



# Graph Diffusion Transformers are In-Context Molecular Designers

Gang Liu<sup>1</sup>, Jie Chen<sup>3</sup>, Yihan Zhu<sup>1</sup>, Michael Sun<sup>2</sup>, Tengfei Luo<sup>1</sup>, Nitesh V. Chawla<sup>1</sup>, Meng Jiang<sup>1</sup>



<sup>1</sup>University of Notre Dame <sup>2</sup>MIT CSAIL <sup>3</sup>MIT-IBM Watson AI Lab, IBM Research



Code & Model

## ▼ In-Context Learning ability: the ability unlocked by a molecular foundation model from millions of tasks

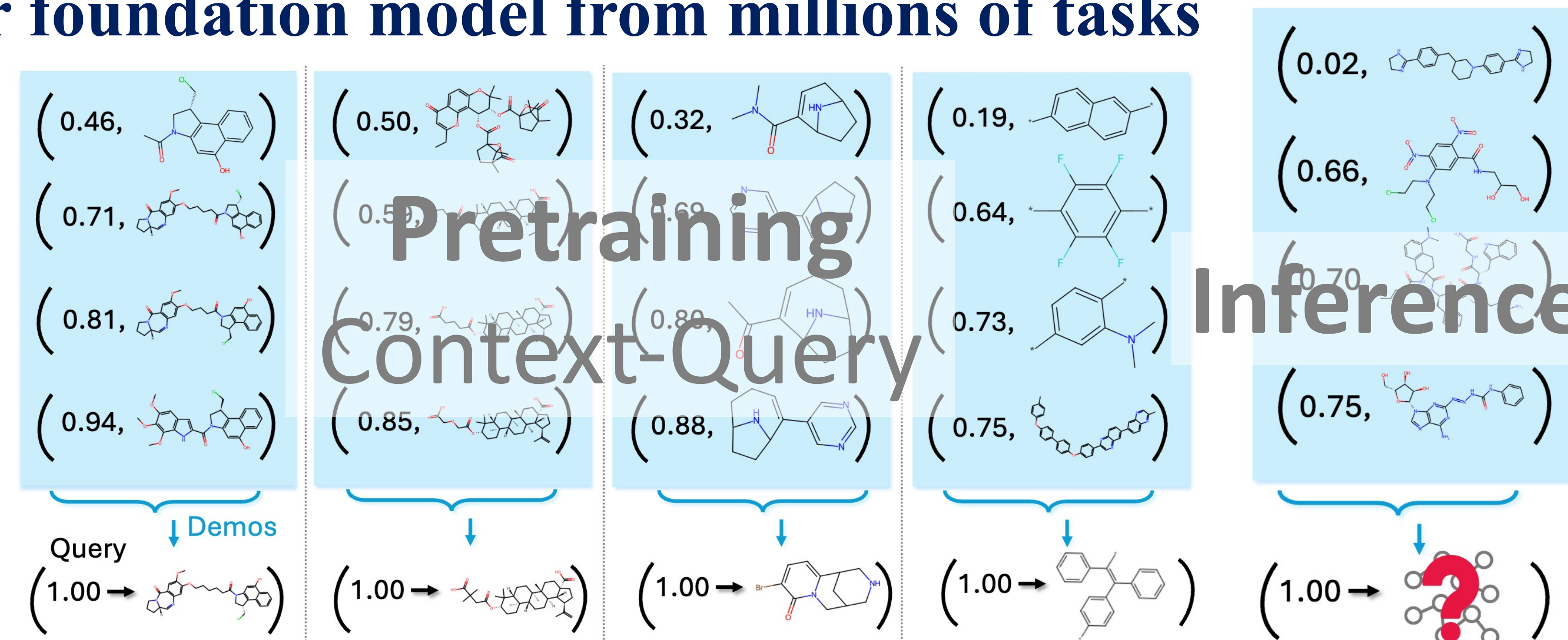
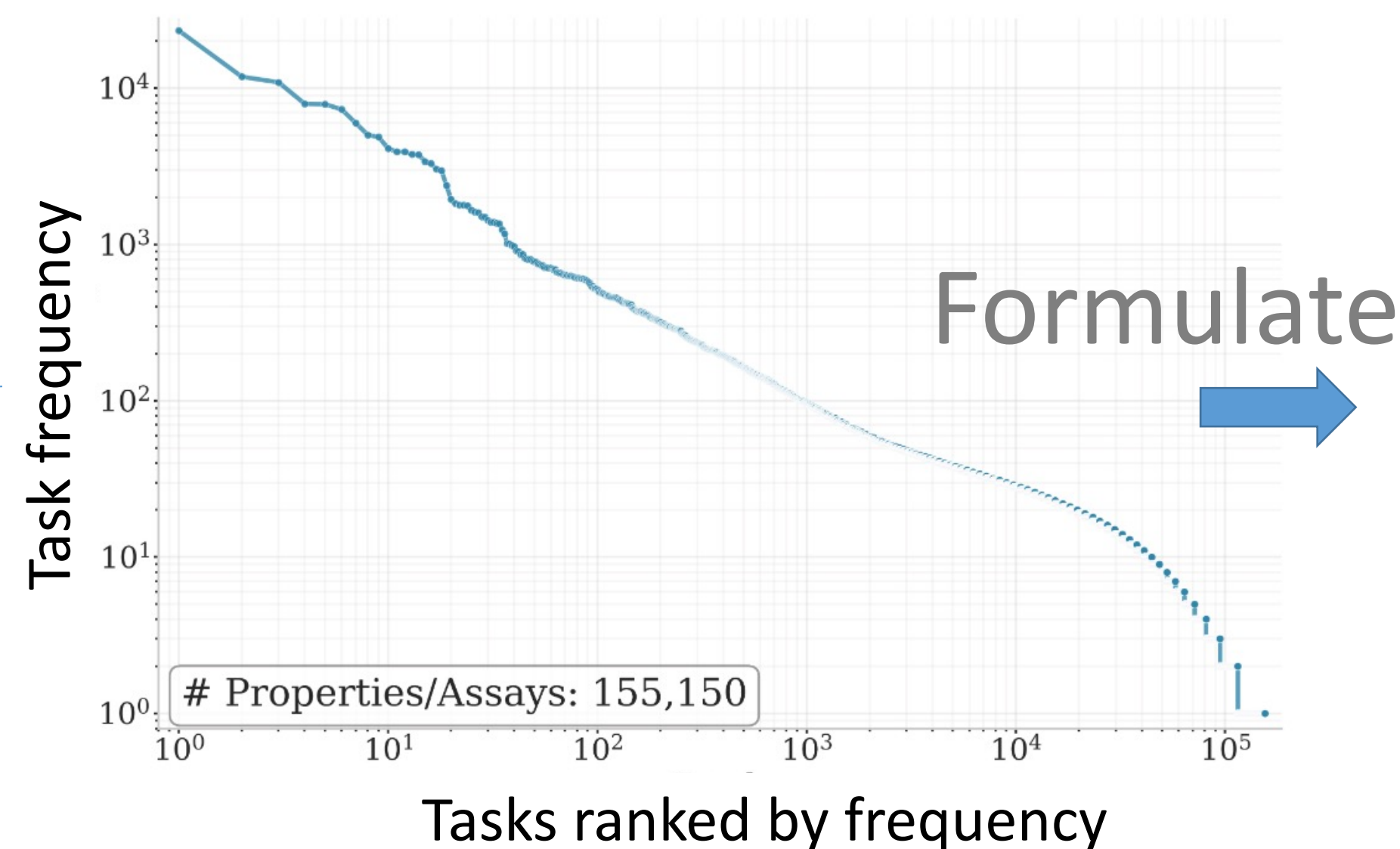
Functional: In vitro cell toxicity against chinese hamster AA8 cell line after 4hr of compound exposure

ADME: Cytotoxicity against human H9 cells after 4 days by coulter counter

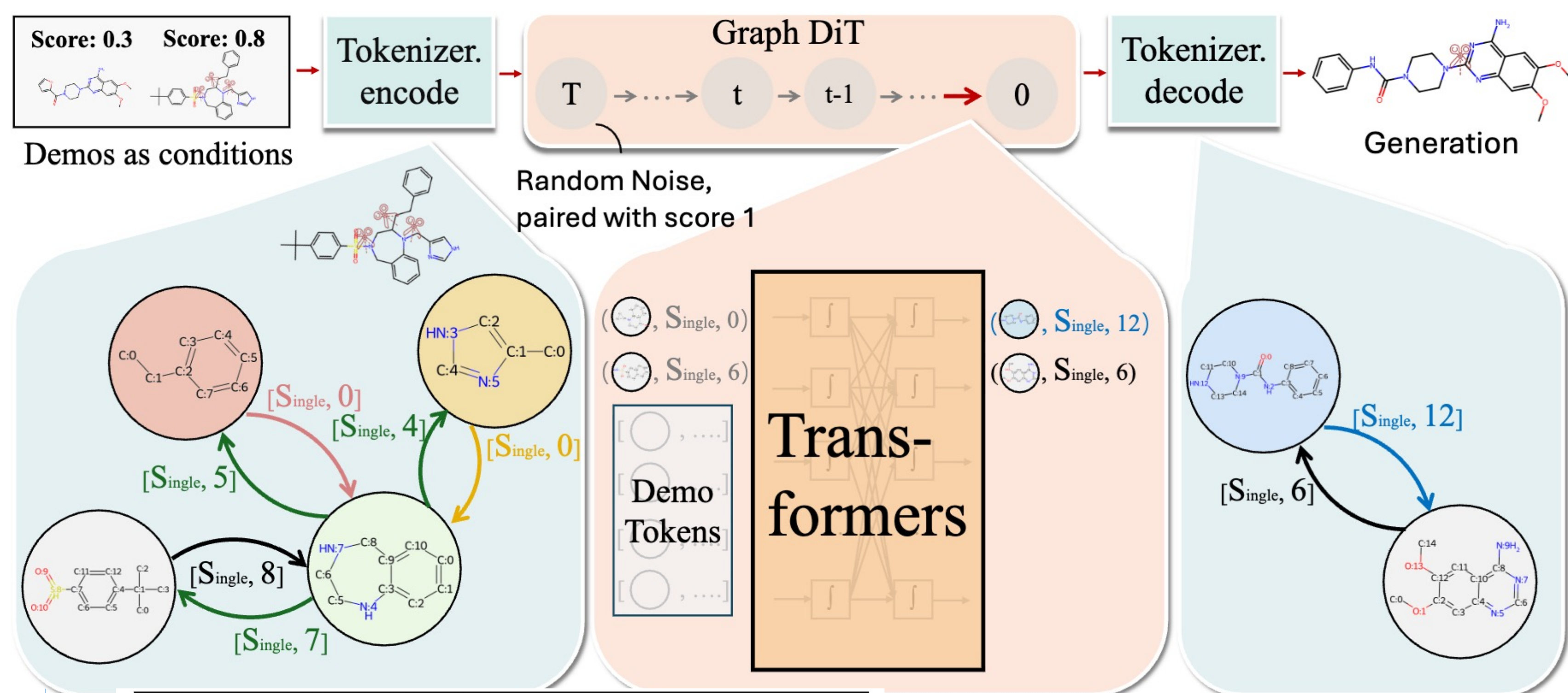
Binding: In vitro binding affinity towards (alpha-4)2(beta-2)3 neuronal nicotinic acetylcholine receptor in P2 membrane fractions of rat forebrain

Materials: Thermal conductivity [W/(m\*K)] for polymeric materials

- Tasks (bioassays, material properties) from ChEMBL, Polymer membrane databases, etc,



## ▼ DemoDiff: Learning Diffusion Models from Demonstrations



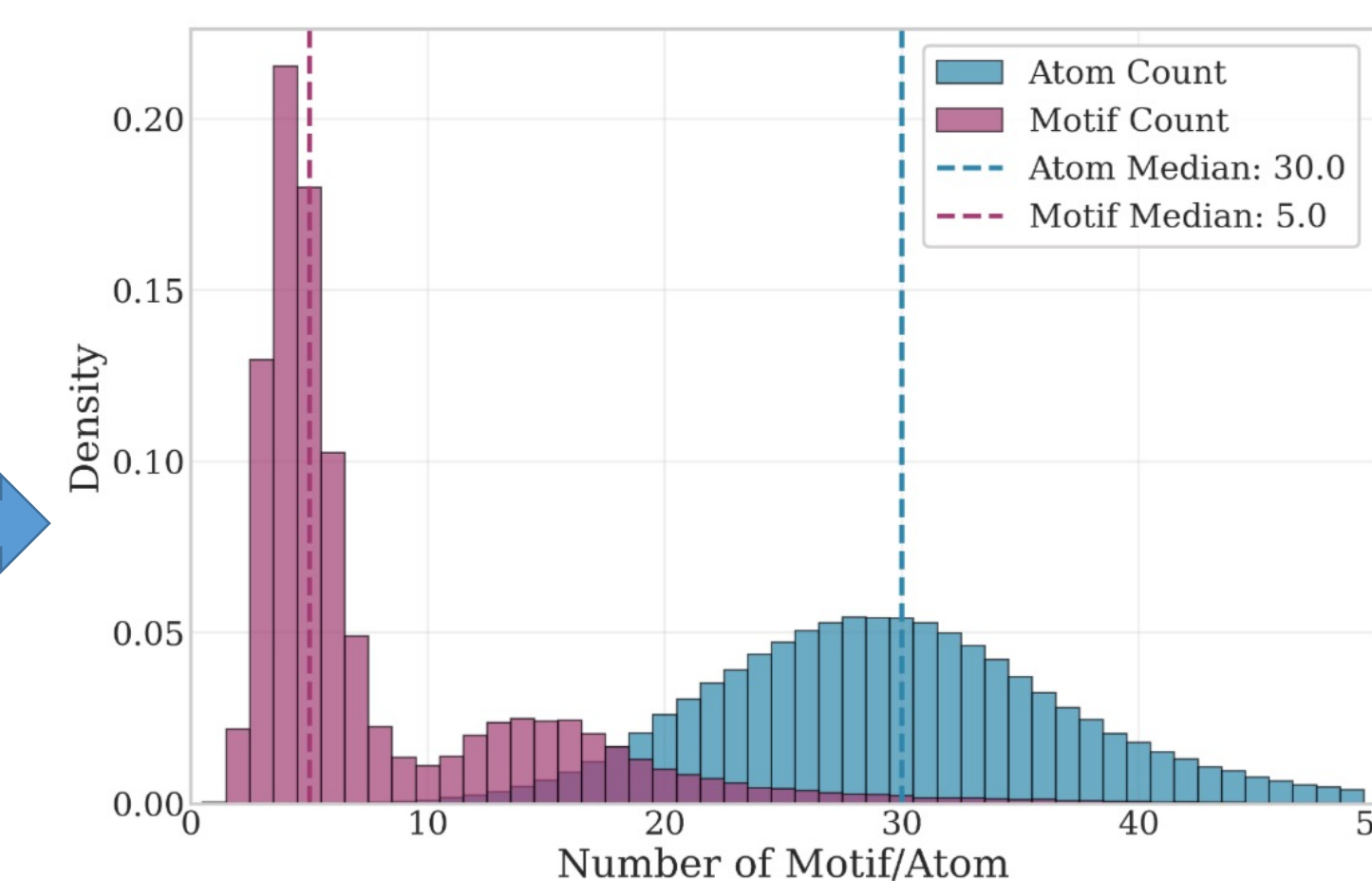
### Algorithm 1 Node Pair Encoding (NPE) with Constraints

Require: molecule list  $\mathcal{D}$ , motif vocabulary  $\mathcal{M} = \emptyset$ , max size  $K$ , ring count threshold  $N_{ring}$

Ensure: motif vocabulary  $\mathcal{M}$

- 1: Initialize each molecule  $X \in \mathcal{D}$  with atom-level and ring-based motifs
- 2: Count frequencies of ring-based motifs across  $\mathcal{D}$
- 3: Add all periodic-table elements, polymerization "\*", and top- $N_{ring}$  frequent rings to  $\mathcal{M}$
- 4: while  $|\mathcal{M}| < K$  do
- 5: (1) Merge Neighbor: Initialize empty multiset  $\mathcal{S} \leftarrow \emptyset$
- 6: for each molecule  $X \in \mathcal{D}$  do
- 7: for each motif  $m$  from  $X$  do
- 8: for each adjacent motif  $m'$  in  $X$  such that  $m$  and  $m'$  are mergeable under structural constraints (e.g., rings treated as units) do
- 9: form new motif  $m \leftarrow m \cup m'$
- 10: add  $m$  to multiset  $\mathcal{S}$  with frequency count
- 11: end for
- 12: end for
- 13: end for
- 14: (2) Frequency Selection: Find most frequent motif  $m^* \in \mathcal{S}$
- 15: (3) Update Graph:
- 16: for each molecule  $X \in \mathcal{D}$  do
- 17: for each pair of adjacent motifs  $(m, m')$  in  $X$  do
- 18: if their merged form equals  $m^*$  then
- 19: replace  $m$  and  $m'$  with  $m^*$  in  $X$
- 20: end if
- 21: end for
- 22: end for
- 23: Add  $m^*$  to  $\mathcal{M}$  if not already in it
- 24: end while
- 25: return  $\mathcal{M}$

NPE for motif vocabulary construction



Molecular size after tokenization (cut by 50 nodes)

## ▼ Results (More Analysis)

We compute the oracle and diversity scores from the Top-10 generations and report the harmonic mean of these two values. We group 33 tasks into six categories and report the mean  $\pm$  std within each category. The best results in each column are bolded.

Task	Drug Rediscovery	Drug MPO	Structure Constrained	Drug Design	Target Based	Material Design	Avg Rank	Total Sum
# Tasks	7	7	5	4	5	5	33	33
Molecular Optimization Methods with 100 Oracle Calls								
GraphGA	0.36±0.07	0.52±0.19	0.43±0.21	0.41±0.32	0.76±0.04	0.58±0.11	6.56	16.65
REINVENT	0.37±0.08	0.52±0.17	0.43±0.21	0.42±0.32	0.76±0.03	0.00±0.00	8.44	13.84
GPBO	0.37±0.07	0.51±0.18	0.42±0.22	0.39±0.33	0.76±0.03	0.60±0.21	6.89	16.65
STONED	0.36±0.07	0.52±0.19	0.39±0.24	0.37±0.36	0.76±0.04	No pSMILES	8.18	13.42
Genetic GFN	0.36±0.10	0.51±0.22	0.38±0.29	0.33±0.37	0.76±0.03	No pSMILES	9.28	13.04
GenMol	0.42±0.08	0.51±0.17	0.42±0.25	0.55±0.21	0.69±0.03	0.62±0.08	7.98	17.36
Molecular Optimization Methods with 10000 Predictor Calls								
GraphGA	0.37±0.09	0.50±0.18	0.45±0.29	0.49±0.26	0.64±0.06	0.55±0.14	9.30	16.30
REINVENT	0.30±0.13	0.23±0.23	0.25±0.24	0.38±0.26	0.17±0.13	0.51±0.16	13.60	9.82
GPBO	0.33±0.10	0.45±0.22	0.43±0.29	0.49±0.15	0.74±0.03	0.42±0.24	9.87	15.35
STONED	0.33±0.10	0.40±0.20	0.50±0.28	0.27±0.10	0.20±0.27	No pSMILES	12.01	9.66
Genetic GFN	0.14±0.08	0.08±0.16	0.08±0.11	0.12±0.11	0.15±0.15	No pSMILES	17.95	3.14
GenMol	0.34±0.07	0.28±0.23	0.21±0.29	0.39±0.19	0.52±0.05	0.21±0.23	12.63	10.37
Conditional Generation Models								
LSTM	0.39±0.25	0.16±0.07	0.55±0.32	0.33±0.35	0.72±0.04	0.16±0.11	11.77	12.30
Graph-DiT	0.43±0.21	0.50±0.18	<b>0.58±0.34</b>	0.48±0.37	0.71±0.04	0.55±0.17	8.53	17.64
Learning from In-Context Demonstrations								
DeepSeek-V3	0.45±0.18	0.51±0.20	0.49±0.24	0.65±0.18	0.64±0.06	0.39±0.24	8.08	16.90
GPT-4o	<b>0.47±0.21</b>	0.53±0.20	0.52±0.30	0.48±0.40	0.73±0.05	0.43±0.16	7.89	17.25
Qwen-Max	0.15±0.21	0.17±0.15	0.32±0.32	0.29±0.29	0.19±0.26	0.10±0.18	15.39	6.46
Llama3.1-8B-FT	0.21±0.13	0.24±0.23	0.24±0.34	0.31±0.31	0.02±0.05	0.29±0.40	14.91	6.04
Qwen3-8B-FT	0.37±0.19	0.27±0.16	0.26±0.34	0.46±0.22	0.67±0.07	0.44±0.39	10.96	12.23
DemoDiff (Ours)	0.44±0.21	<b>0.54±0.23</b>	0.56±0.33	<b>0.79±0.11</b>	<b>0.78±0.05</b>	<b>0.67±0.11</b>	<b>4.10</b>	<b>20.10</b>

## ▼ Acknowledgement

This work was partially supported by NSF IIS-2142827, IIS-2146761, IIS-2234058, and CBET-2332270. We also appreciate the support from the Foundation Models and Applications Lab of Lucy Institute, ND-IBM Tech Ethics Lab, and OpenAI Researcher Access Program. Contact: Gang Liu (gliu7@nd.edu)