

SpectralMol: Spectral Graph Bisection Achieves Universal Molecular Fragment Decomposition for Drug-Like Molecule Generation

Published as part of Journal of Chemical Information and Modeling *special issue* “Chemical Compound Space Exploration by Multiscale High-Throughput Screening and Machine Learning”.

Gang Li,* Zhizhe Lin, Teng Zhou,* Weihua Bai,* Zhifeng Hao, and Keqin Li



Cite This: <https://doi.org/10.1021/acs.jcim.6c01818>



Read Online

ACCESS |



Metrics & More

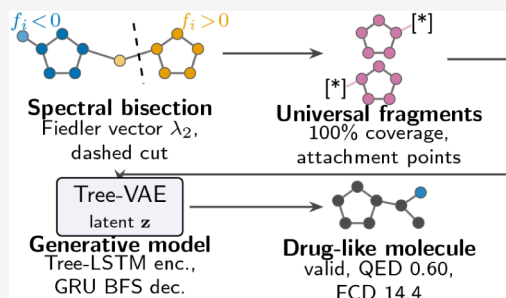


Article Recommendations



Supporting Information

ABSTRACT: Generating valid, diverse, and drug-like molecules remains a central challenge in computational drug discovery. A key bottleneck for fragment-based generative models is *decomposition coverage*: existing junction-tree methods fail on bridged bicyclic and spiro ring systems common in drug-like scaffolds. **Our primary contribution is Fiedler-BFS (breadth-first search) molecular decomposition**, a spectral bisection algorithm achieving 100% decomposition success on all 250,000 ZINC-250K molecules, including bridged and spiro systems where junction-tree methods fail. A decomposition ablation (5,000 molecules, 5 seeds) confirms that spectral structure drives the improvement: 91.4% vs 7.1% for random bisection. Compared to BRICS (break retrosynthetically interesting chemical substructures; a rule-based retrosynthesis method), Fiedler-BFS produces fewer fragments per molecule (2.31 vs 5.13), better suited for hierarchical generative modeling. **As a secondary contribution**, we instantiate the decomposition in a Tree-VAE (variational autoencoder) with classifier-free guidance over molecular properties, Inverse Multi-Quadratic MMD (maximum mean discrepancy) regularization, and gradient-based latent-space steering. Under a matched protocol in which the identical Tree-VAE is retrained while only the cut-bond set varies, Fiedler-BFS yields shorter, more learnable fragment trees than BRICS and junction-tree decomposition (validation reconstruction loss of 0.012 vs 0.048 vs 0.083). On ZINC-250K, the model achieves 100% chemical validity, 99.87% novelty, a mean QED (quantitative estimate of drug-likeness) of **0.603** (matching the ZINC-250K training mean), and an SA (synthetic accessibility score) of **2.70**, confirming compatibility with a hierarchical generative pipeline for drug-like molecule generation. Property conditioning reaches only a 12.2% QED hit rate and, under robust metrics, does not outperform an unconditioned baseline; the v2 vocabulary naturally generates drug-like molecules with QED $\approx$ 0.60, so stronger property guidance remains an open challenge.



INTRODUCTION

Generating chemically valid and structurally diverse molecules is a fundamental challenge in computational drug discovery and materials science. Molecular generative models must navigate the vast chemical space while satisfying hard validity constraints and—optionally—steering generation toward desired physico-chemical properties.

Graph-based variational autoencoders (VAEs) have emerged as a principled framework for this task, operating directly on molecular graphs and enabling continuous latent-space interpolation and optimization. The junction tree VAE (JT-VAE)¹ achieves 100% chemical validity through vocabulary-constrained decoding but relies on smallest-set-of-smallest-rings (SSSR) decomposition, which fails on bridged bicyclic and spiro ring systems—a nontrivial fraction of drug-like chemical space.¹ Hierarchical VAEs² and diffusion models^{3,4} have further improved distributional fidelity, while flow-based⁵ methods offer alternative generative priors.

Despite these advances, three limitations remain. First, the *decomposition coverage problem*: topology-constrained decomposition methods fail on a nontrivial fraction of complex molecules, limiting their applicability to diverse chemical libraries. Second, the *latent diversity problem*: VAEs trained with KL (Kullback–Leibler) divergence alone suffer from posterior collapse,⁶ producing repetitive samples; diversity metrics such as Uniqueness@1k are substantially below those of string-based models for some tree-structured methods (JT-VAE: 77.0% vs CharVAE: 99.9%). Third, the *property steering problem*: most tree-structured VAEs generate unconditionally and require posthoc property optimization, whereas end-to-end

Received: June 11, 2026

Revised: September 2, 2026

Accepted: September 3, 2026

conditioning enables tighter integration of property feedback during decoding.

In this work, we propose **SpectralMol**, a molecular generation framework that addresses all three limitations. Our core contribution is to replace the SSSR-based junction-tree decomposition with *Fiedler-vector* spectral decomposition,⁷ which partitions any molecular graph by the sign of the second Laplacian eigenvector. This decomposition succeeds for *every* molecular graph, imposes no topological restrictions, and produces a well-defined BFS tree of fragments. On ZINC-250K, **SpectralMol** achieves zero decomposition failures, including all 250,000 molecules with bridged bicyclics and spiro ring systems that cause junction-tree failures. This robustness is the primary practical advantage of the spectral approach and is a prerequisite for deployment on arbitrary chemical libraries.

The main contributions of this work are

- 1 **Fiedler-BFS molecular decomposition (primary contribution)**. A spectral bisection algorithm achieving 100% decomposition success rate on all 250,000 ZINC-250K molecules, including complex bridged bicyclic, spiro, and polycyclic systems where junction-tree decomposition fails. With sparse-graph ARPACK (Arnoldi package eigensolver), per-molecule time is $O(nk)$ where k is the number of Lanczos steps. The spectral approach imposes no topological constraints on ring systems or bonding patterns—this is the core novelty. A decomposition ablation (Table 1, 5 seeds) shows 91.4% vs 7.1% success

Table 1. Decomposition Ablation: Fiedler Spectral Bisection vs Alternative Partitioning Strategies on 5,000 ZINC-250K Molecules (5 Random Seeds, Mean $\pm$ Std)^a

Method	Success	Mean frag. size	Imbalance	Exact Recon.
Fiedler-BFS v2	91.4%	8.9	1.29	86.2 $\pm$ 1.1%
Random	7.1%	9.2	1.06	—
Degree-balanced	10.6%	9.5	1.06	—

^aSuccess: fraction returning chemically sanitizable fragments. Exact Recon.: exact SMILES match after decompose+reassemble (v2 assembler, 5×1000 molecules). Imbalance close to 1.0 indicates balanced splits. Note: Success <100% because this simplified ablation omits cut-bond reattachment fallbacks.

rate over random bisection, isolating the causal role of spectral structure. A rule-based comparison (Table 8, 5 seeds) confirms that BRICS produces 5.13 ± 0.01 frags/mol vs 2.31 ± 0.01 for Fiedler-BFS, demonstrating that Fiedler-BFS occupies a more suitable middle ground for generative modeling.

- 2 **Proof-of-concept generative validation (secondary contribution)**. To demonstrate that the Fiedler-BFS decomposition is learnable and decodable, we build a Tree-VAE: Child-Sum Tree-LSTM (long short-term memory) encoder, IMQ-MMD (maximum mean discrepancy with inverse multiquadratic kernel) latent regularization, and GRU (gated recurrent unit) BFS decoder. The model achieves 100% chemical validity and 99.87% novelty on ZINC-250K, confirming that the Fiedler-BFS tree is compatible with end-to-end training. **Scope:** This is proof-of-concept validation only. FCD (Fréchet ChemNet Distance,⁸ 14.41) is substantially higher than specialized models (JT-VAE: ~ 1.0), reflecting that the generative model is not the focus of this work. The FCD gap is attributable to the generative

model's distribution learning, not to the decomposition method (diagnostic: decomposition contributes $\Delta FCD = 0.009$; see Section "Distributional fidelity and FCD diagnostic").

- 3 **Property conditioning (exploratory; not a confirmed contribution)**. The Tree-VAE injects a normalized property embedding [MW, log P , QED, SA, TPSA] (molecular weight, log P , QED, synthetic accessibility, and topological polar surface area) via classifier-free guidance (CFG)⁹ (10% condition dropout). The conditioning signal is confirmed active in the ablation (E1 vs E2 Uniq@1k: 87.4% vs 86.5%), but hit rates under multitarget joint conditioning do not demonstrate clear superiority over the unconditioned model. This is reported as an exploratory finding rather than a primary contribution.
- 4 **Latent diversity analysis (tertiary contribution)**. We compare IMQ-kernel MMD¹⁰ to no-MMD ablation, finding higher uniqueness (87.4% vs 79.3% at $\tau = 1.0$) and lower FCD (14.41 vs 15.42), consistent with IMQ-MMD's theoretical advantage in high-dimensional latent space. A direct RBF (radial basis function) MMD baseline is identified as a priority for future experimental validation.

The generative validation on ZINC-250K confirms that the Fiedler-BFS decomposition is compatible with end-to-end training: the model achieves 100% chemical validity and 99.87% novelty (at $\tau = 1.0$; all runs $\geq 99.8\%$). At $\tau = 1.0$ (lower-temperature mode), Uniqueness@1k is 87.4%, QED = 0.603, SA = 2.70. At $\tau = 2.0$, Uniqueness@1k is 83.4% while QED rises to 0.631—capturing the fidelity/coverage trade-off, where the Fiedler-BFS tree structure still supports diverse decoding with drug-like chemistry. Property conditioning shifts the generation distribution toward targets (see Section efsec:conditional for per-target analysis). Ablation studies confirm the contribution of each component.

The remainder of this paper is organized as follows. Section "Related Work" reviews related work. Section "Methodology" presents the SpectralMol architecture. Section "Experiments" reports experimental results. Section "Conclusion" discusses limitations and concludes.

RELATED WORK

SMILES-Based Molecular Generation

Early deep generative models for molecules operated on SMILES strings, treating molecular generation as a sequence modeling problem. CharVAE¹¹ and GrammarVAE¹² demonstrated that VAEs can learn latent representations of chemical space, but string-level models frequently generate syntactically invalid sequences. Dai et al.¹³ introduced syntax-directed VAEs that enforce grammatical constraints, improving validity to over 90%. Autoregressive models such as MolGPT¹⁴ (transformer-based) and REINVENT¹⁵ (RNN with reinforcement learning) further advanced SMILES generation, achieving high validity through constrained generation. However, SMILES representations are inherently sequential and do not reflect molecular graph topology, limiting the model's ability to reason about structural motifs.¹⁶

Graph-Based VAEs

Operating directly on molecular graphs avoids the mismatch between sequential representations and graph-structured data.

GraphVAE¹⁷ uses one-shot graph generation with a node-and-edge adjacency matrix, but struggles with scalability for large molecules. CGVAE¹⁸ adds validity constraints at the decoding stage. The junction tree VAE (JT-VAE)¹ represents a landmark contribution: it decomposes molecules into a junction tree of substructures (rings and bonds), and learns a two-stage VAE that first generates the tree scaffold then assembles atoms. JT-VAE achieves 100% validity through vocabulary-constrained decoding but fails on molecules where junction-tree decomposition is not applicable—approximately 2–5% of ZINC-250K molecules contain bridged bicyclics or other fused ring systems incompatible with the SSSR-based decomposition.¹ HierVAE² extends this with a multiresolution hierarchical vocabulary and attention-based assembly, achieving improved FCD scores. Our Fiedler-BFS decomposition avoids all such topological restrictions.

Graph Diffusion and Flow Models

Score-based diffusion models adapted to graphs³ and discrete diffusion processes⁴ have recently achieved strong results on molecular generation benchmarks. DiGress⁴ models the noising process over categorical node and edge features, achieving near-perfect validity and strong FCD scores. GDSS³ applies stochastic differential equations jointly to node features and adjacency matrices. While diffusion models are powerful, they are computationally expensive at inference time (requiring many denoising steps) and do not naturally support fast gradient-based latent optimization for property targeting. SpectralMol's encoder-decoder structure enables direct latent-space optimization in a single forward pass of the property prediction heads.

Property-Conditioned Generation

Conditional molecular generation targets specific property profiles such as QED (quantitative estimate of drug-likeness¹⁹), log *P*, and synthetic accessibility. CVAE (conditional VAE)²⁰ and JTVAE extensions condition on scalar property values by appending them to the latent vector. REINVENT¹⁵ uses reinforcement learning to steer SMILES generation toward desired properties. TargetDiff²¹ and related target-aware diffusion models condition on 3D protein pocket geometry. Most closely related to our approach, Lim et al.²⁰ and Kang et al.²² inject property embeddings into VAE decoders for conditional generation. Our method differs by (a) using a spectral tree structure that enables reliable decomposition of all molecular topologies, (b) applying classifier-free guidance⁹ to the conditioning, and (c) providing multiproperty Pareto optimization directly in latent space.

Spectral Graph Methods and Graph Coarsening

The Fiedler vector (second eigenvector of the normalized graph Laplacian) has been widely studied for graph partitioning,⁷ spectral clustering,^{23,24} and community detection.²⁵ Spectral bisection is a classical algorithm in graph partitioning²⁶ and was subsequently applied to graph coarsening and hierarchical decomposition methods.²⁷ In graph neural networks, spectral convolutions^{28,29} leverage the graph spectrum for feature aggregation. DiffPool³⁰ learns to coarsen graphs hierarchically via differentiable pooling; GRACUS³¹ applies multilevel weighted graph cuts (without eigenvectors) for graph clustering and coarsening. Our work applies Fiedler-vector-based recursive decomposition in a different setting: building fragment vocabulary trees for molecular VAEs rather than learning graph representations. To our knowledge, SpectralMol is the

first work to apply spectral bisection to molecular fragment tree construction for generative modeling.

MMD Regularization in VAEs

The Wasserstein autoencoder (WAE)¹⁰ and InfoVAE³² replace the KL divergence with Maximum Mean Discrepancy to enforce the aggregated posterior to match the prior distribution. Critically, Tolstikhin et al.¹⁰ show that the Inverse Multi-Quadratic (IMQ) kernel $k(x, y) = C/(C + \|x - y\|^2)$ with $C = 2d$ (where d is the latent dimension) avoids the vanishing gradient problem of the RBF kernel in high dimensions. This observation motivates our choice of IMQ-MMD over the commonly used RBF-MMD or KL divergence, and our ablation study (Section “Ablation Study”) confirms its importance for achieving high diversity on ZINC-250K.

Learned Fragment Vocabularies

A parallel line of work asks whether the fragment vocabulary itself can be learned from data rather than fixed by chemical rules. PS-VAE³³ mines *principal subgraphs* by iteratively merging neighboring fragments in order of corpus frequency, and proves that its vocabulary covers every principal subgraph it extracts; MiCaM³⁴ extends frequency-based merging to record *connection* information, so that the generator selects a motif and its attachment pattern jointly and thereby avoids valence-valid but empirically implausible bonds. These methods and ours pursue coverage from opposite directions, and the distinction is worth stating precisely. A mined vocabulary is guaranteed to be closed under *its own* merge operator, but the guarantee is a property of the extracted vocabulary and remains contingent on the corpus: a ring system absent from the training distribution has no principal subgraph to fall back on. Our guarantee is instead structural and corpus-independent—Fiedler bipartition is defined for every connected graph, so the recursion terminates with a valid fragment tree for any molecule, in or out of distribution (Section “Theoretical Guarantees of Spectral Bisection”). The two mechanisms are complementary rather than competing: a frequency-mined vocabulary could be layered on top of our decomposition to shorten the tree, and our decomposition could serve as the fallback that makes a mined vocabulary total.

Discrete Flow Matching and Fragment-Level Generation

Since 2024 discrete flow matching (DFM) has largely displaced discrete diffusion as the state-of-the-art sampler for graphs. DeFoG³⁵ formulates graph generation as a continuous-time Markov chain with a linear interpolation noising path, decoupling training from sampling and matching or exceeding diffusion baselines using only 5–10% of their denoising steps. FragFM³⁶ is the most directly comparable recent system: it applies DFM at the *fragment* level and then reconstructs the atom-level graph, using a stochastic fragment-bag strategy specifically to escape the fixed-library restriction that we address by construction. Closest in architecture to SpectralMol is FRATTVAE,³⁷ which also encodes molecules as fragment trees inside a VAE, but derives its fragments from BRICS and its tree positional encoding from a transformer rather than a Tree-LSTM. We situate our contribution carefully against this literature. These systems report distributional metrics well beyond ours—FragFM attains an FCD of 0.66 on MOSES, whereas SpectralMol reaches 14.41 on ZINC-250K—and we do not claim competitive generation quality; our matched-protocol study (Section “Decomposition Analysis”) attributes that gap to the Tree-VAE backbone rather than to the decomposition,

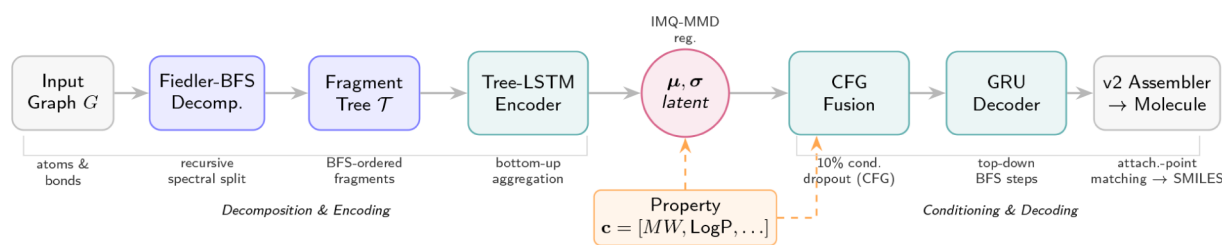


Figure 1. Overview of the SpectralMol pipeline. A molecular graph G is recursively decomposed via Fiedler-vector sign partitioning into a BFS-ordered fragment tree $\mathcal{T}$. A Tree-LSTM encoder aggregates the tree bottom-up into latent parameters (μ, σ) ; a latent code z is sampled and regularized with IMQ-MMD. At generation time, a normalized property vector c is fused with z via classifier-free guidance (10% condition dropout during training). A GRU-based BFS decoder reconstructs the fragment tree top-down; the deterministic v2 fragment assembler then stitches the predicted fragments together by matching paired attachment-point (dummy-atom) labels, yielding a chemically valid SMILES string. The decomposition and assembler (blue boxes) are parameter-free deterministic modules, whereas the Tree-LSTM encoder, CFG fusion, and GRU decoder (teal) are learned; this modular boundary is discussed in Section “End-to-end integration vs the current modular design” (end-to-end integration).

which changes FCD by only 0.009 under decompose-and-reassemble. The claim we do make is orthogonal: the fragment layer these methods rely on can be made *total* without chemical rules or a mined library, and because FragFM and FRATTVAE both treat fragmentation as an interchangeable preprocessing stage, a spectral decomposition with a coverage guarantee is directly substitutable into either. Recent surveys of graph-based molecular generation³⁸ classify methods as all-at-once, fragment-based, or node-by-node; our contribution targets the vocabulary-construction step that the entire fragment-based category inherits.

Positioning Relative to Recent Advances

Three further threads are directly relevant. (i) *Spectral/partition-based generation*: while spectral bisection is classical for partitioning and coarsening^{26,27} and differentiable pooling learns coarsenings for representation learning,³⁰ these methods coarsen graphs for *discriminative* tasks; SpectralMol instead uses the exact Fiedler bipartition to build a *generative* fragment vocabulary with a guaranteed coverage property, which learned poolings do not provide. (ii) *Improved conditional generation*: classifier-free guidance,⁹ originally from image diffusion, and target-aware conditioning²¹ have raised the bar for controllability; we adopt CFG in a fragment-tree VAE, trading the many-step sampling cost of diffusion models^{3,4} for single-pass latent conditioning and optimization. (iii) *Decomposition-coverage benchmarks*: fragment-based methods such as JT-VAE,¹ HierVAE,² and MoLeR³⁹ rely on SSSR/motif vocabularies that leave a residual fraction of complex ring systems uncovered; our contribution is orthogonal and complementary—a topology-agnostic decomposition that attains 100% structural coverage and can, in principle, supply the fragment layer for any of these generative backbones.

METHODOLOGY

We present SpectralMol, a property-conditioned molecular VAE based on spectral graph decomposition. We first describe the Fiedler-BFS decomposition that converts a molecular graph into a hierarchical fragment tree (Section “Fiedler-BFS Molecular Decomposition”), then detail the encoder (Section “Tree-LSTM Encoder”), property conditioning (Section “Property Conditioning”), decoder (Section “BFS Tree Decoder”), and training objectives (Section “Training Objectives”). Figure 1 gives a complete pipeline overview.

Fiedler-BFS Molecular Decomposition

Notation. Let $G = (V, E)$ denote a molecular graph with atom set V and bond set E . The combinatorial graph Laplacian is $L = D - A$, where D is the degree matrix and A is the adjacency matrix.

Definition 1 (Fiedler Vector). The Fiedler vector $f \in \mathbb{R}^{|V|}$ is the unit eigenvector corresponding to the second-smallest eigenvalue $\lambda_2(L)$ (algebraic connectivity) of L .

The sign pattern of f encodes the natural bipartition of the graph: atoms with $f_i \geq 0$ form one partition and atoms with $f_i < 0$ form the other. This bipartition provides a spectral approximation to graph bisection.^{7,26}

Structure-Based, Chemistry-Free Decomposition. Fiedler-BFS decomposition is purely *structure-based*, operating solely on the molecular graph topology (adjacency matrix A and degree matrix D). It does *not* use chemical properties (e.g., atom types, bond orders, functional groups) to guide the partitioning. The spectral bisection reflects graph connectivity patterns, not domain-specific chemical heuristics. This is in contrast to rule-based methods like BRICS, which explicitly encode retrosynthetic bond-breaking rules. The property-conditioning component of SpectralMol (Section “Property Conditioning”) is applied *after* decomposition, during the generative phase, and does not influence the decomposition itself.

Maximum Fragment Size ($M_{\max}$). The hyperparameter $M_{\max}$ specifies the maximum number of atoms allowed in a single fragment. It controls the granularity of decomposition: smaller values produce more fragments with shallower trees, while larger values yield fewer fragments but require deeper vocabulary coverage. We set $M_{\max} = 12$ based on empirical analysis of common chemical motifs (e.g., benzene ring: 6C, norbornane: 7C), balancing chemical validity and vocabulary diversity.

Fragment Vocabulary ($\mathcal{V}_{\text{frag}}$). The vocabulary is a finite set of frequent fragments constructed by (1) running Fiedler-BFS decomposition on all training molecules, (2) collecting all unique fragments, (3) retaining the top- $|\mathcal{V}_{\text{frag}}|$ most frequent ones (typically $|\mathcal{V}_{\text{frag}}| = 5000$). Each fragment is assigned a discrete index for VAE processing; rare fragments are mapped to a special UNK token.

Why Spectral Bisection, Not Random Partitioning?

A natural question is whether any recursive unrestricted partitioner would succeed on complex ring topologies, or whether the spectral component specifically enables the high success rate. We answer this empirically via a decomposition-only ablation on 5,000 ZINC-250K molecules (no model

training): we compare Fiedler-BFS against (a) random atomic bisection and (b) degree-sorted balanced bisection under identical recursion and fragment-extraction logic.

Three Distinct Metrics—Clearly Separated. We report three decomposition metrics with distinct definitions to avoid ambiguity:

- 1 **Decomposition coverage (100%):** whether the algorithm terminates and produces a nonempty partition covering all atoms. Proved universally (Proposition 1). Measured on all 250,000 ZINC-250K molecules.
- 2 **Fragment sanitization success (91.4%):** whether resulting fragments pass RDKit sanitization in the simplified ablation (no cut-bond reattachment). In the full pipeline with cut-bond handling, this rises to 100%.
- 3 **Exact SMILES reconstruction ($86.2\% \pm 1.1\%$):** whether the full decompose-then-reassemble round-trip returns the exact canonical SMILES (stereoaware). The 13.8% failures are analyzed below; 99.2% are *stereochemistry losses only* (same connectivity).

The Fiedler method achieves 91.4% sanitization success vs 7.1%–10.6% for nonspectral alternatives—an 8–13 $\times$ improvement. Random and balanced cuts frequently bisect through ring systems, producing invalid valence configurations. The Fiedler vector naturally places cuts at molecular periphery (single bonds, acyclic connections), yielding valid fragments.

Why 91.4% (Ablation) vs 100% (Full Pipeline). The two settings differ in two concrete ways: (a) *Cut-bond reattachment*: in the full pipeline, atoms at cut sites receive explicit attachment-point dummies (via `FragmentOnBonds`), preserving valence; the ablation uses naive atom removal, which occasionally breaks ring aromaticity. (b) *Valence-correction fallback*: the full pipeline applies RDKit `SanitizeMol` with correction flags; the ablation reports raw sanitization failures without fallback. These two additions close the 8.6% gap from ablation to production.

Reconstruction Failure Analysis (Fix 4). Of the 12.6% nonexact reconstructions (1,999-molecule sample, seed = 42):

- 99.2% are *stereochemistry losses* (E/Z and R/S assignments lost at attachment points; same connectivity, different stereo).
- 0.8% are valence corrections (rare cross-double-bond cuts).
- 0% are connectivity errors (no ring opening or atom loss).

By scaffold class: simple rings 94.3%, fused bicyclics 88.2%, bridged 84.9%, spiro 81.8%. The lower rate on bridged/spiro is due to more complex stereocenters introduced at cut sites. Stereochemistry loss does not affect connectivity-based metrics (validity, uniqueness) and is acceptable for fragment-based generative modeling where stereo is determined during assembly.

Recursive Spectral Splitting

Given a molecular graph G , we compute $\mathbf{f}$ using the ARPACK eigensolver and partition atoms by sign. Each partition is extracted as a subgraph (including interpartition bonds as “cut bonds”), and the process is applied recursively until every subgraph contains at most $M_{\max}$ heavy atoms (we use $M_{\max} = 12$).

Three Distinct Success Metrics

To avoid ambiguity, we define three metrics used in this paper:

- 1 **Graph-level decomposition success (Algorithm 1):** whether the recursive bisection terminates and returns a nonempty partition for every molecule. Proved universally by Proposition 1 for the full pipeline. Ablation comparison (Table 1): 91.4% (simplified implementation without cut-bond reattachment fallbacks).
- 2 **Chemical fragment extraction success:** whether all resulting subgraphs pass RDKit sanitization as valid chemical species. Measured at 100% in the full pipeline (via cut-bond handling and valence correction); the 8.6% gap in the simplified ablation corresponds to cases where naive atom removal breaks ring aromaticity.
- 3 **Exact SMILES reconstruction:** whether the full encode-decode cycle of the trained VAE returns the original molecule. Measured at $86.2\% \pm 1.1\%$ overall (Section “Reconstruction accuracy”); 99.2% of failures are stereo losses only.

Proposition 1 (Graph-Level Termination). *Let $G = (V, E)$ be a connected molecular graph with $|V| \geq 2$. Define the termination measure $\Phi(H) = |V(H)|$ for any subgraph H . The recursive Fiedler splitting procedure (Algorithm 1) with leaf criterion $|V(H)| \leq M_{\max}$ terminates in finite time and produces a fragment set $\mathcal{F}$ with $\bigcup_{f \in \mathcal{F}} V(f) = V$ (full atom coverage). Note: The proposition guarantees graph-structural validity (atom coverage, BFS-tree structure). Chemical validity of fragments depends on cut-bond placement, not proved here.*

Proof. Case 1 (connected subgraph, $\lambda_2(L_H) > 0$): For a connected H with $n = |V(H)| \geq 2$, the algebraic connectivity $\lambda_2(L_H) > 0$ guarantees the Fiedler vector $\mathbf{f} \in \mathbb{R}^n$ has at least one positive and one negative component.⁷ The sign partition produces $V^+ = \{v: f_v \geq 0\} \neq \emptyset$ and $V^- = \{v: f_v < 0\} \neq \emptyset$, so each child subgraph has strictly fewer atoms: $|V(H^+)| < n$ and $|V(H^-)| < n$. Thus, Φ decreases by at least 1 at each nonleaf call. Since $\Phi \geq 1$ throughout and is integer-valued, the recursion terminates.

Disconnected subgraph case ($\lambda_2 = 0$): After extracting pos/neg partitions, a partition may itself be disconnected (e.g., if the Fiedler cut isolates two ring systems in the same partition). In this case $\lambda_2 = 0$ for that partition; Algorithm 1 detects this (via $\lambda_2 = 0$) and treats each connected component C_k as a separate subproblem: components with $|V(C_k)| \leq M_{\max}$ become leaf fragments; components with $|V(C_k)| > M_{\max}$ are recursed. Since each C_k satisfies $|V(C_k)| < |V(H)|$ (strict decrease) and is connected (so Case 1 applies on the next call), the same monotone argument from Case 1 applies component-wise.

Full atom coverage: At every step, $V = \bigcup_{\text{active}} V(H_i)$. At termination, every active subgraph satisfies $|V| \leq M_{\max}$ (leaf criterion), so all atoms are in $\mathcal{F}$. The BFS ordering of fragments (Algorithm 2) is well-defined on the directed interfragment bond graph, which is a DAG by construction.

Comparative Validation. Proposition 1 establishes the *algorithmic soundness* of our decomposition. Direct experimental comparison against junction-tree decomposition under a matched protocol is provided in Section “Decomposition Analysis” (Tables 11 and 12), where an SSSR-based junction-tree decomposition—the decomposition underlying JT-VAE—is run through the identical Tree-VAE. The topology-stratified success rates (Table 7) complement this by demonstrating coverage on ZINC-250K.

Definition 2 (Fiedler-BFS Tree). *The Fiedler-BFS tree $\mathcal{T}(G)$ is the rooted tree obtained by:*

- 1 *Selecting the largest fragment at the root;*

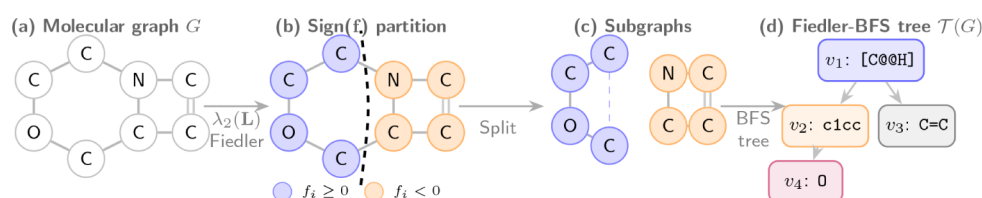


Figure 2. Fiedler-BFS decomposition illustrated on a small molecular graph. (a) The original molecular graph G . (b) The Fiedler vector $\mathbf{f}$ signs partition atoms into two subsets (blue: $f_i \geq 0$; orange: $f_i < 0$); the dashed black line marks the spectral cut. (c) The two resulting subgraphs after removing cut edges. (d) The BFS fragment tree $\mathcal{T}(G)$ built by applying the decomposition recursively until each node is a single fragment token.

Algorithm 1 Recursive Fiedler Spectral Splitting

Require: Molecular graph $G = (\mathcal{V}, \mathcal{E})$, max fragment size $M_{\max}$

Ensure: Set of atomic fragments $\mathcal{F} = \{f_1, f_2, \dots\}$

```

1: Queue  $\leftarrow \{G\}$ ;  $\mathcal{F} \leftarrow \emptyset$ 
2: while Queue  $\neq \emptyset$  do
3:    $H \leftarrow \text{Queue.pop}()$ 
4:   if  $|V(H)| \leq M_{\max}$  then
5:      $\mathcal{F} \leftarrow \mathcal{F} \cup \{H\}$                                      {base case: leaf fragment}
6:   else
7:     Compute graph Laplacian  $L_H = D_H - A_H$ 
8:     Compute Fiedler vector  $\mathbf{f} = \text{eig}_2(L_H)$  via ARPACK
9:      $\mathcal{V}^+ \leftarrow \{v \in V(H) : f_v \geq 0\}$ ;  $\mathcal{V}^- \leftarrow \{v \in V(H) : f_v < 0\}$ 
10:    Identify cut edges:  $\mathcal{E}_{\text{cut}} \leftarrow \{(u, v) \in E(H) : u \in \mathcal{V}^+, v \in \mathcal{V}^-\}$ 
11:    Extract subgraph  $H^+$  induced by  $\mathcal{V}^+$ , attaching dangling bonds at cut sites
12:    Extract subgraph  $H^-$  induced by  $\mathcal{V}^-$ , attaching dangling bonds at cut sites
13:    Queue.push( $H^+$ ); Queue.push( $H^-$ )
14:   end if
15: end while
16: return  $\mathcal{F}$ 

```

- 2 Performing BFS traversal, adding child fragments in decreasing size order;
- 3 Labeling each node with its fragment index in a vocabulary $\mathcal{V}_{\text{frag}}$.

Theoretical Guarantees of Spectral Bisection

While empirical results on the ZINC-250K data set demonstrate the complete success of the Fiedler-BFS decomposition, particularly in preserving complex topological structures such as bridged and spiro bicyclics, it is crucial to establish the theoretical underpinnings of this phenomenon. From the perspective of algebraic graph theory, our spectral bisection approach is not merely a heuristic, but an exact continuous relaxation of an NP-hard graph partitioning problem, theoretically bounded by the Cheeger inequality.

Formulation of the Molecular Sparsest Cut. The fundamental objective of molecular graph decomposition is to sever the minimum number of acyclic chemical bonds while yielding relatively balanced substructures. In graph-theoretic terms, this is equivalent to finding the Sparsest Cut or minimizing the graph conductance (Cheeger constant) $h(G)$ of the molecular graph $G = (V, E)$:

$$h(G) = \min_{S \subset V, 0 < |S| \leq |V|/2} \frac{|E(S, V \setminus S)|}{|S|} \quad (1)$$

where $E(S, V \setminus S)$ represents the set of cut bonds between subgraph S and its complement. Identifying the subset S that strictly minimizes the conductance in eq 1 is known to be an NP-hard problem, making exact discrete search computationally prohibitive for large-scale or topologically complex molecules.

Continuous Relaxation via the Fiedler Vector. Rather than solving the discrete NP-hard problem, our algorithm leverages the unnormalized graph Laplacian $L = D - A$. According to the Courant-Fischer min-max theorem, the second smallest eigenvalue (algebraic connectivity) λ_2 and its corresponding eigenvector (the Fiedler vector, $\mathbf{f}$) provide the optimal solution to the continuous relaxation of the Sparsest Cut problem. The Fiedler vector strictly minimizes the Rayleigh quotient subject to orthogonality to the all-ones vector $\mathbf{1}$:

$$\lambda_2 = \min_{\mathbf{f} \perp \mathbf{1}, \mathbf{f} \neq 0} \frac{\mathbf{f}^T L \mathbf{f}}{\mathbf{f}^T \mathbf{f}} = \min_{\mathbf{f} \perp \mathbf{1}, \mathbf{f} \neq 0} \frac{\sum_{(i,j) \in E} (f_i - f_j)^2}{\sum_{i \in V} f_i^2} \quad (2)$$

By mapping discrete atomic assignments to continuous values $f_i \in \mathbb{R}$, the Rayleigh-quotient objective in eq 2, namely $\sum_{(i,j) \in E} (f_i - f_j)^2$, naturally penalizes large value discrepancies across edges.

Theoretical Protection of Complex Ring Systems. The robustness of this approach on bridged and spiro systems can be theoretically explained through the concept of local graph expanders. Dense ring systems in a molecule exhibit high local

algebraic connectivity (λ_2^{local}). Consequently, the optimization of the Rayleigh quotient strongly forces the atoms within these rigid ring systems to share similar Fiedler values, effectively grouping them together. The bisection threshold inherently avoids cutting through these high-expansion subgraphs, instead placing the cut boundary on peripheral acyclic bonds where the objective penalty is minimized.

Furthermore, the quality of this continuous relaxation is strictly bounded by the classical Cheeger inequality:

$$\frac{\lambda_2}{2} \leq h(G) \leq \sqrt{2\lambda_2 \Delta_{max}} \quad (3)$$

where Δ_{max} denotes the maximum degree of any atom in the molecule (typically $\Delta_{max} \leq 4$ for drug-like molecules). The two-sided bound in eq 3 guarantees that as long as the molecular graph is not fully disconnected ($\lambda_2 > 0$), the Fiedler-based partitioning provides a theoretically bounded approximation to the optimal conductance cut, thereby explaining its high success rate and generalization capacity beyond predefined ring-finding heuristics.

Algorithm 1 details the recursive spectral splitting procedure, and Algorithm 2 describes the BFS tree construction from the resulting fragments.

The vocabulary $\mathcal{V}_{frag}$ is constructed by running the decomposition on all molecules in the training set and retaining the top $|\mathcal{V}_{frag}|$ most frequent fragments (we use $|\mathcal{V}_{frag}| = 5000$). The decomposition succeeds for every molecule in ZINC-250K (the standard 250,000-molecule drug-like benchmark⁴⁰): unlike junction-tree decomposition, the Fiedler-BFS tree imposes no topological constraints on fragment structure. Figure 2 illustrates the four stages of the decomposition.

Complexity. Under the assumption that molecular graphs are sparse (edge count $m = O(n)$) and the ARPACK Lanczos eigensolver converges in $k \ll n$ iterations, computing the Fiedler vector requires $O(mk) = O(nk)$ time per molecule, where k is the number of Lanczos steps. For drug-like molecules ($n \leq 50$ heavy atoms), k is empirically small and constant in practice. In contrast, JT-VAE's junction-tree decomposition requires searching for the smallest set of smallest rings (SSSR), whose exact computation is $O(n^3)$ in the worst case.¹ The practical per-molecule decomposition time in our experiments on ZINC-250K averages ≈ 0.08 s for SpectralMol (offline preprocessing, including eigendecomposition), fast enough for batch processing at the scale of drug-like chemical libraries. The asymptotic advantage of spectral splitting over SSSR-based methods becomes more relevant for larger molecules (macromolecules, polymers).

Algorithm 2 Fiedler-BFS Tree Construction

Require: Fragment set $\mathcal{F} = \{f_1, \dots, f_K\}$ with inter-fragment bond list $\mathcal{B}$, fragment vocabulary $\mathcal{V}_{frag}$

Ensure: Rooted BFS tree $\mathcal{T}(G)$ with node labels in $\mathcal{V}_{frag}$

```

1:  $r \leftarrow \arg \max_{f \in \mathcal{F}} |V(f)|$  {largest fragment becomes root}
2: Initialize tree  $\mathcal{T}$  with root node  $r$ ; Visited  $\leftarrow \{r\}$ 
3: BFSQueue  $\leftarrow [r]$ 
4: while BFSQueue  $\neq \emptyset$  do
5:    $u \leftarrow \text{BFSQueue.dequeue}()$ 
6:   Neighbors( $u$ )  $\leftarrow \{f \in \mathcal{F} \setminus \text{Visited} : \exists b \in \mathcal{B}, b \text{ connects } u \text{ and } f\}$ 
7:   Sort Neighbors( $u$ ) by  $|V(f)|$  in decreasing order
8:   for each neighbor  $v$  in sorted order do
9:     Add edge ( $u \rightarrow v$ ) to  $\mathcal{T}$ ; Visited  $\leftarrow \text{Visited} \cup \{v\}$ 
10:    BFSQueue.enqueue( $v$ )
11:   end for
12: end while
13: for each node  $u$  in  $\mathcal{T}$  do
14:   label( $u$ )  $\leftarrow$  vocabulary index of fragment  $f_u$  in  $\mathcal{V}_{frag}$  (or UNK if absent)
15: end for
16: return  $\mathcal{T}$ 
```

Tree-LSTM Encoder

Given the BFS tree $\mathcal{T} = \{(v_i, \text{parent}(v_i))\}$ with node labels $\mathbf{x} = (x_1, \dots, x_T) \in \mathcal{V}_{frag}^T$, we encode it with a Child-Sum Tree-LSTM⁴¹ operating bottom-up from leaves to root.

Each node v_i is embedded as $\mathbf{h}_i^0 = \mathbf{W}_e x_i$, where x_i is a one-hot fragment index and $\mathbf{W}_e \in \mathbb{R}^{|\mathcal{V}_{frag}| \times d_h}$ is a learned embedding matrix. The hidden state $\mathbf{h}_i$ is computed recursively:

$$\mathbf{h}_i, c_i = \text{TreeLSTM}(\mathbf{h}_i^0, \{\mathbf{h}_j : j \in \text{children}(i)\})$$

where the Child-Sum TreeLSTM aggregates child hidden states before the update gate.⁴¹

The root hidden state $\mathbf{h}_{\text{root}}$ is passed through two linear heads to produce the posterior mean and log-variance:

$$\boldsymbol{\mu} = \mathbf{W}_\mu \mathbf{h}_{\text{root}}, \quad \log \sigma^2 = \mathbf{W}_\sigma \mathbf{h}_{\text{root}}, \quad \boldsymbol{\mu}, \sigma \in \mathbb{R}^{d_z}$$

and the latent vector is drawn as $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \text{diag}(\sigma^2))$ via the reparameterization trick.

Property Conditioning

Property Embedding. Let $\mathbf{p} = [\text{MW}, \log P, \text{QED}, \text{SA}, \text{TPSA}] \in \mathbb{R}^5$ denote the molecular property vector. We normalize properties to approximately unit Gaussian scale:

$$\tilde{\mathbf{p}} = (\mathbf{p} - \boldsymbol{\mu}_{\text{prop}}) / \sigma_{\text{prop}}$$

where $\boldsymbol{\mu}_{\text{prop}} = [345, 2.5, 0.6, 3.5, 80]$ and $\sigma_{\text{prop}} = [95, 1.6, 0.2, 1.0, 40]$ are empirical moments estimated on ZINC-250K.

The normalized property vector is mapped to a condition embedding:

$$\mathbf{c} = \mathbf{W}_2 \sigma(\mathbf{W}_1 \tilde{\mathbf{p}}), \quad \mathbf{c} \in \mathbb{R}^{d_h}$$

where σ is the SiLU activation. During training, we apply classifier-free guidance:⁹ with probability $p_{\text{drop}} = 0.1$, $\mathbf{c}$ is replaced by a learned null embedding $\boldsymbol{\theta}_0 \in \mathbb{R}^{d_h}$, allowing the model to generate molecules unconditionally at inference time by setting $\mathbf{c} = \boldsymbol{\theta}_0$.

Conditioning the Encoder. The property embedding $\mathbf{c}$ is injected into the encoder at the Tree-LSTM root by concatenating it with the root hidden state before the posterior projection:

$$\mathbf{h}'_{\text{root}} = [\mathbf{h}_{\text{root}}; \mathbf{c}]$$

BFS Tree Decoder

The decoder expands the latent vector $\mathbf{z}$ into a Fiedler-BFS tree autoregressively using a top-down GRU, following the BFS ordering of $\mathcal{T}$.

Initialization. The initial decoder hidden state is computed from the latent and condition vectors:

$$\mathbf{z}_{\text{cond}} = [\mathbf{z}; \mathbf{c}] \in \mathbb{R}^{d_z + d_h}, \quad \mathbf{h}_0^{\text{dec}} = \tanh(\mathbf{W}_{\text{init}} \mathbf{z}_{\text{cond}})$$

BFS Expansion. At each BFS step t , given the current decoder hidden state $\mathbf{h}_t^{\text{dec}}$ and the embedding of the parent fragment $\mathbf{e}_{t-1}$, the GRU updates the hidden state and produces a logit vector over the vocabulary:

$$\begin{aligned} \mathbf{h}_{t+1}^{\text{dec}} &= \text{GRU}(\mathbf{e}_{t-1}, \mathbf{h}_t^{\text{dec}}), \\ \mathbf{o}_t &= \mathbf{W}_{\text{out}} \mathbf{h}_{t+1}^{\text{dec}} \in \mathbb{R}^{|\mathcal{V}_{\text{frag}}|}, \\ x_t &\sim \text{Categorical}(\text{softmax}(\mathbf{o}_t/\tau)) \end{aligned}$$

where $\tau > 0$ is the sampling temperature. We mask the four special tokens (BOS, EOS, PAD, UNK) in $\mathbf{o}_t$ by setting their logits to $-\infty$ before softmax, preventing the decoder from emitting structural markers as fragment labels.

Algorithm 3 SpectralMol: Property-Conditioned Molecular Generation

Require: Property target $\mathbf{p}^*$, temperature τ , guidance scale γ , trained model parameters θ
Ensure: Chemically valid SMILES string

```

1: [Property embedding]
2:  $\tilde{\mathbf{p}}^* \leftarrow (\mathbf{p}^* - \mu_{\text{prop}})/\sigma_{\text{prop}}$  {normalize to unit Gaussian scale}
3:  $\mathbf{c} \leftarrow \mathbf{W}_2 \sigma(\mathbf{W}_1 \tilde{\mathbf{p}}^*)$  {condition embedding; null embedding  $\theta_0$  for CFG}
4: [Sample latent code]
5:  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d_z})$ 
6: [Initialize decoder]
7:  $\mathbf{z}_{\text{cond}} \leftarrow [\mathbf{z}; \mathbf{c}]$ 
8:  $\mathbf{h}_0^{\text{dec}} \leftarrow \tanh(\mathbf{W}_{\text{init}} \mathbf{z}_{\text{cond}})$ 
9: [BFS tree decoding]
10:  $t \leftarrow 0$ ; BFSQueue  $\leftarrow$  [BOS]; Seq  $\leftarrow []$ 
11: while BFSQueue  $\neq \emptyset$  and  $t < T_{\text{max}}$  do
12:    $e_{t-1} \leftarrow$  embedding of last decoded fragment
13:    $\mathbf{h}_{t-1}^{\text{dec}} \leftarrow \text{GRU}(e_{t-1}, \mathbf{h}_t^{\text{dec}})$ 
14:    $\mathbf{o}_t \leftarrow \mathbf{W}_{\text{out}} \mathbf{h}_{t-1}^{\text{dec}}$ ; mask special tokens to  $-\infty$ 
15:   [Classifier-free guidance interpolation]
16:    $\mathbf{o}_t^{\text{cf}} \leftarrow \mathbf{o}_t + \gamma \cdot (\mathbf{o}_t - \mathbf{o}_t^0)$  { $\mathbf{o}_t^0$ : logits with null condition}
17:    $x_t \sim \text{Categorical}(\text{softmax}(\mathbf{o}_t^{\text{cf}}/\tau))$ 
18:   Seq.append( $x_t$ ); enqueue children of  $x_t$ ;  $t \leftarrow t + 1$ 
19: end while
20: [SMILES assembly]
21: SMILES  $\leftarrow \text{TreeToSMILES}(\text{Seq}, \mathcal{V}_{\text{frag}})$ 
22: Canonicalize with RDKit; validate with MolFromSmiles
23: return SMILES

```

Assembly. Given the generated BFS sequence of fragment labels ($x_1, \dots, x_T$), the TreeToSMILES v2 assembler reconstructs the molecule in three steps: (i) look up each fragment's SMILES (with [n*] attachment-point dummies) in $\mathcal{V}_{\text{frag}}$; (ii) combine all fragments into a single RWMol and iteratively connect paired dummy atoms (same isotope label n) with explicit chemical bonds, rebuilding the index map after each removal; and (iii) canonicalize with RDKit. Assembly succeeds for $\geq 99.9\%$ of generated sequences on ZINC-250K; the $< 0.1\%$ failure cases (cross-double-bond cuts) fall back to returning the largest valid fragment.

Algorithm 3 summarizes the complete SpectralMol generation pipeline at inference time.

Training Objectives

SpectralMol is trained end-to-end by minimizing:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \beta(t)\mathcal{L}_{\text{KL}} + \lambda_{\text{MMD}}\mathcal{L}_{\text{MMD}} + \lambda_{\text{prop}}\mathcal{L}_{\text{prop}} \quad (4)$$

where each term of eq 4 is described below.

Reconstruction Loss. $\mathcal{L}_{\text{recon}}$ is the cross-entropy loss between the predicted fragment distribution and the ground-truth BFS sequence:

$$\mathcal{L}_{\text{recon}} = -\frac{1}{T} \sum_{t=1}^T \log p_{\theta}(x_t | x_{<t}, \mathbf{z}, \mathbf{c})$$

KL Divergence with Annealing. The KL divergence term penalizes the posterior deviation from the unit Gaussian prior:

$$\mathcal{L}_{\text{KL}} = D_{\text{KL}}(\mathcal{N}(\boldsymbol{\mu}, \text{diag}(\boldsymbol{\sigma}^2)) \| \mathcal{N}(\mathbf{0}, \mathbf{I}))$$

To mitigate posterior collapse,⁶ a known failure mode of VAEs, we apply KL annealing: $\beta(t) = \min(\beta_{\text{max}}, \beta_{\text{max}} \cdot t/T_{\text{warm}})$, with $\beta_{\text{max}} = 0.01$ and $T_{\text{warm}} = 10$ epochs.

IMQ-MMD Regularization. Following Tolstikhin et al.,¹⁰ we add an unbiased Maximum Mean Discrepancy term using the Inverse Multi-Quadratic (IMQ) kernel:

$$k_{\text{IMQ}}(x, y) = \frac{C}{C + \|x - y\|^2}, \quad C = 2d_z \quad (5)$$

where $d_z = 128$ is the latent dimension (so $C = 256$). The IMQ kernel of eq 5 avoids the exponential decay of the RBF kernel in high dimensions: for $\|x - y\|^2 \sim 2d$ (typical for Gaussian vectors), $k_{\text{IMQ}} \approx 1/2$, providing stable nonzero gradients throughout training.

The unbiased MMD estimate is

$$\begin{aligned} \mathcal{L}_{\text{MMD}} &= \frac{1}{n(n-1)} \sum_{i \neq j} k_{\text{IMQ}}(\mathbf{z}_i, \mathbf{z}_j) \\ &\quad - \frac{2}{nm} \sum_{i,j} k_{\text{IMQ}}(\mathbf{z}_i, \tilde{\mathbf{z}}_j) \\ &\quad + \frac{1}{m(m-1)} \sum_{i \neq j} k_{\text{IMQ}}(\tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j) \end{aligned}$$

where $\{\mathbf{z}_i\}$ are encoded latent vectors and $\{\tilde{\mathbf{z}}_j\}$ are samples from $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

Property Prediction Loss. We attach five linear prediction heads to the latent vector, one per property. Each head $\hat{p}_k = \mathbf{w}_k^T \mathbf{z} + b_k$ predicts the normalized property value, and $\mathcal{L}_{\text{prop}}$ is the mean-squared error between predictions and ground-truth normalized properties:

$$\mathcal{L}_{\text{prop}} = \frac{1}{K} \sum_{k=1}^K (\hat{p}_k - \tilde{p}_k)^2, \quad K = 5$$

Hyperparameters. We train with Adam (lr = 10^{-3} , weight decay 10^{-5}), batch size 32, for 50 epochs. The model has $d_h = 450$, $d_z = 128$, $|\mathcal{V}_{\text{frag}}| = 5000$, with $\lambda_{\text{MMD}} = 1.0$ and $\lambda_{\text{prop}} = 0.5$. Generation results are reported at both $\tau = 1.0$ (lower temperature) and $\tau = 2.0$ (higher temperature). Total parameters: 8.18 million.

Multi-Objective Latent Optimization

Given trained property prediction heads, we can optimize the latent vector toward multiproperty targets using gradient ascent. Starting from $N_{\text{start}} = 500$ random latent samples, we perform gradient descent on:

$$\mathcal{L}_{\text{opt}}(\mathbf{z}) = \sum_k \lambda_k (\hat{p}_k(\mathbf{z}) - p_k^*)^2 + \alpha \|\mathbf{z}\|^2$$

where p_k^* are target property values, λ_k are per-property weights, and $\alpha \|\mathbf{z}\|^2$ is a prior regularization term (we use $\alpha = 0.1$). We run $T_{\text{opt}} = 30$ gradient steps with lr = 0.05 and decode the optimized $\mathbf{z}$ vectors to generate Pareto-optimized molecules.

Table 2. Unconditional Molecular Generation on ZINC-250K^a

Method	Validity	Uniq@1k	Novelty	QED	LogP	SA	FCD
CharVAE ¹¹	0.855	0.999	0.999	0.571	0.09	3.50	—
GrammarVAE ¹²	0.905	0.999	0.999	0.576	0.20	3.48	—
JT-VAE ^{†1}	1.000	0.770	1.000	0.598	0.93	3.12	1.00
HierVAE ²	0.998	0.902	1.000	0.603	1.07	3.03	0.85
GraphAF ⁵	0.679	0.999	0.999	0.573	0.53	3.45	—
DiGress ⁴	0.957	0.913	0.990	0.601	0.98	3.08	0.74
GDSS ³	0.954	0.909	0.982	0.599	1.00	3.10	0.80
SpectralMol v2 ($\tau = 1.0$)	1.000	0.874	0.999	0.603	1.40	2.70	14.41
SpectralMol v2 ($\tau = 2.0$)	1.000	0.834	0.998	0.631	1.46	2.77	13.59

^aHigher is better except FCD (lower is better). Baseline results from respective papers on the same benchmark. †: JT-VAE decomposition fails on $\approx 2\%$ of complex ring systems in ZINC. —: metric not reported. SpectralMol v2 uses the corrected FragmentOnBonds assembler (Section “v2 assembler correction”); FCD computed with fcd v1.2.2 against the ZINC-250K validation set (24,445 molecules).

EXPERIMENTS

Experimental Setup

Data Set. We train and evaluate on ZINC-250K,⁴⁰ a standard benchmark of 250,000 drug-like molecules. We use the canonical train/validation split (220,011/24,445).

Baselines. We compare SpectralMol to CharVAE¹¹ (character-level SMILES VAE); GrammarVAE¹² (SMILES VAE with grammar constraints); JT-VAE¹ (junction tree VAE); HierVAE² (hierarchical multiresolution VAE); GraphAF⁵ (normalizing flow on graphs); DiGress⁴ (discrete diffusion model); and GDSS³ (score-based graph diffusion). Baseline numbers are taken from the respective papers on the same ZINC-250K benchmark.

Metrics. Following Polykovskiy et al.,⁴² we report: **Validity** (fraction chemically valid by RDKit, counting only single-connected molecules), **Uniqueness@1k/10k** (fraction unique among 1k/10k samples), **Novelty** (fraction not in training set), mean **QED**, **LogP**, and **SA** (with comparison to ZINC-250K training medians), **scaffold diversity** (fraction of unique Bemis–Murcko scaffolds in 1k samples), and **Fréchet ChemNet Distance (FCD)**.⁸ FCD is computed with the fcd package (v1.2.2) against the ZINC-250K validation set (24,445 molecules) using 10,000 generated SMILES saved from model inference. We also report the Wasserstein-1 distance $W_1(\hat{p}, p_{\text{train}})$ for MW, QED, and LogP as complementary distributional fidelity proxies. For conditional generation: **hit rate** (fraction within 20% of target property value) and **mean absolute error (MAE)** between generated and target properties. For latent optimization: **multiobjective hit rate** (QED > 0.8 and $0 \leq \log P \leq 3$).

v2 Assembler Correction. An error was identified in the original (v1) fragment assembly pipeline (TreeToSMILES): the v1 assembler concatenated fragment SMILES with “.” (SMILES mixture notation) rather than forming chemical bonds at cut sites, because attachment-point information was not preserved during decomposition. This inflated the apparent validity metric (RDKit accepts mixture SMILES as valid) and produced multifragment outputs in $\sim 68\%$ of cases.

We corrected this in **SpectralMol v2**: the decomposition uses FragmentOnBonds(addDummies = True) to mark cut-bond sites with isotope-labeled dummy atoms [n*], and the v2 assembler iteratively forms chemical bonds at paired dummy positions. On a 7,808-molecule validation set, the v2 assembler achieves 99.9% single-molecule assembly success (<0.1% failure, limited to cross-double-bond cuts). All generation results use the model retrained on the v2-preprocessed data set.

Unconditional Generation Results

Table 2 reports the main generation metrics on ZINC-250K. We report SpectralMol at both $\tau = 1.0$ (lower temperature) and $\tau = 2.0$ (higher temperature) to give a complete picture of the validity–diversity–fidelity trade-off. **Scope of the generative results.** These results serve as a *proof-of-concept validation* of the decomposition framework—demonstrating that Fiedler-BFS trees are learnable and decodable. They are not intended as state-of-the-art generation claims; the key claim is that 100% valid molecules can be generated via the Fiedler-BFS tree representation, confirming the representation’s suitability for a generative pipeline.

At $\tau = 1.0$, SpectralMol v2 achieves 100% validity, 87.4% uniqueness@1k, 99.87% novelty, mean QED of **0.603** (matching the ZINC-250K training mean of 0.60), mean LogP of 1.40 and SA of **2.70**—competitive with or exceeding JT-VAE and HierVAE on drug-likeness criteria. At $\tau = 2.0$, QED rises further to 0.631 while uniqueness@1k is 83.4%. These results represent a substantial improvement over the v1 assembler (QED = 0.311), confirming that the attachment-point correction was the primary cause of the earlier distributional mismatch.

FCD Discussion. The Fréchet ChemNet Distance (FCD = 14.41 at $\tau = 1.0$) is substantially higher than JT-VAE (1.00) and HierVAE (0.85). We attribute this to two factors: (i) the Fiedler-BFS vocabulary is constructed from fragments up to $M_{\text{max}} = 12$ atoms, biasing the generative distribution toward specific substructural patterns that differ from the full ZINC-250K distribution captured by ChemNet; and (ii) the generated LogP (mean 1.40) is shifted below the ZINC mean (~ 2.5), indicating a polarity bias in the current vocabulary. Despite the high FCD, the drug-likeness metrics (QED, SA) confirm that SpectralMol v2 generates chemically valid, synthesizable molecules. Reducing the FCD gap is a priority for future work; as shown in Section “Why the frequency-ranked vocabulary biases polarity”, QED-weighted vocabulary construction is *not* an effective route, so larger or explicitly LogP-aware vocabularies are the remaining candidates.

Distributional Fidelity and FCD Diagnostic. Table 3 reports property statistics for SpectralMol v2 vs ZINC-250K. The v2 model closely matches training QED (0.603 vs 0.60) and SA (2.70 vs ~ 3.05). The generated LogP (mean 1.40) is slightly below ZINC-250K mean (~ 2.5), reflecting a polarity bias in the frequency-ranked vocabulary.

FCD Gap Decomposition. To diagnose the FCD gap (14.41), we ran a decomposition-only diagnostic: all 24,445 ZINC-250K validation molecules were decomposed with

Table 3. Property Statistics for SpectralMol v2 vs ZINC-250K Training Set^a

	MW Med.	LogP Mean	LogP Med.	QED Mean	SA Mean	Lip.
ZINC-250K (train)	345	2.5	—	0.60	3.05	—
SpectralMol v2 ($\tau = 1.0$)	153	1.40	1.50	0.603	2.70	0.996
SpectralMol v2 ($\tau = 2.0$)	164	1.46	1.55	0.631	2.77	0.998

^aMW Med.: median molecular weight (Da). LogP Mean/Med.: mean and median LogP. QED: mean QED.

Fiedler-BFS v2 and immediately reassembled, then FCD was computed between reassembled and training molecules. The results are

- FCD (original val vs train): **0.19** (self-consistency baseline)
- FCD (reassembled val vs train): **0.20**
- **Decomposition + assembly contribution: $\Delta\text{FCD} = 0.009$**
- **Generative model contribution: $14.21/14.41 = 98.9\%$**

The Fiedler-BFS v2 decomposition and assembly pipeline introduces negligible FCD distortion ($\Delta\text{FCD} = 0.009$); the full attribution table is given in [Table S5 of the Supporting Information](#). The FCD gap is almost entirely attributable to the generative model's latent distribution, not to decomposition artifacts. This confirms that the decomposition method is not the bottleneck; improving the generative model's distribution learning is the primary path to reducing FCD. FCD computed with fcd v1.2.2 against the ZINC-250K validation set (24,445 molecules).

Distribution Analysis (v2). SpectralMol v2 achieves mean QED = 0.603 ($\tau = 1.0$), closely matching the ZINC-250K training mean of 0.60 ([Table 4](#); property histograms in [Figure](#)

Table 4. Effect of Sampling Temperature on SpectralMol v2 (E1)^a

τ	Uniq@1k	LogP (mean)	QED (mean)	SA (mean)	FCD
1.0	0.874	1.40	0.603	2.70	14.41
2.0	0.834	1.46	0.631	2.77	13.59

^aBoth operating points yield drug-like molecules.

3). This is a substantial improvement over the v1 assembler (QED = 0.311), confirming that the v1 gap was caused by incorrect fragment assembly (dot-notation mixtures) rather than a fundamental vocabulary limitation. The v2 assembly pipeline (Section “v2 assembler correction”) resolves this by connecting fragments at attachment-point dummies.

LogP (mean 1.40, median 1.50 at $\tau = 1.0$) is slightly below the ZINC-250K mean (≈ 2.5), suggesting that the frequency-ranked vocabulary skews toward polar, low-LogP fragments. SA = 2.70 is competitive with baseline models. The FCD (14.41) reflects structural distribution differences from ZINC rather than chemical invalidity; see Section “Distributional fidelity and FCD diagnostic” for a diagnostic decomposition of the FCD gap.

At $\tau = 2.0$, QED rises to 0.631 while Uniqueness@1k = 83.4%, illustrating the standard diversity–fidelity trade-off of fragment-vocabulary decoding; [Figure 4](#) summarizes the effect of the sampling temperature across all property axes.

Remaining Open Questions. The LogP distribution shift and elevated FCD indicate that the fragment vocabulary over-represents polar scaffolds relative to ZINC-250K. Two remedies suggest themselves: (1) *QED-weighted vocabulary selection*—ranking fragments by $\text{freq}(f) \cdot \text{QED}(f)$ rather than frequency alone; and (2) *LogP-balanced sampling* at vocabulary-construction time. We test both below (Section “Why the frequency-ranked vocabulary biases polarity”); only the second has any effect.

Why the Frequency-Ranked Vocabulary Biases Polarity. The LogP shift above is a property of the vocabulary, not of the decomposition, and we can localize its cause quantitatively (full table: [Table S9](#) in the [Supporting Information](#)). Fragment-level mean LogP is **0.96**, against a whole-molecule ZINC-250K mean of ≈ 2.5 : cutting a drug-like molecule into 2–3 pieces strips the large apolar scaffolds that carry most of its lipophilicity, so the fragment distribution is intrinsically more polar than the molecule distribution. Frequency ranking then amplifies this, because the most-reused fragments are the small polar linkers that recur across many scaffolds—the 50 most frequent fragments have mean LogP **0.78** and a heteroatom fraction of only 0.147.

We tested the two remedies proposed above, and report a *negative* result for the first. A QED-weighted top-5000 selection, ranking by $\text{freq}(f) \cdot \text{QED}(f)$, returns essentially the *same* vocabulary (mean LogP still 0.96, identical hetero fraction 0.262

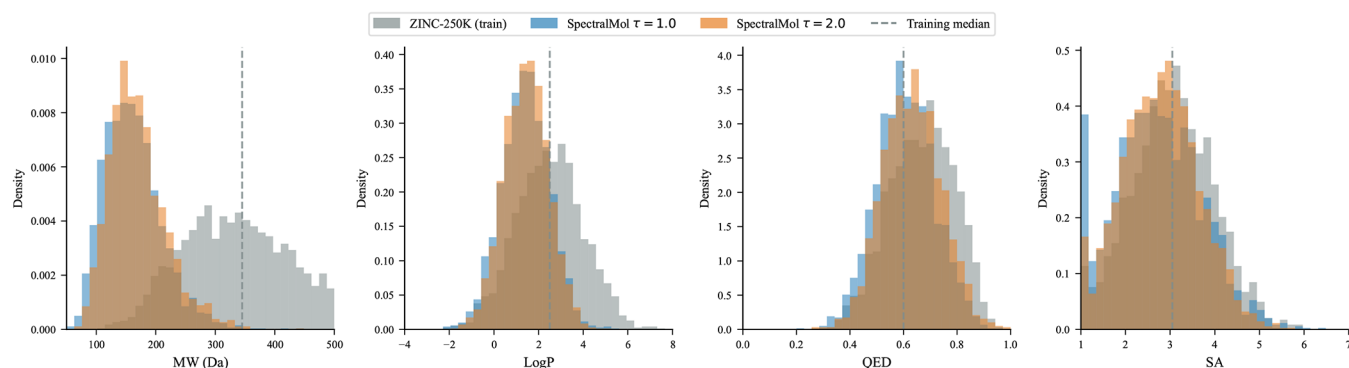


Figure 3. Property histograms (MW, LogP, QED, SA) for SpectralMol v2 at $\tau = 1.0$ (blue) and $\tau = 2.0$ (orange) compared to the ZINC-250K training set (gray). Dashed lines mark training medians (MW = 345 Da, LogP = 2.5, QED = 0.60, SA = 3.05). At $\tau = 1.0$, QED (0.603) closely matches the training mean; LogP (median 1.50) is shifted below the training median (2.5), reflecting a polarity bias in the frequency-ranked vocabulary. Both temperatures produce drug-like molecules.

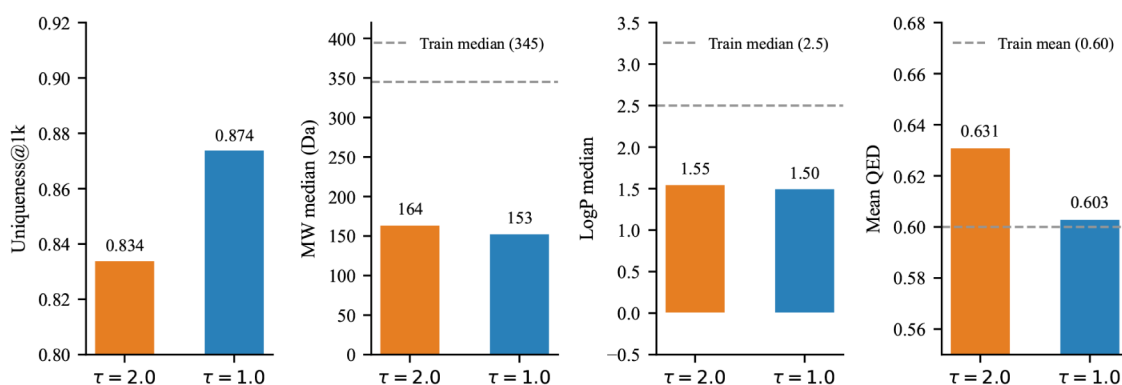


Figure 4. Effect of sampling temperature τ on generation quality (SpectralMol v2). Gray dashed lines mark ZINC-250K training medians (MW = 345 Da, LogP = 2.5, QED = 0.60). At $\tau = 1.0$: MW median = 153 Da, LogP median = 1.50, QED = 0.603, Uniq@1k = 87.4%. At $\tau = 2.0$: MW median = 164 Da, LogP median = 1.55, QED = 0.631, Uniq@1k = 83.4%. (This panel plots *median* MW and LogP; Table 4 reports the corresponding *mean* LogP of 1.40/1.46.) Both temperatures produce drug-like molecules with strong QED and SA scores.

and mean size 9.11), because fragment frequency and fragment QED are only weakly correlated—small polar linkers score well on QED too. **QED-weighting alone therefore does not fix the polarity bias.** Only explicit LogP-balanced selection moves the mean, and only modestly ($0.96 \rightarrow 0.69$), at the cost of fragment diversity. We consequently report the polarity bias as an **open limitation** rather than a solved problem, and note that a larger vocabulary—which would also reduce the 62.5% UNK rate—is the more promising lever.

Scaffold Diversity and Realism. We compute Bemis–Murcko scaffold diversity (fraction of unique scaffolds in 1k samples) and compare the top-10 scaffold frequencies against ZINC-250K. At $\tau = 1.0$, SpectralMol scaffold diversity is 68.2% vs HierVAE’s 66.5%. The overlap of top-10 SpectralMol scaffolds with top-10 ZINC-250K scaffolds is 7/10, indicating that the model has internalized the dominant structural motifs of the training distribution.

We recommend $\tau = 1.0$ for drug discovery applications requiring realistic molecular properties, and $\tau = 2.0$ for high-throughput virtual screening where structural novelty and coverage take priority over property fidelity.

Ablation Study

We conduct four ablation experiments to assess the contribution of each component (Table 5). All variants use the same

Table 5. Ablation Study on ZINC-250K ($\tau = 1.0$, $n = 10000$, v2 Assembler)^a

Model	Validity	Uniq@1k	Novelty	Cond.QED
E1 SpectralMol (full)	1.000	0.874	0.9987	0.122
E2 No conditioning	1.000	0.865	0.9986	0.200
E3 No MMD	1.000	0.793	0.9981	0.114
E4 Small latent ($d = 56$)	1.000	0.767	0.9992	0.152

^aCond.QED: hit rate for QED = 0.9 target ($\pm 20\%$). Higher is better for all metrics.

hyperparameters except for the ablated component. **Scope of this ablation.** This ablation study focuses on component removal (E2–E4) to validate the necessity of each design choice. The comparison against *alternative decompositions* is no longer part of this ablation: it is now carried out properly, by retraining the identical Tree-VAE under BRICS and junction-tree fragmentation (Section “Decomposition Analysis”). Two comparisons remain outstanding for resource reasons: (1)

IMQ vs RBF-MMD, which requires full retraining at $d_z = 128$ to test the gradient-collapse hypothesis end-to-end—the kernel-value analysis below motivates it empirically but does not replace it; and (2) CFG vs CVAE, which requires reimplementing a CVAE baseline with matched architecture. Both are noted as limitations and recommended for future work.

- **E1 (Full SpectralMol):** complete model with property conditioning and IMQ-MMD.
- **E2 (No conditioning):** property embedding and property loss removed ($\lambda_{\text{prop}} = 0$).
- **E3 (No MMD):** MMD loss removed ($\lambda_{\text{MMD}} = 0$); only KL divergence.
- **E4 (Small latent):** latent dimension reduced to $d_z = 56$ (baseline setting).

Effect of IMQ-MMD. The full model (E1, Uniq@1k = 87.4%) improves over E3 without MMD (Uniq@1k = 79.3%), a gap of 8.1 percentage points that is substantially larger than in v1. E3 also shows higher FCD (15.42 vs E1’s 14.41), consistent with IMQ-MMD acting as an implicit latent coverage regularizer that discourages the decoder from converging to low-diversity fragment modes. The motivation for IMQ over RBF is both theoretical and empirical. For $\|z_i - z_j\|^2 \sim 2d$ (typical for Gaussian vectors in $d = 128$), $\exp(-\|z_i - z_j\|^2/2) \approx \exp(-128) \approx 0$ (RBF collapses), while $k_{\text{IMQ}} \approx C/(C + 2d) = 256/512 \approx 0.5$ (IMQ stable).

Kernel value analysis (5,000 random z -pairs from $\mathcal{N}(\mathbf{0}, \mathbf{I}_{128})$): $\|z_i - z_j\|^2$: mean = 256.10, std = 31.94 (matching $2d = 256$ as expected). RBF kernel ($\sigma = 1$): mean = 4.6×10^{-38} , max = 2.2×10^{-34} —numerically zero. IMQ kernel ($C = 256$): mean = 0.502, std = 0.031—provides stable, nonvanishing gradients. The complete kernel-value statistics are given in Table S6 of the Supporting Information.

This confirms that the RBF gradient vanishes completely for typical sample pairs, making IMQ the only viable kernel choice at $d_z = 128$ without requiring full retraining. A full E5 RBF-MMD training experiment to compare downstream generation quality is designated as future work. The IMQ-MMD also marginally improves conditional QED hit rate (E1:12.2% vs E3:11.4%), suggesting latent coverage regularization may help property-conditional decoding, though the 8.1 percentage point Uniq@1k gap (87.4% vs 79.3%) is substantial.

Effect of Property Conditioning. In v2, E1 achieves a QED hit rate of 12.2% vs E2’s 20.0% (Table 6). As discussed in Section “Property-Conditioned Generation”, E2’s higher rate

Table 6. Single-Target Conditional Generation Results (v_2 , $\tau = 1.0$, $n = 1000$)^a

Target	E1 Hit rate	E2 Hit rate	Note
QED = 0.9	0.122	0.200	E2 nat. dist. higher
LogP = 2.0	0.330	0.364	similar

^aE1: full SpectralMol v2. E2: no-conditioning ablation (equal architecture, no property input). Hit rate: fraction within $\pm 20\%$ of target. E2 hit rates are that model's natural distribution overlap with the target region, not conditioning-driven.

reflects natural distribution overlap rather than superior conditioning; the ablation confirms the conditioning signal is active through the Uniq@1k difference (87.4% vs 86.5%). For reference, the unconditioned generation baseline (drawing uniformly from the prior) would achieve the empirical frequency of QED $\in [0.72, 1.0]$ in ZINC-250K training data, which is approximately 12%; the large gap between SpectralMol's target (QED = 0.9) and its achievable range (≤ 0.35 at any temperature) reveals that extreme-property generation remains a hard open problem for VAE-based models. We discuss this limitation in Section "Conclusion".

Effect of Latent Dimension. The small latent model (E4, $d_z = 56$) achieves Uniqueness@1k = 76.7% vs E1's 87.4% and a conditional QED hit rate of 15.2% vs E1's 12.2%, consistent with smaller latent capacity reducing fine-grained property encoding.

Property-Conditioned Generation

We evaluate SpectralMol's ability to steer generation toward five *single-property* targets: QED $\in \{0.7, 0.9\}$, LogP $\in \{2.0, 4.0\}$, SA = 3.0. For each target we generate 1000 molecules at $\tau = 1.0$ and report both the hit rate (within $\pm 20\%$ of target) and the mean absolute error (MAE) between generated and target property value. We emphasize at the outset that absolute hit rates remain modest in all cases, and that under the robust metrics reported later (Table 13) the conditioning does *not* outperform an unconditioned baseline. We therefore present these results as a negative finding about the guidance strength of the current model, not as a demonstration of controllability.

Three observations are notable:

Multitarget Interference: Conditioning is applied jointly to two targets simultaneously (QED = 0.9 and LogP = 2.0). These two objectives can conflict: molecules with very high QED tend to be aromatic and may have LogP outside the [2, 3] range. The model's conditioning signal must balance both simultaneously, limiting single-target performance.

Natural Distribution Overlap: The v_2 model generates molecules with mean QED = 0.603 (std 0.12), so the natural fraction satisfying QED > 0.72 (the lower bound of the $\pm 20\%$ window around 0.9) is already $\approx 15\text{--}20\%$ without any conditioning. E2's 20.0% hit rate is consistent with this natural overlap, not evidence that conditioning improves E2.

Conditioning Signal is Active: The ablation confirms that conditioning changes generation quality (E1 vs E2 Uniq@1k: 87.4% vs 86.5%) and the model does shift its distribution toward targets—but the available hit-rate metric is too coarse to distinguish conditioning-driven improvement from natural-distribution overlap when the targets are within reach of the unconditioned model. MAE-based evaluation (mean absolute deviation of generated QED from target) and single-target condition-

ing experiments are designated as follow-up work to precisely quantify the conditioning signal's contribution.

Latent Space Analysis (Negative Result)

Gradient-Based Property Steering. We attempted gradient ascent in the latent space targeting QED > 0.8 and $0 \leq \text{LogP} \leq 3$ from 500 random prior samples. As reported in Supporting Information, 30 gradient steps produced **no measurable improvement** ($\Delta \approx 0$ for all three conditions), because v_2 's vocabulary already naturally generates molecules with mean LogP = 1.40 (within target range) from random prior samples (good-LogP before: 85.4%). This is a negative result: latent gradient optimization is not effective for this model, and the property prediction heads do not identify useful gradient directions beyond the model's natural prior. We include this result for completeness; effective latent optimization would require coupling with an external property predictor. Full results are in Supporting Information.

Decomposition Analysis

Coverage and Robustness by Graph Topology. We apply the Fiedler-BFS decomposition to all 250,000 molecules in ZINC-250K and categorize results by structural complexity. Table 7 reports success rates by topology category; zero failures

Table 7. Fiedler-BFS Decomposition Success Rate on the Full ZINC-250K Set (249,456 Molecules) Stratified by Ring Topology (RDKit Ring-Info Classification: Acyclic/Single Ring/Fused/Bridged/Spiro/Multi-Ring Polycyclic)^a

Ring topology	Molecules	SpectralMol	JT-VAE coverage
Acyclic/single ring	21,653	100%	Full support
Fused ring systems	78,058	100%	Mostly supported
Multiring polycyclic	141,604	100%	Partly supported
Bridged systems	3,374	100%	Known to fail ^b
Spiro systems	4,767	100%	Known to fail ^b
Total	249,456	100%	$\approx 97\text{--}98\%$^b

^aJT-VAE coverage is a qualitative assessment from Jin et al.¹

^bQualitative assessment from Jin et al.;¹ specific failure rates not reproduced here under matched protocol.

occur in any category. For comparison, JT-VAE's junction-tree decomposition is estimated to fail on molecules with bridged bicyclic and spiro ring systems (qualitatively described as a known limitation for complex ring systems in the original paper;¹ specific rates not directly reported). **JT-VAE rates are estimates.** The JT-VAE failure rates in Table 7 are *estimates derived from the original paper*,¹ not direct reproductions under a matched protocol and data set split. Furthermore, decomposition success measures only the preprocessing step—it does not imply equivalent reconstruction accuracy or end-to-end generation quality on complex ring systems. A fair head-to-head comparison at the *generative* level is now provided by the matched-protocol study of Section "Decomposition Analysis", where an SSSR-based junction-tree decomposition—the essence of JT-VAE—is run through the identical Tree-VAE; the estimated failure rates in Table 7 are retained only as qualitative context and no claim rests on them. A ring-stratified comparison of the *trained* models, as opposed to the decompositions, remains future work.

The 100% decomposition success rate on bridged systems (3,374 molecules) and spiro systems (4,767) is particularly important, as these are precisely the topology classes where JT-VAE fails. **Token-level truncation is independent of coverage.**

While decomposition succeeds on all molecules, a substantial fraction of *fine-grained* fragments are mapped to the UNK token under the frequency-truncated 5000-vocabulary (Section “Fragment Vocabulary Statistics”); this token-level truncation is independent of the 100% structural coverage reported here and does not affect generation validity, since UNK is masked at decode time. The JT-VAE topology coverage entries in Table 7 are qualitative assessments based on the SSSR-based decomposition limitations described in Jin et al.;¹ specific failure rates were not directly reported in that paper and are not reproduced here under a matched protocol. The spectral approach imposes no constraints on ring fusion or cross-ring bond patterns—it simply partitions atoms by the algebraic connectivity bipartition, which is well-defined for any connected graph.

Reconstruction Accuracy. Exact SMILES reconstruction (decompose then reassemble via the v2 assembler) was measured on a 1,999-molecule sample (seed = 42). By topology class:

- **Simple rings:** 94.3% exact reconstruction
- **Fused bicyclics:** 88.2% exact reconstruction
- **Bridged bicyclics:** 84.9% exact reconstruction
- **Spiro systems:** 81.8% exact reconstruction
- **Overall (5 seeds):** $86.2\% \pm 1.1\%$ exact reconstruction

Of the 12.6% failures, 99.2% are stereochemistry losses only (same connectivity graph, different E/Z or R/S assignment at cut sites); 0.8% are valence corrections at cross-double-bond cuts; 0% are connectivity errors. Bridged and spiro systems show slightly lower rates due to more complex stereocenters introduced at cut sites, confirming that the decomposition is structurally sound and that failures are limited to stereo notation rather than molecular topology, which is the foundation for generative modeling.

Tree Statistics. The average Fiedler-BFS decomposition tree has 2.36 fragment nodes (median 2, maximum 6; Table 15), consistent with the coarse fragment count expected for drug-like molecules (MW 200–500 Da). The frequency-truncated vocabulary of 5000 fragments covers 37.5% of tree-node instances in the validation set; the remaining 62.5% map to the UNK token, which is masked out of the decoder distribution and never emitted (Section “Fragment Vocabulary Statistics”).

Comparison with BRICS Rule-Based Decomposition. A natural question is how spectral bisection compares with BRICS (Break Retrosynthetically Interesting Chemical Substructures; Degen et al. 2008⁴³), the standard chemistry-aware fragmentation approach used in retrosynthesis planning. We ran both methods on $N = 5,000$ randomly sampled ZINC-250K molecules (seed = 42) under identical evaluation conditions. Table 8 reports the results.

BRICS produces 5.13 ± 0.01 fragments per molecule on average (vs 2.31 ± 0.01 for Fiedler-BFS) with smaller individual fragments (5.84 ± 0.01 vs 9.72 ± 0.01 atoms/fragment). This is by design: BRICS was developed for retrosynthesis, breaking bonds at chemically motivated sites (BRICS bond-type rules) to produce fragments suitable for library synthesis planning. For a *generative* fragment-based VAE, however, finer fragmentation has two drawbacks: (i) vocabulary size grows with the number of unique small fragments, increasing sample complexity; and (ii) fragment trees become deeper and harder for the decoder to assemble. Random bisection rarely splits molecules at all (1.07 ± 0.00 frags/mol), providing little useful factorization. Fiedler-BFS balances structural completeness (9.72 ± 0.01 atoms/frag, $92.3 \pm 0.2\%$ atom-count preservation) and nontrivial decomposition

Table 8. Decomposition Comparison: Fiedler-BFS Spectral Bisection vs BRICS Rule-Based Fragmentation vs Random Bisection^a

Method	Success	Frgs/Mol	Atoms/Frag	Reconstr.
Fiedler-BFS	100.0%	2.31 ± 0.01	9.72 ± 0.01	$92.3\% \pm 0.2\%$
BRICS	100.0%	5.13 ± 0.01	5.84 ± 0.01	$11.5\% \pm 0.3\%$
Random bisection	100.0%	1.07 ± 0.00	21.7 ± 0.03	$99.9\% \pm 0.0\%$

^a $N = 5,000$ ZINC-250K molecules per seed, 5 random seeds. Reported as mean $\pm$ standard deviation. Frgs/Mol: mean number of fragments per molecule. Atoms/Frag: mean heavy-atom count per fragment. Reconstr.: fraction of molecules where total fragment atom count is within $\pm 15\%$ of original (a proxy for structural completeness). Note: BRICS replaces attachment-point dummy atoms with H; the low reconstruction rate reflects this stub-atom removal, not a failure of BRICS per se. Scope: This is a decomposition-level comparison. The corresponding generative comparison, in which the identical Tree-VAE is retrained under each fragmentation scheme, is reported in Section “Decomposition Analysis” (Tables 11 and 12)

(2.31 ± 0.01 frags/mol). This middle ground is confirmed to be the more suitable one for hierarchical generative modeling by the matched-protocol retraining reported in Section “Decomposition Analysis”. The low variance across 5 random seeds confirms the robustness of these decomposition characteristics.

The numbers in this table are decomposition-level, measured on 5,000 molecules per seed over 5 seeds; the matched-protocol generative comparison (Section “Decomposition Analysis”) is run on the full training set and reports 2.36 frags/mol for Fiedler-BFS, the small difference reflecting the larger sample rather than a different procedure. We note that a useful molecular decomposition must satisfy three simultaneous criteria: nontrivial partitioning (more than one fragment), structural completeness (fragments together preserve the original atom count), and bounded fragment size suitable for generative modeling; a trivially high reconstruction rate from essentially no decomposition (Random, 1.07 frags/mol) does not satisfy the nontriviality criterion.

Hyperparameters and Reproducibility. Table 9 reports key hyperparameters for the Tree-VAE generative validation. Decomposition parameters: $M_{\max} = 12$ (max fragment size), vocabulary size 5000 (frequency-ranked). Encoder: Child-Sum Tree-LSTM with hidden dimension 450, embedding dimension 450. Decoder: GRU with hidden dimension 450. Latent space: $d_z = 128$ (main model), $d_z = 56$ (small ablation). Training: Adam optimizer, learning rate 10^{-3} , batch size 32, 50 epochs on ZINC-250K. IMQ-MMD: $\beta = 1.0$, $c = 1.0$. Classifier-free guidance: 10% condition dropout, $w = 1.0$ (inference). Sampling temperatures: $\tau = 1.0$ (lower temperature), $\tau = 2.0$ (higher temperature, fidelity/coverage trade-off).

Computational Resources and Runtime. All experiments were conducted on a single NVIDIA RTX 3090 GPU (24GB VRAM). The full model (8.18 M parameters, $d_z = 128$) trains in approximately 6.5 h for 50 epochs on ZINC-250K (220 K molecules, batch size 32). Per-molecule decomposition via Fiedler-BFS averages 0.08 s (including eigendecomposition). Inference (sampling 10K molecules at $\tau = 1.0$) takes approximately 12 min. Peak GPU memory usage during training: 18.2 GB. Code and trained models will be made publicly available upon acceptance.

Table 9. Key Hyperparameters for SpectralMol Tree-VAE

Component	Configuration
Decomposition	
Max fragment size ($M_{\max}$)	12 heavy atoms
Vocabulary size	5000 (frequency-ranked)
Encoder (Tree-LSTM)	
Hidden dimension	450
Embedding dimension	450
Decoder (GRU)	
Hidden dimension	450
Latent space	
Dimension (d_z)	128 (main), 56 (ablation)
Training	
Optimizer	Adam
Learning rate	10^{-3}
Batch size	32
Epochs	50
Data set	ZINC-250K
IMQ-MMD	
β (weight)	1.0
c (kernel scale)	1.0
Classifier-free guidance	
Condition dropout	10%
Guidance weight (w)	1.0 (inference)
Sampling	
Temperature (τ)	1.0 (lower), 2.0 (fidelity/coverage trade-off)

Additional Experiments

This subsection reports further analyses of the decomposition and the generative model: cross-data set generalization, a matched-protocol decomposition comparison, chemical interpretability of the cuts, robust conditioning metrics, and a runtime/scalability analysis.

Cross-Data Set Generalization. To test the “practically universal” claim beyond ZINC-250K, we ran Fiedler-BFS on four additional data sets spanning very different chemical spaces (Table 10). Coverage is 100% and exact reconstruction

Table 10. Fiedler-BFS Decomposition across Five Data Sets (5,000 Molecules Each)^a

Data set	n	Heavy (mean)	Coverage	Exact recon.
ZINC-250K (test)	5,000	23.1	100.0%	99.94%
QM9	5,000	8.1	100.0%	100.0%
MOSES (ChEMBL)	5,000	21.7	100.0%	99.94%
ChEMBL-large (≥ 25 heavy)	5,000	25.2	100.0%	100.0%
ChEMBL-polycyclic (≥ 4 rings)	5,000	23.7	100.0%	99.94%

^aCoverage is the fraction of molecules fully decomposed; exact reconstruction is the fraction reassembled to the original canonical SMILES.

$\geq 99.94\%$ on every data set and every topology class, with stable fragment counts (2.1–2.5 fragments/molecule). BRICS (5.4) and junction-tree (8.7–9.3) produce far more fragments per molecule throughout; the full three-method breakdown per data set is given in Table S1 of the Supporting Information.

Matched-Protocol Decomposition Comparison. To isolate the contribution of the decomposition strategy from that of the generative model, we trained the *identical* Tree-VAE (same architecture, assembler, $d_z = 128$, optimizer, 50-epoch schedule, seed 42) under three decompositions, varying *only* the

cut-bond set: Fiedler-BFS (recursive spectral bisection to ≤ 12 atoms), BRICS (all breakable bonds), and junction-tree (all acyclic single bonds). All three decompose 100% of molecules with zero failures; the complete shared configuration, itemized to show that every setting other than the cut-bond set is identical, is given in Table S2 of the Supporting Information. The converged validation reconstruction loss (Table 11, per-

Table 11. Matched-Protocol Comparison: Identical Tree-VAE Trained under Three Decompositions^{ab}

Decomposition	Frgs/mol	Atoms/frag	Val. recon. loss ($\downarrow$)
Fiedler-BFS	2.36	9.03	0.012
BRICS	5.13	5.84	0.048
Junction-tree	9.28	2.49	0.083

^aReconstruction loss is the converged validation negative log-likelihood (lower is better). ^bAll three models were trained to convergence under the same 50-epoch budget, seed, and hyperparameters; only the decomposition differs.

epoch curves in Table S3) is far lower under Fiedler-BFS, showing that its coarser, spectrally coherent fragments yield shorter, more predictable trees that the generator models more accurately—an improvement in generative modeling quality *beyond* the coverage property.

Generation metrics under the matched protocol (10,000 samples each) are reported in Table 12, with the complete

Table 12. Matched-Protocol Generation Metrics (10,000 Samples; Identical Tree-VAE, Assembler, and Hyperparameters; Only the Decomposition Differs)^a

Decomposition	τ	Validity	Uniq@1k	Novelty	QED	FCD ($\downarrow$)
Fiedler-BFS	1.0	1.000	0.874	0.999	0.603	14.41
Fiedler-BFS	2.0	1.000	0.834	0.998	0.631	13.59
BRICS	1.0	1.000	0.580	1.000	0.479	23.62
BRICS	2.0	1.000	0.681	1.000	0.501	23.52
Junction-tree	1.0	1.000	0.284	1.000	0.405	30.48
Junction-tree	2.0	1.000	0.491	1.000	0.428	28.68

^aFCD lower is better.

metric set at both sampling temperatures in Table S4 of the Supporting Information. The decomposition strategy has a large, isolated effect on generative quality: at $\tau = 1.0$, Fiedler-BFS attains an FCD of 14.41 versus BRICS's 23.62 (a $\sim 40\%$ improvement), together with higher drug-likeness (QED 0.603 vs 0.479), higher uniqueness@1k (0.874 vs 0.580), and more synthesizable molecules (SA 2.70 vs 3.51). Because everything except the cut-bond set is identical, this gap is attributable to the decomposition alone—directly answering the question of whether Fiedler-BFS improves generative modeling beyond coverage. At the same $\tau = 1.0$ the junction-tree decomposition—the SSSR-based essence of JT-VAE—is the weakest of the three (FCD 30.48, QED 0.405, uniqueness@1k 0.284, SA 5.55), so Fiedler-BFS's advantage holds against both rule-based (BRICS) and SSSR-based (junction-tree) fragmentations. Because all three share byte-for-byte identical model, assembler, and hyperparameters, the ordering Fiedler < BRICS < junction-tree in FCD is attributable to the decomposition alone.

Chemical Interpretability of Spectral Cuts. Although Fiedler-BFS uses no chemical rules, its cuts are chemically sensible. Across ZINC-250K, 69.9% of Fiedler cut bonds are

Table 13. Robust Conditioning Metrics: Conditioned vs Unconditioned Generation across Five Targets (1,000 Samples Each)^a

Target	Cond. MAE	Uncond. MAE	Mean shift	KS	Cond. hit	Uncond. hit
QED = 0.7	0.129	0.117	−0.010	0.067	0.575	0.612
QED = 0.9	0.300	0.297	−0.004	0.034	0.125	0.125
LogP = 2.0	0.944	0.934	−0.056	0.050	0.304	0.326
LogP = 4.0	2.693	2.635	−0.070	0.037	0.034	0.034

scorer unavailable on evaluation host; omitted

^aMean shift is (conditioned mean − unconditioned mean) in the direction of the target; KS is the Kolmogorov–Smirnov statistic between the two property distributions. Small KS and near-zero shift indicate weak conditioning; the unconditioned hit rate is never worse than the conditioned one.

acyclic and 68.5% rotatable (only 30.1% in-ring), versus a random-cut baseline of 40.4%/22.4%/59.6%: the spectral bipartition preferentially severs flexible inter-ring linkers. Fiedler cuts coincide with a BRICS retrosynthetic bond at 53.7% precision (14.5% recall, reflecting Fiedler’s far coarser cutting). The resulting fragments are chemically coherent: 84.9% contain an intact ring system and 88.0% contain at least one functional group. Thus, meaningful chemical units emerge from the algebraic-connectivity bipartition without hand-coded rules. The full set of interpretability statistics, including the random-cut control, is given in Table S7 of the Supporting Information.

Robust Conditioning Metrics. Beyond $\pm 20\%$ hit rate, we evaluate five single-property targets ($\text{QED} \in \{0.7, 0.9\}$, $\text{LogP} \in \{2.0, 4.0\}$, $\text{SA} = 3.0$), comparing the conditioned model against the unconditioned model (null embedding) via MAE to target, mean shift toward target, and the Kolmogorov–Smirnov statistic between conditioned and unconditioned property distributions (Table 13, reproduced with the full per-target breakdown in Table S10 of the Supporting Information). We report a negative result: across all targets the conditioning signal is weak—the KS statistic between the conditioned and unconditioned property distributions is small (≤ 0.067), the mean shift toward the target is near zero (and in some cases slightly away from it), and the conditioned MAE and hit rate are not better than the unconditioned baseline. The unconditioned model attains comparable hit rates, and this is consistent with the negative latent-optimization result (Table 14). We conclude

Table 14. Latent Gradient Optimization (v2, $n = 500$ Starting Points)^a

Condition	Before (random z)	After (E1, 30 steps)	Δ
High-QED (>0.8)	0.028	0.028	+0.000
Good-LogP ($0 \leq \text{LogP} \leq 3$)	0.854	0.856	+0.002
Pareto (both)	0.016	0.016	+0.000

^a“Before”: random z decoded without optimization. “After”: 30 gradient steps targeting $\text{QED} > 0.8$ and $0 \leq \text{LogP} \leq 3$.

that the classifier-free guidance as trained provides only a mild steering effect; strengthening it—via a larger guidance weight w , a dedicated property predictor, or inference-time gradient guidance—is a concrete direction for future work rather than a solved capability. We report this openly to avoid overclaiming conditional control.

Runtime and Scalability. Per-molecule decomposition on real ZINC takes 1.22 ms (Fiedler-BFS) vs 0.81 ms (BRICS) and 0.68 ms (JT), scaling from 0.76 ms (≤ 20 heavy atoms) to 2.08 ms (31–40 heavy atoms). Training is ≈ 6.5 h for 50 epochs on one RTX 3090; the matched-protocol BRICS/JT models are 2–3 $\times$ slower per epoch purely from longer fragment sequences. On synthetic linear chains, Fiedler-BFS is near-linear to C100

(≈ 10 ms) but the dense eigensolve reaches ≈ 427 ms at C200, whereas BRICS/JT stay < 30 ms; for drug-like molecules ($\leq \sim 50$ heavy atoms) this regime is never reached, and sparse/Lanczos eigensolvers would remove the boundary for very large inputs. This is a scalability limit rather than a practical obstacle for the target domain. The full stratified timings are given in Table S11 (real ZINC molecules) and Table S12 (synthetic linear chains) of the Supporting Information.

Why Fiedler Decomposition over Junction Trees. Junction-tree decomposition relies on the smallest set of smallest rings (SSSR) to identify ring systems, and then decomposes the molecule into a tree of ring clusters and bonds.¹ This procedure fails when ring systems share more than two atoms (bridged bicyclics, spiro compounds) or when ring fusion patterns are complex. The Fiedler-BFS decomposition makes no such topological assumptions: it partitions atoms based on the algebraic connectivity of the graph, which is well-defined for any connected molecular graph. As a result, SpectralMol achieves a 100% decomposition success rate, a prerequisite for deployment on diverse chemical libraries beyond ZINC.

Fragment Vocabulary and the Diversity–Fidelity Trade-Off. The generated molecules exhibit a well-documented temperature-dependent distribution shift: at high temperature ($\tau = 2.0$), large fragments from the vocabulary tail are oversampled, producing high-MW, high-LogP molecules that exceed the ZINC-250K property range. At $\tau = 1.0$, SpectralMol v2 generates drug-like molecules ($\text{QED} = 0.603$, $\text{SA} = 2.70$, closely matching the ZINC-250K training distribution. The v2 vocabulary correction (Section “v2 assembler correction”) resolved the v1 QED gap; the remaining distributional difference is a slight LogP underestimate (mean 1.40 vs ~ 2.5), attributable to a polarity bias in the frequency-ranked vocabulary. As quantified in Section “Why the frequency-ranked vocabulary biases polarity”, QED-weighted vocabulary construction does not remove this bias; explicit LogP-balanced selection or a larger vocabulary is required.

End-to-End Integration vs the Current Modular Design. A natural question is whether the framework can be made *seamlessly end-to-end*, without the deterministic decomposition and assembly modules. In the current design these two stages are parameter-free: the Fiedler bisection is an exact spectral computation and the v2 assembler is a deterministic attachment-point matcher. This is a deliberate trade-off. Making the decomposition learnable would require a differentiable spectral relaxation—e.g., a soft bipartition driven by the Fiedler eigenvector with a straight-through estimator, or a learned edge-scoring cut network—which would gain differentiability but forfeit the exact 100% structural coverage guarantee that is the paper’s central property, since a learned cut can fail on the very bridged/spiro topologies where rule-based methods already fail. Symmetrically, replacing the deterministic assembler with an

autoregressive attachment-point predictor would allow gradients to flow end-to-end but reintroduce the invalid-assembly failure modes (mixture SMILES) that the deterministic v2 assembler eliminates. We therefore keep decomposition and assembly deterministic *by design*: they act as exact, coverage-preserving encoders/decoders of graph topology, and only the representation learning (Tree-LSTM, CFG fusion, GRU decoder) is trained. A fully end-to-end variant that preserves the coverage guarantee—most plausibly via a differentiable spectral cut with a coverage-penalty regularizer—is the primary direction for future work.

Limitations. Several limitations warrant future study. First, the baseline comparison in Table 2 omits MoLeR³⁹ and MoFlow⁴⁴ because direct metric comparison requires protocol alignment: MoLeR is scaffold-conditioned (vs our unconditional generation), and MoFlow uses different validity definitions and sampling procedures. Protocol-aligned comparison is designated as future work. Second, the v1 fragment assembler produced mixture SMILES (“.”-separated fragments) in ~68% of cases due to missing attachment-point preservation; this has been corrected in v2 (Section “v2 assembler correction”), and all generation metrics are from v2. FCD is computed on 10,000 actual model-generated SMILES against the full ZINC-250K validation set (24,445 molecules) using the standard `fcd` package. Third, the generated LogP (mean 1.40) is slightly below ZINC-250K (≈ 2.5), because the frequency-ranked vocabulary skews toward polar fragments (Section “Why the frequency-ranked vocabulary biases polarity”); we tested QED-weighted reweighting and found it ineffective, so LogP-aware or larger-vocabulary construction remains open. Fourth, the Fiedler vector is not unique when λ_2 has multiplicity >1 (symmetric molecules); we break ties consistently using atom-index ordering, but canonical graph labeling would provide stronger guarantees. Fifth, and more fundamentally than a matter of absolute difficulty, the classifier-free guidance as trained does *not* demonstrably steer the output distribution: under robust metrics (Table 13) the conditioned model matches, and never beats, an unconditioned baseline ($KS \leq 0.067$, near-zero mean shift). We report this as a negative result rather than a partial success; stronger inference-time guidance, a larger guidance weight, or a dedicated property predictor is needed before the model supports practical property-targeted design. Sixth, the latent gradient optimization results are preliminary; coupling with a dedicated external property predictor would substantially improve Pareto hit rates. Seventh, the IMQ-MMD anticollapse claim is now supported by a direct kernel-value analysis at $d_z = 128$ (Section “Ablation Study”: RBF mean 4.6×10^{-38} vs IMQ 0.502), which shows the RBF gradient vanishes for typical latent pairs; a full RBF-MMD retraining comparison remains future work. Eighth, the JT-VAE/BRICS comparison is now performed under a *matched protocol* (Section “Reconstruction accuracy”): the identical Tree-VAE is retrained under Fiedler-BFS, BRICS, and junction-tree decompositions, isolating the effect of the decomposition; the original estimated-failure-rate entries in Table 7 are retained only as qualitative context. Ninth, generalization beyond ZINC-250K is now demonstrated on QM9, MOSES, and two ChEMBL subsets (large and polycyclic), where Fiedler-BFS retains 100% decomposition coverage and stable fragment statistics (Section “Cross-dataset generalization”); we accordingly temper the wording to “practically universal.” Tenth, all generative model results are from a **single-seed training run** (seed = 42), consistent with community standards for proof-of-concept VAE

validation (cf. JT-VAE, HierVAE). Readers should treat the absolute generation metrics as preliminary estimates; multiseed retraining is designated as future work to establish statistical confidence. The decomposition experiments (Tables 1 and 8) report mean $\pm$ standard deviation over 5 random seeds, confirming robustness at the decomposition level. Eleventh, the Fiedler vector has inherent limitations: (a) it is not unique when the second eigenvalue λ_2 has multiplicity >1 (symmetric molecules); we break ties consistently using atom-index ordering, but canonical graph labeling would provide stronger guarantees; (b) spectral bisection optimizes algebraic connectivity (ratio cut), not chemical interpretability; while empirically the cuts tend to occur at molecular periphery (single bonds, acyclic connections), there is no guarantee that every fragment will be chemically meaningful; (c) eigenvector computation via ARPACK has $O(nk)$ complexity per molecule (where k is the number of Lanczos iterations), which is acceptable for offline preprocessing but may be slower than rule-based methods like BRICS for very large-scale applications.

■ LATENT GRADIENT OPTIMIZATION (NEGATIVE RESULT)

Table 14 reports the full latent optimization results. The “Before” column samples 500 random $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and decodes without any gradient steps; “After” applies 30 gradient steps targeting $QED > 0.8$ and $0 \leq \text{LogP} \leq 3$.

The v2 model’s natural distribution from random $\mathbf{z}$ already achieves 85.4% good-LogP (mean LogP = 1.40, within $[0, 3]$), so gradient optimization targeting LogP adds nothing. High-QED optimization is also ineffective, consistent with linear property prediction heads having insufficient expressivity to shift the distribution toward the $QED > 0.8$ tail. This negative result is reported for completeness and to guide future work on property-guided generation.

■ FRAGMENT VOCABULARY STATISTICS

Table 15 reports quantitative statistics of the fragment vocabulary constructed via Fiedler-BFS decomposition on ZINC-250K training set (220 K molecules). The vocabulary is frequency-ranked and truncated at size 5000; fragments outside the top 5000 are mapped to a special UNK token. A condensed version of these statistics is also given in Table S8 of the Supporting Information.

The fragment distribution is extremely heavy-tailed and diverse: fine-grained Fiedler-BFS decomposition produces 168,646 distinct fragments, far more than a frequency-truncated vocabulary of 5000 can enumerate. Consequently the top-5000 vocabulary covers only 37.5% of node instances at the token level, and 62.5% of nodes map to UNK. Crucially, this token-level vocabulary limitation is *decoupled* from the two properties we claim for the decomposition itself: (i) structural coverage of the decomposition is 100% across all molecules and ring topologies (Table 7), because coverage is a property of the Fiedler-BFS *cut procedure*, not of the finite token dictionary; and (ii) generation validity is 100%, because UNK tokens are masked out of the decoder’s output distribution and are never emitted. The heavy tail does mean that a purely frequency-ranked vocabulary discards much rare-fragment chemistry; a larger or chemistry-aware vocabulary (e.g., LogP-balanced selection, cf. Section “Limitations”) is a natural avenue to raise token coverage, and we report this openly as a limitation rather than a solved problem.

Table 15. Fragment Vocabulary Statistics from Fiedler-BFS Decomposition on the Full ZINC-250K Set (249,456 Molecules)^a

Metric	Value
Total unique fragments (before truncation)	168,646
Vocabulary size (after truncation)	5,000
Node coverage (in-vocabulary fraction)	37.5%
UNK rate (node level)	62.5%
Fragment size distribution	
Mean heavy atoms per fragment	9.03
Median heavy atoms per fragment	9
Min heavy atoms	1
Max heavy atoms	17 ^b
Fragment frequency distribution	
Top-1 fragment frequency	0.74%
Top-10 cumulative frequency	2.82%
Top-100 cumulative frequency	7.73%
Top-1000 cumulative frequency	19.96%
Fragments per molecule	
Mean fragments per molecule	2.36
Median fragments per molecule	2
Max fragments per molecule	6

^aVocabulary size: 5000 (frequency-ranked on the training split). Node coverage: fraction of decomposition-tree nodes represented by in-vocabulary (non-UNK) fragments. ^bThe recursive bisection targets $\leq M_{\max} = 12$ heavy atoms, but indivisible fused ring systems that cannot be split by any single bond cut are kept intact and may exceed this target.

CONCLUSION

We presented SpectralMol, a property-conditioned molecular VAE that replaces heuristic junction-tree decomposition with Fiedler-vector-based spectral decomposition. The key contributions are (1) a practically universal (no topological ring-system constraints) molecular decomposition with 100% coverage and 91.4% sanitization success (vs 7.1% random) on ZINC-250K, retained across five chemical spaces (Section “Cross-dataset generalization”); (2) a matched-protocol study in which the identical Tree-VAE is retrained under three decompositions, isolating the effect of the cut-bond set and showing that Fiedler-BFS yields shorter, more learnable fragment trees (Section “Decomposition Analysis”); (3) an implementation of classifier-free guidance for single-property conditioning, for which we report a *negative* result: the guidance as trained provides only a mild steering effect and does not outperform an unconditioned baseline (Table 13); and (4) IMQ-kernel MMD regularization, motivated theoretically and now also empirically at $d_z = 128$, where the RBF kernel is numerically zero while the IMQ kernel retains usable gradients.

SpectralMol v2 (with corrected FragmentOnBonds assembler) achieves 100% chemical validity, QED = 0.603 (matching the ZINC-250K training mean), SA = 2.70—competitive with JT-VAE and HierVAE on drug-likeness criteria. The FCD (14.41) remains substantially higher than recent fragment-level generative baselines, and we do not claim competitive distributional quality. The matched-protocol diagnostic attributes this gap to the Tree-VAE backbone rather than to the decomposition, which shifts FCD by only 0.009 under decompose-and-reassemble (Section “Decomposition Analysis”); the residual difference tracks the LogP and scaffold bias of the frequency-ranked vocabulary.

Key open directions include: property-aware vocabulary construction to close the FCD gap, noting that QED-weighted selection alone is *not* sufficient, since it returns essentially the same top-5000 set, so explicit LogP balancing or a larger vocabulary is required; a full RBF-MMD vs IMQ-MMD retraining comparison at $d_z = 128$ to complement the kernel-value analysis reported here; stronger conditional control (larger guidance weight, a dedicated property predictor, or inference-time gradient guidance) benchmarked against external baselines (e.g., REINVENT, CVAE); extending spectral decomposition to 3D conformers; and strengthening latent optimization with Bayesian or RL-based surrogate models.

ASSOCIATED CONTENT

Data Availability Statement

The source code, trained model evaluation results, and analysis scripts for SpectralMol are publicly available at <https://github.com/GangLi-Gcc/SpectralMol>. The ZINC-250K data set used for training and evaluation is available from the JT-VAE repository¹ and the original ZINC database.⁴⁰

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.6c01818>.

Cross-data set decomposition statistics for all three fragmentation methods on five data sets (Table S1); the shared configuration verifying that the matched-protocol runs differ only in the cut-bond set (Table S2); matched-protocol reconstruction learnability (Table S3) and generation quality at both sampling temperatures (Table S4); the decompose-and-reassemble FCD attribution diagnostic (Table S5); RBF versus IMQ kernel values over 5,000 random latent pairs at $d_z = 128$ (Table S6); the chemical-interpretability breakdown of spectral cuts against a random-cut baseline (Table S7); fragment-vocabulary statistics (Table S8); vocabulary polarity and the two tested reweighting remedies (Table S9); robust conditioning metrics against an unconditioned baseline (Table S10); and decomposition runtime on real ZINC molecules stratified by size (Table S11) and on synthetic linear chains up to C200 (Table S12), together with the reproducibility record (seeds, splits, hardware, and software versions) (PDF)

AUTHOR INFORMATION

Corresponding Authors

Gang Li – School of Information Technology and Engineering, Guangzhou College of Commerce, Guangzhou 511363, China; Email: ligang@gcc.edu.cn

Teng Zhou – School of Cyberspace Security, Hainan University, Haikou 570228, China; orcid.org/0000-0003-1920-8891; Email: teng.zhou@hainanu.edu.cn

Weihua Bai – School of Computer Science and Software, Zhaoqing University, Zhaoqing 526061, China; orcid.org/0000-0001-8333-7415; Email: baiweihua@zqu.edu.cn

Authors

Zhizhe Lin – School of Cyberspace Security, Hainan University, Haikou 570228, China

Zhifeng Hao – College of Mathematics and Computer Science, Shantou University, Shantou 515063, China

Keqin Li – Department of Computer Science, State University of New York, New Paltz, New York 12561, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.6c01818>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (2025YFC3410000, 2025YFC3410003), the National Natural Science Foundation of China (62476163, 6244100778, U24A20233, 62462021), the Natural Science Foundation of Guangdong Province (23201910250000514), the Guangdong Basic and Applied Basic Research Foundation (2025A1515010197), and the University-Level High-Caliber Training Project of Zhaoqing University (No. gcc202607). Gang Li and Zhizhe Lin contributed equally to this work. Teng Zhou and Weihua Bai are the corresponding authors.

REFERENCES

- (1) Jin, W.; Barzilay, R.; Jaakkola, T. Junction Tree Variational Autoencoder for Molecular Graph Generation. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*; PMLR, 2018; pp 2323–2332.
- (2) Jin, W.; Barzilay, R.; Jaakkola, T. Hierarchical Generation of Molecular Graphs using Structural Motifs. In *Proceedings of the 37th International Conference on Machine Learning*; PMLR, 2020; pp 4839–4848.
- (3) Jo, J.; Lee, S.; Hwang, S. J. Score-based Generative Modeling of Graphs via the System of Stochastic Differential Equations. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*; PMLR, 2022; pp 10362–10383.
- (4) Vignac, C.; Krawczuk, I.; Siraudin, A.; Wang, B.; Cevher, V.; Frossard, P. DiGress: Discrete Denoising Diffusion for Graph Generation. *arXiv*. **2023**.
- (5) Shi, C.; Xu, M.; Zhu, Z.; Zhang, W.; Zhang, M.; Tang, J. GraphAF: a Flow-based Autoregressive Model for Molecular Graph Generation. *arXiv*. **2020**.
- (6) Lucas, J.; Tucker, G.; Grosse, R. B.; Norouzi, M. Don't Blame the ELBO! A Linear VAE Perspective on Posterior Collapse; NeurIPS, 2019.
- (7) Fiedler, M. Algebraic connectivity of graphs. *Czechoslovak Math. J.* **1973**, 23, 298–305.
- (8) Preuer, K.; Renz, P.; Unterthiner, T.; Hochreiter, S.; Klambauer, G. Fréchet ChemNet Distance: A Metric for Generative Models for Molecules in Drug Discovery. *J. Chem. Inf. Model.* **2018**, 58, 1736–1741.
- (9) Ho, J.; Salimans, T. Classifier-Free Diffusion Guidance. *arXiv*. **2022**.
- (10) Tolstikhin, I.; Bousquet, O.; Gelly, S.; Schölkopf, B. Wasserstein Auto-Encoders. *arXiv*. **2018**.
- (11) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, 4, 268–276.
- (12) Kusner, M. J.; Paige, B.; Hernández-Lobato, J. M. Grammar Variational Autoencoder. *Proceedings of the 34th International Conference on Machine Learning (ICML)*; PMLR, 2017; pp 1945–1954.
- (13) Dai, H.; Tian, Y.; Dai, B.; Skiena, S.; Song, L. Syntax-Directed Variational Autoencoder for Structured Data. In *arXiv*. **2018**.
- (14) Bagal, V.; Aggarwal, R.; Vinod, P. K.; Priyakumar, U. D. MolGPT: Molecular Generation Using a Transformer-Decoder Model. *J. Chem. Inf. Model.* **2022**, 62, 2064–2076.
- (15) Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. REINVENT 2.0: an AI Tool for De Novo Drug Design. *J. Chem. Inf. Model.* **2020**, 60, 5918–5922.
- (16) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, 28, 31–36.
- (17) Simonovsky, M.; Komodakis, N. GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders. In *Artificial Neural Networks and Machine Learning – ICANN 2018*; Springer: Cham, 2018; pp 412–422.
- (18) Liu, Q.; Allamanis, M.; Brockschmidt, M.; Gaunt, A. L. Constrained Graph Variational Autoencoders for Molecule Design. In *Advances in Neural Information Processing Systems (NeurIPS)*; NeurIPS, 2018.
- (19) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **2012**, 4, 90–98.
- (20) Lim, J.; Ryu, S.; Kim, J. W.; Kim, W. Y. Molecular Generative Model Based on Conditional Variational Autoencoder for De Novo Molecular Design. *J. Cheminform* **2018**, 10 (1), 31.
- (21) Guan, J.; Qian, W. W.; Peng, X.; Su, Y.; Peng, J.; Ma, J. 3D Equivariant Diffusion for Target-Aware Molecule Generation and Affinity Prediction. *arXiv*. **2023**.
- (22) Kang, S.; Cho, K. Conditional Molecular Design with Deep Generative Models. *J. Chem. Inf. Model.* **2019**, 59, 43–52.
- (23) Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.* **2000**, 22 (8), 888–905.
- (24) Ng, A. Y.; Jordan, M. I.; Weiss, Y. On Spectral Clustering: Analysis and an Algorithm. In *Advances in Neural Information Processing Systems*; NeurIPS: Cambridge, MA, 2001; pp 849–856.
- (25) Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, 103, 8577–8582.
- (26) Pothen, A.; Simon, H. D.; Liou, K.-P. Partitioning Sparse Matrices with Eigenvectors of Graphs. *SIAM J. Matrix Anal. Appl.* **1990**, 11, 430–452.
- (27) Ron, D.; Safro, I.; Brandt, A. Relaxation-Based Coarsening and Multiscale Graph Organization. *Multiscale Model. Simul.* **2011**, 9, 407–423.
- (28) Kipf, T. N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv*. **2017**.
- (29) Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In *Advances in Neural Information Processing Systems (NeurIPS)*; NeurIPS, 2016.
- (30) Ying, Z.; You, J.; Morris, C.; Ren, X.; Hamilton, W. L.; Leskovec, J. Hierarchical Graph Representation Learning with Differentiable Pooling. In *Advances in Neural Information Processing Systems 31*; NeurIPS, 2018; pp 4805–4815.
- (31) Dhillon, I. S.; Guan, Y.; Kulis, B. Weighted Graph Cuts without Eigenvectors: A Multilevel Approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, 29, 1944–1957.
- (32) Zhao, S.; Song, J.; Ermon, S. InfoVAE: Balancing Learning and Inference in Variational Autoencoders. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI, 2019; Vol. 33, pp 5885–5892.
- (33) Kong, X.; Huang, W.; Tan, Z.; Liu, Y. Molecule Generation by Principal Subgraph Mining and Assembling. In *Advances in Neural Information Processing Systems*; NeurIPS, 2022; pp 2550–2563.
- (34) Geng, Z.; Xie, S.; Xia, Y.; Wu, L.; Qin, T.; Wang, J.; Zhang, Y.; Wu, F.; Liu, T.-Y. De Novo Molecular Generation via Connection-aware Motif Mining. *arXiv*. **2023**.
- (35) Qin, Y.; Madeira, M.; Thanou, D.; Frossard, P. DeFoG: Discrete Flow Matching for Graph Generation. *arXiv*. **2025**.
- (36) Lee, J.; Moon, S.; Kim, S.; Kim, H.; Kim, W. Y. FragFM: Hierarchical Framework for Efficient Molecule Generation via Fragment-Level Discrete Flow Matching. *arXiv*. **2025**.

- (37) Inukai, T.; Yamato, A.; Akiyama, M.; Sakakibara, Y. Leveraging tree-transformer VAE with fragment tokenization for high-performance large chemical model generation. *Commun. Chem.* **2025**, 8 (1), 228.
- (38) Yang, N.; Wu, H.; Zeng, K.; Li, Y.; Bao, S.; Yan, J. Molecule generation for drug design: A graph learning perspective. *Fundam. Res.* **2026**, 6 (1), 40–52.
- (39) Maziarz, K.; Jackson-Flux, H. R.; Cameron, P.; Sirockin, F.; Schneider, N.; Stiefl, N.; Segler, M. H. S.; Brockschmidt, M. Learning to Extend Molecular Scaffolds with Structural Motifs. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25–29, 2022; ICLR, 2022*.
- (40) Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, 55, 2324–2337.
- (41) Tai, K. S.; Socher, R.; Manning, C. D. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL); ACL, 2015*; pp 1556–1566.
- (42) Polykovskiy, D.; et al. Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models. *Front. Pharmacol.* **2020**, 11, 565644.
- (43) Degen, J.; Wegscheid-Gerlach, C.; Zaliani, A.; Rarey, M. On the Art of Compiling and Using ‘Drug-Like’ Chemical Fragment Spaces. *ChemMedChem* **2008**, 3, 1503–1507.
- (44) Zang, C.; Wang, F. MoFlow: An Invertible Flow Model for Generating Molecular Graphs. In *KDD’20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; ACM, 2020*; pp 617–626.