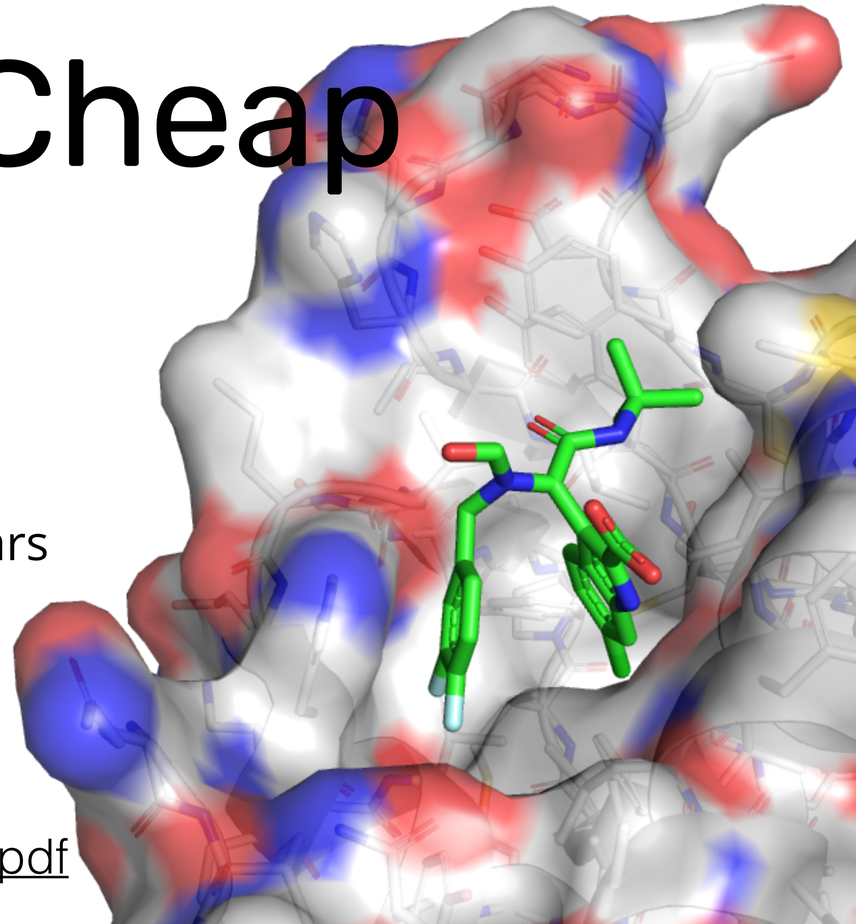
University of Pittsburgh

Structure-Based Virtual Screening on the Cheap

David Ryan Koes

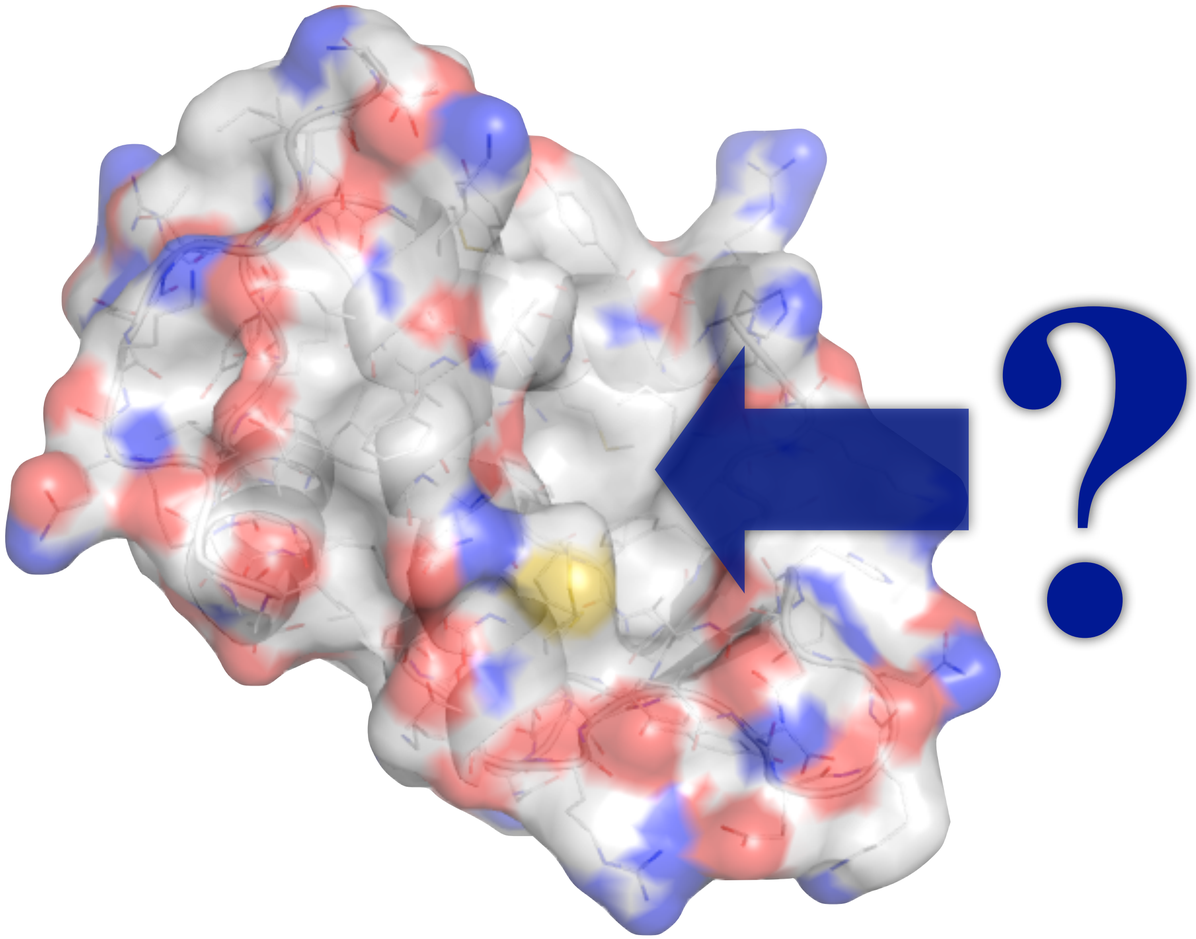
@david_koes

Global Health Compound Design Webinars
11/15/2021

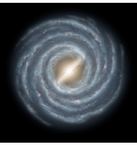


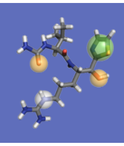
https://bits.csb.pitt.edu/files/on_the_cheap.pdf

Structure Based Drug Design

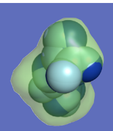


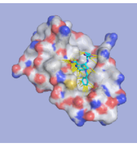
Structure Based Drug Design Funnel

Purchasable  Accessible

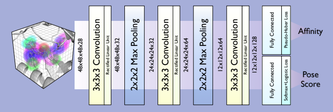


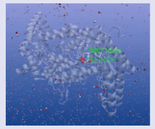
Matching



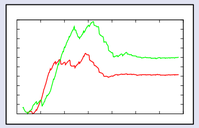


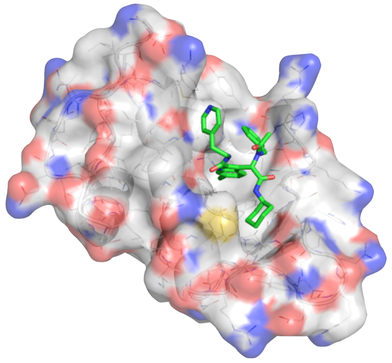
Scoring

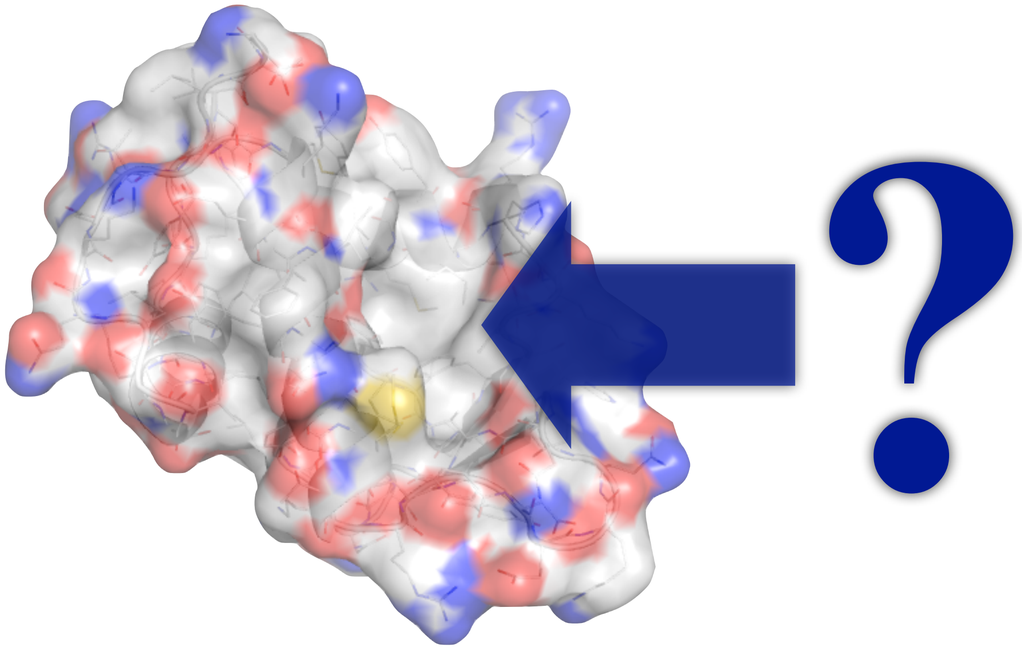




Dynamics

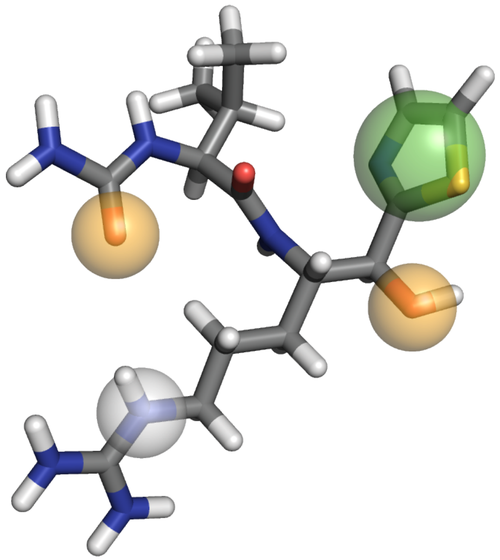


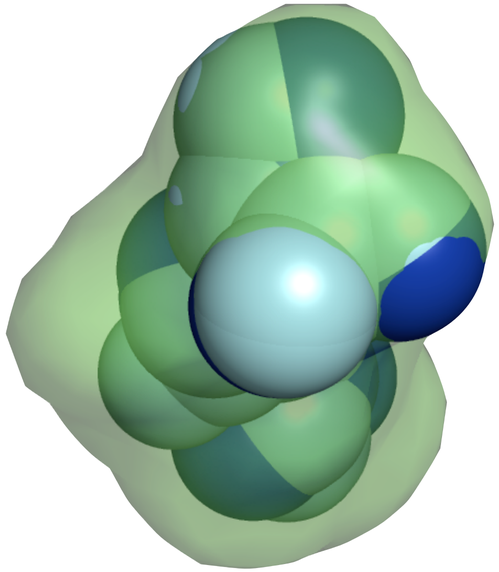


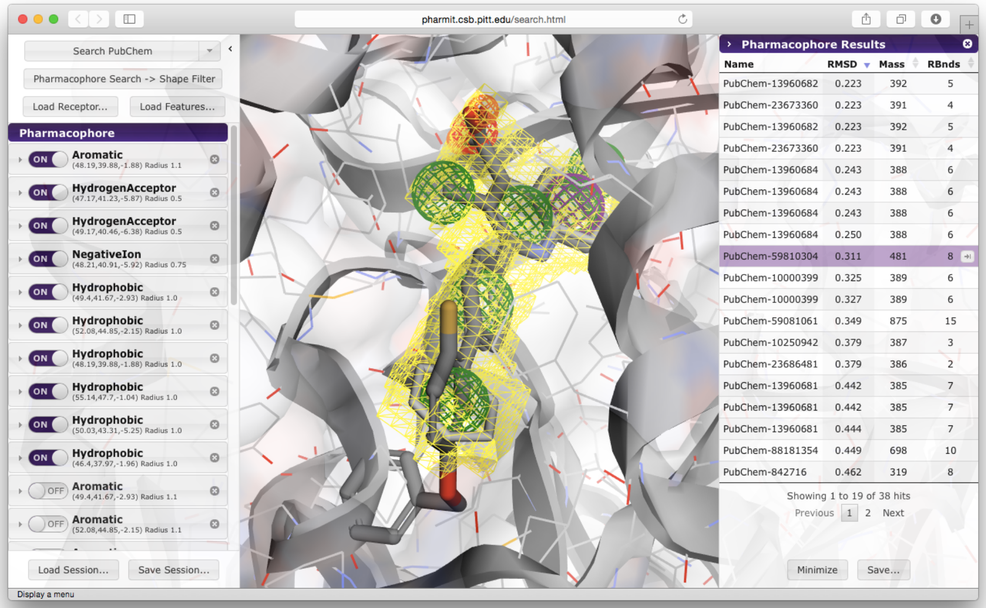


How does it bind? (pose prediction)
Does it bind? (virtual screening)
How well does it bind? (affinity prediction)

Matching







pharmer.sf.net

github.com/dkoes/shapedb

pharmit.csb.pitt.edu

Koes DR, Camacho CJ. Pharmer: efficient and exact pharmacophore search. J Chem Inf Model. 2011 Jun 27;51(6):1307-14.

Koes DR, Camacho CJ. Shape-based virtual screening with volumetric aligned molecular shapes. J Comput Chem. 2014 Sep 30;35(25):1824-34.

Hain E, Camacho CJ, Koes DR. Fragment oriented molecular shapes. J Mol Graph Model. 2016 May;66:143-54.

Koes DR, Camacho CJ. Indexing Volumetric Shapes with Matching and Packing. Knowl Inf Syst. 2015 Apr 1;43(1):157-180.

Rego N, Koes D. 3Dmol.js: molecular visualization with WebGL. Bioinformatics. 2015 Apr 15;31(8):1322-4.

Koes D. The Pharmit Backend: A Computer Systems Approach to Enabling Interactive Online Drug Discovery. IBM Journal of Research and Development. 2018 Nov 28.

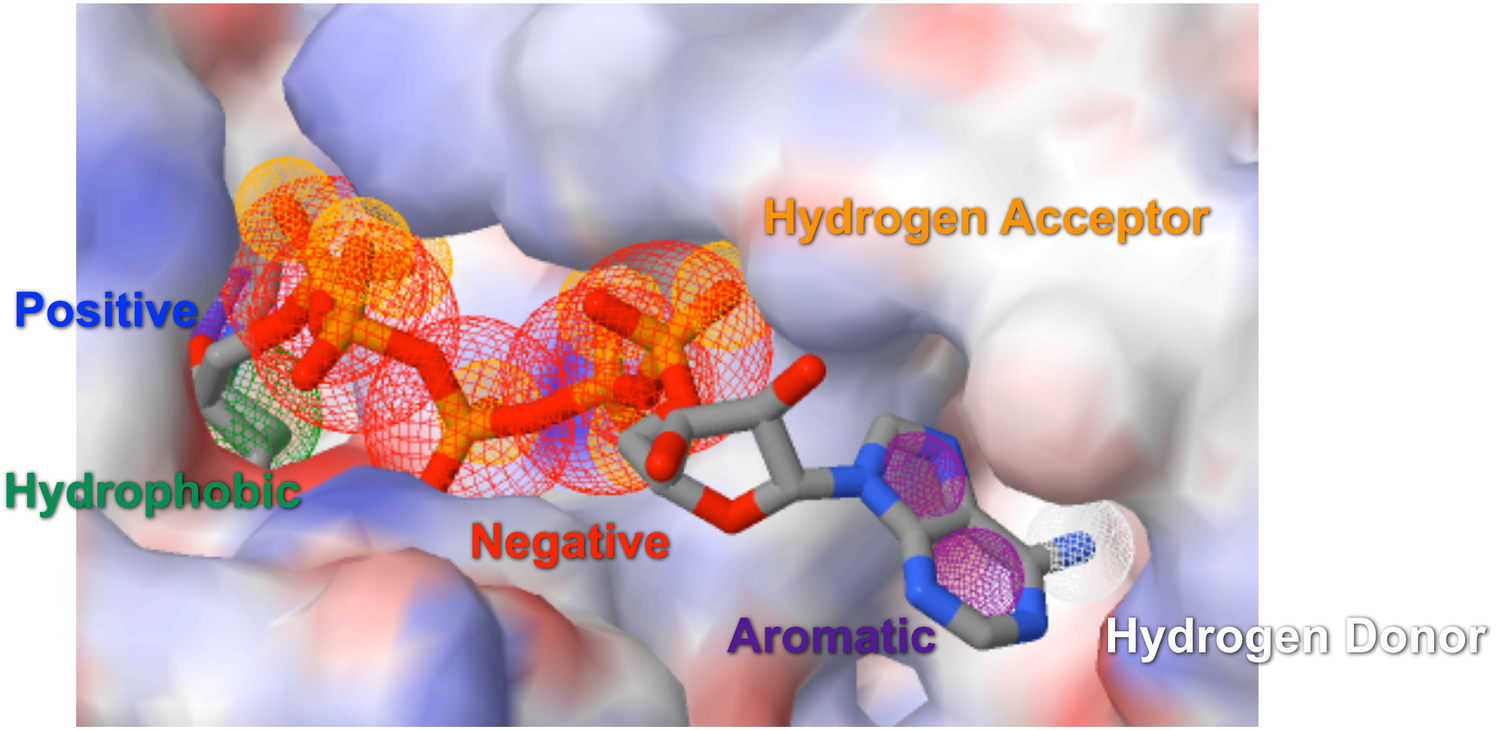
Sunseri J, Koes DR. Pharmit: interactive exploration of chemical space. Nucleic Acids Res. 2016 Jul 8;44(W1):W442-8.

Pharmacophore

IUPAC: The ensemble of steric and electronic features that is necessary to ensure the optimal supra-molecular interactions with a specific biological target structure and to trigger (or to block) its biological response.

Common Features:

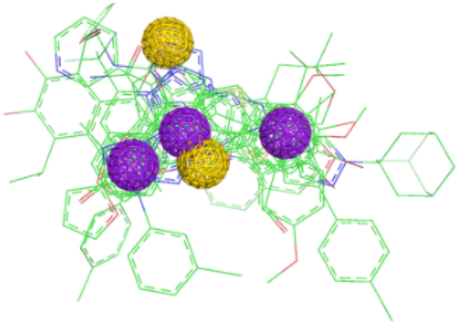
aromatic ring
hydrophobic area
positive ionizable
negative ionizable
hydrogen bond donor
hydrogen bond acceptor



Pharmacophore Elucidation

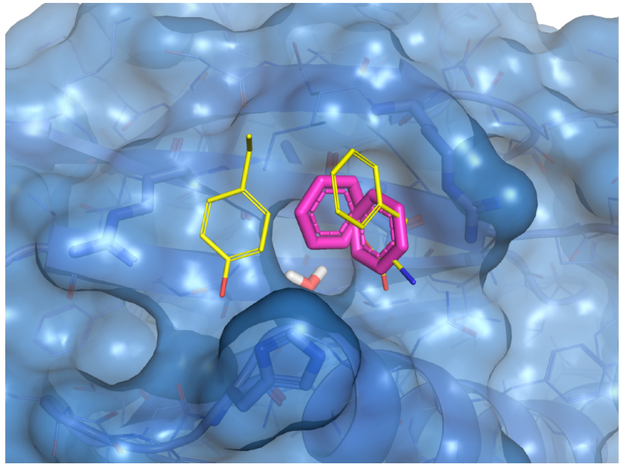
Ligand Only

e.g. PharmaGist, PENG, GAPE, MOGA

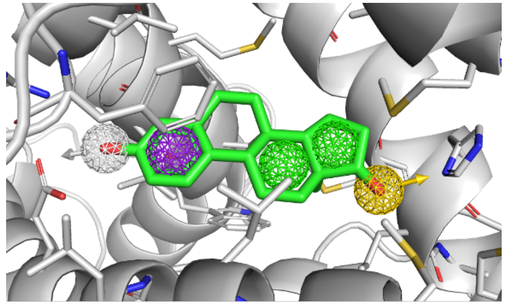


Receptor Only

Simulation: Pharmmaker, SILCS, PyRod
Docking: FTMap
Grids: GRAIL

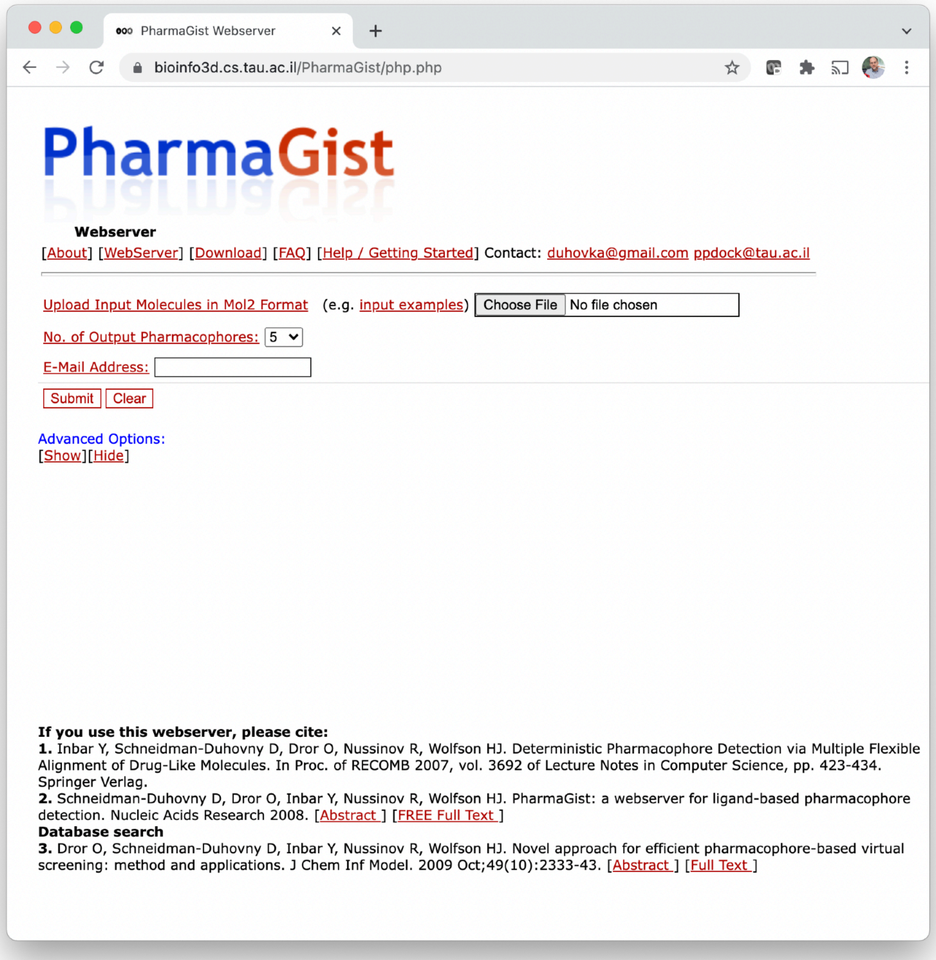


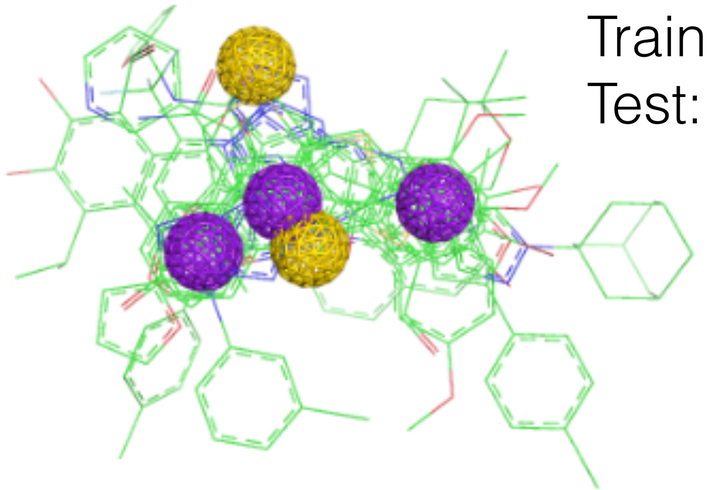
Ligand-Receptor



Koes DR. Pharmacophore modeling: methods and applications. Computer-Aided Drug Discovery. 2015:167-88.

Ligand Only

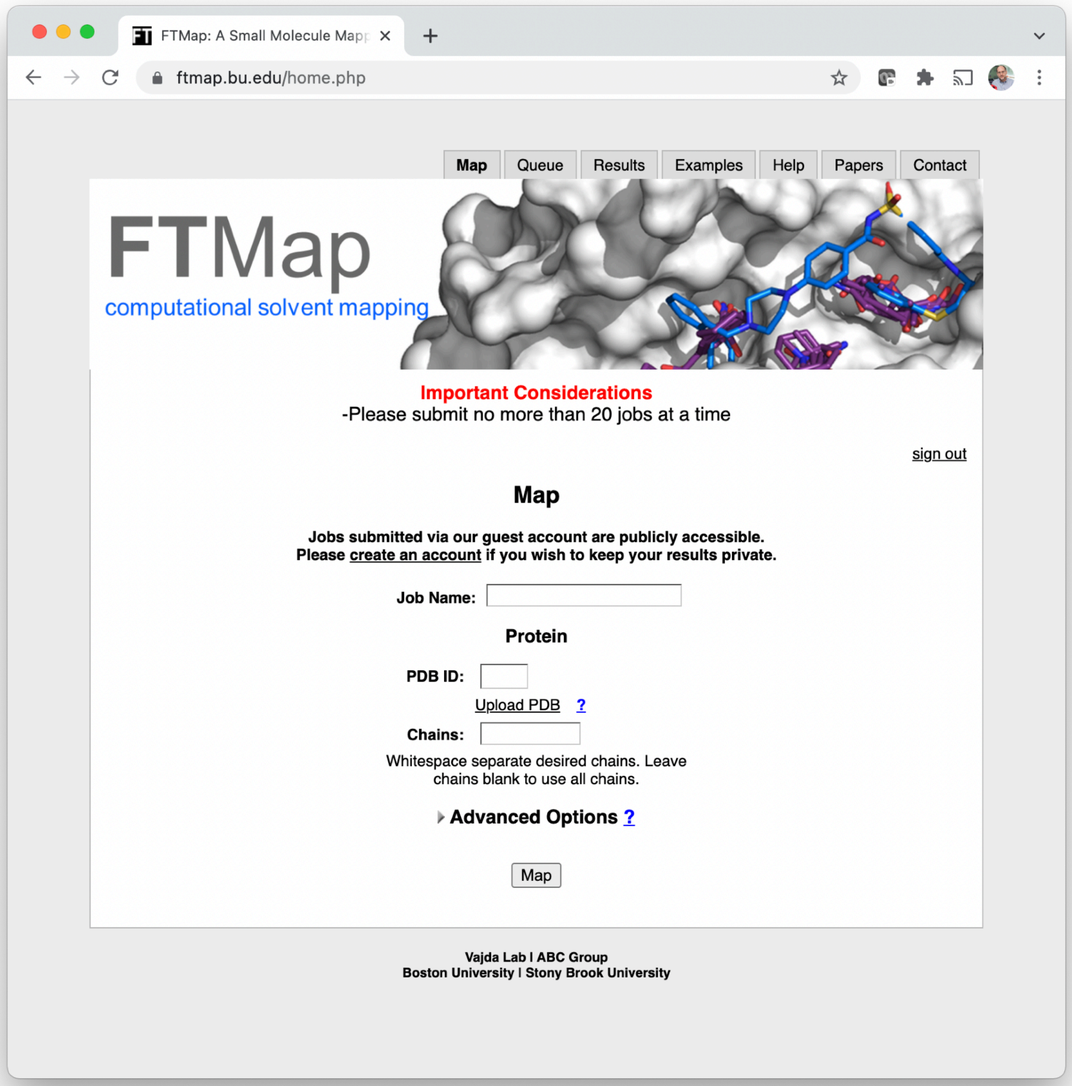


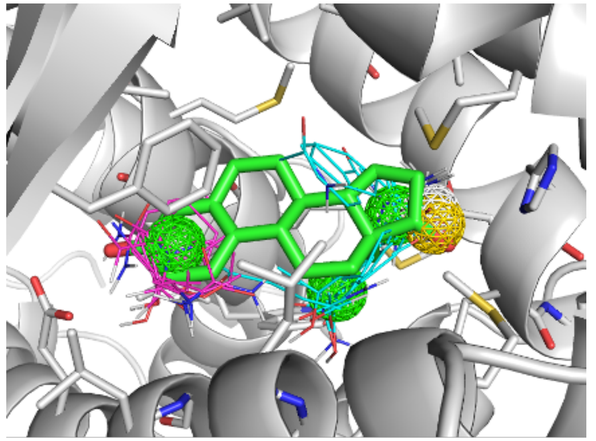


Estrogen Receptor Alpha
Training: PubChem Assay 713
Test: DUD-E esr1

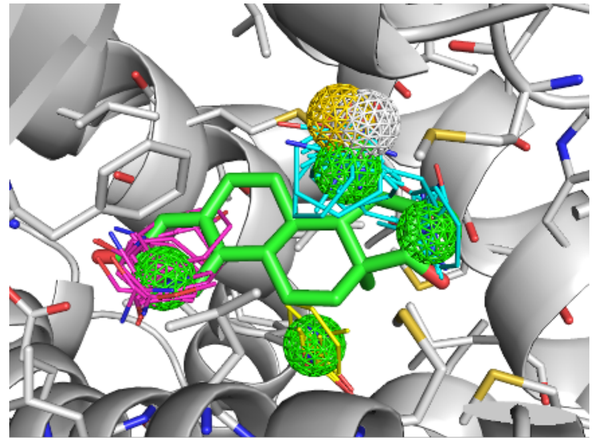
Features	Ligands	TP	FP	Precision	Recall	F1	EF
5	9	1	178	0.56%	0.26%	0.004	**0.309**
4	18	41	2166	1.86%	10.70%	0.032	**1.041**
3	27	131	8083	1.59%	34.20%	0.031	**0.892**

Receptor Only





Bound (PDB 1QKU)

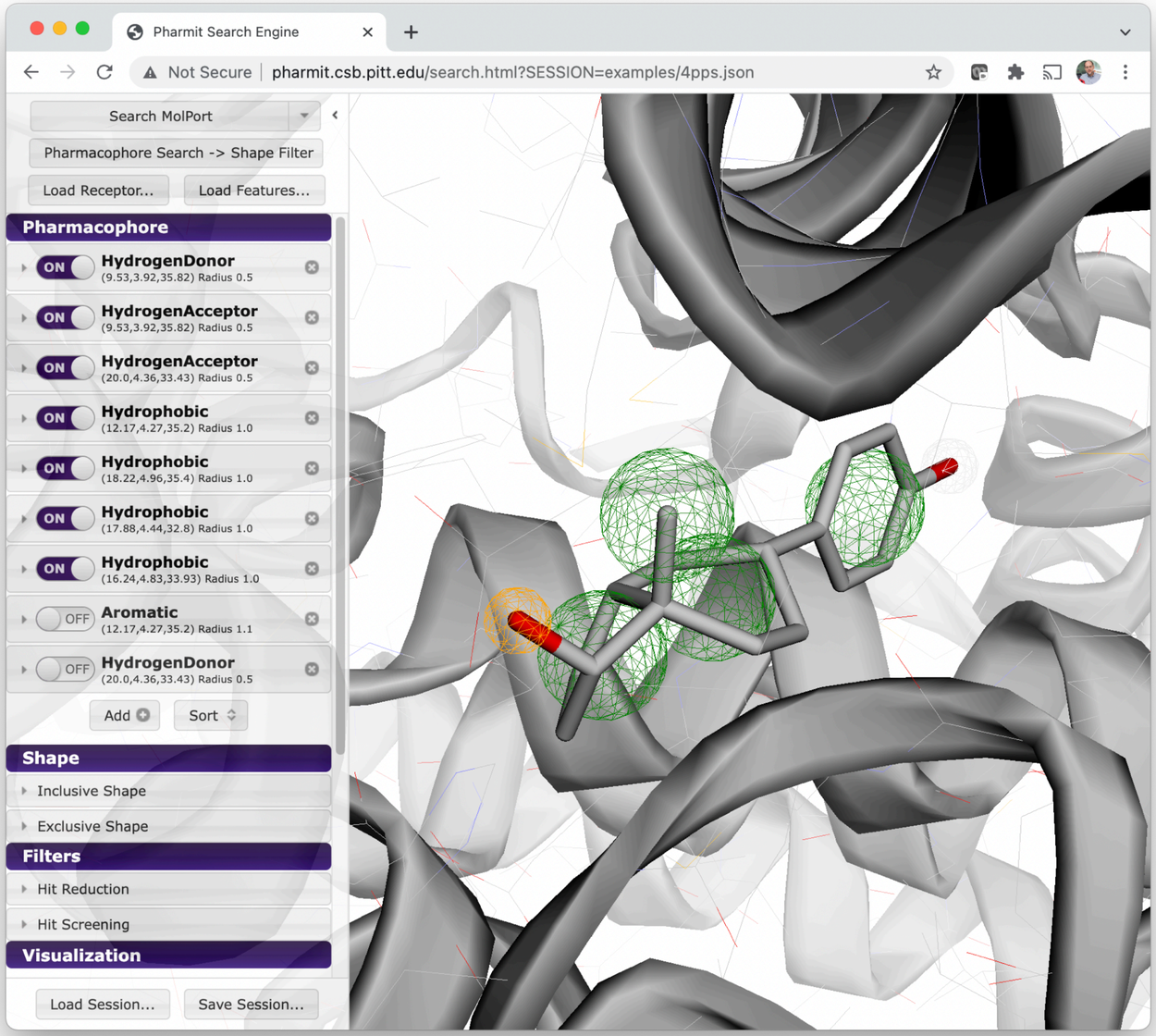


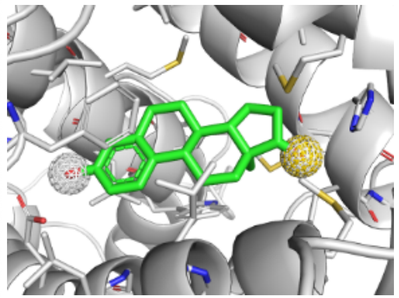
Unbound (PDB 2B23)

estradiol shown for reference

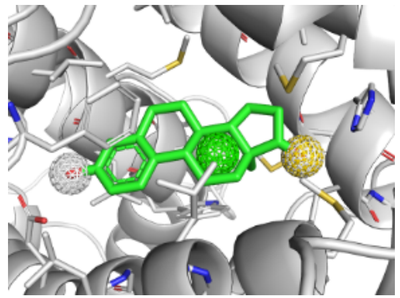
	Features	TP	FP	Precision	Recall	F1	EF
Bound	5	54	265	16.93%	14.10%	0.154	**11.209**
Unbound	6	5	13	27.78%	1.31%	0.025	**21.157**

Ligand-Receptor

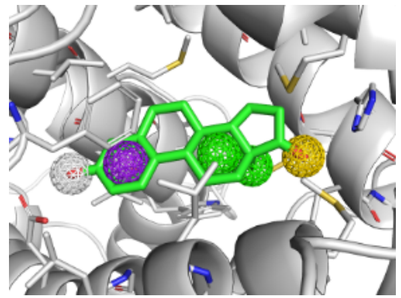




3



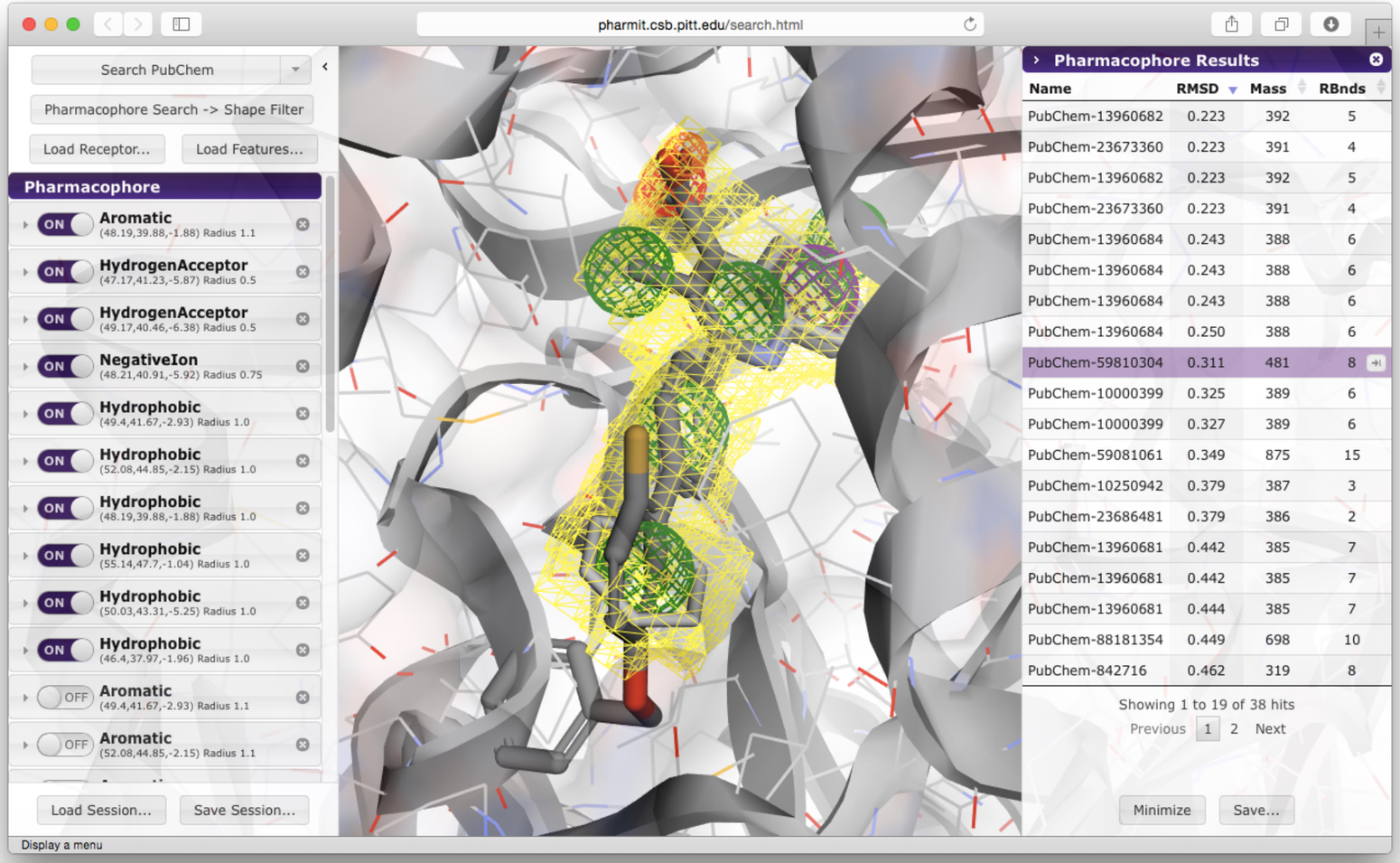
4



5

Features	Train F1	TP	FP	Precision	Recall	F1	EF
3	0.189	214	537	28.50%	55.87%	0.377	**21.921**
4	0.301	140	139	50.18%	36.55%	0.423	**55.404**
5	0.326	71	50	58.68%	18.54%	0.282	**78.111**

Pharmit Demo



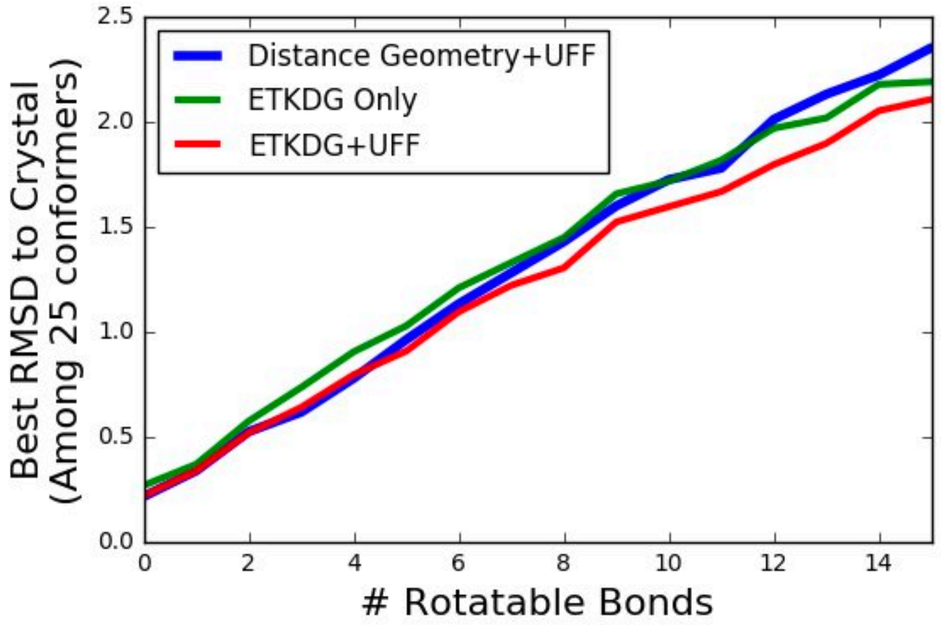
Key Points

- Molecule Loading
- Feature Definition
- Shape Constraints
- Filters
- Visualization
- Minimization

Pharmit Database Prep

Input: SMILES (length < 250)

- Tautomers: What's a tautomer?
  - default OpenBabel protonation
- Stereoisomers
  - Respect SMILE specification
  - Sample unspecified steriochemistry implicitly during conformer generation
- Conformers
  - Sample maximum of 20 conformers using RDKit



Protein-Ligand Scoring

AutoDock Vina

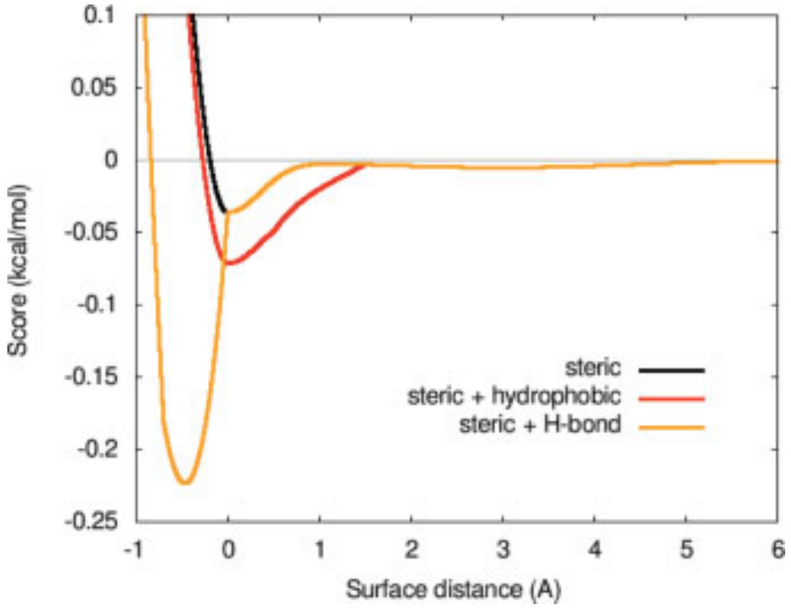
$$\text{gauss}_1(d) = w_{\text{guass}_1} e^{-(d/0.5)^2}$$

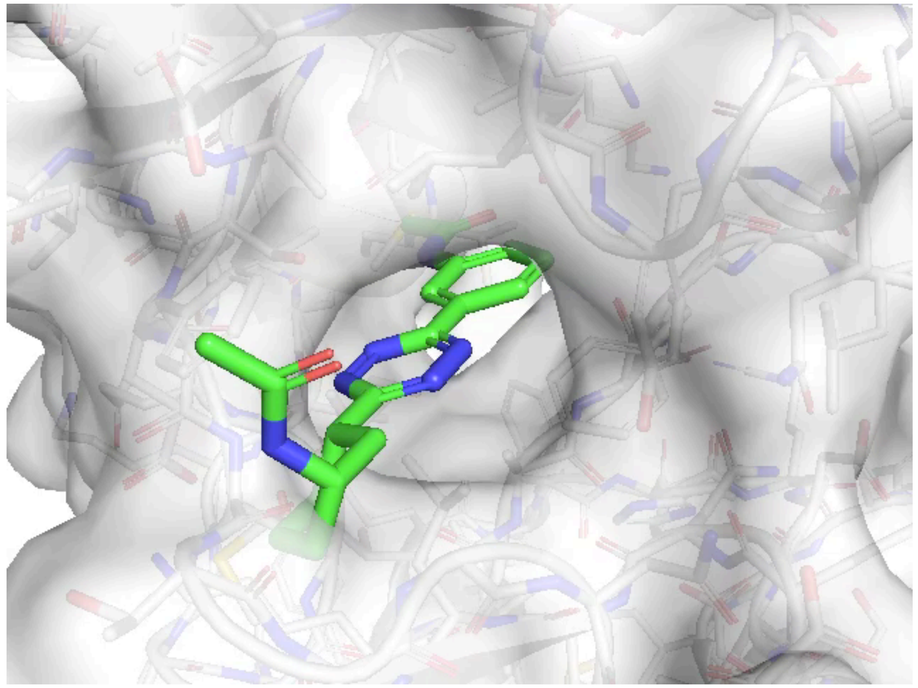
$$\text{gauss}_2(d) = w_{\text{guass}_2} e^{-((d-3)/2)^2}$$

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

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

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





Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*. 2010 Jan 30;31(2):455-61.

Protein-Ligand Scoring

AutoDock Vina

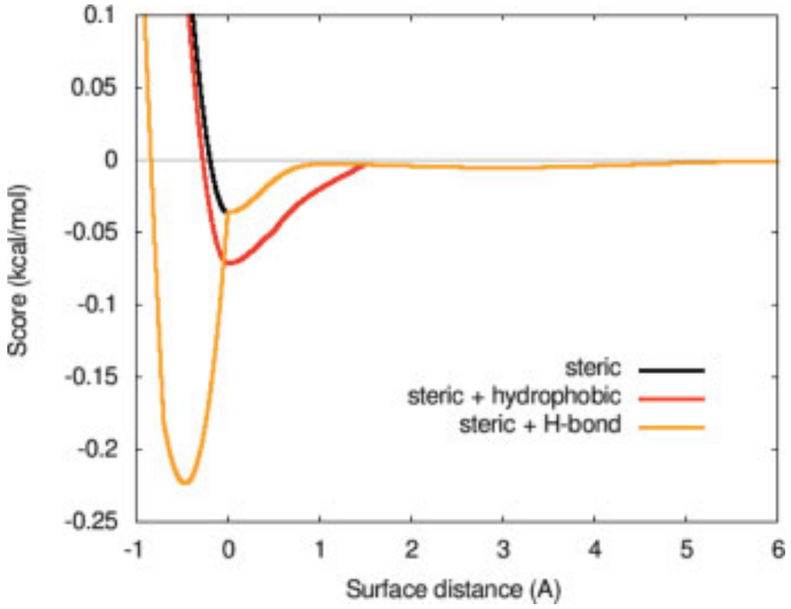
$$\text{gauss}_1(d) = w_{\text{guass}_1} e^{-(d/0.5)^2}$$

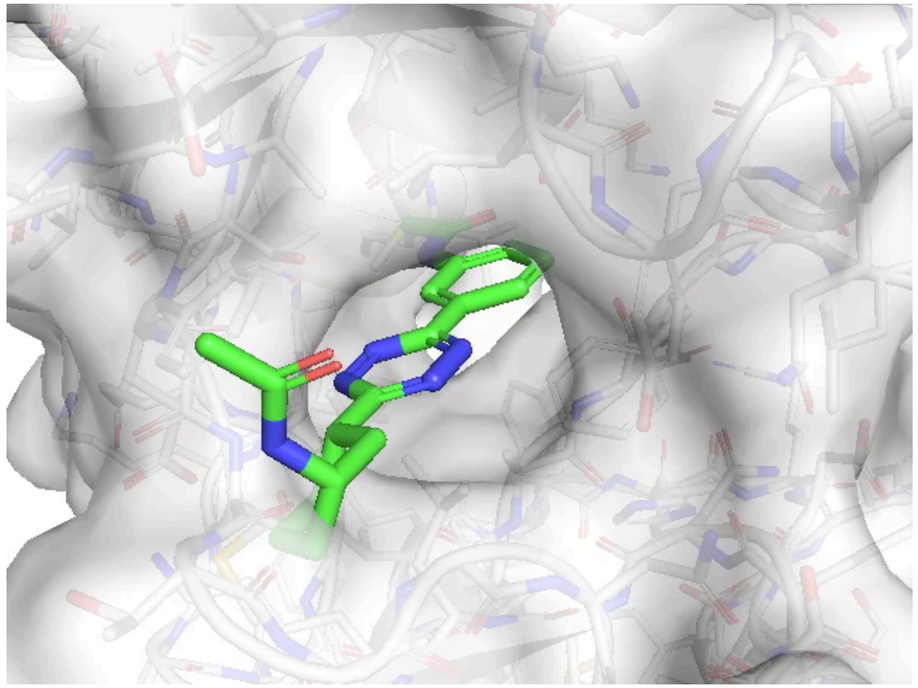
$$\text{gauss}_2(d) = w_{\text{guass}_2} e^{-((d-3)/2)^2}$$

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

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

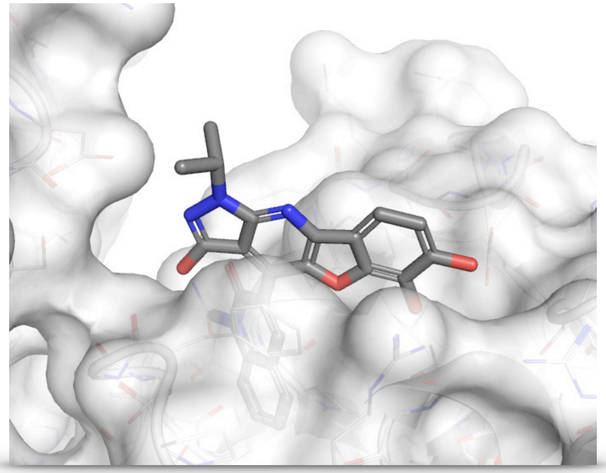
$$\text{hbond}(d) = \begin{cases} w_{\text{hbond}} & d < -0.7 \\ 0 & d > 0 \\ w_{\text{hbond}}(-\frac{10}{7} d) & otherwise \end{cases}$$





Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*. 2010 Jan 30;31(2):455-61.

Protein-Ligand Scoring



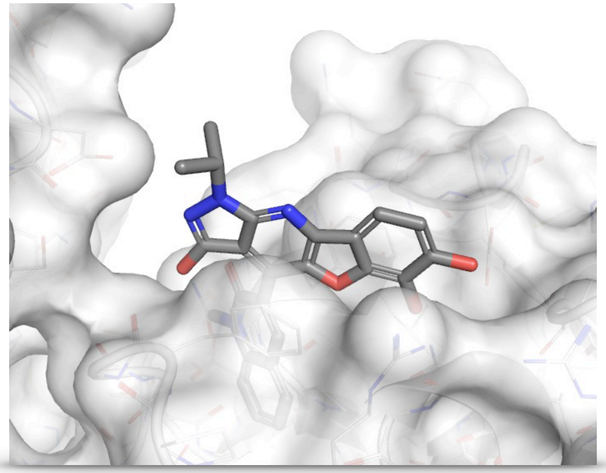
Model

Pose Prediction

Binding Discrimination

Affinity Prediction

Protein-Ligand Scoring



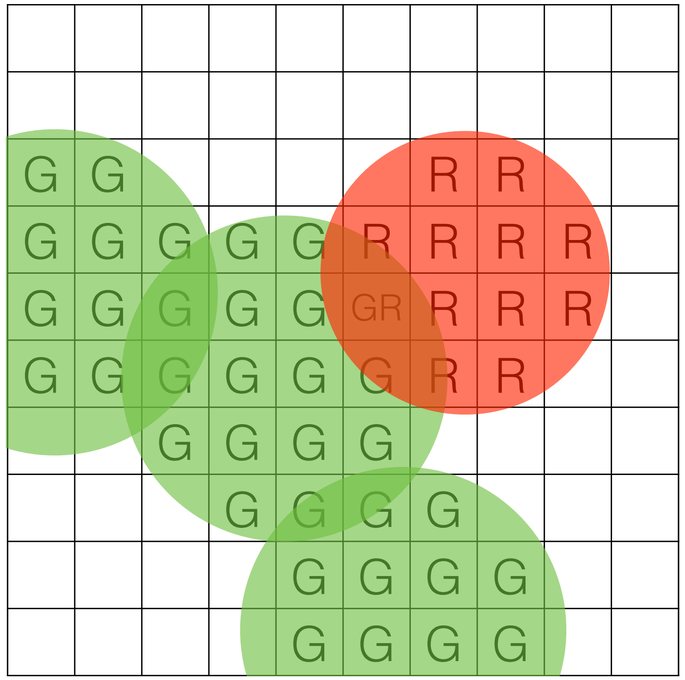
Model

Pose Prediction

Binding Discrimination

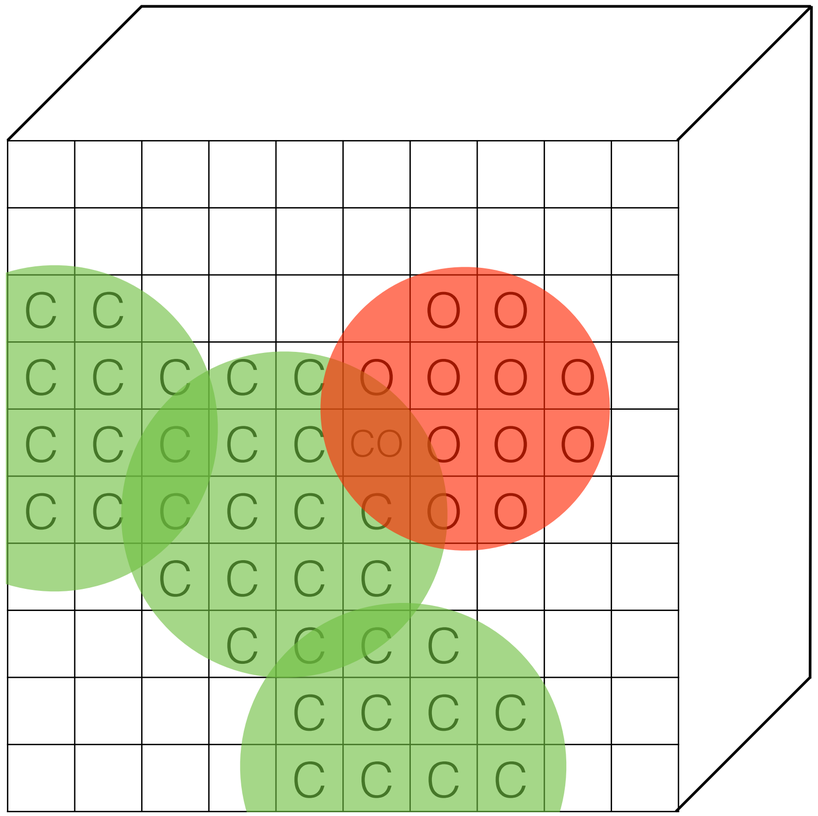
Affinity Prediction

Protein-Ligand Representation



(R,G,B) pixel

Protein-Ligand Representation



(R,G,B) pixel →

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

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

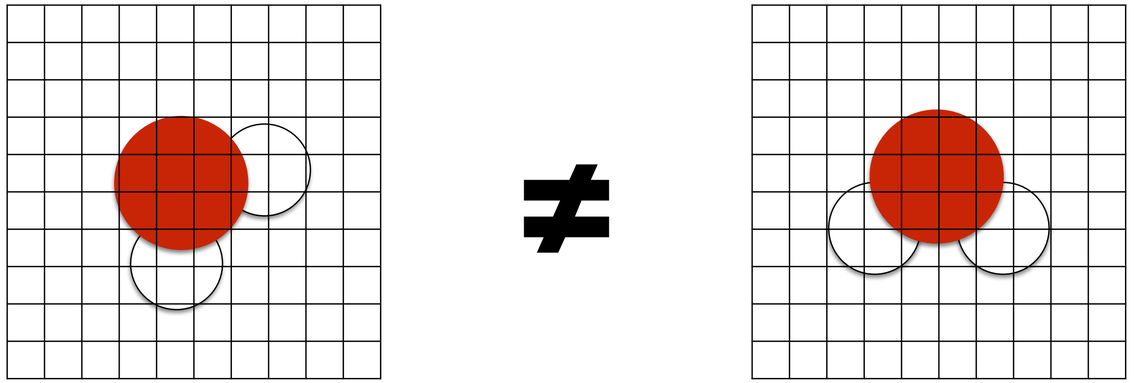
Why Grids?

Cons

- *coordinate frame dependent*
  - but can be input into equivariant networks
- pairwise interactions not explicit

Pros

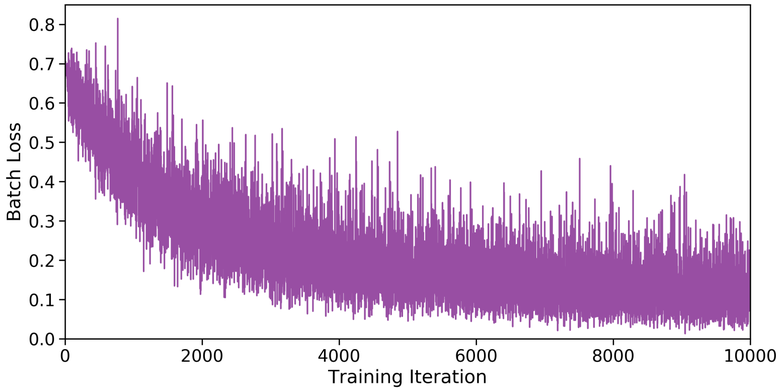
- clear spatial relationships
- amazingly parallel
- easy to interpret



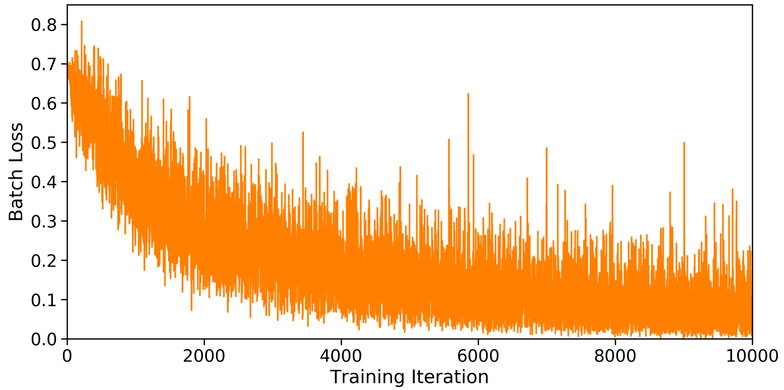
libmolgrid

Dr. Jocelyn Sunseri

Caffe Training



PyTorch Training



JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

libmolgrid: Graphics Processing Unit Accelerated Molecular Gridding for Deep Learning Applications

Jocelyn Sunseri and David R. Koes*

Cite this: *J. Chem. Inf. Model.* 2020, 60, 3, 1079–1084

Publication Date: February 12, 2020

https://doi.org/10.1021/acs.jcim.9b01145

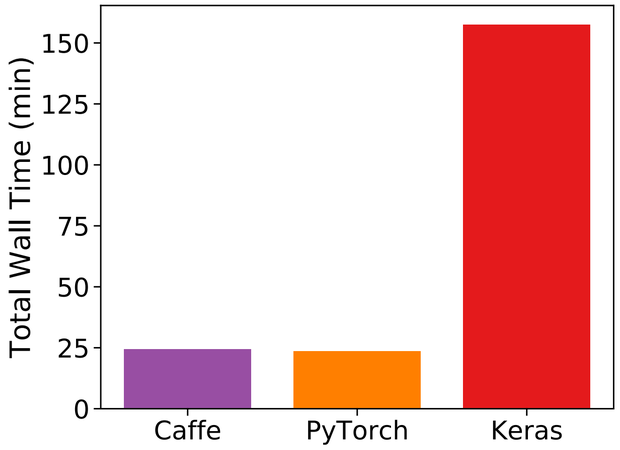
Copyright © 2020 American Chemical Society

RIGHTS & PERMISSIONS ACS Editors' Choice

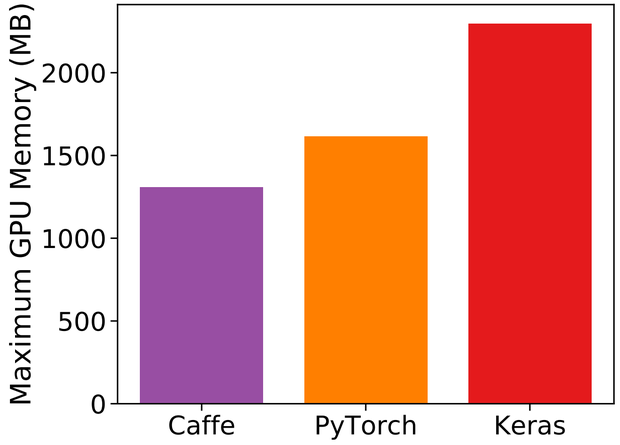
Article Views	Altmetric	Citations
3038	13	2

LEARN ABOUT THESE METRICS

GPU Performance



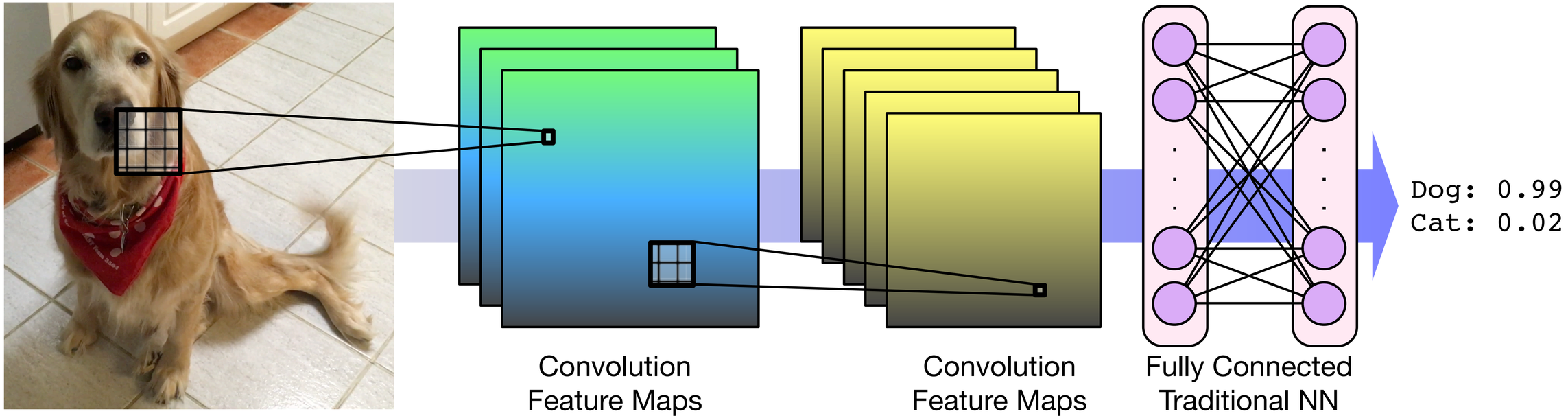
GPU Memory Utilization



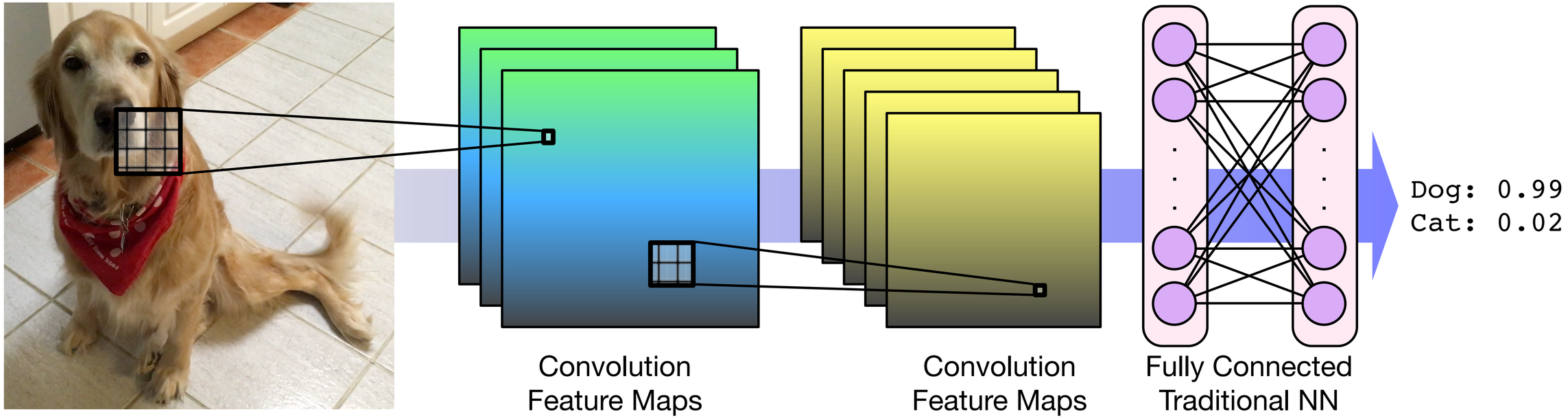

```
e = molgrid.ExampleProvider(balanced=True,shuffle=True)
e.populate('examples.txt')

gmaker = molgrid.GridMaker(resolution=0.5,
                           dimension=23.5)

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

github.com/gnina/libmolgrid

Convolutional Neural Networks



Convolutional Neural Networks



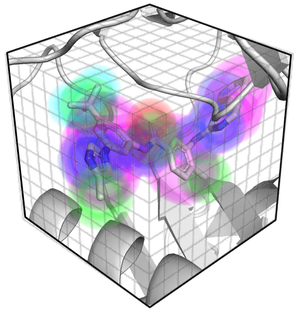
Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR. Protein–ligand scoring with convolutional neural networks. Journal of chemical information and modeling. 2017 Apr 11;57(4):942-57.

Sunseri J, King JE, Francoeur PG, Koes DR. Convolutional neural network scoring and minimization in the D3R 2017 community challenge. Journal of computer-aided molecular design. 2019 Jan 15;33(1):19-34.

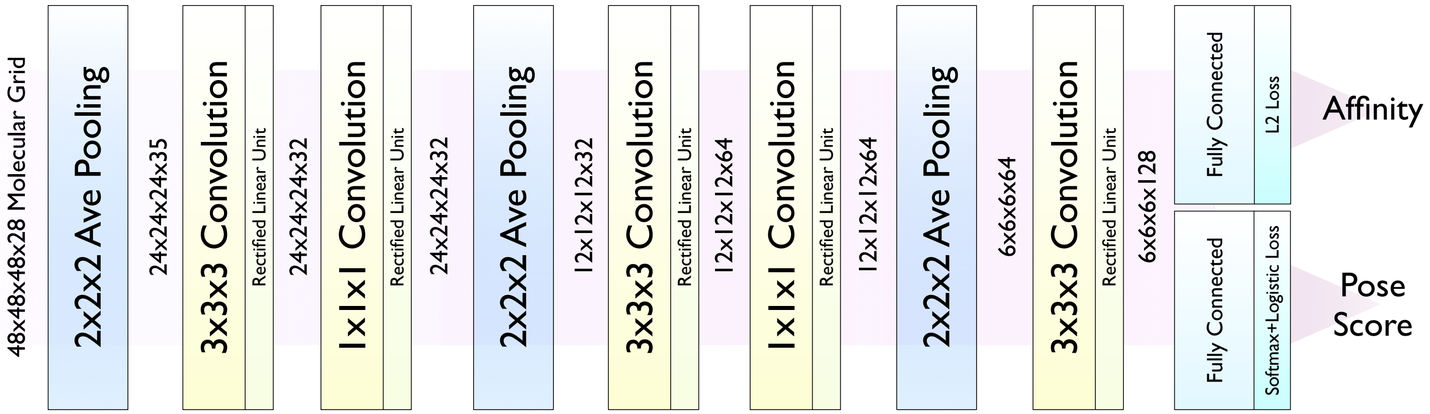
Hochuli J, Helbling A, Skaist T, Ragoza M, Koes DR. Visualizing convolutional neural network protein-ligand scoring. Journal of Molecular Graphics and Modelling. 2018 Sep 1;84:96-108.

Francoeur PG, Masuda T, Sunseri J, Jia A, Iovanisci RB, Snyder I, Koes DR. Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design. Journal of Chemical Information and Modeling. 2020 Aug 31;60(9):4200-15.

Protein Ligand Scoring

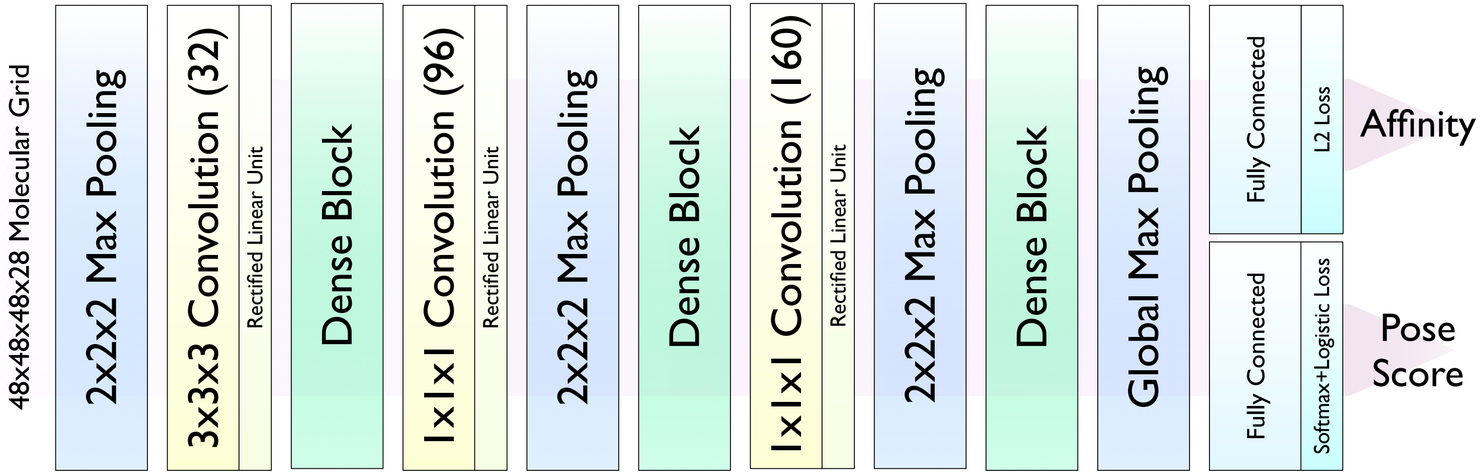


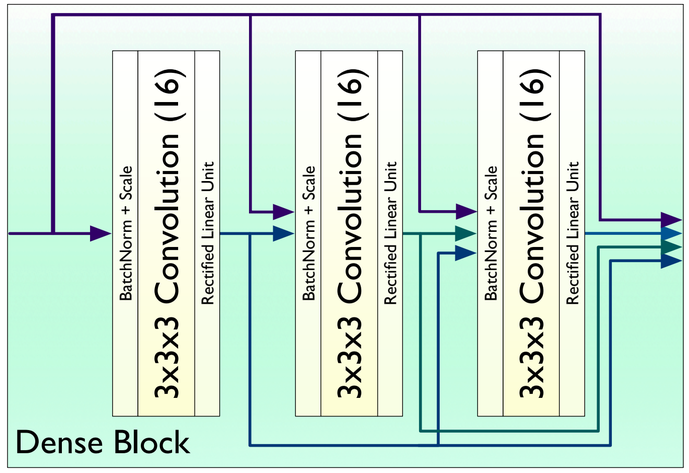
Def2018



388,736 Parameters

Dense





684,640 Parameters

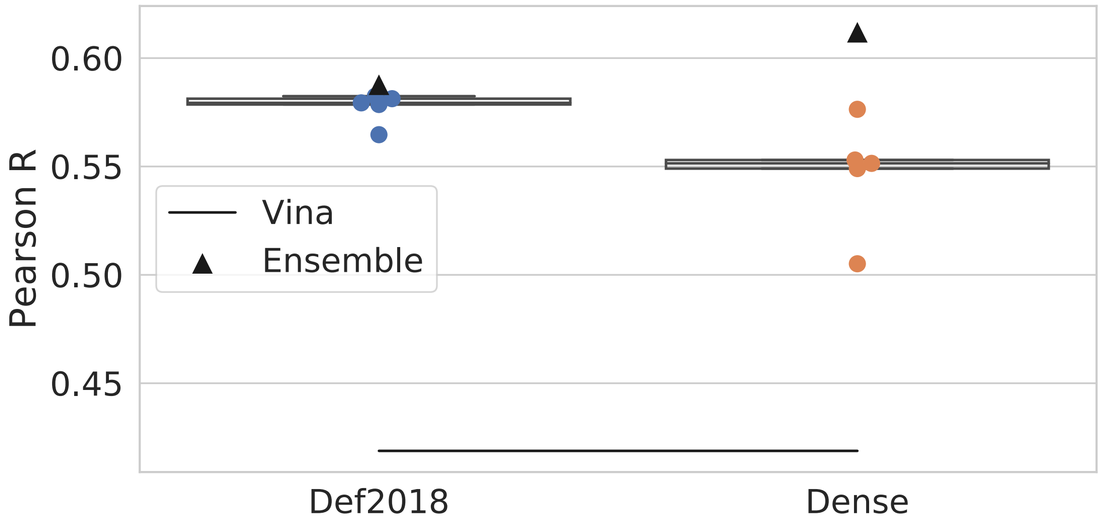
Cross-Docked Protein Ligand Scoring

https://github.com/gnina/models/tree/master/data/CrossDocked2020

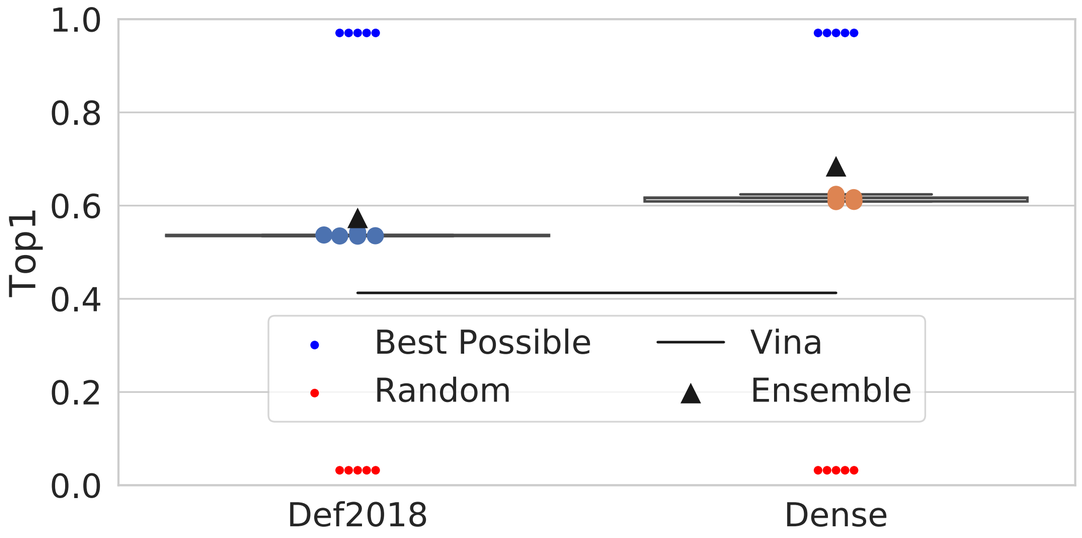
Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design

Paul G. Francoeur, Tomohide Masuda, Jocelyn Sunseri, Andrew Jia, Richard B. Iovanisci, Ian Snyder, and David R. Koes*

Cite this: *J. Chem. Inf. Model.* 2020, 60, 9, 4200–4215
Publication Date: August 31, 2020
https://doi.org/10.1021/acs.jcim.0c00411



Affinity Prediction



Pose Selection

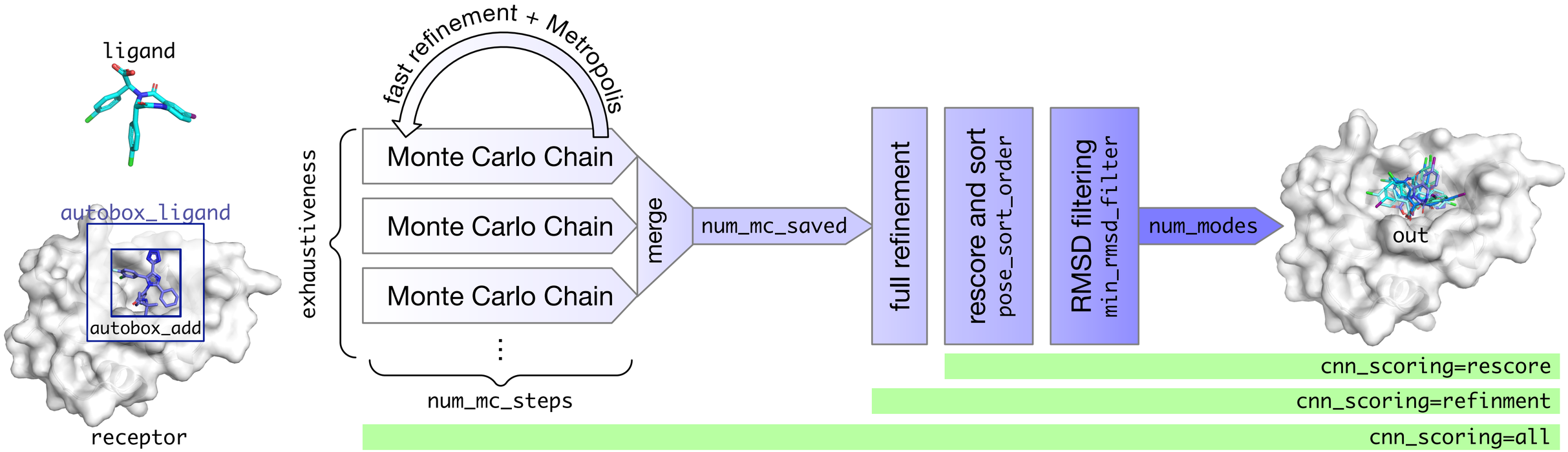
GNINA 1.0

https://github.com/gnina/gnina

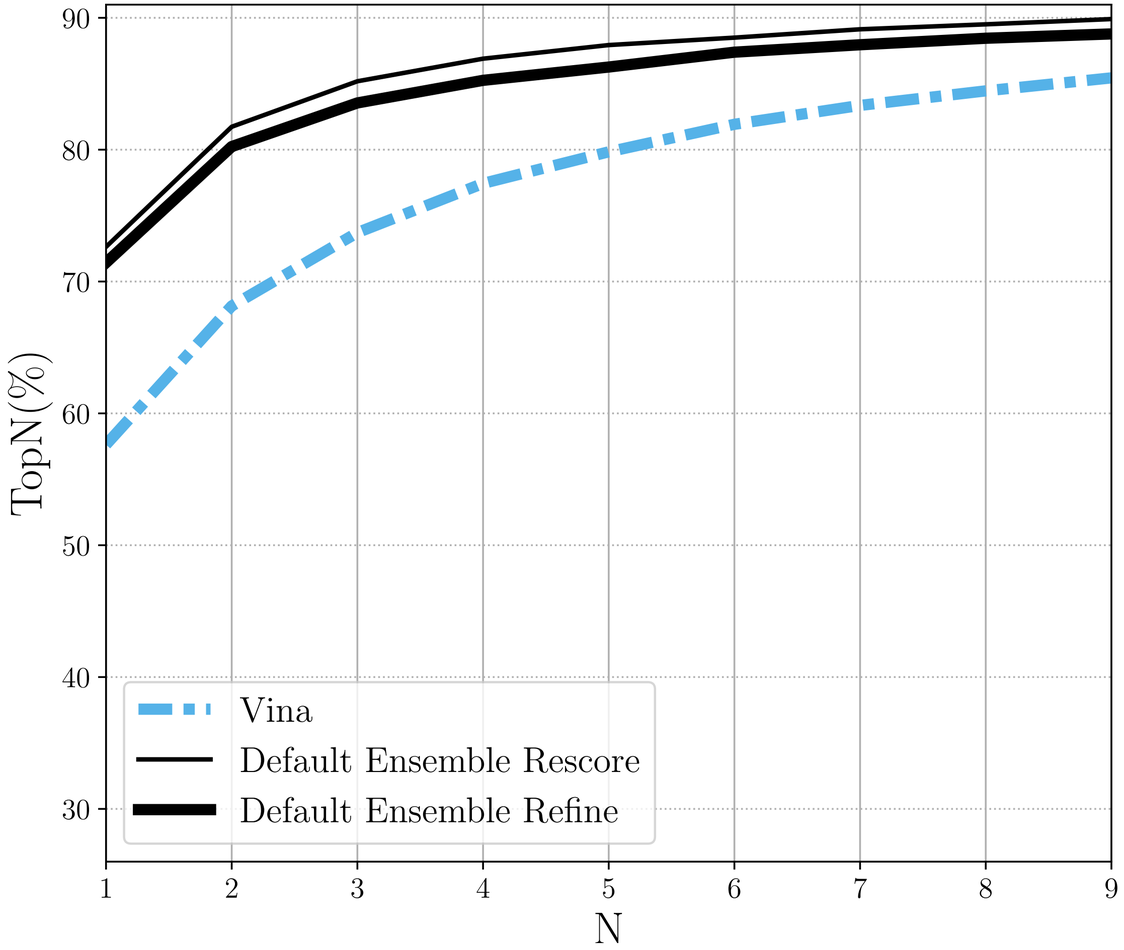
SOFTWARE Open Access

GNINA 1.0: molecular docking with deep learning

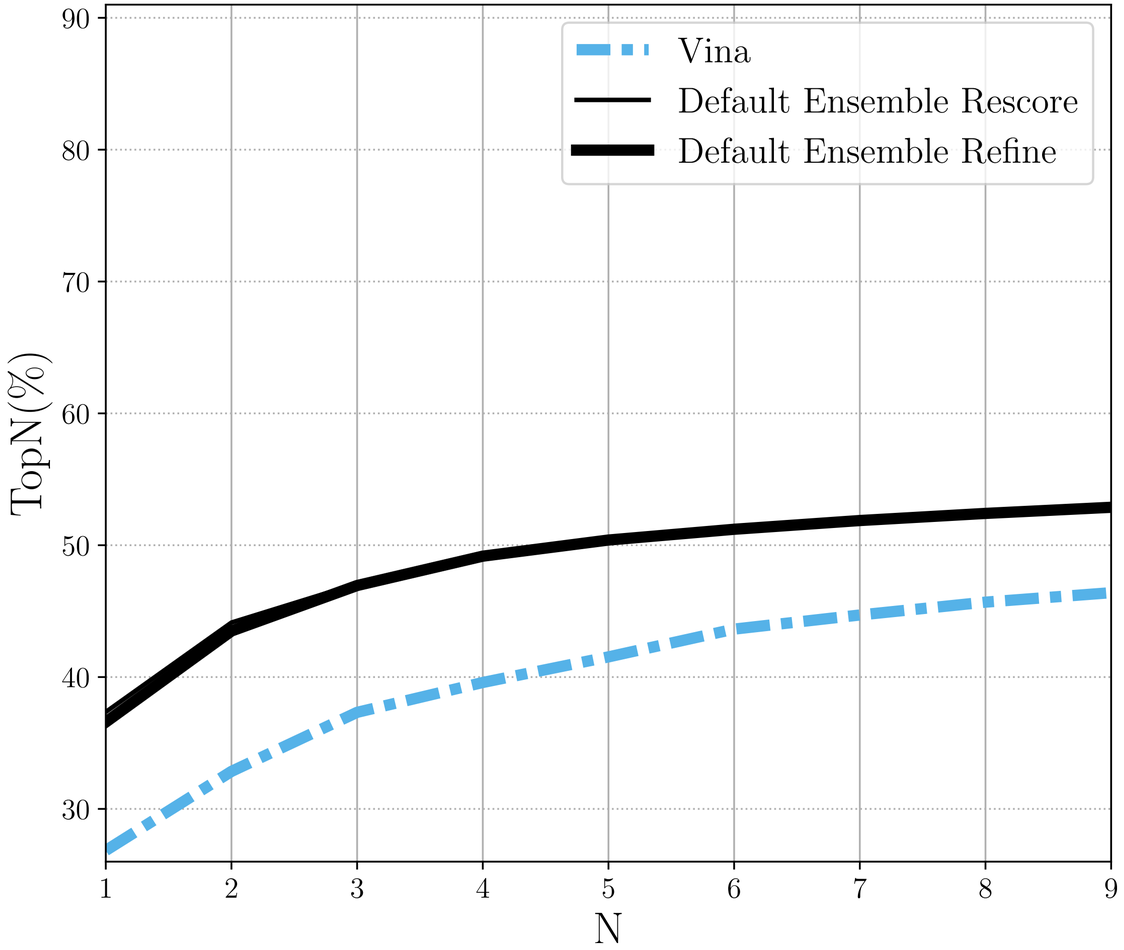
Andrew T. McNutt[1], Paul Francoeur[1], Rishal Aggarwal[2], Tomohide Masuda[1], Rocco Meli[3], Matthew Ragoza[1], Jocelyn Sunseri[1] and David Ryan Koes[1*]



Using the CNN

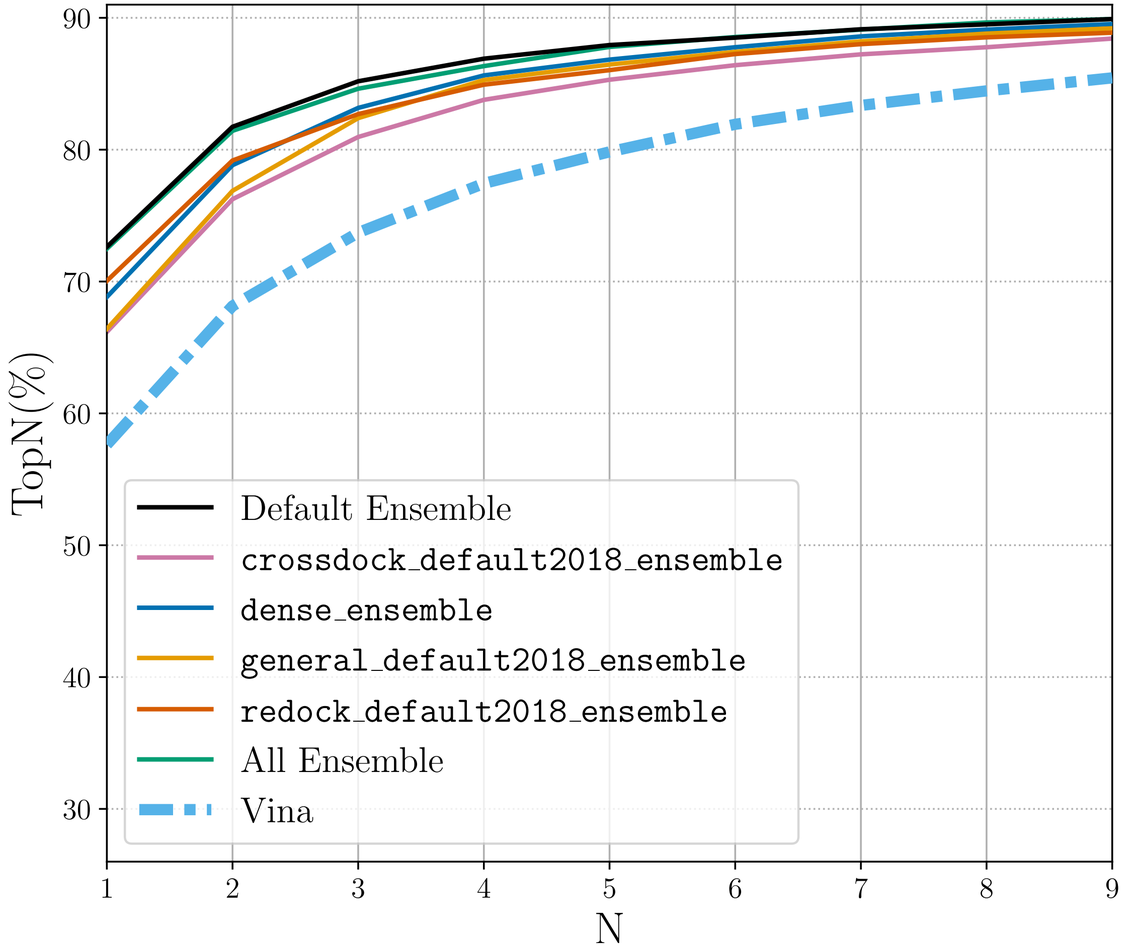


Redocking

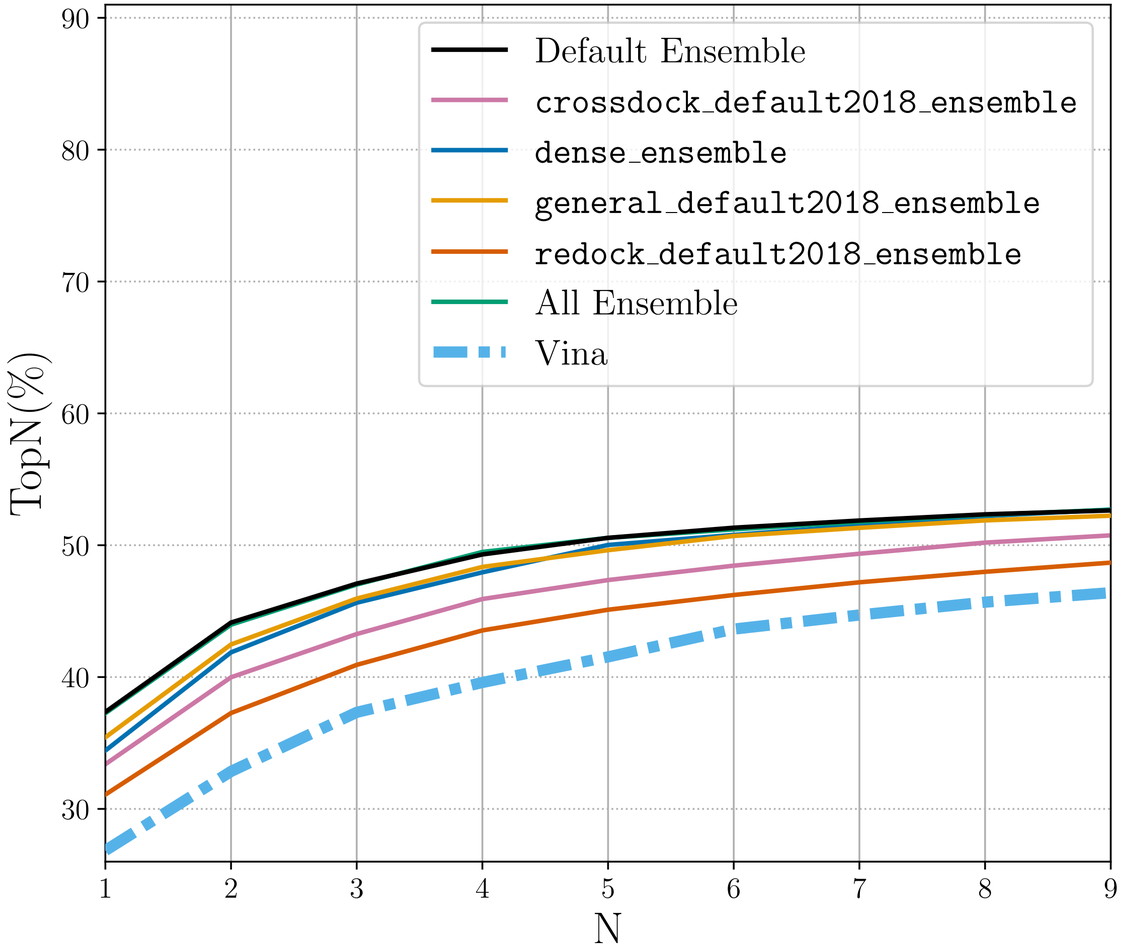


Crossdocking

Default CNN Ensemble



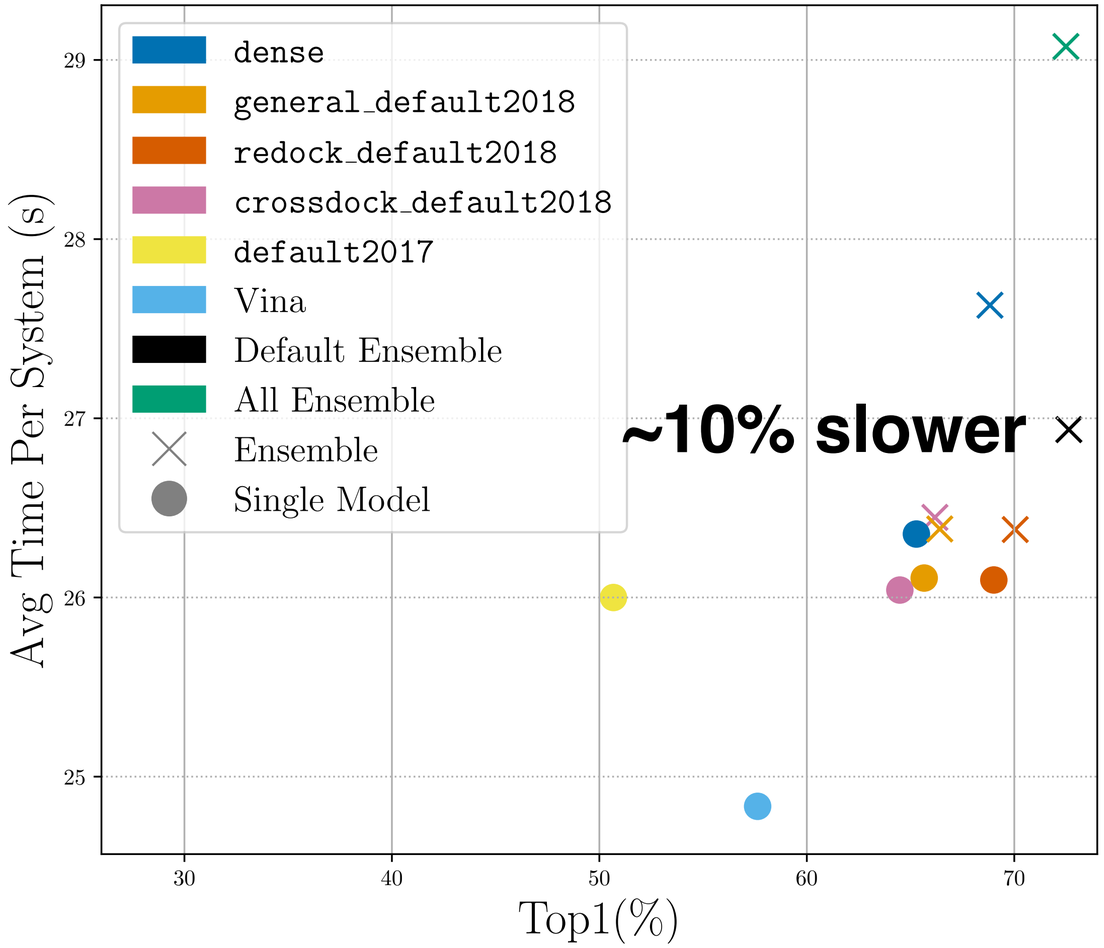
Redocking



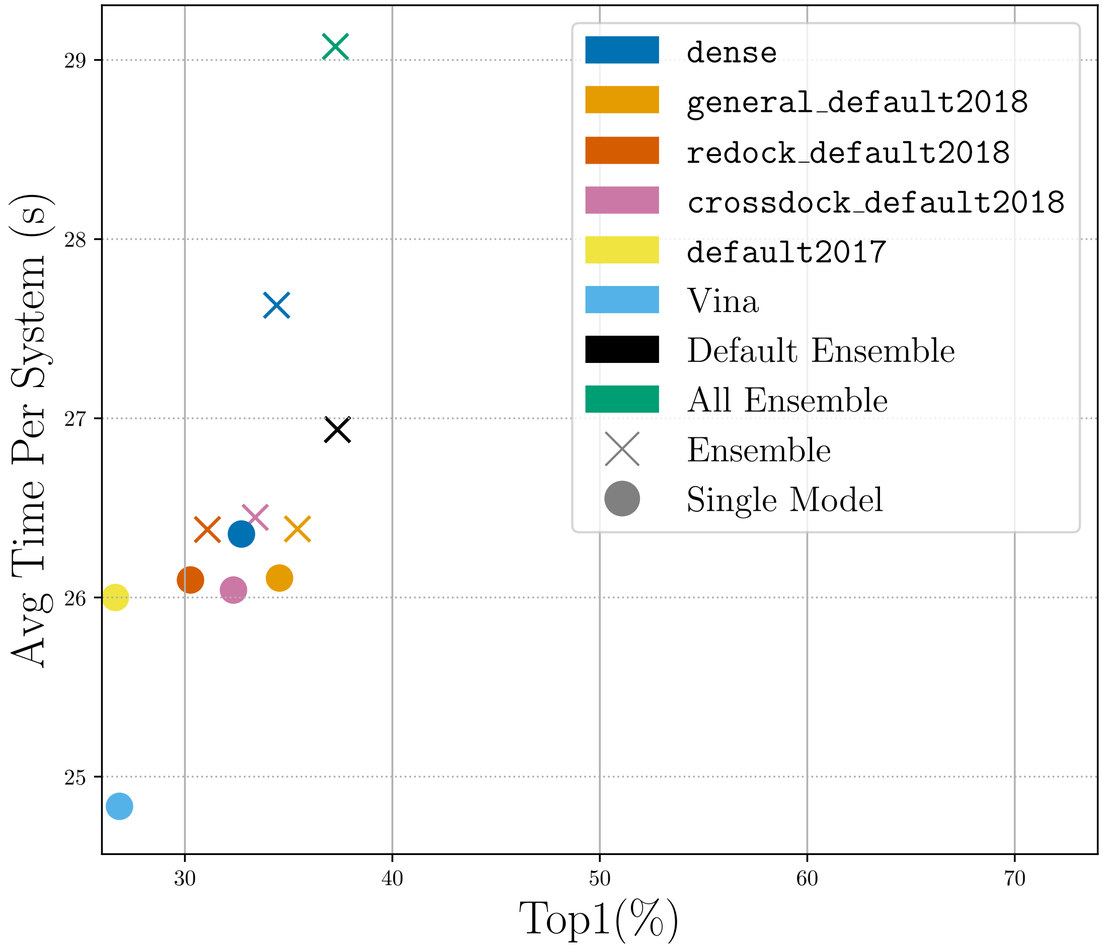
Crossdocking

Default CNN Ensemble

GPU Times



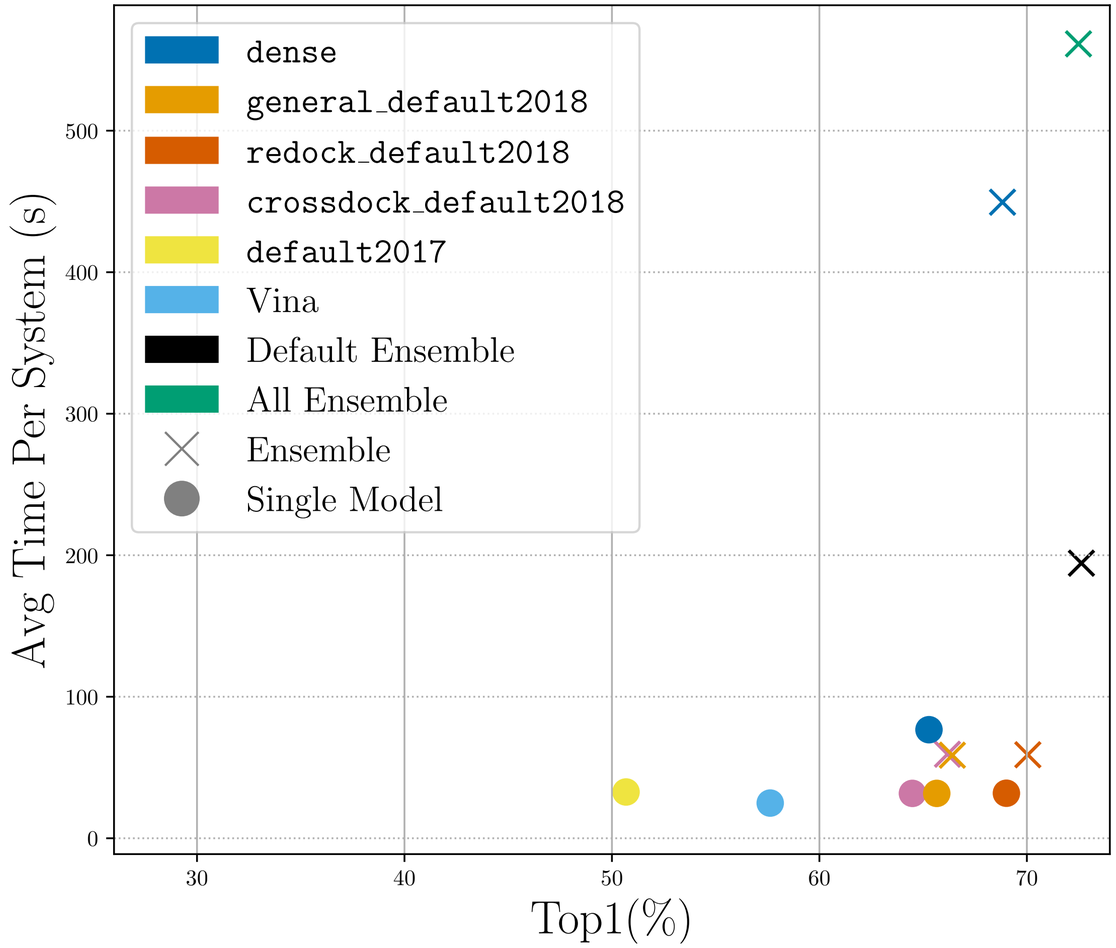
Redocking



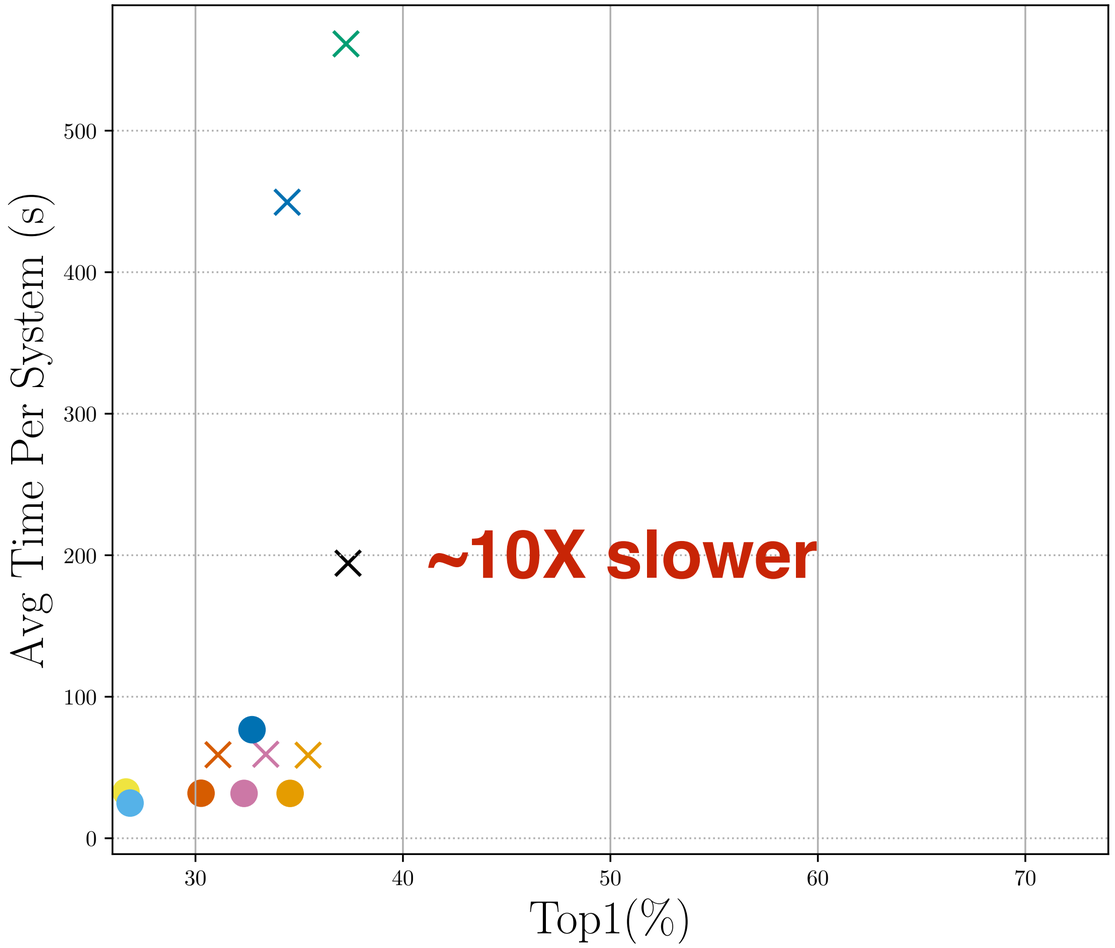
Crossdocking

Default CNN Ensemble

CPU Times

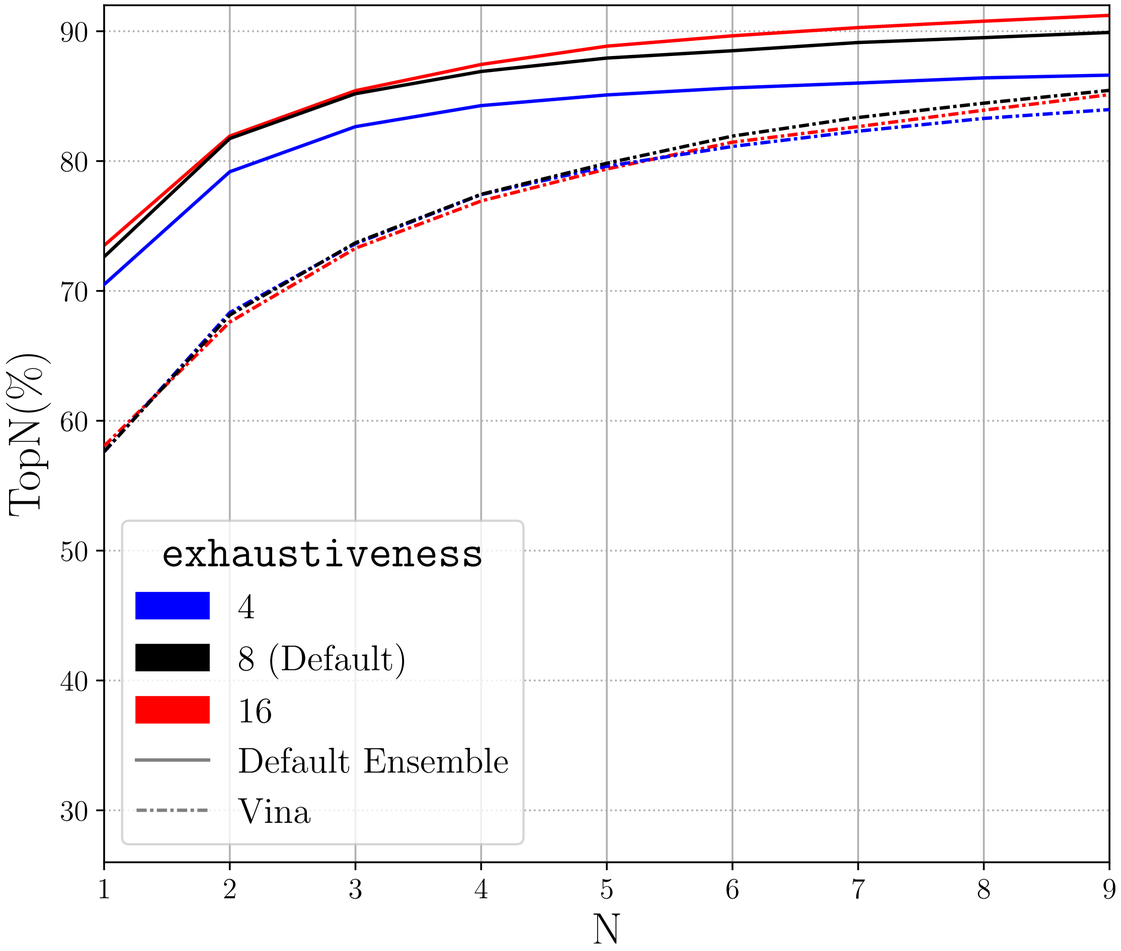


Redocking

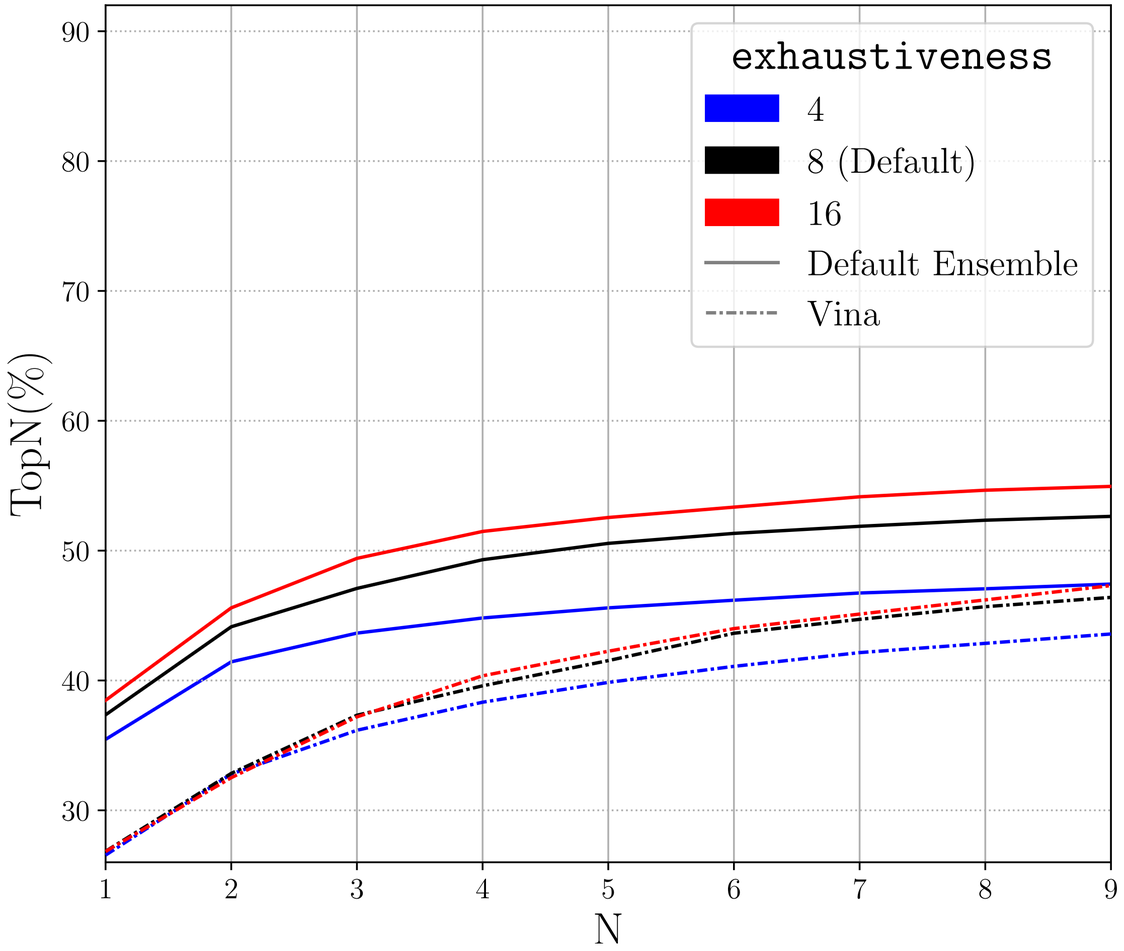


Crossdocking

Exhaustiveness



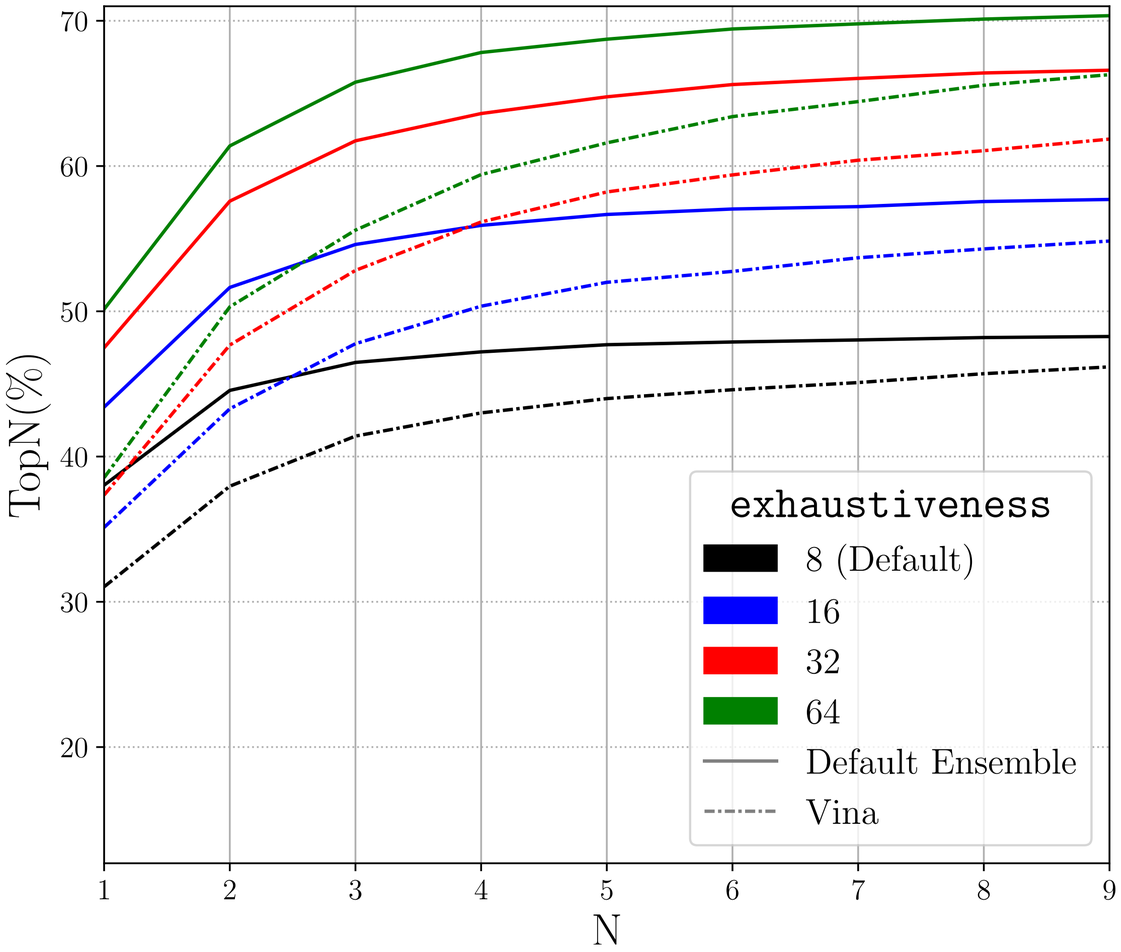
Redocking



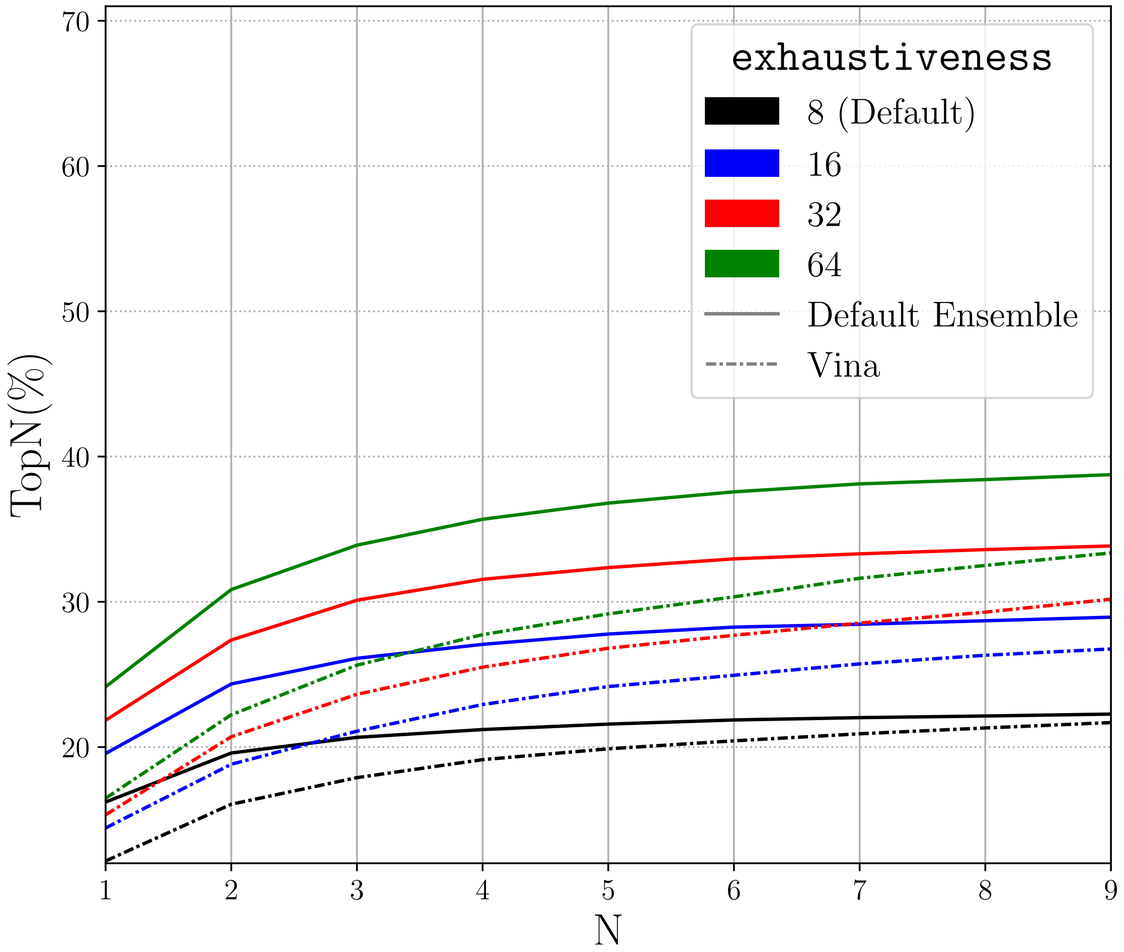
Crossdocking

Whole Protein Docking

```
gnina -r rec.pdb -l lig.pdb --autobox_ligand rec.pdb -o docked.sdf.gz
```



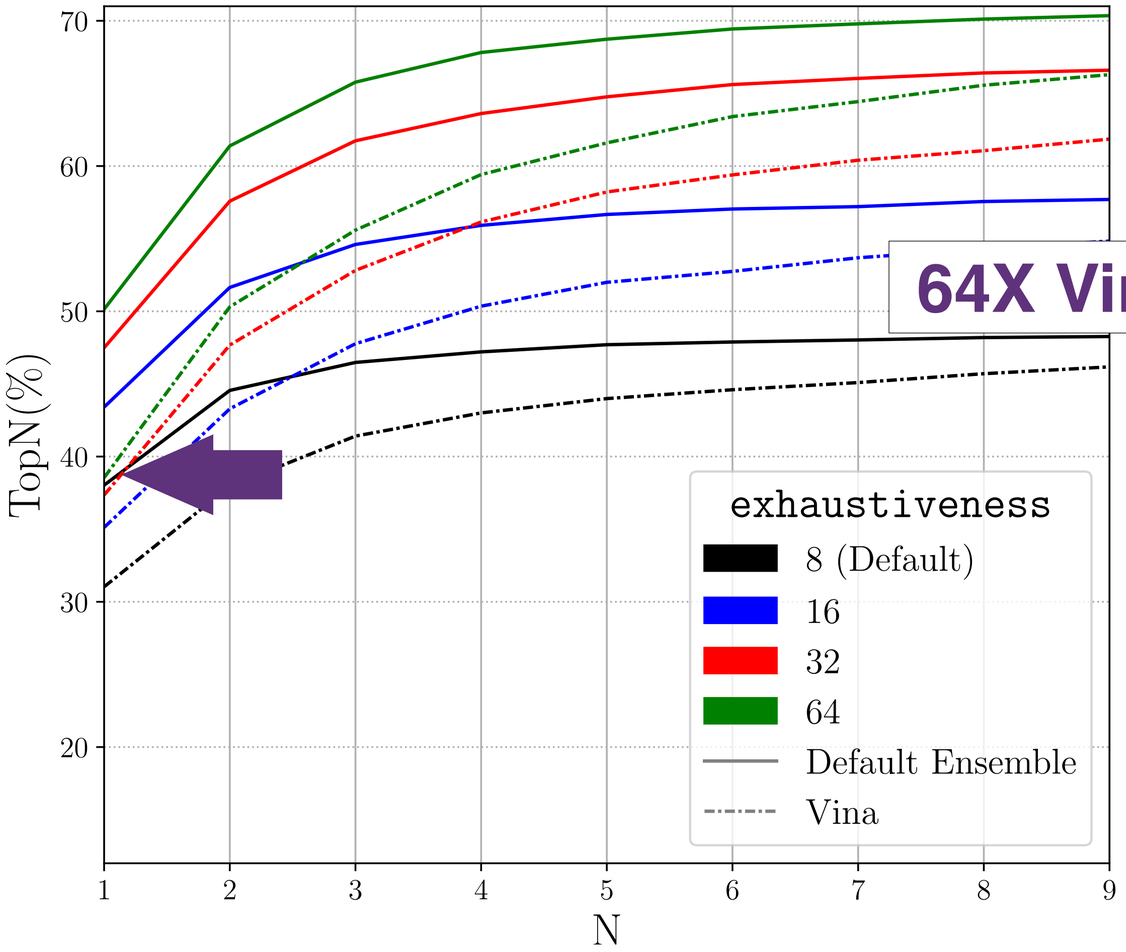
Redocking



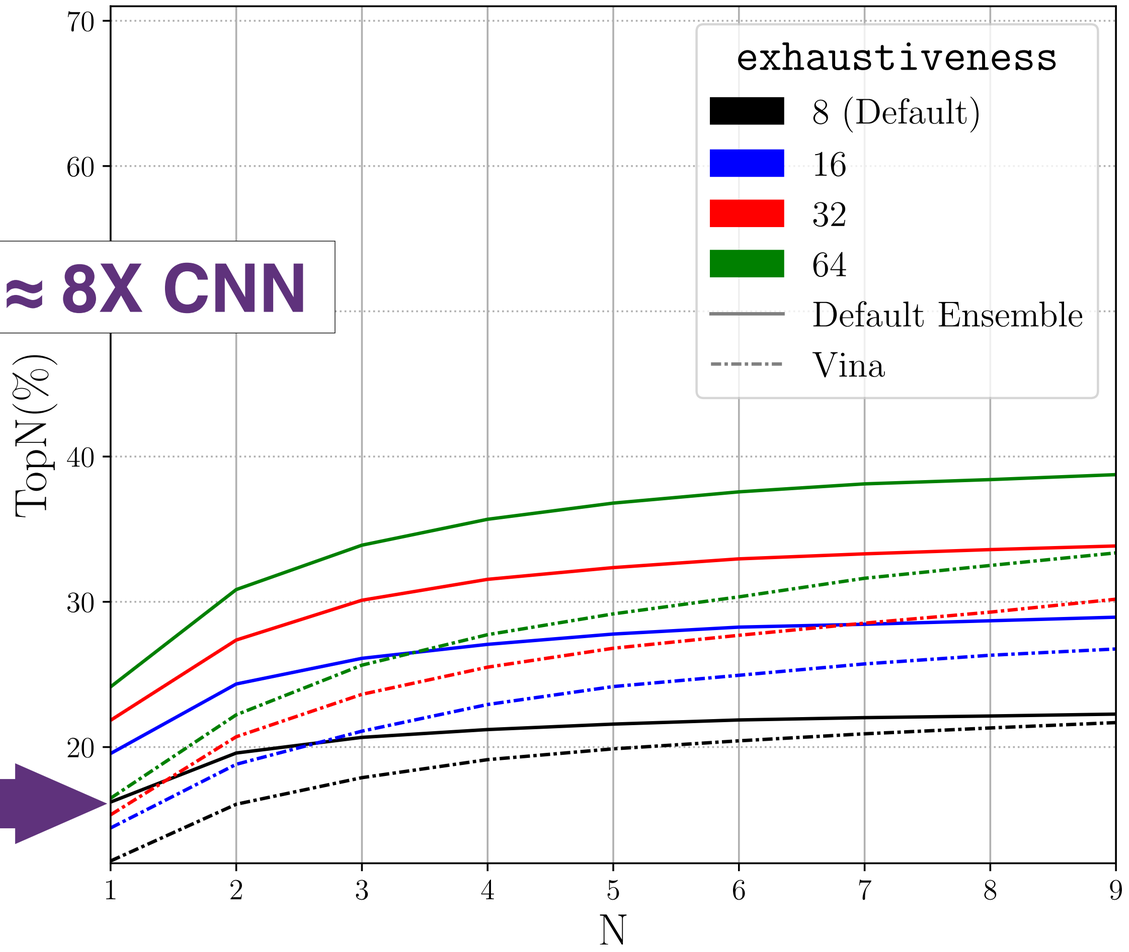
Crossdocking

Whole Protein Docking

```
gnina -r rec.pdb -l lig.pdb --autobox_ligand rec.pdb -o docked.sdf.gz
```



64X Vina ≈ 8X CNN

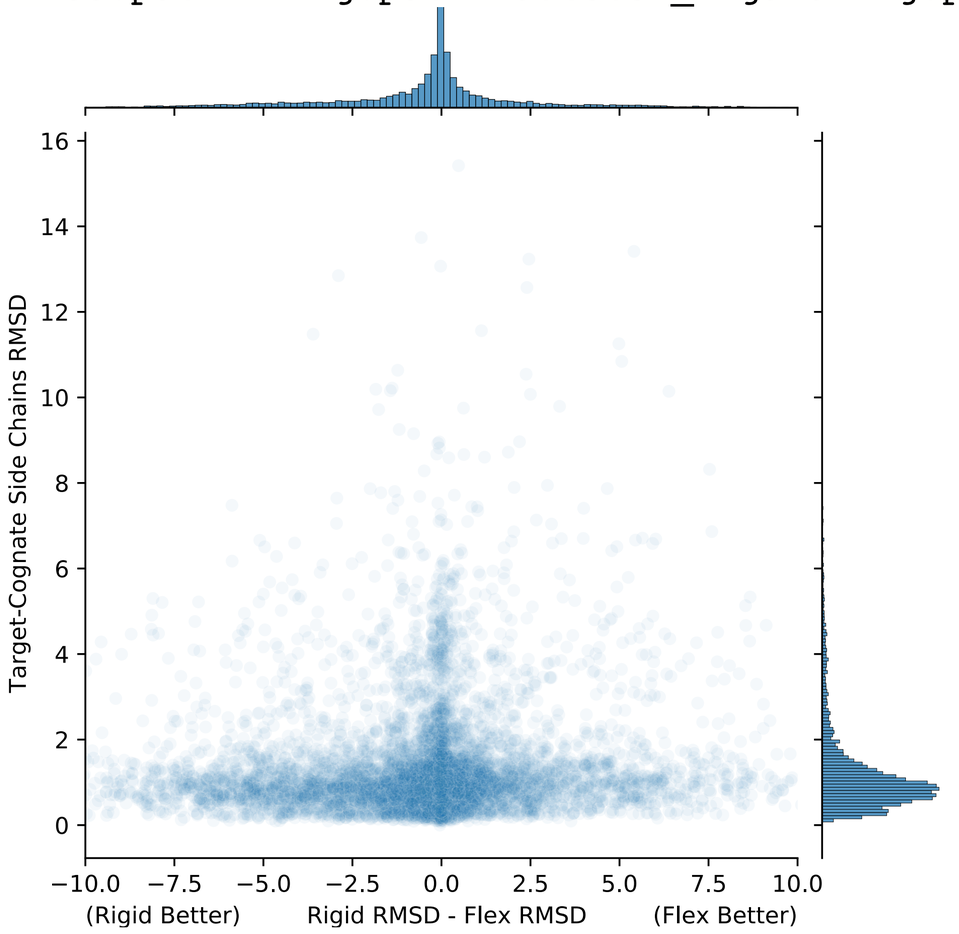


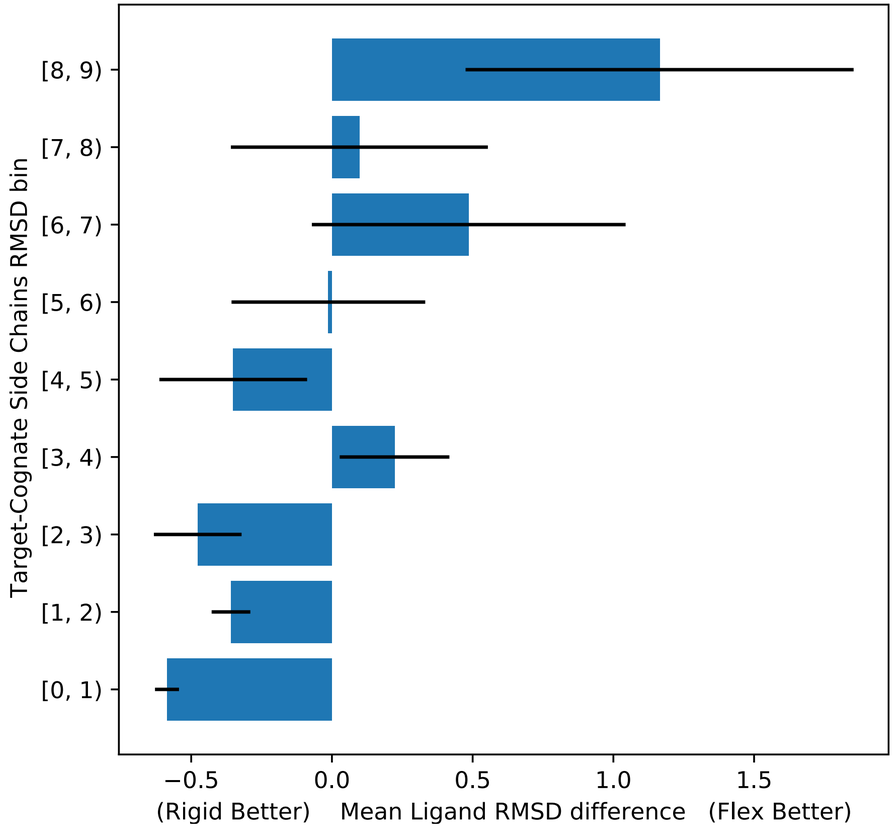
Redocking

Crossdocking

Flexible Docking

```
gnina -r rec.pdb -l lig.pdb --autobox_ligand lig.pdb -o docked.sdf.gz --flexdist 3.5 --flexdist_ligand lig.pdb
```

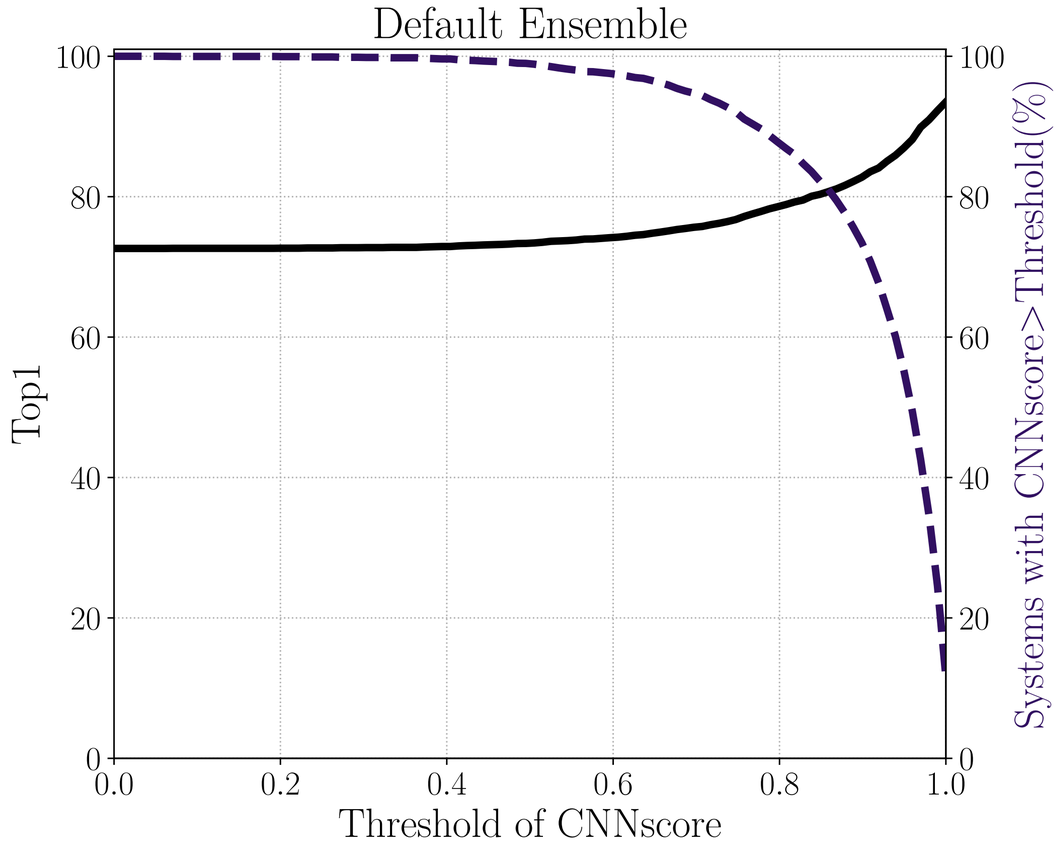




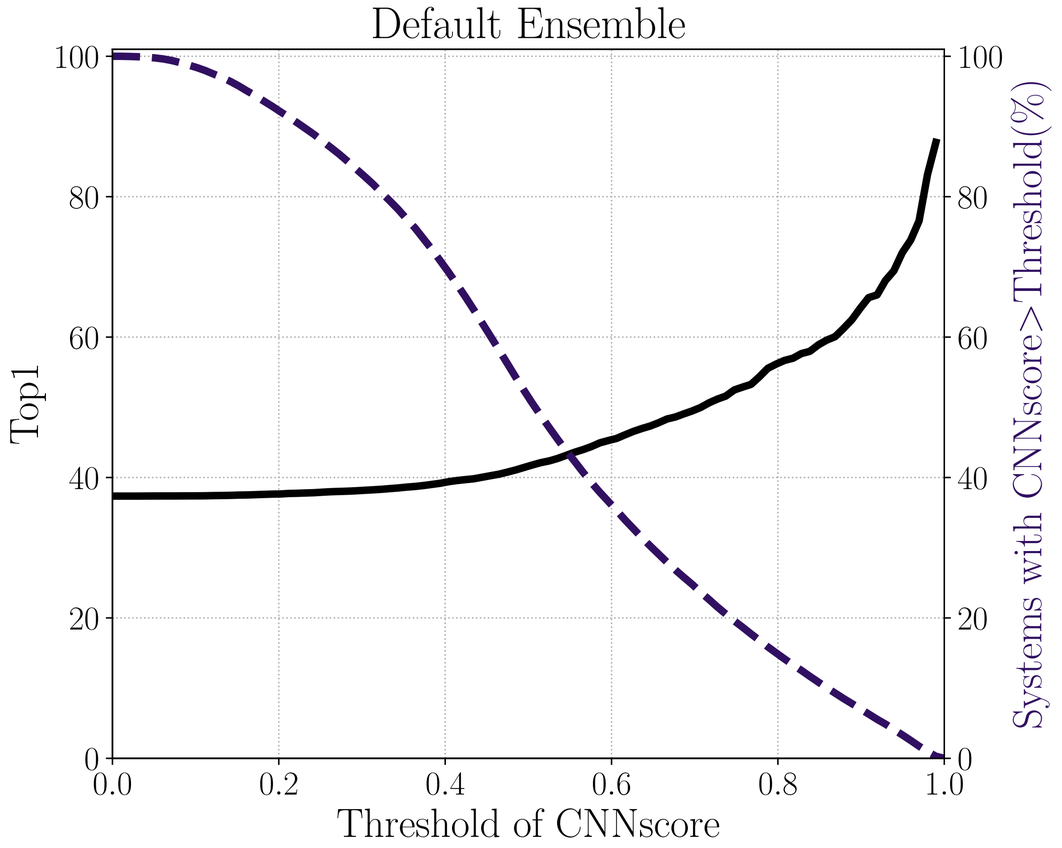
Recommendation: Use flexible docking for specific, intelligently selected residues

```
gnina -r rec.pdb -l lig.pdb --autobox_ligand lig.pdb -o docked.sdf --flexres A:84,A:88
```

CNN Scores

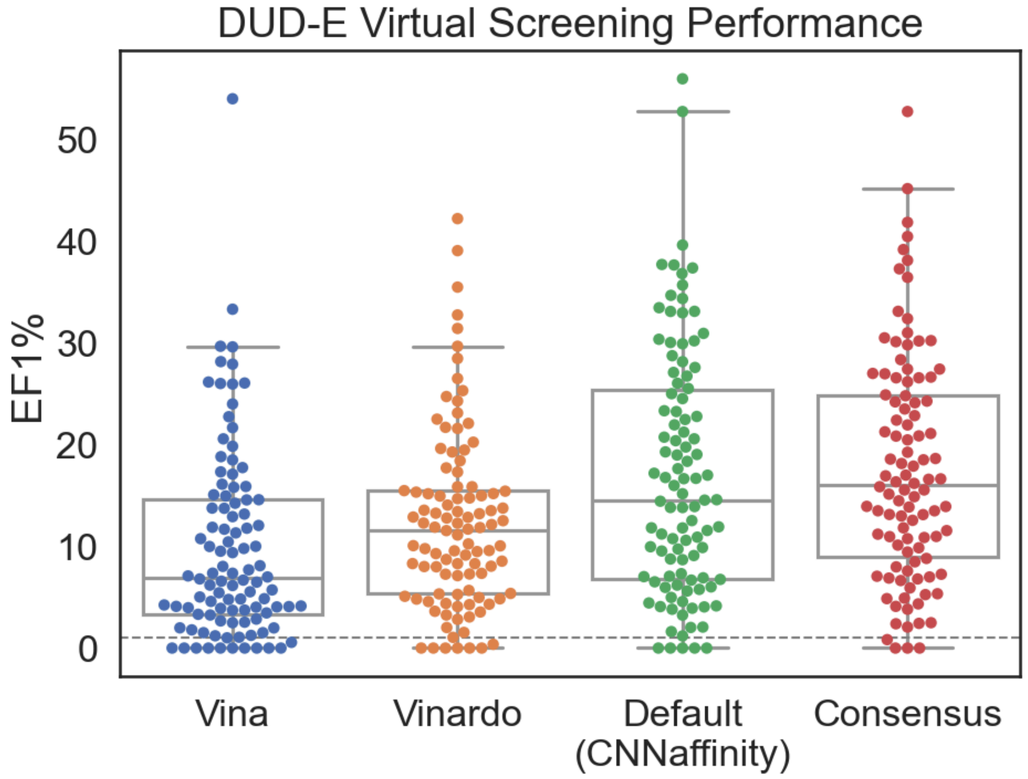


Redocking

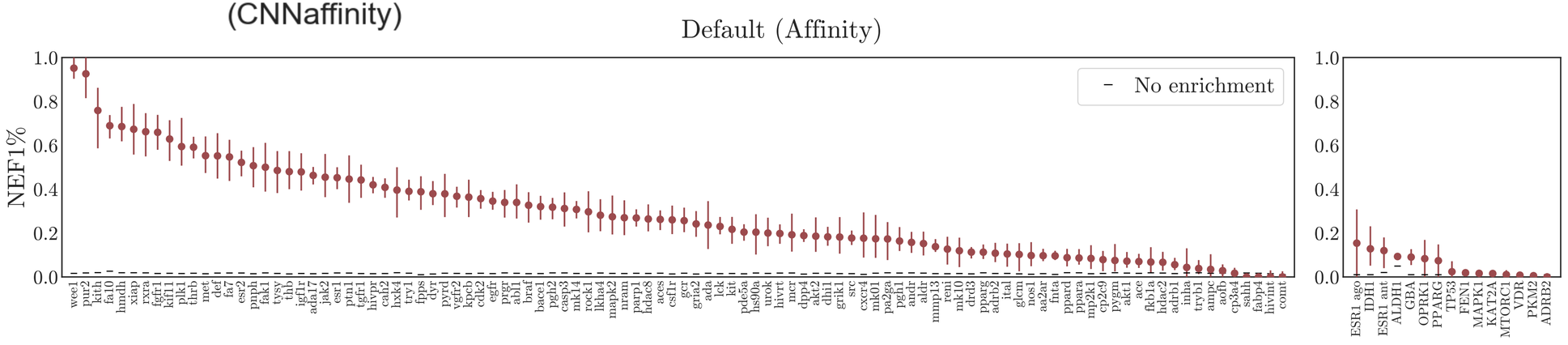


Crossdocking

Virtual Screening



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



Using GNINA

Colab Notebook

gnina.ipynb - Colaboratory

colab.research.google.com/drive/1QYo5QLUE80N_G28PlpYs6OKGddhhd931?usp=sharing#scrollTo=sMTknhig...

gnina.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk

```
!./gnina -r rec.pdb -l lig.pdb --autobox_ligand lig.pdb -o docked.sdf --seed 0
```

```
                 _
                (_)
   __ _ _ __  _ _ __   __ _
  / _` | '_ \| | '_ \ / _` |
 | (_| | | | | | | | | (_| |
  \__, |_| |_|_|_| |_|\__,_|
   __/ |
  |___/

gnina v1.0.1 HEAD:aa41230   Built Mar 23 2021.
gnina is based on smina and AutoDock Vina.
Please cite appropriately.

Commandline: ./gnina -r rec.pdb -l lig.pdb --autobox_ligand lig.pdb -o docked.sdf --seed 0
================================
*** Open Babel Warning  in PerceiveBondOrders
  Failed to kekulize aromatic bonds in OBMol::PerceiveBondOrders

Using random seed: 0

0%   10   20   30   40   50   60   70   80   90   100%
|----|----|----|----|----|----|----|----|----|----|
**************************************************

mode |  affinity  |     CNN      |   CNN
     | (kcal/mol) | pose score | affinity
-----+------------+------------+----------
    1       -8.50       0.8963      6.783
    2       -8.08       0.6049      6.600
    3       -8.31       0.4536      6.456
    4       -6.59       0.2824      6.080
    5       -6.75       0.2072      5.836
```

48s completed at 12:16 AM

gnina.ipynb - Colaboratory

colab.research.google.com/drive/1QYo5QLUE80N_G28PlpYs6OKGddhhd931?usp=sharing#scrollTo=WctyMpd...

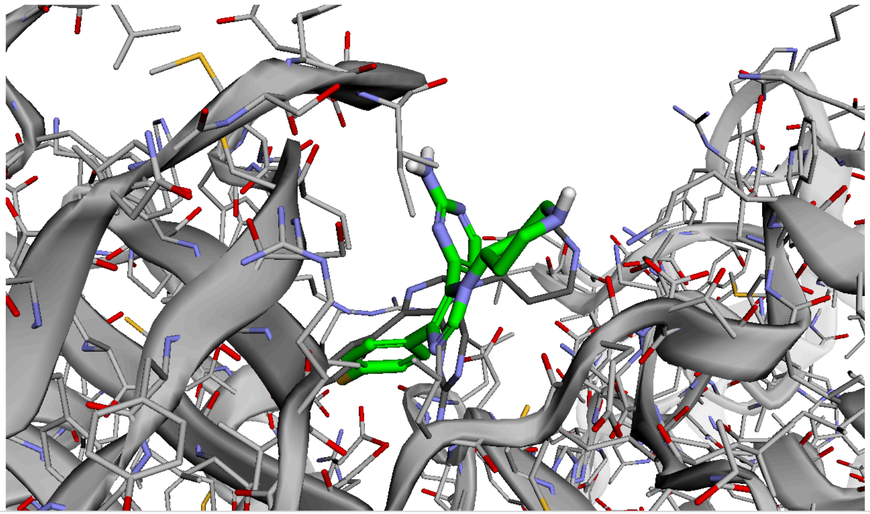
gnina.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text

RAM Disk

```
import py3Dmol
v = py3Dmol.view()
v.addModel(open('rec.pdb').read())
v.setStyle({'cartoon':{},'stick':{'radius':.1}})
v.addModel(open('lig.pdb').read())
v.setStyle({'model':1},{'stick':{'colorscheme':'dimgrayCarbon','radius':.125}})
v.addModelsAsFrames(open('docked.sdf').read())
v.setStyle({'model':2},{'stick':{'colorscheme':'greenCarbon'}})
v.animate({'interval':1000})
v.zoomTo({'model':1})
v.rotate(90)
```



48s completed at 12:16 AM

Software Availability

Name	License	URL
gnina	Apache/GNU	https://github.com/gnina/gnina
pharmit	BSD/GNU	http://pharmit.csb.pitt.edu/
libmolgrid	Apache/GNU	https://gnina.github.io/libmolgrid/
smina	Apache/GNU	http://smina.sf.net/
3Dmol.js	BSD	https://3dmol.org/
sdsorter	GNU	https://sourceforge.net/projects/sdsorter/

Acknowledgements

Drew McNutt
Jocelyn Sunseri
Matthew Ragoza
Paul Francoeur
Jonathan King
Tomohide Masuda

Rocco Meli
Rishal Aggarwal

Josh Hochuli
Elisa Idrobo
Lily Turner
Alec Helbling
Andrew Jia
Rich Iovanisci
Ian Snyder

National Institute of General Medical Sciences
R01GM108340
R35GM140753

NSF
CHE-1800435

RELAY THERAPEUTICS