ESDM: EFFICIENT SPECTRUM-CONDITIONED MOLECULAR DIFFUSION MODELS FOR MOLECULAR STRUCTURE ELUCIDATION

Anonymous authors
Paper under double-blind review

ABSTRACT

De novo molecular generation from tandem mass spectrometry (MS/MS) has become an important direction for unknown compound identification and molecular discovery. Diffusion-based methods have demonstrated strong molecular reconstruction capability by modeling conditional graph generation processes. However, their inference procedure still relies on long sequential reverse diffusion, resulting in substantial computational overhead and limited practical applicability in high-throughput scenarios. To address this issue, we propose an efficient inference framework for diffusion-based molecular generation conditioned on mass spectra, incorporating two lightweight inference optimization modules: Sampling Acceleration and Sampling Corrector. The Sampling Acceleration module reduces redundant denoising evaluations through jump-step reverse diffusion and non-uniform timestep scheduling, while the Sampling Corrector improves sampling stability via lightweight posterior refinement on edge logits. Since both Sampling Acceleration and Sampling Corrector are implemented exclusively during inference, the proposed framework is potentially applicable to other diffusion-based molecular generation models without requiring retraining. Experimental results on the NPLIB1 and MassSpecGym benchmarks demonstrate that ESDM significantly reduces inference cost (reducing inference time by up to 58.9% on NPLIB1 and 62.8% on MassSpecGym) while maintaining competitive molecular reconstruction performance. Compared with DiffMS, ESDM achieves a favorable balance between generation quality and computational efficiency under multiple accelerated sampling configurations, highlighting its practical value for large-scale molecular generation and high-throughput molecular annotation.

1 INTRODUCTION

In metabolomics, drug discovery, and environmental exposure analysis, reconstructing molecular structures from mass spectrometry (MS) data is critical for bridging analytical signals and chemical interpretation Hufsky et al. (2014); Hu et al. (2024); Gentry et al. (2024). Compared with spectral library matching, de novo methods directly generate candidate structures from tandem mass spectrometry (MS/MS) data, offering greater potential for identifying unknown compounds, underrepresented chemical classes, and novel natural products Stravs et al. (2021); Schneider et al. (2026). This task can be formulated as a high-dimensional inverse problem that infers molecular graphs from fragment observations by jointly modeling spectral representations, chemical constraints, and combinatorial search Mqawass et al. (2026). Its practical value lies in translating unknown peaks into testable structural hypotheses for downstream experimental design and decision-making Nie & Gao (2026); Dorigatti et al. (2025).

However, direct molecular generation from MS/MS remains challenging. ***First***, fragmentation spectra are intrinsically incomplete and ambiguous: structurally distinct molecules, particularly isomers, may produce similar spectra, while the same molecule can vary across collision energies, ionization modes, and instruments Schneider et al. (2026); Nie & Gao (2026). ***Second***, the enormous chemical space requires models to balance structural validity, spectral consistency, and generative diversity Mqawass et al. (2026).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Existing studies have developed several paradigms to address these challenges. Early approaches such as CSI:FingerID Dührkop et al. (2015) follow spectrum-to-fingerprint or retrieval pipelines, but their dependence on predefined candidate libraries limits identification of truly novel molecules. This motivated de novo generation methods such as MSNovelist Stravs (2024), which predicts molecular fingerprints before generating structures. Although this reduces reliance on reference databases, reconstruction remains sensitive to intermediate fingerprint errors that can propagate through decoding. End-to-end methods such as MassGenie Shrivastava et al. (2021a) and Spec2Mol Litsa et al. (2021) therefore directly map spectra to molecular representations without handcrafted intermediate descriptors or multi-stage prediction pipelines Neo et al. (2025). Together, these advances have shifted molecular annotation from in-database identification toward out-of-database generation Schneider et al. (2026), enabling broader chemical-space exploration and applications such as metabolite identification and natural product discovery. As such systems move toward large-scale analytical workflows, both generation quality and computational efficiency become increasingly important.

Despite this progress, the spectrum-to-structure mapping remains inherently ambiguous, as a single spectrum may correspond to multiple chemically plausible structures. To better capture this uncertainty and structural diversity, diffusion-based generative models have recently been introduced into spectrum-conditioned molecular generation. DiffMS Bohde et al. (2025), for example, formulates molecular reconstruction as conditional graph diffusion and achieves state-of-the-art performance on multiple benchmarks. However, its fixed and relatively long reverse diffusion process may perform redundant denoising evaluations, while multiple-candidate generation further increases inference latency Yang et al. (2023).

This overhead mainly arises from iterative reverse diffusion, whose inference cost grows approximately linearly with the number of sampling steps. Fixed timestep schedules also allocate identical computational budgets to samples of different reconstruction difficulty. Simply reducing sampling steps without reliable cross-step transition modeling can introduce posterior estimation errors and unstable reverse transitions, degrading reconstruction quality Ikeda et al. (2025); Ho et al. (2020); Song et al. (2021). Therefore, the key challenge is to reduce sampling cost while preserving both sampling stability and competitive reconstruction performance Karras et al. (2022); Song et al. (2023).

To address this challenge, we propose ESDM, an efficient spectrum-conditioned molecular diffusion framework built upon DiffMS. ESDM introduces two complementary techniques: Sampling Acceleration and Sampling Corrector. Sampling Acceleration reduces denoising evaluations through non-uniform timestep scheduling and jump-step reverse diffusion, while Sampling Corrector performs lightweight posterior refinement on edge logits to mitigate structural instability caused by large timestep intervals. Together, they improve inference efficiency and sampling stability while maintaining competitive reconstruction performance, with a lightweight and modular design that facilitates integration into broader diffusion-based molecular generation frameworks.

The main contributions of this work are summarized as follows:

1. We propose ESDM, an efficient spectrum-conditioned molecular diffusion framework that jointly improves inference efficiency and sampling stability while maintaining competitive reconstruction performance.
2. We design a Sampling Acceleration module using non-uniform timestep scheduling and jump-step reverse diffusion to reduce redundant denoising evaluations.
3. We develop a lightweight Sampling Corrector that refines posterior predictions during accelerated sampling to improve reverse diffusion stability without additional denoising evaluations.

2 Background and Related Work

2.1 De Novo Molecular Structure Generation from MS/MS

De novo molecular structure generation from tandem mass spectrometry (MS/MS) aims to infer molecular structures directly from fragmentation spectra without relying on reference databases Bohde et al. (2025); Stravs et al. (2021); Schneider et al. (2026), offering greater potential than library-

search-based annotation for identifying unknown compounds and novel chemical entities. However, the task is inherently challenging because spectrum-to-structure mapping is ambiguous: different molecules, especially structural isomers, may produce similar fragmentation patterns, while incomplete and noisy spectra further complicate reconstruction Hufsky et al. (2014); Dührkop et al. (2021); **?**. The combinatorial growth of molecular search space further makes exhaustive exploration infeasible.

Early approaches relied mainly on spectrum retrieval and fingerprint prediction. CSI:FingerID Dührkop et al. (2015); Hájek et al. (2023) predicts molecular fingerprints from spectra and retrieves candidate structures from databases, but its performance remains constrained by database coverage. MSNovelist Stravs et al. (2021) combines fingerprint prediction with neural structure generation, although errors in intermediate representations may propagate during reconstruction. More recent end-to-end methods, including Spec2Mol Litsa et al. (2021), MassGenie Shrivastava et al. (2021a), and FlowMS Nie & Gao (2026), directly learn spectrum-to-molecule mappings with neural generative models, reducing dependence on intermediate predictions and extending exploration beyond reference databases Litsa et al. (2021); Neo et al. (2025).

2.2 Diffusion-Based Molecular Generation

Diffusion models have shown strong capability in molecular generation by modeling complex structural dependencies and molecular topologies Yang et al. (2023); Ho et al. (2020). DiffMS Bohde et al. (2025) extends this paradigm to spectrum-conditioned molecular generation by formulating reconstruction as conditional discrete graph diffusion with a transformer-based spectrum encoder and GraphTransformer decoder, achieving state-of-the-art performance on several benchmarks. However, repeated reverse denoising introduces substantial inference cost, with latency growing approximately linearly with the number of sampling timesteps Yang et al. (2023); Ikeda et al. (2025), limiting scalability in high-throughput annotation.

2.3 Inference Acceleration for Diffusion Models

Existing diffusion acceleration methods mainly follow three directions: reducing sampling steps with non-Markovian methods such as DDIM Song et al. (2021), using efficient diffusion ODE solvers such as DPM-Solver Lu et al. (2022), and compressing multi-step denoising through progressive distillation or consistency learning Salimans & Ho (2022); Song et al. (2023). However, these methods are primarily designed for continuous diffusion in Euclidean domains and do not directly transfer to discrete molecular graph diffusion, where discrete graph transitions and chemical validity impose additional constraints. Efficient inference for spectrum-conditioned discrete graph diffusion therefore remains challenging. We address this gap through inference-only sampling acceleration and posterior refinement to improve diffusion efficiency and stability.

3 Method

3.1 Problem Formulation

Given an input tandem mass spectrum $\mathbf{s}$, the objective of spectrum-conditioned molecular generation is to reconstruct the corresponding molecular structure conditioned on spectral observations. Following DiffMS Bohde et al. (2025), molecular structures are represented as discrete molecular graphs $G = (\mathbf{X}, \mathbf{E})$, where $\mathbf{X} \in \mathbb{R}^{N \times d_X}$ denotes atom representations and $\mathbf{E} \in \mathbb{R}^{\mathbb{N} \times \mathbb{N} \times \mathbb{E}}$ denotes bond representations. Here, N denotes the maximum number of atoms, while d_X and d_E indicate the dimensions of atom and bond feature spaces, respectively.

A spectrum encoder first extracts conditional information from the input spectrum $\mathbf{c} = E(\mathbf{s})$, where $\mathbf{s}$ denotes the input mass spectrum, $E(\cdot)$ denotes the spectrum encoder, and $\mathbf{c}$ represents the spectrum-conditioned latent representation. Following the discrete graph diffusion framework, a forward diffusion process gradually perturbs molecular graphs into noisy latent states: $\mathbf{z}_t = (\mathbf{X}_t, \mathbf{E}_t)$, where $t \in [0, T]$ denotes the diffusion timestep and T represents the total number of diffusion steps, and $\mathbf{z}_0 = G$. A denoising network parameterized by θ learns the reverse diffusion transition conditioned on spectral information: $p_\theta\left(\mathbf{z}_{t-1)} \mid \mathbf{z}_t, \mathbf{c}\right)$. The objective of this work is to improve reverse diffusion efficiency while preserving molecular reconstruction quality.

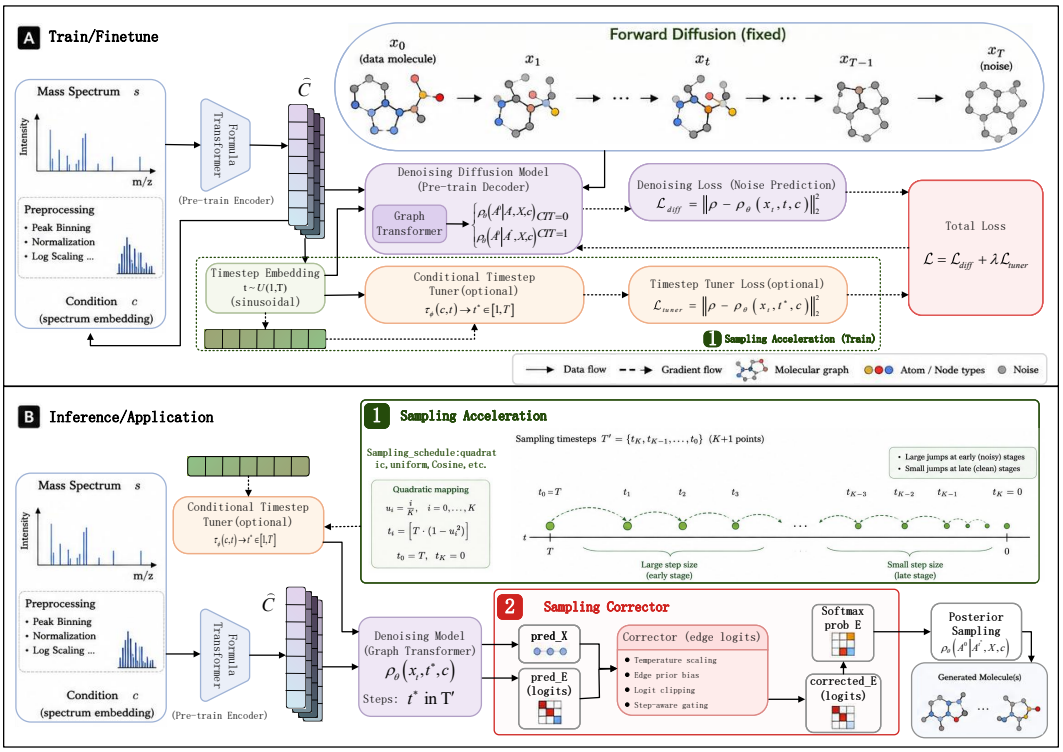


Figure 1: Overview of the proposed ESDM framework. (A) In the training stage, the Conditional Timestep Tuner learns spectrum-conditioned timestep preferences jointly with the diffusion denoising model. (B) In the inference stage, ESDM consists of two components: **1) Sampling Acceleration**, which constructs a reduced timestep trajectory for jump-step reverse diffusion, and **2) Sampling Corrector**, which refines edge predictions before posterior sampling.

3.2 Framework of ESDM

ESDM is built upon the conditional discrete graph diffusion framework of DiffMS and focuses on improving inference efficiency during reverse diffusion sampling. DiffMS adopts a standard sequential reverse diffusion procedure:

$$\mathbf{z}_T \rightarrow \mathbf{z}_{T-1} \rightarrow \ldots \rightarrow \mathbf{z}_0, \tag{1}$$

where molecular graphs are iteratively reconstructed through multiple denoising transitions. Although this iterative process provides strong molecular generation capability, repeated denoising evaluations introduce considerable computational overhead. Since each reverse transition requires decoder(Graph Transformer) inference, inference cost grows approximately linearly with the sampling length. This limitation becomes increasingly critical in large-scale molecular annotation and deployment-oriented scenarios. Moreover, directly reducing sampling steps often introduces posterior approximation errors during accelerated reverse diffusion, particularly for bond prediction, resulting in unstable molecular structures and degraded reconstruction quality.

To address these issues, ESDM introduces two complementary components: Sampling Acceleration and Sampling Corrector. Sampling Acceleration reduces redundant denoising evaluations through non-uniform timestep scheduling and jump-step reverse diffusion. Sampling Corrector further improves accelerated sampling stability through lightweight posterior refinement on bond prediction logits. By jointly integrating these two components, ESDM improves diffusion inference efficiency while preserving competitive molecular reconstruction capability.

The overall framework consists of a training stage and an inference stage, as illustrated in Fig. 1. During training (Fig. 1-A), a Conditional Timestep Tuner is introduced to learn spectrum-dependent timestep schedules from spectral conditions and diffusion states. The learned tuner is then deployed

Algorithm 1 Sampling Acceleration & Sampling Corrector

Input: Mass spectrum s, sampling steps K, diffusion model ε_θ
Output: Generated molecular graph G

$\mathbf{c} \leftarrow E_s(s)$
$\mathbf{z}_T \sim p(\mathbf{z}_T)$
$\mathcal{T}' = \{t_K, \ldots, t_0\} \leftarrow \mathrm{BuildSchedule}(T, K)$
for $(t, u) \in \mathcal{T}'$ **do**
  $\mathrm{predX}, \mathrm{predE}_{\mathrm{logits}} \leftarrow \varepsilon_\theta(\mathbf{z}_t, t, \mathbf{c})$
  if `use_sampling_corrector` **then**
    $\mathrm{predE}_{\mathrm{logits}} \leftarrow \mathrm{Corrector}(\mathrm{predE}_{\mathrm{logits}}, t)$
  end if
  $\mathbf{z}_u \sim p(\mathbf{z}_u \mid \mathbf{z}_t, \mathrm{predX}, \mathrm{predE}_{\mathrm{logits}})$
  $\mathbf{z}_t \leftarrow \mathbf{z}_u$
end for
return G

during inference (Fig. 1-B) to construct reduced timestep trajectories for Sampling Acceleration. Based on these trajectories, Sampling Corrector further refines bond predictions to improve generation stability under accelerated diffusion.

3.3 Sampling Acceleration

To improve inference efficiency, ESDM introduces a Sampling Acceleration module consisting of a Conditional Timestep Tuner (CTT) and an adaptive schedule construction strategy. During training (Fig. 1-A-1), the CTT is optimized using the proposed tuner loss to learn condition-aware sampling difficulty estimation. During inference (Fig. 1-B-1), the trained CTT generates a condition-dependent timestep allocation, which is subsequently used by the schedule construction strategy to build a shortened sampling schedule for accelerated denoising. Unlike DiffMS, which relies on a fixed diffusion schedule, the proposed tuner learns sample-specific timestep preferences conditioned on the input spectrum and diffusion state. Specifically, the tuner takes the spectrum embedding and timestep embedding as input and predicts adaptive timestep positions that are optimized jointly with the diffusion denoising objective through an auxiliary timestep tuning loss. During inference, the learned timestep schedules can be directly used to construct reduced sampling trajectories, providing a data-driven foundation for the proposed Sampling Acceleration strategy. And instead of traversing all diffusion timesteps, the reverse process is conducted over a subset of timesteps:

$$T^{'} = \{t_K, t_{K-1}, \ldots, t_0\}, \tag{2}$$

where $t_K = T, t_0 = 0, K \approx (T/8, T/5)$ and T is the total number of diffusion timesteps, K is the number of accelerated sampling steps used in ESDM. Accordingly, the reverse diffusion process becomes:

$$\mathbf{z}_{t_K} \rightarrow \mathbf{z}_{t_{K-1}} \rightarrow \ldots \rightarrow \mathbf{z}_{t_0}, \tag{3}$$

and the pseudocode of the inference procedure is shown in Algorithm 1. Therefore, the proposed jump-step reverse diffusion strategy enables the model to traverse only a subset of timesteps, substantially reducing the number of denoising evaluations and improving inference efficiency.

In our implementation, the timestep schedule is constructed before sampling through the function: $BuildSchedule(\mathbb{S}, T, K)$, where $\mathbb{S}$ denotes a selectable predefined strategy, such as quadratic, cosine, etc, which generates non-uniform(except for the default strategy uniform) reverse timesteps according to the predefined sampling schedule. For example, when ESDM adopts a quadratic timestep schedule to allocate more denoising steps to the late-stage diffusion process. Specifically, the timestep positions are generated as:

$$u_i = \frac{i}{K}, \quad t_i = T\left(1 - u_i^2\right). \tag{4}$$

This design allows larger timestep skipping during the early denoising stage while preserving finer reverse transitions near the final reconstruction stage. Since molecular structures become increasingly sensitive to edge prediction errors during late-stage denoising, the quadratic schedule improves structural refinement stability under accelerated sampling.

Unlike standard Denoising Diffusion Probabilistic Model (DDPM) Ho et al. (2020) sampling, accelerated reverse diffusion in ESDM requires transitions between non-adjacent timesteps. Specifically, instead of only performing: $\mathbf{z}_t \rightarrow \mathbf{z}_{t-1}$, ESDM allows direct transitions: $\mathbf{z}_t \rightarrow \mathbf{z}_s, \ \ s < t$, where t and s denote two arbitrary timesteps in the accelerated schedule T.

To support timestep skipping, ESDM constructs the corresponding forward transition matrix between arbitrary timesteps. Let:

$$\overline{\alpha}_t = \prod_{i=1}^{t} \alpha_i, \tag{5}$$

denote the cumulative noise retention coefficient at timestep t. The conditional retention coefficient between timestep s and timestep t is defined as:

$$\alpha_{t|s} = \frac{\overline{\alpha}_t}{\overline{\alpha}_s}. \tag{6}$$

Following the discrete diffusion formulation used in DiffMS, the transition matrix between timestep s and timestep t is constructed as:

$$Q_{t|s} = \alpha_{t|s} I + \left(1 - \alpha_{t|s}\right) K, \tag{7}$$

where I denotes the identity transition matrix and K denotes the categorical noise prior matrix. This formulation enables diffusion-consistent jump-step transitions under non-uniform timestep intervals.

By reducing the number of denoising evaluations from T to K, the computational complexity of sampling decreases from $O(T \cdot C_d)$ to $O(K \cdot C_d)$, where C_d denotes the computational cost of a single Graph Transformer denoising operation. Therefore, the theoretical acceleration ratio of ESDM is T/K.

3.4 Sampling Corrector

Although the Sampling Acceleration module significantly reduces the number of reverse diffusion evaluations, aggressive timestep skipping may introduce distribution distortion during reverse diffusion, particularly for edge generation. In practice, large timestep intervals tend to produce unstable edge posterior distributions, leading to over-confident bond predictions, invalid molecular structures, and degraded structural consistency during late-stage denoising.

To alleviate this issue, ESDM introduces a lightweight Sampling Corrector module that performs logit-level posterior refinement during reverse diffusion. Different from additional denoising networks or iterative refinement strategies, the proposed corrector directly operates on the predicted edge logits $\widehat{E}_{logits}$, generated by the GraphTransformer denoiser $\varepsilon_\theta\left(z_t, t, c\right)$. The corrected edge logits are denoted as:

$$\widetilde{E}_{logits} = \mathcal{C}\left(\widehat{E}_{logits}, t\right), \tag{8}$$

where $\mathcal{C}()$ represents the correction operation conditioned on the current timestep t.

The Sampling Corrector mainly consists of three lightweight operations: temperature calibration, prior-guided edge biasing, and logit clipping. First, temperature scaling is applied to adjust the sharpness of the posterior distribution:

$$\widehat{E}_{logits} \leftarrow \frac{\widehat{E}_{logits}}{\tau}, \tag{9}$$

where τ denotes the temperature coefficient. This operation prevents over-confident edge predictions caused by accelerated timestep skipping and improves posterior smoothness during late-stage denoising. Second, ESDM introduces a prior-guided correction term to stabilize edge generation:

$$\widehat{E}_{logits} \leftarrow \widehat{E}_{logits} + \lambda P_{edge}, \tag{10}$$

where P_{edge} denotes the edge prior distribution and λ controls the correction strength. This operation biases the posterior distribution toward chemically plausible edge patterns and suppresses unrealistic bond configurations under accelerated sampling. Afterward, logit clipping is performed to avoid numerical instability caused by excessively large posterior logits:

$$\widehat{E}_{logits} \leftarrow clip\left(\widehat{E}_{logits}, -c, c\right), \tag{11}$$

where c denotes the clipping threshold. This operation stabilizes the subsequent softmax normalization and reduces posterior saturation during reverse diffusion. Unlike globally applied refinement strategies, the Sampling Corrector in ESDM is only activated during selected reverse diffusion stages. Specifically, correction is performed only when the current timestep satisfies $t < t_{corr}$, where t_{corr} denotes the maximum timestep for correction. In addition, the corrector is executed every n reverse diffusion steps to further reduce computational overhead. This design is motivated by the observation that early-stage diffusion states mainly contain random noise, while late-stage denoising is considerably more sensitive to edge prediction quality. Finally, the corrected logits $\widetilde{E}_{logits}$ are used for posterior sampling:

$$\mathbf{z}_s \sim p\left(\mathbf{z}_s \mid \mathbf{z}_t, predX, \widetilde{E}_{logits}\right), \tag{12}$$

thereby improving molecular validity and structural stability under accelerated reverse diffusion.

Since the Sampling Corrector only involves lightweight tensor operations without additional denoising evaluations, its computational overhead is negligible compared with the GraphTransformer inference cost. Consequently, ESDM can effectively stabilize accelerated sampling while preserving high inference efficiency. The complete inference procedure of ESDM, including both Sampling Acceleration and Sampling Corrector, is summarized in Algorithm 1, corresponding to the inference stage shown in Fig. 1-B.

4 Experiments

4.1 Experimental Settings

Datasets: We evaluate ESDM on two widely used MS/MS-based molecular identification benchmarks: NPLIB1 and MassSpecGym. NPLIB1 contains high-quality reference spectra collected from natural products, while MassSpecGym provides a large-scale benchmark for spectrum-conditioned molecular structure elucidation. The datasets are publicly available at the corresponding repositories[1][2].

Baselines: We compare ESDM with representative spectrum-conditioned molecular generation methods, including retrieval-based, sequence-based, graph-based, and diffusion-based approaches. Specifically, we consider Spec2Mol Schneider et al. (2026), MADGEN Shrivastava et al. (2021b), MIST+NeuralDecipher Hájek et al. (2023), MIST+MSNovelist Stravs (2024), and DiffMS Bohde et al. (2025) as baselines.

Evaluation Metrics: Following previous works Bohde et al. (2025); Bushuiev et al. (2024), we adopt Top-1 accuracy, Top-10 accuracy, and Tanimoto similarity as the main evaluation metrics. In addition, inference time and sampling steps are further reported to evaluate the acceleration capability of ESDM.

Implementation Details: ESDM is implemented based on the official DiffMS framework while keeping the original GraphTransformer denoiser and training strategy unchanged. During inference, accelerated timestep schedules are constructed using the Sampling Acceleration module, and the Sampling Corrector is enabled during reverse diffusion. Unless otherwise specified, the quadratic timestep schedule is adopted by default. On NPLIB1, all experiments were conducted using a single 12GB NVIDIA RTX 2080Ti GPU, following the original DiffMS experimental setting. Due to the substantially larger scale of the MassSpecGym dataset, experiments on MassSpecGym were performed using four 12GB NVIDIA RTX 2080Ti GPUs for efficient evaluation. Our code is available at GitHub.

4.2 Main Results

Table 1 summarizes the de novo structural elucidation performance on NPLIB1 and MassSpecGym.

On NPLIB1, ESDM substantially reduces inference cost while preserving competitive generation quality. Compared with the reproduced DiffMS baseline, the 150/500→100 configuration improves

[1] https://zenodo.org/record/8316682/files/canopus_train_export_v2.tar

[2] https://zenodo.org/records/15008938/files/msg_preprocessed.tar.gz

Table 1: De novo structural elucidation performance on the NPLIB1 Dührkop et al. (2021) and MassSpecGym Bushuiev et al. (2024) datasets. Bold and underlined values indicate the best and second-best results, respectively. $*$ denotes baseline results reproduced in our experimental environment; other results are adopted from existing literature or benchmarks.

Model	Top-1		Top-10		Time	Sample/Step
	Accuracy	Tanimoto	Accuracy	Tanimoto		
NPLIB1						
Spec2Mol	0.00%	0.12	0.00%	0.16	–	–
MADGEN	2.10%	0.22	2.39%	0.27	–	–
MIST + Neuraldecipher	2.32%	**0.35**	6.11%	0.43	–	–
MIST + MSNovelist	5.40%	<u>0.34</u>	11.04%	**0.44**	–	–
DiffMS*	<u>7.098%</u>	0.324	13.574%	0.430	61h 30m 39s	100/500
ESDM(Ours)	5.600%	0.318	12.204%	0.420	14h 10m 21s	100/500→q(80)
	6.475%	0.320	**15.442%**	<u>0.435</u>	16h 17m 28s	120/500→q(80)
	7.596%	0.326	<u>14.944%</u>	<u>0.435</u>	25h 17m 03s	150/500→q(100)
MassSpecGym						
SMILES Transformer	0.00%	0.03	0.00%	0.10	–	–
MIST + MSNovelist	0.00%	0.06	0.00%	0.15	–	–
SELFIES Transformer	0.00%	0.08	0.00%	0.13	–	–
Spec2Mol	0.00%	0.12	0.00%	0.16	–	–
MIST + Neuraldecipher	0.00%	0.14	0.00%	0.16	–	–
MADGEN	1.31%	0.20	1.54%	0.26	–	–
DiffMS*	<u>1.703%</u>	<u>0.232</u>	<u>3.716%</u>	<u>0.331</u>	347h 51m 3s	100/500
ESDM(Ours)	**2.019%**	**0.286**	**4.176%**	**0.349**	129h 30m 23s	150/500→q(120)

Top-1 accuracy from 7.098% to 7.596% and Tanimoto similarity from 0.324 to 0.326, while reducing inference time from 61h 30m 39s to 25h 17m 03s, corresponding to a 58.9% reduction and approximately $2.43\times$ acceleration. Under more aggressive acceleration, the 120/500→80 configuration achieves the best Top-10 accuracy of 15.442%, exceeding DiffMS at 13.574%, while using fewer denoising evaluations. The fastest 100/500→80 configuration further reduces inference overhead while maintaining competitive Tanimoto similarity.

On MassSpecGym, ESDM shows a similar speed-quality trade-off. Compared with DiffMS, it improves Top-1 accuracy from 1.703% to 2.019% and Top-10 Tanimoto similarity from 0.331 to 0.349, while reducing inference time from 347h 51m 03s to 129h 30m 23s. This corresponds to a 62.8% reduction in inference cost and approximately $2.69\times$ acceleration. These results indicate that ESDM consistently reduces the high inference overhead of diffusion-based molecular generation while maintaining or improving reconstruction quality.

4.3 Ablation Study

Effect of Sampling Acceleration: Comparing fs100_quadratic with fs100_uniform under the same sampling budget shows that quadratic scheduling improves both Top-1 accuracy (5.230%→5.604%) and Tanimoto similarity (0.306→0.318), supporting finer allocation of denoising steps to later reverse stages. Comparing fs120_quadratic_C and fs150_quadratic_C further shows that a larger sampling budget improves reconstruction quality (6.475%→7.596% Top-1 accuracy) at the cost of increased runtime. Under the aggressive 10-sample setting, fs10_quadratic_C outperforms fs10_cosine_C (3.860% vs. 3.486% Top-1 accuracy), indicating that the quadratic schedule is more effective than the cosine schedule under the same 80-step budget.

Effect of Sampling Corrector: The effect of the Sampling Corrector is isolated by comparing fs10_quadratic_C with fs10_quadratic on NPLIB1, where enabling the corrector improves Top-1 accuracy from 2.739% to 3.860% and Tanimoto similarity from 0.261 to 0.272 with only a small runtime increase. On MassSpecGym, msg_fs10_TT_C also improves over msg_fs10_TT in both Top-1 accuracy (0.280%→0.293%) and Tanimoto similarity (0.189→0.190). For the full-step setting, msg_fs10_C slightly improves Top-1 accuracy over msg_fs10_org (0.295%→0.298%) but decreases

Table 2: Ablation study of ESDM on the NPLIB1 dataset and MSG dataset. In 'Sample/Step", the value before the slash indicates the number of generated molecular samples per spectrum, while the value after the slash specifies the timestep schedule used for reverse diffusion. Specifically, $u(N)$ denotes a uniform timestep schedule with N sampling steps, and $q(N)$ denotes a quadratic timestep schedule with N sampling steps. 'Corrector" indicates whether the Sampling Corrector is enabled to refine bond prediction logits under accelerated sampling. The results demonstrate the individual and combined contributions of the two proposed components.

	Top 1 Acc/Top 1 Tan	Time	Sample/Step	Corrector
NPLIB1				
fs100_uniform	5.230% / 0.306	14h 14m 40s	100/u(80)	×
fs100_quadratic	5.604% / 0.318	14h 10m 21s	100/q(80)	×
fs120_quadratic_C	6.475% / 0.320	16h 17m 28s	120/q(80)	✓
fs150_quadratic_C	7.596% / 0.326	25h 17m 03s	150/q(100)	✓
fs10_quadratic_C	3.860% / 0.272	2h 5m 12s	10/q(80)	✓
fs10_quadratic	2.739% / 0.261	1h 59m 15s	10/q(80)	×
fs10_cosine_C	3.486% / 0.266	2h 15m 17s	10/c(80)	✓
MassSpecGym				
msg_fs10_TT_C	0.293% / 0.190	31h 08m 10s	10/q(80)	✓
msg_fs10_TT	0.280% / 0.189	30h 32m 04s	10/q(80)	×
msg_fs10_C	0.298% / 0.192	122h 49m 39s	10/u(500)	✓
msg_fs10_org	0.295% / 0.197	121h 04m 27s	10/u(500)	×

Tanimoto similarity (0.197→0.192), suggesting that the benefit of posterior correction is more evident under accelerated sampling.

4.4 Hyperparameter Sensitivity Analysis

We further investigate the sensitivity of the Sampling Corrector to four key hyperparameters, including the temperature coefficient τ, prior-guidance weight λ, clipping threshold c, and correction timestep t_{corr}. The results show that ESDM maintains stable performance across a broad range of parameter configurations, indicating the robustness of the proposed correction mechanism. Detailed optimization procedures and parameter-wise analysis are provided in Appendix A.1.

4.5 Case Study

To qualitatively evaluate the molecular structures generated by ESDM, we visualize representative reconstruction examples on the NPLIB1 dataset. The results show that ESDM can recover exact molecular structures or chemically similar candidates under accelerated sampling. More qualitative examples are presented in Appendix A.2.

5 Conclusion and Future Work

In this work, we propose ESDM, an efficient inference framework for diffusion-based molecular generation conditioned on mass spectra. By introducing Sampling Acceleration and Sampling Corrector, ESDM reduces redundant reverse diffusion evaluations while improving accelerated sampling stability. Experimental results on NPLIB1 and MassSpecGym demonstrate that ESDM significantly reduces inference cost while maintaining competitive molecular reconstruction performance. Compared with DiffMS, ESDM achieves favorable speed-quality tradeoff under multiple accelerated sampling configurations. In future work, we plan to further explore adaptive timestep scheduling strategies and chemically-aware correction mechanisms to improve accelerated molecular generation quality. Extending ESDM to larger molecular discovery benchmarks and Multimodal generation tasks is also a promising direction.

AI USE STATEMENT

Generative AI tools were used to assist with language polishing and text condensation during manuscript preparation. The authors reviewed and verified all AI-assisted content and take full responsibility for the final manuscript. The AI tools were not used to generate experimental data, conduct experiments, or independently determine the scientific conclusions of this work.

REFERENCES

Montgomery Bohde, Mrunali Manjrekar, Runzhong Wang, Shuiwang Ji, and Connor W. Coley. DiffMS: Diffusion generation of molecules conditioned on mass spectra. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 4737–4756. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/bohde25a.html.

Roman Bushuiev, Anton Bushuiev, Niek F. de Jonge, Adamo Young, Fleming Kretschmer, Raman Samusevich, Janne Heirman, Fei Wang, Luke Zhang, Kai Dührkop, Marcus Ludwig, Nils A. Haupt, Apurva Kalia, Corinna Brungs, Robin Schmid, Russell Greiner, Bo Wang, David S. Wishart, Li-Ping Liu, Juho Rousu, Wout Bittremieux, Hannes Rost, Tytus D. Mak, Soha Hassoun, Florian Huber, Justin J.J. van der Hooft, Michael A. Stravs, Sebastian Böcker, Josef Sivic, and Tomáš Pluskal. Massspecgym: A benchmark for the discovery and identification of molecules. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 110010–110027. Curran Associates, Inc., 2024. doi: 10.52202/079017-3491. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/c6c31413d5c53b7d1c343c1498734b0f-Paper-Datasets_and_Benchmarks_Track.pdf.

Emilio Dorigatti, Jonathan Groß, Jonas Kühlborn, Robert Möckel, Frank Maier, and Julian Keupp. Enhancing automated drug substance impurity structure elucidation from tandem mass spectra through transfer learning and domain knowledge††electronic supplementary information (esi) available. see doi: https://doi.org/10.1039/d5dd00115c. *Digital Discovery*, 4(9):2454–2464, 2025. ISSN 2635-098X. doi: https://doi.org/10.1039/d5dd00115c. URL https://www.sciencedirect.com/science/article/pii/S2635098X25001573.

Kai Dührkop, Huibin Shen, Marvin Meusel, Juho Rousu, and Sebastian Böcker. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Science*, 112(41):12580–12585, October 2015. doi: 10.1073/pnas.1509788112.

Kai Dührkop, {Louis Félix} Nothias, Markus Fleischauer, Raphael Reher, Marcus Ludwig, {Martin A.} Hoffmann, Daniel Petras, {William H.} Gerwick, Juho Rousu, {Pieter C.} Dorrestein, and Sebastian Böcker. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nature Biotechnology*, 39(4):462–471, April 2021. ISSN 1087-0156. doi: 10.1038/s41587-020-0740-8.

Emily C. Gentry, Stephanie L. Collins, Morgan Panitchpakdi, Pedro Belda-Ferre, Allison K. Stewart, Marvic Carrillo Terrazas, Hsueh-han Lu, Simone Zuffa, Tingting Yan, Julian Avila-Pacheco, Damian R. Plichta, Allegra T. Aron, Mingxun Wang, Alan K. Jarmusch, Fuhua Hao, Mashette Syrkin-Nikolau, Hera Vlamakis, Ashwin N. Ananthakrishnan, Brigid S. Boland, Amy Hemperly, Niels Vande Casteele, Frank J. Gonzalez, Clary B. Clish, Ramnik J. Xavier, Hiutung Chu, Erin S. Baker, Andrew D. Patterson, Rob Knight, Dionicio Siegel, and Pieter C. Dorrestein. Reverse metabolomics for the discovery of chemical structures from humans. *Nature*, 626(7998):419–426, February 2024. doi: 10.1038/s41586-023-06906-8.

Adam Hájek, Michal Starý, Filip Jozefov, Helge Hecht, Elliott Price, and Aleš Křenek. De-novo Identification of Small Molecules from Their GC-EI-MS Spectra. *arXiv e-prints*, art. arXiv:2304.01634, April 2023. doi: 10.48550/arXiv.2304.01634.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. *arXiv e-prints*, art. arXiv:2006.11239, June 2020. doi: 10.48550/arXiv.2006.11239.

Maowei Hu, Lei Yang, Nathaniel Twarog, Jason Ochoada, Yong Li, Eirinaios I Vrettos, Arnaldo X Torres-Hernandez, James B Martinez, Jiya Bhatia, Brandon M Young, Jeanine Price, Kevin McGowan, Theresa H Nguyen, Zhe Shi, Matthew Anyanwu, Mary Ashley Rimmer, Shea Mercer, Zoran Rankovic, Anang A Shelat, and Daniel J Blair. Continuous collective analysis of chemical reactions. *Nature*, 636(8042):374—379, December 2024. ISSN 0028-0836. doi: 10.1038/s41586-024-08211-4. URL https://europepmc.org/articles/PMC11823688.

Franziska Hufsky, Kerstin Scheubert, and Sebastian Böcker. Computational mass spectrometry for small-molecule fragmentation. *TrAC Trends in Analytical Chemistry*, 53:41–48, 2014. ISSN 0165-9936. doi: https://doi.org/10.1016/j.trac.2013.09.008. URL https://www.sciencedirect.com/science/article/pii/S016599361300232X.

Kotaro Ikeda, Tomoya Uda, Daisuke Okanohara, and Sosuke Ito. Speed-accuracy relations for diffusion models: Wisdom from nonequilibrium thermodynamics and optimal transport. *Phys. Rev. X*, 15:031031, Jul 2025. doi: 10.1103/x5vj-8jq9. URL https://link.aps.org/doi/10.1103/x5vj-8jq9.

Tero Karras, Miika Aittala, Samuli Laine, and Timo Aila. Elucidating the design space of diffusion-based generative models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Eleni Litsa, Vijil Chenthamarakshan, Payel Das, and Lydia Kavraki. Spec2mol: An end-to-end deep learning framework for translating ms/ms spectra to de-novo molecules. *ChemRxiv*, 2021 (0913), 2021. doi: 10.26434/chemrxiv-2021-6rdh6. URL https://chemrxiv.org/doi/abs/10.26434/chemrxiv-2021-6rdh6.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: a fast ode solver for diffusion probabilistic model sampling in around 10 steps. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Ghaith Mqawass, Tuan-Dung Le, Fabian J. Theis, and Djork-Arné Clevert. De novo molecular structure elucidation from mass spectra via flow matching. *ArXiv*, abs/2602.19912, 2026. URL https://api.semanticscholar.org/CorpusID:285973322.

Neng Kai Nigel Neo, Lim Jing, Ngoui Yong Zhau Preston, Koh Xue Ting Serene, and Bingquan Shen. One Small Step with Fingerprints, One Giant Leap for De Novo Molecule Generation from Mass Spectra. *arXiv e-prints*, art. arXiv:2508.04180, August 2025. doi: 10.48550/arXiv.2508.04180.

Jianan Nie and Peng Gao. FlowMS: Flow matching for de novo structure elucidation from mass spectra. In *ICLR 2026 Workshop on Foundation Models for Science: Real-World Impact and Science-First Design*, 2026. URL https://openreview.net/forum?id=7H0fSu6gz1.

Tim Salimans and Jonathan Ho. Progressive Distillation for Fast Sampling of Diffusion Models. *arXiv e-prints*, art. arXiv:2202.00512, February 2022. doi: 10.48550/arXiv.2202.00512.

Mark Yu. Schneider, Daniil D. Kholmanskikh, Kirill Ya. Romanov, Elena A. Perekina, Sergei A. Nikolenko, Ruslan Yu. Lukin, and Ivan V. Golov. De novo structure prediction from tandem mass spectra: Algorithms, benchmarks, and limitations. *Molecules*, 31(5), 2026. ISSN 1420-3049. URL https://www.mdpi.com/1420-3049/31/5/769.

Aditya Divyakant Shrivastava, Neil Swainston, Soumitra Samanta, Ivayla Roberts, Marina Wright Muelas, and Douglas B. Kell. Massgenie: a transformer-based deep learning method for identifying small molecules from their mass spectra. *bioRxiv*, 2021a. doi: 10.1101/2021.06.25.449969. URL https://www.biorxiv.org/content/early/2021/06/26/2021.06.25.449969.

Aditya Divyakant Shrivastava, Neil Swainston, Soumitra Samanta, Ivayla Roberts, Marina Wright Muelas, and Douglas B. Kell. Massgenie: A transformer-based deep learning method

for identifying small molecules from their mass spectra. *Biomolecules*, 11(12), 2021b. ISSN 2218-273X. doi: 10.3390/biom11121793. URL https://www.mdpi.com/2218-273X/11/12/1793.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=St1giarCHLP.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency Models. *arXiv e-prints*, art. arXiv:2303.01469, March 2023. doi: 10.48550/arXiv.2303.01469.

Michael A. Stravs. A decade of computational mass spectrometry from reference spectra to deep learning. *CHIMIA*, 78(7-8):525–530, Aug. 2024. doi: 10.2533/chimia.2024.525. URL https://www.chimia.ch/chimia/article/view/2024_525.

Michael A. Stravs, Kai Dührkop, Sebastian Böcker, and Nicola Zamboni. Msnovelist: De novo structure generation from mass spectra. *bioRxiv*, 2021. doi: 10.1101/2021.07.06.450875. URL https://www.biorxiv.org/content/early/2021/07/07/2021.07.06.450875.

Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Comput. Surv.*, 56(4), November 2023. ISSN 0360-0300. doi: 10.1145/3626235. URL https://doi.org/10.1145/3626235.

A Appendix

A.1 Hyperparameter Sensitivity Analysis

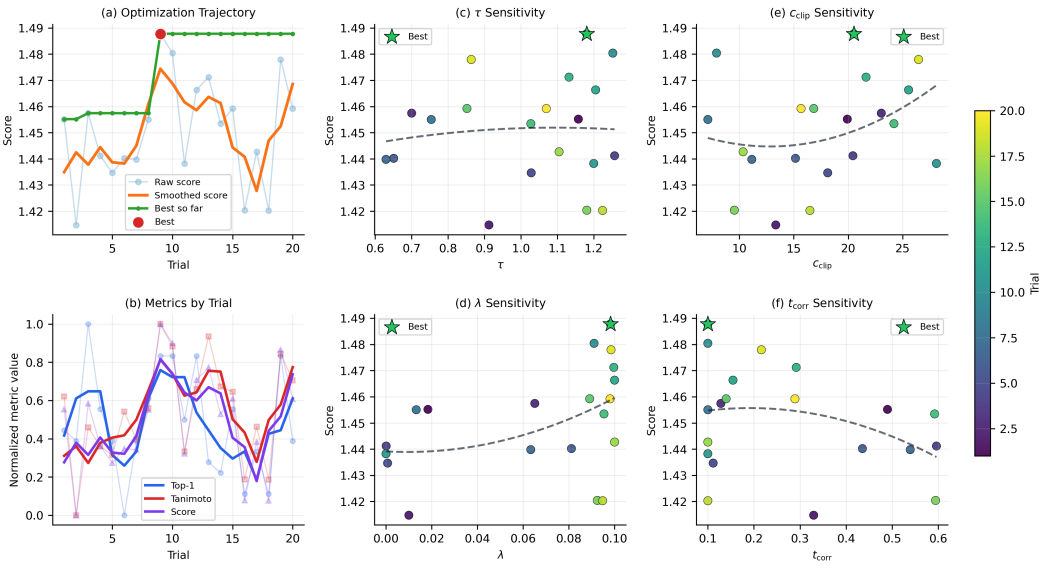


Figure 2: Hyperparameter Analysis.

We further investigate the sensitivity of ESDM to four key hyperparameters introduced in the Sampling Corrector, namely the temperature coefficient τ, the prior-guidance weight λ, the clipping threshold c_{clip}, and the correction activation timestep t_{corr}. Hyperparameter optimization is conducted on the CANOPUS dataset using Gaussian-process-based Bayesian optimization over the predefined search space. Following the evaluation protocol, each candidate configuration is assessed in

test-only mode with 10 generated molecular samples per spectrum. To jointly consider reconstruction accuracy and structural similarity, a composite score is adopted as the optimization objective: $Score = Top - 1 + 5 \times Tanimoto$.

The optimization results are summarized in Fig. 2. Fig. 2(a) presents the optimization trajectory across 20 trials. The raw score, smoothed score, and best-so-far curve indicate that the search converges rapidly and identifies a near-optimal region after approximately the first half of the evaluations. Fig. 2(b) shows the evolution of Top-1 accuracy, Tanimoto similarity, and the composite score throughout the optimization process. Although fluctuations are observed among different trials, the overall trend remains stable, suggesting that the proposed Sampling Corrector is reasonably robust to moderate hyperparameter variations.

Figs. 2(c)–(f) further analyze the influence of individual hyperparameters on the final score. The best-performing configuration is obtained at $\tau = 1.18$, $\lambda = 0.098$, $c_{clip} = 20.52$, and $t_{corr} = 0.10$, achieving a composite score of approximately 1.49. As shown in Fig. 2(c), moderate-to-high values of τ generally lead to better performance, indicating that an appropriate degree of posterior smoothing helps stabilize the denoising process. Fig. 2(d) reveals a positive correlation between λ and the optimization score, suggesting that stronger structural prior guidance can effectively improve molecular reconstruction quality.

Regarding the clipping threshold, Fig. 2(e) shows that excessively small clipping values tend to restrict the corrective effect, whereas moderately large values around 20–25 yield more favorable results. Fig. 2(f) demonstrates that activating the correction mechanism at earlier denoising stages generally produces higher scores than late-stage activation. This observation indicates that early structural guidance is more effective for steering the reverse diffusion trajectory toward chemically plausible molecular structures.

Overall, the optimization landscape exhibits relatively smooth behavior without abrupt performance degradation across the explored parameter ranges. These results demonstrate that the proposed Sampling Corrector is not overly sensitive to hyperparameter selection and that a balanced combination of posterior smoothing, prior-guided correction, and logit clipping contributes to improved molecular generation performance.

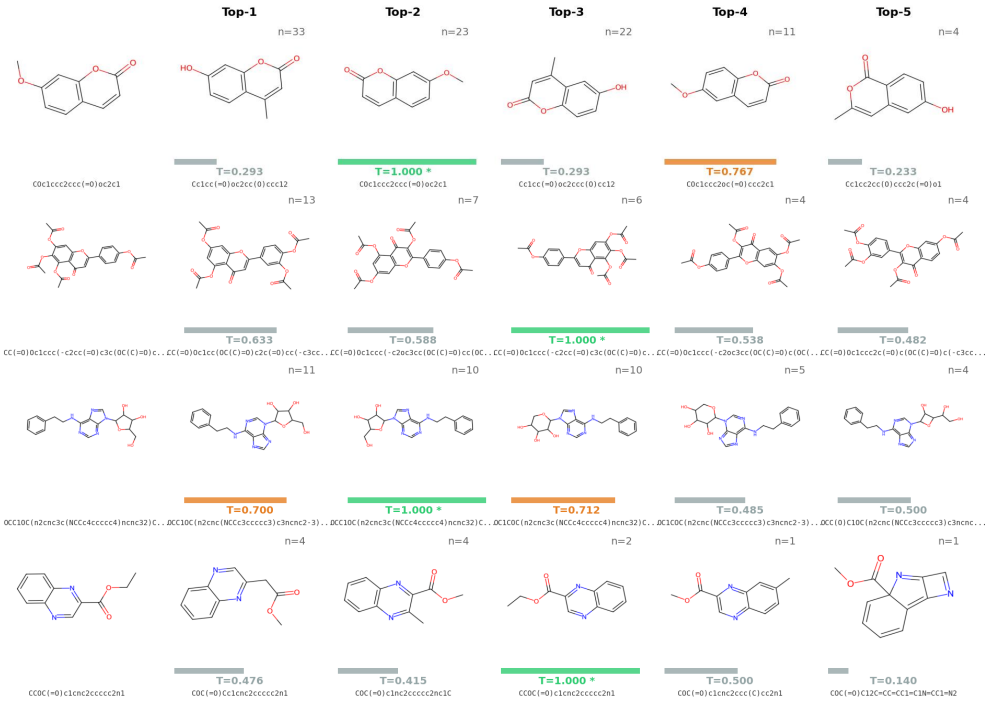
A.2 Additional Case Study

To further investigate the qualitative behavior of ESDM, we conduct a case study on representative samples from the NPLIB1 test set. Figure 3 presents several molecular reconstruction examples generated by ESDM under accelerated diffusion sampling.

For each spectrum, the leftmost molecule corresponds to the ground-truth structure, while the remaining molecules represent the Top-5 generated candidates ranked according to their occurrence frequency across repeated sampling runs. The value (n) above each candidate denotes the number of times that the corresponding molecular structure appears during repeated generations. The colored bars indicate the similarity between the generated structure and the reference molecule, where green denotes an exact match, orange indicates a highly similar structure with Tanimoto similarity greater than 0.675, and gray represents lower-similarity candidates.

Several observations can be drawn from these examples. First, ESDM is capable of accurately recovering the target molecular structure for a subset of samples even under accelerated sampling settings, demonstrating that the proposed acceleration strategy preserves essential spectrum-to-structure mapping information. Second, when exact reconstruction is not achieved, many generated candidates still exhibit high structural similarity to the ground-truth molecule, suggesting that the generated molecular graphs capture meaningful chemical substructures and functional group patterns. Third, the occurrence frequencies indicate that chemically plausible structures are repeatedly sampled, reflecting the stability of the proposed Sampling Corrector during accelerated reverse diffusion.

Overall, the qualitative results demonstrate that ESDM can maintain competitive molecular reconstruction capability while substantially reducing inference cost. The generated candidates remain chemically meaningful and structurally consistent with the underlying spectral observations, supporting the effectiveness of the proposed acceleration and correction mechanisms.


Top-1
Top-2
Top-3
Top-4
Top-5
n=33
n=23
n=22
n=11
n=4
T=0.293
T=1.000 *
T=0.293
T=0.767
T=0.233
COc1ccc2ccc(=O)oc2c1
Cc1cc(=O)oc2cc(O)ccc12
COc1ccc2ccc(=O)oc2c1
Cc1cc(=O)oc2ccc(O)cc12
COc1ccc2oc(=O)ccc2c1
Cc1cc2cc(O)ccc2c(=O)o1
n=13
n=7
n=6
n=4
n=4
T=0.633
T=0.588
T=1.000 *
T=0.538
T=0.482
CC(=O)Oc1ccc(-c2cc(=O)c3c(OC(C)=O)c...
CC(=O)Oc1cc(OC(C)=O)c2c(=O)cc(-c3cc...
CC(=O)Oc1ccc(-c2oc3cc(OC(C)=O)cc(OC...
CC(=O)Oc1ccc(-c2cc(=O)c3c(OC(C)=O)c...
CC(=O)Oc1ccc(-c2oc3cc(OC(C)=O)c(OC(...
CC(=O)Oc1ccc2c(=O)c(OC(C)=O)c(-c3cc...
n=11
n=10
n=10
n=5
n=4
T=0.700
T=1.000 *
T=0.712
T=0.485
T=0.500
OCC1OC(n2cnc3c(NCCc4ccccc4)ncnc32)C...
OCC1OC(n2cnc(NCCc3ccccc3)c3ncnc2-3)...
OCC1OC(n2cnc3c(NCCc4ccccc4)ncnc32)C...
OC1COC(n2cnc3c(NCCc4ccccc4)ncnc32)C...
OC1COC(n2cnc(NCCc3ccccc3)c3ncnc2-3)...
OCC(O)C1OC(n2cnc(NCCc3ccccc3)c3ncnc...
n=4
n=4
n=2
n=1
n=1
T=0.476
T=0.415
T=1.000 *
T=0.500
T=0.140
CCOC(=O)c1cnc2ccccc2n1
COC(=O)Cc1cnc2ccccc2n1
COC(=O)c1nc2ccccc2nc1C
CCOC(=O)c1cnc2ccccc2n1
COC(=O)c1cnc2ccc(C)cc2n1
COC(=O)C12C=CC=CC1=C1N=CC1=N2

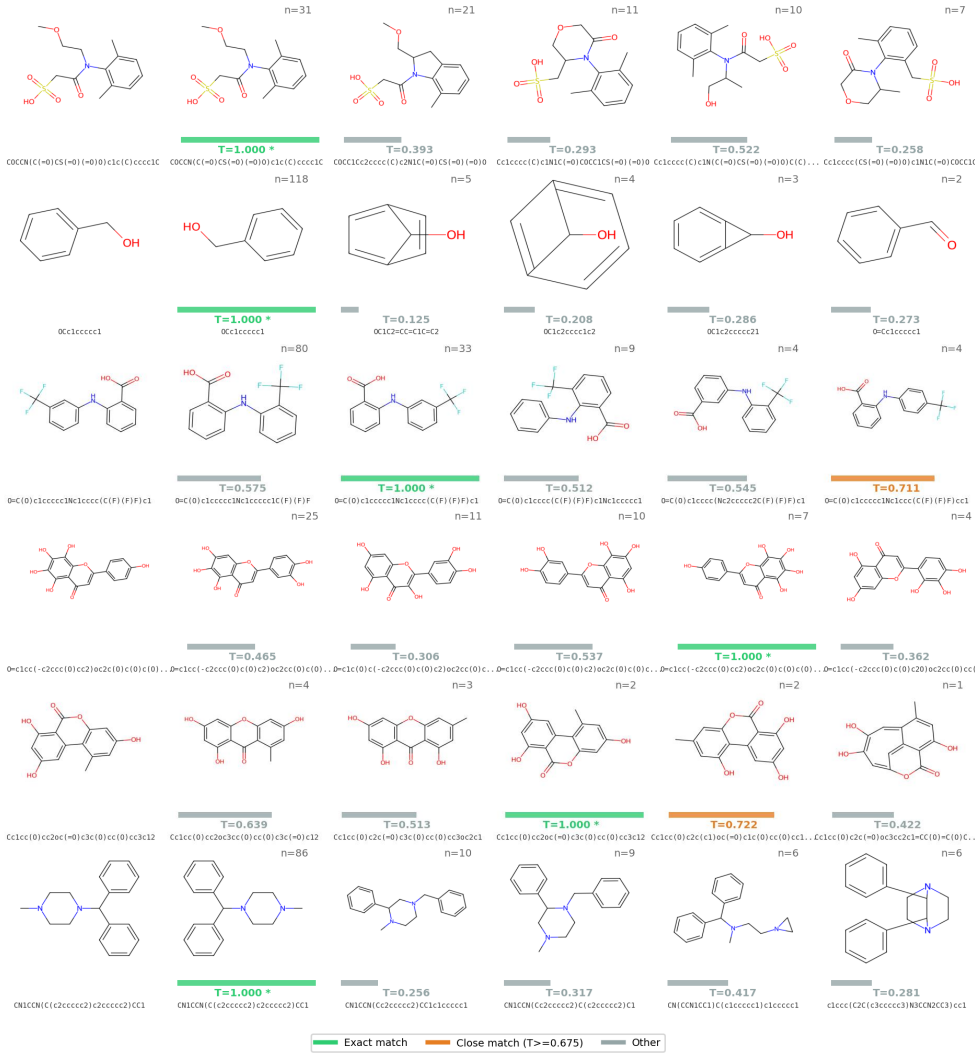


Figure 3: Qualitative case studies on the NPLIB1 dataset. For each test sample, the leftmost structure corresponds to the ground-truth molecule, while the remaining columns show the Top-5 candidate structures generated by ESDM and ranked according to their occurrence frequency. The value n above each candidate denotes the number of times that the structure appears among repeated generations. The colored bars indicate the similarity between the generated molecule and the ground truth: green denotes an exact structural match, orange denotes a close match with Tanimoto similarity greater than 0.675, and gray denotes all other generated structures. These examples demonstrate the ability of ESDM to recover correct or chemically similar molecular structures under accelerated diffusion sampling.