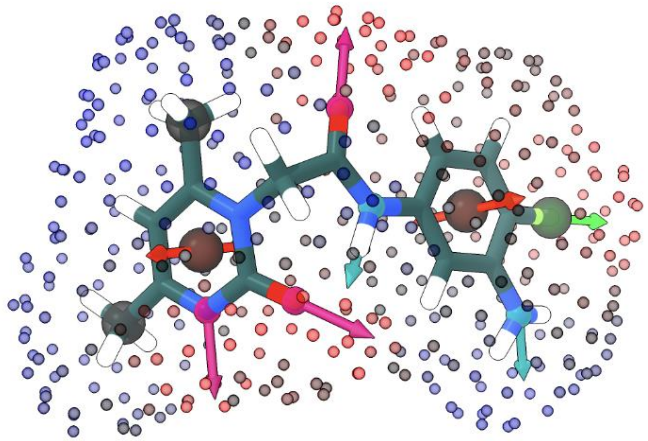
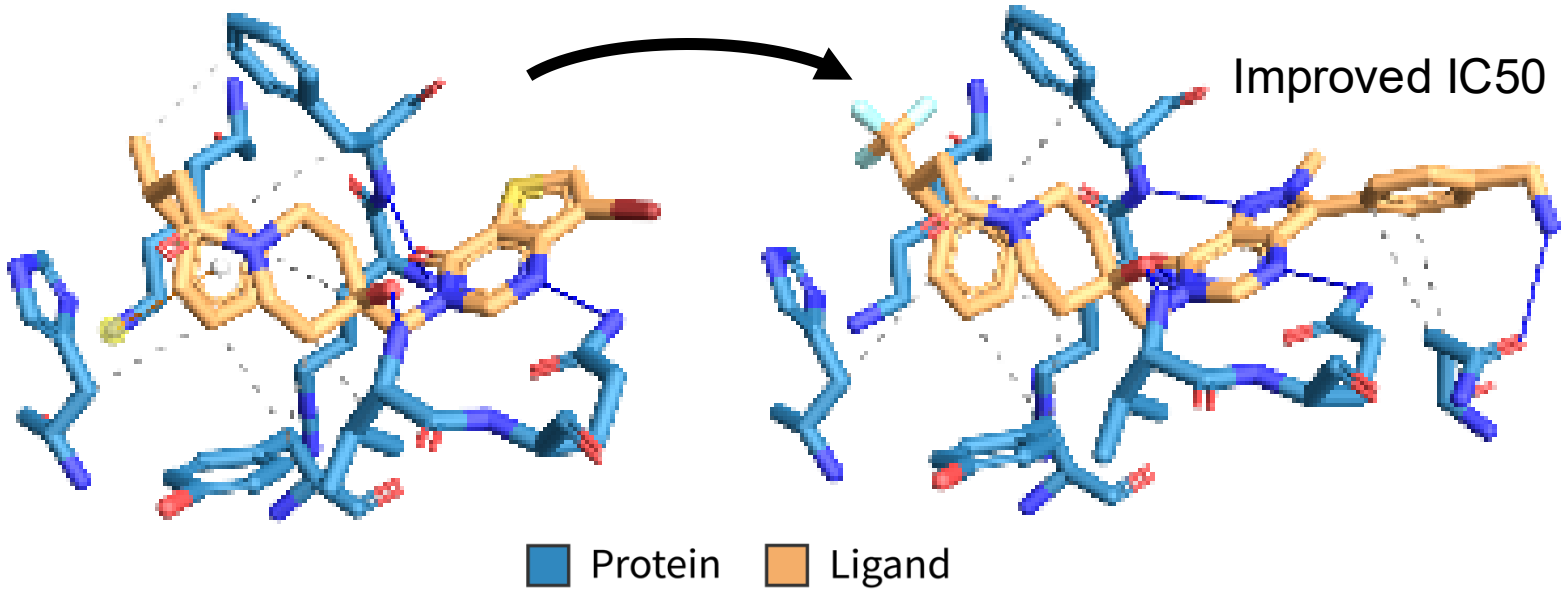
ShEPhERD: Diffusing shape, electrostatics, and pharmacophores for bioisosteric drug design



Keir Adams*, **Kento Abeywardane***, Jenna Fromer, and Connor Coley
Coley Research Group, MIT
ICLR 2025

Engineering molecules to engage in precise 3D intermolecular interactions underpins chemical design

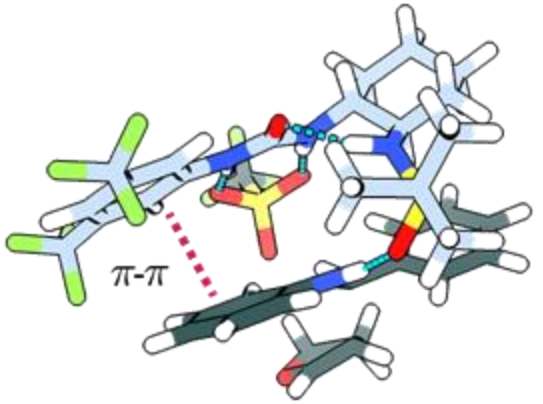
Small molecule drug design
(today's focus)



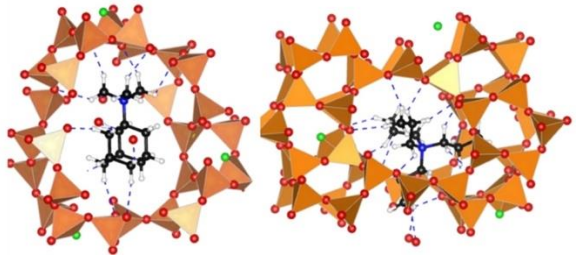
Improving protein-ligand interactions can improve potency

Other chemistries
(not discussed in this talk)

Organocatalyst design

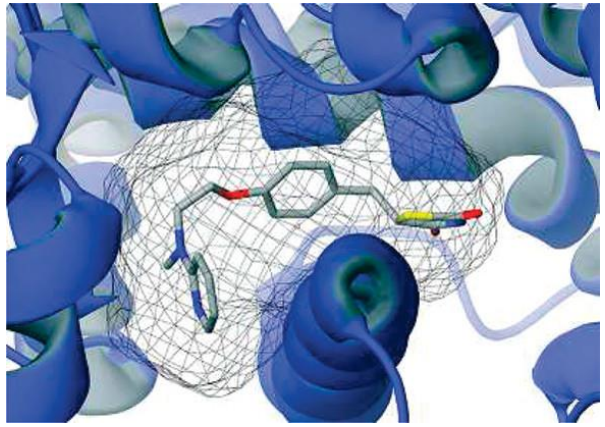


Functional materials design

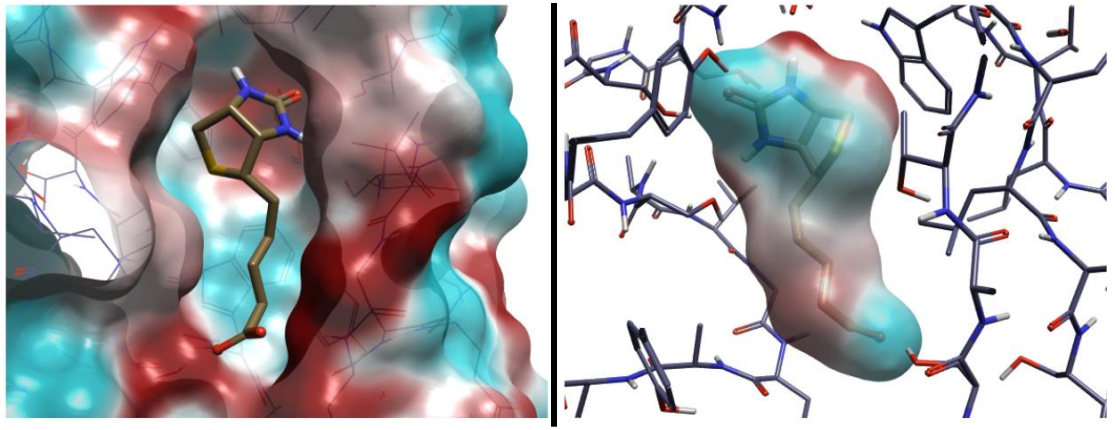


What factors into intermolecular interactions?

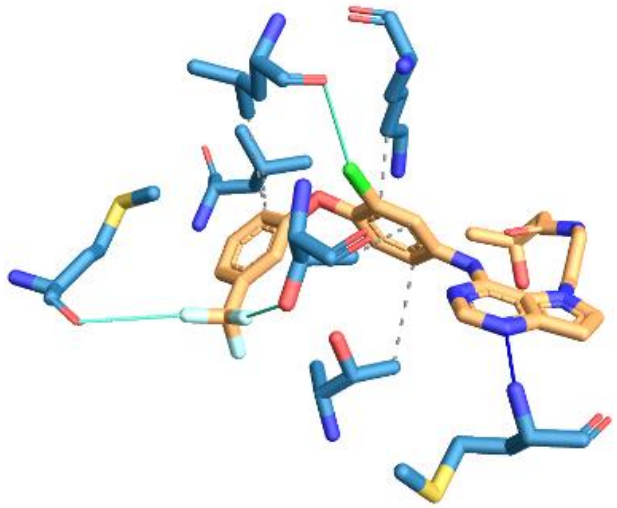
Shape



Electrostatic Potential (ESP)



Noncovalent interactions

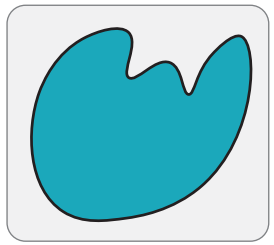


Lee et al. *J. Chem. Inf. Model*. (2009), 49, 11, 2419-2428
https://projects.volkamerlab.org/

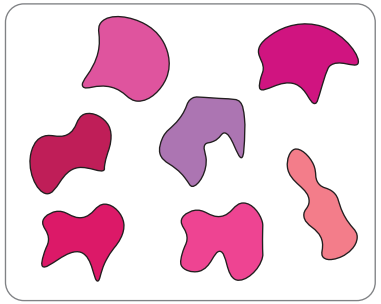
https://cresset-group.com/science/overview/

Contrasting structure-based vs. ligand-based drug design

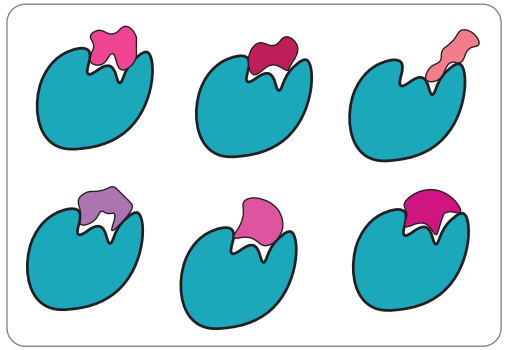
Structure-Based Drug Design (SBDD)



***given* protein target**

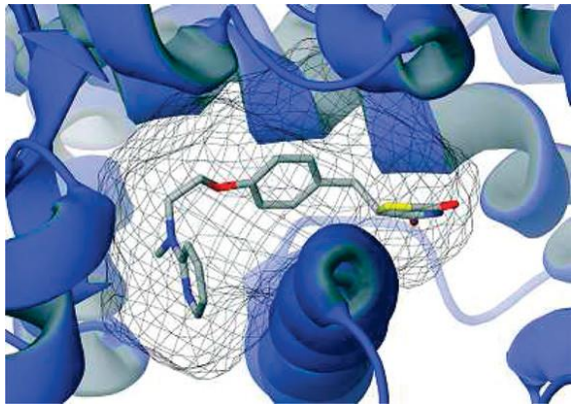


define molecular search space



screen for favorable drug-target interactions

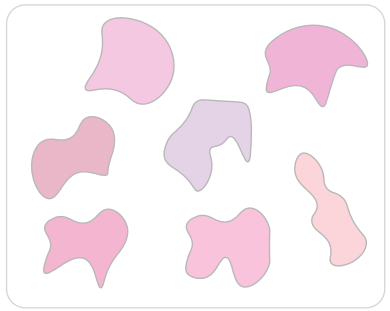
evaluate protein-ligand binding



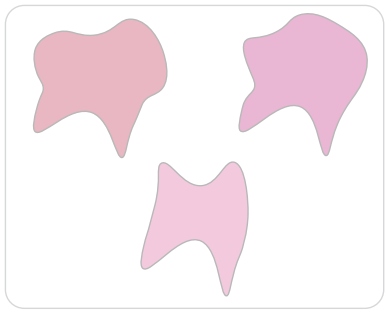
** Focus of this work

Ligand-Based Drug Design (LBDD)

***given* known ligand**



define molecular search space



screen for ligand analogues based on 2D/3D similarity

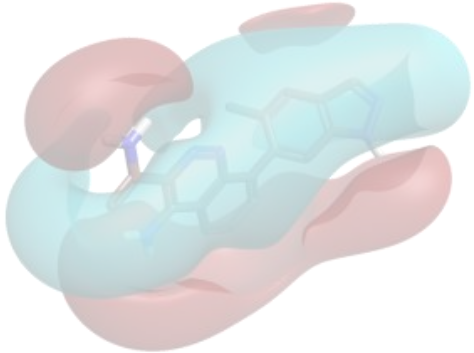
evaluate 3D similarity to known bioactive ligand

Lee et al. *J. Chem. Inf. Model*. (2009), 49, 11, 2419-2428
Kumar et al., *Front. Chem.* (2018), 6, 315
https://cresset-group.com/science/overview/
Qing et al. *Lig. Chann. Res.* (2014)

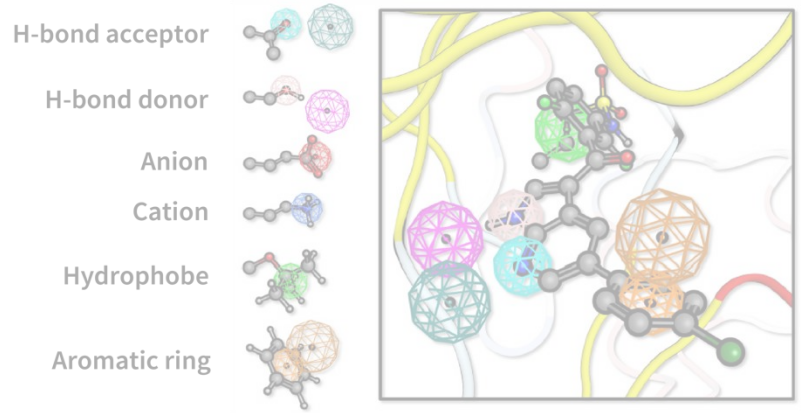
Bioisosteric substructures or ligands are identified via molecular or interaction similarity

Volumetric shape

Electrostatic potential



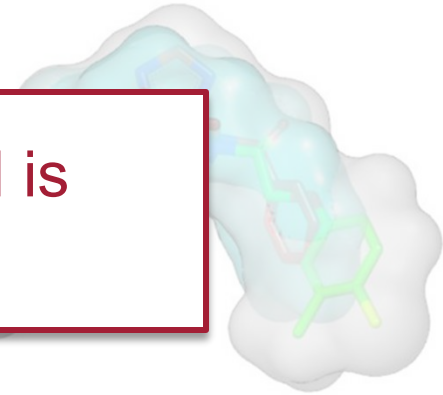
Pharmacophoric features



** Focus of this work

Ligand-Based Drug Desig (LBDD)

evaluate 3D similarity to known bioactive ligand

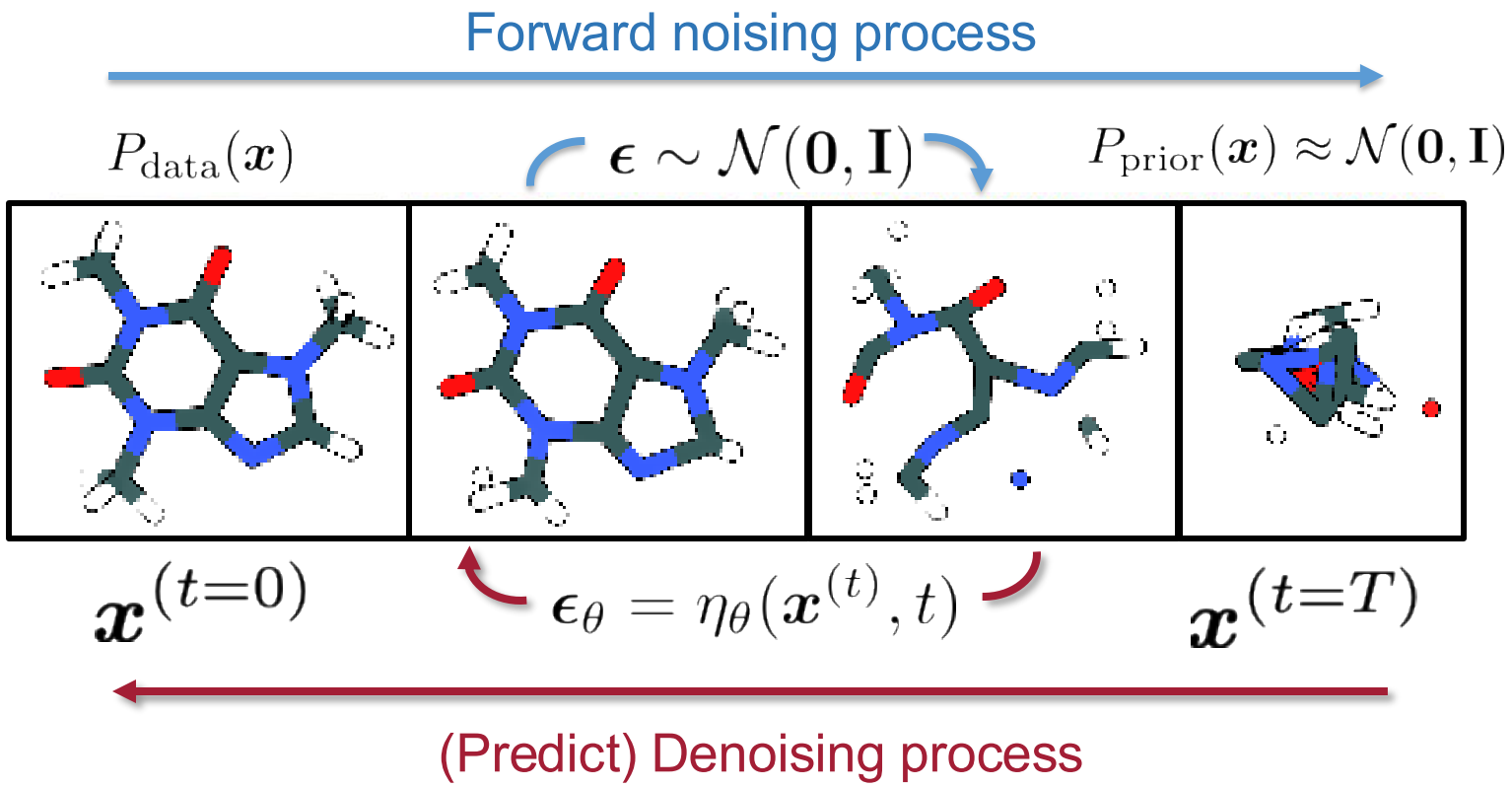


Traditional LBDD is inefficient due to random search and is restricted to pre-enumerated molecules.

Lee et al. *J. Chem. Inf. Model*. (2009), 49, 11, 2419-2428
Kumar et al., *Front. Chem.* (2018), 6, 315

https://cresset-group.com/science/overview/
Qing et al. *Lig. Chann. Res.* (2014)

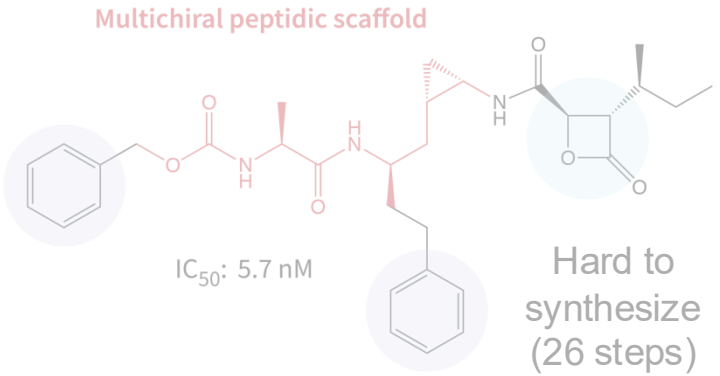
Generative models enables efficient search of the full chemical space around the optimal solution through conditional generation



Denoising 3D molecular structure

Exemplary LBDD challenges require the preservation of interactions

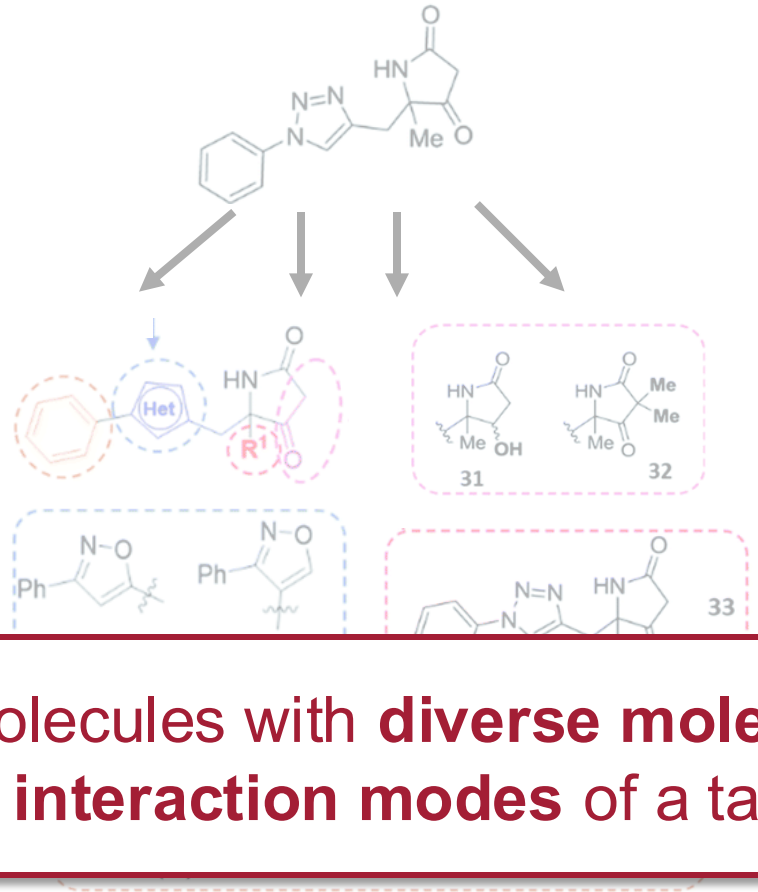
Natural product ligand hopping



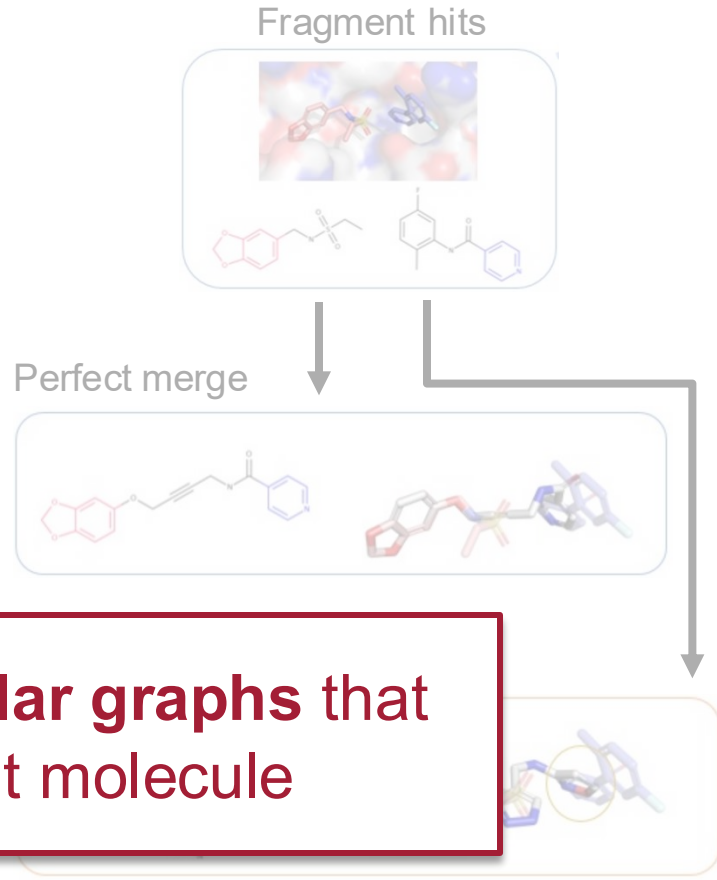
Hard to synthesize (26 steps)

Easy to synthesize (3 steps)

Bioactive hit diversification



Bioisosteric fragment merging



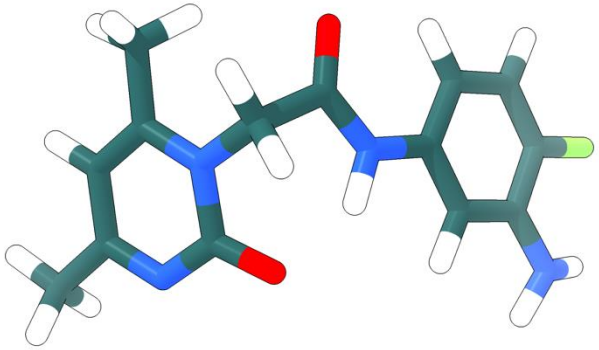
Goal: Generate molecules with **diverse molecular graphs** that **preserve the interaction modes** of a target molecule

Kawamura et al. *Chem. Comm.* (2014) **50,** 2445-2447
Kidd et al. *Chem. Sci.* (2020) **11**, 10792-10801

Wills et al. *BioRxiv.* (2024). 10.1101/2024.08.02.606367

Point cloud representation of a molecule

Molecular Structure (x_1)



$$\boldsymbol{x}_1 = (\boldsymbol{a}, \boldsymbol{C}, \boldsymbol{f}, \boldsymbol{B})$$

$\boldsymbol{a} \in \mathbb{R}^{n_1 \times N_a}$ (one-hot atom types)

$\boldsymbol{C} \in \mathbb{R}^{n_1 \times 3}$ (atomic positions)

$\boldsymbol{f} \in \mathbb{R}^{n_1 \times 5}$ (one-hot atomic formal charges)

$\boldsymbol{B} \in \mathbb{R}^{n_1 \times n_1 \times 5}$ (one-hot bond adjacency matrix / types)

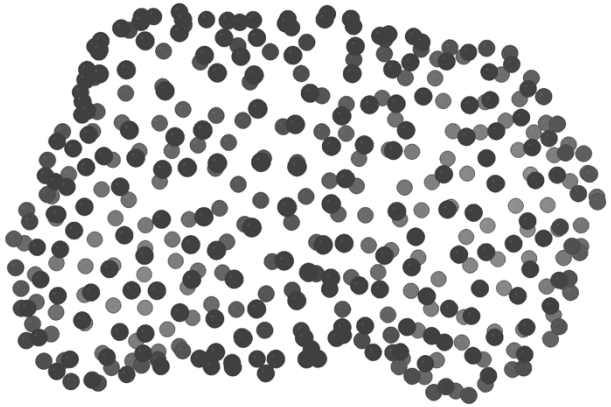
Requirements for interaction profiles

1. Expressively captures potential interaction modes
2. Decoupled from molecular structure

Point clouds can represent essential molecular interaction features

Shape (x_2)

Points sampled on solvent-accessible surface

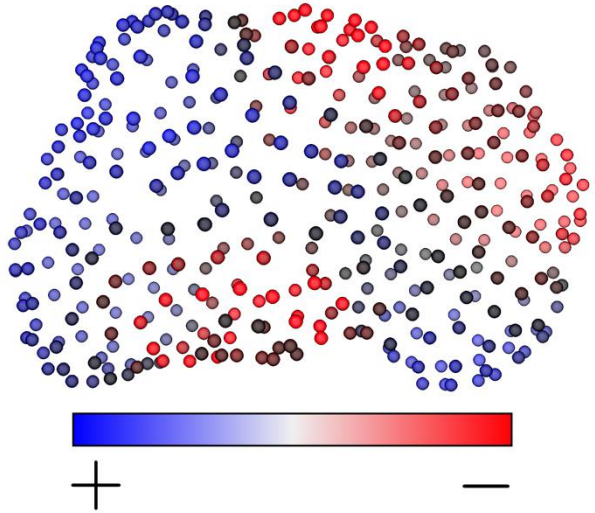


$$\boldsymbol{x}_2 = \boldsymbol{S}_2$$

$\boldsymbol{S}_2 \in \mathbb{R}^{n_2 \times 3}$ (positions)

Electrostatics (x_3)

Coulombic potential on surface points



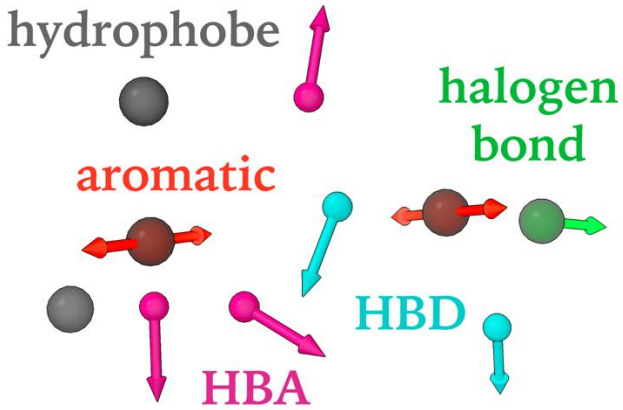
$$\boldsymbol{x}_3 = (\boldsymbol{S}_3, \boldsymbol{v})$$

$\boldsymbol{S}_3 \in \mathbb{R}^{n_3 \times 3}$ (positions)

$\boldsymbol{v} \in \mathbb{R}^{n_3}$ (ESP)

Pharmacophores (x_4)

Composed of pharmacophore position and vector point clouds



$$\boldsymbol{x}_4 = (\boldsymbol{p}, \boldsymbol{P}, \boldsymbol{V})$$

$\boldsymbol{p} \in \mathbb{R}^{n_4 \times N_p}$ (one-hot types)

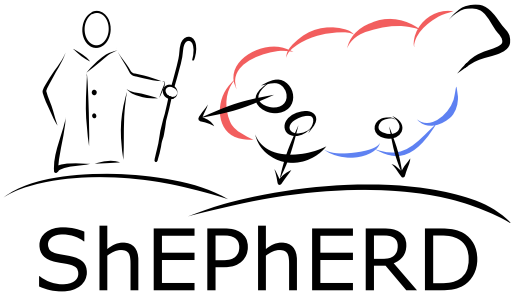
$\boldsymbol{P} \in \mathbb{R}^{n_4 \times 3}$ (positions)

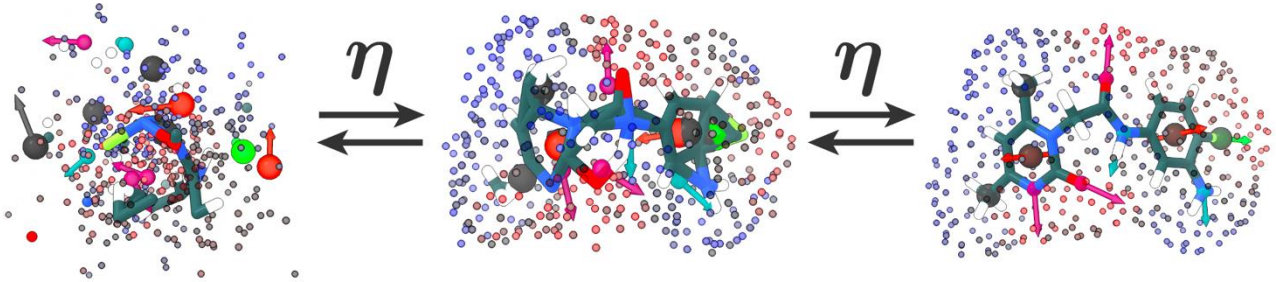
$\boldsymbol{V} \in \{\mathbb{S}^2, \boldsymbol{0}\}^{n_4}$ (unit vectors)

or $\in \mathbb{R}^{n_4 \times 3}$

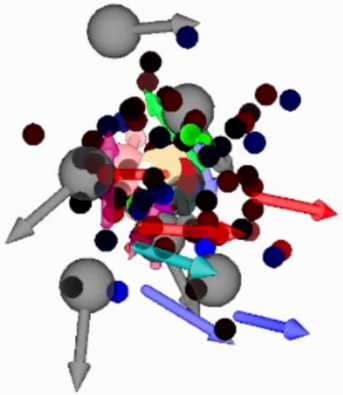
ShEPhERD defines a **joint** diffusion model over 3D molecules and explicit representations of their **shapes**, **electrostatics**, and **pharmacophores**

ShEPhERD **=** ***Sh****ape,* ***E****lectrostatics, and* ***Ph****armacophore* ***E****xplicit* ***R****epresentation* ***D****iffusion*





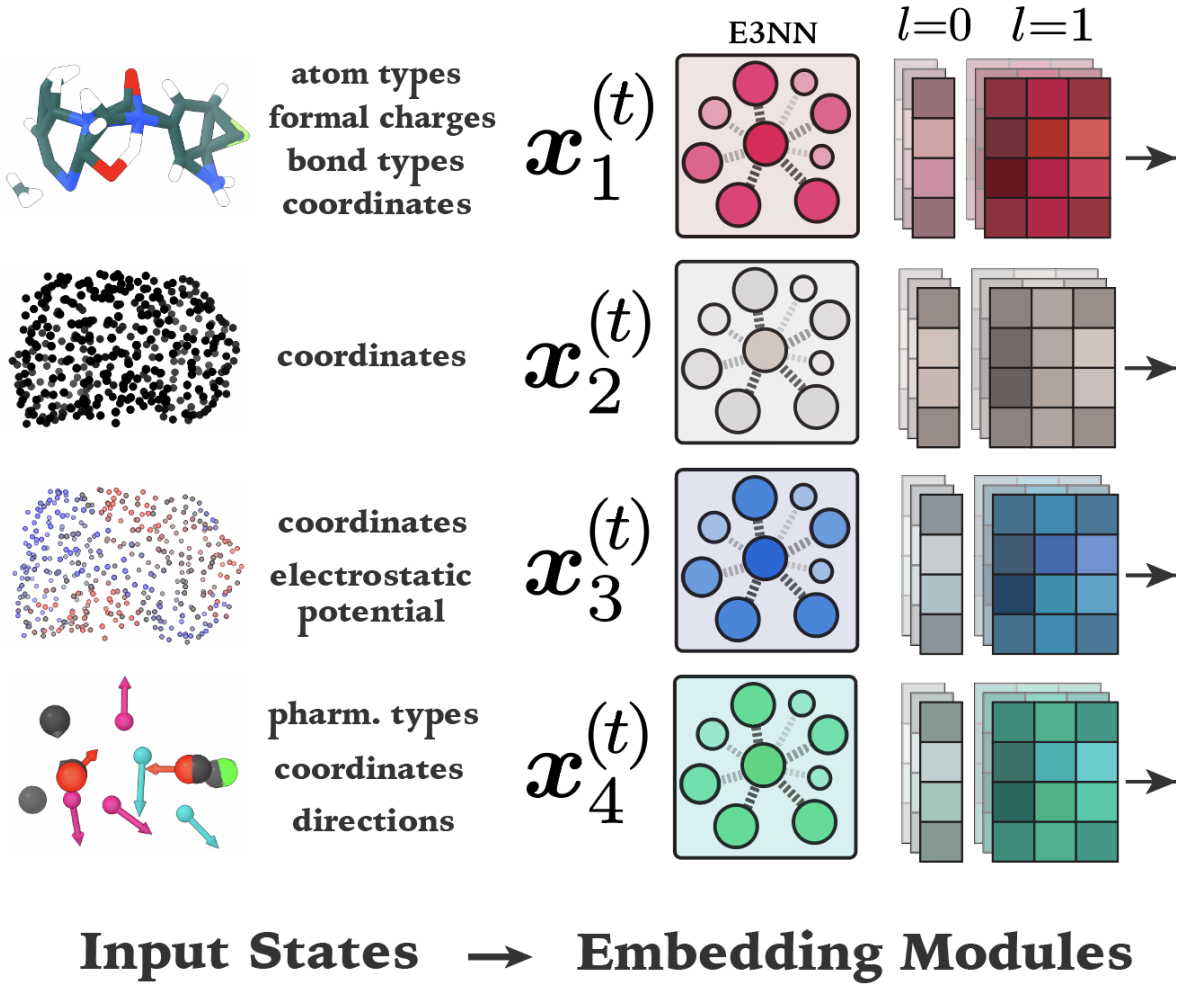
- Trained on 1.6M molecules from **MOSES**
- All data-types are **treated as continuous** and noised with isotropic Gaussian noise $\boldsymbol{\epsilon} \in N(\mathbf{0}, \boldsymbol{I})$
- **Simultaneously** denoises each point cloud
- Conditionally sample by **inpainting** interaction profiles



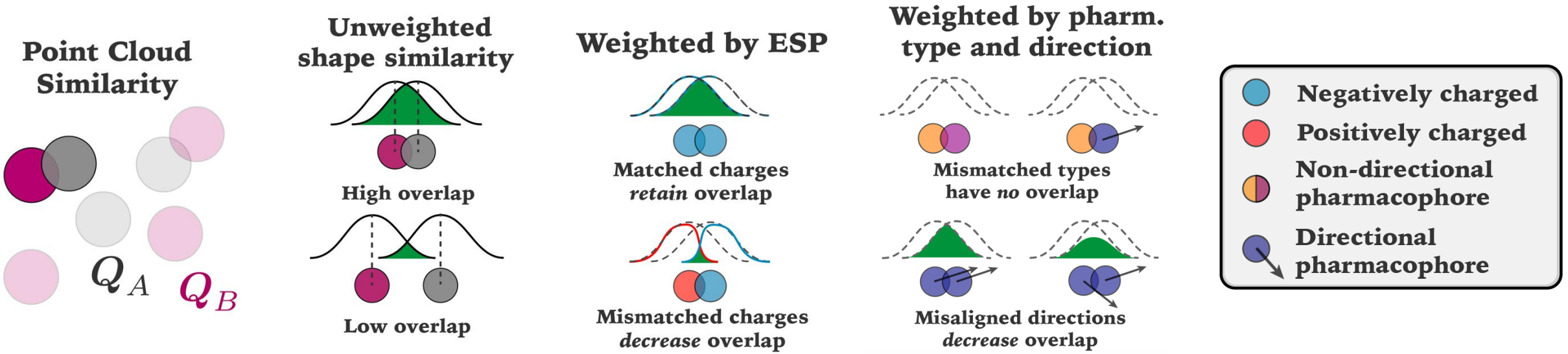
Joint denoising with ***ShEPhERD***
structure, shape, electrostatics, and pharmacophore

ShEPhERD's SE(3)-equivariant denoising architecture

Goal: Obtain the denoised state $(\boldsymbol{x}_1^{(t-1)}, \boldsymbol{x}_2^{(t-1)}, \boldsymbol{x}_3^{(t-1)}, \boldsymbol{x}_4^{(t-1)})$ by predicting the (forward) noises



Generated samples conditioned on a target interaction profiles **maintain high 3D interaction similarity** as measured by **Gaussian overlap**



Tanimoto similarity

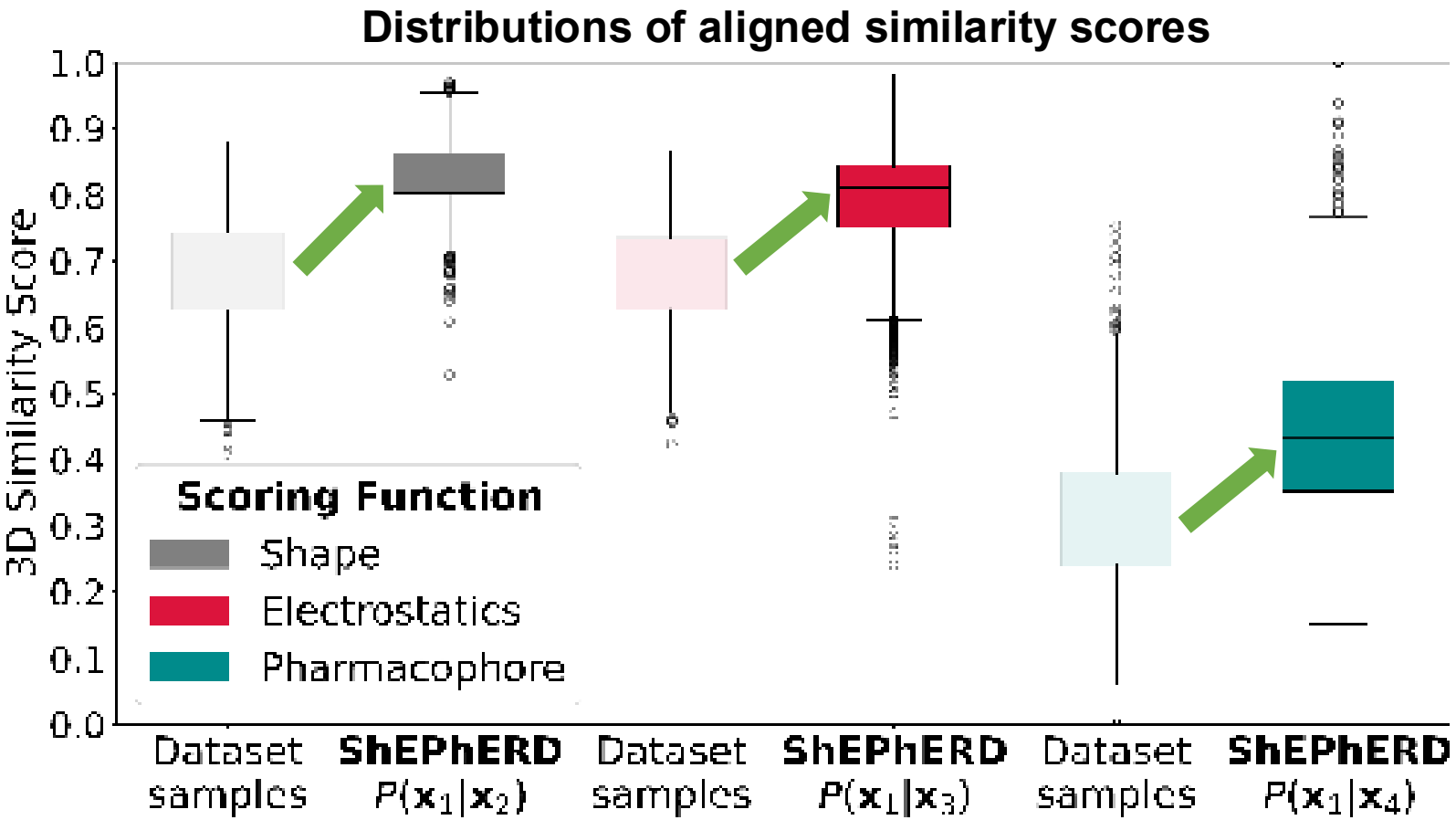
$$\mathrm{sim}^*(\boldsymbol{Q}_A, \boldsymbol{Q}_B) = \frac{O_{AB}}{O_{AA} + O_{BB} - O_{AB}}$$

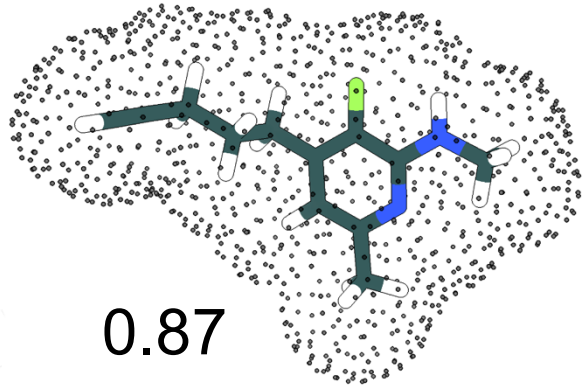
Gaussian overlap

$$O_{AB} = \sum_{a \in \boldsymbol{Q}_A} \sum_{b \in \boldsymbol{Q}_B} w_{ab} \left(\frac{\pi}{2\alpha}\right)^{\frac{3}{2}} \exp\left(-\frac{\alpha}{2}||\boldsymbol{r}_a - \boldsymbol{r}_b||^2\right)$$

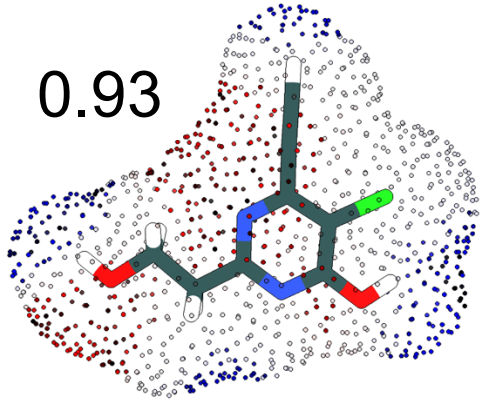
Grant et al., *J. Comp. Chem.* (1996). 17, 14, 1653-1666
Wahl et al. *JCIM*. (2024). 64, 15 5944-5953
Adams, K.*, **Abeywardane, K.***, Fromer, J., Coley, C.W. *ICLR*, (2025), arXiv:2411.04130

ShEPhERD enriches 3D similarity distributions of interaction profiles

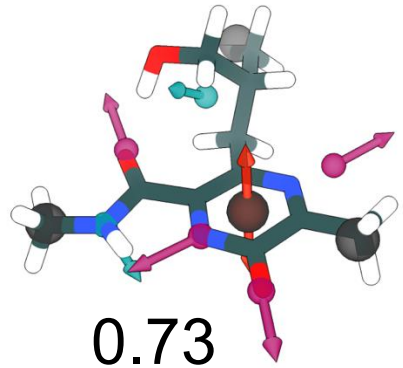




Shape-conditioned



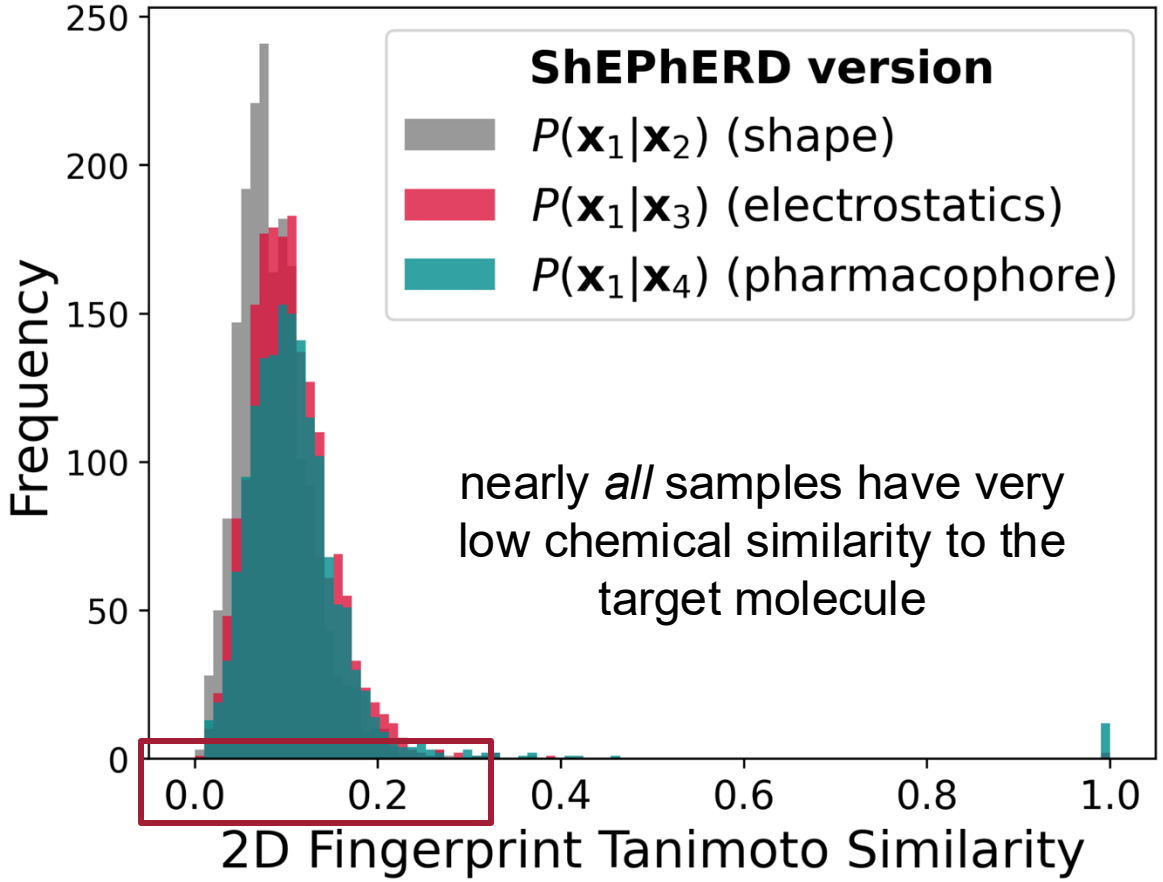
ESP-conditioned



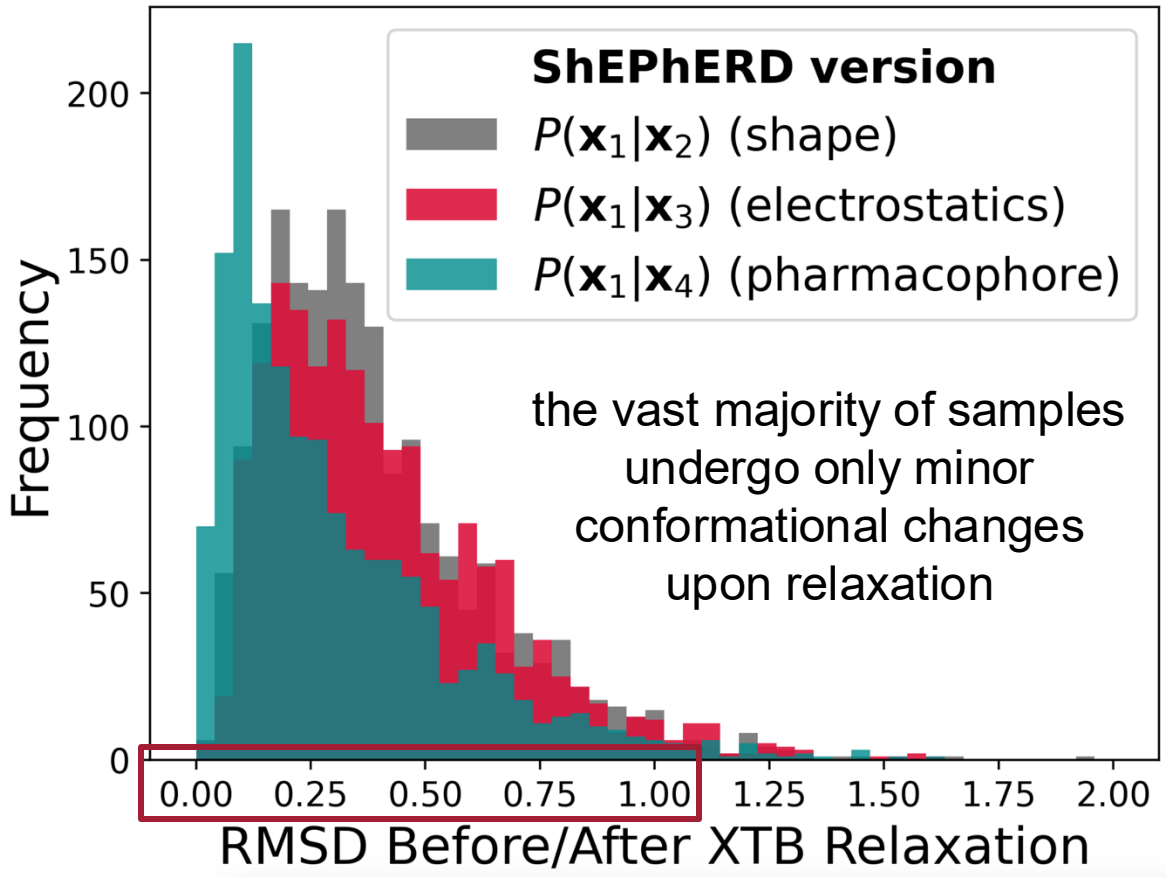
Pharmacophore-conditioned

ShEPhERD generates **chemically diverse** and **locally stable** conformers

Graph similarity

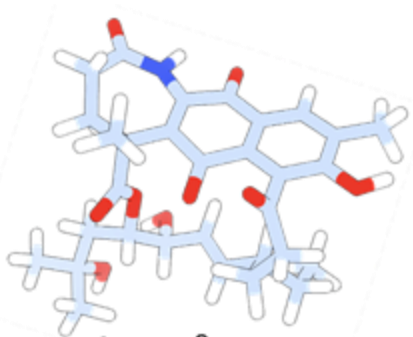


RMSD after relaxation



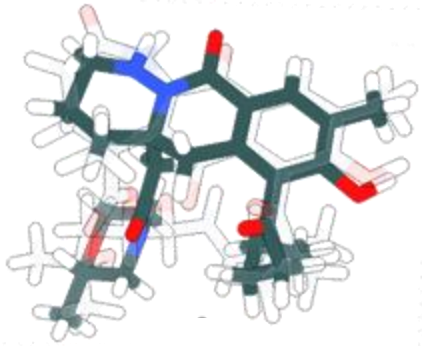
ShEPhERD enables natural product ligand hopping

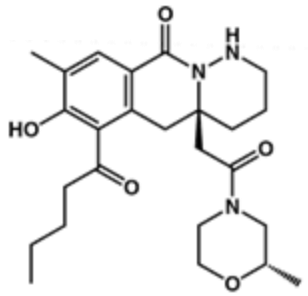
Target



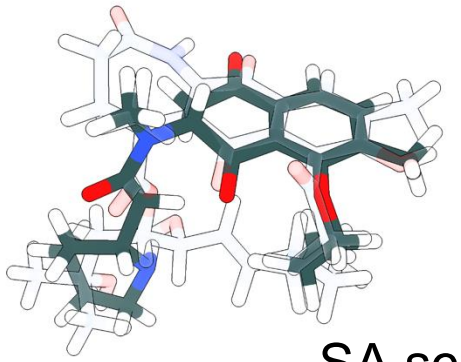
SA score: 6.6

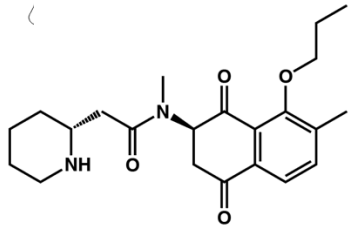
Example analogues



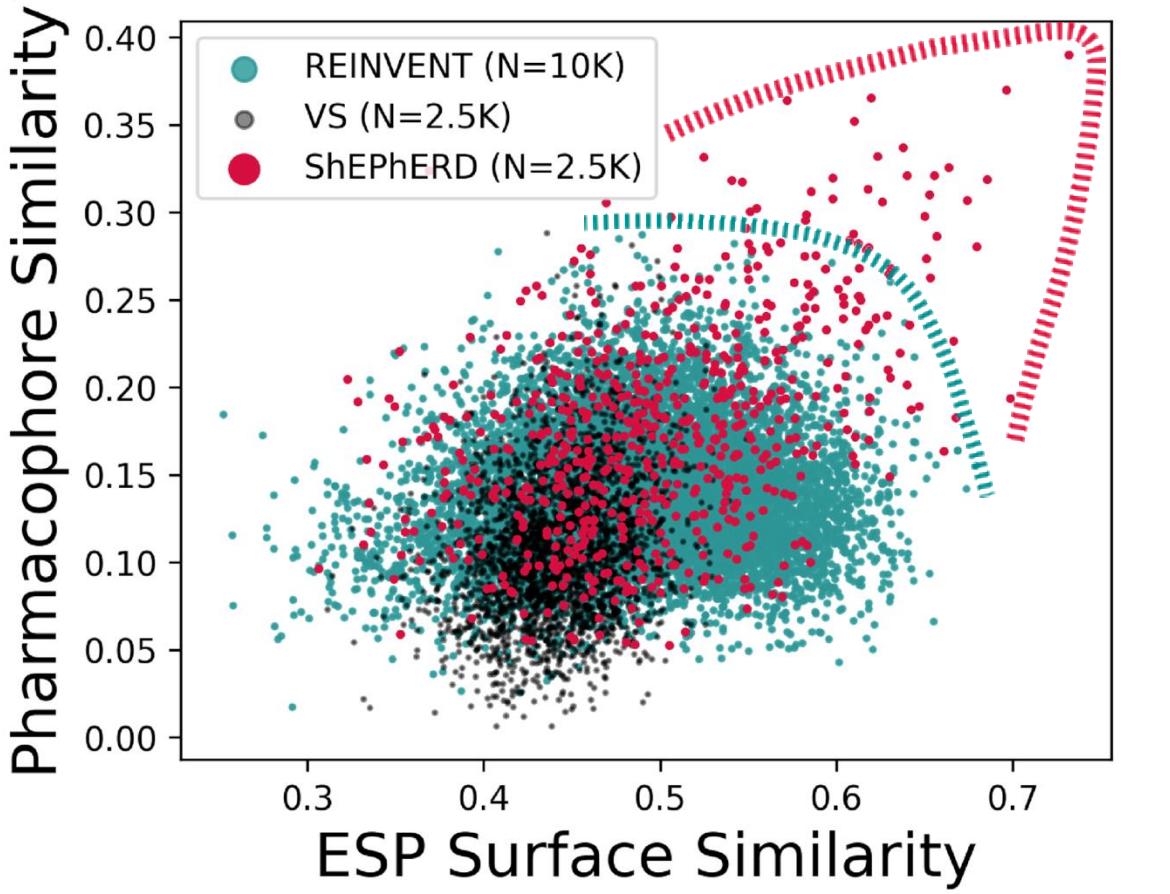


SA score: 4.2
ESP sim: 0.73
Pharm. sim: 0.39





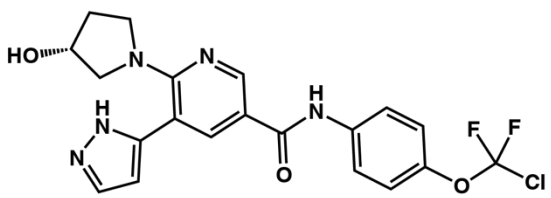
SA score: 3.7
ESP sim: 0.62
Pharm. sim: 0.37



ShEPhERD diversifies ligands + preserves their binding modes

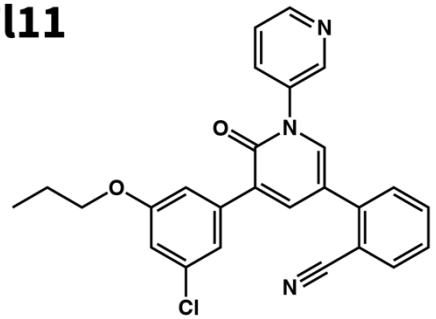
PDB Ligand

5mo4



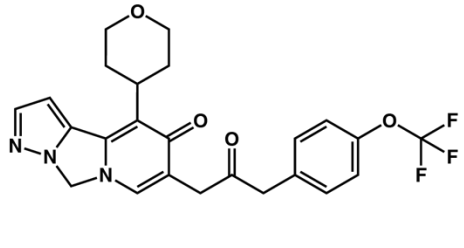
Vina: -9.9

7l11

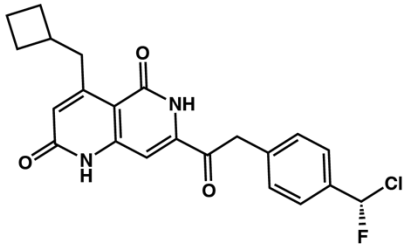


Vina: -9.1

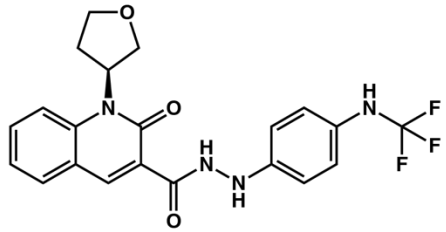
Top *ShEPhERD* analogues



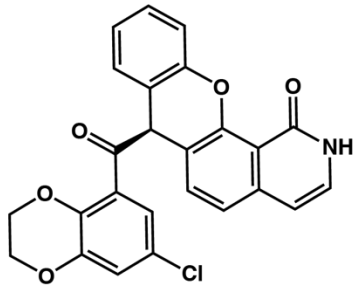
Vina: -10.3



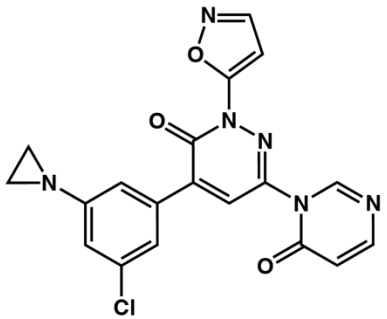
Vina: -10.1



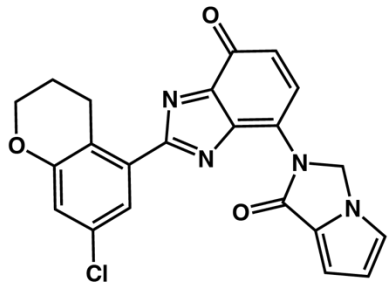
Vina: -10.1



Vina: -9.6

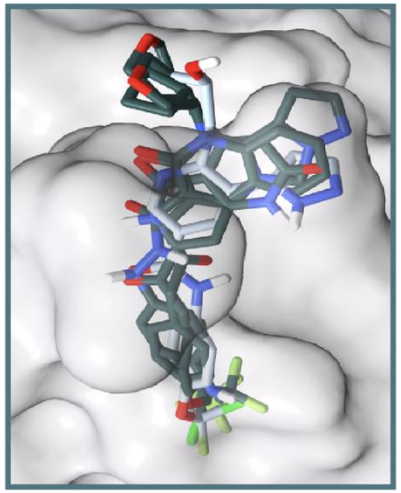


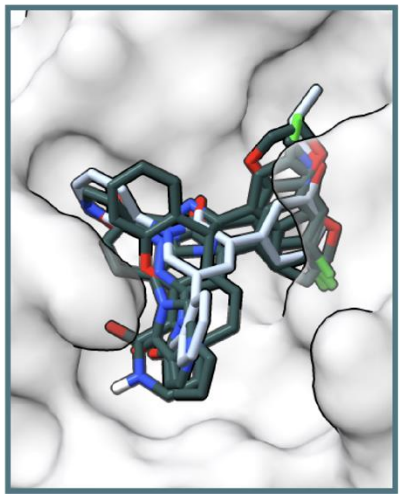
Vina: -9.3



Vina: -9.1

Overlaid docked poses



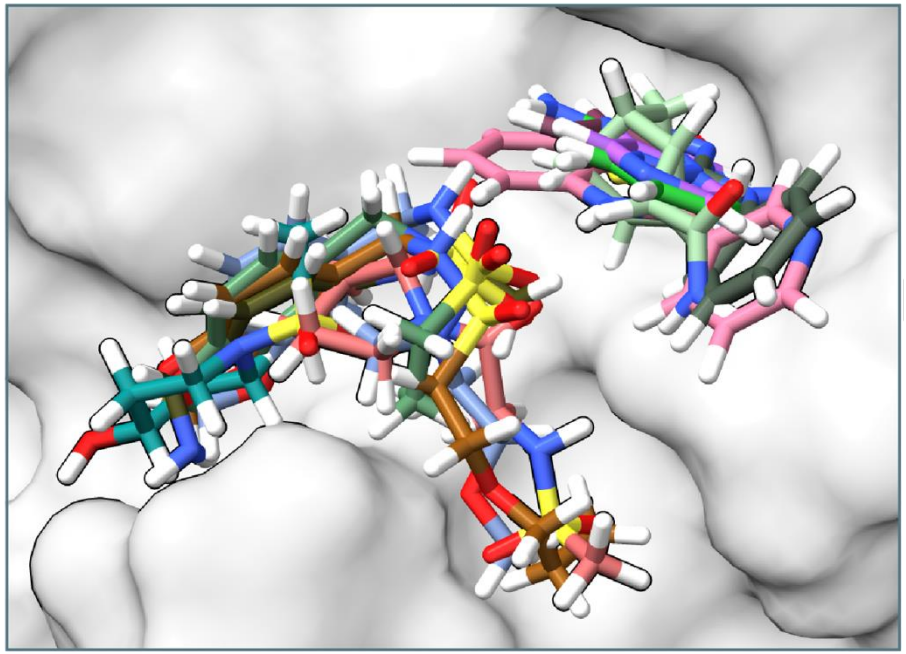


*note docking is used as a poor, *in silico* surrogate for binding affinity

ShEPhERD enables bioisosteric fragment merging

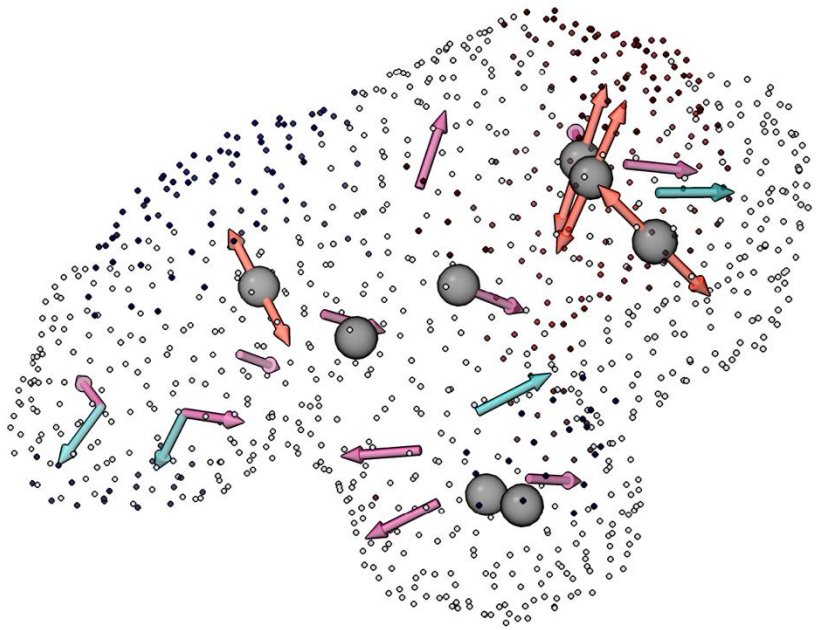
- **Bioisosteric fragment merging** seeks to merge fragments into a new ligand that *preserves the fragments' binding interactions,* without necessarily containing the exact fragments themselves

Fragment Screen
(13 overlapping fragments)



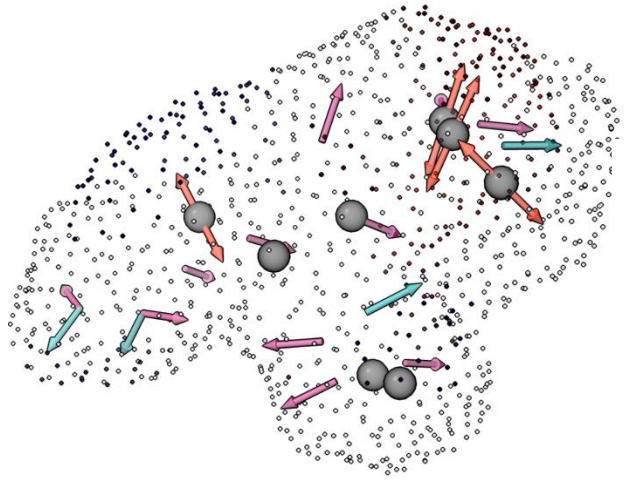
Extract interactions

Aggregate ESP surface and pharmacophore

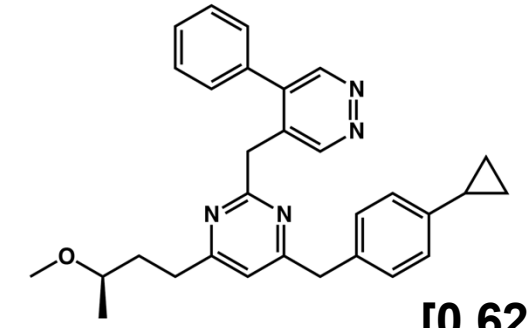


ShEPhERD enables bioisosteric fragment merging

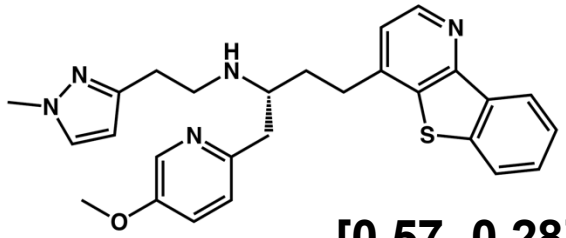
- **Bioisosteric fragment merging** seeks to merge fragments into a new ligand that *preserves the fragments' binding interactions,* without necessarily containing the exact fragments themselves



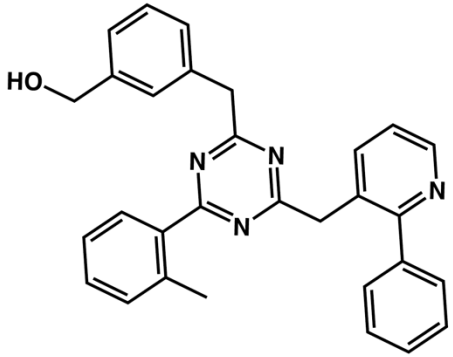
ShEPhERD



[0.62, 0.44]



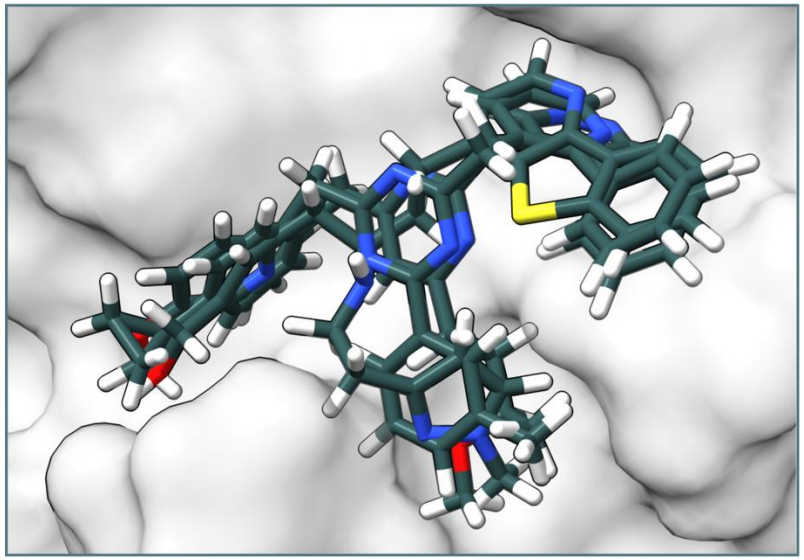
[0.57, 0.28]



[0.53, 0.30]

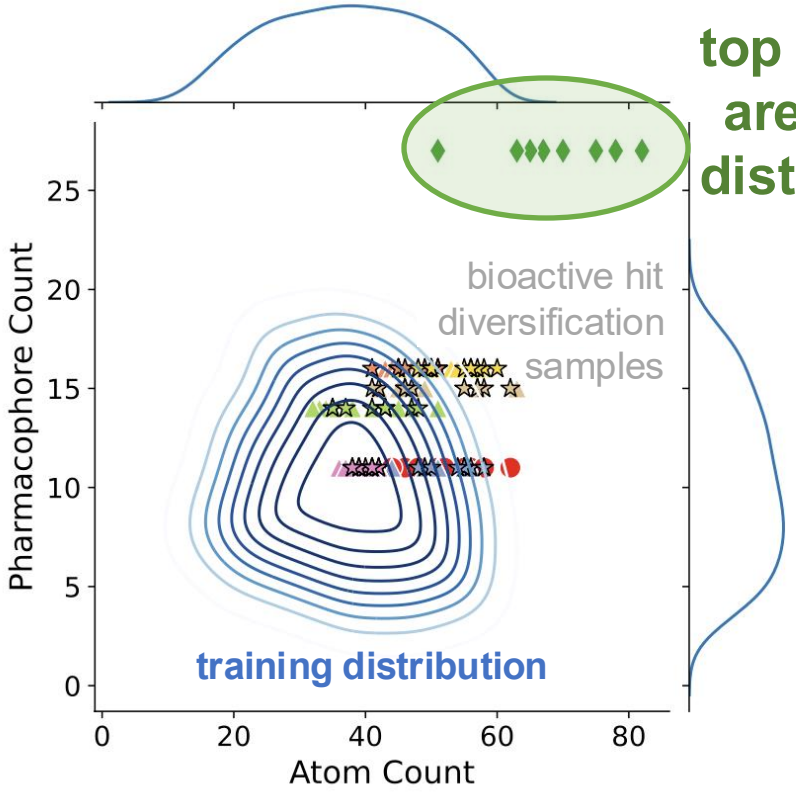
[ESP Sim (↑), Pharm. Sim (↑)]

Aligned poses (post-relaxation)

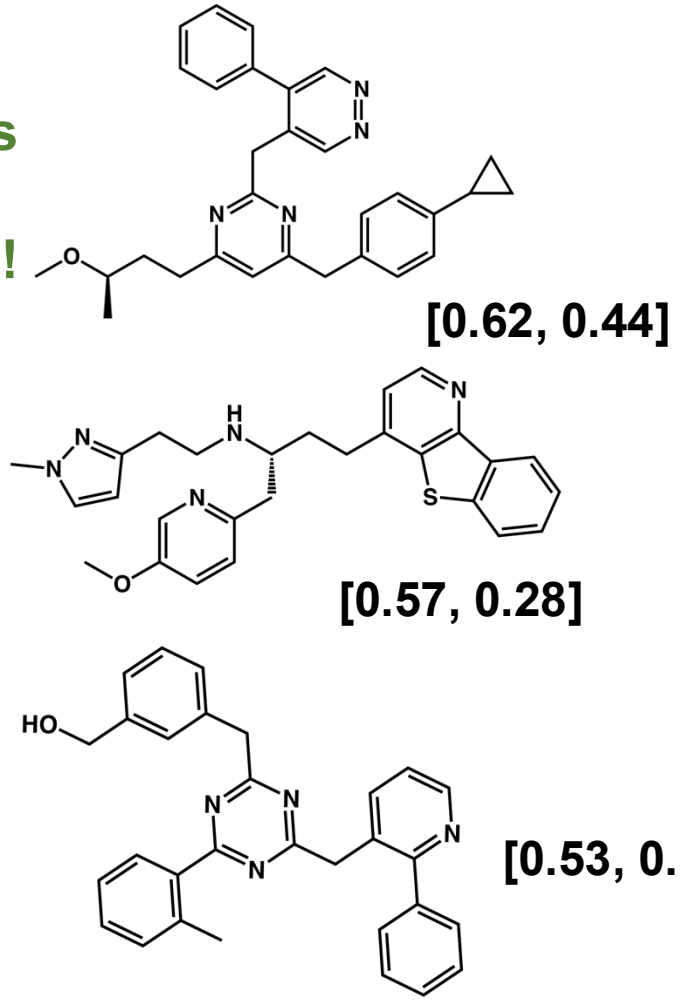


ShEPhERD enables bioisosteric fragment merging

- **Bioisosteric fragment merging** seeks to merge fragments into a new ligand that *preserves the fragments' binding interactions,* without necessarily containing the exact fragments themselves



top samples are out-of-distribution!

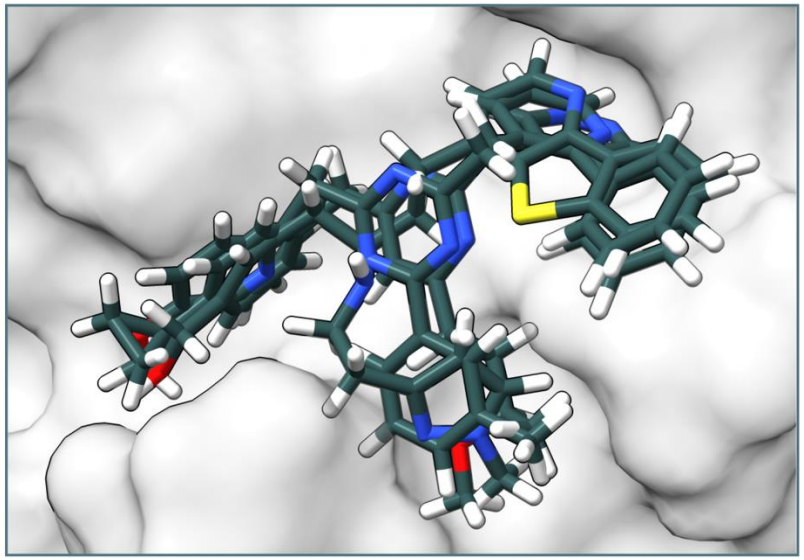


[0.62, 0.44]

[0.57, 0.28]

[0.53, 0.30]

Aligned poses (post-relaxation)

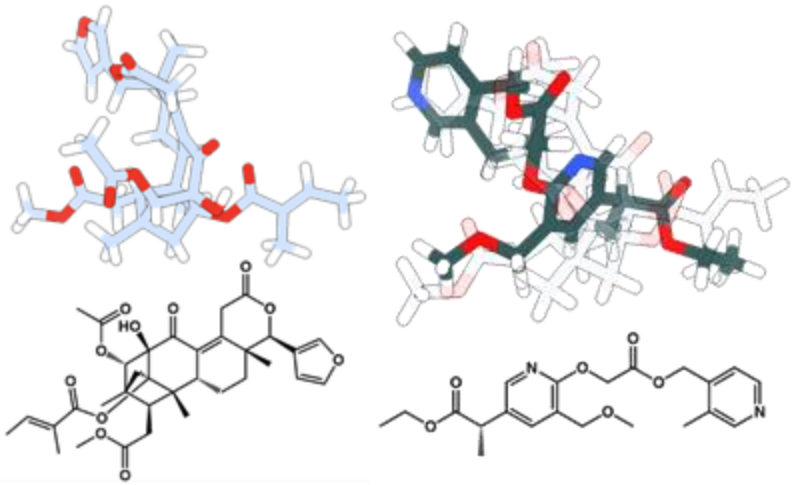


Outlook

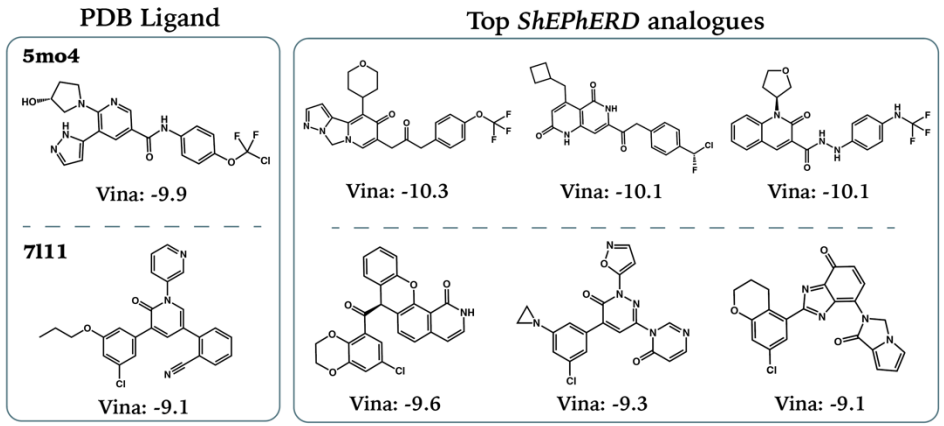
ShEPhERD shows promise for challenging tasks in 3D ligand-based drug design

- Can generate diverse molecules that maintain interactions

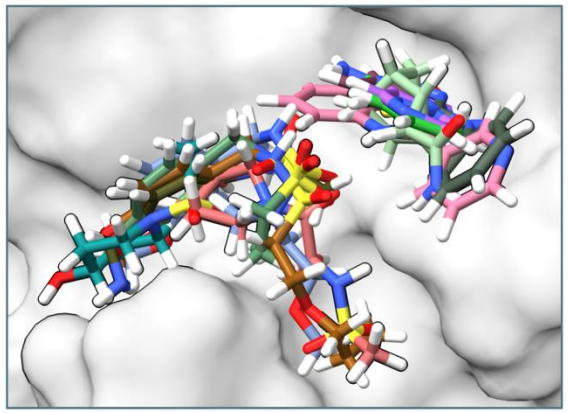
Ligand hopping



Hit diversification



Fragment merging



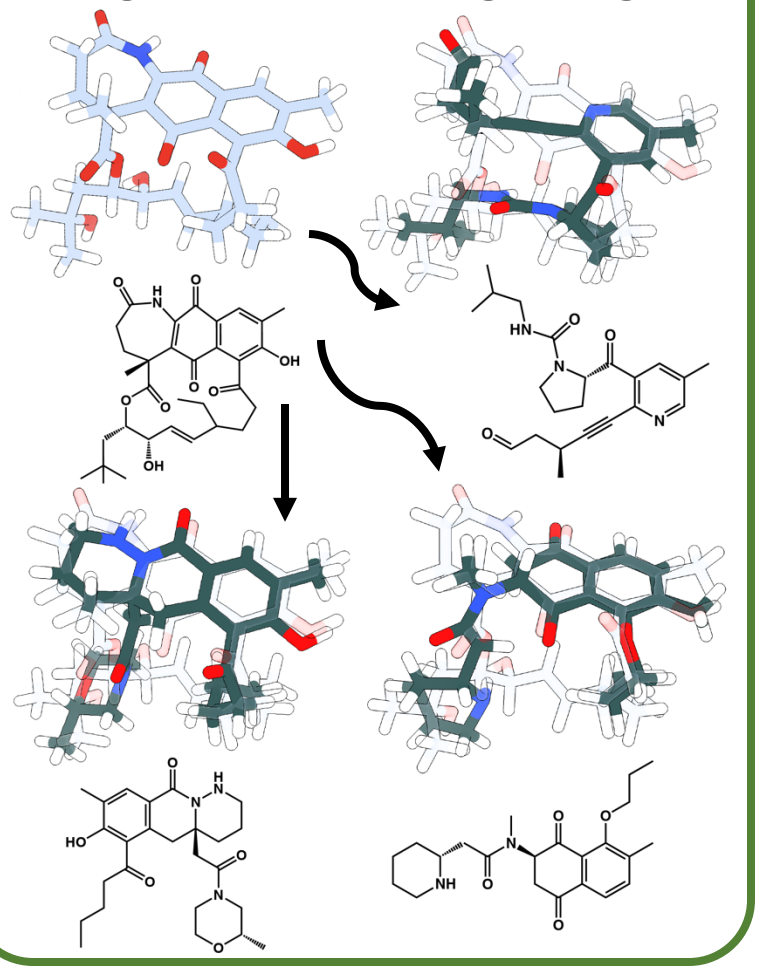
Future directions

- Inference-time optimization strategies
- Redesigning model to accelerate sampling (e.g., flow-matching vs. diffusion)
- Scaling up to larger drug-like datasets (ZINC, Enamine, ChEMBL, etc.)

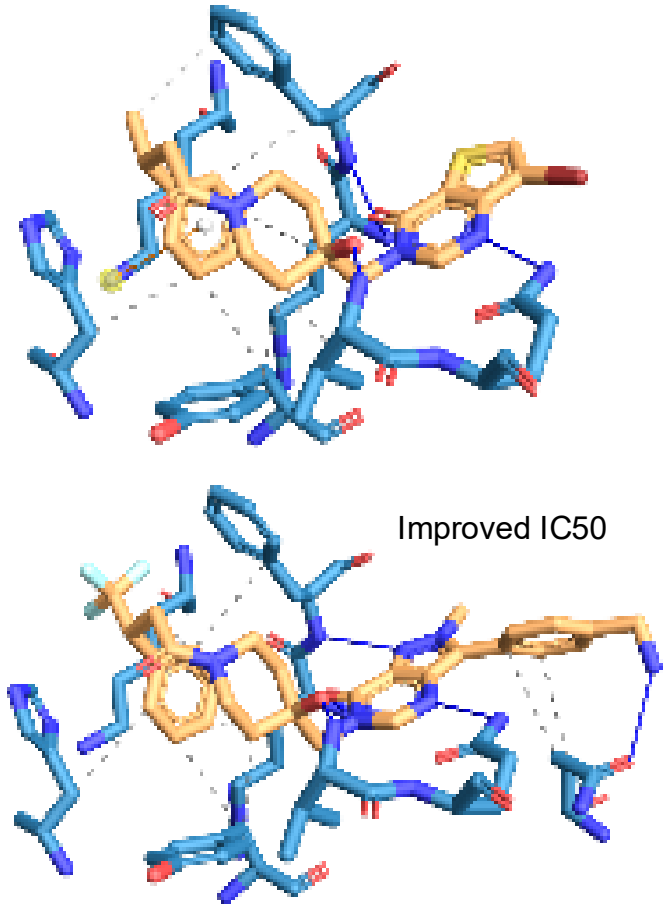
Outlook

ShEPhERD may also be extended to address interaction-dependent tasks beyond LBDD

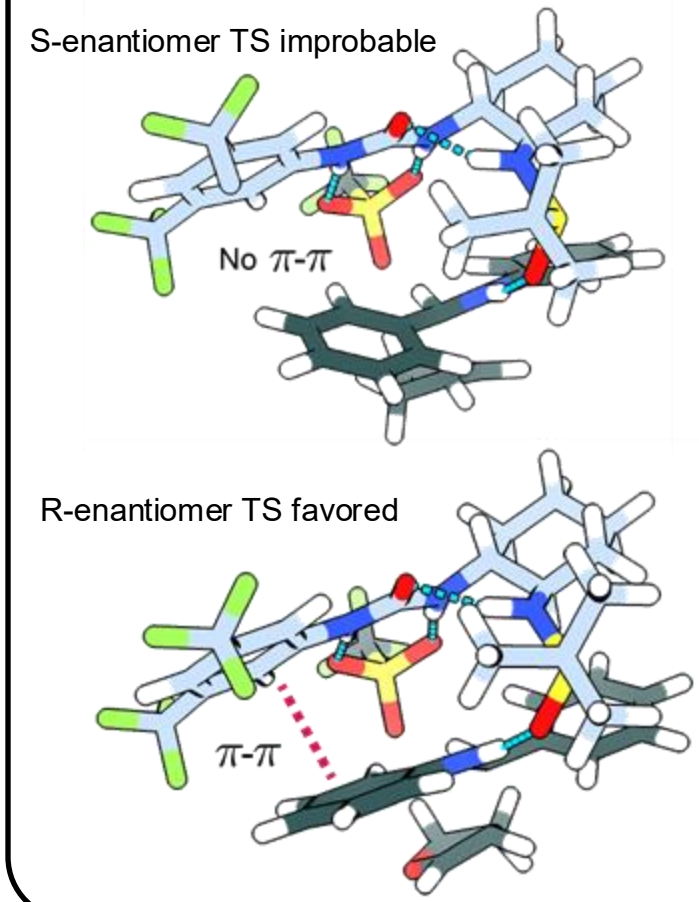
Ligand-based drug design



Structure-based drug design



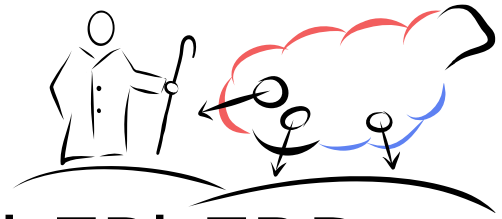
Organocatalyst design



** Focus of this work

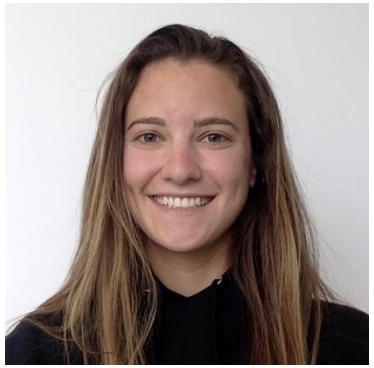
Gavory et al. *Nat. Chem. Bio.* (2017) 14, 118–125
Xu et al. *Science.* (2010) 327, 5968, 986-990

Acknowledgements



ShEPhERD team:

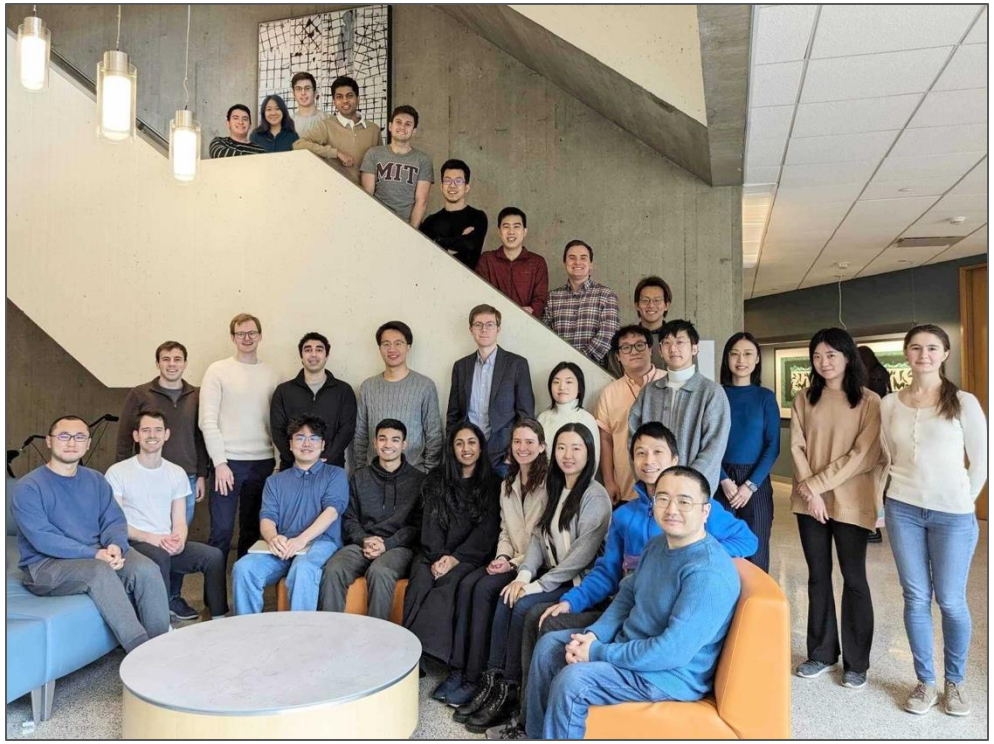
Keir Adams



Jenna Fromer

PI: Connor W. Coley

Coley Research Group:



GitHub

Funding and computational resources:

Open-source code, data, and weights found at: **github.com/coleygroup/shepherd**