# Deep Learning for Drug Discovery

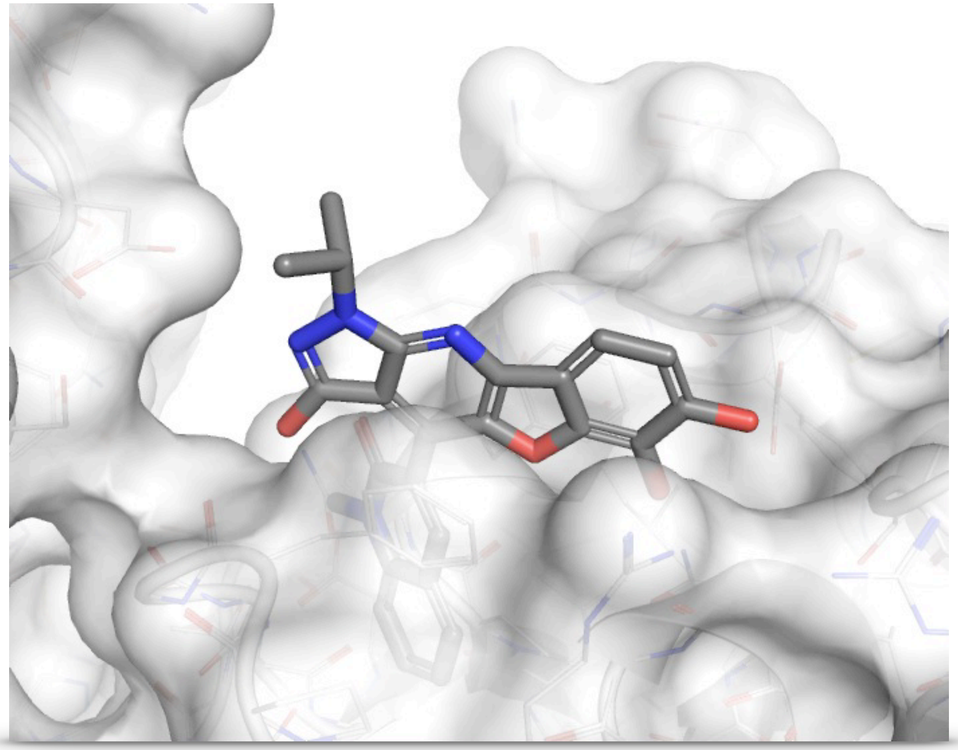
David Koes

@david_koes

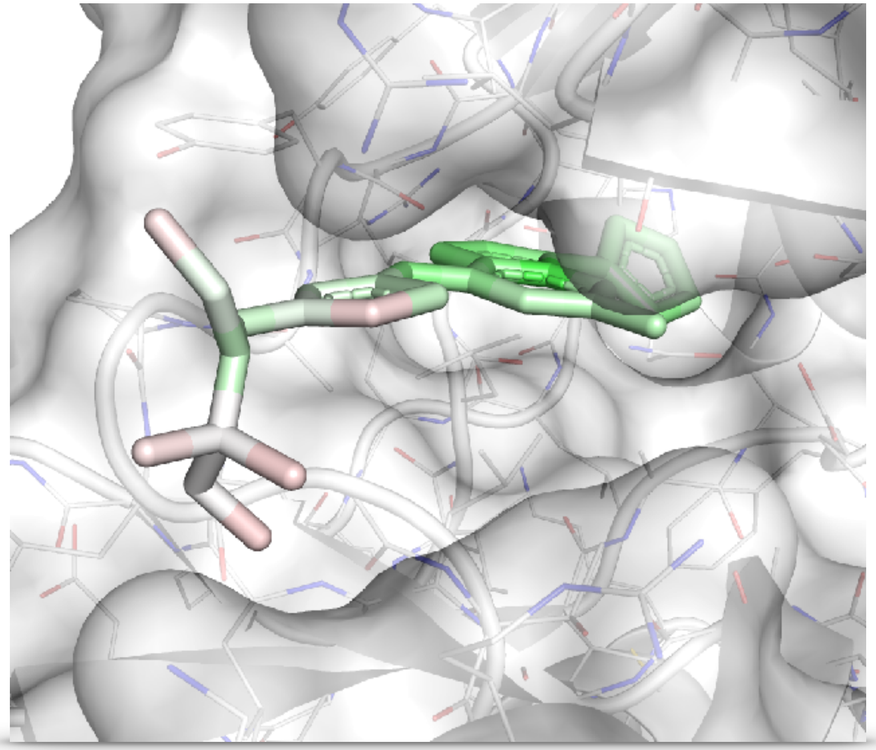
2nd RSC-BMCS / RSC-CICAG
Artificial Intelligence in Chemistry
Cambridge, UK
September 3, 2019

# Structure Based Drug Design

Pose Prediction → Binding Discrimination → Affinity Prediction



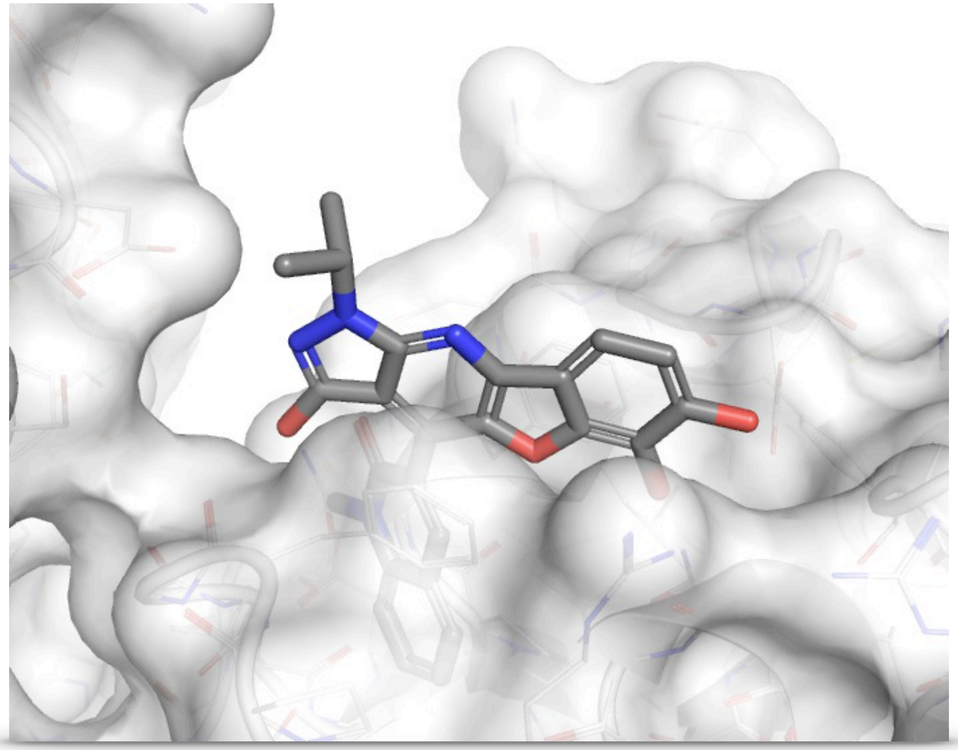
**Virtual Screening**



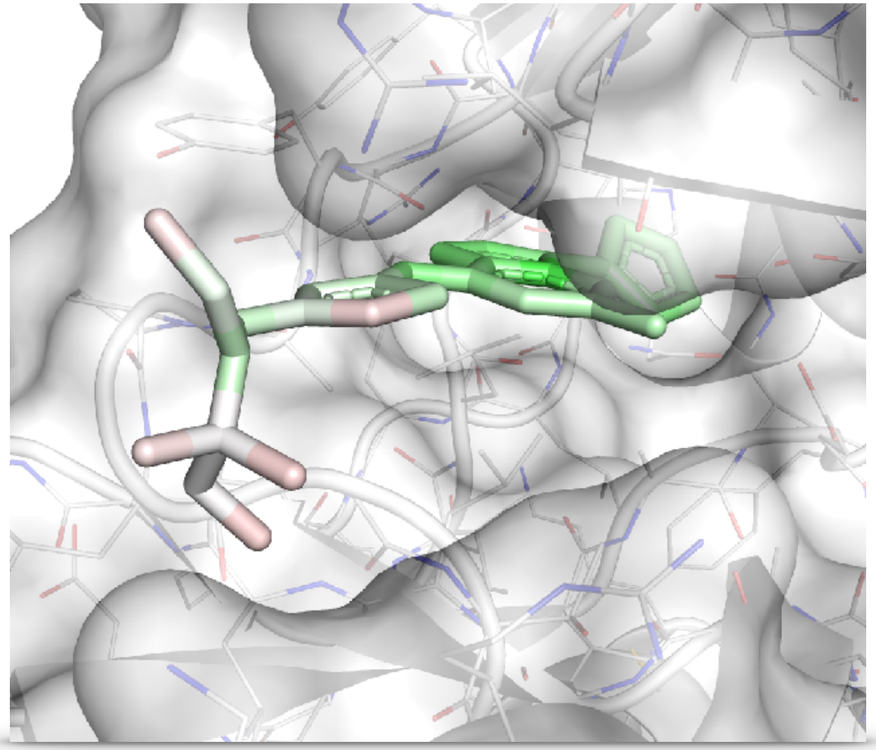
**Lead Optimization**

# Structure Based Drug Design

Pose Prediction → Binding Discrimination → Affinity Prediction

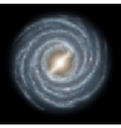


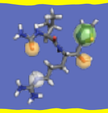
**Virtual Screening**



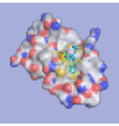
**Lead Optimization**

# Drug Discovery Funnel

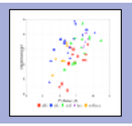
Purchasable  Accessible

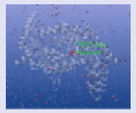


**Matching**

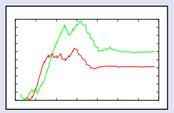


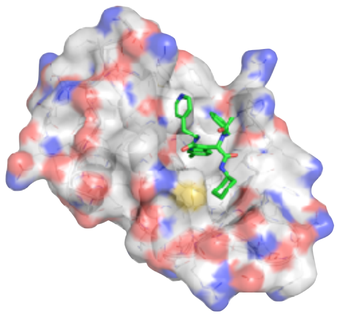
**Scoring**

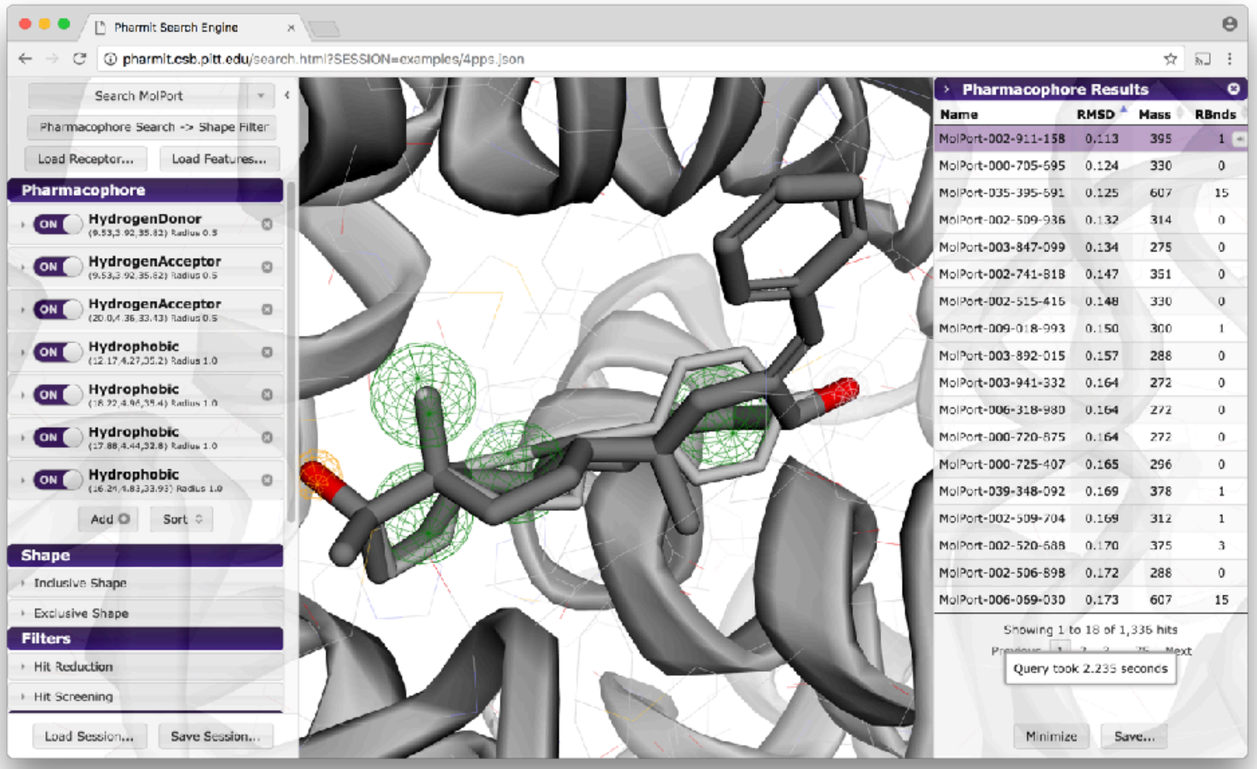




**Dynamics**

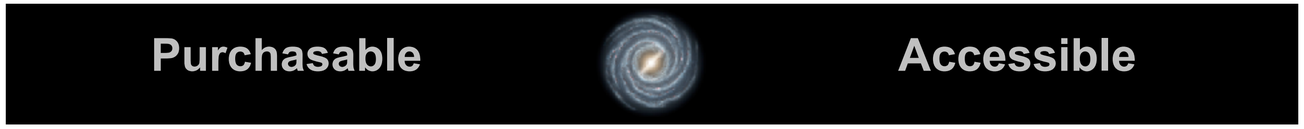






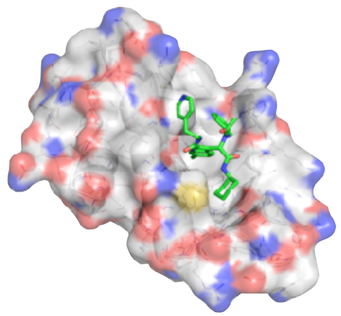

http://pharmit.csb.pitt.edu

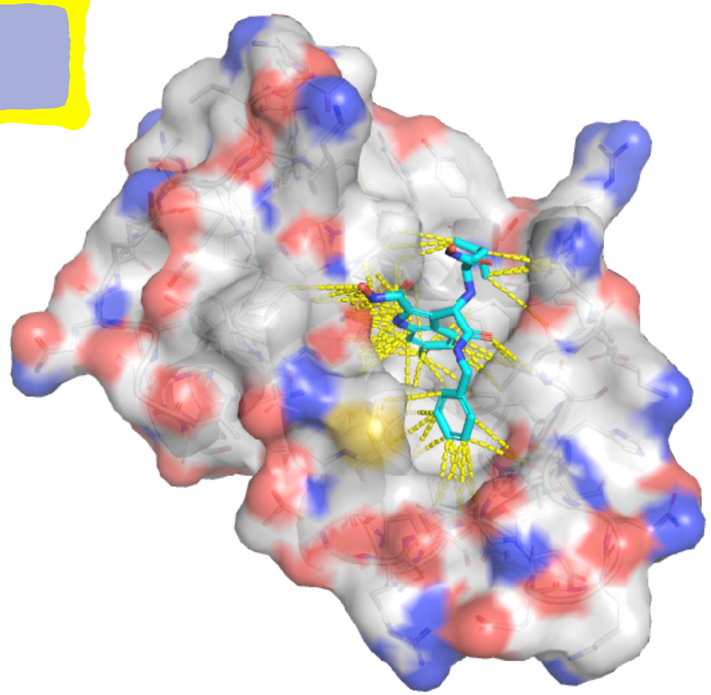
# Drug Discovery Funnel







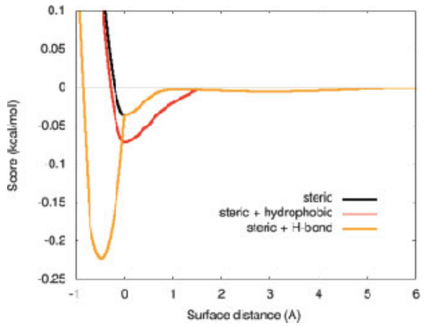




$$\text{gauss}_1(d) = w_{\text{guass}_1} e^{-(d/0.5)^2}$$
$$\text{gauss}_2(d) = w_{\text{guass}_2} e^{-((d-3)/2)^2}$$
$$\text{repulsion}(d) = \begin{cases} w_{\text{repulsion}} d^2 & d < 0 \\ 0 & d \geq 0 \end{cases}$$

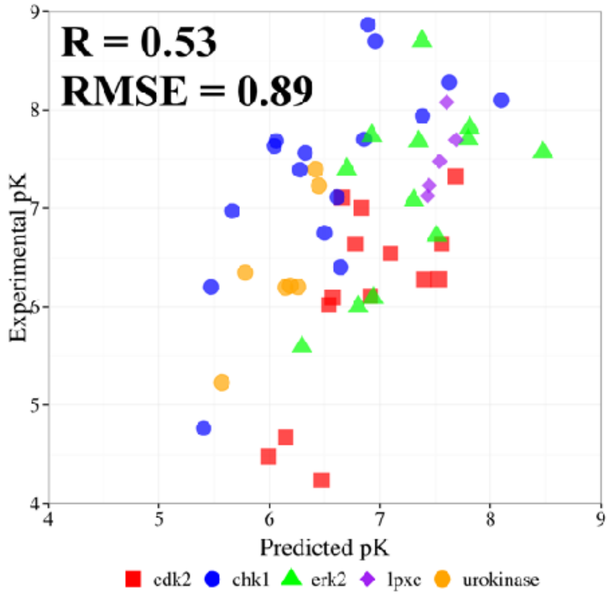
$$\text{hydrophobic}(d) = \begin{cases} w_{\text{hydrophobic}} & d < 0.5 \\ 0 & d > 1.5 \\ w_{\text{hydrophobic}}(1.5 - d) & otherwise \end{cases}$$

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

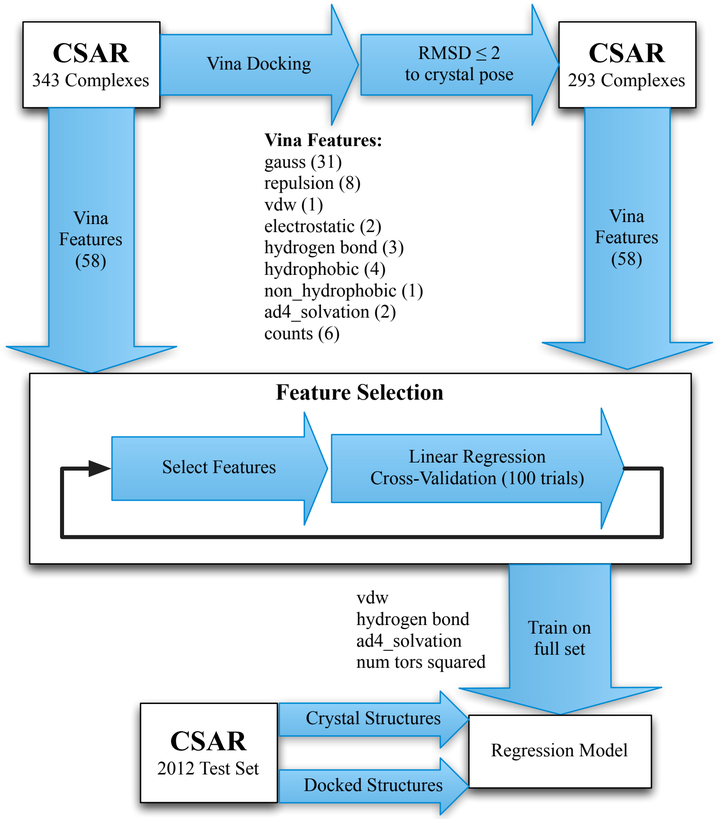


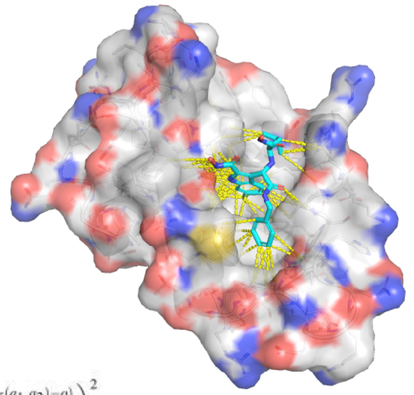

O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry* 31 (2010) 455-461

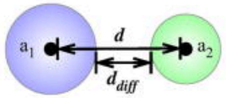
# Scoring

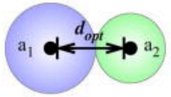



smina.sf.net






$$\text{gauss}(a_1,a_2,d) = e^{\left(\frac{-(d_{diff}(a_1,a_2)-o)}{w}\right)^2}$$

$$\text{repulsion}(a_1,a_2,d) = \begin{cases}(d_{diff}(a_1,a_2)-o)^2, & d_{diff}(a_1,a_2) < o\\ 0, & \text{otherwise}\end{cases}$$

$$\text{vdw}(a_1,a_2,d) = \left(\frac{d_{opt}(a_1,a_2)}{d}\right)^8 - 2\left(\frac{d_{opt}(a_1,a_2)}{d}\right)^4$$

$$\text{electrostatic}(a_1,a_2,d) = \frac{partial\_charge(a_1)\cdot partial\_charge(a_2)}{d^x}$$

$$\text{hydrogen\_bond}(a_1,a_2,d) = \begin{cases}0, & (a_1,a_2)\text{ do not form hydrogen bond}\\ 1, & d_{diff}(a_1,a_2) < -0.7\\ 0, & d_{diff}(a_1,a_2) \geq b\\ \frac{d_{diff}(a_1,a_2)-b}{-0.7-b}, & \text{otherwise}\end{cases}$$

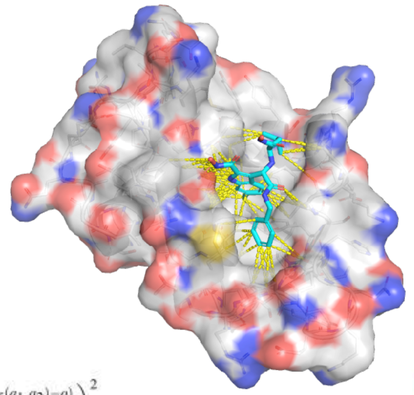
$$\text{hydrophobic}(a_1,a_2,d) = \begin{cases}0, & not\_hydrophobic(a_1)\text{ or }not\_hydrophobic(a_2)\\ 1, & d_{diff}(a_1,a_2) < 0.5\\ 0, & d_{diff}(a_1,a_2) \geq b\\ \frac{d_{diff}(a_1,a_2)-b}{0.5-b}, & \text{otherwise}\end{cases}$$

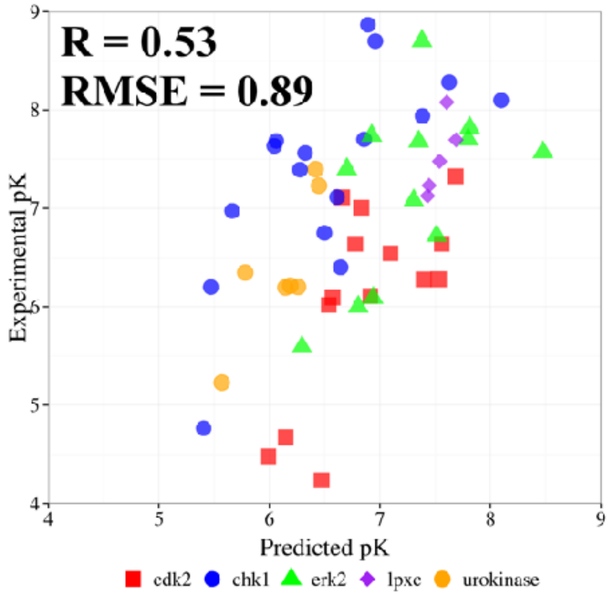
$$\text{non\_hydrophobic}(a_1,a_2,d) = \begin{cases}0, & is\_hydrophobic(a_1)\text{ or }is\_hydrophobic(a_2)\\ 1, & d_{diff}(a_1,a_2) < 0.5\\ 0, & d_{diff}(a_1,a_2) \geq 1.5\\ 1.5 - d_{diff}(a_1,a_2), & \text{otherwise}\end{cases}$$

$$\text{ad4\_solvation}(a_1,a_2,d) = [(solv(a_1)+q\cdot partial\_charge(a_1))\,volume(a_2)+ (solv(a_2)+q\cdot partial\_charge(a_2))\,volume(a_1)]\,e^{-\left(\frac{d}{7.2}\right)^2}$$

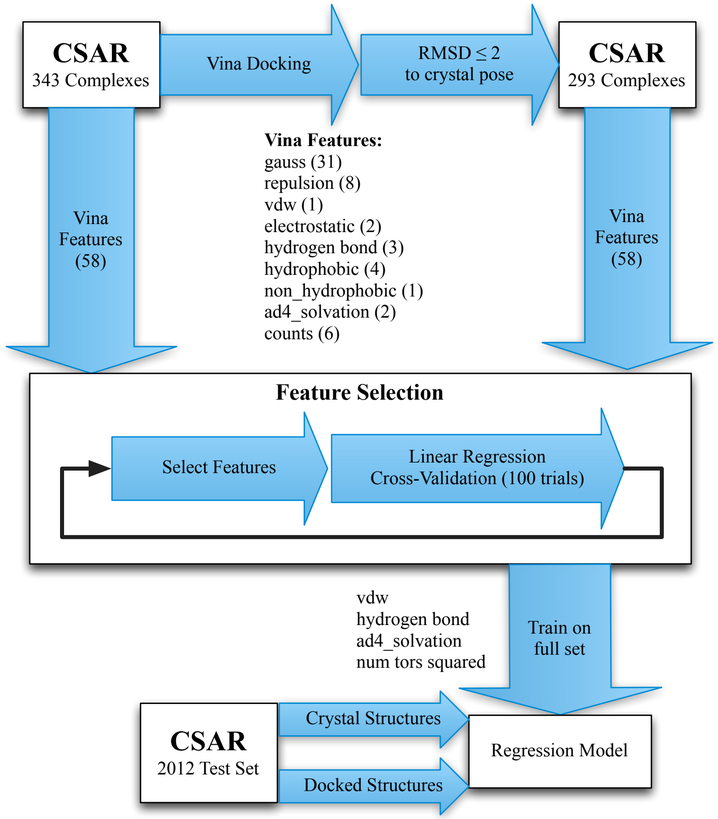
$$\text{counts}: \quad \#heavy\_atoms,\ ligand\_length,\ \#hydrophobic\_atoms,\ \#torsions,\ \#torsions^2,\ \sqrt{\#torsions}$$

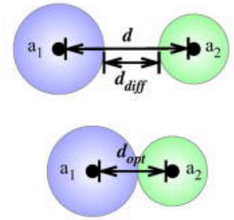
# Scoring






smina.sf.net




$$\text{gauss}(a_1,a_2,d) = e^{\left(\frac{-(d_{diff}(a_1,a_2)-o)}{w}\right)^2}$$

$$\text{repulsion}(a_1,a_2,d) = \begin{cases} (d_{diff}(a_1,a_2)-o)^2, & d_{diff}(a_1,a_2) < o \\ 0, & \text{otherwise} \end{cases}$$

$$\text{vdw}(a_1,a_2,d) = \left(\frac{d_{opt}(a_1,a_2)}{d}\right)^8 - 2\left(\frac{d_{opt}(a_1,a_2)}{d}\right)^4$$

$$\text{electrostatic}(a_1,a_2,d) = \frac{partial\_charge(a_1)\cdot partial\_charge(a_2)}{d^x}$$

$$\text{hydrogen\_bond}(a_1,a_2,d) = \begin{cases} 0, & (a_1,a_2) \text{ do not form hydrogen bond} \\ 1, & d_{diff}(a_1,a_2) < -0.7 \\ 0, & d_{diff}(a_1,a_2) \geq b \\ \frac{d_{diff}(a_1,a_2)-b}{-0.7-b}, & \text{otherwise} \end{cases}$$

$$\text{hydrophobic}(a_1,a_2,d) = \begin{cases} 0, & not\_hydrophobic(a_1) \text{ or } not\_hydrophobic(a_2) \\ 1, & d_{diff}(a_1,a_2) < 0.5 \\ 0, & d_{diff}(a_1,a_2) \geq b \\ \frac{d_{diff}(a_1,a_2)-b}{0.5-b}, & \text{otherwise} \end{cases}$$

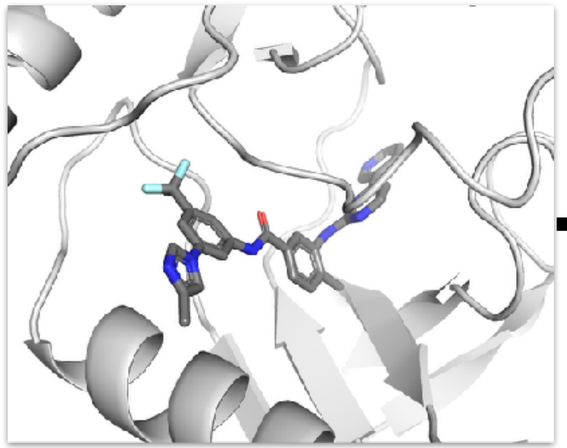
$$\text{non\_hydrophobic}(a_1,a_2,d) = \begin{cases} 0, & is\_hydrophobic(a_1) \text{ or } is\_hydrophobic(a_2) \\ 1, & d_{diff}(a_1,a_2) < 0.5 \\ 0, & d_{diff}(a_1,a_2) \geq 1.5 \\ 1.5 - d_{diff}(a_1,a_2), & \text{otherwise} \end{cases}$$

$$\text{ad4\_solvation}(a_1,a_2,d) = [(solv(a_1)+q\cdot partial\_charge(a_1))\,volume(a_2)+ (solv(a_2)+q\cdot partial\_charge(a_2))\,volume(a_1)]\,e^{-\left(\frac{d}{7.2}\right)^2}$$

$$\text{counts}: \quad \#heavy\_atoms,\ ligand\_length,\ \#hydrophobic\_atoms,\ \#torsions,\ \#torsions^2,\ \sqrt{\#torsions}$$

Koes DR, Baumgartner MP, Camacho CJ. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. J Chem Inf Model. 2013 Aug 26;53(8):1893-904.

# Protein-Ligand Scoring



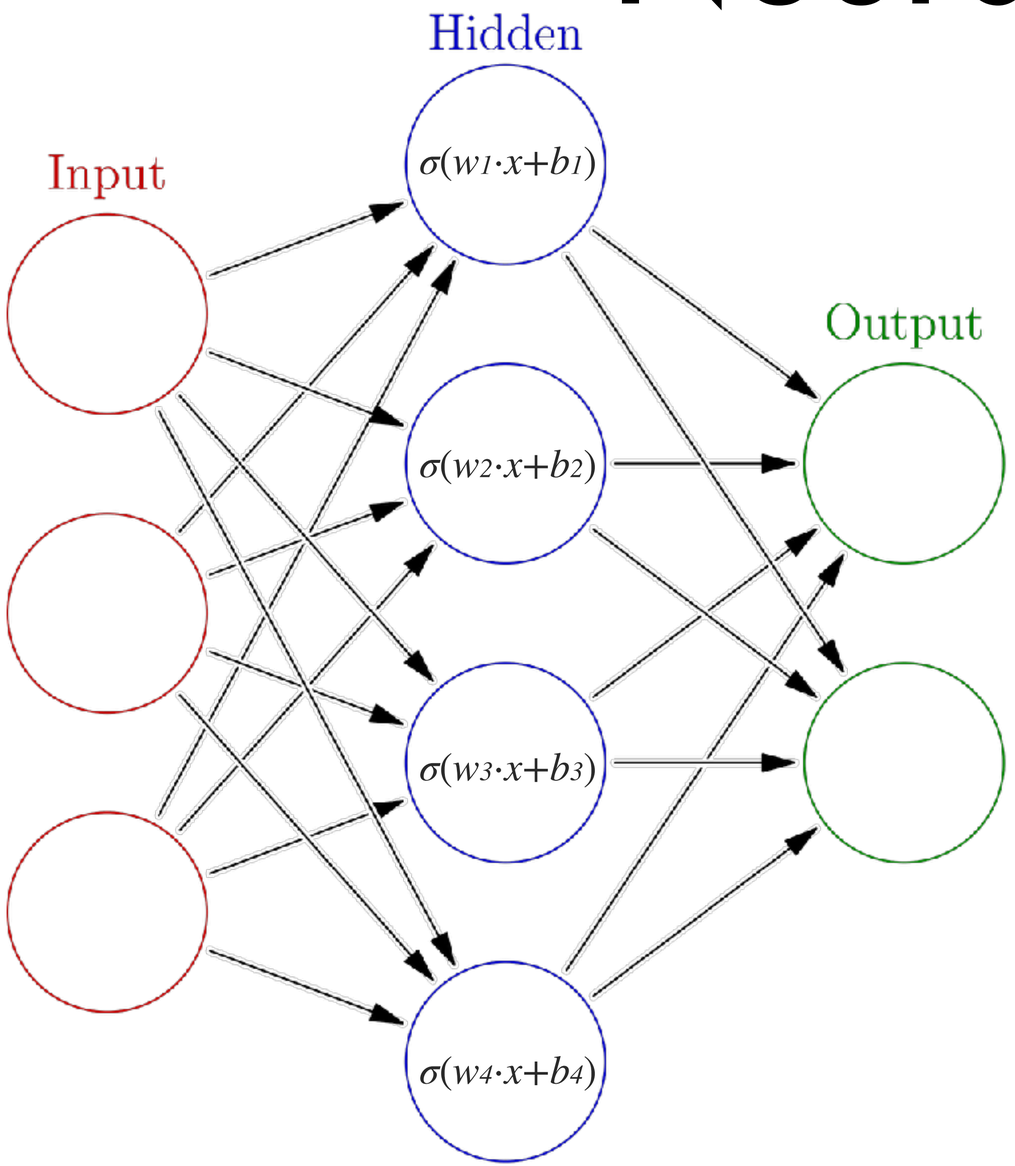
**Model**

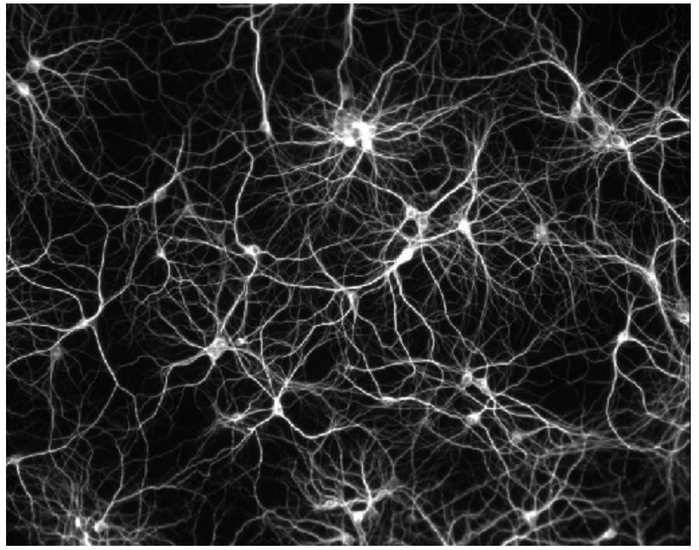
Pose Prediction

Binding Discrimination

Affinity Prediction

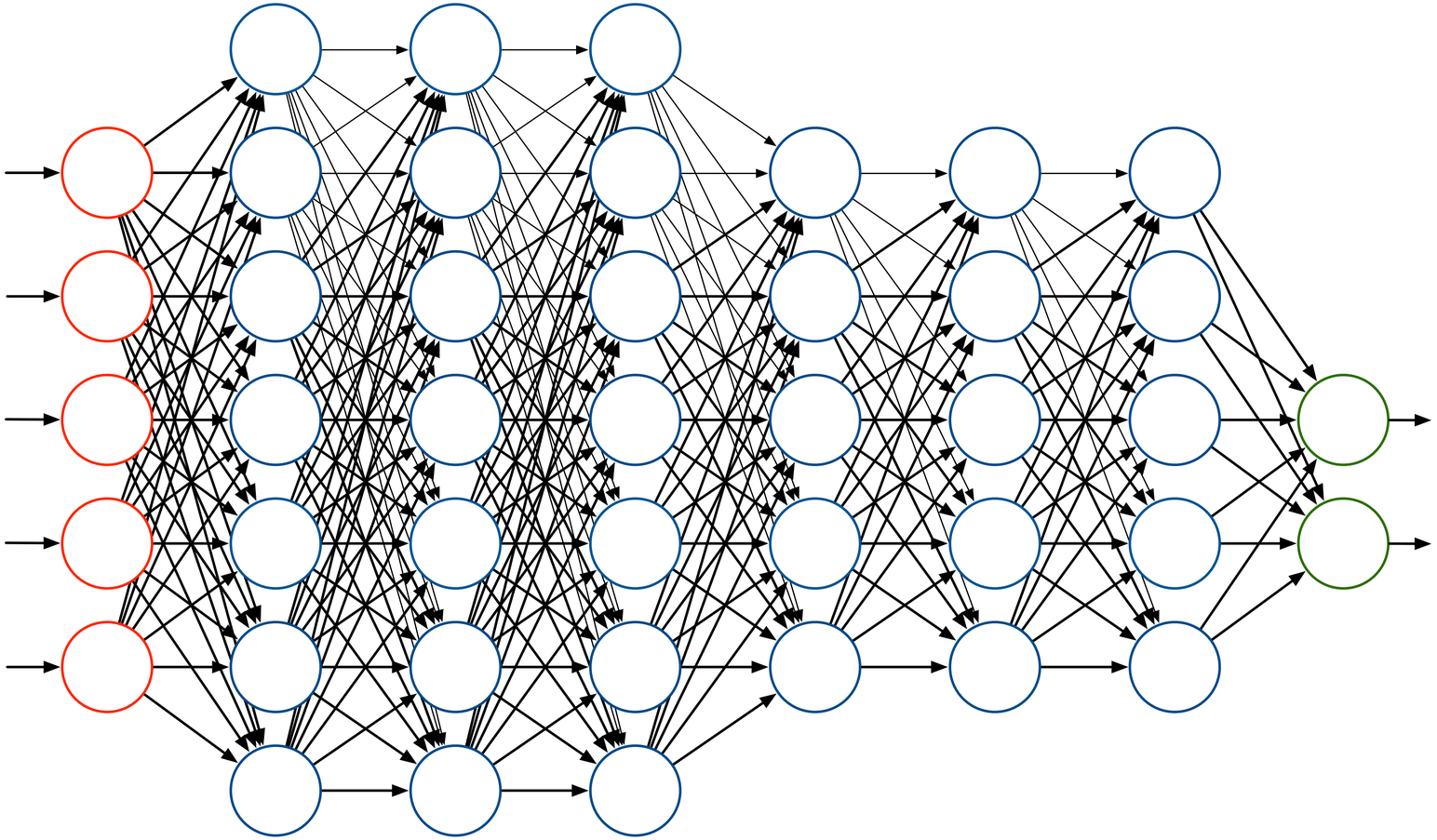
# Neural Networks



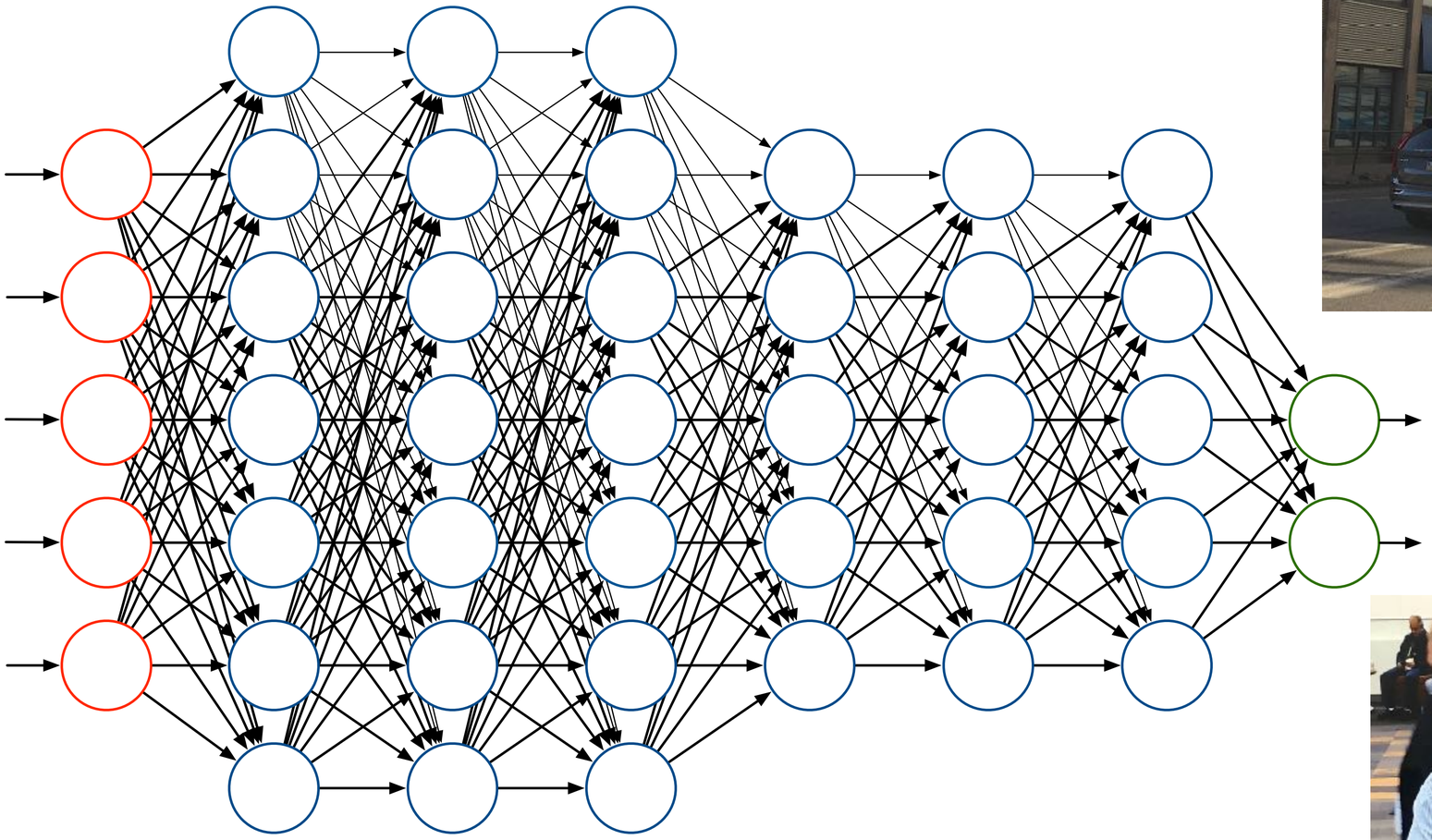



The **universal approximation theorem** states that, under reasonable assumptions, a feedforward **neural network** with a finite number of nodes **can approximate any continuous** function to within a given error over a bounded input domain.

# Deep Learning



# Deep Learning

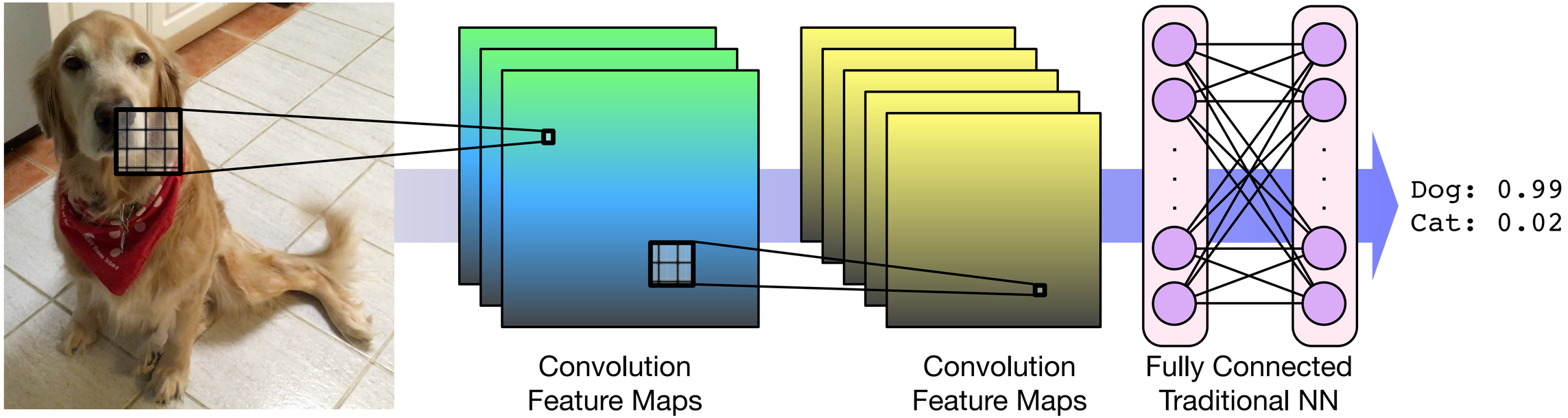


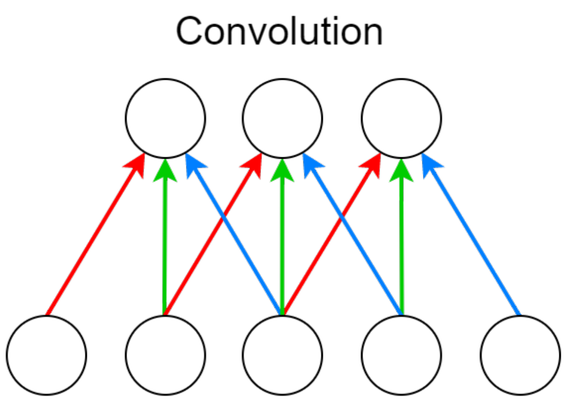

# Deep Learning

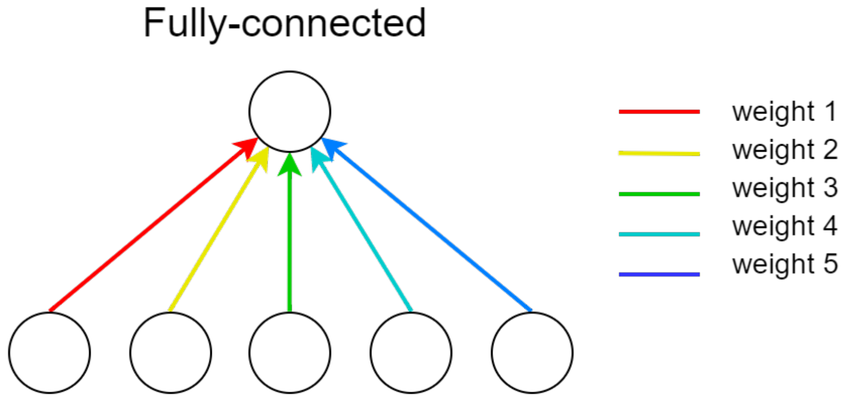
$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

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

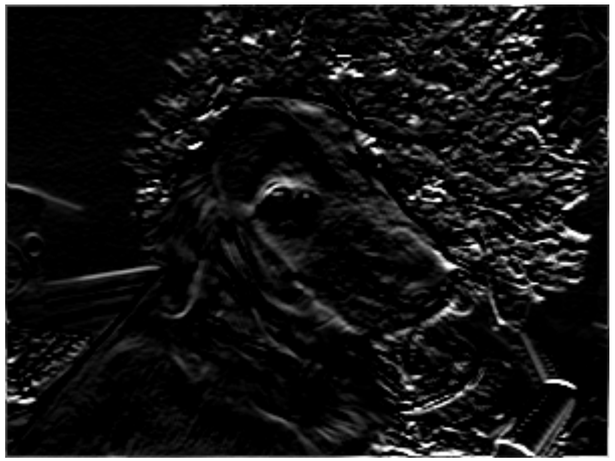

# Convolutional Neural Networks



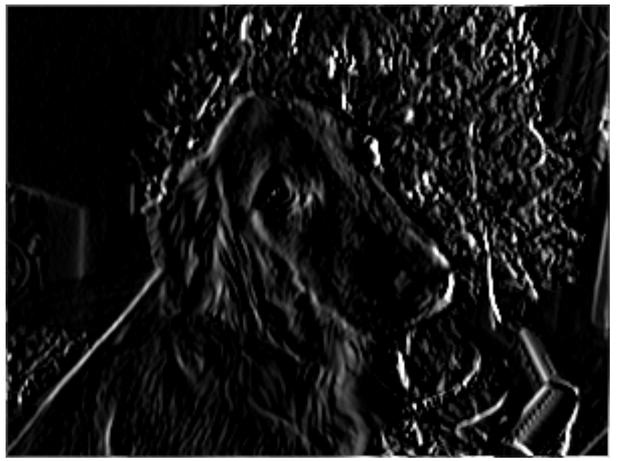



# Convolutional Filters



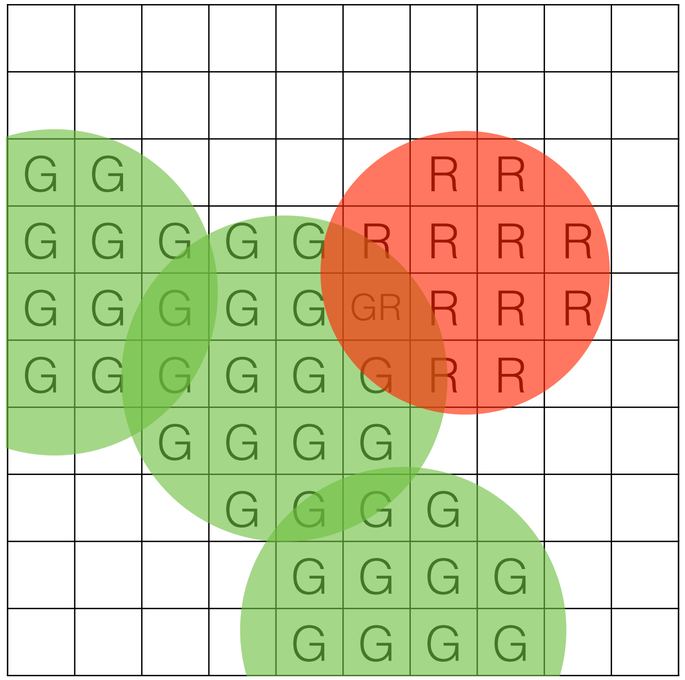
| | | |
|---|---|---|
| -1 | -1 | -1 |
| 0 | 0 | 0 |
| 1 | 1 | 1 |



| | | |
|---|---|---|
| -1 | 0 | 1 |
| -1 | 0 | 1 |
| -1 | 0 | 1 |

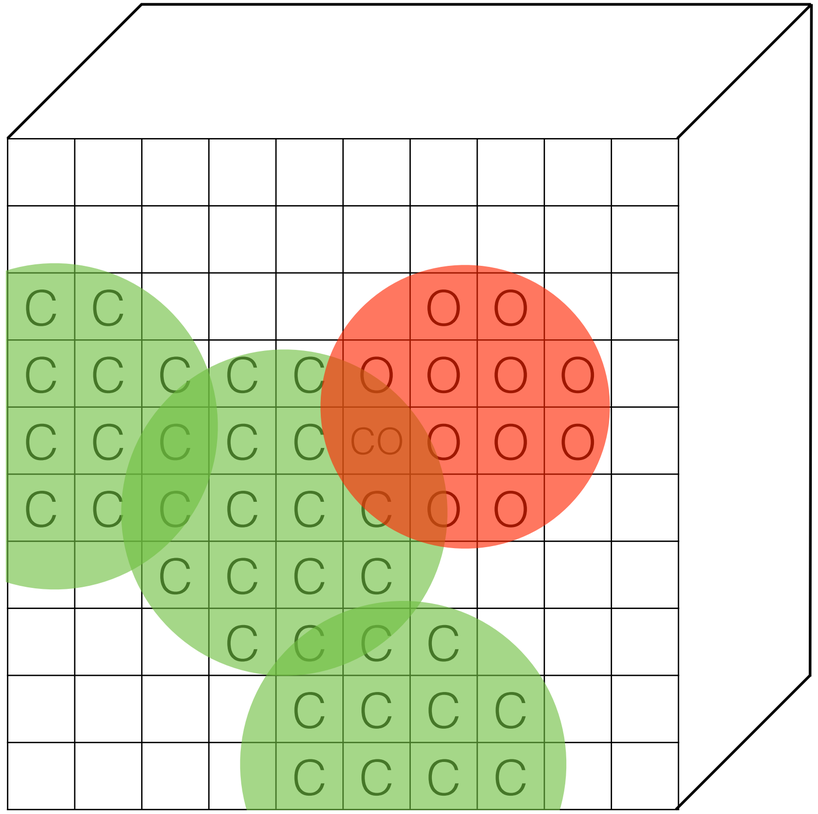
| | | |
|---|---|---|
| -1 | -1 | -1 |
| -1 | 8 | -1 |
| -1 | -1 | -1 |

# Protein-Ligand Representation




(R,G,B) pixel

# 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**.

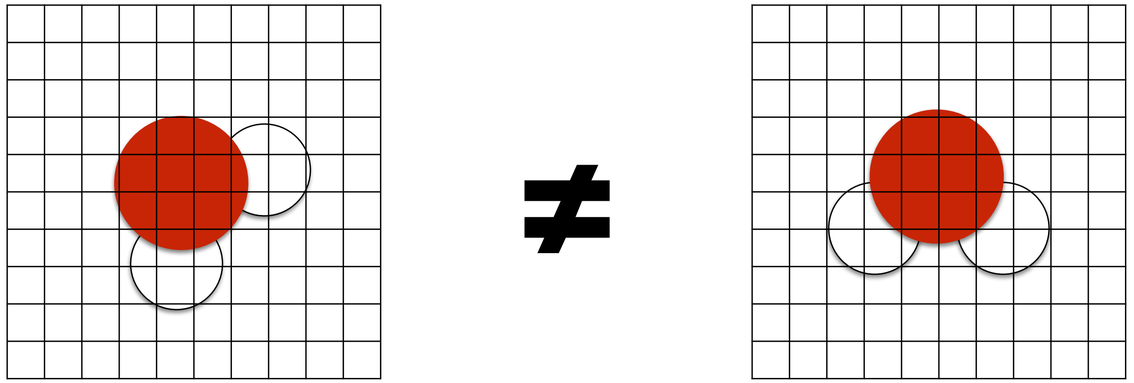
# Why Grids?

**Cons**

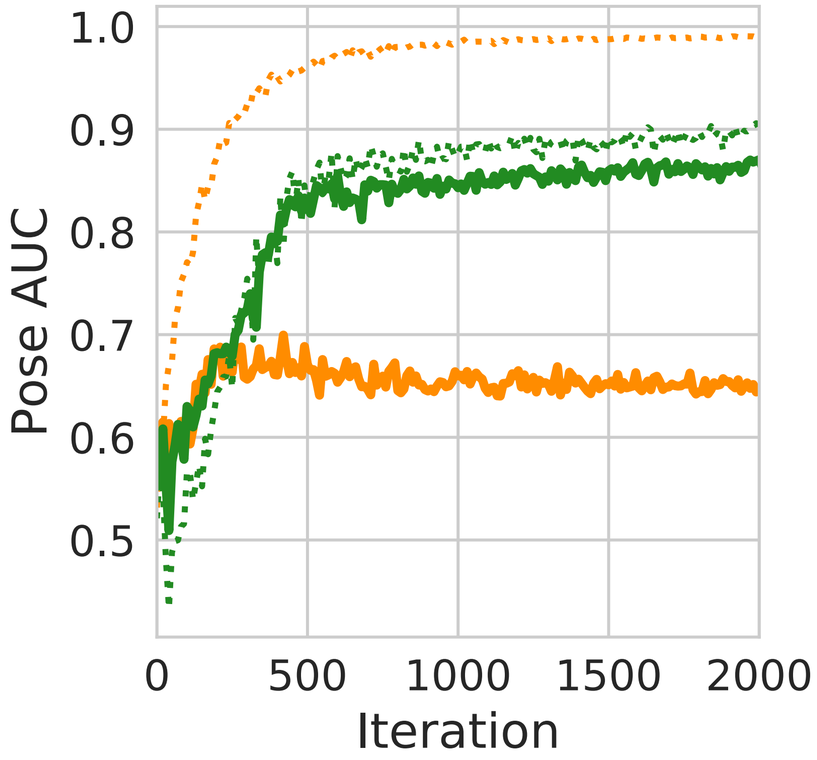
- *coordinate frame dependent*
- pairwise interactions not explicit

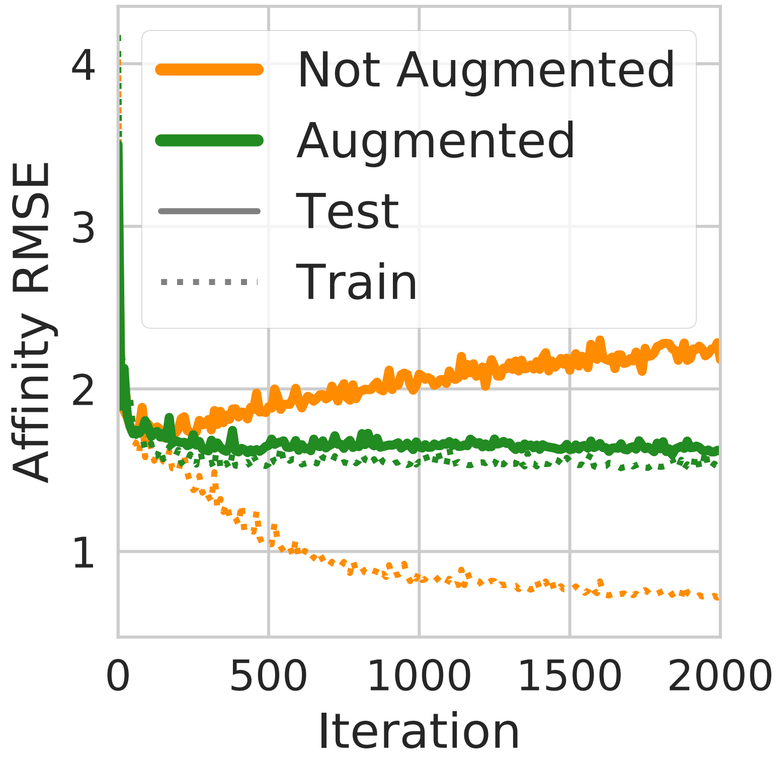
**Pros**

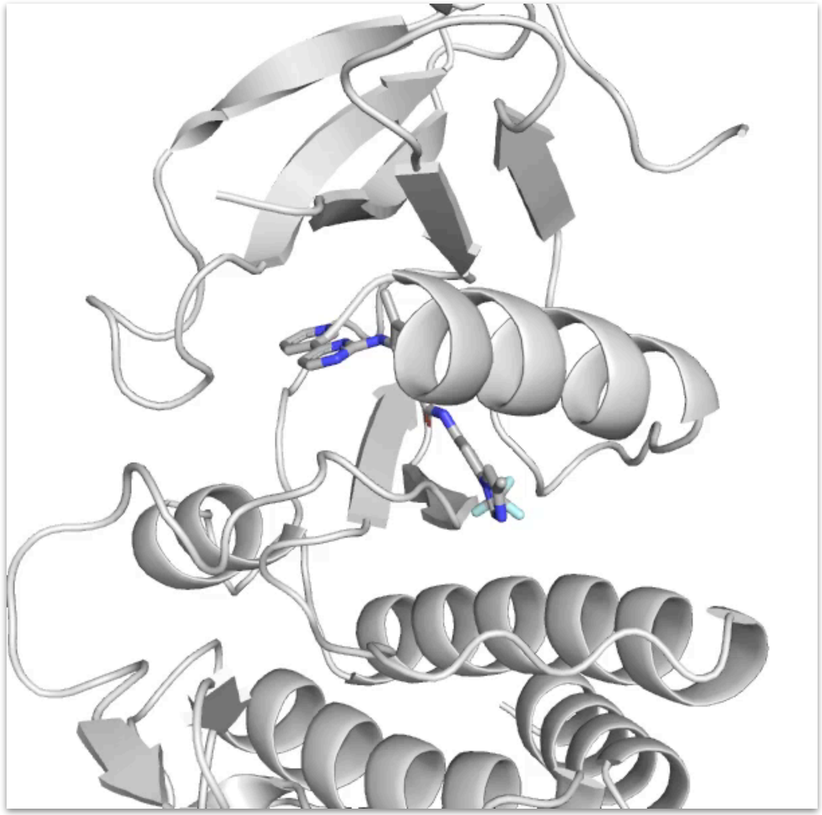
- clear spatial relationships
- amazingly parallel
- easy to interpret



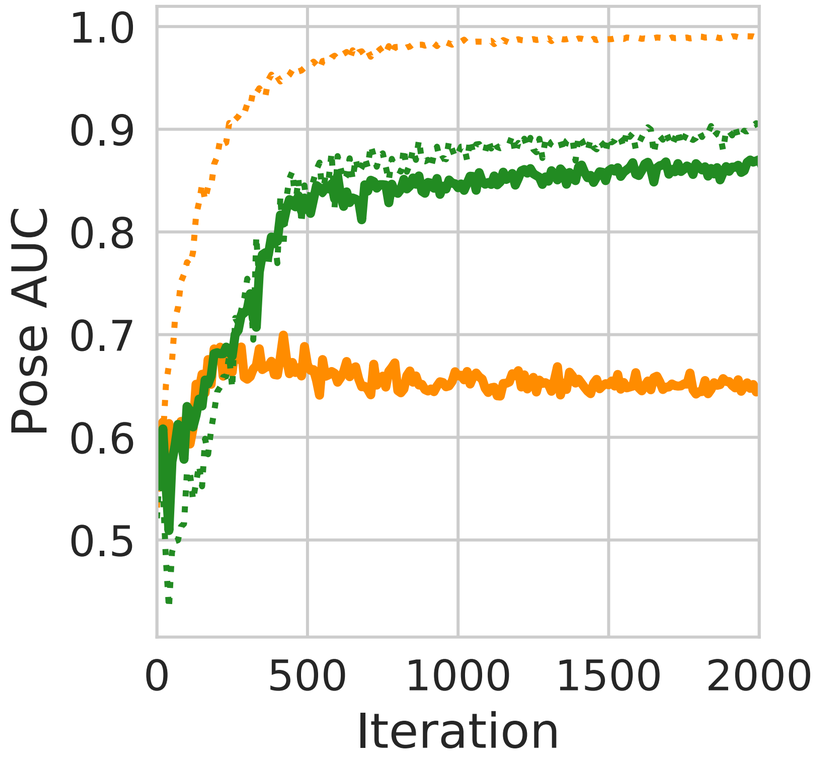
# Data Augmentation

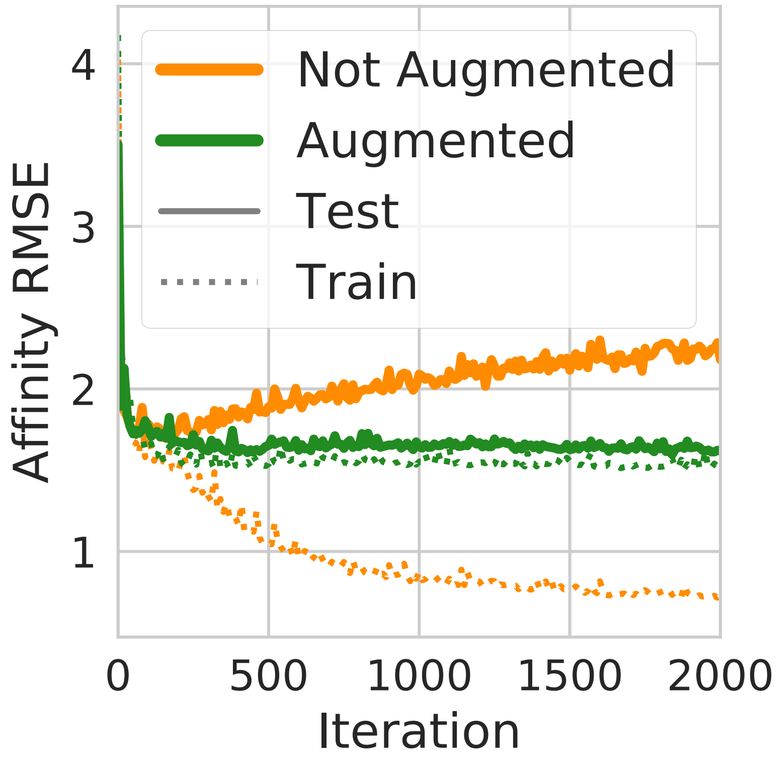


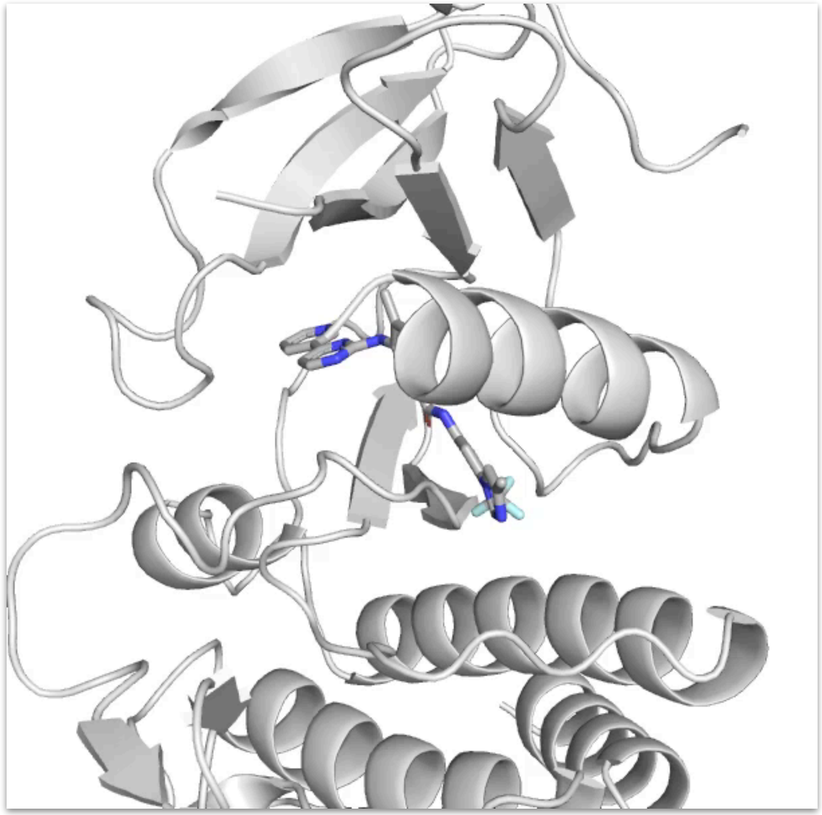



# Data Augmentation



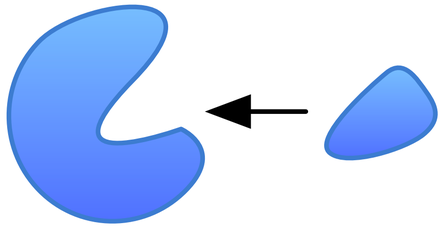



# Training

PDBbind 2016 refined set

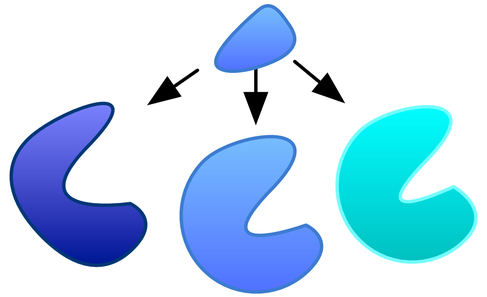
- 4056 protein-ligand complexes
- diverse targets
- wide range of affinities
- generate poses with AutoDock Vina
- include minimized crystal pose



**Redocked Training Set**

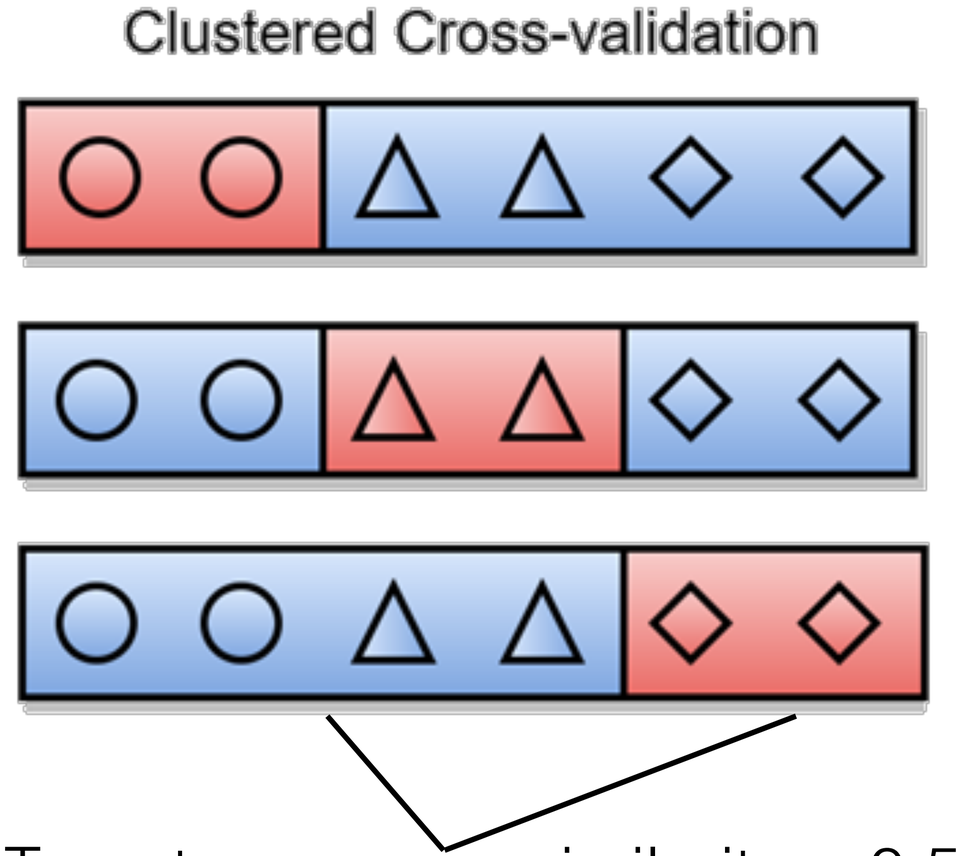
Pocketome

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



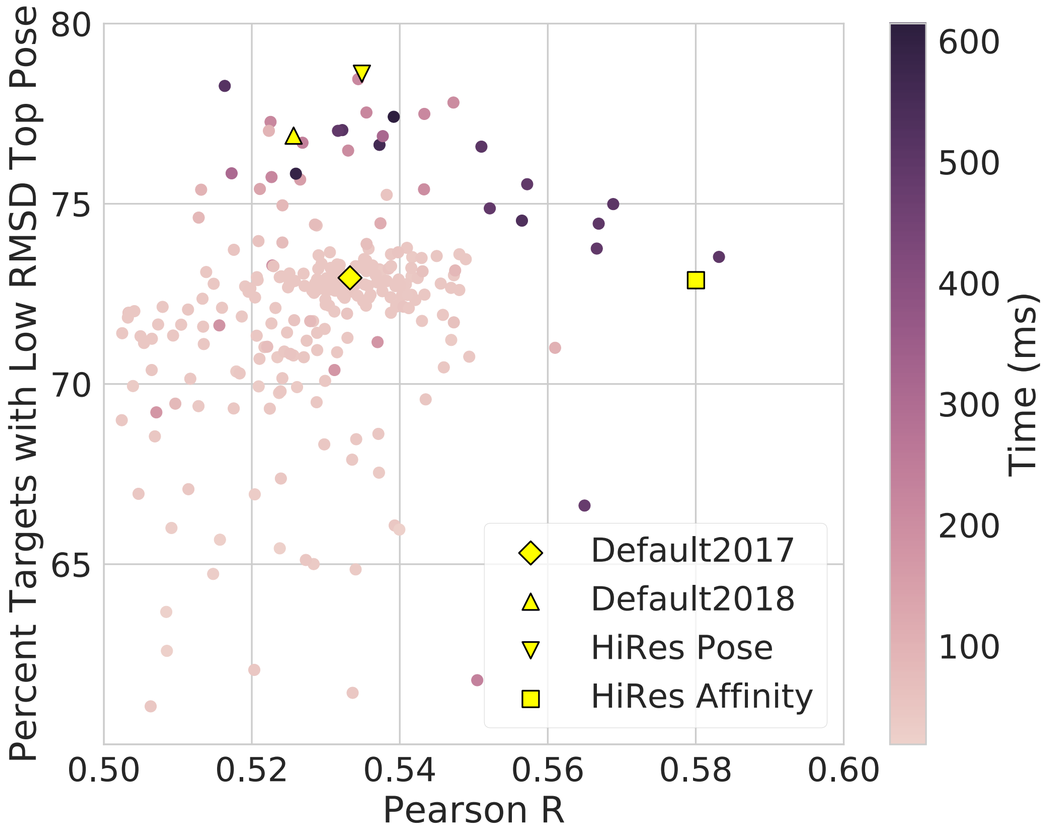
**Crossdocked Training Set**

# Training

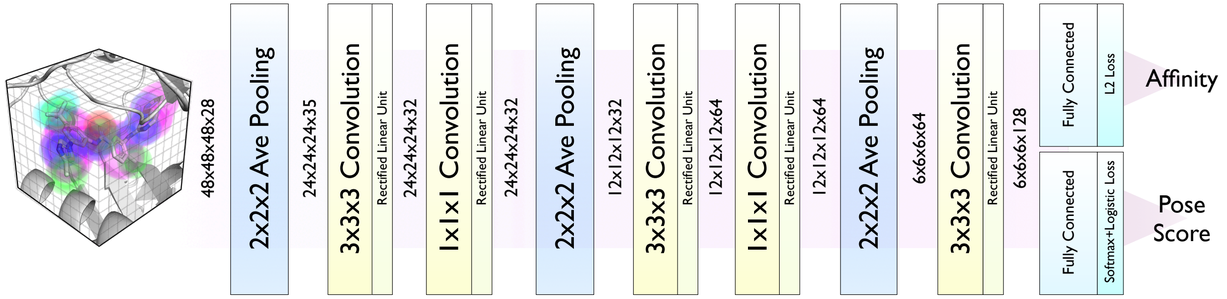



Target sequence similarity < 0.5
AND
Ligand similarity < 0.9

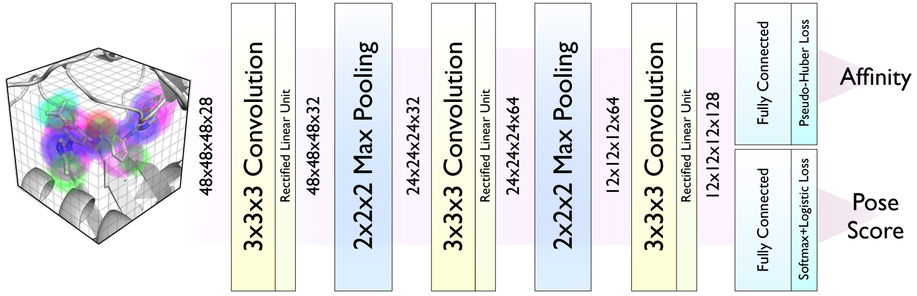
# Optimized Models



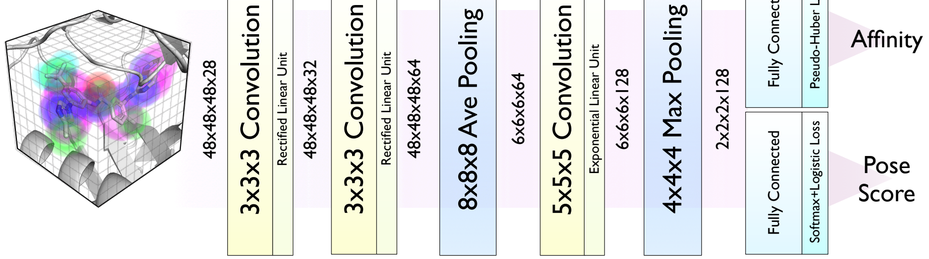

## Default2018



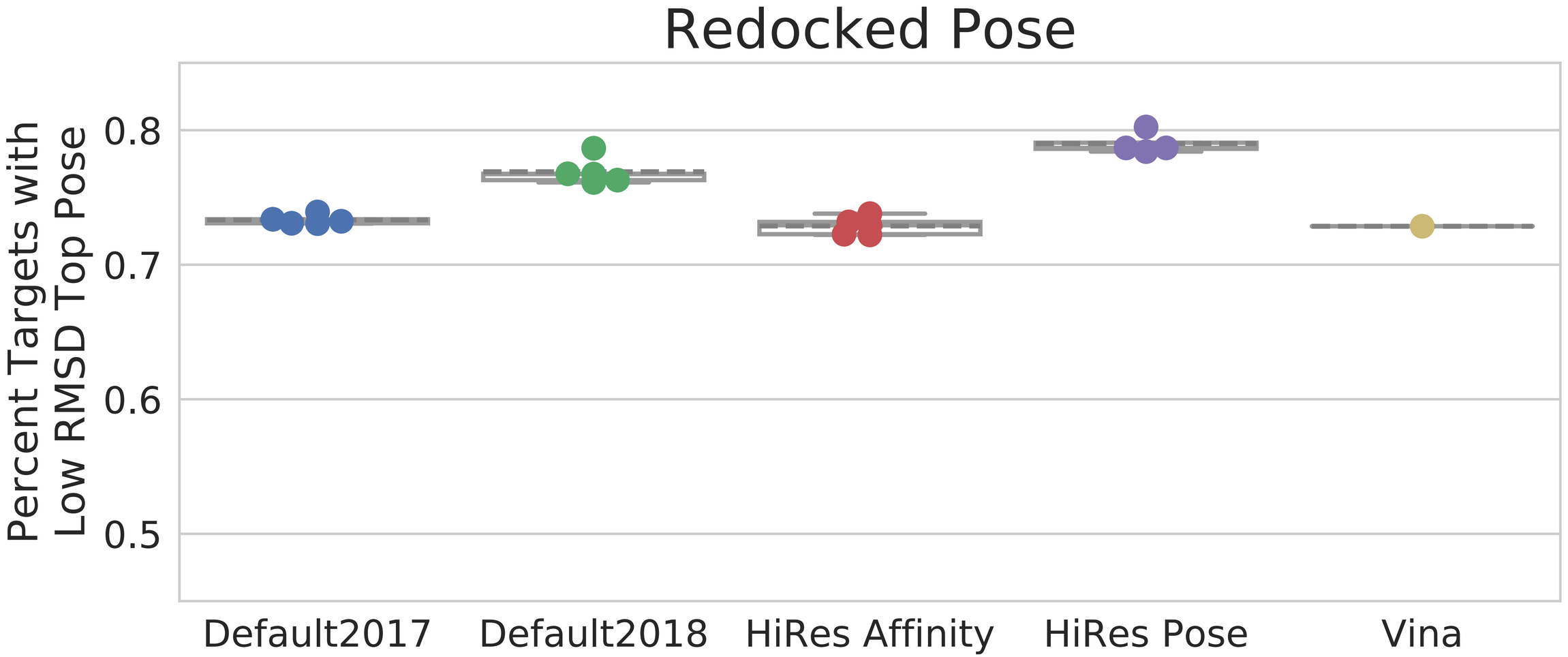

## HiRes Pose



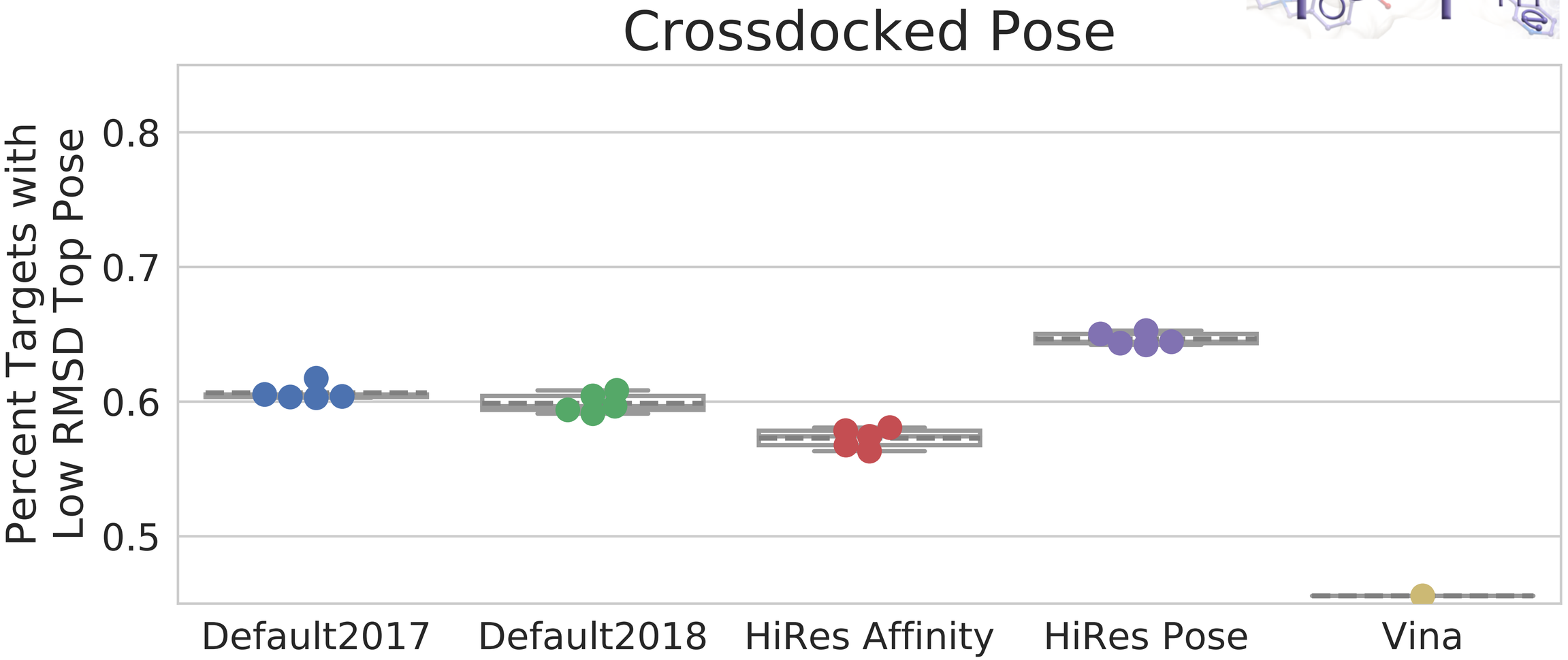

## HiRes Affinity



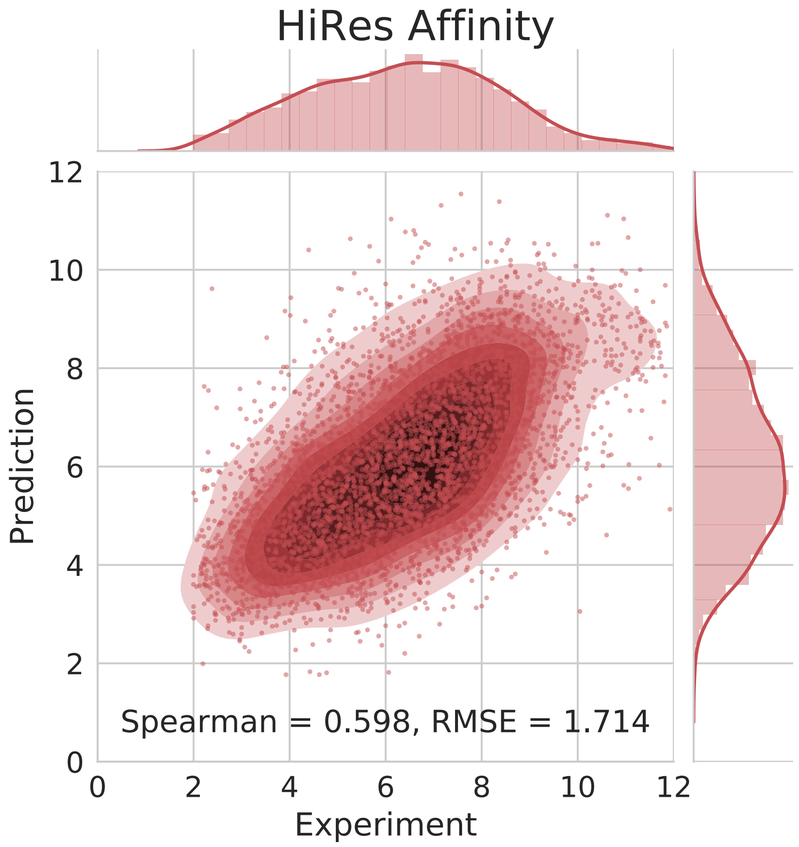
# Pose Results

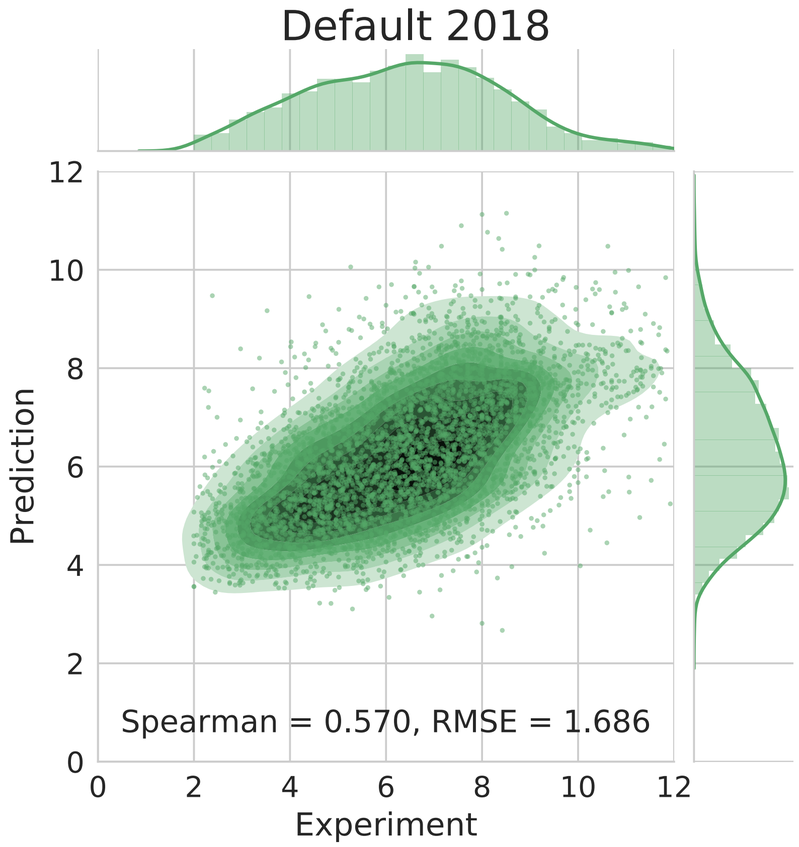


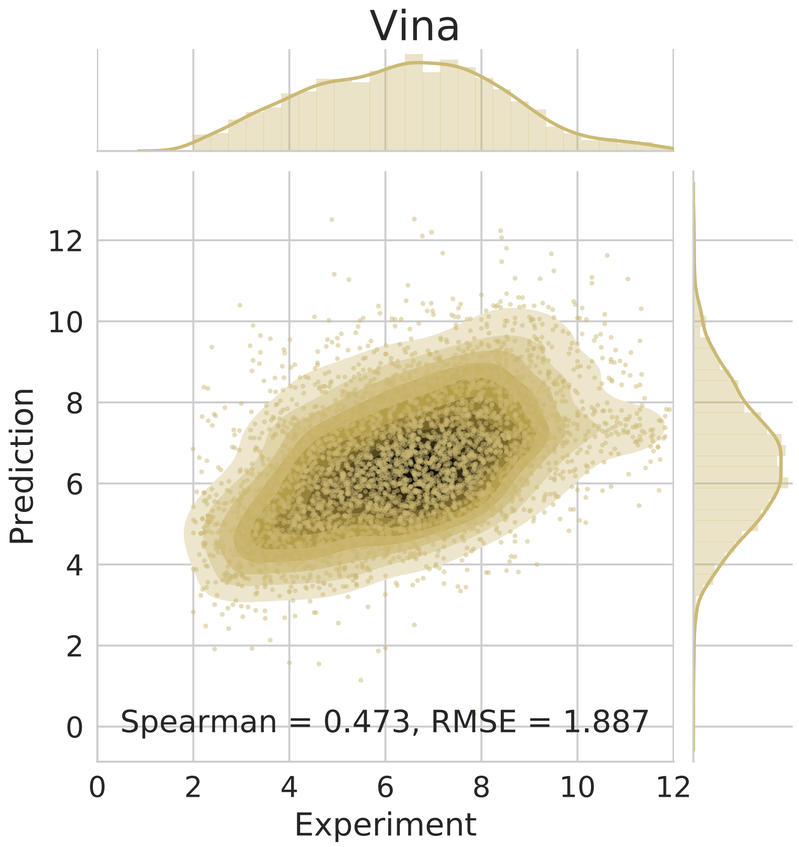
# Pose Results



# Affinity Results

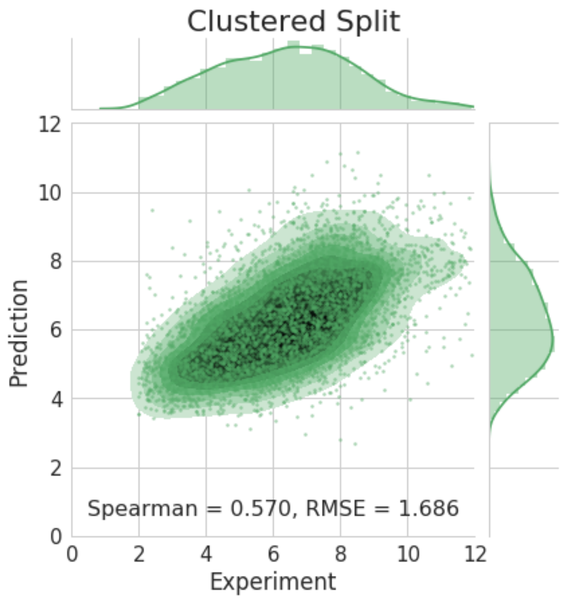


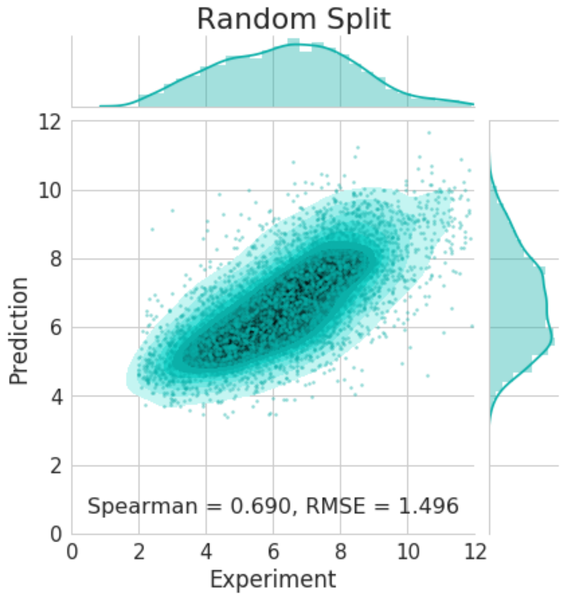



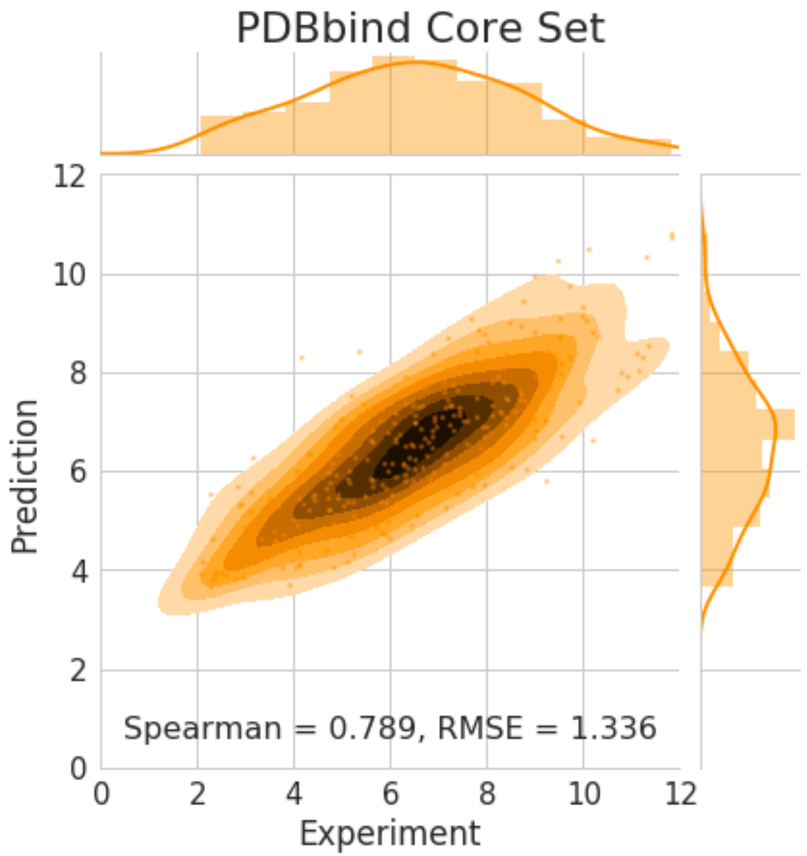



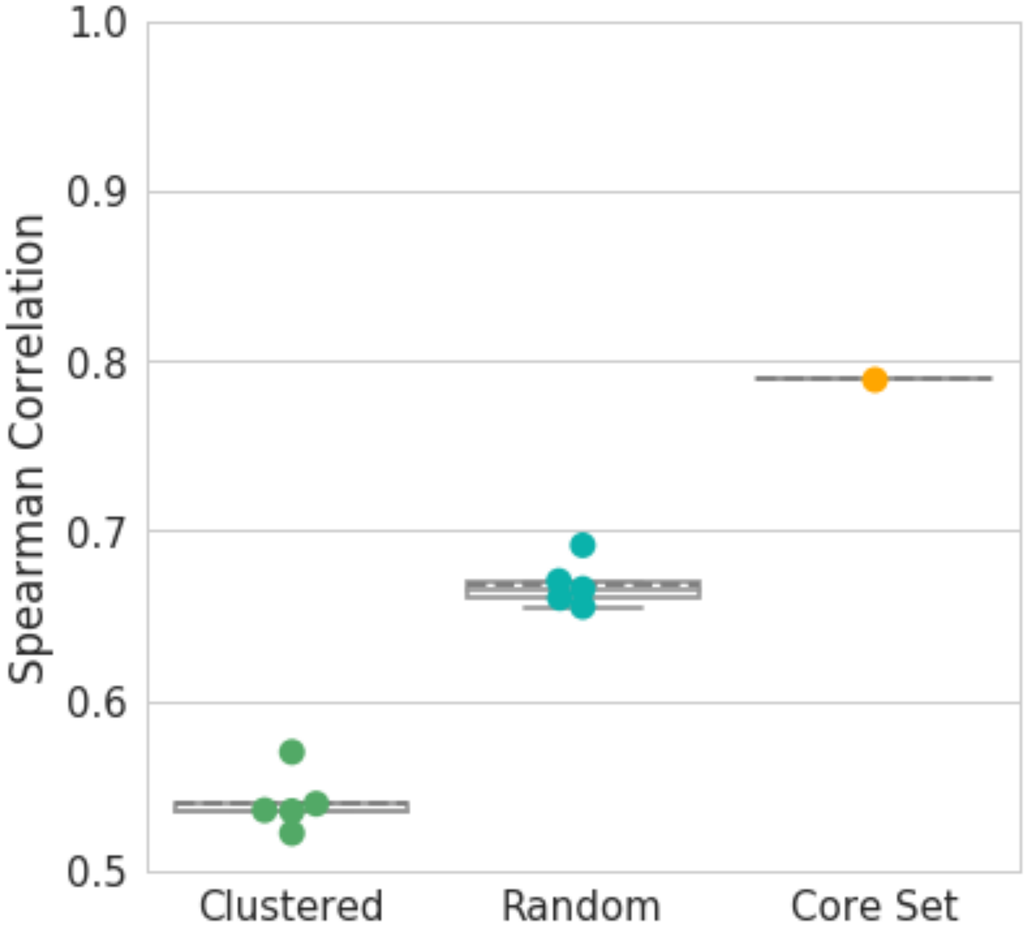

Clustered Cross Validation

# Affinity Results

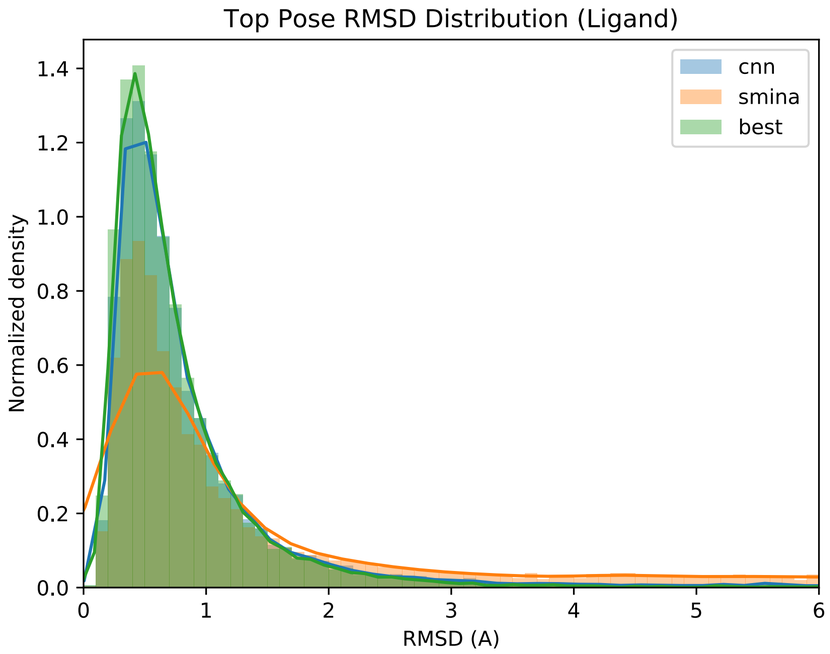


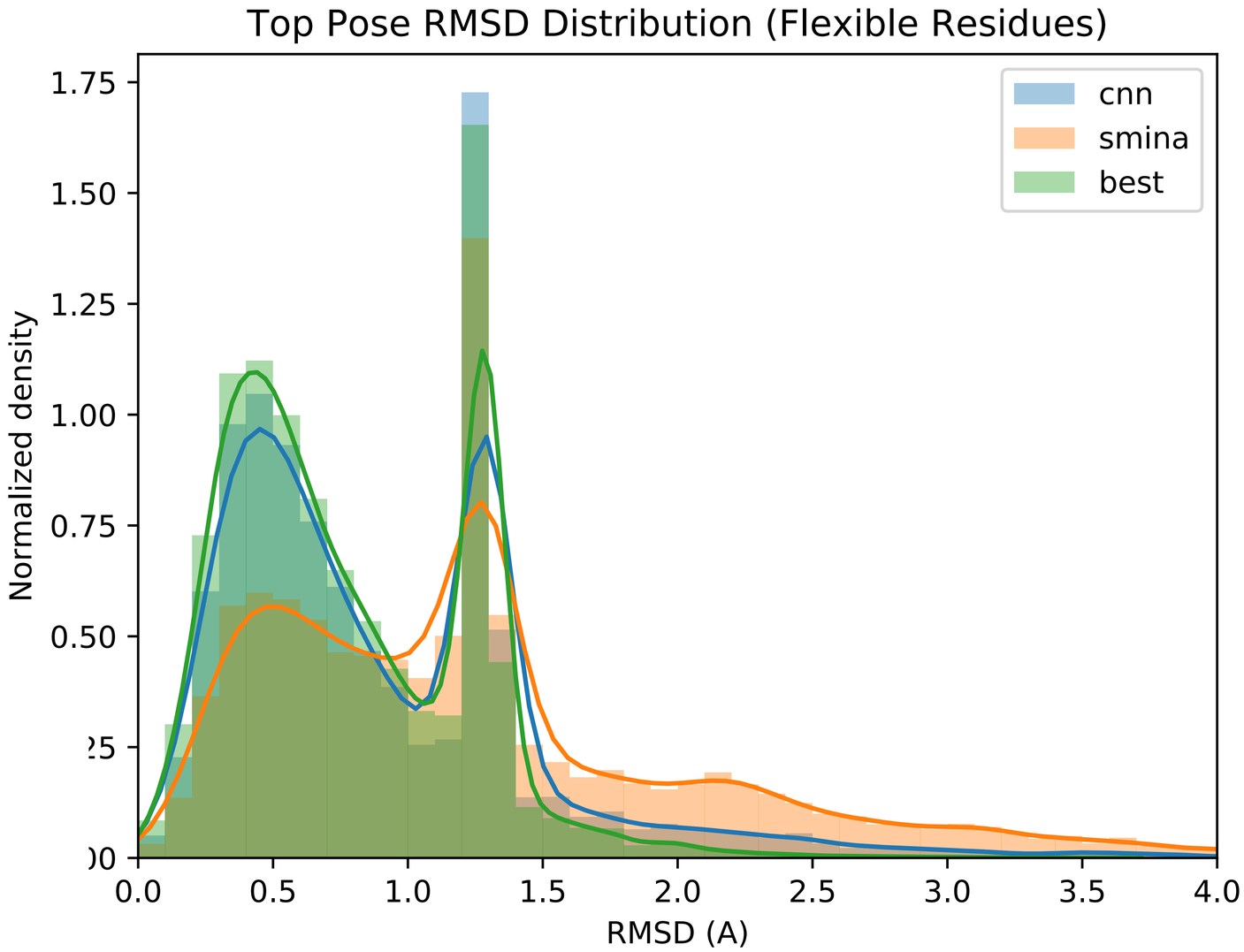



# Flexible Docking Scoring



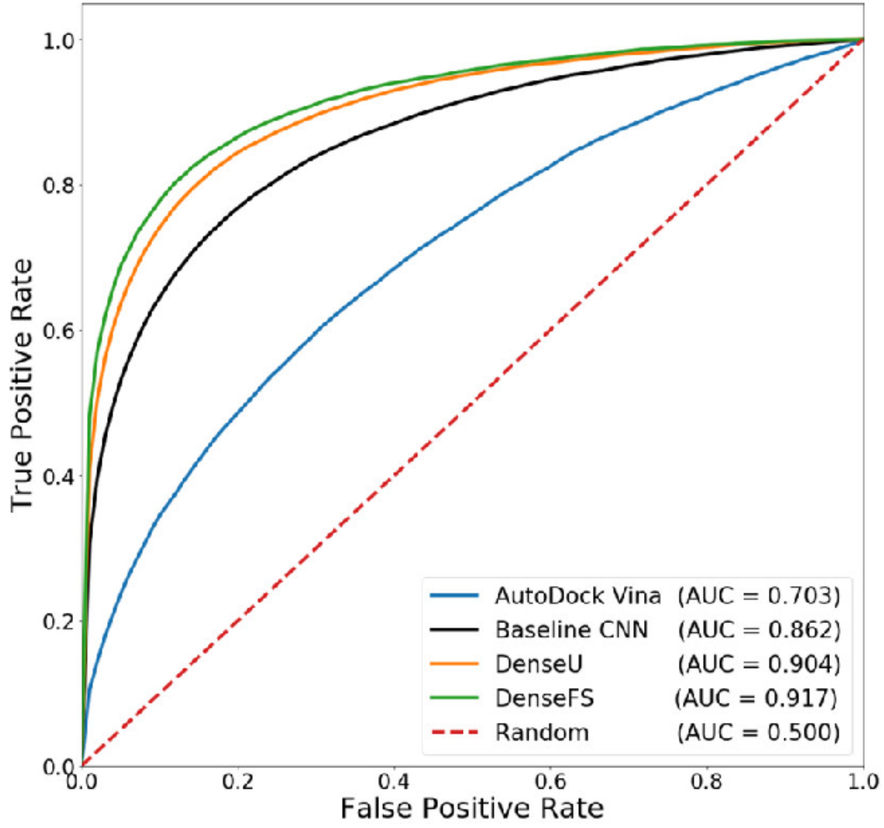


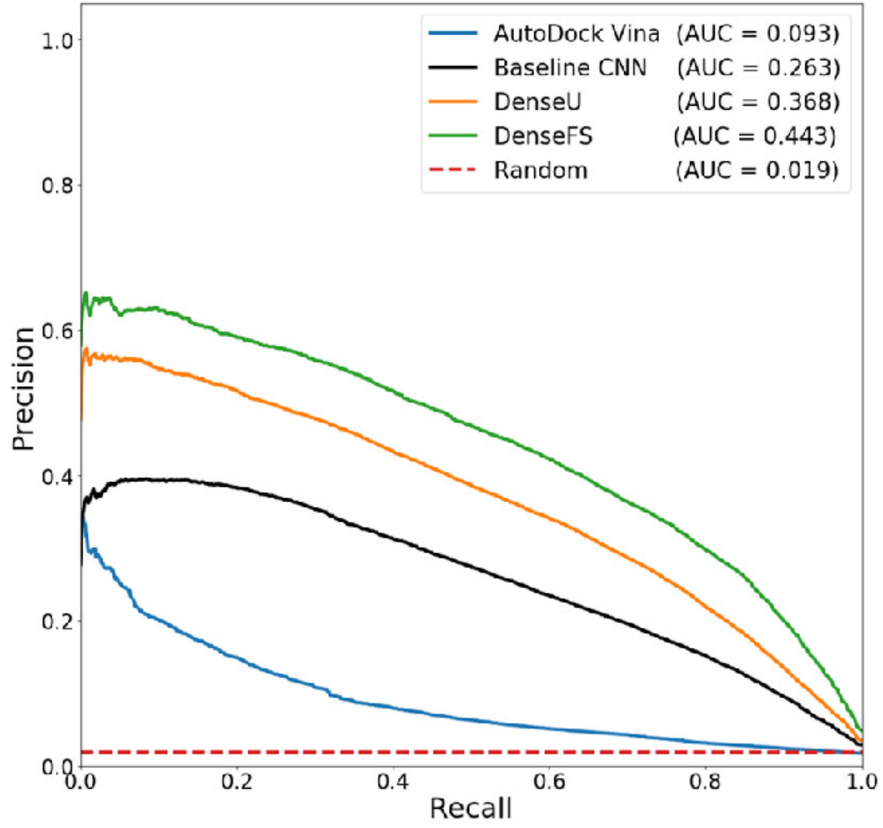
Google Summer of Code

# Virtual Screening

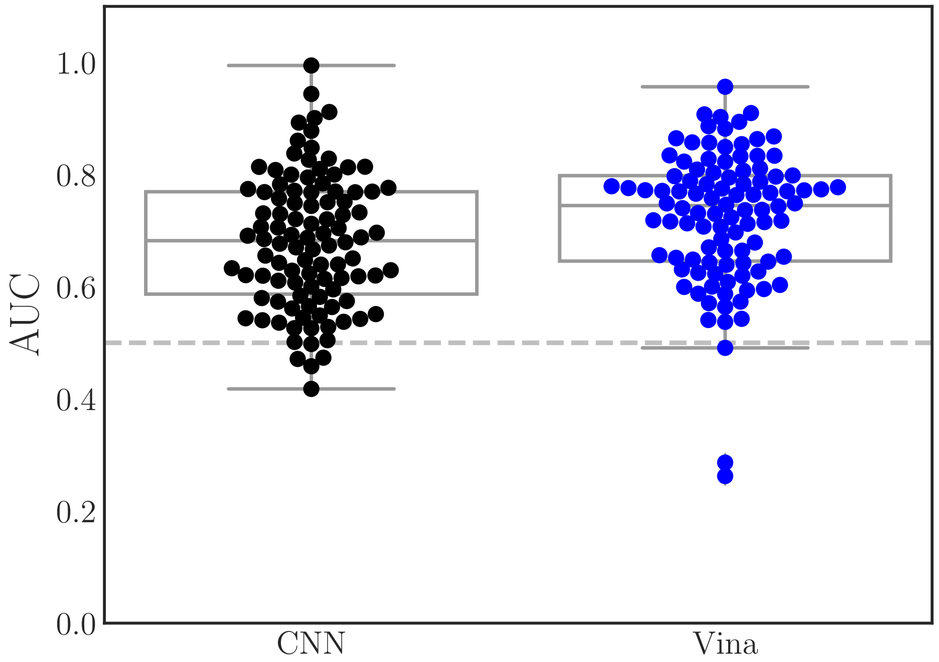
**Protein Family-Specific Models Using Deep Neural Networks and Transfer Learning Improve Virtual Screening and Highlight the Need for More Data**

Fergus Imrie[†], Anthony R. Bradley[‡¶§], Mihaela van der Schaar[†⊥], and Charlotte M. Deane*[†]
[†] Oxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford OX1 3LB, U.K.
[‡] Structural Genomics Consortium, University of Oxford, Oxford OX3 7DQ, U.K.
[¶] Department of Chemistry, University of Oxford, Oxford OX1 3TA, U.K.
[§] Diamond Light Source Ltd., Didcot OX11 0DE, U.K.
[⊥] Department of Engineering, University of Oxford, Oxford OX1 3PJ, U.K.
– Alan Turing Institute, London NW1 2DB, U.K.

# Virtual Screening



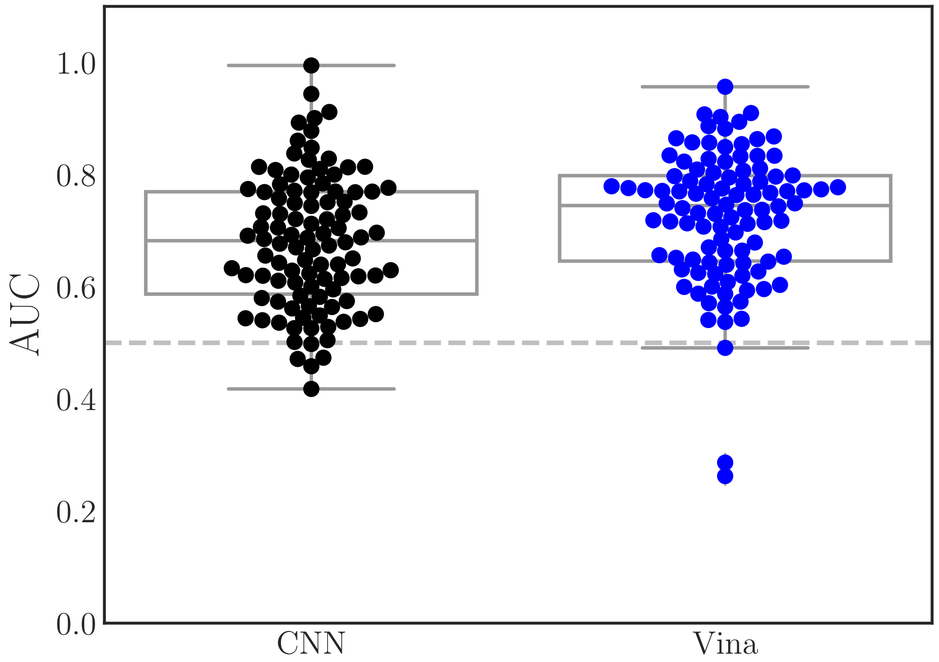

**In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening**

Jochen Sieg, Florian Flachsenberg, and Matthias Rarey*
Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design, Bundesstraße 43, 20146 Hamburg, Germany

**Hidden Bias in the DUD-E Dataset Leads to Misleading Performance of Deep Learning in Structure-Based Virtual Screening**

Preprint submitted on 24.03.2019, 15:39 and posted on 25.03.2019, 12:58 by Lieyang Chen, Anthony Cruz, Steven Ramsey, Callum J. Dickson, José S. Duca, Viktor Hornak, David R. Koes, Tom Kurtzman

# Virtual Screening




**In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening**

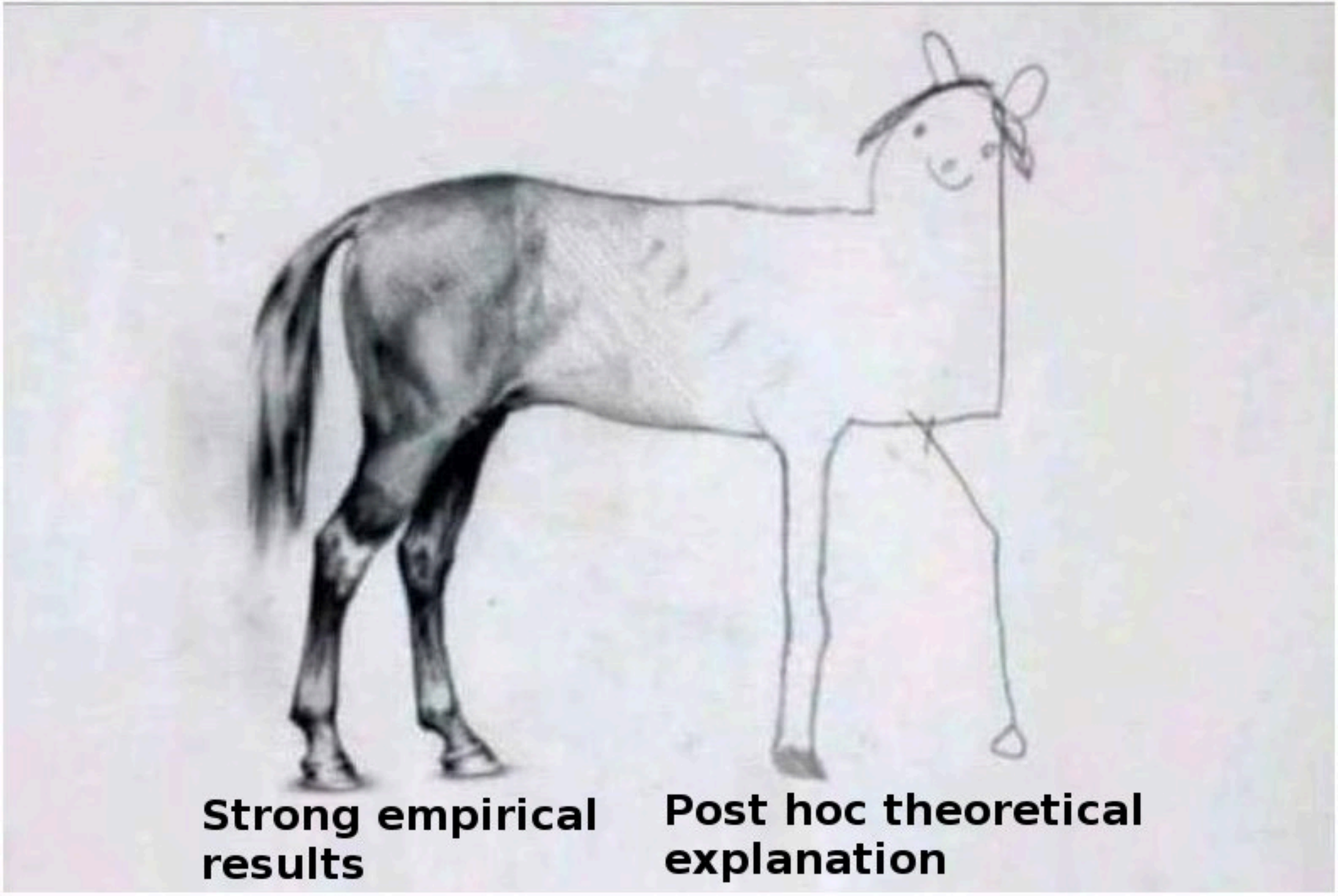
Jochen Sieg, Florian Flachsenberg, and Matthias Rarey*
Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design, Bundesstraße 43, 20146 Hamburg, Germany

**Hidden Bias in the DUD-E Dataset Leads to Misleading Performance of Deep Learning in Structure-Based Virtual Screening**

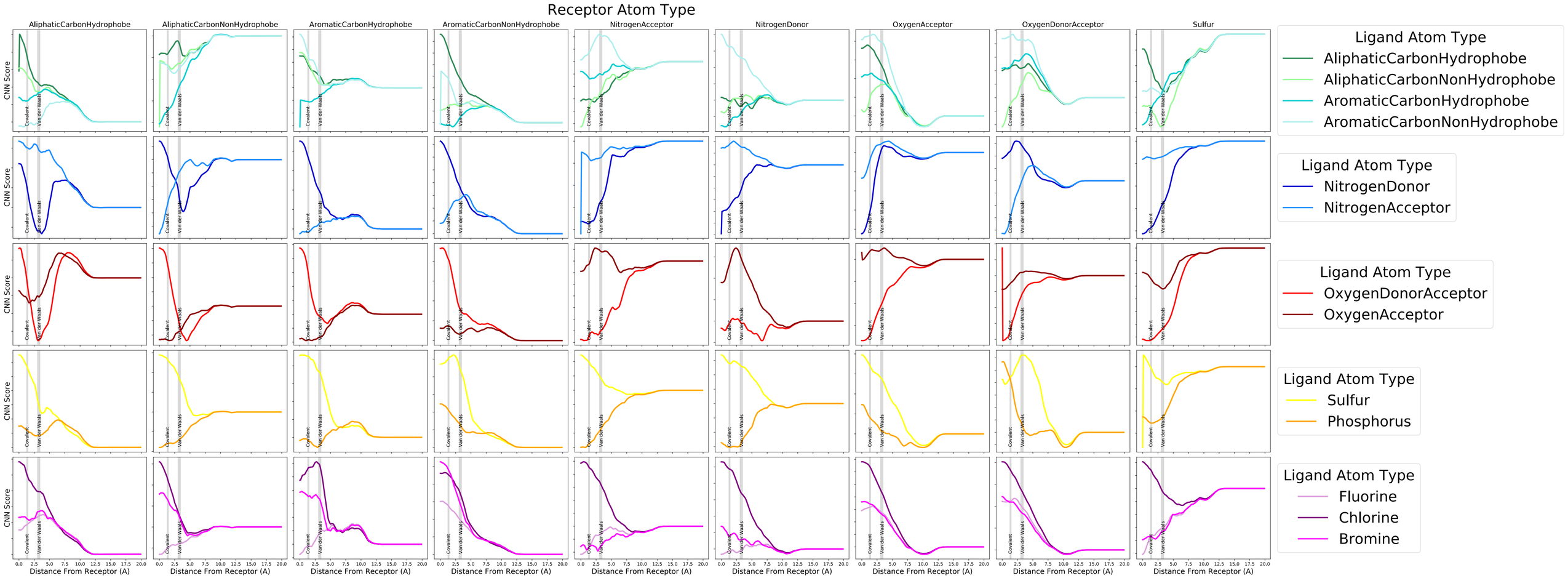
Preprint submitted on 24.03.2019, 15:39 and posted on 25.03.2019, 12:58 by Lieyang Chen, Anthony Cruz, Steven Ramsey, Callum J. Dickson, José S. Duca, Viktor Hornak, David R. Koes, Tom Kurtzman

# Visualization

# Anatomy of a deep learning paper

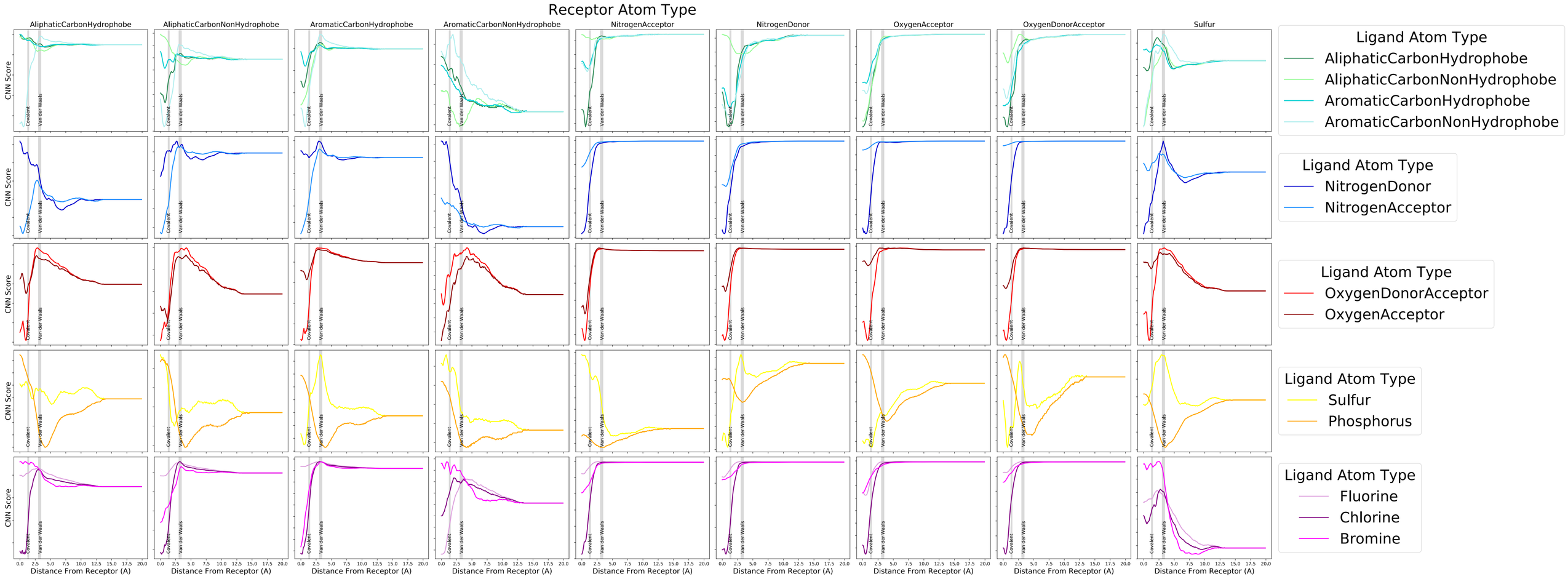


# Visualizing with Atomistic Probes



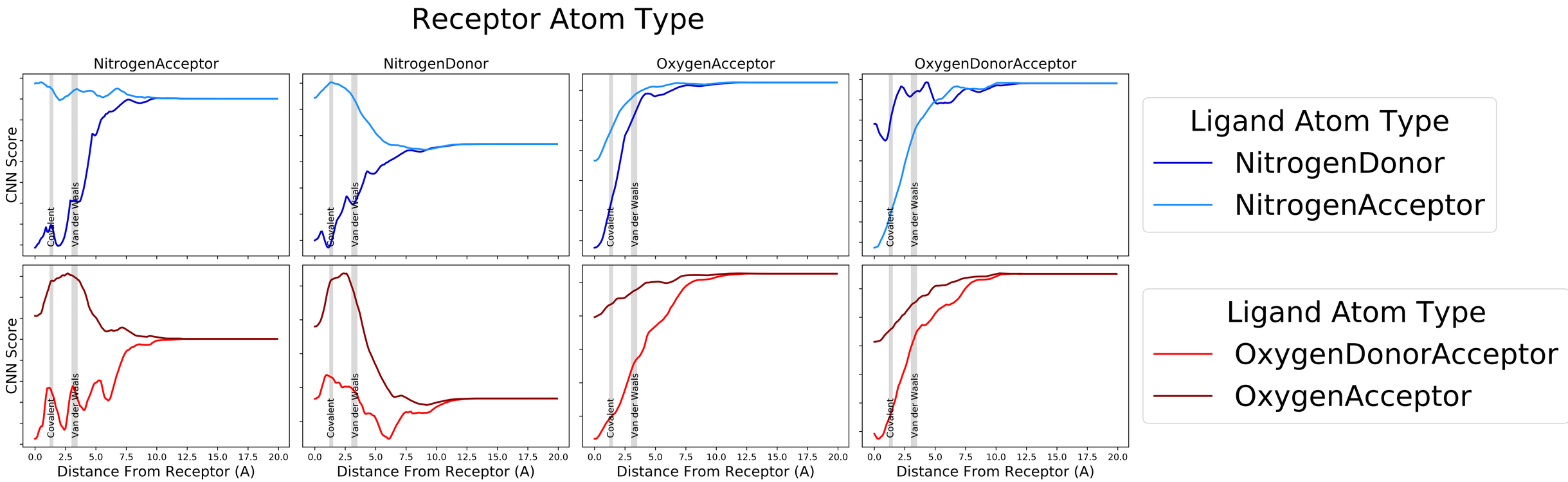

**Redocked Training Set**

# Visualizing with Atomistic Probes



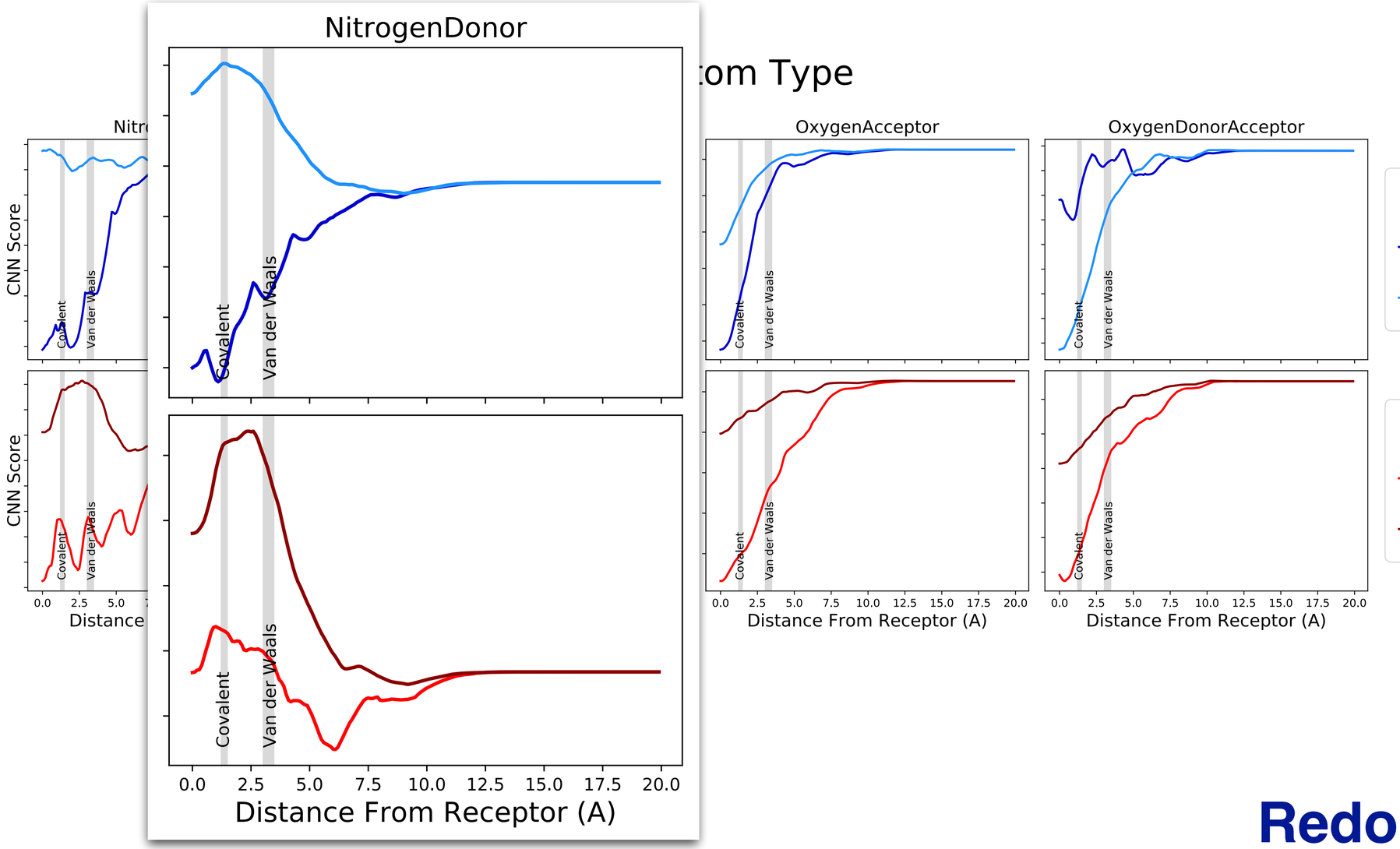

**Crossdocked Training Set**

# Hydrogen Bonds… or Not




**Redocked Training Set**

# Hydrogen Bonds... or Not

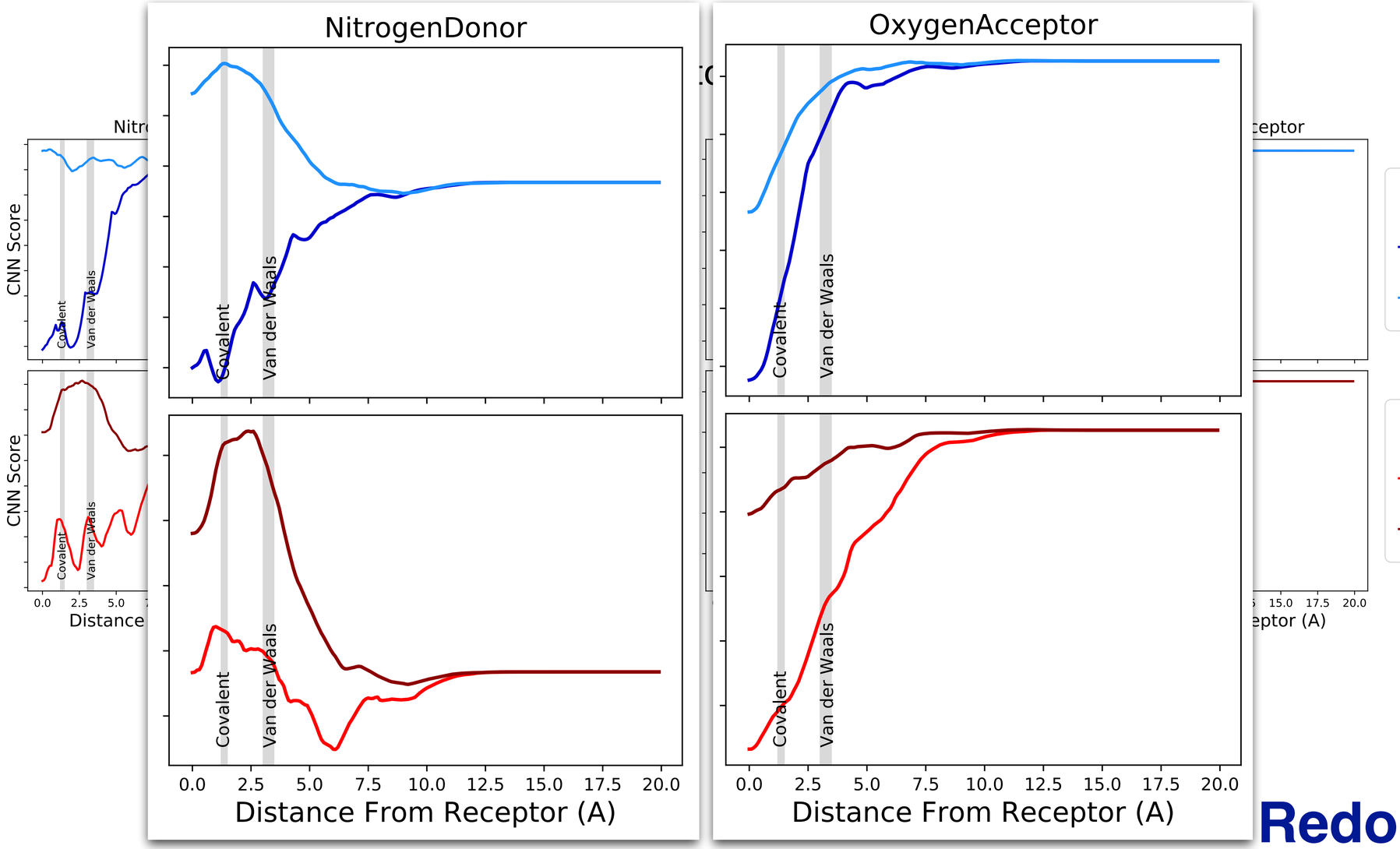



Ligand Atom Type
- NitrogenDonor
- NitrogenAcceptor

Ligand Atom Type
- OxygenDonorAcceptor
- OxygenAcceptor

**Redocked Training Set**

# Hydrogen Bonds... or Not

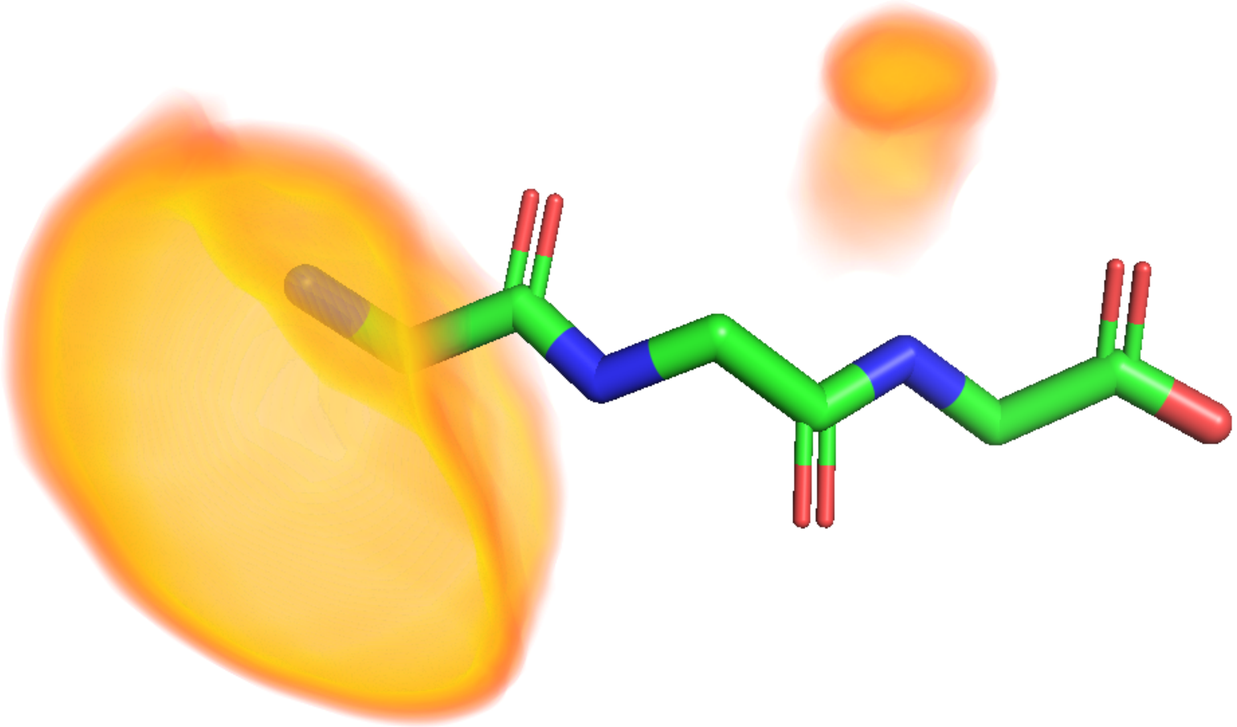



Ligand Atom Type
- NitrogenDonor
- NitrogenAcceptor

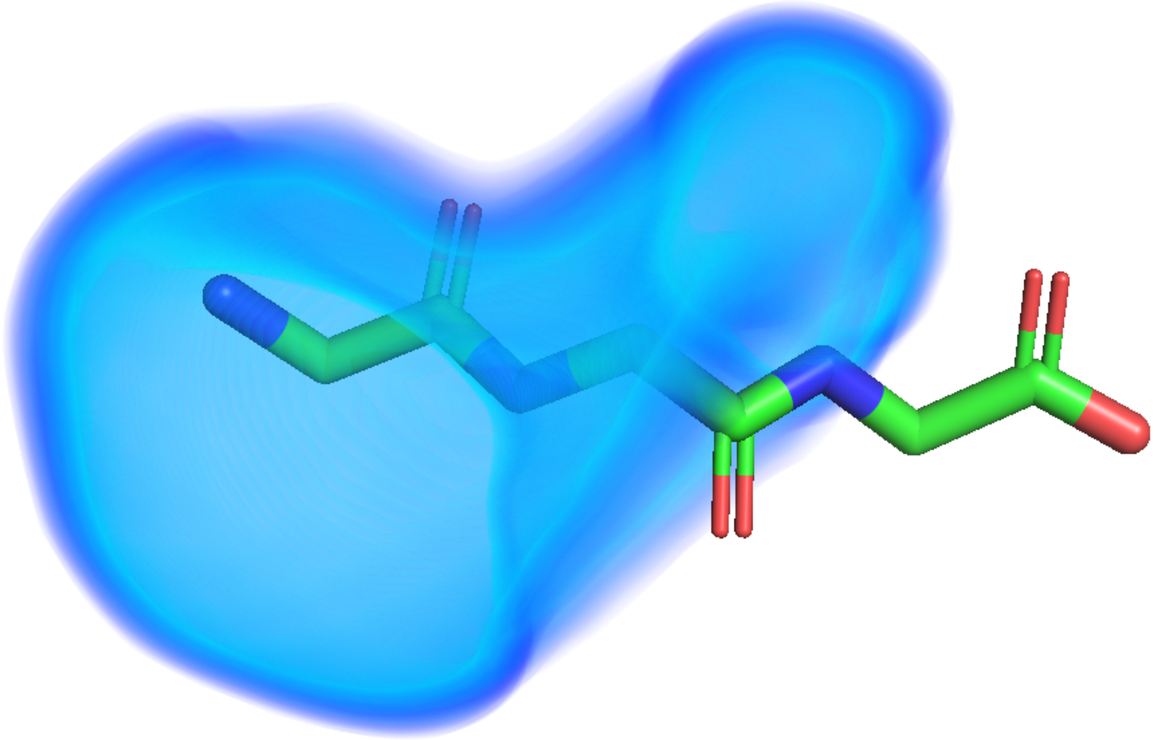
Ligand Atom Type
- OxygenDonorAcceptor
- OxygenAcceptor

**Redocked Training Set**

# Visualizing with Atomistic Probes

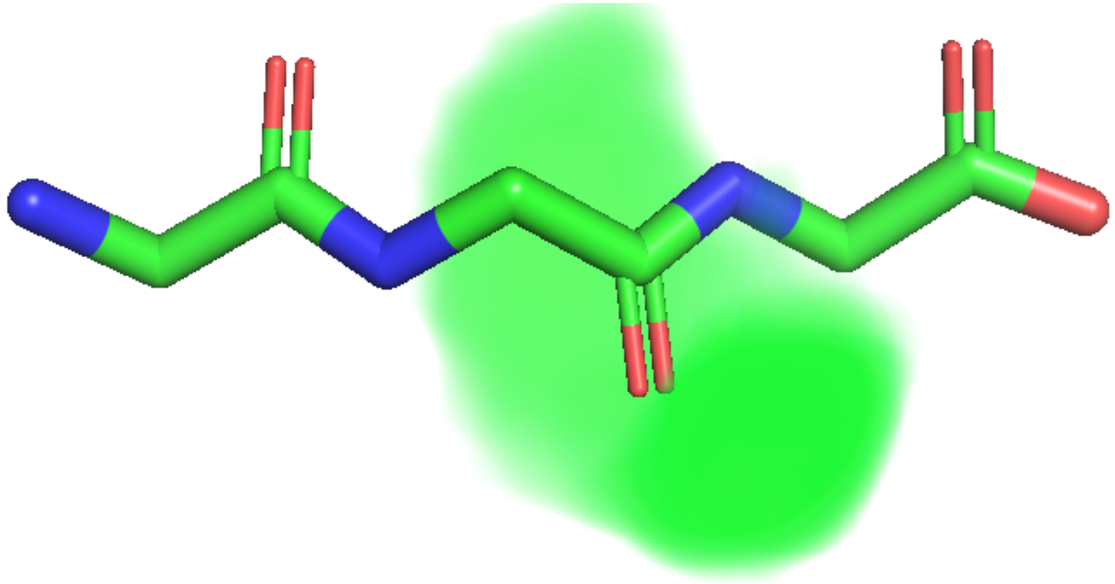


Oxygen Acceptor

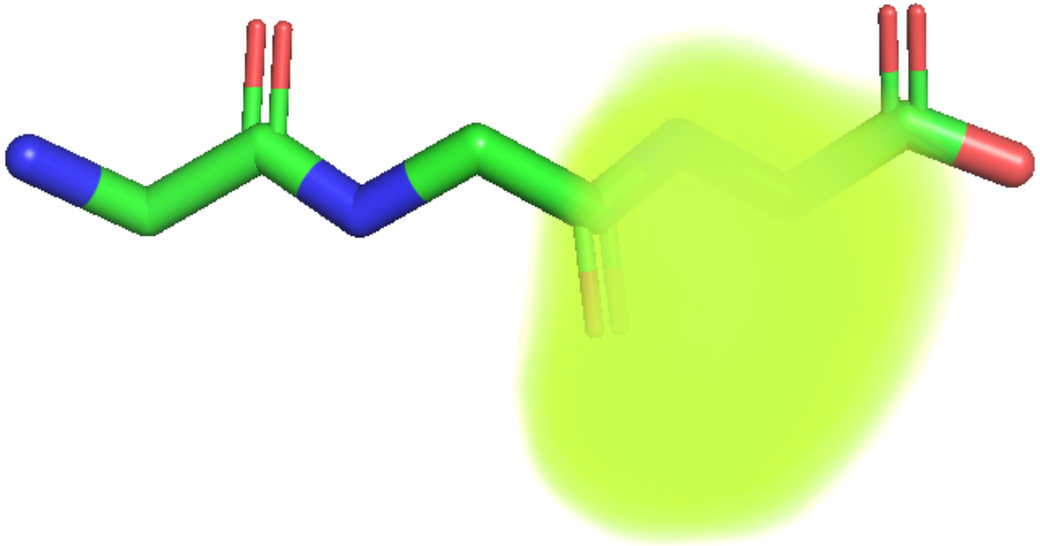


Nitrogen Acceptor

# Visualizing with Atomistic Probes

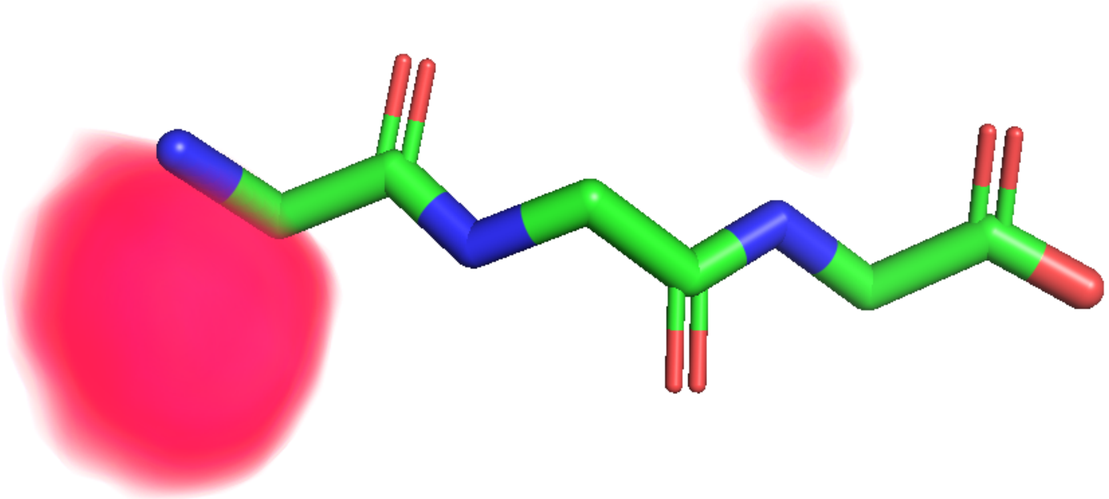


Aliphatic Carbon

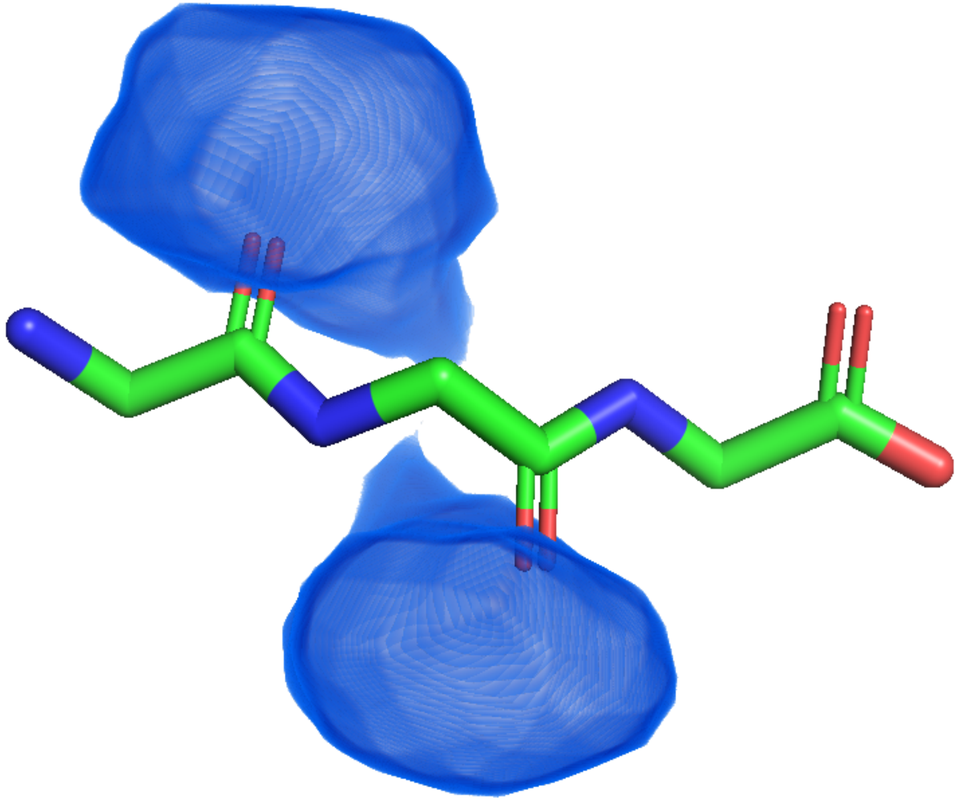


Aromatic Carbon

# Visualizing with Atomistic Probes

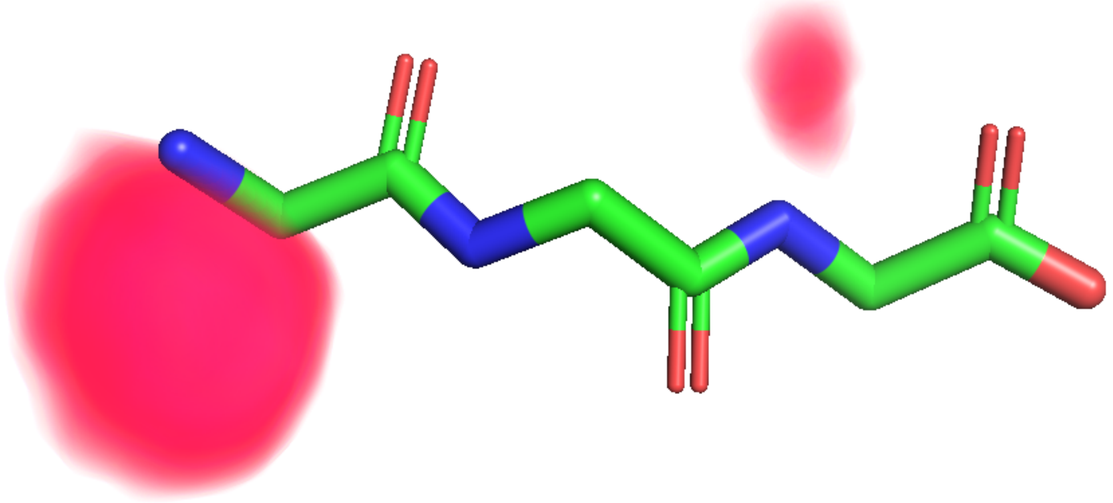


Oxygen Donor/Acceptor

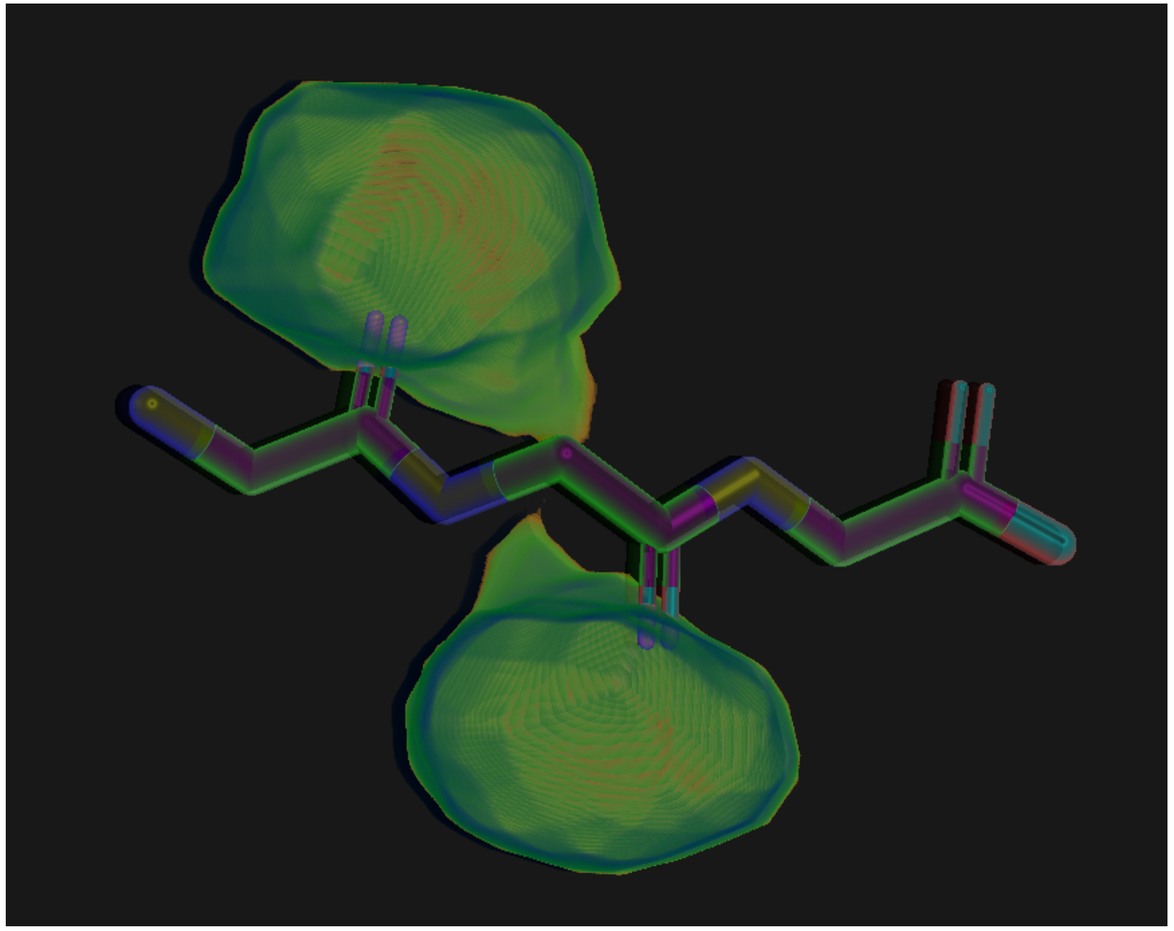


Nitrogen Donor

# Visualizing with Atomistic Probes

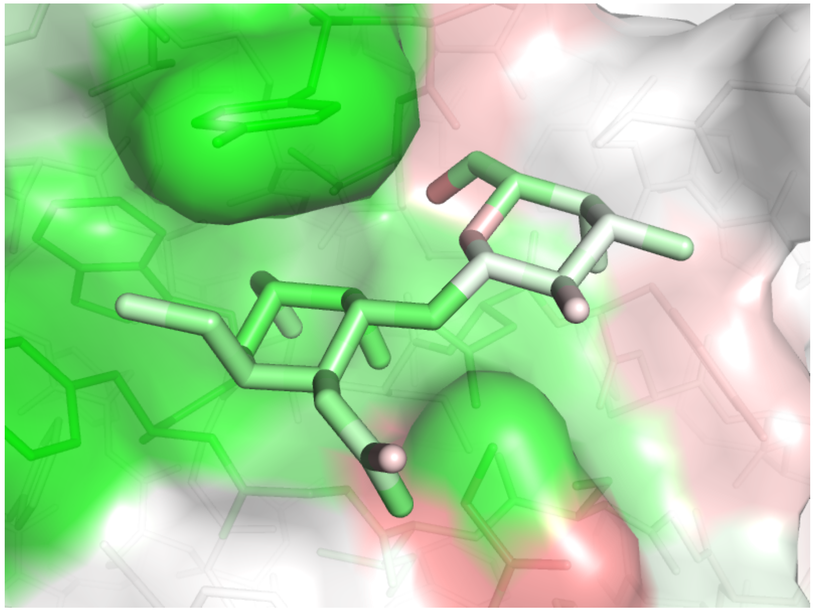


Oxygen Donor/Acceptor

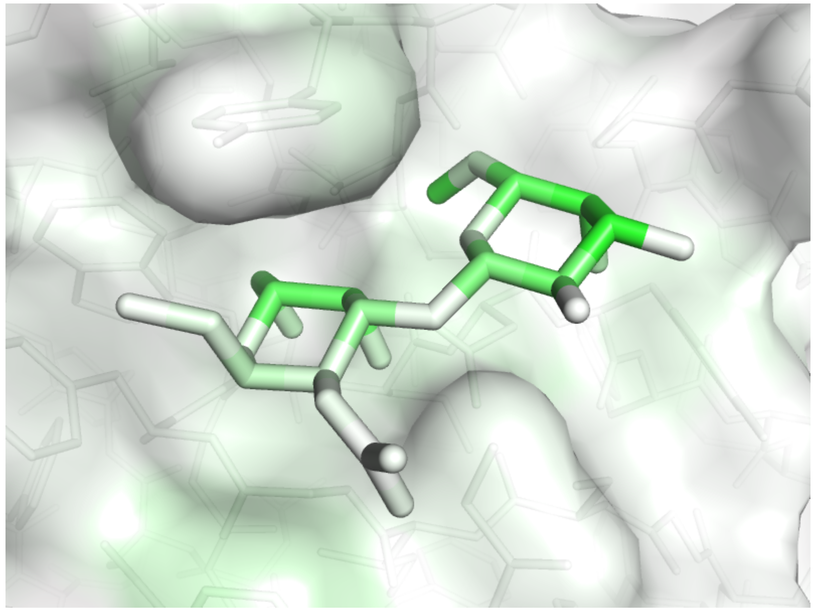


Nitrogen Donor

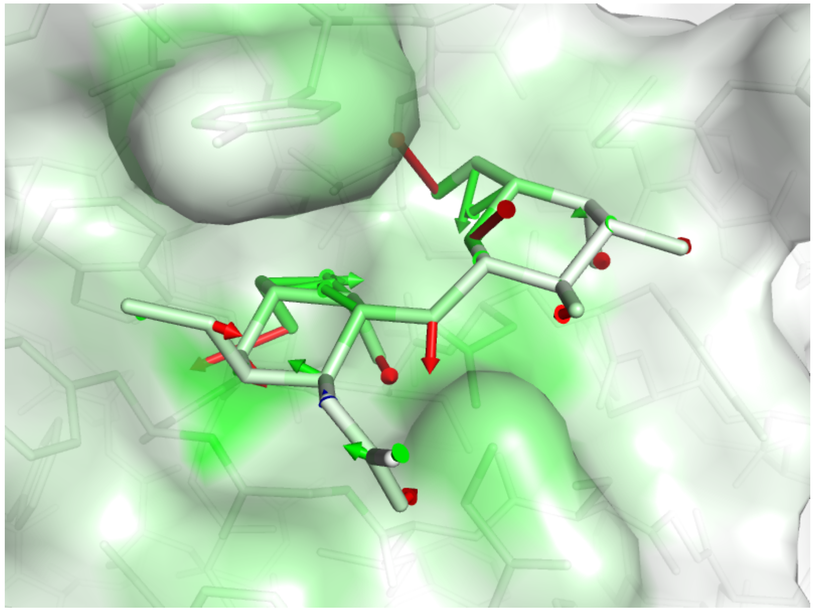
# Visualizing Network Decisions



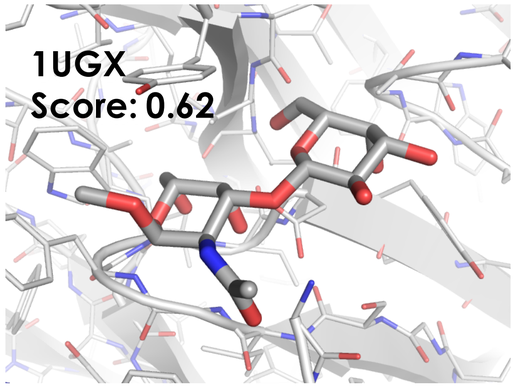
masking



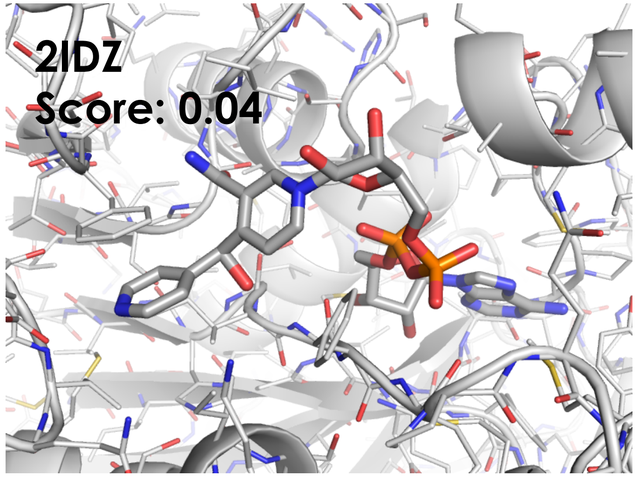
layer-wise relevance

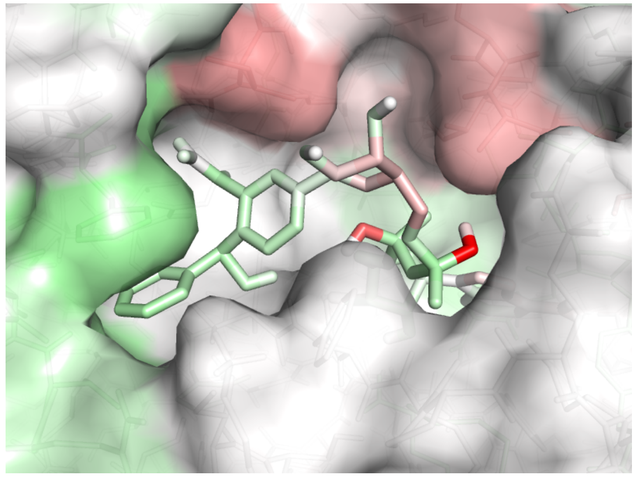


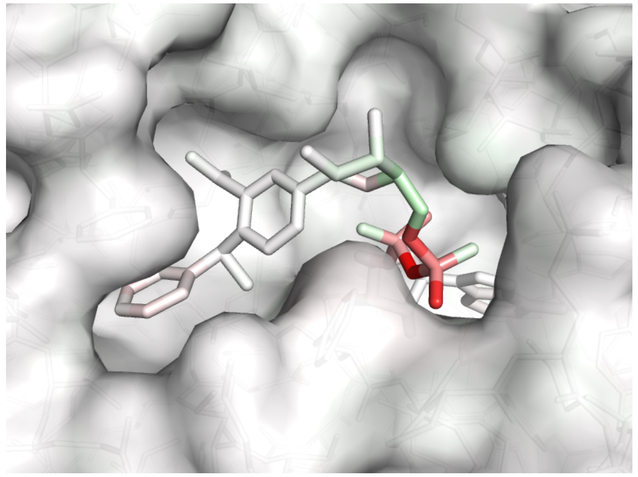
gradients

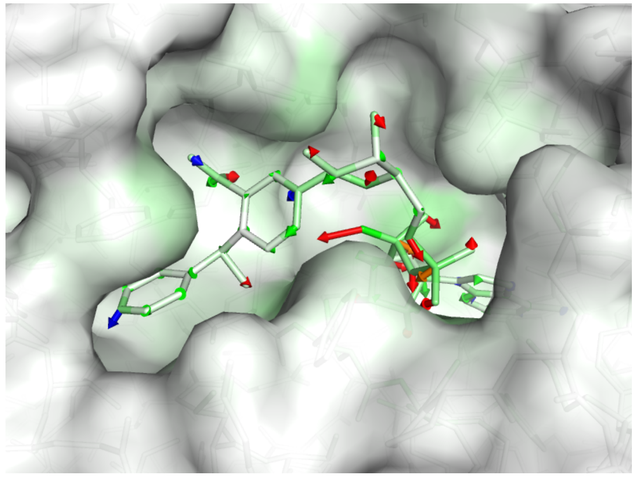


# Visualizations



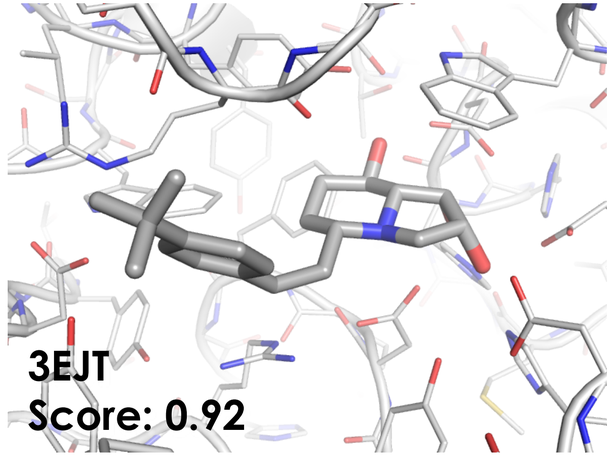


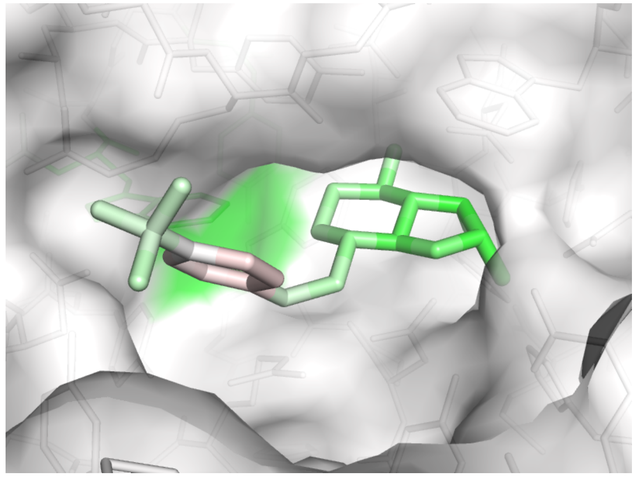


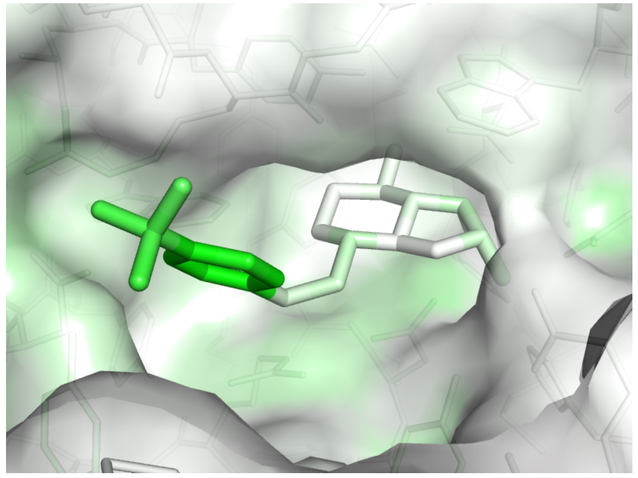
Masking

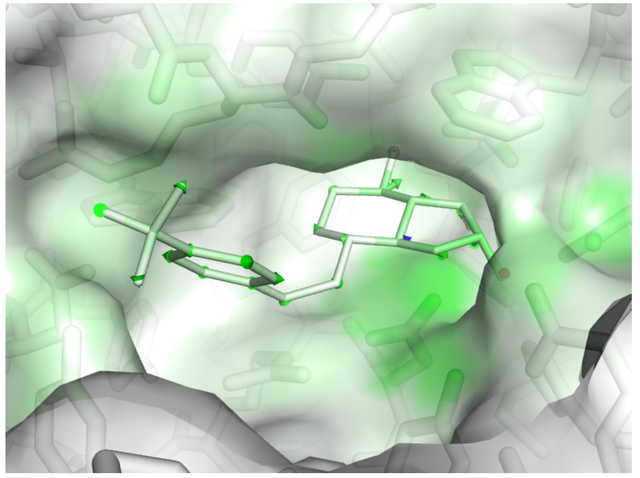
LRP

Gradients

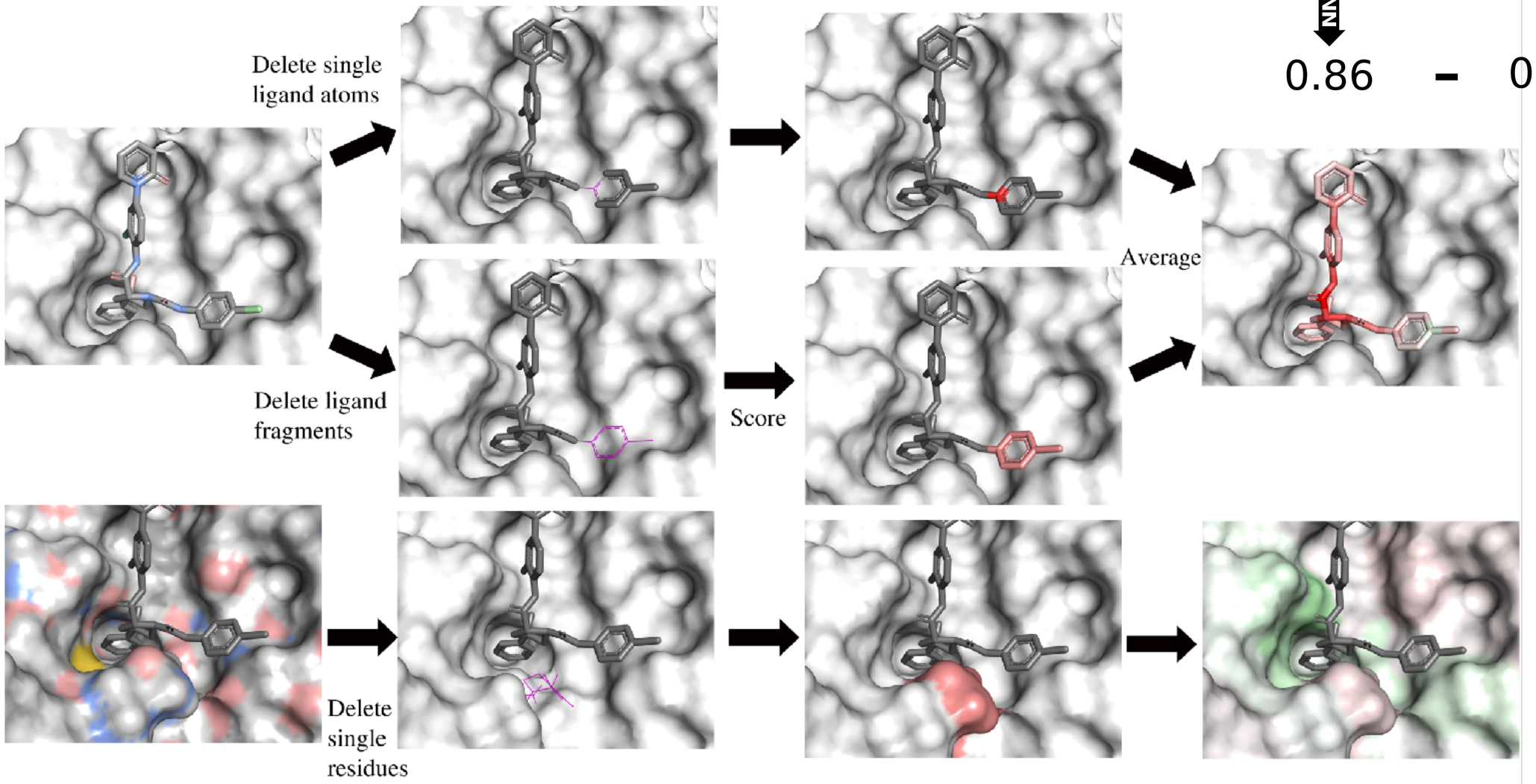


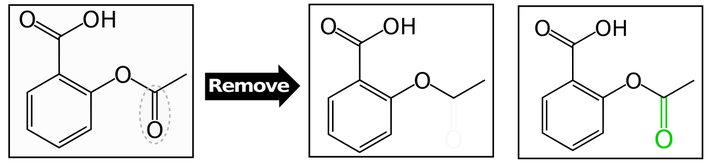







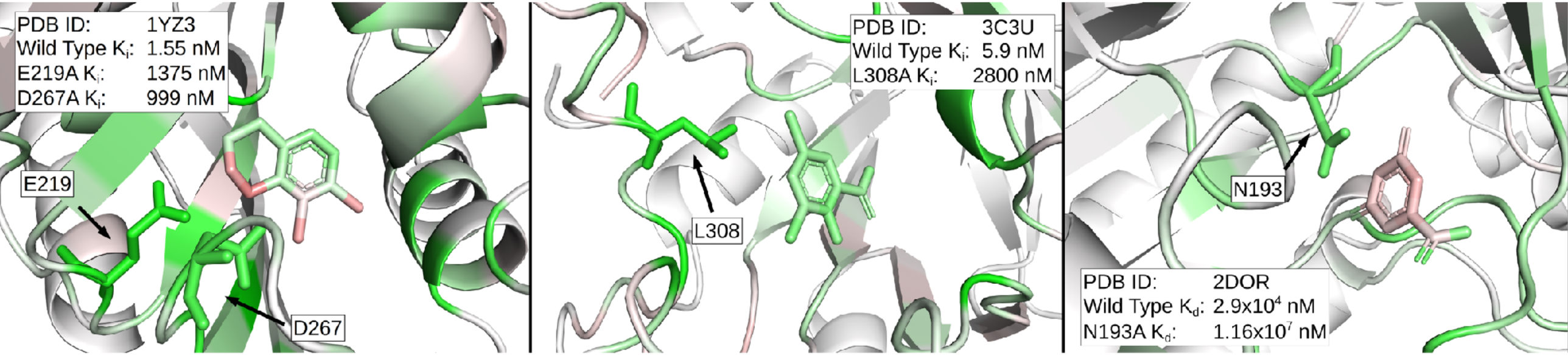
# Masking



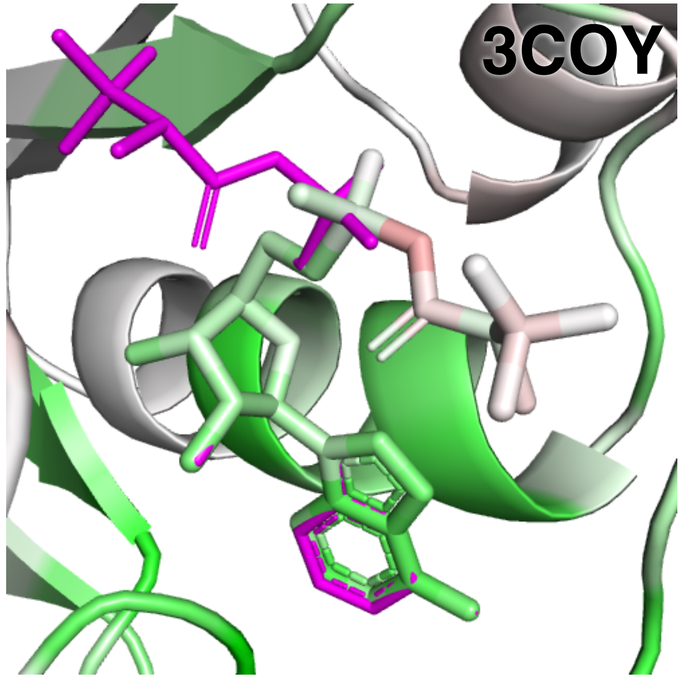



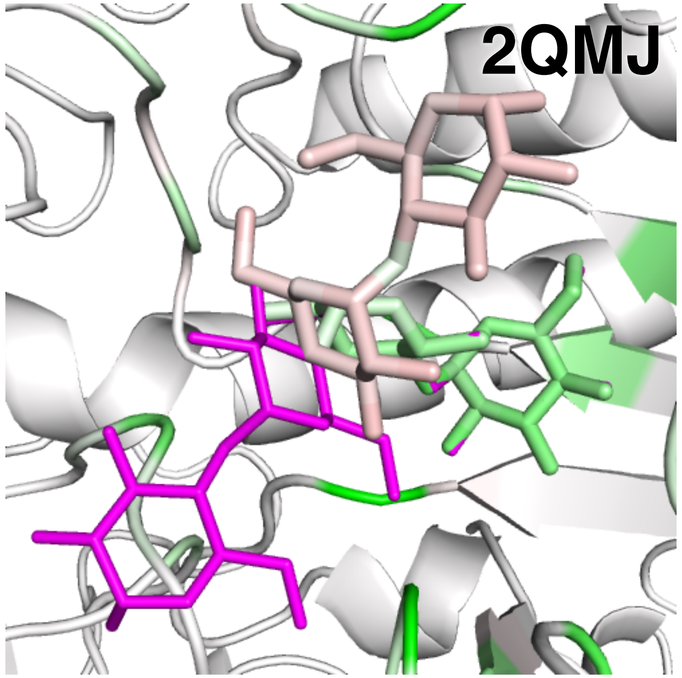

0.86 - 0.84 = 0.02

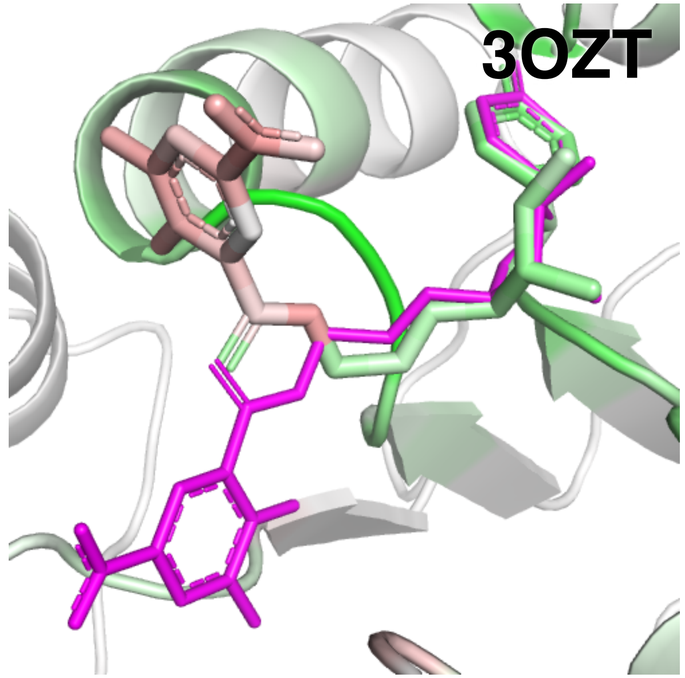
# Masking: Enzyme Mutants



# Pose Sensitivity



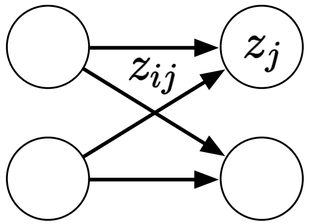




Partially Aligned Poses

# Layer-wise Relevance

$$f(x) = \ldots = \sum_{d \in l+1} R_d^{(l+1)} = \sum_{d \in l} R_d^{(l)} = \ldots = \sum_d R_d^{(1)}$$

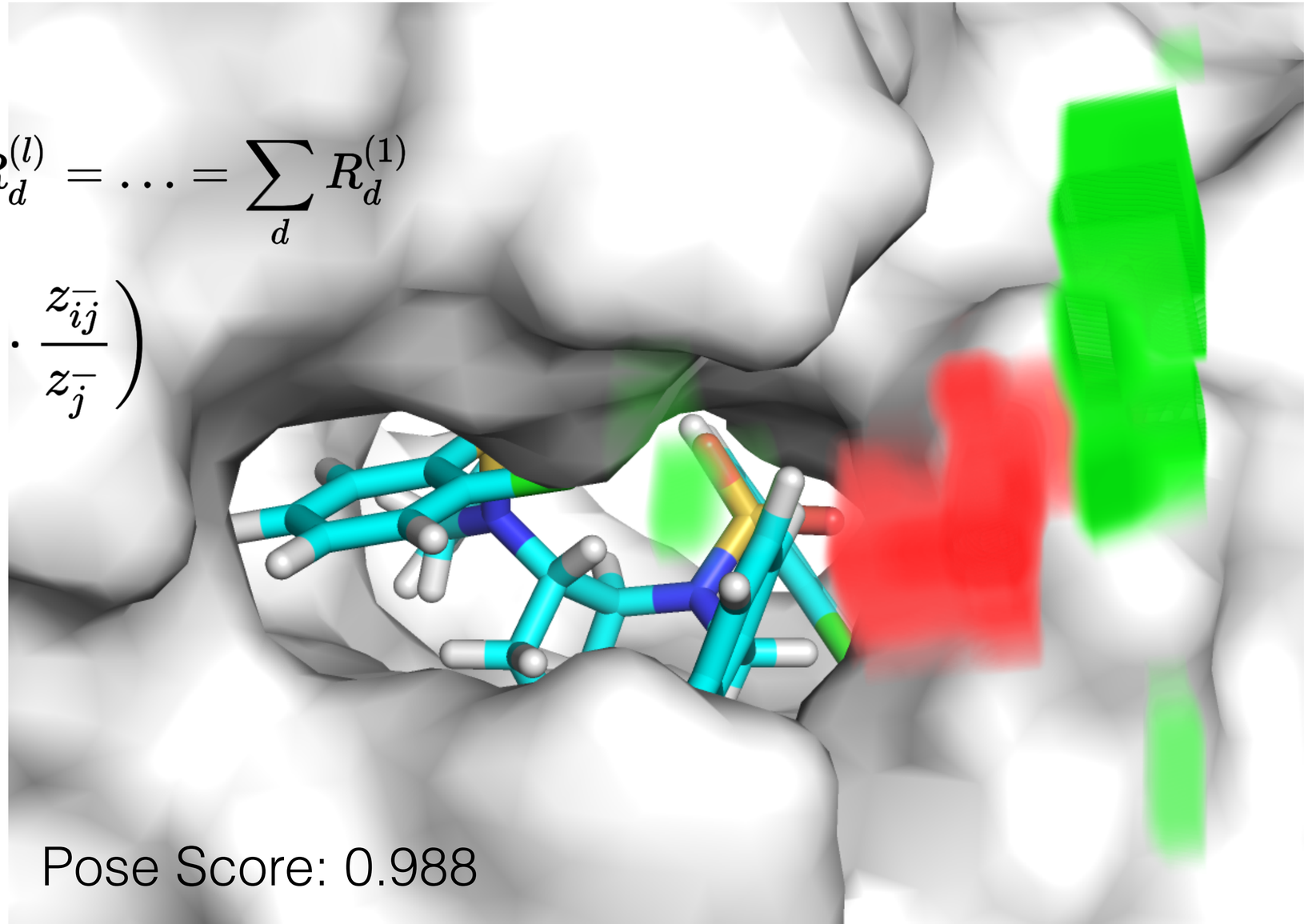
$$R_{i \leftarrow j}^{(l,l+1)} = R_j^{(l+1)} \cdot \left( \alpha \cdot \frac{z_{ij}^+}{z_j^+} + \beta \cdot \frac{z_{ij}^-}{z_j^-} \right)$$



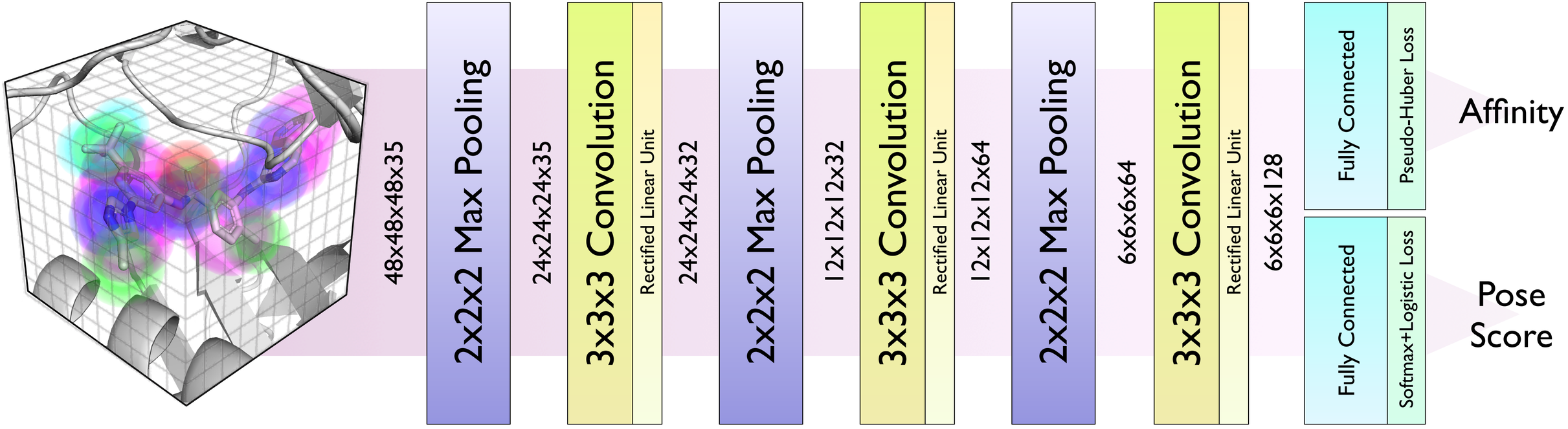

**On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation**

Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, Wojciech Samek

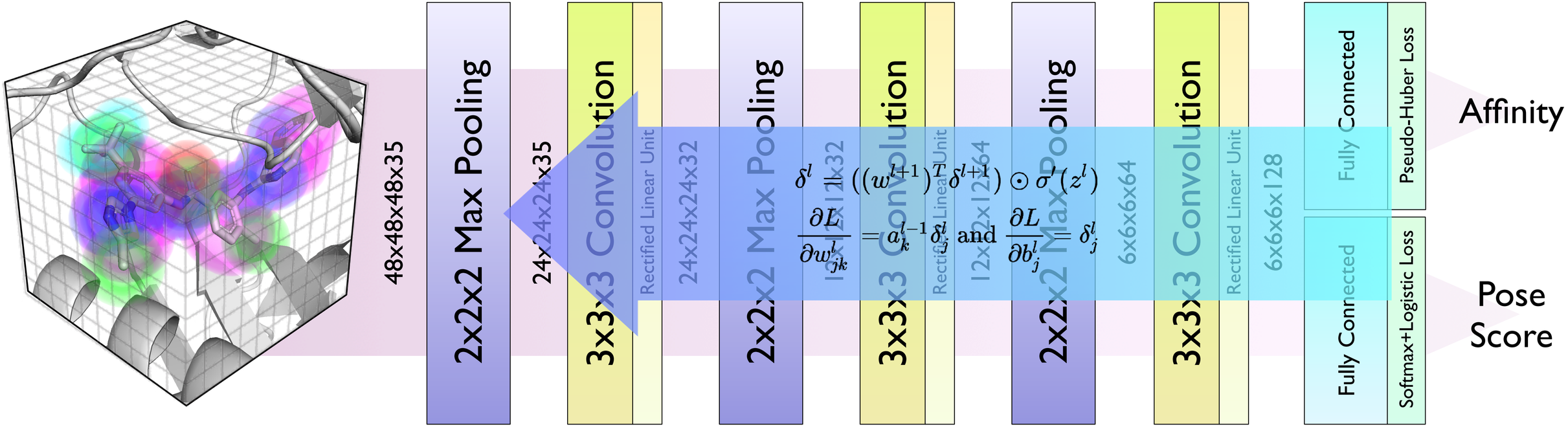
Published: July 10, 2015 • https://doi.org/10.1371/journal.pone.0130140



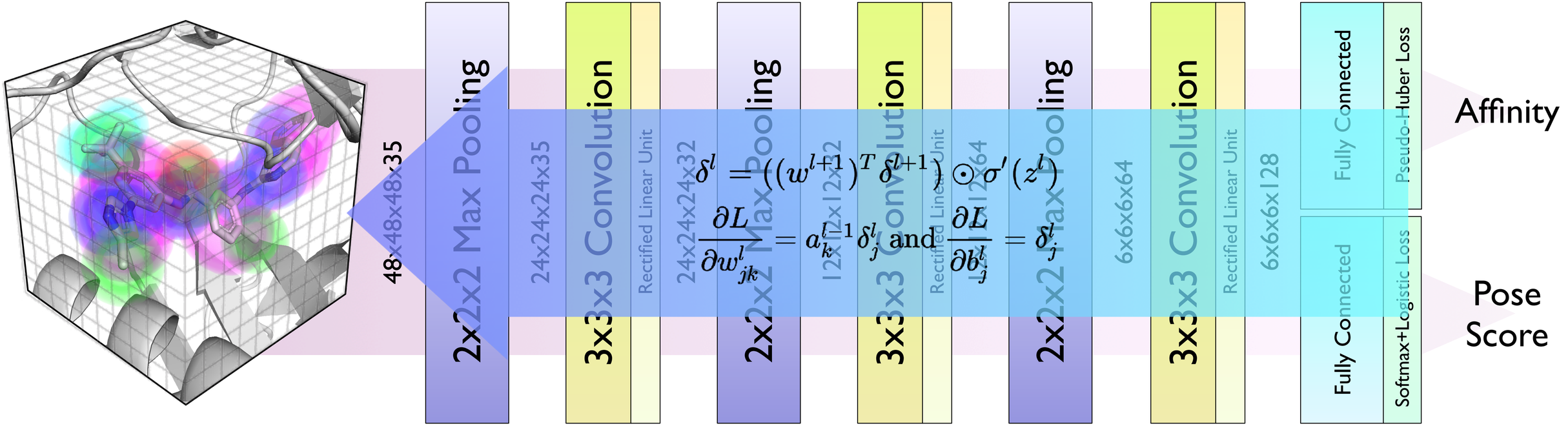
# Gradients


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

# Gradients


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
$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$
$\frac{\partial L}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j$ and $\frac{\partial L}{\partial b^l_j} = \delta^l_j$

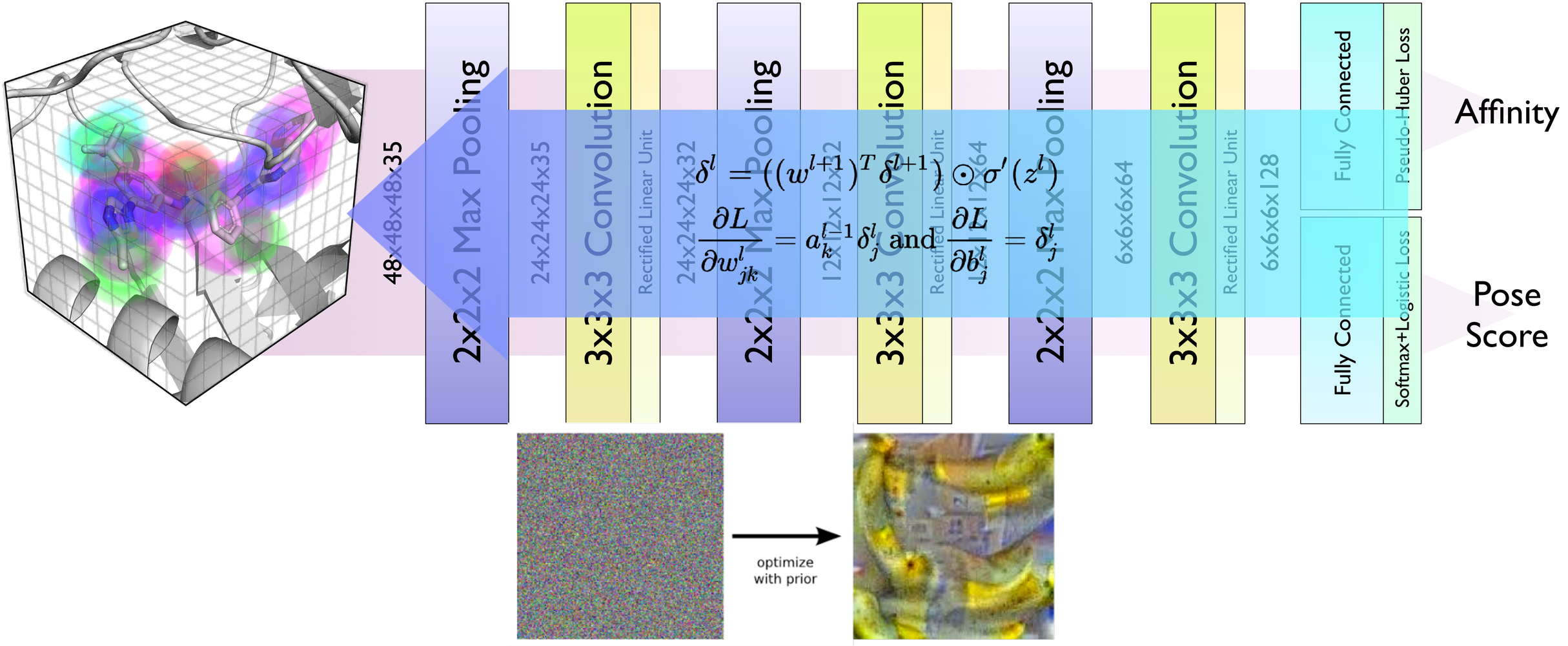
# Gradients



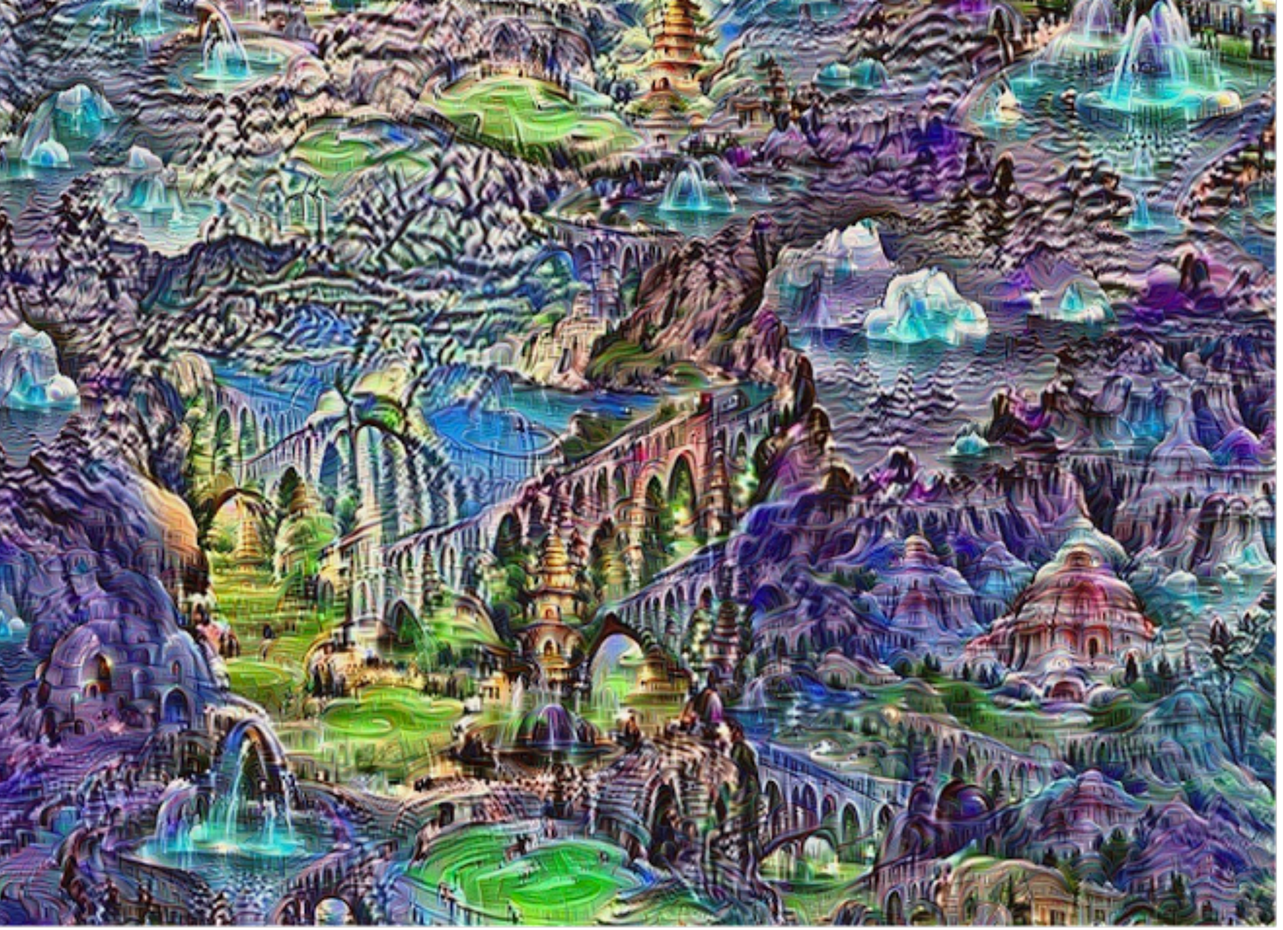
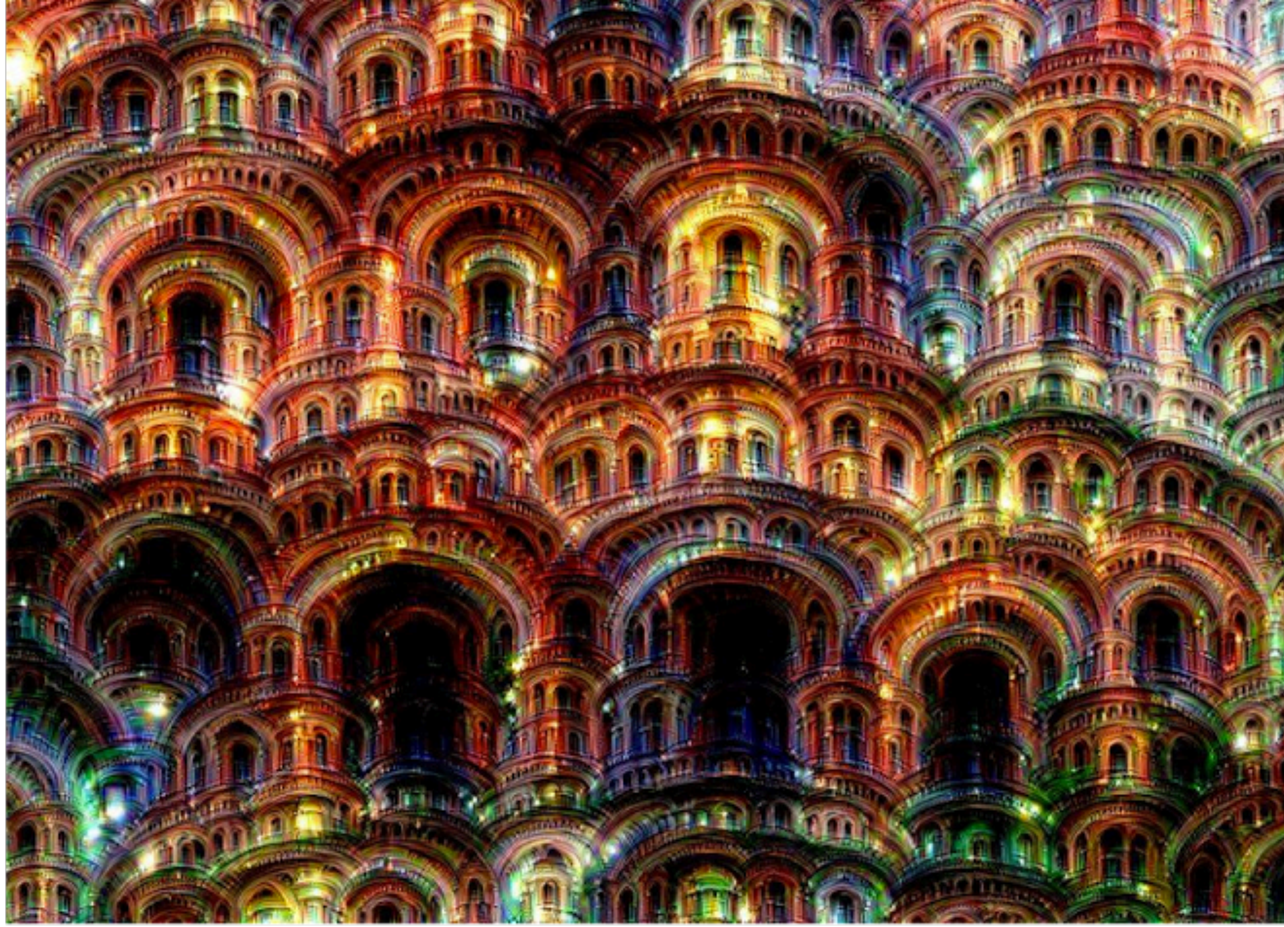
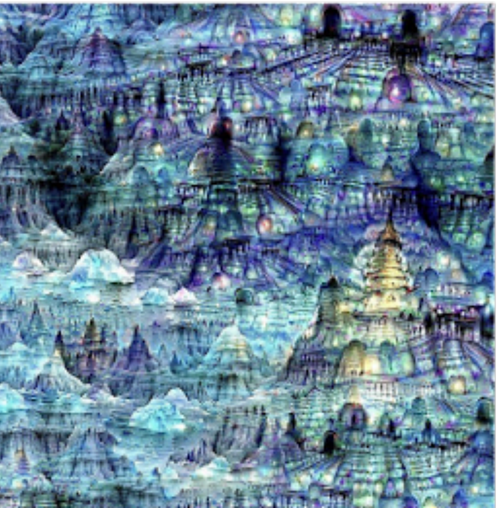
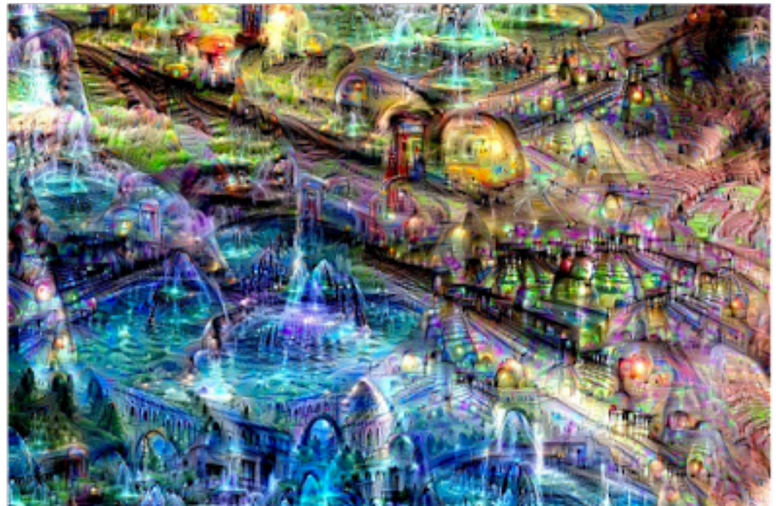

$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

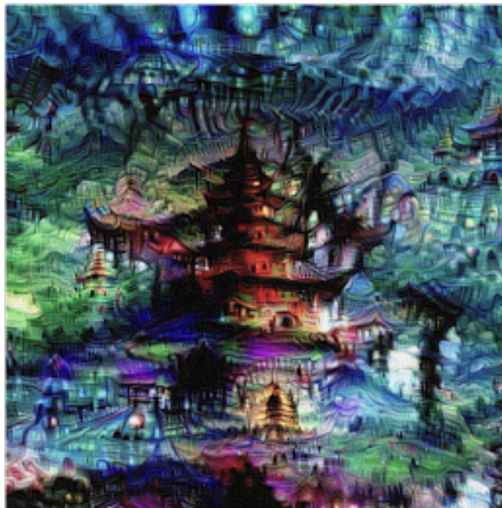
$$\frac{\partial L}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j \text{ and } \frac{\partial L}{\partial b^l_j} = \delta^l_j$$

# Gradients

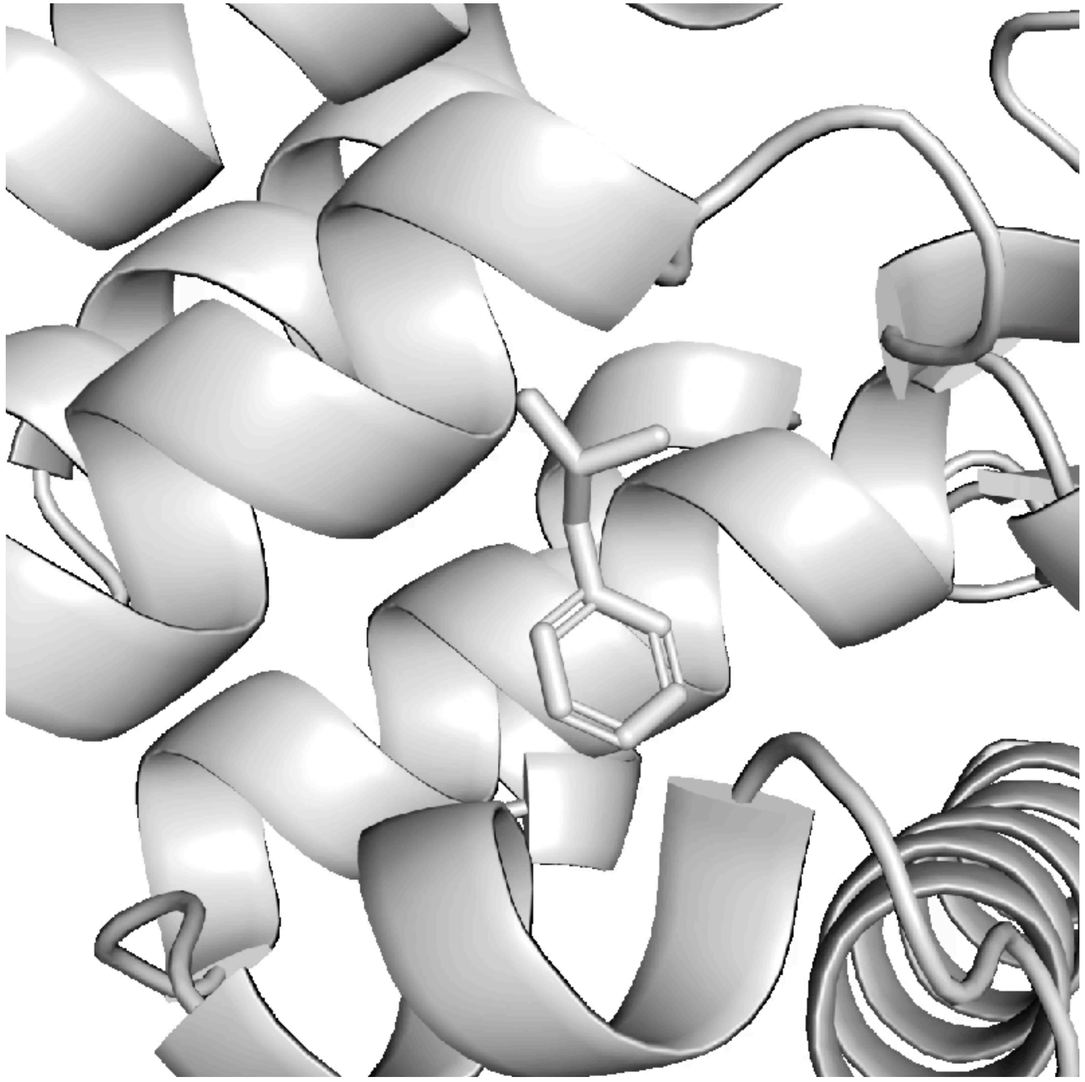


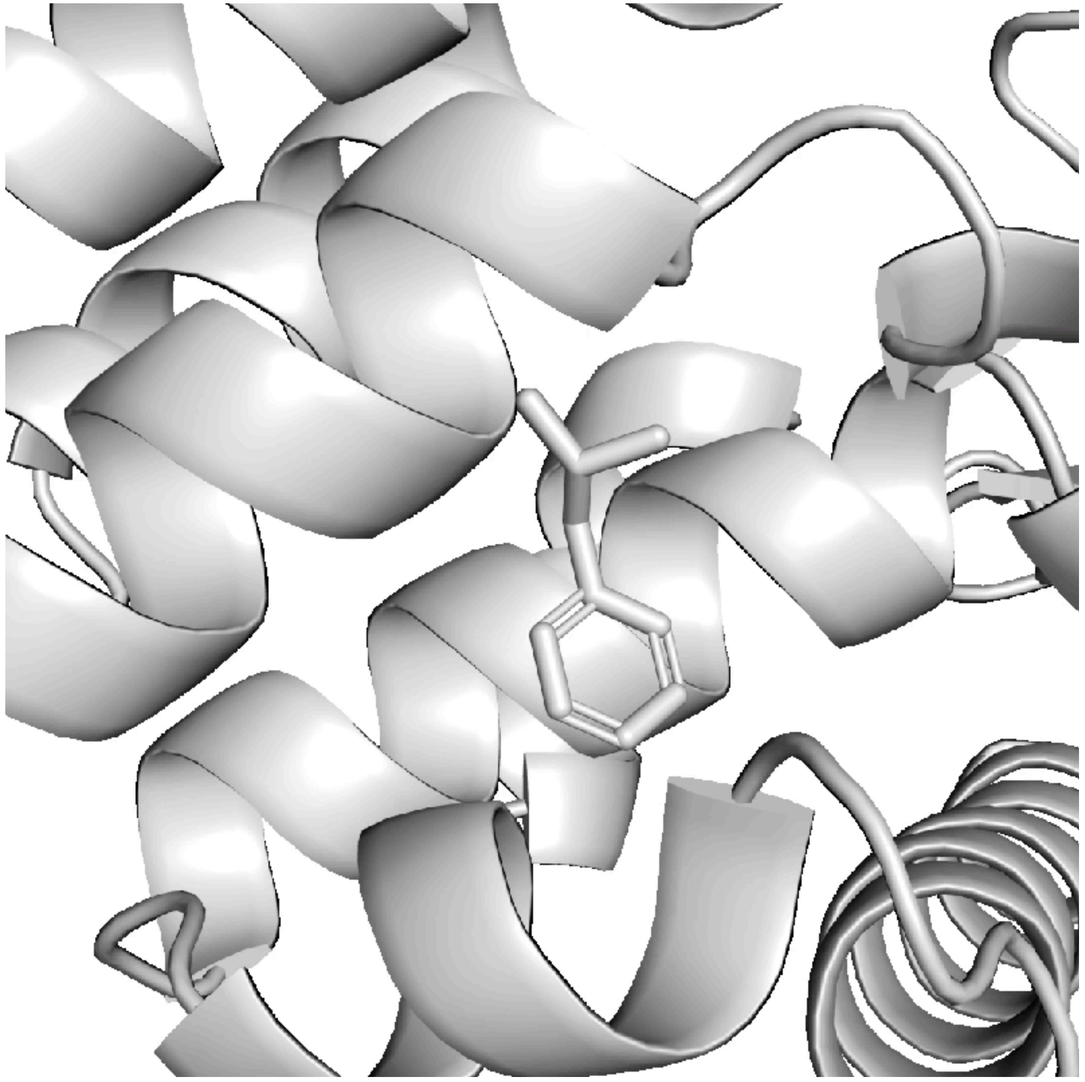

https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html



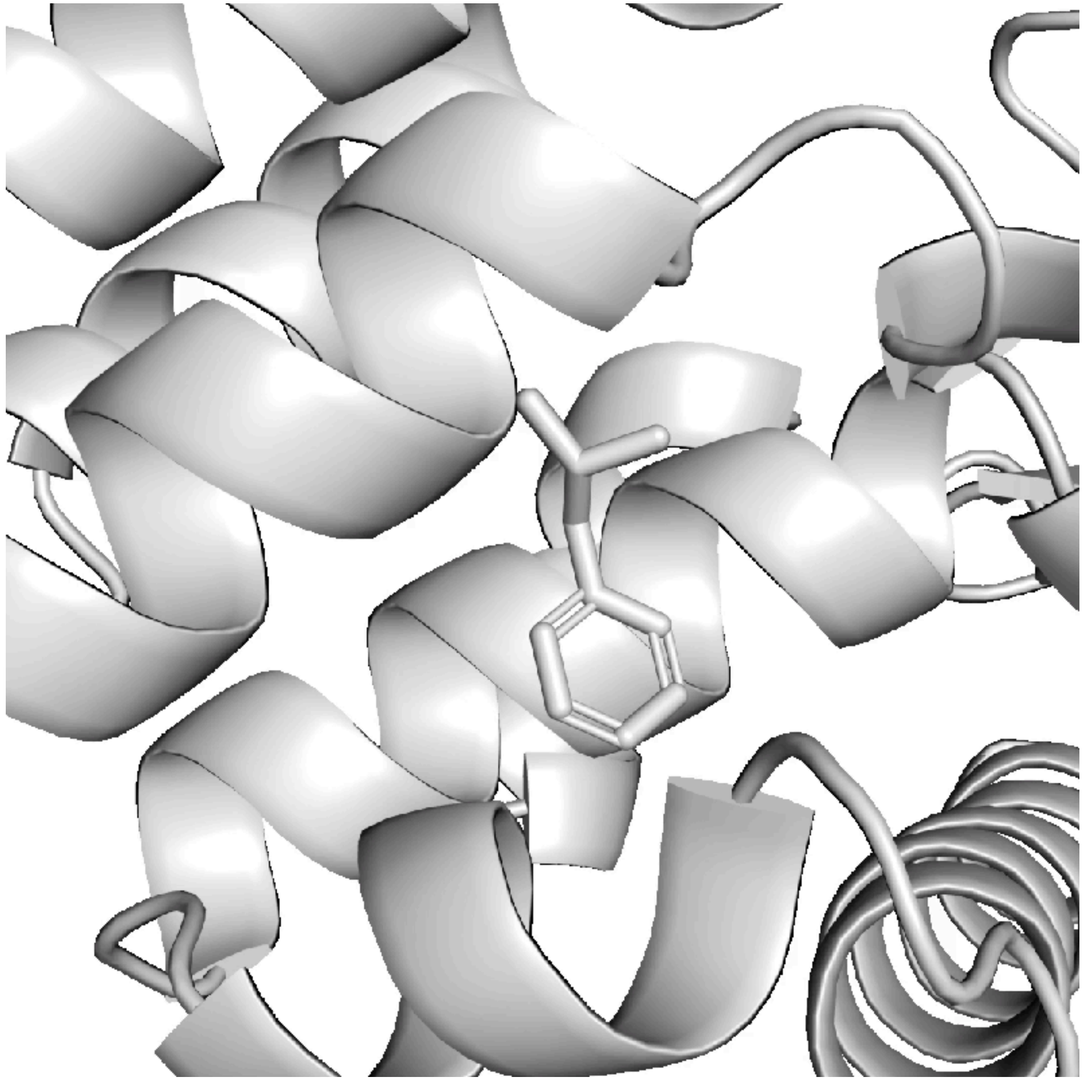
# Deep Dreams of Molecules

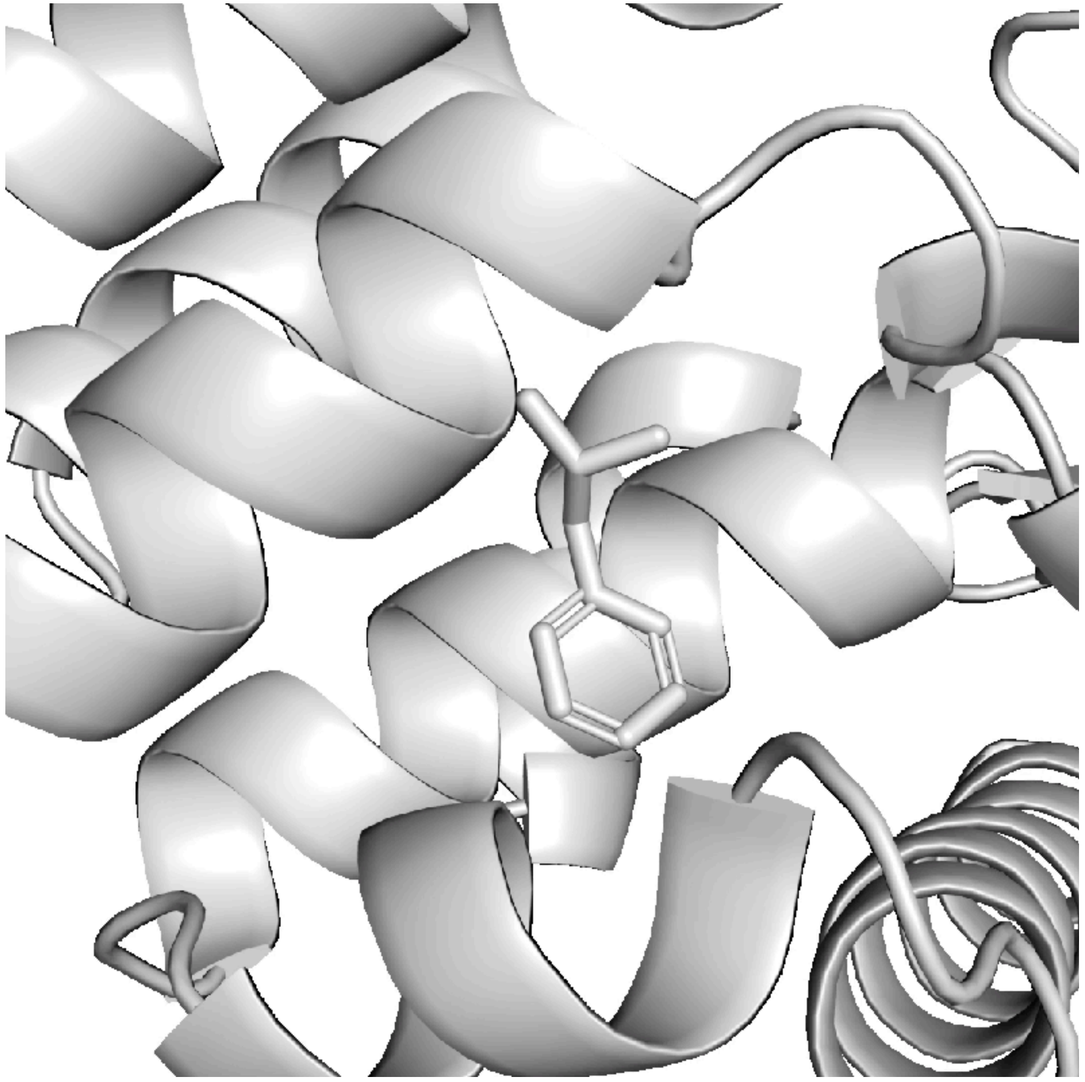




184l

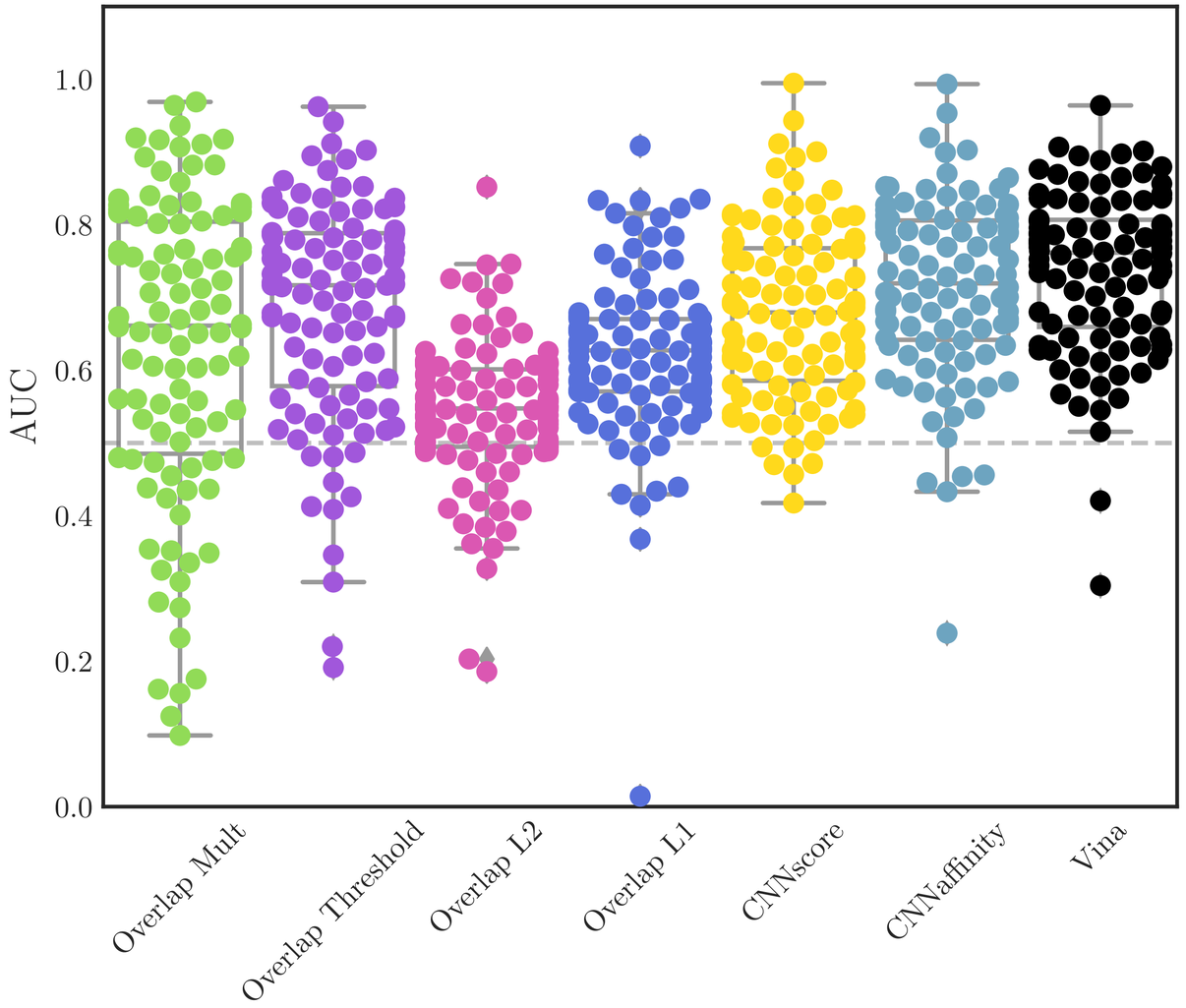
# Deep Dreams of Molecules

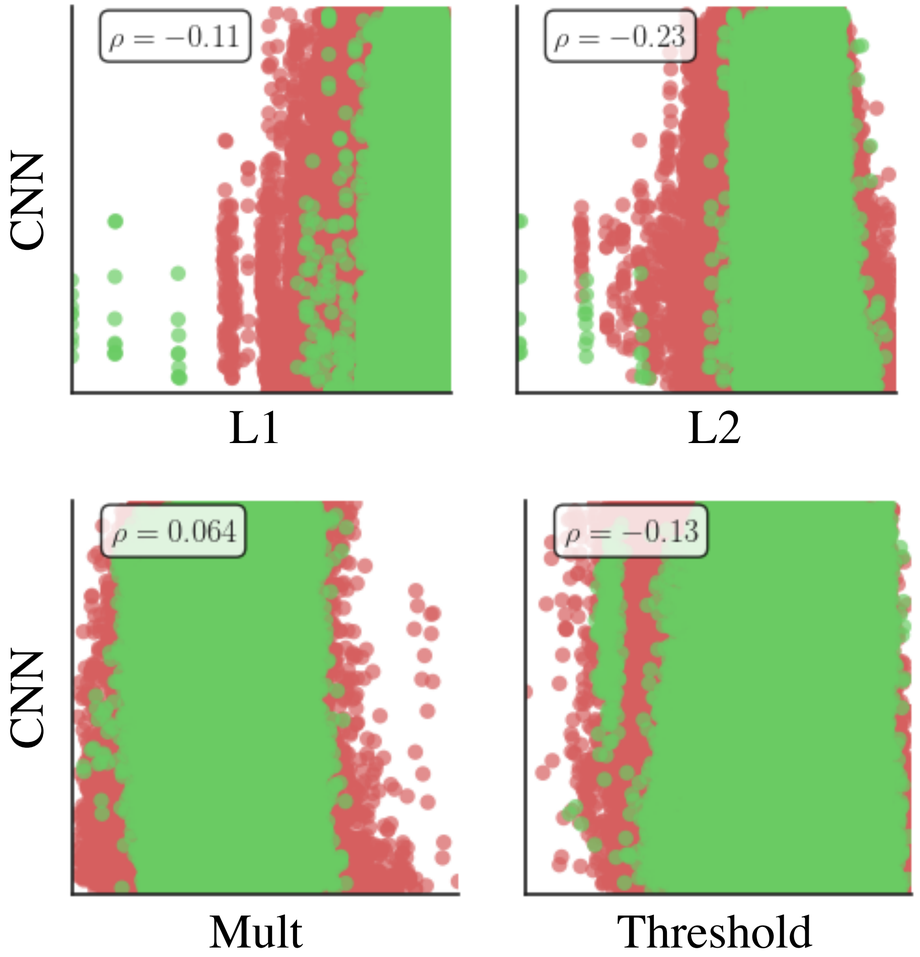




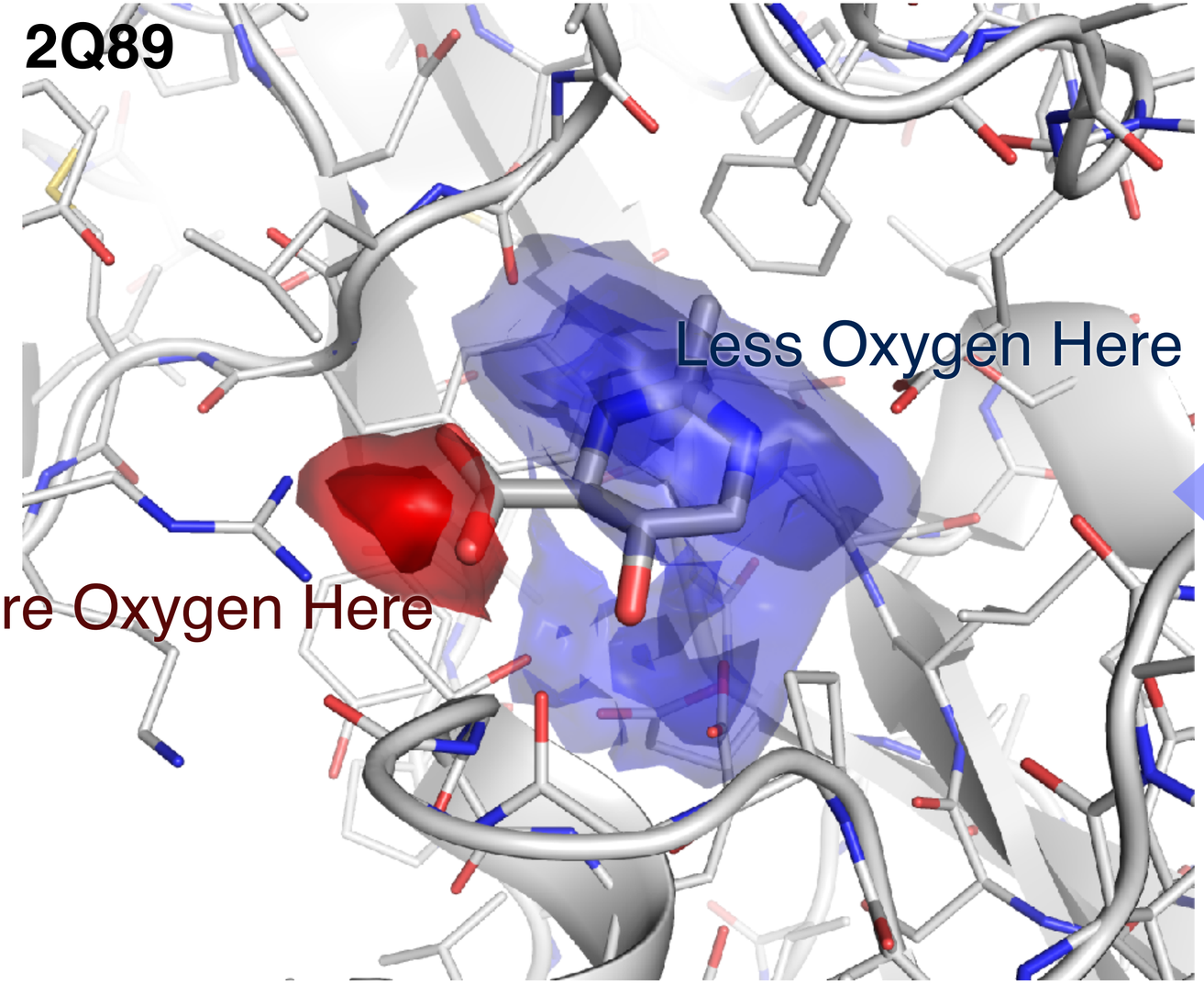
184l

# Screening with Pseudo Ligands



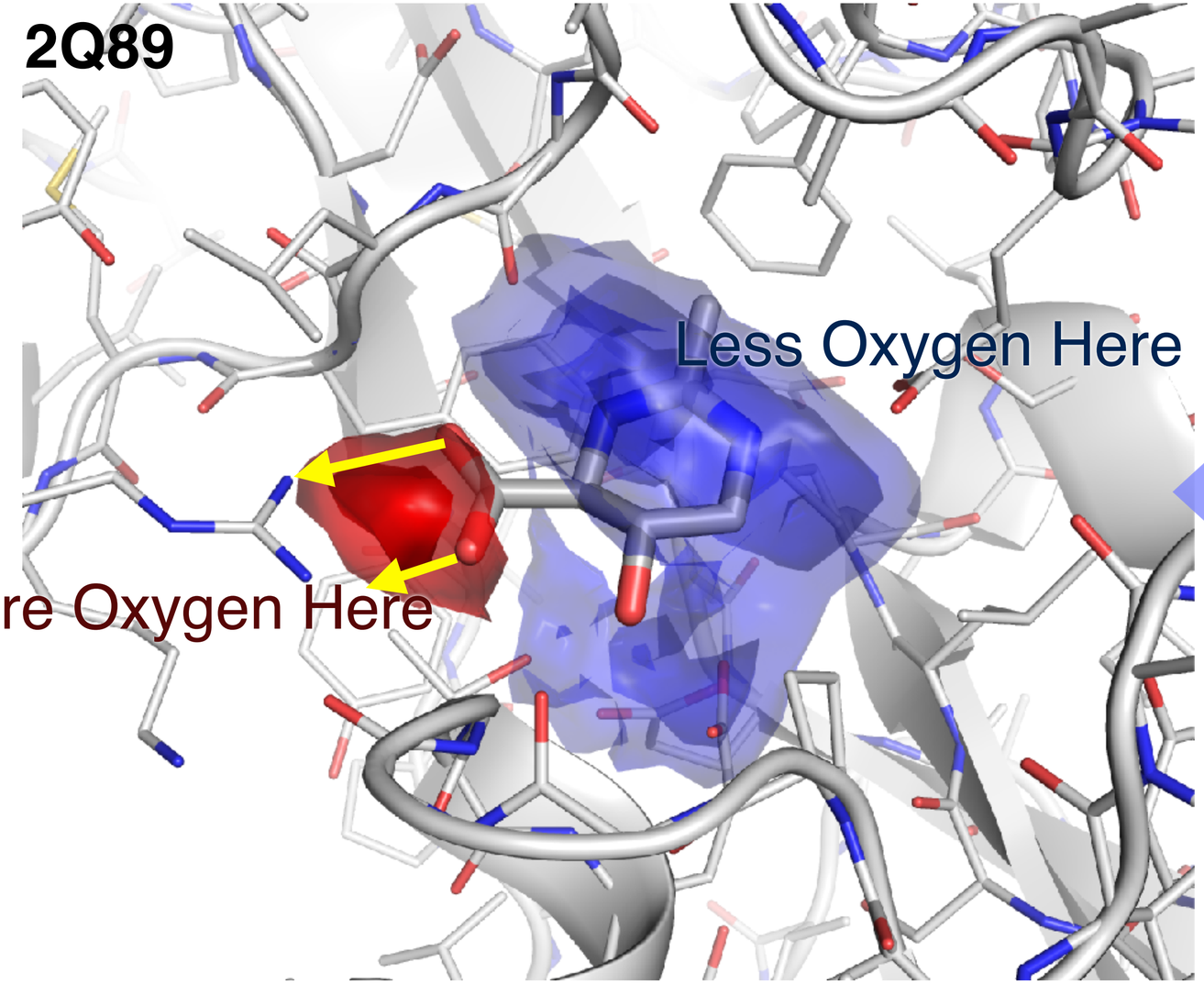



# Gradients: Beyond Scoring

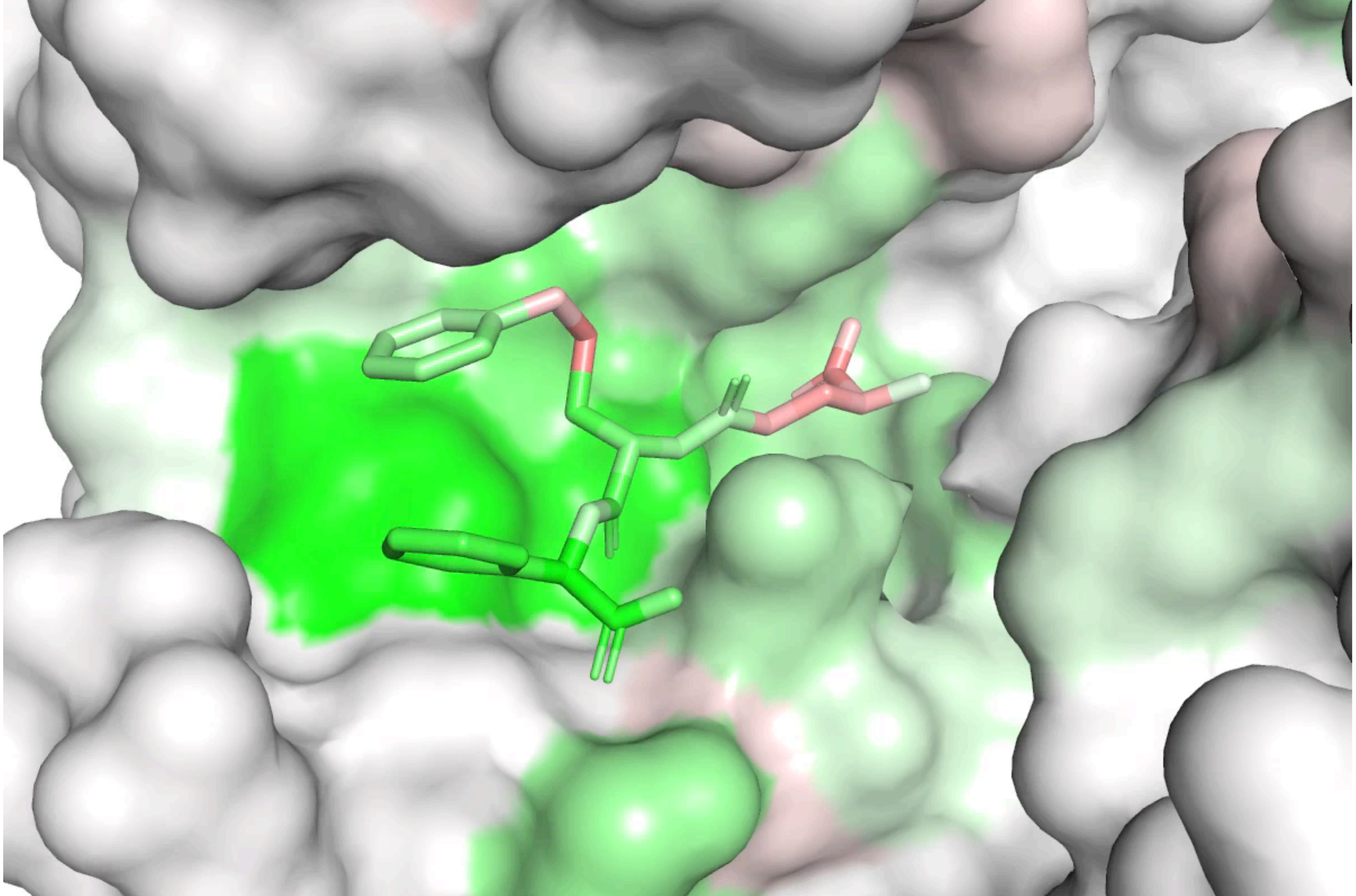


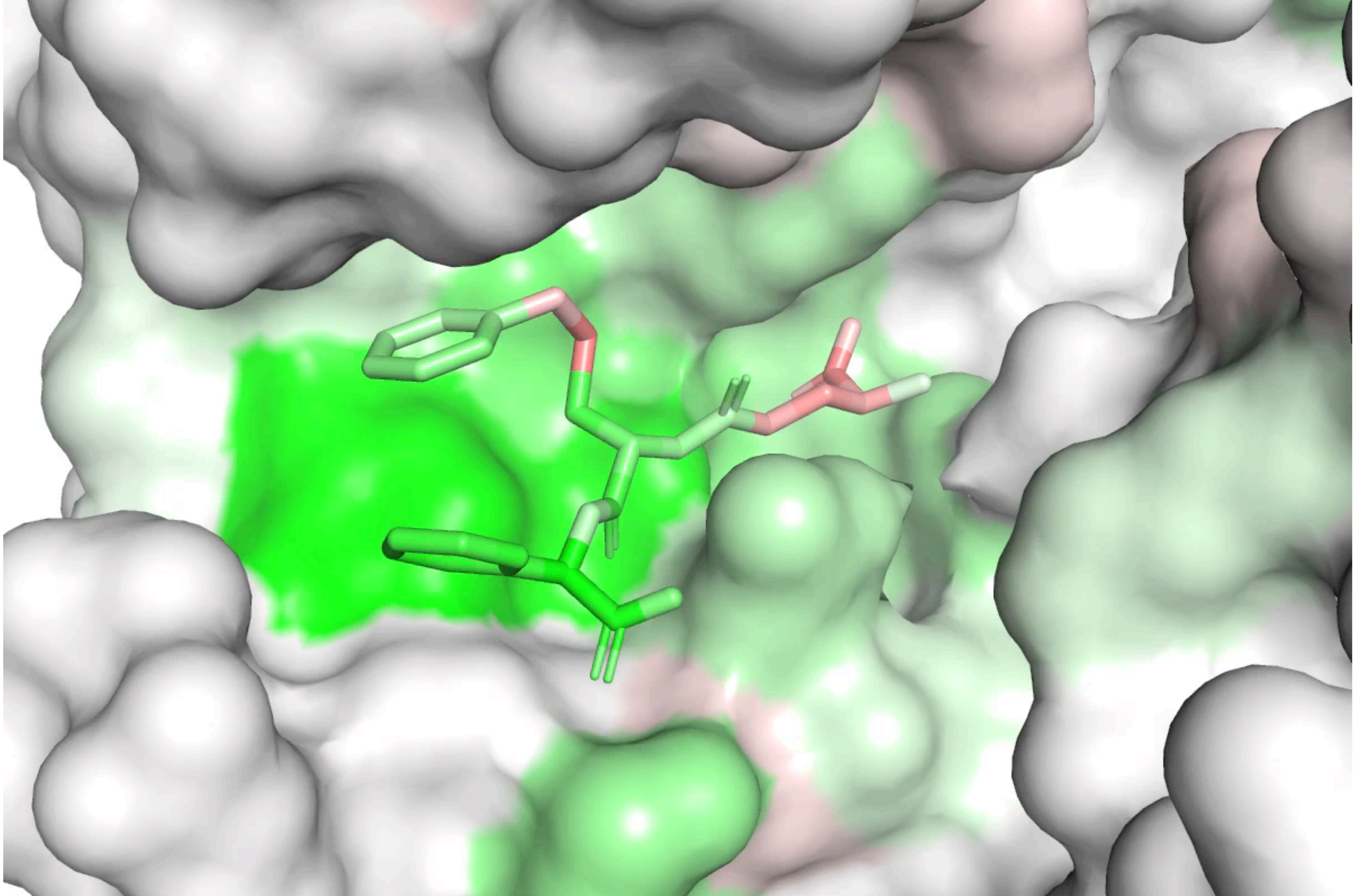

$$\frac{\partial L}{\partial A} = \sum_{i \in G_A} \frac{\partial L}{\partial G_i} \frac{\partial G_i}{\partial D} \frac{\partial D}{\partial A}$$

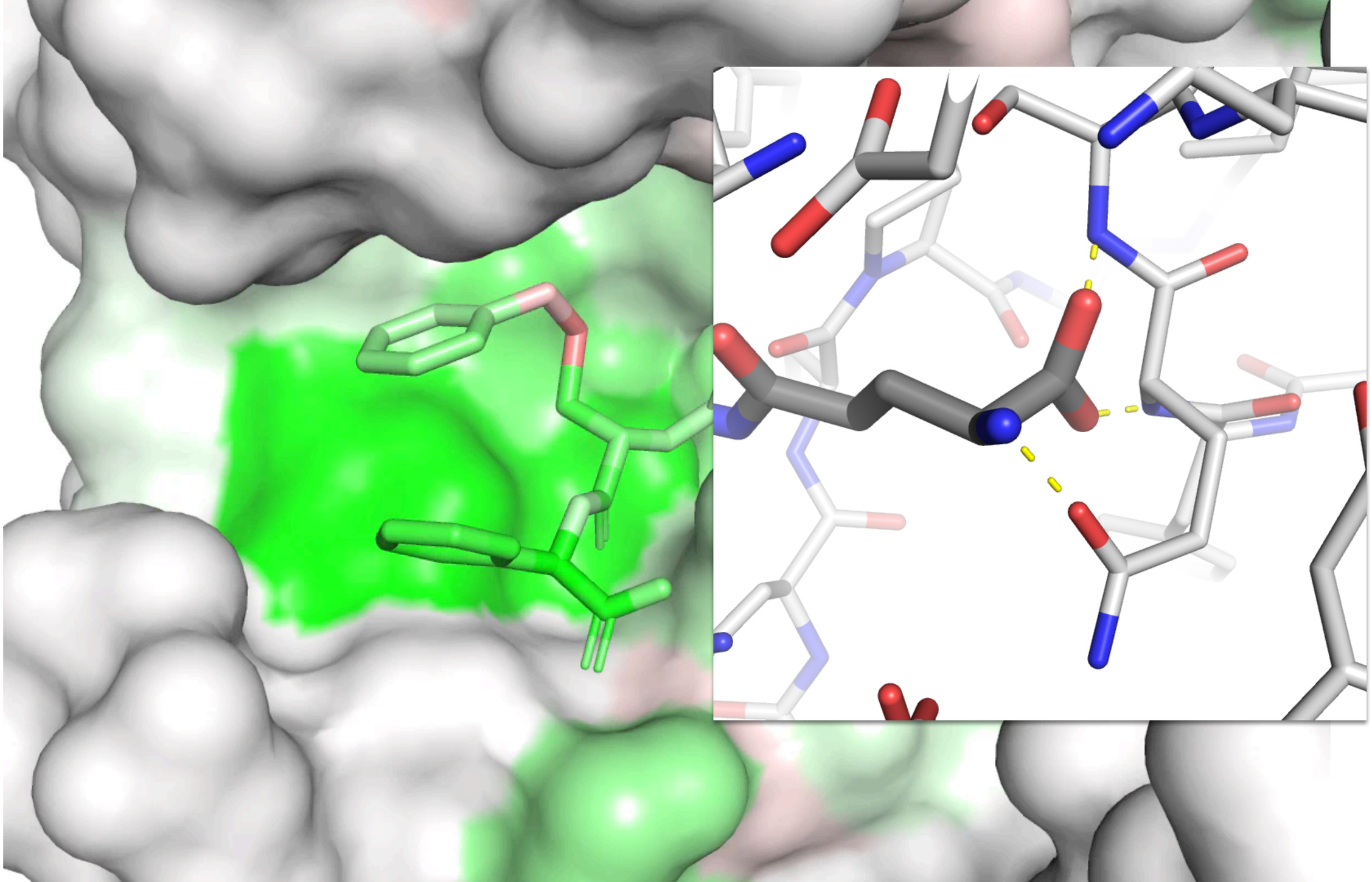
# Gradients: Beyond Scoring

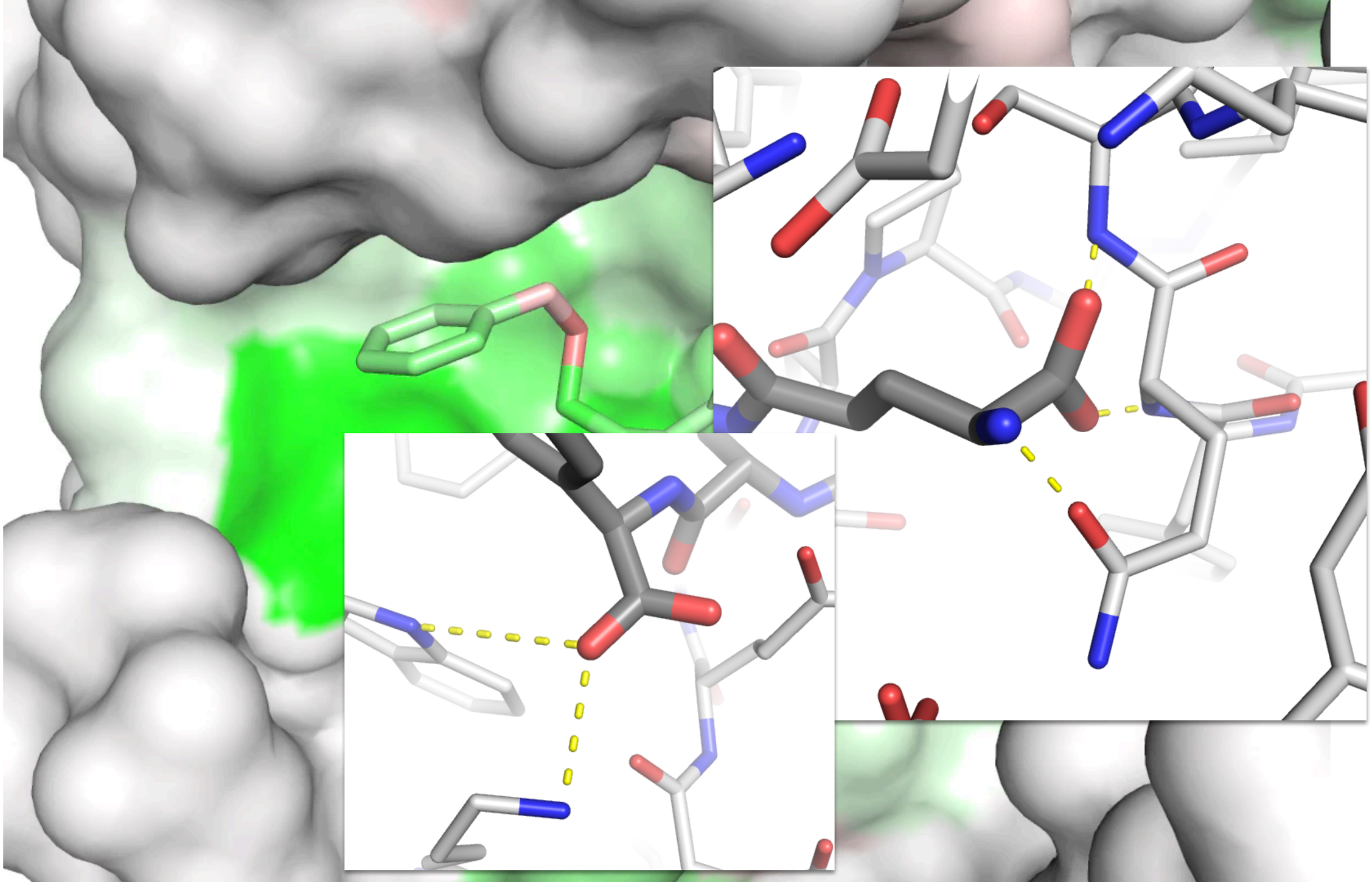



$$\frac{\partial L}{\partial A} = \sum_{i \in G_A} \frac{\partial L}{\partial G_i} \frac{\partial G_i}{\partial D} \frac{\partial D}{\partial A}$$

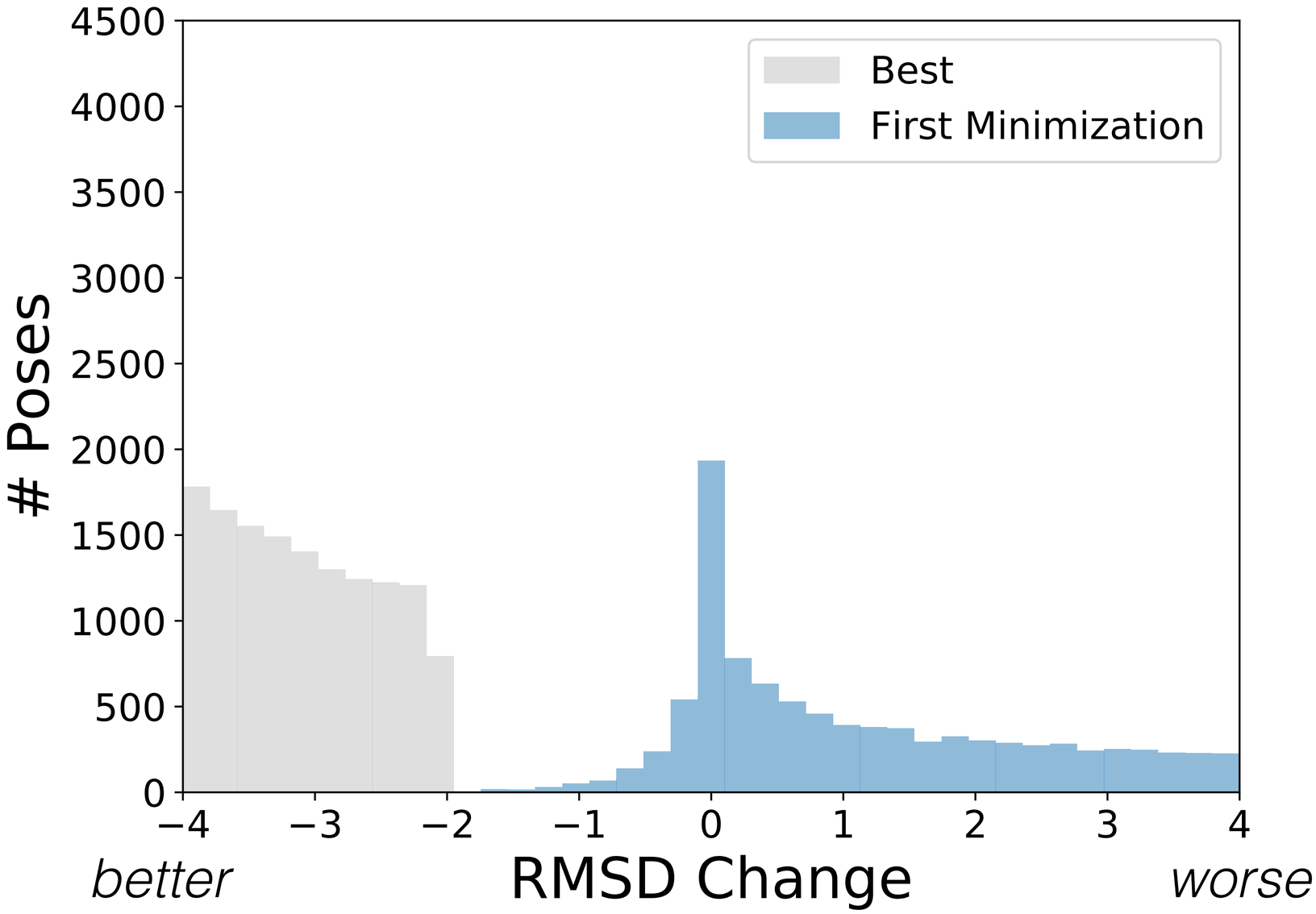




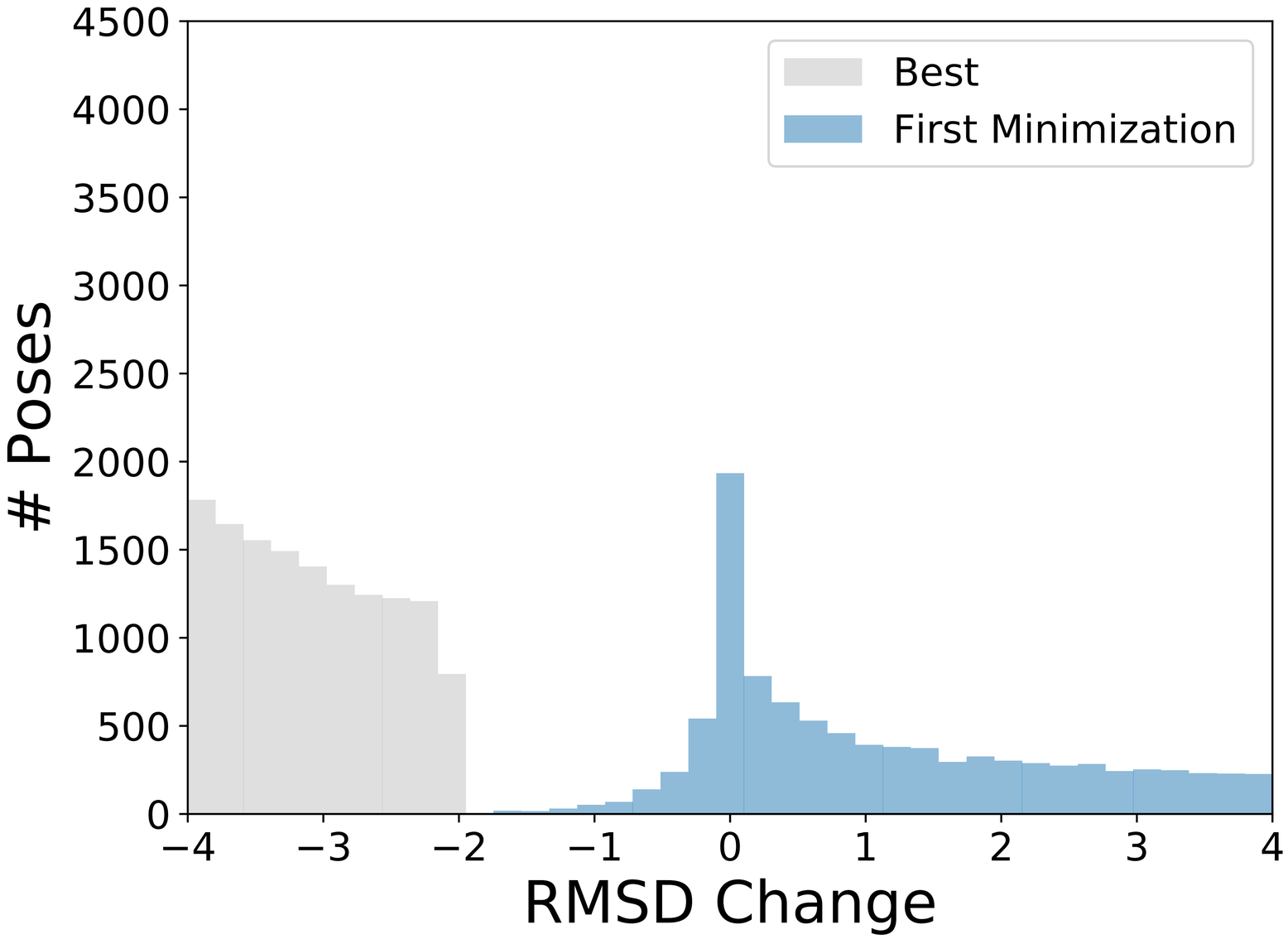




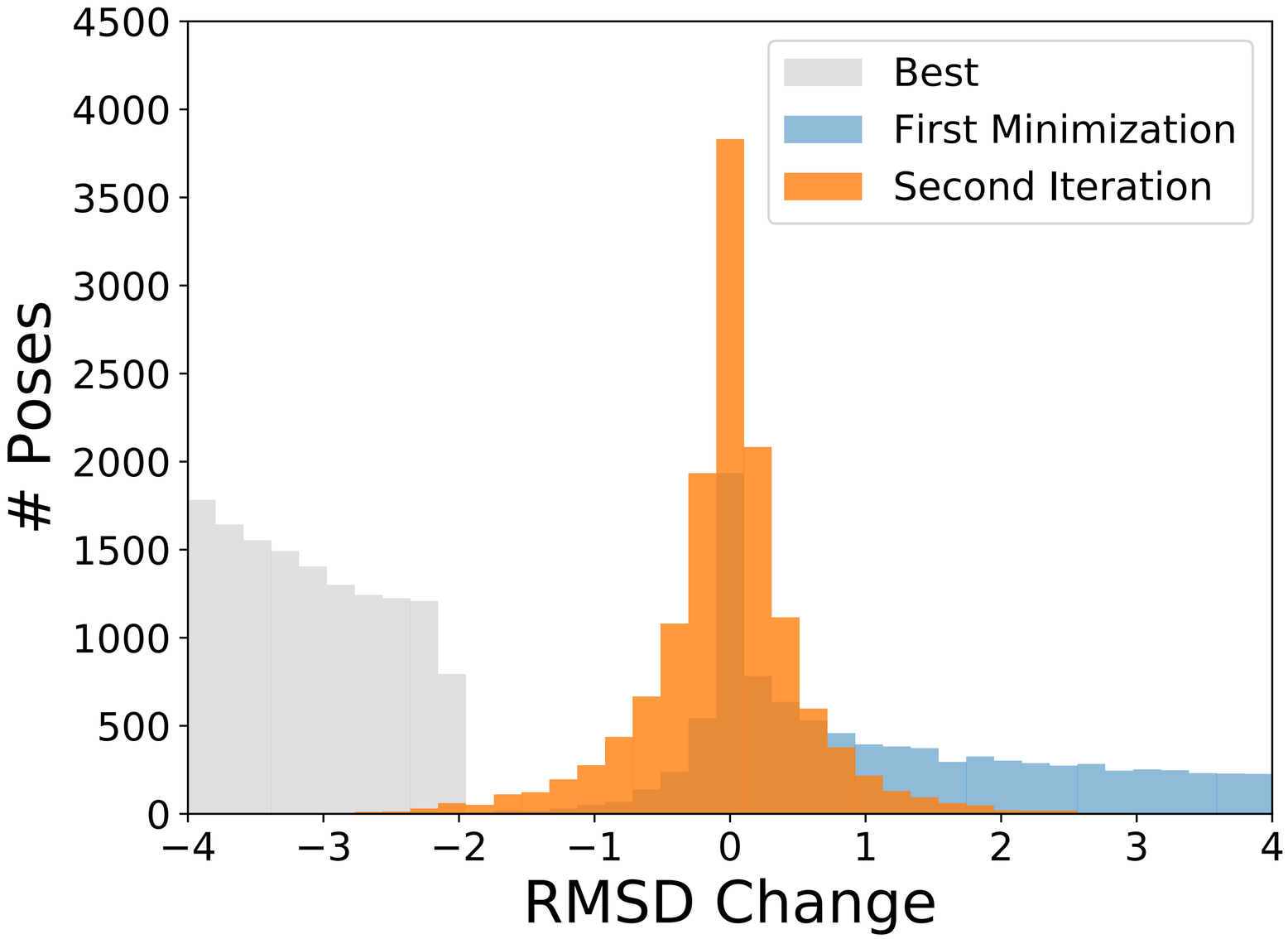
# Minimizing Low RMSD Poses



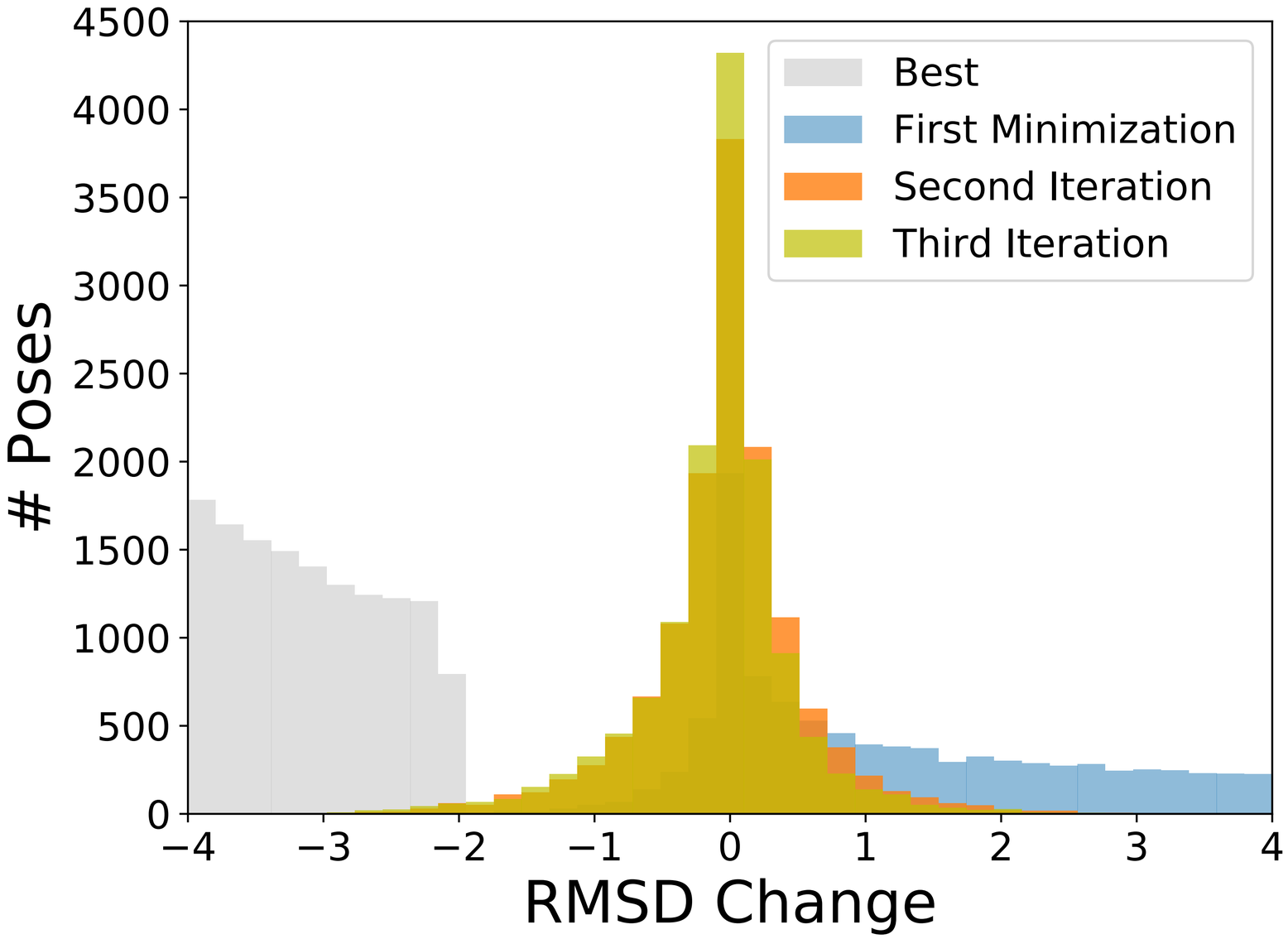
# Iterative Refinement

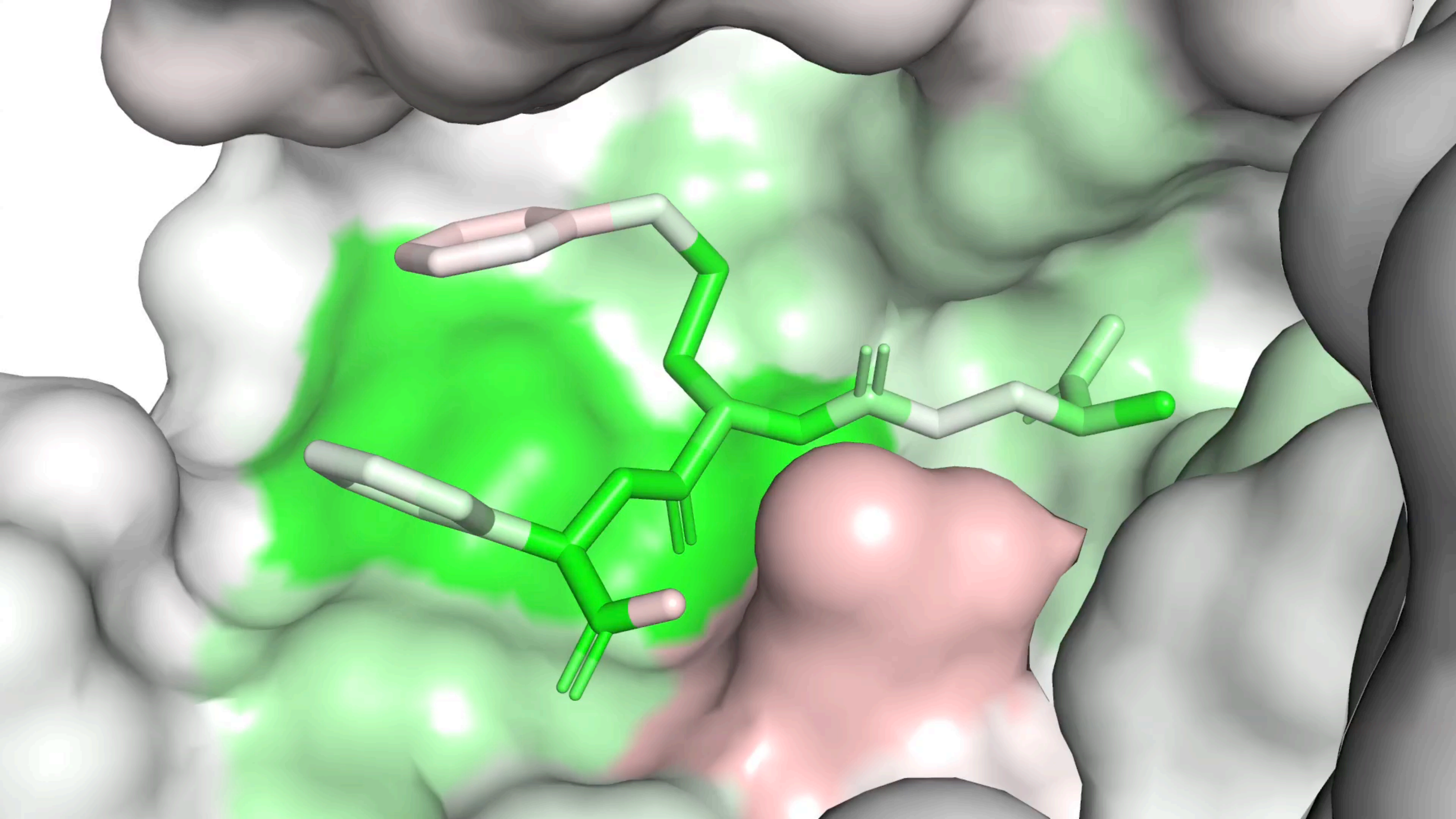


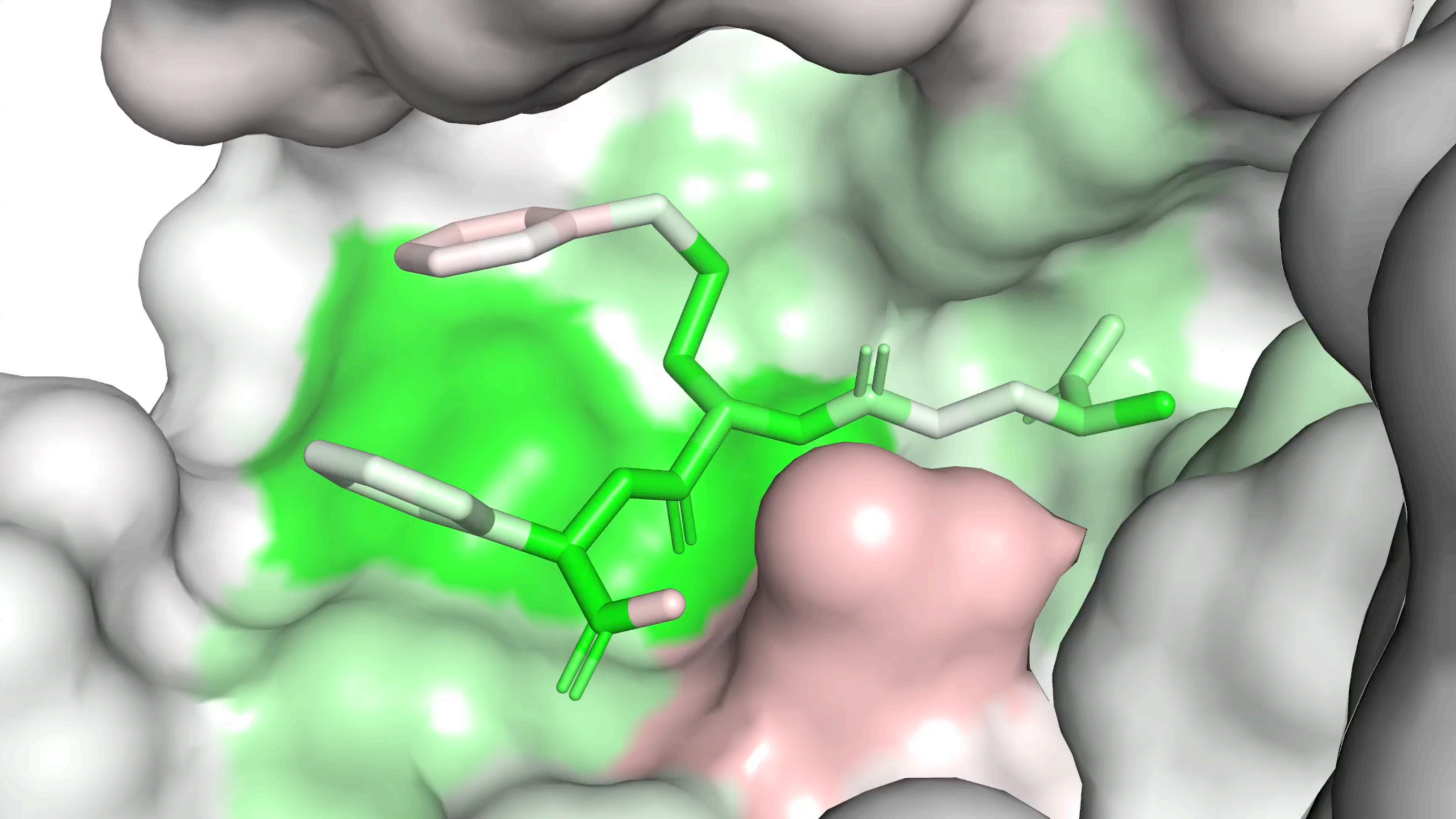
# Iterative Refinement

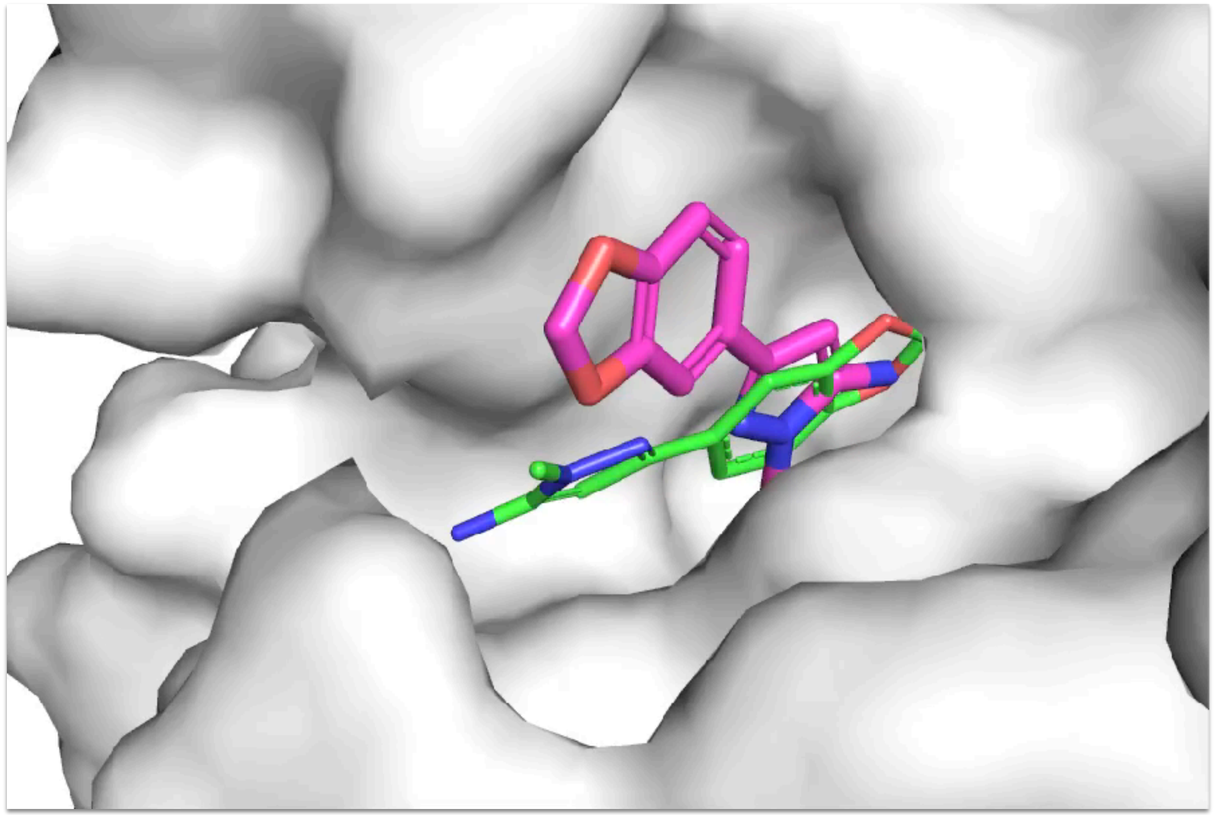


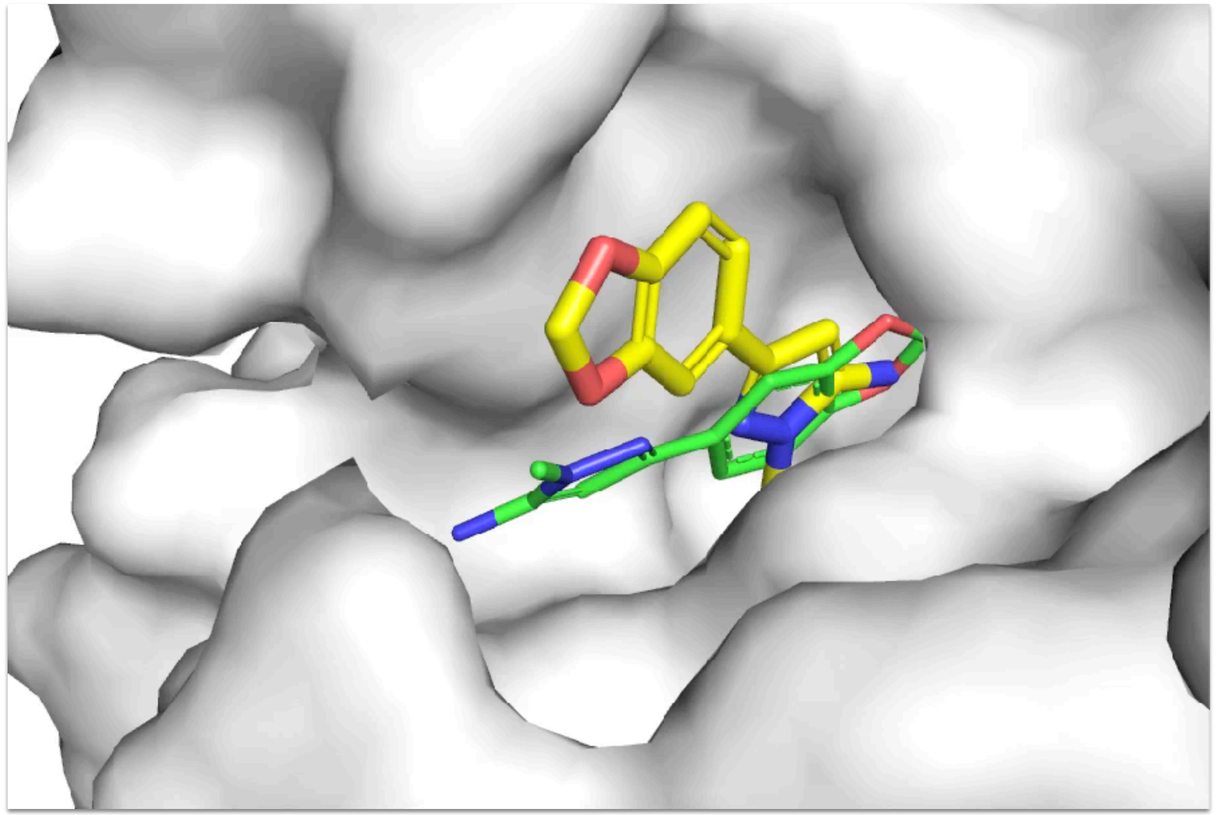
# Iterative Refinement

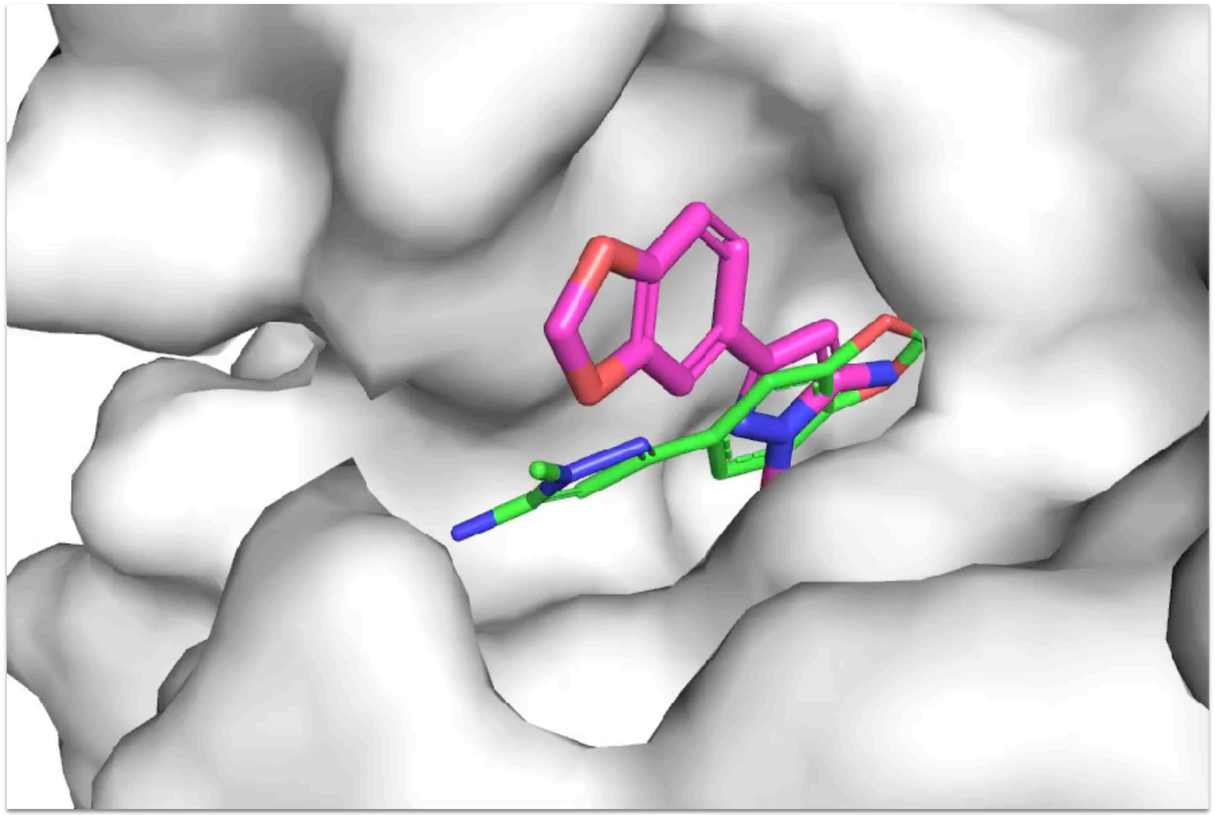


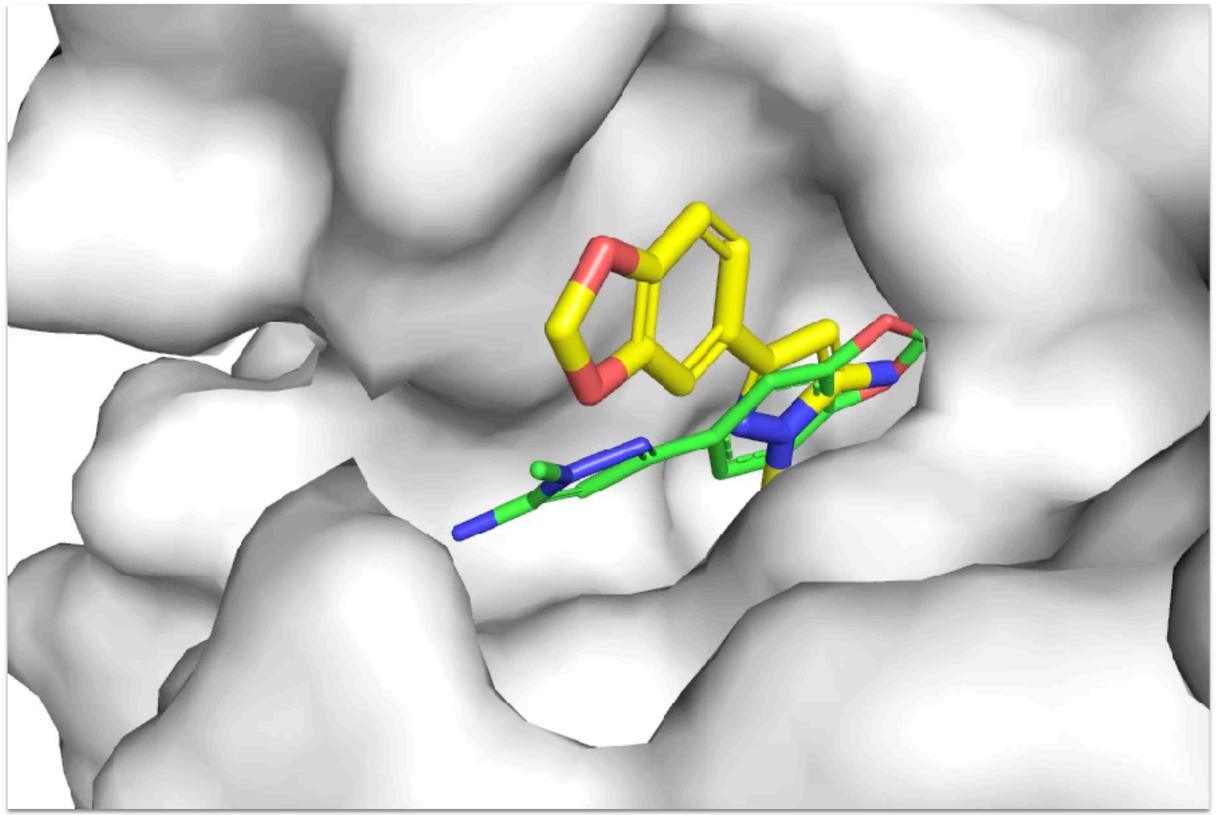




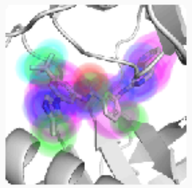








Search or jump to... / Pull requests Issues Marketplace Explore



# gnina

Repositories 4 People 9 Teams 1 Projects 0 Settings

Find a repository... Type: All Language: All

## gnina

gnina is not smina/vina

cheminformatics computational-chemistry drug-discovery

C++ ★ 35 26 Updated 11 days ago

## scripts

Jupyter Notebook ★ 1 13 BSD-3-Clause Updated on Sep 17, 2018

## models

Trained caffe models



★ 6 9 Updated on Aug 23, 2018

## Protein–Ligand Scoring with Convolutional Neural Networks

Matthew Ragoza†‡, Joshua Hochuli‡¶, Elisa Idrobo§, Jocelyn Sunseri∥, and David Ryan Koes*∥
†Department of Neuroscience, ‡Department of Computer Science, ¶Department of Biological Sciences, and ∥Department 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


**Cite this:** *J. Chem. Inf. Model.* 2017, 57, 4, 942-957
RIS Citation GO

arXiv.org > stat > arXiv:1710.07400

Statistics > Machine Learning

## Ligand Pose Optimization with Atomic Grid-Based Convolutional Neural Networks

Matthew Ragoza, Lillian Turner, David Ryan Koes

*(Submitted on 20 Oct 2017)*

Journal of Molecular Graphics and Modelling
Volume 84, September 2018, Pages 96-108

## Visualizing convolutional neural network protein-ligand scoring

Joshua Hochuli, Alec Helbling, Tamar Skaist, Matthew Ragoza, David Ryan Koes

Journal of Computer-Aided Molecular Design
pp 1–15 | Cite as

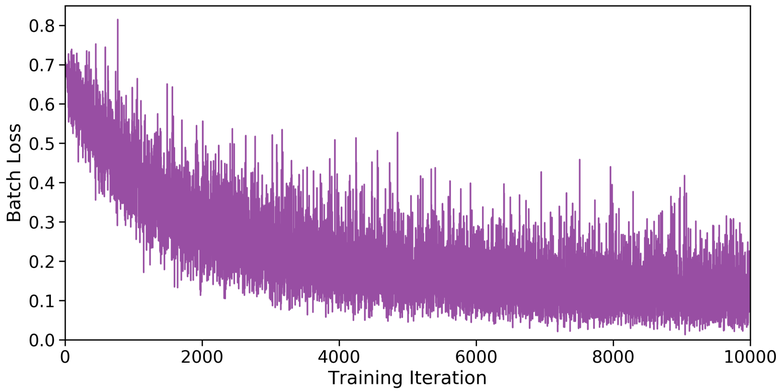
## Convolutional neural network scoring and minimization in the D3R 2017 community challenge

Authors Authors and affiliations

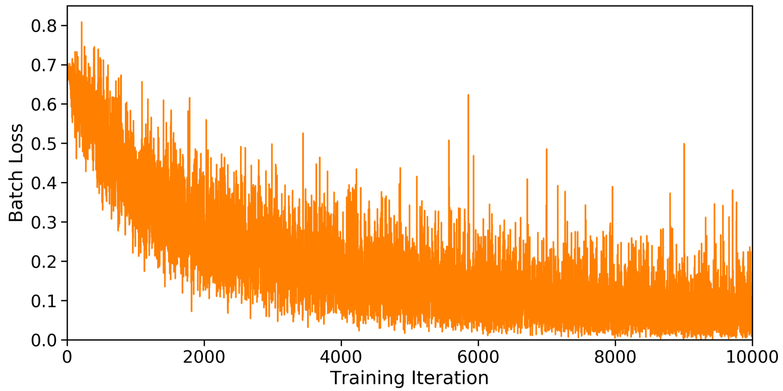
Jocelyn Sunseri, Jonathan E. King, Paul G. Francoeur, David Ryan Koes

# libmolgrid

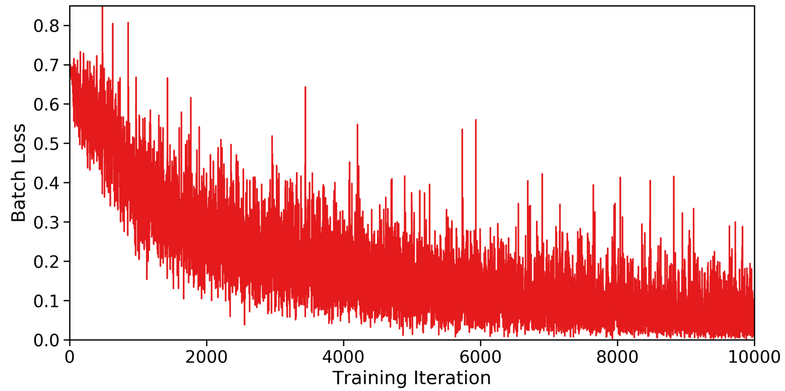
### Caffe Training



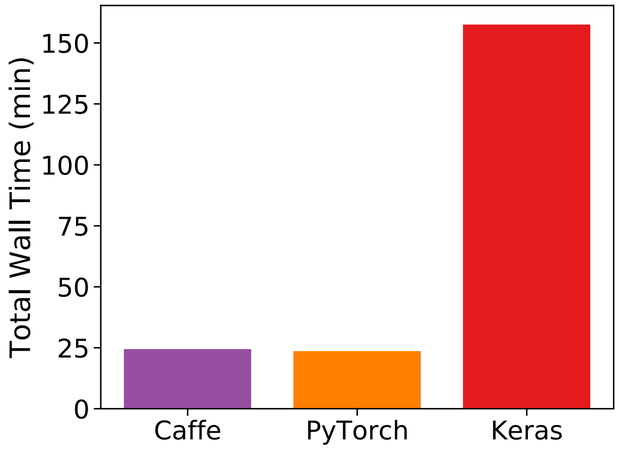

### PyTorch Training



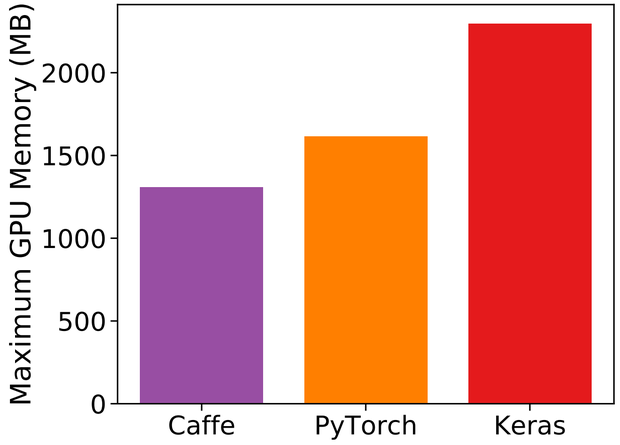

### Keras Training




### GPU Performance




### GPU Memory Utilization



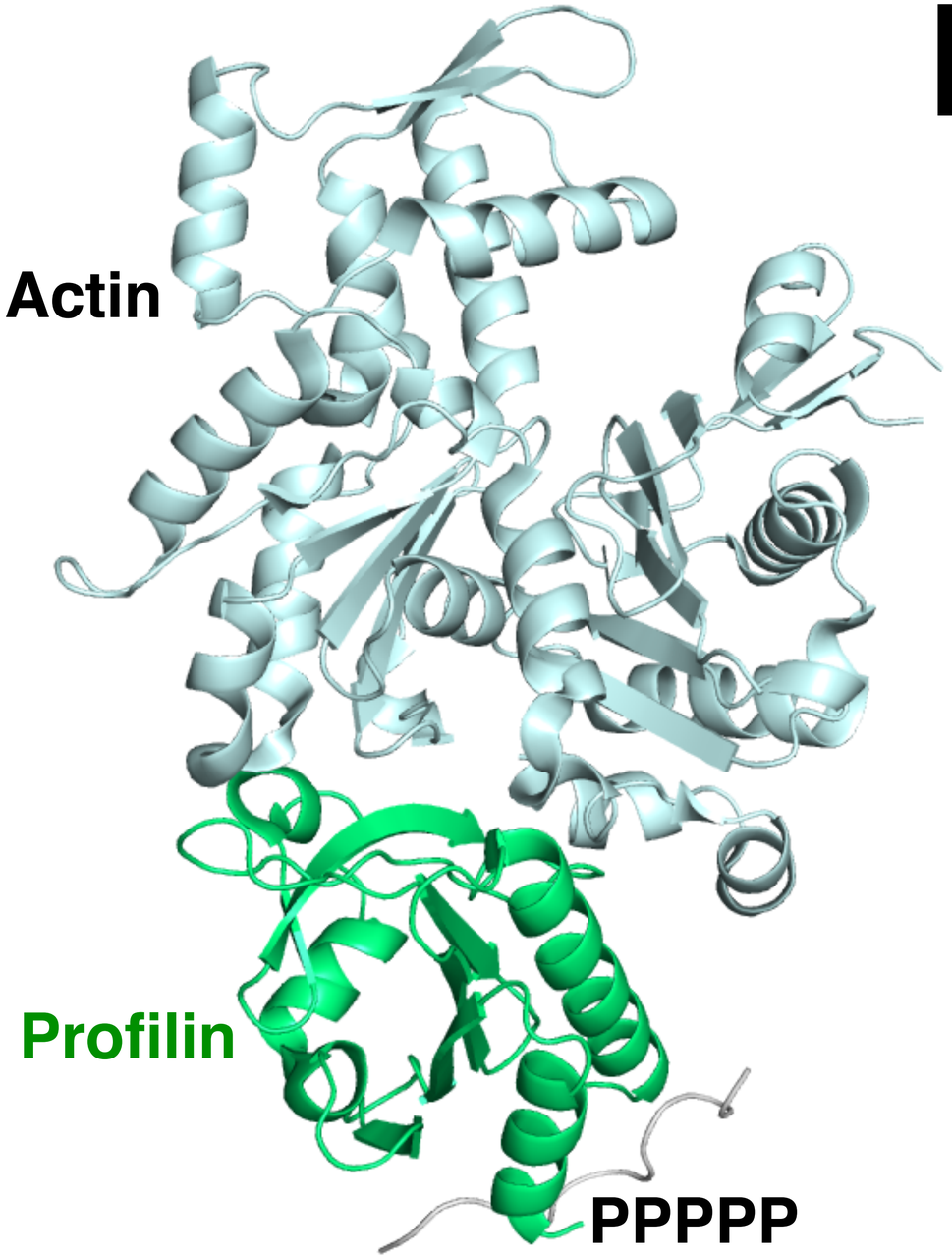

```
e = molgrid.ExampleProvider(balanced=True,shuffle=True)
e.populate('examples.txt')

gmaker = molgrid.GridMaker()

batch = e.next_batch(batch_size)
gmaker.forward(batch, input_tensor,
               random_translation=0, random_rotation=True)
```

github.com/gnina/libmolgrid

# Case Studies

# Case 1: Profilin-Actin

# Profilin



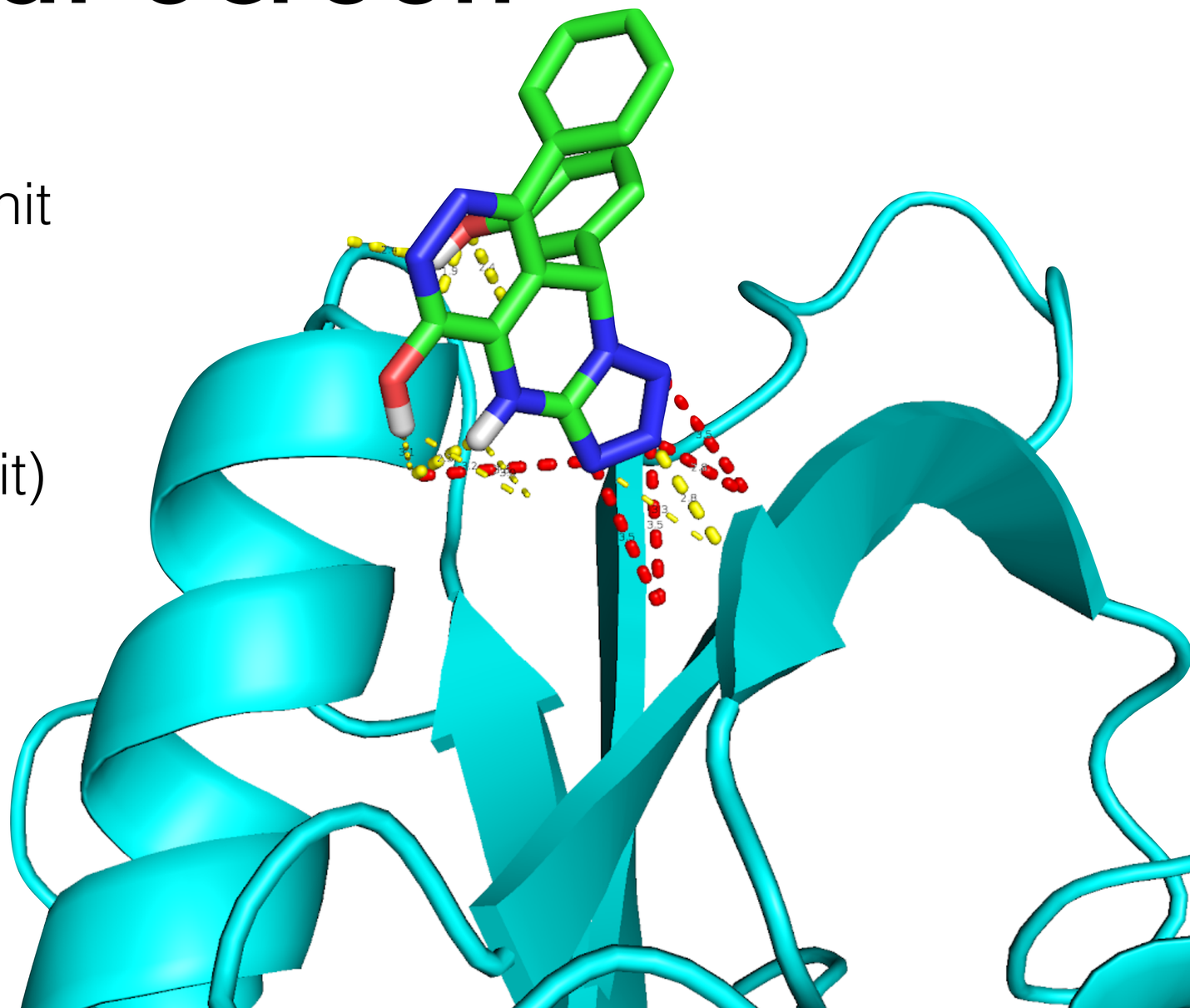

- Actin-binding protein
- Accelerates actin polymerization in presence of proline-rich proteins (e.g. formin, WASP, VASP)
- Sequesters actin otherwise

Dave Gau

Partha Roy

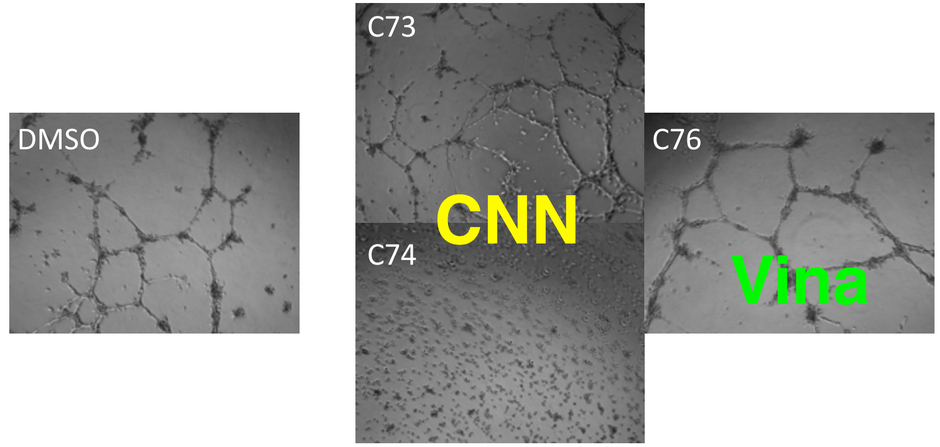
# Virtual Screen

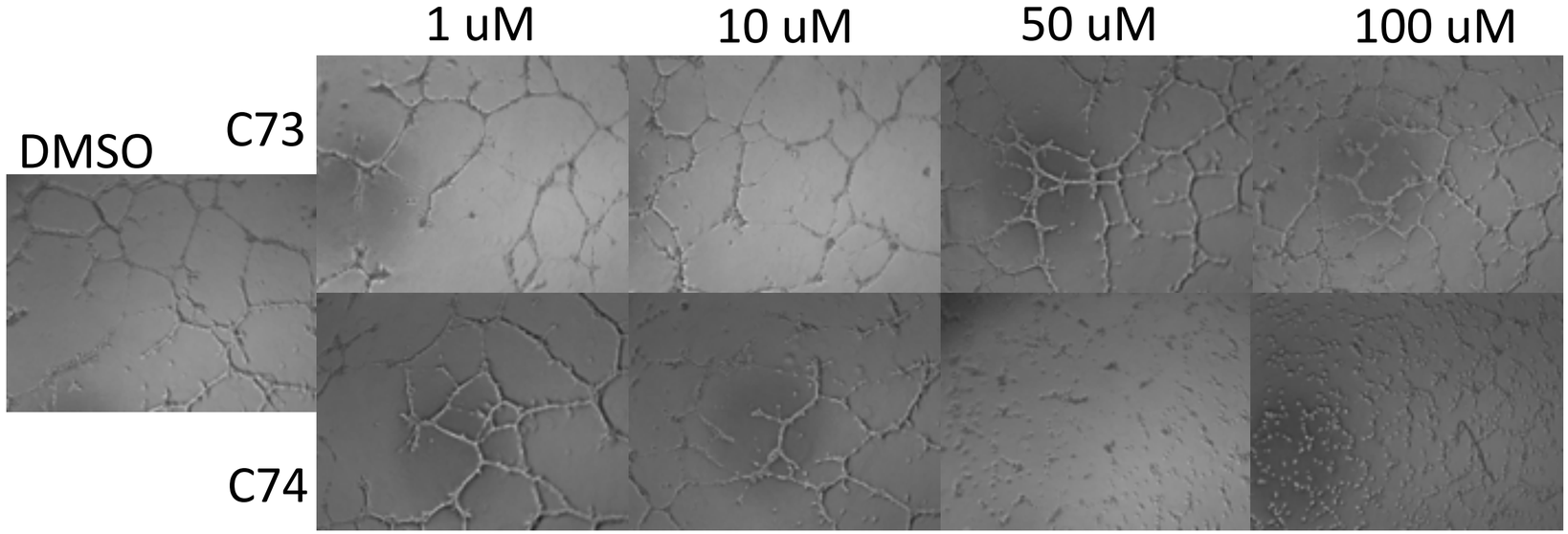
- Whole protein docking of early hit
- Identified 5 sites
- Pharmacophore screen (Pharmit)
- Ranked with Vina and CNN



# Results

57 compounds tested, **3 actives** identified

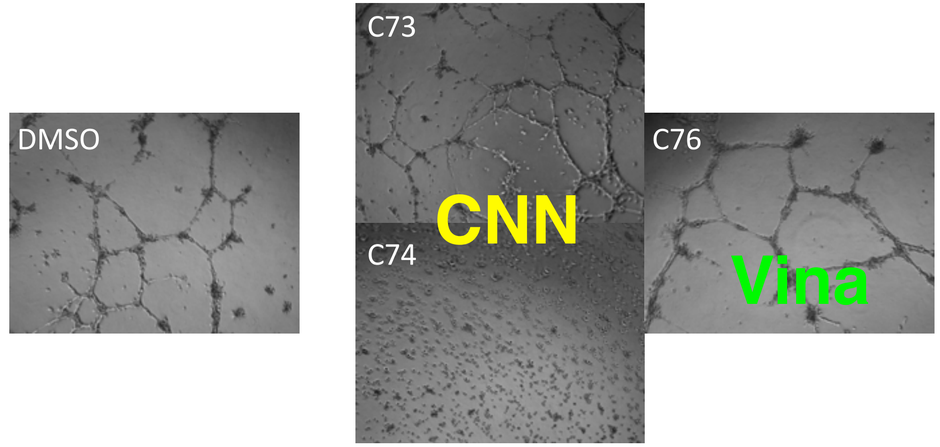


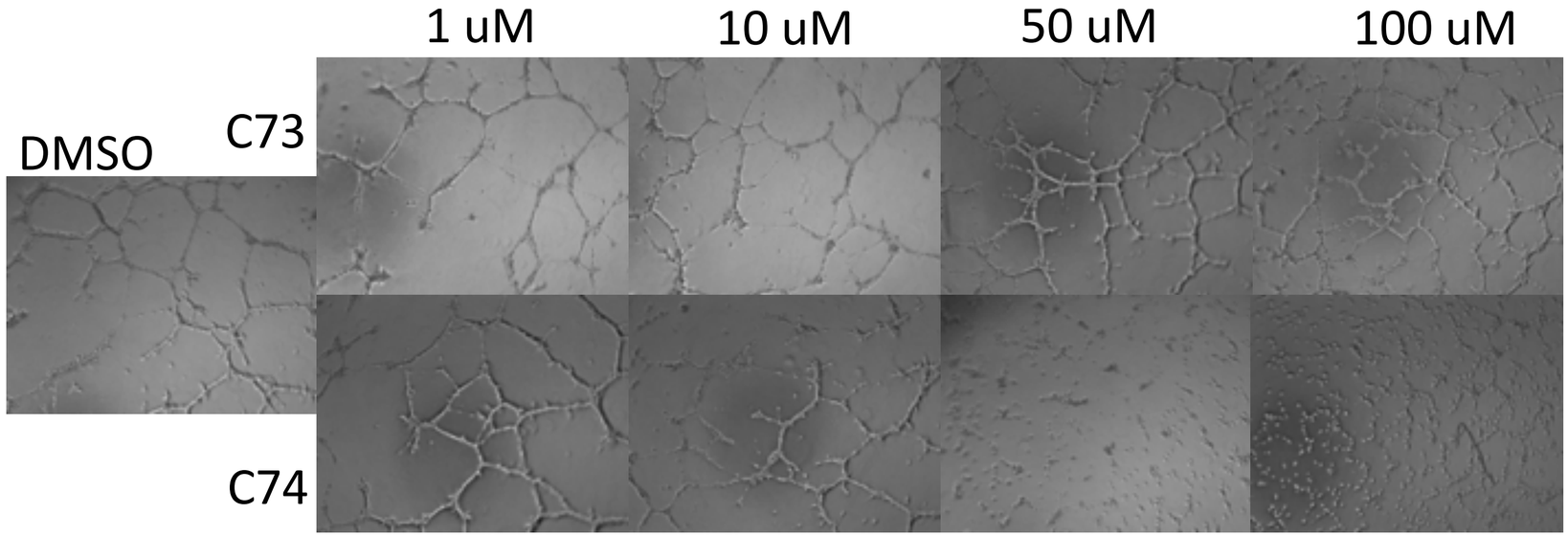



# Results

57 compounds tested, **3 actives** identified

1 (Vina) didn't work in cells



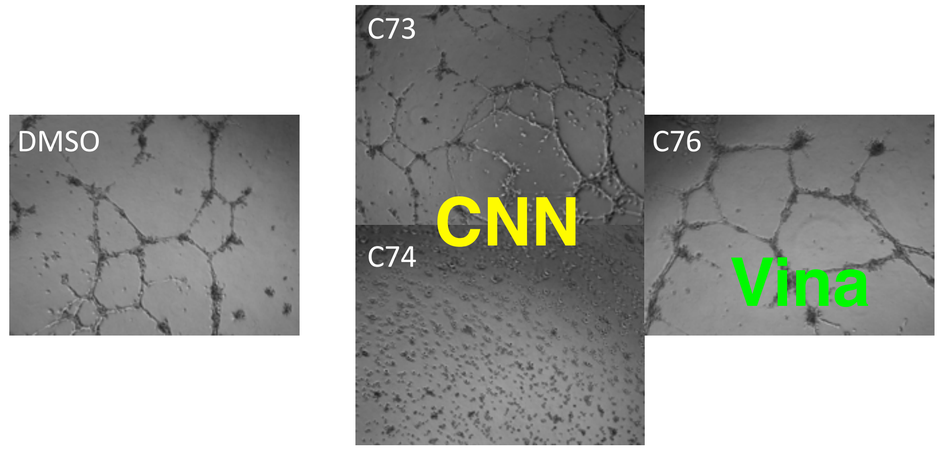



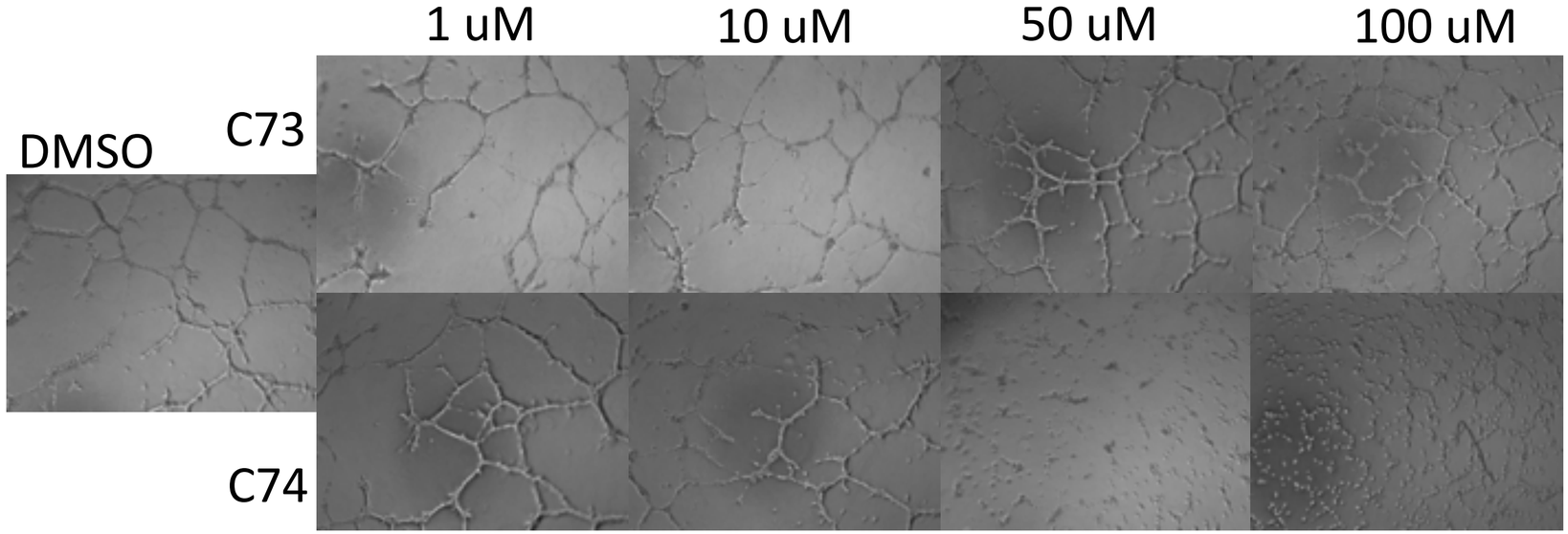
# Results

57 compounds tested, **3 actives** identified

1 (Vina) didn't work in cells

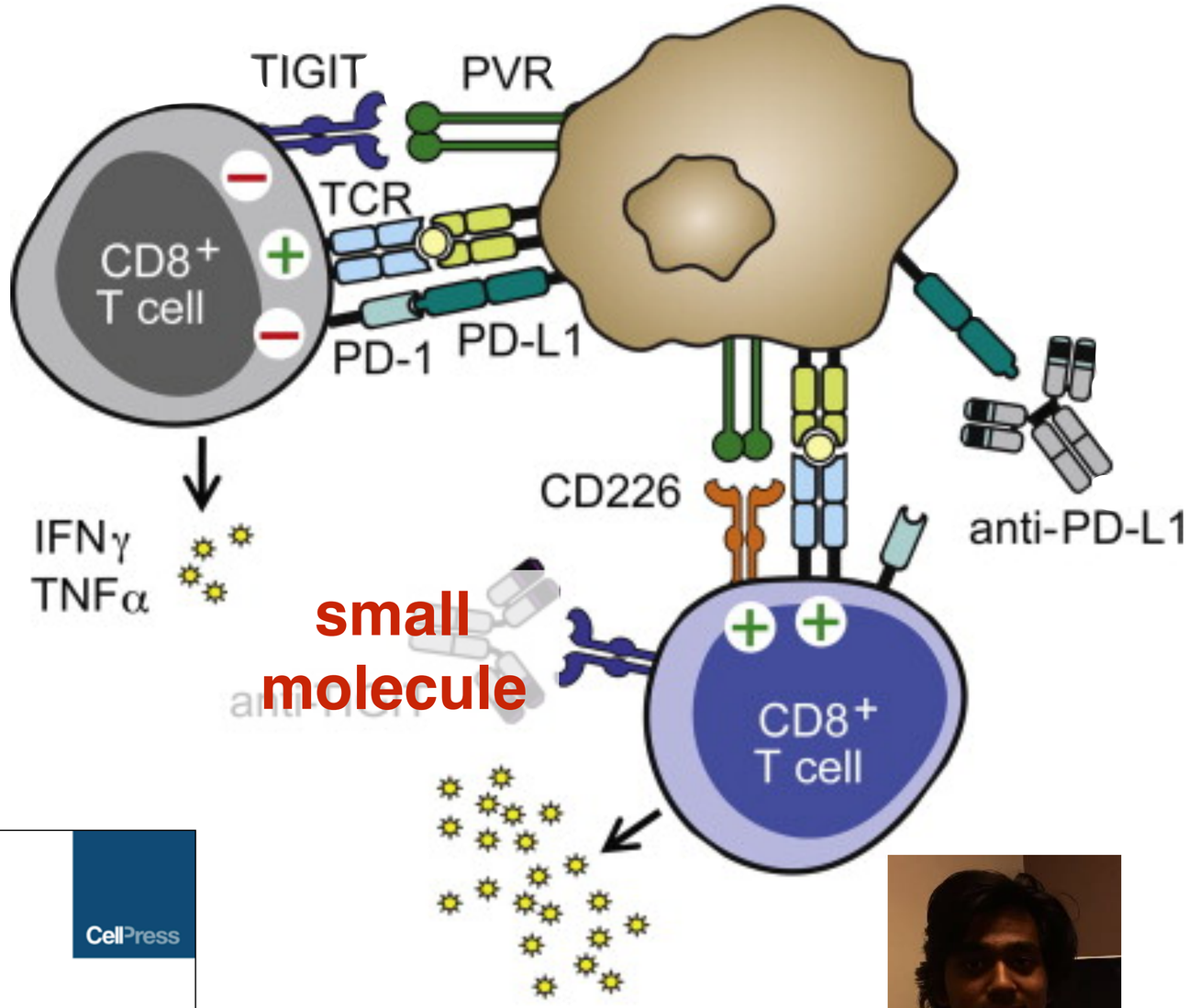
All predicted to bind to different sites



# Case 2: TIGIT

# Can we block TIGIT/ PVR interaction with a small molecule?




Cancer Cell
Article

CellPress

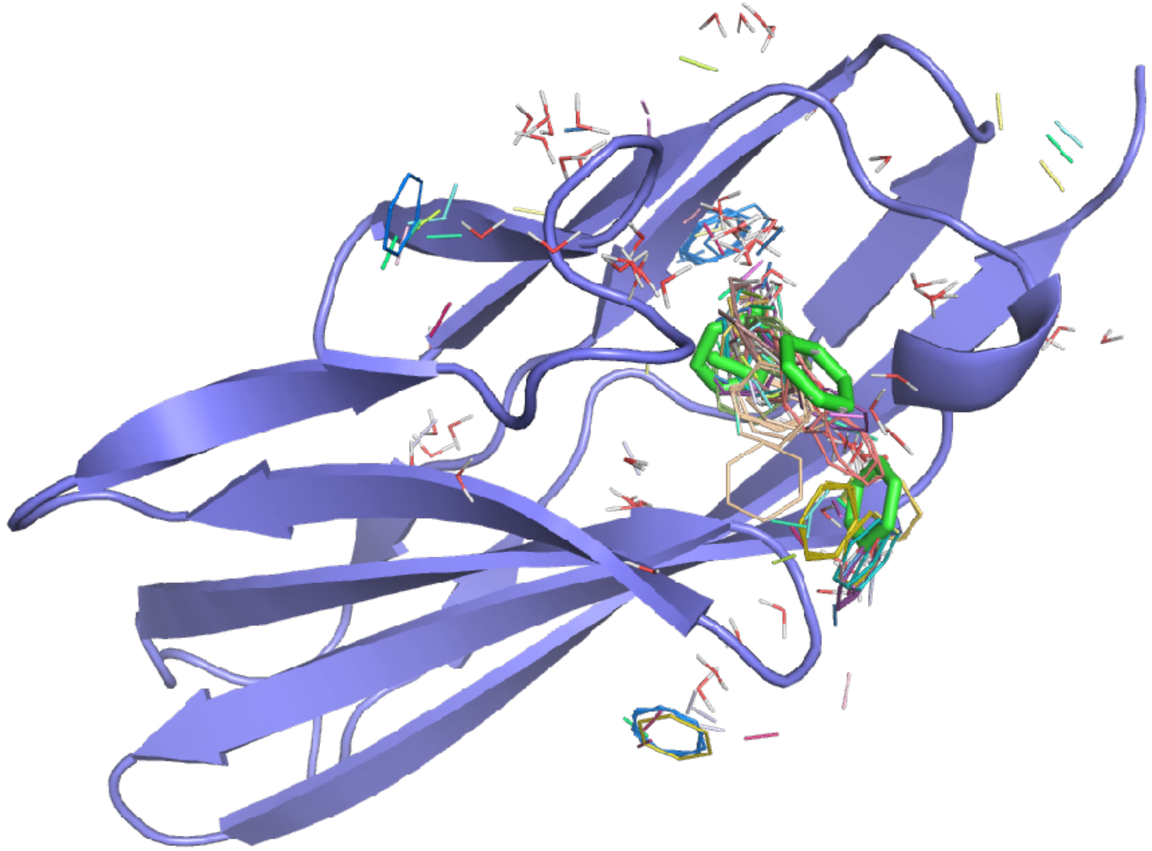
## The Immunoreceptor TIGIT Regulates Antitumor and Antiviral CD8+ T Cell Effector Function

Robert J. Johnston,[1] Laetitia Comps-Agrar,[2] Jason Hackney,[3] Xin Yu,[1] Mahrukh Huseni,[4] Yagai Yang,[5] Summer Park,[6] Vincent Javinal,[5] Henry Chiu,[7] Bryan Irving,[1] Dan L. Eaton,[2] and Jane L. Grogan[1,*]

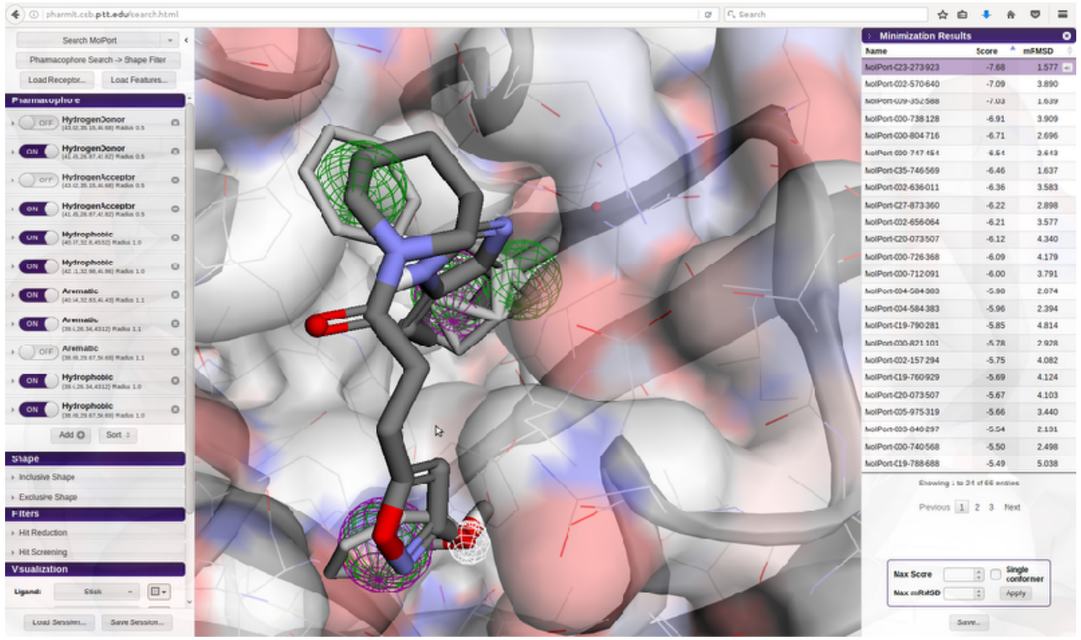


Gibran Biswas

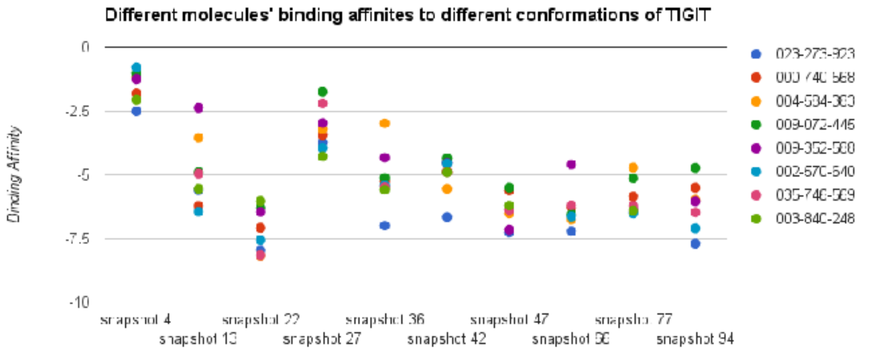
# Does anything bind to this pocket?



Fragment Docking

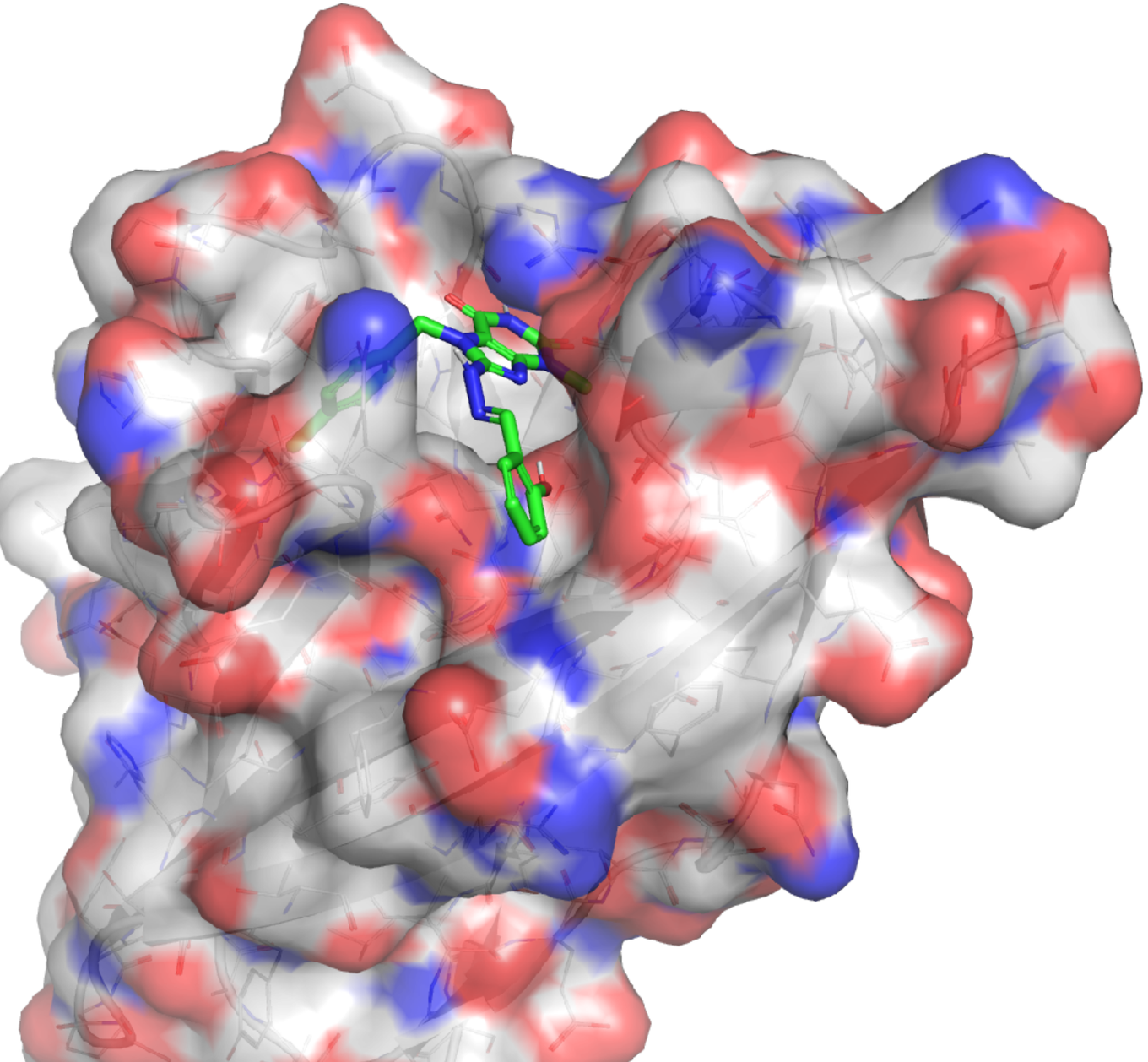



Pharmacophore Search




Consensus Scoring (CNN and Vina)

# Screening



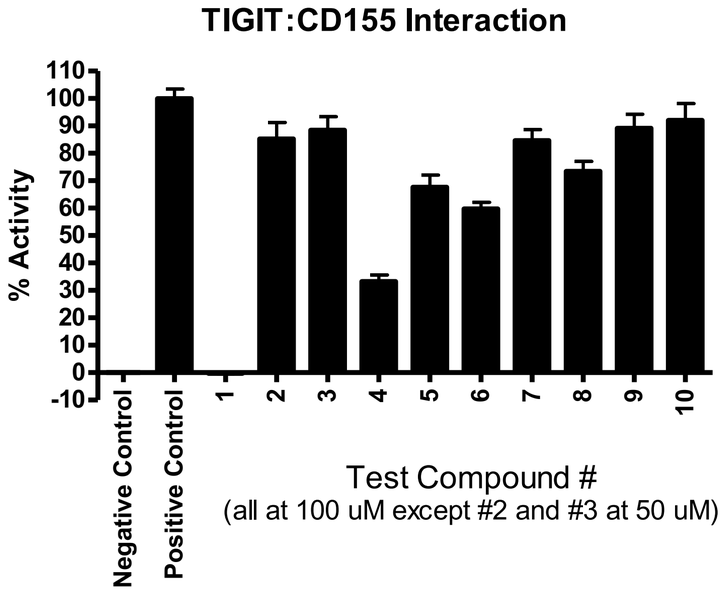
**10 diverse compounds selected for screening**

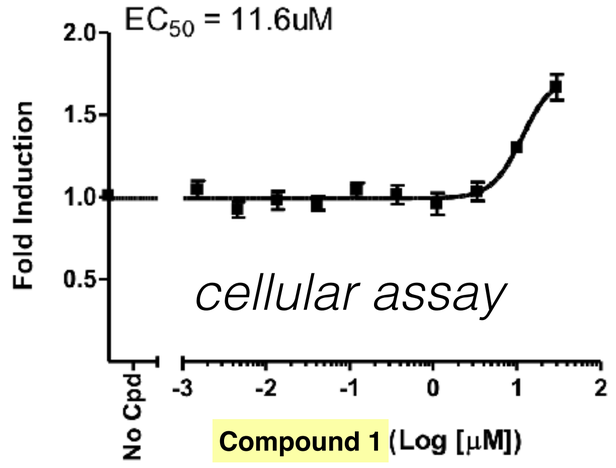
- **top ranked by Vina**
- **top ranked by CNN**

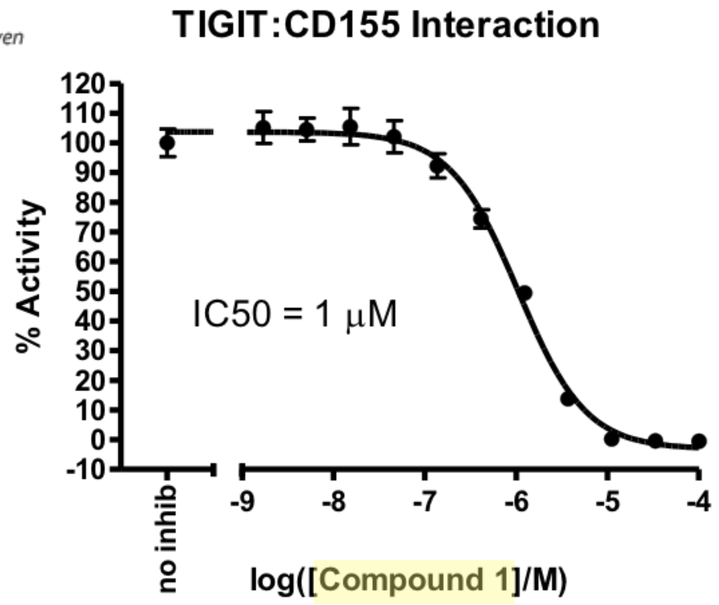
| Name | CNN Affinity | CNN Score | Vina |
|---|---|---|---|
| Compound 1 | 7.69807 | 0.994763 | 85.95 |
| Compound 2 | 5.57909 | 0.0180277 | -8.12632 |
| Compound 3 | 6.73692 | 0.0624742 | -9.81935 |
| Compound 4 | 6.87897 | 0.953488 | -3.81378 |
| Compound 5 | 6.32813 | 0.209807 | -8.60293 |
| Compound 6 | 5.689 | 0.0437 | -8.991 |
| Compound 7 | 4.368 | 0.022 | -9.34722 |
| Compound 8 | 4.81 | 0.072 | -6.81787 |
| Compound 9 | 5.22 | 0.032 | -6.264 |
| Compound 10 | 6.67 | 0.361 | 6.1053 |

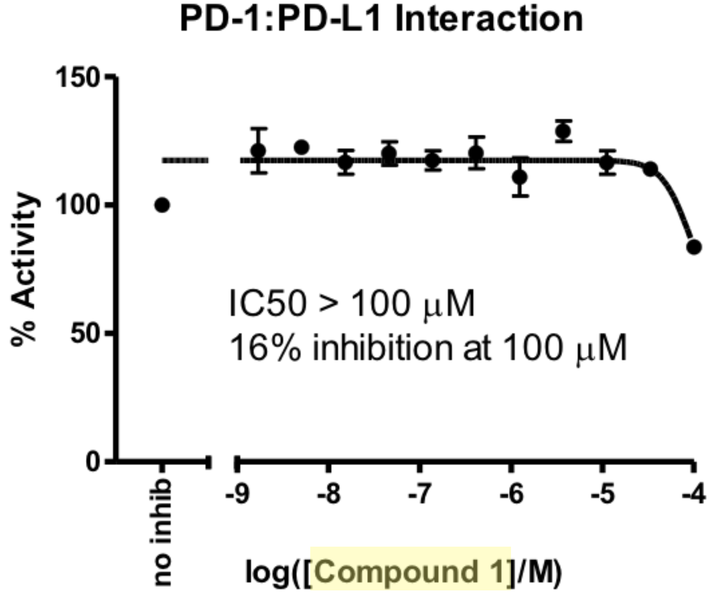
# Results

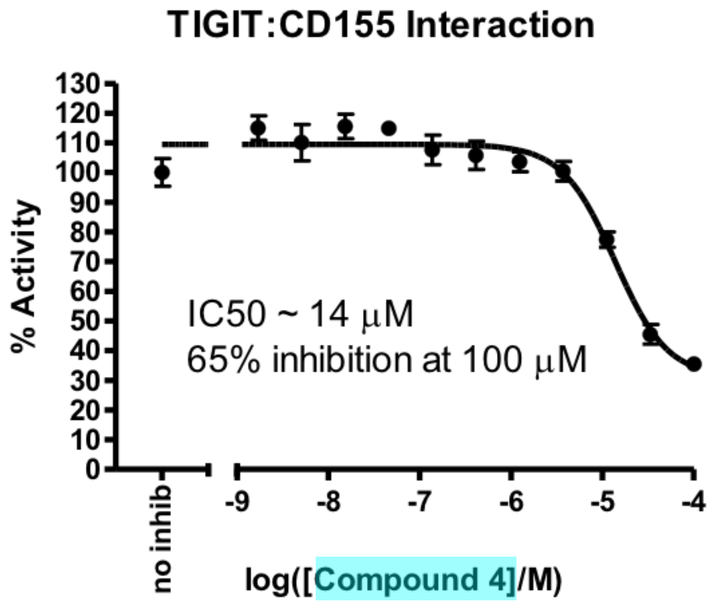
BPS Bioscience
Scientist Founded, Scientist Driven

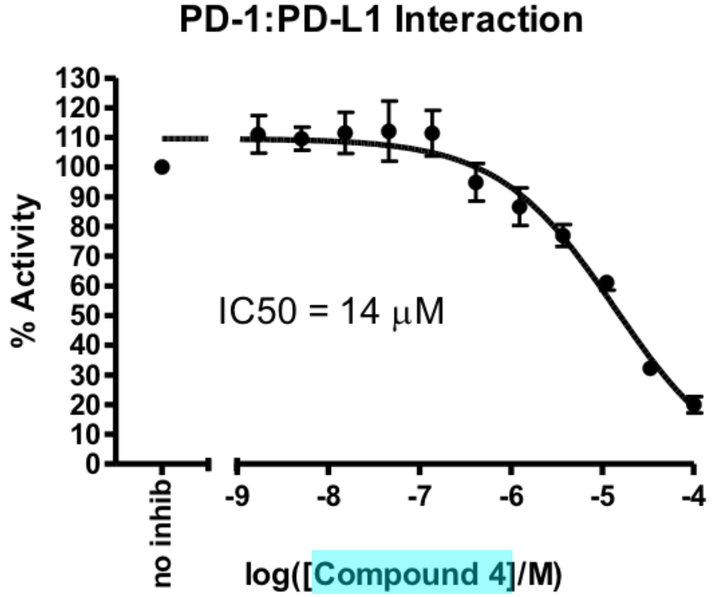


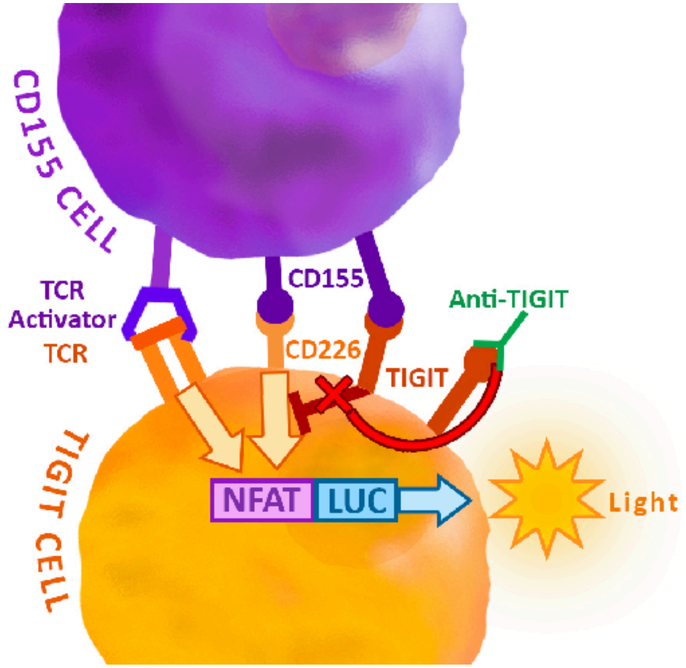



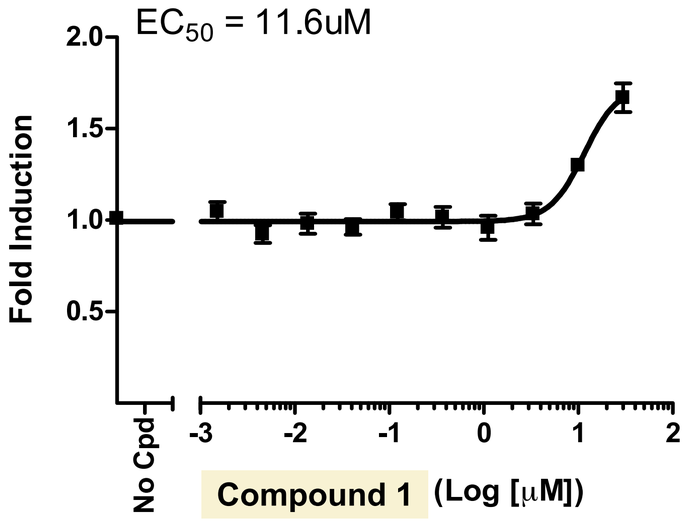



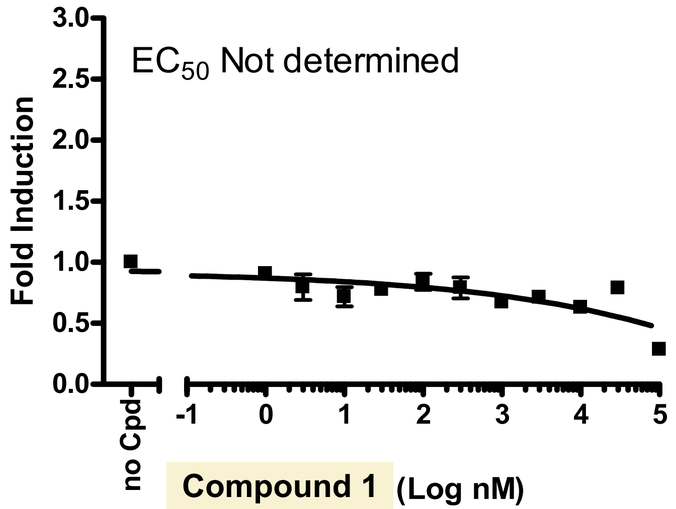



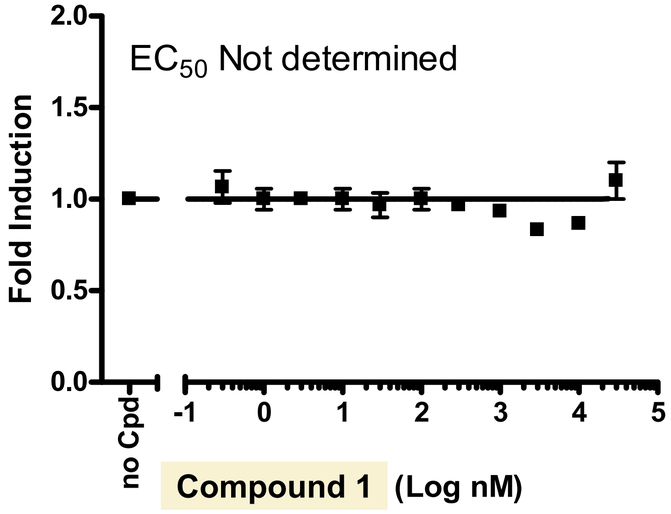
# But...

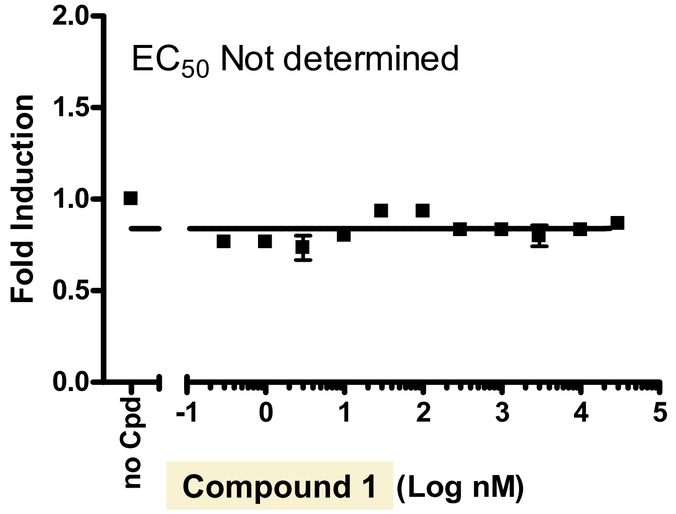



The first trial was promising, but the maximum does was limited by DMSO concentration. Future trials at appropriate dosages showed no response.

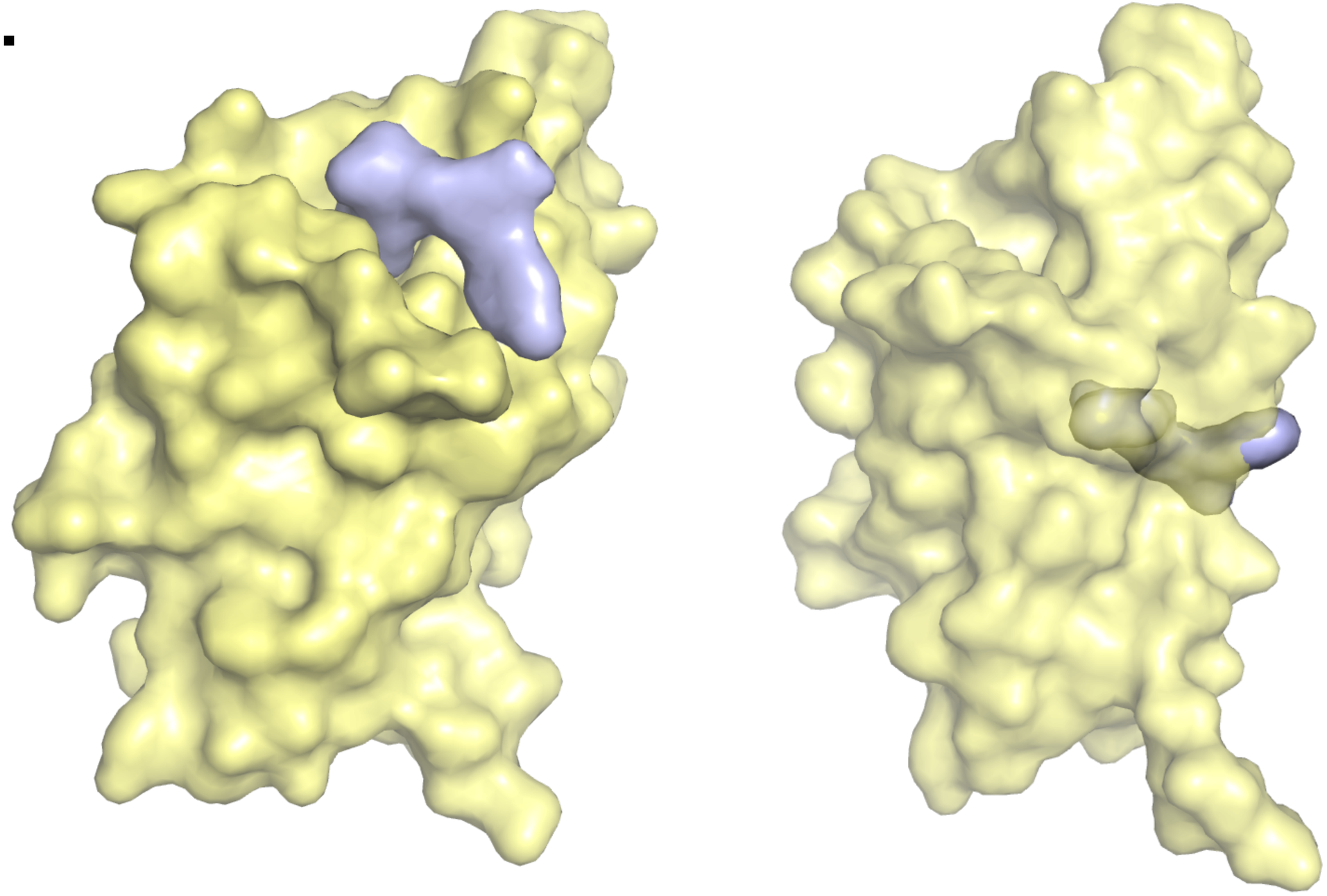


# But...

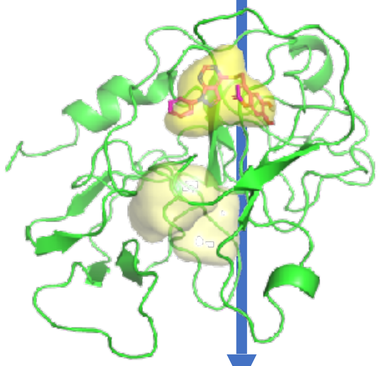


# Case 3: Mystery Target

# Approach

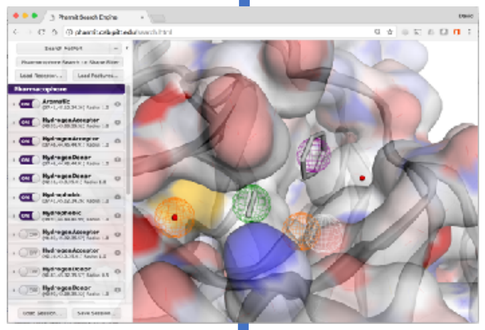
unbiased MD simulations

pocket identification (mdpocket)



select receptor using pocket volume/druggability

pharmacophore query from fragment docking



minimization w/Vina

filters:
energy
minimized RMSD
**Lipinski**
**PAINS**

HBD <= 5
HBA <= 10
MW <= 500
logP <= 5

hits

consensus scoring

CNN | Vina | GlideXP

best min-rank | best max-rank

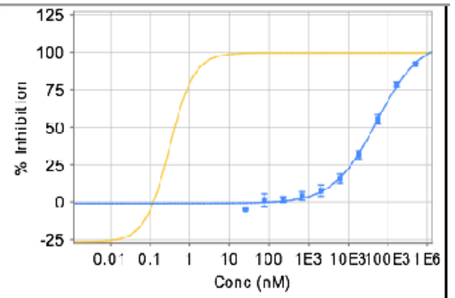
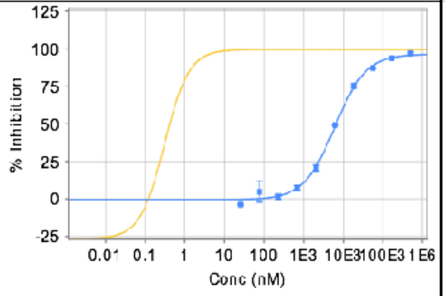
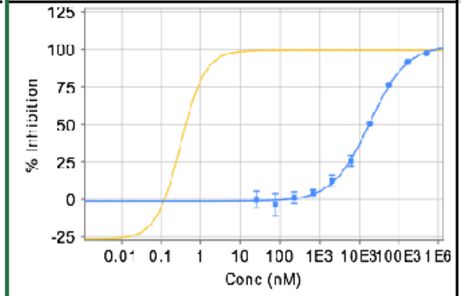
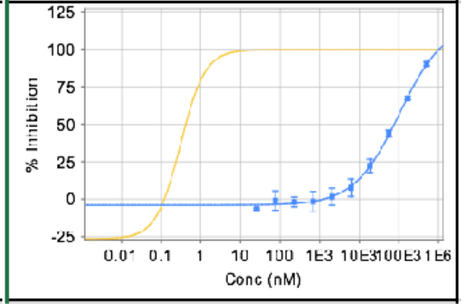
*select ~20 final hits per pocket*

sort by sum-rank

diversity filter
*at most 2 cmpds w/ Tanimoto > 0.7*

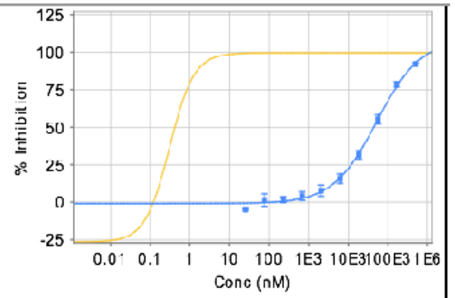
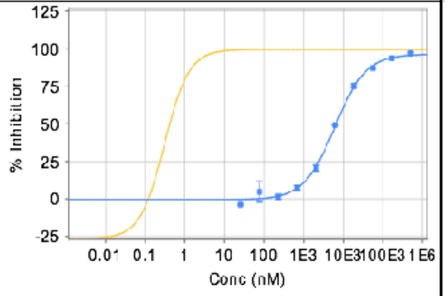
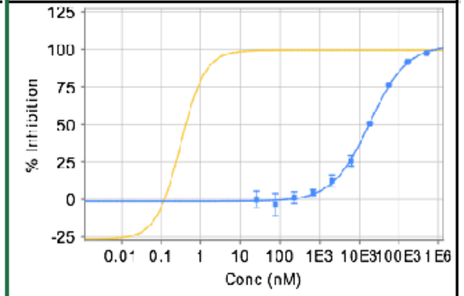
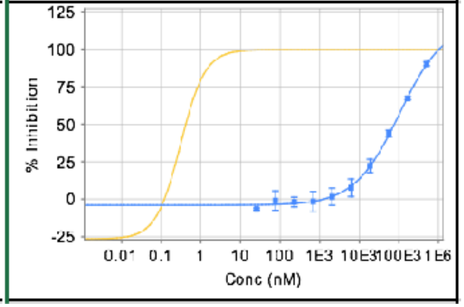
# Screening Hits

- 50 compounds tested
- designed against 3 putative allosteric pockets
- 4 hits (3 from P2, 1 from P4)
- P2 was potentially a very desirable pocket to hit for target-specific reasons

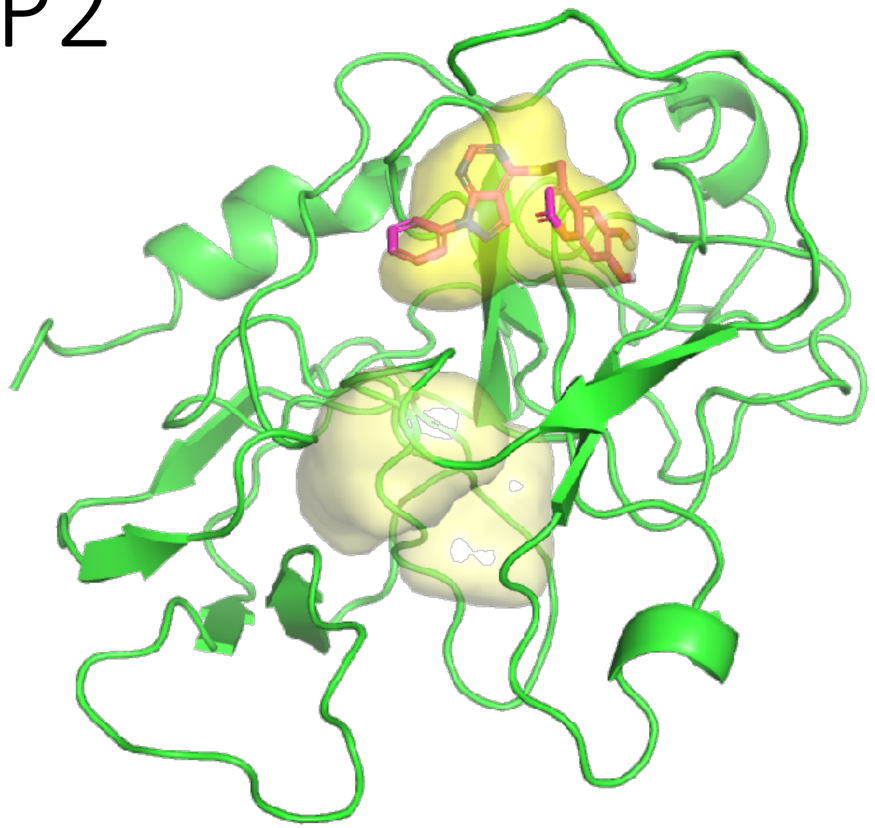
| | | |
|---|---|---|
| P2 | 51800 |  |
| P2 | 6080 |  |
| P4 | 18100 |  |
| P2 | 102000 |  |

# Screening Hits

- 50 compounds tested
- designed against 3 putative allosteric pockets
- 4 hits (3 from P2, 1 from P4)
- P2 was potentially a very desirable pocket to hit for target-specific reasons

| | | |
|---|---|---|
| P2 | 51800 |  |
| P2 | 6080 |  |
| P4 | 18100 |  |
| P2 | 102000 |  |

# P2



| | | MolPort ID | CNNScore | -Vina | -GlideXP | cnn_rank | vina_rank | glide_rank | sum_rank |
|---|---|---|---|---|---|---|---|---|---|
| 6μM | 0 | | 7.35853 | 9.73614 | 6.68 | 3.0 | 5.0 | 7.5 | 15.5 |
| | 1 | | 6.36562 | 9.25733 | 6.56 | 4.0 | 7.0 | 9.0 | 20.0 |
| 102μM | 2 | | 6.14087 | 8.73366 | 6.94 | 6.0 | 12.0 | 5.0 | 23.0 |
| | 3 | | 5.66284 | 9.40722 | 6.76 | 13.0 | 6.0 | 6.0 | 25.0 |
| | 4 | | 6.02615 | 8.82225 | 6.21 | 7.0 | 8.0 | 13.0 | 28.0 |
| | 5 | | 5.32736 | 10.18550 | 6.51 | 14.0 | 4.0 | 10.0 | 28.0 |
| | 6 | | 5.80743 | 7.26154 | 8.24 | 10.0 | 18.0 | 1.0 | 29.0 |
| | 7 | | 5.93622 | 8.75104 | 6.38 | 8.0 | 11.0 | 11.0 | 30.0 |
| 52μM | 8 | | 6.20744 | 6.18992 | 6.68 | 5.0 | 19.0 | 7.5 | 31.5 |
| | 9 | | 5.06070 | 10.32670 | 5.20 | 15.0 | 1.0 | 16.0 | 32.0 |
| | 10 | | 4.77866 | 10.32050 | 5.32 | 16.0 | 2.0 | 15.0 | 33.0 |
| | 11 | | 5.72094 | 8.76388 | 6.24 | 11.0 | 10.0 | 12.0 | 33.0 |
| | 12 | | 7.55155 | 8.26183 | 4.26 | 1.0 | 15.0 | 18.0 | 34.0 |
| | 13 | | 3.84854 | 10.19640 | 5.99 | 18.0 | 3.0 | 14.0 | 35.0 |
| | 14 | | 4.29961 | 7.43586 | 8.18 | 17.0 | 17.0 | 3.0 | 37.0 |
| | 15 | | 3.34137 | 7.79760 | 8.21 | 19.0 | 16.0 | 2.0 | 37.0 |
| | 16 | | 2.77562 | 8.26298 | 8.17 | 20.0 | 14.0 | 4.0 | 38.0 |
| | 17 | | 5.70688 | 8.80796 | 4.77 | 12.0 | 9.0 | 17.0 | 38.0 |
| | 18 | | 5.92500 | 8.53183 | 4.25 | 9.0 | 13.0 | 19.0 | 41.0 |
| | 19 | | 7.52880 | 6.05508 | 3.26 | 2.0 | 20.0 | 20.0 | 42.0 |

# But...

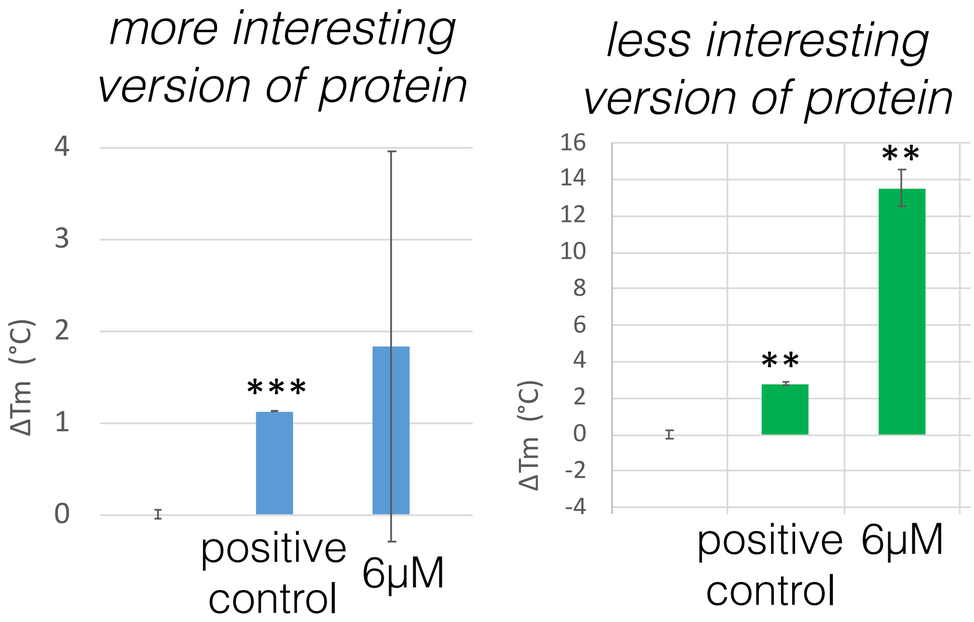
*Sorry to be the bearer of potentially bad news but ... it seems that there may have been some interference (quenching of the product fluorophore) with the compounds/samples.*

# But...

*Sorry to be the bearer of potentially bad news but ... it seems that there may have been some interference (quenching of the product fluorophore) with the compounds/samples.*

# But but...

Thermal shift assays show binding



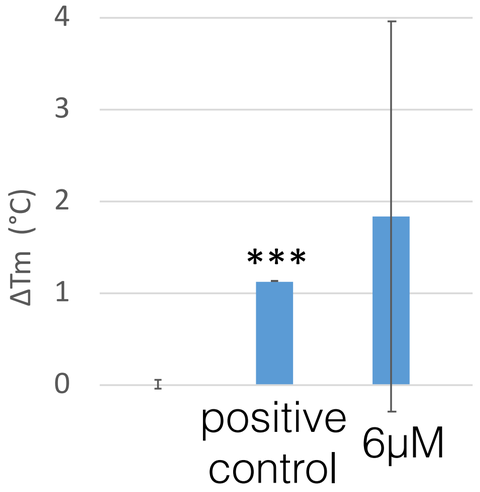
# But...

*Sorry to be the bearer of potentially bad news but ... it seems that there may have been some interference (quenching of the product fluorophore) with the compounds/samples.*

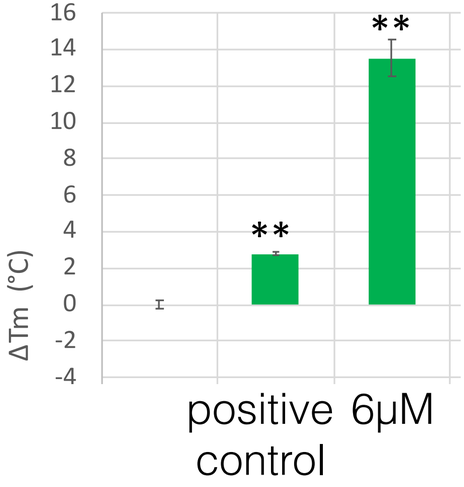
# But but...

Thermal shift assays show binding

*more interesting version of protein*




*less interesting version of protein*




# But but but...

Those error bars

# Case 4: DUSP6

# Dual specificity phosphatase 6

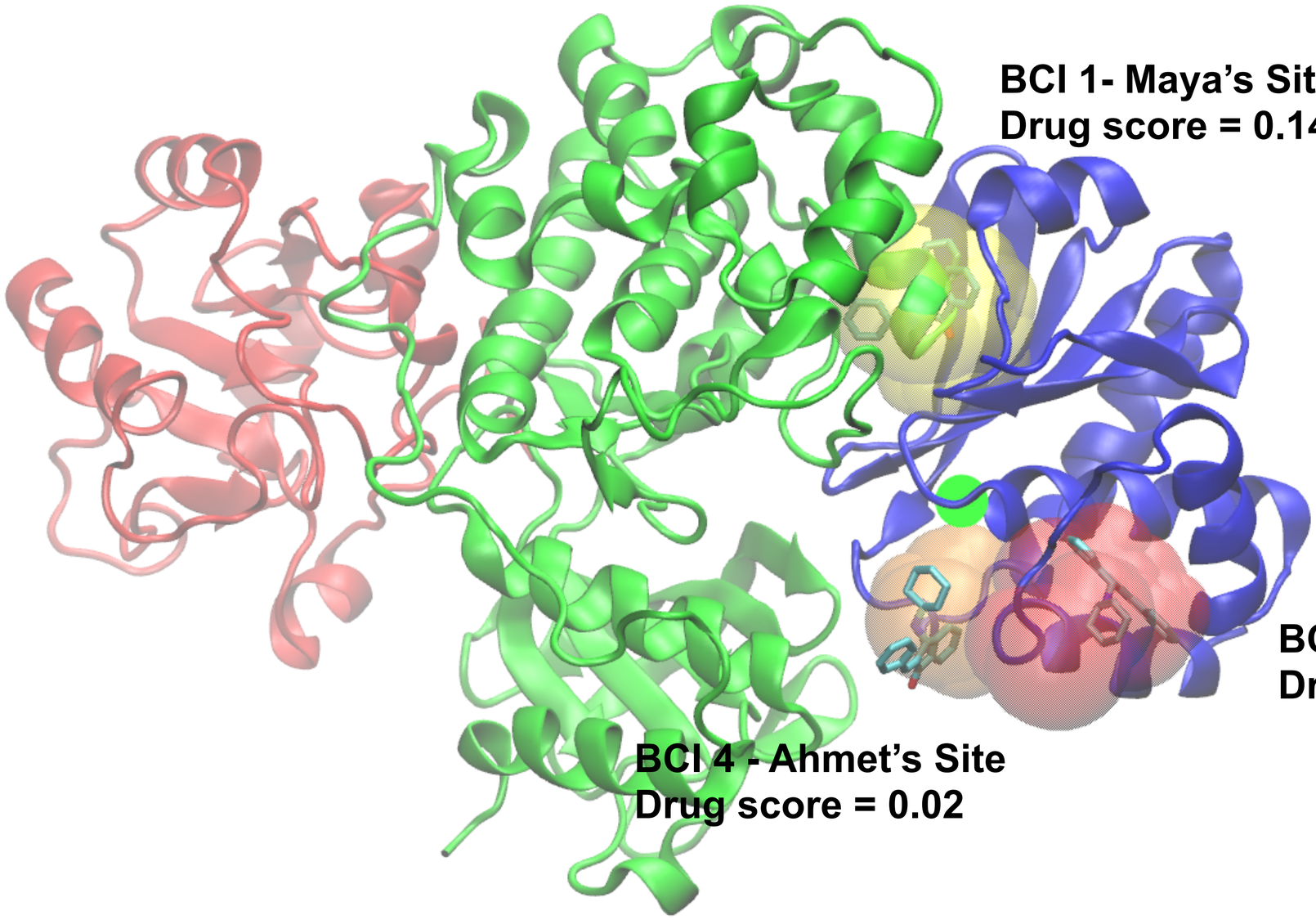
Andreas Vogt

**Students:**

Aaron Zheng

Tamar Skaist

Maya AlMoussa


BCI 1- Maya's Site
Drug score = 0.14
BCI 2- Tamar's Site
Drug score = 0.9
BCI 4 - Ahmet's Site
Drug score = 0.02

# Results

34 compounds tested at the highest possible concentration (300 μM or 75 μM depending on solubility) and 1/10 that

24 hour exposure

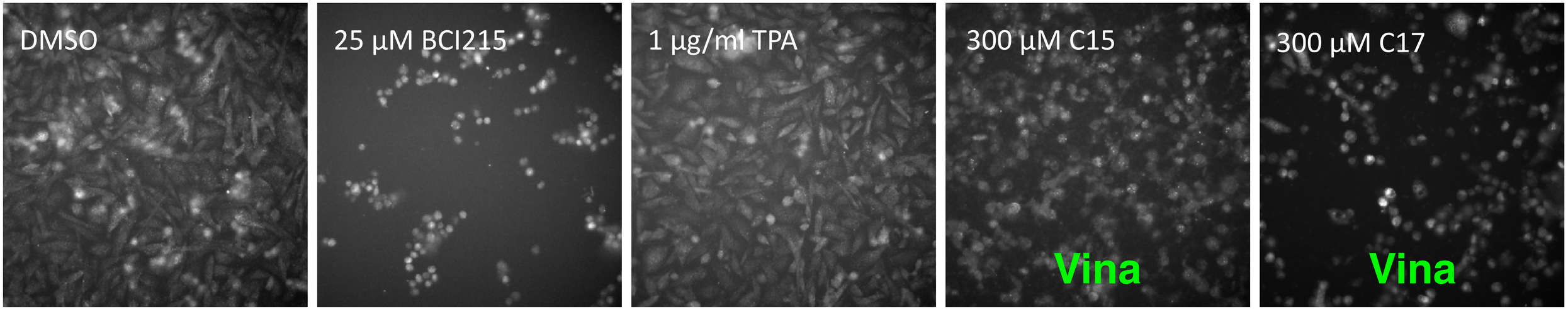
Stain for phospho-ERK

Selected hits with >1.5-fold increase in pERK over DMSO

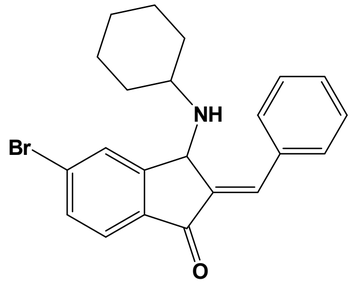
Of the **six** visually **possibles**, three were from the Maya (BCI1) and three from the Ahmet site (BCI4)

**5** hits selected by Vina and **1** by the CNN

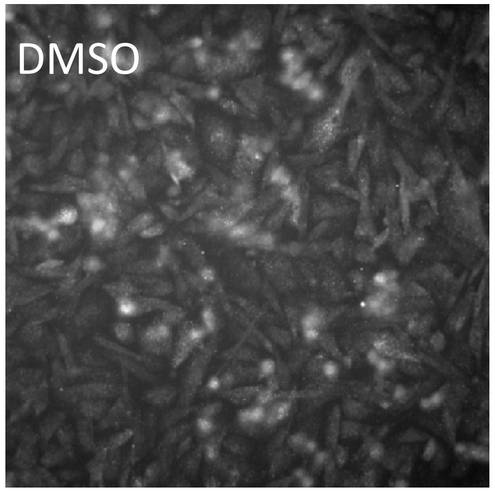
# Possible positives

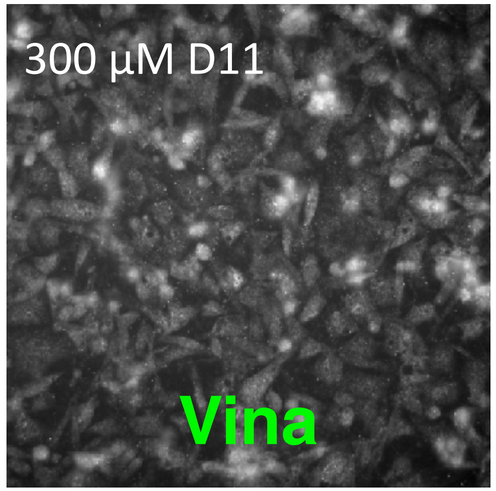


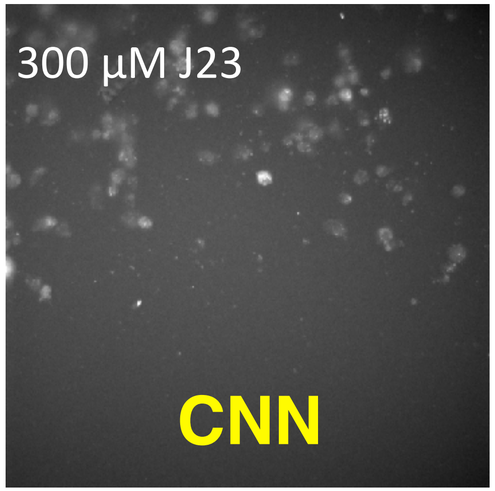

PKC activator

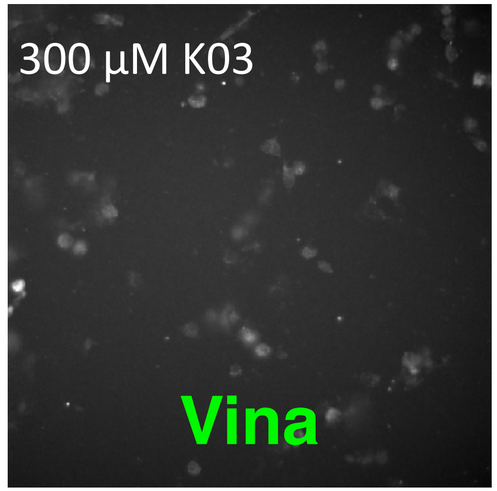


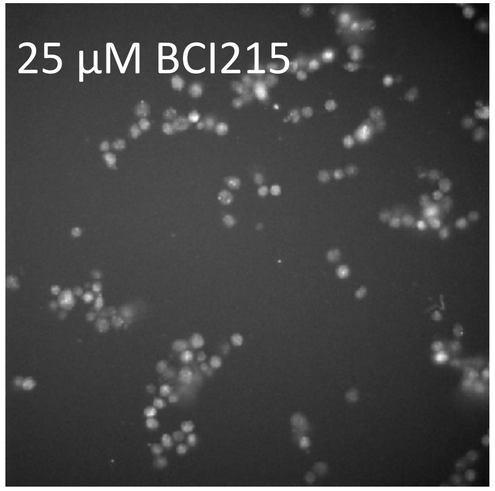
# Possible positives



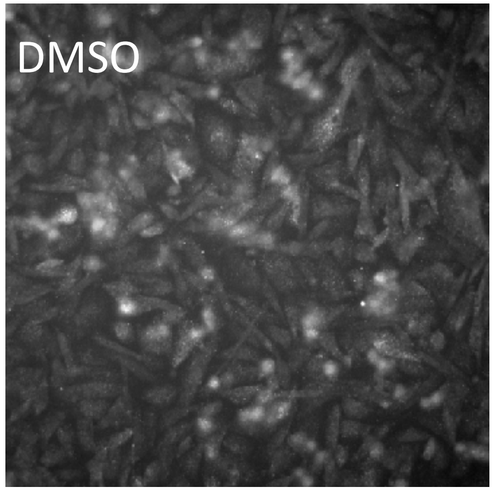



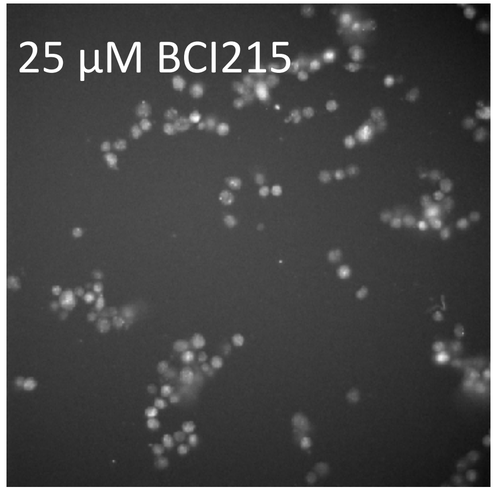



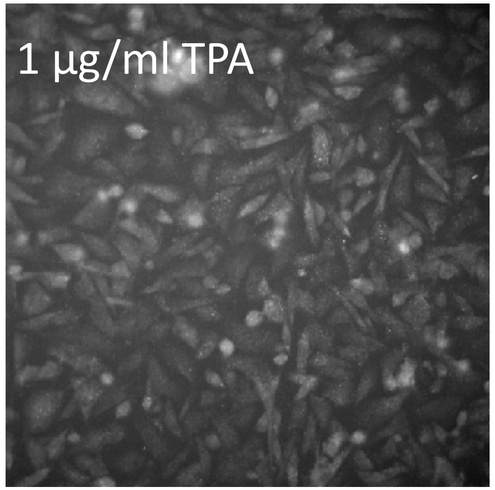



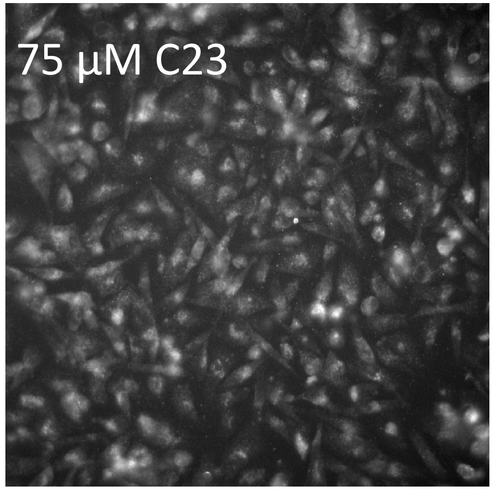



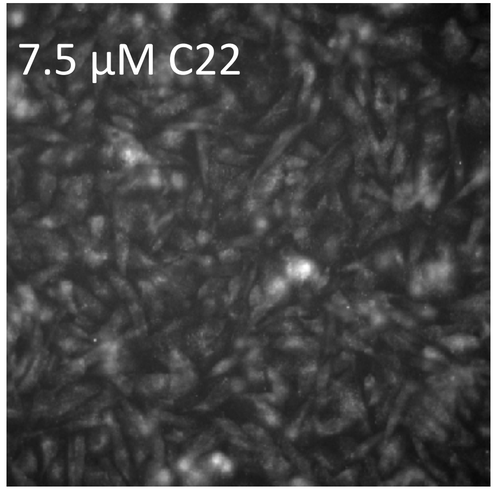
# Possible positives

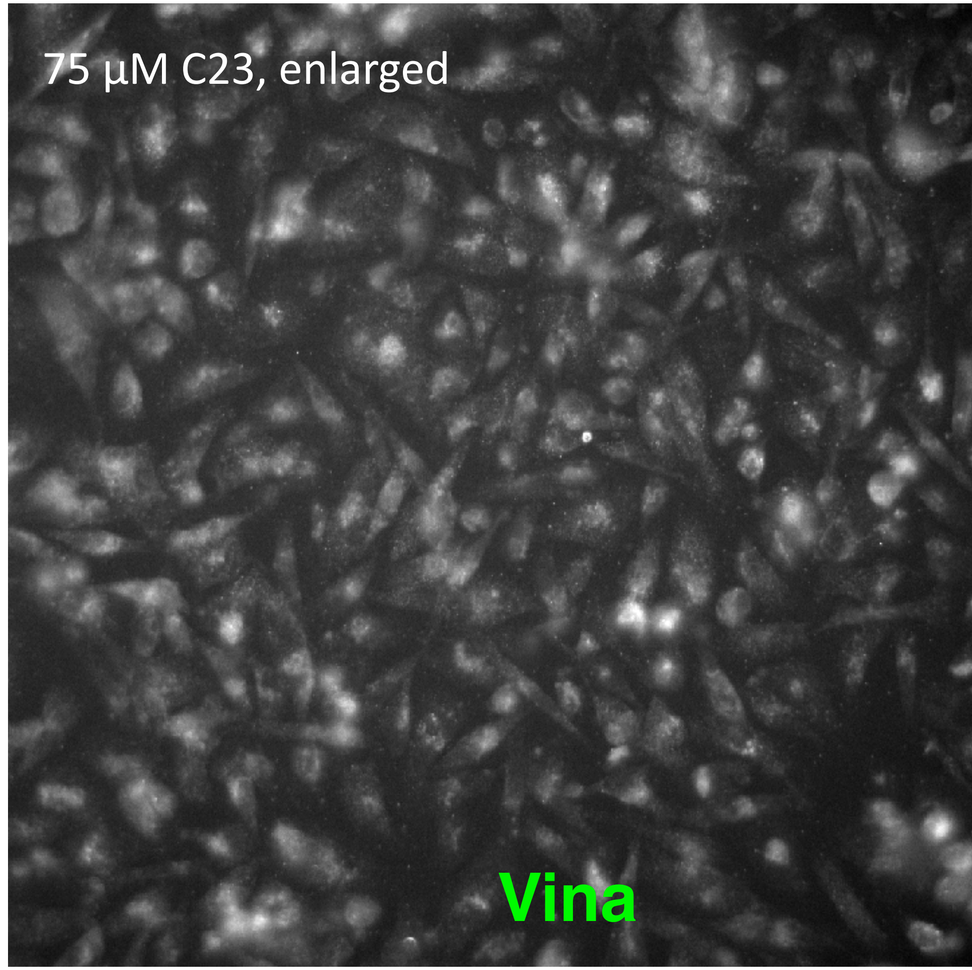


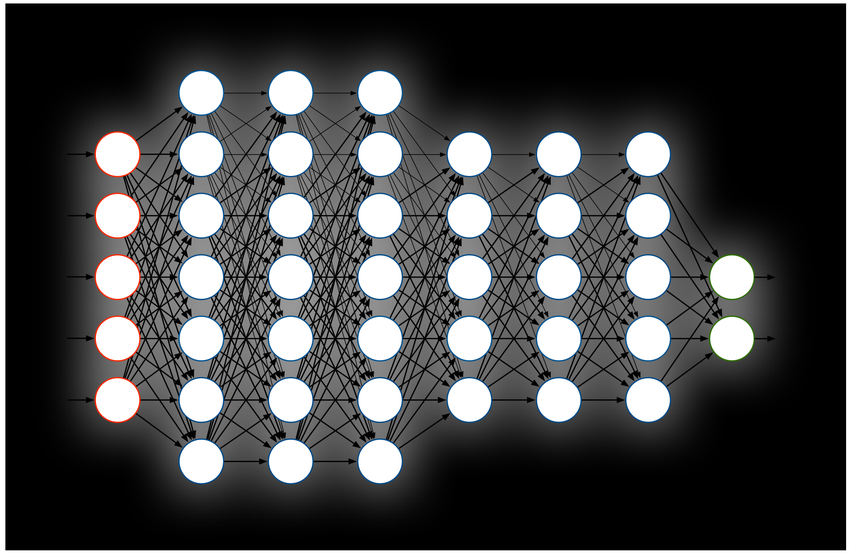



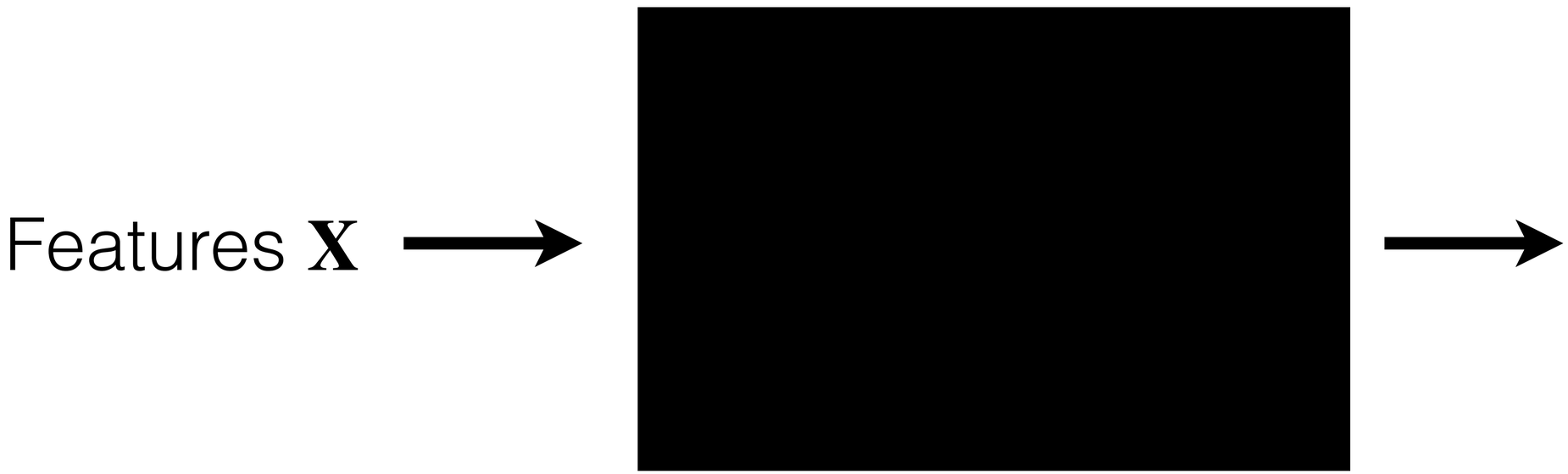



# Generative Modeling

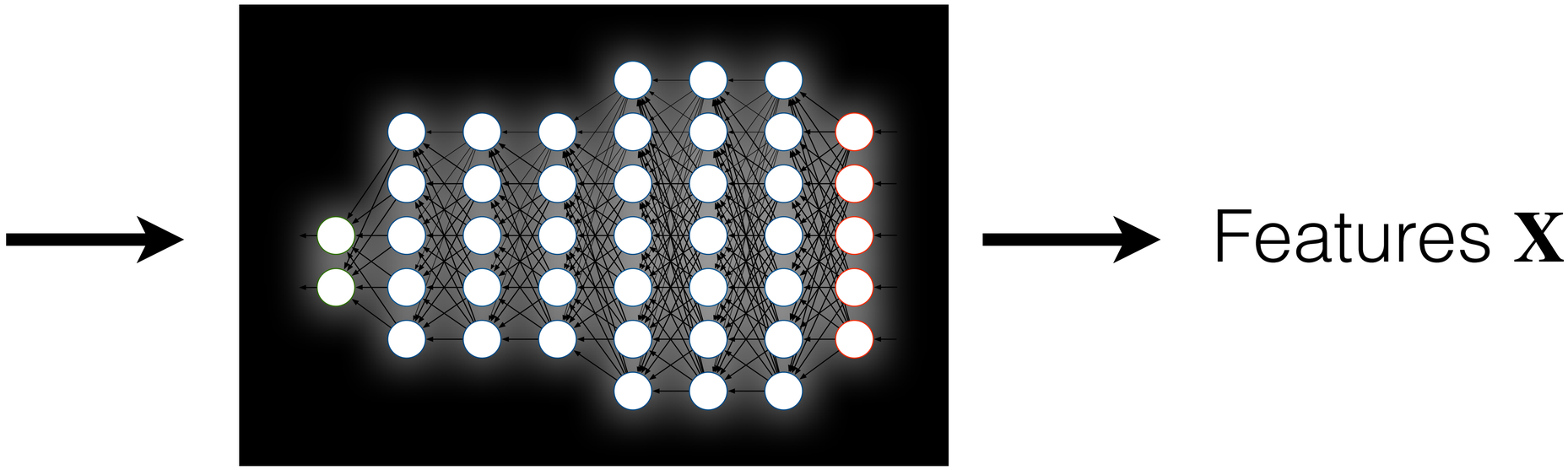
# Discriminative Model

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

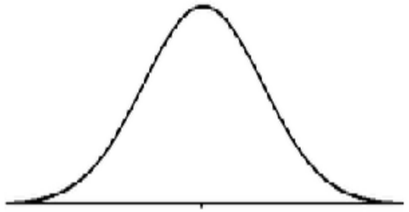
# Generative Model



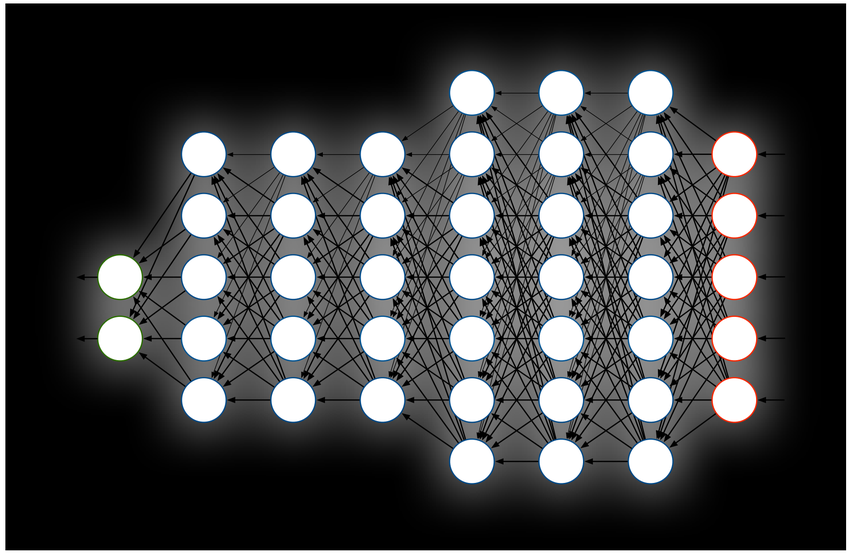
# Generative Model



# Generative Model

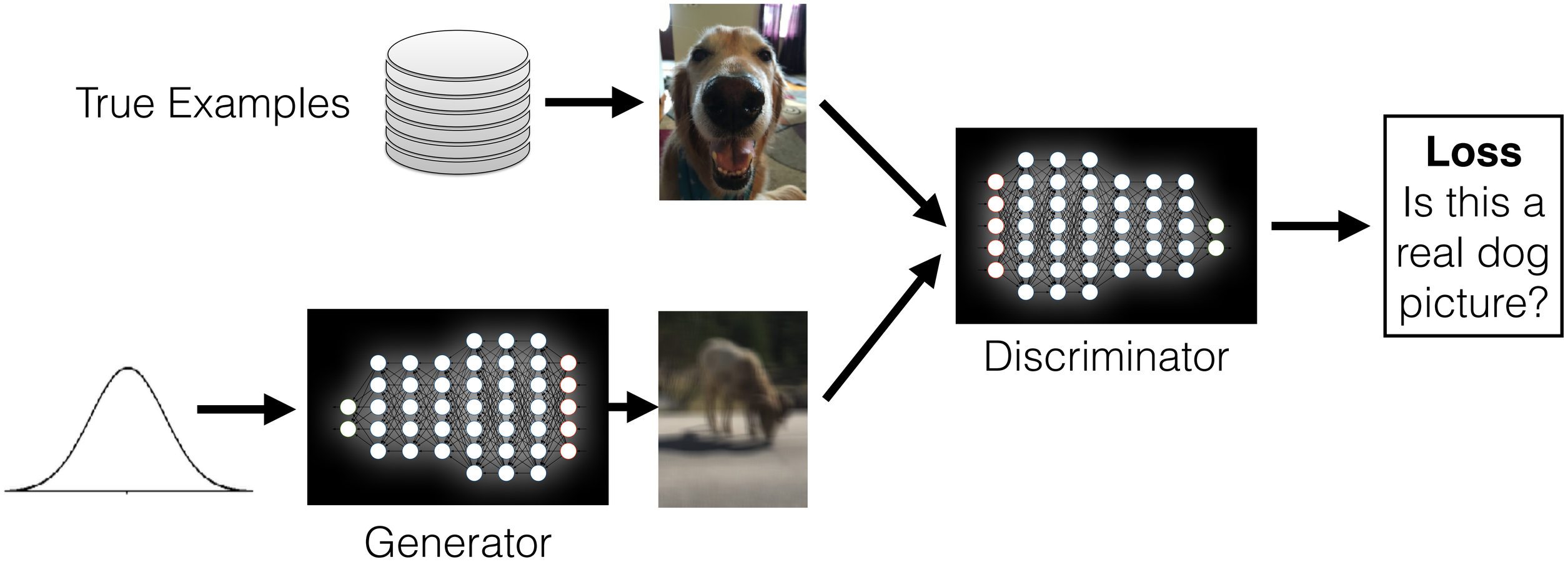


**y?**



Features **X**

# Generative Adversarial Networks



# Generative Adversarial Networks

2014

2015

2016

2017

2018

**Ian Goodfellow** @goodfellow_ian · 2h

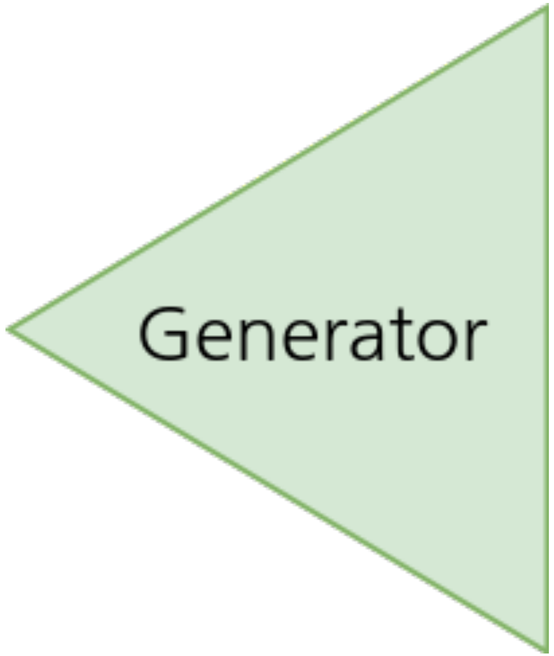
4.5 years of GAN progress on face generation. arxiv.org/abs/1406.2661 arxiv.org/abs/1511.06434 arxiv.org/abs/1606.07536 arxiv.org/abs/1710.10196 arxiv.org/abs/1812.04948

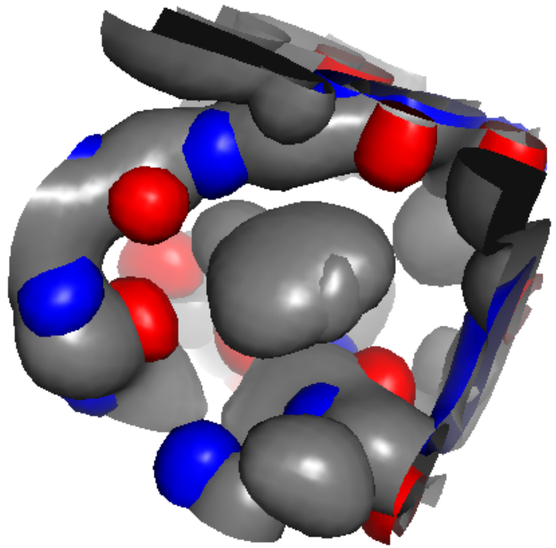
https://thispersondoesnotexist.com

# 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.

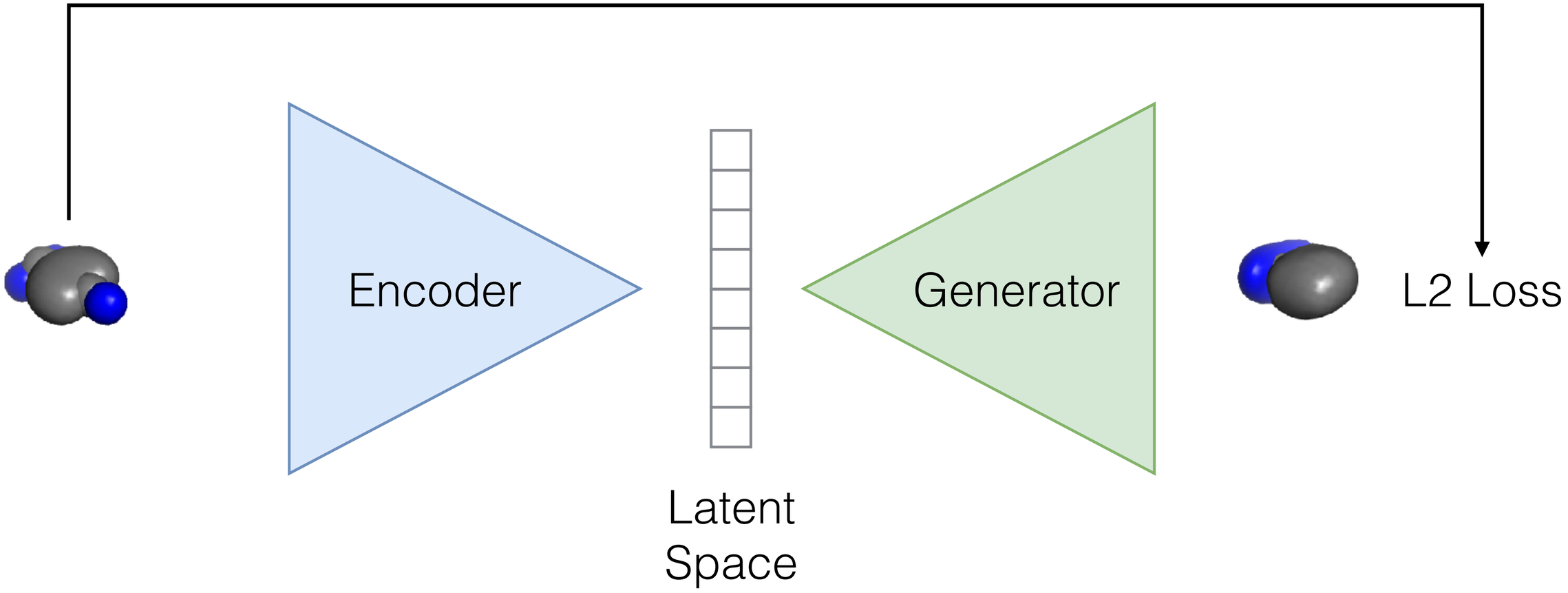
*noise sample*



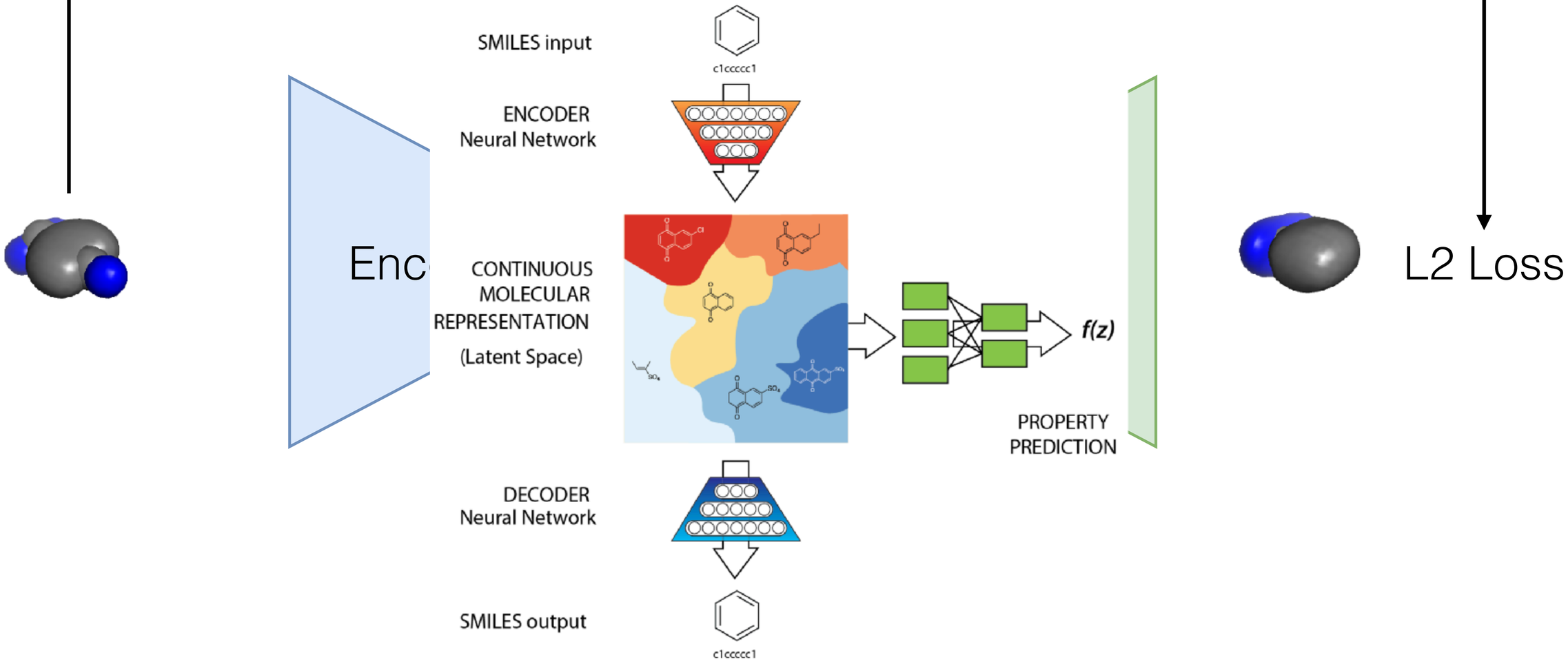
*generated receptor & ligand grid*

# Autoencoding

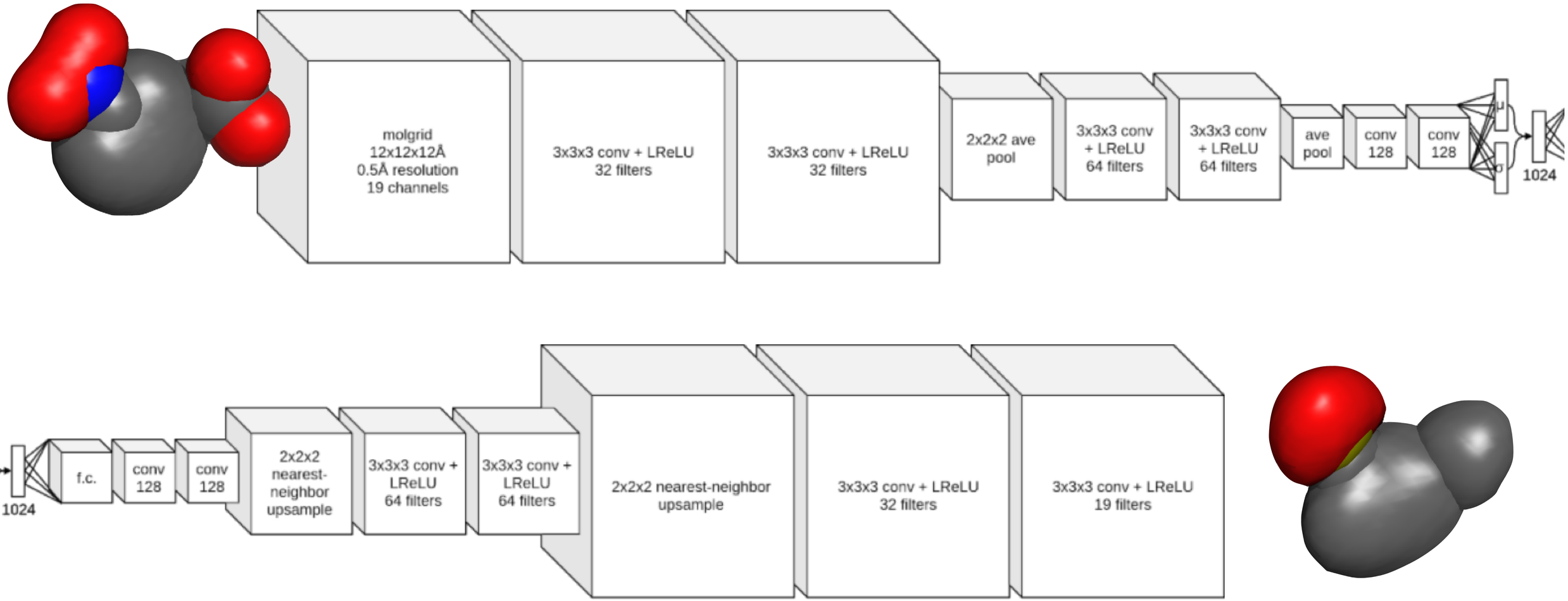


# Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules

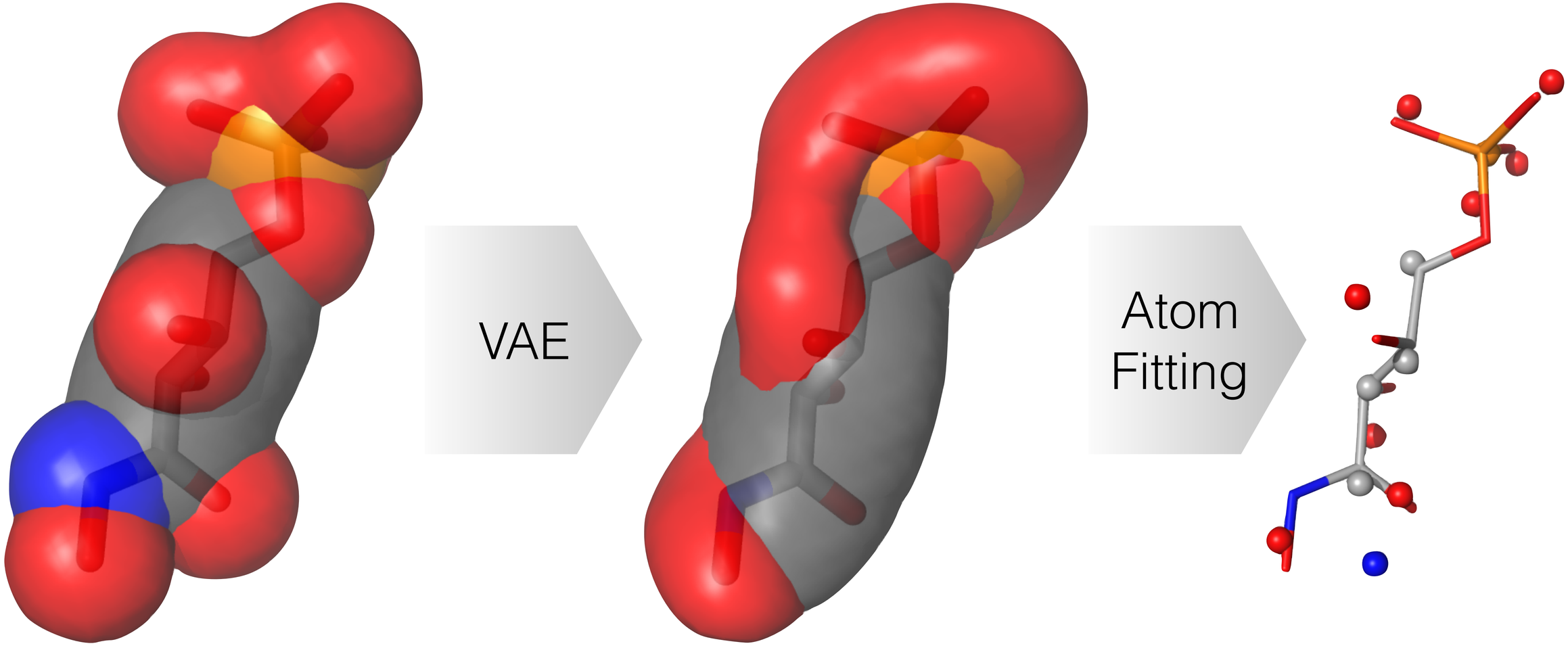
Rafael Gómez-Bombarelli[†#], Jennifer N. Wei[‡#], David Duvenaud[¶#], José Miguel Hernández-Lobato[§#], Benjamín Sánchez-Lengeling[‡], Dennis Sheberla[‡], Jorge Aguilera-Iparraguirre[†], Timothy D. Hirzel[†], Ryan P. Adams[∇⊥], and Alán Aspuru-Guzik[*‡⊥]



# Model

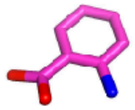
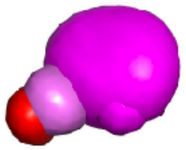
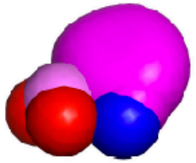
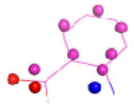
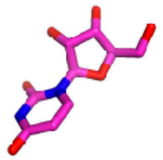
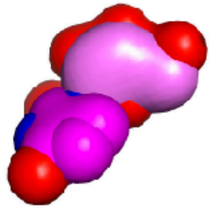
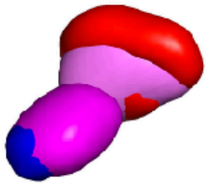
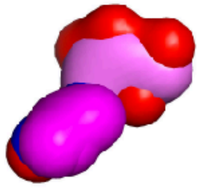
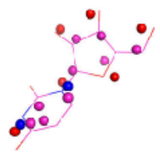
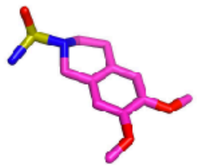
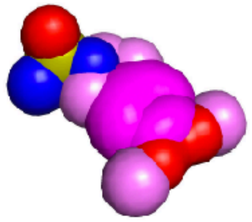
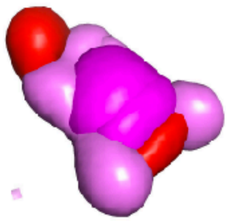
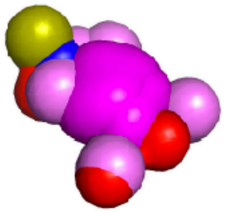
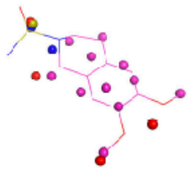
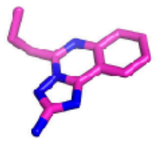
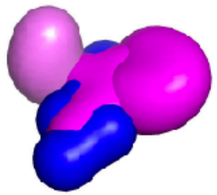
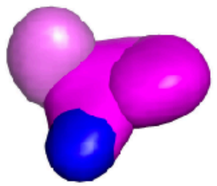
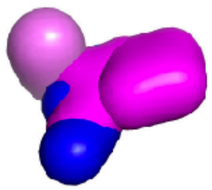
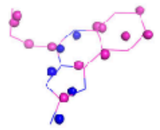

molgrid
12x12x12Å
0.5Å resolution
19 channels
3x3x3 conv + LReLU
32 filters
3x3x3 conv + LReLU
32 filters
2x2x2 ave
pool
3x3x3 conv
+ LReLU
64 filters
3x3x3 conv
+ LReLU
64 filters
ave
pool
conv
128
conv
128
1024
1024
f.c.
conv
128
conv
128
2x2x2
nearest-
neighbor
upsample
3x3x3 conv +
LReLU
64 filters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

# Variational Autoencoding Examples

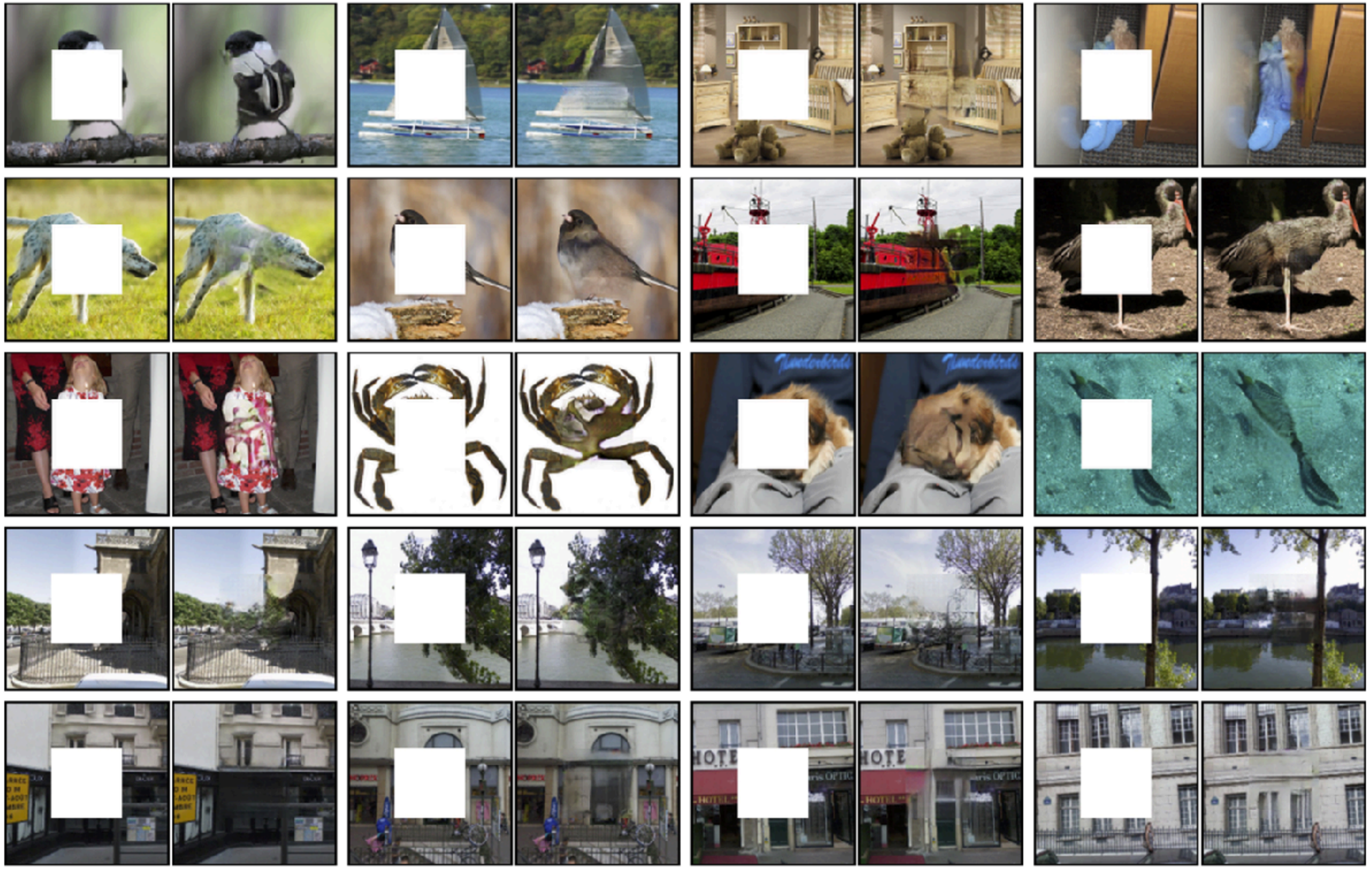



2BES

# Variational Autoencoding Examples

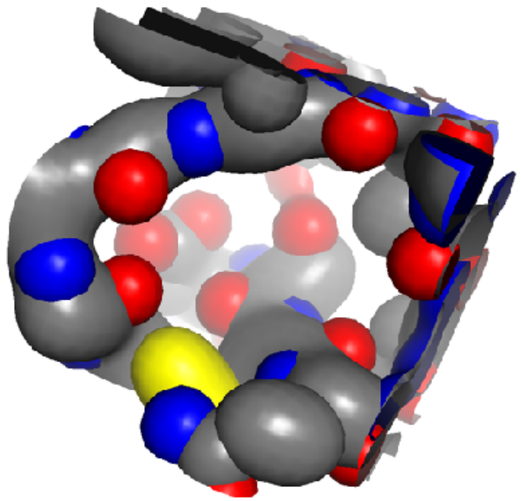
| PDB | True structure | True density | Gen. density | Fit density | Fit structure | Gen. L2 distance | Fit L2 distance | Fit RMSD |
|---|---|---|---|---|---|---|---|---|
| 3h78 |  |  |  |  |  | 9.4053 | 8.3141 | 0.6160 |
| 4jx9 |  |  |  |  |  | 13.8545 | 9.7198 | 0.8820 |
| 3igp |  |  |  |  |  | 14.8525 | 12.5245 | 1.2066 |
| 4cwf |  |  |  |  |  | 11.4730 | 9.0564 | 0.6725 |

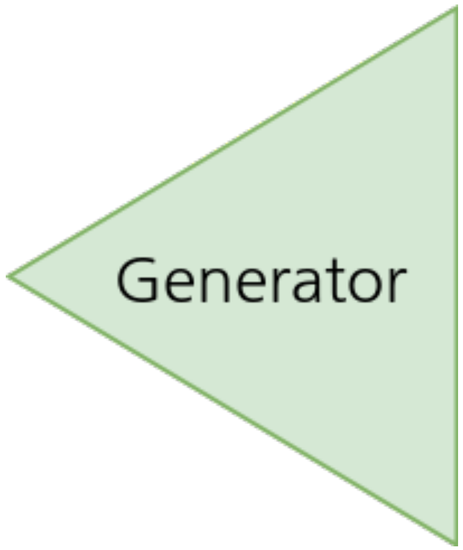
# Context Encoding

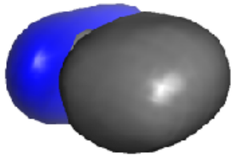


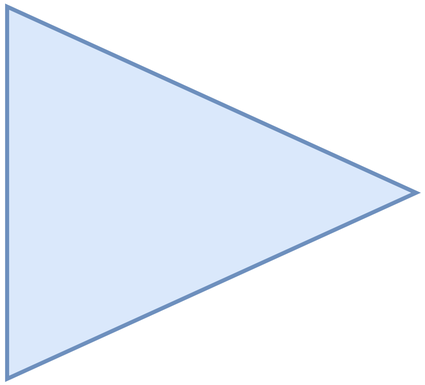
http://people.eecs.berkeley.edu/~pathak/context_encoder/

# Context Encoding





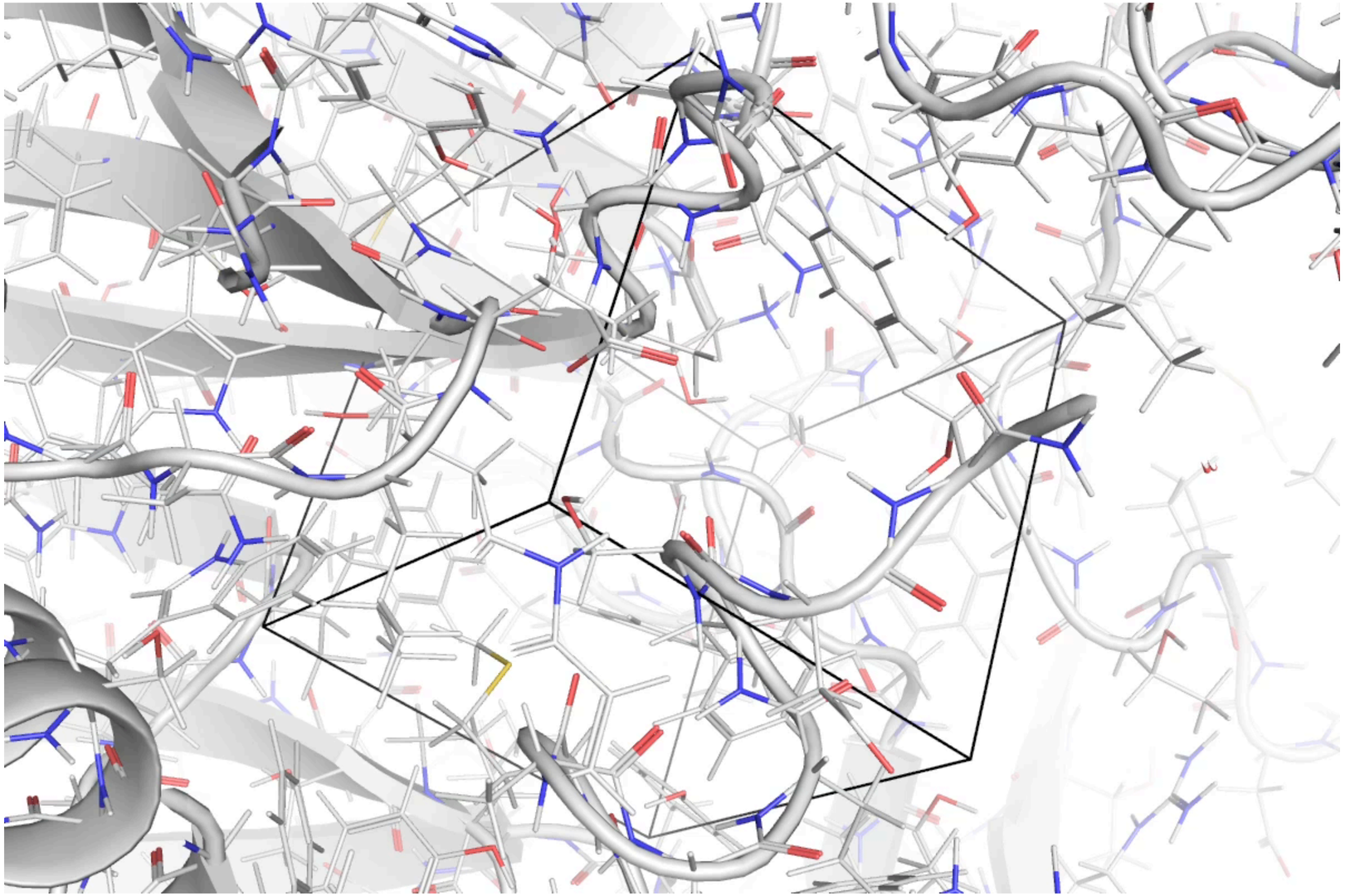





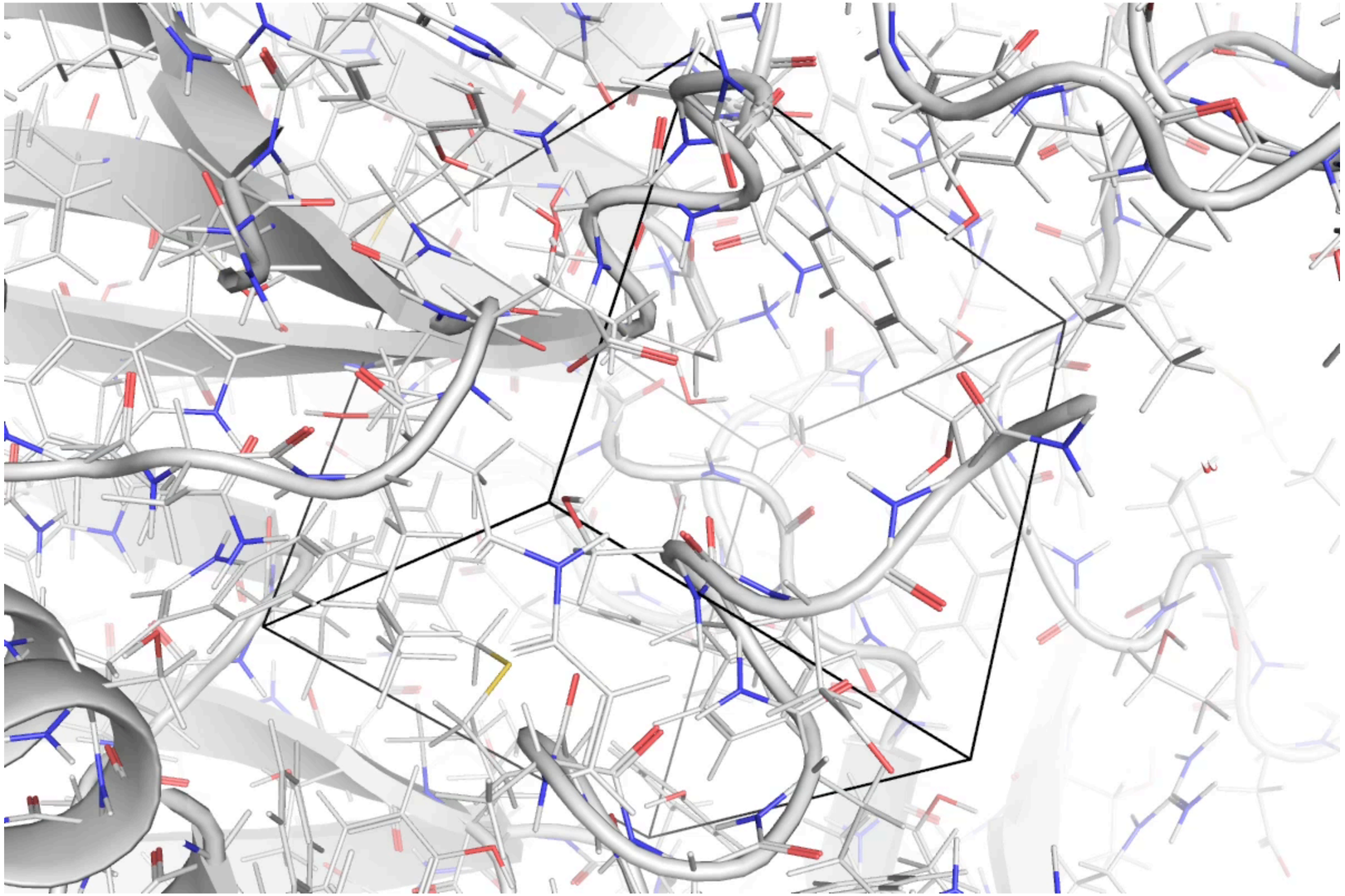

*receptor grid*

*generated ligand grid*

# Conditioning on the Receptor

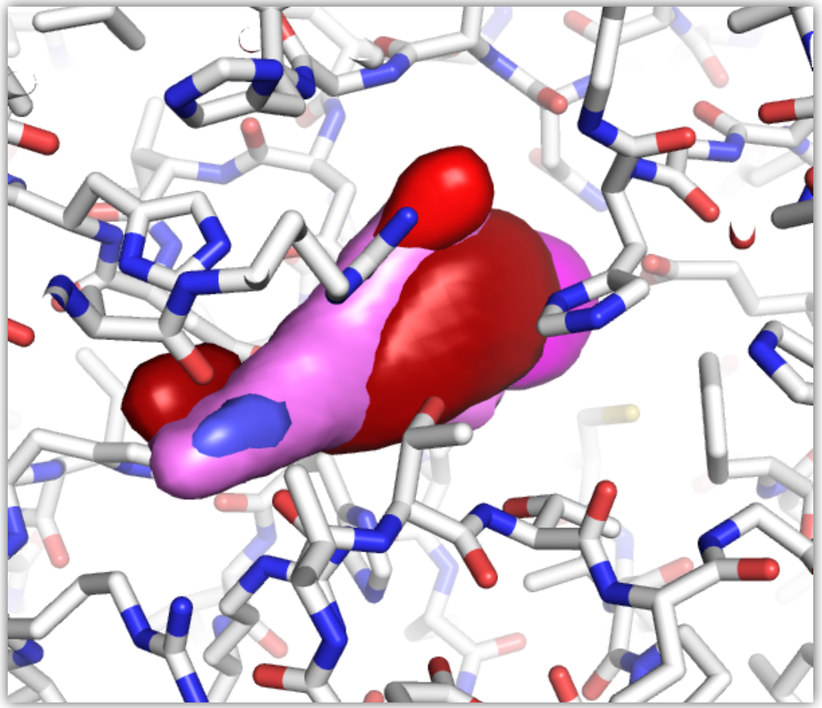


# Conditioning on the Receptor

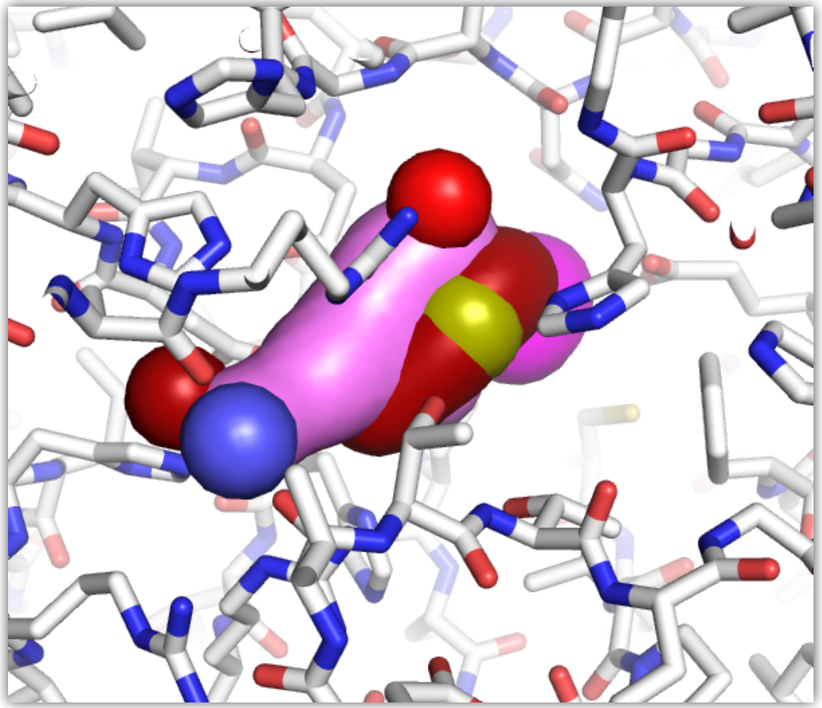


# Context Encoding with Fully Convolutional Network

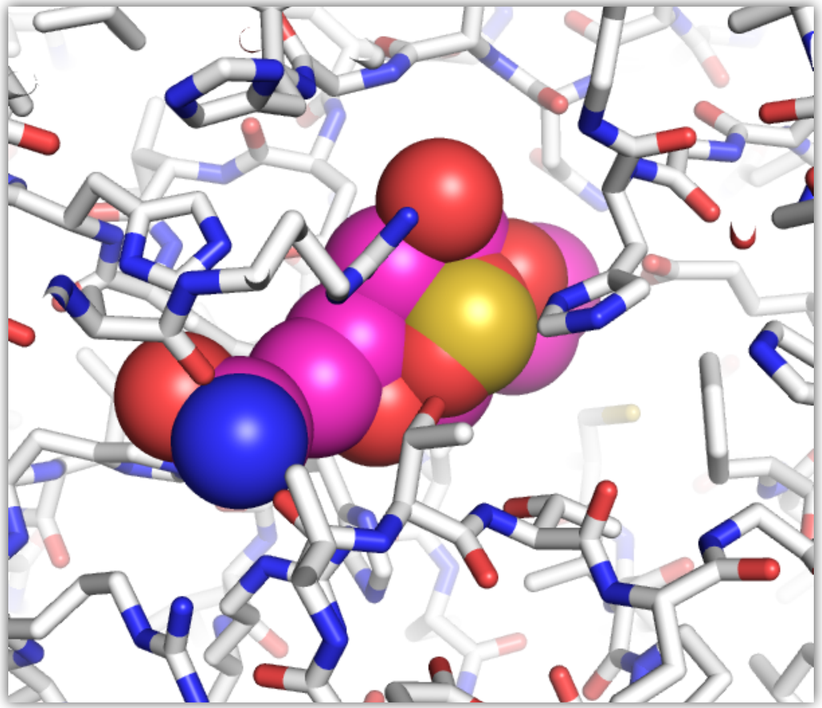
1m5w



Generated



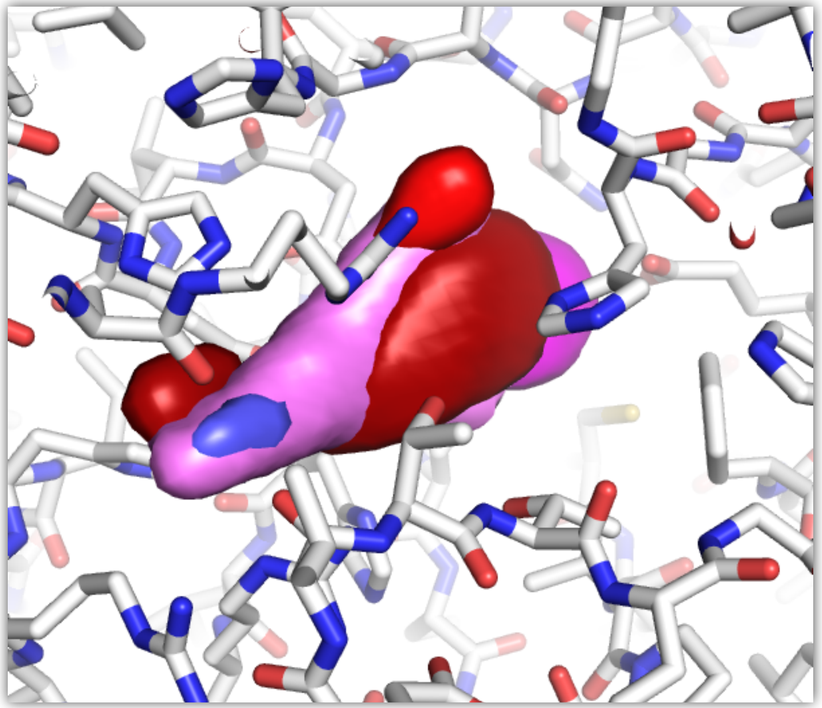
Fit Densities



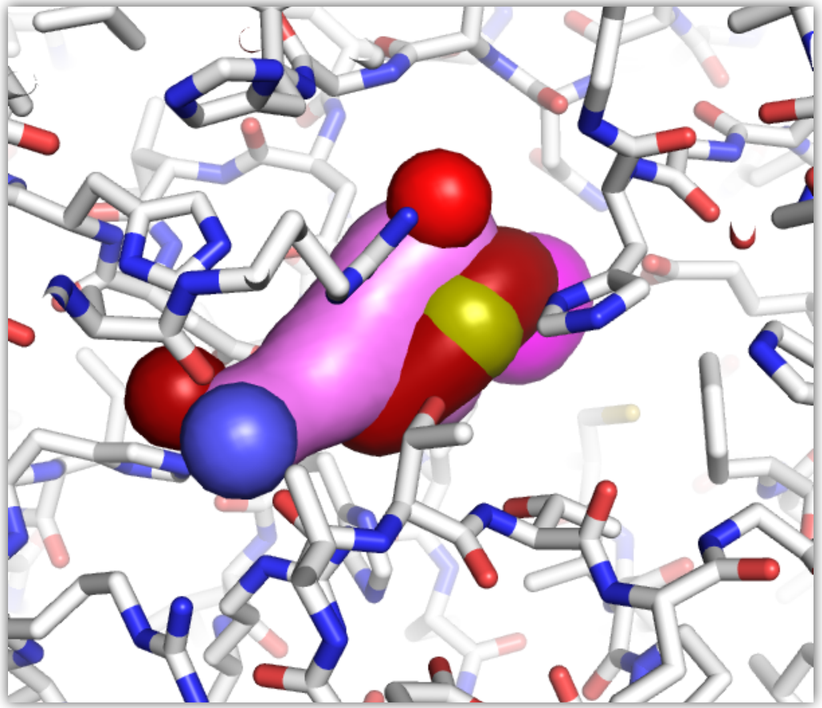
Fit Atoms

# Context Encoding with Fully Convolutional Network

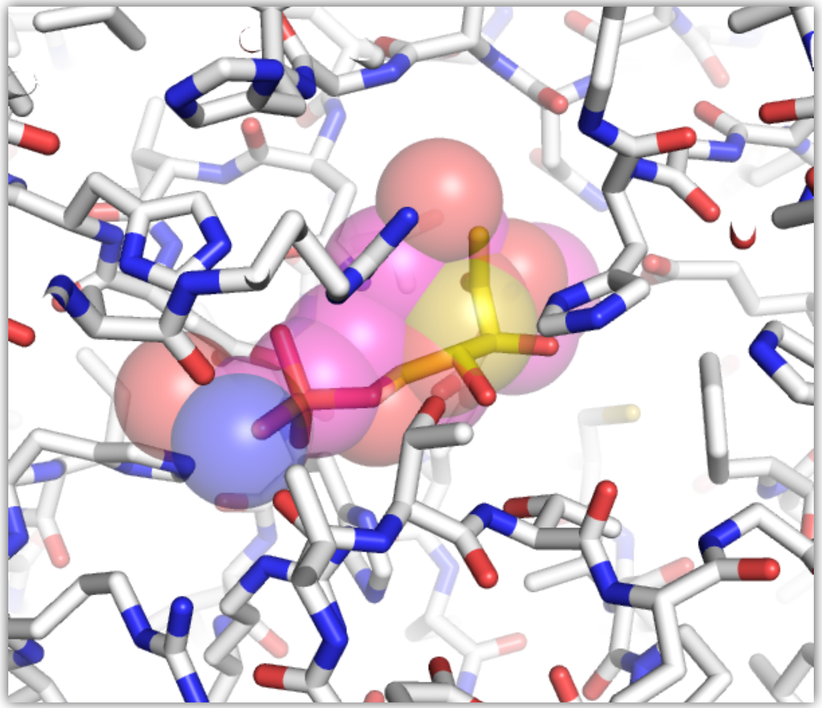
1m5w



Generated



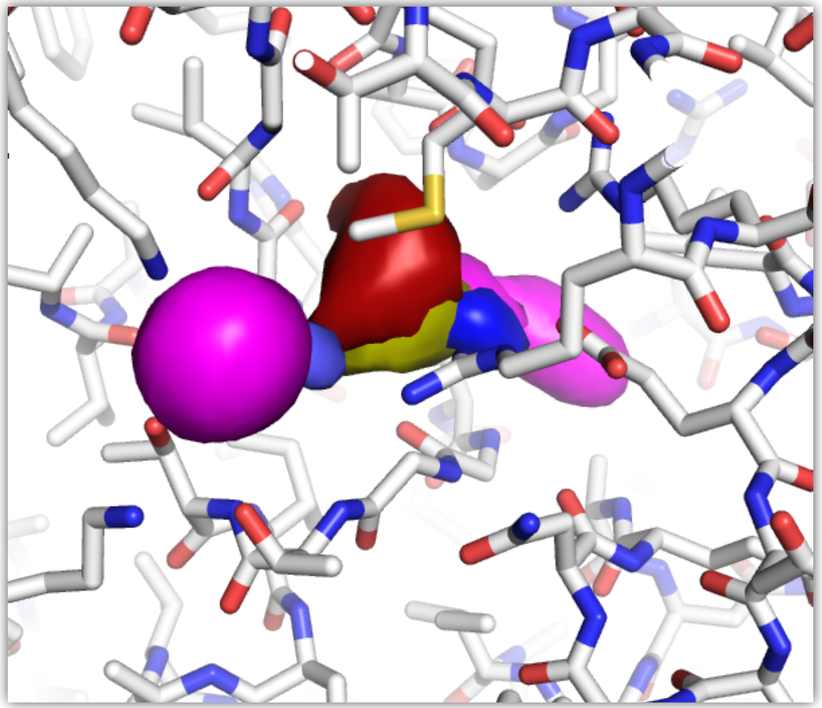
Fit Densities



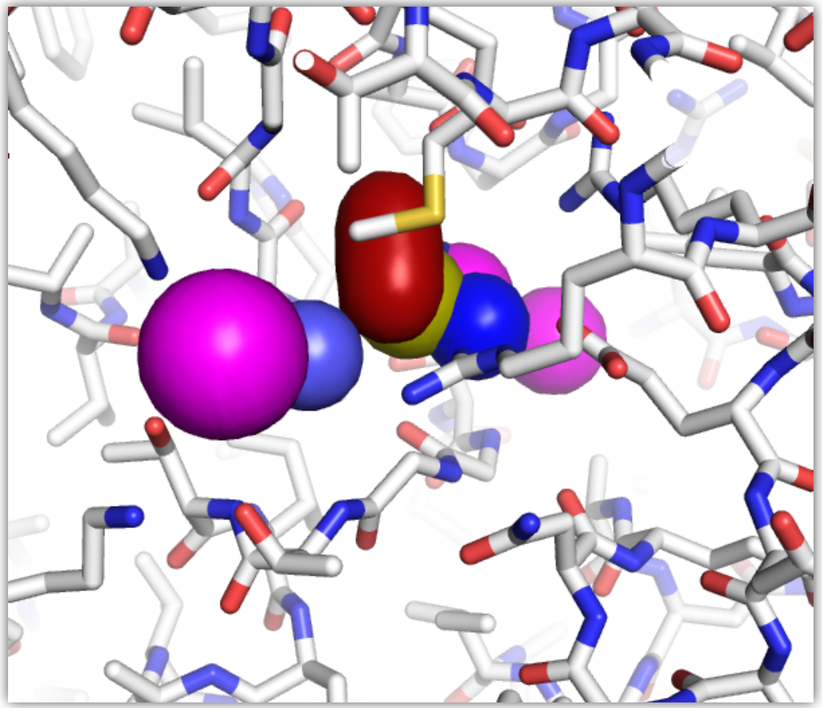
Fit Atoms

# Context Encoding with Fully Convolutional Network

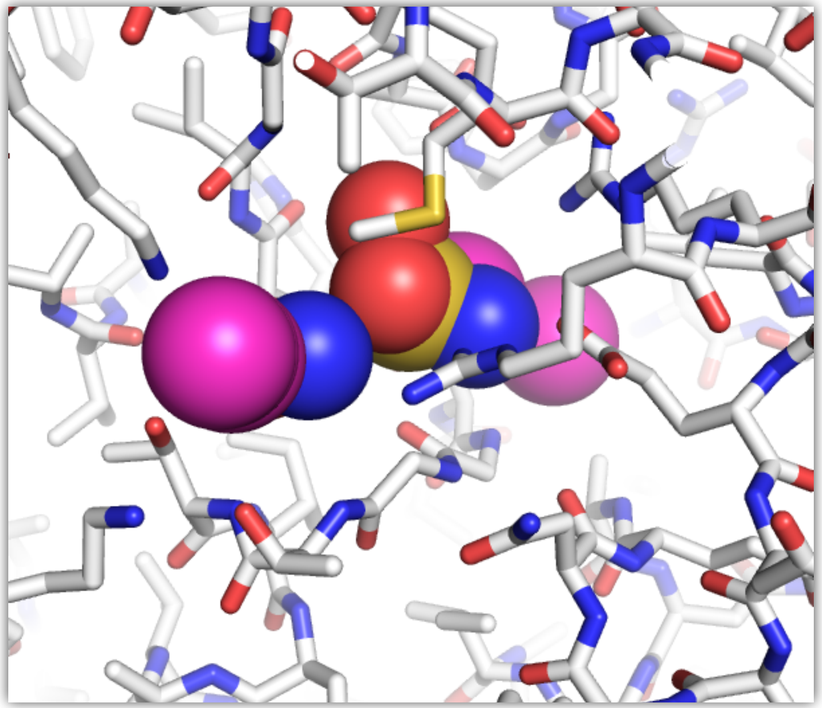
3bxg



Generated



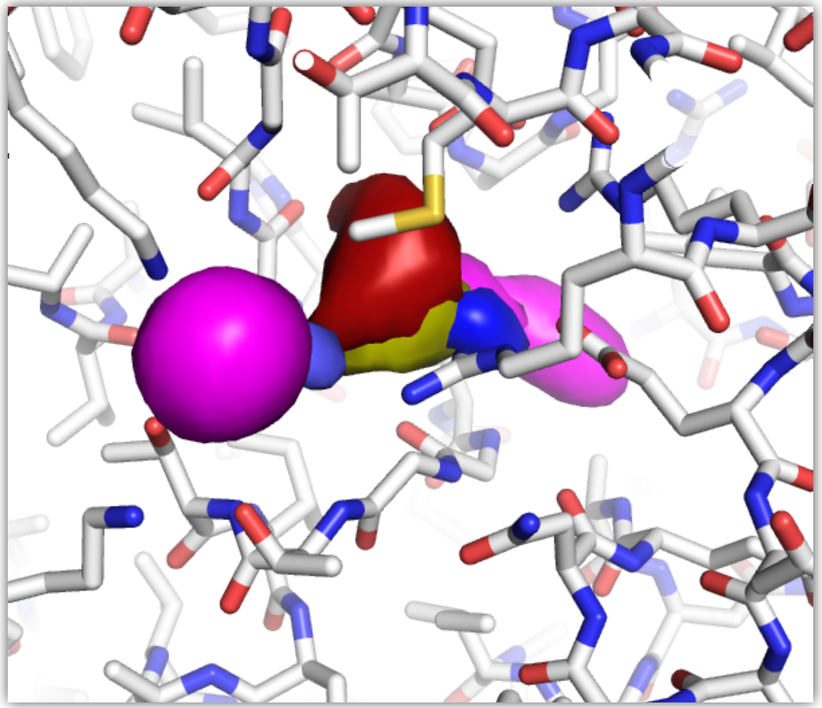
Fit Densities



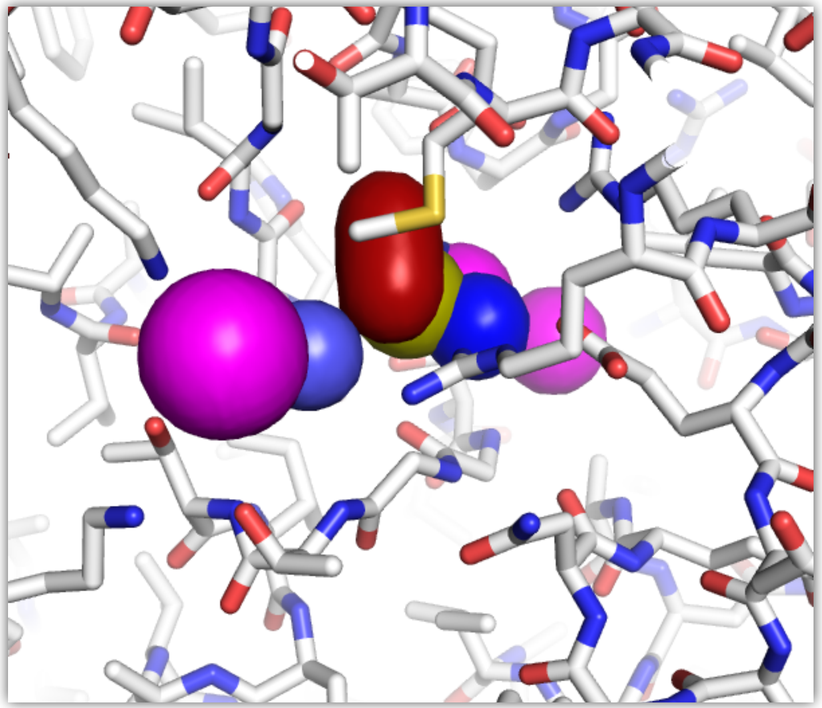
Fit Atoms

# Context Encoding with Fully Convolutional Network

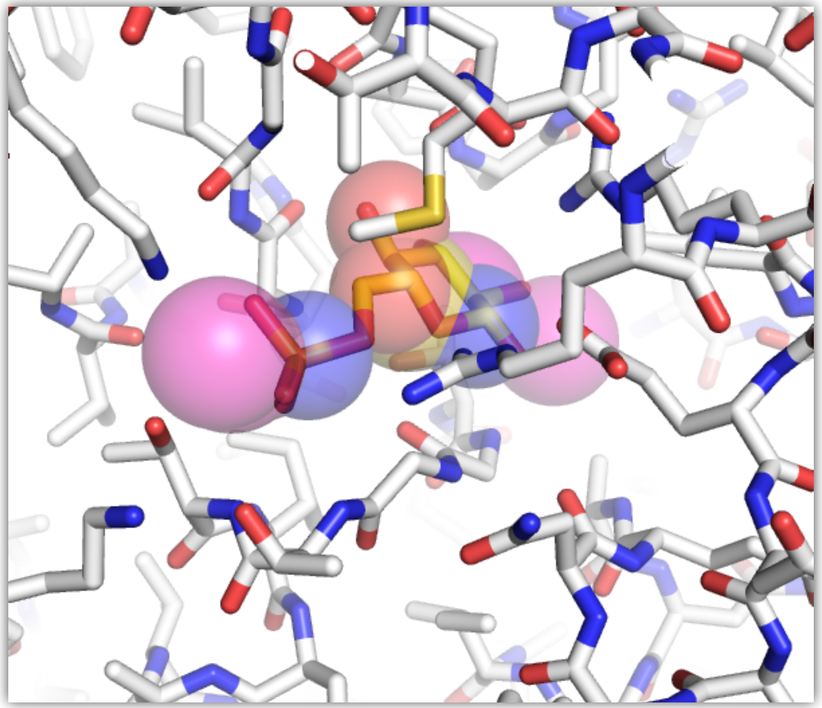
3bxg



Generated



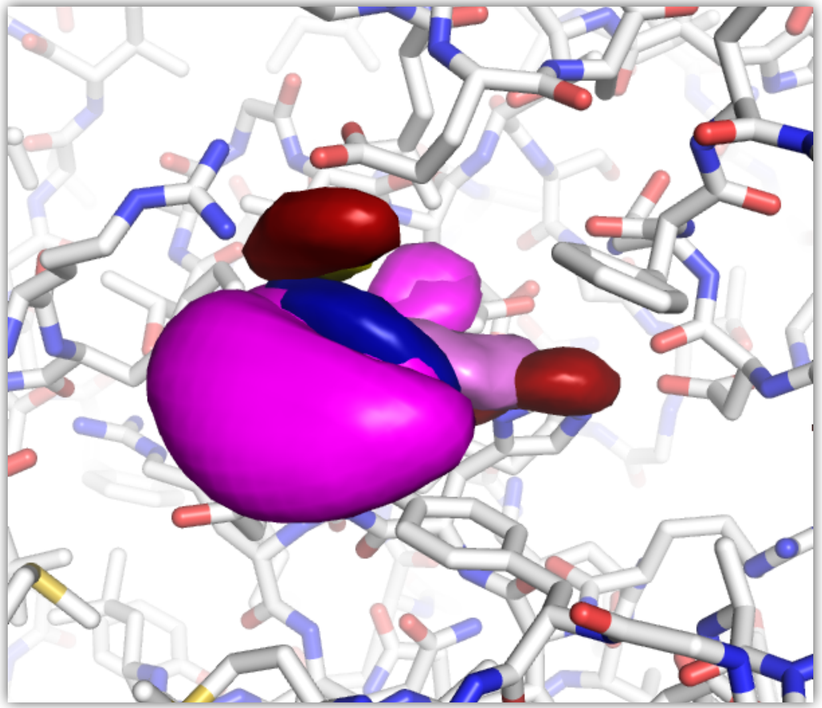
Fit Densities



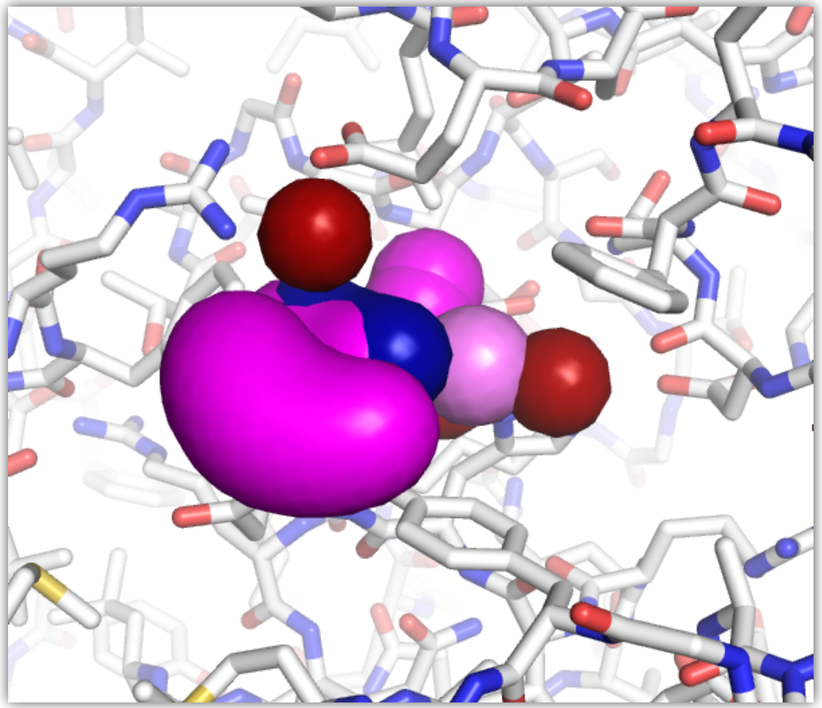
Fit Atoms

# Context Encoding with Fully Convolutional Network

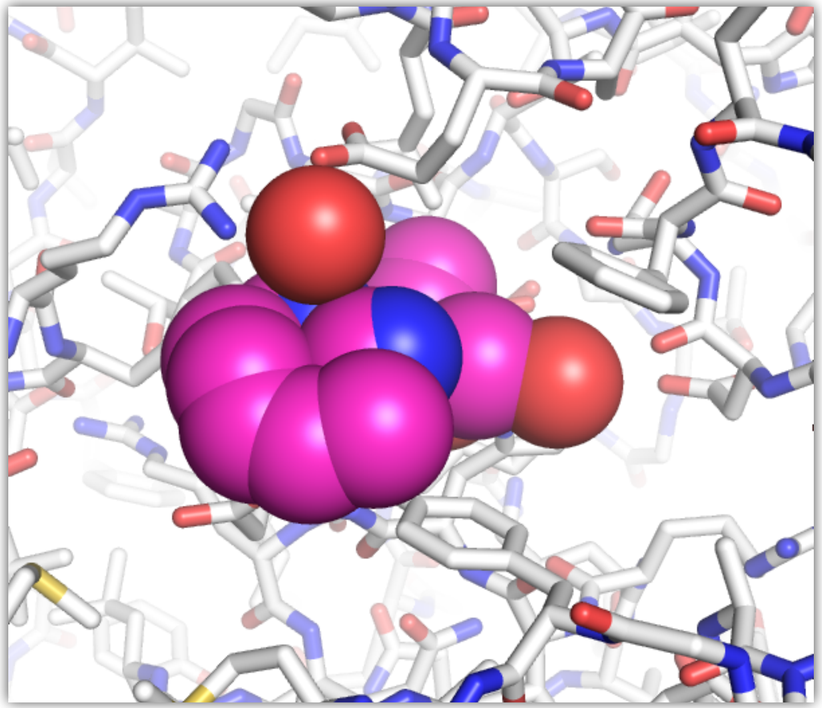
3ebp



Generated



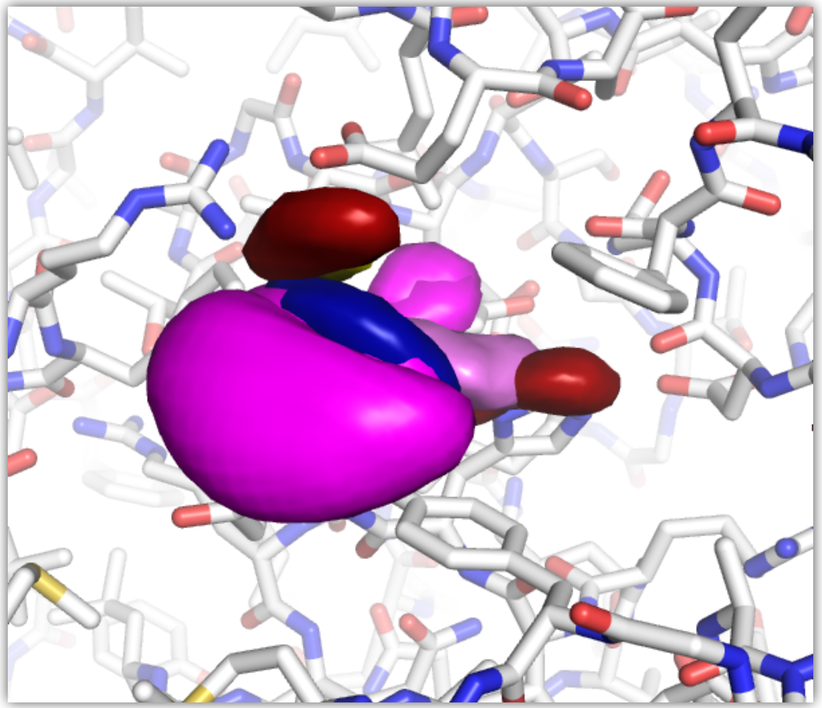
Fit Densities



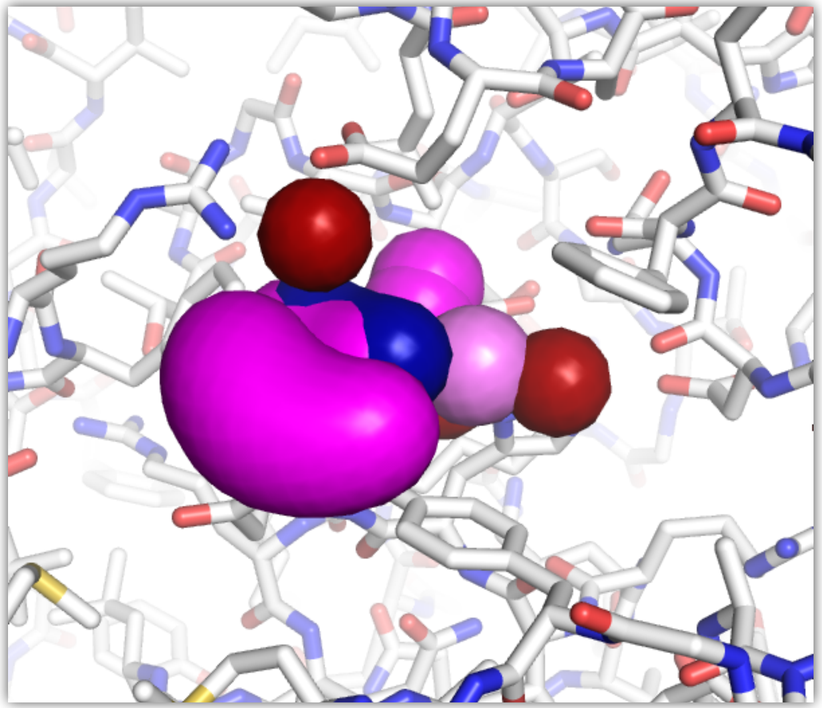
Fit Atoms

# Context Encoding with Fully Convolutional Network

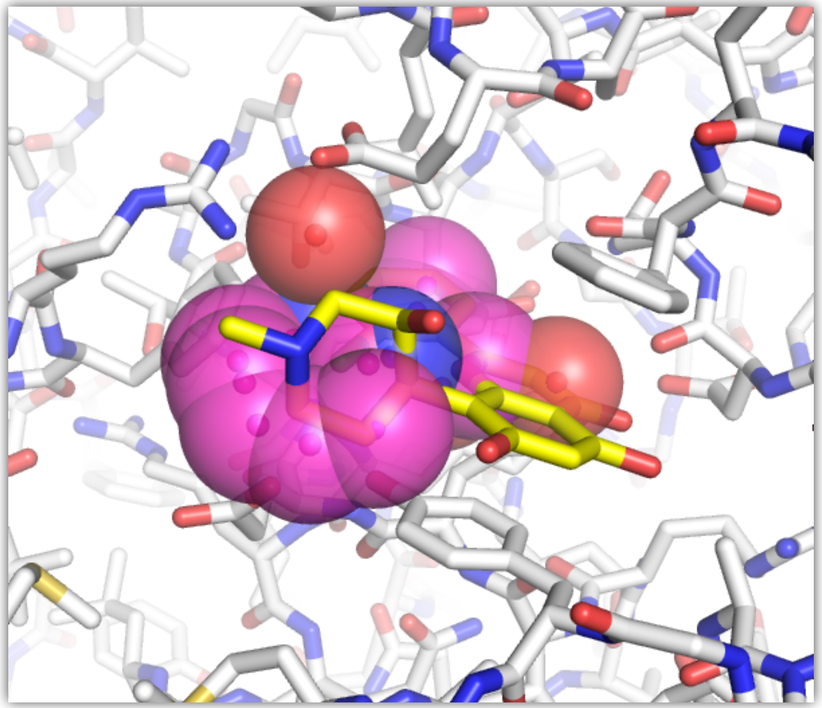
3ebp



Generated

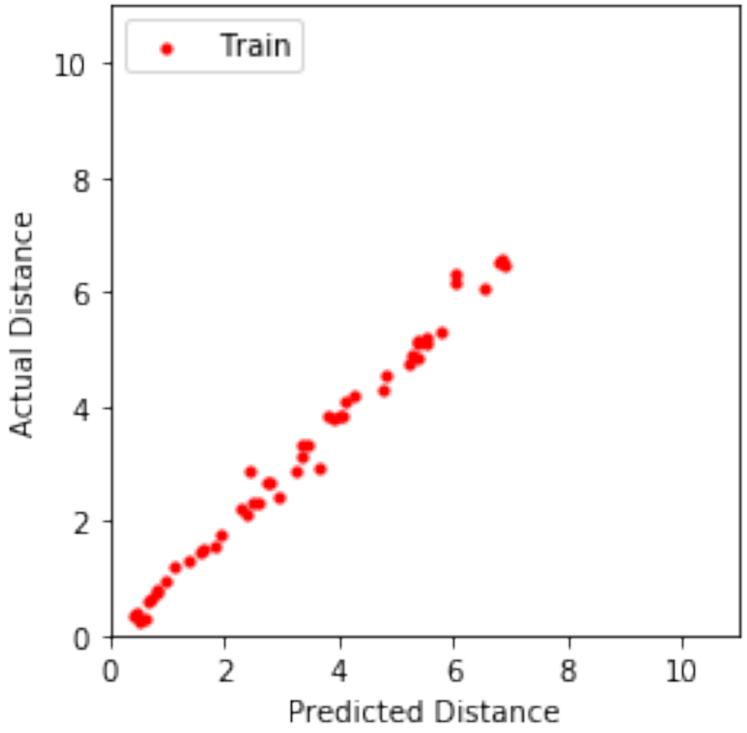


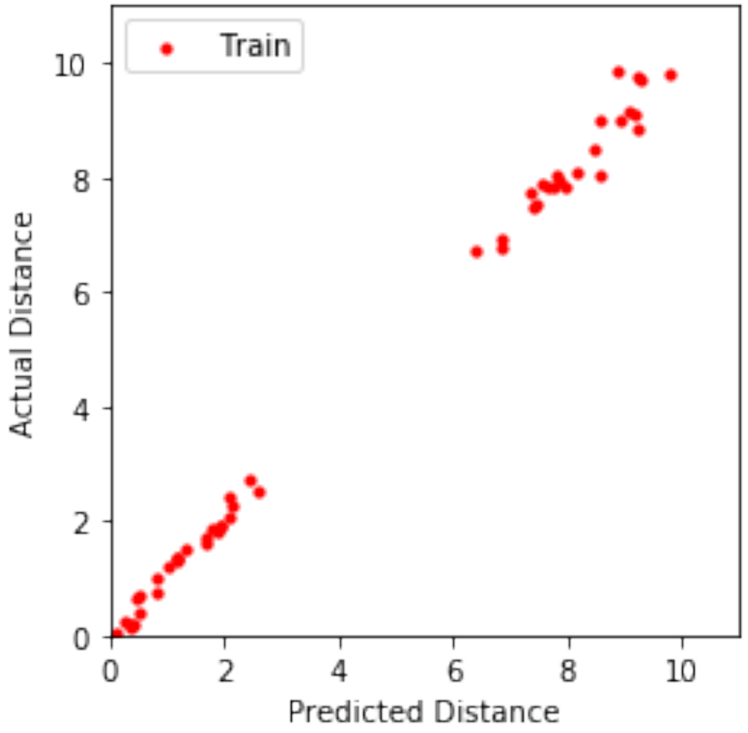
Fit Densities

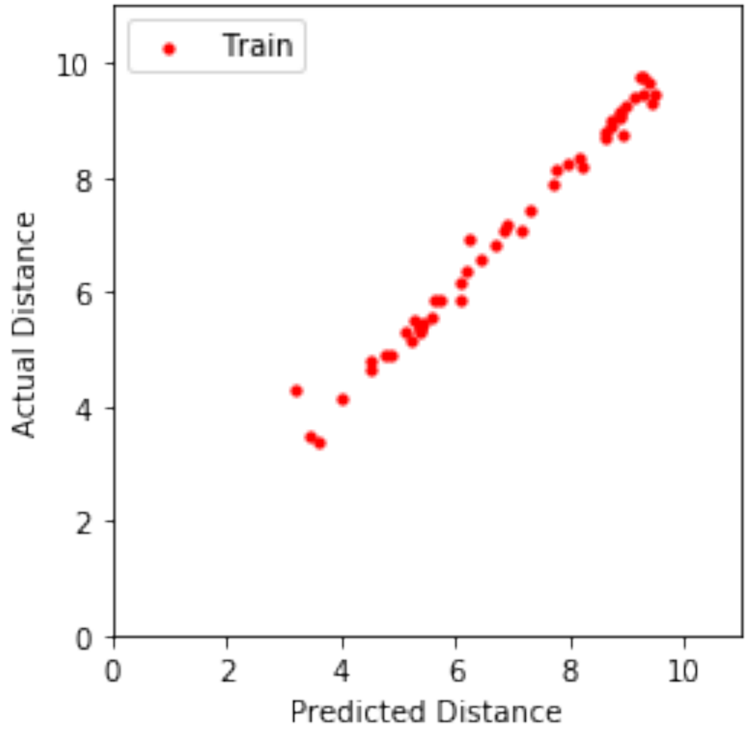


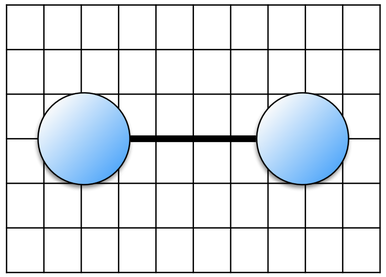
Fit Atoms

# An Aside…



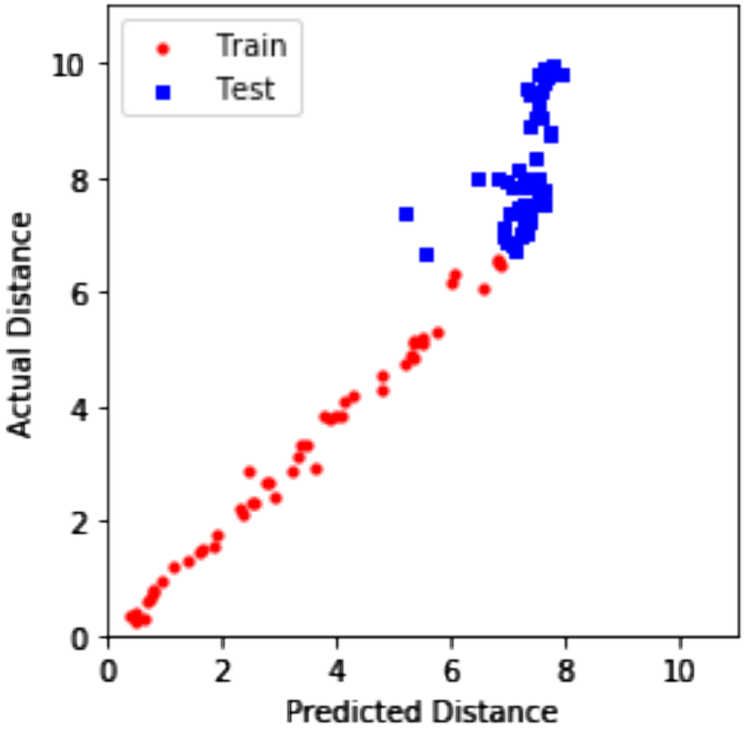



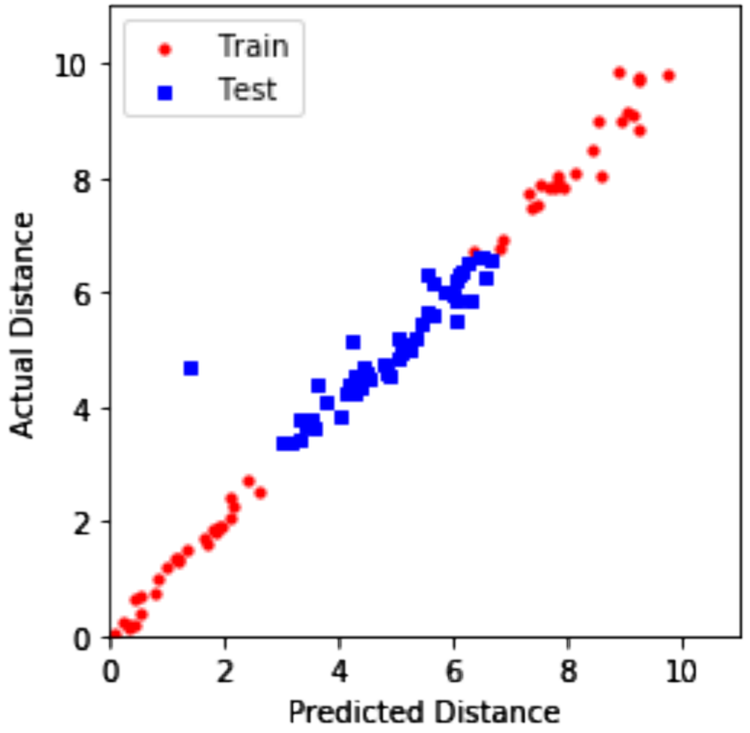



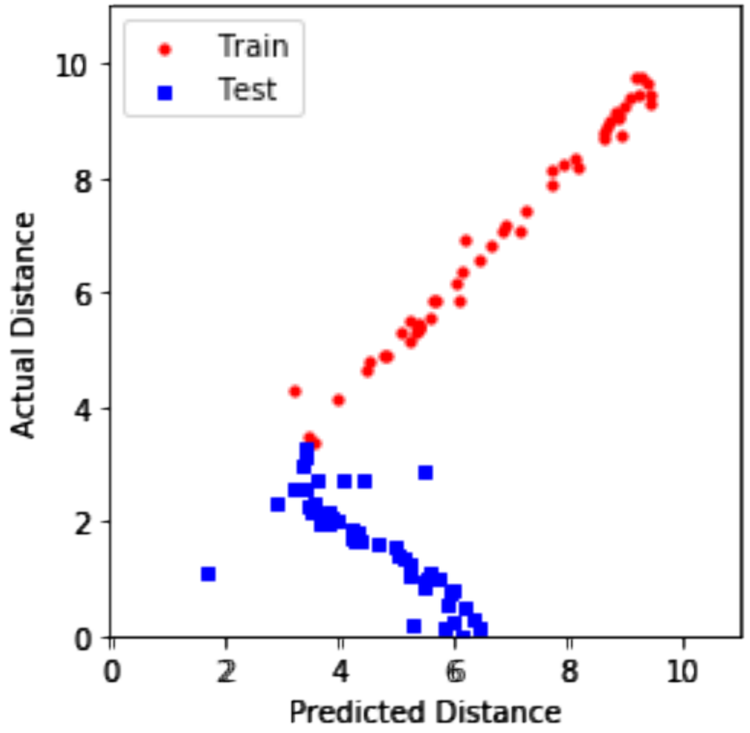



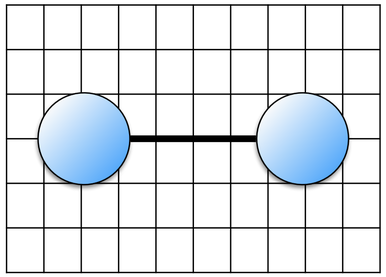
Distance

# An Aside…



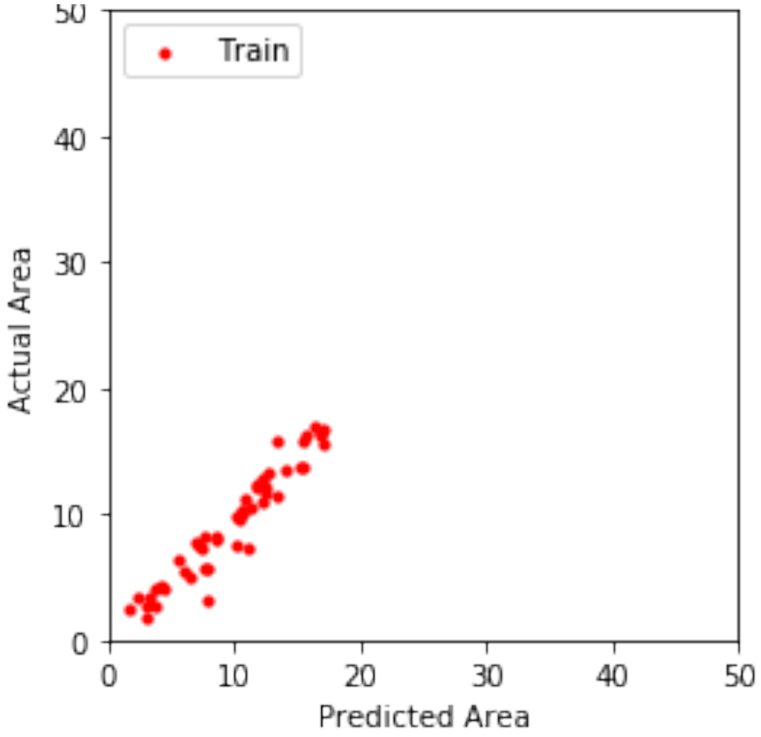



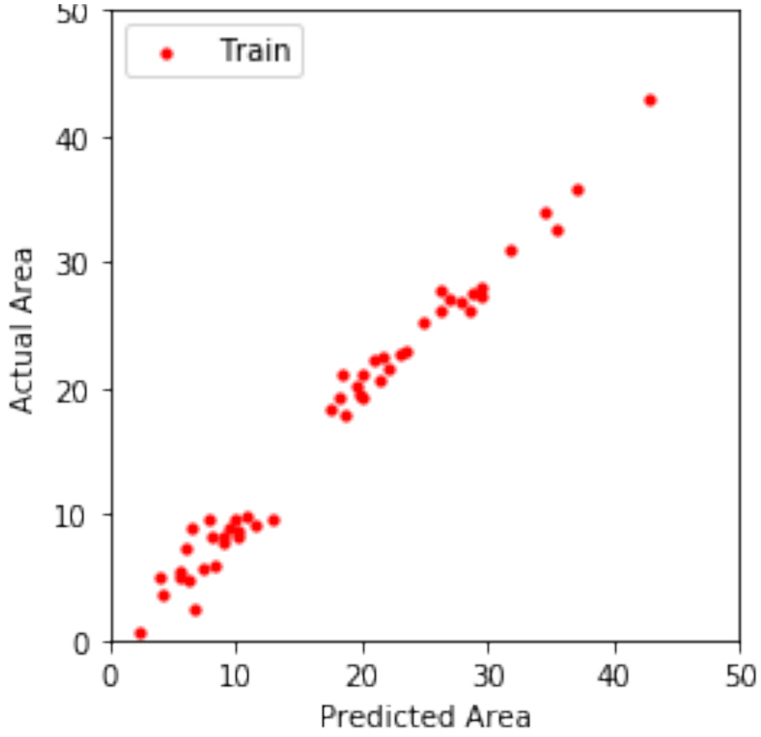



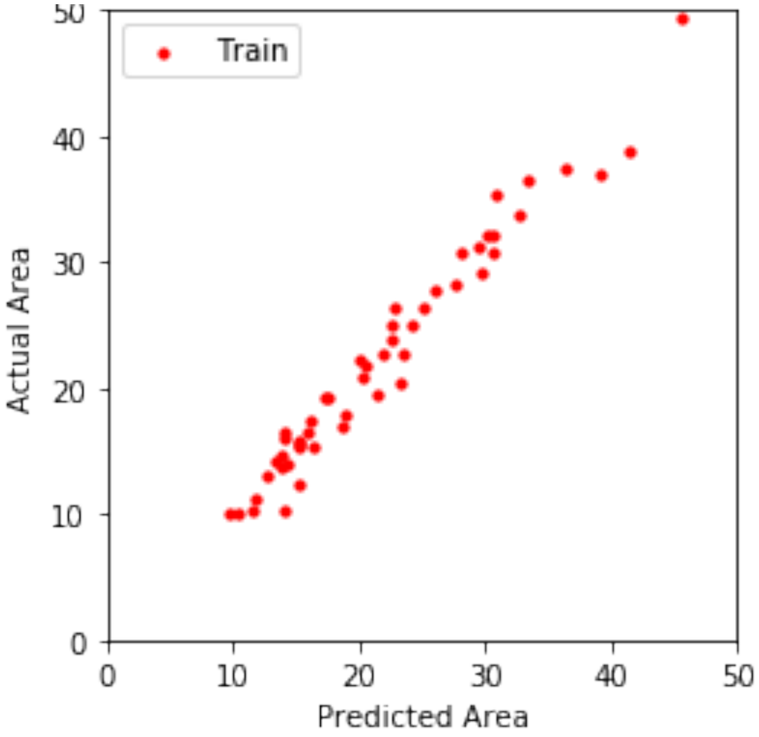



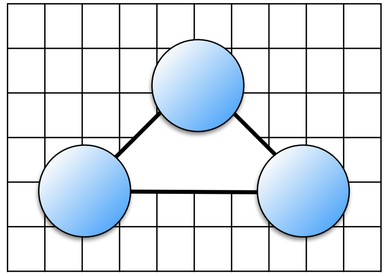
Distance

# An Aside…



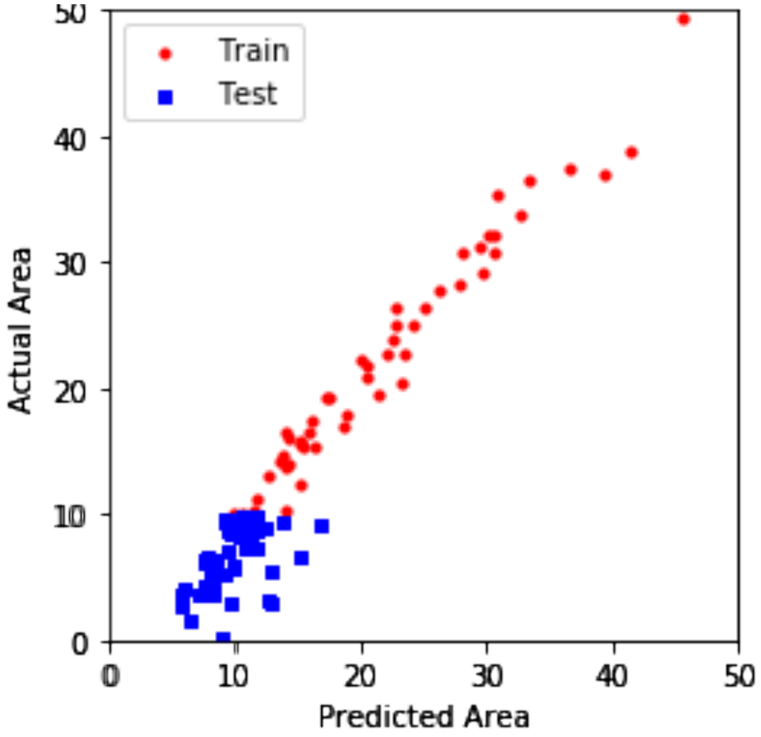



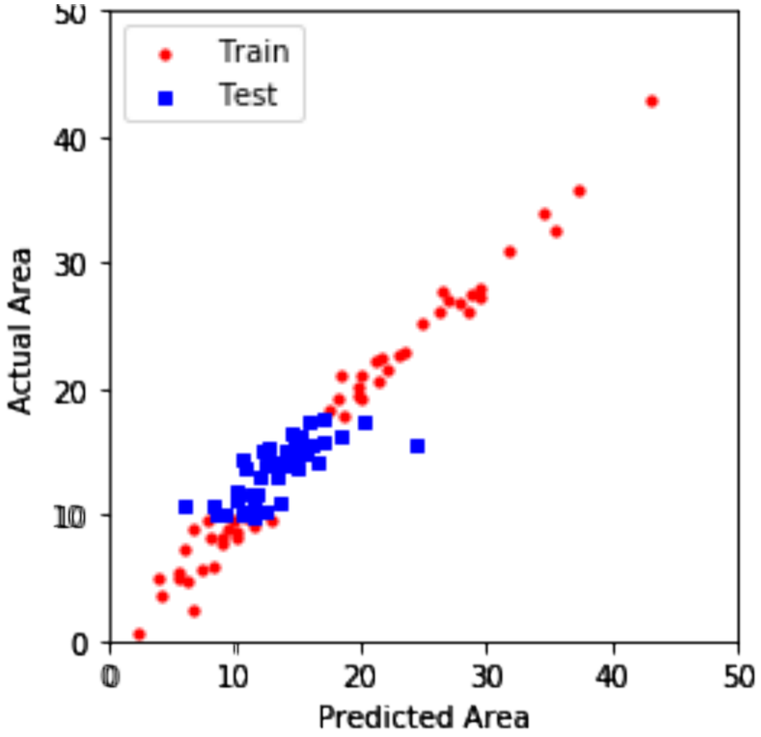



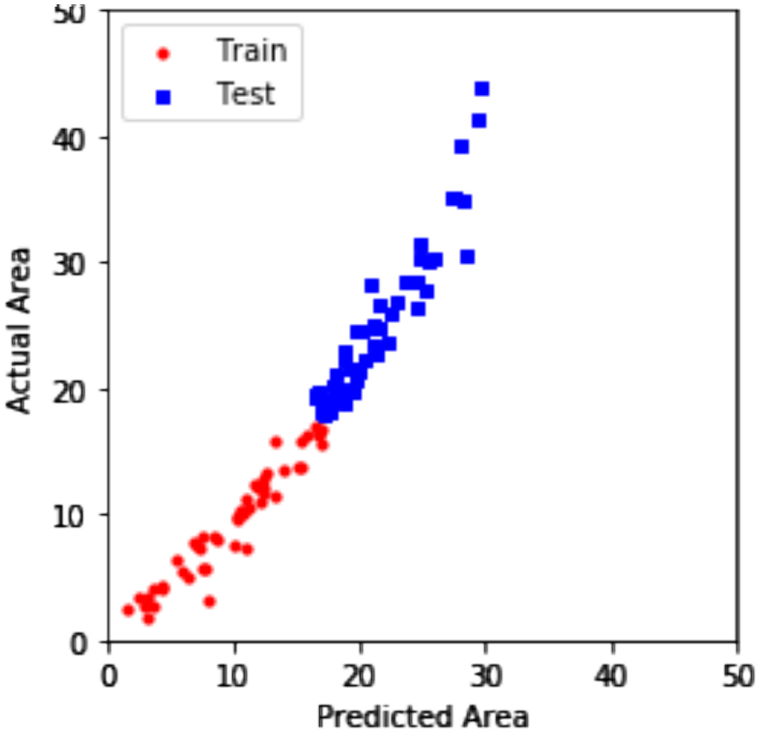



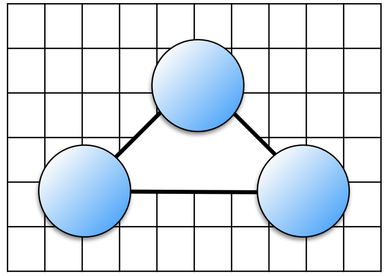
Area

# An Aside…



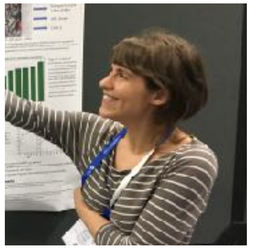



Area

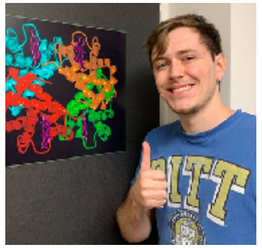
# Acknowledgements



Jocelyn Sunseri

Paul Francoeur

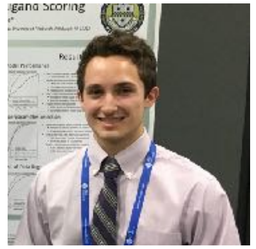
Josh Hochuli



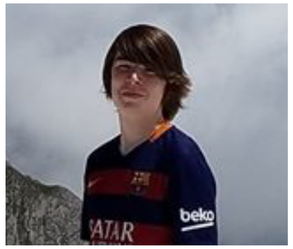
Hunter Haaf

Rachel Rosenzweig

Keshavan Seshadri



Matt Ragoza



Alec Helbling

Department of Computational and Systems Biology

National Institute of General Medical Sciences
R01GM108340

NVIDIA

Google Cloud

RELAY THERAPEUTICS

Do I have more time?
Do you care about chemistry education?

# Molecular Active Learning

ELSEVIER

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect

Nathaniel J. Hunsu*, Olusola Adesope, Dan James Bayly

*Educational Leadership, Sport Studies, Educational and Counseling Psychology, Washington State University, Pullman, WA 99164-4530, USA*

"Overall, we found small but significant effects of using ARS-based technologies on a number of desirable cognitive and non-cognitive learning outcomes."


Open Chemistry

Google Summer of Code

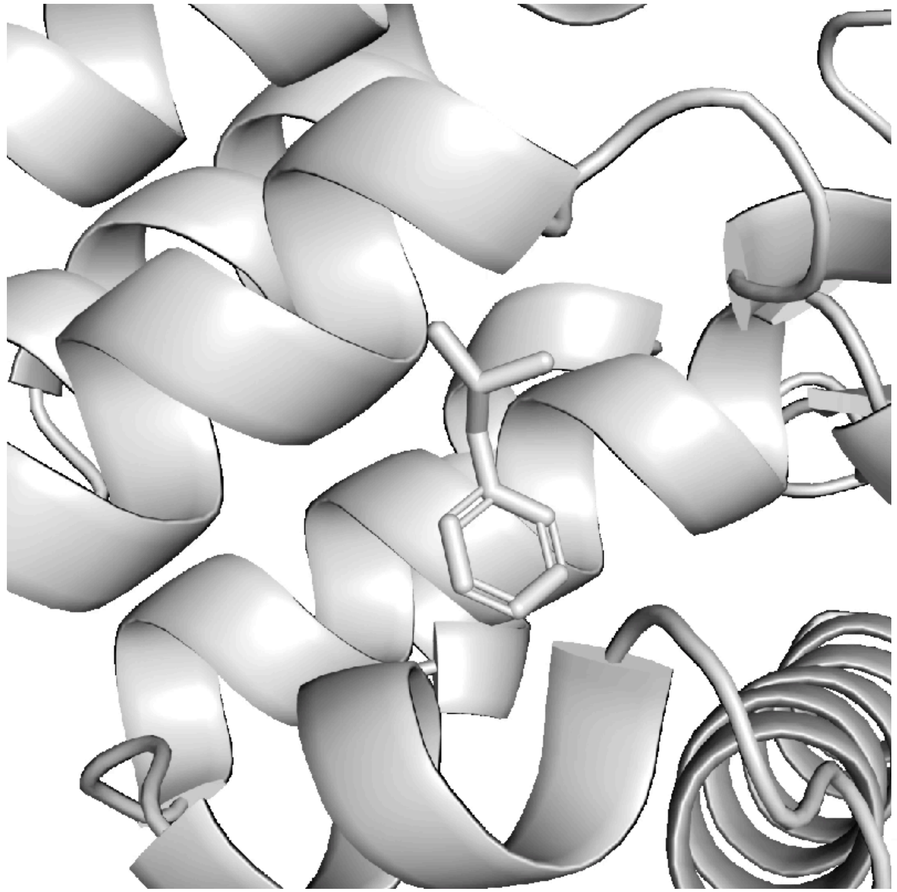
Go to this URL: http://3dmol.csb.pitt.edu/viewer.html

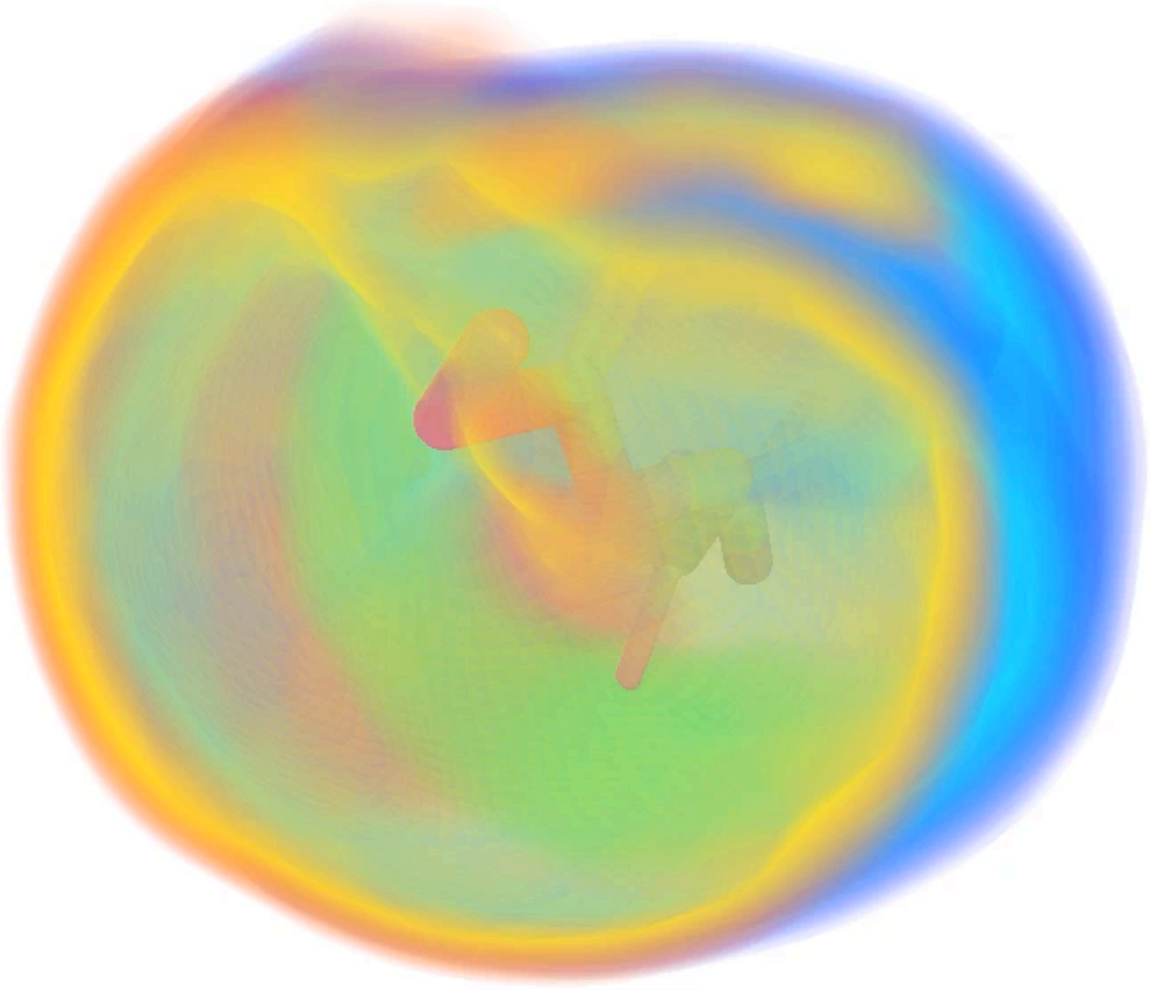
github.com/gnina

github.com/3dmol

http://bits.csb.pitt.edu

@david_koes





github.com/gnina

github.com/3dmol

http://bits.csb.pitt.edu

@david_koes