Data-Efficient Molecular Generation with Hierarchical Textual Inversion

Seojin Kim[1], Jaehyun Nam[1], Sihyun Yu[1], Younghoon Shin[2], Jinwoo Shin[1]

[1]Korea Advanced Institute of Science and Technology (KAIST)
[2]Korea University

ICML 2024

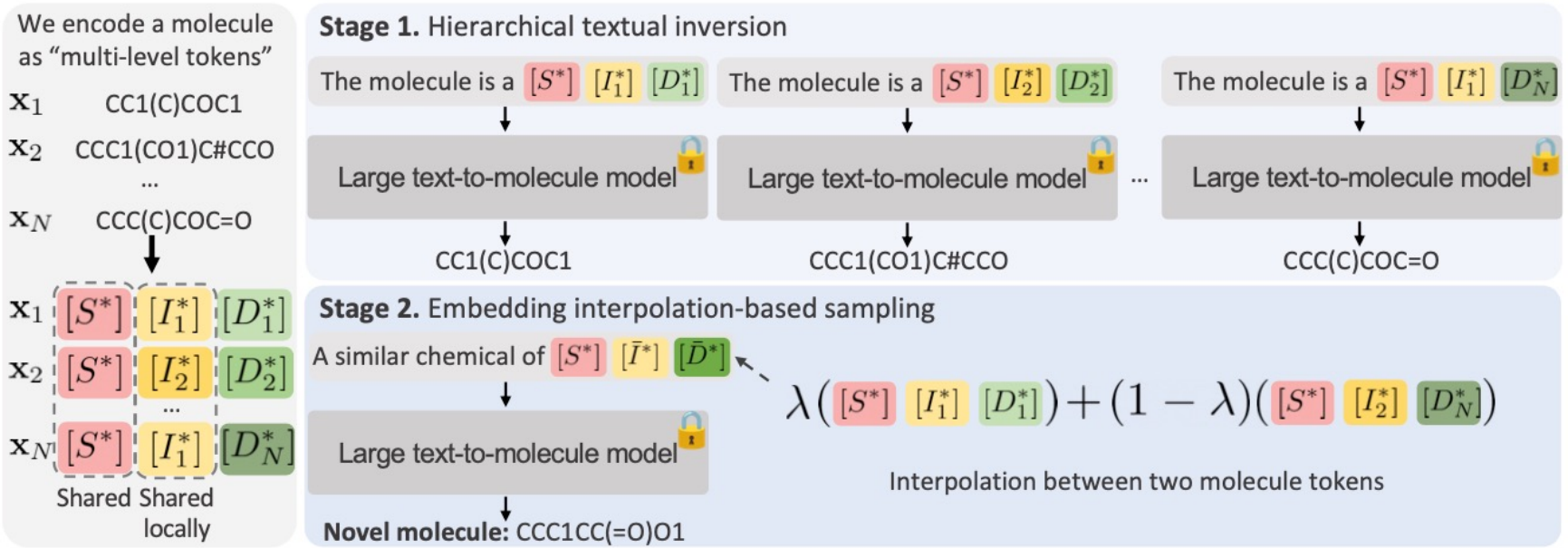
Data-Efficient Molecular Generation

Data-efficient molecular generation is important in chemical application:

- **Prior molecular generation approaches:** Assume abundant training data, e.g., QM9 contains 130k molecules
- **Difficulty of the collection from target distribution**: Require experimental verification, e.g., drug-likeness on HIV

- **Hi-Mol**: We propose **a novel framework for data-efficient molecular generation**

Few-shot adaptation of molecular foundation models generates high-quality molecules in a data-efficient manner



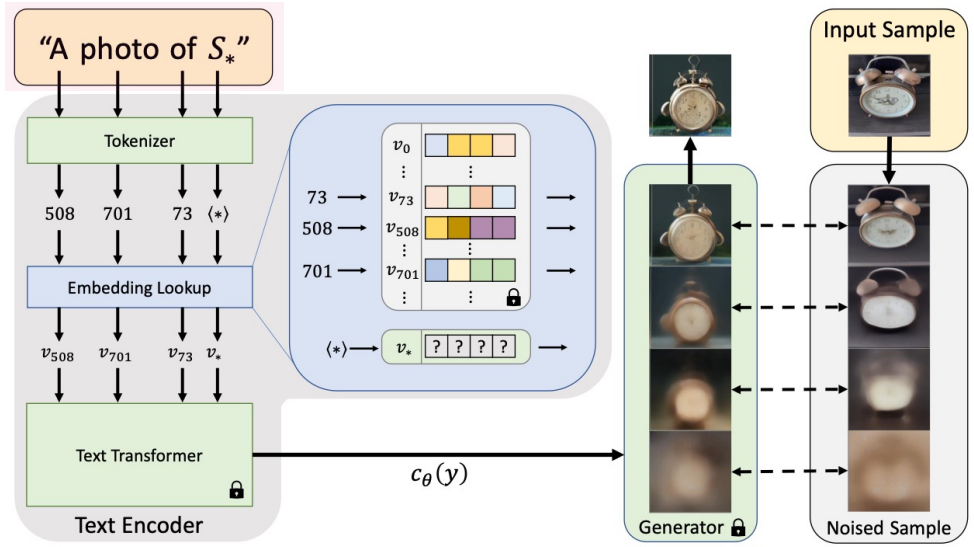
Textual Inversion in Text-to-Image Foundation model

Textual inversion introduces a **new word** that represents a **common concept** **from** **a few images**:

- The embedding of new word [S*] is optimized to reconstruct given images of a common concept

- **Hi-Mol**: We propose **a novel textual inversion framework for data-efficient molecular generation**

Hi-Mol effectively captures the distribution of limited number of molecules with molecular foundation model



$$v_* = \arg\min_v \mathbb{E}_{z\sim\mathcal{E}(x),y,\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta(z_t, t, c_\theta(y))\|_2^2\right]$$

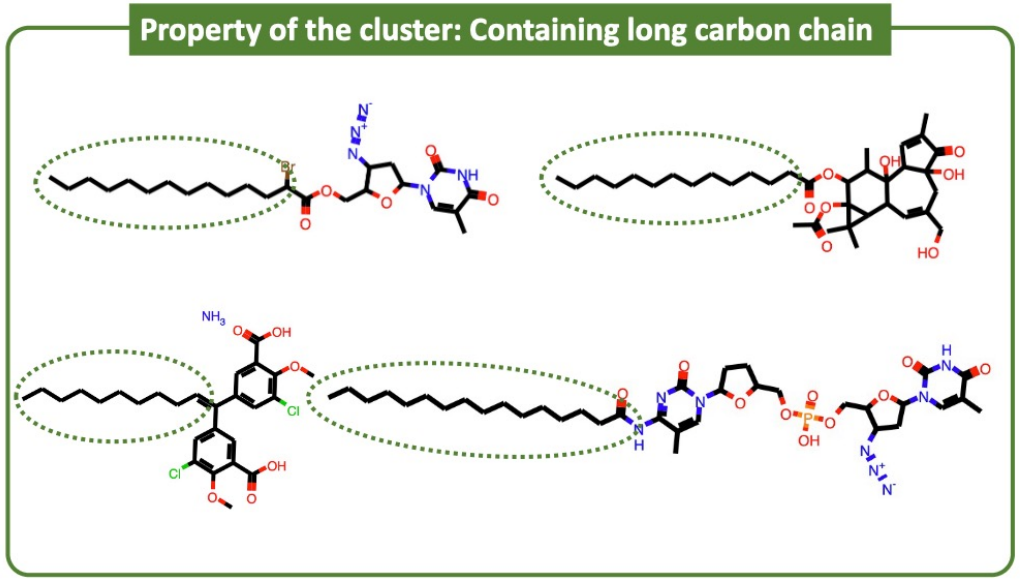
[Gal et al., 2022] An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, arXiv 2022

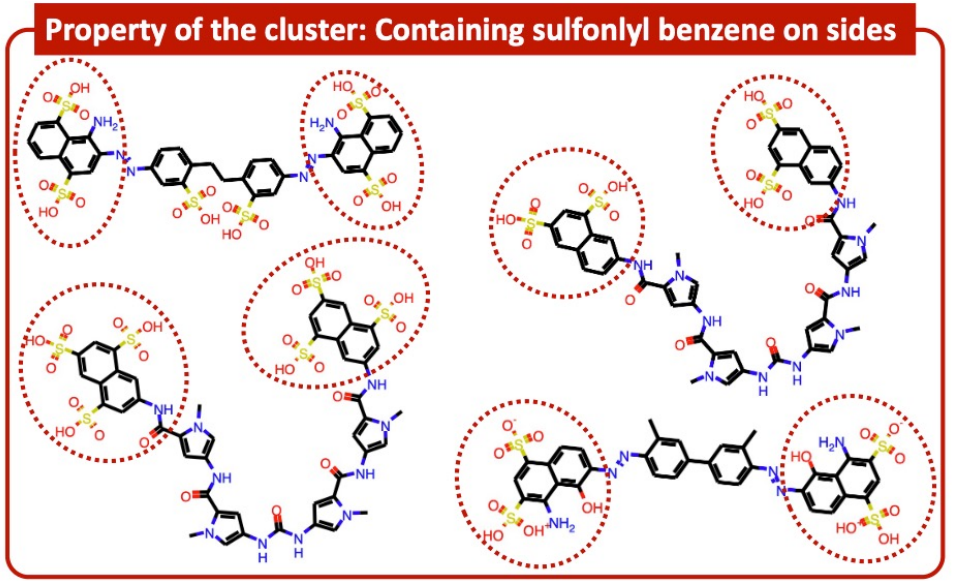
Textual Inversion for Molecular Generation?

Naïve application of **textual inversion** into **molecular foundation models** does not work well in general

- Molecules with the common property do not necessarily share visually similar molecular structures
- Modeling such molecules with only "shared" word cannot effectively capture the common concept of molecules

Training prompt	FCD ↓	NSPDK ↓	Valid. ↑	Unique. ↑	Novelty ↑
The molecule is a $[S^*]$.	7.913	0.041	**96.2**	19.3	39.5
The molecule is a $[S^*][D_n^*]$.	0.486	0.002	93.8	70.8	72.3
The molecule is a $[S^*][I_{c_n}^*][D_n^*]$.	**0.434**	**0.001**	90.7	**75.8**	**73.5**

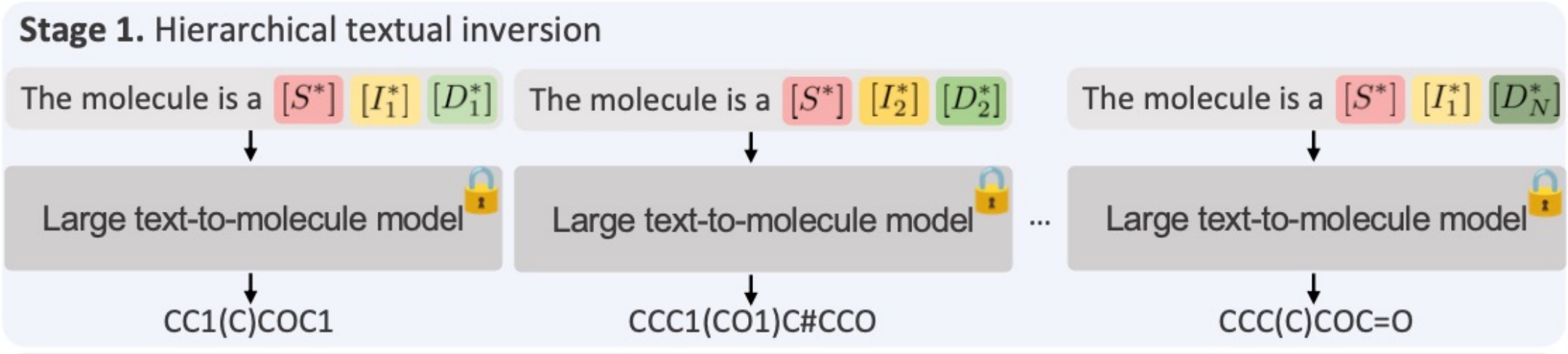




Textual Inversion for Molecular Generation?

Hierarchical textual inversion: low- (or high-) level features of molecules are separately learned

- We introduce 3 kinds of words: [S*] (shared), [I*] (intermediate), [D*] (detail)
- Shared word [S*]: A single word for whole dataset, captures overall semantic of target molecules

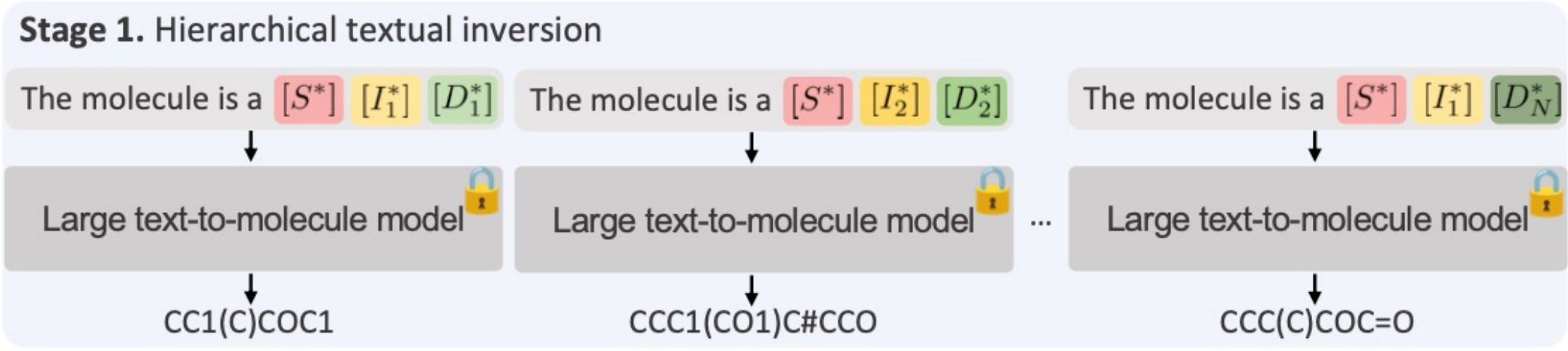


$$\mathcal{L}(\theta; \mathbf{x}_n) := \min_{k \in [1,K]} \mathcal{L}_{\text{CE}}\Big(\texttt{softmax}\big(f(\text{“The molecule is a } [S^*][I_k^*][D_n^*]\text{”})\big),\ \texttt{SMILES}(\mathbf{x}_n)\Big)$$

Textual Inversion for Molecular Generation?

Hierarchical textual inversion: low- (or high-) level features of molecules are separately learned

- We introduce 3 kinds of words: [S*] (shared), [I*] (intermediate), [D*] (detail)
- Intermediate words [I_k*]: Word assigned for k-th cluster, captures cluster-wise semantics with 1 <= k <= K



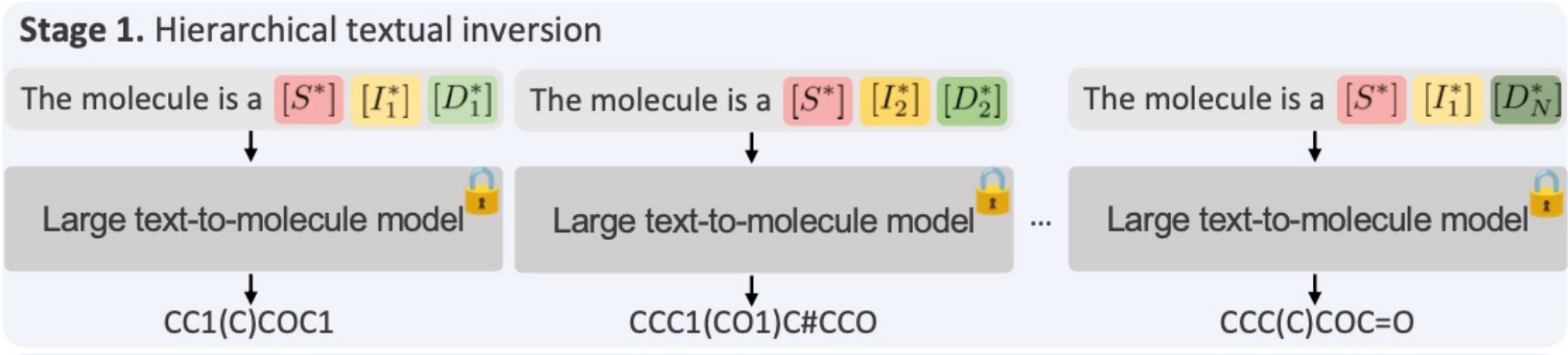
$$\mathcal{L}(\theta; \mathbf{x}_n) := \min_{k \in [1,K]} \mathcal{L}_{\text{CE}}\Big(\texttt{softmax}\big(f(\text{“The molecule is a } [S^*][I_k^*][D_n^*]\text{”})\big),\ \texttt{SMILES}(\mathbf{x}_n)\Big)$$

k is automatically assigned by multiple-choice learning

Textual Inversion for Molecular Generation?

Hierarchical textual inversion: low- (or high-) level features of molecules are separately learned

- We introduce 3 kinds of words: [S*] (shared), [I*] (intermediate), [D*] (detail)
- Detail words [D_n*]: Word assigned for n-th molecule, captures molecule-wise detailed semantics with 1 <= n <= N

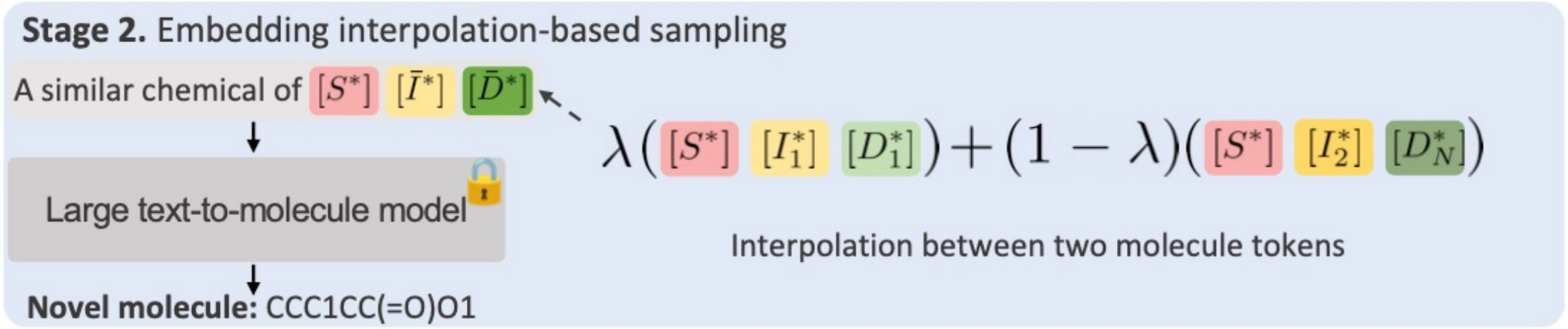


$$\mathcal{L}(\theta; \mathbf{x}_n) := \min_{k \in [1,K]} \mathcal{L}_{\text{CE}}\Big(\texttt{softmax}\big(f(\text{“The molecule is a } [S^*][I_k^*][D_n^*]\text{”})\big),\ \texttt{SMILES}(\mathbf{x}_n)\Big)$$

Textual Inversion for Molecular Generation?

Embedding interpolation-based sampling: Generation via interpolation of learned word embeddings

- We randomly interpolate detail token embeddings and the corresponding intermediate token embeddings
- Then, we forward obtained interpolated token embeddings into the original molecular foundation model



$$(\bar{\mathbf{i}}, \bar{\mathbf{d}}) := \lambda(\mathbf{i}_{c_i}, \mathbf{d}_i) + (1-\lambda)(\mathbf{i}_{c_j}, \mathbf{d}_j),$$
$$\mathbf{x} := f(\text{“A similar chemical of } [S^*][\bar{I}^*][\bar{D}^*]\text{”})$$

Experiments: HI-Mol is effective for limited training samples

We report the molecular generation performance of active (label-1) molecules in MoleculeNet benchmark

- MoleculeNet contains limited active molecules compared to conventional datasets, e.g., 1,232 molecules in HIV.
- Best scores are marked bold; Grammar is marked when the method forces the validity of generated molecules.

Dataset	Method	Class	Grammar	Active. ↑	FCD ↓	NSPDK ↓	Valid. ↑	Unique. ↑	Novelty ↑
HIV	JT-VAE [13]	Fragment	✓	0.0	38.8	0.221	**100**	25.4	**100**
	PS-VAE [33]	Fragment	✓	3.7	21.8	0.053	**100**	91.4	**100**
	STGG [8]	SMILES	✓	1.6	20.2	0.033	**100**	95.8	**100**
	CRNN [3]	SMILES	✗	3.3	29.7	0.064	30.0	**100**	**100**
	GDSS [9]	Graph	✗	0.0	34.1	0.080	69.4	**100**	**100**
	GSDM [27]	Graph	✗	0.0	32.7	0.124	52.0	90.4	**100**
	DiGress [7]	Graph	✗	0.0	26.2	0.067	17.8	**100**	**100**
	HI-Mol (Ours)	SMILES	✗	**11.4**	**19.0**	**0.019**	60.6	94.1	**100**
BBBP	JT-VAE [13]	Fragment	✓	80.6	37.4	0.202	**100**	10.8	**100**
	PS-VAE [33]	Fragment	✓	84.9	17.3	0.039	**100**	91.6	**100**
	STGG [8]	SMILES	✓	89.1	14.4	0.019	99.8	95.8	**100**
	CRNN [3]	SMILES	✗	88.8	20.2	0.026	54.0	**100**	**100**
	GDSS [9]	Graph	✗	0.0	35.7	0.065	88.4	99.2	**100**
	GSDM [27]	Graph	✗	11.6	26.4	0.086	64.0	94.3	**100**
	DiGress [7]	Graph	✗	8.2	17.4	0.033	43.8	94.6	**100**
	HI-Mol (Ours)	SMILES	✗	**94.4**	**11.2**	**0.011**	78.8	92.9	**100**

Experiments: HI-Mol is effective for limited training samples

We report the molecular generation performance on small subset of QM9 dataset

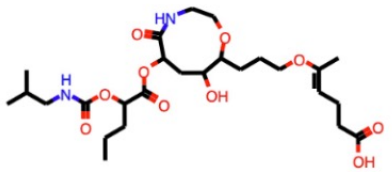
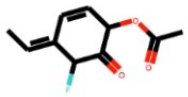
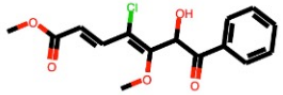
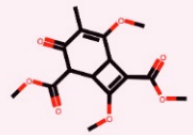
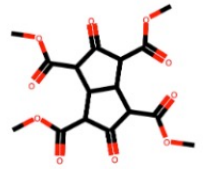
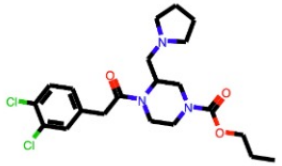
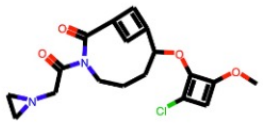
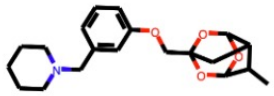
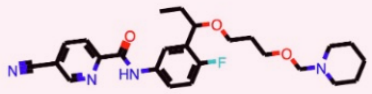
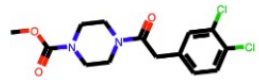
- Hi-Mol outperforms prior methods when trained only with 2% of the entire training set of QM9 dataset.
- Best scores are marked bold; Grammar is marked when the method forces the validity of generated molecules.

Method	Class	Grammar	FCD ↓	NSPDK ↓	Valid. ↑	Unique. ↑	Novelty ↑
STGG† [8]	SMILES	✓	0.585	-	**100**	95.6	69.8
CG-VAE† [34]	Graph	✓	1.852	-	**100**	98.6	94.3
GraphAF [28]	Graph	✗	5.268	0.020	67	94.5	88.8
GraphDF [29]	Graph	✗	10.82	0.063	82.7	97.6	**98.1**
MoFlow [30]	Graph	✗	4.467	0.017	91.4	98.7	94.7
EDP-GNN [31]	Graph	✗	2.680	0.005	47.5	**99.3**	86.6
GraphEBM [32]	Graph	✗	6.143	0.030	8.22	97.8	97.0
GDSS [9]	Graph	✗	2.900	0.003	95.7	98.5	86.3
GSDM* [27]	Graph	✗	2.650	0.003	99.9	-	-
HI-Mol (Ours; 2%)	SMILES	✗	0.434	**0.001**	90.7	75.8	73.5
HI-Mol (Ours; 5%)	SMILES	✗	**0.412**	**0.001**	89.4	85.8	70.4

Experiments: HI-Mol generates more faithful molecules

Molecules from Hi-Mol framework introduces novel substructures preserving some original substructures

- For example, Hi-Mol introduces 4- and 6- membered rings while preserving several ester groups
- We report the most similar molecule with given query molecule in terms of Tanimoto similarity

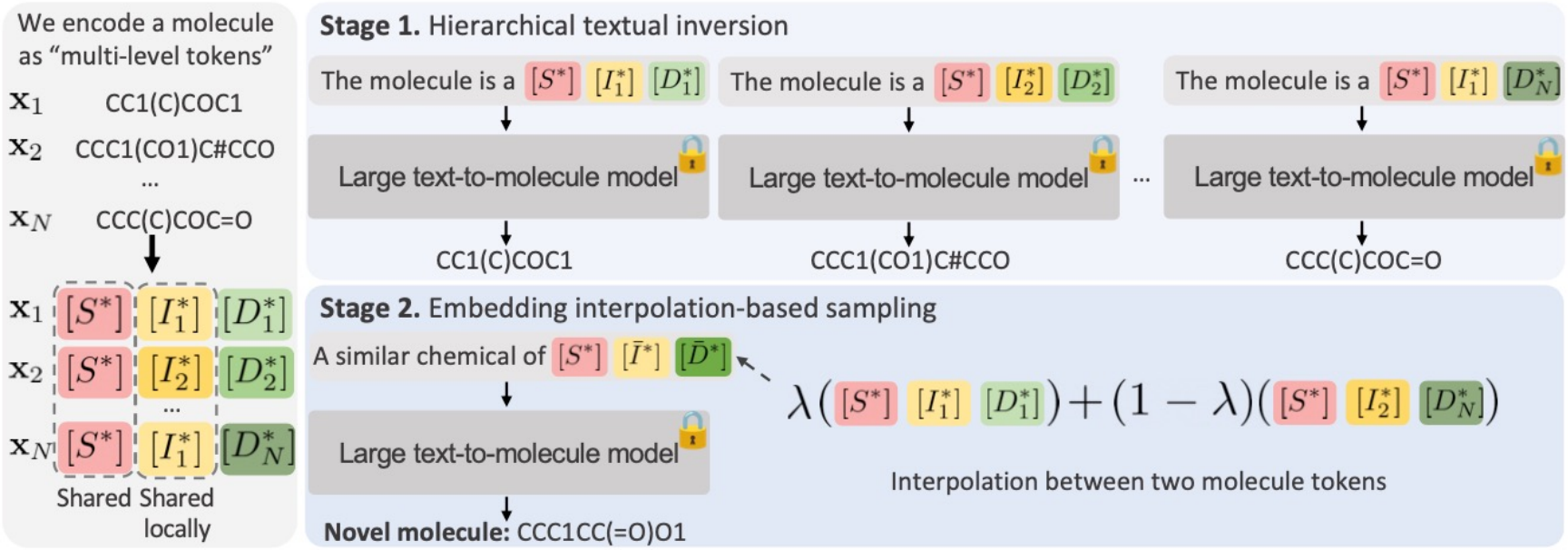
Dataset	DiGress [7]	PS-VAE [33]	STGG [8]	**HI-Mol (Ours)**	Train
HIV	 0.154	 0.141	 0.157	 **0.326**	
BBBP	 0.238	 0.238	 0.246	 **0.505**	

Hi-Mol: Data-Efficient Framework for Molecular Generation

Summary: We propose a simple yet effective data-efficient framework for molecular generation

We propose Hi-Mol = Hierarchical textual inversion for molecular generation

1. Hierarchical textual inversion with text-to-molecule model and embedding interpolation sampling
2. Molecular generation: Hi-Mol shows superior performance even with smaller number of training samples
3. Compositionality: Can be used to generate molecules with more flexibility



Thank you