CIKM 2026, ROME, NOVEMBER 2026

S³-Prompt

Resolving Activity Cliffs via Structure-Disentangled Molecular Graph-Text Alignment

Wooseong Syz Cho, **Ha-Myung Park**

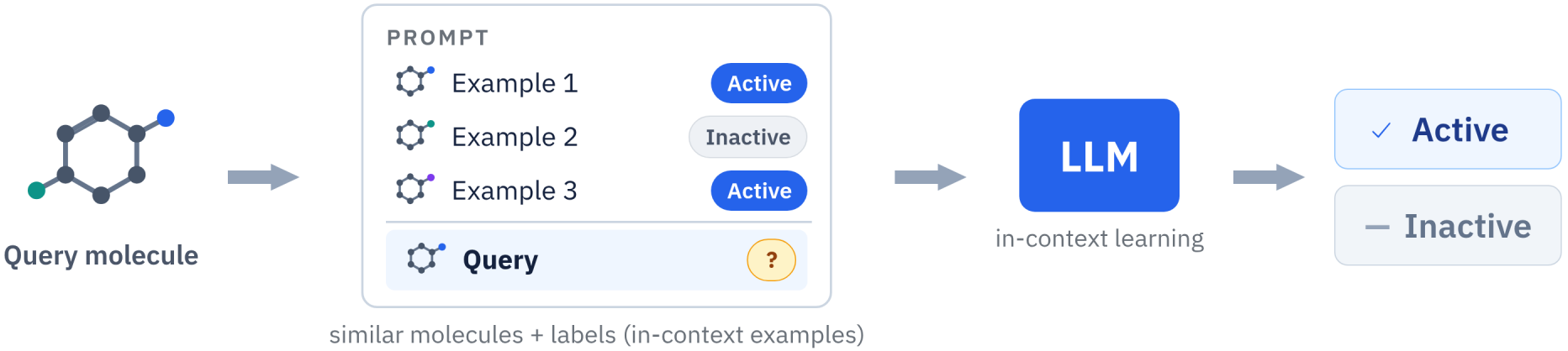
Kookmin University, Seoul, Korea





THE PROBLEM

Classifying molecules as active or inactive



Two target tasks (datasets)

BBBP

Blood–brain barrier penetration

Active — the molecule crosses into the brain

BACE

β-secretase 1 — an Alzheimer-related enzyme

Active — the molecule inhibits the enzyme

THREE CHALLENGES

Why naïve in-context learning breaks on molecules

1 Structural entanglement



SMILES ↓

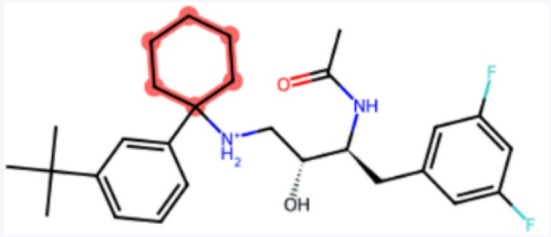
`CC(=O)Oc1ccccc1C(=O)O`

Global and **local** structures are mixed in one string.

Our fix: split into scaffold + side-chain, embed with a GNN.

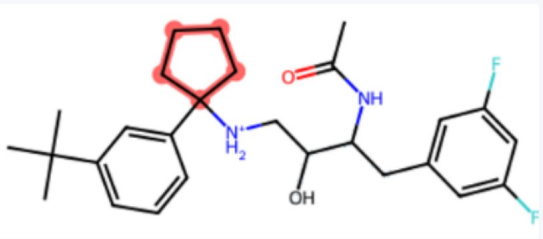
2 Activity cliffs

BACE_73 — Active

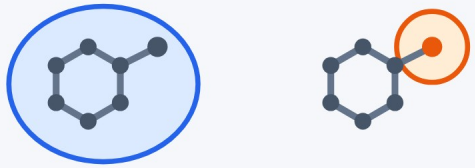


Tanimoto 0.983 — opposite label

BACE_926 — Inactive



3 Task–retrieval misalignment



Whole molecule matters

BBBP, ESOL

One local group matters

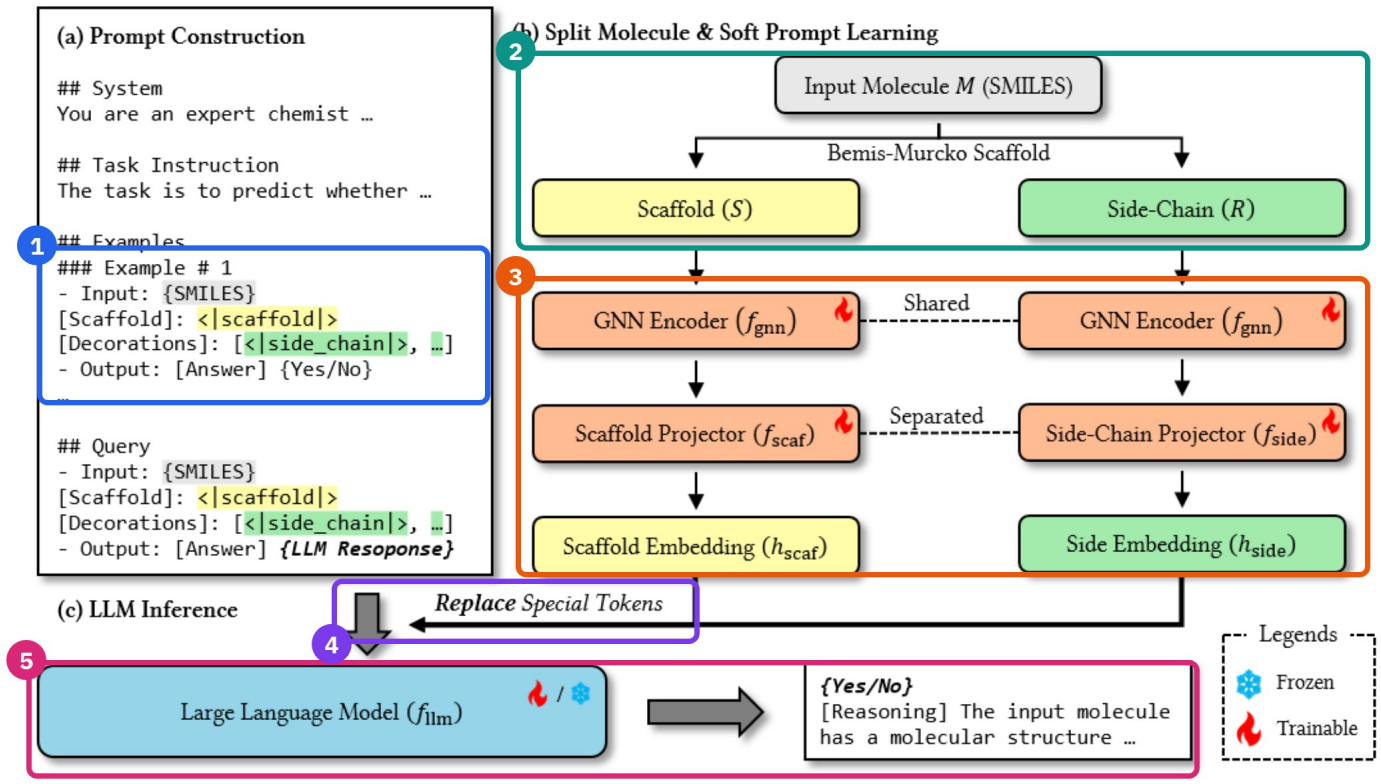
BACE, LIPO

⚠ **Prior ICL methods use one similarity for every task**

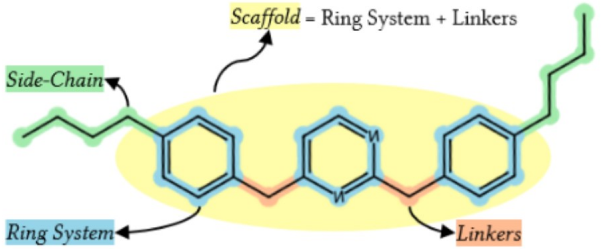
Our fix: choose the similarity per task — global or local.

METHOD

S³-Prompt: overall architecture



1. For a **query** molecule, retrieve the **8 most similar** molecules as examples.
2. Split each molecule into **scaffold** and **side-chain** (Bemis–Murcko).



3. Embed each part with a **GNN**.
4. Insert the embeddings into the prompt as **soft embeddings**, replacing the special tokens.
5. The **LLM** reads the prompt and predicts **active / inactive**.

All modules are trained with contrastive learning.

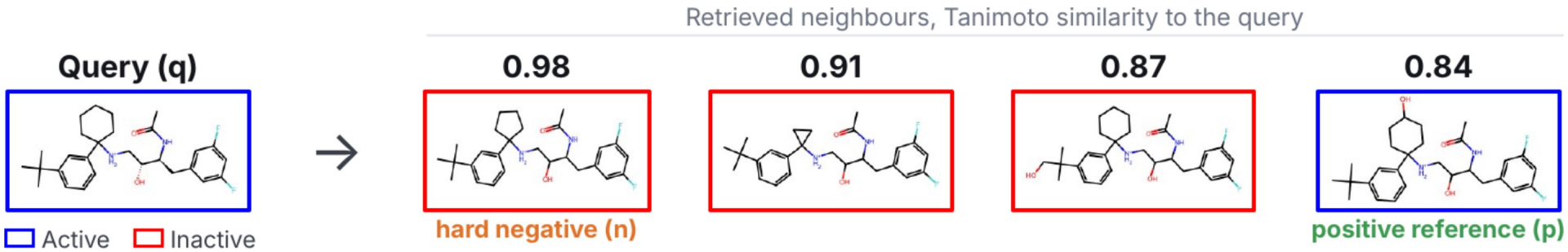
METHOD · CONTRASTIVE LEARNING

Learning the activity cliff on purpose

In the LLM's **final hidden space**, pull the **positive** sample close to the query and push the **negative** sample far away.



From the 8 retrieved samples, we pick p and n like this:



Positive p

Same activity as the query, but the **lowest** similarity (0.84, Active)

Negative n

Different activity, but the **highest** similarity (0.98, Inactive)

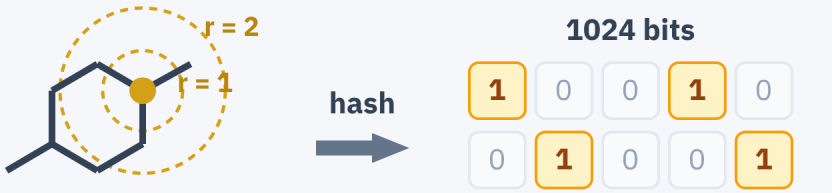
→ **The model is trained to tell apart near-identical molecules with opposite labels — the activity cliff.**

METHOD · STRATEGIC RETRIEVAL

Match the retrieval to the task

Prior in-context methods use **one fingerprint similarity for every task**. We found that a different similarity fits each task.

Topological retrieval

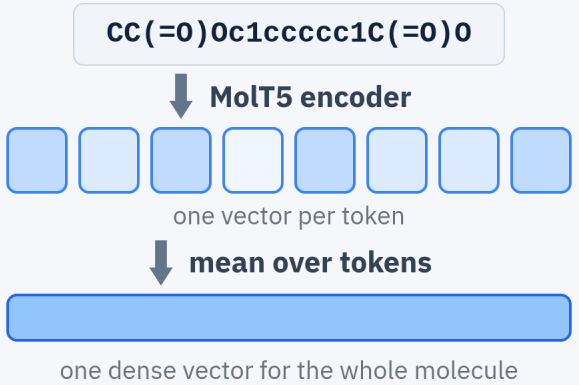


Morgan fingerprint (FP) + Tanimoto similarity

Encodes local **substructures** → use when **local topology** matters.

BACE

Semantic retrieval



MolT5 embedding + cosine similarity

Encodes the **whole molecule** → use when **global information** matters.

BBBP

EXPERIMENTS · MAIN RESULTS

S³-Prompt gives the best AUC on both BACE and BBBP

Category	Model	BACE (↑)	BBBP (↑)
Graph Based	GIN	0.701	0.658
	GraphMVP	0.812	0.724
	MoAMA	0.801	0.728
	MolCLR	0.835	0.735
	KPGT	0.818	0.714
	ACANet	0.691	0.897
Language Based	Galactica-6.7B	0.584	0.535
	MoLFormer	0.882	0.937
	MoleculeSTM	0.808	0.700
	GIMLET	0.696	0.594
	InstructMol	0.821	0.724
	MolCoMA	0.852	0.735
	HIGHT	0.771	0.618
ICL Based	MolRAG	0.754	0.572
	MolecularGPT	0.733	0.726
Ours	**S³-Prompt (FP) †**	**0.906**	0.903
	S³-Prompt (Sem) ‡	0.831	**0.965**

† FP = topological retrieval, frozen LLM (index E) · ‡ Sem = semantic retrieval with LoRA (index G). Both fixed settings are reported on every dataset.

BACE 0.906 with topological retrieval (FP)

BBBP 0.965 with semantic retrieval (Sem)

EXPERIMENTS · ABLATION

What each component adds (settings A to G)

Index	Decomposition	LoRA	Loss	BACE	BBBP
A	–	–	–	0.812	0.921
B	Text-based	–	–	0.882	0.958
C	Text-based	✓	–	0.876	0.957
D	Soft embedding	–	–	0.868	0.921
E	Soft embedding	–	✓	0.906	0.953
F	Soft embedding	✓	–	0.878	0.953
G	Soft embedding	✓	✓	0.889	0.965

Decomposition as SMILES text (A → B)

Scaffold + side-chain in the prompt: BACE 0.812 → 0.882 **(+0.07)**

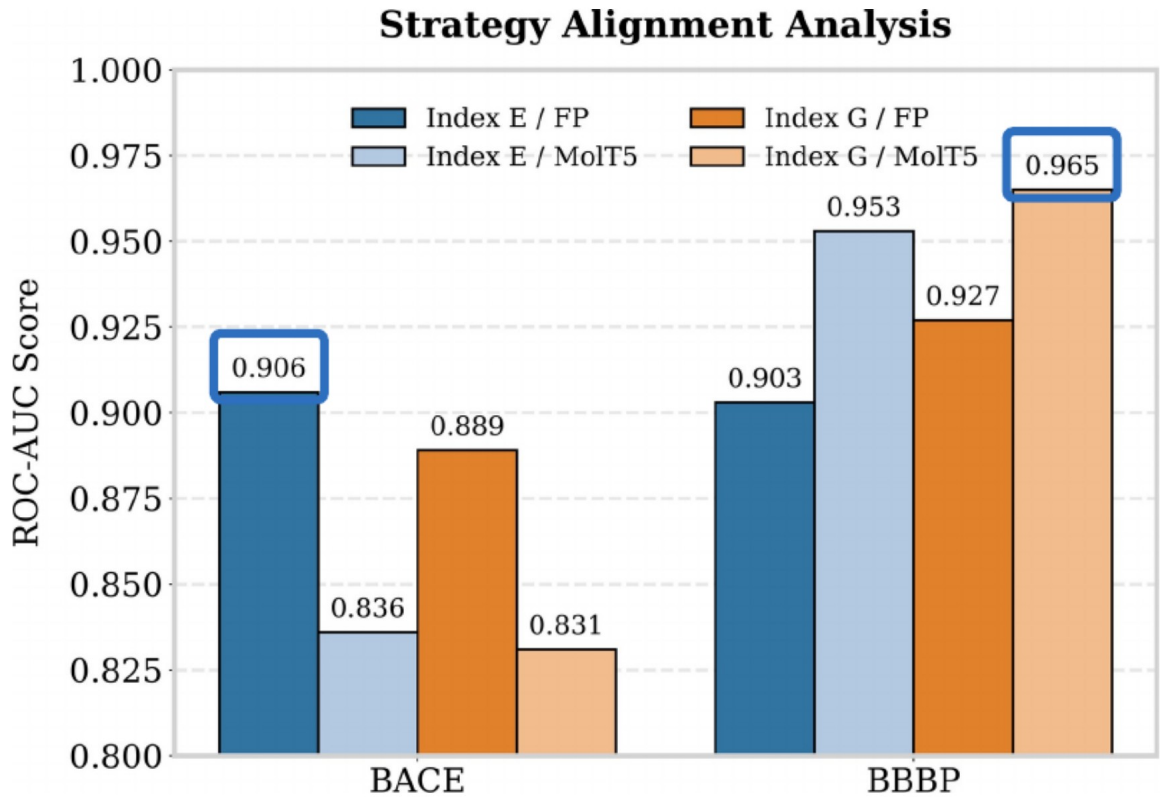
Soft embedding + contrastive loss (D → E)

Soft embedding alone drops to 0.868, but with the contrastive loss BACE reaches 0.906 **(+0.02 over text)**

LoRA fine-tuning (E → G)

Tuning the LLM together adds about **+0.01** on BBBP: 0.953 → 0.965

The best retrieval method differs by task



Same model, **only the retrieval method is swapped**: topological (FP) vs semantic (MolT5).

BACE prefers topological

FP **0.906** vs MolT5 0.836 (Index E)

BBBP prefers semantic

MolT5 **0.965** vs FP 0.927 (Index G)

→ **Picking the wrong retrieval costs up to 0.07 AUC.**

CONCLUSION

Takeaways for in-context molecular classification

01 · REPRESENTATION

Decompose, then embed

Split the molecule into **scaffold** and **side-chain**, embed with a **GNN**, and insert as **soft embeddings**.

02 · TRAINING

Contrastive learning

Hard positive / negative pairs from the retrieved samples relieve the **activity-cliff** problem.

03 · RETRIEVAL

Match similarity to the task

Topological for **BACE, LIPO** · Semantic for **BBBP, ESOL**.

S³-Prompt reaches state-of-the-art — BACE **0.906** · BBBP **0.965** AUC

```
code github.com/Syzseisus/S3-Prompt
model huggingface.co/Syzseisus/S3-Prompt
dataset huggingface.co/datasets/Syzseisus/S3-Prompt
```

Thank you



