Deep Learning for Structure-Based Drug Discovery: From Scoring to Generative Design

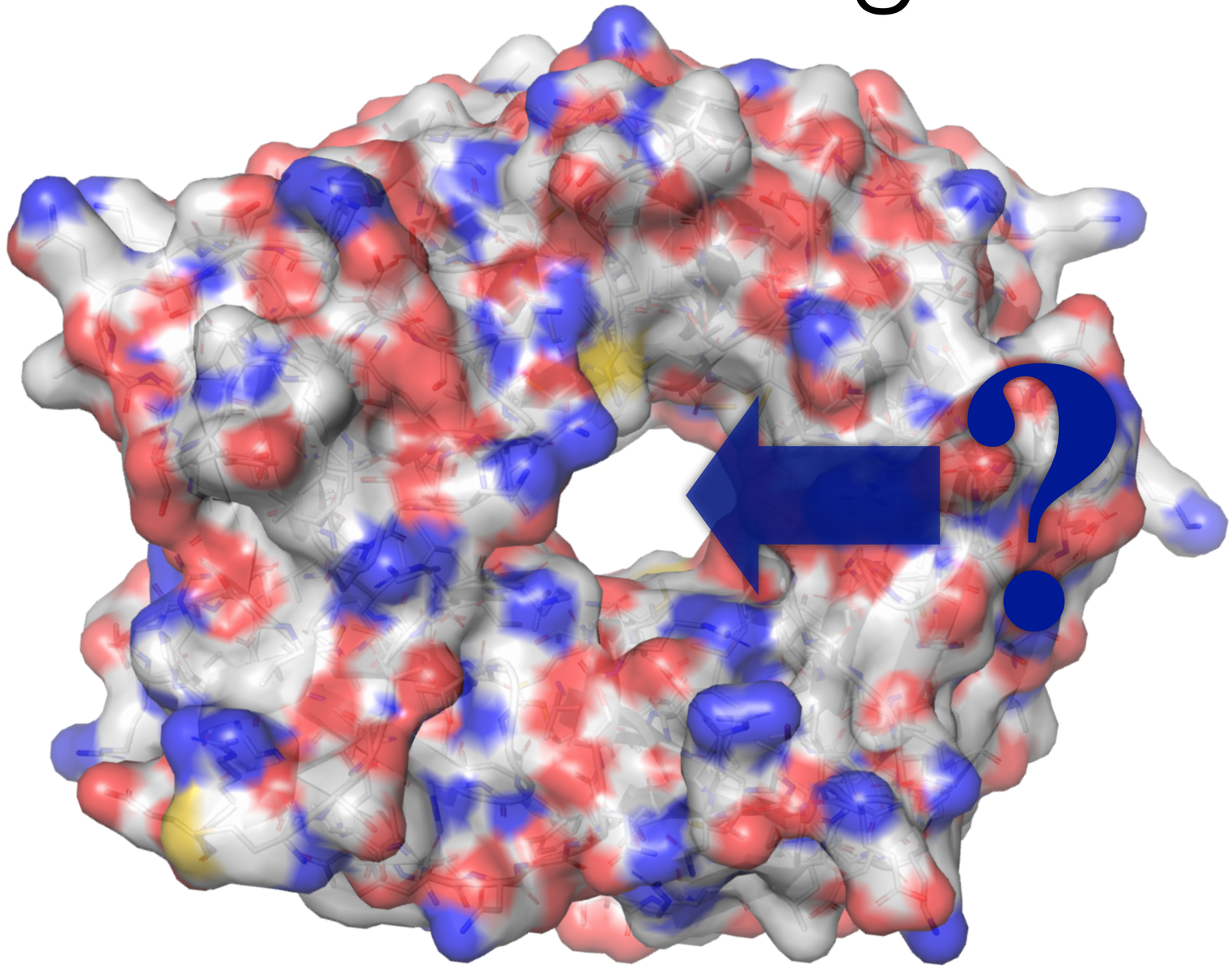
David Ryan Koes

October 17, 2025

AbbVie SMTPT Seminar Series

University of Pittsburgh

Structure Based Drug Discovery



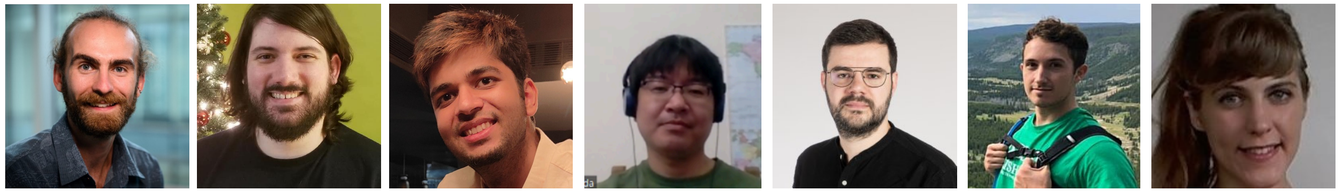
GNINA 1.0

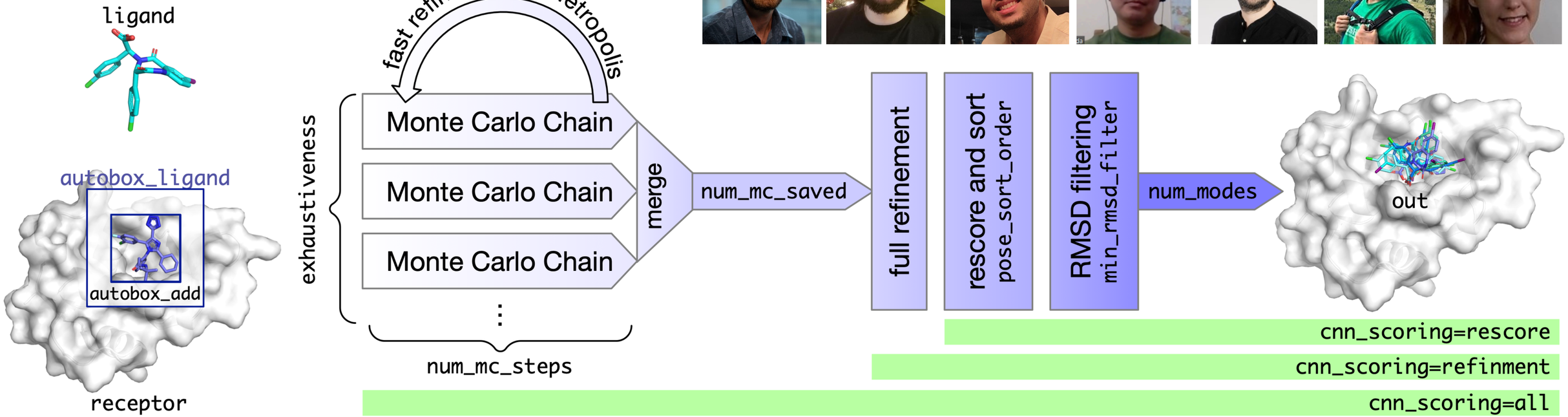
https://github.com/gnina/gnina

SOFTWARE Open Access

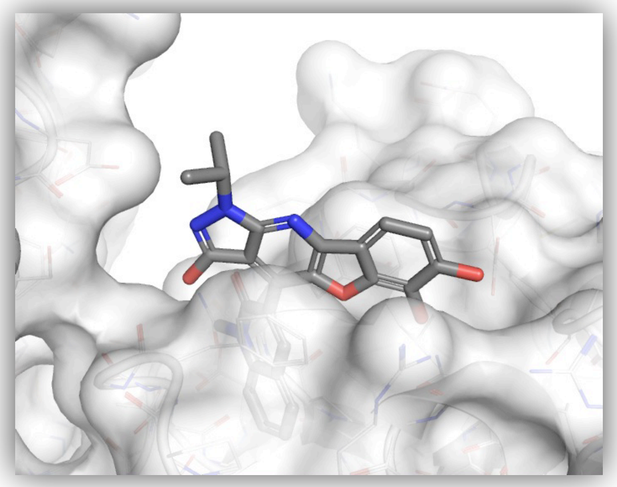
GNINA 1.0: molecular docking with deep learning

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*]





Protein-Ligand Scoring



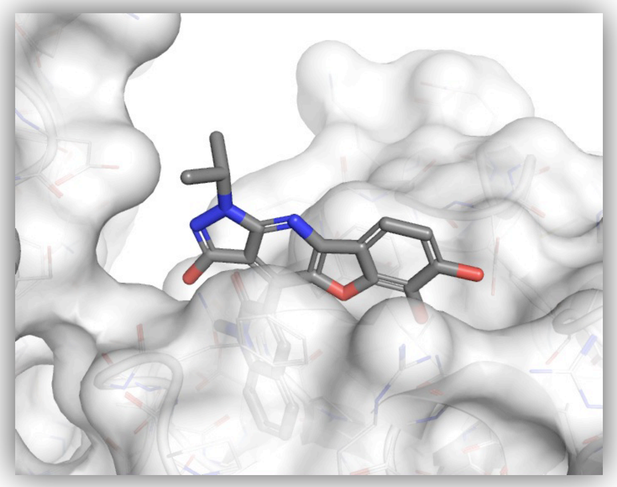
Model

Pose Prediction

Binding Discrimination

Affinity Prediction

Protein-Ligand Scoring



Model

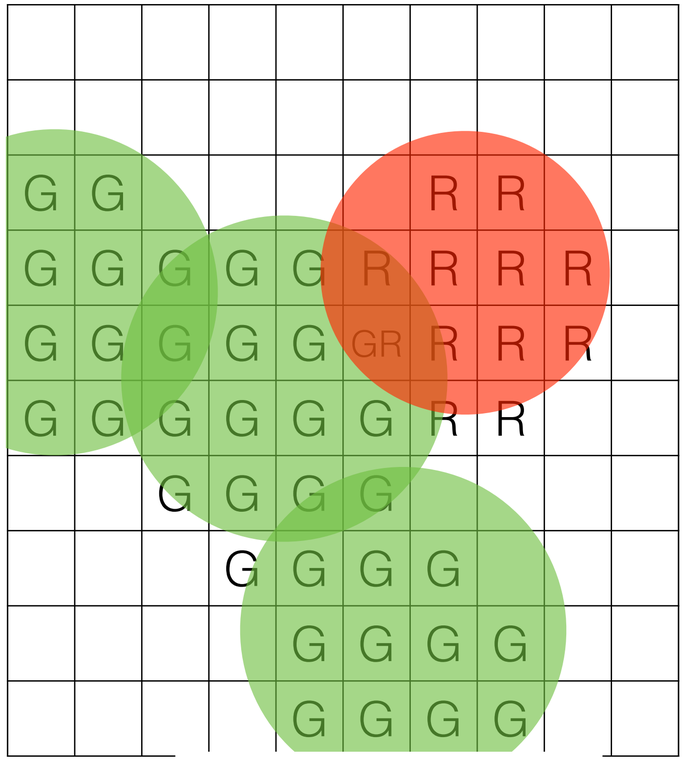
Pose Prediction

Binding Discrimination

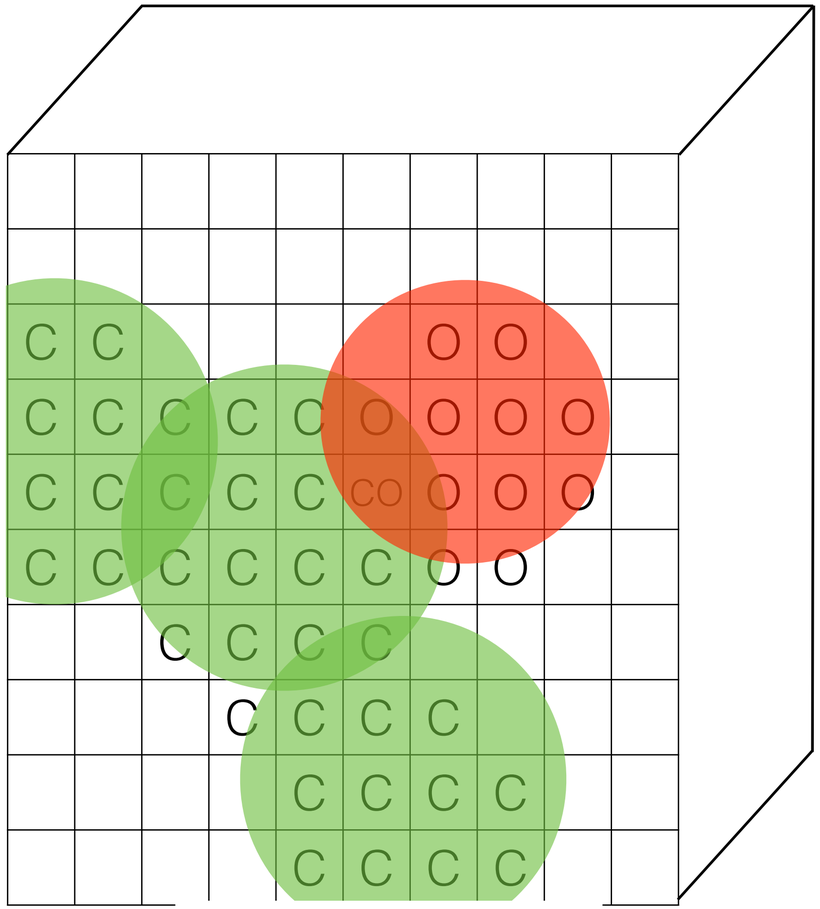
Affinity Prediction

Protein-Ligand Representation

(R,G,B) pixel



Protein-Ligand Representation



(R,G,B) pixel →

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

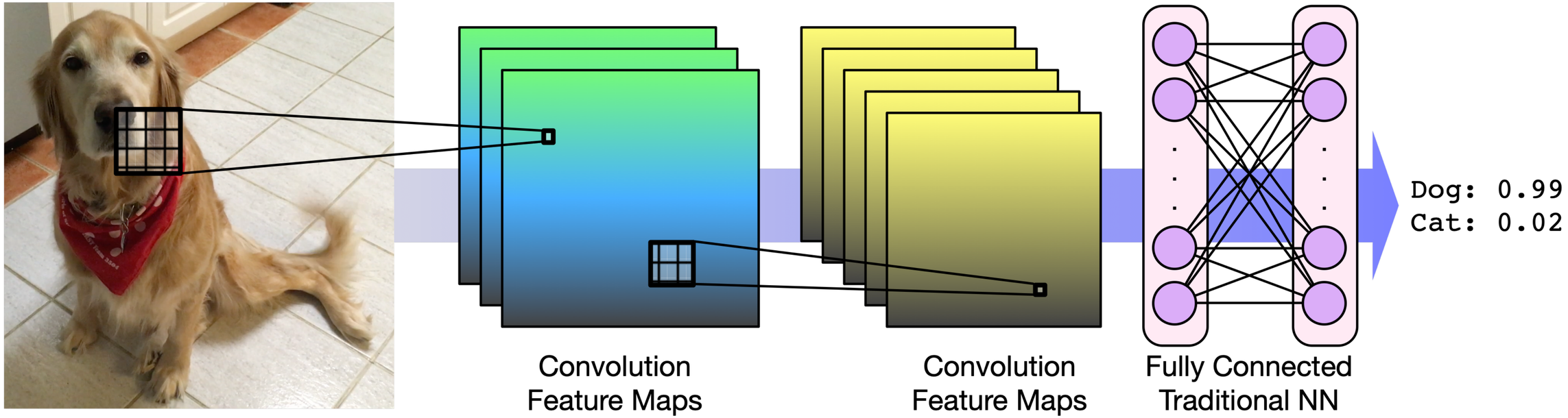
The only parameters for this representation are the choice of **grid resolution**, **atom density**, and **atom types**.

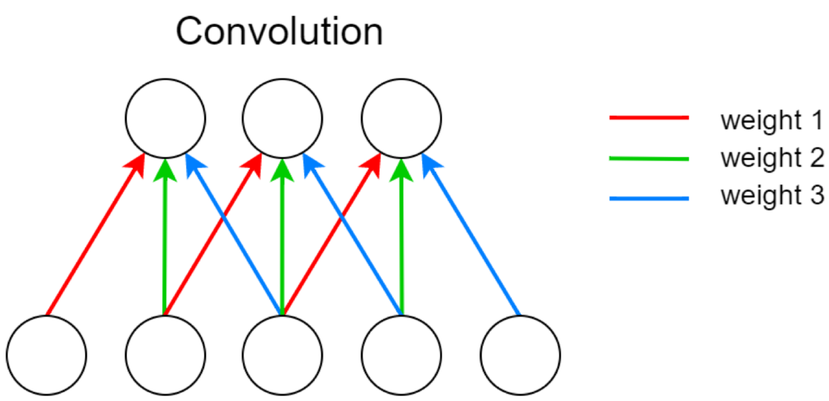
JOURNAL OF CHEMICAL INFORMATION AND MODELING
Article
pubs.acs.org/jcim

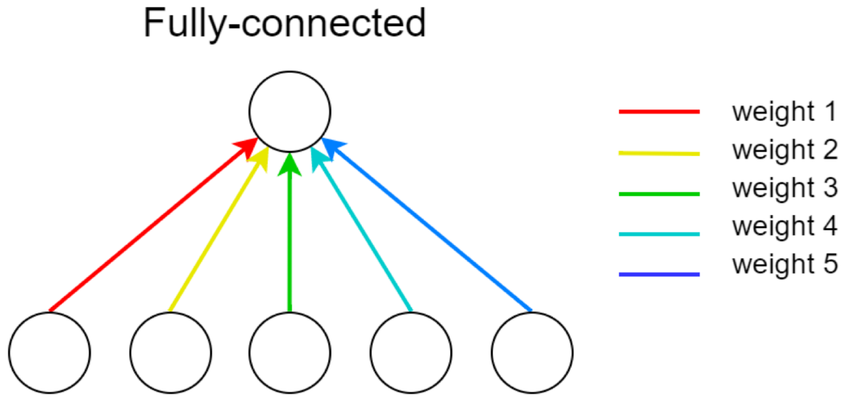
Protein–Ligand Scoring with Convolutional Neural Networks

Matthew Ragoza,[†,‡] Joshua Hochuli,[‡,¶] Elisa Idrobo,[§] Jocelyn Sunseri,[∥] and David Ryan Koes*[,∥]

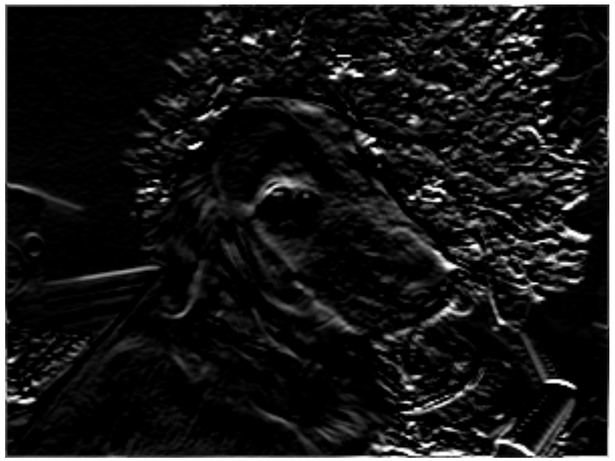
Convolutional Neural Networks







Convolutional Filters

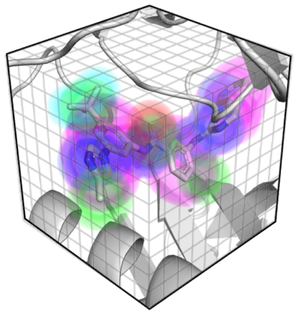


-1	-1	-1
0	0	0
1	1	1

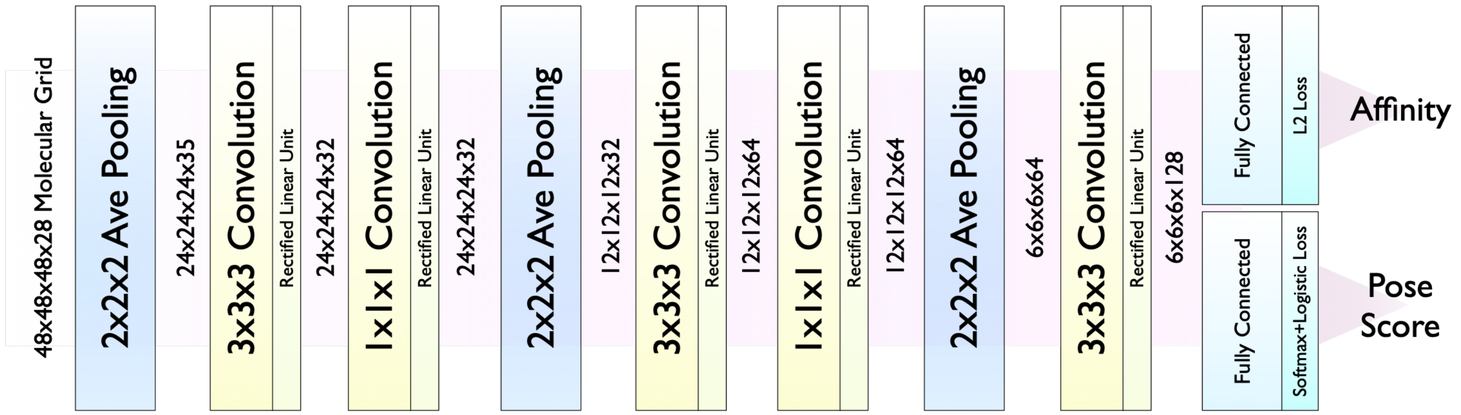
-1	0	1
-1	0	1
-1	0	1

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

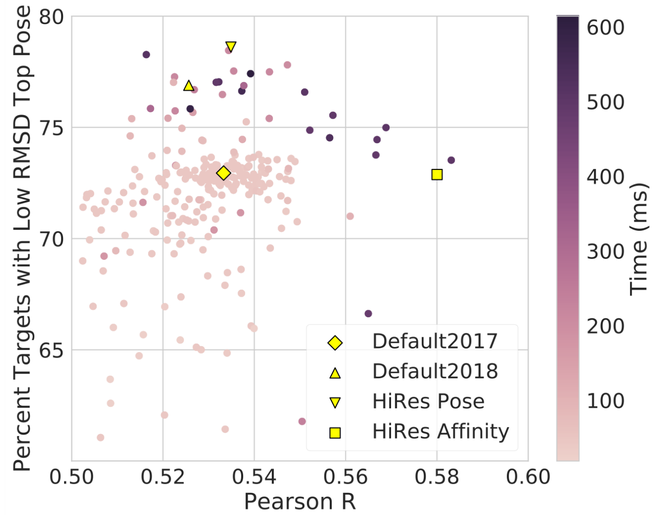
Protein Ligand Scoring



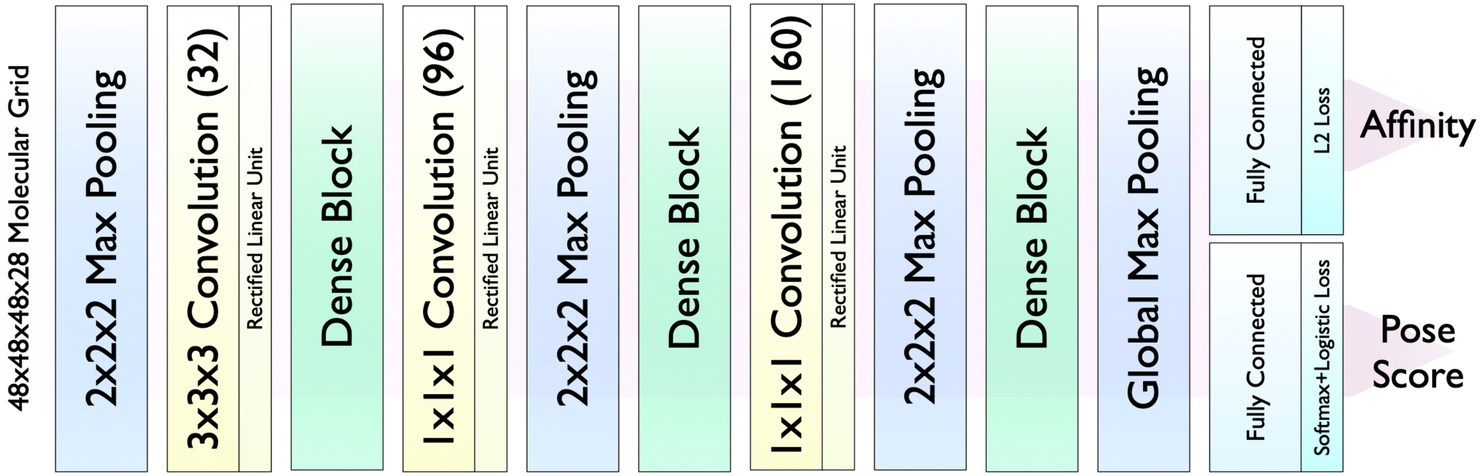
Def2018



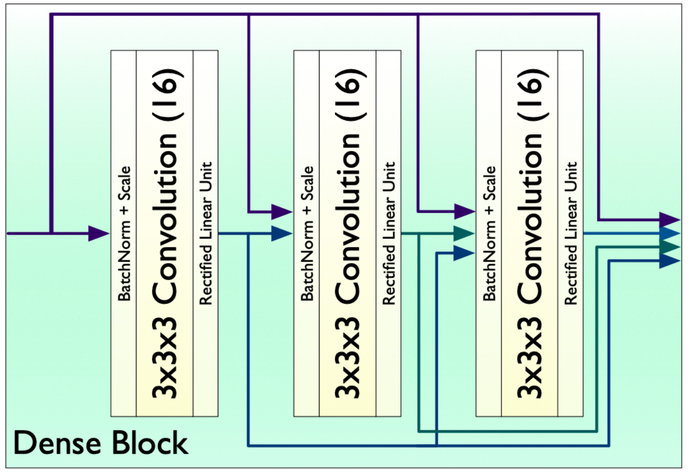
388,736 Parameters



Dense



684,640 Parameters

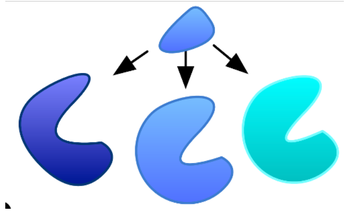


Cross-Docked Protein Ligand Scoring

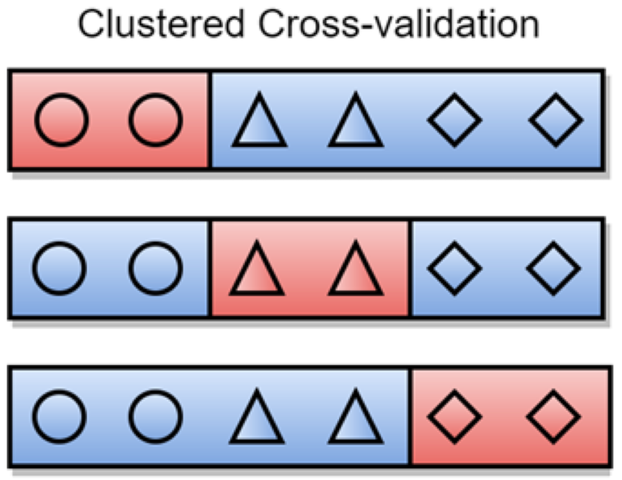
Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design

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

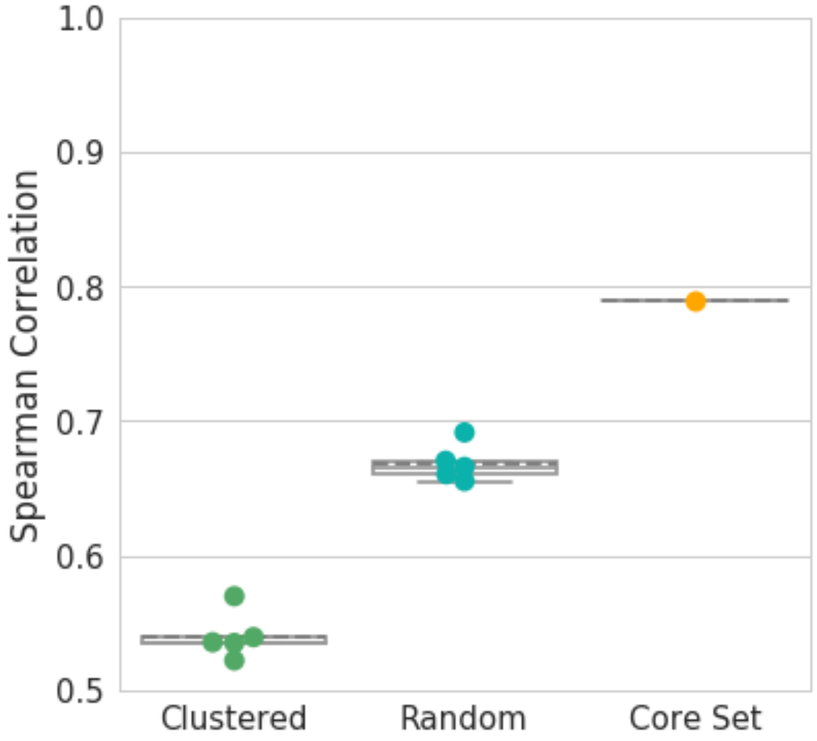
Cite this: *J. Chem. Inf. Model.* 2020, 60, 9, 4200–4215
Publication Date: August 31, 2020
https://doi.org/10.1021/acs.jcim.0c00411

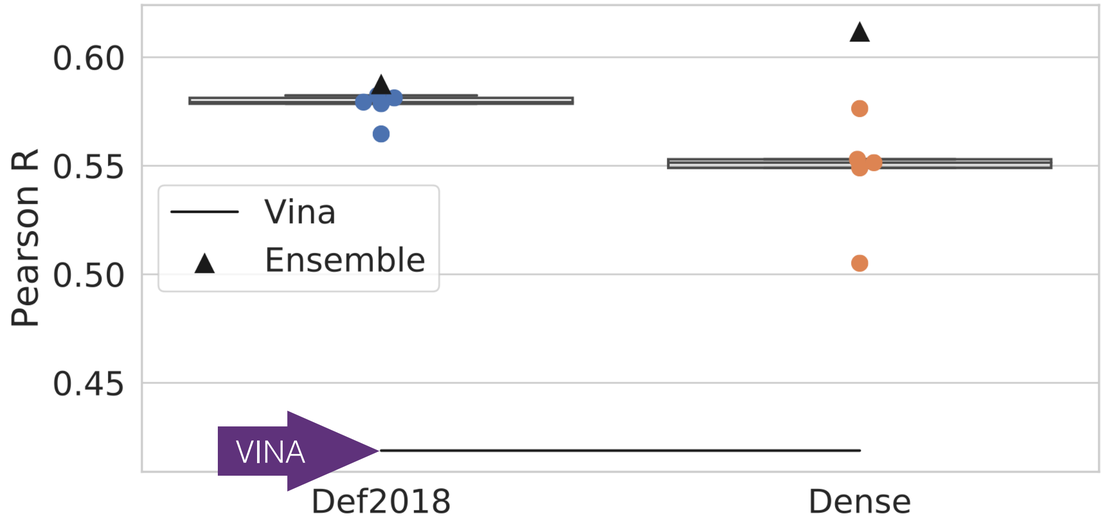


18,450 complexes
22.6 million poses

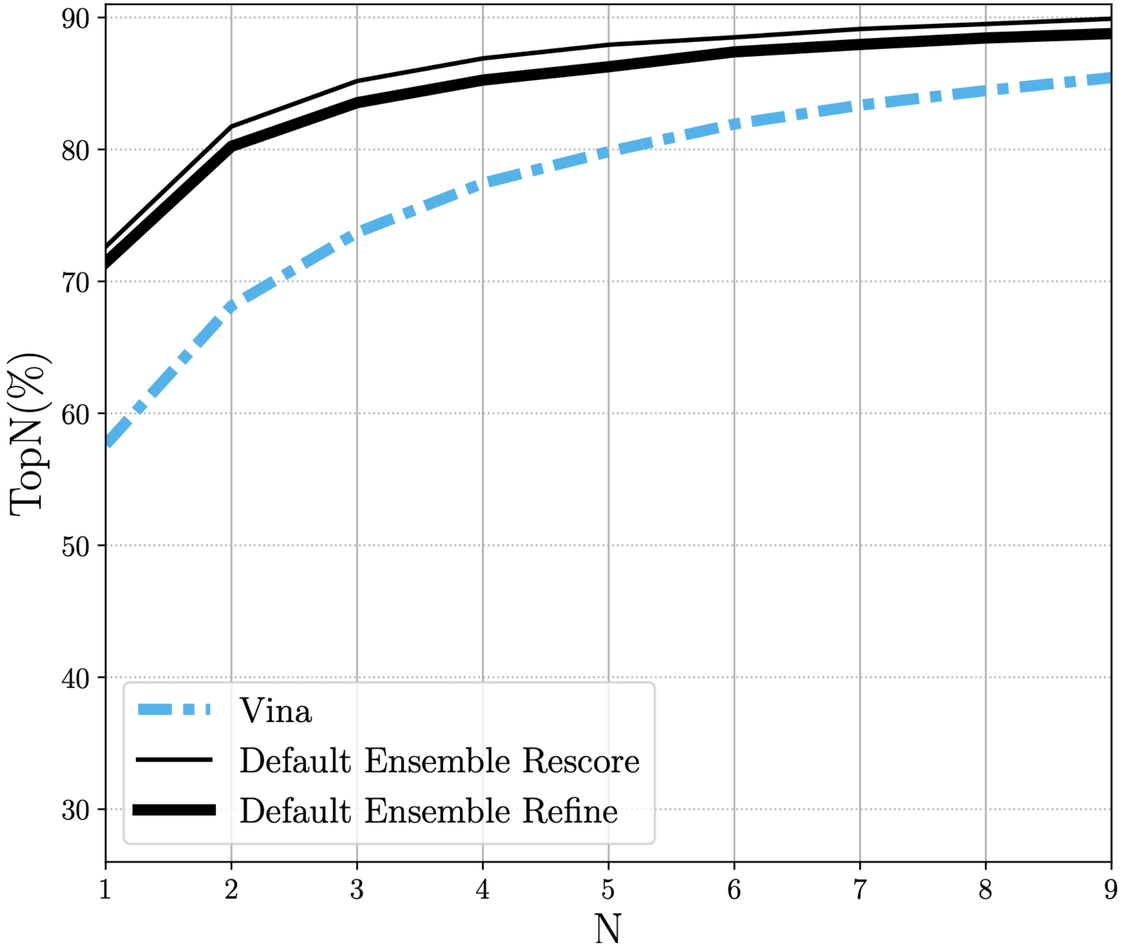


Affinity Prediction

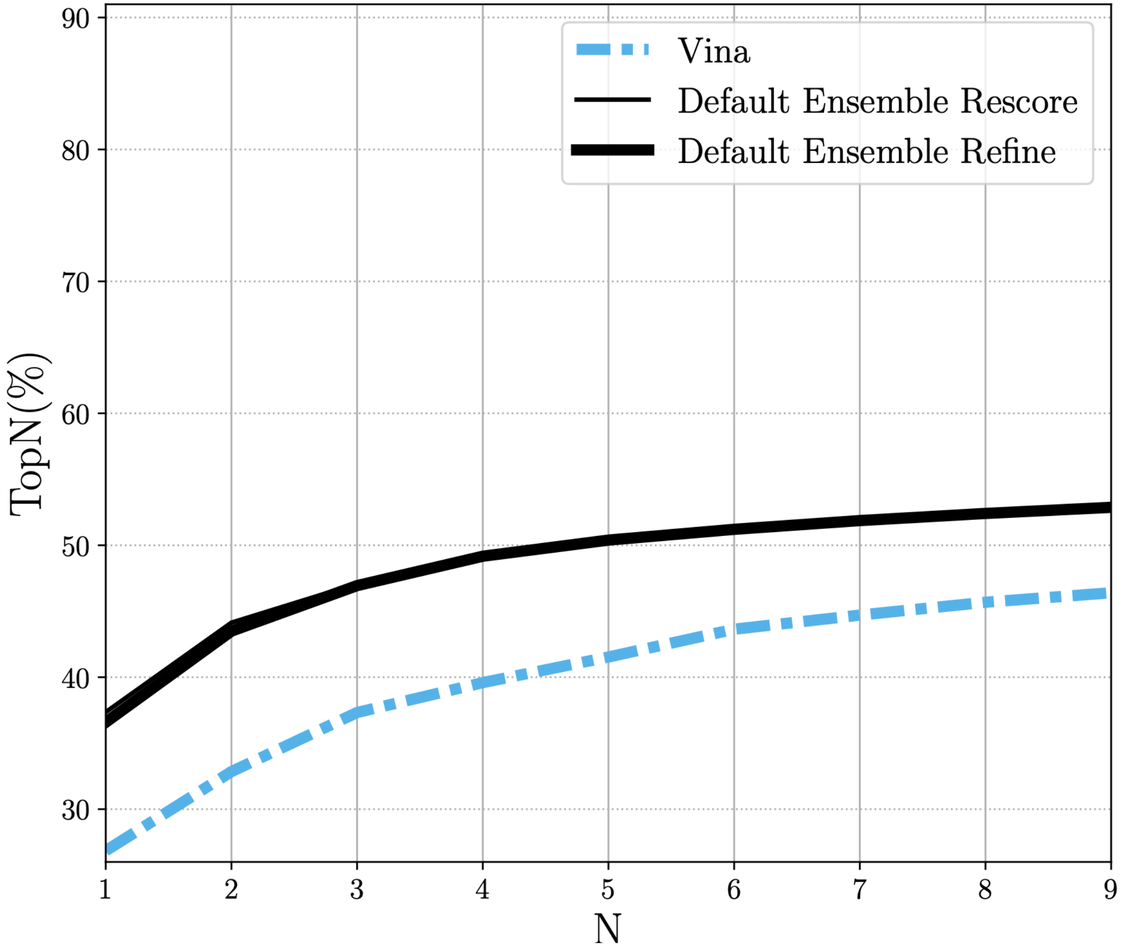




Docking Performance



Redocking



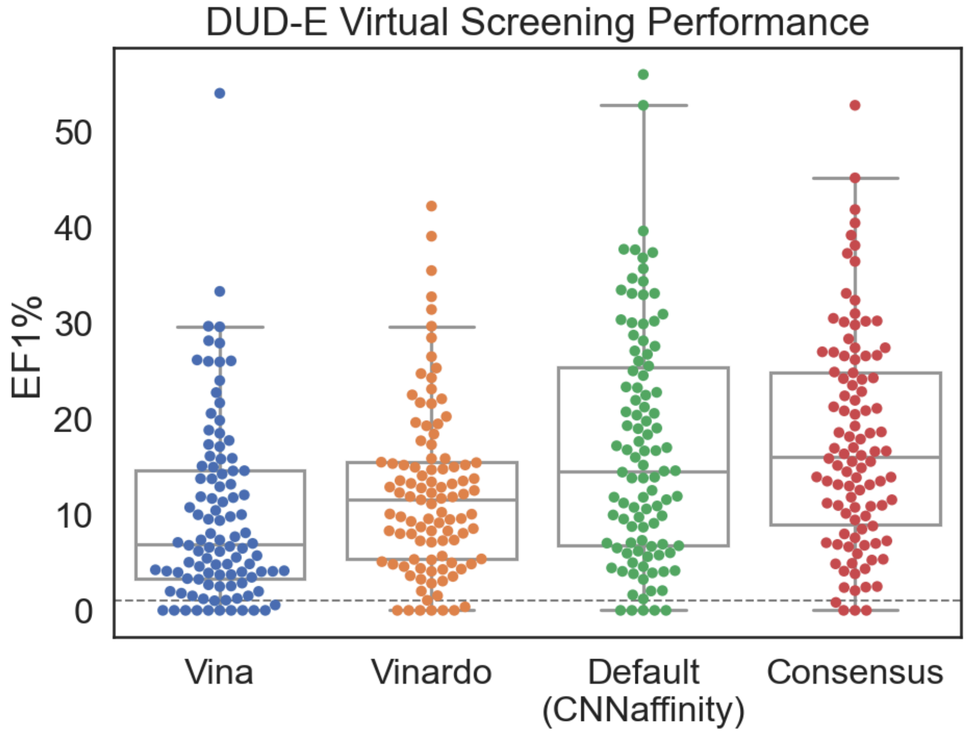
Crossdocking

molecules

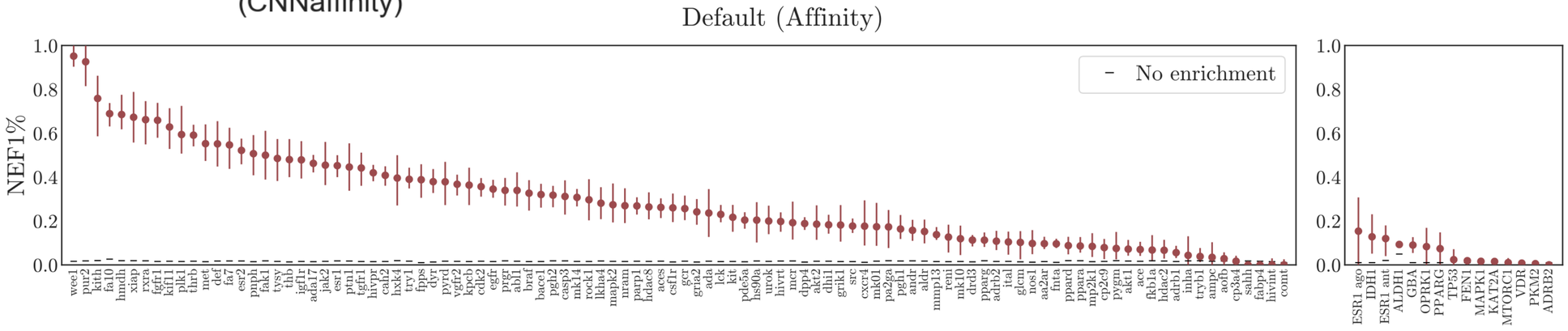
Open Access Article

Virtual Screening with GNINA 1.0

by Jocelyn Sunseri and David Ryan Koes *



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*



GNINA 1.3

https://github.com/gnina/gnina

GNINA 1.3: the next increment in molecular docking with deep learning

Andrew T. McNutt, Yanjing Li, Rocco Meli, Rishal Aggarwal & David Ryan Koes

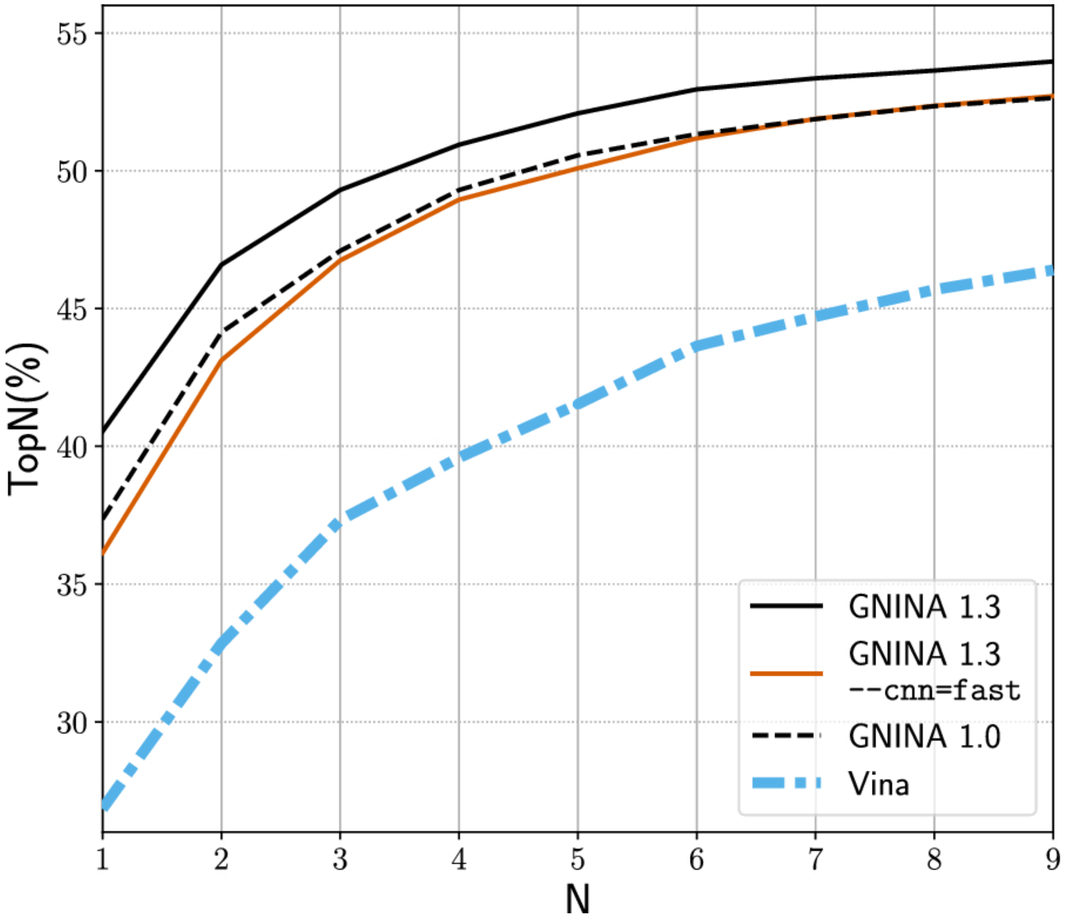
Journal of Cheminformatics **17**, Article number: 28 (2025) | Cite this article

Caffe → Torch

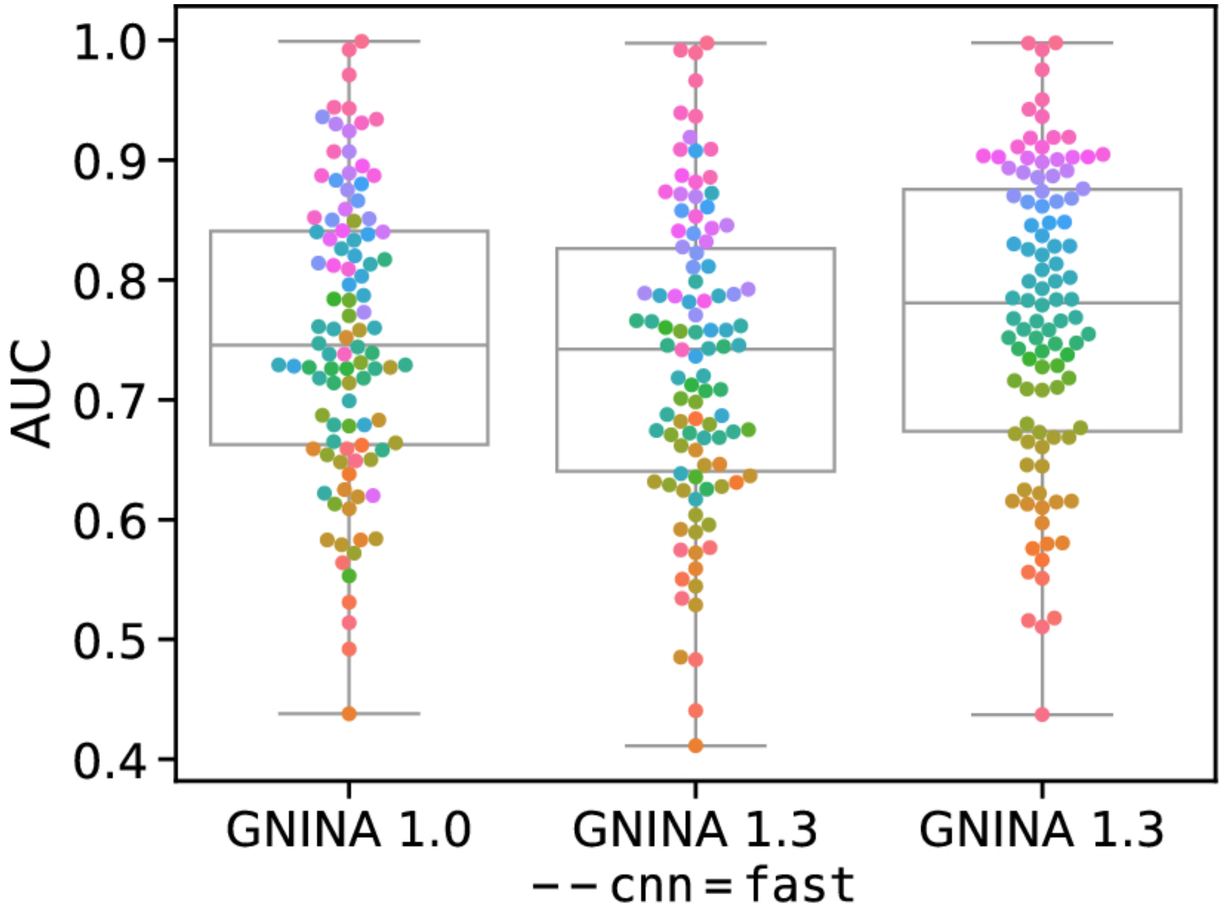
easy covalent docking

retrained models

GNINA 1.3 Performance

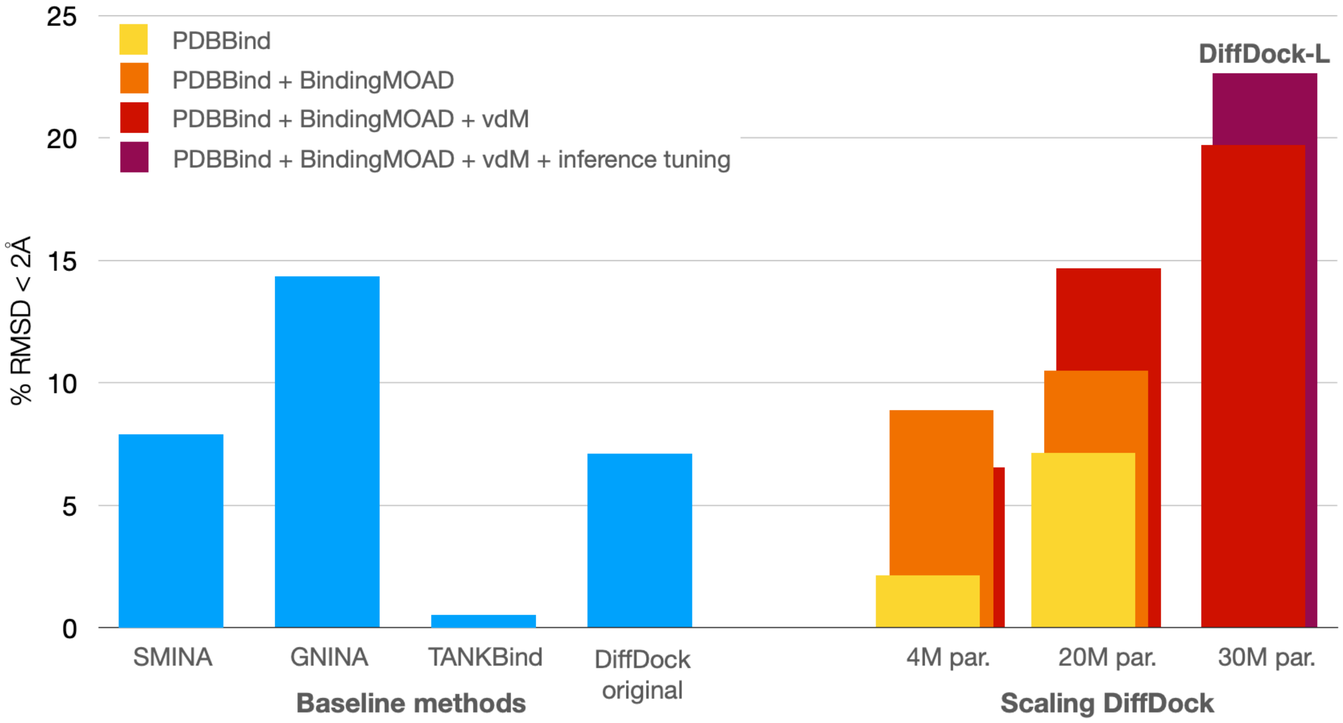


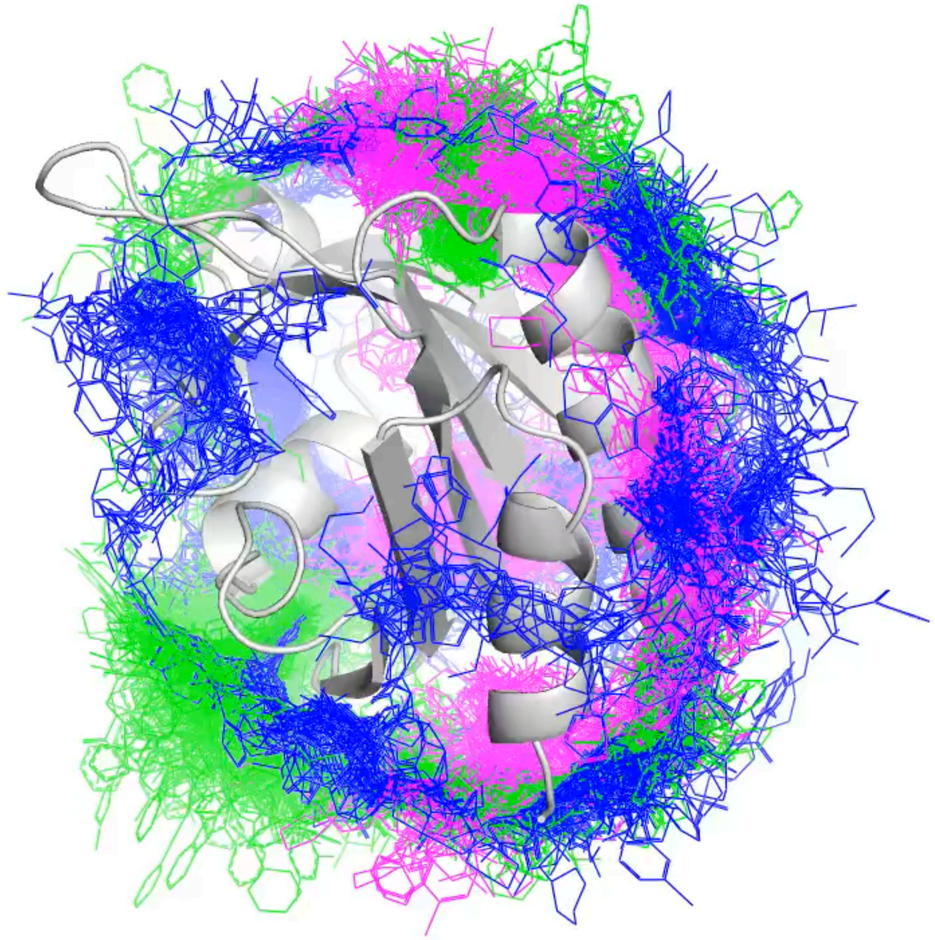
Crossdocking



Virtual Screening (DUD-E)

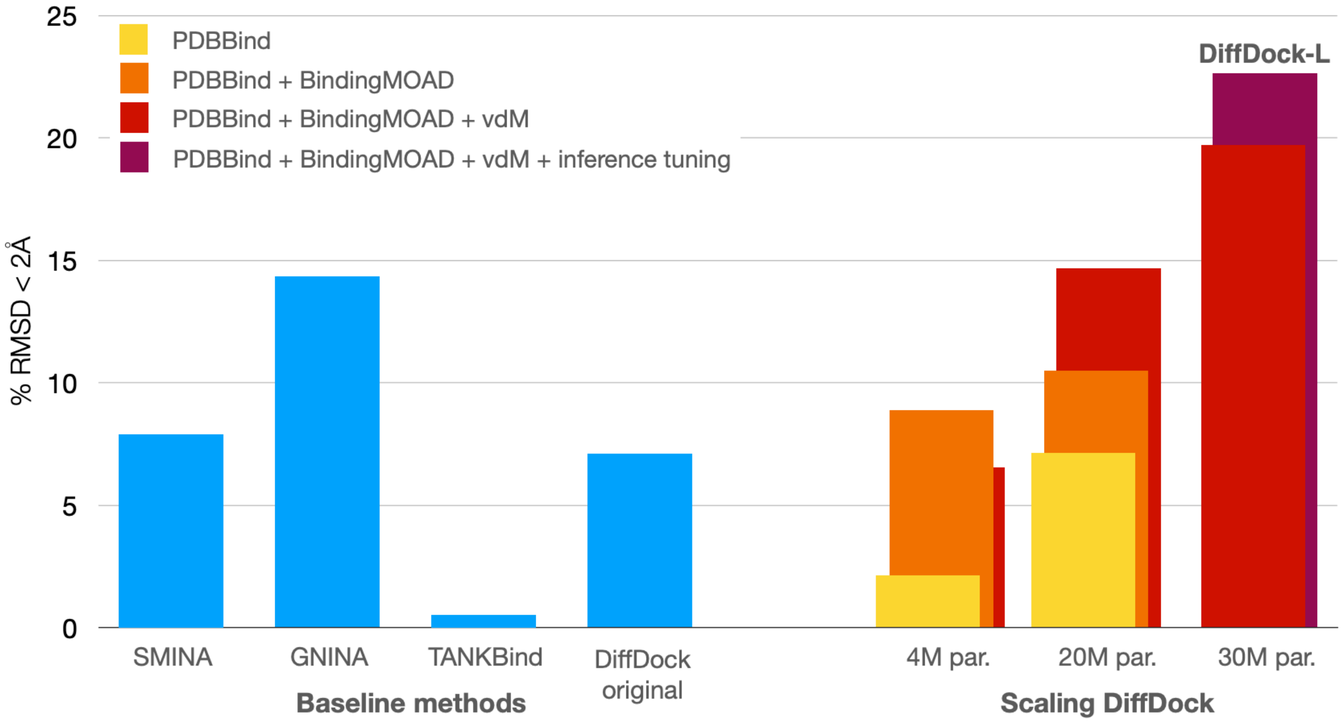
GNINA vs End-to-end Deep Docking

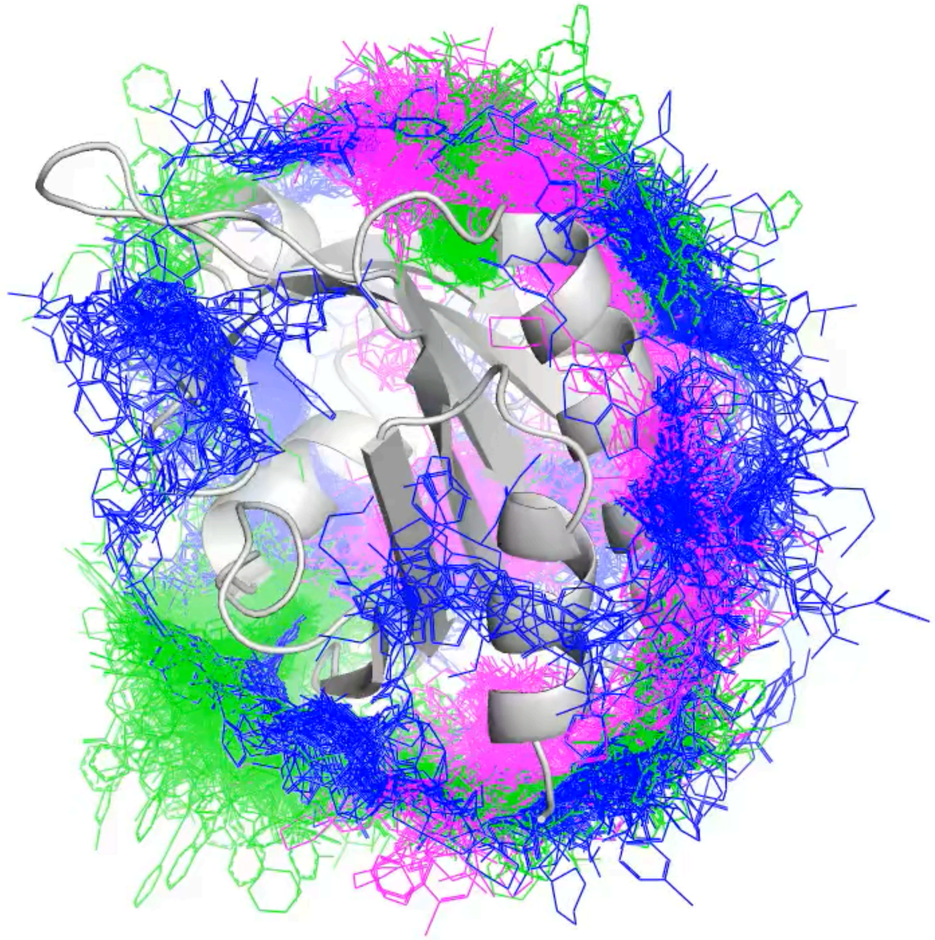




GNINA DynamicBind Boltz1

GNINA vs End-to-end Deep Docking

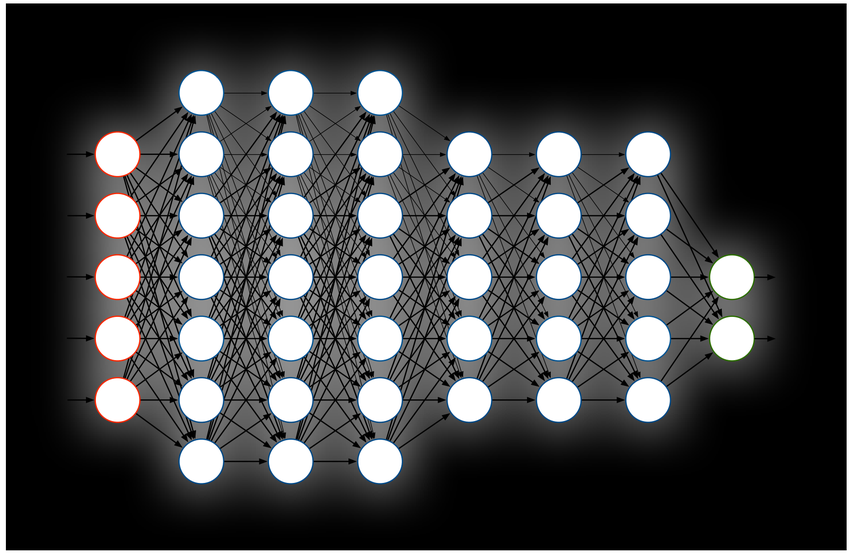




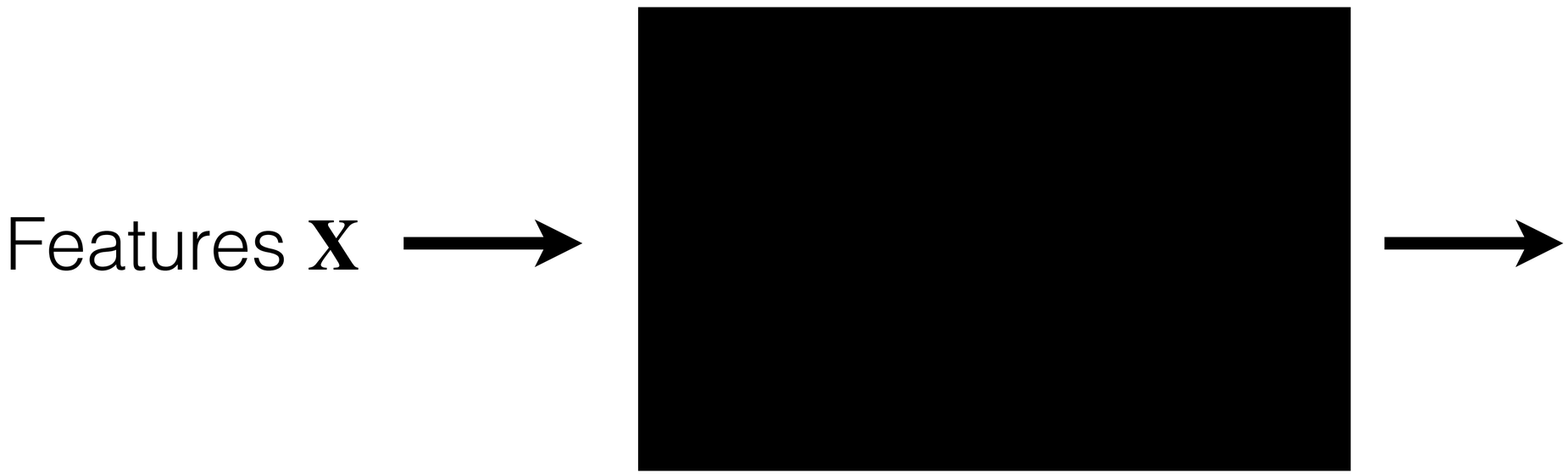
GNINA DynamicBind Boltz1

Generative Modeling

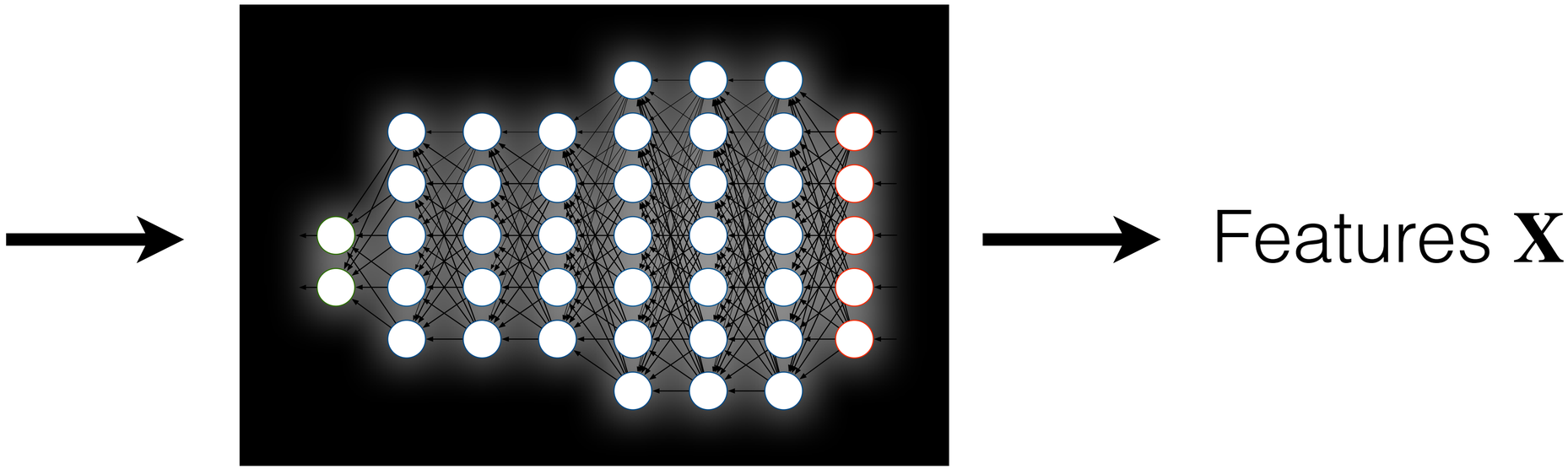
Discriminative Model

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

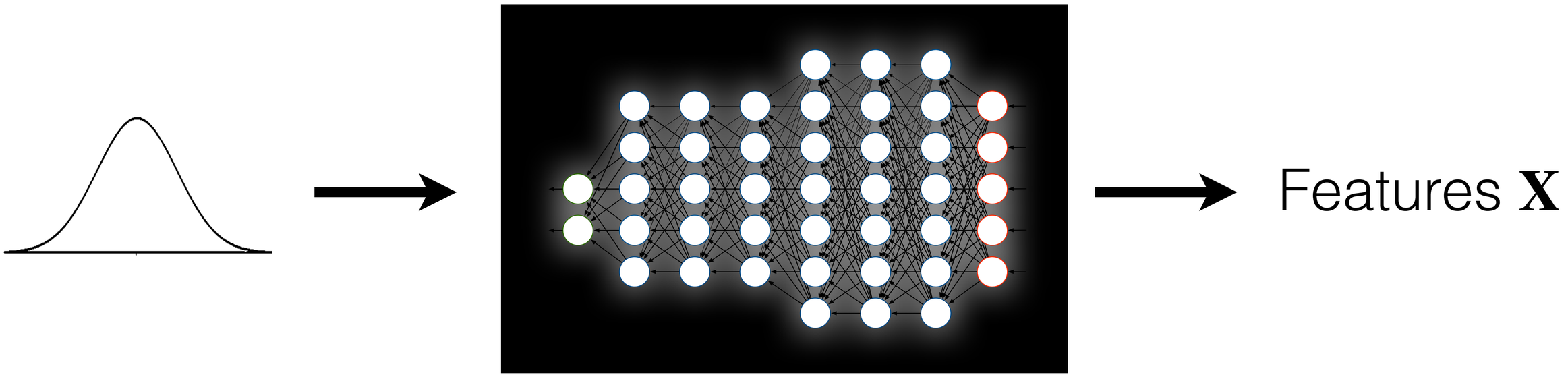
Generative Model



Generative Model



Generative Model



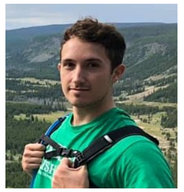
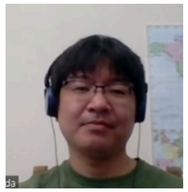
Learning a Continuous Representation of 3D Molecular Structures with Deep Generative Models

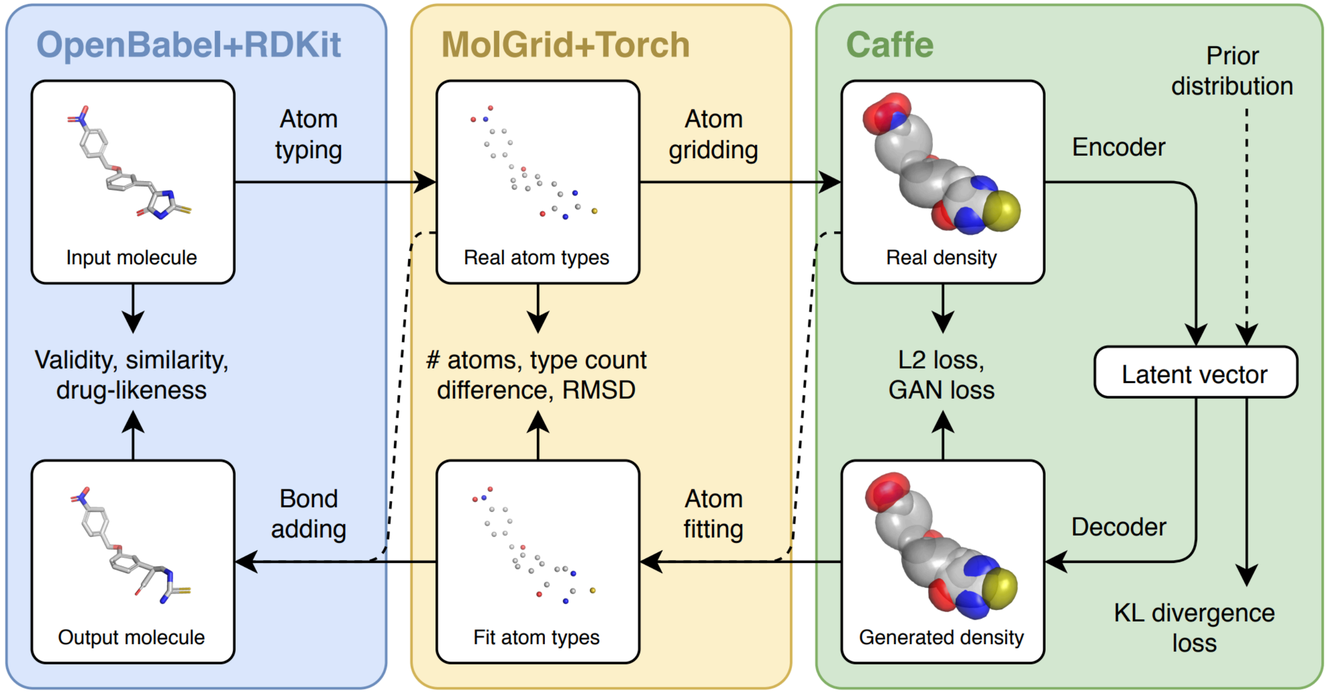
NeurIPS 2020 Workshop
Machine Learning for Structural Biology

Matthew Ragoza*
Comp. & Systems Biology
University of Pittsburgh
Pittsburgh, PA 15213
mtr22@pitt.edu

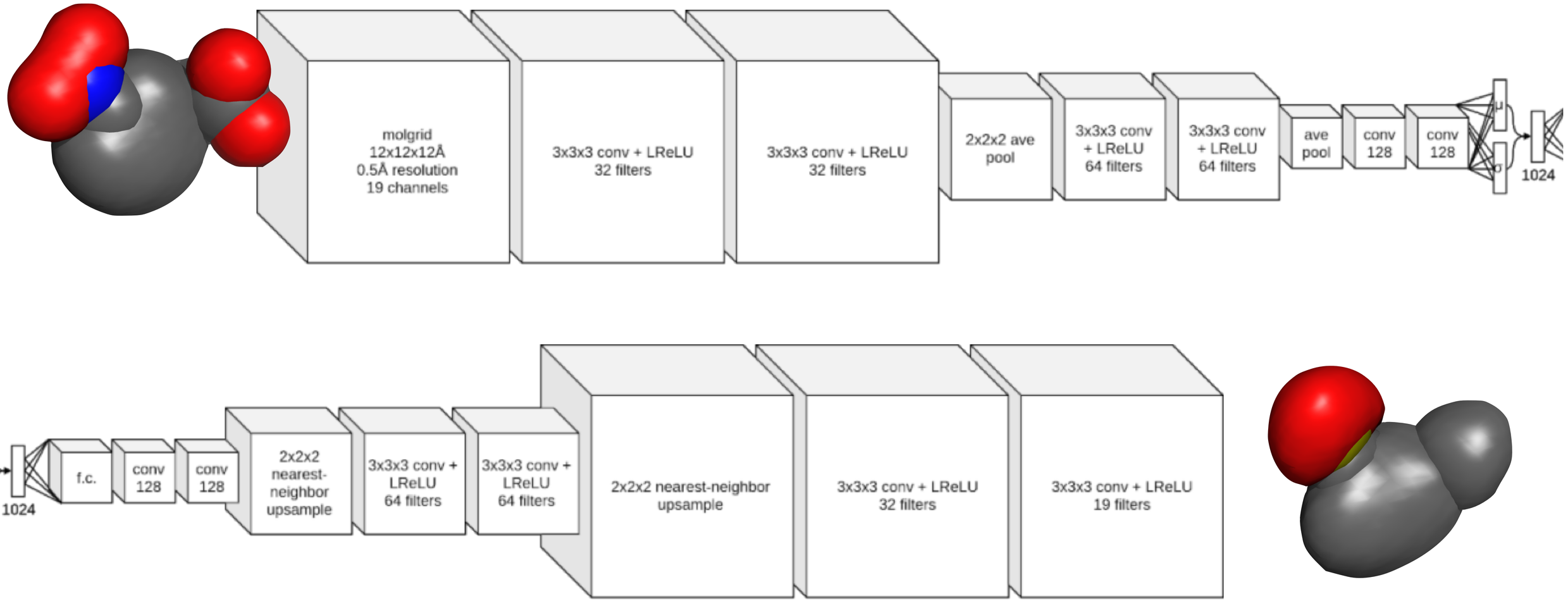
Tomohide Masuda*
Comp. & Systems Biology
University of Pittsburgh
Pittsburgh, PA 15213
tmasuda@pitt.edu

David Ryan Koes
Comp. & Systems Biology
University of Pittsburgh
Pittsburgh, PA 15213
dkoes@pitt.edu

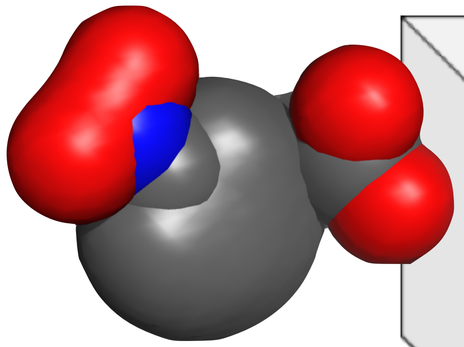


Model


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
μ
σ
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

Model

Variational
Latent Space



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

μ

σ

1024

1024

f.c.

conv
128

conv
128

2x2x2
nearest-
neighbor
upsample

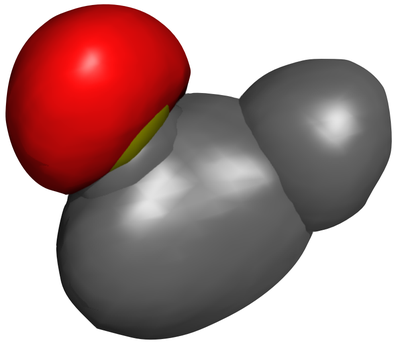
3x3x3 conv +
LReLU
64 filters

3x3x3 conv +
LReLU
64 filters

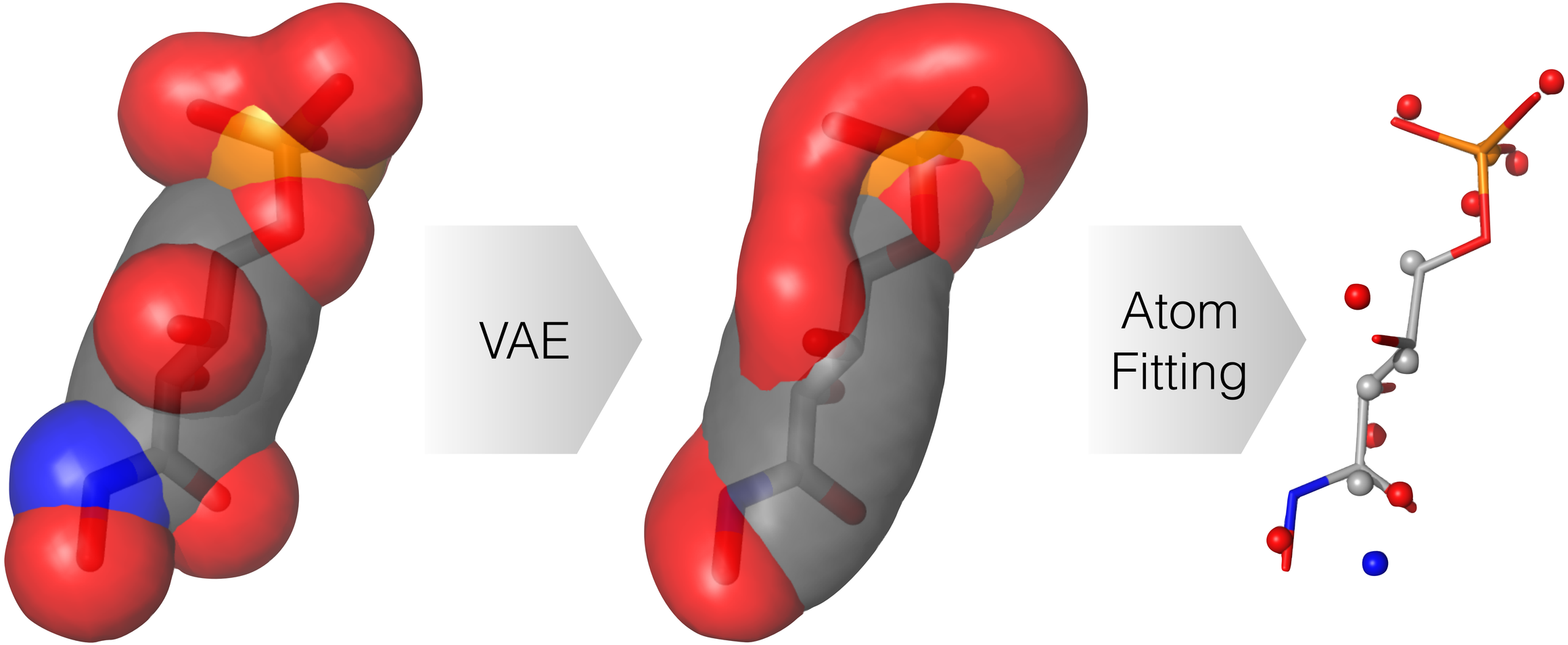
2x2x2 nearest-neighbor
upsample

3x3x3 conv + LReLU
32 filters

3x3x3 conv + LReLU
19 filters



Variational Autoencoding Examples



2BES

Different input molecules

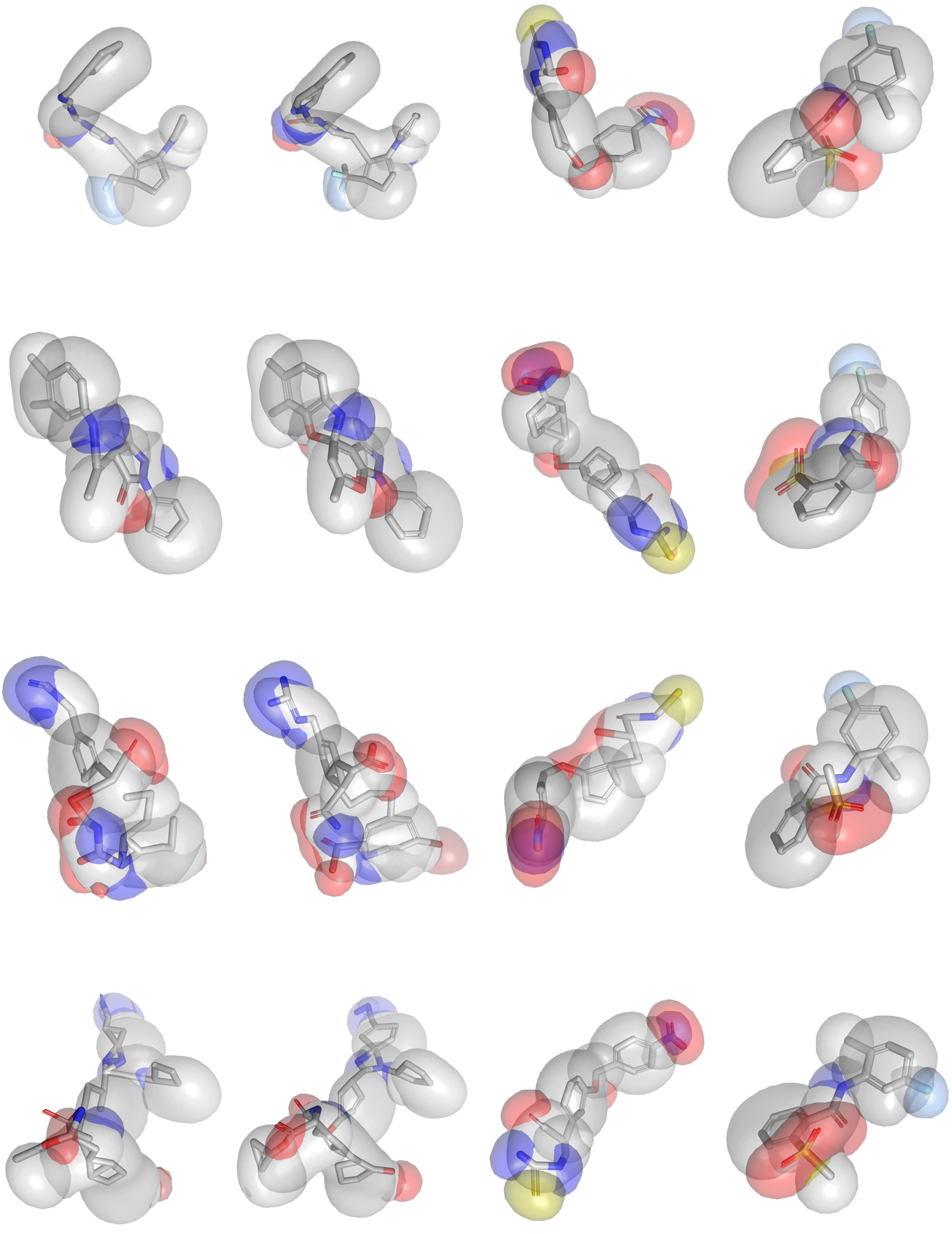
Same input molecules

VAE

AE

VAE

AE



Different input molecules

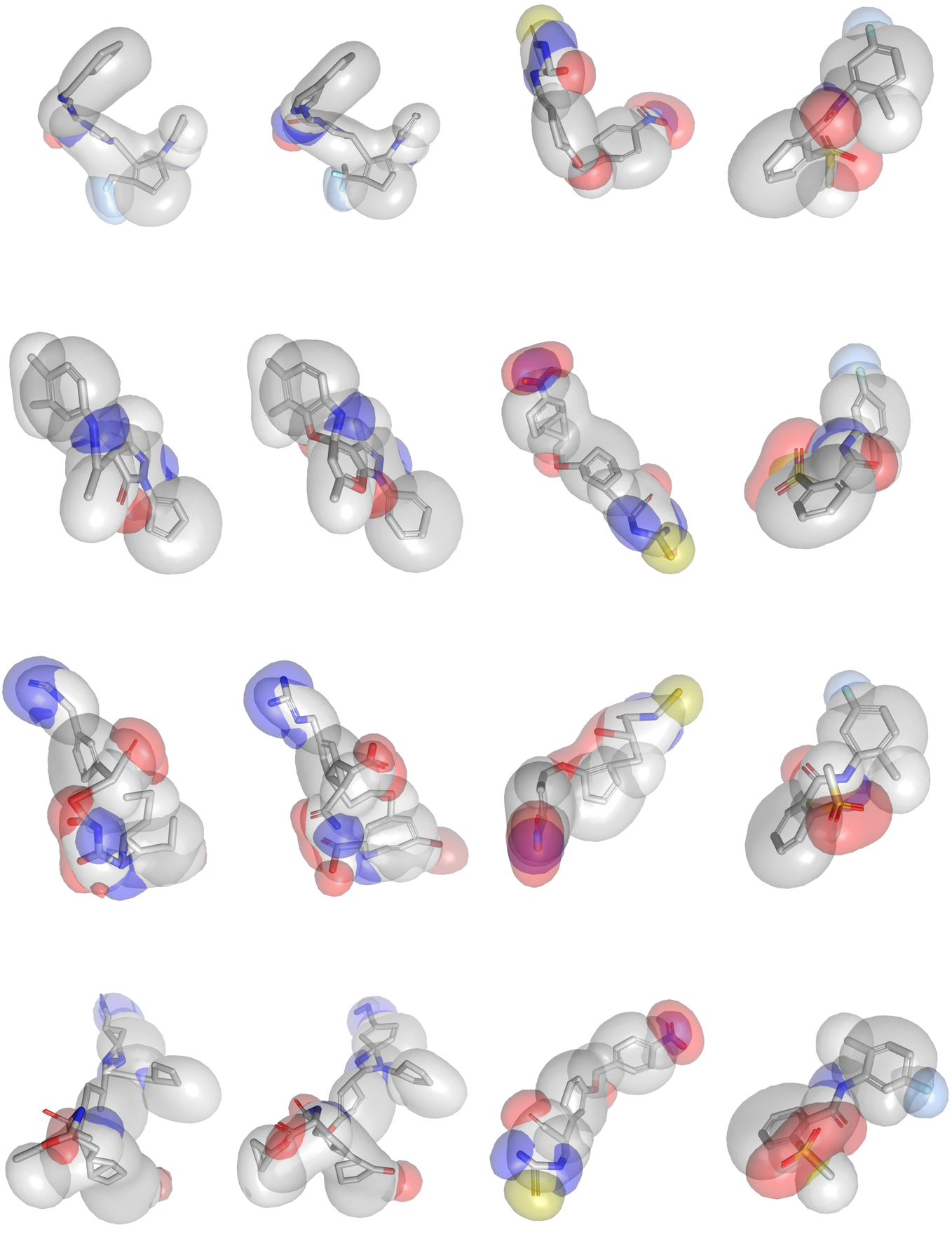
Same input molecules

VAE

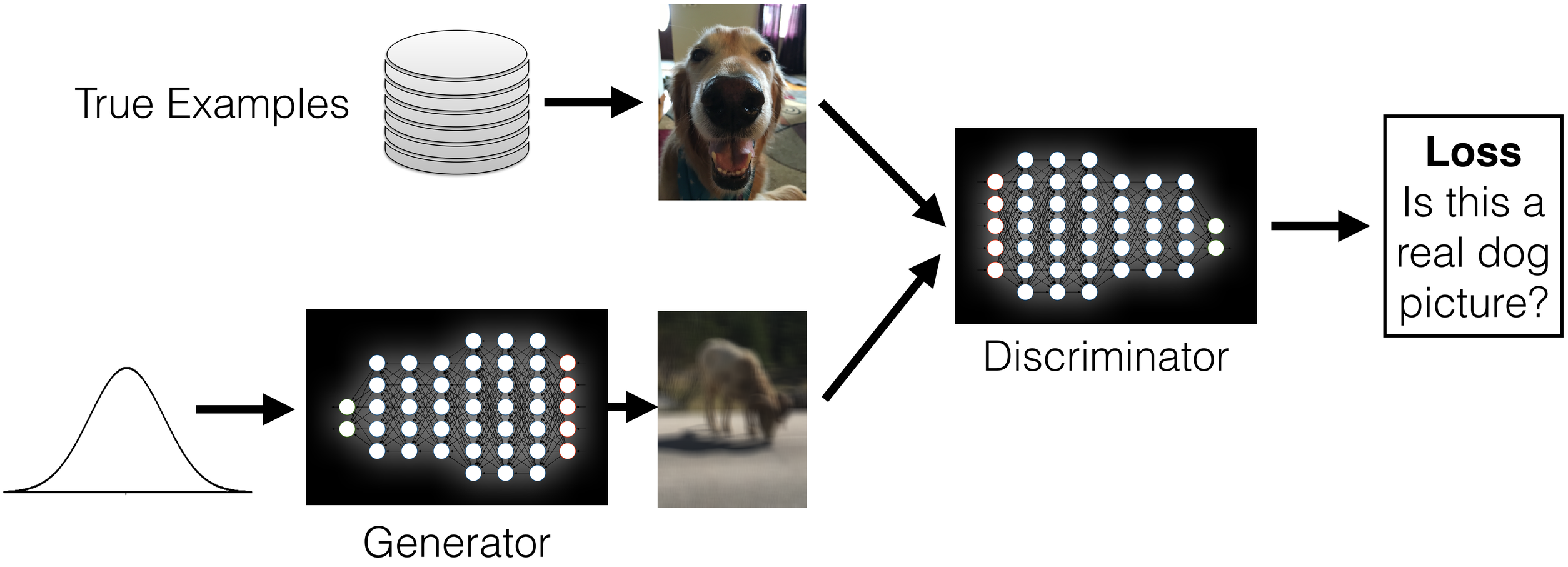
AE

VAE

AE

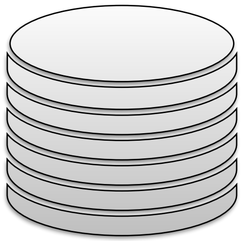


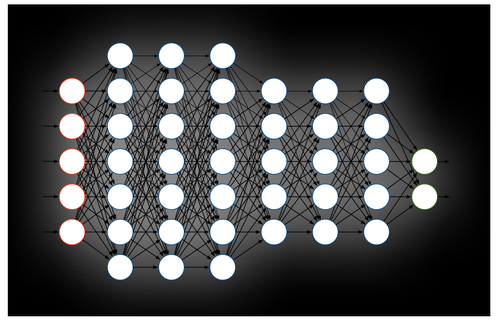
Generative Adversarial Networks



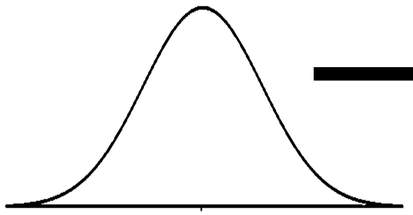
Generative Adversarial Networks

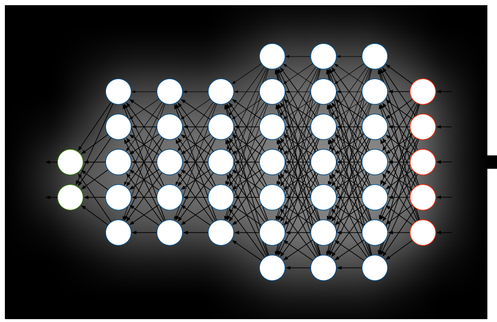
True Examples





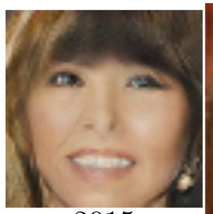
Loss
Is this a real dog picture?



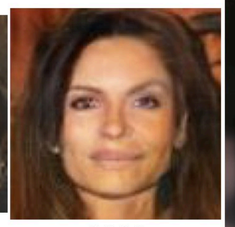


Generator

2014



2015

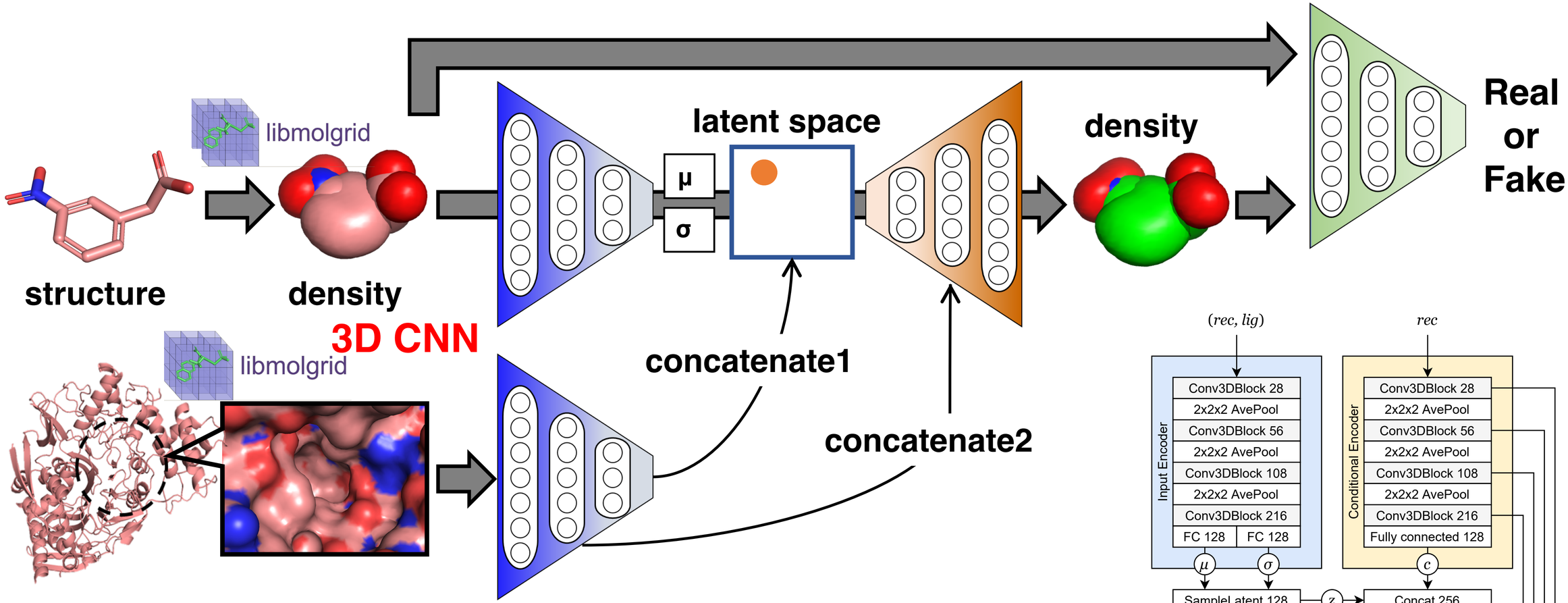


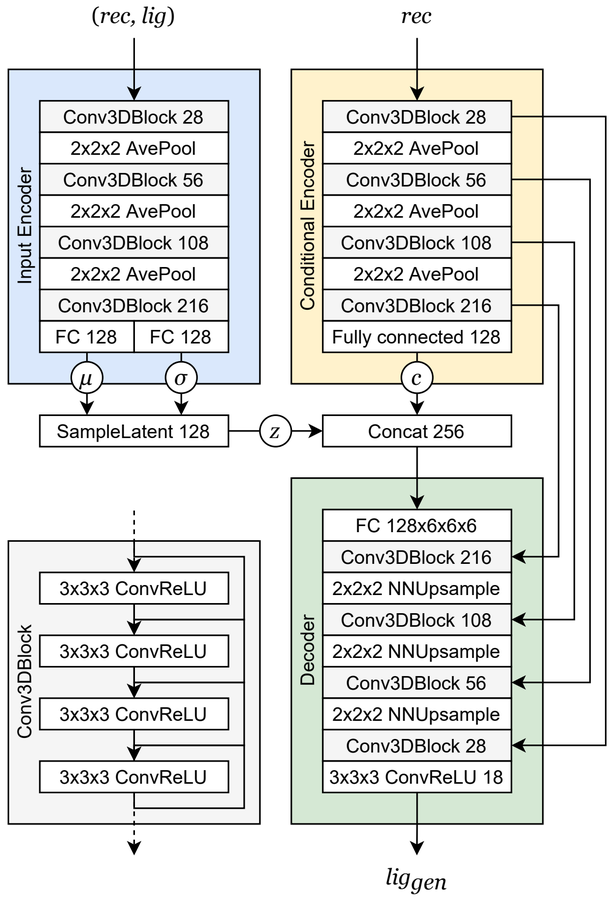
2016

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





Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

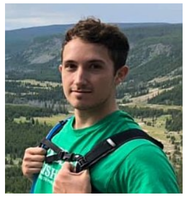
Check for updates

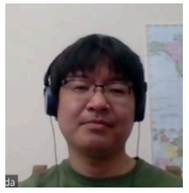
Cite this: *Chem. Sci.*, 2022, **13**, 2701

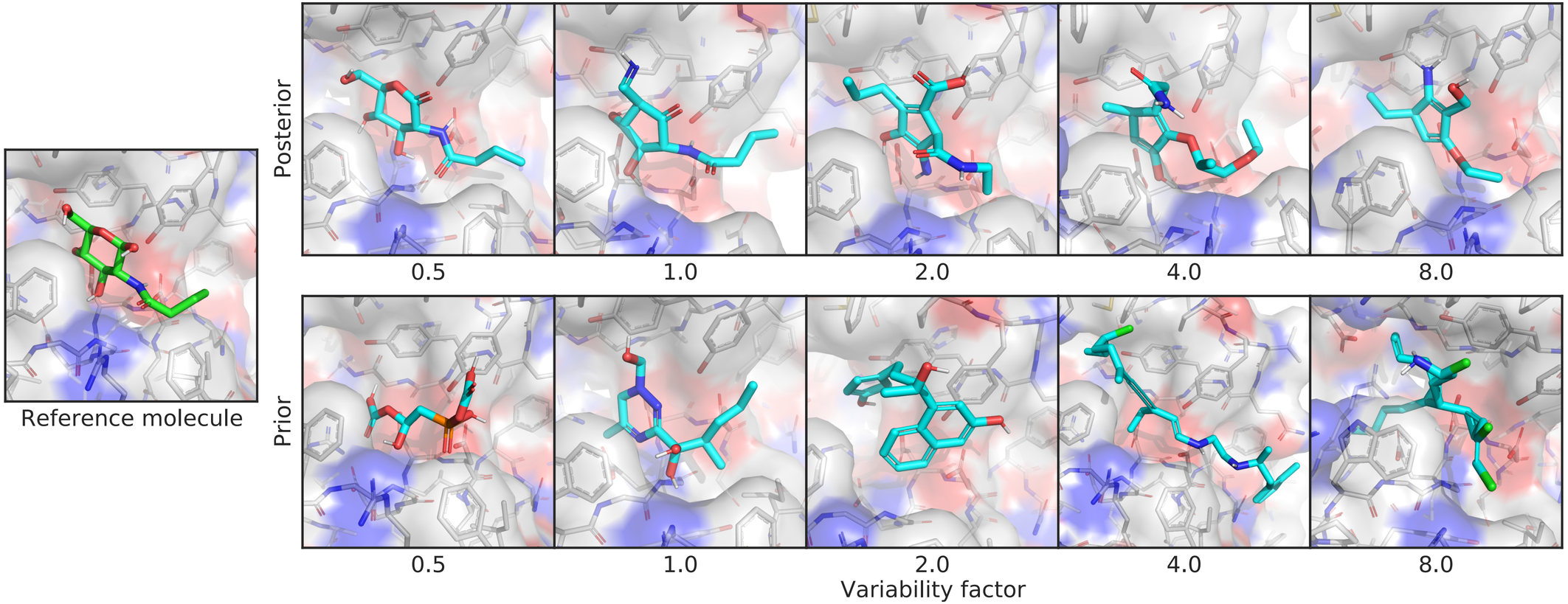
All publication charges for this article have been paid for by the Royal Society

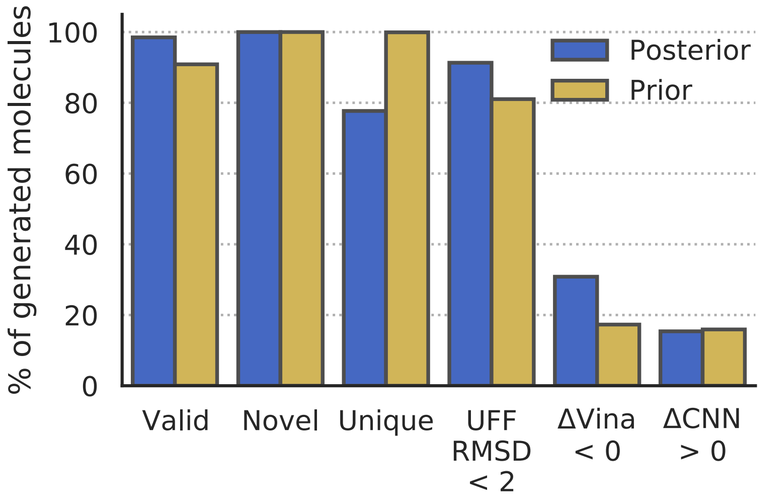
Generating 3D molecules conditional on receptor binding sites with deep generative models†

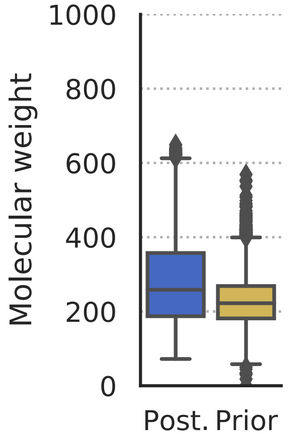
Matthew Ragoza, *[a] Tomohide Masuda[b] and David Ryan Koes[c]

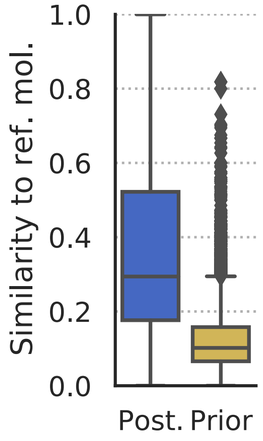


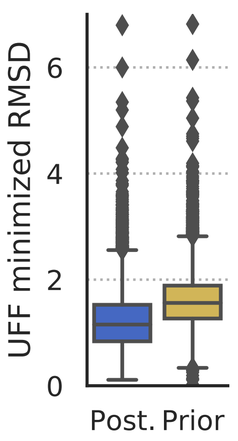


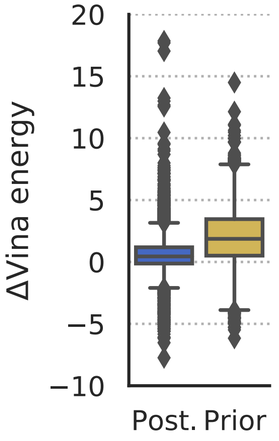

Reference molecule
Posterior
Prior
0.5
1.0
2.0
4.0
8.0
Variability factor

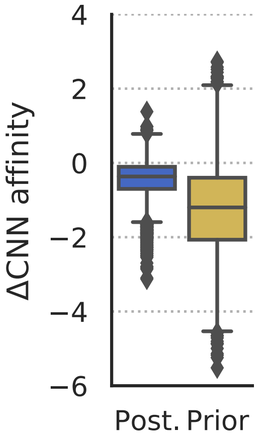

% of generated molecules
Posterior
Prior
Valid
Novel
Unique
UFF RMSD < 2
ΔVina < 0
ΔCNN > 0

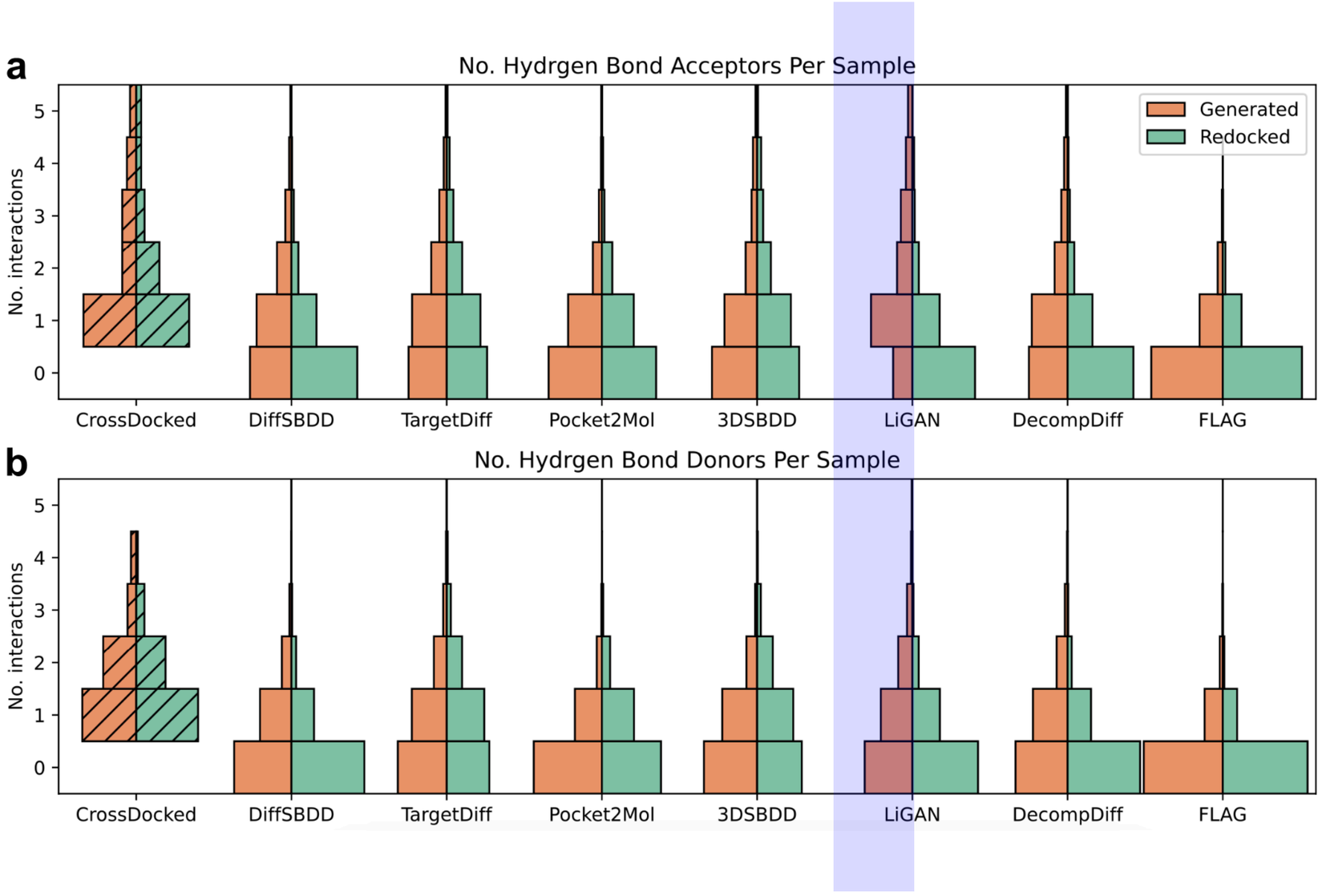

Molecular weight
Post. Prior


Similarity to ref. mol.
Post. Prior


UFF minimized RMSD
Post. Prior


ΔVina energy
Post. Prior


ΔCNN affinity
Post. Prior



PoseCheck: Generative Models for 3D Structure-based Drug Design Produce Unrealistic Poses

Charles Harris*
University of Cambridge
cch57@cam.ac.uk

Kieran Didi
University of Cambridge
ked48@cam.ac.uk

Arian R. Jamasb
University of Cambridge
arj39@cam.ac.uk

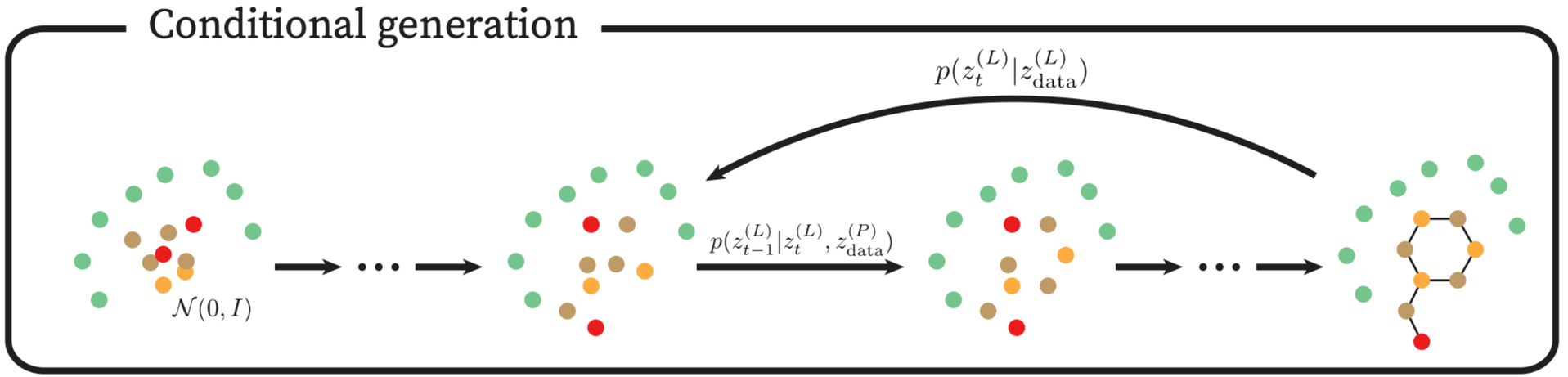
Chaitanya K. Joshi
University of Cambridge
ckj24@cam.ac.uk

Simon V. Mathis
University of Cambridge
svm34@cam.ac.uk

Pietro Lio
University of Cambridge
pl219@cam.ac.uk

Tom L. Blundell
University of Cambridge
tlb20@cam.ac.uk

Diffusion Models



Structure-based Drug Design with Equivariant Diffusion Models

Arne Schneuing[1]*, Yuanqi Du[2]*, Charles Harris[3], Arian Jamasb[3], Ilia Igashov[1], Weitao Du[4], Tom Blundell[3], Pietro Lió[3], Carla Gomes[2], Max Welling[5], Michael Bronstein[6] & Bruno Correia[1]
[1]École Polytechnique Fédérale de Lausanne, [2]Cornell University, [3]University of Cambridge, [4]USTC, [5]Microsoft Research AI4Science, [6]University of Oxford

arXiv > q-bio > arXiv:2210.01776

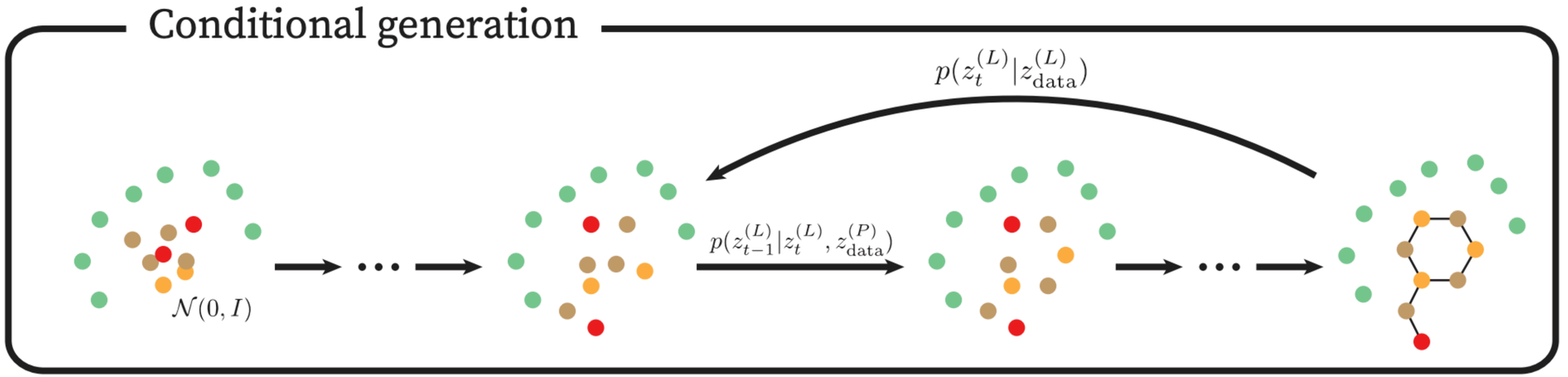
Quantitative Biology > Biomolecules

[Submitted on 4 Oct 2022 (v1), last revised 11 Feb 2023 (this version, v2)]

DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, Tommi Jaakkola

Diffusion Models



Structure-based Drug Design with Equivariant Diffusion Models

Arne Schneuing[1]*, Yuanqi Du[2]*, Charles Harris[3], Arian Jamasb[3], Ilia Igashov[1], Weitao Du[4], Tom Blundell[3], Pietro Lió[3], Carla Gomes[2], Max Welling[5], Michael Bronstein[6] & Bruno Correia[1]

[1]École Polytechnique Fédérale de Lausanne, [2]Cornell University, [3]University of Cambridge, [4]USTC, [5]Microsoft Research AI4Science, [6]University of Oxford

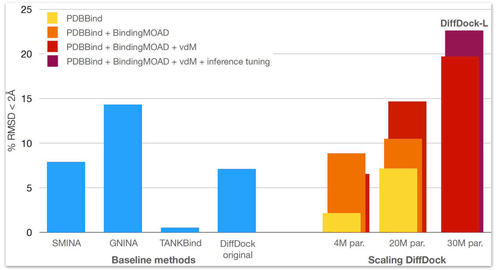
arXiv > q-bio > arXiv:2210.01776

Quantitative Biology > Biomolecules

[Submitted on 4 Oct 2022 (v1), last revised 11 Feb 2023 (this version, v2)]

DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, Tommi Jaakkola



Keypoint Conditioned Diffusion

Pocket Identification

↓

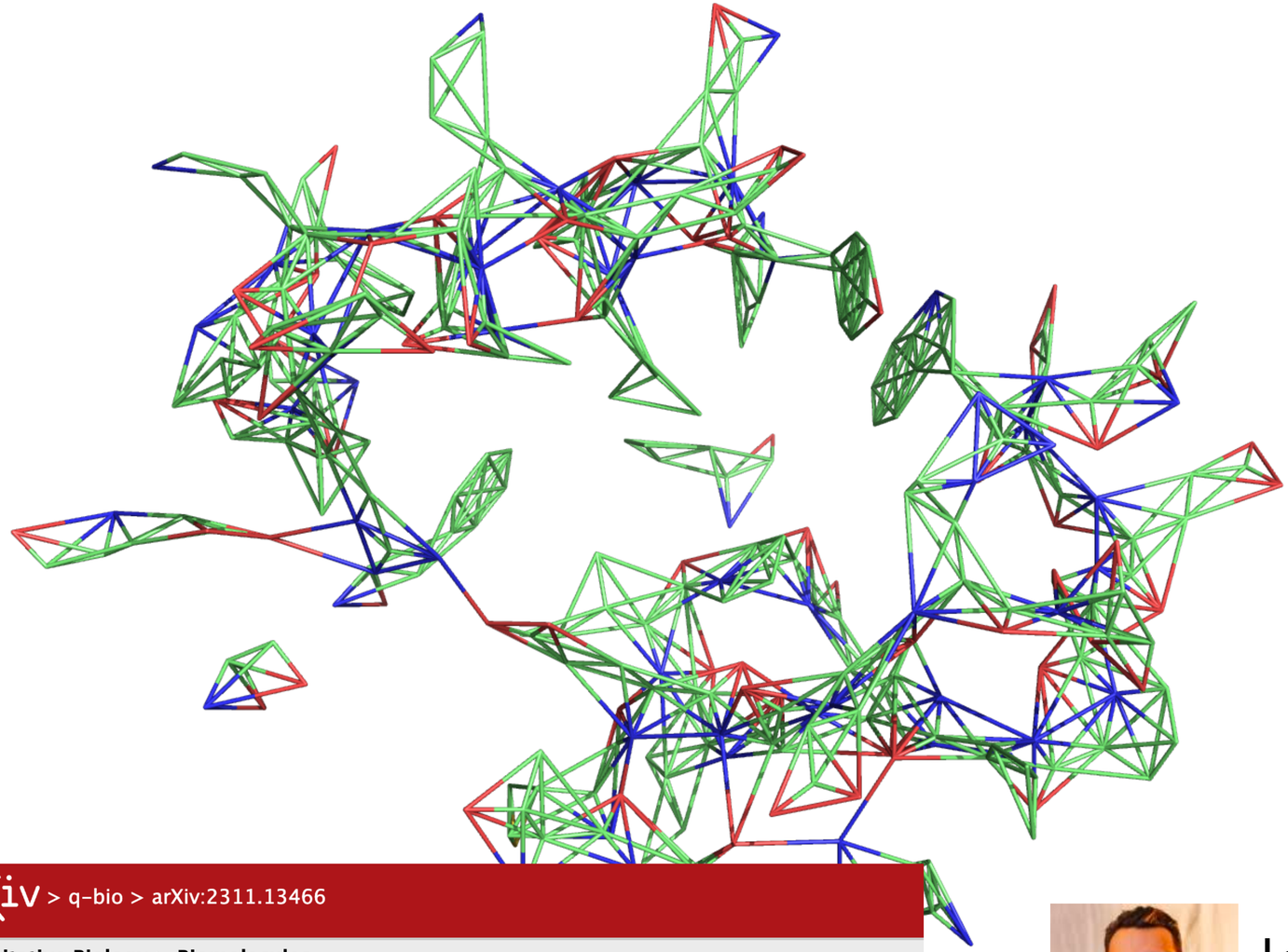
Atomic Message Passing

↓

Keypoint Generation

↓

Keypoint-Conditioned Diffusion



arXiv > q-bio > arXiv:2311.13466

Quantitative Biology > Biomolecules

[Submitted on 22 Nov 2023 (v1), last revised 8 May 2024 (this version, v2)]

Accelerating Inference in Molecular Diffusion Models with Latent Representations of Protein Structure

Ian Dunn, David Ryan Koes

Ian Dunn

Keypoint Conditioned Diffusion

Pocket Identification

↓

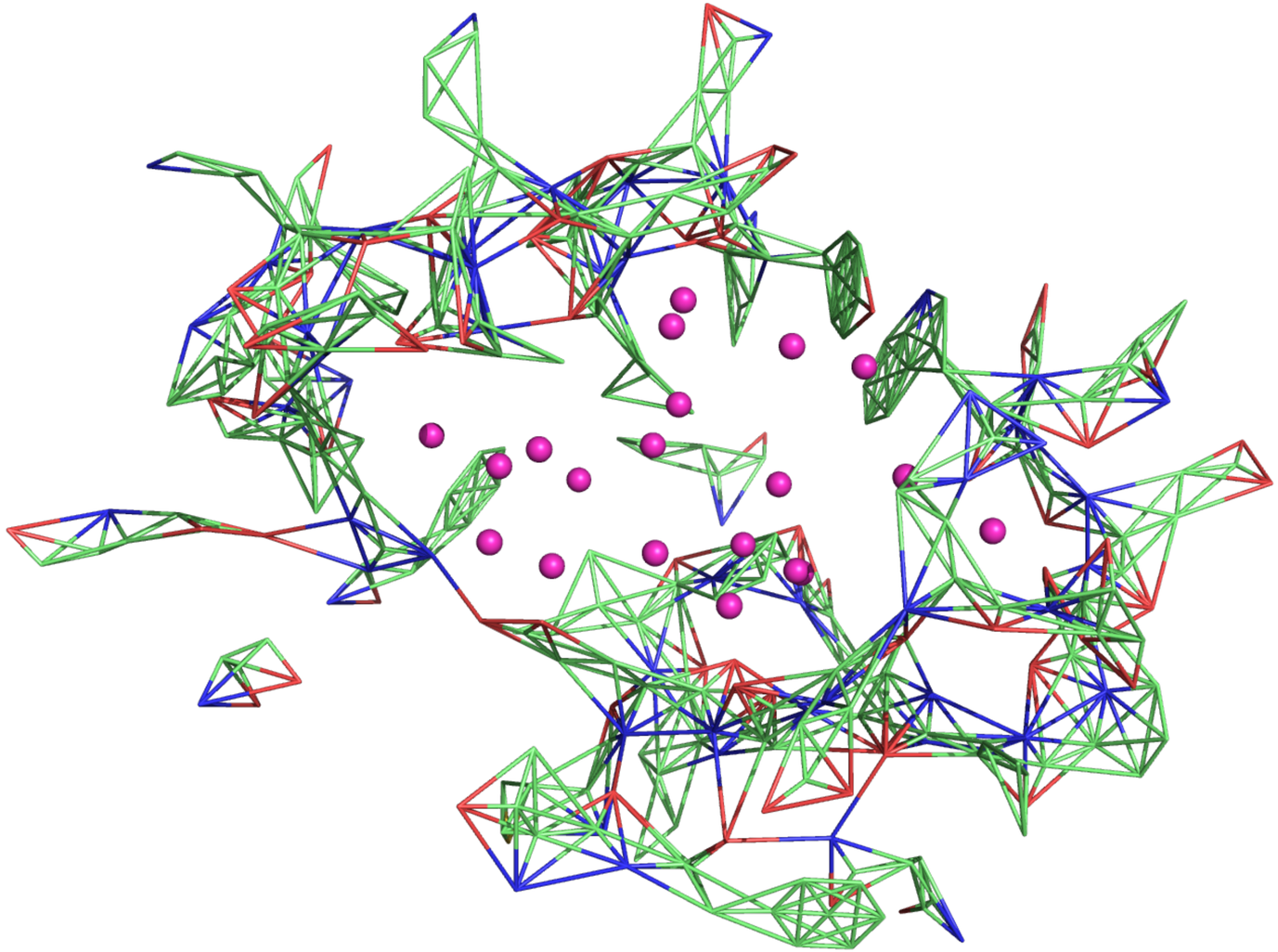
Atomic Message Passing

↓

Keypoint Generation

↓

Keypoint-Conditioned Diffusion



Keypoint Conditioned Diffusion

Pocket Identification

↓

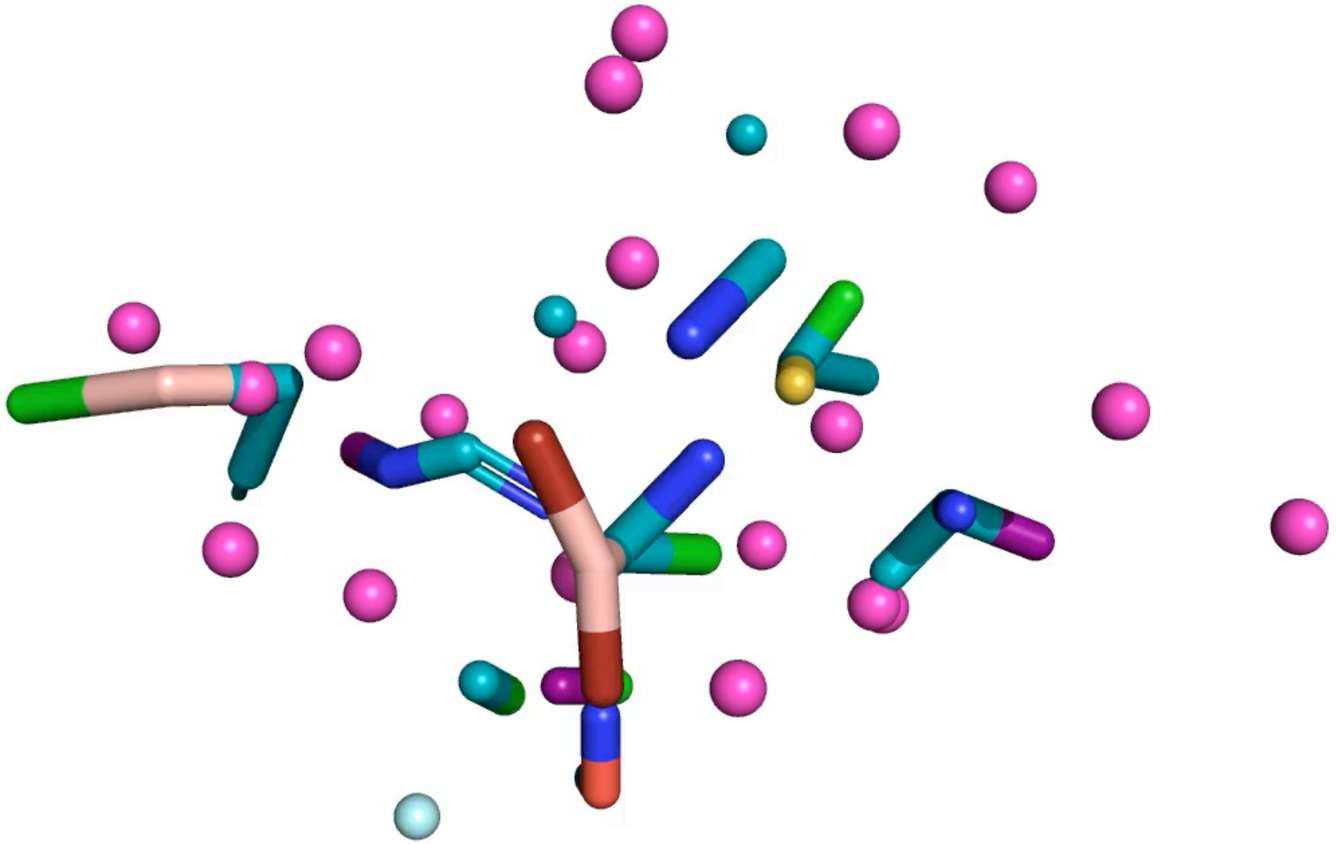
Atomic Message Passing

↓

Keypoint Generation

↓

Keypoint-Conditioned Diffusion



Keypoint Conditioned Diffusion

Pocket Identification

↓

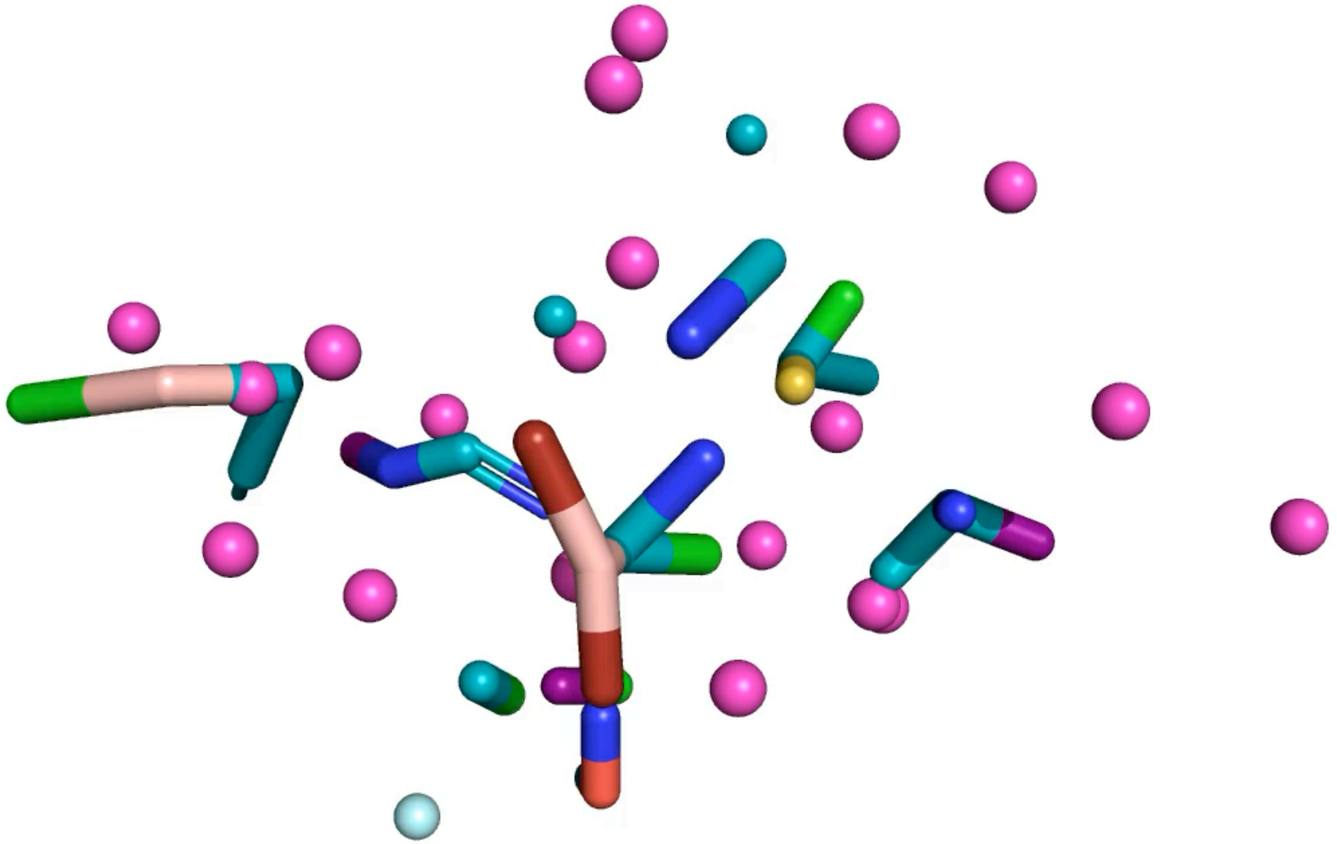
Atomic Message Passing

↓

Keypoint Generation

↓

Keypoint-Conditioned Diffusion



Keypoint Conditioned Diffusion

Pocket Identification

↓

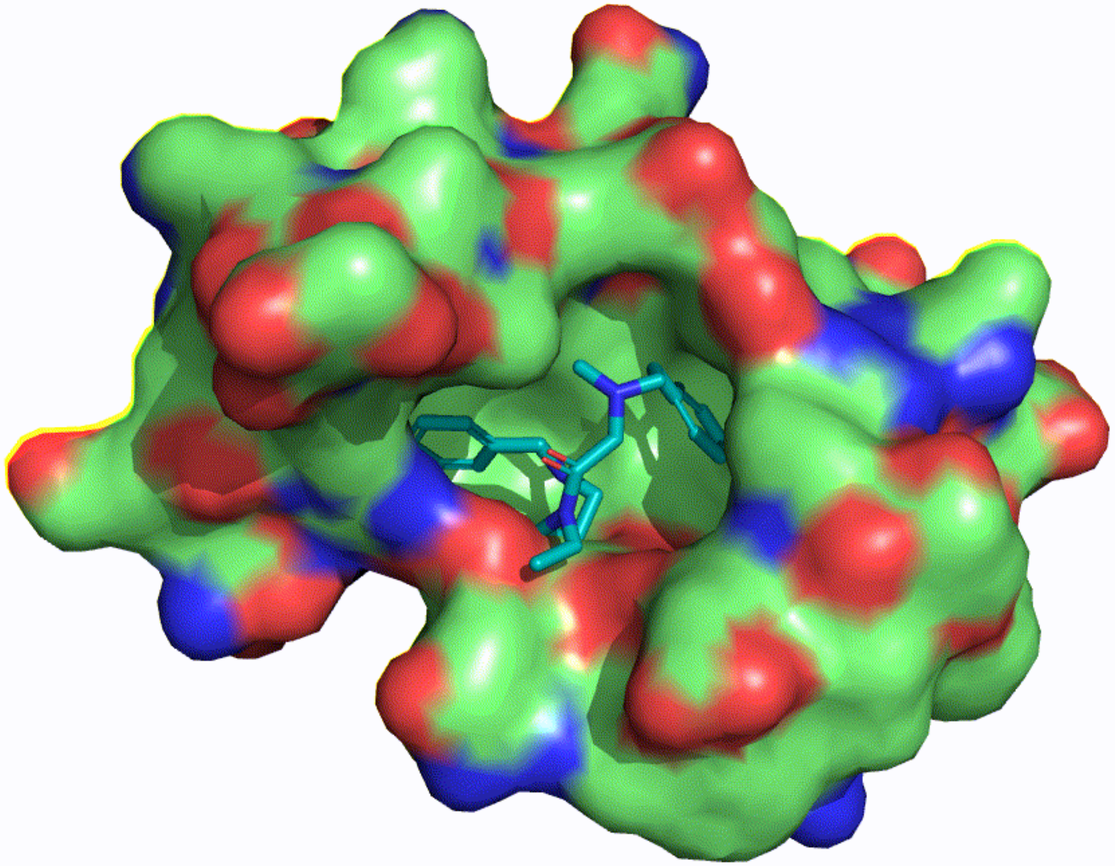
Atomic Message Passing

↓

Keypoint Generation

↓

Keypoint-Conditioned Diffusion



Keypoint Conditioned Diffusion

Pocket Identification

↓

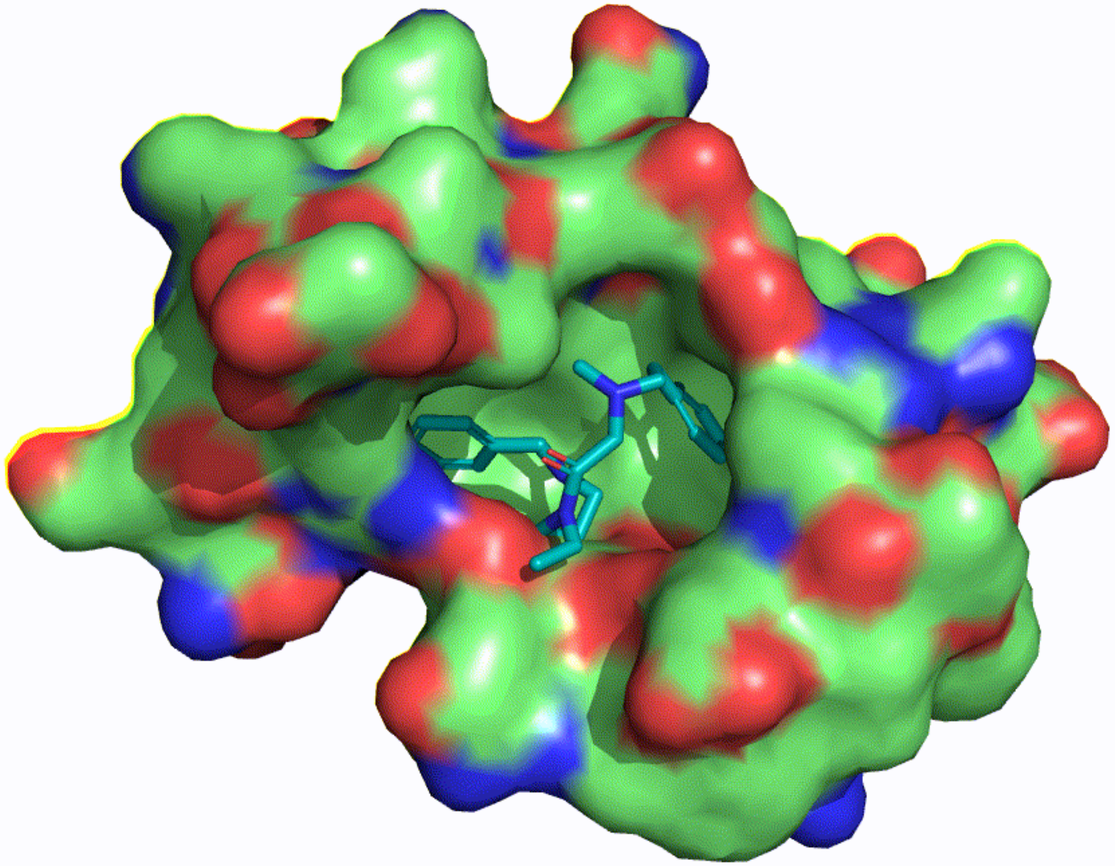
Atomic Message Passing

↓

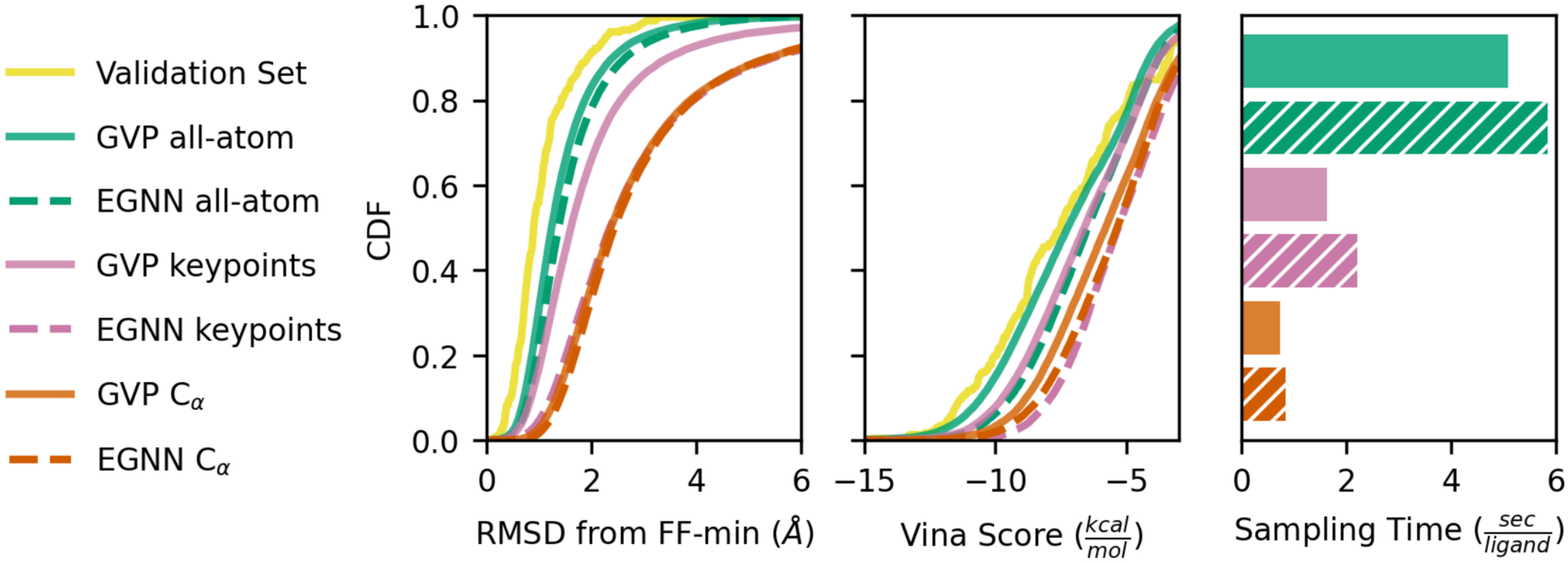
Keypoint Generation

↓

Keypoint-Conditioned Diffusion

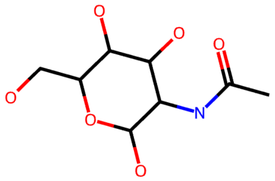


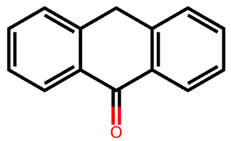
Keypoint Conditioned Diffusion

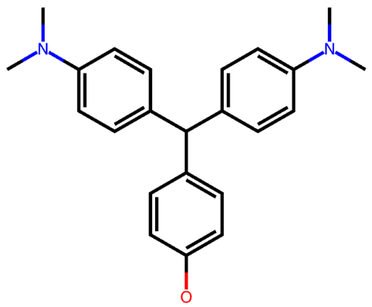


Generated molecules are not synthetically accessible or physically plausible

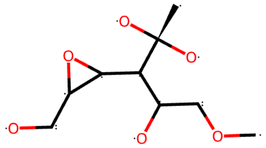
Reference Molecules

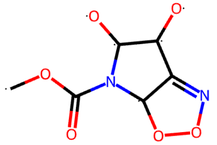


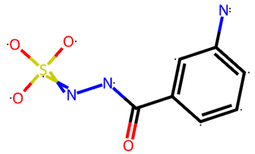


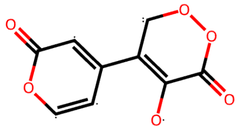


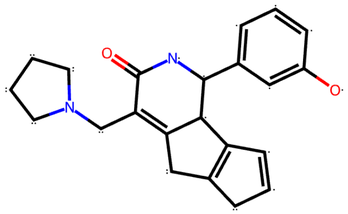
Generated Molecules

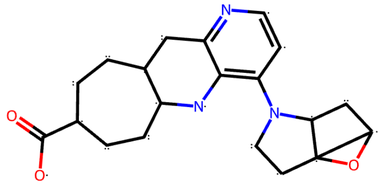






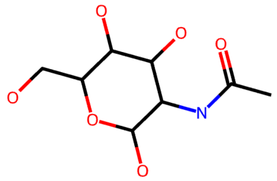


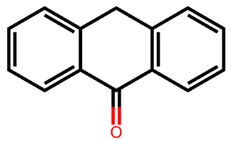


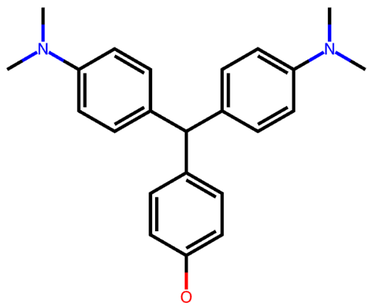


Generated molecules are not synthetically accessible or physically plausible

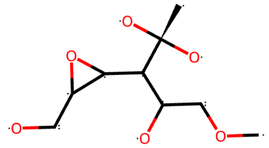
Reference Molecules

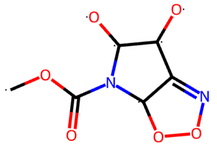




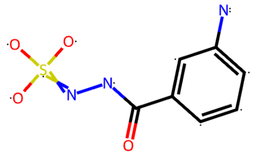


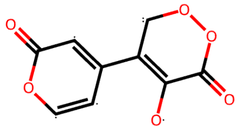
Generated Molecules

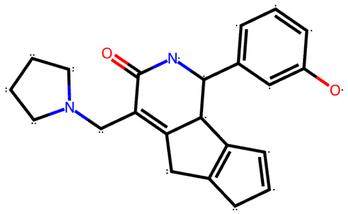


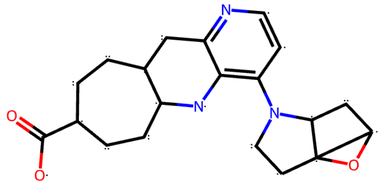


"epoxide, good luck with stability"



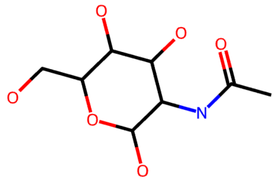


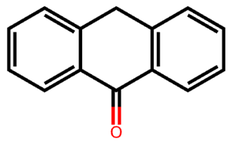


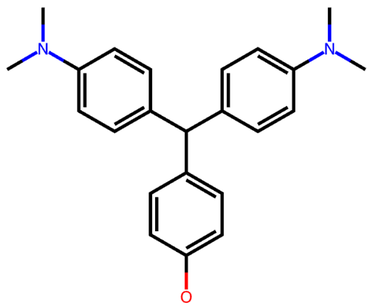


Generated molecules are not synthetically accessible or physically plausible

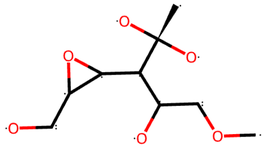
Reference Molecules

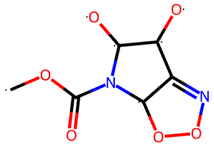




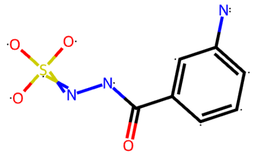


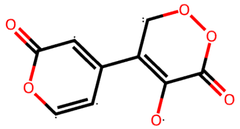
Generated Molecules

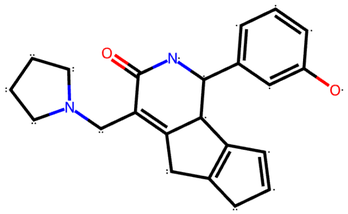


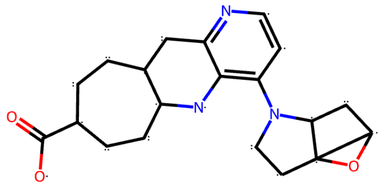


"epoxide, good luck with stability"





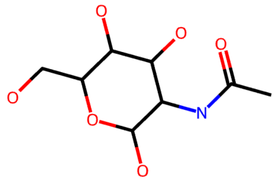


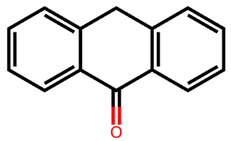


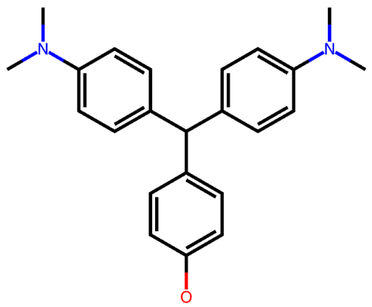
"certainly unstable"

Generated molecules are not synthetically accessible or physically plausible

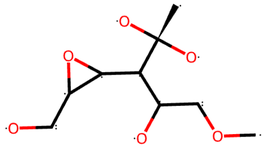
Reference Molecules

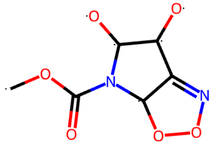




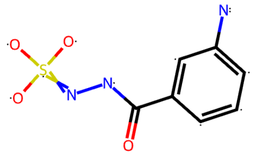


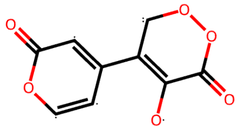
Generated Molecules



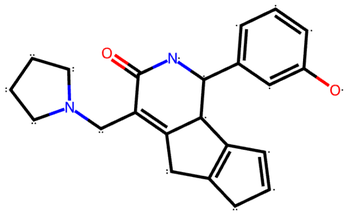


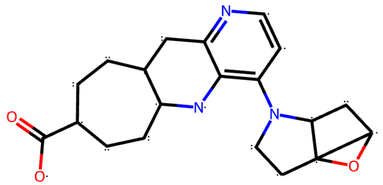
"epoxide, good luck with stability"





"no"

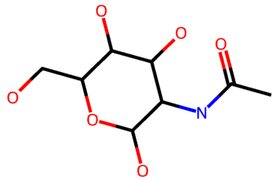


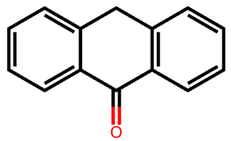


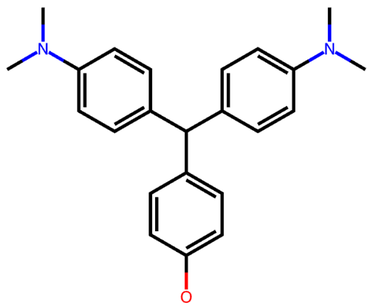
"certainly unstable"

Generated molecules are not synthetically accessible or physically plausible

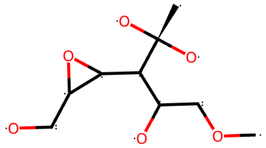
Reference Molecules



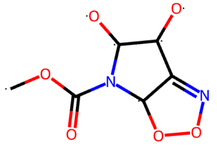




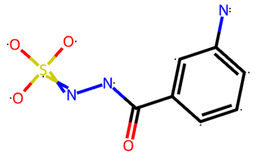
Generated Molecules



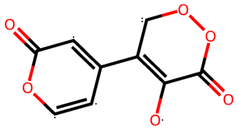
"epoxide, good luck with stability"

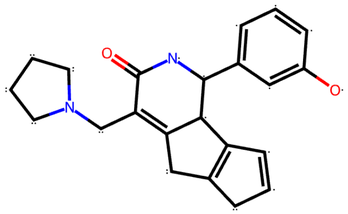


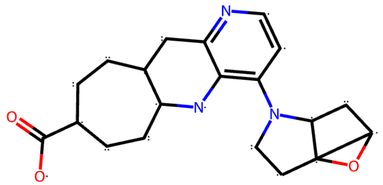
"f*** no"



"no"







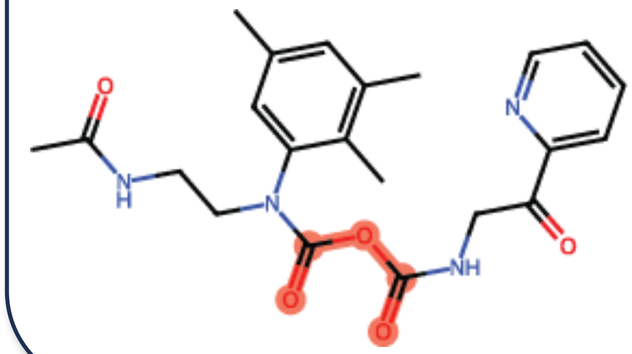
"certainly unstable"

Practical Measures of Molecule Quality

Existing literature primarily focuses on validity/valency; necessary but insufficient dimensions of molecule quality

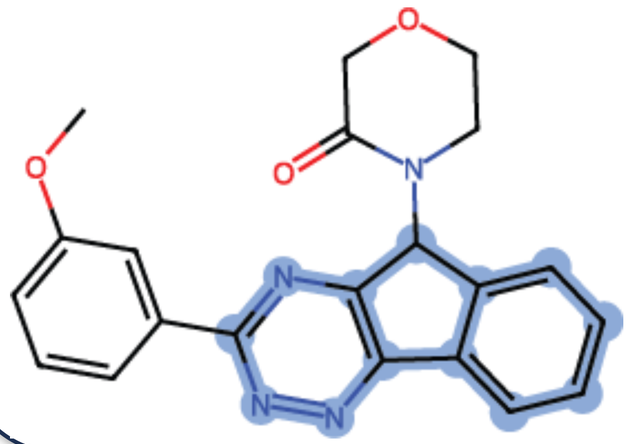
We propose to evaluate molecule quality at the level of **functional groups** and **ring systems**

Structural Alerts



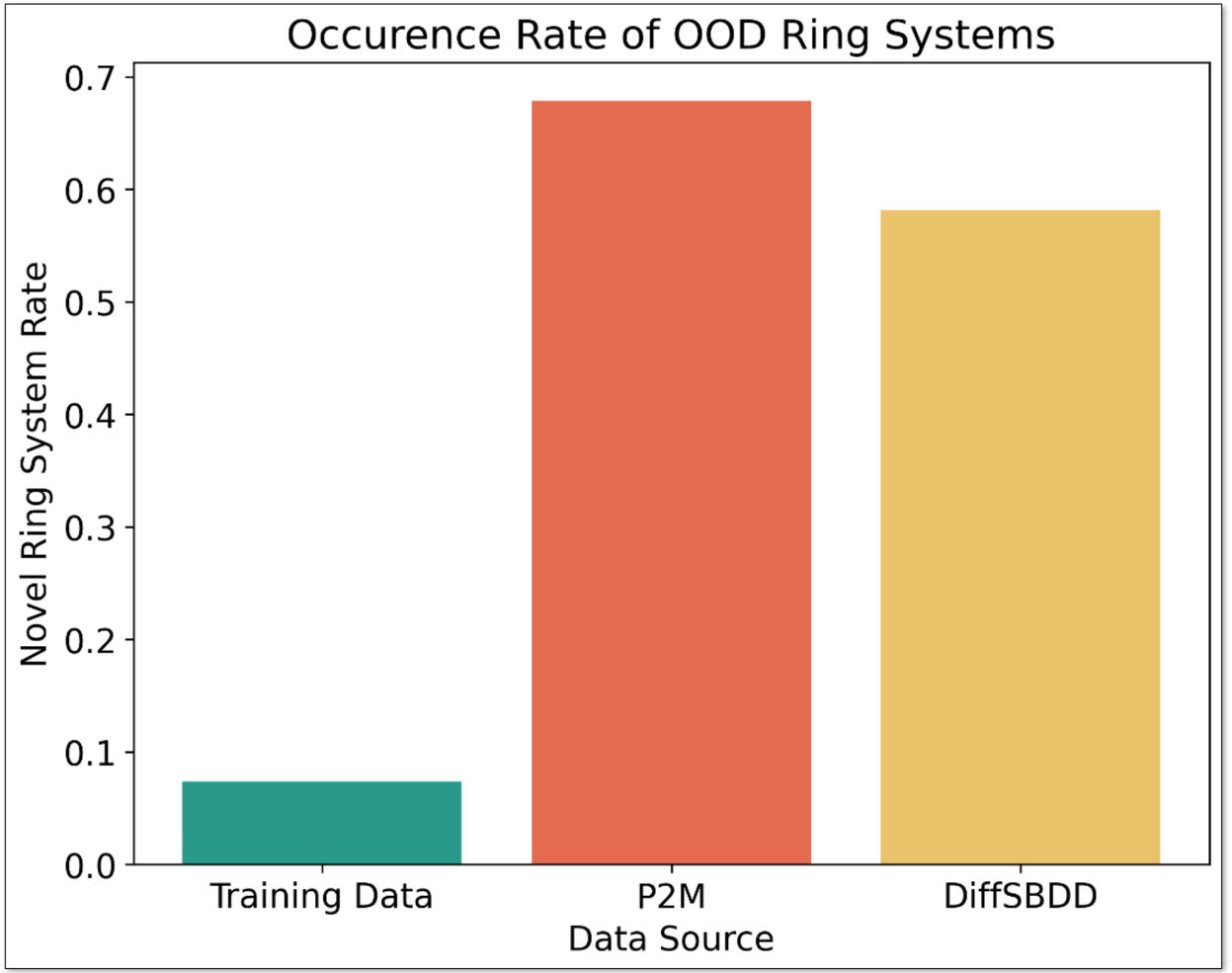
Functional groups typically used to filter libraries in virtual screening workflows

OOD Rings

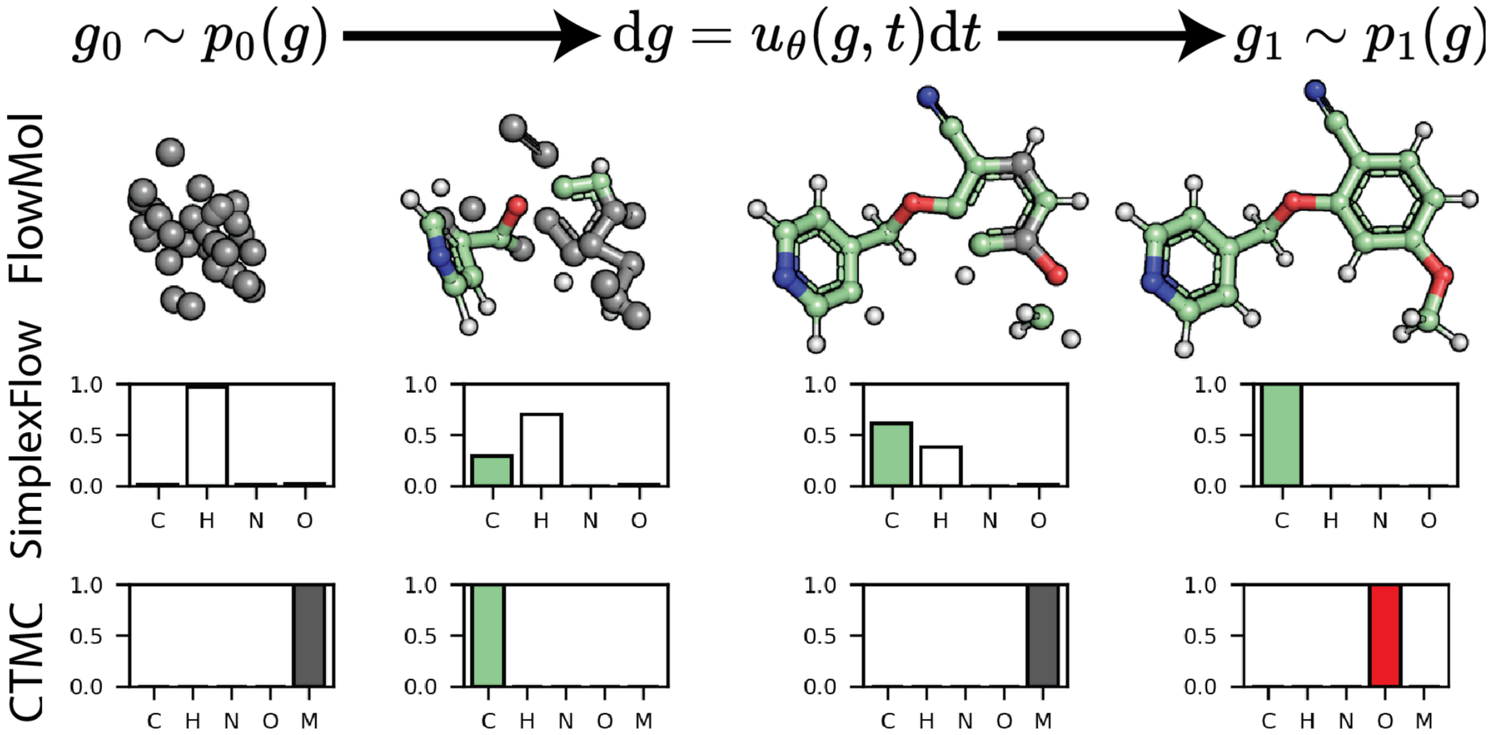


Ring systems that never appear in ChEMBL

Practical Measures of Molecule Quality



Unconditional Generation with FlowMol

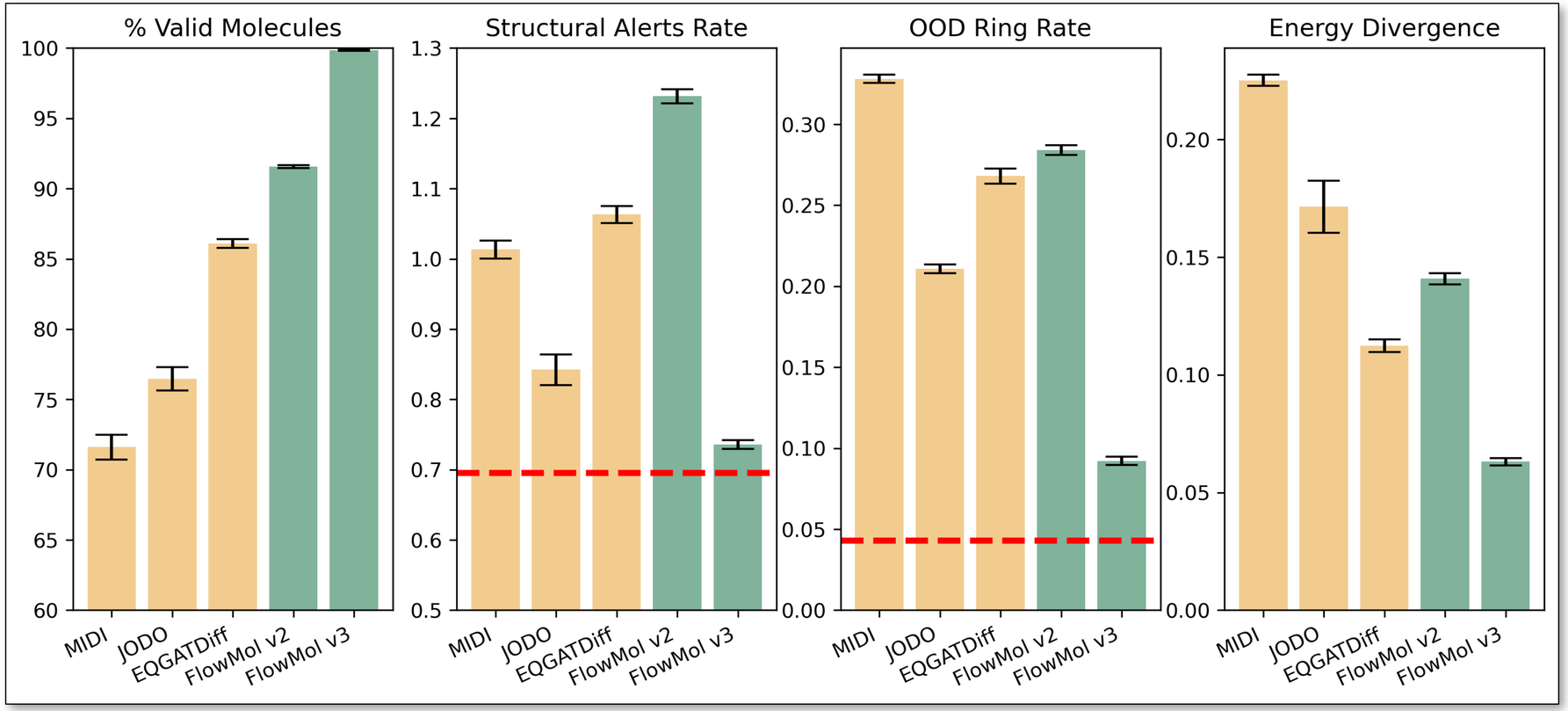


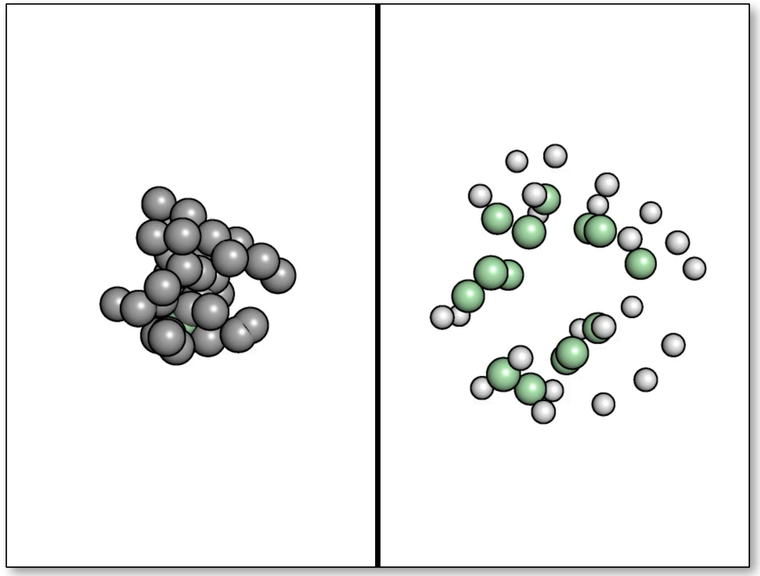
Exploring Discrete Flow Matching for 3D De Novo Molecule Generation

Ian Dunn
Dept. of Computational & Systems Biology
University of Pittsburgh
Pittsburgh, PA 15260
ian.dunn@pitt.edu

David Ryan Koes
Dept. of Computational & Systems Biology
University of Pittsburgh
Pittsburgh, PA 15260
dkoes@pitt.edu

FlowMol v3





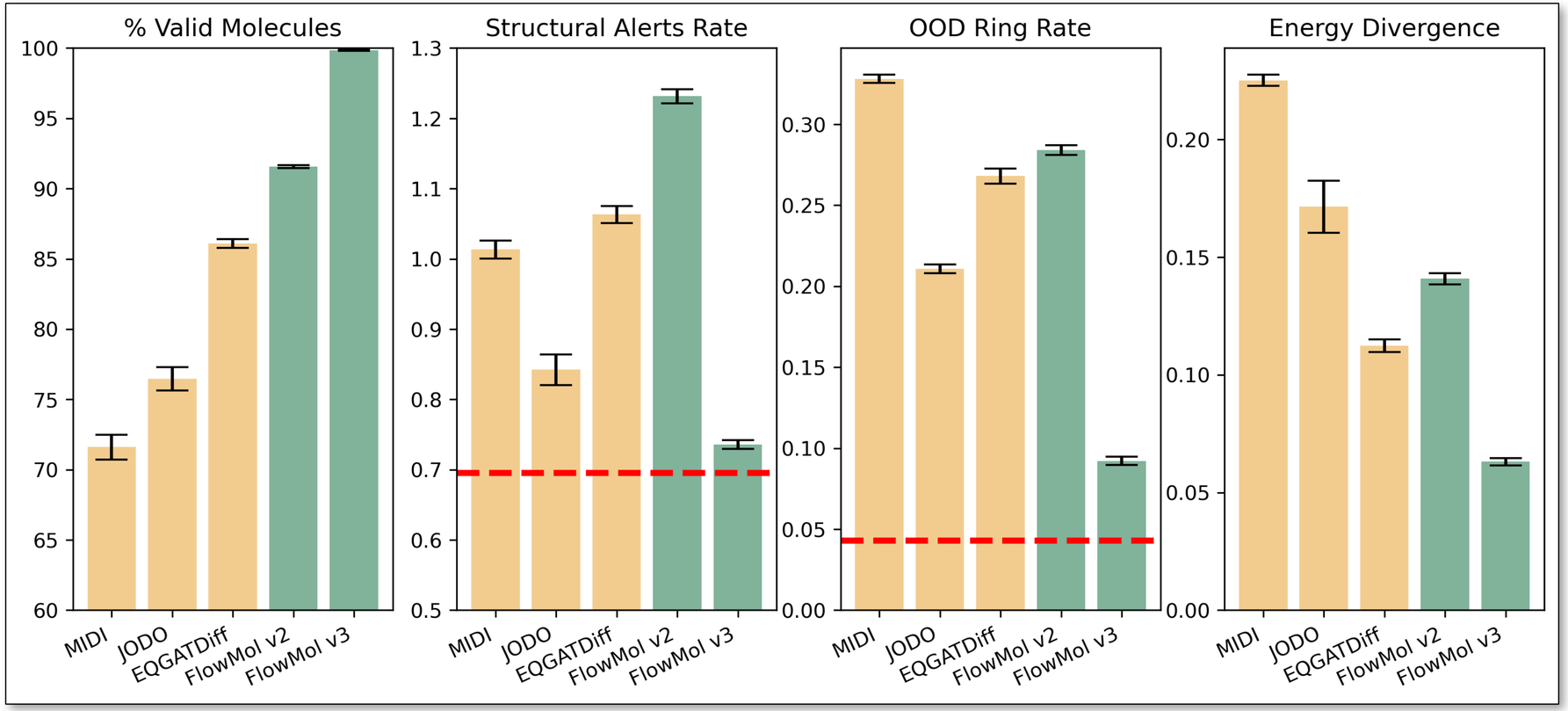
FlowMol Public

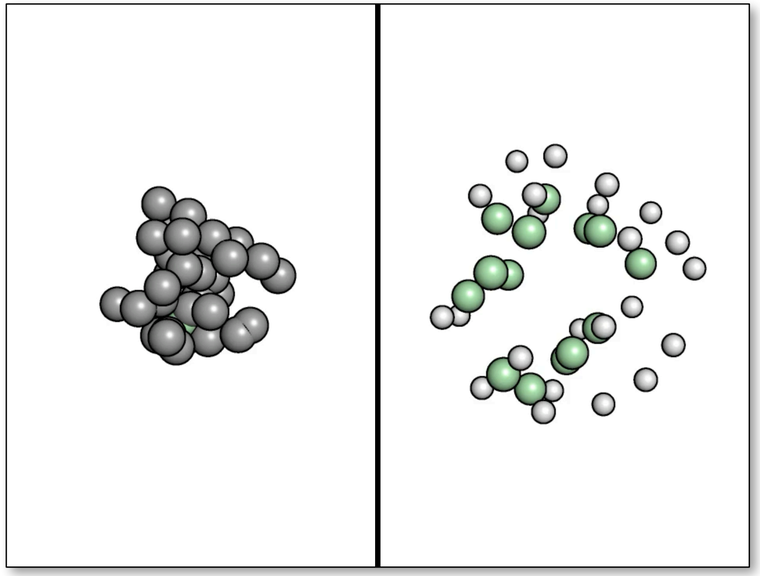
Mixed continous/categorical flow-matching model for de novo molecule generation.

Python ☆ 105 5

- State of the art validity
- Improves chemical plausibility and synthetic accessibility

FlowMol v3





FlowMol Public

Mixed continous/categorical flow-matching model for de novo molecule generation.

Python ☆ 105 5

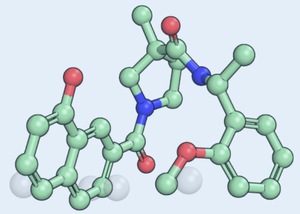
- State of the art validity
- Improves chemical plausibility and synthetic accessibility

OMTRA

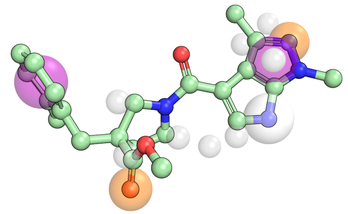
FlowMol3

De novo Generation

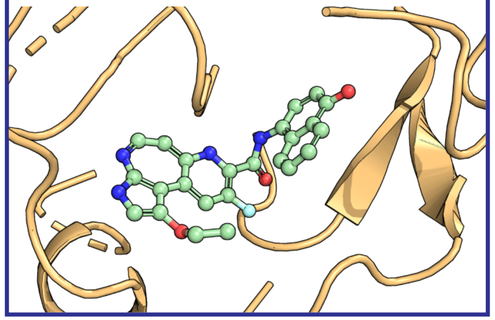
Unconditional



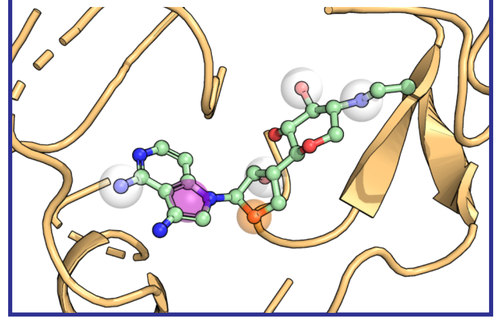
Pharmacophore Conditioned



Pocket Conditioned

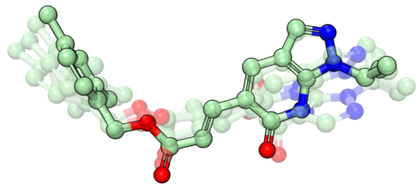


Pocket & Pharm. Conditioned

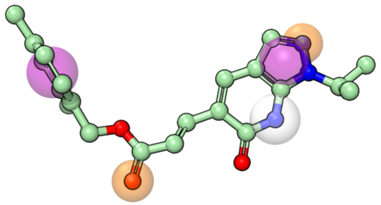


Conformer Generation

Unconditional

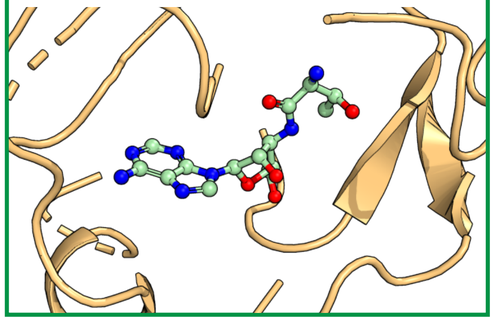


Pharmacophore Conditioned

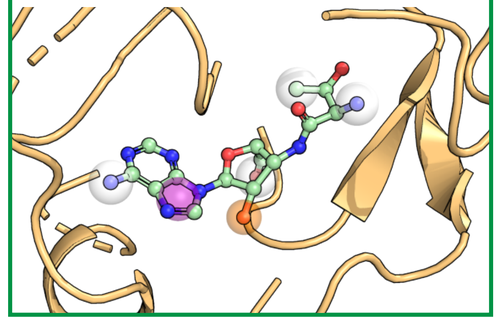


Docking

Pocket Conditioned

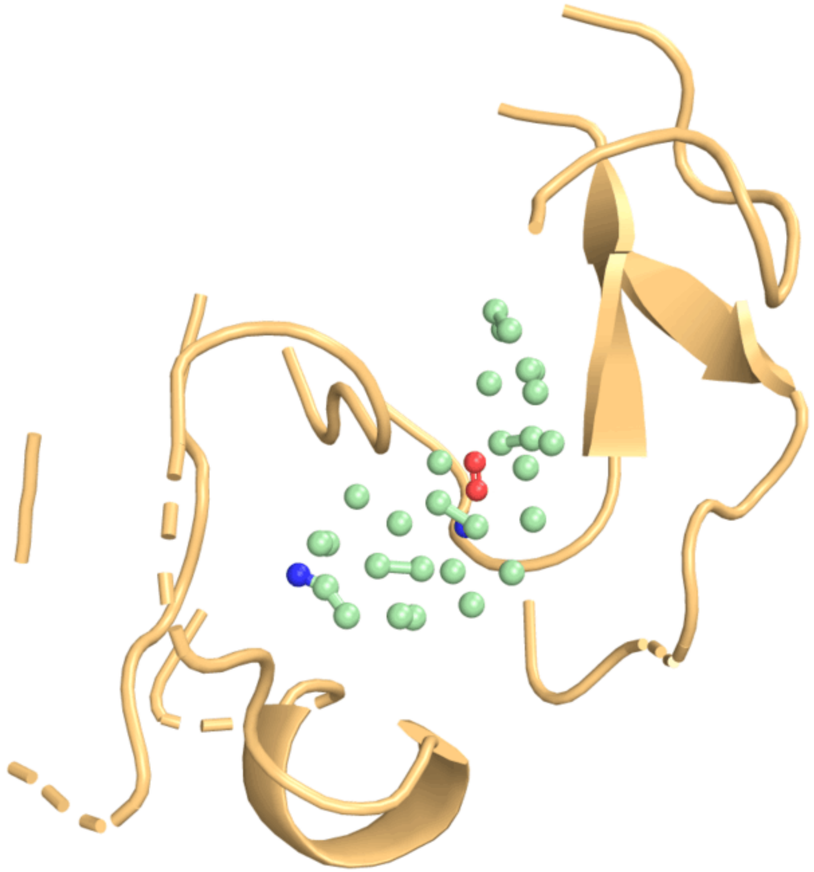


Pocket & Pharm. Conditioned

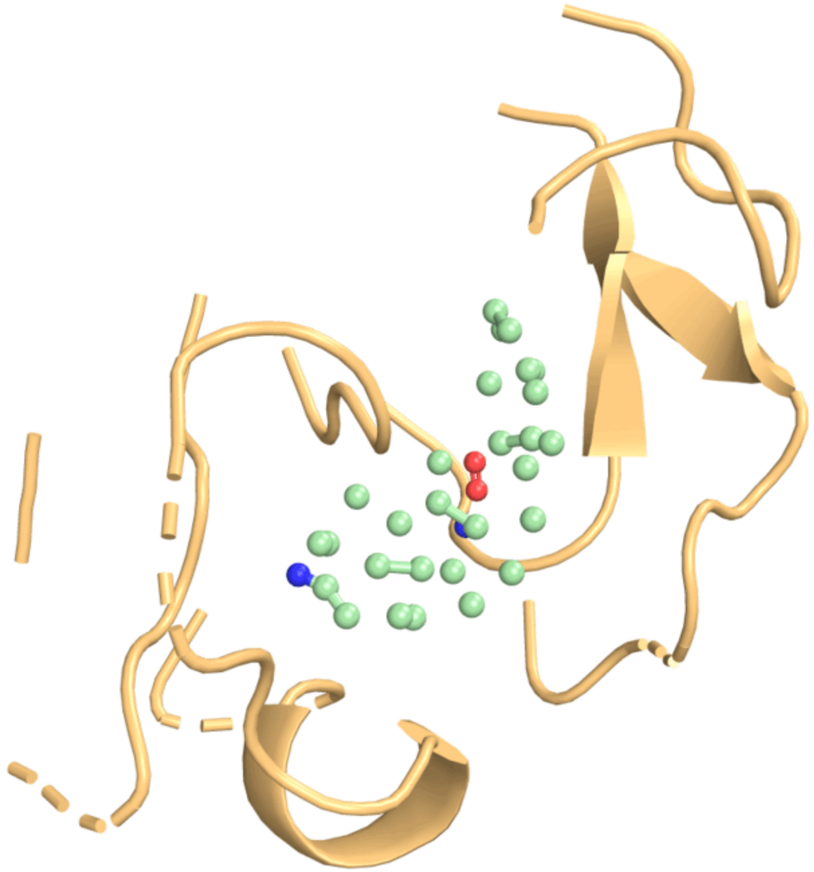


Ian Dunn Tyler Katz Liv Toft Riya Shah Juhi Gupta Ramith Hettiarachchi

OMTRA: De Novo Design

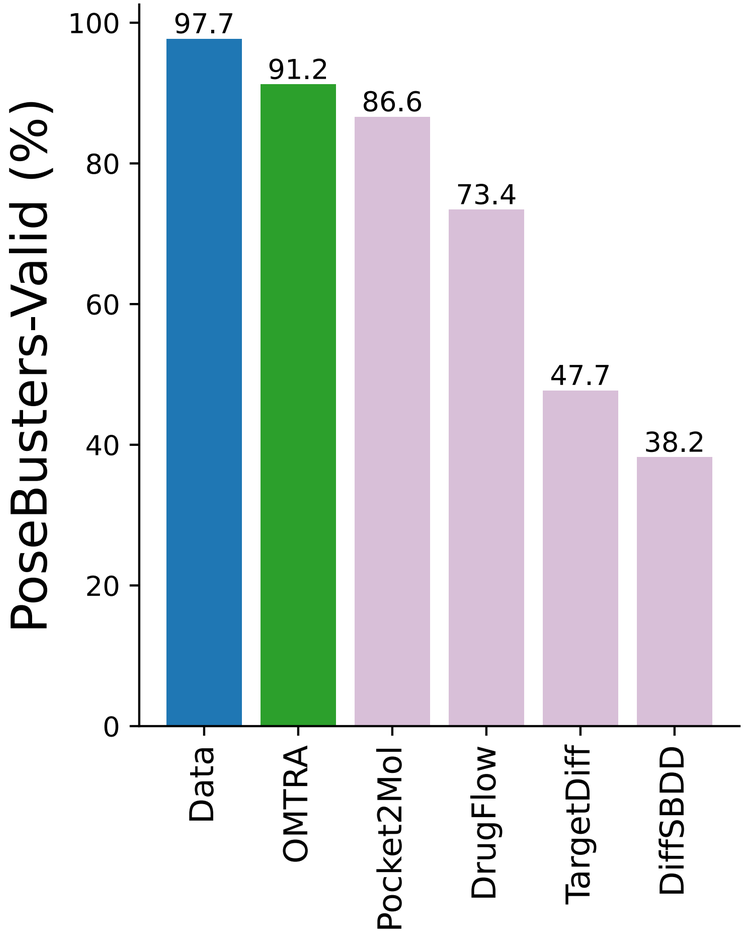


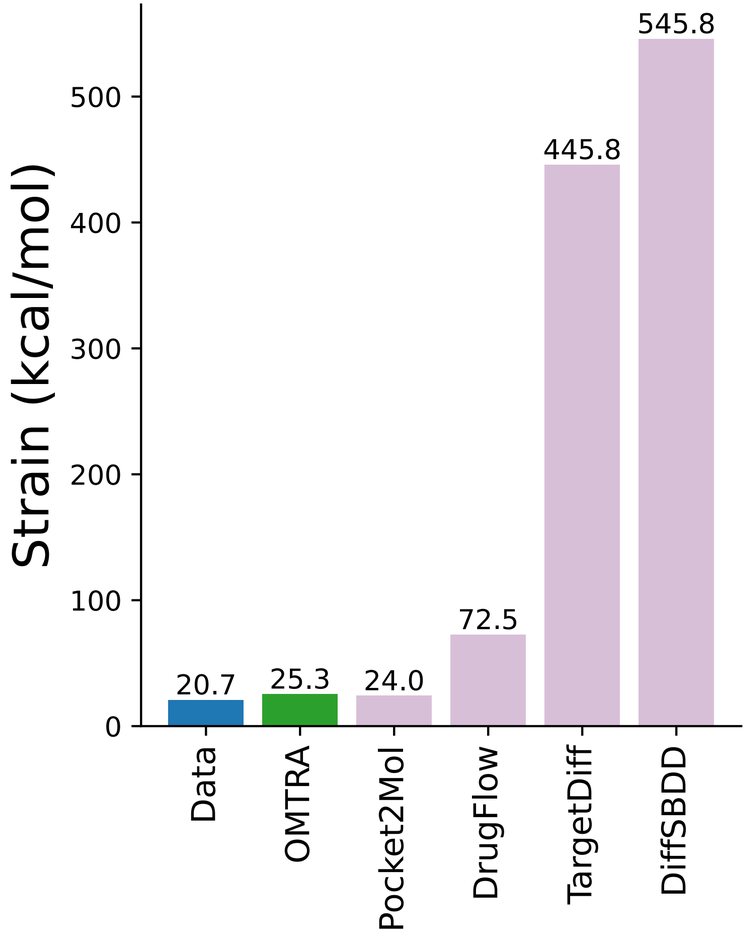
OMTRA: De Novo Design

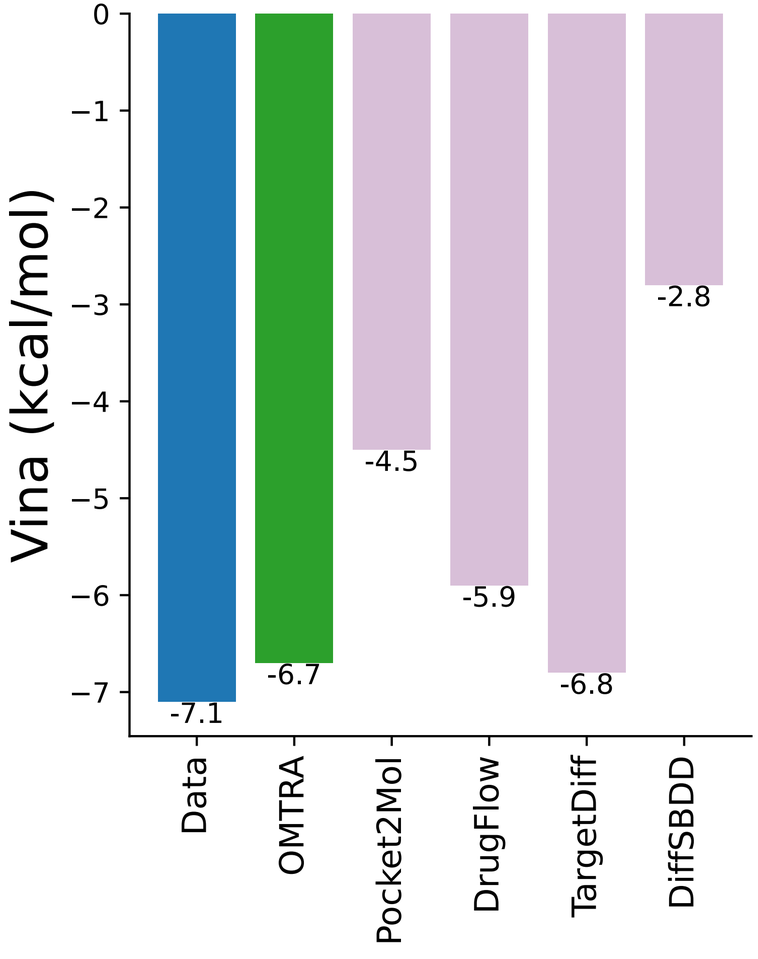


OMTRA: De Novo Design

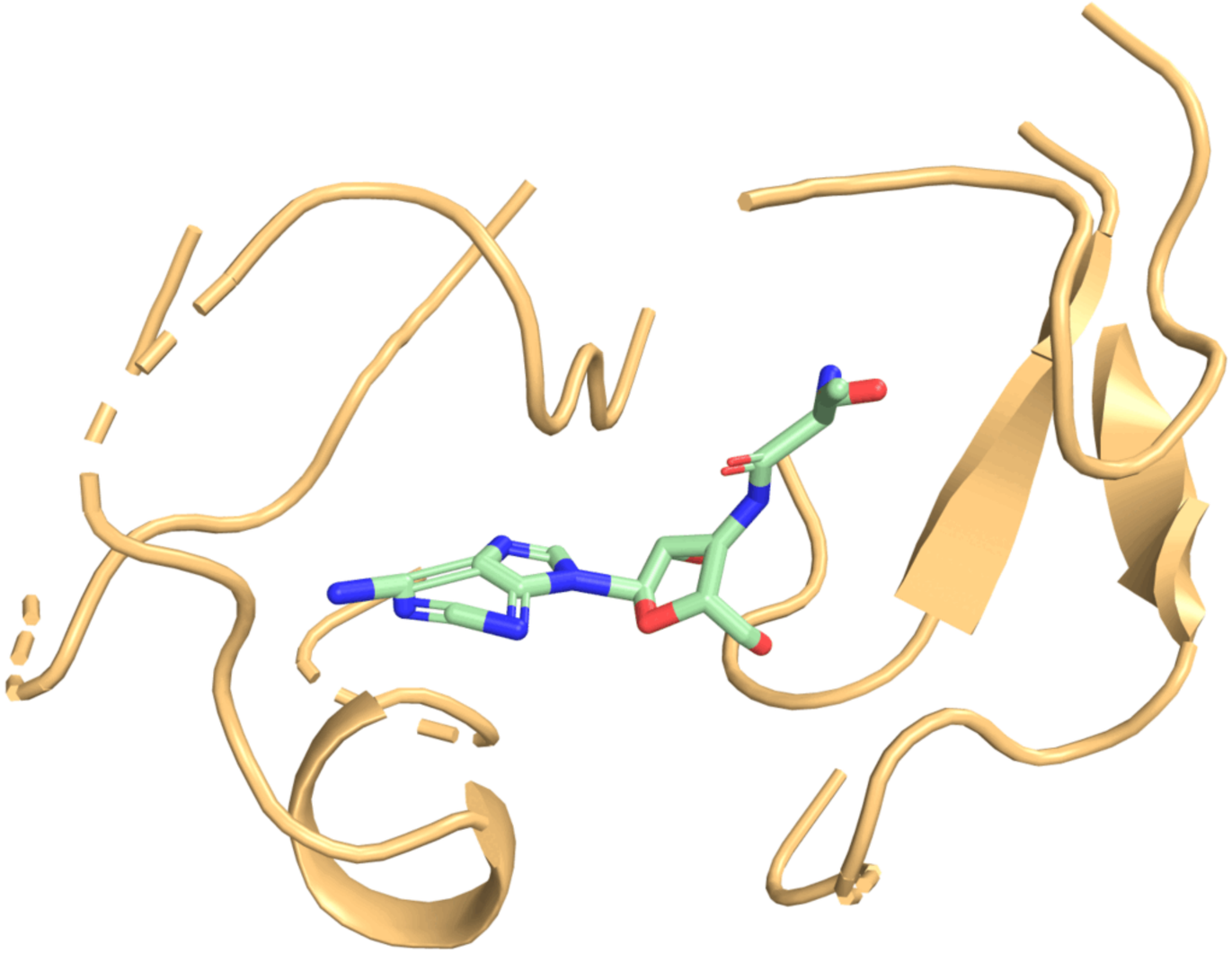
Evaluated on Luo et al CrossDocked test set.



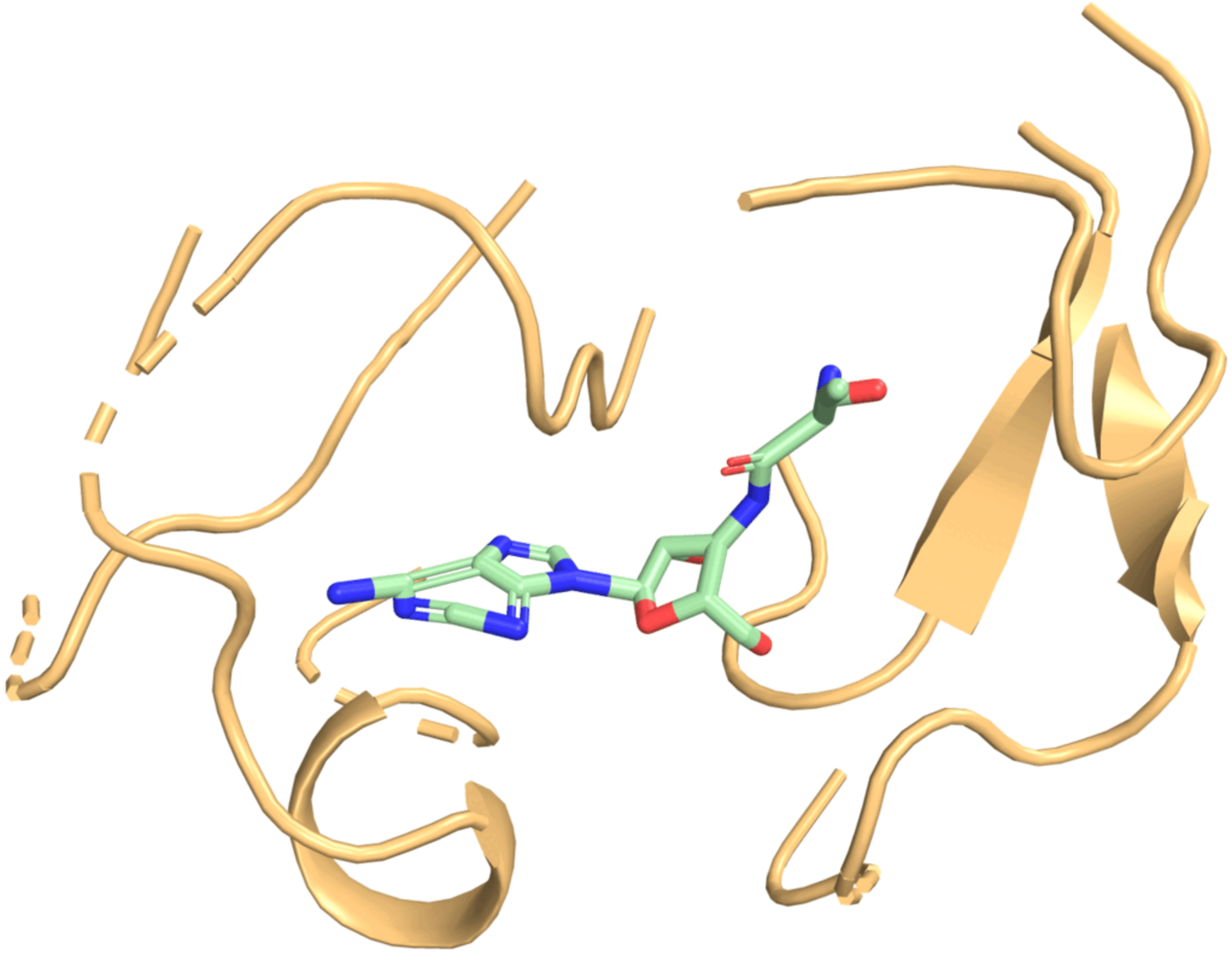




OMTRA: Docking

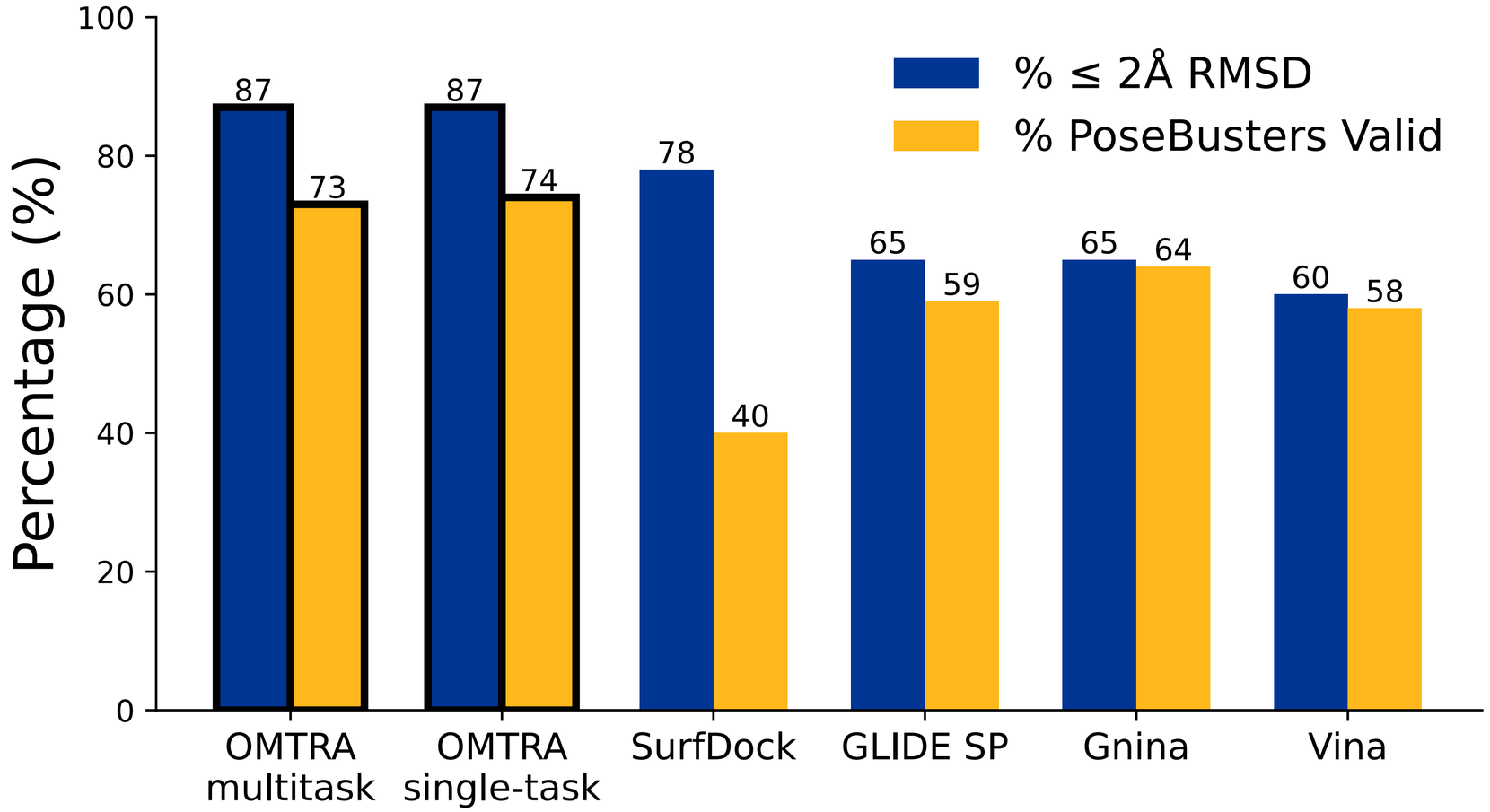


OMTRA: Docking



OMTRA: Docking

Evaluated on PoseBusters test set.

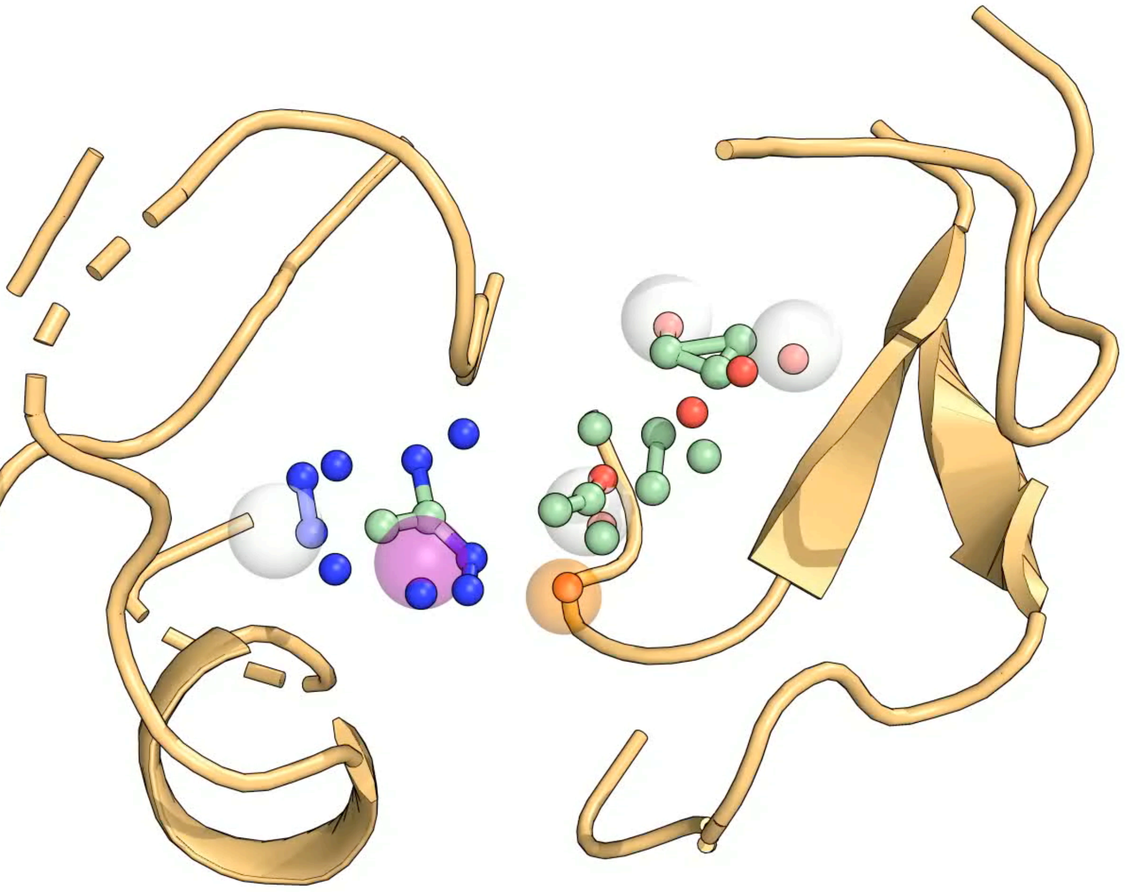


OMTRA: Transfer Learning?

Evaluated on PLINDER test set.

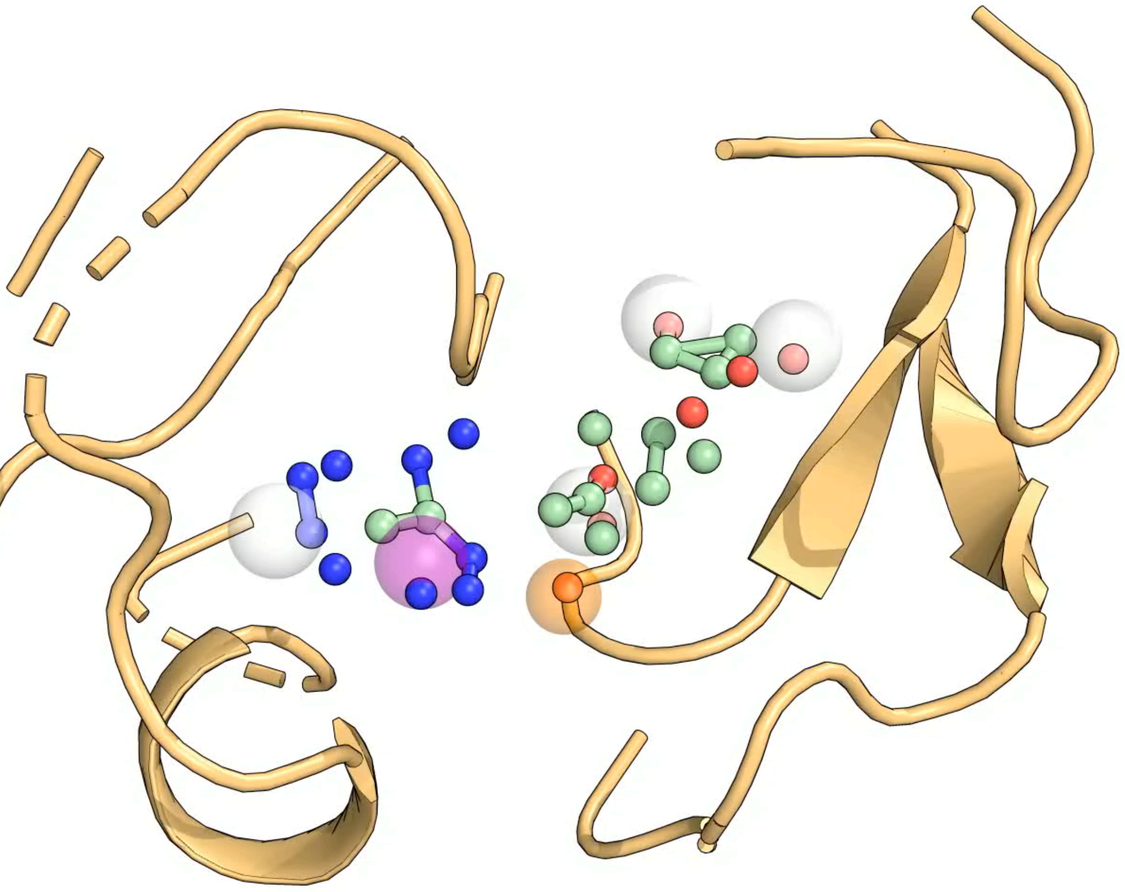
Multitask	Pretrained	*De Novo* Design			Docking	
		% Valid	%PB-Valid	strain	% $\leq$ 2Å	%PB-Valid
Data				94.0	39.2	
✗	✗	98.4	71.2	96.2	92.0	75.0
✗	✓	96.6	67.4	125.6	91.0	73.0
✓	✓	97.6	69.7	69.6	95.0	75.0

OMTRA: Pharmacophore Conditioning



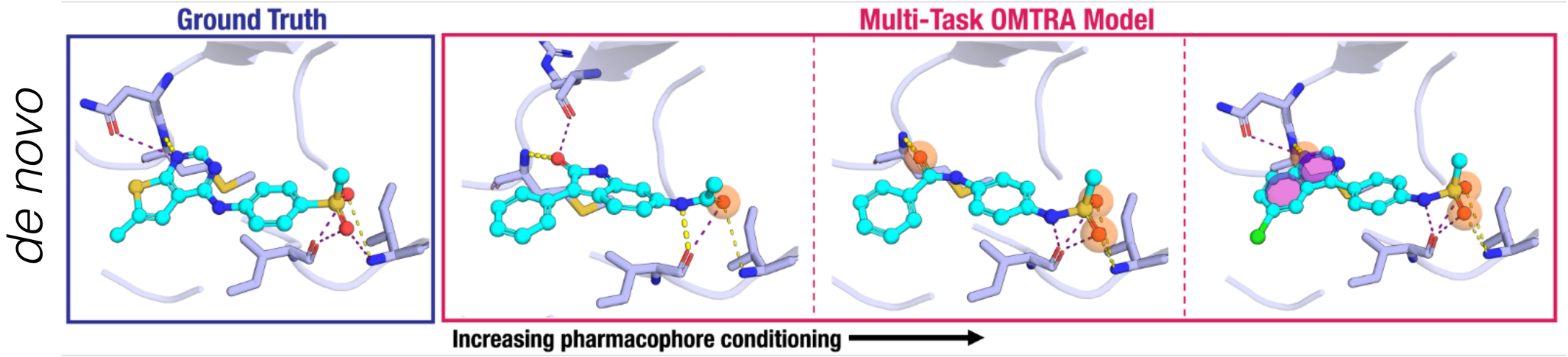
		Prot Conditioning	Prot + Pharm Conditioning
de novo design	%PB-Valid	67.5	66.2
	interaction recovery	51.0	67.4
	% Pharm Matches	-	96.9
docking	% RMSD $\leq$ 2Å	93.0	99.0
	%PB-Valid	73.0	81.0
	% Pharm Matches	-	99.5

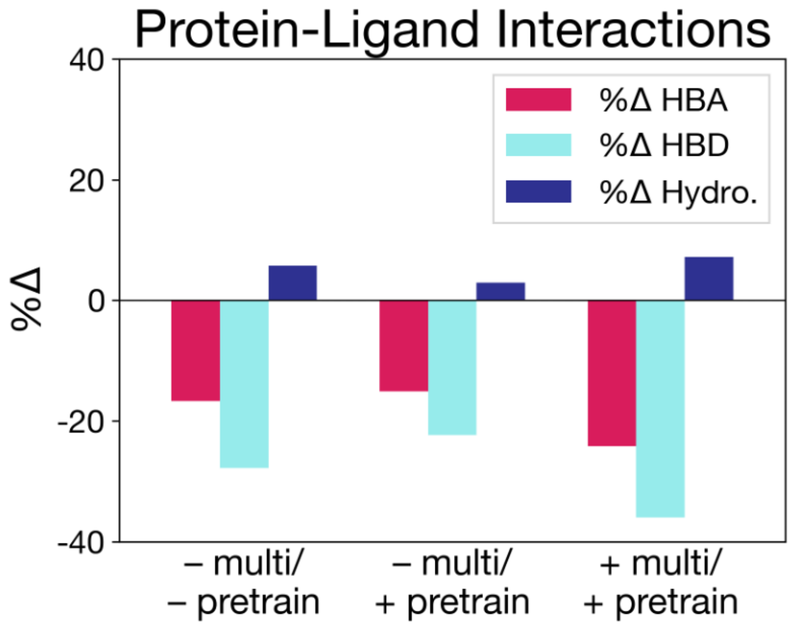
OMTRA: Pharmacophore Conditioning

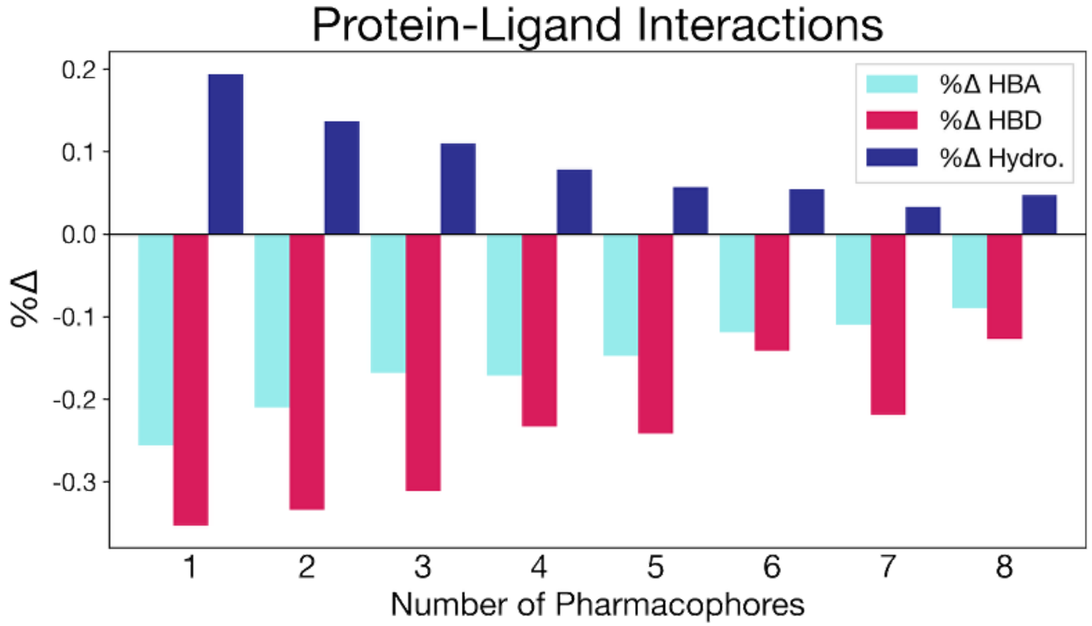


		Prot Conditioning	Prot + Pharm Conditioning
de novo design	%PB-Valid	67.5	66.2
	interaction recovery	51.0	67.4
	% Pharm Matches	-	96.9
docking	% RMSD $\leq$ 2Å	93.0	99.0
	%PB-Valid	73.0	81.0
	% Pharm Matches	-	99.5

OMTRA: Pharmacophore Conditioning

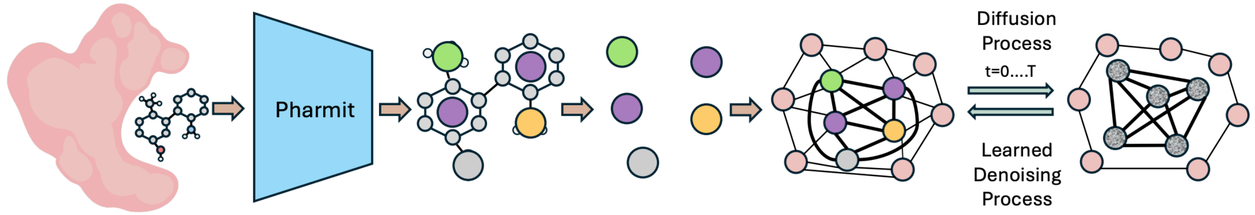


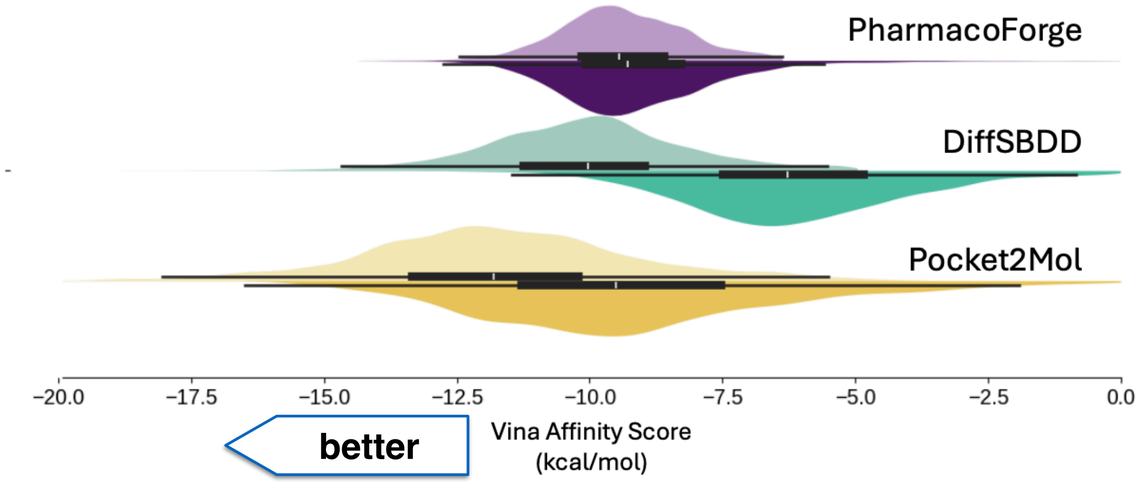




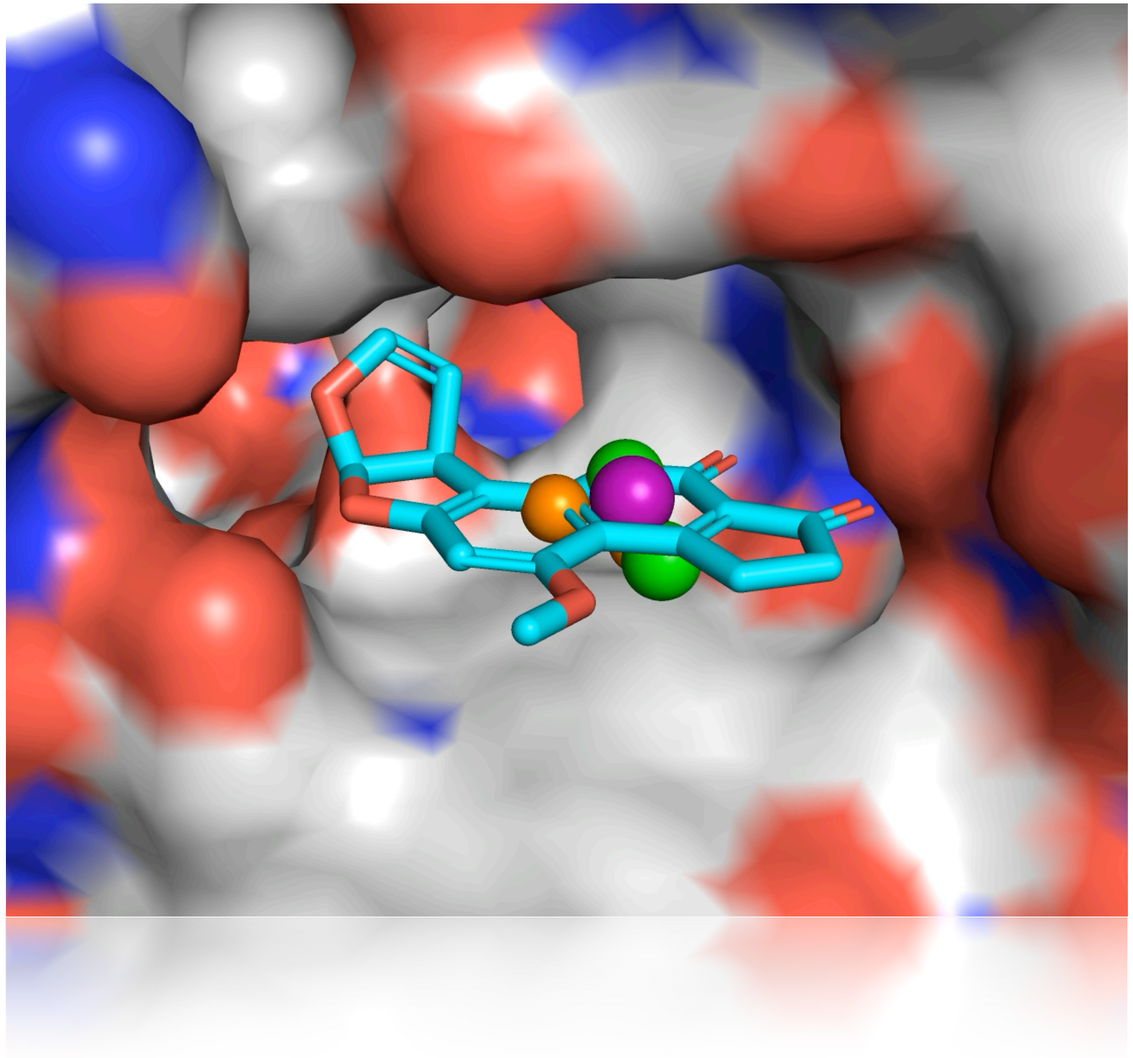
Generating *Descriptions* of Molecules

PharmacoForge







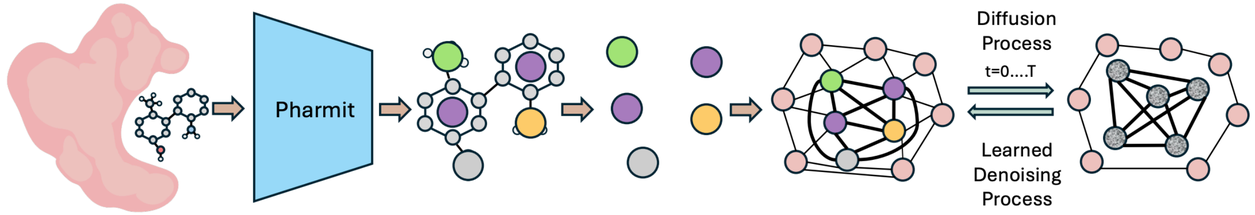



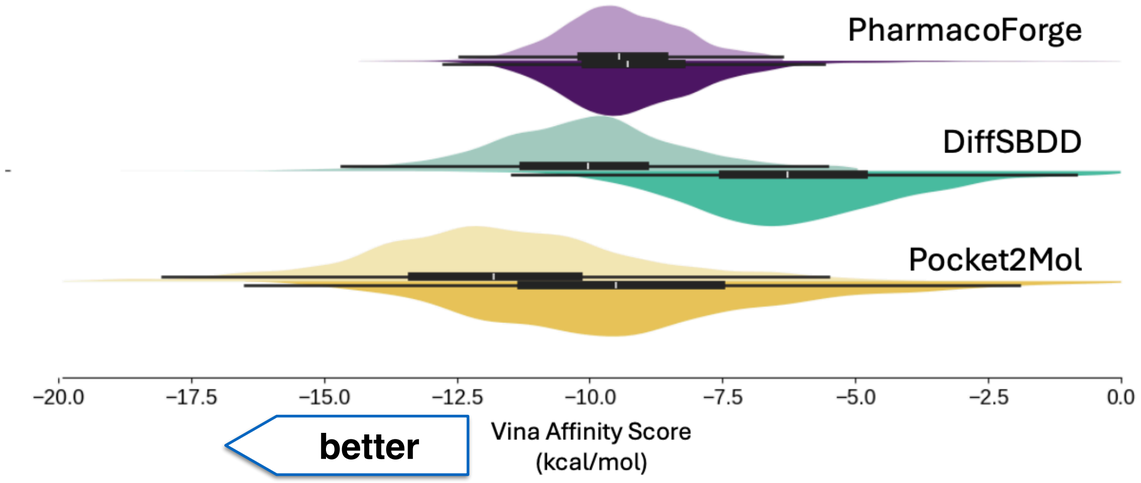
PharmacoForge: pharmacophore generation with diffusion models

Emma L. Flynn[1,2], Riya Shah [1], Ian Dunn [1,2], Rishal Aggarwal [1,2] and David Ryan Koes [1]*

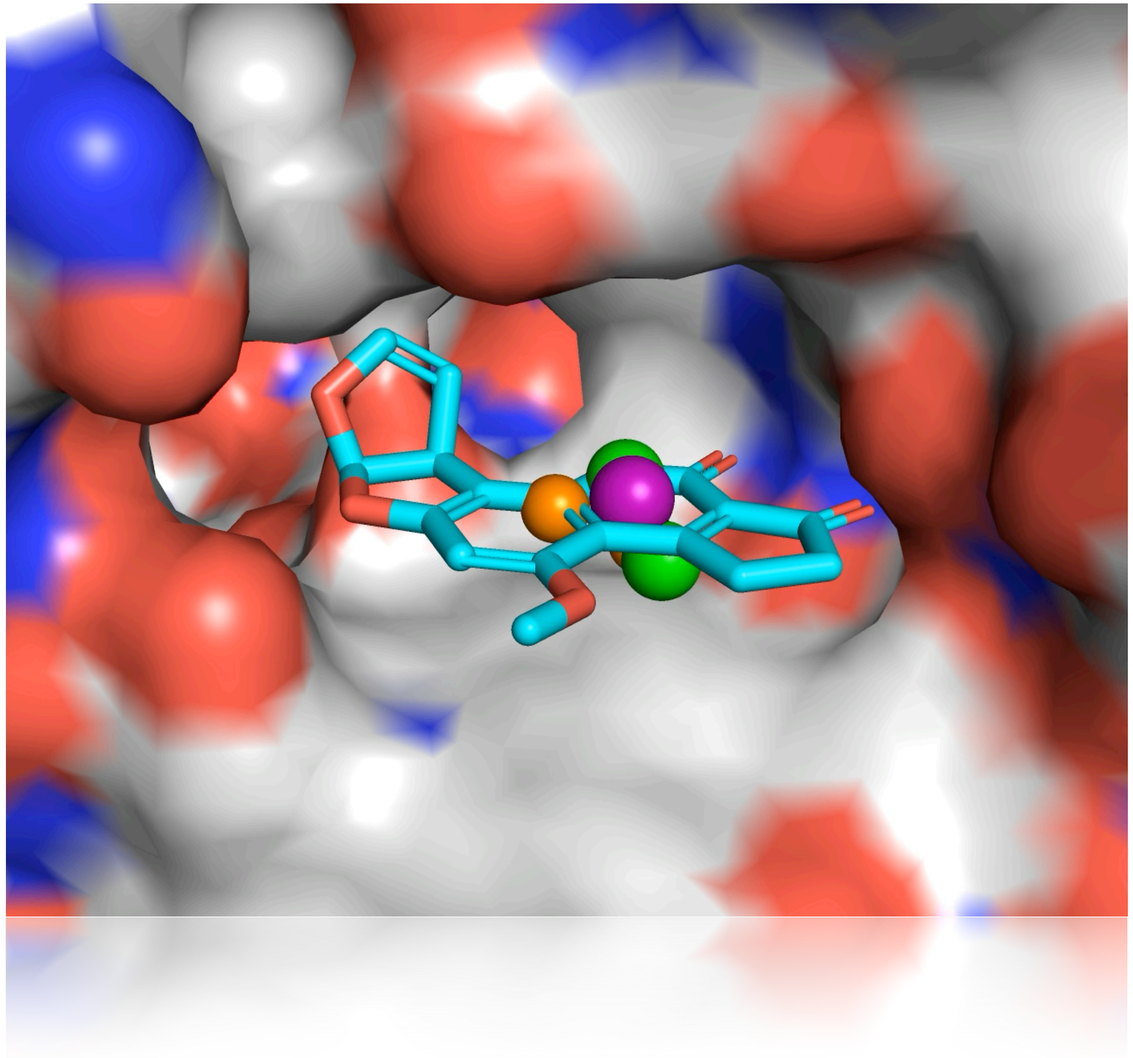
frontiers | Frontiers in Bioinformatics

PharmacoForge







PharmacoForge: pharmacophore generation with diffusion models

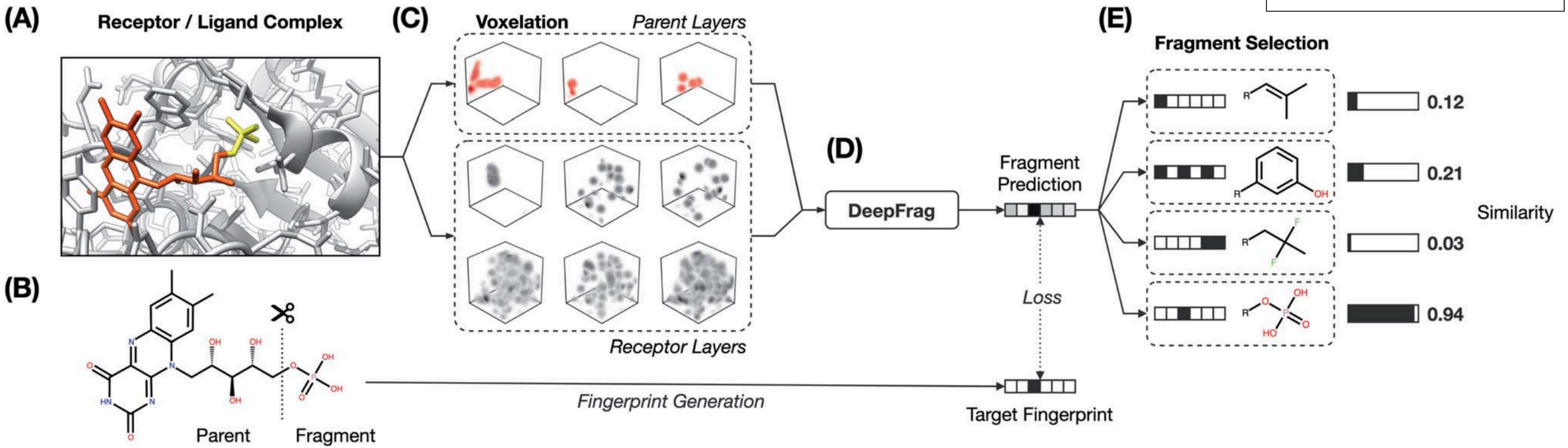
Emma L. Flynn[1,2], Riya Shah [1], Ian Dunn [1,2], Rishal Aggarwal [1,2] and David Ryan Koes [1]*

frontiers | Frontiers in Bioinformatics

DeepFrag: a deep convolutional neural network for fragment-based lead optimization†

Harrison Green, [a] David R. Koes[b] and Jacob D. Durrant *[a]

https://durrantlab.pitt.edu/deepfrag/

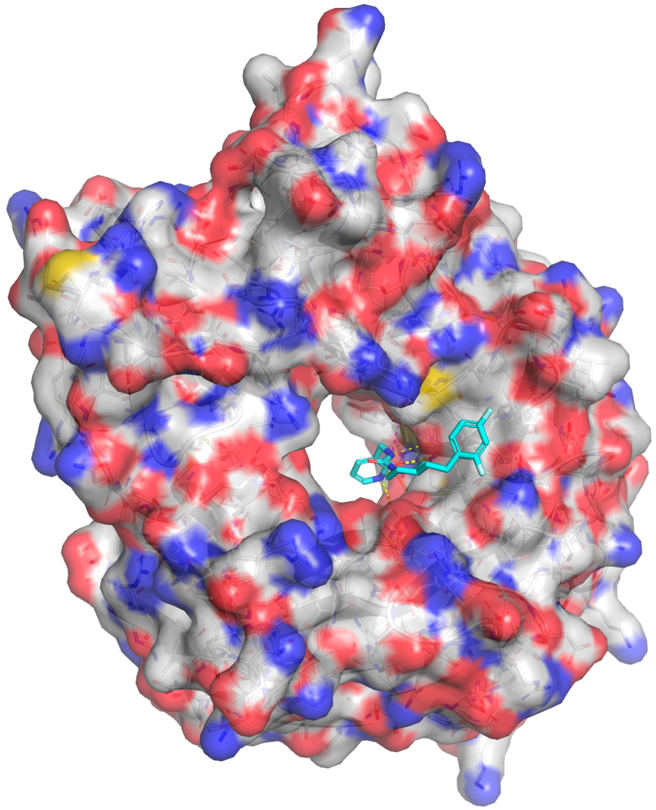


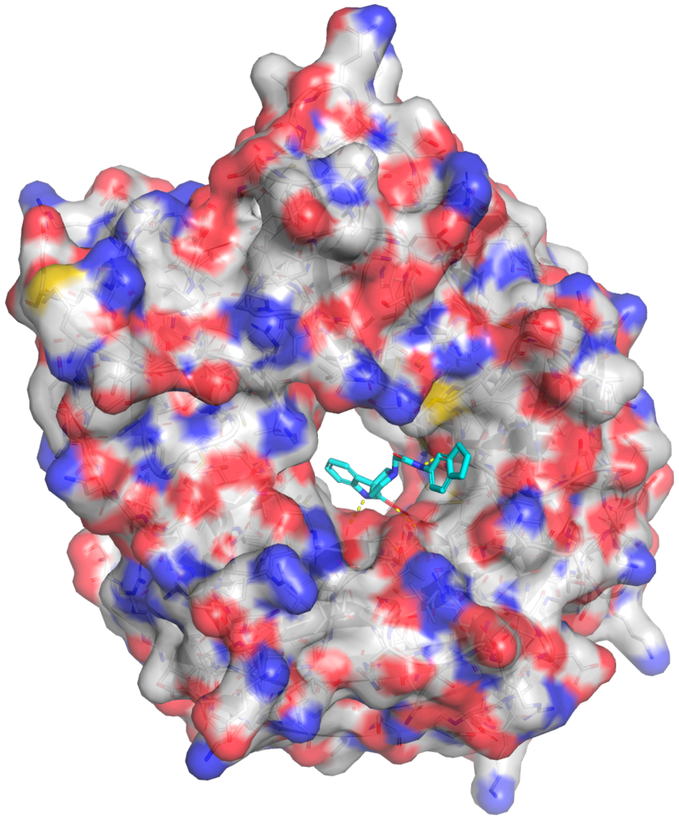
Prospective Evaluation

CACHE CHALLENGE

CRITICAL ASSESSMENT OF COMPUTATIONAL HIT-FINDING EXPERIMENTS

CACHE Challenge #1

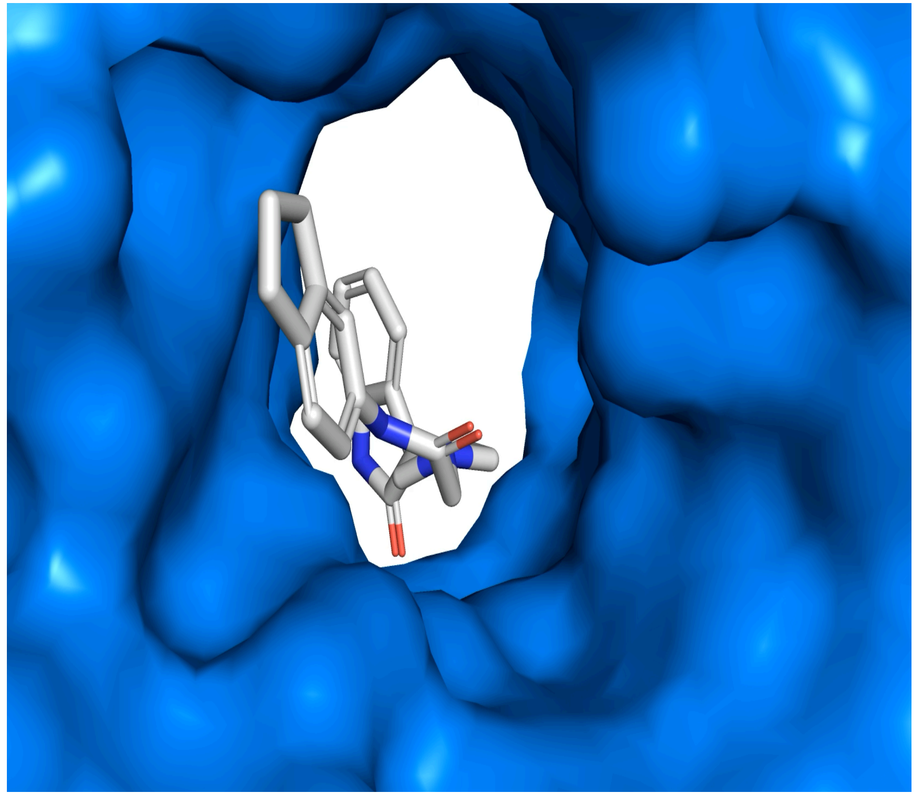




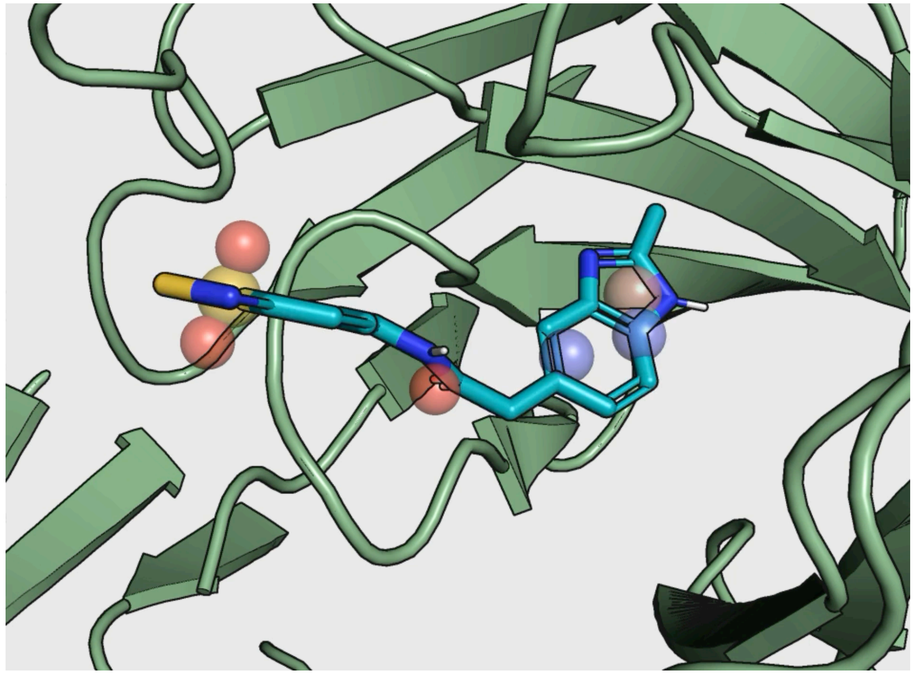
Ian Dunn

A Tale of Two Methods

Large-Scale Docking with GNINA

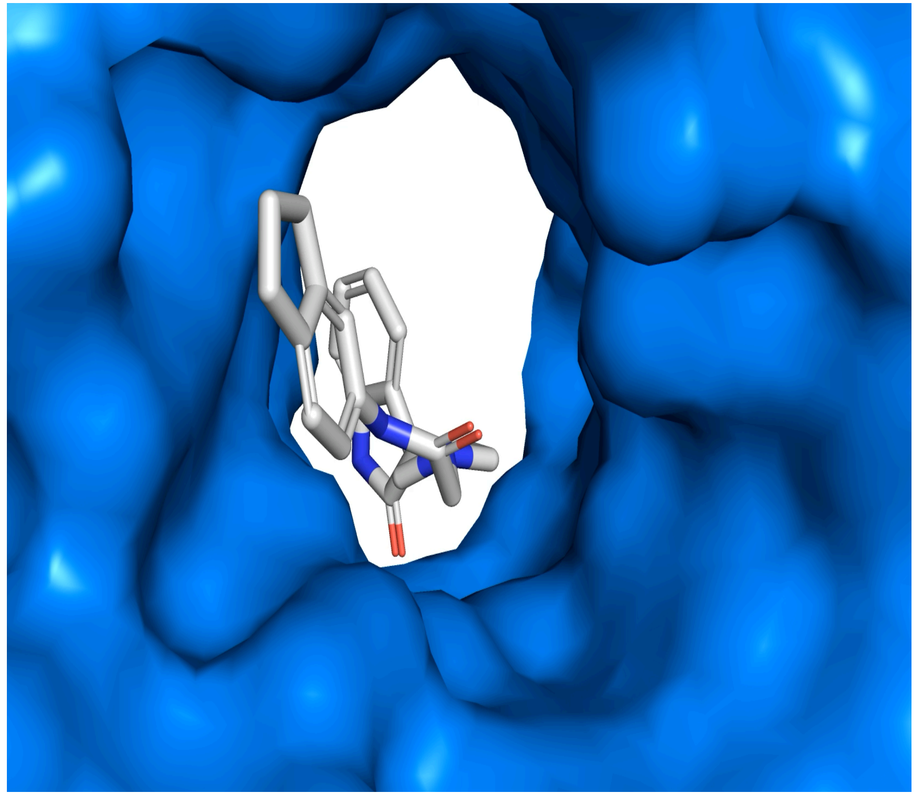


Pharmacophore Screening with Pharmit

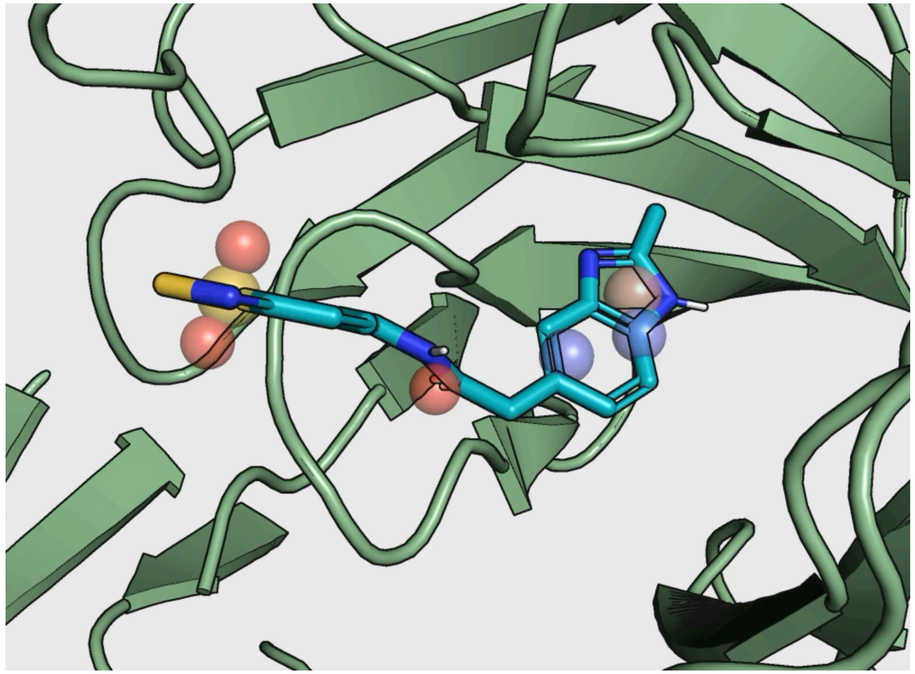


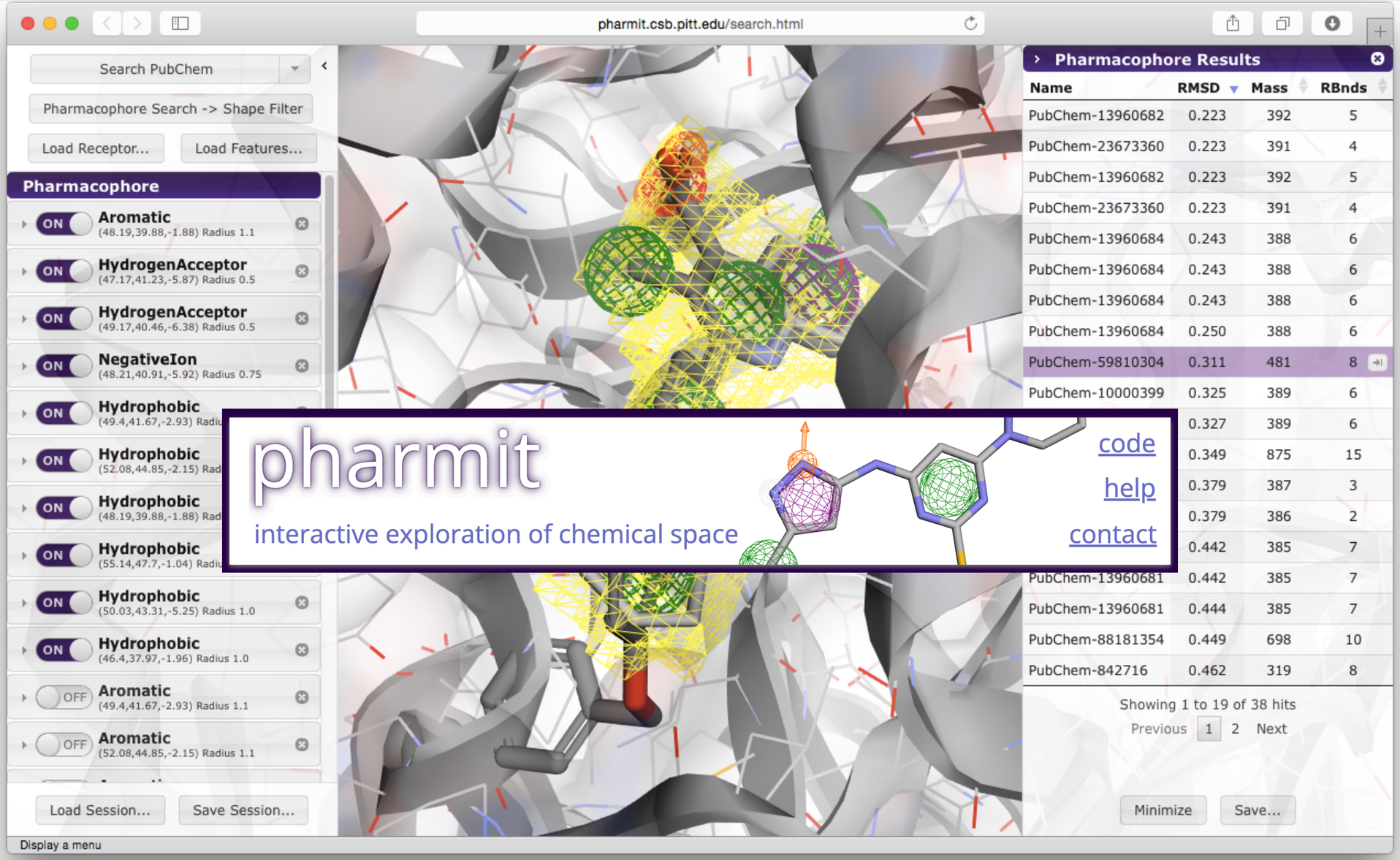
A Tale of Two Methods

Large-Scale Docking with GNINA



Pharmacophore Screening with Pharmit



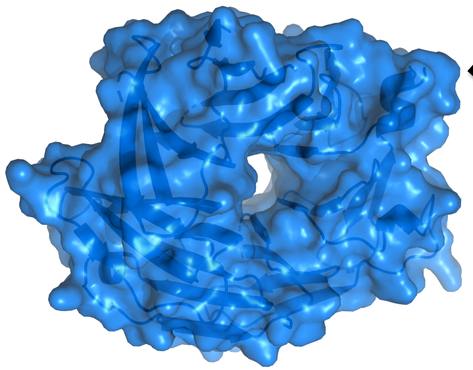


http://pharmit.csb.pitt.edu/

High-throughput Docking Pipeline

molport

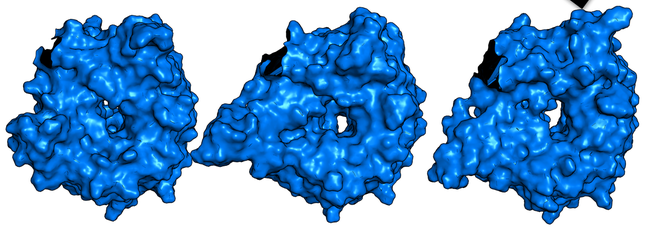
~7 million molecules



Crystal Structure

Docking with GNINA

Top 1k Molecules by GNINA score

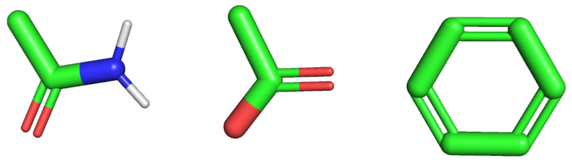


MD Ensemble Structures

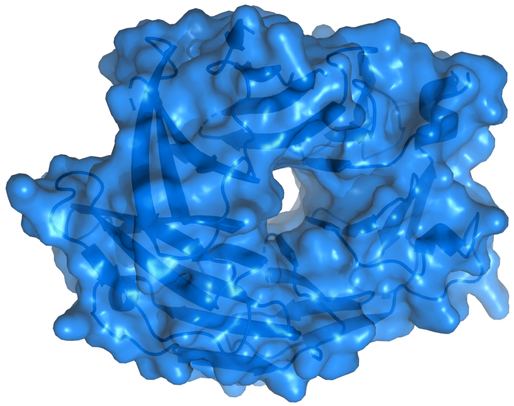
Docking with GNINA

1000 molecules, + GNINA Scores

Pharmacophore Generation via Fragment Docking

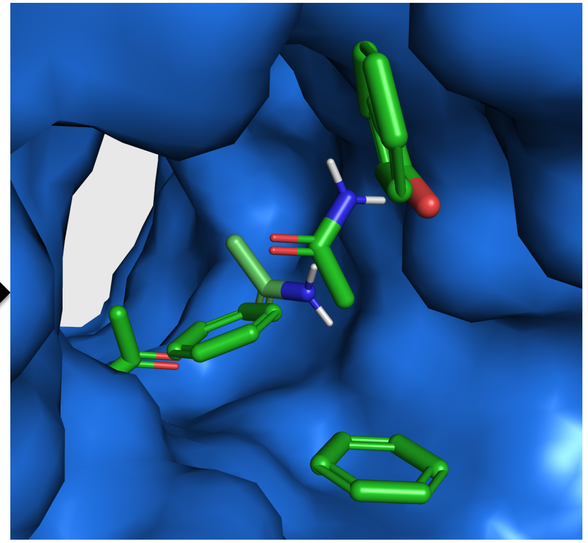


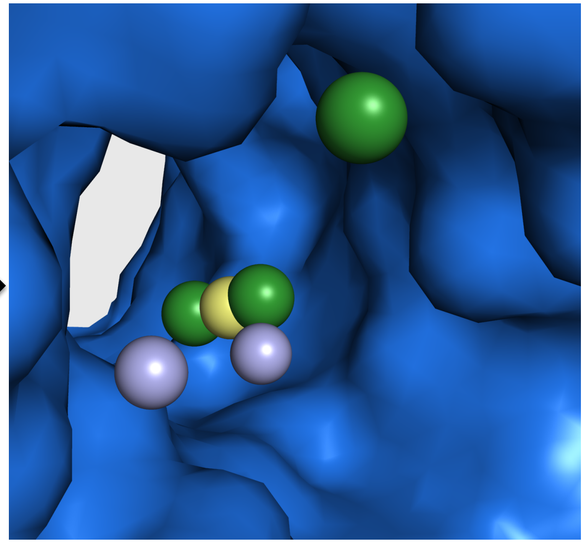
Molecule Fragments



Crystal Structure

Docking with GNINA

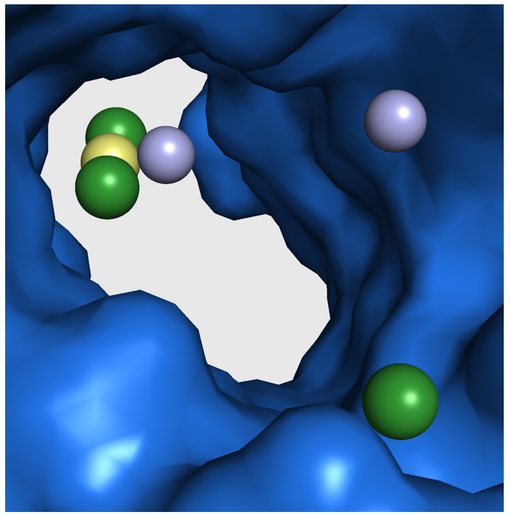




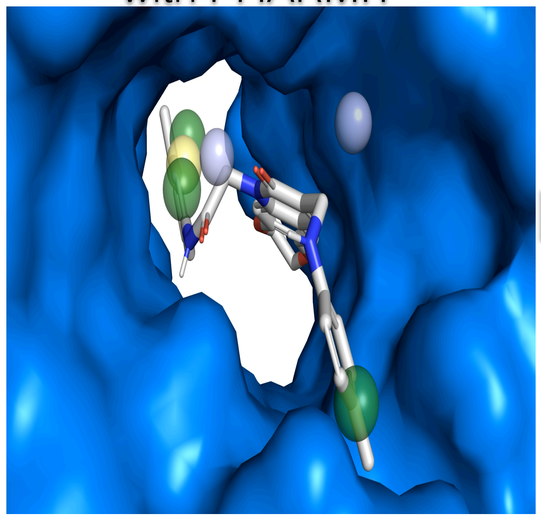
Pharmacophore Generation

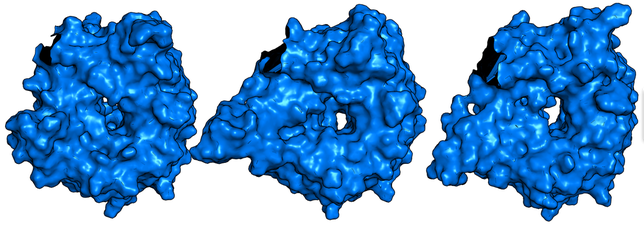
Pharmacophore Pipeline

Pharmacophore Generation



Pharmacophore Screening with PHARMIT





MD Ensemble Docking with GNINA

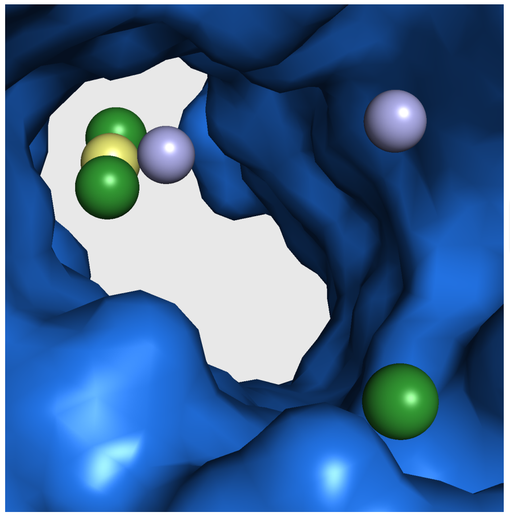
3572 molecules
+
GNINA scores

- ZINC20: 20 mil molecules
- MCULE: 45 mil molecules
- MCULE-ULTIMATE: 126 mil molecules

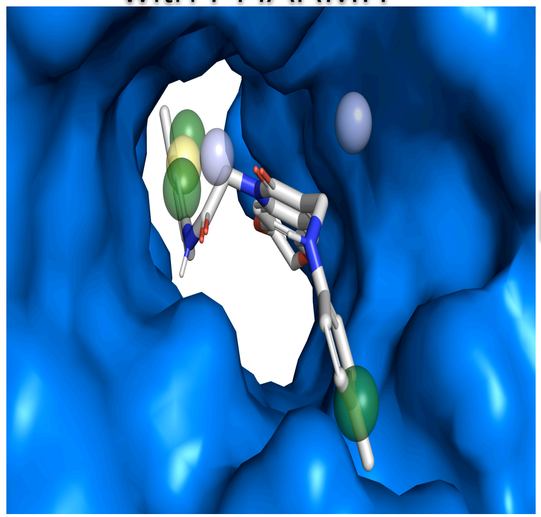
Molecule Libraries

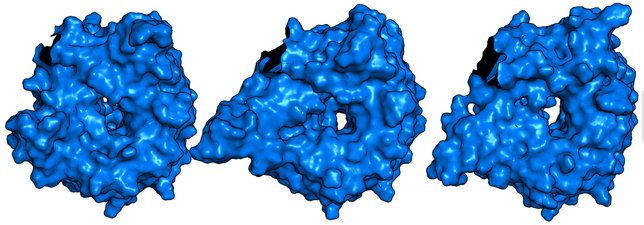
Pharmacophore Pipeline

Pharmacophore Generation



Pharmacophore Screening with PHARMIT





MD Ensemble Docking with GNINA

3572 molecules
+
GNINA scores

- ZINC20: 20 mil molecules
- MCULE: 45 mil molecules
- MCULE-ULTIMATE: 126 mil molecules

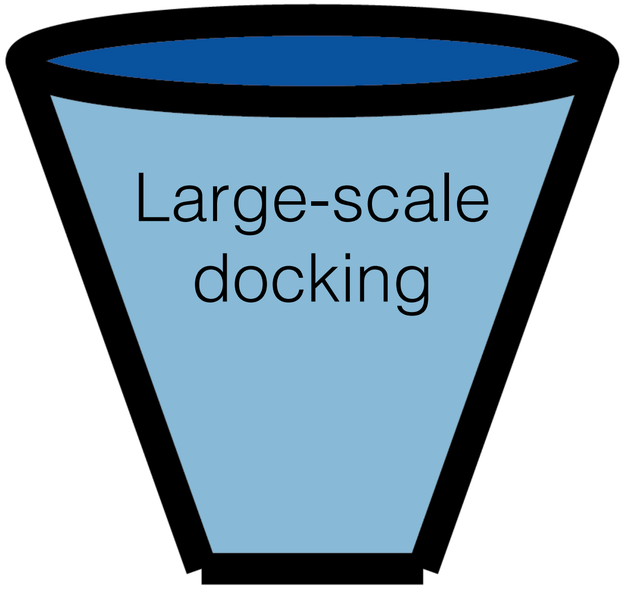
Molecule Libraries

Round 1 Submission

molport

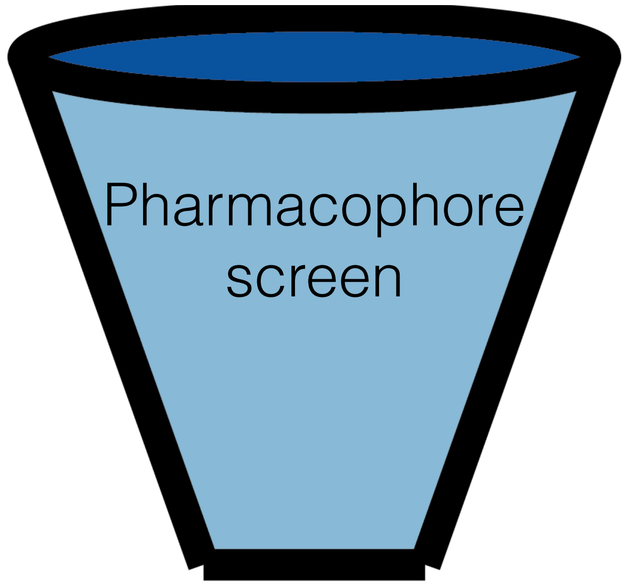
- ZINC20: 20 mil molecules
- MCULE: 45 mil molecules
- MCULE-ULTIMATE: 126 mil molecules

Molecule Libraries



2 screening methods

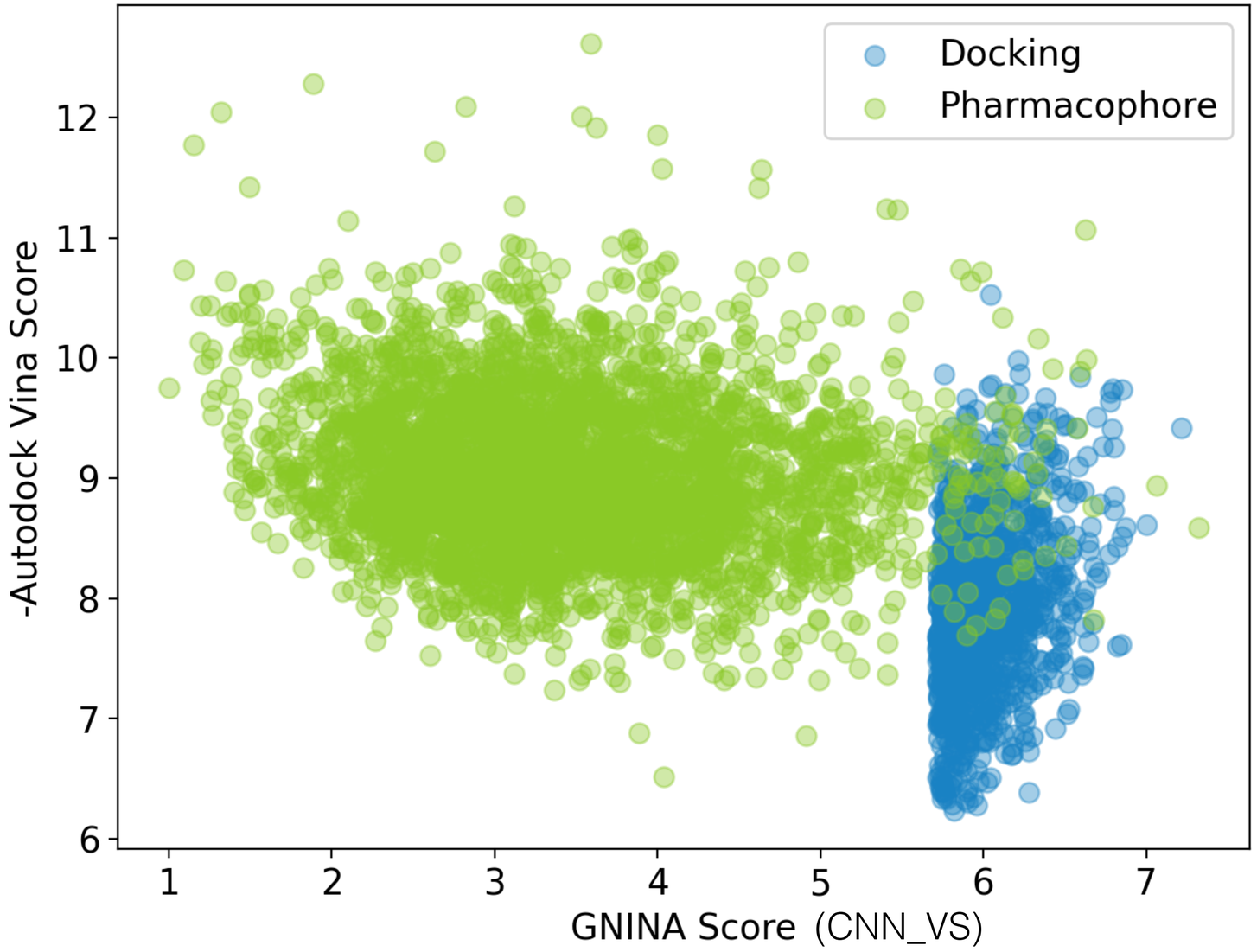
2 scoring methods



1k ligands
gnina scores
vina scores

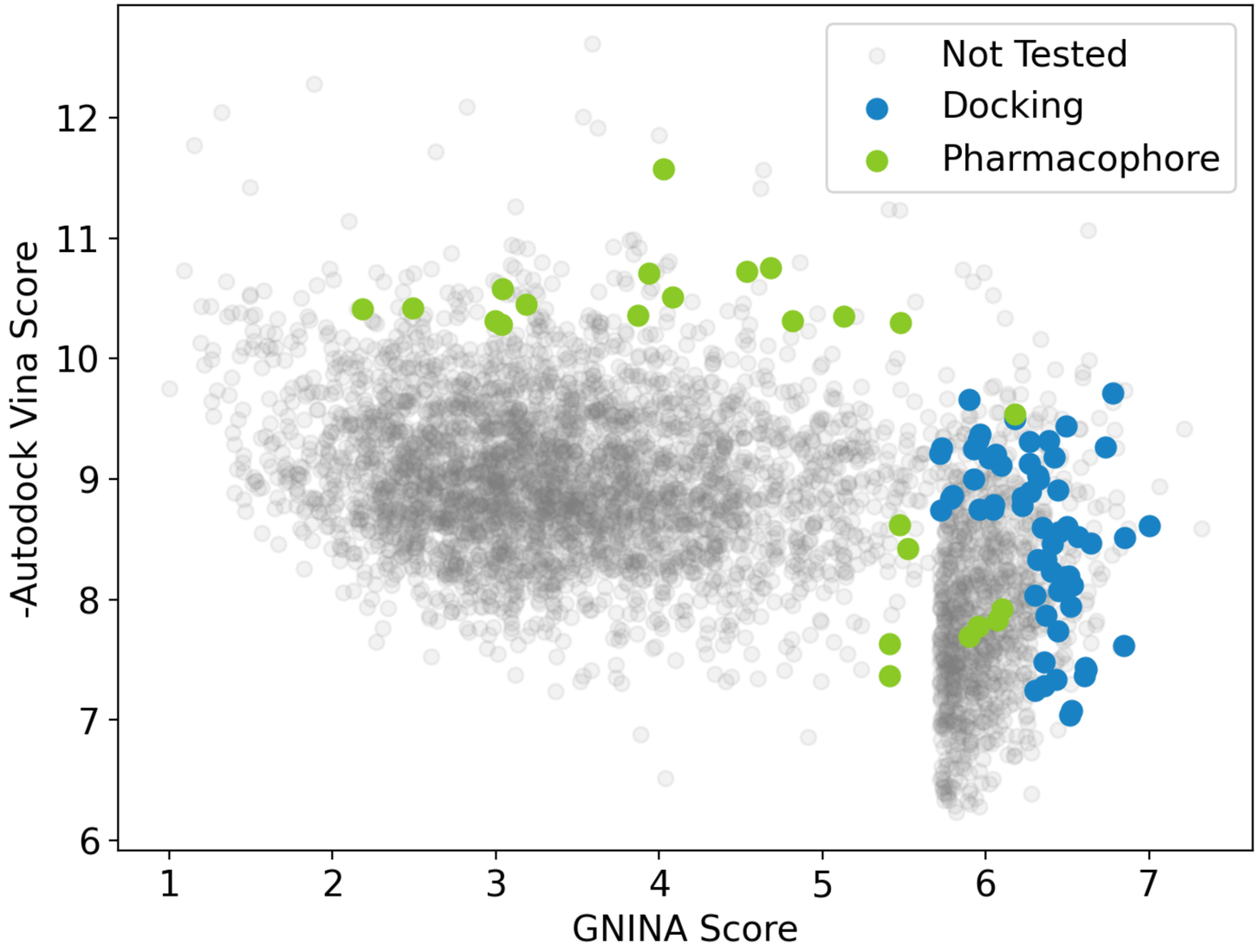
3.5k ligands
gnina scores
vina scores

Round 1 Results



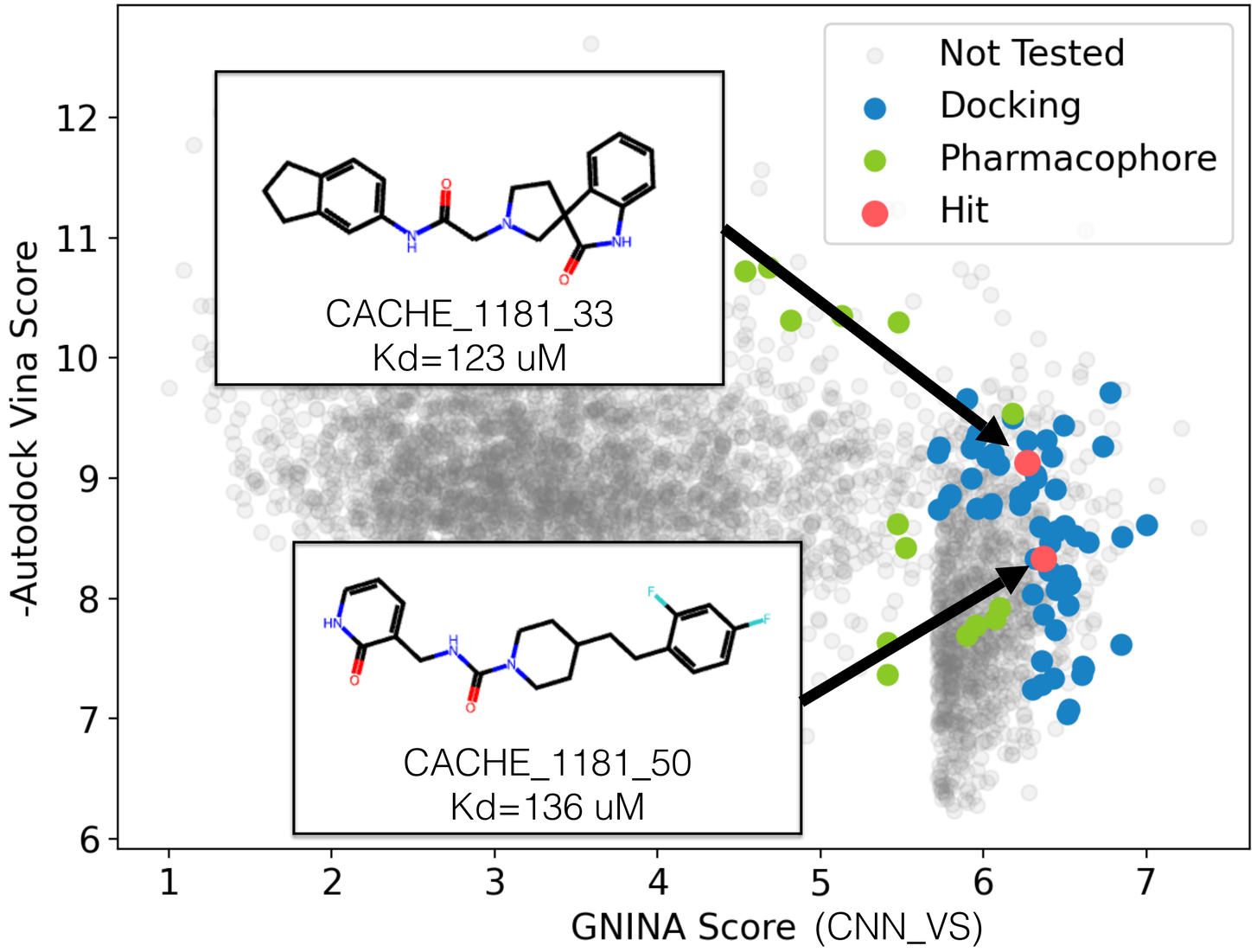
- Selection limited/ skewed by database availability
- 84 ligands tested
  - 59 from docking
  - 24 from pharm screen

Round 1 Results



- Selection limited/ skewed by database availability
- 84 ligands tested
  - 59 from docking
  - 24 from pharm screen

Round 1 Results

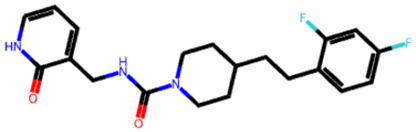


- 2/84 were hits
  - Both from docking

Round 2: Hit Optimization

Hit Optimization Pipeline

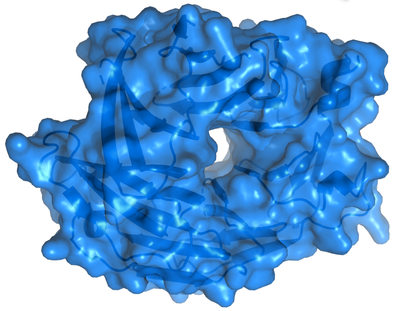
Parent Compound



Similarity screen against
Enamine REAL
Return 5000 most similar
ligands by tanimoto score

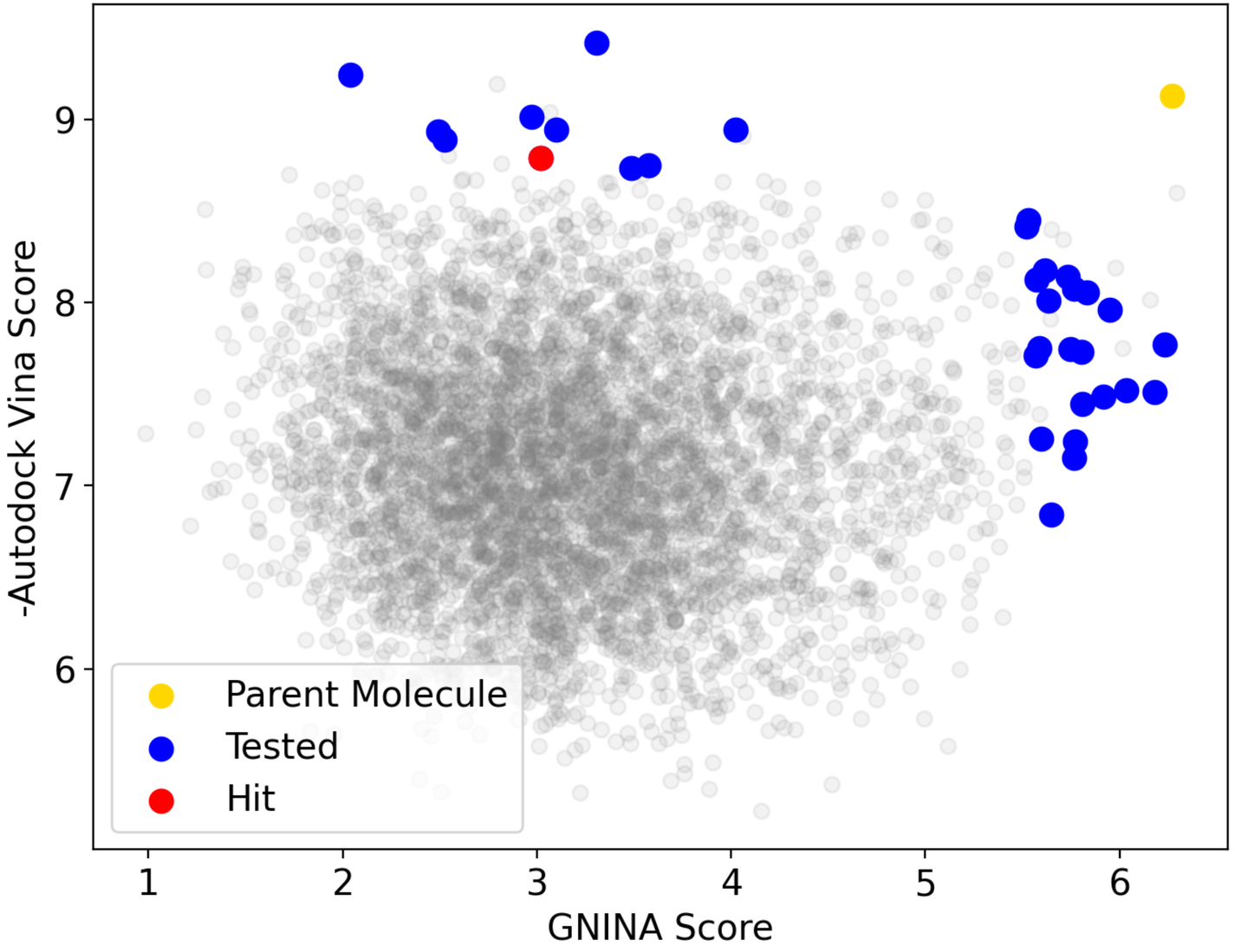
Docking with
GNINA

5000 molecules
+
GNINA scores

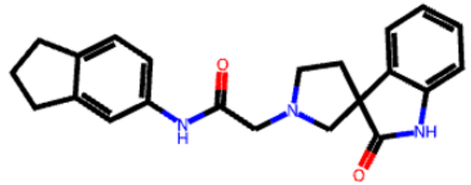


Crystal Structure

Hit Optimization Results

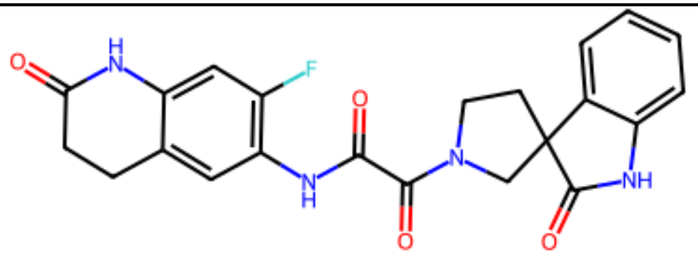


Parent Compound



CACHE_1181_33
Kd=123 uM

Hit Compound



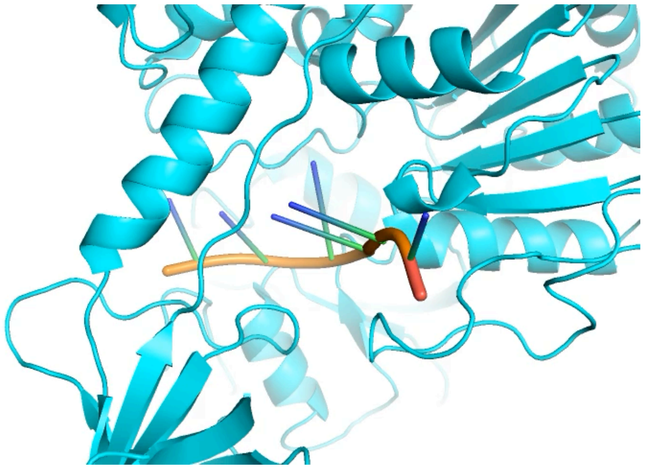
CACHE-HO_1181_24
Kd=56 uM

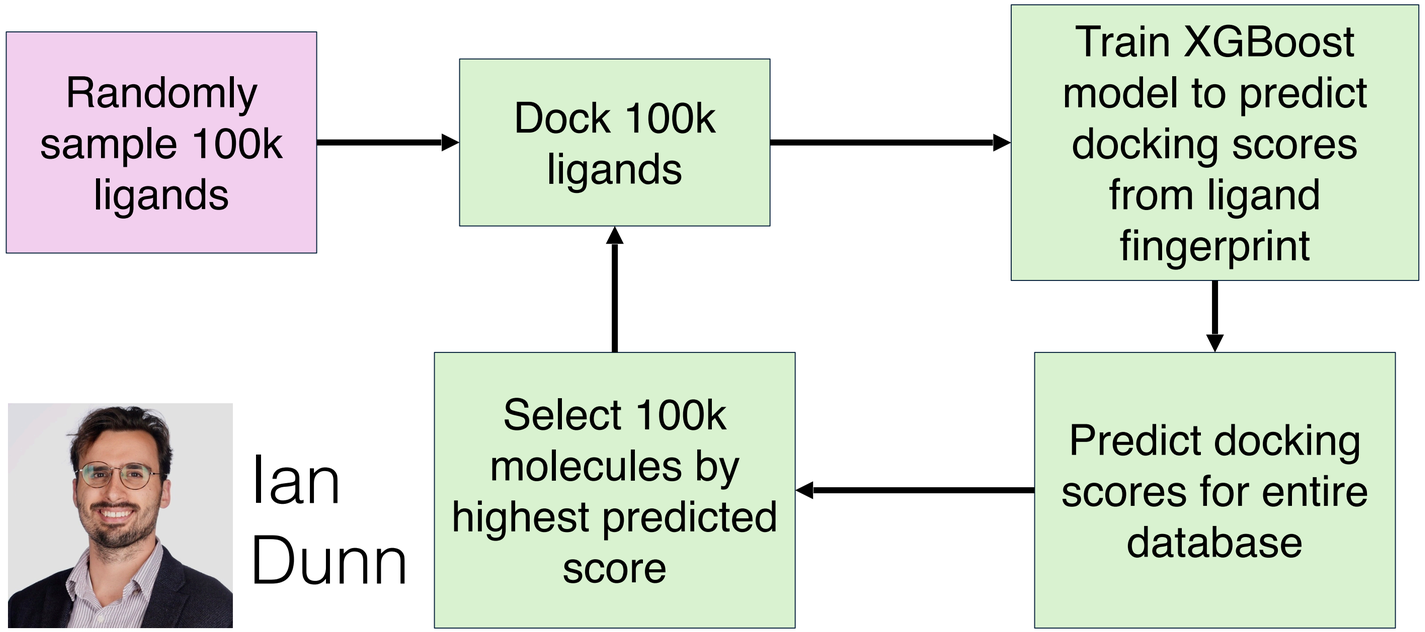
Final Results

Participant	Participant ID	Aggregated score
David Koes, University of Pittsburgh	1181	18
Olexandr Isayev & Maria Kurnikova, Carnegie Mellon University & Artem Cherkasov, University of British Columbia	1209	18
Christina Schindler, Merck KGaA	1193	17
Dmitri Kireev, University of Missouri	1183	16
Christoph Gorgulla, St. Jude Children's Research Hospital and Harvard University	1195	16
Didier Rognan, Université Strasbourg	1202	16
Pavel Polishchuk, Palacky University	1210	16
Kam Zhang, Centre for Biosystems Dynamic Research, RIKEN	1188	15
Shuangjia Zheng, Shanghai Jiao Tong University (previously Galixir)	1187	14
Carlos Zepeda, Treventis/UHN	1200	14
Fabian Liessmann, Leipzig University	1201	14
	1179	13

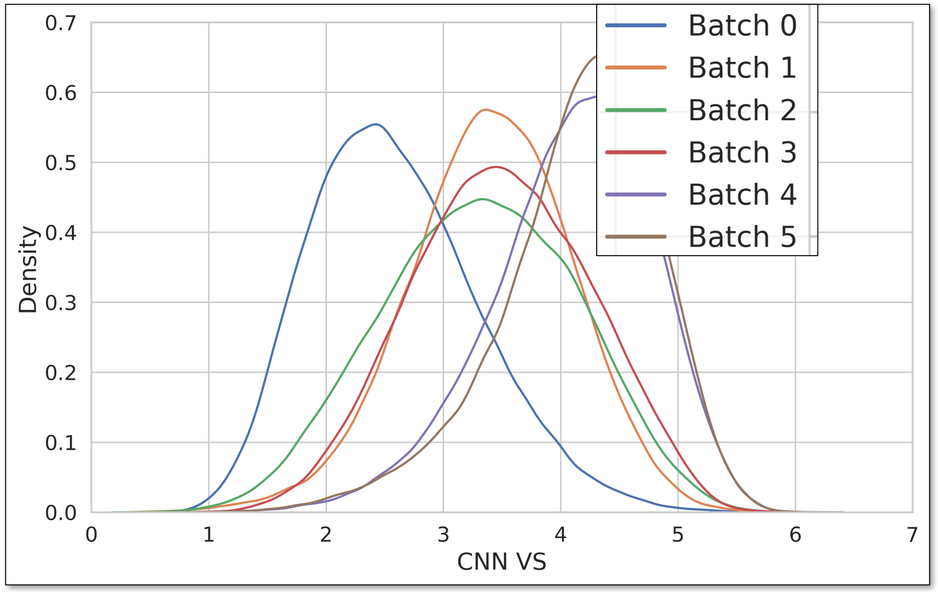
CACHE Challenge #2

- RNA binding site of SARS-COV2 NSP13
- “Deep Docking” of Enamine (4B)



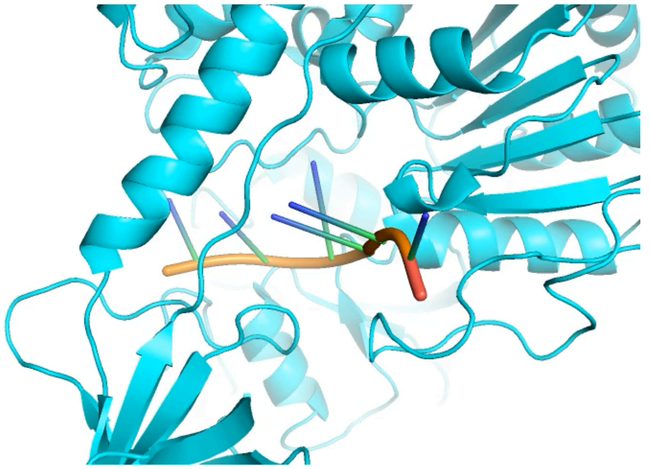


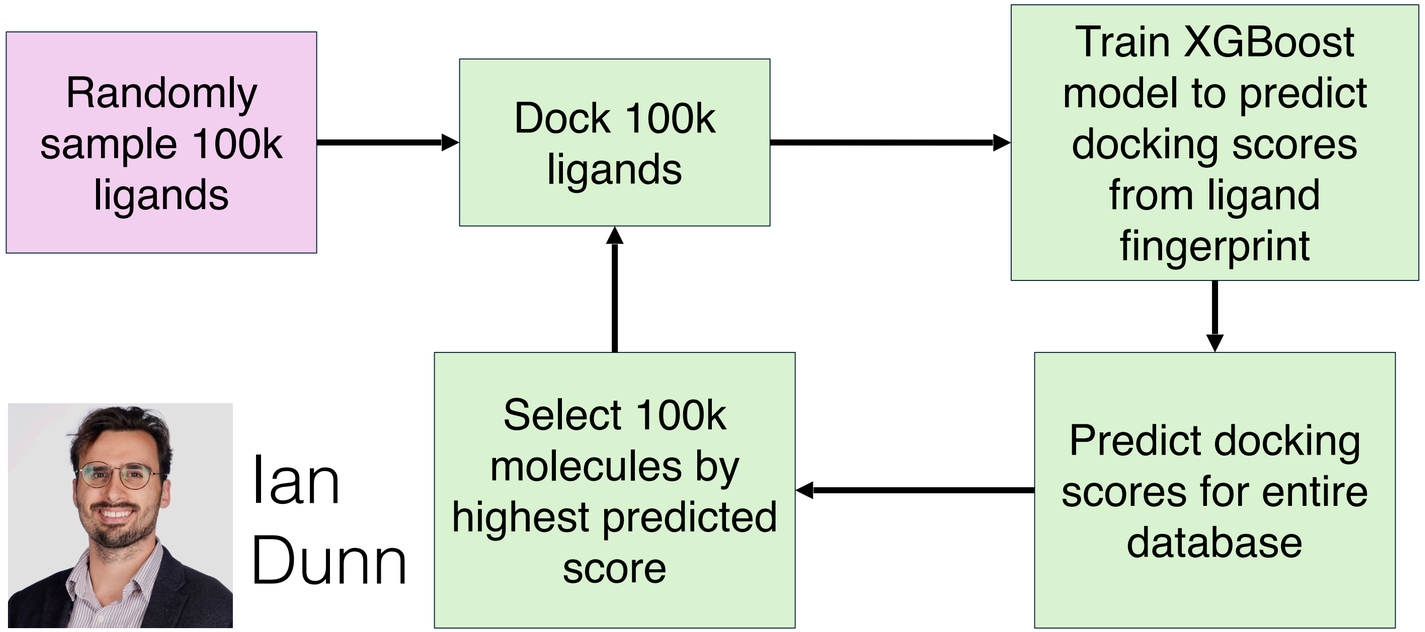
Ian Dunn



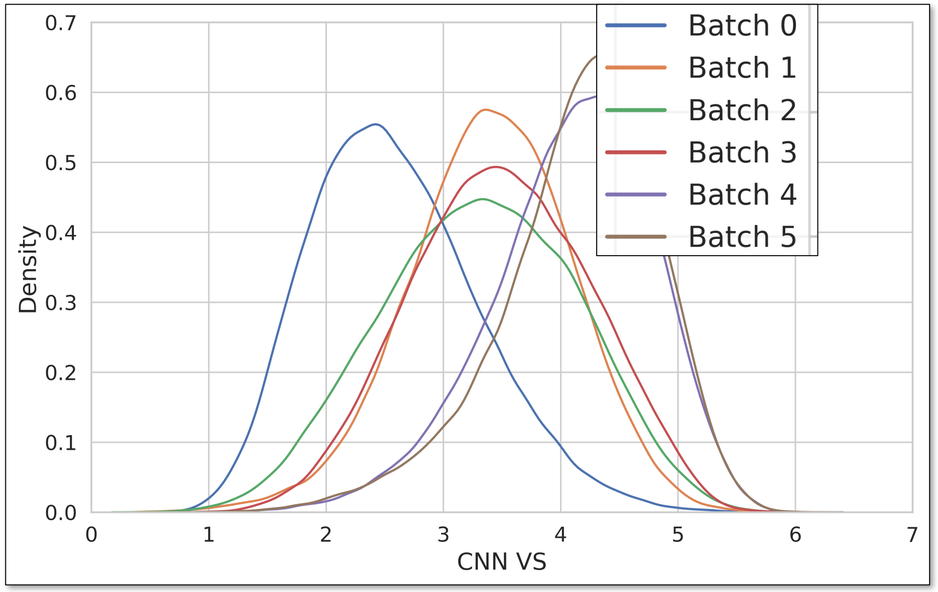
CACHE Challenge #2

- RNA binding site of SARS-COV2 NSP13
- “Deep Docking” of Enamine (4B)





Ian Dunn



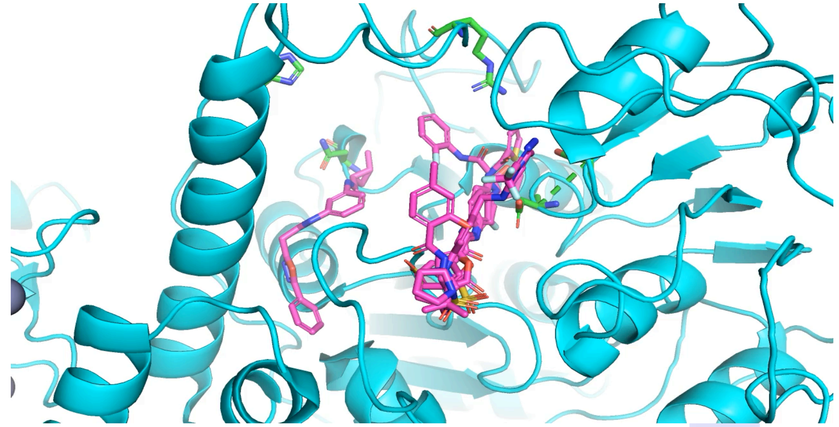
CACHE #2 Results

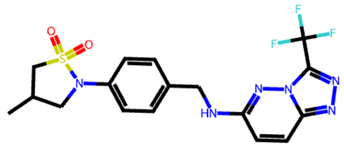
5/50 compounds identified as potential hits

>2x the average hit rate

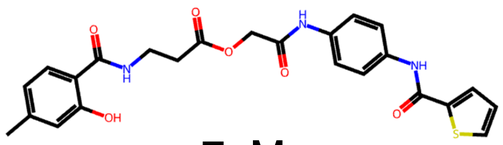
4/5 hits from last round of active learning

Highest affinity round 1 hit in the competition

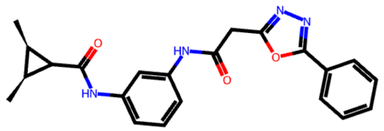




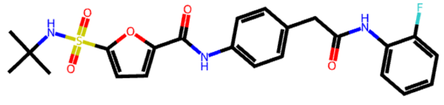
43μM



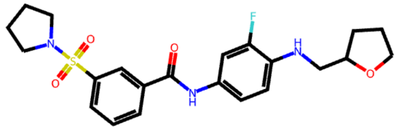
7μM



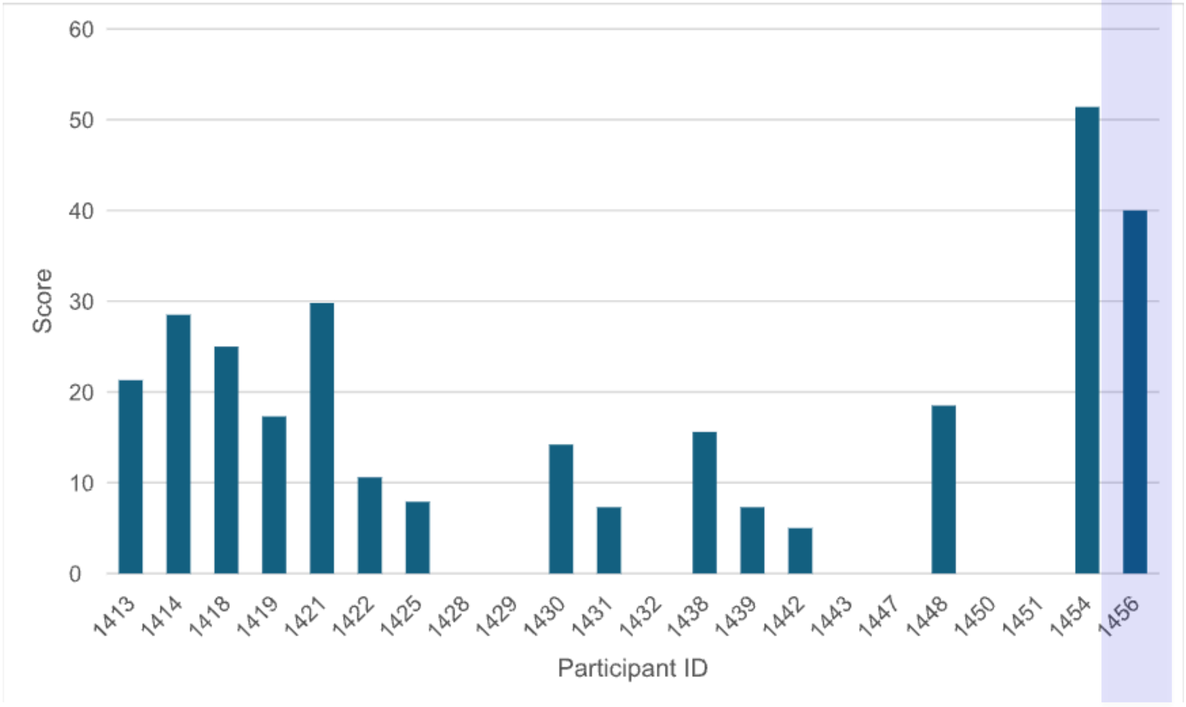
5μM



45μM



96μM



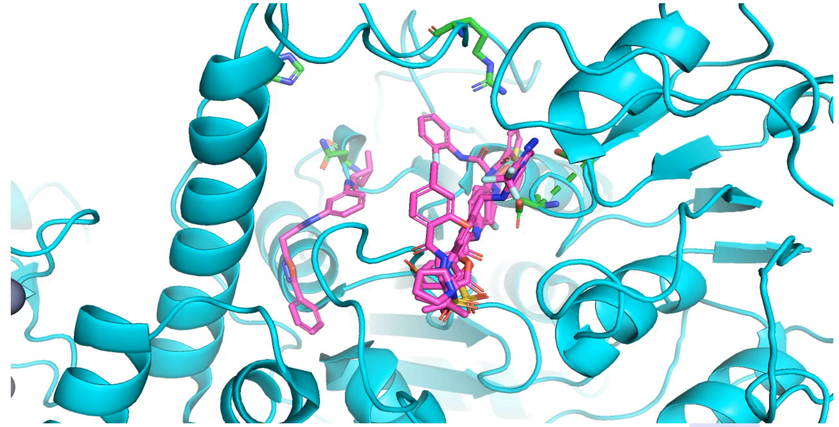
CACHE #2 Results

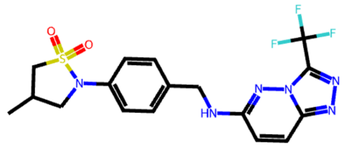
5/50 compounds identified as potential hits

>2x the average hit rate

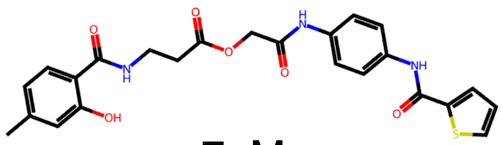
4/5 hits from last round of active learning

Highest affinity round 1 hit in the competition

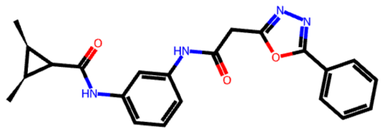




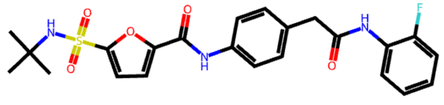
43μM



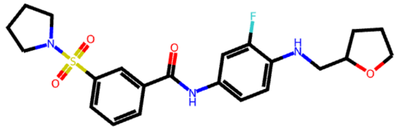
7μM



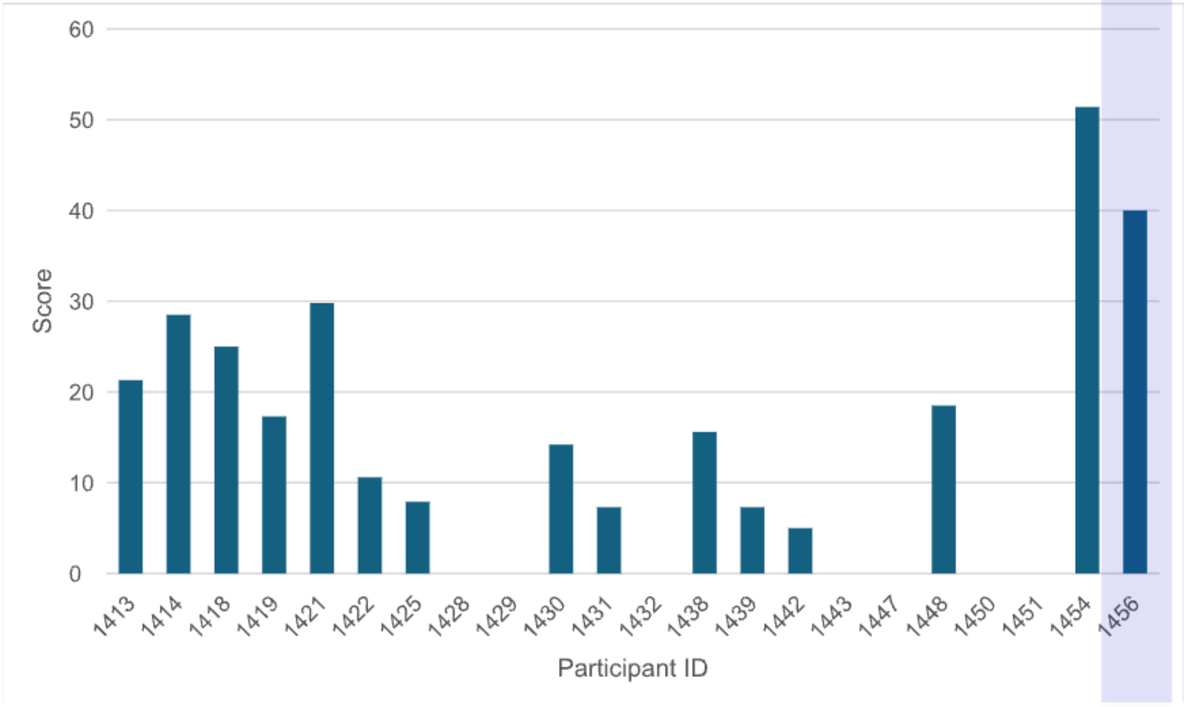
5μM



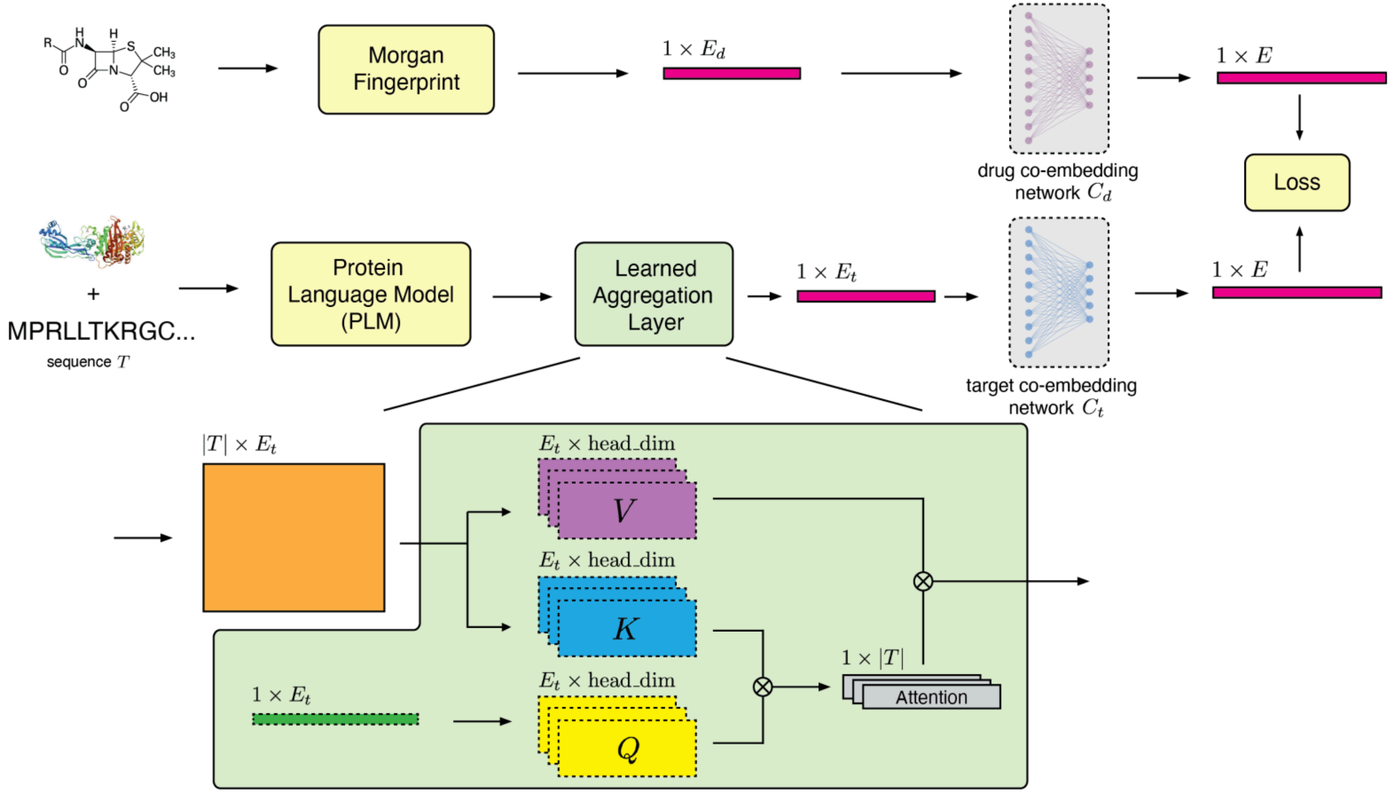
45μM



96μM



Scalable Screening of Ultra Large Libraries



Learn *co-embedding* of ligands and proteins

Ligands are "close" to proteins they bind to

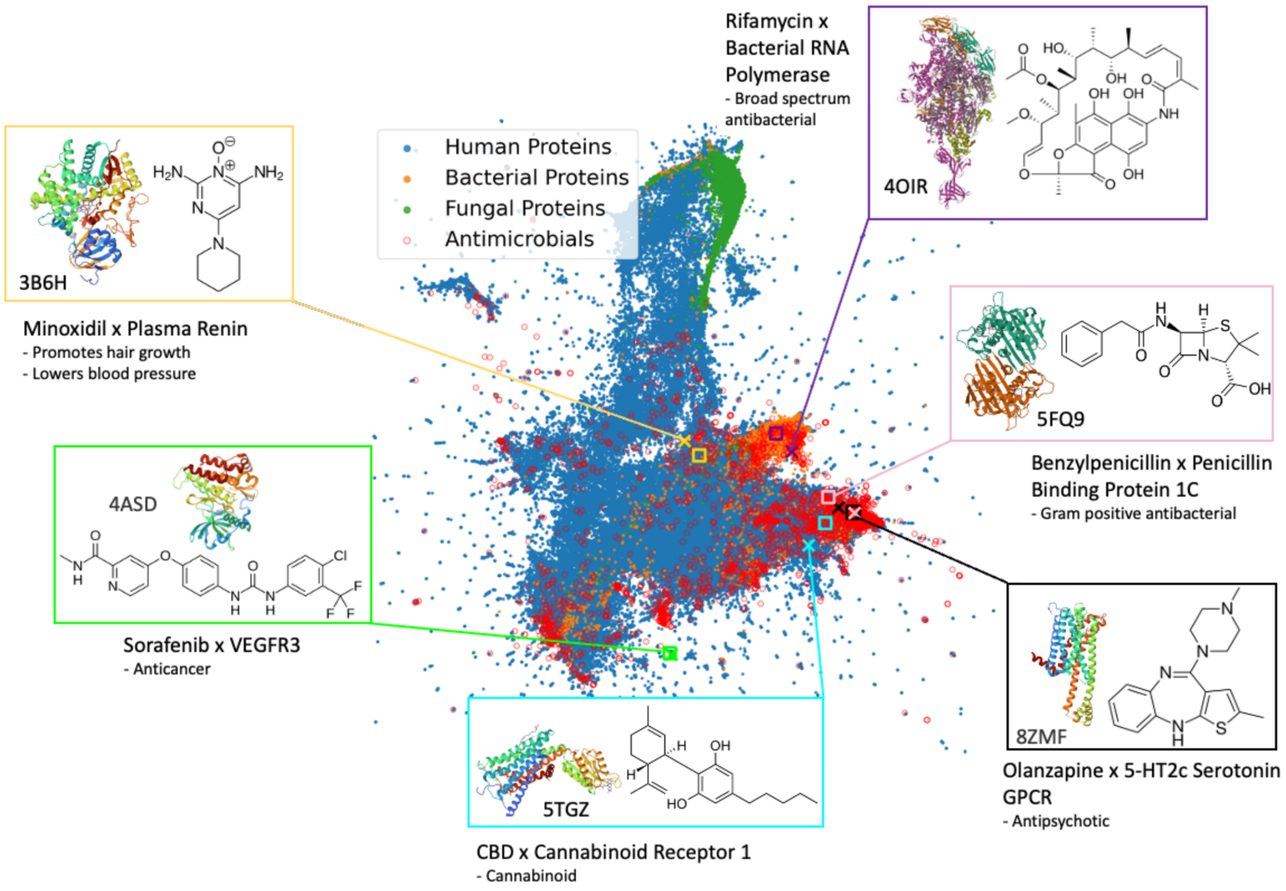
Quantitative Biology > Biomolecules

[Submitted on 23 Nov 2024 (v1), last revised 20 Jan 2025 (this version, v2)]

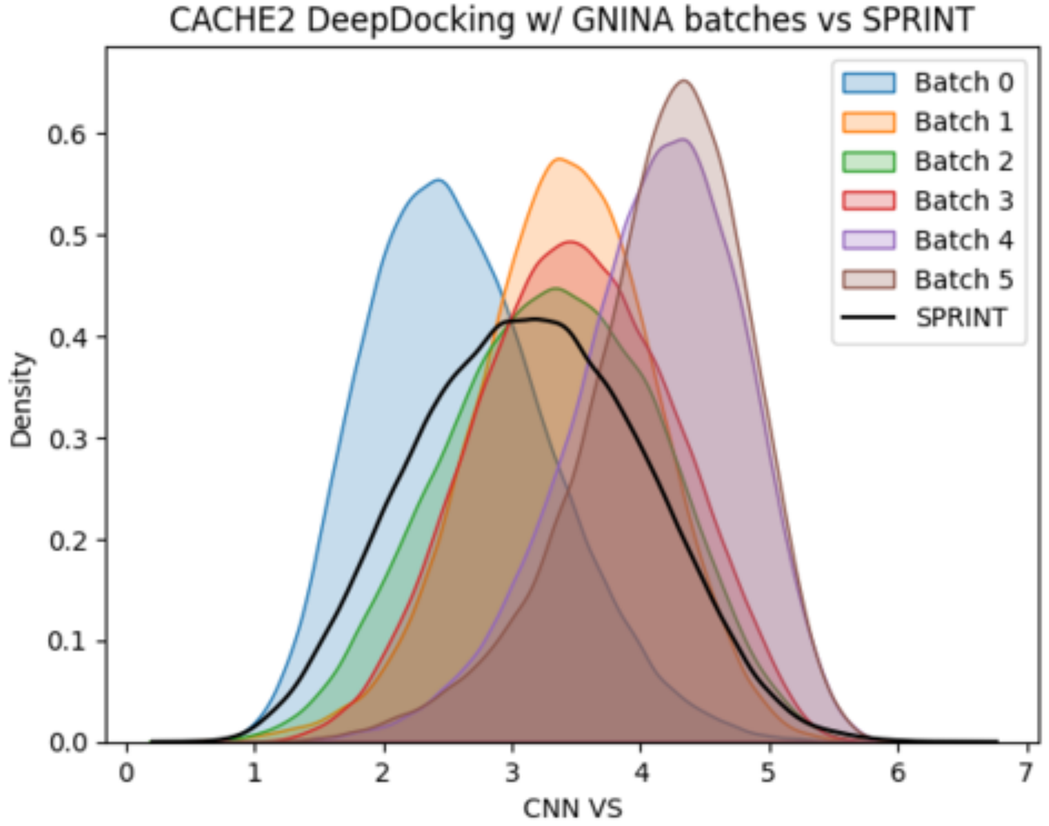
Scaling Structure Aware Virtual Screening to Billions of Molecules with SPRINT

Andrew T. McNutt, Abhinav K. Adduri, Caleb N. Ellington, Monica T. Dayao, Eric P. Xing, Hosein Mohimani, David R. Koes

Enhanced Deep Docking with SPRINT

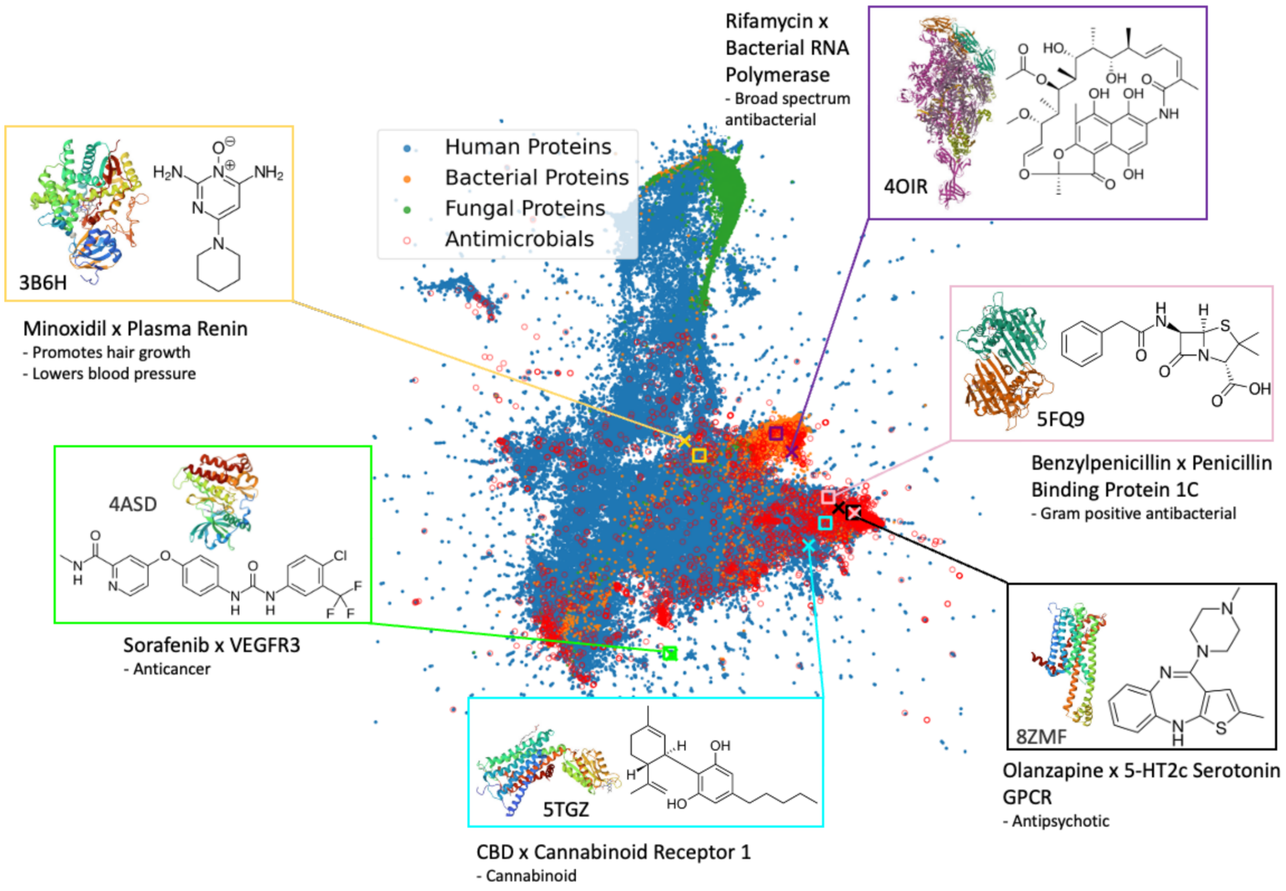


Retrospective CACHE2 Deep Docking

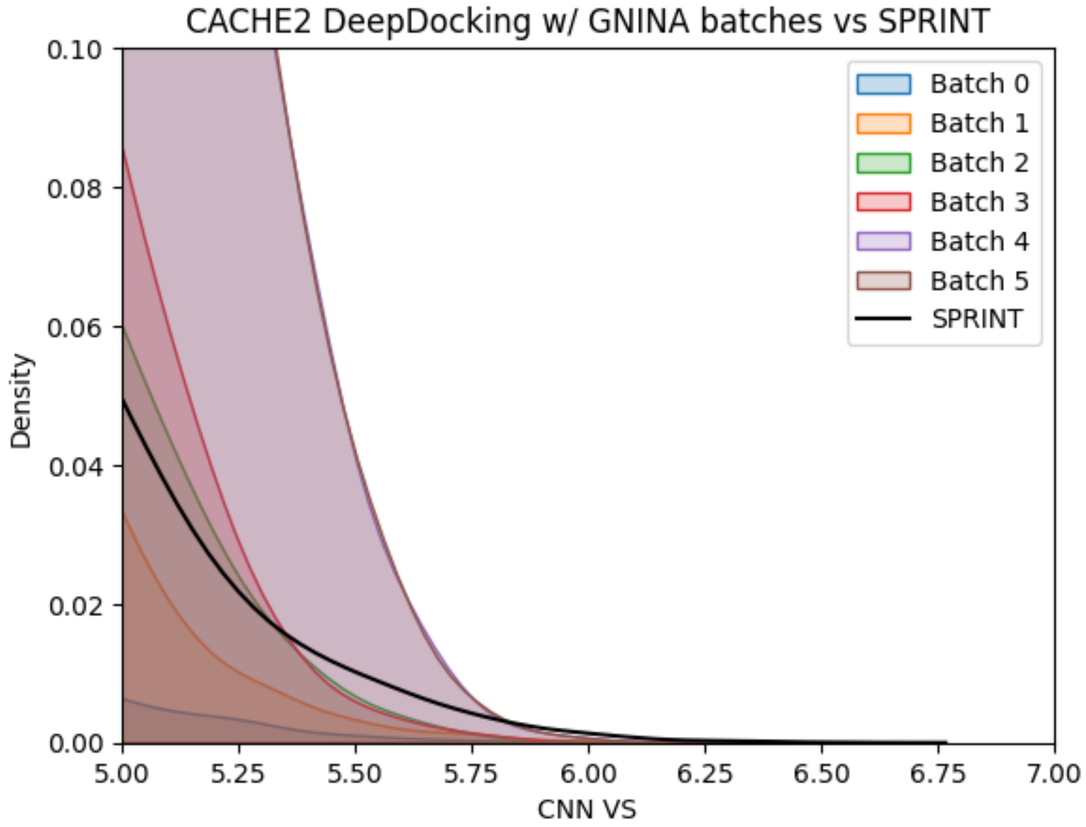


https://github.com/abhinadduri/panspecies-dti

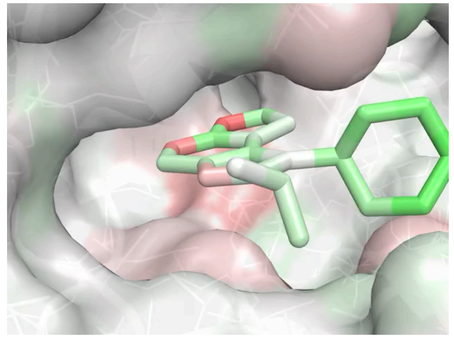
Enhanced Deep Docking with SPRINT



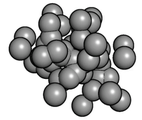
Retrospective CACHE2 Deep Docking



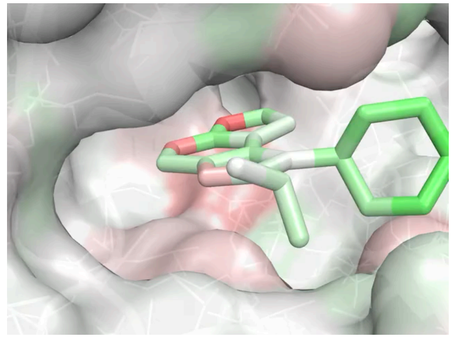
https://github.com/abhinadduri/panspecies-dti



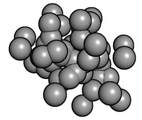
Takeaways



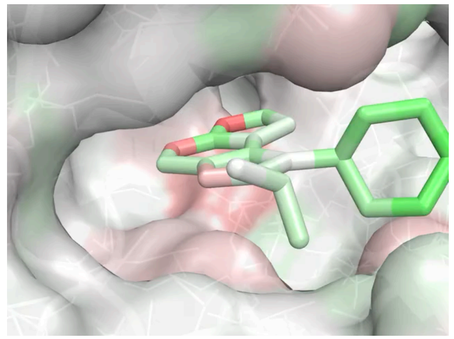
Deep learning works for molecules!



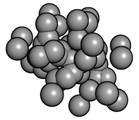
Takeaways



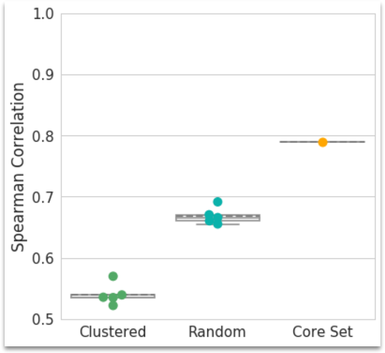
Deep learning works for molecules!

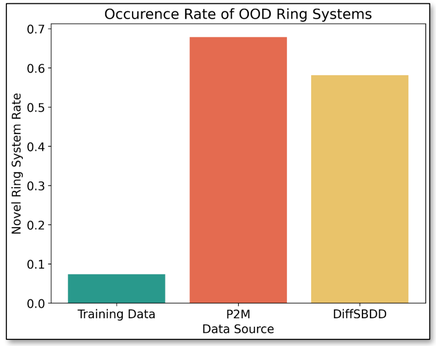


Takeaways

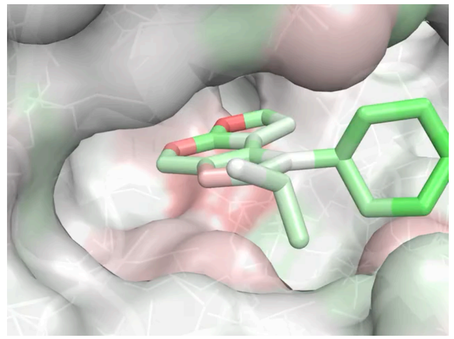


Deep learning works for molecules!

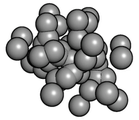




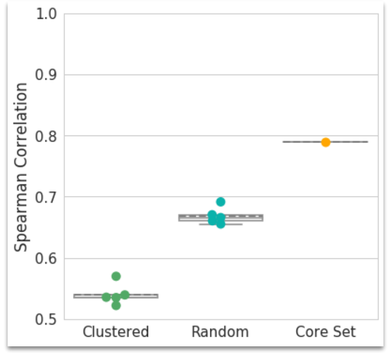
Rigorous evaluation is essential

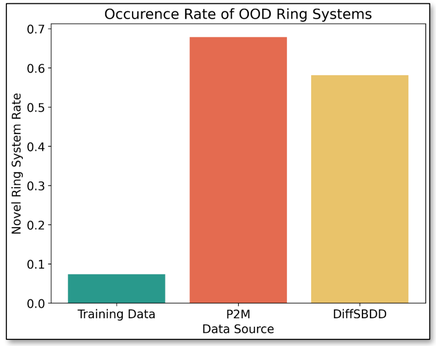


Takeaways



Deep learning works for molecules!

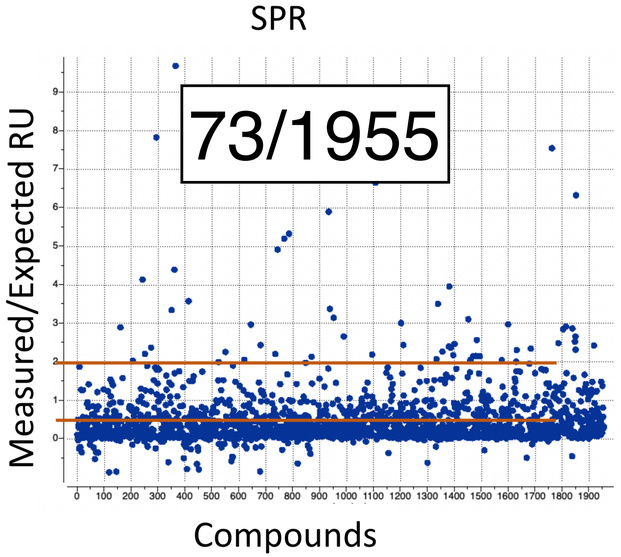


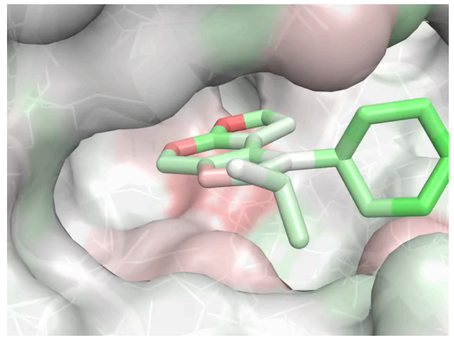


Rigorous evaluation is essential

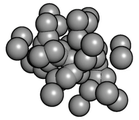
No replacement for prospective evaluation

Still pretty bad





Takeaways

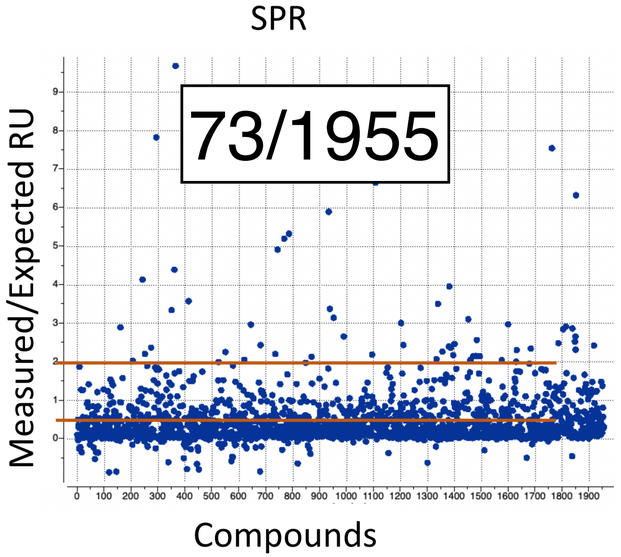


Deep learning works for molecules!

It is pronounced NEE-na

No replacement for prospective evaluation

Still pretty bad



Acknowledgements

Current

Ian Dunn
Emma Flynn
Riya Shah
Rishal Aggarwal
Drew McNutt
Daniel Penaherrera
Jacky Chen
Somayeh Pirhadi
Fareeda Abu-Juam
Ben Krummenacher

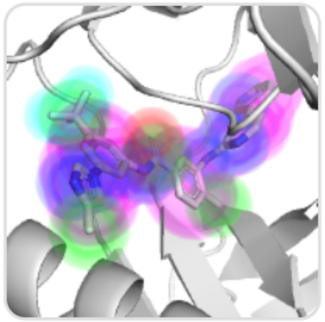
Previous

Jocelyn Sunseri
Matthew Ragoza
Tomohide Masuda
Paul Francoeur
Jonathan King
Rocco Meli
Josh Hochuli
Elisa Idrobo
Lily Turner
Alec Helbling
Andrew Jia
Rich Iovanisci
Ian Snyder
Nick Rego

National Institute of General Medical Sciences R01GM108340 R35GM140753

NSF CHE-1800435

Shameless Plugs

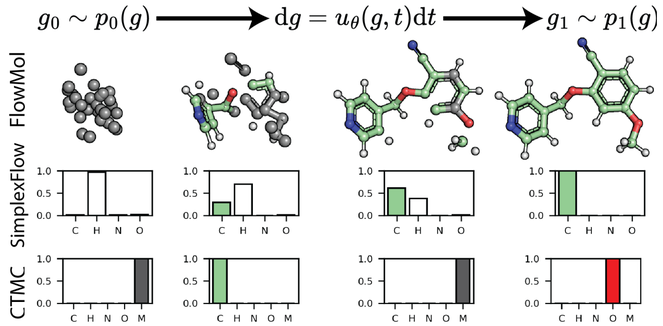


gnina

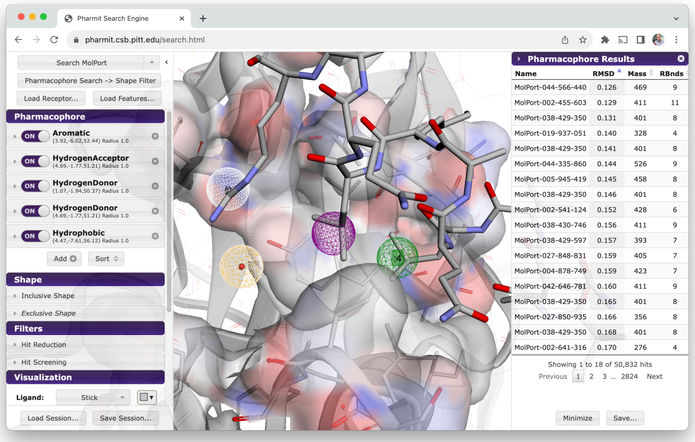
https://github.com/gnina

FlowMol

https://github.com/Dunni3/FlowMol

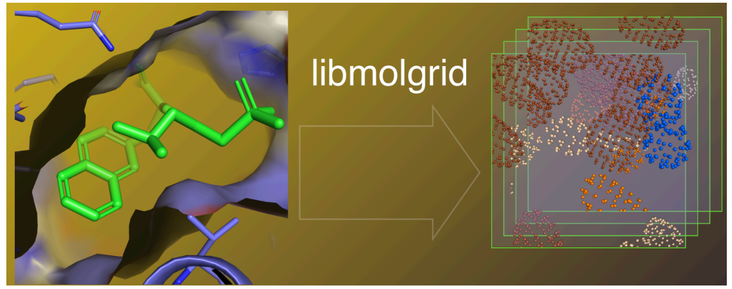


Pharmit



https://pharmit.csb.pitt.edu

@dkoes.compstruct.org



https://gina.github.io/libmolgrid/

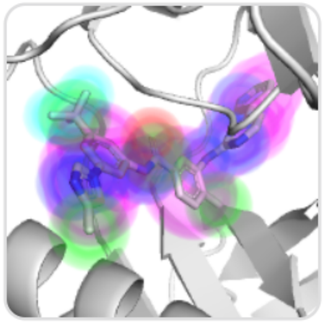
py3Dmol

```
[ ] import py3Dmol

[ ] p = py3Dmol.view(query='mmtf:1ycr')
    p.setStyle({'cartoon': {'color':'spectrum'}})
    p
```

https://3dmol.org/

Shameless Plugs



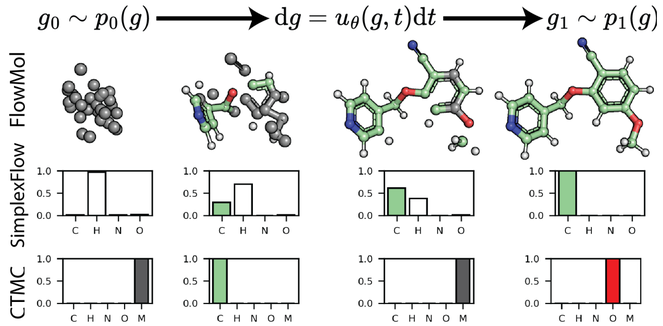
gnina

(pronounced NEE-na)

https://github.com/gnina

FlowMol

https://github.com/Dunni3/FlowMol



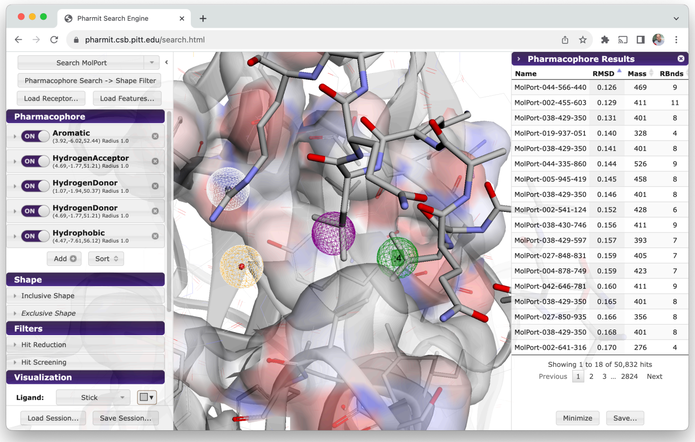
py3Dmol

```
[ ] import py3Dmol

[ ] p = py3Dmol.view(query='mmtf:1ycr')
    p.setStyle({'cartoon': {'color':'spectrum'}})
    p
```

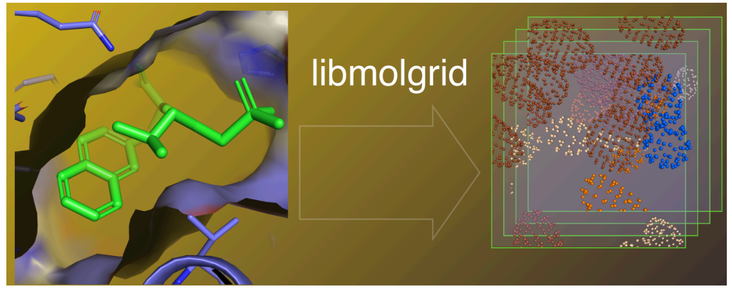
https://3dmol.org/

Pharmit



https://pharmit.csb.pitt.edu

@dkoes.compstruct.org



https://gina.github.io/libmolgrid/