2025 NeurIPS Spotlight

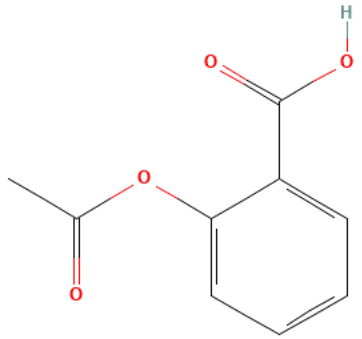
3D Interaction Geometric Pre-training for Molecular Relational Learning

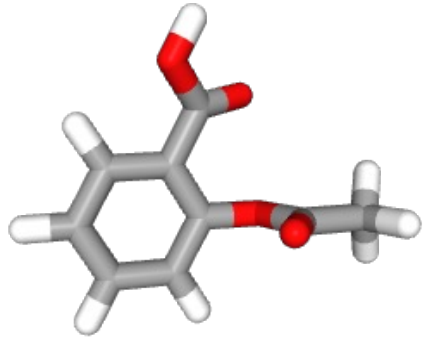
Namkyeong Lee, Yunhak Oh, Heewoong Noh, Gyoung S. Na
Minkai Xu, Hanchen Wang, Tianfan Fu, Chanyoung Park

KAIST KRICT Stanford University 南京大学 NANJING UNIVERSITY NEURAL INFORMATION PROCESSING SYSTEMS

BACKGROUND REPRESENTATION OF MOLECULE

CC(=O)OC1=CC=CC=C1C(=O)O

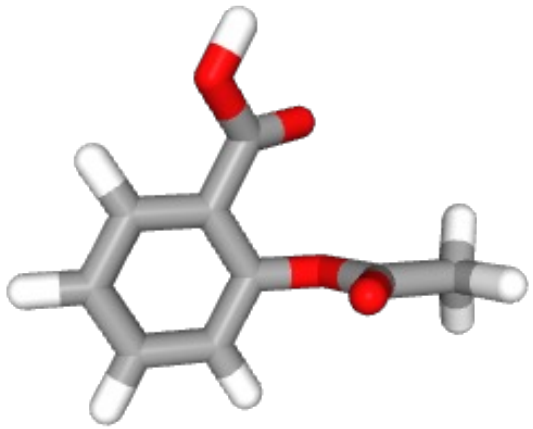




SMILES String | **2D Topology** | **3D Geometry**

Diverse molecular representations exist!

BACKGROUND REPRESENTATION OF MOLECULE



3D Geometry

Why 3D Geometry Matters?

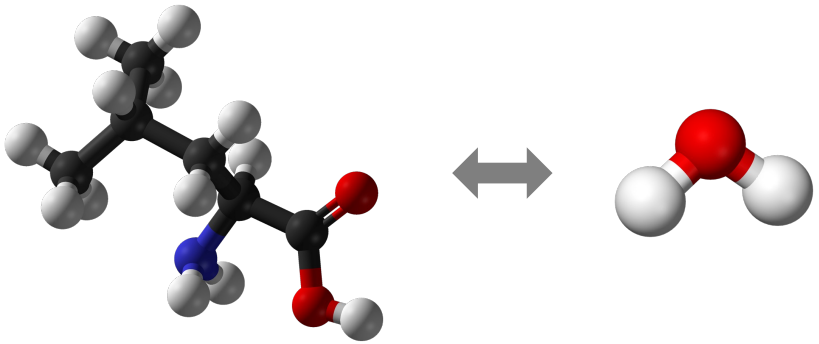
- Essential for predicting molecular properties
- Controls chemical reactivity and interactions
- Key to biological recognition (Lock-and-Key)

Challenges

Trade-off between cost and accuracy

- ETKDG algorithm → Fast but **less accurate**
- CREST algorithm → More balanced, but still **very slow**

BACKGROUND MOLECULAR RELATIONAL LEARNING



Molecular Relational Learning

Learning the interaction behavior between a pair of molecules

Examples

Materials Discovery

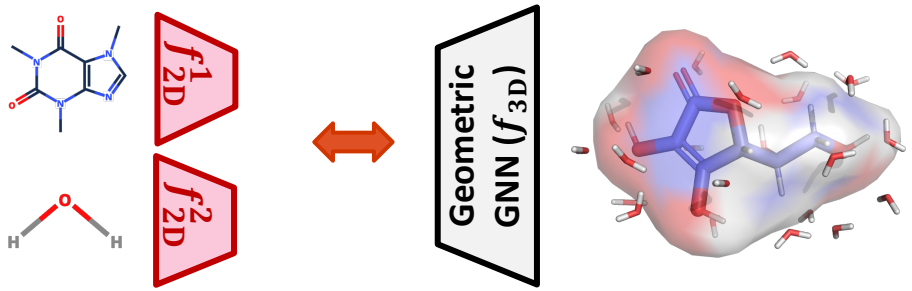
- Predicting optical properties when a **Chromophore** and **Solvent** react
- Predicting solubility when a **solute** and **solvent** react
- Predicting side effects when taking **two types of drugs** simultaneously

Drug Discovery

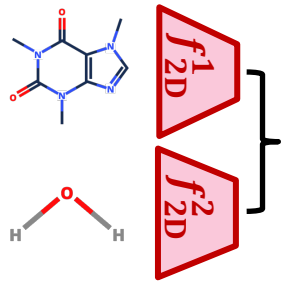
Can we learn utilize 3D geometry for molecular relational learning?

BACKGROUND MOLECULAR RELATIONAL LEARNING

Step1. Pre-training Stage



Step2. Fine-tuning Stage



Various properties
Absorption
Solvation Free Energy
Drug-Drug Interaction
…

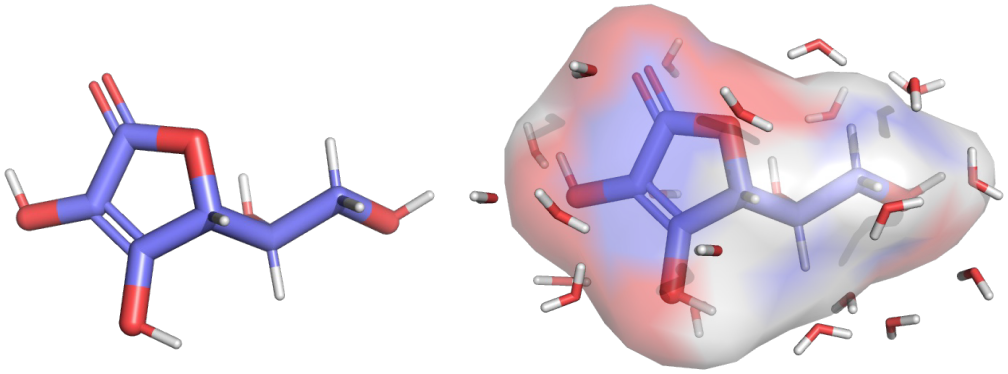
Pre-training & Fine-tuning Framework

- Learn interaction geometry during Pre-training stage
- Predict target properties from 2D structure in Fine-tuning stage

Challenges

- How to get interaction geometry of molecules?
- Pre-training strategy for molecular relational learning?

BACKGROUND 3D GEOMETRY FOR MOLECULAR RELATIONAL LEARNING



Single Molecule

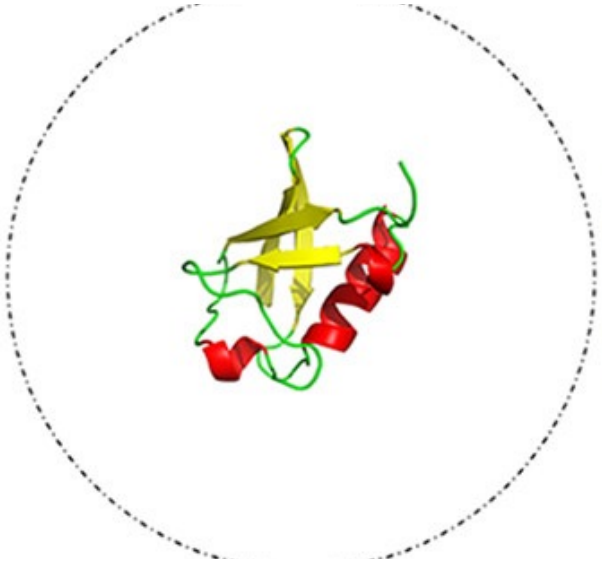
Molecular Interaction Environment

How to calculate interaction geometry?

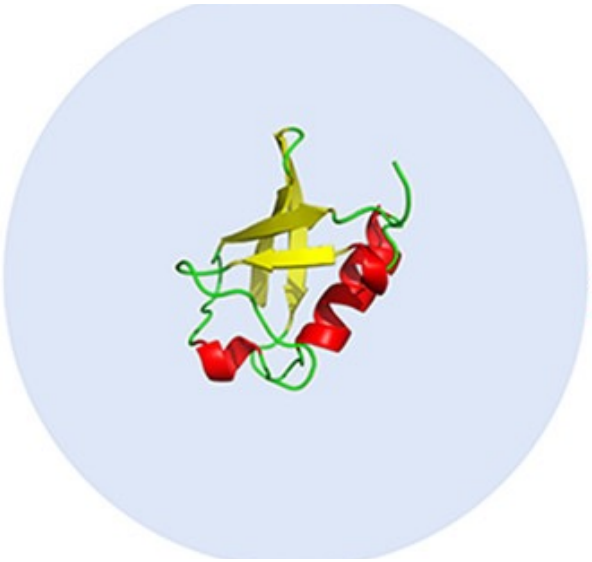
- The geometry of individual molecules
- The relative spatial arrangements between multiple molecules

BACKGROUND 3D GEOMETRY FOR MOLECULAR RELATIONAL LEARNING

In Molecular Dynamics Simulation…

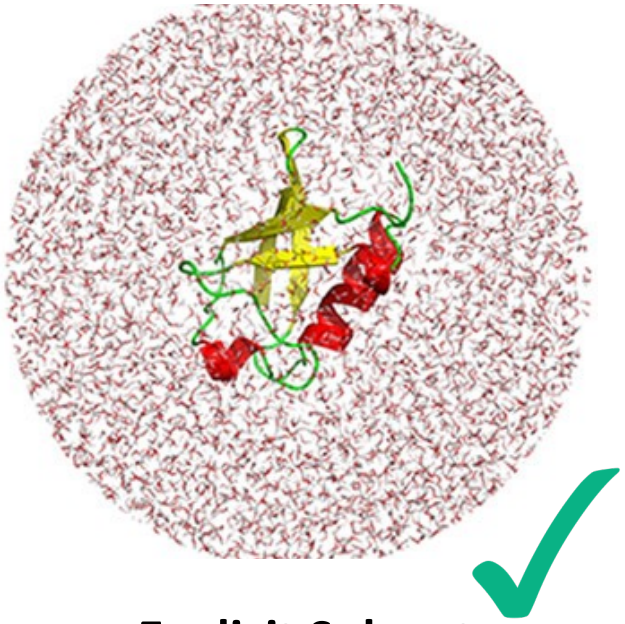


Vacuum



Implicit Solvent

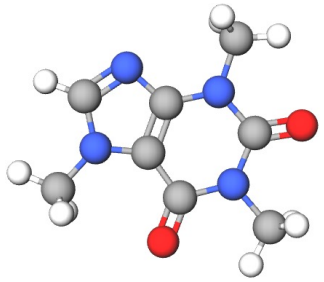
- Less accurate
- Cheap



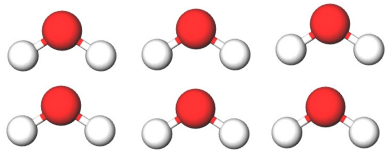
Explicit Solvent

- More accurate
- Expensive

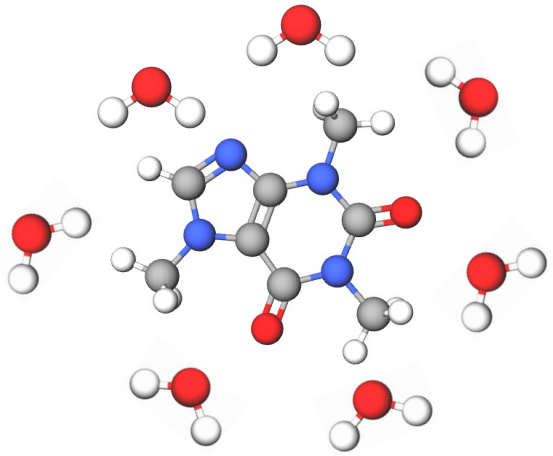
METHODOLOGY VIRTUAL INTERACTION GEOMETRY CONSTRUCTION



Larger Molecule

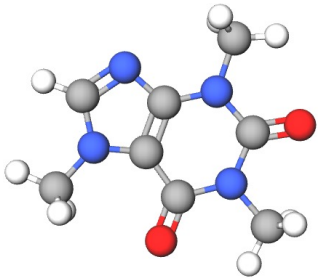


Smaller Molecule

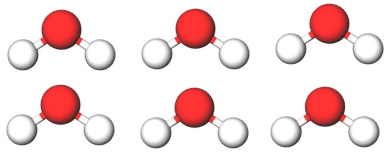


Virtual Interaction Geometry

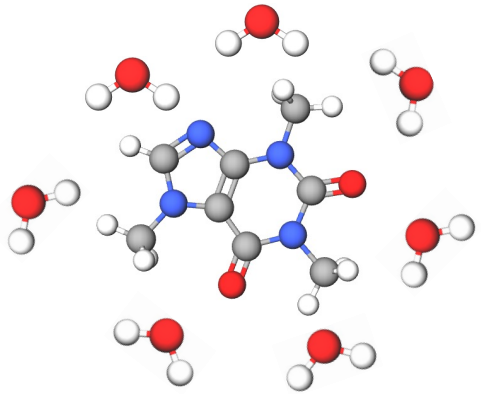
METHODOLOGY VIRTUAL INTERACTION GEOMETRY CONSTRUCTION



Larger Molecule



Smaller Molecule



Virtual Interaction Geometry

[Step1] Select target atom from larger molecule

[Step2] Position smaller molecules near target atom

Create random direction vector ε

Scale the direction vector by the smaller molecule's radius **(r^2)**

$$\mathbf{R}^{2,i} = \mathbf{R}^2 + \varepsilon_i * r^2 + \mathbf{R}_i^1$$

Transit smaller molecules near target atom

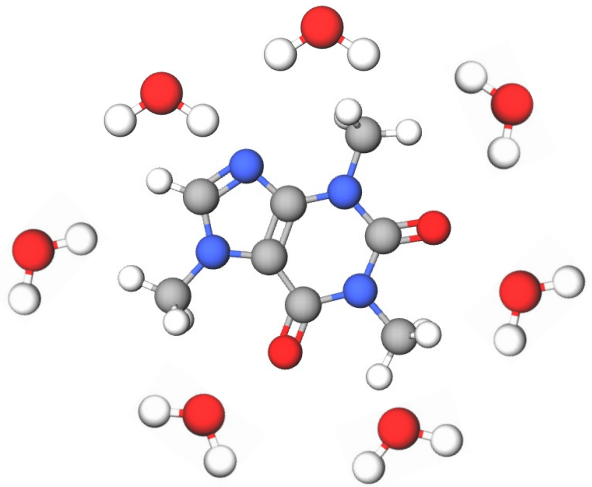
[Step3] Interaction Geometry

$$\mathbf{R}_{\text{vr}} = (\mathbf{R}^1 \| \underbrace{\mathbf{R}^{2,1} \| \ldots \| \mathbf{R}^{2,i} \| \ldots \| \mathbf{R}^{2,n}}_{\text{Smaller molecules}}) \in \mathbb{R}^{(N^1 + n \cdot N^2) \times 3}$$

$$\mathbf{X}_{\text{vr}} = (\mathbf{X}^1 \| \mathbf{X}^2 \| \ldots \| \mathbf{X}^2) \in \mathbb{R}^{(N^1 + n \cdot N^2) \times F}$$

$$g_{\text{vr}} = (\mathbf{X}_{\text{vr}}, \mathbf{R}_{\text{vr}})$$

METHODOLOGY PRE-TRAINING FOR 3D INTERACTION GEOMETRY



Virtual Interaction Geometry

SE(3) Invariant Global Geometry

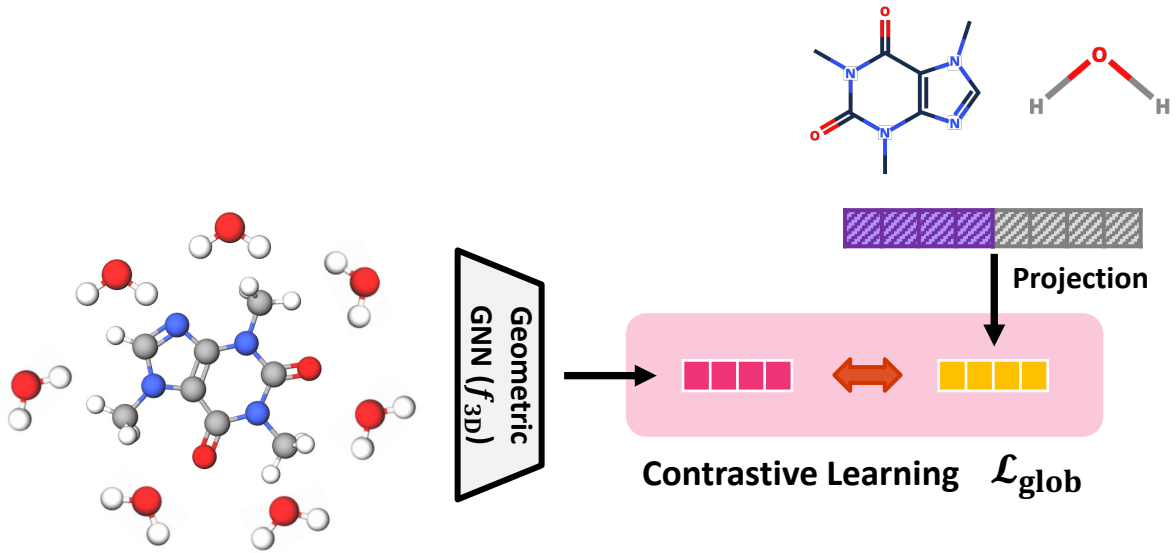
Reveals overall geometry of interaction environment

SE(3) Equivariant Local Geometry

Fine-grained relative geometry between large and small molecules

- Utilize **Local frame** for equivariance

METHODOLOGY PRE-TRAINING FOR 3D INTERACTION GEOMETRY



SE(3) Invariant Global Geometry

Reveals overall geometry of interaction environment

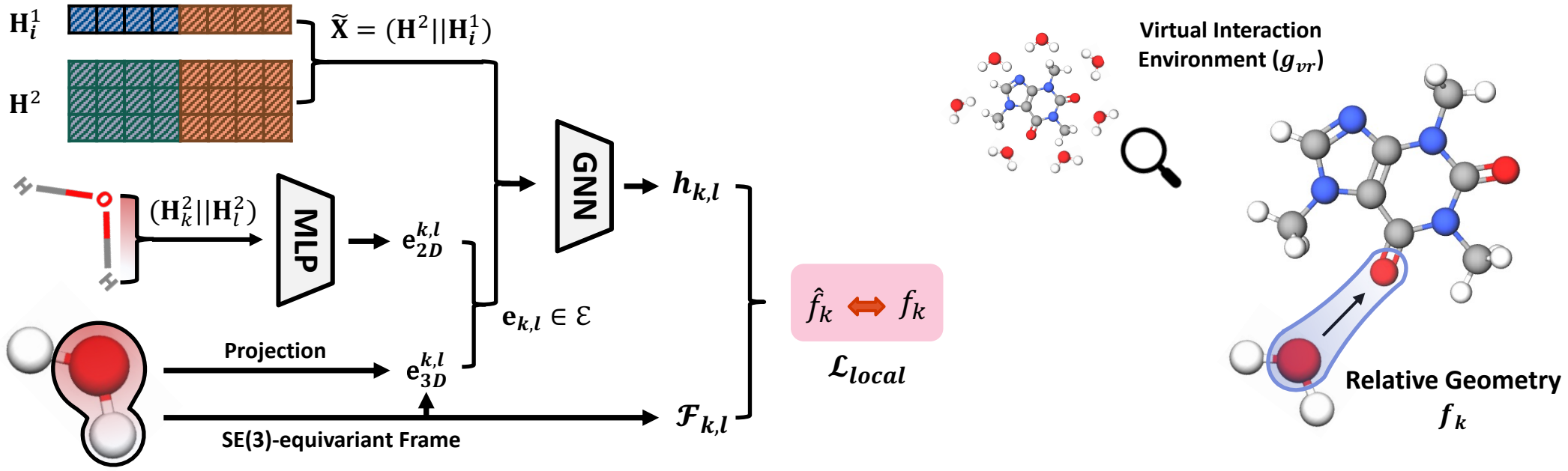
$$\mathcal{L}_{\text{glob}} = -\frac{1}{N_{\text{batch}}} \sum_{i=1}^{N_{\text{batch}}} \left[\log \frac{e^{\texttt{sim}(\mathbf{z}_{2D,i}, \mathbf{z}_{3D,i})/\tau}}{\sum_{k=1}^{N_{\text{batch}}} e^{\texttt{sim}(\mathbf{z}_{2D,i}, \mathbf{z}_{3D,k})/\tau}} + \log \frac{e^{\texttt{sim}(\mathbf{z}_{3D,i}, \mathbf{z}_{2D,i})/\tau}}{\sum_{k=1}^{N_{\text{batch}}} e^{\texttt{sim}(\mathbf{z}_{3D,i}, \mathbf{z}_{2D,k})/\tau}} \right]$$

METHODOLOGY PRE-TRAINING FOR 3D INTERACTION GEOMETRY

SE(3) Equivariant Local Geometry

Fine-grained relative geometry between large and small molecules

→ Training 2D molecule encoder to predict geometry of the larger molecule

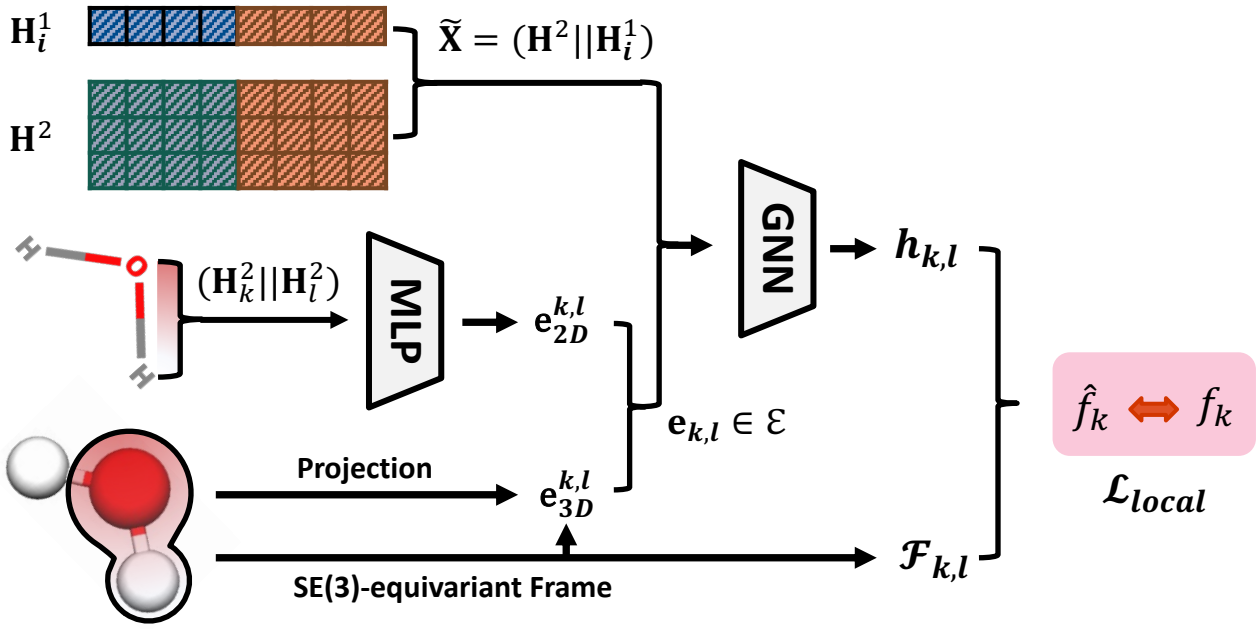


METHODOLOGY PRE-TRAINING FOR 3D INTERACTION GEOMETRY

SE(3) Equivariant Local Geometry

Fine-grained relative geometry between large and small molecules

→ Training 2D molecule encoder to predict geometry of the larger molecule



Local Frame for Equivariance

Orthogonal frame between atoms k and l in small molecule

$$\mathcal{F}_{k,l} = \left(\frac{\mathbf{r}_k - \mathbf{r}_l}{||\mathbf{r}_k - \mathbf{r}_l||}, \frac{\mathbf{r}_k \times \mathbf{r}_l}{||\mathbf{r}_k \times \mathbf{r}_l||}, \frac{\mathbf{r}_k - \mathbf{r}_l}{||\mathbf{r}_k - \mathbf{r}_l||} \times \frac{\mathbf{r}_k \times \mathbf{r}_l}{||\mathbf{r}_k \times \mathbf{r}_l||} \right)$$

$$\mathbf{e}_{3\mathrm{D}}^{k,l} = \mathrm{Projection}_{\mathcal{F}_{k,l}}(\mathbf{r}_k, \mathbf{r}_l) \in \mathbb{R}^d$$

Projection to invariant feature

$$\mathbf{e}_{2\mathrm{D}}^{k,l} = \mathrm{MLP}(\mathbf{H}_k^2||\mathbf{H}_l^2) \in \mathbb{R}^d$$

$$\mathbf{h}_{k,l} = \mathrm{GNN}(\tilde{\mathbf{X}}, \mathcal{E})$$

$$\hat{f}_k = \sum_l \mathbf{h}_{k,l} \odot \mathcal{F}_{k,l}$$

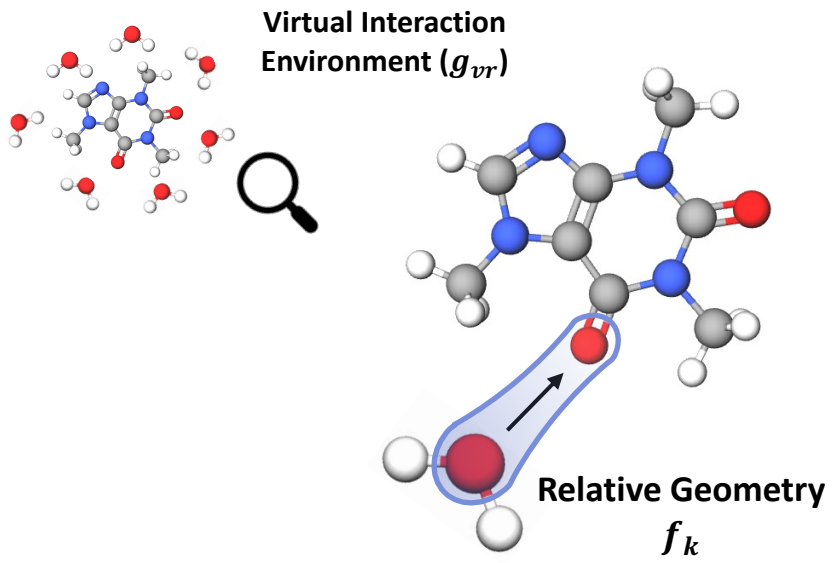
Restore to equivariant feature

METHODOLOGY PRE-TRAINING FOR 3D INTERACTION GEOMETRY

SE(3) Equivariant Local Geometry

Fine-grained relative geometry between large and small molecules

→ Training 2D molecule encoder to predict geometry of the larger molecule



Local relative geometry as Pseudo Force

Since solvent atoms are placed near specific solute atoms,
the dominant interaction direction aligns with the interatomic vector

→ Reasonable proxy for the net force axis

EXPERIMENTS

Dataset		$\mathcal{G}^1$	$\mathcal{G}^2$	# $\mathcal{G}^1$	# $\mathcal{G}^2$	# Pairs	Task
Chromophore [1]	Absorption	Chrom.	Solvent	6416	725	17276	reg.
	Emission	Chrom.	Solvent	6412	1021	18141	reg.
	Lifetime	Chrom.	Solvent	2755	247	6960	reg.
MNSol [2]		Solute	Solvent	372	86	2275	reg.
FreeSolv [3]		Solute	Solvent	560	1	560	reg.
CompSol [4]		Solute	Solvent	442	259	3548	reg.
Abraham [5]		Solute	Solvent	1038	122	6091	reg.
CombiSolv [6]		Solute	Solvent	1495	326	10145	reg.
ZhangDDI [7]		Drug	Drug	544	544	40255	cls.
ChChMiner [8]		Drug	Drug	949	949	21082	cls.

Dataset statistics

Chromophore dataset

- Absorption max
- Emission max
- Lifetime

Solvation Free Energy dataset

- MNSol
- FreeSolv
- CompSol
- Abraham
- CombiSolv

Drug-Drug Interaction dataset

- ZhangDDI
- ChChMiner

EXPERIMENTS

Model	Chromophore			MNSol	FreeSolv	CompSol	Abraham	CombiSolv
	Absorption	Emission	Lifetime					
MPNN	22.00 (0.30)	26.34 (0.41)	0.789 (0.021)	0.643 (0.005)	1.127 (0.110)	0.420 (0.018)	0.640 (0.008)	0.614 (0.031)
+ 3DMRL	19.96 (0.12)	25.21 (0.31)	0.753 (0.018)	0.609 (0.008)	1.068 (0.087)	0.377 (0.020)	0.550 (0.051)	0.599 (0.025)
Improvement	9.27%	4.29%	4.56%	5.28%	5.24%	10.24%	14.06%	2.44%
AttentiveFP	22.86 (0.30)	28.70 (0.23)	0.871 (0.010)	0.570 (0.021)	1.019 (0.070)	0.350 (0.008)	0.426 (0.042)	0.471 (0.028)
+ 3DMRL	22.80 (0.61)	28.54 (1.97)	0.784 (0.013)	0.562 (0.031)	0.901 (0.059)	0.271 (0.009)	0.378 (0.027)	0.448 (0.011)
Improvement	0.26%	0.55%	9.99%	1.40%	11.57%	22.57%	11.26%	4.88%
CIGIN	19.66 (0.69)	25.84 (0.23)	0.821 (0.017)	0.582 (0.022)	0.958 (0.116)	0.369 (0.018)	0.421 (0.018)	0.464 (0.002)
+ 3DMRL	18.00 (0.17)	24.21 (0.09)	0.729 (0.014)	0.528 (0.019)	0.839 (0.105)	0.277 (0.006)	0.371 (0.031)	0.435 (0.006)
Improvement	8.44%	6.30%	11.20%	9.28%	12.42%	24.93%	11.87%	6.25%
CGIB	18.37 (0.35)	24.52 (0.25)	0.808 (0.015)	0.562 (0.008)	0.876 (0.037)	0.321 (0.002)	0.404 (0.037)	0.448 (0.008)
+ 3DMRL	17.93 (0.35)	23.92 (0.29)	0.733 (0.009)	0.538 (0.020)	0.842 (0.078)	0.274 (0.002)	0.370 (0.027)	0.442 (0.015)
Improvement	2.40%	5.90%	9.28%	4.27%	3.88%	14.64%	8.42%	1.33%
$CGIB_{Cont}$	18.59 (0.24)	24.68 (0.49)	0.803 (0.019)	0.561 (0.012)	0.897 (0.098)	0.333 (0.005)	0.404 (0.039)	0.452 (0.015)
+ 3DMRL	17.90 (0.17)**	23.94 (0.24)	0.720 (0.020)	0.524 (0.018)*	0.863 (0.075)	0.284 (0.007)	0.372 (0.021)	0.441 (0.022)
Improvement	3.71%	3.00%	10.33%	6.59%	3.79%	14.71%	7.92%	2.43%

Observations

- 3DMRL obtains consistent improvements over the base GNNs in all 40 tasks

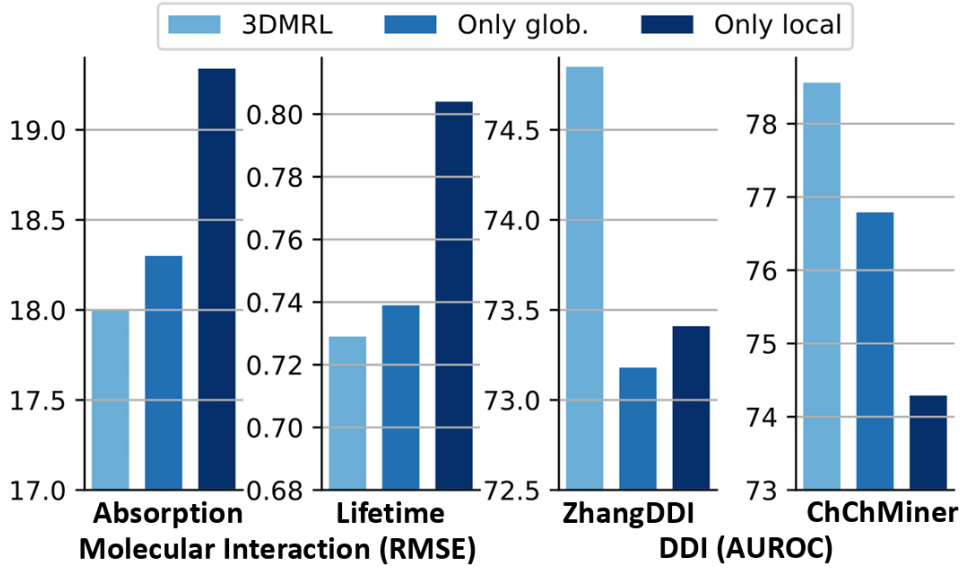
EXPERIMENTS

Strategy	(a) Molecular Interaction Tasks (RMSE ↓)							
	Chromophore			MNSol	FreeSolv	CompSol	Abraham	CombiSolv
	Absorption	Emission	Lifetime					
No Pre-training	19.66 (0.69)	25.84 (0.23)	0.821 (0.017)	0.567 (0.014)	0.884 (0.074)	0.331 (0.029)	0.412 (0.028)	0.458 (0.002)
MPP (molecular property prediction) Pre-training								
3D Infomax	18.71 (0.61)	24.59 (0.22)	0.790 (0.022)	0.585 (0.015)	0.873 (0.103)	0.321 (0.041)	0.426 (0.036)	0.464 (0.004)
GraphMVP	18.40 (0.62)	24.73 (0.14)	0.797 (0.022)	0.561 (0.025)	1.010 (0.115)	0.301 (0.025)	0.418 (0.020)	0.437 (0.015)
MoleculeSDE	18.56 (0.24)	24.91 (0.10)	0.836 (0.040)	0.564 (0.018)	0.971 (0.122)	0.308 (0.024)	0.426 (0.028)	0.454 (0.012)
MRL (molecular relational learning) Pre-training								
3DMRL	**18.00** (0.17)	**24.21** (0.09)	**0.729** (0.014)	**0.528** (0.019)	**0.839** (0.105)	**0.277** (0.006)	**0.371** (0.031)	**0.435** (0.006)

Observations

- Single molecule pre-training methods did not yield satisfactory results in molecular relational learning
- 3DMRL consistently delivers significant performance improvements

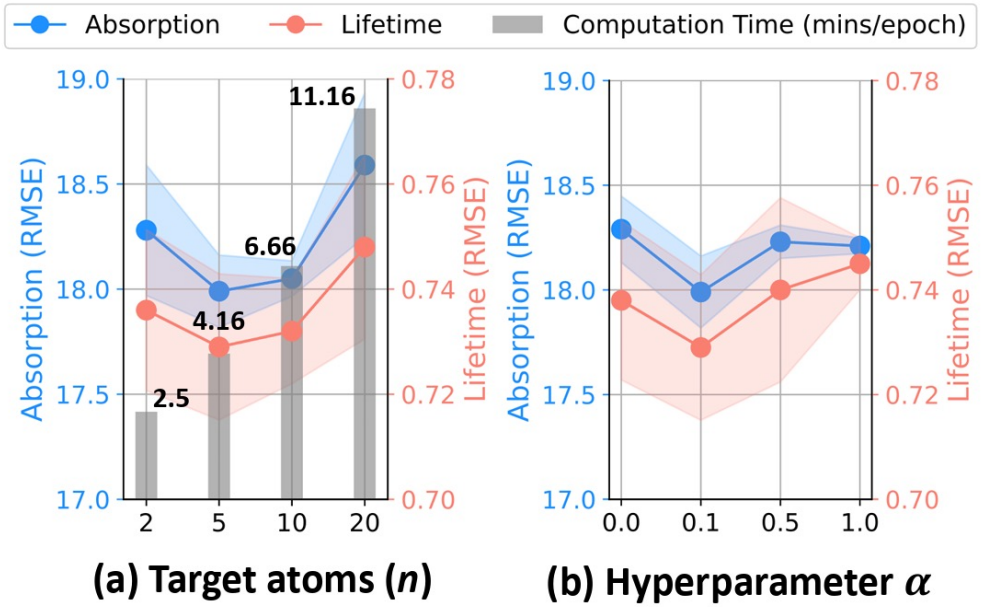
EXPERIMENTS



Ablation Studies

- Learning the global geometry plays a critical role
- Combining both losses (3DMRL) yields the best results

EXPERIMENTS



(a) Target atoms (n)

(b) Hyperparameter α

Sensitivity Analysis (a)

Selecting appropriate number of target atoms is crucial for both model performance and computational efficiency

Sensitivity Analysis (b)

Selecting appropriate weight of local geometry loss is crucial

THANK YOU!

[Full Paper] https://openreview.net/forum?id=PZaxCfLGLA

[Source Code] https://github.com/Namkyeong/3DMRL

[Author Email] namkyeong96@kaist.ac.kr