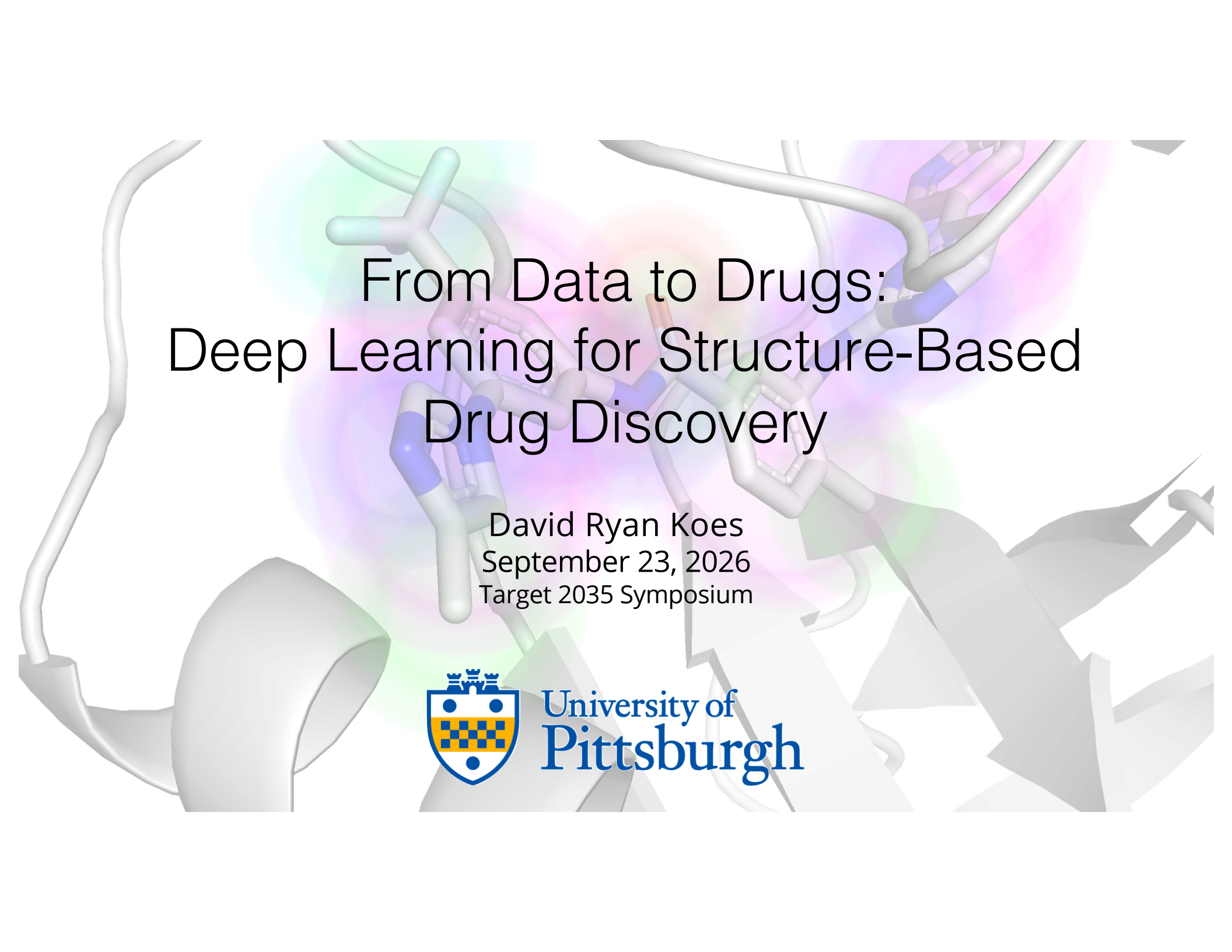


A complex 3D molecular structure, likely a protein-ligand complex, rendered in a semi-transparent style. The structure is composed of various colored spheres (blue, red, green, yellow) and sticks, with some parts highlighted in a translucent purple and pink glow. The overall shape is intricate, with many loops and folds.

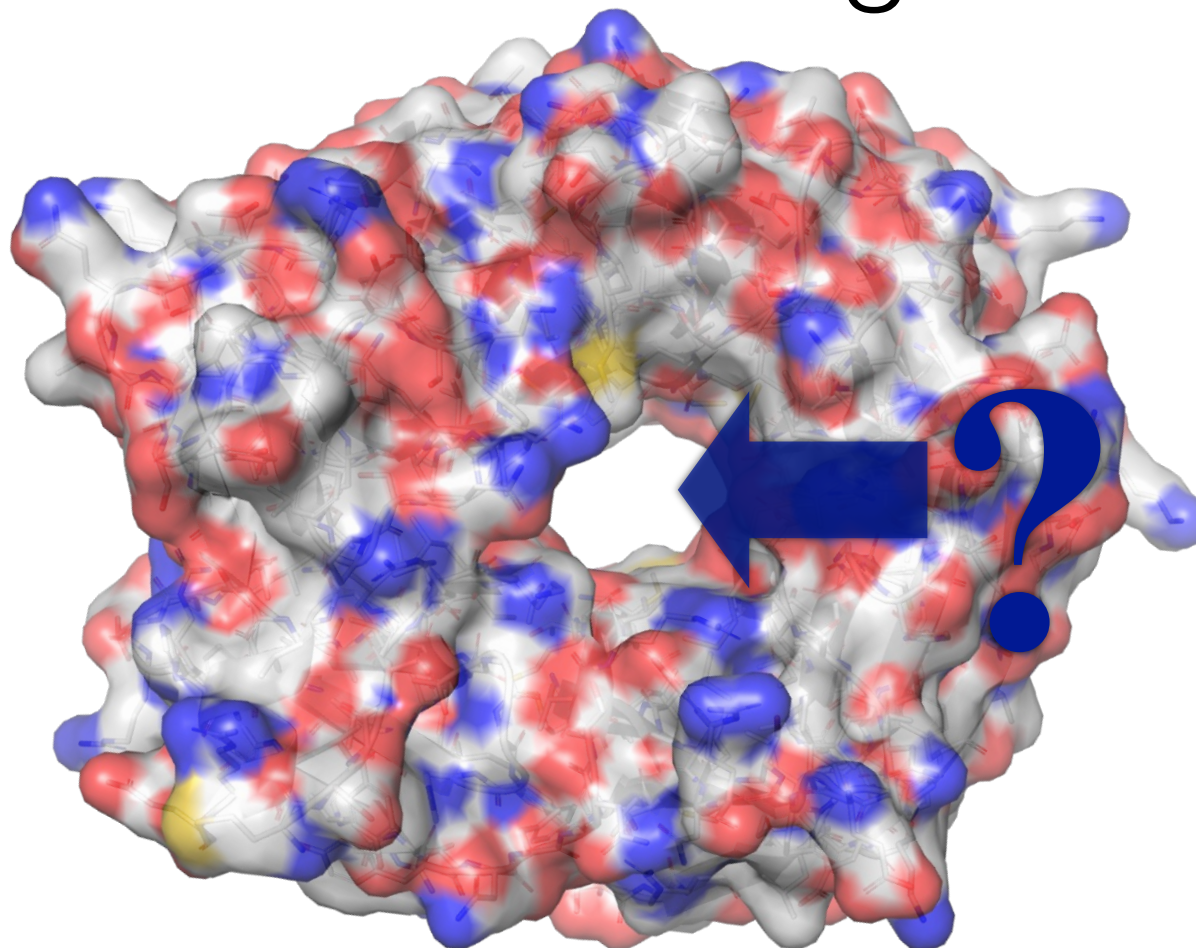
From Data to Drugs: Deep Learning for Structure-Based Drug Discovery

David Ryan Koes
September 23, 2026
Target 2035 Symposium



University of
Pittsburgh

Structure Based Drug Discovery



GNINA 1.0

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

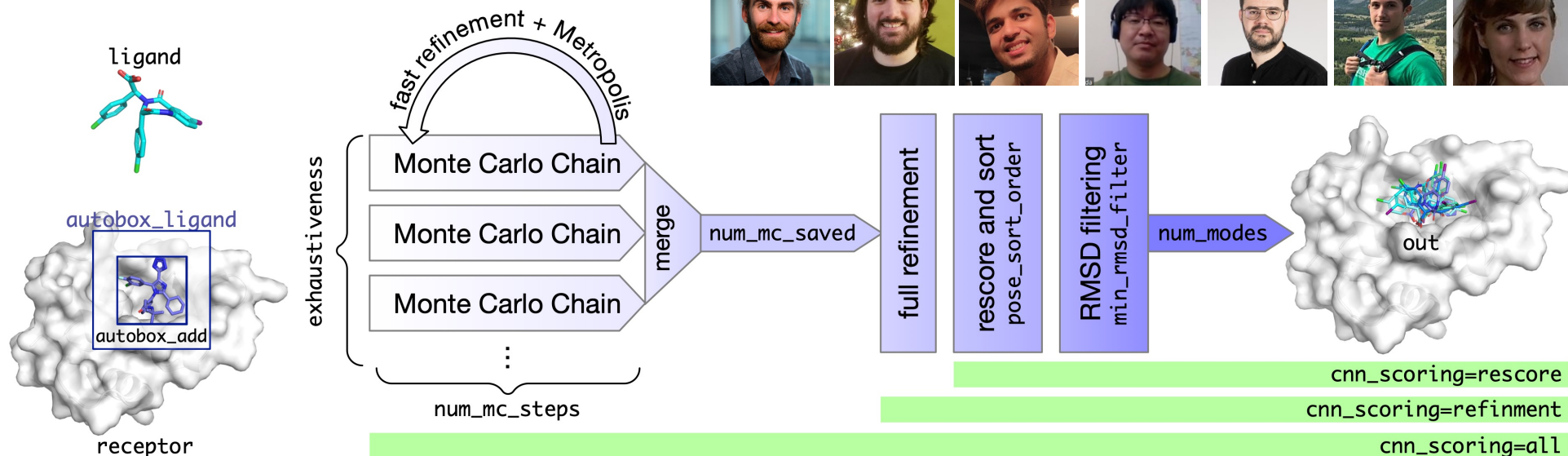
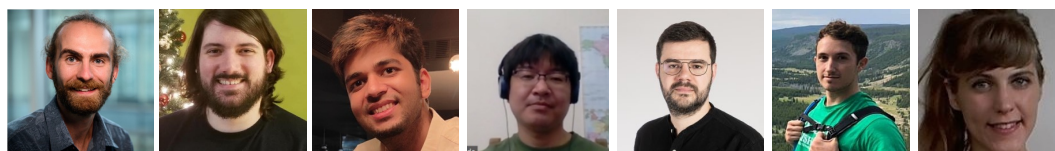
SOFTWARE

Open Access

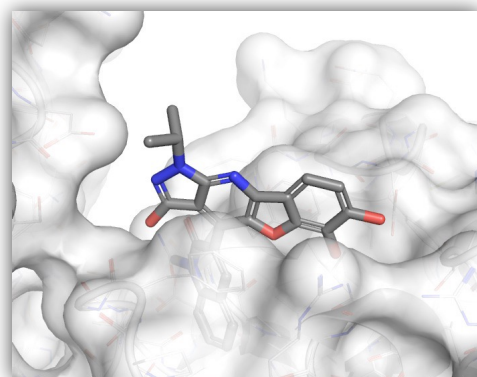


GNINA 1.0: molecular docking with deep learning

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



Protein-Ligand Scoring



Model

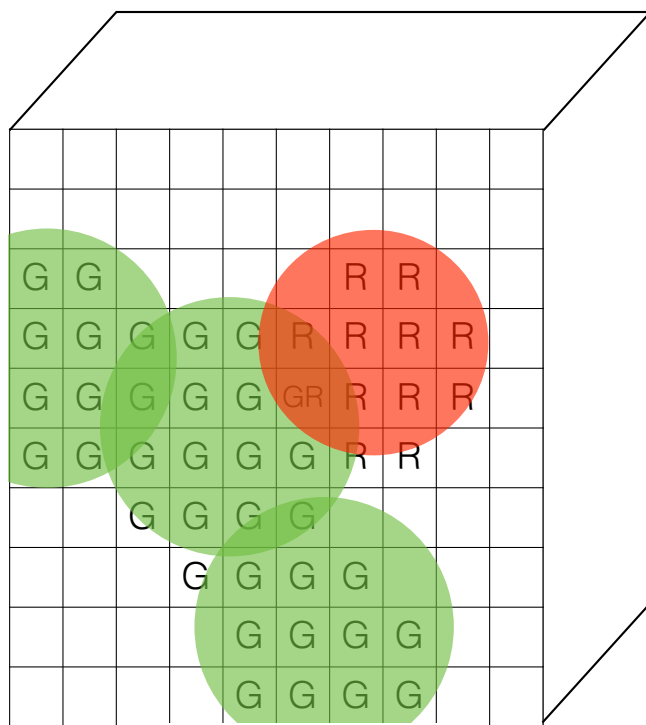


Pose Prediction

Binding
Discrimination

Affinity Prediction

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**.

JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

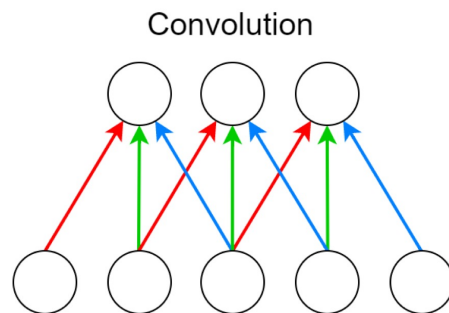
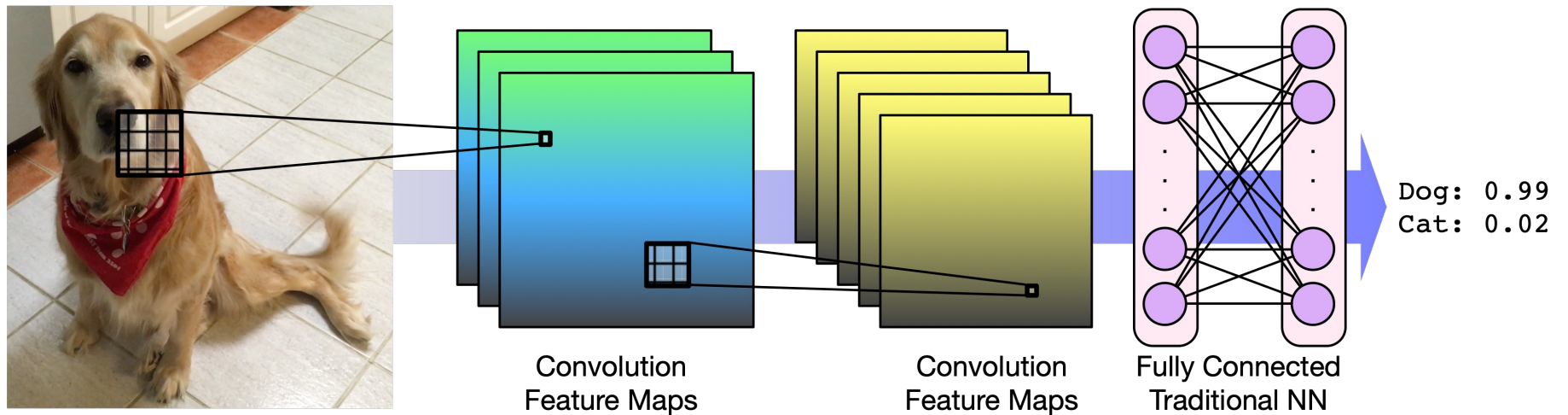
Article

pubs.acs.org/jcim

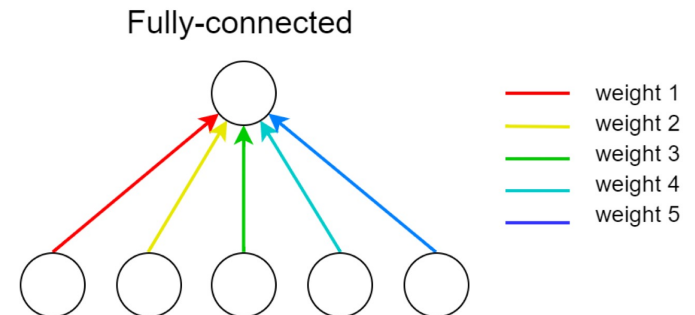
Protein–Ligand Scoring with Convolutional Neural Networks

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

Convolutional Neural Networks



— weight 1
— weight 2
— weight 3



— weight 1
— weight 2
— weight 3
— weight 4
— weight 5

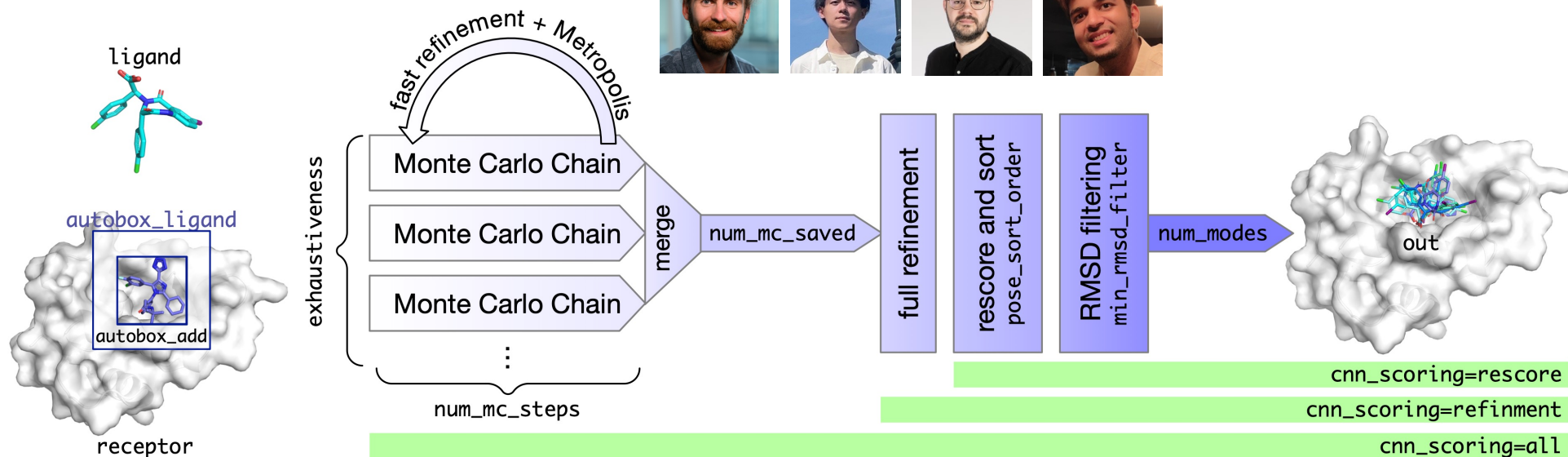
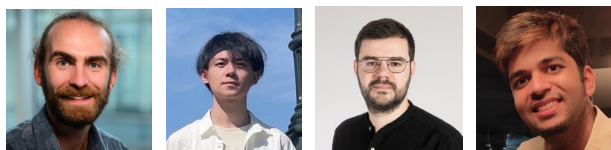
GNINA 1.3

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

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

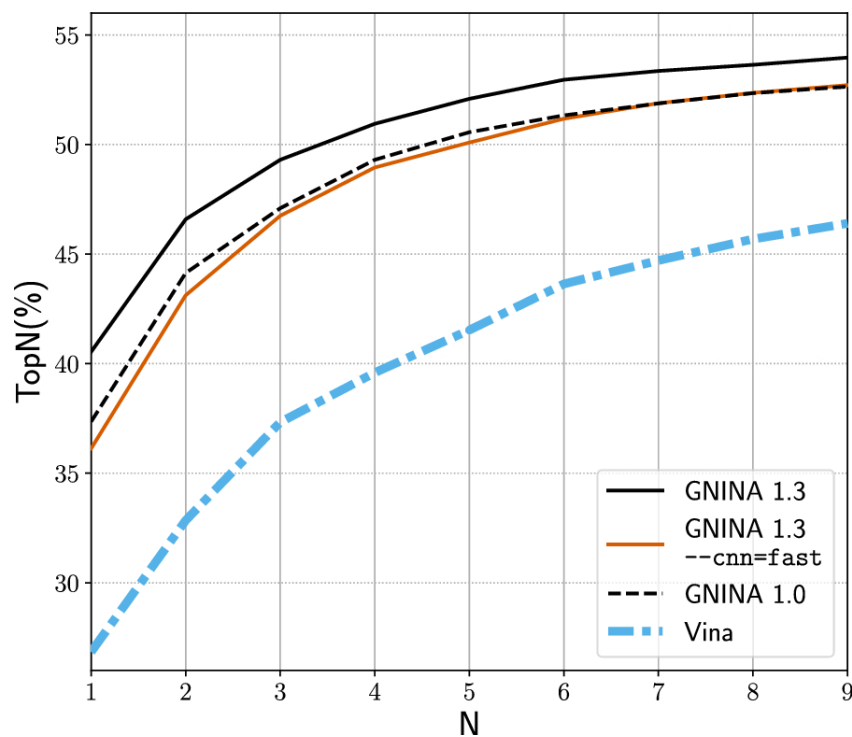
Andrew T. McNutt, Yanjing Li, Rocco Meli, Rishal Aggarwal & David Ryan Koes ✉

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

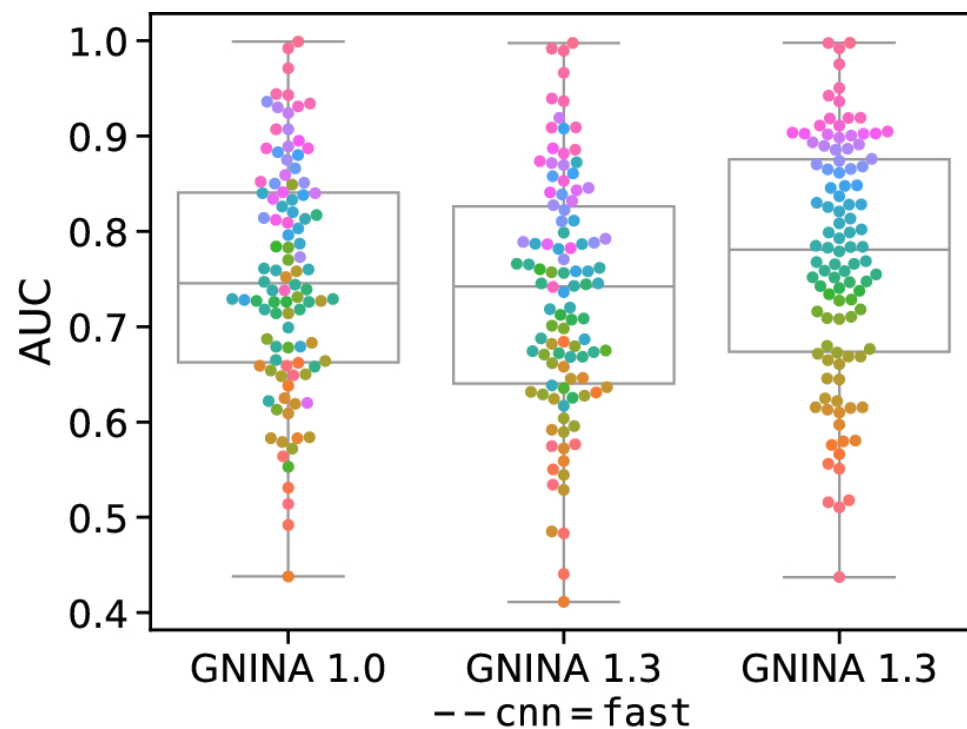


Caffe → Torch
easy covalent docking
retrained models

GNINA 1.3 Performance



Crossdocking



Virtual Screening (DUD-E)

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. Iovaniisci, 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

Altmetric

17

Citations

5

[LEARN ABOUT THESE METRICS](#)

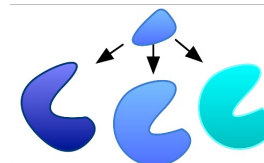
Share



Add to



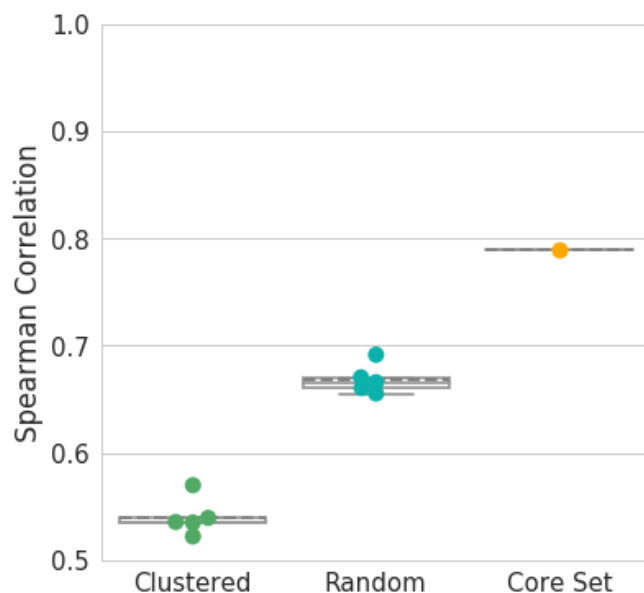
Export



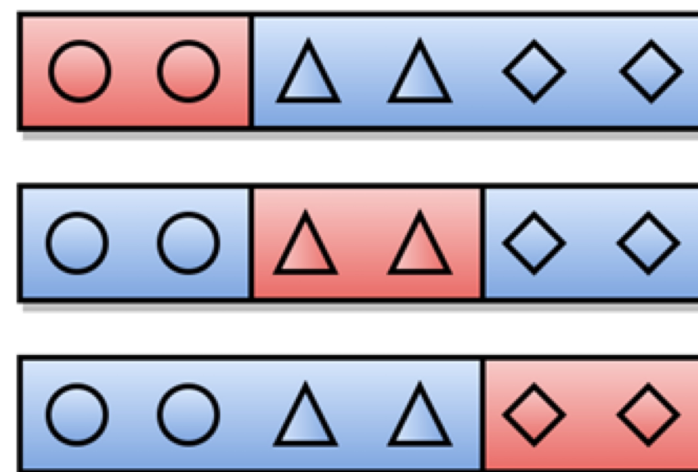
18,450 complexes

22.6 million poses

Affinity
Prediction



Clustered Cross-validation



**Isomorphic Labs**

73,687 followers

1w •

Isomorphic Labs has reached an exciting new milestone in computational drug discovery. In a technical report published today, we show how our drug design engine represents a massive leap forward in accurately predicting ...more



You and 1,890 others

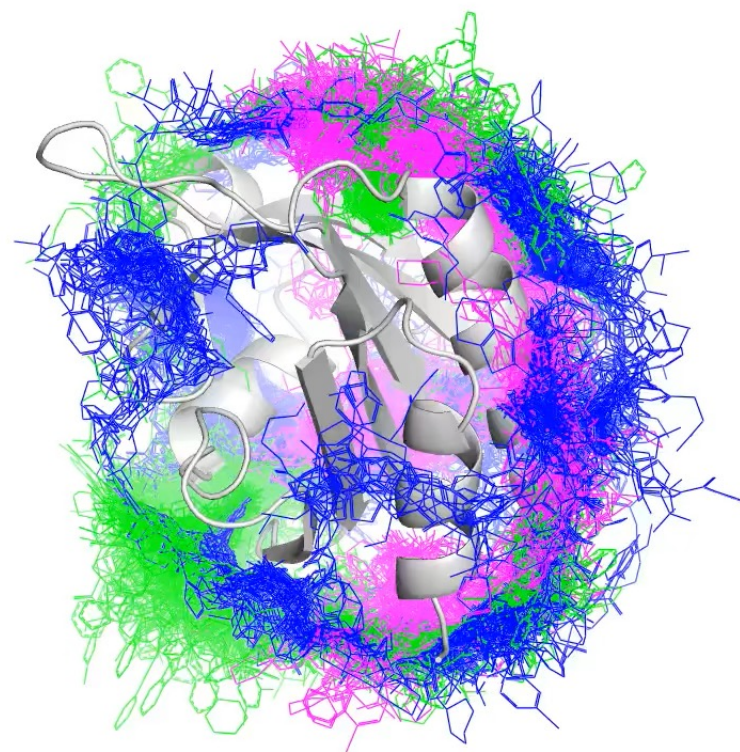
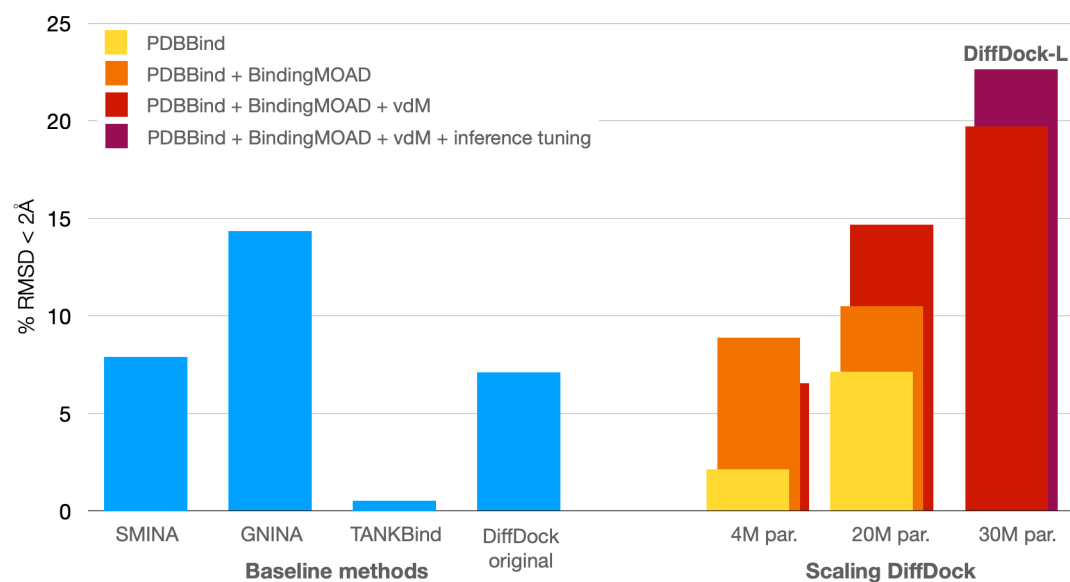
16 comments · 203 reposts

[nature](#) > [news](#) > [article](#)**NEWS** | 19 February 2026

‘An AlphaFold 4’ – scientists marvel at DeepMind drug spin-off’s exclusive new AI

Isomorphic Lab’s proprietary drug-discovery model is a major advance, but scientists developing open-source tools are left guessing how to achieve similar results.

GNINA vs End-to-end Deep Docking



GNINA DynamicBind Boltz1

Prospective Evaluation

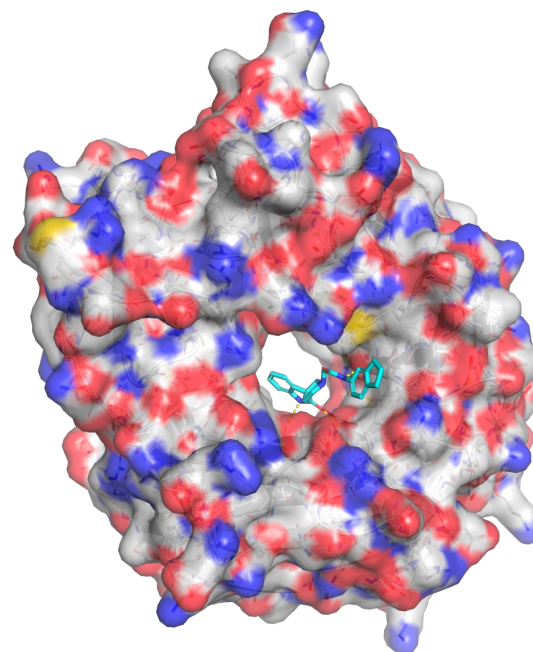
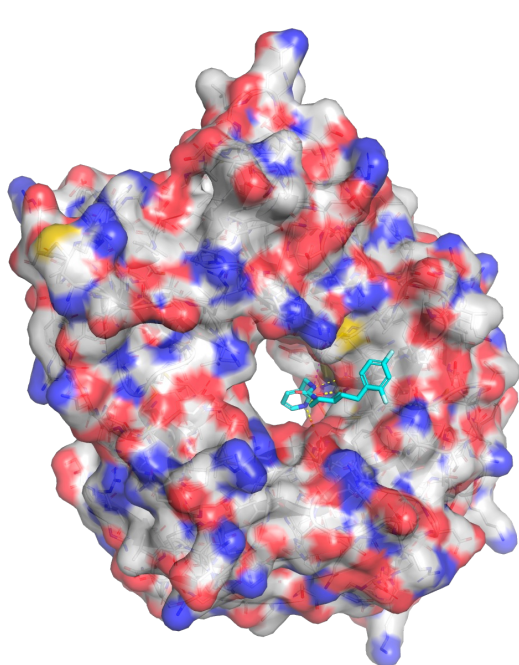


CRITICAL ASSESSMENT OF COMPUTATIONAL
HIT-FINDING EXPERIMENTS

CACHE Challenge #1

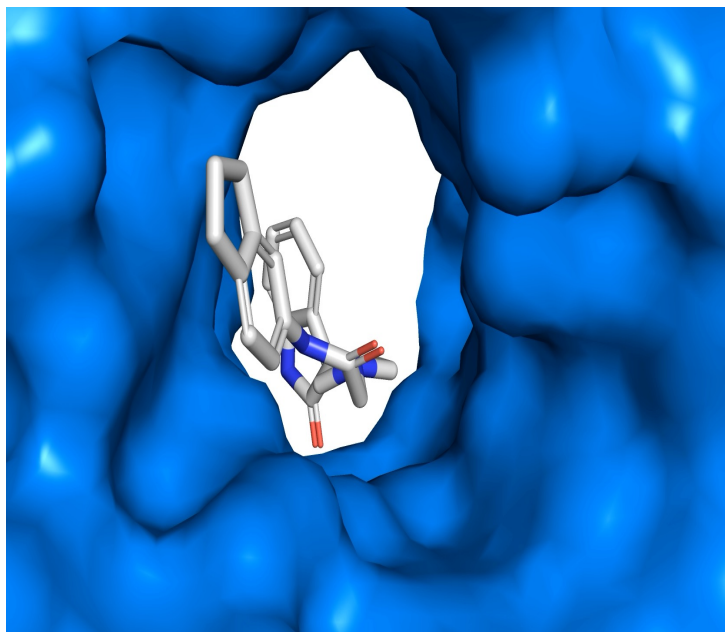


Ian Dunn

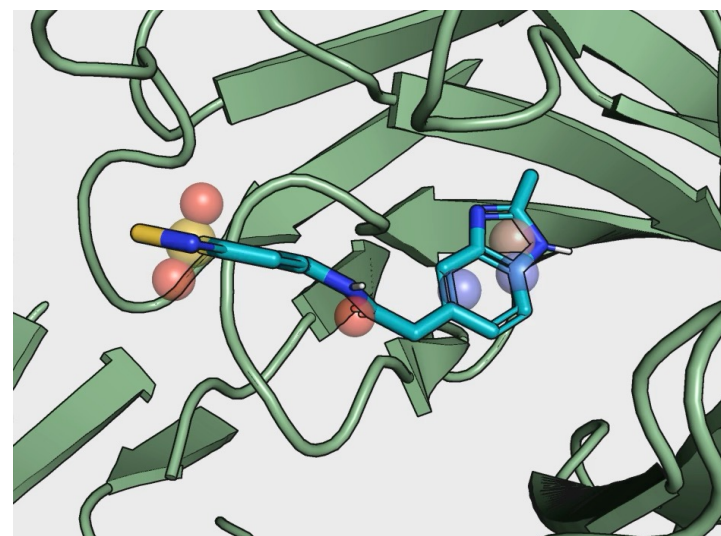


A Tale of Two Methods

Large-Scale Docking with GNINA



Pharmacophore Screening with Pharmit



The screenshot displays the pharmit web application interface. The central area shows a 3D molecular docking visualization with a ligand (yellow sticks) bound to a protein (grey surface). The left sidebar contains a 'Pharmacophore' section with various filters (Aromatic, HydrogenAcceptor, NegativeIon, Hydrophobic) and their coordinates and radii. The right sidebar shows a 'Pharmacophore Results' table with columns for Name, RMSD, Mass, and RBnds. A central banner for 'pharmit' includes the text 'interactive exploration of chemical space' and links for 'code', 'help', and 'contact'.

pharmit
interactive exploration of chemical space

[code](#)
[help](#)
[contact](#)

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
	0.327	389	6
	0.349	875	15
	0.379	387	3
	0.379	386	2
	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

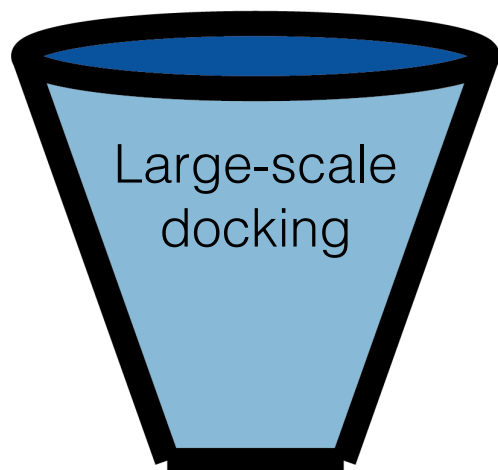
Showing 1 to 19 of 38 hits
Previous 1 2 Next

Minimize Save...

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

Round 1 Submission

molport



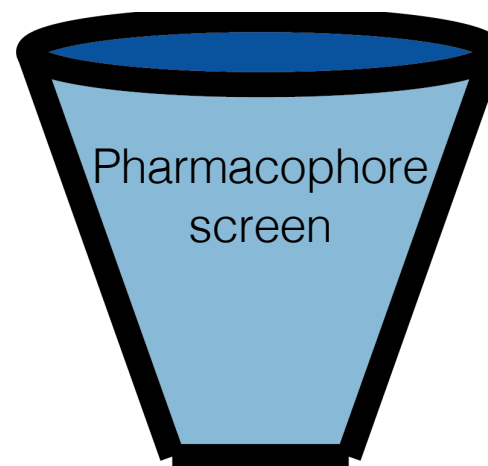
1k ligands
gnina scores
vina scores

2 screening methods
2 scoring methods



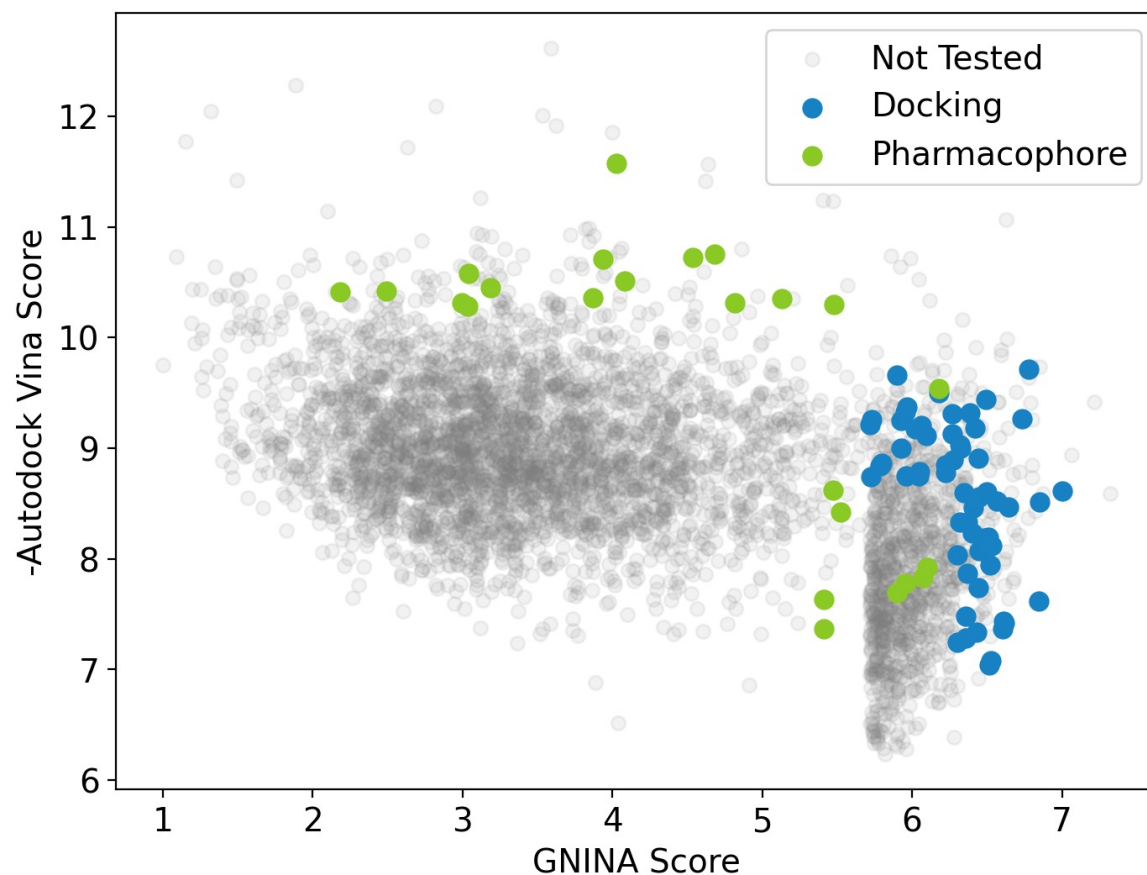
- ZINC20: 20 mil molecules
- MCULE: 45 mil molecules
- MCULE-ULTIMATE: 126 mil molecules

Molecule Libraries



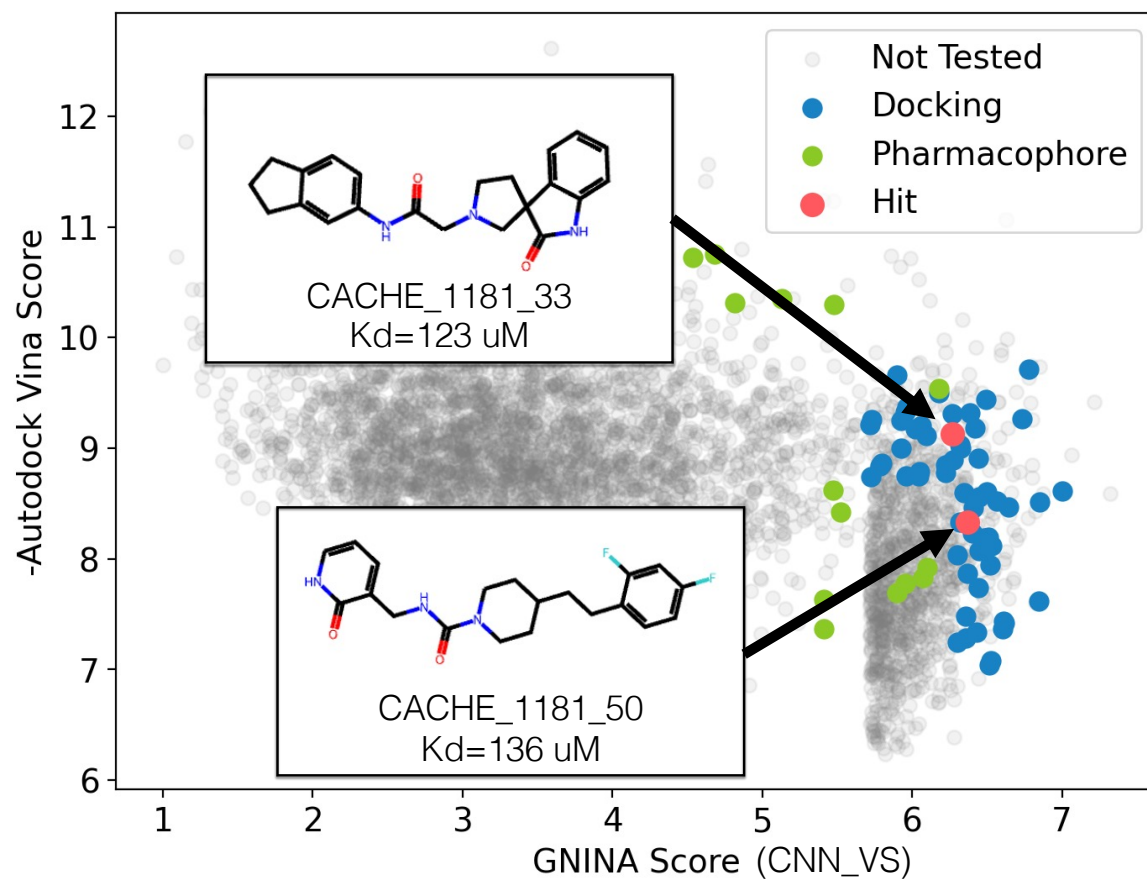
3.5k ligands
gnina scores
vina scores

Round 1 Results



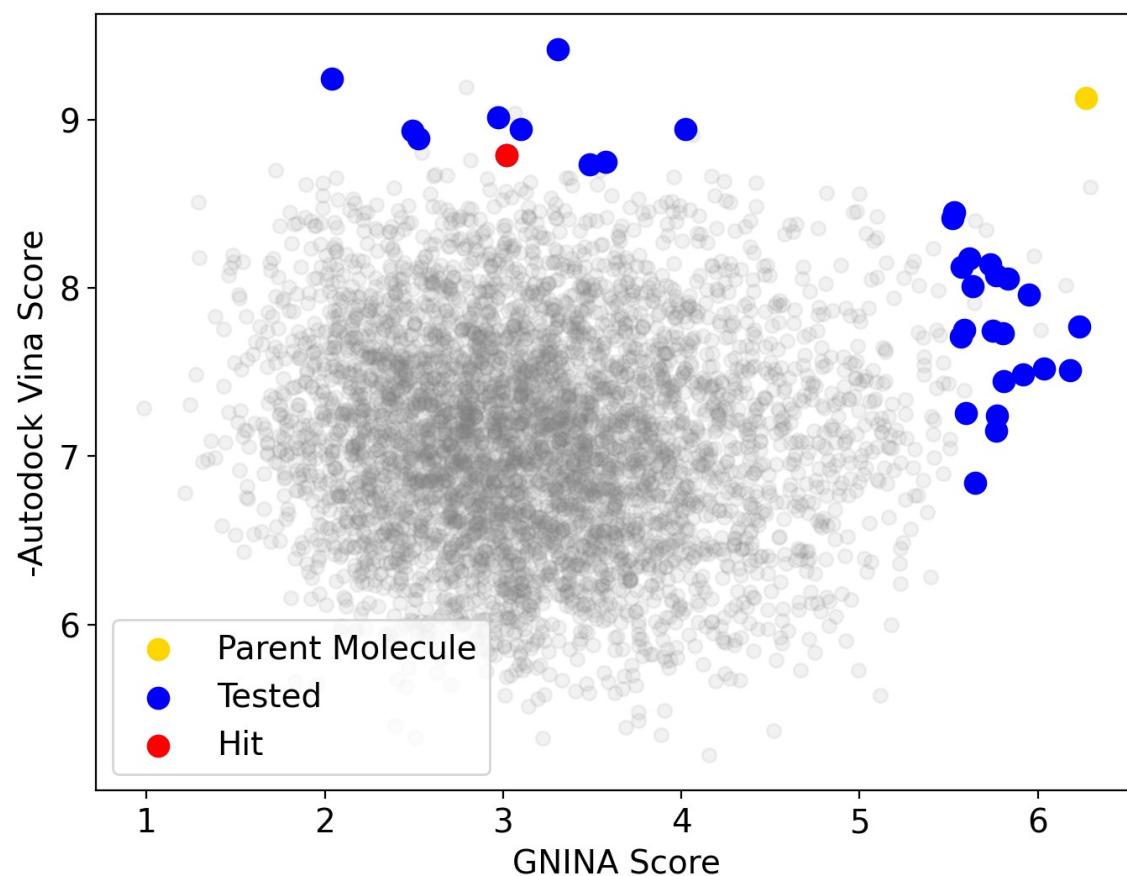
- Selection limited/skewed by database availability
- 84 ligands tested
 - 59 from docking
 - 24 from pharm screen

Round 1 Results

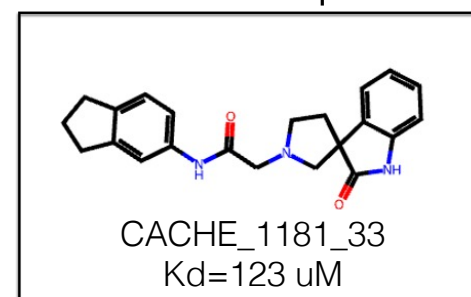


- 2/84 were hits
- Both from docking

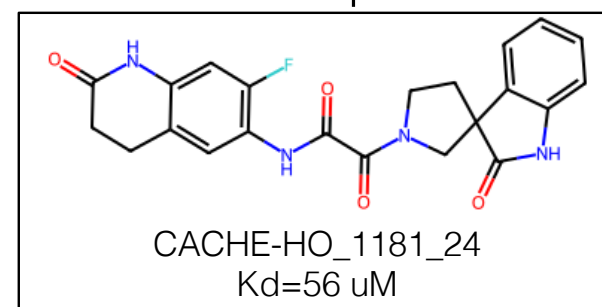
Hit Optimization Results



Parent Compound



Hit Compound

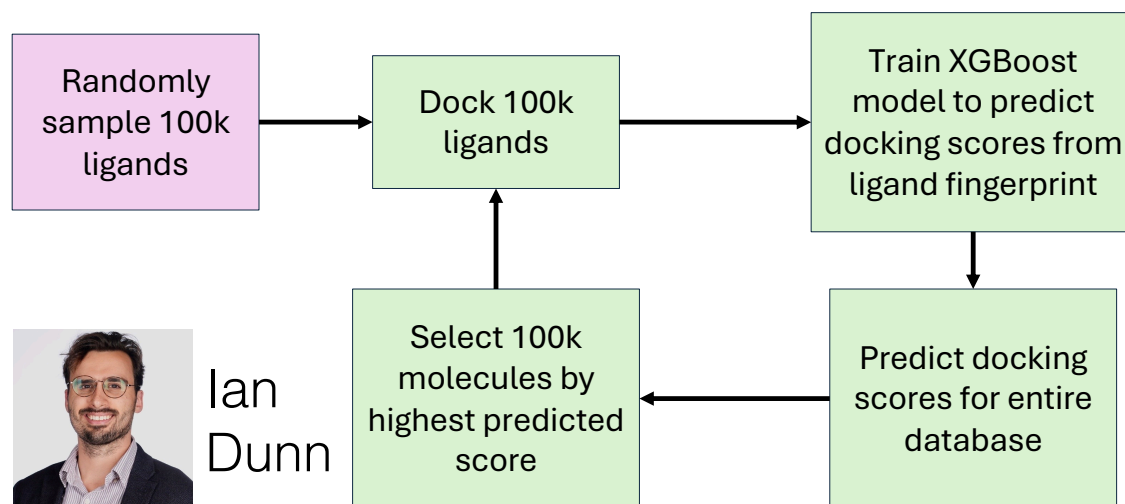
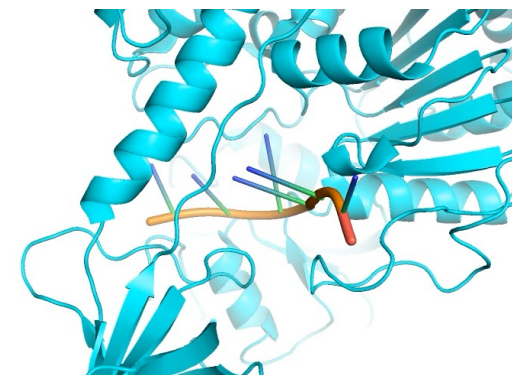


Final Results

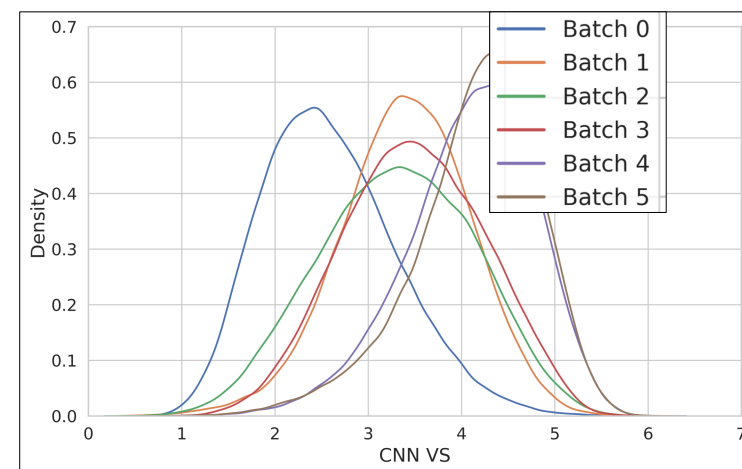
Participant	Participant ID	Aggregated score
David Koes, University of Pittsburgh	1181	18
Olexandr Isayev & Maria Kurnikova, Carnegie Mellon University & Artem Cherkasov, University of British Columbia	1209	18
Christina Schindler, Merck KGaA	1193	17
Dmitri Kireev, University of Missouri	1183	16
Christoph 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



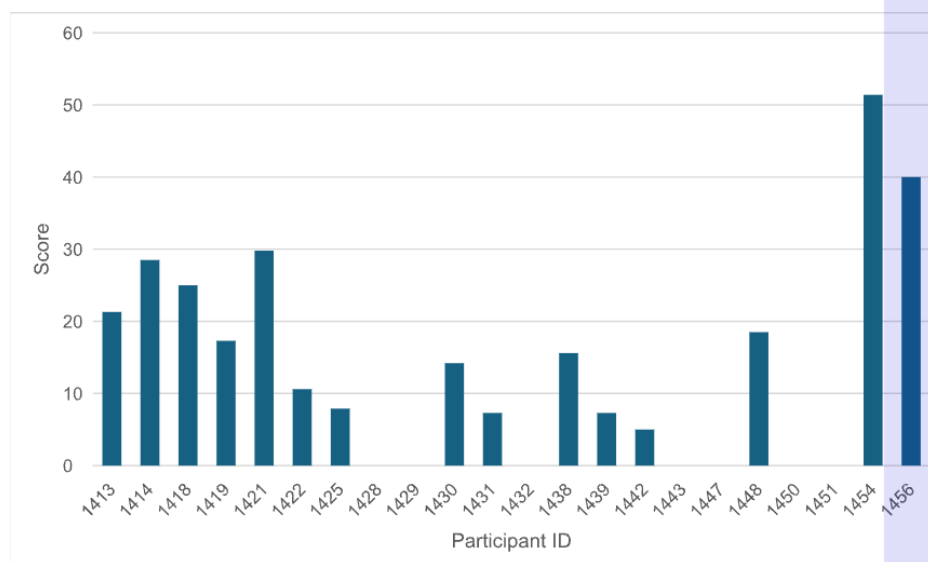
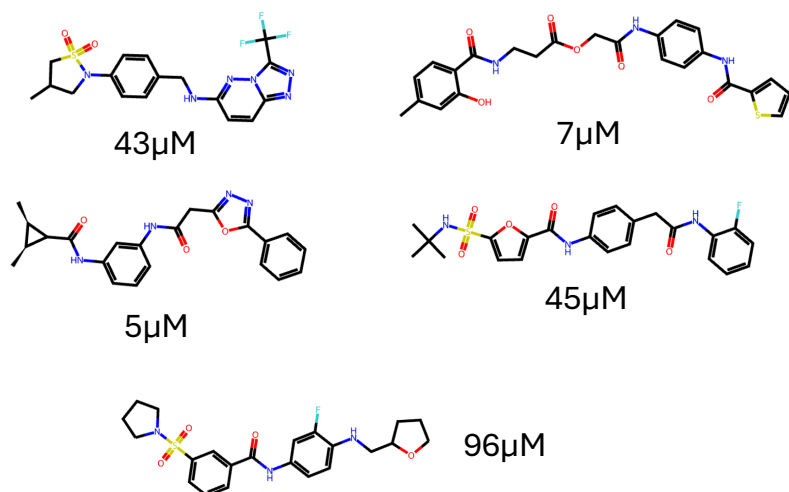
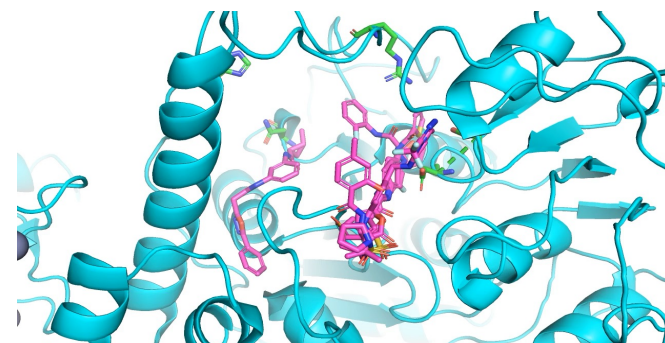
CACHE #2 Results

5/50 compounds identified as potential hits

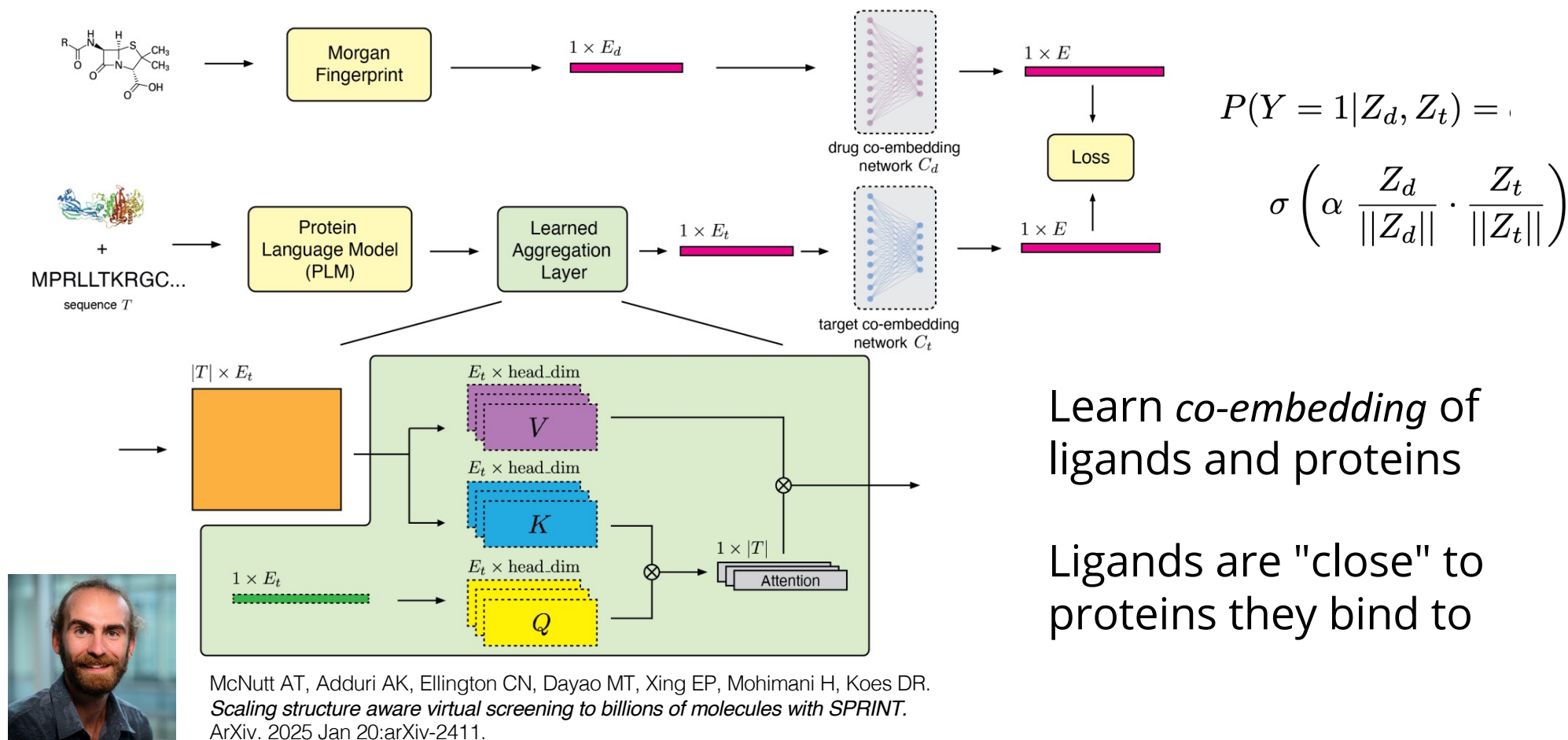
>2x the average hit rate

4/5 hits from last round of active learning

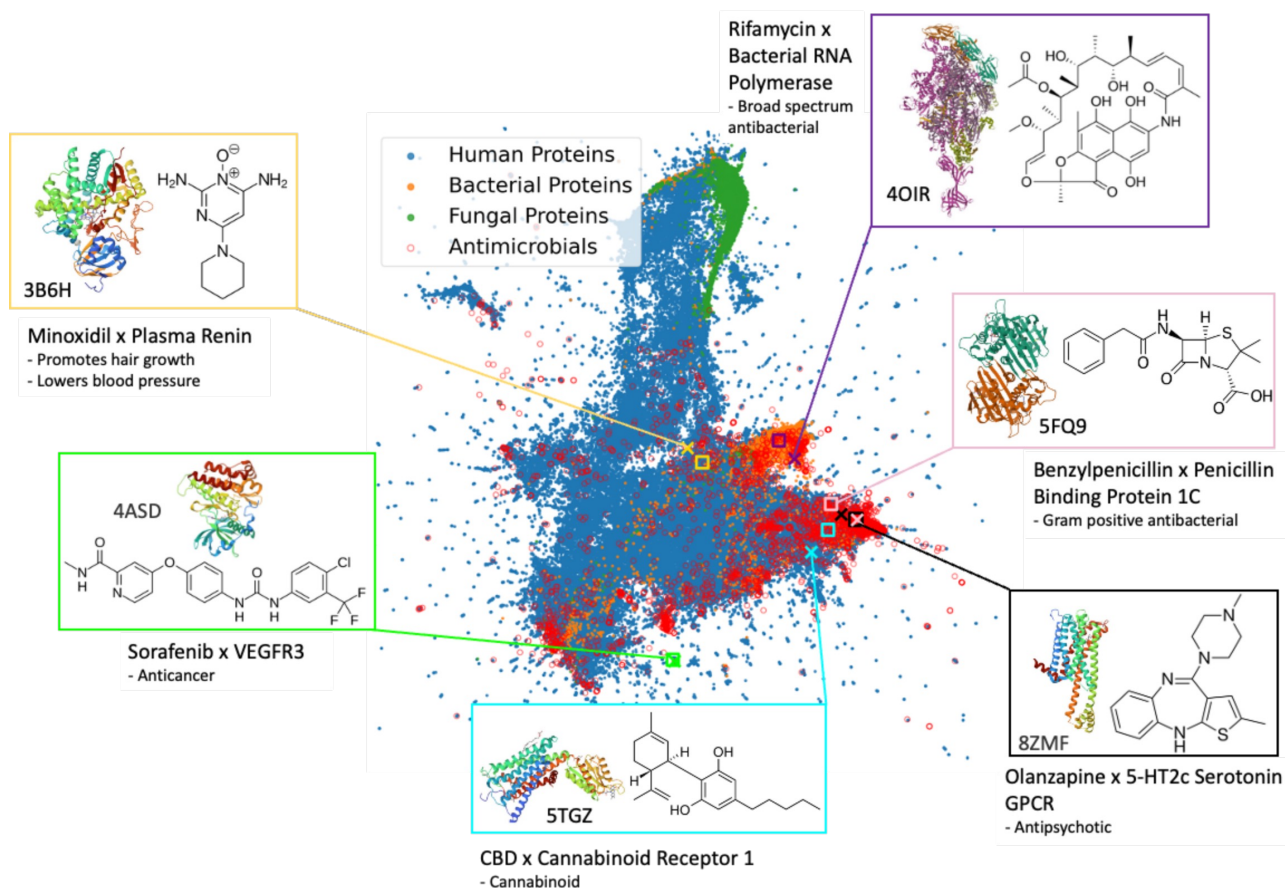
Highest affinity round 1 hit in the competition



Scalable Screening of Ultra Large Libraries



SPRINT Co-embeddings

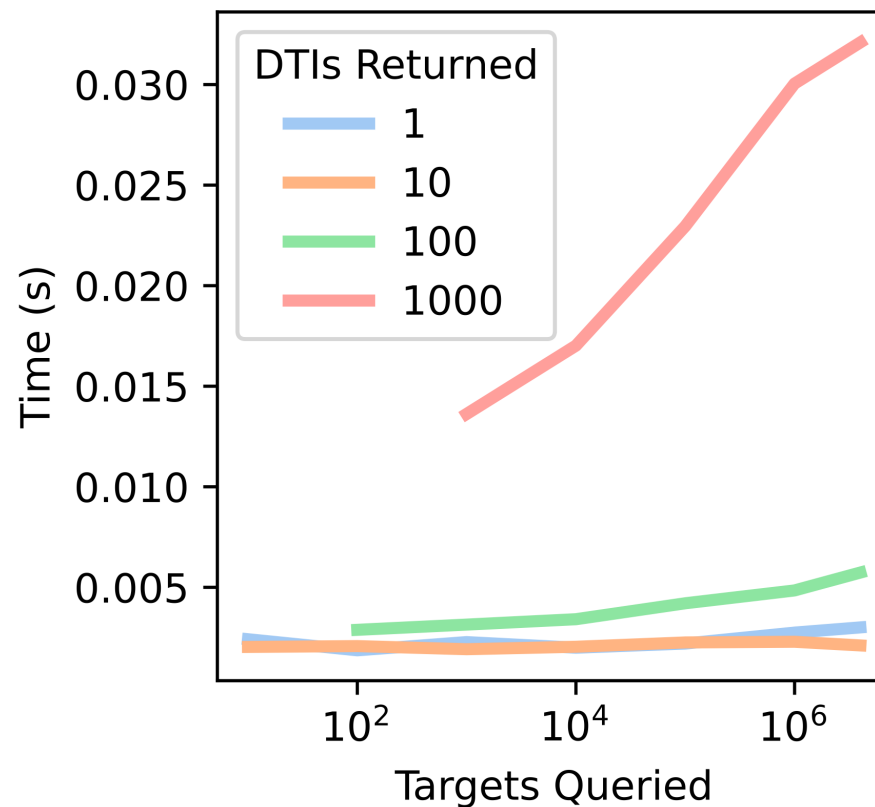


Suspiciously Good Retrospective Virtual Screening

Table 2: Virtual Screening results on LIT-PCBA. SPRINT-ProtBert replaces SaProt with the ProtBert model, and SPRINT-Average replaces learned aggregation with average pooling and additional MLP layers. Parameter counts are shown in parentheses.

	AUROC (%)	BEDROC (%)	EF		
			0.5%	1%	5%
Surflex [20]	51.47	-	-	2.50	-
Glide-SP [21]	53.15	4.00	3.17	3.41	2.01
Planet [22]	57.31	-	4.64	3.87	2.43
GNINA [23]	60.93	5.40	-	4.63	-
DeepDTA [24]	56.27	2.53	-	1.47	-
BigBind [25]	60.80	-	-	3.82	-
DrugCLIP [9]	57.17	6.23	8.56	5.51	2.27
SPRINT-Average (15.7M)	67.49	7.80	7.23	6.26	3.71
SPRINT-ProtBert (15.9M)	73.4	11.9	11.68	10.19	5.27
SPRINT (16M)	73.4	12.3	15.90	10.78	5.29

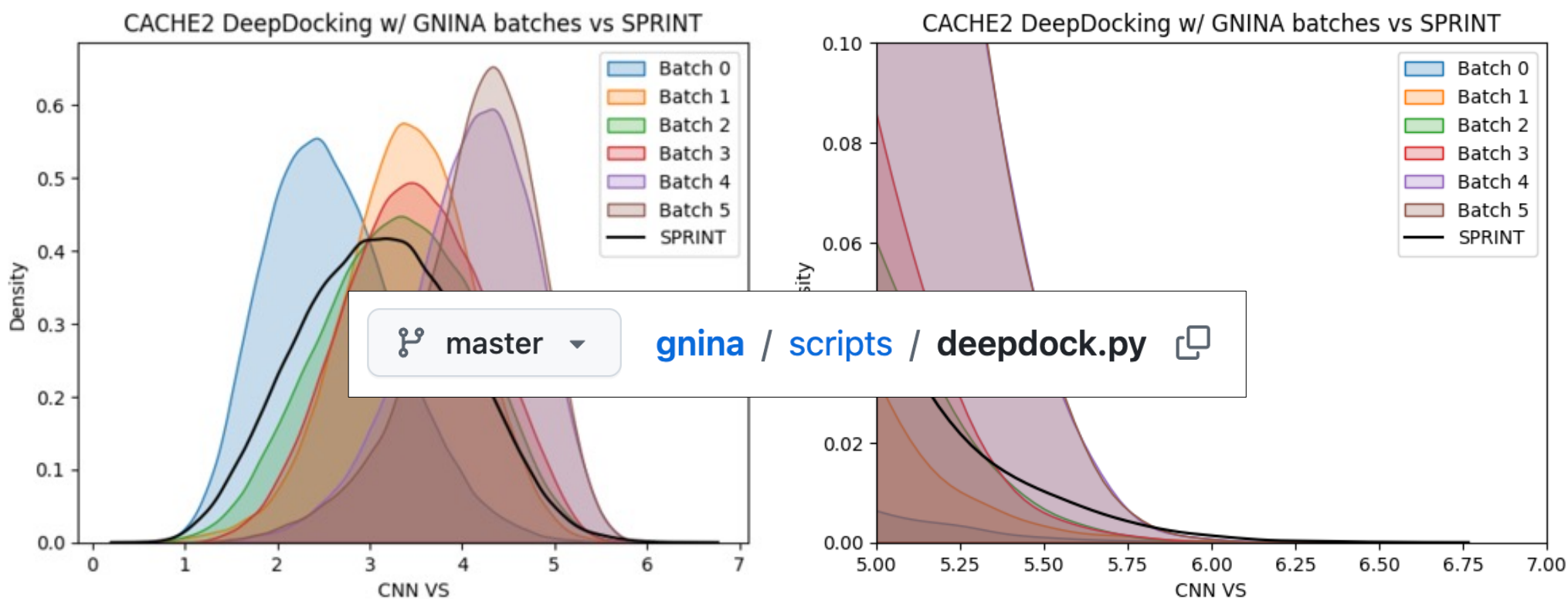
SPRINT can be crazy fast



 **Chroma**
Vector Database

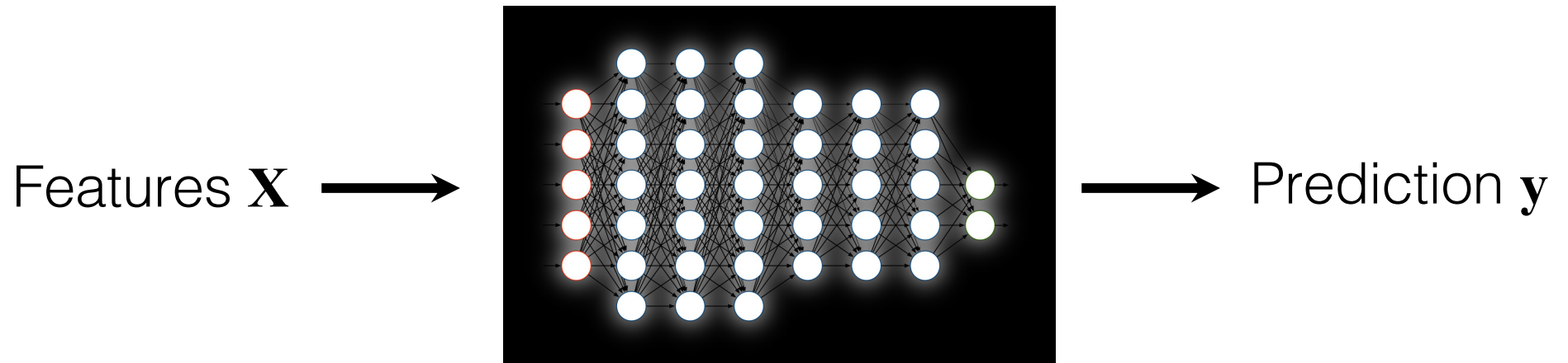
Enhanced Deep Docking with SPRINT

Retrospective CACHE2 Deep Docking

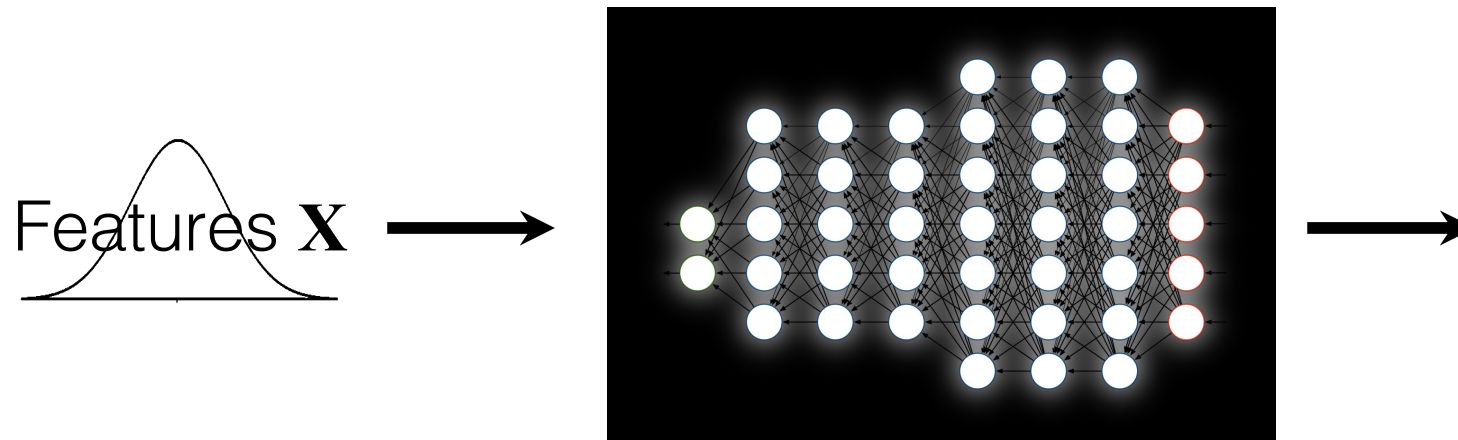


Generative Modeling

Discriminative Model



Generative Model



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

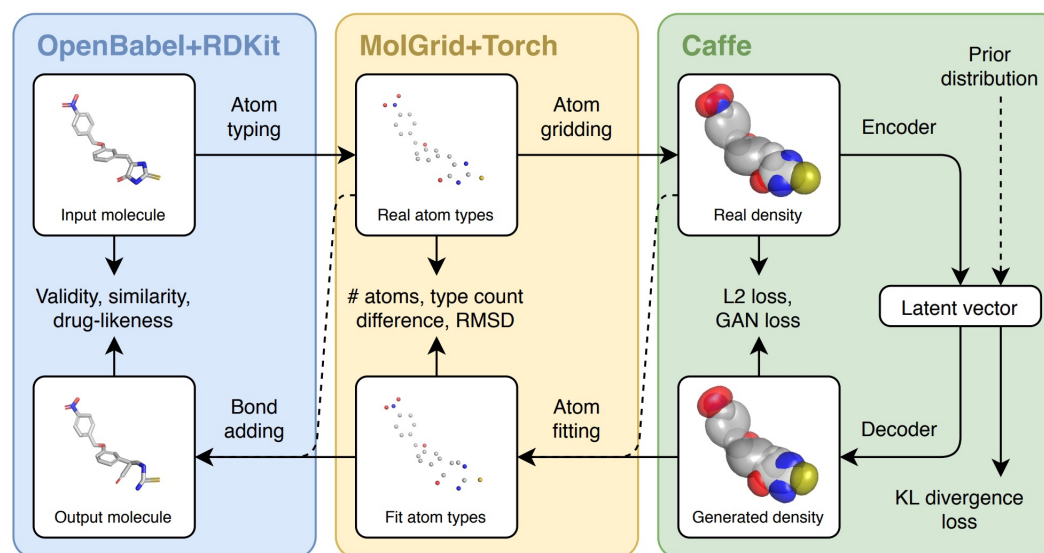
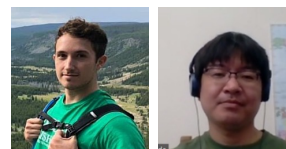
NeurIPS 2020 Workshop

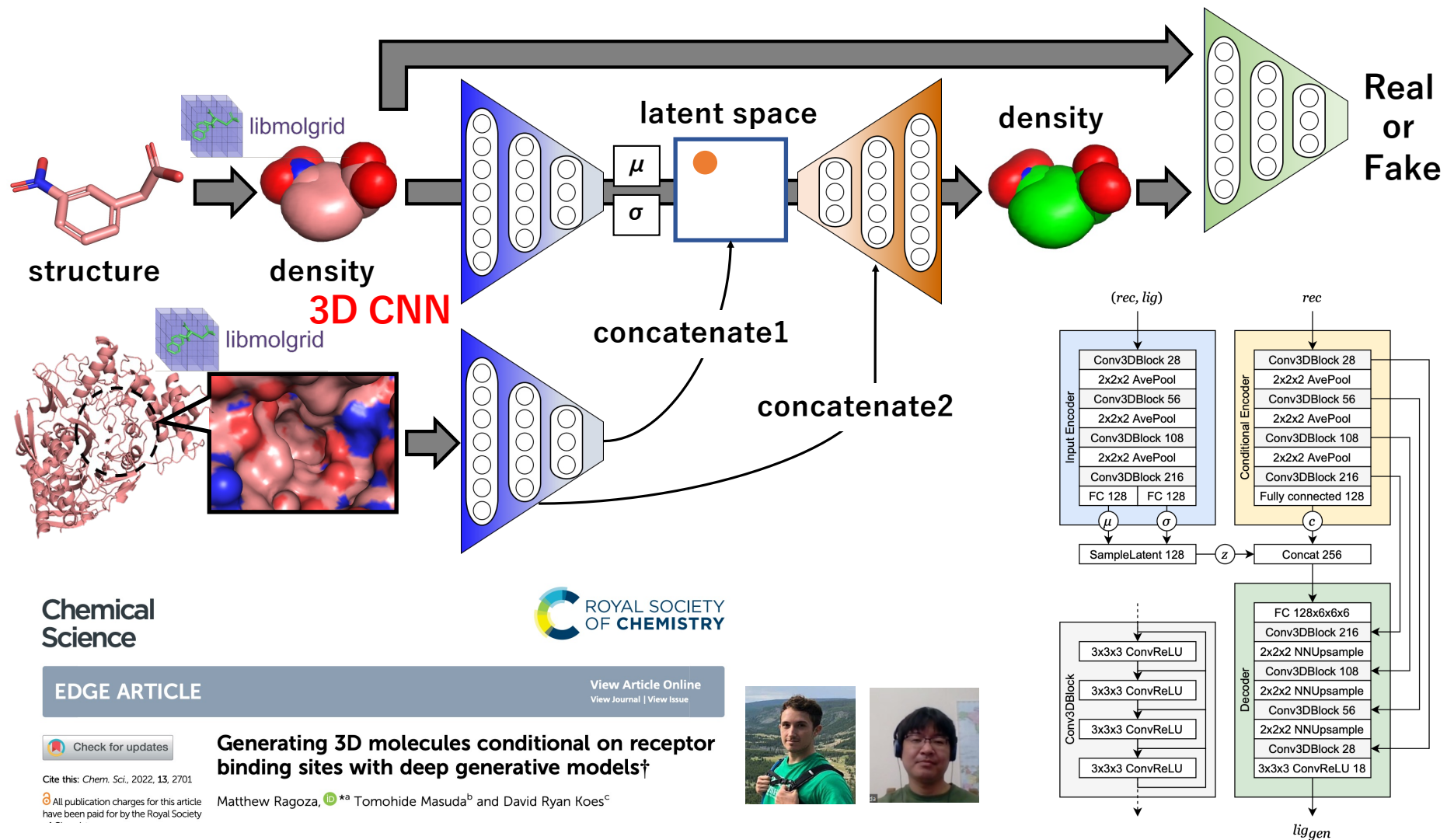
Machine Learning for Structural Biology

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



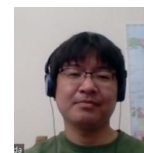
Chemical
ScienceROYAL SOCIETY
OF CHEMISTRY

EDGE ARTICLE

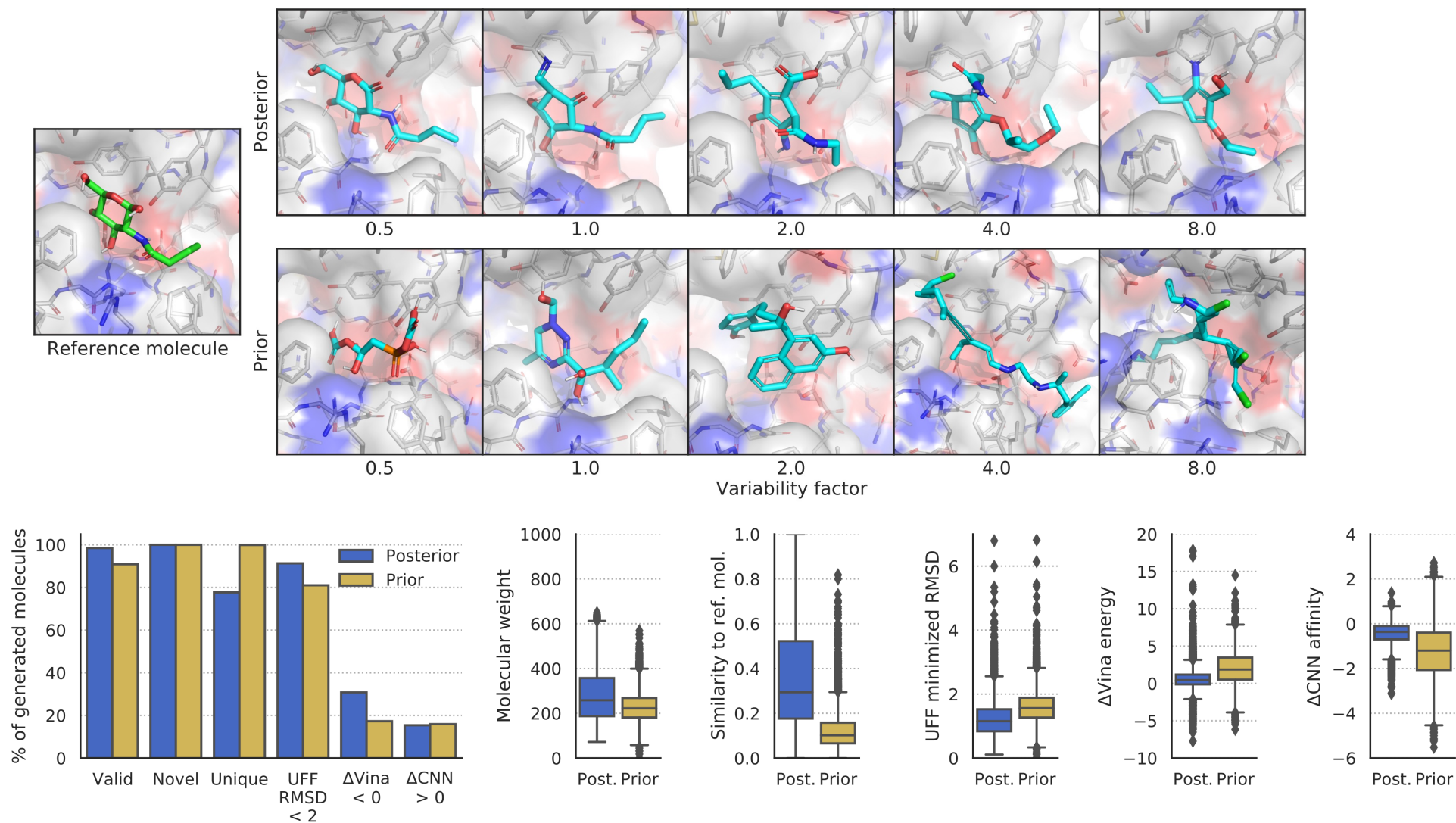
View Article Online
View Journal | View IssueCite this: *Chem. Sci.*, 2022, 13, 2701

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

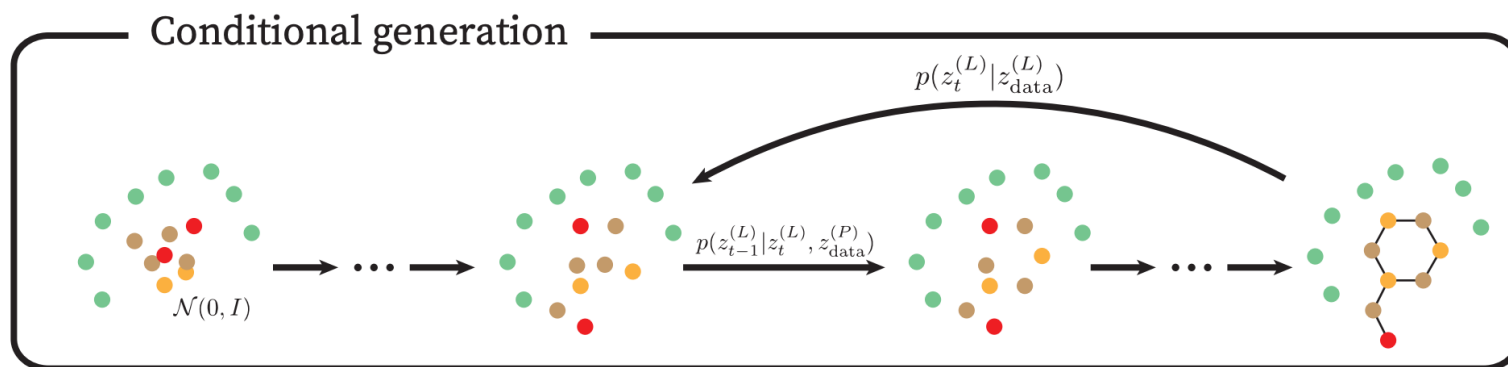
Generating 3D molecules conditional on receptor binding sites with deep generative models†

Matthew Ragoza, ^{lib} ^{**a} Tomohide Masuda^b and David Ryan Koes^c

liggen



Diffusion Models



STRUCTURE-BASED DRUG DESIGN WITH EQUIVARIANT DIFFUSION MODELS

Arne Schneuing^{1*}, Yuanqi Du^{2*}, Charles Harris³, Arian Jamasb³, Ilia Igashov¹, Weitao Du⁴, Tom Blundell³, Pietro Lió³, Carla Gomes², Max Welling⁵, Michael Bronstein⁶ & Bruno Correia¹

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

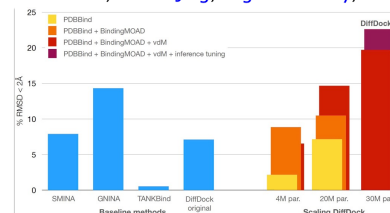
arXiv > q-bio > arXiv:2210.01776

Quantitative Biology > Biomolecules

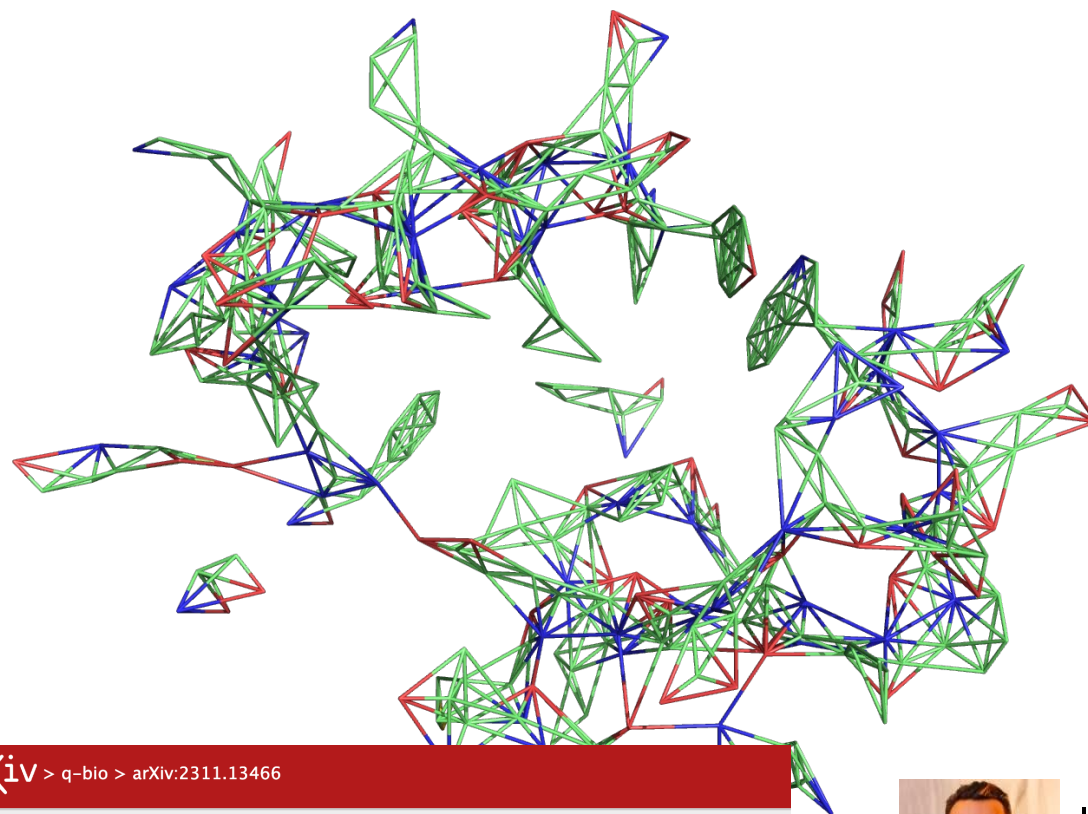
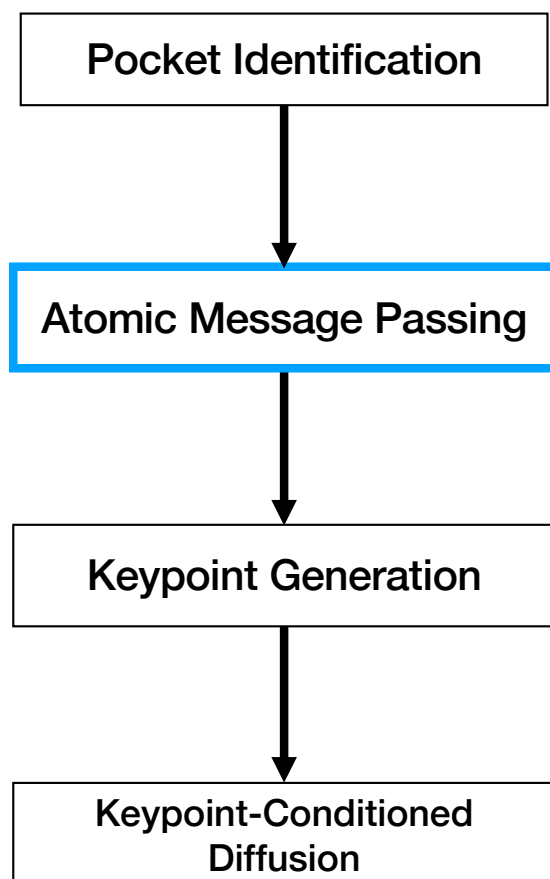
[Submitted on 4 Oct 2022 (v1), last revised 11 Feb 2023 (this version, v2)]

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

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



Keypoint Conditioned Diffusion



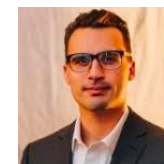
arXiv > q-bio > arXiv:2311.13466

Quantitative Biology > Biomolecules

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

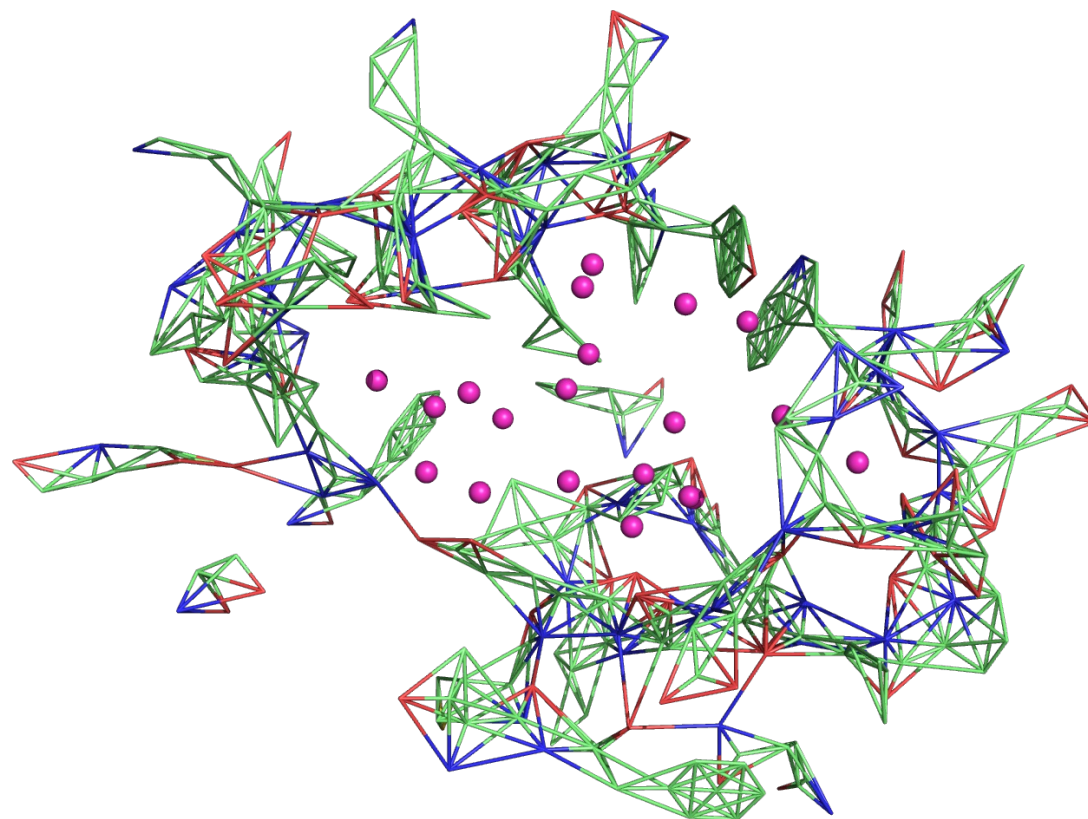
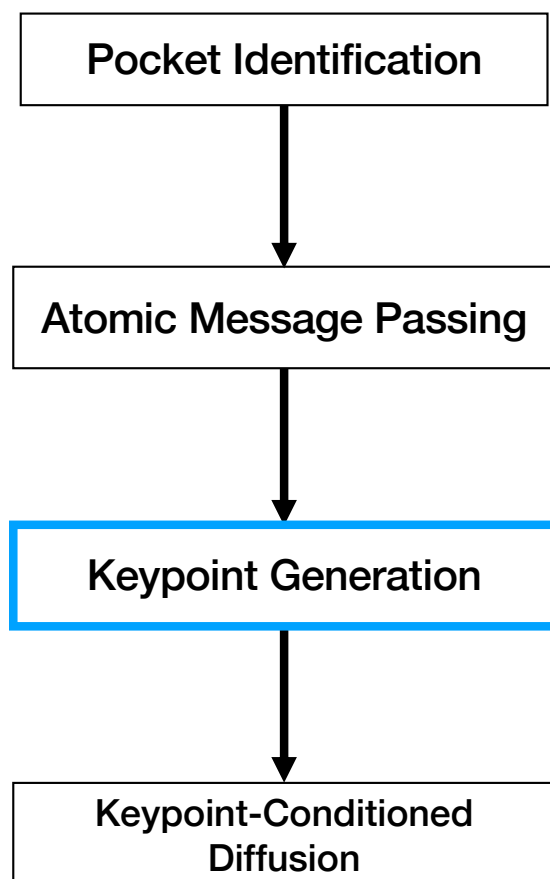
Accelerating Inference in Molecular Diffusion Models with Latent Representations of Protein Structure

Ian Dunn, David Ryan Koes

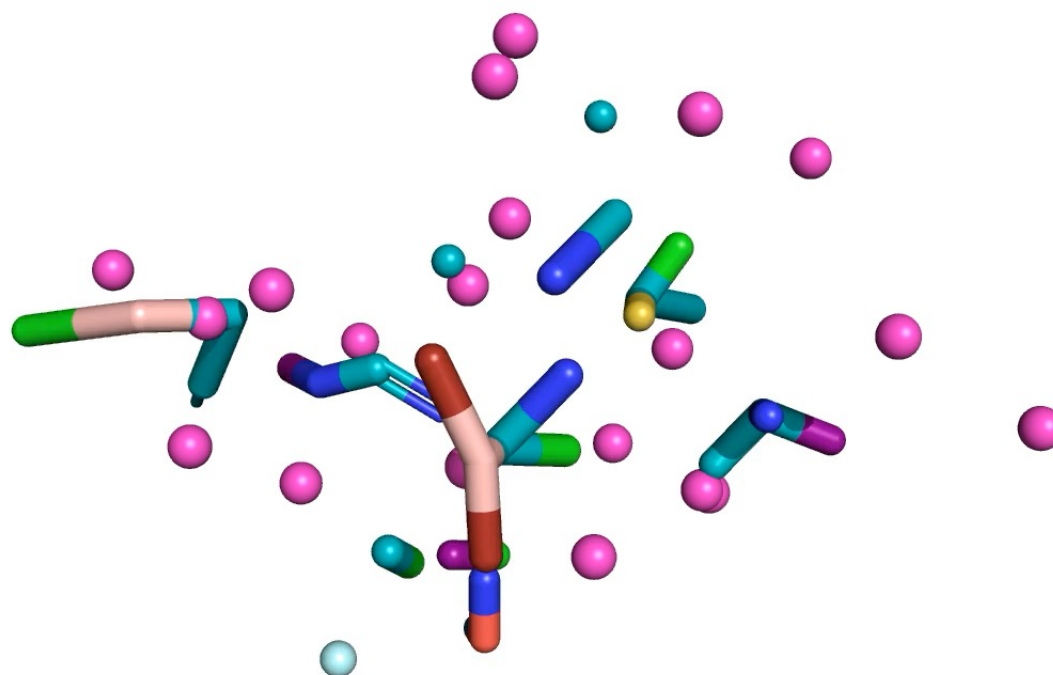
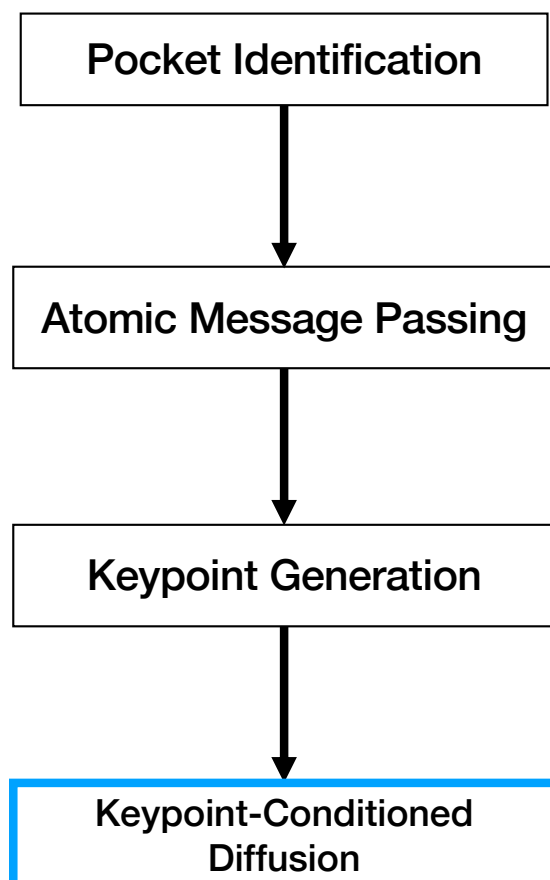


Ian
Dunn

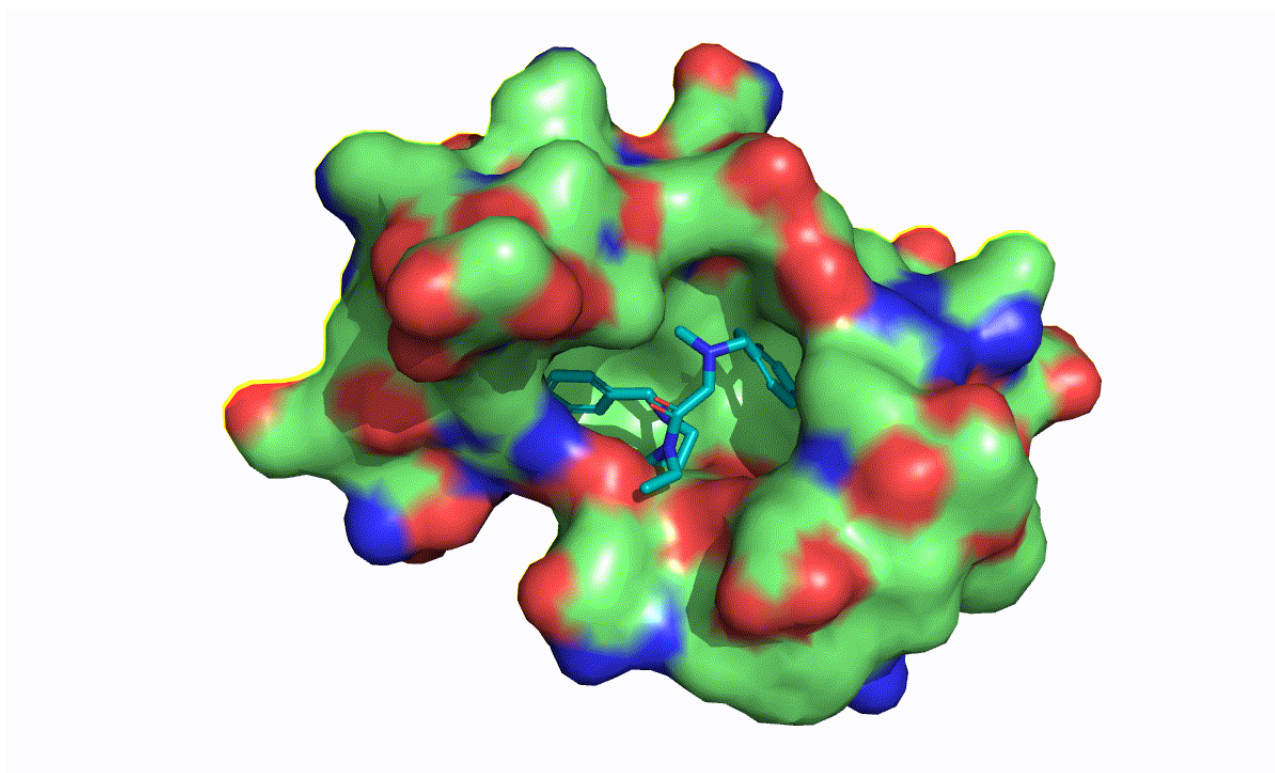
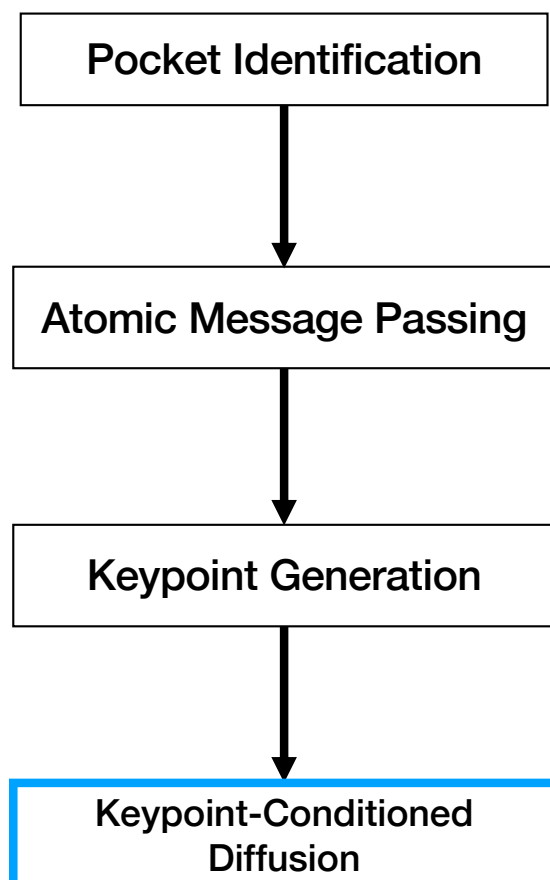
Keypoint Conditioned Diffusion



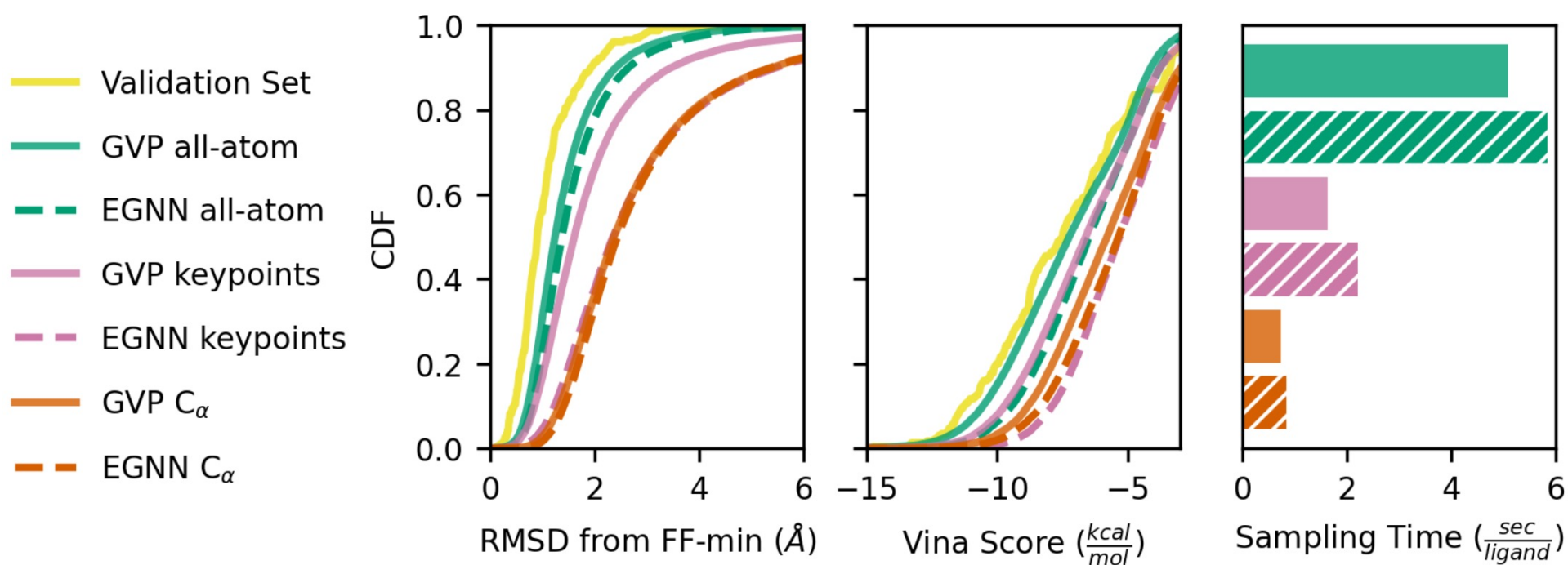
Keypoint Conditioned Diffusion



Keypoint Conditioned Diffusion

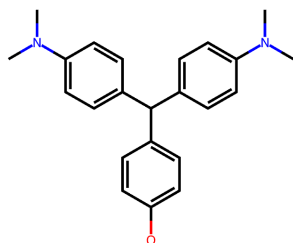
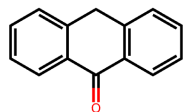
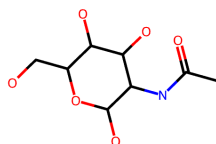


Keypoint Conditioned Diffusion

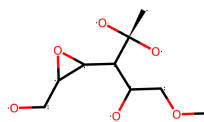


Generated molecules are not synthetically accessible or physically plausible

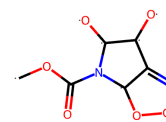
Reference Molecules



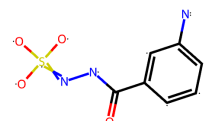
Generated Molecules



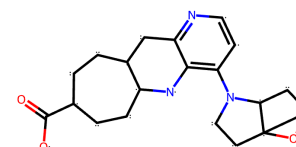
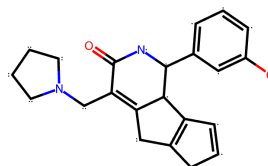
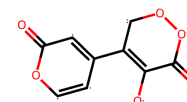
"epoxide, good luck with stability"



"f*** no"



"no"



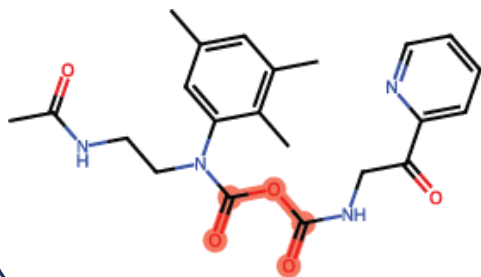
"certainly unstable"

Practical Measures of Molecule Quality

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

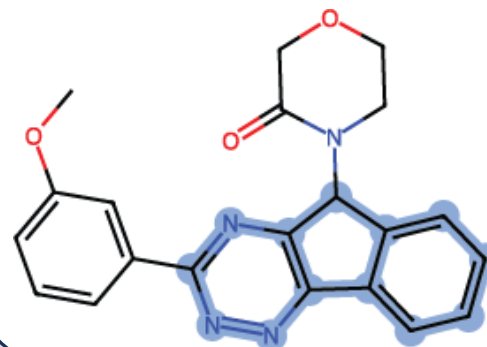
We propose to evaluate molecule quality at the level
of **functional groups** and **ring systems**

Structural Alerts



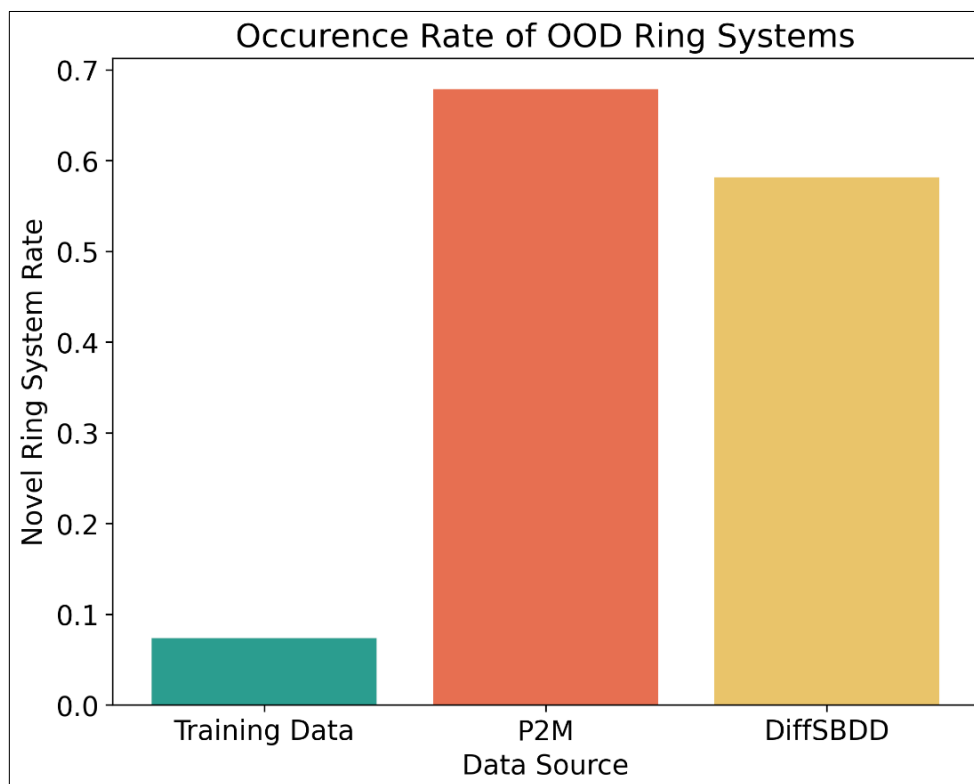
Functional groups
typically used to
filter libraries in
virtual screening
workflows

OOD Rings

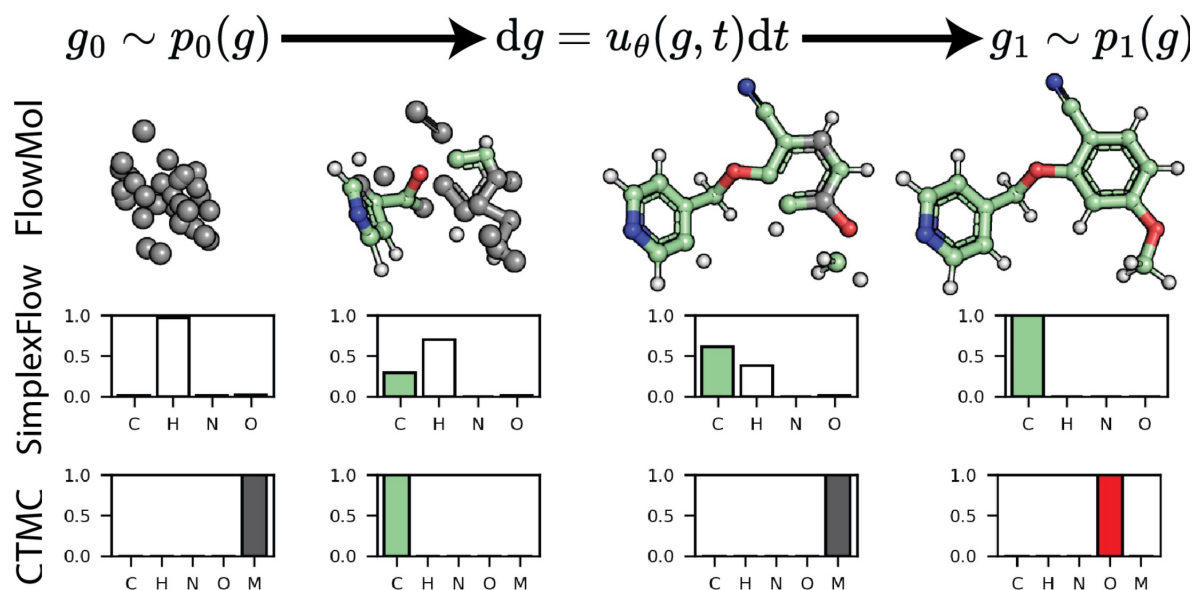


Ring systems that
never appear in
ChEMBL

Practical Measures of Molecule Quality



Unconditional Generation with FlowMol



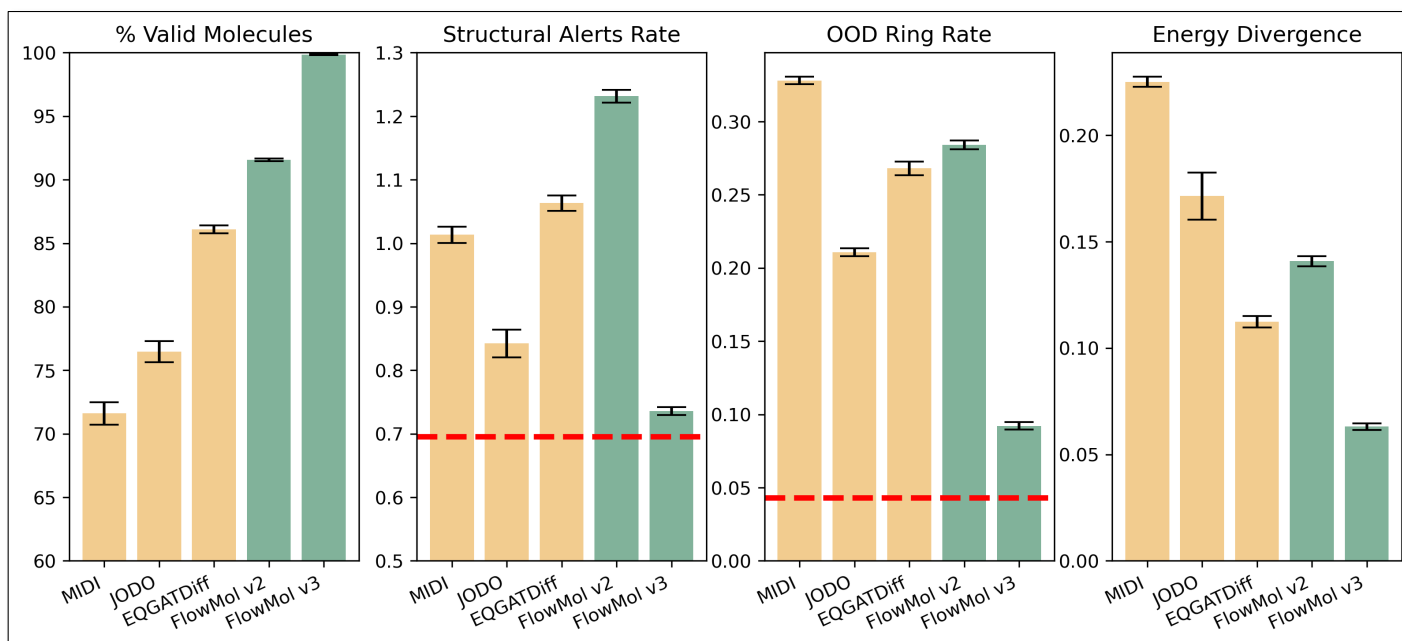
Exploring Discrete Flow Matching for 3D De Novo Molecule Generation



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



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

FlowMol Public

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

Python ☆ 105 🍴 5

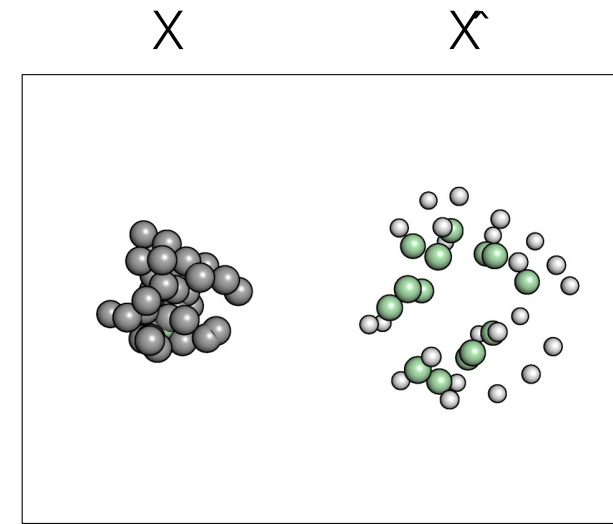
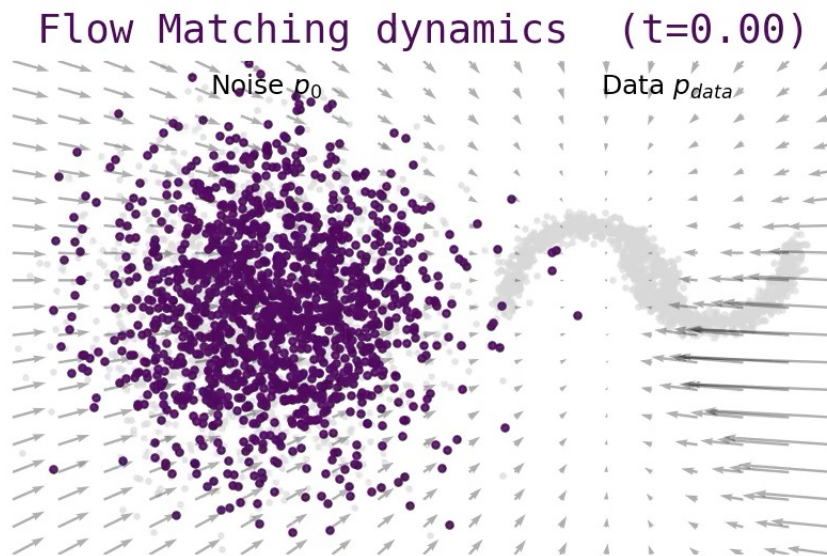
FlowMol3: Flow Matching for 3D De Novo Small-Molecule Generation

[Ian Dunn](#) and [David Ryan Koes](#)



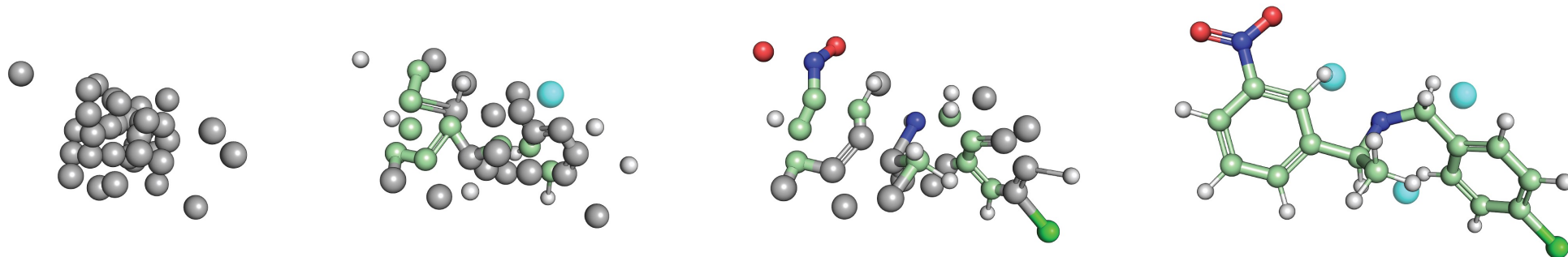
Why is FlowMol3 Better?

$$g_0 \sim p_0(g) \longrightarrow dg = u_\theta(g, t)dt \longrightarrow g_1 \sim p_1(g)$$

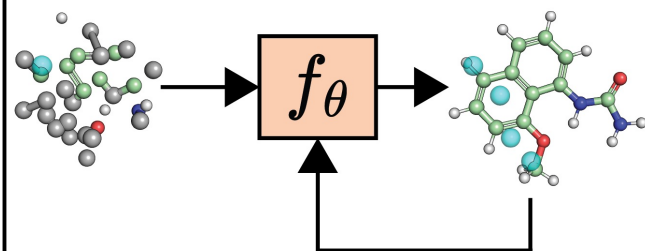


Why is FlowMol3 Better?

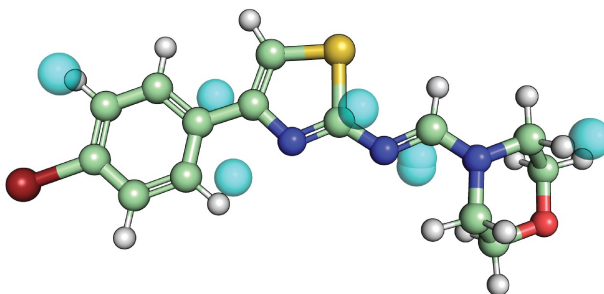
$$g_0 \sim p_0(g) \longrightarrow dg = u_\theta(g, t)dt \longrightarrow g_1 \sim p_1(g)$$



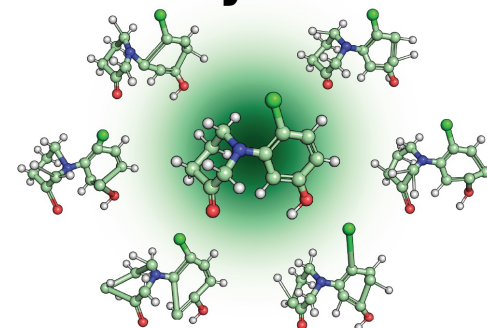
Self-Conditioning



Fake Atoms



Geometry Distortion

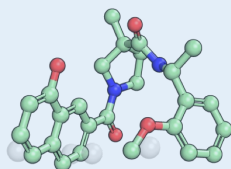


OMTRA

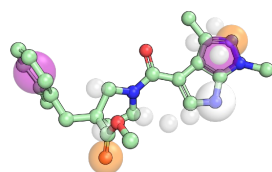
De novo Generation

FlowMol3

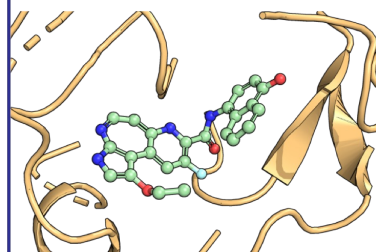
Unconditional



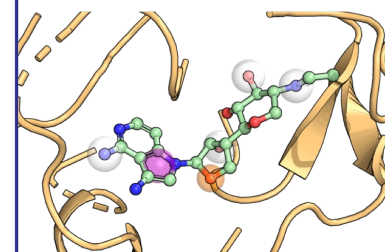
Pharmacophore Conditioned



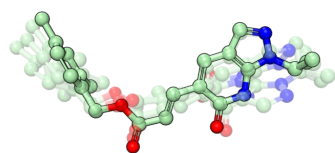
Pocket Conditioned



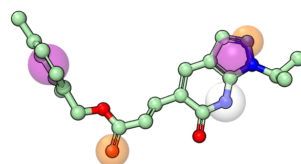
Pocket & Pharm. Conditioned



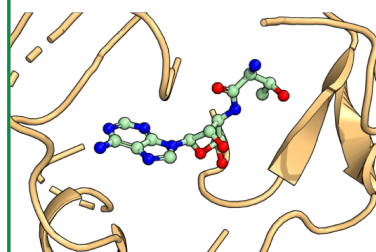
Unconditional



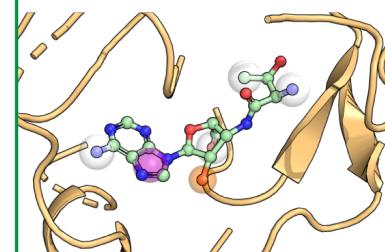
Pharmacophore Conditioned



Pocket Conditioned



Pocket & Pharm. Conditioned



Conformer Generation

Docking

Ian Dunn



Tyler Katz



Liv Toft



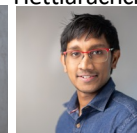
Riya Shah



Juhi Gupta



Ramith
Hettiarachchi

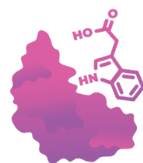


*2025 Workshop on Machine
Learning in Structural Biology*

Data

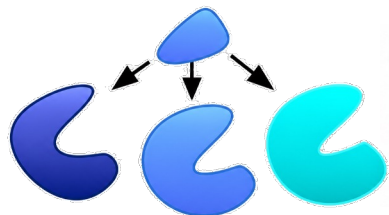
pharmit

>500 million 3D molecules



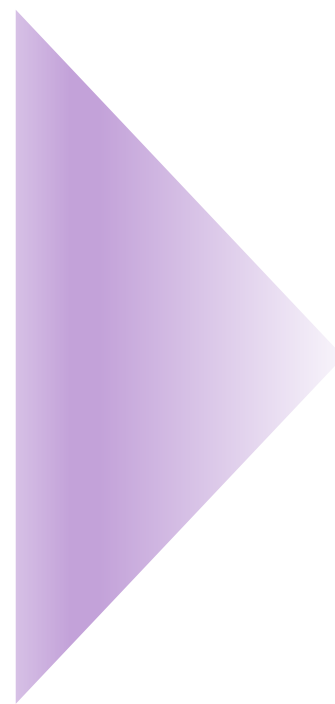
plinder

450k Protein-Ligand Interactions



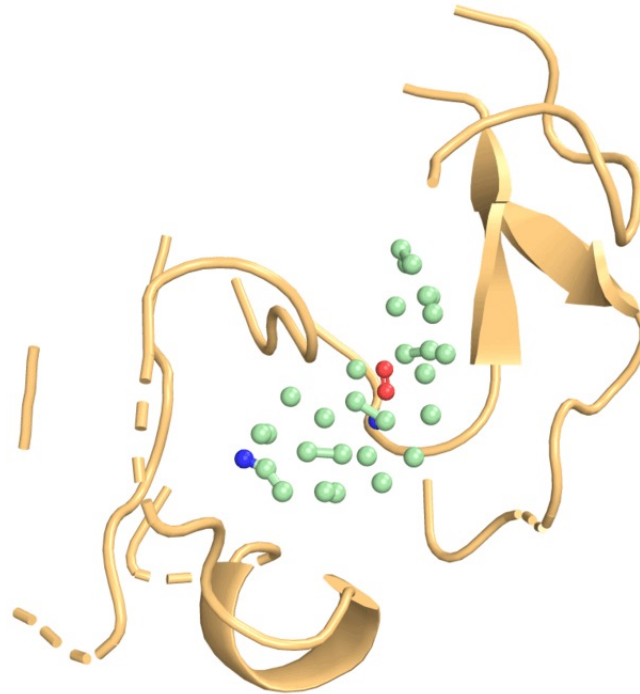
CrossDocked2020

18,450 complexes
22.6 million poses



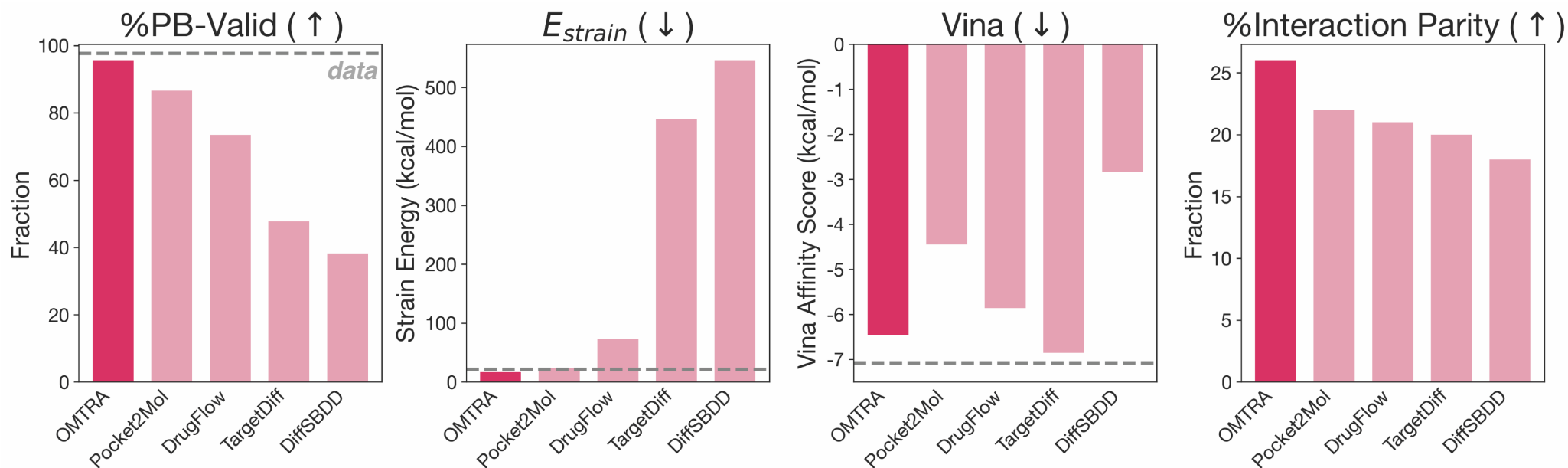
OMTRA

OMTRA: De Novo Design

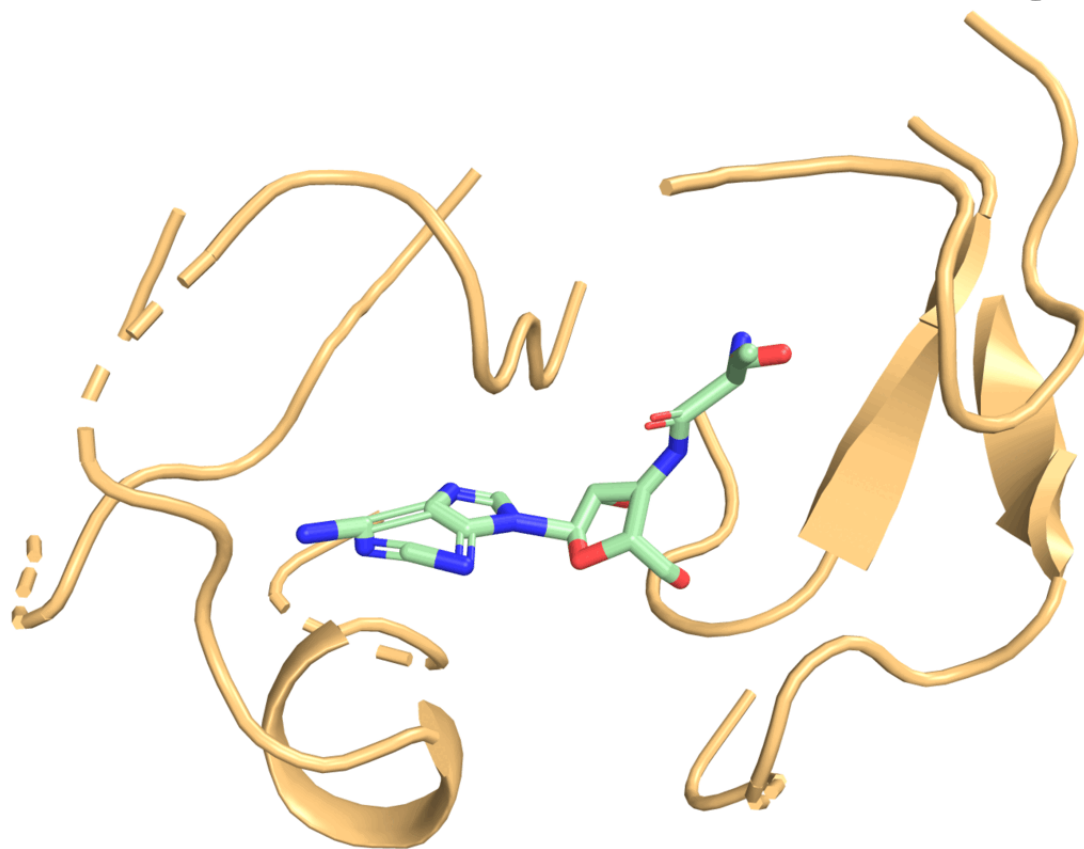


OMTRA: De Novo Design

Evaluated on Luo et al CrossDocked test set.



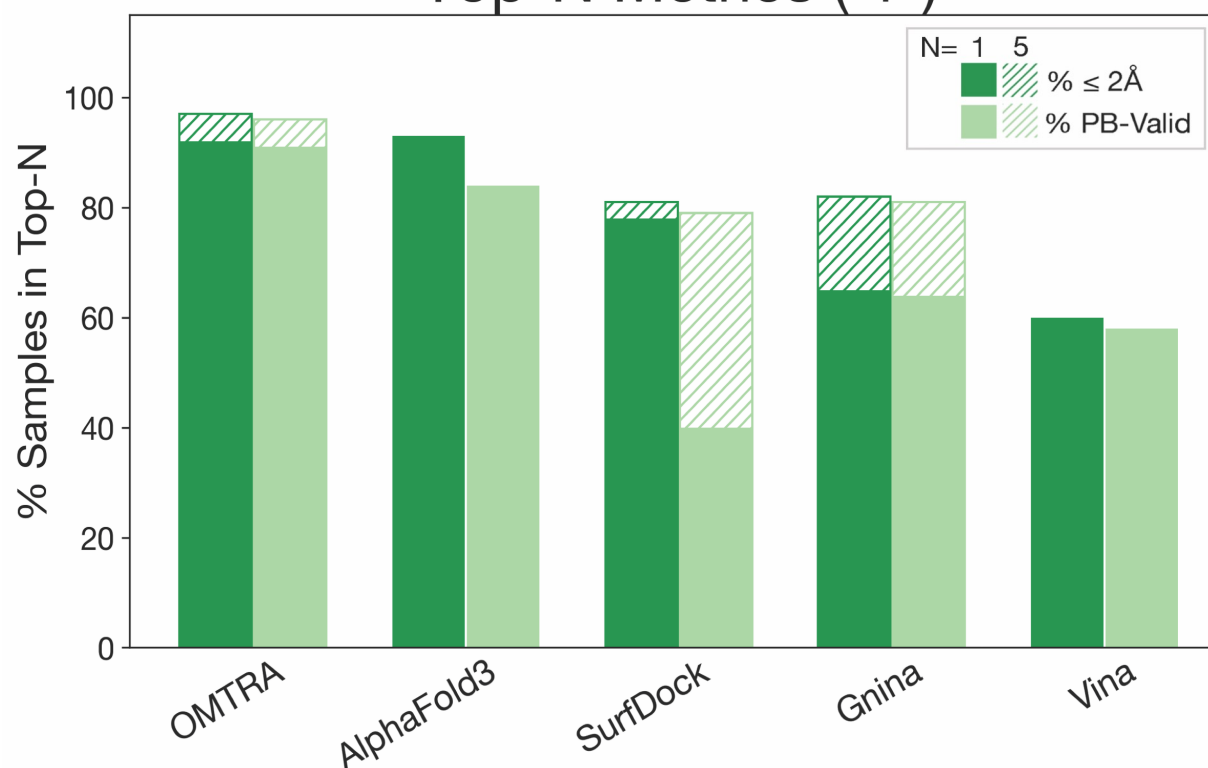
OMTRA: Docking



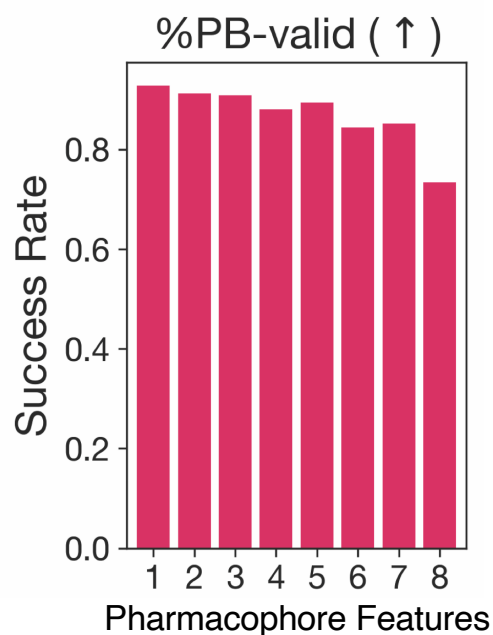
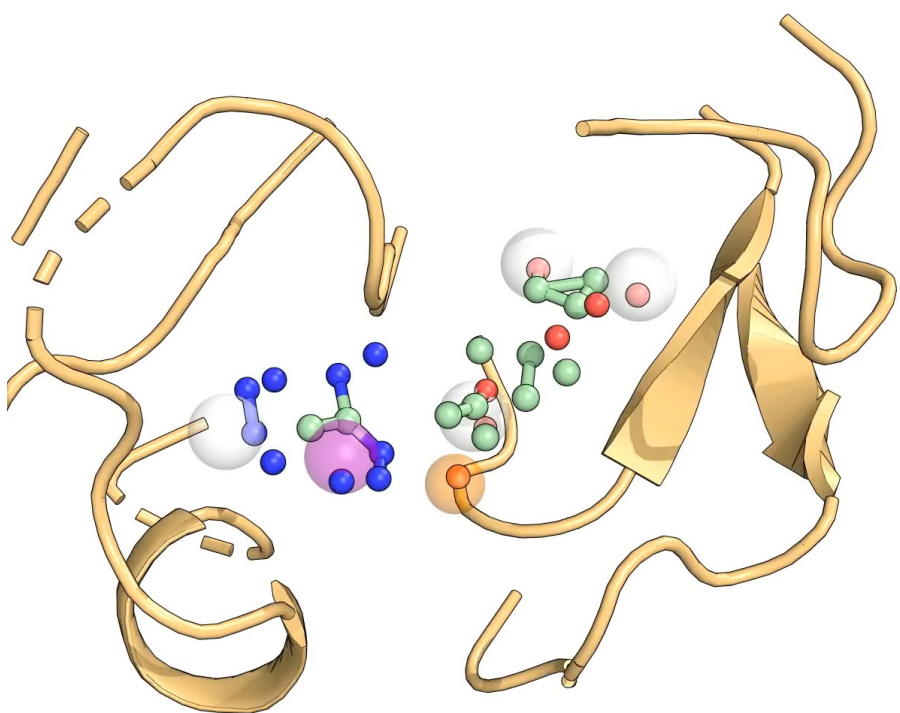
OMTRA: Docking

Evaluated on PoseBusters test set.

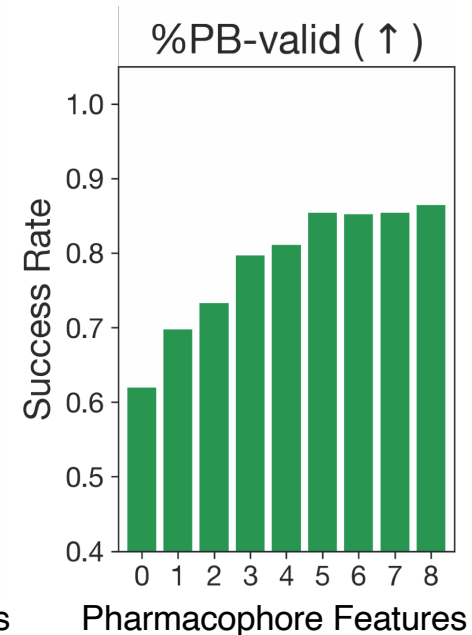
Top-N Metrics (↑)



OMTRA: Pharmacophore Conditioning



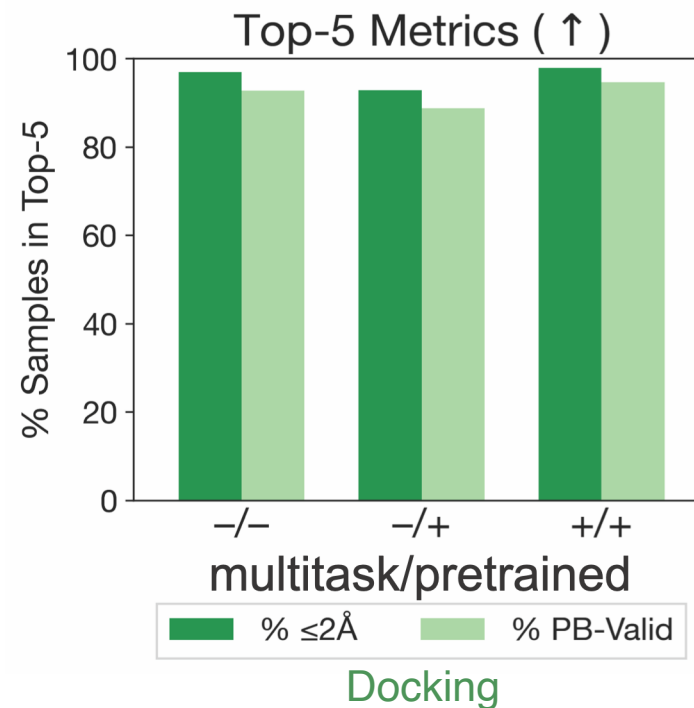
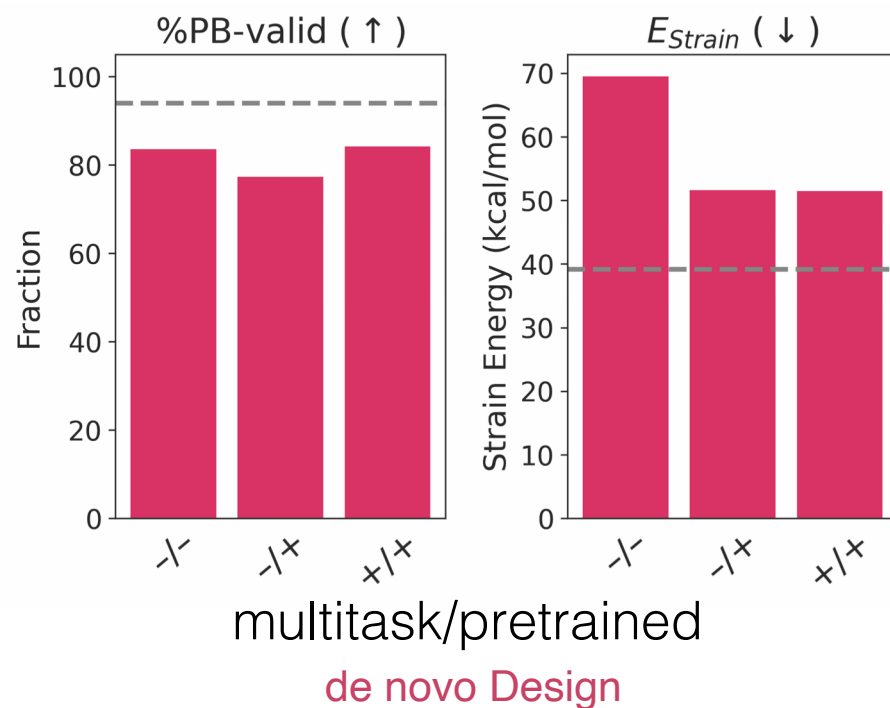
de novo Design



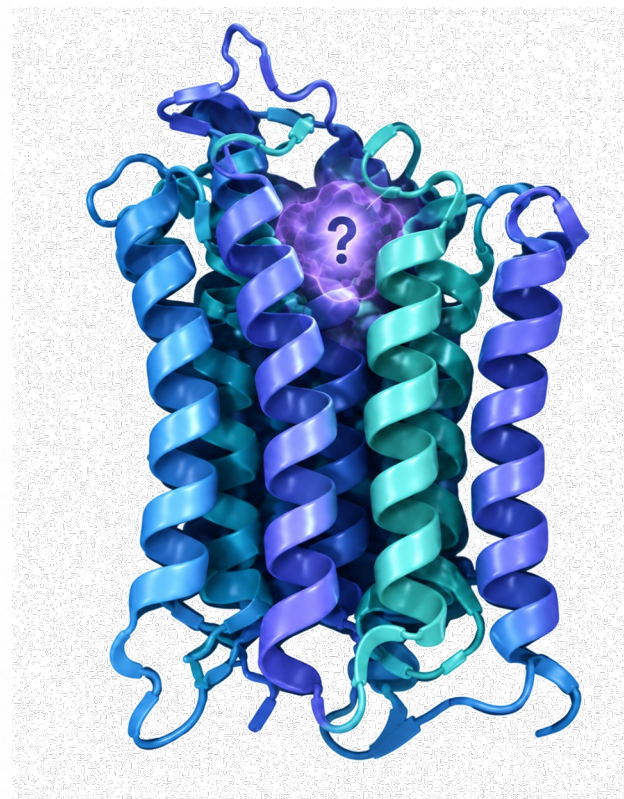
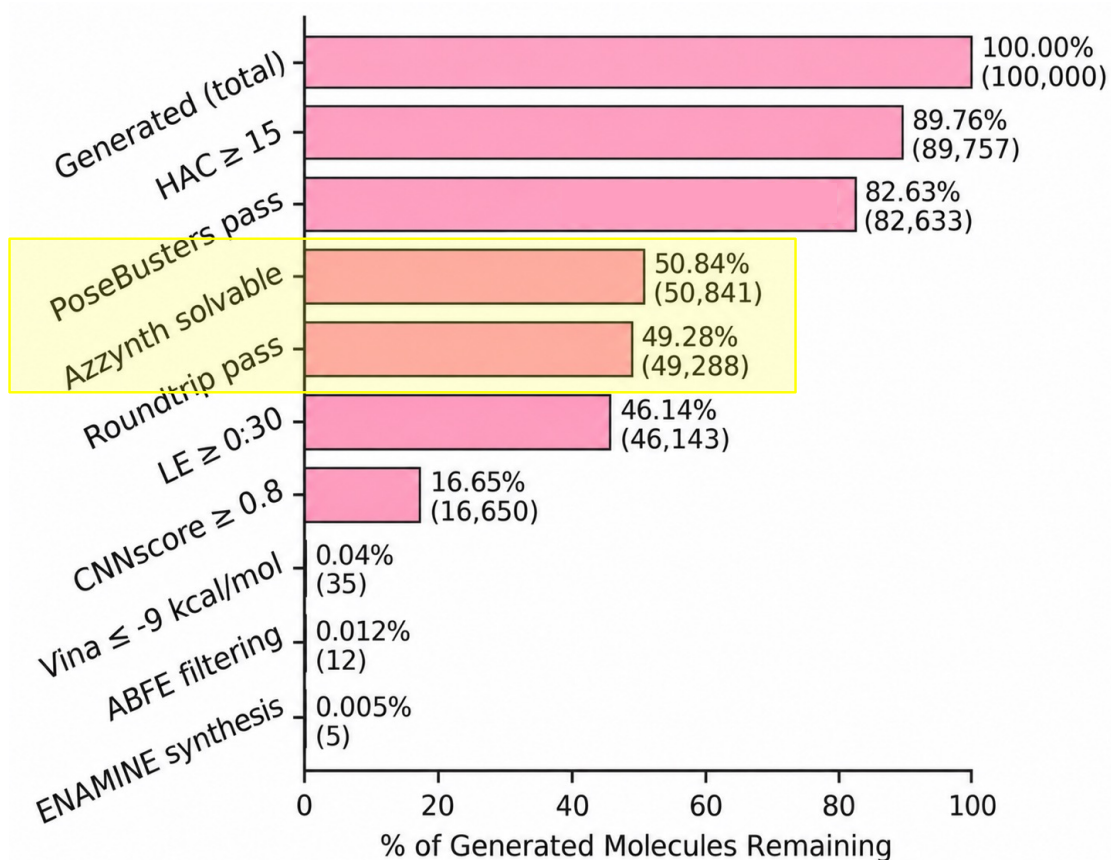
Docking

OMTRA: Transfer Learning?

Evaluated on PLINDER test set.



Synthesizeability



Web App

Input Files

1. Upload Protein

Upload PDB or CIF file defining the target.

receptor.pdb
PROCESSED

2. Binding Site Definition

Recommend deselecting protein surface to view pocket.

Ref Ligand Detected Manual

Note: Defines the pocket as residues 8Å from reference ligand atoms.

ligand.sdf
PROCESSED

CENTER
[16.75, 23.93, 4.05]

Style
☐ Sticks
☐ Surface
☒ Backbone (Cartoon)
☒ Pockets

Fix atoms (optional) 14

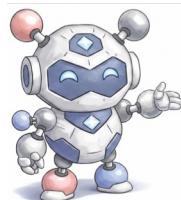
Fixed atoms keep the same position during generation.

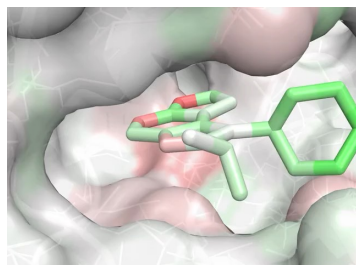
By fragment By atom

Click fragments in the viewer or toggle below. Click to toggle · Shift+drag to box-select.

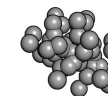
☐ Frag 0 (27) ☐ Frag 1 (1)
☒ Frag 2 (2) ☒ Frag 3 (2)
☐ Frag 4 (1) ☒ Frag 5 (10)

All None
3 of 6 fragments fixed
14 atoms fixed Clear Invert

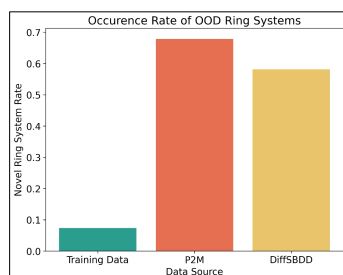
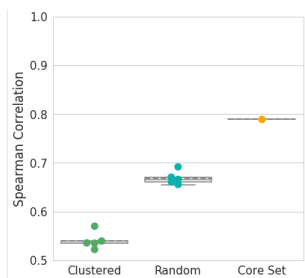




Takeaways



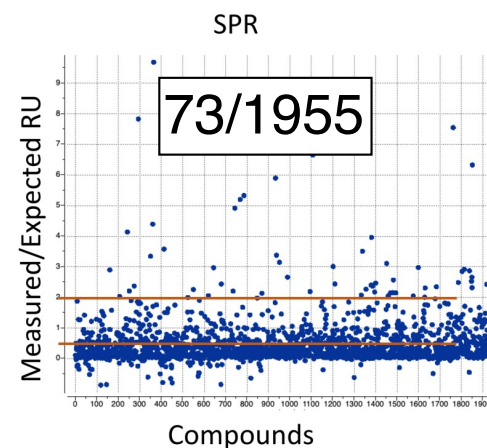
Deep learning works for molecules!



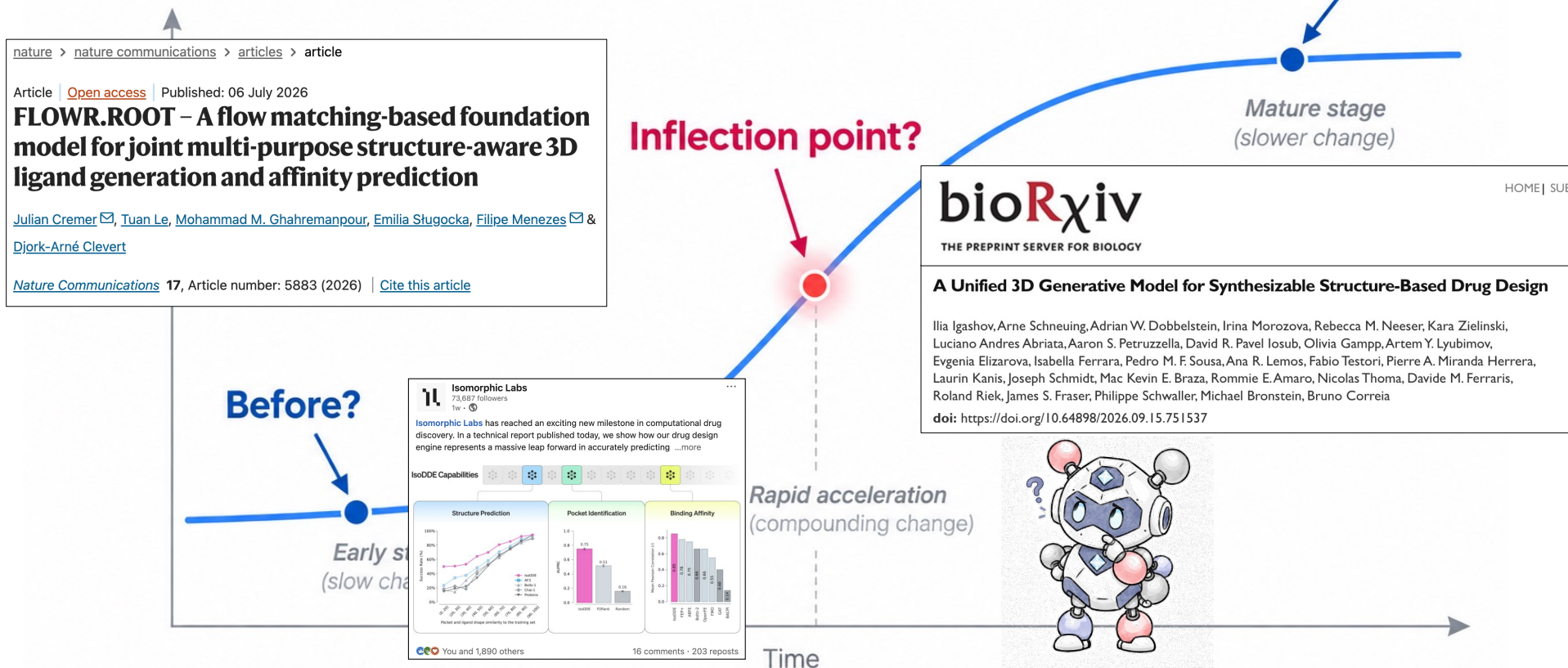
Rigorous evaluation is essential

No replacement for prospective evaluation

Still pretty bad ... but ...



Are we at an inflection point?



Acknowledgements



Current

Emma Flynn
Riya Shah
Rishal Aggarwal
Daniel Penaherrera
Jacky Chen
Fareeda Abu-Juam
Ben Krummenacher
Marios Gavrielatos

Previous

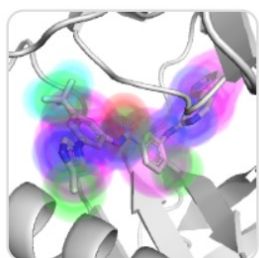
Ian Dunn
Drew McNutt
Jocelyn Sunseri
Matthew Ragoza
Tomohide Masuda
Paul Francoeur
Jonathan King
Rocco Meli
Josh Hochuli
Elisa Idrobo
Lily Turner
Alec Helbling
Andrew Jia
Rich Iovanisci
Ian Snyder
Nick Rego



National Institute of
General Medical Sciences



Shameless Plugs

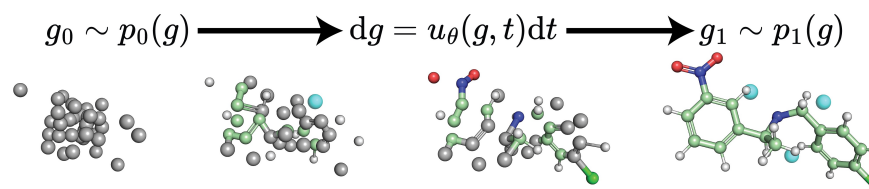


gnina
(NEE-na)

<https://github.com/gnina>

FlowMol

<https://github.com/Dunni3/FlowMol>



py3Dmol

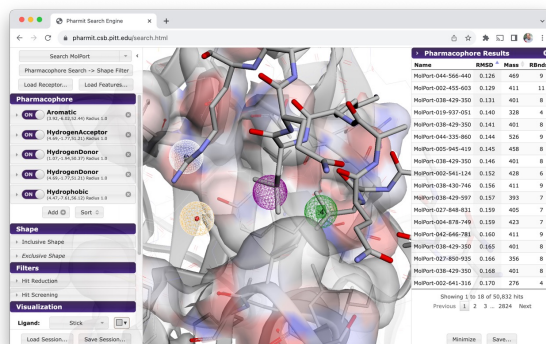
```
[ ] import py3Dmol

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



<https://3dmol.org/>

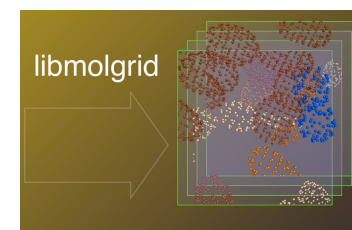
Pharmit



<https://pharmit.csb.pitt.edu>

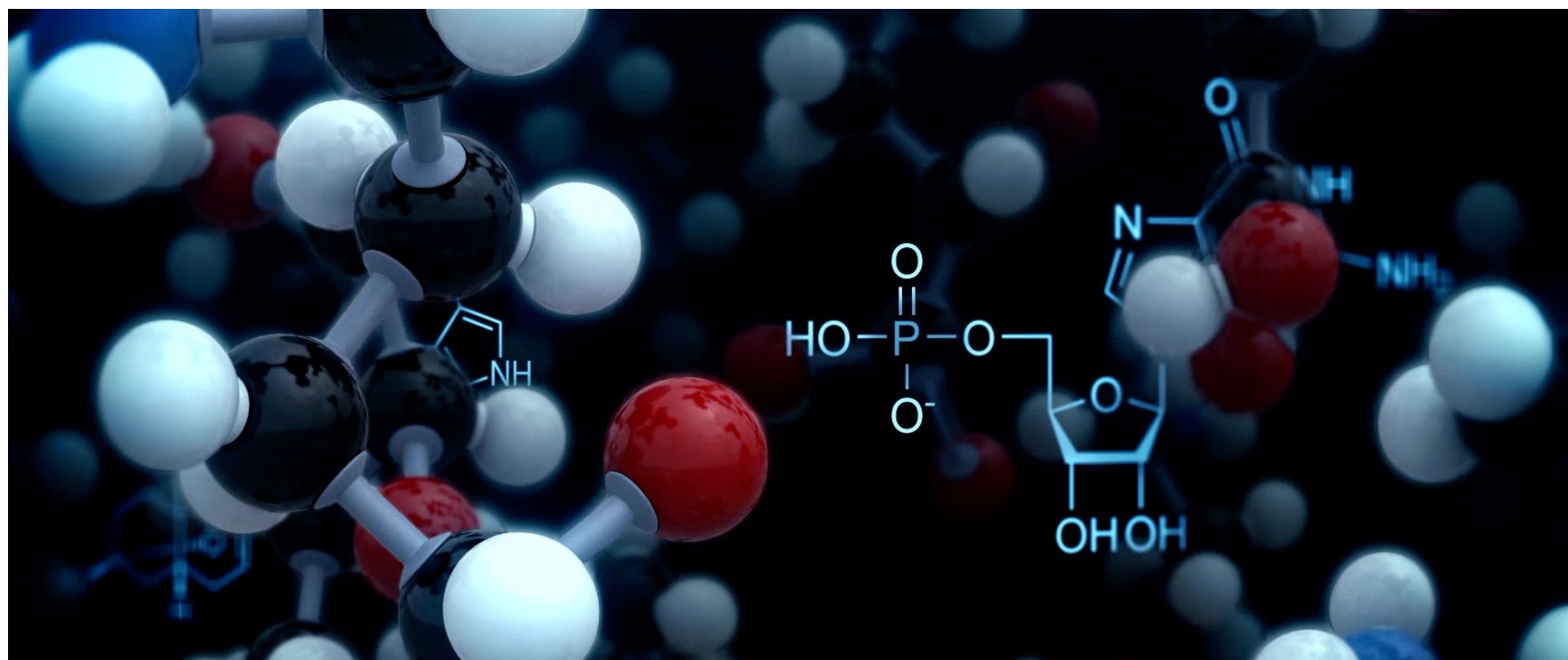


<https://github.com/gnina/OMTRA>

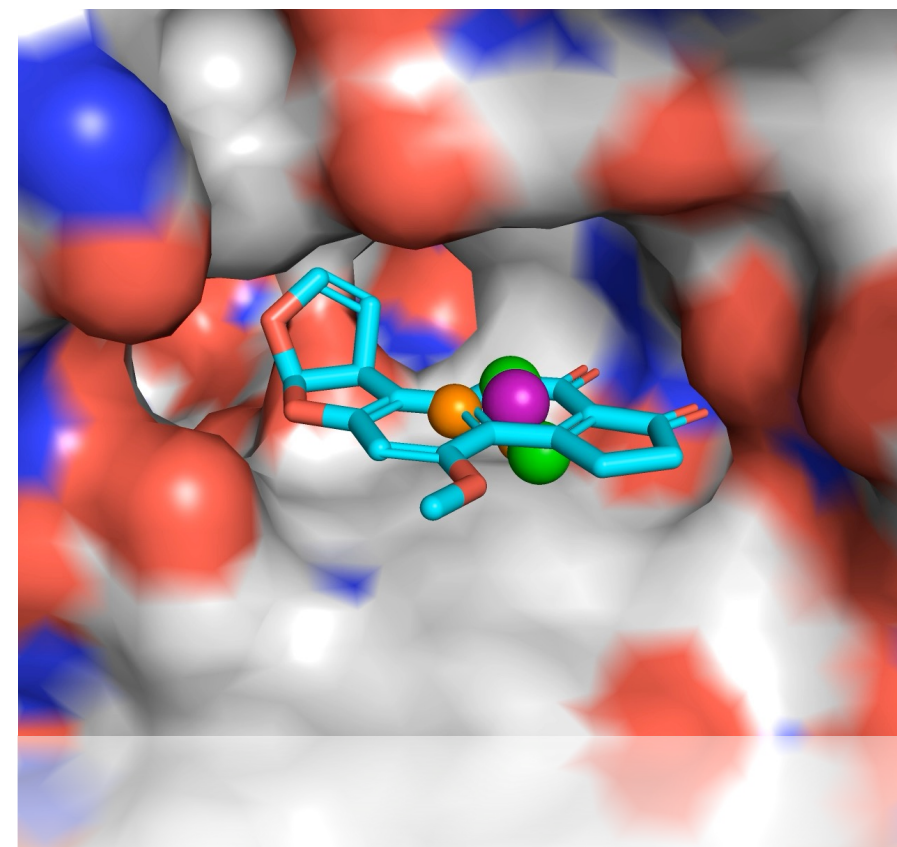
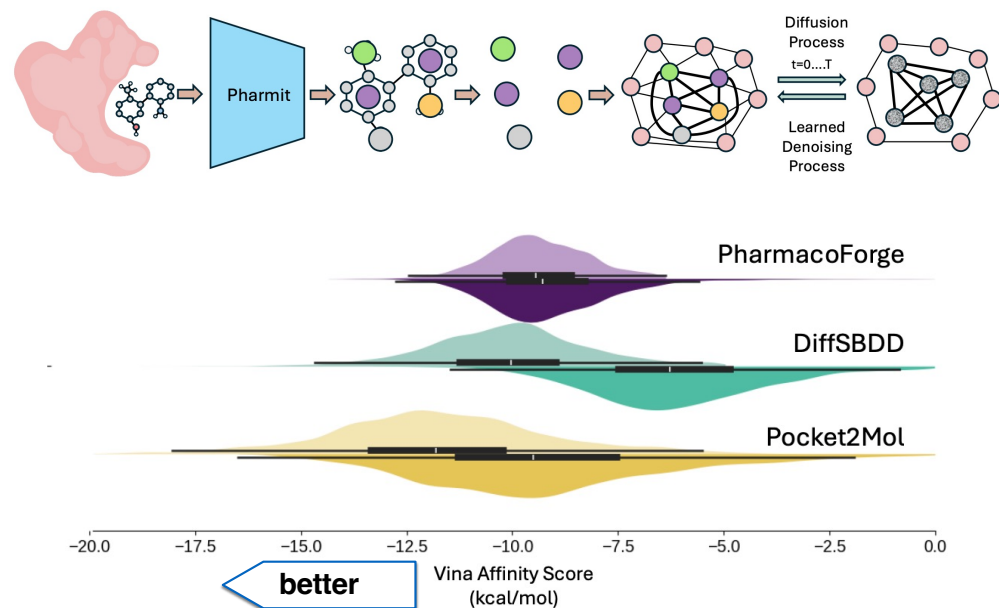


<https://gina.github.io/libmolgrid/>

Generating *Descriptions* of Molecules



PharmacoForge



PharmacoForge: pharmacophore generation with diffusion models

Emma L. Flynn^{1,2}, Riya Shah¹, Ian Dunn^{1,2}, Rishal Aggarwal^{1,2} and David Ryan Koes^{1*}

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/>

