HYPERGRAPH-GUIDED COMPLEX STRUCTURE ENHANCEMENT FOR ACCURATE 3D MOLECULE GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

To accelerate the development of drug discovery, de novo 3D molecule generation has achieved many advances, aiming to synthesise accurate and reasonable 3D molecules from the given noise. Since molecules usually involve complex and diverse structures, it is challenging to precisely construct atoms and corresponding relations. Currently, most methods mainly employ flow- or diffusion-based techniques to effectively improve the quality of 3D molecules. However, they usually rely on atom-level and pairwise relation modelling, which may be insufficient for explicitly capturing higher-order chemical structures, e.g., aromatic rings, resulting in unstable patterns, unrealistic scaffolds, and degraded chemical validity in the generated molecules. To this end, inspired by the powerful relationship modelling capabilities of hypergraphs, we explore exploiting hypergraphs to perform higher-order structure enhancement. And a method of HSA-Mol3D is proposed. Concretely, HSA-Mol3D introduces a dynamic hypergraph branch into the generative backbone to model higher-order chemical structures. We further construct a staged and stabilised hypergraph fusion strategy in conjunction to make such higher-order information stable and effective throughout generation. Our proposed method enhances the stability of higher-order structures by progressively integrating hypergraph information into the generation process of the primary flow field model. Experiments show that our method achieves competitive performance on drug-like molecular generation, with particularly strong results in scaffold quality and the investigation of novel, higher-order skeletal medicinal molecules with superior properties.

1 INTRODUCTION

De novo molecular generation is an important component of AI-driven drug discovery. Building on the progress of 2D molecular graph generation (Jin et al., 2018; Irwin et al., 2022; Polykovskiy et al., 2018; Shi et al., 2020), 3D molecular generation has attracted increasing attention in recent years (Irwin et al., 2024; Ketata et al., 2024; Huang et al., 2024; Hoogeboom et al., 2022). Accurate 3D molecule generation has important applications in structure-based drug design (Zhang et al., 2023) and inverse molecule design targeting quantum properties (Bao et al., 2023).

However, 3D molecular generation is inherently more challenging because it requires the joint modeling of multiple interdependent factors, while still preserving chemical validity and realistic geometry. To address this problem, many existing 3D molecular generation methods are based on density-based 3D generation schemes (Kirchmeyer et al., 2024; Pinheiro et al., 2023), diffusion-based models (Huang et al., 2024; Vignac et al., 2023; Luo et al., 2025), and flow-based models (Hassan et al., 2024; Irwin et al., 2025; Dunn & Koes, 2026). These frameworks are typically built upon equivariant neural networks and graph neural networks, with molecular structures represented mainly as pairwise molecular graphs (Ketata et al., 2024; Satorras et al., 2021).

While graph-based models have made progress, they frequently treat higher-order structures, such as rings and aromatic systems, as secondary features or predefined fragments. This leads to a representational gap where the inherent multi-atom synergy is decomposed into simplified pairwise constraints, potentially failing to capture the holistic electronic and geometric properties of complex

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.

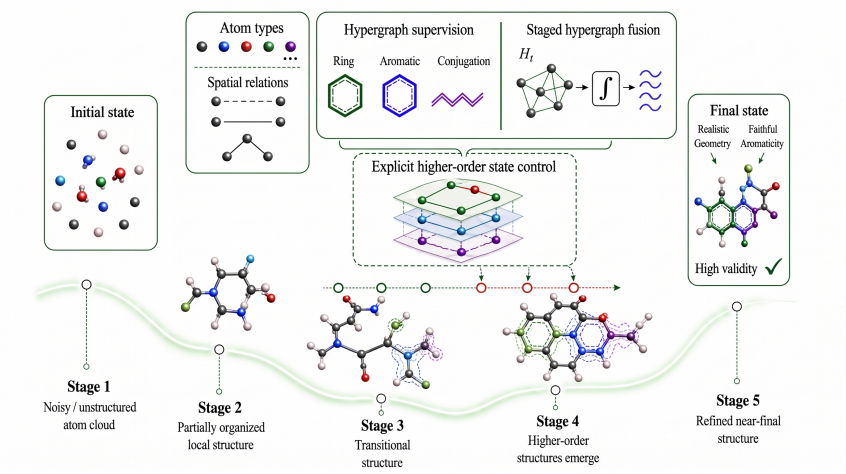

Figure 1: **Illustration of hypergraph-enhanced 3D molecular generation.** The dynamic hypergraph branch models ring, aromatic, and conjugation states and integrates them into the generation process through staged hypergraph fusion, enabling more complete and stable higher-order structures along the flow trajectory.

pharmacophores. As pointed out in recent studies (Restrepo, 2026), molecular systems are inherently governed by higher-order relationships that go beyond binary atomic interactions, while many important chemical states, such as rings, aromaticity, and conjugation, cannot be explicitly represented as higher-order structures by ordinary pairwise molecular graphs alone (Restrepo, 2022). Previous work has focused on controlling molecular topology through isotropic networks or modeling parts of molecules directly as motifs (Zhang et al., 2023). However, motif/subgraph-based methods depend heavily on how fragments are assembled or connected; if connection information is not properly modeled, generation quality and chemical reasonableness may be affected (Kong et al., 2022; Geng et al., 2023). Therefore, hypergraph-based representations provide a natural way to explicitly encode higher-order molecular structures by modeling interactions among sets of atoms rather than only atom pairs (Restrepo, 2026; 2022), as shown in Fig. 1.

Hypergraphs provide a suitable representation for such structures and have shown promise in molecular learning tasks related to conjugated and aromatic systems (Antelmi et al., 2023). Existing research further indicates that hypergraphs can better represent higher-order relationships among molecules or molecular fragments than regular graphs (Jo et al., 2021; Murgas et al., 2022; Dang & Hy, 2026). Nevertheless, previous studies have mainly explored hypergraphs in discriminative molecular learning rather than generation. More recently, HOG-Diff introduces higher-order topological skeletons into graph diffusion through a coarse-to-fine generation process (Huang & Birdal, 2026). These studies demonstrate the value of higher-order generation, but they do not address the same setting considered here: dynamically inferring chemistry-typed set representations from an evolving joint atom–charge–bond–coordinate state and injecting them into an equivariant 3D generator. This motivates us to design a hypergraph-enhanced framework for improving higher-order structural modeling in 3D molecular generation.

In this paper, we propose HSA-Mol3D, a hypergraph state-augmented framework for unified 3D molecular generation. HSA-Mol3D jointly generates atom types, formal charges, bond types, and atomic coordinates, while dynamically constructing typed hypergraph states for rings, aromatic systems, and conjugated components. Rather than using hypergraphs as fixed or oracle structural priors, the proposed branch infers higher-order structures from the evolving molecular state and introduces them as confidence-controlled late-stage corrections. Our main contributions are summarized as follows:

- We introduce dynamic chemistry-typed hypergraph states into unified 3D molecular generation, enabling ring, aromatic, and conjugation structures to be represented explicitly as evolving set-level states rather than only through pairwise relations.

- We propose a confidence-controlled staged hypergraph fusion mechanism that treats higher-order structural states as dynamic feedback variables in the generation process. By delaying their intervention until the intermediate molecular states become sufficiently reliable and injecting them through confidence-weighted residual corrections, the proposed mechanism enables stable, state-dependent refinement of the base generative dynamics while reducing early-stage error amplification.
- We evaluate HSA-Mol3D on QM9 (Ramakrishnan et al., 2014) and GEOM-Drugs (Axelrod & Gomez-Bombarelli, 2022). The results demonstrate competitive overall generation quality, with particularly strong scaffold fidelity and ring-distribution fidelity on drug-like molecules. Additional analyses further verify the reliability of typed higher-order state recovery and the robustness of the proposed dynamic feedback mechanism.

2 Related Works

2.1 3D molecule generation

3D molecular generation aims to produce chemically valid molecules with realistic conformations. Recent methods usually represent molecules as sets of atoms with coordinates and atom types, and generate molecular structures using autoregressive or normalizing-flow-based models (Gebauer et al., 2019; 2021; Luo & Ji, 2022; Shi et al., 2020; Pirnay et al., 2025). With the success of diffusion models (Vignac et al., 2023; Huang et al., 2024), equivariant diffusion has become a major paradigm for 3D molecular generation (Hoogeboom et al., 2022; Satorras et al., 2021; Hassan et al., 2024; Ketata et al., 2024). In addition, point-cloud diffusion, voxel-based methods represent molecules as 3D atomic density grids, generalize them for better scalability and sampling efficiency (Kirchmeyer et al., 2024; Pinheiro et al., 2023). There are also lossless consistency frameworks and efficient molecular generation models based on joint flow models (Dunn & Koes, 2026) to improve efficiency and the effectiveness of molecular generation.

2.2 Hypergraph Representation

Molecular graphs have long served as the standard representation for chemical structures, where atoms are represented as vertices and chemical bonds are represented as pairwise edges (Fialkowski et al., 2005; Bishop et al., 2006; Fujita, 2016; Rao et al., 2013). This representation is simple and effective for encoding local connectivity, but it is intrinsically limited to binary relations (Restrepo, 2022; Casetti et al., 2024; Liu et al., 2020; Schwaller et al., 2020). Hypergraphs provide a natural extension of molecular graphs by allowing each hyperedge to connect an arbitrary number of atoms, making them suitable for representing higher-order chemical structures (Corsini & Leoreanu, 2013; Berge, 1973; Jo et al., 2021). In addition, hypergraph-based molecular representations have shown promise in molecular property prediction and reactivity modeling, where higher-order structural information can provide advantages over conventional graph-based representations (Casetti et al., 2024; Murgas et al., 2022; Saifuddin et al., 2023). Hypergraphs extend ordinary graphs by allowing each hyperedge to connect an arbitrary number of nodes, and have been widely studied through hypergraph neural networks, hypergraph convolution, and hypergraph attention for modeling high-order relations beyond pairwise edges (Yadati et al., 2019; Bai et al., 2021; Antelmi et al., 2023; Dang & Hy, 2026). In molecular learning, hypergraph-based methods have been used to represent conjugated structures and encode chemical constraints for property prediction and molecular optimization (Chen & Schwaller, 2024; Zheng et al., 2026; Dang & Hy, 2026; Kajino, 2019), but their integration into unified 3D molecular de novo generation remains underexplored. HSA-Mol3D addresses this gap by dynamically constructing typed hyperedges for ring, aromatic, and conjugation components and fusing them into the joint atom-charge-bond-coordinate generation backbone.

3 Hypergraph State-Augmented Framework

In this paper, we formulate de novo 3D molecule generation as a joint generative problem (Dunn & Koes, 2026) over discrete chemical attributes and continuous coordinates. For a molecule with N atoms, we represent its state as $\mathcal{M} = (a, q, b, \boldsymbol{x})$, where $a = (a_i)_{i=1}^N$ and $q = (q_i)_{i=1}^N$ denote the atom types and formal charges, $b = (b_{ij})_{1 \le i < j \le N}$ denotes the bond types, and $\boldsymbol{x} \in \mathbb{R}^{N \times 3}$

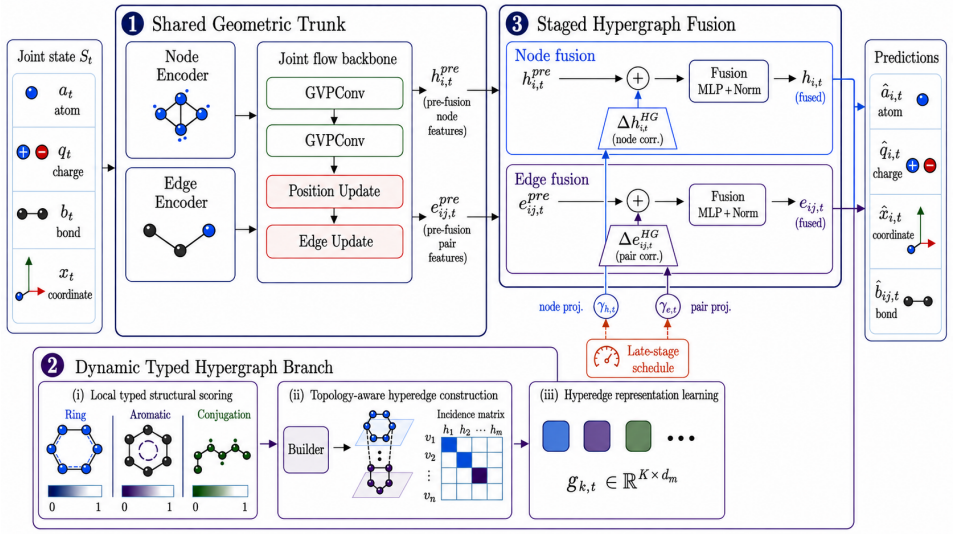

Figure 2: **Architectural Overview of the Full HSA-Mol3D Model:** (1) a shared geometric trunk, (2) a dynamic typed hypergraph branch, and (3) staged hypergraph fusion. Particularly, the shared geometric trunk extracts equivariant node, pair, and coordinate representations from the intermediate molecular state. Next, the dynamic typed hypergraph branch constructs evolving higher-order structural states (Section 3.1) and learns hyperedge-level representations from the current molecular state (Section 3.2). Finally, the staged hypergraph fusion module progressively injects reliable higher-order structural information into the node and pair representations during generation (Section 3.3).

denotes the atomic coordinate matrix. At generation time $t \in [0, 1]$, we define the molecular state as $\mathcal{M}_t = (a_t, q_t, b_t, \boldsymbol{x}_t)$, with its discrete chemical component denoted by $\mathcal{D}_t = (a_t, q_t, b_t)$, leading to the following multimodal probability path factorization:

$$p_t(\mathcal{M}_t \mid \mathcal{M}_0, \mathcal{M}_1) = p_t^{\mathrm{c}}(\boldsymbol{x}_t \mid \boldsymbol{x}_0, \boldsymbol{x}_1)\, p_t^{\mathrm{d}}(\mathcal{D}_t \mid \mathcal{D}_0, \mathcal{D}_1), \tag{1}$$

where p_t^{c} and p_t^{d} denote the continuous coordinate path and discrete chemical-state path. The coordinate state evolves according to $\boldsymbol{x}_t = (1 - \alpha_x(t))\, \boldsymbol{x}_0 + \alpha_x(t)\boldsymbol{x}_1$, where $\boldsymbol{x}_0$ is sampled from the coordinate prior, $\boldsymbol{x}_1$ denotes the target molecular structure, and $\alpha_x(t)$ controls the progression along the generative trajectory. The intermediate molecular state $\mathcal{M}_t$ is processed by an L-layer GVP-based equivariant geometric trunk (Jing et al., 2021), which produces the pre-fusion node, pair, and coordinate representations used by the hypergraph branch: $(\mathbf{h}_t^{\mathrm{pre}}, \mathbf{e}_t^{\mathrm{pre}}, \mathbf{x}_t^{\mathrm{pre}}) = F_\theta^{\mathrm{trunk}}(\mathcal{M}_t, t)$, where $\mathbf{h}_t^{\mathrm{pre}}$ and $\mathbf{e}_t^{\mathrm{pre}}$ denote the pre-fusion node and pair representations and $\mathbf{x}_t^{\mathrm{pre}}$ denotes the updated atomic coordinates. We then develop a framework to construct, represent, and progressively integrate dynamic hypergraph states into the generative process.

3.1 Dynamic Typed Hypergraph Construction

As illustrated in Fig. 2, our framework first performs dynamic hypergraph state construction to characterize the higher-order structural state of the current molecule. Given the pre-fusion node and pair representations $\mathbf{h}_t^{\mathrm{pre}}$ and $\mathbf{e}_t^{\mathrm{pre}}$ produced by the shared geometric trunk, the hypergraph branch considers three types of higher-order chemical structures: $\mathcal{R} = \{\mathrm{ring}, \mathrm{arom}, \mathrm{conj}\}$. Different from prior higher-order and hypergraph-based molecular models (Huang & Birdal, 2026; Ma et al., 2026), We introduce dynamic typed hypergraph states to complement the atom- and pair-level representations of the shared backbone. For each higher-order structural type $r \in \mathcal{R}$, we introduce a type-specific node scoring head $S_{\mathrm{node},r}$ and pair scoring head $S_{\mathrm{pair},r}$. Both heads are lightweight two-layer MLPs and operate on the pre-fusion node and pair representations produced by the shared trunk. They predict local evidence to determine whether an atom or a pair of atoms supports a higher-order structure of type r:

$$s_{i,t}^{(r)} = \sigma\big(S_{\mathrm{node},r}(\mathbf{h}_{i,t}^{\mathrm{pre}})\big), \qquad s_{ij,t}^{(r)} = \sigma\big(S_{\mathrm{pair},r}(\mathbf{e}_{ij,t}^{\mathrm{pre}})\big), \tag{2}$$

where $s_{i,t}^{(r)}$ estimates the participation of atom i in the type r structure, and $s_{ij,t}^{(r)}$ measures whether atoms i and j jointly support the same higher-order structure.

Topology-aware proposal construction. Considering that bond states and atomic geometry evolve jointly during generation (Dunn & Koes, 2026), we constrain higher-order structural evidence to remain consistent with the current molecular connectivity. The typed pair evidence is modulated as $\tilde{s}_{ij,t}^{(r)} = s_{ij,t}^{(r)} p_{ij,t}^{\text{bond}}$, where $p_{ij,t}^{\text{bond}}$ denotes the predicted probability that atoms i and j are chemically connected. The $s_{i,t}^{(r)}$ and $\tilde{s}_{ij,t}^{(r)}$ are then used to form a sparse typed candidate graph for each structural type r. We retain node and pair evidence above thresholds a and b and construct as follows:

$$\mathcal{E}_t^{(r)} = \left\{(i,j) \mid \tilde{s}_{ij,t}^{(r)} \geq b\right\}, \quad \mathcal{V}_t^{(r)} = \left\{i \mid s_{i,t}^{(r)} \geq a\right\} \cup \text{End}\left(\mathcal{E}_t^{(r)}\right), \quad \mathcal{G}_t^{(r)} = \left(\mathcal{V}_t^{(r)}, \mathcal{E}_t^{(r)}\right). \quad (3)$$

where $\mathcal{E}_t^{(r)}$ represents the retained pair set for structure type r, and $\mathcal{V}_t^{(r)}$ represents the corresponding candidate atom set, including atoms with sufficiently high node scores and the endpoints of retained pairs. $\text{End}(\cdot)$ returns the atoms appearing as endpoints of the retained pairs, $\mathcal{V}_t^{(r)}$ and $\mathcal{E}_t^{(r)}$ define the sparse typed candidate graph $\mathcal{G}_t^{(r)}$.

Dynamic hypergraph state. Existing hierarchical molecular generators often rely on predefined substructures, fragmentation rules, or fixed higher-level units (Jin et al., 2018; 2020; Qiang et al., 2023; Lee et al., 2025). Such representations are well defined once the target molecular structure is available. However, the intermediate molecular state continuously evolves, and its higher-order organization is therefore not known a priori. To this end, we construct the hypergraph from the current state $\mathcal{M}_t$. We extract the connected components of each $\mathcal{G}_t^{(r)}$ and apply lightweight type-specific structural filters to remove components that are inconsistent with the corresponding chemical topology. The resulting typed proposal set is defined as

$$\widehat{\mathcal{P}}_t = \bigcup_{r \in \mathcal{R}_t} \left\{(\mathcal{S}, r) \;\middle|\; \mathcal{S} \in \mathcal{F}_r\left(\text{Comp}\left(\mathcal{G}_t^{(r)}\right)\right)\right\}, \quad (4)$$

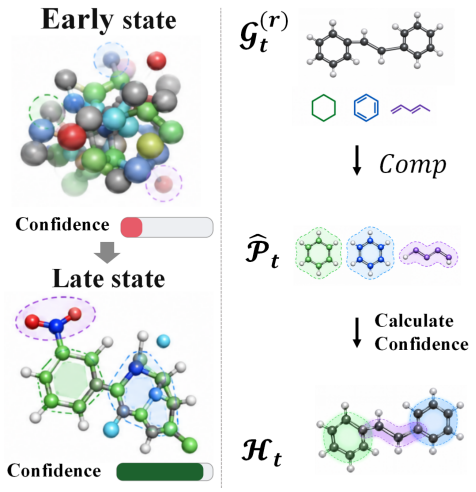

Figure 3: Dynamic Hypergraph State

where $\text{Comp}(\cdot)$ extracts the connected components of the typed candidate graph, and $\mathcal{F}_r(\cdot)$ denotes the type-specific structural filtering operation for structure type r. Each $(\mathcal{S}, r) \in \widehat{\mathcal{P}}_t$ represents a candidate higher-order structure specified by an atom set $\mathcal{S}$ and its corresponding structural type r. At generation time t, $\mathcal{R}_t \subseteq \mathcal{R}$ denotes the structural types considered in the current hypergraph state.

As the molecular state becomes progressively more stable along the generation trajectory, the inferred higher-order structures become increasingly reliable. Each proposal in $\widehat{\mathcal{P}}_t$ is assigned a structural confidence score, and low-confidence or redundant proposals are removed. The retained proposals form the dynamic typed hypergraph state $\mathcal{H}_t = \{(\mathcal{S}_{k,t}, r_{k,t}, c_{k,t})\}_{k=1}^{K_t}$, where $\mathcal{S}_{k,t}$ denotes the atom set of the k-th retained hyperedge, $r_{k,t} \in \mathcal{R}_t$ denotes its higher-order structural type, $c_{k,t} \in [0,1]$ denotes its structural confidence, and K_t is the number of retained hyperedges at generation time t. Fig. 3 shows the schematic process of hypergraph construction.

3.2 Hyperedge Representation Learning

For each hyperedge $(\mathcal{S}_{k,t}, r_{k,t}, c_{k,t}) \in \mathcal{H}_t$, we follow the permutation-invariant principle for set-structured data (Lee et al., 2019) and encode its internal node and pair features into a unified hyperedge representation $\mathbf{z}_{k,t}$ using mean pooling followed by a lightweight MLP. This representation provides a compact higher-order context for selectively retrieving the local features associated with the current hyperedge.

Node- and pair-level attention. Following the content-dependent aggregation principle of attention (Lee et al., 2019), we use the hyperedge-level representation $\mathbf{z}_{k,t}$ as the query, while the associated pre-fusion node representations provide the keys and values. This allows each hyperedge to selectively retrieve node information relevant to its current higher-order structure:

$$\mathbf{q}^{\mathrm{n}}_{k,t} = W^{\mathrm{n}}_{Q}\mathbf{z}_{k,t}, \quad \mathbf{k}^{\mathrm{n}}_{i,t} = W^{\mathrm{n}}_{K}\mathbf{h}^{\mathrm{pre}}_{i,t}, \quad \mathbf{v}^{\mathrm{n}}_{i,t} = W^{\mathrm{n}}_{V}\mathbf{h}^{\mathrm{pre}}_{i,t},$$
$$\alpha^{\mathrm{n}}_{i,k,t} = \mathrm{softmax}_{i}\left(\frac{(\mathbf{q}^{\mathrm{n}}_{k,t})^{\top}\mathbf{k}^{\mathrm{n}}_{i,t}}{\sqrt{d}}\right), \qquad \mathbf{a}^{\mathrm{n}}_{k,t} = \sum_{i\in\mathcal{S}_{k,t}} \alpha^{\mathrm{n}}_{i,k,t}\mathbf{v}^{\mathrm{n}}_{i,t}. \tag{5}$$

Similarly, pair-level structural information is aggregated using a separate attention module:

$$\mathbf{q}^{\mathrm{p}}_{k,t} = W^{\mathrm{p}}_{Q}\mathbf{z}_{k,t}, \quad \mathbf{k}^{\mathrm{p}}_{ij,t} = W^{\mathrm{p}}_{K}\mathbf{e}^{\mathrm{pre}}_{ij,t}, \quad \mathbf{v}^{\mathrm{p}}_{ij,t} = W^{\mathrm{p}}_{V}\mathbf{e}^{\mathrm{pre}}_{ij,t},$$
$$\alpha^{\mathrm{p}}_{ij,k,t} = \mathrm{softmax}_{(i,j)}\left(\frac{(\mathbf{q}^{\mathrm{p}}_{k,t})^{\top}\mathbf{k}^{\mathrm{p}}_{ij,t}}{\sqrt{d}}\right), \qquad \mathbf{a}^{\mathrm{p}}_{k,t} = \sum_{(i,j)\in\mathcal{P}_{k,t}} \alpha^{\mathrm{p}}_{ij,k,t}\mathbf{v}^{\mathrm{p}}_{ij,t}. \tag{6}$$

Finally, the hyperedge-level query and the attended node- and pair-level contexts are concatenated and projected as

$$\mathbf{g}_{k,t} = F_{\mathrm{hyper}}\big[\,\mathrm{Concat}(\mathbf{z}_{k,t}, \mathbf{a}^{\mathrm{n}}_{k,t}, \mathbf{a}^{\mathrm{p}}_{k,t})\big], \tag{7}$$

where F_{hyper} denotes a two-layer MLP that maps the concatenated hyperedge, node, and pair contexts into a unified hyperedge representation space, with SiLU activation applied after each linear transformation. In this case, $\mathbf{a}^{\mathrm{n}}_{k,t}$ and $\mathbf{a}^{\mathrm{p}}_{k,t}$ encode node-level and pair-level local structural information.

3.3 Staged Hypergraph Fusion

Since the dynamic hypergraph is inferred from the current intermediate molecular state, its higher-order predictions can be unreliable at early generation stages. Directly feeding such information back into the generative backbone may amplify structural flaws (Dunn & Koes, 2026). We therefore adopt a confidence-controlled staged fusion mechanism, in which higher-order states are gradually introduced according to the generation time and their structural confidence.

For each predicted hyperedge, we assign an effective fusion weight $\omega^{(s)}_{k,t} \in [0,1]$ to control its contribution to the generative backbone. The weight jointly accounts for the current generation stage and the reliability of the inferred higher-order structure, suppressing uncertain hyperedges while progressively strengthening reliable structural feedback as generation proceeds.

Higher-order residual fusion. Each learned hyperedge representation $\mathbf{g}_{k,t}$ contributes higher-order context to its associated atoms and atom pairs. The corresponding corrections are aggregated as

$$\Delta\mathbf{h}^{\mathrm{HG}}_{i,t} = \sum_{k:\, i\in\mathcal{S}_{k,t}} \omega^{(s)}_{k,t}\Psi_{\mathrm{n}}\left(\mathbf{g}_{k,t}, \mathbf{h}^{\mathrm{pre}}_{i,t}\right), \qquad \Delta\mathbf{e}^{\mathrm{HG}}_{ij,t} = \sum_{k:\, i,j\in\mathcal{S}_{k,t}} \omega^{(s)}_{k,t}\Psi_{\mathrm{p}}\left(\mathbf{g}_{k,t}, \mathbf{e}^{\mathrm{pre}}_{ij,t}\right). \tag{8}$$

where Ψ_{n} and Ψ_{p} are lightweight learnable projections that map the hyperedge representation into the node and pair feature spaces, respectively.

The higher-order corrections are then injected into the shared backbone through residual fusion:

$$\mathbf{h}_{i,t} = \mathrm{LN}\left(\mathbf{h}^{\mathrm{pre}}_{i,t} + F_{\mathrm{n}}\left[\mathbf{h}^{\mathrm{pre}}_{i,t} \| \Delta\mathbf{h}^{\mathrm{HG}}_{i,t}\right]\right), \qquad \mathbf{e}_{ij,t} = \mathrm{LN}\left(\mathbf{e}^{\mathrm{pre}}_{ij,t} + F_{\mathrm{p}}\left[\mathbf{e}^{\mathrm{pre}}_{ij,t} \| \Delta\mathbf{e}^{\mathrm{HG}}_{ij,t}\right]\right). \tag{9}$$

where F_{n} and F_{p} denote the node- and pair-level fusion MLPs. The fused representations are subsequently used by the prediction heads to update the atom, charge, bond, and coordinate states.

Training objective. The overall model is trained jointly with the molecular flow objective and higher-order hypergraph supervision:

$$\mathcal{L} = \mathcal{L}_{\mathrm{flow}} + \mathcal{L}_{\mathrm{local}} + \mathcal{L}_{\mathrm{type}} + \mathcal{L}_{\mathrm{obj}} + \mathcal{L}_{\mathrm{mem}}, \tag{10}$$

where $\mathcal{L}_{\mathrm{flow}}$ denotes the joint molecular generation loss, and $\mathcal{L}_{\mathrm{local}}$ supervises the local typed node- and pair-level evidence used for dynamic hypergraph construction. $\mathcal{L}_{\mathrm{type}}$ supervises the predicted higher-order structural type, $\mathcal{L}_{\mathrm{obj}}$ distinguishes valid hyperedge proposals from spurious ones, and $\mathcal{L}_{\mathrm{mem}}$ supervises atom membership within each matched hyperedge. Detailed definitions are provided in Sections B.5, B.6, and B.7.

Table 1: Performance of *de novo* 3D molecule generation on GEOM-Drugs. Higher values indicate better performance. *Data* denotes molecules randomly sampled from the test set. *Early fusion* applies hypergraph fusion from the initial generation step. Bold values indicate the best performance among the compared generative models.

Method	AtomStable	MolStable	V&C	V&U	V&U&N	SNN	Frag	Scaf
Data	0.999	0.999	0.998	0.998	0.000	0.579	0.995	0.584
CDGS (Huang et al., 2023)	0.991	0.706	0.285	0.285	0.285	0.262	0.789	0.022
JODO (Huang et al., 2024)	**1.000**	0.981	0.874	0.905	0.902	0.417	**0.993**	0.483
MiDi (Vignac et al., 2023)	0.968	0.822	0.633	0.654	0.652	0.392	0.951	0.196
EQGAT (Le et al., 2022)	**1.000**	0.989	0.845	0.863	0.859	0.377	0.983	0.161
FlowMol3 (Dunn & Koes, 2026)	**1.000**	0.982	0.946	0.913	0.913	0.466	0.838	0.232
UDM-3D (Luo et al., 2025)	**1.000**	0.925	0.879	0.913	0.907	**0.525**	0.990	0.540
Early fusion	1.000	0.968	0.959	0.951	0.951	0.385	0.793	0.389
HSA-Mol3D (Ours)	**1.000**	**0.995**	**0.976**	**0.964**	**0.964**	0.464	0.902	**0.542**

Table 2: Results on additional GEOM-Drugs metrics. ↑ and ↓ indicate that higher and lower values are better, respectively. *Data* denotes molecules randomly sampled from the test set. *Early fusion* applies hypergraph fusion from the initial generation step. Bold values indicate the best performance among the compared generative models.

Method	Atoms/mol TV ↓	Bond len. W_1 ↓	Bond ang. W_1 ↓	Valency W_1 ↓	1-frag. (%) ↑	Energy ↓	Ring size TV ↓	QED ↑	logP
Data	0.000	0.000	0.005	0.001	100.0	54.5	0.011	0.658	2.95
EDM (Hoogeboom et al., 2022)	0.604	**0.002**	6.42	0.285	42.2	951.3	0.976	0.472	1.11
GeoLDM (Xu et al., 2023)	0.469	0.009	2.96	0.254	51.6	461.5	0.644	0.497	1.05
FuncMol (Kirchmeyer et al., 2024)	1.050	0.003	2.49	0.245	70.5	109.7	0.264	0.659	2.73
Early fusion	0.284	0.030	4.137	0.063	98.1	186.8	0.131	0.587	3.370
HSA-Mol3D (Ours)	**0.145**	0.009	**1.65**	**0.060**	**99.9**	**70.6**	**0.092**	**0.665**	2.91

4 EXPERIMENTS

In this section, we present the experimental results of our proposed HSA-Mol3D on 3D molecule generation tasks. Our experiments demonstrate competitive overall generation quality, with particular improvements in validity, novelty, scaffold fidelity, and ring-size distribution.

4.1 EXPERIMENTAL SETUP

Datasets. We evaluate HSA-Mol3D on two datasets: QM9 (Ramakrishnan et al., 2014), GEOM-drugs (Axelrod & Gomez-Bombarelli, 2022) .QM9 contains an enumeration of all possible molecules up to 9 heavy atoms (29 including hydrogens) satisfying some constraints (Ramakrishnan et al., 2014).We model hydrogen explicitly and consider 5 chemical elements for QM9 (C, H, O, N, F) with 100K/20K/13K molecules train/val/test splits, following (Kirchmeyer et al., 2024).GEOM-drugs contains multiple conformations for 430K drug-sized molecules (computed with semi-empirical density functional theory), with an average of 44 atoms per molecule.We model hydrogen explicitly and consider 8 for GEOM-drugs (C, H, O, N, F, S, Cl and Br) and 80%/10%/10% splits and retain the lowest-energy conformer per molecule.

Baselines. For *de novo* 3D molecule generation, we compare HSA-Mol3D with CDGS (Huang et al., 2023), JODO (Huang et al., 2024), MiDi (Vignac et al., 2023), EDM (Hoogeboom et al., 2022), GeoLDM (Xu et al., 2023), EQGAT (Le et al., 2022), UDM-3D (Luo et al., 2025),SemlaFlow (Irwin et al., 2025),FlowMol3 (Dunn & Koes, 2026), FuncMol (Kirchmeyer et al., 2024). For baselines that generate atom identities and 3D coordinates without explicit bond types, we reconstruct molecular bonds using Open Babel (O'Boyle et al., 2011), following prior evaluation protocols (Vignac et al., 2023; Pinheiro et al., 2023; Ragoza et al., 2020; Guan et al., 2023). For baselines that directly predict bond types, we follow the original reconstruction procedure described in their papers.

Table 3: Performance of *de novo* 3D molecule generation on QM9. Higher values indicate better performance. *Data* denotes molecules randomly sampled from the test set. Bold values indicate the best performance among the compared generative models.

Method	AtomStable	MolStable	V&C	V&U	V&U&N	SNN	Frag	Scaf
Data	0.999	0.988	0.989	0.989	0.000	0.490	0.992	0.946
CDGS (Huang et al., 2023)	0.997	0.951	0.951	0.936	0.860	0.493	0.973	0.784
JODO (Huang et al., 2024)	**0.999**	**0.988**	0.990	0.960	0.780	**0.522**	0.986	**0.934**
MiDi (Vignac et al., 2023)	0.998	0.976	0.980	0.954	0.769	0.501	0.979	0.882
EQGAT (Le et al., 2022)	**0.999**	0.971	0.965	0.950	0.891	0.482	0.950	0.703
SemlaFlow (Irwin et al., 2025)	0.995	0.949	0.857	0.821	0.821	-	-	-
UDM-3D (Luo et al., 2025)	**0.999**	**0.988**	0.983	**0.973**	0.950	0.508	**0.987**	0.898
HSA-Mol3D (Ours)	**0.999**	**0.988**	0.985	**0.973**	**0.973**	0.501	0.902	0.902

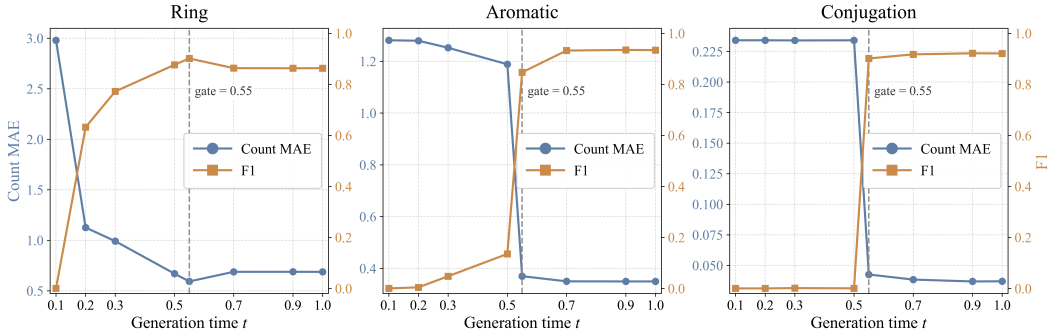

Figure 4: Evolution of typed hyperedge recovery along the generation trajectory. For ring, aromatic, and conjugation structures, F1 generally increases while Count MAE decreases as generation proceeds, indicating that the inferred higher-order states become progressively more reliable. The dashed line marks the threshold at t_{mid}.

4.2 *De Novo* 3D Molecule Generation

Experimental Setting. We consider several metrics used in previous work (Huang et al., 2024; Pinheiro et al., 2023) to benchmark molecule generation for the standard QM9 and GEOM-drugs datasets: Atom stability, validity & completeness (V&C), validity & uniqueness (V&U), validity & uniqueness & novelty (V&U&N), similarity to nearest neighbor (SNN), fragment similarity (Frag), scaffold similarity (Scaf).

To further investigate the quality of molecular conformations and other molecular properties on GEOM-drugs, we consider some additional metrics in previous work (Kirchmeyer et al., 2024; Pinheiro et al., 2023; Xu et al., 2023). We additionally evaluate atom-count and ring-size total variation, bond-length, bond-angle, and valency Wasserstein distances, single-fragment rate, strain energy, QED, and logP.

Table 1 shows how HSA-Mol3D did on the de novo generation task using the GEOM-Drugs datasets. Our model achieves state-of-the-art performance on most of the metrics in this datasets. The results show significant key observations:

Generation quality and structural fidelity. As shown in Tables 1 and 2, HSA-Mol3D achieves the best V&C, V&U, and V&U&N scores among the compared models on GEOM-Drugs while maintaining perfect atom stability. It also obtains the strongest scaffold similarity, atom-count TV, ring-size TV, and QED results among the generative baselines. On QM9, as shown in Table 3, HSA-Mol3D also remains competitive with established 3D molecular generators. The comparison with *Early fusion* further highlights the importance of staged hypergraph integration. The staged hypergraph fusion design of HSA-Mol3D effectively improves overall generation quality by introducing higher-order structural information at more reliable stages of the generation process.

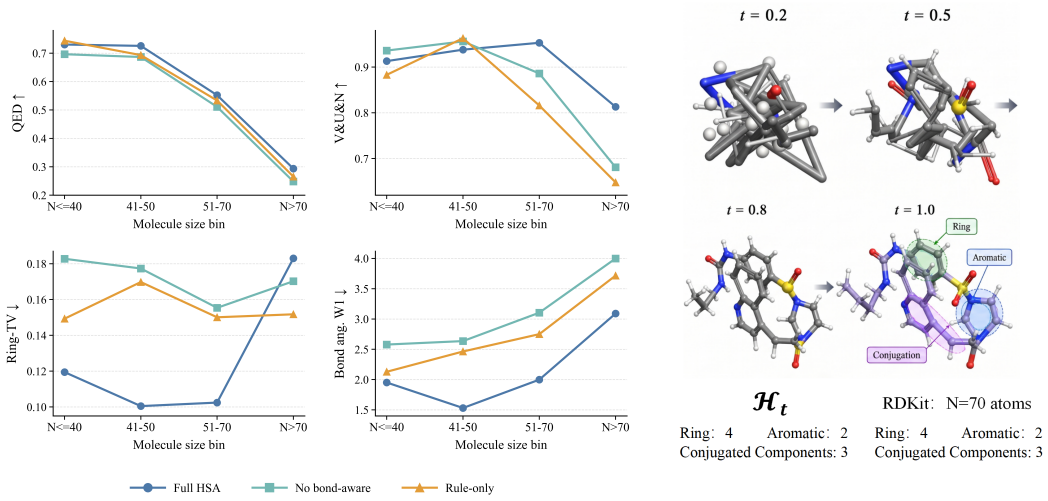

Figure 5: **Generation performance across molecular complexity and representative complex-molecule generation dynamics.** Left: Size-conditioned QED, V&U&N, Ring-TV, and bond-angle W_1 are compared across Full HSA, No bond-aware, and Rule-only. Right: A representative $N_{\text{atom}} = 70$ generation trajectory from $t = 0.2$ to $t = 1.0$. At $t = 1.0$, the predicted dynamic hypergraph state $\mathcal{H}_t$ is consistent with the RDKit analysis, with matching counts of rings, aromatic systems, and conjugated components (4, 2, and 3, respectively).

4.3 Ablations and Analysis Studies

Typed higher-order structure recovery analysis. Fig. 4 analyzes how the recovery quality of the three typed higher-order structures evolves along the generation trajectory.As generation proceeds, the F1 score generally improves while Count MAE decreases across ring, aromatic, and conjugation structures, indicating that the inferred hypergraph states become increasingly more reliable as the molecular state becomes more structured.

Component ablation. We separately ablate bond-aware construction and explicit hypergraph fusion to examine their contributions. As shown in Table 4, Full HSA achieves the most balanced generation quality, demonstrating the complementary roles of connectivity-aware construction and higher-order structural feedback. Additional ablation studies are provided in Appendix A.1.

Table 4: Ablation of bond-aware construction and hypergraph fusion on GEOM-Drugs.

Bond-aware	Fusion	MolStable↑	V&U&N↑	Ring-TV↓	Bond ang. W_1 ↓
–	×	0.980	0.913	0.170	3.506
×	✓	0.994	0.950	0.172	2.813
✓	✓	**0.995**	**0.964**	**0.092**	**1.650**

Size-conditioned generation analysis. Fig. 5 further examines how the generation behavior changes with molecular size. HSA-Mol3D shows clear advantages in V&U&N and bond-angle W1 for medium-to-large molecules, while maintaining higher QED. The representative $N_{\text{atom}} = 70$ trajectory further demonstrates the ability of HSA-Mol3D to generate structurally complex molecules. At the final state, the predicted hypergraph state $\mathcal{H}_t$ contains four rings, two aromatic systems, and three conjugated components, matching the corresponding RDKit analysis and supporting the effectiveness of the proposed higher-order modeling for complex molecule generation.

5 Conclusion

HSA-Mol3D is a hypergraph-enhanced framework for 3D molecular generation that introduces dynamic hypergraph fusion into joint flow modeling. Extensive experiments on GEOM-Drugs and QM9 demonstrate strong performance in *de novo* 3D molecule generation. This work highlights that combining hypergraph-based structural enhancement with joint flow modeling provides a promising direction for generating structurally complex drug-like molecules with improved chemical and geometric fidelity.

AI USE STATEMENT

In this work, we used generative AI tools for the refinement of research ideas, discussion of methodological and experimental design choices, revision of the manuscript for clarity and presentation. We have not used generative AI tools for data collection or annotation or for the interpretation of experimental results, and the rest of the required disclosure tasks are not applicable to this work. We have reviewed all AI-assisted work. AI-assisted text and figures were manually checked by the authors for scientific accuracy, consistency, and clarity before inclusion in the manuscript. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide detailed information to facilitate reproducibility of our results. Appendix B presents the complete mathematical formulation of HSA-Mol3D, together with the model configuration, training and evaluation settings, key hyperparameters, and implementation details used in the experiments. Additional experimental protocols, ablation settings, and evaluation procedures are also documented in the appendix and supplementary materials. These materials are intended to enable independent reproduction of the reported results.

REFERENCES

Alessia Antelmi, Gennaro Cordasco, Mirko Polato, Vittorio Scarano, Carmine Spagnuolo, and Dingqi Yang. A survey on hypergraph representation learning. *ACM Computing Surveys*, 56 (1):1–38, 2023.

Simon Axelrod and Rafael Gomez-Bombarelli. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific data*, 9(1):185, 2022.

Song Bai, Feihu Zhang, and Philip HS Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, 2021.

Fan Bao, Min Zhao, Zhongkai Hao, Peiyao Li, Chongxuan Li, and Jun Zhu. Equivariant energy-guided SDE for inverse molecular design. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.

Claude Berge. Graphs and hypergraphs. north-holl math. libr, 1973.

Kyle JM Bishop, Rafal Klajn, and Bartosz A Grzybowski. The core and most useful molecules in organic chemistry. *Angewandte Chemie International Edition*, 45(32), 2006.

Nicholas Casetti, Pragnay Nevatia, Junwu Chen, Philippe Schwaller, and Connor W Coley. Comment on "molecular hypergraph neural networks"[j. chem. phys. 160, 144307 (2024)]. *The Journal of Chemical Physics*, 161(20), 2024.

Junwu Chen and Philippe Schwaller. Molecular hypergraph neural networks. *The Journal of Chemical Physics*, 160(14), 2024.

Piergiulio Corsini and Violeta Leoreanu. *Applications of hyperstructure theory*, volume 5. Springer Science & Business Media, 2013.

Tien Dang and Truong-Son Hy. Equihgnn: Scalable rotationally equivariant hypergraph neural networks. *The Journal of Chemical Physics*, 164(14), 2026.

Ian Dunn and David R Koes. Flowmol3: flow matching for 3d de novo small-molecule generation. *Digital Discovery*, 5(5):2052–2066, 2026.

Marcin Fialkowski, Kyle JM Bishop, Victor A Chubukov, Christopher J Campbell, and Bartosz A Grzybowski. Architecture and evolution of organic chemistry. *Angewandte Chemie International Edition*, 44(44):7263–7269, 2005.

Shinsaku Fujita. Half-century journey from synthetic organic chemistry to mathematical stereochemistry through chemoinformatics. *Iranian Journal of Mathematical Chemistry*, 7(2):155–221, 2016.

Niklas Gebauer, Michael Gastegger, and Kristof Schütt. Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules. *Advances in neural information processing systems*, 32, 2019.

Niklas W. A. Gebauer, Michael Gastegger, Stefaan Simon Pierre Hessmann, Klaus-Robert Müller, and Kristof T. Schütt. Inverse design of 3d molecular structures with conditional generative neural networks. *CoRR*, abs/2109.04824, 2021.

Zijie Geng, Shufang Xie, Yingce Xia, Lijun Wu, Tao Qin, Jie Wang, Yongdong Zhang, Feng Wu, and Tie-Yan Liu. De novo molecular generation via connection-aware motif mining. *arXiv preprint arXiv:2302.01129*, 2023.

Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3d equivariant diffusion for target-aware molecule generation and affinity prediction. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.

Majdi Hassan, Nikhil Shenoy, Jungyoon Lee, Hannes Stärk, Stephan Thaler, and Dominique Beaini. Et-flow: Equivariant flow-matching for molecular conformer generation. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024.

Emiel Hoogeboom, Victor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato (eds.), *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, Proceedings of Machine Learning Research, pp. 8867–8887. PMLR, 2022.

Han Huang, Leilei Sun, Bowen Du, and Weifeng Lv. Conditional diffusion based on discrete graph structures for molecular graph generation. In Brian Williams, Yiling Chen, and Jennifer Neville (eds.), *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pp. 4302–4311. AAAI Press, 2023.

Han Huang, Leilei Sun, Bowen Du, and Weifeng Lv. Learning joint 2-d and 3-d graph diffusion models for complete molecule generation. *IEEE Trans. Neural Networks Learn. Syst.*, 35(9): 11857–11871, 2024.

Yiming Huang and Tolga Birdal. Hog-diff: Higher-order guided diffusion for graph generation. In *The Fourteenth International Conference on Learning Representations*, 2026.

Ross Irwin, Spyridon Dimitriadis, Jiazhen He, and Esben Jannik Bjerrum. Chemformer: a pre-trained transformer for computational chemistry. *Mach. Learn. Sci. Technol.*, 3(1):15022, 2022.

Ross Irwin, Alessandro Tibo, Jon Paul Janet, and Simon Olsson. Efficient 3d molecular generation with flow matching and scale optimal transport. In *ICML 2024 AI for Science Workshop*, 2024.

Ross Irwin, Alessandro Tibo, Jon Paul Janet, and Simon Olsson. Semlaflow - efficient 3d molecular generation with latent attention and equivariant flow matching. In Yingzhen Li, Stephan Mandt, Shipra Agrawal, and Mohammad Emtiyaz Khan (eds.), *International Conference on Artificial Intelligence and Statistics, AISTATS 2025, Mai Khao, Thailand, 3-5 May 2025*, Proceedings of Machine Learning Research, pp. 3772–3780. PMLR, 2025.

Wengong Jin, Regina Barzilay, and Tommi S. Jaakkola. Junction tree variational autoencoder for molecular graph generation. In Jennifer G. Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm,*

Sweden, July 10-15, 2018, volume 80 of *Proceedings of Machine Learning Research*, pp. 2328–2337. PMLR, 2018.

Wengong Jin, Regina Barzilay, and Tommi S. Jaakkola. Hierarchical generation of molecular graphs using structural motifs. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 4839–4848. PMLR, 2020.

Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael John Lamarre Townshend, and Ron O. Dror. Learning from protein structure with geometric vector perceptrons. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

Jaehyeong Jo, Jinheon Baek, Seul Lee, Dongki Kim, Minki Kang, and Sung Ju Hwang. Edge representation learning with hypergraphs. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 7534–7546, 2021.

Hiroshi Kajino. Molecular hypergraph grammar with its application to molecular optimization. In *International Conference on Machine Learning*, pp. 3183–3191. PMLR, 2019.

Mohamed Amine Ketata, Nicholas Gao, Johanna Sommer, Tom Wollschläger, and Stephan Günnemann. Lift your molecules: Molecular graph generation in latent euclidean space. *arXiv preprint arXiv:2406.10513*, 2024.

Matthieu Kirchmeyer, Pedro O. Pinheiro, and Saeed Saremi. Score-based 3d molecule generation with neural fields. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024.

Xiangzhe Kong, Wenbing Huang, Zhixing Tan, and Yang Liu. Molecule generation by principal subgraph mining and assembling. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

Tuan Le, Frank Noé, and Djork-Arné Clevert. Representation learning on biomolecular structures using equivariant graph attention. In Bastian Rieck and Razvan Pascanu (eds.), *Learning on Graphs Conference, LoG 2022, 9-12 December 2022, Virtual Event*, Proceedings of Machine Learning Research, pp. 30. PMLR, 2022.

Joongwon Lee, Seonghwan Kim, and Wou Youn Kim. Fragfm: Efficient fragment-based molecular generation via discrete flow matching. *CoRR*, abs/2502.15805, 2025. doi: 10.48550/ARXIV.2502.15805.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam R. Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3744–3753. PMLR, 2019.

Xiangxiang Liu, Ligong Wang, and Xihe Li. The wiener index of hypergraphs. *Journal of Combinatorial Optimization*, 39(2):351–364, 2020.

Yanchen Luo, Zhiyuan Liu, Yi Zhao, Sihang Li, Kenji Kawaguchi, Tat-Seng Chua, and Xiang Wang. Towards unified latent space for 3d molecular latent diffusion modeling. *CoRR*, abs/2503.15567, 2025.

Youzhi Luo and Shuiwang Ji. An autoregressive flow model for 3d molecular geometry generation from scratch. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

Jiani Ma, Qi Yang, Lin Zhang, Hui Liu, and Yuanting Zheng. Hyg-mol: An interpretable multimodal hypergraph framework for molecular property prediction. *Computational and Structural Biotechnology Journal*, 35(1):0036, 2026.

Kevin A Murgas, Emil Saucan, and Romeil Sandhu. Hypergraph geometry reflects higher-order dynamics in protein interaction networks. *Scientific Reports*, 12(1):20879, 2022.

Noel M O'Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1):33, 2011.

Pedro O. Pinheiro, Joshua A. Rackers, Joseph Kleinhenz, Michael Maser, Omar Mahmood, Andrew M. Watkins, Stephen Ra, Vishnu Sresht, and Saeed Saremi. 3d molecule generation by denoising voxel grids. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

Jonathan Pirnay, Jan G Rittig, Alexander B Wolf, Martin Grohe, Jakob Burger, Alexander Mitsos, and Dominik G Grimm. Graphxform: graph transformer for computer-aided molecular design. *Digital Discovery*, 4(4):1052–1065, 2025.

Daniil Polykovskiy, Alexander Zhebrak, Benjamín Sánchez-Lengeling, Sergey Golovanov, Oktai Tatanov, Stanislav Belyaev, Rauf Kurbanov, Aleksey Artamonov, Vladimir Aladinskiy, Mark Veselov, Artur Kadurin, Sergey I. Nikolenko, Alán Aspuru-Guzik, and Alex Zhavoronkov. Molecular sets (MOSES): A benchmarking platform for molecular generation models. *CoRR*, abs/1811.12823, 2018.

Bo Qiang, Yuxuan Song, Minkai Xu, Jingjing Gong, Bowen Gao, Hao Zhou, Wei-Ying Ma, and Yanyan Lan. Coarse-to-fine: a hierarchical diffusion model for molecule generation in 3d. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 28277–28299. PMLR, 2023.

Matthew Ragoza, Tomohide Masuda, and David Ryan Koes. Learning a continuous representation of 3d molecular structures with deep generative models. *arXiv preprint arXiv:2010.08687*, 2020.

Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

Shodhan Rao, Arjan van der Schaft, and Bayu Jayawardhana. A graph-theoretical approach for the analysis and model reduction of complex-balanced chemical reaction networks. *Journal of Mathematical Chemistry*, 51(9):2401–2422, 2013.

Guillermo Restrepo. Chemical space: limits, evolution and modelling of an object bigger than our universal library. *Digital Discovery*, 1(5):568–585, 2022.

Guillermo Restrepo. Higher order structures in chemistry: hypergraphs reshape the molecule and the reaction. *Digital Discovery*, 2026.

Khaled Mohammed Saifuddin, Briana Bumgardner, Farhan Tanvir, and Esra Akbas. Hygnn: Drug-drug interaction prediction via hypergraph neural network. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pp. 1503–1516. IEEE, 2023.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Philippe Schwaller, Riccardo Petraglia, Valerio Zullo, Vishnu H Nair, Rico Andreas Haeuselmann, Riccardo Pisoni, Costas Bekas, Anna Iuliano, and Teodoro Laino. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chemical science*, 11(12):3316–3325, 2020.

Chence Shi, Minkai Xu, Zhaocheng Zhu, Weinan Zhang, Ming Zhang, and Jian Tang. Graphaf: a flow-based autoregressive model for molecular graph generation. *arXiv preprint arXiv:2001.09382*, 2020.

Clément Vignac, Nagham Osman, Laura Toni, and Pascal Frossard. Midi: Mixed graph and 3d denoising diffusion for molecule generation. In Danai Koutra, Claudia Plant, Manuel Gomez Rodriguez, Elena Baralis, and Francesco Bonchi (eds.), *Machine Learning and Knowledge Discovery in Databases: Research Track - European Conference, ECML PKDD 2023, Turin, Italy, September 18-22, 2023, Proceedings, Part II*, Lecture Notes in Computer Science, pp. 560–576. Springer, 2023.

Minkai Xu, Alexander S. Powers, Ron O. Dror, Stefano Ermon, and Jure Leskovec. Geometric latent diffusion models for 3d molecule generation. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, Proceedings of Machine Learning Research, pp. 38592–38610. PMLR, 2023.

Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha P. Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 1509–1520, 2019.

Zaixi Zhang, Yaosen Min, Shuxin Zheng, and Qi Liu. Molecule generation for target protein binding with structural motifs. In *The eleventh international conference on learning representations*, 2023.

Liwei Zheng, Bin Dai, Fan Yang, Jie Xia, Kaibiao Lin, and Zhongqi Cai. MFHNN-DDI: molecular fragment-based hypergraph neural networks for drug-drug interaction prediction. *Neurocomputing*, 661:131958, 2026.

A More Experimental Results

A.1 Additional Ablation Studies and Analysis

Effect of staged hypergraph fusion. To evaluate the effect of the proposed staged hypergraph fusion strategy, we compare three variants: an auxiliary-only baseline that retains hypergraph-related auxiliary supervision but disables hypergraph message fusion(Aux-only), an early-fusion variant that injects hypergraph messages from the beginning of the generation trajectory, and the proposed time-gated fusion strategy. The model backbone and the remaining training settings are kept unchanged across these variants.

As shown in Tables 5 and 6, early fusion improves some structure-related statistics, including novelty and ring-size distribution, but degrades molecular stability, distributional similarity, and several geometric metrics. This behavior suggests that injecting predicted higher-order messages while the intermediate molecular state remains highly noisy can amplify unreliable structural evidence.

In contrast, the proposed time-gated strategy postpones hypergraph intervention until the intermediate atom–bond–coordinate state becomes more reliable. It therefore provides a more balanced result across molecular validity, scaffold fidelity, geometric realism, and chemical consistency. In particular, time-gated fusion achieves the strongest scaffold similarity, bond-length and bond-angle distributions, single-fragment rate, strain energy, and QED among the three variants.

Table 5: Ablation analysis of generation quality under different hypergraph fusion strategies on GEOM-Drugs. Early fusion denotes applying hypergraph fusion from the initial time step, whereas time-gated fusion activates hypergraph fusion only at later generation stages. Higher values are better for all metrics.

Variant	Atom Stable	Mol Stable	V&C	V&U	V&U&N	SNN	Frag	Scaf
Aux-only (no hypergraph fusion)	**1.000**	0.980	0.964	0.904	0.913	0.462	0.898	0.233
+ Early fusion	**1.000**	0.968	0.959	0.951	0.951	0.385	0.793	0.389
+ Time-gated fusion (HSA-Mol3D)	**1.000**	**0.995**	**0.976**	**0.964**	**0.964**	**0.464**	**0.902**	**0.542**

Table 6: Ablation analysis of geometric quality, chemical consistency, and molecular properties under different hypergraph fusion strategies on GEOM-Drugs. Lower values are better for TV, W_1, and median energy, while higher values are better for single-fragment rate and QED.

Variant	atms/mol TV↓	bond len W_1 ↓	bond ang W_1 ↓	valency W_1 ↓	single frag %↑	median energy↓	ring sz TV↓	QED ↑	logP
Aux-only (no hypergraph fusion)	0.288	0.023	3.506	0.076	96.4	144.9	0.170	0.660	3.737
+ Early fusion	0.284	0.030	4.137	0.063	98.1	186.8	**0.131**	0.587	3.370
+ Time-gated fusion (HSA-Mol3D)	**0.145**	**0.009**	**1.65**	**0.060**	**99.9**	**70.6**	**0.092**	**0.665**	2.91

Controlled ablations of hypergraph construction. We further perform controlled component ablations to isolate the effects of bond-aware candidate scoring and learned proposal refinement. All three variants use the same annotations, training and validation splits, shared generative backbone, loss configuration. The *No bond-aware* variant removes only the predicted bond-existence modulation from the typed pair score:

$$\widetilde{s}_{ij,t}^{(r)} = s_{ij,t}^{(r)} \qquad \text{instead of} \qquad \widetilde{s}_{ij,t}^{(r)} = s_{ij,t}^{(r)} p_{ij,t}^{\mathrm{bond}}. \tag{11}$$

It otherwise retains topology-aware proposal construction, learned set refinement, set-level supervision, and staged hypergraph fusion.

The *Rule-only* variant retains the local typed node and pair scores and topology-aware connected-component construction, but directly uses the resulting deterministic proposals as hyperedges. Learned membership refinement, objectness prediction, and the corresponding set-level refinement losses are disabled. The resulting rule-based hyperedges are still encoded and injected into the generative backbone through the same staged fusion module.

The *Full HSA* variant uses both bond-aware candidate scoring and learned proposal refinement. Thus, the three variants differ only in the controlled components under evaluation.

Table 7: Controlled comparison of typed hyperedge construction on the GEOM-Drugs validation set at $t = 1.0$. For all variants, we evaluate the raw typed proposals produced by their corresponding hyperedge-construction modules before set-level membership refinement and hypergraph fusion. A predicted hyperedge and a ground-truth hyperedge are considered matched when they have the same structural type and their atom-set IoU is at least $\delta_{\text{match}} = 0.50$. P denotes prediction precision, and R_{cov} denotes non-exclusive ground-truth coverage recall. Count MAE is the per-molecule absolute error in the number of typed hyperedges. Pred. and GT denote the average numbers of predicted and ground-truth hyperedges per molecule. Higher values are better for P, R_{cov}, and F1, whereas lower values are better for Count MAE. Best values within each structural type are shown in bold.

Variant	Type	$P \uparrow$	$R_{\text{cov}} \uparrow$	F1↑	Count MAE↓	Pred.	GT
No bond-aware	Ring	**0.9995**	0.8945	0.9441	0.6823	2.3215	3.0038
No bond-aware	Aromatic	0.9869	**0.9751**	**0.9810**	0.0736	1.9276	1.9521
No bond-aware	Conjugation	0.9969	**0.9366**	0.9658	0.1277	1.8053	1.9209
Rule-only	Ring	0.9989	**0.9731**	**0.9858**	**0.0857**	2.9181	3.0038
Rule-only	Aromatic	0.9603	0.9714	0.9659	0.1698	2.0666	1.9521
Rule-only	Conjugation	0.7545	0.7332	0.7437	0.5699	1.8666	1.9209
Full HSA	Ring	**0.9995**	0.8708	0.9307	0.7416	2.2622	3.0038
Full HSA	Aromatic	**0.9902**	0.9720	**0.9810**	**0.0730**	1.9153	1.9521
Full HSA	Conjugation	**0.9990**	0.9365	**0.9668**	**0.1233**	1.8008	1.9209

Typed hyperedge matching metrics. For structural type $r \in \{\text{ring}, \text{arom}, \text{conj}\}$, let

$$\widehat{H}^{(r)} = \left\{ \widehat{E}_k^{(r)} \right\}_{k=1}^{\widehat{M}_r}, \qquad H^{\star(r)} = \left\{ E_m^{\star(r)} \right\}_{m=1}^{M_r^\star}.$$

denote the predicted and ground-truth hyperedge sets, respectively. Their atom-set overlap is measured by

$$\text{IoU}\left(\widehat{E}_k^{(r)}, E_m^{\star(r)}\right) = \frac{\left|\widehat{E}_k^{(r)} \cap E_m^{\star(r)}\right|}{\left|\widehat{E}_k^{(r)} \cup E_m^{\star(r)}\right|}. \tag{12}$$

A predicted and a ground-truth hyperedge are considered matched when they have the same structural type and their atom-set IoU is at least δ_{match}. We use non-exclusive matching, so multiple predictions may cover the same ground-truth component. Precision and ground-truth coverage recall are defined as

$$P_r = \frac{1}{\widehat{M}_r} \sum_{k=1}^{\widehat{M}_r} \mathbb{I}\left[\max_m \text{IoU}\left(\widehat{E}_k^{(r)}, E_m^{\star(r)}\right) \geq \delta_{\text{match}}\right], \tag{13}$$

and

$$R_{\text{cov},r} = \frac{1}{M_r^\star} \sum_{m=1}^{M_r^\star} \mathbb{I}\left[\max_k \text{IoU}\left(\widehat{E}_k^{(r)}, E_m^{\star(r)}\right) \geq \delta_{\text{match}}\right]. \tag{14}$$

The corresponding F1 score is

$$F_{1,r} = \frac{2 P_r R_{\text{cov},r}}{P_r + R_{\text{cov},r}}. \tag{15}$$

Count MAE measures the absolute difference between the numbers of predicted and ground-truth hyperedges per molecule. Pred. and GT denote their corresponding average counts. All variants are evaluated on the same GEOM-Drugs validation molecules at $t = 1.0$.

Table 7 compares the three construction mechanisms using a common raw-proposal protocol. The comparison between Full HSA and No bond-aware isolates the effect of multiplying the learned typed pair evidence by the predicted bond-existence probability. The comparison with Rule-only contrasts the learned score-based, bond-aware proposal constructor with deterministic topology-based proposal construction. Because this table evaluates proposals before set-level refinement, it

Table 8: Controlled ablation of the main components of HSA-Mol3D on GEOM-Drugs. *No fusion* removes explicit hypergraph feedback from the generative backbone; *Rule-only* uses deterministic topology-based hyperedges without learned proposal refinement; *No bond-aware* removes bond-existence modulation while retaining learned hypergraph construction and staged fusion. Higher values are better for MolStable, V&U&N, and QED, whereas lower values are better for Ring-TV, bond-angle W_1, bond-length W_1, Energy, and Atoms/mol TV.

Variant	Learned Refinement	Bond-aware	Fusion	Mol Stable↑	V&U&N ↑	Ring-TV ↓	Bond ang. W_1 ↓	Bond len. W_1 ↓	Energy ↓	QED ↑	Atoms/mol TV↓
No fusion	–	–	×	0.980	0.913	0.170	3.506	0.023	144.9	0.660	0.288
Rule-only	×	✓	✓	**0.995**	0.917	0.162	2.621	0.038	138.2	**0.676**	**0.118**
No bond-aware	✓	×	✓	0.994	0.950	0.172	2.813	0.026	152.9	0.649	0.123
Full HSA	✓	✓	✓	**0.995**	**0.964**	**0.092**	**1.650**	**0.009**	**70.6**	0.665	0.145

does not directly measure the effects of learned objectness, membership refinement, or staged fusion; those components are evaluated through the downstream generation results.

The three construction mechanisms exhibit distinct type-specific behaviors. Rule-only obtains the strongest ring reconstruction, achieving an F1 score of 0.9858 and a Count MAE of only 0.0857. It also recovers aromatic structures accurately, but its conjugation performance is substantially weaker, with an F1 score of 0.7437. This indicates that deterministic topology-based rules provide a particularly effective prior for cyclic structures but do not capture all conjugation relations equally well.

Full HSA achieves the strongest or tied-best performance for aromatic and conjugation proposals. For aromatic components, it obtains the highest precision (0.9902), a tied-best F1 score of 0.9810, and the lowest Count MAE of 0.0730. For conjugation components, Full HSA achieves the highest precision (0.9990), F1 score (0.9668), and lowest Count MAE (0.1233). These results show that the learned bond-aware constructor provides more balanced recovery across aromatic and conjugated structures, although it does not outperform Rule-only for ring reconstruction.

The effect of bond-aware modulation is comparatively small and type-dependent. Removing the predicted bond-existence factor slightly improves ring coverage and F1, while Full HSA provides slightly better aromatic precision and count accuracy as well as better conjugation precision, F1, and count accuracy.

Component ablation. Table 8 summarizes the contribution of the main components of HSA-Mol3D to downstream generation quality. The *No fusion* baseline performs reasonably well in molecular stability, but shows clearly weaker V&U&N, ring-distribution fidelity, geometric quality, and energy. Introducing rule-based hypergraph construction improves several global molecular statistics, including MolStable, QED, and atom-count TV, indicating that deterministic higher-order priors already provide useful structural guidance. However, Rule-only remains substantially weaker than Full HSA in V&U&N, Ring-TV, bond-angle W_1, bond-length W_1, and energy.

Removing bond-aware modulation also degrades the overall balance of the model, particularly in ring-distribution fidelity, geometry, and energetic quality. In contrast, Full HSA achieves the strongest overall performance across validity–novelty, higher-order structural fidelity, and 3D geometric quality, with the best V&U&N, Ring-TV, bond-angle W_1, bond-length W_1, and energy among the controlled variants. These results indicate that learned proposal refinement, bond-aware construction, and explicit hypergraph fusion provide complementary benefits, while the complete HSA design yields the most balanced generation behavior.

A.2 Hypergraph time-step construction analysis

To further understand how the dynamic hypergraph state evolves during generation, we visualize representative sampling trajectories under the proposed time-gated fusion strategy. We select six intermediate time steps, i.e., $t = 0.10, 0.30, 0.51, 0.70, 0.90$, and 1.00, to cover the early, middle, and late stages of the generation process. These time steps are chosen to show how the molecular state changes from a noisy intermediate configuration to a more chemically meaningful structure, and how the predicted higher-order hyperedges emerge along this process.

As shown in Fig. 6, the early stages contain highly uncertain atom and coordinate states, and the predicted hypergraph structure is either absent or weak. When the generation process moves to later stages, the molecular topology becomes clearer, and the dynamic hypergraph branch begins to identify meaningful ring and aromatic components. In particular, ring and aromatic hyperedges appear mainly at later time steps, where the corresponding cyclic structures are already partially formed. This observation supports the motivation of delaying hypergraph fusion, since early hypergraph construction can be unreliable when the molecular state is still noisy.

Fig. 7 further shows the temporal evolution of aromatic and conjugation-related hyperedges. The results indicate that conjugation patterns are not stable at the beginning of generation, but become more coherent after the molecular skeleton and local bonding patterns are gradually established. At the final stage, the predicted aromatic and conjugation hyperedges are aligned with chemically meaningful substructures, suggesting that the dynamic hypergraph branch can capture high-order molecular relations from the evolving joint state.

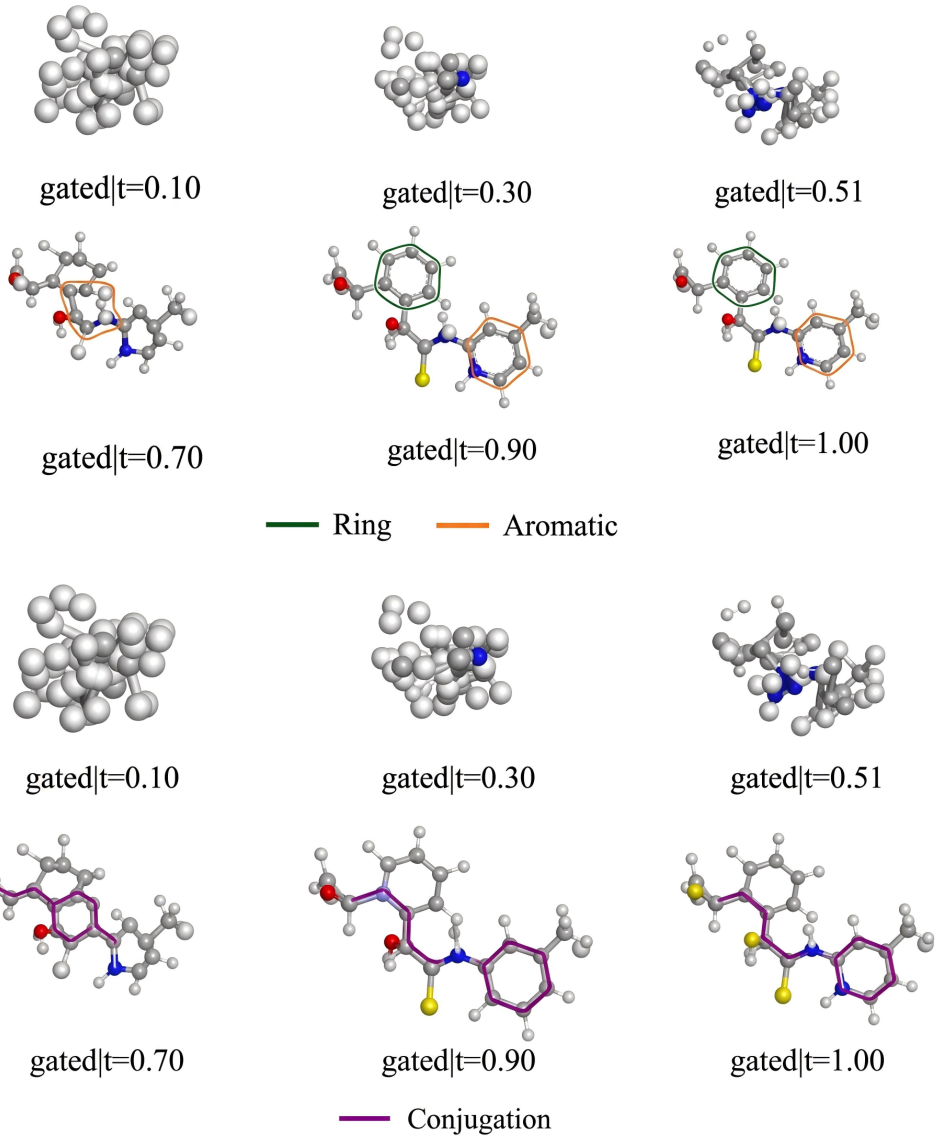

Figure 6: Temporal evolution of predicted dynamic hypergraphs under the time-gated fusion strategy. The upper part visualizes the emergence of ring and aromatic hyperedges during the generation process. Green contours denote predicted ring hyperedges, orange contours denote predicted aromatic hyperedges, and purple contours denote predicted conjugation hyperedges.

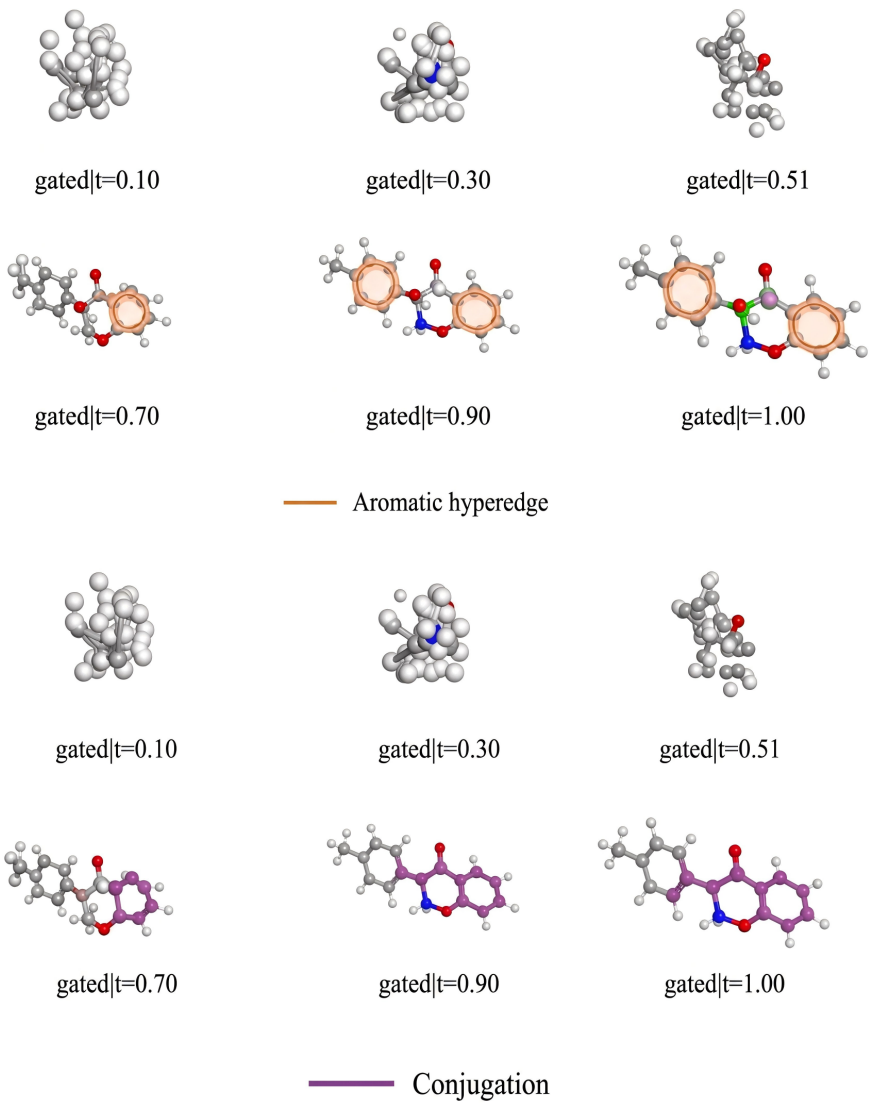

Figure 7: Temporal evolution of predicted dynamic hypergraphs under the time-gated fusion strategy. The figure shows representative generation trajectories where high-order molecular patterns gradually become clearer in later denoising steps. Green contours denote predicted ring hyperedges, orange contours denote predicted aromatic hyperedges, and purple contours denote predicted conjugation hyperedges.

B IMPLEMENTATION DETAILS OF HSA-MOL3D

This section provides implementation details of the shared joint-flow backbone, the dynamic typed hypergraph branch, set-level supervision, staged fusion, training objectives, and sampling procedure.

B.1 BACKBONE AND TRAINING CONFIGURATION

We model eight chemical atom types for GEOM-Drugs (C, H, O, N, F, S, Cl, and Br) and five for QM9 (C, H, O, N, and F). Formal charge and Kekulized bond vocabularies contain six and four classes, respectively. The scalar node, pair, and hyperedge dimensions are 256, 128, and 256. The geometric trunk uses 16 vector channels, a 64-dimensional time embedding, a 32-dimensional radial basis embedding, and a radial cutoff of 10.0. The remaining backbone settings are summarized in Table 10.

Cosine interpolation. We use the cosine schedule $\alpha(t) = 1 - \cos^2\left(\frac{\pi}{2}t^\nu\right)$ with $\nu = 1$. The coordinate state is constructed as $\mathbf{x}_t = (1 - \alpha_x(t))\mathbf{x}_0 + \alpha_x(t)\mathbf{x}_1$, where $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, followed by center-of-mass removal. For each discrete variable, a target token is revealed with probability $\alpha(t)$ and otherwise replaced by its absorbing mask token. The atom, coordinate, charge, and bond losses are time weighted by $w(t) = \mathrm{clip}_{[0.05,1.5]}\left(\frac{\alpha(t)}{1-\alpha(t)+\epsilon}\right)$.

B.2 DYNAMIC TYPED HYPERGRAPH CONSTRUCTION

Time-dependent type activation. The dynamic builder is disabled for highly noisy states. The active higher-order type set is

$$\mathcal{R}_t = \begin{cases} \varnothing, & t < t_{\min}, \\ \{\mathrm{ring}, \mathrm{arom}\}, & t_{\min} \le t < t_{\mathrm{mid}}, \\ \{\mathrm{ring}, \mathrm{arom}, \mathrm{conj}\}, & t \ge t_{\mathrm{mid}}, \end{cases} \tag{16}$$

where $t_{\min} = 0.20$ and $t_{\mathrm{mid}} = 0.55$.

Local typed evidence. For each type $r \in \mathcal{R}$, the local heads predict node and pair probabilities $s_{i,t}^{(r)} = \sigma(\ell_{i,t}^{(r)})$ and $s_{ij,t}^{(r)} = \sigma(\ell_{ij,t}^{(r)})$. The shared bond head predicts a categorical distribution $\boldsymbol{\pi}_{ij,t}^{\mathrm{bond}}$ over four Kekulized bond classes, from which the bond-existence probability is computed as $p_{ij,t}^{\mathrm{bond}} = \sum_{c=1}^{K_b-1}(\boldsymbol{\pi}_{ij,t}^{\mathrm{bond}})_c$. For the full model, pair evidence is modulated by the predicted bond probability, $\widetilde{s}_{ij,t}^{(r)} = s_{ij,t}^{(r)} p_{ij,t}^{\mathrm{bond}}$, whereas the *No bond-aware* variant directly uses $\widetilde{s}_{ij,t}^{(r)} = s_{ij,t}^{(r)}$.

For each active type, candidate nodes and pairs are retained when $s_{i,t}^{(r)} > a$ and $\widetilde{s}_{ij,t}^{(r)} > b$, with $a = 0.45$ and $b = 0.35$. Endpoints of retained pairs are included even when their individual node scores do not exceed a. The retained pairs form a sparse candidate graph $\mathcal{G}_t^{(r)}$, whose connected components define the coarse typed proposals.

Threshold sensitivity. To examine whether the hypergraph construction depends on a narrowly tuned pair threshold, we vary b over a broad range while keeping the remaining settings fixed. As shown in Fig. 8, the recovery performance is highly stable across all three structural types. The maximum F1 variations are only 0.003, 0.003, and 0.004 for ring, aromatic, and conjugation hyperedges, respectively. The predicted-to-ground-truth count ratios also remain nearly unchanged, indicating that moderate changes in b do not lead to either excessive hyperedge proliferation or substantial structural loss.

B.3 TYPE-SPECIFIC COMPONENT FILTERING.

The connected components extracted from each typed candidate graph are filtered using structure-specific criteria before being converted into hyperedge proposals.

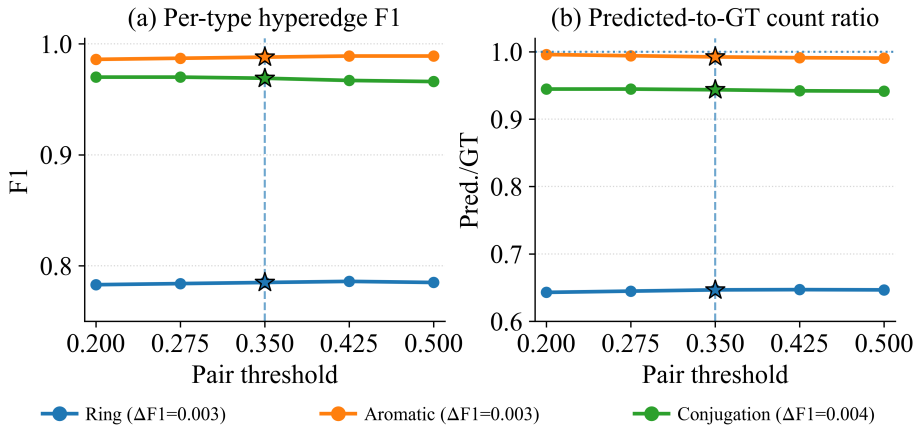

Figure 8: Sensitivity of typed hyperedge construction to the pair threshold b. (a) Per-type hyperedge F1 and (b) predicted-to-ground-truth hyperedge count ratio across different threshold values. The dashed line and star markers indicate the default setting $b = 0.35$. Recovery performance remains stable over the evaluated range, with maximum F1 variations of only 0.003, 0.003, and 0.004 for ring, aromatic, and conjugation structures, respectively.

Ring components. Ring proposals are filtered using topology rather than a hard minimum-density threshold. For each connected candidate $\mathcal{C}$, we extract its exact 2-core, $\mathcal{C}^{\text{core}} = 2\text{Core}(\mathcal{C})$. Let $V_{\mathcal{C}}$ and $E_{\mathcal{C}}$ denote the vertices and retained candidate edges in the resulting core. Its cycle rank is $\mu(\mathcal{C}^{\text{core}}) = |E_{\mathcal{C}}| - |V_{\mathcal{C}}| + 1$. A ring proposal is retained only if

$$3 \le |V_{\mathcal{C}}| \le 12, \qquad \mu(\mathcal{C}^{\text{core}}) \ge 1. \tag{17}$$

Thus, density is not used as a hard rejection criterion for ring candidates, preventing ordinary sparse cycles from being discarded solely because of their complete-graph density.

Aromatic and conjugation components. For $r \in \{\text{arom}, \text{conj}\}$, a connected proposal $E_{k,t}$ is retained when

$$3 \le |E_{k,t}| \le m_{\max}^{(r)}, \qquad \text{dens}(E_{k,t}) \ge \rho_{\min}^{(r)}, \tag{18}$$

with $m_{\max}^{(\text{arom})} = 18$, $m_{\max}^{(\text{conj})} = 64$, $\rho_{\min}^{(\text{arom})} = 0.22$, and $\rho_{\min}^{(\text{conj})} = 0.12$. For any candidate atom set E, $\text{dens}(E) = |\mathcal{P}_t^{(r)}(E)|/[|E|(|E|-1)/2]$, where $\mathcal{P}_t^{(r)}(E)$ denotes the retained typed pairs whose endpoints both lie in E. Aromatic and conjugation proposals are constructed independently and may overlap.

Soft confidence and top-k selection. For a retained proposal $E_{k,t}$ of type r, we compute the mean node and pair scores as $\bar{s}_{k,t}^{\text{node}} = |E_{k,t}|^{-1} \sum_{i \in E_{k,t}} s_{i,t}^{(r)}$ and $\bar{s}_{k,t}^{\text{pair}} = |\mathcal{P}_t^{(r)}(E_{k,t})|^{-1} \sum_{(i,j) \in \mathcal{P}_t^{(r)}(E_{k,t})} \tilde{s}_{ij,t}^{(r)}$. The size penalty is $p_{k,t}^{\text{size}} = \max\left(\frac{|E_{k,t}| - 0.6 m_{\max}^{(r)}}{m_{\max}^{(r)}}, 0\right)$. The resulting proposal confidence is

$$c_{k,t} = \text{clip}_{[0,1]}\left(0.50\bar{s}_{k,t}^{\text{node}} + 0.35\bar{s}_{k,t}^{\text{pair}} + 0.20\,\text{dens}(E_{k,t}) - 0.10\min(p_{k,t}^{\text{size}}, 0.5)\right). \tag{19}$$

For ring proposals, density is used only as a soft ranking feature rather than a hard topology constraint. We retain at most $k_{\text{ring}} = 4$, $k_{\text{arom}} = 4$, and $k_{\text{conj}} = 6$ proposals per type, merge them by confidence, and keep at most $M_{\max} = 12$ hyperedges per molecule.

B.4 Learned Set Refinement

For each retained proposal $E_{k,t}$, we first pool its pre-fusion node and pair representations to obtain a compact proposal descriptor. To avoid trivial type leakage, the supervision heads do not use the builder-assigned hard type embedding. Instead, type, objectness, and membership predictions are conditioned on pooled structural features, generation time, proposal size, and base confidence.

The proposal-level heads predict the structural type and objectness, while the membership head assigns each atom a soft membership score $m_{k,i,t} = \sigma(\ell^{\mathrm{mem}}_{k,i,t})$. For each atom i, the membership logit is computed as

$$\ell^{\mathrm{mem}}_{k,i,t} = \frac{\langle W_q \mathbf{z}_{k,t}, W_k \mathbf{h}^{\mathrm{pre}}_{i,t} \rangle}{\sqrt{d_h}} + \beta_{\mathrm{cand}} \mathbb{I}[i \in E_{k,t}], \qquad \beta_{\mathrm{cand}} = 1.0. \tag{20}$$

The learned membership probabilities are then used to refine the node aggregation associated with each proposal:

$$\mathbf{a}^{\mathrm{ref,node}}_{k,t} = \sum_{i=1}^{N} \frac{m_{k,i,t}}{\sum_j m_{k,j,t} + \epsilon} W_v \mathbf{h}^{\mathrm{pre}}_{i,t}. \tag{21}$$

We combine this refined aggregation with the original candidate-based aggregation as

$$\mathbf{a}^{\mathrm{node}}_{k,t} = (1 - \alpha_{\mathrm{mem}}) \mathbf{a}^{\mathrm{cand,node}}_{k,t} + \alpha_{\mathrm{mem}} \mathbf{a}^{\mathrm{ref,node}}_{k,t}, \qquad \alpha_{\mathrm{mem}} = 0.5. \tag{22}$$

The refined node representation is combined with the corresponding pair-level aggregation to form the final hyperedge representation, following the attention-based encoding described in Section 3.2.

B.5 Set Matching and Supervision

Predicted hyperedge proposals are matched to ground-truth typed hyperedges using Hungarian assignment based on atom-set overlap. For a predicted proposal $\widehat{E}_{k,t}$ and a ground-truth hyperedge $E^\star_m$, the overlap is measured by $\mathrm{IoU}(\widehat{E}_{k,t}, E^\star_m) = |\widehat{E}_{k,t} \cap E^\star_m| / |\widehat{E}_{k,t} \cup E^\star_m|$. A matched pair is treated as positive only when

$$\mathrm{IoU}(\widehat{E}_{k,t}, E^\star_{\pi(k)}) \geq \delta_{\mathrm{train}}, \qquad \delta_{\mathrm{train}} = 0.10. \tag{23}$$

Positive proposals receive the ground-truth structural type, objectness target one, and dense atom-membership labels, while unmatched valid proposals are assigned the background type and objectness target zero.

Because set-level predictions are less reliable at highly noisy states, their supervision is weighted by $w_{\mathrm{set}}(t) = \sigma((t - t_{\mathrm{set}})/\epsilon_{\mathrm{set}})$, with $t_{\mathrm{set}} = 0.50$ and $\epsilon_{\mathrm{set}} = 0.05$. Type and objectness are optimized with cross-entropy and binary cross-entropy losses, respectively.

Membership supervision is applied only to positive matches. We combine binary cross-entropy and soft Dice losses as

$$\begin{aligned} \mathcal{L}_{\mathrm{mem}} = & -\frac{1}{N} \sum_i \left[y^{\mathrm{mem}}_{k,i} \log m_{k,i,t} + (1 - y^{\mathrm{mem}}_{k,i}) \log(1 - m_{k,i,t}) \right] \\ & + 1 - \frac{2 \sum_i m_{k,i,t} y^{\mathrm{mem}}_{k,i} + \epsilon}{\sum_i m_{k,i,t} + \sum_i y^{\mathrm{mem}}_{k,i} + \epsilon}. \end{aligned} \tag{24}$$

For the reported typed-hyperedge recovery metrics, we use the stricter threshold $\delta_{\mathrm{metric}} = 0.50$, which is used only for evaluation and is intentionally higher than the training assignment threshold.

B.6 Staged Hypergraph Fusion

The staged fusion mechanism follows the formulation in Section 3.3. Here we provide the implementation details of the proposal gate and curriculum schedule. Let $s \in \{A, B, C, D\}$ denote the current training stage. A hyperedge proposal is activated only when both its generation time and base confidence satisfy the stage-specific thresholds,

$$A^{(s)}_{k,t} = \mathbb{I}[t \geq t^{(s)}_{\mathrm{on}}] \mathbb{I}[c_{k,t} \geq \tau^{(s)}_c]. \tag{25}$$

Learned objectness further rescales each activated proposal:

$$\omega^{(s)}_{k,t} = A^{(s)}_{k,t} \left[(1 - \alpha_{\mathrm{obj}}) + \alpha_{\mathrm{obj}} \sigma(\ell^{\mathrm{obj}}_{k,t}) \right], \qquad \alpha_{\mathrm{obj}} = 0.5. \tag{26}$$

The resulting weight is applied during hypergraph-message aggregation. The overall fusion strength is additionally controlled by the learned time gate $g_t = \sigma(F_{\text{time}}(\text{Emb}(t)))$ and the stage coefficient η_s. The aggregated node- and pair-level messages are then injected into the backbone through the residual fusion defined in Section 3.3.

Table 9: Stage-specific settings of the hypergraph fusion curriculum.

Stage	Epochs	$t_{\text{on}}^{(s)}$	$\tau_c^{(s)}$	η_s
A	1–10	1.01	1.01	0.00
B	11–25	0.70	0.75	0.10
C	26–45	0.60	0.70	0.18
D	46–80	0.50	0.60	0.25

Stage A disables hypergraph fusion by setting $\eta_s = 0$. During the subsequent stages, the activation time and confidence thresholds are gradually relaxed while the fusion coefficient is increased, allowing higher-order structural information to contribute progressively as the predicted molecular state becomes more reliable.

B.7 Training Objectives

The complete training objective follows the formulation in the main text:

$$\mathcal{L} = \mathcal{L}_{\text{flow}} + \mathcal{L}_{\text{local}} + \mathcal{L}_{\text{type}} + \mathcal{L}_{\text{obj}} + \mathcal{L}_{\text{mem}}. \tag{27}$$

The molecular generation term is

$$\mathcal{L}_{\text{flow}} = \mathcal{L}_{\text{atom}} + \mathcal{L}_{\text{coord}} + \mathcal{L}_{\text{charge}} + \mathcal{L}_{\text{bond}} + 0.12\mathcal{L}_{\text{valence}}. \tag{28}$$

Atom, charge, and bond endpoint predictions use cross-entropy losses, while coordinates are trained with a time-weighted endpoint regression loss and valence with a Smooth-L_1 objective.

The local higher-order supervision is

$$\begin{aligned}\mathcal{L}_{\text{local}} = {} & 0.08\mathcal{L}_{\text{atom,arom}} + 0.05\mathcal{L}_{\text{atom,ring}} + 0.08\mathcal{L}_{\text{atom,conj}} \\ & + 0.10\mathcal{L}_{\text{pair,arom}} + 0.05\mathcal{L}_{\text{pair,ring}} + 0.08\mathcal{L}_{\text{pair,conj}},\end{aligned} \tag{29}$$

where all six terms are binary cross-entropy-with-logits losses.

The set-level terms $\mathcal{L}_{\text{type}}$ and $\mathcal{L}_{\text{obj}}$ use cross-entropy and binary cross-entropy, respectively, as defined in Section B.5. Membership supervision is applied only to positive matched proposals and combines binary cross-entropy and soft Dice losses, $\mathcal{L}_{\text{mem}} = \mathcal{L}_{\text{mem,BCE}} + \mathcal{L}_{\text{mem,Dice}}$.

The type, objectness, and membership terms are further weighted by the time-dependent supervision factor $w_{\text{set}}(t) = \sigma((t - t_{\text{set}})/\epsilon_{\text{set}})$, with $t_{\text{set}} = 0.50$ and $\epsilon_{\text{set}} = 0.05$.

B.8 Training and Sampling Procedures

Algorithm 1 summarizes the training procedure of HSA-Mol3D, while Algorithm 2 describes the iterative sampling and decoding process.

Sampling configuration. Sampling starts from centered Gaussian coordinates together with mask tokens for atom, charge, and bond states. We use 250 sampling time points,purity-based token selection, and categorical temperature 0.05 for discrete variables.

For two consecutive times $t < s$, the coordinate step size is $\gamma_x(t, s) = \text{clip}_{[0,1]}\left(\frac{\alpha_x(s) - \alpha_x(t)}{1 - \alpha_x(t) + \epsilon}\right)$, and coordinates are updated by

$$\mathbf{x}_s = \text{Center}\left[\mathbf{x}_t + \gamma_x(t, s)\left(\widehat{\mathbf{x}}_1 - \mathbf{x}_t\right)\right]. \tag{30}$$

For discrete variables, revealed tokens are remasked with probability $p_{\text{remask}} = \text{clip}_{[0,1]}(\lambda_{\text{mask}}\Delta t)$, while masked tokens are selected for revelation with probability $p_{\text{unmask}} = \text{clip}_{[0,1]}\left(\frac{\Delta t(1+\lambda_{\text{mask}}t)}{1-t+\epsilon}\right)$. Purity sampling prioritizes masked positions with the largest maximum class probabilities, after which token values are sampled from the temperature-scaled categorical distributions.

Algorithm 1: Training Procedure of HSA-Mol3D

Input: training set $\mathcal{D}$, model parameters θ
Output: updated model parameters θ
while *not converged* **do**
  Sample a mini-batch $\mathcal{B} \subset \mathcal{D}$ and $t \sim \mathcal{U}(0,1)$;
  Sample centered prior coordinates $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and construct the intermediate molecular state $\mathcal{M}_t$;
  Compute pre-fusion node and pair representations using the shared geometric trunk;
  Predict local typed node/pair evidence and bond-existence probabilities;
  Determine the active structural type set $\mathcal{R}_t$;
  for $r \in \mathcal{R}_t$ **do**
    Compute the bond-aware pair evidence $\widetilde{s}_{ij,t}^{(r)}$;
    Retain candidate nodes and pairs using the node and pair thresholds;
    Build the sparse typed candidate graph $\mathcal{G}_t^{(r)}$ and extract its connected components;
    if $r = \text{ring}$ **then**
      Extract the exact 2-core and retain cyclic components satisfying the ring criterion;
    else
      Apply the type-specific size and density filtering criteria;
    Compute the base proposal confidence $c_{k,t}$;
    Retain the top-k_r proposals for structural type r;
  Merge all typed proposals and retain at most $M_{\max} = 12$ hyperedges to form $\mathcal{H}_t$;
  Refine retained proposals and predict hyperedge type, objectness, and atom membership;
  Match predicted proposals to ground-truth hyperedges using Hungarian assignment;
  Apply staged hypergraph fusion;
  Predict endpoint coordinates, atom types, formal charges, and bond states;
  Compute $\mathcal{L}$ using Eq. equation 27;
  Update $\theta \leftarrow \theta - \eta_{lr} \nabla_\theta \mathcal{L}$;

Algorithm 2: Sampling and Decoding with HSA-Mol3D

Input: trained model f_θ, sampled from a learned atom-count prior $\{N_b\}$, sampling steps K
Output: generated molecular states
Initialize centered Gaussian coordinates $\mathbf{x}_0$;
Initialize atom, charge, and bond states as mask tokens;
Initialize the self-conditioning cache;
for $k \leftarrow 0$ **to** $K - 2$ **do**
  Set $t = t_k$, $s = t_{k+1}$, and $\Delta t = s - t$;
  Encode the current molecular state with the shared geometric trunk;
  Determine the active structural type set $\mathcal{R}_t$;
  for $r \in \mathcal{R}_t$ **do**
    Predict typed node/pair evidence and bond-existence probabilities;
    Build the sparse typed candidate graph and extract connected components;
    Apply type-specific component filtering and proposal scoring;
    Retain the top-k_r proposals for type r;
  Merge retained proposals to construct the dynamic hypergraph $\mathcal{H}_t$;
  Refine hyperedge representations and apply staged hypergraph fusion;
  Predict endpoint coordinates, atom types, formal charges, and bond states;
  Update coordinates using Eq. equation 30;
  Update discrete states using remasking and purity-based unmasking;
  Store detached endpoint predictions for self-conditioning;
Perform the final prediction at $t = 1$ and resolve any remaining masks;
Decode atom types, charges, bonds, and coordinates into molecular structures with RDKit;
return generated molecular states;

B.9 HYPERPARAMETER SUMMARY

The main implementation hyperparameters are summarized in Tables 10 and 11. The former reports the shared geometric-backbone configuration, while the latter lists the key parameters specific to dynamic typed hypergraph construction and refinement. Stage-dependent fusion parameters are reported separately in Table 9 and are therefore not repeated here.

Table 10: Key hyperparameters of the shared generation backbone.

Parameter	Value
Scalar node dimension	256
Pair dimension	128
Hyperedge dimension	256
Vector channels	16
Time embedding dimension	64
RBF dimension	32
RBF cutoff	10.0
GVPConv blocks	3×2
Self-conditioning probability	0.5
Cosine schedule exponent ν	1.0

Table 11: Key settings of the dynamic typed hypergraph branch.

Parameter	Value
Maximum hyperedges $M_{\max}$	12
Minimum component size	3
Node threshold a	0.45
Pair threshold b	0.35
Builder start time $t_{\min}$	0.20
Conjugation start time t_{mid}	0.55
Top-k (ring/arom./conj.)	4 / 4 / 6
Minimum aromatic density	0.22
Minimum conjugation density	0.12
Maximum size (ring/arom./conj.)	12 / 18 / 64
Training match threshold δ_{train}	0.10
Set-weight center t_{set}	0.50
Set-weight softness ϵ_{set}	0.05
Membership blend α_{mem}	0.50
Objectness blend α_{obj}	0.50

C ADDITIONAL QUALITATIVE RESULTS

We present some generated molecules in Fig. 9 and 10.

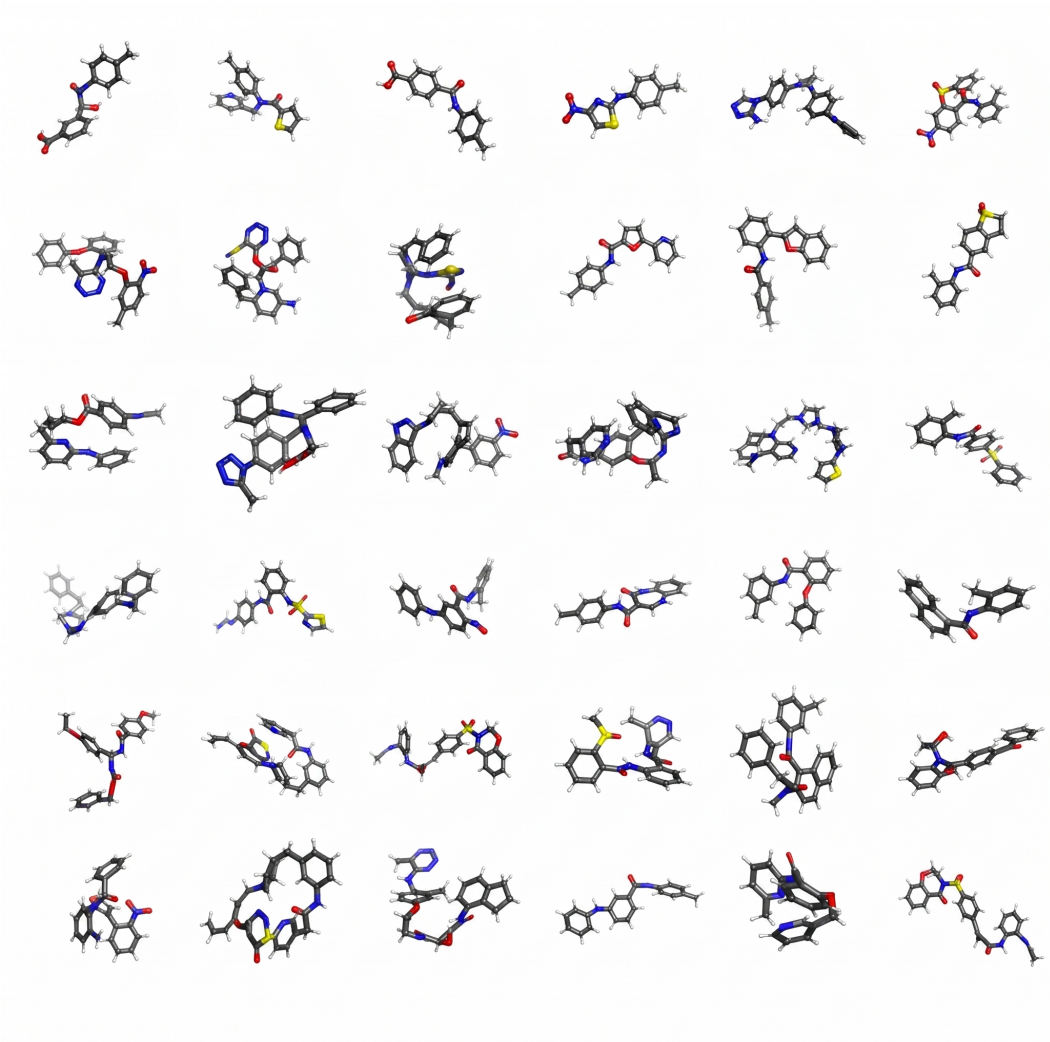

Figure 9: Visualization of random samples generated by HSA-Mol3D on GEOM-Drugs.

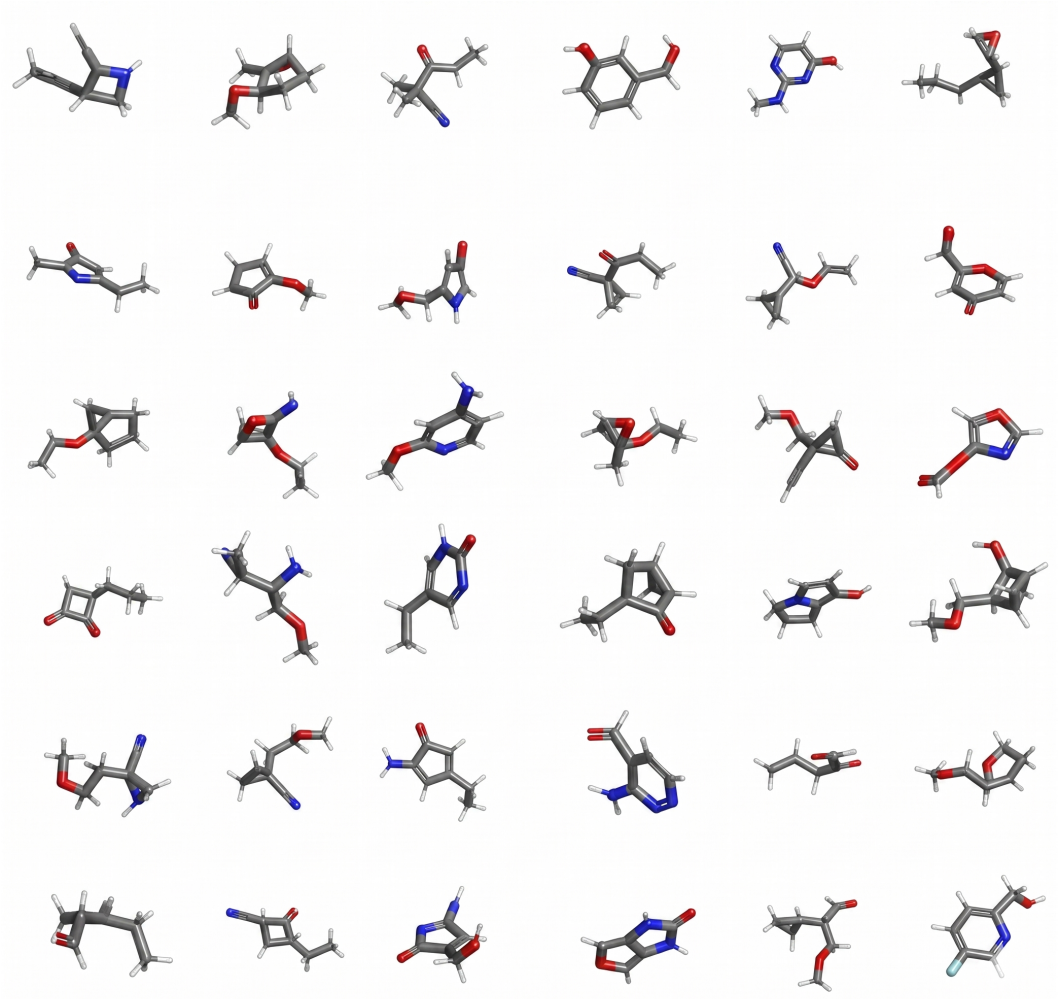

Figure 10: Visualization of random samples generated by HSA-Mol3D on QM9.