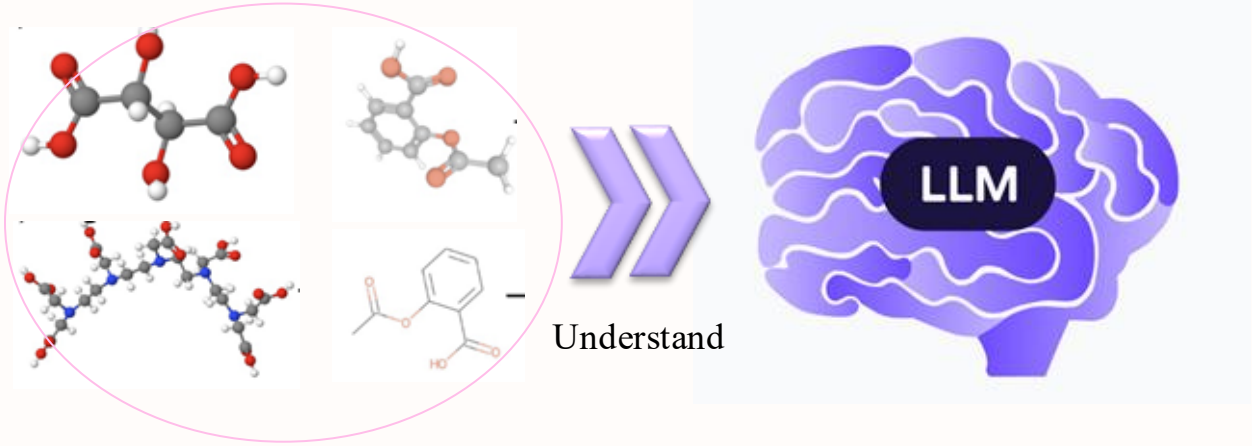
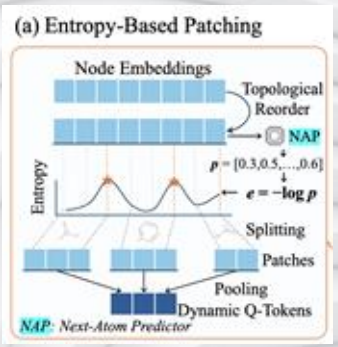
Entropy-Guided Dynamic Tokens for Graph-LLM Alignment in Molecular Understanding

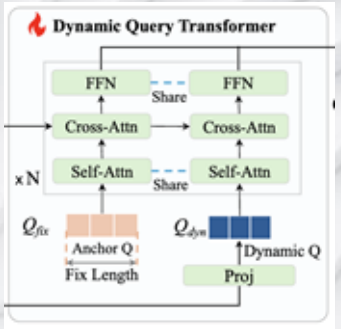
– A bridge for LLMs to understand molecular graphs.



Novelty No.1



Novelty No.2



Zihao Jing

Date 2026 Mar 6

⇨ Problem Statement

Hard to Understand

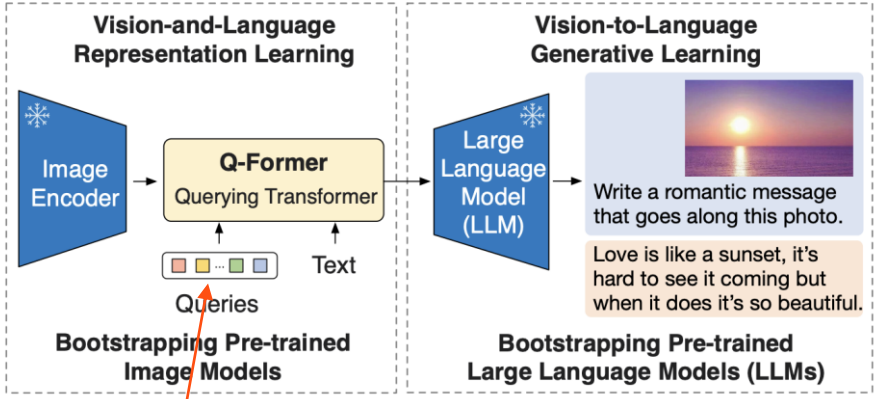
As they are graphs or coordinates

L L M

SMILES

c1ccccc1

Easy to understand as text



Novelty of Q-Former:

1. LLM can understand images from now on;

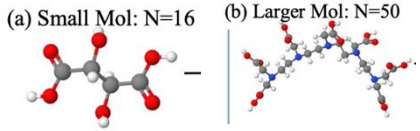
2. The image encoder and LLM kept frozen during training, which save a lot of computing resources

Some researchers found the idea from vision-language model, from **Q-Former** (BLIP-2, the milestone for LLM understand images)

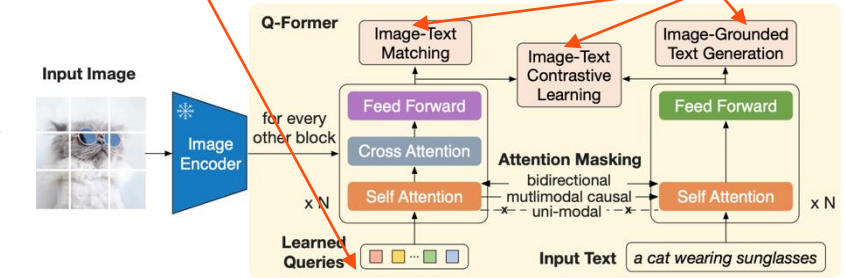
Learnable Modality Anchor

Pre-Training Tasks

The image can usually be split into fixed-number of patches:



However, graphs that have vary length can not fit this very well.

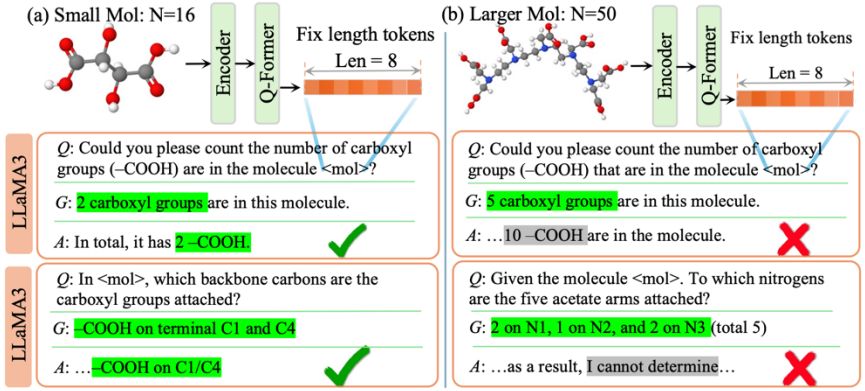


Output will be concatenated with text inputs for LLM inference

These learnable (when training) anchor tokens will capture the feature of the image through cross attention with image embeddings

⇨ Current Challenges on LLM for Molecular Graph

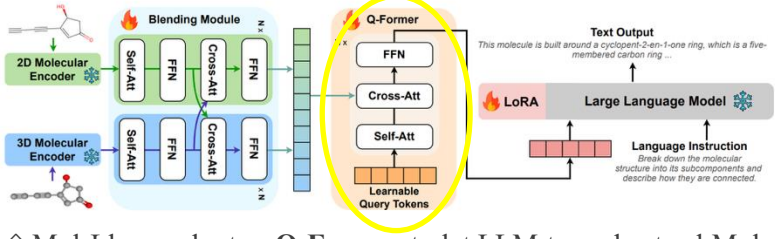
1. Loss of Structure



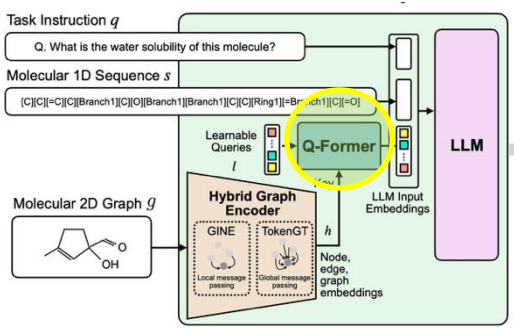
2. Heavy Fine-tuning

Table 1: Computing costs comparison between frozen/unfrozen backbone settings of Mol-Llama.

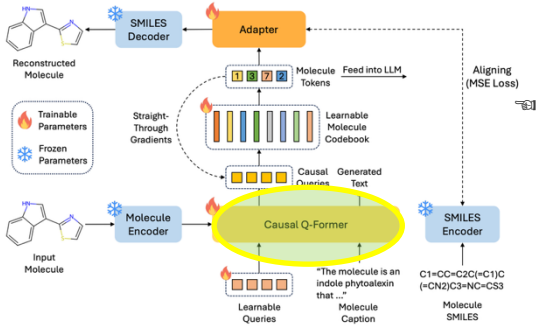
Connector	LLM	Trainable	FLOPs/Token	Time/Step
Finetune	Finetune	8.1B	4.9e10	0.93
Finetune	Frozen	84M	1.7e10	0.23



⇧ Mol-Llama adapts a **Q-Former** to let LLM to understand Molecular 2D&3D, has gained great performance on property, mol captioning tasks.



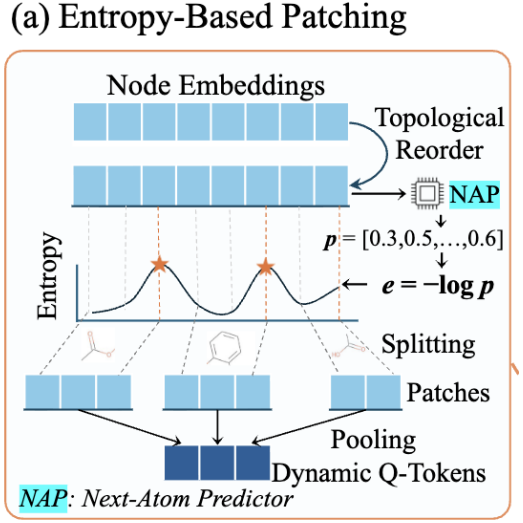
☜ Mol-LLM adapts a **Q-Former** as well, using which to process the output from a hybrid graph encoder, and concatenate with instruction and SELFISH for LLM



☜ 3DMol-T5 adapts a **Q-Former** as well, using which to process the output from a hybrid graph encoder, and concatenate with instruction and SELFISH for LLM

➪Novelty 1 — Entropy Guided Patching

To collaborate with dynamic tokens, capture the substructure features to overcome the ‘loss of structure’ issue



Logits of the correct atom, from the **Next-Atom Prediction** model

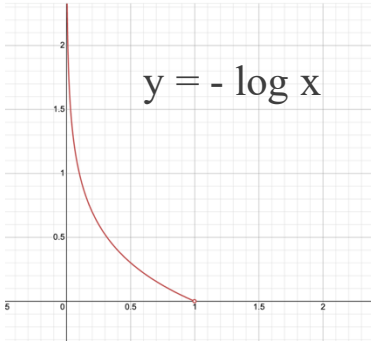
Next Atom

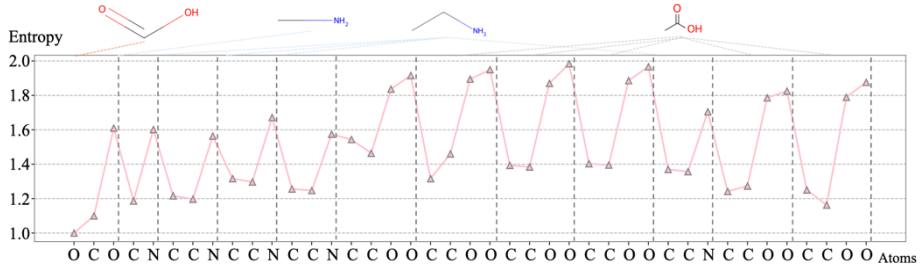
$$p_t = \text{softmax}(L_t)[a_{t+1}], \qquad t = 1, \ldots, T-1.$$

Probability of the correct next atom

$$e_t = -\log p_t$$

Entropy

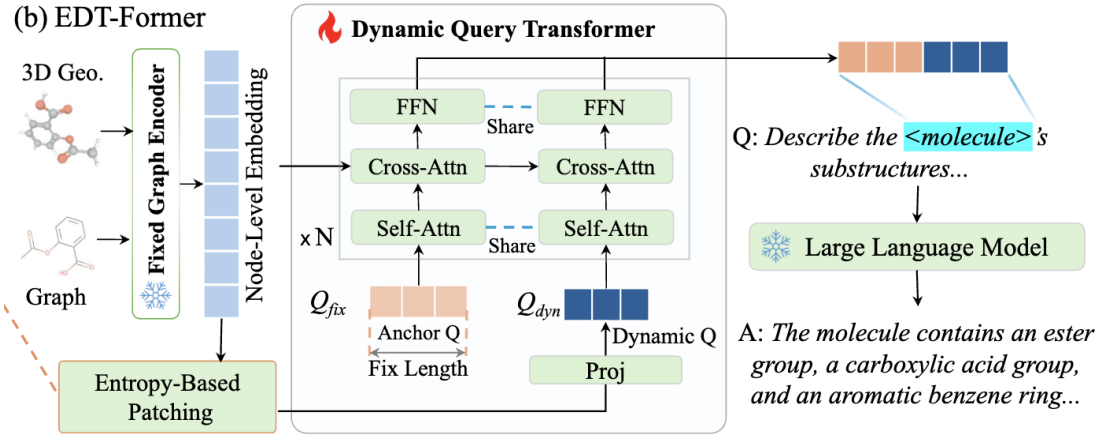




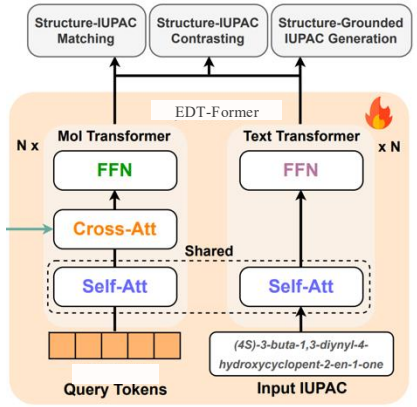
We split patches at each entropy peak. That’s the position that Transformer model is hard to predict next atom.

➪Novelty 2 — Dynamic Query Former

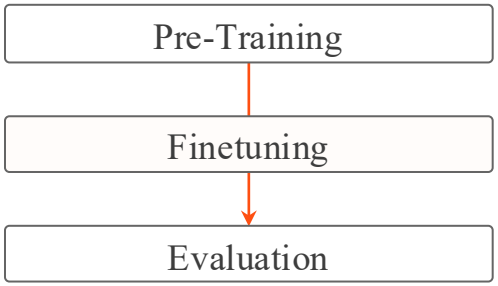
To overcome the loss of structure issue caused by fixed length token



Structure of EDT-Former



Pre-Training Tasks of EDT-Former



Take the original training tasks from Q-Former using the Mol-Llama-Instruct dataset. (Graph encoder + Dynamic Query Former)

End-to-end finetuning (joint with LLM) on conversation datasets.

Inference and Evaluation

Table 13: Fine-tuning dataset. Four types of instruction data from Mol-Llama-Instruct, the data types and amounts are listed.

Category	Amount
Detailed Structural Descriptions	77,239
Structure-to-Chemical Features	73,712
Structure-to-Biological Features	73,645
Comprehensive Conversations	60,147

➪ Benchmarks

Baselines: General LLM, Molecular LLM, Large-scale LLM

Baselines: General LLM, Molecular LLM, Large-scale LLM

Datasets: MoleculeQA, TDC, Mol-Instructions

Efficiency

Table 7: Estimated memory usage and training time per step for EDT-Former with Llama3.1-8B.

EDT-Former	Llama3.1-8B	Mem (GB)	Time/Step (s)
Train	Frozen	37	0.26
Train	LoRA	77	0.93
Train	Train	>200	–

MoleculeQA – Multichoice Questions

Table 3: Performance on the MoleculeQA benchmark. Models are compared across four tasks (Structure, Source, Property, Application) with accuracy (%) reported. The best (pink) and second-best (lightpink) results are highlighted. Modality T/G/3D represents text, graph, and 3D geometry.

Model	Modality	Imp.	Size	Strct.	Src.	Prop.	App.	Avg.	Total
Random									
Random	-	-	-	24.41	22.30	23.04	24.57	23.58	24.03
Molecular LLMs									
BioMedGPT-LM	Text	SFT	7B	54.19	60.01	38.85	40.90	48.49	52.23
Mol-Inst.-Llama3.1	T/G	SFT	8B	72.79	70.82	43.08	41.22	56.98	65.31
3D-MoLM	T/G/3D	SFT	8B	73.17	70.50	44.79	44.19	58.16	65.96
LLaMo	Text	SFT	8B	70.56	66.63	44.60	45.18	56.74	63.74
Mol-Llama3.1	Text/Graph/3D	SFT	8.2B	73.16	70.22	45.70	46.18	58.82	66.21
General LLMs									
Galactica	Text	SFT	6.7B	32.35	41.92	31.05	28.21	33.38	33.96
BLOOM	Text	SFT	7.1B	35.01	47.51	31.46	33.56	36.89	37.31
Pythia	Text	SFT	6.9B	42.79	58.90	38.58	39.07	44.84	45.61
Llama-2-chat	Text	SFT	7B	28.75	39.84	31.33	27.71	31.91	31.54
Vicuna-v1.5	Text	SFT	13B	37.01	43.19	30.64	31.55	35.60	37.07
Large Scale General LLMs									
GPT-3.5	Text	10-Shot	-	25.60	37.60	28.04	32.22	30.87	29.29
GPT-4	Text	10-Shot	-	60.94	50.19	35.57	43.91	47.65	53.47
GPT-5	Text	10-Shot	-	62.78	53.22	36.42	46.91	49.83	56.12
Ours									
EDT-Former	T/G/3D	10-Shot	8.3B	66.46	60.98	40.35	36.40	51.05	58.78
EDT-Former	T/G/3D	SFT	8.3B	74.55	72.39	50.71	48.58	61.56	68.34

Mol-Instructions

Table 4: Results on the Mol-Instructions dataset. Models are finetuned and evaluated on two tasks: molecular description generation (BLEU, ROUGE, and METEOR scores) and molecular property prediction (MAE). The best (pink) and second-best (lightpink) results are highlighted.

Models	Molecular Description						Property (MAE)
	BLUE-2	BLUE-4	ROUGE-1	ROUGE-2	ROUGE-L	METEOR	
Alpaca-7B	0.068	0.014	0.178	0.041	0.136	0.107	322.109
Baize-7B	0.064	0.015	0.189	0.053	0.148	0.106	261.343
LLaMA-2-7B	0.059	0.014	0.164	0.066	0.148	0.184	5.553
Vicuna-v1.5-13B	0.052	0.011	0.151	0.055	0.130	0.168	860.051
Galatica-6.7B	0.024	0.008	0.074	0.015	0.063	0.065	0.568
Qwen3-8B	0.098	0.029	0.207	0.050	0.157	0.173	4.737
Mol-Inst.-Llama2-7B	0.217	0.143	0.337	0.196	0.291	0.254	0.013
Mol-Inst.-Llama3.1-8B	0.419	0.361	0.719	0.646	0.709	0.637	15.059
Mol-LLaMA2-7.2B	0.433	0.385	0.711	0.649	0.601	0.601	0.0087
Mol-LLaMA3.1-8.2B	0.445	0.398	0.717	0.656	0.709	0.617	0.0079
MolReasoner-7B	0.438	0.322	0.553	0.366	0.482	0.475	10.323
EDT-Former-8.3B	0.424	0.402	0.726	0.652	0.717	0.631	0.0062

Property – BBBP&Pampa

Table 2: Accuracy (%) on zero-shot molecular property prediction benchmarks (Pampa and BBBP). Models are evaluated with three prompting strategies: Direct, Reasoning, and Rich Instructions (see App. D.3). The best (pink) and second-best (lightpink) results are highlighted.

Models	Size	Pampa				BBBP			
		Direct	Reasoning	RichInst.	Avg.	Direct	Reasoning	RichInst.	Avg.
General LLMs									
GPT-4o	-	48.65	58.23	47.17	51.35	60.82	61.34	64.43	62.20
Llama2	7B	57.14	57.53	84.52	66.40	37.37	51.56	53.09	47.34
Llama3.1	8B	56.51	46.19	63.64	55.45	57.07	51.03	55.15	54.42
Molecular LLMs									
Mol-Inst.-Llama2	7B	49.63	31.16	38.18	39.66	52.58	52.58	51.34	52.17
Mol-LLaMA-2	7.2B	75.68	79.61	67.90	74.40	53.37	52.58	52.58	52.84
Mol-Inst.-Llama3.1	8B	55.91	33.50	70.47	53.29	53.44	55.31	54.91	54.55
3D-MoLM	8B	46.93	50.00	64.86	53.93	49.14	51.65	51.91	50.90
LLaMo	8B	49.25	64.37	48.51	54.04	55.44	55.45	56.91	55.93
Mol-LLaMA3.1	8.2B	63.55	64.37	72.48	66.80	59.54	55.56	59.08	58.06
Ours									
EDT-Former	8.3B	81.57	81.57	83.78	82.31	74.44	74.69	75.86	75.00

➪ Ablation and Insights

Ablation on Each Component:

Shows each module has its contribution.

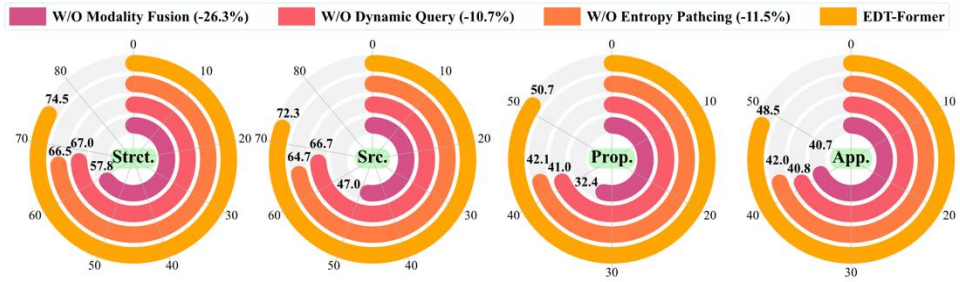
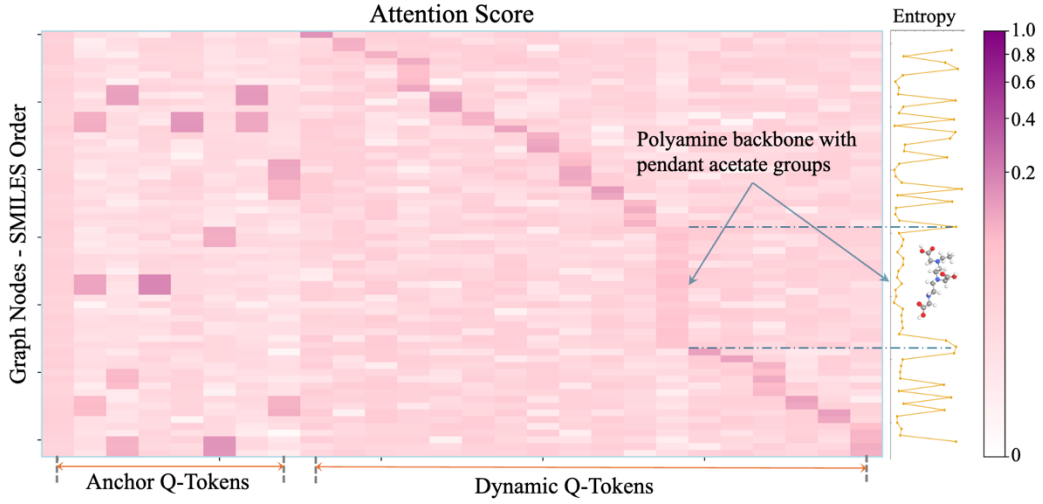


Figure 5: Ablation study of components on the MoleculeQA dataset. Accuracy is reported across four task types (Structure, Source, Property, and Application) when removing each component (modality fusion, Entropy-Guided Patching, or Dynamic Query Transformer).

Attention Visualization:

Shows our Entropy-Guided Patching can capture substructures.



⇨ Future Work More general; More Useful; AI for Science.

Plan to extend this model to scientific graph from molecular graph.

- Knowledge Graphs
- Protein Interaction Networks
- Protein Graphs
- Material & Crystal Graphs
- Reaction Graphs
- Pathway / Metabolic Graphs

Welcome Collaboration