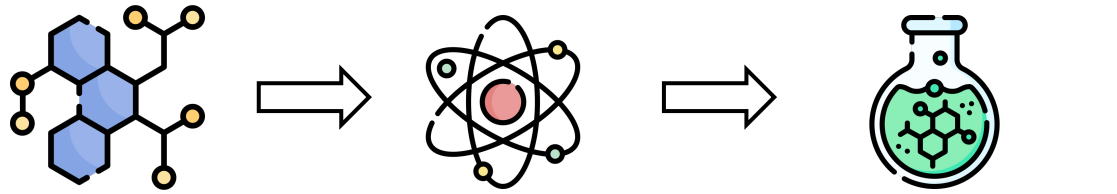
TRIDENT: Tri-Modal Molecular Representation Learning with Taxonomic Annotations and Local Correspondence

Feng Jiang[1], Mangal Prakash[2], Hehuan Ma[1], Jianyuan Deng[2], Yuzhi Guo[1], Amina Mollaysa[2], Tommaso Mansi[2], Rui Liao[2], Junzhou Huang[1]

[1]University of Texas at Arlington [2]Johnson & Johnson Innovative Medicine

Background

- Molecular Representation Learning



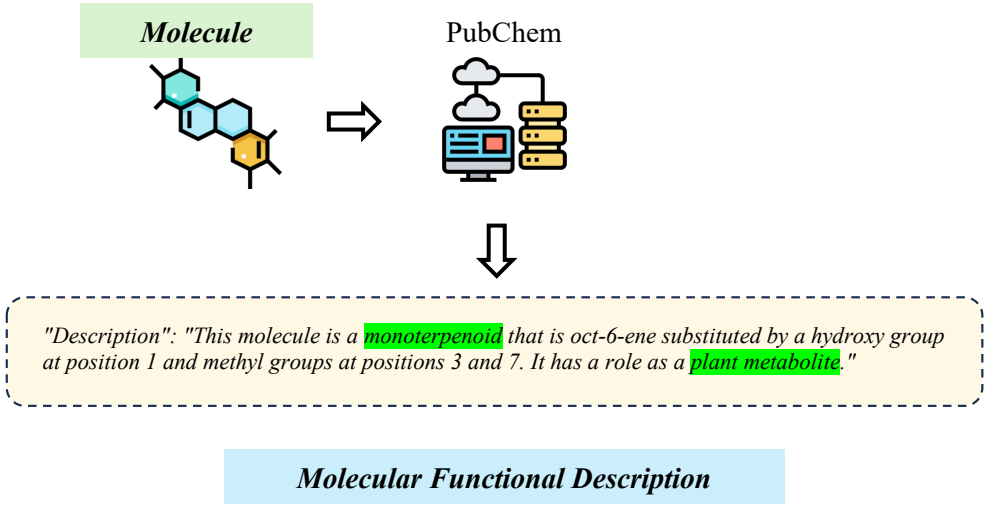
Molecule *Model* *Molecule property prediction*

- Multimodal Molecular Representation Learning

Molecule *Molecular function description* *Model*

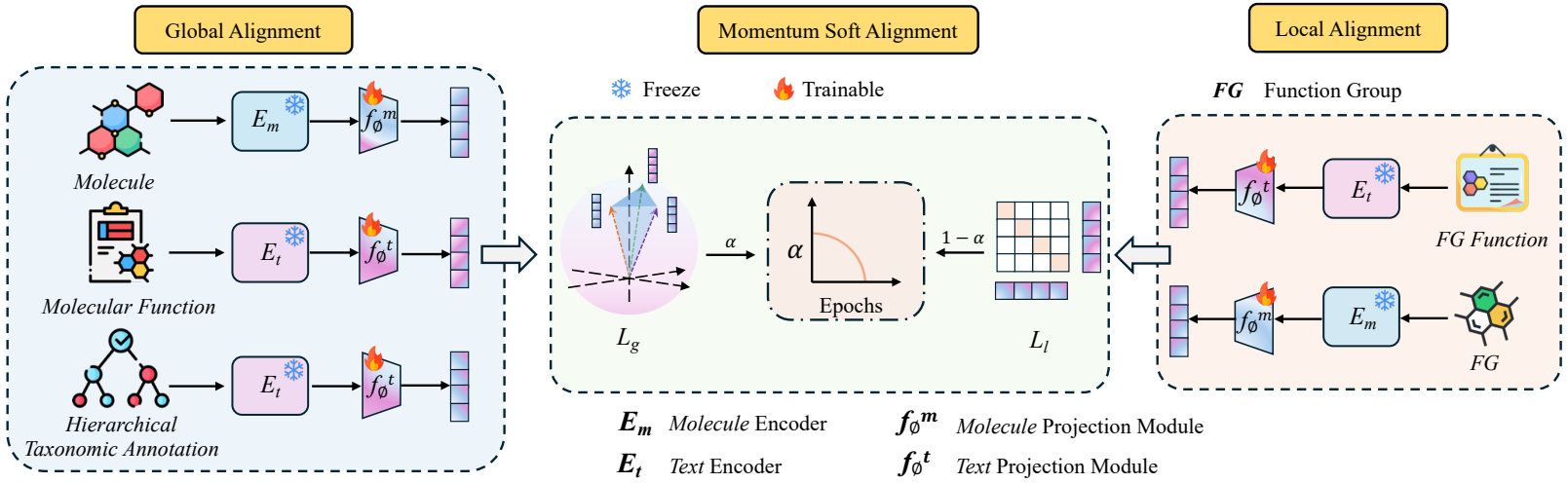
Limitations of Current Methods

- Flat Representations: Existing methods simplify molecules with unified functional descriptions, ignoring nuanced multi-taxonomy annotations
- Alignment Challenges: Pairwise alignment schemes struggle to model interdependencies across multiple modalities with hierarchical information.
- Missing Fine-Grained Details: Focus on molecule-level alignment neglects relationships between substructures and sub-textual descriptions.



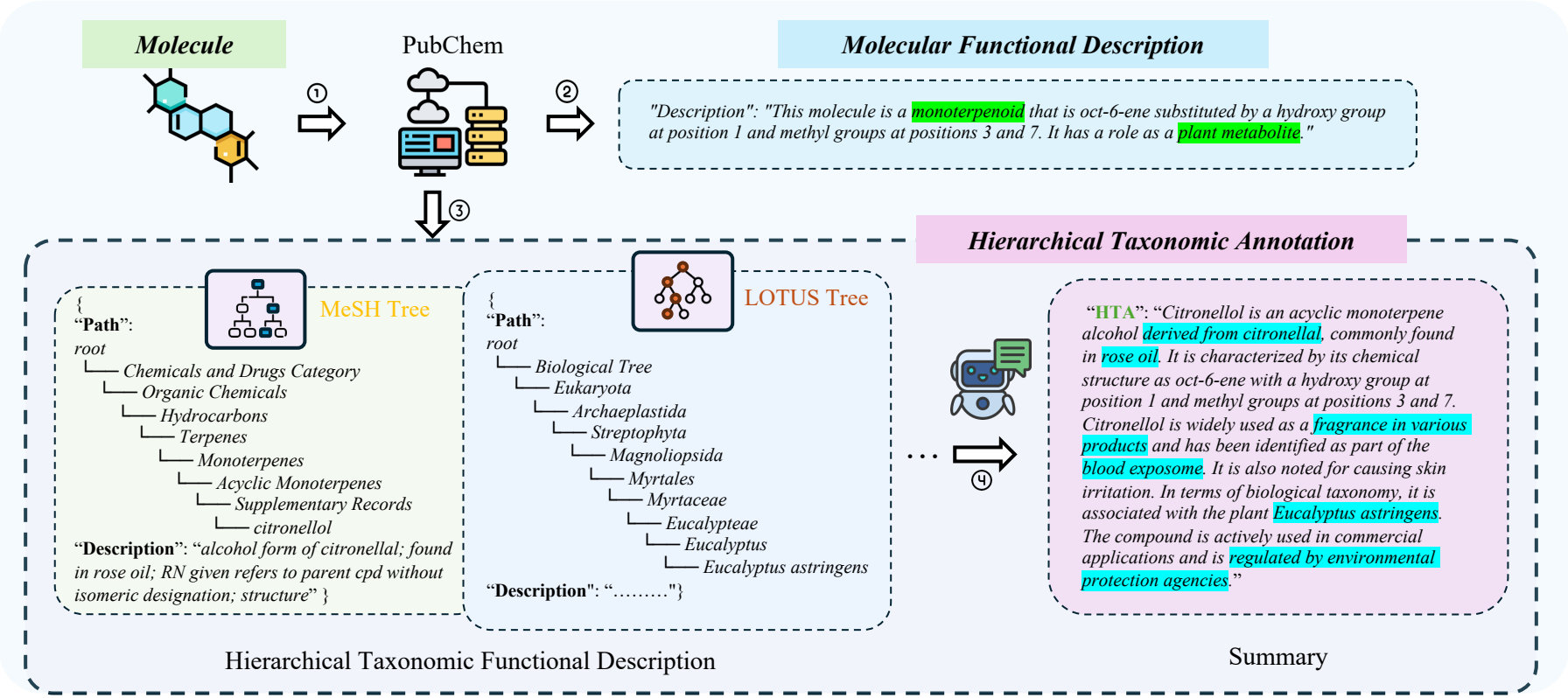
Key Contributions

- Flat Representations => Hierarchical Taxonomic Annotation
- Alignment Challenges => Volume-Based Global Alignment
- Missing Fine-Grained Details => Fine-Grained Local Alignment
- Momentum-Based Dynamic Integration (Global & Local Alignment)



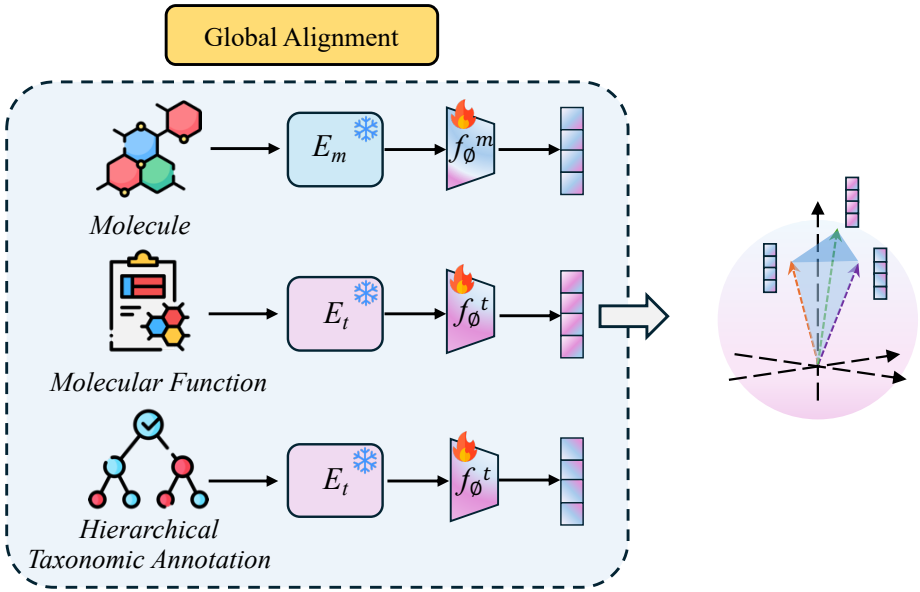
Hierarchical Taxonomic Annotation

Curated 47,269 <SMILES, Text, HTA> triplets with 32 diverse classification systems, enabling structured multi-level molecular understanding.



Volume-Based Global Alignment

Geometry-aware tri-modal alignment captures higher-order relationships across all modalities simultaneously.



$$\mathcal{L}_{\text{M2TH}} = -\frac{1}{B}\sum_{i=1}^{B}\log\frac{\exp(-\text{Vol}(m_i,t_i,h_i)/\tau)}{\sum_{j=1}^{B}\exp(-\text{Vol}(m_i,t_j,h_j)/\tau)},$$

$$\mathcal{L}_{\text{TH2M}} = -\frac{1}{B}\sum_{i=1}^{B}\log\frac{\exp(-\text{Vol}(m_i,t_i,h_i)/\tau)}{\sum_{j=1}^{B}\exp(-\text{Vol}(m_j,t_i,h_i)/\tau)}.$$

$$\mathcal{L}_{\text{g}} = \frac{1}{2}(\mathcal{L}_{\text{M2TH}} + \mathcal{L}_{\text{TH2M}}).$$

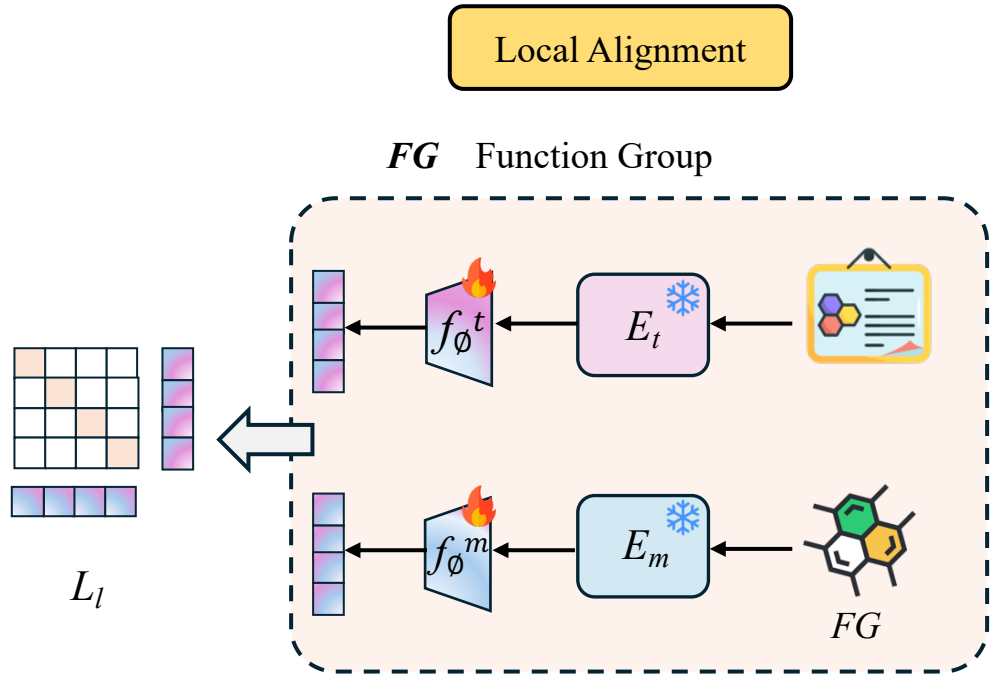
Fine-Grained Local Alignment

Contrastive learning between functional groups and their textual descriptions for each molecule.

$$\mathcal{L}_{FG2T} = -\frac{1}{B}\sum_{i=1}^{B}\log\frac{\exp(fg_{pooled,i}\cdot fgt_{pooled,i}/\tau)}{\sum_{j=1}^{B}\exp(fg_{pooled,i}\cdot fgt_{pooled,j}/\tau)},$$

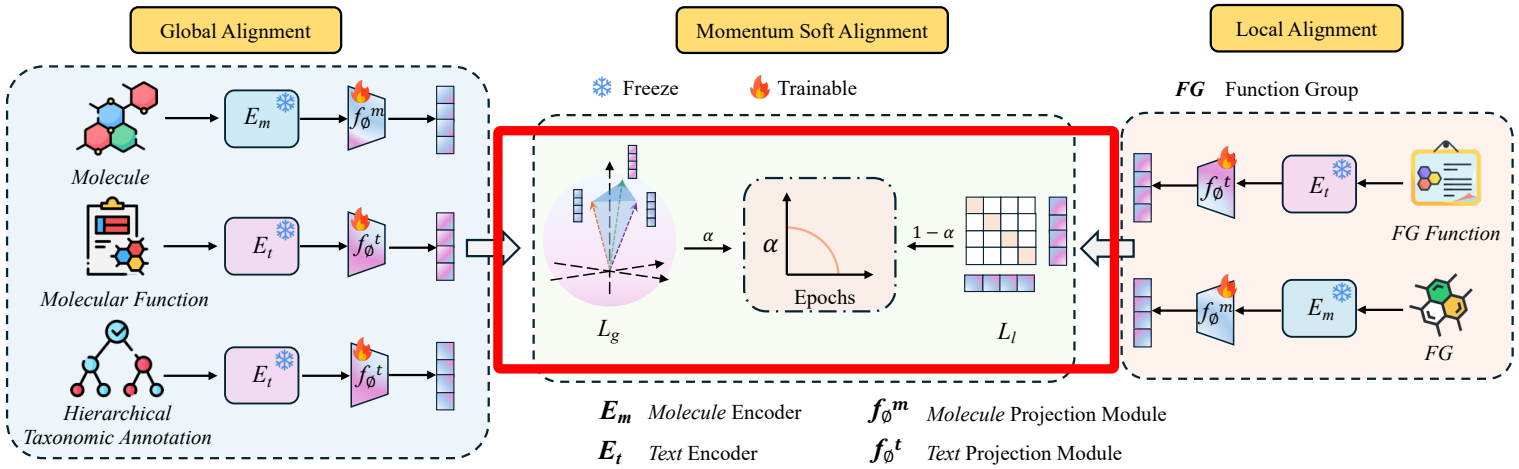
$$\mathcal{L}_{T2FG} = -\frac{1}{B}\sum_{i=1}^{B}\log\frac{\exp(fg_{pooled,i}\cdot fgt_{pooled,i}/\tau)}{\sum_{j=1}^{B}\exp(fg_{pooled,j}\cdot fgt_{pooled,i}/\tau)},$$

$$\mathcal{L}_{\mathrm{l}} = \frac{1}{2}(\mathcal{L}_{FG2T} + \mathcal{L}_{T2FG}),$$



Momentum-Based Dynamic Integration

Adaptive balancing mechanism dynamically adjusts weights between global and local losses.



$$\mathcal{L} = \alpha \mathcal{L}_{\mathrm{g}} + (1-\alpha)\mathcal{L}_{\mathrm{l}},$$

$$\alpha_t = \beta \alpha_{t-1} + (1-\beta) \cdot \frac{\mathcal{L}_{\mathrm{g}}^{(t)}}{\mathcal{L}_{\mathrm{g}}^{(t)} + \mathcal{L}_{\mathrm{l}}^{(t)}},$$

SOTA Performance

Performance comparison on MoleculeNet molecule property prediction

Method	BBBP	Tox21	ToxCast	Sider	ClinTox	MUV	HIV	Bace	Avg
MOLFORMER	70.74±1.34	74.74±0.56	65.51±0.63	61.75±1.23	77.64±0.98	67.58±1.01	75.64±1.76	78.64±2.35	71.53
KV-PLM	70.50±0.54	72.12±1.02	55.03±1.65	59.83±0.56	89.17±2.73	54.63±4.81	65.40±1.69	75.80±2.73	67.81
MegaMolBART	68.89±0.17	73.89±0.67	63.32±0.79	59.52±1.79	78.12±4.62	61.51±2.75	71.04±1.70	82.46±0.84	69.84
MoleculeSTM-SMILES	70.75±1.90	75.71±0.89	65.17±0.37	63.70±0.81	86.60±2.28	65.69±1.46	77.02±0.44	81.99±0.41	73.33
MolFM	72.90±0.10	77.20±0.70	64.40±0.20	64.20±0.90	79.70±1.60	76.00±0.80	78.80±1.10	83.90±1.10	74.64
MoMu	70.50±2.00	75.60±0.30	63.40±0.50	60.50±0.90	79.90±4.10	70.50±1.40	75.90±0.80	76.70±2.10	71.63
Atomas	73.72±1.67	77.88±0.36	66.94±0.90	64.40±1.90	93.16±0.50	76.30±0.70	80.55±0.43	83.14±1.71	77.01
MolCA-SMILES	70.80±0.60	76.00±0.50	56.20±0.70	61.10±1.20	89.00±1.70	-	-	79.30±0.80	72.10
TRIDENT (M-S)	73.14±0.44	78.23±0.12	67.79±0.56	**64.62±0.47**	**95.75±0.71**	82.88±1.41	79.64±1.15	**84.19±0.95**	78.28
TRIDENT (M-M)	**73.95±1.01**	**79.36±0.13**	**67.80±0.37**	63.64±0.56	95.41±0.66	**83.51±0.48**	**81.63±0.52**	82.39±0.56	**78.46**

SOTA Performance

Performance comparison on TDC

Method	DILI (475 drugs)		Carcinogens (278 drugs)		Skin Reaction (404 drugs)	
	AUC	ACC	AUC	ACC	AUC	ACC
MOLFORMER	85.59±1.39	76.39±5.24	77.27±0.76	77.32±1.47	63.75±1.41	60.98±3.44
KV-PLM	73.46±0.61	62.50±2.08	75.18±3.71	76.01±1.75	62.88±2.30	59.76±5.17
MolT5	77.37±1.15	69.44±1.20	86.89±1.00	84.45±1.11	68.67±3.99	62.22±1.41
MoMu	80.44±2.47	75.00±4.17	80.11±1.50	78.00±2.62	61.63±1.94	56.10±3.45
MolCA-SMILES	88.34±1.28	80.56±2.40	82.00±1.80	78.76±0.52	65.13±0.88	62.20±1.72
MoleculeSTM-SMILES	91.20±2.02	84.72±2.41	83.87±1.30	81.05±0.63	67.72±0.50	61.60±0.73
MolXPT	91.67±0.76	84.03±3.19	75.76±2.73	80.90±2.06	61.08±1.28	62.60±1.40
BioT5	82.45±1.81	76.39±3.18	82.83±4.31	76.19±2.06	68.27±4.41	62.21±1.06
BioT5+	82.58±1.65	80.56±1.20	86.62±2.32	77.41±2.00	65.25±0.66	62.27±1.20
Atomas	90.17±1.30	85.08±2.16	82.47±2.11	80.75±0.50	70.33±0.88	61.79±6.14
TRIDENT (M-S)	**95.08±0.70**	**86.81±2.40**	83.42±1.10	81.47±0.92	70.33±0.63	**63.42±4.22**
TRIDENT (M-M)	94.56±0.88	86.80±3.18	**87.07±0.77**	**84.62±1.07**	**72.00±1.09**	62.60±1.40

Summary

TRIDENT replaces conventional contrastive loss with a volume-based global alignment objective for soft, geometry-aware tri-modal alignment.

It further introduces a local alignment objective linking molecular substructures to sub-textual descriptions, while a momentum mechanism dynamically balances global and local learning for both semantic and structural alignment.

Thank you!