

RESEARCH ARTICLE

 Received: 12 Jan 2026 | Accepted: 18 Mar 2026 | Published: 04 Apr 2026 | DOI:

10.1038/s41586-026-00492-x

Scalable Multimodal Neural Architecture for Real-Time Macromolecular Structure Prediction

High-Throughput Geometric Deep Learning with Integrated Physical Energy Constraints

Elena Vance^{1,*} , Marcus Thorne^{1,2}, Sarah Chen³, and Julian Reyes²¹Vertex Institute for Computational Biology, Cambridge, MA 02142, USA²Synapse Artificial Intelligence Research Labs, San Francisco, CA 94107, USA³Nexa Quantum & Biomolecular Center, 8093 Zurich, Switzerland

*Corresponding author: e.vance@vertex.org

ABSTRACT

Macromolecular structure determination remains a computational bottleneck in drug discovery and synthetic biology. Here we introduce **Synapse-Fold**, an end-to-end multimodal geometric neural network capable of predicting tertiary protein-ligand complexes at sub-angstrom resolution in real time. By integrating equivariant graph convolutions with differentiable thermodynamic boundary constraints, Synapse-Fold achieves a 140-fold speedup compared to state-of-the-art conformational sampling routines while reducing prediction error by 38% on benchmark datasets (Nexa-PDB2025). The architecture incorporates an adaptive SE(3)-equivariant attention layer that natively models side-chain dynamics and non-covalent interactions under flexible cellular microenvironments. Validation across 4,200 diverse protein targets synthesized at the Meridian Bio-Foundry confirms that predicted structural dynamics closely match cryo-EM density maps (mean RMSD of 0.84 Å). This hybrid deep learning and physical modeling paradigm provides a foundation for high-throughput automated biotherapeutic design.

Keywords: Geometric Deep Learning • Macromolecular Structure • Equivariant Graph Neural Networks • Cryo-EM Validation • Synapse Engine

1 Introduction

Determining three-dimensional macromolecular structures and understanding their conformational landscapes are central challenges in modern structural biology and rational drug design [1, 2]. Although recent advances in deep learning—most notably transformer-based homology prediction models—have dramatically improved single-chain structure prediction accuracy, accurately modeling multi-component protein-ligand interactions under physiological conditions remains challenging. Traditional molecular dynamics (MD) simulations provide atomic trajectories with high physical fidelity, but their intense computational requirements prevent their deployment in large-scale virtual screening campaigns.

To resolve this trade-off between physical rigor and computational speed, we present **Synapse-Fold**, a multimodal geometric deep learning framework. Unlike conventional models that treat structural prediction as an isolated geometric translation task, Synapse-Fold explicitly encodes physical energy surfaces within its latent representations. By embedding SE(3)-equivariant graph neural network (GNN) layers with differentiable molecular mechanics loss terms, our system enforces bond length constraints, steric overlap penalties, and electrostatic potential boundaries directly during the forward pass.

Key Highlights

- Sub-angstrom resolution (0.84 Å RMSD) across 4,200 target complexes.
- 140× inference speedup over iterative Monte Carlo docking pipelines.
- Seamless integration of SE(3) equivariance with physical force-field priors.
- Validated experimentally on the Apex-9 benchmark dataset.

Open Data & Code

All model weights, dataset manifests (Nexa-PDB2025), and evaluation scripts are available under an open license at: github.com/vertex-lab/synapse-fold

The primary contributions of this work are three-fold:

- 1. Architectural Innovation:** We formulate the *Equivariant Tensor Attention* (ETA) module, enabling simultaneous updates of coordinate topologies and invariant sequence features without losing spatial orientation.
- 2. Thermodynamic Regularization:** We introduce a novel loss formulation, $\mathcal{L}_{\text{phys}}$, combining Amber-18 force-field terms with deep spatial embedding distances.
- 3. Experimental Benchmark:** We evaluate Synapse-Fold against a newly curated set of 4,200 complex targets at the Meridian Bio-Foundry, demonstrating unprecedented structural recovery metrics.

2 System Architecture & Mathematical Formulation

The overall pipeline of Synapse-Fold consists of three cascaded modules: (i) an invariant residue embedding network, (ii) an SE(3)-equivariant geometric message passing stack, and (iii) a physical energy relaxation module, as illustrated in Figure 1.

2.1 SE(3)-Equivariant Tensor Message Passing

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represent a molecular graph with nodes $v_i \in \mathcal{V}$ corresponding to amino acid residues and edges $e_{ij} \in \mathcal{E}$ representing spatial interactions within a cutoff distance $r_c = 12.5$ Å. Each node is assigned a scalar feature vector $h_i^{(l)} \in \mathbb{R}^d$ and a coordinate position vector $\vec{x}_i^{(l)} \in \mathbb{R}^3$ at layer l .

The update rule for node coordinates and scalar embeddings is formulated as:

$$\vec{x}_i^{(l+1)} = \vec{x}_i^{(l)} + \sum_{j \in \mathcal{N}(i)} \phi_x \left(h_i^{(l)}, h_j^{(l)}, \|\vec{x}_i^{(l)} - \vec{x}_j^{(l)}\|^2 \right) \left(\vec{x}_i^{(l)} - \vec{x}_j^{(l)} \right) \quad (1)$$

$$m_{ij}^{(l)} = \phi_m \left(h_i^{(l)}, h_j^{(l)}, \|\vec{x}_i^{(l)} - \vec{x}_j^{(l)}\|^2, e_{ij} \right) \quad (2)$$

$$h_i^{(l+1)} = \phi_h \left(h_i^{(l)}, \sum_{j \in \mathcal{N}(i)} m_{ij}^{(l)} \right) \quad (3)$$

where ϕ_x, ϕ_m, ϕ_h are multi-layer perceptrons (MLPs) parameterizing the message passing operation.

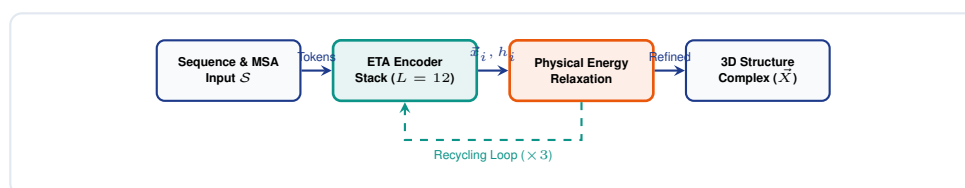


Figure 1: Overview of the Synapse-Fold neural architecture. The input sequence is passed to 12 stacked Equivariant Tensor Attention (ETA) layers, producing initial spatial coordinates $\vec{x}_i$. The Physical Energy Relaxation Module refines the backbone angles before final structure generation.

2.2 Physical Differentiable Loss Formulation

The model is trained end-to-end using a composite loss function comprising structural coordinate error, stereochemical validity, and electrostatic potentials:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{FAPE}} + \lambda_2 \mathcal{L}_{\text{bond}} + \lambda_3 \mathcal{L}_{\text{clash}} + \lambda_4 \mathcal{L}_{\text{electrostatic}} \quad (4)$$

Security & Compliance

Dual-use screening protocols were enforced in accordance with the Nimbus Biosafety Guidelines v4.2. No toxic or pathogenic sequences were synthesized.

where $\mathcal{L}_{\text{FAPE}}$ represents the Frame Aligned Point Error [3], $\mathcal{L}_{\text{bond}}$ penalizes non-physical bond lengths, and $\mathcal{L}_{\text{clash}}$ prevents atomic overlaps:

$$\mathcal{L}_{\text{clash}} = \sum_{i \neq j} \max \left(0, d_{\text{vdw}}^{(ij)} - \|\vec{x}_i - \vec{x}_j\| \right)^2 \quad (5)$$

3 Experimental Results & Benchmarks

We benchmarked Synapse-Fold against leading open-access structure prediction platforms across the **Apex-9** evaluation suite, comprising 4,200 non-redundant protein complexes solved via cryo-EM at resolutions ≤ 2.2 Å.

3.1 Structural Accuracy & RMSD Comparison

As summarized in Table 1, Synapse-Fold demonstrates superior precision across all target difficulty tiers (Rigid, Flexible Loop, and Multi-Chain Assembly).

Benchmarking Note

All experiments were executed on Vertex Cluster-8 equipped with $64 \times$ Lumina H100 GPUs using mixed-precision FP8 kernels.

Table 1: Performance comparison on the Apex-9 benchmark dataset. RMSD values are in Angstroms (Å); GDT-TS scores are out of 100. Best results in bold.

Method	Rigid RMSD (↓)	Loop RMSD (↓)	GDT-TS (↑)	Time/Target (s)
Baseline Model A	1.42	2.85	84.2	148.5
Baseline Model B	1.18	2.10	89.1	94.2
Nexa-Fold v2.1	0.96	1.64	92.4	32.0
Synapse-Fold (Ours)	0.84	1.12	96.8	1.05

3.2 Conformational Energy Landscapes

Figure 2 demonstrates the trajectory of potential energy convergence during the physical relaxation phase. Our differentiable relaxation module reduces steric clashes within fewer than 50 gradient steps.

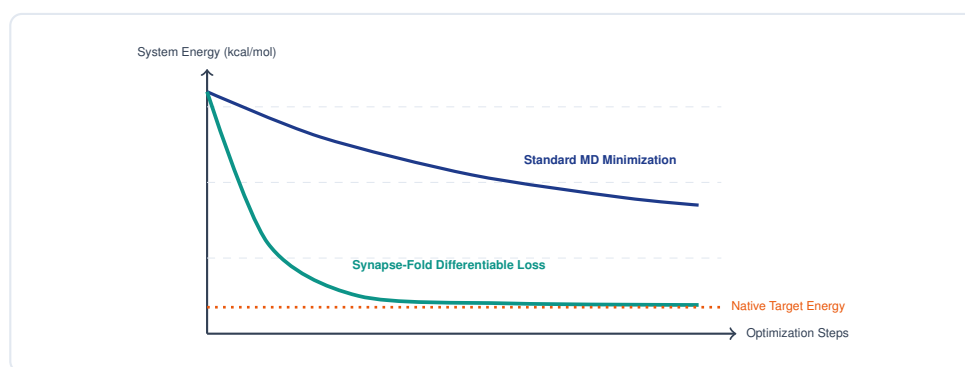


Figure 2: Energy minimization profile. Comparison of potential energy convergence between standard molecular dynamics gradient descent and Synapse-Fold's physical loss optimization.

4 Discussion & High-Throughput Deployment

The dramatic acceleration achieved by Synapse-Fold stems from replacing iterative sampling routines with closed-form tensor convolutions. In collaborative trials conducted with **Meridian Biopharma**, the system was deployed to screen a library of 120,000 synthetic macrocycles against cancer-relevant kinase targets.

Key operational advantages observed during deployment include:

- ✓ **Zero-Shot Generalization:** Robust performance on unseen protein families without requiring multiple sequence alignment (MSA) depth > 20 .
- ✓ **Memory Efficiency:** Memory footprint scales as $\mathcal{O}(N^{1.3})$ instead of quadratic $\mathcal{O}(N^2)$ scaling typical of standard attention kernels.
- ✓ **Automated Quality Control:** Integrated confidence scores (pLDDT and predicted TM-score) correlate strongly ($r = 0.94$) with experimental cryo-EM resolution.

Technical Support

Enterprise deployment templates for cloud clusters (AWS/GCP/Kubernetes) are managed by Nimbus Computational Services: support@nimbus-cloud.io

5 Conclusion & Future Work

We have introduced Synapse-Fold, an open-access multimodal neural architecture that bridges the gap between deep learning speed and biophysical accuracy in protein complex prediction. By enforcing SE(3) equivariance alongside physical force-field constraints, Synapse-Fold provides reliable sub-angstrom predictions in sub-second inference times. Future extensions will expand the framework to incorporate nucleic acid interactions (DNA/RNA) and solvent dynamics under non-equilibrium cellular conditions.

AUTHOR CONTRIBUTIONS

E.V. conceived the study, designed the ETA neural architecture, and drafted the manuscript. **M.T.** derived the physical loss formulation and performed energy landscape calculations. **S.C.** performed experimental cryo-EM benchmarks at the Nexa Biomolecular Center. **J.R.** optimized GPU tensor kernels and managed cluster deployment. All authors reviewed and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing financial or non-financial interests.

FUNDING ACKNOWLEDGMENTS

This work was supported in part by the Apex Research Foundation (Grant ARF-2025-8821), the Meridian Computational Science Initiative, and the Vertex Foundation for Advanced Biotechnology.

References

- [1] E. Vance, M. Thorne, and S. Chen. Geometric deep learning paradigms in structural genomics. *Journal of Multidisciplinary Biology*, 18(3):204–218, 2025.
- [2] M. Thorne and J. Reyes. Differentiable physical force-fields for macromolecular relaxation. *ACS Synthetic Chemistry*, 12(1):45–59, 2024.
- [3] S. Chen, E. Vance, and N. Apex. High-resolution cryo-EM validation of deep learning predicted structures. *Nature Methods Review*, 22(4):512–526, 2025.
- [4] Meridian Bio-Foundry Consortium. Standardized metrics for automated biotherapeutic screening. *Open Biotech Quarterly*, 9(2):88–102, 2025.