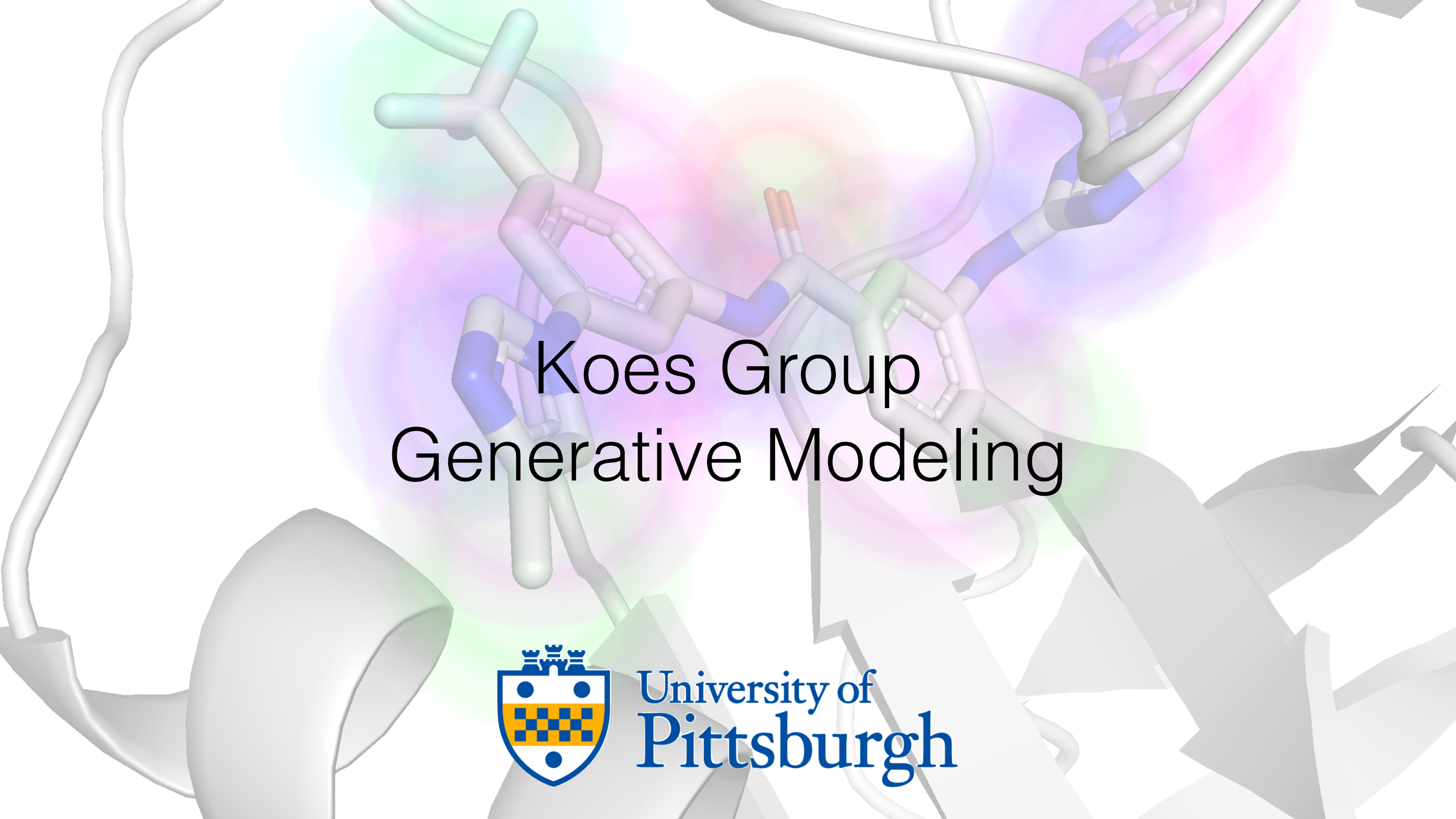
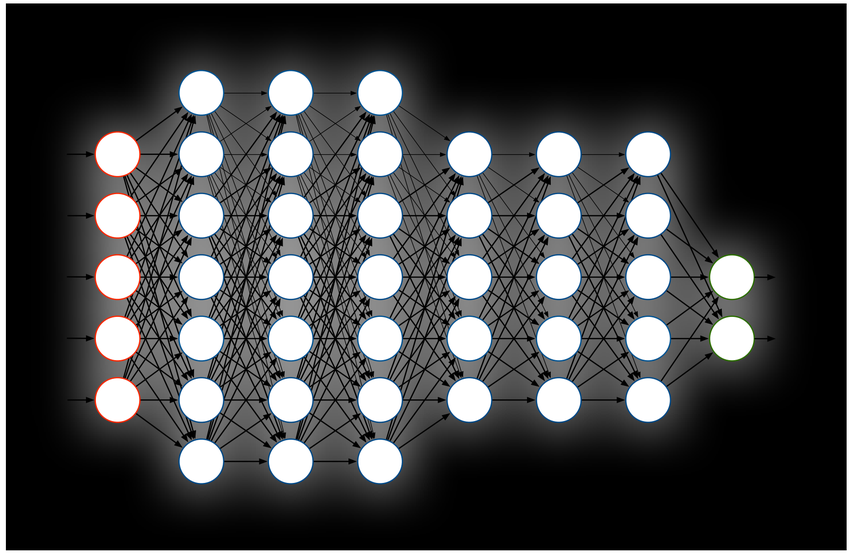


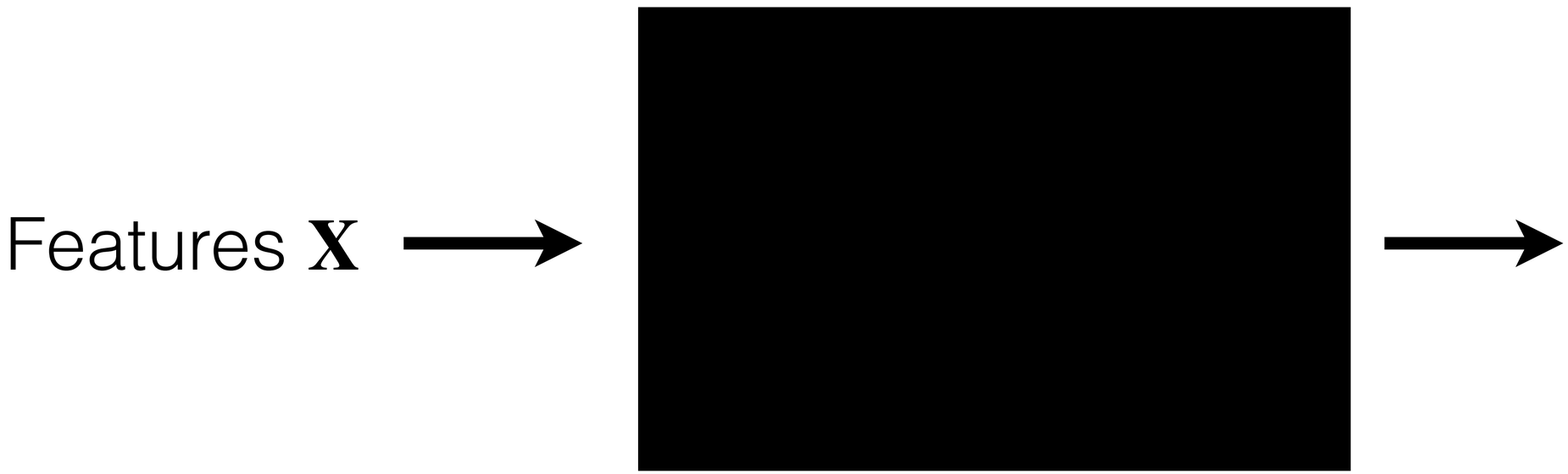
Koes Group
Generative Modeling

University of Pittsburgh

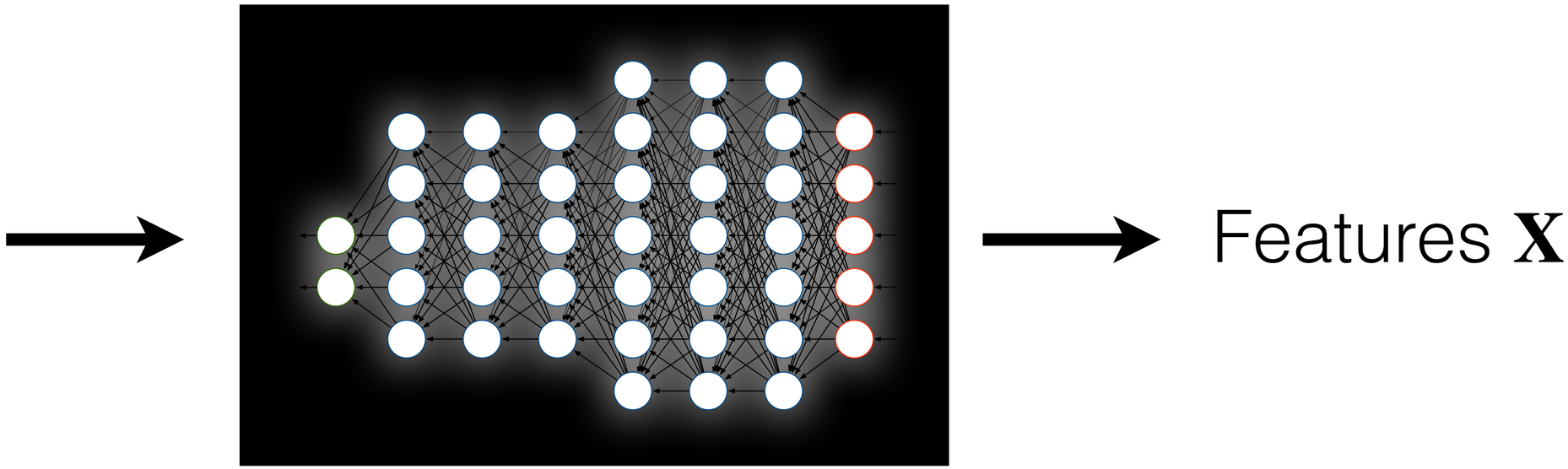
Discriminative Model

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

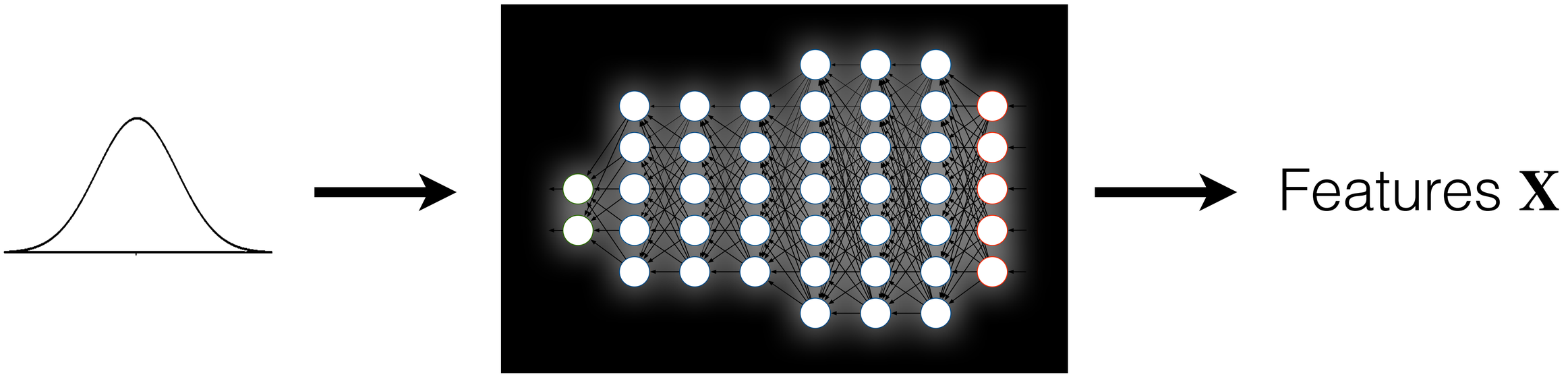
Generative Model



Generative Model



Generative Model



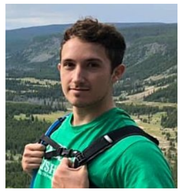
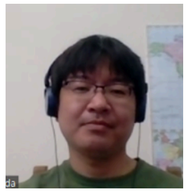
Learning a Continuous Representation of 3D Molecular Structures with Deep Generative Models

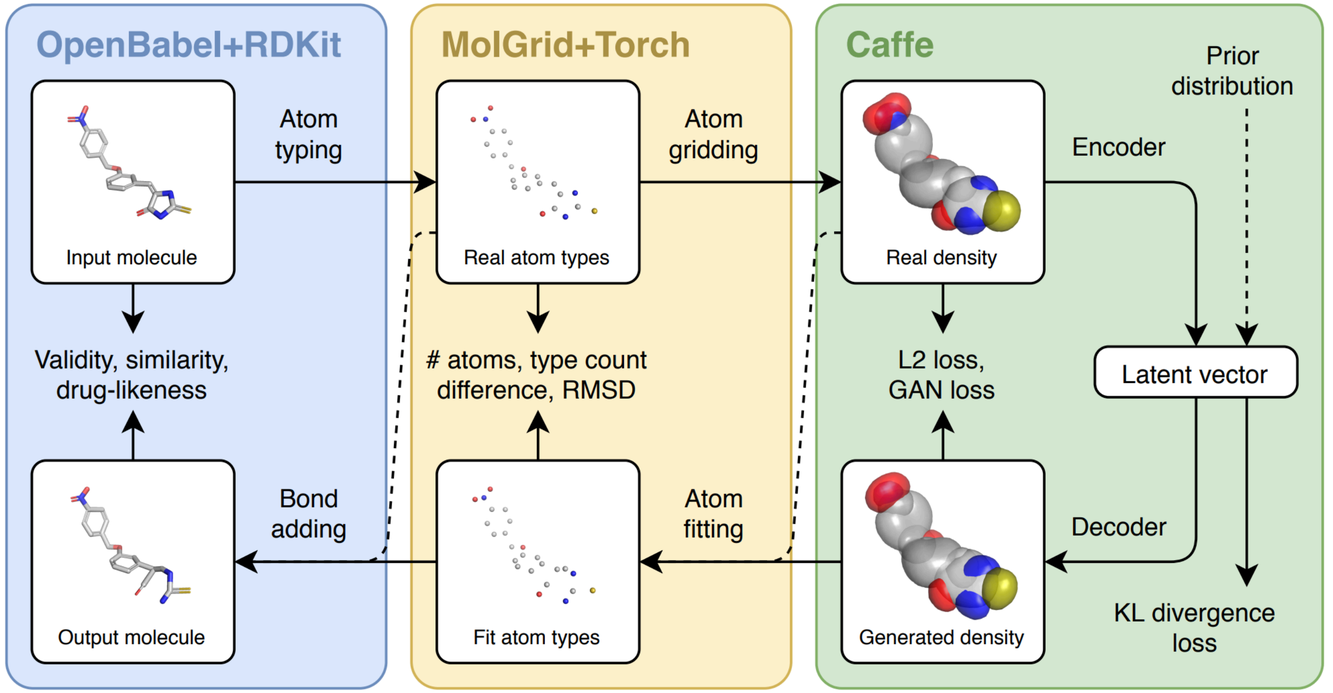
NeurIPS 2020 Workshop
Machine Learning for Structural Biology

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

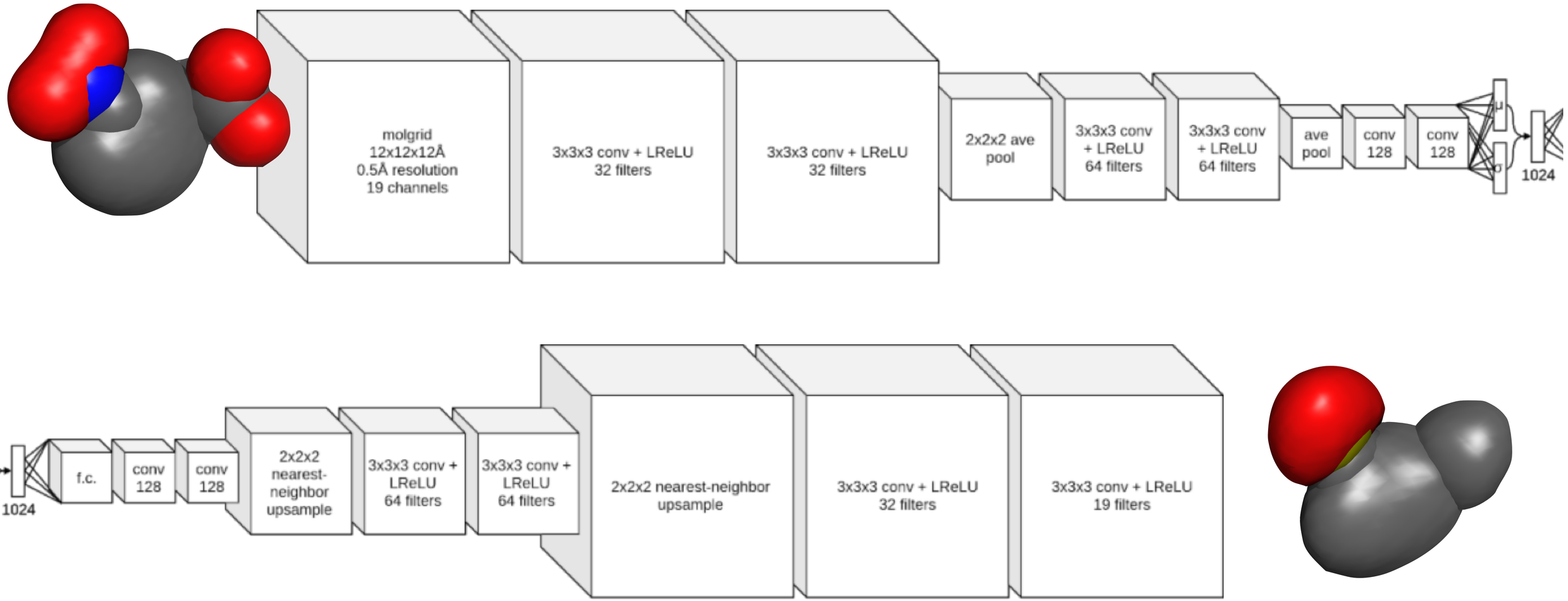
Tomohide Masuda*
Comp. & Systems Biology
University of Pittsburgh
Pittsburgh, PA 15213
tmasuda@pitt.edu

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

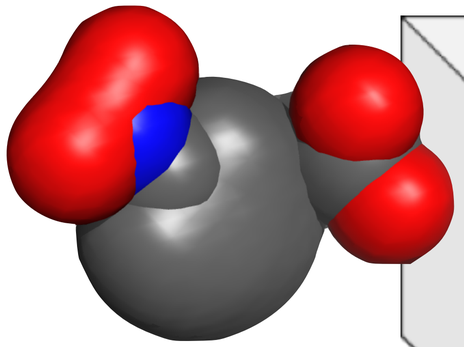


Model


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

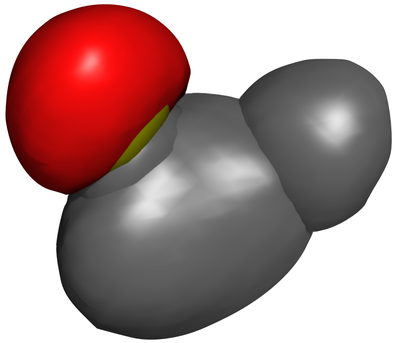
3x3x3 conv +
LReLU
64 filters

3x3x3 conv +
LReLU
64 filters

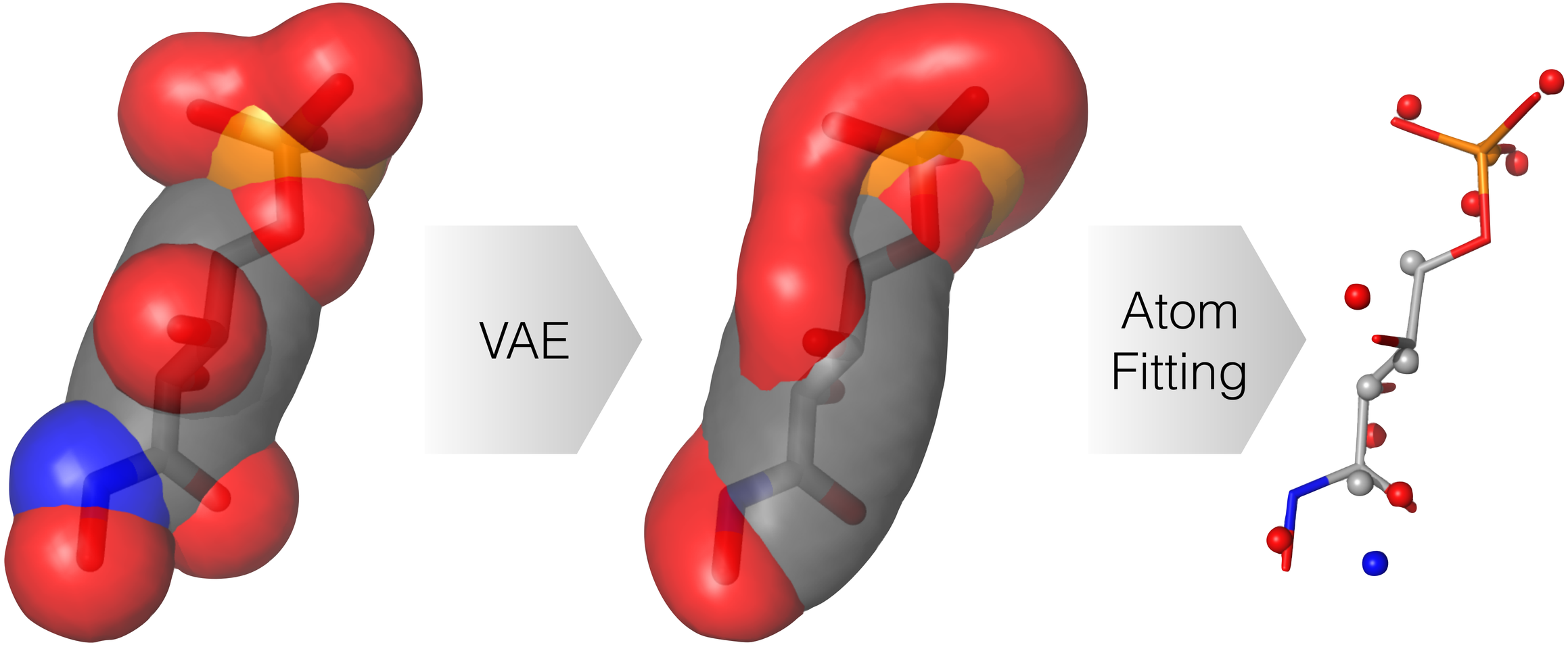
2x2x2 nearest-neighbor
upsample

3x3x3 conv + LReLU
32 filters

3x3x3 conv + LReLU
19 filters



Variational Autoencoding Examples



2BES

Different input molecules

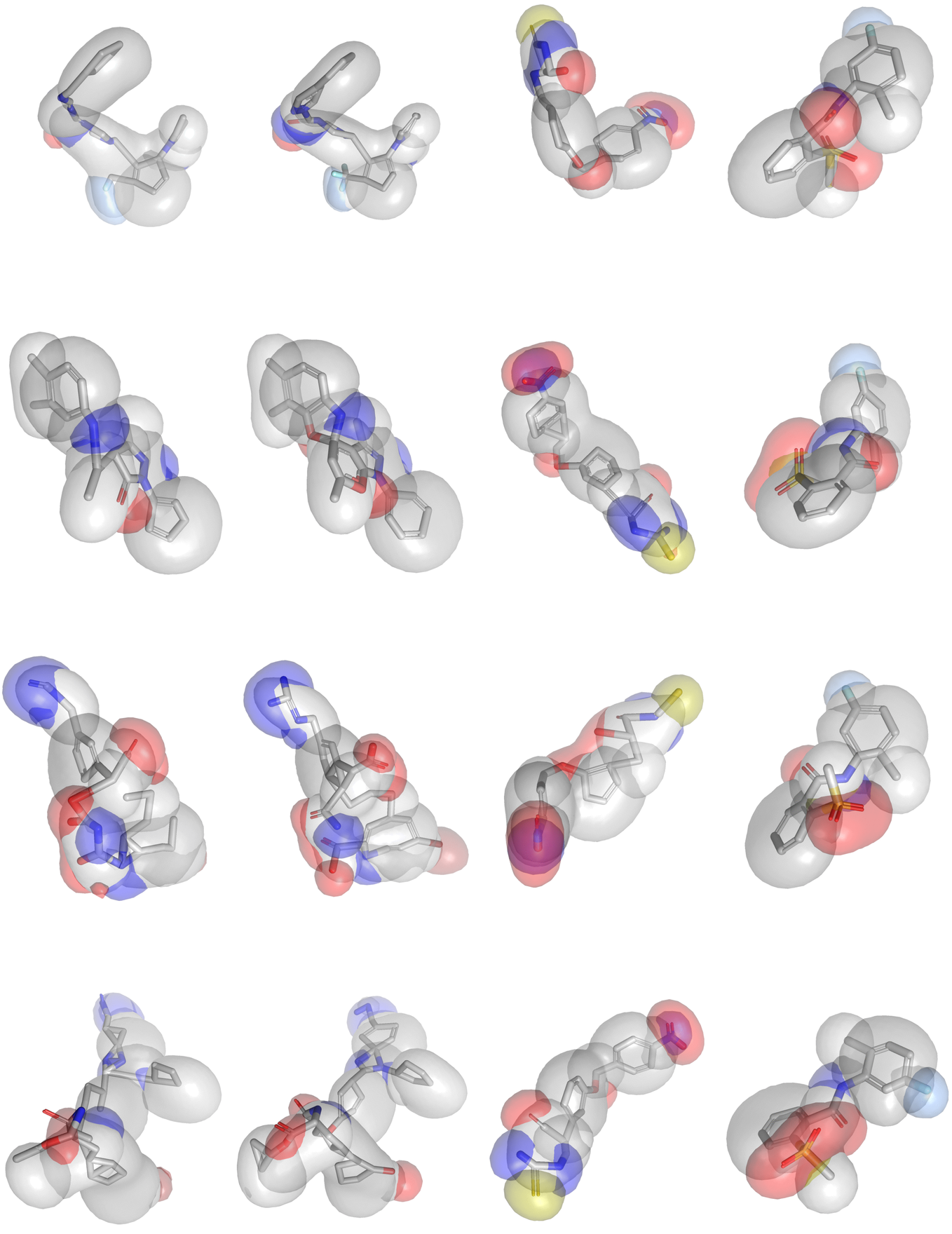
Same input molecules

VAE

AE

VAE

AE



University of Pittsburgh

Computational and Systems Biology

Different input molecules

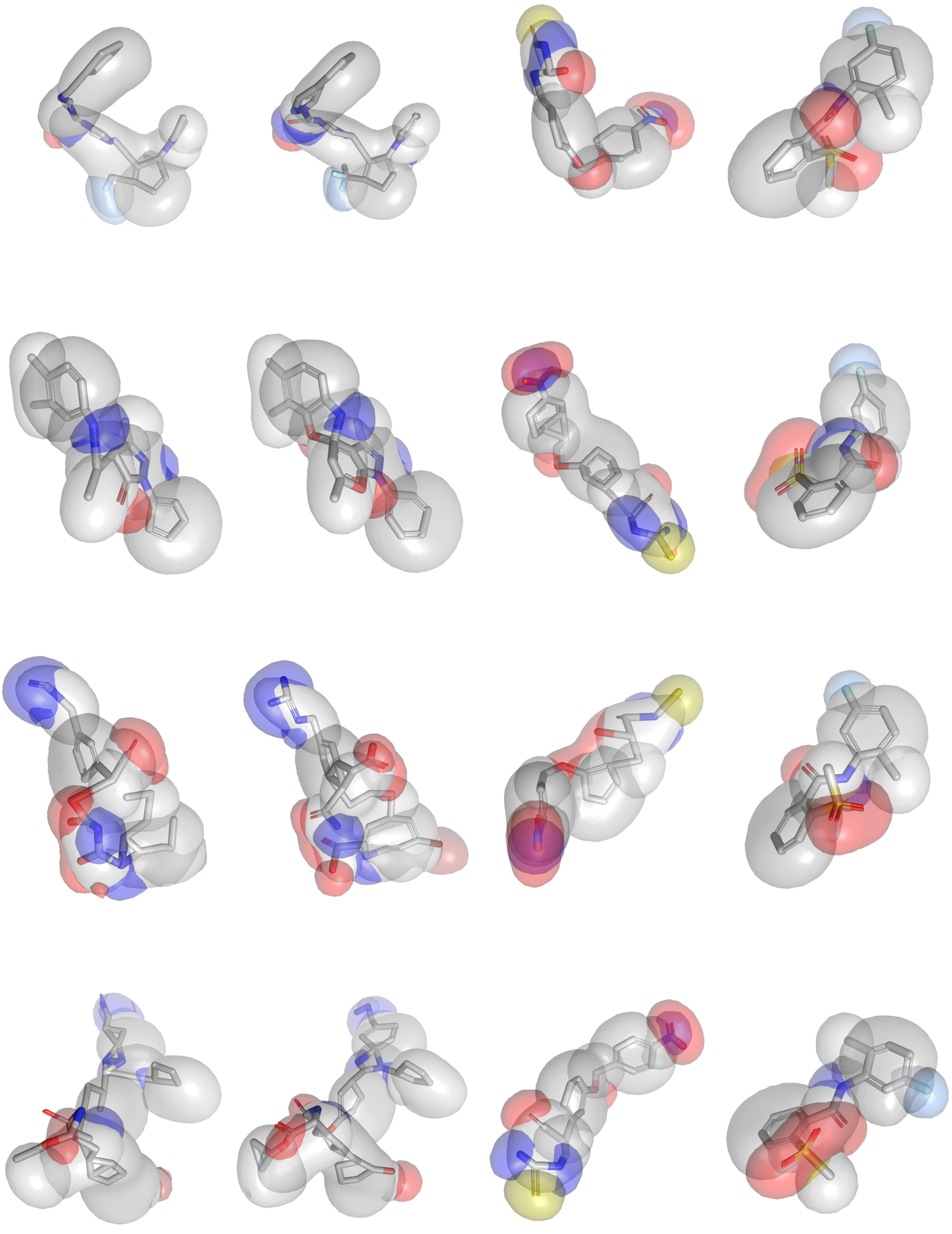
Same input molecules

VAE

AE

VAE

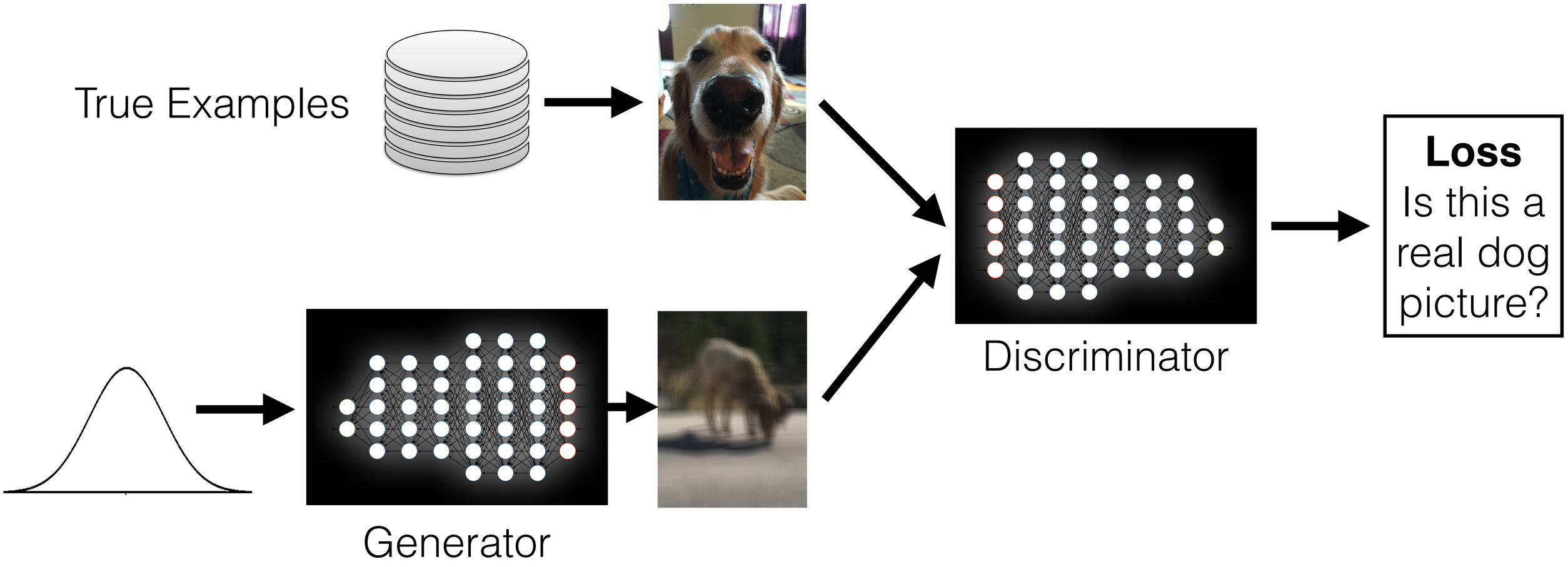
AE



University of Pittsburgh

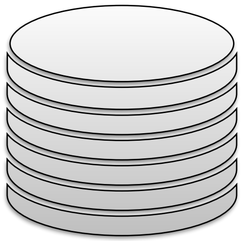
Computational and Systems Biology

Generative Adversarial Networks

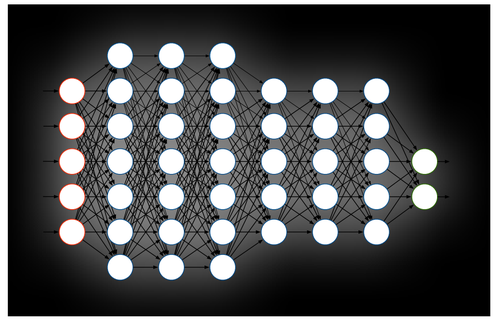


Generative Adversarial Networks

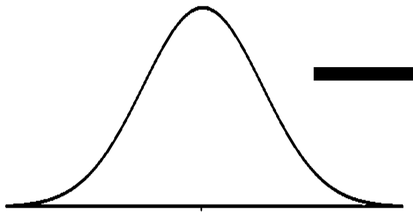
True Examples

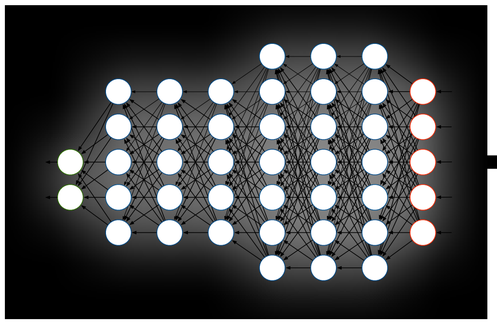






Loss
Is this a real dog picture?



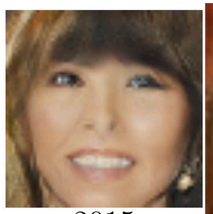


Generator

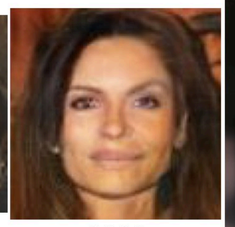




2014



2015



2016

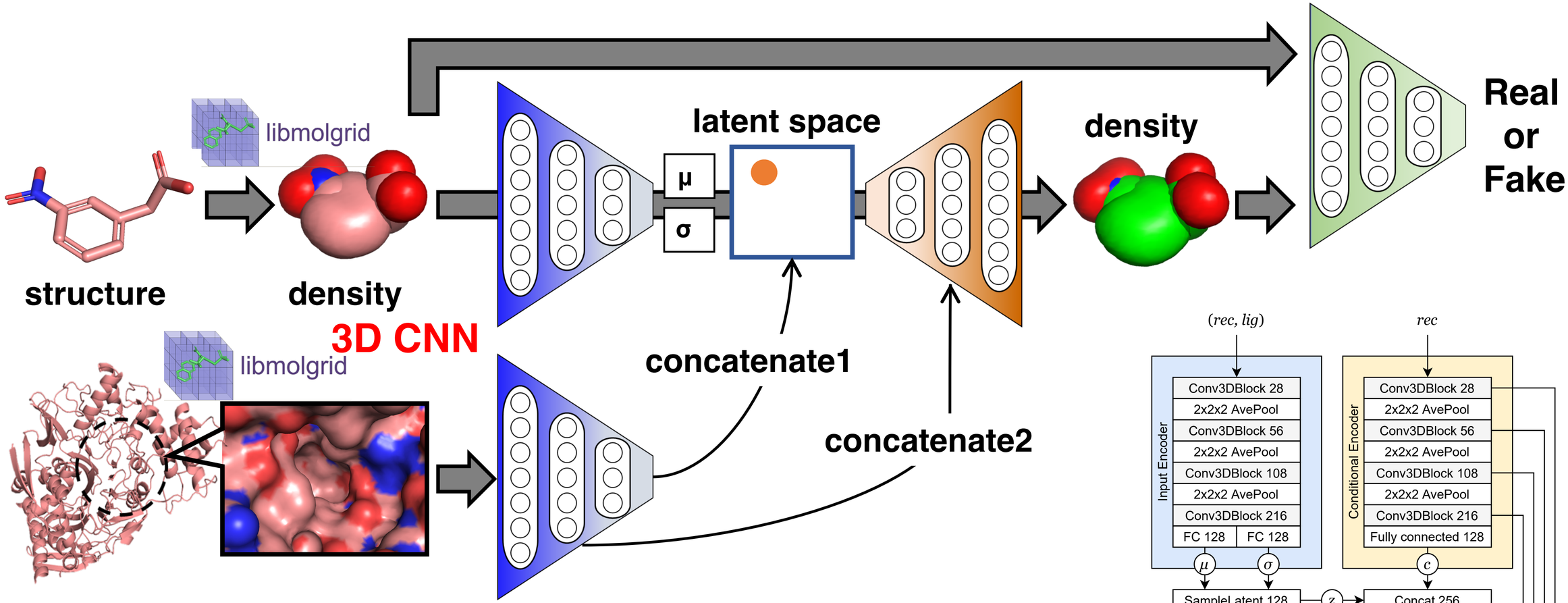


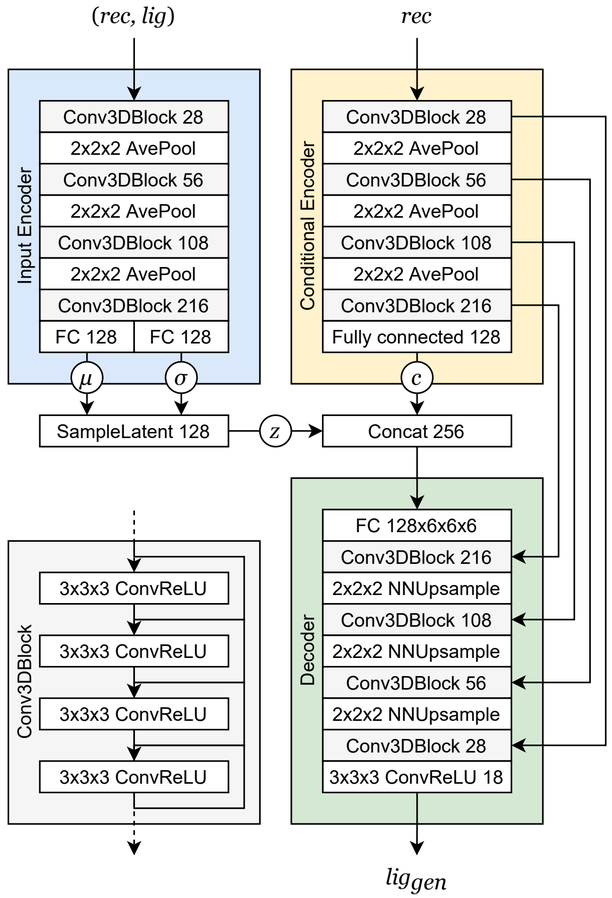
2017



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

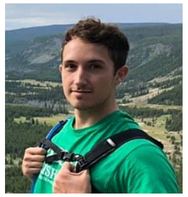
Check for updates

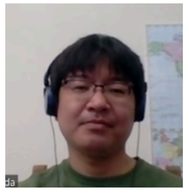
Cite this: *Chem. Sci.*, 2022, **13**, 2701

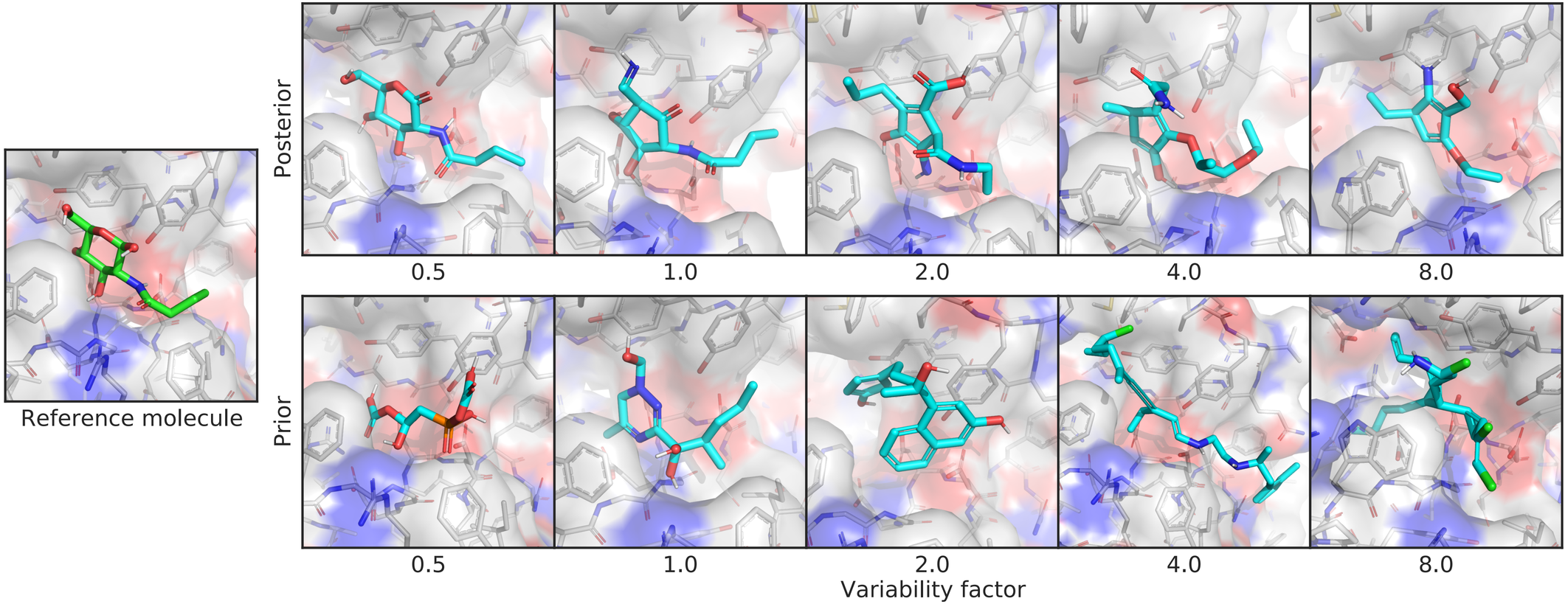
All publication charges for this article have been paid for by the Royal Society

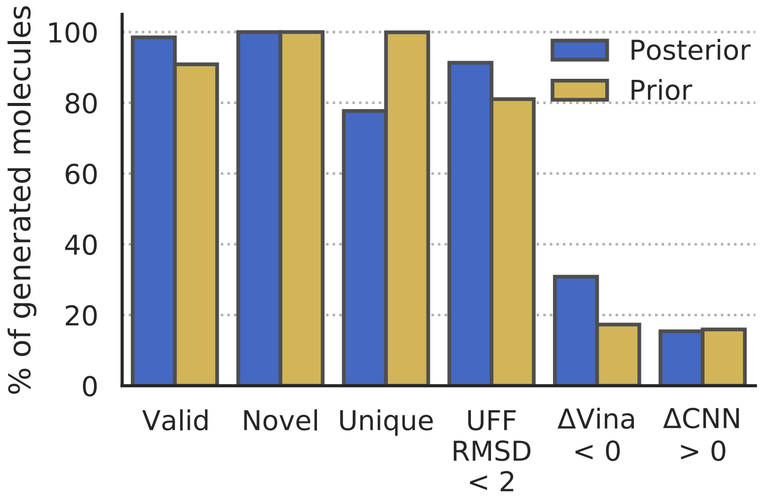
Generating 3D molecules conditional on receptor binding sites with deep generative models†

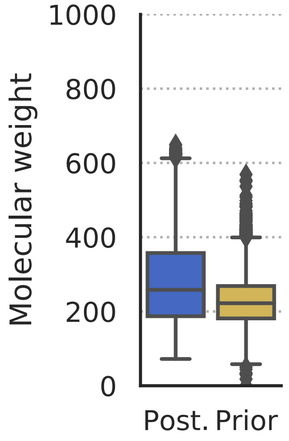
Matthew Ragoza, *[a] Tomohide Masuda[b] and David Ryan Koes[c]

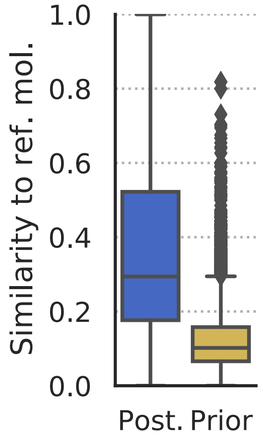


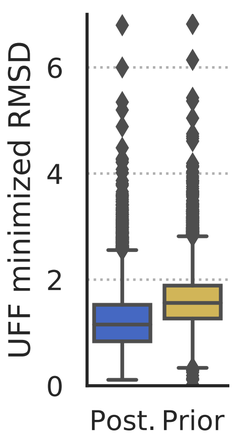


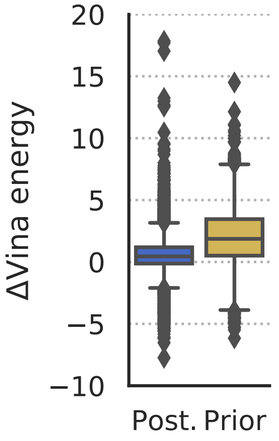

Reference molecule
Posterior
Prior
0.5
1.0
2.0
4.0
8.0
Variability factor

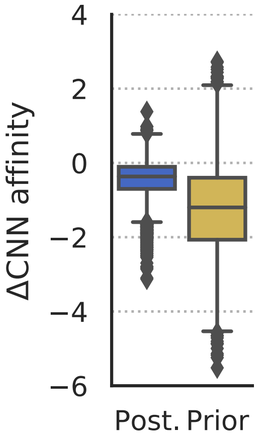

% of generated molecules
Posterior
Prior
Valid
Novel
Unique
UFF RMSD < 2
ΔVina < 0
ΔCNN > 0

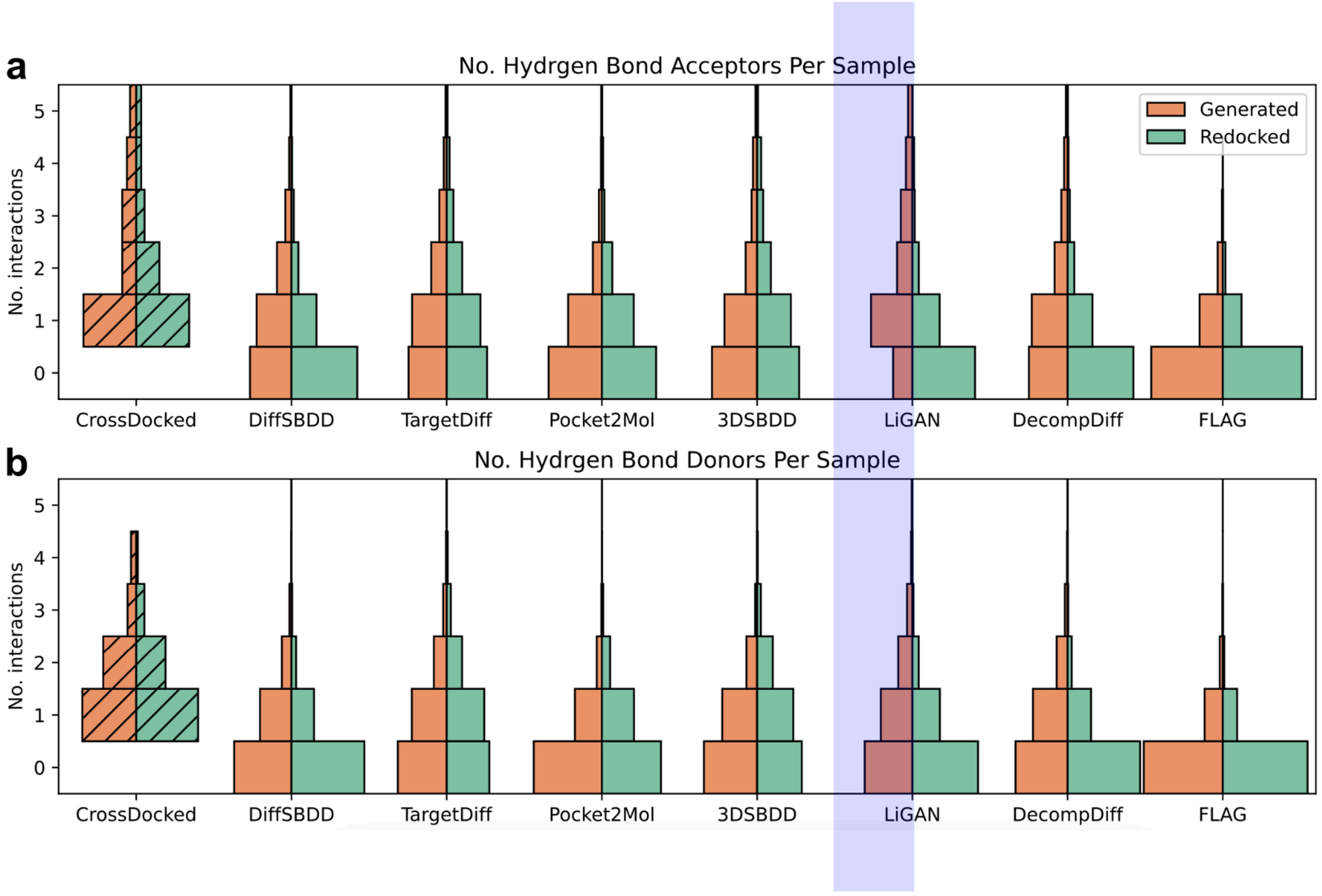

Molecular weight
Post. Prior


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

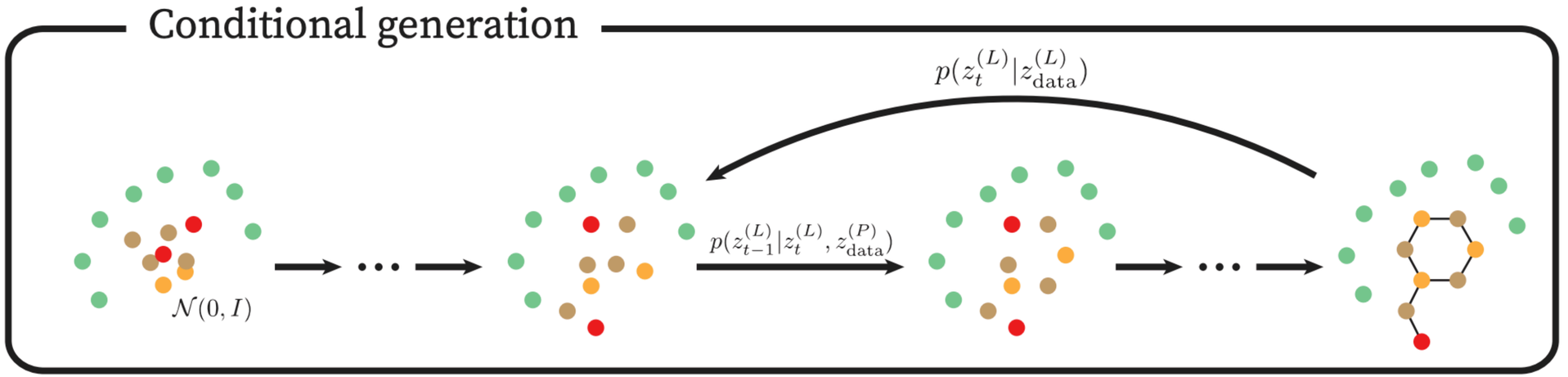
Chaitanya K. Joshi
University of Cambridge
ckj24@cam.ac.uk

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

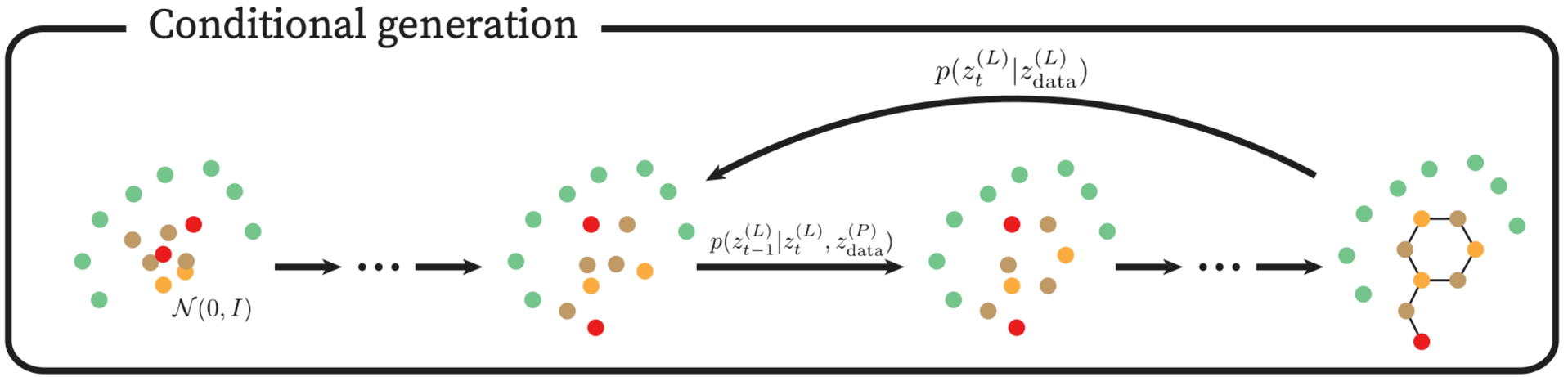
Quantitative Biology > Biomolecules

[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

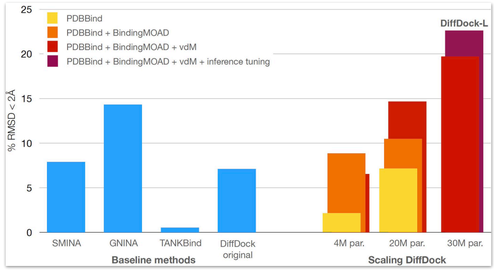
arXiv > q-bio > arXiv:2210.01776

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

↓

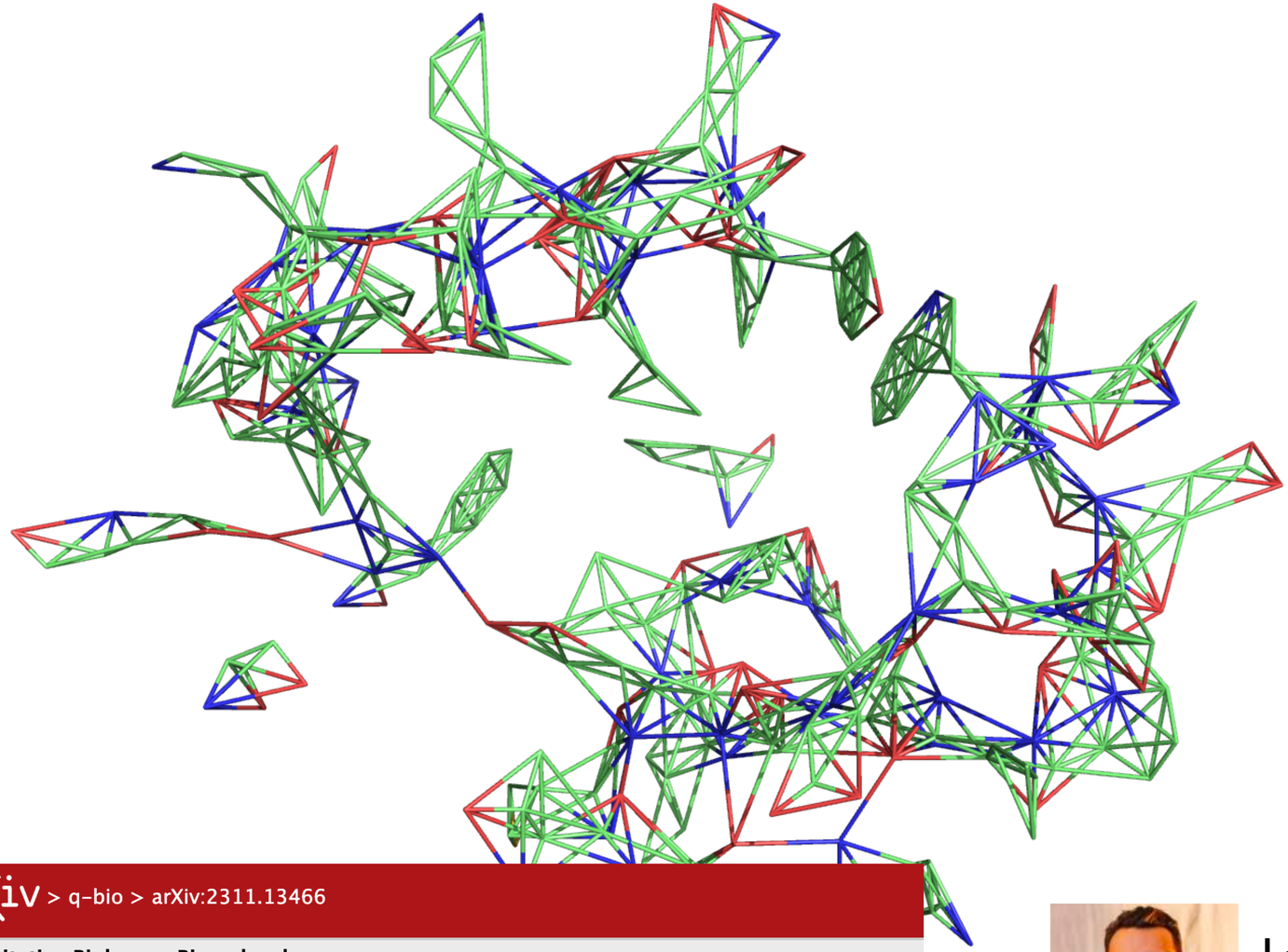
Atomic Message Passing

↓

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

↓

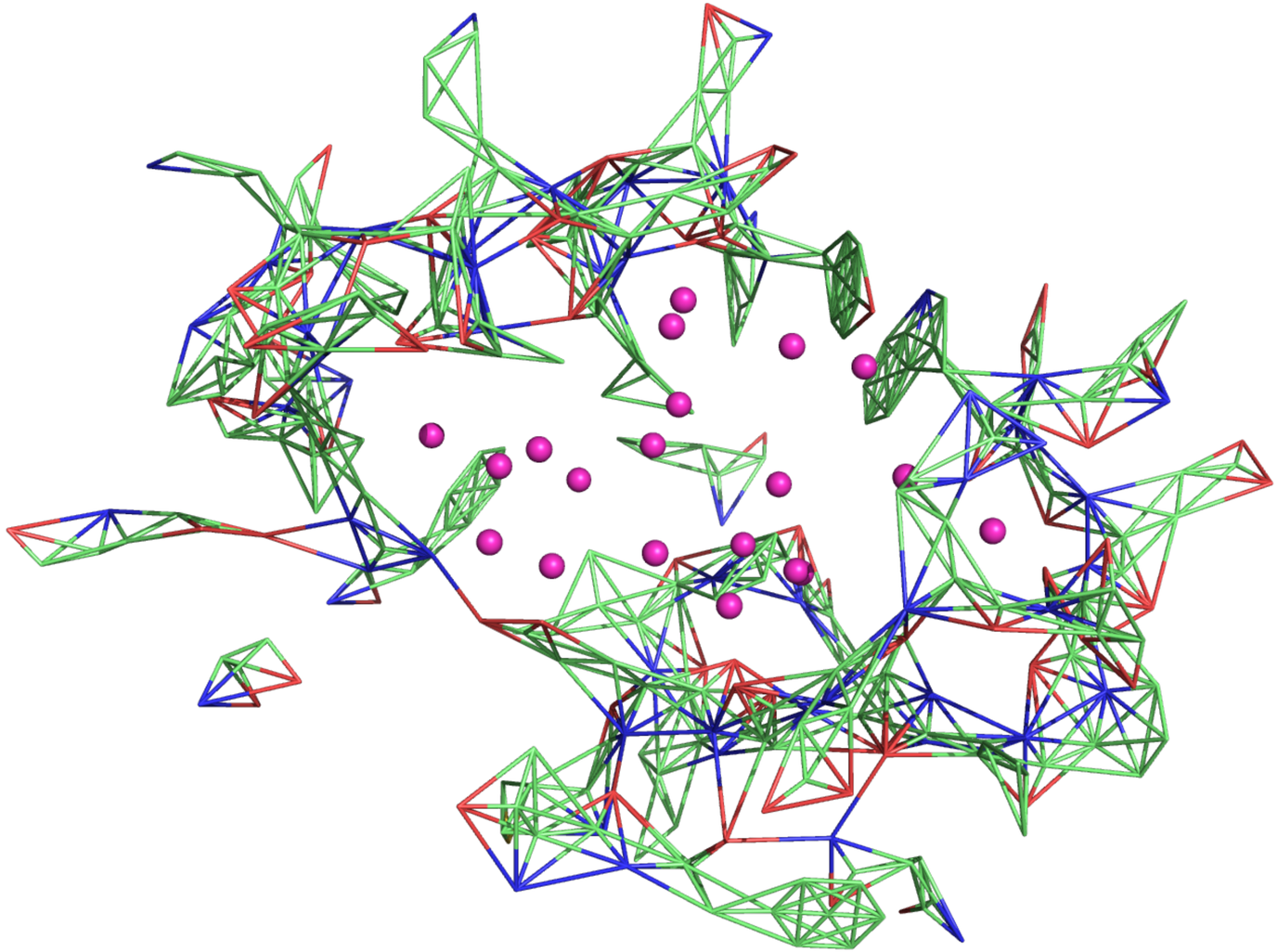
Atomic Message Passing

↓

Keypoint Generation

↓

Keypoint-Conditioned Diffusion



Keypoint Conditioned Diffusion

Pocket Identification

↓

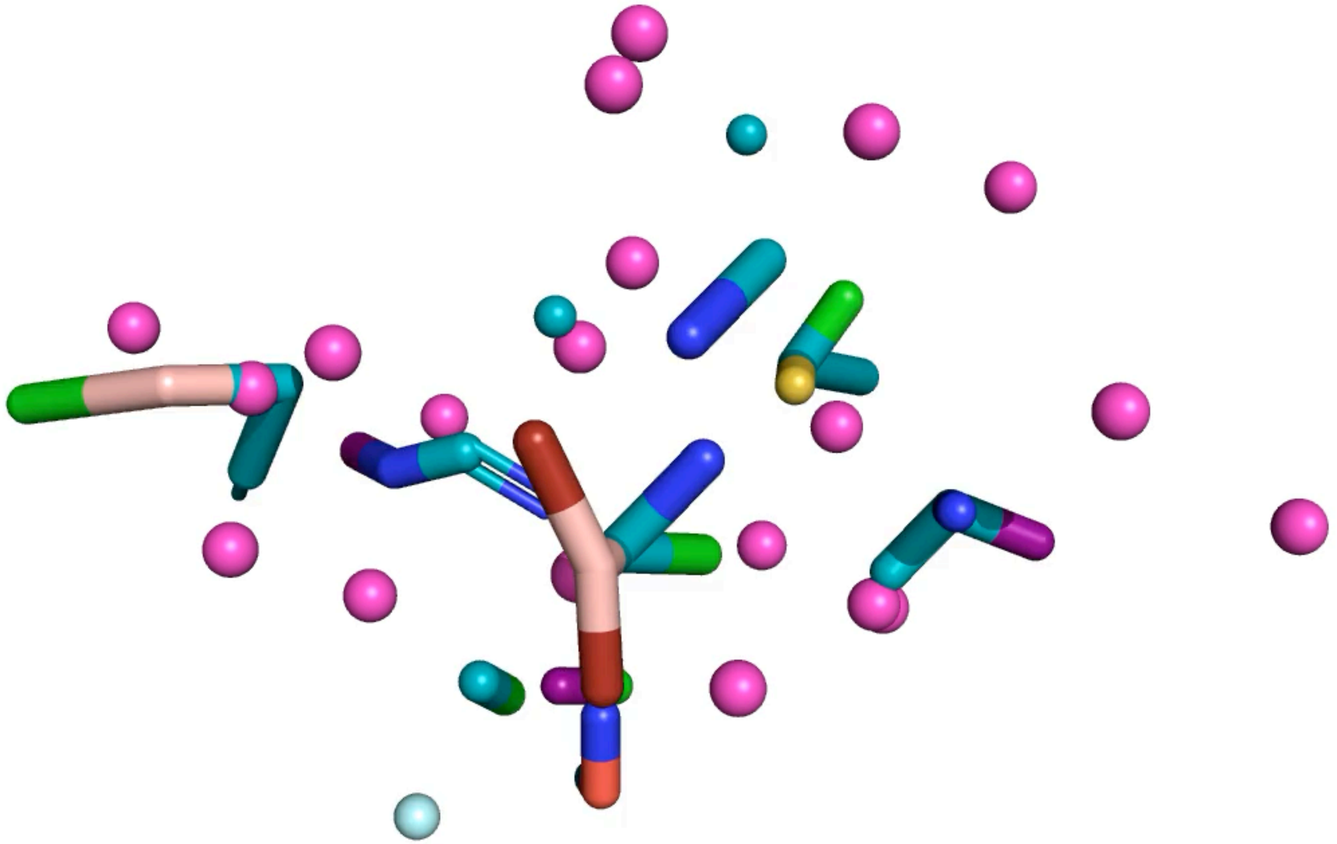
Atomic Message Passing

↓

Keypoint Generation

↓

Keypoint-Conditioned Diffusion



Keypoint Conditioned Diffusion

Pocket Identification

↓

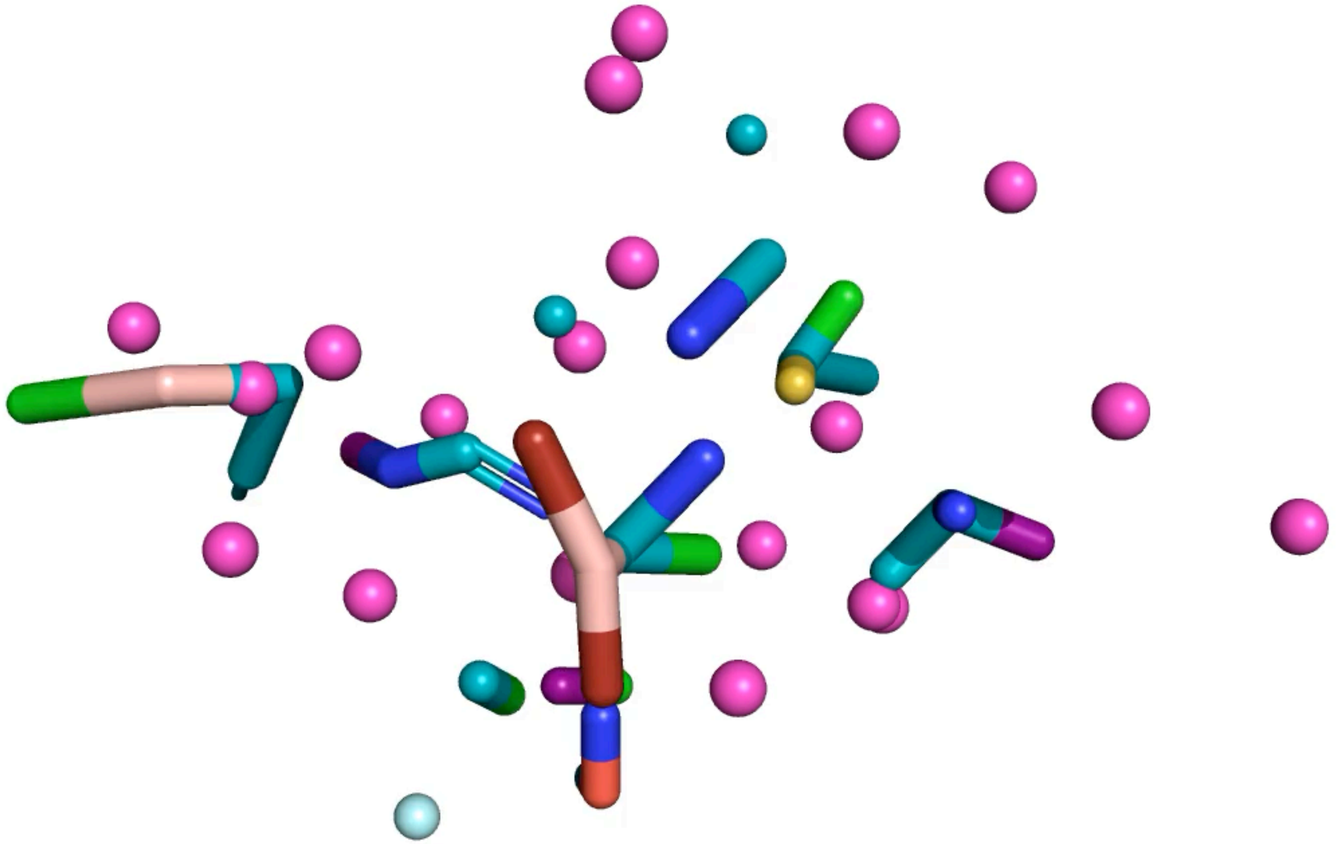
Atomic Message Passing

↓

Keypoint Generation

↓

Keypoint-Conditioned Diffusion



Keypoint Conditioned Diffusion

Pocket Identification

↓

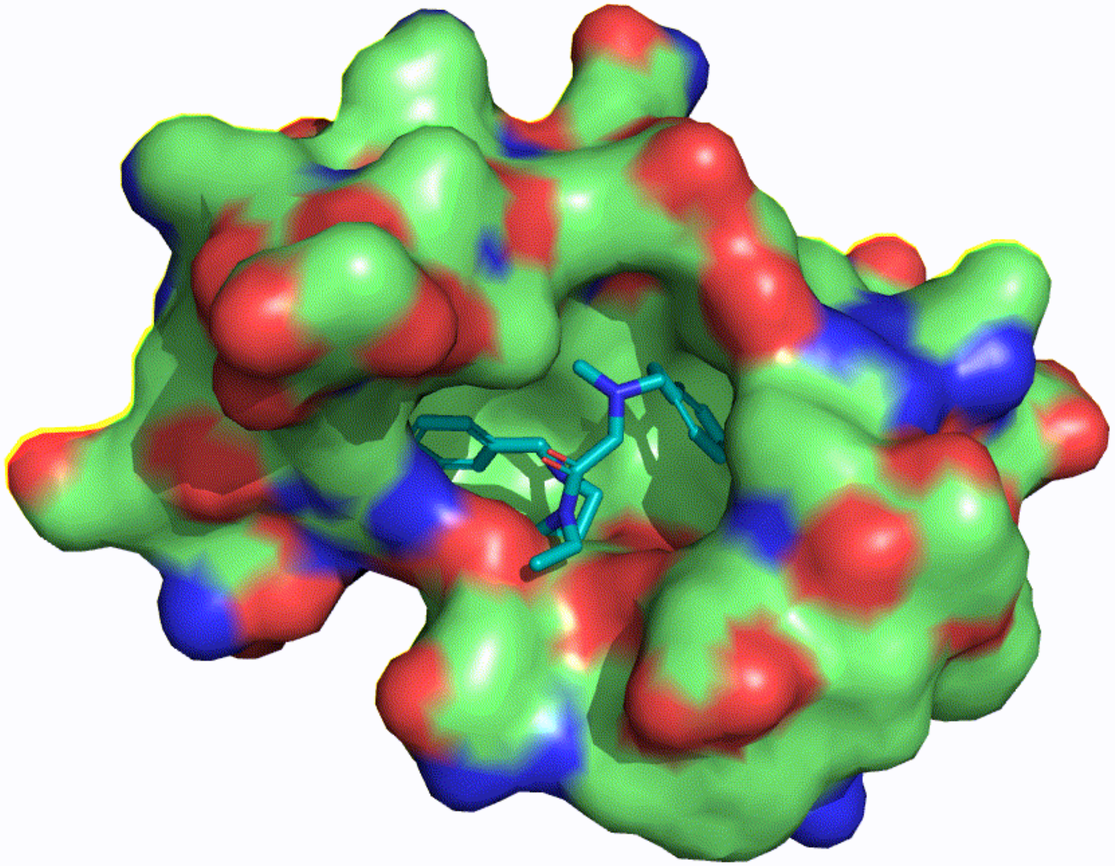
Atomic Message Passing

↓

Keypoint Generation

↓

Keypoint-Conditioned Diffusion



Keypoint Conditioned Diffusion

Pocket Identification

↓

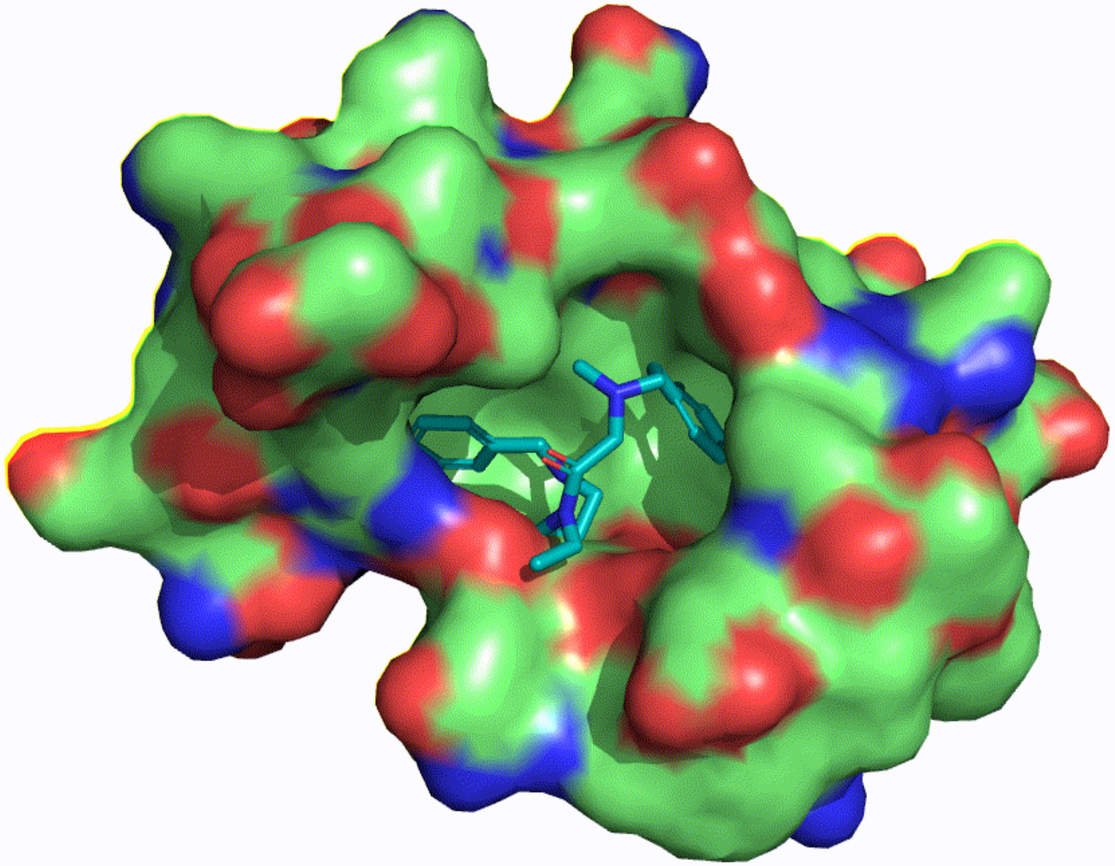
Atomic Message Passing

↓

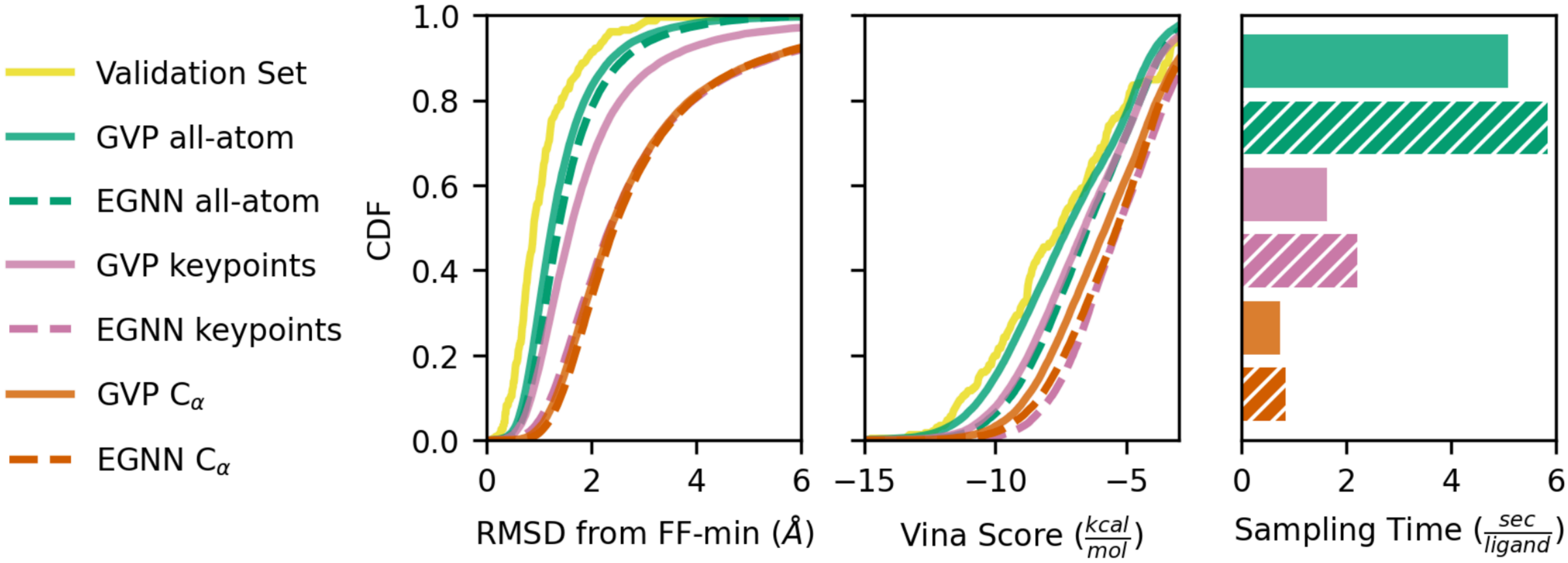
Keypoint Generation

↓

Keypoint-Conditioned Diffusion

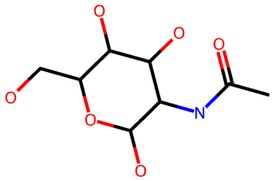


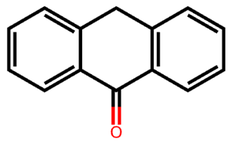
Keypoint Conditioned Diffusion

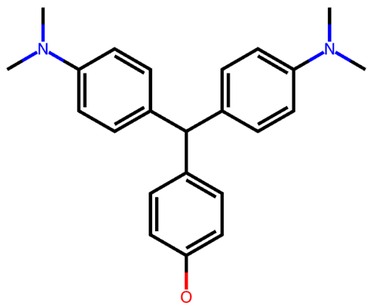


Generated molecules are not synthetically accessible or physically plausible

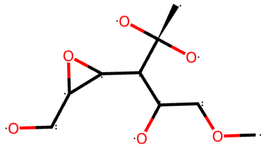
Reference Molecules

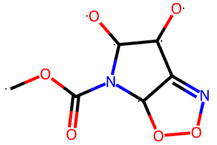


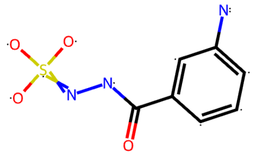


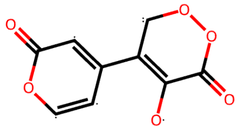


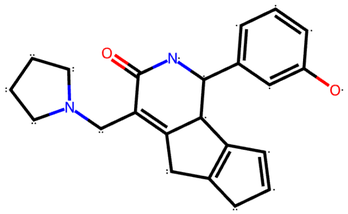
Generated Molecules

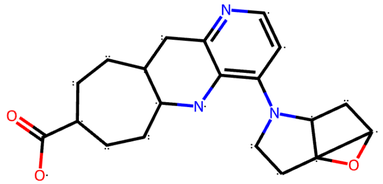






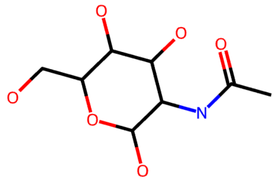


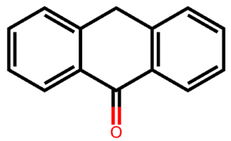


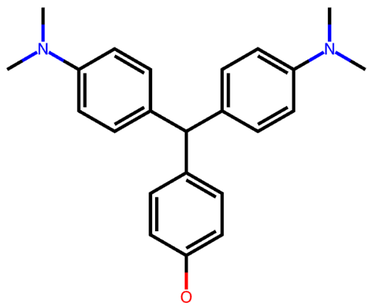


Generated molecules are not synthetically accessible or physically plausible

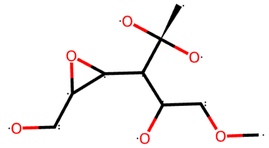
Reference Molecules

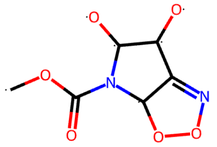




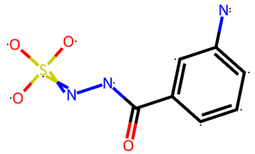


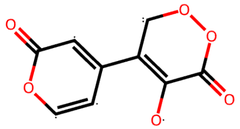
Generated Molecules

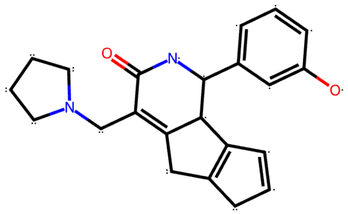


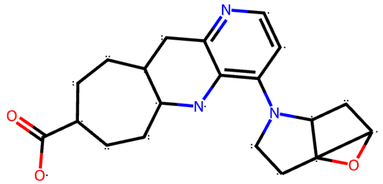


"epoxide, good luck with stability"



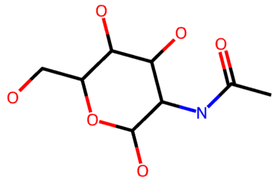


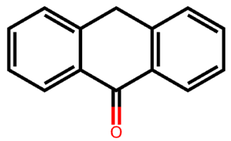


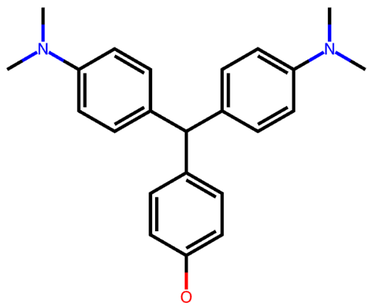


Generated molecules are not synthetically accessible or physically plausible

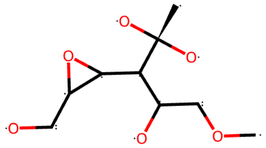
Reference Molecules

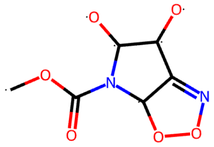




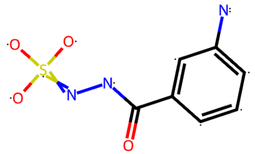


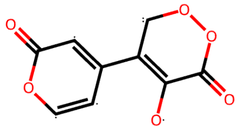
Generated Molecules

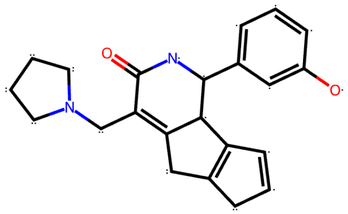


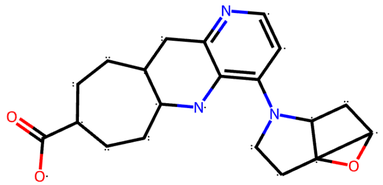


"epoxide, good luck with stability"





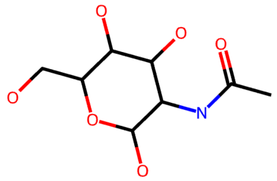


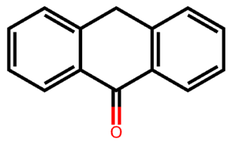


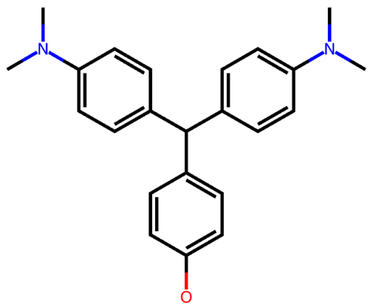
"certainly unstable"

Generated molecules are not synthetically accessible or physically plausible

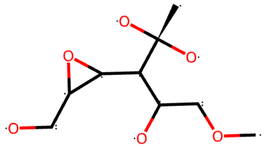
Reference Molecules

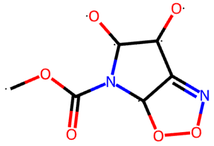




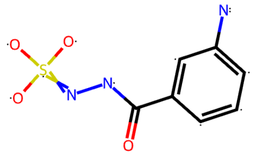


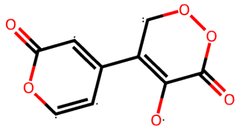
Generated Molecules



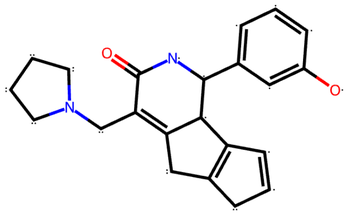


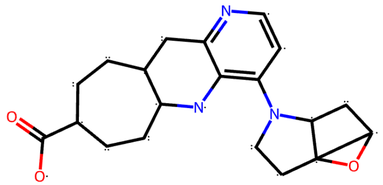
"epoxide, good luck with stability"





"no"

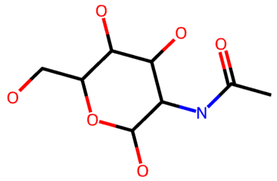


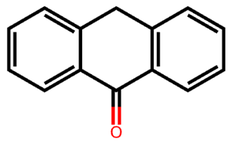


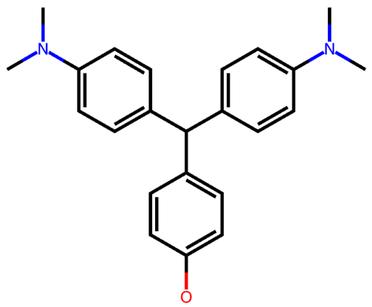
"certainly unstable"

Generated molecules are not synthetically accessible or physically plausible

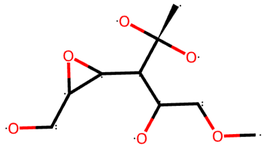
Reference Molecules



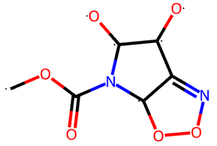




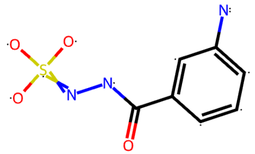
Generated Molecules



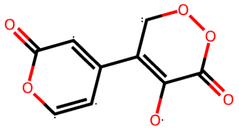
"epoxide, good luck with stability"

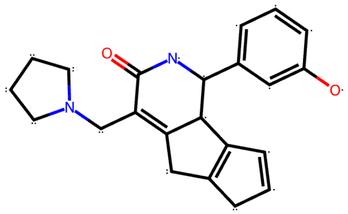


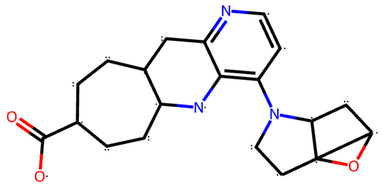
"f*** no"



"no"







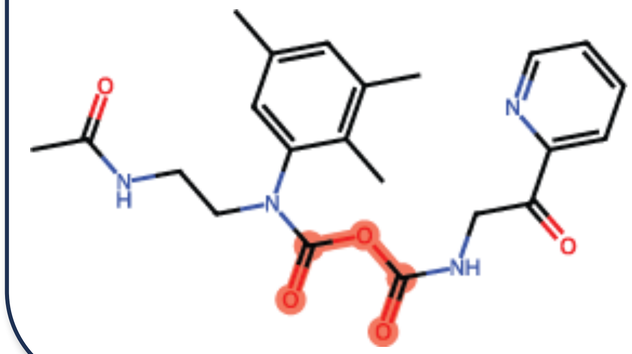
"certainly unstable"

Practical Measures of Molecule Quality

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

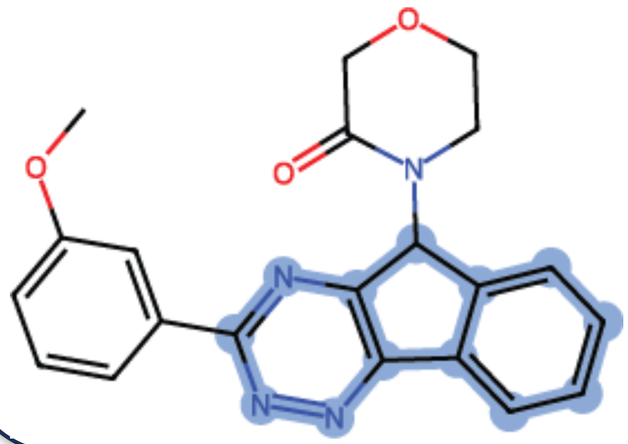
We propose to evaluate molecule quality at the level of **functional groups** and **ring systems**

Structural Alerts



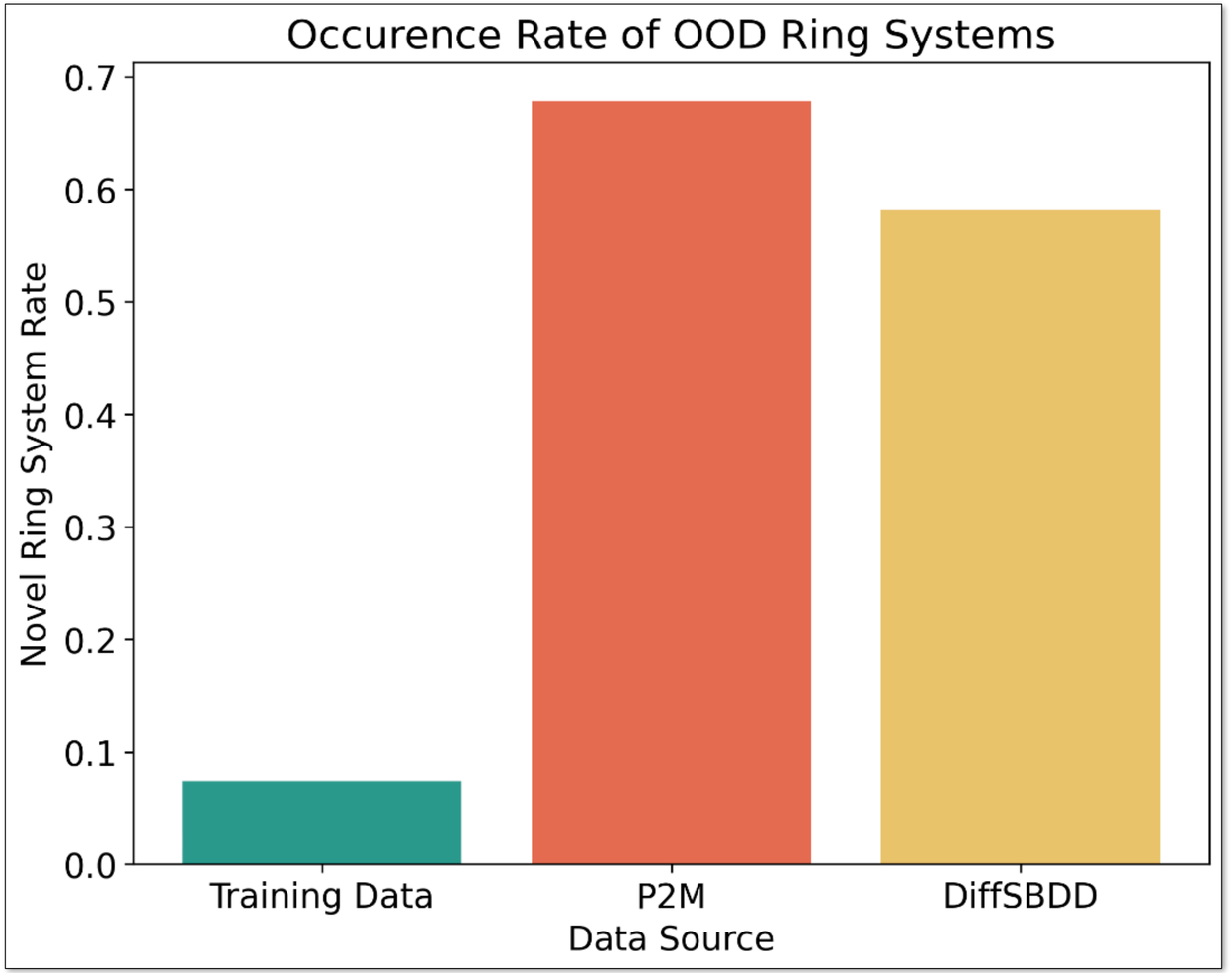
Functional groups typically used to filter libraries in virtual screening workflows

OOD Rings

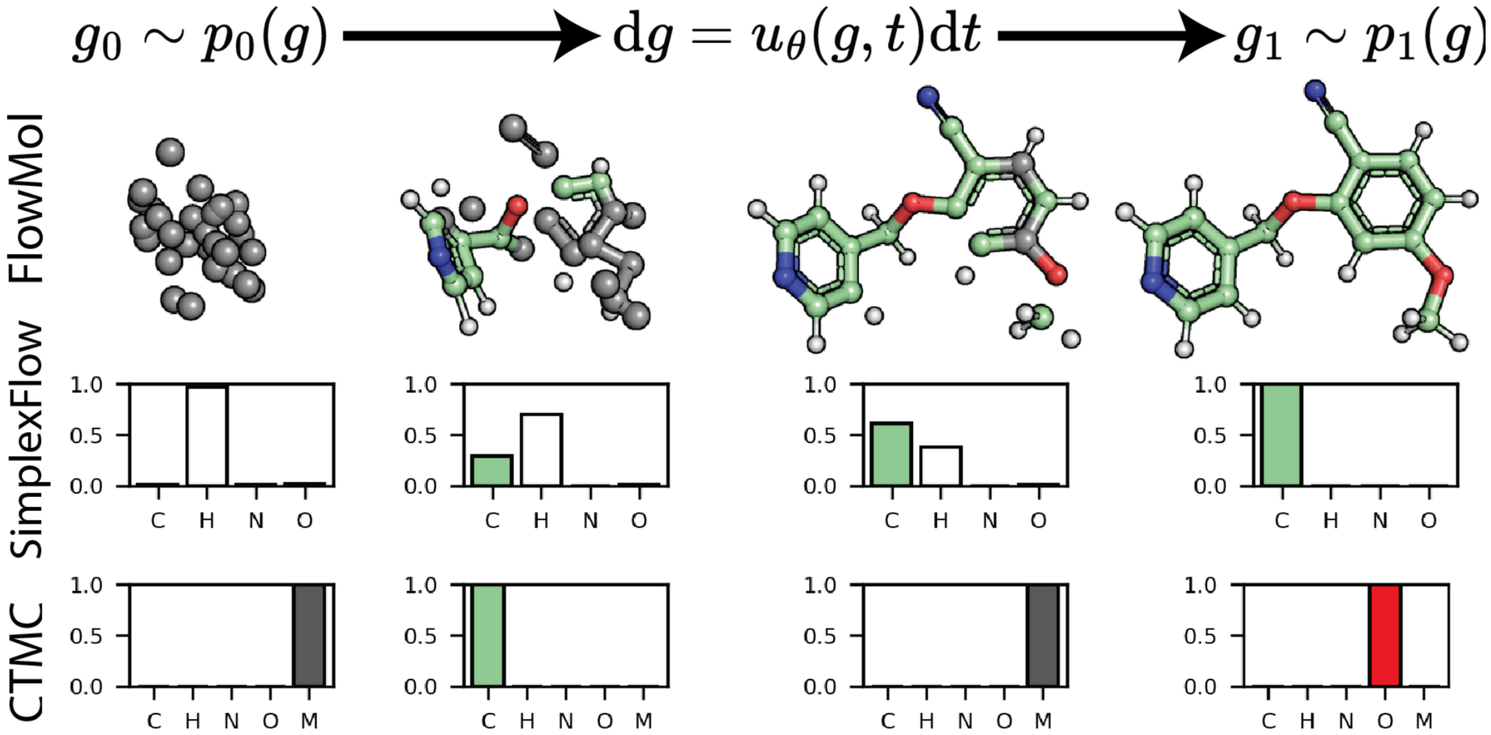


Ring systems that never appear in ChEMBL

Practical Measures of Molecule Quality



Unconditional Generation with FlowMol



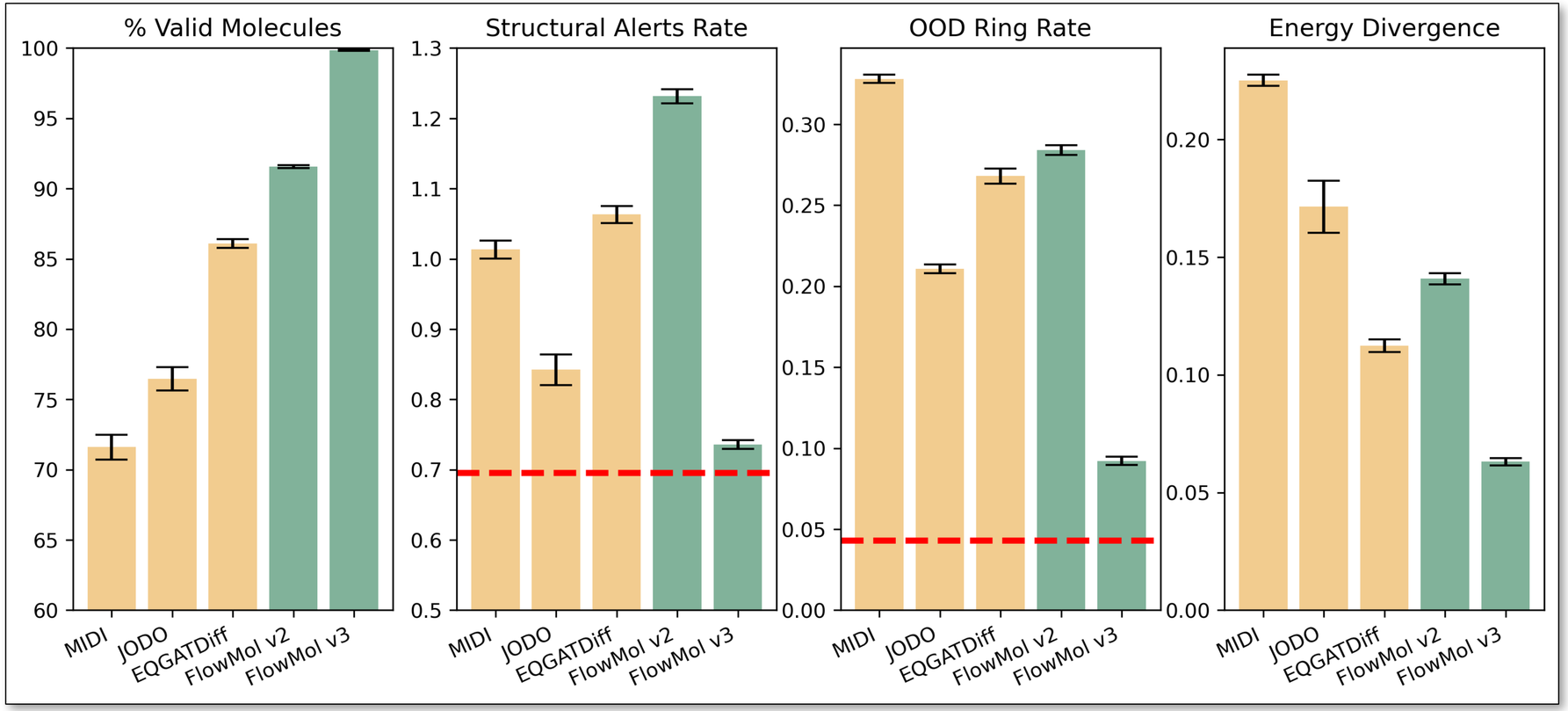
Exploring Discrete Flow Matching for 3D De Novo Molecule Generation

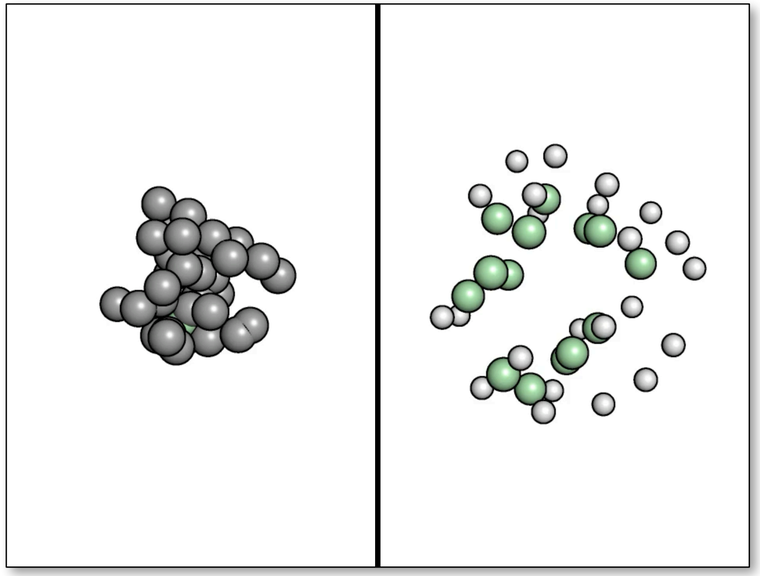


Ian Dunn
Dept. of Computational & Systems Biology
University of Pittsburgh
Pittsburgh, PA 15260
ian.dunn@pitt.edu

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

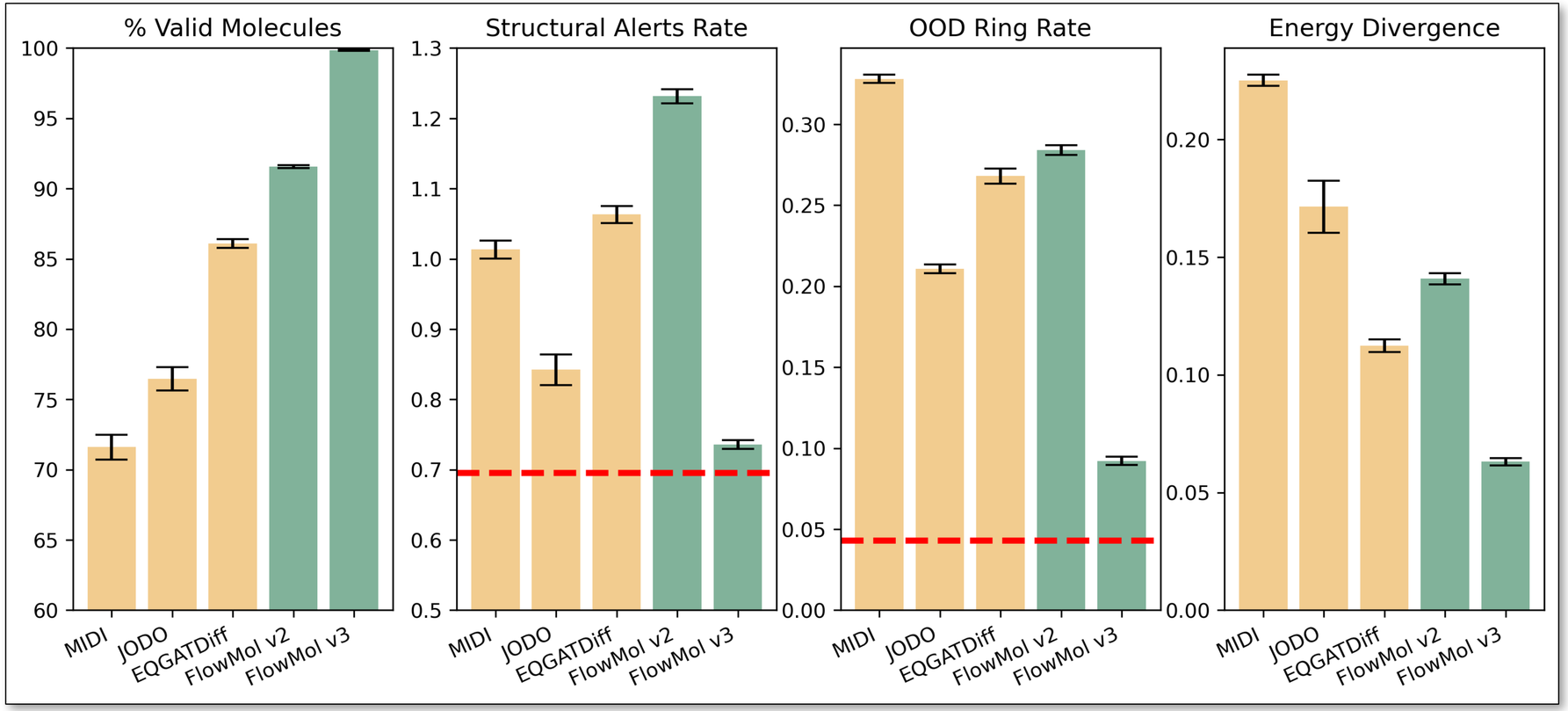
- State of the art validity
- Improves chemical plausibility and synthetic accessibility

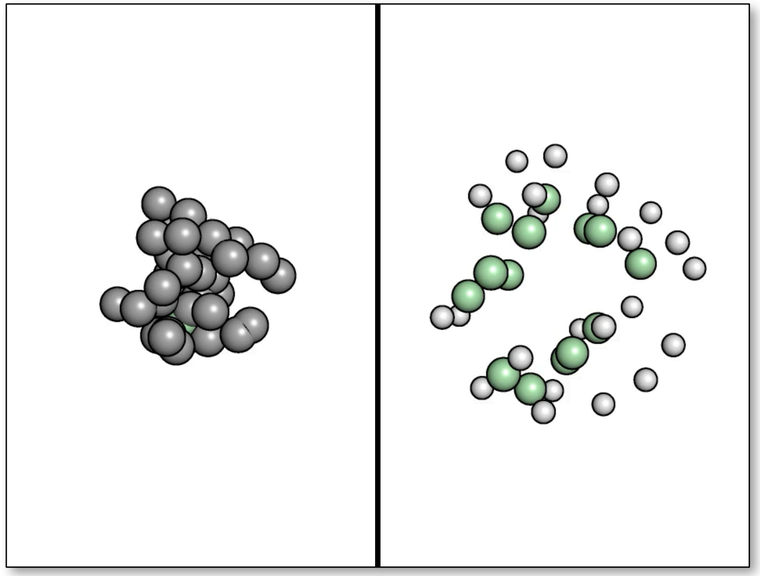
FlowMol3: Flow Matching for 3D De Novo Small-Molecule Generation

Ian Dunn and David Ryan Koes



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

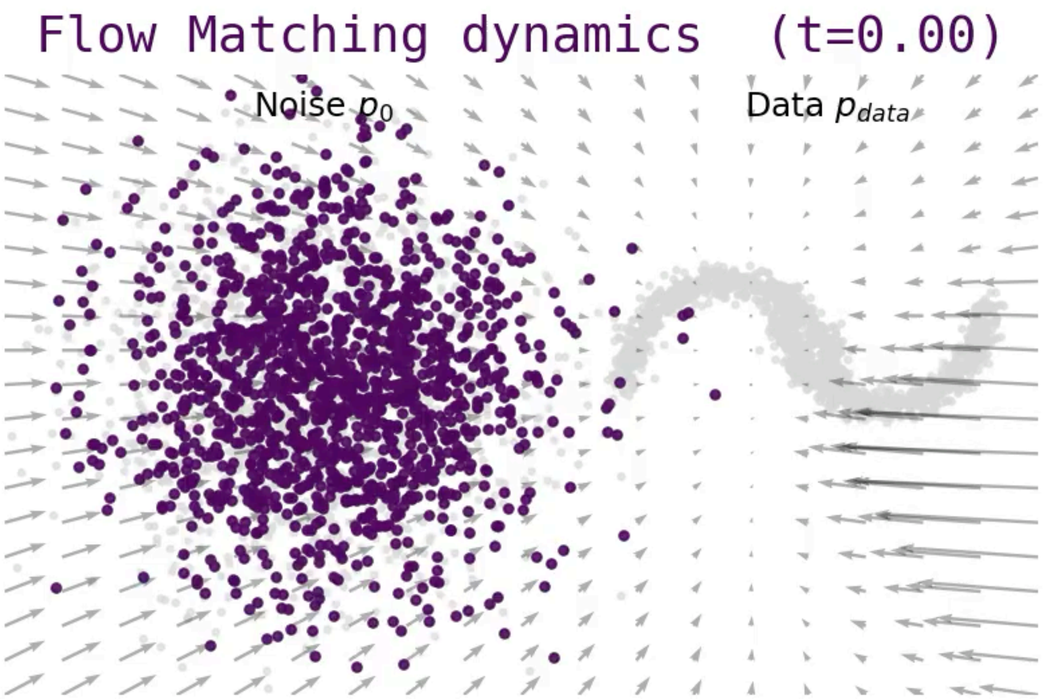
FlowMol3: Flow Matching for 3D De Novo Small-Molecule Generation

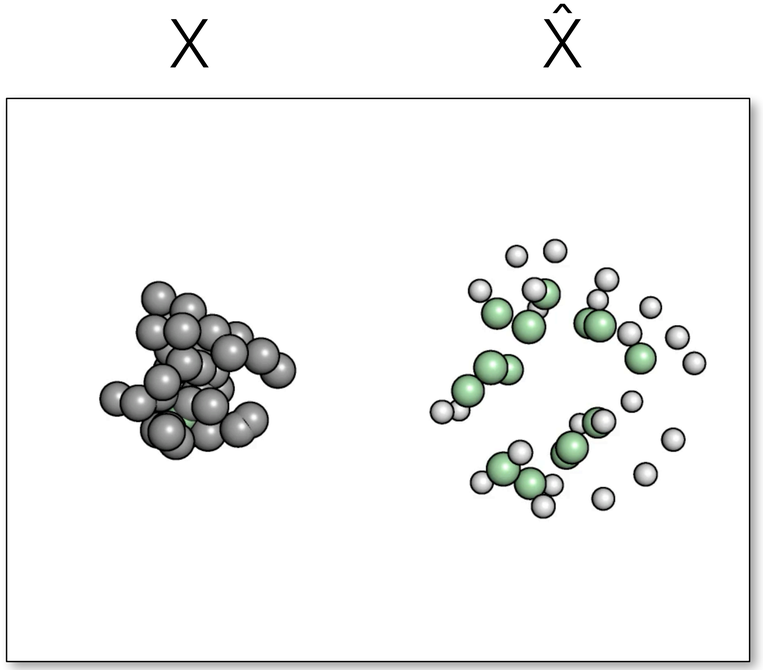
Ian Dunn and David Ryan Koes



Why is FlowMol3 Better?

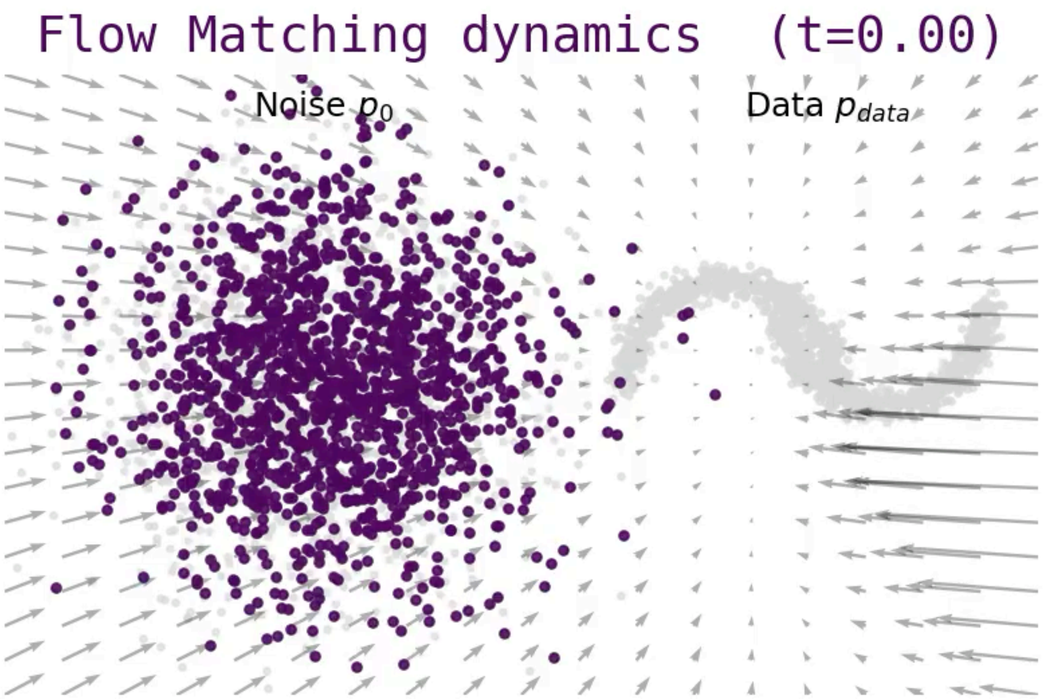
$$g_0 \sim p_0(g) \longrightarrow \mathrm{d}g = u_\theta(g, t)\mathrm{d}t \longrightarrow g_1 \sim p_1(g)$$

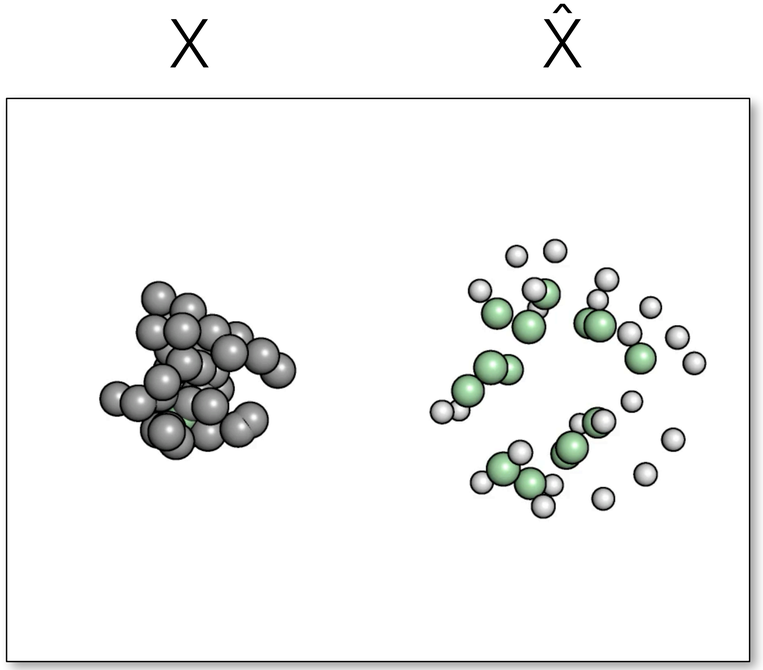




Why is FlowMol3 Better?

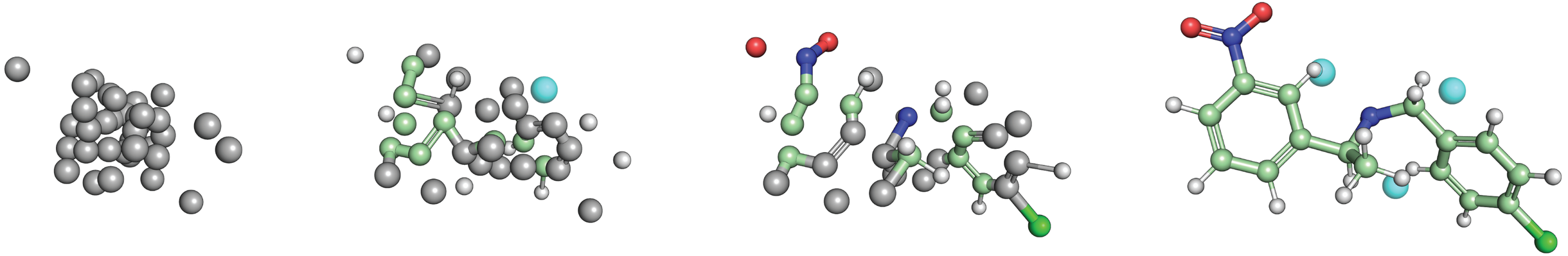
$$g_0 \sim p_0(g) \longrightarrow \mathrm{d}g = u_\theta(g, t)\mathrm{d}t \longrightarrow g_1 \sim p_1(g)$$



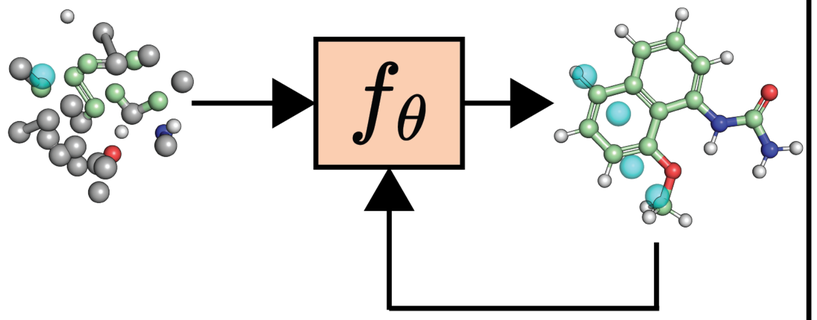


Why is FlowMol3 Better?

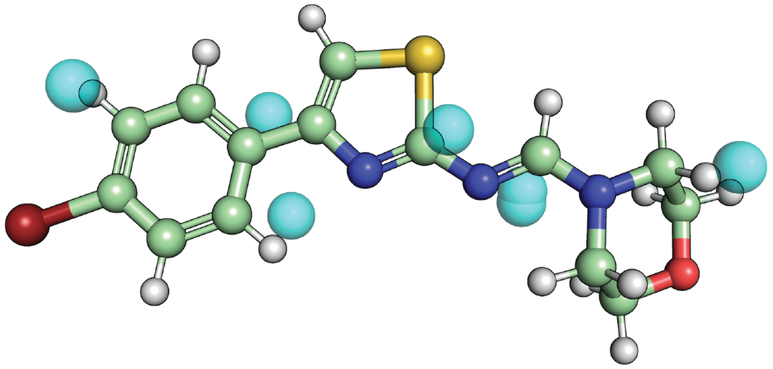
$$g_0 \sim p_0(g) \longrightarrow \mathrm{d}g = u_\theta(g, t)\mathrm{d}t \longrightarrow g_1 \sim p_1(g)$$



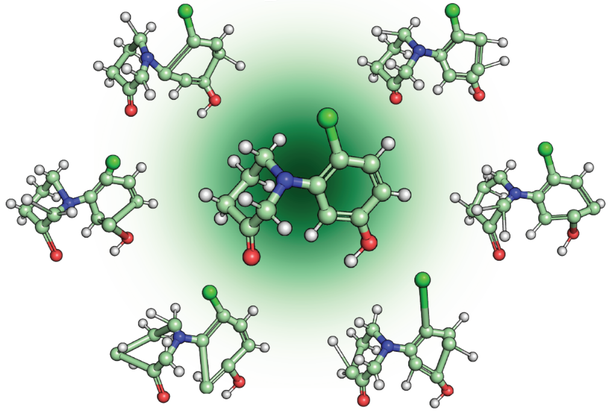
Self-Conditioning



Fake Atoms



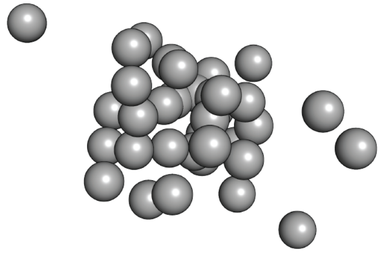
Geometry Distortion

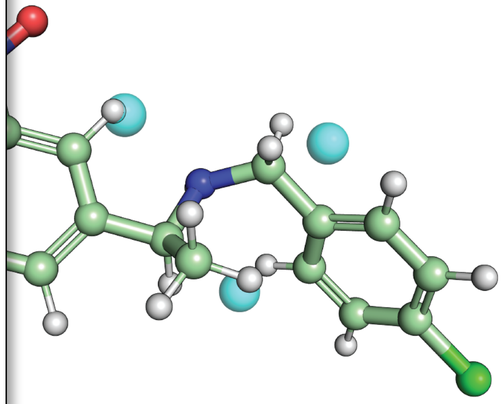


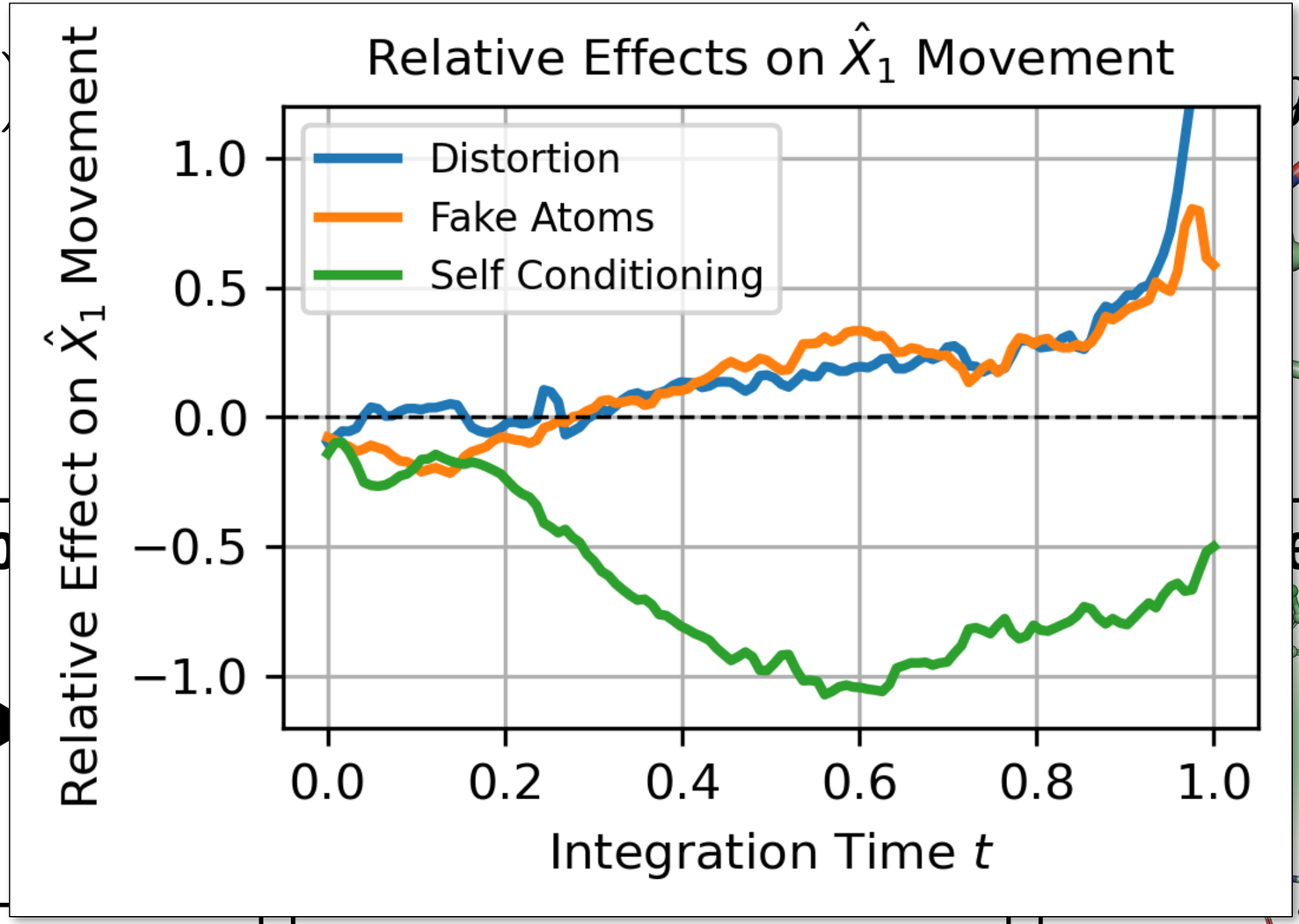
Why is FlowMol3 Better?

$g_0 \sim p_0(g)$

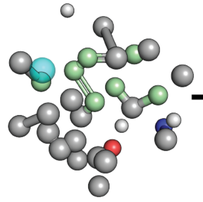
$g_1 \sim p_1(g)$





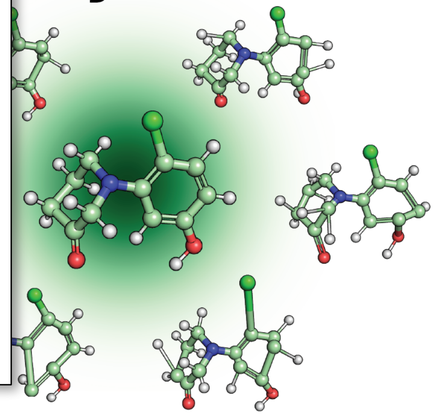


Self-Conditio



f_θ

etry Distortion

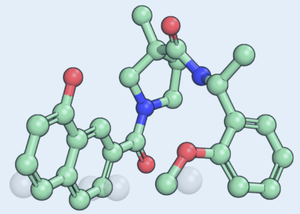


OMTRA

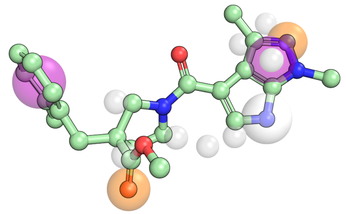
FlowMol3

De novo Generation

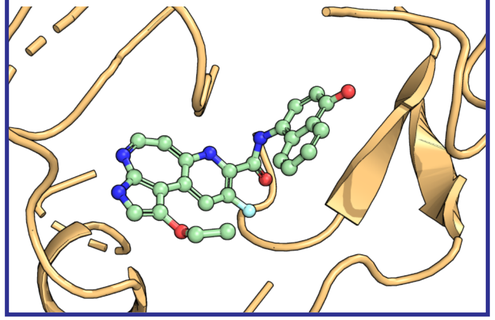
Unconditional



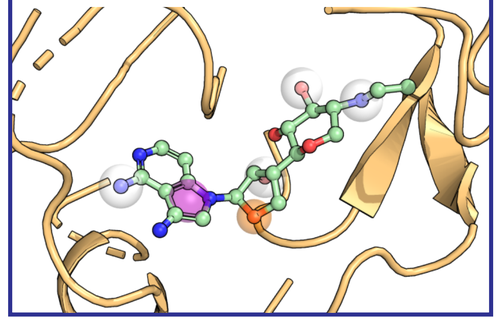
Pharmacophore Conditioned



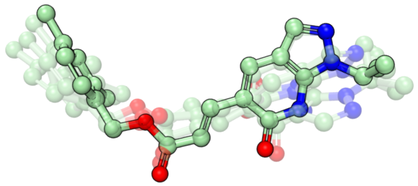
Pocket Conditioned



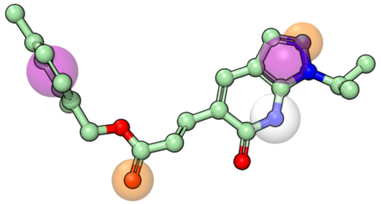
Pocket & Pharm. Conditioned



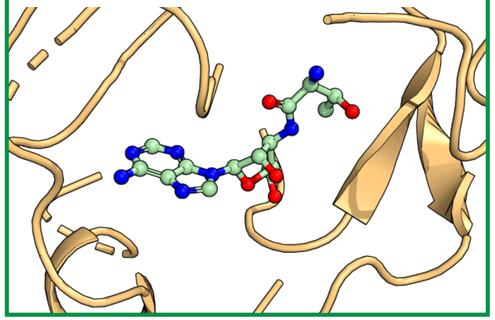
Unconditional



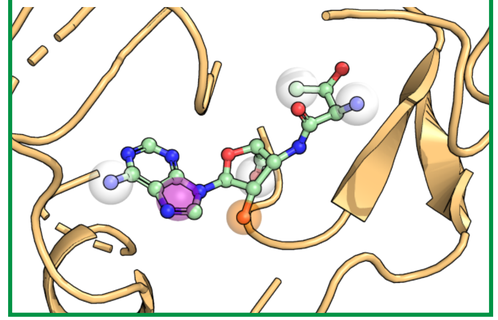
Pharmacophore Conditioned



Pocket Conditioned



Pocket & Pharm. Conditioned



Conformer Generation

Docking

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

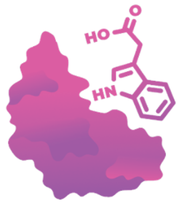


2025 Workshop on Machine Learning in Structural Biology

Data

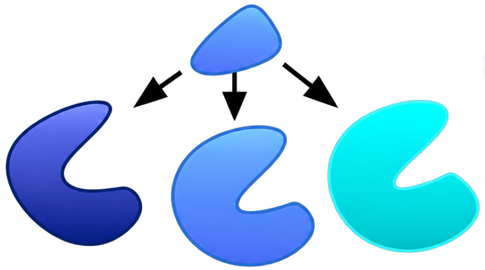
pharmit

>500 million 3D molecules



plinder

450k Protein-Ligand Interactions



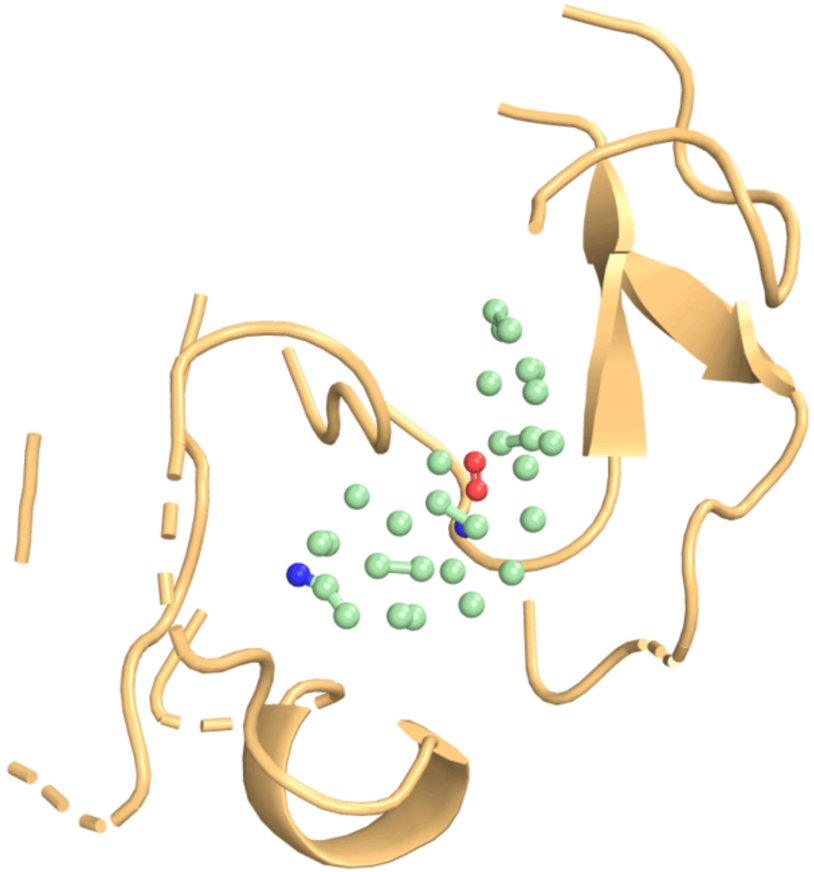
CrossDocked2020

18,450 complexes

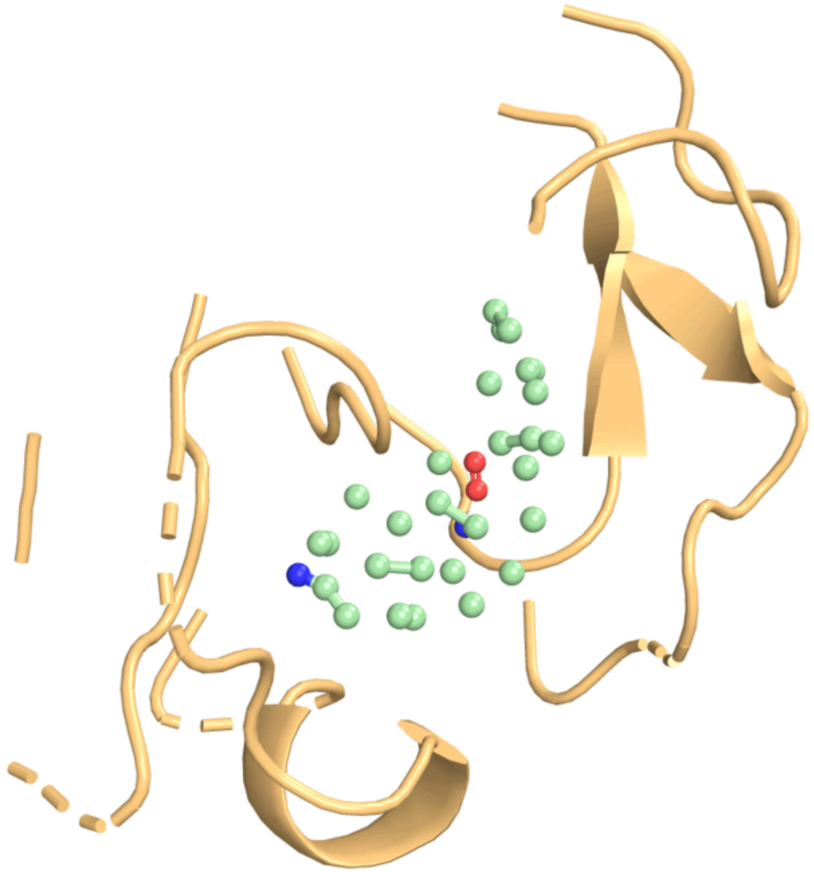
22.6 million poses

OMTRA

OMTRA: De Novo Design

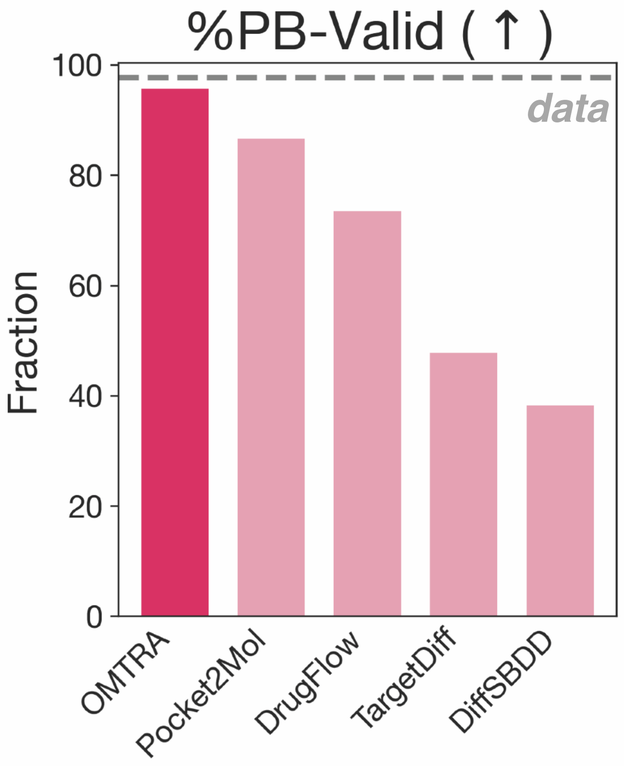


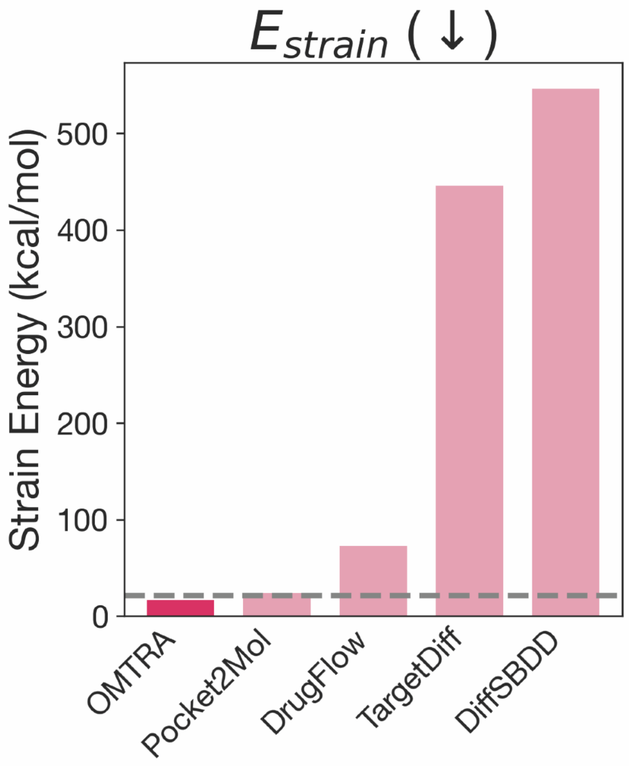
OMTRA: De Novo Design

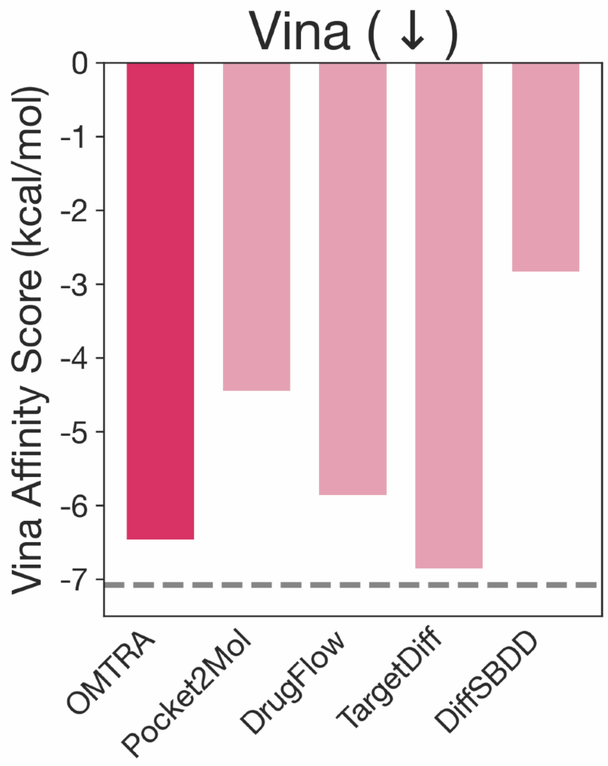


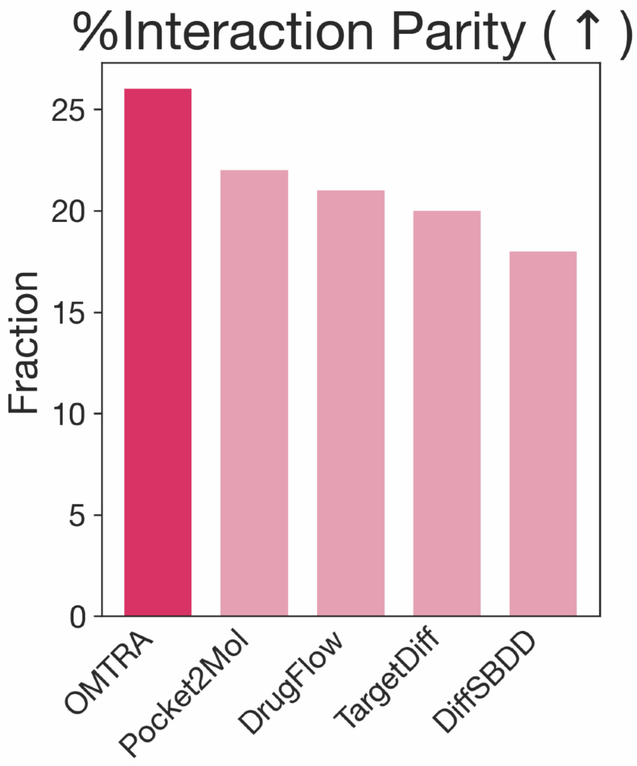
OMTRA: De Novo Design

Evaluated on Luo et al CrossDocked test set.

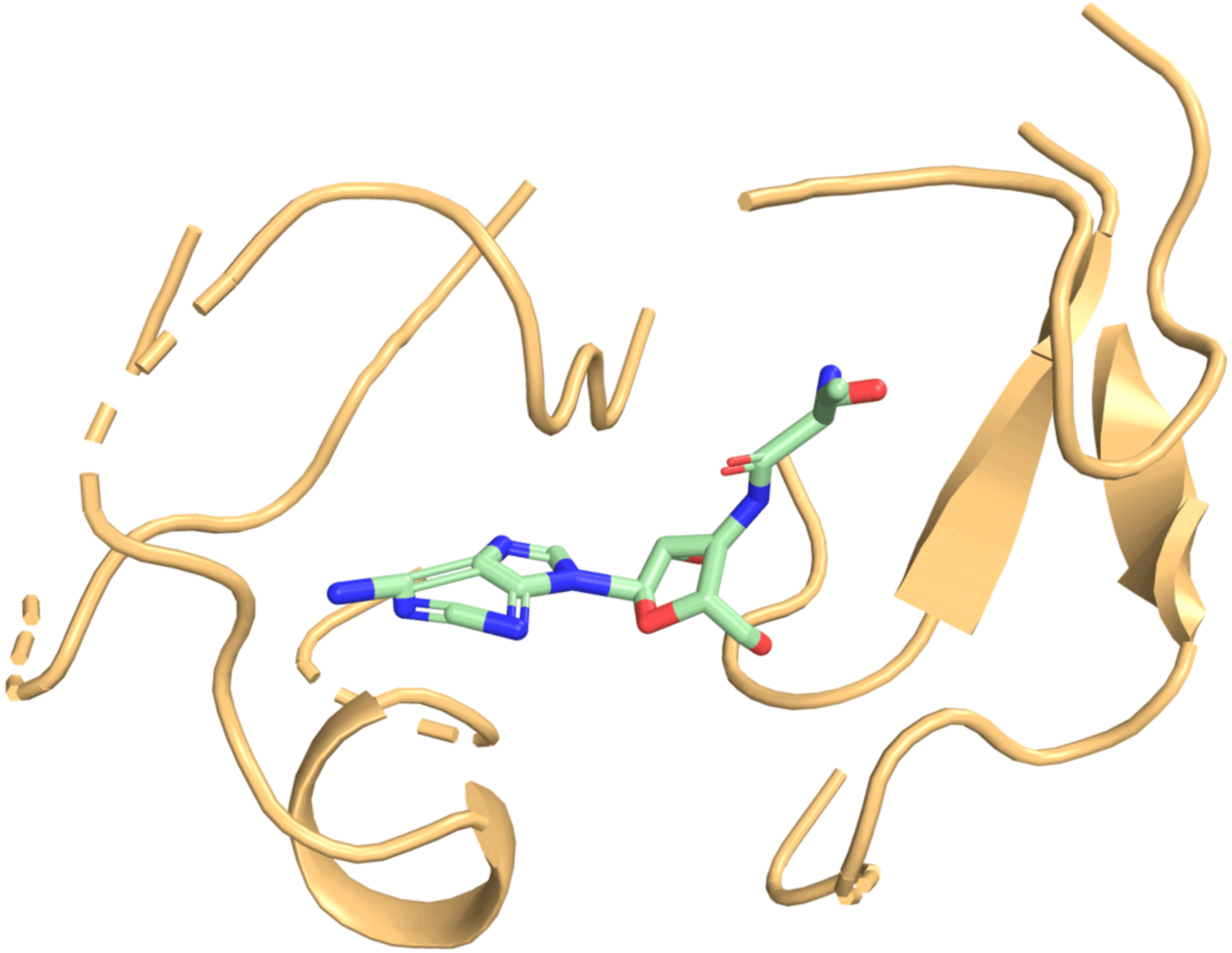




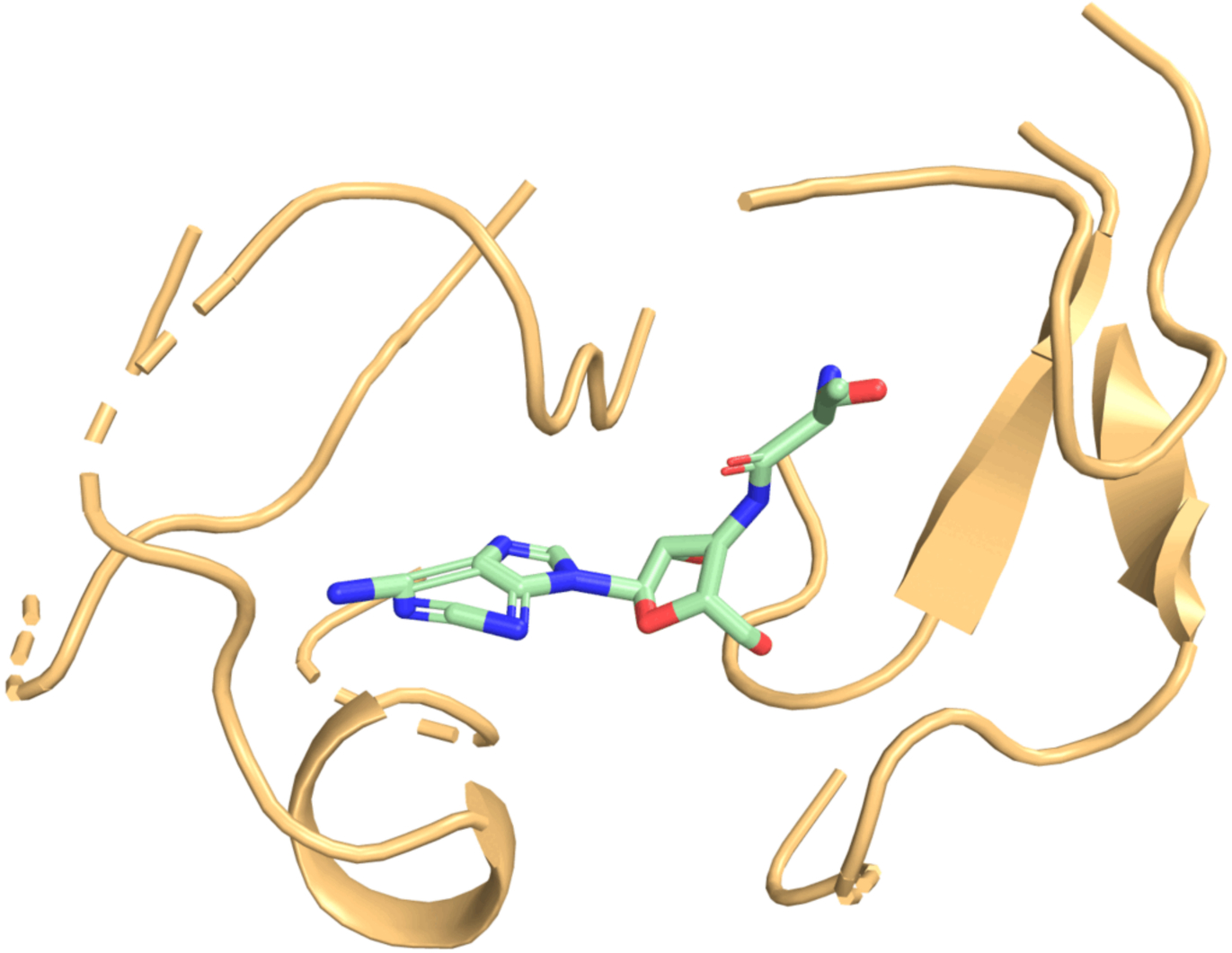




OMTRA: Docking

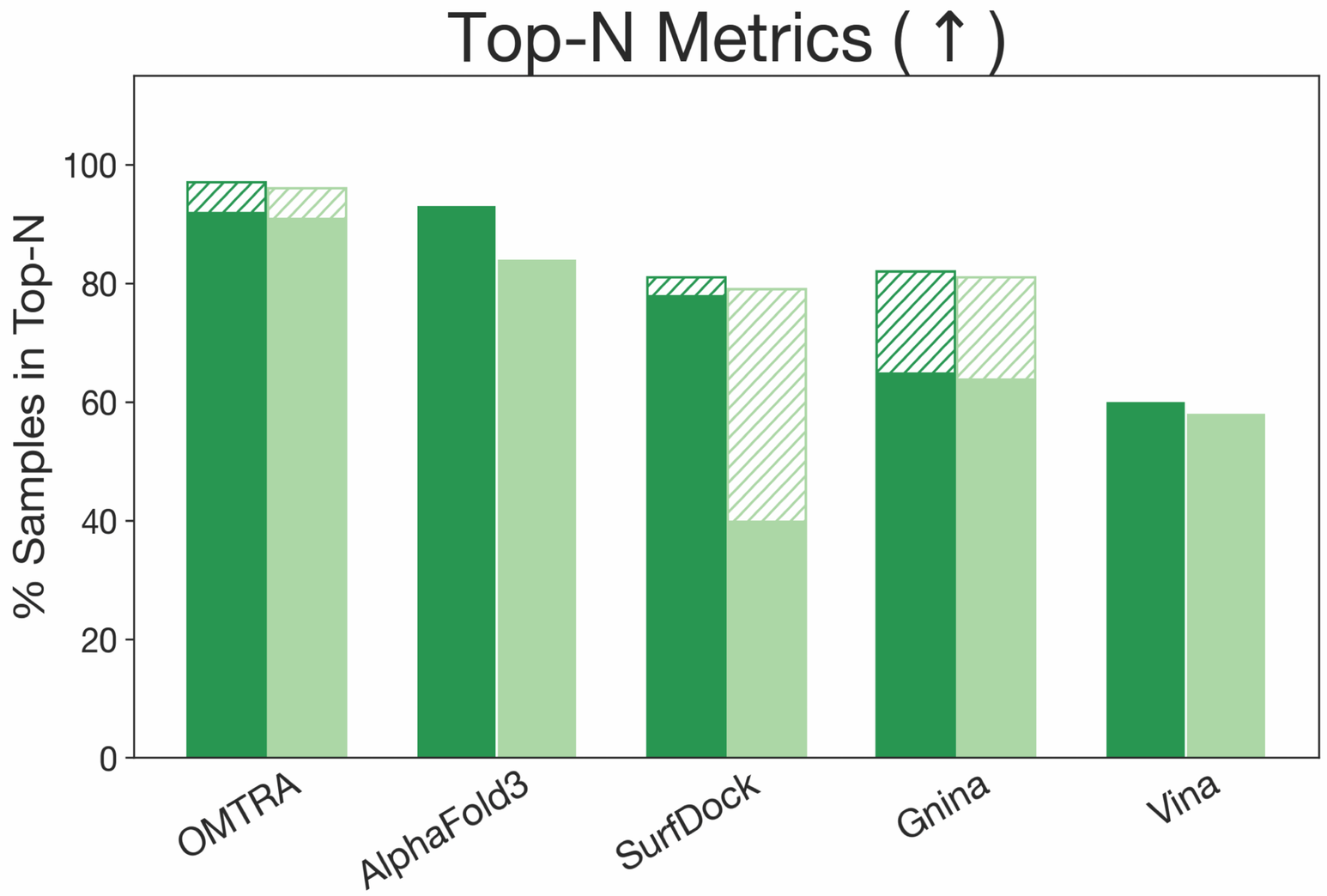


OMTRA: Docking

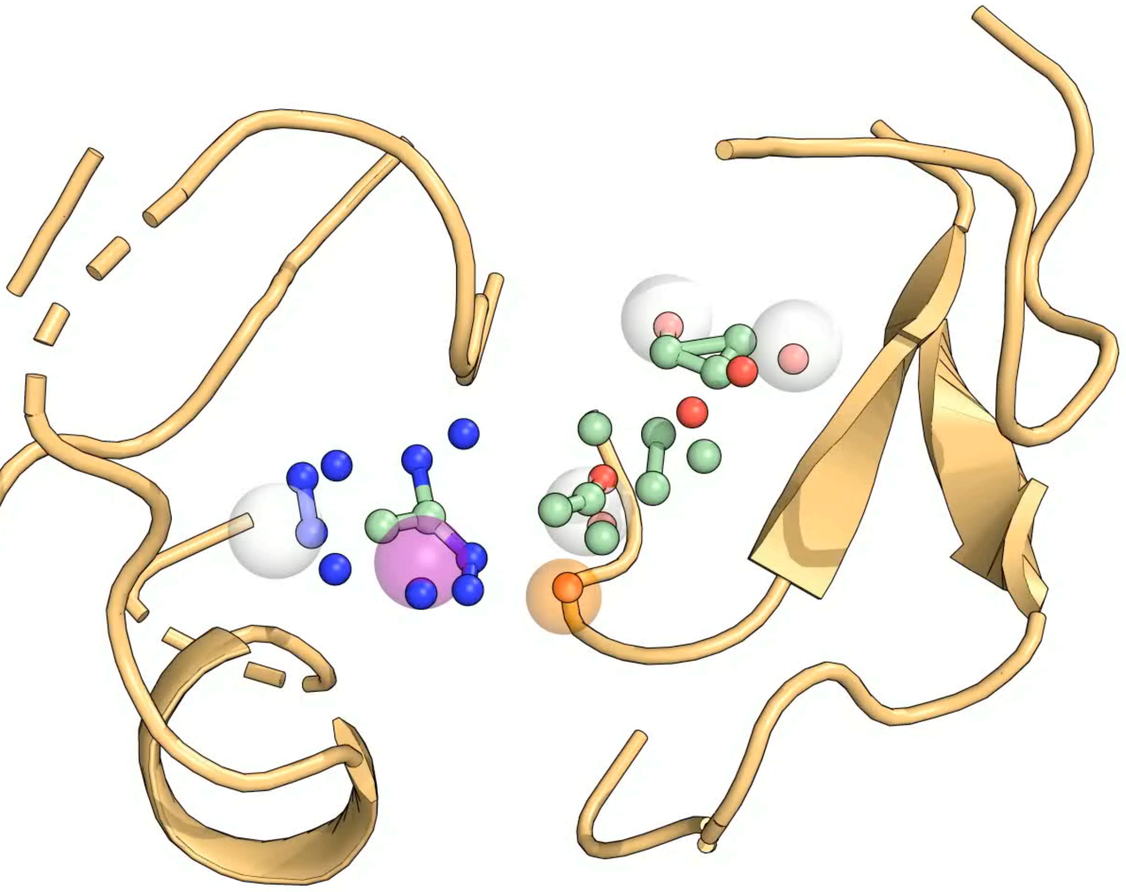


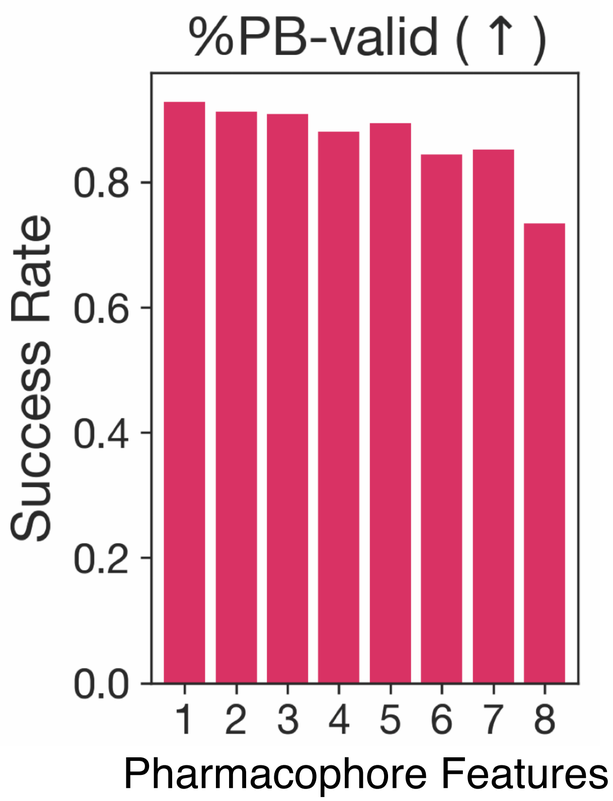
OMTRA: Docking

Evaluated on PoseBusters test set.

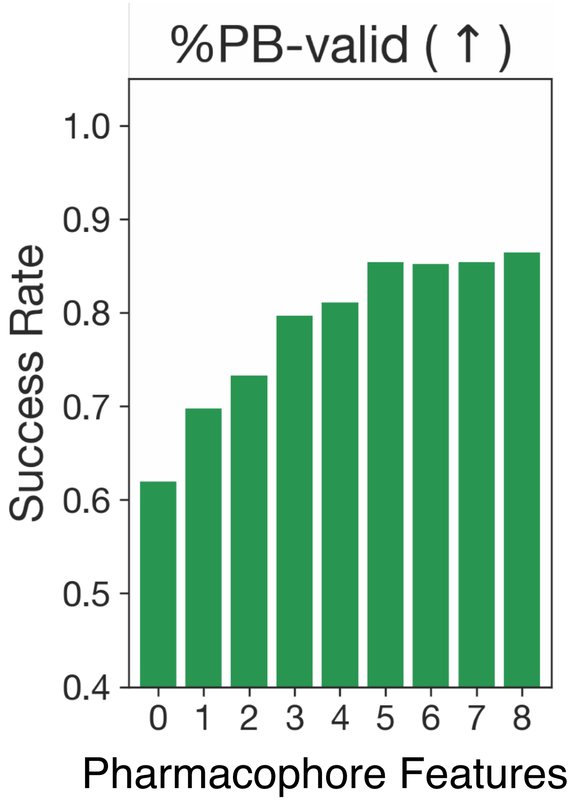


OMTRA: Pharmacophore Conditioning



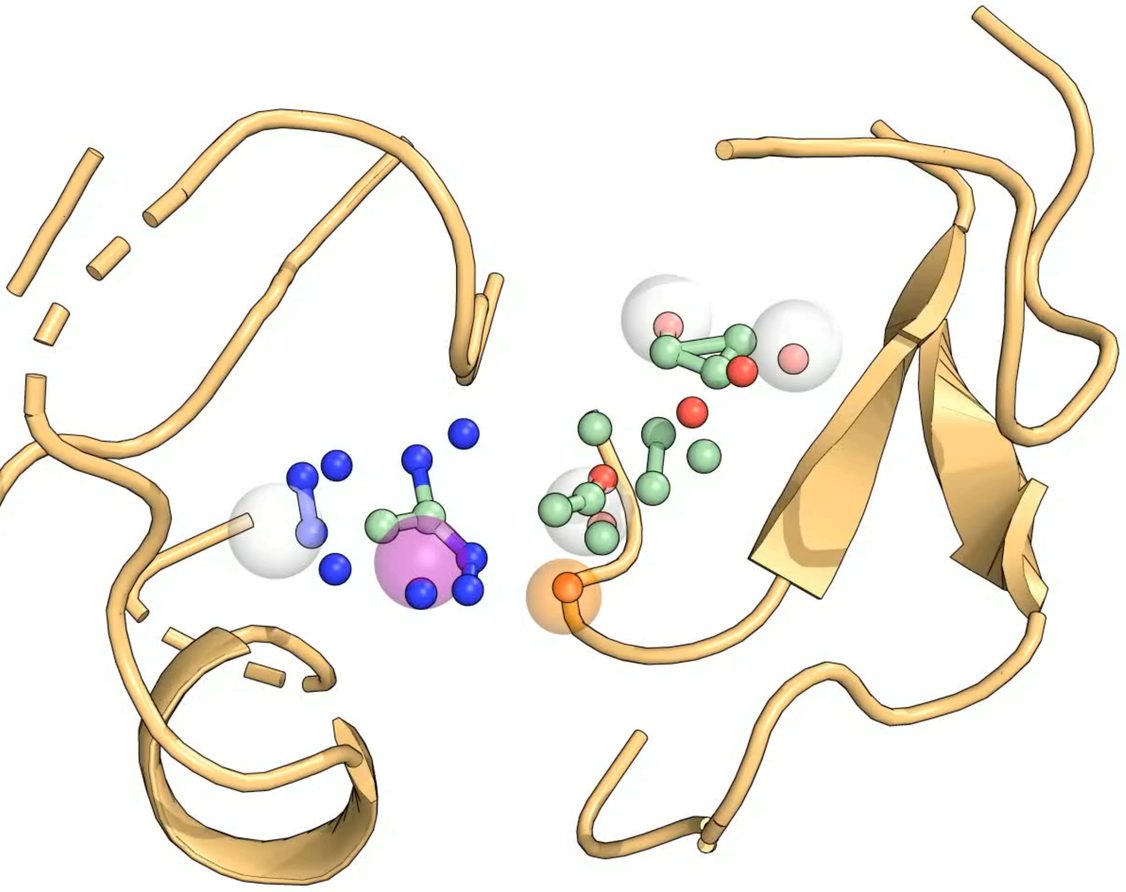


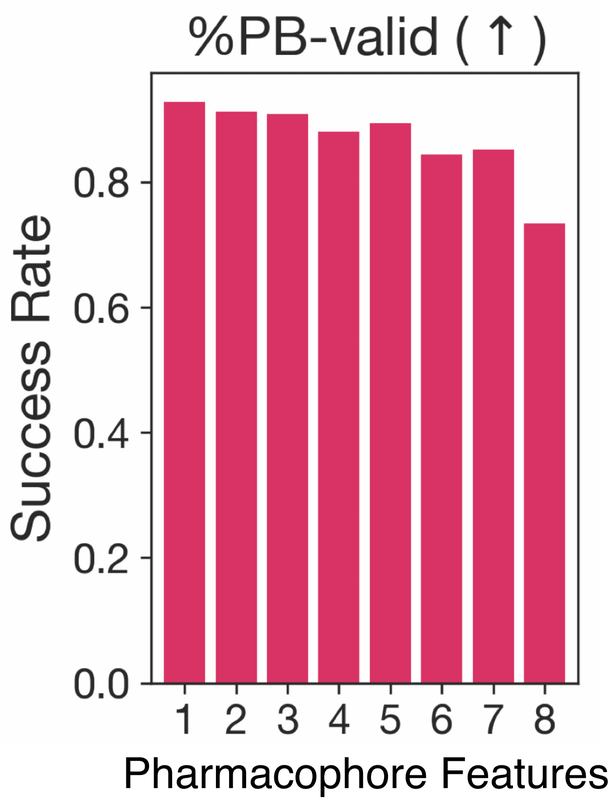
de novo Design



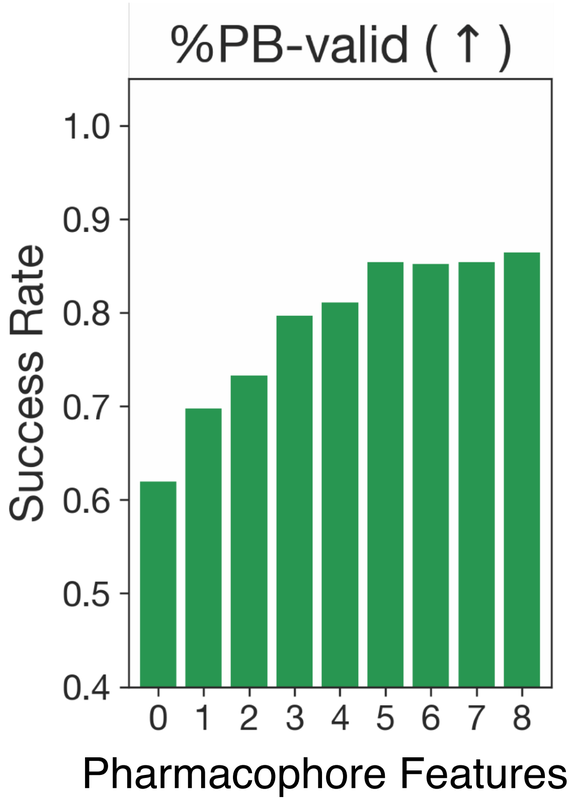
Docking

OMTRA: Pharmacophore Conditioning





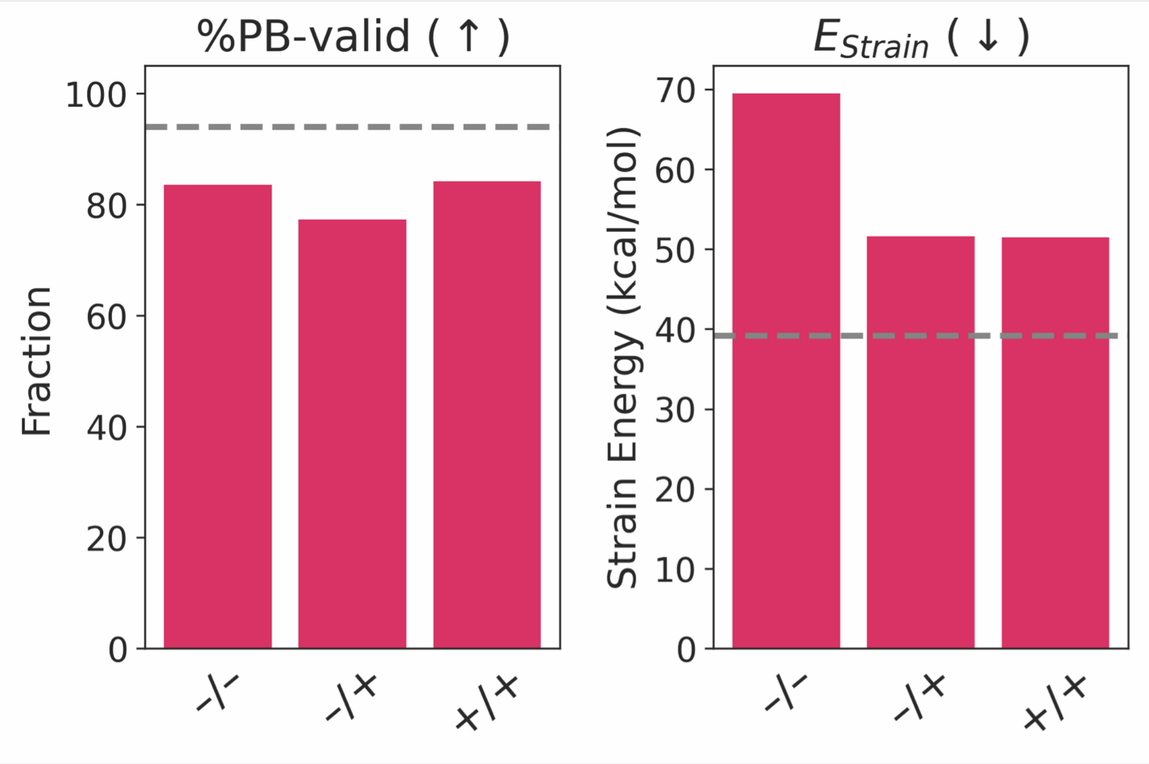
de novo Design



Docking

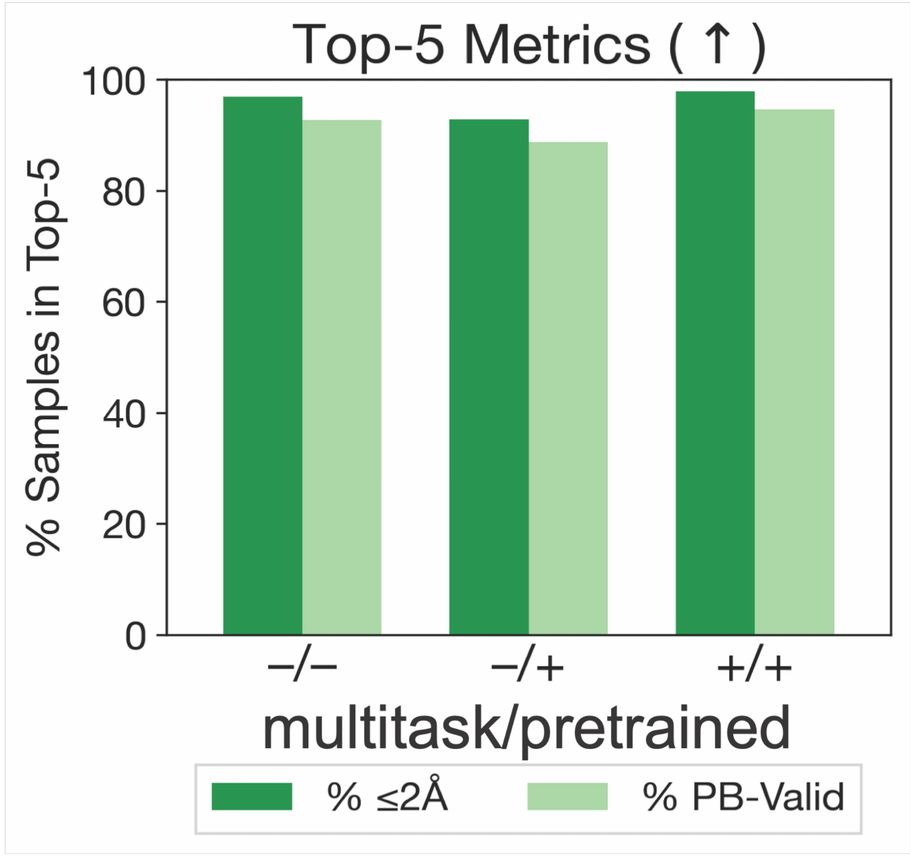
OMTRA: Transfer Learning?

Evaluated on PLINDER test set.



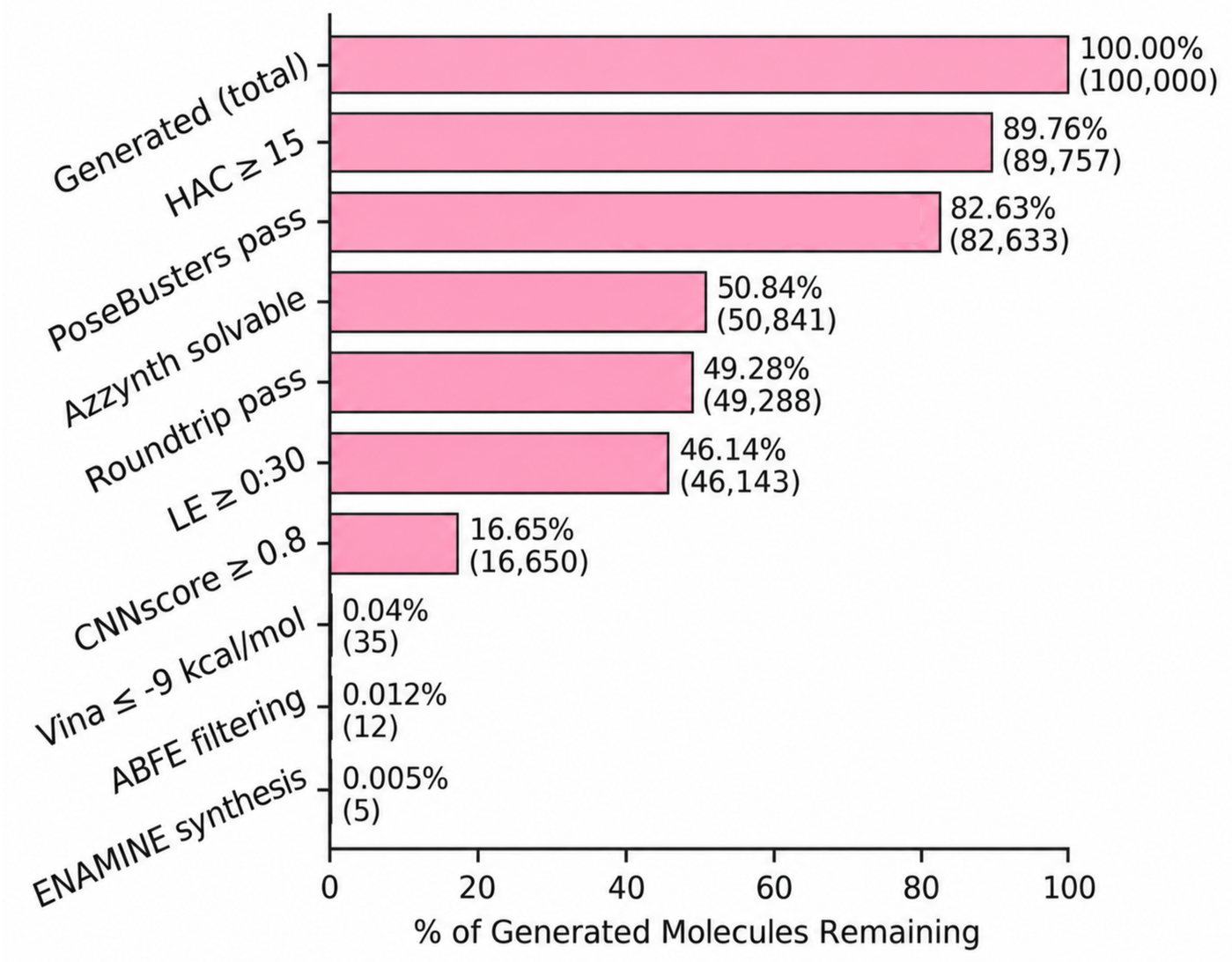
multitask/pretrained

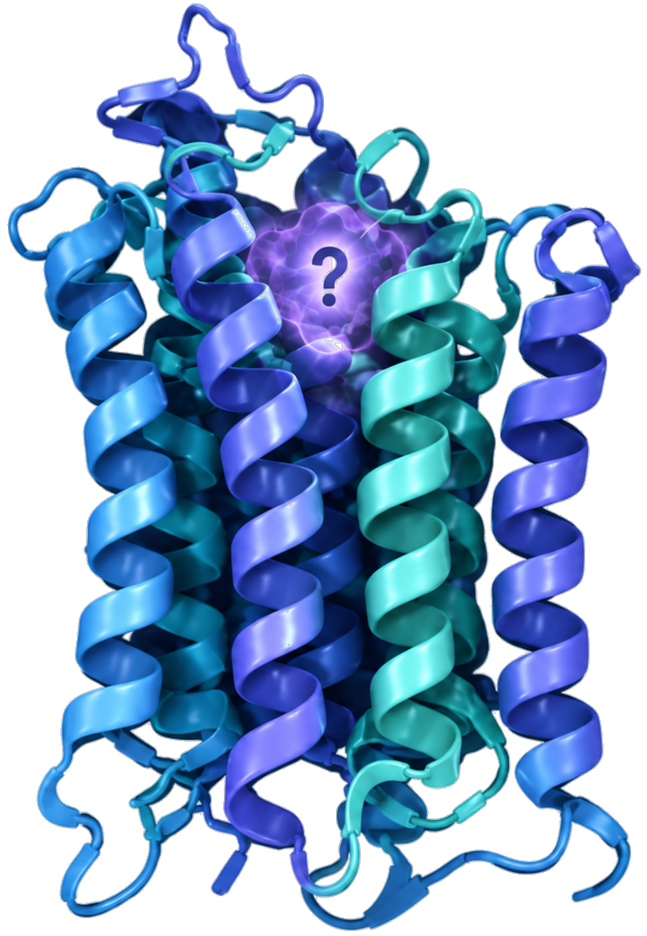
de novo Design

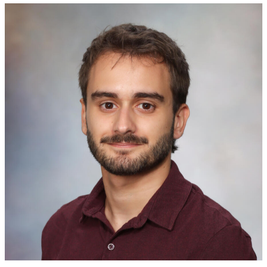


Docking

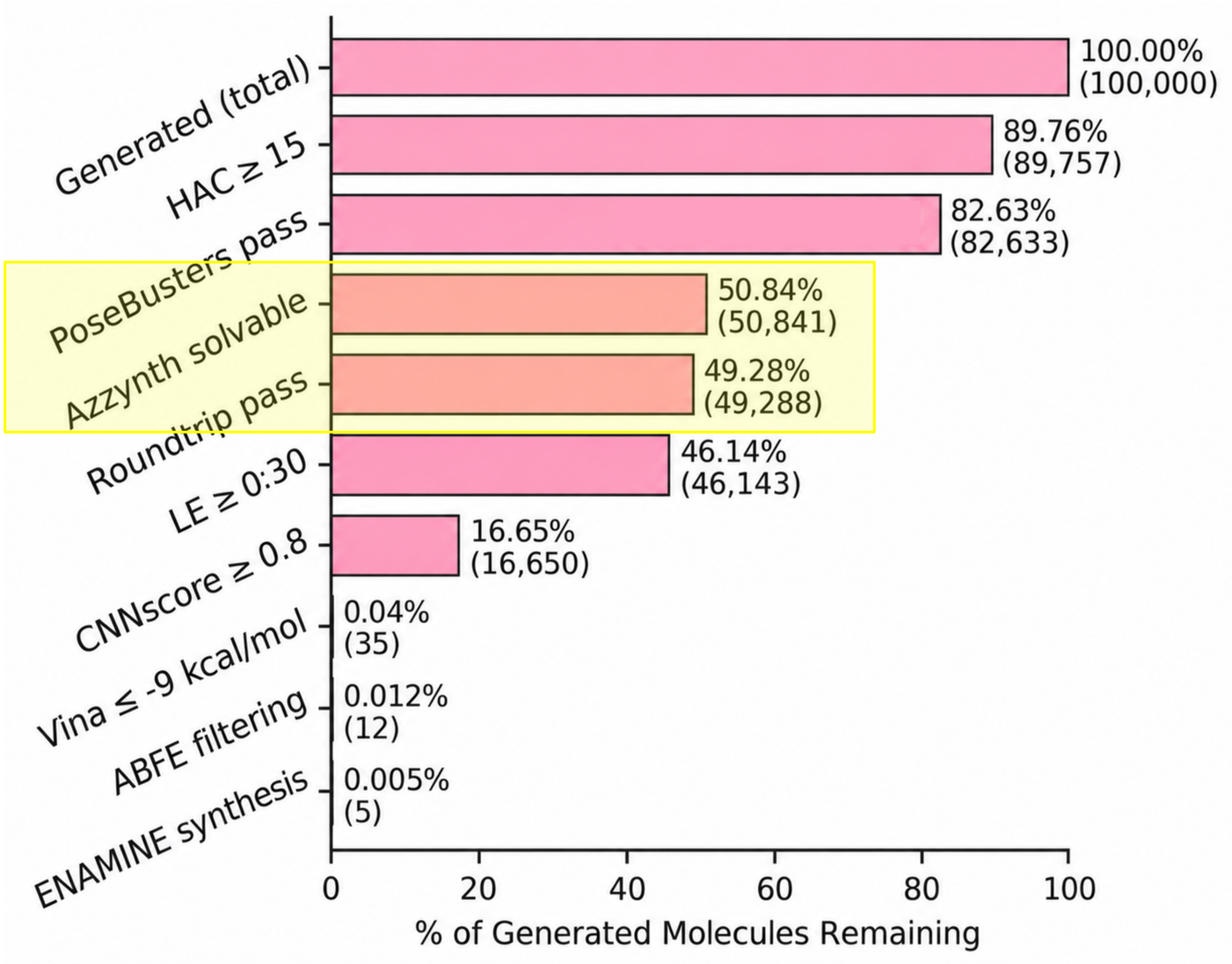
Synthesizeability

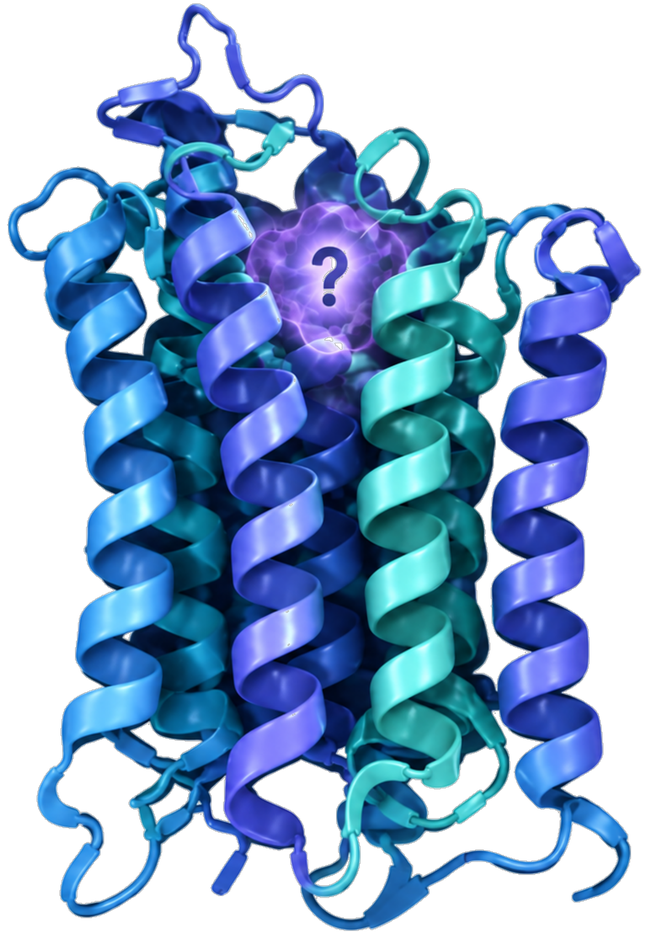


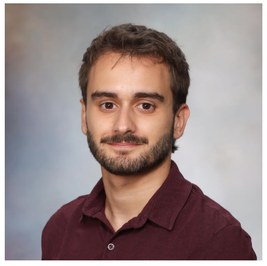




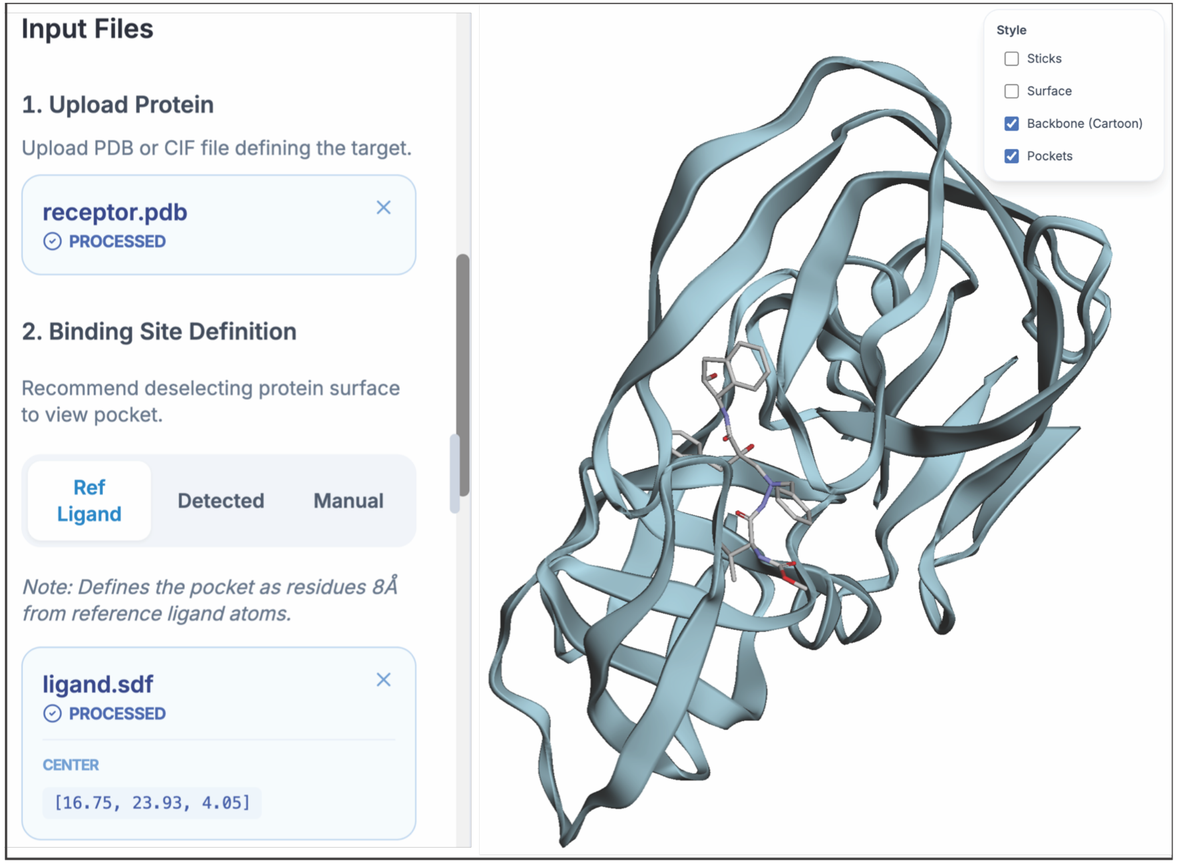
Synthesizeability





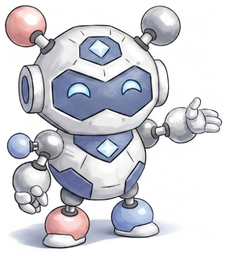


Web App


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

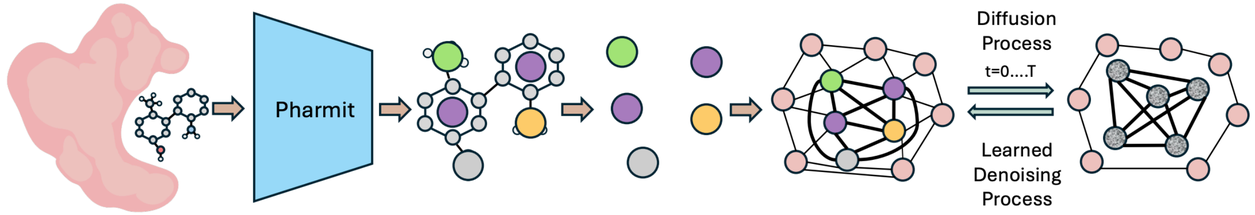

Fix atoms (optional) 14
Fixed atoms keep the same position during generation.
By fragment
By atom
Click fragments in the viewer or toggle below. Click to toggle · Shift+drag to box-select.
Frag 0 (27)
Frag 1 (1)
Frag 2 (2)
Frag 3 (2)
Frag 4 (1)
Frag 5 (10)
All None
3 of 6 fragments fixed
14 atoms fixed
Clear Invert

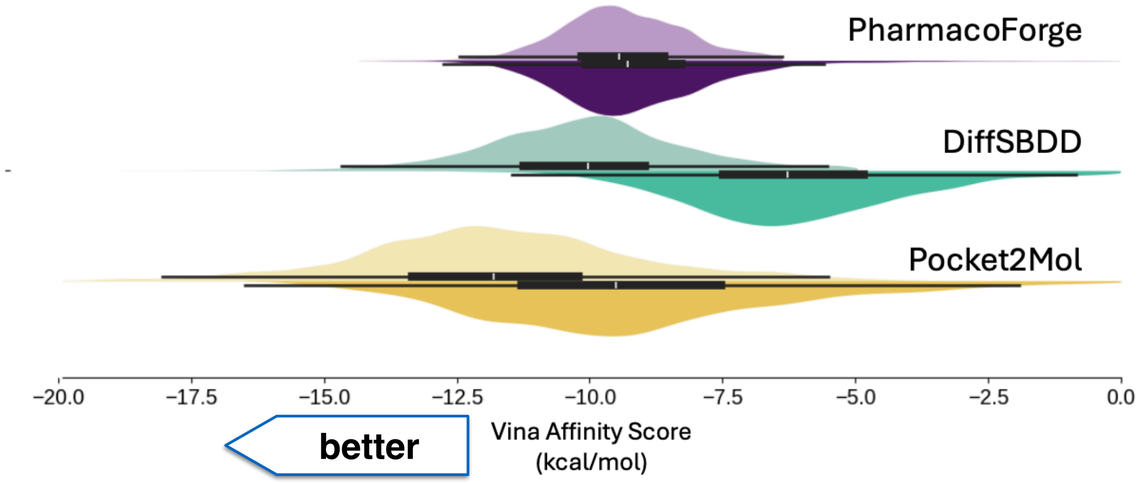




Generating *Descriptions* of Molecules

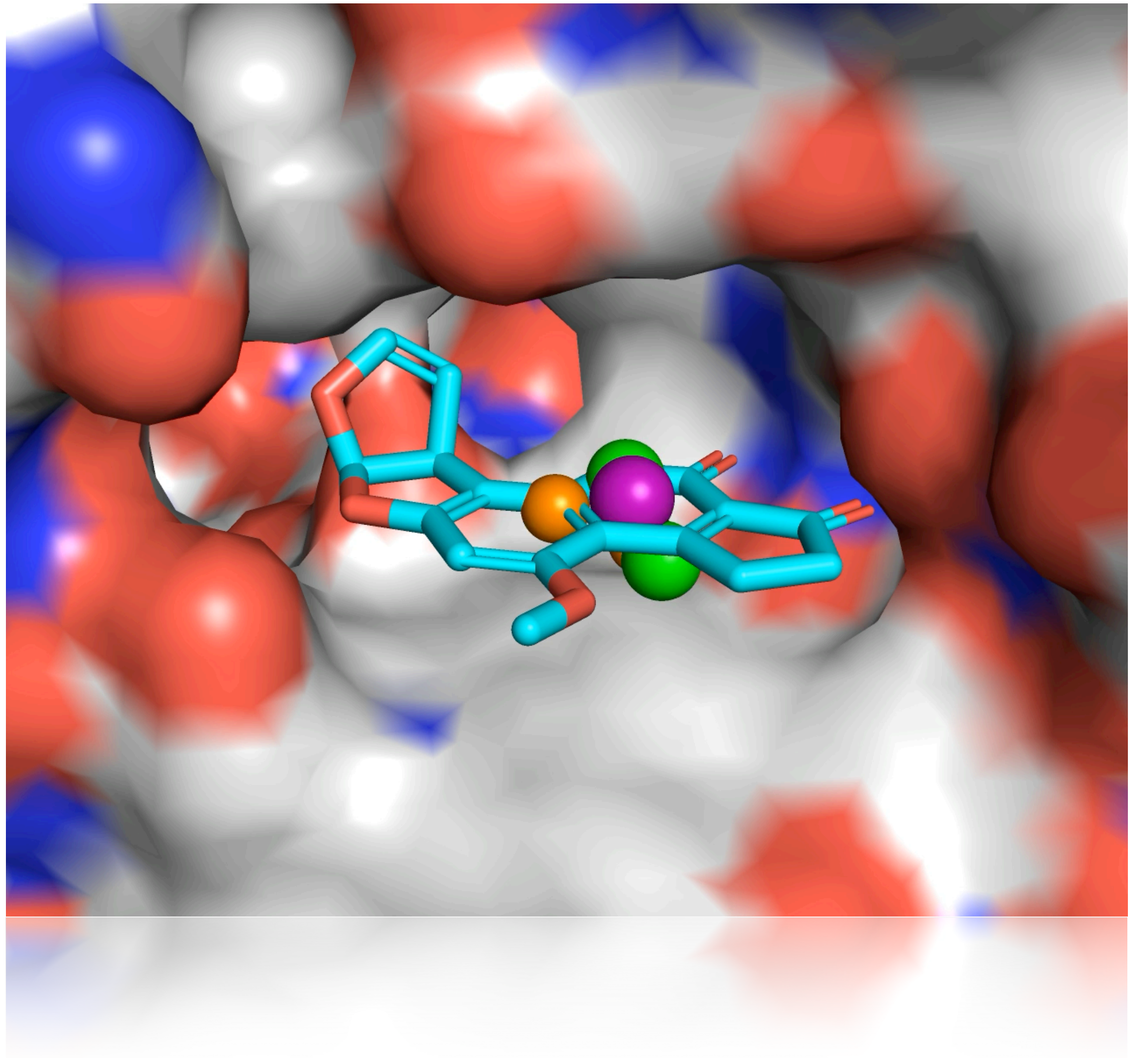
PharmacoForge









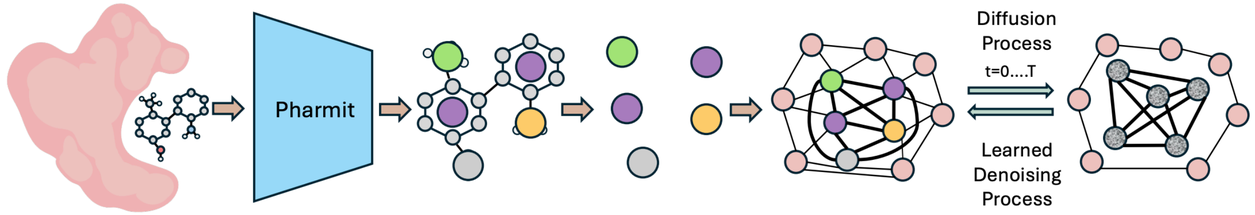



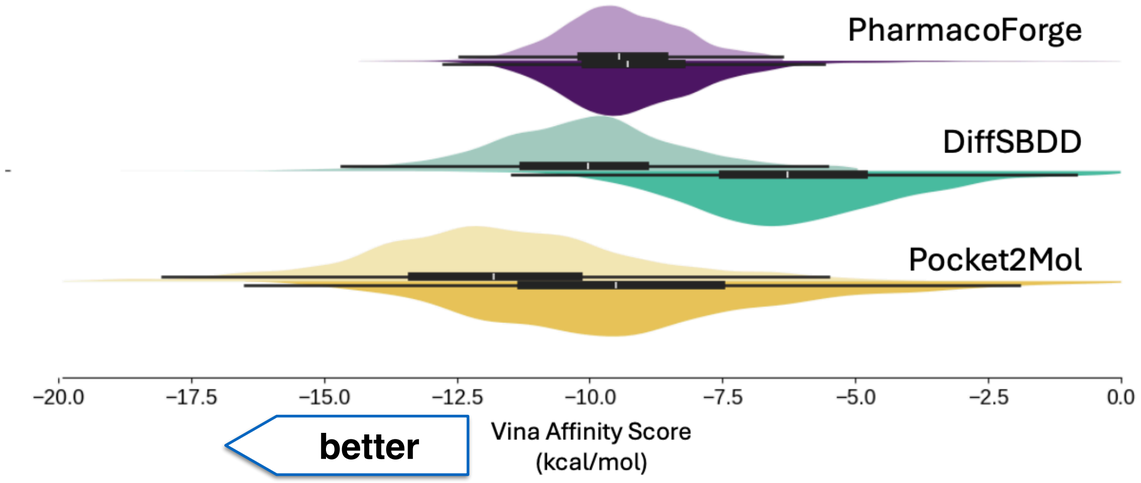
PharmacoForge: pharmacophore generation with diffusion models

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

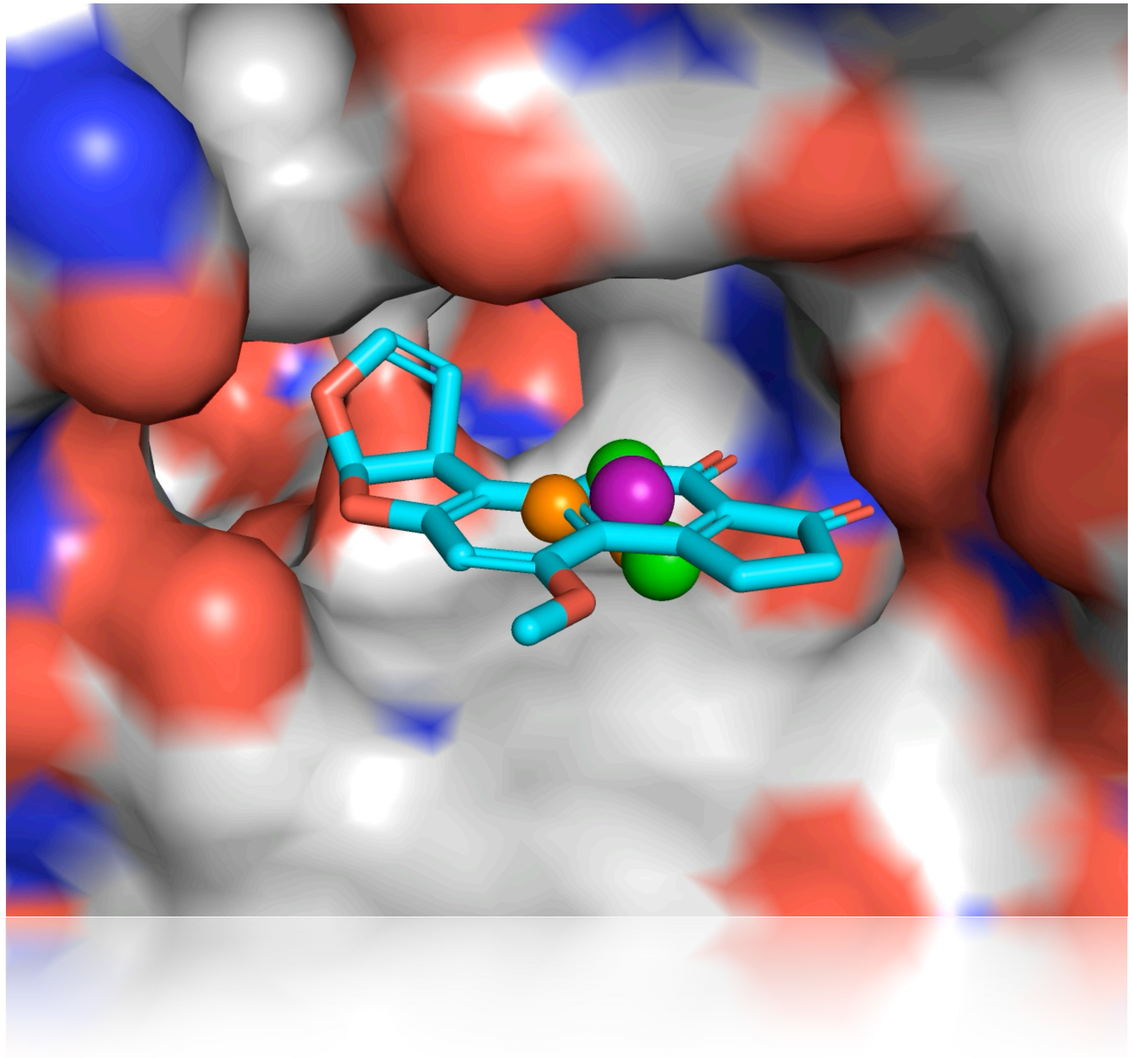
PharmacoForge









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]*

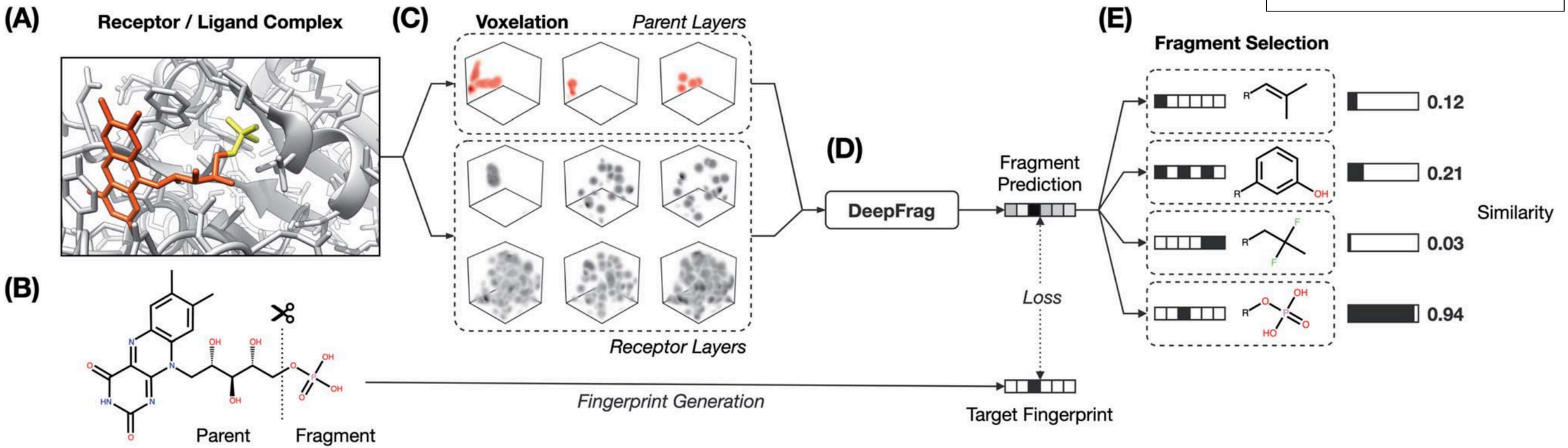
frontiers | Frontiers in Bioinformatics

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

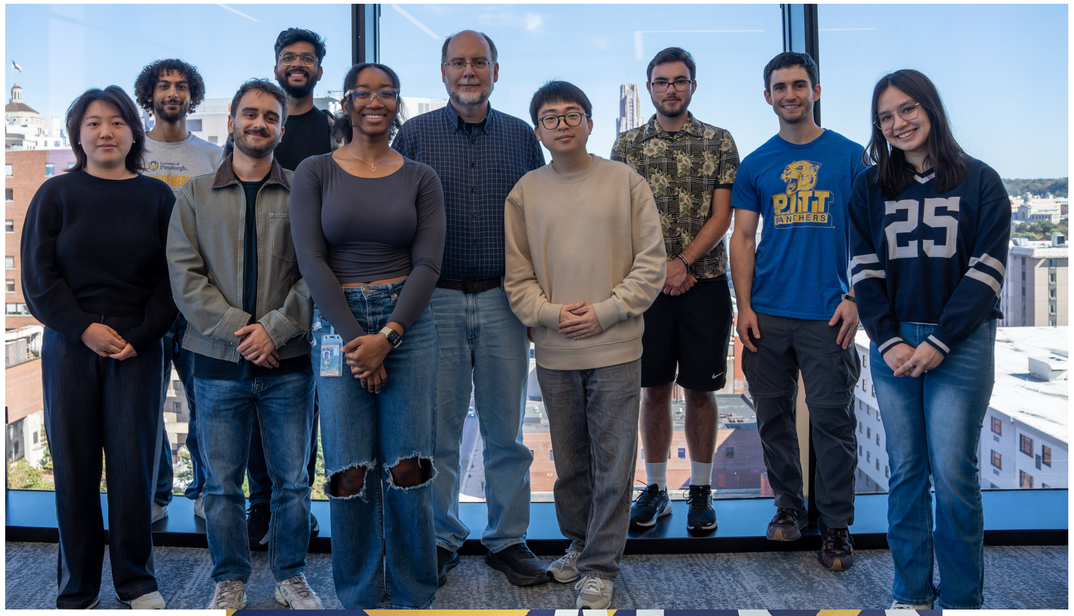
Harrison Green, [a] David R. Koes[b] and Jacob D. Durrant *[a]

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





Acknowledgements





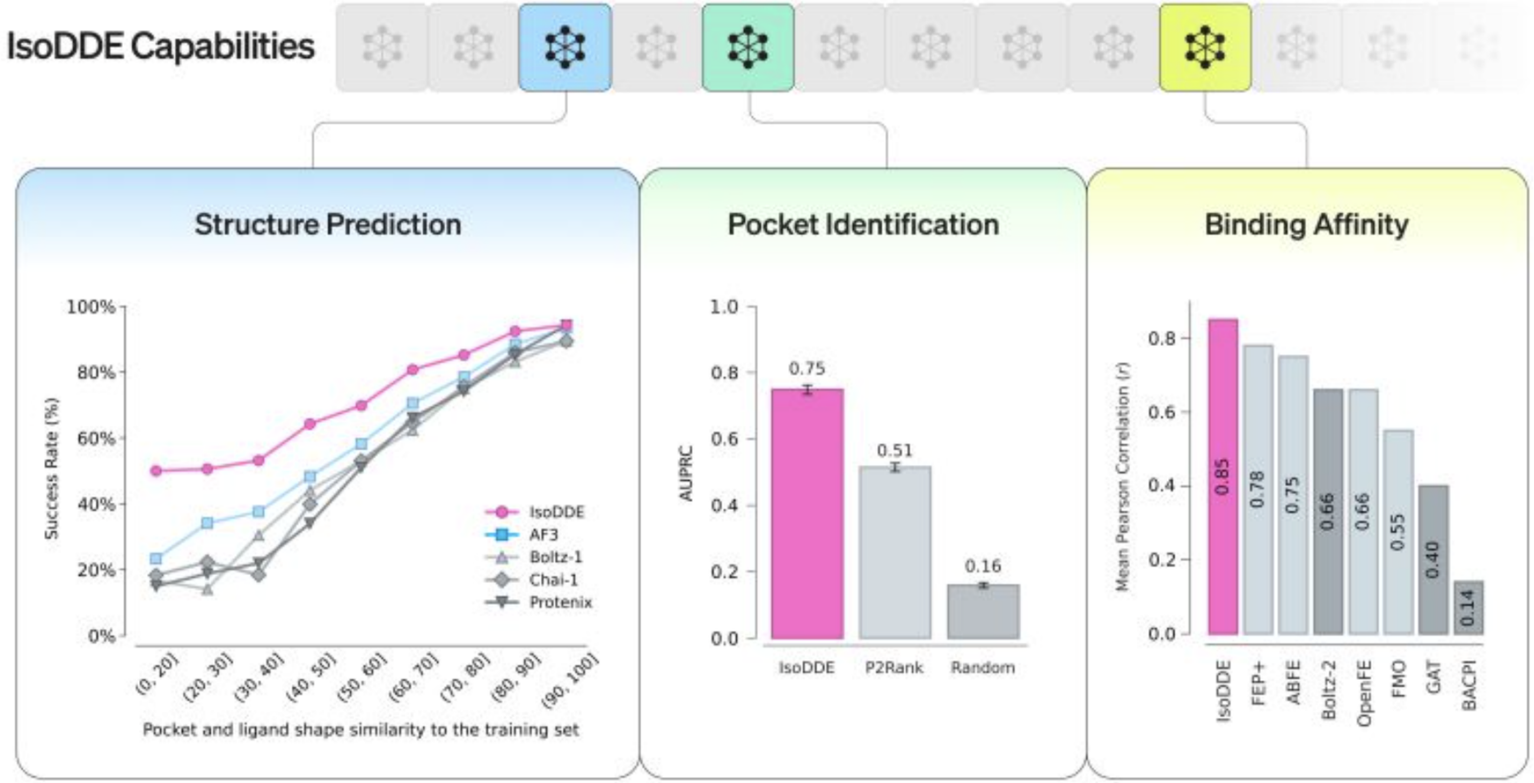
Current
Emma Flynn
Riya Shah
Rishal Aggarwal
Daniel Penaherrera
Jacky Chen
Fareeda Abu-Juam
Ben Krummenacher
Marios Gavrielatos

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
NSF

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