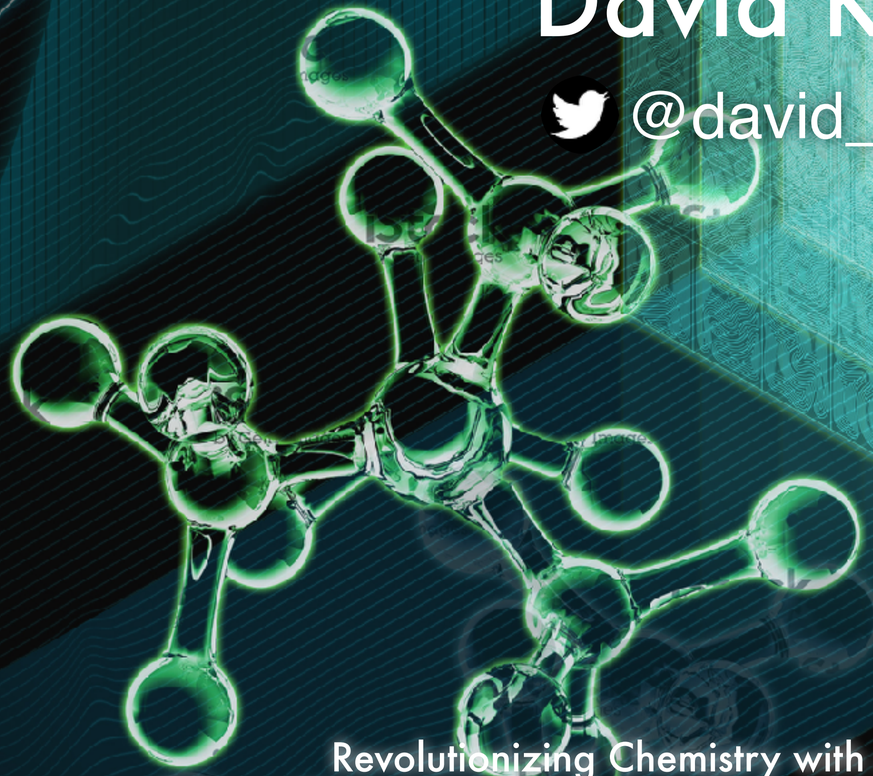
Generative models for structure-based drug design

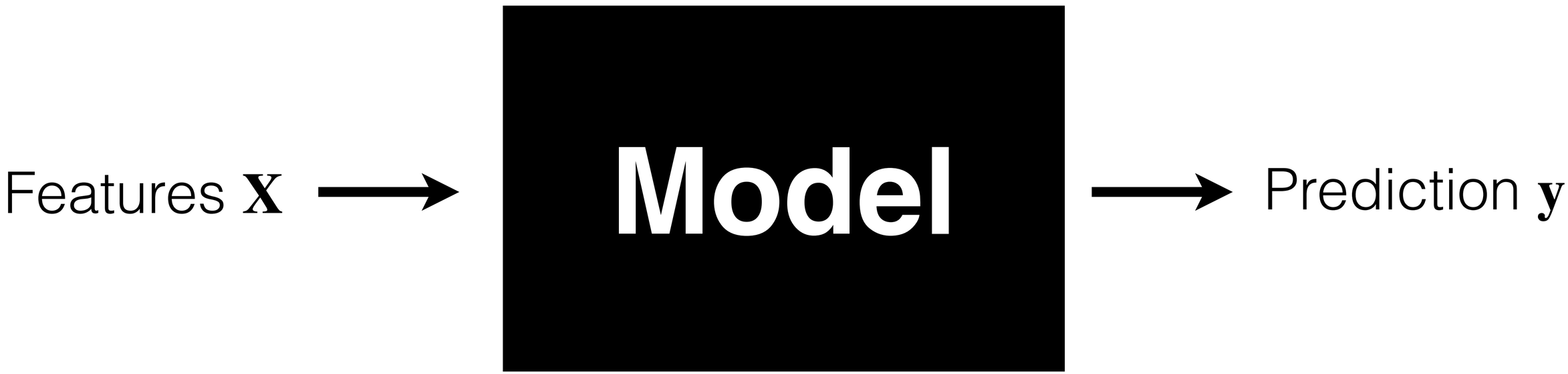
David Koes

@david_koes

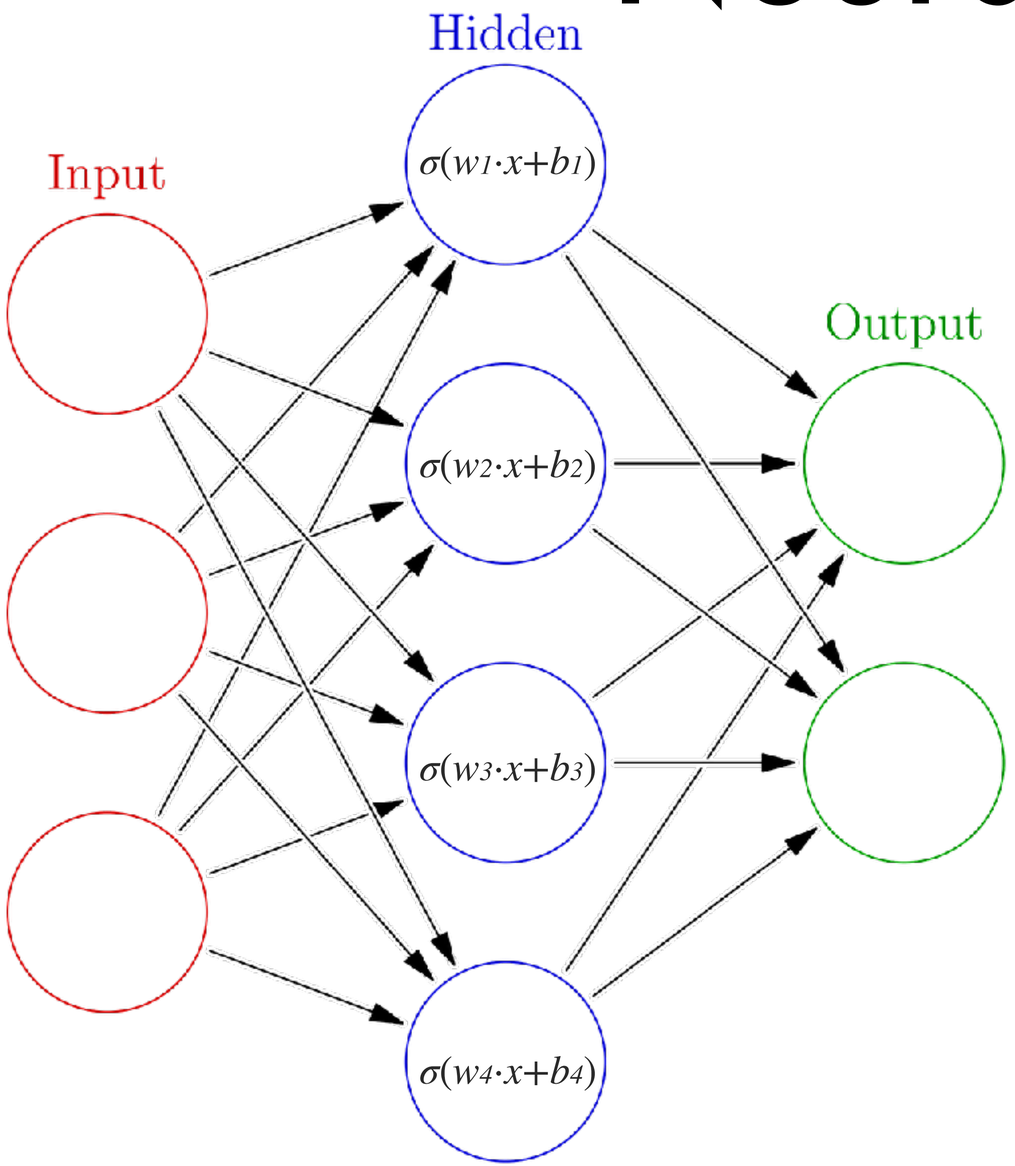


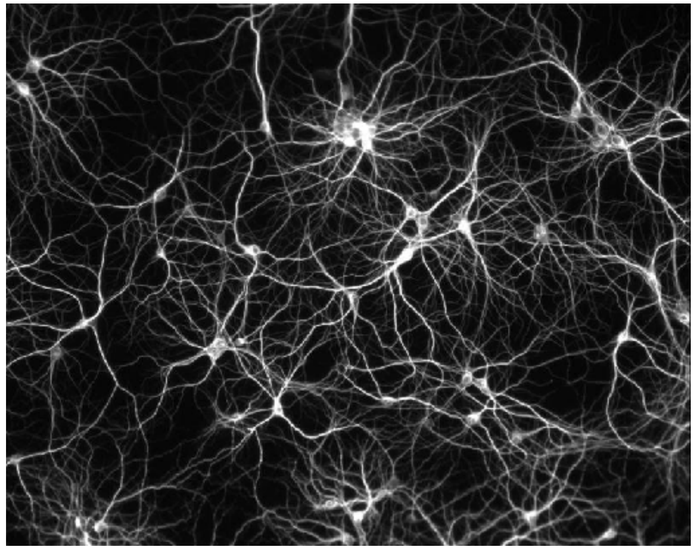
Revolutionizing Chemistry with Artificial Intelligence
American Chemical Society
Boston, MA
August 20, 2018

Machine Learning



Neural Networks





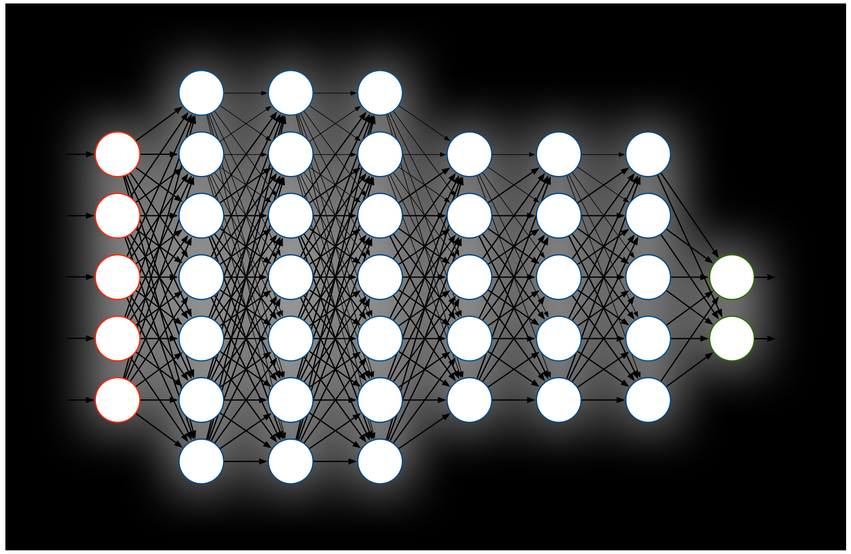
The **universal approximation theorem** states that, under reasonable assumptions, a feedforward **neural network** with a finite number of nodes **can approximate any continuous** function to within a given error over a bounded input domain.

Deep Learning

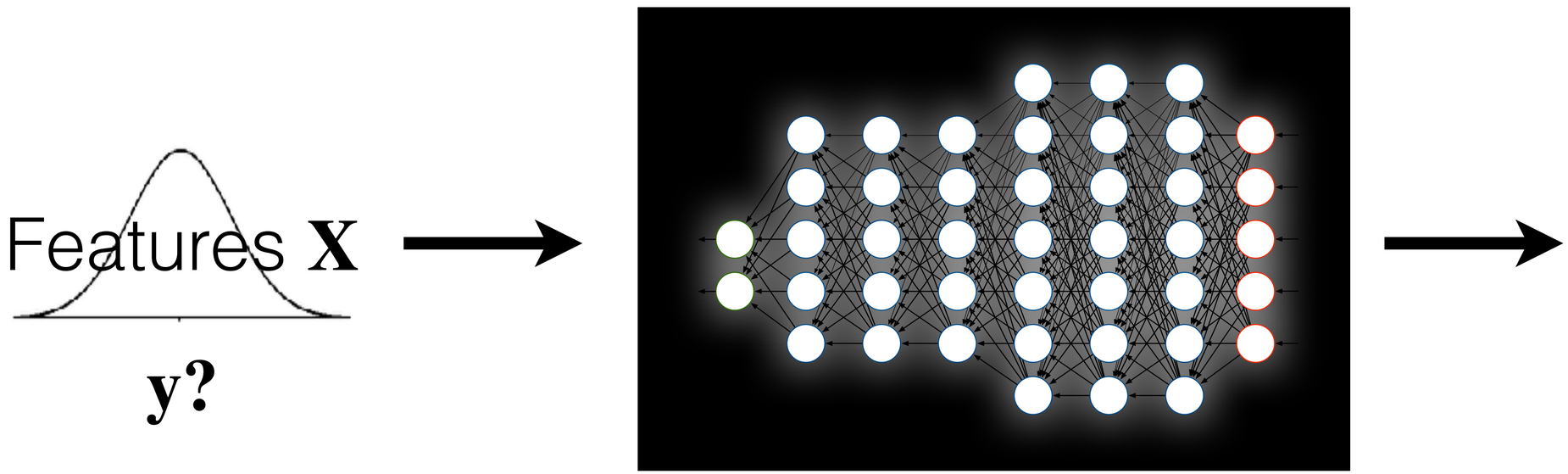
$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

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

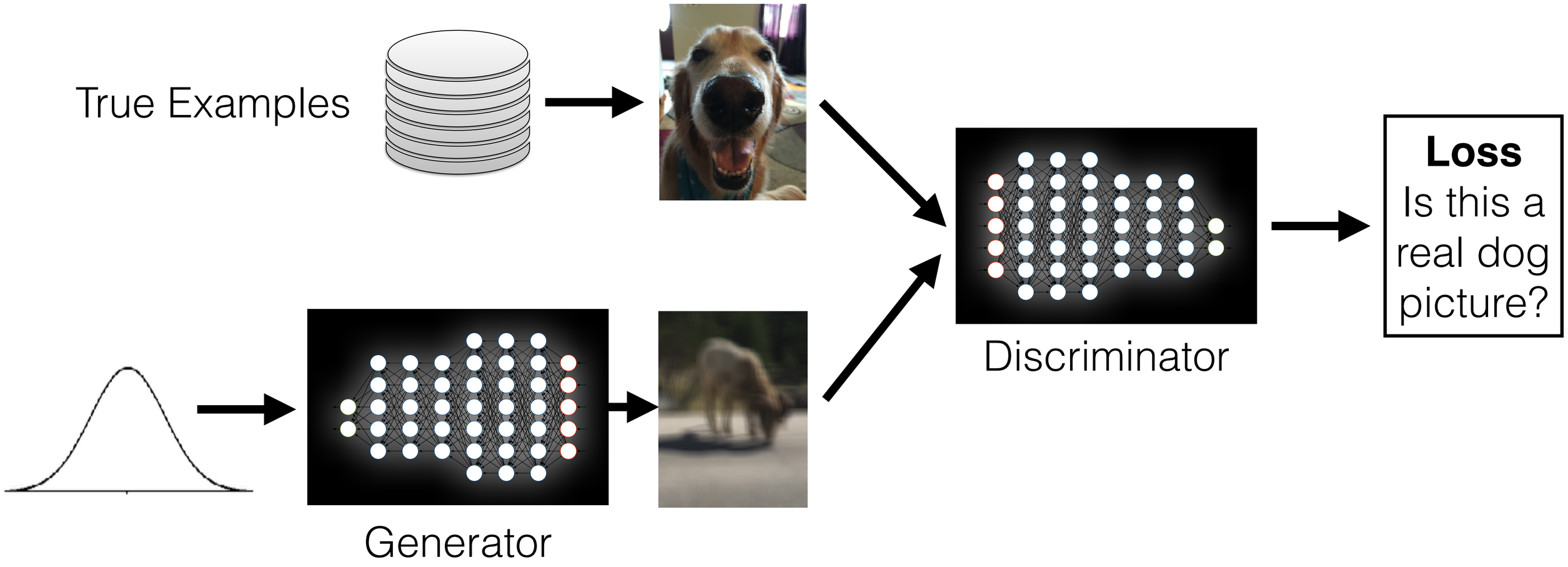
Discriminative Model

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

Generative Model



Generative Adversarial Networks



Generative Adversarial Networks

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

Generative Adversarial Networks

https://arxiv.org › stat ▾

by IJ Goodfellow - 2014 - Cited by 4339 - Related articles

Jun 10, 2014 - Submission history. From: Ian **Goodfellow** [view email] [v1] Tue, 10 Jun **2014** 18:58:17 GMT (1257kb,D). Which authors of this paper are ...

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

Tero Karras
NVIDIA

Timo Aila
NVIDIA

Samuli Laine
NVIDIA

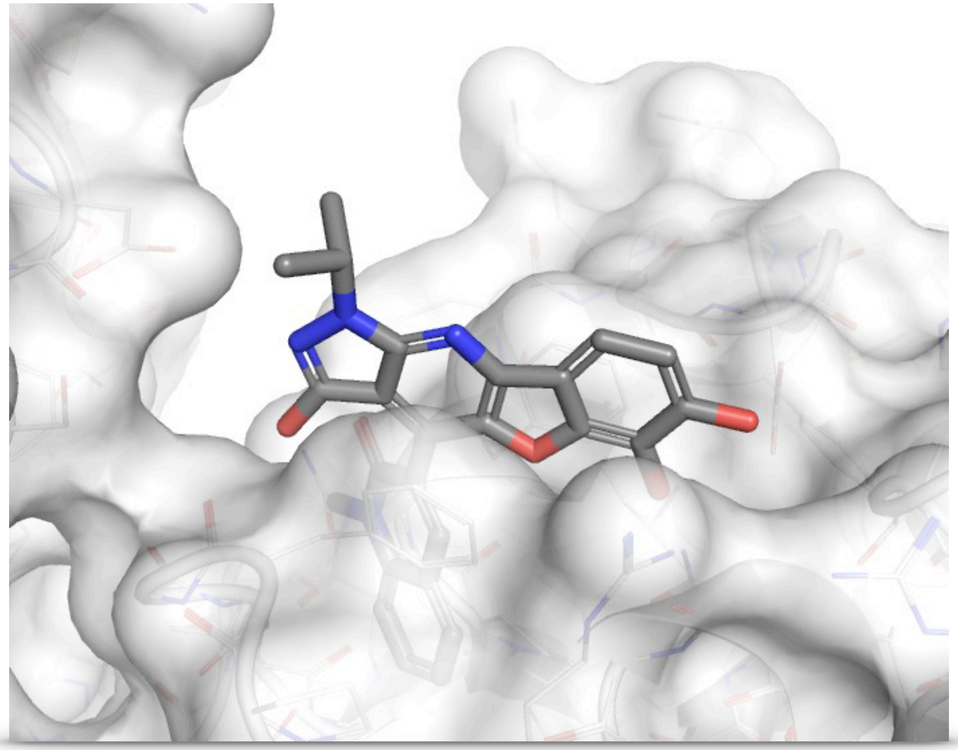
Jaakko Lehtinen
NVIDIA
Aalto University

https://youtu.be/G06dEcZ-QTg

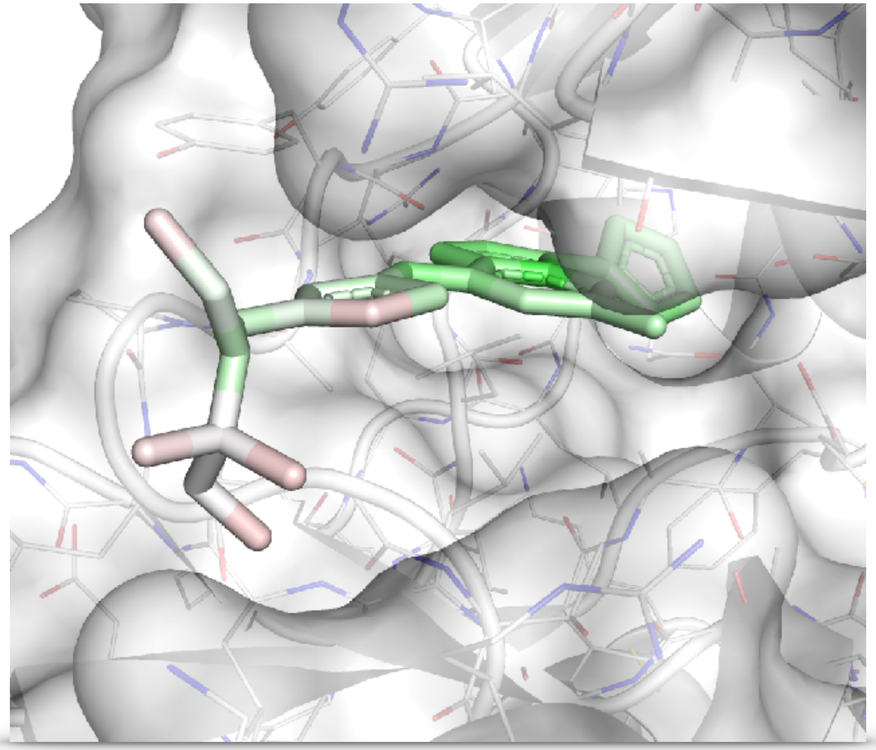
Generative models for structure-based drug design

Structure Based Drug Design

Pose Prediction Binding Discrimination Affinity Prediction

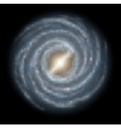


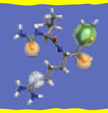
Virtual Screening



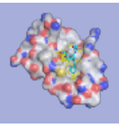
Lead Optimization

Drug Discovery Funnel

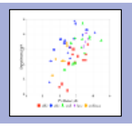
Purchasable  Accessible

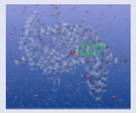


Matching

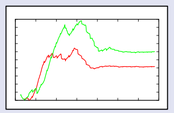


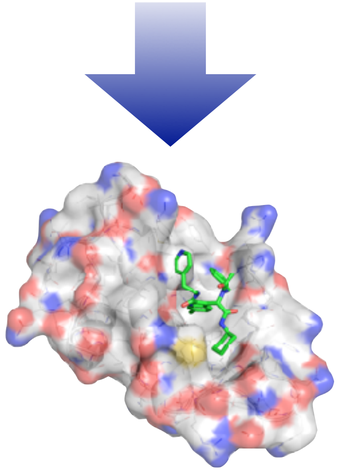
Scoring

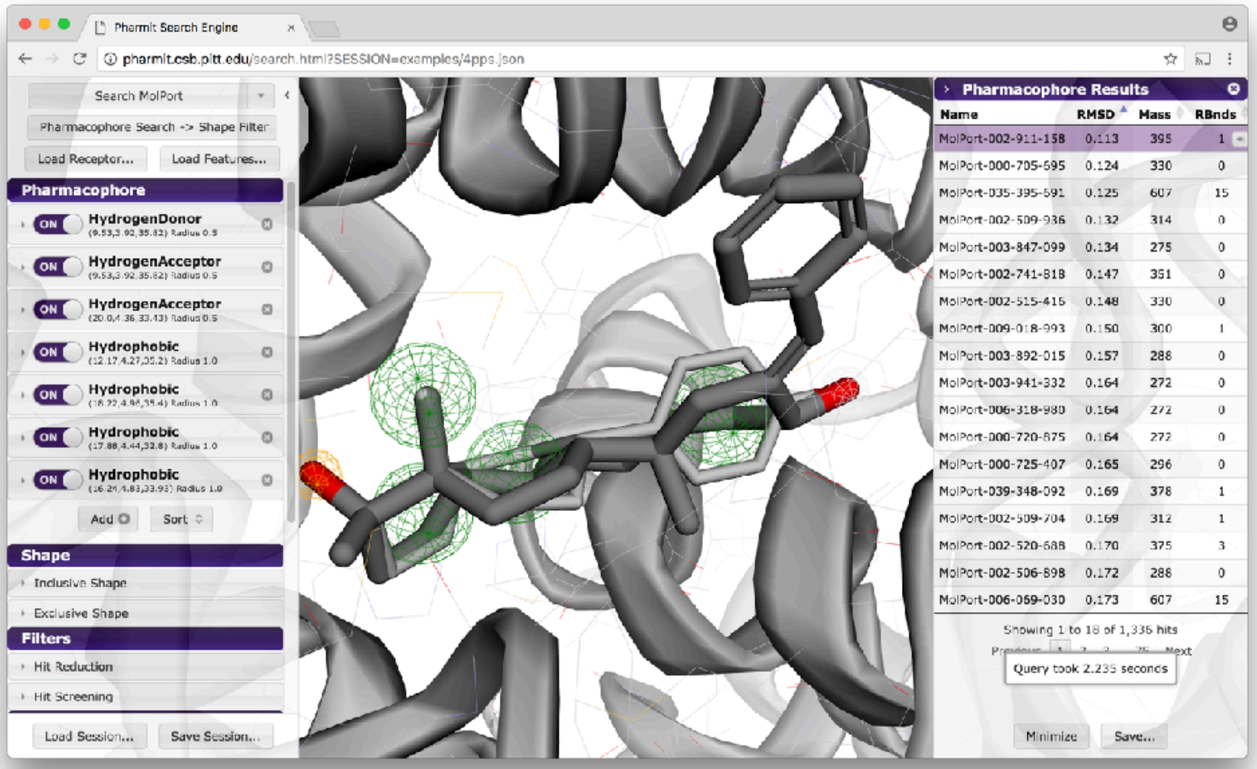




Dynamics



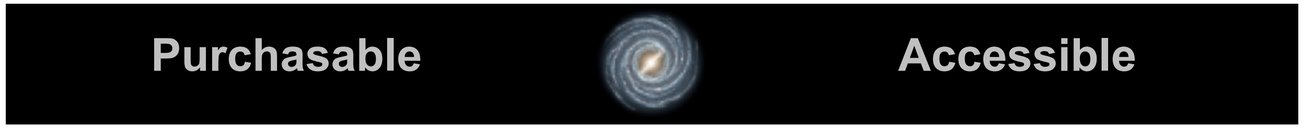


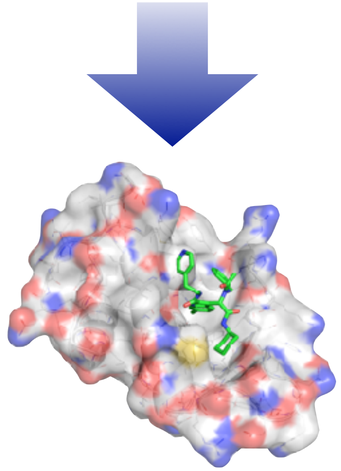


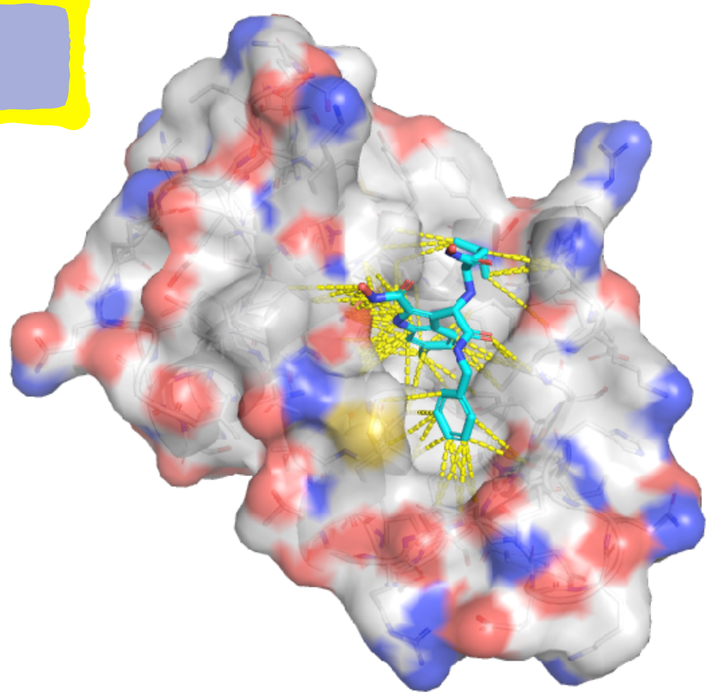
http://pharmit.csb.pitt.edu

COMP 528: Structure-based searching of chemical space with Pharmit
Thursday, Aug 23 9:20 AM
Douglass, Westin Boston Waterfront

Drug Discovery Funnel



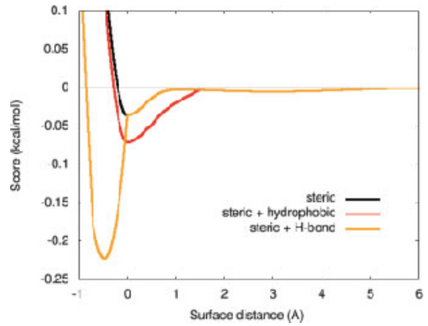




$$\text{gauss}_1(d) = w_{\text{guass}_1} e^{-(d/0.5)^2}$$
$$\text{gauss}_2(d) = w_{\text{guass}_2} e^{-((d-3)/2)^2}$$
$$\text{repulsion}(d) = \begin{cases} w_{\text{repulsion}} d^2 & d < 0 \\ 0 & d \geq 0 \end{cases}$$

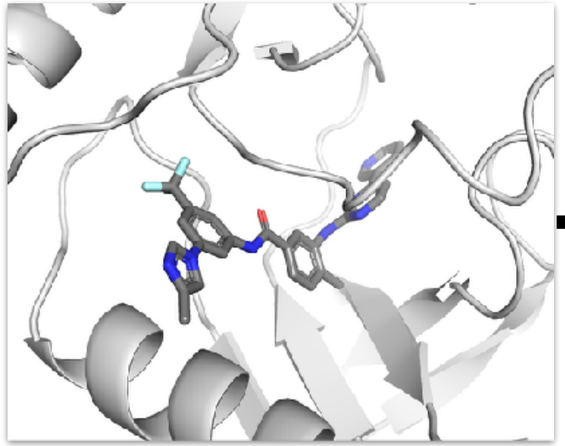
$$\text{hydrophobic}(d) = \begin{cases} w_{\text{hydrophobic}} & d < 0.5 \\ 0 & d > 1.5 \\ w_{\text{hydrophobic}}(1.5 - d) & otherwise \end{cases}$$

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



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



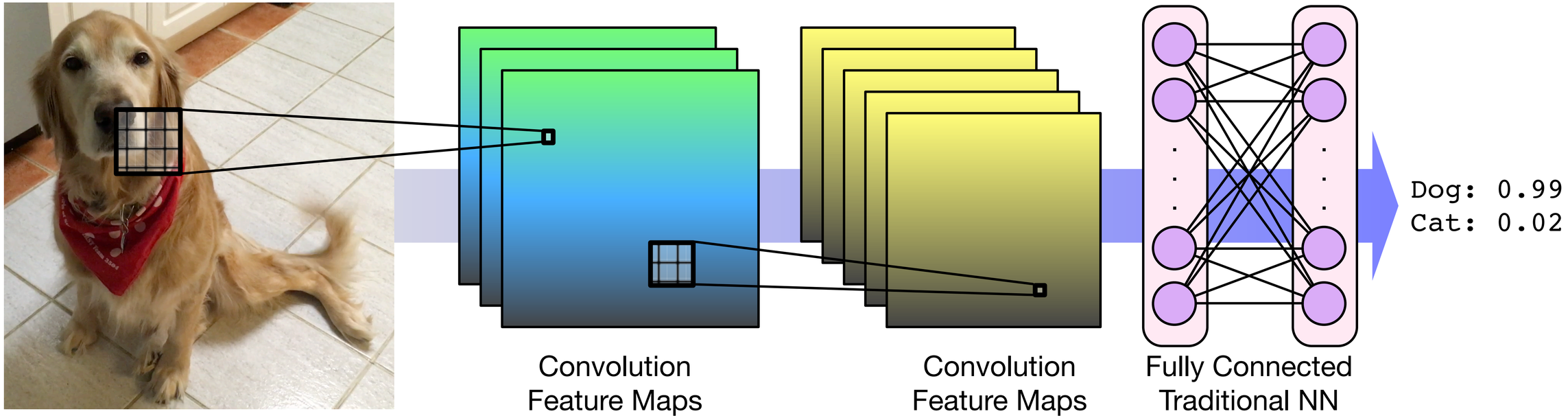
Model

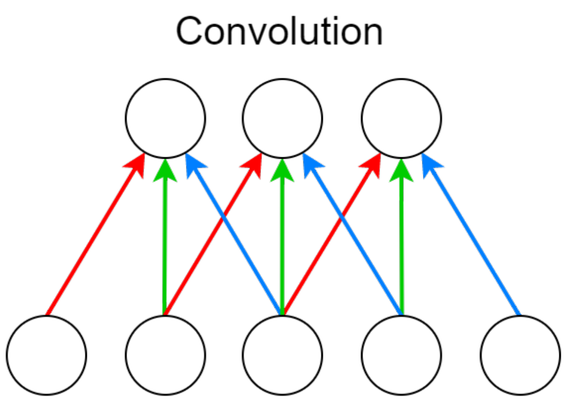
Pose Prediction

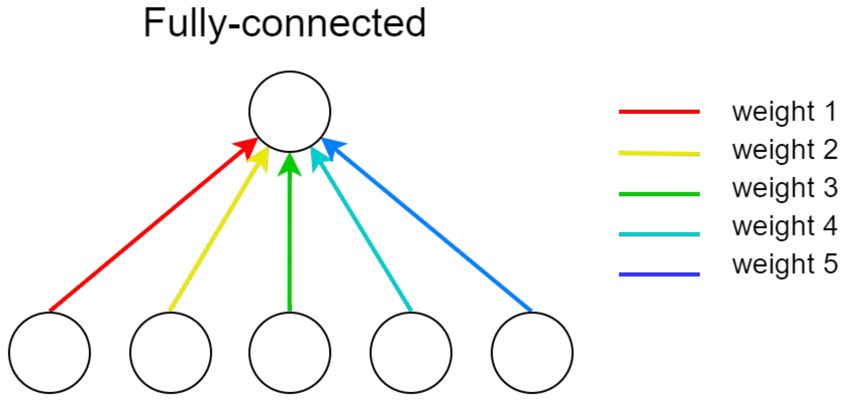
Binding Discrimination

Affinity Prediction

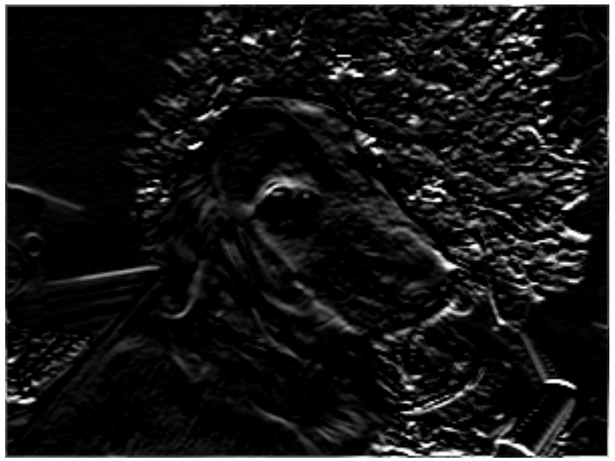
Convolutional Neural Networks



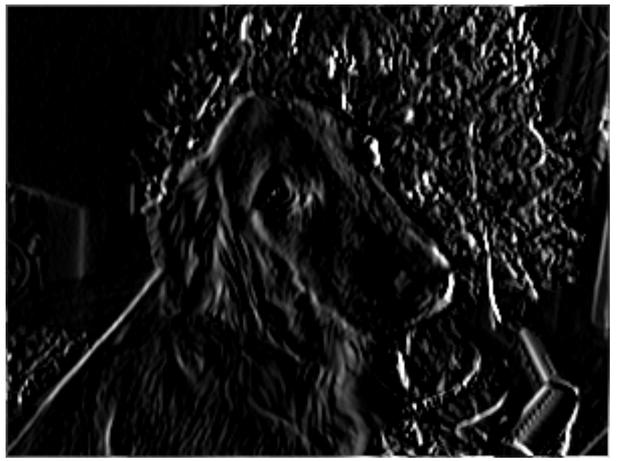




Convolutional Filters



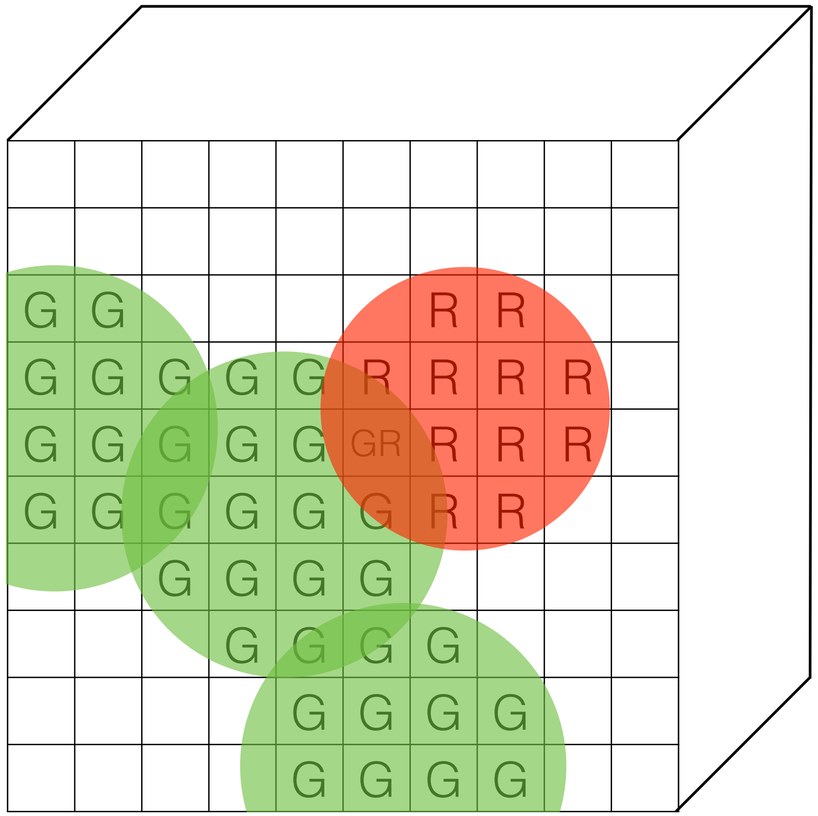
-1	-1	-1
0	0	0
1	1	1



-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**.

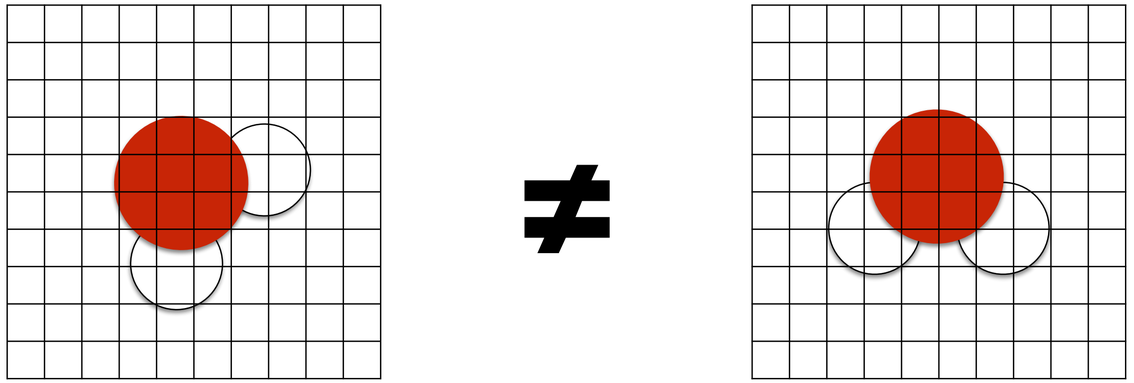
Why Grids?

Cons

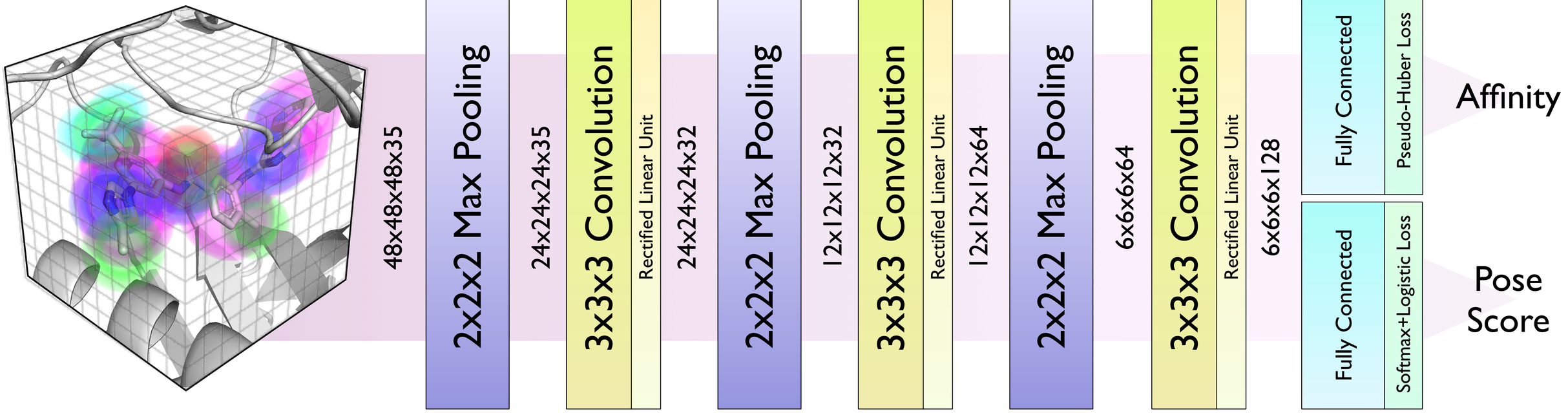
- *coordinate frame dependent*
- pairwise interactions not explicit

Pros

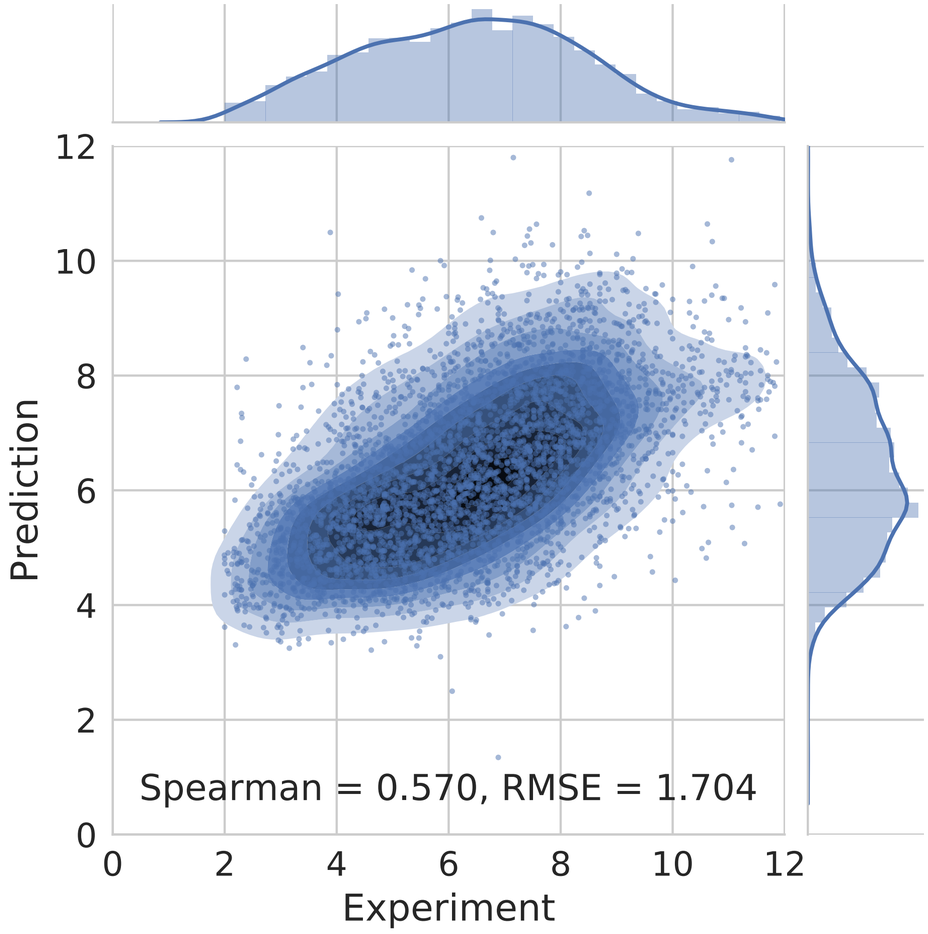
- clear spatial relationships
- amazingly parallel
- easy to interpret

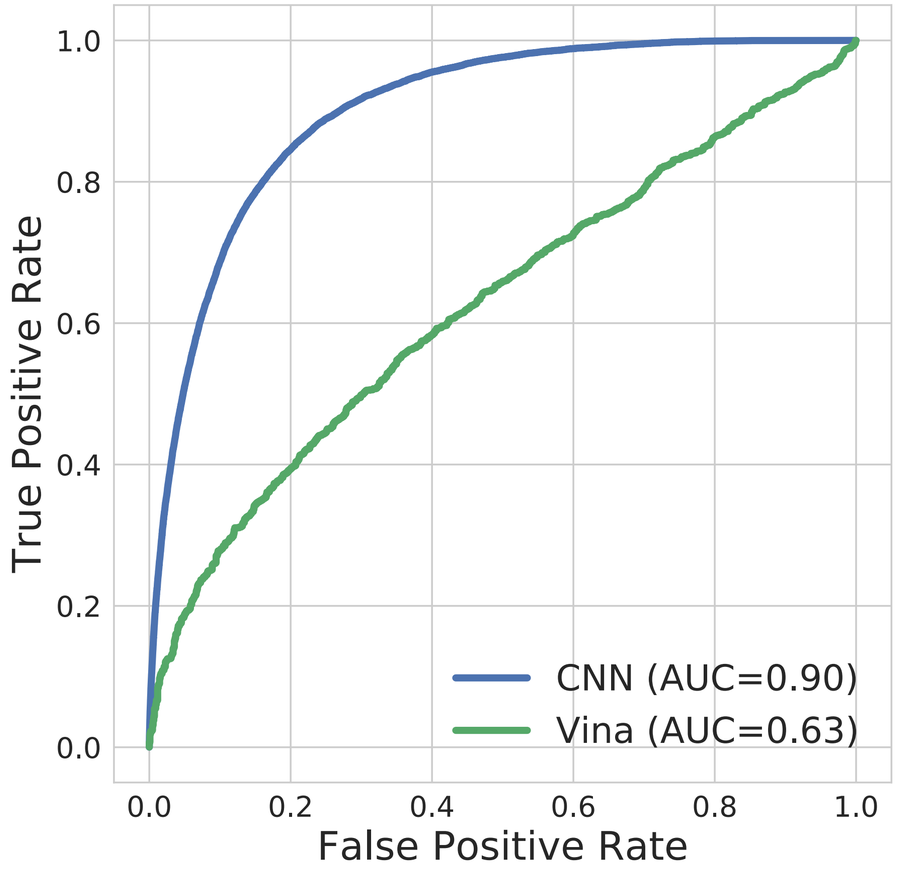


Model


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

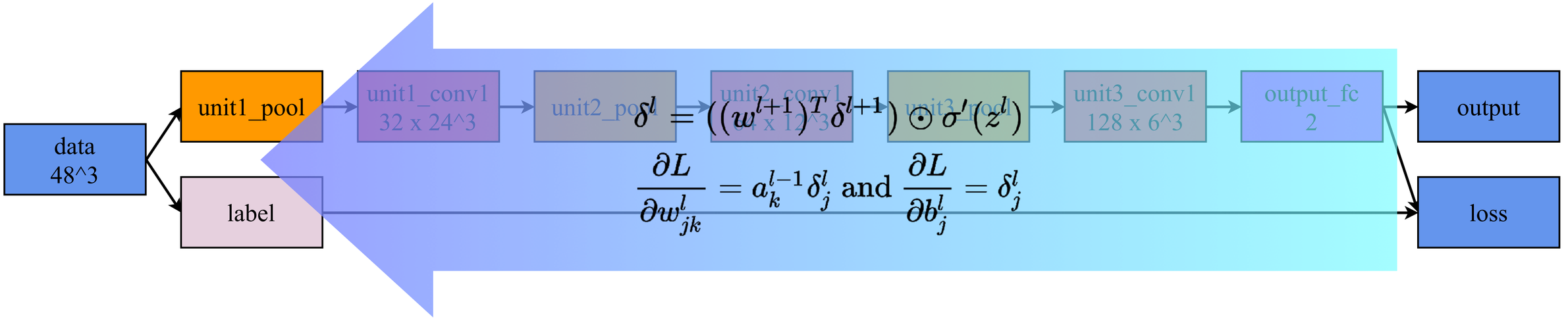




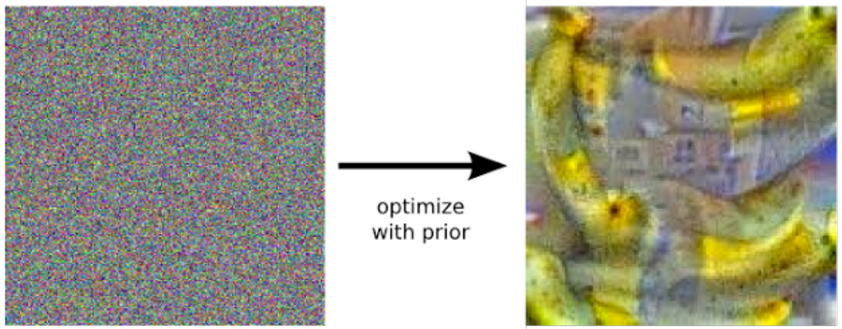
COMP 410: GNINA: Deep learning for molecular docking
Monday, Aug 20 8:00 PM
Exhibit Hall C, Boston Convention & Exhibition Center

COMP: Poster Session
Tuesday, Aug 21 6:00 PM
Exhibit Hall B1, Boston Convention & Exhibition Center

Beyond Scoring

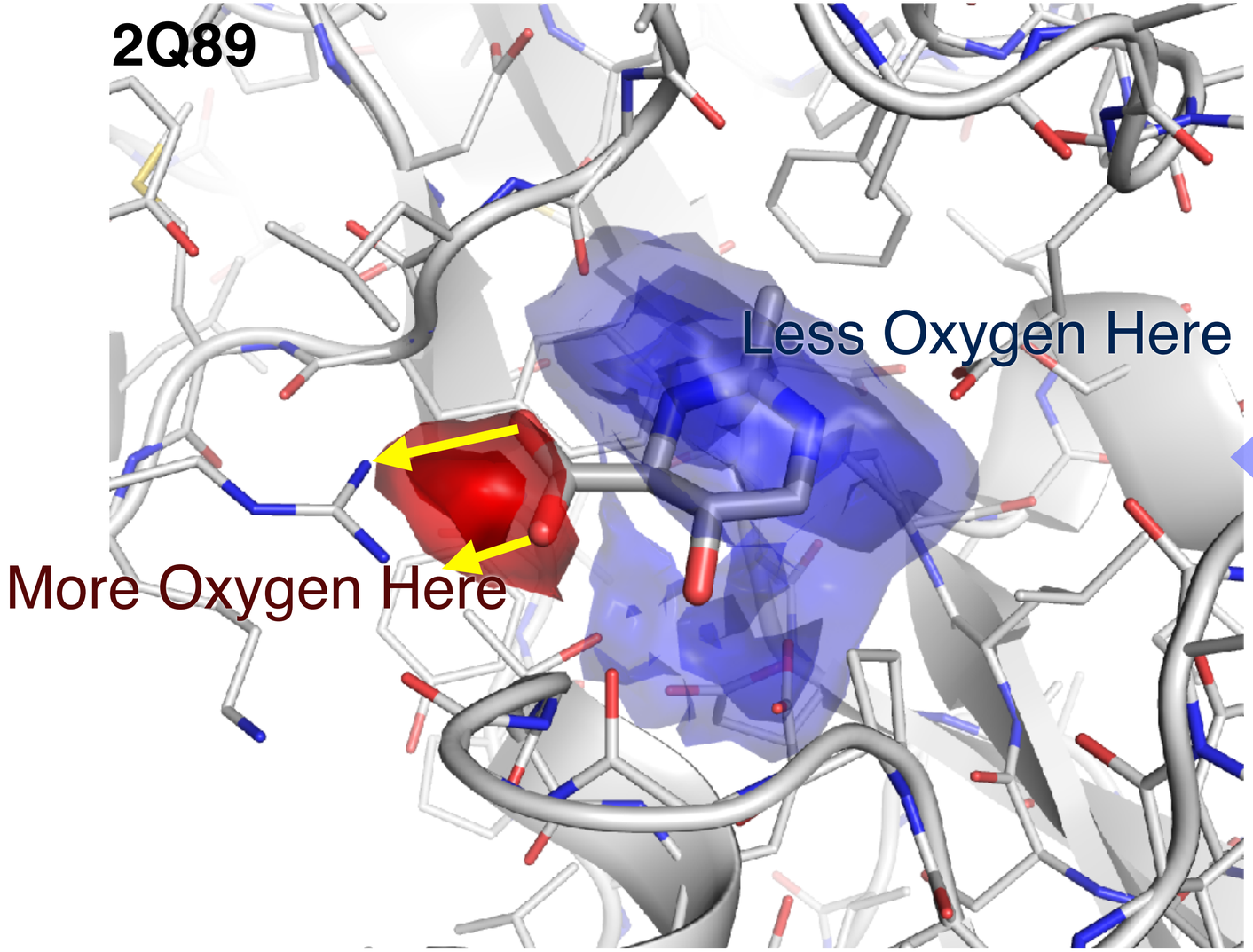


Deep Dreams

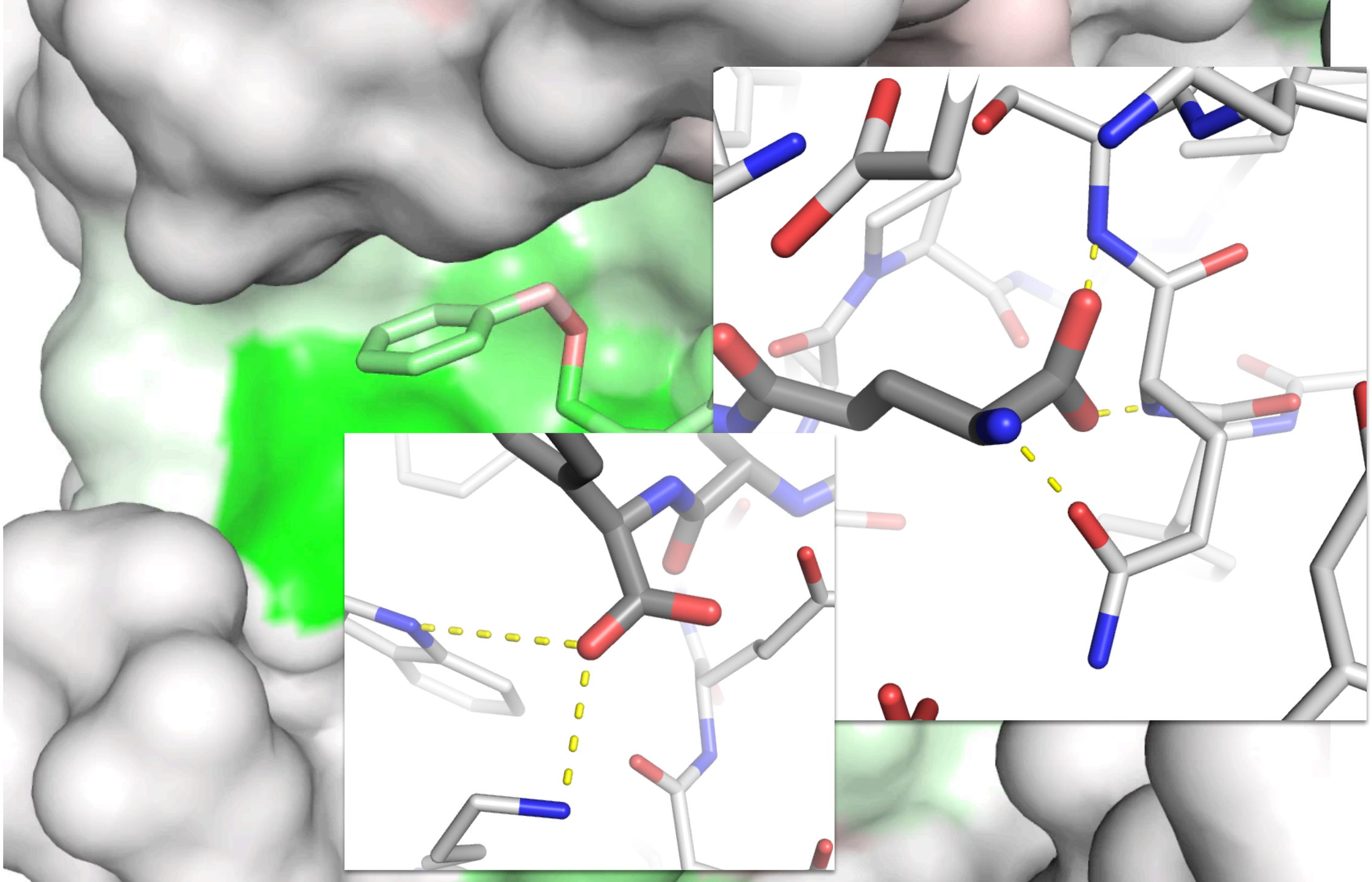


https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

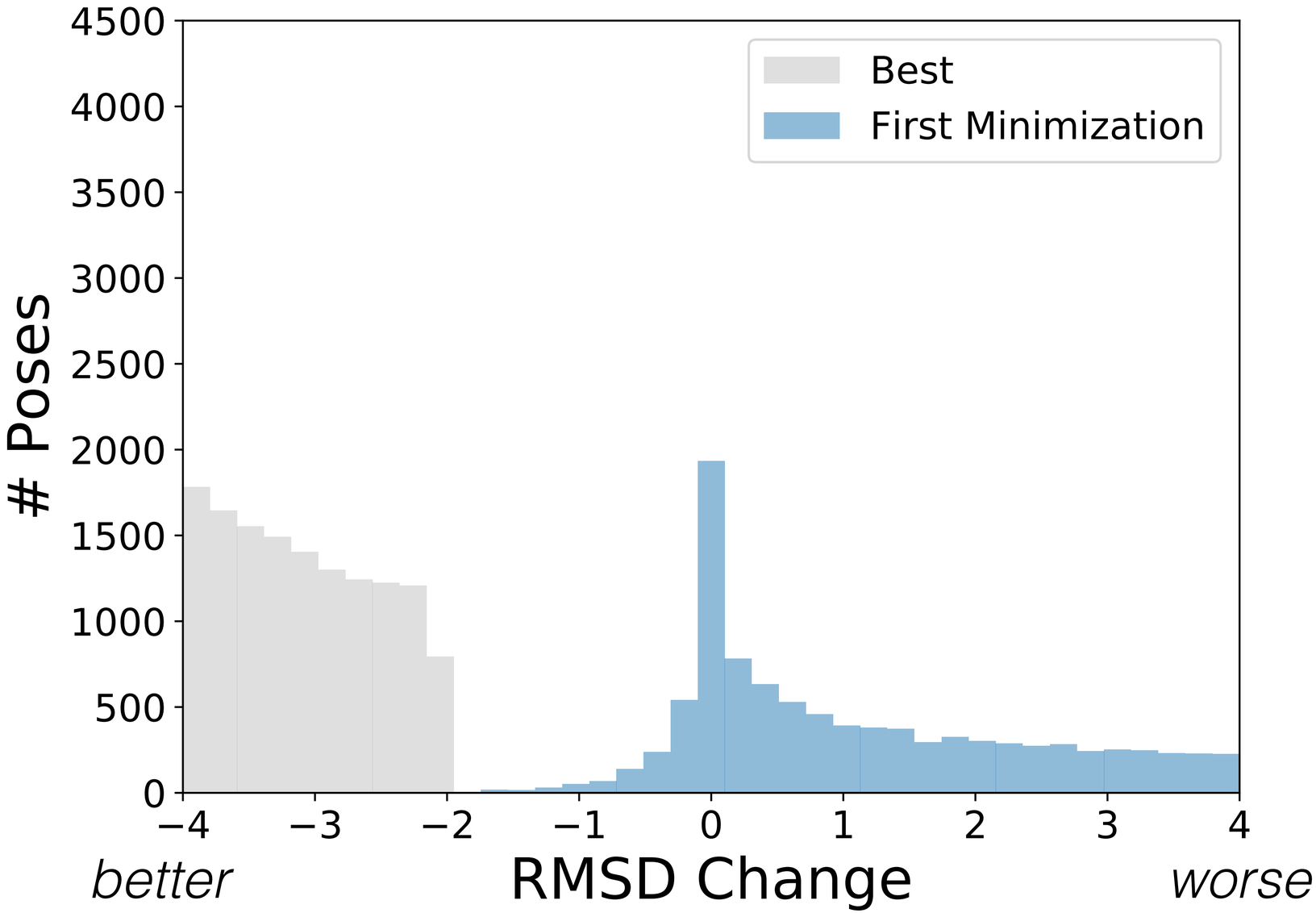
Beyond Scoring



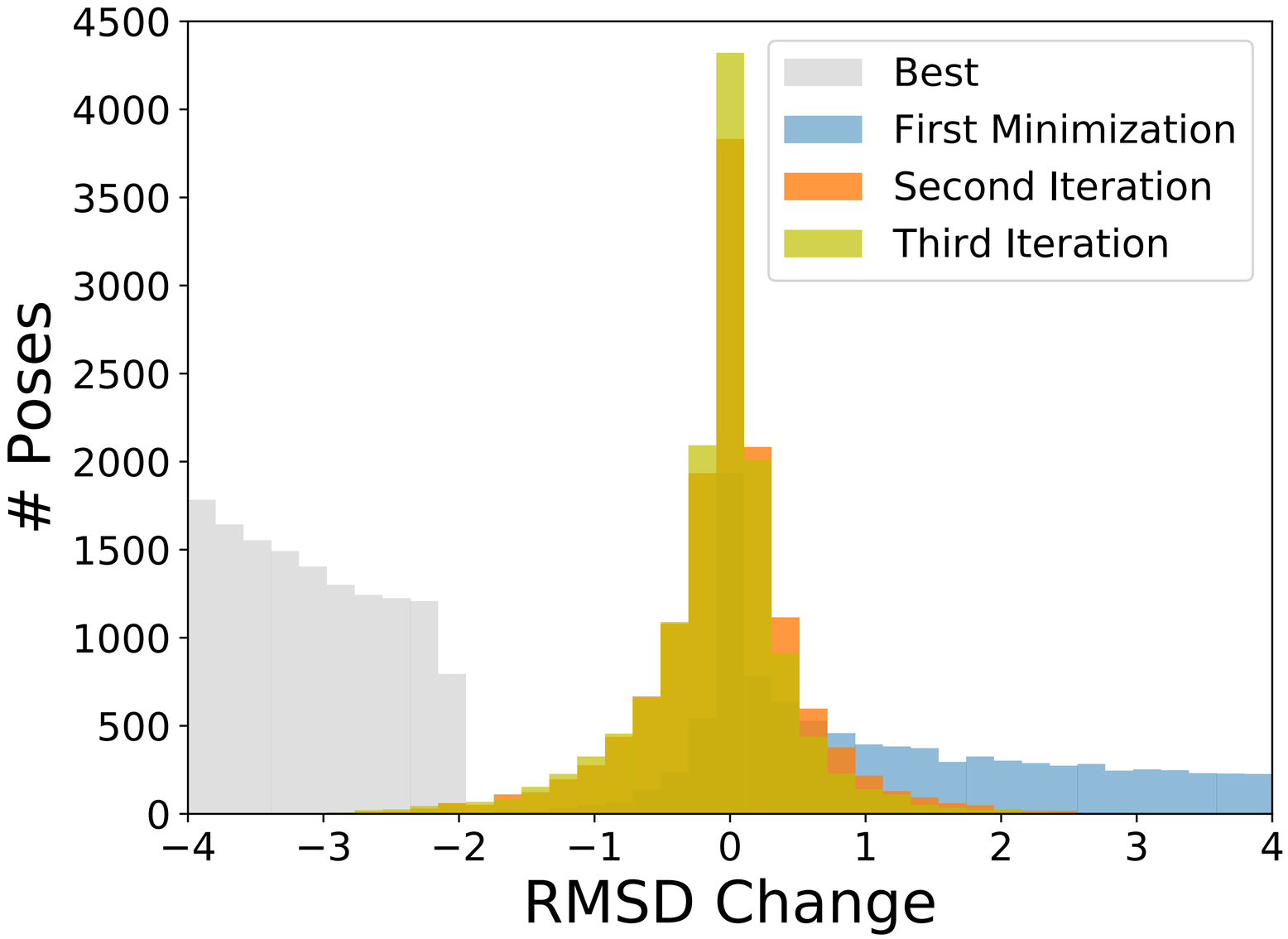
$$\frac{\partial L}{\partial A} = \sum_{i \in G_A} \frac{\partial L}{\partial G_i} \frac{\partial G_i}{\partial D} \frac{\partial D}{\partial A}$$

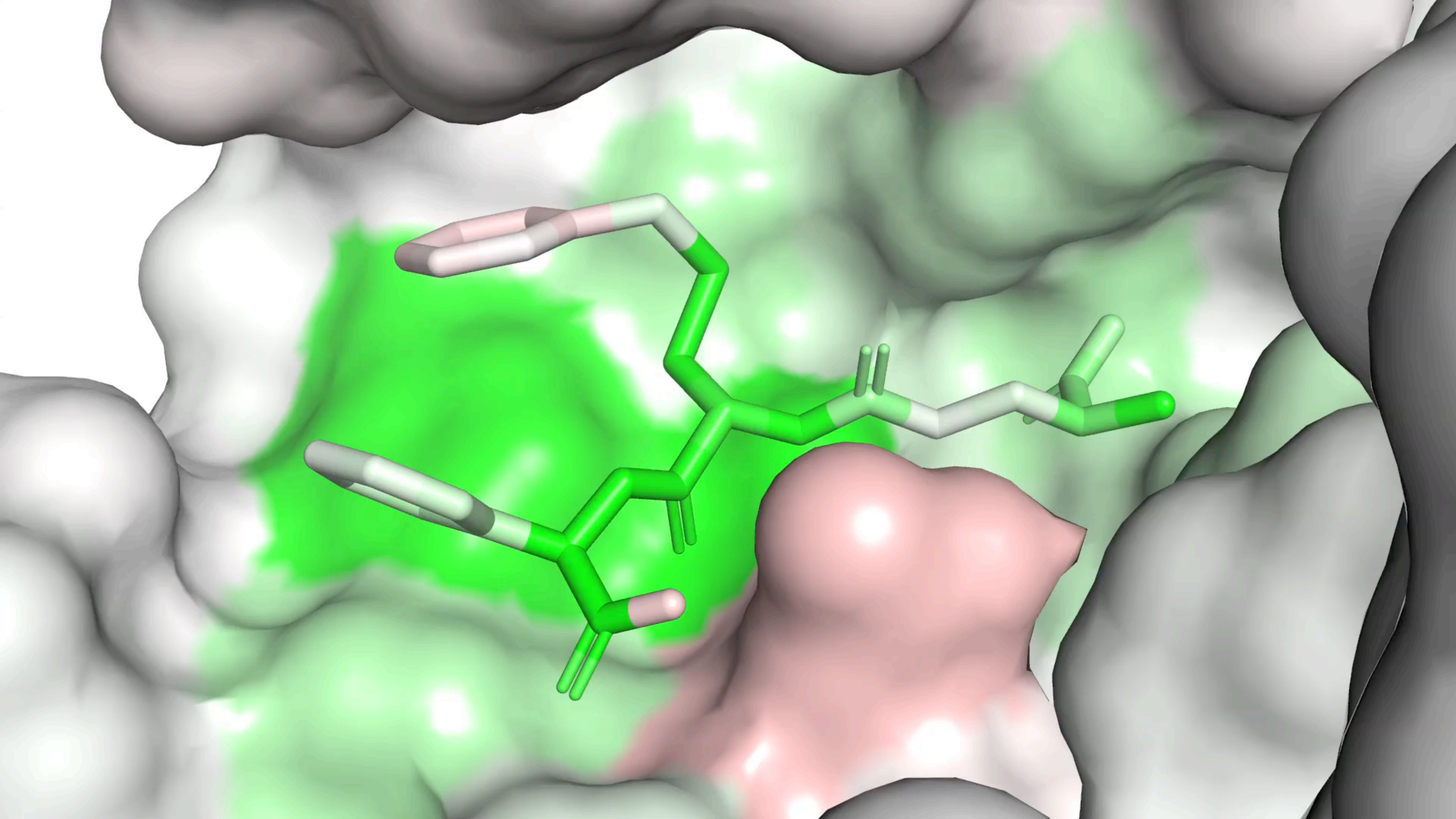


Minimizing Low RMSD Poses



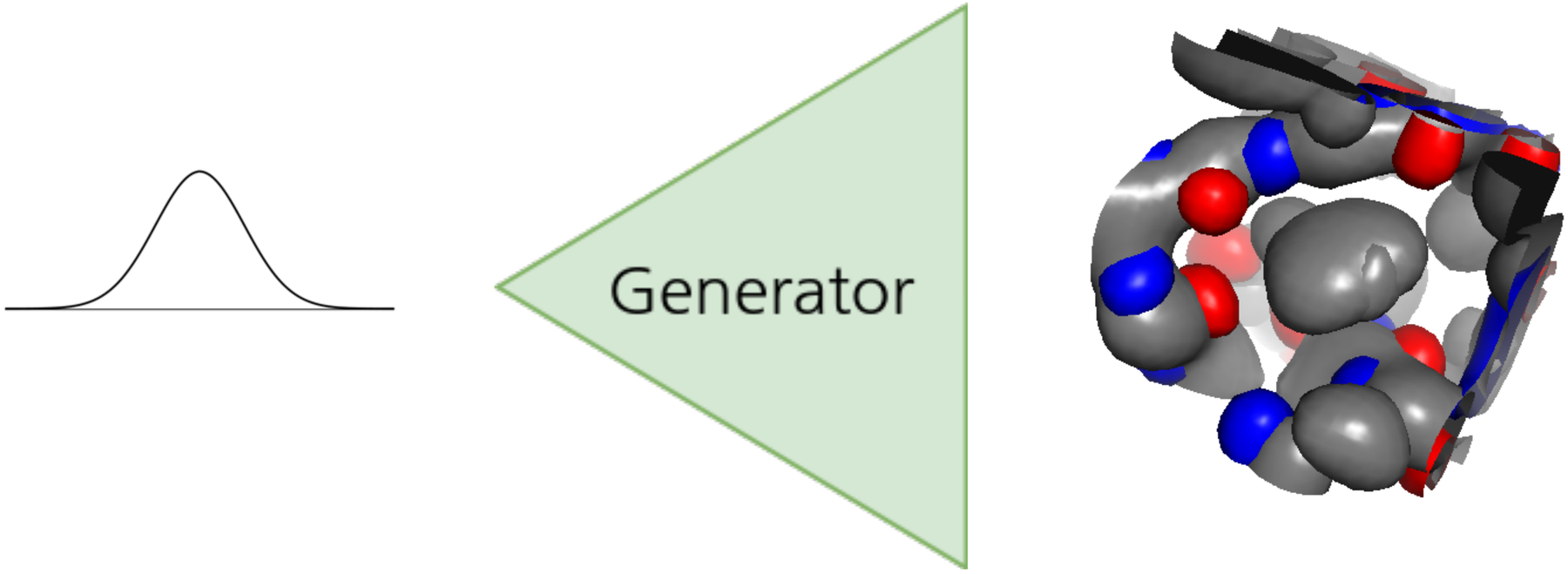
Iterative Refinement





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.

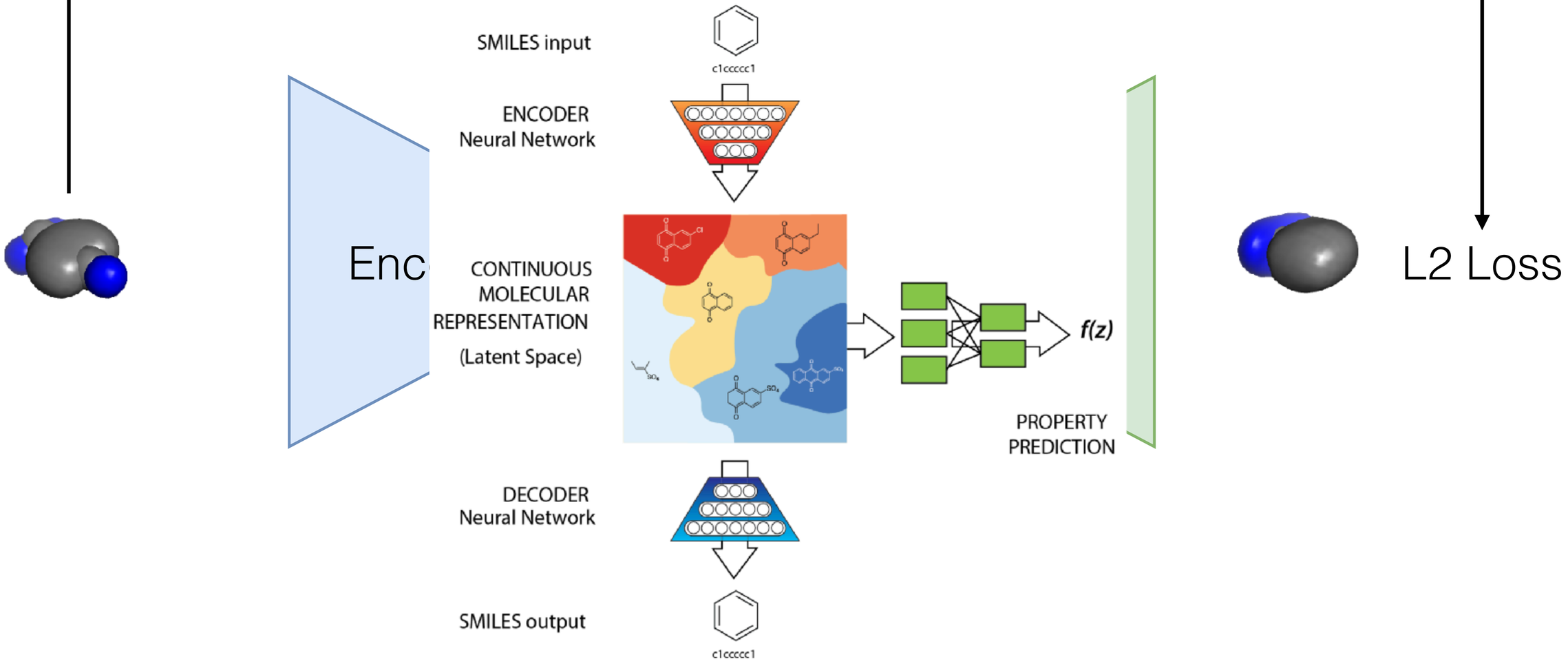


noise sample

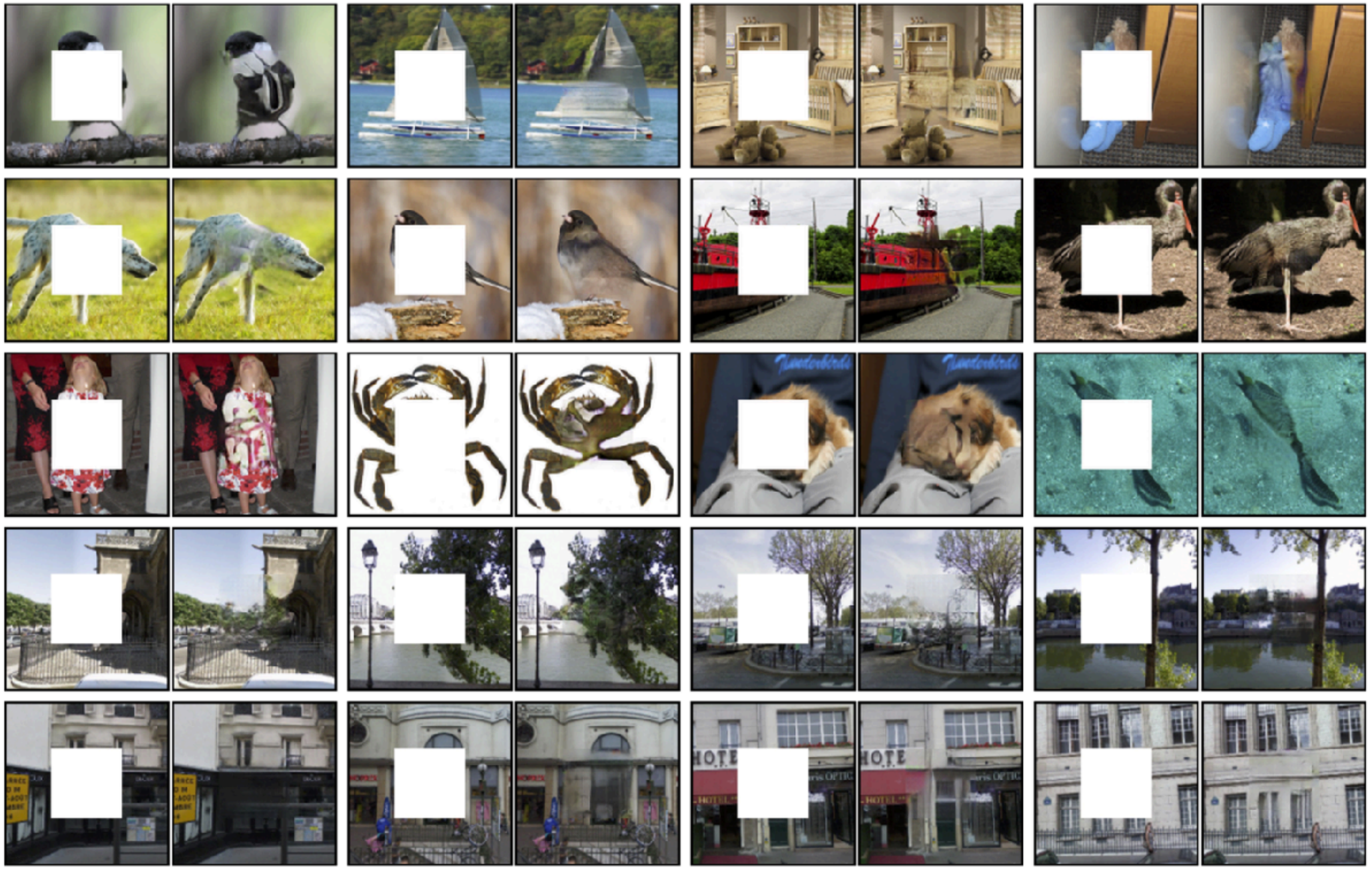
generated receptor & ligand grid

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[∇ı], and Alán Aspuru-Guzik[*‡⊥]

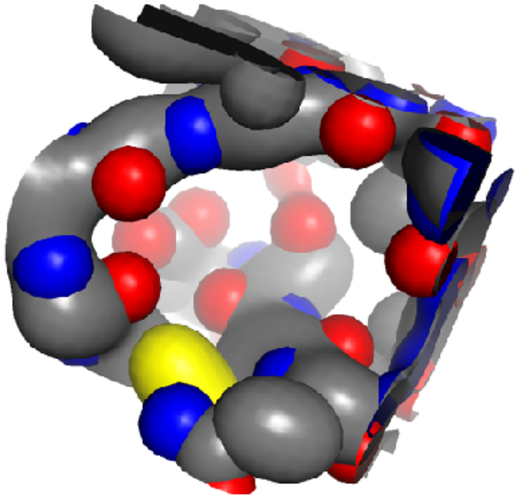


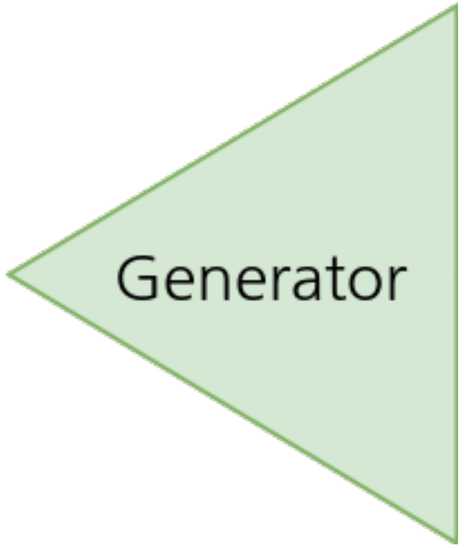
Context Encoding

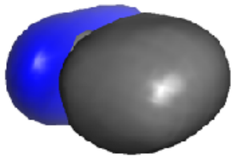


http://people.eecs.berkeley.edu/~pathak/context_encoder/

Context Encoding



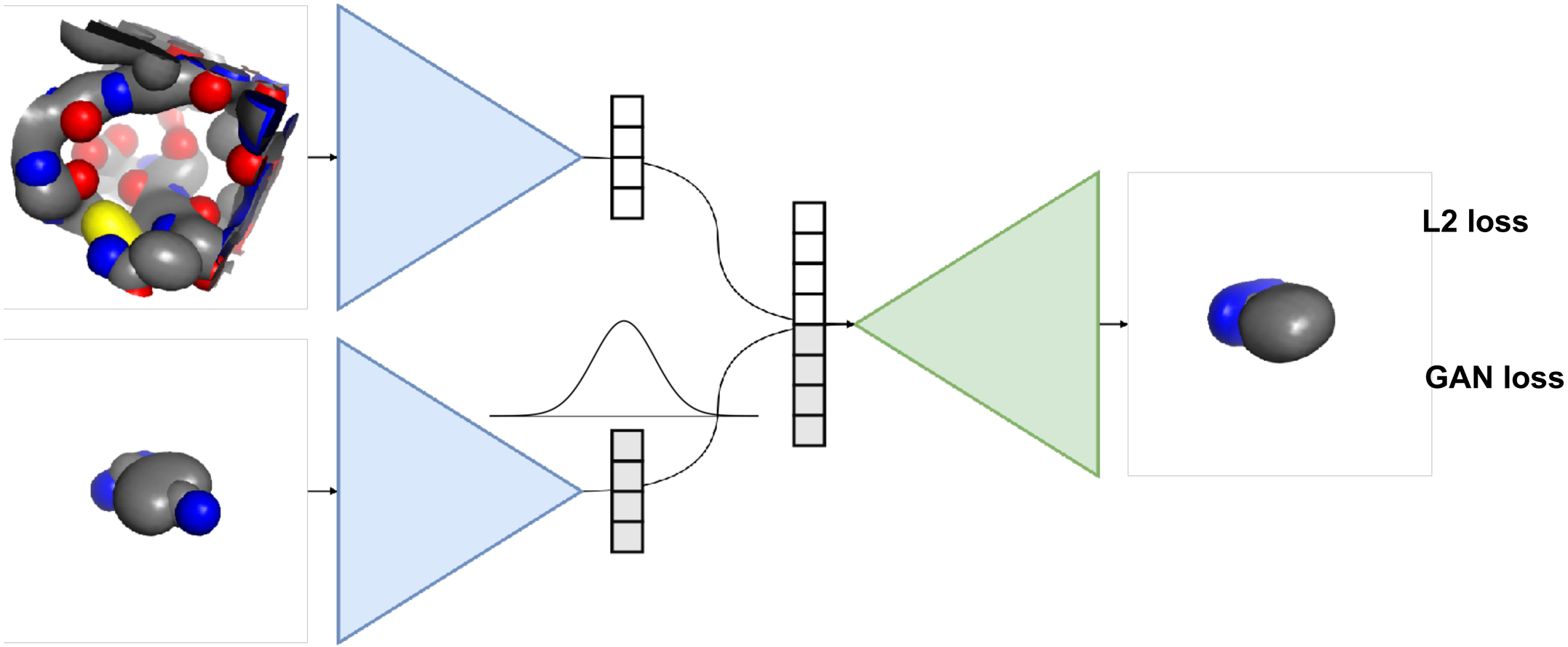




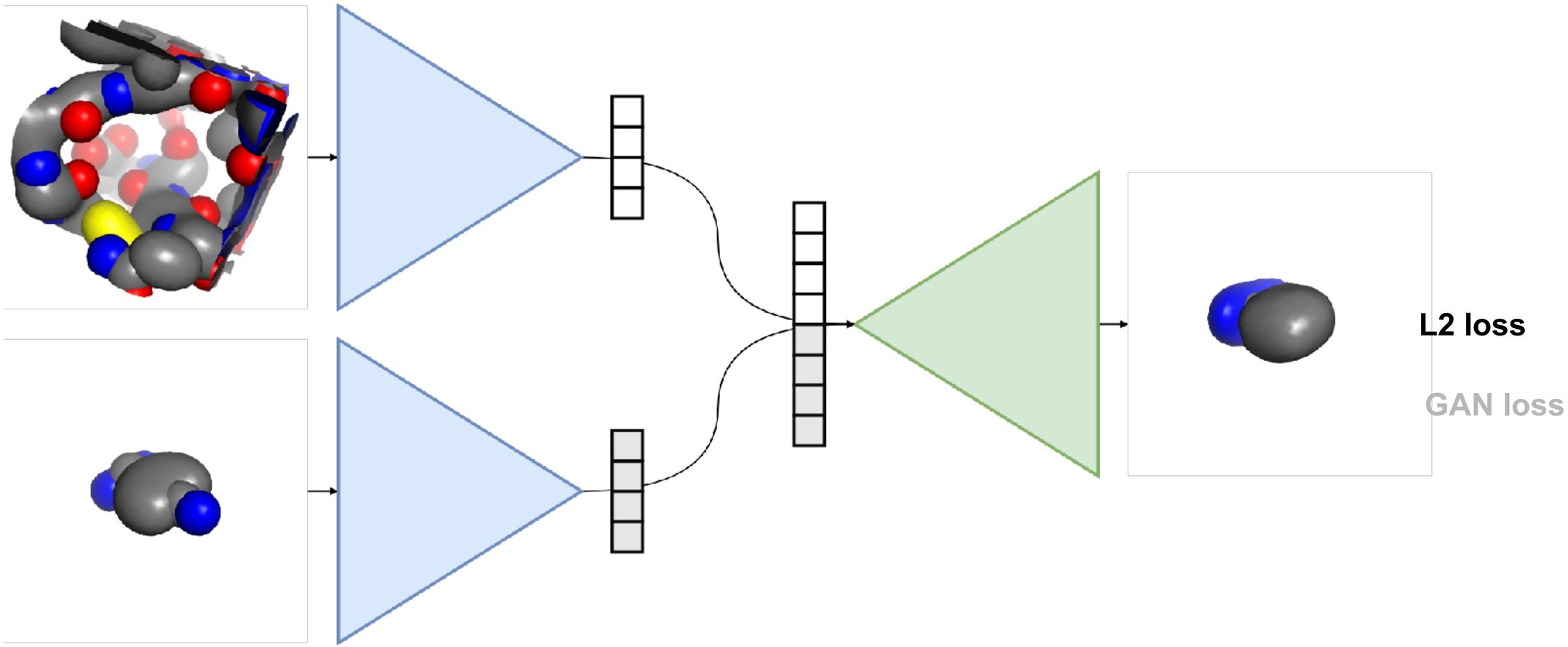
receptor grid

generated ligand grid

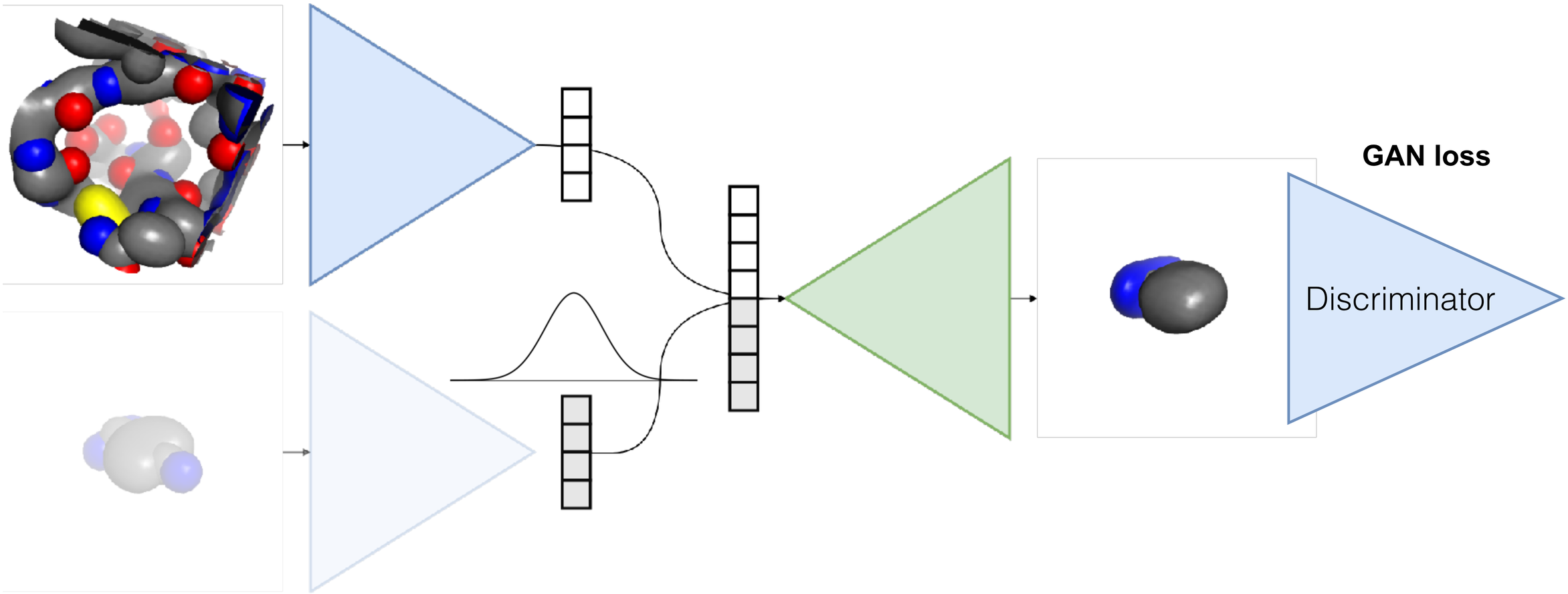
Receptor-Conditional Ligand-Variational Model



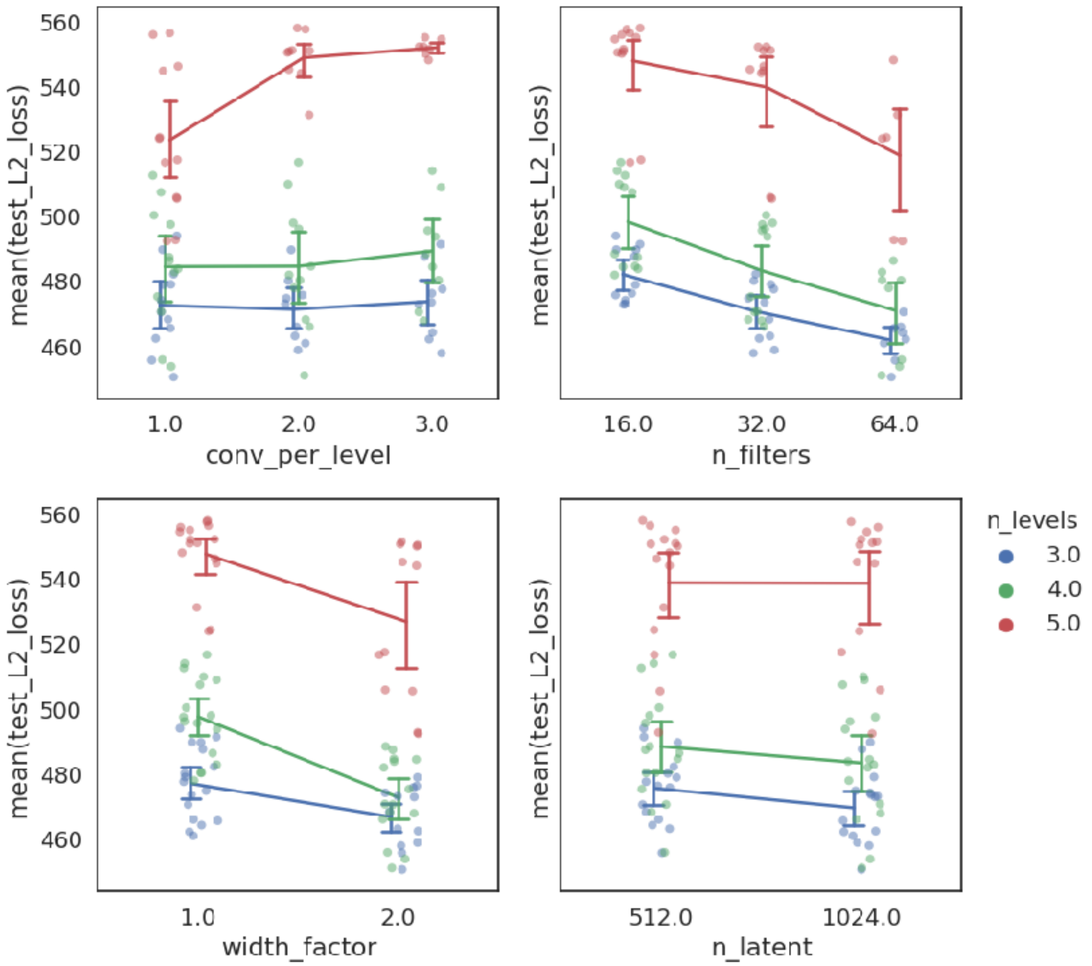
Receptor-Conditional Ligand-Variational Model

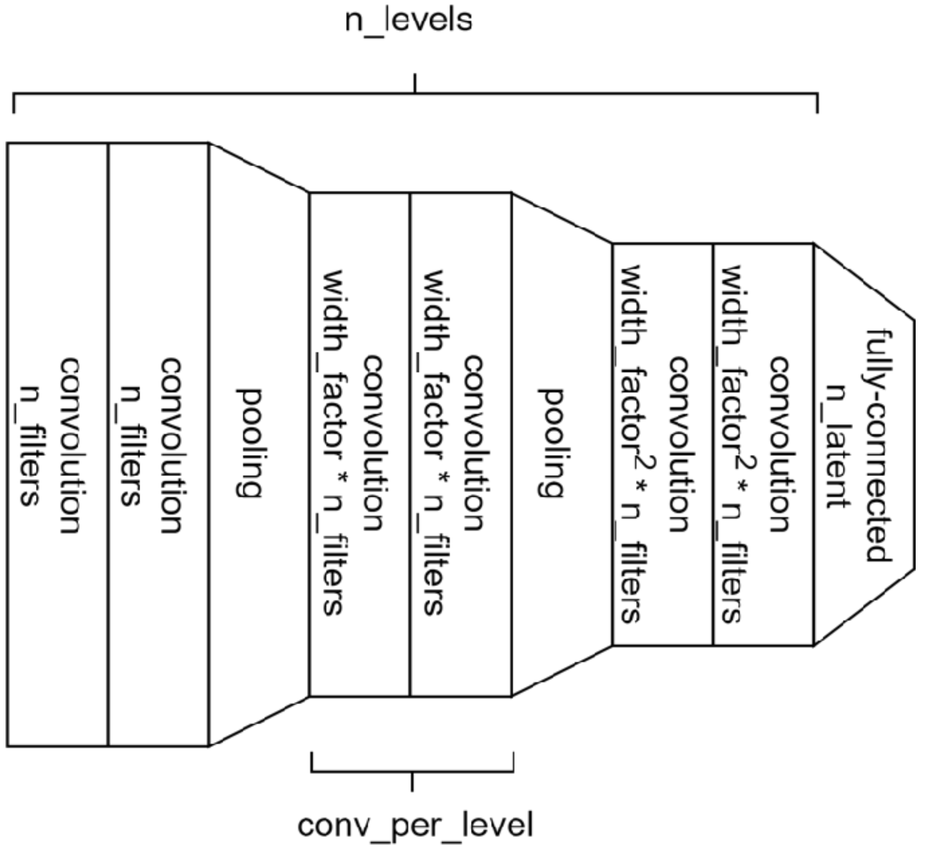


Receptor-Conditional Ligand-Variational Model



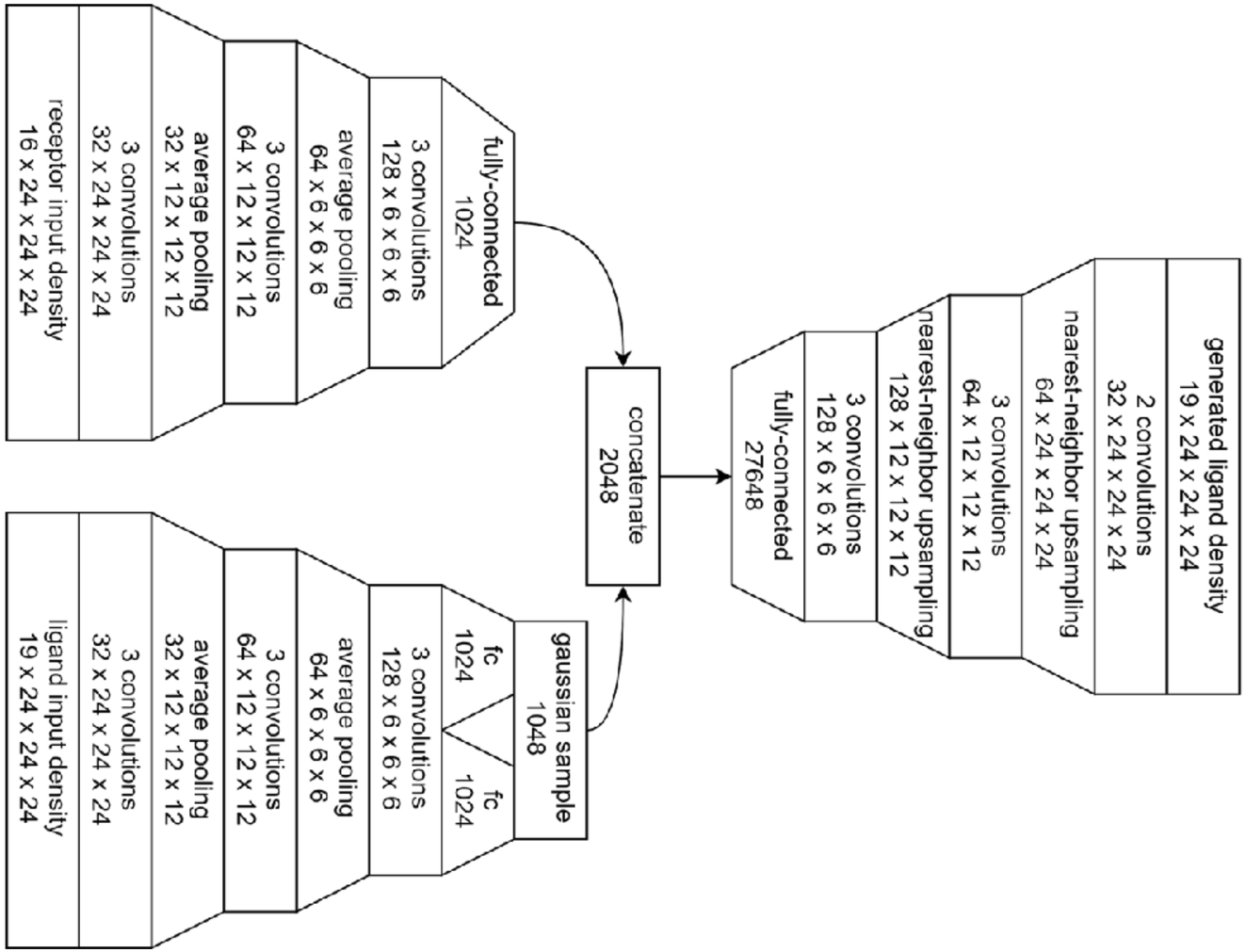
Model Architecture





Model Architecture

n_levels = 3
conv_per_level = 3
n_filters = 32
width_factor = 2
n_latent = 1024



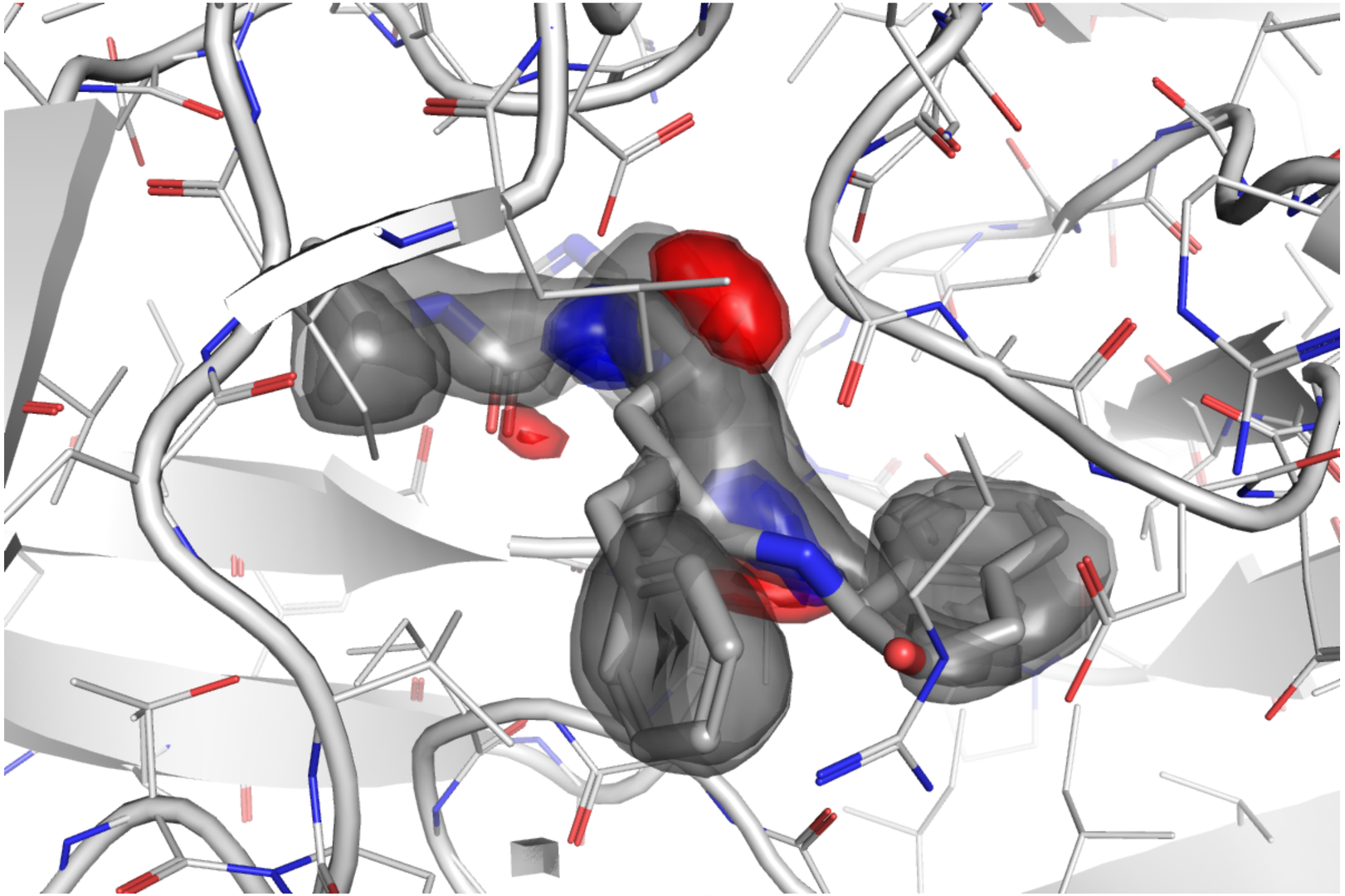
Training Procedure

2016 PDBbind refined set
3765 crystal structures
Vina docking
RMSD < 2 Å from crystal pose
8648 poses (~2.3 per target)
random rotation & translation

Adam optimization
base_lr = 0.00001
momentum = 0.9
momentum2 = 0.999
max_iter = 100000
batch_size = 50

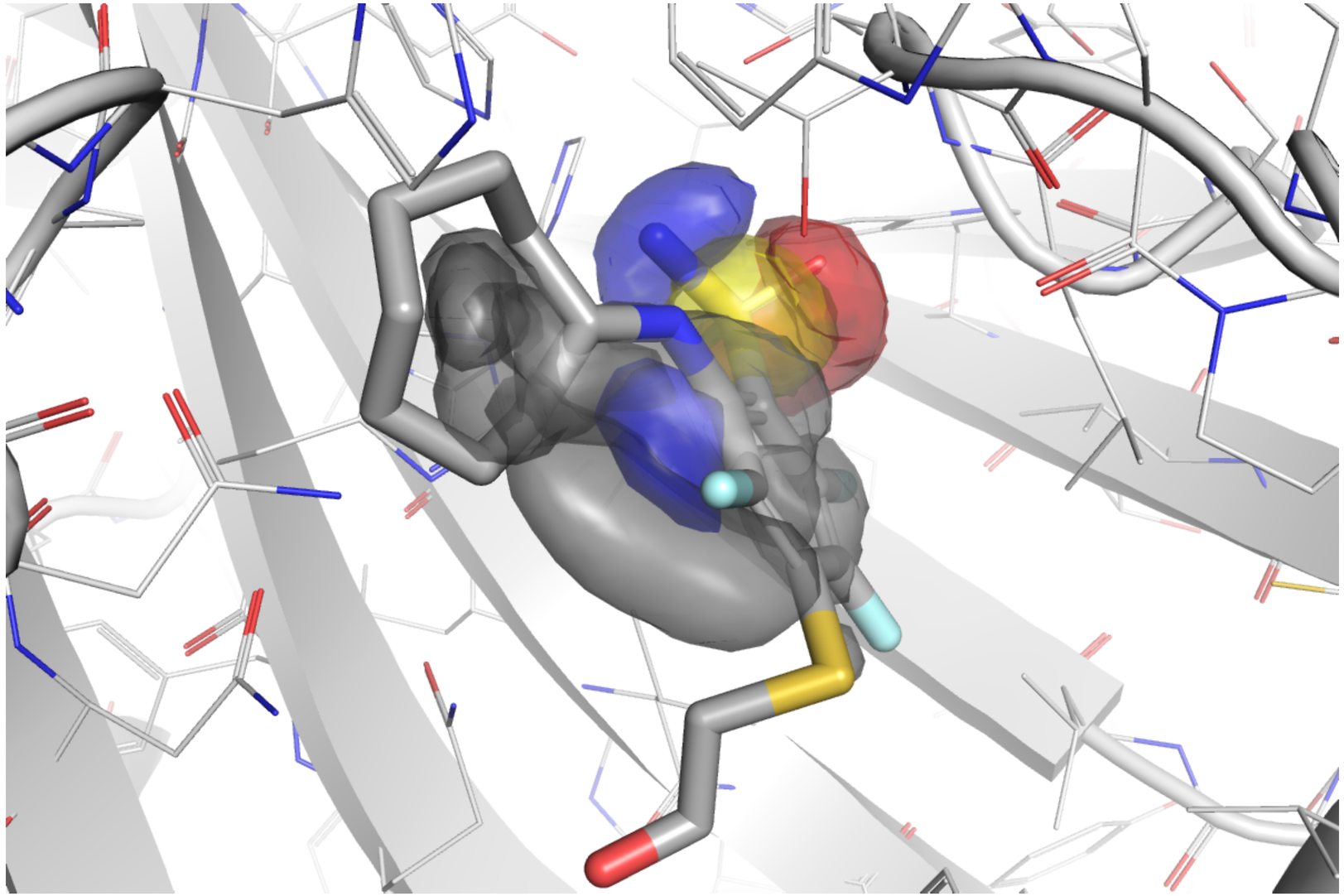
Caffe

Autoencoding Examples



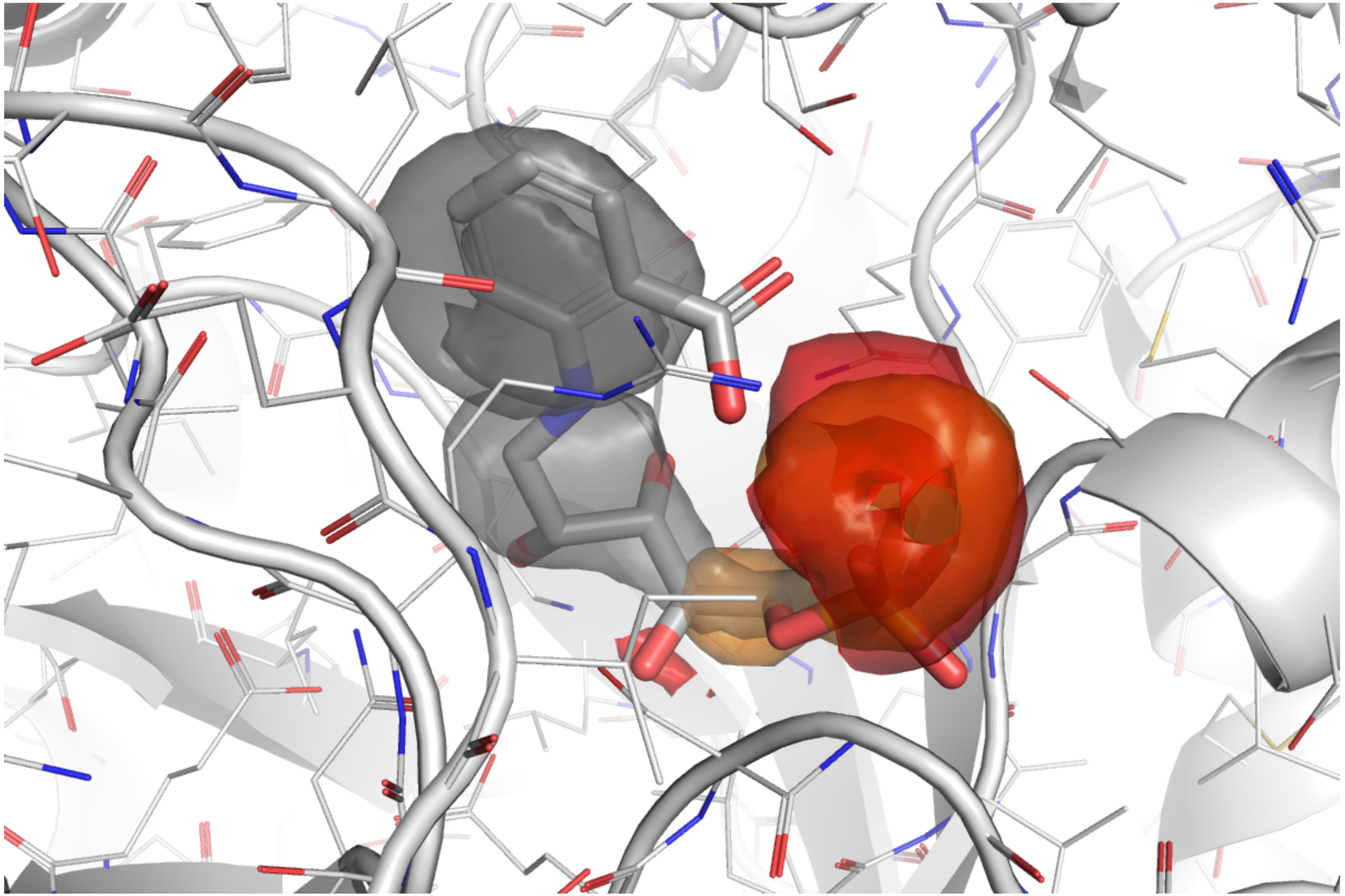
2AVO

Autoencoding Examples



4PYX

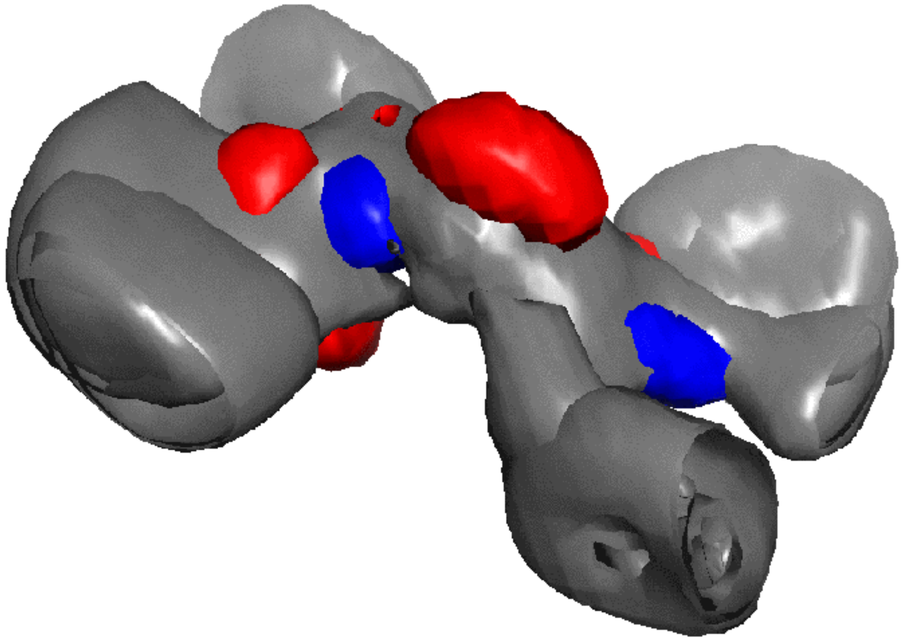
Autoencoding Examples



1LBF

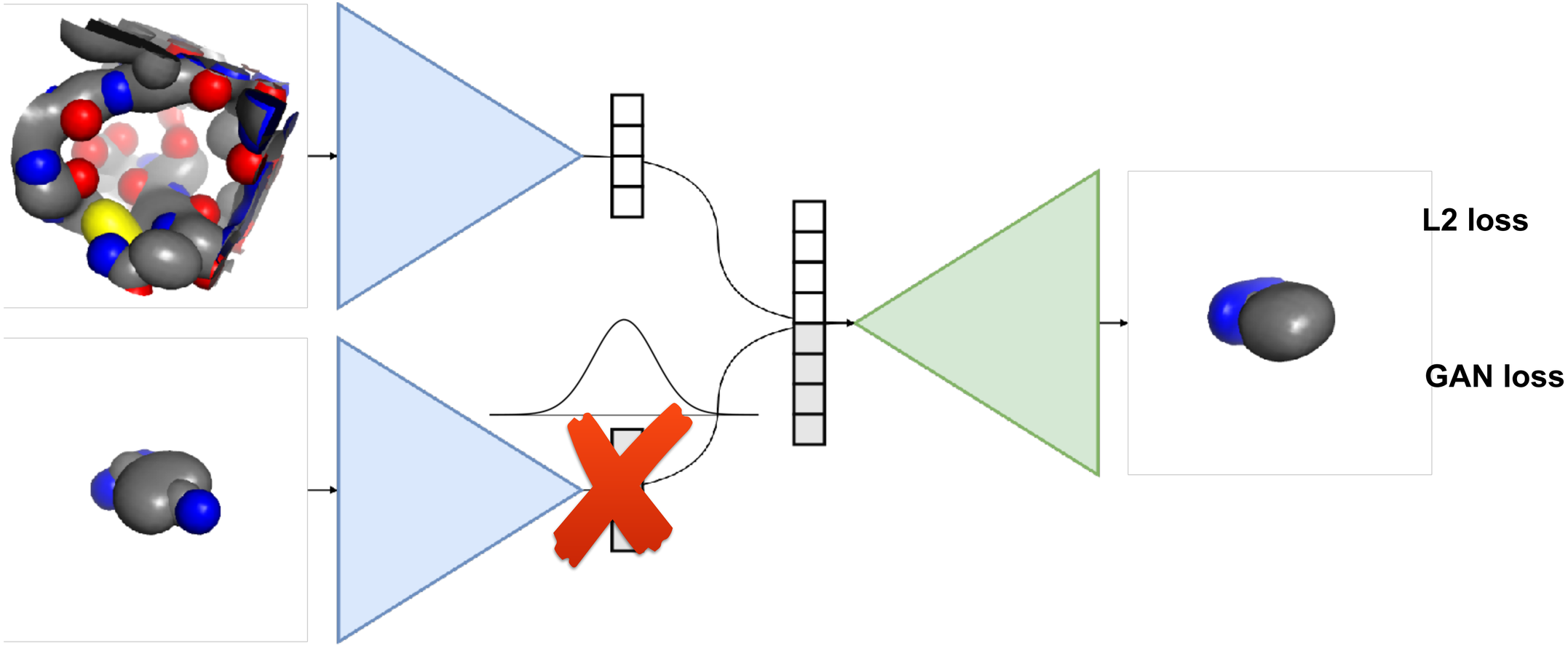
Atom Fitting

$$a^* = \underset{a}{\operatorname{argmin}} \|d - D(a)\|_2^2 + \lambda E(a)$$

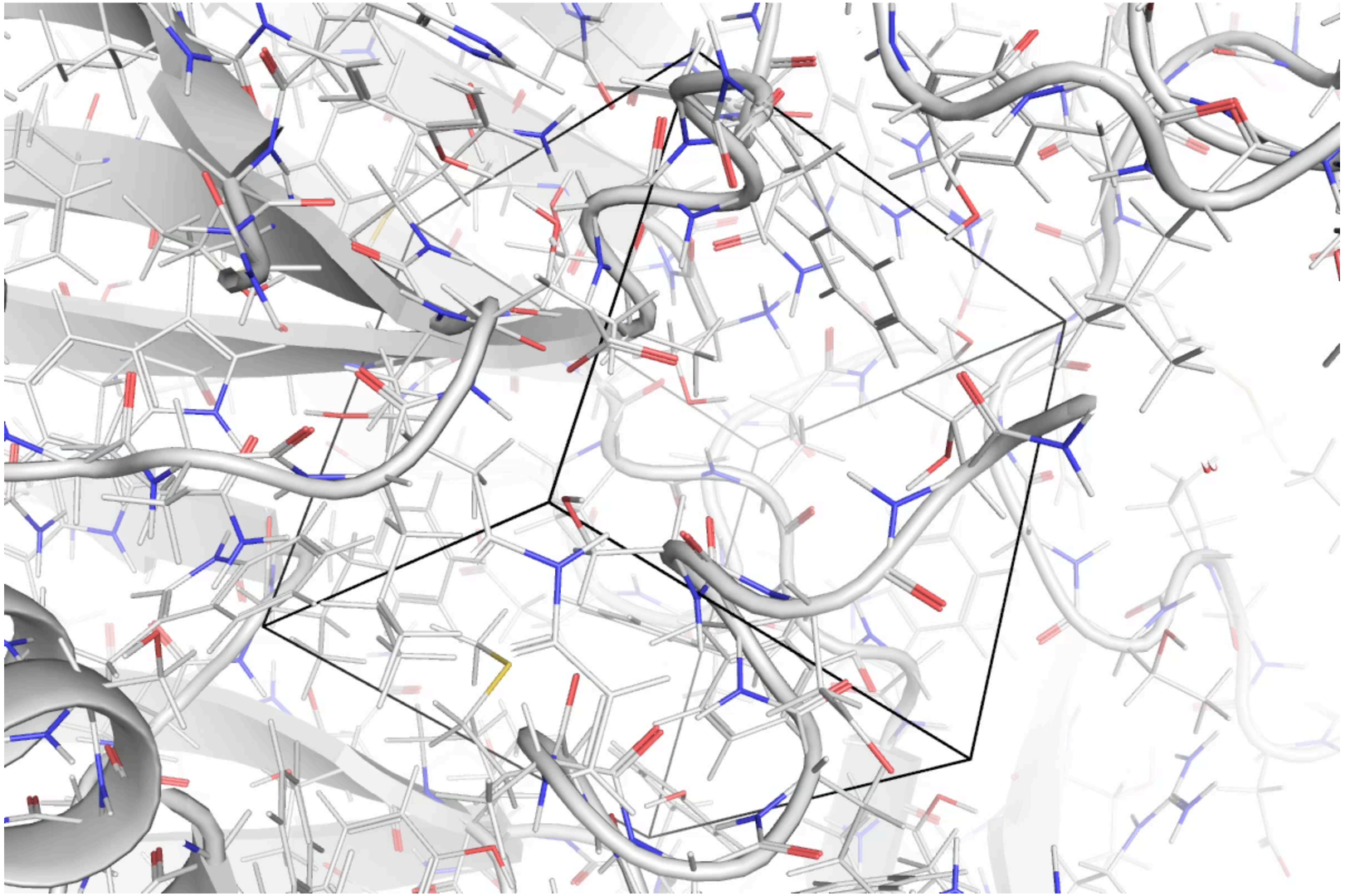


Ligand Variation

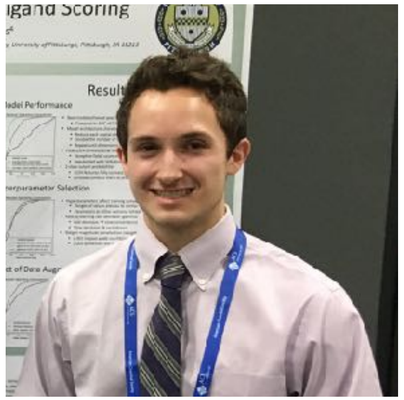
to be continued…



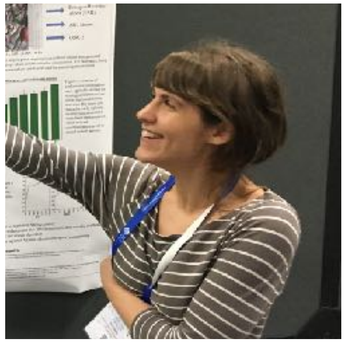
Conditioning on the Receptor



Acknowledgements



Matt Ragoza



Jocelyn Sunseri

Paul Francoeur

Department of Computational and Systems Biology

AI GRANT

National Institute of General Medical Sciences
R01GM108340

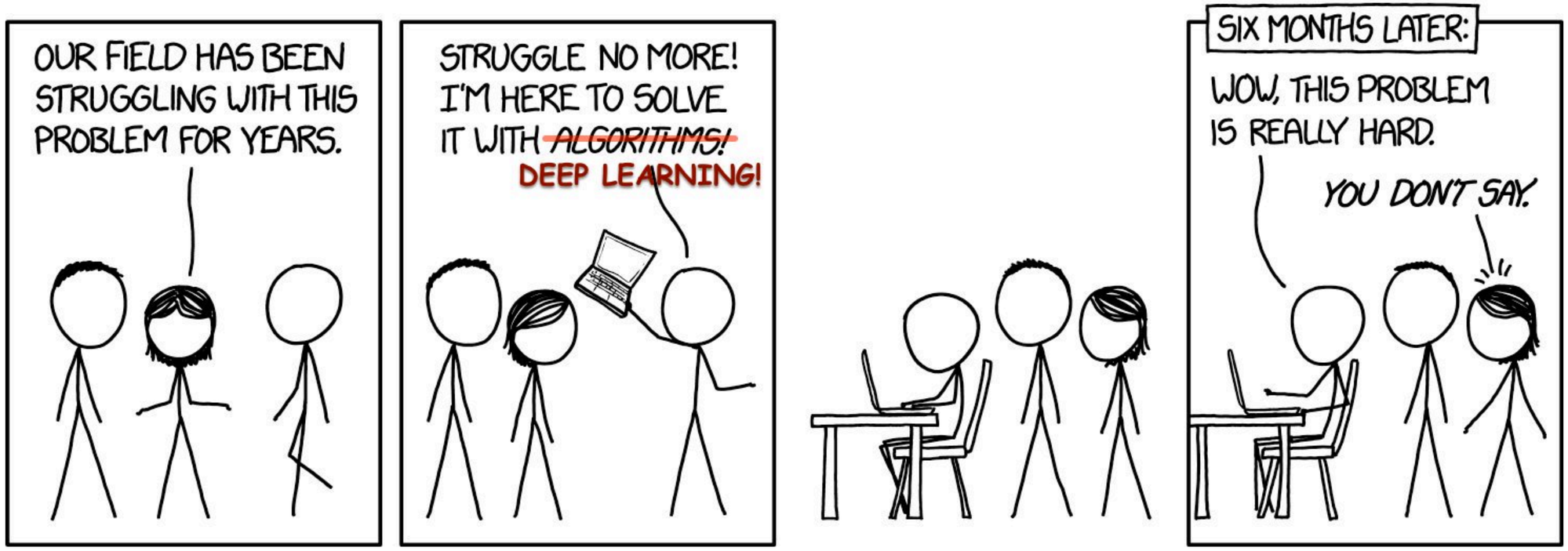
NVIDIA

Google Cloud

github.com/gnina

http://bits.csb.pitt.edu

@david_koes



COMP: Poster Session
Tuesday, Aug 21 6:00 PM
Exhibit Hall B1, Boston Convention & Exhibition Center

COMP 528: Structure-based searching of chemical space with Pharmit
Thursday, Aug 23 9:20 AM
Douglass, Westin Boston Waterfront