Learning Inter-Atomic Potentials without Explicit Equivariance

Anonymous authors
Paper under double-blind review

Abstract

Accurate and scalable machine-learned inter-atomic potentials (MLIPs) are essential for molecular simulations ranging from drug discovery to new material design. Current state-of-the-art models enforce roto-translational symmetries through equivariant neural network architectures, a hard-wired inductive bias that can often lead to reduced flexibility, computational efficiency, and scalability. In this work, we introduce **TransIP**: **Trans**former-based **I**nter-Atomic **P**otentials, a novel training paradigm for interatomic potentials achieving symmetry compliance without explicit architectural constraints. Our approach guides a generic non-equivariant Transformer-based model to learn SO(3)-equivariance by optimizing its representations in the embedding space. Trained on the recent Open Molecules (OMol25) collection, a large and diverse molecular dataset built specifically for MLIPs and covering different types of molecules (including small organics, biomolecular fragments, and electrolyte-like species), TransIP attains comparable performance in machine-learning force fields versus state-of-the-art equivariant baselines. Further, compared to a data augmentation baseline, TransIP achieves 40% to 60% improvement in performance across varying OMol25 dataset sizes. More broadly, our work shows that learned equivariance can be a powerful and efficient alternative to equivariant or augmentation-based MLIP models.

1 Introduction

Atomistic simulations are a fundamental task in chemistry and materials science (Zhang et al., 2018; Deringer et al., 2019), with Density Functional Theory (DFT) serving as a basis for accurately calculating interatomic forces and energies. However, the utility of DFT is severely restricted by its computational costs, which typically scale cubically with system size, rendering large-scale or long-timescale simulations intractable. This has motivated machine-learned interatomic potentials (MLIPs) to overcome this limitation by learning the potential energy surface from data, offering orders-of-magnitude speed-ups compared to DFT calculations (Noé et al., 2020; Batzner et al., 2022; Batatia et al., 2022; Jacobs et al., 2025; Leimeroth et al., 2025).

Equivariant neural networks have become a central paradigm for MLIPs due to their ability to encode the three-dimensional structure of molecular graphs (Anderson et al., 2019; Thölke & Fabritiis, 2022; Liao et al., 2024; Fu et al., 2025). These architectures are designed to explicitly respect roto-translational symmetries (SE(3) equivariance) by construction, often employing compute-intensive mechanisms like spherical harmonics or equivariant message passing (Fuchs et al., 2020; Passaro & Zitnick, 2023a; Liao & Smidt, 2023; Maruf et al., 2025). However, due to the design difficulties and limited expressive power of these architectures (Joshi et al., 2023; Cen et al., 2024), a recent trend in predictive and generative modeling is to use unconstrained models when enough data is available (Wang et al., 2024; Abramson et al., 2024; Zhang et al., 2025; Joshi et al., 2025).

In this paper, we introduce **TransIP** (**Trans**former-based **I**nteratomic **P**otentials), a training paradigm that achieves molecular symmetry for interatomic potentials *without* imposing architectural SO(3) constraints. TransIP steers a standard transformer toward SO(3) equivariance via an additional contrastive objective, allowing the model to retain the scalability and hardware efficiency of attention mechanisms while learning symmetry from data.

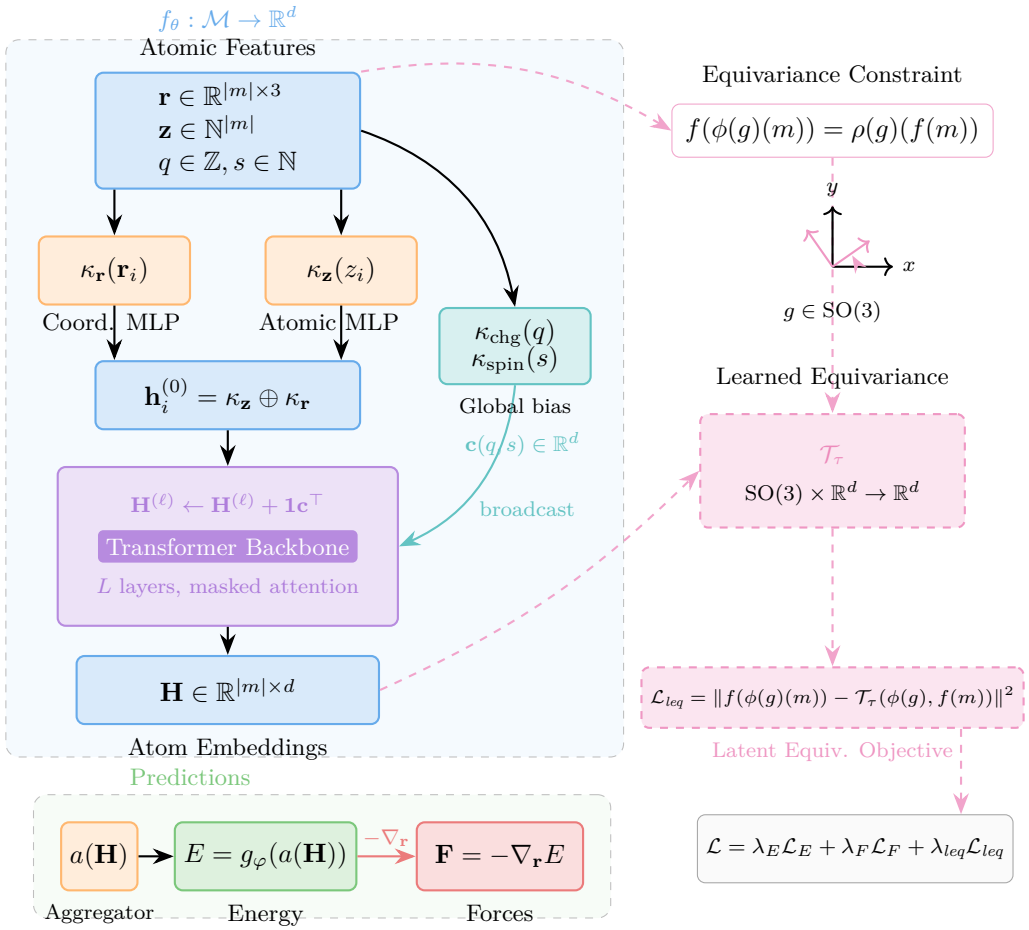

Figure 1: TransIP: Transformer-based Interatomic Potentials.

Our contributions are as follows:

- We propose a single-stage MLIP training pipeline with a general transformer-based model to obtain SO(3) equivariance through training, rather than hard-wired equivariant layers or a separate pretraining–fine-tuning framework.
- We introduce an architecture-agnostic contrastive loss function that promotes SO(3) equivariance in the embedding space of an unconstrained model. By aligning latent features across SO(3) transformations in the model's backbone, we show that TransIP scales better across different datasets and model sizes compared to traditional data augmentation techniques.
- On a diverse molecular benchmark, Open Molecules 25 (Levine et al., 2025) (that includes small organics, biomolecular fragments, electrolyte-like species), we show that TransIP outperforms data augmentation techniques often by a large margin, and achieves comparable performance versus current state-of-the-art MLIP baselines.

2 Symmetry in Embedding Space

2.1 Problem Formulation

Molecular representations. Let $\mathcal{M}$ denote the space of molecular configurations. Each molecule $m \in \mathcal{M}$ is represented by atomic features $\mathbf{x} = (\mathbf{r}, \mathbf{z}, q, s)$, where $\mathbf{r} \in \mathbb{R}^{|m|\times 3}$ are atomic coordinates, $\mathbf{z} \in \mathbb{N}^{|m|}$ are

atomic numbers, $q \in \mathbb{Z}$ is the total molecular charge, and $s \in \mathbb{N}$ is the spin multiplicity, with $|m|$ denoting the number of atoms in the molecule m.

Our goal is to learn an embedding function $f_\theta : \mathcal{M} \to \mathbb{R}^d$ that maps molecular configurations to a d-dimensional latent space, and a prediction function $g_\varphi : \mathbb{R}^d \to \mathbb{R}$ that acts in the embedding space $\mathbb{R}^d$ and outputs molecular properties (e.g., energy). Both f_θ and g_φ are neural networks parameterized by θ and φ, respectively.

Symmetry groups. We define a symmetry group G that acts on a set $\mathcal{X}$ as a group of bijective functions from $\mathcal{X}$ to itself, and the group operation is function composition. We say a function f is *equivariant* w.r.t. the group G if for every transformation $g \in G$ and every input $x \in X$,

$$f(\phi(g)(x)) = \rho(g)(f(x)) \tag{1}$$

The group representations ϕ and ρ specify how we apply the elements of the group G on input and output data. As a concrete case, we can define G as a rotation group SO(3) over molecular configurations $\mathcal{M}$, with $g \in \mathrm{SO}(3)$ representing an element of G that acts on a molecule m by rotating the coordinates of each atom in 3D space. Formally, for a molecule $m = (\mathbf{r}, \mathbf{z}, q, s)$ with coordinates $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_{|m|})$, $\mathbf{r}_i \in \mathbb{R}^3$, the input action rotates each atom:

$$\big(\phi(g)\, m\big) = \big((R\mathbf{r}_1, \ldots, R\mathbf{r}_{|m|}),\, \mathbf{z},\, q,\, s\big).$$

Here R is a 3×3 rotation matrix (orthogonal with $\det R = 1$); $\mathbf{z}, q, s$ are unchanged. An associated output representation rotates vector-valued quantities—e.g., for forces $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_{|m|})$, $\rho(g)\mathbf{F} = (R\mathbf{F}_1, \ldots, R\mathbf{F}_{|m|})$—while scalar outputs such as energies remain invariant, $\rho(g)E = E$.

2.2 Implicit Equivariance in Embedding Space

We seek an embedding function f that behaves equivariantly with respect to the symmetry group G, meaning there exists a transformation $\rho(g) : \mathbb{R}^d \to \mathbb{R}^d$ such that:

$$f(\phi(g)(m)) = \rho(g)(f(m)) \quad \forall g \in G, m \in \mathcal{M} \tag{2}$$

Common approaches enforce equivariance constraints through specialized architectures. Instead, we want the embedding function f to learn symmetry without equivariance constraints. However, with G being the rotation group SO(3) on $\mathcal{M}$ and the output of f being a high-dimensional vector, there is no direct representation of $\rho(g)$ to act in the space of $\mathbb{R}^d$. Thus, rather than specifying $\rho(g)$ analytically, we propose to learn the group transformation on an embedding vector in $\mathbb{R}^d$ using a neural network $\mathcal{T}_\tau : \mathrm{SO}(3) \times \mathbb{R}^d \to \mathbb{R}^d$ parameterized by τ. $\mathcal{T}$ can be understood as a non-linear function that learns the group action implicitly on a latent vector, by providing the group representation on the input data.

3 Learning Inter-Atomic Potentials without Explicit Equivariance

In this section, we introduce our training framework: TransIP (Transformer-based Inter-atomic Potentials), a new approach that achieves SO(3)-equivariance through learned transformations in an embedding space without explicit equivariance constraints. Our method, illustrated in Figure 1, consists of three key components: (i) an unconstrained Transformer backbone that processes molecular configurations, (ii) a learned transformation network that performs group actions in the embedding space, and (iii) a contrastive objective that enforces latent equivariance (equiv.) during training.

3.1 TransIP: Transformer-based Interatomic Potentials

Atom as tokens. We model each molecule as a variable-length sequence of tokens, where each token represents an atom. Unlike conventional graph neural networks that construct edges based on distance cutoffs or neighbours' atoms, we process all atoms within a molecule through self-attention, bounded by a maximum context length N_{ctx}. For batch processing, we use padding masks to prevent cross-molecule attention, ensuring each molecule is processed independently.

Transformer Backbone. We implement the embedding function $f_\theta : \mathcal{M} \to \mathbb{R}^d$ as a Transformer encoder that processes atom-level tokens. Each atom i is initialized with a token representation:

$$\mathbf{h}_i^{(0)} = \kappa_{\mathbf{z}}(z_i) \oplus \kappa_{\mathbf{r}}(\mathbf{r}_i)$$

where $\kappa_{\mathbf{z}} : \mathbb{N} \to \mathbb{R}^d$ and $\kappa_{\mathbf{r}} : \mathbb{R}^3 \to \mathbb{R}^d$ are learnable MLPs that embed atomic numbers and centered coordinates (with $\mathbf{r}_i \leftarrow \mathbf{r}_i - \frac{1}{|m|}\sum_j \mathbf{r}_j$), and $\oplus$ denotes concatenation. These tokens are processed through L Transformer layers with masked self-attention within each molecule, producing final per-atom embeddings $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_{|m|}]^\top \in \mathbb{R}^{|m| \times d}$.

Global Molecular Properties. Following Levine et al. (2025), we incorporate global molecular properties (total charge q and spin multiplicity s of a molecule m) through learnable embeddings, and form a graph-level bias:

$$\mathbf{c}(q, s) = \kappa_{\text{chg}}(q) + \kappa_{\text{spin}}(s) \in \mathbb{R}^d$$

where κ_{chg} and κ_{spin} are learnable embedding functions for charge and spin, respectively. This global bias is broadcast-added at each Transformer layer: $\mathbf{H}^{(\ell)} \leftarrow \mathbf{H}^{(\ell)} + \mathbf{1}\mathbf{c}(q, s)^\top$.

Energy and Force Predictions. For molecular property prediction, we employ a permutation-invariant aggregator $a : \mathbb{R}^{|m| \times d} \to \mathbb{R}^d$ followed by an energy prediction head $g_\varphi : \mathbb{R}^d \to \mathbb{R}$:

$$E_\varphi(m) = g_\varphi(a(\mathbf{H}))$$

Forces are computed as conservative gradients of the energy with respect to atomic positions:

$$\mathbf{F}(m) = -\nabla_{\mathbf{r}} E_\varphi(m) \in \mathbb{R}^{|m| \times 3}$$

3.2 Learned Latent Equivariance

Transformation Network. We propose a transformation network $\mathcal{T}_\tau : \mathrm{SO}(3) \times \mathbb{R}^d \to \mathbb{R}^d$ that learns how group actions (e.g., rotations) act on molecular embeddings. We implement $\mathcal{T}_\tau$ as a multilayer perceptron that takes as input the group representation in the input domain $\phi(g)$ and the molecular embedding $f(m)$. Formally,

$$\mathcal{T}_\tau(\phi(g), f(m)) = \mathrm{MLP}_\tau([\phi(g), f(m)])$$

where $[\cdot, \cdot]$ denotes concatenation and MLP_τ is a multilayer perceptron with parameters τ.

Contrastive Objective for Latent Equivariance: To learn the molecular symmetry without architectural constraints, we define our latent equivariance loss as:

$$\mathcal{L}_{leq}(\phi(g), m, f, \mathcal{T}) = \|f(\phi(g)(m)) - \mathcal{T}_\tau(\phi(g), f(m))\|^2 \tag{3}$$

This loss encourages the embedding function f to behave equivariantly with respect to the symmetry group G, as mediated by the transformation network $\mathcal{T}_\tau$. During training, we sample a molecule m from the dataset and a rotation element g uniformly from $\mathrm{SO}(3)$ and minimize the expected latent loss:

$$\min \mathbb{E}_{m \sim \mathcal{M}, g \sim \mathrm{SO}(3)}[\mathcal{L}_{leq}(\phi(g), m, f, \mathcal{T})] \tag{4}$$

3.3 Training Objective

Our training objective combines three complementary losses in a *single-stage* framework for accurate prediction of energy and forces as well as implicit learning of molecular symmetry.

Prediction Losses. For energy and force predictions, we use:

$$\mathcal{L}_E = \tfrac{1}{|m|}|E_\varphi(m) - E^\star| \quad \text{(per-atom mean absolute error (MAE))} \tag{5}$$

$$\mathcal{L}_F = \tfrac{1}{3|m|}\|\mathbf{F}(m) - \mathbf{F}^\star\|_F^2 \quad \text{(per-molecule mean squared error (MSE))} \tag{6}$$

where $E^\star$ and $\mathbf{F}^\star$ are ground-truth energies and forces, and $\|\cdot\|_F$ denotes the Frobenius norm. For energies, we use referenced targets as described by Levine et al. (2025).

Combined Objective. Training combines three weighted terms: (i) the latent equivariance target $\mathcal{L}_{leq}$ defined in Eq. 3; (ii) energy loss $\mathcal{L}_E$; and (iii) force loss $\mathcal{L}_F$. The total objective is

$$\mathcal{L}_{\text{total}} = \lambda_E \mathcal{L}_E + \lambda_F \mathcal{L}_F + \lambda_{leq} \mathcal{L}_{leq} \tag{7}$$

where λ_E, λ_F, and λ_{leq} are hyperparameters for each loss. The optimal hyperparameters are given in Table 5 of Appendix A.

4 Related Work

4.1 ML Interatomic Potentials

Using machine learning (ML) methods to predict energies and forces of different molecular systems and materials has been an active area of research (Schütt et al., 2017; Chmiela et al., 2022; Musaelian et al., 2023; Liao et al., 2024; Yang et al., 2025; Yuan et al., 2025). Due to the intricate 3D structures of atomistic systems, equivariant designs such as steerable convolution (Cohen & Welling, 2017; Brandstetter et al., 2022) and higher-order tensors (Thomas et al., 2018), as well as covariant representation (Anderson et al., 2019), have been essential backbones for modeling molecular systems. For example, Gasteiger et al. (2020); Klicpera et al. (2021) introduced equivariant directional message passing between pairs of atoms with a spherical harmonics representation. In contrast, Batzner et al. (2022) developed equivariant convolution with tensor-products and Batatia et al. (2022) built higher-order messages with equivariant graph neural networks (Satorras et al., 2021). Additionally, Passaro & Zitnick (2023b) reduced the computational complexity of SO(3) convolution and replaced it with SO(2) convolutions, which have been used as a backbone for MLIPs (Fu et al., 2025). More recently, Rhodes et al. (2025) presented Orb-v3 models with improved computational efficiency, built on Graph Network Simulators (Sanchez-Gonzalez et al., 2020).

4.2 Unconstrained ML models

While current-state-of-the-art MLIP models primarily rely on equivariant GNNs, unconstrained models are actively used in other domains. For example, integrating data augmentation via image transformations has been used in different vision tasks, from classification (Inoue, 2018; Dosovitskiy et al., 2021; Rahat et al., 2024) to segmentation (Negassi et al., 2022; Yu et al., 2023). For geometric data, the use of unconstrained models and diffusion Transformers (without explicit equivariance constraints) has been a recent trend in generative tasks, e.g., AlphaFold 3 for biomolecular structure prediction (Abramson et al., 2024) as well as molecular conformation and materials generation (Wang et al., 2024; Zhang et al., 2025; Joshi et al., 2025). In contrast, several works have been introduced to overcome the limitations of strictly equivariant GNNs by enforcing symmetry via frame averaging over geometric inputs (Puny et al., 2022; Duval et al., 2023; Lin et al., 2024; Huang et al., 2024; Dym et al., 2024); learning canonicalization functions that map inputs to a canonical orientation before prediction (Kaba et al., 2022; Baker et al., 2024; Ma et al., 2024; Lippmann et al., 2025); or learning equivariance through data augmentation with molecule-specific graph-based architectures (Qu & Krishnapriyan, 2024; Mazitov et al., 2025). However, in this work, we demonstrate that an unconstrained general-purpose Transformer model can serve as a backbone for MLIPs. We replace graph-based inductive biases with a scalable latent equivariance objective that implicitly learns equivariant features in a single training stage, without explicit equivariance constraints or a pretraining-finetuning framework.

5 Experimental Setup

Dataset. We train and evaluate our proposed method **TransIP** on the Open Molecules 2025 (OMol25) collection (Levine et al., 2025), a large-scale molecular DFT dataset for ML interatomic potentials. OMol25 covers 83 atomic elements and diverse chemistries including: metal complexes, electrolytes, biomolecules, SPICE, neutral organic, and reactivity. Following Levine et al. (2025), we use the official `4M` training split

(3,986,754) and the out-of-distribution composition validation split *Val-Comp* (2,762,021). *Val-Comp* consists of molecules gathered from various datasets and domains, such as biomolecules, neutral organics, and metal complexes.

Model Configurations. We evaluate TransIP across three model scales: Small (14M parameters), Medium (85M parameters), and Large (302M parameters). All models use MLP-based coordinate embeddings. The transformation network $\mathcal{T}_\tau$ is a 2-layer MLP with GELU activations and 2d hidden dimension.

Training Setup. Using the standardized FAIRCHEM Python package (Shuaibi et al., 2025), we train TransIP on the OMol25 dataset using an AdamW optimizer with learning rate 5×10^{-4}, weight decay 10^{-3}, and gradient norm clipping at 100. We use a cosine learning rate schedule with linear warmup over the first 1% of training, followed by cosine decay down to 1% of the initial `lr`. The loss weights are set to $\lambda_E = 5$ for energies and $\lambda_F = 15$ for forces. For the latent equivariance objective λ_{leq}, we sweep the values in $\{1, 5, 10, 100\}$ and selected $\lambda_{leq} = 5$ based on validation performance.

Scalability Experiments. We conduct three sets of experiments to assess TransIP's scaling behavior:

- **Data scaling**: We train the Small (14M parameter) model on three dataset sizes (1M, 2M, 4M molecules) for 5 epochs using 8 NVIDIA 80GB GPUs, comparing TransIP with learned equivariance against an unconstrained Transformer version with SO(3) data augmentation (TransAug).
- **Model size scaling.** We compare TransIP and TransAug with different model sizes (Small/Medium/Large) trained on the same number of samples from the OMol25 4M dataset and report the evaluation metrics as a function of the processed number of atoms per second.
- **Extended training**: We train TransIP models (Small, Medium, Large) on the OMol25 4M dataset for 80 epochs using 32 NVIDIA 80GB GPUs to evaluate its performance against current state-of-the-art equivariant baselines.

Baselines. We compare TransIP against: (i) an *unconstrained* TransIP variant trained with SO(3) rotation augmentation to assess the impact of learned latent equivariance versus data augmentations, and (ii) state-of-the-art equivariant models on OMol25: `eSCN` (Fu et al., 2025) in small/medium configurations with both direct and energy-conserving force variants as well as `GemNet-OC` (Gasteiger et al., 2022).

Evaluation metrics. Following the OMol25 official benchmark, we report: Force MAE (meV/Å), Force cosine similarity, Energy per atom MAE (meV/atom), and Total energy MAE (meV). Detailed metric definitions are provided in Appendix A.5.

6 Results and Discussion

6.1 Scaling data size

To assess how performance scales with different training dataset sizes, we compare our latent equivariance-based model (TransIP) against an unconstrained baseline that uses SO(3) data augmentation (TransAug). Both models use a (small) 14M parameter Transformer architecture. Given our tight compute budget, we train on 1M, 2M, and 4M OMol25 molecules for 5 epochs and report validation (*Val-Comp*) results.

Performance in a limited data regime. Figure 2 shows that TransIP delivers large gains when trained on 1M samples and outperforms TransAug across all evaluation metrics with a large margin on the total validation split. We also include the performance comparison for each molecule category in Appendix B. In Figure 2, the learned latent equivariance objective provides substantial improvements in force MAE (255 meV/Å vs 600 meV/Å MAE) and directional consistency (0.7 vs 0.44 force cosine similarity). Energy predictions also benefit from the latent equivariance objective, with TransIP achieving 58 meV/atom compared to TransAug's 120 meV/atom. These results suggest that learning equivariance in a latent space is a more effective scheme to incorporate molecular symmetry than data augmentation, particularly when training data is limited.

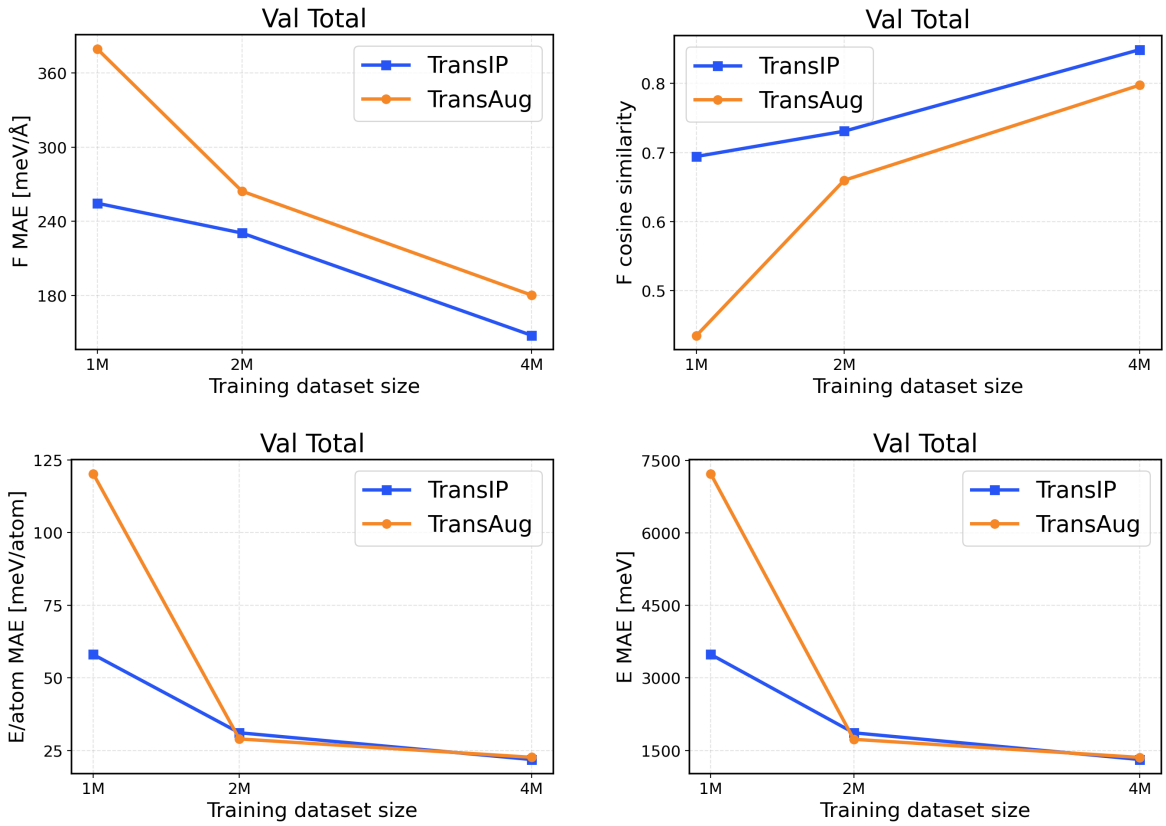

Figure 2: Val-Comp performance across different dataset sizes (1M / 2M / 4M): The top row presents force metrics, while the bottom row reports energy metrics.

Performance in a larger data regime. As we scale to 2M and 4M molecules, both models (TransIP and TransAug) improve across the evaluation metrics. However, on larger datasets, TransIP still achieves better force MAEs and cosine similarity metrics compared to TransAug. This might indicate that the learned transformation network can successfully capture the geometric relationships necessary for accurate force predictions. Notably, energy prediction performance converges between the two at larger data scales, with both methods achieving comparable per-atom MAE values. This convergence suggests that while learned equivariance provides crucial benefits for force-related metrics in all data regimes, its advantages for energy prediction become less pronounced as the model can learn invariant energy representations from sufficient augmented data.

6.2 Learned latent equivariance

We investigate how learned equivariance affects the embedding space in relation to validation performance as the data scale increases. Figure 3 plots each metric against latent equiv. error for TransIP (Small) trained for 5 epochs on 1M, 2M, and 4M molecules (see Table 3 for a detailed definition of each model configuration).

Lower latent equivariance error leads to better accuracy. We found that the learned equiv. error serves as a strong predictor of model performance. Across all metrics, we observe a clear monotonic trend: lower equiv. error is associated with better performance (Figure 3). However, energy and force predictions respond differently to improvements in equivariance. Energy predictions show near-linear scaling with equiv. error, indicating that energy accuracy is directly limited by equivariance quality. This strong coupling aligns with energies being scalar invariants that depend primarily on learning correct symmetry-preserving features. In contrast, force predictions exhibit a two-regime behavior: initial improvements in equivariance (1M→2M) yield modest force improvements, while further tightening of equivariance (2M→4M) produces

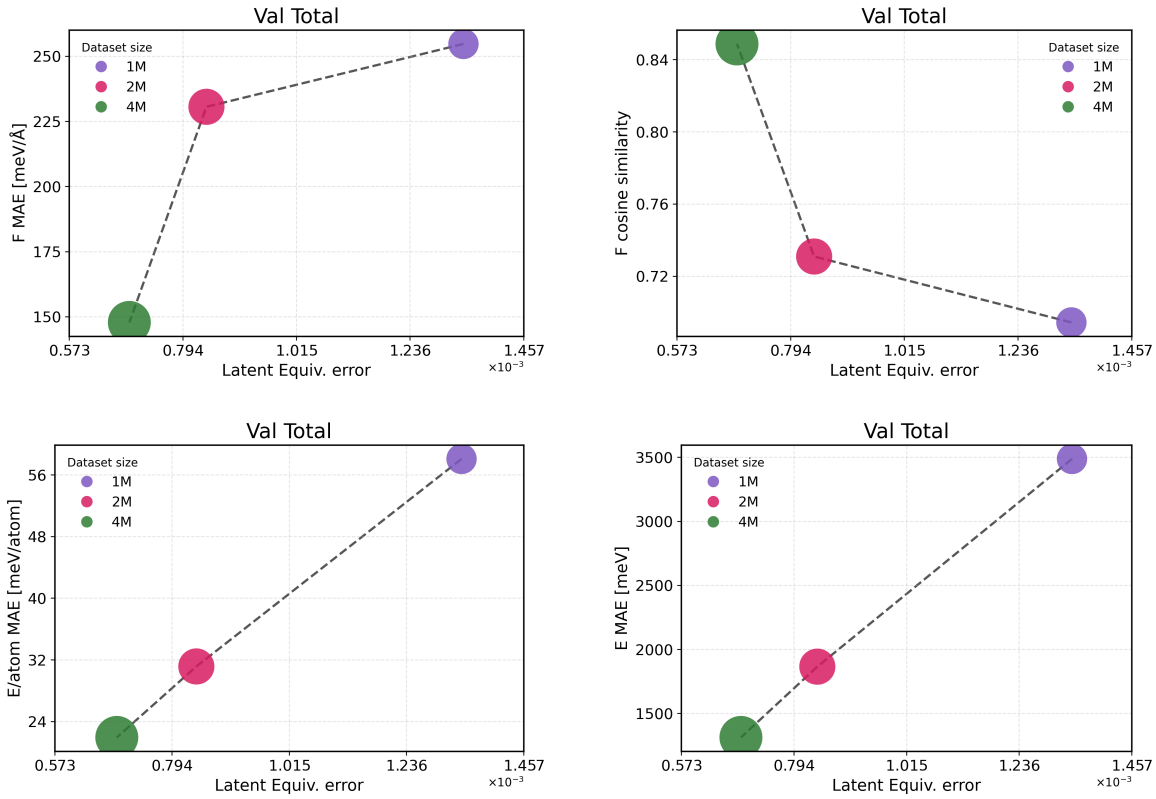

Figure 3: Latent equivariance (embedding) error versus validation performance. The top row reports force metrics, while the bottom row presents energy metrics.

disproportionate gains. This might indicate that forces require both accurate equivariant features and sufficient data diversity to learn the energy landscape's geometry.

These results demonstrate that implicitly learning equivariance through our learned transformation network provides an efficient inductive bias, accelerating learning. The 48% reduction in equiv. error from 1M to 4M training examples translates to 40-60% performance improvements, being more efficient than what would be expected from data scaling alone.

Learning equivariance leads to faster inference. To measure the inference efficiency of our method, we compare TransIP and TransAug with different model sizes (Small/Medium/Large) trained on 4M samples and report the evaluation metrics as a function of the processed number of atoms per second. However, due to limited compute, we compare models under a *fixed training budget* (i.e., with the same number of samples), which is 10k, 25k, and 100k steps for our Small, Medium, and Large models, respectively.

From the results in Figure 4, we see that TransIP scales smoothly with parameter count despite limited training: As model size grows, performance improves across all metrics. In contrast, TransAug exhibits poorer scaling—larger models perform worse than smaller ones, with the Large model configuration yielding the lowest performance. This might indicate that augmentation alone does not provide a sufficiently informative and stable inductive bias for large-capacity models trained for molecular force field prediction.

6.3 Architectural equivariance versus learned equivariance

Table 1 compares the energy and force prediction performance of TransIP-L trained for 80 epochs against several well-known equivariant baselines on the OMol 2M Val-Comp evaluation dataset. For a fair comparison, we use Gaussian radial basis functions (RBF) encodings of interatomic distances (Behler & Parrinello, 2007) in the 80-epoch runs, which are widely used in MLIPs models, including equivariant baselines (Fu

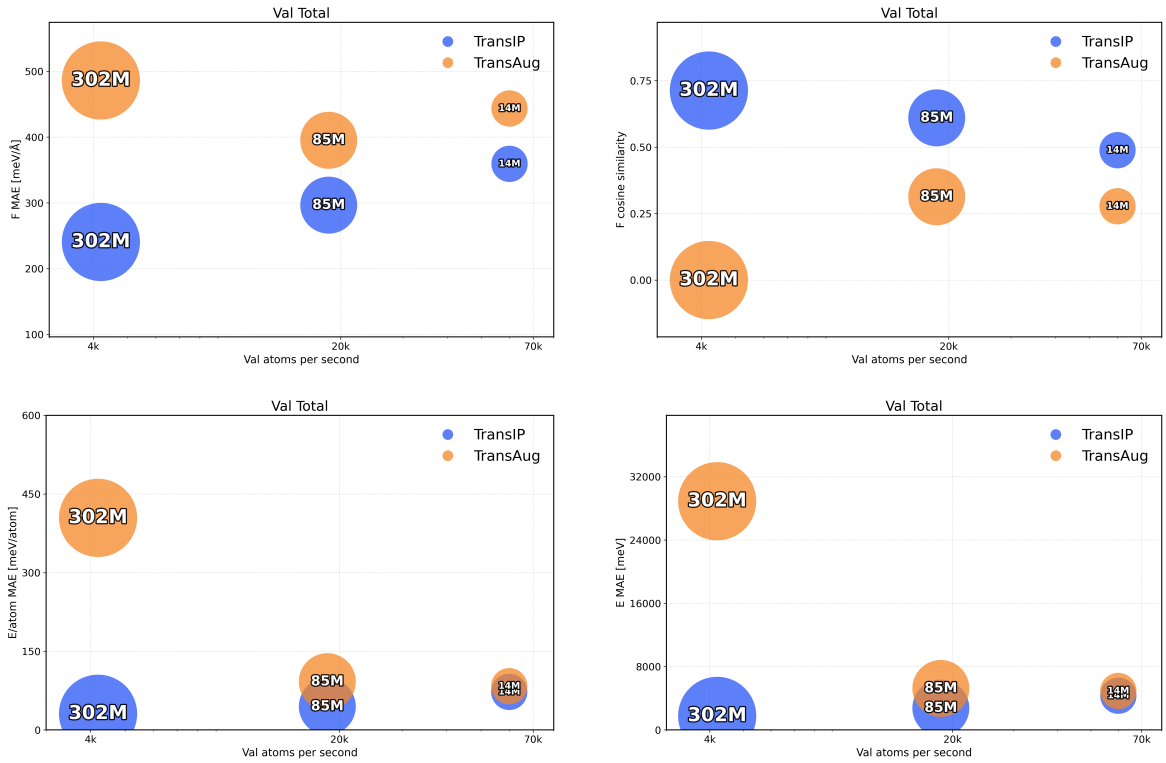

Figure 4: Validation total inference trade-off (atoms/s versus performance). The top row presents force metrics, while the bottom row represents energy metrics.

et al., 2025; Levine et al., 2025). We incorporate RBF features at two levels: as aggregated local distance features added to each atom token, and as additive bias on the attention logits. We report additional ablations on the effect of RBF features in Appendix B.1. We also include the performance of all TransIP variants in Table 8.

The results in Table 1 demonstrate that TransIP-L outperforms eSEN-sm in total energy and is competitive in total force MAE. Furthermore, TransIP-L outperforms MACE in total energy and competes in total force metrics despite MACE being trained with the total OMol25 dataset (`102M`) (Levine et al., 2025) and TransIP-L being trained with the `4M` subset.

We report the inference speed for our TransIP model and eSEN baselines on the same hardware using a single H200 NVIDIA GPU in Table 2. For eSEN, we follow the small and medium versions indicated by Levine et al. (2025) with hyperparameters in Table 4. Our results show that TransIP-L is 1.5× faster than eSEN-small and 5× faster than eSEN-medium.

Table 1: Comprehensive Val-Comp. E/atom MAE (`meV/atom`), E MAE (`meV`), and F MAE (`meV/Å`) results.

	Biomolecules			Electrolytes			Metal Complexes			Neutral Organics			Total		
Model	**E/atom ↓**	**E ↓**	**F ↓**	**E/atom ↓**	**E ↓**	**F ↓**	**E/atom ↓**	**E ↓**	**F ↓**	**E/atom ↓**	**E ↓**	**F ↓**	**E/atom ↓**	**E ↓**	**F ↓**
eSEN-sm-d.	0.88	127.43	8.12	1.93	134.78	12.64	3.37	192.77	40.44	2.16	59.84	20.17	2.19	129.77	13.01
eSEN-sm-cons.	0.86	125.74	6.17	1.61	117.33	11.16	2.72	156.54	35.33	1.50	42.07	16.92	1.89	114.81	11.10
eSEN-md-d.	0.47	70.13	3.38	1.18	78.67	6.51	2.53	142.55	27.31	1.21	33.37	9.26	1.32	77.13	6.78
GemNet-OC-r6	0.40	63.57	5.84	1.39	82.08	9.37	2.74	152.87	33.60	1.88	47.96	16.55	1.41	79.67	9.83
GemNet-OC	0.25	39.58	5.20	1.04	56.32	8.42	2.66	148.49	32.76	1.64	40.98	15.59	1.13	57.82	8.98
MACE-OMol-L-0[1]	–	265.43	6.63	–	200.36	11.06	–	162.53	31.72	–	56.98	15.10	–	197.74	10.94
TransIP-L	0.60	89.40	11.00	1.30	86.30	14.20	3.30	197.60	38.50	2.40	69.40	23.80	1.60	91.20	14.70

Table 2: Inference speed for TransIP-L and eSEN baselines.

	TransIP-L	eSEN-small	eSEN-medium
Approx. atoms/sec	9,500	6,300	1,800

7 What TransIP Learns

To understand the structure of learned equivariance, we ask whether the effect of rotating different inputs can be explained by a *single* group action in the latent space; i.e., whether there exists a representation $\rho(g) : \mathbb{R}^d \to \mathbb{R}^d$ such that $f\big(\phi(g)(m)\big) \approx \rho(g)\, f(m)$, where $f_\theta : \mathcal{M} \to \mathbb{R}^d$ denotes the embedding network, and $g \in \mathrm{SO}(3)$ acts on a molecule m via the input representation $\phi(g)$ (rotation of atomic coordinates). Because $\rho(g)$ is unknown, we compute an approximate group action $\widehat{\rho}(g) \in \mathrm{O}(d)$ by solving an orthogonal Procrustes problem on embeddings from 100 validation samples (obtained from a trained TransIP model). Writing $Z = [\, f(m_1)^\top, \ldots, f(m_n)^\top \,], \quad Z_g = [\, f(\phi(g)(m_1))^\top, \ldots, f(\phi(g)(m_n))^\top \,]$, we first pool-whiten the two views (shared mean and standard deviation per channel) and then solve $\widehat{\rho}(g) = \arg\min_{Q \in \mathrm{O}(d)} \| \widetilde{Z} Q - \widetilde{Z}_g \|_F^2$, which has the closed form $\widehat{\rho}(g) = UV^\top$ for the SVD of $\widetilde{Z}^\top \widetilde{Z}_g = U \Sigma V^\top$.

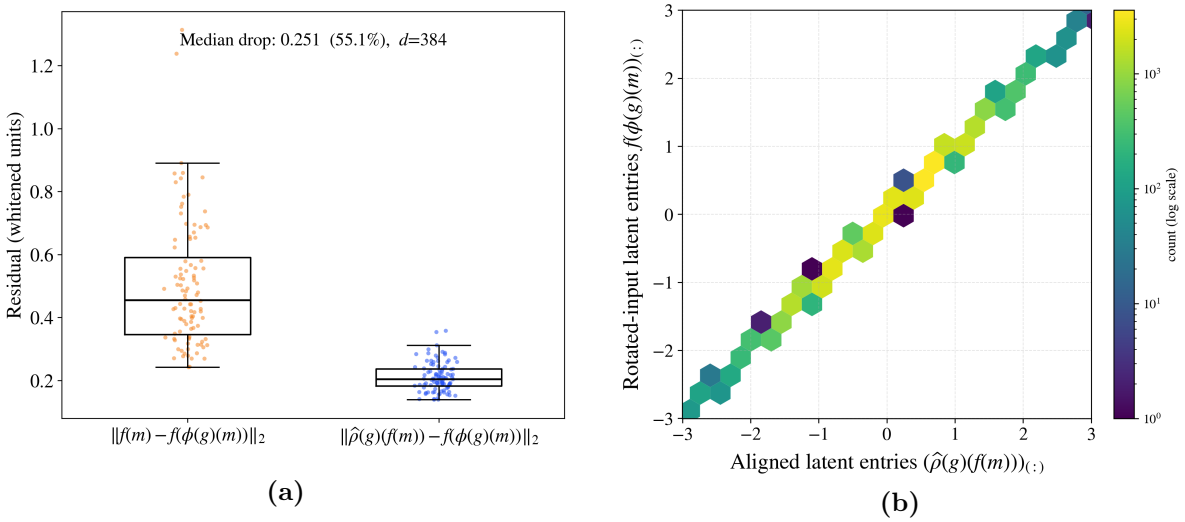

Figure 5: **Group action in the embedding space. (a)** Per-molecule residuals before alignment, $\|f(m) - f(\phi(g)(m))\|_2$, and after applying a global orthogonal map $\widehat{\rho}(g)$ on pool-whitened latents, $\|\widehat{\rho}(g) f(m) - f(\phi(g)(m))\|_2$. **(b)** Entrywise comparison: hexbin density of $(\widehat{\rho}(g) f(m))_{(:)}$ vs. $f(\phi(g)(m))_{(:)}$, pooled over molecules' embeddings.

In Figure 5a, we report per-molecule residuals before alignment, $\| f(m) - f(\phi(g)(m)) \|_2$, and after applying the global orthogonal map, $\| \widehat{\rho}(g) f(m) - f(\phi(g)(m)) \|_2$. A left→right drop in the distribution indicates that a single orthogonal transform explains most of the rotation-induced change in the embedding. In Figure 5b, we compare the channel-level relation by plotting a hexbin density of all pairs $(\widehat{\rho}(g) f(m))_k, \quad (f(\phi(g)(m)))_k, \quad k = 1, \ldots, d, \; m \in \text{val.}$ where color encodes the log count of points in each hexagonal bin. A tight diagonal concentration after the single global alignment $\widehat{\rho}(g)$ might suggest that the two views are almost identical at entrywise-level and the group action in latent space is *approximately orthogonal* and shared across different molecules.

Takeaways

Figure 5a shows that the magnitude of the rotation-induced discrepancy of different molecules drops after a single orthogonal alignment, and Figure 5b shows that the aligned channels match entrywise, concentrating along the identity. These results indicate that TransIP learns an embedding where input rotations act approximately as a shared orthogonal transformation, even though explicit equivariance was not enforced in the architecture.

8 Conclusion

In this work, we introduced TransIP for modeling interatomic potentials with a modern Transformer-based architecture and a scalable latent equivariance objective. Empirical results across a variety of chemical systems as well as model and dataset scales suggest that TransIP's latent equivariance objective enables better performance scaling than popular data augmentation-based alternatives to learning geometric equivariance. Further, we find that improvements in learning latent equivariance are strongly related to improved modeling of interatomic potentials, suggesting a complementary nature between the two prediction objectives. With sufficient compute, future work could involve studying the performance of TransIP in larger data, modeling, and runtime regimes in addition to the behavior of TransIP in a context amenable to the double-descent phenomenon (Power et al., 2022).

While equivariant models for molecular machine learning have recently gained much research interest, with the large amount of data being generated and the need for larger model sizes, it is also important that models used for interatomic potentials be highly scalable. Through our work, we have shown that the generic Transformer is capable of modeling molecules accurately but is also able to learn equivariance effectively through our novel latent objective, all while being highly scalable. By making our code openly available to the research community, we hope that our work inspires future research that explores ways to leverage the simpler and more scalable Transformer architecture to better model equivariant molecular properties through learned equivariance.

References

Jacob Abramson, Jane Adler, John Dunger, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630:493–500, 2024. doi: 10.1038/s41586-024-07487-w. URL https://doi.org/10.1038/s41586-024-07487-w.

Brandon Anderson, Truong Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural networks. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/03573b32b2746e6e8ca98b9123f2249b-Paper.pdf.

Justin Baker, Shih-Hsin Wang, Tommaso de Fernex, and Bao Wang. An explicit frame construction for normalizing 3d point clouds. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=SZ0JnRxi0x.

Ilyes Batatia, David Peter Kovacs, Gregor N. C. Simm, Christoph Ortner, and Gabor Csanyi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=YPpSngE-ZU.

Simon Batzner, Albert Musaelian, Lixin Sun, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13:2453, 2022. doi: 10.1038/s41467-022-29939-5. URL https://doi.org/10.1038/s41467-022-29939-5.

Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters*, 98(14):146401, 2007. doi: 10.1103/PhysRevLett.98.146401.

Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J Bekkers, and Max Welling. Geometric and physical quantities improve e(3) equivariant message passing. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=_xwr8gOBeV1.

Jiacheng Cen, Anyi Li, Ning Lin, Yuxiang Ren, Zihe Wang, and Wenbing Huang. Are high-degree representations really unnecessary in equivariant graph neural networks? In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=M0ncNVuGYN.

Stefan Chmiela, Valentin Vassilev-Galindo, Oliver T. Unke, Adil Kabylda, Huziel E. Sauceda, Alexandre Tkatchenko, and Klaus-Robert Müller. Accurate global machine learning force fields for molecules with hundreds of atoms, 2022. URL https://arxiv.org/abs/2209.14865.

Taco S. Cohen and Max Welling. Steerable CNNs. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=rJQKYt5ll.

Volker L. Deringer, Miguel A. Caro, and Gábor Csányi. Machine learning interatomic potentials as emerging tools for materials science. *Advanced Materials*, 2019.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Alexandre Agm Duval, Victor Schmidt, Alex Hernández-García, Santiago Miret, Fragkiskos D. Malliaros, Yoshua Bengio, and David Rolnick. FAENet: Frame averaging equivariant GNN for materials modeling. In *Proceedings of the 40th International Conference on Machine Learning*, 2023. URL https://proceedings.mlr.press/v202/duval23a.html.

Nadav Dym, Hannah Lawrence, and Jonathan W. Siegel. Equivariant frames and the impossibility of continuous canonicalization. In *ICML*, 2024. URL https://openreview.net/forum?id=4iy0q0carb.

Xiang Fu, Brandon M Wood, Luis Barroso-Luque, Daniel S. Levine, Meng Gao, Misko Dzamba, and C. Lawrence Zitnick. Learning smooth and expressive interatomic potentials for physical property prediction. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=R0PBjxIbgm.

Fabian B. Fuchs, Daniel E. Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 2020.

Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=B1eWbxStPH.

Johannes Gasteiger, Muhammed Shuaibi, Anuroop Sriram, Stephan Günnemann, Zachary Ward Ulissi, C. Lawrence Zitnick, and Abhishek Das. Gemnet-OC: Developing graph neural networks for large and diverse molecular simulation datasets. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=u8tvSxm4Bs.

Tinglin Huang, Zhenqiao Song, Rex Ying, and Wengong Jin. Protein-nucleic acid complex modeling with frame averaging transformer. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=Xngi3Z3wkN.

Hiroshi Inoue. Data augmentation by pairing samples for images classification, 2018. URL https://arxiv.org/abs/1801.02929.

Ryan Jacobs, Dane Morgan, Siamak Attarian, Jun Meng, Chen Shen, Zhenghao Wu, Clare Yijia Xie, Julia H. Yang, Nongnuch Artrith, Ben Blaiszik, Gerbrand Ceder, Kamal Choudhary, Gabor Csanyi, Ekin Dogus Cubuk, Bowen Deng, Ralf Drautz, Xiang Fu, Jonathan Godwin, Vasant Honavar, Olexandr Isayev,

Anders Johansson, Boris Kozinsky, Stefano Martiniani, Shyue Ping Ong, Igor Poltavsky, KJ Schmidt, So Takamoto, Aidan P. Thompson, Julia Westermayr, and Brandon M. Wood. A practical guide to machine learning interatomic potentials – status and future. *Current Opinion in Solid State and Materials Science*, 35:101214, March 2025. ISSN 1359-0286. doi: 10.1016/j.cossms.2025.101214. URL http://dx.doi.org/10.1016/j.cossms.2025.101214.

Chaitanya K. Joshi, Cristian Bodnar, Simon V Mathis, Taco Cohen, and Pietro Lio. On the expressive power of geometric graph neural networks. In *Proceedings of the 40th International Conference on Machine Learning*, 2023. URL https://proceedings.mlr.press/v202/joshi23a.html.

Chaitanya K. Joshi, Xiang Fu, Yi-Lun Liao, Vahe Gharakhanyan, Benjamin Kurt Miller, Anuroop Sriram, and Zachary W. Ulissi. All-atom diffusion transformers: Unified generative modelling of molecules and materials. In *International Conference on Machine Learning*, 2025.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *NeurIPS 2022 Workshop on Symmetry and Geometry in Neural Representations*, 2022. URL https://openreview.net/forum?id=pVD1k8ge25a.

Johannes Klicpera, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=HS_sOaxS9K-.

Niklas Leimeroth, Linus C. Erhard, Karsten Albe, and Jochen Rohrer. Machine-learning interatomic potentials from a users perspective: A comparison of accuracy, speed and data efficiency, 2025. URL https://arxiv.org/abs/2505.02503.

Daniel S. Levine, Muhammed Shuaibi, Evan Walter Clark Spotte-Smith, Michael G. Taylor, Muhammad R. Hasyim, Kyle Michel, Ilyes Batatia, Gábor Csányi, Misko Dzamba, Peter Eastman, Nathan C. Frey, Xiang Fu, Vahe Gharakhanyan, Aditi S. Krishnapriyan, Joshua A. Rackers, Sanjeev Raja, Ammar Rizvi, Andrew S. Rosen, Zachary Ulissi, Santiago Vargas, C. Lawrence Zitnick, Samuel M. Blau, and Brandon M. Wood. The open molecules 2025 (omol25) dataset, evaluations, and models, 2025. URL https://arxiv.org/abs/2505.08762.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=KwmPfARgOTD.

Yi-Lun Liao, Brandon Wood, Abhishek Das*, and Tess Smidt*. EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=mCOBKZmrzD.

Yuchao Lin, Jacob Helwig, Shurui Gui, and Shuiwang Ji. Equivariance via minimal frame averaging for more symmetries and efficiency. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=guFsTBXsov.

Peter Lippmann, Gerrit Gerhartz, Roman Remme, and Fred A Hamprecht. Beyond canonicalization: How tensorial messages improve equivariant message passing. In *International Conference on Representation Learning*, 2025.

George Ma, Yifei Wang, Derek Lim, Stefanie Jegelka, and Yisen Wang. A canonicalization perspective on invariant and equivariant learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=jjcY92FX4R.

Moin Uddin Maruf, Sungmin Kim, and Zeeshan Ahmad. Learning long-range interactions in equivariant machine learning interatomic potentials via electronic degrees of freedom. *The Journal of Physical Chemistry Letters*, 16(35):9078–9087, August 2025. ISSN 1948-7185. doi: 10.1021/acs.jpclett.5c02352. URL http://dx.doi.org/10.1021/acs.jpclett.5c02352.

Arslan Mazitov, Filippo Bigi, Matthias Kellner, Paolo Pegolo, Davide Tisi, Guillaume Fraux, Sergey Pozdnyakov, Philip Loche, and Michele Ceriotti. Pet-mad, a lightweight universal interatomic potential for advanced materials modeling, 2025. URL https://arxiv.org/abs/2503.14118.

Albert Musaelian, Sebastian Batzner, Anders Johansson, Simon Kozinsky, and Boris Kozinsky. Learning local 3d energetics with graph neural networks. *Nature Communications*, 2023.

Misgana Negassi, Diane Wagner, and Alexander Reiterer. Smart(sampling)augment: Optimal and efficient data augmentation for semantic segmentation. *Algorithms*, 15(5), 2022. ISSN 1999-4893. doi: 10.3390/a15050165. URL https://www.mdpi.com/1999-4893/15/5/165.

Frank Noé, Alexandre Tkatchenko, Klaus-Robert Müller, and Cecilia Clementi. Machine learning for molecular simulation. *Annual Review of Physical Chemistry*, 71(1):361–390, April 2020. ISSN 1545-1593. doi: 10.1146/annurev-physchem-042018-052331. URL http://dx.doi.org/10.1146/annurev-physchem-042018-052331.

Saro Passaro and C. Lawrence Zitnick. Reducing so(3) convolutions to so(2) for efficient equivariant gnns. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023a.

Saro Passaro and C. Lawrence Zitnick. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR, 2023b.

Alethea Power, Yuri Burda, Harri Edwards, Igor Babuschkin, and Vedant Misra. Grokking: Generalization beyond overfitting on small algorithmic datasets. *arXiv preprint arXiv:2201.02177*, 2022.

Omri Puny, Matan Atzmon, Edward J. Smith, Ishan Misra, Aditya Grover, Heli Ben-Hamu, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=zIUyj55nXR.

Eric Qu and Aditi Krishnapriyan. The importance of being scalable: Improving the speed and accuracy of neural network interatomic potentials across chemical domains. *Advances in Neural Information Processing Systems*, 37:139030–139053, 2024.

Fazle Rahat, M Shifat Hossain, Md Rubel Ahmed, Sumit Kumar Jha, and Rickard Ewetz. Data augmentation for image classification using generative ai, 2024. URL https://arxiv.org/abs/2409.00547.

Benjamin Rhodes, Sander Vandenhaute, Vaidotas Šimkus, James Gin, Jonathan Godwin, Tim Duignan, and Mark Neumann. Orb-v3: atomistic simulation at scale, 2025. URL https://arxiv.org/abs/2504.06231.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter W. Battaglia. Learning to simulate complex physics with graph networks. In *Proceedings of the 37th International Conference on Machine Learning*. JMLR.org, 2020.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *Proceedings of the 38rd International Conference on Machine Learning*, 2021.

Kristof T. Schütt, Huziel E. Sauceda, P.-J. Kindermans, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *NeurIPS*, 2017.

Muhammed Shuaibi, Abhishek Das, Anuroop Sriram, Misko, Luis Barroso-Luque, Ray Gao, Siddharth Goyal, Zachary Ulissi, Brandon Wood, Tian Xie, Junwoong Yoon, Brook Wander, Adeesh Kolluru, Richard Barnes, Ethan Sunshine, Kevin Tran, Xiang, Daniel Levine, Nima Shoghi, Ilias Chair, , Janice Lan, Kaylee Tian, Joseph Musielewicz, clz55, Weihua Hu, , Kyle Michel, willis, and vbttchr. FAIRChem, 2025. URL https://github.com/facebookresearch/fairchem.

Philipp Thölke and Gianni De Fabritiis. Equivariant transformers for neural network based molecular potentials. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=zNHzqZ9wrRB.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds, 2018. URL https://arxiv.org/abs/1802.08219.

Yuyang Wang, Ahmed A. Elhag, Navdeep Jaitly, Joshua M. Susskind, and Miguel Ángel Bautista. Swallowing the bitter pill: Simplified scalable conformer generation. In *Forty-first International Conference on Machine Learning*, 2024.

Ziduo Yang, Xian Wang, Yifan Li, Qiujie Lv, Calvin Yu-Chian Chen, and Lei Shen. Efficient equivariant model for machine learning interatomic potentials. *npj Computational Materials*, 11(1):49, 2025. doi: 10.1038/s41524-025-01535-3. URL https://doi.org/10.1038/s41524-025-01535-3.

Xinyi Yu, Guanbin Li, Wei Lou, Siqi Liu, Xiang Wan, Yan Chen, and Haofeng Li. Diffusion-based data augmentation for nuclei image segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, 2023.

Eric C-Y Yuan, Yunsheng Liu, Junmin Chen, Peichen Zhong, Sanjeev Raja, Tobias Kreiman, Santiago Vargas, Wenbin Xu, Martin Head-Gordon, Chao Yang, et al. Foundation models for atomistic simulation of chemistry and materials. *arXiv preprint arXiv:2503.10538*, 2025.

Leo Zhang, Kianoosh Ashouritaklimi, Yee Whye Teh, and Rob Cornish. Symdiff: Equivariant diffusion via stochastic symmetrisation. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=i1NNCrRxdM.

Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and Weinan E. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Physical Review Letters*, 120(14), April 2018. ISSN 1079-7114. doi: 10.1103/physrevlett.120.143001. URL http://dx.doi.org/10.1103/PhysRevLett.120.143001.

A Implementation Details

A.1 Model Architecture

Table 3 provides the complete architectural specifications for TransIP's model versions, as well as eSEN hyperparameters for the inference test in Table 2. For eSEN, we follow the small version reported by Levine et al. (2025).

Table 3: TransIP model configurations. All versions share the same embedding method and activation functions.

Configuration	**Small (S)**	**Medium (M)**	**Large (L)**
Hidden dimension (d)	384	768	1024
Number of layers (L)	8	12	24
Number of heads	6	12	16
Total parameters	14M	85M	302M
Shared configurations:			
Coordinate embedding		MLP	
Activation function		GELU	
Context length		1024	
Projection dropout		0.01	
Attention dropout		0.0	
Transformation network $\mathcal{T}_\tau$*:*			
Number of layers		2	
Hidden dimension		$2 \times$ d	
Activation		GELU	

Table 4: eSEN hyperparameters for inference test in Table 2.

Configuration	Small	Medium
sphere_channels	128	128
lmax	2	4
mmax	2	2
edge_channels	128	128
distance_function	gaussian	gaussian
num_distance_basis	64	128
num_layers	4	10
hidden_channels	128	128
max_neighbors	30	30
cutoff_radius	6	6
normalization_type	rms_norm_sh	rms_norm_sh
activation_type	gate	gate
ff_type	spectral	spectral

A.2 Training Hyperparameters

Table 5 provides TransIP's optimal hyperparameters.

A.3 Data Processing and Augmentation

TransIP processes molecular data with the following pipeline:

Table 5: Training hyperparameters used for TransIP experiments.

Hyperparameter	**Value**
Optimization:	
Optimizer	AdamW
Learning rate	$\{5, 3\} \times 10^{-4}$
Weight decay	1×10^{-3}
Gradient clip norm	{200, 100}
Learning rate schedule:	
Scheduler type	Cosine
Warmup fraction	0.01
Min LR factor	0.01
Loss weights:	
Energy (λ_E)	5
Forces (λ_F)	15
Equivariance (λ_{leq})	5 (selected from {1, 5, 10, 100})

- **Coordinate centering**: Atomic coordinates are centered by subtracting the center of mass: $\mathbf{r}_i \leftarrow \mathbf{r}_i - \frac{1}{|m|}\sum_j \mathbf{r}_j$
- **Equivariance pairs**: For training with learned equivariance, we create pairs $(m, \phi(g)(m))$ where g is sampled uniformly from SO(3) per molecule.

A.4 Radial basis functions

For the 80-epoch runs in Section 6.3, we use Gaussian RBF (radial basis functions) features (following Behler & Parrinello (2007)), defined as:

$$\psi_k(r_{ij}) = \exp\left(-\gamma(r_{ij} - \mu_k)^2\right), \qquad k = 1, \ldots, K \tag{8}$$

where r_{ij} is the Euclidean distance between node i and node j, and K is the total number of RBF channels. The centers $\{\mu_k\}_{k=1}^K$ are chosen uniformly between $r_{\min}$ and $r_{\max}$, while γ determines the width of the basis functions. We include RBF features at two levels. First, at the node level, we aggregate them as local features for each atom:

$$\mathbf{d}_i = \sum_{j \neq i,\ r_{ij} < r_c} \psi(r_{ij}), \tag{9}$$

where r_c is a cutoff radius and $\psi(r_{ij}) = [\psi_1(r_{ij}), \ldots, \psi_K(r_{ij})]^\top \in \mathbb{R}^K$. The aggregated feature is then projected to the model dimension and added to the atom token representation. Second, at the attention level, we use them as additive biases for the attention heads:

$$\mathbf{b}_{ij} = W_{\text{rbf}}\psi(r_{ij}) \in \mathbb{R}^{N_h}, \tag{10}$$

where N_h is the number of attention heads. This bias is added to the attention logits before the softmax:

$$\ell_{ij}^{(h)} = \frac{\mathbf{q}_i^{(h)\top}\mathbf{k}_j^{(h)}}{\sqrt{d_h}} + b_{ij}^{(h)} + \mathcal{M}_{ij}, \tag{11}$$

$$\alpha_{ij}^{(h)} = \text{softmax}_j\left(\ell_{ij}^{(h)}\right), \tag{12}$$

where h denotes the attention head, and $\mathcal{M}_{ij}$ is the masking term.

A.5 Evaluation Metrics

We evaluate model performance using the following metrics:

Force Mean Absolute Error (MAE):

$$\text{Force MAE} = \frac{1}{3|m|}\sum_{i=1}^{N}\sum_{\alpha\in\{x,y,z\}}|\mathbf{F}_{i,\alpha} - \mathbf{F}^*_{i,\alpha}| \quad (\text{meV/Å}) \tag{13}$$

Force Cosine Similarity:

$$\text{Force CosSim} = \frac{1}{|m|}\sum_{i=1}^{|m|}\frac{\mathbf{F}_i \cdot \mathbf{F}^*_i}{\|\mathbf{F}_i\|\|\mathbf{F}^*_i\|} \tag{14}$$

Energy per Atom MAE:

$$\text{Energy/atom MAE} = \frac{1}{|m|}|E - E^*| \quad (\text{meV/atom}) \tag{15}$$

Total Energy MAE:

$$\text{Total Energy MAE} = |E - E^*| \quad (\text{meV}) \tag{16}$$

where $\mathbf{F}$ and E denote predicted forces and energies, $\mathbf{F}^*$ and E^* are ground truth values, and $|m|$ is the total number of atoms. For energies, we use referenced targets following Levine et al. (2025).

A.6 Computational Resources

- 5-epoch experiments: 8 NVIDIA 80GB GPUs
- 80-epoch experiments: 32 NVIDIA 80GB GPUs

A.7 Validation Splits

For 5-epoch runs, we evaluate on domain-specific validation subsets sampled from the OMol25 validation (Val-Comp) dataset:

- Metal complexes: 20,000 samples
- Electrolytes: 20,000 samples
- Biomolecules: 20,000 samples
- SPICE: 9,630 samples (complete subset)
- Neutral organics: 20,000 samples (including ANI2x, OrbNet-Denali, GEOM, Trans1x, RGD)
- Reactivity: 20,000 samples
- Full validation set: 20,000 samples.

We use the full (2M) Val Comp dataset to evaluate TransIP and TransAug in Table 1.

B Additional Results

B.1 RBF features

To study the effect of RBF features, we ran additional experiments using TransIP-S trained for 5 epochs with different numbers of RBF channels. We report the performance of energy and force predictions on OMol Val-Comp splits in Table 6. Here, 0 RBF channels indicates TransIP without RBF features. Other hyperparameters are the same as Appendix A. The results show that increasing the number of RBF channels improves the performance across all splits and metrics.

Table 6: TransIP-S with different number of RBF channels. Each model is trained for 5 epochs and evaluated on the Val-Comp splits.

	Biomolecules		Electrolytes		Metal Complexes		Neutral Organics		Total	
RBF channels	**Energy ↓**	**Forces ↓**	**Energy ↓**	**Forces ↓**	**Energy ↓**	**Forces ↓**	**Energy ↓**	**Forces ↓**	**Energy ↓**	**Forces ↓**
0	6.92	93.97	8.35	73.76	10.58	90.26	12.64	105.33	10.74	82.30
8	5.32	56.92	6.81	51.87	9.34	78.44	10.31	83.19	8.60	55.87
16	4.18	43.90	5.69	42.48	8.42	70.93	9.08	69.98	7.10	45.34
32	4.36	43.36	5.72	42.04	8.57	70.88	8.99	69.46	7.30	44.84
64	4.01	39.26	5.48	39.12	8.32	68.45	8.56	65.01	6.84	41.49

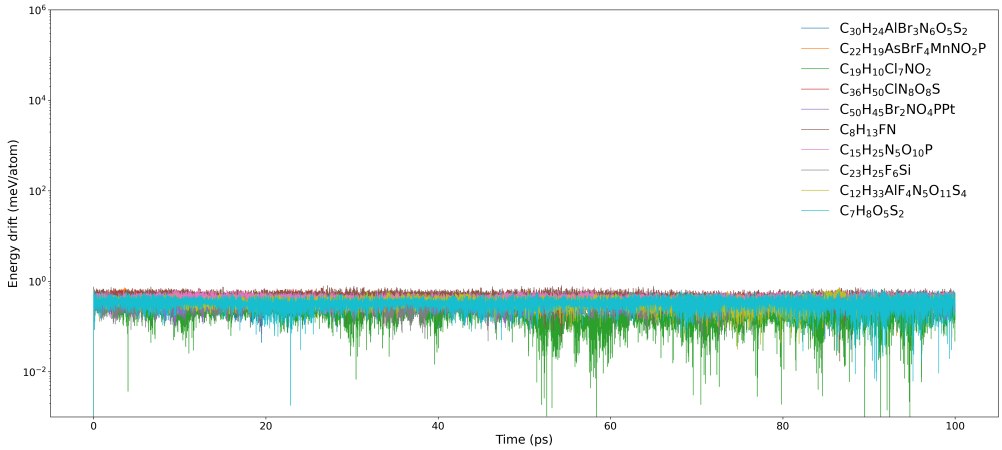

Figure 6: NVE stability test. Total energy drift for TransIP-M over 100 ps trajectories.

B.2 MD simulation

We evaluate the simulation stability of TransIP using NVE molecular dynamics on 10 molecules randomly selected from the OMol25 out-of-distribution validation subset (Val-Comp). For this experiment, we use the TransIP-M model and propagate the dynamics of each molecule for 100 ps with a 1 fs timestep using a velocity-Verlet integrator. We report the total energy drift along each trajectory as a measure of simulation stability; the results are shown in Figure 6. TransIP exhibits small energy drift across these trajectories, suggesting stable short-timescale dynamics. This behavior is qualitatively consistent with reports for equivariant MLIP baselines such as eSEN (Fu et al., 2025). Although this is a preliminary experiment, it provides initial evidence that TransIP may be suitable for broader molecular dynamics evaluations and motivates further studies on longer trajectories and more diverse simulation conditions.

B.3 OMol25 splits

In this section, we include additional dataset scaling results on OMol25 splits for TransIP and TransAug as well as the 80-epoch runs of all TransIP variants (Small, Medium, Large).

Table 7: Val-Comp energy and force MAE for SPICE and Reactivity splits.

		SPICE		Reactivity	
Model	**Epochs**	**Energy ↓**	**Forces ↓**	**Energy ↓**	**Forces ↓**
TransAug-S	5	11.5	151.3	23.0	179.7
TransIP-S	5	8.7	121.8	17.8	136.4

Table 8: Comprehensive Val-Comp. E/atom MAE (`meV/atom`), E MAE (`meV`), and F MAE (`meV/Å`) results.

	Biomolecules			Electrolytes			Metal Complexes			Neutral Organics			Total		
Model	**E/atom ↓**	**E ↓**	**F ↓**	**E/atom ↓**	**E ↓**	**F ↓**	**E/atom ↓**	**E ↓**	**F ↓**	**E/atom ↓**	**E ↓**	**F ↓**	**E/atom ↓**	**E ↓**	**F ↓**
TransIP-S	2.02	273.83	29.00	3.22	225.60	30.31	5.71	335.87	59.28	5.38	156.62	51.91	3.99	233.20	31.90
TransIP-M	0.84	115.32	16.92	1.88	129.03	19.50	3.95	227.75	45.86	3.77	103.03	34.24	2.22	128.08	20.42
TransIP-L	0.60	89.40	11.00	1.30	86.30	14.20	3.30	197.60	38.50	2.40	69.40	23.80	1.60	91.20	14.70

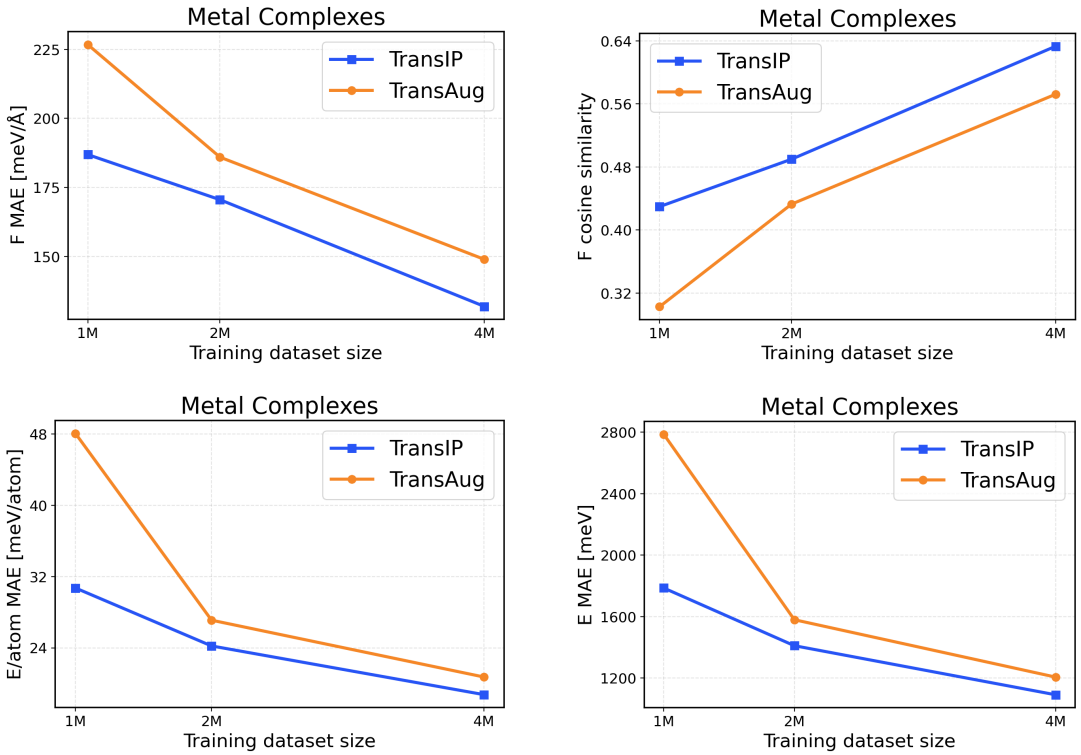

Figure 7: Metal Complexes scaling across training dataset sizes (1M / 2M / 4M). The top row presents force metrics, while the bottom row displays energy metrics.

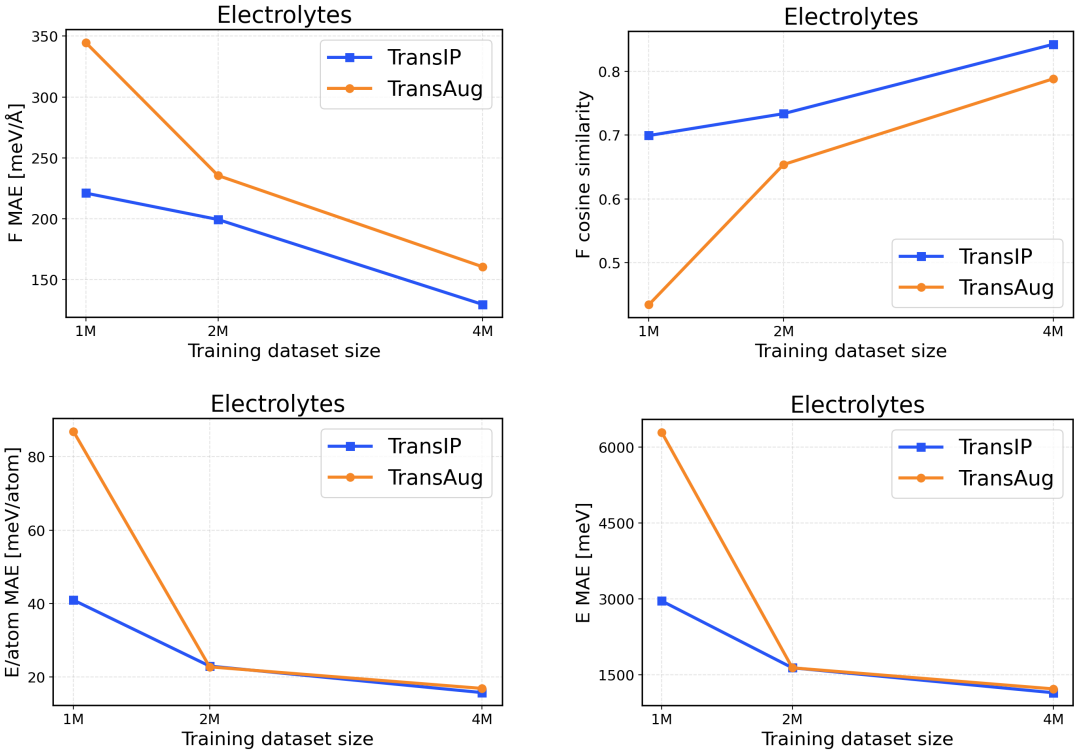

Figure 8: Electrolytes scaling across training dataset sizes (1M / 2M / 4M). The top row presents force metrics, while the bottom row displays energy metrics.

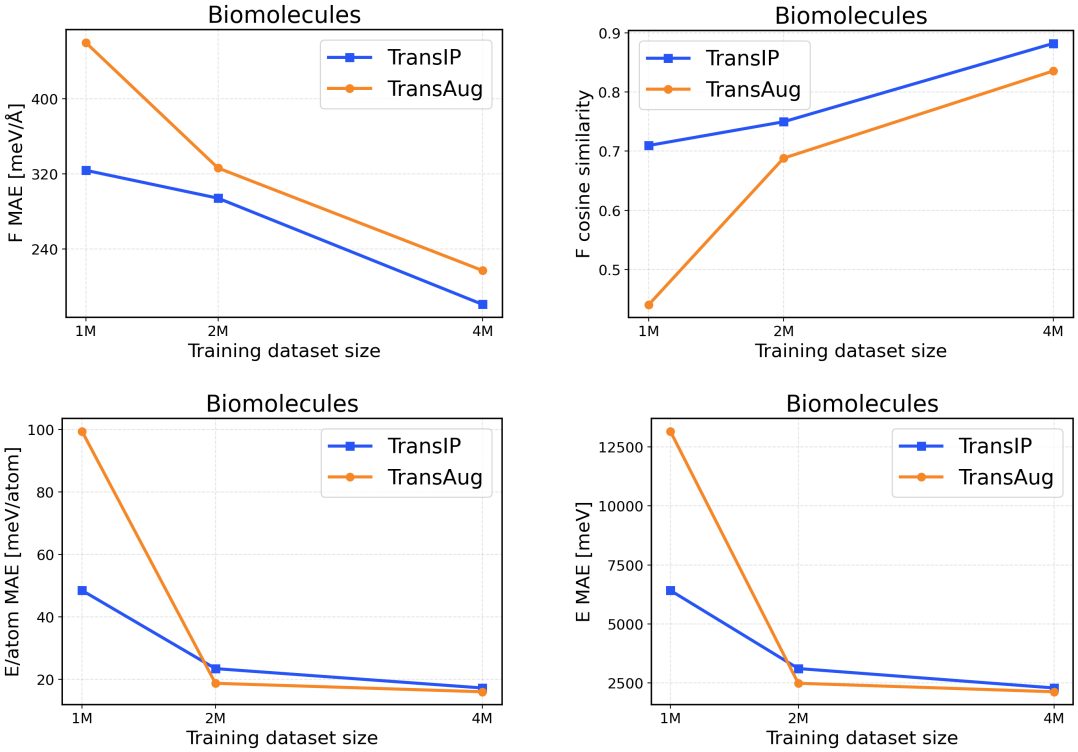

Figure 9: Biomolecules scaling across training dataset sizes (1M / 2M / 4M). The top row presents force metrics, while the bottom row displays energy metrics.

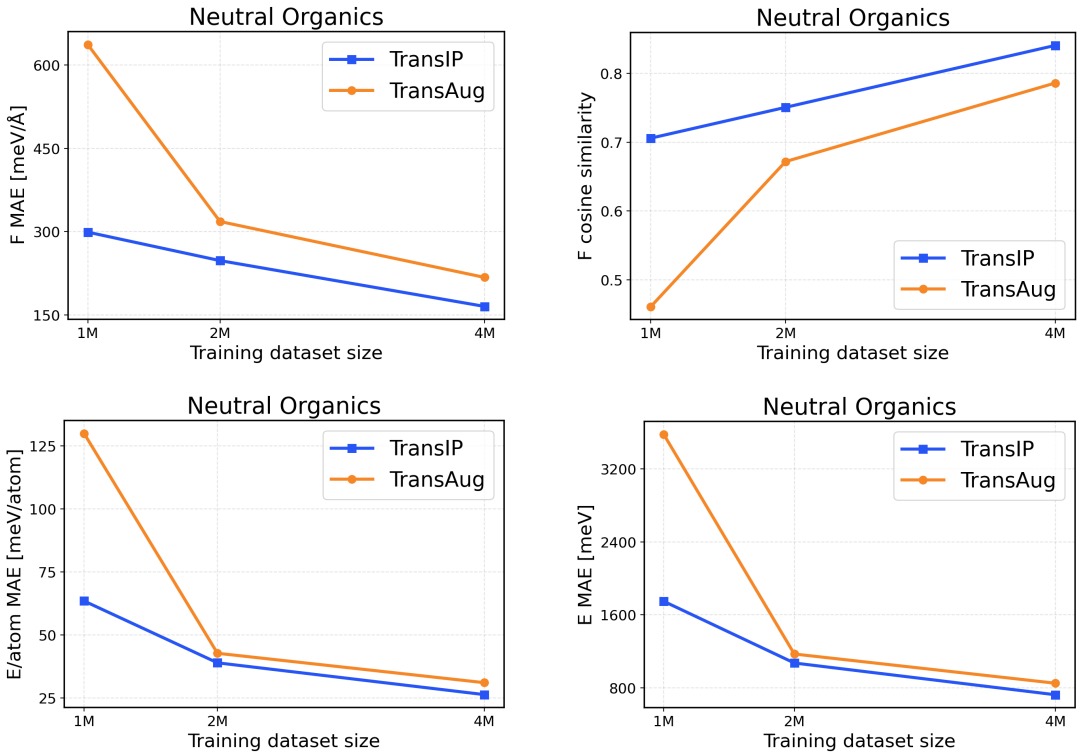

Figure 10: Neutral Organics scaling across training dataset sizes (1M / 2M / 4M). The top row presents force metrics, while the bottom row displays energy metrics.

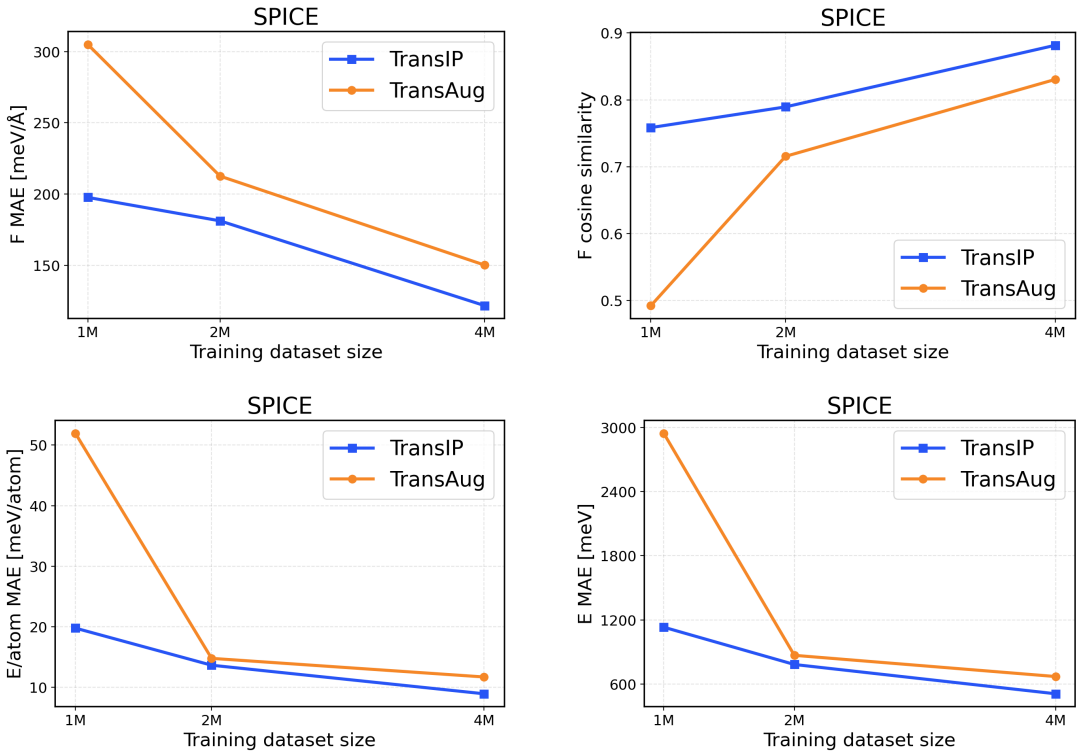

Figure 11: SPICE scaling across training dataset sizes (1M / 2M / 4M). The top row presents force metrics, while the bottom row displays energy metrics.

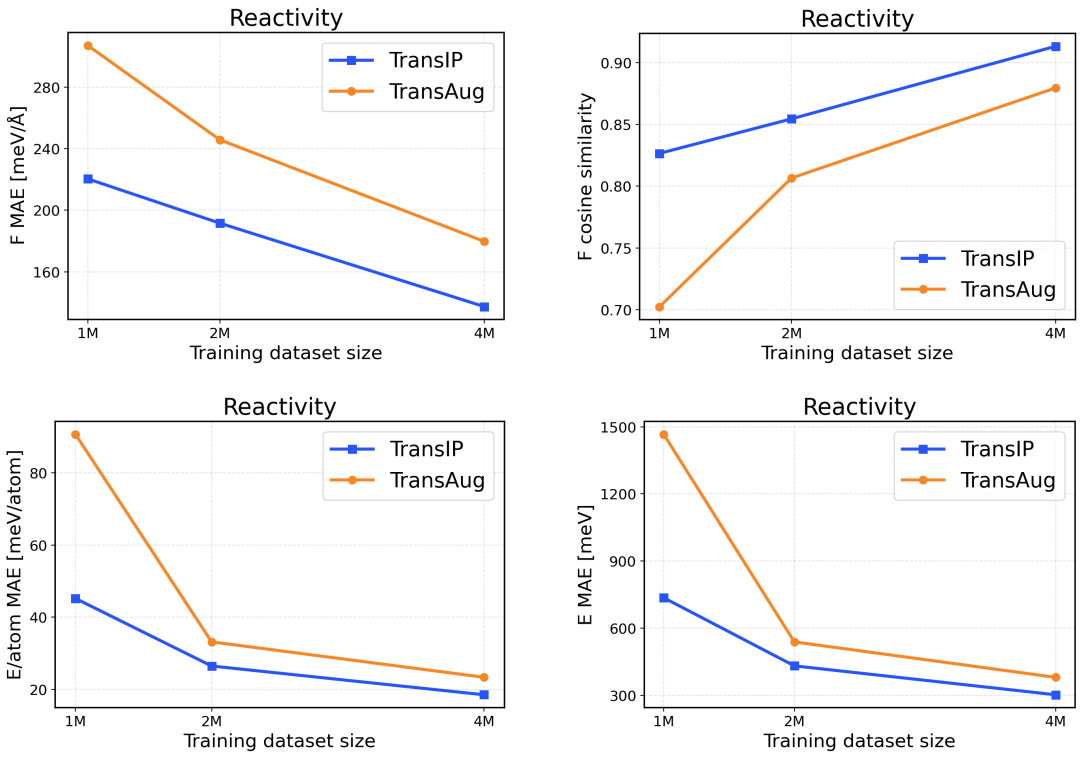

Figure 12: Reactivity scaling across training dataset sizes (1M / 2M / 4M). The top row presents force metrics, while the bottom row displays energy metrics.