Uncertainty-Aware Multi-Objective Reinforcement Learning-Guided Diffusion Models for 3D De Novo Molecular Design

Lianghong Chen, Dongkyu Kim, Mike Domaratzki, Pingzhao Hu*

Western University, Ontario, Canada

*Corresponding author (Email: phu49@uwo.ca)

Background

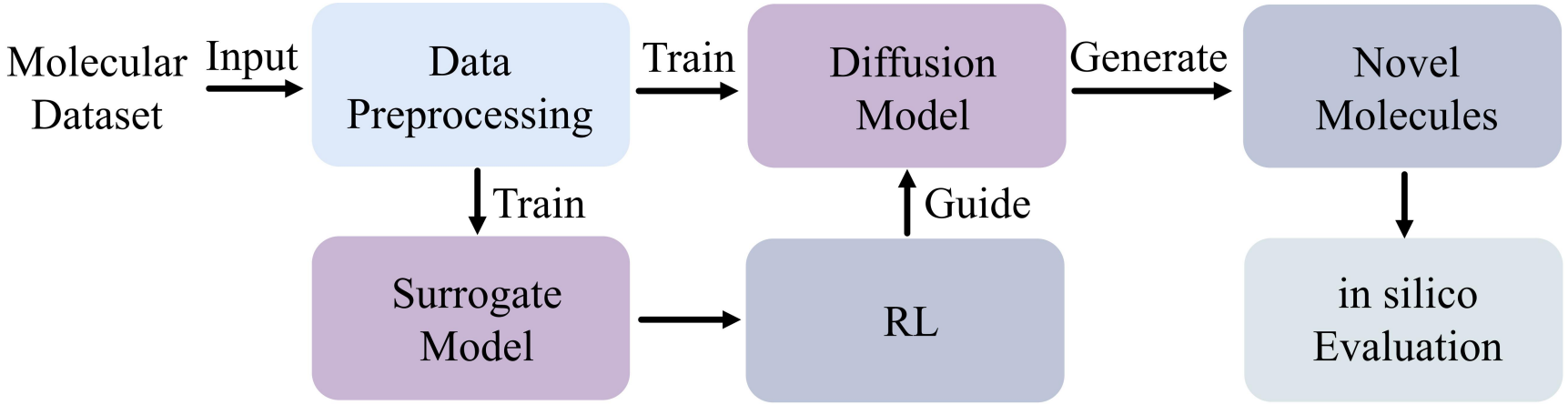
- Lack of a **unified framework** for generating molecules with **multiple drug-relevant properties** and strong **protein binding**
- Reinforcement Learning (RL) has been explored for 1D/2D molecular generative models, but 3D molecular diffusion models with RL guidance remain **underexplored**
- Previous RL methods for diffusion model optimization suffer from challenges such as **reward sparsity** and **mode collapse**
- When **multiple conflicting optimization objectives** exist, it difficult to **balance** these objectives, and algorithms tend to **bias** toward objectives that are easier to optimize

Contributions

- **First unified framework** integrating RL, diffusion models, 3D molecular generation, and uncertainty-based multi-objective optimization
- **Uncertainty-aware reward** design enables **balanced** multi-objective optimization for drug-relevant properties
- Three auxiliary mechanisms address RL–diffusion challenges such as mode collapse and reward sparsity:
  - **Reward boosting** for validity, uniqueness, and novelty
  - **Diversity penalty** to enhance exploration
  - **Dynamic cutoff** strategy for stable, adaptive reward signals
- Consistent performance **gains** across datasets

Study Design

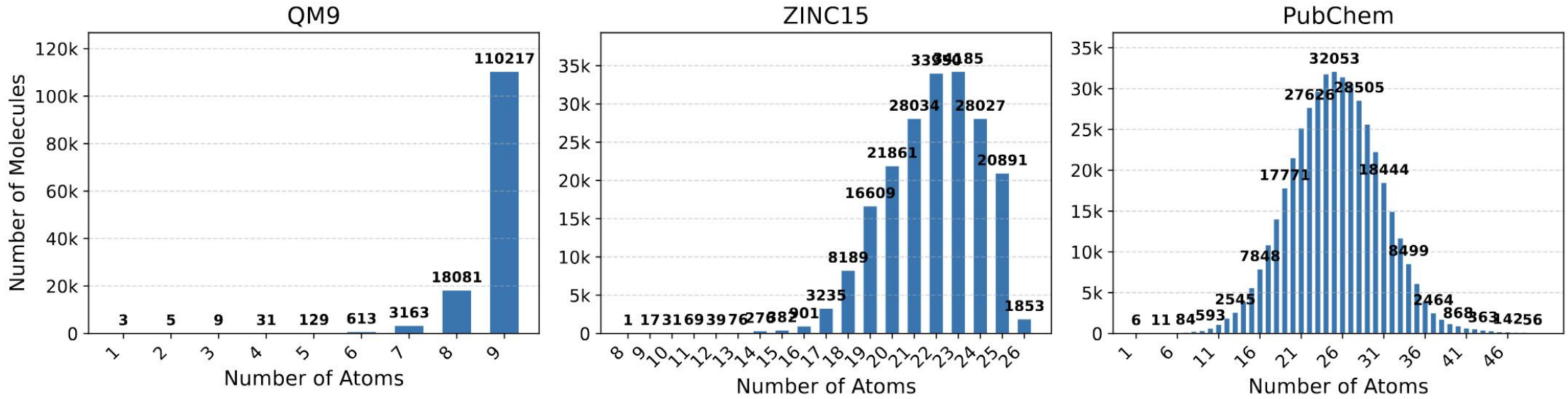
- Train surrogate models to **predict uncertainty**
- Pretrain the diffusion model to learn **3D molecular generation**
- RL guides the diffusion model to generate molecules with **desired drug-relevant objectives**
- Perform in silico evaluation to further verify the **drug potential** and **binding stability**



Datasets

- QM9: **small-sized** molecules used for basic **3D molecular generation** tasks
- ZINC15: **medium-sized**, **drug-like** molecules
- PubChem: **large-scale** collection of **diverse** and **complex** molecules

Dataset	Unique Atom Types
QM9	C, N, O, F
ZINC15	C, N, O, F, P, S, Cl, Br, I
PubChem	C, N, O, F, P, S, Cl, Br, I



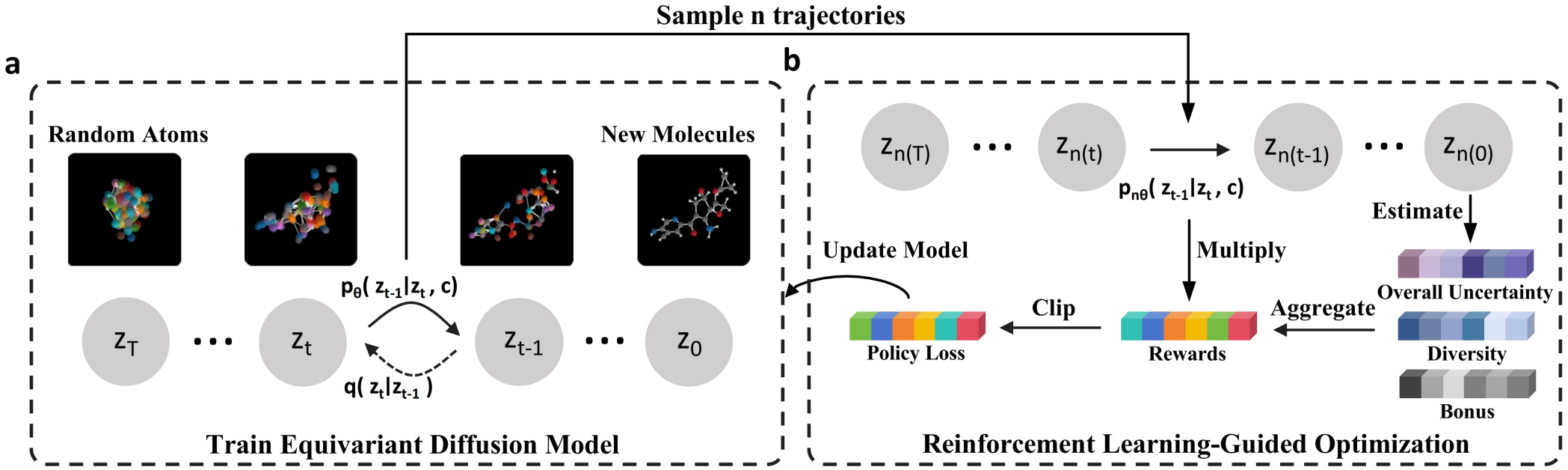
Method

- **Generate** n candidate molecules and **evaluate** them to obtain RL **rewards**
- Reward design:

generated molecule — property threshold — similarity-based penalty

$$R_{\text{total}}(m;\delta_1,\ldots,\delta_k,t_{\text{episode}}) = U_{\text{multi}}(m;\delta_1,\ldots,\delta_k)\cdot R_{\text{bonus}}(m) - \lambda(t_{\text{episode}})\cdot D(m) \quad (1)$$

uncertainty — # properties — quality-based bonus

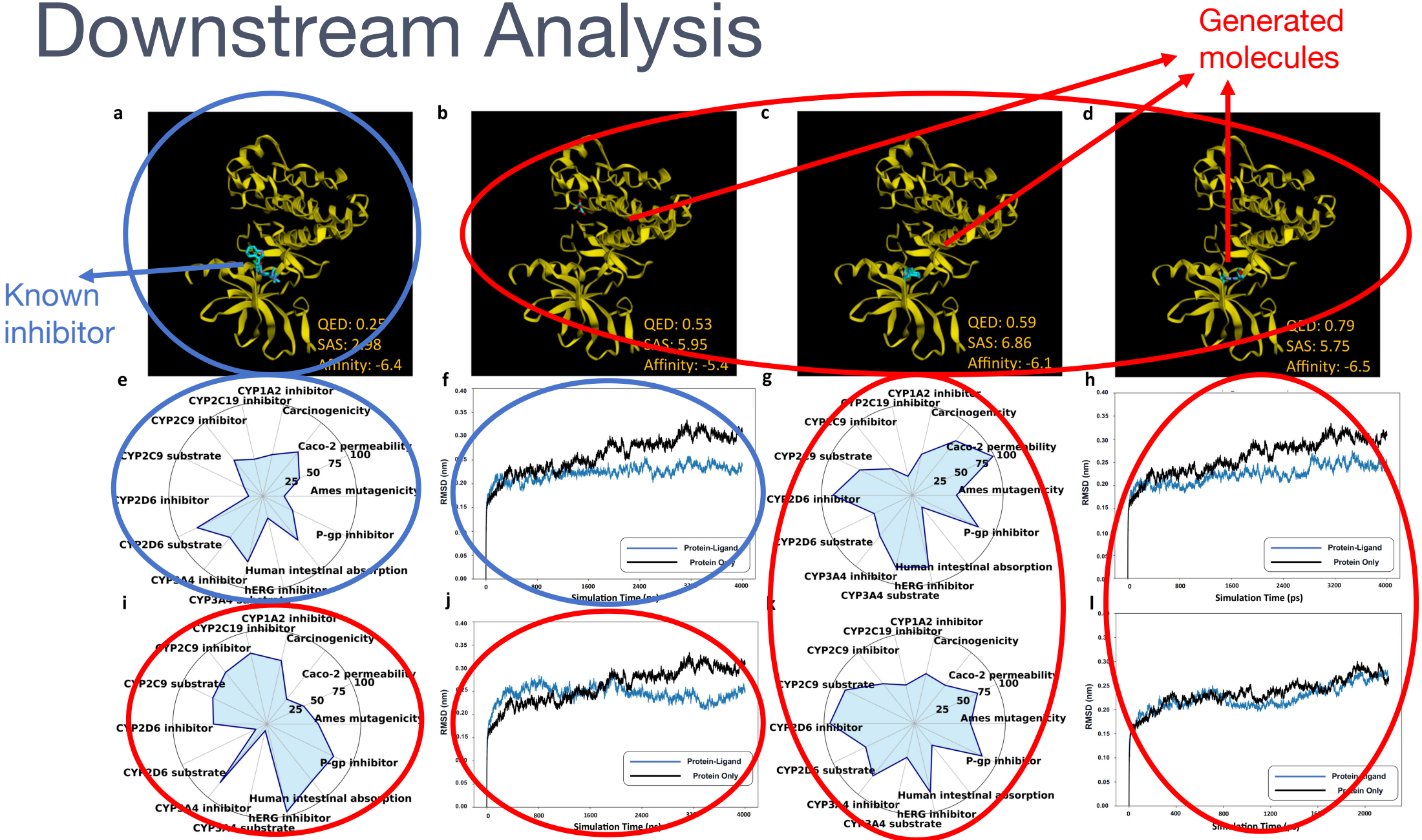


Results

Dataset	Method	Val (%) (↑)	Uni (%) (↑)	Nov (%) (↑)	VUN (%) (↑)	ASta (%) (↑)	MSta (%) (↑)	Top (%) (↑)
QM9	W/O RL	88.55 ± 0.65	**97.57 ± 0.30**	99.75 ± 0.15	86.19 ± 1.02	99.34 ± 0.11	95.90 ± 0.34	25.17 ± 1.17
	SFT-PG	88.57 ± 1.33	96.80 ± 0.33	99.80 ± 0.15	85.57 ± 1.60	99.24 ± 0.10	95.62 ± 0.58	25.58 ± 1.65
	DDPO-SF	88.65 ± 1.05	97.39 ± 0.48	99.79 ± 0.20	86.16 ± 1.59	99.25 ± 0.09	95.50 ± 0.34	25.65 ± 1.51
	DDPO-IS	88.82 ± 1.03	96.59 ± 0.71	99.39 ± 0.04	85.27 ± 1.30	97.31 ± 0.09	86.10 ± 0.64	25.77 ± 1.29
	DPOK	88.10 ± 0.37	97.52 ± 0.42	**99.81 ± 0.15**	85.75 ± 0.80	99.18 ± 0.03	95.28 ± 0.18	25.20 ± 1.44
	Ours	**98.17 ± 0.07**	90.90 ± 0.72	99.63 ± 0.04	**88.90 ± 0.68**	**99.87 ± 0.03**	**99.17 ± 0.27**	**28.33 ± 0.61**
ZINC15	W/O RL	30.05 ± 1.34	**100.00 ± 0.00**	**100.00 ± 0.00**	30.05 ± 1.34	88.36 ± 0.40	12.00 ± 1.33	8.02 ± 0.46
	SFT-PG	41.25 ± 1.48	**100.00 ± 0.00**	**100.00 ± 0.00**	41.25 ± 1.48	91.71 ± 0.08	25.55 ± 0.65	10.43 ± 0.73
	DDPO-SF	30.25 ± 1.56	**100.00 ± 0.00**	**100.00 ± 0.00**	30.25 ± 1.56	88.37 ± 0.40	11.97 ± 1.41	8.05 ± 0.61
	DDPO-IS	30.47 ± 1.39	**100.00 ± 0.00**	**100.00 ± 0.00**	30.47 ± 1.39	88.35 ± 0.44	12.02 ± 1.57	8.13 ± 0.60
	DPOK	30.13 ± 2.28	**100.00 ± 0.00**	**100.00 ± 0.00**	30.13 ± 2.28	88.42 ± 0.61	12.01 ± 1.38	8.02 ± 0.78
	Ours	**99.02 ± 0.46**	99.75 ± 0.06	**100.00 ± 0.00**	**98.77 ± 0.49**	**99.86 ± 0.03**	**98.08 ± 0.63**	**33.40 ± 0.89**
PubChem	W/O RL	7.18 ± 4.78	99.67 ± 0.65	**100.00 ± 0.00**	7.17 ± 4.80	94.51 ± 0.16	38.18 ± 0.92	2.23 ± 1.65
	SFT-PG	7.47 ± 1.40	99.57 ± 0.85	**100.00 ± 0.00**	7.44 ± 1.37	82.99 ± 0.49	33.25 ± 0.34	2.03 ± 0.76
	DDPO-SF	7.98 ± 2.96	**100.00 ± 0.00**	**100.00 ± 0.00**	7.98 ± 2.96	94.49 ± 0.98	44.22 ± 0.32	2.40 ± 0.37
	DDPO-IS	10.50 ± 6.19	99.90 ± 0.20	**100.00 ± 0.00**	10.48 ± 6.16	95.36 ± 0.99	45.37 ± 1.85	2.52 ± 1.22
	DPOK	7.65 ± 1.75	99.67 ± 0.64	**100.00 ± 0.00**	7.62 ± 1.76	94.51 ± 0.20	38.17 ± 0.64	2.42 ± 0.46
	Ours	**16.23 ± 9.72**	**100.00 ± 0.00**	**100.00 ± 0.00**	**16.23 ± 9.72**	**99.04 ± 0.13**	**88.65 ± 0.59**	**2.97 ± 1.60**

Val", "Uni", and "Nov" represent the percentages of valid, unique, and novel molecules, respectively. "VUN" is their joint metric computed as Val × Uni × Nov, representing the percentage of molecules that are simultaneously valid, unique, and novel. "ASta" and "MSta" denote atom-level and molecule-level stability. "Top" indicates the proportion of generated molecules that simultaneously satisfy all three property constraints

Downstream Analysis



Conclusion

- Proposed an uncertainty-aware RL framework to guide diffusion models for **3D drug-like molecular** generation
- Achieved superior performance across three benchmark datasets, demonstrating **strong stability** and **generalization**
- Molecular dynamic simulation and ADMET analysis **varified** the structural stability and drug-like potential of generated molecules

Future Work

- Extend the framework to **broader drug-related property objectives**
- Incorporate chemical prior knowledge to tackle **more complex drug targets**

Acknowledge

This work was supported in part by the Canada Research Chairs Tier II Program (CRC-2021-00482), the Canadian Institutes of Health Research (PLL 185683, PJT 190272, PJT204042), the Natural Sciences, Engineering Research Council of Canada (RGPIN-2021-04072) and The Canada Foundation for Innovation (CFI) John R. Evans Leaders Fund (JELF) program (\#43481), the Studentship/Fellowship funded by Breast Cancer Canada, and the Vector Scholarship in Artificial Intelligence provided through the Vector Institute.