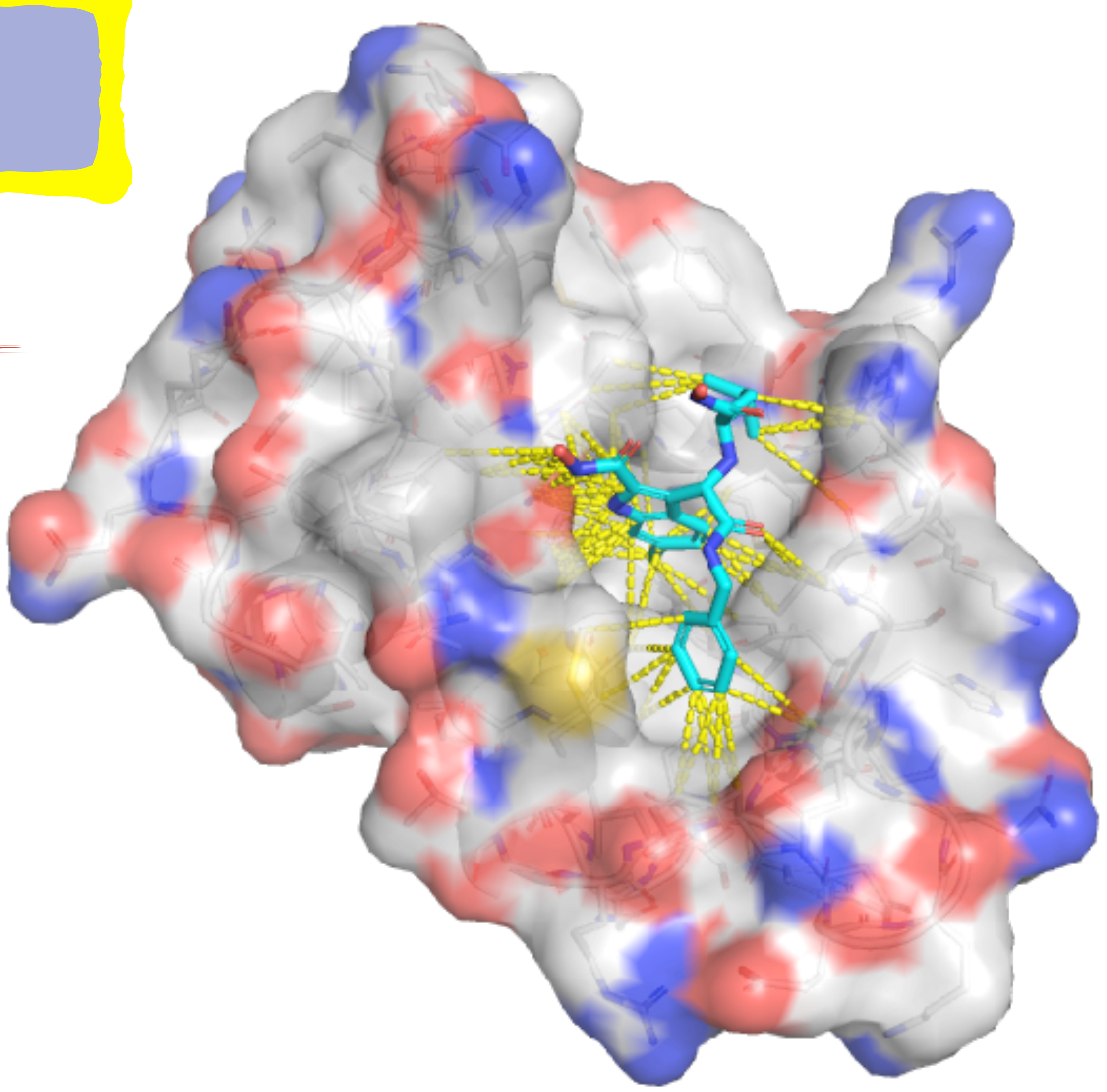




O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry* 31 (2010) 455-461

Drug Discovery Funnel

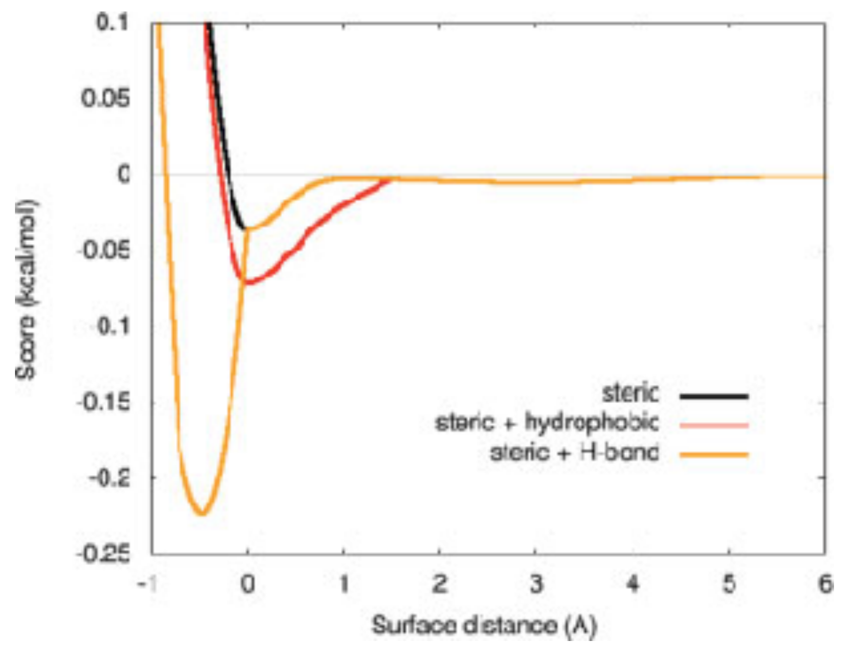




$$
\begin{aligned}
\text{gauss}_1(d) &= w_{\text{guass}_1} e^{-(d/0.5)^2} \\
\text{gauss}_2(d) &= w_{\text{guass}_2} e^{-((d-3)/2)^2} \\
\text{repulsion}(d) &= \begin{cases} w_{\text{repulsion}} d^2 & d < 0 \\ 0 & d \geq 0 \end{cases}
\end{aligned}
$$

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

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



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

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?

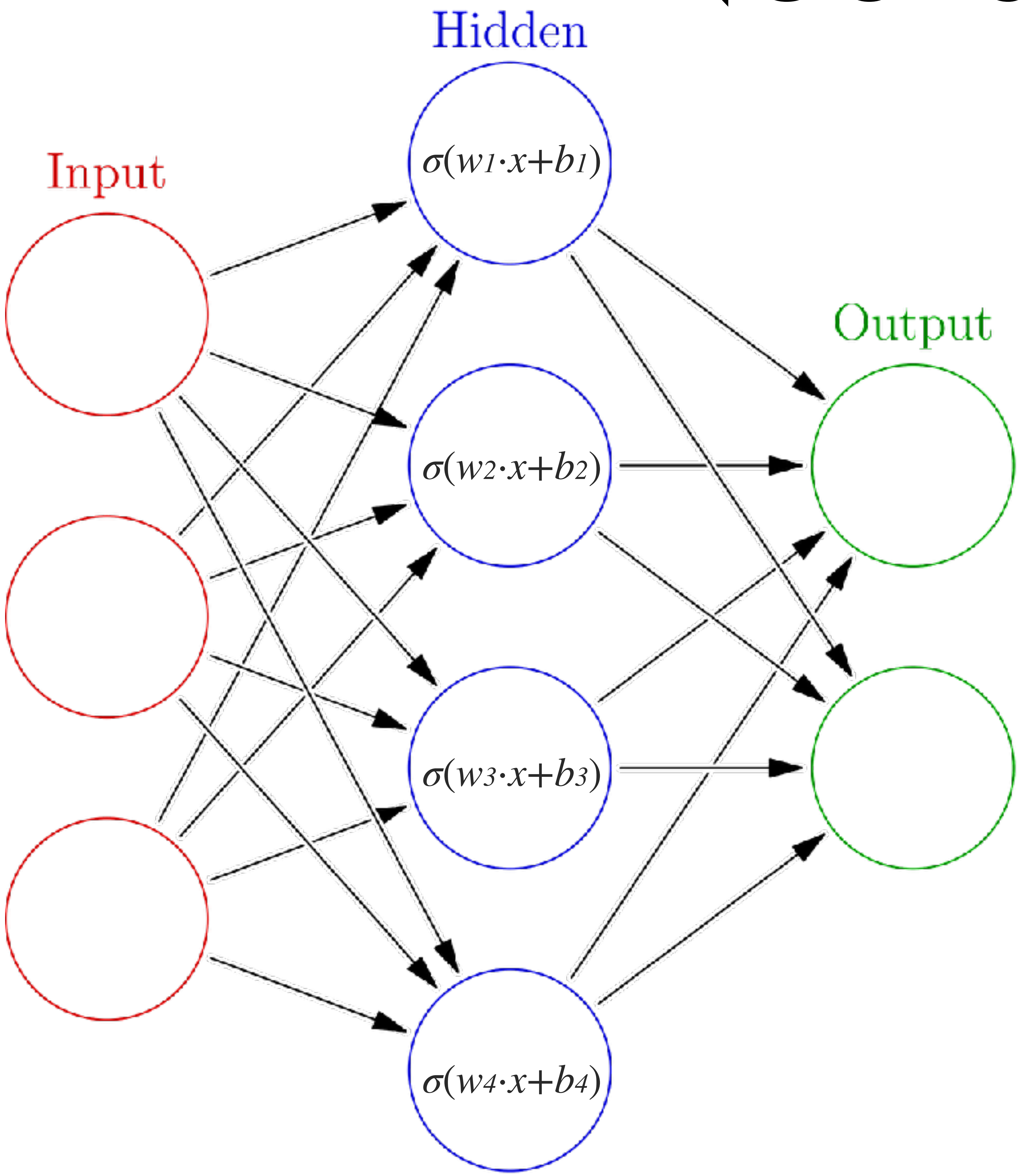
Key Idea: Leverage "big data"

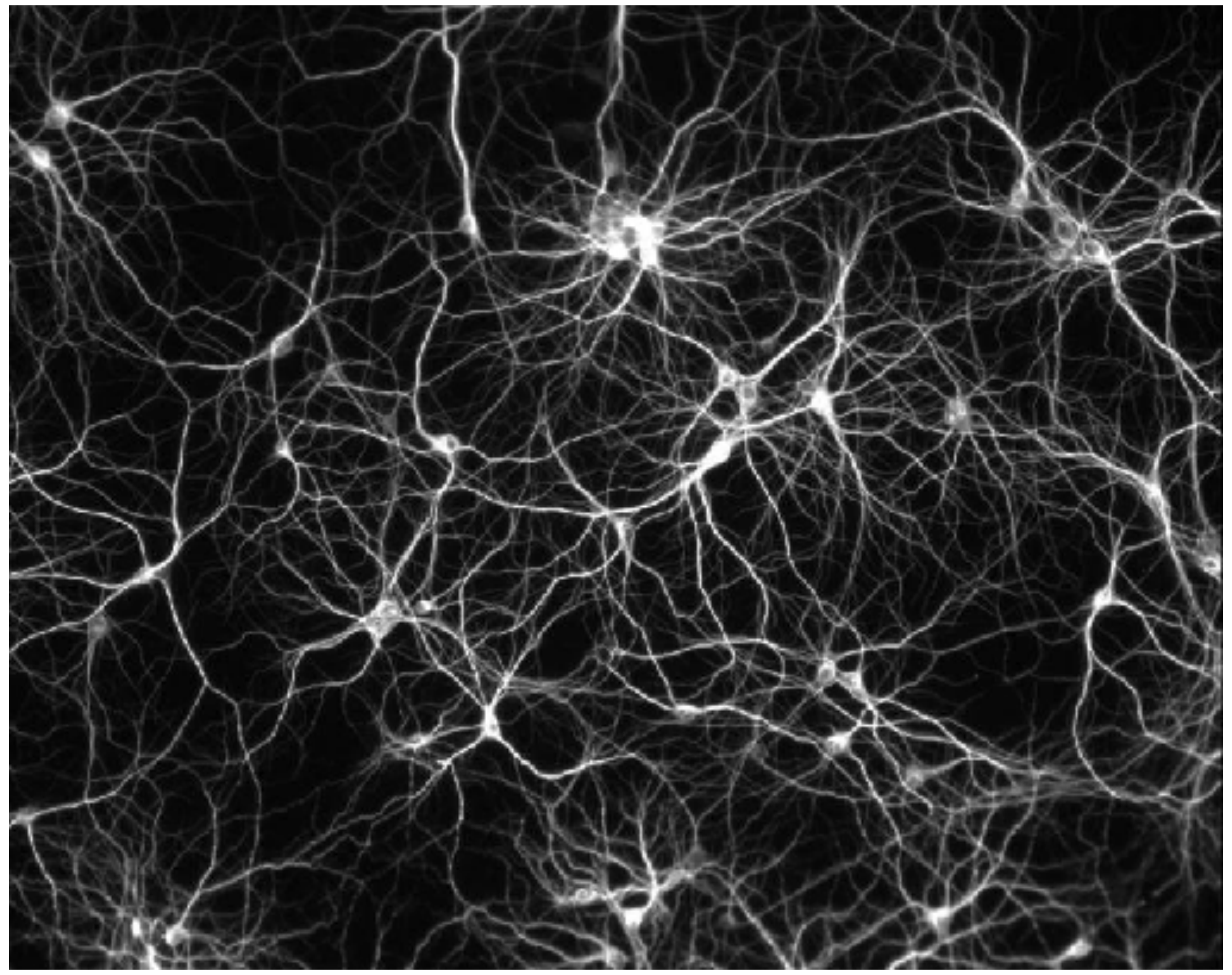
- 231,655,275 bioactivities in PubChem
- 125,526 structures in the PDB
- 16,179 annotated complexes in PDBbind

Machine Learning

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

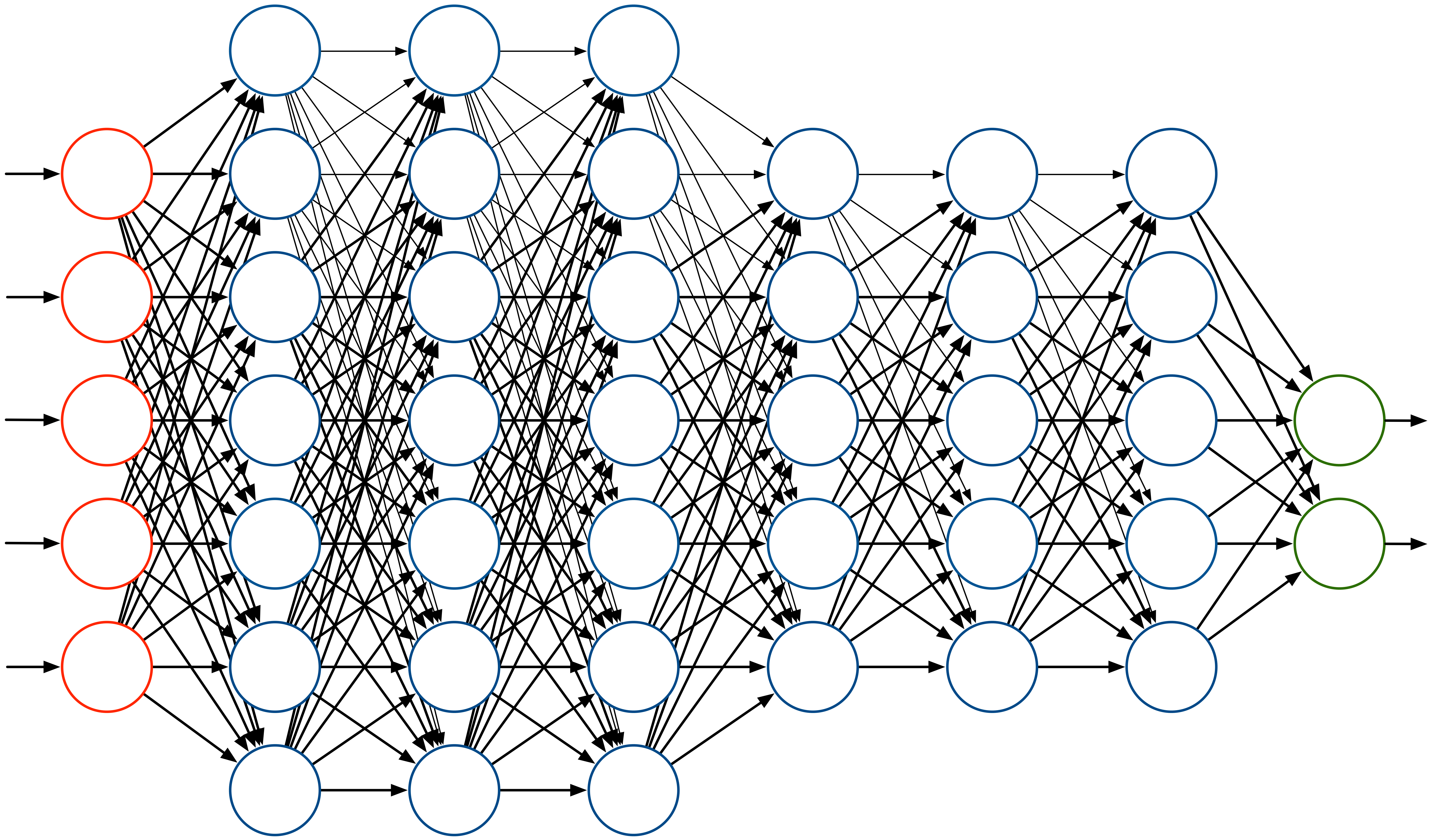
Neural Networks



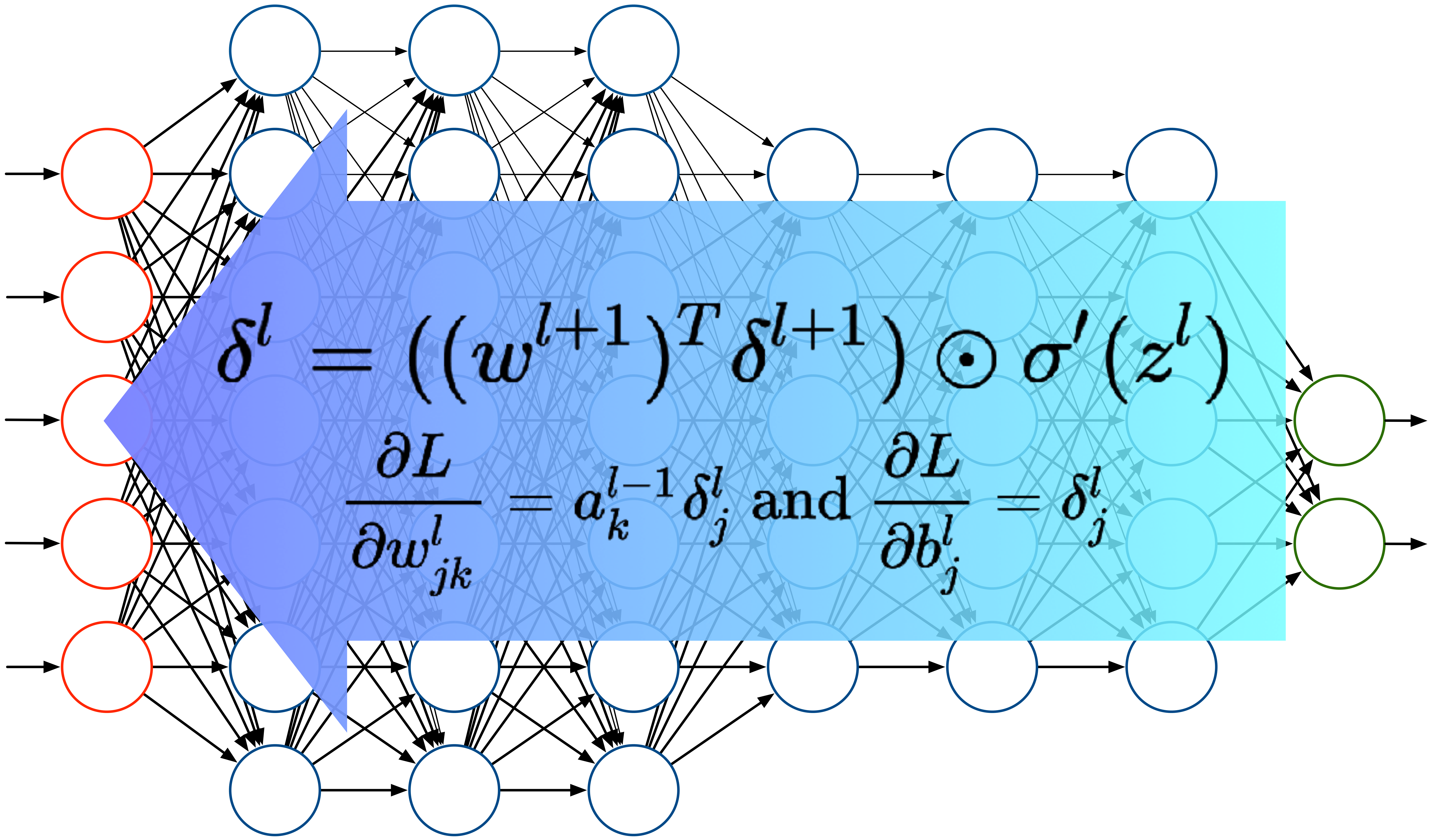


The **universal approximation theorem** states that, under reasonable assumptions, a feedforward **neural network** with a finite number of nodes **can approximate any continuous** function to within a given error over a bounded input domain.

Deep Learning



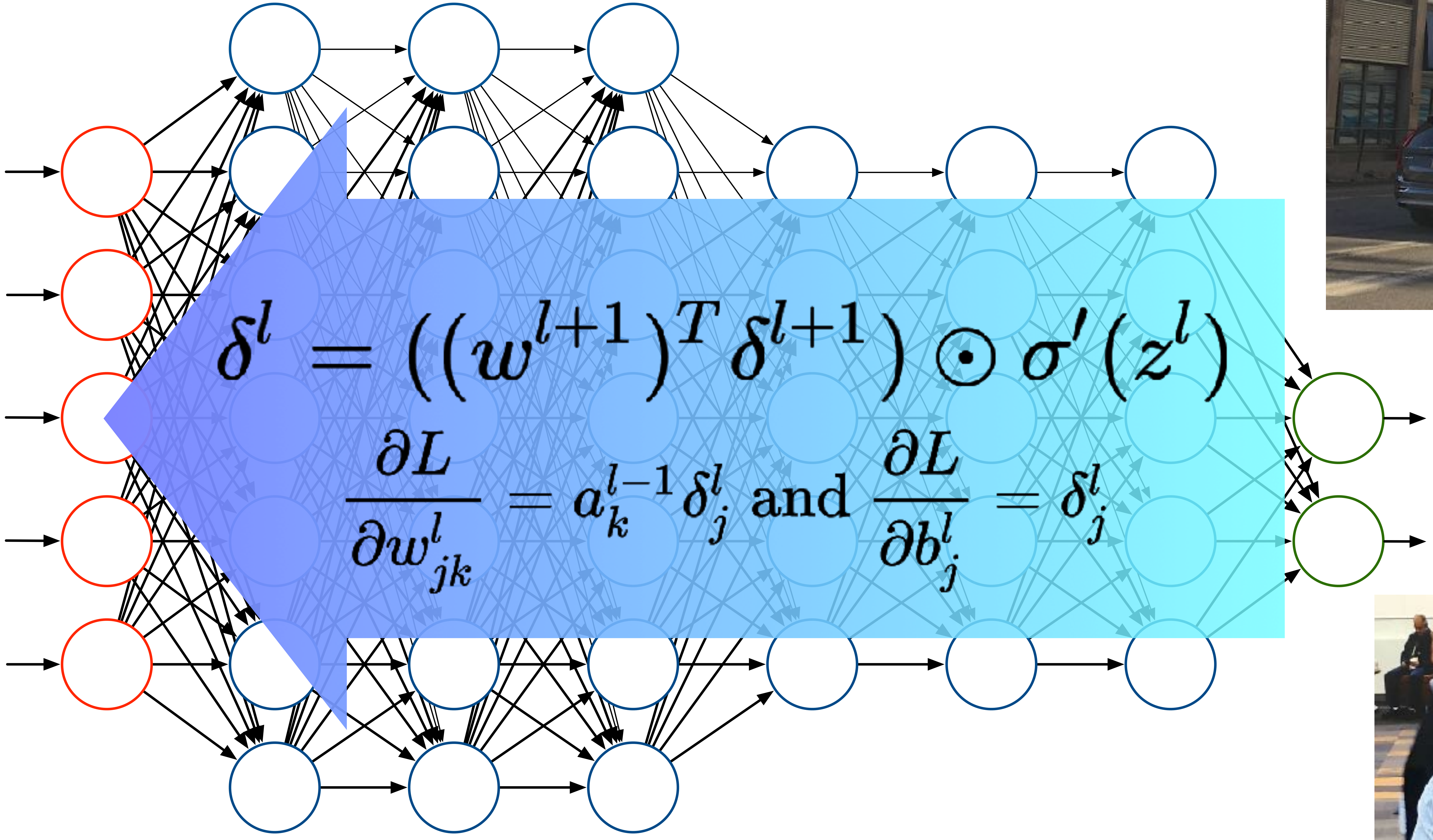
Deep Learning



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

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

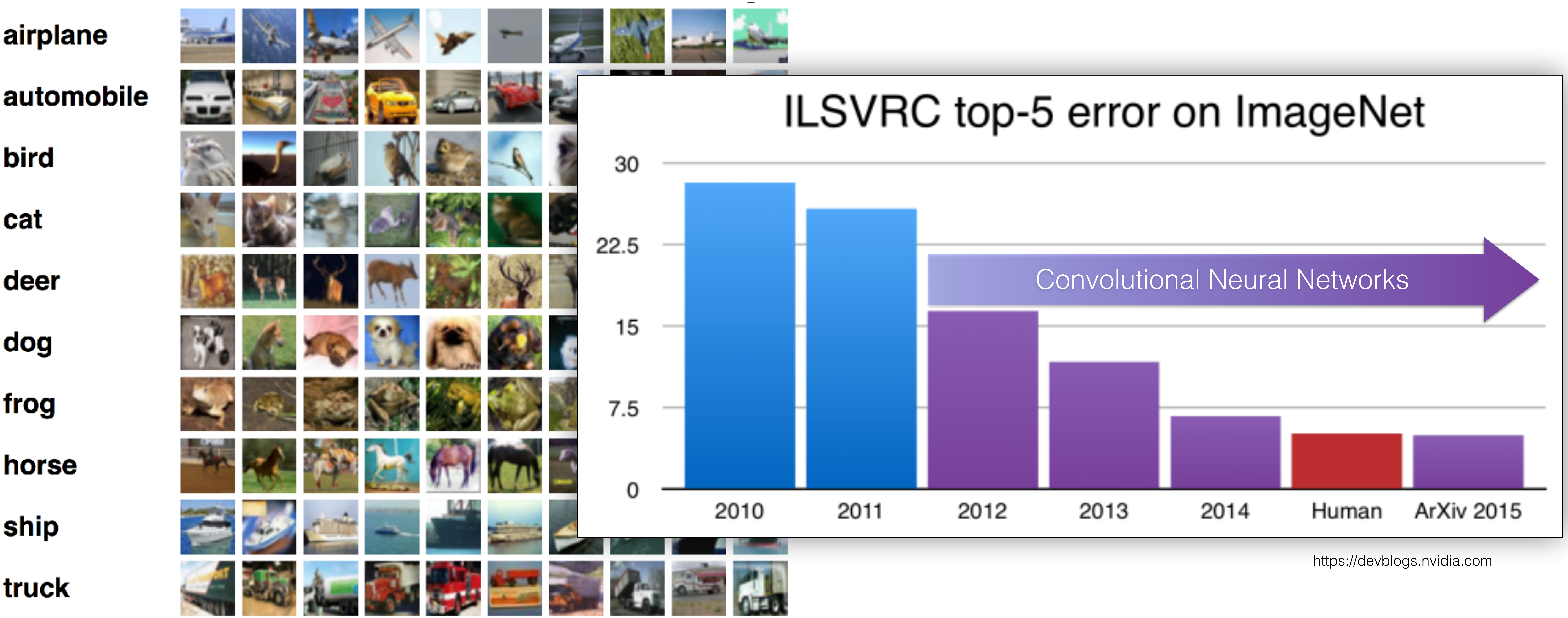
Deep Learning



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

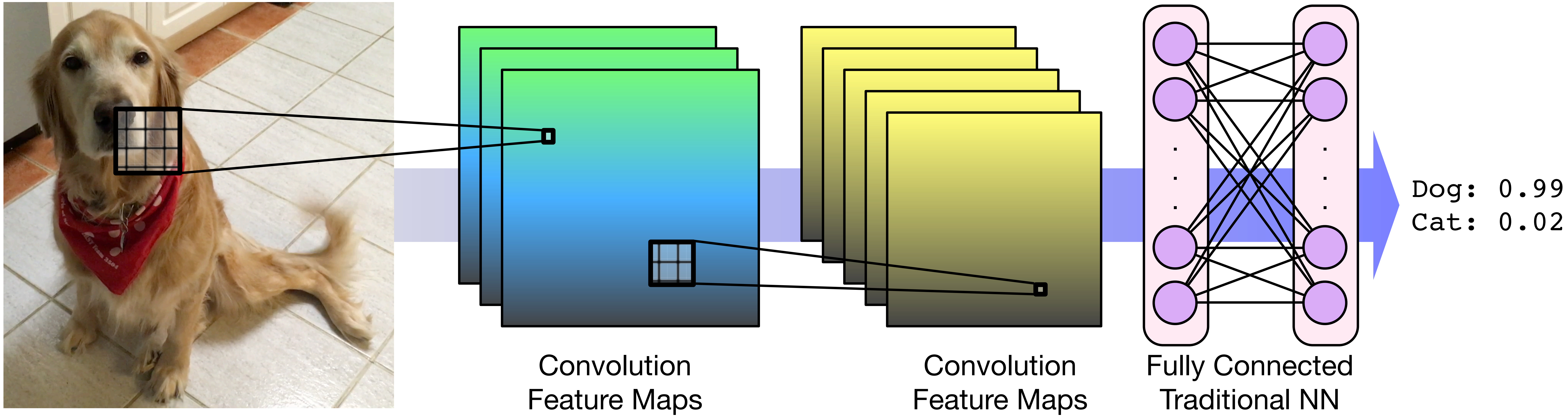
$$\frac{\partial L}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j \text{ and } \frac{\partial L}{\partial b^l_j} = \delta^l_j$$

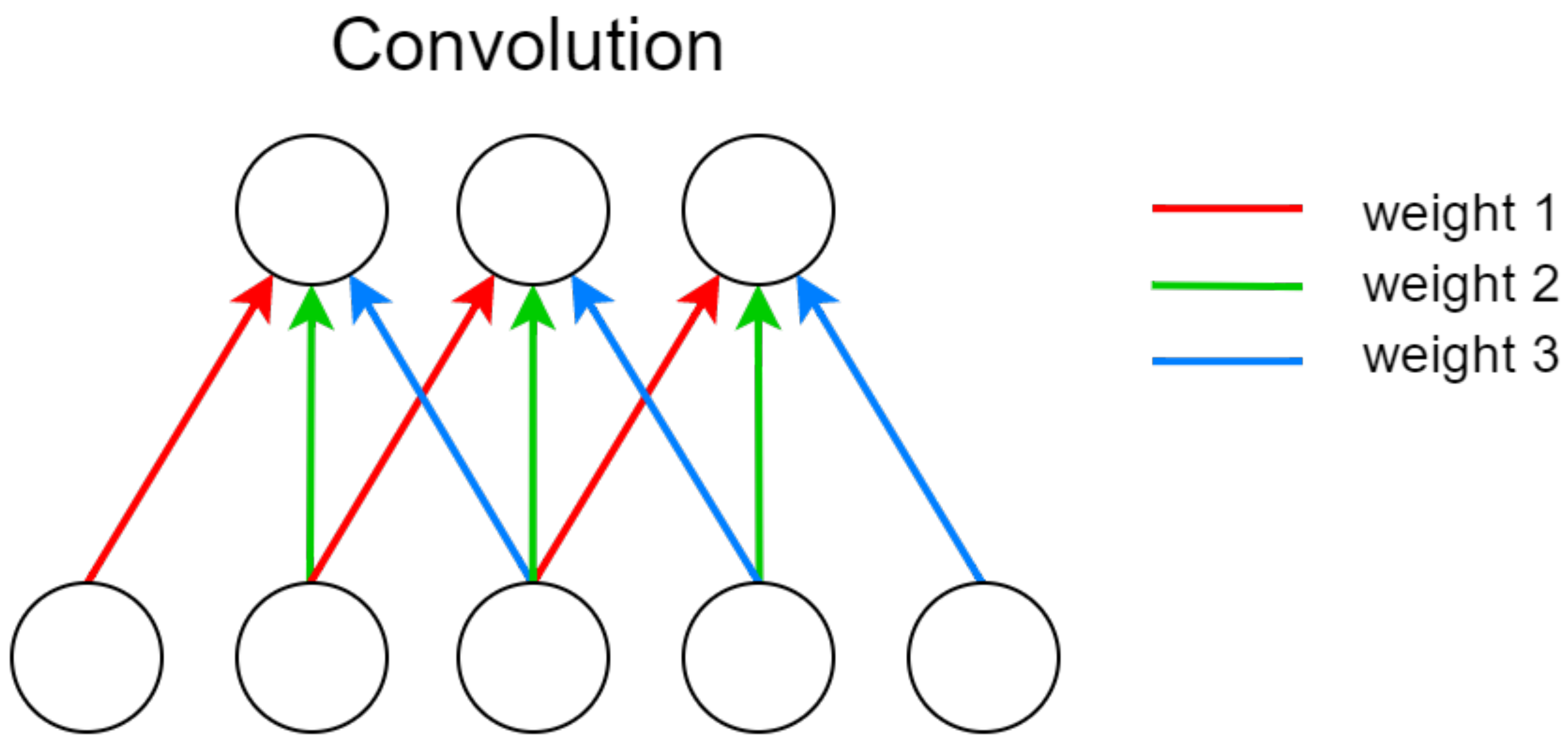
Image Recognition

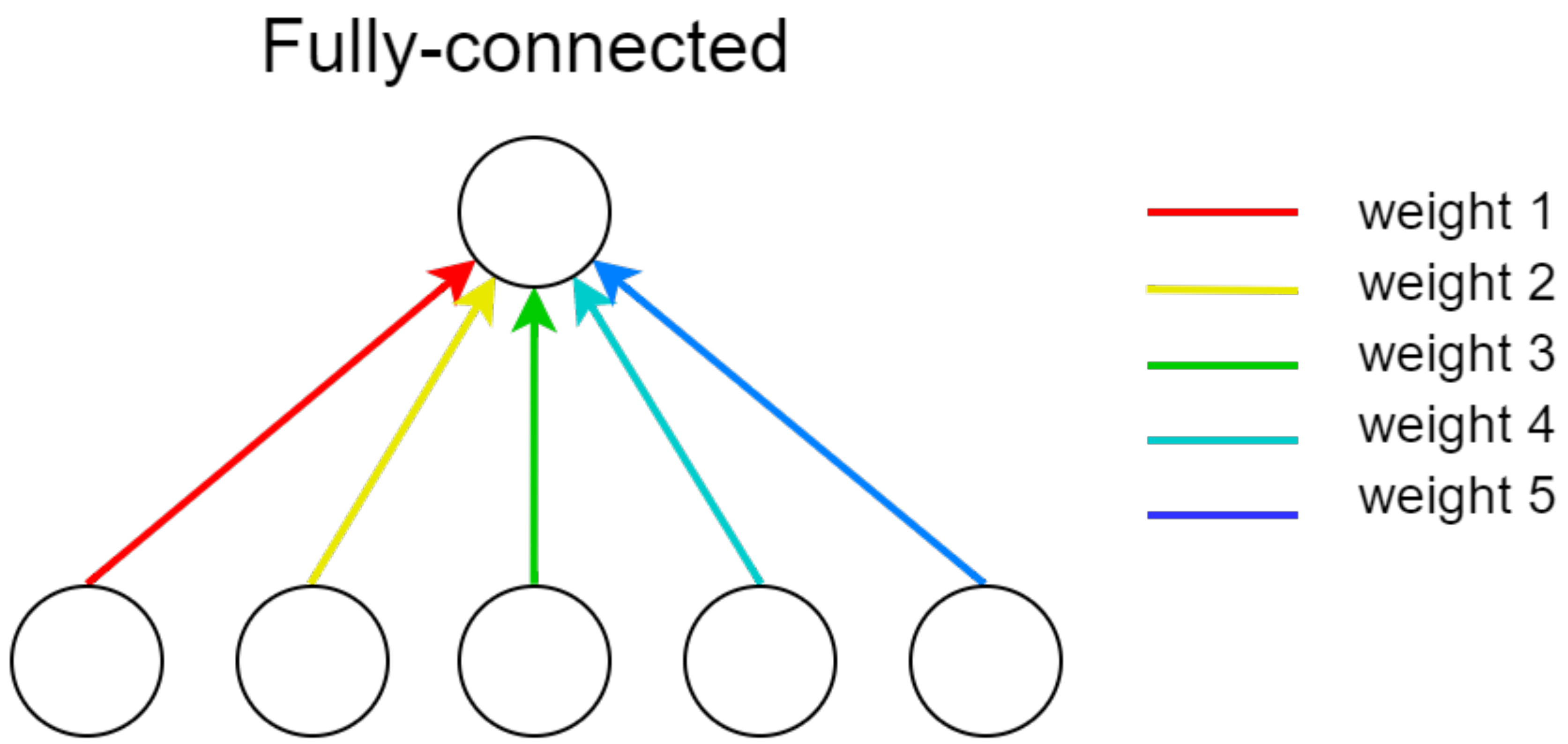


https://devblogs.nvidia.com

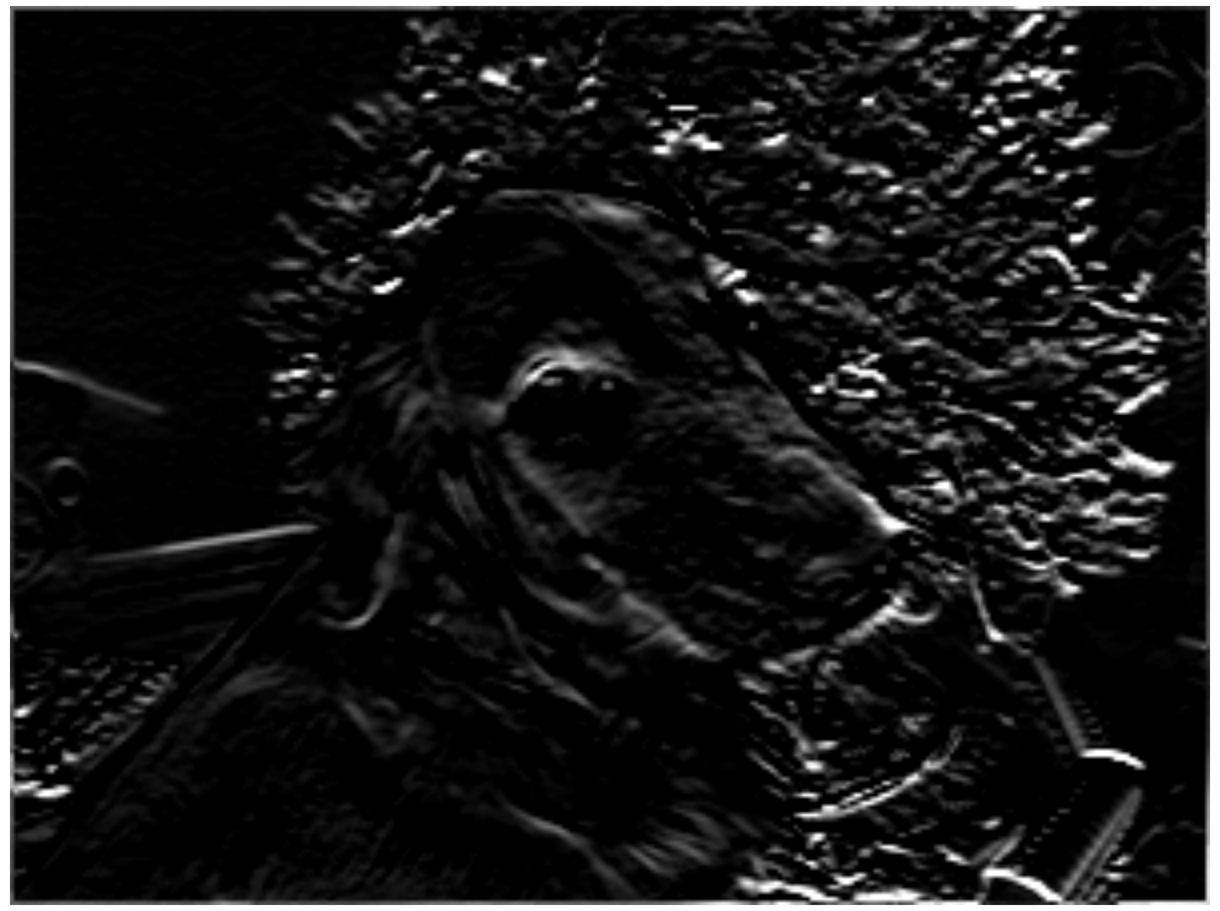
Convolutional Neural Networks







Convolutional Filters

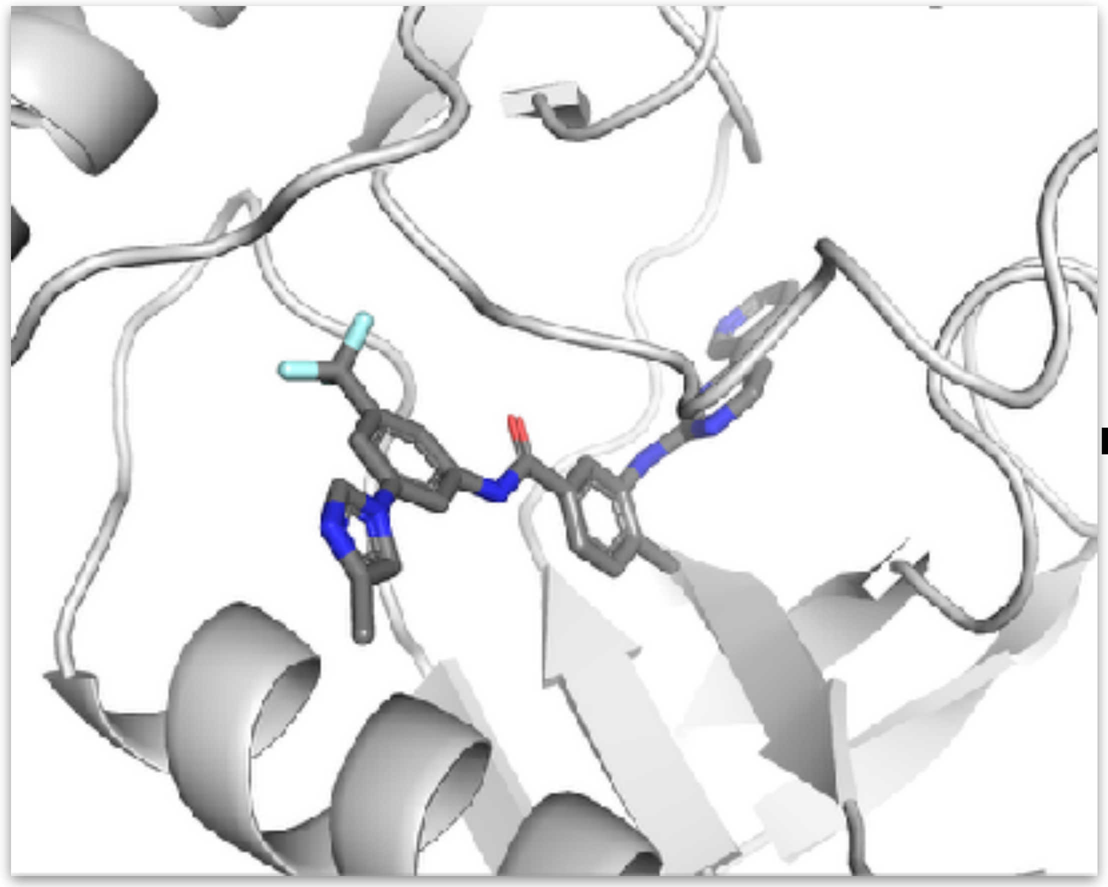


-1	-1	-1
0	0	0
1	1	1

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

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

CNNs for Protein-Ligand Scoring



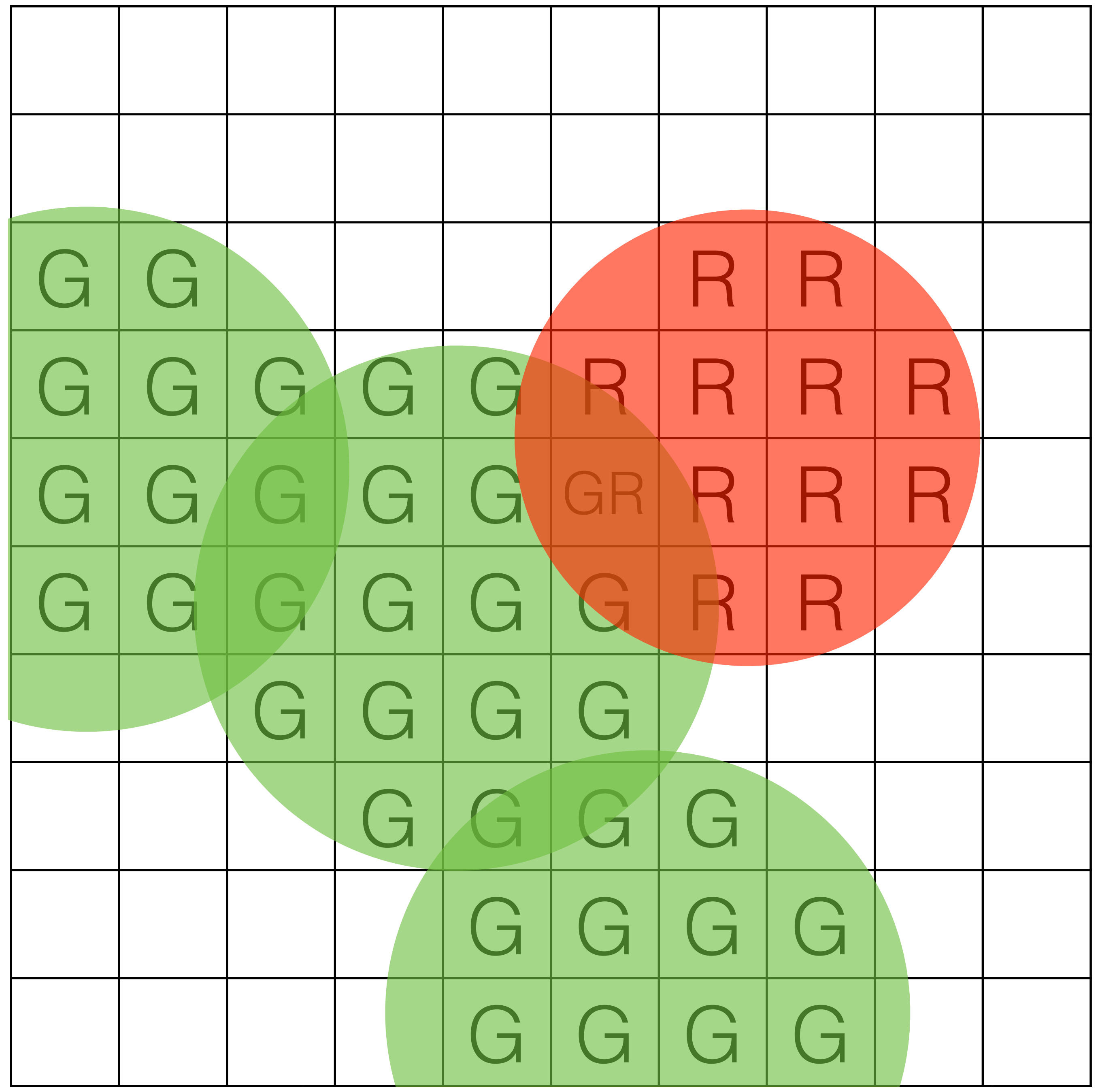
CNN

Pose Prediction

Binding Discrimination

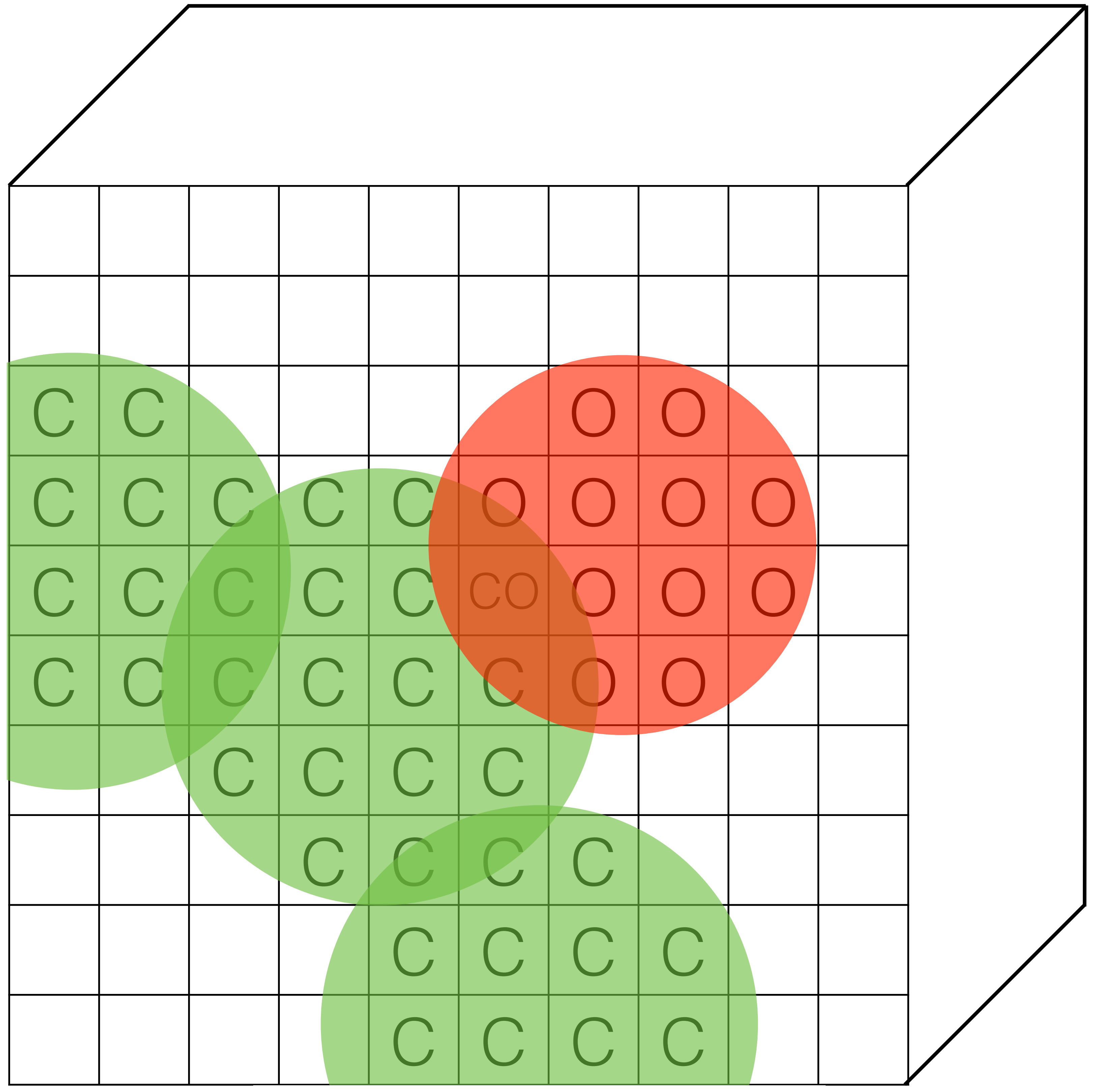
Affinity Prediction

Protein-Ligand Representation



(R,G,B) pixel

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 Data

PDBbind

Pose Prediction

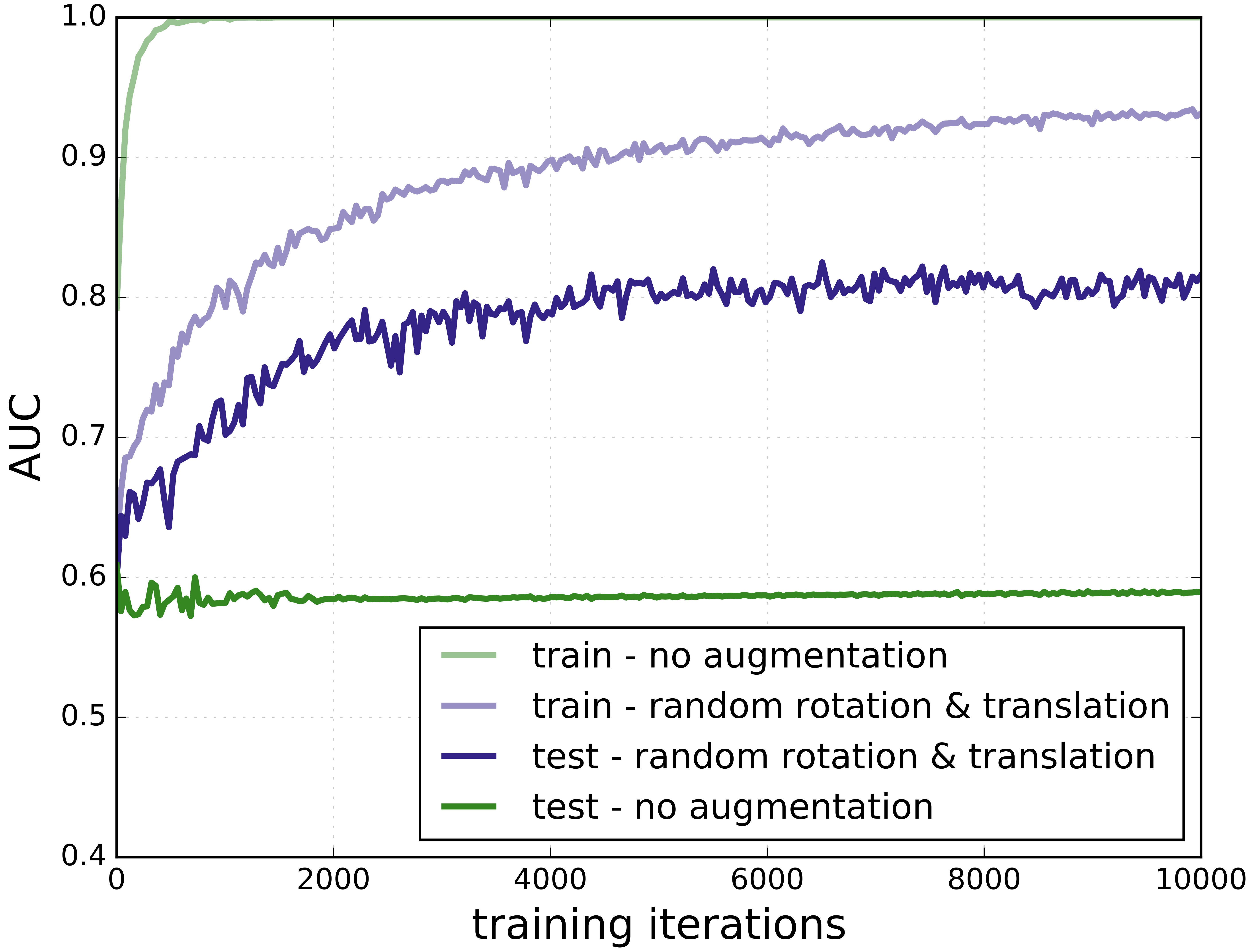
4056 protein-ligand complexes

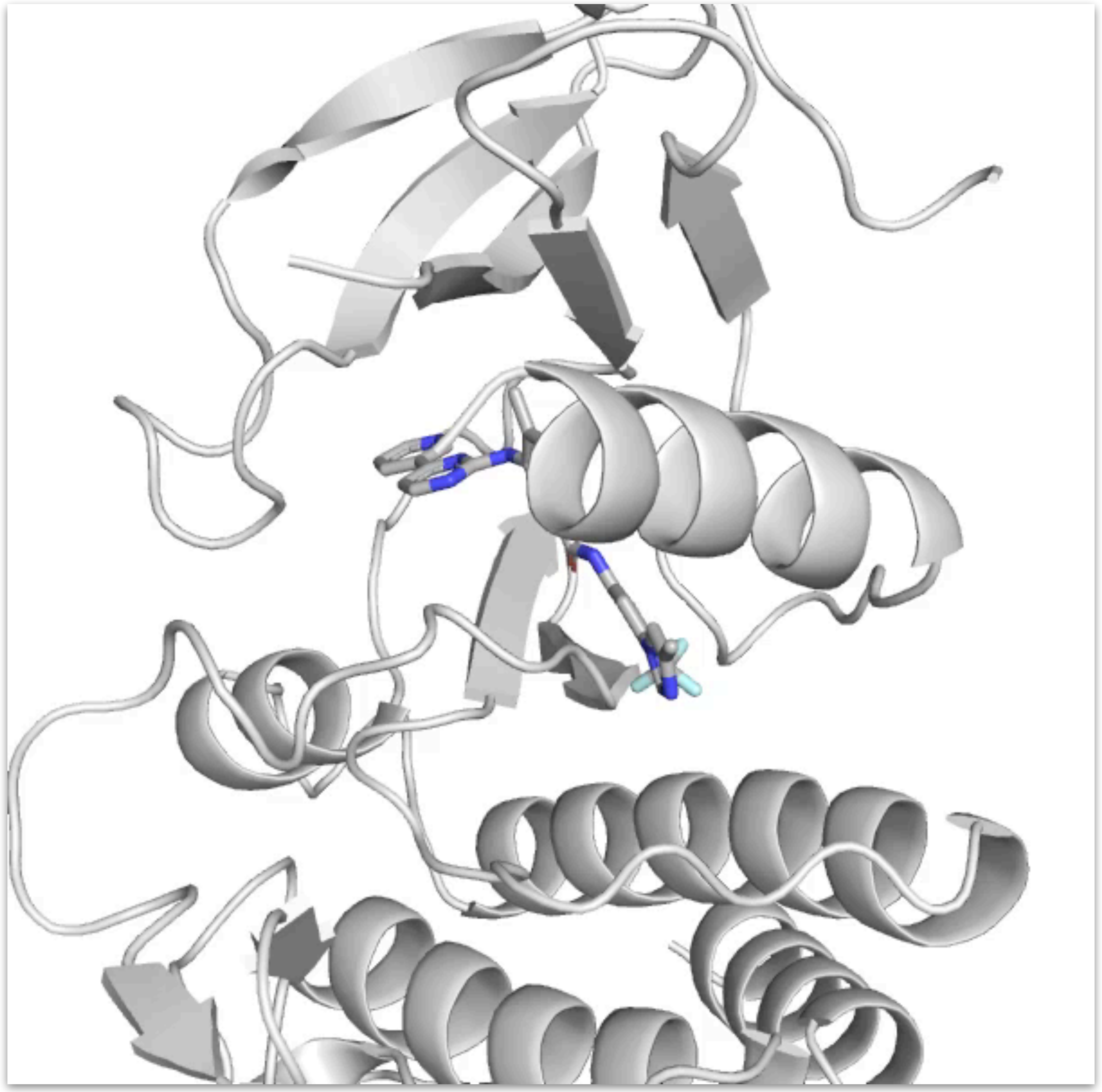
- diverse targets
- wide range of affinities
- generate poses with AutoDock Vina
- include minimized crystal pose
  - 8,688 <2Å RMSD (actives)
  - 76,743 >4Å RMSD (decoys)

Affinity Prediction

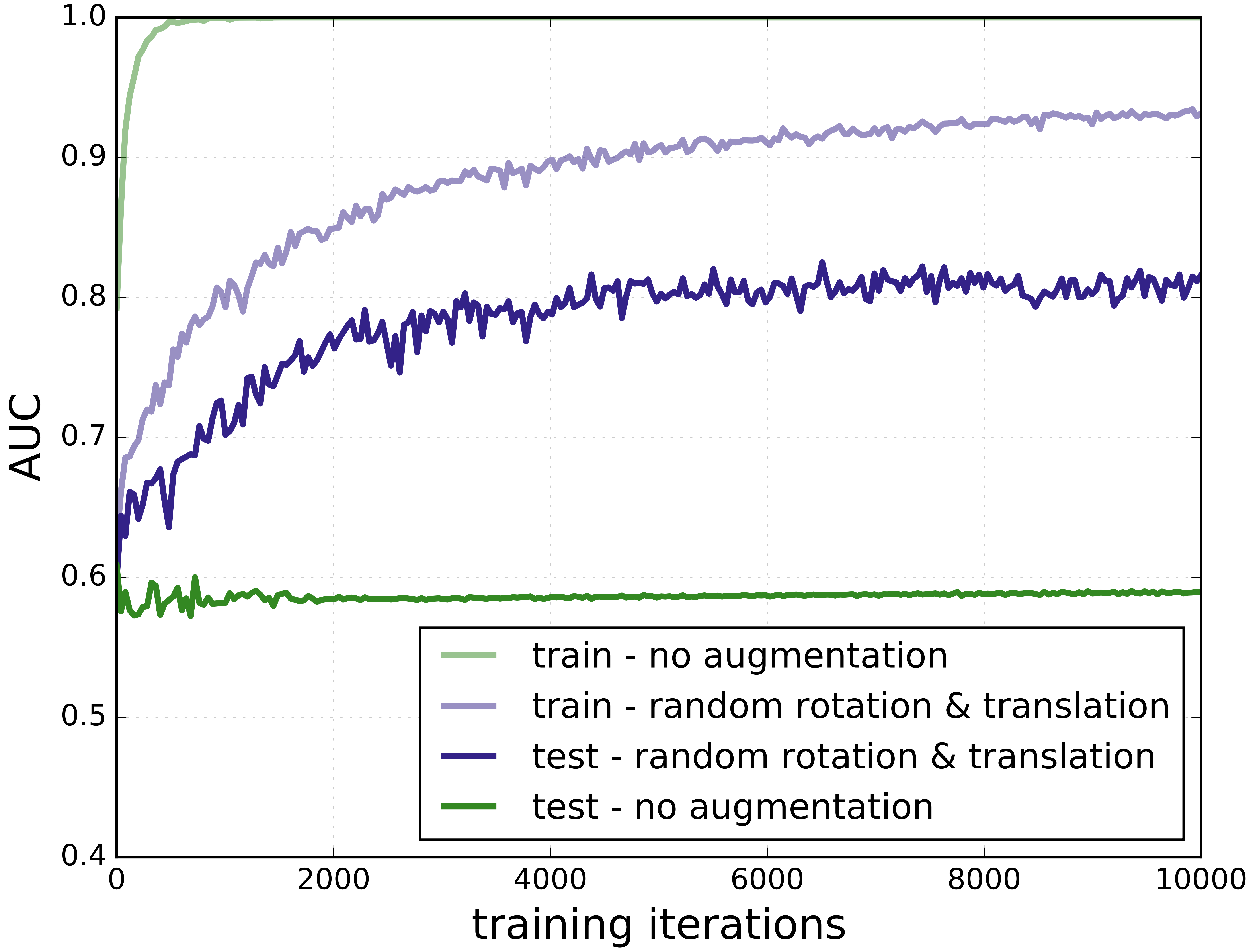
- 8,688 low RMSD poses
- assign known affinity
- **regression problem**

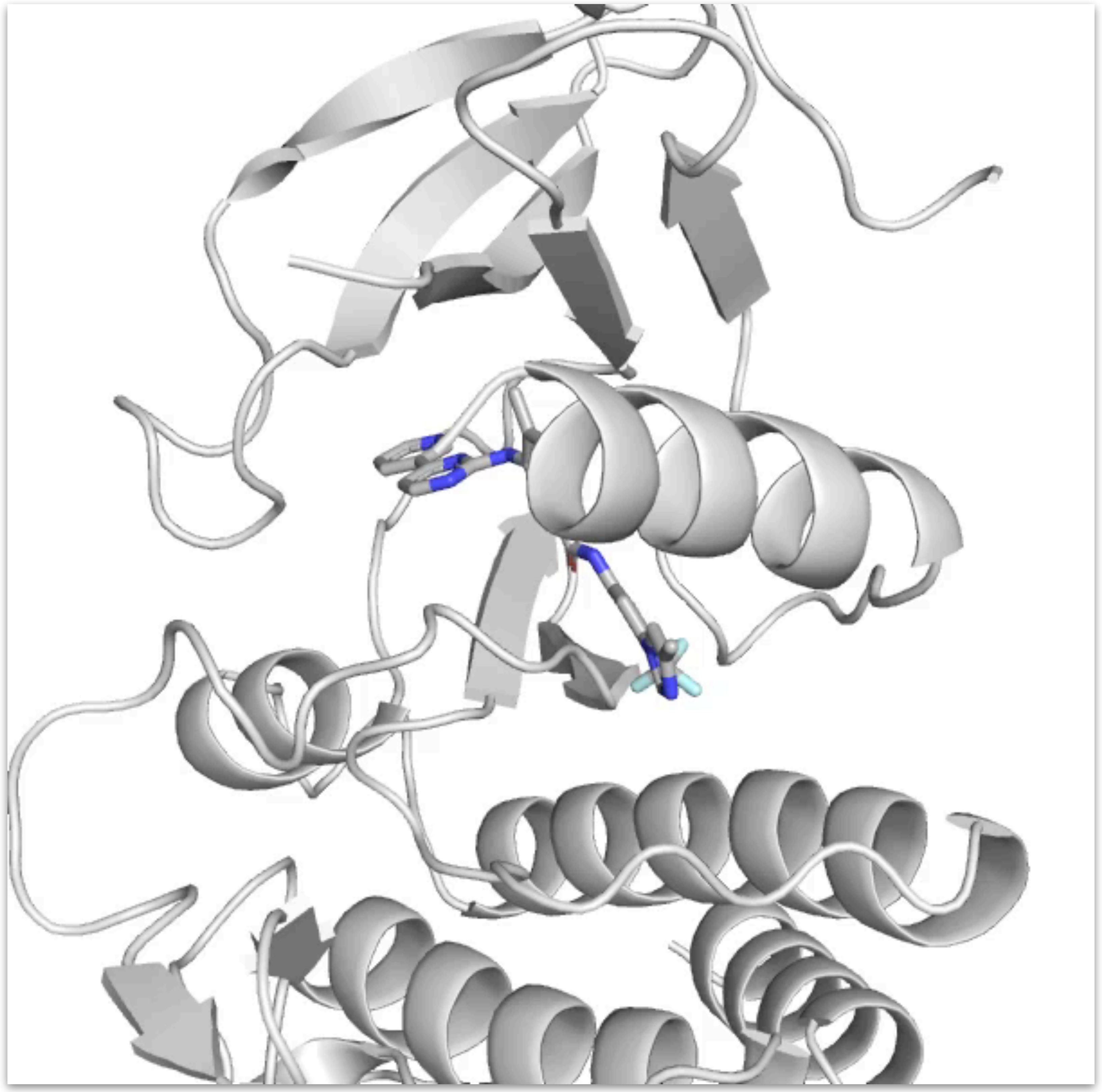
Data Augmentation





Data Augmentation





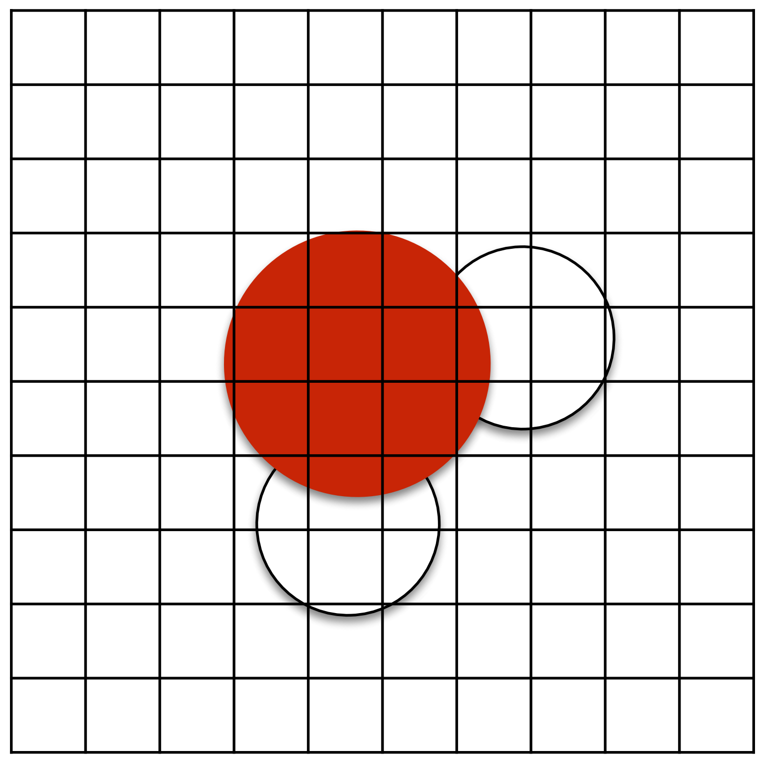
Why Grids?

Cons

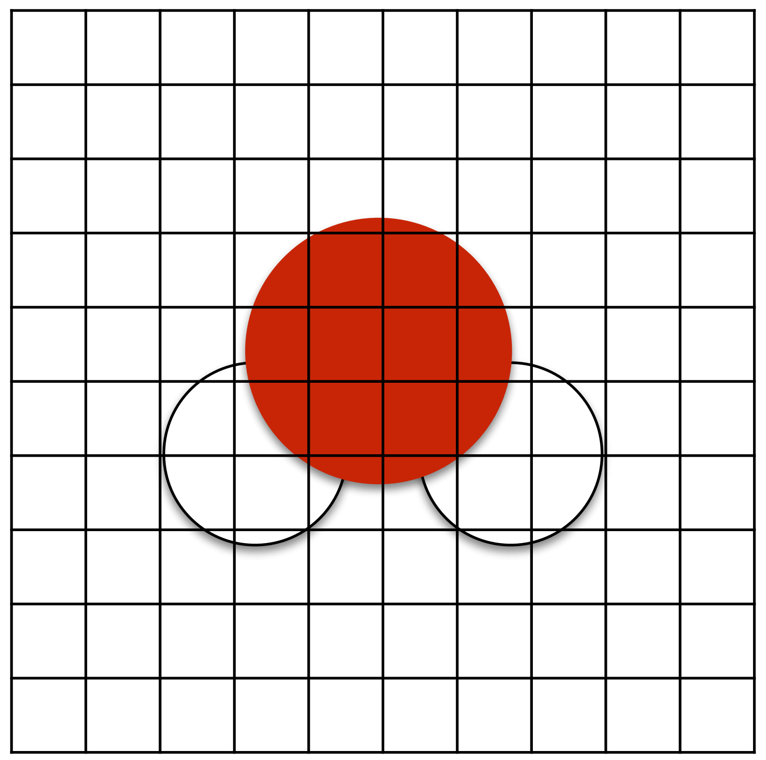
- *coordinate frame dependent*
- pairwise interactions not explicit

Pros

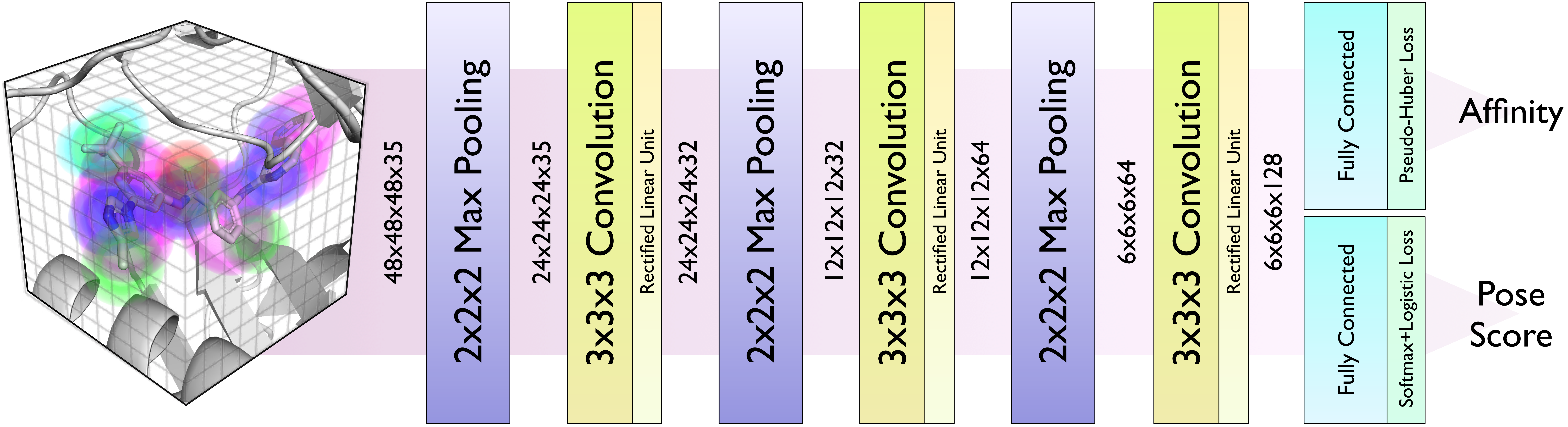
- clear spatial relationships
- amazingly parallel
- easy to interpret



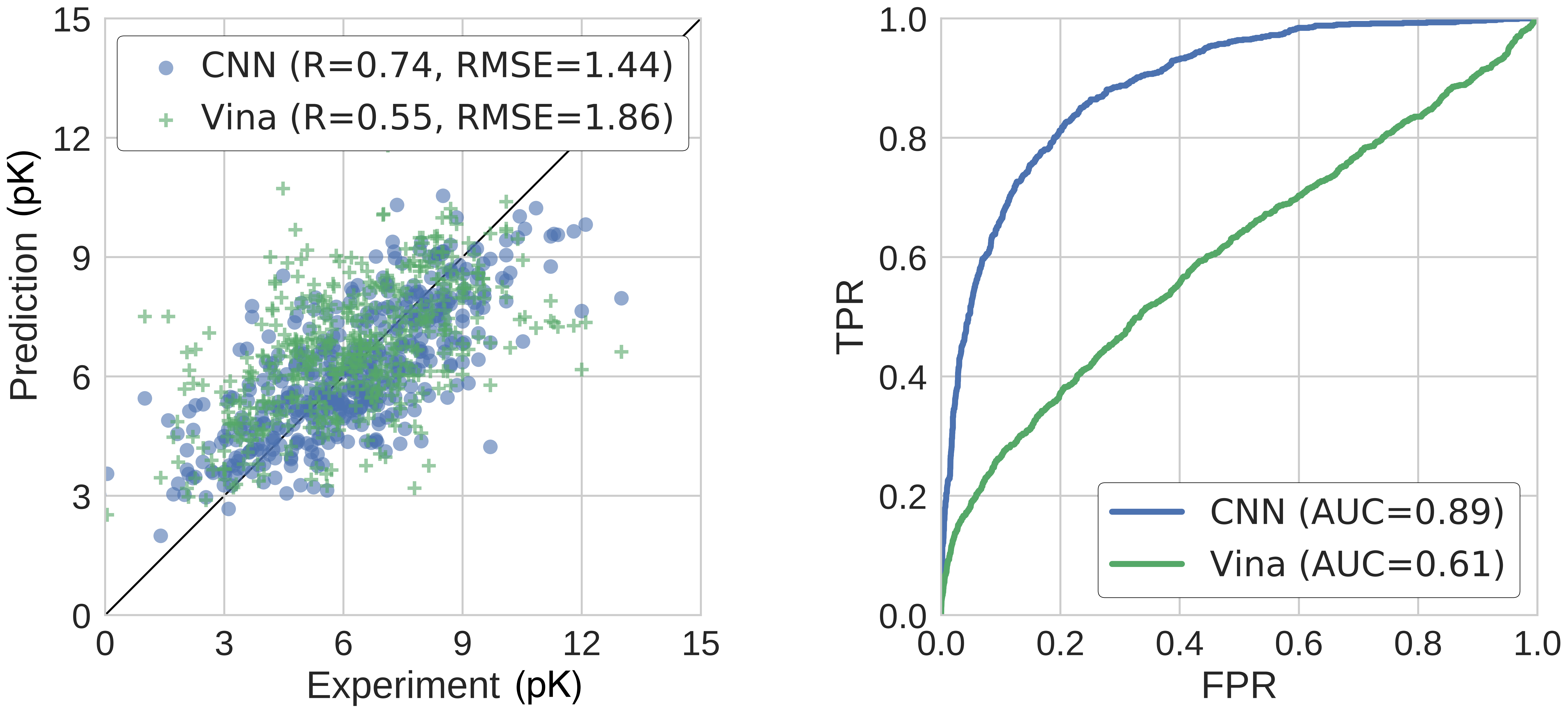
$\neq$



Model

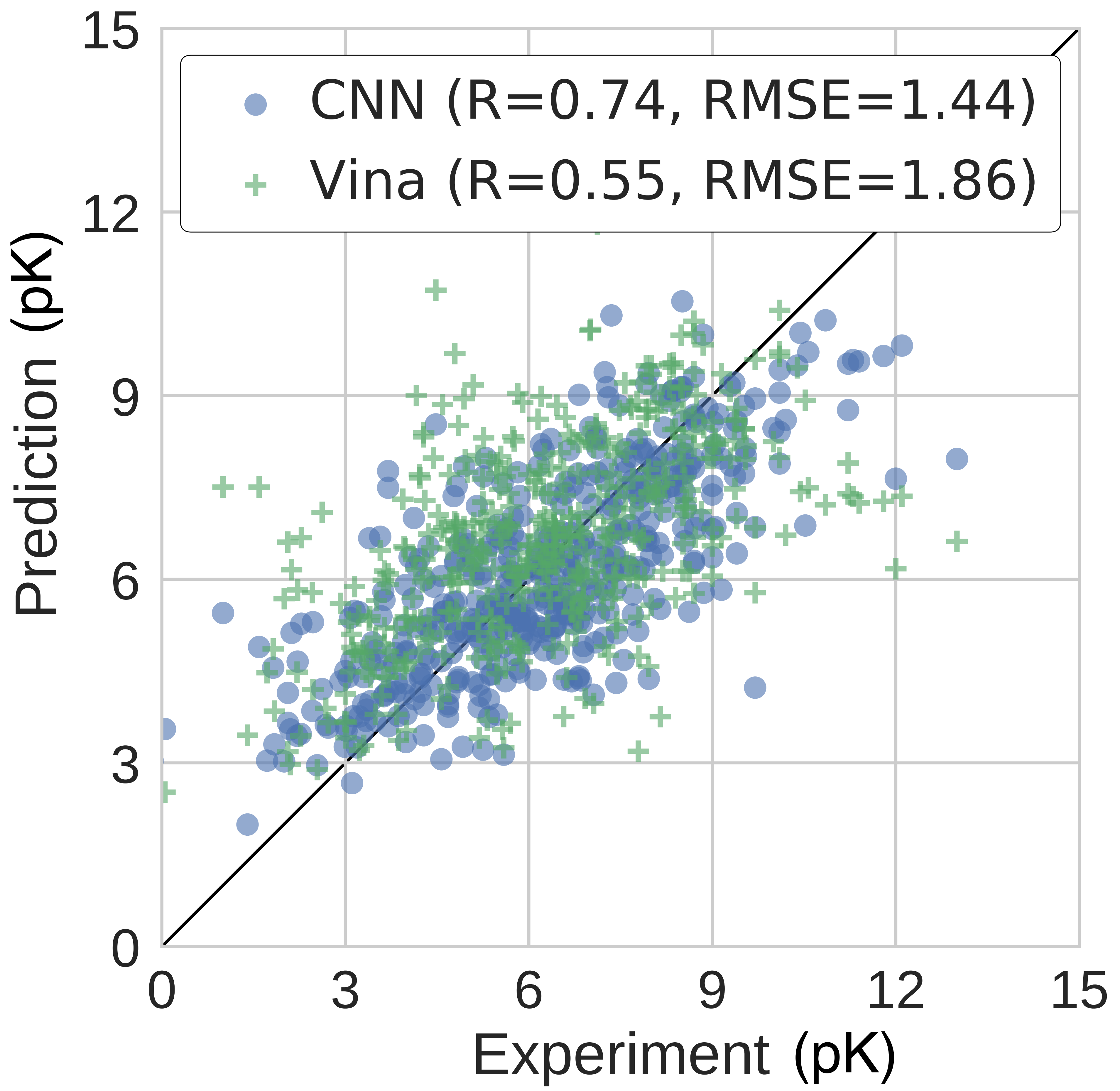

48x48x48x35
2x2x2 Max Pooling
24x24x24x35
3x3x3 Convolution
Rectified Linear Unit
24x24x24x32
2x2x2 Max Pooling
12x12x12x32
3x3x3 Convolution
Rectified Linear Unit
12x12x12x64
2x2x2 Max Pooling
6x6x6x64
3x3x3 Convolution
Rectified Linear Unit
6x6x6x128
Fully Connected
Pseudo-Huber Loss
Affinity
Fully Connected
Softmax+Logistic Loss
Pose Score

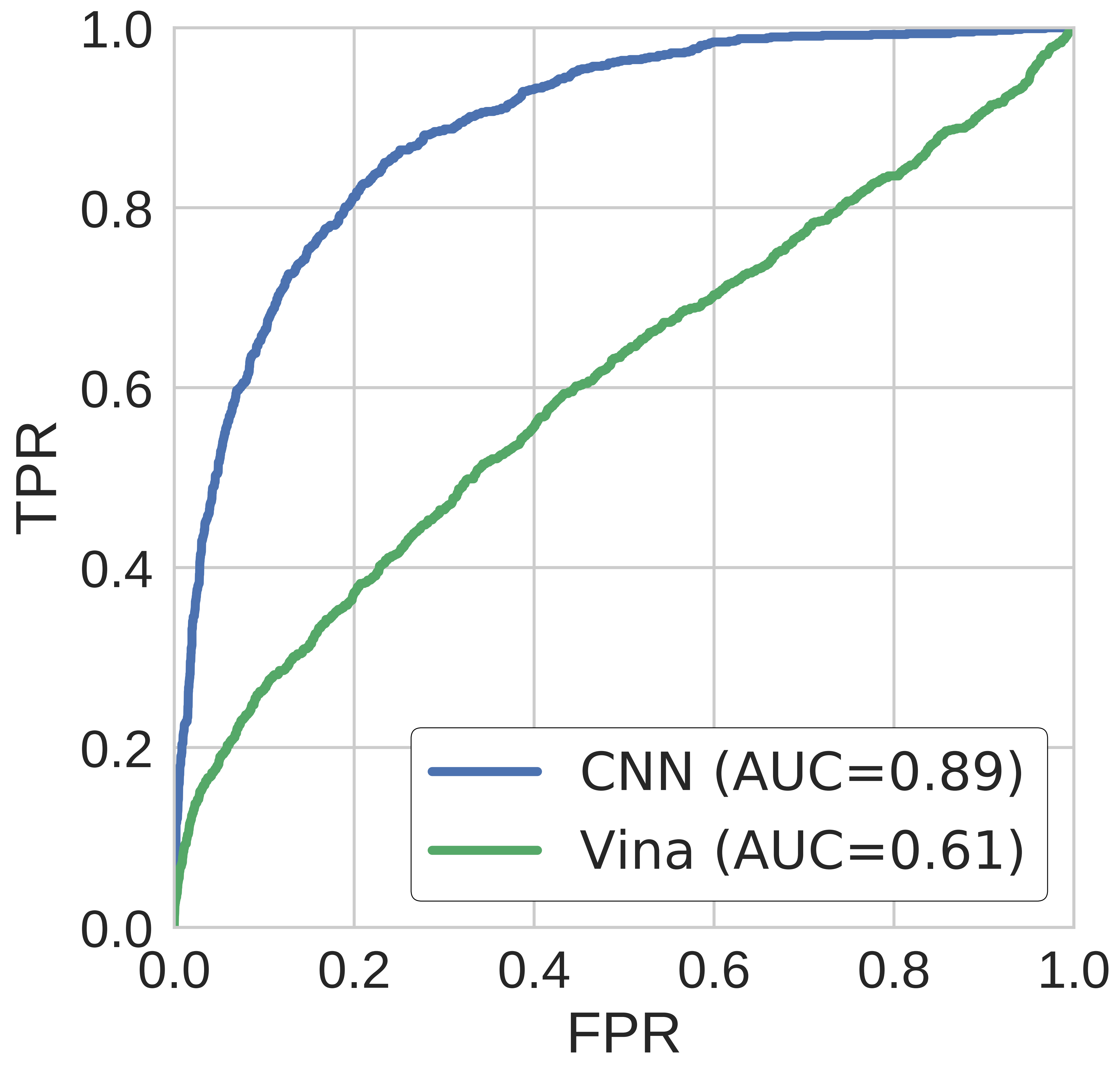
Results



Trained on PDBbind refined; tested on CSAR

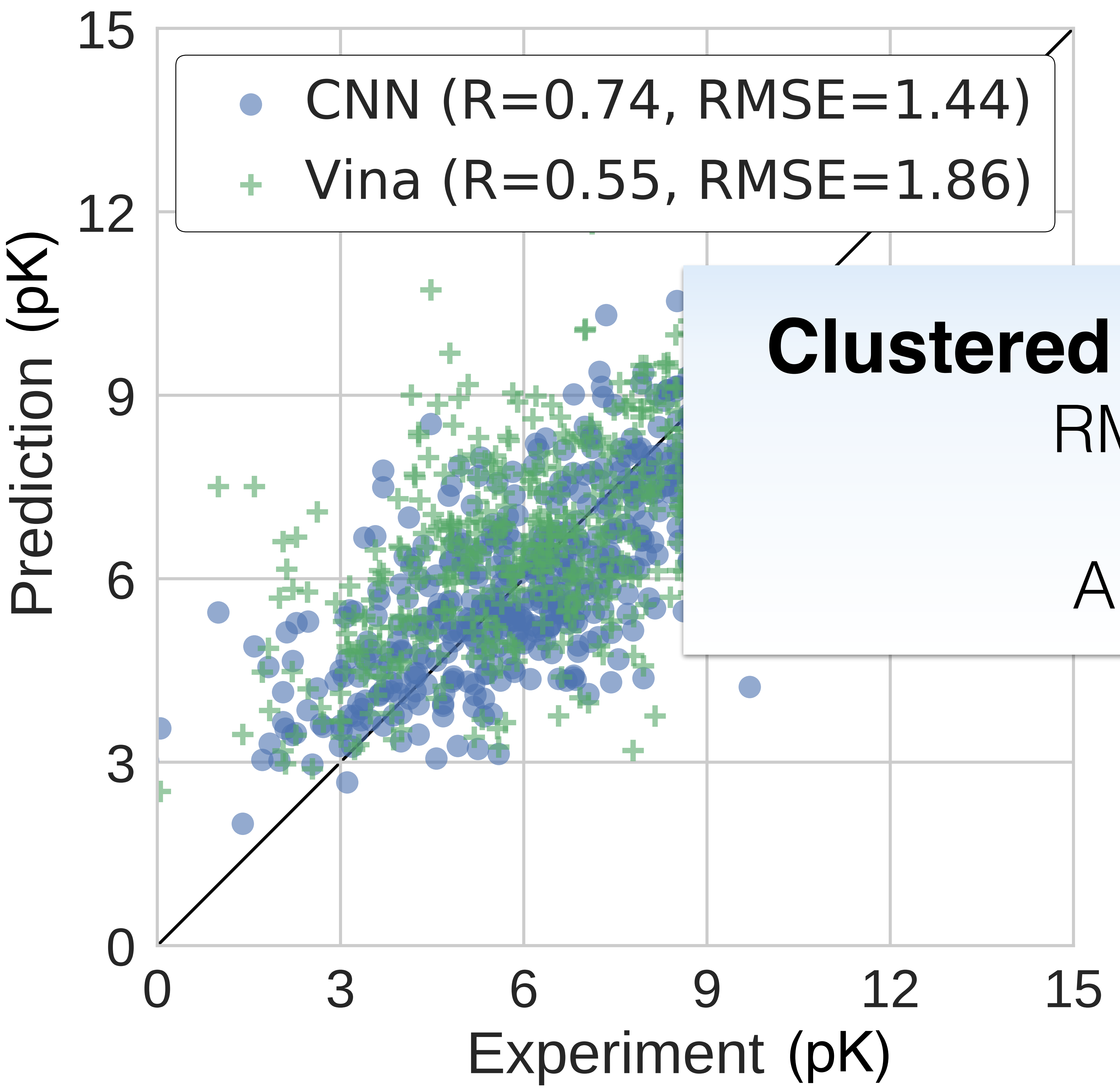
Results

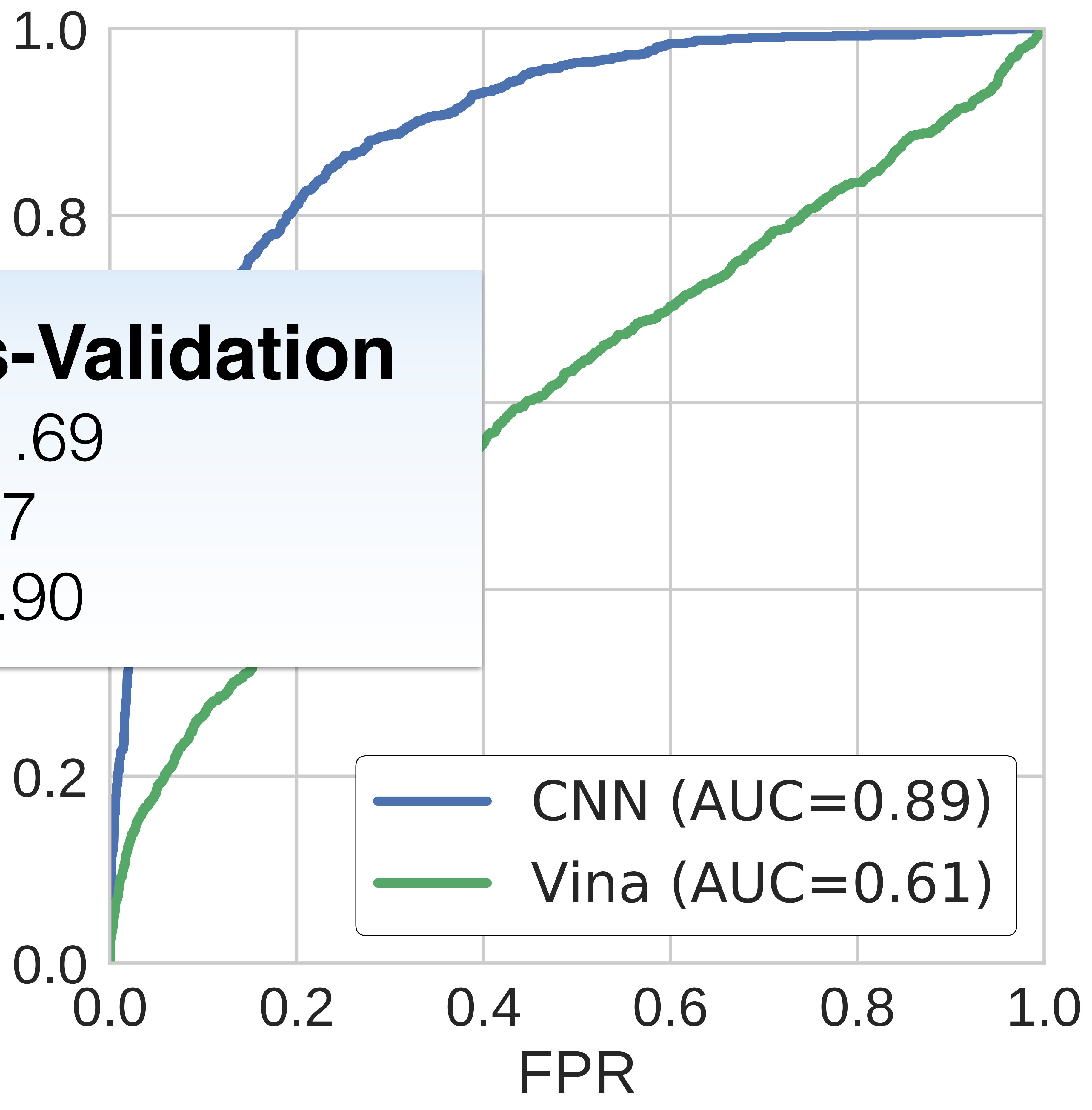




Trained on PDBbind refined; tested on CSAR

Results





Clustered Cross-Validation

RMSE = 1.69

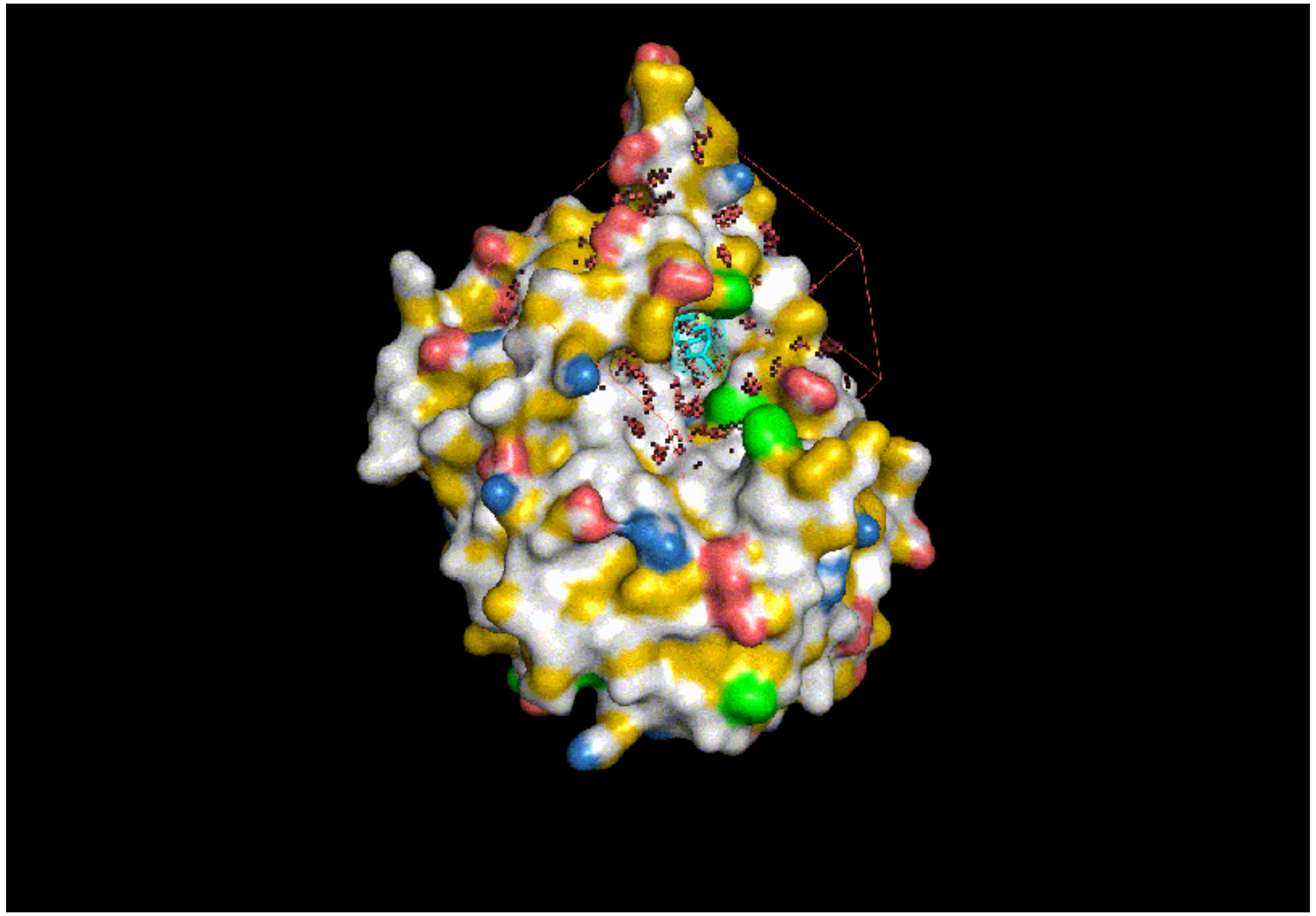
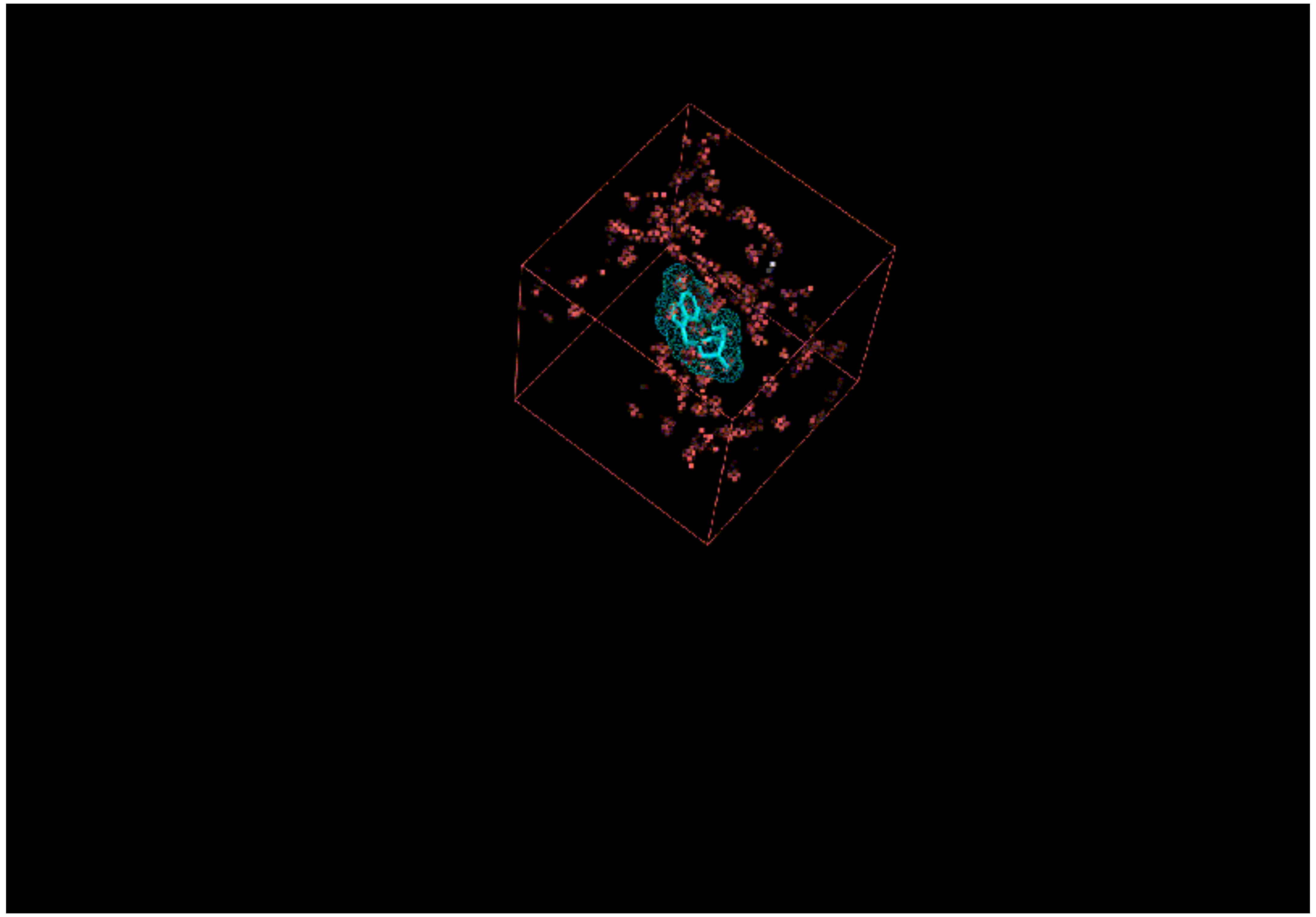
R = 0.57

AUC = 0.90

Trained on PDBbind refined; tested on CSAR

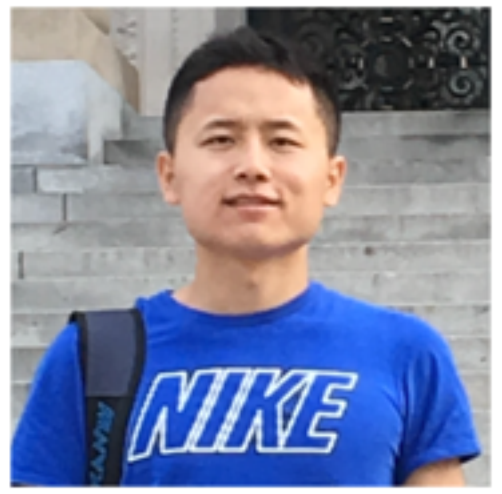
What about water?

Grid Inhomogeneous Solvation Theory

GIST analysis of 3BGS (purine nucleoside phosphorylase) active site

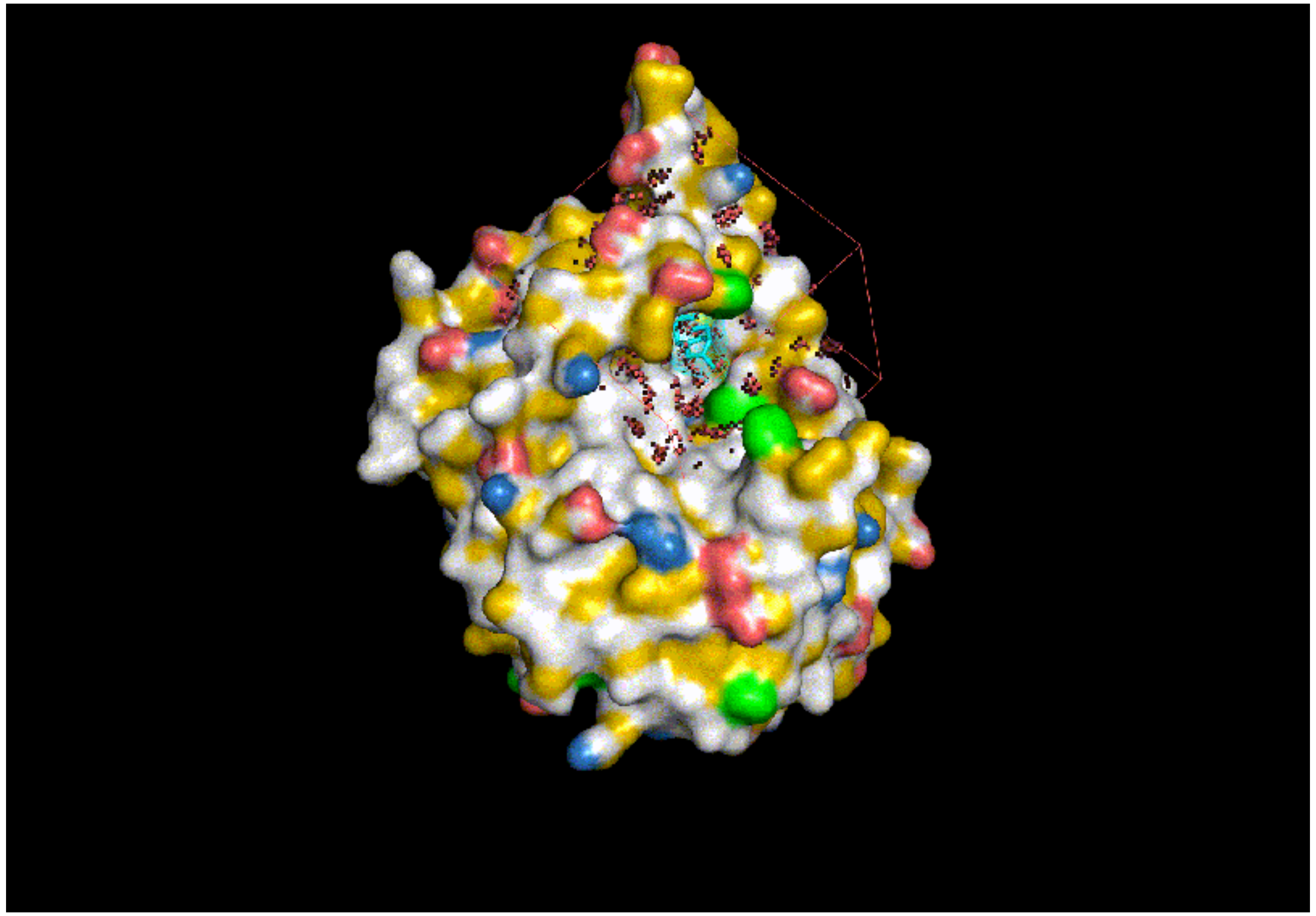

Tom Kurtzman

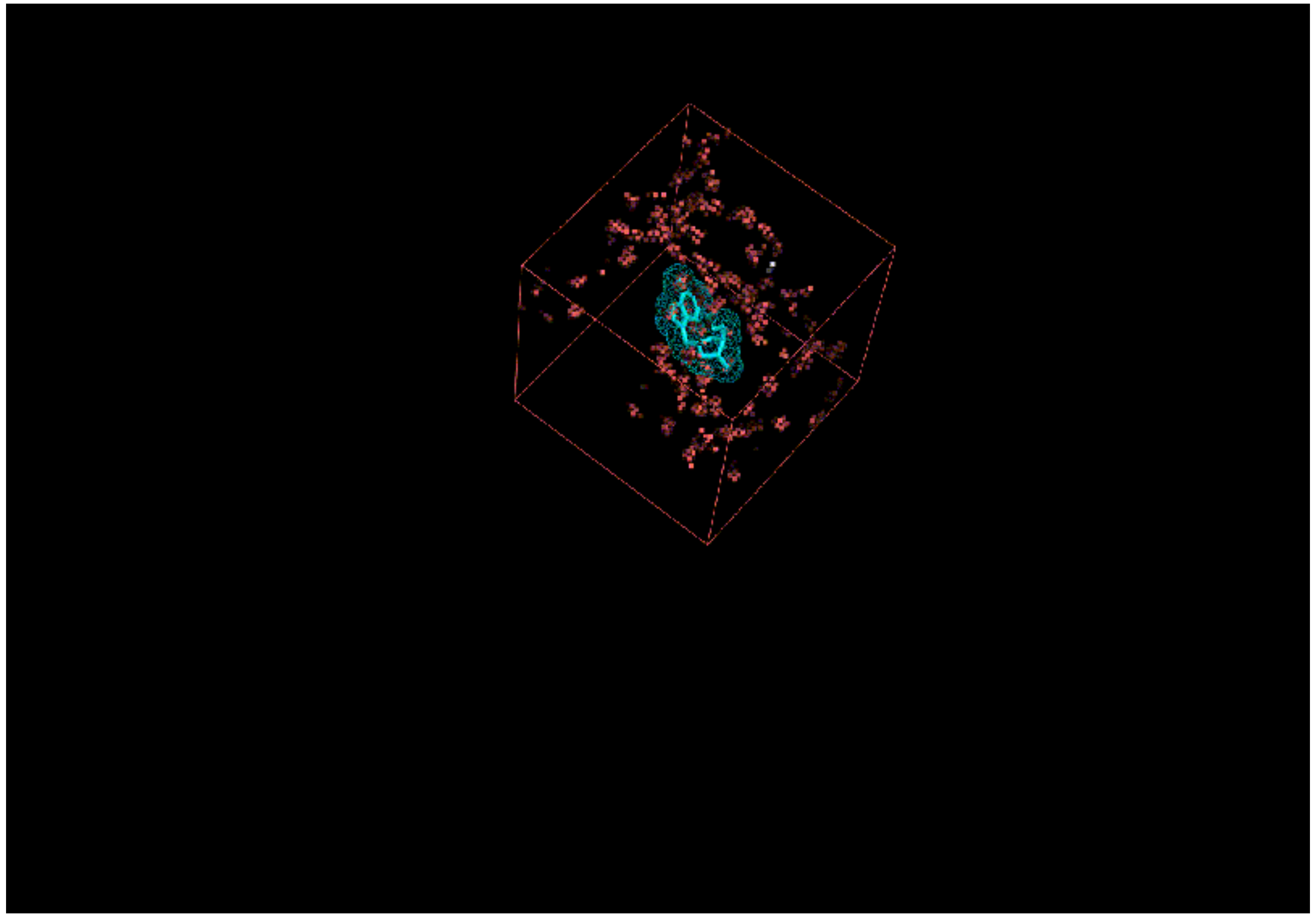

Eric Chen


Steven Ramsey


Anthony Cruz-Balberdy

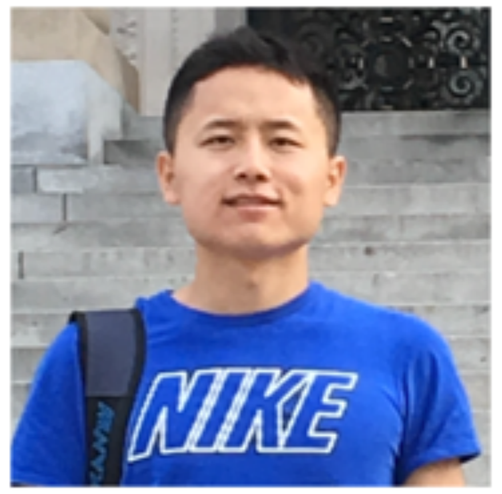
Grid Inhomogeneous Solvation Theory





GIST analysis of 3BGS (purine nucleoside phosphorylase) active site

Tom Kurtzman

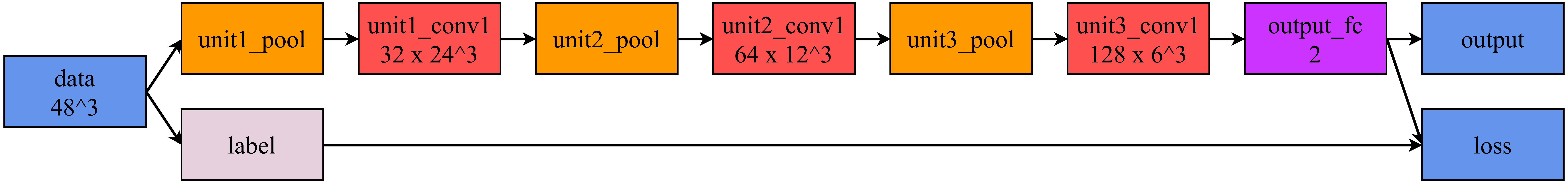


Eric Chen

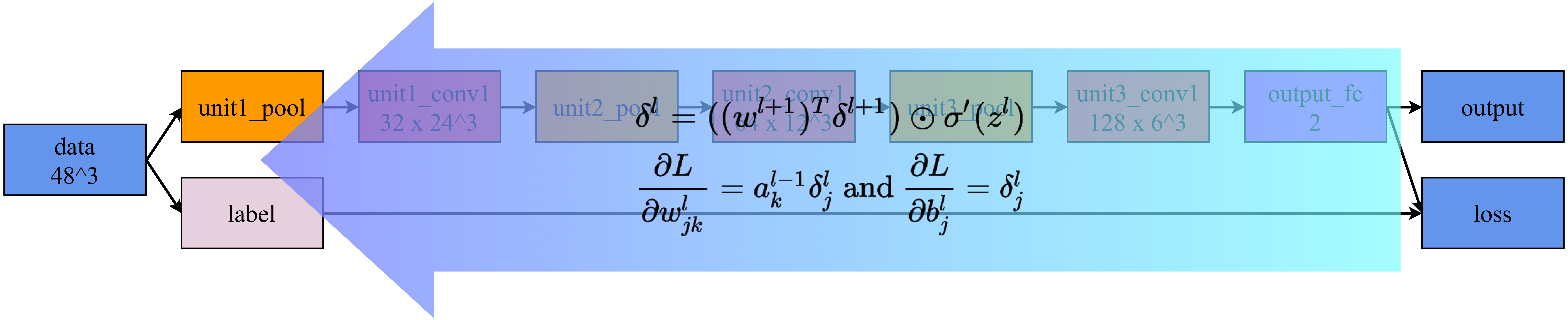
Steven Ramsey

Anthony Cruz-Balberdy

Beyond Scoring



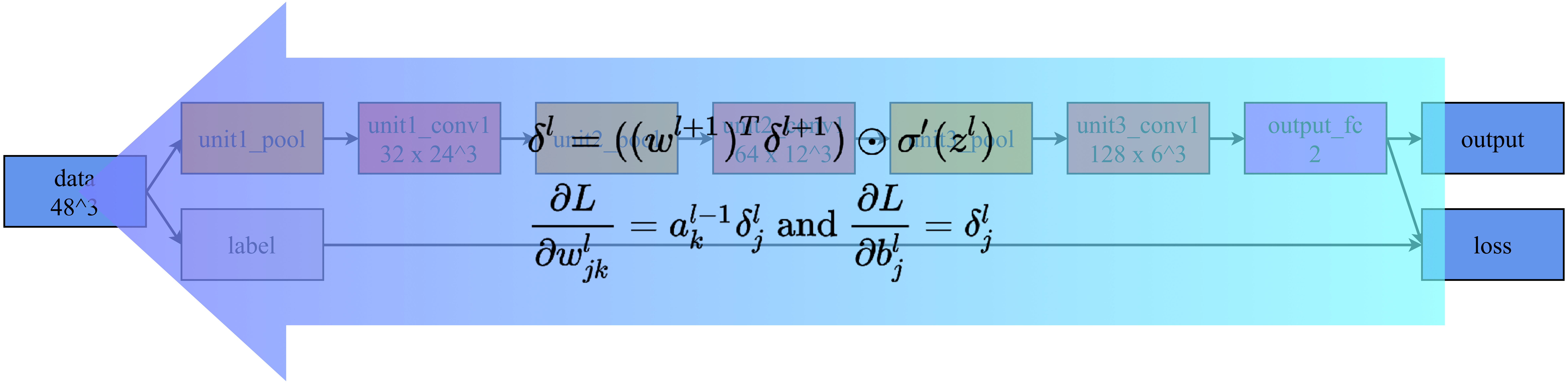
Beyond Scoring



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

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

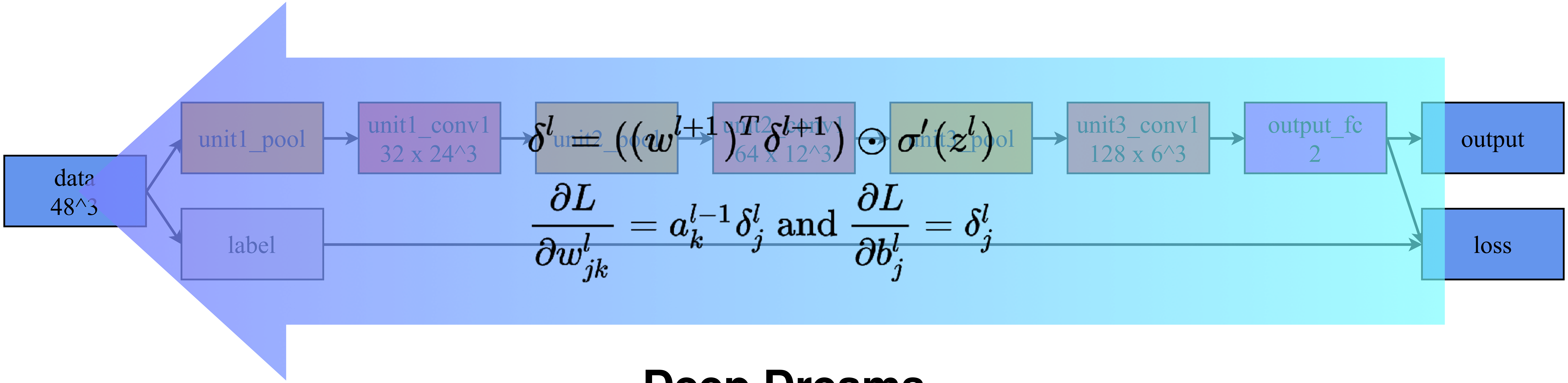
Beyond Scoring



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

$$\frac{\partial L}{\partial w^l_{jk}} = a_k^{l-1} \delta_j^l \text{ and } \frac{\partial L}{\partial b^l_j} = \delta_j^l$$

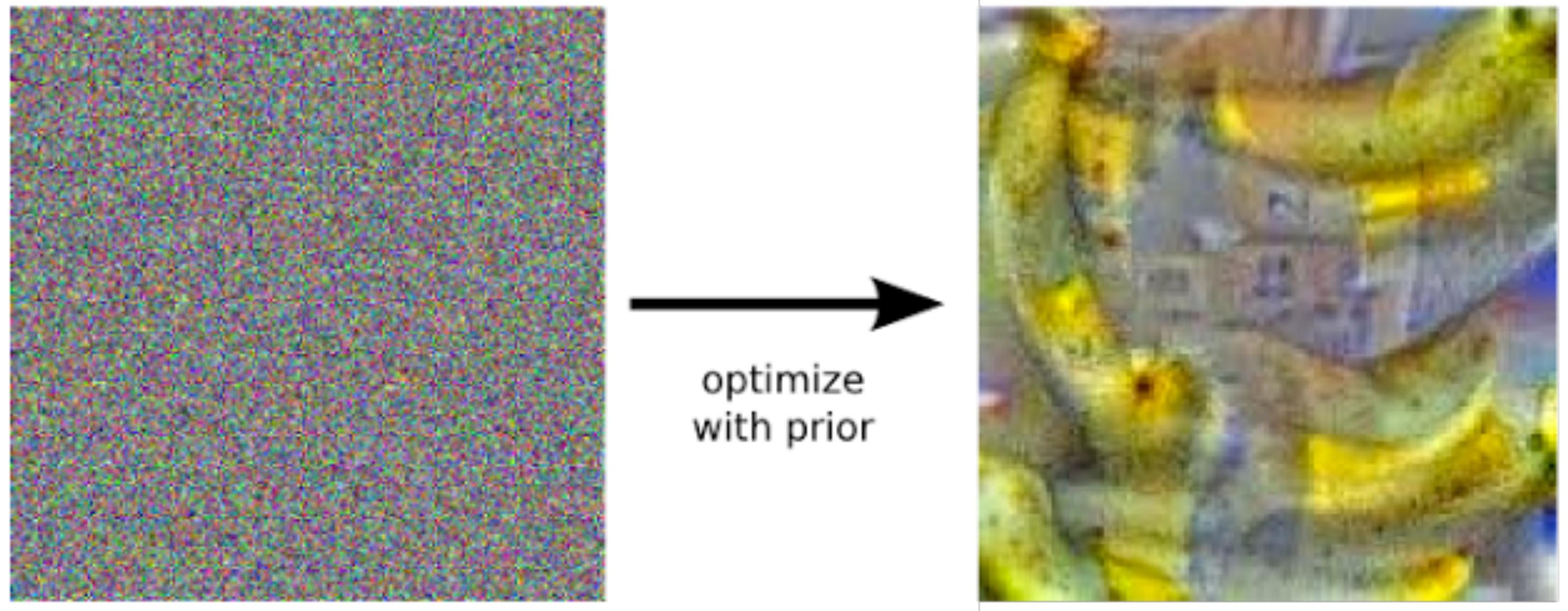
Beyond Scoring



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

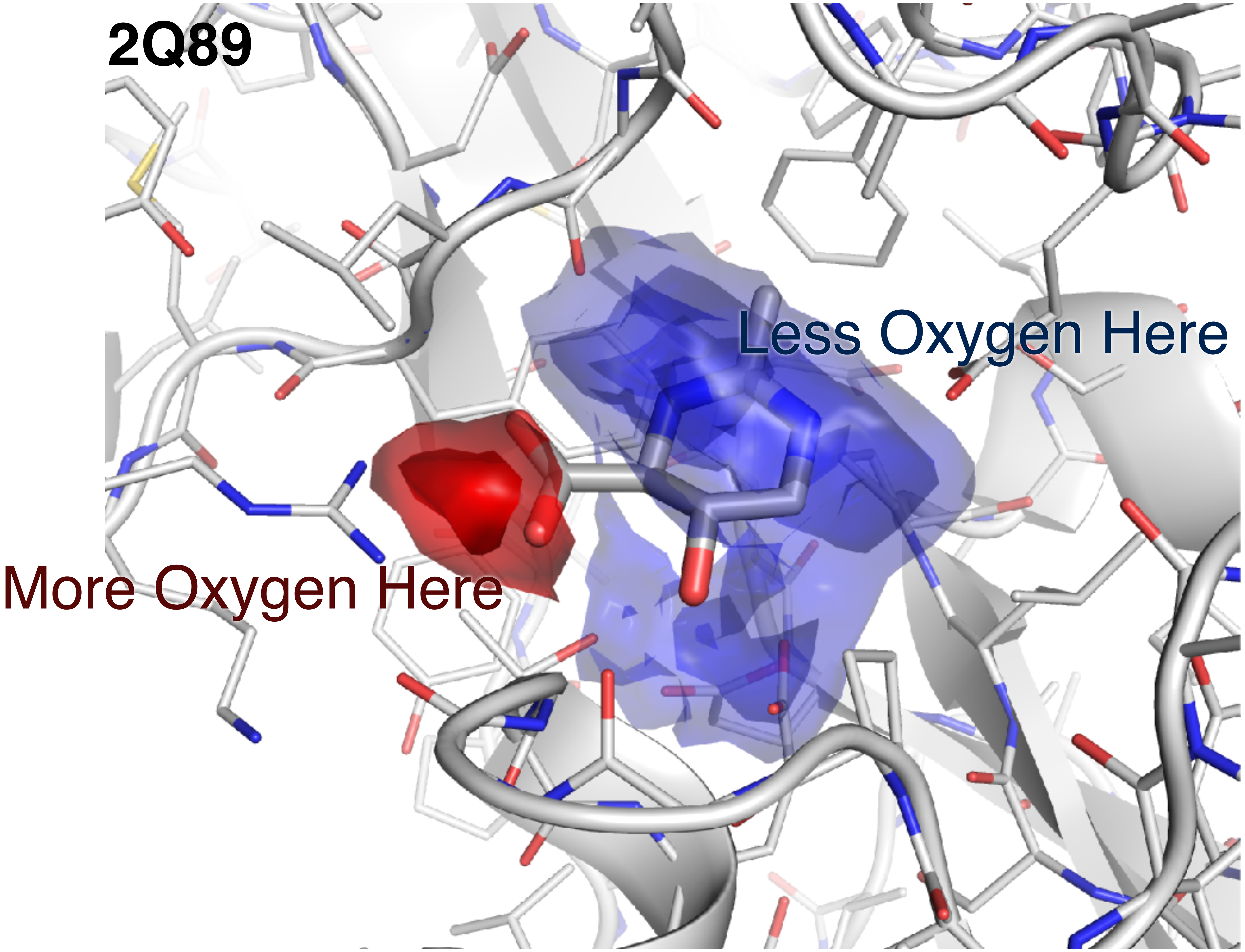
$$\frac{\partial L}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j \text{ and } \frac{\partial L}{\partial b^l_j} = \delta^l_j$$

Deep Dreams

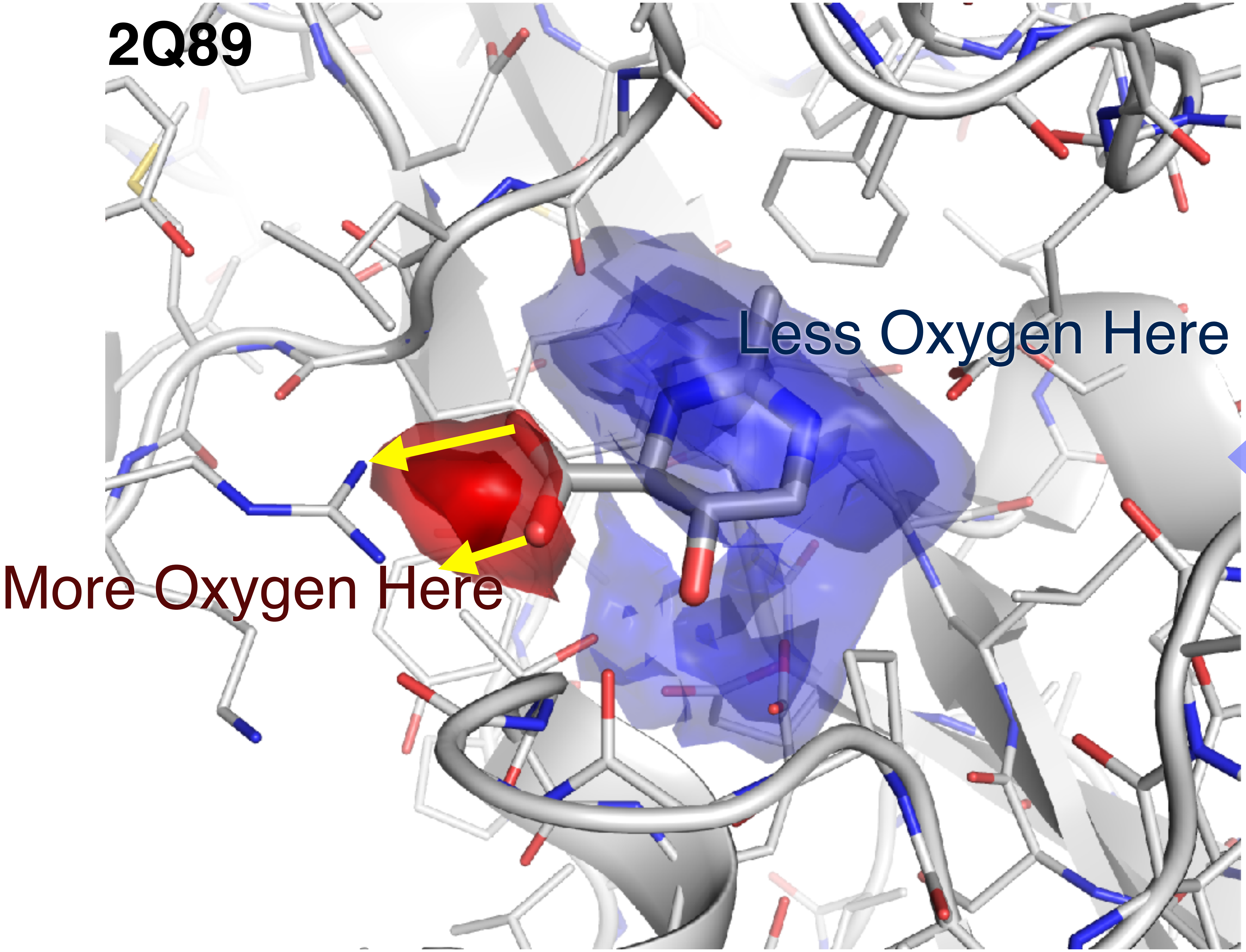


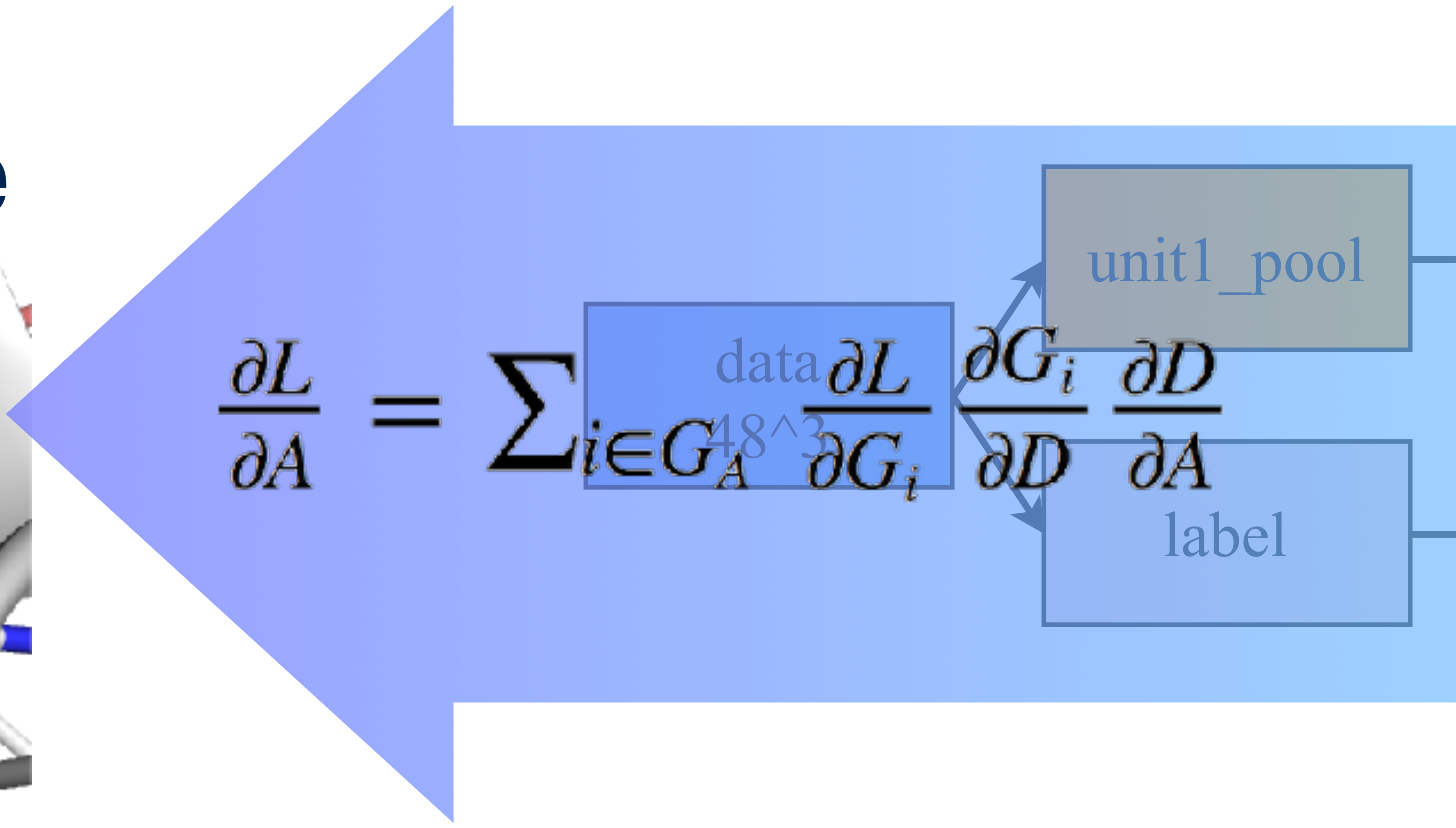
https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Beyond Scoring

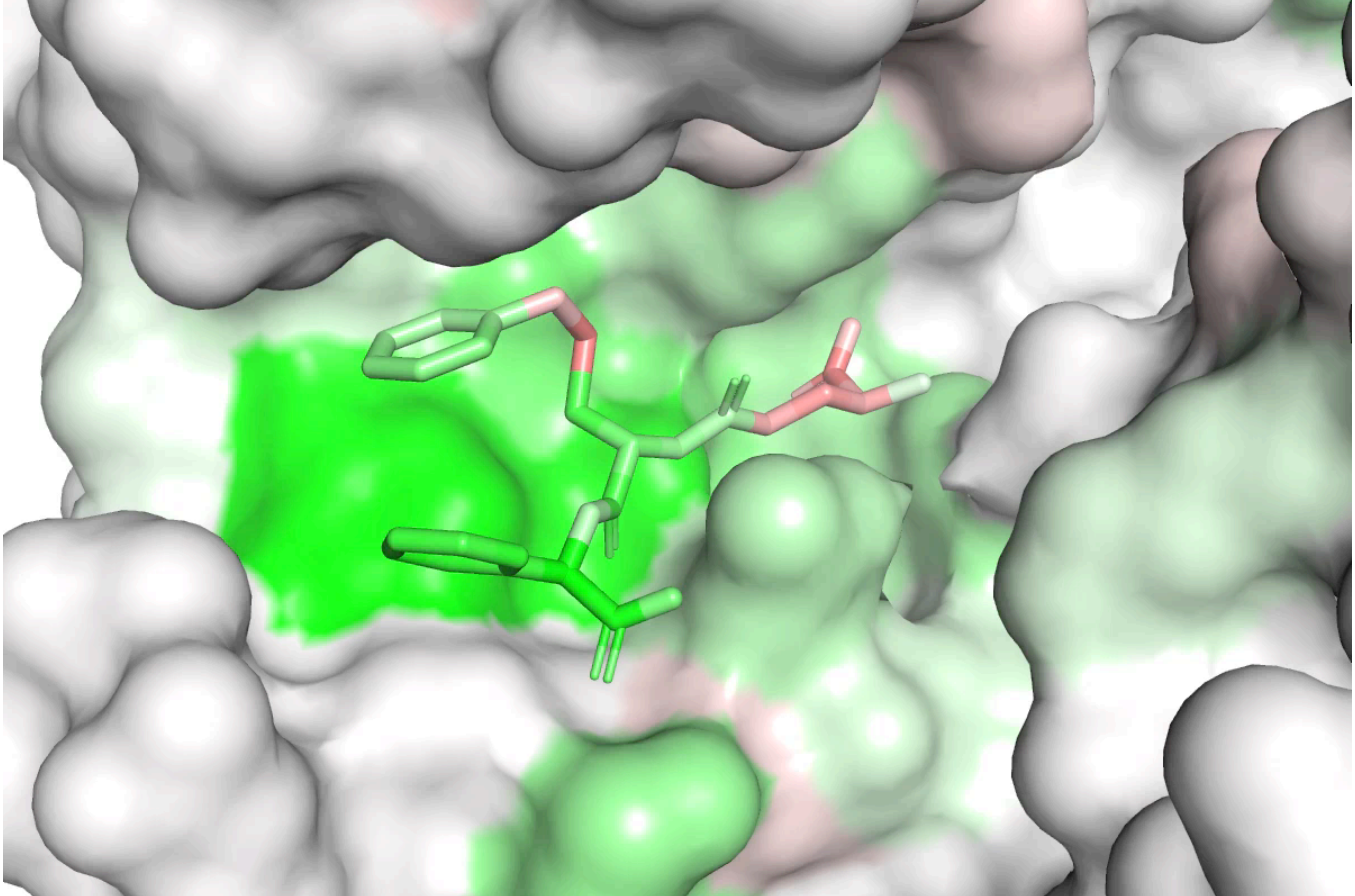


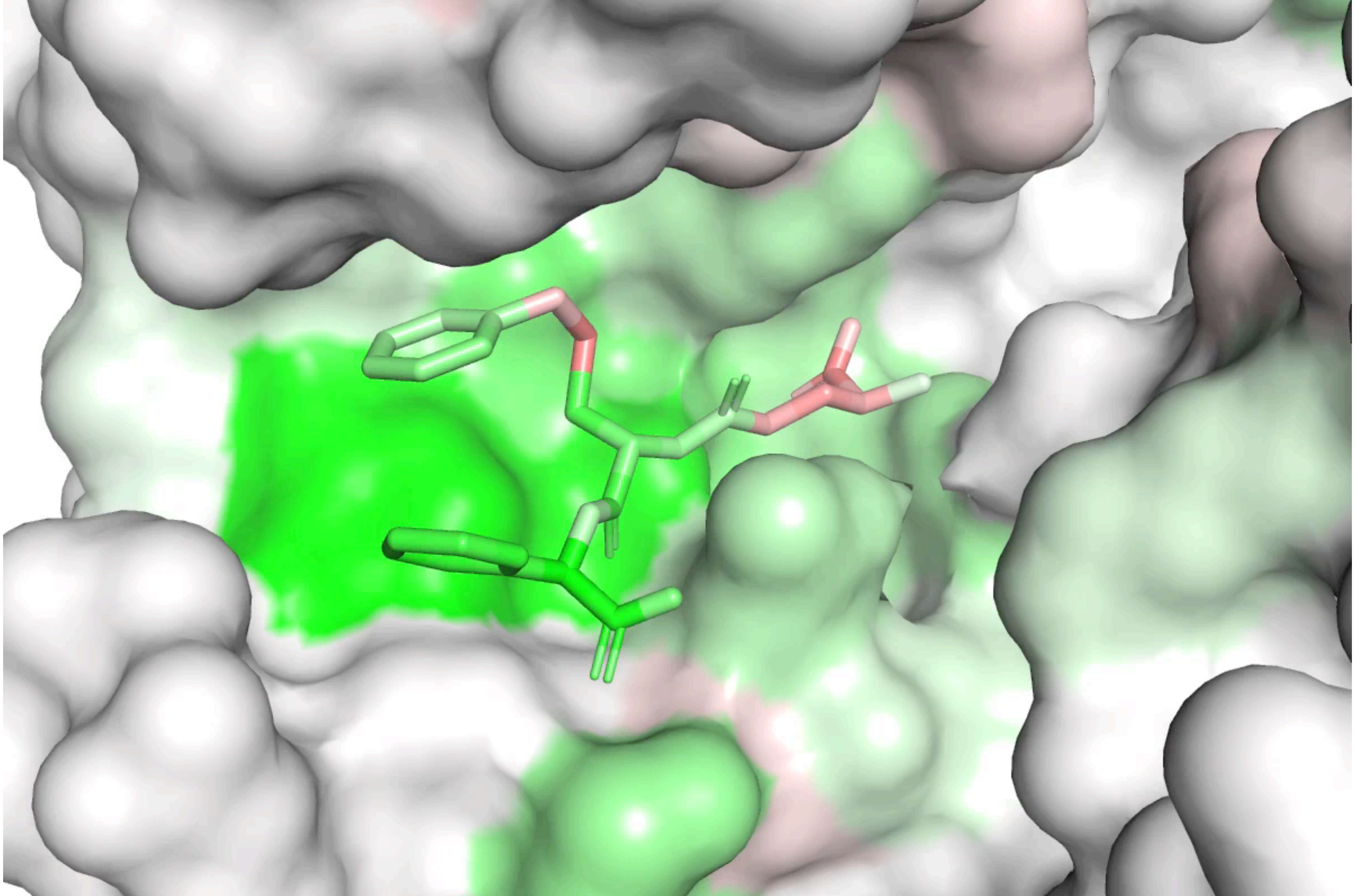
Beyond Scoring

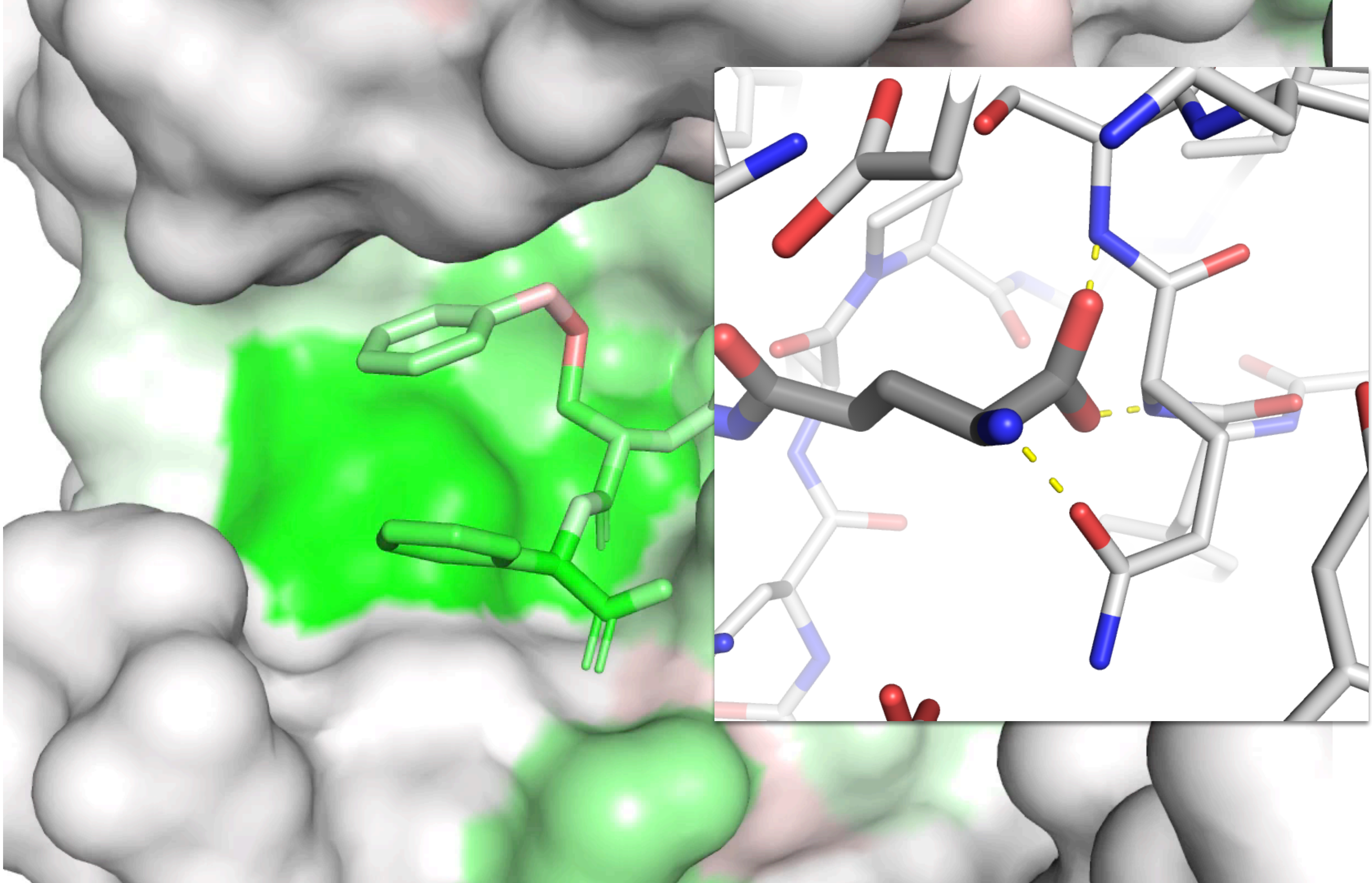


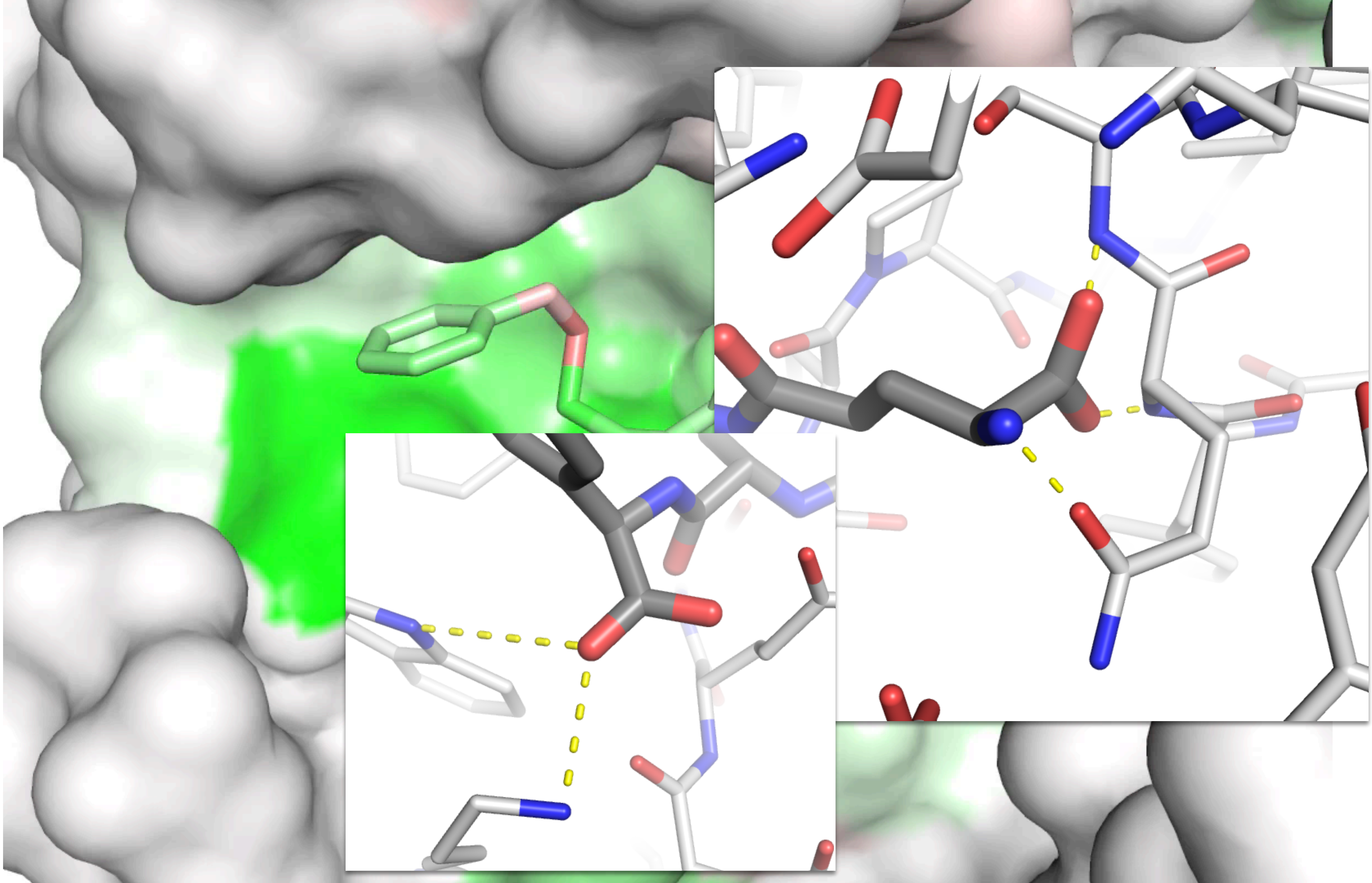


$$\frac{\partial L}{\partial A} = \sum_{i \in G_A} \frac{\partial L}{\partial G_i} \frac{\partial G_i}{\partial D} \frac{\partial D}{\partial A}$$

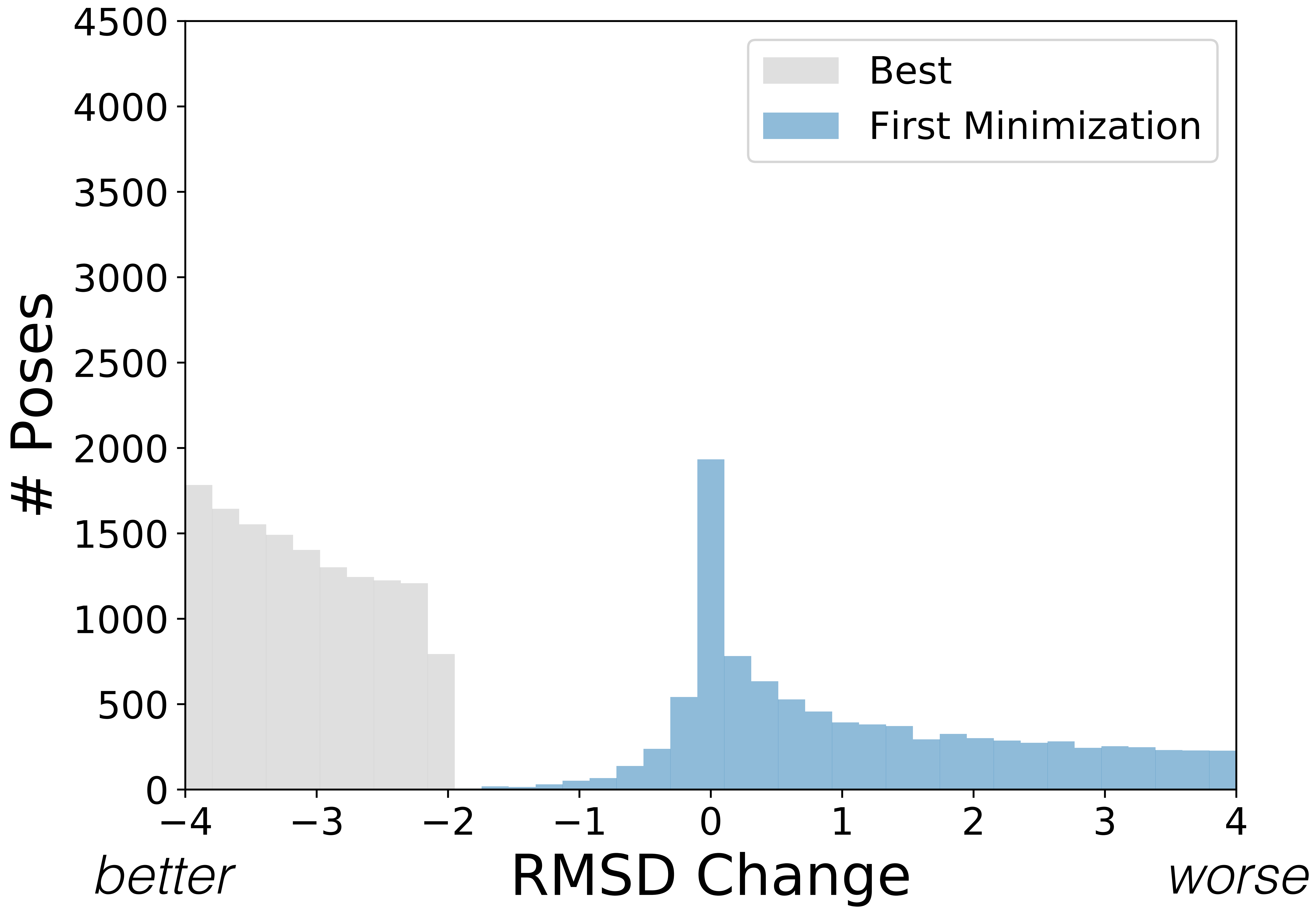




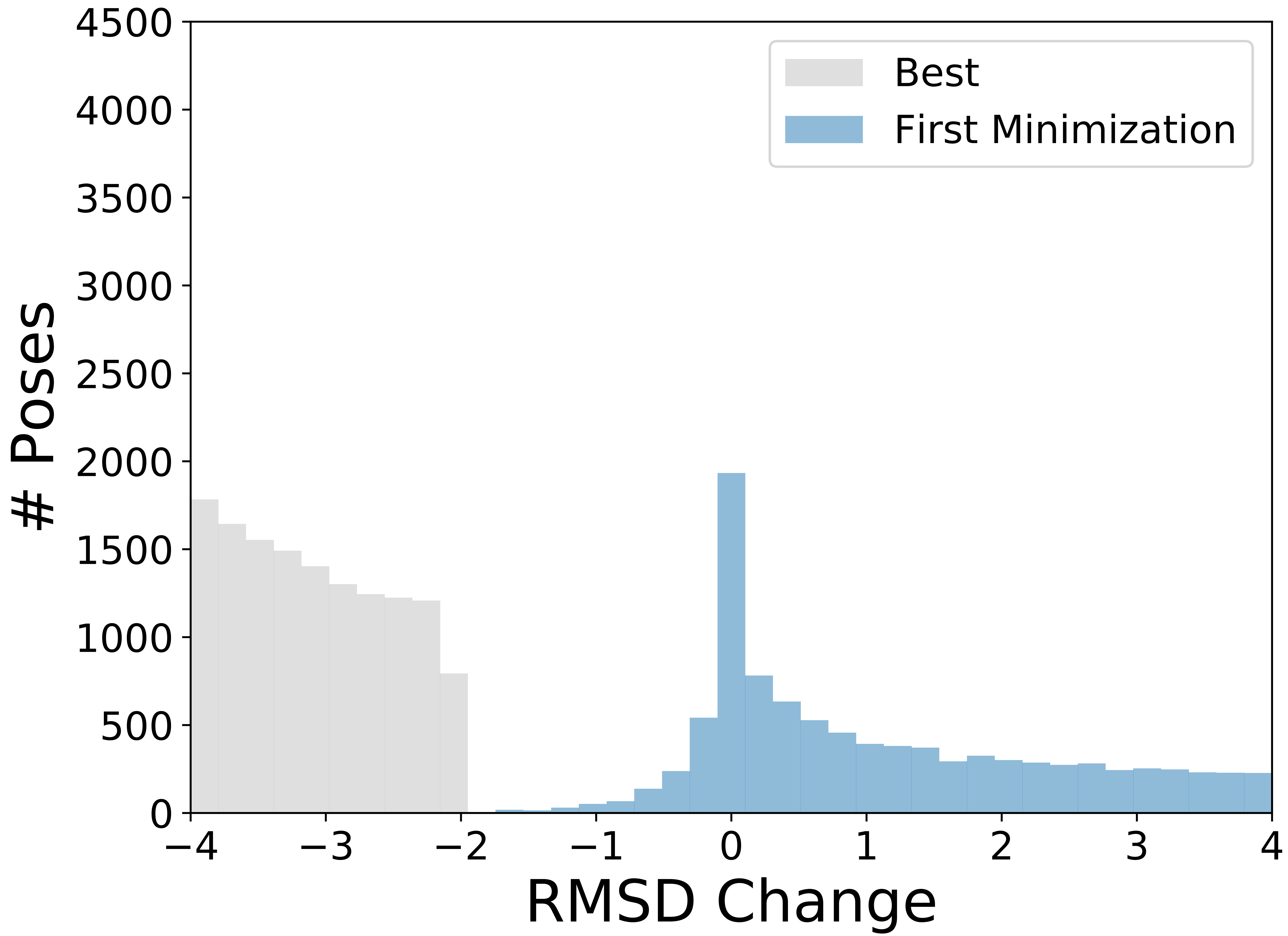




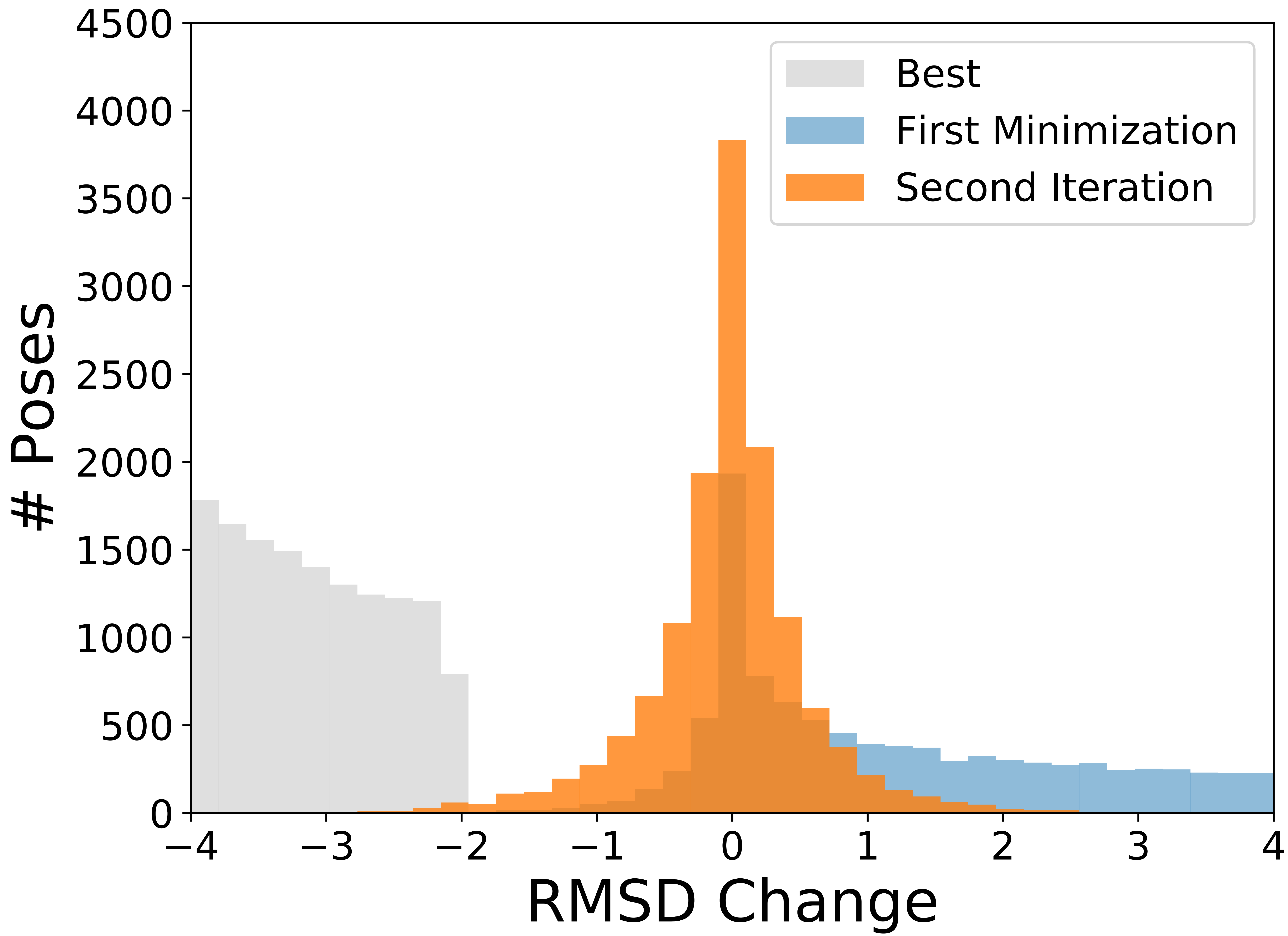
Minimizing Low RMSD Poses



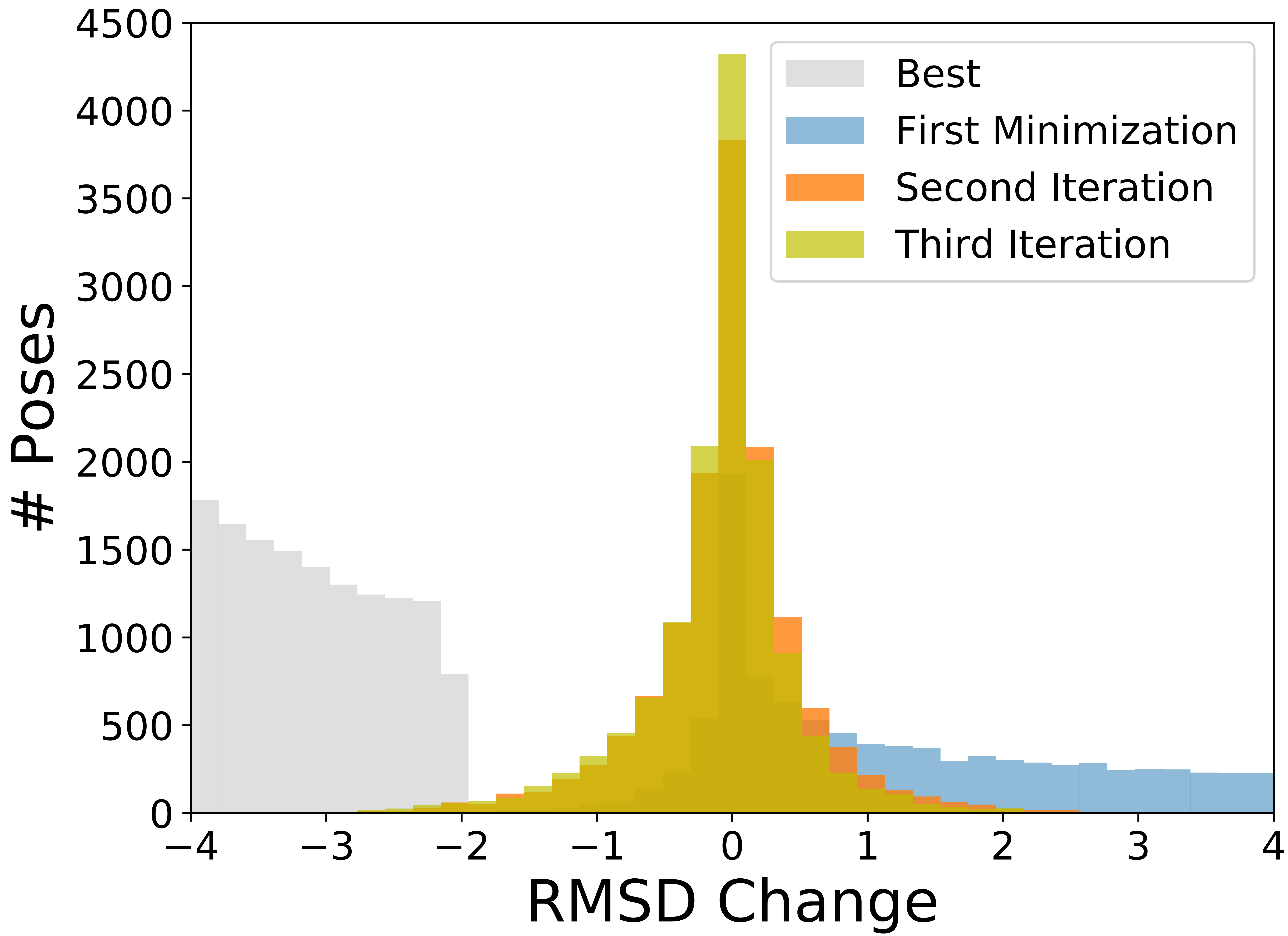
Iterative Refinement



Iterative Refinement



Iterative Refinement



Docking

vina/smina/gnina

Sampling

MCMC

MCMC

MCMC

MCMC

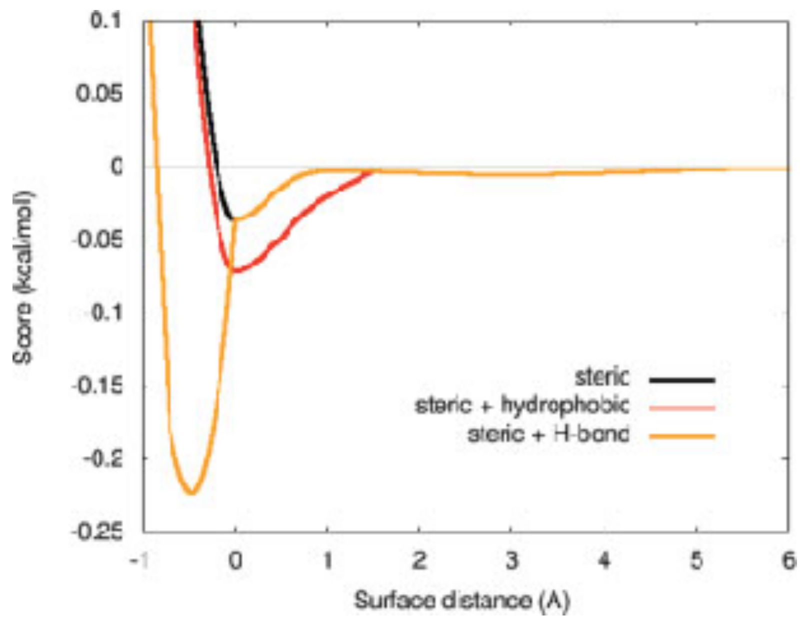
MCMC

⋮

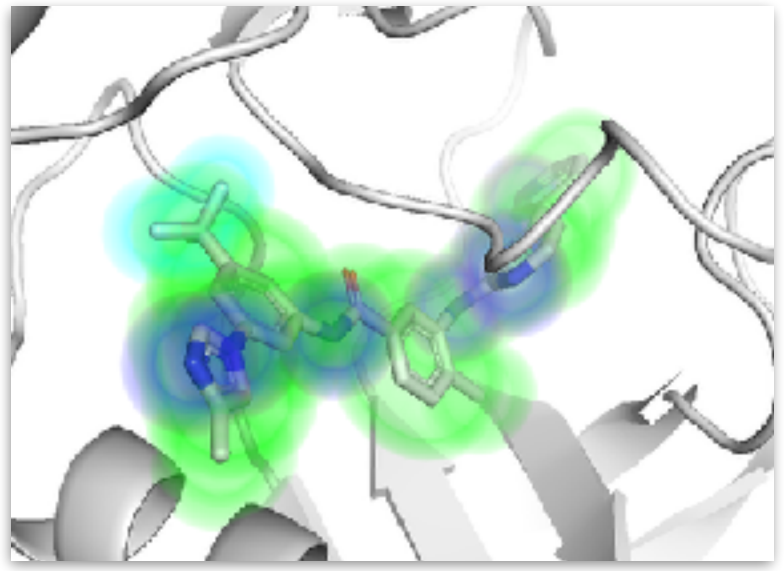
N (50) independent Monte Carlo chains
Scored with grid-accelerated Vina
Best identified pose retained

best poses

Refinement



Vina



CNN

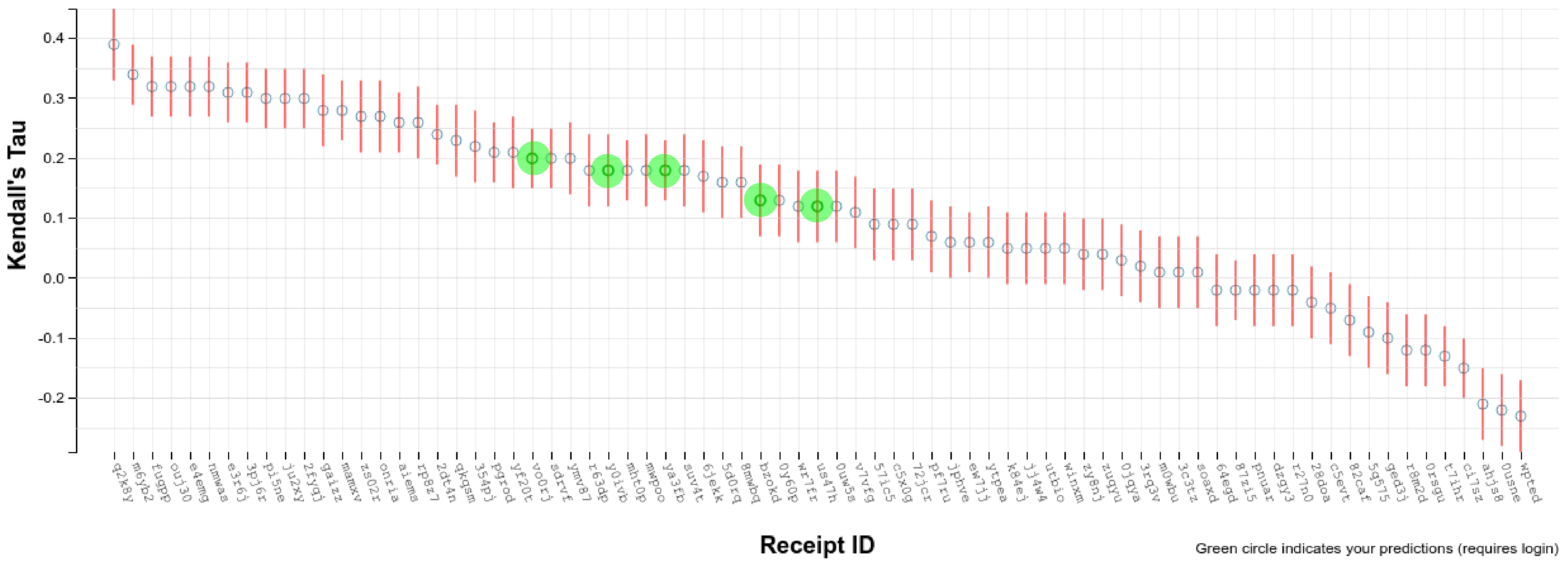
Rescoring

CNN
pose
affinity

D3R Grand Challenge 3

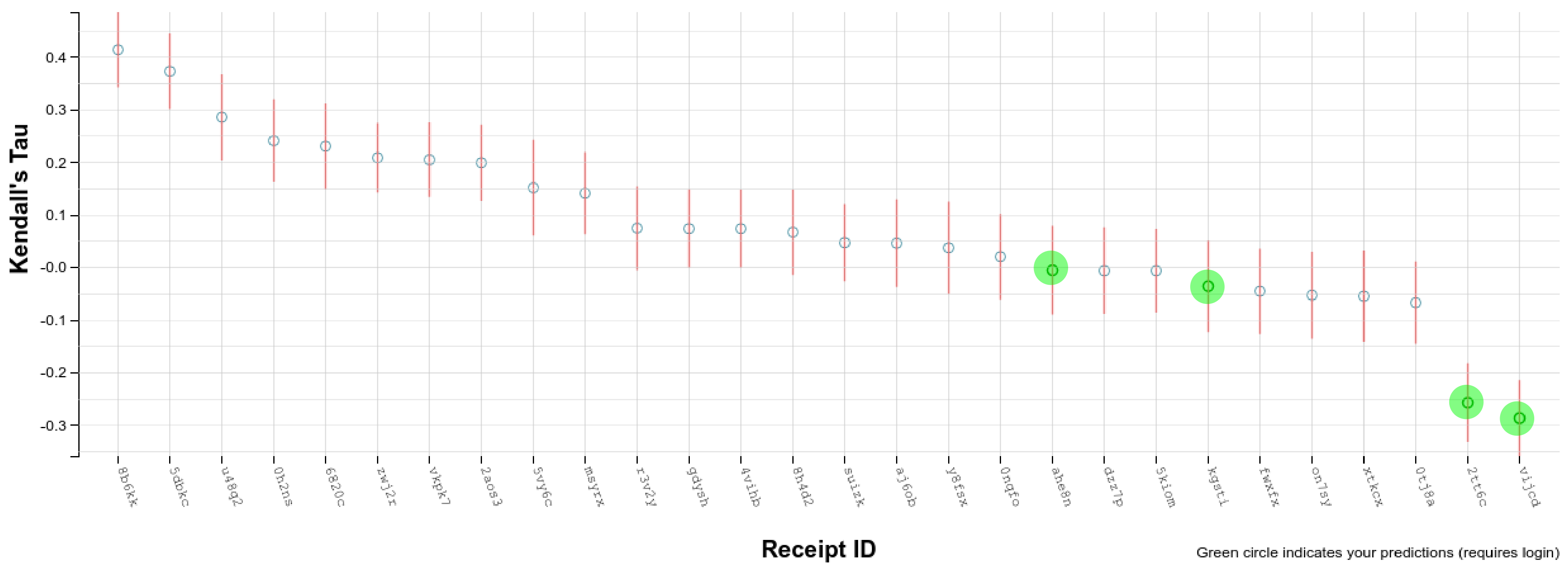
Grand Challenge 3 - CatS_stage2

Affinity Ranking - Kendall's Tau



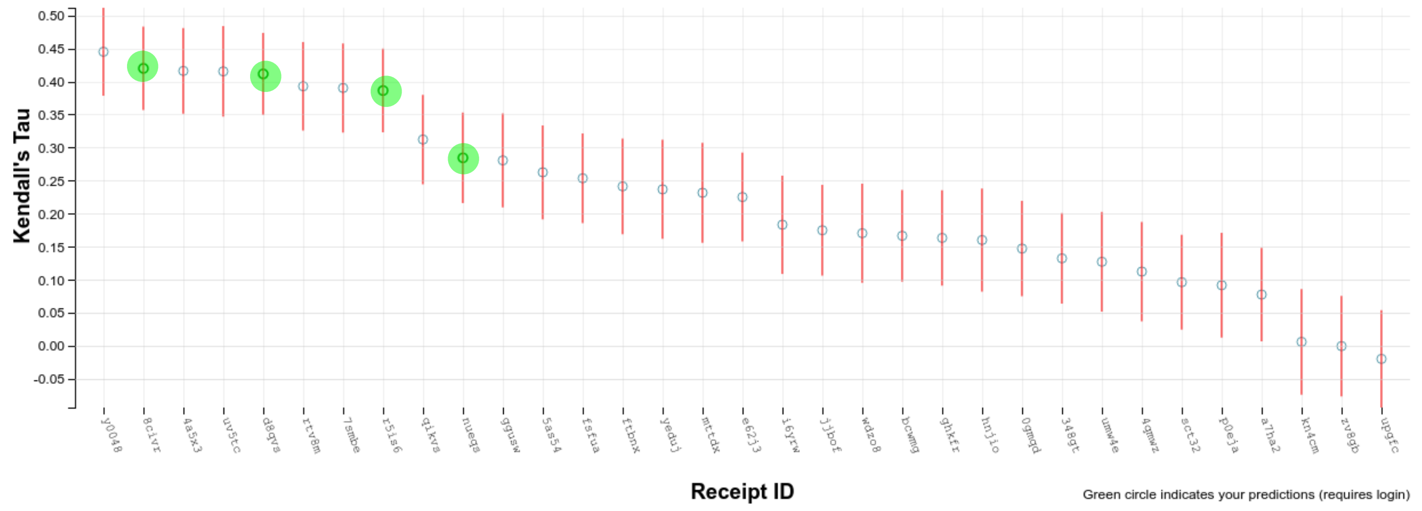
Grand Challenge 3 - p38a

Affinity Ranking - Kendall's Tau



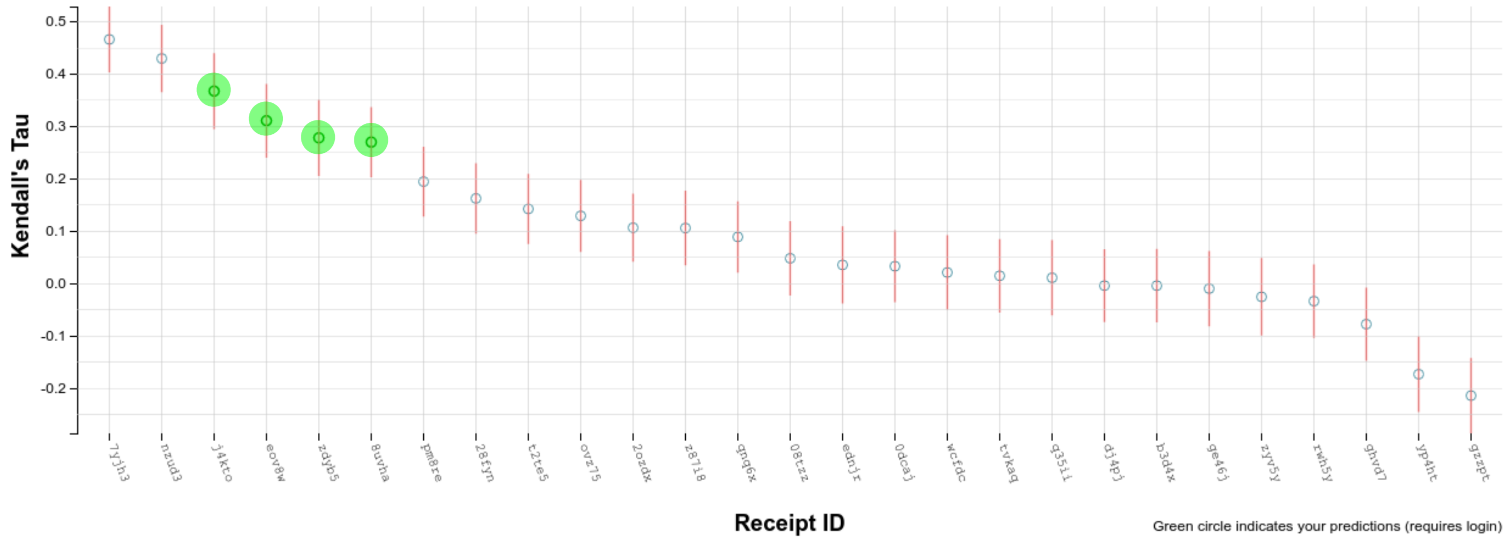
Grand Challenge 3 - VEGFR2

Affinity Ranking - Kendall's Tau



Grand Challenge 3 - JAK2_SC2

Affinity Ranking - Kendall's Tau



Grand Challenge 3 - TIE2

Affinity Ranking - Kendall's Tau

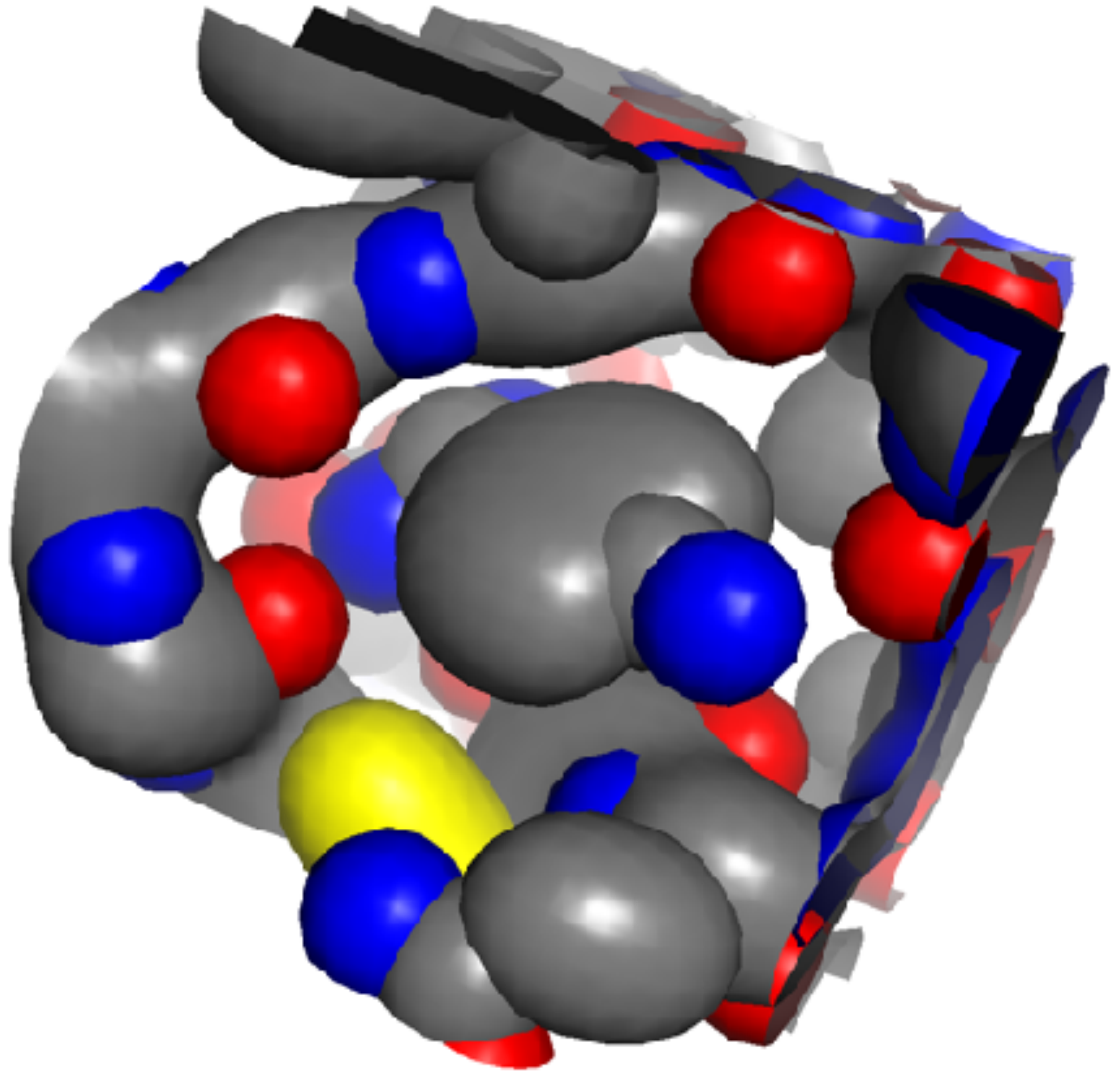
Grand Challenge 3

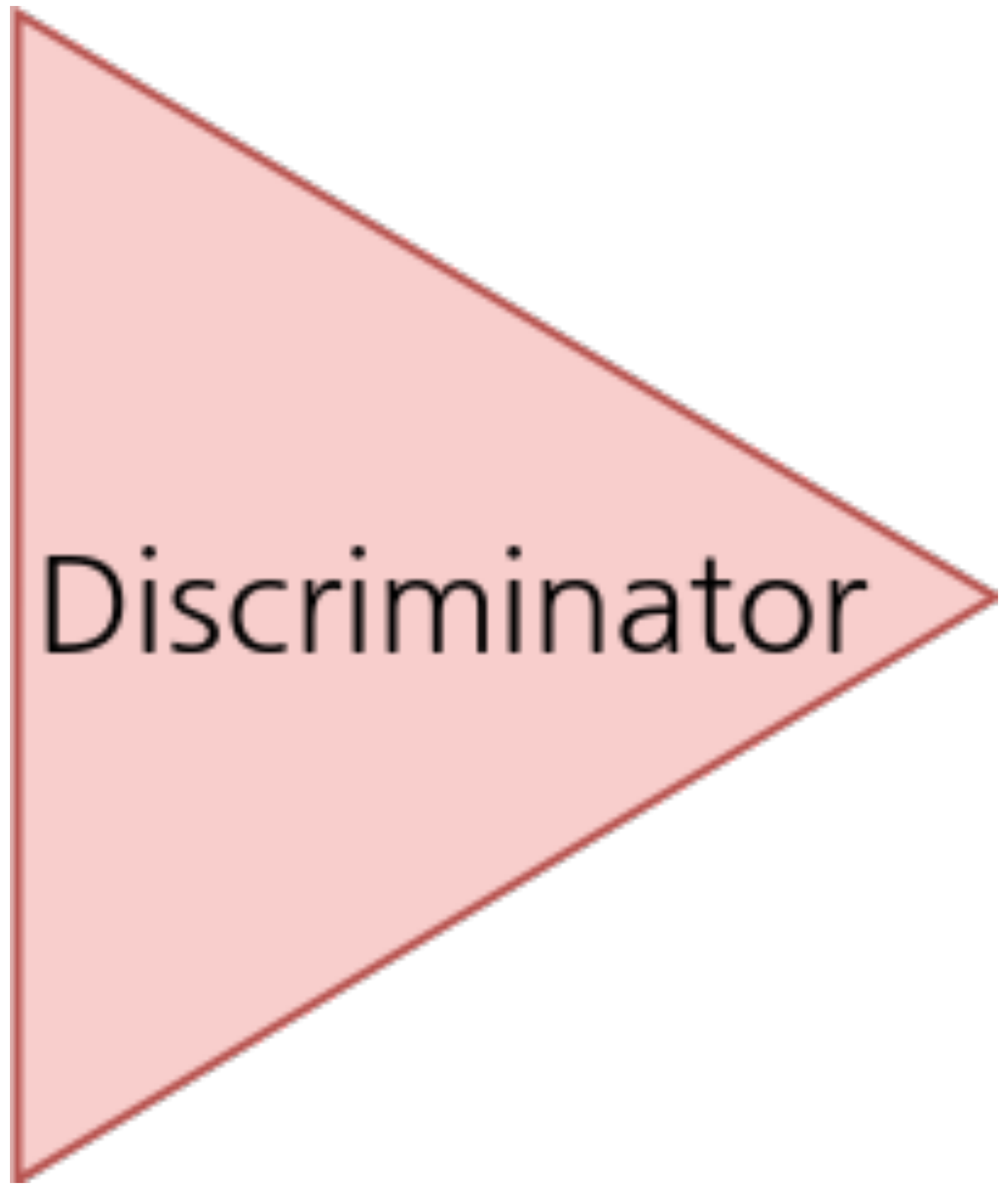
Spearman Correlation

	cnn_docked_affinity	cnn_rescore_affinity	cnn_docked_scoring	cnn_rescore_scoring	vina
cat	0.0701	0.154	-0.0351	0.178	**0.179**
p38a	-0.0784	-0.116	-0.329	-0.305	**-0.0631**
vegfr2	0.366	**0.484**	0.434	0.448	0.414
jak2	**0.428**	0.338	0.39	0.27	0.106
jak2_sub3	**0.68**	0.369	-0.372	0.159	-0.633
tie2	0.648	**0.835**	0.136	-0.078	0.561
abl1	0.634	**0.745**	0.005	0.182	0.713

and now for something completely different…

Discriminative Models





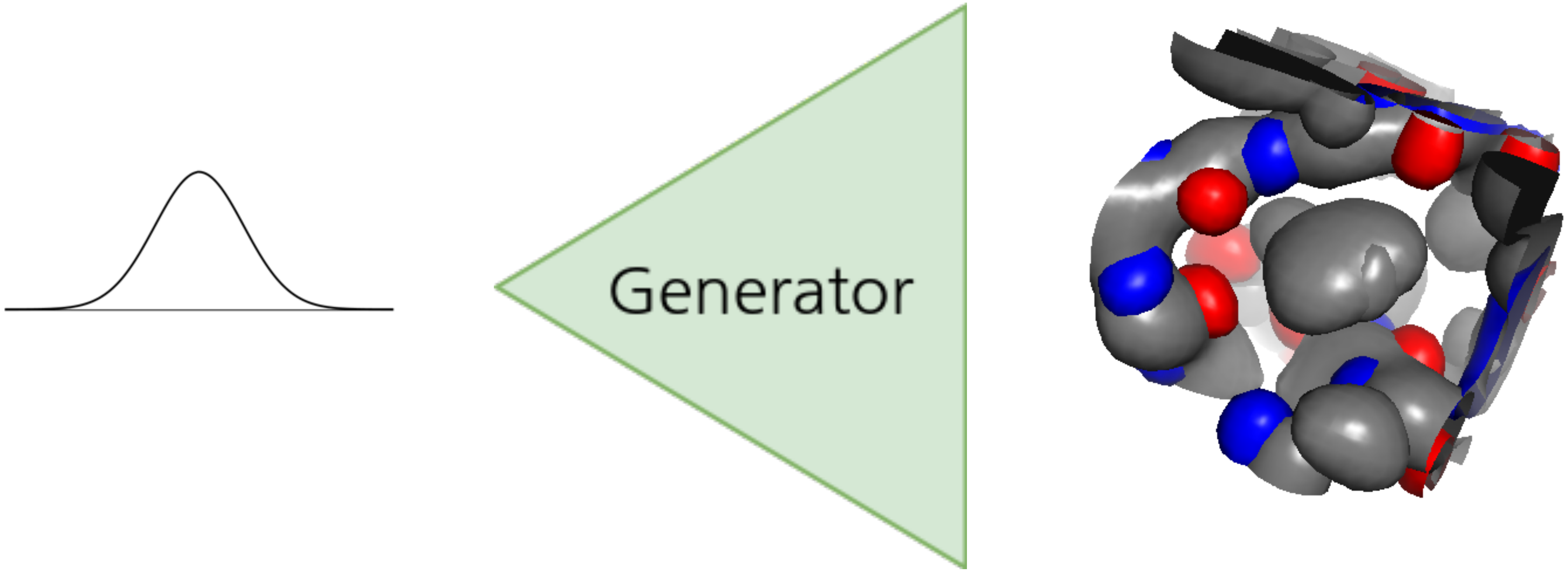
active/decoy

receptor & ligand grid

predicted class

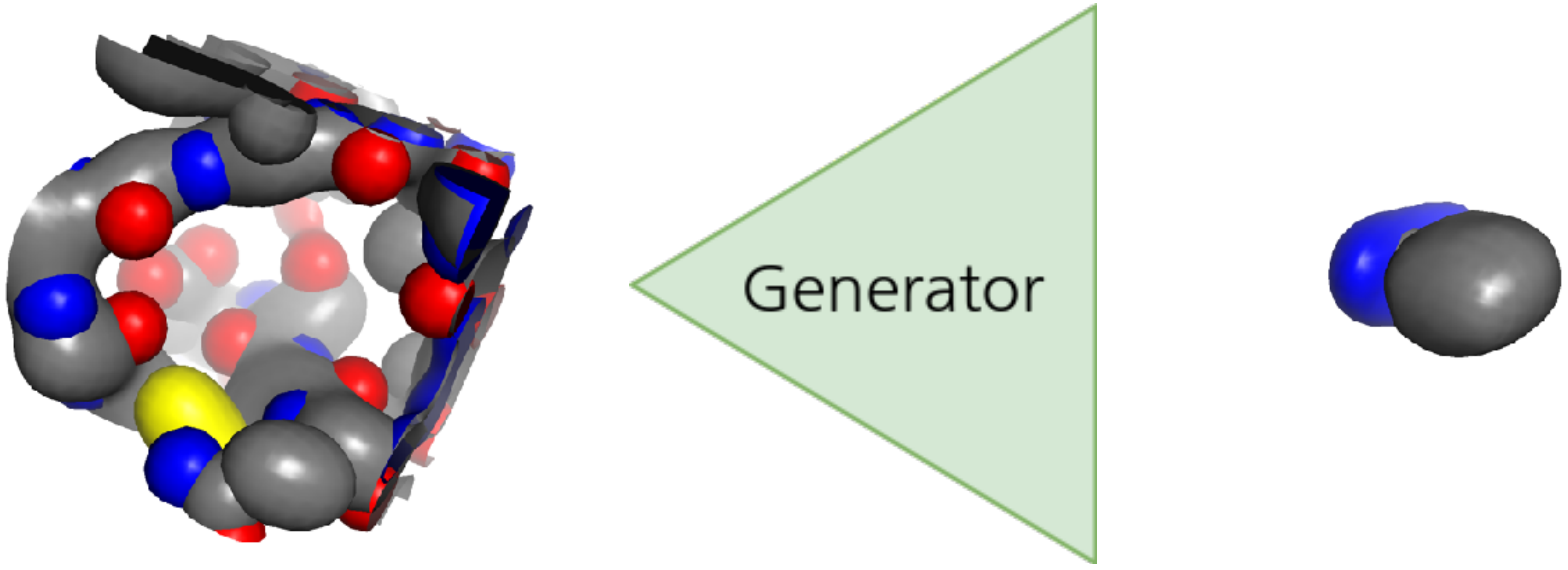
Generative Models

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*

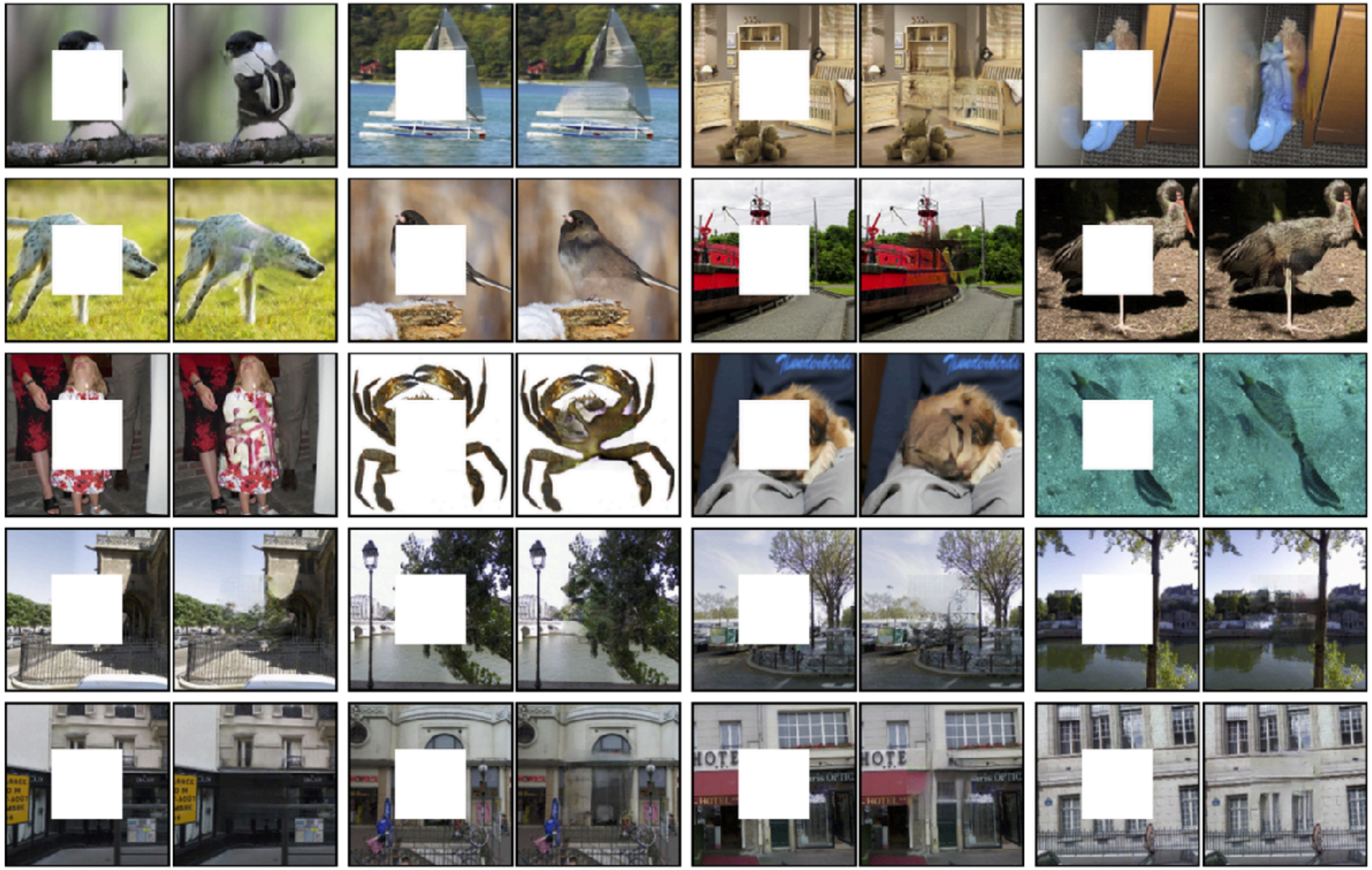
Context Encoding



receptor grid

generated ligand grid

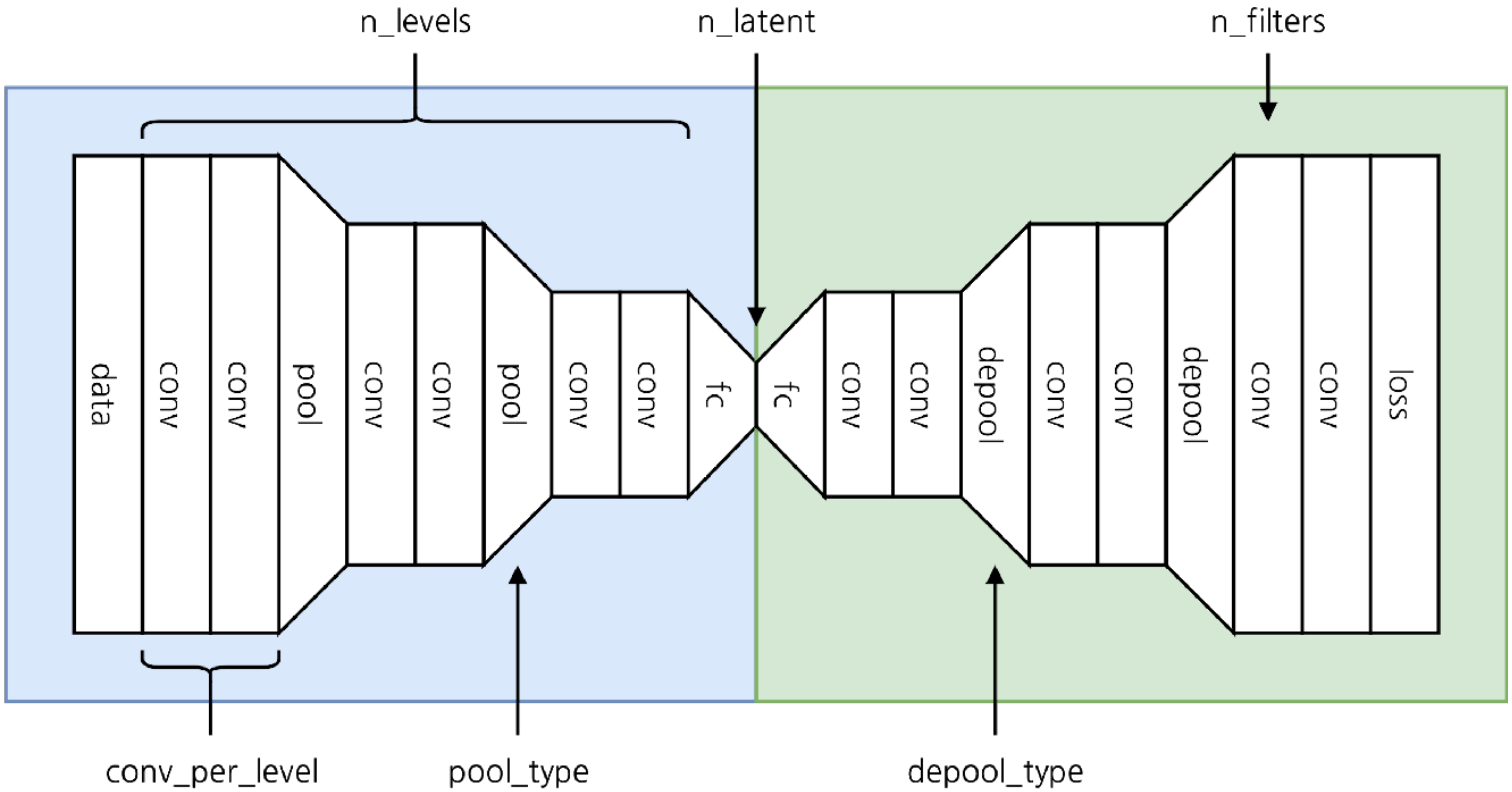
Context Encoding



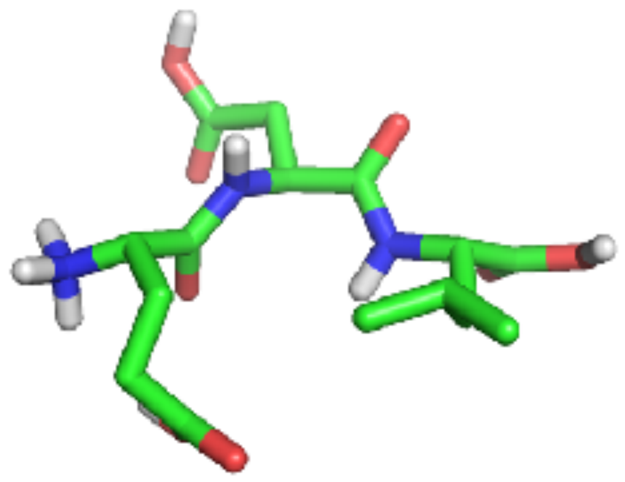
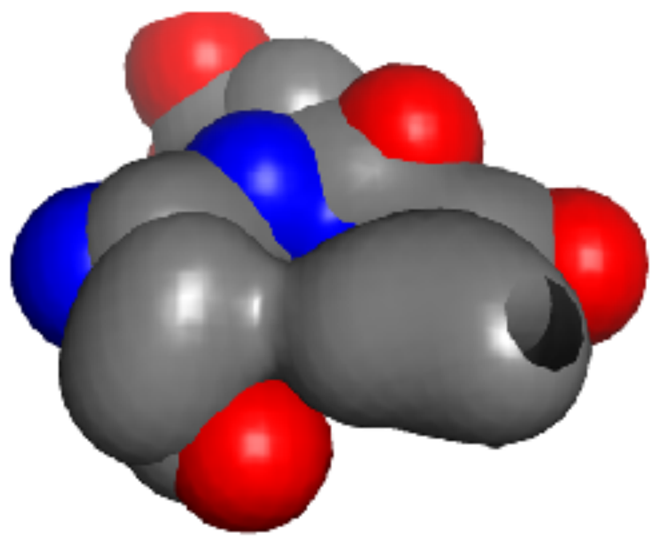
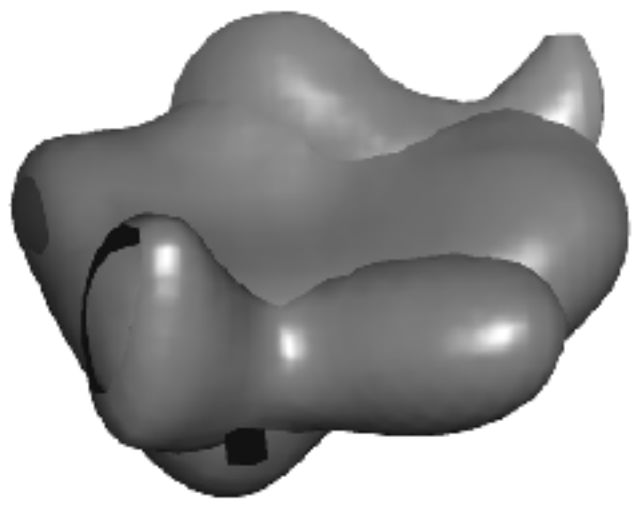
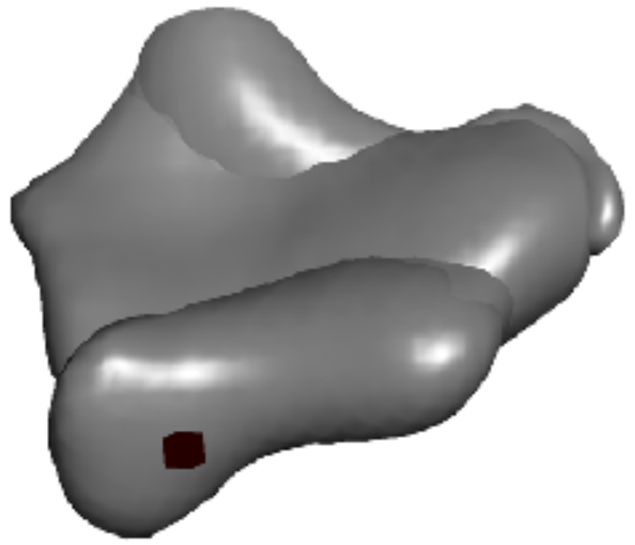
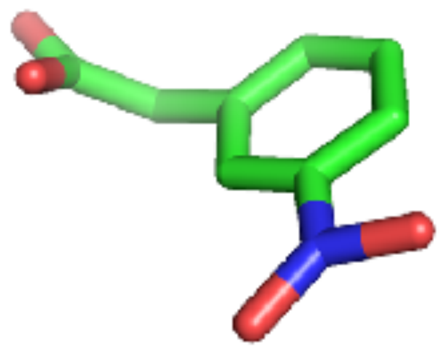
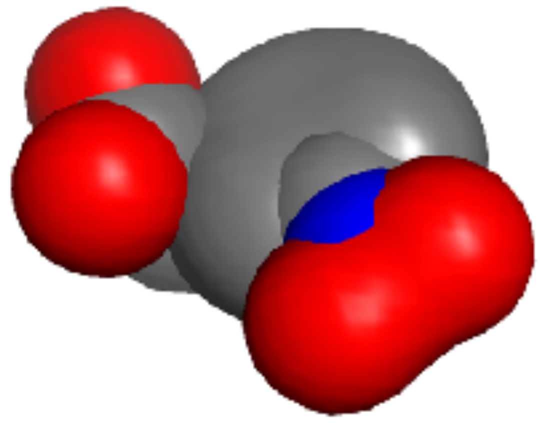
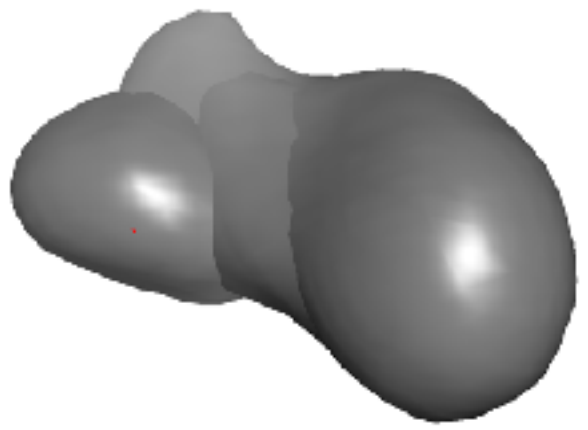
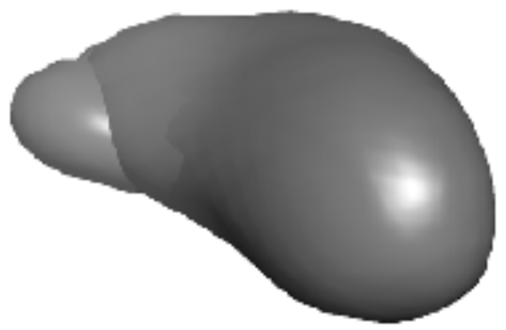
http://people.eecs.berkeley.edu/~pathak/context_encoder/

Model Architecture

- data_dim (24)
- resolution (0.5, 1.0)
- n_levels (3, 4, 5)
- conv_per_level (1, 2, 3)
- n_filters (16, 32, 64)
- width_factor (1, 2)
- n_latent (512, 1024)
- pool_type
  - max pooling
  - average pooling
  - strided convolution
- depool_type
  - nearest-neighbor
  - strided deconvolution
- loss_types
  - L2 loss



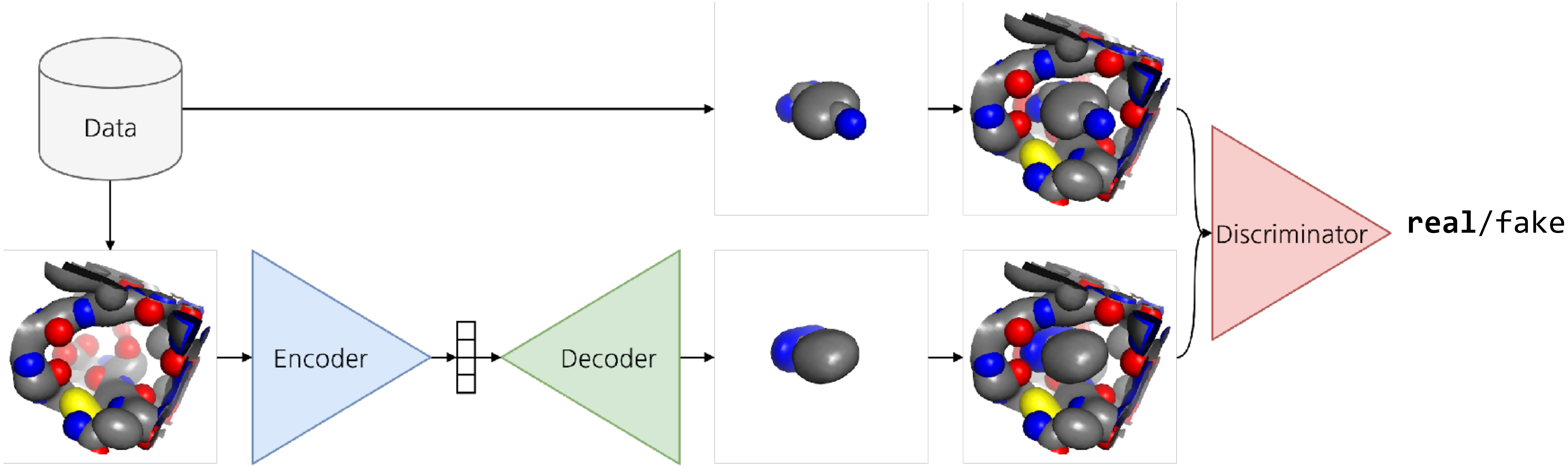
Context Encoder Examples

	sdf file	ground truth	ce11_24_2_3_64_m_n	ce11_24_3_3_64_a_n
1a30				
1ai5_ligand				

Generative Adversarial Networks

A discriminator network is trained to distinguish real vs. fake receptor-ligand grids

A generator network (context encoder) is trained to produce output that fools the discriminator



Progressive Growing of GANs for Improved Quality, Stability, and Variation

Tero Karras
NVIDIA

Timo Aila
NVIDIA

Samuli Laine
NVIDIA

Jaakko Lehtinen
NVIDIA
Aalto University

NVIDIA

https://youtu.be/G06dEcZ-QTg

Progressive Growing of GANs for Improved Quality, Stability, and Variation

Tero Karras
NVIDIA

Timo Aila
NVIDIA

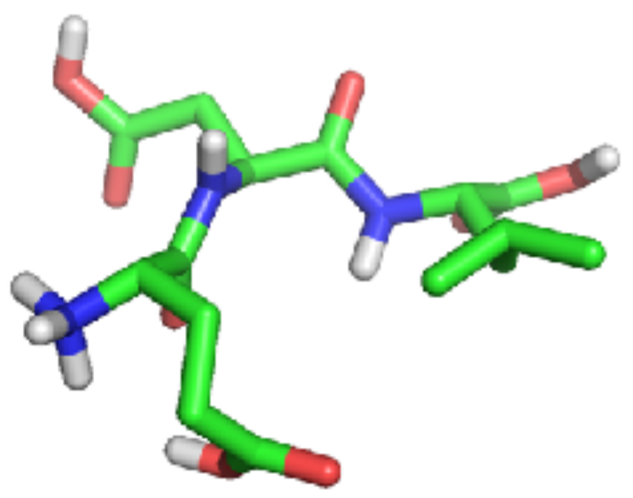
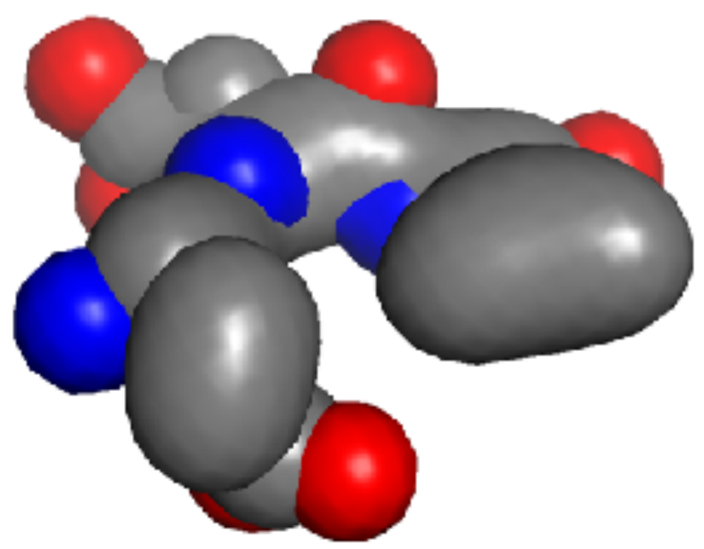
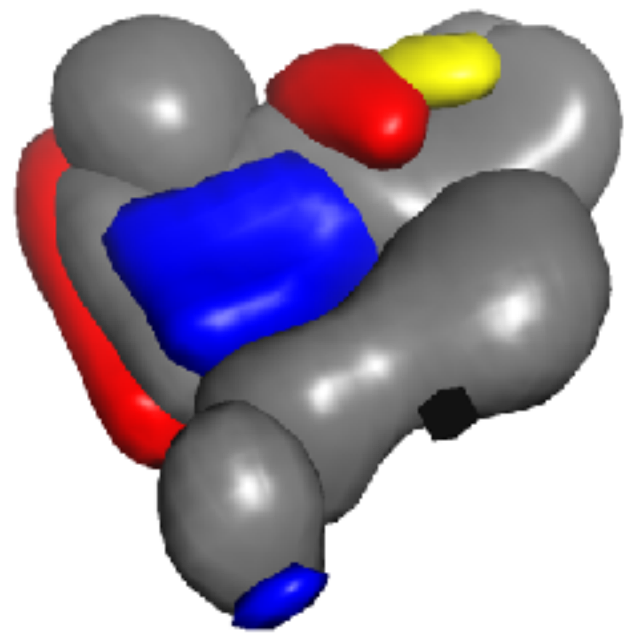
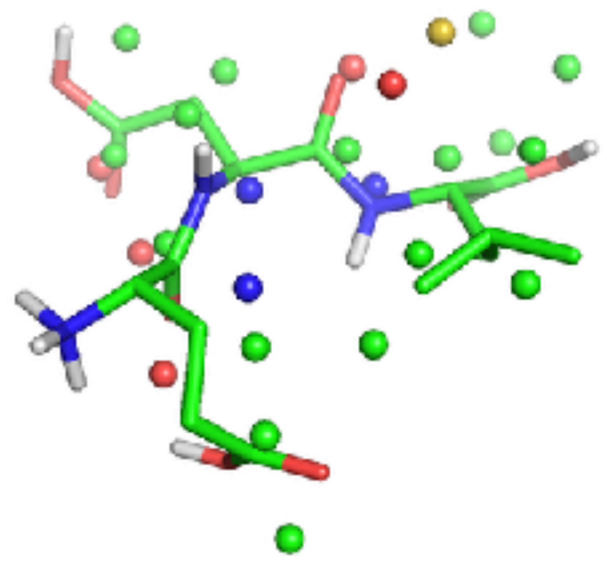
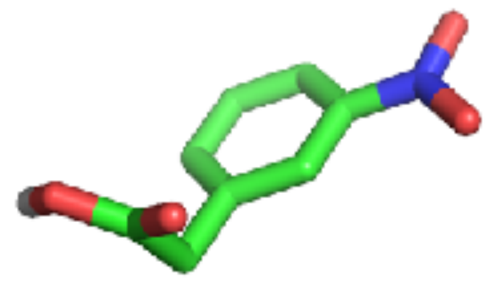
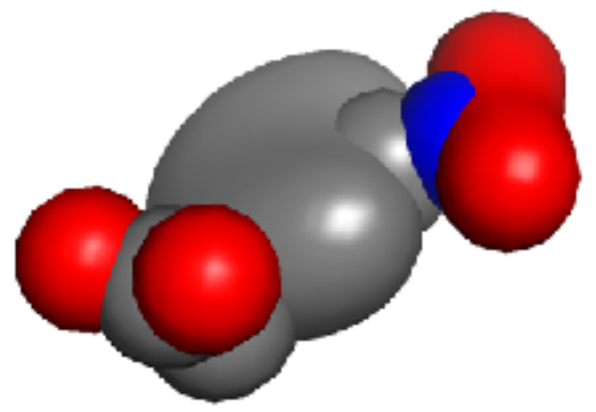
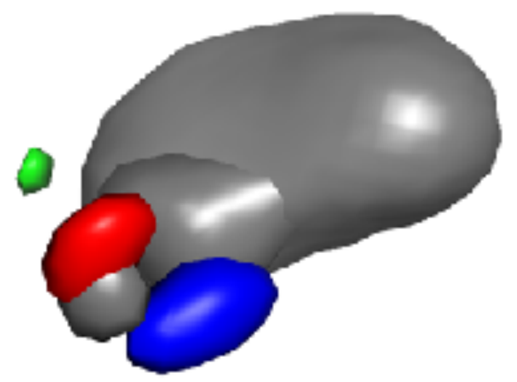
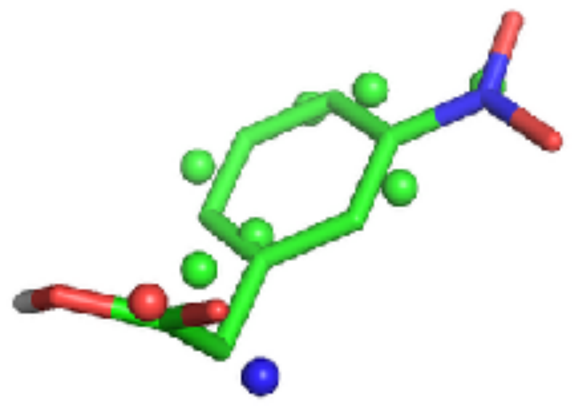
Samuli Laine
NVIDIA

Jaakko Lehtinen
NVIDIA
Aalto University

NVIDIA

https://youtu.be/G06dEcZ-QTg

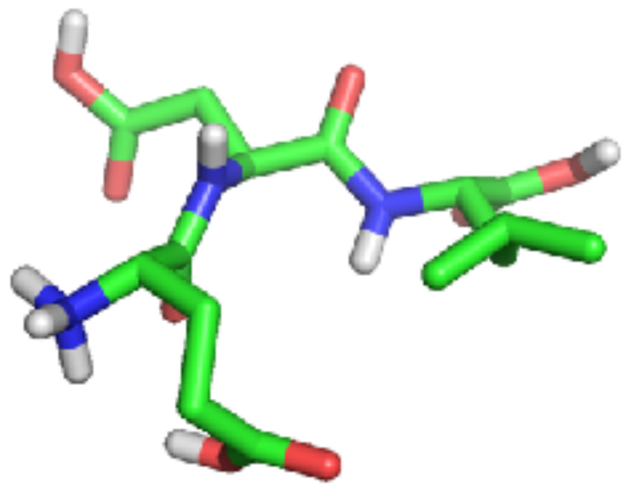
Preliminary GAN Examples

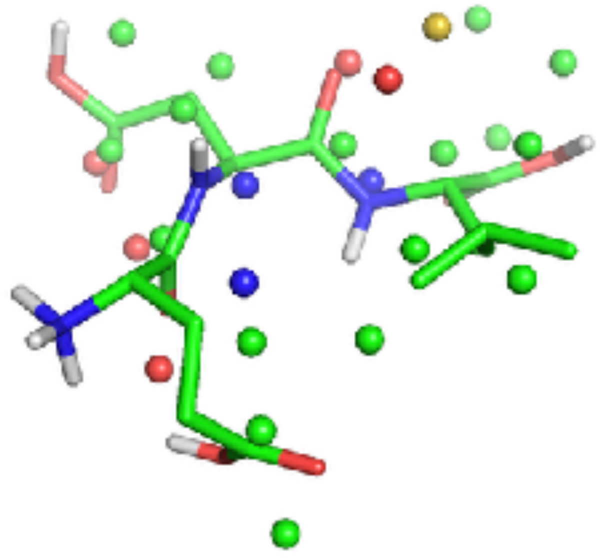
	ligand	ground truth	GAN	fit atoms
1a30				
1ai5				

Preliminary GAN Examples

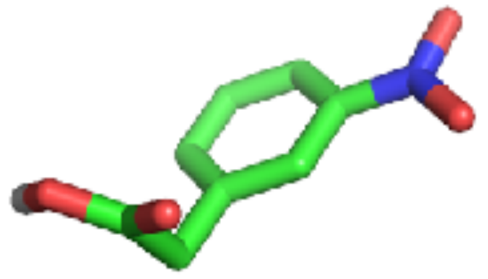
ligand | ground truth | GAN | fit atoms

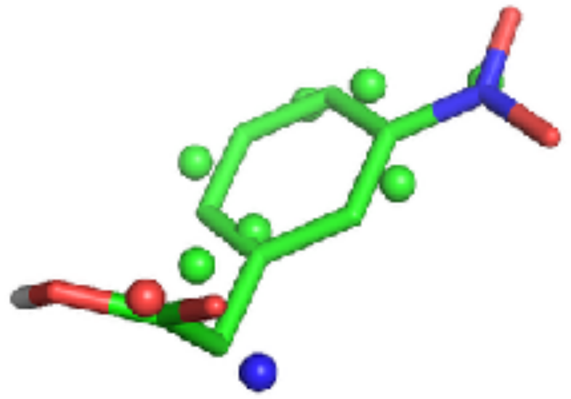
1a30





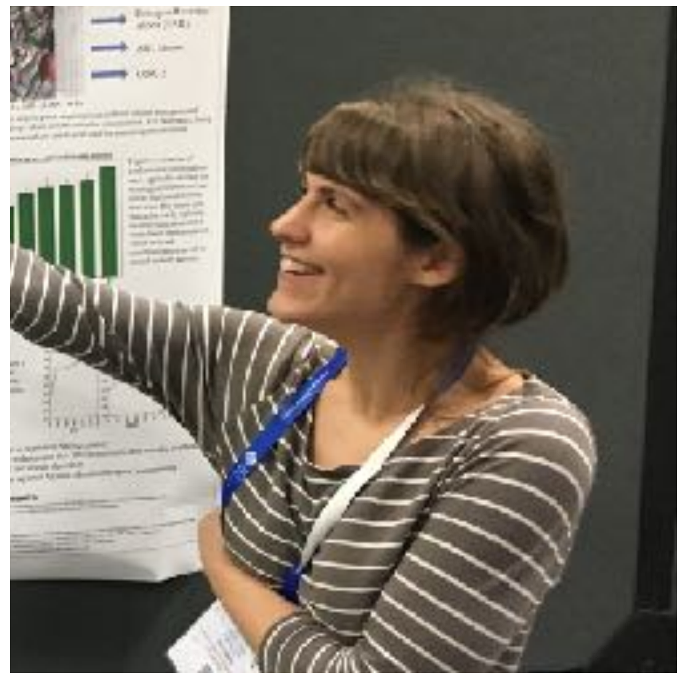
1ai5





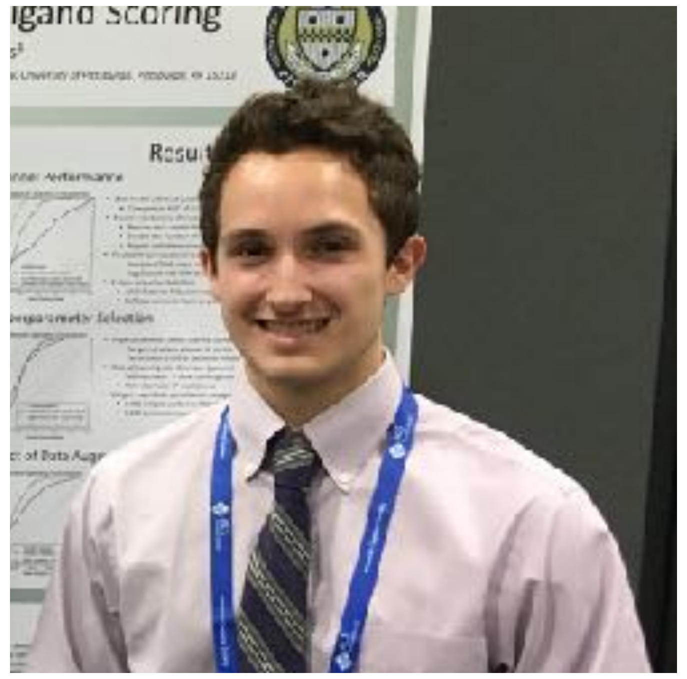
http://torch.ch/blog/2015/11/13/gan.html

Acknowledgements



Jocelyn Sunseri

Josh Hochuli



Matt Ragoza

Group Members

Jocelyn Sunseri
Jonathan King
Paul Francoeur
Matt Ragoza
Josh Hochuli
Pulkit Mittal
Alec Helbling
Gibran Biswas
Sharanya Bandla
Faiha Khan
Lily Turner

UNIVERSITY OF PITTSBURGH
VERITAS VIRTUS 1787

Department of Computational and Systems Biology

TECBio REU

AI GRANT

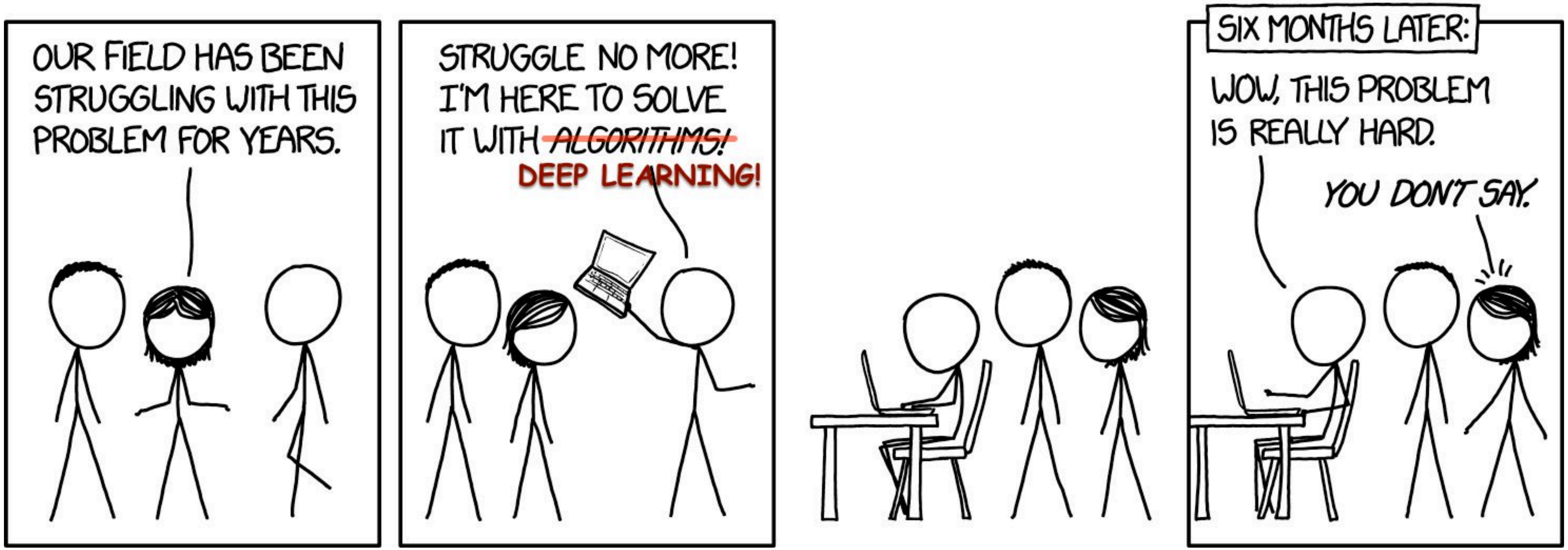
National Institute of General Medical Sciences
R01GM108340

NVIDIA

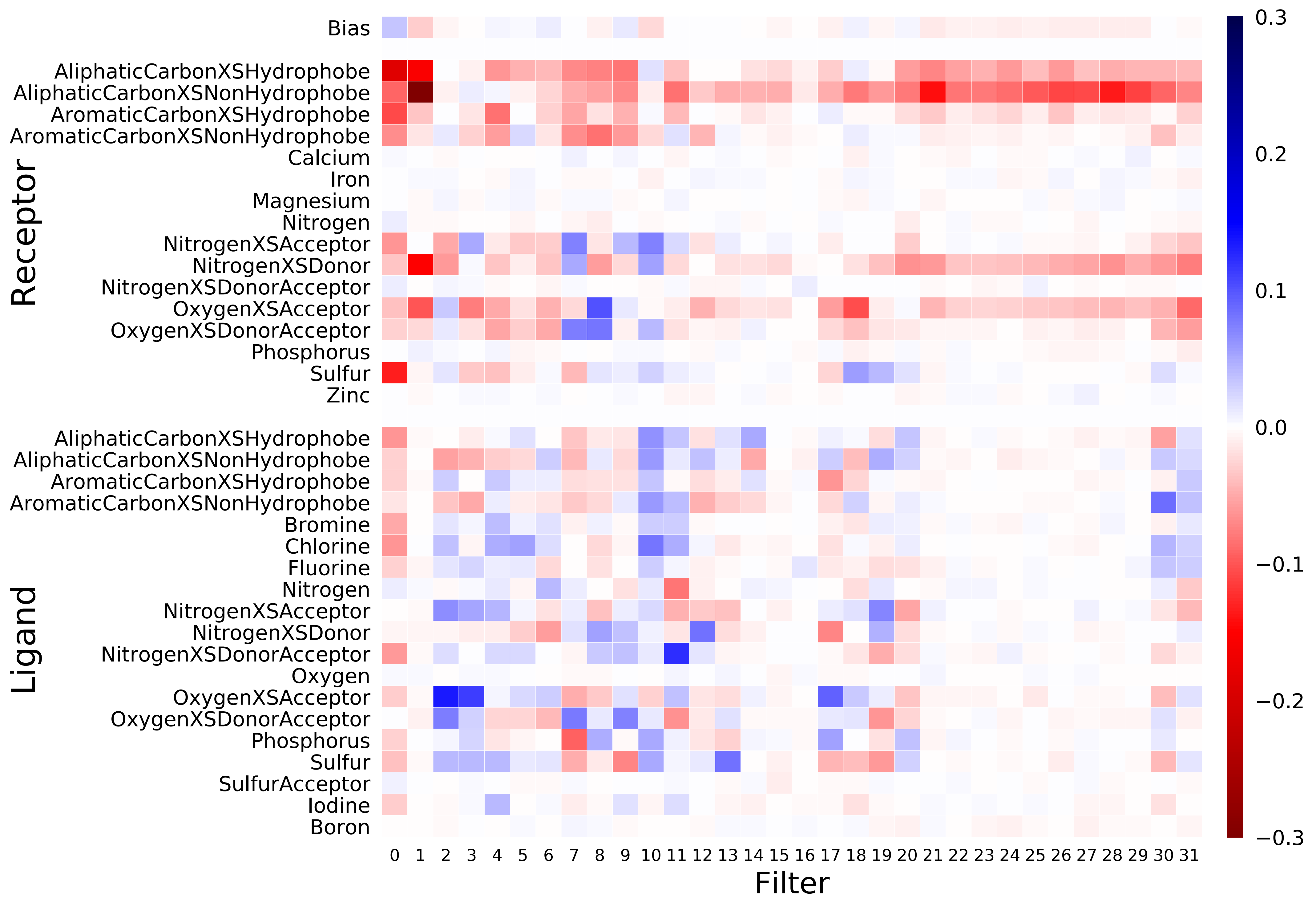
github.com/gnina

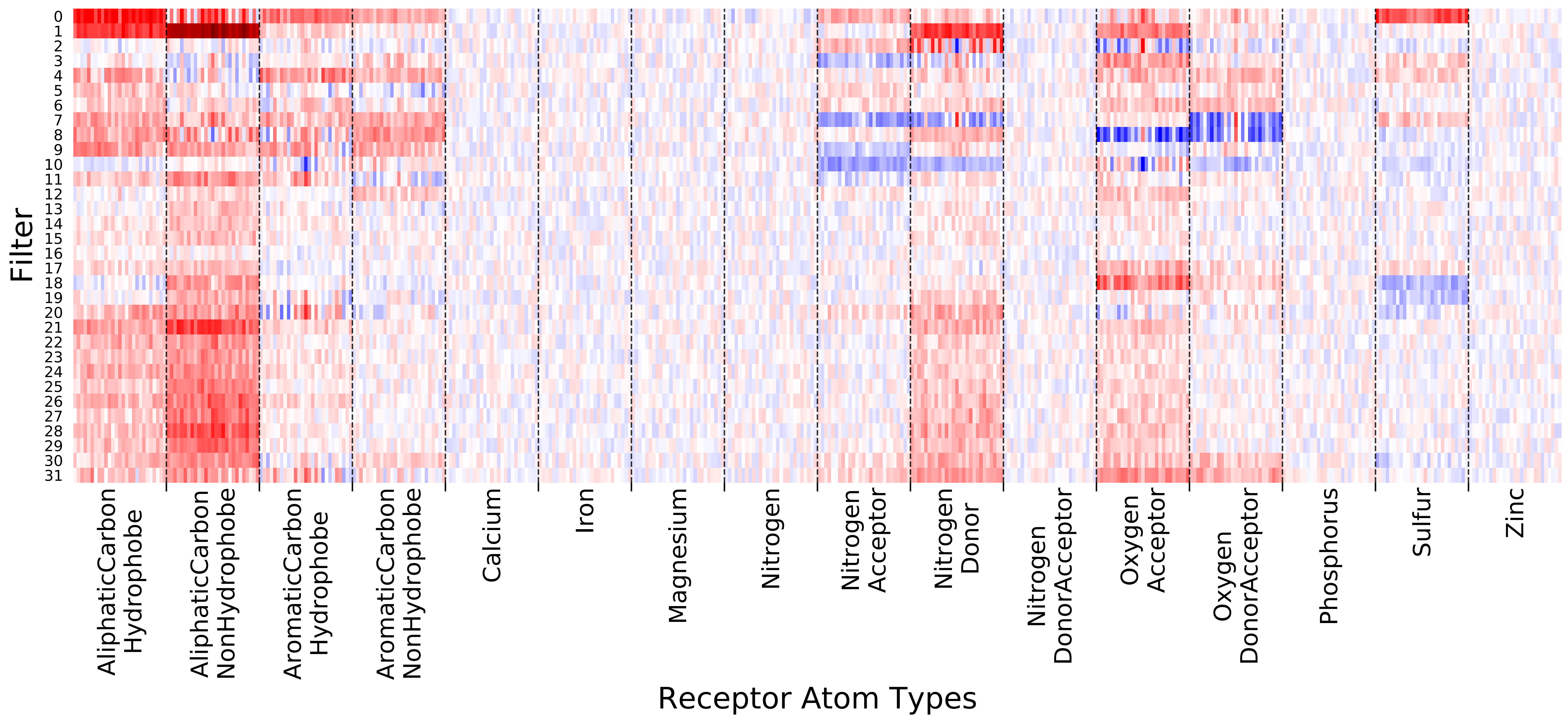
http://bits.csb.pitt.edu

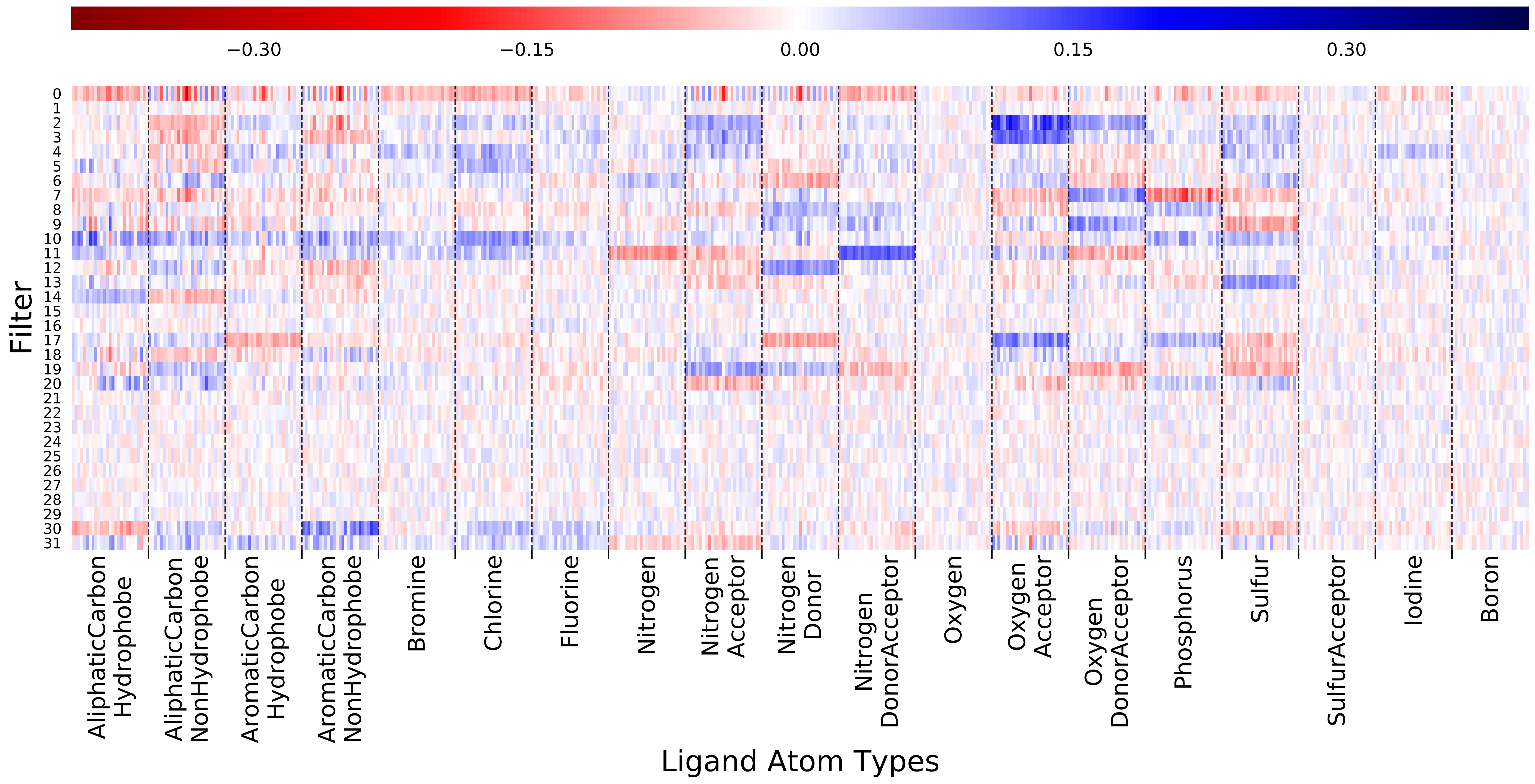
@david_koes



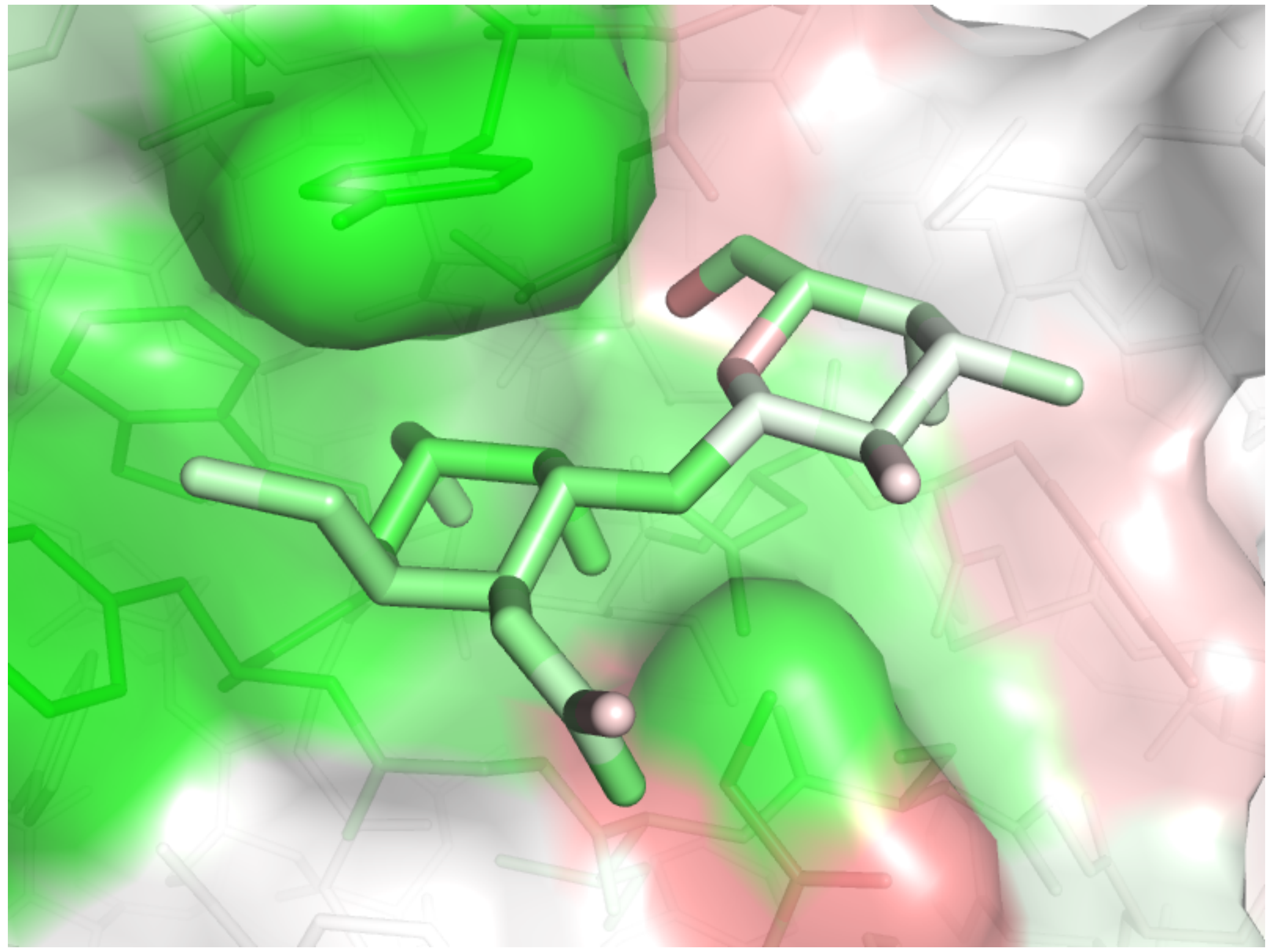
Filter Visualization



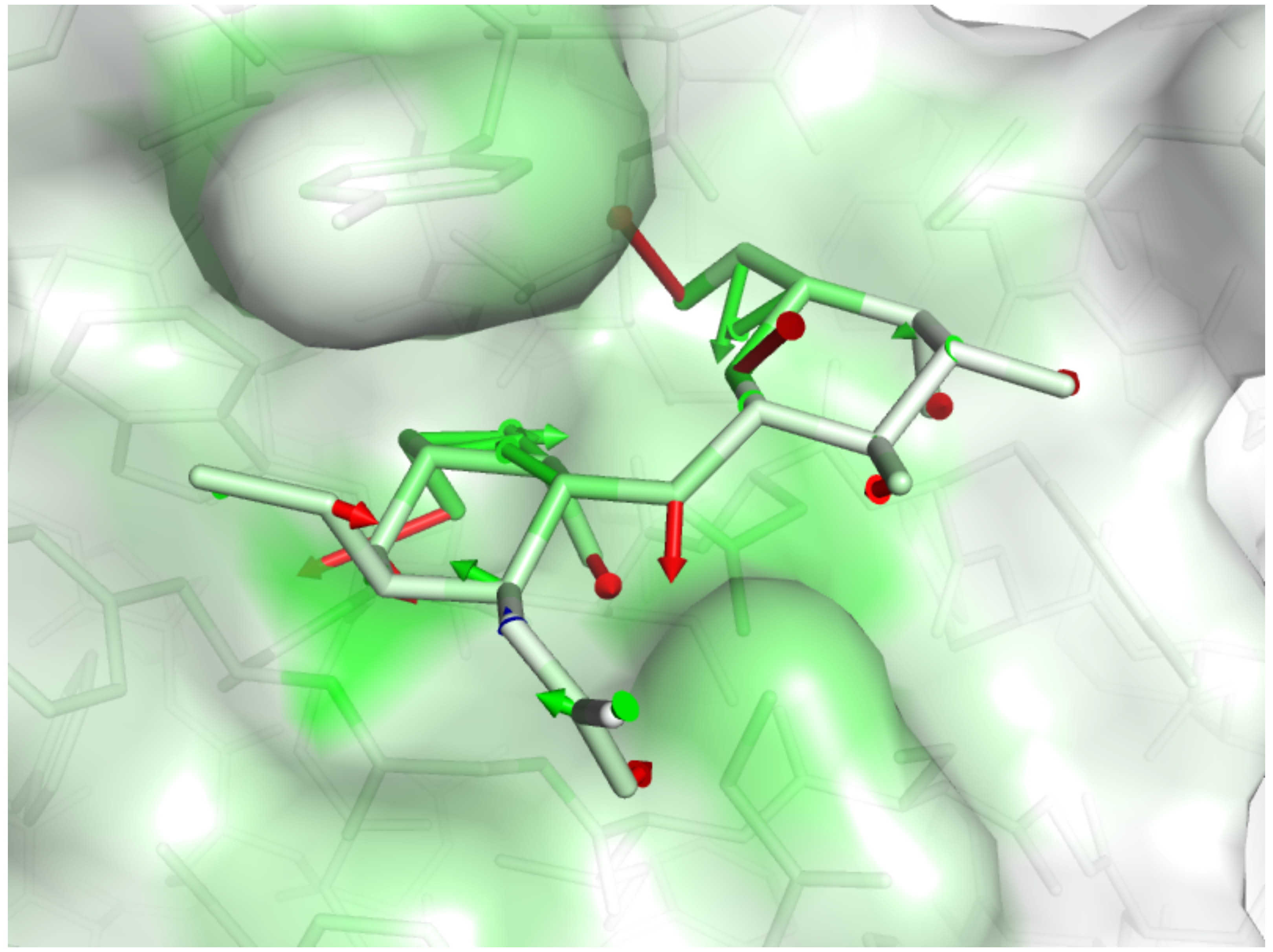




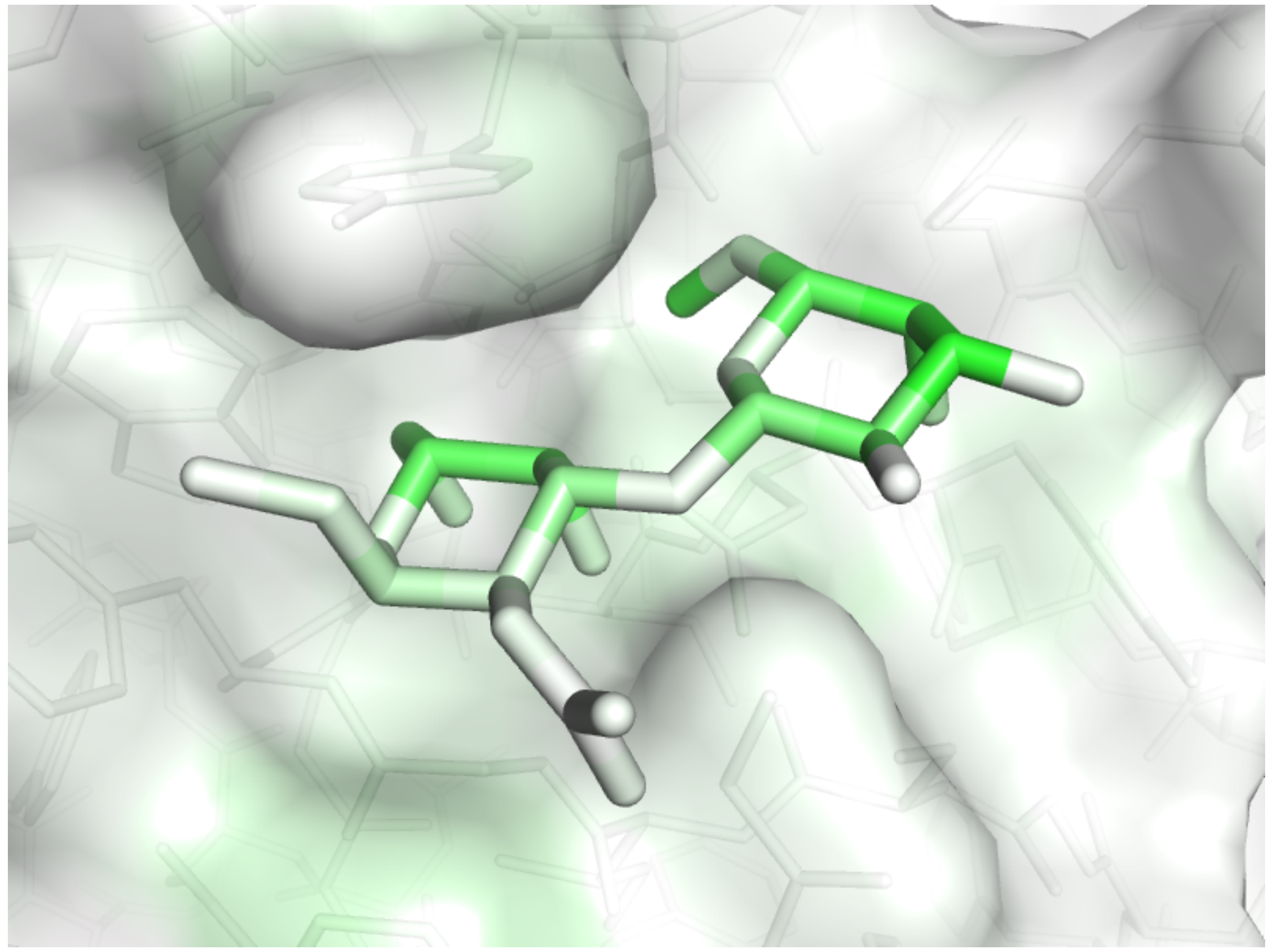
Visualization



masking



gradients



layer-wise relevance