RETROINTEXT: A Multimodal Large Language Model Enhanced Framework For Retrosynthetic Planning Via In-Context Representation Learning

ICLR

The 13th International Conference on Learning Representation 2025

Chenglong Kang[1] Xiaoyi Liu[2*] Fei Guo[1*]

[1] School of Computer Science and Engineering, Central South University

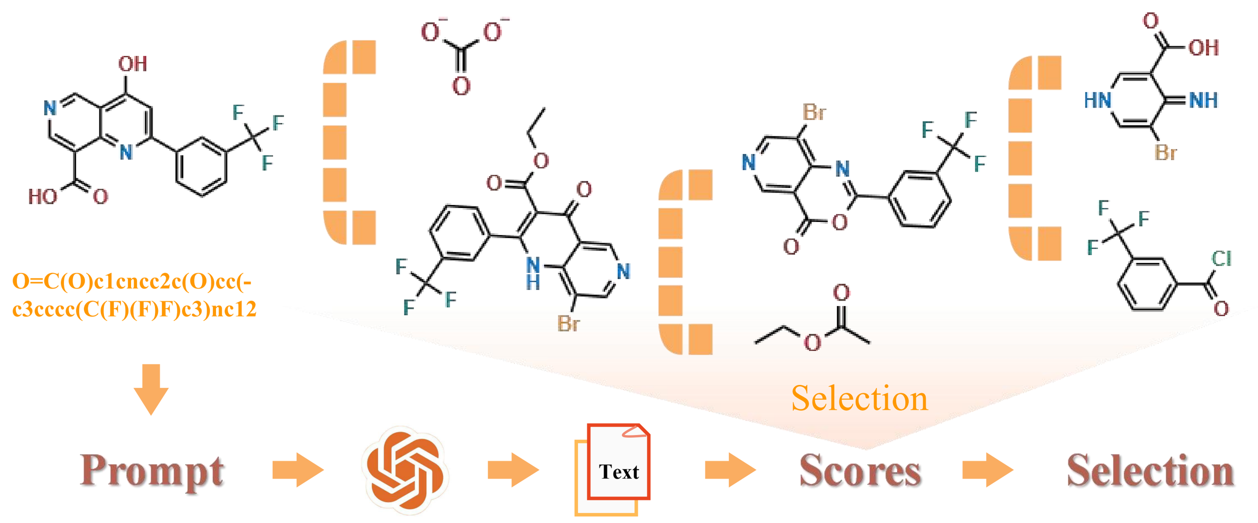
[2] School of Chinese Materia Medica, Beijing University of Chinese Medicine

* Corresponding Authors

Code

INTRODUCTION

Retrosynthetic planning: Finding retrosynthetic routes for a target molecule M_t from the starting material set S.



Contributions:

- We propose the RetroInText framework as a template-free approach for retrosynthesis prediction. When predicting subsequent steps in retrosynthesis, this framework integrates in-context textual information from previous steps.
- With RetroInText, we leverage the advantages of large language models (LLMs) and ChatGPT as generative models and evaluate reactions based on their molecular descriptions. A combination of textual information, molecular graphs, and 3D geometric information is used to select the optimal molecule in the selection phase.
- Extensive experiments demonstrate that RetroInText achieves competitive performance. Furthermore, RetroInText is tested in experiments to show its ability to predict complex reactions.

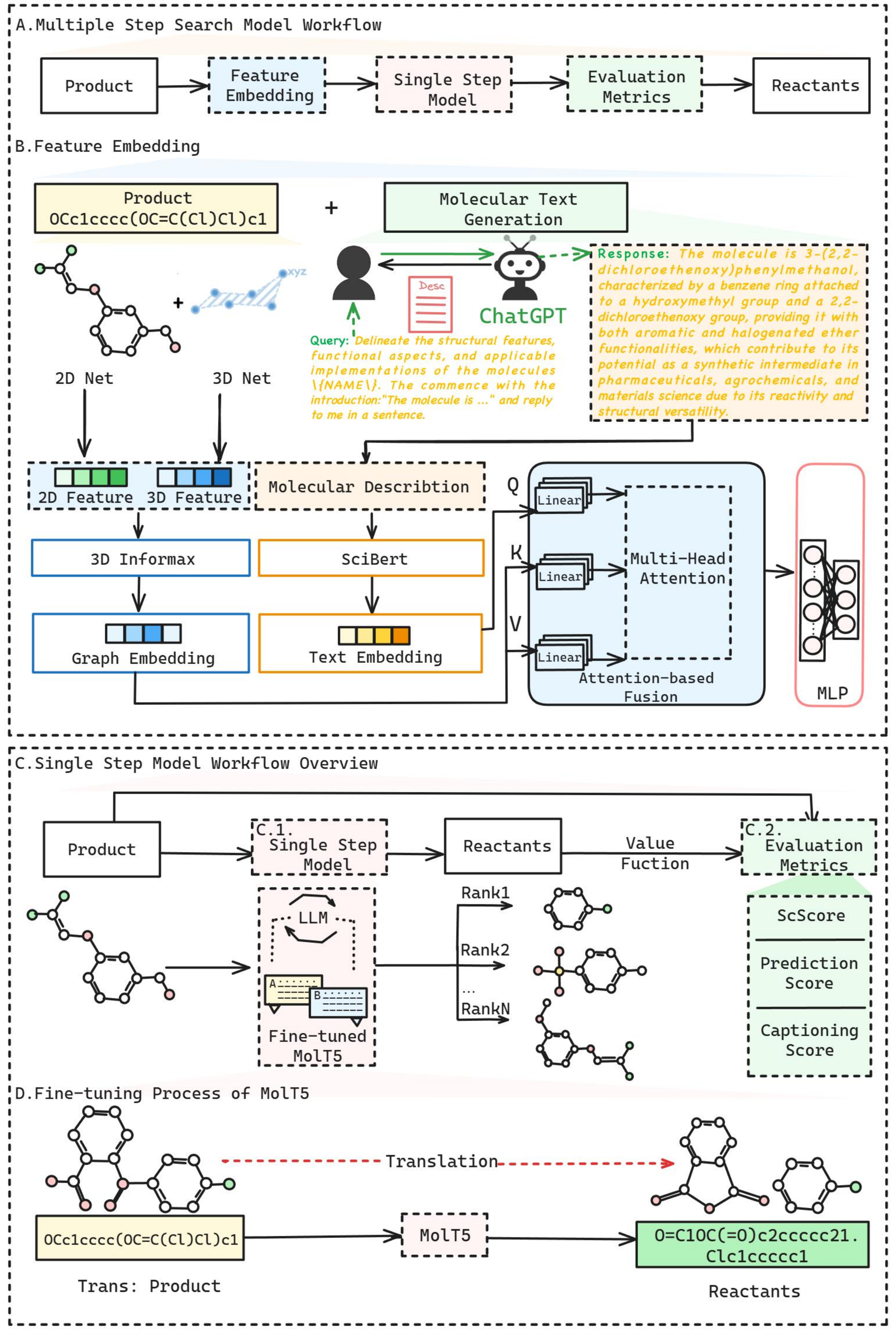
DATASET

For evaluation, we utilize the publicly available RetroBench dataset, which consists of 46,458 molecules for training, 5,803 for validation, and 5,838 for testing. The corresponding synthetic pathways for each molecule are extracted from the USPTO-full reaction network.

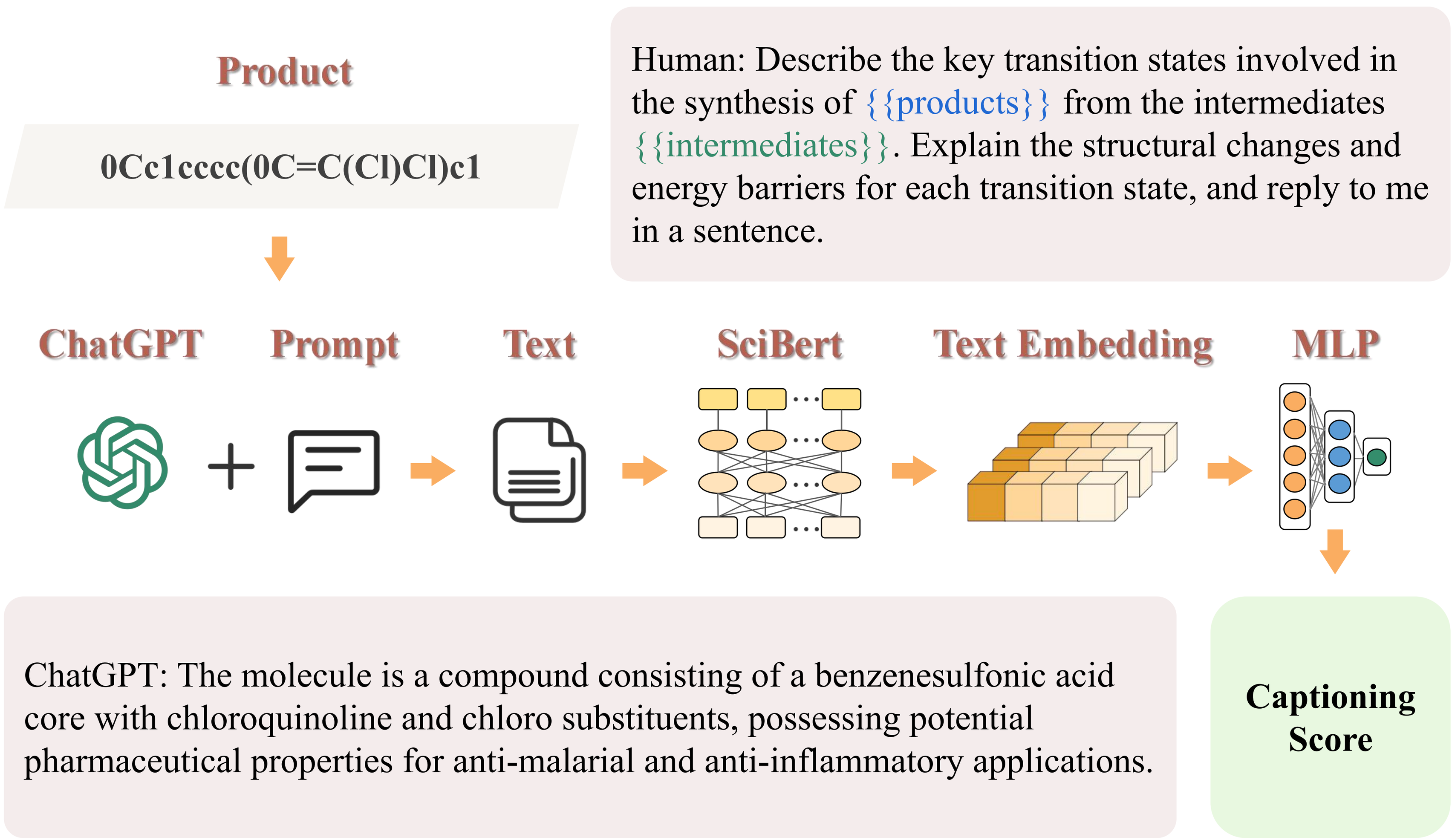
#Molecules / Depth — Datasets	2	3	4	5	6	7	8	9	10	11	12	13
Training	22,903	12,004	5,849	3,268	1,432	594	276	107	25	0	0	0
Validation	2,862	1,500	731	408	179	74	34	13	2	0	0	0
Test	2,862	1,500	731	408	179	74	34	13	2	32	2	1

Table 1: Statistics of molecules at various depths summarized from the RetroBench dataset.

METHODOLOGY



Overview of RetroInText: **A.** Multiple Step Search Model Workflow of RetroInText. **B.** Feature Embedding. The product is represented as a molecular graph and 3D geometry features. It is combined with text embeddings generated by ChatGPT and processed through SciBert for multimodal integration. **C.** Single Step Model Workflow. **C.1.** A fine-tuned MolT5 model generates potential reactants from the product, ranked by **C.2.** Evaluation Metrics. Reactants are evaluated using ScScore, captioning score, and prediction score to determine synthetic routes' quality and feasibility. **D.** MolT5 transforms the product SMILES into potential reactant structures.



Algorithm 1 Retrosynthesis Planning Algorithm

Input: target molecule M_t, starting material set S, textual information $\mathcal{T}$
Initialize: reactants set $\mathcal{R} = \{\}$, path set $\mathcal{P} = \{M_t\}$
while $\mathcal{P}$ is not empty **do**
  Take path p from $\mathcal{P}$, predict reactants $\mathcal{I}_p$ for expansion given p by $O(\cdot)$
  for reactant $\mathcal{I}_p^{(i)}$ in $\mathcal{I}_p$ **do**
    if $\mathcal{I}_p^{(i)} \in \mathcal{S}$ **then** Put $\mathcal{I}_p^{(i)}$ into $\mathcal{R}$
    else
      rank $p' = p + [\mathcal{I}_p^{(i)}]$ by computing captioning score of $\mathcal{T}$
      put ranked p' into $\mathcal{P}$
    end if
  end for
end while
return predicted reactant set $\mathcal{R}$

EXPERIMENTS

Search Algorithm / Single-step Models	Retro* Top-1	Retro* Top-2	Retro* Top-3	Retro* Top-4	Retro* Top-5	Retro*-0 Top-1	Retro*-0 Top-2	Retro*-0 Top-3	Retro*-0 Top-4	Retro*-0 Top-5	Greedy DFS Top-1
Template-based											
Retrosim (Coley et al., 2017)	35.1	40.5	42.9	44.0	44.6	35.0	40.5	43.0	44.1	44.6	31.5
Neuralsym (Segler & Waller, 2017)	**41.7**	**49.2**	52.1	53.6	54.4	**42.0**	**49.3**	52.0	53.6	54.3	**39.2**
GLN (Dai et al., 2019)	39.6	48.9	**52.7**	**54.6**	**55.7**	39.5	48.7	**52.6**	**54.5**	**55.6**	38.0
Semi-template-based											
G2Gs (Shi et al., 2020)	5.4	8.3	9.9	10.9	11.7	4.2	6.5	7.6	8.3	8.9	3.8
GraphRetro (Somnath et al., 2021)	15.3	19.5	21.0	21.9	22.4	15.3	19.5	21.0	21.9	22.2	**14.4**
GraphRetro+CREBM (Liu et al., 2024b)	**16.3**	**20.1**	**21.6**	**22.3**	**22.7**	**16.3**	**20.2**	**21.6**	**22.3**	**22.7**	-
Template-free											
Transformer (Karpov et al., 2019)	31.3	40.4	44.7	47.2	48.9	31.2	40.5	45.1	47.3	48.7	26.7
Transformer+CREBM	35.0	43.4	46.7	48.7	49.7	34.0	43.1	46.4	48.3	49.4	-
Megan (Sacha et al., 2021)	18.8	27.9	32.7	36.6	38.1	18.6	27.7	32.6	36.4	38.5	32.9
FusionRetro (Liu et al., 2023a)	37.5	45.0	48.3	50.6	51.5	37.4	45.0	48.4	50.4	51.1	35.2
FusionRetro+CREBM (Liu et al., 2024b)	39.4	46.6	49.3	50.7	51.5	39.6	46.7	49.5	51.0	51.7	33.8
RetroInText (Ours)	**41.2**	**48.7**	**51.8**	**53.3**	**54.2**	**42.1**	**49.9**	**53.0**	**54.7**	**55.7**	**39.8**

Table 2: Summary of retrosynthetic planning results for exact match accuracy (%).

Methods	Top-1	Top-2	Top-3	Top-4	Top-5
MolT5 (SMILES)	37.2	43.7	46.2	47.4	48.3
RetroInText(1D SMILES)	35.6	41.6	44.1	45.4	46.2
RetroInText(2D+3D Graph)	37.5	45.0	48.2	50.0	50.9
RetroInText(w/o text)	40.2	47.3	50.2	51.7	52.7
RetroInText	**41.2**	**48.7**	**51.8**	**53.3**	**54.2**

Table 3: Ablation study of RetroInText for exact match accuracy (%).

Depth	Retro*(w/o text) Top-1	Top-2	Top-3	Top-4	Top-5	Retro*(with text) Top-1	Top-2	Top-3	Top-4	Top-5
Depth2	**45.0**	**52.4**	55.4	57.2	58.3	44.9	52.3	**55.4**	**57.3**	**58.3**
Depth3	38.9	45.9	49.3	50.5	51.5	**40.0**	**47.9**	**51.5**	**53.0**	**53.9**
Depth4	33.7	40.9	42.5	43.6	43.6	**36.1**	**43.6**	**46.4**	**47.7**	**48.3**
Depth5	35.5	41.7	43.4	44.4	44.4	**39.0**	**47.8**	**50.3**	**51.2**	**51.7**
Depth6	33.0	36.3	36.9	38.0	38.0	**36.3**	**40.8**	**41.9**	**43.0**	**44.1**
Depth7	25.7	31.1	31.1	31.1	31.1	**28.4**	**33.8**	**35.1**	**35.1**	**35.1**
Depth8	29.4	41.2	41.2	41.2	41.2	**32.4**	**41.2**	**44.1**	**47.1**	**47.1**

Table 4: Exact match accuracy (%) at different depths of ground truth synthetic routes.

CONCLUSIONS

- RetroInText integrates multimodal data by combining contextual information from ChatGPT, molecular structure, and 3D data through MolT5, a large language model, for innovative retrosynthetic planning.
- Experimental results show that RetroInText outperforms existing methods on the RetroBench dataset, achieving state-of-the-art performance.