



University of
Pittsburgh

Structure-Based Virtual Screening on the Cheap

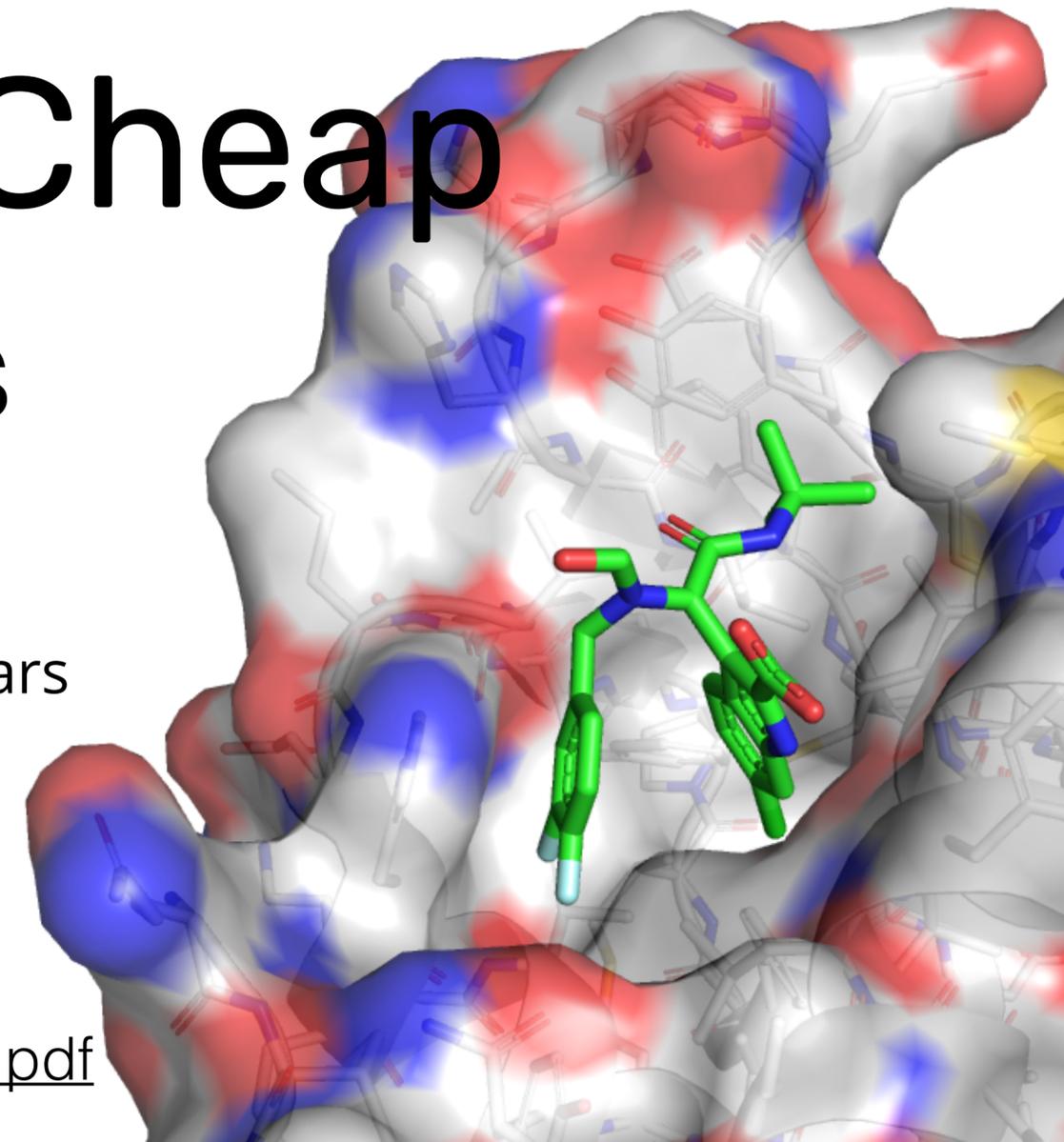
David Ryan Koes

 @david_koes

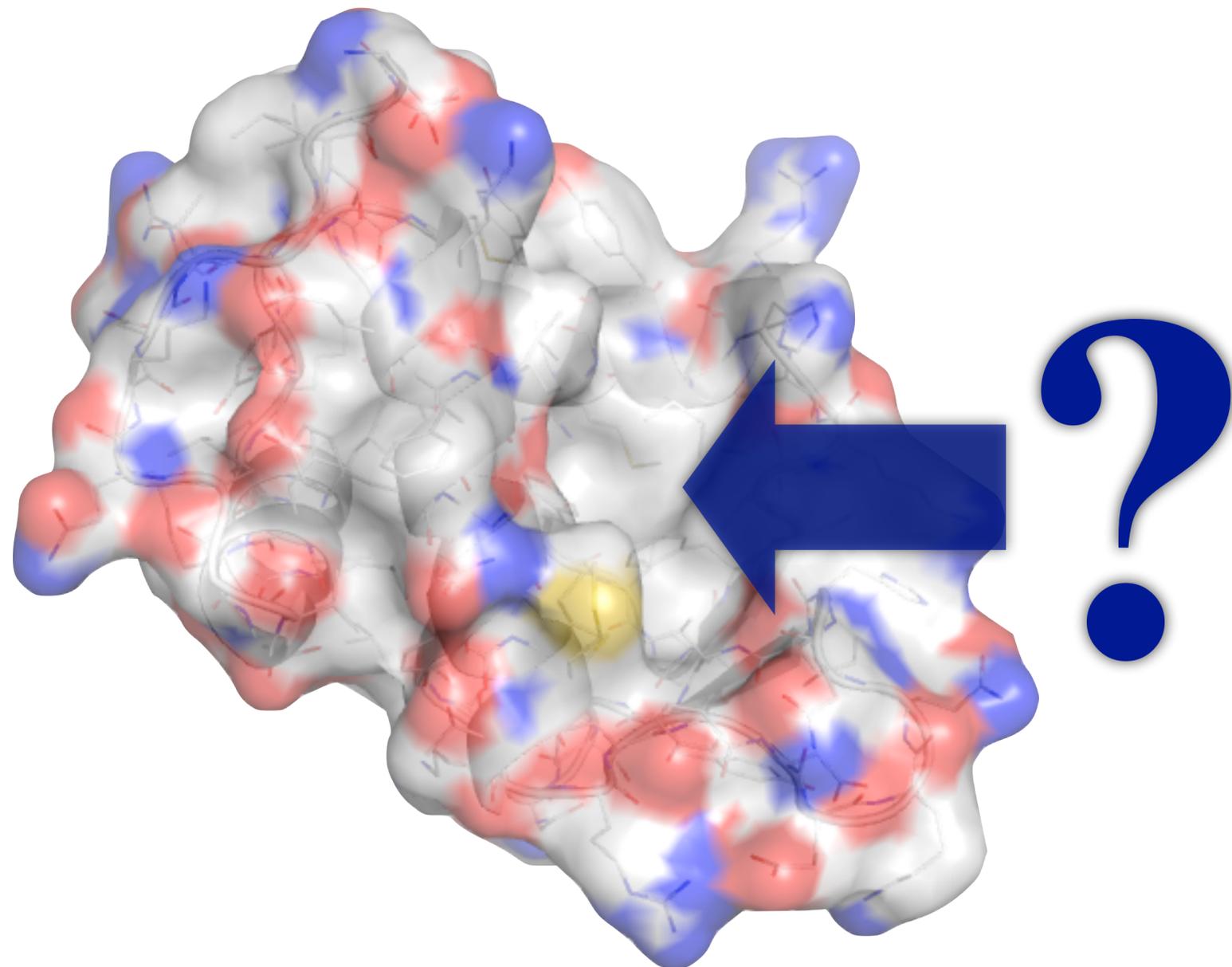
Global Health Compound Design Webinars

11/15/2021

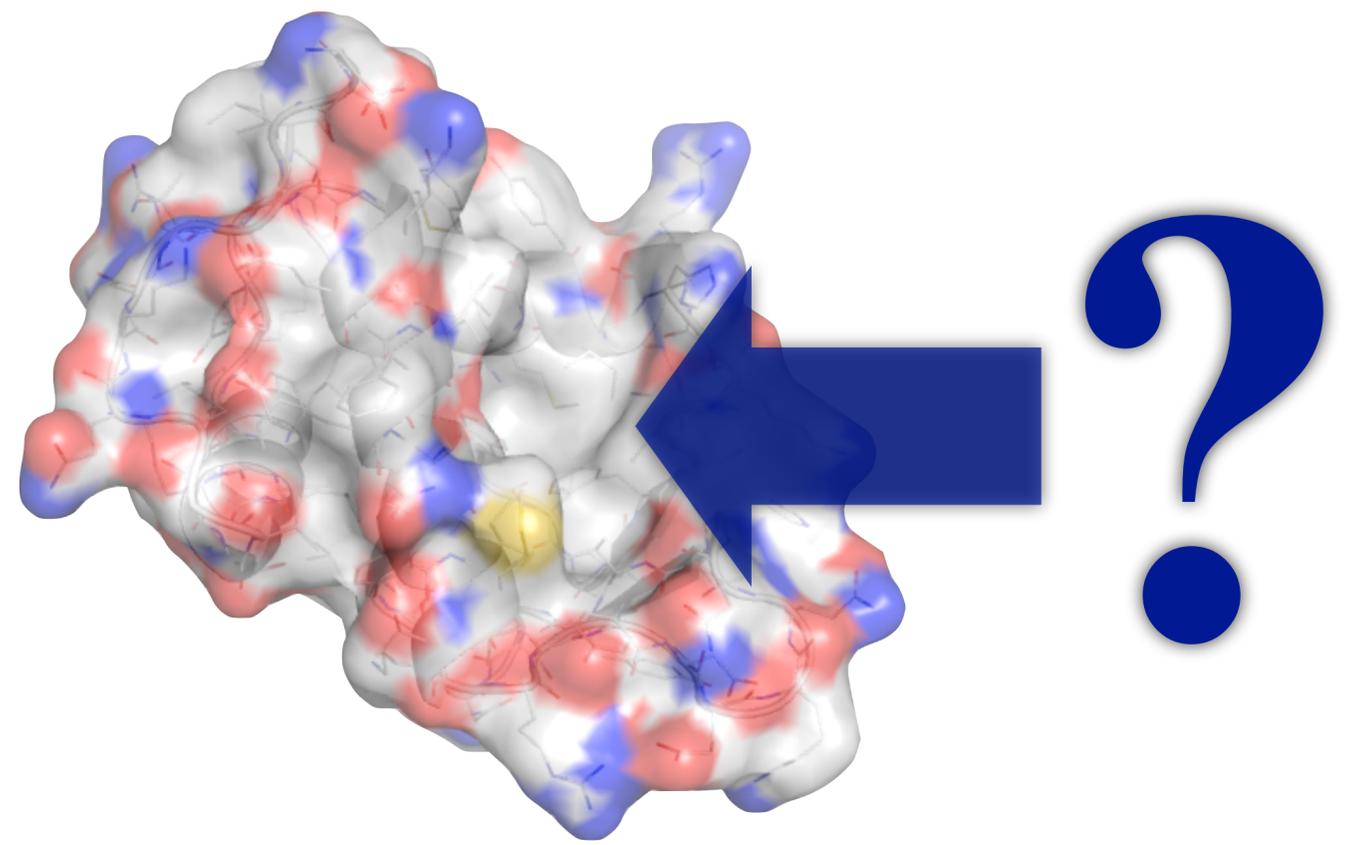
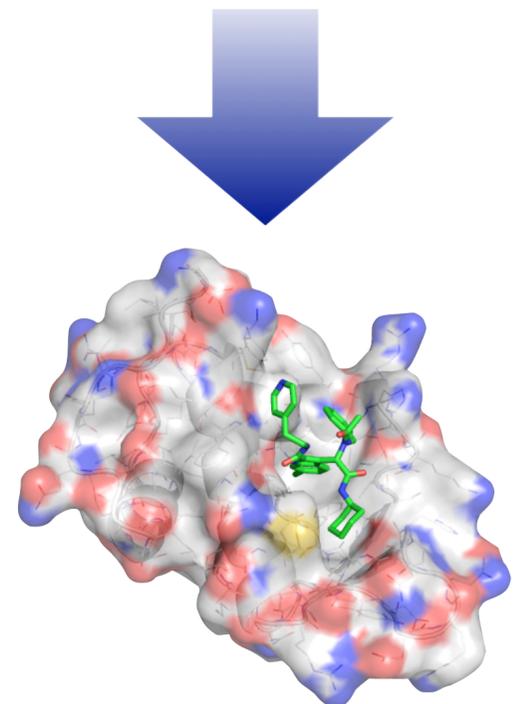
https://bits.csb.pitt.edu/files/on_the_cheap.pdf



Structure Based Drug Design

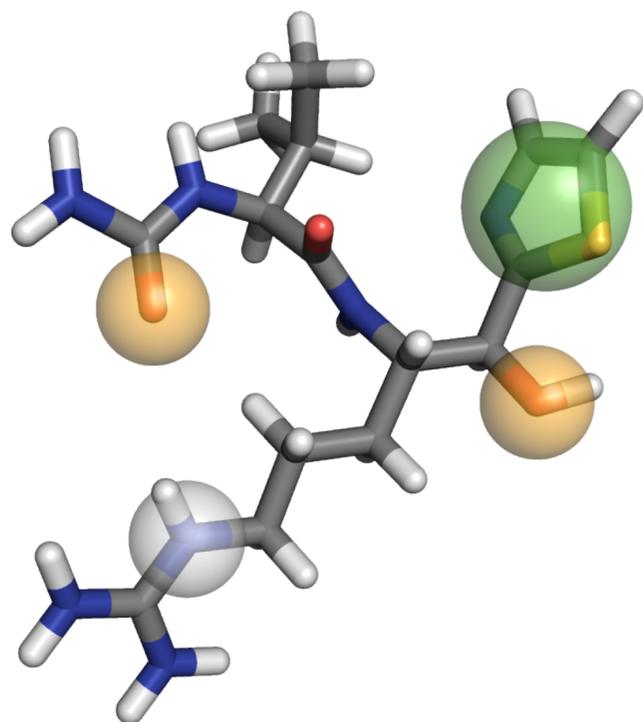


Structure Based Drug Design Funnel



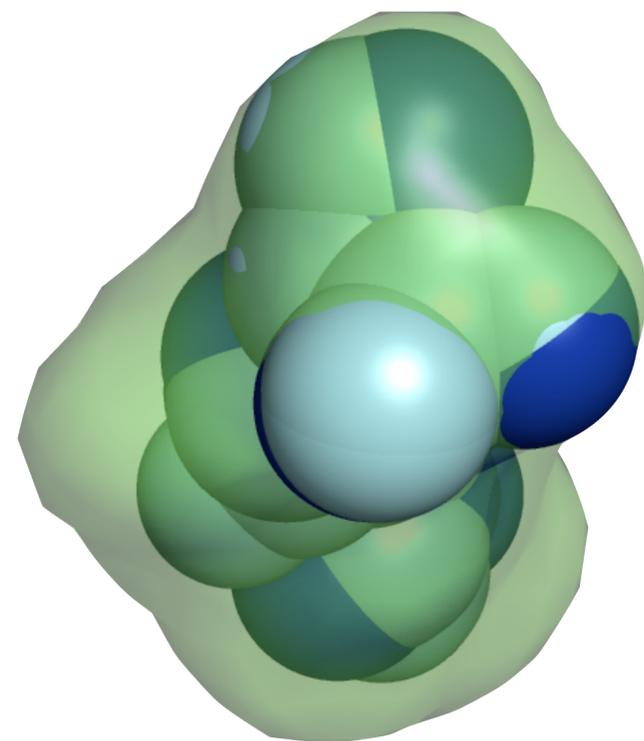
How does it bind? (pose prediction)
 Does it bind? (virtual screening)
 How well does it bind? (affinity prediction)

Matching



pharmer.sf.net

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



github.com/dkoes/shapedb

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.



Name	RMSD	Mass	RBnds
PubChem-13960682	0.223	392	5
PubChem-23673360	0.223	391	4
PubChem-13960682	0.223	392	5
PubChem-23673360	0.223	391	4
PubChem-13960684	0.243	388	6
PubChem-13960684	0.243	388	6
PubChem-13960684	0.243	388	6
PubChem-13960684	0.243	388	6
PubChem-13960684	0.250	388	6
PubChem-59810304	0.311	481	8
PubChem-10000399	0.325	389	6
PubChem-10000399	0.327	389	6
PubChem-59081061	0.349	875	15
PubChem-10250942	0.379	387	3
PubChem-23686481	0.379	386	2
PubChem-13960681	0.442	385	7
PubChem-13960681	0.442	385	7
PubChem-13960681	0.444	385	7
PubChem-88181354	0.449	698	10
PubChem-842716	0.462	319	8

pharmit.csb.pitt.edu

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.

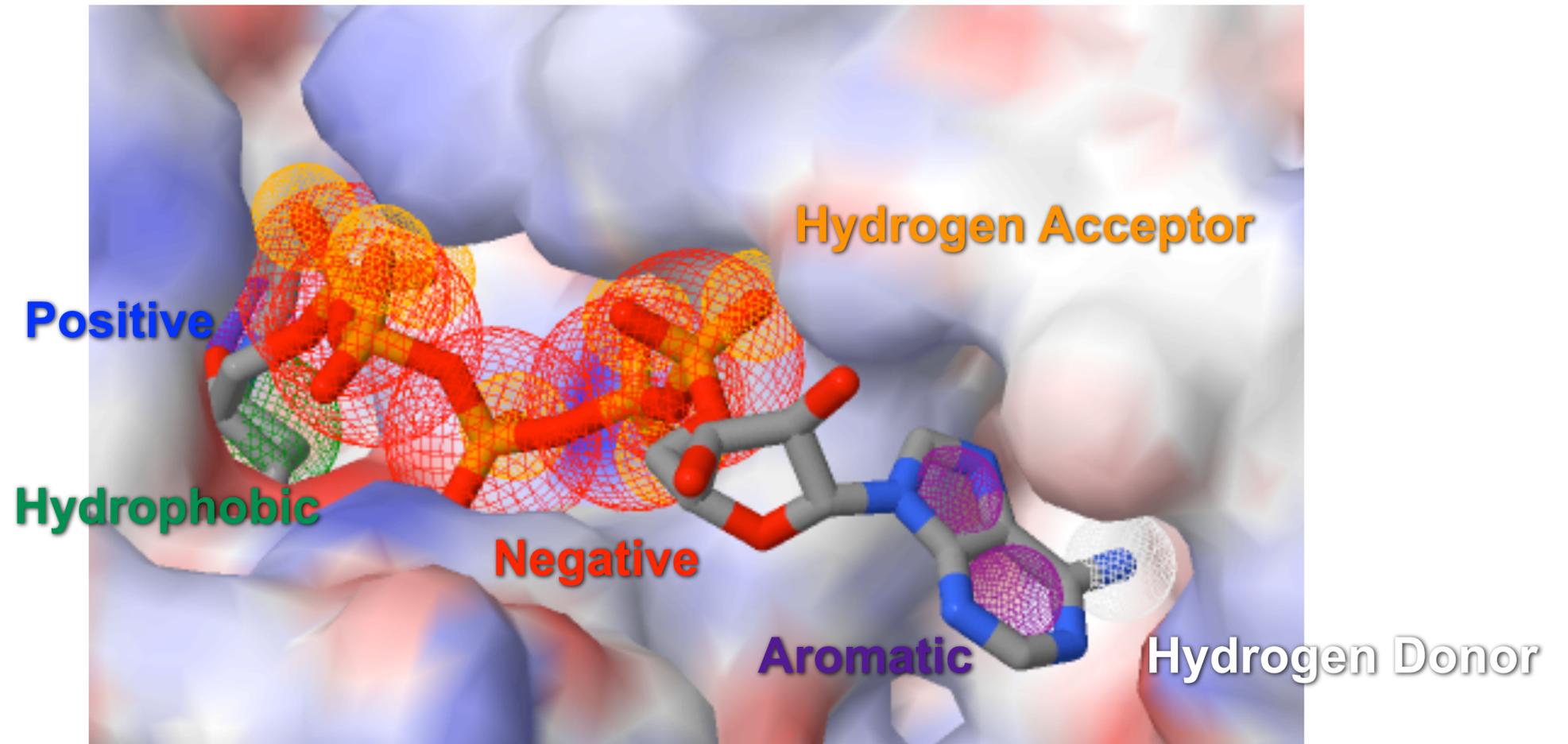
Sunseri J, Koes DR. [Pharmit: interactive exploration of chemical space](#). Nucleic Acids Res. 2016 Jul 8;44(W1):W442-8.

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:

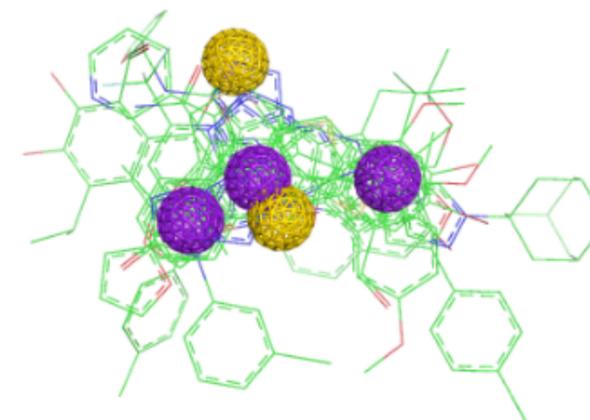
- aromatic ring
- hydrophobic area
- positive ionizable
- negative ionizable
- hydrogen bond donor
- hydrogen bond acceptor



Pharmacophore Elucidation

Ligand Only

e.g. PharmaGist, PENG, GAPE, MOGA

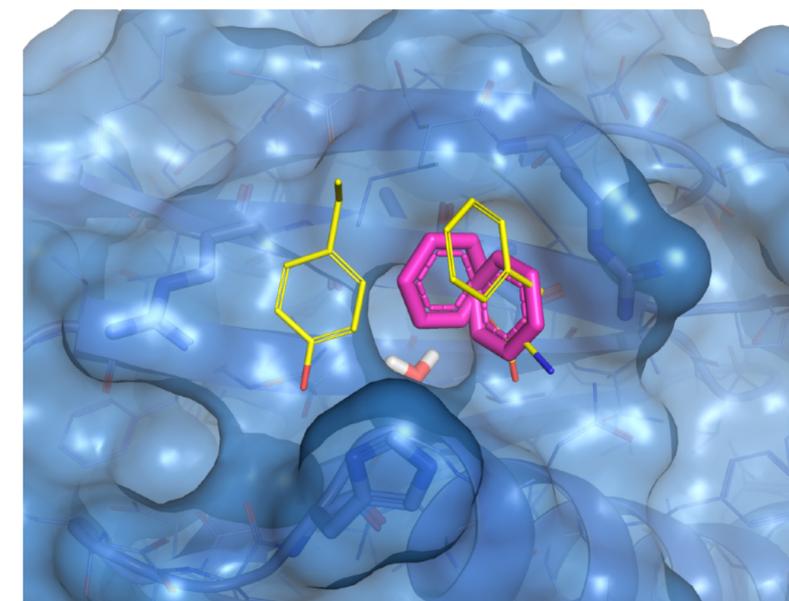


Receptor Only

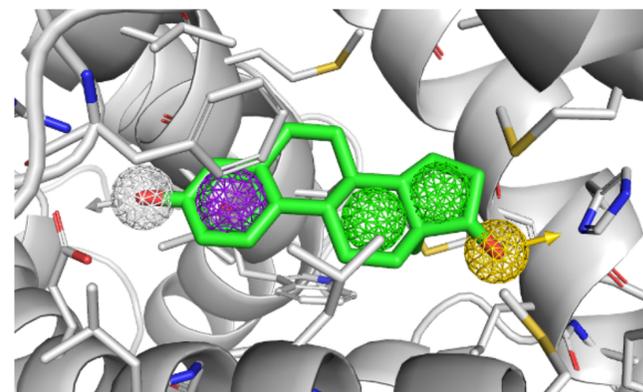
Simulation: Pharmmaker, SILCS, PyRod

Docking: FTMap

Grids: GRAIL



Ligand-Receptor



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

Ligand Only

PharmaGist
Webservice

[About] [WebServer] [Download] [FAQ] [Help / Getting Started] Contact: duhovka@gmail.com ppdock@tau.ac.il

Upload Input Molecules in Mol2 Format (e.g. [input examples](#)) No file chosen

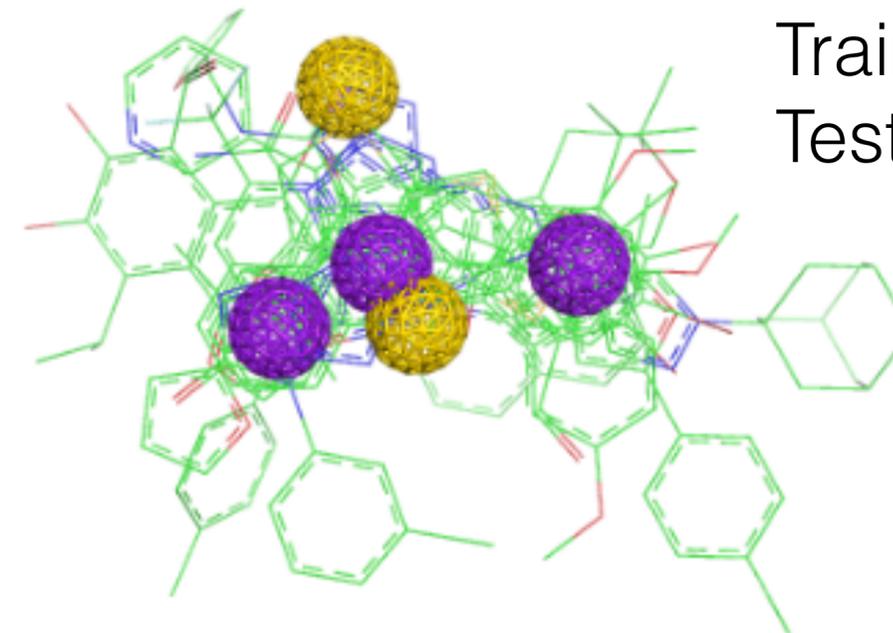
No. of Output Pharmacophores:

E-Mail Address:

Advanced Options:
[\[Show\]](#)[\[Hide\]](#)

If you use this webservice, please cite:

1. Inbar Y, Schneidman-Duhovny D, Dror O, Nussinov R, Wolfson HJ. Deterministic Pharmacophore Detection via Multiple Flexible Alignment of Drug-Like Molecules. In Proc. of RECOMB 2007, vol. 3692 of Lecture Notes in Computer Science, pp. 423-434. Springer Verlag.
2. Schneidman-Duhovny D, Dror O, Inbar Y, Nussinov R, Wolfson HJ. PharmaGist: a webservice for ligand-based pharmacophore detection. Nucleic Acids Research 2008. [\[Abstract\]](#) [\[FREE Full Text\]](#)
3. Dror O, Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. Novel approach for efficient pharmacophore-based virtual screening: method and applications. J Chem Inf Model. 2009 Oct;49(10):2333-43. [\[Abstract\]](#) [\[Full Text\]](#)



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

FTMap: A Small Molecule Mapper

ftmap.bu.edu/home.php

Map Queue Results Examples Help Papers Contact

FTMap

computational solvent mapping

Important Considerations
-Please submit no more than 20 jobs at a time

[sign out](#)

Map

Jobs submitted via our guest account are publicly accessible. Please [create an account](#) if you wish to keep your results private.

Job Name:

Protein

PDB ID:

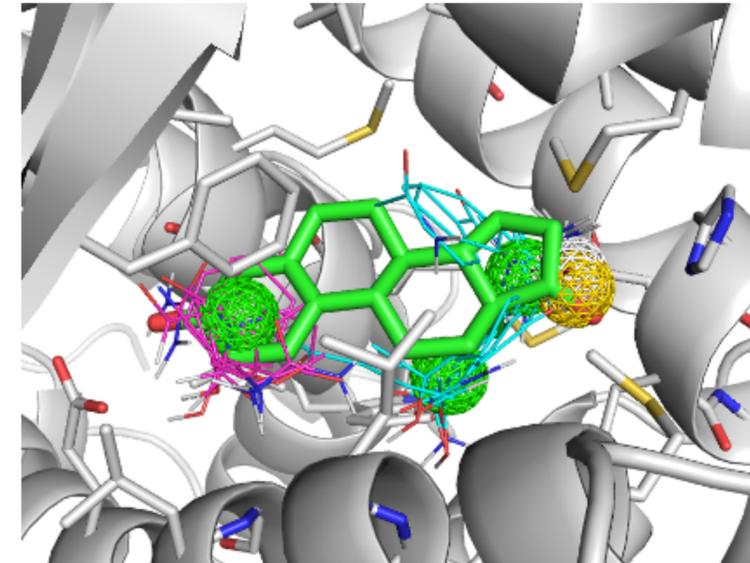
[Upload PDB](#) [?](#)

Chains:

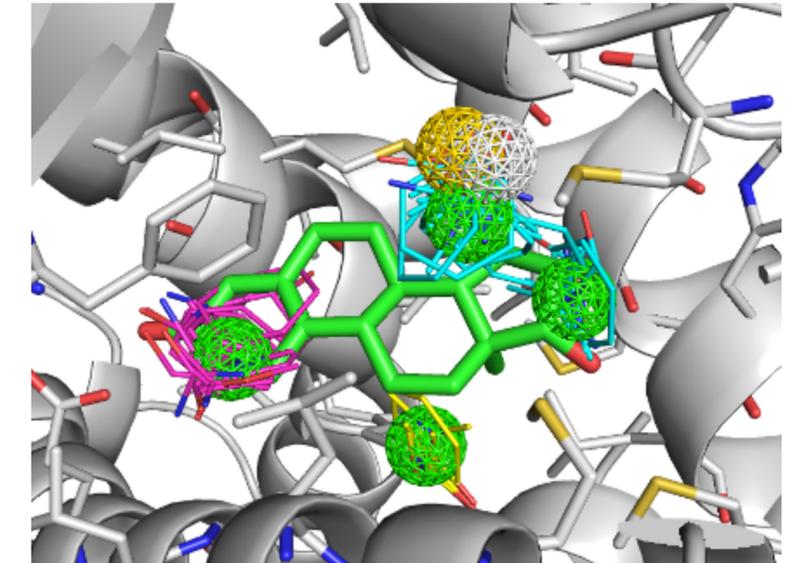
Whitespace separate desired chains. Leave chains blank to use all chains.

[Advanced Options](#) [?](#)

Vajda Lab | ABC Group
Boston University | Stony Brook University



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

Pharmit Search Engine

Search MolPort

Pharmacophore Search -> Shape Filter

Load Receptor... Load Features...

Pharmacophore

- HydrogenDonor** (9.53,3.92,35.82) Radius 0.5
- HydrogenAcceptor** (9.53,3.92,35.82) Radius 0.5
- HydrogenAcceptor** (20.0,4.36,33.43) Radius 0.5
- Hydrophobic** (12.17,4.27,35.2) Radius 1.0
- Hydrophobic** (18.22,4.96,35.4) Radius 1.0
- Hydrophobic** (17.88,4.44,32.8) Radius 1.0
- Hydrophobic** (16.24,4.83,33.93) Radius 1.0
- Aromatic** (12.17,4.27,35.2) Radius 1.1
- HydrogenDonor** (20.0,4.36,33.43) Radius 0.5

Add + Sort ▾

Shape

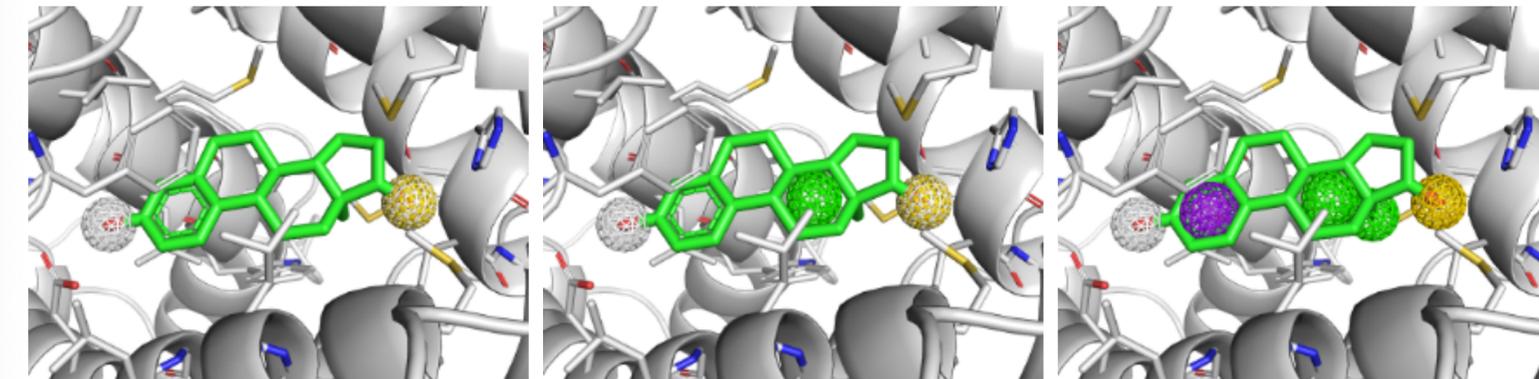
- Inclusive Shape
- Exclusive Shape

Filters

- Hit Reduction
- Hit Screening

Visualization

Load Session... Save Session...



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

pharmit.csb.pitt.edu/search.html

Search PubChem

Pharmacophore Search -> Shape Filter

Load Receptor... Load Features...

Pharmacophore

- Aromatic** (48.19,39.88,-1.88) Radius 1.1
- HydrogenAcceptor** (47.17,41.23,-5.87) Radius 0.5
- HydrogenAcceptor** (49.17,40.46,-6.38) Radius 0.5
- NegativeIon** (48.21,40.91,-5.92) Radius 0.75
- Hydrophobic** (49.4,41.67,-2.93) Radius 1.0
- Hydrophobic** (52.08,44.85,-2.15) Radius 1.0
- Hydrophobic** (48.19,39.88,-1.88) Radius 1.0
- Hydrophobic** (55.14,47.7,-1.04) Radius 1.0
- Hydrophobic** (50.03,43.31,-5.25) Radius 1.0
- Hydrophobic** (46.4,37.97,-1.96) Radius 1.0
- Aromatic** (49.4,41.67,-2.93) Radius 1.1
- Aromatic** (52.08,44.85,-2.15) Radius 1.1

Load Session... Save Session...

Display a menu

Pharmacophore Results

Name	RMSD	Mass	RBnds
PubChem-13960682	0.223	392	5
PubChem-23673360	0.223	391	4
PubChem-13960682	0.223	392	5
PubChem-23673360	0.223	391	4
PubChem-13960684	0.243	388	6
PubChem-13960684	0.243	388	6
PubChem-13960684	0.243	388	6
PubChem-13960684	0.250	388	6
PubChem-59810304	0.311	481	8
PubChem-10000399	0.325	389	6
PubChem-10000399	0.327	389	6
PubChem-59081061	0.349	875	15
PubChem-10250942	0.379	387	3
PubChem-23686481	0.379	386	2
PubChem-13960681	0.442	385	7
PubChem-13960681	0.442	385	7
PubChem-13960681	0.444	385	7
PubChem-88181354	0.449	698	10
PubChem-842716	0.462	319	8

Showing 1 to 19 of 38 hits

Previous 1 2 Next

Minimize Save...

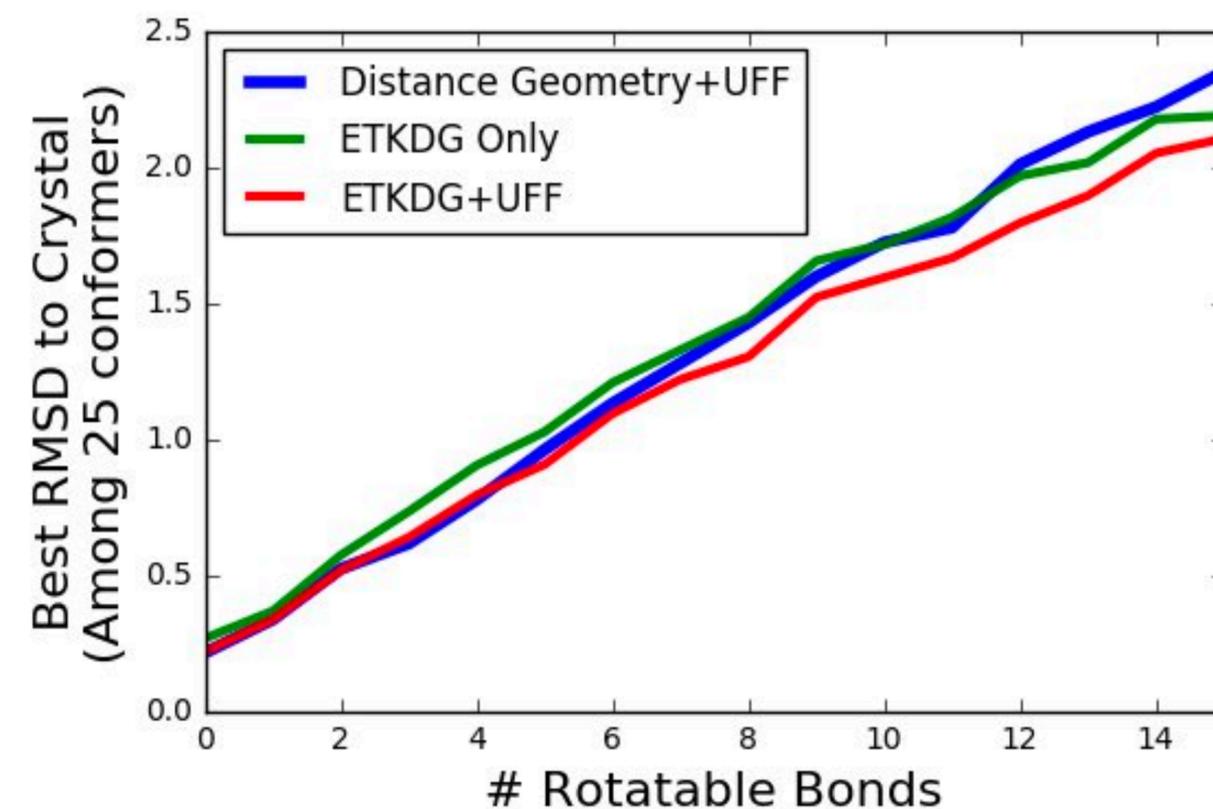
Key Points

- 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 stereochemistry implicitly during conformer generation
- Conformers
 - Sample maximum of 20 conformers using RDKit



Protein-Ligand Scoring

AutoDock Vina

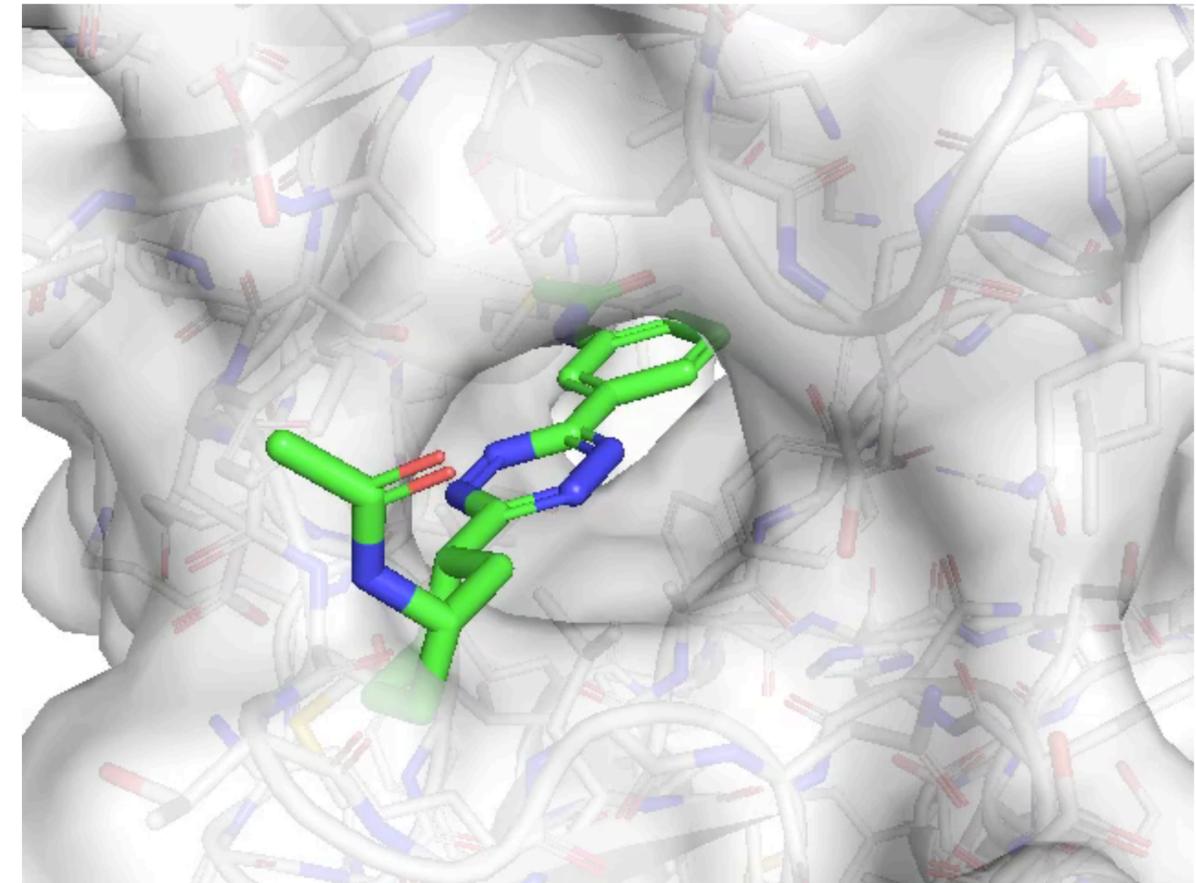
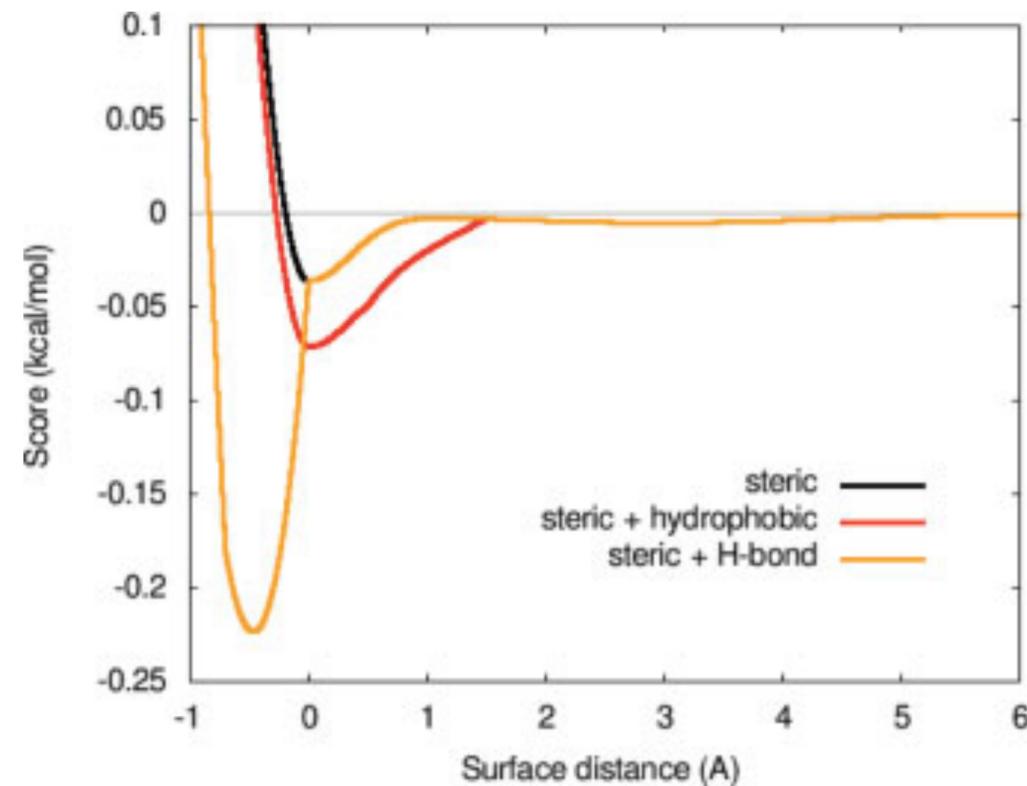
$$\text{gauss}_1(d) = w_{\text{gauss}_1} e^{-(d/0.5)^2}$$

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

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

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

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



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

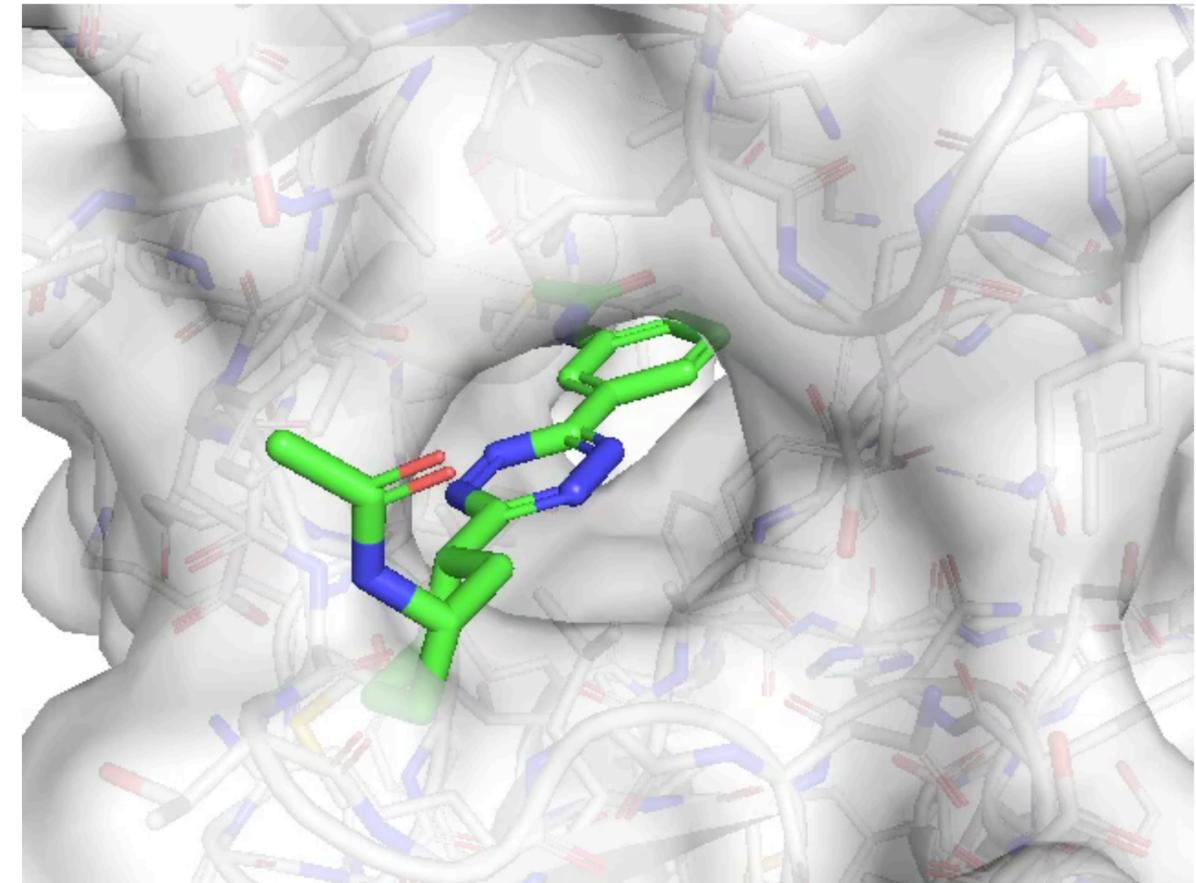
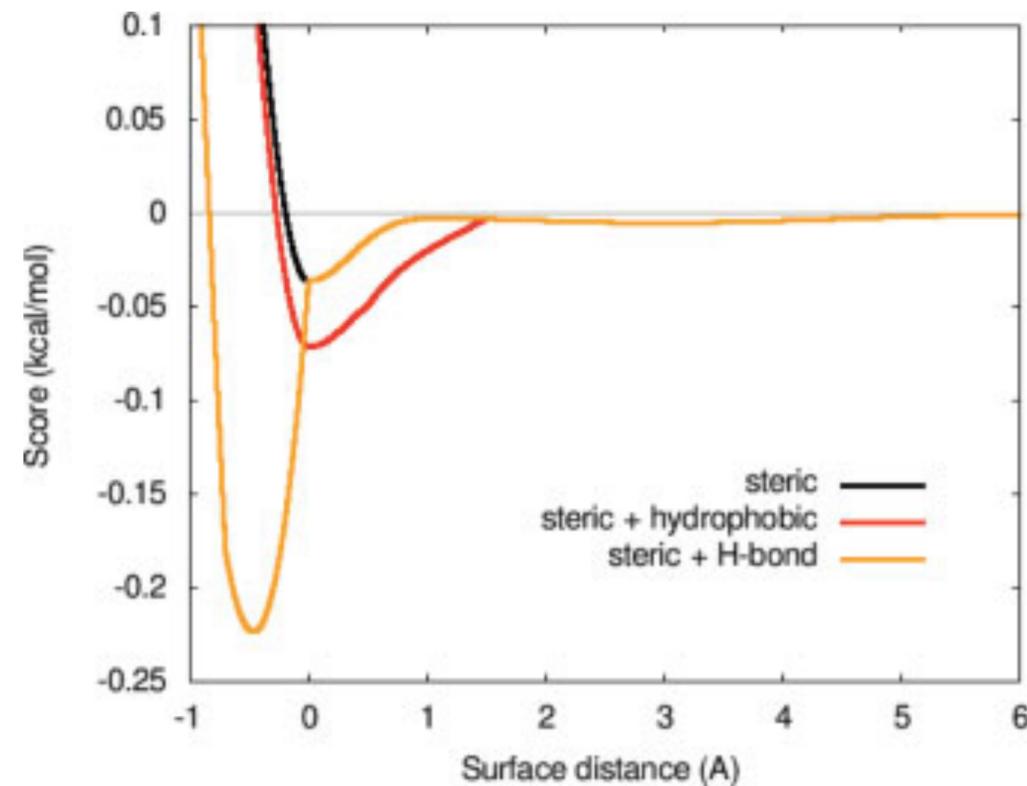
$$\text{gauss}_1(d) = w_{\text{gauss}_1} e^{-(d/0.5)^2}$$

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

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

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

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



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.

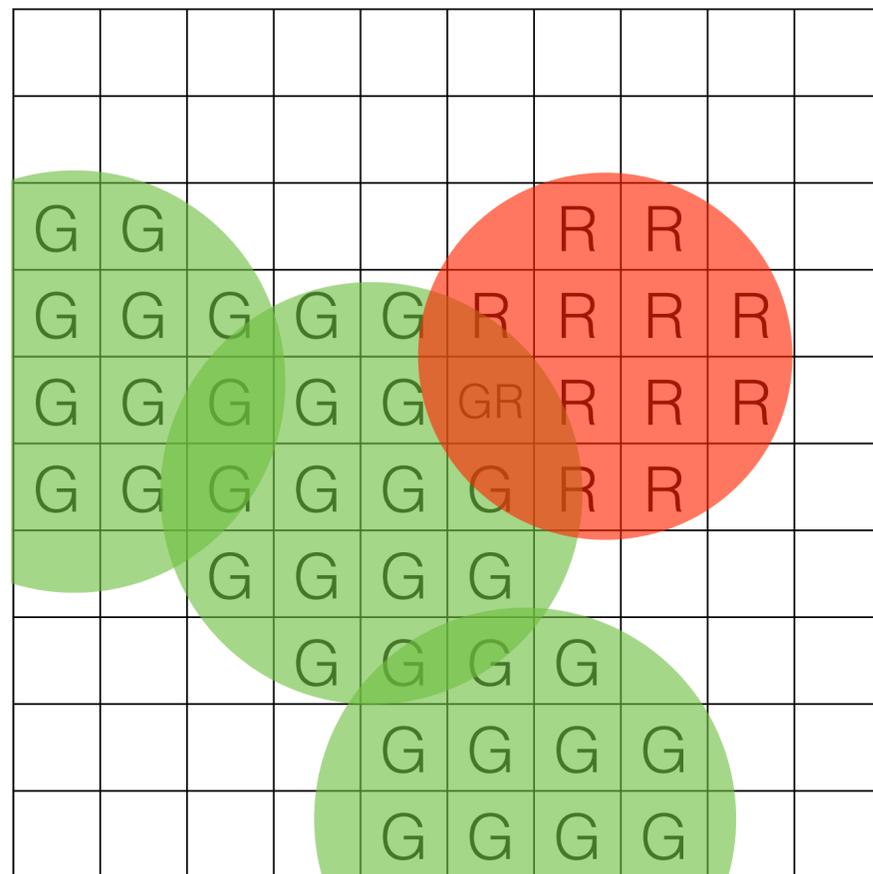
Protein-Ligand Scoring



Protein-Ligand Scoring

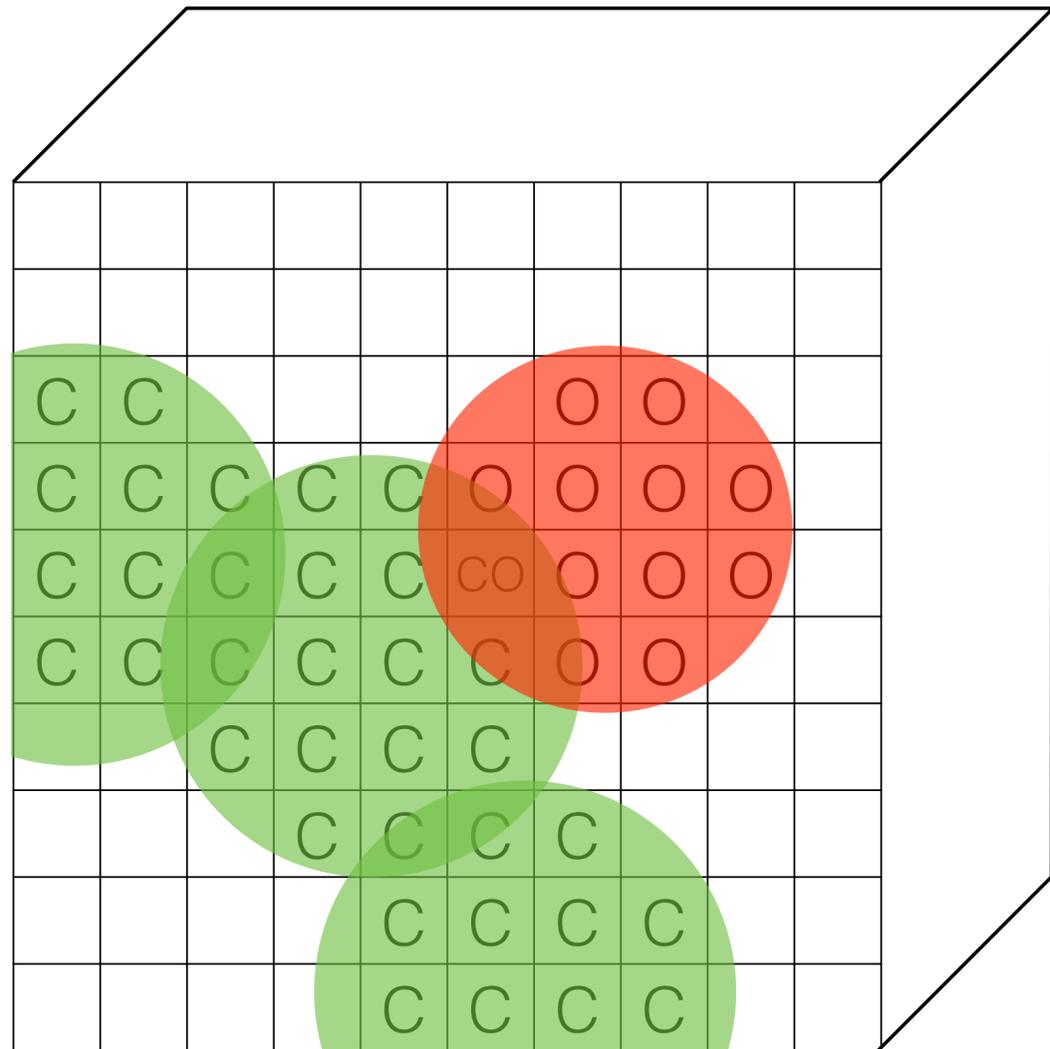


Protein-Ligand Representation



(R,G,B) pixel

Protein-Ligand Representation



(R,G,B) pixel \rightarrow

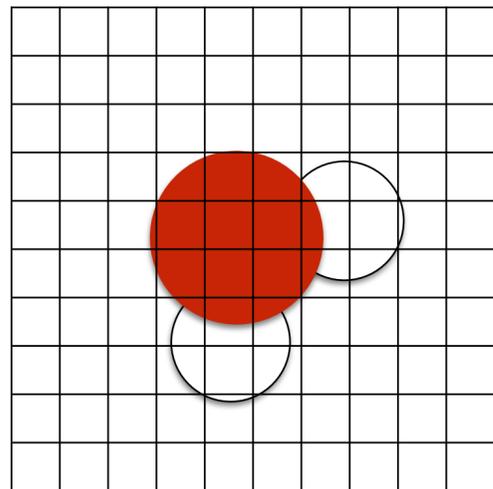
(Carbon, Nitrogen, Oxygen,...) **voxel**

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

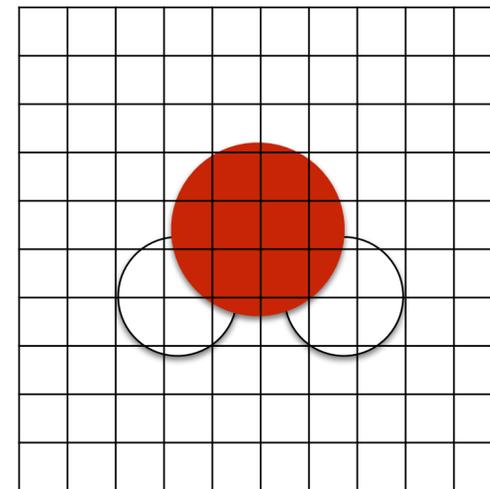
Why Grids?

Cons

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



≠



Pros

- clear spatial relationships
- amazingly parallel
- easy to interpret

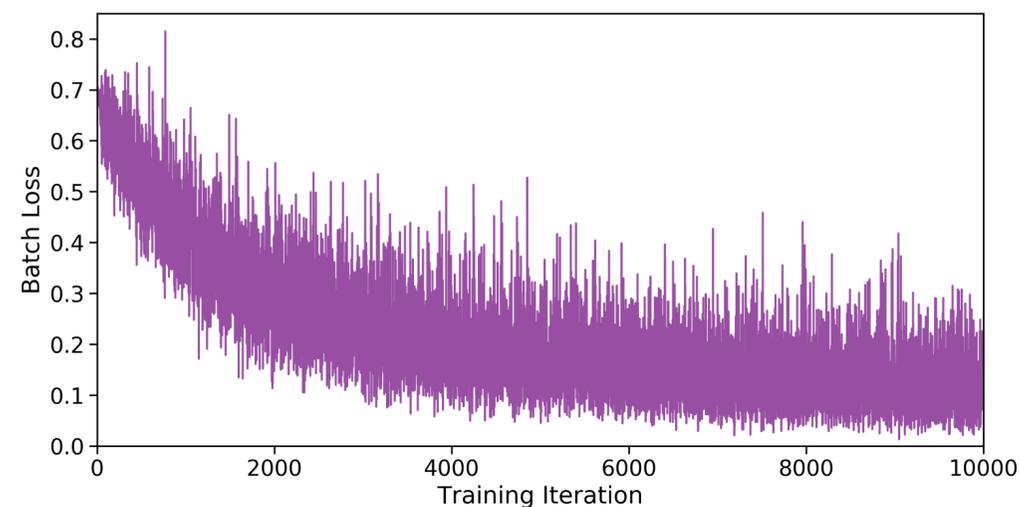
libmolgrid



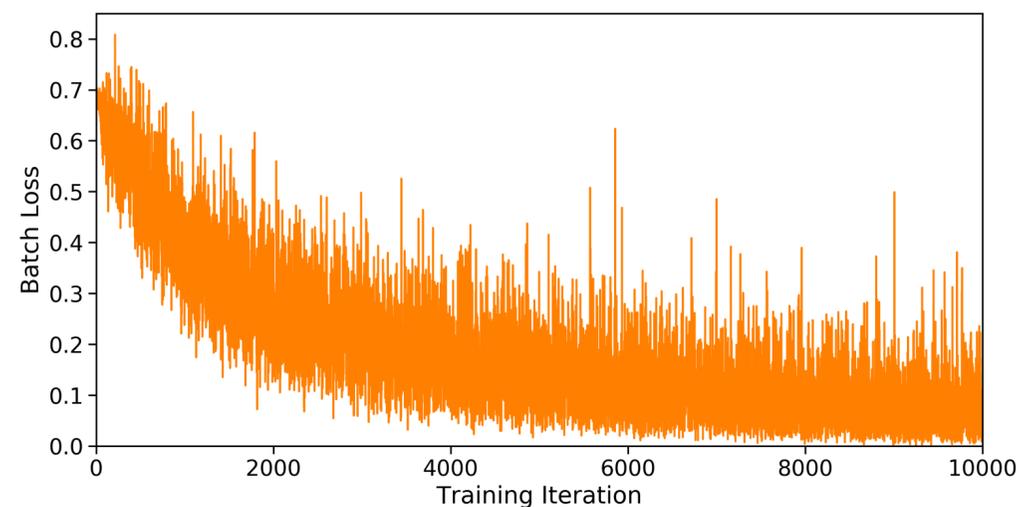
Sunseri, Jocelyn S

Dr. Jocelyn Sunseri

Caffe Training



PyTorch Training



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

3038

Altmetric

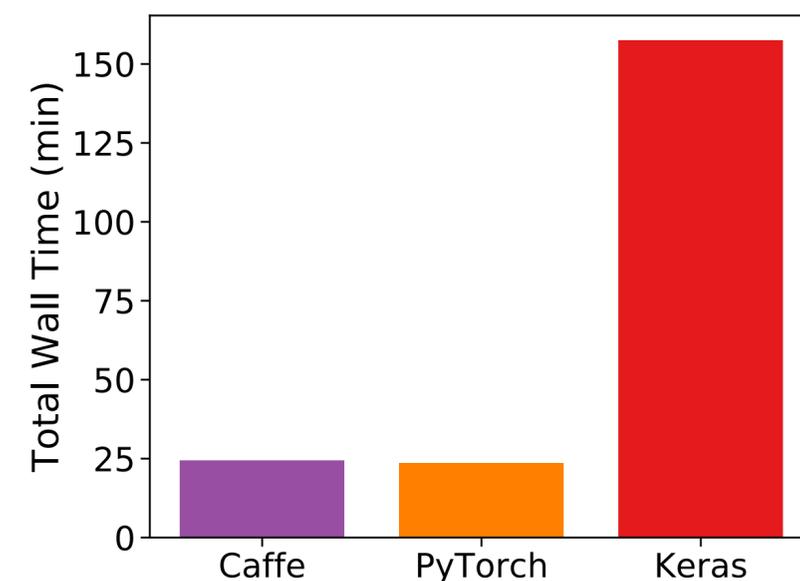
13

Citations

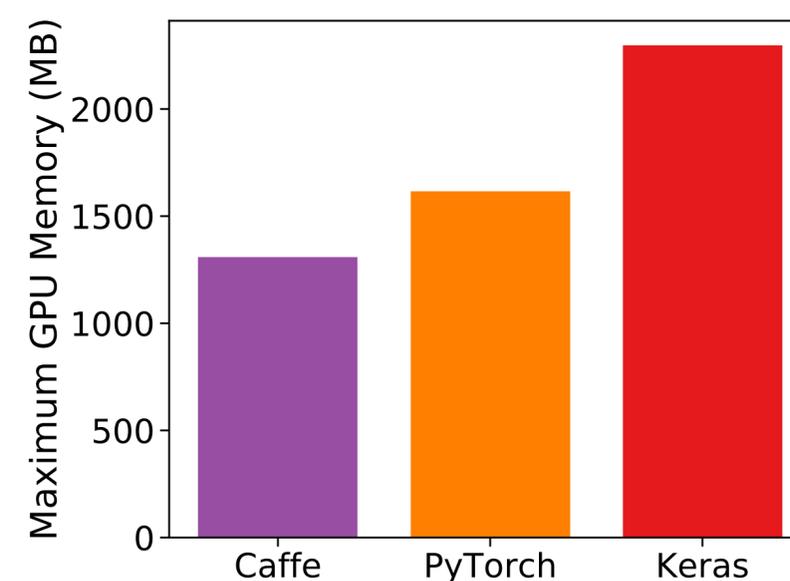
2

[LEARN ABOUT THESE METRICS](#)

GPU Performance



GPU Memory Utilization



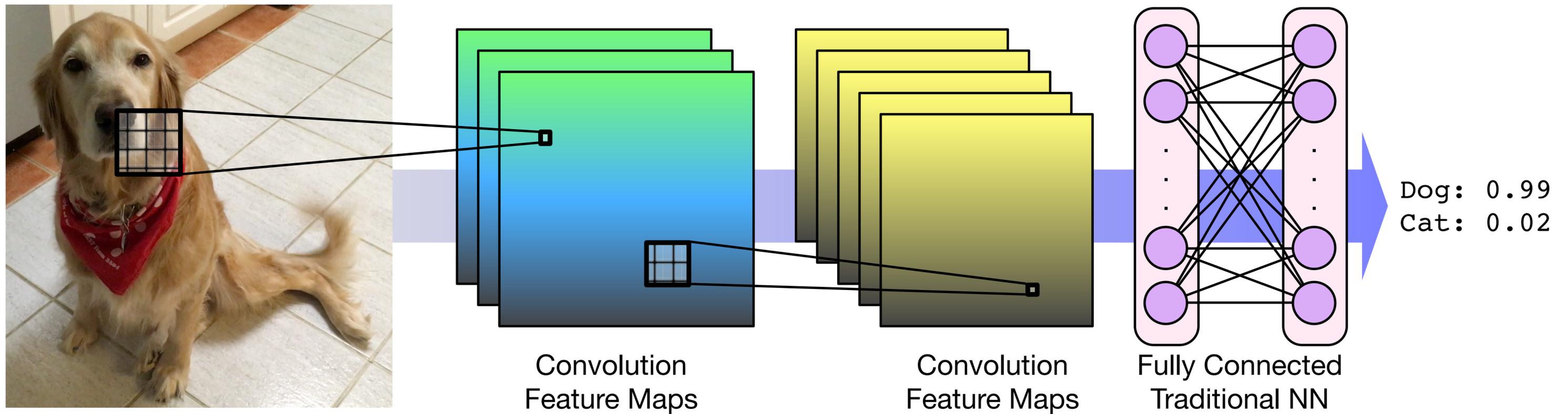
```
e = molgrid.ExampleProvider(balanced=True, shuffle=True)
e.populate('examples.txt')
```

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

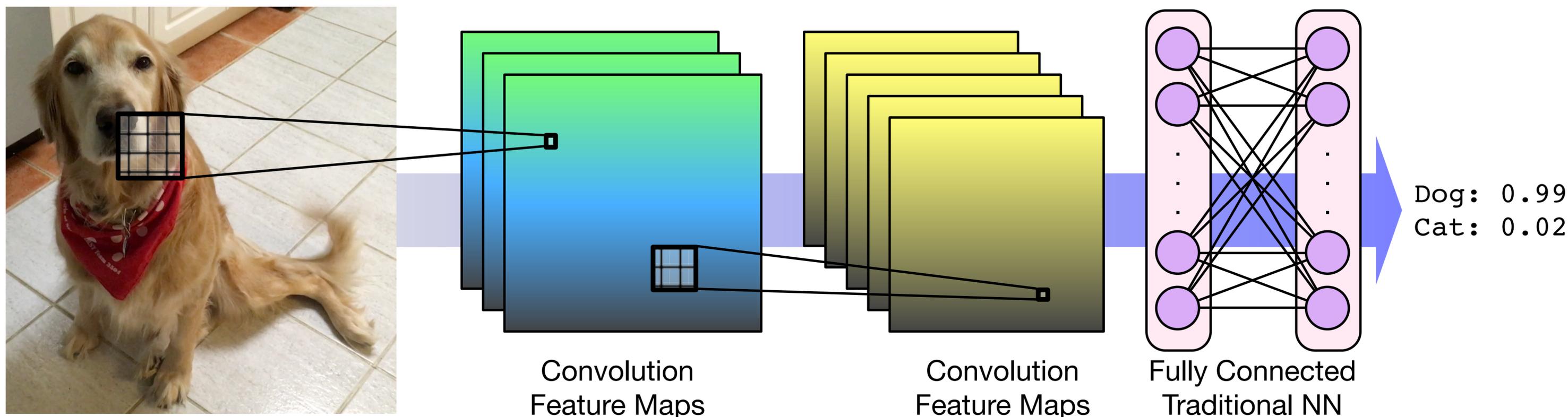
```
batch = e.next_batch(batch_size)
gmaker.forward(batch, input_tensor,
                random_translation=0, random_rotation=True)
```

 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.

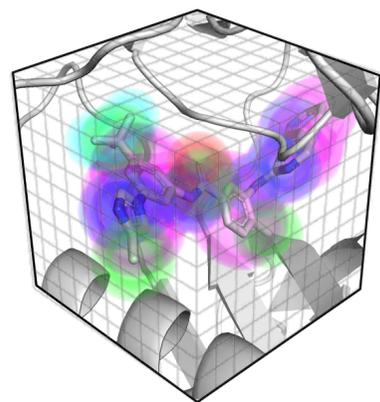
Sunseri J, King JE, Francoeur PG, Koes DR. Convolutional neural network scoring and minimization in the D3R 2017 community challenge. *Journal of computer-aided molecular design*. 2019 Jan 15;33(1):19-34.

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.

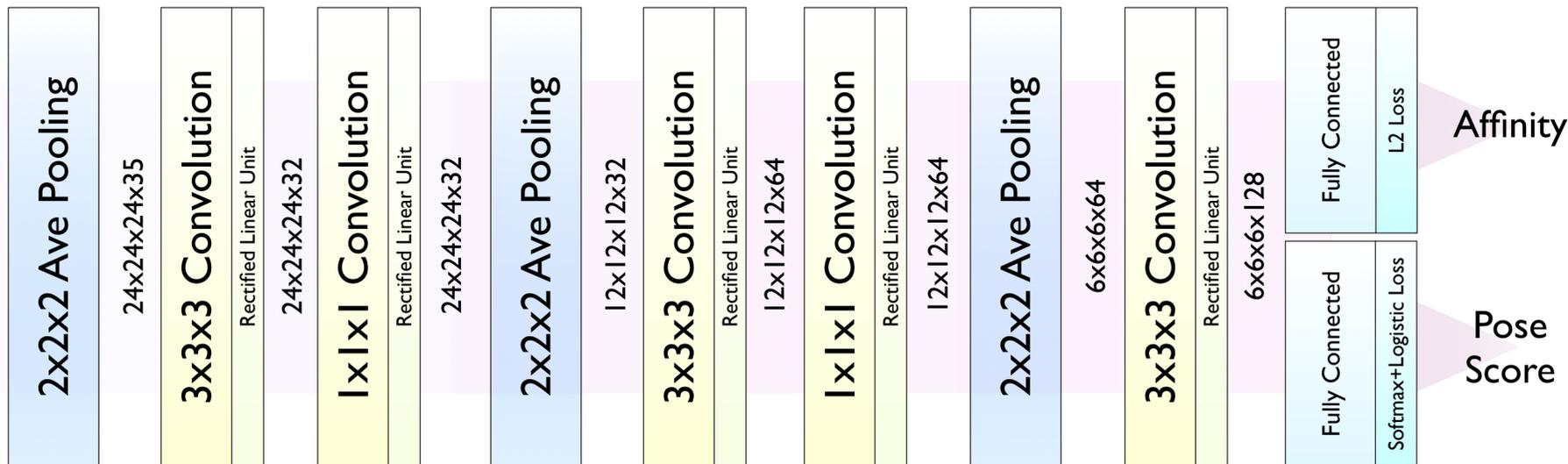
Francoeur PG, Masuda T, Sunseri J, Jia A, Iovanisci RB, Snyder I, Koes DR. Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design. *Journal of Chemical Information and Modeling*. 2020 Aug 31;60(9):4200-15.

Protein Ligand Scoring

Def2018



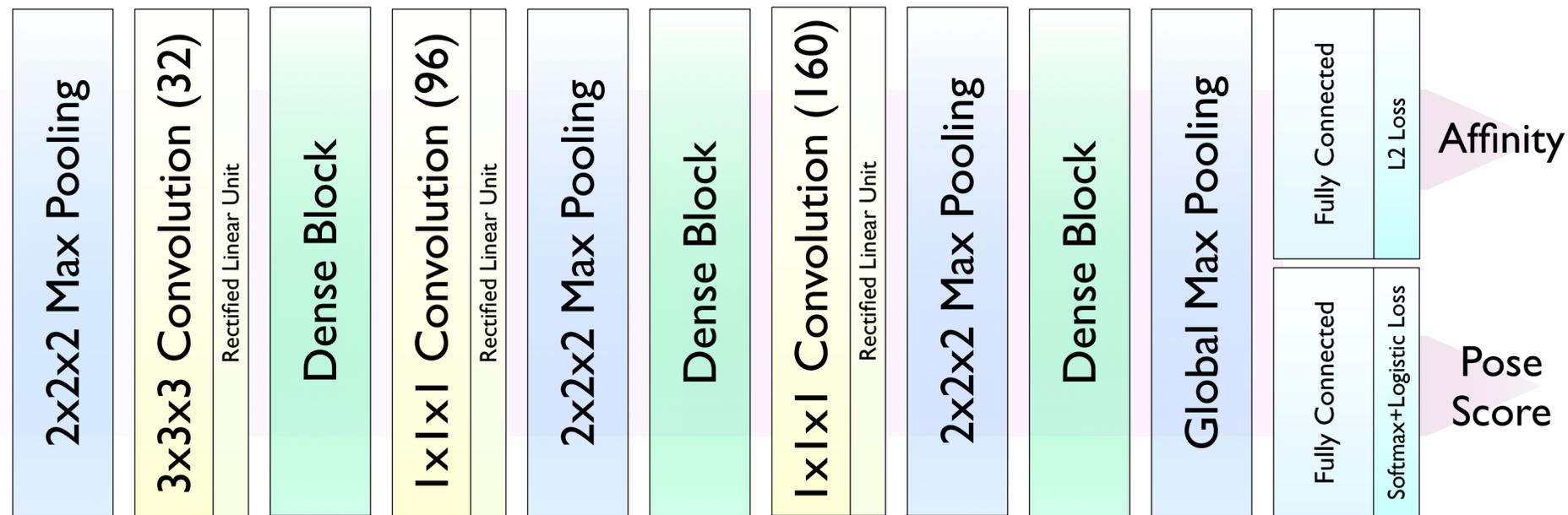
48x48x28 Molecular Grid



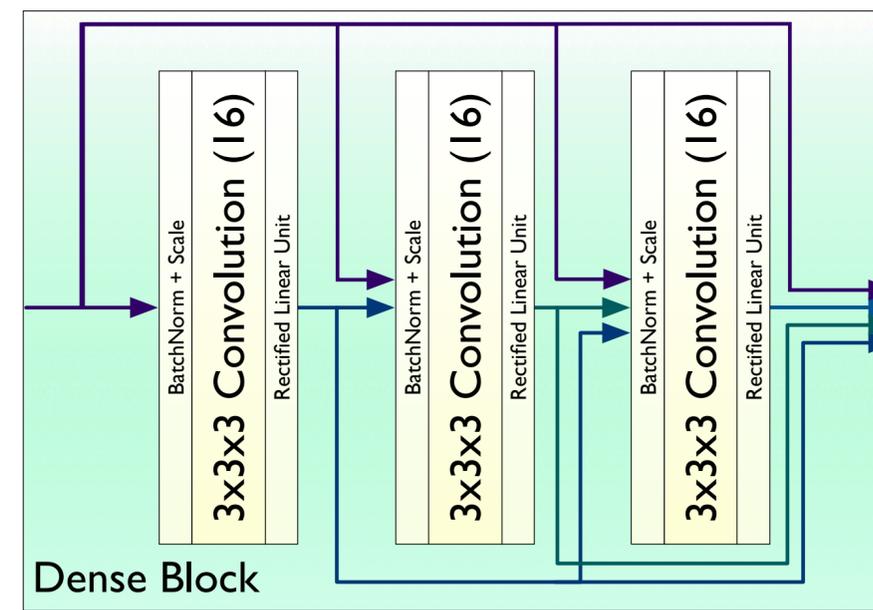
388,736 Parameters

Dense

48x48x48x28 Molecular Grid



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>

Copyright © 2020 American Chemical Society

[RIGHTS & PERMISSIONS](#)

Article Views

1420

[LEARN ABOUT THESE METRICS](#)

Altmetric

17

Citations

5

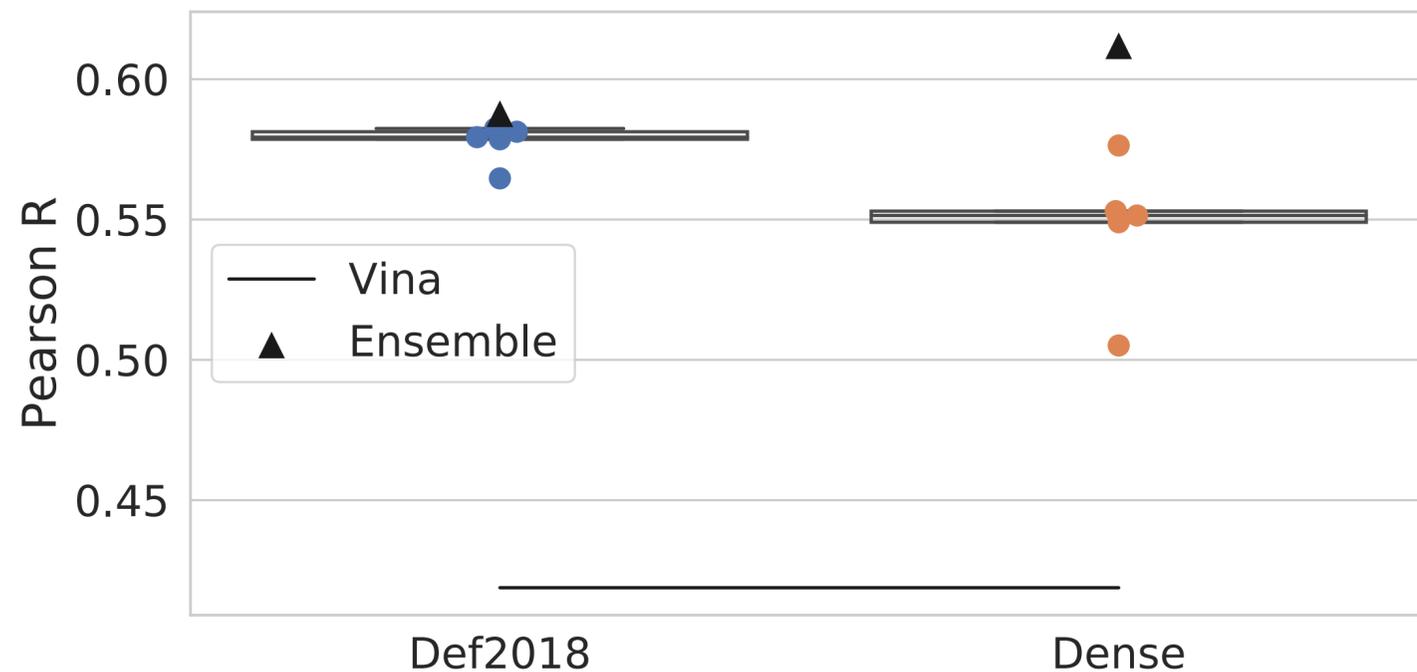
Share



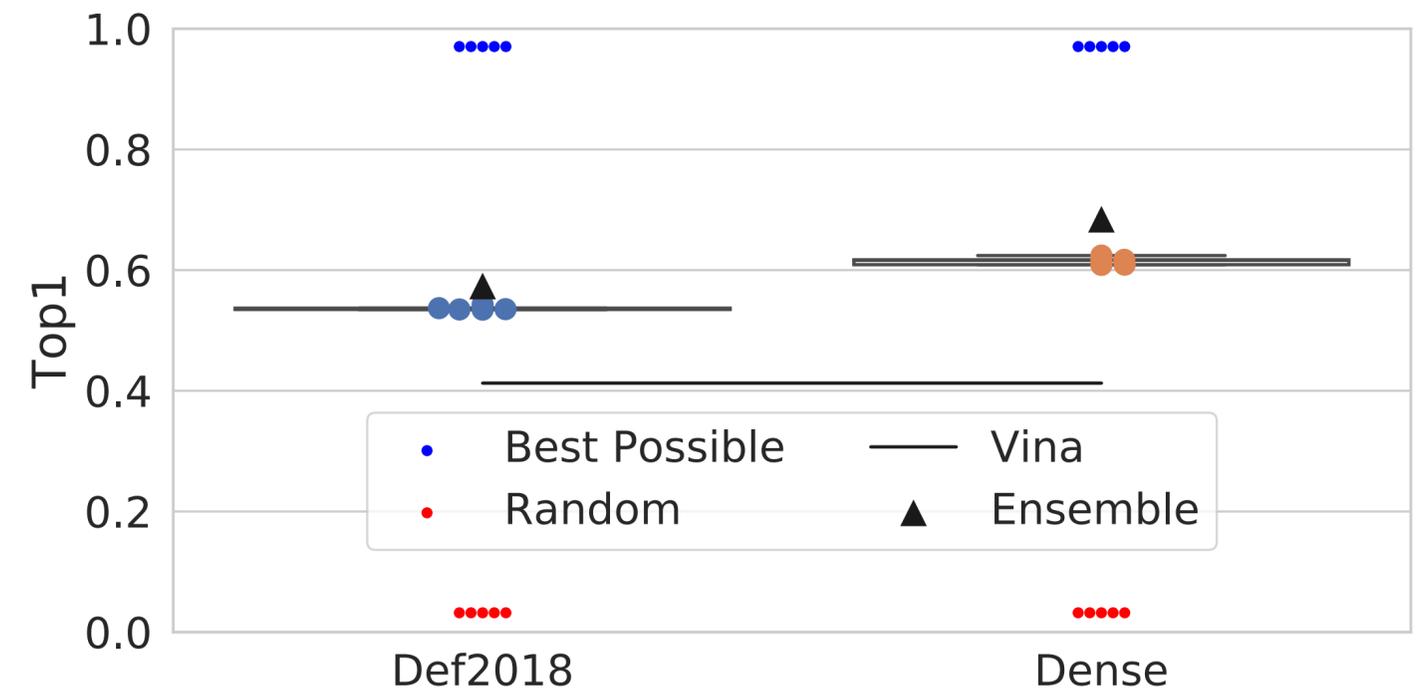
Add to



Export



Affinity Prediction



Pose Selection

GNINA 1.0

<https://github.com/gnina/gnina>

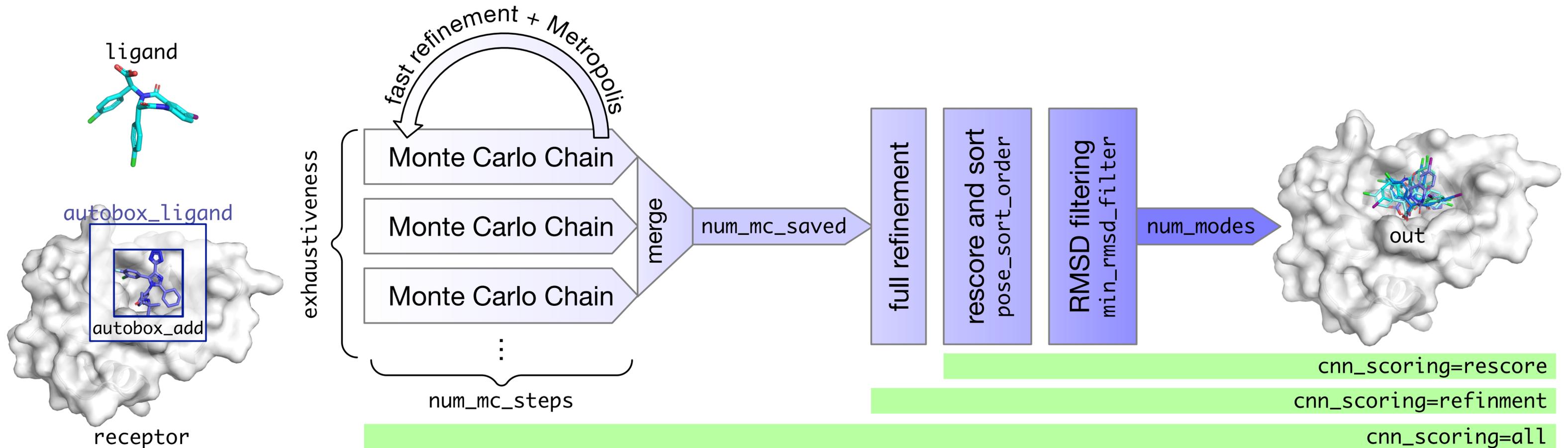
SOFTWARE

Open Access

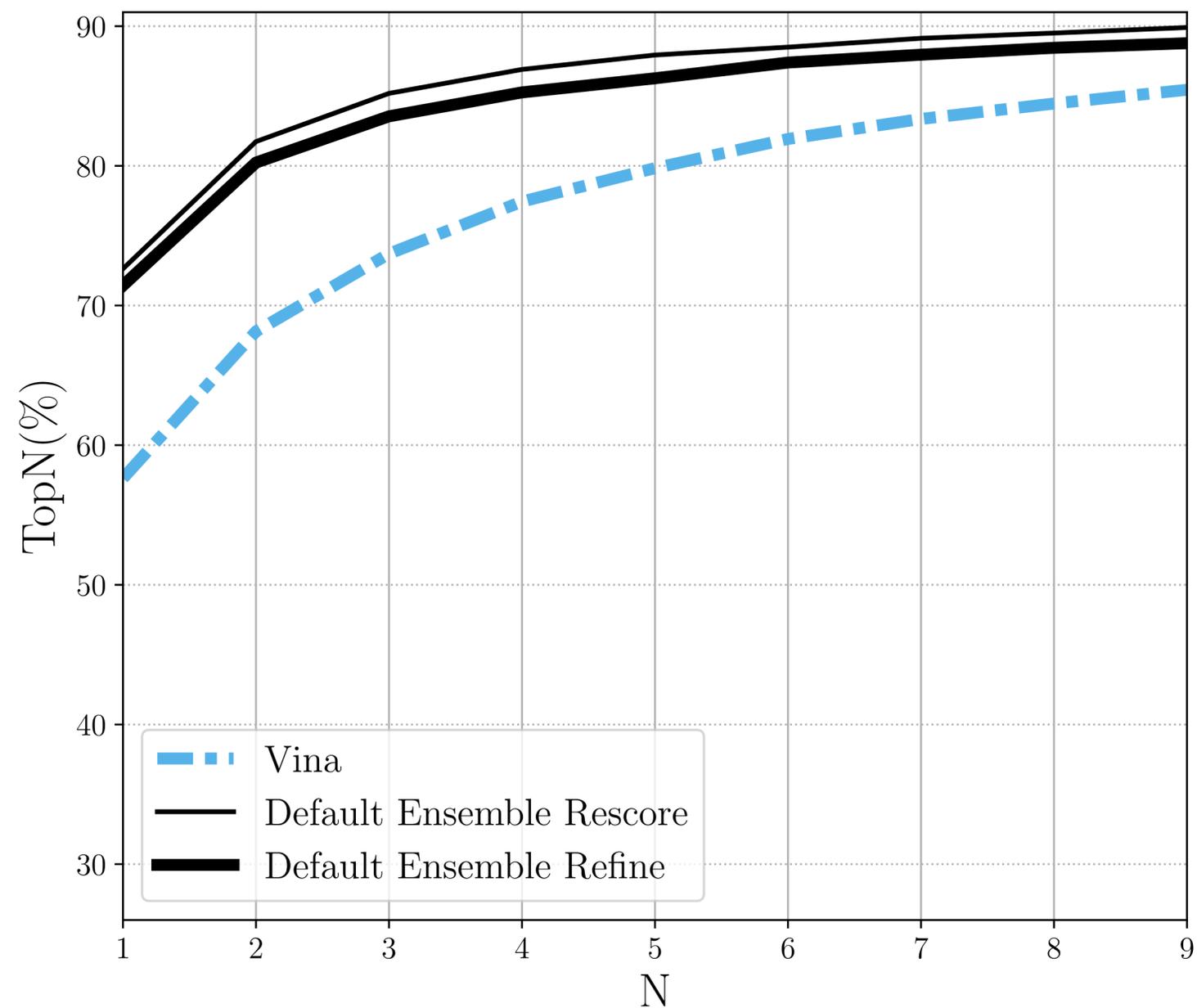


GNINA 1.0: molecular docking with deep learning

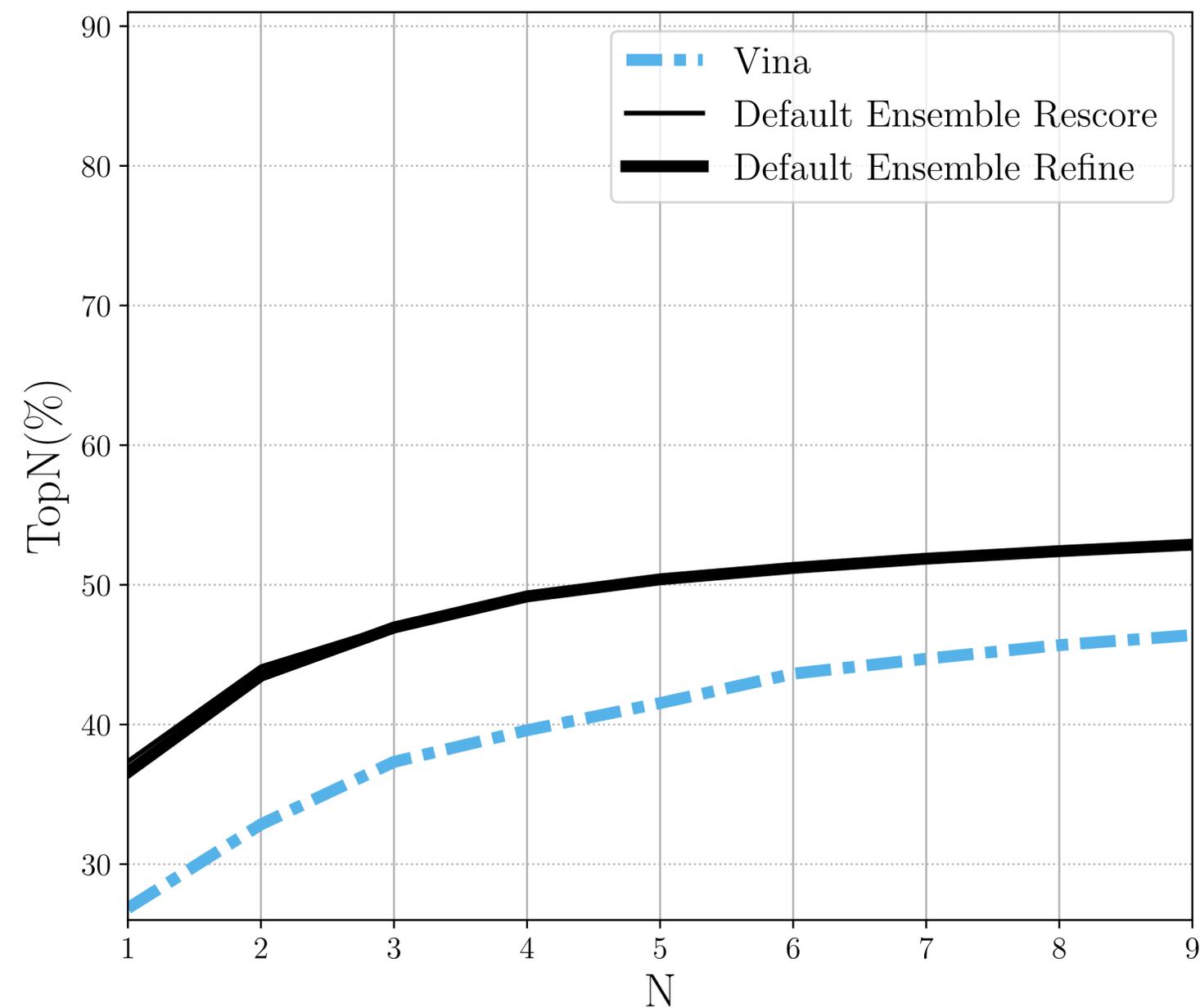
Andrew T. McNutt¹, Paul Francoeur¹, Rishal Aggarwal², Tomohide Masuda¹, Rocco Meli³, Matthew Ragoza¹, Jocelyn Sunseri¹ and David Ryan Koes^{1*}



Using the CNN

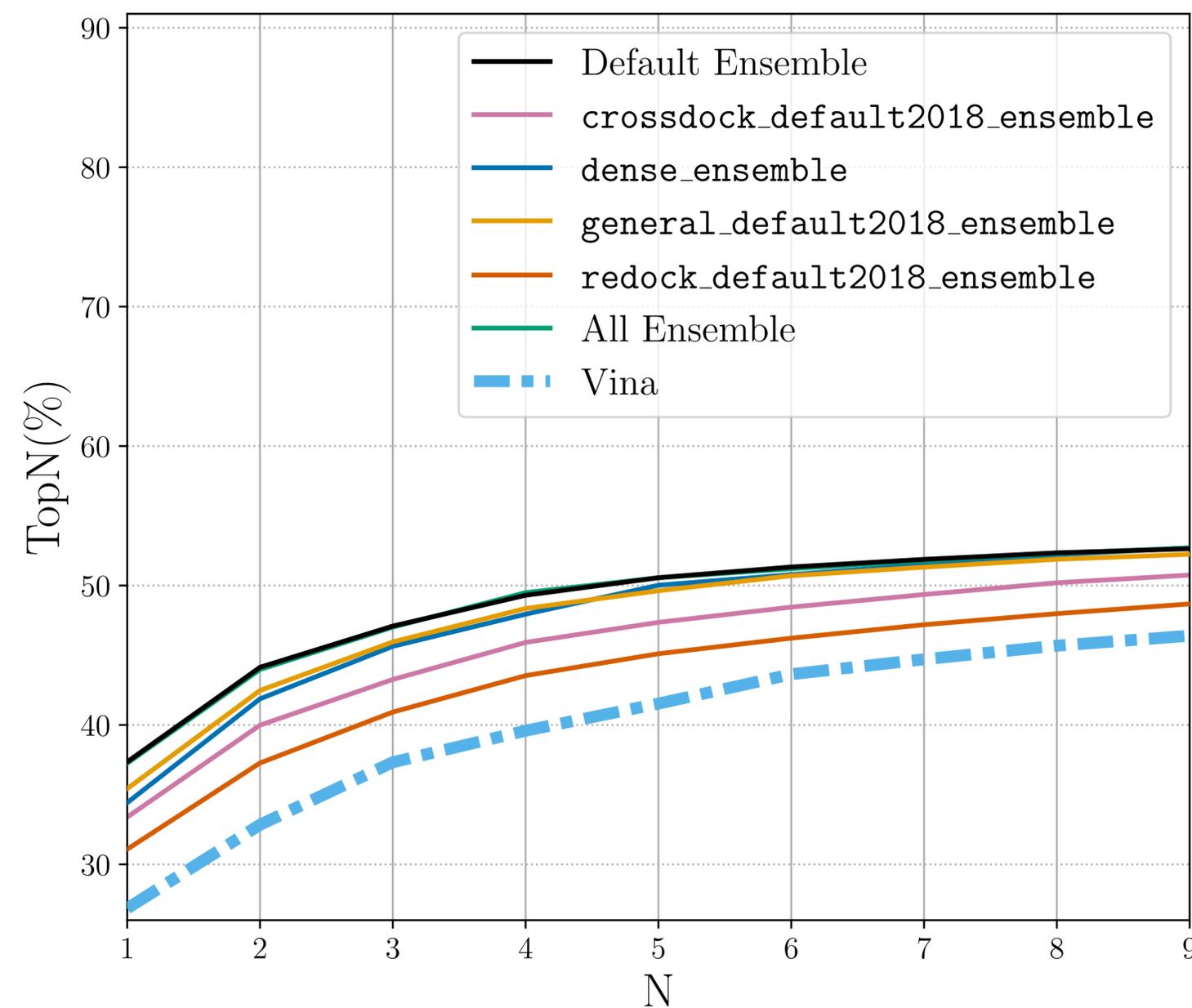
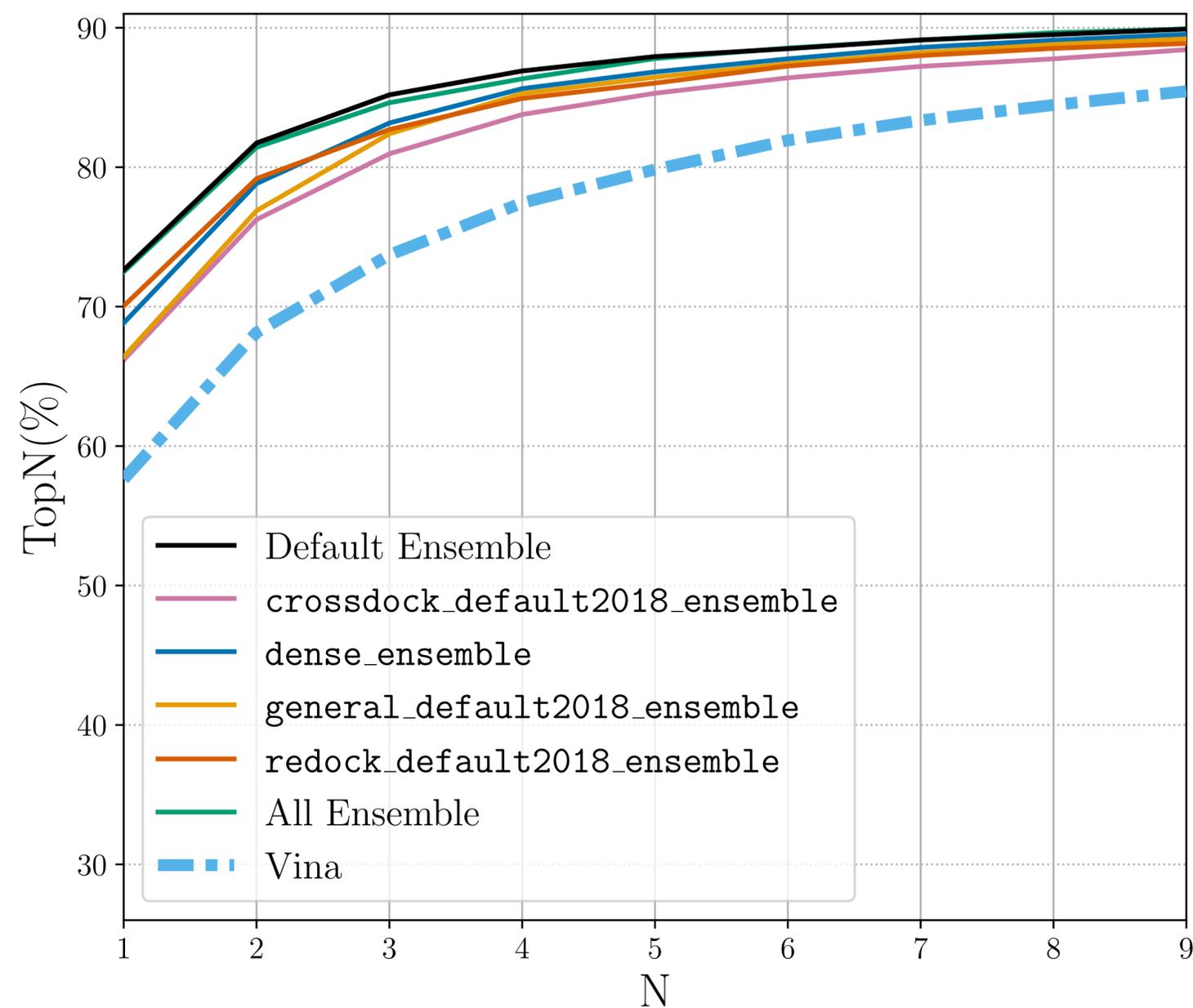


Redocking



Crossdocking

Default CNN Ensemble

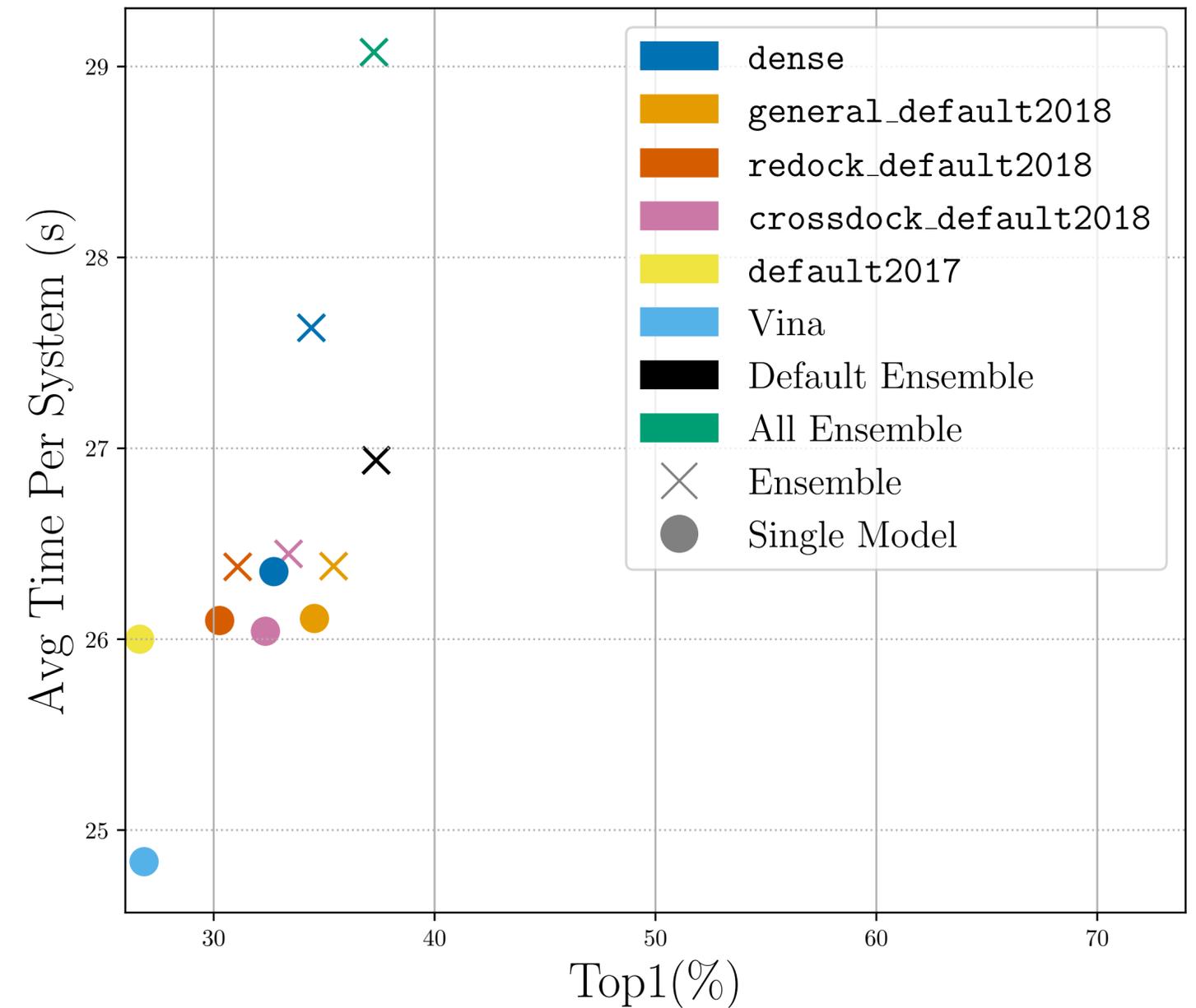
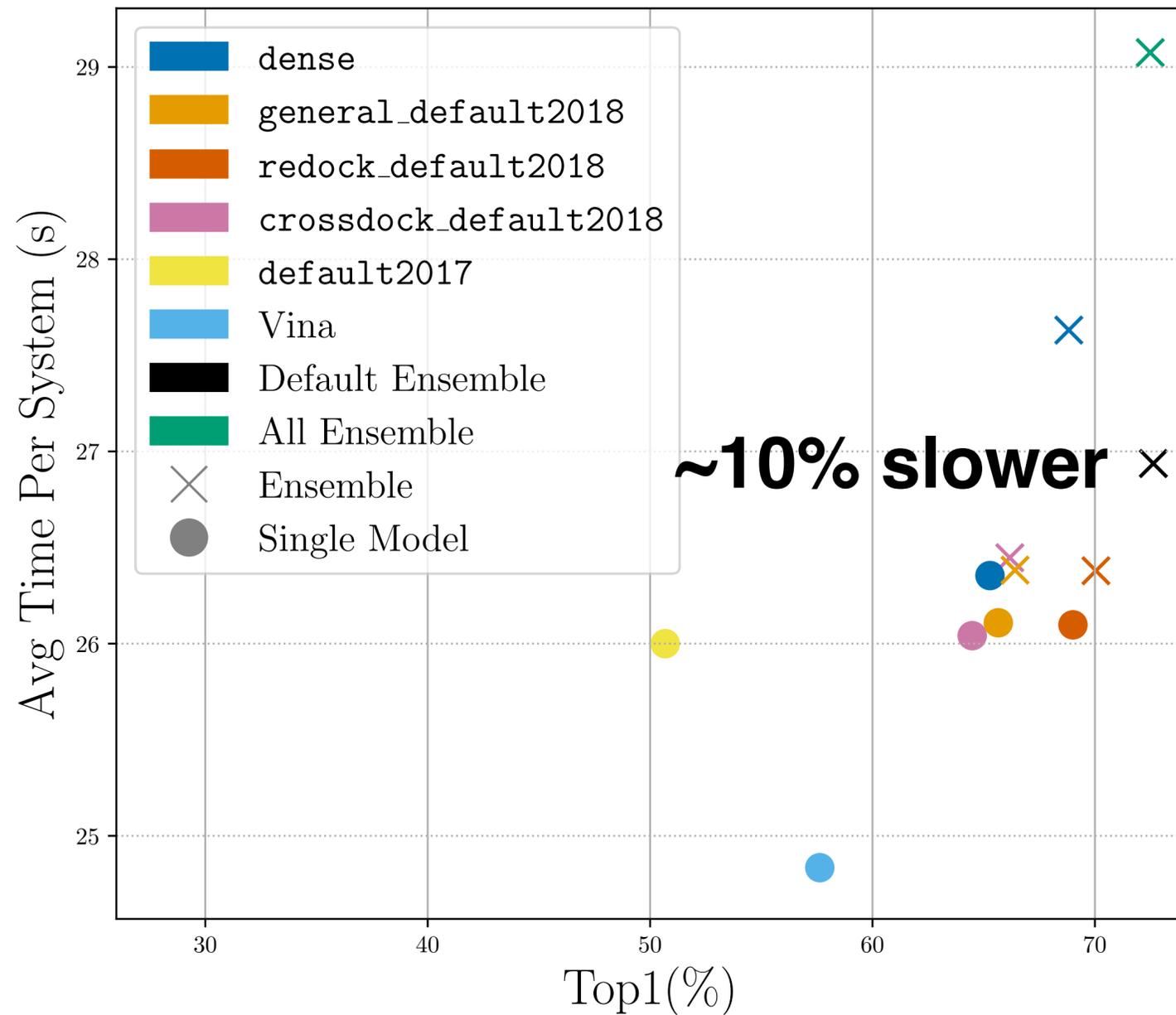


Redocking

Crossdocking

Default CNN Ensemble

GPU Times

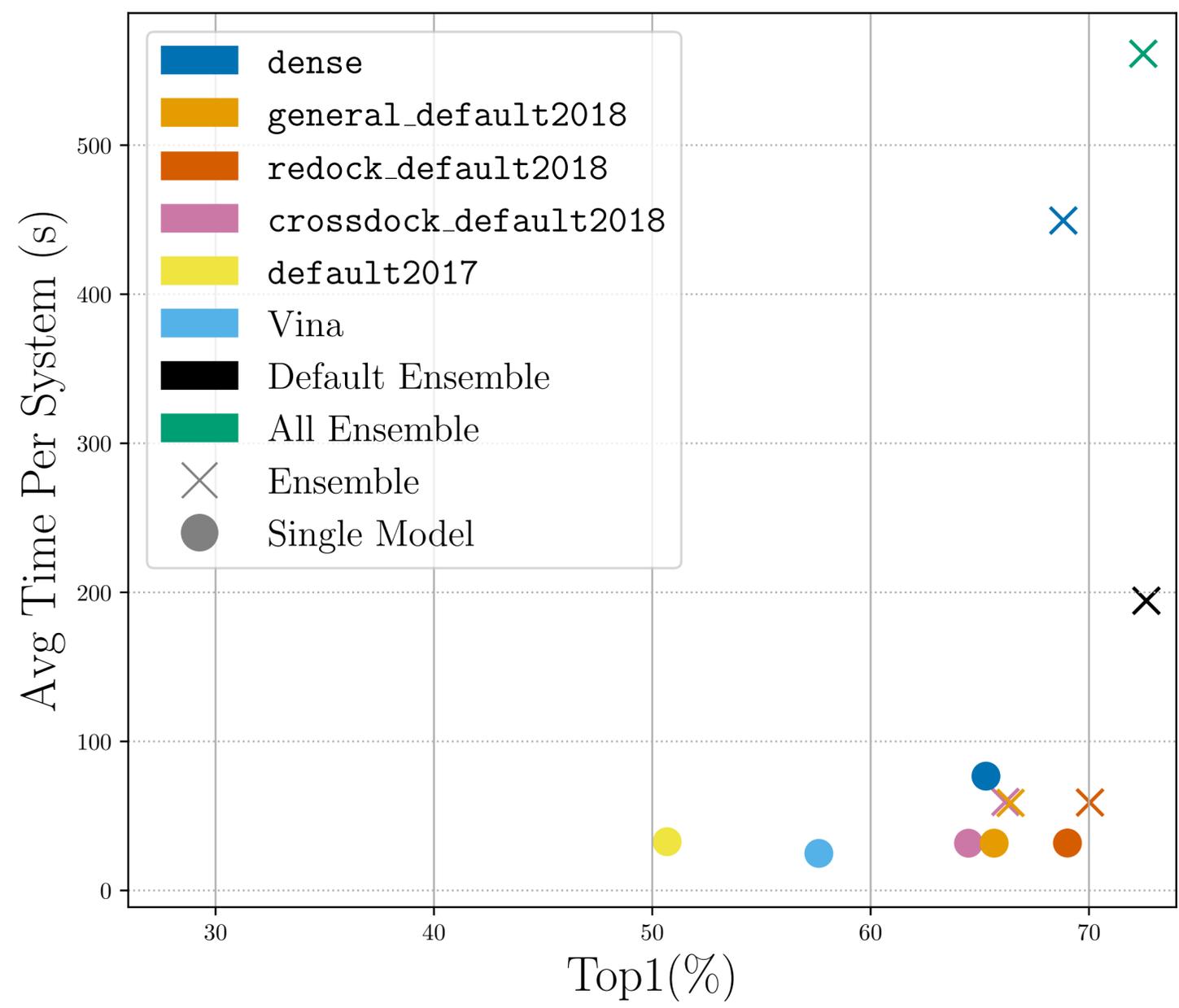


Redocking

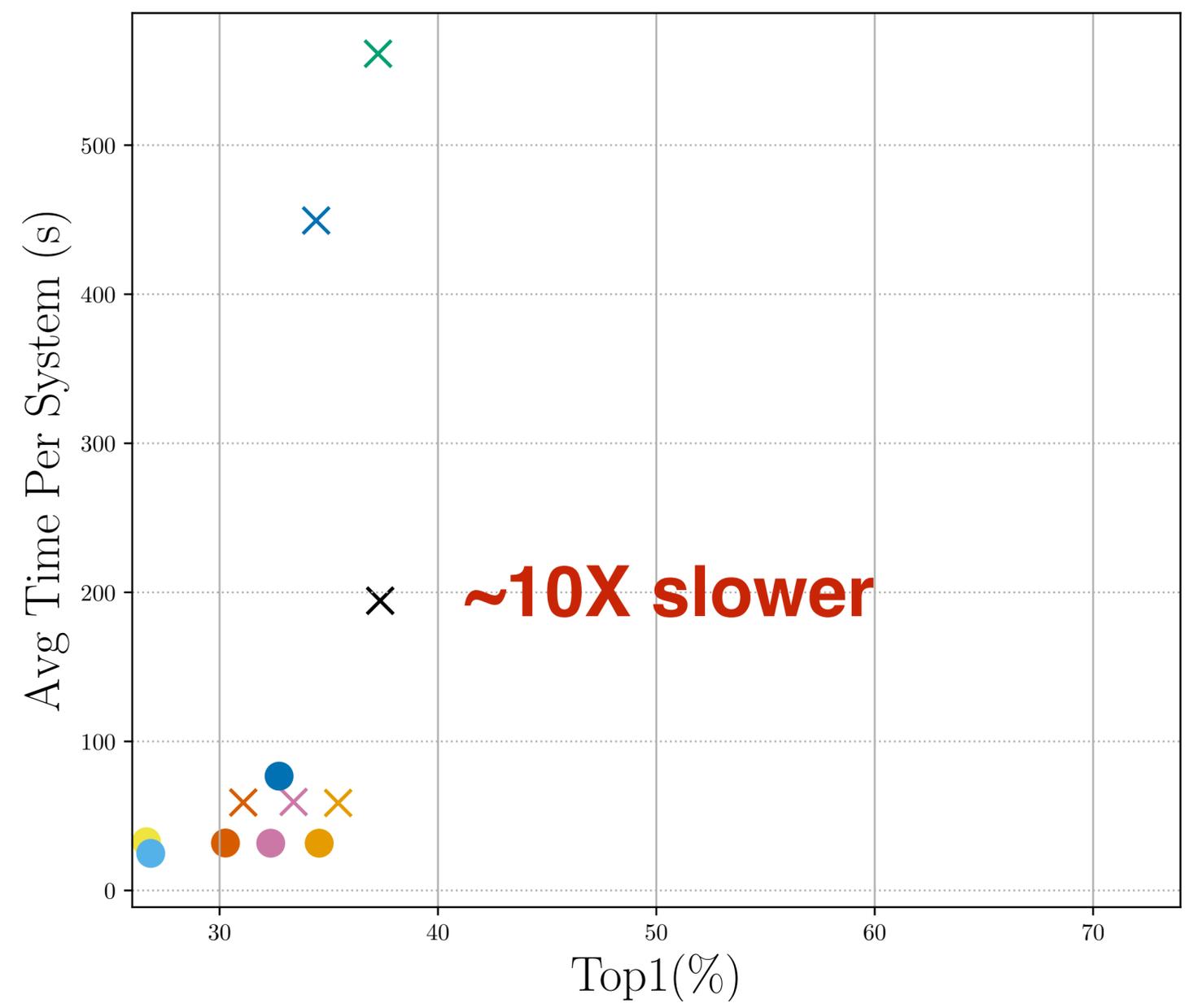
Crossdocking

Default CNN Ensemble

CPU Times



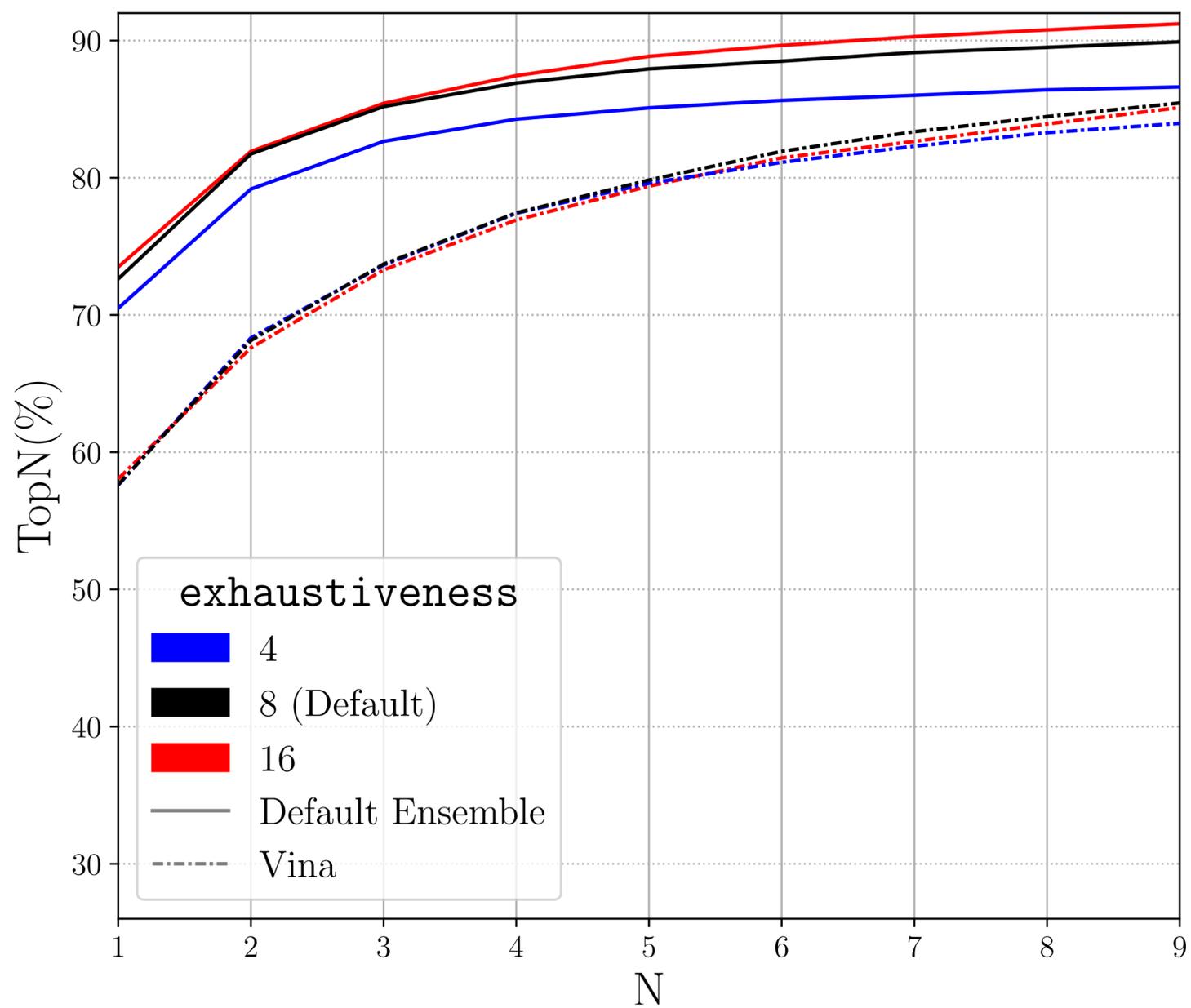
Redocking



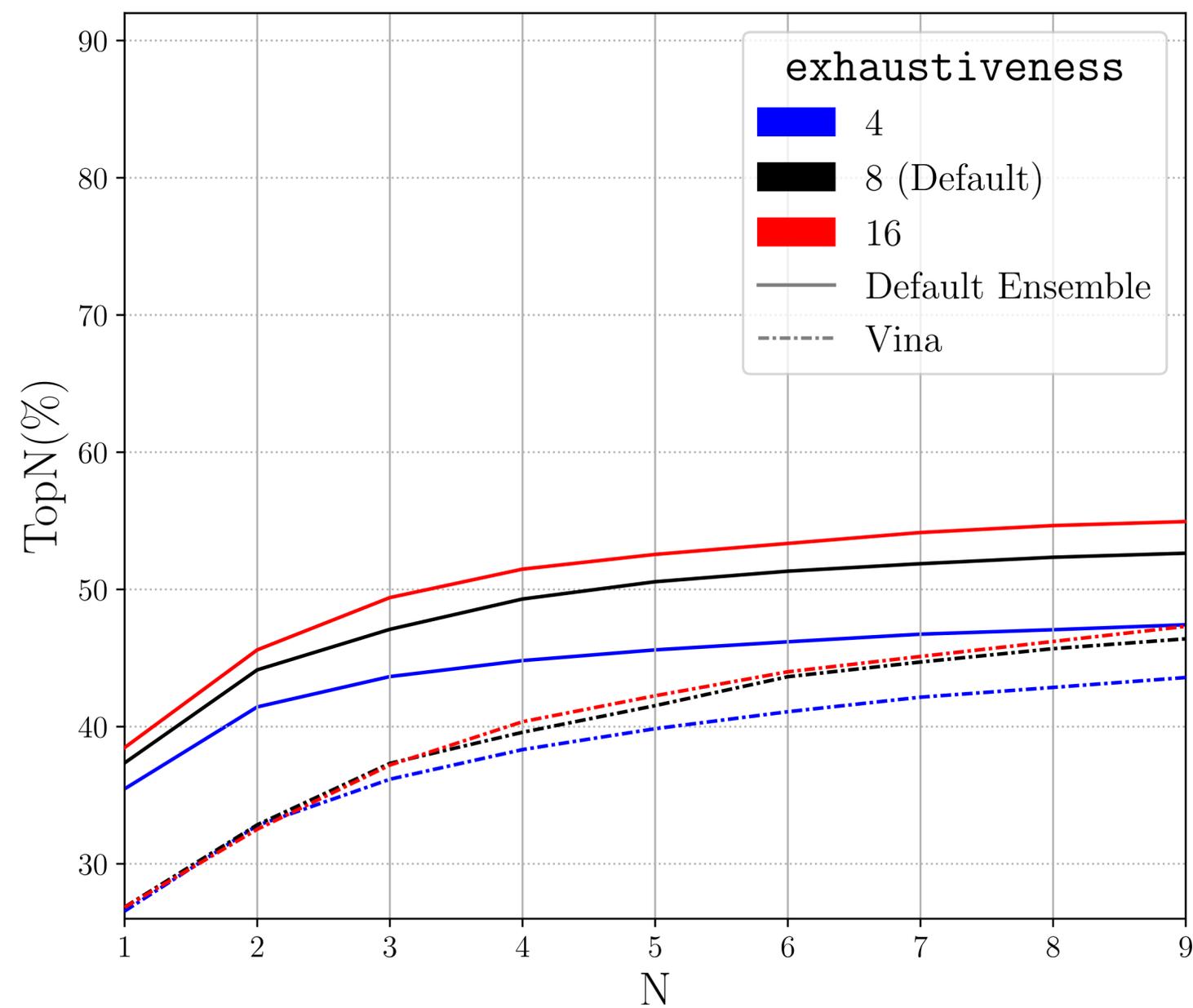
Crossdocking

~10X slower

Exhaustiveness



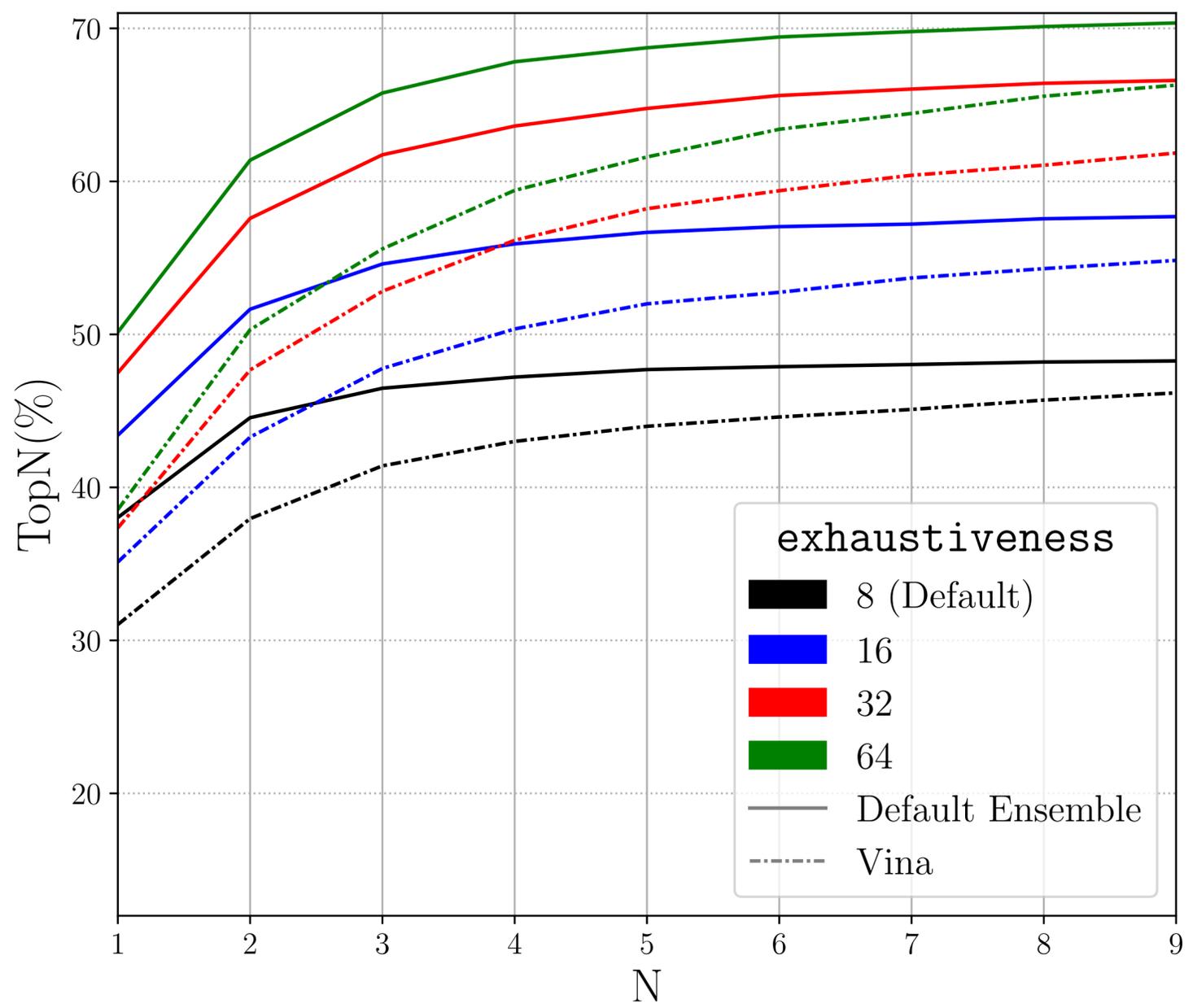
Redocking



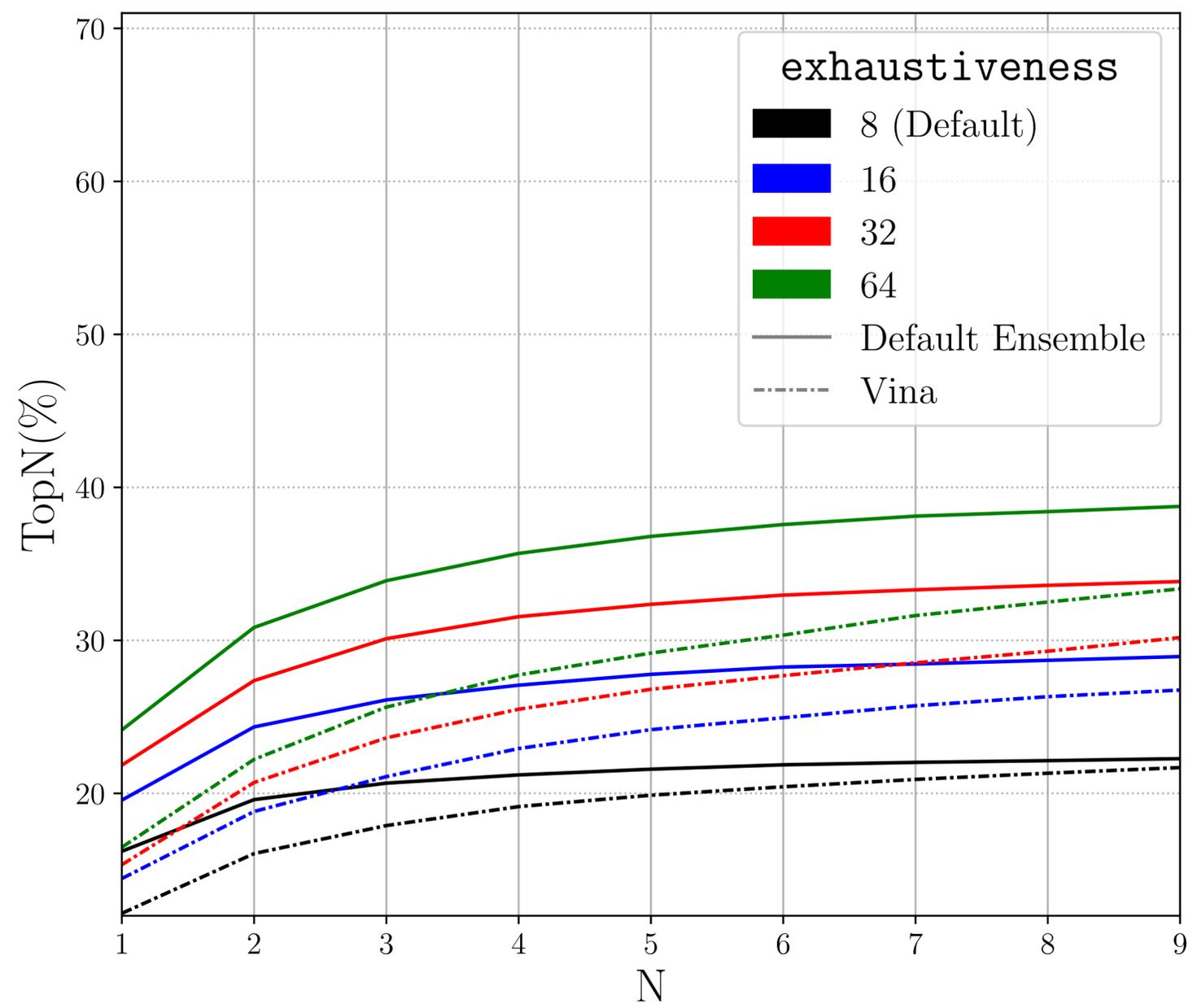
Crossdocking

Whole Protein Docking

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



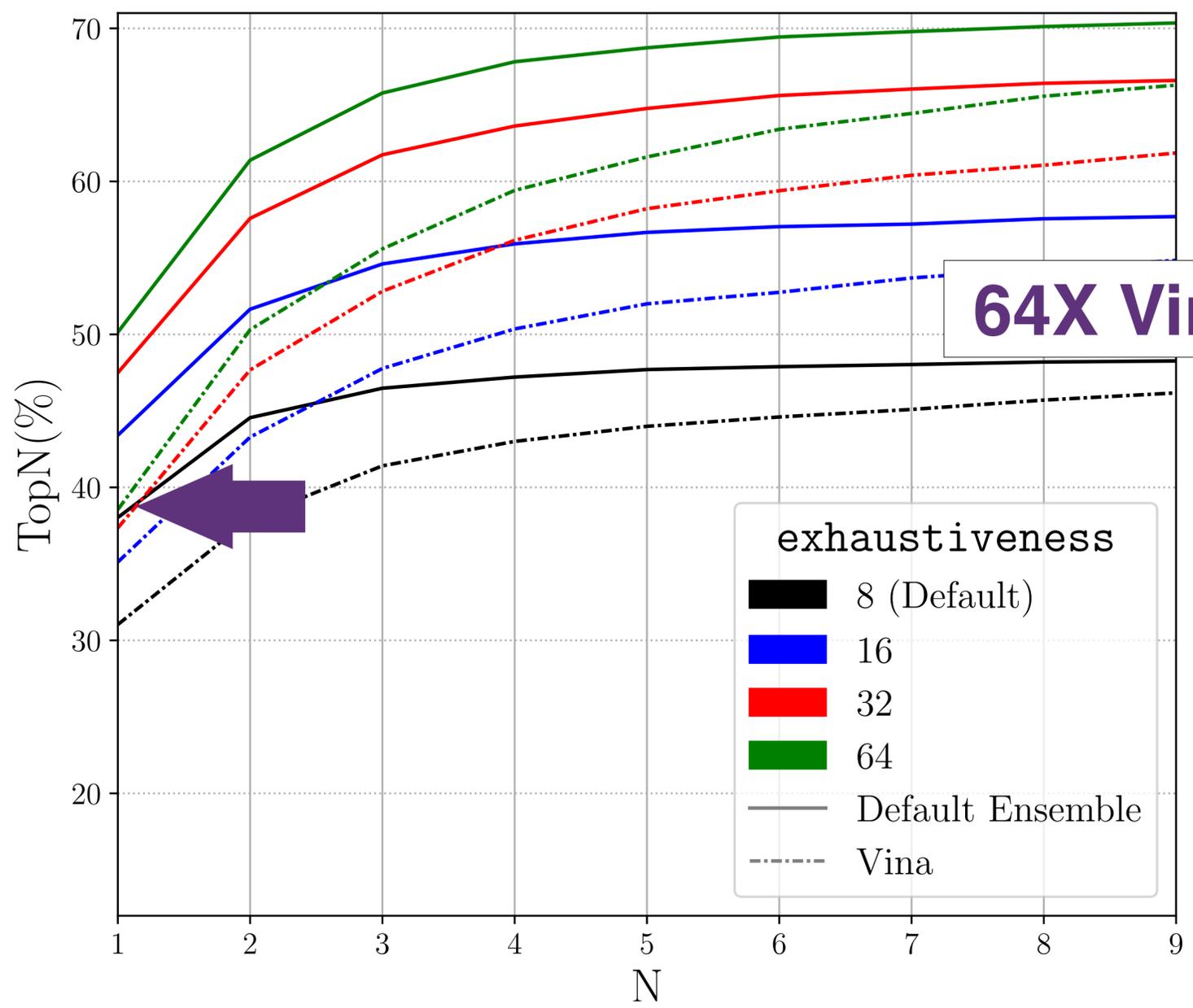
Redocking



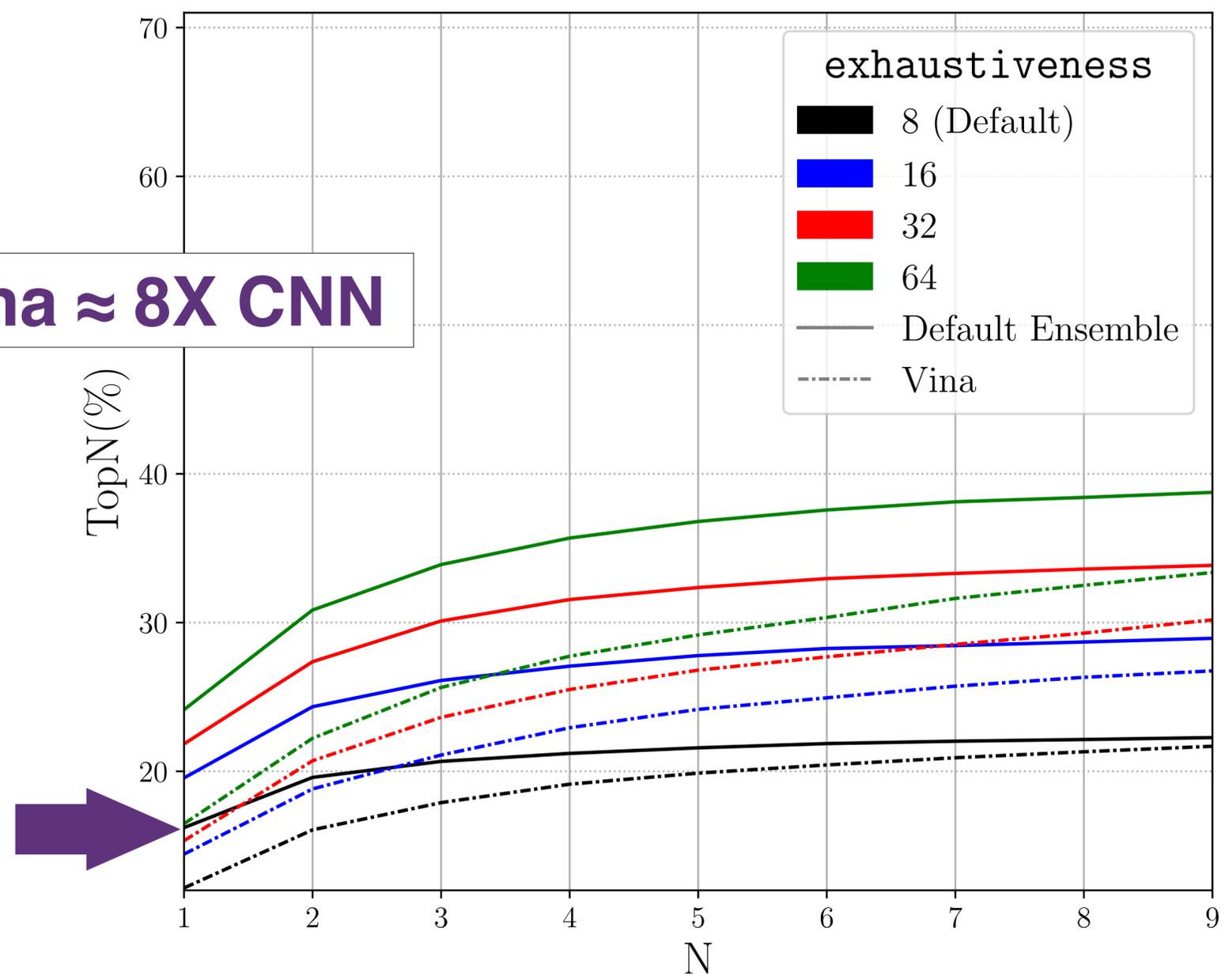
Crossdocking

Whole Protein Docking

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



64X Vina \approx 8X CNN

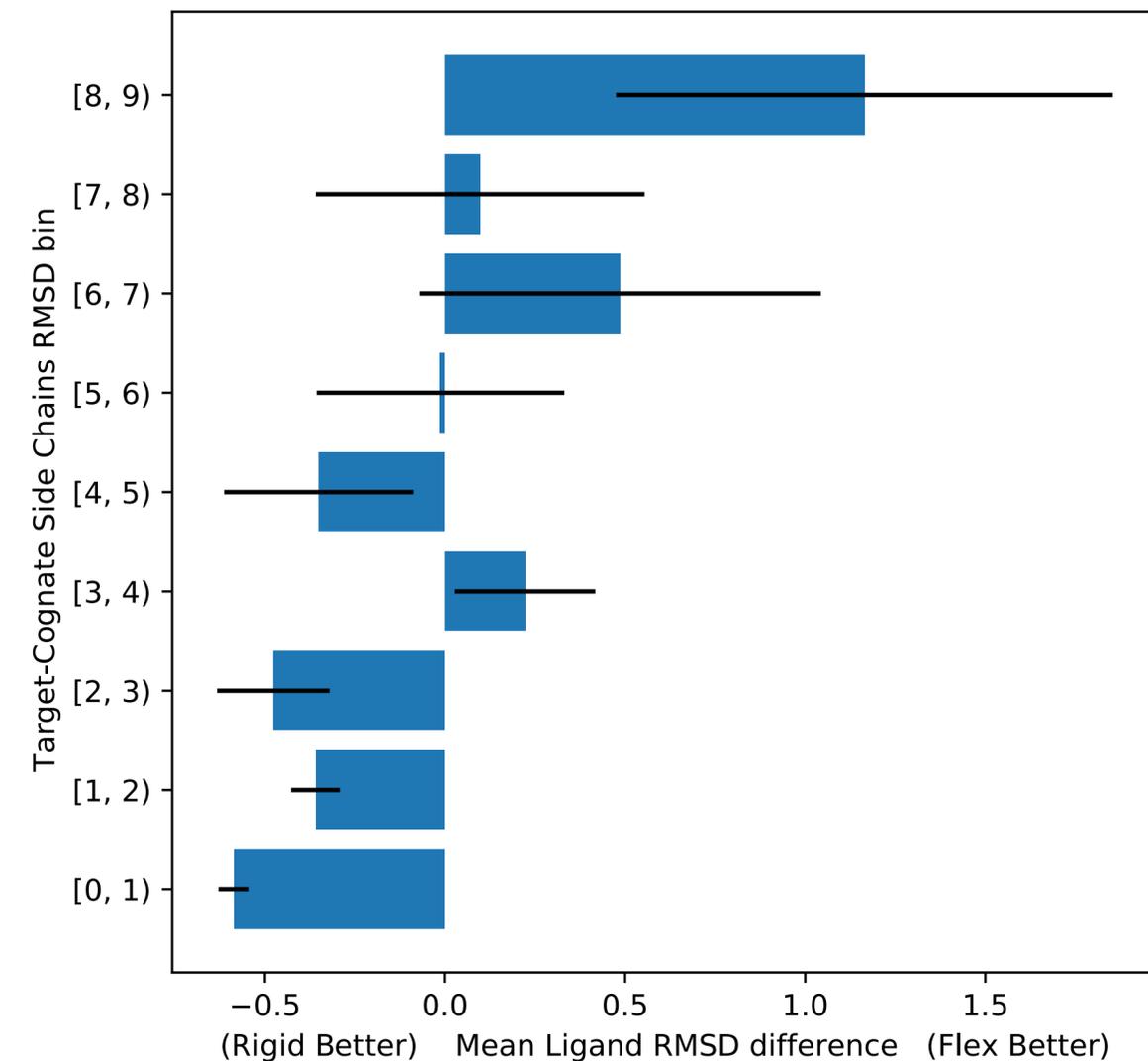
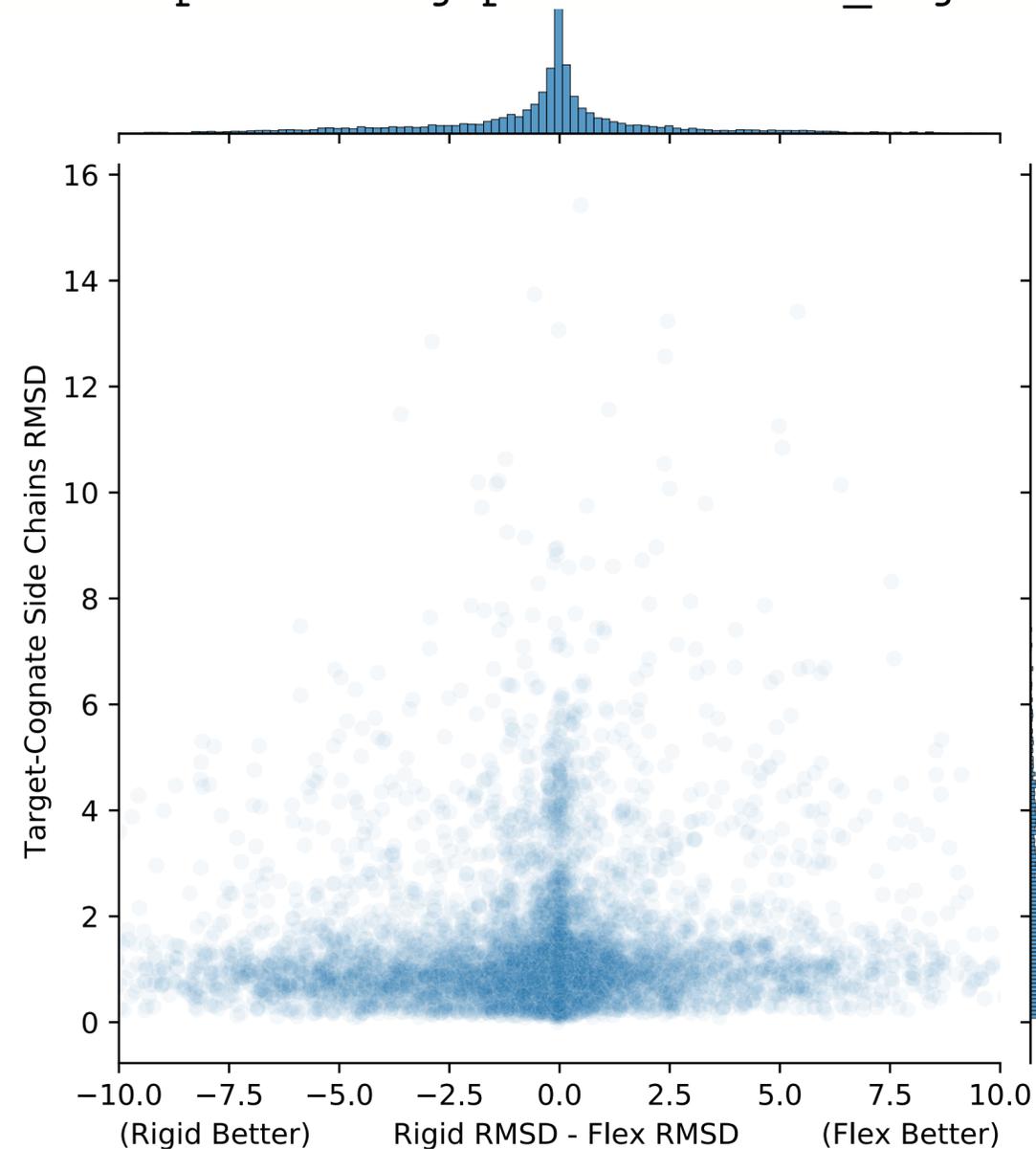


Redocking

Crossdocking

Flexible Docking

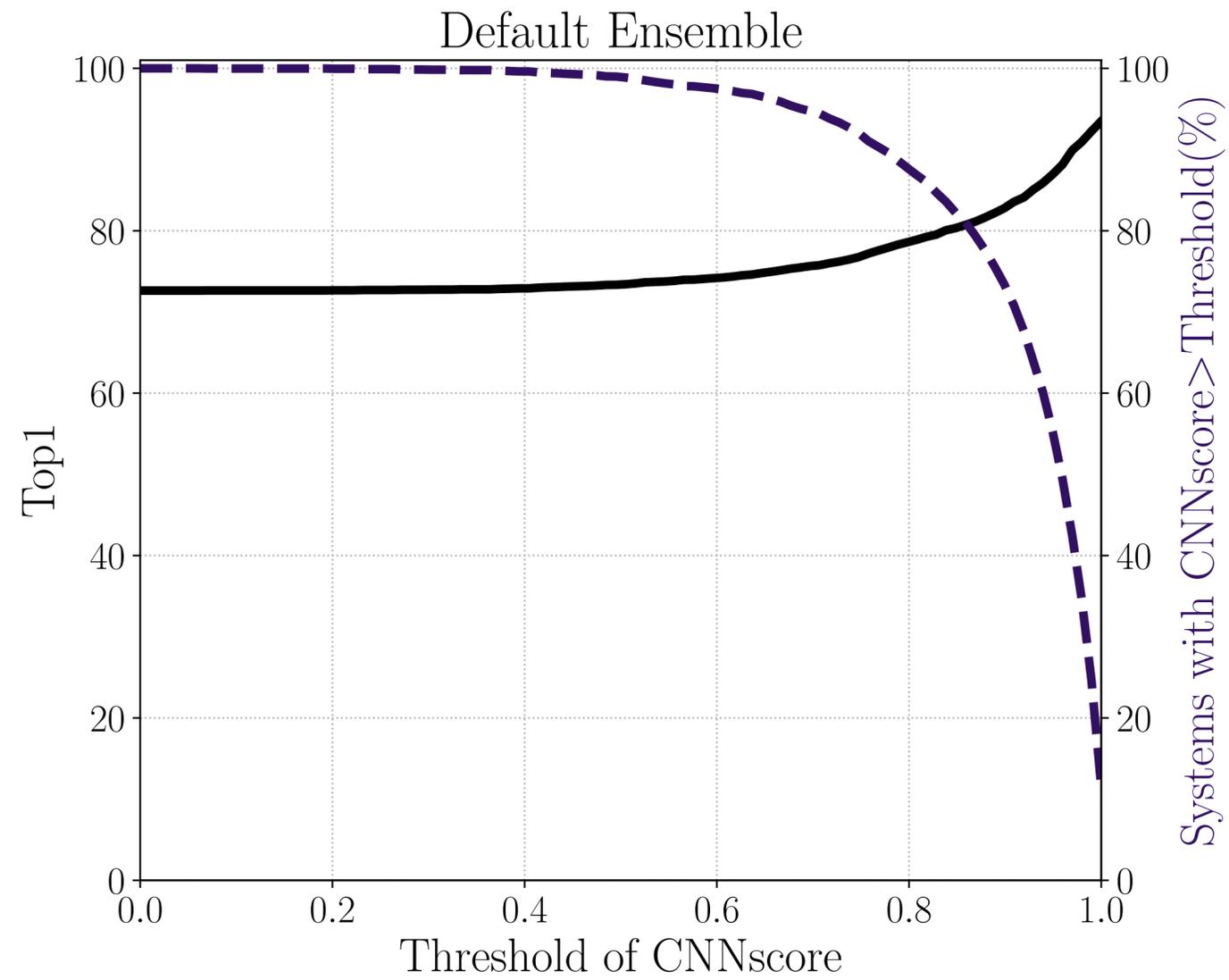
```
gnina -r rec.pdb -l lig.pdb --autobox_ligand lig.pdb -o docked.sdf.gz --flexdist 3.5 --flexdist_ligand lig.pdb
```



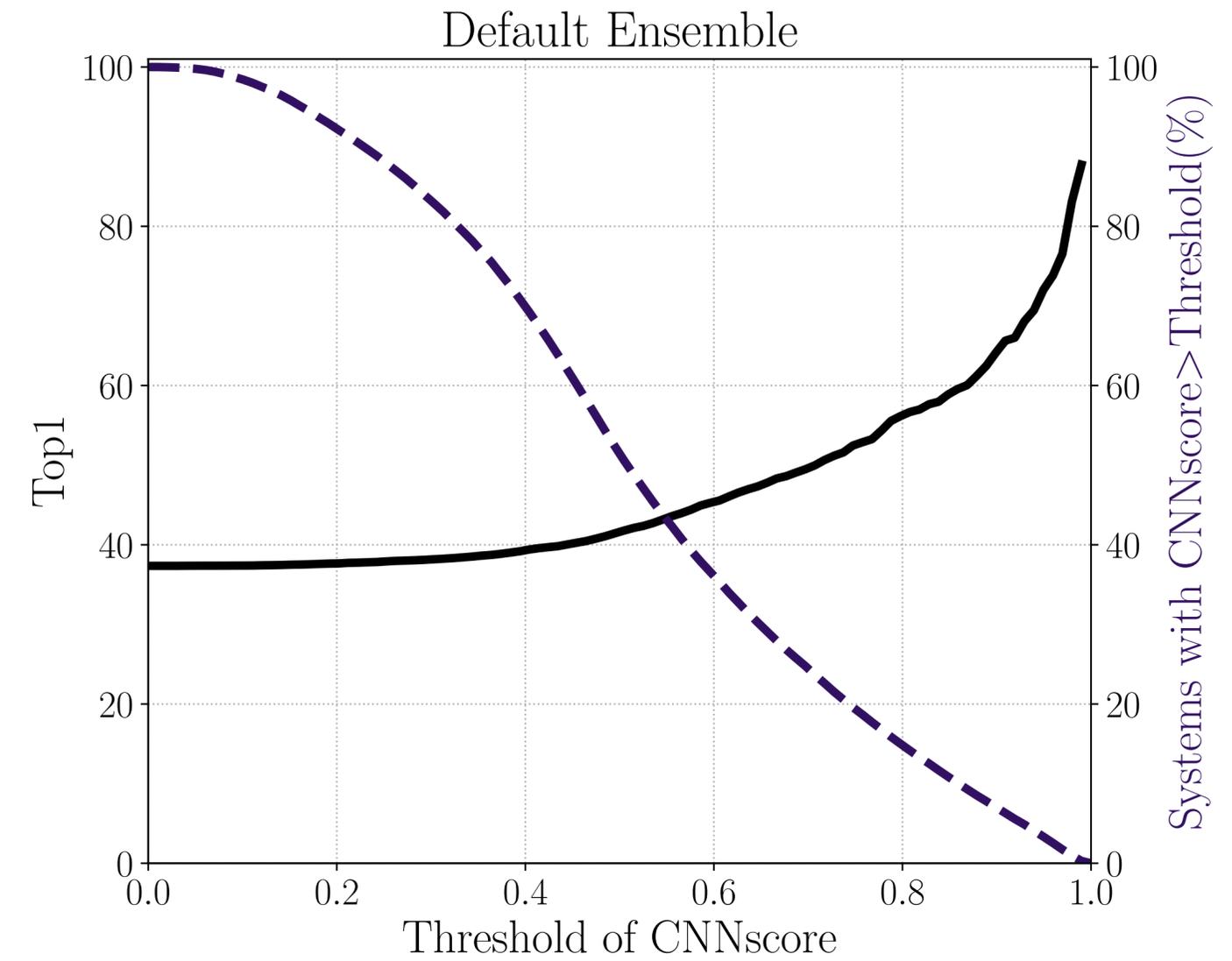
Recommendation: Use flexible docking for specific, intelligently selected residues

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

CNN Scores



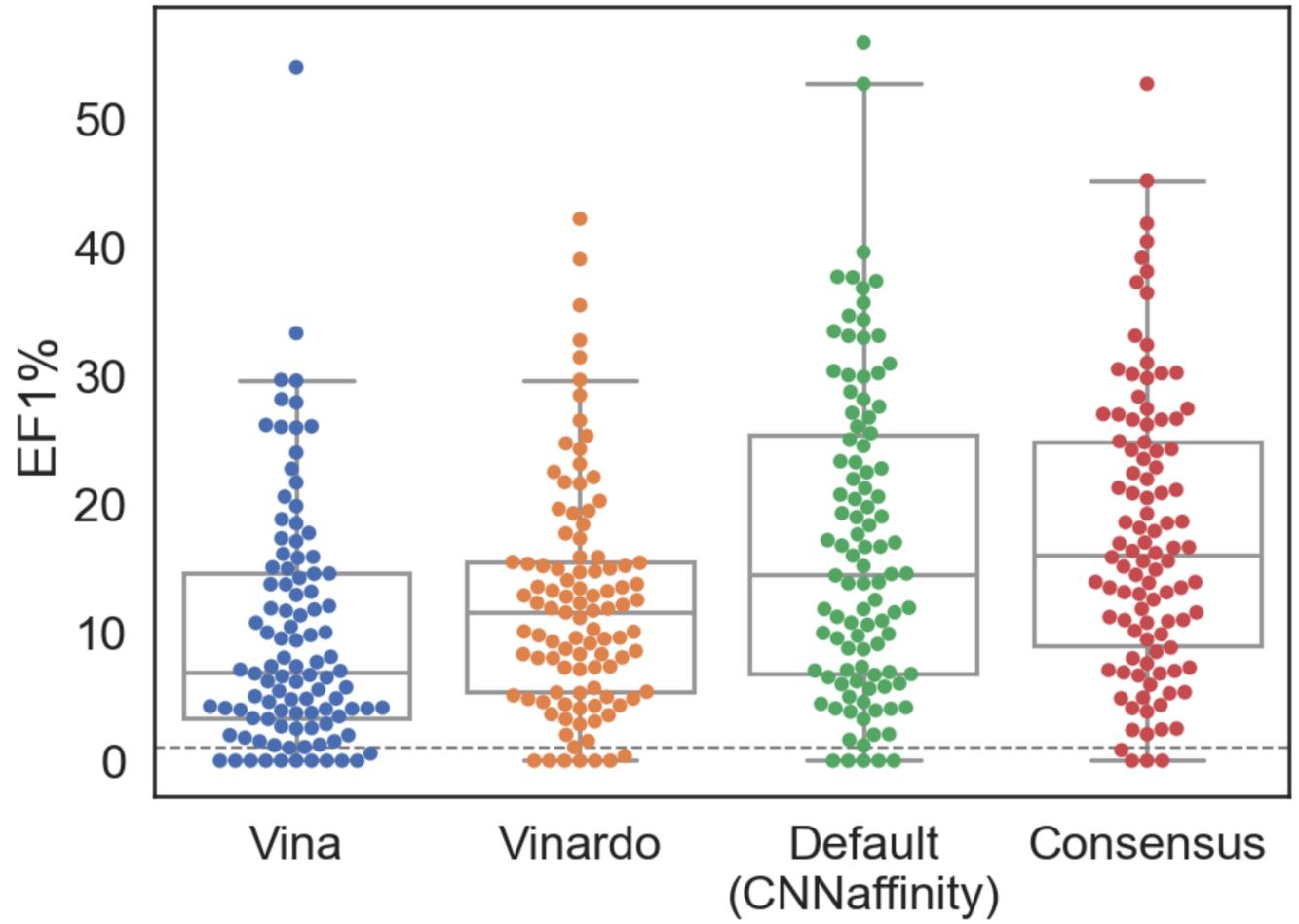
Redocking



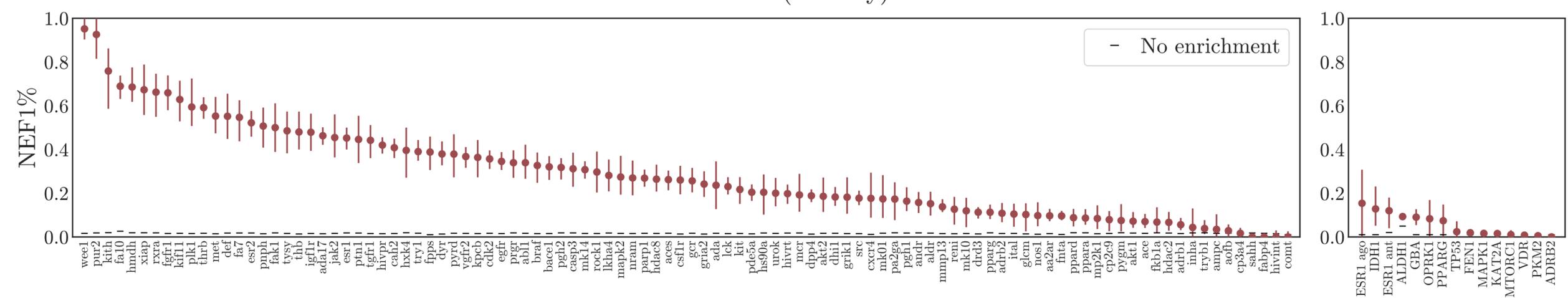
Crossdocking

Virtual Screening

DUD-E Virtual Screening Performance



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...

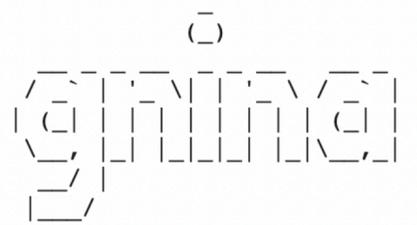
gnina.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ 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 (kcal/mol)	CNN pose score	CNN 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...

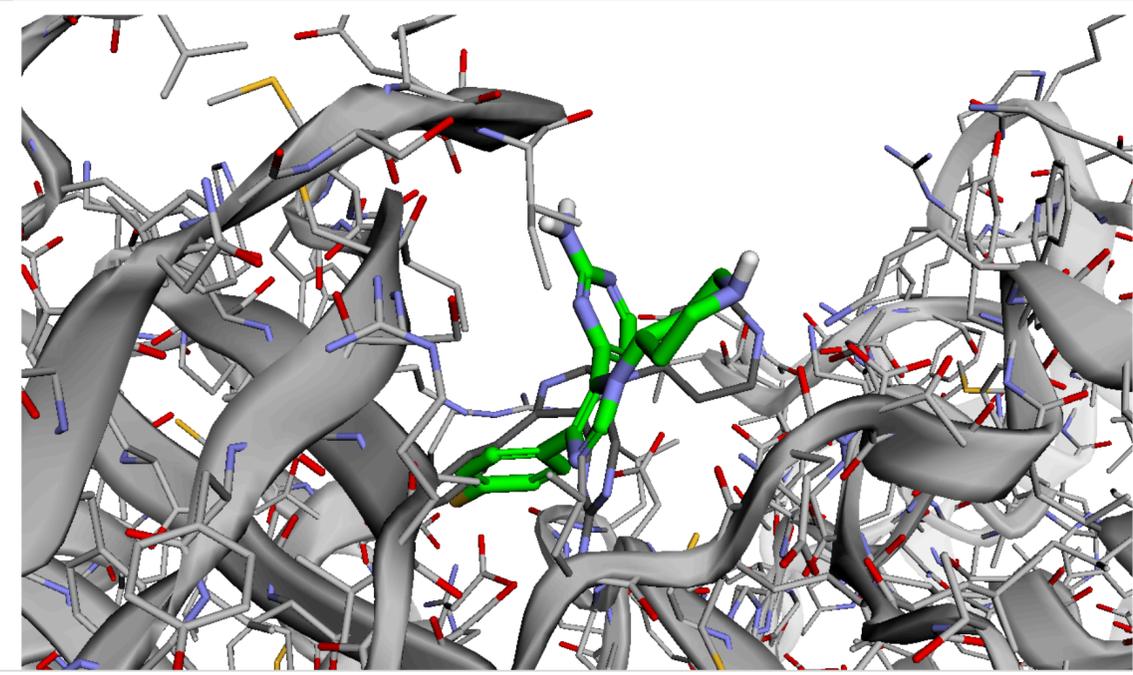
gnina.ipynb

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

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

Rocco Meli
 Rishal Aggarwal

