Accepted by ICLR 2025

Fragment and Geometry Aware Tokenization of Molecules for Structure-Based Drug Design Using Language Models

Cong Fu, Xiner Li, Blake Olson, Heng Ji, Shuiwang Ji

Texas A&M University

Yijingxiu Lu

Table of Contents

- Introduction
- Background
- Preliminary
  - Rigid transformation
  - SE(3)
- Model Architecture
  - Translate molecule into sequence of fragments
  - Overall framework
- Experiments
- Summary

Introduction

In this work, authors propose **Frag2Seq**:

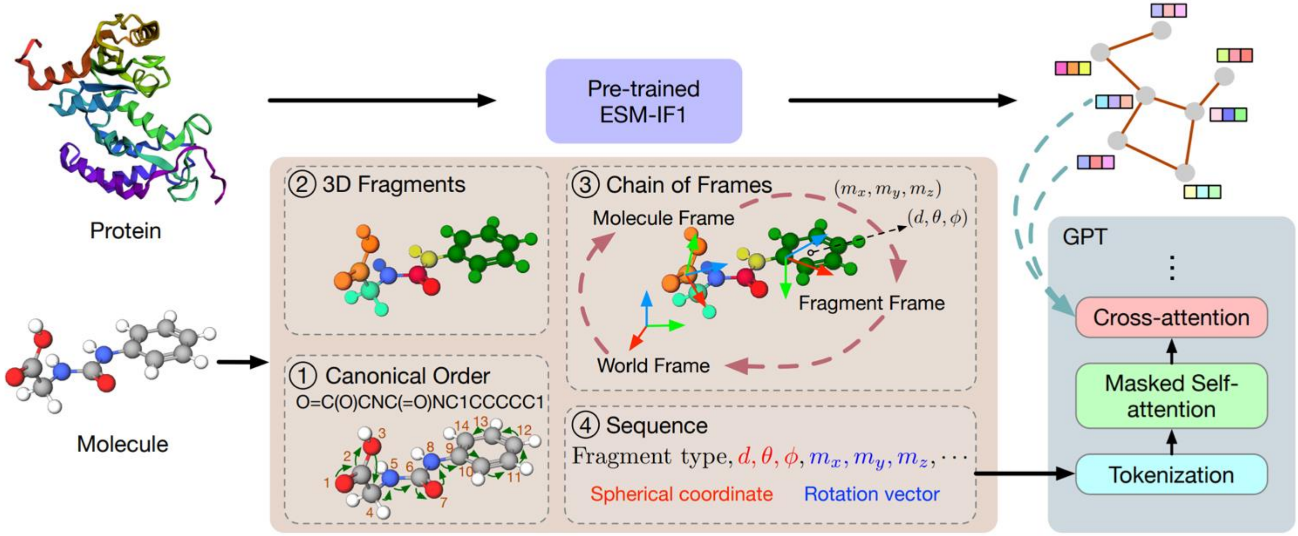
- Frag2Seq: Fragment-based molecule generation (target-aware) with language model.
  - ensure language model’s ability in simulating the physical and chemical properties of molecules.
  - encode protein context information in LMs for efficiently capturing interaction information.
- SE(3)-invariant sequences that preserve geometric information of 3D fragments.

Input:

- **Molecule**: sequence representation Frag2Seq
- **Protein**: node embedding of backbone from ESM-IF1

Model Architecture:

- **GPT:** learn distribution of molecule fragment tokens.
- **Cross attention:** protein node embedding (k,v) x ligand token embedding (q)



Background

Structure-based drug design (SBDD):

Definition:

- Design and optimize molecules to interact specifically and effectively with biological targets.

Challenges:

- Requires the model to capture complicated protein-ligand interaction while improving drug-likeness of designed molecules.
- Current methods only consider atom-wise generation.
- Diffusion models are inefficient.

Language Model (LM):

Advantages:

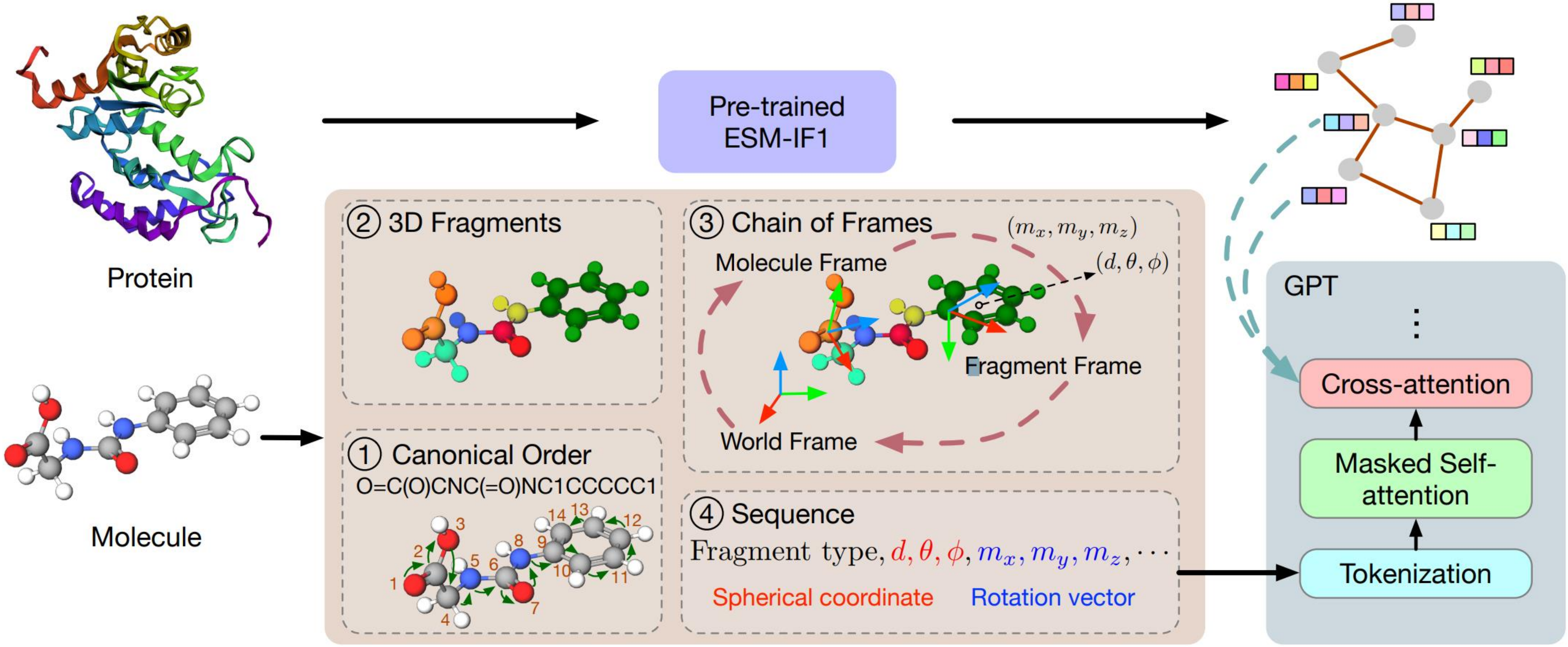
- Can handle large datasets with prominent efficiency over diffusion-based methods.
- Learn from massive biological and chemical texts for diverse potential tasks.

Limitation:

- Difficulty in applying LM on geometric graph data.

Pipelines

1. Convert 3D molecules into fragment-informed sequences.
   - Split 3D molecules into 3D fragments.
   - Construct a bijective mapping between 3D fragments and SE(3)-invariant sequences.
2. Extract protein pocket embedding from pre-trained folding model (ESM-2).
3. Use cross-attention mechanism to generate target-aware molecules with language model.



Preliminary

Rigid transformation:

$$A = \begin{pmatrix} R & \mathbf{t} \\ 0 & 1 \end{pmatrix}$$

where R is a 3 x 3 rotation matrix, and t is a translation vector

Inverse of a general rigid transformation:

$$\begin{pmatrix} R & \mathbf{t} \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} R^T & -R^T\mathbf{t} \\ 0 & 1 \end{pmatrix}$$

Special Euclidean Group SE(3)

- The group of rigid transformation (Rotation + Translation)
- Inner product:

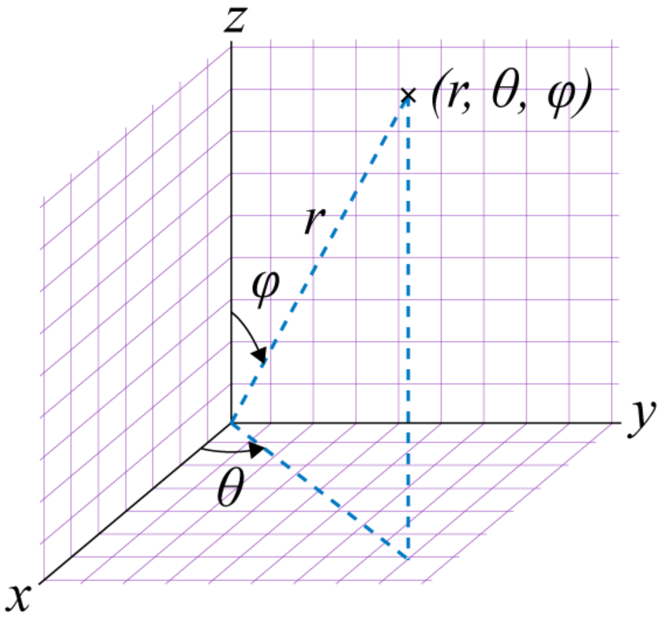
$$\begin{pmatrix} R_2 & \mathbf{t}_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R_1 & \mathbf{t}_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} R_2 R_1 & R_2\mathbf{t}_1 + \mathbf{t}_2 \\ 0 & 1 \end{pmatrix}$$

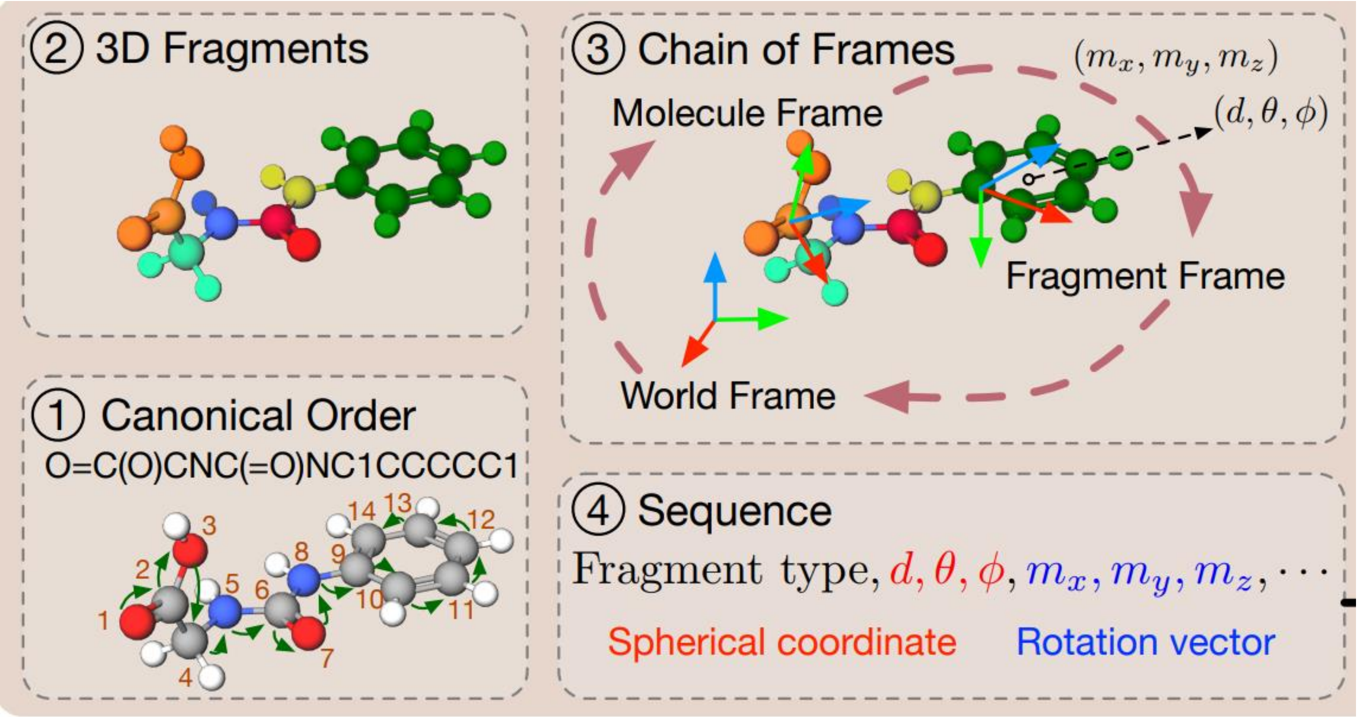
Preliminary

Spherical Coordinates:

- conversion of rectangular coordinates to spherical coordinate:

$$x = r\,(\sin\theta)\,(\cos\Phi)$$
$$y = r\,(\sin\theta)\,(\sin\Phi)$$
$$z = r\,(\cos\theta)$$

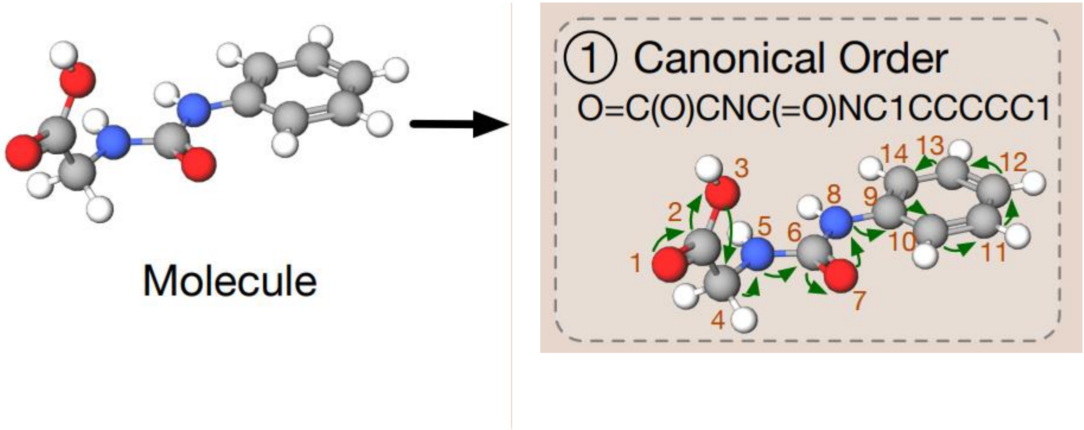




Atom Ordering based on 3D Graph Isomorphism

- SMILES -> Canonical SMILES
  - Let L : M → L be a function that maps a molecule M ∈ M, the set of all finite 3D molecular graphs, to its canonical order L(M) ∈ L, the set of all possible canonical orders,

$$L(M_1) = L(M_2) \Leftrightarrow M_1 \cong_{3D} M_2$$

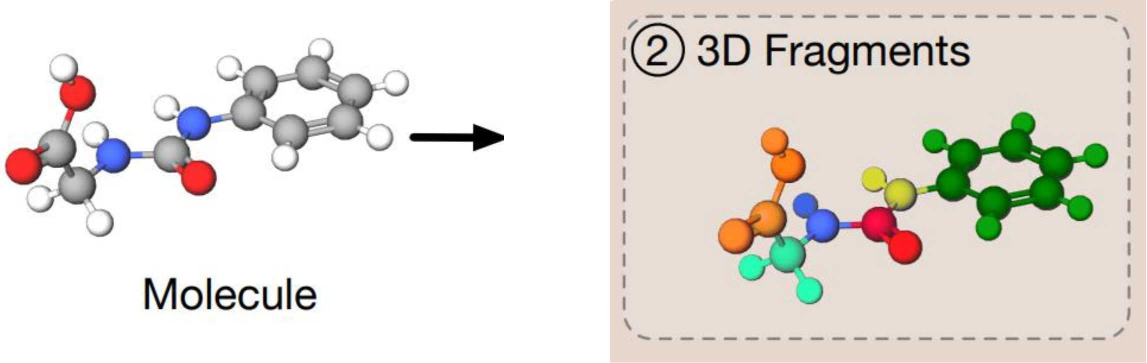


3D Molecule Fragmentation

Molecule to Fragments:

- Cutting rotatable chemical bonds iff (to prevent breaking the functional groups):
    1) The bond is not in a ring.
    2) The bond type is single.
    3) The degree of the beginning and end atom on the bond is larger than 1.

Sort fragments based on the order of their appearance in the canonical SMILES representation.



Fragment-based 3D Molecule Tokenization

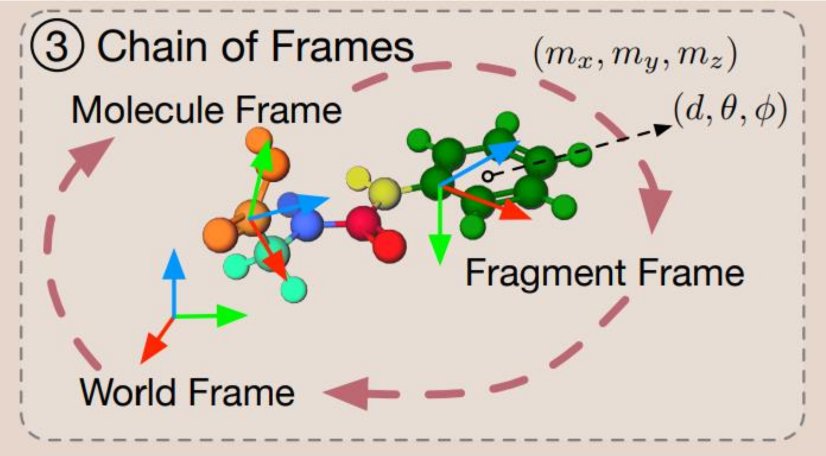
SE(3)-Equivariant Molecule and Fragment Frames Construction:

- Calculate SE(3)-invariance property:
  - Sort the fragments based on their first ranked atom in the canonical order $L(M)$.
  - **Fragment center**: calculated as the average of atom coordinates.
  - **Molecule local frame**: calculated with the first three non-collinear fragment centers (l_1, l_2, l_m) as:

$$\boldsymbol{x} = \text{normalize}(\boldsymbol{v}_{\ell_2} - \boldsymbol{v}_{\ell_1}), \quad \boldsymbol{y} = \text{normalize}\left((\boldsymbol{v}_{\ell_m} - \boldsymbol{v}_{\ell_1}) \times \boldsymbol{x}\right), \quad \boldsymbol{z} = \boldsymbol{x} \times \boldsymbol{y},$$

Where $m = (x, y, z)$ is defined as the molecule local frame

  - **Fragment local frame**: calculated with the first three non-collinear atoms in a fragment.



Fragment-based 3D Molecule Tokenization

Homogeneous transformation:

- Construct homogeneous transformation matrices from rotation matrices R and translation vectors t:

$$T_{\mathfrak{m}\to\mathfrak{w}} = \begin{bmatrix} R_{\mathfrak{m}\to\mathfrak{w}} & \boldsymbol{t}_{\mathfrak{m}\to\mathfrak{w}} \\ \mathbf{0} & 1 \end{bmatrix}, \quad T_{\mathfrak{g}\to\mathfrak{w}} = \begin{bmatrix} R_{\mathfrak{g}\to\mathfrak{w}} & \boldsymbol{t}_{\mathfrak{g}\to\mathfrak{w}} \\ \mathbf{0} & 1 \end{bmatrix},$$

From molecule coordinates to fragment coordinates:

$$\begin{aligned} T_{\mathfrak{g}\to\mathfrak{m}} = T_{\mathfrak{m}\to\mathfrak{w}}^{-1} T_{\mathfrak{g}\to\mathfrak{w}} &= \begin{bmatrix} R_{\mathfrak{m}\to\mathfrak{w}}^{T} & -R_{\mathfrak{m}\to\mathfrak{w}}^{T}\boldsymbol{t}_{\mathfrak{m}\to\mathfrak{w}} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} R_{\mathfrak{g}\to\mathfrak{w}} & \boldsymbol{t}_{\mathfrak{g}\to\mathfrak{w}} \\ \mathbf{0} & 1 \end{bmatrix} \\ &= \begin{bmatrix} R_{\mathfrak{m}\to\mathfrak{w}}^{T} R_{\mathfrak{g}\to\mathfrak{w}} & R_{\mathfrak{m}\to\mathfrak{w}}^{T}(\boldsymbol{t}_{\mathfrak{g}\to\mathfrak{w}} - \boldsymbol{t}_{\mathfrak{m}\to\mathfrak{w}}) \\ \mathbf{0} & 1 \end{bmatrix}. \end{aligned}$$

$$R_{\mathfrak{g}\to\mathfrak{m}} = R_{\mathfrak{m}\to\mathfrak{w}}^{T} R_{\mathfrak{g}\to\mathfrak{w}}, \quad \boldsymbol{t}_{\mathfrak{g}\to\mathfrak{m}} = R_{\mathfrak{m}\to\mathfrak{w}}^{T}(\boldsymbol{t}_{\mathfrak{g}\to\mathfrak{w}} - \boldsymbol{t}_{\mathfrak{m}\to\mathfrak{w}}).$$

Convert atom local coordinates from fragment local frame back to the world frame:

$$\boldsymbol{t}_{\mathfrak{g}\to c(\mathcal{G})} = V_{c(\mathcal{G})}^{\mathfrak{m}} - \boldsymbol{t}_{\mathfrak{g}\to\mathfrak{m}},$$

SE(3)-Invariant Fragment Local Representations

Represent fragment center with spherical coordinates:

- Convert the coordinates of each fragment center to spherical coordinates d, θ, ϕ under the molecule frame $m = (x, y, z)$:

$$d_{\ell_i} = ||\boldsymbol{v}_{\ell_i} - \boldsymbol{v}_{\ell_1}||_2, \ \ \theta_{\ell_i} = \arccos\left((\boldsymbol{v}_{\ell_i} - \boldsymbol{v}_{\ell_1}) \cdot \boldsymbol{z} / d_{\ell_i}\right),$$
$$\phi_{\ell_i} = \operatorname{atan2}\left((\boldsymbol{v}_{\ell_i} - \boldsymbol{v}_{\ell_1}) \cdot \boldsymbol{y}, (\boldsymbol{v}_{\ell_i} - \boldsymbol{v}_{\ell_1}) \cdot \boldsymbol{x}\right).$$

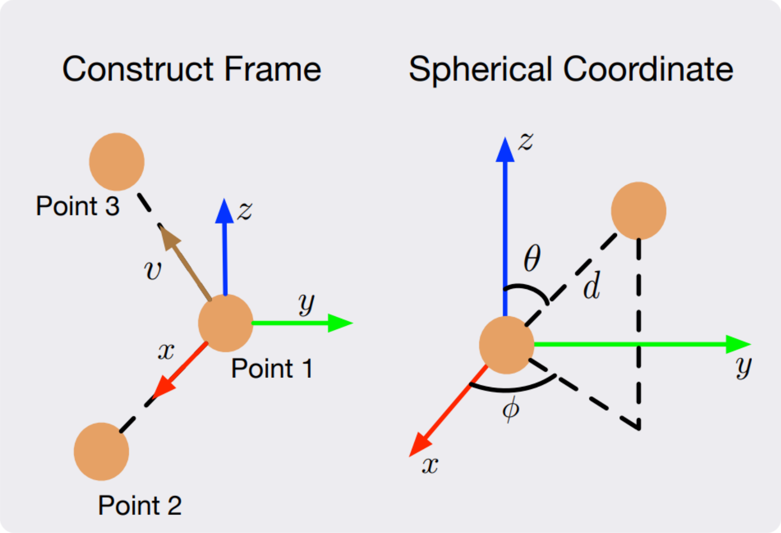
Represent fragment local frame with rotation vector:

- Obtain the rotation angle ψ and rotation axis $a = (m_{xi}, m_{yi}, m_{zi})$ from the rotation matrix

Final fragment-position vector:

$$x_{l_i}^* = [s_i, d_i, \theta_i, \phi_i, m_{xi}, m_{yi}, m_{zi}]$$

- where s_i is the canonical SMILES string of fragment g_i

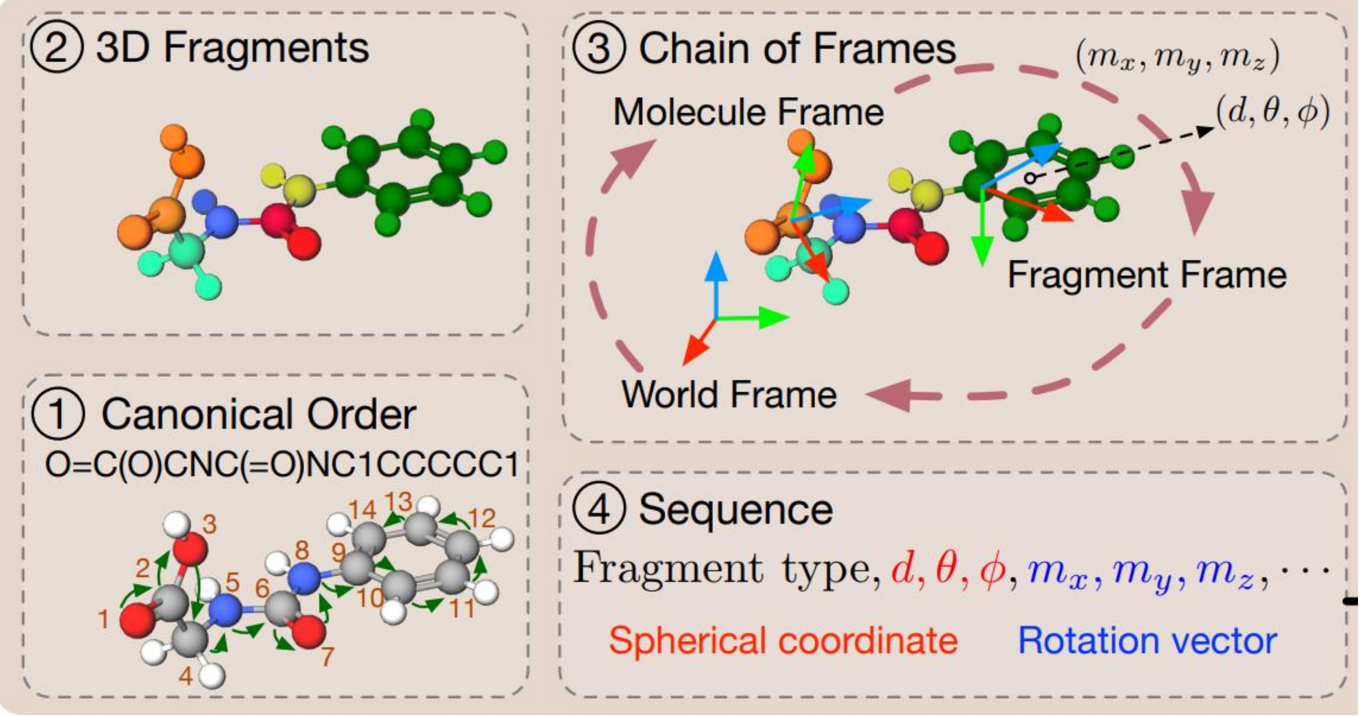


Fragment and Geometry Aware Tokenization

Frag2Seq:

- Given a molecule M with k fragments, frag2seq is defined as the concatenation of fragment-position vectors in canonical order:

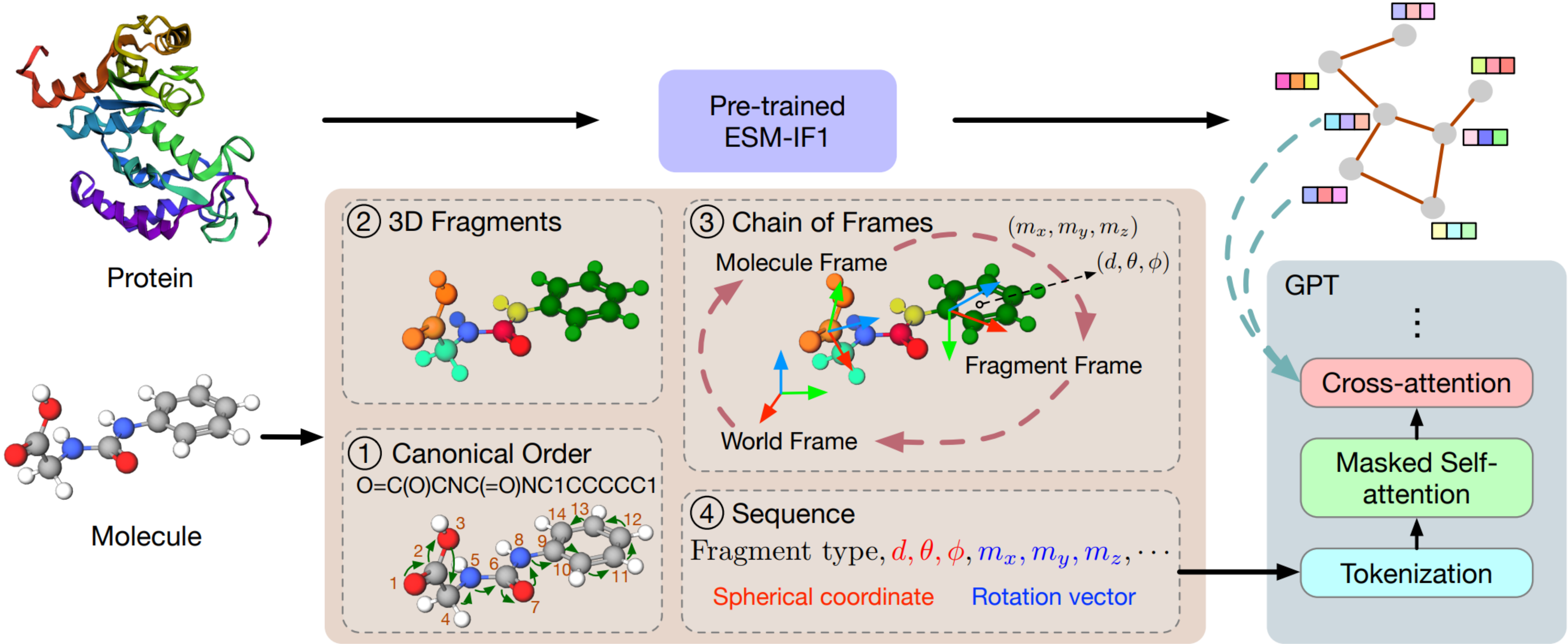
$$Frag2Seq(M) = concat(x^*_{l_1}, \cdots, x^*_{l_k})$$



Overview of Frag2Seq pipeline

Objective function:

- **Next token prediction:** $$\mathcal{L}(U) = \sum_i \log p_\theta(u_i | u_{i-1}, \cdots, u_1).$$



Experiments

Dataset:

- CrossDocked: docking pose dataset that curated from PDBbind
  - **Train**: 100,000 protein-ligand pairs.
  - **Test**: 100 proteins.

Baselines:

- **Auto-regressive methods:**
  3D-SBDD, Pocket2Mol, GraphBP
- **Diffusion methods:**
  TargetDiff, DecompDiff, DiffSBDD

Evaluation Metrics:

- **Vina Score**: estimated binding affinity.
- **High Affinity**: percentage of generated molecules that have higher binding affinity than reference.
- **QED**: measure of drug-likeness.
- **SA**: synthetic feasibility.
- **Diversity**: pairwise diversity of generated molecules for a binding pocket.
- **Lipinski**: measure of drug-likeness (Lipinski's rule of five).
- **Time**: time cost to generate.

Results

Overall comparison:

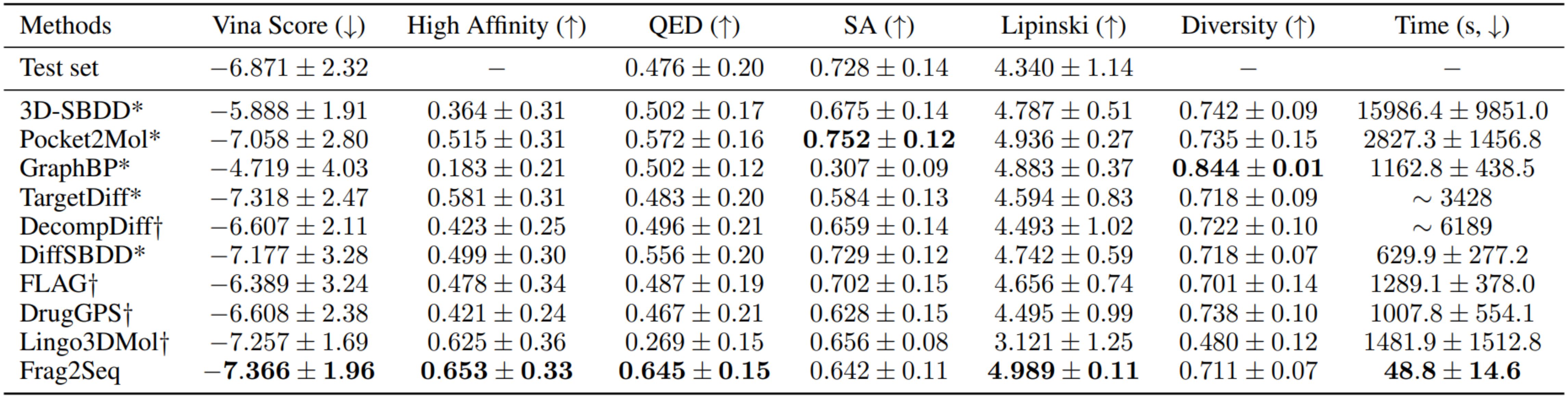
- Achieves better performance compared to baseline methods

Drug-likeness:

- **QED** and **Lipinski**
  - Frag2Seq generated molecules have better drug-like properties

Binding Affinity:

- **Vina score** and **High Affinity**
  - Frag2Seq method achieves the best binding affinity



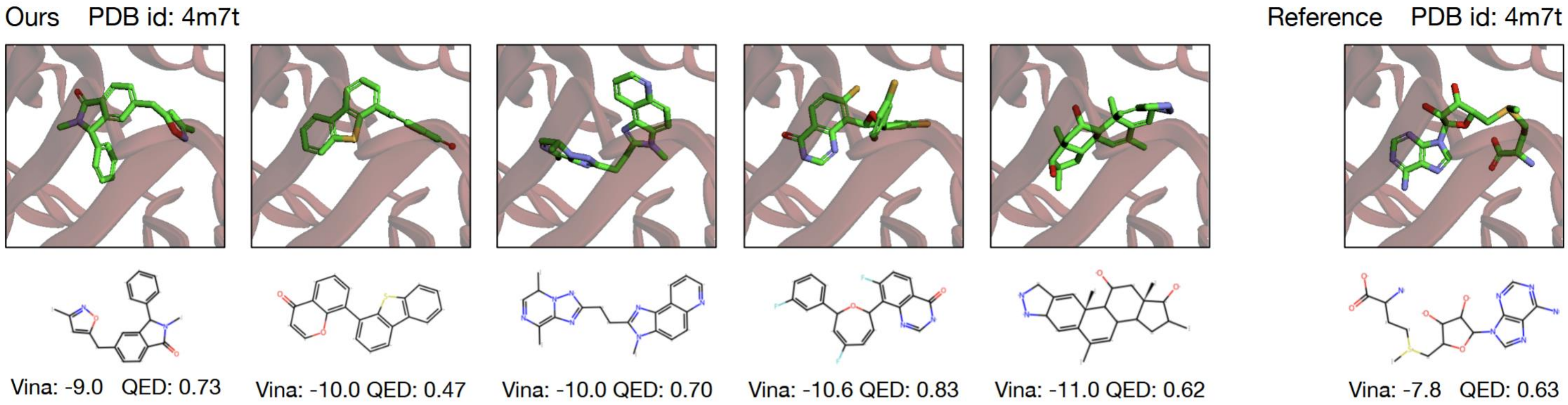
Methods	Vina Score (↓)	High Affinity (↑)	QED (↑)	SA (↑)	Lipinski (↑)	Diversity (↑)	Time (s, ↓)
Test set	−6.871 ± 2.32	–	0.476 ± 0.20	0.728 ± 0.14	4.340 ± 1.14	–	–
3D-SBDD*	−5.888 ± 1.91	0.364 ± 0.31	0.502 ± 0.17	0.675 ± 0.14	4.787 ± 0.51	0.742 ± 0.09	15986.4 ± 9851.0
Pocket2Mol*	−7.058 ± 2.80	0.515 ± 0.31	0.572 ± 0.16	**0.752 ± 0.12**	4.936 ± 0.27	0.735 ± 0.15	2827.3 ± 1456.8
GraphBP*	−4.719 ± 4.03	0.183 ± 0.21	0.502 ± 0.12	0.307 ± 0.09	4.883 ± 0.37	**0.844 ± 0.01**	1162.8 ± 438.5
TargetDiff*	−7.318 ± 2.47	0.581 ± 0.31	0.483 ± 0.20	0.584 ± 0.13	4.594 ± 0.83	0.718 ± 0.09	∼ 3428
DecompDiff†	−6.607 ± 2.11	0.423 ± 0.25	0.496 ± 0.21	0.659 ± 0.14	4.493 ± 1.02	0.722 ± 0.10	∼ 6189
DiffSBDD*	−7.177 ± 3.28	0.499 ± 0.30	0.556 ± 0.20	0.729 ± 0.12	4.742 ± 0.59	0.718 ± 0.07	629.9 ± 277.2
FLAG†	−6.389 ± 3.24	0.478 ± 0.34	0.487 ± 0.19	0.702 ± 0.15	4.656 ± 0.74	0.701 ± 0.14	1289.1 ± 378.0
DrugGPS†	−6.608 ± 2.38	0.421 ± 0.24	0.467 ± 0.21	0.628 ± 0.15	4.495 ± 0.99	0.738 ± 0.10	1007.8 ± 554.1
Lingo3DMol†	−7.257 ± 1.69	0.625 ± 0.36	0.269 ± 0.15	0.656 ± 0.08	3.121 ± 1.25	0.480 ± 0.12	1481.9 ± 1512.8
Frag2Seq	**−7.366 ± 1.96**	**0.653 ± 0.33**	**0.645 ± 0.15**	0.642 ± 0.11	**4.989 ± 0.11**	0.711 ± 0.07	**48.8 ± 14.6**

Results

Examples of generated 3D molecules for a specific protein pocket (PDB id: 4m7t):

- **Reference Molecule:** Provided in the test set.
- **Ours Molecule:** Generated by Frag2Seq.
  - **Vina:** lower, indicates higher binding affinity
  - **QED:** higher, indicates more drug-likeness

Confirms the method's effectiveness in protein-ligand interaction modeling and molecule generation.



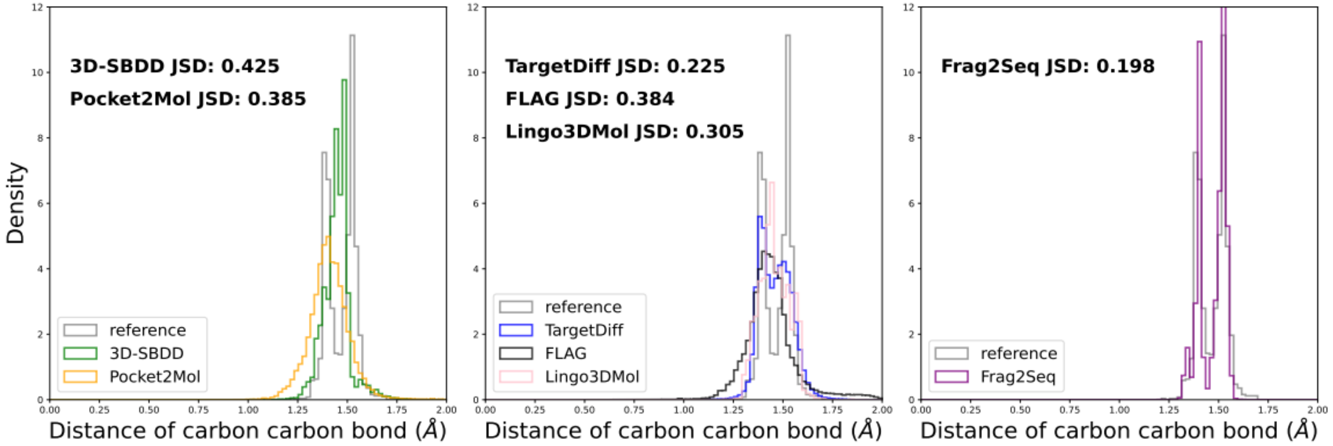
Results

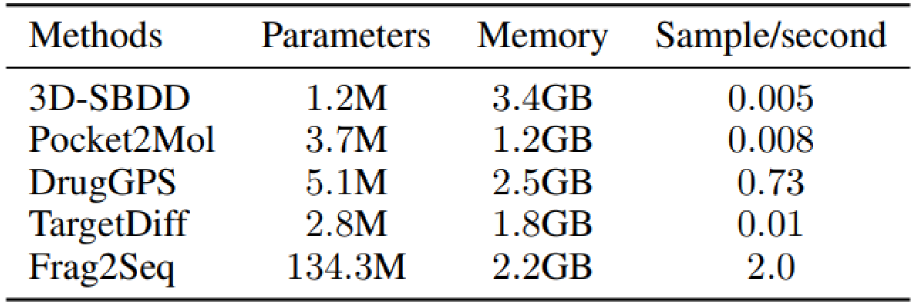
Empirical distribution of carbon-carbon bond distances analyze:

- **Reference Distribution:**
  - Exhibits **two** distinct modes.
- **Performance of Other Methods:**
  - **Most methods:** Can only capture **one** mode due to mode collapse.
  - **TargetDiff:** Exhibits two modes but suffers from the oversmoothness issue.
  - **Frag2Seq:** Better captures the two modes in the reference distribution.

Sampling Efficiency:

- Significantly better sampling efficiency than baseline methods.
- **Due to:** Simplified generation pipeline and Fragment-based generation strategy





Methods	Parameters	Memory	Sample/second
3D-SBDD	1.2M	3.4GB	0.005
Pocket2Mol	3.7M	1.2GB	0.008
DrugGPS	5.1M	2.5GB	0.73
TargetDiff	2.8M	1.8GB	0.01
Frag2Seq	134.3M	2.2GB	2.0

Summary

Strengths:

- **SE(3)-Invariant Tokenization Method**
  - Preserves important 3D geometric information
  - Mathematically rigorous proof
- **Fragment-Based Generation**
  - Reduces computational complexity
  - Enhances drug-likeness
- **Applying Language Models to Structure-Based Drug Design**
  - Novel integration of LLMs with SBDD

Weaknesses:

- **Simple Canonical SMILES-Based Sequence Construction**
  - Relies on Simple Canonical SMILES
    - Structurally similar molecules may have significantly different token representations.
- **Direct Cross-Attention Integration**
  - Lacks additional optimization strategies
  - Reduced interpretability