August 3-7, 2025
KDD2025

# Advancing Molecular Graph-Text Pre-training via Fine-grained Alignment

**Yibo Li**[1], Yuan Fang[2]*, Mengmei Zhang[3], Chuan Shi[1]*,

[1]Beijing University of Posts and Telecommunications

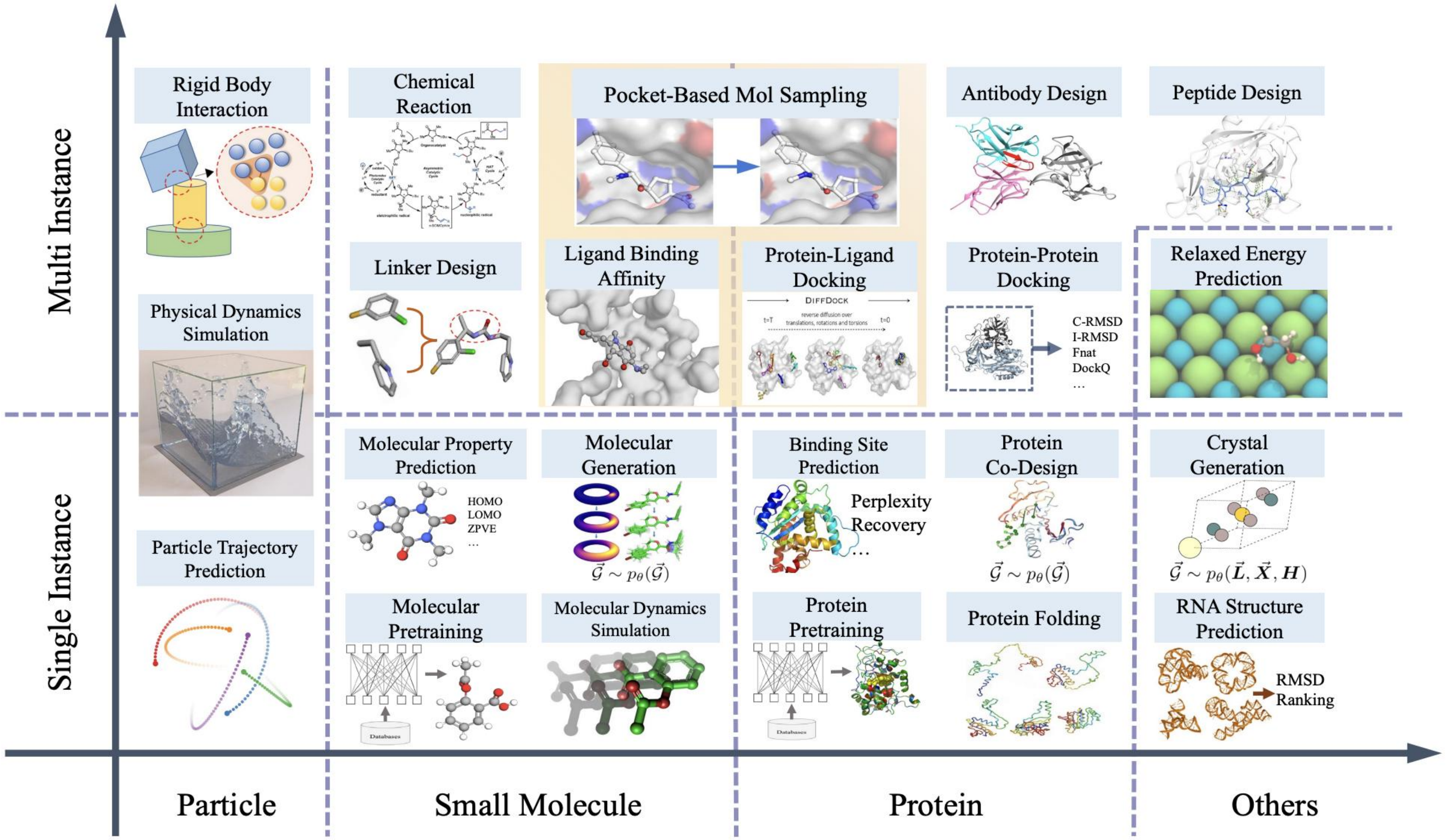
[2]Singapore Management University

[3]China Telecom Bestpay

CONTENTS

## Molecular Graph-Text Pre-training

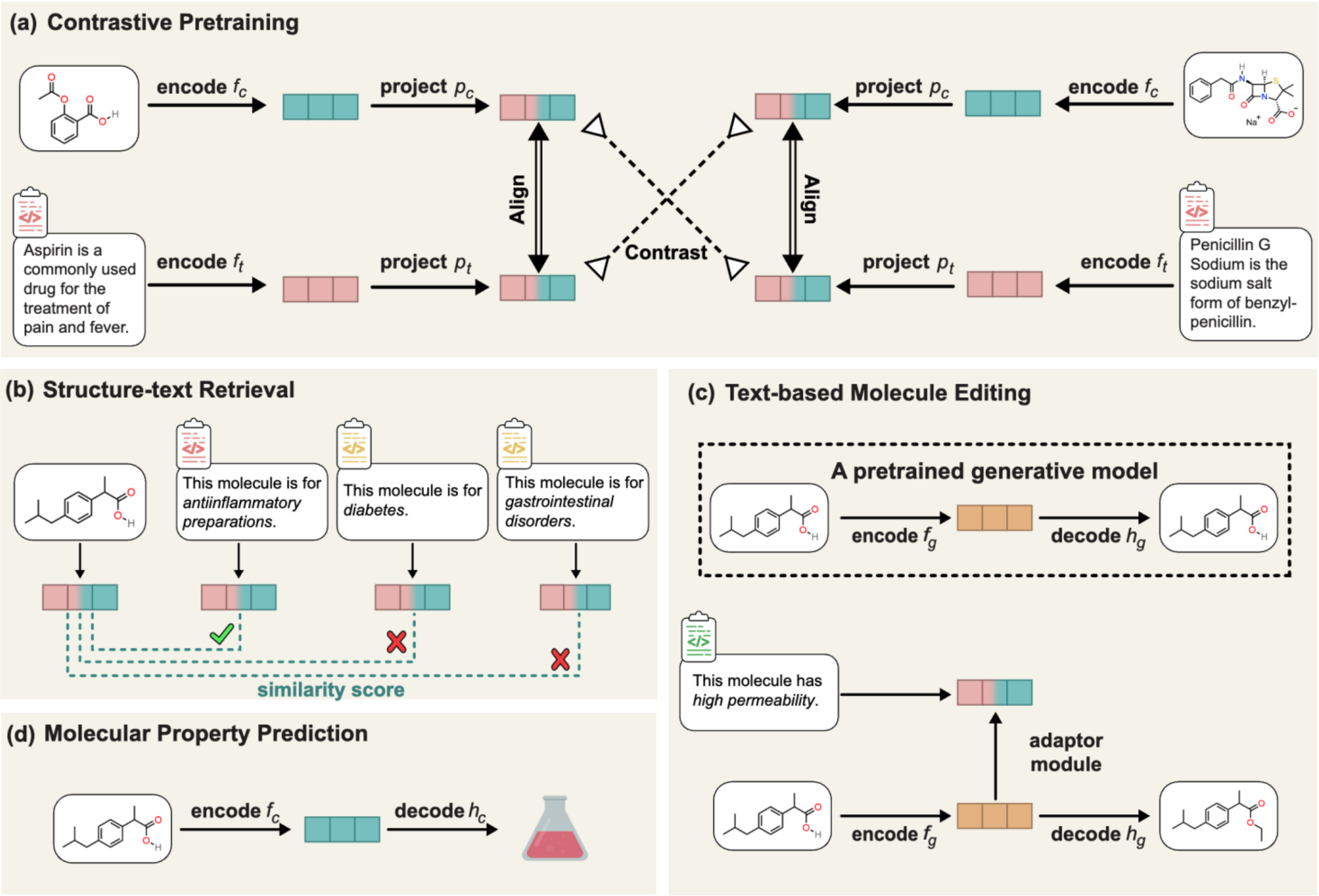
- **Comprehending molecular structure and related knowledge is pivotal in scientific investigations spanning diverse fields**




[1] A survey of geometric graph neural networks: data structures, models and applications

# Molecular Graph-Text Pre-training

- **Several studies explore molecular structures along with their corresponding descriptions.**

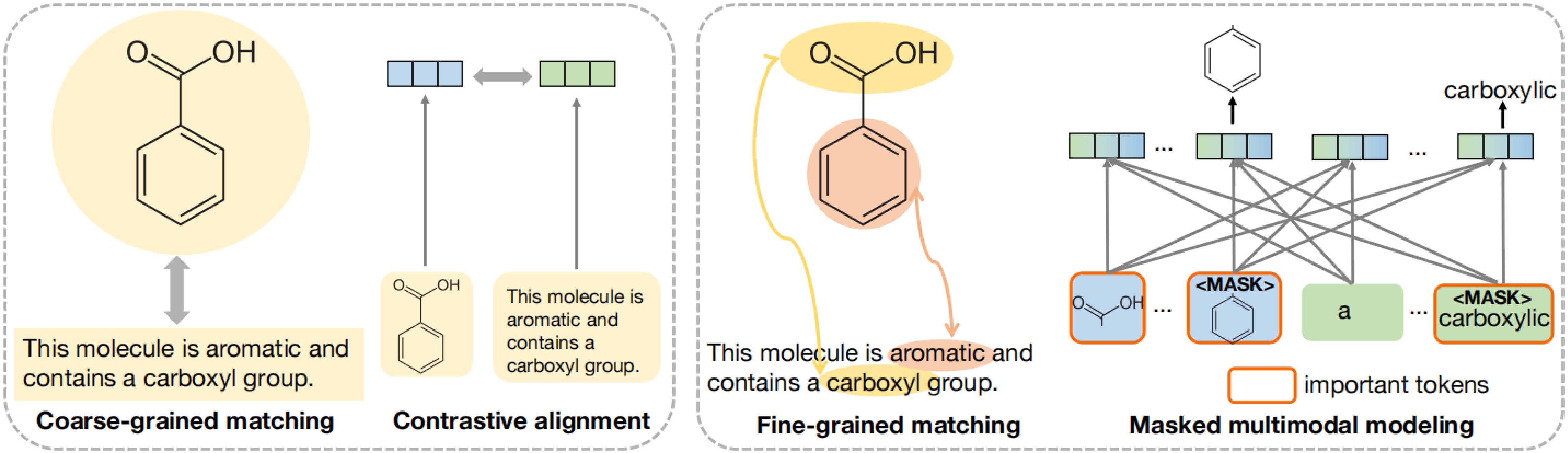



[1] Multi-modal Molecule Structure-text Model for Text-based Retrieval and Editing

CONTENTS

2

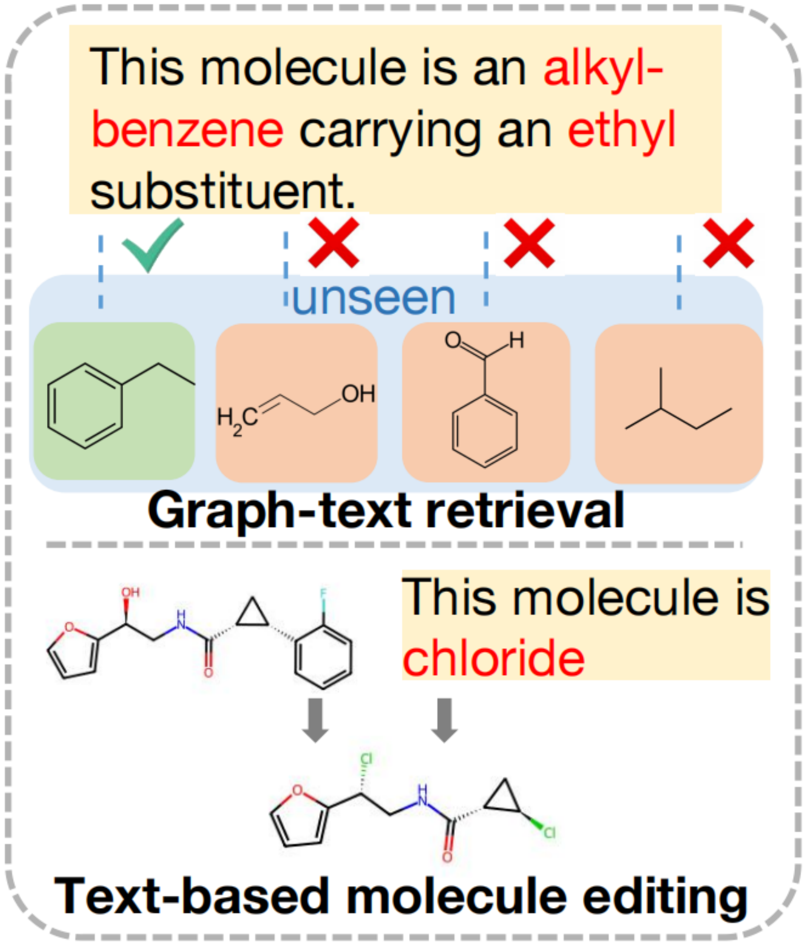
# Coarse-grained Alignment & Fine-grained Alignment




(a) Molecule-level learning

(b) Motif-level learning

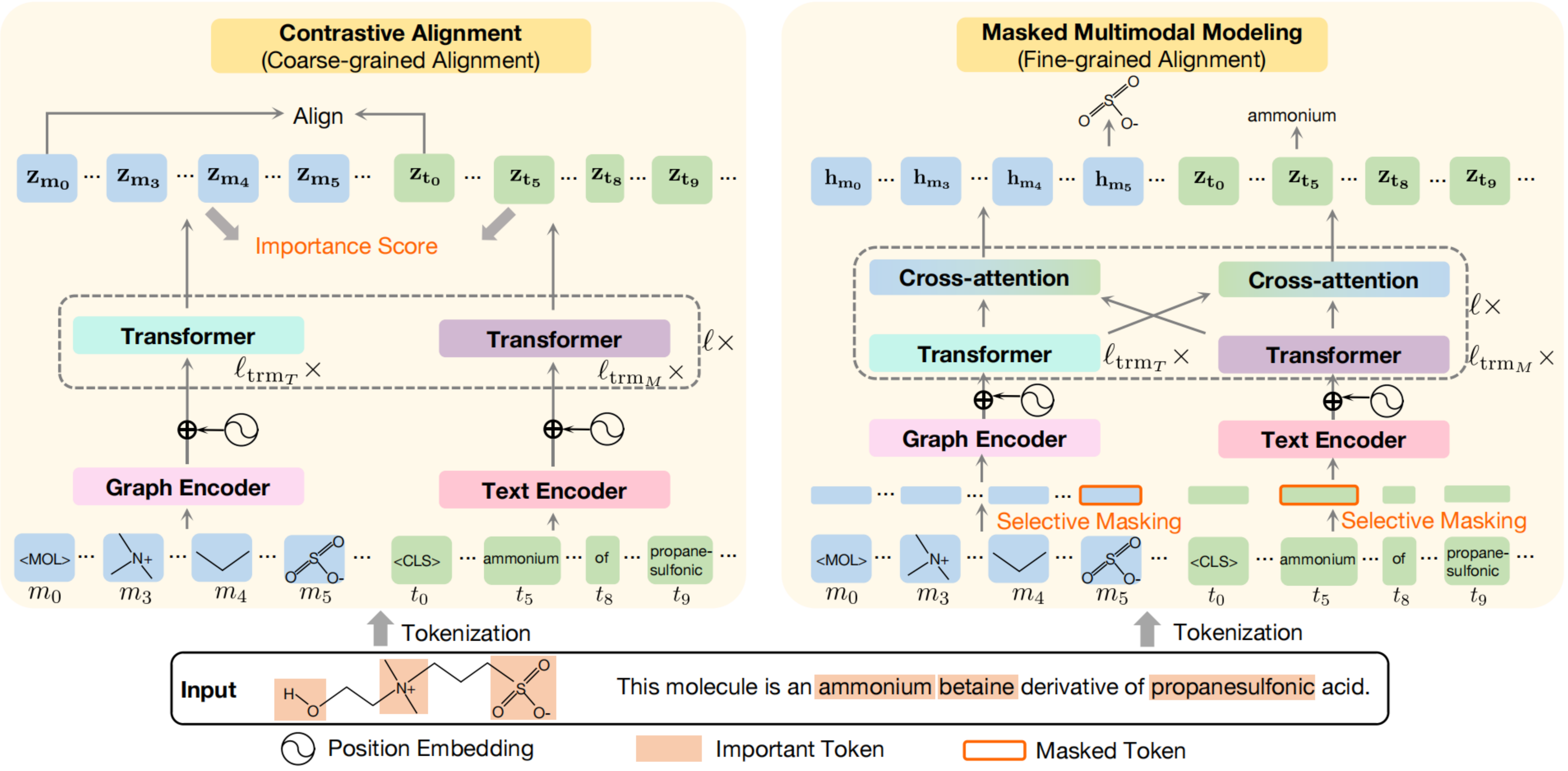
# 2 Fine-grained Alignment Is Crucial



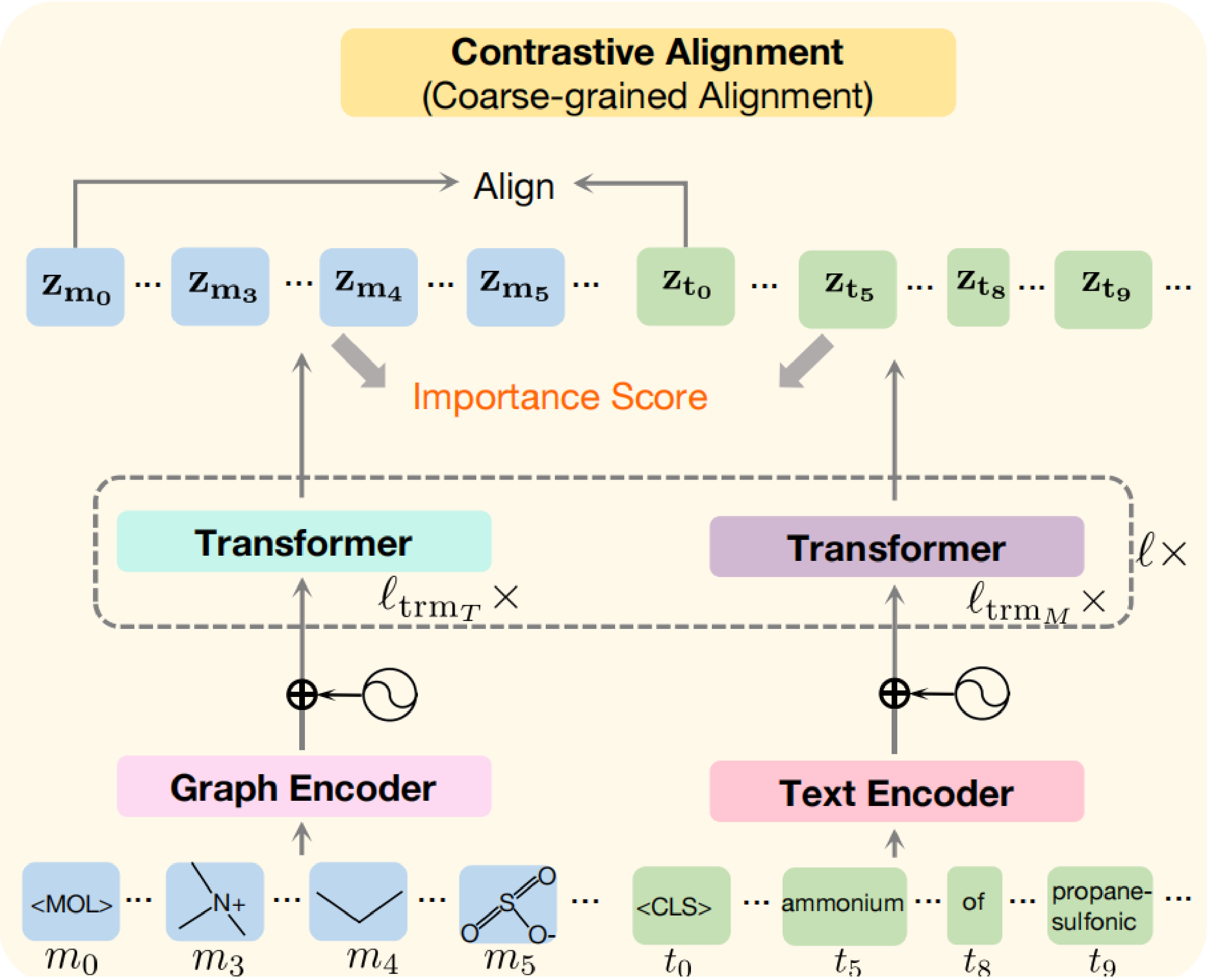

(c) Downstream tasks

# CONTENTS

# 3 FineMolTex Framework

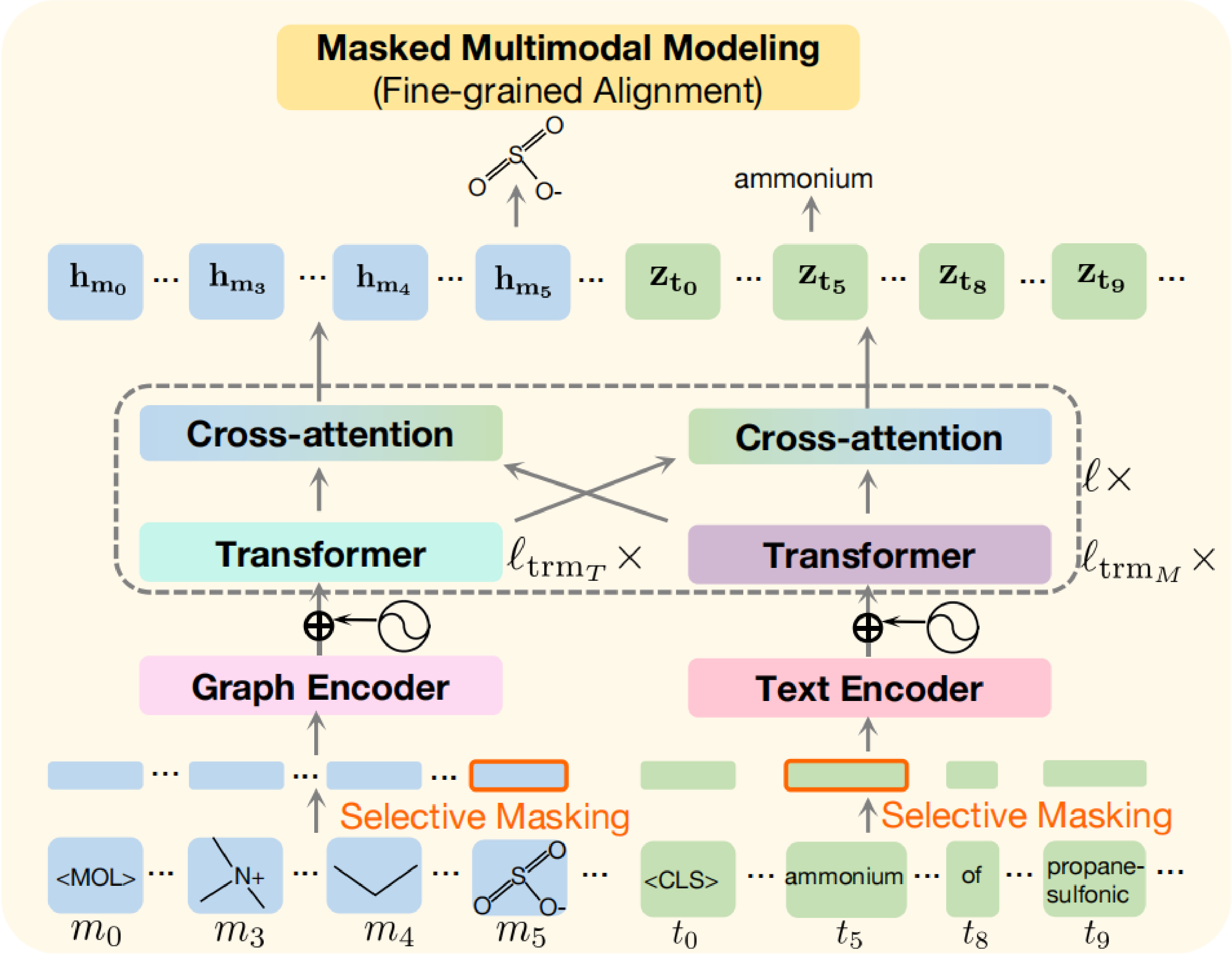


# 3 FineMolTex Framework



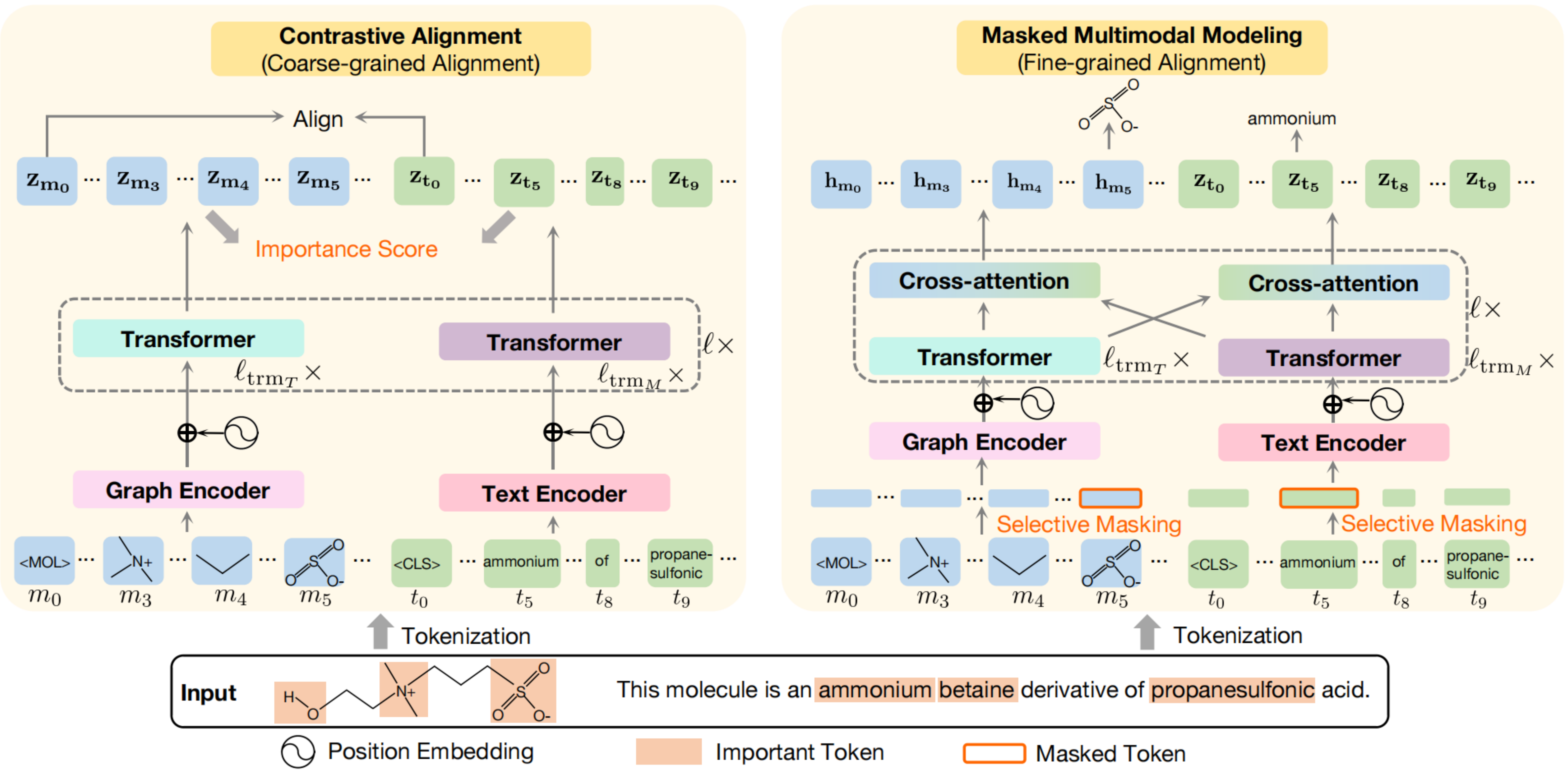

$$L_{\mathrm{con}} = -\frac{1}{2}\mathbb{E}_{m_0,t_0}\left[\log\frac{\exp(\cos(\mathbf{z}_{m_0},\mathbf{z}_{t_0})/\tau)}{\exp(\cos(\mathbf{z}_{m_0},\mathbf{z}_{t_0})/\tau)+\sum_{t_0'}\exp(\cos(\mathbf{z}_{m_0},\mathbf{z}_{t_0'})/\tau)} + \log\frac{\exp(\cos(\mathbf{z}_{t_0},\mathbf{z}_{m_0})/\tau)}{\exp(\cos(\mathbf{z}_{t_0},\mathbf{z}_{m_0})/\tau)+\sum_{m_0'}\exp(\cos(\mathbf{z}_{t_0},\mathbf{z}_{m_0'})/\tau)}\right],$$

## Framework




$$L_{\mathrm{pre}} = \beta \sum_i \mathrm{CE}(\hat{y}_{m_i}, y_{m_i}) + \alpha \sum_j \mathrm{CE}(\hat{y}_{t_j}, y_{t_j}).$$

# Framework



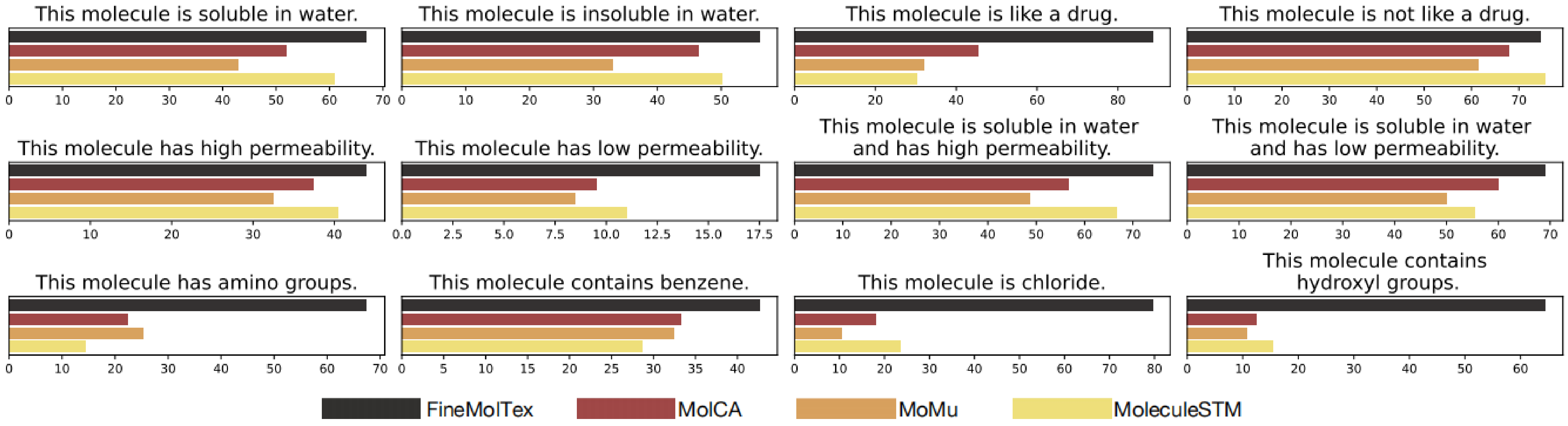
Contrastive Alignment
(Coarse-grained Alignment)
Align
Importance Score
Transformer
Transformer
Graph Encoder
Text Encoder
<MOL>
<CLS>
ammonium
of
propane-sulfonic
Masked Multimodal Modeling
(Fine-grained Alignment)
ammonium
Cross-attention
Cross-attention
Selective Masking
Selective Masking
Tokenization
Tokenization
Input
This molecule is an ammonium betaine derivative of propanesulfonic acid.
Position Embedding
Important Token
Masked Token

CONTENTS

## Can FineMolTex better generalize to unseen molecules?

| $T$ | Given Molecular Graph | | | Given Text | | |
|---|---|---|---|---|---|---|
| | 4 | 10 | 20 | 4 | 10 | 20 |
| KV-PLM | 68.38±0.03 | 47.59±0.03 | 36.54±0.03 | 67.68±0.03 | 48.00±0.02 | 34.66±0.02 |
| MolCA | 83.75±0.54 | 74.25±0.26 | 66.14±0.21 | 81.27±0.33 | 69.46±0.17 | 62.13±0.16 |
| MoMu-S | 70.51±0.04 | 55.20±0.15 | 43.78±0.10 | 70.71±0.22 | 54.70±0.31 | 44.25±0.43 |
| MoMu-K | 69.40±0.11 | 53.14±0.26 | 42.32±0.28 | 68.71±0.03 | 53.29±0.05 | 43.83±0.12 |
| 3D-MoLM | 81.35±0.14 | 73.65±0.13 | 64.79±0.15 | 79.78±0.22 | 62.38±0.16 | 53.43±0.11 |
| MV-Mol | 92.24±0.26 | 85.38±0.19 | 79.41±0.43 | 91.28±0.13 | 85.32±0.15 | 80.37±0.22 |
| MoleculeSTM | 92.14±0.02 | 86.27±0.02 | 81.08±0.05 | 91.44±0.02 | 86.76±0.03 | 81.68±0.03 |
| FineMolTex | **96.78±0.05** | **92.48±0.02** | **87.94±0.14** | **96.29±0.12** | **91.65±0.15** | **85.07±0.11** |

## Can FineMolTex bridge the gap to tasks centered on motif-level knowledge?



## Can FineMolTex perform better on single-modality tasks?

| Model | BBBP | Tox21 | ToxCast | Sider | ClinTox | MUV | HIV | Bace | Avg |
|---|---|---|---|---|---|---|---|---|---|
| AttrMask | 67.8±2.6 | 75.0±0.2 | 63.6±0.8 | 58.1±1.2 | 75.4±8.8 | 73.8±1.2 | 75.4±0.5 | 80.3±0.0 | 71.2 |
| ContextPred | 63.1±3.5 | 74.3±0.2 | 61.6±0.5 | 60.3±0.8 | 80.3±3.8 | 71.4±1.4 | 70.7±3.6 | 78.8±0.4 | 70.1 |
| InfoGraph | 64.8±0.6 | 76.2±0.4 | 62.7±0.7 | 59.1±0.6 | 76.5±7.8 | 73.0±3.6 | 70.2±2.4 | 77.6±2.0 | 70.0 |
| MolCLR | 67.8±0.5 | 67.8±0.5 | 64.6±0.1 | 58.7±0.1 | 84.2±1.5 | 72.8±0.7 | 75.9±0.2 | 71.1±1.2 | 71.3 |
| GraphMVP | 68.1±1.4 | 77.1±0.4 | 65.1±0.3 | 60.6±0.1 | 84.7±3.1 | 74.4±2.0 | 77.7±2.5 | 80.5±2.7 | 73.5 |
| GraphCL | 69.7±0.7 | 73.9±0.7 | 62.4±0.6 | 60.5±0.9 | 76.0±2.7 | 69.8±2.7 | 78.5±1.2 | 75.4±1.4 | 70.8 |
| KV-PLM | 70.5±0.5 | 72.1±1.0 | 55.0±1.7 | 59.8±0.6 | 89.2±2.7 | 54.6±4.8 | 65.4±1.7 | 78.5±2.7 | 68.2 |
| MoMu-S | 70.5±2.0 | 75.6±0.3 | 63.4±0.5 | 60.5±0.9 | 79.9±4.1 | 70.5±1.4 | 75.9±0.8 | 76.7±2.1 | 71.6 |
| MoMu-K | 70.1±1.4 | 75.6±0.5 | 63.0±0.4 | 60.4±0.8 | 77.4±4.1 | 71.1±2.7 | 76.2±0.9 | 77.1±1.4 | 71.4 |
| MolCA | 70.0±0.5 | **77.2±0.5** | 64.5±0.8 | 63.0±1.7 | 89.5±0.7 | 72.1±1.3 | 77.2±0.6 | 79.8±0.5 | 74.2 |
| MoleculeSTM | 70.0±0.5 | 76.9±0.5 | 65.1±0.4 | 61.0±1.1 | 92.5±1.1 | 73.4±2.9 | 77.0±1.8 | 80.8±1.3 | 74.6 |
| FineMolTex | **73.5±1.6** | 77.1±1.2 | **68.6±0.9** | **64.8±1.4** | **92.5±0.8** | **76.3±1.2** | **79.0±1.4** | **84.0±1.5** | **76.9** |

## Has FineMolTex learned fine-grained knowledge?

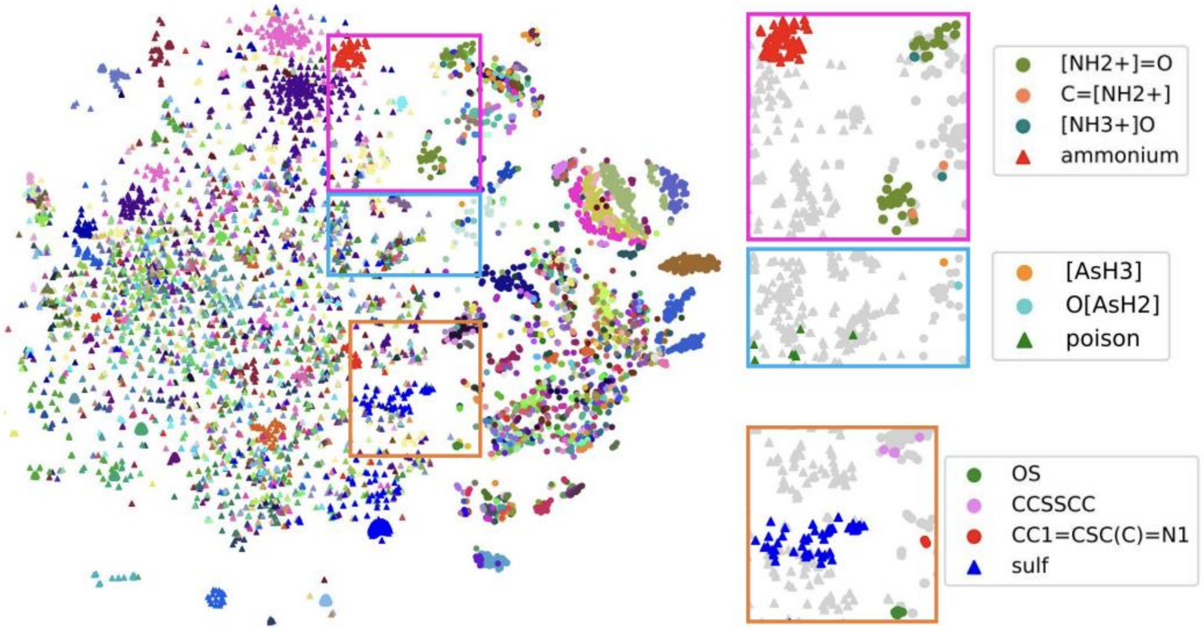



Figure 1: Visualization of motif tokens and word tokens using $t$-SNE. Triangles denote word tokens; circles denote motif tokens.

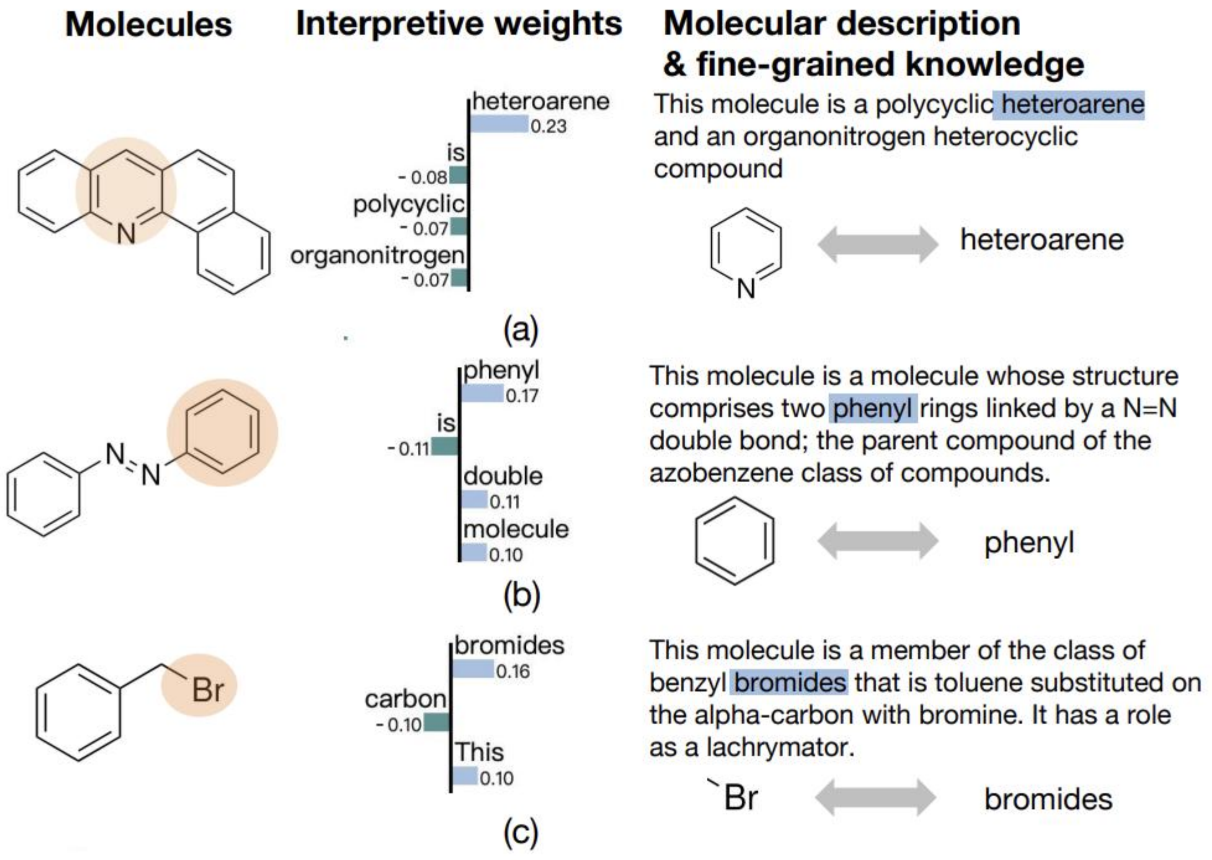



Figure 2: Explaination of the prediction of certain masked motifs based on text tokens utilizing LIME.

CONTENTS

# 5 Conclusion

- We reveal that fine-grained motif-level knowledge is crucial for molecular representation learning.
- We propose FineMolTex to jointly learn both coarse- and fine-grained knowledge through a contrastive alignment task and a masked multimodal learning task, respectively.
- By selectively masking the important motif/word tokens and predicting their labels using tokens from the other modality, we can effectively learn fine-grained alignment between motifs and words.
- Experimental results on three downstream tasks and two case studies demonstrate the effectiveness of FineMolTex.

# Thank you !

Email: liyibo@u.nus.edu
WeChat: liushi32992