TAGMOL: Target-Aware Gradient-guided Molecule Generation

MoleculeAI

Vineeth Dorna* D. Subhalingam* Keshav Kolluru Shreshth Tuli
Mrityunjay Singh Saurabh Singal N. M. Anoop Krishnan Sayan Ranu

Indian Institute of Technology Delhi

CONTRIBUTIONS

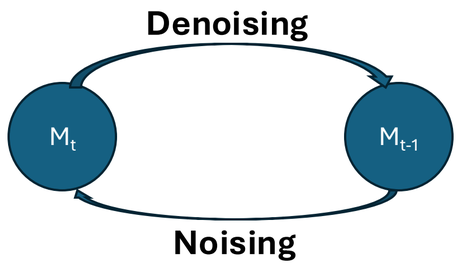
Goal
Generate ligand molecules that tightly bind to the protein pocket and possess acceptable pharmacological properties.

Algorithm Design
We introduce TAGMOL, decoupling molecular generation and property prediction. The latter guides diffusion sampling and create molecules with desired properties.

Rigorous empirical evaluation
TAGMOL outperforms SOTA by *22%* in *Vina Score*, all the while being guided by binding affinity and crucial pharmacological properties such as *QED* & *SA*.

PRELIMINARY

- **Protein:** $\mathbf{P} = [\mathbf{X}^{\mathcal{P}}, \mathbf{V}^{\mathcal{P}}]$
- **Ligand:** $\mathbf{M} = [\mathbf{X}^{\mathcal{M}}, \mathbf{V}^{\mathcal{M}}]$
- **Number of atoms:** $N_{\mathcal{P}}$ / $N_{\mathcal{M}}$
- **Number of features:** N_f / N_K
- **3D Coordinates:** $\mathbf{X}^{\mathcal{P}} \in \mathbb{R}^{N_{\mathcal{P}} \times 3}$ / $\mathbf{X}^{\mathcal{M}} \in \mathbb{R}^{N_{\mathcal{M}} \times 3}$
- **Atom Features:** $\mathbf{V}^{\mathcal{P}} \in \mathbb{R}^{N_{\mathcal{P}} \times N_f}$ / $\mathbf{V}^{\mathcal{M}} \in \mathbb{R}^{N_{\mathcal{M}} \times N_K}$
- **Set of guided properties:** $\mathbb{Y}$
- **Property Guide:** *SE(3)* Invariant GNN ϕ_y to predict property $\mathbf{y} \in \mathbb{Y}$
- **Generative backbone:** Diffusion model θ
- $\mathbf{M_t}$: Noisy molecule $\mathbf{M}$ at diffusion time step t



ALGORITHM

Guide Training Objective

$$NLL = -\mathbb{E}_{p(\mathbf{P},\mathbf{M}_{0:T})} \sum_{t=0}^{T} \log(p_{\phi_y}(\mathbf{y}|\mathbf{M}_t, \mathbf{P}, t))$$

$$= \mathbb{E}_{p(\mathbf{P},\mathbf{M}_{0:T})} \sum_{t=0}^{T} \frac{(\mathbf{y} - \phi_{\mathbf{y}}(\mathbf{M}_t, \mathbf{P}, t))^2}{2}$$

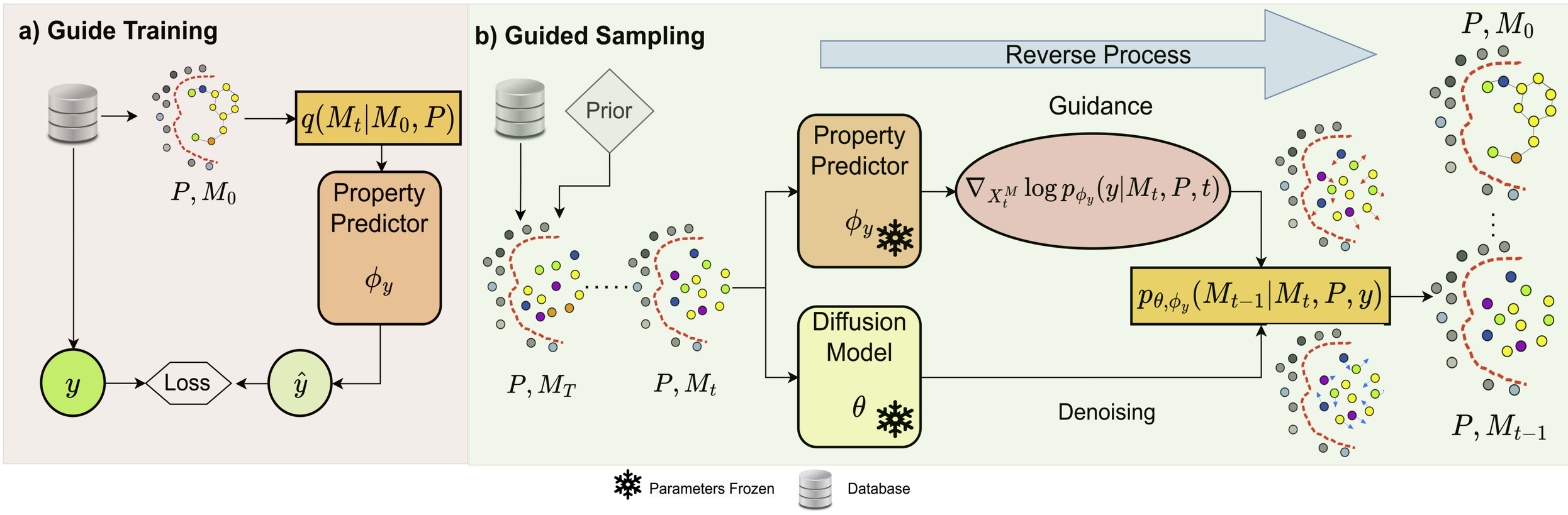
Guided Sampling

$$p_{\theta,\phi_{\mathbb{Y}}}(\mathcal{M}_{t-1}|\mathcal{M}_t, \mathcal{P}, \mathbb{Y}) = Z p_\theta(\mathcal{M}_{t-1}|\mathcal{M}_t, \mathcal{P}) \cdot \prod_{\mathbf{y}\in\mathbb{Y}} p_{\phi_{\mathbf{y}}}(\mathbf{y}|\mathcal{M}_{t-1}, \mathcal{P}, t)$$

$$\mathbf{X}^{\mathcal{M}}_{t-1} \sim \mathcal{N}(\tilde{\mu}_\theta(\mathcal{M}_t, \mathcal{P}, t) + \delta, \tilde{\beta}\mathbf{I})$$

$$\mathbf{V}^{\mathcal{M}}_{t-1} \sim \mathcal{C}\left(\tilde{c}_\theta\left(\mathcal{M}_t, \mathcal{P}, t\right)\right)$$

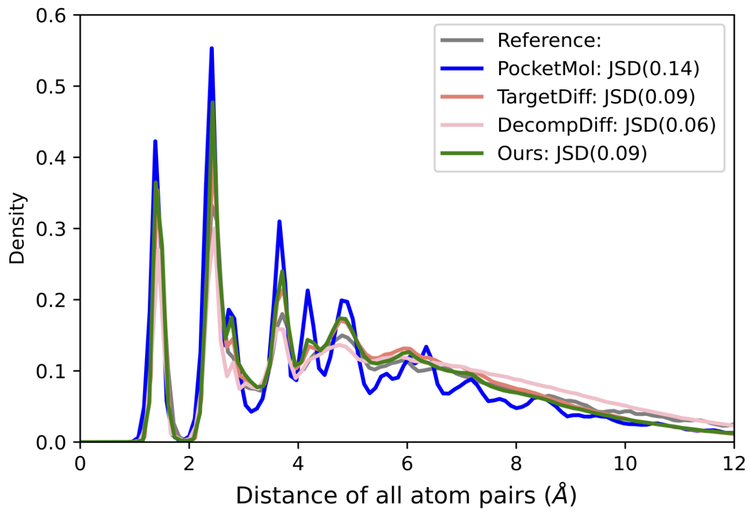
$$\delta = \sum_{\mathbf{y}\in\mathbb{Y}} s_{\mathbf{y}} \tilde{\beta} \nabla_{\mathbf{X}^{\mathcal{M}}_t} \log p_{\phi_{\mathbf{y}}}(\mathbf{y}|\mathcal{M}_t, \mathcal{P}, t)$$

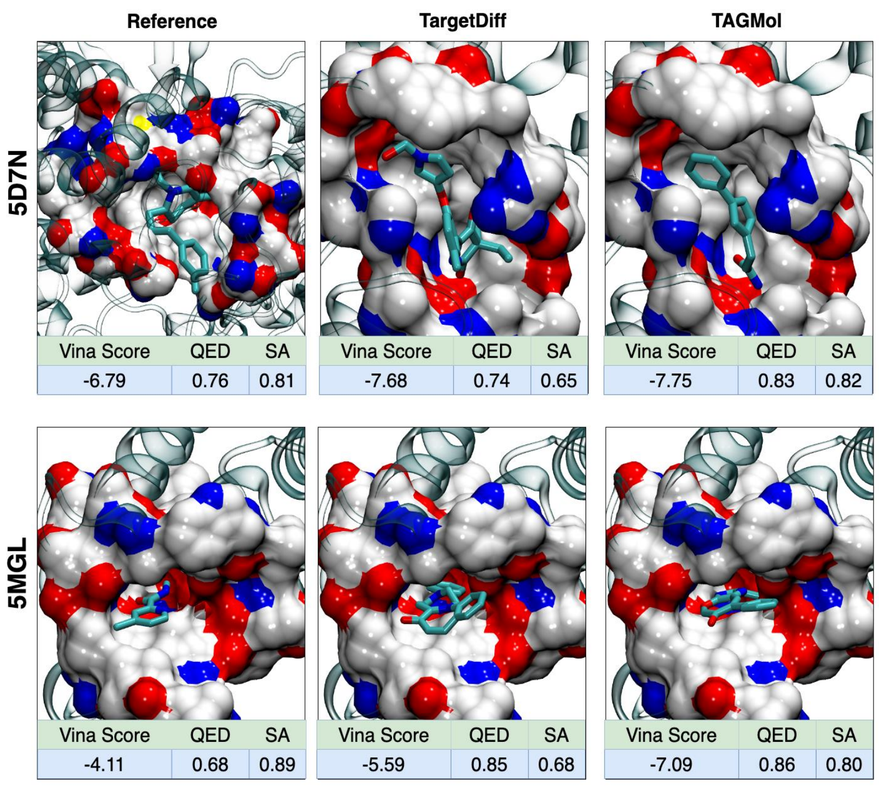


EXPERIMENTS & RESULTS

Methods	Vina Score (↓) Avg.	Vina Score (↓) Med.	Vina Min (↓) Avg.	Vina Min (↓) Med.	Vina Dock (↓) Avg.	Vina Dock (↓) Med.	High Affinity (↑) Avg.	High Affinity (↑) Med.	QED (↑) Avg.	QED (↑) Med.	SA (↑) Avg.	SA (↑) Med.	Hit(↑) Rate %
Reference	-6.36	-6.46	-6.71	-6.49	-7.45	-7.26	-	-	0.48	0.47	0.73	0.74	21
liGAN	-	-	-	-	-6.33	-6.20	21.1%	11.1%	0.39	0.39	0.59	0.57	13.2
AR	<u>-5.75</u>	-5.64	-6.18	-5.88	-6.75	-6.62	37.9%	31.0%	0.51	0.50	<u>0.63</u>	<u>0.63</u>	12.9
Pocket2Mol	-5.14	-4.70	-6.42	-5.82	-7.15	-6.79	48.4%	51.0%	**0.56**	**0.57**	**0.74**	**0.75**	24.3
TargetDiff	-5.47	<u>-6.30</u>	-6.64	-6.83	-7.80	-7.91	58.1%	59.1%	0.48	0.48	0.58	0.58	20.5
DecompDiff	-4.85	-6.03	-6.76	<u>-7.09</u>	<u>-8.48</u>	<u>-8.50</u>	<u>64.8%</u>	**78.6%**	0.44	0.41	0.59	0.59	<u>24.9</u>
TAGMOL	**-7.02**	**-7.77**	**-7.95**	**-8.07**	**-8.59**	**-8.69**	**69.8%**	<u>76.4%</u>	<u>0.55</u>	<u>0.56</u>	0.56	0.56	**27.7**

- **Dataset:** CrossDocked2020 [1]
- **Properties optimized:** Vina Score, QED & SA
- **Generative Backbone:** TargetDiff [2]
- **Evaluation Metric:** Hit rate *(QED ≥ 0.4, SA ≥ 0.5 & Vina Dock ≤ 8.18 kcal/mol)*





REFERENCES

[1] Francoeur, P. G., Masuda, T., Sunseri, J., Jia, A., Iovanisci, R. B., Snyder, I., & Koes, D. R. (2020). Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. Journal of chemical information and modeling, 60(9), 4200-4215.

[2] Guan, J., Qian, W. W., Peng, X., Su, Y., Peng, J., & Ma, J. 3D Equivariant Diffusion for Target-Aware Molecule Generation and Affinity Prediction. In The Eleventh International Conference on Learning Representations.

PAPER CODE