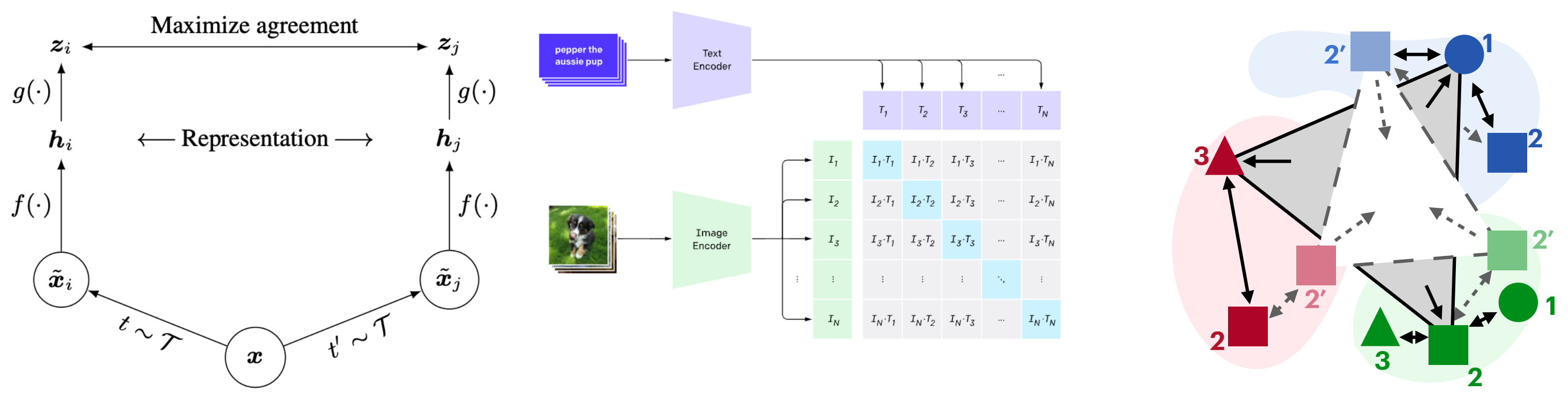
Triangular Contrastive Learning on Molecular Graphs

MinGyu Choi, Wonseok Shin,
Yijingxiu Lu, Sun Kim

Goal: Trimodal Representation Learning



Unimodal (SimCLR) **Bimodal** (CLIP) **Trimodal** (Ours)

Molecular object has multimodality (sequence, graph, and structure), while often labels often correspond with unimodal data. For example, protein property data usually contain protein sequences but their structure are usually unknown. To distill unapproachable high-order information to the low-order approachable modality, bimodal contrastive pretraining could be a good choice. Considering 1D/2D/3D nature of molecular data, molecular pretraining must contrast three modalities - however, ***contrast on trimodality has rarely been discussed***. Here, we propose ***geometry-aware contrastive framework - Triangular Contrastive Learning***, which minimize and maximize the areas of Triangles, instead of pairwise distances.

Observations on Trimodal Embedding Space

Alignment and Uniformity

- **Alignment**: Positive pairs are mapped closely in the embedding space.
- **Uniformity**: Embeddings are uniformly distributed, preserving as much information as possible.

Transformation of embedding spaces.

Intramodal - '***hypersphere***': distributes the embedding space
Intermodal - 'line': compresses the embedding space
Joint Joint intra- and intermodal: '*cones*'

Triangle Area Loss - '*angular diversified cones*'

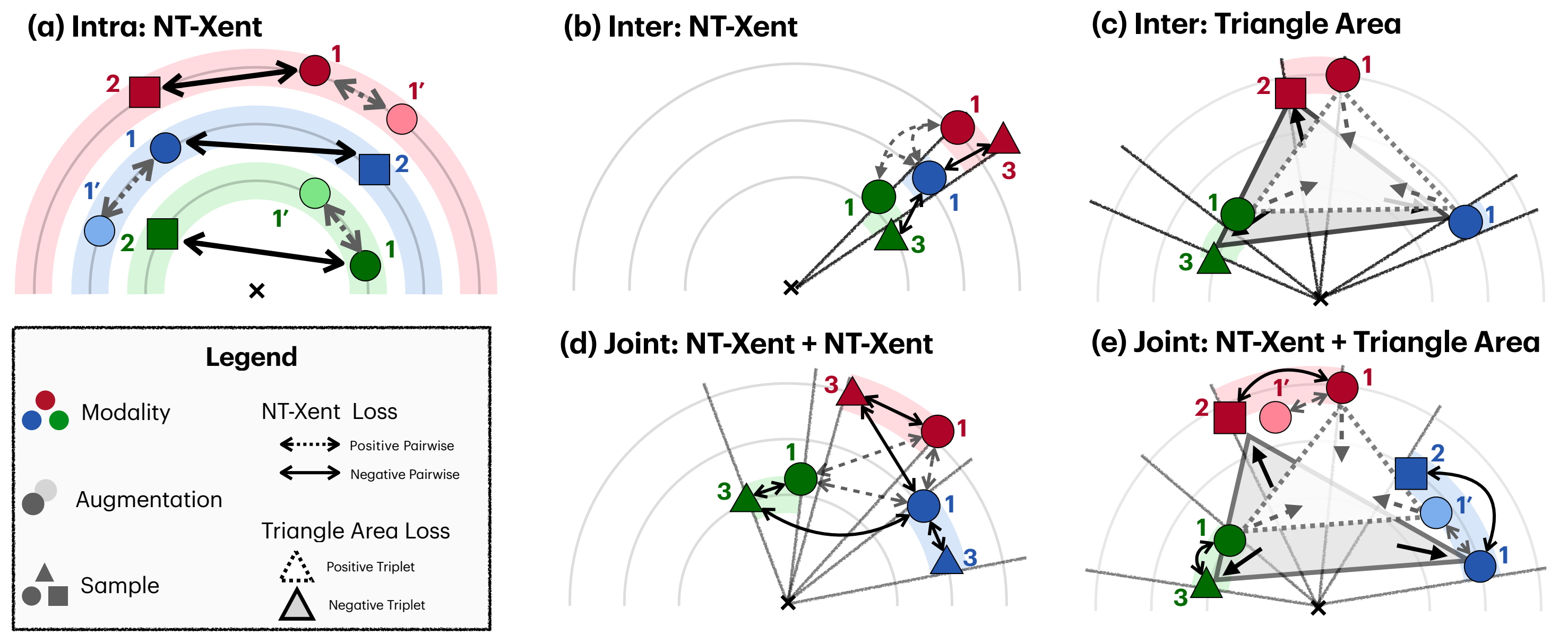


Figure 1. Illustration of embedding space after trimodal contrastive learning. Specific loss function and geometry of each space: (a) NT-Xent as intramodal loss: 'hypersphere' (b) NT-Xent as intermodal loss: 'line' (c) Triangle Area Loss as intermodal loss: 'line' (d) NT-Xent as intra- and intermodal loss: 'cone' (e) Triangle Area Loss as intermodal loss, NT-Xent as intramodal loss: 'cone'. Angles within the space and angles between them are not to scale.

TriCL: Triangular Contrastive Learning

TriCL Framework

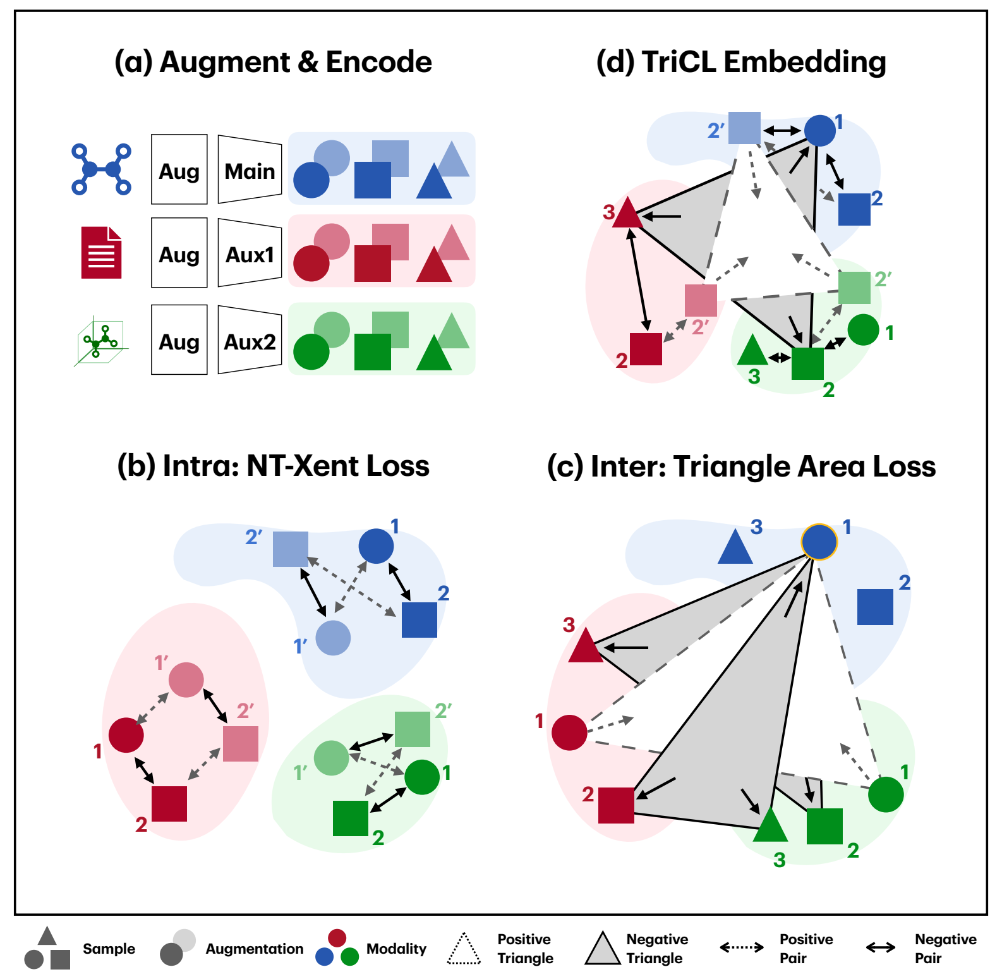


Figure 2.The TriCL framework. (a) Each sample is represented as three distinct format; after augmented twice then encoded generating six reprsentations per sample. (b) Representations in different modalities are contrasted using Triangle Area Loss. (c) Representations in the same modality are contrasted using pairwise NT-Xent loss. (d) TriCL build the embedding space by carefully balancing intramodal and intermodal contrastive loss.

Geometry-aware Triangular Area Loss

Triangular Area Loss for inter-model contrastive learning.

$$\mathcal{L}_{\text{inter}} = \underbrace{\mathbb{E}\left[\text{Area}(\mathbf{z}_i^{\text{main}}, \mathbf{z}_j^{\text{aux1}}, \mathbf{z}_k^{\text{aux2}})^2 \mid \mathbf{P}\right]}_{\text{intermodal alignment}} - \underbrace{\mathbb{E}\left[\text{Area}(\mathbf{z}_i^{\text{main}}, \mathbf{z}_j^{\text{aux1}}, \mathbf{z}_k^{\text{aux2}})^2 \mid \mathbf{N}\right]}_{\text{intermodal uniformity}}$$

For intra-modal contrastive learning, we use pairwise NT-Xent loss.

$$\mathcal{L}_{\text{intra}}^{\text{enc}} = \frac{1}{2B}\sum_{k=1}^{B}\left(\ell(2k-1, 2k) + \ell(2k, 2k-1)\right)$$

$$\ell(i,j) = \underbrace{-\frac{1}{n\tau}\sum_{i,j}\text{sim}(\boldsymbol{z}_i^{\text{enc}}, \boldsymbol{z}_j^{\text{enc}})}_{\text{intramodal alignment}} + \underbrace{\frac{1}{n}\sum_{i}\log\sum_{k=1}^{2n}\mathbb{1}_{k\neq i}\exp(\text{sim}(\boldsymbol{z}_i^{\text{enc}}, \boldsymbol{z}_k^{\text{enc}})/\tau)}_{\text{intramodal uniformity}}$$

Then TriCL optimizes: $\mathcal{L} = \lambda_{\text{intra}}^{\text{main}}\mathcal{L}_{\text{intra}}^{\text{main}} + \mathcal{L}_{\text{intra}}^{\text{aux1}} + \mathcal{L}_{\text{intra}}^{\text{aux2}} + \lambda_{\text{inter}}\mathcal{L}_{\text{inter}}$

Comparison with other methods

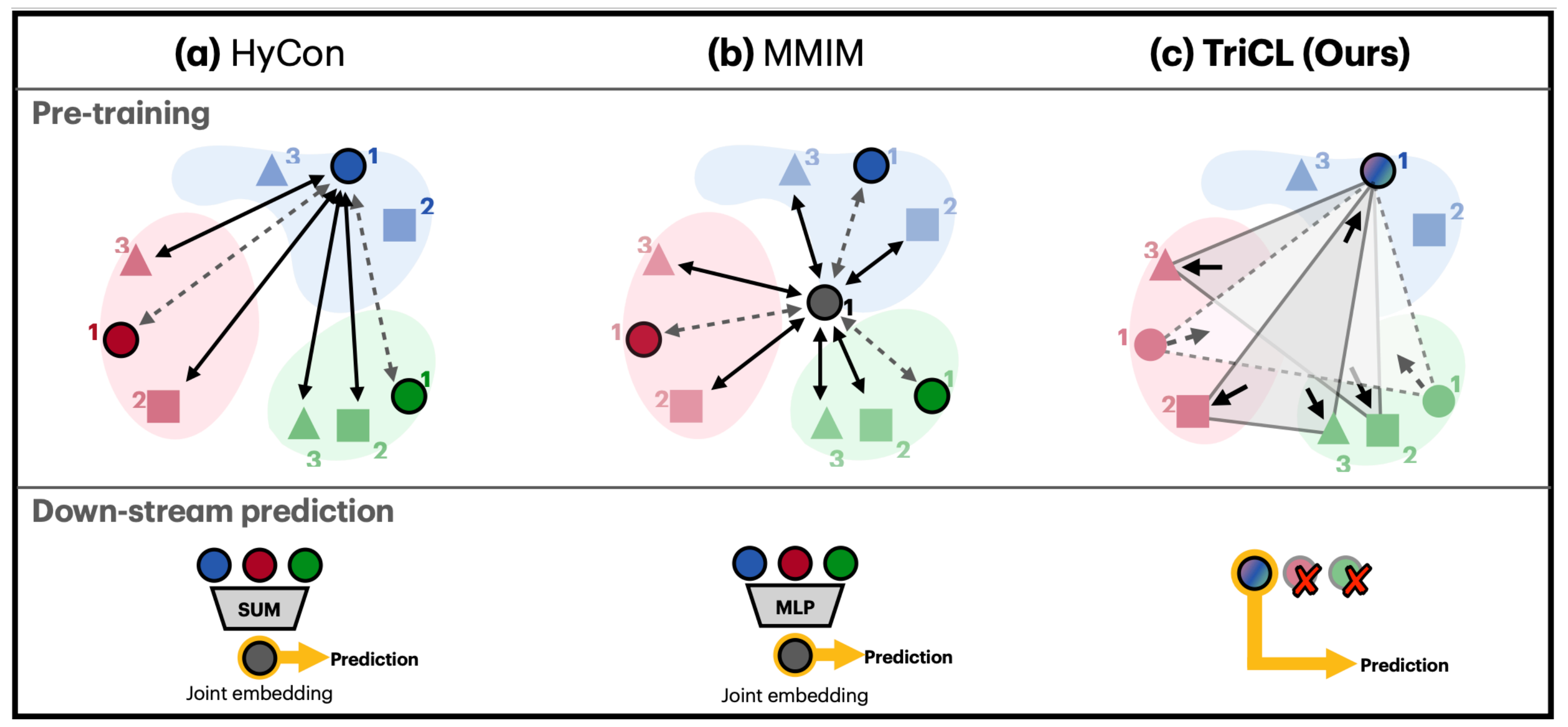


Figure 3.Comparison with previous trimodal models. (a) HyCon uses pairwise contrastive learning with two auxiliary modalities, but fine-tune all three models on downstream tasks. (b) MMIM generates a unified representation via pairwise contrastive learning, and also uses all three modalities for downstream prediction. (c) TriCL (Ours) exploits geometry-aware Triangular contrastive learning, while uses only one encoder for downstream task.

Molecular Property Prediction (MoleculeNet)

Pre-training	BBBP	Tox21	ToxCast	SIDER	ClinTox	MUV	HIV	BACE	AVG
-	65.4(2.4)	74.9(0.8)	61.6(1.2)	58.0(2.4)	58.8(5.5)	71.0(2.5)	75.3(0.5)	72.6(4.9)	67.21
EdgePred	64.5(3.1)	74.5(0.4)	60.8(0.5)	56.7(0.1)	55.8(6.2)	73.3(1.6)	75.1(0.8)	64.6(4.7)	65.64
AttrMask	70.2(0.5)	74.2(0.8)	62.5(0.4)	60.4(0.6)	68.6(9.6)	73.9(1.3)	74.3(1.3)	77.2(1.4)	70.16
GPT-GNN	64.5(1.1)	74.2(0.8)	62.5(0.4)	60.4(0.6)	68.6(9.6)	73.9(1.3)	74.3(1.3)	77.2(1.4)	68.27.
InfoGraph	69.2(0.8)	73.0(0.7)	62.0(0.3)	59.2(0.2)	75.1(5.0)	74.0(1.5)	74.5(1.8)	73.9(2.5)	70.10
ContextPred	71.2(0.9)	73.3(0.5)	62.8(0.3)	59.3(1.4)	73.7(4.0)	72.5(2.2)	75.8(1.1)	78.6(1.4)	70.89
GraphLoG	67.8(1.7)	73.0(0.3)	62.2(0.4)	57.4(2.3)	62.0(1.8)	73.1(1.7)	73.4(0.6)	78.8(0.7)	68.47
G-Motif	66.4(3.4)	73.2(0.8)	62.6(0.5)	60.6(1.1)	77.8(2.0)	73.3(2.0)	73.8(1.4)	73.4(4.0)	70.14
GraphCL	67.5(3.3)	75.0(0.3)	62.8(0.2)	60.1(1.3)	78.9(4.2)	77.1(1.0)	75.0(0.4)	68.7(7.8)	70.64
JOAO	66.0(0.6)	74.4(0.7)	62.7(0.6)	60.7(1.0)	66.3(3.9)	77.0(2.2)	76.6(0.5)	72.9(2.0)	69.57
GraphMVP-G	70.8(0.5)	**75.9(0.5)**	63.1(0.2)	60.2(1.1)	79.1(2.8).	**77.7(0.6)**	76.0(0.1).	79.3(1.5)	72.76
GraphMVP-C	**72.4(1.6)**	74.4(0.2)	63.1(0.4)	**63.9(1.2)**	77.5(4.2)	75.0(1.0)	77.0(1.2)	81.2(0.9)	73.07
TriCL(OURS)	**72.4(0.4)**	75.5(0.3)	**63.9(0.4)**	62.0(1.0)	**85.4(1.9)**	77.0(0.8)	**78.9(0.5)**	**82.5(1.2)**	**74.71**

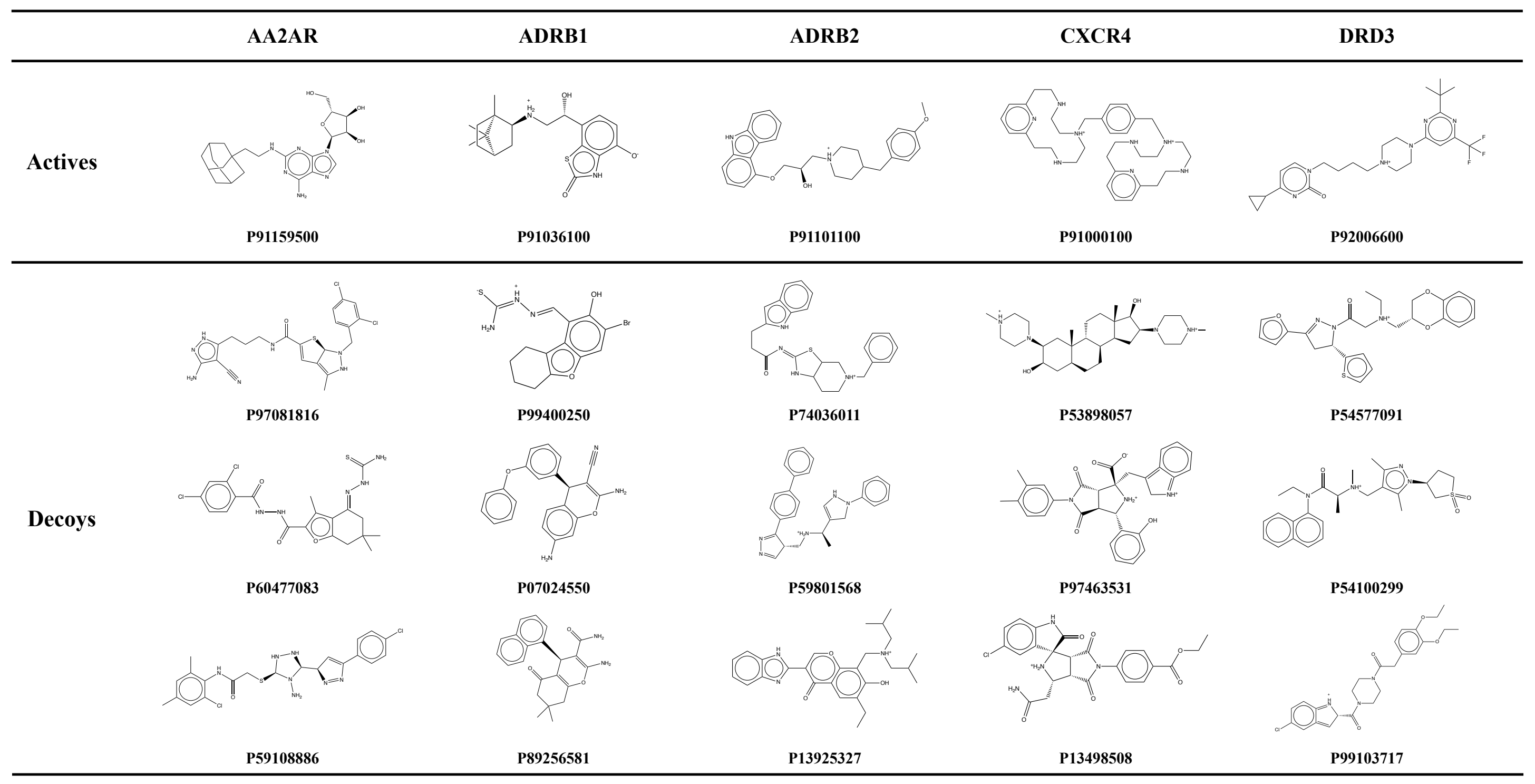
Table 1. Results on the molecular property prediction classification tasks. We report an average test AUC-ROC on 8 downstream tasks with standard deviation inside the parenthesis. Top 1 AUC-ROC score for each task is underlined and bolded. Datasets were scaffold splitted. Finetuning was repeated under 3 independent seeds {0, 1, 42}. We report the test AUC-ROC at the epoch which validation AUC-ROC was the highest.

Ligand/Decoy Discrimination (GPCR)

Embedding space assessment using GPCR active / decoy examples

	GPCR active compounds			Target Instances (Alignment)				
	Align	Uniform	Combined	AA2AR	ADRB1	ADRB2	CXCR4	DRD3
GNN (Unimodal CL)	0.574	0.546	0.028	**0.317**	0.324	0.324	0.233	0.388
TriCL	**0.602**	**0.316**	**0.286**	0.299	**0.368**	**0.384**	**0.381**	**0.458**

Table 2. Case study on GPCR-binding compounds. Alignment metric is the average cosine similarity between all active compounds targeting GPCRs or the same GPCR (higher is better). Uniformity metric is the average cosine similarity between GPCR-targeting compounds and others (close to 0 is better). Combined metric refers to (Align − |Uniform|) (higher is better).



	AA2AR	ADRB1	ADRB2	CXCR4	DRD3
Actives	P91159500	P91036100	P91101100	P91000100	P92006600
Decoys	P97081816	P99400250	P74036011	P53898057	P54577091
	P60477083	P07024550	P59801568	P97463531	P54100299
	P59108886	P89256581	P13925327	P13498508	P99103717

Table 3. Selected active and decoy compounds in DUD-E GPCR subset. Labels below each structures are protonation codes, provided in DUD-E dataset.

References

Chen et. al. (2020). "A Simple Framework for Contrastive Learning of Visual Representations." ICML2020

Radford et. al. (2021). "Learning Transferable Visual Models From Natural Language Supervision.". ICML2021

Liu et. al. (2021). "Pre-training Molecular Graph Representation with 3D Geometry.". ICLR2022

Choi et. al. (2022). "Triangular Contrastive Learning on Molecular Graphs.". Arxiv