Appendix

A Experiment setting

ChemDual is implemented using Pytorch v2.4.1 and CUDA 12.1. In the training phase, all experiments are conducted on the same machine with Intel Xeon(R) Platinum 8352V CPU @ 2.10GHz and 4 GPUs (NVIDIA RTX 4090 GPUs 24G) for a total of 212 GPU hours and processed 852,518,047 tokens. We use the AdamW optimizer to facilitate the optimization process. Each GPU was assigned a batch size of 2, with gradient accumulation steps set to 8. The pre-training phase was conducted for 1 epoch with a learning rate of 5e-5, which enable stable and gradual learning during the initial stages. The fine-tuning phase was carried out over 2 epochs with a learning rate of 1e-4. In the inference phase, the model was quantized to INT4 and ran on a single 4090 GPU at a speed of 98.16 tokens/s.

B Experiment on ChemLLMBench

As shown in Table 3, ChemDual demonstrates superior performance on both reaction prediction and retrosynthesis tasks within the ChemLLMBench. Specifically, ChemDual achieves the highest accuracy of 67.0 in reaction prediction, outperforming ChemDFM (49.0), BioT5+ (9.0), and Mol-Instruction (4.5) by a substantial margin. Similarly, in retrosynthesis prediction, ChemDual achieves an accuracy of 33.0, which also surpasses other baselines, indicating the model's improved reasoning ability on backward chemical tasks. While maintaining competitive validity scores across both tasks, ChemDual exhibits a strong balance between correctness and chemical plausibility, highlighting the effectiveness of the dual-task learning framework in enhancing structural understanding and generalization in complex chemical scenarios.

Table 3: Performance comparison on ChemLLMBench.

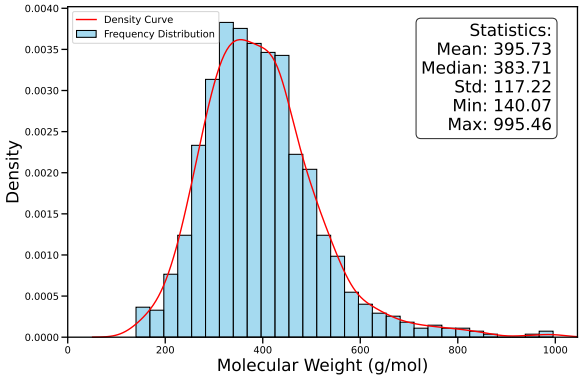
Task	Model	Accuracy	Validity
Reaction Prediction	Mol-Instruction [Fang *et al.*, 2024]	4.5	**100.0**
	BioT5+ [Pei *et al.*, 2024]	9.0	**100.0**
	ChemDFM [Zhao *et al.*, 2025]	49.0	**98.0**
	ChemDual	**67.0**	99.0
Retrosynthesis	Mol-Instruction [Fang *et al.*, 2024]	9.0	**100.0**
	BioT5+ [Pei *et al.*, 2024]	26.0	**100.0**
	ChemDFM [Zhao *et al.*, 2025]	12.0	**91.0**
	ChemDual	**33.0**	97.7

C Experiment on fragment and recombination

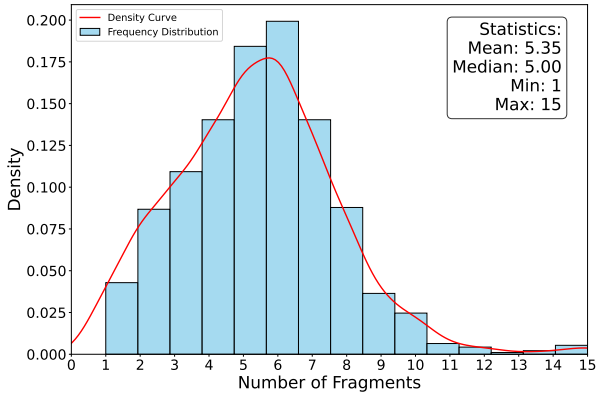
Result on fragment and recombination. The results are presented in Figures 7a and 7b. In the recombination task, the model generated 996 valid molecules with most molecular weights (54.89%) falling in the 320–480 g/mol range. Notably, 79.94% of the molecules were within the 250–500 g/mol range, considered drug-like and suggesting good drug-likeness in generated compounds. This reflects the model's effectiveness in producing compounds with high drug potential. For the fragment task, the model primarily generated fragment counts between 1 and 10, with fragmentation over 10 parts occurring in fewer than 2% of cases. This indicates the model's capacity to fragment molecules into manageable parts while also handling larger molecules when needed, demonstrating versatility in fragmenting diverse molecular sizes.

D Case study on molecule recombination

The molecules generated by ChemDual are shown in Figure 8, arranged in the order of their generation as determined by the docking score. The generated compounds exhibit a high degree of similarity to original molecule in terms of molecular fingerprints, reflecting the effective inheritance of key structural and chemical features from the original fragments.



(a) Distribution of molecular weights after recombination. The histogram reveals that the molecular weights of recombined products primarily fall within the 320–480 g/mol range, accounting for 54.89% of the molecules. 79.94% of the molecules lie within the 250–500 g/mol range, which is considered optimal for drug-likeness.



(b) Distribution of molecular fragment numbers. The histogram shows that most molecules were fragmented into 1–10 parts, with the highest frequencies observed between 5 and 7 fragments. Fragment counts above 10 are rare but present, with less than 2% of molecules exceeding this number.

Figure 7: Analysis of molecular properties.

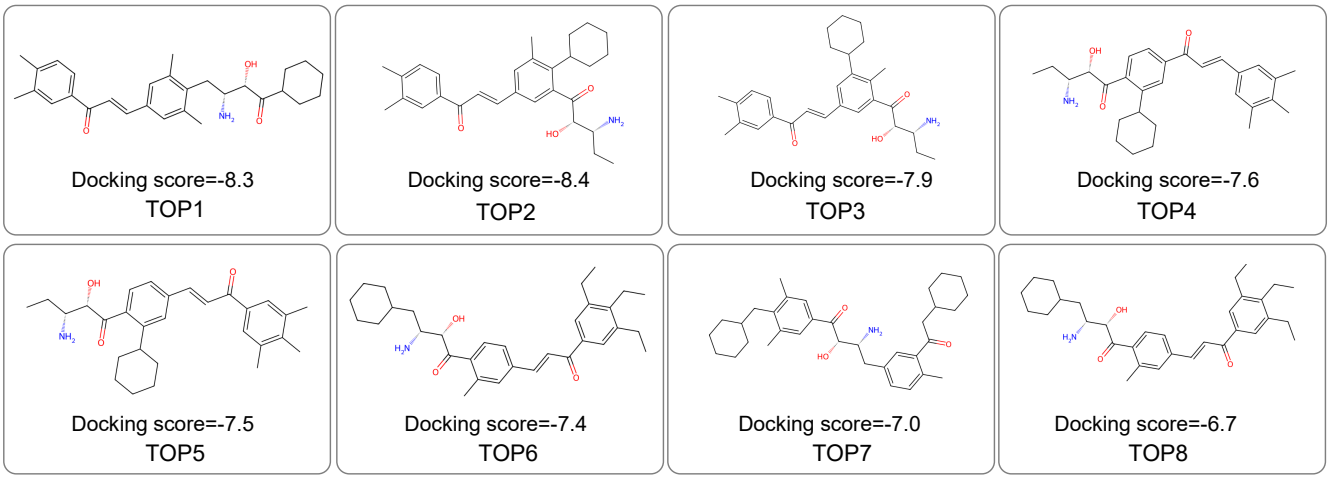


Figure 8: Different molecules generated by ChemDual. Docking score represent the maximum binding affinities (kcal/mol) between the ligands and MAP2 obtained through molecular docking.