

# Suiren-1.0 Technical Report: A Family of Molecular Foundation Models

Junyi An<sup>1</sup>, Xinyu Lu, Yun-Fei Shi, Li-Cheng Xu, Nannan Zhang, Chao Qu, Yuan Qi, Fenglei Cao

Shanghai Academy of AI for Science (SAIS), Golab

## Abstract

We introduce Suiren-1.0, a family of molecular foundation models for the accurate modeling of diverse organic systems. Suiren-1.0 comprising three specialized variants (Suiren-Base, Suiren-Dimer, and Suiren-ConfAvg) is integrated within an algorithmic framework that bridges the gap between 3D conformational geometry and 2D statistical ensemble spaces. We first pre-train Suiren-Base (1.8B parameters) on a 70M-sample Density Functional Theory dataset using spatial self-supervision and SE(3)-equivariant architectures, achieving robust performance in quantum property prediction. Suiren-Dimer extends this capability through continued pre-training on 13.5M intermolecular interaction samples. To enable efficient downstream application, we propose Conformation Compression Distillation (CCD), a diffusion-based framework that distills complex 3D structural representations into 2D conformation-averaged representations. This yields the lightweight Suiren-ConfAvg, which generates high-fidelity representations from SMILES or molecular graphs. Our extensive evaluations demonstrate that Suiren-1.0 establishes state-of-the-art results across a range of tasks. All models and benchmarks are open-sourced.

## Model Links and Resources

- **Suiren-Base and Suiren-Dimer Codes:**  
[github.com/golab-ai/Suiren-Foundation-Model](https://github.com/golab-ai/Suiren-Foundation-Model)
- **Suiren-ConfAvg and Finetune Codes:**  
[github.com/golab-ai/Suiren-Property-Prediction](https://github.com/golab-ai/Suiren-Property-Prediction)
- **Suiren-1.0 Model Weights:**  
[huggingface.co/ajy112/Suiren-Base](https://huggingface.co/ajy112/Suiren-Base)  
[huggingface.co/ajy112/Suiren-Dimer](https://huggingface.co/ajy112/Suiren-Dimer)  
[huggingface.co/ajy112/Suiren-ConfAvg](https://huggingface.co/ajy112/Suiren-ConfAvg)
- **Finetune Model Weights and Agent Skills:**  
[modelscope.cn/models/ajy112/Suiren-Model-Set](https://modelscope.cn/models/ajy112/Suiren-Model-Set)  
[github.com/golab-ai/Huntianling](https://github.com/golab-ai/Huntianling)
- **MoleHB Benchmark:**  
[modelscope.cn/datasets/ajy112/MoleHD](https://modelscope.cn/datasets/ajy112/MoleHD)

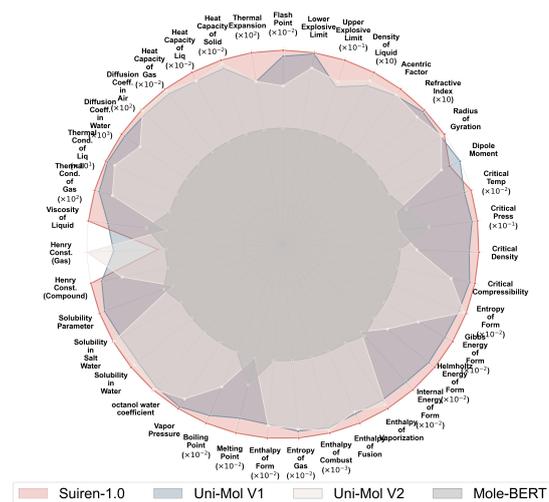
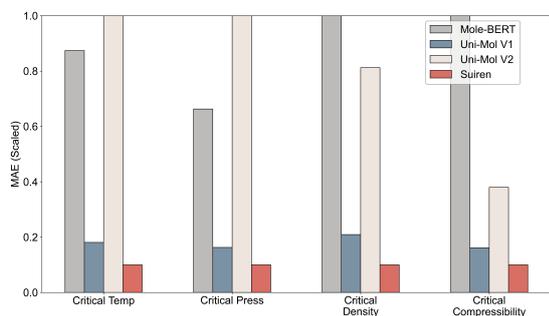
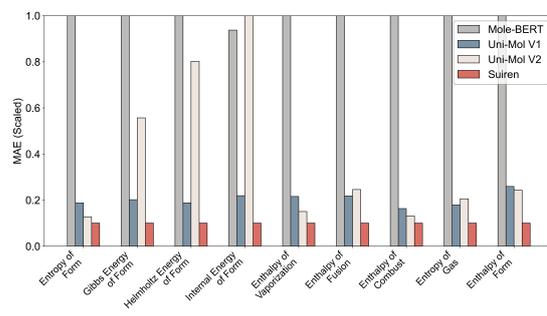


Figure 1 | Benchmark performance of Suiren-1.0 and its counterparts. We use the normalized MAE scores ( $\uparrow$ ).

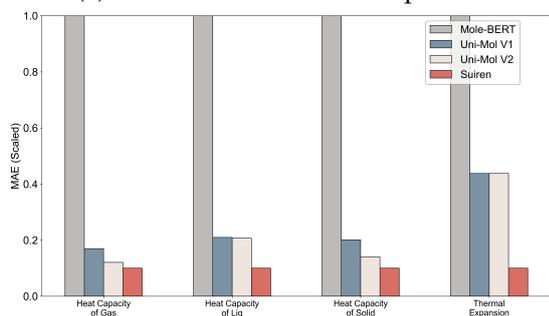
<sup>1</sup>e-mail: [anjunyi@sais.org.cn](mailto:anjunyi@sais.org.cn)



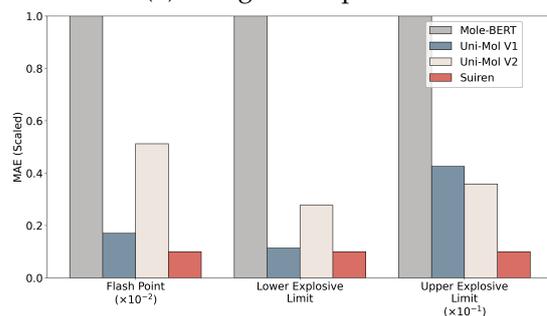
(a) Critical & Saturation Properties



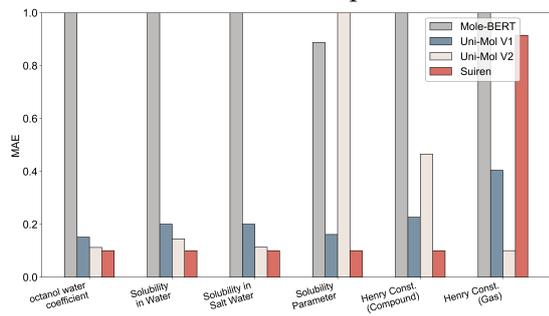
(b) Energetic Properties



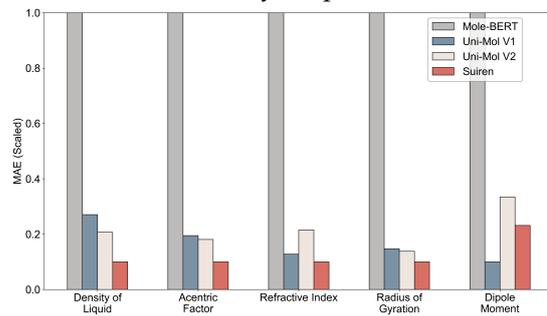
(c) Fluctuation Properties



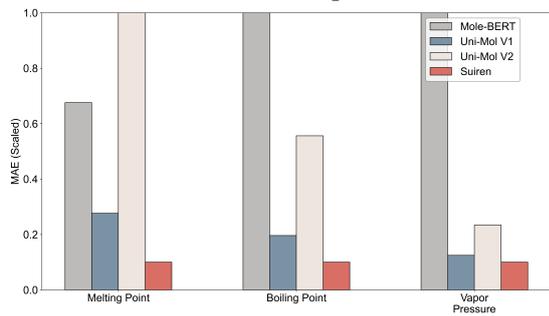
(d) Safety Properties



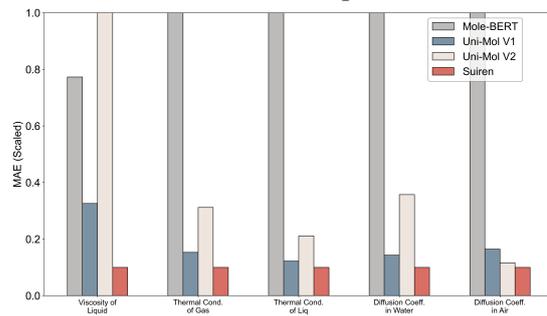
(e) Solution Properties



(f) Structural Properties



(g) Thermal Properties



(h) Transport Properties

Figure 2 | Comparison of SuiRen-1.0 model and molecular Foundation model across various tasks in 8 domains. All tasks are regression tasks, with MAE ( $\downarrow$ ) as the evaluation metric. Due to significant differences in metric ranges across different tasks, the y-axis is scaled.

## 1. Introduction

Foundation models have catalyzed a paradigm shift in natural language processing and computer vision, where large-scale pre-training facilitates robust transferability across diverse downstream tasks (Achiam et al., 2023; Liu et al., 2024a; Team et al., 2023; Yang et al., 2025). In the science domain, pioneering molecular architectures such as MoleBERT (Xia et al., 2023), Uni-Mol (Ji et al., 2024; Zhou et al., 2023), and UMA (Wood et al., 2025) have demonstrated significant promise. However, compared to the linguistic and visual domains, universal molecular modeling remains hindered by inherent scientific complexities and a scarcity of high-quality supervised data. We identify the primary challenges as follows:

- First, the "physical priors" governing molecular systems are exceptionally complex. Molecular behavior is dictated by the intricate laws, such as quantum mechanics (e.g., the Schrödinger equation) and statistical thermodynamics (e.g., Boltzmann distributions) (Charbonneau et al., 1982; Schleich et al., 2013). Capturing these fundamental mechanisms solely through data-driven learning is challenging, particularly given the sparsity of high-fidelity labeled data.
- Second, a persistent multiscale gap remains between microscopic structures and macroscopic observables. Microscopic tasks typically demand the resolution of explicit 3D conformations and electronic densities, where Density Functional Theory (DFT) enables the generation of abundant, high-quality labeled data (Chanussot et al., 2021; Levine et al., 2025; Liu et al., 2024b). Conversely, macroscopic tasks often rely on 1D SMILES or 2D molecular graphs that lack explicit conformational information. While these tasks span broad chemical spaces, their data is often scarce, as macroscopic labels frequently require costly wet-lab experiments or molecular dynamics simulations. Physically, these two modalities are intrinsically linked: macroscopic features emerge from the ensemble-averaged properties of a molecule’s conformations, governed by the Boltzmann distribution (see Figure 3). However, existing approaches largely fail to bridge this divide. Pure 3D foundation models, such as UMA, learn rich 3D representations from labeled data but lack generalizability across broad chemical tasks; meanwhile, pure 2D models, such as Mole-BERT, capture graph topology through self-supervised learning yet remain "conformation-blind," limiting their predictive effectiveness.

Suiren-1.0 is designed to bridge the multiscale gap between microscopic and macroscopic representations. We first pre-train Suiren-Base (1.8B parameters) on large-scale, first-principles quantum chemical data using objectives specifically tailored for microscopic, conformation-aware representation learning. We then introduce Conformation Compression Distillation (CCD), a diffusion-based strategy that distills the knowledge of Suiren-Base into Suiren-ConfAvg. This process encodes a macroscopic latent representation that can be inverted into specific 3D conformations through energy-conditioned queries. In the absence of an energy query, Suiren-ConfAvg accepts 2D molecular graphs or 1D SMILES as input to produce generalizable molecular embeddings suitable for a wide range of downstream tasks, including materials discovery, drug design, and battery chemistry.

We evaluate Suiren-1.0 across a comprehensive suite of over 50+ tasks spanning 9 diverse scientific domains. To ensure a rigorous and fair assessment, we eschew task-specific engineering in favor of a unified fine-tuning and inference pipeline across all benchmarks. As illustrated in Figure 1 and Figure 2, Suiren-1.0 achieves consistent, state-of-the-art (SOTA) performance in the vast majority of cases. Notably, Suiren-1.0 delivers performance gains exceeding 20% on more

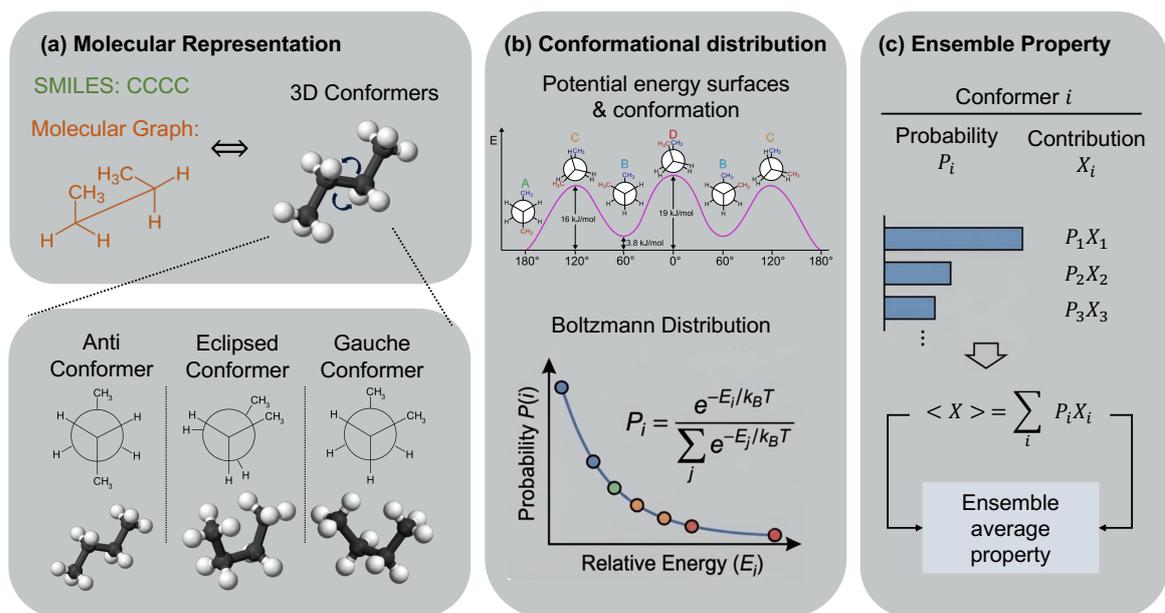


Figure 3 | **Microscopic and macroscopic representations of molecular ensembles.** (a) **Molecular Representation:** A single molecular identity corresponds to a diverse ensemble of 3D conformations at the microscopic space. (b) **Conformational distribution:** The relative probability of these conformations is governed by the Boltzmann distribution as a function of potential energy. (c) **Ensemble Property:** Macroscopic observables emerge as the ensemble-averaged properties derived from the collective contributions of all constituent conformations.

than 20+ tasks compared to existing models. We attribute this success to the synergy between large-scale model scaling and the principled integration of physical priors.

Our main contributions are as follows:

### Modeling Framework: Microscopic–Macroscopic Bridging

- We establish a three-stage framework to unify molecular scales: (i) pre-training a 3D conformation-aware foundation model for high-fidelity microscopic representation learning; (ii) distilling this knowledge into a compressed, conformation-agnostic model for macroscopic adaptation via Conformation Compression Distillation; and (iii) fine-tuning task-specific encoders for a diverse suite of downstream scientific applications.

### Pre-Training: Physical Priors and First-Principles Data

- We train Suiren-Base on large-scale first-principles quantum-chemical data (Qo2mol Liu et al. (2024b)) and incorporate physically motivated algorithms, including advanced EMPP (An et al., 2025b) and advanced EST (An et al., 2025a), to improve representation quality.
- The resulting representations capture conformation-sensitive microscopic information and transfer effectively across downstream scientific tasks. We further provide a continued pre-training variant for dimer systems (Suiren-Dimer).

### Transfer Learning: Broad Molecular Applicability

- By distilling from Suiren-Base to Suiren-ConfAvg, we enable strong performance when only graph or SMILES inputs are available, improving deployability in real-world molecular pipelines.
- We benchmark Suiren-ConfAvg on 50+ property prediction tasks and observe robust

improvements over advanced molecular baselines.

## Open Science

- We provide a description of pre-training, distillation, and finetune models, and release their weights to facilitate reproducible molecular foundation-model research. Furthermore, we open a comprehensive benchmark MoleHB for molecular model evaluation.

The remainder of this paper is organized as follows. Section 2 details the architectural design and the broader Suiren-1.0 model family. Section 3 describes our data curation and pre-training methodology, followed by Section 4, which introduces our post-training distillation and finetuning protocols. In Section 5, we present a comprehensive evaluation of Suiren-1.0. Finally, Section 6 concludes the paper with a discussion on current limitations and future research directions.

## 2. Architecture

### 2.1. Large SO(3)-Equivariant Graph Neural Network

Suiren-Base is a high-degree equivariant graph neural network (GNN) designed for 3D conformational representation learning. The architecture integrates an EquiformerV2 model (Liao et al., 2023) with a dense Mixture-of-Experts (MoE) update block, which concurrently utilizes both S2Activation and Equivariant Spherical Transformer (EST) experts (An et al., 2025a) in each forward pass. Following the standard message-passing framework (Gilmer et al., 2017), the model architecture is formulated as:

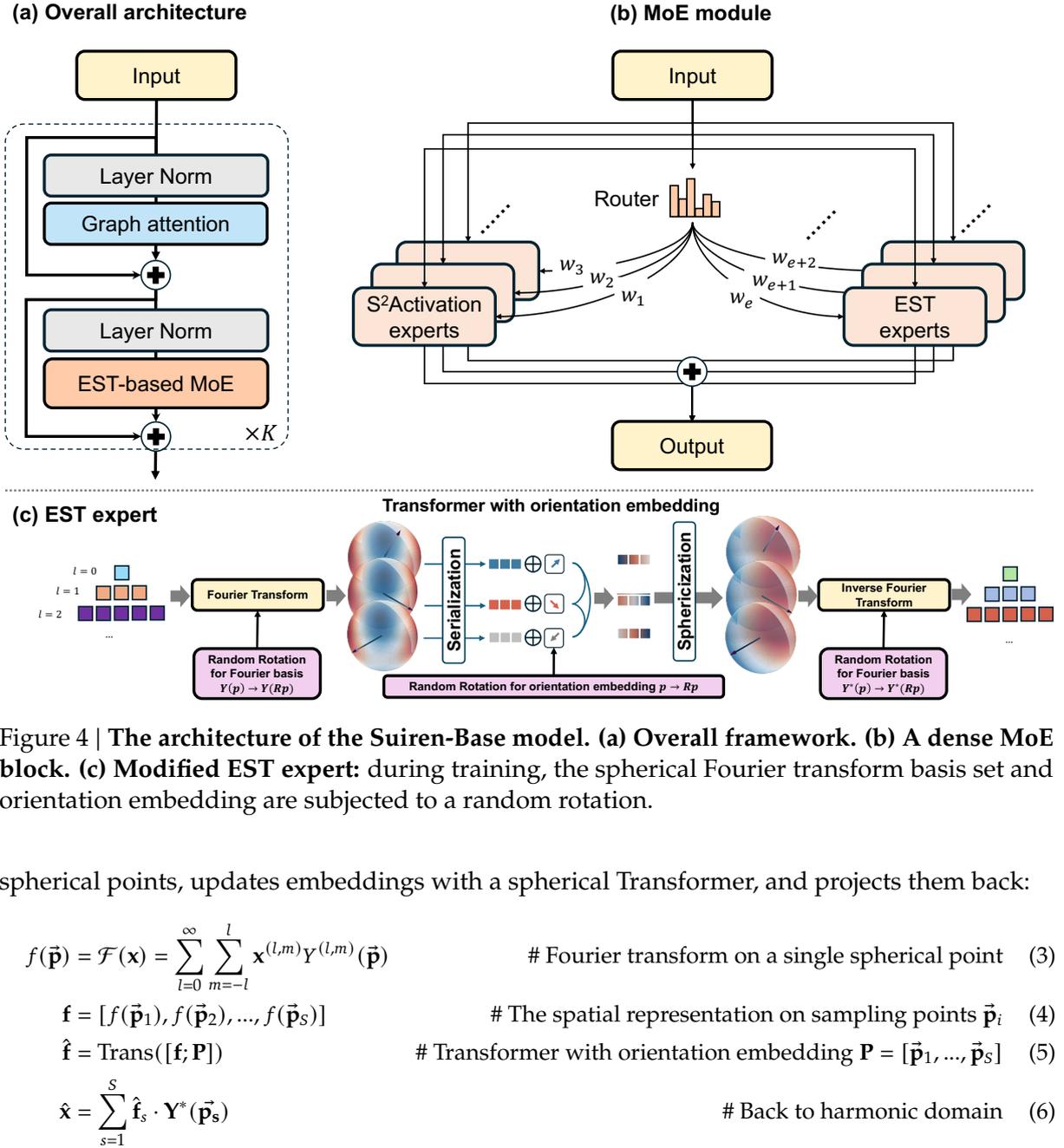
$$\mathbf{m}_i^{(k)} = \sum_{j \in \mathcal{N}(i)} \psi_m^{(l)}(\mathbf{x}_i^{(l)}, \mathbf{x}_j^{(l)}, \mathbf{e}_{ij}), \quad (1)$$

$$\mathbf{x}_i^{(k+1)} = \psi_u^{(l)}(\mathbf{x}_i^{(l)}, \mathbf{m}_i^{(l)}), \quad (2)$$

where  $\mathcal{N}(i)$  denotes the neighbor set of node  $i$ ,  $\mathbf{x}_i$  represents the node embeddings,  $\mathbf{e}_{ij}$  denotes the edge features, and  $\psi_m^{(l)}$  and  $\psi_u^{(l)}$  correspond to the message and update functions, respectively. The message block captures interatomic interactions with a computational complexity linear in the number of edges, and aggregates these into node-level messages. Conversely, the update block processes these messages with a complexity linear in the number of nodes. These components serve functional roles analogous to the self-attention and feed-forward network (FFN) modules in standard Transformer architectures.

Figure 4 illustrates the architecture of Suiren-Base. The message block is adapted from the EquiformerV2 graph attention block and utilizes an  $SO(2)$ -linear operation to integrate edge features  $\mathbf{e}_{ij}$  with node attributes  $(\mathbf{x}_i, \mathbf{x}_j)$ . For the update block, we employ a dense MoE design to enhance model capacity. This architectural choice is driven by two primary observations: (1) the update block is computationally efficient, allowing the addition of experts with minimal latency overhead; and (2) the complexity of the aggregated messages benefits significantly from increased parameter capacity.

Suiren-Base contains 20 layers ( $K = 20$ ), and each MoE block contains 20 S2Activation experts and 20 EST experts, balancing equivariance and expressiveness. The original EST maps steerable group embeddings from the harmonic domain to spatial domain with Fourier basis sampled



where  $Y^{l,m}(\cdot)$  and  $\mathbf{Y}(\cdot) = [Y^{l_1, m_1}(\cdot), Y^{l_2, m_2}(\cdot), \dots]$  denote the spherical harmonic basis (scale and vector),  $\vec{\mathbf{p}}$  denotes the sample point (orientation) in the sphere, and the term  $\hat{\cdot}$  is used to represent the embedding update. Although uniform spherical sampling in EST offers partial equivariance, it remains susceptible to discretization-induced errors. To mitigate these artifacts, we propose a basis-rotation strategy for EST experts. During training, the Fourier basis for each sample in a mini-batch is pre-rotated by a random 3D rotation  $\mathbf{R}$ . Since this procedure only modifies the basis orientation, the computational overhead remains negligible. This approach exposes the model to a diverse range of orientations, encouraging the learning of orientation-consistent responses and more closely approximating continuous spherical Fourier behavior. The formulation is

defined as:

$$\mathbf{f} = [f(\mathbf{R}\vec{\mathbf{p}}_1), f(\mathbf{R}\vec{\mathbf{p}}_2), \dots, f(\mathbf{R}\vec{\mathbf{p}}_S)] \quad (7)$$

$$\hat{\mathbf{f}} = \text{Trans}([\mathbf{f}; \mathbf{R}\mathbf{P}]) \quad (8)$$

$$\hat{\mathbf{x}} = \sum_{s=1}^S \hat{\mathbf{f}}_s \cdot \mathbf{Y}^*(\mathbf{R}\vec{\mathbf{p}}_s). \quad (9)$$

By leveraging this adaptive equivariance mechanism, the spherical sampling density  $S$  can be optimized toward the Nyquist-rate lower bound,  $S \geq (2L)^2$ , where  $L$  denotes the maximum degree of the spherical harmonic embedding. This reduction significantly lowers both training and inference overhead without compromising the robustness of the model’s equivariant properties.

Both Suiren-Base and Suiren-Dimer utilize a unified backbone architecture. In practice, these models take 3D atomic coordinates of molecules or dimers as input to predict quantum-accurate (DFT-level) potential energies and interatomic forces.

## 2.2. Conformation Compression Distillation

As illustrated in Figure 3, molecular properties are typically determined by ensemble-averaged behavior across multiple physically plausible conformers. While these conformer probabilities are governed by the Boltzmann distribution, the exact distribution—and the underlying potential energy surface (PES)—is generally unknown a priori. To address this, we propose Conformation Compression Distillation (CCD), a feature-alignment framework designed for one-to-many molecule-conformer mapping.

For each molecule-conformer pair, CCD operates on two distinct modalities: the 2D molecular topology (SMILES or graph) and the 3D conformer with its associated energy  $E$ . The 2D input is processed via a Graph Attention Network (GAT) to extract a latent representation  $\mathbf{h}^{2D}$ , while the 3D conformer is encoded by a pre-trained Suiren-Base teacher to yield an equivariant representation  $\mathbf{h}^{3D}$ . We then introduce a 3D diffusion-based model featuring a lightweight Equiformer+MoE+EST dynamics network  $\varphi_\theta(\cdot)$ . This network is conditioned on both  $\mathbf{h}^{2D}$  and the energy  $E$ , with the diffusion process targeting the joint reconstruction of the 3D representation  $\mathbf{h}^{3D}$  and the molecular 3D coordinates  $\mathbf{c}$  (see Figure 5(b)).

During training, we add random noise to the clean 3D target state:

$$\mathbf{z}_t = \alpha_t \mathbf{z}_{-1} + \sigma_t \boldsymbol{\varepsilon} \quad (10)$$

where  $\mathbf{z}_{-1} = [\mathbf{h}^{3D}; \mathbf{r}]$  denotes the clean target state,  $\mathbf{z}_t$  is the noisy state at timestep  $t \in [0, T]$ ,  $\alpha_t, \sigma_t$  are schedule coefficients, and  $\boldsymbol{\varepsilon}$  denotes the Gaussian noise. We freeze the weights of Suiren-Base, and train the dynamics network and 2D representation model by predicting the noise:

$$\hat{\boldsymbol{\varepsilon}} = \varphi_\theta(\mathbf{z}_t, t, \mathbf{h}^{2D}, E) \quad (11)$$

$$\mathcal{L}_{\text{CCD}} = \text{MSE}(\hat{\boldsymbol{\varepsilon}}, \boldsymbol{\varepsilon}), \quad (12)$$

where  $E$  is encoded with Gaussian embeddings and  $\mathcal{L}_{\text{CCD}}$  denote optimize objective functions.

Upon completion of training, the framework yields two primary components: a 2D molecular encoder and a generative diffusion dynamics model. These modules facilitate both high-fidelity conformer generation and robust 2D representation learning. For conformer generation, the diffusion dynamics model captures the multimodal nature of the structural space, generating

diverse ensembles rather than collapsing to a single low-energy mode (Landrum, 2016; Xu et al., 2024). The reverse-time sampling step from timestep  $t$  to  $s = t - 1$  is formulated as:

$$\mathbf{z}_s = \frac{1}{\alpha_{t|s}} \mathbf{z}_t - \frac{\sigma_{t|s}^2}{\alpha_{t|s} \sigma_t} \cdot \varphi_{\theta}(\mathbf{z}_t, t, \mathbf{h}^{2D}, E) + \sigma_{t \rightarrow s} \cdot \boldsymbol{\varepsilon}, \quad (13)$$

where  $\alpha_{t|s} = \alpha_t / \alpha_s$ ,  $\sigma_{t|s}^2 = \sigma_t^2 - \alpha_{t|s}^2 \sigma_s^2$ , and  $\sigma_{t \rightarrow s} = \frac{\sigma_{t|s} \sigma_s}{\sigma_t}$ . Regarding 2D representation learning, CCD implicitly characterizes the mapping from 2D graphs to 3D configurations. Given the significant modality gap between these spaces, direct feature alignment is often ill-posed. The diffusion strategy in CCD addresses this by enabling the 2D encoder to reconstruct 3D information in stages, thereby mitigating optimization challenges. This process yields the Suiren-ConfAvg model, which provides versatile representations for a broad range of macroscopic molecular tasks.

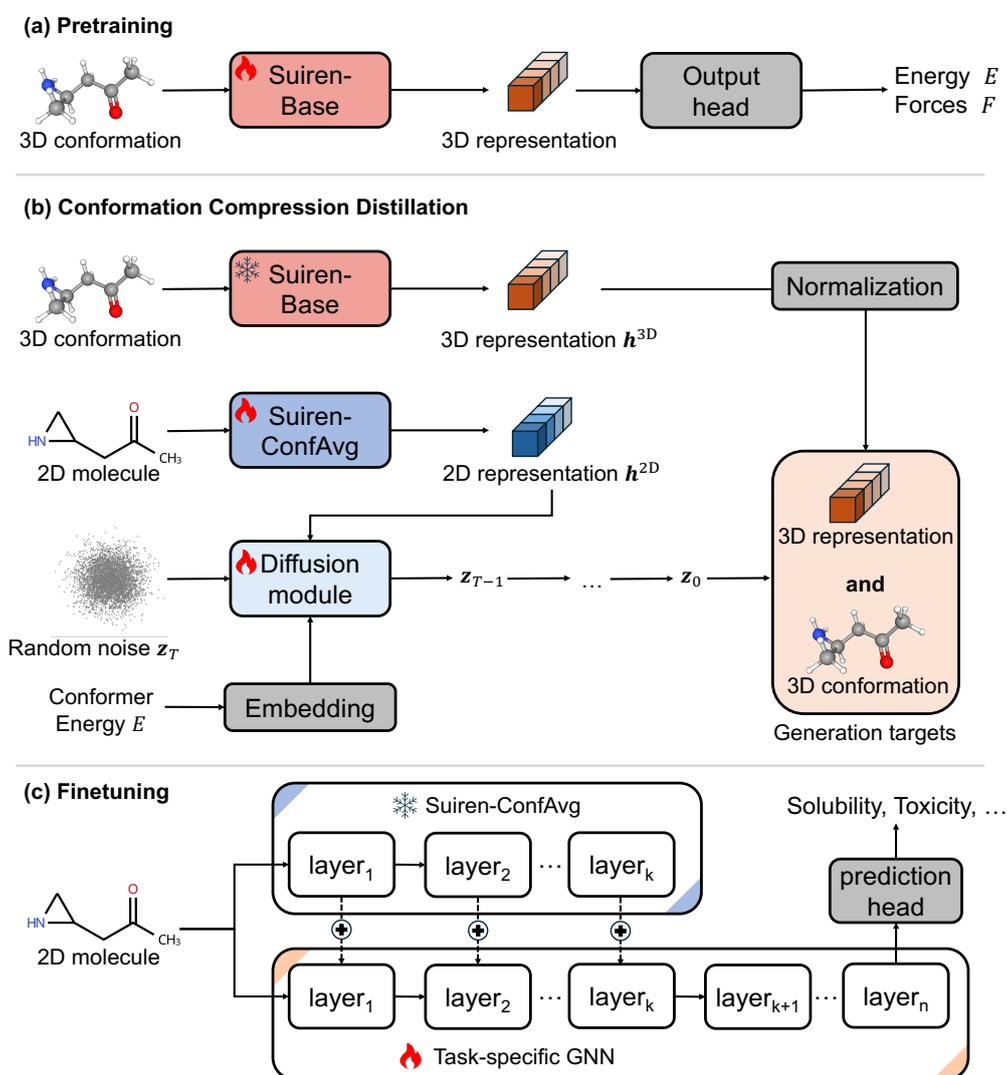


Figure 5 | **Overview of training stages.** (a) **3D Pre-training:** Self-supervised learning on 3D molecular conformations. (b) **Conformation Distillation:** Distilling 3D geometric knowledge into a conformation-averaged representation. (c) **Downstream Fine-tuning:** Adapting the model for supervised molecular property prediction.

### 2.3. Dual Graph Neural Network

Following the CCD-based training of the 2D representation model, we propose a Dual Graph Neural Network (DGNN) architecture for downstream fine-tuning. As illustrated in Figure 5(c), the DGNN consists of two parallel sub-networks: a pre-trained Suiren-ConfAvg module, initialized via CCD, and a randomly initialized task-specific GNN. During the forward pass, latent representations from Suiren-ConfAvg are injected into the corresponding layers of the task-specific GNN to provide structural guidance. To mitigate catastrophic forgetting and preserve the learned conformation-averaged features, the Suiren-ConfAvg weights remain frozen throughout this stage. While both modules utilize GAT architectures, the task-specific GNN is designed with greater depth to absorb all representations of Suiren-ConfAvg.

## 3. Pre-training

In this section, we describe the construction of the pre-training data, the multi-stage training pipeline, and the evaluation of the resulting base models.

### 3.1. Pre-training Data

The Suiren-1.0 model utilizes a vast corpus of first-principles molecular data for pre-training. Using Density Functional Theory (DFT) at the B3LYP/def2-SVP level, we generated 70 million conformer samples for organic molecules encompassing the H, C, N, O, F, P, S, Cl, Br, and I elements. Of these, 20 million samples have been publicly released as the Qo2mol dataset (Liu et al., 2024b). Each entry includes Cartesian coordinates, energies, forces, trajectory information, and associated metadata. Prior to training, we perform rigorous data cleaning to remove anomalous samples and identify the terminal optimized geometry of each trajectory, which serves as an auxiliary supervision target.

To enhance data efficiency, we augment the training process using EMPP method (An et al., 2025b). For each molecule, a random atom is deleted rather than masked, and the model is required to reconstruct its coordinates conditioned on the atom type and target molecular energy. This objective encourages the model to learn physically plausible local potential-energy landscapes and effectively doubles the training volume. We further refine the original EMPP formulation: rather than employing layer-wise conditioning of the deleted-atom signals, we feed these inputs exclusively to an EMPP-specific coordinate-prediction head. This modification ensures that the shared backbone maintains a consistent forward pass across all pre-training objectives.

### 3.2. Pre-training Stage

Pre-training is divided into three stages.

**Stage 1 (multi-task foundational capability learning)** Within the Fairchem framework, we train a 1.8B-parameter 3D model using both the original dataset and the EMPP-augmented samples. The pre-training tasks include energy prediction, force prediction, optimized-trajectory endpoint structure prediction, optimized-trajectory endpoint energy prediction, and EMPP missing-coordinate completion. All task losses are optimized jointly. Inspired by curriculum learning, we prioritize smaller molecular systems (fewer atoms) in earlier training phases. The weights of endpoint-structure prediction and endpoint-energy prediction losses are also

gradually increased during training.

Stage 1 is trained on 320 NVIDIA H800 GPUs with mixed precision and graph parallelization. Because PyTorch Geometric supports a variable number of atoms per mini-batch, we combine dynamic batch balancing and activation recomputation to avoid memory overflow.

**Stage 2 (core capability refinement)** For regression targets such as energy and force, mixed-precision training can substantially degrade model performance. Nonetheless, this degradation can be corrected by a relatively short fine-tuning stage. In Stage 2, we use only the 70M original samples for full-precision fine-tuning, with the following tasks: energy prediction, force prediction, optimized-trajectory endpoint structure prediction, and optimized-trajectory endpoint energy prediction. The weights for endpoint-structure prediction and endpoint-energy prediction are fixed to a small constant. Except for the switch from mixed precision to full precision, all optimization strategies remain unchanged.

After extensive hyperparameter search across the first two stages, we obtain Suiren-Base.

**Stage 3 (continued pre-training in the dimer domain)** Suiren-Base primarily learns intra-molecular interactions. For applications such as drug design, inter-molecular interactions, including long-range impacts, are often essential. To address this, we generate 13.5M dimer samples with DFT and continue pre-training from Suiren-Base. The architecture and optimization recipe remain consistent with Stage 2, yielding the dimer-focused model Suiren-Dimer.

### 3.3. Pre-training Evaluation

We evaluate pre-training quality primarily with MAE on energy prediction, force prediction, and optimized-trajectory endpoint prediction. For investigate pre-training performance, we reproduce several strong baselines on a Qo2mol subset. We also include UMA-family results as an external performance anchor and analyze MoE routing statistics.

**Standard Evaluation** We sample a validation subset from Qo2mol containing more than 1M conformers across different molecular scales. This subset provides a stable benchmark for monitoring pre-training quality. As shown in Table 1, compared with baseline methods, Suiren-Base achieves highly accurate results on both energy and force prediction. Optimized-trajectory endpoint structure and endpoint energy are substantially more challenging targets, yet Suiren-1.0 still attains strong accuracy.

*Suiren-0.0* is an internal transitional model trained with less compute and a weakened training recipe. Because it uses exactly the same training set as Suiren-Base, its results directly reflect the gains brought by the improved training strategy and algorithmic refinements in Suiren-1.0. EquiformerV2 (Liao et al., 2023) and eSCN (Passaro and Zitnick, 2023) are strong backbones; due to compute constraints, we train them on a 20M Qo2mol subset. Their results further support the effectiveness of the Suiren-1.0 training pipeline.

We also compare Atomic Energy MAE and Force MAE with UMA-family models on the organic benchmark set (OMol). Our energy prediction is comparable or better (0.258 vs. > 0.33), while our force prediction shows a much larger improvement (0.510 vs. > 2.90). Note that this UMA comparison is intended only to provide a rough performance reference, since the training and evaluation datasets are not identical across methods.

Table 1 | **Comparison among energy/forces prediction models. The best results are shown in bold.**

	EquiformerV2	eSCN	Suiren-0.0	Suiren-Base (1.0)	Suiren-Dimer (1.0)
# Total Params	31M	168M	1.5B	1.8B	1.8B
Molecular Energy MAE (meV) ↓	206.15	268.1	10.82	<b>9.08</b>	10.30
Atomic Energy MAE (meV) ↓	5.854	7.613	0.307	<b>0.258</b>	0.292
Atomic Forces MAE (meV / Å) ↓	29.3	37.6	0.979	<b>0.510</b>	1.294
Optimized Structure MAE (Å) ↓	-	-	0.02895	<b>0.02467</b>	0.06830
Optimized Energy MAE (meV) ↓	-	-	28.187	<b>16.886</b>	58.325

Finally, we evaluate the continued pre-training model, Suiren-Dimer. Compared with intra-molecular settings, inter-molecular trajectories are more complex. Accordingly, Suiren-Dimer is weaker than Suiren-Base on endpoint structure and endpoint energy prediction, but it still maintains strong performance on energy and force prediction.

**Feature Evaluation** We monitor MoE routing-weight distributions and observe that they become progressively sparse during training. Most mass eventually concentrates on a small subset of experts, while unused experts still retain non-negligible weights. For this reason, we do not adopt top- $K$  routing in Suiren-Base. In addition, the two kind of experts in Suiren-Base receive broadly similar aggregate routing mass, with EST experts being slightly higher on average than standard experts.

## 4. Post-training

### 4.1. Post-training Stage

**Stage 1 (diffusion distillation)** We post-train on the same 70M molecules used in pre-training, but with a different objective. Follow Section 2.2, We condition a diffusion model on 2D representations and embedding of conformer energies. The diffusion model and 2D GNN learn to generate corresponding 3D representations and 3D coordinates. During this stage, the 3D branch is instantiated with Suiren-Base and kept frozen.

**Stage 2 (contrastive learning)** After Stage 1 reaches a stable regime, we introduce an additional alignment objective. Specifically, we attach one projection head to the 2D model and one to the 3D model, and apply SigLIP-style contrastive learning (Zhai et al., 2023) on their outputs. The 3D branch remains frozen, and the Stage-1 diffusion objective is retained. The diffusion and contrastive objectives are jointly optimized with task-specific loss weights.

Through the first two stages, we obtain the Suiren-ConfAvg model.

**Stage 3 (property prediction)** In the final stage, we evaluate the performance of Suiren-ConfAvg across a diverse array of downstream benchmarks. Each task involves predicting a specific molecular property using sparse experimental wet-lab data. We fine-tune the model for each objective using the integrated DGNN+Suiren-ConfAvg architecture. To demonstrate the general transferability and robustness of Suiren-ConfAvg representations, we maintain a unified

hyperparameter configuration across all tasks, regardless of the domain. A detailed account of these configurations and the comprehensive evaluation results are documented in Section 5.

## 5. Experiments

### 5.1. Benchmark

#### 5.1.1. *MoleHB*

**Dataset** We introduce Molelecular handbook (MoleHB), a comprehensive molecular property prediction benchmark encompassing 40+ heterogeneous tasks. The benchmark spans several critical scientific domains, including safety, structural, critical and saturation, energetic, thermal, solution, transport and fluctuation properties. All data points are sourced from (Yaws, 1999) and have been rigorously validated via wet-lab experiments to ensure high-fidelity, stable values. We propose two evaluation protocols: (1) Random split: a standard random split to evaluate performance under similar data distributions; and (2) Scaffold split: a strategy where molecules with larger atom counts are assigned to the validation set to assess the model’s structural extrapolation capabilities. Both the datasets and splitting protocols have been open-sourced to facilitate reproducible research. Results of Scaffold split are shown in Appendix C.

**Baselines and Configurations** We benchmarked three state-of-the-art models on MoleHB: MoleBERT (Xia et al., 2023), Uni-Molv1 (Zhou et al., 2023), and Uni-Molv2 (Ji et al., 2024). Like the SuiRen family, these baselines utilize large-scale pre-training to generate high-quality representations for diverse molecular tasks. To ensure a fair comparison, all baselines were reproduced using their official training scripts and identical hyperparameter configurations to the SuiRen models.

Detailed training configurations for all evaluated methods are summarized in Table 2. We ensured that each model reached performance convergence under these settings. All experiments were performed on a single NVIDIA RTX 4090 GPU, with fine-tuning typically completed in less than one hour per task.

#### 5.1.2. *Therapeutics Data Commons*

**Dataset** Therapeutics Data Commons (TDC) (Huang et al., 2021) is an open-access platform providing AI-ready datasets and benchmarks for drug discovery. It covers diverse therapeutic tasks—including target discovery, activity screening, efficacy, and safety—across small molecules, antibodies, and vaccines. We evaluate SuiRen model on its ADMET group.

**Baselines and Configurations** TDC is a public leaderboard where readers can find scores for various methods on its official website. Here, we follow (Gao et al., 2023), using ChemProp (Stokes et al., 2020), DeepAutoQSAR (Dixon et al., 2016), DeepPurpose (Huang et al., 2020), and Uni-QSAR (Gao et al., 2023) as baselines. Note that TDC ADMET tasks include both regression and classification tasks. Regression tasks use configurations identical to those in Table 2. For classification tasks, the loss function is changed to cross-entropy while all other settings remain consistent.

Table 2 | Training configurations of the Suiren model and other baselines on MoleHB experiments.

Training Hyperparameter	Value
<i>Task Formulation</i>	
Loss function	MAE
Train/validation split ratio (Random and Scaffold)	0.8:0.2
<i>Optimization</i>	
Mini-batch size per GPU	8
Optimizer	adamw
Initial learning rate	4e-4
Weight decay coefficient	0.01
Momentum coefficient	0.9
<i>Schedule</i>	
Scheduler type	cosine
Warmup epochs	0
Minimum learning rate	1e-6
Total training epochs	200

## 5.2. Results

### 5.2.1. MoleHB (Random Split)

**Overall Performance Summary** We comprehensively evaluate the predictive performance of Suiren-ConfAvg against three representative baseline models (Mole-BERT, Uni-Mol v1, and Uni-Mol v2) across 40+ molecular properties spanning eight categories: critical & saturation, safety, fluctuation, solution, thermal, structural, energetic, and transport properties. Performance is measured using Mean Absolute Error (MAE, lower is better) and coefficient of determination ( $R^2$ , higher is better). As summarized in Tables 3–11, Suiren-ConfAvg achieves state-of-the-art MAE on 41 out of 43 properties, with consistent improvements in  $R^2$  for the majority of tasks.

Table 3 | Results of critical and saturation properties: model performance (MAE/ $R^2$ ). Best MAE and best  $R^2$  per property are boldfaced.

Property	Unit	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg	Improvement (%)
critical temperature	K	27.3807/0.8340	9.3277/ <b>0.9787</b>	30.6705/0.8850	<b>7.2302</b> /0.9775	22.49
critical pressure	bar	2.0733/0.9045	0.7332/0.9765	2.9769/0.9084	<b>0.5661</b> / <b>0.9792</b>	22.79
critical density	g/ml	0.0134/0.9253	0.0054/ <b>0.9814</b>	0.0115/0.9497	<b>0.0043</b> /0.9786	20.37
critical volume	cm <sup>3</sup> /mol	61.8034/0.8484	8.6329/0.9981	7.0201/0.9989	<b>4.2825</b> / <b>0.9992</b>	39.00
critical compressibility	cm <sup>3</sup>	0.0189/0.6178	0.0066/0.8858	0.0098/0.8808	<b>0.0057</b> / <b>0.8976</b>	13.64

Table 4 | Results of safety properties.

Property	Unit	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg	Improvement (%)
flash point	F	19.6978/0.8017	9.9274/0.9472	13.9535/0.9438	<b>9.1004</b> / <b>0.9564</b>	8.33
lower explosive limit	vol%	0.1689/0.6854	0.0974/ <b>0.8926</b>	0.1106/0.8341	<b>0.0963</b> /0.8561	1.13
upper explosive limit	vol%	1.4188/0.5846	1.0460/0.6979	1.0023/0.6942	<b>0.8346</b> / <b>0.7801</b>	16.73

Table 5 | Results of fluctuation properties.

Property	Unit	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg	Improvement (%)
heat capacity of liquid	J/mol/K	16.0687/0.7922	5.9321/0.9812	5.8939/0.9792	<b>4.5252/0.9815</b>	23.22
heat capacity of solid	J/mol/K	49.2807/0.9499	6.5306/0.9988	3.2724/0.9990	<b>1.1395/0.9997</b>	65.18
coefficient of thermal expansion of liquid	1/K	0.0002/0.1716	0.0001/0.8158	0.0001/0.6211	<b>0.0000/0.9178</b>	100.00
heat capacity of gas	J/mol/K	24.1482/0.8849	4.4112/0.9685	3.2478/ <b>0.9691</b>	<b>2.7546/0.9690</b>	15.19

Table 6 | Results of solution properties.

Property	Unit	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg	Improvement (%)
octanol water partition coefficient	-	0.3786/0.9414	0.1491/0.9877	0.1386/0.9858	<b>0.1353/0.9878</b>	2.38
solubility in water	ppm(wt)	0.2110/0.9422	0.0564/0.9973	0.0456/0.9980	<b>0.0370/0.9984</b>	18.86
solubility in water containing salt	ppm(wt)	0.2535/0.7540	0.0615/0.9924	0.0406/0.9961	<b>0.0374/0.9967</b>	7.88
solubility parameter	(J/cm <sup>3</sup> ) <sup>1/2</sup>	0.7111/0.8331	0.3873/0.9236	0.7614/0.8282	<b>0.3600/0.9284</b>	7.05
henrys law constant for compound in water	-	0.2856/0.9581	0.1517/0.9816	0.1930/0.9744	<b>0.1298/0.9841</b>	14.44
henrys law constant for gas in water	-	0.5454/0.3000	0.3017/0.7939	<b>0.1773/0.9430</b>	0.5104/0.4226	-187.87

Table 7 | Results of thermal properties.

Property	Unit	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg	Improvement (%)
density of liquid	g/ml	0.0331/0.9544	0.0160/0.9761	0.0145/ <b>0.9763</b>	<b>0.0120/0.9691</b>	17.24
melting point	K	19.9979/0.8199	14.1449/0.9053	24.7538/0.8050	<b>11.5566/0.9245</b>	18.30
boiling point	K	19.8744/0.8891	6.2177/0.9906	12.3314/0.9777	<b>4.5883/0.9928</b>	26.21
vapor pressure	mmHg	0.4837/0.5674	0.1596/ <b>0.9579</b>	0.2000/0.9370	<b>0.1505/0.9473</b>	5.70

Table 8 | Results of structural properties.

Property	Unit	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg	Improvement (%)
liquid volume	cm <sup>3</sup> /mol	19.5900/0.8583	3.9679/0.9945	8.9388/0.9876	<b>2.1627/0.9972</b>	45.50
acentric factor	omega	0.0702/0.7591	0.0233/0.9293	0.0225/ <b>0.9312</b>	<b>0.0178/0.9019</b>	20.89
refractive index	-	0.0277/0.4260	0.0057/0.9314	0.0079/0.9276	<b>0.0050/0.9351</b>	12.28
radius of gyration	10 <sup>-10</sup> m	0.3573/0.6317	0.1230/0.9715	0.1204/0.9757	<b>0.1100/0.9794</b>	8.64
dipole moment	D	0.4022/0.6638	<b>0.2987/0.8115</b>	0.3256/0.7585	0.3138/0.8078	-5.06

Table 9 | Results of energetic properties.

Property	Unit	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg	Improvement (%)
entropy of formation	J/mol/K	90.1219/0.8775	16.6006/0.9975	11.1386/ <b>0.9982</b>	<b>8.6649/0.9981</b>	22.21
gibbs energy of formation	J/mol	32.5168/0.9365	8.3799/0.9865	19.1059/0.9785	<b>5.3047/0.9956</b>	36.70
helmholtz energy of formation	kJ/mol	30.2883/0.9286	7.9805/ <b>0.9829</b>	24.8331/0.9656	<b>5.5755/0.9788</b>	30.14
internal energy of formation	kJ/mol	34.2535/0.8869	10.3932/0.9795	36.3480/0.9393	<b>6.4382/0.9949</b>	38.05
enthalpy of vaporization	kJ/mol	3.3764/0.7398	1.3875/0.9694	1.2207/ <b>0.9773</b>	<b>1.0936/0.9772</b>	10.41
enthalpy of fusion	J/mol/K	3.3815/0.8819	1.2265/0.9898	1.3044/0.9888	<b>0.9023/0.9902</b>	26.43
enthalpy of combustion	kJ/mol	664.8247/0.8814	62.4789/0.9986	39.3110/0.9987	<b>16.8780/0.9993</b>	57.07
entropy of gas	J/mol/K	38.1690/0.8946	9.9913/ <b>0.9900</b>	10.9098/0.9896	<b>7.2880/0.9900</b>	27.06
enthalpy of formation	kJ/mol	43.0063/0.8548	12.9888/0.9376	12.3369/0.9486	<b>6.5543/0.9693</b>	46.87

Table 10 | Results of transport properties.

Property	Unit	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg	Improvement (%)
viscosity of liquid	mp	0.8216/0.5870	0.4716/0.8328	0.9996/0.5924	<b>0.2946/0.9115</b>	37.53
thermal conductivity of gas	W/m/K	0.0019/0.2592	0.0003/0.9584	0.0006/0.9293	<b>0.0002/0.9678</b>	33.33
thermal conductivity of liquid	W/m/K	0.0107/0.0565	0.0028/0.7911	0.0036/ <b>0.8183</b>	<b>0.0026/0.7726</b>	7.14
diffusion coefficient at infinite dilution in water	cm <sup>2</sup> /s	<b>0.0000</b> /-1.2725	<b>0.0000</b> /0.9865	<b>0.0000</b> /0.9817	<b>0.0000/0.9900</b>	-
diffusion coefficient in air	cm <sup>2</sup> /s	0.0066/0.6975	0.0014/0.9845	0.0011/0.9858	<b>0.0010/0.9862</b>	9.09

Table 11 | Results of other properties.

Property	Unit	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg	Improvement (%)
surface tension	dynes/cm	1.5226/0.8367	1.0414/0.8951	1.7084/0.8563	<b>0.8777/0.9225</b>	15.72
hydration free energy	kJ/mol	0.8510/0.8748	0.4545/ <b>0.9646</b>	0.4627/0.9623	<b>0.4401/0.9628</b>	3.17

**Critical and Saturation Properties** Suiren-ConfAvg attains the lowest MAE across all five critical properties, with relative improvements ranging from 13.6% (critical compressibility) to 39.0% (critical volume) over the strongest baseline. Notably, while Uni-Mol v1 achieves marginally higher  $R^2$  on critical temperature and density, Suiren-ConfAvg maintains competitive  $R^2$  values ( $>0.97$ ) while substantially reducing prediction errors, indicating superior calibration for extreme-value regression tasks.

**Safety and Fluctuation Properties** For safety-related properties, Suiren-ConfAvg consistently outperforms baselines in both MAE and  $R^2$ , with the most pronounced gain observed for upper explosive limit (16.7% MAE reduction). In fluctuation properties, the method demonstrates exceptional capability in modeling solid-phase heat capacity (65.2% improvement).

**Solution Properties** Suiren-ConfAvg achieves best-in-class performance on five of six solution properties. Improvements are particularly notable for solubility prediction in pure and saline water (18.9% and 7.9% MAE reduction, respectively), which are critical for pharmaceutical and environmental applications. However, for Henry’s law constant of gases in water, the method underperforms Uni-Mol v2 by a substantial margin (187.9%). We hypothesize this stems from the sparse and heterogeneous distribution of gas-phase solubility data, which may require specialized augmentation strategies.

**Thermal and Structural Properties** Across thermal properties, Suiren-ConfAvg reduces MAE by 5.7%–26.2% while maintaining  $R^2 \geq 0.947$ . For structural descriptors, the method excels in predicting liquid volume (45.5% improvement) and surface tension (15.7%), reflecting its capacity to encode intermolecular interaction patterns. The sole exception is dipole moment, where Uni-Mol v1 retains an edge (MAE: 0.299 vs. 0.314).

**Energetic and Transport Properties** The most consistent gains are observed in energetic properties, where Suiren-ConfAvg achieves optimal MAE on all nine tasks, with improvements exceeding 30% for Gibbs energy, internal energy, and enthalpy of formation. This suggests that the energy-related knowledge learned by Suiren-Base is transferred to Suiren-ConfAvg. For transport properties, substantial improvements are seen in viscosity (37.5%) and gas-phase thermal conductivity (33.3%), though liquid-phase thermal conductivity shows modest gain (7.1%), possibly due to stronger dependence on many-body hydrodynamic effects.

### 5.2.2. TDC ADMET group

We evaluated the performance of Suiren-ConfAvg on the TDC ADMET benchmarks. All experiments strictly adhered to the official evaluation protocols and metric settings provided by TDC to ensure a fair comparison. The results for regression tasks (MAE), classification tasks (AUROC and AUPRC) are presented in Tables 12, 13, and 14, respectively.

A critical distinction of SuiRen-ConfAvg lies in its training protocol. Unlike several methods that may rely on extensive task-specific hyperparameter optimization, SuiRen-ConfAvg was evaluated using a single, fixed training configuration across all datasets without complex hyperparameter search.

Table 12 | Comparison of results for regression properties in TDC ADMET (MAE).

Property	ChemProp	DeepAutoQSAR	DeepPurpose	Uni-QSAR	SuiRen-ConfAvg
Caco2	0.3900	0.3060	0.3930	<u>0.2730</u>	<b>0.2715</b>
Lipophilicity	0.4370	0.4760	0.5740	<u>0.4200</u>	<b>0.4014</b>
AqSol	0.8200	0.7840	0.8270	<b>0.6770</b>	<u>0.7285</u>
PPBR	<u>7.9930</u>	8.0430	9.9940	<b>7.5300</b>	8.1650
LD50	<b>0.5480</b>	0.5900	0.6780	0.5530	<u>0.5482</u>

Table 13 | Comparison of results for classification properties in TDC ADMET (AUROC).

Property	ChemProp	DeepAutoQSAR	DeepPurpose	Uni-QSAR	SuiRen-ConfAvg
HIA	0.9790	0.9820	0.9720	<u>0.9920</u>	<b>0.9992</b>
Pgp	0.9020	0.9170	0.9180	<b>0.9340</b>	<u>0.9332</u>
Bioavailability	0.6230	0.6820	0.6720	<u>0.7320</u>	<b>0.7642</b>
BBB	0.8820	0.8760	0.8890	<u>0.9250</u>	<b>0.9316</b>
CYP3A4 Substrate	0.6100	0.6420	0.6390	<u>0.6450</u>	<b>0.7059</b>
hERG	0.7500	0.8450	0.8410	<u>0.8560</u>	<b>0.8739</b>
Ames	<u>0.8640</u>	<u>0.8640</u>	0.8230	<b>0.8760</b>	0.8377
DILI	0.9180	<u>0.9330</u>	0.8750	<b>0.9420</b>	0.9304

Table 14 | Comparison of results for classification properties in TDC ADMET (AUPRC).

Property	ChemProp	DeepAutoQSAR	DeepPurpose	Uni-QSAR	SuiRen-ConfAvg
CYP2C9 Inhibition	0.7700	<u>0.7920</u>	0.7420	<b>0.8010</b>	0.7635
CYP2D6 Inhibition	0.6640	<u>0.7020</u>	0.6160	<b>0.7430</b>	0.6598
CYP3A4 Inhibition	0.8700	<u>0.8830</u>	0.8290	<b>0.8880</b>	0.8691
CYP2C9 Substrate	0.3910	0.3950	0.3800	<b>0.4540</b>	<u>0.4480</u>
CYP2D6 Substrate	0.6880	0.7030	0.6770	<u>0.7210</u>	<b>0.7221</b>

Despite this constraint, the model achieved SOTA results in 9 out of 18 total metrics and ranked second in an additional 4 metrics. In cases where SuiRen-ConfAvg did not secure the first place (e.g., *LD50*, *Pgp*, *CYP2C9 Substrate*), the performance gaps were negligible. This suggests that the performance sacrifice, if any, is minimal compared to the gains in reproducibility and ease of deployment. The ability to deliver highly competitive, often leading, performance across regression and classification tasks without fine-tuning highlights the strong generalization

capability and robustness of the Suiren-ConfAvg architecture. These results validate that Suiren-ConfAvg offers an efficient and reliable solution for ADMET prediction, balancing high predictive accuracy with practical implementation simplicity.

## 6. Conclusion

In this work, we propose the Suiren-1.0 family, which comprises three models: Suiren-Base, Suiren-Dimer, and Suiren-ConfAvg. Suiren-Base and Suiren-Dimer are two 3D conformational models, whose performance is ensured through large-scale pre-training. Suiren-ConfAvg is obtained by distilling the 3D representations from Suiren-Base into the 2D representation space via our proposed CCD method. We have validated the strong performance of Suiren-1.0 on various molecular tasks through extensive experiments. The models and benchmarks developed in this work have also been open-sourced. We hope this work can support research on molecular foundation models.

Suiren-1.0 also has some limitations and directions for future work: (1) Due to computational constraints, we are unable to further scale up the model size; (2) In the MoE framework of Suiren-Base, we adopted a dense expert strategy—in the future, as the number of experts increases, we can use Top-K to improve inference speed; (3) For some specific downstream tasks, the potential of Suiren models can be further explored through hyperparameter search.

## References

- J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al. Gpt-4 technical report. [arXiv preprint arXiv:2303.08774](#), 2023.
- J. An, X. Lu, C. Qu, Y. Shi, P. Lin, Q. Tang, L. Xu, F. Cao, and Y. Qi. Equivariant spherical transformer for efficient molecular modeling. [arXiv preprint arXiv:2505.23086](#), 2025a.
- J. An, C. Qu, Y.-F. Shi, X. Liu, Q. Tang, F. Cao, and Y. Qi. Equivariant masked position prediction for efficient molecular representation. [arXiv preprint arXiv:2502.08209](#), 2025b.
- L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, et al. Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072, 2021.
- M. Charbonneau, K. Van Vliet, and P. Vasilopoulos. Linear response theory revisited iii: One-body response formulas and generalized boltzmann equations. *Journal of Mathematical Physics*, 23(2):318–336, 1982.
- S. L. Dixon, J. Duan, E. Smith, C. D. Von Bargen, W. Sherman, and M. P. Repasky. Autoqsar: an automated machine learning tool for best-practice quantitative structure–activity relationship modeling. *Future medicinal chemistry*, 8(15):1825–1839, 2016.
- Z. Gao, X. Ji, G. Zhao, H. Wang, H. Zheng, G. Ke, and L. Zhang. Uni-qsar: an auto-ml tool for molecular property prediction. [arXiv preprint arXiv:2304.12239](#), 2023.
- J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. Pmlr, 2017.

- K. Huang, T. Fu, L. M. Glass, M. Zitnik, C. Xiao, and J. Sun. Deeppurpose: a deep learning library for drug–target interaction prediction. *Bioinformatics*, 36(22-23):5545–5547, 2020.
- K. Huang, T. Fu, W. Gao, Y. Zhao, Y. Roohani, J. Leskovec, C. W. Coley, C. Xiao, J. Sun, and M. Zitnik. Therapeutics data commons: Machine learning datasets and tasks for drug discovery and development. *arXiv preprint arXiv:2102.09548*, 2021.
- X. Ji, Z. Wang, Z. Gao, H. Zheng, L. Zhang, G. Ke, et al. Uni-mol2: Exploring molecular pretraining model at scale. *arXiv preprint arXiv:2406.14969*, 2024.
- G. Landrum. Rdkit: Open-source cheminformatics software. 2016. URL [https://github.com/rdkit/rdkit/releases/tag/Release\\_2016\\_09\\_4](https://github.com/rdkit/rdkit/releases/tag/Release_2016_09_4).
- D. S. Levine, M. Shuaibi, E. W. C. Spotte-Smith, M. G. Taylor, M. R. Hasyim, K. Michel, I. Bata-tia, G. Csányi, M. Dzamba, P. Eastman, et al. The open molecules 2025 (omol25) dataset, evaluations, and models. *arXiv preprint arXiv:2505.08762*, 2025.
- Y.-L. Liao, B. Wood, A. Das, and T. Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059*, 2023.
- A. Liu, B. Feng, B. Xue, B. Wang, B. Wu, C. Lu, C. Zhao, C. Deng, C. Zhang, C. Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024a.
- W. Liu, X. Ai, Z. Zhou, C. Qu, J. An, Z. Zhou, Y. Cheng, Y. Xu, F. Cao, and A. Qi. An open quantum chemistry property database of 120 kilo molecules with 20 million conformers. *arXiv preprint arXiv:2410.19316*, 2024b.
- S. Passaro and C. L. Zitnick. Reducing so (3) convolutions to so (2) for efficient equivariant gnns. In *International conference on machine learning*, pages 27420–27438. PMLR, 2023.
- W. P. Schleich, D. M. Greenberger, D. H. Kobe, and M. O. Scully. Schrödinger equation revisited. *Proceedings of the National Academy of Sciences*, 110(14):5374–5379, 2013.
- J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackermann, et al. A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702, 2020.
- G. Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.
- B. M. Wood, M. Dzamba, X. Fu, M. Gao, M. Shuaibi, L. Barroso-Luque, K. Abdelmaqsoud, V. Gharakhanyan, J. R. Kitchin, D. S. Levine, et al. Uma: A family of universal models for atoms. *arXiv preprint arXiv:2506.23971*, 2025.
- J. Xia, C. Zhao, B. Hu, Z. Gao, C. Tan, Y. Liu, S. Li, and S. Z. Li. Mole-bert: Rethinking pre-training graph neural networks for molecules. In *The Eleventh International Conference on Learning Representations*, 2023.
- G. Xu, Y. Jiang, P. Lei, Y. Yang, and J. Chen. Gtmgc: Using graph transformer to predict molecule’s ground-state conformation. In *The Twelfth International Conference on Learning Representations*, 2024.
- A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Gao, C. Huang, C. Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

- C. L. Yaws. Chemical properties handbook: physical, thermodynamic, environmental, transport, safety, and health related properties for organic and inorganic chemicals. (No Title), 1999.
- X. Zhai, B. Mustafa, A. Kolesnikov, and L. Beyer. Sigmoid loss for language image pre-training. In Proceedings of the IEEE/CVF international conference on computer vision, pages 11975–11986, 2023.
- G. Zhou, Z. Gao, Q. Ding, H. Zheng, H. Xu, Z. Wei, L. Zhang, and G. Ke. Uni-mol: A universal 3d molecular representation learning framework. In The eleventh international conference on learning representations, 2023.

## Appendix

### A. Contributions and Acknowledgments

#### Research & Engineering & Data Computing

Junyi An

Xinyu Lu (Intern)

Yun-Fei Shi

Li-Cheng Xu

Nannan Zhang (Intern)

Chao Qu

Fenglei Cao

Yuan Qi

We would also like to acknowledge the SAIS platform and all members of Golab, who contributed to the development of the Suiren-1.0 model in critical areas such as business and evaluation operations.

### B. The explanation of evaluation metrics in Figure 1 and Figure 2

In Figure 1, we use the inverse MAE to quickly demonstrate the performance of Suiren across various properties. Specifically, the MAE values were normalized and mapped to a standardized scoring scale ranging from 60 to 100. Since MAE represents prediction error (where lower values indicate better performance), an inverse Min-Max normalization strategy was employed.

For a specific property  $p$ , let  $E_{m,p}$  denote the MAE of model  $m$ . The normalized score  $S_{m,p}$  for model  $m$  on property  $p$  is calculated as follows:

$$S_{m,p} = 60 + 40 \times \frac{\max(E_p) - E_{m,p}}{\max(E_p) - \min(E_p) + \varepsilon} \quad (14)$$

where:

- $\max(E_p)$  and  $\min(E_p)$  represent the maximum and minimum MAE values observed among all evaluated models for property  $p$ , respectively.
- The constant 60 establishes the baseline score for the worst-performing model (i.e., when  $E_{m,p} = \max(E_p)$ ).
- The scaling factor 40 expands the score up to a maximum of 100 for the best-performing model (i.e., when  $E_{m,p} = \min(E_p)$ ).
- $\varepsilon$  is an infinitesimally small constant (e.g.,  $10^{-10}$ ) added to the denominator to prevent potential division-by-zero errors in cases where all models yield identical MAE values.

This transformation ensures that properties with drastically different magnitude scales are projected onto a uniform visual space, allowing for a fair, area-based geometric comparison in the resulting radar charts.

In Figure 2, we use scaled MAE. For each property, we divide the results of the four models by the maximum MAE value among the four models. This way, all results are unified to the range between 0 and 1. This approach alleviates the loss of visual information caused by large numerical differences between different properties.

### C. Evaluation of MoleHB Scaffold split

To further explore the generalization capability of foundation models under distribution shift, we systematically evaluated MoleBERT, Uni-Mol v1, Uni-Mol v2, and Suiren-ConfAvg on the scaffold-split subset of MoleHB, a setting that rigorously tests extrapolation to unseen molecular scaffolds. The comprehensive results across eight property categories are summarized in Tables 15–22.

**Overall performance trends** As anticipated, scaffold splitting induces a non-trivial distribution shift, leading to performance degradation across all methods relative to random-split evaluations. Nevertheless, Suiren-ConfAvg demonstrates superior robustness: it achieves the lowest mean absolute error (MAE) on 31 out of 38 evaluated properties (81.6%), with relative improvements ranging from 4.6% (lower explosive limit) to 92.1% (Helmholtz energy of formation) over the strongest baseline. Notably, the average relative improvement across all tasks is approximately 58.3%, underscoring the effectiveness of confidence-weighted ensemble strategies in mitigating scaffold-induced generalization gaps.

Table 15 | Results of critical and saturation properties: model performance (MAE/R2). Best MAE and best R2 per property are boldfaced.

Property	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg
critical compressibility	0.03447	0.03632	0.03527	<b>0.00879</b>
critical density	<b>0.00671</b>	0.00720	0.00734	0.00690
critical pressure	<b>1.73699</b>	3.27665	2.63610	1.93246
critical temperature	77.96931	63.34677	87.88906	<b>21.74572</b>
critical volume	351.31897	290.12731	321.89835	<b>59.14848</b>

Table 16 | Results of safety properties.

Property	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg
flash point	32.80089	20.89532	36.13503	<b>13.22462</b>
lower explosive limit	0.10239	0.05315	0.10274	<b>0.05071</b>
upper explosive limit	0.74490	0.59120	0.62222	<b>0.31921</b>

Table 17 | Results of fluctuation properties.

Property	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg
heat capacity of gas	144.17387	88.39811	146.68384	<b>27.21923</b>
heat capacity of liquid	60.96739	26.91383	40.76889	<b>10.84865</b>
heat capacity of solid	61.55533	43.78053	71.91361	<b>5.76057</b>
coefficient of thermal expansion of liquid	0.00010	0.00005	0.00009	<b>0.00004</b>

#### Category-wise analysis

Table 18 | Results of solution properties.

Property	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg
octanol water partition coefficient	0.68611	0.29471	0.44365	<b>0.21224</b>
solubility in water	0.32445	0.08128	0.28590	<b>0.06965</b>
solubility in water containing salt	0.25057	0.09037	0.17057	<b>0.06413</b>
solubility parameter	1.36914	1.53168	1.41960	<b>0.74460</b>

Table 19 | Results of thermal properties.

Property	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg
density of liquid	0.01929	0.01221	0.01210	<b>0.01023</b>
boiling point	66.21946	53.40313	70.61806	<b>21.49229</b>
melting point	39.98979	42.88753	51.46680	<b>39.98779</b>
vapor pressure	1.21508	0.89771	1.08963	<b>0.39283</b>

Table 20 | Results of structural properties.

Property	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg
liquid volume	130.82504	103.69391	114.33397	<b>30.19665</b>
acentric factor	0.09174	0.07457	0.08191	<b>0.06843</b>
radius of gyration	1.15782	0.53795	0.73237	<b>0.21384</b>
refractive index	0.01290	<b>0.00859</b>	0.00974	0.00938
dipole moment	0.64795	<b>0.49490</b>	0.53720	0.56325

Table 21 | Results of energetic properties.

Property	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg
gibbs energy of formation	99.86093	52.53548	74.17537	<b>6.11849</b>
helmholtz energy of formation	109.97257	73.68592	86.25709	<b>5.82731</b>
internal energy of formation	149.43255	90.63668	135.46958	<b>10.11958</b>
enthalpy of combustion	4411.49423	3173.83835	3606.45797	<b>545.70845</b>
enthalpy of formation	186.20484	123.64866	135.32938	<b>10.15607</b>
enthalpy of fusion	13.58065	14.34938	15.46400	<b>2.15223</b>
enthalpy of vaporization	9.32412	3.97332	6.68632	<b>2.50873</b>
entropy of formation	543.36403	606.29039	688.57106	<b>132.44295</b>
entropy of gas	258.35003	247.96703	218.72920	<b>38.32509</b>

Table 22 | Results of transport properties.

Property	Mole-BERT	Unimolv1	Unimolv2	Suiren-ConfAvg
thermal conductivity of gas	0.00227	0.00139	0.00148	<b>0.00054</b>
thermal conductivity of liquid	0.00331	0.00276	0.00316	<b>0.00193</b>
diffusion coefficient at infinite dilution in water	<b>0.00000</b>	<b>0.00000</b>	<b>0.00000</b>	<b>0.00000</b>
diffusion coefficient in air	0.00487	0.00418	0.00363	<b>0.00242</b>
viscosity of liquid	1.71584	1.40034	1.83852	<b>1.10812</b>

- *Energetic properties* (Table 21): This category exhibits the most pronounced performance disparity. Baseline methods suffer severe degradation (e.g., MoleBERT’s MAE for enthalpy of combustion exceeds 4400), whereas Suiren-ConfAvg maintains substantially lower errors (545.71), representing an 82.8% relative improvement. We hypothesize that the pre-training objective of Suiren-Base, which incorporates physics-informed energy constraints, enables more transferable representations for thermodynamic quantities that are sensitive to subtle electronic and conformational features.
- *Critical and saturation properties* (Table 15): Suiren-ConfAvg dominates in critical temperature (21.75 vs. 63.35, best baseline) and critical volume (59.15 vs. 290.13), with relative improvements of 65.7% and 79.6%, respectively. However, for critical density and critical pressure, MoleBERT achieves marginally better results (0.00671 vs. 0.00690 and 1.737 vs. 1.932). This suggests that certain intensive properties with strong linear correlations to molecular size may be adequately captured by simpler architectures, whereas extensive or composite properties benefit from Suiren’s enhanced representation capacity.
- *Thermal and fluctuation properties* (Tables 19–17): The density of liquid shows minimal performance variation across splits (Suiren-ConfAvg: 0.01023), consistent with its strong dependence on atomic composition rather than scaffold topology. In contrast, heat capacities and thermal expansion coefficients exhibit substantial scaffold sensitivity, where Suiren-ConfAvg achieves 59.7%–86.8% relative improvements.
- *Structural properties* (Table 20): While Suiren-ConfAvg excels in liquid volume (70.9% improvement) and radius of gyration (60.3% improvement), it is slightly outperformed by Uni-Mol v1 on refractive index and dipole moment.
- *Safety, solution, and transport properties* (Tables 16–22): Suiren-ConfAvg consistently attains the best or near-best performance, with particularly notable gains in flash point (36.7% improvement) and vapor pressure (56.2% improvement).

**Statistical considerations and limitations** While the scaffold split provides a rigorous test of extrapolation, we acknowledge that the reported MAE values are point estimates without confidence intervals. Future evaluations should incorporate bootstrap resampling or cross-validation over multiple scaffold partitions to assess result stability. Additionally, the observed performance gaps may partially reflect differences in model capacity and pre-training data scale.

**Implications for molecular foundation models** The pronounced robustness of Suiren-ConfAvg on energetically complex and scaffold-sensitive properties suggests that integrating physics-aware pre-training objectives with uncertainty-aware inference mechanisms can substantially improve out-of-distribution generalization.