Graph Learning on Ensembles of Cyclic Peptides: An Investigation of Molecular Ensemble Modeling

Aaron Feller [1 2] **Kris Deibler** [2] **Maxim Secor** [2]

Abstract

Molecular property prediction from structure often uses a single representative conformation, even though many molecules exist as conformational ensembles in solution. We introduce **EnsembleEGNN**, a molecular ensemble foundation model that encodes an ensemble by first encoding each conformer with shared Equivariant Graph Neural Network (EGNN) layers, then pooling the resulting conformer representations with a Set Attention Block. We pretrain the model on CREMP, a cyclic peptide ensemble dataset, using a multi-task self-supervised objective combining masked token recovery, noisy-coordinate reconstruction, and pairwise distance reconstruction. On the CREMP-CycPeptMPDB dataset, training EnsembleEGNN from scratch fails entirely ($R^2 = 0.005$). However, the pretrained model reaches $R^2 = 0.477$ and Pearson $r = 0.699$, outperforming the sequence-only BERT baseline ($R^2 = 0.439$, Pearson $r = 0.667$). When EnsembleEGNN is co-trained end-to-end with the BERT sequence encoder, the hybrid model improves further to $R^2 = 0.538$ and Pearson $r = 0.737$. These results demonstrate that encoding conformational ensembles into a single thermodynamically informed embedding improves cyclic-peptide property prediction.

1. Introduction

Graph foundation models (GFMs) aim to pretrain once and transfer across node, edge, and graph-level tasks with minimal fine-tuning (Liu et al., 2025). In the molecular domain, GFMs have shown promise for small molecules and proteins (Méndez-Lucio et al., 2024; Hsu et al., 2022), yet a fundamental representational mismatch persists. In solution, many molecules do not occupy a single static structure but are understood as Boltzmann-weighted ensembles of conformers (Chandler, 1987; Tuckerman, 2023). This mismatch is particularly detrimental for cyclic peptides (Hui et al., 2025; Bayaraa et al., 2026; Damjanovic et al., 2021). These scaffolds are inherently flexible, and their thermodynamic ensembles ultimately dictate critical properties like membrane permeability, protease resistance, and target binding (Yudin, 2015; Calvo-Barreiro et al., 2025).

The conformer aggregation problem. Structure-based methods typically process single conformers (Otero et al., 2025; Liu et al., 2026) or apply naive pooling (Zhu et al., 2024). For flexible cyclic peptides, this discards critical thermodynamic context regarding how atoms move across accessible states. The core challenge remains: given C conformers of N atoms with Boltzmann weights (w_c), how can a model efficiently aggregate this geometric information into a single embedding? With conformer-generation pipelines improving (Aranganathan et al., 2025), models that can encode at the ensemble level will become increasingly useful.

Our approach. EnsembleEGNN encodes each conformer separately with a shared equivariant graph neural network (EGNN) (Satorras et al., 2021) and then combines those conformer-level representations into a single ensemble embedding. In this view, the same ordered set of atoms is observed under multiple 3D conformations, allowing the model to explicitly capture the inherent flexibility of the molecule. By pretraining on Boltzmann-weighted atom embeddings across C conformer states, the model is forced to aggregate geometric context across the ensemble (Figure 1). The conformers are then pooled using a set attention block (Lee et al., 2019), allowing for the attention mechanism to learn prioritization of conformers that are important on a per-task basis. This design maintains low complexity while still learning from the full thermodynamic distribution rather than from a single static structure.

Contributions. We present three main contributions: **(1)** an ensemble geometric encoder that learns a single embedding from multiple conformers using set attention pooling;

[1]Interdisciplinary Life Sciences, University of Texas at Austin, Austin, TX. [2]Molecular AI, Novo Nordisk, Lexington, MA. Correspondence to: Aaron Feller <aaron.feller@utexas.edu>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

(2) a multi-task pretraining objective with masked token recovery, noisy-coordinate reconstruction, and distance reconstruction that leverages Boltzmann-informed atom pooling, evaluated with CREMP (Grambow et al., 2024); and **(3)** cross-validation on CREMP-CycPeptMPDB to predict membrane permeability, demonstrating that a) pretraining is necessary and b) conformer-aware geometric modeling improves over a sequence-encoder baseline.

2. Method

2.1. Data Representation for Conformer Ensembles

To bridge 1D token identity and 3D geometry, we represent each peptide as a set of atom tokens, with multiple conformations that reuse the same nodes with different position coordinates.

Each molecule contains three pieces of information: node (atom) identities (a), 3D conformer coordinates (X), and conformer weights (w). In the all-atom experiments used in this paper, the node tokens are atom-level tokens and the coordinates are all-atom conformer coordinates. The goal of the model is to a) formulate pretraining tasks for ensembles, and b) compress an ensemble into a single embedding that can be used for downstream prediction.

To construct these inputs from CREMP (Grambow et al., 2024), we extract all available conformers, rank them by their Boltzmann weights, retain the top $C = 5$ states, and renormalize those weights to sum to one before serialization. Node (atom) tokens (a) correspond directly to atom identities, giving a coordinate tensor of shape $C \times a \times 3$ together with per-conformer weights.

2.2. Architecture

Conformer-wise encoding. Each conformer is processed through L EGNN layers (Satorras et al., 2021). The model updates both the node features and coordinates by passing messages exclusively over the local k-nearest-neighbor graph within that specific conformer, shown here:

$$\mathbf{h}_i^{(c,\ell+1)}, \mathbf{x}_i^{(c,\ell+1)} = \mathrm{EGNNLayer}\Big(\mathbf{h}_i^{(c,\ell)},\, \mathbf{x}_i^{(c,\ell)},\, \mathcal{N}_k(i,c)\Big) \tag{1}$$

where the initial node features are $\mathbf{h}_i^{(c,0)} = \mathrm{Embed}(a_i)$, the initial coordinates are $\mathbf{x}_i^{(c,0)} = \mathbf{X}_i^{(c)}$, and $\mathcal{N}_k(i,c)$ denotes the k nearest spatial neighbors of node i in conformer c. This restriction to a specified set of neighbors allows for the model to extend to any molecule size and enables straightforward pretraining.

The choice of EGNN layers ensures that the per-conformer message passing is strictly equivariant to rotations and translations in 3D space. This property allows the model to learn robust geometric representations without relying on expensive coordinate frame alignments or spatial data augmentation.

Conformer fusion. Let $\tilde{\mathbf{h}}^{(c)} \in \mathbb{R}^{N\times d}$ denote the final post-EGNN node (i.e. atom) features for conformer c. To pool these separate structures into a unified representation, each conformer receives an attention weight driven by two signals: a learned score from the mean ensemble embedding and a thermodynamic prior from the Boltzmann weight. We compute these attention weights as:

$$\alpha_c = \frac{\exp\big(f_\phi(\bar{\mathbf{h}}^{(c)}) + \log w_c\big)}{\sum_{c'} \exp\big(f_\phi(\bar{\mathbf{h}}^{(c')}) + \log w_{c'}\big)} \tag{2}$$

where $\bar{\mathbf{h}}^{(c)} = \frac{1}{N}\sum_i \tilde{\mathbf{h}}_i^{(c)}$ is the mean node feature for conformer c and f_ϕ is a 2-layer MLP.

Using these attention weights, we compute two distinct ensemble embeddings:

$$\mathbf{h}_i^{\mathrm{ens}} = \sum_{c=1}^{C} \alpha_c\, \tilde{\mathbf{h}}_i^{(c)}, \qquad \mathbf{z}^{\mathrm{ens}} = \mathrm{SAB}\Big(\{\bar{\mathbf{h}}^{(c)}\}_c\Big) \tag{3}$$

The first term, $\mathbf{h}_i^{\mathrm{ens}}$, is a per-atom embedding obtained by taking the attention-weighted average of each atom across all conformers. The second term, $\mathbf{z}^{\mathrm{ens}} \in \mathbb{R}^d$, is a global graph-level embedding generated by passing the conformer embeddings through a Set Attention Block (SAB) (Lee et al., 2019).

Utilizing SAB with inducing points reduces the conformer fusion complexity from $O(N^2)$ with quadratic attention to $O(nm)$, enabling efficient scaling across large conformational ensembles.

2.3. pretraining Objective

During pretraining, we corrupt both the node identities and the coordinates. We randomly select 15% of valid node positions for masked-token supervision. Following a BERT-style corruption scheme, 80% of those selected positions are replaced with a mask token, 10% are replaced with a random token, and 10% are left unchanged.

We also add Gaussian noise to valid node and conformer positions, producing a noisy coordinate tensor that the model must refine back toward the clean structure.

The composite pretraining loss is:

$$\mathcal{L} = 0.3(\mathcal{L}_{\mathrm{tok}}) + 0.5(\mathcal{L}_{\mathrm{coord}}) + 0.2(\mathcal{L}_{\mathrm{dist}}) \tag{4}$$

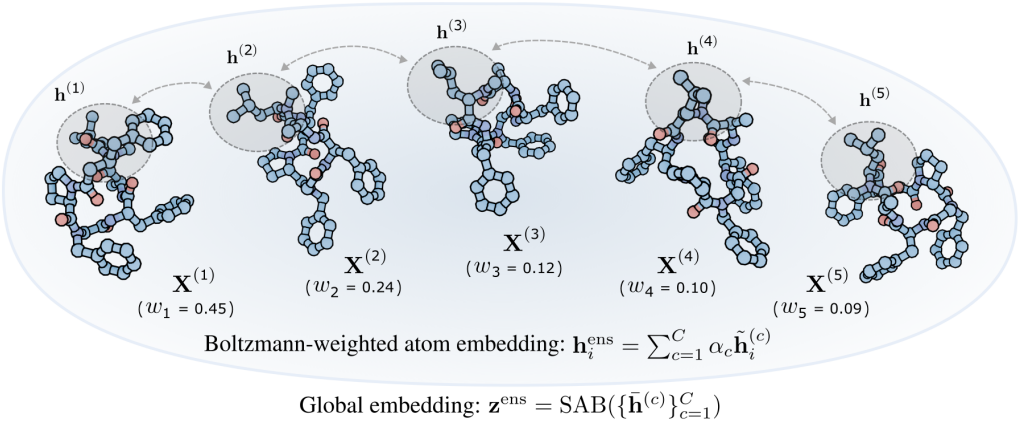

Figure 1. Information flow and conformer aggregation in EnsembleEGNN. The model processes a thermodynamic ensemble of C distinct spatial states, each defined by 3D coordinates $\mathbf{X}^{(c)}$ and a Boltzmann weight w_c. The shaded regions highlight the local k-nearest-neighbor graph of a single shared atom (node i) across the ensemble. After per-conformer EGNN layers update the local geometry, the finalized node features ($\tilde{\mathbf{h}}_i^{(c)}$) are aggregated via a Boltzmann-informed attention mechanism (α_c) to produce a single unified atom embedding ($\mathbf{h}_i^{\text{ens}}$). Finally, a Set Attention Block (SAB) reasons over the entire collection of conformers to generate a macroscopic global embedding ($\mathbf{z}^{\text{ens}}$) for downstream property prediction.

Algorithm 1 EnsembleEGNN Conformer-Aware Fusion Forward Pass

Input: Peptide with N atoms having:
atom types $\{a_i\}_{i=1}^{N}$,
per-conformer atom coordinates $\{\mathbf{X}^{(c)}\}_{c=1}^{C}$,
and Boltzmann weights $\{w_c\}_{c=1}^{C}$
Initialize atom features $\mathbf{h}_i^{(c,0)} = \text{Embed}(a_i)$
for $\ell = 0$ **to** $L-1$ **do**
  Update $\mathbf{h}_i^{(c,\ell+1)}, \mathbf{x}_i^{(c,\ell+1)}$ via EGNN
end for
Let $\tilde{\mathbf{h}}^{(c)} = \mathbf{h}^{(c,L)}$ denote final atom features
for $c = 1$ **to** C **do**
  Conformer token: $\bar{\mathbf{h}}^{(c)} = \frac{1}{N}\sum_{i=1}^{N} \tilde{\mathbf{h}}_i^{(c)}$
  Attention logit: $u_c = f_\phi(\bar{\mathbf{h}}^{(c)}) + \log w_c$
end for
Attention weights: $\alpha_c = \frac{\exp(u_c)}{\sum_{c'} \exp(u_{c'})}$
Ensemble atoms: $\mathbf{h}_i^{\text{ens}} = \sum_{c=1}^{C} \alpha_c \tilde{\mathbf{h}}_i^{(c)}$
Global embedding: $\mathbf{z}^{\text{ens}} = \text{SAB}(\{\bar{\mathbf{h}}^{(c)}\}_{c=1}^{C})$
Output: $\mathbf{h}^{\text{ens}}$ and $\mathbf{z}^{\text{ens}}$

The masked-token term ($\mathcal{L}_{\text{tok}}$) is a cross-entropy loss computed at the masked valid nodes, training the per-atom embeddings to recover the original token identities from corrupted inputs. The coordinate term ($\mathcal{L}_{\text{coord}}$) is an MSE loss between the final EGNN-refined coordinates and the clean conformer coordinates, training the geometric backbone to denoise each conformer.

The distance term ($\mathcal{L}_{\text{dist}}$) compares pairwise distances computed from projected per-atom embeddings with pairwise distances computed from a Boltzmann-weighted mean coordinate set, acting as a coarse regularizer on the per-atom embeddings. Importantly, this target is based on distances in the Boltzmann-weighted mean structure, not on the average of pairwise distances across conformers.

3. Experiments

3.1. Pretraining optimization and convergence monitoring.

During self-supervised pretraining, we log train and validation total loss throughout optimization and track convergence on a train-scaled global-step axis. The model was trained with all-atom representations and 12 nearest neighbors. These learning dynamics showed a smoothed train-loss trajectory and validation loss decreasing with training loss (Figure 2).

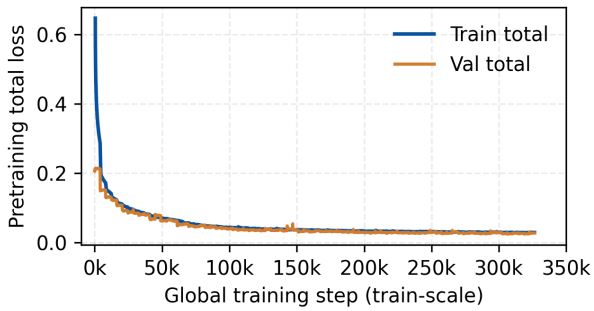

Figure 2. EnsembleEGNN pretraining loss dynamics. Trajectories of the training loss (50-step moving average) and validation loss plotted against global training steps, illustrating stable model convergence on the CREMP dataset.

3.2. Dataset and Protocol

For quantitative evaluation, we used CREMP-CycPeptMPDB (Grambow et al., 2024), which contains cyclic peptide conformers paired with artificial membrane permeation scores, yielding $n = 3003$ labeled measurements before boundary filtering. We removed exact boundary targets at $\log(P_{\text{app}})$=-10.0, -8.0, and -4.0, resulting in a final evaluation set of $n = 2979$. We ran 5-fold cross-validation with fixed splits and report fold-wise and aggregate metrics over all held-out examples. In each run, 3 folds were used for training, 1 fold for validation, and 1 fold is kept as an unseen test fold used only for post-training prediction. EnsembleEGNN uses all-atom representations (i.e., atom-level node tokens and conformer coordinates) with 12 nearest neighbors, and BERT uses 1D sequence. The target is $\log(P_{\text{app}})$, and we evaluated MAE, RMSE, Pearson r, and R^2. All models used the same data splits and fine-tuning protocol (Table S1).

3.3. Baselines

We compared four transfer variants under identical training and evaluation splits: (1) **BERT-only**, a sequence-only baseline utilizing the PeptideCLM-2- encoder (Feller et al., 2026); (2) **EnsembleEGNN-random-init**, the geometric model trained from random initialization without pretraining; (3) **EnsembleEGNN**, the pretrained geometric model; and (4) **Hybrid**, the full co-training EnsembleEGNN and PeptideCLM-2 with concatenated embeddings.

3.4. Results

Overall performance on the PAMPA dataset details the necessity of the pretrained conformer-aware approach (Table 1). While the randomly initialized EnsembleEGNN performed near random (R^2 = 0.005), pretraining the model on CREMP outperformed the sequence-only BERT baseline (R^2 = 0.477 vs. 0.439). The co-trained Hybrid model ultimately achieved the strongest aggregate metrics, improving the final R^2 to 0.538, resulting in a Pearson r of 0.737.

The Hybrid model consistently achieved the strongest mean R^2 performance across all 5 individual cross-validation folds (Table 2). Visualizations of the overall metrics further demonstrate this performance gap (Fig. 3), supported by per-model held-out prediction scatter plots (Fig. 4). Notably, the EnsembleEGNN model requires pretraining in order to achieve performance above random for 20 epoch finetuning.

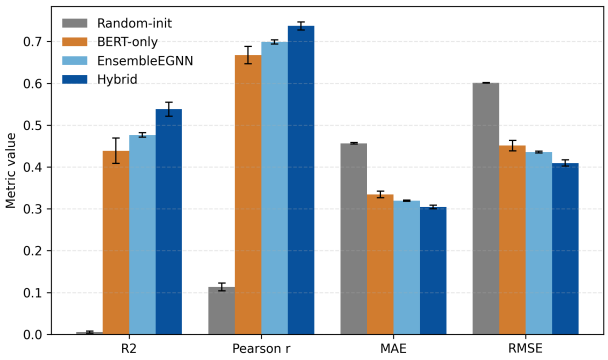

Figure 3. Aggregate performance across cross-validation folds. Comparison of mean evaluation metrics for the EnsembleEGNN w/o pretraining, BERT-only, EnsembleEGNN pretrained, and Hybrid architectures on the held-out PAMPA dataset. The co-trained Hybrid model consistently achieves the strongest overall performance across all evaluated metrics. Error bars represent the standard error across three independent training runs.

4. Discussion

Ensemble representations as a GFM primitive. The conformer aggregation problem is not unique to peptides. Protein loop flexibility (Marks et al., 2018), RNA secondary-structure ensembles (Bauskar et al., 2025), and small-molecule tautomers (Greenwood et al., 2010) all involve distributions over molecular structures. Our results suggest that explicitly learning from multiple conformers, rather than collapsing to a single structure upfront, can be a useful design principle for molecular foundation models.

Transfer learning and scalability. Attempting to train the geometric conformer model from random initialization on the downstream task failed entirely ($R^2 = 0.005$). The shared EGNN backbone must be pretrained on CREMP with self-supervised objectives to establish meaningful physical priors before transferring to downstream regression tasks. This gives hope that with expanded study, the model may be able to generalize outside the training distribution, as has been shown with transfer learning from a model pretrained on a larger corpus. Consistent with the foundation model paradigm, we anticipate that by scaling both conformer datasets and model capacity, the representation advantages of self-supervised geometric pretraining will compound. Notably, the current architecture at 11.3M parameters outperforms a BERT-style architecture that contains 114M parameters.

Computational efficiency and scalability. Utilizing a SAB with inducing points makes conformer fusion effectively linear in the number of conformers when the number of inducing points is fixed, rather than quadratic as in standard self-attention. Should future applications require modeling larger conformer sets, this separation between per-conformer geometric encoding and ensemble pooling should remain useful. The same decomposition also allows the model to scale independently with molecule size and with the number of conformers.

Table 1. Model performance metrics. Cross validation of predictive accuracy on heldout CREMP-CycPeptMPDB data ($n = 2979$). Values are reported as the mean $\pm$ standard error across 3 independent training replicates.

Model	R^2	Pearson r	MAE	RMSE
Random-init	0.005 ± 0.004	0.113 ± 0.016	0.457 ± 0.003	0.601 ± 0.001
BERT-only	0.439 ± 0.053	0.667 ± 0.035	0.334 ± 0.014	0.451 ± 0.022
EnsembleEGNN	0.477 ± 0.009	0.699 ± 0.009	0.319 ± 0.002	0.436 ± 0.004
Hybrid	$\mathbf{0.538 \pm 0.029}$	$\mathbf{0.737 \pm 0.016}$	$\mathbf{0.304 \pm 0.008}$	$\mathbf{0.409 \pm 0.013}$

Table 2. Detailed fold-wise R^2 evaluation. Comparison of predictive performance across five distinct test folds. Results represent the mean $\pm$ standard error of three independent replicates, illustrating the relative stability of the BERT-only, EnsembleEGNN, and Hybrid architectures on different slices of the data.

Fold	BERT-only	EnsembleEGNN	Hybrid
1	0.426 ± 0.019	0.503 ± 0.007	$\mathbf{0.565 \pm 0.011}$
2	0.472 ± 0.026	0.456 ± 0.012	$\mathbf{0.494 \pm 0.035}$
3	0.430 ± 0.061	0.509 ± 0.031	$\mathbf{0.553 \pm 0.018}$
4	0.462 ± 0.057	0.451 ± 0.016	$\mathbf{0.526 \pm 0.054}$
5	0.402 ± 0.060	0.460 ± 0.012	$\mathbf{0.552 \pm 0.018}$

Limitations and future work. Our approach has three primary limitations. Architecturally, the current EGNN layers perform message passing only within each conformer; interactions across conformers are introduced later through conformer pooling rather than through explicit cross-conformer message passing. Data-wise, this study was strictly bounded by the coverage and accuracy of the precomputed CREMP ensembles. Finally, while our hybrid model outperforms the sequence baseline on this benchmark, the CREMP-CycPeptMPDB dataset remains a low-data regime ($n = 2979$), which warrants caution when making generalized cross-paper comparisons. Future work will explore richer cross-conformer interaction mechanisms and integrate scalable, high-fidelity conformer generation to expand beyond existing library constraints.

5. Conclusion

We present EnsembleEGNN, an ensemble geometric foundation model that encodes a molecular ensemble with shared EGNN layers, pooling conformer representations into a single thermodynamically informed embedding. Through multi-task self-supervised pretraining on CREMP, the model learns to recover masked tokens, refine noisy coordinates, and preserve coarse ensemble geometry. With transfer learning to the CREMP-CycPeptMPDB benchmark ($n = 2979$), EnsembleEGNN outperforms a pretrained BERT architecture. A hybrid model further improves over the sequence-only baseline by $+0.099$ in R^2 and $+0.070$ in Pearson r, with correspondingly lower MAE and RMSE. More broadly, this work shows that directly learning from conformer ensembles can improve cyclic-peptide representation learning.

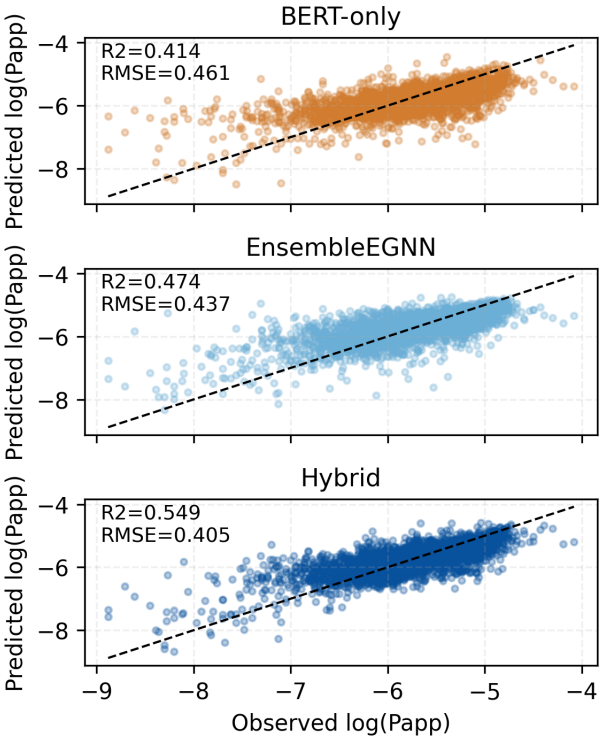

Figure 4. Predictive performance on held-out PAMPA measurements. Parity plot comparing predicted versus experimental $\log(P_{\text{app}})$ values across the complete evaluation set ($n = 2979$). Proximity to the diagonal identity line illustrates the prediction accuracy, highlighting the tighter correlation achieved by the Hybrid model.

Impact

This paper presents work whose goal is to advance the field of machine learning for molecular design and early stage drug discovery. By improving property prediction for highly flexible cyclic peptides, architectures like EnsembleEGNN can help accelerate the development of novel therapeutics and reduce the time and resource costs associated with high throughput physical screening. While advancing computational drug design has overwhelmingly positive societal benefits, we acknowledge that machine learning models in this domain should not be utilized in isolation. Predictions generated by our foundation model are intended to serve as hypothesis generating tools to guide wet lab validation, rather than definitive claims of biological efficacy or safety. We do not foresee any direct negative ethical consequences arising specifically from the methodological advancements presented in this work.

LLM usage

LLMs were used during the preparation of the manuscript solely as writing aids to catch typographical and grammatical errors and to improve writing style and latex formatting.

Code availability

In order to ensure reproducibility and allow the scientific community to build on this concept, we release the model architecture, pretrained checkpoint, and training code, as well as generation of manuscript figures for this paper at https://github.com/AaronFeller/EnsembleEGNN.

Acknowledgments

A.L.F. would like to thank the team at Novo Nordisk for supporting open research, which has led to the completion of this manuscript and the release of all code under MIT license.

References

Aranganathan, A., Gu, X., Wang, D., Vani, B. P., and Tiwary, P. Modeling boltzmann-weighted structural ensembles of proteins using artificial intelligence–based methods. *Current opinion in structural biology*, 91:103000, 2025.

Bauskar, S., Jiao, J., Kannan, N., Kimm, A., Baker, J. M., Tyler, M. J., Bertozzi, A. L., and Andrews, A. M. Boltzmann graph ensemble embeddings for aptamer libraries. In *2025 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 1–6, 2025. doi: 10.1109/ICDMW69685.2025.00125.

Bayaraa, N., Secor, M., Descoteaux, M. L., and Lin, Y.-S. Fast generation of simulation-quality structural ensembles of mixed-chirality cyclic peptides via diffusion models. *Journal of Chemical Theory and Computation*, 22(6): 3103–3113, 2026.

Calvo-Barreiro, L., Secor, M., Damjanovic, J., Abdel-Rahman, S. A., Lin, Y.-S., and Gabr, M. Computational design of a bicyclic peptide inhibitor targeting the icos/icos-l protein–protein interaction. *Chemical Biology & Drug Design*, 105(5):e70117, 2025.

Chandler, D. Introduction to modern statistical mechanics. *Mechanics. Oxford University Press, Oxford, UK*, 5(449): 11, 1987.

Damjanovic, J., Miao, J., Huang, H., and Lin, Y.-S. Elucidating solution structures of cyclic peptides using molecular dynamics simulations. *Chemical reviews*, 121(4):2292–2324, 2021.

Feller, A. L., Secor, M., Swanson, S., Wilke, C. O., and Deibler, K. Scaling smiles-based chemical language models for therapeutic peptide engineering. *Biorxiv: the Preprint Server for Biology*, 2026.

Grambow, C. A., Weir, H., Cunningham, C. N., Biancalani, T., and Chuang, K. V. CREMP: Conformer-rotamer ensembles of macrocyclic peptides for machine learning. *Scientific Data*, 11(1):859, 2024.

Greenwood, J. R., Calkins, D., Sullivan, A. P., and Shelley, J. C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *Journal of computer-aided molecular design*, 24(6):591–604, 2010.

Hsu, C., Verkuil, R., Liu, J., Lin, Z., Hie, B., Sercu, T., Lerer, A., and Rives, A. Learning inverse folding from millions of predicted structures. In *International Conference on Machine Learning*, pp. 8946–8970. PMLR, 2022.

Hui, T., Secor, M., Ho, M. N., Bayaraa, N., and Lin, Y.-S. Molecular dynamics (md)-derived features for canonical and noncanonical amino acids. *Journal of Chemical Information and Modeling*, 65(4):1837–1849, 2025.

Lee, J., Lee, Y., Kim, J., Kosiorek, A., Choi, S., and Teh, Y. W. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International Conference on Machine Learning*, pp. 3744–3753. PMLR, 2019.

Liu, J., Yang, C., Lu, Z., Chen, J., Li, Y., Zhang, M., Bai, T., Fang, Y., Sun, L., Yu, P. S., and Shi, C. Graph foundation models: Concepts, opportunities and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Liu, J., Pan, T., Guo, X., Ran, Z., Hao, Y., Yang, Y., Ng, A. P., Pan, S., Song, J., and Li, F. Molx: A geometric foundation model for protein–ligand modelling. *bioRxiv*, pp. 2026–02, 2026.

Marks, C., Shi, J., and Deane, C. M. Predicting loop conformational ensembles. *Bioinformatics*, 34(6):949–956, 2018.

Méndez-Lucio, O., Nicolaou, C. A., and Earnshaw, B. Mole: a foundation model for molecular graphs using disentangled attention. *Nature communications*, 15(1):9431, 2024.

Otero, D. G., Akbari, O., and Bilodeau, C. Pepmnet: a hybrid deep learning model for predicting peptide properties using hierarchical graph representations. *Molecular Systems Design & Engineering*, 10(3):205–218, 2025.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. In *International Conference on Machine Learning*, pp. 9323–9332. PMLR, 2021.

Tuckerman, M. E. *Statistical mechanics: theory and molecular simulation*. Oxford university press, 2023.

Yudin, A. K. Macrocycles: lessons from the distant past, recent developments, and future directions. *Chemical Science*, 6(1):30–49, 2015.

Zhu, Y., Hwang, J., Adams, K., Liu, Z., Nan, B., Stenfors, B., Du, Y., Chauhan, J., Wiest, O., Isayev, O., Coley, C. W., Sun, Y., and Wang, W. Learning over molecular conformer ensembles: Datasets and benchmarks. In *Proceedings of the Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=vXm9S99v8X`.

Supplemental Table

Table S1. EnsembleEGNN architecture and training hyperparameters. Summary of the verified backbone, conformer-fusion, pretraining, and downstream optimization settings used in the reported experiments. Values reflect the effective configuration of the reported all-atom EnsembleEGNN runs.

Category	Parameter	Value
EGNN backbone	Parameters	7.85M
	Number of EGNN layers (L)	6
	Effective hidden dimension (d)	448
	Number of neighbors (k)	12
	Dropout	0.1
Conformer fusion	Parameters	3.42M
	Fusion mode	`setattn`
	Set-attention blocks	2
	Inducing points	4
	Attention heads	8
	Conformer-prior bias	enabled
Pretraining: EnsembleEGNN	Epochs	80
	Batch size	8
	Learning rate	2×10^{-4}
	Weight decay	1×10^{-4}
	Masking ratio	15%
	Corruption probabilities (mask/random/keep)	0.8 / 0.1 / 0.1
	Coordinate noise std (σ)	0.15 Å
	Loss weights (λ_{tok}, λ_{coord}, λ_{dist})	0.3 / 0.5 / 0.2
Fine-tuning: EnsembleEGNN	Epochs	20
	Batch size	8
	Learning rate	1×10^{-4}
	Weight decay	1×10^{-4}
	Regression head layers	2
	Head dropout	0.1
	Early stopping patience	3
Fine-tuning: BERT	Encoder	`aaronfeller/peptideclm-2-hybrid-base`
	Epochs	20
	Batch size	16
	Learning rate	3×10^{-4}
	Weight decay	1×10^{-2}
	Regression head layers	2
	Head dropout	0.1
	Early stopping patience	3
Fine-tuning: Hybrid	Geometric init checkpoint	`best_pretrain.pt`
	Epochs	20
	Batch size	8
	Learning rates (head / ensemble / CLM)	2×10^{-4} / 1×10^{-4} / 2×10^{-5}
	Weight decay	1×10^{-4}
	Regression head layers	2
	Head dropout	0.1
	Early stopping patience	3