Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d6dd00096g

ConforFormer: representation for molecules through understanding of conformers

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

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.

Received 26th February 2026
Accepted 14th May 2026

DOI: 10.1039/d6dd00096g

rsc.li/digitaldiscovery

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

[a]Department of Chemical Engineering, TU Delft, Delft, Netherlands. E-mail: mpklein@tudelft.nl; E.A.Pidko@tudelft.nl
[b]Avride Inc., Austin, TX, USA. E-mail: irina.vl.rudenko@gmail.com
[c]Perplexity AI, Belgrade, Serbia. E-mail: ivan.bushmarinov@perplexity.ai
† Equal contribution.

behavior. In reality, each compound exists as an ensemble of 3D structures (conformers) whose distribution determines such properties as 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 19 M unique SMILES strings and around 209 M 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 (ref. 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 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 pre-training. 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 (Fig. 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 connectivity rather efficiently. However, they do not uniquely specify the molecular 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

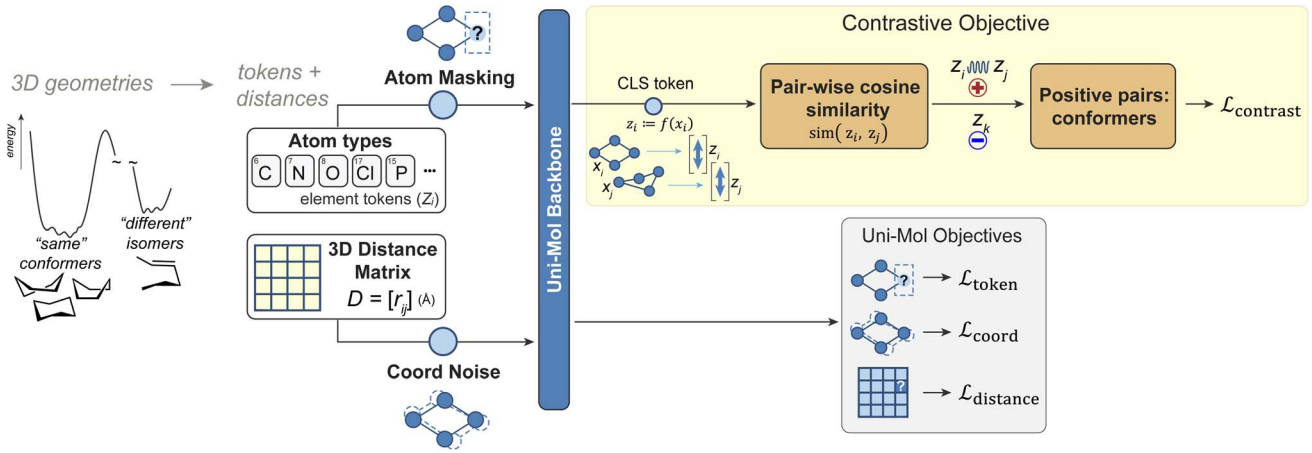

Fig. 1 Schematic illustration of the ConforFormer framework: model architecture with pretraining objectives.

and supramolecular compounds, where bonding patterns may be ambiguous, fluxional behavior is common, and chemically relevant distinctions often arise from "subtle" geometric rearrangements. 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. Fig. 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 $\mathcal{M}$ denote the set of molecules in the training corpus and let $\mathscr{C}(\Updownarrow)$ denote the set of conformers associated with molecule $m \in \mathcal{M}$. 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 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 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

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

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 $\mathbf{z}_i := f(x_i)$.

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:

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

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

$$\mathscr{L}_{\mathrm{contrast}} := \sum_{(i,j) \in \mathscr{P}} \ell_{i,j}.$$

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 SI, further SI (Section E).

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

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

Here $\mathscr{L}_{\mathrm{token}}$ corresponds to the loss for masked token prediction, $\mathscr{L}_{\mathrm{coord}}$ is the loss associated with the coordinates denoising task, and $\mathscr{L}_{\mathrm{distance}}$ is the loss associated with the masked distance prediction. These objectives and their original batching protocol were introduced in ref. 4 and further detailed in ref. 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 Fig. 2. We consider multiple training configurations, including the original Uni-Mol pre-trained model and variants trained with the conformer-alignment contrastive objective introduced in 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.

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

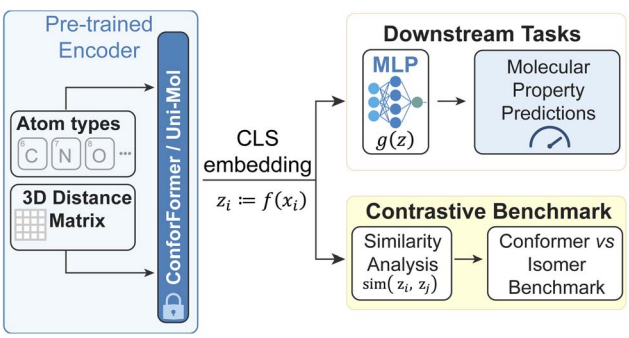

Fig. 2 Schematic illustration of the ConforFormer framework: fine-tuning and evaluation scheme.

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 (Tables S5 and S6). This, along with other ablation studies, can be found in section E of the SI.

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 × 256 × 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. Fig. 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 SI).

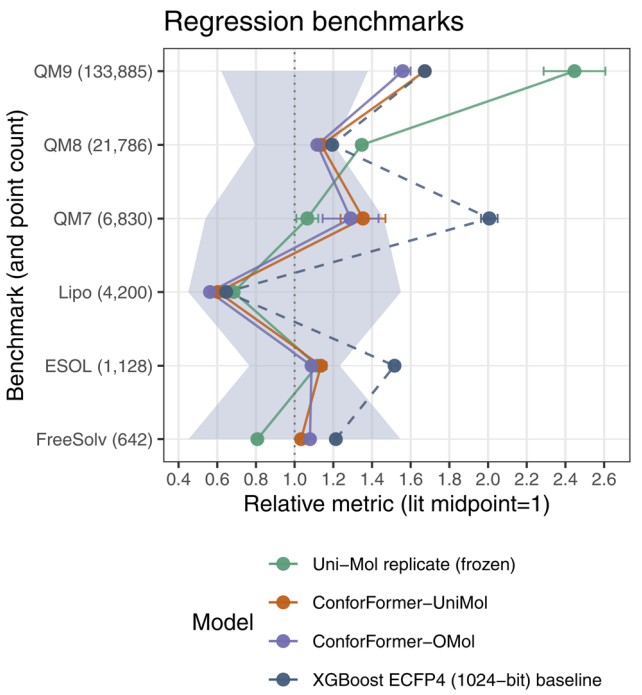

Fig. 3 Regression MoleculeNet benchmark results (rescaled error) for the XGBoost ECFP4 (1024-bit) baseline and embedding-based models; less is better (↓). The shaded area indicates the literature range for post-2019 models. The task-specific errors are rescaled linearly so that the midpoint of the literature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

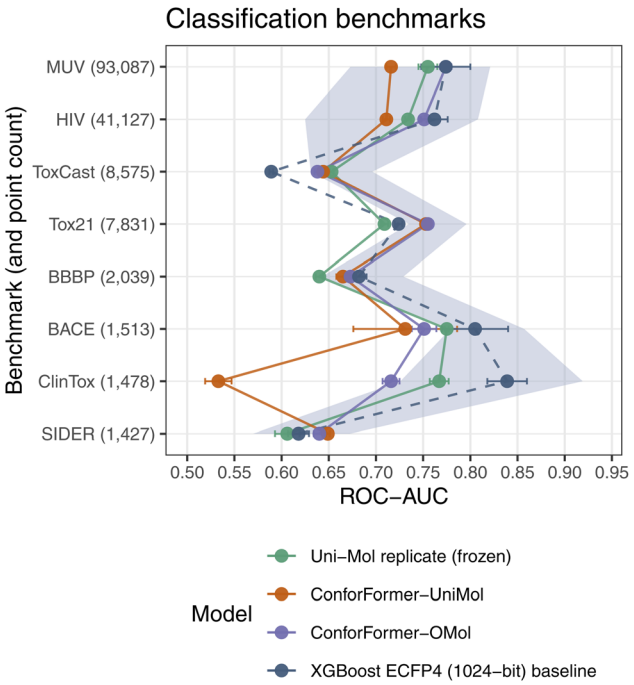

Fig. 4 Classification MoleculeNet benchmark results (ROC-AUC) for the XGBoost ECFP4 (1024-bit) baseline and embedding-based models; more is better (↑). The shaded area indicates the literature range for post-2019 models.

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 (denoted as the XGBoost ECFP4 (1024-bit) baseline). We additionally screened RDKit Morgan ECFP4 and ECFP6 fingerprints folded into 2048 and 16 384 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, Fig. 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.

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 8.2 M 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). In contrast, adding the conformer-alignment objective yields consistently better embeddings. ConforFormer-OMol performs best on 4 out of 6 quantum-chemical regression benchmarks (Fig. 3) and 5 out of 8 classification benchmarks (Fig. 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.

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 (ref. 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 SI 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 (Fig. 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,

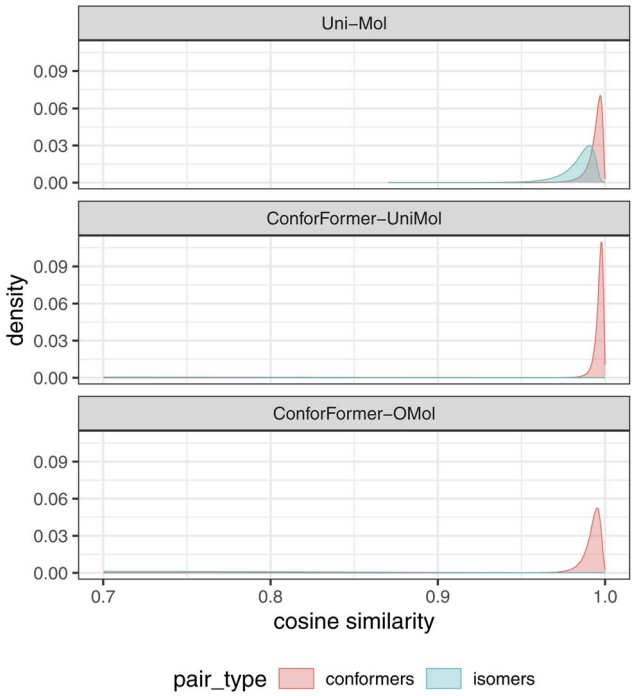

Fig. 5 Distribution of cosine similarities between CLS token values extracted from Uni-Mol, ConforFormer-UniMol and ConforFormer-OMol, as measured on the Pharmaisomer benchmark.

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

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

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

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

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:

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

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

The precision/recall curves constructed by sweeping over $\theta \in [0, 1]$ can be found on Fig. 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 (Fig. 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 Fig. 7 for an example of conformers with very dissimilar shape and Fig. 8 for a pair of isomers with an overall similar one. Both have the same similarity of 0.93 in the Uni-

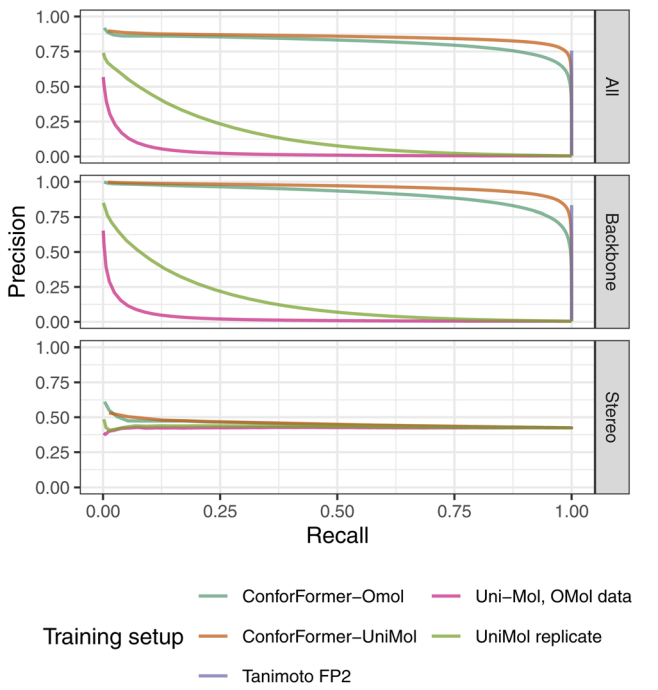

Fig. 6 Precision and recall curves for different frozen representations on Pharmaisomer benchmarks.

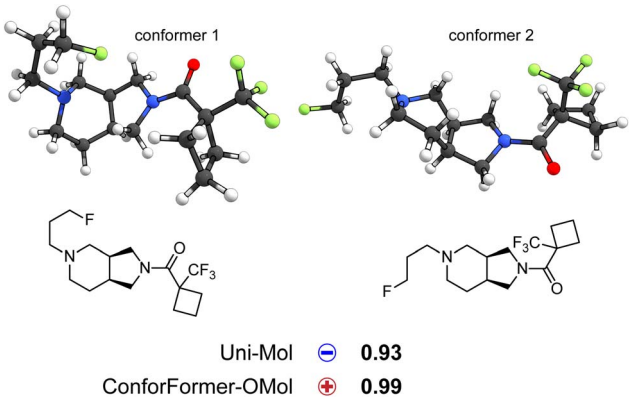

Fig. 7 A pair of conformers of the same molecule having similarity of 0.93 in the Uni-Mol embedding space and 0.99 in ConforFormer-OMol (indicating they belong to the same molecule).

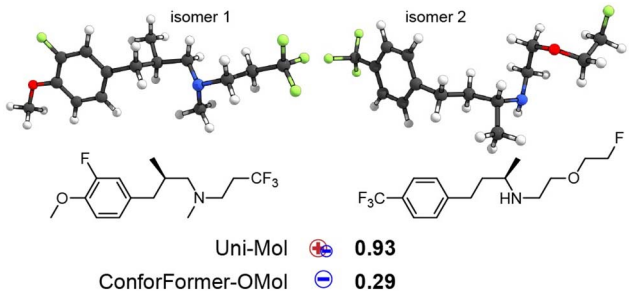

Fig. 8 A pair of isomers (distinct molecules) with similarity of 0.93 in the Uni-Mol embedding space but 0.29 in ConforFormer-OMol (indicating they are distinct).

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 isomer 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, physicochemical 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 finger-print 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.

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 Fig. 3 and 4, and plot Fig. 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 supplementary information (SI). Supplementary information: extended details regarding the PharmIsomer dataset, raw benchmark scores, and additional visual examples of model performance. See DOI: https://doi.org/10.1039/d6dd00096g.

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, *arXiv*, 2025, preprint, arXiv:2509.13805, DOI: 10.48550/arXiv.2509.13805, https://arxiv.org/abs/2509.13805.
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, *ChemRxiv*, 2022, preprint, https://chemrxiv.org/doi/full/10.26434/chemrxiv-2022-jjm0j-v4.
5 W. Ahmad, E. Simon, S. Chithrananda, G. Grand and B. Ramsundar, ChemBERTa-2: Towards Chemical Foundation Models, *arXiv*, 2022, preprint, arXiv:2209.01712, DOI: 10.48550/arXiv.2209.01712, https://arxiv.org/abs/2209.01712.
6 Y. Wang, J. Wang, Z. Cao and A. Barati Farimani, *Nat. Mach. Intell.*, 2022, **4**, 279–287.
7 S. Chithrananda, G. Grand and B. Ramsundar, ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction, *arXiv*, 2020, preprint, arXiv:2010.09885, DOI: 10.48550/arXiv.2010.09885, https://arxiv.org/abs/2010.09885.
8 X. Fang, L. Liu, J. Lei, D. He, S. Zhang, J. Zhou, F. Wang, H. Wu and H. Wang, *Nat. Mach. Intell.*, 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, *J. Chem. Inf. Model.*, 2024, **64**, 3610–3620.
11 R. Laplaza, M. D. Wodrich and C. Corminboeuf, *J. Phys. Chem. Lett.*, 2024, **15**, 7363–7370.
12 S. Finta, A. V. Kalikadien and E. A. Pidko, *J. Chem. Theory Comput.*, 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, in *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, *arXiv*, 2024, preprint, arXiv:2212.03533, DOI: 10.48550/arXiv.2212.03533, https://arxiv.org/abs/2212.03533.
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, *J. Chem. Inf. Model.*, 2019, **59**, 3370–3388.
17 Y. Rong, Y. Bian, T. Xu, W. Xie, Y. WEI, W. Huang and J. Huang, *Adv. Neural Inf. Process. Syst.*, 2020, 12559–12571.
18 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
19 J. Li and X. Jiang, *Wireless Commun. Mobile Comput.*, 2021, **2021**, 7181815.
20 C. Liu, Y. Sun, R. Davis, S. T. Cardona and P. Hu, *J. Cheminf.*, 2023, **15**, 1–14.
21 S. Lu, Z. Gao, D. He, L. Zhang and G. Ke, *Nat. Commun.*, 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 Trans. Pattern Anal. Mach. Intell.*, 2020, **42**, 824–836.
29 A. Martin Pendas and E. Francisco, *Nat. Commun.*, 2022, **13**, 3327.
30 M. Weisberg, *Philos. Sci.*, 2008, **75**, 932–946.
31 S. Shaik, *J. Comput. Chem.*, 2007, **28**, 51–61.
32 G. Frenking and A. Krapp, *J. Comput. Chem.*, 2007, **28**, 15–24.
33 S. Alvarez, R. Hoffmann and C. Mealli, *Chem.–Eur. J.*, 2009, **15**, 8358–8373.
34 R. Hoffmann and P. Laszlo, *Angew Chem. Int. Ed. Engl.*, 1991, **30**, 1–16.
35 G. N. Lewis, *J. Am. Chem. Soc.*, 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, *J. Phys. Chem. A*, 1998, **102**, 7314–7323.
38 M. Brookhart, M. L. H. Green and G. Parkin, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 6908–6914.
39 S. Bougueroua, A. A. Kolganov, C. Helain, C. Zens, D. Barth, E. A. Pidko and M.-P. Gaigeot, *Phys. Chem. Chem. Phys.*, 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, *Chem. Sci.*, 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, *arXiv*, 2025, preprint, arXiv:2505.08762, DOI: 10.48550/arXiv.2505.08762, https://arxiv.org/abs/2505.08762.
43 S. Riniker and G. A. Landrum, *J. Chem. Inf. Model.*, 2015, **55**, 2562–2574.
44 T. A. Halgren, *J. Comput. Chem.*, 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, *J. Chem. Inf. Model.*, 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, *J. Chem. Inf. Model.*, 2020, **60**, 6065–6073.
48 M. Klein, I. Rudenko, E. Pidko and I. Bushmarinov, *PharmaIsomer*, 2026, https://zenodo.org/records/18739668.