GEMS: Molecular Structure Identification via Geodesic Navigation of the Isomer Manifold

Utku Umur Acikalin [1] **Yingheng Wang** [1] **Goncalo J. Gouveia** [1] **Tyler Schwertfeger** [1] **Frank C Schroeder** [1] **Carla P Gomes** [1]

Abstract

Identifying small-molecule structures from tandem mass spectra (MS2) is a central bottleneck in untargeted metabolomics. Existing approaches either retrieve candidates from spectral libraries, limiting discovery to known compounds, or generate ranked final samples without maintaining validated molecular-hypothesis trajectories. We introduce **GEMS** (Geodesic Elucidation of Molecular Structures), a seeded isomer-navigation method that searches a chemically validated isomer manifold with valence-preserving 2-switch operations. Rather than mapping a spectrum directly to a structure, GEMS maintains a current molecular hypothesis, compares its forward-predicted spectrum to the observation, and uses the resulting correspondence to select a local edit or HALT action. Each retained rollout state is a valid isomer, yielding inspectable trajectories from high-quality retrieved seeds. We train the policy with PubChem-grounded shortest-path supervision certified under the stated atom-alignment and validity model. At inference time, multi-start search records every visited valid molecule and ranks unique candidates by forward spectral cosine and policy halt probability. With target-excluded starts, GEMS reaches the target in at least one trajectory for 72.19% of held-out NIST unique spectra and obtains 63.39% top-10 accuracy; without additional training, it reaches the target for 32.02% of CASMI spectra. GEMS achieves substantially higher NIST top-1 accuracy than recent generative baselines while using roughly two to three orders of magnitude fewer iterative model calls per query.

[1]Cornell University. Correspondence to: Utku Umur Acikalin <ua45@cornell.edu>.

Accepted at the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026)

1. Introduction

Tandem mass spectrometry (MS2) is a primary tool for identifying small molecules in biological samples, environmental monitoring, and drug metabolism. Given an MS2 spectrum, the task is to infer the molecular structure that produced it. This *structure identification* problem remains a key bottleneck in untargeted metabolomics, with public repositories containing millions of unidentified spectra (da Silva et al., 2015).

The dominant paradigm is **spectral library search**: match the query against reference spectra from characterized compounds. This is effective when the target is in the library, but cannot identify novel or rare molecules, and current libraries cover only a small fraction of biologically relevant chemical space (Wang et al., 2016). A second line of work predicts **molecular fingerprints** from spectra and searches structural databases (Dührkop et al., 2015), but remains limited to molecules already catalogued.

Recent **de novo generation** methods attempt to bypass database limitations by translating spectra to molecular strings (Stravs et al., 2022; Litsa et al., 2023; Butler et al., 2023) or by sampling molecular graphs with diffusion or flow models (Bohde et al., 2025; Nie & Gao, 2026; Sun et al., 2026). These methods formulate structure elucidation as direct generation of final candidate molecules, either by sequence decoding or by iterative graph refinement from a global spectral condition. GEMS uses a different interface between spectral evidence and molecular generation: it maintains a connected valid isomer as the current molecular hypothesis, predicts the MS2 spectrum of that hypothesis with a frozen forward model, merges the observed and predicted spectra into a spectral correspondence graph, and chooses a local valence-preserving edit or HALT action from this observed–predicted correspondence. Thus, the inverse problem is expressed as an inspectable trajectory of chemically meaningful molecular hypotheses, where both spectral agreement and disagreement inform editing and termination, rather than only as a ranked set of generated final samples.

We propose **GEMS** (Geodesic Elucidation of Molecular Structures), a different approach grounded in chemical

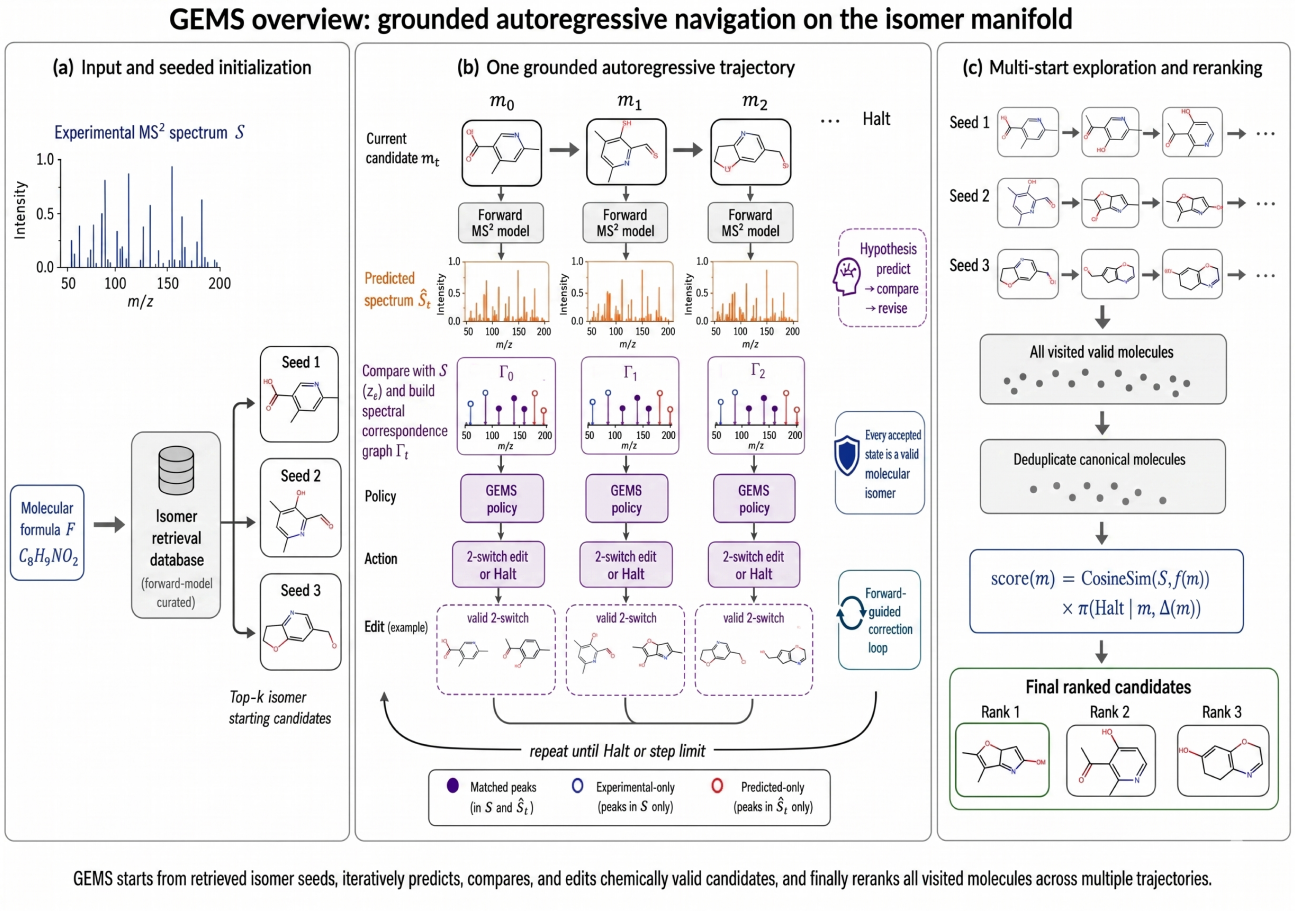

Figure 1. **GEMS overview.** Given an MS2 spectrum S and formula $\mathcal{F}$, GEMS retrieves candidate isomer seeds from a forward-model-curated database **(a)**. From each seed, GEMS runs an autoregressive trajectory **(b)**: at step t, the frozen forward model predicts $\hat{S}_t = f(m_t)$, the observed and predicted spectra are merged into a spectral correspondence graph Γ_t, and a policy selects a validated 2-switch edit or HALT. Multiple trajectories run from different seeds **(c)**; all visited valid molecules are deduplicated and ranked by $\text{score}(m) = \text{CosineSim}(S, f(m)) \cdot \pi(\text{HALT} \mid m, \Gamma(m))$. The output is a ranked list of valid candidate structures.

graph theory. We assume the precursor molecular formula is provided, as in formula-conditioned benchmarks and high-resolution MS workflows with confident formula assignment. GEMS navigates the resulting isomer manifold with valence-preserving **2-switch operations**, so every intermediate state remains chemically feasible.

GEMS treats structure elucidation as *forward-model-guided inverse search*: from an isomer candidate, it predicts the spectrum with a forward MS2 model, compares that prediction to the experimental spectrum, and uses the spectral correspondence signal to choose a 2-switch edit or learned HALT action. Every admitted molecule is a valid isomer, so rollouts produce chemically meaningful hypotheses rather than latent intermediates. Because the policy is *Markovian*—conditioned on the current molecule and spectral correspondence graph, not history—the same policy applies after any valid edit; recovery is measured by autoregressive inference. Typical trajectories require only a few validated edits, unlike iterative samplers that refine non-final graph states over many steps.

Our contributions are:

1. We formalize seeded structure identification as navigation on an isomer manifold with validity-checked valence-preserving edits (Section 3.1).

2. We introduce a forward-guided inverse policy conditioned on the current candidate and observed–predicted spectral correspondence graph (Section 3.2).

3. We develop PubChem-grounded multi-positive supervision from certified shortest trajectories and show strong NIST and out-of-distribution CASMI results (Sections 3.2–4).

2. Related work

MS2 spectrum-to-structure approaches The dominant practical approach is candidate retrieval against experimental spectral libraries such as NIST (NIST Mass Spectrometry Data Center, 2020), MassBank (Horai et al., 2010), and GNPS (Wang et al., 2016), or structural databases ranked by forward spectrum predictors such as CFM-ID (Allen et al., 2014), NEIMS (Wei et al., 2019), MassFormer (Young et al., 2024a), FraGNNet (Young et al., 2024b), and ICEBERG (Goldman et al., 2024); all are closed-world when the target is absent from the searched set. Fingerprint and substructure prediction methods such as CSI:FingerID (Dührkop et al., 2015), SIRIUS 4 (Dührkop et al., 2019), and MS2Prop (Voronov et al., 2022) extend coverage but remain dependent on structural databases. De novo methods escape fixed candidate sets by generating molecules directly from spectra, either as SMILES strings (MSNovelist (Stravs et al., 2022), Spec2Mol (Litsa et al., 2023), MS2Mol (Butler et al., 2023)) or molecular graphs (DiffMS (Bohde et al., 2025), FlowMS (Nie & Gao, 2026), MBGen (Sun et al., 2026)); however, to our knowledge, none maintains a validated current candidate repeatedly compared with the experimental spectrum through a forward model and locally edited from the resulting spectral evidence. Seeded approaches such as MADGEN (Wang et al., 2025) and SEISMiQ (Dorigatti et al., 2025) narrow search with a retrieved scaffold or known fragment, but the anchor is fixed rather than iteratively revised through forward-model guidance.

Molecular graph editing and 2-switch graph theory In molecular design, JT-VAE (Jin et al., 2018), HierVAE (Jin et al., 2020), and graph-to-graph translation networks (Do et al., 2019) learn to construct or modify molecules for property optimization; GEMS uses graph editing for structure identification, with each action a degree-preserving move on the isomer manifold driven by observed–predicted spectral correspondence. Degree-preserving 2-switches are classical for sampling graphs with fixed degree sequence (Senior, 1951; Hakimi, 1962), with connectedness-preserving variants (Taylor, 1981; Fernandes, 2022); shortest 2-switch paths relate to sorting by transpositions (Bafna & Pevzner, 1998), NP-hard in general, motivating the bounded oracle in Section 3.2. The molecular encoder of GEMS adopts Graphormer-style attention biases over molecular graphs (Ying et al., 2021).

Our method Forward spectrum models are typically used for candidate scoring, library expansion, or post-hoc reranking. GEMS instead places the forward model *inside* the inverse search loop: at each step the current molecule is scored by its observed–predicted spectral correspondence graph, and the policy uses this evidence to choose a local 2-switch edit or halt. This separates GEMS from retrieval methods, which never leave the candidate set, and from de novo decoders, which generally do not validate and repair intermediate molecular hypotheses with forward spectral guidance.

3. Method

3.1. Preliminaries

Tandem mass spectrometry In tandem mass spectrometry (MS2), a precursor ion is fragmented, yielding fragment m/z values and relative intensities. These patterns are informative but indirect structural signatures. We assume the precursor formula is known, limiting search to isomers.

Molecule representation We represent a molecule as a heavy-atom multigraph $\mathbf{A}$, with off-diagonal bond multiplicities in $\{0, 1, 2, 3\}$ and diagonal implicit hydrogen counts. These counts are equivalent to explicit degree-one hydrogen vertices compressed onto the heavy-atom diagonal. An *isomer* has the same formula, charge, and valence representation as the target; stereoisomers are not distinguished. A molecule is treated as chemically valid if it passes RDKit sanitization under this representation. Valid 2-switch edits preserve each aligned atom's total degree, formula, and total implicit hydrogens.

2-switch operations and the isomer manifold A **2-switch** is a local edit on $\mathbf{A}$ that preserves aligned atom valences and molecular formula. TYPEA switches two heavy–heavy unit bonds by deleting one unit from each bond and reconnecting the four endpoints in one of two pairings. TYPEB is the compressed form of switching a heavy–hydrogen unit bond with a heavy–heavy unit bond: a hydrogen is transferred from a donor heavy atom to one endpoint of the heavy–heavy bond, the old heavy–heavy bond is removed, and the donor is connected to the other endpoint. Candidate edits are retained only when bond multiplicities remain in $\{0, 1, 2, 3\}$, atom types and formal charges are unchanged, and the edited molecule remains a valid member of the same isomer manifold.

Definition 1 (Isomer manifold)**.** *For a molecular formula $\mathcal{F}$ with fixed charge and valence representation, the **isomer manifold** $\mathcal{G}_{\mathcal{F}} = (V_{\mathcal{F}}, E_{\mathcal{F}})$ has vertices $V_{\mathcal{F}}$ given by connected atom-colored molecular multigraphs with the prescribed atom multiset, formal charges, and element-preserving degree sequence, with implicit hydrogens compressed onto the diagonal. An edge $(m, m') \in E_{\mathcal{F}}$ exists iff m' is obtained from m by a valid 2-switch.*

Classical switch-chain results imply that graph realizations with a fixed degree sequence are connected by degree-preserving switches (Senior, 1951; Hakimi, 1962),

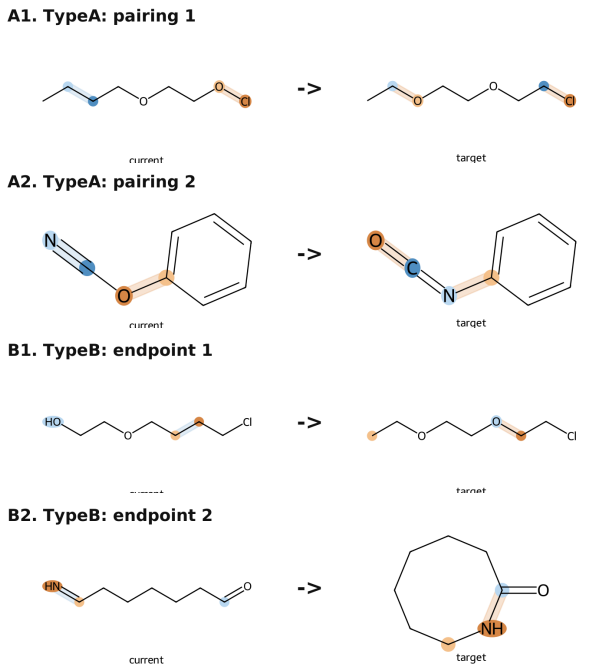

Figure 2. Two 2-switch operations. **(A1–A2)** TYPEA switches two heavy–heavy bonds, with the two panels showing the two possible endpoint pairings. **(B1–B2)** TYPEB transfers a hydrogen from a donor heavy atom to one endpoint of a heavy–heavy bond, removes that bond, and connects the donor to the other endpoint; the two panels show the two receiver choices. All displayed edits are certified one-step optimal actions from PubChem-grounded training states, with the post-edit molecule equal to the target. Highlighted atoms and bonds indicate the local changes.

with connectedness-preserving variants for connected graphs (Taylor, 1981; Fernandes, 2022). In the all-atom expansion, TYPEA corresponds to switching two heavy–heavy unit bonds, while TYPEB corresponds to switching one heavy–hydrogen and one heavy–heavy unit bond. Thus, abstract connected realizations with the same element-preserving degree sequence are connected by the compressed TYPEA/TYPEB moves, while the implemented GEMS state space further restricts moves by allowed bond orders, formula checks, and chemical validity.

Problem formulation Given an MS2 spectrum S, formula $\mathcal{F}$, initial candidate $m_0 \in V_{\mathcal{F}}$, and forward spectrum predictor $f : V_{\mathcal{F}} \times \Theta \to \mathcal{S}$, we seek a path $m_0 \to m_1 \to \cdots \to m_T = m^*$ on the isomer manifold $\mathcal{G}_{\mathcal{F}}$ whose endpoint best explains the observed spectrum under the forward model: $m^* \in \arg\min_{m \in V_{\mathcal{F}}} d\big(S, f(m;\Theta)\big)$. We learn a policy $\pi(a \mid m_t, \Gamma_t)$ over valid 2-switch actions $a \in \mathcal{A}(m_t)$, where Γ_t is the observed–predicted spectral correspondence graph relating S and $f(m_t;\Theta)$.

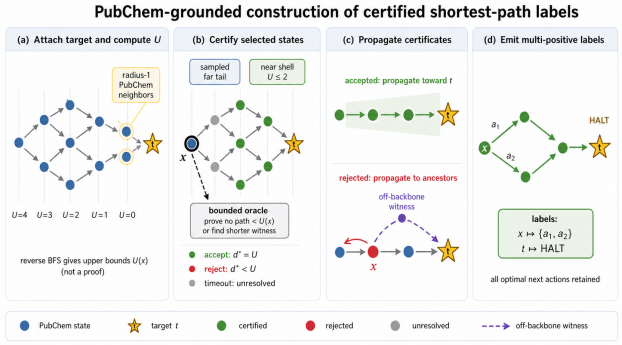

Figure 3. PubChem-grounded construction of certified shortest-path supervision. GEMS builds a PubChem isomer backbone, attaches each target through exact radius-1 PubChem neighbors, computes PubChem-distance upper bounds, and lazily certifies selected states with a bounded oracle. Retained states yield all optimal next 2-switch actions while the target emits HALT.

3.2. Geodesic Elucidation of Molecular Structures

GEMS has four components: (1) PubChem-grounded shortest-path supervision on an isomer 2-switch backbone (Section 3.2); (2) a graph-transformer policy with edge-token action scoring from the current molecule and spectral correspondence graph (Section 3.2); (3) behavior cloning with marginalized cross-entropy over all certified optimal next actions (Equation 2); and (4) multi-start autoregressive inference with visited-molecule reranking by forward cosine times halt probability (Section 3.2). The forward spectrum model is frozen: GEMS learns only $\pi(a \mid m_t, \Gamma_t)$, while the forward model constructs correspondence graphs and scores visited molecules at inference.

Data Curation. To obtain supervision, GEMS constructs PubChem-grounded shortest-path labels on the 2-switch isomer graph. For each formula, we build a reusable PubChem backbone whose nodes are PubChem isomers and whose edges are valid one-step 2-switches between PubChem molecules. For each target, we attach the target to this backbone through exact radius-1 PubChem neighbors and run reverse BFS, producing a PubChem-backbone distance upper bound $U(x)$ and a shortest-path DAG over PubChem states.

Because the full isomer manifold may contain non-PubChem shortcuts, $U(x)$ is not automatically exact. We certify retained states with a bounded oracle under the stated atom-alignment and validity model: a state is accepted if no path shorter than $U(x)$ exists, and rejected if the oracle finds a shorter off-backbone witness. Accepted certificates propagate toward the target because subpaths of shortest paths remain shortest; rejected certificates propagate to ancestors whose PubChem route depends on the rejected node.

The final corpus uses a sampled-tail certification policy. We retain the exact near shell, sample unresolved farther

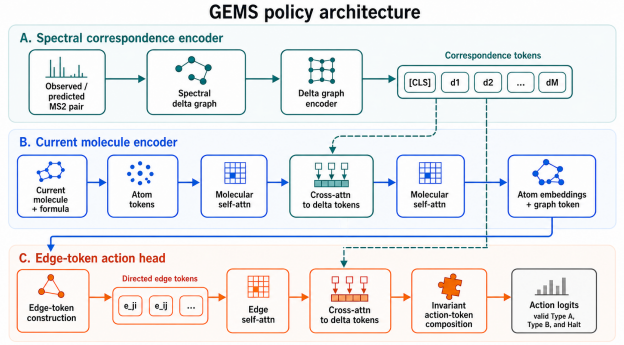

Figure 4. GEMS model architecture. A correspondence encoder embeds the observed–predicted spectral graph into correspondence tokens. The molecular encoder cross-attends to these tokens, while the edge-token action head first applies edge self-attention and then cross-attends to the same correspondence tokens before scoring valid TypeA, TypeB, and Halt actions.

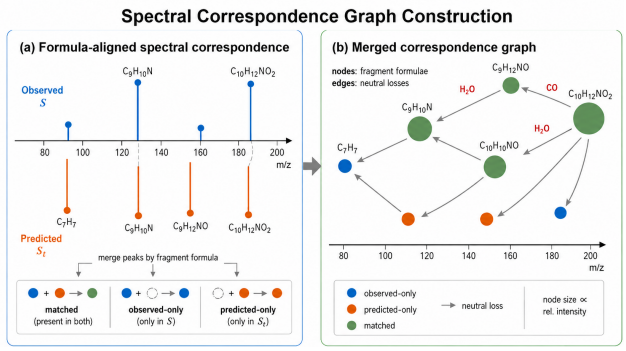

Figure 5. Spectral correspondence representation. Experimental and predicted MS2 neutral-loss graphs are merged by peak formula into a correspondence graph with matched, experimental-only, and predicted-only nodes.

PubChem-distance layers, halt after two consecutive all-timeout layers, and keep support spines connecting accepted far-tail states to the exact shell. Each retained state contributes all certified optimal next actions in the DAG, while the target contributes Halt. For every retained non-target state x, the label set contains one-step successors y certified to satisfy $d^*(x,t) = 1 + d^*(y,t)$ under this model. States not certified before timeout are not used. Additional data-generation details are given in Appendix A.1.

Framework. The model takes two inputs: (1) the current molecular graph $\mathbf{A}_t$ with atom features, and (2) the spectral correspondence graph Γ_t between experimental MS2 and the forward prediction on m_t. It outputs a distribution over the action space $\mathcal{A}(\mathbf{A}_t) \cup \{\text{Halt}\}$. The policy has three modules: a graph transformer over the spectral correspondence graph, a Graphormer-style molecular encoder whose molecular tokens cross-attend to correspondence tokens, and an edge-token action head that scores every enumerable TypeA and TypeB action plus Halt.

To build the correspondence graph, peaks in experimental and predicted spectra are assigned formulae, connected by neutral-loss edges, and merged by formula; nodes mark matched, experimental-only, and predicted-only peaks and carry intensity, m/z, formula, and neutral-loss features.

The molecular encoder embeds atom and pairwise graph features, adds a molecular-formula embedding to the graph `[CLS]` token, and integrates the correspondence tokens by cross-attention. The action head converts current bonds into directed edge tokens, refines them with self-attention and correspondence cross-attention, and composes invariant TypeA, TypeB, and Halt action tokens. For molecular edits, equivalent edge orders and directions are averaged before the scalar logit projection. Invalid or padded actions are masked to $-\infty$ and passed through SM:

$$\pi(a \mid m_t, \Gamma_t) = \mathrm{SM}\big([\ell_A^{1:n_A}, \ell_B^{1:n_B}, \ell_H]\big). \tag{1}$$

Here, $\ell_A^{1:n_A}$, $\ell_B^{1:n_B}$, and ℓ_H denote the TypeA, TypeB, and halt logits. Appendix A.2 gives the full feature inventory.

The model is trained with marginalized cross-entropy over all certified optimal next actions. Let $\mathcal{A}^+(m_t)$ denote the positive set defined by the oracle-certified shortest-path DAG.

$$\mathcal{L} = -\mathbb{E}_{(m_t,\Gamma_t)\sim\mathcal{D}}\Big[\log \sum_{a\in\mathcal{A}^+(m_t)} \pi(a \mid m_t, \Gamma_t)\Big], \tag{2}$$

This reduces to negative log-likelihood for a single positive action, but does not penalize another certified shortest-path successor. Training uses behavior cloning (Pomerleau, 1988; Bain & Sammut, 1995) on these examples.

Inference. At inference time, GEMS operates autoregressively. Starting from m_0, each step runs the forward model on m_t, constructs Γ_t, enumerates candidate actions, and scores them with the policy. The primary correspondence graph is recomputed at every step because each 2-switch changes the molecule and hence the forward model's prediction. All visited molecules are valid isomers. Algorithmic details are given in Appendix A.3.

For robustness, we use *multi-start search and re-ranking*: GEMS runs from multiple starting candidates, producing a diverse visited set. For each visited molecule, we record the forward-model spectral cosine and the model's Halt probability at that state. Post-hoc ranking deduplicates canonical molecules for each query spectrum and scores each candidate by

$$\mathrm{score}(m) = \mathrm{CosineSim}(S, f(m)) \cdot \pi(\text{Halt} \mid m, \Gamma(m)).$$

This ranking uses both spectral agreement and the inverse policy's estimate that the current state should stop. The product requires both spectral agreement and policy confidence:

a spectral match that the policy expects to edit further is ranked lower, as is a high-halt molecule with poor spectral agreement.

4. Experiments

4.1. Training setup

Training data. Training data use the PubChem-grounded sampled-tail procedure in Section 3.2 and Appendix A.1. Each retained state is paired with a spectral correspondence graph and a positive mask over all certified optimal actions. Retained states have certified shortest-path labels under the stated atom-alignment and validity model; sampled-tail selection of far-tail PubChem states is heuristic. We split examples by source trajectory group to avoid evaluating on duplicate steps from the same generated path.

Model. The architecture supports up to 40 heavy atoms and 500 correspondence-graph nodes. The model variant used in the main experiments uses the edge-token action head; embedding dimension, layer counts, attention heads, and dropout are treated as experimental hyperparameters.

Optimization. We train with AdamW (lr $= 10^{-4}$, weight decay 0.01), linear warmup with cosine decay, AMP, and gradient clipping at 1.0. The policy objective is marginalized cross-entropy over certified positive actions. Behavior cloning is run for one epoch.

Forward spectrum model. All GEMS runs use a frozen forward spectrum predictor. It takes the candidate structure, precursor type, and collision-energy features and outputs a positive-mode MS2 spectrum. The same forward model constructs correspondence graphs during training, rollout, and reranking; it is not updated during policy training.

4.2. Evaluation protocol and metrics

Benchmarks. We evaluate on two main structure-identification settings. **NIST** is the in-distribution benchmark: supervision is generated from NIST training spectra and evaluated online on held-out spectra whose target INCHIKEY14 connectivities are disjoint from the NIST training split. The PubChem isomer backbones used for oracle construction are external formula-level chemical-space graphs: they define same-formula states and 2-switch connectivity, but do not provide held-out NIST spectra. Here, common denotes the intersection of query spectra supported by all compared methods under our preprocessing. We report both a full common subset with 17,703 spectra and a 1,762-spectrum one-per-target-connectivity subset, denoted NIST unique, used for the 100-sample DiffMS comparison. **CASMI** contains 481 challenge spectra and tests generalization to compounds outside the NIST training distribution.

Exact-match metric. All top-k structure-identification results use INCHIKEY14 connectivity equality as the exact-match criterion. This compares the first 14 INCHIKEY characters, checking connectivity but not stereochemistry. A query is counted correct at top-k if the target connectivity appears among the first k ranked unique structures for that query.

Ranking protocol. For GEMS, each query is run from the stated number of isomer starting trajectories. The main protocol excludes the target from these starts. During inference we record every valid molecule visited by any trajectory, deduplicate canonical molecules, and rank them by the score in Section 3.2.

Baselines. DiffMS (Bohde et al., 2025), MADGEN (Wang et al., 2025), Spec2Mol (Litsa et al., 2023), FlowMS (Nie & Gao, 2026), and MolSpecFlow (Wang et al., 2026) rank generated candidates, with sample counts shown in each table. MADGEN rows are local reruns using the predictive scaffold retriever, and preprocessing drops count as failures. Spec2Mol uses a NIST-trained spectrum encoder with the pretrained decoder and top-100 decoding. FlowMS and MolSpecFlow are recent flow-based spectrum-to-molecule baselines. All baselines use the same query spectra and exact-match protocol as GEMS.

4.3. Results

Structure-identification results. Table 1 reports structure-identification accuracy under the protocol above. During inference, GEMS records every valid visited molecule, deduplicates canonical structures, and ranks them by the score in Section 3.2. On the NIST unique common subset, four-start GEMS obtains 30.48% top-1 and 59.02% top-10 accuracy, compared with DiffMS at 6.53% top-1 and 21.00% top-10 under the same formula-conditioned exact-match protocol. On the NIST full common subset, four-start GEMS obtains 32.08% top-1 and 56.64% top-10 over 17,703 spectra. With 16 starts, GEMS reaches the target in at least one trajectory for 72.65% of spectra and raises top-10 to 65.02%, while top-1 remains 30.62%, consistent with the navigation-versus-ranking pattern below.

CASMI is the strongest out-of-distribution test: without additional training, GEMS obtains 12.47% top-1 and 23.70% top-10 with four starts, and visits the target in 32.02% of spectra with up to 16 starts. These results position GEMS as seeded valid-molecule navigation that improves over direct generative baselines while retaining valid hypotheses throughout rollout. Full top-5/top-25 tables, ablations, and qualitative examples are in Appendix A.6.

More search reaches the target, but top-1 ranking saturates. To separate navigation from ranking, we vary the

Table 1. Structure-identification results. Query counts are NIST unique: 1,762; NIST full: 17,703; CASMI: 481. Top-k entries report INCHIKEY14 connectivity accuracy in percent; MCES is lower-is-better.

Dataset	Method	Budget	Any	Final	Top-1	Top-5	Top-10	Top-50	MCES@1	MCES@10
NIST unique	Spec2Mol	100 dec.	–	–	0.00	0.06	0.06	0.06	30.49	24.72
	MADGEN	50 samp.	–	–	2.21	2.67	3.41	9.82	9.12	7.14
	DiffMS	100 samp.	–	–	6.53	17.08	21.00	29.23	22.27	15.69
	FlowMS	100 samp.	–	–	7.45	18.38	23.92	31.78	20.12	13.38
	MolSpecFlow	100 samp.	–	–	8.22	19.17	25.08	32.41	19.61	12.42
	GEMS	4 starts	60.56	34.05	30.48	54.65	59.02	60.56	8.87	3.16
	GEMS	16 starts	72.19	40.81	29.17	54.82	63.39	71.91	9.54	3.05
NIST full	MADGEN	20 samp.	–	–	2.26	2.85	4.24	7.41	9.81	6.67
	Spec2Mol	100 dec.	–	–	0.01	0.03	0.05	0.10	29.76	24.58
	DiffMS	20 samp.	–	–	5.63	12.47	16.13	18.34	23.15	16.41
	FlowMS	20 samp.	–	–	6.72	13.14	17.42	19.63	22.34	15.28
	MolSpecFlow	20 samp.	–	–	7.14	13.85	18.03	20.25	21.73	14.59
	GEMS	4 starts	57.23	33.31	32.08	53.62	56.64	57.23	7.87	3.17
	GEMS	16 starts	72.65	42.26	30.62	56.76	65.02	72.51	9.14	2.98
CASMI	GEMS	4 starts	24.95	10.60	12.47	20.79	23.70	24.95	22.46	13.76
	GEMS	up to 16 starts	32.02	14.35	10.81	19.13	23.70	32.02	23.60	12.65
	DiffMS	100 samp.	–	–	0.42	2.29	2.70	3.95	39.76	32.33
	FlowMS	100 samp.	–	–	0.55	2.84	3.35	4.27	38.24	31.15
	MolSpecFlow	100 samp.	–	–	0.63	3.31	3.89	4.81	37.61	30.79
	Spec2Mol	100 dec.	–	–	0.00	0.00	0.00	0.00	43.60	29.29
	MADGEN	50 samp.	–	–	0.62	0.83	0.83	1.87	16.54	12.23

number of deterministic starts on the NIST unique common subset. Increasing K from 1 to 16 raises mean unique scored structures from 6.2 to 42.2, target reached in at least one trajectory from 45.57% to 72.19%, and final-target success from 23.67% to 40.81%. Top-10 accuracy also increases from 45.46% to 63.39%. In contrast, top-1 ends at 29.17% for $K = 16$ and ranges only from 28.94% to 30.48% across the entire sweep. Thus, additional search increasingly finds the target, but selecting it from the enlarged visited pool becomes limiting (Figure 6).

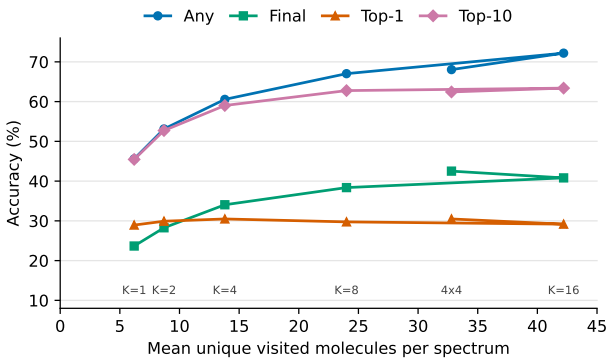

Figure 6. Increasing search improves the fraction of spectra where at least one trajectory reaches the target, final-target success, and top-10 accuracy on NIST unique, but top-1 saturates. Larger search shifts the bottleneck from reaching the target to ranking the enlarged visited set.

Halt probability, not forward cosine alone, drives visited-molecule ranking. We decompose the visited-molecule score into its forward-cosine and policy-halt components (Table 2). On NIST unique with four starts, cosine-only ranking gives 16.00% top-1, halt-only ranking gives 30.36%, and the product score used by GEMS gives 30.48%. With 16 starts, the corresponding values are 13.96%, 28.77%, and 29.17%. Conditioned on the target being reached, cosine-only ranking selects it at rank 1 in only 26.43% of four-start cases and 19.34% of 16-start cases. The highest-cosine visited molecule also has higher mean forward cosine than the target: 0.802 versus 0.715 for four starts, and 0.816 versus 0.709 for 16 starts. Expanding the visited pool exposes high-cosine decoys, so forward-cosine alone is poorly calibrated for final identification among GEMS candidates.

The learned halt probability carries most target-selection signal, and the small product-score gain suggests reranking calibration is central once navigation reaches the target.

A natural same-seed alternative is to replace the learned policy with greedy or beam search over valid 2-switch neighbors, using the same retrieved seed and frozen forward model. This isolates the role of the learned proposal policy, but it is computationally unfavorable in our setting: unlike GEMS, which evaluates the forward model only on

Table 2. Visited-molecule reranking decomposition on the NIST unique common subset. Product denotes the default score, forward cosine times halt probability.

K	Rule	Top-1	Top-10	Mean rank if visited	Rank-1 if visited
4	cosine only	16.00	57.32	4.02	26.43
4	halt only	30.36	58.57	2.71	50.14
4	product	30.48	59.02	2.61	50.33
16	cosine only	13.96	52.50	8.56	19.34
16	halt only	28.77	62.94	5.24	39.86
16	product	29.17	63.39	4.85	40.41

Table 3. Spectral-correspondence ablations on the NIST unique common subset with $K = 4$. All values are percentages.

Variant	Any	Final	Top-1	Top-5	Top-10
Full GEMS	60.56	34.05	30.48	54.65	59.02
Recomputed corr.	60.44	34.05	30.42	54.71	58.97
Stale initial corr.	60.84	34.05	26.45	51.87	58.68
No intensity/source	50.91	17.20	15.83	39.44	47.56

the current state before scoring all actions with the policy, exhaustive local search must instantiate many neighboring molecules and run the forward model on each candidate neighbor. Since the TYPEA proposal space grows with pairs of current bonds, this can require tens to hundreds of additional forward-model calls per step and becomes especially costly for larger action spaces. It also does not remove the main ranking failure mode observed in Table 2: high-cosine decoys can outrank the true target under the frozen forward model. Thus, exhaustive cosine-greedy or beam search is a useful diagnostic but not a scalable substitute for learned forward-guided navigation.

Spectral correspondence supports termination and ranking. We next ablate the correspondence graph used by the policy (Table 3). A recomputed-correspondence control at four starts gives 30.42% top-1 and 58.97% top-10, matching the default run. Reusing the start-state correspondence graph leaves target-reaching and final success nearly unchanged, but reduces top-1 to 26.45%. Removing correspondence intensities and source labels is much more damaging: top-1 drops to 15.83%, top-10 to 47.56%, and final success to 17.20%. GEMS can still navigate locally from molecular context and correspondence topology, but accurate halting and top ranking depend strongly on current correspondence features. Because this ablation removes intensity magnitudes and source labels together, channel-isolated controls are deferred to Appendix A.7; there, source labels retain much of the signal, labels plus experimental intensity stay near the full control, and zeroing both is far more destructive. Appendix A.7 also reports fixed-budget sampling and support-aware reranking.

Post-hoc multi-trajectory support suggests reranking headroom. Finally, we test whether agreement across trajectories can improve post-hoc ranking. For each unique structure in the default 16-start NIST run, we count distinct rollout trajectories that visit it and multiply the product score by $\log_2(1 + n_{\text{rollouts}})$. This post-hoc support-aware score improves NIST unique top-1 from 29.17% to 32.92%, top-5 from 54.82% to 57.04%, and top-10 from 63.39% to 64.93%. Count alone is insufficient on NIST: rollout-count-only ranking gives 28.49% top-1 and occurrence-count-only ranking gives 21.00%. On CASMI 16-start, support-aware scoring also improves top-1 from 10.81% to 12.27% and top-10 from 23.70% to 24.74%, showing directional transfer. Appendix A.7 shows that occurrence support and rollout-count ranking are also competitive on CASMI, suggesting repeated discovery may be informative out of distribution. Because CASMI has only 481 spectra and this analysis is post-hoc, we report support as reranking headroom rather than the default GEMS score.

5. Discussion and limitations

The diagnostic ablations reveal a division of labor between search and scoring. As starts increase, GEMS visits and halts on the target more often, but top-1 accuracy is limited by calibration over the enlarged valid-molecule set: on NIST unique, high-cosine decoys often outrank the target. Halt probability partially corrects this, and post-hoc multi-trajectory support suggests that repeated discovery can add confidence. Future work should therefore prioritize calibrated reranking–fine-tuning the forward-cosine component, incorporating forward-model uncertainty, and learning multi-trajectory support–rather than increasing rollout count alone.

A natural same-seed alternative is greedy or beam search over valid 2-switch neighbors using the same retrieved initialization and frozen forward model. This isolates the role of the learned proposal policy, but it is computationally unfavorable in this setting. Unlike GEMS, which evaluates the forward model on the current state and then scores valid edits with the learned policy, exhaustive local search must instantiate and forward-score many neighboring molecules at each step. Since the TYPEA proposal space grows with pairs of current bonds, this can require many additional forward-model calls and becomes increasingly expensive as the local action space grows. Moreover, cosine-driven local search is vulnerable to the same high-cosine decoys exposed by our reranking analysis; beam search can amplify this issue by expanding many plausible but incorrect neighbors. Thus, exhaustive cosine-greedy or beam search is a useful diagnostic, but not a performant and scalable replacement for learned forward-guided navigation.

GEMS produces inspectable, chemically valid trajectories

by combining local valence-preserving edits with post-edit validity and isomer checks. At inference, its main dependency is the frozen forward model, which shapes the correspondence graph and reranking score. The PubChem-grounded oracle is used only offline for supervision construction, not during rollout or ranking.

The experiments consider positive-mode MS2 and molecules up to 40 heavy atoms; stereoisomers are not distinguished. At inference, the action space grows with candidate bond pairs, which may be expensive for larger molecules. The offline sampled-tail construction limits certification cost by excluding unresolved states during supervision construction. Despite these limitations, forward-guided local navigation substantially improves formula-conditioned identification over direct generative baselines, while exposing reranking calibration as the main bottleneck for converting target-reaching into top-1 accuracy.

References

Allen, F., Pon, A., Wilson, M., Greiner, R., and Wishart, D. S. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Research*, 42(W1):W94–W99, 2014.

Bafna, V. and Pevzner, P. A. Sorting by transpositions. *SIAM Journal on Discrete Mathematics*, 11(2):224–240, 1998.

Bain, M. and Sammut, C. A framework for behavioural cloning. In *Machine intelligence 15*, pp. 103–129, 1995.

Bohde, M., Manjrekar, M., Wang, R., Ji, S., and Coley, C. W. DiffMS: Diffusion generation of molecules conditioned on mass spectra. *arXiv preprint arXiv:2502.09571*, 2025.

Butler, T., Frandsen, A., Lightheart, R., et al. MS2Mol: A transformer model for illuminating dark chemical space from mass spectra. *ChemRxiv*, 2023.

da Silva, R. R., Dorrestein, P. C., and Quinn, R. A. Illuminating the dark matter in metabolomics. *Proceedings of the National Academy of Sciences*, 112(41):12549–12550, 2015.

Do, K., Tran, T., and Venkatesh, S. Graph transformation policy network for chemical reaction prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, 2019.

Dorigatti, E., Groß, J., Kühlborn, J., Möckel, R., Maier, F., and Keupp, J. Enhancing automated drug substance impurity structure elucidation from tandem mass spectra through transfer learning and domain knowledge. *Digital Discovery*, 4:2454–2464, 2025.

Dührkop, K., Shen, H., Meusel, M., Rousu, J., and Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Sciences*, 112(41):12580–12585, 2015.

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., Dorrestein, P. C., Rousu, J., and Böcker, S. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, 16(4):299–302, 2019.

Fernandes, R. On the switch-length of two connected graphs with the same degree sequence. *Australasian Journal of Combinatorics*, 83(1):87–100, 2022.

Goldman, S., Li, J., and Coley, C. W. Generating molecular fragmentation graphs with autoregressive neural networks. *Analytical Chemistry*, 96(8):3419–3428, 2024.

Hakimi, S. L. On realizability of a set of integers as degrees of the vertices of a linear graph. I. *Journal of the Society for Industrial and Applied Mathematics*, 10(3):496–506, 1962.

Horai, H., Arita, M., Kanaya, S., et al. MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, 45(7):703–714, 2010.

Jin, W., Barzilay, R., and Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. In *International Conference on Machine Learning (ICML)*, 2018.

Jin, W., Barzilay, R., and Jaakkola, T. Hierarchical generation of molecular graphs using structural motifs. In *International Conference on Machine Learning (ICML)*, 2020.

Litsa, E. E., Chenthamarakshan, V., Das, P., and Kavraki, L. E. An end-to-end deep learning framework for translating mass spectra to de-novo molecules. *Communications Chemistry*, 6:132, 2023.

Nie, J. and Gao, P. FlowMS: Flow matching for de novo structure elucidation from mass spectra. *arXiv preprint arXiv:2603.18397*, 2026.

NIST Mass Spectrometry Data Center. NIST/EPA/NIH mass spectral library. National Institute of Standards and Technology, 2020.

Pomerleau, D. A. Alvinn: An autonomous land vehicle in a neural network. *Advances in neural information processing systems*, 1, 1988.

Senior, J. K. Partitions and their representative graphs. *American Journal of Mathematics*, 73(3):663–689, 1951.

Stravs, M. A., Dührkop, K., Böcker, S., and Zamboni, N. MSNovelist: de novo structure generation from mass spectra. *Nature Methods*, 19(7):865–870, 2022.

Sun, X., Wei, W., Rao, J., Xie, J., and Yang, Y. De novo molecular generation from mass spectra via many-body enhanced diffusion. *arXiv preprint arXiv:2602.01643*, 2026.

Taylor, R. Constrained switchings in graphs. In *Combinatorial Mathematics VIII*, volume 884 of *Lecture Notes in Mathematics*, pp. 314–336. Springer, 1981.

Voronov, G., Lightheart, R., Davison, J., Krettler, C. A., Healey, D., and Butler, T. MS2Prop: A machine learning model that directly predicts chemical properties from mass spectrometry data for novel compounds. *bioRxiv*, 2022. doi: 10.1101/2022.10.09.511482.

Wang, M., Carver, J. J., Phelan, V. V., et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, 34(8):828–837, 2016.

Wang, Y., Chen, X., Liu, L., and Hassoun, S. MADGEN: Mass-Spec attends to de novo molecular generation. In *International Conference on Learning Representations (ICLR)*, 2025.

Wang, Y., Yang, F., Xu, K., Yuan, L., Zhu, J., Zhang, J., Tang, Z., Bian, Y., Chang, C., Tian, Y., et al. Molspecflow: Mass-constrained hybrid flow matching for joint molecular-spectral analysis. *bioRxiv*, pp. 2026–01, 2026.

Wei, J. N., Belanger, D., Adams, R. P., and Sculley, D. Rapid prediction of electron-ionization mass spectrometry using neural networks. *ACS Central Science*, 5(4):700–708, 2019.

Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., and Liu, T.-Y. Do Transformers really perform badly for graph representation? In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Young, A., Röst, H., and Wang, B. Tandem mass spectrum prediction for small molecules using graph transformers. *Nature Machine Intelligence*, 6:404–416, 2024a.

Young, A., Wang, F., Wishart, D., Wang, B., Röst, H., and Greiner, R. FraGNNet: A deep probabilistic model for tandem mass spectrum prediction. *arXiv preprint arXiv:2404.02360*, 2024b.

A. Additional details

A.1. PubChem-grounded data generation

Formula-level PubChem backbone. For each molecular formula in the training set, we collect all PubChem isomers and expand each by enumerating its valid 2-switch neighbors. We keep only one-step edges whose endpoints are both PubChem molecules, yielding a reusable formula-level PubChem backbone. This graph is not the full isomer manifold: it deliberately excludes non-PubChem states except for target-local attachment states used during certification.

Target attachment and PubChem distances. For each target molecule, we build the target-centered exact radius-1 2-switch ball and attach the target to PubChem backbone nodes that occur among those radius-1 neighbors. A reverse BFS from the target over this attached backbone gives a PubChem-backbone distance $U(x)$ for each reachable PubChem node x. The directed shortest-path DAG contains edges $x \to y$ with $U(y) = U(x) - 1$. Because the full isomer manifold may contain non-PubChem shortcuts, $U(x)$ is only an upper bound on the true shortest 2-switch distance $d^*(x, t)$ until certified.

Bounded certification oracle: atom mapping via ILP. When a PubChem-backbone bound must be certified, the oracle compares a source molecule m_s to the target molecule m_t and first establishes a correspondence between their atoms. We solve an integer linear program (ILP) to find a bijection $\sigma : [n] \to [n]$ between the heavy atoms that maximizes edge overlap:

$$\max_{\sigma} \sum_{i<j} \min\big(\mathbf{A}^{(s)}_{ij},\ \mathbf{A}^{(t)}_{\sigma(i)\sigma(j)}\big) + \sum_{i} \min\big(\mathbf{A}^{(s)}_{ii},\ \mathbf{A}^{(t)}_{\sigma(i)\sigma(i)}\big), \tag{3}$$

subject to σ being a permutation that maps each atom to one of the same element. The first term maximizes heavy-atom bond overlap; the second maximizes hydrogen count overlap. Multiple optimal mappings may achieve the same ILP objective but lead to different 2-switch distances. We use Gurobi's solution pool to enumerate a bounded pool of optimal alignments (`PoolGap` $= 0$; 20 solutions in the current data-generation runs), deduplicate alignments that induce the same bounded-search root, and search the unique alignments whose admissible lower bound is below the PubChem bound. Consequently, accepted labels are alignment-pool-certified rather than certified over every possible element-preserving atom correspondence when the number of optimal alignments exceeds the pool limit.

Bounded shortest-path search via bidirectional A*. Given the aligned adjacency matrices $\mathbf{A}^{(s)}$ and $\mathbf{A}^{(t)}$ under the chosen mapping, we search for a 2-switch path using

bidirectional A* with the admissible heuristic

$$h(\mathbf{A}) = \max\Big(h_{\mathrm{H}},\ \Big\lceil \frac{h_{\text{edge}} + h_{\mathrm{H}}}{2} \Big\rceil\Big), \tag{4}$$

where $h_{\text{edge}} = \sum_{i<j} \max(0, \mathbf{A}_{ij} - \mathbf{A}^{(t)}_{ij})$ counts excess heavy-atom bond multiplicities and $h_{\mathrm{H}} = \sum_i \max(0, \mathbf{A}_{ii} - \mathbf{A}^{(t)}_{ii})$ counts excess hydrogen assignments. In sampled-tail certification, the search is bounded by the PubChem-backbone distance $U(x)$: finding a path of length $< U(x)$ rejects the backbone certificate, while proving that no such path exists certifies $U(x)$ as the shortest distance under the searched alignment pool and validity model.

Sampled-tail certification. We first retain exact near-shell labels for $U(x) \leq 2$. The data-generation procedure then probes the tail by PubChem-distance layer rather than drawing one global random pool: for each layer $d \in \{4, 5, \ldots, 10\}$, it selects up to 10 unresolved PubChem nodes with $U(x) = d$ and runs bounded-oracle certification with a short per-node timeout. Tail nodes are selected by unresolved ancestor coverage in the current shortest-path DAG, so each successful certificate can support many upstream examples. If every selected node in a layer times out, that layer is marked as a complete timeout layer; after two consecutive complete timeout layers, the procedure stops probing farther layers. For a selected node x, the oracle either proves that no path shorter than $U(x)$ exists within the searched alignment pool and validity model, in which case $d^*(x, t) = U(x)$ under that model, or finds an explicit shorter witness path. If $d^*(x, t) = U(x)$, then any path of length $U(x)$ from x to the target is optimal, and every subpath on such a path is also optimal; we therefore propagate accepted exact distances downward along the shortest-path DAG. If the oracle finds a shorter witness for x, then any ancestor whose PubChem-bound route through x is strictly beaten by that witness also loses its certification, and this rejection is propagated upward along the DAG. Accepted far-tail nodes retain one deterministic support spine down to the already exact near shell; this can pull intermediate layers such as $U = 3$ into the retained trajectory support even though the scheduled probes start at $d = 4$. The resulting supervision is certified for retained accepted examples under the searched alignment pool and validity model, while corpus construction remains heuristic because unscheduled or timed-out far-tail PubChem nodes are left unresolved.

Training example extraction. Each certified state contributes all valid one-step actions that move to a certified DAG child at distance one smaller, rather than a single arbitrarily chosen path action. The target state contributes a HALT action. Thus each example contains a positive action set, used by the marginalized training objective in Equation 2.

A.2. Architecture details

Spectral correspondence graph features. Given the experimental MS2 spectrum and the forward model's prediction on the current candidate m_t, molecular formulae are assigned to each peak by constrained decomposition of the observed m/z values against the precursor formula. Peaks are then connected by directed neutral-loss edges: an edge from peak u to peak v exists if v's formula is a sub-formula of u's, with the edge labeled by the loss formula $u - v$. This is done separately for the experimental and predicted spectra, producing two neutral-loss graphs. These graphs are merged into the spectral correspondence graph Γ_t: peaks with matching molecular formulae are unified into single nodes, and unmatched peaks become separate nodes. Each node carries source type $s \in \{$experimental-only, predicted-only, both$\}$, m/z value, experimental intensity, predicted intensity, and an 11-element formula vector over C, H, N, O, S, P, F, Cl, Br, I, and Si. Edge features are neutral-loss formula vectors over the same elements.

Molecular features. Each heavy atom is described by categorical features for atomic number, hydrogen count, formal charge, total valence, ring membership, degree, and precursor type, plus continuous collision energy. Categorical features are embedded via lookup tables and summed; continuous features are linearly projected. The attention bias in molecular self-attention incorporates shortest-path distance, bond type, ring membership for bonded pairs, and effective resistance distance.

Global representation. After the final molecular self-attention phase, we extract the `[CLS]` token embedding $\mathbf{h}_{\texttt{CLS}}$ and compute three pooled representations over atom embeddings $\{\mathbf{h}_i\}$:

$$\mathbf{h}_{\text{global}} = \big[\mathbf{h}_{\texttt{CLS}} \parallel \alpha \cdot \textstyle\sum_i \mathbf{h}_i \parallel \mathrm{mean}_i(\mathbf{h}_i) \parallel \max_i(\mathbf{h}_i)\big] \in \mathbb{R}^{4d}, \tag{5}$$

where $\alpha = 0.01$ is a fixed scaling factor that prevents the sum from dominating. This pooled vector is used by the HALT token, while candidate edit tokens use $\mathbf{h}_{\texttt{CLS}}$ as graph-level context.

Directed edge tokens. For each directed current bond $i \to j$, we embed bond features and form a directed edge token

$$\mathbf{e}^{(0)}_{i\to j} = \mathrm{MLP}_{\text{edge}}\big([\mathbf{h}_i \parallel \mathbf{h}_j \parallel \mathbf{e}^{\text{bond}}_{ij} \parallel \mathbf{h}_{\texttt{CLS}}]\big) \in \mathbb{R}^d. \tag{6}$$

All directed edge tokens for the molecule are updated by two self-attention layers. These edge tokens also cross-attend to the correspondence tokens $\mathbf{z}_\Gamma$ before action scoring.

TYPEA edit tokens. A TYPEA(u, v, x, y) action removes two existing bonds and reconnects their endpoints. Because

the removed bonds are the local edit context, the head gathers the refined directed edge tokens for both orientations of both bonds: $u \to v$, $v \to u$, $x \to y$, and $y \to x$. It then composes the two bond tokens with a TYPEA action-type embedding and $\mathbf{h}_{\texttt{CLS}}$ through a shared pair MLP. The implementation averages the four equivalent presentations obtained by swapping the two bonds and reversing both directed bonds, giving an order-invariant action token $\mathbf{t}_A(u, v, x, y)$.

TYPEB edit tokens. A TYPEB action moves one hydrogen from a donor atom to a recipient endpoint of an existing heavy–heavy bond, removes that bond, and connects the donor to the other endpoint. The head first constructs a transfer token from the donor embedding, recipient embedding, and $\mathbf{h}_{\texttt{CLS}}$. It combines that transfer token with the directed token for the affected recipient-adjacent bond, averaging the two edge directions to obtain an action token $\mathbf{t}_B$.

HALT token. The HALT action is represented by an MLP over the pooled global representation and a HALT action-type embedding:

$$\mathbf{t}_{\text{HALT}} = \text{MLP}_{\text{HALT}}\big([\mathbf{h}_{\text{global}} \,\|\, \mathbf{e}^{\text{type}}_{\text{HALT}}]\big) \in \mathbb{R}^d. \tag{7}$$

All TYPEA, TYPEB, and HALT tokens share a final scalar projection to logits.

A.3. Inference details

Deferred validity checking. The policy scores the action set, but inference does not necessarily apply the top-scoring edit. It scans actions in descending policy score and accepts the first action whose application yields a valid isomer within the stated molecular-size bound. If the HALT action is selected, the current molecule is recorded as the final state; if no ranked edit is valid, the trajectory terminates without adding an invalid molecule. Thus all candidate molecules are validated before they enter the ranked set.

A.4. Action enumeration

For a molecule with n heavy atoms and E edges (counting bond multiplicities), the number of valid Type A actions is $O(E^2)$—each pair of edges that can be "swapped" without creating duplicate edges. TYPEB actions choose a hydrogen donor h_d, a heavy–heavy edge (e_s, e_e) not incident to h_d, and one endpoint of that edge as the hydrogen receiver, yielding $O(n \cdot E)$ candidates. In practice, the action space ranges from tens to low thousands of actions for molecules with 10–25 heavy atoms.

A.5. Heuristic admissibility

Let $d^*(\mathbf{A}, \mathbf{A}^{(t)})$ be the shortest valid 2-switch distance to the target under a fixed atom alignment. Each valid TYPEA or TYPEB action can reduce the hydrogen-excess count by at most one, so $d^*(\mathbf{A}, \mathbf{A}^{(t)}) \geq h_{\text{H}}$. Each action can reduce the combined edge-plus-hydrogen excess by at most two, so $d^*(\mathbf{A}, \mathbf{A}^{(t)}) \geq \lceil (h_{\text{edge}} + h_{\text{H}})/2 \rceil$. Therefore their maximum is also a lower bound on the remaining distance and is admissible.

Algorithm 1 GEMS Inference

Require: Experimental spectrum S, starting candidates $\{m_0^{(1)}, \ldots, m_0^{(K)}\}$, forward model f, policy π, max steps T
Ensure: Ranked list of identified structures

$\mathcal{R} \leftarrow \emptyset$ {Visited valid molecules and scores}
for $k = 1, \ldots, K$ **do**
  $m \leftarrow \text{ValidateStart}(m_0^{(k)})$
  for $t = 0, 1, \ldots, T-1$ **do**
    $\hat{S}_t \leftarrow f(m)$ {Forward model prediction}
    $\Gamma_t \leftarrow \text{BuildCorrespondenceGraph}(S, \hat{S}_t)$ {Spectral correspondence}
    $\mathcal{A}_t \leftarrow \text{EnumerateActions}(m) \cup \{\text{HALT}\}$
    $\mathbf{p}_t \leftarrow \pi(\cdot \mid m, \Gamma_t, \mathcal{A}_t)$
    $\mathcal{R} \leftarrow \mathcal{R} \cup \{(m, \text{CosineSim}(S, \hat{S}_t) \cdot p_t(\text{HALT}))\}$
    ranked $\leftarrow \text{SortByScore}(\mathcal{A}_t, \mathbf{p}_t)$
    $m' \leftarrow \perp$
    for a in ranked **do**
      if $a = \text{HALT}$ **then**
        break
      end if
      $\tilde{m} \leftarrow \text{Apply}(m, a)$
      if $\tilde{m}$ is a valid isomer **then**
        $m' \leftarrow \tilde{m}$
        break
      end if
    end for
    if $m' = \perp$ **then**
      break
    end if
    $m \leftarrow m'$
  end for
end for
return unique canonical molecules in $\mathcal{R}$ ranked by maximum recorded score

A.6. Additional NIST and CASMI results

Table 4. NIST structure-identification results with full top-k columns. Full DiffMS and full MADGEN top-25/top-50 equal their best generated sample budgets when fewer than 50 samples were generated.

Method	Spectra	Samples/Traj.	Top-1	Top-5	Top-10	Top-25	Top-50
GEMS, unique	1,762	4 traj.	30.48%	54.65%	59.02%	60.50%	60.56%
GEMS, unique	1,762	16 traj.	29.17%	54.82%	63.39%	70.66%	71.91%
DiffMS, unique	1,762	100 samples	6.53%	17.08%	21.00%	25.77%	29.23%
MADGEN, unique	1,762	50 samples	2.21%	2.67%	3.41%	5.68%	9.82%
Spec2Mol, unique	1,762	100 decodes	0.00%	0.06%	0.06%	0.06%	0.06%
GEMS, full	17,703	4 traj.	32.08%	53.62%	56.64%	57.22%	57.23%
GEMS, full	17,703	16 traj.	30.62%	56.76%	65.02%	71.14%	72.51%
DiffMS, full	17,703	20 samples	5.63%	12.47%	16.13%	18.34%	18.34%
MADGEN, full	17,703	20 samples	2.26%	2.85%	4.24%	7.41%	7.41%
Spec2Mol, full	17,703	100 decodes	0.01%	0.03%	0.05%	0.07%	0.10%

Table 5. CASMI structure-identification results with full top-k columns.

Method	Spectra	Samples/Traj.	Top-1	Top-5	Top-10	Top-25	Top-50
GEMS	481	4 traj.	12.47%	20.79%	23.70%	24.95%	24.95%
GEMS	481	up to 16 traj.	10.81%	19.13%	23.70%	29.73%	32.02%
DiffMS	481	100 samples	0.42%	2.29%	2.70%	3.95%	3.95%
Spec2Mol	481	100 decodes	0.00%	0.00%	0.00%	0.00%	0.00%

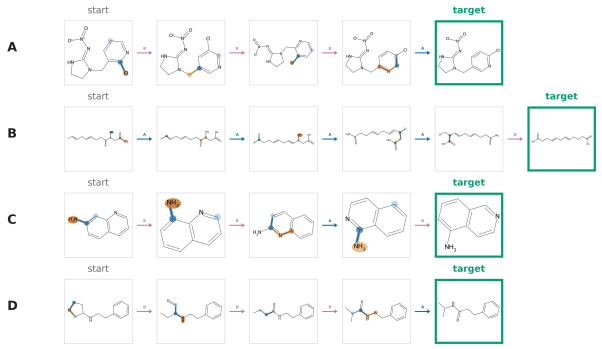

Figure 7. Autoregressive recovery examples on CASMI. Each row shows a model rollout from an isomer starting candidate to the first target-connectivity hit; colored atoms mark the selected 2-switch action and arrow colors denote TypeA or TypeB moves. These trajectories are not oracle-shortest under the alignment and validity model used by our oracle: rows A–D have shortest-path distances $1, 3, 2, 3$, while the learned policy reaches the target in $4, 5, 4, 4$ edits, respectively. The examples illustrate empirical recovery from valid but non-optimal intermediate edits.

A.7. Additional diagnostic ablations

Table 6. Spectral-correspondence channel isolation on the NIST unique common subset with 16 starts. All variants use the same request set and product reranking. The source labels distinguish experimental-only, common/matched, and predicted-only peaks.

Variant	Any	Final	Top-1	Top-10	Mean unique	Interpretation
Full recomputed correspondence	72.02	40.86	29.11	63.39	42.2	Full online correspondence control
Stale initial correspondence	72.64	43.98	26.22	60.78	43.7	Reuses the start-state correspondence graph
Zero intensity/source	64.76	24.69	13.34	46.82	47.1	Removes labels and intensity magnitudes
Scale intensity max to 1	69.18	43.93	28.04	61.63	37.4	Preserves labels but changes absolute scale
Source labels only	69.30	43.70	27.92	61.75	37.4	Labels alone retain much of the signal
Labels + experimental intensity	71.91	38.59	28.15	61.69	43.5	Observed-side intensity preserves coverage
Labels + predicted intensity	71.11	36.04	23.95	58.34	44.1	Predicted-side intensity is weakest for top-1

Table 7. Fixed eight-rollout exploration/exploitation sweep on the NIST unique common subset. Sampling rows use action temperature 0.5 and action top-$k = 5$.

Variant	Starts	Rollouts/start	Any	Final	Top-1	Top-5	Top-10	Mean unique
sampled 1×8	1	8	51.36	33.09	29.57	46.88	49.60	14.4
sampled 2×4	2	4	57.89	36.95	30.31	51.31	55.62	17.8
sampled 4×2	4	2	64.07	38.93	30.02	54.03	60.16	23.3
greedy K=8	8	1	67.03	38.37	29.74	55.51	62.77	24.0

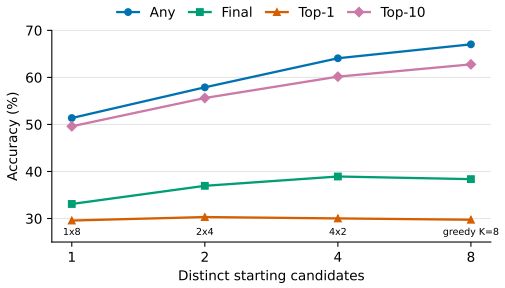

Figure 8. Fixed eight-rollout sampling sweep on the NIST unique common subset. Allocating the same rollout budget across more starts improves target-reaching and top-10 accuracy, while top-1 remains nearly flat.

Table 8. Post-hoc support-aware reranking on recorded traces. Product denotes forward cosine times halt probability. Support counts are computed after inference and were not used by the policy.

Dataset	Reranker	Top-1	Top-5	Top-10	Mean rank if visited	Median rank if visited
NIST unique K16	product	29.17	54.82	63.39	4.85	2
NIST unique K16	rollout-count only	28.49	52.84	62.32	5.16	2
NIST unique K16	occurrence-count only	21.00	40.81	53.23	7.85	4
NIST unique K16	product × $\sqrt{n_{\text{rollouts}}}$	32.92	56.64	64.70	4.43	2
NIST unique K16	product × $\log_2(1 + n_{\text{rollouts}})$	32.92	57.04	64.93	4.38	2
NIST unique K16	product × $\log_2(1 + n_{\text{occurrences}})$	32.35	56.58	64.36	4.49	2
CASMI K4	product	12.47	20.79	23.70	3.11	–
CASMI K4	product × $\log_2(1 + n_{\text{rollouts}})$	12.47	20.79	23.70	2.98	–
CASMI K4	product × $\log_2(1 + n_{\text{occurrences}})$	13.10	20.58	23.70	–	–
CASMI K16	product	10.81	19.13	23.70	7.71	–
CASMI K16	rollout-count only	12.47	20.79	24.95	7.77	–
CASMI K16	product × $\log_2(1 + n_{\text{rollouts}})$	12.27	20.58	24.74	6.78	–
CASMI K16	product × $\log_2(1 + n_{\text{occurrences}})$	12.68	20.58	24.74	–	–

A.8. Inference timing and model size

The main inverse-policy checkpoint contains 16.7M parameters; inference also uses a frozen 51.6M-parameter forward spectrum predictor that maps candidate structures and acquisition metadata to predicted MS2 spectra, which GEMS then uses to construct correspondence graphs and compute reranking cosines. For comparison, the NIST DiffMS model has approximately 76.7M parameters, and the Spec2Mol model used here has 52.1M parameters. Table 9 reports observed training wall-clock times and single-GPU-equivalent inference times, except where CPU workers are explicitly stated.

On the NIST unique comparison, GEMS K4 requires about 9.4× less single-GPU-equivalent inference time than the 100-sample DiffMS run, while GEMS K16 requires about 5.0× less. Spec2Mol is computationally lighter than both graph-generation methods, but its exact-match accuracy is substantially lower in Table 1.

Table 9. Model sizes and observed timing. Training rows report observed wall-clock hardware. Inference rows report single-GPU-equivalent time, except for the CPU MCES pass.

Method / task	Model size	Workload	Time
GEMS policy training	16.7M policy	1 epoch	6h05m, 4 H100
DiffMS NIST training	76.7M	1 / 20 / 75 epochs	9–11 min / 3–4 h / 12–14 h A100
GEMS K4, no forward cache	16.7M policy + 51.6M forward	1,762 spectra, 7,035 rollouts	5.2 GPU-h
GEMS K16, forward cache	16.7M policy + 51.6M forward	1,762 spectra, 27,909 rollouts	9.8 GPU-h
DiffMS NIST eval, 100 samples	76.7M	1,762 spectra	48.9 A100 GPU-h
DiffMS full NIST eval, 1 sample	76.7M	full NIST	4h45m A100
DiffMS full NIST eval, 20 samples	76.7M	full NIST	$\sim$4 days A100
DiffMS full NIST eval, 100 samples	76.7M	full NIST	$\sim$20 days A100
Spec2Mol top-100 decode + top-k	52.1M	full NIST	20–25 min A100
Spec2Mol MCES pass	–	full NIST	$\sim$6 min, 32 CPU workers