Atomwise

Can simple pre-training be as effective as self-supervised learning (in ADME modeling)?

Hossam Ashtawy, PhD

August 24, 2021

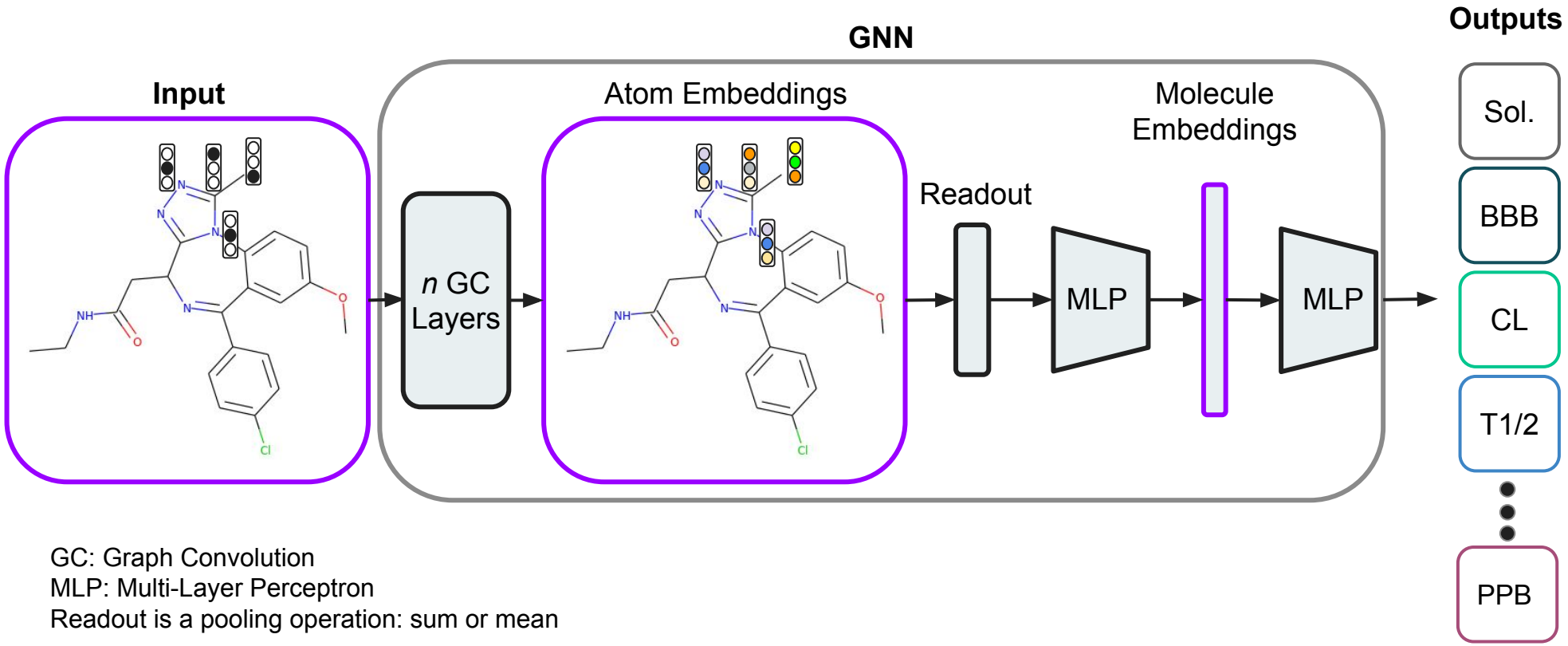
Atomwise Co-Authors: Brandon Anderson, Jon Sorenson, Izhar Wallach

The Problem

- Build generalizable ADMET predictors
- from datasets with varying sizes & diversity for several ADMET endpoints
- Learn & predict the endpoints **simultaneously**

The Problem

Multi-Task Graph Neural Network (GNN) Model

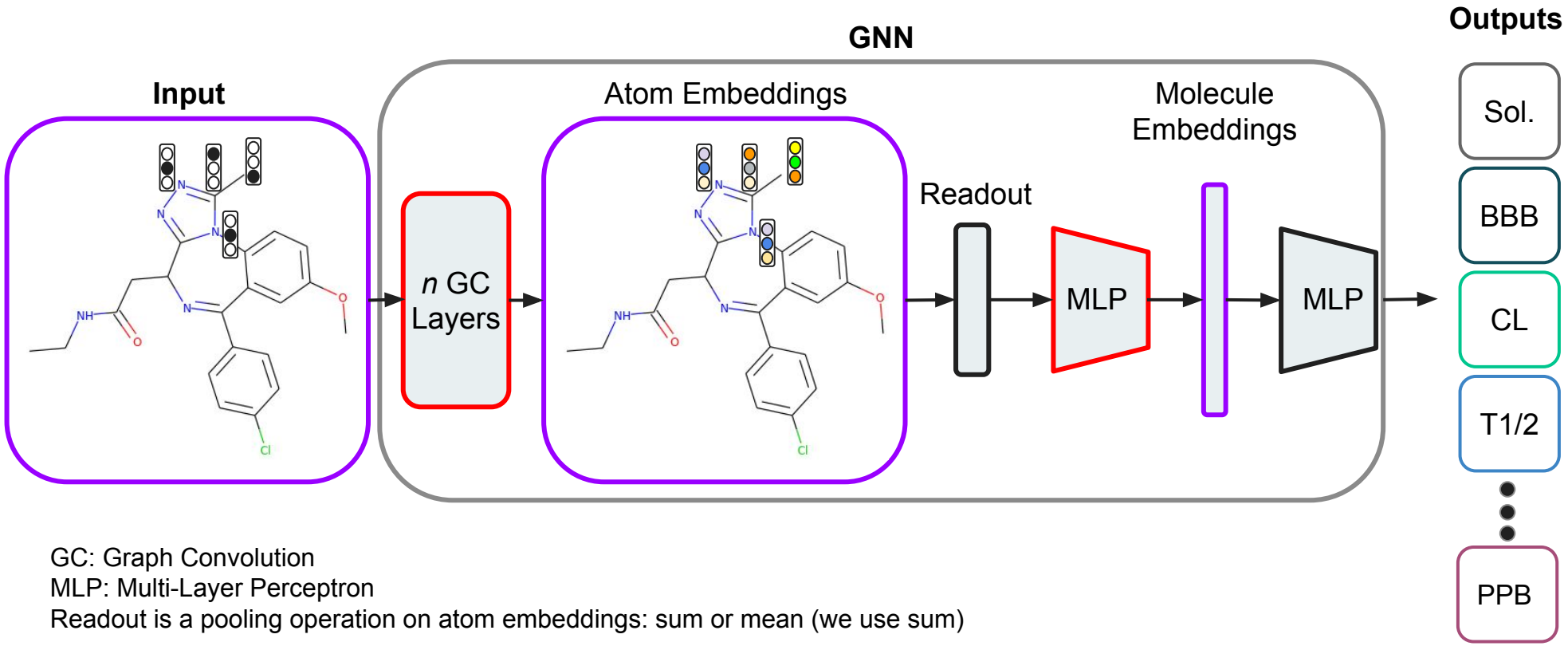


GC: Graph Convolution
MLP: Multi-Layer Perceptron
Readout is a pooling operation: sum or mean

- **GNN** can in theory produce **expressive** representations
- Sub-optimal representations due to small ADMET datasets

Mitigation: Pre-Training

Multi-Task Graph Neural Network (GNN) Model



GC: Graph Convolution
MLP: Multi-Layer Perceptron
Readout is a pooling operation on atom embeddings: sum or mean (we use sum)

Examples of Self-Supervised Learning

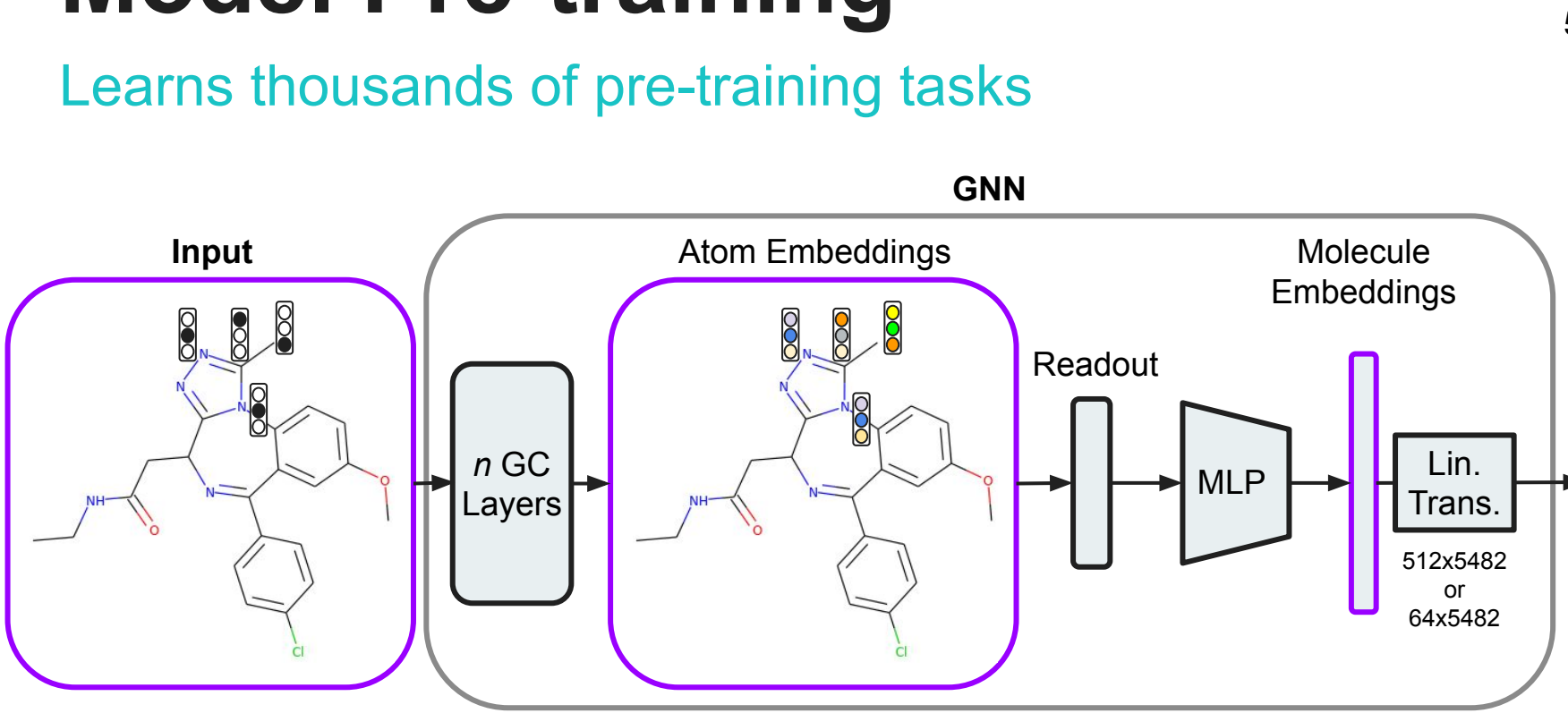
Recent publications

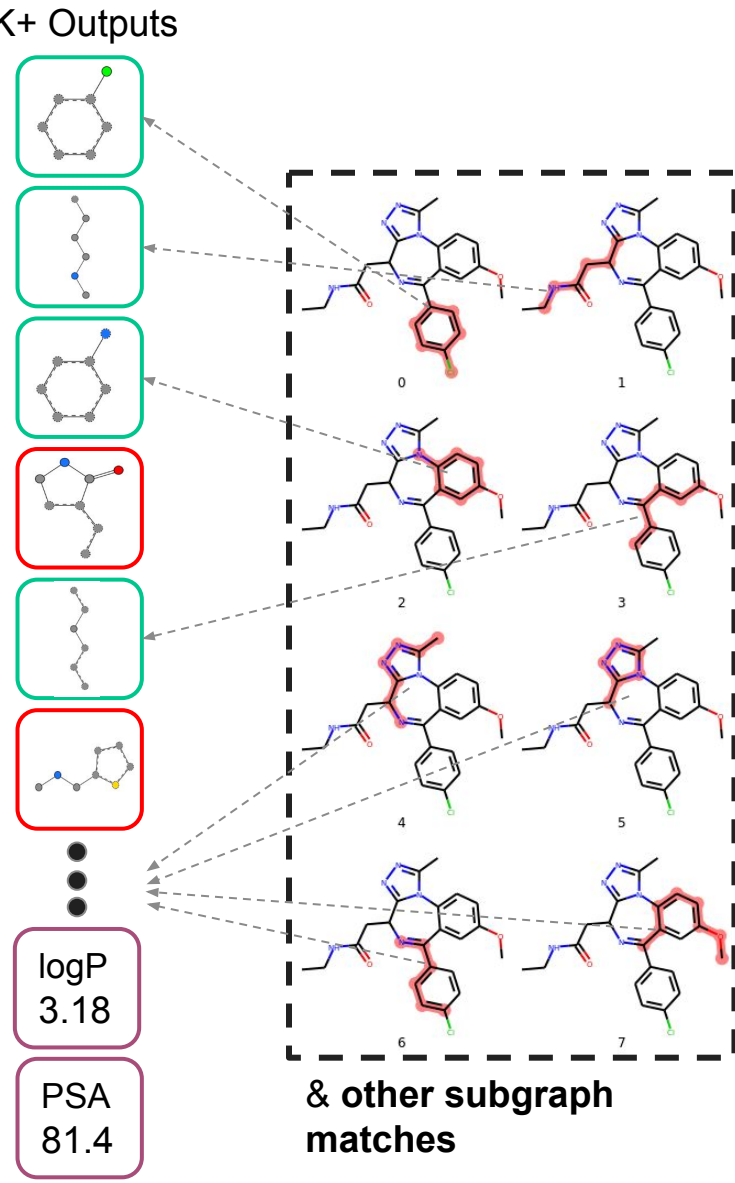
- Context Prediction of Atoms (GNN-based)
  - Hu, Weihua, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. 2019. “Strategies for Pre-Training Graph Neural Networks.” *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/1905.12265.
  - Rong, Yu, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. 2020. “Self-Supervised Graph Transformer on Large-Scale Molecular Data.” *Advances in Neural Information Processing Systems* 33.
- Data Augmentation & Contrastive Loss (GNN-based)
  - Wang, Yuyang, Jianren Wang, Zhonglin Cao, and Amir Barati Farimani. 2021. “MolCLR: Molecular Contrastive Learning of Representations via Graph Neural Networks.” *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/2102.10056.
- Language Models (LSTM-based)
  - Winter, Robin, Floriane Montanari, Frank Noé, and Djork-Arné Clevert. n.d. “Learning Continuous and Data-Driven Molecular Descriptors by Translating Equivalent Chemical Representations.” https://doi.org/10.26434/chemrxiv.6871628.v1.

A **simpler** pre-training strategy ...

Model Pre-training

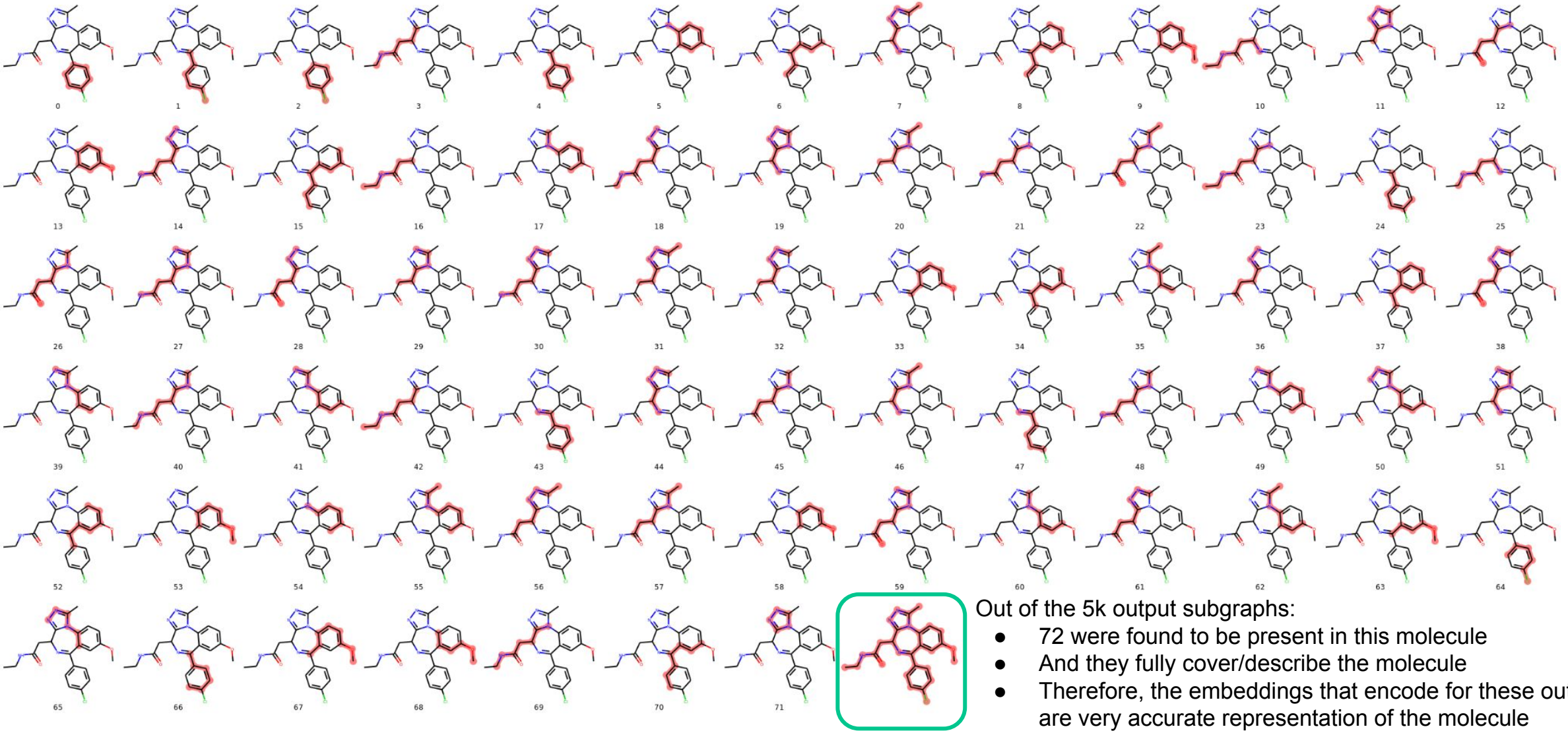
Learns thousands of pre-training tasks





GC: Graph Convolution
MLP: Multi-Layer Perceptron
Readout is a pooling operation: sum or mean

Sub-structures provide full coverage

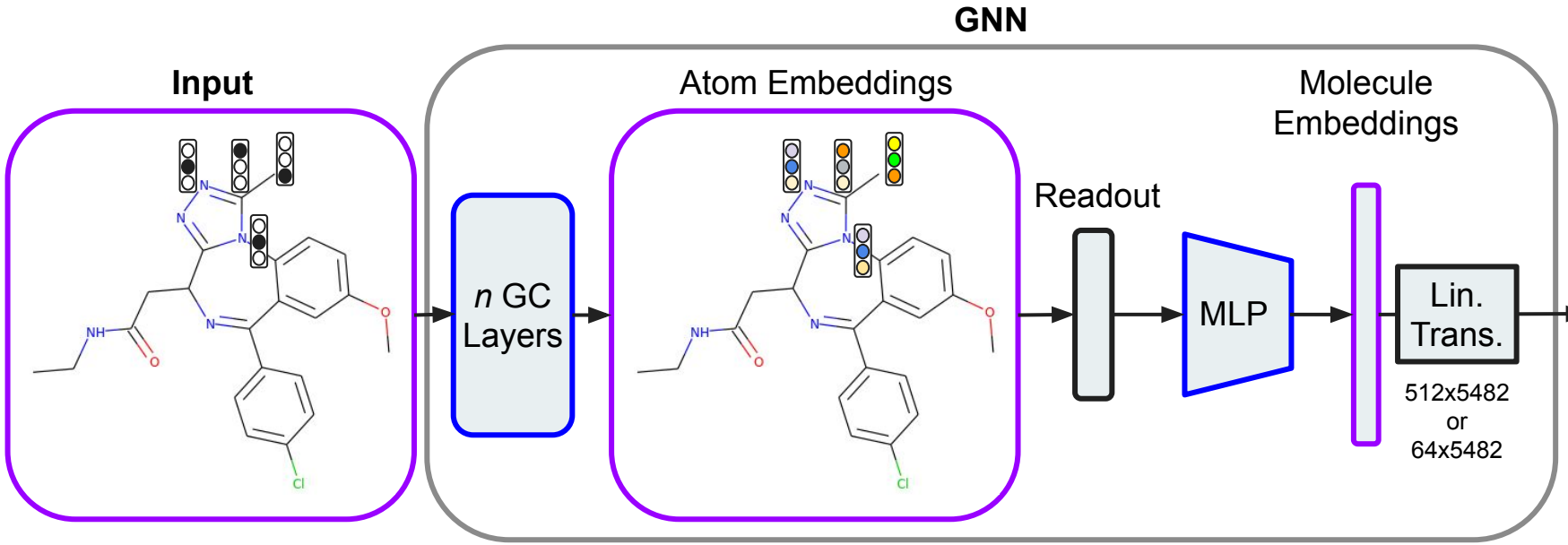


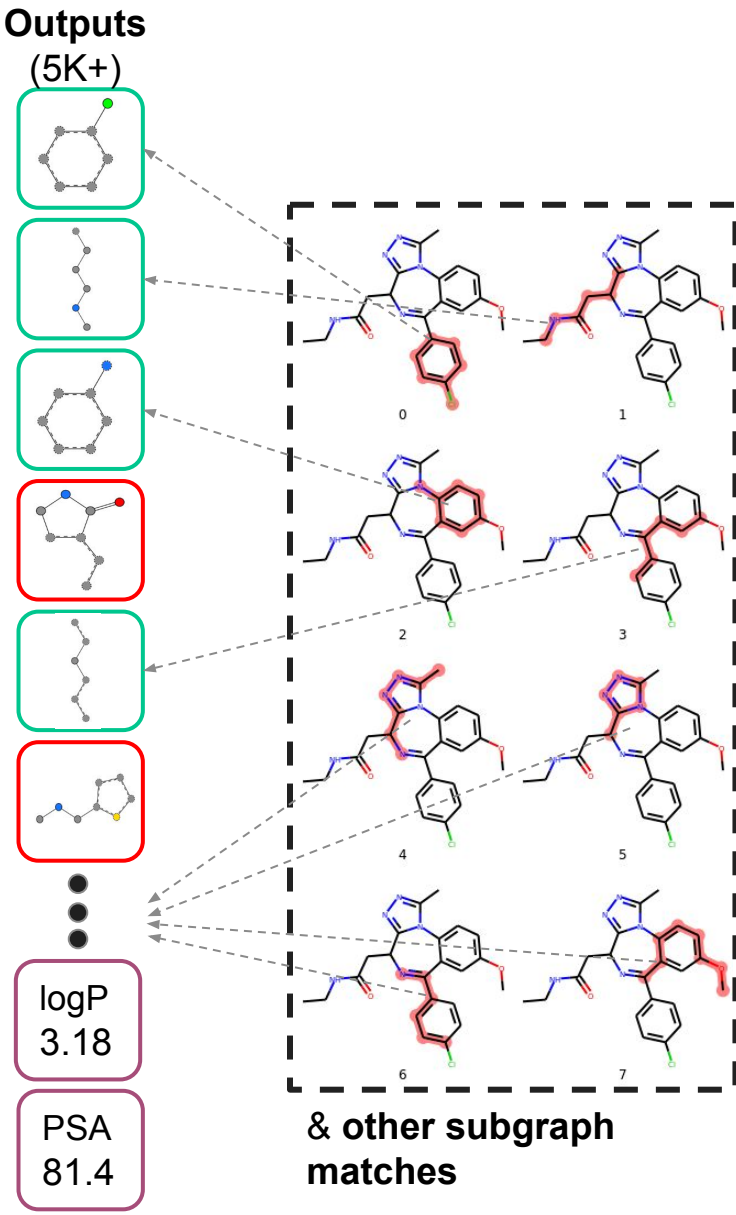
Out of the 5k output subgraphs:

- 72 were found to be present in this molecule
- And they fully cover/describe the molecule
- Therefore, the embeddings that encode for these outputs, are very accurate representation of the molecule

Model Pre-training

Learns thousands of pre-training tasks

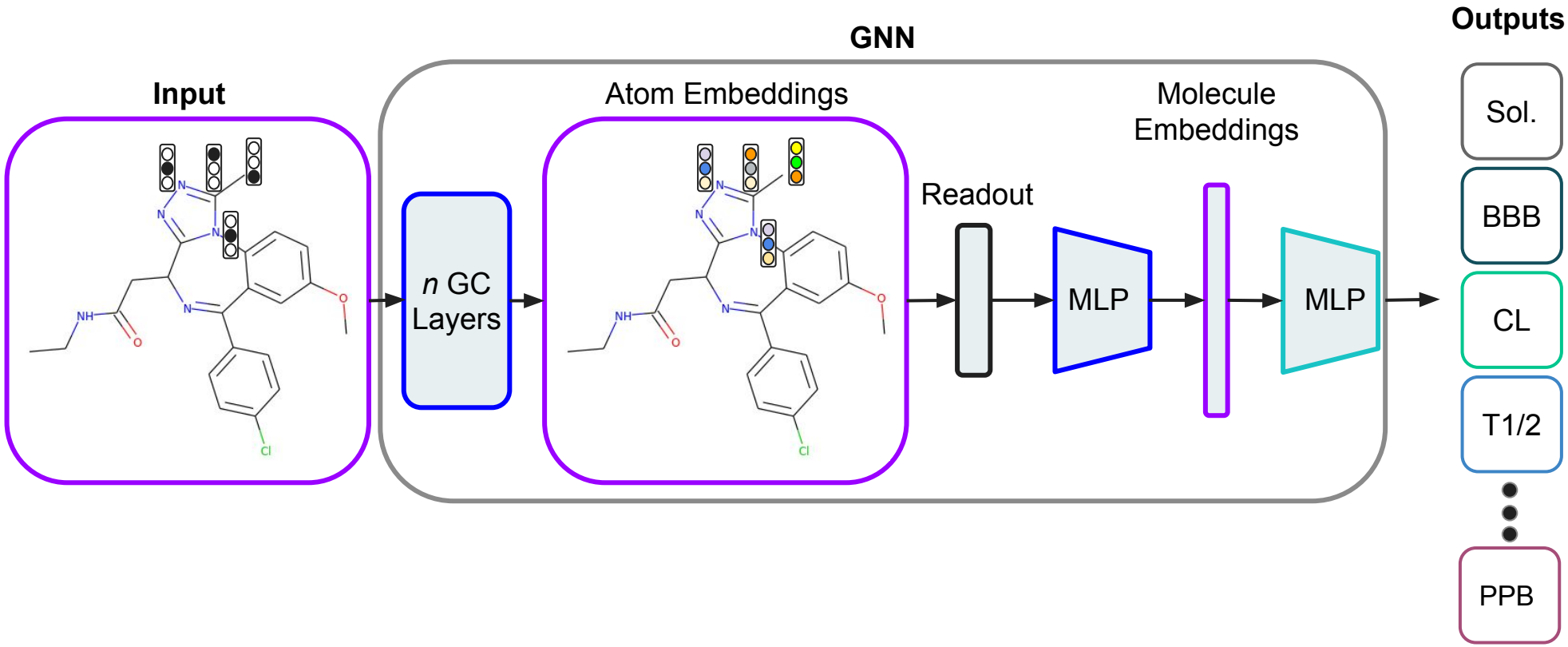




GC: Graph Convolution
MLP: Multi-Layer Perceptron
Readout is a pooling operation: sum or mean

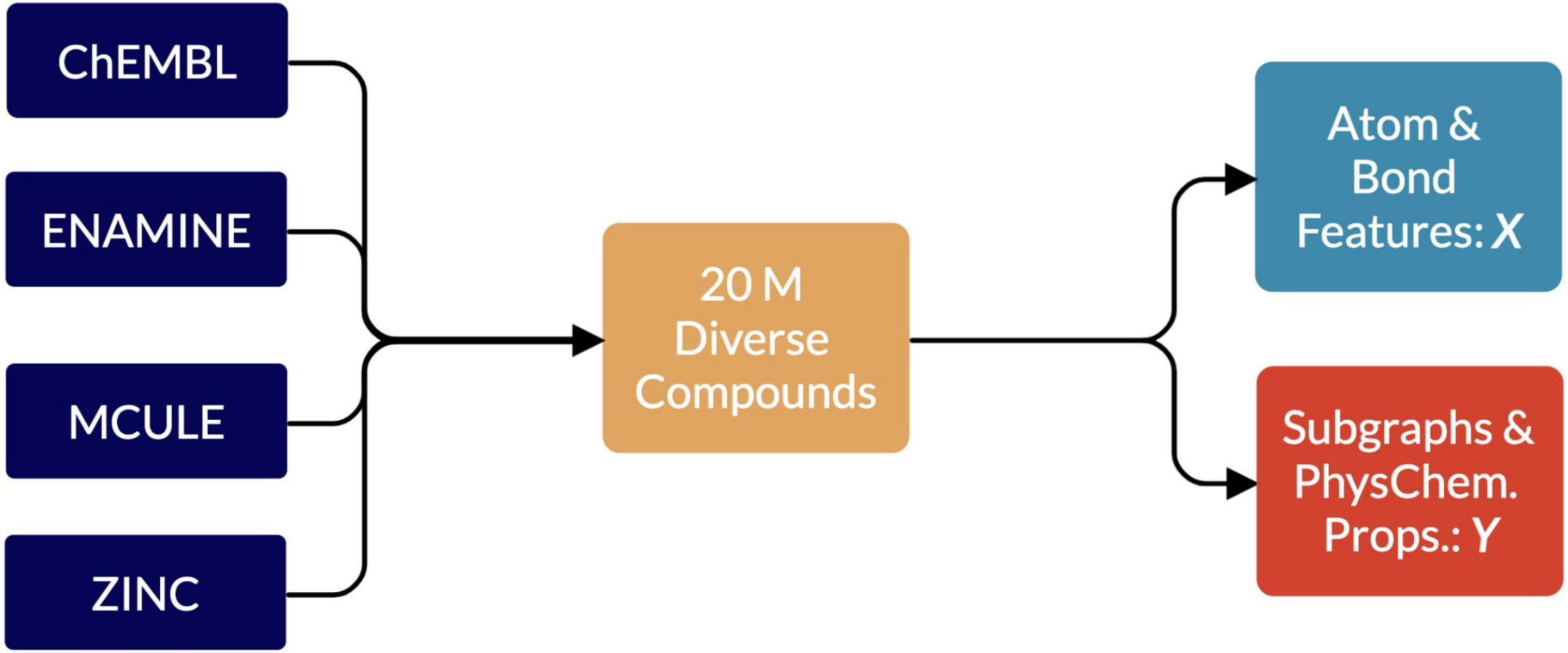
Model Fine-tuning

Fine-tune the pre-trained model on downstream tasks (ADMET)

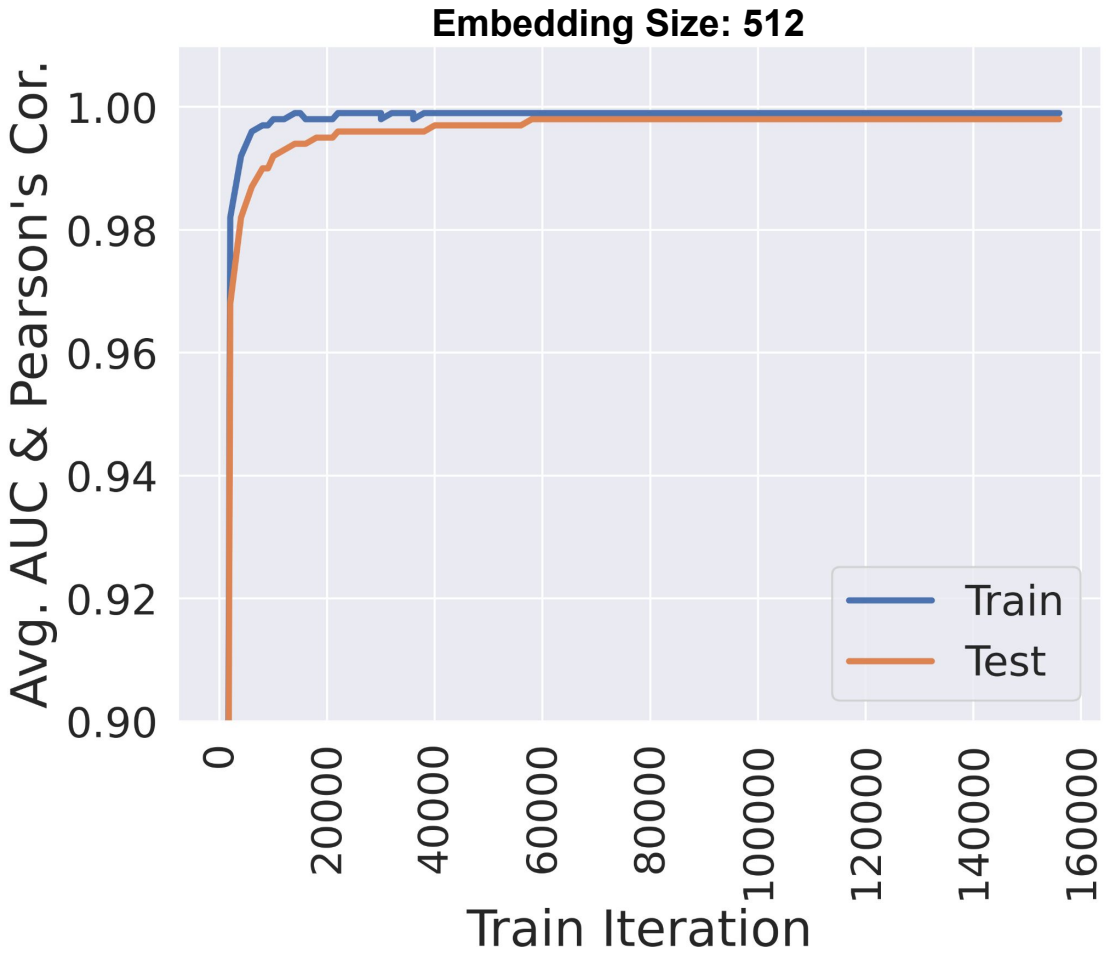


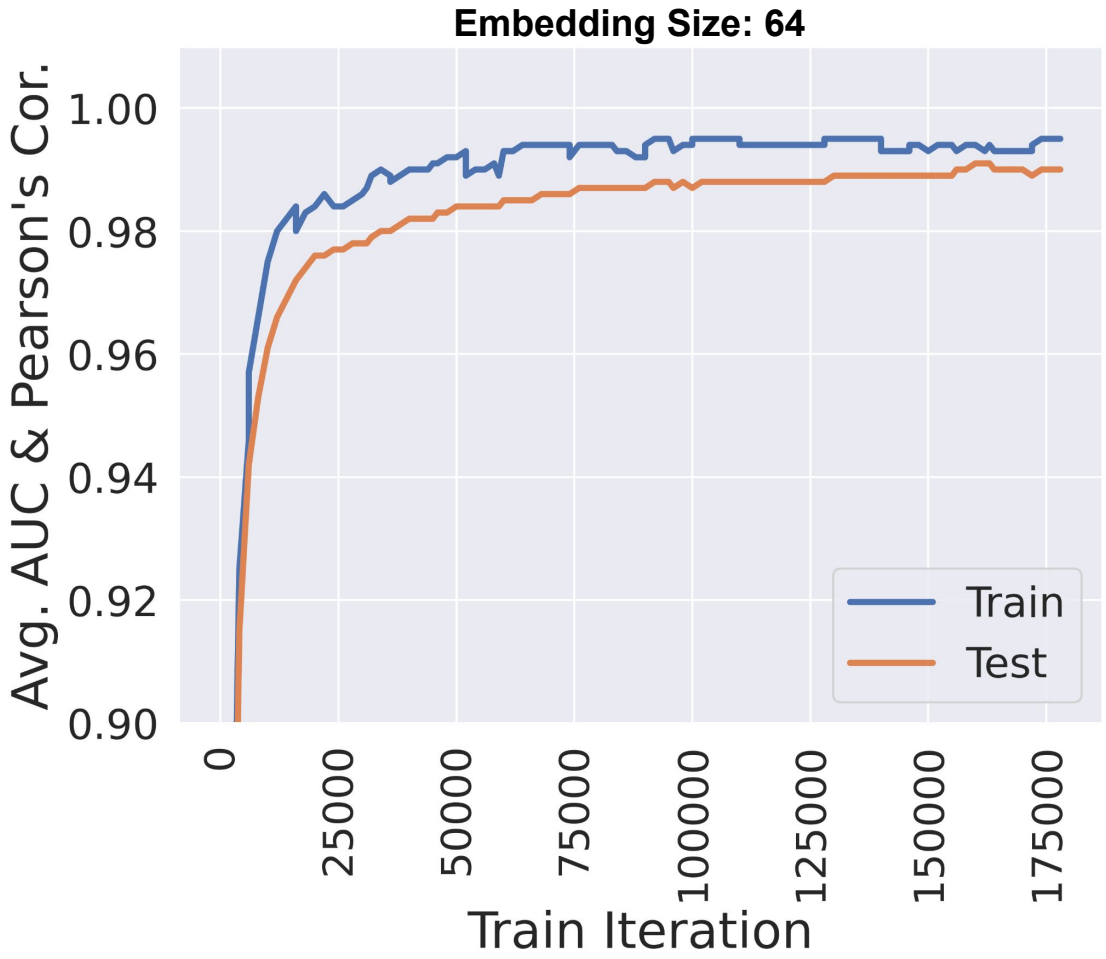
Pre-training Data

Millions of diverse compounds from public and commercial sources

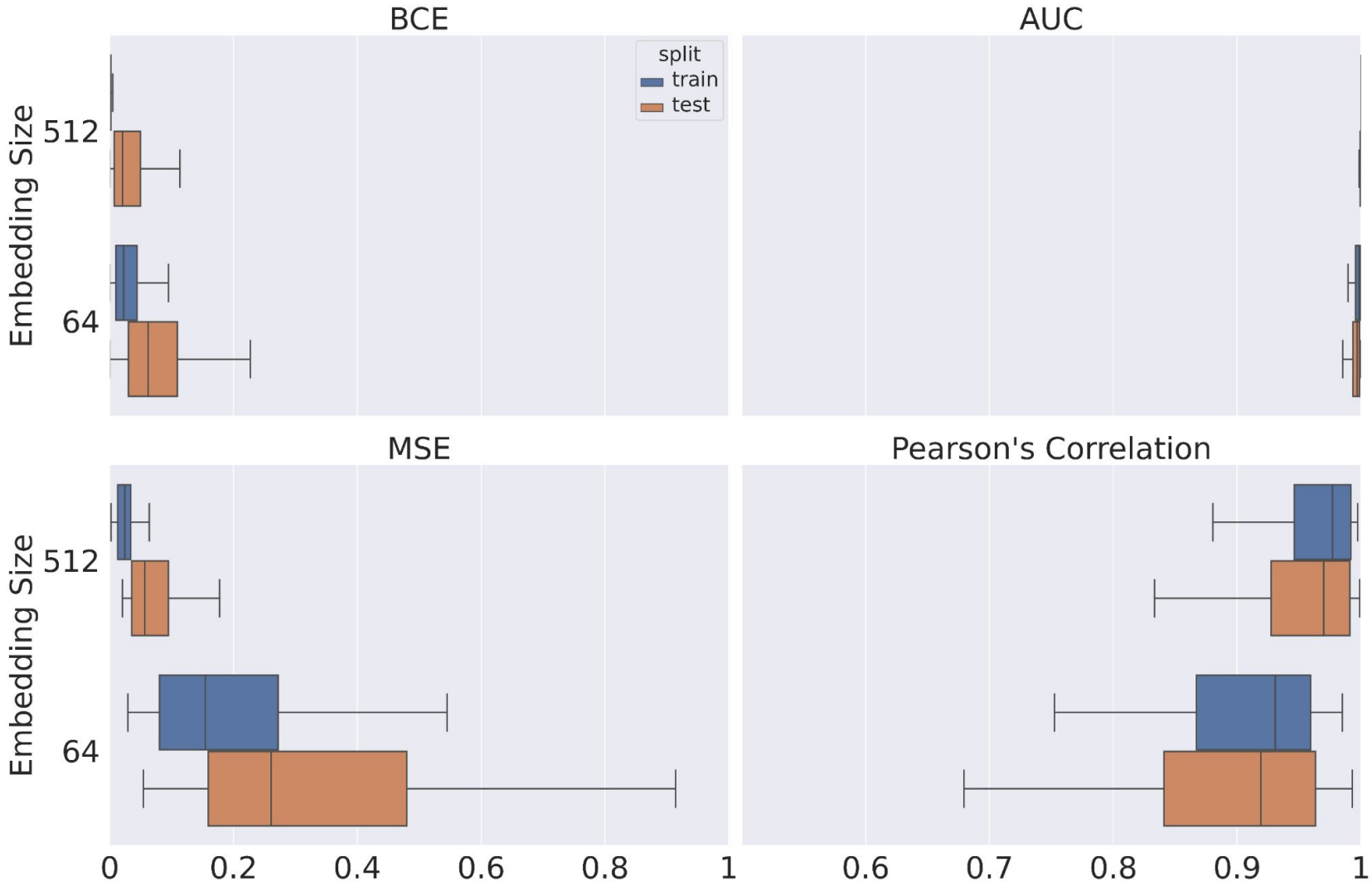


Can a GNN learn thousands of tasks?



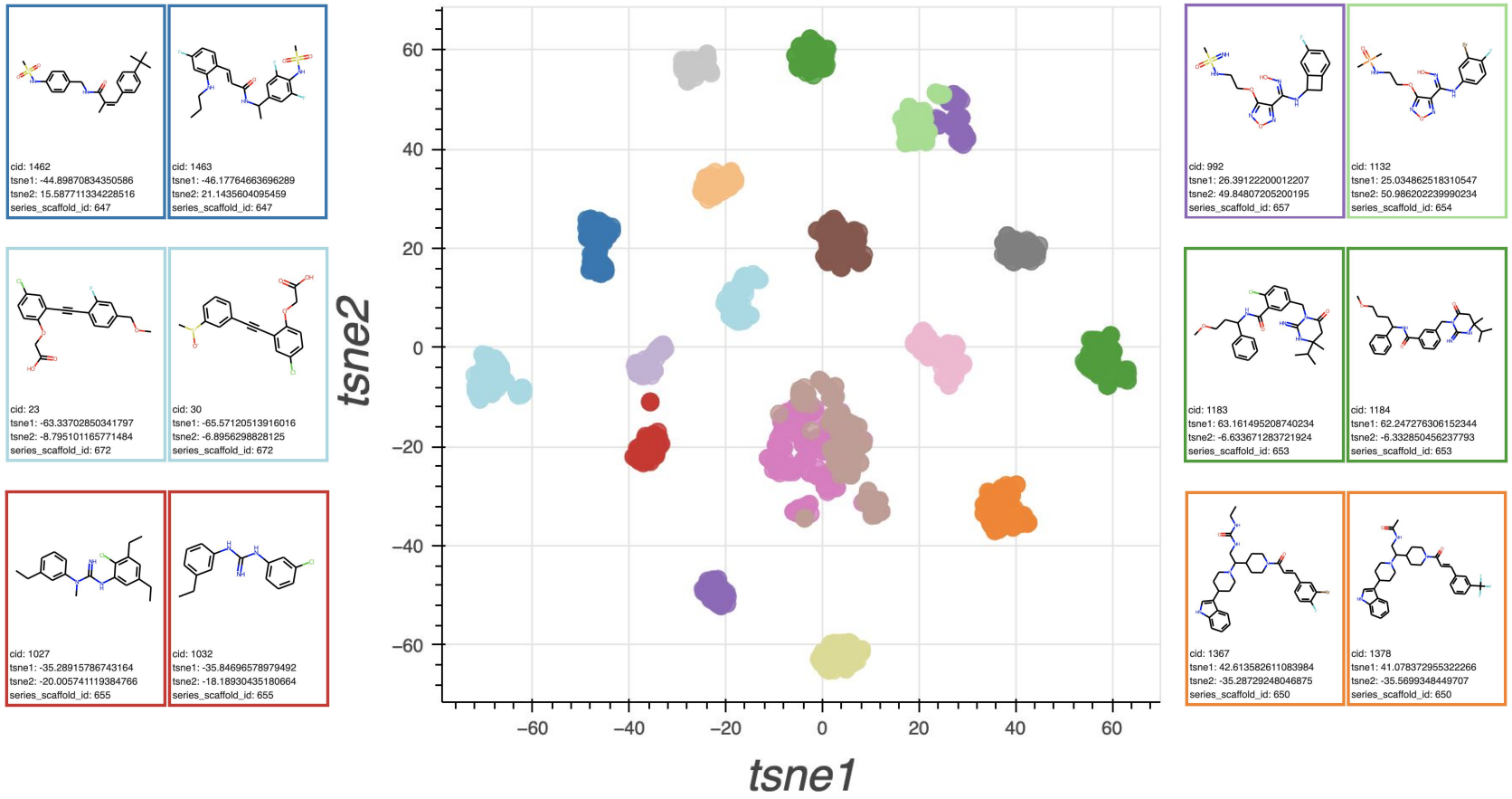


Can a GNN learn thousands of tasks?

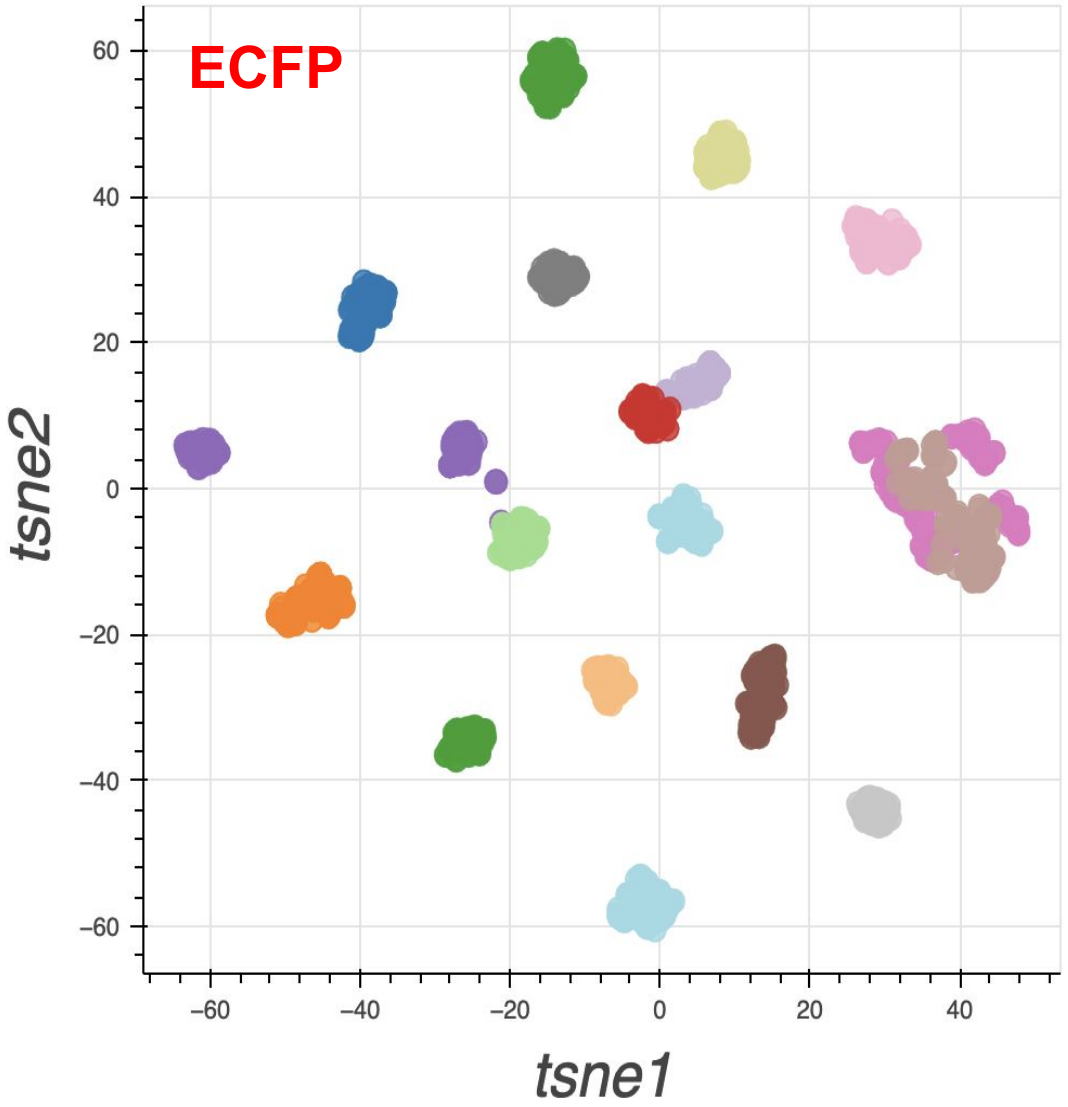


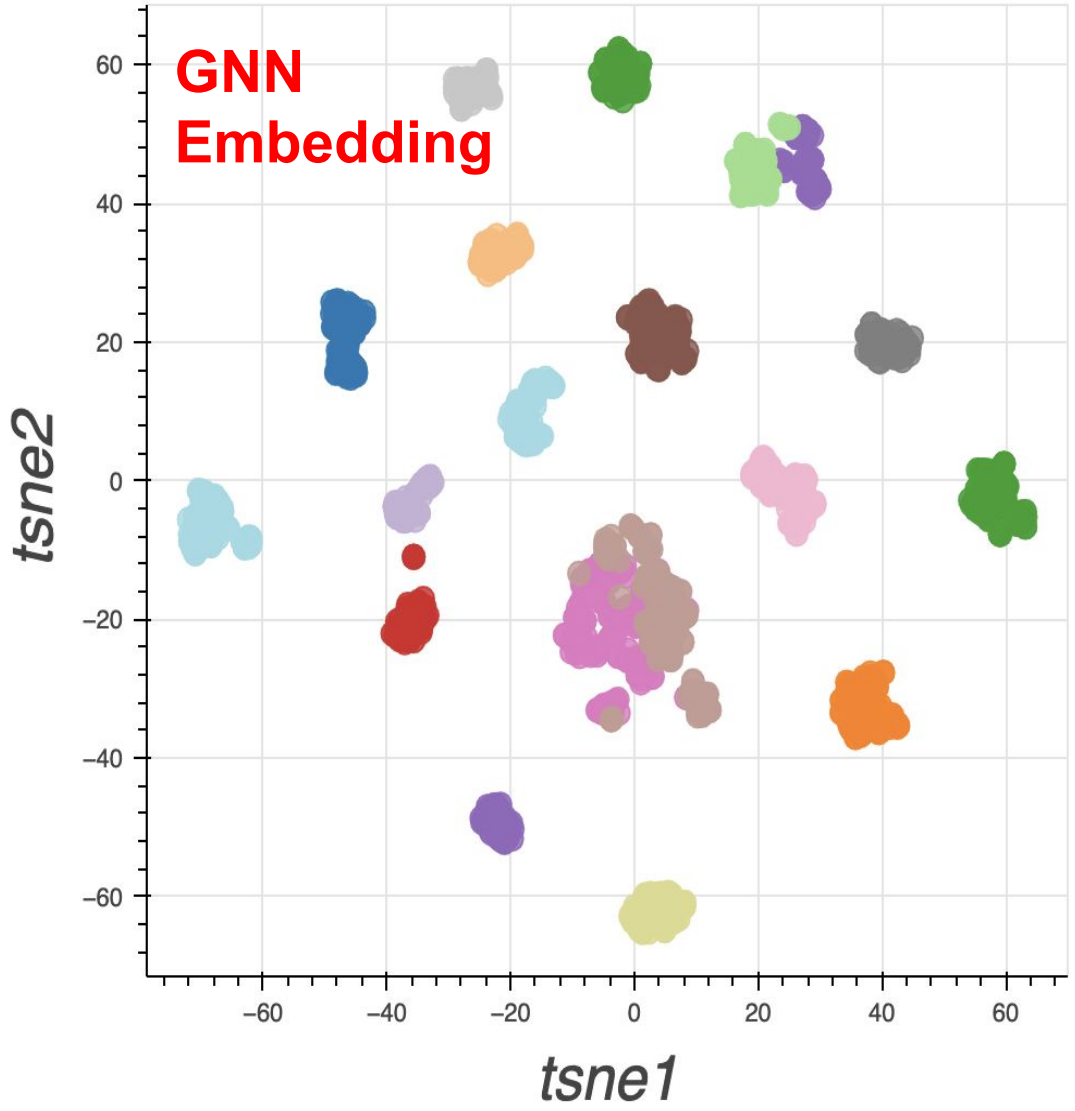
Regression labels are standardized

Clustering with the learned embeddings

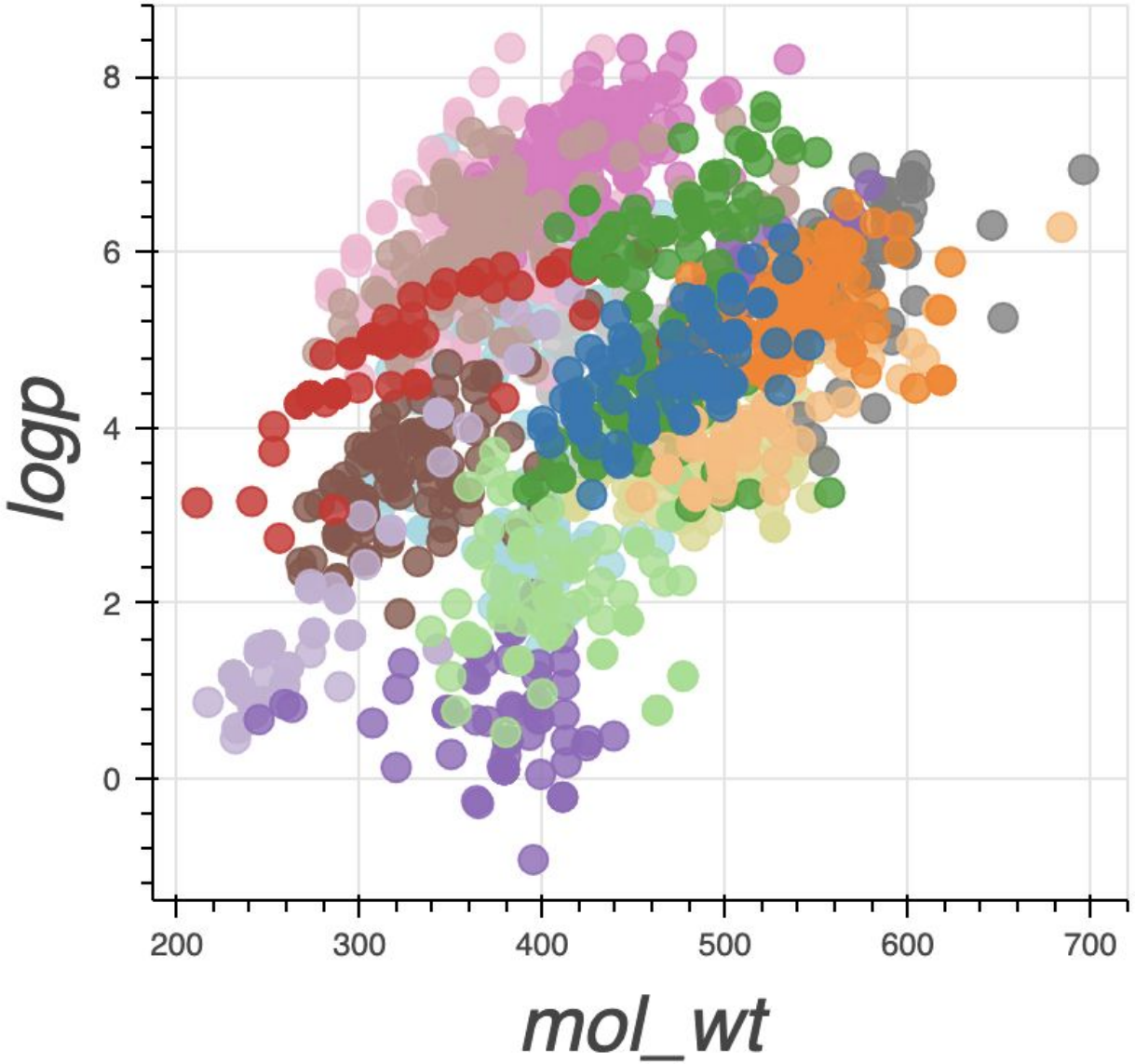


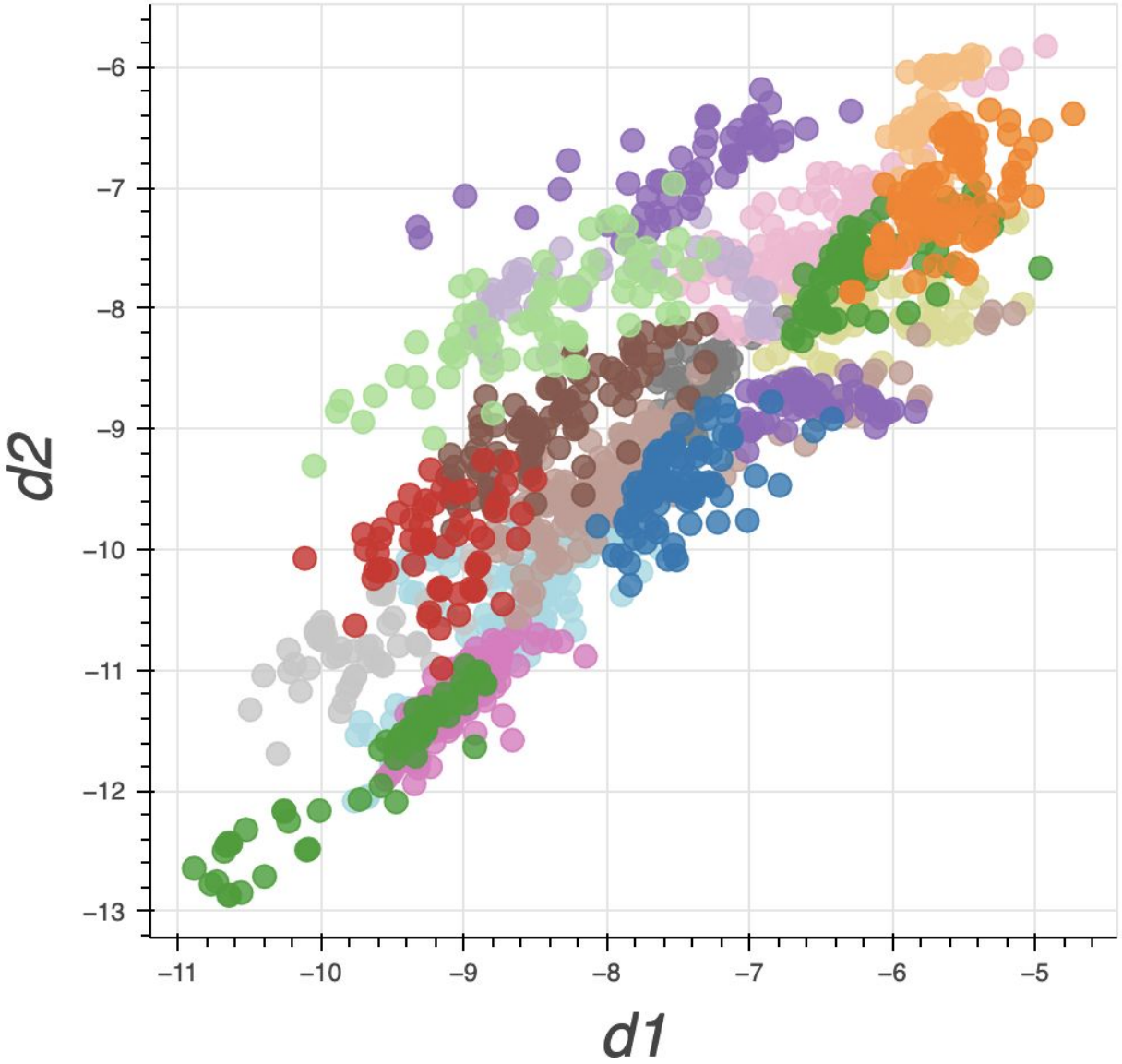
Clustering: ECFP vs. GNN Embedding



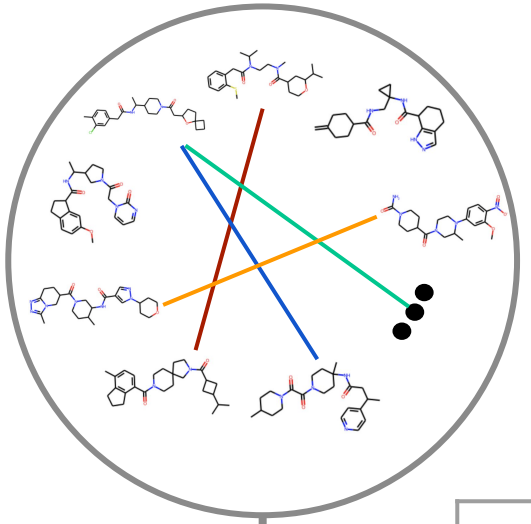


2D Embedding





How different embedding methods compare?

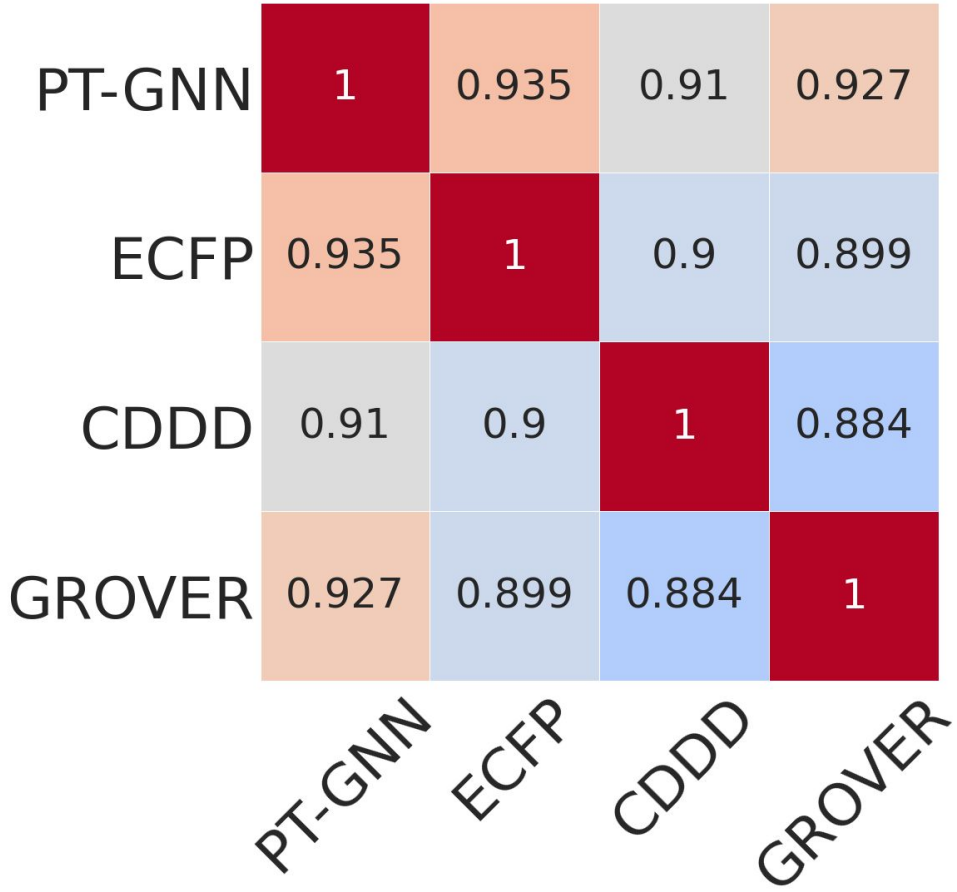


Pairwise distances

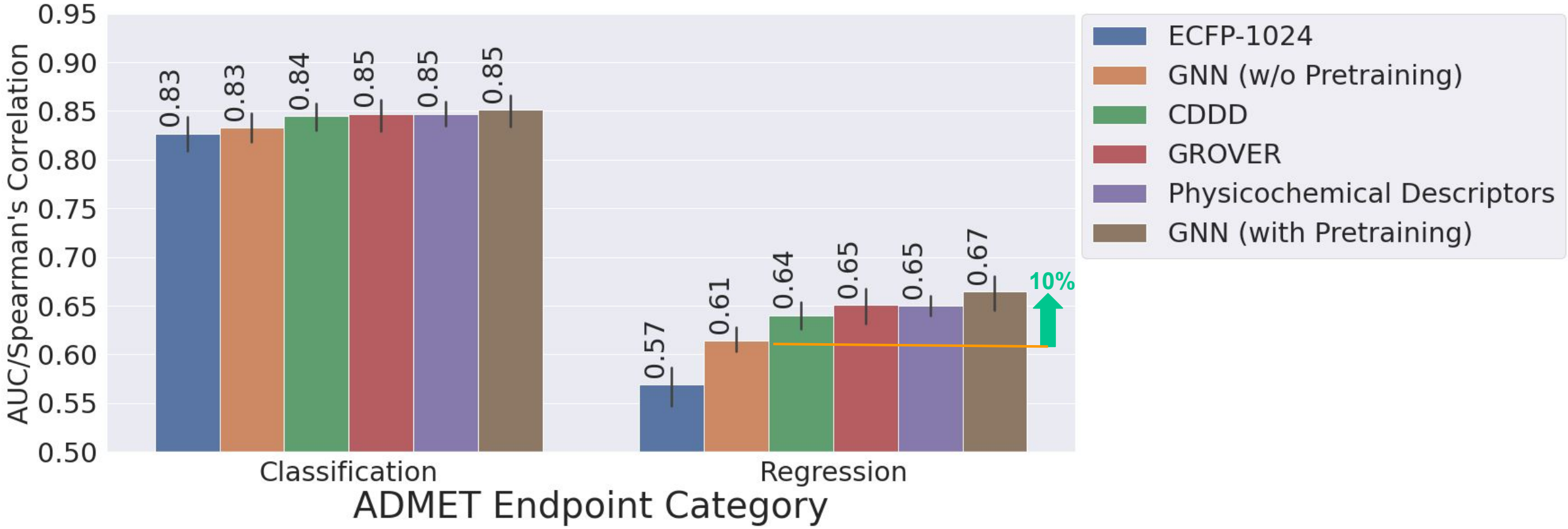
ECFP	PT-GNN	CDDD	GROVER
0.51	0.42	0.49	0.38
0.44	0.33	0.40	0.39
0.60	0.51	0.70	0.54
...	...	...	...
0.30	0.23	0.19	0.25

All-to-all correlation

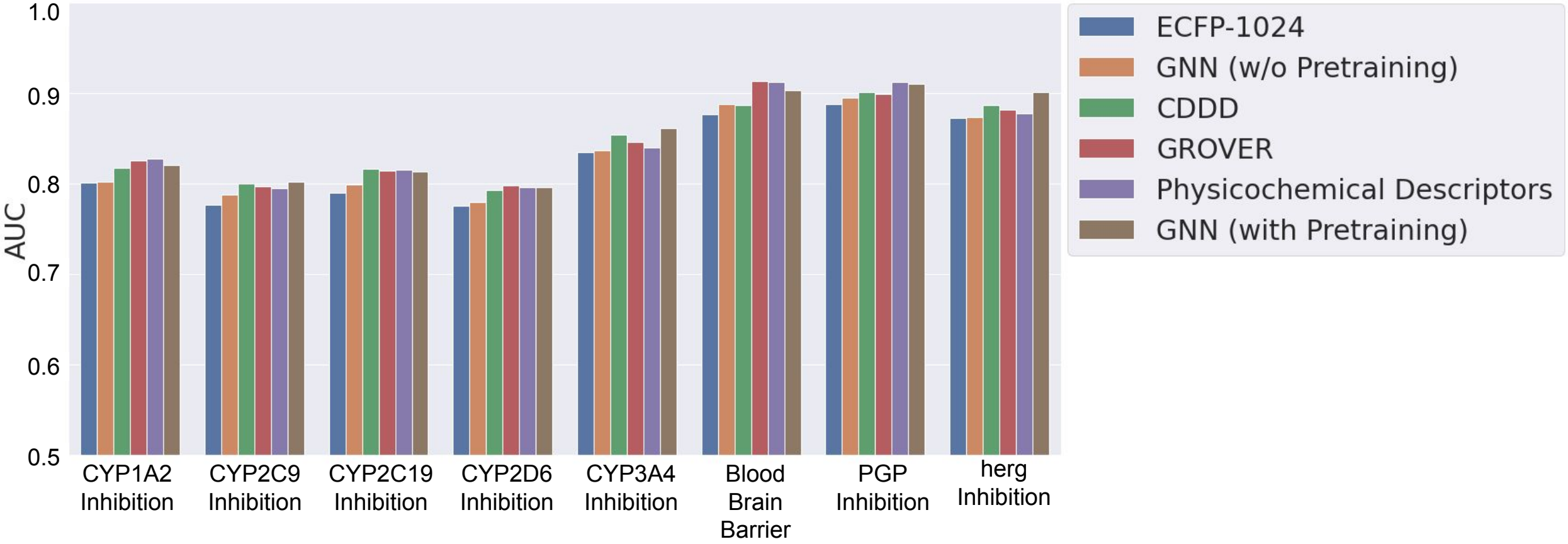
column-wise



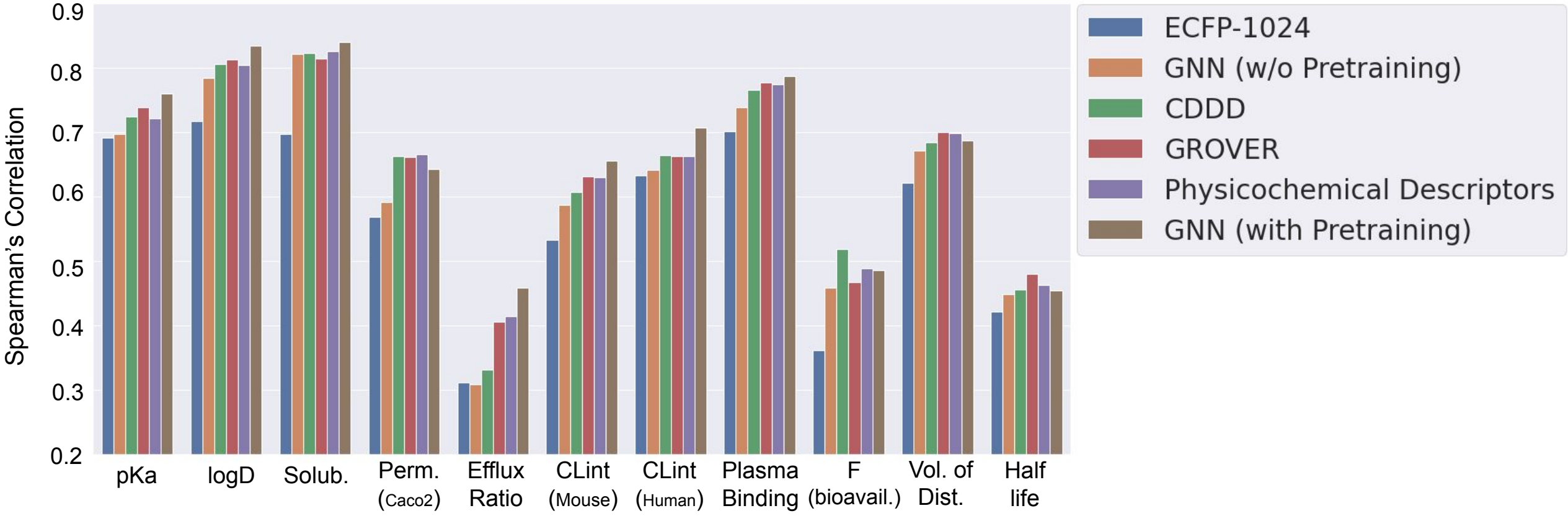
Average Performance on 20 ADMET Endpoints



Ranking Performance: Classification



Ranking Performance: Regression



Key takeaways

- Pre-training tasks align with human intuition
- Learning them forces the model to generate meaningful embeddings
- Generated embeddings are predictive for ADME
- Simple & effective

Acknowledgments

In collaboration with:

Brandon Anderson

Jon Sorenson

Izhar Wallach

and the Atomwise team!

Our Other Talks & Posters

https://info.atomwise.com/acs_fall2021

Join Us!

https://www.atomwise.com/careers/