PILOT: Equivariant Diffusion for Target-Aware De Novo Ligand Generation with Multi-Objetive Guidance via Importance Sampling

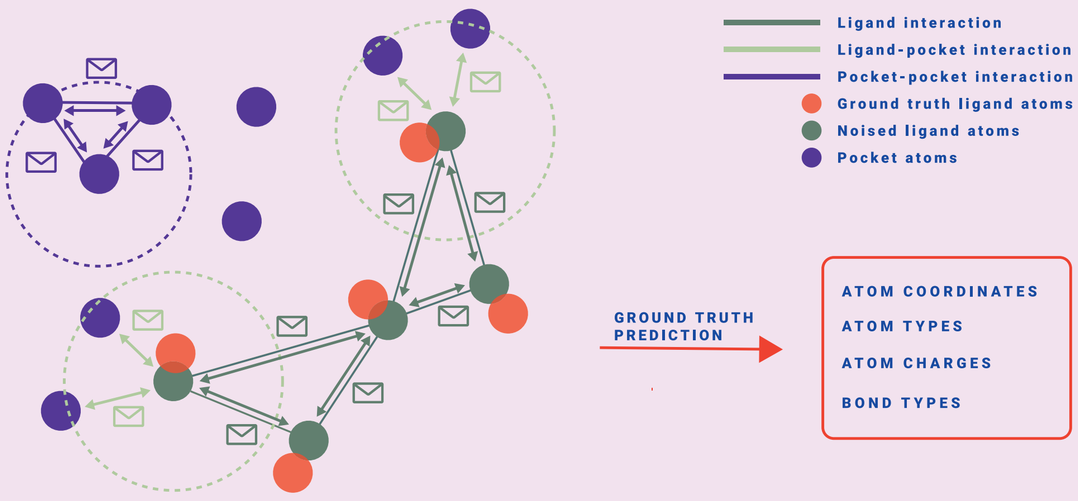
JULIAN CREMER*[1,3], TUAN LE*[1,2], FRANK NOÉ[2,4], DJORK-ARNÉ CLEVERT[1] AND KRISTOF T. SCHÜTT[1]

[1] Pfizer Research & Development
[2] Freie Universität Berlin
[3] Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB)
[4] Microsoft Research

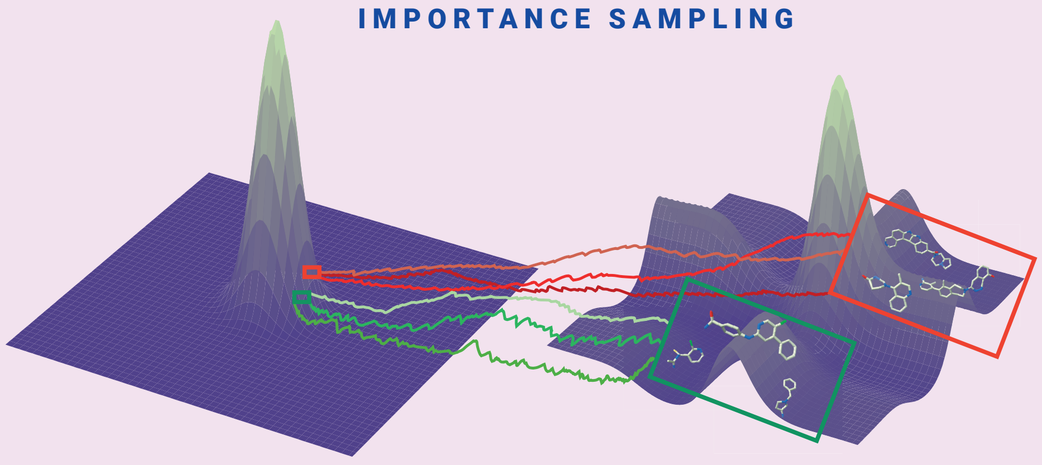
* Equal contribution

Overview

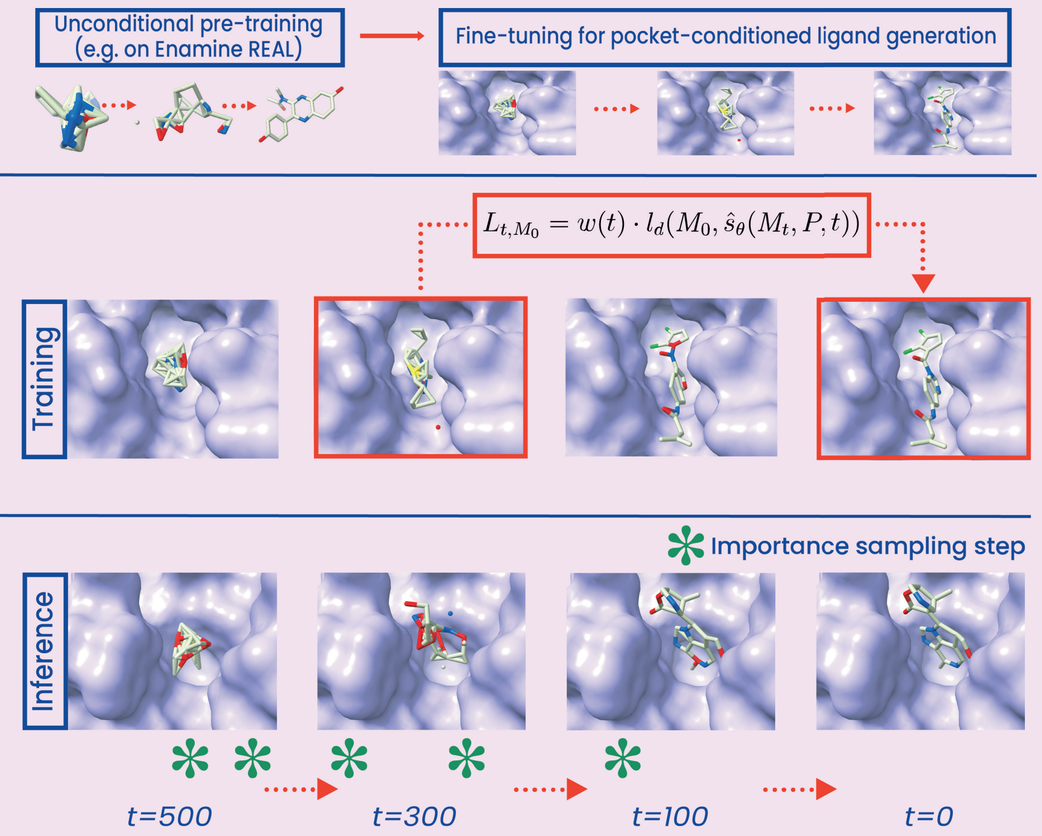
EQUIVARIANT TARGET-AWARE DIFFUSION



IMPORTANCE SAMPLING



TRAINING AND INFERENCE



MOTIVATION AND BACKGROUND

1. Prior works utilizing equivariant autoregressive or diffusion models for target-aware 3D ligand design demonstrated suboptimal performance regarding synthetic accessibility. Is there a way to alleviate this issue?
2. How can we design molecules that have optimized pharmacological properties specifically tailored to given protein targets?
3. Can we increase performance and generality by pre-training the diffusion model on large corpora of unconditional molecular data derived at a low level of theory? How does this transfer to target-conditioned drug design?

We introduce PILOT, a novel *de novo* generative model, denoted as $p_\theta(M|P,c)$, designed for generating 3D ligands represented by $x = (\mathbf{H}, \mathbf{X}, \mathbf{E})$, with $\mathbf{H} \in \{0,1\}^{N \times K_a}$, $\mathbf{E} \in \{0,1\}^{N \times N \times K_b}$, $\mathbf{X} \in \mathbb{R}^{N \times 3}$. This model processes both continuous and discrete variables and is conditioned on a specific protein pocket P. We propose trajectory-based importance sampling to optimize multiple objectives, including synthetic accessibility, docking score, and predicted half-maximal inhibitory concentration (IC_{50}), thereby enhancing the targeted design of ligands.

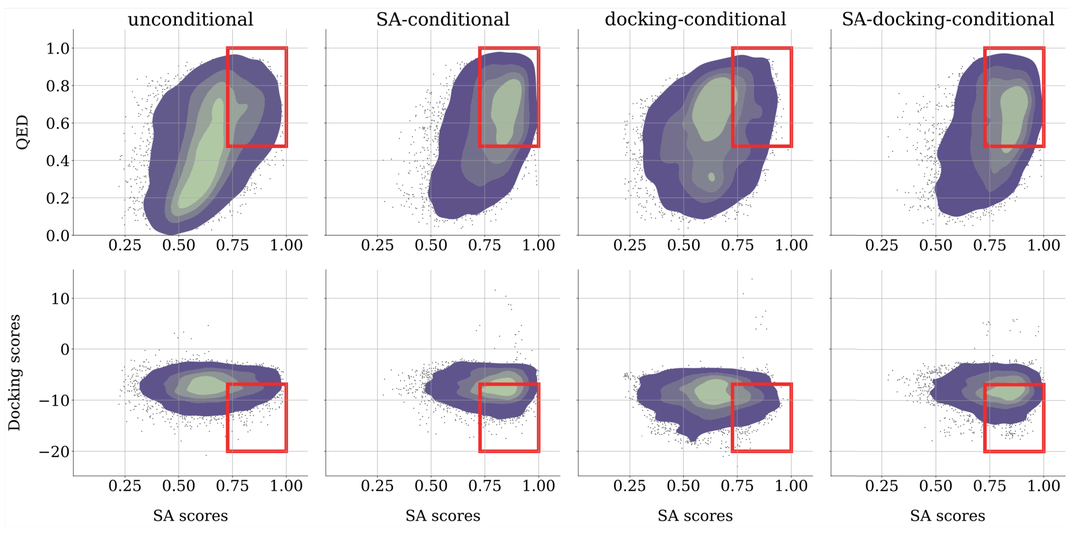
Experiments

THE EFFECT OF IMPORTANCE SAMPLING

Performance comparison among unconditional, SA-conditional, and SA-docking-conditional sampling using the CrossDocked2020 test set comprising 100 protein targets for 100 sampled ligands each. Evaluation based on mean docking score as proxy for binding affinity using QVina2 and other RDKit-based chemical properties like drug-likeness (QED) and synthetic accessibility (SA)

Model	QVina2 (All) ↓	QVina2 (Top-10%) ↓	QED ↑	SA ↑	Lipinski ↑	Diversity ↑
Training set	-7.57$_{\pm 2.09}$	-	0.53$_{\pm 0.20}$	0.75$_{\pm 0.10}$	4.57$_{\pm 0.91}$	-
Test set	-6.88$_{\pm 2.33}$	-	0.47$_{\pm 0.20}$	0.72$_{\pm 0.13}$	4.34$_{\pm 1.14}$	-
TargetDiff	-7.318$_{\pm 2.47}$	**-9.669**$_{\pm 2.55}$	0.483$_{\pm 0.20}$	0.584$_{\pm 0.13}$	4.594$_{\pm 0.83}$	0.718$_{\pm 0.09}$
DiffSBDD-cond	-6.950$_{\pm 2.06}$	-9.120$_{\pm 2.16}$	0.469$_{\pm 0.21}$	0.578$_{\pm 0.13}$	4.562$_{\pm 0.89}$	0.728$_{\pm 0.07}$
un-conditional	-7.33$_{\pm 2.19}$	-9.28$_{\pm 2.26}$	0.49$_{\pm 0.22}$	0.64$_{\pm 0.13}$	4.40$_{\pm 1.05}$	0.69$_{\pm 0.07}$
SA-conditional	-7.32$_{\pm 2.25}$	-8.91$_{\pm 2.29}$	**0.58**$_{\pm 0.19}$	**0.77**$_{\pm 0.10}$	**4.82**$_{\pm 0.54}$	**0.73**$_{\pm 0.08}$
SA-docking-conditional	**-8.35**$_{\pm 2.75}$	**-10.36**$_{\pm 2.62}$	**0.58**$_{\pm 0.17}$	0.72$_{\pm 0.12}$	**4.88**$_{\pm 0.44}$	0.68$_{\pm 0.09}$
SA-docking-conditional (norm)	-7.92$_{\pm 2.44}$	-9.85$_{\pm 2.33}$	0.56$_{\pm 0.19}$	**0.78**$_{\pm 0.11}$	**4.84**$_{\pm 0.47}$	0.75$_{\pm 0.13}$

Scatter plots with Gaussian kernel density estimation illustrating the evolution of QED, SA and docking scores for all sampled ligands across test targets for unconditional, SA-conditional and docking-conditional samplingRed rectangles within these plots highlight regions where sampled ligands demonstrate superior QED, SA, and docking scores compared to the test set



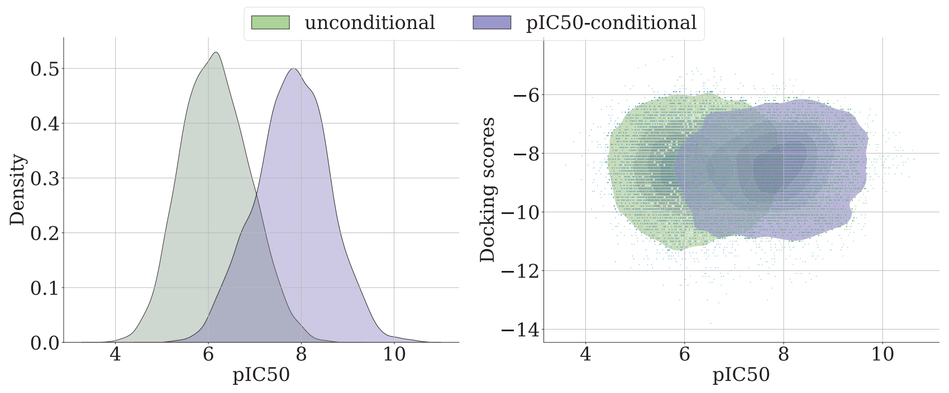
POTENCY OPTIMIZATION

Performance comparison among unconditional and pIC50-conditional sampling using the Kinodata3D test set comprising 10 kinase targets for 100 sampled ligands each. Evaluation based on mean docking score as proxy for binding affinity using QVina2, predicted pIC50 and other RDKit-based chemical properties like drug-likeness (QED) and synthetic accessibility (SA)

Model	Vina (All) ↓	Vina (Top-10%) ↓	pIC50 ↑	QED ↑	SA ↑	Lipinski ↑	Diversity ↑
Training set	-9.20$_{\pm 1.13}$	-	7.05$_{\pm 1.28}$	0.49$_{\pm 0.16}$	0.75$_{\pm 0.07}$	4.73$_{\pm 0.52}$	-
Test set	-8.78$_{\pm 1.13}$	-	6.41$_{\pm 1.56}$	0.61$_{\pm 0.14}$	0.79$_{\pm 0.05}$	**4.96**$_{\pm 0.22}$	-
unconditional	-8.49$_{\pm 1.05}$	-9.79$_{\pm 0.87}$	6.28$_{\pm 0.68}$	**0.63**$_{\pm 0.14}$	**0.75**$_{\pm 0.13}$	**4.95**$_{\pm 0.25}$	**0.65**$_{\pm 0.06}$
pIC50-conditional	-8.60$_{\pm 0.98}$	-9.75$_{\pm 0.86}$	**7.65**$_{\pm 0.78}$	**0.62**$_{\pm 0.16}$	0.67$_{\pm 0.09}$	**4.94**$_{\pm 0.28}$	0.57$_{\pm 0.06}$

left: Density plot comparing unconditional with pIC50-conditional sampling

right: Scatter heatmap overlap of unconditional and pIC50-conditional sampling comparing docking scores and predicted pIC50 values



CDK2 - EXAMPLES