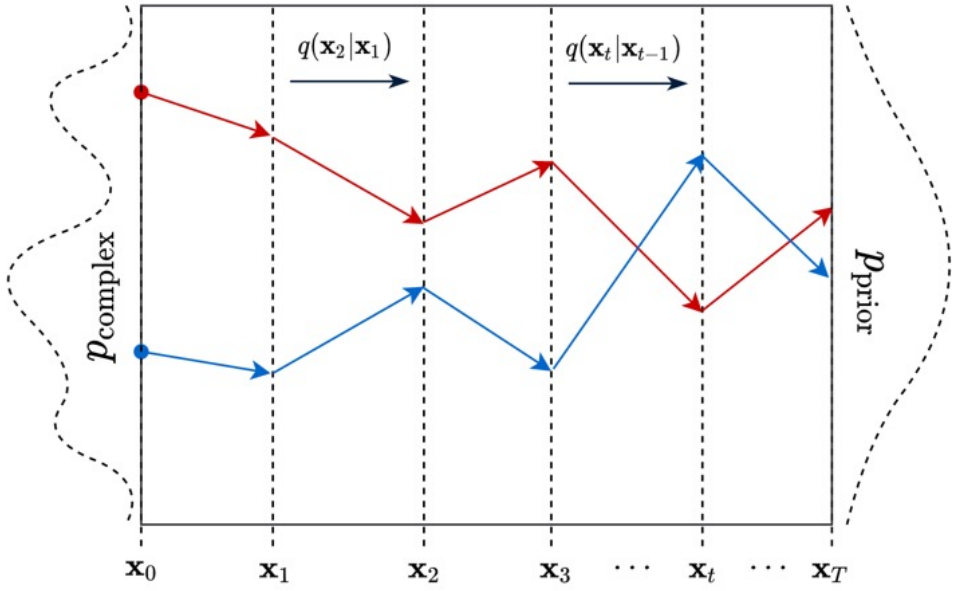
LATENT-CONDITIONED EQUIVARIANT DIFFUSION FOR STRUCTURE-BASED DE NOVO LIGAND GENERATION

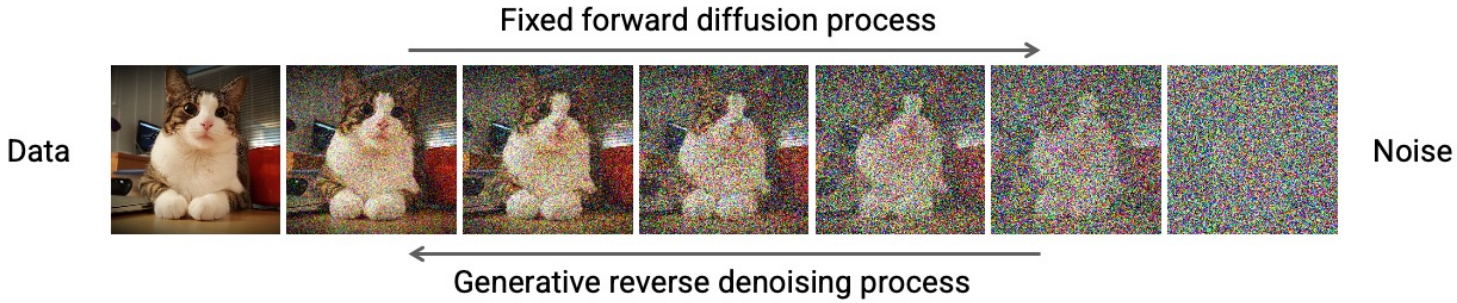
Tuan Le*, Julian Cremer*, Djork-Arné Clevert, Kristof T. Schütt

***Equal contribution**

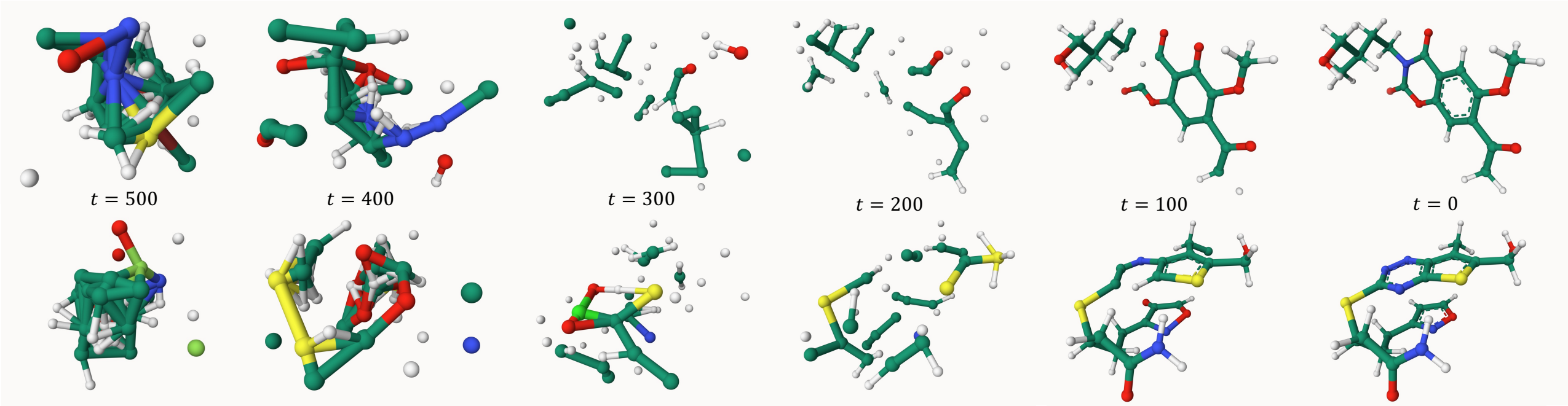
What is denoising diffusion?

Principles of denoising diffusion



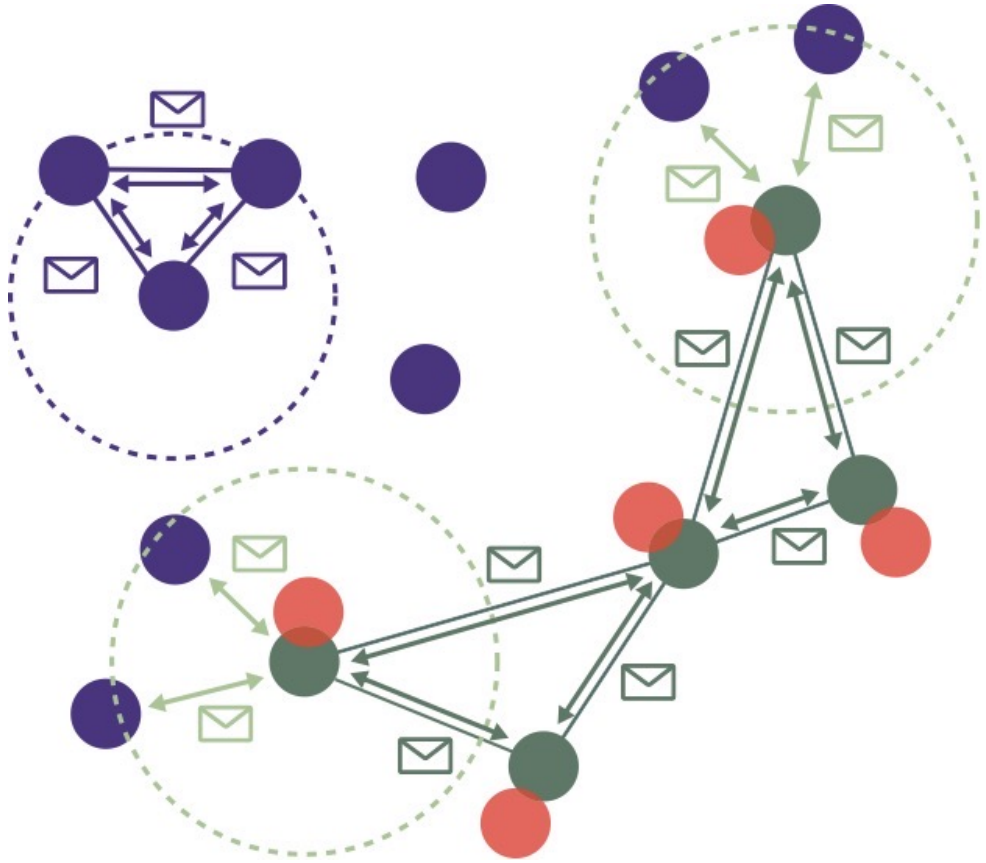


…for molecules



Denoising diffusion for structure-based drug discovery

The network: Equivariant GNN

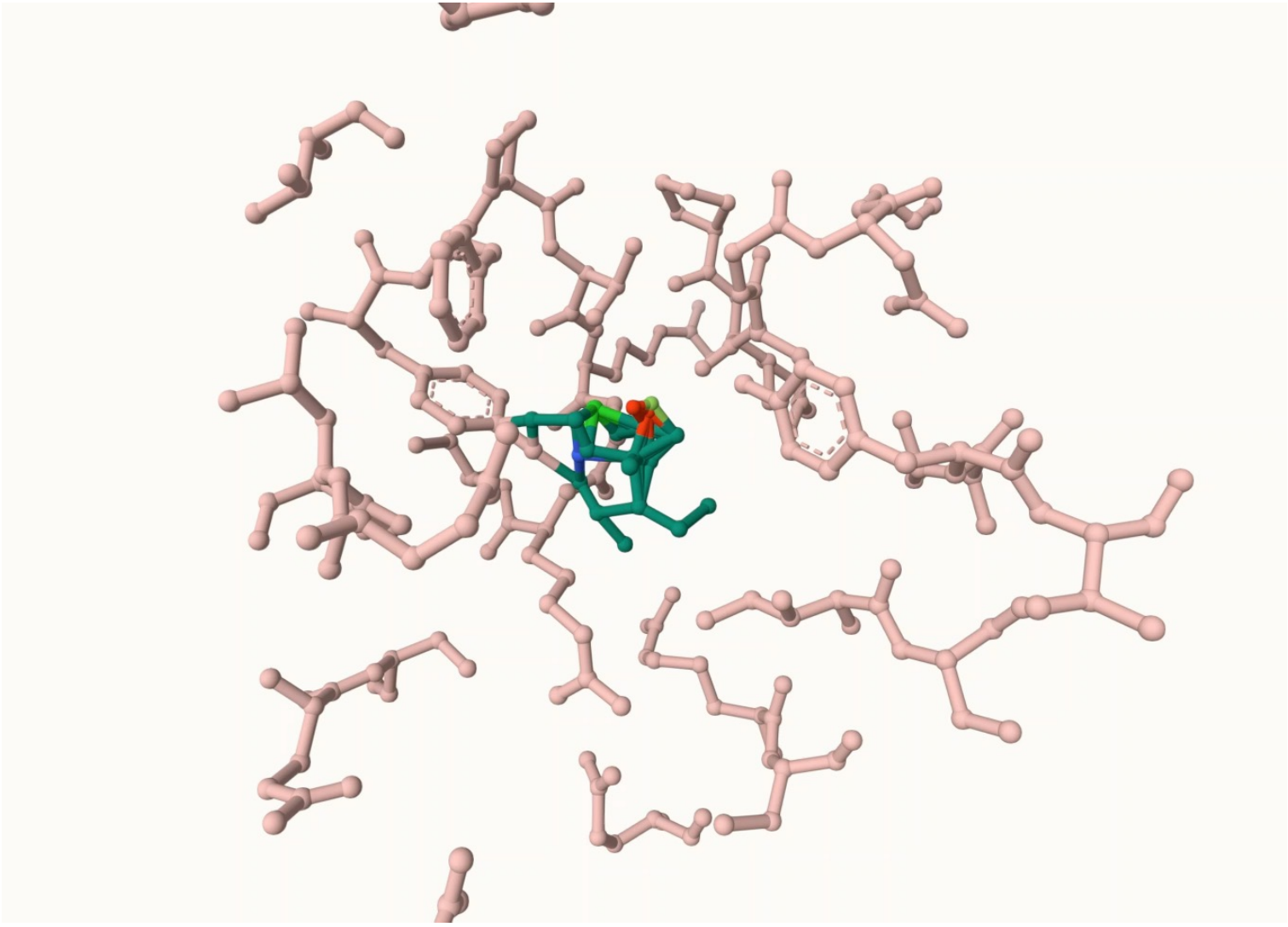


Ligand interaction
Ligand-pocket interaction
Pocket-pocket interaction
Ground truth ligand atoms
Noised ligand atoms
Pocket atoms

GROUND TRUTH
PREDICTION

ATOM COORDINATES
ATOM TYPES
ATOM CHARGES
BOND TYPES

Denoising diffusion for molecules with pocket condition

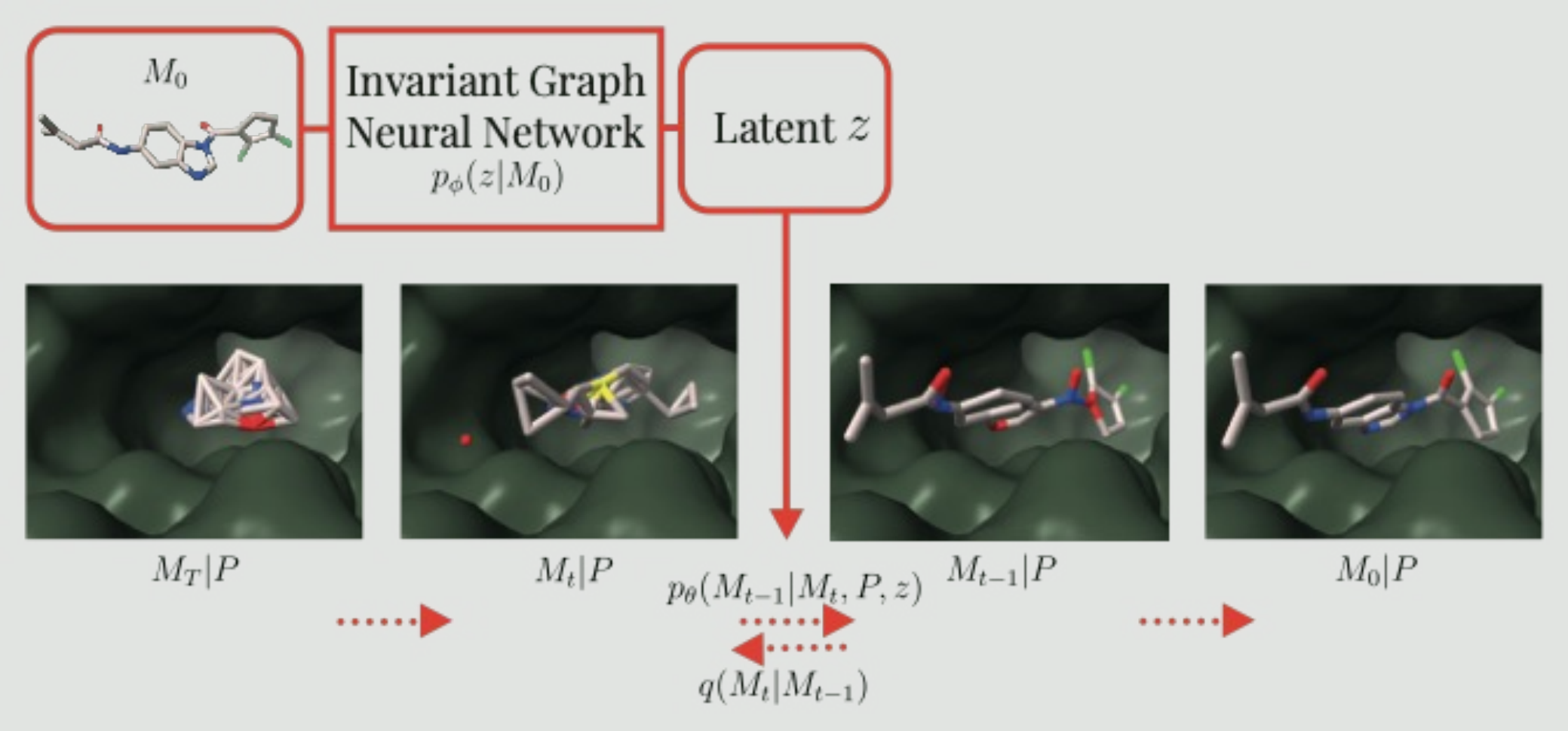


What are the problems with current generative *de novo* models?

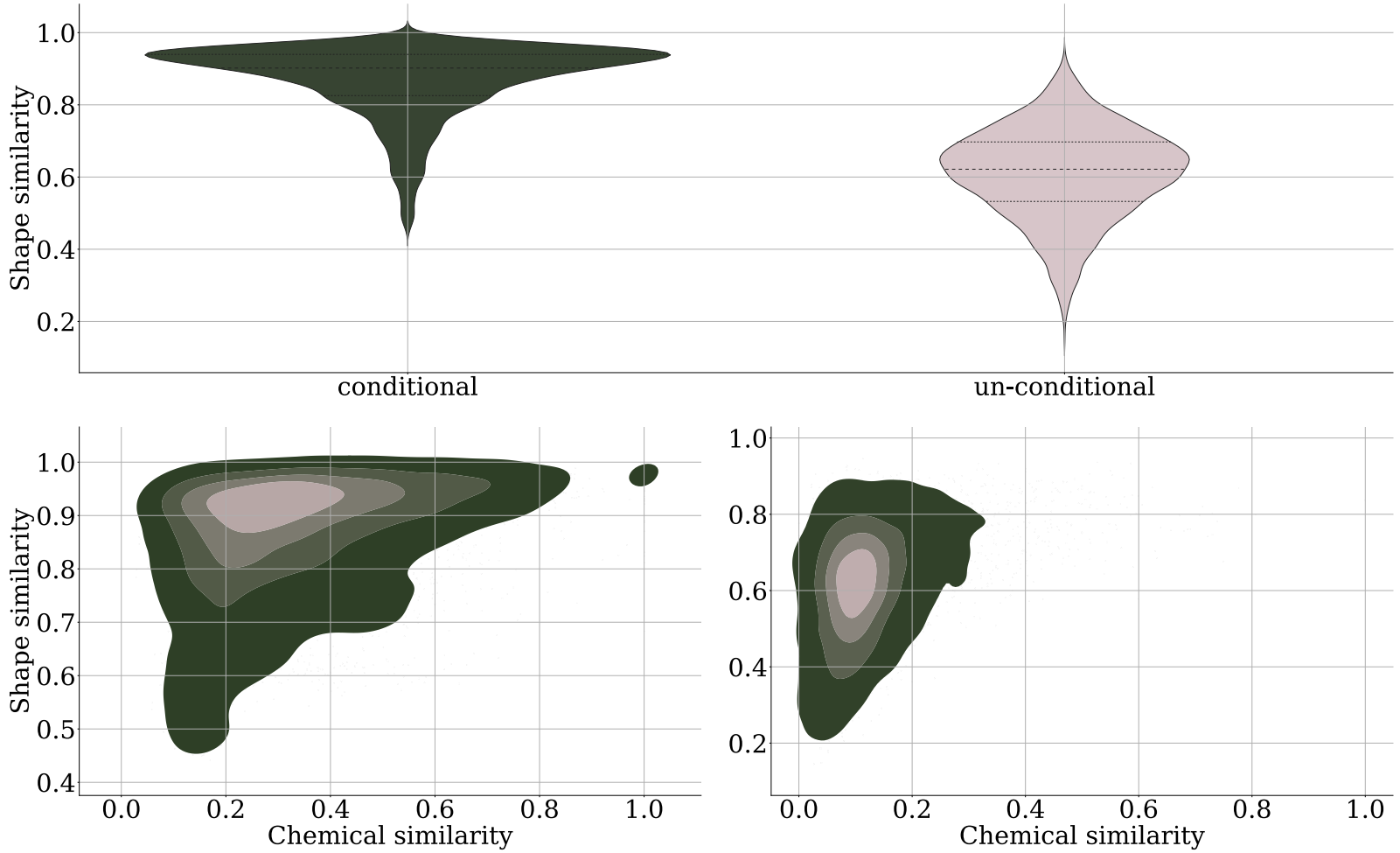
- In principle, current SOTA models show promising results
- However, the dataset limitations are striking
  - No efficient chemical space coverage possible
  - Many flaws in the dataset lead to bias propagation
  - Many drug discovery campaigns have very specific needs
  - Hence, ligand generation from scratch suboptimal

- How about constraining the generation in chemical and/or property space?
  - We came up with an easy-to-use latent-conditional approach (besides the pocket condition) to have better control over the generation process
  - We applied the approach to hit expansion
    - Chemical diversification of already existing hits without losing potential activity
      - E.g., preserve the shape of the hit molecule and diversify its chemical composition

The pipeline of PoLiGenX



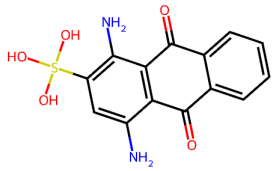
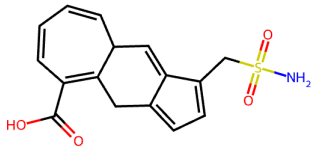
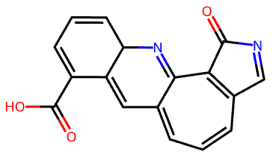
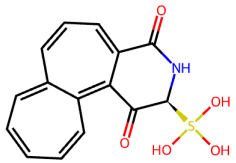
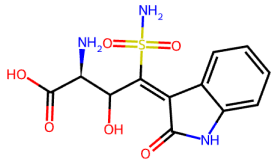
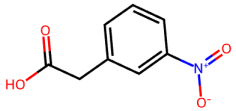
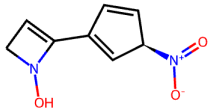
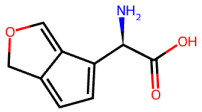
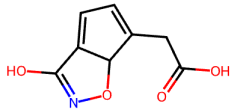
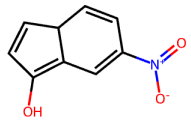
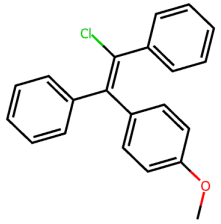
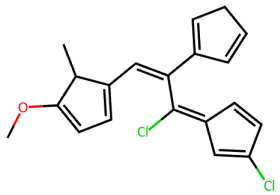
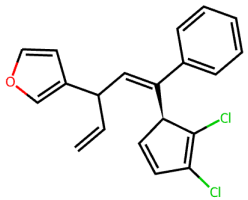
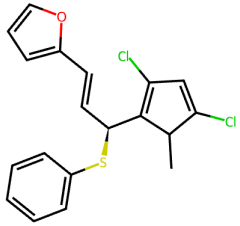
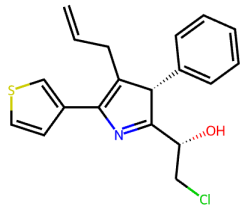
Model evaluation: Shape and chemical composition



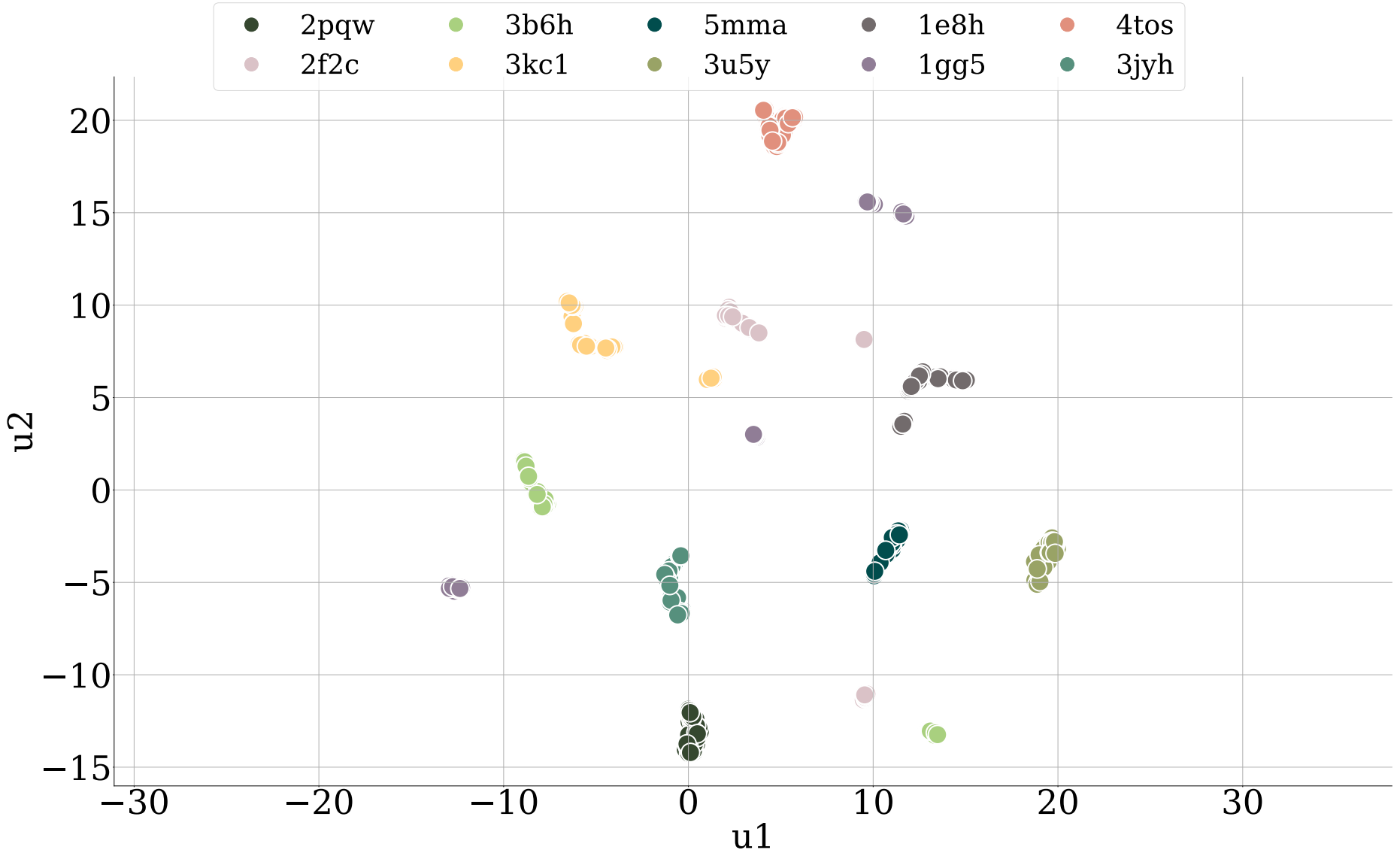
Model evaluation: Docking and druglikeness

Data	QVina2 (All) ↓	QVina2 (Top-10%) ↓	QED ↑	logP ↑	MolWt ↑	H-acceptors ↑	H-donors ↑	Lipinski ↑
CrossDocked test set	-6.85$_{\pm 2.33}$	-	0.47$_{\pm 0.20}$	0.79	0.85	0.84	0.8	3.35$_{\pm 1.14}$
PoLiGenX	**-7.21**$_{\pm 2.22}$	**-8.04**$_{\pm 2.44}$	**0.59**$_{\pm 0.20}$	**0.91**	**0.87**	**0.85**	**0.91**	**3.57**$_{\pm 0.93}$

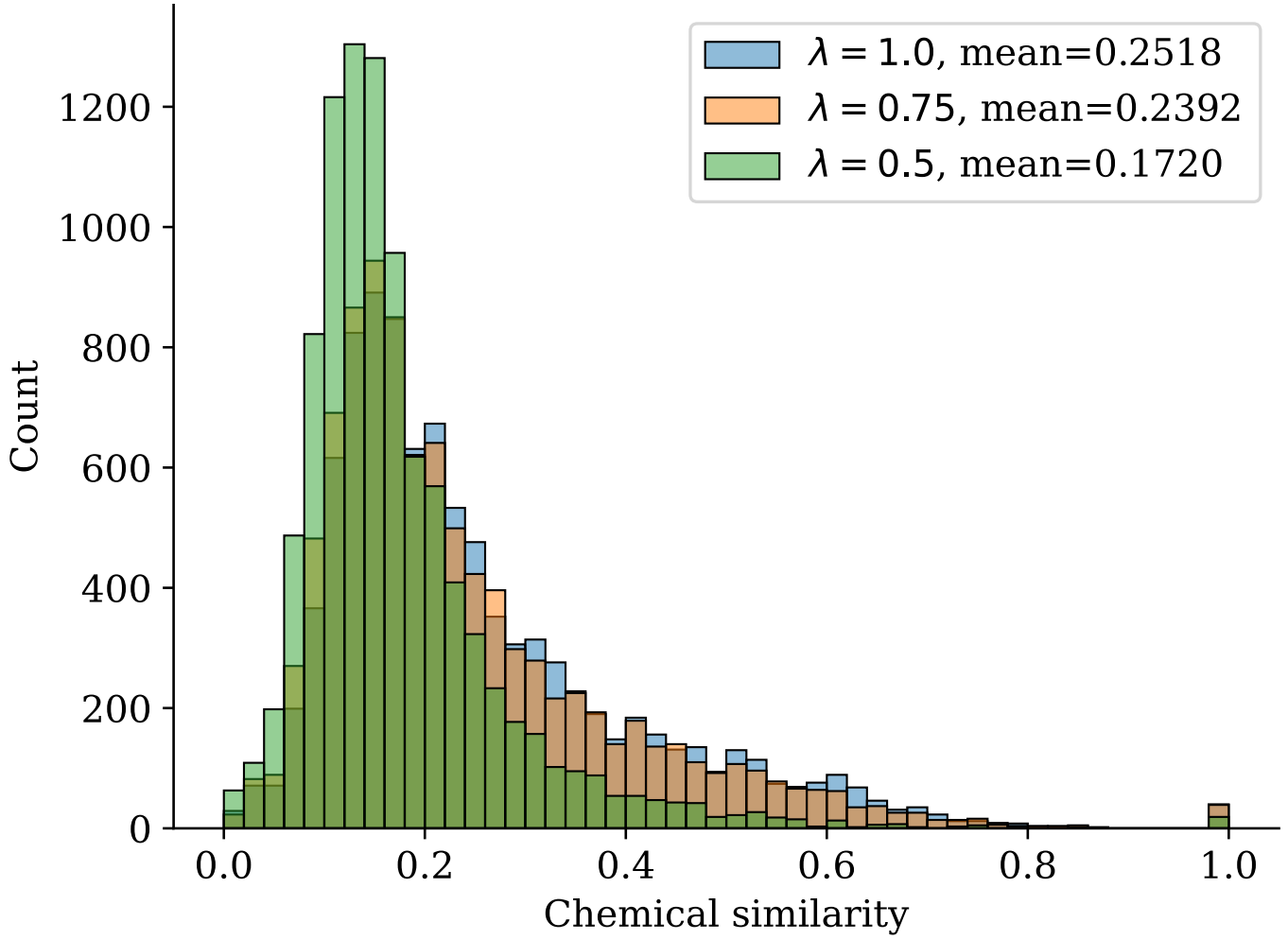
Examples

Reference	Generated			
	 0.101	 0.07	 0.115	 0.246
	 0.151	 0.132	 0.224	 0.191
	 0.138	 0.164	 0.129	 0.106

The expressiveness of the latents



Control over the latents



Summary

- Introduction of an easy and efficient way to incorporate a constraining mechanism into diffusion-based SBDD models via latent conditioning
- High shape similarity with reference ligands, while chemical diversification is guaranteed (no mode collaps etc.)
- Docking scores and druglikeness are favorable
- The latent model learns expressive embeddings (that could be also used downstream for other purposes)
- Flexible control over the latent strength

Thanks!

Questions?