



中国科学院自动化研究所
模式识别实验室
New Laboratory of Pattern Recognition



中国科学院自动化研究所
Institute of Automation
Chinese Academy of Sciences



[NeurIPS 2024]

Pin-Tuning: Parameter-Efficient In-Context Tuning for Few-Shot Molecular Property Prediction

Liang Wang^{1,2}, Qiang Liu^{1,2,†}, Shaozhen Liu¹, Xin Sun³, Shu Wu^{1,2}, Liang Wang^{1,2,3}

¹Institute of Automation, Chinese Academy of Sciences

²University of Chinese Academy of Sciences

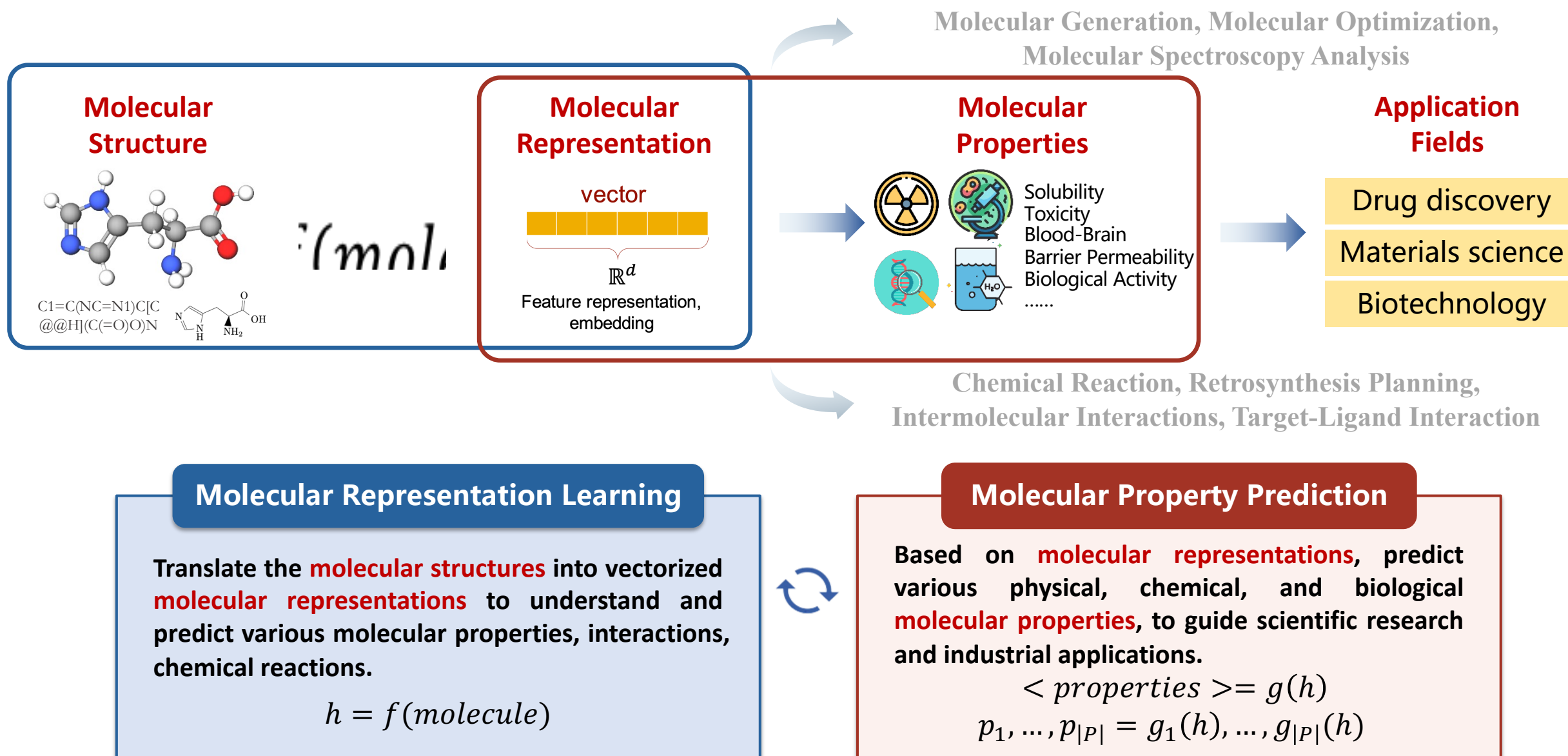
³University of Science and Technology of China

[†]Correspondence



12 November 2024

Research Background



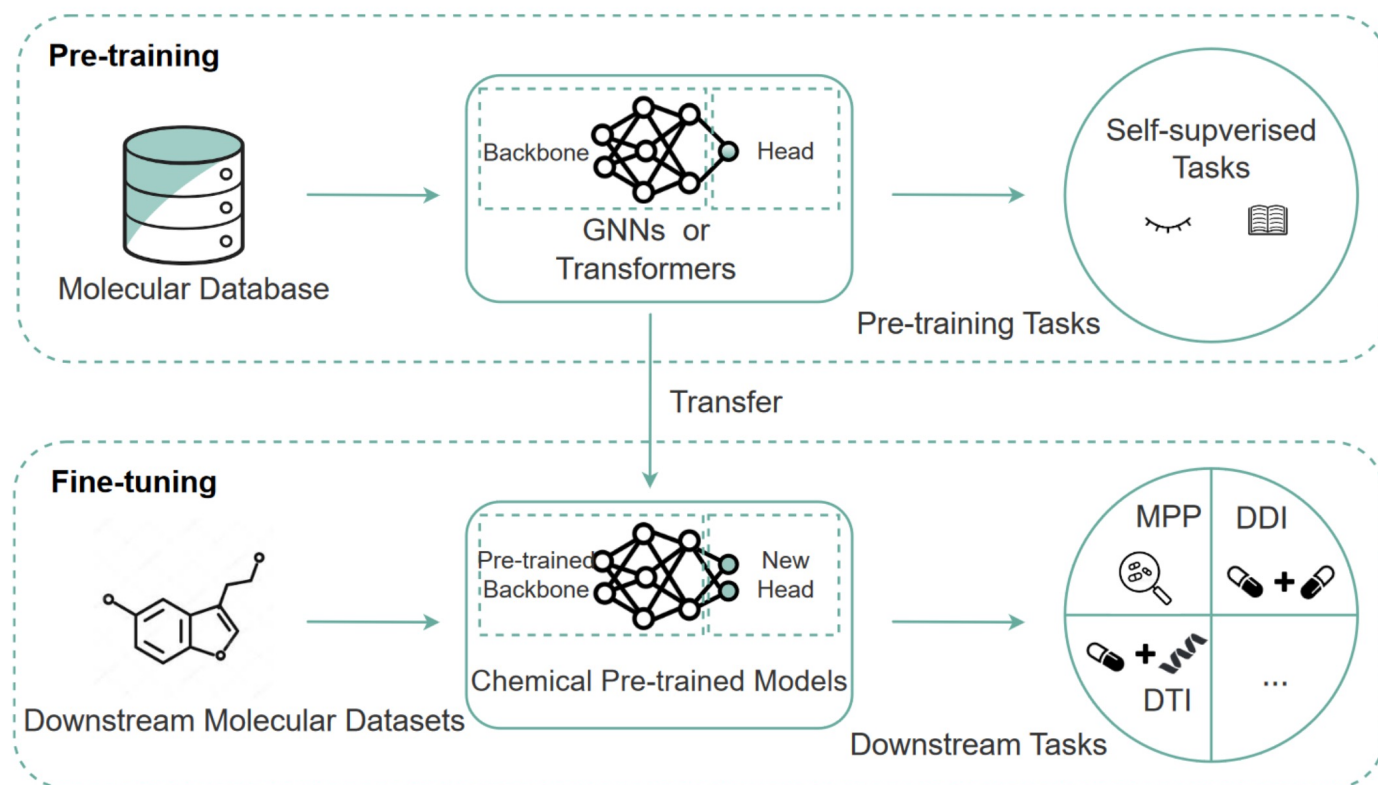
Research Background

Challenges of supervised molecular representation learning

(1) Scarcity of labeled data.

(2) Poor out-of-distribution generalization capability.

Pipeline of Molecular Representation Pre-training



✓ **Pre-trained** on large-scale unlabeled molecules.

✓ **Fine-tuned** on various downstream tasks.

Few-Shot Molecular Property Prediction

Key Elements underlying Molecular Property Prediction

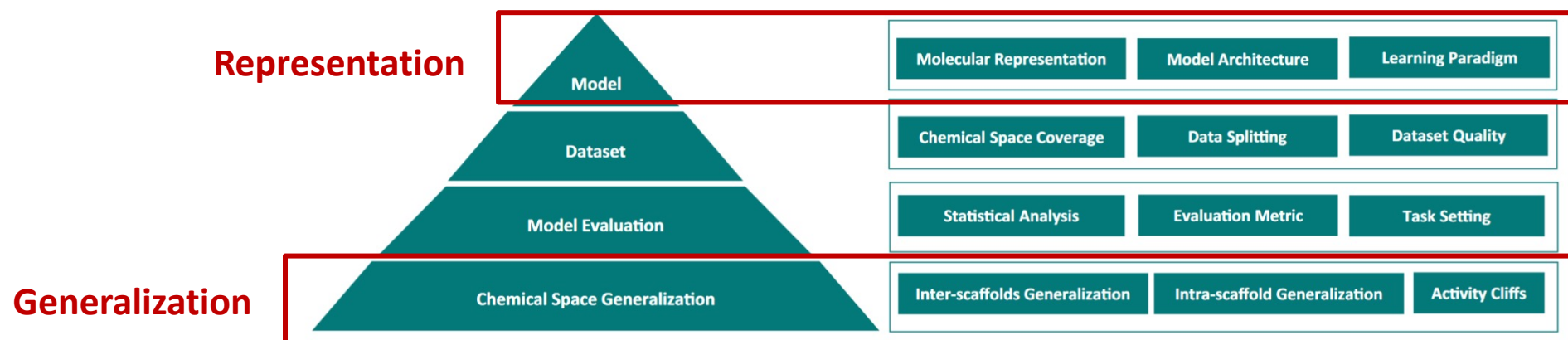
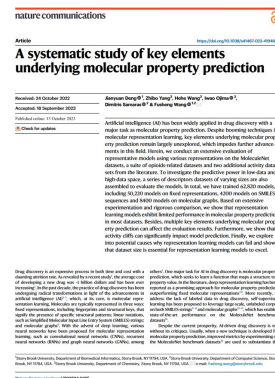


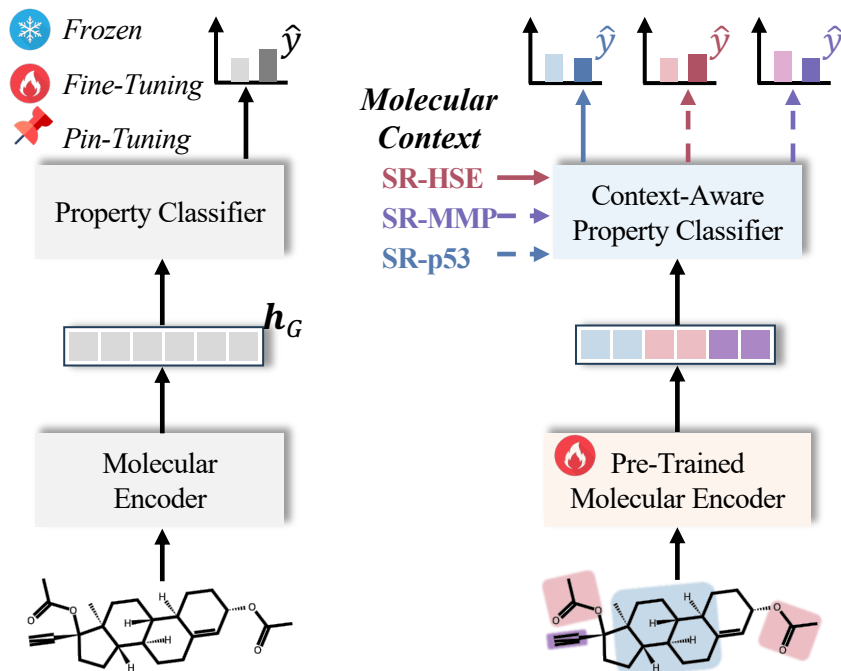
Fig. 1 | Key elements underlying molecular property prediction. There are four aspects involved: model, dataset, model evaluation, and generalization. Currently, the focus is more on the model, which aims at developing novel learning paradigms or model architectures on certain molecular representations. However, it is also necessary to consider other crucial elements, pertaining to (1) what the model is built upon, (2) how the model is evaluated, and (3) eventually what the model is capable of. For the dataset, its chemical space coverage (w.r.t. both structures and

labels), and scrutiny of its quality, including dataset size and label accuracy (e.g., duplicates, contradictories, and noise), as well as data splitting, is essential before developing a model for a specific property prediction task. For the model evaluation, thoughtful consideration of statistical analysis, evaluation metrics, and task settings is critical as they impact the observed prediction performance. For the chemical space generalization, it is important to clarify the model's applicability and if the activity-cliffs issue is addressed.



[1] “A Systematic Study of Key Elements Underlying Molecular Property Prediction.” *Nature Communications*, 2023

Few-Shot Molecular Property Prediction



(a) Vanilla MPP framework.

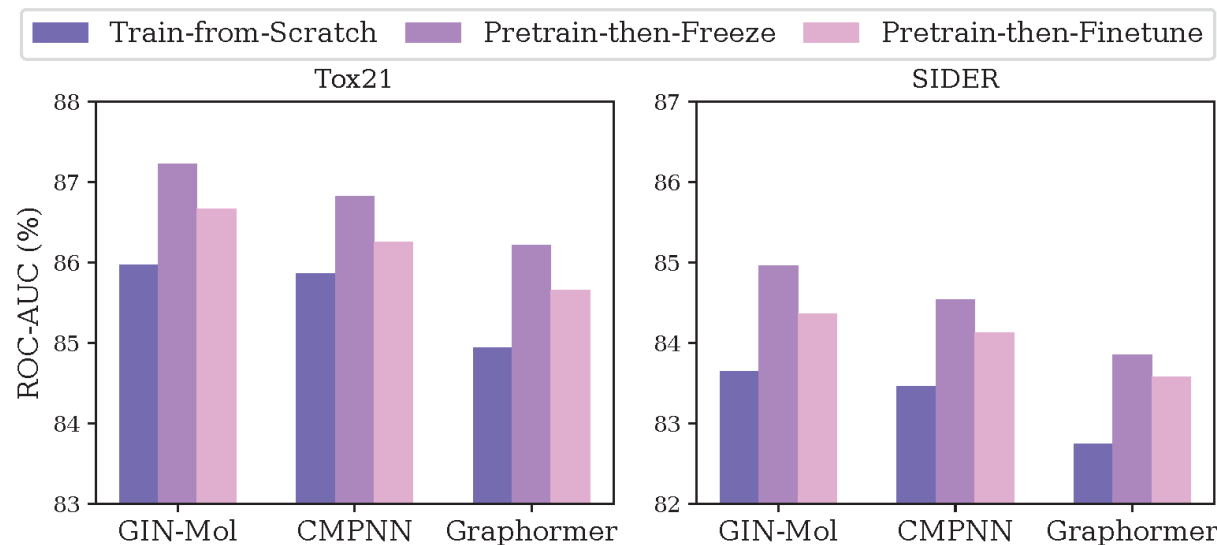
(b) Existing FSMPP framework.

Observation

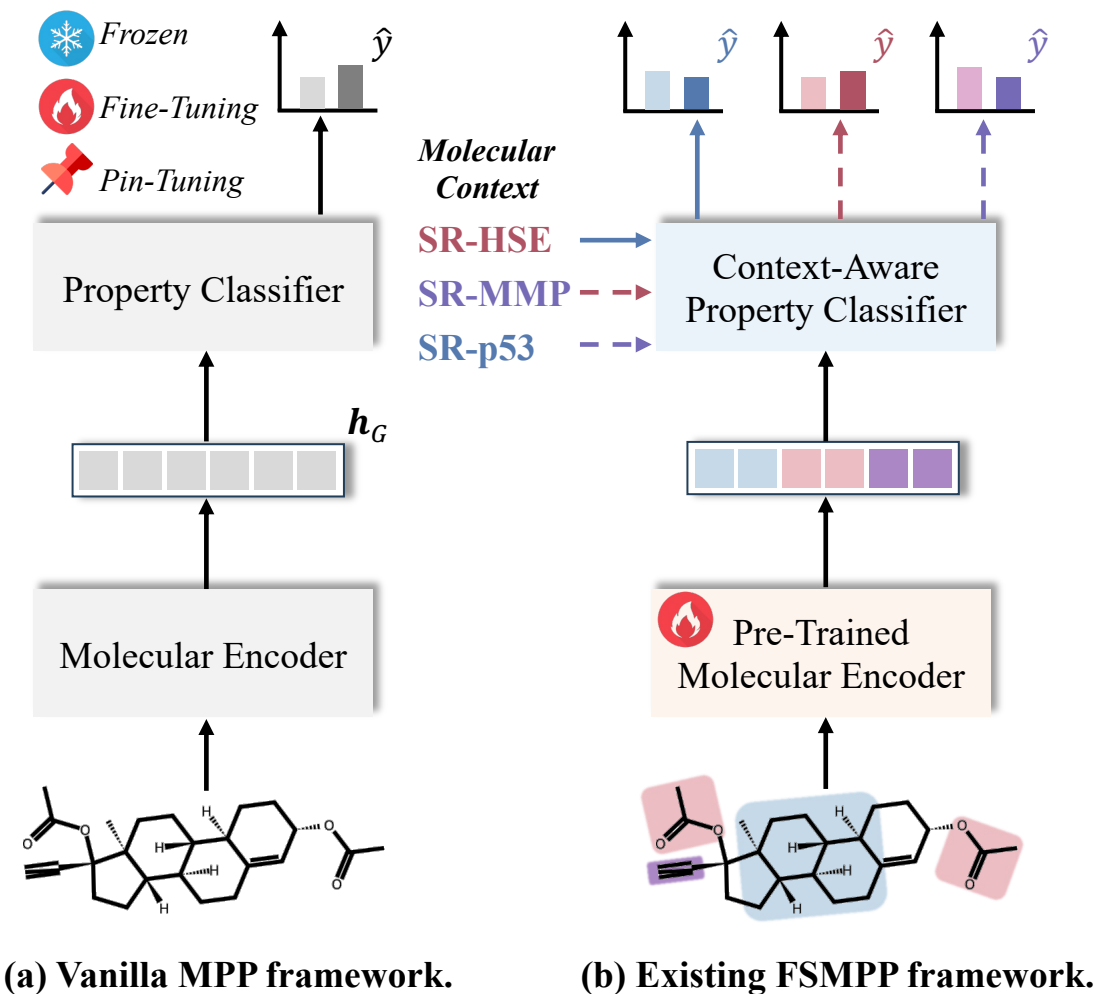
Train-from-Scratch < Pretrain-then-Finetune ≤ Pretrain-then-Freeze

Pre-training is effective, but fine-tuning is ineffective.

How to adapt molecular pre-trained models to downstream tasks, especially in few-shot scenarios?



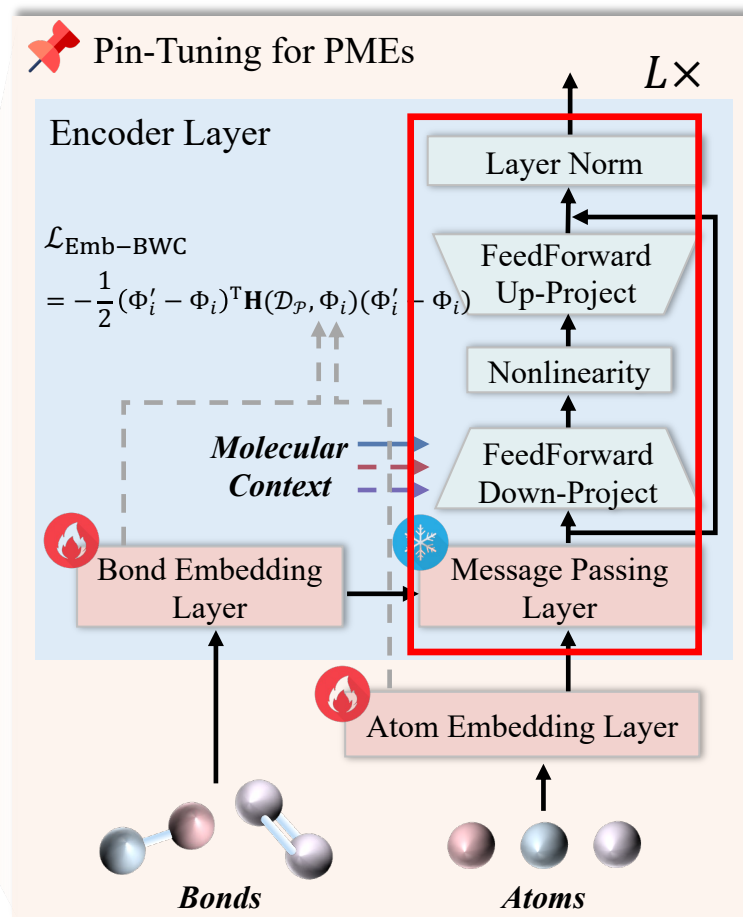
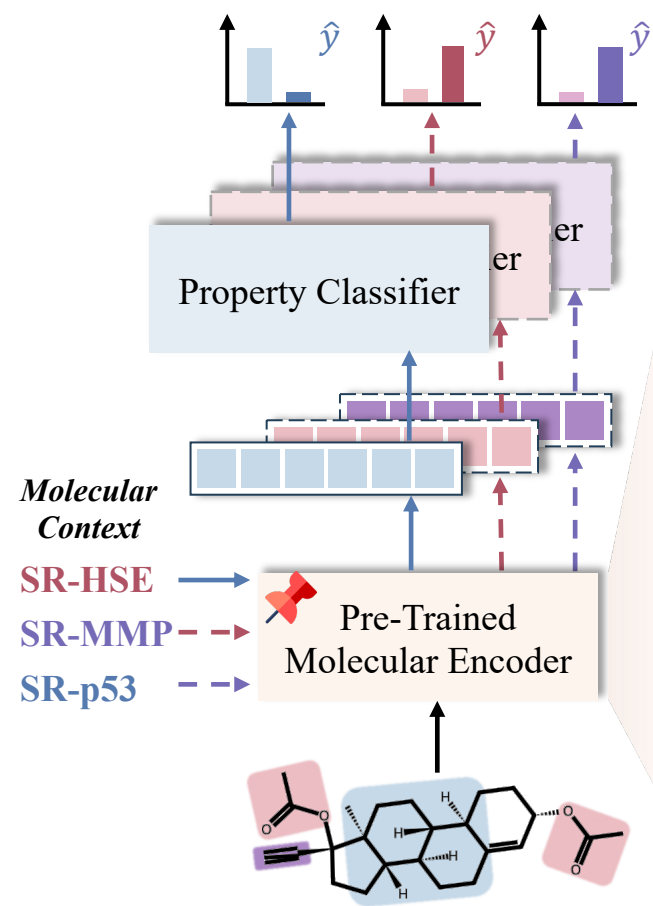
Few-Shot Molecular Property Prediction



Reasons:

1. **Imbalance between the abundance of tunable parameters and the scarcity of labeled molecules.**
2. **Limited contextual perceptiveness in the encoder.**

Pin-Tuning: Parameter-Efficient In-Context Tuning for Few-Shot Molecular Property Prediction



MP-Adapter: message passing layer-oriented adapter

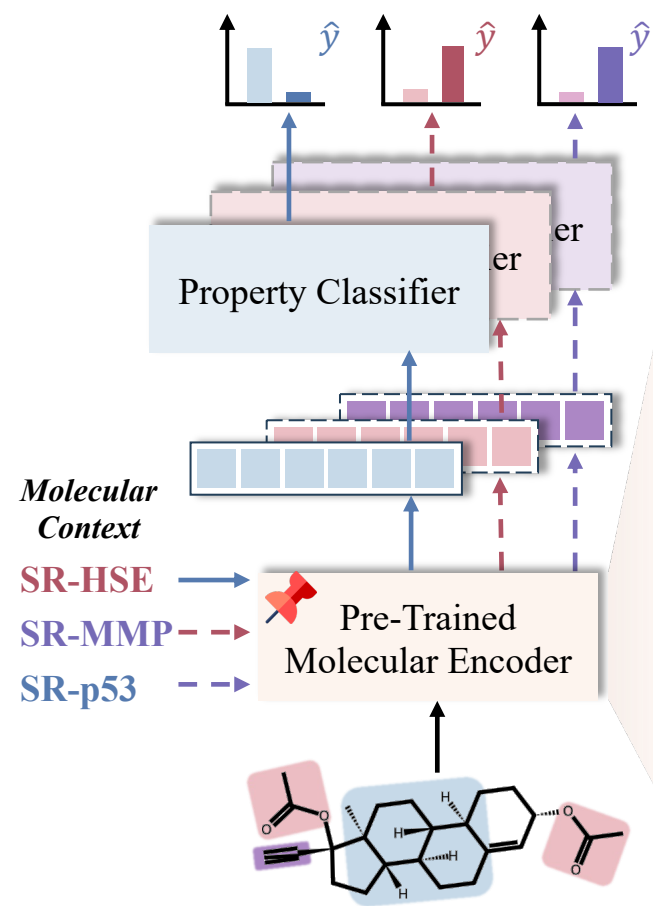
$$z_v^{(l)} = \text{FeedForward}_{\text{down}}(h_v^{(l)}) \in \mathbb{R}^{d_2},$$

$$\Delta h_v^{(l)} = \text{FeedForward}_{\text{up}}(\phi(z_v^{(l)})) \in \mathbb{R}^d,$$

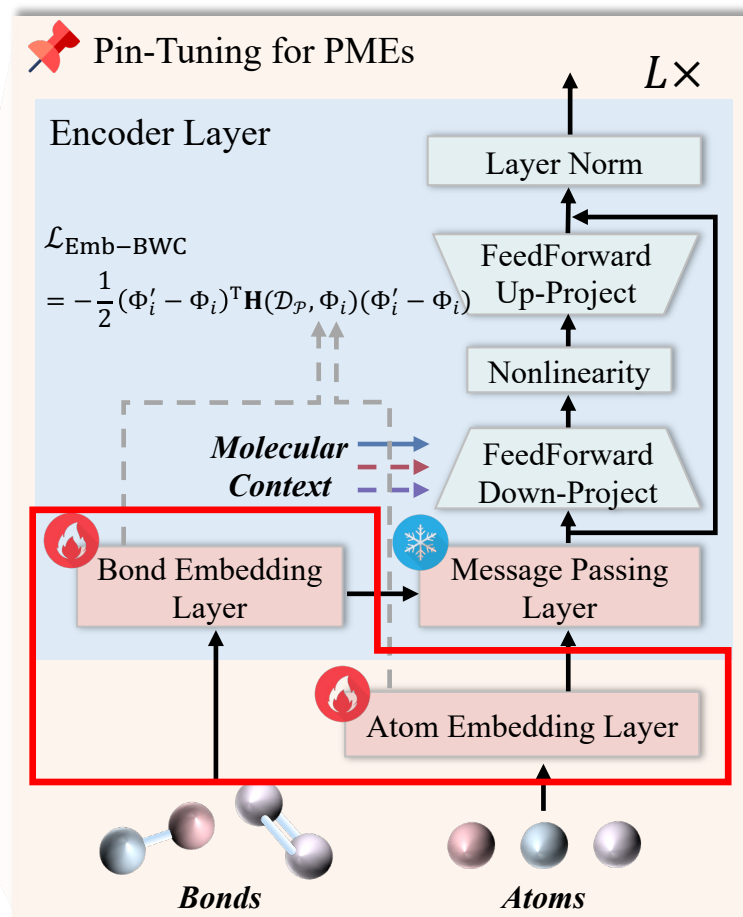
$$\tilde{h}_v^{(l)} = \text{LayerNorm}(h_v^{(l)} + \Delta h_v^{(l)}) \in \mathbb{R}^d,$$

- *Bottleneck*
- *Near-zero initialization*
- *Skip-connection*

Pin-Tuning: Parameter-Efficient In-Context Tuning for Few-Shot Molecular Property Prediction



Our FSMPP framework.



Our Pin-Tuning method for PME.

Emb-BWC: embedding layer-oriented Bayesian weight consolidation

$$\mathcal{L}_{\text{Emb-BWC}} = -\frac{1}{2} \sum_{i=1}^E (\Phi'_i - \Phi_i)^T \mathbf{H}(\mathcal{D}_P, \Phi_i) (\Phi'_i - \Phi_i),$$

- *Maximum a posteriori (MAP) estimation*
- *Bayesian learning theory*
- *Second-order Taylor expansion*

Three choices of diagonal approximation of Hessian

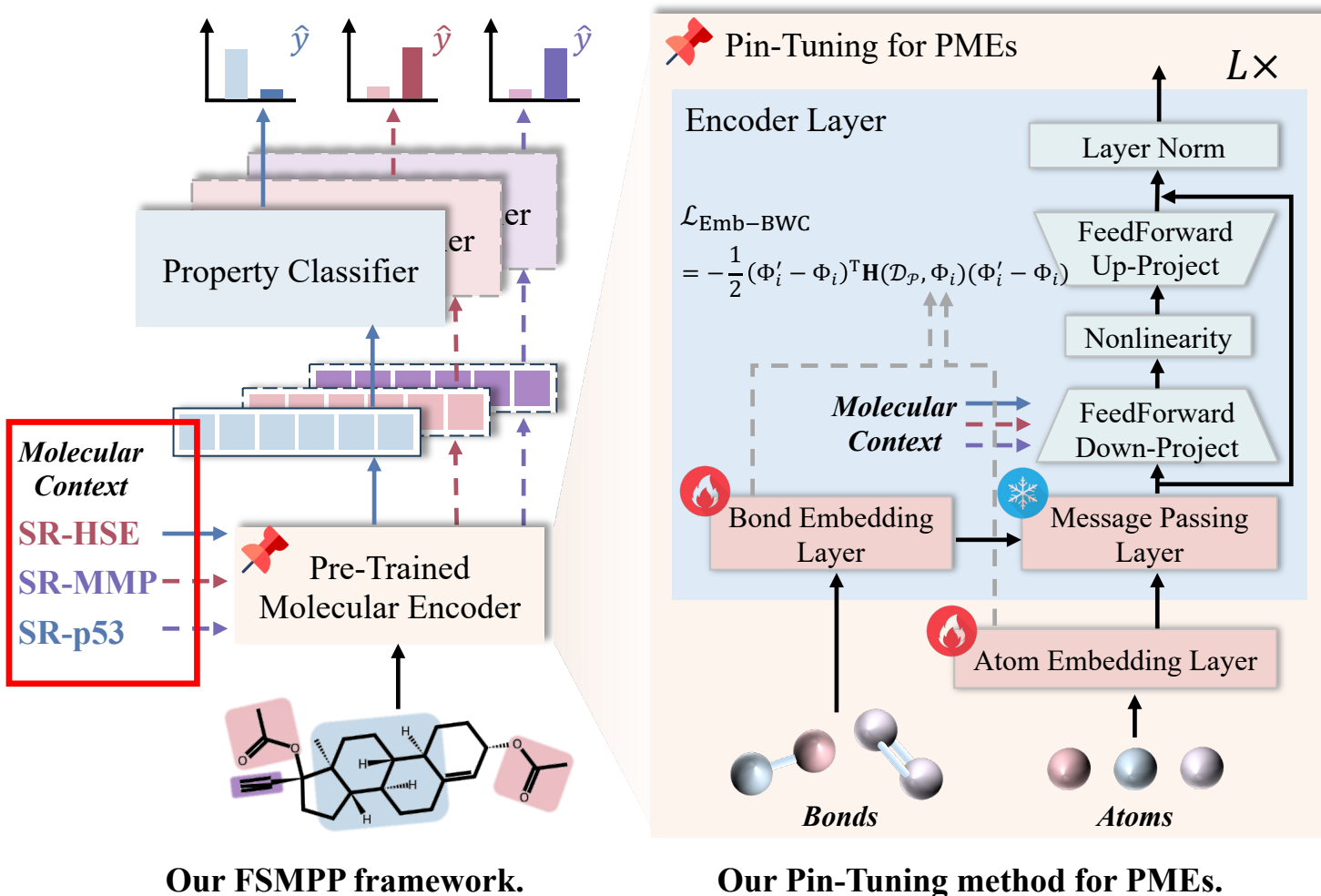
$$\mathcal{L}_{\text{Emb-BWC}}^{\text{IM}} = \frac{1}{2} \sum_{i=1}^E \sum_{j=1}^d (\Phi'_{i,j} - \Phi_{i,j})^2$$

$$\mathcal{L}_{\text{Emb-BWC}}^{\text{FIM}} = \frac{1}{2} \sum_{i=1}^E \hat{\mathbf{F}}_i (\Phi'_i - \Phi_i)^2$$

$$\mathcal{L}_{\text{Emb-BWC}}^{\text{EFIM}} = \frac{1}{2} \sum_{i=1}^E \tilde{\mathbf{F}}_i (\tilde{\Phi}'_i - \tilde{\Phi}_i)^2$$

- *Identity matrix.*
- *Diagonal of Fisher information matrix.*
- *Diagonal of embedding-wise Fisher information matrix.*

Pin-Tuning: Parameter-efficient In-Context Tuning for Few-Shot Molecular Property Prediction



Enabling contextual perceptiveness in MP-Adapter

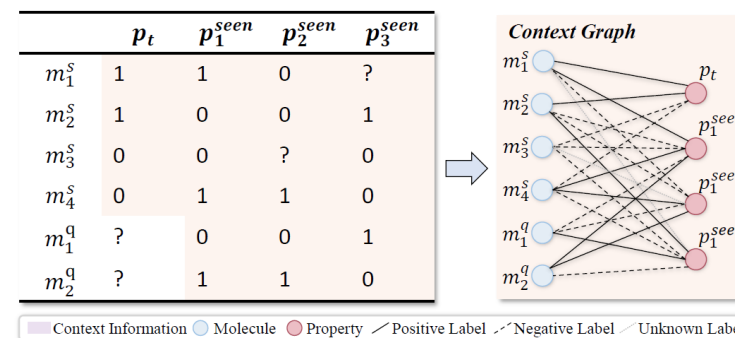


Figure 3: Convert the context information of a 2-shot episode into a context graph.

$$\mathbf{C} = \text{ContextEncoder}(\mathcal{V}_t, \mathbf{A}_t, \mathbf{X}_t)$$

$$z^{(l)} = \text{FeedForward}_{\text{down}}(h_v^{(l)} \| c_m \| c_p),$$

Experiment Results

Table 1: ROC-AUC scores (%) on benchmark datasets, compared with methods trained from scratch (first group) and methods that leverage pre-trained molecular encoder (second group). The best is marked with **boldface** and the second best is with underline. $\Delta Improve.$ indicates the relative improvements over the baseline models in percentage.

Model	Tox21		SIDER		MUV		ToxCast		PCBA	
	10-shot	5-shot	10-shot	5-shot	10-shot	5-shot	10-shot	5-shot	10-shot	5-shot
Siamese	80.40(0.35)	-	71.10(4.32)	-	59.96(5.13)	-	-	-	-	-
ProtoNet	74.98(0.32)	72.78(3.93)	64.54(0.89)	64.09(2.37)	65.88(4.11)	64.86(2.31)	68.87(0.43)	66.26(1.49)	64.93(1.94)	62.29(2.12)
MAML	80.21(0.24)	69.17(1.34)	70.43(0.76)	60.92(0.65)	63.90(2.28)	63.00(0.61)	68.30(0.59)	67.56(1.53)	66.22(1.31)	65.25(0.75)
TPN	76.05(0.24)	75.45(0.95)	67.84(0.95)	66.52(1.28)	65.22(5.82)	65.13(0.23)	69.47(0.71)	66.04(1.14)	67.61(0.33)	63.66(1.64)
EGNN	81.21(0.16)	76.80(2.62)	72.87(0.73)	60.61(1.06)	65.20(2.08)	63.46(2.58)	74.02(1.11)	67.13(0.50)	69.92(1.85)	67.71(3.67)
IterRefLSTM	81.10(0.17)	-	69.63(0.31)	-	49.56(5.12)	-	-	-	-	-
Pre-GNN	82.14(0.08)	82.04(0.30)	73.96(0.08)	76.76(0.53)	67.14(1.58)	70.23(1.40)	75.31(0.95)	74.43(0.47)	76.79(0.45)	75.27(0.49)
Meta-MGNN	82.97(0.10)	76.12(0.23)	75.43(0.21)	66.60(0.38)	68.99(1.84)	64.07(0.56)	76.27(0.56)	75.26(0.43)	72.58(0.34)	72.51(0.52)
PAR	84.93(0.11)	83.95(0.15)	78.08(0.16)	77.70(0.34)	<u>69.96</u> (1.37)	<u>68.08</u> (2.42)	79.41(0.08)	76.89(0.32)	73.71(0.61)	72.79(0.98)
GS-Meta	86.67(0.41)	<u>86.43</u> (0.02)	84.36(0.54)	84.57(0.01)	66.08(1.25)	64.50(0.20)	83.81(0.16)	82.65(0.35)	79.40(0.43)	77.47(0.29)
Pin-Tuning	91.56 (2.57)	90.95 (2.33)	93.41 (3.52)	92.02 (3.01)	73.33 (2.00)	70.71 (1.42)	84.94 (1.09)	83.71 (0.93)	81.26 (0.46)	79.23 (0.52)
$\Delta Improve.$	5.64%	5.23%	10.73%	8.81%	4.82%	3.86%	1.35%	1.28%	2.34%	2.27%

Tunable Parameter Size Analysis

$$N_{Fine-Tuning} = |E_n|d + L(|E_e|d + 2dd_1 + 3d + d_1).$$

$$N_{Pin-Tuning} = |E_n|d + L(|E_e|d + 2dd_2 + 3d + d_2).$$

$$\Delta N = (d_1 - d_2)L(2d + 1).$$

Ours (14.2% parameters, higher performance)

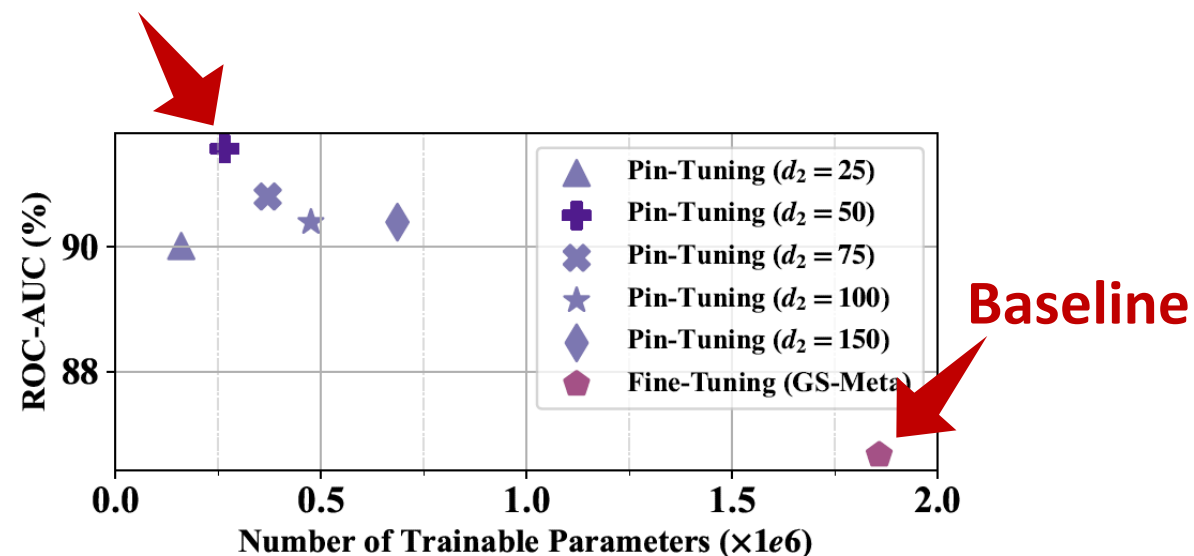


Figure 5: ROC-AUC (%) and number of trainable parameters of Pin-Tuning with varied value of d_2 and full Fine-Tuning method (e.g., GS-Meta) on the Tox21 dataset.

Visualization

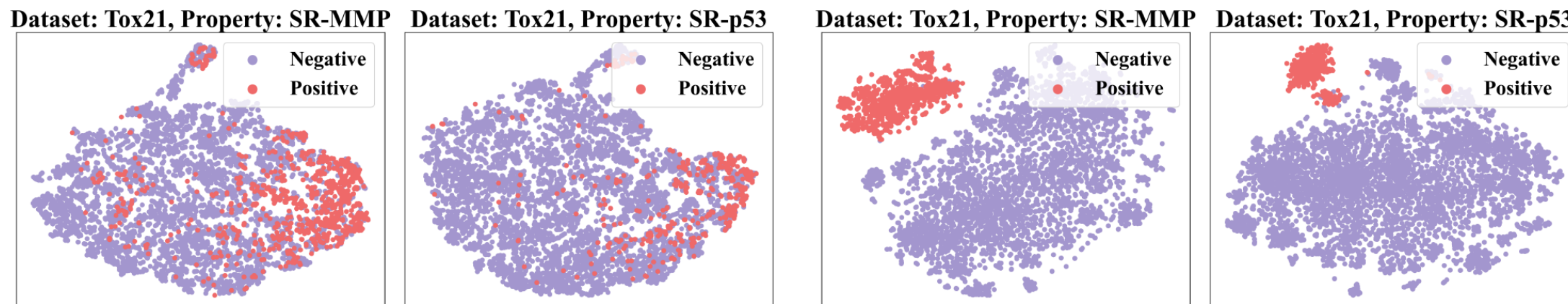


Figure 6: Molecular representations encoded by GS-Meta [58].

Figure 7: Molecular representations encoded by Pin-Tuning.



Thank you for your attention!

Contact: liang.wang@cripac.ia.ac.cn