Digital Discovery

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: M. P. Klein, I. Rudenko, E. A. Pidko and I. Bushmarinov, *Digital Discovery*, 2026, DOI: 10.1039/D6DD00096G.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/digitaldiscovery

View Article Online
DOI: 10.1039/D6DD00096G

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

ConforFormer: representation for molecules through understanding of conformers

Mas Pieter Klein,[a] Irina Rudenko,[b‡] Evgeny A. Pidko,[a‡] and Ivan Bushmarinov[c‡]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Molecular properties of chemical compounds are governed not by a single unique arrangement of atoms (2D molecular graph) but by ensembles of three-dimensional conformers, yet most molecular representations for machine learning approaches either ignore conformational diversity or use it implicitly to augment molecular graphs. Here we introduce ConforFormer, a geometry-first foundation model capable of learning conformation-robust molecular embeddings directly from the 3D atomic coordinates. By aligning representations across multiple conformers of the same molecules through a novel contrastive objective, ConforFormer produces compact, task-agnostic embeddings that can be generated once and directly applied to downstream tasks, including property prediction and structural similarity, without extensive fine-tuning. Across a range of quantum-chemical and bioactivity benchmarks, these frozen embeddings achieve competitive performance without task-specific fine-tuning, while offering improved stability on small datasets. Beyond property prediction, the learned embedding space allows to discriminate with high-precision molecular conformers and isomers, substantially outperforming classical fingerprint-based similarity measures. This implies that explicit exposure to conformational relationships induces representations that generalize beyond the conformer recognition task itself, capturing chemically meaningful structural constraints directly from 3D geometries. More broadly, our results suggest that incorporating conformation-awareness as a foundational learning task provides a fundamental route towards transferable, geometry-centered molecular representations particularly relevant for complex chemical systems, where conventional graph-based representations are ambiguous or ill-defined.

1 Introduction

The properties and function of chemical compounds are governed not by a single static arrangement of atoms, but by ensembles of three-dimensional conformations that interconvert on accessible timescales and which equilibrium can be manipulated by varying external stimuli. This conformational complexity greatly impacts properties ranging from catalytic reactivity to molecular recognition and biochemical functions, yet it presents a fundamental challenge for machine learning applications in chemistry. Molecules having similar chemical notations (e.g. brutto formulae, SMILES strings, 2D molecular graphs, etc.) may differ substantially in their 3D geometries. Recent years have seen growing interest in the development of large foundation models as a means to address such complexity across scientific domains. Pre-training large foundational models via self-supervised learning has proven highly effective in text and vision tasks, motivating analogous approaches in the natural sciences, including chemistry[1], physics[2], and applied meteorology[3].

In chemistry, such models aim to learn transferable internal representations of the chemical space during pre-training that can be reused across a wide range of prediction tasks, reducing reliance on task-specific supervision. Importantly, the usefulness of a chemical foundation model is determined by the representations it learns during pre-training. Existing pre-trained chemical models are typically used to initialize weights for supervised prediction tasks, which are solved by fine-tuning the whole model for the objective[4–8]. While this approach can achieve state-of-the-art performance on benchmarks, it often shows limited robustness on real-world chemical datasets, which in laboratory settings rarely exceed a few hundred experimentally measured points[9]. This suggests that the common pre-training objectives do not always yield representations that are sufficiently stable or transferable under realistic chemical data constraints.

Most chemical foundation models still operate on simplified 2D representations of molecules, ignoring the conformation and configurational diversity that governs their real chemical behavior. In reality, each compound exists as an ensemble of 3D structures (conformers) whose distribution determines such properties as

[a] Department of Chemical Engineering, TU Delft, Delft, Netherlands. E-mail: mp-klein@tudelft.nl; E.A.Pidko@tudelft.nl
[b] Avride Inc., Austin, TX, USA. E-mail: irina-rud@avride.ai
[c] Perplexity AI, Belgrade, Serbia. E-mail: ivan.bushmarinov@perplexity.ai
‡ Corresponding authors. Equal contribution.

View Article Online
DOI: 10.1039/D6DD00096G

binding affinities, docking poses, and chemical reactivity[10–12]. Typically, conformers differ from each other by rotations around single bonds, inversion of nitrogen lone pairs and other movements allowed by molecular flexibility. Conformers are distinct from isomers, which also are 3D geometries with the same composition, but one isomer cannot be produced from another without rearranging chemical bonds, i.e. a chemical reaction happening. Capturing the distribution of 3D geometries possible for a molecule is essential for the property prediction task, yet explicit incorporation of understanding conformations as a foundational learning objective remains largely unexplored in current chemical foundation models.

From a chemical perspective, conformers of the same molecule represent distinct geometrical realizations of an equivalent chemical entity, making their alignment a natural target for contrastive learning that would enforce equivalence across conformational space. Contrastive learning has emerged as a powerful strategy to enhance foundation models and refine embeddings without explicit labels by regularizing the embedding space in a way that it becomes organized so that distance correlates with semantic similarity. By structuring the embedding space to bring similar objects closer while pushing dissimilar ones apart, models learn more informative, general-purpose representations. Methods developed at Amazon[13,14] illustrate how contrastive approaches can refine embeddings across modalities, improving downstream task performance. A notable example is Microsoft E5[15], trained in a weakly supervised manner on naturally occurring document pairs such as questions and answers from forums.

To our knowledge, no chemical embedding model incorporated conformational equivalence into its foundational learning objectives. Here we introduce ConforFormer, a foundational model that explicitly accounts for this diversity by aligning embeddings across multiple conformations of a molecule to produce compact, informative representations suitable for downstream tasks. In this work, we present 1) a new compact embedding for chemical structures, learnable from 3D geometries, 2) a novel contrastive learning process necessary to build it, 3) a benchmark evaluating the model's ability to distinguish pharmaceutically relevant molecules, and 4) the performance of the resulting embeddings on established chemical benchmarks.

2 Technical and Chemical Preliminaries

In this section, we first summarize the popular technical approaches to molecular representation learning and their use in the current foundation model, before outlining the chemical perspective that motivates the conformer-based learning strategy introduced herein.

2.1 Backbone Models for Molecular Representations

Backbone architectures for molecular embeddings can be broadly grouped into three categories based on how molecular structure is represented. Graph-based models such as message-passing neural networks (MPNNs)[16] and GROVER[17] represent molecules as atom-bond graphs and capture local connectivity through message passing. Sequence-based transformers adapt methods originally developed for natural language processing to string-based molecular representations such as SMILES[18]. Examples include MolBERT[19], ChemBERTa[7] and ChemBERTa-2[5]. By treating individual atoms in the structure-encoding string as tokens, these models benefit from scalable pre-training on very large datasets. However, similar to graph-based approaches, the richness of the resulting representations is constrained by the two-dimensional encoding of the molecular structures. To better account for spatial effects, a growing class of models has emerged that explicitly incorporates atomic coordinates into the representation. Methods such as GEM[8] and ABT-MPNN[20] augment 2D graph-based representations with 3D geometrical information, resulting in improved performance on tasks that depend on the molecular shape. Among these, Uni-Mol family of models[4,21,22] has established itself as a leading framework. Built on a transformer backbone with explicit encoding of atomic positions in 3D space, Uni-Mol achieves state-of-the-art performance across multiple benchmarks, including molecular property prediction, conformer generation, and docking.

2.2 Uni-Mol Architecture

Uni-Mol is a representative example of a geometry-first molecular foundation model, designed to learn transferable molecular representations directly from 3D atomic structures. The architecture is based on an E(3)-equivariant transformer (the distinction between SE(3) and E(3) is discussed in Dumitrescu *et al.*[23]). Each atom is represented as a token embedding that incorporates its element type as a categorical feature, while spatial information is encoded through pairwise interatomic distances. 3D geometry is introduced via a distance matrix representation that is integrated into the attention mechanism of the multi-layer, multi-head transformer encoder as an initial attention mask. The Uni-Mol model is pre-trained using a combination of self-supervised objectives, including masked atom prediction, masked interatomic distance prediction, and coordinate denoising. Pre-training is carried out on a large dataset introduced in the original Uni-Mol paper[4] containing 19M unique SMILES strings and around 209M associated 3D molecular geometries with up to 10 conformers generated per molecule. This large-scale pre-training enables the model to learn generalizable representations that capture both chemical connectivity and 3D geometric relationships. Subsequent developments within the Uni-Mol family of methods focused on extending model capacity and improving performance on selected benchmarks. Uni-Mol+[21] augments the original framework with additional atom and molecular graph features, as well as low-cost geometry data obtained from energy minimization trajectories. Uni-Mol2[22] further explores model scaling, extending the architecture up to a 1B parameter model.

Following ChemBERTa[19], Uni-Mol introduced a special *CLS* token to aggregate global molecular information. This token is assigned an "empty" atom type and placed at the geometric center of the molecule. It is processed alongside atomic tokens by the transformer but is excluded from the atom masking or distance prediction tasks. During downstream use, the embedding associated with the *CLS* token serves as a fixed-length representation

View Article Online
DOI: 10.1039/D6DD00096G

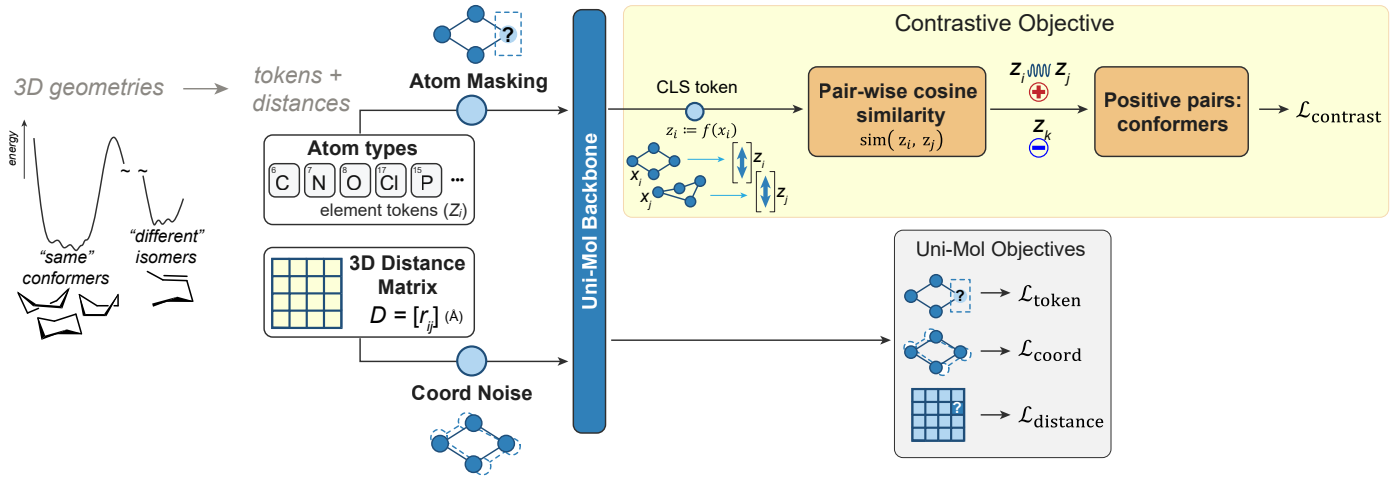

Fig. 1 Schematic illustration of the ConforFormer framework: Model architecture with pretraining objectives.

of the entire molecule and is commonly passed to task-specific prediction heads.

We should note that Uni-Mol does not treat a molecule as a single fixed geometry at downstream inference time. Within the standard Uni-Mol protocol, molecular property predictions are obtained by averaging the predictions over up to 10 conformers generated per molecule. The resulting conformer ensemble reduces sensitivity to any particular RDKit-generated geometry and partially accounts for conformational variability within the limited sampled set.

Herein, we selected the original Uni-Mol as the backbone architecture, because it relies exclusively on the atom types and their 3D geometrical arrangements as input. This geometry-centric design makes Uni-Mol a suitable foundation for exploring learning objectives that explicitly account for the conformational diversity and fluxionality at the representation level.

2.3 Technical Practices in Transfer Learning with Foundational Transformer Models

In practice, large transformer models pretrained with self-supervised objectives (e.g., masked token prediction, contrastive learning, or distance-based representations) are rarely fully retrained for downstream benchmarks. Instead, transfer learning is typically performed by freezing most of the pre-trained layers and updating only a small subset of parameters, such as task-specific prediction heads or a limited number of upper transformer layers. This strategy significantly reduces computational cost while retaining the general-purpose representations learned during pretraining. Such practices are well-established in natural language processing, where models such as BERT[24] are commonly adapted using lightweight fine-tuning schemes, e.g. the use of adapter modules or by partial unfreezing[25]. Such approaches achieve strong performance on benchmarks such as GLUE[26] and SuperGLUE[27], while requiring only modest task-specific optimization.

A practical next step is to use the pre-trained model to produce a fixed representation of a data item ("embedding") that can be reused across downstream tasks and for similarity search. General-purpose embeddings reduce computational cost to solve regression and classification tasks, for example, when combined with graph-based match algorithms such as HNSW[28]. Chemistry applications span both data-scarce regimes (e.g. reactivity and selectivity datasets with hundreds to thousands labeled examples) to data-rich regimes (e.g. large screening candidate libraries or long molecular dynamics trajectories containing billions of structures). In both settings, repeatedly fine-tuning large models end-to-end is often impractical. Herein, we therefore focus on building compact, task-agnostic molecular embeddings that remain useful across multiple downstream tasks.

2.4 Chemical perspective for molecular representations

The representation of molecules as atoms connected by well-defined chemical bonds is a foundational abstraction in chemistry and it underlies most molecular representations used in computational modeling and machine learning. Structural formulae and molecular graphs provide an efficient and chemically intuitive labeling scheme that allows for systematic reasoning about reactivity, selectivity and molecular function. This abstraction is deeply embedded in the way chemists have been trained for generations and how chemistry is practiced, and, naturally, it has been readily adopted in data-driven chemistry modeling approaches. However, from a physical perspective, chemical bonds are not, strictly speaking, directly observable entities but rather conceptual constructs used to rationalize the behavior of interacting atoms and more intuitively approach the electronic structure.

Real molecules sample distributions of geometries within extended regions of a multidimensional potential energy surface (Figure 1, left panel). For stable organic molecules in the ground state, the molecular graph (in other words, atom connectivity) is generally preserved across the accessible conformational ensemble. Distinct conformers correspond to different local minima with the same topology separated only by low energy barriers due to e.g. bond rotations or pyramidal inversions. Chemically distinct species correspond to regions separated by sufficiently high energy barriers and their interconversion involves a chemical reaction. Structural formulae label these regions, and for most ground-state organic compounds, they capture connectiv-

View Article Online
DOI: 10.1039/D6DD00096G

ity rather efficiently. However, they do not uniquely specify the molecule geometry and do not readily encode the intrinsic variability in the 3D configurational fluxionality that often determines the chemical and physical properties of interest.

The notion of the chemical bond as a physical or conceptual entity has been the subject of long-standing discussion in the chemistry community[29–34]. Although more than a century-old Lewis model[35] provides an exceptionally successful language for chemical reasoning, the bond assignment is ultimately a model-dependent interpretation of the underlying electronic structure. Bader's Quantum Theory of Atoms in Molecules[36,37] provides an influential electron density-based framework for analyzing molecular structure and interatomic interactions. At the same time, discussions in the theoretical chemistry community emphasized the role of Lewis structures and connectivity-based reasoning as conceptual scaffolds with remarkable practical resilience[31–33].

In the present proof of concept, we deliberately focus on organic molecules, where graph- or string-based representations provide a sufficiently accurate and robust description. This makes the comparison conservative, because ConforFormer is evaluated in a regime where conventional graph-based methods are expected to work well. However, their limitations become apparent for more chemically complex systems, particularly organometallic and supramolecular compounds, where bonding patterns may be ambiguous, fluxional behavior is common, and chemically relevant distinctions often arise from "subtle" geometric rearrangements.[33] The broader relevance of such systems should therefore be understood as motivation for future extensions rather than as a demonstrated application of the current workflow. Organometallic compounds with agostic, η^n-coordinated, or fluxional bonding illustrate this point. A 3D structure or structural ensemble provides a more direct description of such compounds, while the conventions for assigning and drawing individual bonds may be ambiguous[38]. These challenges are especially pronounced for tasks such as predicting catalytic activity and selectivity, where 3D structure and conformational accessibility play a decisive role[39,40]. We note that extending the present approach to such systems will require structural ensembles and chemically meaningful positive/negative labels from e.g. quantum-chemical sampling or molecular dynamics.

One may alternatively address these challenges by introducing richer molecular-graph encodings that better capture coordination, stereochemistry, and fluxional bonding. This is an important avenue of work. Here we explore a complementary approach, in which molecular graphs are not used as model input. This does not imply that molecular identity is defined without graph information during dataset construction. Instead, the model learns representations directly from atom types and 3D geometries. Graph information is only used during data preparation to associate conformers of the same molecule and construct contrastive training pairs. We show that the resulting representation retains discriminative properties typically associated with graph-based fingerprints, while being explicitly defined on structural ensembles.

3 Methods: Conformer-Based Contrastive Learning for Molecular Representations

Guided by the chemical and technical considerations outlined above, we develop ConforFormer, a machine learning framework, in which conformational equivalence is explicitly enforced at the representation level. Figure 1 schematically illustrates the ConforFormer framework and its relation to the Uni-Mol backbone. As input, only the 3D coordinates of the conformers and the corresponding atomic numbers are considered for inputs. The central idea is to treat different 3D conformers of the same molecule as distinct geometric realizations of a single chemical entity, and to align their representations during pre-training. In contrast to task-specific fine-tuning strategies, this approach aims to produce compact, task-agnostic molecular embeddings that are robust to conformational variability and can be directly applied for various downstream applications.

3.1 Problem Formulation

We formulate molecular representation learning at the level of molecular identity, treating different 3D conformers of the same molecule as equivalent with respect to representation learning. Each conformer is processed independently by the Uni-Mol backbone encoder, producing an embedding vector in a shared latent space.

Formally, let $\mathscr{M}$ denote the set of molecules in the training corpus and let $\mathscr{C}(\updownarrow)$ denote the set of conformers associated with molecule $m \in \mathscr{M}$. During training, pairs of conformers sampled from the same molecule are treated as positive pairs, while conformers originating from different molecules form negative pairs. The learning objective enforces alignment of embeddings within each equivalence class $\mathscr{C}(\updownarrow)$, while maintaining separation between different molecular identities. Such a formulation does not assume functional equivalence of individual conformers. Instead, it defines a representation space, in which molecular identity is stable with respect to geometric variability. The contrastive learning objective implementing this formulation is introduced in the next section.

3.2 Contrastive Learning Objective

Following contrastive learning approaches established in text and image processing domains[14,15], we introduce a contrastive learning objective to improve the stability and transferability of frozen molecular embeddings. The contrastive learning is implemented as an additional task during pre-training and operates on 3D molecular conformers via a novel conformer-alignment target.

The model is trained to distinguish pairs of conformers among various molecules. Importantly, molecular graph information is not provided as the input to the model and it is only used at the data generation stage to label positive and negative pairs. This design enforces alignment based exclusively on 3D geometries and atomic identity.

We employ the normalized temperature-scaled cross-entropy (NT-Xent) loss function[41] to teach the model to put the embeddings of different 3D representations close in the embedding

View Article Online
DOI: 10.1039/D6DD00096G

space. This objective explicitly regularizes the embedding space such that representations of distinct 3D realizations of the same molecular identity are brought closer together, while embeddings of different molecules remain separated.

Let $\mathscr{X}$ denote the set of molecules and $f : \mathscr{X} \to \mathbb{R}^d$ an embedding function with embedding dimension $d = 512$. For vectors $\mathbf{v}, \mathbf{u} \in \mathbb{R}^d$, we define a cosine-style similarity

$$\mathrm{sim}(\mathbf{u}, \mathbf{v}) \coloneqq \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|} \in [0, 1].$$

The $[0, 1]$ range is imposed by the embedding normalization. For molecules $x, x' \in \mathscr{X}$, we write $\mathrm{sim}(f(x), f(x'))$ and denote $\mathbf{z}_i \coloneqq f(x_i)$.

In each training batch, we sample $n = 128$ unique molecules and generate two distinct 3D representations (views) for each, yielding $2n$ embeddings $\{\mathbf{z}_i\}_{i \in \mathscr{B}}$ with index set $\mathscr{B} = \{1, \ldots, 2n\}$. Let $\mathscr{P} \subset \mathscr{B} \times \mathscr{B}$ denote the set of ordered positive pairs, where $(i, j) \in \mathscr{P}$ if and only if $i \neq j$ and both indices correspond to two conformers of the same molecule. For each positive pair (i, j) NT-Xent loss is defined as:

$$\ell_{i,j} \coloneqq -\log \frac{\exp\big(\mathrm{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau\big)}{\sum_{k \in \mathscr{B} \setminus \{i\}} \exp\big(\mathrm{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau\big)},$$

where $\tau > 0$ is a temperature parameter. The total contrastive loss is obtained by summing over all positive pairs,

$$\mathscr{L}_{\text{contrast}} \coloneqq \sum_{(i,j) \in \mathscr{P}} \ell_{i,j}.$$

Higher values of τ reduce sensitivity to small embedding differences by flattening the softmax distribution. In all experiments reported here, we set $\tau = 0.07$; additional temperature ablations are provided in the Supporting Information, further SI (Section E).

The contrastive loss is combined with the original Uni-Mol pre-training objectives to yield the total loss:

$$\mathscr{L}_{total} = \mathscr{L}_{token} + 5 \cdot \mathscr{L}_{coord} + 10 \cdot \mathscr{L}_{distance} + 2 \cdot \mathscr{L}_{contrast}$$

Here $\mathscr{L}_{token}$ corresponds to the loss for masked token prediction, $\mathscr{L}_{coord}$ is the loss associated with the coordinates denoising task, and $\mathscr{L}_{distance}$ is the loss associated with the masked distance prediction. These objectives and their original batching protocol were introduced in[4] and further detailed in[22]. Models trained with the additional contrastive objective are referred to as ConforFormer throughout the work.

3.3 Training Protocol and Model Variants

All models in this work are based on the Uni-Mol backbone architecture described in Section 2.2. The downstream fine-tuning and evaluation workflow is summarized in Figure 2. We consider multiple training configurations, including the original Uni-Mol pre-trained model and variants trained with the conformer-alignment contrastive objective introduced Section 3.2. Training is performed on molecular conformer datasets from (i) the Uni-Mol corpus[4] and (ii) the OpenMolecules (OMol) dataset.[42] The Uni-Mol dataset was used as-provided,[4] with 10 conformations generated per molecule using rdkit,[43] and optimized using MFF94.[44] The OMol dataset provides higher-quality molecular geometries and includes a subset designed specifically for conformer analysis (see section B.2 of the SI for the data preparation details). Details of conformer generation, filtering, and pre-processing (including handling of degenerate geometries) are provided in the SI. The conformer-generation workflow used here is designed for organic molecules and should not be interpreted as a general protocol for organometallic or supramolecular systems, where conformer generation and molecular-identity assignment require additional chemical conventions.

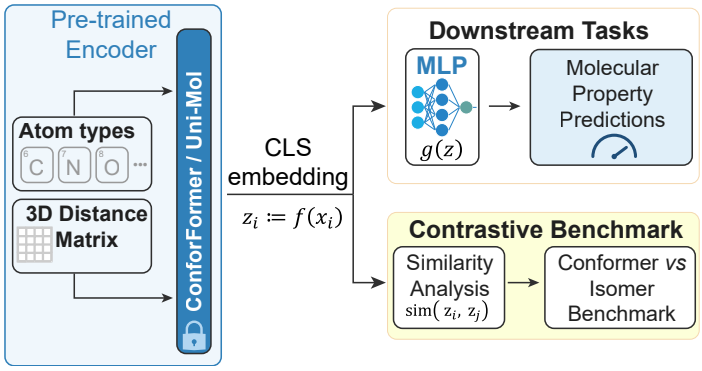

Fig. 2 Schematic illustration of the ConforFormer framework: Finetuning and evaluation scheme.

Models trained with the additional contrastive objective on these datasets are referred to as ConforFormer-UniMol and ConforFormer-OMol, respectively. All model variants are trained using identical backbone architectures, embedding dimensions and optimization settings to ensure that observed differences arise solely from the pre-training objectives and data sources rather than architectural or procedural changes.

For downstream evaluation, we focus on the quality and stability of learned molecular embeddings. Unless explicitly stated otherwise, the Uni-Mol encoder parameters are frozen and representations are extracted from the CLS token. These frozen embeddings are then used as input to lightweight task-specific models or similarity analyses without additional fine-tuning of the backbone. This evaluation protocol follows standard transfer-learning practices for large pre-trained transformer models, as discussed in Section 2.3. Under this framework, task-specific evaluation is conducted using MoleculeNet. Metrics of ROC-AUC are used for classification benchmarks, and root-mean-squared deviation (RMSD) for regression benchmarks. The model's predictions are compared against classes or true values, which are identical across all conformers. Reducing the number of conformers used when training task-specific models was found to have little effect on final performance across all benchmarks (Table S5 and Table S5). This, along with other ablation studies, can be found in section E of the ESI.

View Article Online
DOI: 10.1039/D6DD00096G

4 Results and Discussion

4.1 Representation Quality under Frozen Evaluation

We evaluated the quality and transferability of the molecular representations learned by ConforFormer across quantum-chemical, physico-chemical, and biological benchmarks. We focus on the representation-level differences and the impact of the conformer-aligned pre-training, with a detailed analysis of how the pre-training objective and the quality of the training data affect the downstream performance. A three-layer $512 \times 256 \times 128$ multi-layer perceptron (MLP) on top of frozen embeddings (see SI, Section A) was trained during the fine-tuning process to obtain predictions. This setting limits the ability of downstream optimization to compensate for the intrinsic shortcomings of the representation. Figures 3 and 4 summarize the results on quantum-chemical regression and biological classification benchmarks, respectively, shown alongside the recent literature ranges (the more extensive benchmarking results are tabulated in Tables S1 and S2 of the Supporting Information).

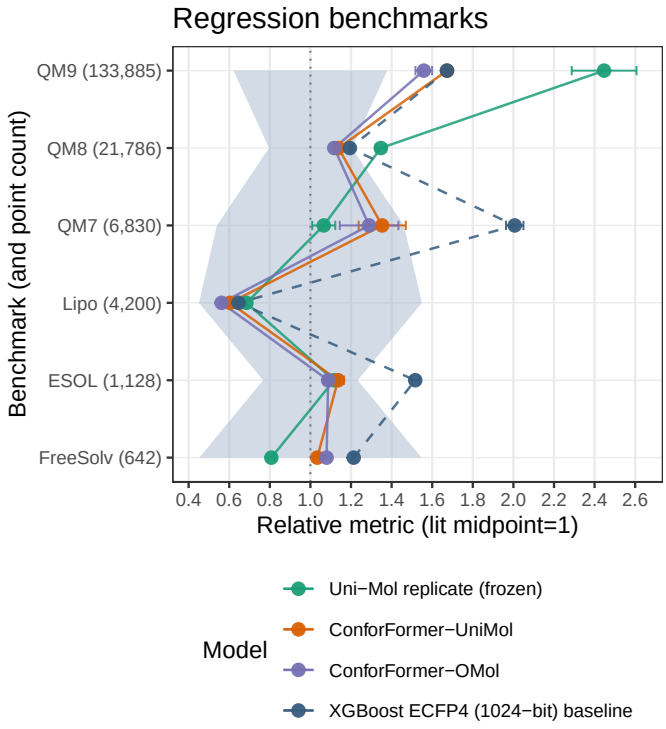

Fig. 3 Regression MoleculeNet benchmark results (rescaled error) for the XGBoost ECFP4 (1024-bit) baseline and embedding-based models; less is better (↓). The shaded area indicates the literature range for post-2019 models. The task-specific errors are rescaled linearly so that the midpoint of the literature range equals 1. The unscaled data is available in the Table S1 of the SI.

As a reference, we also report the results for XGBoost[45] trained on RDKit ECFP4 (1024-bit) Morgan fingerprints[46], which we identified in our screening as the strongest 2D baseline used in the main text (denoted as the XGBoost ECFP4 (1024-bit) baseline). We additionally screened RDKit Morgan ECFP4 and ECFP6 fingerprints folded into 2048 and 16384 bits, as well as Open Babel FP2, FP3, and FP4 fingerprints. This baseline predicts directly from the molecular graph via engineered local substructure features, whereas the Uni-Mol and ConforFormer embeddings are learned from 3D structures and then decoded by a lightweight predictor. Even without any end-to-end training aimed at obtaining a useful representation, the Uni-Mol backbone produces embeddings that are competitive with the fingerprint baseline on multiple tasks. Nevertheless, ConforFormer contrastive loss further improves performance on the regression tasks (ConforFormer-UniMol, Figure 3). Specifically, models trained with the conformer-alignment contrastive objective consistently yield higher-quality frozen embeddings than the Uni-Mol baselines trained without this objective. This effect is most pronounced for geometry-sensitive datasets, including QM8 and QM9, where contrastively trained models achieve markedly lower error than Uni-Mol.

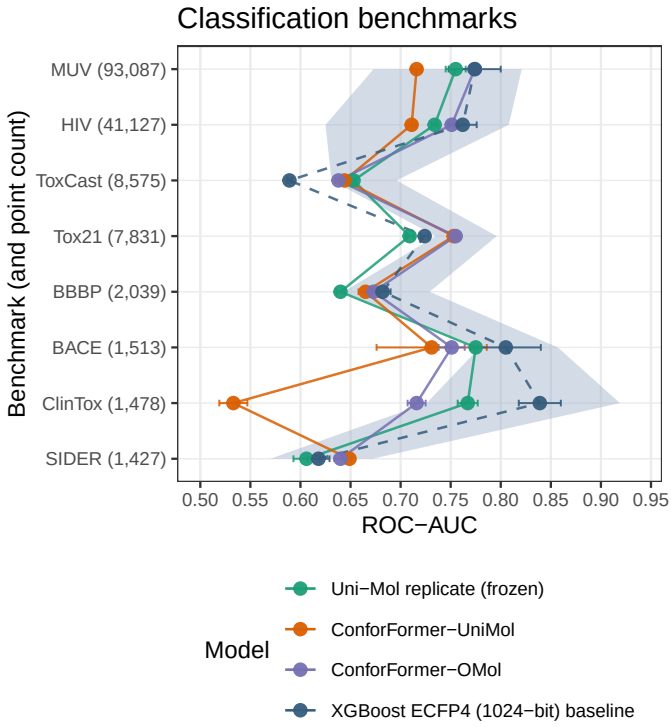

Fig. 4 Classification MoleculeNet benchmark results (ROC-AUC) for the XGBoost ECFP4 (1024-bit) baseline and embedding-based models; more is better (↑). The shaded area indicates the literature range for post-2019 models.

The Uni-Mol dataset is substantially skewed towards organic compounds and utilizes low-quality RDKit MMFF geometries. This limitation can be addressed by training on the recently released OMol dataset[42] which is computed at ωB97M-V/def2-TZVPD level and has conformation data for 8.2M unique molecules (see SI, section B for details). We therefore analyzed whether the improved quality of geometries alone or explicitly learning conformational relationships during pre-training could give rise to systematic improvements in frozen transfer. When trained on the OMol subset without the conformer-alignment contrastive objective (UniMol-OMol), performance remains broadly comparable to the Uni-Mol replicate under frozen regime (see SI).

View Article Online
DOI: 10.1039/D6DD00096G

In contrast, adding the conformer-alignment objective yields consistently better embeddings. ConforFormer-OMol performs best on 4 out of 6 quantum-chemical regression benchmarks (Figure 3) and 5 out of 8 classification benchmarks (Figure 4), while remaining below the frozen Uni-Mol replicate baseline on BACE and ClinTox. These two datasets have a low number of datapoints and the best model for those in our setup was in fact the XGBoost ECFP4 (1024-bit) baseline. Importantly, ConforFormer-OMol shows a significant improvement over Uni-Mol frozen embeddings and ConforFormer-UniMol on challenging MUV and HIV benchmarks, as well as much more stable performance than ConforFormer-UniMol, with no benchmark demonstrating numbers significantly outside the literature range except for QM9. This indicates that a diverse pre-train data with better quality of the molecular geometries, available via the OMol dataset, is beneficial for the quality of the embeddings downstream.

To ensure a controlled downstream comparison, we evaluated ConforFormer-OMol using the geometries from the MoleculeNet benchmark released by the Uni-Mol team (Table S4). This avoids conflating representation differences with differences in geometry generation pipelines at evaluation time. While higher-quality conformer generation could plausibly improve absolute metrics for all 3D-based approaches, a systematic study of geometry generation protocols is beyond the scope of the present work. We therefore emphasize representation-level differences under a consistent evaluation setup rather than maximizing absolute task performance. In this context, it is also expected that fully fine-tuned large models can achieve higher absolute performance on some tasks because end-to-end optimization can adapt the representation directly to downstream labels and better exploit task-specific supervision.

We also note that the Uni-Mol baseline already accounts for conformational variability at inference through prediction averaging over a conformer ensemble (up to 10 conformers per molecule). Thus, this does not contribute to the improvements observed for ConforFormer. Instead, the consistent gains under frozen evaluation are aligned with a representation-level mechanism. Conformer alignment during pre-training strengthens the representation itself and regularizes the embedding space, increasing robustness to geometric variability while remaining sensitive to conformational diversity. We believe that the ConforFormer-OMol representation captures some fluxional behavior beyond the 10 explicitly supplied conformers.

To summarize, the presented frozen evaluation demonstrates that conformer-aligned pre-training yields denser, more transferable 3D-derived molecular embeddings. Without task-specific fine-tuning, they provide competitive prediction of quantum-chemical properties and show strong performance on multiple pharmaceutically relevant classification benchmarks. This provides direct evidence that learning conformational relationships improves the chemistry captured by the representation, rather than simply improving downstream optimization.

4.2 Emergent Structure of the Embedding Space: Isomers vs Conformers

As a result of the contrastive loss applied, we would expect the model to gain the ability to distinguish molecules better. Even if a molecular graph is only used to generate a sample of 3D geometries and distinct labels for the contrastive loss, the model should be able to learn which transformations of the molecular geometry are "allowed" under the constraint of structure remaining the same. However, we did not construct the training objective in a way to specifically distinguish conformers from isomers. So, in this section we explore the emergent behavior of the obtained embeddings in generalizing beyond the supplied conformations to distinguish between isomers.

4.2.1 Isomer/conformer distinguishing benchmark

We introduce a new benchmark dataset PharmaIsomer to validate the models' capability to distinguish between conformers and isomers and explore the resulting embedding space.

To construct this benchmark, we used a portion of ZINC20[47] not overlapping with Uni-Mol or OMol datasets, selected subsets of isomeric molecules and pre-generated batches containing isomers and conformers for consistent evaluation. Specifically, each batch contained 128 unique molecules, which are all isomers to each other. Each isomer had exactly 2 conformers, resulting in 256 datapoints per batch. An 80/10/10 train/validation/test split was employed for the dataset so that the performance of models trained specifically on it could be evaluated; metrics below are all reported on the test split. Overall, PharmaIsomer contains 3,261,807,960 datapoints in 12,741,440 batches (see Section C of the Supporting Information for details). The dataset is freely available under CC-BY license[48].

The dataset contains four types of molecular pairs: backbone isomers where the molecules have a different bond order with the same composition (99.50% of all pairs); conformers (0.39%), optical isomers where molecules are mirror images of each other (0.05%); and diastereomers where the molecular topology is the same but the relative configuration of optical centers and/or double bonds is different (0.06%).

4.2.2 Isomer similarity

As the initial step, we plotted the distributions of cosine similarity densities for the embeddings obtained from Uni-Mol replicate and from ConforFormer–OMol. Interestingly, the CLS token directly from our replication of the Uni-Mol already showed some level of separation between conformers and isomers, with cosine similarity of embeddings for conformers being closer to 1 than for isomers (Figure 5). This suggested from the start that a correctly trained model could learn to distinguish between those.

After including a contrastive objective, ConforFormer–Unimol and ConforFormer–OMol learn to cleanly separate conformers and isomers without any additional training. So, besides the embeddings becoming more useful for property prediction, they can be competitive for the tasks of similarity search as well. For that, we needed a more formal evaluation of the model capability to distinguish conformers and isomers.

Let $\mathbb{D} := \{(x_i, x_i', y_i)\}_{i=1}^{N}$ be a dataset of molecule pairs, where

View Article Online
DOI: 10.1039/D6DD00096G

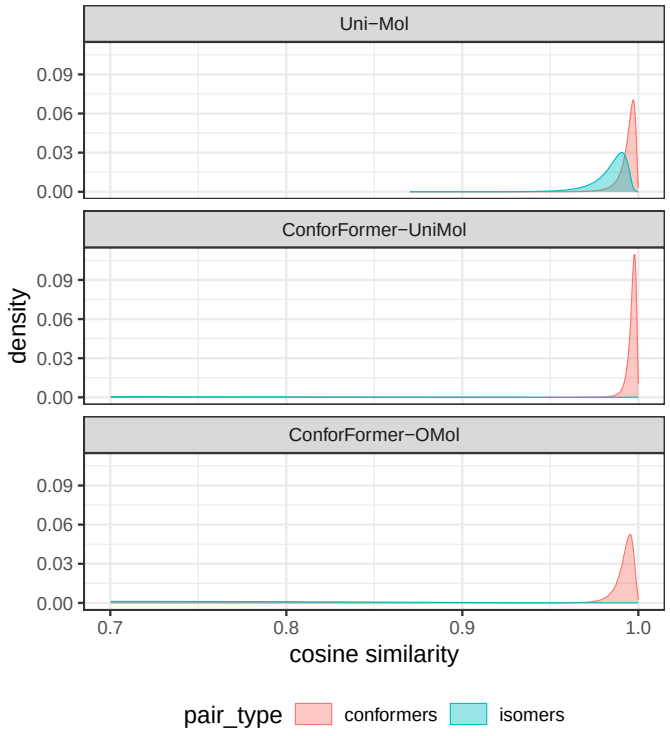

Fig. 5 Distribution of cosine similarities between CLS token values extracted from Uni-Mol, ConforFormer–UniMol and ConforFormer–OMol, as measured on the Pharmaisomer benchmark.

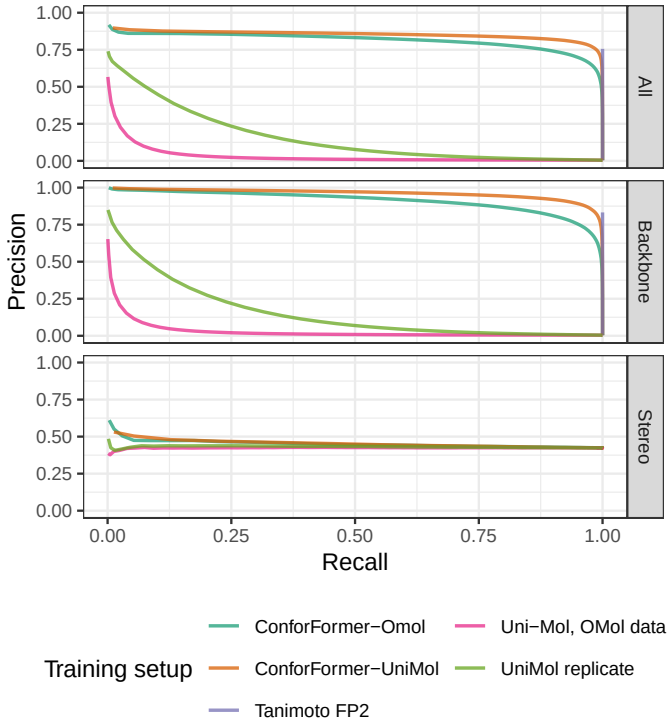

Fig. 6 Precision and recall curves for different frozen representations on Pharmaisomer benchmarks.

$x_i, x_i' \in \mathbb{X}$ and $y_i \in \{0,1\}$ indicates the pair type: $y_i = 1$ for *conformers* and $y_i = 0$ for *isomers*. Define index sets

$$\mathbb{C} := \{\, i \in \{1,\dots,N\} : y_i = 1 \,\}, \qquad \mathbb{I} := \{\, i \in \{1,\dots,N\} : y_i = 0 \,\},$$

with counts $N_C := |\mathbb{C}|$ and $N_I := |\mathbb{I}|$ (so $N = N_C + N_I$).

Reusing the same similarity as in pretraining, define

$$s_i \;:=\; \mathrm{sim}\big(f(x_i), f(x_i')\big) \in [0,1].$$

For a threshold $\theta \in [0,1]$, predict *conformer* when $s_i \geq \theta$:

$$\hat{y}_i(\theta) \;:=\; \mathbf{1}_{s_i \geq \theta}\,.$$

Confusion counts and metrics are then defined as follows:

$$\mathrm{TP}(\theta) := \sum_{i\in\mathbb{C}} \mathbf{1}_{s_i\geq\theta}\,, \qquad \mathrm{FN}(\theta) := \sum_{i\in\mathbb{C}} \mathbf{1}_{s_i<\theta}\,,$$

$$\mathrm{FP}(\theta) := \sum_{i\in\mathbb{I}} \mathbf{1}_{s_i\geq\theta}\,, \qquad \mathrm{TN}(\theta) := \sum_{i\in\mathbb{I}} \mathbf{1}_{s_i<\theta}\,.$$

$$\mathrm{Prec}(\theta) := \frac{\mathrm{TP}(\theta)}{\mathrm{TP}(\theta)+\mathrm{FP}(\theta)}, \qquad \mathrm{Rec}(\theta) := \frac{\mathrm{TP}(\theta)}{\mathrm{TP}(\theta)+\mathrm{FN}(\theta)} = \frac{\mathrm{TP}(\theta)}{N_C}.$$

The precision/recall curves constructed by sweeping over $\theta \in [0,1]$ can be found on Figure 6. In this analysis, we treat enantiomers (mirror isomers) as the same molecule; the Uni-Mol backbone is based on a distance matrix, therefore has E(3) symmetry[23] and treats enantiomers as the same by design. For Uni-Mol replicate, the precision at 50% recall was just 8%; for ConforFormer–OMol it was above 83%, with most of the errors coming from the low capability of the model to recognize diastereomers (on backbone isomers its precision at 50% recall was 94%).

Notably, post-training the model on the train part of the PharmaIsomer dataset saturates the backbone part of the benchmark with 99.9% precision at 50% recall but still reaches just 56% precision at 50% recall for diastereomers. For both isomers and diastereomers, the precision of the model is on par or higher than of a baseline utilizing the Tanimoto similarity between the FP2 fingerprints of the molecule pairs as the similarity score s_i^T. This representation has 100% recall by design at $s_i^T = 1$, but it cannot be adjusted to obtain higher precision.

The precision and recall curves (Figure 6) for recognizing isomers of molecules outside both Uni-Mol and OMol training datasets conclusively show that our model has obtained the capability to make inference about unique chemical structures without being directly trained on molecular graphs. While Uni-Mol replicate model seems to consider overall shape of the molecule more in making these assessments, ConforFormer–OMol recognizes similarity based on underlying molecular graph which it inferred from training with the novel contrastive objective. See Figure 7 for an example of conformers with very dissimilar shape and Figure 8 for a pair of isomers with an overall similar one. Both have the same similarity of 0.93 in the Uni-Mol embedding space but differ strongly (0.99 vs 0.26) in the ConforFormer–OMol one. Section F of the SI contains other examples of the models' disagreements in similarity evaluations for conformer and iso-

View Article Online
DOI: 10.1039/D6DD00096G

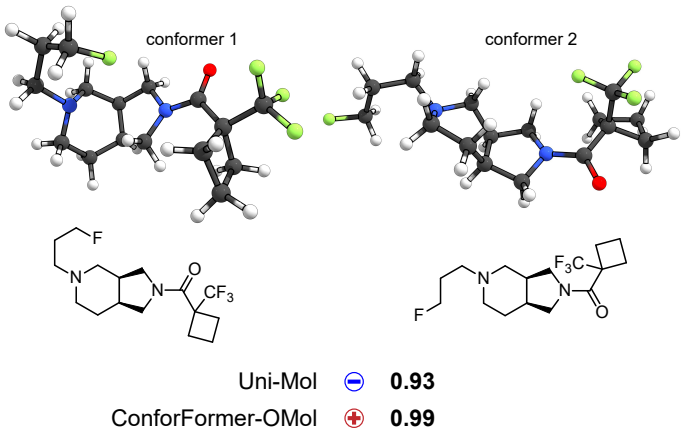

Fig. 7 A pair of conformers of the same molecule having similarity of 0.93 in the Uni-Mol embedding space and 0.99 in ConforFormer–OMol (indicating they belong to the same molecule)

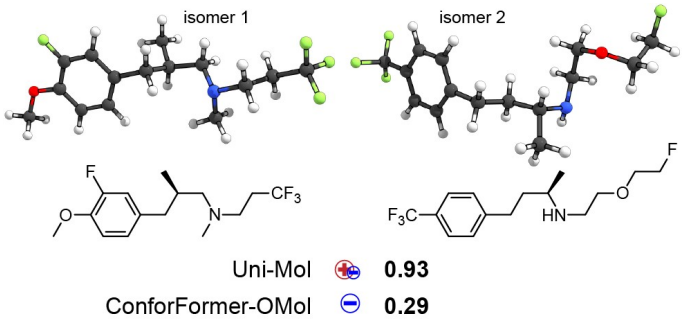

Fig. 8 A pair of isomers (distinct molecules) with similarity of 0.93 in the Uni-Mol embedding space but 0.29 in ConforFormer–OMol (indicating they are distinct)

mer pairs. Exploring the similarity relationships beyond the conformer/isomer pairs, molecules with close embeddings but not isomers tend to be chemically similar. At cosine similarity of ConforFormer-OMol embeddings > 0.90 the molecular pairs invariably share most of the backbone; at similarities > 0.75 the molecules typically share some major structural motif. A detailed study of the practical applicability of searching for such close neighbors in drug discovery and related tasks can be a focus of a future study.

5 Conclusions

In this paper, we introduce ConforFormer, a geometry-first framework that turns 3D molecular structures into compact, task-agnostic embeddings by explicitly enforcing conformational equivalence during pre-training. Built on the Uni-Mol backbone, ConforFormer adds a novel contrastive objective that aligns representations of different conformers of the same molecule while maintaining separation between different molecular identities, yielding a compact 512-dimensional vector representation directly from atomic identities and coordinates.

These frozen embeddings are directly usable and show competitive performance across quantum-chemical, physico-chemical and bioactivity benchmarks. We observe systematic improvements when conformer alignment is included in the training objectives, especially when high-quality geometries are used during the pre-training (ConforFormer-OMol). Our results suggest that the conformation-aware pre-training can produce representations that transfer robustly beyond the pre-training objective, including in small-data regimes where fully unfrozen fine-tuning would give rise to instabilities.

Beyond property prediction, we find that the learned embedding space readily supports chemist-interpretable similarity analysis without task-specific retraining. On the PharmaIsomer benchmark, ConforFormer embeddings cleanly separate conformers from isomers and substantially outperform classical fingerprint similarity and embeddings obtained without the contrastive objective in terms of precision at a given recall. For example, at 50% recall, precision increases from about 8% for a Uni-Mol replicate to >83% for ConforFormer-OMol, with most residual errors attributable to challenging sterochemical cases (notably, diastereomers). These results point to an emergent ability to encode graph-like structural constraints from 3D geometries alone, even though the model is not trained with an explicit objective to distinguish conformers from isomers.

From the practical perspective, the direct access to such robust molecular embeddings provides a computationally efficient alternative to retraining large backbone for each downstream application. For example, similarity search and screening can be performed directly in the learned embedding space, avoiding the need for task-specific objectives or dedicated fingerprint engineering. In line with this, our similarity analysis task on the PharmaIsomer dataset shows that nearest-neighbor relationships inferred directly from the embeddings enable efficient and chemically reasonable notion of closeness, while remaining far cheaper compared to full large-scale retraining on often proprietary pharma-related molecular datasets. Future work will focus on improving stereochemical sensitivity, where the current E(3)-invariant design is limiting, better modeling of conformational distributions and geometry quality, and extending conformer/isomer labeling via molecular dynamics simulations to more complex and fluxional chemical systems (including organometallic and coordination compounds), potentially augmented by additional training objectives.

Author contributions

M.P. Klein: methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, visualization. I. Rudenko: conceptualization, methodology, formal analysis, investigation, writing – original draft, writing – review & editing. I. Bushmarinov: conceptualization, supervision, methodology, formal analysis, investigation, writing – original draft, writing – review & editing. E. A. Pidko: supervision, conceptualization, resources, funding acquisition, writing – review & editing, visualization, project administration.

Conflicts of interest

There are no conflicts to declare.

View Article Online
DOI: 10.1039/D6DD00096G

Data availability

All of the code used to pre-train the models, fine-tune them, build the contrastive benchmarks and datasets, measure the results reported in Figures 4 and 3, and plot Figures 5 and 6 is available at a GitHub repository `https://github.com/EPiCs-group/ConforFormer`. The model weights are published to HuggingFace `https://huggingface.co/ConforFormer/ConforFormer`. A sample of the PharmaIsomer dataset is available in the ConforFormer GitHub repository and the full dataset in Zenodo[48]. The model training hyperparameters, dataset descriptions, and ablation study details can be found in the Supporting Information.

Acknowledgements

E.A.P. thanks NWO Domain Science for support in the framework of DynaCat VICI grant (VI.C.242.082, https://doi.org/10.61686/WNCYG94137). The use of supercomputer facilities was sponsored by NWO Domain Science (2024.008).

Notes and references

1 J. Choi, G. Nam, J. Choi and Y. Jung, *JACS Au*, 2025, **5**, 1499–1518.

2 F. Wiesner, M. Wessling and S. Baek, *Towards a Physics Foundation Model*, 2025, `http://arxiv.org/abs/2509.13805`, arXiv:2509.13805 [cs].

3 C. Bodnar, W. P. Bruinsma, A. Lucic, M. Stanley, A. Allen, J. Brandstetter, P. Garvan, M. Riechert, J. A. Weyn, H. Dong, J. K. Gupta, K. Thambiratnam, A. T. Archibald, C.-C. Wu, E. Heider, M. Welling, R. E. Turner and P. Perdikaris, *Nature*, 2025, **641**, 1180–1187.

4 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*, 2022, `https://chemrxiv.org/doi/full/10.26434/chemrxiv-2022-jjm0j-v4`, ChemRxiv.

5 W. Ahmad, E. Simon, S. Chithrananda, G. Grand and B. Ramsundar, *ChemBERTa-2: Towards Chemical Foundation Models*, 2022, `http://arxiv.org/abs/2209.01712`, arXiv:2209.01712 [cs].

6 Y. Wang, J. Wang, Z. Cao and A. Barati Farimani, *Nature Machine Intelligence*, 2022, **4**, 279–287.

7 S. Chithrananda, G. Grand and B. Ramsundar, *ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction*, 2020, `http://arxiv.org/abs/2010.09885`, arXiv:2010.09885 [cs].

8 X. Fang, L. Liu, J. Lei, D. He, S. Zhang, J. Zhou, F. Wang, H. Wu and H. Wang, *Nature Machine Intelligence*, 2022, **4**, 127–134.

9 D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher and A. G. Doyle, *Science*, 2018, **360**, 186–190.

10 M. Kuznetsov, F. Ryabov, R. Schutski, R. Shayakhmetov, Y.-C. Lin, A. Aliper and D. Polykovskiy, *Journal of Chemical Information and Modeling*, 2024, **64**, 3610–3620.

11 R. Laplaza, M. D. Wodrich and C. Corminboeuf, *The Journal of Physical Chemistry Letters*, 2024, **15**, 7363–7370.

12 S. Finta, A. V. Kalikadien and E. A. Pidko, *Journal of Chemical Theory and Computation*, 2025, **21**, 5334–5345.

13 Q. Jiang, C. Chen, H. Zhao, L. Chen, Q. Ping, S. D. Tran, Y. Xu, B. Zeng and T. Chilimbi, 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 7661–7671.

14 K. E. Ak, J. Mohta, D. Dimitriadis, S. Manchanda, Y. Xu and M. Shen, Computer Vision – ECCV 2024 Workshops: Milan, Italy, September 29–October 4, 2024, Proceedings, Part XVIII, Berlin, Heidelberg, 2025, pp. 32–45.

15 L. Wang, N. Yang, X. Huang, B. Jiao, L. Yang, D. Jiang, R. Majumder and F. Wei, *Text Embeddings by Weakly-Supervised Contrastive Pre-training*, 2024, `http://arxiv.org/abs/2212.03533`, arXiv:2212.03533 [cs].

16 K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen and R. Barzilay, *Journal of Chemical Information and Modeling*, 2019, **59**, 3370–3388.

17 Y. Rong, Y. Bian, T. Xu, W. Xie, Y. WEI, W. Huang and J. Huang, Advances in Neural Information Processing Systems, 2020, pp. 12559–12571.

18 D. Weininger, *Journal of Chemical Information and Computer Sciences*, 1988, **28**, 31–36.

19 J. Li and X. Jiang, *Wireless Communications and Mobile Computing*, 2021, **2021**, 7181815.

20 C. Liu, Y. Sun, R. Davis, S. T. Cardona and P. Hu, *Journal of Cheminformatics*, 2023, **15**, 1–14.

21 S. Lu, Z. Gao, D. He, L. Zhang and G. Ke, *Nature Communications*, 2024, **15**, 7104.

22 X. Ji, Z. Wang, Z. Gao, H. Zheng, L. Zhang, G. Ke and W. E, The Thirty-eighth Annual Conference on Neural Information Processing Systems, 2024.

23 A. Dumitrescu, D. Korpela, M. Heinonen, Y. Verma, V. Iakovlev, V. Garg and H. Lähdesmäki, The Thirteenth International Conference on Learning Representations, 2024.

24 J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, 2019, pp. 4171–4186.

25 N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. D. Laroussilhe, A. Gesmundo, M. Attariyan and S. Gelly, Proceedings of the 36th International Conference on Machine Learning, 2019, pp. 2790–2799.

26 A. Wang, A. Singh, J. Michael, F. Hill, O. Levy and S. Bowman, Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Brussels, Belgium, 2018, pp. 353–355.

27 P.-E. Sarlin, D. DeTone, T. Malisiewicz and A. Rabinovich, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 4937–4946.

28 Y. A. Malkov and D. A. Yashunin, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **42**, 824–836.

29 A. Martin Pendas and E. Francisco, *Nature Communications*,

View Article Online
DOI: 10.1039/D6DD00096G

2022, **13**, 3327.

30 M. Weisberg, *Philosophy of Science*, 2008, **75**, 932–946.

31 S. Shaik, *Journal of Computational Chemistry*, 2007, **28**, 51–61.

32 G. Frenking and A. Krapp, *Journal of Computational Chemistry*, 2007, **28**, 15–24.

33 S. Alvarez, R. Hoffmann and C. Mealli, *Chemistry – A European Journal*, 2009, **15**, 8358–8373.

34 R. Hoffmann and P. Laszlo, *Angewandte Chemie International Edition in English*, 1991, **30**, 1–16.

35 G. N. Lewis, *Journal of the American Chemical Society*, 1916, **38**, 762–785.

36 R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, New York, 1990.

37 R. F. W. Bader, *The Journal of Physical Chemistry A*, 1998, **102**, 7314–7323.

38 M. Brookhart, M. L. H. Green and G. Parkin, *Proceedings of the National Academy of Sciences*, 2007, **104**, 6908–6914.

39 S. Bougueroua, A. A. Kolganov, C. Helain, C. Zens, D. Barth, E. A. Pidko and M.-P. Gaigeot, *Physical Chemistry Chemical Physics*, 2025, **27**, 1298–1309.

40 A. V. Kalikadien, C. Valsecchi, R. v. Putten, T. Maes, M. Muuronen, N. Dyubankova, L. Lefort and E. A. Pidko, *Chemical Science*, 2024, **15**, 13618–13630.

41 T. Chen, S. Kornblith, M. Norouzi and G. Hinton, Proceedings of the 37th International Conference on Machine Learning, 2020, pp. 1597–1607.

42 D. S. Levine, M. Shuaibi, E. W. C. Spotte-Smith, M. G. Taylor, M. R. Hasyim, K. Michel, I. Batatia, G. Csányi, M. Dzamba, P. Eastman, N. C. Frey, X. Fu, V. Gharakhanyan, A. S. Krishnapriyan, J. A. Rackers, S. Raja, A. Rizvi, A. S. Rosen, Z. Ulissi, S. Vargas, C. L. Zitnick, S. M. Blau and B. M. Wood, *The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models*, 2025, `http://arxiv.org/abs/2505.08762`, arXiv:2505.08762 [physics].

43 S. Riniker and G. A. Landrum, *Journal of Chemical Information and Modeling*, 2015, **55**, 2562–2574.

44 T. A. Halgren, *Journal of Computational Chemistry*, 1996, **17**, 490–519.

45 T. Chen and C. Guestrin, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

46 D. Rogers and M. Hahn, *Journal of Chemical Information and Modeling*, 2010, **50**, 742–754.

47 J. J. Irwin, K. G. Tang, J. Young, C. Dandarchuluun, B. R. Wong, M. Khurelbaatar, Y. S. Moroz, J. Mayfield and R. A. Sayle, *Journal of Chemical Information and Modeling*, 2020, **60**, 6065–6073.

48 M. Klein, I. Rudenko, E. Pidko and I. Bushmarinov, *PharmaIsomer*, 2026, `https://zenodo.org/records/18739668`.

View Article Online
DOI: 10.1039/D6DD00096G

Data availability

All of the code used to pre-train the models, fine-tune them, build the contrastive benchmarks and datasets, measure the results re ported in Figures 4 and 3, and plot Figures 5 and 6 is avail able at a GitHub repository https://github.com/EPiCs-group/ ConforFormer. The model weights are published to HuggingFace https://huggingface.co/ConforFormer/ConforFormer. A sample of the PharmaIsomer dataset is available in the Confor Former GitHub repository and the full dataset in Zenodo48. The model training hyperparameters, dataset descriptions, and abla tion study details can be found in the Supporting Information.

View Article Online
DOI: 10.1039/D6DD00096G

CONFORFORMER: REPRESENTATION FOR MOLECULES THROUGH UNDERSTANDING OF CONFORMERS

Mas Pieter Klein[a], Irina Rudenko[b], Evgeny A. Pidko[a], Ivan Bushmarinov[c]

Supporting Information

[a] Department of Chemical Engineering, TU Delft, Delft, Netherlands
[b] Avride Inc., Austin, TX, USA
[c] Perplexity AI, Belgrade, Serbia

View Article Online
DOI: 10.1039/D6DD00096G

A Training details

A.1 General Remarks

All code is available at the GitHub repository `https://github.com/EPiCs-group/ConforFormer`. The model architecture throughout this whole document is that of the standard Uni-Mol up until contrastive learning is done. This can be found in Appendix C (Table 6) in Zhou *et al.*[1]. Set-up is kept identical, regardless of whether the Uni-Mol, OMol, or contrastive benchmark is used as training data. The pre-trained models parameters can be found in the following HuggingFace repository `https://huggingface.co/ConforFormer/ConforFormer` The settings for the Uni-Mol replication and ConforFormer models are listed in A.2 and A.3 respectively. A three-layer $512 \times 256 \times 128$ MLP was used for fine-tuning with exactly the same settings as in Zhou *et al.*[1]. See the Ablation Studies (section E in the Supporting Information) for details of other experiments.

A.2 Uni-Mol replication

Training was carried out using the Uni-Core framework (version 0.0.3).

Hyperparameters

- `masked_token_loss = 1`
- `masked_coord_loss = 5`
- `masked_dist_loss = 10`
- `x_norm_loss = 0.01`
- `delta_pair_repr_norm_loss = 0.01`
- `mask_prob = 0.15`
- `noise_type = "uniform"`
- `noise = 1.0`
- `only_polar = 0` (no hydrogens on the molecule)
- `dropout = 0.1` (applied to FFN, attention heads, etc.)
- Activation functions are always GeLU
- `batch size = 128`

View Article Online
DOI: 10.1039/D6DD00096G

Training details

- Linear learning rate schedule
- 10,000 warm-up steps
- 1,000,000 total steps
- Validation every 10,000 steps
- Adam optimizer
- $\epsilon = 1 \times 10^{-6}$
- $\beta = (0.9, 0.99)$
- Weight decay of 1×10^{-4}

A.3 ConforFormer models

All hyperparameters remained the same as in A.2 except for the learning rate and the learning rate scheduling:

- Learning rate schedule altered to `ReduceLROnPlateau`
  - Patience = 3
  - $\epsilon = 0.25$
  - lr-shrink = 0.5
- 5,000 warm-up steps
- Validation was done every 5000 steps, starting after warm-up
- Peak learning rate of 5×10^{-4}
- Batch size of 128

A.4 Reduced UniMol dataset

For quick iterations and experiments, we used a setup which could be trained to convergence overnight on a single NVIDIA H100 GPU. For that, we chose 1/8 of the Uni-Mol dataset by simply iterating over it and selecting every 8th data point:

```
for i, datapoint in enumerate(dataset):
    if i % 8 == 0:
        new_dataset.put(datapoint)
```

View Article Online
DOI: 10.1039/D6DD00096G

For these experiments, the hyperparameters remained the same as in A.2 except for the learning rate, learning rate schedule, and batch size, unless specified otherwise.

- Learning rate schedule altered to `ReduceLROnPlateau`
  - Patience = 2
  - $\epsilon = 0.25$
  - lr-shrink = 0.5
- 5,000 warm-up steps
- Validation was done every 5000 steps, starting after warm-up
- Peak learning rate of 5×10^{-4}
- batch size = 384

B Datasets and benchmarks

B.1 Uni-Mol

The training split of Uni-Mol[a], as detailed in Zhou *et al.*[1], consists of 18.8M unique molecules each with 10 conformations, resulting in ca. 190M datapoints. These 10 conformations were all generated using RDKit. On average, each datapoint has 27 heavy atoms. A large portion of the dataset consists of organic molecules. The dataset contains 67 unique heavy atoms, with C, O, and N making up greater than 95%. The remaining 5% consists almost exclusively of the halogens (F, Cl, Br, and I) along with P and S. Consequently, there are only 9 heavy atom types that have a share greater than 0.01%. The validation split consists of ≈100K unique molecules, again with 10 conformations each. Of the 21 unique heavy atoms, the same 9 atoms have a share greater than 0.01% of the dataset. After the reduction of the dataset, taking every eighth datapoint, the relative distribution of all heavy atoms remains the same.

B.2 OMol

The full Open Molecules (OMol) dataset[2] consists of various molecules which are relevant to homogenous catalysis, electrolytes, and biomolecular systems. Structures were calculated at the ωB97M-V/def2-TZVPD level of theory, resulting in unquestionably higher quality data than that found in Uni-Mol. In total there are 101M unique datapoints[b]. On average the molecules have 26 heavy atoms. C, O, and N make up 91% of all heavy atoms. OMol has a larger atom variety than Uni-Mol with 83 unique atom types, 59 having a share larger than 0.01% divided amongst various charge and spin states.

[a] Accessed on 2025/24/04
[b] Accessed on 2025/26/06

View Article Online
DOI: 10.1039/D6DD00096G

Reducing the dataset to that of only molecules with at least 2 heavy atoms and at least two conformations (the setup used to train ConforFormer–OMol) results in a dataset of 8.25M unique molecules, 55M total datapoints (further OMol-conf). On average, each molecule has 6-7 unique conformations. All 83 atom types remain in the dataset and now 64 have at least a 0.01% share of all heavy atoms.

B.3 MoleculeNet

MoleculeNet[2] is a collection of molecular benchmarks that contains tasks relevant to Physiology, Biophysics, Physical chemistry, or Quantum mechanics. While we acknowledge the documented limitations of the MoleculeNet benchmarks - for example, suboptimal dataset splitting, inconsistent molecular representations, and curation errors[3] — these datasets remain the most widely adopted standards in the field. Consequently, they provide an essential framework for benchmarking against a vast body of existing literature and established models, as well as the UniMol architecture we utilize. Classification tasks are always evaluated using ROC–AUC score, while regression tasks are either evaluated using RMSE or MAE. Generally, the benchmarks contain primarily organic molecules (compositions of C, N, O), with these atoms accounting for anywhere between 70% to almost 100% of all heavy atoms (all atoms excluding H) within a benchmark. For all tasks, SMILES strings are provided alongside the targets, with only a select few (QMx, $x \in \{7, 8, 9\}$, the quantum mechanics based benchmarks) having provided 3D coordinates. As such, the structures generated for every 3D coordinate are those already made by the team of UniMol. In the cases where 3D coordinates are provided, only one per molecule is present, and 9 more were generated. The benchmarks are used as provided by the UniMol team [c].

B.3.1 BACE

BACE is a classification benchmark with 1 target and contains slightly over 1500 molecules. The target is a binary label which qualitatively describes a molecule's ability to inhibit the human beta-secretase 1 (BACE-1). The molecules within the benchmark are purely organic, containing on average 34 heavy atoms, primarily C, N, O, and S. The halogens account for slightly over 1% of heavy atoms.

B.3.2 BBBP

BBBP is a classification benchmark with labels indicating whether a molecule can or cannot penetrate the blood-brain barrier. The $\approx$ 2000 molecules are primarily organics and occasionally halogenated. There are small amounts of salts, specifically alkai (earth) metals with 21 Na atoms and 1 Ca atom. On average, they contain 24 heavy atoms.

[c] Accessed on 2025/25/04

View Article Online
DOI: 10.1039/D6DD00096G

B.3.3 Clintox

Clintox is a classification task, describing drug-like molecules using qualitative data of those approved by the FDA or failed due to toxicity. It contains ≈ 1500 molecules, primarily of organic nature, with 26 heavy atoms. There are small amounts of main-group and d-block atoms. Curiously, all of this atomic variety is found in either the train or test splits.

B.3.4 ESOL

ESOL is a benchmark of 1100 small (≈ 13 heavy atoms) organic molecules, occasionally halogenated (F, Cl, Br and I accounting for ≈ 5% of heavy atoms, primarily Cl). The target is the log solubility of a molecule in water (in mol/L). It is evaluated using RMSE. Disproportionally few heavier halogens (Br and I) are in the test benchmark, specifically none.

B.3.5 FreeSolv

FreeSolv is a benchmark of small organic molecules and their experimental or calculated solvation energy in water (in kcal/mol). Performance is evaluated using RMSE. It contains 642 small organic molecules (on average 9 heavy atoms), occasionally halogenated.

B.3.6 HIV

HIV is a classification benchmark asking a model to distinguish between molecules which do or do not inhibit HIV replication. It is primarily organic molecules (C, O, N accounting for 95% of all heavy atoms), however still contains a large variety of alkali (earth) metals, d-block metals (frequently containing almost all occurrences of a d-block metal through the whole of MoleculeNet), and main-group elements. On average, the 41K datapoints contain approx. 25 heavy atoms.

B.3.7 Lipo

Lipo is a regression benchmark requiring a model to predict experimentally determined log(P) (octanol/water partition coefficient) at a pH of 7.4. It 4,200 organic molecules with on average 27 heavy atoms. The performance of a model is evaluated using RMSE.

B.3.8 MUV

MUV is a benchmark of 93K organic molecules of 24 heavy atoms on average. It contains the classification task of 17 targets, determined through high-through put experiments on BioAssays. It is a subset of datapoints contained in PCBA, refined through nearest-neighbor analysis and is meant to validated virtual screening methods.

B.3.9 PCBA

PCBA is a benchmark of selected PubChem BioAssay consisting of results of high-throughput experiments on the biological activity of small molecules. It has 128 targets. On average, each

View Article Online
DOI: 10.1039/D6DD00096G

datapoint has 26 heavy atoms, primarily C, N ,O (accounting for 95% of all heavy atoms in the benchmark), and various main-group and d-block elements. It contains 438K datapoints. The fine-tuning result of ConforFormer–OMol on this benchmark was 0.829 (on par with literature data) but due to its size we did not run it for most of the models in the study and do not include it in the tables.

B.3.10 QM7, QM8, QM9

The QMx benchmarks consist of small organic (QM8 and QM9 occasionally halogenated) molecules with 7, 8, or 9 heavy atoms. They are regression benchmarks, performance measured in MAE, with the aim to predict various quantum-mechanically properties, such as atomization energy and HOMO-LUMO gap. They contain approx. 7K, 22K, and 130K datapoints, respectively.

B.3.11 Sider

SIDER is a classification benchmark with 27 targets, aiming to predict if a marketed drug has adverse drug reactions to 27 system organ classes. It consists of generally large molecules (average of 33 heavy atoms), primarily of organic nature. However, it does contain a variety of main-group and d-block elements. This atomic variety finds itself almost exclusively in the training split, frequently over 95% of these atoms. It contains 1427 datapoints.

B.3.12 Tox21

Tox21 is a classification benchmark of 7831 datapoints and has 12 targets. These targets are binary labels as to whether a molecule has any qualitative toxicity on 12 biological systems. Datapoints contain 18.5 heavy atoms on average, primarily organic in nature. There are also small amount of main-group and d-block elements, primarily found within the training data ($> 90\%$ occurrence in training split).

B.3.13 Toxcast

Toxcast has 617 binary classification targets representing qualitative toxicology data generated using in-vitro high-throughput experimentation. The ca. 8600 datapoints usually contain 19 heavy atoms, primarily C, O, N, and halogens. Main-group and d-block elements make up about 1% of all heavy atoms.

C PharmaIsomer dataset

To construct this benchmark, we used a portion of ZINC20[4] [d]. This database consists of commercially available medicinal molecules for virtual screening. For our purposes, it provides a large selection of relevant and valid molecular structrues from which to generate molecular conformations.

[d] Accessed on 2025/14/07

View Article Online
DOI: 10.1039/D6DD00096G

From the ZINC20 dataset, we took mildly reactive, relatively easy to purchase molecules, which in the database are marked as having reactivity up to and including "standard" and purchase up to and including "wait OK." We removed all overlap with Uni-Mol and OMol datasets using SMILES-identified unique formulas and also removed molecules that do not have a chemical formula. Ten conformers of each SMILES string were generated using ETKDGv1,[5] implemented in rdKit, and optimized with the MMFF94 force field.[6] To filter conformations, a root-mean-squared threshold of 0.5 was used. Only structurally distinct conformers were used and special attention was put to avoid duplicate conformers.

To ensure that only isomeric structures are assessed for similarity, simplify inference, and make metrics between runs comparable, batches were statically constructed beforehand rather than dynamically produced at inference time. Specifically, each batch contained 128 unique molecules, all isomers to each other. Each isomer had exactly 2 conformers, resulting in 256 datapoints per batch. An 80/10/10 train/test/validation split was employed so that the performance of models trained on this dataset could be evaluated; metrics in the main text are reported on the validation split.

The dataset contains 3,261,807,960 datapoints in 12,741,440 batches and is freely available under a CC-BY license[7].

D LLM usage

Large language models were used for grammar checking, LaTeX formatting, initial literature search, generation of the SQLite processing code used to support isomer classification analysis, and generation of code for the CatBoost/FP4 and XGBoost/ECFP4 (1024-bit) baselines. The initial literature search was performed with OpenAI o1 (Deep Research), code generation and LaTeX editing with OpenAI GPT 5.1 (initial draft), 5.2 (second draft) and 5.4 (final draft). The edits to the paper text performed by LLMs were limited to LaTeX syntax corrections and table formatting.

E Ablation studies

E.1 Overview

The table S3 contains details of the specific experiments we ran, with the following table S4 containing benchmark results for various pre-training and post-training setups with different number of unfrozen layers (15 corresponding to a fully unfrozen model). The"ConforFormer" objective refers to the loss $\mathcal{L}_{total}$ as described in 3.2. Abbreviated names for experiments are used throughout these tables, with the following mapping to the main text entries:

- **U**: Uni-Mol replicate
- **U-no-flat**: Uni-Mol no "flat"
- **O-c**: Uni-Mol, OMol data

View Article Online
DOI: 10.1039/D6DD00096G

- **CF-O-c**: ConforFormer–OMol
- **CF-U**: ConforFormer–UniMol
- **Random-w**: Uni-Mol no pretrain

An important note from the ablation study is that the addition of contrastive learning in-post provides no improvement to the model. In fact, it actually worsens performance. Comparing the Cpost runs in Table S4, an increase in temperature (τ) is seen to worsen finetuning results. Taking the results of BBBP as an example, the ROC–AUC score changes as $0.656 \rightarrow 0.562 \rightarrow 0.544$ as τ goes from 0.01 to 0.1 and finally 0.5. We used these results to guide our choice of τ for ConforFormer, but we checked that increasing the temperature of the NT-Xent loss in pre-training worsens the performacne as well (see **CF-U-r-0.25**). As the model is made to differentiate more, it performs worse. Compared to the results of "ConforFormer–OMol", the best performing model with contrastive pose-training results in worse metrics after finetuning.

A contrastive-only pre-training on the Unimol dataset (**Contrast-U-r**), on the other hand, results in a below average quality of the embeddings but still performs surprisingly well, suggesting that contrastive objective alone could also be potentially viable strategy for model pre-training.

Sanity checks were run to validate that the ConforFormer loss and not changes in the training setup were actually driving the metric improvements. However, changing the batch size (**U-r-128**) or adding more conformers in each batch (**U-r-256-conf**) did not lead to any noticeable improvements. Training the Uni-Mol model without additional contrastive loss on the full OpenMolecules dataset (**O**) did not improve benchmarks beyond ConforFormer–OMol **CF-O-c** or Uni-Mol replicate **U** either.

Given the results of models trained using contrastive learning on PharmaIsomer, it was hypothesized that doing training that mimics this benchmark would improve performance. For this, the Uni-Mol data set was tailored to allow for the construction of batches that guarantee a minimum number of isomers. The results of these pre-training can be found in the **CF-U-xI** setups, $x \in \{10, 25, 40\}$ in Table S4 . These 3 rows represent a minimum of 40%, 25%, and 10% of datapoints within a batch having an isomer pair. No noticeable improvement is observed within the classification tasks and no significant trend is observed for the regression tasks. Thus, the hypothesis of additional isomers in the batch is not seen to be true for the Uni-Mol dataset with the contrastive training set-up outlined in Section A.3.

Each line in Table S4 corresponds to a single fine-tuning run. For the stability-analysis context, refer to the transfer-learning discussion in Section 2.3 of the main text.

As a final ablation, the number of conformers used during finetuning was reduced from 10 to 1, 2, or 5. The results are shown in Tables S5 and S6. Regardless of setup,

View Article Online
DOI: 10.1039/D6DD00096G

E.2 Full run data

The labels for pre-training runs from Table S3 are used throughout Table S4, which contains fine-tuning results with different number of model layers unfrozen. Layers were frozen starting from the last. Finetuning was performed using the hyperparameters specified in the Uni-Mol GitHub repository `https://github.com/deepmodeling/Uni-Mol/tree/main/unimol`. A typical fine-tuning job took ≈ 3 hours on an A100 GPU. For the fine-tuning starting from **Random-w**, all batch sizes were set to 128, and training was stopped after no improvement was made after 40 epochs (18 hours on an H100). The remaining hyperparameters are identical to those under standard finetunings.

Table S1: Performance on quantum-chemical and physicochemical regression benchmarks. Values denote root-mean-square deviation (RMSD), with lower values indicating better agreement with reference data. Literature results are reproduced from the Uni-Mol benchmark study. Results from the current work are grouped into unfrozen models, where the encoder is fine-tuned for each task, frozen models, where representations from a fixed pre-trained encoder are used without backbone fine-tuning, and 2D fingerprint baselines. Dataset sizes (N) indicate the number of molecules included in each benchmark. See Section B for dataset descriptions.

Model	ESOL	FreeSolv	Lipo	QM7	QM8	QM9
N points	1128	642	4200	6830	21786	133885
Literature data						
D-MPNN	1.05(1)	2.08(8)	0.683(16)	103.5(86)	0.0190(1)	0.00814(1)
Attentive FP	0.88(3)	2.07(18)	0.721(1)	72.0(27)	0.0179(10)	0.00812(1)
N-GramRF	1.07(11)	2.69(8)	0.812(28)	92.8(40)	0.0236(6)	0.01037(16)
N-GramXGB	1.08(8)	5.06(74)	2.072(30)	81.9(19)	0.0215(5)	0.00964(31)
PretrainGNN	1.10(1)	2.76(0)	0.739(3)	113.2(6)	0.0200(1)	0.00922(4)
$\text{GROVER}_{\text{base}}$	0.98(9)	2.18(5)	0.817(8)	94.5(38)	0.0218(4)	0.00984(55)
$\text{GROVER}_{\text{large}}$	0.90(2)	2.27(5)	0.823(10)	92.0(9)	0.0224(3)	0.00986(25)
GraphMVP	1.03(3)	–	0.681(10)	–	–	–
MolCLR	1.27(4)	2.59(25)	0.691(4)	66.8(23)	0.0178(3)	–
GEM	0.80(3)	1.88(9)	0.660(8)	58.9(8)	0.0171(1)	0.00746(1)
Uni-Mol	0.79(3)	1.48(5)	0.603(10)	41.8(2)	0.0156(1)	0.00467(4)
Current work						
Unfrozen models						
Uni-Mol replicate	0.83(3)	1.80(11)	0.608(9)	58.8(30)	0.0160(1)	0.00520(0)
ConforFormer–OMol	0.91(2)	1.99(5)	0.642(11)	53.8(18)	0.0159(0)	0.00542(4)
Uni-Mol no pretrain	0.98(5)	2.49(23)	0.787(22)	83.6(156)	0.0186(6)	0.00618(7)
Frozen models						
Uni-Mol replicate	1.15(3)	**2.64(6)**	0.916(4)	**82.6(44)**	0.0264(5)	0.0184(12)
Uni-Mol no "flat"	1.23(3)	2.92(4)	0.935(4)	88.5(32)	0.0263(2)	0.01910(19)
Uni-Mol, OMol data	1.18(1)	3.00(6)	0.949(6)	89.9(57)	0.0274(2)	0.0202(3)
ConforFormer–UniMol	1.17(3)	3.38(5)	0.807(5)	104.9(90)	0.0223(2)	0.01258(17)
ConforFormer–OMol	**1.12(2)**	3.53(7)	**0.752(7)**	99.9(112)	**0.0219(2)**	**0.01172(31)**
2D baselines						
XGBoost ECFP4 (1024-bit) baseline	1.562(19)	3.967(24)	0.864(5)	155.5(33)	0.02341(5)	0.01258(3)
CatBoost FP4 baseline	1.786(64)	3.400(32)	1.045(8)	129.2(58)	0.02632(18)	0.01704(62)

View Article Online
DOI: 10.1039/D6DD00096G

Table S2: Performance on biological activity classification benchmarks. Values report the mean ROC–AUC over multiple random splits (standard deviation in parentheses), with higher values indicating better performance. Literature results are taken from the original Uni-Mol study for reference. Results from the current work are reported separately for unfrozen models, where the encoder is fine-tuned together with the task-specific prediction head, frozen models, where molecular representations are extracted from a fixed pre-trained encoder and only lightweight task-specific models are trained, and 2D fingerprint baselines. Dataset sizes (N) correspond to the number of labeled molecules in each benchmark. See Section B for dataset descriptions.

Model	BBBP	BACE	ClinTox	Tox21	ToxCast	SIDER	HIV	MUV
N points	2039	1513	1478	7831	8575	1427	41127	93087
Literature data								
D-MPNN	0.710(3)	0.809(6)	0.906(6)	0.759(7)	0.655(3)	0.570(7)	0.771(5)	0.786(14)
Attentive FP	0.643(18)	0.784(0)	0.847(3)	0.761(5)	0.637(2)	0.606(32)	0.757(14)	0.766(15)
N-GramRF	0.697(6)	0.779(15)	0.775(40)	0.743(4)	–	0.668(7)	0.772(1)	0.769(7)
N-GramXGB	0.691(8)	0.791(13)	0.875(27)	0.758(9)	–	0.655(7)	0.787(4)	0.748(2)
PretrainGNN	0.687(13)	0.845(7)	0.726(15)	0.781(6)	0.657(6)	0.627(8)	0.799(7)	0.813(21)
$\text{GROVER}_{\text{base}}$	0.700(1)	0.826(7)	0.812(30)	0.743(1)	0.654(4)	0.648(6)	0.625(9)	0.673(18)
$\text{GROVER}_{\text{large}}$	0.695(1)	0.810(14)	0.762(37)	0.735(1)	0.653(5)	0.654(1)	0.682(11)	0.673(18)
GraphMVP	0.724(16)	0.812(9)	0.791(28)	0.759(5)	0.631(4)	0.639(12)	0.770(12)	0.777(6)
MolCLR	0.722(21)	0.824(9)	0.912(35)	0.750(2)	–	0.589(14)	0.781(5)	0.796(19)
GEM	0.724(4)	0.856(11)	0.901(13)	0.781(1)	0.692(4)	0.672(4)	0.806(9)	0.817(5)
Uni-Mol	0.729(6)	0.857(2)	0.919(18)	0.796(5)	0.696(1)	0.659(13)	0.808(3)	0.821(13)
Current work								
Unfrozen models								
Uni-Mol replicate	0.705(26)	0.832(25)	0.857(22)	0.788(3)	0.685(7)	0.644(14)	0.784(8)	0.784(9)
ConforFormer–OMol	0.691(21)	0.820(29)	0.686(26)	0.787(4)	0.689(7)	0.634(13)	0.786(4)	0.758(30)
Uni-Mol, no pretrain	0.655(11)	0.775(39)	0.639(46)	0.735(15)	0.635(12)	0.607(18)	0.739(18)	0.616(11)
Frozen models								
Uni-Mol replicate	0.640(4)	0.775(2)	**0.767(10)**	0.709(4)	**0.653(2)**	0.606(13)	0.734(3)	0.755(10)
Uni-Mol no "flat"	0.651(1)	0.778(1)	0.725(11)	0.711(1)	0.637(1)	0.606(6)	0.746(2)	0.742(8)
Uni-Mol, OMol data	0.664(4)	**0.783(4)**	0.698(11)	0.710(1)	0.629(1)	0.610(6)	**0.756(5)**	0.695(7)
ConforFormer–UniMol	0.665(7)	0.731(55)	0.533(14)	0.753(1)	0.644(3)	**0.649(2)**	0.711(4)	0.716(5)
ConforFormer–OMol	**0.673(6)**	0.751(13)	0.716(9)	**0.755(1)**	0.638(3)	0.640(6)	0.751(4)	**0.774(6)**
2D baselines								
XGBoost ECFP4 (1024-bit)	0.682(8)	0.805(35)	0.839(21)	0.724(4)	0.589(2)	0.618(11)	0.762(14)	0.774(26)
CatBoost FP4	0.675(6)	0.800(8)	0.816(53)	0.682(7)	0.577(5)	0.616(8)	0.720(16)	0.641(31)

View Article Online
DOI: 10.1039/D6DD00096G

Table S3: Training setups used in the experiments. Batch size refers to the pre-training procedure; when written as n_u (n_t), it denotes n_t geometries total in the batch, with n_u used to compute the Uni-Mol losses $\mathcal{L}_{token}$, $\mathcal{L}_{coord}$, and $\mathcal{L}_{distance}$. PharmaIsomer post-training batch size, where applicable, is always 256.

Training	Dataset	GPU	Hours	Objective	Batch size
CF-2U-r	1/8 Uni-Mol	H100	14	ConforFormer + Uni-Mol loss on full batch, $\tau = 0.07$	256
CF-O-c	OMol-conf	4×H100	24	ConforFormer, $\tau = 0.07$	128 (256)
CF-U	Uni-Mol	H100	48	ConforFormer, $\tau = 0.07$	128 (256)
CF-U-10I	Uni-Mol, 10% isomers	H100	48	ConforFormer, $\tau = 0.07$	128 (256)
CF-U-25I	Uni-Mol, 25% isomers	H100	48	ConforFormer, $\tau = 0.07$	128 (256)
CF-U-40I	Uni-Mol, 40% isomers	H100	48	ConforFormer, $\tau = 0.07$	128 (256)
CF-U-H	Uni-Mol with H	A100	72	ConforFormer, $\tau = 0.07$	128 (256)
CF-U-r	1/8 Uni-Mol	A100	14	ConforFormer, $\tau = 0.07$	128 (256)
CF-U-r-0.25	1/8 Uni-Mol	A100	14	ConforFormer, $\tau = 0.25$	128 (256)
Contrast-U-r	1/8 Uni-Mol	A100	10	Contrast loss only, $\tau = 0.07$	256
O	OMol	A100	120	Uni-Mol	384
O-c	OMol-conf	A100	120	Uni-Mol	128
O-Cpost-0.01	OMol	A100	120	Uni-Mol + contrast on PharmaIsomer in post-train, $\tau = 0.01$	384
O-Cpost-0.05	OMol	A100	120	Uni-Mol + contrast on PharmaIsomer in post-train, $\tau = 0.05$	384
Random-w	–	A100	–	Random weight initialization	–
U	Uni-Mol	A100	72	Uni-Mol	128
U-no-flat	Uni-Mol, no "flat"	A100	72	Uni-Mol	128
U-r	1/8 Uni-Mol	A100	10	Uni-Mol	384
U-r-128	1/8 Uni-Mol	A100	14	Uni-Mol	128
U-r-256	1/8 Uni-Mol	H100	14	Uni-Mol	256
U-r-256-conf	1/8 Uni-Mol, 2 conformers each	H100	14	Uni-Mol	256
U-r-Cpost-0.01	1/8 Uni-Mol	H100	14	Uni-Mol + contrast on PharmaIsomer in post-train, $\tau = 0.01$	384
U-r-Cpost-0.1	1/8 Uni-Mol	H100	14	Uni-Mol + contrast on PharmaIsomer in post-train, $\tau = 0.1$	384
U-r-Cpost-0.5	1/8 Uni-Mol	H100	14	Uni-Mol + contrast on PharmaIsomer in post-train, $\tau = 0.5$	384

Table S4: Benchmark results on MoleculeNet as detailed in B.3. "Unfrozen" refers to the number of model layers unfrozen diring the fine-tuning procedure. Left block: classification benchmarks. Right block: regression benchmarks.

Training	Unfrozen	ROC-AUC, ↑								RMSE, ↓					
		BBBP	BACE	ClinTox	Tox21	ToxCast	SIDER	HIV	MUV	ESol	FreeSolv	Lipo	QM7	QM8	QM9
CF-2U-r	0	0.671	0.710	0.589	0.715	0.603	0.646	0.678	0.771	1.24	3.52	0.94	88.2	0.0236	0.0143
CF-O-c	0	0.673	0.763	0.724	0.753	0.638	0.643	0.724	0.739	1.04	3.67	0.75	163.0	0.0223	0.0123
CF-U	0	0.626	0.785	0.596	0.736	0.617	0.623	0.603	0.747	1.06	3.66	0.87	93.4	0.0243	0.0140
CF-U-10I	0	0.657	0.754	0.586	0.732	0.620	0.603	0.738	0.744	1.34	3.39	0.91	105.4	0.0262	0.0163
CF-U-25I	0	0.646	0.752	0.651	0.709	0.612	0.593	0.702	0.746	1.48	3.87	0.88	98.7	0.0250	0.0152
CF-U-40I	0	0.660	0.732	0.518	0.694	0.606	0.596	0.685	0.780	1.51	4.24	0.91	103.5	0.0254	0.0170
CF-U-H	0	0.641	0.799	0.430	0.721	0.626	0.638	0.707	0.709	1.18	3.60	0.84	119.3	0.0248	0.0140
CF-U-r	0	0.658	0.760	0.530	0.752	0.645	0.650	0.707	0.714	1.13	3.36	0.80	97.3	0.0227	0.0129
CF-U-r-0.25	0	0.583	0.657	0.685	0.674	0.564	0.608	0.722	0.644	1.57	3.93	1.00	97.3	0.0296	0.0242
Contrast-U-r	0	0.635	0.751	0.566	0.688	0.590	0.589	0.695	0.684	1.35	4.02	0.98	117.2	0.0259	0.0171
O	0	0.655	0.775	0.630	0.693	0.639	0.583	0.723	0.742	1.17	2.95	1.03	109.5	0.0309	0.0257
O-c	0	0.659	0.787	0.680	0.710	0.629	0.613	0.758	0.695	1.18	3.01	0.95	85.6	0.0278	0.0200
O-Cpost-0.01	0	0.672	0.722	0.562	0.679		0.609	0.666	0.573	1.52	4.29	1.03	91.4	0.0270	0.0191
O-Cpost-0.05	0	0.613	0.783	0.768	0.682	0.574	0.548	0.576	0.603	1.71	4.30	1.06	96.9	0.0280	0.0221
U	0	0.633	0.778	0.753	0.717	0.652	0.583	0.732	0.738	1.16	2.59	0.92	89.3	0.0274	0.0208
U-no-flat	0	0.650	0.778	0.727	0.712	0.636	0.598	0.744	0.734	1.27	2.89	0.93	91.4	0.0264	0.0191
U-r	0		0.728	0.742	0.671	0.626		0.757	0.603	1.19	2.86	1.03	113.9	0.0297	0.0241
U-r-128	0	0.673	0.777	0.746	0.655		0.582	0.743	0.534	1.51	3.19	1.04	127.1	0.0315	0.0266

Continued on next page

Training	Unfrozen	ROC-AUC, ↑								RMSE, ↓					
		BBBP	BACE	ClinTox	Tox21	ToxCast	SIDER	HIV	MUV	ESol	FreeSolv	Lipo	QM7	QM8	QM9
U-r-256	0	0.654	0.771	0.626	0.677	0.612	0.602	0.734	0.641	1.29	2.75	1.02	143.8	0.0297	0.0255
U-r-256-conf	0	0.654	0.764	0.626	0.677	0.612	0.607	0.734	0.641	1.33	2.75	1.04	116.1	0.0287	0.0232
U-r-Cpost-0.01	0	0.656	0.730	0.438	0.692	0.601		0.660	0.626	1.24	3.37	0.99	81.1	0.0273	0.0207
U-r-Cpost-0.1	0	0.562	0.679	0.574	0.591	0.522		0.559	0.539	1.89	4.00	1.12	126.5	0.0318	0.0274
U-r-Cpost-0.5	0	0.544	0.590	0.500	0.603	0.540	0.540	0.518	0.640	1.99	4.39	1.11	126.4	0.0320	0.0283
CF-O-c	1	0.676	0.811	0.682	0.774	0.667	0.655	0.773	0.824	1.08	2.38	0.67	96.0	0.0188	0.0083
CF-U	1	0.659	0.802	0.650	0.773	0.662	0.638	0.763	0.793	0.95	2.92	0.75	101.9	0.0216	0.0100
CF-U-H	1	0.641	0.828	0.727	0.751	0.663	0.652	0.766	0.805	1.00	2.62	0.73	98.9	0.0217	0.0101
CF-U-r	1	0.683	0.805	0.656	0.775	0.682	0.590	0.757	0.775	1.05	2.16	0.74	88.1	0.0187	0.0082
CF-U-r-0.25	1	0.643	0.673	0.641	0.718	0.624	0.624	0.766	0.698	1.31	2.63	0.85	78.9	0.0188	0.0091
O	1	0.651	0.799	0.596	0.751	0.654	0.671	0.736	0.730	1.10	2.64	0.80	105.9	0.0232	0.0131
O-Cpost-0.01	1	0.700		0.672	0.685	0.618	0.615	0.668	0.609	1.34	3.94	0.97	79.8	0.0251	0.0137
O-Cpost-0.05	1	0.614	0.831	0.722	0.722	0.598	0.559	0.598	0.622		4.24	1.02	91.8	0.0251	0.0145
U	1	0.727	0.779	0.826	0.780	0.695	0.632	0.780	0.813	0.83	1.88	0.65	85.5	0.0181	0.0078
U-no-flat	1	0.723	0.788	0.889	0.789	0.683	0.635	0.790	0.787	0.86	2.12	0.65	62.8	0.0174	0.0072
U-r	1		0.747	0.708	0.755	0.665		0.766	0.704	1.00	2.90	0.77	113.3	0.0219	0.0118
U-r-Cpost-0.01	1	0.674	0.787	0.543	0.735	0.630	0.614	0.667		1.06	2.84		77.3	0.0235	0.0119
U-r-Cpost-0.1	1	0.650	0.786	0.584	0.650	0.594	0.577	0.568	0.497	1.20	3.55		85.8	0.0266	0.0163
U-r-Cpost-0.5	1	0.650	0.786	0.584	0.710	0.602	0.596	0.568	0.497	1.20	3.55		96.0	0.0254	0.0148
CF-O-c	3	0.702	0.816	0.670	0.779	0.668	0.654	0.784	0.791	1.02	2.21	0.67	72.7	0.0180	0.0067
CF-U	3	0.665	0.803	0.671	0.777	0.668	0.641	0.766	0.795	0.94	2.59	0.74	101.2	0.0192	0.0079
CF-U-H	3	0.639	0.829	0.723	0.760	0.672	0.647	0.762	0.790	0.96	2.40	0.70	90.5	0.0191	0.0079
CF-U-r	3	0.689	0.821	0.646	0.777	0.684	0.630	0.769	0.760	0.95	2.25	0.72	78.7	0.0177	0.0069

Continued on next page

Training	Unfrozen	ROC-AUC, ↑								RMSE, ↓					
		BBBP	BACE	ClinTox	Tox21	ToxCast	SIDER	HIV	MUV	ESol	FreeSolv	Lipo	QM7	QM8	QM9
CF-U-r-0.25	3	0.680	0.740	0.643	0.724	0.646	0.647	0.762	0.675	1.13	2.20	0.77	69.7	0.0176	0.0068
O	3	0.728	0.837	0.630	0.751	0.654	0.671	0.745	0.797	1.11	2.25	0.71	82.4	0.0200	0.0087
O-Cpost-0.01	3		0.752	0.681	0.713	0.622	0.604	0.687	0.663	1.20	3.79	0.93	64.4	0.0225	0.0098
O-Cpost-0.05	3	0.653	0.825	0.775	0.720	0.609	0.566	0.664	0.679	1.23	4.13	0.97	68.5	0.0263	0.0102
U	3	0.717	0.785	0.842	0.782	0.692	0.636	0.791	0.807	0.88	1.63	0.61	74.3	0.0165	0.0060
U-no-flat	3	0.719	0.791	0.902	0.785	0.690	0.633	0.794	0.798	0.91	1.84	0.62	56.7	0.0168	0.0057
U-r	3	0.698	0.832	0.749	0.779	0.676	0.638	0.778	0.795	0.88	2.30	0.69	90.6	0.0185	0.0080
U-r-Cpost-0.01	3	0.696	0.785	0.579		0.629	0.624	0.693		0.98	2.42		80.8	0.0205	0.0088
U-r-Cpost-0.1	3	0.671	0.785		0.683	0.608	0.584	0.613	0.582	1.10	3.09	1.02	78.0	0.0234	0.0105
U-r-Cpost-0.5	3	0.661	0.661	0.503	0.729	0.605	0.595	0.578	0.602	1.08	3.74	0.93	93.6	0.0231	0.0100
CF-O-c	15	0.650	0.807	0.716	0.778	0.691	0.649	0.782	0.772	0.94	2.15	0.64	60.3	0.0159	0.0054
CF-U-r	15	0.699	0.815	0.745				0.789	0.819	0.88	1.89	0.66	59.4	0.0168	
O	15	0.680	0.859	0.704	0.785	0.680	0.618	0.785	0.711	0.87	1.95	0.66	52.0	0.0159	0.0054
Random-w	15	0.639	0.823	0.592	0.746	0.648	0.617	0.761	0.619	0.98	2.83	0.77	97.3	0.0178	0.0063
U	15	0.723	0.811	0.861	0.786	0.684	0.646	0.771	0.789	0.81	1.93	0.62	59.2	0.0161	0.0052
U-no-flat	15	0.694	0.796	0.883	0.797	0.688	0.648	0.800	0.883	0.74	3.08	0.38	64.2	0.0160	0.0053
U-r	15	0.725	0.825	0.854	0.782	0.677	0.637	0.780	0.792	0.85	2.15	0.64	58.1	0.0166	0.0060

View Article Online
DOI: 10.1039/D6DD00096G

Table S5: Performance on quantum-chemical and physicochemical regression benchmarks when reducing the number of conformations during finetuning from 10 to 1, 2, or 5. Values denote root-mean-square deviation (RMSD), with lower values indicating better agreement with reference data. Standard deviation in parentheses.

Model	Unfrozen	num. conf.	ESOL	Lipo	QM7	QM8	QM9
CF-O-c	0	1	1.14(3)	0.746(4)	105.9(136)	0.0218(1)	0.01160(7)
CF-O-c	0	2	1.14(3)	0.746(4)	105.9(136)	0.0218(1)	0.01160(7)
CF-O-c	0	5	1.12(2)	0.743(3)	98.6 (115)	0.0218(1)	0.01158(11)
U	0	1	1.15(2)	0.925(8)	80.8 (35)	0.0262(4)	0.01776(11)
U	0	2	1.15(2)	0.925(8)	80.8 (35)	0.0262(4)	0.01776(11)
U	0	5	1.13(2)	0.919(6)	80.8 (39)	0.0262(4)	0.01776(5)
U	15	1	0.84(4)	0.609(13)	56.2 (10)	0.0159(5)	0.00524(5)
U	15	2	0.84(4)	0.609(13)	56.2 (10)	0.0159(5)	0.00524(5)
U	15	5	0.85(3)	0.606(11)	57.8 (28)	0.0159(4)	0.00522(4)

Table S6: Performance on biological activity classification benchmarks when reducing the number of conformations during finetuning from 10 to 1, 2, or 5. Values report the mean ROC–AUC over multiple random splits (standard deviation in parentheses), with higher values indicating better performance.

Model	Unfrozen	num. conf.	BBBP	BACE	ClinTox	Tox21	ToxCast	SIDER	HIV	MUV
CF-O-c	0	1	0.662(9)	0.744(22)	0.711(9)	0.755(1)	0.638(2)	0.636(2)	0.749(3)	0.778(4)
CF-O-c	0	2	0.662(9)	0.744(22)	0.712(9)	0.755(1)	0.638(2)	0.636(1)	0.749(3)	0.778(4)
CF-O-c	0	5	0.666(6)	0.744(13)	0.712(10)	0.755(1)	0.638(2)	0.636(2)	0.754(3)	0.776(5)
U	0	1	0.639(5)	0.775(4)	0.786(8)	0.705(1)	0.652(5)	0.608(3)	0.735(4)	0.748(10)
U	0	2	0.639(5)	0.775(4)	0.786(7)	0.705(1)	0.652(5)	0.608(3)	0.735(4)	0.748(10)
U	0	5	0.640(4)	0.777(4)	0.788(8)	0.708(1)	0.652(3)	0.609(5)	0.735(3)	0.763(10)
U	15	1	0.705(17)	0.812(12)	0.819(31)	0.777(10)	0.677(10)	0.629(13)	0.787(7)	0.792(29)
U	15	2	0.705(18)	0.812(12)	0.865(36)	0.777(10)	0.677(10)	0.636(11)	0.781(12)	0.772(61)
U	15	5	0.700(23)	0.799(53)	0.878(31)	0.786(7)	0.688(6)	0.643(23)	0.790(10)	0.799(26)

F More visual examples of model performance

F.1 Isomers

View Article Online
DOI: 10.1039/D6DD00096G

Figure S1: A pair of isomers (distinct molecules) having similarity of 0.99 in the Uni-Mol embedding space and 0.04 in ConforFormer–OMol

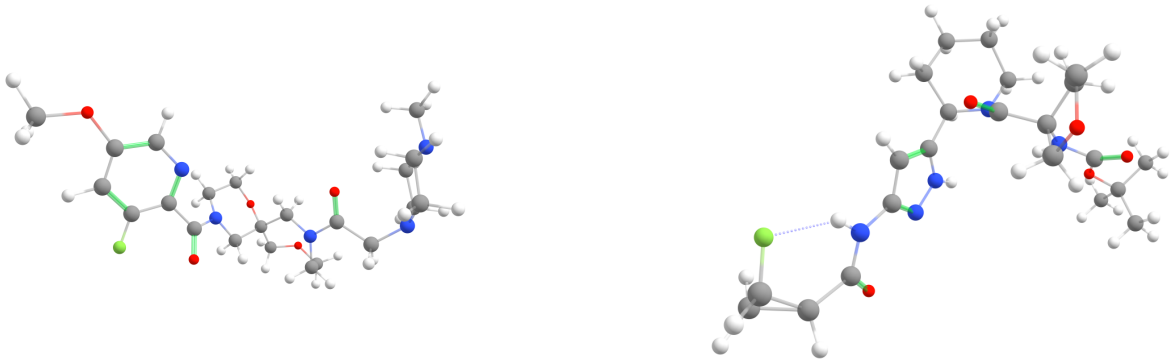

Figure S2: A pair of isomers (distinct molecules) having similarity of 0.99 in the Uni-Mol embedding space and 0.20 in ConforFormer–OMol

Figure S3: A pair of isomers (distinct molecules) having similarity of 0.99 in the Uni-Mol embedding space and 0.70 in ConforFormer–OMol

Figure S4: A pair of isomers (distinct molecules) having similarity of 0.98 in the Uni-Mol embedding space and 0.40 in ConforFormer–OMol

View Article Online
DOI: 10.1039/D6DD00096G

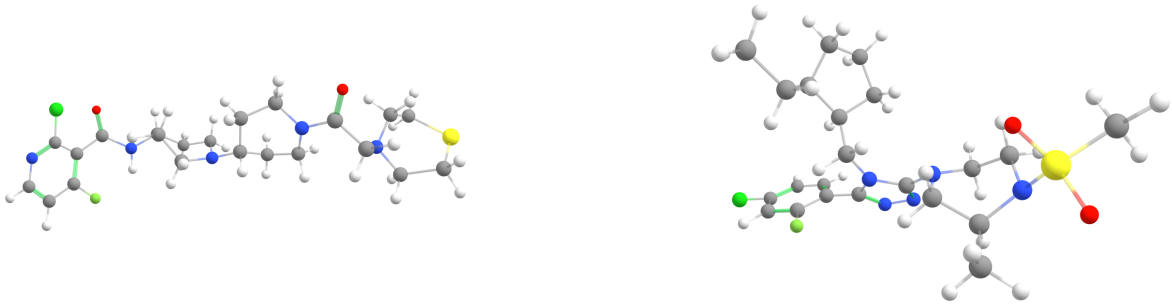

Figure S5: A pair of isomers (distinct molecules) having similarity of 0.98 in the Uni-Mol embedding space and 0.20 in ConforFormer–OMol

Figure S6: A pair of isomers (distinct molecules) having similarity of 0.93 in the Uni-Mol embedding space and 0.14 in ConforFormer–OMol

View Article Online
DOI: 10.1039/D6DD00096G

F.2 Conformers

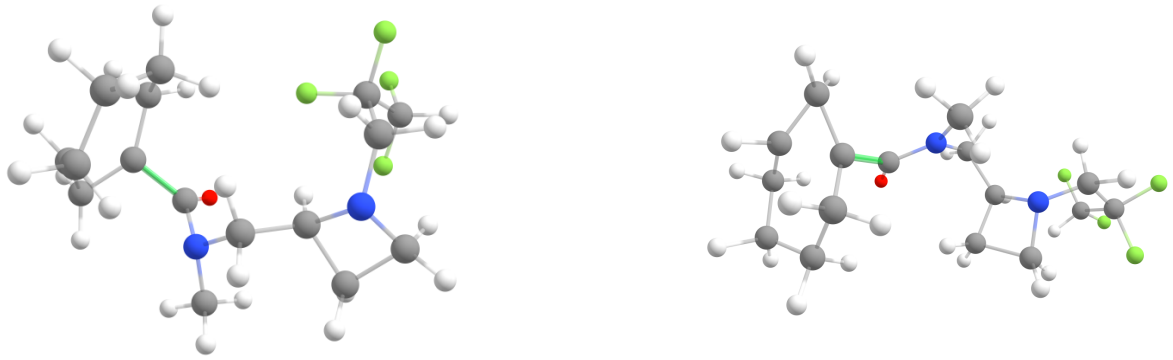

Figure S7: A pair of conformers of the same molecule having similarity of 0.95 (very low, below the 2nd percentile of all pairs) in the Uni-Mol embedding space and 0.99 in ConforFormer–OMol

Figure S8: A pair of conformers of the same molecule having similarity of 0.95 (very low, below the 2d percentile ofd all pairs) in the Uni-Mol embedding space and 0.99 in ConforFormer–OMol

View Article Online
DOI: 10.1039/D6DD00096G

Figure S9: A pair of conformers of the same molecule having similarity of 0.96 in the Uni-Mol embedding space and 0.64 in ConforFormer–OMol. *Note* the distorted ring in the right image is an indication of improperly generated data and is technically not a conformer of the image left. This chemically significant distortion in the structure is not detected by Uni-Mol.

Figure S10: A pair of conformers of the same molecule having similarity of 0.94 in the Uni-Mol embedding space and 0.99 in ConforFormer–OMol.

View Article Online
DOI: 10.1039/D6DD00096G

References

[1] 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*, 2022, https://chemrxiv.org/doi/full/10.26434/chemrxiv-2022-jjm0j-v4, ChemRxiv.

[2] D. S. Levine, M. Shuaibi, E. W. C. Spotte-Smith, M. G. Taylor, M. R. Hasyim, K. Michel, I. Batatia, G. Csányi, M. Dzamba, P. Eastman, N. C. Frey, X. Fu, V. Gharakhanyan, A. S. Krishnapriyan, J. A. Rackers, S. Raja, A. Rizvi, A. S. Rosen, Z. Ulissi, S. Vargas, C. L. Zitnick, S. M. Blau and B. M. Wood, *The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models*, 2025, http://arxiv.org/abs/2505.08762, arXiv:2505.08762 [physics].

[3] P. Walters, *We need better benchmarks for machine learning in drug discovery*, Practical Cheminformatics Blog, 2023, https://practicalcheminformatics.blogspot.com/2023/08/we-need-better-benchmarks-for-machine.html, Accessed: 2026-04-29.

[4] J. J. Irwin, K. G. Tang, J. Young, C. Dandarchuluun, B. R. Wong, M. Khurelbaatar, Y. S. Moroz, J. Mayfield and R. A. Sayle, *Journal of Chemical Information and Modeling*, 2020, **60**, 6065–6073.

[5] S. Riniker and G. A. Landrum, *Journal of Chemical Information and Modeling*, 2015, **55**, 2562–2574.

[6] T. A. Halgren, *Journal of Computational Chemistry*, 1996, **17**, 490–519.

[7] M. Klein, I. Rudenko, E. Pidko and I. Bushmarinov, *PharmaIsomer*, 2026, https://zenodo.org/records/18739668.