GeoPMR: Preserving Relational and Hierarchical Geometry in Multimodal Molecular Representation Learning

Lusheng Li[1], Menglin Yang[2*]
[1] University of Nebraska Medical Center
[2] The Hong Kong University of Science and Technology (Guangzhou)
* Corresponding author

Motivation

Geometric Heterogeneity Across Modalities

A single compound can be described at multiple levels:
atomic identity (1D fingerprint), graph topology (2D), conformational geometry (3D), gene perturbation profile, cellular morphology, and transcriptomic response.

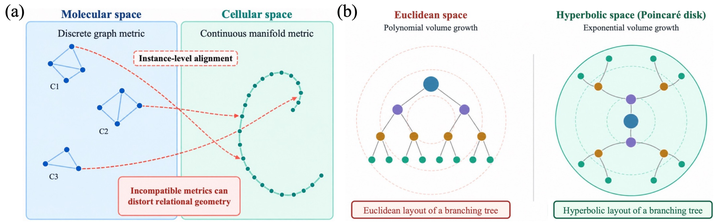
The Problem: Current methods typically perform **instance-level alignment** (using contrastive losses like InfoNCE or non-contrastive regularizers like VICReg) within a shared, symmetric Euclidean embedding space.

The Consequence: Directly pulling individual point embeddings together without accounting for these differences can **distort the relational structure** and neighborhood relationships

Lack of Hierarchical Depth Structure and Modality Ordering

The Problem: Molecular-biological systems naturally exhibit multi-scale, hierarchical dependencies flowing from atomic/structural chemical descriptions down to downstream cellular and phenotypic responses.

The Consequence: Previous multimodal alignment methods treat modalities symmetrically and fail to capture this inherent hierarchical depth and multi-scale organization.

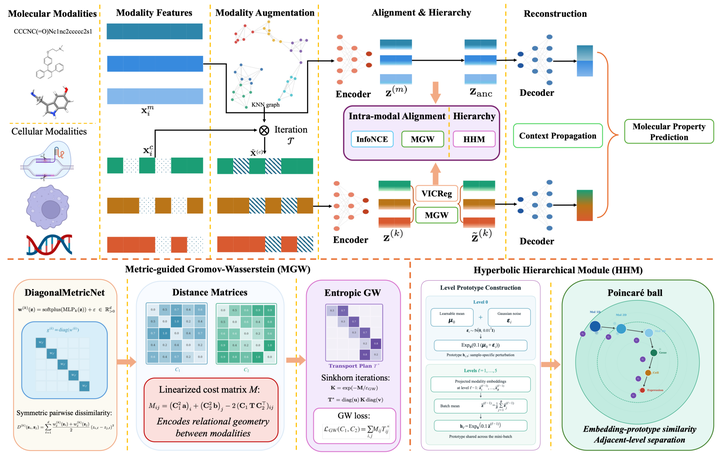


Structural properties across modalities. (a) Molecular and cellular representations carry different intrinsic geometries. Instance-level alignment can distort the within-modality relational structure. (b) Schematic layouts of the same branching tree in Euclidean and hyperbolic spaces.

Summary of the benchmark datasets.

Dataset	Type	Tasks	Molecules	Atoms (Max./Avg.)	Edges (Max./Avg.)
ChEMBL	Classif.	41	2355	61 / 23.7	68 / 25.6
Broad	Classif.	32	6567	74 / 34.1	82 / 36.8
ToxCast	Classif.	617	8576	124 / 18.7	268 / 38.5
ClinTox	Classif.	2	1477	136 / 26.1	286 / 55.7
SIDER	Classif.	27	1427	492 / 33.6	1010 / 70.7
HIV	Classif.	1	41127	222 / 25.5	502 / 54.9
Biogen	Regr.	6	3521	78 / 23.2	84 / 25.3

GeoPMR framework



We propose **GeoPMR** (Geometry-Preserving Molecular Representations), a unified framework that captures structural properties through **metric-guided Gromov-Wasserstein (MGW)** alignment and a **hyperbolic hierarchical module (HHM)**, respectively.

Metric-guided Gromov-Wasserstein (MGW)

The total MGW term aggregates over cross-modal pairs (molecular anchor vs. each cellular modality) and augmentation pairs:

$$\mathcal{L}_{\mathrm{MGW}} = \sum_{k=4}^{6} \mathcal{L}_{\mathrm{GW}}\left(\mathbf{C}_{\mathrm{anc}}, \mathbf{C}^{(k)}\right) + \sum_{k=4}^{6} \mathcal{L}_{\mathrm{GW}}\left(\mathbf{C}^{(k)}, \tilde{\mathbf{C}}^{(k)}\right)$$

Hyperbolic hierarchical module (HHM)

the *embedding-prototype similarity loss* penalizes the squared geodesic distance between each modality embedding and its corresponding level prototype:

$$\mathcal{L}_{\mathrm{sim}} = \frac{1}{6B} \sum_{\ell=0}^{5} \sum_{i=1}^{B} \frac{d_{\mathbb{B}}^2\left(\tilde{\mathbf{z}}_i^{(\ell)}, \mathbf{h}_{i,\ell}\right)}{\tau_{\mathrm{hier}}^2}$$

the *adjacent-level separation loss* compares embeddings of the same molecule at consecutive modality levels:

$$\mathcal{L}_{\mathrm{dis}} = \frac{1}{5B} \sum_{\ell=0}^{4} \sum_{i=1}^{B} \mathrm{softplus}\left(m - d_{\mathbb{B}}\left(\tilde{\mathbf{z}}_i^{(\ell)}, \tilde{\mathbf{z}}_i^{(\ell+1)}\right)\right)$$

The total pretraining objective:

$$\mathcal{L} = \mathcal{L}_{\mathrm{CPR}} + \lambda_1 \underbrace{\left(\sum_{m \in \{g,c,e\}} \mathcal{L}_{\mathrm{NCE}}(\mathbf{z}^{\mathrm{anc}}, \mathbf{z}^{(m)}) + \sum_{m \in \{g,c,e\}} \mathcal{L}_{\mathrm{VIC}}(\mathbf{z}^{(m)}, \tilde{\mathbf{z}}^{(m)}) + \alpha \sum_{m} \mathcal{L}_{\mathrm{MGW}} \right)}_{\mathcal{L}_{\mathrm{align}}} + \lambda_2 \mathcal{L}_{\mathrm{hier}}$$

Experiments

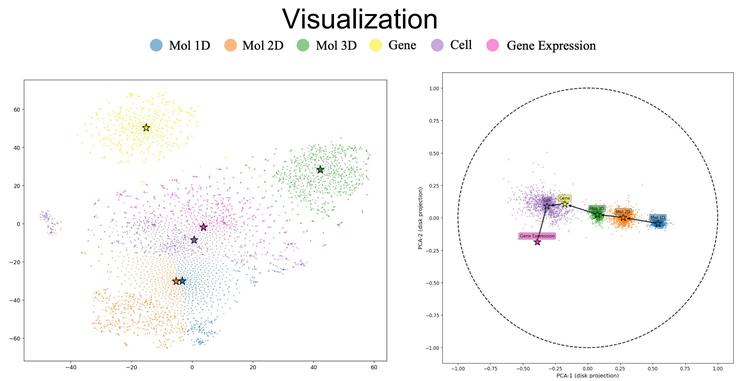
Main Results on Molecular Property Prediction

	Method	ChEMBL (2355/41)	ToxCast (8576/617)	Broad (6567/32)	Biogen (3521/6) — MAE ×10² ↓						
		AUC% ↑	AUC% ↑	AUC% ↑	Avg.	HLM	RLM	ER	Sol.	hPPB	rPPB
Single-modal	MLP	76.8 ± 2.2	57.6 ± 1.0	63.3 ± 0.3	66.2 ± 2.4	66.1 ± 2.6	69.5 ± 3.0	56.8 ± 2.3	56.5 ± 4.2	74.6 ± 6.2	73.7 ± 7.3
	RF	54.7 ± 0.7	52.3 ± 0.1	55.5 ± 0.1	52.8 ± 0.2	44.2 ± 0.1	51.6 ± 0.1	44.2 ± 0.1	42.0 ± 0.2	67.7 ± 0.7	66.9 ± 0.9
	GP	51.0 ± 0.0	OOM	50.6 ± 0.0	60.0 ± 0.0	51.3 ± 0.0	61.6 ± 0.0	59.5 ± 0.0	49.7 ± 0.0	68.8 ± 0.0	69.3 ± 0.0
	AttrMask [20]	73.9 ± 0.5	63.1 ± 0.8	59.8 ± 0.2	67.3 ± 0.3	82.4 ± 1.1	99.1 ± 1.2	49.8 ± 0.7	51.7 ± 1.0	57.9 ± 0.6	62.6 ± 0.5
	ContextPred [20]	77.0 ± 0.5	63.0 ± 0.6	60.0 ± 0.2	68.5 ± 0.9	85.0 ± 7.9	96.5 ± 3.7	49.7 ± 0.4	55.1 ± 2.7	61.4 ± 1.8	63.1 ± 0.5
	EdgePred [20]	75.6 ± 0.5	63.5 ± 1.1	59.9 ± 0.2	67.8 ± 0.9	81.2 ± 10.2	99.1 ± 6.9	48.0 ± 0.5	53.5 ± 2.8	62.2 ± 1.8	62.9 ± 0.7
	GraphCL [51]	75.6 ± 1.6	52.2 ± 0.2	67.2 ± 0.5	53.9 ± 0.6	43.8 ± 0.3	49.6 ± 0.3	45.4 ± 0.6	40.6 ± 0.5	76.7 ± 1.0	67.1 ± 2.2
	GROVER [21]	73.3 ± 1.4	53.1 ± 0.4	66.2 ± 0.1	54.9 ± 1.6	44.5 ± 0.4	52.6 ± 0.3	46.5 ± 0.7	41.7 ± 0.6	73.2 ± 5.7	71.0 ± 4.3
	JOAO [52]	75.1 ± 1.0	52.3 ± 0.2	67.3 ± 0.4	55.0 ± 0.8	44.5 ± 0.5	51.4 ± 0.6	47.6 ± 0.5	40.6 ± 0.2	74.3 ± 2.8	71.5 ± 2.6
	MGSSL [53]	75.1 ± 1.1	64.2 ± 0.2	66.9 ± 0.5	53.2 ± 0.3	44.8 ± 0.6	41.5 ± 0.2	65.6 ± 1.8	64.6 ± 0.5	52.7 ± 0.5	49.7 ± 0.3
	GraphLoG [54]	73.5 ± 0.7	58.6 ± 0.4	62.9 ± 0.4	56.9 ± 0.4	49.3 ± 0.3	58.8 ± 0.5	54.8 ± 0.5	42.6 ± 0.3	66.8 ± 1.7	69.0 ± 1.3
	GraphMAE [55]	74.7 ± 0.1	53.3 ± 0.1	66.8 ± 0.3	52.8 ± 0.8	43.3 ± 0.9	50.9 ± 1.4	51.2 ± 0.8	40.9 ± 0.3	64.4 ± 2.7	65.9 ± 3.8
	D-SLA [56]	69.3 ± 1.0	57.8 ± 0.5	63.3 ± 0.3	57.9 ± 0.7	50.4 ± 0.7	60.9 ± 0.6	53.6 ± 1.7	43.3 ± 0.9	68.6 ± 1.2	70.8 ± 2.0
Multimodal	RoBERTa-102M [57]	74.7 ± 1.9	64.2 ± 0.8	59.8 ± 0.7	69.0 ± 2.6	71.4 ± 14.5	76.7 ± 13.2	65.1 ± 19.2	63.7 ± 24.6	67.5 ± 5.2	69.9 ± 4.9
	GPT2-87M [57]	71.0 ± 3.4	61.5 ± 1.1	60.6 ± 0.3	74.0 ± 8.5	65.4 ± 12.9	81.8 ± 25.5	73.1 ± 20.8	54.1 ± 12.9	83.2 ± 21.5	86.1 ± 19.8
	MolT5 [58]	69.9 ± 0.8	64.7 ± 0.9	55.1 ± 0.9	65.1 ± 0.5	76.7 ± 2.1	65.3 ± 1.7	55.9 ± 1.1	49.2 ± 1.0	70.3 ± 0.8	73.1 ± 1.0
	Uni-Mol [22]	76.8 ± 0.4	64.6 ± 0.2	65.4 ± 0.1	55.8 ± 2.8	50.1 ± 5.2	59.9 ± 6.6	49.9 ± 5.6	43.6 ± 1.1	65.4 ± 4.9	65.8 ± 1.2
	CLOOME [4]	66.7 ± 1.8	54.2 ± 0.9	61.7 ± 0.4	64.3 ± 0.4	65.2 ± 1.5	75.0 ± 2.1	56.9 ± 0.8	44.2 ± 0.8	70.7 ± 0.4	73.6 ± 0.8
	InfoCORE (GE) [5]	79.3 ± 0.9	65.3 ± 0.2	60.2 ± 0.2	69.9 ± 1.2	79.9 ± 3.6	80.3 ± 0.9	51.6 ± 1.8	51.3 ± 2.1	78.6 ± 0.3	77.8 ± 1.9
	InfoCORE (CP) [5]	73.8 ± 2.0	62.4 ± 0.4	61.1 ± 0.2	71.0 ± 0.6	74.5 ± 4.9	84.4 ± 1.0	53.5 ± 0.7	53.6 ± 2.1	80.8 ± 1.5	79.4 ± 3.4
	InfoAlign [6]	81.3 ± 0.6	66.4 ± 1.1	69.5 ± 0.3	49.4 ± 0.2	39.7 ± 0.4	48.4 ± 0.6	39.2 ± 0.3	40.5 ± 0.6	66.7 ± 1.7	62.0 ± 1.5
	CHMR [7]†	84.5 ± 0.2	69.3 ± 0.5	68.5 ± 0.2	41.2 ± 0.3	33.7 ± 0.2	39.8 ± 0.1	**34.3 ± 0.1**	35.4 ± 0.2	53.9 ± 0.9	49.9 ± 1.0
	GeoPMR (Ours)	**85.7 ± 0.6**	**70.1 ± 0.2**	**69.9 ± 0.3**	**40.5 ± 0.3**	**33.4 ± 0.3**	**39.6 ± 0.3**	35.1 ± 0.6	**35.3 ± 0.3**	**51.8 ± 0.5**	**47.9 ± 1.2**

Ablation Studies

Models	ChEMBL↑	ToxCast↑	Broad↑	Biogen↓
a. **GeoPMR**	**85.7 ± 0.6**	**70.1 ± 0.2**	**69.9 ± 0.3**	**40.5 ± 0.3**
Intra-Modal Alignment				
b. w/o MGW	83.7 ± 0.4	69.2 ± 0.2	69.5 ± 0.3	42.2 ± 0.3
Hyperbolic Hierarchical Module				
c. w/o HHM	84.3 ± 0.3	69.4 ± 0.2	69.5 ± 0.2	41.9 ± 0.3
d. Eucl. HHM	84.8 ± 0.2	69.2 ± 0.4	68.9 ± 0.1	41.9 ± 0.3
SOTA baseline				
e. CHMR	84.5 ± 0.2	69.3 ± 0.5	68.5 ± 0.2	41.2 ± 0.3

Visualization



(a) t-SNE visualization. (b) Poincaré disk view.

Code and datasets
https://github.com/lulu-beta/GeoPMR