PERSISTENT HOMOLOGY WITH VINEYARDS FOR ROBUST SPECTRAL REPRESENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Automated molecular structure elucidation from experimental spectra is fundamentally challenged by noise, baseline distortions, overlapping signals, and modality-specific artifacts. Existing neural encoders typically represent spectra either as dense intensity grids, entangling experimental artifacts with chemical information, or as sparse peak lists that depend on heuristic detection thresholds. We introduce VINEA, a modality-general framework that represents spectra through their multiscale topological structure, capturing persistent spectral features while suppressing transient noise and preserving chemically meaningful relationships. VINEA provides a unified representation across 1D NMR, HSQC, IR, and Raman spectra. With approximately 1M parameters, VINEA provides the strongest overall robustness on ^{1}H NMR and HSQC retrieval, leads Raman robustness when trained on clean spectra, and leads IR functional-group prediction under corruption-aware training. After fine-tuning on experimental spectra, VINEA matches or exceeds the strongest specialized architectures on NMR, HSQC, and Raman while remaining competitive on IR. These results establish spectral topology as a compact, robust, and modality-agnostic representation for automated structure elucidation.

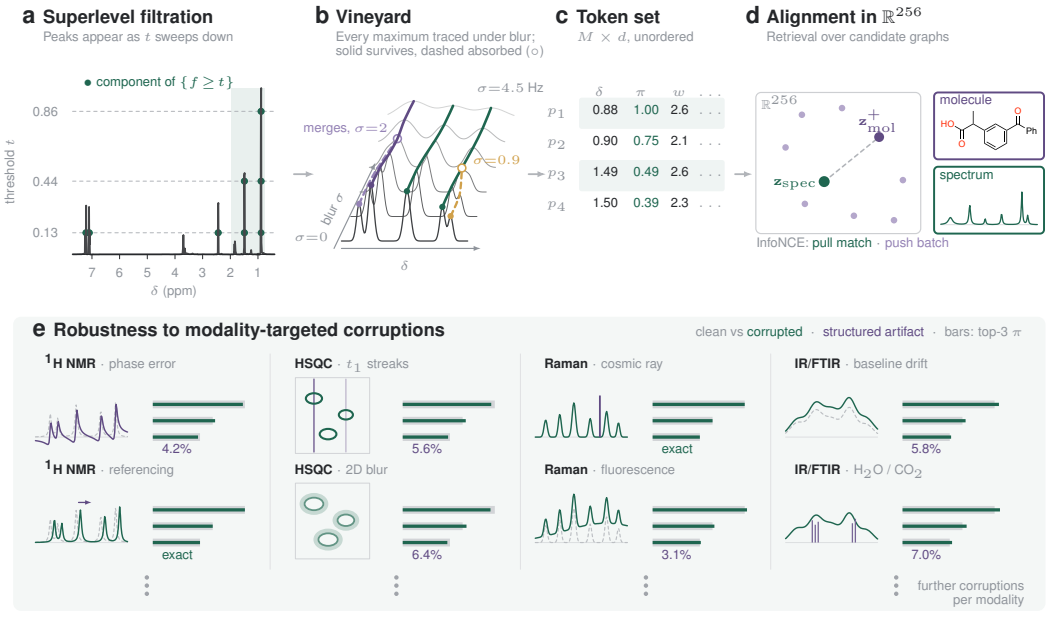


Figure 1: **VINEA spectral representation and embedding pipeline. a**. Superlevel filtration extracts H_0 features and their persistence. **b**. Features are tracked across Gaussian blur scales to form a vineyard. **c**. Topological and geometric descriptors are assembled into an unordered token set. **d**. The spectral embedding is aligned with the molecular embedding using InfoNCE for retrieval. **e**. Illustration of persistence stability under representative spectral perturbations. Percentages indicate the maximum deviation of the three most persistent features.

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

1 INTRODUCTION

Determining molecular structure from spectroscopic measurements is a central challenge in chemistry. Nuclear Magnetic Resonance (NMR) spectroscopy, often paired with two-dimensional heteronuclear correlation experiments such as HSQC, remains the gold standard for unambiguous structural assignment (Fang et al., 2026; Saurí, 2026), while infrared (IR) and Raman spectroscopy provide complementary vibrational fingerprints of functional groups. Despite their different physical origins, all four modalities can be represented as sampled intensity fields over physically calibrated spectral coordinates. Experimental chemists interpret them through the position, prominence, width, and arrangement of peaks above the baseline, rather than through absolute intensity. In recent years, deep learning has accelerated the shift from manual interpretation to automated structural prediction (Hu et al., 2026; Han et al., 2026; Xu et al., 2026; Elyashberg & Argyropoulos, 2021), yet most models are trained and evaluated on simulated or heavily preprocessed spectra. Experimental spectra, however, carry acquisition artifacts, such as thermal noise, baseline drift, and phase errors in NMR (Jahangiri et al., 2025; Xi & Rocke, 2008; Aurich & Schymanski, 2026), acquisition ridges in HSQC, atmospheric absorptions in FTIR, and fluorescence in Raman spectroscopy. A robust spectral representation must therefore preserve the chemically informative pattern while remaining insensitive to these artifacts.

Existing representations follow two paradigms, dense intensity arrays and discrete peak lists, and although seemingly opposite, both define spectral information through decisions made on absolute intensity. Dense representations treat the spectrum as a binned intensity array (Hu et al., 2024), retaining baseline, background, and noise alongside the chemical signal and binding the model to a fixed sampling grid. Peak-based representations reduce the spectrum to detected resonances or cross-peaks (Yang et al., 2026b; Xu et al., 2026), discarding many distortions but depending on a detection threshold that becomes unreliable at low signal-to-noise or under overlap. How well either preserves spectral information under realistic perturbations remains largely unquantified. To reduce dependence on absolute intensity, we build on persistent homology (Chazal & Michel, 2021; Chazal et al., 2013; Dłotko & Wanner, 2018), which tracks connected structures as an intensity threshold is lowered through the spectrum (Edelsbrunner & Harer). The resulting persistence depends on relative intensity structure rather than a fixed detection threshold and is stable to bounded perturbations (Chazal et al., 2013). Prior applications to analytical spectra have considered persistence at a single resolution (Klaila et al., 2023). We instead embed the spectrum in Gaussian scale space and track features across smoothing scales, forming a persistence vineyard (Cohen-Steiner et al., 2006; Turner, 2023). Fine-scale fluctuations disappear rapidly under smoothing, while broader spectral structures persist across larger scales, yielding a multiscale representation without explicit peak picking.

Herein, we introduce VINEA (Figure 1), which represents each spectrum as an unordered set of H_0 branches, peaks in one-dimensional spectra and connected cross-peak regions in HSQC, described by their position, persistence, local geometry, and evolution through scale space. A permutation-invariant encoder then maps this branch set to a learned spectral representation. Because the construction operates on scalar fields over one- or two-dimensional spectral domains, the same framework applies to ^{1}H NMR, HSQC, IR, and Raman spectra. Our main contributions are

- **A peak-picking-free, modality-general representation.** VINEA tracks superlevel persistence across a Gaussian scale-space vineyard, removing the need for a fixed intensity grid or hand-set peak-picking threshold. The same representation and set-encoder template apply across four spectroscopy modalities.
- **A unified benchmark for spectral corruption.** We introduce physics-based, modality-specific artifacts at five calibrated severity levels for ^{1}H NMR, HSQC, Raman, and IR/FTIR (Figure 3). VINEA and 15 published baselines are retrained under matched protocols and evaluated over ten seeds.
- **Compact robustness that transfers beyond simulation.** With approximately 1M parameters, VINEA is among the smallest models evaluated. Without corruption training, it exceeds the strongest ^{1}H NMR baseline by 7–16$\times$ in Recall@1 across S1–S5. Corruption-aware training produces the strongest overall robustness on NMR, HSQC, and IR/FTIR while remaining competitive with dedicated Raman encoders. These gains extend to field-strength shifts and fine-tuning on experimental spectra.

2 Related Work

Dense spectral representations. Spectral learning methods commonly operate directly on sampled intensity grids. In NMR, NMR2Struct (Hu et al., 2024) and SECS (Mirza et al., 2026) process discretized spectra with convolutional architectures, while NMRViT (Han et al., 2026) partitions the signal into local patches for transformer encoding. Similar strategies are used in vibrational spectroscopy, including the IR-focused ACT model (Huang et al., 2025). Dense representations preserve weak features and line-shape information, but also expose the encoder directly to acquisition artifacts, baseline variation, and changes in spectral resolution.

Sparse and structured representations. Sparse approaches instead represent spectra through detected resonances or learned spectral tokens. NMRTrans encodes NMR spectra as permutation-invariant peak sets, while NMR-Text serializes chemical shifts, multiplicities, and integrations as language-like tokens (Yang et al., 2026b). DiffNMR (Yang et al., 2025) and UltraNMR (Yang et al., 2026a) further embed chemical-shift coordinates using radial-basis or Fourier features (Zhou et al., 2025). Two-dimensional NMR naturally admits related set-based formulations based on cross-peak coordinates and intensities. These representations introduce useful spectroscopic priors, but depend on the preprocessing or peak-picking procedure used to construct the discrete feature set.

Topological spectral representations. Persistent homology provides a complementary view in which spectral structures are characterized by their evolution across intensity thresholds rather than by a fixed sampling grid or a single peak-detection threshold. Topological descriptors have previously been explored for MALDI imaging and Raman analysis, including denoising and classification (Klaila et al., 2023). VINEA extends this perspective by combining persistent spectral features with physical geometry and scale-space vineyards, yielding an unordered representation that retains both topological prominence and local spectral structure. We evaluate this representation against dense and sparse encoders under controlled, modality-specific perturbations described in Section 4 and Appendix D.

3 Persistent Scale-Space Spectral Representation

VINEA represents a processed spectrum $f : \Omega_m \to \mathbb{R}$ as a finite unordered set of persistent multiscale structures, where $\Omega_m \subset \mathbb{R}^{d_m}$ is the physical coordinate domain of modality m, with $d_m = 1$ for ^{1}H NMR, Raman, and IR and $d_m = 2$ for HSQC. VINEA extracts H_0 persistent branches, decorates them with physical geometry, follows them through Gaussian scale space, and encodes the resulting unordered token set with permutation-invariant attention. Modality-specific preprocessing and coordinate normalization are given in Appendix C.

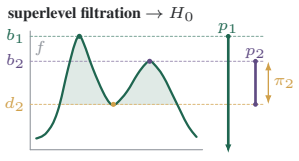
3.1 Persistent Spectral Features

Regular sampling of Ω_m induces a cubical complex $\mathcal{K}$ (Kaczynski et al., 2004). We extend sampled vertex values to every cell $\kappa \in \mathcal{K}$ by the superlevel upper-star convention $f(\kappa) = \min_{v \prec \kappa} f(v)$, where $v \prec \kappa$ ranges over the vertices of κ. For threshold $t \in \mathbb{R}$, define

$$\mathcal{K}^t = \{\kappa \in \mathcal{K} : f(\kappa) \geq t\}. \tag{3.1}$$

If $t_1 > t_2$, then $\mathcal{K}^{t_1} \subseteq \mathcal{K}^{t_2}$, so decreasing t defines a filtration. Applying H_0 yields a persistence module that records the birth and merger of connected superlevel components (Edelsbrunner et al., 2002).

For a one-dimensional spectrum, each local maximum creates an H_0 component as the threshold descends. When two components meet at an intervening minimum, the Elder Rule preserves the component with larger birth value and terminates the younger component. Let $\mathcal{B}(f)$ denote the resulting set of persistent branches. A finite branch $p \in \mathcal{B}(f)$ is associated with birth b_p, death d_p, and persistence



$$\pi_p = b_p - d_p > 0. \tag{3.2}$$

Its corresponding persistence-diagram point is (b_p, d_p), so $\mathrm{Dgm}_0(f) = \{(b_p, d_p) : p \in \mathcal{B}(f)\}$. For finite branches in one dimension, π_p coincides with the topographic prominence of the generating

maximum. The global-maximum component is the unique essential H_0 class. For the finite descriptor its death is assigned to the minimum filtration threshold $t_{\min}$ specified in Appendix C in details. For two-dimensional fields such as HSQC, the same H_0 construction tracks connected cross-peak regions rather than intervals.

Persistent homology is stable to bounded perturbations of the scalar field. For fields f and g defined on the same finite cubical complex,

$$d_B(\mathrm{Dgm}_0(f), \mathrm{Dgm}_0(g)) \le \|f - g\|_\infty, \tag{3.3}$$

where d_B is the bottleneck distance (Cohen-Steiner et al., 2007). Thus the topological descriptor itself cannot change arbitrarily under a uniformly small intensity perturbation. This guarantee applies to the persistence diagrams, not to the complete learned encoder. End-to-end robustness is evaluated experimentally in Section 5.

To retain positive- and negative-going structures, we define

$$\beta = \min(0, Q_{0.25}(f)), \qquad f_s = s(f - \beta), \quad s \in \{+1, -1\}, \tag{3.4}$$

and compute persistent branches independently on each oriented field f_s. Every branch carries its orientation s_p, and all subsequent branch geometry is evaluated on the same oriented field f_{s_p} from which the branch was obtained.

Persistent homology captures prominence but deliberately omits physical geometry. Let C_p denote the branch-owned connected component immediately before the merge event at d_p, and let c_p denote its physical position: the generating maximum coordinate in one dimension and the component centroid in two dimensions. Its area-like mass is

$$A_p = \Delta V_m \sum_{x_i \in C_p} \left(f_{s_p}(x_i) - d_p\right)_+, \tag{3.5}$$

where ΔV_m is the physical measure represented by one grid cell and $(a)_+ = \max(a, 0)$. To describe shape relative to the branch's own prominence, define

$$t_p^{(\ell)} = b_p - \ell \pi_p, \qquad C_p^{(\ell)} = \{x_i \in C_p : f_{s_p}(x_i) \ge t_p^{(\ell)}\}, \tag{3.6}$$

for $\ell \in \{0.25, 0.50, 0.75\}$. In one dimension, the physical extent of $C_p^{(\ell)}$ gives the prominence-relative widths $w_p^{(\ell)}$. We additionally retain half-prominence asymmetry and distances to neighboring persistent branches. In two dimensions, the same construction uses component centroids and coordinate-wise physical extents. We collect these non-positional geometric quantities in g_p.

3.2 Scale-Space Vineyards

A native-resolution persistence diagram describes the spectrum at only one effective resolution. VINEA therefore embeds the spectral field in Gaussian scale space,

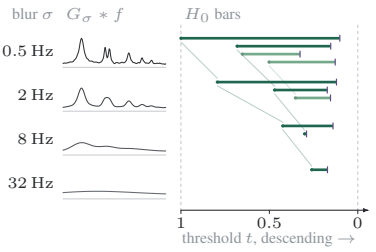
$$f_\sigma(\mathbf{x}) = (G_\sigma * f)(\mathbf{x}), \qquad G_\sigma(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{d_m/2}} \exp\left(-\frac{\|\mathbf{x}\|_2^2}{2\sigma^2}\right), \tag{3.7}$$

where $\sigma > 0$ is the Gaussian standard deviation in the coordinate metric used for modality m. For multidimensional spectra, the relative scaling of coordinate axes used by the Gaussian metric is fixed by the modality-specific preprocessing described in the supplementary material.

At each fixed σ, the same upper-star construction yields $\mathrm{Dgm}_0(f_\sigma)$. The continuous Gaussian family induces

$$\sigma \longmapsto \mathrm{Dgm}_0(f_\sigma), \tag{3.8}$$

a persistence vineyard whose vines trace persistence points through the birth–death plane (Cohen-Steiner et al., 2006; Turner, 2023). Here t is the filtration parameter and σ is the scale-space continuation parameter. Superlevel sets need not be nested in σ, so this is not a canonical bifiltration.



For continuous one-dimensional Gaussian scale space, increasing σ cannot create new local extrema (Lindeberg, 1994). Existing maxima and minima move and may annihilate in maximum–minimum pairs. A native-scale branch can therefore be continued toward coarser scales until the

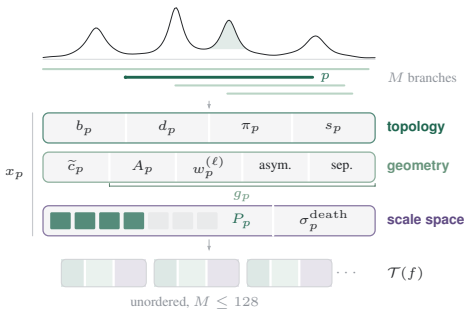


Figure 2: **From persistent branches to VINEA tokens.** Each branch p is summarized by topology $\tau_p = (b_p, d_p, \pi_p, s_p)$, physical position c_p, local geometry g_p, and vineyard profile $v_p = (P_p, \sigma_p^{\text{death}})$. Their concatenation forms the branch descriptor x_p, and the resulting unordered set is processed by the set encoder.

Algorithm 1: VINEA tokenization

Input: Normalized field f, physical coordinates x, modality scale ladder Σ_m, token budget M
Output: Unordered branch set $\mathcal{T}(f)$

$\beta \leftarrow \min(0, Q_{0.25}(f))$
$\mathcal{T} \leftarrow \varnothing$
for $s \in \{+1, -1\}$ **do**
  $f_s \leftarrow s(f - \beta)$
  $\mathcal{B}^s \leftarrow \textsc{PersistentBranches}_0(f_s)$
  $\mathcal{V}^s \leftarrow \textsc{TrackVineyard}(f_s, \Sigma_m)$
  foreach $p \in \mathcal{B}^s$ **do**
    $\pi_p \leftarrow b_p - d_p$
    $(c_p, g_p) \leftarrow \textsc{Geometry}(p, f_s, x)$
    $(P_p, \sigma_p^{\text{death}}) \leftarrow \mathcal{V}^s[p]$
    $\mathcal{T} \leftarrow \mathcal{T} \cup \{(b_p, d_p, \pi_p, s, c_p, g_p, P_p, \sigma_p^{\text{death}})\}$
$\mathcal{T} \leftarrow \text{TopM}_\pi(\mathcal{T}, M)$
return $\mathcal{T}$, $|\mathcal{T}| \leq M$

GEOMETRY returns physical position and local branch geometry. TRACKVINEYARD samples the Gaussian-scale vineyard on Σ_m, associates branches between adjacent scales, and returns each sampled persistence profile and scale-death statistic.

associated extremum annihilates, after which the branch remains inactive. VINEA samples this continuous construction on the finite modality-specific ladder $\Sigma_m = \{\sigma_1, \ldots, \sigma_L\}$ and follows branches between adjacent sampled scales. For HSQC, persistence diagrams are sampled in the same way and corresponding components are associated between adjacent scales by the deterministic matching procedure specified in Appendix C. The one-dimensional non-creation theorem is not used to justify that matching.

At scale σ, let $\pi_p(\sigma)$ denote the persistence of the continuation of native branch p. Because the absolute persistence scale varies between spectra and across blur scales, we convert it to a bounded relative-persistence score

$$r_p(\sigma) = \frac{\pi_p(\sigma)}{\pi_p(\sigma) + \tau_\pi(\sigma)}, \qquad \tau_\pi(\sigma) = \text{median}_{q \in \mathcal{A}_\sigma}\, \pi_q(\sigma)\, \min\left(1, \frac{M_\sigma}{\rho N}\right), \tag{3.9}$$

where $\mathcal{A}_\sigma$ is the set of active finite branches at scale σ, $M_\sigma = |\mathcal{A}_\sigma|$, N is the number of sampled grid locations, and $\rho = 0.05$. If no finite branch is present, the corresponding profile entries are set to zero. For $\tau_\pi(\sigma) > 0$, the transformation is monotone in persistence, lies in $[0, 1]$, equals $1/2$ when $\pi_p = \tau_\pi$, and approaches zero as $\pi_p \to 0$. For every native branch, VINEA records the sampled scale-space persistence profile

$$P_p = \big(r_p(\sigma_1), \ldots, r_p(\sigma_L)\big) \in [0, 1]^L, \tag{3.10}$$

with entries set to zero after the branch disappears. We additionally store the largest sampled scale at which the branch remains active,

$$\sigma_p^{\text{death}} = \max\{\sigma_\ell : p \text{ is active at } \sigma_\ell\}. \tag{3.11}$$

This is a discrete scale-death statistic rather than the exact continuous annihilation scale. Branches surviving through σ_L are right-censored at the end of the ladder. We call $(P_p, \sigma_p^{\text{death}})$ the branch's vineyard profile. For 1D NMR, $\Sigma_{\text{NMR}} = \{0.5, 1, 2, 4, 8, 16, 32\}$ Hz. Other modalities use ladders expressed in their corresponding coordinate systems.

3.3 PERMUTATION-INVARIANT SPECTRAL ENCODING

Algorithm 1 maps the original grid to the unordered set $\mathcal{T}(f) = \{x_p\}_{p=1}^M$ with $M \leq 128$. We write the descriptor of branch p as

$$x_p = \big[b_p,\, d_p,\, \pi_p,\, s_p,\, c_p,\, g_p,\, P_p,\, \sigma_p^{\text{death}}\big]. \tag{3.12}$$

No dense spectral grid is reconstructed after this stage.

Absolute spectral position receives a dedicated multiscale encoding. Let $\psi_m : \Omega_m \to [0,1]^{d_m}$ denote affine normalization of the modality-specific acquisition window and define $\xi_p = \psi_m(c_p)$. We use coordinate-wise dyadic Fourier features (Tancik et al., 2020),

$$\gamma_F(\xi_p) = \bigparallel_{j=1}^{d_m} \bigparallel_{k=0}^{F-1} \left[\sin(2^k \pi \xi_{p,j}), \cos(2^k \pi \xi_{p,j})\right], \qquad F = 8. \tag{3.13}$$

Let $u_p = [x_p \| \gamma_F(\xi_p)]$. A shared embedding network maps u_p to the latent dimension, while a learned scalar gate $q_\theta(u_p)$ modulates the explicit persistence weighting,

$$z_p^{(0)} = m_p \, \pi_p \, \mathrm{sigmoid}\big(q_\theta(u_p)\big) \, \mathrm{MLP}_{\mathrm{emb}}(u_p), \tag{3.14}$$

where $m_p \in \{0,1\}$ masks padding and q_θ is scalar-valued. The factor π_p provides an architectural persistence bias: branches with vanishing prominence enter the attention stack with vanishing latent norm without requiring a hand-selected detection threshold. The encoder retains absolute spectral position while supplying self-attention with a separate relative geometric pathway. Let

$$\Lambda_m = \mathrm{diag}(\lambda_{m,1}, \ldots, \lambda_{m,d_m}) \tag{3.15}$$

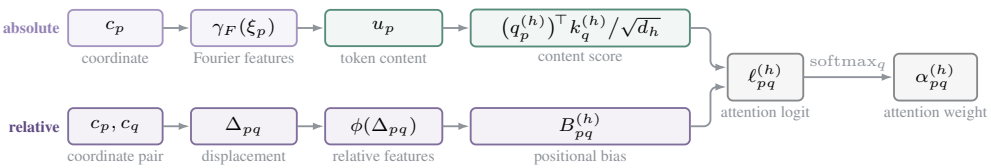
contain fixed modality-specific physical normalization scales, and define the normalized displacement

$$\Delta_{pq} = \Lambda_m^{-1}(c_p - c_q) \in \mathbb{R}^{d_m}. \tag{3.16}$$

Its coordinate-wise Fourier embedding $\phi(\Delta_{pq})$ is mapped by a learned head-specific scalar function B_h to an additive attention bias. For attention head h,

$$\ell_{pq}^{(h)} = \underbrace{\frac{(q_p^{(h)})^\top k_q^{(h)}}{\sqrt{d_h}}}_{\text{content}} + \underbrace{B_h(\phi(\Delta_{pq}))}_{\text{relative geometry}} + \underbrace{M_{pq}}_{\text{mask}}, \tag{3.17}$$

where $M_{pq} = 0$ when token q is valid and $M_{pq} = -\infty$ when q is padding. The corresponding attention weights are $\alpha_{pq}^{(h)} = \mathrm{softmax}_q(\ell_{pq}^{(h)})$. Padded query outputs are masked after each block.



The two positional pathways encode complementary inductive biases. The absolute encoding $\gamma_F(\xi_p)$ retains chemically informative resonance location. In contrast, the relative pathway depends only on coordinate differences and is exactly invariant to a common translation:

$$\Lambda_m^{-1}\big[(c_p + \delta) - (c_q + \delta)\big] = \Lambda_m^{-1}(c_p - c_q) = \Delta_{pq}. \tag{3.18}$$

VINEA is therefore not globally translation invariant. Instead, self-attention receives an exactly translation-invariant relational channel while absolute position remains available as complementary information. We use two Set Transformer blocks with four attention heads (Lee et al., 2019). The final token set is reduced by three symmetric readouts, pooling by multi-head attention (PMA), masked mean, and masked coordinate-wise maximum, and projected to a normalized 256-dimensional spectral embedding,

$$z_{\mathrm{spec}} = \mathrm{Head}\left[z_{\mathrm{PMA}} \,\|\, \frac{\sum_p m_p z_p}{\sum_p m_p} \,\|\, \max_{p:m_p=1} z_p\right], \qquad \widehat{z}_{\mathrm{spec}} = \frac{z_{\mathrm{spec}}}{\|z_{\mathrm{spec}}\|_2}. \tag{3.19}$$

Because branchwise maps are shared, self-attention is permutation equivariant and the readout is symmetric, so the encoder is invariant to branch ordering. The spectral tower contains approximately 0.6M trainable parameters and operates on at most 128 persistent branches rather than on the original spectral grid. The encoder architecture is shared across modalities, while modality-specific coordinate maps, physical normalization scales, and scale ladders account for their different domains. Retrieval tasks align $\widehat{z}_{\mathrm{spec}}$ contrastively with a molecular embedding, whereas classification tasks attach a task-specific prediction head. Architecture and feature compositions were ablated in Table 16.

4 MULTIMODAL ROBUSTNESS BENCHMARK

We evaluate whether spectral representations preserve molecular information under perturbations introduced by acquisition, processing, and sample conditions. Corruptions are applied to processed spectra as controlled stress tests rather than complete instrument simulations. Exact transformations and severity schedules are provided in Appendix D.

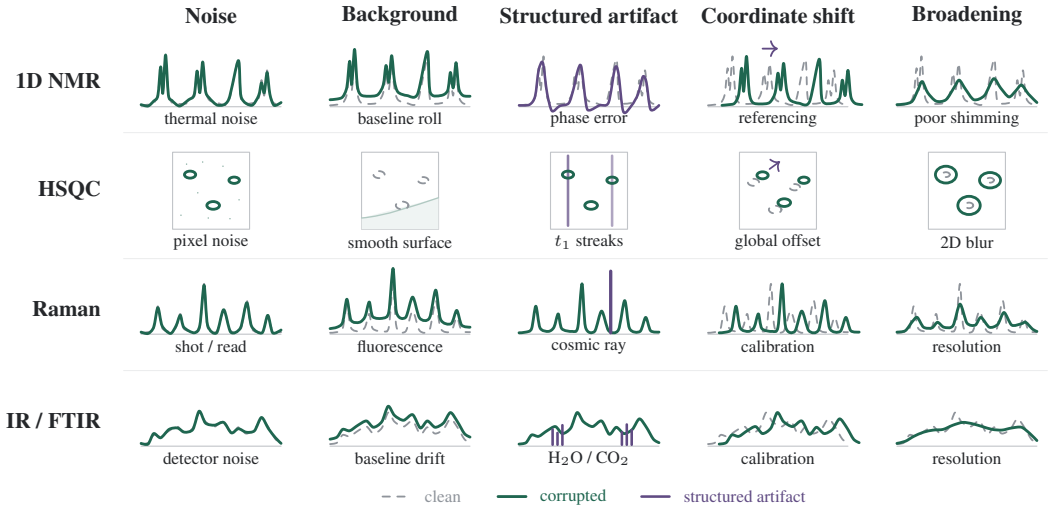


Figure 3: **Unified synthetic corruption benchmark across spectral modalities.** Rows denote modalities and columns group transformations by their effect on the processed signal. Dashed gray denotes the clean signal, VINEA green the corrupted observation, and violet a modality-specific structured artifact. The schematics are illustrative. Exact transformations and severity schedules are reported in Appendix D.

Spectral Modalities and Datasets We evaluate VINEA on four spectral modalities. For ^{1}H NMR and HSQC, we use the simulated spectra of the Molecular Spectra Dataset (MSD) (Alberts et al., 2024), with 1D retrieval against 18,493 candidate structures and HSQC evaluated on 5,000 molecules under a scaffold split. For Raman, we perform spectrum-to-molecule retrieval on Raman-ChEMBL (Liang et al., 2025), and for IR, 17-label functional-group prediction on the 177,461 patent-derived molecules of Zipoli et al. (2025). Retrieval is scored with Recall@1 and IR with macro-AUROC. For NMR, we additionally evaluate transfer to experimental spectra and magnetic-field transfer. Dataset construction, splits, and candidate pools are detailed in Appendix E.

Corruption suite. Severity increases from S1 (mild) to S5 (very severe). The suite covers instrumental noise, structured interference, calibration shifts, and resolution changes. NMR includes thermal noise, baseline distortion, phase error, referencing shifts, and line broadening. Field transfer are evaluated separately. HSQC additionally includes smooth 2D backgrounds and t_1 streaking. Raman includes shot/read noise, fluorescence, cosmic-ray spikes, calibration drift, and broadening, while IR includes detector noise, baseline drift, calibration error, broadening, and atmospheric H_2O/CO_2 residuals. We evaluate on clean spectra and on combined corruptions at each severity S1-S5, with single-artifact results at S3 are reported in Appendix Table 10. Corruption draws are deterministic for each sample, condition, and benchmark seed.

Baselines and protocol. We compare VINEA with 15 modality-specific convolutional, patch-based, and transformer baselines on identical partitions, applying corruptions before any model-specific preprocessing. In the clean-trained setting, models are trained only on clean spectra and evaluated on unseen S1-S5 corruptions. In the corruption-aware setting, the first 20% of training is clean, after which the maximum severity rises progressively from S1 to S3 and each corruption is applied independently with probability 0.5. Results are averaged over 10 paired seeds. Architectures, optimization, and statistical details are given in Appendices E and F.

5 RESULTS

We evaluate VINEA along three axes: robustness to corruptions unseen during training, robustness acquired through corruption-aware training, and transfer from simulated to experimental spectra.

5.1 ROBUSTNESS TO UNSEEN CORRUPTIONS AND SEVERITIES

We separate robustness induced by the representation from robustness learned through noise augmentation (Figure 4). Clean-trained models (top row) see only clean spectra, whereas corruption-aware models (bottom row) are trained on a linear S1–S3 ramp with 20% clean spectra, so S4-S5 test extrapolation. Under clean-only training, VINEA is markedly more robust on retrieval. On ^{1}H NMR, it exceeds the strongest baseline by 7-16$\times$ across S1-S5 and remains $\sim$20$\times$ above chance at S5, where all baselines approach chance (5.4×10^{-5}). On Raman, despite comparable clean performance (0.781 vs. 0.797 for SANet), VINEA retains $0.509/0.264/0.063$ at S1–S3, compared with $0.099/0.016/0.004$ for the strongest baseline. HSQC is the most fragile modality: VINEA leads on clean spectra (0.465 vs. 0.326 Recall@1 for DeepSAT) and retains the highest mean through S4, but all methods approach chance at S4-S5. On IR, VINEA and IRCNN are the leading pair under corruption, differing by at most 0.005 macro-AUROC (0.917 vs. 0.916 at S1. 0.660 vs. 0.665 at S5), while the next-best baseline reaches only 0.604.

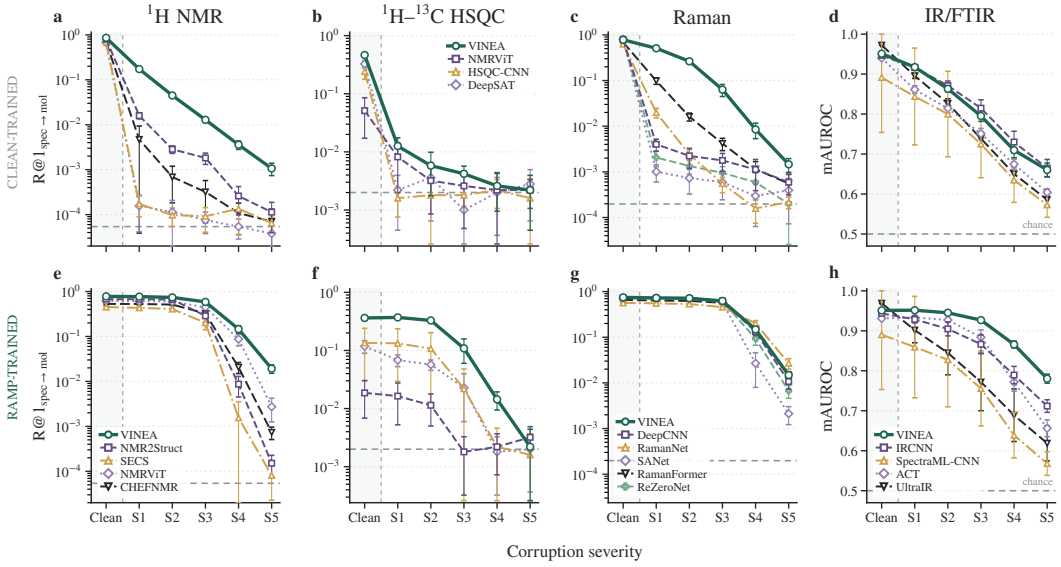


Figure 4: **Robustness under clean-only and corruption-aware training.** Performance from clean spectra through combined corruption severities S1–S5 for ^{1}H NMR, ^{1}H-^{13}C HSQC, Raman, and IR/FTIR. *Top:* clean-trained models. *Bottom:* models trained with a linear S1-S3 corruption ramp and 20% clean spectra. S4-S5 are unseen during training. Metrics are spectrum-to-molecule Recall@1 (log scale) and macro-AUROC for IR/FTIR. Points show mean $\pm$ s.d. over ten seeds, and dashed lines indicate chance. VINEA uses the same representation and set-encoder template across all modalities.

Corruption-aware training improves robustness across architectures, and VINEA remains the most robust model in three of the four modalities. At S3, it attains a Recall@1 of 0.588 on ^{1}H NMR, compared with 0.446 for the strongest baseline, and 0.109 on HSQC, compared with 0.023, an almost five-fold improvement. For IR functional-group prediction, it increases macro-AUROC over the strongest baseline by 4.3 and 6.8 percentage points at S3 and S5, respectively. Raman is the principal exception: VINEA is on par with the strongest published baselines through S3 but is overtaken at S4–S5, marking the present limit of this robustness. Notably, all results use a single persistence-based representation and set-encoder template shared across the four modalities, indicating that this robustness does not depend on modality-specific feature engineering.

5.2 Experimental transfer beyond simulated corruption

Robustness to simulated corruptions translates into competitive performance on experimental spectra. After S1–S3 ramp pretraining and fine-tuning on experimental data (Figure 5a), VINEA matches or slightly exceeds the strongest specialized architecture on ^{1}H NMR (0.382 vs. 0.375 for NMRViT), HSQC (0.336 vs. 0.332 for DeepSAT), and Raman (0.4% vs. 0.7% error for RamanFormer), and comes within 0.021 macro-AUROC of UltraIR on IR, all with a single shared design. Corruption-aware pretraining already benefits VINEA before any fine-tuning: it raises zero-shot Recall@1 on measured ^{1}H NMR by 2.4–3.2× across all four repositories and increases IR macro-AUROC from 0.620 to 0.684, the largest gain among IR models, while VINEA leads zero-shot Raman retrieval under both pretraining regimes. On measured ^{1}H NMR and HSQC, dense encoders remain stronger in the zero-shot setting, a gap that target-domain fine-tuning closes (Appendix Tables 13 and 14).

Under simulated field-strength shifts on GISSMO (Dashti et al., 2017) (Figure 5b), VINEA trained at 400 MHz outperforms every baseline from 200 to 1300 MHz without adaptation. It reaches a Recall@1 of 0.293 at the training field, 1.45× the strongest baseline (0.202), and still leads at 1300 MHz (0.214 vs. 0.195). Below 200 MHz, NMR2Struct is comparable to or slightly ahead of VINEA. The t-SNE projections in Figure 5c show that VINEA's embedding organization is preserved under corruption, and CKA, effective rank, and neighborhood retention are analyzed in Appendix I.

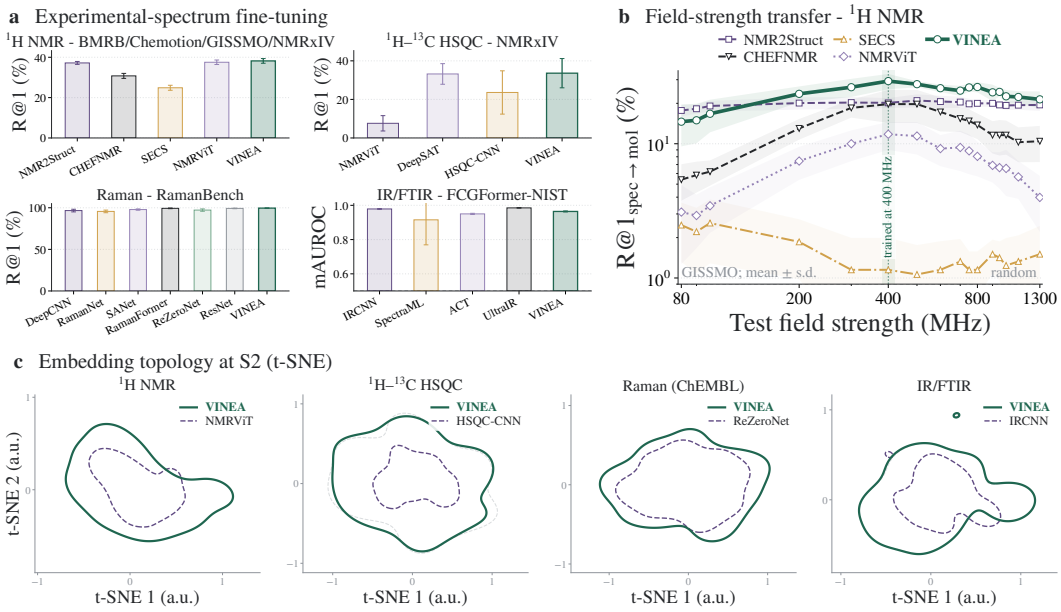


Figure 5: **Experimental transfer and embedding geometry across four spectroscopy modalities.** **a.** Experimental fine-tuning after S1-S3 corruption-ramp pretraining with 20% clean spectra. Bars show mean ± s.d. over ten seeds. NMR results are pooled across BMRB, Chemotion, GISSMO, and NMRxIV. **b.** Transfer from training at 400 MHz to GISSMO spectra measured between 80 and 1300 MHz. Recall@1 is shown on a logarithmic scale. **c.** S2 embedding topology for 350 matched spectra. Contours enclose 80% of the t-SNE density and the faint contour shows clean VINEA. VINEA uses the same representation and encoder template across all modalities.

6 Conclusion

We introduced VINEA, a compact, modality-agnostic representation that tracks persistent spectral structure across a Gaussian scale-space vineyard, together with a multimodal corruption benchmark for ^{1}H NMR, HSQC, Raman, and IR. Against 15 retrained published encoders, VINEA gives the largest robustness gains without corruption training and the largest chance-adjusted robustness with it, lying on the robustness-parameter Pareto frontier in all four modalities (Appendix Figure 12), and remains competitive after fine-tuning on experimental spectra. Closing the gap in zero-shot experimental transfer is the main open direction.

AI USE STATEMENT

Generative AI tools were used to assist with research software development, including coding and debugging, and with the writing and editing of the manuscript. They were also used to assist with the design and implementation of schematic figures and LaTeX/TikZ code. No generative image models were used, and no experimental results, benchmark values, or scientific conclusions were generated by AI.

All AI-assisted code, text, equations, and figures were reviewed and, where applicable, tested or verified by the authors. The authors take full responsibility for the final content of the manuscript and Supplementary Information.

ETHICS STATEMENT

This work develops machine-learning methods for robust molecular spectroscopy and does not involve human participants, personal data, or clinical decision making. Experiments use simulated or previously collected NMR, HSQC, Raman, and IR datasets.

VINEA is intended to improve the reliability of molecular identification and spectral representation under realistic acquisition variability. As with other computational chemistry tools, downstream use may extend to applications beyond those evaluated here and should remain subject to appropriate scientific, institutional, and regulatory oversight.

The synthetic corruptions used in our benchmark are controlled stress tests on processed spectra rather than exhaustive simulations of experimental instruments. Performance may therefore vary across laboratories, instruments, molecular classes, and acquisition protocols not represented in the benchmark.

REPRODUCIBILITY STATEMENT

The main paper and Supplementary Information provide the complete mathematical definition of VINEA and specify the datasets, preprocessing, corruption operators, baselines, training procedures, evaluation metrics, and ablations used in the experiments. We include the source code and experiment configurations required to reproduce the reported results. Random seeds are fixed, synthetic corruptions are generated deterministically where applicable, and reported results are aggregated over repeated seeded runs. The software environment is managed with `uv` using a locked dependency specification to ensure consistent package versions. Dataset splits and all modality-specific preprocessing and benchmark parameters are documented in the Supplementary Information.

REFERENCES

Marvin Alberts, Oliver Schilter, Federico Zipoli, Nina Hartrampf, and Teodoro Laino. Unraveling Molecular Structure: A Multimodal Spectroscopic Dataset for Chemistry. In *Advances in Neural Information Processing Systems*, volume 37, pp. 125780–125808. Curran Associates, Inc., 2024. doi: 10.52202/079017-3996. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/e38e60b33bb2c6993e0865160cdb5cf1-Abstract.html.

Dagny Aurich and Emma L. Schymanski. Advances and perspectives in computer-assisted structure elucidation: A review. *Digital Discovery*, 5(3):1023–1036, March 2026. ISSN 2635-098X. doi: 10.1039/d5dd00438a. URL https://doi.org/10.1039/d5dd00438a.

Vincenzo Barone, Malgorzata Biczysko, and Julien Bloino. Fully anharmonic IR and Raman spectra of medium-size molecular systems: Accuracy and interpretation. *Phys. Chem. Chem. Phys.*, 16(5):1759–1787, 2014. ISSN 1463-9076, 1463-9084. doi: 10.1039/C3CP53413H. URL https://pubs.rsc.org/cp/article/16/5/1759-1787/397149.

F. Bloch, W. W. Hansen, and Martin Packard. Nuclear Induction. *Physical Review*, 69(3-4): 127–127, February 1946. ISSN 0031-899X. doi: 10.1103/PhysRev.69.127. URL https://link.aps.org/doi/10.1103/PhysRev.69.127.

Geoffrey Bodenhausen and David J. Ruben. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, 69(1):185–189, January 1980. ISSN 0009-2614. doi: 10.1016/0009-2614(80)80041-8. URL https://www.sciencedirect.com/science/article/pii/0009261480800418.

M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 389 (20):457–484, 1927. ISSN 1521-3889. doi: 10.1002/andp.19273892002. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19273892002.

Frédéric Chazal and Bertrand Michel. An introduction to Topological Data Analysis: Fundamental and practical aspects for data scientists, February 2021. URL http://arxiv.org/abs/1710.04019.

Frederic Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules, March 2013. URL http://arxiv.org/abs/1207.3674.

David Cohen-Steiner, Herbert Edelsbrunner, and Dmitriy Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*, SCG '06, pp. 119–126, New York, NY, USA, June 2006. Association for Computing Machinery. ISBN 978-1-59593-340-9. doi: 10.1145/1137856.1137877. URL https://dl.acm.org/doi/10.1145/1137856.1137877.

David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of Persistence Diagrams. *Discrete & Computational Geometry*, 37(1):103–120, January 2007. ISSN 1432-0444. doi: 10.1007/s00454-006-1276-5. URL https://doi.org/10.1007/s00454-006-1276-5.

Hesam Dashti, William M. Westler, Marco Tonelli, Jonathan R. Wedell, John L. Markley, and Hamid R. Eghbalnia. Spin System Modeling of Nuclear Magnetic Resonance Spectra for Applications in Metabolomics and Small Molecule Screening. *Analytical Chemistry*, 89(22):12201–12208, November 2017. ISSN 0003-2700, 1520-6882. doi: 10.1021/acs.analchem.7b02884. URL https://pubs.acs.org/doi/10.1021/acs.analchem.7b02884.

Lin Deng, Yuzhong Zhong, Maoning Wang, Xiujuan Zheng, and Jianwei Zhang. Scale-Adaptive Deep Model for Bacterial Raman Spectra Identification. *IEEE Journal of Biomedical and Health Informatics*, 26(1):369–378, January 2022. ISSN 2168-2208. doi: 10.1109/JBHI.2021.3113700. URL https://ieeexplore.ieee.org/document/9541099.

Paweł Dłotko and Thomas Wanner. Rigorous cubical approximation and persistent homology of continuous functions. *Computers & Mathematics with Applications*, 75(5):1648–1666, March 2018. ISSN 08981221. doi: 10.1016/j.camwa.2017.11.027. URL http://arxiv.org/abs/1610.03833.

Edelsbrunner, Letscher, and Zomorodian. Topological Persistence and Simplification. *Discrete & Computational Geometry*, 28(4):511–533, November 2002. ISSN 1432-0444. doi: 10.1007/s00454-002-2885-2. URL https://doi.org/10.1007/s00454-002-2885-2.

Herbert Edelsbrunner and John Harer. *COMPUTATIONAL TOPOLOGY AN INTRODUCTION*.

Mikhail Elyashberg and Dimitris Argyropoulos. Computer Assisted Structure Elucidation (CASE): Current and future perspectives. *Magnetic Resonance in Chemistry*, 59(7):669–690, 2021. ISSN 1097-458X. doi: 10.1002/mrc.5115. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrc.5115.

Richard R. Ernst. Nuclear Magnetic Resonance Fourier Transform Spectroscopy (Nobel Lecture). *Angewandte Chemie International Edition in English*, 31(7):805–823, 1992. ISSN 1521-3773. doi: 10.1002/anie.199208053. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.199208053.

Zheng Fang, Chen Yang, Yusen Tan, Yunpeng Zhao, Fanjie Xu, Hongxin Xiang, Hanyu Sun, Hanyu Gao, Xiaojian Wang, Wenjie Du, Yuqiang Li, and Jun Xia. Towards Generalizable and Evidential Nuclear Magnetic Resonance-Based Molecular Structure Elucidation via Large Language Model Agent, June 2026. URL http://arxiv.org/abs/2606.29776.

Rekha Gautam, Sandeep Vanga, Freek Ariese, and Siva Umapathy. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation*, 2(1):8, June 2015. ISSN 2195-7045. doi: 10.1140/epjti/s40485-015-0018-6. URL https://doi.org/10.1140/epjti/s40485-015-0018-6.

Chao Han, Xiaolin Pan, and Yingkai Zhang. End-to-end molecular structure elucidation from multimodal NMR spectra images using vision transformers. *Chemical Science*, 17(26):13072–13082, 2026. ISSN 2041-6520, 2041-6539. doi: 10.1039/D6SC02352E. URL https://xlink.rsc.org/?DOI=D6SC02352E.

Frank Hu, Michael S. Chen, Grant M. Rotskoff, Matthew W. Kanan, and Thomas E. Markland. Accurate and Efficient Structure Elucidation from Routine One-Dimensional NMR Spectra Using Multitask Machine Learning. *ACS Central Science*, 10(11):2162–2170, November 2024. ISSN 2374-7943, 2374-7951. doi: 10.1021/acscentsci.4c01132. URL https://pubs.acs.org/doi/10.1021/acscentsci.4c01132.

Frank Hu, Jonathan M. Tubb, Dimitris Argyropoulos, Sergey Golotvin, Mikhail Elyashberg, Grant M. Rotskoff, Matthew W. Kanan, and Thomas E. Markland. Pushing the Limits of One-Dimensional NMR Spectroscopy for Automated Structure Elucidation Using Artificial Intelligence. *Journal of Chemical Information and Modeling*, 66(12):6992–7002, June 2026. ISSN 1549-9596. doi: 10.1021/acs.jcim.6c00905. URL https://doi.org/10.1021/acs.jcim.6c00905.

Shiluo Huang, Yining Jin, Wei Jin, and Ying Mu. Analytical-Chemistry-Informed Transformer for Infrared Spectra Modeling. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(16):17440–17448, April 2025. ISSN 2374-3468. doi: 10.1609/aaai.v39i16.33917. URL https://ojs.aaai.org/index.php/AAAI/article/view/33917.

Nabil Ibtehaz, Muhammad E. H. Chowdhury, Amith Khandakar, Serkan Kiranyaz, M. Sohel Rahman, and Susu M. Zughaier. RamanNet: A generalized neural network architecture for Raman spectrum analysis. *Neural Computing and Applications*, 35(25):18719–18735, September 2023. ISSN 1433-3058. doi: 10.1007/s00521-023-08700-z. URL https://doi.org/10.1007/s00521-023-08700-z.

Amir Jahangiri, Tatiana Agback, Ulrika Brath, and Vladislav Orekhov. Towards Ultimate NMR Resolution with Deep Learning, February 2025. URL http://arxiv.org/abs/2502.20793.

Guwon Jung, Son Gyo Jung, and Jacqueline M. Cole. Automatic materials characterization from infrared spectra using convolutional neural networks. *Chemical Science*, 14(13):3600–3609, April 2023. ISSN 2041-6520. doi: 10.1039/d2sc05892h. URL https://doi.org/10.1039/d2sc05892h.

Tomasz Kaczynski, Konstantin Mischaikow, and Marian Mrozek. Cubical Homology. In Tomasz Kaczynski, Konstantin Mischaikow, and Marian Mrozek (eds.), *Computational Homology*, pp. 39–92. Springer, New York, NY, 2004. ISBN 978-0-387-21597-6. doi: 10.1007/0-387-21597-2_2. URL https://doi.org/10.1007/0-387-21597-2_2.

Hyun Woo Kim, Chen Zhang, Raphael Reher, Mingxun Wang, Kelsey L. Alexander, Louis-Félix Nothias, Yoo Kyong Han, Hyeji Shin, Ki Yong Lee, Kyu Hyeong Lee, Myeong Ji Kim, Pieter C. Dorrestein, William H. Gerwick, and Garrison W. Cottrell. DeepSAT: Learning Molecular Structures from Nuclear Magnetic Resonance Data. *Journal of Cheminformatics*, 15(1):71, August 2023. ISSN 1758-2946. doi: 10.1186/s13321-023-00738-4. URL https://doi.org/10.1186/s13321-023-00738-4.

Gideon Klaila, Vladimir Vutov, and Anastasios Stefanou. Supervised topological data analysis for MALDI mass spectrometry imaging applications. *BMC Bioinformatics*, 24(1):279, July 2023. ISSN 1471-2105. doi: 10.1186/s12859-023-05402-0. URL http://arxiv.org/abs/2302.13948.

Mario Koddenbrock, Christoph Lange, Robin Legner, Martin Jäger, Martin Kögler, Mariano N. Cruz Bournazou, Peter Neubauer, Felix Biessmann, and Erik Rodner. RamanBench: A Large-Scale Benchmark for Machine Learning on Raman Spectroscopy, May 2026. URL http://arxiv.org/abs/2605.02003.

Onur Can Koyun, Reyhan Kevser Keser, Safa Onur Şahin, Damla Bulut, Mustafa Yorulmaz, Veysel Yücesoy, and Behçet Uğur Töreyin. RamanFormer: A Transformer-Based Quantification Approach for Raman Mixture Components. *ACS Omega*, 9(22):23241–23251, June 2024. ISSN 2470-1343. doi: 10.1021/acsomega.3c09247. URL https://doi.org/10.1021/acsomega.3c09247.

Stefan Kuhn, Eda Tumer, Simon Colreavy-Donnelly, and Ricardo Moreira Borges. A pilot study for fragment identification using 2D NMR and deep learning. *Magnetic Resonance in Chemistry*, 60(11):1052–1060, 2022. ISSN 1097-458X. doi: 10.1002/mrc.5212. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrc.5212.

Greg Landrum, Paolo Tosco, Brian Kelley, Ricardo Rodriguez, David Cosgrove, Riccardo Vianello, sriniker, Peter Gedeck, Gareth Jones, Dan Nealschneider, Eisuke Kawashima, NadineSchneider, tadhurst-cdd, Andrew Dalke, Niels Maeder, Matt Swain, Yakov Pechersky, Brian Cole, Samo Turk, Aleksandr Savelev, Rachel Walker, Alain Vaucher, Maciej Wójcikowski, Hussein Faara, Ichiru Take, Vincent F. Scalfani, Daniel Probst, Kazuya Ujihara, Steven Kearnes, and Axel Pahl. Rdkit/rdkit: 2026_03_5 (Q1 2026) Release. August 2026. doi: 10.5281/zenodo.21741729. URL https://zenodo.org/records/21741729.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam R. Kosiorek, Seungjin Choi, and Yee Whye Teh. Set Transformer: A Framework for Attention-based Permutation-Invariant Neural Networks, May 2019. URL http://arxiv.org/abs/1810.00825.

Malcolm H Levitt. *Spin Dynamics: Basics of Nuclear Magnetic Resonance, 2nd Edition*. 2008. ISBN 978-0-470-51118-3.

Jiechun Liang, Jack Ling, Limin Xu, and Xi Zhu. A Dataset of Raman and Infrared Spectra as an Extension to the ChEMBL. *Scientific Data*, 12(1):939, June 2025. ISSN 2052-4463. doi: 10.1038/s41597-025-05289-x. URL https://www.nature.com/articles/s41597-025-05289-x.

Tony Lindeberg. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, 21(1-2):225–270, January 1994. ISSN 0266-4763. doi: 10.1080/757582976. URL https://www.tandfonline.com/doi/abs/10.1080/757582976.

Jinchao Liu, Margarita Osadchy, Lorna Ashton, Michael Foster, Christopher J. Solomon, and Stuart J. Gibson. Deep convolutional neural networks for Raman spectrum recognition: A unified solution. *Analyst*, 142(21):4067–4074, November 2017. ISSN 0003-2654. doi: 10.1039/c7an01371j. URL https://doi.org/10.1039/c7an01371j.

Derek A. Long. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*. Wiley, May 2002. ISBN 978-0-471-49028-9.

Thomas G. Mayerhöfer, Susanne Pahlow, and Jürgen Popp. The Bouguer-Beer-Lambert Law: Shining Light on the Obscure. *ChemPhysChem*, 21(18):2029–2046, 2020. ISSN 1439-7641. doi: 10.1002/cphc.202000464. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cphc.202000464.

Adrian Mirza, Luc Patiny, and Kevin Maik Jablonka. End-to-end multimodal structure elucidation from raw spectra combining contrastive learning and evolutionary algorithms. *Nature Communications*, 17(1):5013, June 2026. ISSN 2041-1723. doi: 10.1038/s41467-026-73846-y. URL https://www.nature.com/articles/s41467-026-73846-y.

Huaping Mo, John S. Harwood, Danzhou Yang, and Carol Beth Post. A simple method for NMR $t1$ noise suppression. *Journal of Magnetic Resonance*, 276:43–50, March 2017. ISSN 1090-7807. doi: 10.1016/j.jmr.2016.12.014. URL https://www.sciencedirect.com/science/article/pii/S1090780716302786.

Gareth A. Morris and Ray Freeman. Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society*, 101(3):760–762, January 1979. ISSN 0002-7863. doi: 10.1021/ja00497a058. URL https://doi.org/10.1021/ja00497a058.

David A. Naylor and Margaret K. Tahic. Apodizing functions for Fourier transform spectroscopy. *Journal of the Optical Society of America A*, 24(11):3644, November 2007. ISSN 1084-7529, 1520-8532. doi: 10.1364/JOSAA.24.003644. URL https://opg.optica.org/abstract.cfm?URI=josaa-24-11-3644.

W. G. Proctor and F. C. Yu. The Dependence of a Nuclear Magnetic Resonance Frequency upon Chemical Compound. *Physical Review*, 77(5):717–717, March 1950. ISSN 0031-899X. doi: 10.1103/PhysRev.77.717. URL https://link.aps.org/doi/10.1103/PhysRev.77.717.

E. M. Purcell, H. C. Torrey, and R. V. Pound. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review*, 69(1-2):37–38, January 1946. doi: 10.1103/PhysRev.69.37. URL https://link.aps.org/doi/10.1103/PhysRev.69.37.

C. V. Raman and K. S. Krishnan. A New Type of Secondary Radiation. *Nature*, 121(3048):501–502, March 1928. ISSN 1476-4687. doi: 10.1038/121501c0. URL https://www.nature.com/articles/121501c0.

Laura Hannemose Rieger, Max Wilson, Tejs Vegge, and Eibar Flores. Understanding the patterns that neural networks learn from chemical spectra. *Digital Discovery*, 2(6):1957–1968, December 2023. ISSN 2635-098X. doi: 10.1039/d3dd00203a. URL https://doi.org/10.1039/d3dd00203a.

Josep Saurí. The Evolving Landscape of NMR Structural Elucidation. *Molecules*, 31(5):888, January 2026. ISSN 1420-3049. doi: 10.3390/molecules31050888. URL https://www.mdpi.com/1420-3049/31/5/888.

Yusen Tan, Yixuan Chen, Zheng Fang, Pan Liu, Yifan Li, Qinyu Guo, Zhedong Lin, Yuqiang Li, Xiangxiang Zeng, Tong Wang, and Jun Xia. Simulation-to-real transfer learning for infrared spectroscopic chemical sensing and analysis from molecules to complex samples, August 2026. URL http://arxiv.org/abs/2608.13341.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains, June 2020. URL http://arxiv.org/abs/2006.10739.

Katharine Turner. Representing Vineyard Modules, July 2023. URL http://arxiv.org/abs/2307.06020.

Dong Wei, Shuo Chen, and Quan Liu. Review of Fluorescence Suppression Techniques in Raman Spectroscopy. *Applied Spectroscopy Reviews*, 50(5):387–406, May 2015. ISSN 0570-4928. doi: 10.1080/05704928.2014.999936. URL https://doi.org/10.1080/05704928.2014.999936.

Yuanxin Xi and David M. Rocke. Baseline Correction for NMR Spectroscopic Metabolomics Data Analysis. *BMC Bioinformatics*, 9(1):324, July 2008. ISSN 1471-2105. doi: 10.1186/1471-2105-9-324. URL https://doi.org/10.1186/1471-2105-9-324.

Ziyu Xiong, Yichi Zhang, Foyez Alauddin, Chu Xin Cheng, Joon An, Mohammad Seyedsayamdost, and Ellen Zhong. Atomic Diffusion Models for Small Molecule Structure Elucidation from NMR Spectra. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 115995–116031. Curran Associates, Inc., 2025. doi: 10.52202/085713-3868. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/a845fdc3f87751710218718adb634fe7-Abstract-Conference.html.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How Powerful are Graph Neural Networks?, February 2019. URL http://arxiv.org/abs/1810.00826.

Wangdong Xu, Byeol Ryu, Anthony Tong, Huanru Henry Mao, Hyunwoo Kim, James Jialun Zhao, Chen Zhang, Yiran Xu, Mingxun Wang, William H. Gerwick, and Garrison W. Cottrell. SPECTRE: A Multimodal Spectral Transformer for Small Molecule Annotation. *Journal of Chemical Information and Modeling*, 66(5):2501–2512, March 2026. ISSN 1549-9596. doi: 10.1021/acs.jcim.5c02444. URL https://doi.org/10.1021/acs.jcim.5c02444.

Chen Yang, Zheng Fang, Hanyu Sun, Fanjie Xu, Hongxin Xiang, Hanyu Gao, Xiangxiang Zeng, Yuqiang Li, Xiaojian Wang, and Jun Xia. A large-scale foundation model enables simulation-to-real adaptation for nuclear magnetic resonance-based molecular structure analysis, June 2026a. URL http://arxiv.org/abs/2606.20756.

Liujia Yang, Zhuo Yang, Jiaqing Xie, Yubin Wang, Ben Gao, Tianfan Fu, Xingjian Wei, Jiaxing Sun, Jiang Wu, Conghui He, Yuqiang Li, and Qinying Gu. NMRTrans: Structure Elucidation from Experimental NMR Spectra via Set Transformers, February 2026b. URL http://arxiv.org/abs/2602.10158.

Qingsong Yang, Binglan Wu, Xuwei Liu, Bo Chen, Wei Li, Gen Long, Xin Chen, and Mingjun Xiao. DiffNMR: Diffusion Models for Nuclear Magnetic Resonance Spectra Elucidation, July 2025. URL http://arxiv.org/abs/2507.08854.

Zhouao Zhou, Xinli Liao, Xu Qiu, Yue Zhang, Jiyang Dong, Xiaobo Qu, and Donghai Lin. NMR-former: A Transformer-Based Deep Learning Framework for Peak Assignment in 1D 1H NMR Spectroscopy. *Analytical Chemistry*, 97(1):904–911, January 2025. ISSN 0003-2700. doi: 10.1021/acs.analchem.4c05632. URL https://doi.org/10.1021/acs.analchem.4c05632.

Federico Zipoli, Marvin Alberts, and Teodoro Laino. IR–NMR Multimodal Computational Spectra Dataset for 177K Patent-Extracted Organic Molecules. *ChemRxiv*, 2025(0619), 2025. doi: 10.26434/chemrxiv-2025-l0zqz. URL https://chemrxiv.org/doi/full/10.26434/chemrxiv-2025-l0zqz.

Supplementary Information

Persistent Homology with Vineyards for Robust Spectral Representation

CONTENTS

A Problem Formulation and Global Notation **18**
A.1 Spectral Observations and Paired Data 18
A.2 Learning Tasks 19
A.3 Nuisance Transformations and Robustness 19

B Spectral Modalities **19**
B.1 One-Dimensional Nuclear Magnetic Resonance Spectroscopy 20
B.2 Two-Dimensional Heteronuclear Single-Quantum Correlation NMR 21
B.3 Molecular Vibrations 22
B.4 Raman Spectroscopy 23
B.5 Infrared Spectroscopy 24

C VINEA: Persistent Scale-Space Spectral Representation **25**
C.1 Persistent Homology of a Sampled Spectrum 25
C.2 Signed Persistent Branches 26
C.3 Geometric Decoration of Persistent Branches 27
C.4 Gaussian Scale-Space Vineyard 28
C.5 Branch Descriptor and Token Set 29
C.6 Permutation-Invariant Spectral Encoder 29
C.7 Computational Structure 31
C.8 Persistent Structure of Ideal First-Order NMR Multiplets . . . 31

D Synthetic Corruption Benchmark **34**
D.1 Evaluation Conditions 34
D.2 One-Dimensional NMR 34
D.3 Two-Dimensional HSQC 35
D.4 Raman Spectroscopy 37
D.5 Infrared Spectroscopy 38
D.6 Scope and Determinism 39

E Datasets and Models **39**
E.1 Tasks and comparison framework 39
E.2 VINEA Benchmark Instantiations 40
E.3 Published ^{1}H NMR Baselines 41
E.4 Published HSQC Baselines 41
E.5 Published Raman Baselines 41
E.6 Published IR Baselines 42
E.7 Internal and Non-Neural Controls 42
E.8 VINEA Architecture Variants 43
E.9 Transfer and Representation Probes 43
E.10 Consolidated results across datasets and modalities 43

F Architectural Ablations **45**

G Training **50**

H Evaluation protocol **51**
H.1 Retrieval pools 51
H.2 Metrics 51
H.3 Experimental and acquisition transfer 52
H.4 Representation-geometry analysis 53
H.5 Robustness–efficiency analysis 54
H.6 Probing protocol 55

I Embedding Geometry over corruptions **55**

Supplementary Table 1: **Notation used throughout the paper and Supplementary Information.** Colors follow the visual semantics of the method figures.

Symbol	Meaning	Symbol	Meaning
Spectral observations			
$m \in \mathcal{M}$	spectral modality	d_m	dimension of the physical coordinate domain
Ω_m	physical coordinate domain	Γ_m	acquisition or resampling grid
N_m	number of sampled grid locations	$f, \mathbf{f}$	scalar field and sampled spectral array
$\mathcal{X}_m$	processed spectra for modality m	$\mathcal{K}_m$	cubical complex induced by the grid
t	superlevel filtration threshold		
Persistent topology			
p	persistent H_0 branch	b_p, d_p	birth and death values
π_p	persistence $b_p - d_p$	s_p	branch orientation in $\{+1, -1\}$
C_p	branch component before its merge	τ_p	topology $[b_p, d_p, \pi_p, s_p]$
Geometric decoration			
c_p	physical position or centroid	A_p	integrated mass above the death level
$C_p^{(\ell)}$	component at relative depth ℓ	$w_p^{(\ell)}$	prominence-relative width or extent
asym_p	half-persistence asymmetry	g_p	local geometric descriptor
ΔV_m	physical measure of one grid cell		
Scale-space vineyard			
σ	Gaussian smoothing scale	Σ_m	modality-specific scale ladder
f_σ	field smoothed at scale σ	$r_p(\sigma)$	bounded relative-persistence score
P_p	sampled vineyard profile	$\sigma_p^{\mathrm{death}}$	largest sampled scale where p remains active
v_p	vineyard descriptor $[P_p, \sigma_p^{\mathrm{death}}]$		
Tokens and coordinates			
x_p	branch descriptor $[\tau_p \Vert c_p \Vert g_p \Vert v_p]$	$\mathcal{T}(f)$	unordered branch-token set
$M, M_{\max}$	retained branches and token budget	m_p	token validity mask
ξ_p	normalized absolute coordinate $\psi_m(c_p)$	Λ_m	physical coordinate normalization matrix
Δ_{pq}	normalized displacement $\Lambda_m^{-1}(c_p - c_q)$		
Learned representations			
E_θ	spectral encoder	G_ϕ	molecular encoder for retrieval
$z_f, \widehat{z}_f$	spectral embedding and normalized form	$z_y, \widehat{z}_y$	molecular embedding and normalized form
D	embedding dimension	τ	contrastive temperature
Robustness benchmark			
$a \in \mathcal{A}_m$	corruption family for modality m	$\eta \in \Xi_{m,a}$	corruption parameters
$\Pi_{m,a,s}$	parameter distribution at severity s	$\mathcal{C}_{m,a,\eta}$	label-preserving corruption operator
d_{emb}	clean–corrupted embedding displacement	$\mathcal{R}_{\mathrm{emb}}$	expected embedding drift
$P_h(s)$	task score of model h at severity s	$\rho_h(s)$	retention $P_h(s)/P_h(0)$

A Problem Formulation and Global Notation

This appendix defines notation shared across modalities, learning tasks, and robustness experiments. Modality-specific acquisition physics is described in Appendix B, the complete VINEA construction in Appendix C, synthetic corruption operators in Appendix D, and optimization objectives in Appendix G.

A.1 Spectral Observations and Paired Data

Let $\mathcal{M}$ denote the set of spectral modalities and $m \in \mathcal{M}$ a particular modality. Its calibrated coordinate domain is $\Omega_m \subset \mathbb{R}^{d_m}$, with $d_m = 1$ for ^{1}H NMR, Raman, and IR and $d_m = 2$ for HSQC. A processed spectrum is represented conceptually as a scalar field $f : \Omega_m \to \mathbb{R}$ sampled on a finite acquisition or resampling grid $\Gamma_m = \{x_j\}_{j=1}^{N_m}$. The corresponding discrete observation is $\mathbf{f} = (f(x_1), \ldots, f(x_{N_m}))$. For multidimensional spectra, j indexes the flattened Cartesian grid. We distinguish physical coordinates $x \in \Omega_m$ from array indices throughout.

A supervised observation is a triple (f, m, y), where $y \in \mathcal{Y}$ denotes the associated semantic target. Depending on the task, y may represent a molecular graph, a class label, or a vector of molecular attributes. A dataset is

$$\mathcal{D} = \{(f_i, m_i, y_i)\}_{i=1}^{N}. \tag{A.1}$$

Multiple spectral observations may correspond to the same molecule, so the measurement instance f_i and semantic target y_i are treated as distinct objects.

A.2 Learning Tasks

The spectral encoder is written abstractly as

$$z_i = E_\theta(f_i, m_i) \in \mathbb{R}^D, \qquad \widehat{z}_i = \frac{z_i}{\|z_i\|_2}. \tag{A.2}$$

For VINEA, E_θ operates on the persistent branch set defined in Appendix C, but the benchmark notation below applies equally to dense and structured baselines.

For spectrum-to-molecule retrieval, a molecular encoder $G_\phi : \mathcal{Y} \to \mathbb{R}^D$ produces $\widehat{z}_y = G_\phi(y)/\|G_\phi(y)\|_2$. Spectrum–molecule similarity is cosine similarity,

$$\mathrm{sim}(f, y) = \widehat{z}_f^\top \widehat{z}_y. \tag{A.3}$$

Given a candidate set $\mathcal{Y}_{\mathrm{cand}}$ containing the correct molecule y_i, its retrieval rank is

$$\mathrm{rank}(f_i, y_i) = 1 + \sum_{y \in \mathcal{Y}_{\mathrm{cand}} \setminus \{y_i\}} 1[\mathrm{sim}(f_i, y) > \mathrm{sim}(f_i, y_i)]\,. \tag{A.4}$$

Recall@K is the fraction of queries satisfying $\mathrm{rank}(f_i, y_i) \le K$. The principal retrieval metric is Recall@1.

For predictive tasks, a task-specific head H_ω acts on $\widehat{z}_f$. In the IR functional-group experiment, $H_\omega(\widehat{z}_f) \in [0,1]^C$ predicts $C = 17$ binary labels and performance is summarized by macro-AUROC. Training objectives and optimization details are given in Appendix G.

A.3 Nuisance Transformations and Robustness

For modality m, let $\mathcal{A}_m$ denote the set of corruption families. A corruption $a \in \mathcal{A}_m$ with random parameters η acts through

$$\widetilde{f} = \mathcal{C}_{m,a,\eta}[f]. \tag{A.5}$$

At severity level $s \in \{1, \dots, 5\}$, parameters are sampled from the modality- and corruption-specific distribution $\eta \sim \Pi_{m,a,s}$. The clean condition is represented by the identity transformation. All benchmark corruptions are designed to preserve the semantic target,

$$y(\mathcal{C}_{m,a,\eta}[f]) = y(f), \tag{A.6}$$

so robustness refers specifically to nuisance variation rather than invariance to arbitrary changes in spectral structure. Let $P_h(0)$ denote the clean task performance of model h and $P_h(s)$ its performance under corruption severity s. We report absolute task performance as the primary quantity. When normalized degradation is useful, retention is

$$R_h(s) = \frac{P_h(s)}{P_h(0)}. \tag{A.7}$$

Representation drift between a clean observation and its corrupted counterpart is measured by cosine displacement,

$$d_{\mathrm{emb}}(f, \widetilde{f}) = 1 - \widehat{z}_f^\top \widehat{z}_{\widetilde{f}}, \tag{A.8}$$

where smaller values indicate greater stability of the learned representation. Corruption definitions, severity parameters, and combined corruption protocols are specified in Appendix D.

B Spectral Modalities

The four modalities used in this work produce sampled scalar fields but arise from different microscopic interactions. One-dimensional NMR measures the coherent response of nuclear magnetic moments to an external magnetic field. HSQC extends NMR to a two-dimensional experiment in which scalar spin coupling transfers coherence between two nuclear species. Raman spectroscopy measures inelastic photon scattering by molecular vibrations. Infrared spectroscopy measures resonant absorption by vibrational transitions. Their common mathematical form is a measured field $f : \Omega_m \subset \mathbb{R}^{d_m} \to \mathbb{R}$, with $d_m = 1$ for ^{1}H NMR, Raman, and IR and $d_m = 2$ for HSQC. The coordinate has a direct physical interpretation, whereas the measured intensity depends on molecular structure together with concentration, relaxation, excitation conditions, instrument response, and signal processing (Bloch et al., 1946; Purcell et al., 1946; Bodenhausen & Ruben, 1980; Raman & Krishnan, 1928).

Supplementary Table 2: **Spectral modalities considered in this work.** Peak coordinates and intensities arise from different physical interactions even though all four measurements are represented as scalar fields.

Modality	Coordinate	Interaction	Structural information
^{1}H NMR	chemical shift or frequency	nuclear spin resonance	local electronic environment and J coupling
HSQC	(δ_H, δ_X)	heteronuclear coherence transfer	directly coupled nuclear environments
Raman	Raman shift (cm^{-1})	inelastic light scattering	polarizability-active vibrations
IR/FTIR	wavenumber (cm^{-1})	resonant dipole absorption	dipole-active vibrations

B.1 One-Dimensional Nuclear Magnetic Resonance Spectroscopy

B.1.1 Nuclear spin and the Zeeman interaction

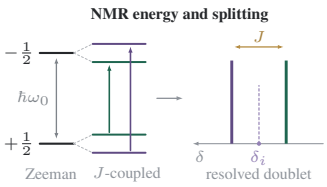
Atomic nuclei with nonzero spin possess intrinsic angular momentum and an associated magnetic moment. For a nucleus with spin operator $\hat{\mathbf{I}}$, the magnetic moment is $\hat{\boldsymbol{\mu}} = \gamma\hbar\hat{\mathbf{I}}$, where γ is the isotope-specific gyromagnetic ratio. Placing the nucleus in a homogeneous static magnetic field $\mathbf{B}_0 = B_0\mathbf{e}_z$ gives the Zeeman Hamiltonian

$$\hat{H}_Z = -\hat{\boldsymbol{\mu}} \cdot \mathbf{B}_0 = -\gamma\hbar B_0 \hat{I}_z. \tag{B.1}$$

For a spin-$1/2$ nucleus, the allowed projections are $m_I = \pm 1/2$. Their energy separation has magnitude $\Delta E = \hbar\omega_0$, where

$$\omega_0 = |\gamma| B_0 \tag{B.2}$$

is the Larmor angular frequency. A transverse radiofrequency field near $\omega_0/(2\pi)$ can therefore drive transitions between the two spin states (Bloch et al., 1946; Purcell et al., 1946; Levitt, 2008).



Supplementary Figure 6: **Energy and frequency scales in NMR.** The external field produces the Zeeman splitting. Electronic shielding determines the resonance position, while scalar coupling generates local line splittings separated by J.

At thermal equilibrium the lower-energy state is slightly more populated. For a spin-$1/2$ ensemble,

$$P = \tanh\left(\frac{\hbar\omega_0}{2k_{\mathrm{B}}T}\right) \simeq \frac{\hbar\omega_0}{2k_{\mathrm{B}}T}, \tag{B.3}$$

where the approximation applies when $\hbar\omega_0 \ll k_{\mathrm{B}}T$. The small equilibrium polarization explains the intrinsically weak signal of conventional solution NMR.

B.1.2 Electronic shielding and chemical shift

If all nuclei of the same isotope experienced exactly the external field B_0, they would resonate at nearly the same frequency and would carry little chemical information. The surrounding electrons respond to the applied magnetic field and modify the field at each nucleus. This response is described by the shielding tensor $\boldsymbol{\sigma}_i$. In an isotropic liquid, rapid molecular tumbling averages its anisotropic components and only the isotropic shielding $\sigma_i^{\mathrm{iso}} = \mathrm{Tr}(\boldsymbol{\sigma}_i)/3$ remains. To first order, the resonance frequency is

$$\nu_i \simeq \frac{|\gamma_i| B_0}{2\pi}\left(1 - \sigma_i^{\mathrm{iso}}\right). \tag{B.4}$$

Because the absolute resonance frequency scales with magnetic field, NMR spectra are normally reported using the dimensionless chemical shift

$$\delta_i = 10^6 \frac{\nu_i - \nu_{\mathrm{ref}}}{\nu_0}\ \mathrm{ppm}, \tag{B.5}$$

where ν_{ref} is the reference resonance and ν_0 is the nominal spectrometer frequency (Proctor & Yu, 1950). Chemical shifts in ppm are approximately independent of B_0. A chemical-shift difference $\Delta\delta$ expressed numerically in ppm corresponds to a frequency separation

$$\Delta\nu = \nu_0\, 10^{-6} \Delta\delta \tag{B.6}$$

in hertz. This distinction is important for the representation because absolute chemical position and local frequency separations obey different transformation laws when the magnetic field changes.

B.1.3 SCALAR SPIN COUPLING AND MULTIPLETS

Nuclear spins in the same molecule also interact indirectly through the electronic structure. In an isotropic liquid, the scalar coupling between spins i and j is described by

$$\hat{H}_J = 2\pi\hbar J_{ij}\hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j, \tag{B.7}$$

where J_{ij} is measured in hertz. In the high-field weak-coupling regime, where the resonance-frequency separation is much larger than the coupling, the secular approximation gives $\hat{H}_J \simeq 2\pi\hbar J_{ij}\hat{I}_{iz}\hat{I}_{jz}$ (Levitt, 2008; Ernst, 1992).

Coupling to one spin-$1/2$ partner splits a resonance into a doublet with separation J. Coupling to n magnetically equivalent spin-$1/2$ partners produces an $(n+1)$-line first-order multiplet with adjacent separation J and relative amplitudes

$$a_k = 2^{-n}\binom{n}{k}, \qquad k = 0, \ldots, n. \tag{B.8}$$

Unlike chemical-shift separations expressed in hertz, scalar couplings are approximately independent of B_0. Multiple inequivalent couplings produce nested splitting patterns. Strong coupling, chemical exchange, unresolved lines, and finite linewidth can invalidate the ideal first-order description.

B.1.4 TIME-DOMAIN DETECTION AND LINEWIDTH

A radiofrequency pulse creates transverse magnetization that precesses in the transverse plane and induces a voltage in the receiver coil. A simplified complex free-induction decay is

$$s(t) = \sum_{k=1}^{K} A_k e^{-t/T_{2,k}^*} e^{i(2\pi\nu_k t + \phi_k)} + \epsilon(t), \tag{B.9}$$

where A_k is the complex amplitude, ν_k the resonance frequency, ϕ_k the phase, $T_{2,k}^*$ the effective transverse coherence time, and $\epsilon(t)$ acquisition noise. The frequency-domain spectrum is obtained by Fourier transformation,

$$S(\nu) = \int_0^\infty s(t)e^{-i2\pi\nu t}\, dt. \tag{B.10}$$

For a single exponentially damped coherence, the absorption line is Lorentzian with full width at half maximum $\Gamma = 1/(\pi T_2^*)$ (Ernst, 1992). The effective T_2^* includes intrinsic transverse relaxation and static field inhomogeneity.

Observed NMR intensity is therefore not an intrinsic molecular constant. It also depends on concentration, excitation conditions, longitudinal recovery, transverse relaxation, pulse-sequence timing, receiver gain, and processing. Resonance location and local splitting are generally more directly interpretable as structural coordinates than raw peak amplitude.

B.2 TWO-DIMENSIONAL HETERONUCLEAR SINGLE-QUANTUM CORRELATION NMR

B.2.1 FROM ONE-DIMENSIONAL FREQUENCY ENCODING TO TWO DIMENSIONS

A one-dimensional NMR experiment directly records one time coordinate and Fourier transforms it into one frequency coordinate. A two-dimensional experiment introduces a second evolution time t_1. The pulse sequence is repeated many times while incrementing t_1, producing a signal $s(t_1, t_2)$. Fourier transformation with respect to both variables yields a two-dimensional spectrum. The directly detected coordinate originates from t_2, while the indirectly encoded coordinate originates from the repeated t_1 evolution (Ernst, 1992).

HSQC uses this construction to correlate the chemical shift of a sensitive nucleus, usually ^{1}H, with the chemical shift of a directly coupled heteronucleus such as ^{13}C or ^{15}N. The resulting spectrum contains localized cross-peaks whose coordinates identify pairs of coupled nuclear environments (Bodenhausen & Ruben, 1980).

B.2.2 HETERONUCLEAR SPIN HAMILTONIAN AND COHERENCE TRANSFER

Consider a weakly coupled spin-$1/2$ pair I and S. In a doubly rotating frame, the internal Hamiltonian can be written

$$\frac{\hat{H}}{\hbar} = \Omega_I \hat{I}_z + \Omega_S \hat{S}_z + 2\pi J_{IS} \hat{I}_z \hat{S}_z, \tag{B.11}$$

where Ω_I and Ω_S are carrier-relative angular-frequency offsets and J_{IS} is the heteronuclear scalar coupling. The offset terms generate chemical-shift evolution. The coupling term converts single-spin transverse coherence into correlated two-spin coherence. Under pure scalar-coupling evolution,

$$\hat{I}_x \longrightarrow \hat{I}_x \cos(\pi J_{IS} t) + 2\hat{I}_y \hat{S}_z \sin(\pi J_{IS} t). \tag{B.12}$$

This relation is the basis of INEPT-type polarization transfer (Morris & Freeman, 1979). Delays near $\Delta = 1/(4J_{IS})$ allow radiofrequency pulses and coupling evolution to convert proton coherence into heteronuclear coherence efficiently for the target coupling range.

A simplified proton-detected HSQC pathway can therefore be summarized as

^{1}H $\longrightarrow$ INEPT transfer $\longrightarrow$ X coherence during t_1 $\longrightarrow$ reverse INEPT $\longrightarrow$ ^{1}H detection during t_2.

The indirect period encodes the heteronuclear frequency and the direct period encodes the proton frequency.

B.2.3 TWO-DIMENSIONAL SIGNAL AND CROSS-PEAKS

For isolated correlations, a simplified complex time-domain signal is

$$s(t_1, t_2) \approx \sum_{k=1}^{K} A_k e^{-\lambda_{1,k} t_1 - \lambda_{2,k} t_2} e^{i(\Omega_{X,k} t_1 + \Omega_{H,k} t_2 + \phi_k)}, \tag{B.13}$$

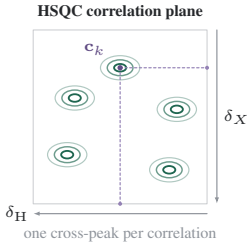
where $\lambda_{1,k}$ and $\lambda_{2,k}$ represent effective coherence-loss rates during the two evolution dimensions. The two-dimensional Fourier transform is

$$S(\omega_X, \omega_H) = \iint s(t_1, t_2) e^{-i(\omega_X t_1 + \omega_H t_2)}\, dt_1\, dt_2. \tag{B.14}$$

An isolated correlation is therefore localized near

$$\mathbf{c}_k = (\delta_{H,k}, \delta_{X,k}). \tag{B.15}$$

Cross-peak position encodes the paired proton and heteronuclear environments. Cross-peak width depends on relaxation, acquisition duration, sampling, apodization, and magnetic-field homogeneity. Cross-peak intensity additionally depends on transfer efficiency, the value of J_{IS}, pulse imperfections, relaxation during the sequence, and concentration. An HSQC spectrum is consequently a physically calibrated two-dimensional scalar field, not a generic image.



Supplementary Figure 7: **HSQC as a two-dimensional spectral field.** Each resolved heteronuclear correlation produces a localized cross-peak centered near (δ_H, δ_X). The two axes have distinct physical meanings and are not interchangeable spatial coordinates.

B.3 MOLECULAR VIBRATIONS

Raman and infrared spectroscopy interrogate the same molecular vibrational degrees of freedom but couple to them through different operators. Their common starting point is the Born–Oppenheimer approximation, in which nuclear motion is treated on an electronic potential-energy surface $V(\mathbf{R})$ (Born & Oppenheimer, 1927).

Let $\mathbf{R}_0$ be an equilibrium nuclear geometry. Expanding the potential around this minimum gives

$$V(\mathbf{R}_0 + \Delta\mathbf{R}) = V_0 + \frac{1}{2}\Delta\mathbf{R}^\top \mathbf{H} \Delta\mathbf{R} + \mathcal{O}(\|\Delta\mathbf{R}\|^3), \tag{B.16}$$

where $\mathbf{H}$ is the Cartesian Hessian of the potential at equilibrium. Let $\mathbf{M}$ be the diagonal matrix of nuclear masses. The mass-weighted Hessian is $\mathbf{F} = \mathbf{M}^{-1/2}\mathbf{H}\mathbf{M}^{-1/2}$, and its eigenvectors satisfy

$$\mathbf{F}\mathbf{l}_k = \omega_k^2 \mathbf{l}_k. \tag{B.17}$$

The eigenvector $\mathbf{l}_k$ defines normal coordinate Q_k, while ω_k is its harmonic angular frequency. A nonlinear molecule containing N atoms has $3N-6$ vibrational normal modes after removing overall translation and rotation. A linear molecule has $3N-5$.

Within the harmonic approximation, mode k has energy levels

$$E_{v_k} = \hbar\omega_k \left(v_k + \frac{1}{2}\right), \qquad v_k = 0, 1, 2, \ldots. \tag{B.18}$$

Real molecular potentials are anharmonic. Anharmonicity shifts fundamental frequencies and permits overtones, combination bands, and resonant interactions between vibrational states (Barone et al., 2014).

B.4 Raman Spectroscopy

B.4.1 Inelastic light scattering

Raman spectroscopy measures the small fraction of incident light that is scattered with a changed photon energy (Raman & Krishnan, 1928). Let an incident photon have angular frequency ω_L. If the molecule changes from state $|i\rangle$ to state $|f\rangle$, conservation of energy gives

$$E_i + \hbar\omega_L = E_f + \hbar\omega_s, \tag{B.19}$$

where ω_s is the angular frequency of the scattered photon.

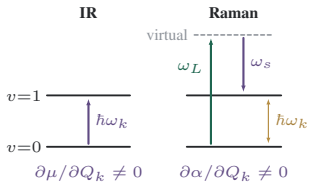
For Stokes scattering, the molecule gains one vibrational quantum in mode k, so $E_f - E_i \simeq \hbar\omega_k$ and

$$\omega_s = \omega_L - \omega_k. \tag{B.20}$$

The experimentally reported coordinate is the Raman shift

$$\Delta\widetilde{\nu} = \frac{\omega_L - \omega_s}{2\pi c} \simeq \frac{\omega_k}{2\pi c}, \tag{B.21}$$

usually expressed in cm^{-1}. A Raman-band position therefore reports a molecular energy difference rather than the absolute optical frequency of the laser.



Supplementary Figure 8: **IR and Raman probe the same vibrational levels through different interactions.** IR absorption requires a changing molecular dipole along the normal mode. Raman scattering requires a changing polarizability.

B.4.2 Scattering amplitude and polarizability

Raman scattering is a second-order light-matter process. In electric-dipole perturbation theory, the Kramers–Heisenberg–Dirac amplitude contains a sum over intermediate molecular states $|n\rangle$ of the form

$$\mathcal{A}_{fi} \propto \sum_n \left[\frac{\langle f|\hat{\mu}_a|n\rangle\langle n|\hat{\mu}_b|i\rangle}{E_n - E_i - \hbar\omega_L - i\Gamma_n} + \frac{\langle f|\hat{\mu}_b|n\rangle\langle n|\hat{\mu}_a|i\rangle}{E_n - E_i + \hbar\omega_s + i\Gamma_n} \right], \tag{B.22}$$

where Γ_n describes the finite lifetime of the intermediate response (Long, 2002). The often-drawn virtual level is therefore not generally a stationary molecular eigenstate.

Away from electronic resonance, the response can be summarized by the molecular polarizability tensor $\boldsymbol{\alpha}$. An electric field induces the dipole $\mathbf{p}_{\mathrm{ind}} = \boldsymbol{\alpha}\mathbf{E}$. Expanding the polarizability along normal coordinate Q_k gives

$$\boldsymbol{\alpha}(Q_k) = \boldsymbol{\alpha}_0 + \left(\frac{\partial\boldsymbol{\alpha}}{\partial Q_k}\right)_0 Q_k + \cdots. \tag{B.23}$$

Within the double-harmonic Placzek approximation, a fundamental vibration is Raman active when

$$\left(\frac{\partial\boldsymbol{\alpha}}{\partial Q_k}\right)_0 \neq 0. \tag{B.24}$$

For incident and detected polarization vectors $\mathbf{e}_i$ and $\mathbf{e}_s$, the Stokes intensity has the schematic dependence

$$I_k^{\mathrm{R}} \propto \omega_s^4 (n_k + 1) \left| \mathbf{e}_s^{\top} \left(\frac{\partial\boldsymbol{\alpha}}{\partial Q_k}\right)_0 \mathbf{e}_i \right|^2, \tag{B.25}$$

where n_k is the thermal occupation of mode k (Long, 2002). Raman intensity therefore depends on the molecular polarizability response together with laser wavelength, optical polarization, temperature, sample orientation, laser power, and collection geometry.

B.5 Infrared Spectroscopy

B.5.1 Electric-dipole absorption

Infrared spectroscopy measures resonant absorption of electromagnetic radiation by molecular vibrations. In the electric-dipole approximation, the interaction Hamiltonian is

$$\hat{H}'(t) = -\hat{\boldsymbol{\mu}} \cdot \mathbf{E}(t), \tag{B.26}$$

where $\hat{\boldsymbol{\mu}}$ is the molecular electric-dipole operator. For a monochromatic field with polarization $\mathbf{e}$, time-dependent perturbation theory gives a transition rate proportional to

$$W_{i\to f} \propto \left|\mathbf{e} \cdot \langle f|\hat{\boldsymbol{\mu}}|i\rangle\right|^2 \delta(E_f - E_i - \hbar\omega). \tag{B.27}$$

Absorption is therefore resonant. The photon energy must match the molecular transition energy.

Expanding the permanent dipole along normal coordinate Q_k gives

$$\boldsymbol{\mu}(Q_k) = \boldsymbol{\mu}_0 + \left(\frac{\partial \boldsymbol{\mu}}{\partial Q_k}\right)_0 Q_k + \cdots. \tag{B.28}$$

For a harmonic oscillator, the coordinate matrix element is nonzero for $\Delta v = \pm 1$. Under the double-harmonic approximation, a fundamental mode is therefore infrared active when

$$\left(\frac{\partial \boldsymbol{\mu}}{\partial Q_k}\right)_0 \neq 0. \tag{B.29}$$

Its integrated intensity scales as

$$I_k^{\mathrm{IR}} \propto \left|\left(\frac{\partial \boldsymbol{\mu}}{\partial Q_k}\right)_0\right|^2. \tag{B.30}$$

Anharmonicity and higher-order dipole derivatives permit transitions beyond the ideal fundamental selection rule (Barone et al., 2014).

Raman and IR thus probe the same normal modes through different molecular response functions. Raman activity depends on the derivative of polarizability with respect to the normal coordinate. IR activity depends on the derivative of the permanent dipole. The two modalities are therefore complementary. For a centrosymmetric molecule, the ideal mutual-exclusion rule gives an important limiting example in which gerade fundamental modes can be Raman active and ungerade fundamental modes can be IR active.

B.5.2 Wavenumber, transmission, and absorbance

Infrared spectra are conventionally parameterized by optical wavenumber

$$\widetilde{\nu} = \frac{\nu}{c} = \frac{1}{\lambda}, \tag{B.31}$$

usually in cm^{-1}. A harmonic normal mode with angular frequency ω_k corresponds to $\widetilde{\nu}_k = \omega_k/(2\pi c)$.

In a transmission experiment, $I_0(\widetilde{\nu})$ is the reference intensity and $I(\widetilde{\nu})$ is the transmitted intensity. The transmittance and absorbance are

$$T(\widetilde{\nu}) = \frac{I(\widetilde{\nu})}{I_0(\widetilde{\nu})}, \qquad A(\widetilde{\nu}) = -\log_{10} T(\widetilde{\nu}). \tag{B.32}$$

Under the ideal Beer–Lambert approximation,

$$A(\widetilde{\nu}) = \varepsilon(\widetilde{\nu}) c \ell, \tag{B.33}$$

where ε is the molar attenuation coefficient, c is concentration, and ℓ is optical path length. The relation is not universally exact. Scattering, reflection, local-field effects, heterogeneous samples, finite spectral resolution, and other optical effects can modify measured band intensities and shapes (Mayerhöfer et al., 2020).

B.5.3 FOURIER-TRANSFORM INFRARED SPECTROSCOPY

Most modern mid-infrared measurements are obtained with Fourier-transform infrared spectroscopy. A broadband source passes through an interferometer, commonly a Michelson interferometer. Different optical path differences modulate the spectral components with different phases. If x denotes optical path difference, the modulated part of an ideal interferogram can be represented as

$$I_{\mathrm{mod}}(x) = \int_0^\infty S(\widetilde{\nu}) \cos(2\pi\widetilde{\nu}x)\, d\widetilde{\nu}. \tag{B.34}$$

The frequency-domain spectrum is recovered by Fourier transformation of the interferogram. Finite maximum optical path difference limits spectral resolution and determines the instrumental line shape. Apodization modifies this line shape by reducing truncation artifacts at the cost of spectral resolution.

The measured IR field is therefore shaped jointly by molecular vibrational transition energies, dipole derivatives, sample concentration, optical geometry, atmospheric absorption, instrumental resolution, and interferogram processing. As in NMR and Raman spectroscopy, spectral position and measured amplitude should not be interpreted as equivalent types of information.

C VINEA: PERSISTENT SCALE-SPACE SPECTRAL REPRESENTATION

VINEA maps a sampled spectral field to an unordered set of persistent multiscale objects. The construction is shared by one-dimensional NMR, Raman, and IR spectra and by two-dimensional HSQC fields. Zeroth persistent homology first decomposes the field into connected spectral structures ordered by prominence. Each structure is then decorated with physical geometry and followed through Gaussian scale space. The resulting persistent branches are converted to tokens and encoded by a permutation-invariant attention network. This appendix gives the complete mathematical construction. The physical interpretation of each spectral coordinate is given in Appendix B and the modality-specific preprocessing and corruption models are described in Appendix D.

C.1 PERSISTENT HOMOLOGY OF A SAMPLED SPECTRUM

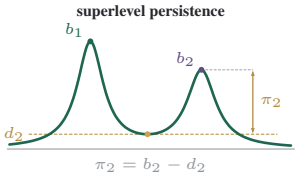
C.1.1 FROM THRESHOLDING TO PERSISTENCE

Persistent homology can be viewed as a systematic way of recording how structures appear and merge while a scalar field is thresholded. Consider first a one-dimensional spectrum. At a high intensity threshold, only neighborhoods of the strongest local maxima are retained. Lowering the threshold expands these neighborhoods. A new connected component appears when the threshold passes a local maximum and two components merge when it passes the minimum separating them. Persistent homology records the threshold at which each component appears, the threshold at which it merges into an older component, and the difference between the two. In this work only zeroth homology H_0 is required. Its rank is the number of connected components, so no higher-dimensional topological interpretation is needed for the spectral representation.

Let $m \in \mathcal{M}$ denote the modality, let $\Omega_m \subset \mathbb{R}^{d_m}$ be its physical coordinate domain, and let $f : \Omega_m \to \mathbb{R}$ be the processed and amplitude-normalized spectral field sampled on a regular grid. The dimensionality is $d_m = 1$ for ^{1}H NMR, Raman, and IR and $d_m = 2$ for HSQC.

C.1.2 CUBICAL SUPERLEVEL FILTRATION

The sampling grid induces a cubical complex $\mathcal{K}_m$ whose vertices are sampled coordinates and whose higher-dimensional cells connect adjacent samples (Kaczynski et al., 2004). In one dimension these cells are vertices and intervals. In two dimensions they are vertices, edges, and pixels. Sampled values are defined on vertices. To obtain a valid superlevel filtration, the value of a cell $\kappa \in \mathcal{K}_m$ is extended by the upper-star rule



$$f(\kappa) = \min_{v \prec \kappa} f(v), \tag{C.1}$$

where $v \prec \kappa$ ranges over the vertices of κ. A cell therefore enters only after all of its vertices have entered, which guarantees that every thresholded object is a subcomplex.

For threshold $t \in \mathbb{R}$, the superlevel subcomplex is

$$\mathcal{K}_m^t = \{\kappa \in \mathcal{K}_m : f(\kappa) \geq t\}. \tag{C.2}$$

If $t_1 > t_2$, then $\mathcal{K}_m^{t_1} \subseteq \mathcal{K}_m^{t_2}$. Decreasing t therefore gives the nested sequence

$$\varnothing \subseteq \mathcal{K}_m^{t_1} \subseteq \mathcal{K}_m^{t_2} \subseteq \cdots \subseteq \mathcal{K}_m^{t_L}.$$

The inclusions induce maps $H_0(\mathcal{K}_m^{t_1}) \to H_0(\mathcal{K}_m^{t_2})$, and the collection of spaces and maps is the zeroth persistence module (Edelsbrunner et al., 2002; Chazal et al., 2013). Equal filtration values are resolved with a fixed deterministic ordering compatible with the cubical filtration.

C.1.3 Birth, death, and the Elder Rule

In one dimension, every local maximum generates a connected component. Suppose a component generated by maximum p appears at threshold b_p. As the threshold decreases, the component grows until it first merges with a component generated by a higher maximum. Under the Elder Rule, the component with larger birth value survives and the younger component terminates. Its merge threshold is its death value d_p, and its persistence is

$$\pi_p = b_p - d_p > 0. \tag{C.3}$$

We denote the set of persistent branches by $\mathcal{B}(f)$. Each finite branch $p \in \mathcal{B}(f)$ corresponds to the persistence-diagram point (b_p, d_p), giving

$$\mathrm{Dgm}_0(f) = \{(b_p, d_p) : p \in \mathcal{B}(f)\}. \tag{C.4}$$

Because we use a descending superlevel filtration, $b_p > d_p$. Equivalently, one may compute ordinary sublevel persistence on $-f$ and transform the resulting coordinates back to the intensity scale of f.

For a one-dimensional field, π_p is the topographic prominence of the corresponding finite maximum. The component generated by the global maximum never merges into an older component and is therefore the unique essential H_0 class. VINEA assigns this class the finite death value $t_{\min}$ used as the lower filtration boundary so that it can be represented by the same finite descriptor as the other branches. In two dimensions, the construction is unchanged, although connected components correspond to cross-peak regions rather than one-dimensional intervals.

C.1.4 Stability

Persistence diagrams are stable with respect to perturbations of the scalar field. For two tame fields f and g on the same finite cubical complex,

$$d_B(\mathrm{Dgm}_0(f), \mathrm{Dgm}_0(g)) \leq \|f - g\|_\infty, \tag{C.5}$$

where d_B is the bottleneck distance (Cohen-Steiner et al., 2007). Hence an intensity perturbation bounded by ε can move the persistence diagram by at most ε. This statement concerns the topological descriptor itself. It does not imply stability of the subsequent geometric measurements, branch truncation, neural attention, or downstream prediction.

C.2 Signed Persistent Branches

Spectral processing can generate informative structures of either orientation. VINEA therefore applies the same superlevel construction to two oriented versions of the field. We estimate

$$\beta = \min(0, Q_{0.25}(f)) \tag{C.6}$$

and define

$$f_s = s(f - \beta), \qquad s \in \{+1, -1\}. \tag{C.7}$$

Branches are extracted independently from f_{+1} and f_{-1}. Each branch carries the orientation s_p of the field from which it was obtained, and every geometric quantity associated with that branch is evaluated on f_{s_p}.

The clipping in Equation (C.6) leaves non-negative baseline estimates unchanged at zero and corrects negative baseline excursions. If two spectra satisfy $\|f - g\|_\infty \leq \varepsilon$, then the empirical quantiles satisfy $|Q_{0.25}(f) - Q_{0.25}(g)| \leq \varepsilon$. Since $x \mapsto \min(0, x)$ is non-expansive, $|\beta_f - \beta_g| \leq \varepsilon$, which gives $\|(f - \beta_f) - (g - \beta_g)\|_\infty \leq 2\varepsilon$. The standard bottleneck bound therefore applies directly to each already-oriented field and gives the conservative bound

$$d_B(\mathrm{Dgm}_0(f_s), \mathrm{Dgm}_0(g_s)) \leq 2\varepsilon \tag{C.8}$$

when the baseline estimate is recomputed independently for the two spectra.

C.3 Geometric Decoration of Persistent Branches

The persistence pair (b_p, d_p) records prominence but not physical location or shape. VINEA therefore augments every branch with measurements defined on its branch-owned connected component immediately before the merge event at d_p. We denote this component by C_p.

C.3.1 Position and integrated branch mass

For a one-dimensional spectrum, c_p is the physical coordinate of the local maximum that generates branch p. For a two-dimensional field, $c_p = (c_{p,1}, c_{p,2})$ is the centroid used for the corresponding connected region. All coordinates are expressed in physical rather than array units.

Let ΔV_m denote the physical measure associated with one grid cell. In one dimension this is the coordinate spacing. In two dimensions it is the corresponding cell area. The integrated elevation of branch p above its merge level is

$$A_p = \Delta V_m \sum_{x_i \in C_p} \big(f_{s_p}(x_i) - d_p\big)_+, \qquad (a)_+ = \max(a, 0). \tag{C.9}$$

The quantity A_p measures branch mass relative to the level at which the component merges. It is a geometric signal descriptor and is not assumed to be universally proportional to concentration, spin count, or molecular abundance.

C.3.2 Prominence-relative geometry

Absolute intensity thresholds would make widths strongly dependent on branch amplitude. We instead measure geometry at levels defined relative to the branch's own persistence. For $\ell \in (0, 1)$, define

$$t_p^{(\ell)} = b_p - \ell \pi_p, \qquad C_p^{(\ell)} = \{x_i \in C_p : f_{s_p}(x_i) \geq t_p^{(\ell)}\}. \tag{C.10}$$

The implementation evaluates $\ell \in \{0.25, 0.50, 0.75\}$. In one dimension,

$$w_p^{(\ell)} = \max_{x_i \in C_p^{(\ell)}} x_i - \min_{x_i \in C_p^{(\ell)}} x_i \tag{C.11}$$

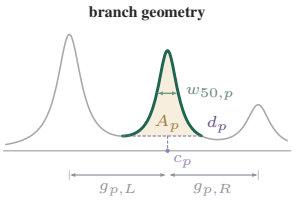
is the physical width at that prominence-relative level. We write these values as $w_{25,p}$, $w_{50,p}$, and $w_{75,p}$. A scale-free broad-tail statistic is

$$\text{shape}_p = \frac{w_{75,p}}{w_{50,p} + \varepsilon_{\text{num}}}, \tag{C.12}$$

where $\varepsilon_{\text{num}} > 0$ prevents division by zero.

At half persistence, let $x_{p,L}^{(1/2)}$ and $x_{p,R}^{(1/2)}$ denote the left and right boundary coordinates of $C_p^{(1/2)}$. The signed asymmetry is

$$\text{asym}_p = \frac{(x_{p,R}^{(1/2)} - c_p) - (c_p - x_{p,L}^{(1/2)})}{x_{p,R}^{(1/2)} - x_{p,L}^{(1/2)} + \varepsilon_{\text{num}}}. \tag{C.13}$$



It is zero for a symmetric branch and changes sign under left to right reflection. For one-dimensional branch centers ordered as $c_1 < \cdots < c_M$, the local separations are

$$g_{p,L} = c_p - c_{p-1}, \qquad g_{p,R} = c_{p+1} - c_p \tag{C.14}$$

when the corresponding neighbor exists. These values represent local geometric spacing. In a resolved first-order NMR multiplet they approach scalar-coupling separations, but VINEA does not impose that physical interpretation during tokenization. For $d_m = 2$, the sets $C_p^{(\ell)}$ are connected planar regions. Coordinate-wise extents are

$$w_{p,j}^{(\ell)} = \max_{x_i \in C_p^{(\ell)}} x_{i,j} - \min_{x_i \in C_p^{(\ell)}} x_{i,j}, \qquad j = 1, \ldots, d_m. \tag{C.15}$$

The resulting centroid, mass, and extent measurements play the same role as their one-dimensional counterparts.

C.4 GAUSSIAN SCALE-SPACE VINEYARD

A native-resolution persistence diagram describes topology at one effective resolution. VINEA additionally measures how each structure evolves as fine-scale detail is progressively removed. Let

$$f_\sigma(\mathbf{x}) = (G_\sigma * f)(\mathbf{x}), \qquad G_\sigma(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{d_m/2}} \exp\left(-\frac{\|\mathbf{x}\|_2^2}{2\sigma^2}\right). \tag{C.16}$$

The kernel has unit integral. For $d_m > 1$, $\mathbf{x}$ denotes the modality-specific coordinate system after the axis normalization used for scale-space smoothing. This normalization fixes the metric in which a scalar σ represents isotropic blur. Gaussian kernels satisfy

$$G_{\sigma_2} = G_{\sqrt{\sigma_2^2 - \sigma_1^2}} * G_{\sigma_1}, \qquad \sigma_2 > \sigma_1, \tag{C.17}$$

so increasing scale corresponds to additional Gaussian smoothing.

C.4.1 THE VINEYARD

For each fixed σ, the smoothed values are extended from vertices to cells by the same upper-star construction,

$$\mathcal{K}_\sigma^t = \{\kappa \in \mathcal{K}_m : f_\sigma(\kappa) \geq t\}, \tag{C.18}$$

and $t \mapsto H_0(\mathcal{K}_\sigma^t)$ is an ordinary one-parameter persistence module. Allowing the Gaussian scale to vary continuously produces

$$\sigma \longmapsto \mathrm{Dgm}_0(f_\sigma), \tag{C.19}$$

which is a persistence vineyard (Cohen-Steiner et al., 2006; Turner, 2023). The trajectories traced by persistence points $(b_p(\sigma), d_p(\sigma))$ in the birth-death plane are the vines.

The threshold t and blur scale σ play different mathematical roles. For fixed σ, decreasing t produces nested superlevel sets. For fixed t, increasing σ need not produce nested sets because smoothing can lower maxima while simultaneously raising nearby minima. The construction is therefore a one-parameter family of ordinary persistence modules rather than a canonical bifiltration in (t, σ).

C.4.2 CONTINUATION THROUGH SCALE

In continuous one-dimensional Gaussian scale space, increasing σ cannot create a new local extremum (Lindeberg, 1994). Existing extrema move continuously away from degenerate events and may disappear through maximum-minimum annihilation. A native-scale maximum therefore has a causal continuation toward coarser scales until its annihilation event, after which it cannot reappear.

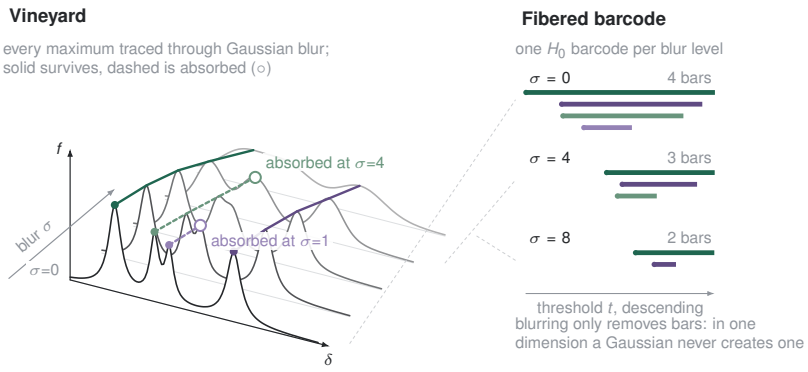
VINEA evaluates this continuous construction on a finite ladder $\Sigma_m = \{\sigma_1, \ldots, \sigma_L\}$. For one-dimensional spectra, active branches are continued between adjacent scales through the corresponding generating maxima. Once a branch can no longer be continued, it is marked inactive and is not reactivated at a coarser scale. For HSQC, the same sequence of cubical persistence diagrams is evaluated and persistent components are associated by a deterministic partial matching between adjacent scales. An unmatched component terminates its sampled vine.

C.4.3 RELATIVE PERSISTENCE ACROSS SCALE

Let $\pi_p(\sigma) = b_p(\sigma) - d_p(\sigma)$ denote the persistence of the continuation of branch p at scale σ. Raw persistence depends on the amplitude distribution of the current spectrum and changes systematically as the field is blurred. VINEA therefore converts it to a bounded relative-persistence score

$$r_p(\sigma) = \frac{\pi_p(\sigma)}{\pi_p(\sigma) + \tau_\pi(\sigma)}, \qquad \tau_\pi(\sigma) = \mathrm{median}_{q \in \mathcal{A}_\sigma}\, \pi_q(\sigma) \min\left(1, \frac{M_\sigma}{\rho N}\right), \tag{C.20}$$

where $\mathcal{A}_\sigma$ is the set of active finite branches, $M_\sigma = |\mathcal{A}_\sigma|$, N is the number of sampled grid locations, and $\rho = 0.05$. The density factor reduces the influence of the empirical median when only a small branch population is available. When $\tau_\pi(\sigma) > 0$, the map is monotone in $\pi_p(\sigma)$, lies in $[0, 1]$, equals $1/2$ when $\pi_p(\sigma) = \tau_\pi(\sigma)$, and tends to zero with persistence. If no finite branch is present at a scale, all profile entries at that scale are defined as zero.



Supplementary Figure 9: **Gaussian scale-space vineyard.** The continuous Gaussian family induces $\sigma \mapsto \mathrm{Dgm}_0(f_\sigma)$. VINEA evaluates this vineyard on a finite modality-specific ladder and records the persistence of each continued branch. In one dimension, increasing Gaussian scale can move and annihilate extrema but cannot create new ones.

For native-scale branch p, the sampled vineyard profile is

$$P_p = \big(r_p(\sigma_1), \ldots, r_p(\sigma_L)\big) \in [0,1]^L. \tag{C.21}$$

Once branch p terminates, all later entries are set to zero. We additionally record

$$\sigma_p^{\mathrm{death}} = \max\{\sigma_\ell : p \text{ is active at } \sigma_\ell\}. \tag{C.22}$$

This is the largest surviving sampled scale and is therefore a discretized scale-death statistic rather than the exact continuous annihilation value. A branch that remains active at σ_L is right-censored by the finite ladder and is stored with $\sigma_p^{\mathrm{death}} = \sigma_L$.

For 1D NMR,

$$\Sigma_{\mathrm{NMR}} = \{0.5, 1, 2, 4, 8, 16, 32\}\ \mathrm{Hz}. \tag{C.23}$$

Other modalities use fixed ladders in their corresponding scale-space coordinate systems. We refer to $(P_p, \sigma_p^{\mathrm{death}})$ as the branch's *vineyard profile*.

C.5 Branch Descriptor and Token Set

For each oriented native-scale branch, define the topological descriptor $\tau_p = (b_p, d_p, \pi_p, s_p)$ and the vineyard descriptor $v_p = (P_p, \sigma_p^{\mathrm{death}})$. The geometric descriptor g_p contains mass, prominence-relative extents, shape statistics, and local separations appropriate to modality m. The complete mathematical branch descriptor is

$$x_p = \big[\tau_p \| c_p \| g_p \| v_p\big]. \tag{C.24}$$

Branches are ranked by native-scale persistence and at most $M_{\max} = 128$ are retained,

$$\mathcal{T}(f) = \mathrm{TopM}_\pi\big(\{x_p\}_{p \in \mathcal{B}(f)}, M_{\max}\big). \tag{C.25}$$

The resulting representation is an unordered set rather than a rasterized spectral grid.

For the principal one-dimensional implementation, the active descriptor fields are summarized in Table 3. The implementation stores both normalized and physical coordinates because the former is convenient for absolute positional encoding while the latter preserves physically meaningful separations.

Five additional stored quantities, `log_precision`, `noise_rel`, `rel_w50`, `rel_area`, and `rel_gap`, are reserved for resolution-aware variants and are not consumed by the principal encoder. No dense grid is reconstructed after tokenization.

C.6 Permutation-Invariant Spectral Encoder

The ordering of persistent branches has no physical meaning. The neural encoder must therefore be permutation equivariant at the token level and permutation invariant after global aggregation (Lee et al., 2019). Let $m_p \in \{0, 1\}$ indicate whether token p is valid rather than padding.

Supplementary Table 3: **Per-branch fields used by the principal one-dimensional VINEA encoder.**

Group	Quantity	Stored field
topology	b_p, d_p, π_p	`height, death, persistence`
relative persistence	$r_p(0)$	`snr`
position	normalized and physical c_p	`position, position_khz`
mass	$\log A_p$	`log_area`
shape	$w_{25,p}, w_{50,p}, w_{75,p}$	`w25,w50,w75`
shape	shape, asymmetry	`shape,asym`
local spacing	$g_{p,L}, g_{p,R}$	`gap_left,gap_right`
orientation	s_p	`sign`
vineyard	$\sigma_p^{\mathrm{death}}, P_p$	`log_scale_death, blur_{...}`

C.6.1 Absolute coordinate encoding

Let $\psi_m : \Omega_m \to [0,1]^{d_m}$ be the affine normalization of the modality-specific acquisition window and define

$$\xi_p = \psi_m(c_p). \tag{C.26}$$

Absolute position is expanded coordinate-wise with $F = 8$ dyadic Fourier octaves (Tancik et al., 2020),

$$\gamma_F(\xi_p) = \Big\|_{j=1}^{d_m} \Big\|_{k=0}^{F-1} \left[\sin(2^k \pi \xi_{p,j}), \cos(2^k \pi \xi_{p,j})\right]. \tag{C.27}$$

Let D_m^{desc} denote the descriptor dimension for modality m. The complete neural input is $u_p = [x_p \| \gamma_F(\xi_p)]$. A shared branchwise embedding maps u_p to latent dimension D,

$$e_p = \mathrm{MLP}_{\mathrm{emb}}(u_p) \in \mathbb{R}^D, \tag{C.28}$$

where the embedding block is linear, LayerNorm, GELU, and linear.

C.6.2 Persistence-gated token embedding

A learned scalar function q_θ provides data-dependent modulation of each branch. The initial token is

$$z_p^{(0)} = m_p \pi_p \,\mathrm{sigmoid}\big(q_\theta(u_p)\big) e_p. \tag{C.29}$$

The explicit factor π_p imposes a topological inductive bias. A branch with vanishing persistence enters the attention stack with vanishing norm, while no manually selected persistence threshold is required.

C.6.3 Relative-coordinate attention

Absolute position is chemically informative, but local relationships between branches should also be available independently of a common coordinate translation. Let

$$\Lambda_m = \mathrm{diag}(\lambda_{m,1}, \ldots, \lambda_{m,d_m}) \tag{C.30}$$

contain fixed physical normalization scales for modality m. The normalized pairwise displacement is

$$\Delta_{pq} = \Lambda_m^{-1}(c_p - c_q). \tag{C.31}$$

With $F' = 6$, define the coordinate-wise relative Fourier features

$$\phi(\Delta_{pq}) = \Big\|_{j=1}^{d_m} \Big\|_{k=0}^{F'-1} \left[\sin(2^k \pi \Delta_{pq,j}), \cos(2^k \pi \Delta_{pq,j})\right]. \tag{C.32}$$

For attention head h, let $q_p^{(h)}$ and $k_q^{(h)}$ denote the query and key vectors obtained from the current token embeddings. The pre-softmax attention logit is

$$\ell_{pq}^{(h)} = \frac{(q_p^{(h)})^\top k_q^{(h)}}{\sqrt{D_h}} + B_h(\phi(\Delta_{pq})) + M_{pq}, \tag{C.33}$$

where B_h is a learned scalar-valued function and

$$M_{pq} = \begin{cases} 0, & m_q = 1, \\ -\infty, & m_q = 0. \end{cases} \tag{C.34}$$

The attention weights are $\alpha_{pq}^{(h)} = \mathrm{softmax}_q(\ell_{pq}^{(h)})$. Padded query outputs are masked again after each block.

The relative pathway is exactly invariant to a common coordinate translation because

$$\Lambda_m^{-1}\big[(c_p + \delta) - (c_q + \delta)\big] = \Lambda_m^{-1}(c_p - c_q) = \Delta_{pq}. \tag{C.35}$$

The complete encoder is not globally translation invariant because $\gamma_F(\xi_p)$ deliberately preserves absolute position. The two pathways therefore encode complementary information.

C.6.4 Set Transformer and invariant readout

The encoder uses two Set Transformer blocks with four attention heads. Each block applies multi-head self-attention, a residual connection, LayerNorm, and a position-wise feed-forward network. The token mask is retained throughout.

Let $z_1, \ldots, z_M \in \mathbb{R}^D$ be the final valid or padded token embeddings. VINEA computes three symmetric summaries. The masked mean is $\bar{z} = \sum_p m_p z_p / \sum_p m_p$. The masked coordinate-wise maximum is $(z_{\max})_j = \max_{p:m_p=1} z_{p,j}$. Pooling by multi-head attention with one learned seed produces z_{PMA}. The three summaries are concatenated and projected to the spectral embedding,

$$z_{\mathrm{spec}} = \mathrm{Head}\left[z_{\mathrm{PMA}} \| \bar{z} \| z_{\max}\right], \qquad \widehat{z}_{\mathrm{spec}} = \frac{z_{\mathrm{spec}}}{\|z_{\mathrm{spec}}\|_2} \in \mathbb{R}^{256}. \tag{C.36}$$

Branchwise transformations are shared across tokens and self-attention is permutation equivariant. PMA, masked mean, and masked maximum are invariant to token ordering. Consequently, for any permutation π,

$$E_\theta\big(\{x_{\pi(p)}\}_{p=1}^M\big) = E_\theta\big(\{x_p\}_{p=1}^M\big). \tag{C.37}$$

C.7 Computational Structure

For a one-dimensional spectrum with N grid points, zeroth persistence can be obtained from the ordered extrema and their merge structure without a general high-dimensional boundary-matrix reduction. The persistence construction is repeated over a fixed number L of Gaussian scales. After tokenization, at most $M_{\max} = 128$ branches enter the neural encoder. Self-attention therefore scales as

$$O(M_{\max}^2 D) \tag{C.38}$$

rather than with the length of the original sampled spectrum. Increasing digitization density can increase the preprocessing cost but does not increase the downstream token sequence once $M_{\max}$ is fixed. The principal spectral tower contains approximately 5.9×10^5 trainable parameters. Retrieval models additionally include the molecular encoder and projection layers.

C.8 Persistent Structure of Ideal First-Order NMR Multiplets

The preceding construction is modality independent. In resolved one-dimensional NMR spectra, several branch descriptors also admit a direct physical interpretation. The results below describe an ideal first-order regime and are not assumptions imposed during tokenization.

C.8.1 Ideal multiplet model

Let $\kappa_w : \mathbb{R} \to [0, 1]$ be an even, continuously differentiable, unimodal line-shape kernel of characteristic width w, with $\kappa_w(0) = 1$ and $\kappa_w(x) \to 0$ as $|x| \to \infty$. Coupling to n equivalent spin-$1/2$ partners with scalar coupling $J > 0$ gives the ideal multiplet

$$f(x) = \sum_{k=0}^{n} a_k \kappa_w(x - \nu_k), \qquad a_k = 2^{-n}\binom{n}{k}, \qquad \nu_k = \nu_c + \left(k - \frac{n}{2}\right) J. \tag{C.39}$$

The line centers satisfy $\nu_{k+1} - \nu_k = J$ and the isolated amplitudes follow the binomial profile.

C.8.2 Resolution of two equal lines

Lemma C.1 (Resolution threshold). *Consider two identical lines of FWHM w centered at $\pm d/2$. For a Lorentzian profile $L_w(x) = \left[1 + (2x/w)^2\right]^{-1}$, the midpoint is a strict local minimum exactly when*

$$d > \frac{w}{\sqrt{3}}.$$

For a Gaussian profile $G_w(x) = \exp(-4 \ln 2\, x^2/w^2)$, the midpoint is a strict local minimum exactly when

$$d > \frac{w}{\sqrt{2 \ln 2}} \approx 0.849\, w.$$

Proof. Symmetry gives a vanishing first derivative at the midpoint. The two lines are resolved exactly when the second derivative of their superposition is positive there. Direct differentiation gives $d^2 > w^2/3$ for the Lorentzian. For the Gaussian, writing $w = 2\sqrt{2 \ln 2}\, \sigma_0$ gives the equivalent condition $d > 2\sigma_0$. □

C.8.3 Elder pairing in a resolved multiplet

Let $x_0 < \cdots < x_n$ be resolved local maxima with interleaving minima $s_1, \ldots, s_n$. Suppose the maximum heights $h_k = f(x_k)$ are strictly unimodal with unique largest value at $k^\star$. Every non-essential maximum then dies at the saddle on the side facing the global maximum,

$$d_k = \begin{cases} f(s_{k+1}), & k < k^\star, \\ f(s_k), & k > k^\star. \end{cases} \tag{C.40}$$

Indeed, moving from x_k toward the global maximum first encounters a neighboring component generated by a larger maximum, so the Elder Rule terminates branch k at that merge. For a binomial amplitude sequence with two equal central coefficients, a deterministic tie rule selects the representative essential branch. The multiset of finite branch persistences is unaffected by this choice.

C.8.4 Weak-overlap limit

Define

$$\varepsilon(J, w) = \sup_{|x| \geq J/2} \kappa_w(x). \tag{C.41}$$

For any decaying line shape, $\varepsilon(J, w) \to 0$ as $J/w \to \infty$.

Theorem C.2 (Persistent structure of a resolved first-order multiplet). *Assume the multiplet in Equation (C.39) contains exactly $n + 1$ resolved local maxima. For fixed n, as $J/w \to \infty$, the maximum associated with line k satisfies $x_k - \nu_k \to 0$, so $x_{k+1} - x_k \to J$. For every finite branch,*

$$b_k = a_k + O(\varepsilon), \qquad d_k = O(\varepsilon), \qquad \pi_k = a_k + O(\varepsilon). \tag{C.42}$$

Hence the finite-branch persistence pattern converges to the corresponding binomial amplitude pattern.

Proof. In a neighborhood of ν_k, the kth component contributes $a_k \kappa_w(x - \nu_k)$ while every other component is uniformly $O(\varepsilon)$. A strict isolated maximum is stable under such a uniform perturbation, so a nearby maximum satisfies $x_k \to \nu_k$ and $b_k = a_k + O(\varepsilon)$. A saddle between neighboring resolved lines lies far from every center relative to w, so its total intensity is $O(\varepsilon)$. Therefore $d_k = O(\varepsilon)$ and subtraction gives the persistence result. □

At finite overlap, maxima of the superposition need not coincide exactly with the isolated line centers and branch gaps need not equal J. VINEA therefore stores measured positions and separations rather than imposing a multiplet model.

Supplementary Table 4: **Ideal four-line first-order patterns.** Local spacing distinguishes patterns with equal line count, while persistence reflects the corresponding isolated amplitude structure in the weak-overlap limit.

Pattern	Neighboring gaps	Relative amplitudes
quartet	(J, J, J)	$1:3:3:1$
doublet of doublets	$(J_2, J_1 - J_2, J_2)$	$1:1:1:1$

C.8.5 Dependence on magnetic-field strength

Scalar coupling constants expressed in hertz are approximately independent of the static field B_0 in the first-order regime. Chemical-shift separations expressed in hertz instead scale with spectrometer frequency. Consequently, when NMR coordinates are expressed in hertz, a change in field strength changes the absolute multiplet position and chemical-shift separations while leaving local J-splittings approximately unchanged. When coordinates are expressed in ppm, chemical-shift positions are approximately field invariant while the same J-splittings occupy a field-dependent distance in ppm. This difference motivates exposing absolute coordinates and local relative geometry through separate encoder pathways.

C.8.6 Scale death of an equal Gaussian doublet

Proposition C.3 (Gaussian scale death)**.** *Consider two equal Gaussian lines of standard deviation σ_0 separated by J. Gaussian smoothing with standard deviation σ_b produces an effective line width*

$$\sigma_{\mathrm{eff}} = \sqrt{\sigma_0^2 + \sigma_b^2}. \tag{C.43}$$

The two maxima remain resolved while $J > 2\sigma_{\mathrm{eff}}$. Their continuous annihilation scale is therefore

$$\sigma_b^\star = \sqrt{\frac{J^2}{4} - \sigma_0^2}, \qquad J > 2\sigma_0. \tag{C.44}$$

Proof. Convolution of Gaussian kernels adds variances. Applying the Gaussian resolution condition from Lemma C.1 to the smoothed pair gives $J = 2\sqrt{\sigma_0^2 + \sigma_b^2}$ at the transition. Solving for σ_b gives Equation (C.44). □

The stored $\sigma_p^{\mathrm{death}}$ is a discretized observation of this type of annihilation scale on the finite ladder Σ_m. Persistence and scale death therefore encode complementary information. Persistence measures separation along the intensity axis, while scale death measures the resolution scale at which neighboring structures cease to remain distinct.

C.8.7 Quartet and doublet-of-doublets example

A first-order quartet generated by one coupling J and a doublet of doublets generated by couplings $J_1 > J_2$ both contain four resolved components, but their geometry and persistence patterns differ.

More generally, coupling to groups of $n_1, \ldots, n_r$ equivalent partners with coupling constants $J_1, \ldots, J_r$ gives ideal line positions and amplitudes

$$\nu_{\mathbf{k}} = \nu_c + \sum_{i=1}^{r} \left(k_i - \frac{n_i}{2}\right) J_i, \qquad a_{\mathbf{k}} = \prod_{i=1}^{r} 2^{-n_i} \binom{n_i}{k_i}. \tag{C.45}$$

When the individual components are sufficiently resolved, the persistent branch set approaches this discrete splitting structure. VINEA therefore exposes multiplicity, amplitude, local spacing, and resolution-scale information without explicit multiplet fitting or coupling assignment.

D SYNTHETIC CORRUPTION BENCHMARK

The robustness benchmark applies controlled acquisition-motivated transformations directly to processed spectra. Each transformation changes the measured signal while preserving the molecular identity and downstream target. Using the notation of Appendix A, corruption family a for modality m acts as $\widetilde{f} = \mathcal{C}_{m,a,\eta}[f]$, where η contains the sampled corruption parameters.

The benchmark is a processed-spectrum stress test rather than a complete simulator of the measurement apparatus. We do not reconstruct the full NMR free-induction decay, HSQC pulse sequence, Raman optical path, FTIR interferogram, or detector electronics. Instead, each operator reproduces a characteristic observable distortion associated with acquisition, calibration, background variation, or finite resolution (Mo et al., 2017; Gautam et al., 2015). This design makes individual failure modes controllable and allows their effects on the representation to be isolated.

We use five ordered severity levels $\mathcal{S} = \{\mathrm{S1}, \ldots, \mathrm{S5}\}$ from mild to very severe. Unless stated otherwise, corruption amplitudes expressed as percentages are relative to the maximum absolute intensity of the clean input,

$$A_f = \|f\|_\infty. \tag{D.1}$$

Severity controls the magnitude of an enabled corruption rather than the probability that it is applied. Random parameters are generated deterministically from the sample identity, benchmark condition, and benchmark seed, so all models are evaluated on the same corrupted realization.

D.1 EVALUATION CONDITIONS

The benchmark evaluates four complementary conditions. CLEAN applies no transformation. COMBINED-Sk applies every corruption defined for the modality at severity Sk. ONLY-a applies corruption a alone at S3 and measures direct sensitivity to that artifact. LOCO-a applies every S3 corruption except a and measures how strongly that artifact contributes to the combined degradation.

For modality m with corruption set $\mathcal{A}_m$, these conditions can be written

$$f_{\text{Only-}a} = \mathcal{C}_{m,a,\eta_a}[f], \qquad f_{\text{LOCO-}a} = \left(\bigcirc_{b \in \mathcal{A}_m \setminus \{a\}} \mathcal{C}_{m,b,\eta_b} \right)[f]. \tag{D.2}$$

The composition order is fixed throughout the benchmark because the transformations do not generally commute. For example, broadening an impulsive artifact is not equivalent to inserting the same artifact after broadening. Enabled corruptions are applied with probability one at evaluation time. Corruption-aware training uses the separate protocol specified in Appendix G.

D.2 ONE-DIMENSIONAL NMR

Let $S(\delta)$ denote a processed one-dimensional ^{1}H NMR spectrum with chemical-shift coordinate δ. The controlled acquisition benchmark includes additive noise, baseline distortion, phase error, line broadening, and global referencing displacement. These operators represent distinct changes to the measured trace rather than independent claims about the microscopic origin of every experimental imperfection (Levitt, 2008; Xi & Rocke, 2008).

D.2.1 ADDITIVE NOISE

Broadband receiver and electronic noise is approximated by independent Gaussian perturbations,

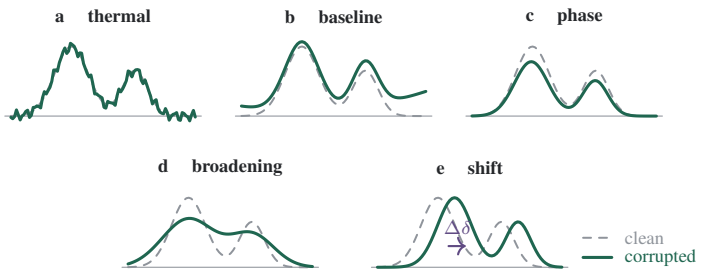
$$S_{\text{aug}}(\delta_i) = S(\delta_i) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2). \tag{D.3}$$

The scale is parameterized through the peak signal-to-noise ratio $\mathrm{SNR} = A_S/\sigma$, where $A_S = \|S\|_\infty$. This is a processed-spectrum approximation to broadband detection noise rather than a time-domain receiver simulation (Levitt, 2008).

D.2.2 BASELINE DISTORTION

Slow baseline variation is modeled by a low-order polynomial on a normalized coordinate $u \in [-1, 1]$,

$$B(u) = \sum_{k=0}^{K} c_k u^k, \qquad K \le 3, \qquad S_{\text{aug}} = S + B. \tag{D.4}$$



Supplementary Figure 10: **Representative one-dimensional NMR corruptions.** The clean trace is shown in gray and the transformed trace in green. The schematic illustrates the action of each processed-spectrum operator, while the benchmark uses the numerical definitions given in the text.

The coefficients are randomly sampled and the resulting curve is normalized before scaling relative to A_S. Low polynomial order produces offsets, slopes, and broad curvature without deliberately inserting narrow resonance-like features. Such baseline distortions are representative of imperfect acquisition and processing corrections (Xi & Rocke, 2008).

D.2.3 PHASE DISTORTION

Incorrect phase correction mixes absorptive and dispersive components. From the real spectrum we form the analytic signal

$$S_A(\delta) = S(\delta) + i\,\mathcal{H}[S](\delta), \tag{D.5}$$

where $\mathcal{H}$ denotes the Hilbert transform. A phase function $\phi(\delta) = \phi_0 + \phi_1\delta$ is then applied,

$$S_{\text{aug}}(\delta) = \text{Re}\left[S_A(\delta)e^{i\phi(\delta)}\right] = S(\delta)\cos\phi(\delta) - \mathcal{H}[S](\delta)\sin\phi(\delta). \tag{D.6}$$

The largest sampled phase parameters are bounded by $\pi/4$. This construction generates dispersive positive and negative lobes while preserving the original spectral coordinate.

D.2.4 LINE BROADENING

Loss of effective resolution caused by magnetic-field inhomogeneity is represented by Lorentzian convolution,

$$S_{\text{aug}} = S * K_\gamma, \qquad K_\gamma(\delta) = \frac{1}{\pi}\frac{\gamma}{\delta^2 + \gamma^2}. \tag{D.7}$$

The parameter γ is the Lorentzian half-width at half maximum, so the corresponding full width at half maximum is 2γ. Increasing γ broadens resonances and progressively removes fine splitting structure.

D.2.5 GLOBAL CHEMICAL-SHIFT DISPLACEMENT

Referencing differences translate the complete spectrum without changing relative spectral geometry. We isolate this effect through

$$S_{\text{aug}}(\delta) = S(\delta - \Delta\delta). \tag{D.8}$$

The documented ^{1}H implementation bounds the displacement by $\Delta\delta \sim \mathcal{U}(-0.05, 0.05)$ ppm. Translation is performed by linear interpolation and samples moved outside the acquisition window are replaced by zero. This corruption represents a global coordinate offset rather than molecule-dependent shifts caused by pH, temperature, or concentration changes.

Solvent interference and magnetic-field transfer are evaluated separately from the S1–S5 acquisition-corruption ladder because they represent structured physical or domain variation rather than generic stochastic corruption.

D.3 TWO-DIMENSIONAL HSQC

Let $I(i, j)$ denote a processed HSQC intensity plane, where i indexes the indirect dimension and j indexes the directly detected dimension. The benchmark applies Gaussian noise, smooth background

distortion, global phase error, structured t_1 streaking, global coordinate displacement, and isotropic resolution broadening. The operators act directly on the processed plane.

D.3.1 GAUSSIAN NOISE

Independent Gaussian noise is added to every pixel,

$$I_{\mathrm{aug}}(i,j) = I(i,j) + \epsilon_{ij}, \qquad \epsilon_{ij} \sim \mathcal{N}\left(0, \frac{I_{\max}^2}{\rho^2}\right), \tag{D.9}$$

where $I_{\max} = \|I\|_\infty$ and ρ is the target peak SNR.

D.3.2 SMOOTH BACKGROUND DISTORTION

A smooth random background is generated from normalized coordinates $u, v \in [-1, 1]$ using

$$B(u,v) = c_0 + c_1 u + c_2 v + c_3 uv + c_4(u^2 + v^2). \tag{D.10}$$

The surface is normalized by its maximum absolute magnitude and then scaled relative to $I_{\max}$ before addition. This construction produces constant offsets, planar trends, coupled slopes, and broad curvature without inserting localized cross-peak-like objects.

D.3.3 GLOBAL PHASE ERROR

A phase angle is sampled independently for the two spectral dimensions,

$$\phi_d \sim \mathcal{N}(0, \sigma_\phi^2), \qquad d \in \{1, 2\}. \tag{D.11}$$

Along each dimension, the corresponding directional analytic signal is constructed with a Hilbert transform, rotated by ϕ_d, and projected back to its real component. This mixes absorptive and dispersive contributions. The current implementation applies one global phase rotation per dimension and does not include a frequency-dependent first-order phase ramp.

D.3.4 t_1 STREAKING

Multidimensional NMR is acquired by repeating the experiment over successive indirect-time increments. Scan-to-scan amplitude or phase modulation across these increments spreads signal along the indirect frequency dimension after Fourier transformation and can produce characteristic ridges near strong peaks (Mo et al., 2017). We reproduce this processed-domain geometry by selecting up to four strong direct-frequency positions j_k and adding narrow ridges,

$$I_{\mathrm{aug}}(i,j) = I(i,j) + A_s I_{\max} \sum_{k=1}^{K} r_k(i)\, \kappa_{w_k}(j - j_k), \qquad K \le 4. \tag{D.12}$$

Here $r_k(i)$ is a smooth random modulation along the indirect dimension and κ_{w_k} is a narrow ridge profile. The ridge width w_k is sampled between approximately 0.7 and 2 pixels.

D.3.5 GLOBAL COORDINATE DISPLACEMENT

One displacement is sampled independently for each axis,

$$\Delta_x, \Delta_y \sim \mathcal{N}(0, \sigma_{\mathrm{shift}}^2). \tag{D.13}$$

The complete plane is translated with bilinear interpolation. Values moved outside the acquisition window are replaced by zero. The operation represents a global referencing offset rather than independent movement of individual cross-peaks.

D.3.6 RESOLUTION BROADENING

The processed plane is convolved with an isotropic Gaussian kernel. The current implementation uses the same standard deviation along both axes. This is a controlled loss-of-resolution operator and is not intended as a complete model of the generally anisotropic linewidth mechanisms of real HSQC spectra.

Supplementary Table 5: **HSQC corruption severity schedule.** Percentage amplitudes are relative to $I_{\max}$.

	S1	S2	S3	S4	S5
SNR	100	50	25	12	6
baseline [%]	0.3	0.8	2	5	10
phase SD [°]	1	3	7	15	30
t_1 ridge [%]	0.3	0.8	2	5	10
shift SD [px]	.15	.35	.70	1.5	3.0
blur SD [px]	.15	.30	.60	1.2	2.0

D.4 Raman Spectroscopy

Let $I(\widetilde{\nu})$ denote a Raman spectrum indexed by Raman shift in cm^{-1}. The benchmark combines smooth fluorescence backgrounds, sparse cosmic-ray spikes, signal-dependent detector noise, global calibration shift, and Gaussian resolution broadening. These operators reflect common processed-spectrum artifacts encountered in Raman measurements (Gautam et al., 2015; Wei et al., 2015).

D.4.1 Fluorescence background

Fluorescence produces a broad contribution that can dominate narrow Raman bands while varying smoothly over wavenumber (Wei et al., 2015). We generate a non-negative background from a constant component, a random linear term, quadratic curvature, and one to three broad Gaussian humps. The resulting curve is normalized to unit maximum and scaled relative to $A_I = \|I\|_\infty$ before addition.

D.4.2 Cosmic-ray spikes

High-energy particle events on CCD detectors can produce narrow high-amplitude impulses that are substantially sharper than molecular Raman bands (Gautam et al., 2015). The benchmark inserts a severity-dependent number of positive impulses at random coordinates. Each impulse is weakly broadened so that it occupies one or a few sampled locations. Its realized height is sampled uniformly between 70% and 130% of the nominal severity-dependent amplitude.

D.4.3 Photon and detector noise

Photon-counting fluctuations grow approximately with the square root of signal intensity. We approximate the combined shot and read-noise scale by

$$\sigma_i = aA_I\sqrt{\frac{\max(I_i - I_{\min}, 0)}{\max_j(I_j - I_{\min})} + 0.05}, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma_i^2). \tag{D.14}$$

The corrupted value is $I_i^{\mathrm{aug}} = I_i + \epsilon_i$. The square-root term approximates signal-dependent shot noise and the additive constant provides non-zero noise near the baseline. This is a Gaussian approximation rather than explicit Poisson photon counting.

D.4.4 Global wavenumber displacement

Instrument calibration or alignment changes are modeled through

$$\Delta\widetilde{\nu} \sim \mathcal{N}(0, \sigma_{\mathrm{shift}}^2), \qquad I_{\mathrm{aug}}(\widetilde{\nu}) = I(\widetilde{\nu} - \Delta\widetilde{\nu}). \tag{D.15}$$

The translated trace is evaluated by interpolation. All bands move coherently, so the operator tests calibration sensitivity rather than molecule-dependent changes in individual vibrational frequencies (Gautam et al., 2015).

D.4.5 Resolution broadening

The spectrum is convolved with a Gaussian kernel whose standard deviation is specified in cm^{-1} and converted to grid units using the spectrum's coordinate spacing. Increasing this width lowers

Supplementary Table 6: **Raman corruption severity schedule.** Percentage amplitudes are relative to A_I.

	S1	S2	S3	S4	S5
fluorescence [%]	3	8	18	35	65
cosmic spikes	1	2	3	5	8
spike height [%]	25	50	90	150	250
noise scale [%]	.4	.8	1.6	3.2	6
shift SD [cm^{-1}]	.25	.5	1	2	4
blur SD [cm^{-1}]	.5	1	2	4	8

narrow peak heights and merges nearby bands, representing variation in effective instrumental resolution (Gautam et al., 2015).

D.5 Infrared Spectroscopy

Let $I(\widetilde{\nu})$ denote an IR or FTIR spectrum indexed by optical wavenumber. The benchmark combines smooth baseline distortion, detector noise, global calibration shift, resolution broadening, and structured atmospheric residuals. Water vapor and carbon dioxide in the optical path can remain after background correction when atmospheric conditions differ between the background and sample measurements. Finite optical-path difference and apodization also modify effective spectral resolution and line shape (Naylor & Tahic, 2007).

D.5.1 Atmospheric CO_2 residual

A normalized three-band template is centered near 2338, 2361, and 667 cm^{-1} with nominal widths of 4, 4, and 7 cm^{-1}. Each center receives random jitter with approximately 0.8 cm^{-1} standard deviation. The complete template is multiplied by a randomly selected positive or negative sign before addition. Allowing both signs reflects imperfect subtraction when the gas contribution is larger in either the sample scan or the reference scan.

D.5.2 Atmospheric H_2O residual

Water contamination is represented by a sparse collection of narrow bands. Eighteen components are placed between 1360 and 1880 cm^{-1} and twelve between 3570 and 3930 cm^{-1}. Individual centers receive approximately 1 cm^{-1} jitter, widths are sampled from 1.5 to 4 cm^{-1}, and relative amplitudes are randomized. The complete residual is assigned either sign before addition.

D.5.3 Baseline distortion

The smooth baseline is generated on normalized coordinate $x \in [-1, 1]$ using

$$b(x) = c_0 + c_1 x + c_2(2x^2 - 1) + c_3(4x^3 - 3x). \tag{D.16}$$

The curve is normalized by its maximum absolute magnitude and scaled relative to A_I before addition. Random coefficients allow offsets, slopes, broad curvature, and S-shaped drift. Smooth baseline variation is a common preprocessing concern in vibrational spectroscopy (Gautam et al., 2015).

D.5.4 Detector noise

Independent Gaussian noise is added with a slowly varying standard deviation,

$$\sigma(\widetilde{\nu}) = aA_I\,(0.75 + 0.5\widetilde{\nu}_{\mathrm{norm}})\,, \qquad \widetilde{\nu}_{\mathrm{norm}} \in [0, 1]. \tag{D.17}$$

The noise level therefore varies from $0.75aA_I$ at one edge of the spectral window to $1.25aA_I$ at the other. This operator approximates wavelength-dependent throughput and detector noise without modeling detector electronics explicitly.

Supplementary Table 7: **Infrared corruption severity schedule.** Percentage amplitudes are relative to A_I.

	S1	S2	S3	S4	S5
CO_2 residual [%]	1	2.5	5	10	18
H_2O residual [%]	1	2.5	5	10	18
baseline [%]	1	2.5	6	12	22
noise scale [%]	.25	.5	1	2.5	5
shift SD [cm^{-1}]	.25	.5	1	2	4
blur SD [cm^{-1}]	.5	1	2	4	8

D.5.5 Global wavenumber displacement

A normally distributed displacement is applied to the complete spectrum through interpolation,

$$\Delta\widetilde{\nu} \sim \mathcal{N}(0, \sigma^2_{\text{shift}}), \qquad I_{\text{aug}}(\widetilde{\nu}) = I(\widetilde{\nu} - \Delta\widetilde{\nu}). \tag{D.18}$$

The transformation tests sensitivity to coordinate calibration while preserving the relative positions of molecular bands.

D.5.6 Resolution broadening

The spectrum is convolved with a Gaussian kernel whose standard deviation is specified in cm^{-1}. The operator represents a generic reduction in spectral resolution rather than the exact FTIR instrumental line shape, which depends on finite optical-path difference and the selected apodization function (Naylor & Tahic, 2007).

D.6 Scope and Determinism

The corruption suite is designed to compare representation robustness under controlled and reproducible perturbations. It should not be interpreted as a complete probability model of experimental spectroscopy. Several real effects are deliberately simplified. HSQC broadening is isotropic in the benchmark even though experimental linewidths can differ across dimensions. Raman shot noise is represented by a Gaussian approximation rather than direct Poisson counting. IR resolution loss is implemented with Gaussian convolution rather than the exact FTIR instrumental line shape. NMR acquisition artifacts are applied to processed spectra rather than generated through a complete time-domain simulation.

All stochastic parameters are derived deterministically from the sample identity, condition, severity, and benchmark seed. The same input under the same condition therefore produces the same corrupted spectrum across model families. This ensures that differences in robustness cannot be attributed to different random corruption realizations.

E Datasets and Models

E.1 Tasks and comparison framework

We evaluate four spectral modalities under a common representation benchmark. ^{1}H NMR and ^{1}H–^{13}C HSQC use the Molecular Spectra Dataset (MSD) (Alberts et al., 2024). The ^{1}H NMR retrieval benchmark uses an 18,493-molecule candidate pool, while HSQC is evaluated on a 5,000-molecule scaffold-disjoint split. Raman uses Raman–ChEMBL (Liang et al., 2025) for spectrum-to-molecule retrieval. IR uses the computational spectra of Zipoli et al. (2025) for 17-label functional-group prediction. Dataset splits, candidate pools, experimental transfer sets, and evaluation metrics are specified in Appendix H.

All published methods are compared at the level of their spectral encoder. We retain each source architecture through its final hidden representation, remove the original task-specific decoder or prediction head, and attach only the projection or prediction layer required by the common benchmark. Unless stated otherwise, models are trained locally on the same training partition and selected on

Supplementary Table 8: **Model families included in the reported experiments.** Internal controls are listed separately and are not counted as published baselines.

Modality and task	Proposed model	Published comparison models
^{1}H NMR retrieval	VINEA	NMR2Struct, SECS, NMRViT, CHEFNMR
^{1}H–^{13}C HSQC retrieval	VINEA-HSQC	NMRViT-HSQC, Kuhn HSQC-CNN, DeepSAT
Raman retrieval	VINEA	DeepCNN, RamanNet, SANet, RamanFormer
IR functional-group prediction	VINEA	IRCNN, SpectraML-CNN, ACT, UltraIR

the same clean validation partition. Parameter counts therefore describe the instantiated benchmark model rather than the complete published end-to-end system.

Retrieval tasks. For ^{1}H NMR, HSQC, and Raman, each spectral encoder is trained against the same molecular encoder. Molecular graphs are constructed from SMILES and processed by a three-layer Graph Isomorphism Network (Xu et al., 2019). Atom features encode atomic number, degree, formal charge, hydrogen count, aromaticity, and ring membership. Global sum pooling is followed by projection into the shared 256-dimensional embedding space. Spectrum and molecule embeddings are ℓ_2 normalized and trained with the symmetric contrastive objective defined in Appendix G. Retrieval uses cosine similarity against the complete held-out candidate pool. Molecular formula is not supplied to the neural spectral encoders.

Corruption is applied to the processed spectrum before construction of the model-specific input view. Dense grids, patches, peak sets, and VINEA tokens are therefore all derived from the same corrupted observation. The molecular encoder, contrastive objective, data split, candidate pool, and evaluation code are otherwise shared across spectral models.

IR prediction. IR encoders produce a 256-dimensional representation followed by a linear 17-label prediction head. Training uses class-weighted binary cross-entropy. Checkpoints are selected on the clean validation split and remain fixed for all corrupted test conditions. Macro-AUROC is the primary evaluation metric.

E.2 VINEA Benchmark Instantiations

The complete VINEA construction is given in Appendix C. Here we specify only the benchmark instantiations.

One-dimensional VINEA. The same one-dimensional branch representation and set encoder are used for ^{1}H NMR, Raman, and IR. Each spectrum is converted to at most 128 persistent branch tokens containing topology, physical position, local geometry, and sampled scale-space persistence. Absolute coordinates are Fourier encoded, pairwise physical displacements enter attention through a learned relative-position bias, and token amplitudes are explicitly weighted by persistence. Two four-head attention blocks are followed by PMA, masked mean, and masked maximum readouts. The spectral encoder contains approximately 0.6 million trainable parameters. Retrieval models additionally include the common molecular tower and projection layers.

VINEA-HSQC. The HSQC model applies the two-dimensional construction in Appendix C to normalized spectral planes. Persistent connected regions provide two-dimensional positions, persistence, sign, geometric extents, and sampled scale-space information. Large input planes are reduced to a 128×128 grid before persistent decomposition while retaining the strongest signed response within each cell. The resulting branch set is processed by the same permutation-invariant encoder template. The complete HSQC retrieval model contains approximately 1.03 million trainable parameters.

E.3 PUBLISHED ^{1}H NMR BASELINES

NMR2Struct. NMR2Struct (Hu et al., 2024) uses the released ^{1}H `H1Embed` convolutional feature extractor. Each spectrum is resampled to 28,000 points and processed by two one-dimensional convolution and max-pooling stages before flattening to a 256-dimensional hidden representation. We retain this spectral encoder and project its output into the shared retrieval space. The benchmark model contains approximately 4.46 million parameters and does not include the original substructure Transformer or structure decoder.

SECS. SECS (Mirza et al., 2026) uses the published large hydrogen-NMR encoder. A strided convolutional stem feeds a deep residual hierarchy followed by global self-attention, global average pooling, and a projection head. Its native 1,024-dimensional latent representation is projected into the common retrieval space. The benchmark model contains approximately 87.5 million parameters and is trained locally rather than initialized from a published cross-modal checkpoint.

NMRViT. NMRViT (Han et al., 2026) resamples each ^{1}H spectrum to 10,000 points and divides it into 200 non-overlapping 50-point patches. Patch embeddings and a learned classification token receive learned positional embeddings and pass through a six-layer, eight-head Transformer of width 512. Patch tokens are dropped with probability 0.25 during training. We retain the normalized classification-token representation and exclude the original molecular decoder. The benchmark model contains approximately 19.7 million parameters.

CHEFNMR. CHEFNMR (Xiong et al., 2025) uses a convolutional tokenization stage followed by a Transformer. The 10,000-point proton trace passes through two convolution, ReLU, and max-pooling blocks. The remaining feature columns are projected to 256-dimensional tokens and processed by four eight-head Transformer blocks with learned positional embeddings. Attention pooling produces the spectral conditioning vector used by the common retrieval projection. We retain only the proton spectral encoder and exclude the carbon branch and diffusion generators. The benchmark model contains approximately 4.41 million parameters.

E.4 PUBLISHED HSQC BASELINES

NMRViT-HSQC. The HSQC form of NMRViT (Han et al., 2026) receives a dense plane resized to 512×512. A stride-32 convolution creates 256 non-overlapping 32×32 patches that are processed with a learned classification token by the same six-layer, eight-head Transformer. Only the spectral encoder is retained. The benchmark model contains approximately 20.19 million parameters.

Kuhn HSQC-CNN. The convolutional model of Kuhn et al. (Kuhn et al., 2022) receives an HSQC plane resized to its native 300×205 grid. Four valid 3×3 convolutions use 32, 64, 64, and 128 channels, with max pooling after the first three layers. The flattened feature map passes through a 64-unit hidden layer before projection into the retrieval space. The original substructure classifier is removed. The benchmark model contains approximately 6.36 million parameters.

DeepSAT. DeepSAT (Kim et al., 2023) consumes a 128×128 one-channel HSQC plane. A two-convolution stem is followed by five densely connected convolutional stages whose widths increase to 2,048 channels. Global maximum pooling produces the spectral representation. The original auxiliary prediction heads are replaced by the common latent projection. The benchmark model contains approximately 149.83 million parameters.

E.5 PUBLISHED RAMAN BASELINES

The Raman architectures are implemented using the same pinned RamanBench framework and are evaluated with the common Raman retrieval objective (Koddenbrock et al., 2026). Baseline spectra are resampled to the native 1,024-point input expected by these models, while VINEA operates directly on the physical Raman-shift axis.

DeepCNN. DeepCNN (Liu et al., 2017) is a one-dimensional convolutional network with kernel widths 21, 11, and 5. Each convolution is followed by batch normalization, a leaky ReLU, and max

pooling. Flattened features pass through a 256-unit tanh layer and dropout before projection to the retrieval embedding. The benchmark model contains approximately 4.27 million parameters.

RamanNet. RamanNet (Ibtehaz et al., 2023) partitions the spectrum into overlapping 50-point windows with stride 25. Each window position has an independent small MLP, which explicitly preserves absolute spectral position. Window representations are concatenated and processed by two dense blocks before projection. The benchmark model contains approximately 0.69 million parameters.

SANet. SANet (Deng et al., 2022) contains five scale-adaptive convolutional blocks. Each block combines six parallel convolutional branches with kernel widths 3, 5, 7, 9, 11, and 13. Cross-scale attention reweights the resulting features before 1×1 fusion. A final projection and global average pooling produce the spectral representation. The benchmark model contains approximately 2.88 million parameters.

RamanFormer. RamanFormer (Koyun et al., 2024) divides the 1,024-point spectrum into eight 128-point patches. Three eight-head Transformer layers process the linearly embedded patches, after which two one-dimensional convolutional layers combine neighboring token features. Global pooling and a dense hidden layer produce the representation used by the retrieval projection. The benchmark model contains approximately 4.31 million parameters.

E.6 Published IR Baselines

IRCNN. IRCNN (Jung et al., 2023) resamples each spectrum to 600 points. Two width-11 convolutions with 31 and 62 channels are followed by batch normalization, ReLU, and max pooling. The flattened representation passes through dense layers of widths 4,927, 2,785, and 1,574 before the common 17-label prediction head. The benchmark model contains approximately 64.37 million parameters.

SpectraML-CNN. The SpectraML-CNN (Rieger et al., 2023) implementation resamples each spectrum to 4,000 points and applies four one-dimensional convolutions arranged in two pooled blocks. The first pair uses four channels and the second pair uses eight. Batch normalization follows the convolutional stages. Flattened features are projected to the shared latent representation before multilabel prediction. The benchmark model contains approximately 2.05 million parameters.

ACT. ACT (Huang et al., 2025) estimates a reversible linear baseline and forms overlapping local-shape tokens from the corrected 650-point spectrum. Its analytical-chemistry-informed attention combines inter-spectrum attention to a fixed clean-training reference set with intra-spectrum attention guided by a training-derived band-correlation map. Learned compression and a two-layer dense representation head precede prediction. Reference spectra, normalization statistics, and correlation maps are fitted only on the clean training split. The public implementation distributes the central SpectrA component as a Windows binary, so our approximately 0.66 million parameter implementation reconstructs the published architecture from the accessible pipeline and paper specification rather than reproducing the binary itself.

UltraIR. UltraIR (Tan et al., 2026) resamples spectra to 1,792 points and constructs raw, first-derivative, and second-derivative channels using a learnable derivative front end. A gated fusion module feeds a residual convolutional hierarchy. Features from several depths are aligned and fused before conversion to overlapping 16-point patches. An eight-layer, 16-head pre-normalized Transformer of width 1,024 produces the final classification-token representation. The benchmark model contains approximately 129.64 million parameters.

E.7 Internal and Non-Neural Controls

Random retrieval. The random control assigns each query a uniformly sampled rank from the candidate pool using a fixed seed. Its expected Recall@1 is therefore $1/N$ for a pool of N candidates.

Supplementary Table 9: **VINEA architecture ablations.** Components not listed in a row remain identical to the full model.

Variant	Change relative to full VINEA
− blur axis	Remove the vineyard profile and sampled scale-death features.
− line shape	Remove area, prominence-relative widths, shape ratio, and asymmetry.
− coupling gaps	Remove left and right neighboring-branch separations.
Unsigned filtration	Retain only positive-orientation branches.
Absolute shifts only	Remove the pairwise relative-coordinate attention bias.
Shuffled position	Permute position-to-feature correspondence within each spectrum while preserving the set of positions.
− learned gate	Remove the learned sigmoid gate while retaining persistence weighting.
One attention block	Reduce the attention stack from two blocks to one.
DeepSets readout	Remove self-attention and PMA and retain masked mean and maximum pooling.
− PMA	Retain self-attention and use masked mean and maximum for readout.
PMA only	Retain only pooling by multi-head attention for readout.
64-d embedding	Reduce both retrieval embeddings from 256 to 64 dimensions.
128-d embedding	Reduce both retrieval embeddings from 256 to 128 dimensions.
512-d embedding	Increase both retrieval embeddings from 256 to 512 dimensions.

Formula-only retrieval. The formula-only control does not inspect the spectrum. Candidates are ranked by the L_1 distance between atom-count vectors with deterministic seeded tie breaking. This control measures the retrieval signal available from molecular formula alone and is reported separately from the formula-blind neural comparisons.

Raw ResNet. The raw-grid control is a compact one-dimensional residual CNN with strided downsampling, residual blocks, global average pooling, and a linear projection. It tests the dense-grid representation paradigm and is not attributed to an external publication.

Spectral Transformer. The patch-based control divides the dense trace into non-overlapping 32-point segments, maps them to tokens, prepends a learned classification token, and applies a four-layer, eight-head Transformer. It is an internal representation control rather than a published baseline.

E.8 VINEA Architecture Variants

The architecture ablation changes one VINEA mechanism at a time while keeping the dataset split, molecular encoder, optimization protocol, checkpoint selection, corruption realizations, and evaluation pool fixed. The reference model uses a 256-dimensional embedding. Table 9 defines each intervention and Table 16 reports its results.

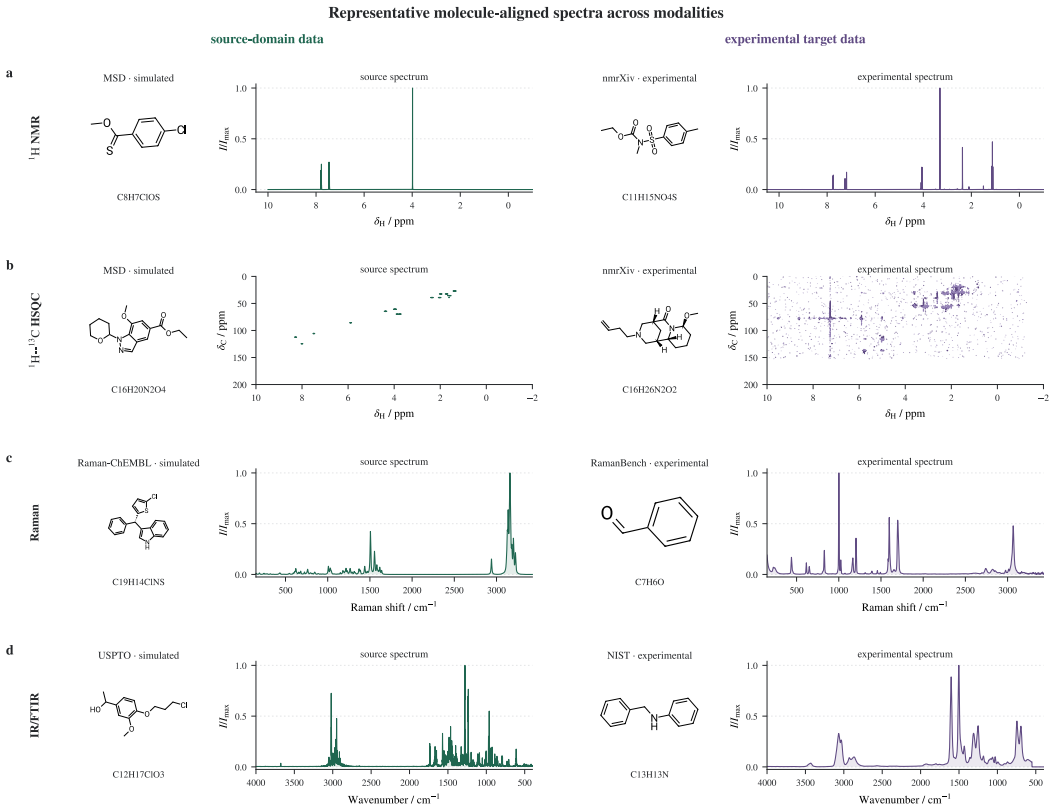
E.9 Transfer and Representation Probes

For measured-spectrum transfer, the same encoder architecture is evaluated under controlled initialization regimes. In the zero-shot setting, a model trained on the synthetic source domain is evaluated directly on the experimental target set. In the fine-tuning setting, the source-domain checkpoint is further optimized on the target training partition. In the scratch setting, the same architecture is randomly initialized and trained only on the target domain. These comparisons separate the effect of source-domain pretraining from architectural differences. Dataset-specific transfer protocols and candidate pools are given in Appendix H.

For representation probing, the pretrained spectral encoder is frozen and the same approximately 66.8k-parameter MLP head is trained on top of every representation. Because encoder parameters remain fixed, these experiments measure the molecular information accessible from the learned spectral representation rather than the capacity of a fully retrained predictor.

E.10 Consolidated results across datasets and modalities

Unless stated otherwise, entries are the mean and sample standard deviation across independently trained seeds. Higher is better for every endpoint. We report one primary endpoint per task to



Supplementary Figure 11: **Multimodal spectral datasets across simulated source and experimental target domains.**

keep the comparison compact: recall@1 (R@1) for retrieval, balanced accuracy (BAcc) for Raman pharmaceutical classification, and macro ROC-AUC (mAUROC) for multilabel IR prediction. The historical result key `topo_scalespace` denotes VINEA.

Artifact-specific robustness. To determine whether robustness under combined corruption is driven by a single easy artifact, we additionally apply each corruption family alone at severity S3. Because the four modalities use different task metrics, we report performance relative to the clean score of the same model and seed. For seed i, relative performance is $r_i = 100\, s_{i,\text{artifact}}/s_{i,\text{clean}}$, and Table 10 reports the mean and sample standard deviation of r_i over ten matched seeds. Ratios are not clipped: values above 100% indicate that the corrupted-test score is slightly higher than the paired clean-test score. We interpret such values as no measurable degradation, not as preservation of more than 100% of the representation. Ramp-trained VINEA retains near-clean performance across most artifact families; thermal noise is the principal remaining failure mode for NMR and HSQC.

Across the clean benchmarks, VINEA has the largest mean on MSD ^{1}H NMR and HSQC, while SANet and UltraIR have the largest means on the clean Raman and IR tasks, respectively. Under S1-S3 ramp training, VINEA has the largest cAUDC in all four modalities. The Raman cAUDC differences among VINEA, ResNet-1D, ReZeroNet, and the Spectral Transformer are small and should not be read as a strict ranking without a paired significance analysis.

Synthetic corruption training generally improves zero-shot transfer to measured data, but the best architecture is dataset-dependent. NMR2Struct or NMRViT leads the measured ^{1}H NMR repositories, SPECTRE leads measured HSQC, VINEA leads measured Raman, and ACT/UltraIR lead measured IR. Thus, robustness to controlled corruption and absolute zero-shot domain transfer are related but not interchangeable criteria.

Supplementary Table 10: **VINEA relative performance under individual corruption families.** Each artifact is applied alone at severity S3. Relative performance is computed independently for each seed as $100\, s_{\mathrm{artifact}}/s_{\mathrm{clean}}$ and then averaged; entries are mean $\pm$ sample standard deviation over ten matched training seeds. Ratios are not clipped. Values above 100% therefore denote a small improvement over the paired clean-test score and are interpreted as no measurable degradation, not as retention of more than 100% of the representation. Clean-only and S1-S3 ramp-trained VINEA checkpoints are evaluated on identical held-out examples and corruption draws. Bold indicates the higher mean within each artifact. The table compares VINEA training regimes, not VINEA against the modality-specific baselines.

Modality	Artifact applied alone at S3	Clean-only relative performance (%)	S1-S3 ramp relative performance (%)
^{1}H NMR			
	Thermal	1.6 ± 0.4	70.5 ± 0.9
	Baseline	69.8 ± 1.3	96.1 ± 0.6
	Phase	68.6 ± 2.3	100.2 ± 0.7
	Shift	87.8 ± 0.9	99.5 ± 0.3
	Shim	6.1 ± 1.3	98.9 ± 1.1
^{1}H-^{13}C HSQC			
	Thermal	1.2 ± 0.8	56.8 ± 21.4
	Baseline	46.9 ± 5.6	99.0 ± 3.1
	Phase	62.4 ± 4.2	95.6 ± 4.5
	Shift	99.6 ± 3.3	109.1 ± 8.2
	Broadening	100.6 ± 3.1	109.2 ± 6.6
	T_1 noise	28.6 ± 3.7	102.1 ± 3.4
Raman			
	Shot noise	5.6 ± 1.9	93.0 ± 1.1
	Fluorescence	25.7 ± 5.4	95.8 ± 1.1
	Cosmic ray	77.0 ± 5.4	98.0 ± 0.8
	Shift	99.3 ± 0.4	99.6 ± 0.3
	Broadening	89.8 ± 0.9	99.5 ± 0.9
IR/FTIR			
	Detector noise	90.9 ± 0.5	98.6 ± 0.1
	Baseline	99.8 ± 0.0	99.9 ± 0.0
	Atmospheric CO_2	99.8 ± 0.1	100.0 ± 0.0
	Atmospheric H_2O	94.6 ± 0.3	99.7 ± 0.0
	Shift	99.8 ± 0.0	100.2 ± 0.0
	Broadening	95.0 ± 0.4	100.1 ± 0.0

F Architectural Ablations

VINEA architecture ablation. We evaluated the contribution of the VINEA descriptor and encoder components using controlled ablations on the simulated MSD ^{1}H NMR test split. All variants used the same scaffold-disjoint split, molecular encoder, optimization budget, and corruption draws, thus only the component named in each row was changed. Models were trained with a linear S1-S3 corruption curriculum over the first 70% of training followed by an S3 plateau. Checkpoints were selected by clean validation MRR from epoch 75 onward and were then evaluated, without further selection, against 18,493 test candidates.

Ablation rationale. The ablations were organized around four questions: which parts of the VINEA descriptor carry chemically useful information, whether peak position must remain associated with the other attributes of the same topological branch, which encoder and readout mechanisms are necessary to combine branches, and whether the default 256-dimensional contrastive space is appropriately sized. Except for the dimension sweep, each intervention changes one mechanism while leaving the descriptor width, 256-dimensional spectrum and molecule outputs, molecular tower, loss, and training protocol unchanged. The dimension sweep changes the terminal projections of both towers together so that their cosine similarity remains well defined.

Descriptor construction. The *− blur axis* arm sets the branch-survival profile across Gaussian blur levels and the blur-death coordinate to zero. It therefore asks whether the second filtration axis

Supplementary Table 11: **Clean in-domain performance.** ^{1}H NMR uses the 18,493-candidate MSD test pool, HSQC uses the 500-candidate held-out MSD test pool, Raman uses the 5,056-candidate ChEMBL retrieval benchmark, and IR uses 17-label USPTO functional-group prediction. Results are mean ± s.d. over ten seeds. Bold marks the largest mean within each task. Parameter counts include the complete trained model.

Dataset and task	Endpoint	Model	Score	Parameters
MSD ^{1}H NMR	R@1	NMR2Struct	$0.798_{\pm 0.005}$	4.46M
		CHEFNMR	$0.691_{\pm 0.018}$	4.41M
		SECS	$0.691_{\pm 0.006}$	87.50M
		NMRViT	$0.676_{\pm 0.013}$	19.70M
		VINEA	$0.859_{\pm 0.002}$	1.05M
MSD HSQC	R@1	NMRViT	$0.051_{\pm 0.034}$	20.19M
		DeepSAT	$0.326_{\pm 0.047}$	149.83M
		HSQC-CNN	$0.242_{\pm 0.065}$	6.36M
		SPECTRE	$0.381_{\pm 0.036}$	117.24M
		VINEA	$0.465_{\pm 0.031}$	1.03M
ChEMBL Raman	R@1	DeepCNN	$0.784_{\pm 0.003}$	4.84M
		RamanNet	$0.629_{\pm 0.007}$	1.27M
		SANet	$0.797_{\pm 0.006}$	3.40M
		RamanFormer	$0.691_{\pm 0.005}$	4.89M
		ReZeroNet	$0.776_{\pm 0.163}$	0.87M
		ResNet-1D	$0.807_{\pm 0.006}$	0.92M
		Spectral Transformer	$0.785_{\pm 0.007}$	1.86M
		VINEA	$0.781_{\pm 0.005}$	1.05M
USPTO IR	mAUROC	IRCNN	$0.947_{\pm 0.001}$	64.37M
		SpectraML-CNN	$0.892_{\pm 0.138}$	2.05M
		ACT	$0.940_{\pm 0.001}$	0.66M
		UltraIR	$0.972_{\pm 0.001}$	129.64M
		VINEA	$0.951_{\pm 0.001}$	0.59M

provides useful evidence that a line persists under smoothing, beyond the persistence information available at the original resolution. The − *line shape* arm zeros the branch area, widths at three relative heights, shape ratio, and asymmetry. This tests whether local peak morphology helps distinguish molecular signals and suppress artifacts whose shape changes under noise, phase error, or shimming. The − *coupling gaps* arm zeros the distances to the nearest branches on either side, testing whether local spacing and multiplet context add information beyond the attributes of an isolated branch. The *unsigned filtration* arm discards sublevel branches while retaining positive superlevel branches, which tests whether negative phase lobes carry complementary information in imperfectly phased spectra. Finally, *absolute shifts only* keeps the absolute chemical-shift coordinates and their Fourier encoding but removes the pairwise relative-shift bias from self-attention. This isolates the proposed reference-tolerant pathway: pairwise separations are unchanged by a global referencing offset, whereas absolute positions are not.

Positional control. The *shuffled position* arm is a deliberately destructive negative control. Within each spectrum, it deterministically reassigns the paired absolute-position coordinates among valid branches while preserving the set of positions, token count, and every non-positional descriptor value. Hence the model sees the same marginal position distribution and the same collection of branch shapes and persistences, but no longer knows which morphology occurred at which chemical shift. This tests whether performance is driven by genuine position-feature correspondence rather than by independent summary statistics of peak locations and shapes.

Encoder and readout. Full VINEA first weights every token by persistence, applies an additional learned sigmoid gate, processes the set with two relative-shift attention blocks, and concatenates pooling-by-multi-head-attention (PMA), masked mean, and masked maximum readouts. The − *learned gate* arm removes only the adaptive sigmoid gate, retaining mandatory persistence weighting, and tests whether the network benefits from learning branch relevance beyond topological prominence. *One attention block* reduces interaction depth from two blocks to one. *DeepSets readout* removes both attention blocks and PMA, leaving masked mean and maximum pooling, which tests whether explicit

Supplementary Table 12: **Robustness after S1-S3 ramp training.** Models are trained with a linear corruption curriculum from severity 1 to 3 and evaluated on the clean, S3, and extrapolative S5 conditions. The Raman rows here are molecular retrieval on ChEMBL (5,056 candidates), not the pharmaceutical classification task in Table 11. cAUDC is the per-seed, chance-adjusted trapezoidal area over clean-S5; chance is $1/N$ for retrieval and 0.5 for mAUROC. Bold marks the largest mean within each modality and column; it does not imply statistical significance.

Modality/task	Model	n	Clean	S3	S5	cAUDC
^{1}H NMR	CHEFNMR	10	0.530 ± 0.025	0.324 ± 0.038	0.001 ± 0.000	0.330 ± 0.015
	NMR2Struct	10	0.677 ± 0.007	0.288 ± 0.085	0.000 ± 0.000	0.394 ± 0.020
	NMRViT	10	0.623 ± 0.016	0.446 ± 0.043	0.003 ± 0.001	0.411 ± 0.021
	SECS	10	0.535 ± 0.020	0.216 ± 0.103	0.000 ± 0.000	0.297 ± 0.028
	VINEA	10	0.778 ± 0.006	0.583 ± 0.022	0.018 ± 0.004	0.525 ± 0.010
HSQC	DeepSAT	10	0.118 ± 0.029	0.023 ± 0.017	0.002 ± 0.002	0.040 ± 0.008
	HSQC-CNN	10	0.135 ± 0.105	0.022 ± 0.026	0.002 ± 0.002	0.065 ± 0.054
	NMRViT	10	0.019 ± 0.012	0.002 ± 0.001	0.003 ± 0.002	0.007 ± 0.004
	VINEA-HSQC	10	0.362 ± 0.034	0.109 ± 0.049	0.002 ± 0.002	0.199 ± 0.019
Raman retrieval	DeepCNN	10	0.720 ± 0.007	0.623 ± 0.016	0.011 ± 0.003	0.507 ± 0.010
	RamanFormer	10	0.659 ± 0.004	0.566 ± 0.006	0.014 ± 0.004	0.463 ± 0.004
	RamanNet	10	0.564 ± 0.006	0.456 ± 0.019	0.027 ± 0.007	0.407 ± 0.012
	ResNet-1D	10	0.739 ± 0.007	0.617 ± 0.019	0.006 ± 0.002	0.510 ± 0.007
	ReZeroNet	10	0.737 ± 0.007	0.621 ± 0.020	0.007 ± 0.002	0.509 ± 0.009
	SANet	10	0.721 ± 0.008	0.578 ± 0.056	0.002 ± 0.001	0.474 ± 0.015
	Spectral Transformer	10	0.751 ± 0.006	0.653 ± 0.018	0.002 ± 0.001	0.509 ± 0.011
	VINEA	10	0.746 ± 0.006	0.623 ± 0.013	0.015 ± 0.003	0.519 ± 0.007
IR	ACT	10	0.931 ± 0.001	0.884 ± 0.019	0.656 ± 0.021	0.724 ± 0.019
	IRCNN	10	0.943 ± 0.001	0.866 ± 0.023	0.712 ± 0.016	0.727 ± 0.031
	SpectraML-CNN	10	0.890 ± 0.137	0.756 ± 0.094	0.568 ± 0.030	0.524 ± 0.186
	UltraIR	10	0.969 ± 0.001	0.772 ± 0.072	0.619 ± 0.044	0.600 ± 0.098
	VINEA	10	0.951 ± 0.001	0.927 ± 0.002	0.780 ± 0.011	0.822 ± 0.007

Supplementary Table 13: **Zero-shot transfer from synthetic MSD ^{1}H NMR to measured repositories.** Spectrum-to-molecule R@1 is evaluated without target-domain training. "Clean" and "ramp" identify clean-only and S1-S3-ramp synthetic pretraining. Candidate pools are 3,966 (Chemotion), 2,495 (BMRB), 552 (GISSMO), and 421 (NMRXiv). Bold marks the largest mean within each pretraining regime and repository. Unequal n in the ramp block reflects the completed seeds available at the freeze date.

Model	Pretraining	n	Chemotion	BMRB	GISSMO	NMRXiv
NMR2Struct	Clean	10	0.037 ± 0.004	0.032 ± 0.002	0.147 ± 0.008	0.186 ± 0.015
CHEFNMR	Clean	10	0.014 ± 0.006	0.017 ± 0.004	0.074 ± 0.014	0.078 ± 0.014
SECS	Clean	10	0.001 ± 0.001	0.002 ± 0.001	0.016 ± 0.005	0.018 ± 0.005
NMRViT	Clean	10	0.001 ± 0.001	0.001 ± 0.001	0.006 ± 0.003	0.007 ± 0.003
VINEA	Clean	10	0.021 ± 0.004	0.019 ± 0.003	0.071 ± 0.010	0.093 ± 0.014
Random	Clean	10	0.000 ± 0.000	0.000 ± 0.000	0.002 ± 0.002	0.003 ± 0.003
NMR2Struct	Ramp	10	0.173 ± 0.006	0.106 ± 0.008	0.278 ± 0.016	0.327 ± 0.019
CHEFNMR	Ramp	10	0.134 ± 0.006	0.070 ± 0.006	0.231 ± 0.008	0.273 ± 0.016
SECS	Ramp	10	0.051 ± 0.006	0.041 ± 0.006	0.136 ± 0.013	0.159 ± 0.007
NMRViT	Ramp	10	0.144 ± 0.007	0.125 ± 0.008	0.283 ± 0.012	0.303 ± 0.016
VINEA	Ramp	10	0.068 ± 0.014	0.047 ± 0.010	0.201 ± 0.027	0.221 ± 0.023
Random	Ramp	10	0.000 ± 0.000	0.000 ± 0.000	0.002 ± 0.002	0.003 ± 0.003

interactions among branches are necessary. The two remaining arms separate the readout components: − *PMA* retains both attention blocks but uses only mean and maximum pooling, whereas *PMA only* removes those fixed summary statistics and retains only learned query pooling. Together, these controls distinguish the contribution of contextual token mixing from the choice of invariant readout.

Shared embedding dimension. The 64-, 128-, and 512-dimensional arms replace the default 256-dimensional terminal projections in both the spectral and molecular towers. This sweep tests whether retrieval is limited by an overly narrow bottleneck, whether a smaller representation is sufficient, and whether additional embedding capacity improves robustness or instead adds unnecessary parameters. All other VINEA mechanisms remain unchanged.

Supplementary Table 14: **Zero-shot transfer to measured HSQC, Raman, and IR data.** Columns compare clean-only and S1-S3-ramp synthetic pretraining, with no target-domain optimization. HSQC reports R@1 over 105 NMRXiv candidates; Raman reports R@1 for 3,510 measured spectra over 31 chemical candidates; IR reports mAUROC on 7,749 measured spectra. All entries use ten seeds. Bold marks the largest mean within each modality and pretraining column.

Modality	Model	Clean pretraining	Ramp pretraining
HSQC	NMRViT	0.012 ±0.010	0.012 ±0.010
	DeepSAT	0.013 ±0.008	0.056 ±0.016
	HSQC-CNN	0.014 ±0.012	0.035 ±0.018
	SPECTRE	0.103 ±0.031	0.097 ±0.021
	VINEA-HSQC	0.021 ±0.013	0.020 ±0.014
	Random	0.010 ±0.010	0.010 ±0.010
Raman	DeepCNN	0.049 ±0.015	0.061 ±0.017
	RamanNet	0.032 ±0.016	0.053 ±0.025
	SANet	0.051 ±0.019	0.090 ±0.020
	RamanFormer	0.045 ±0.022	0.058 ±0.021
	Spectral Transformer	0.059 ±0.015	0.081 ±0.016
	ResNet-1D	0.020 ±0.009	0.084 ±0.021
	ReZeroNet	0.051 ±0.016	0.047 ±0.027
	VINEA	0.080 ±0.016	0.095 ±0.031
IR	IRCNN	0.665 ±0.008	0.674 ±0.005
	SpectraML-CNN	0.640 ±0.051	0.644 ±0.054
	ACT	0.681 ±0.009	0.692 ±0.013
	UltraIR	0.682 ±0.011	0.692 ±0.018
	VINEA	0.620 ±0.017	0.684 ±0.016
	Random	0.501 ±0.004	0.501 ±0.004

Supplementary Table 15: **Robustness and inference cost.** cAUDC is defined as in Table 12. Online parameters exclude the molecular tower for retrieval models; latency is median end-to-end time per clean spectrum at batch size one on an NVIDIA H100. cAUDC uses all completed seeds, whereas cost is measured from one seed-17 checkpoint because it is an architecture property. Only methods with a compatible latency profile are shown. Bold marks the largest mean cAUDC within each modality.

Modality	Model	n	cAUDC	Online params	Latency (ms)
^{1}H NMR	**VINEA**	10	0.525 ±0.010	0.652M	6.890
	NMR2Struct	10	0.394 ±0.020	4.059M	0.655
	SECS	10	0.297 ±0.028	87.150M	12.428
	NMRViT	10	0.411 ±0.021	19.269M	2.386
	CHEFNMR	10	0.330 ±0.015	4.015M	1.762
HSQC	**VINEA-HSQC**	10	0.199 ±0.019	0.631M	455.251
	NMRViT	10	0.007 ±0.004	19.796M	2.976
	HSQC-CNN	10	0.065 ±0.054	5.959M	0.786
	DeepSAT	10	0.040 ±0.008	149.430M	5.896
Raman	**VINEA**	10	0.519 ±0.007	0.652M	2.659
	RamanNet	10	0.407 ±0.012	0.867M	2.164
	SANet	10	0.474 ±0.015	3.006M	2.244
	RamanFormer	10	0.463 ±0.004	4.488M	1.132
	ReZeroNet	10	0.509 ±0.009	0.473M	1.509
	ResNet-1D	10	0.510 ±0.007	0.521M	1.006
IR	**VINEA**	10	0.822 ±0.007	0.591M	3.730
	IRCNN	10	0.727 ±0.031	64.365M	0.793
	SpectraML-CNN	10	0.524 ±0.186	2.047M	0.424
	ACT	10	0.724 ±0.019	0.661M	1.836
	UltraIR	10	0.600 ±0.098	129.643M	3.737

Interpretation. Among the descriptor interventions, removing line-shape information produces the largest robustness decrease, reducing MRR-AUDC from 0.533 to 0.482. The blur-axis ablation causes a pronounced loss on clean spectra (MRR 0.366 to 0.241) but a smaller MRR-AUDC decrease to 0.513, suggesting that multiscale survival is especially useful for retaining clean-spectrum specificity.

The coupling-gap, unsigned-filtration, and absolute-shift-only arms remain closer to the full model at the present seed count. These small differences do not show that the features are chemically irrelevant, which indicate that their incremental benefit is not resolved reliably by this interim aggregate.

The shuffled-position control produces by far the largest change: clean MRR falls from 0.366 to 0.017 and MRR-AUDC from 0.533 to 0.126. Because the arm preserves the marginal values of both position and the remaining descriptors, this result specifically supports the importance of associating each branch's chemical shift with its persistence, shape, and scale-space behavior.

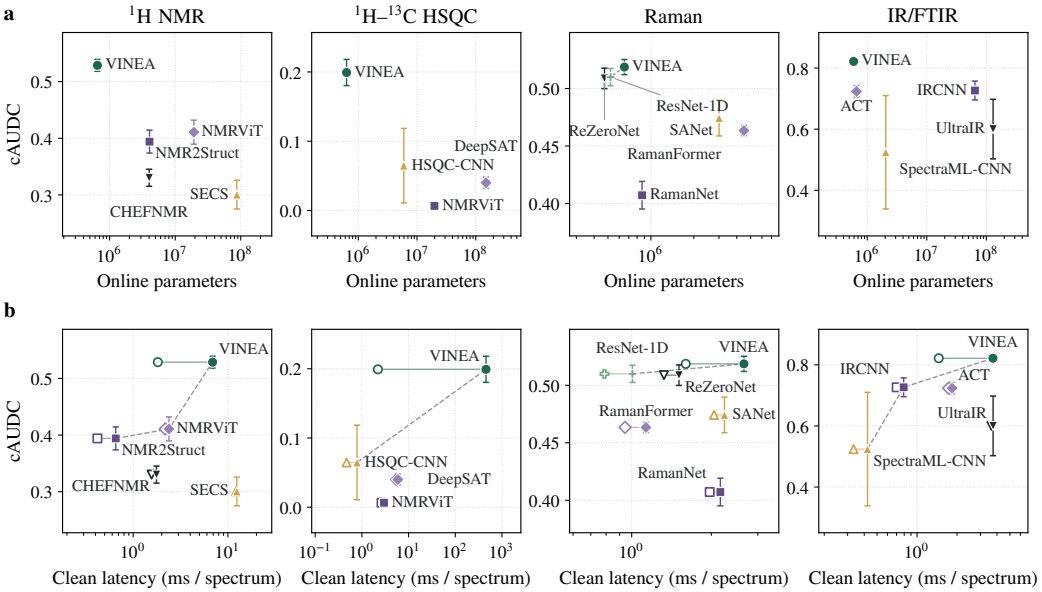
Within the encoder group, removing token interactions and learned pooling in the DeepSets arm gives the second-largest overall reduction (MRR-AUDC 0.392). By comparison, using one rather than two attention blocks reduces MRR-AUDC by 0.010, and removing the learned gate reduces it by 0.023. The no-PMA arm is nominally unchanged (ΔAUDC $= +0.001$), whereas PMA alone loses 0.021. The current means therefore suggest that contextual attention and the fixed mean/maximum summaries are more consequential than PMA itself, although the close readout comparisons require additional seeds before drawing a firm ranking.

Finally, embedding performance is not monotonic with dimension. The 128-dimensional arm has the largest mean MRR-AUDC (0.535), but differs from the 256-dimensional reference by only $+0.003$, and the 64-dimensional arm is also close at 0.528, while expanding to 512 dimensions does not improve the mean (0.524). Thus, these three-seed results provide no evidence that the default representation is bottlenecked by embedding width. The apparent efficiency of the smaller spaces should be treated as a design trend rather than a definitive selection until the remaining seeds are complete.

Supplementary Table 16: **Architecture and representation ablations of VINEA on the simulated MSD ^{1}H NMR test set.** Values are spectrum-to-molecule retrieval mean $\pm$ standard deviation over three training seeds (17, 29, and 41), using a scaffold-disjoint candidate pool of 18,493 molecules. Clean R@1 and MRR measure uncorrupted retrieval, S3 R@1 measures retrieval under the combined severity-3 corruption. MRR-AUDC is the mean reciprocal rank averaged uniformly over the combined S1-S5 corruption curve. Training used an S1-S3 corruption ramp, so S4 and S5 measure extrapolation beyond the training range. ΔAUDC is the paired difference relative to full VINEA. The full model uses a 256-dimensional shared contrastive embedding, and only the final three rows change this dimension. Bold indicates the largest mean in each column and does not imply a statistically significant difference.

Variant	Dim.	Clean R@1	Clean MRR	S3 R@1	MRR-AUDC	ΔAUDC
Full VINEA	256	0.276 ± 0.019	0.366 ± 0.021	0.619 ± 0.030	0.533 ± 0.018	$+0.000$
− blur axis	256	0.162 ± 0.010	0.241 ± 0.009	0.601 ± 0.015	0.513 ± 0.015	-0.020
− line shape	256	0.239 ± 0.017	0.328 ± 0.019	0.492 ± 0.016	0.482 ± 0.011	-0.050
− coupling gaps	256	0.277 ± 0.006	0.368 ± 0.009	0.608 ± 0.022	0.521 ± 0.019	-0.011
Unsigned filtration	256	0.274 ± 0.043	0.364 ± 0.046	0.610 ± 0.016	0.523 ± 0.004	-0.009
Absolute shifts only	256	0.266 ± 0.026	0.354 ± 0.032	0.592 ± 0.009	0.519 ± 0.008	-0.013
Shuffled position	256	0.007 ± 0.002	0.017 ± 0.004	0.117 ± 0.017	0.126 ± 0.009	-0.407
− learned gate	256	0.259 ± 0.014	0.348 ± 0.015	0.599 ± 0.026	0.510 ± 0.025	-0.023
One attention block	256	0.271 ± 0.009	0.362 ± 0.005	0.594 ± 0.005	0.523 ± 0.010	-0.010
DeepSets readout	256	0.193 ± 0.028	0.276 ± 0.031	0.357 ± 0.032	0.392 ± 0.021	-0.140
− PMA	256	0.267 ± 0.023	0.358 ± 0.026	0.621 ± 0.020	0.533 ± 0.011	$+0.001$
PMA only	256	0.217 ± 0.014	0.304 ± 0.020	0.598 ± 0.025	0.511 ± 0.010	-0.021
64-d embedding	64	0.261 ± 0.013	0.350 ± 0.012	0.618 ± 0.013	0.528 ± 0.014	-0.004
128-d embedding	128	0.279 ± 0.037	0.364 ± 0.041	0.625 ± 0.001	0.535 ± 0.005	$+0.003$
512-d embedding	512	0.266 ± 0.007	0.357 ± 0.007	0.612 ± 0.031	0.524 ± 0.016	-0.009

The three-seed results identify positional correspondence as the dominant factor in robust retrieval: shuffling the association between position and the remaining token features reduced MRR-AUDC from 0.533 to 0.126. Replacing the attention-based encoder with a DeepSets readout produced the next-largest drop (0.392), while removing line-shape descriptors reduced MRR-AUDC to 0.482. By contrast, removing PMA or using a 128-dimensional embedding produced performance close to the full 256-dimensional model. Increasing the embedding dimension to 512 did not improve the mean.



Supplementary Figure 12: **Robustness against inference cost for ramp-trained models.** Chance-adjusted area under the degradation curve (cAUDC; clean to S5; Section H.5) against **(a)** online parameter count and **(b)** median clean latency per spectrum at batch size one. Columns correspond to ^{1}H NMR, HSQC, Raman, and IR/FTIR. Points show means over training seeds, and vertical error bars show one standard deviation of cAUDC. In panel **(b)**, filled markers give end-to-end latency and hollow markers give GPU forward-pass latency. The connecting segment therefore represents non-forward overhead, dominated by construction of the model input view. Dashed lines connect the non-dominated models within each modality. In panel **(a)**, VINEA is the only non-dominated model for ^{1}H NMR, HSQC, and IR/FTIR. For Raman, ReZeroNet and ResNet-1D are also non-dominated and lie within one standard deviation of VINEA. Latency was profiled using seed-17 checkpoints on one NVIDIA H100. The largest VINEA latency overhead occurs for HSQC view construction (453 ms of 455 ms end-to-end), while its GPU forward pass requires 1.5–2.2 ms across modalities. The view-construction code was not optimised for speed. For Raman, ResNet-1D is the convolutional baseline that was profiled, and the reported robustness scores are its own.

Given the small number of training seeds, the close differences among full VINEA, the no-PMA variant, and the dimension variants should be interpreted as trends rather than evidence of a reliable ordering.

G Training

Objective. A spectral tower produces z_{spec} and a molecular graph tower z_{mol}, both 256-dimensional and ℓ_2-normalized. For a batch of B matched spectrum–molecule pairs, the models are trained with symmetric InfoNCE over the other examples in the batch,

$$\mathcal{L} = -\frac{1}{2B}\sum_{i=1}^{B}\left[\log\frac{\exp\left(\sigma\, z_{\text{spec},i}^{\top} z_{\text{mol},i}\right)}{\sum_{j=1}^{B}\exp\left(\sigma\, z_{\text{spec},i}^{\top} z_{\text{mol},j}\right)} + \log\frac{\exp\left(\sigma\, z_{\text{mol},i}^{\top} z_{\text{spec},i}\right)}{\sum_{j=1}^{B}\exp\left(\sigma\, z_{\text{mol},i}^{\top} z_{\text{spec},j}\right)}\right]. \tag{G.1}$$

Optimization. Batch size is 32 for every encoder and dataset. In InfoNCE the batch is the negative pool, so it is held constant across the comparison. The optimizer is AdamW at weight decay 10^{-4} with a cosine schedule and a five-epoch linear warmup. Batches with non-finite loss are skipped, gradient-norm clipping is applied where configured, and a diverged model is refused at scoring time (Section H.2).

Learning rates are set per encoder on validation data, because the transformers diverge at the rate that suits the convolutional encoders: 3×10^{-4} for NMR2Struct, CHEFNMR, NMRViT, NMRTrans and NMR-Text, 10^{-4} for SECS, and 3×10^{-5} for UltraNMR. All baselines have been run with a max iteration budget of 100 and only the best checkpoints were collected to measure the experiments.

Implementation details. VINEA uses at most 128 topological tokens, embedded with a two-layer self-attention encoder of width 128 and four attention heads. Chemical positions are represented using eight Fourier frequencies. For one-dimensional spectra, six additional frequencies encode relative spectral shifts. The resulting representation is projected to a 256-dimensional latent space. Models were trained with AdamW over ten random seeds. For ^{1}H NMR, HSQC, Raman, and IR, respectively, we used batch sizes of 32, 4, 64, and 64. Epoch budgets of 100, 30, 50, and 50, and learning rates of 3×10^{-4}, 10^{-4}, 3×10^{-4}, and 10^{-4}. Progressive training began with 20% clean epochs, followed by a linear corruption-severity ramp from S1 to S3, with each corruption applied independently with probability 0.5. For transfer to experimental spectra, the complete encoder was fine-tuned from the progressively trained checkpoint using a learning rate of 3×10^{-5}.

Selection, seeds and hardware. Training runs to a fixed epoch cap. Validation retrieval is measured on *clean* spectra at regular intervals and the best-scoring state is restored, so selection never observes a corrupted spectrum or the test split. Ten seeds are used throughout, fixed across models and datasets: $\{17, 29, 41, 53, 71, 83, 97, 109, 127, 149\}$. A seed controls initialization, shuffling and dropout only, since the test-time corruption is hashed per item and is identical across seeds by construction. All runs used a single NVIDIA H100 with PyTorch 2.13.0 and CUDA 12.6.

H Evaluation protocol

H.1 Retrieval pools

The candidate pool is the entire test split, and chance recall@1 is $1/N$ in that pool. Recall is therefore comparable within a dataset but not across datasets, and each table states its own chance level rather than averaging over pools.

Splits are taken over Bemis–Murcko scaffold groups, so no scaffold spans train, validation and test. Repeated measurements of a molecule are collapsed to one spectrum, since duplicates share a molecular embedding and index-order tie-breaking would otherwise cap attainable recall@1 below 1.0 for reasons unrelated to the representation.

Each condition corrupts a fresh copy of the test split, and every input view, dense trace, persistence landscape and peak list alike, is re-derived from that corrupted trace. The random state for an item is $\mathrm{blake2b}(\mathsf{condition} \mid \mathsf{index} \mid \mathsf{SMILES} \mid \mathsf{seed})$, so a given molecule under a given condition is corrupted byte-identically for every model and all comparisons are paired per molecule.

The molecular formula is zeroed for every method, leaving only nucleus identity on the context path. A formula-conditioned model can rank candidates without reading a spectrum at all and is corruption-invariant by construction, so formula-derived side information is withheld and reported separately as a control.

H.2 Metrics

Exact-match retrieval. Let q_i be a query embedding and c_j a candidate embedding. Both are ℓ_2-normalized before scoring, so

$$s_{ij} = q_i^\top c_j \tag{H.1}$$

is their cosine similarity. Query i is paired with candidate i, and its one-indexed rank is

$$\mathrm{rank}_i = 1 + \#\{j : s_{ij} > s_{ii}\} + \#\{j < i : s_{ij} = s_{ii}\}, \tag{H.2}$$

where the last term makes ties deterministic by candidate index. For a pool of N paired queries and candidates, we compute

$$\mathrm{R@}k = \frac{1}{N}\sum_{i=1}^{N} 1[\mathrm{rank}_i \le k], \qquad k \in \{1, 5, 10\}, \tag{H.3}$$

$$\mathrm{MRR} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\mathrm{rank}_i}, \qquad \mathrm{MeanRank} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{rank}_i. \tag{H.4}$$

Median rank is the median of the same per-query ranks. Higher R@k and MRR are better, whereas lower mean and median rank are better. Under a uniformly random ordering, $\mathbb{E}[\mathrm{R@}k] = \min(k, N)/N$ and $\mathbb{E}[\mathrm{MRR}] = H_N/N$, where $H_N = \sum_{r=1}^{N} 1/r$, in particular, chance R@1 is $1/N$.

The evaluator computes both spectrum-to-molecule and molecule-to-spectrum retrieval by exchanging q and c. Unless a direction is explicitly stated, the figures and headline tables use spectrum-to-molecule retrieval. Retrieval is exact-match: a structurally similar but non-identical candidate is counted as incorrect. The $N \times N$ similarity matrix is not retained in memory, its rank comparisons are accumulated in query blocks of 512 with no approximation to the resulting ranks.

IR/FTIR classification. IR spectra are evaluated as multilabel functional-group prediction. If a_ℓ denotes the one-versus-rest ROC area for label ℓ, the headline metric is

$$\text{macro-AUROC} = \frac{1}{|\mathcal{L}'|}\sum_{\ell \in \mathcal{L}'} a_\ell, \tag{H.5}$$

where $\mathcal{L}'$ contains labels having at least one positive and one negative example in the evaluated split. Each a_ℓ is the probability that a randomly chosen positive receives a larger score than a randomly chosen negative, with a half credit for ties. Thus chance macro-AUROC is 0.5, and the metric does not require a decision threshold. For thresholded secondary metrics, one global sigmoid threshold is selected on the validation split from 19 equally spaced values in $[0.05, 0.95]$ by micro-F1, then fixed for the test split. We additionally retain macro average precision, macro-F1, micro-F1 and per-label recall, but do not use them as the headline robustness metric.

Aggregation and degradation summaries. Metrics are first computed independently for each training seed. Plotted points are the arithmetic mean across available seeds, and error bars are one standard deviation across those seed-level values. The number of contributing seeds is reported whenever it is below ten. For a metric $m_{r,s}$ from seed r and severity s, clean-relative retention is

$$\mathrm{Retention}_{r,s} = 100\,\frac{m_{r,s}}{m_{r,0}}. \tag{H.6}$$

This is a within-model ratio, can exceed 100% when corruption improves a score, and can flatter a model with weak clean performance. When an unadjusted area under the degradation curve is reported, it is the trapezoidal mean over the equally spaced clean–S5 grid,

$$\mathrm{AUDC}_r = \frac{\frac{1}{2}m_{r,0} + m_{r,1} + m_{r,2} + m_{r,3} + m_{r,4} + \frac{1}{2}m_{r,5}}{5}. \tag{H.7}$$

The chance-adjusted cAUDC used for efficiency analysis is defined separately in Section H.5.

Non-finite embeddings make every rank comparison false, placing every query at rank one and reporting a perfect, perfectly noise-invariant score for a diverged model. Such runs are refused and flagged rather than scored.

H.3 Experimental and acquisition transfer

Experimental-domain transfer uses the same checkpoint and preprocessing rule as the corresponding source-domain evaluation, without fitting or selecting on the experimental test set. For NMR, HSQC and Raman, each experimental spectrum is used as a query against the complete paired molecular candidate pool and the headline score is spectrum-to-molecule R@1. For IR/FTIR, the unchanged

multilabel predictor is applied to the experimental spectra and scored by macro-AUROC as defined above. Scores are computed per training seed and then reported as the seed mean and standard deviation. Since candidate-pool size sets retrieval chance, R@1 values are compared only within the same transfer dataset.

Field-strength transfer is evaluated separately on GISSMO. Models trained on 400-MHz spectra are applied without adaptation to clean spectra simulated at each indicated test field strength. At every field, all models use the same 113-molecule candidate pool, and the reported metric is spectrum-to-molecule R@1. Its random baseline is therefore $1/113$. Field strength is not supplied to the encoder as side information.

H.4 Representation-geometry analysis

Geometry is measured on spectral embeddings of the same test examples in the same row order under clean and corrupted conditions. For the local analysis, we select once a deterministic subset of $n = \min(N, 512)$ examples using random seed 17 and reuse it for every severity. After ℓ_2 normalization, let $\mathcal{N}_{10}^{(s)}(i)$ denote the indices of the ten largest cosine similarities to example i at severity s, excluding i itself. Local neighborhood retention is

$$\mathrm{NN}_{10}(s) = \frac{100}{n} \sum_{i=1}^{n} \frac{\left| \mathcal{N}_{10}^{(0)}(i) \cap \mathcal{N}_{10}^{(s)}(i) \right|}{10}. \tag{H.8}$$

It is 100% when every clean ten-neighbor set is preserved and 0% when none is preserved.

To put retained task performance on a common scale across retrieval and classification, the corresponding task score is mapped relative to clean and chance:

$$T_{r,s} = 100 \max\left\{0, \frac{m_{r,s} - c}{m_{r,0} - c}\right\}, \qquad s \in \{1, \ldots, 5\}, \tag{H.9}$$

with $c = 1/N$ for R@1 and $c = 0.5$ for macro-AUROC. The lower clipping prevents below-chance scores from dominating the scale. The values above 100% remain possible if a corrupted score exceeds its clean counterpart. Each faint point in the geometry panel is one checkpoint seed. Curves and error bars show the mean and standard deviation over seeds for each model–severity cell.

The reported ρ_{pool} is Spearman's rank correlation across all model–severity mean points. Because both neighborhood and task retention decline mechanically with severity, we also report $\rho_{\mathrm{within}\,S}$: at each severity, the cross-model mean is subtracted separately from both axes, and Spearman correlation is computed over the pooled residuals. The latter therefore tests whether models that preserve more local geometry than their same-severity peers also retain more task performance, neither coefficient establishes causality.

For global geometry summaries, linear centered-kernel alignment between clean X and corrupted Y is

$$\mathrm{CKA}(X, Y) = \frac{\|X_c^\top Y_c\|_F^2}{\sqrt{\|X_c^\top X_c\|_F^2 \|Y_c^\top Y_c\|_F^2}}, \tag{H.10}$$

where X_c and Y_c are column-centered matched embedding matrices. Linear CKA is invariant to isotropic rescaling and orthogonal transformations, equals one for identical geometry, and is undefined for a constant embedding. To measure dimensional expansion or collapse, let λ_j be the eigenvalues of $Z_c^\top Z_c$ and $p_j = \lambda_j / \sum_k \lambda_k$. The entropy effective rank and its clean-relative value are

$$r_{\mathrm{eff}}(Z) = \exp\left(- \sum_{j : p_j > 0} p_j \log p_j \right), \qquad r_{\mathrm{rel}}(s) = \frac{r_{\mathrm{eff}}(Z_s)}{r_{\mathrm{eff}}(Z_0)}. \tag{H.11}$$

The t-SNE panels are qualitative visualizations, not additional performance metrics. For one representative checkpoint per displayed model, 350 matched clean/S2 examples are sampled with seed 17, concatenated, reduced to at most 40 principal components, and jointly embedded with t-SNE (perplexity 35, PCA initialization, automatic learning rate, 1,000 iterations and random seed 17). Model-specific maps are similarity-aligned through their matched clean points to the VINEA clean

map, removing arbitrary translation, rotation, reflection and global scale. Curves are Gaussian-kernel density contours enclosing 80% of the estimated mass. The pale dashed contour is the clean reference. Because each model is projected separately, contour shape and overlap are interpreted only qualitatively, and t-SNE axis distances are not compared across modalities or treated as scores.

H.5 Robustness–efficiency analysis

Chance-adjusted robustness area. The Pareto analysis uses recall@1 for ^{1}H NMR, HSQC and Raman, and macro-AUROC for IR. Let $m_{r,s}$ be the test score for training seed r at severity $s \in \{0, 1, \ldots, 5\}$, where $s = 0$ denotes the clean condition, and let c be the chance score. We map chance to zero and a perfect score to one using

$$q_{r,s} = \frac{m_{r,s} - c}{1 - c}. \tag{H.12}$$

For retrieval, $c = 1/N$ for the corresponding candidate pool ($N = 18{,}493$, 500 and $5{,}056$ for NMR, HSQC and Raman, respectively). for macro-AUROC, $c = 0.5$. Values below chance are retained rather than clipped. The chance-adjusted area under the degradation curve is the trapezoidal integral over the equally spaced clean–S5 grid,

$$\mathrm{cAUDC}_r = \frac{\frac{1}{2}q_{r,0} + q_{r,1} + q_{r,2} + q_{r,3} + q_{r,4} + \frac{1}{2}q_{r,5}}{5}. \tag{H.13}$$

We average cAUDC_r over the available training seeds. Because the trapezoidal operator is linear, integrating the condition-wise seeded means gives the same mean whenever the same seed set is present at every severity. This cAUDC is used only for the robustness–efficiency figure and is distinct from the unadjusted AUDC defined above.

Parameter counts. We count parameters from the instantiated checkpoint architecture, not from its serialized file size. For retrieval models, the online count comprises the spectral backbone, metadata-context encoder and fusion layer. It excludes the molecular graph tower, contrastive temperature and candidate-library embeddings, which are not evaluated for an incoming spectral query. For IR, the online count includes the spectral encoder and its classification head. The complete trainable parameter count, including the molecular tower where applicable, is retained separately in the profiling record.

Latency profiling. All displayed ramp-trained architectures are profiled from their seed-17 checkpoint on one NVIDIA H100, with batch size one, four CPU threads, `torch.inference_mode()` and compilation disabled. For each model we use the same first 48 held-out clean spectra within a modality, preceded by six unrecorded warm-up queries. Spectra and checkpoints are loaded before timing. The end-to-end timer starts from an in-memory raw spectrum and includes construction of the model-specific view (dense grid, peak set or topological descriptor), host-to-device transfer, the spectral forward pass and output normalization. It excludes disk access, synthetic corruption, the molecular tower, candidate-library encoding, similarity search and metric calculation. Consequently the comparison measures online spectral-query cost rather than the complete offline retrieval experiment.

CPU preprocessing and end-to-end time are measured with a monotonic wall clock. The GPU forward pass is measured with CUDA events, with device synchronization before and after each timed query. We report the median and 95th-percentile latency over the 48 queries, throughput as 1000 divided by median milliseconds per query, and peak allocated GPU memory. The Pareto figure uses median clean end-to-end latency. This initial profile does not claim severity-independent latency.

Pareto frontier. Frontiers are computed independently within each modality. A model i is non-dominated if there is no model j with cost $x_j \le x_i$ and robustness $y_j \ge y_i$, with at least one inequality strict. Cost is either online parameter count or median end-to-end latency, both displayed on a logarithmic axis, and robustness is mean cAUDC. Dashed segments connect the non-dominated models in increasing-cost order. They are a visual guide and do not interpolate unmeasured architectures.

H.6 PROBING PROTOCOL

To test whether the collapse under corruption is specific to nearest-neighbor ranking, the trained spectral towers are reused with the molecular tower removed and a fixed head attached: $\mathrm{LayerNorm} \to \mathrm{Linear}(256, 256) \to \mathrm{GELU} \to \mathrm{Dropout}(0.2) \to \mathrm{Linear}(256, \cdot)$, identical for every architecture and task. Two encoder modes exist: `frozen` loads a trained retrieval checkpoint and updates only the head. `scratch` trains tower and head jointly.

Chemical probes are computed from the reference structure with RDKit (Landrum et al., 2026): five classification targets (aromaticity, carbonyl, heterocycle, halogen, and a three-way aromatic-ring count) and five regression targets (molecular weight, Wildman–Crippen $\log P$, TPSA, rotatable bonds, double bonds). Classification uses cross-entropy with inverse-frequency class weights from the training split and is scored by balanced accuracy,

$$\mathrm{BAcc} = \frac{1}{C} \sum_{c=1}^{C} \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_c}, \tag{H.14}$$

whose chance level is $1/C$ irrespective of class imbalance. Regression uses MSE on targets standardized with training-split statistics. Predictions are mapped back to the original units before reporting

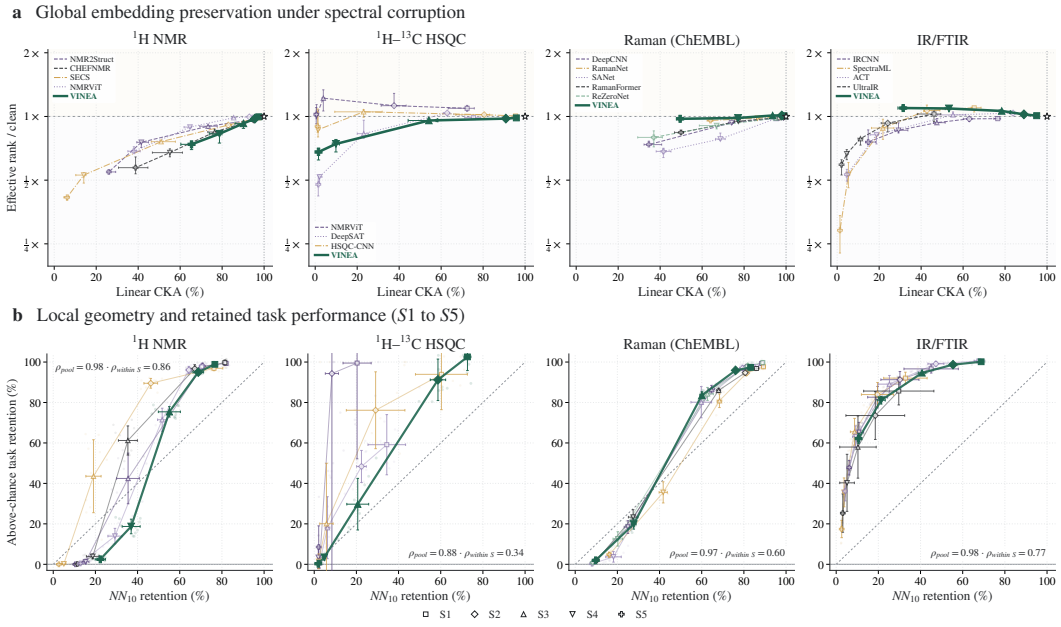
$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\widehat{y}_i - y_i)^2}. \tag{H.15}$$

Every score is reported against a naive baseline, the majority training class or the training mean, evaluated on the same corrupted targets. Signed *lift* is model minus baseline for higher-is-better scores and baseline minus model for error metrics, so positive always means better than guessing. The targets are derived from the answer rather than measured independently, so they establish only whether an embedding carries a given piece of chemistry.

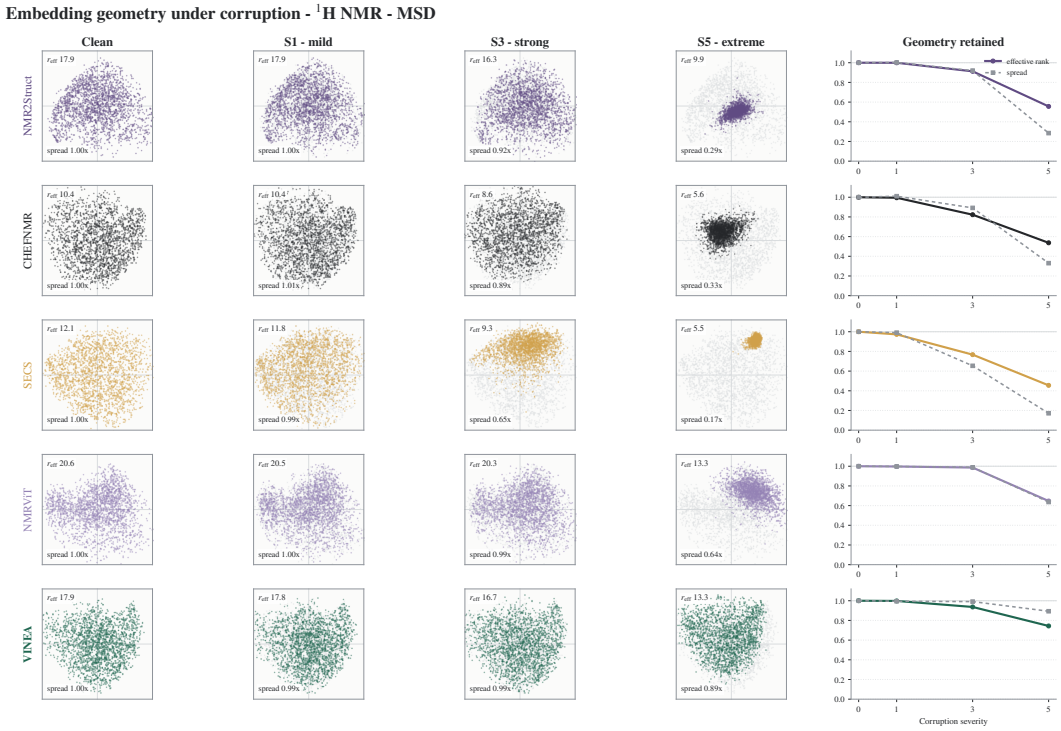
I EMBEDDING GEOMETRY OVER CORRUPTIONS

Geometry of the representation under corruption. VINEA best preserves clean-like global geometry for NMR, HSQC, and IR at S5 (median CKA 65.4% and 31.5% for NMR and IR vs. 38.8% and 14.8%), except on Raman, where RamanNet retains higher CKA at S4-S5; meanwhile, its effective rank stays near or above clean levels (97.5% Raman, 109.6% IR), indicating reorganization rather than collapse, while HSQC degrades most (CKA near zero at S4-S5, 67.9% of clean rank at S5). Task retention tracks local-neighborhood retention closely (pooled Spearman 0.88-0.98; within-severity 0.34-0.86, weakest for HSQC), suggesting that robustness depends mainly on preserving local molecular neighborhoods (Figure 13).

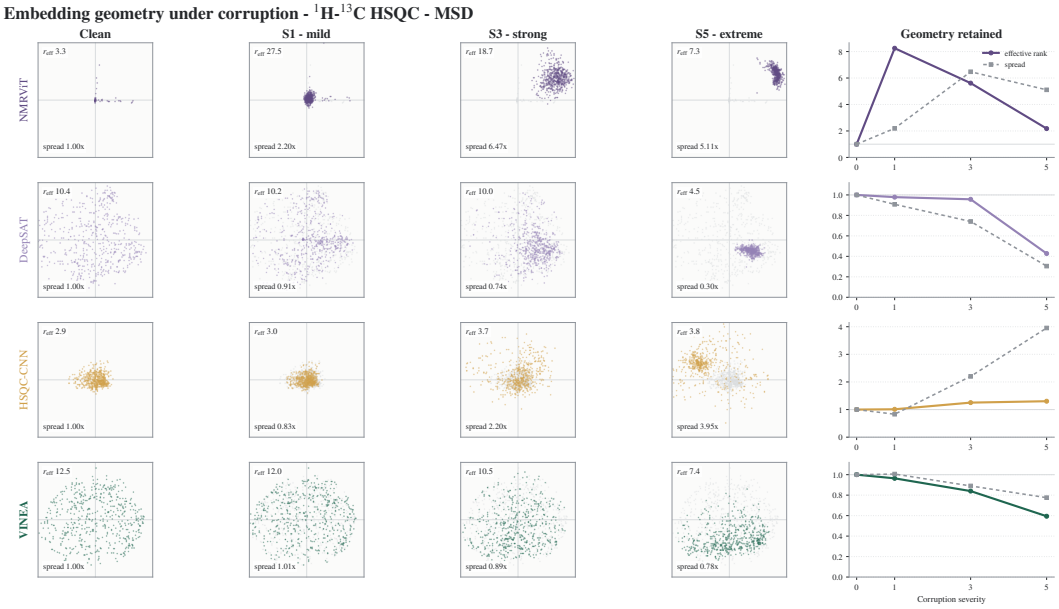
Embedding geometry under corruption. To visualize how corruption changes each learned representation, we project clean and corrupted embeddings onto a fixed two-dimensional basis fitted only to the clean embeddings. The basis and axis limits are held constant within each model row, so contraction, expansion, and displacement remain visible. These projections complement the quantitative CKA and neighbourhood analyses in Figure 13.



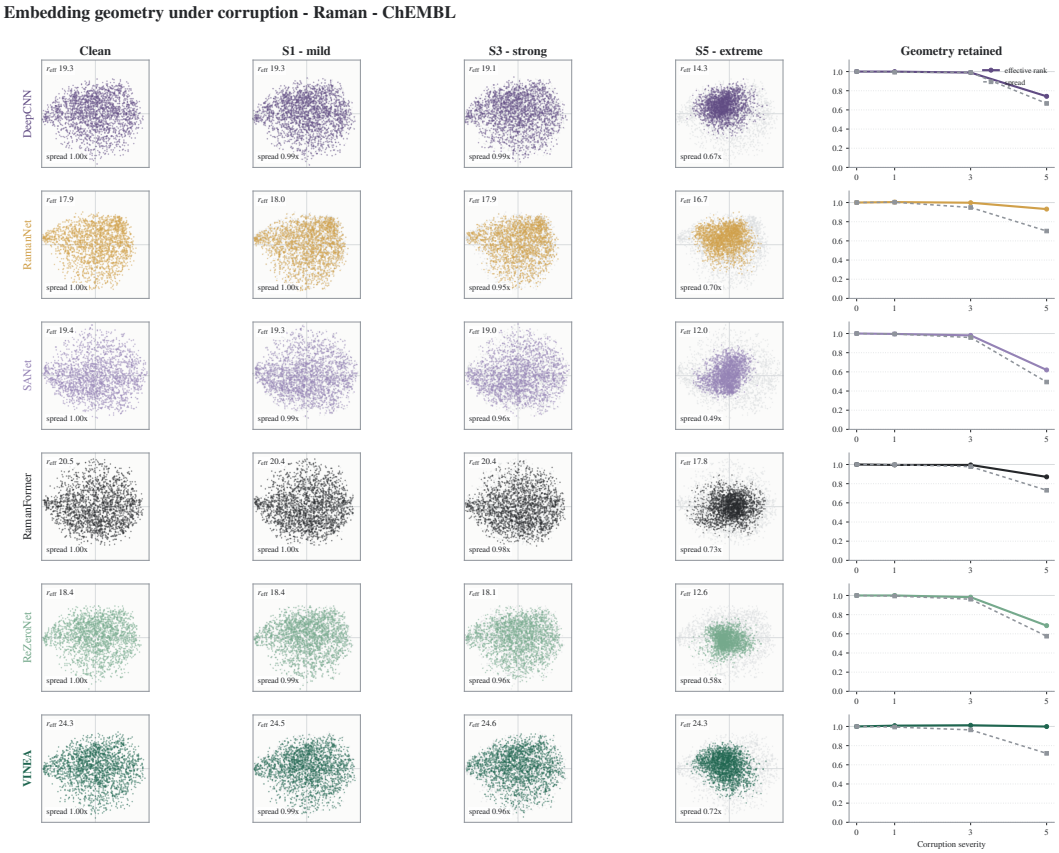
Supplementary Figure 13: **Global and local preservation of learned spectral representations under corruption. a,** Linear CKA between matched clean and corrupted test embeddings is plotted against effective rank relative to the corresponding clean embedding. Trajectories follow corruption levels S1 to S5. Points and whiskers show the median and interquartile range across seeds. The star marks unchanged clean geometry and dimensionality. **b,** Retention of the ten nearest neighbours is compared with retention of above-chance task performance. Pale points show individual seeds and connected symbols show the mean ± standard deviation. Symbol shapes identify corruption severity and model colours are shared between rows. VINEA is shown in green. NMR and HSQC use the MSD test sets, Raman uses ChEMBL, and IR/FTIR uses the UltraIR USPTO fold 1 test set.



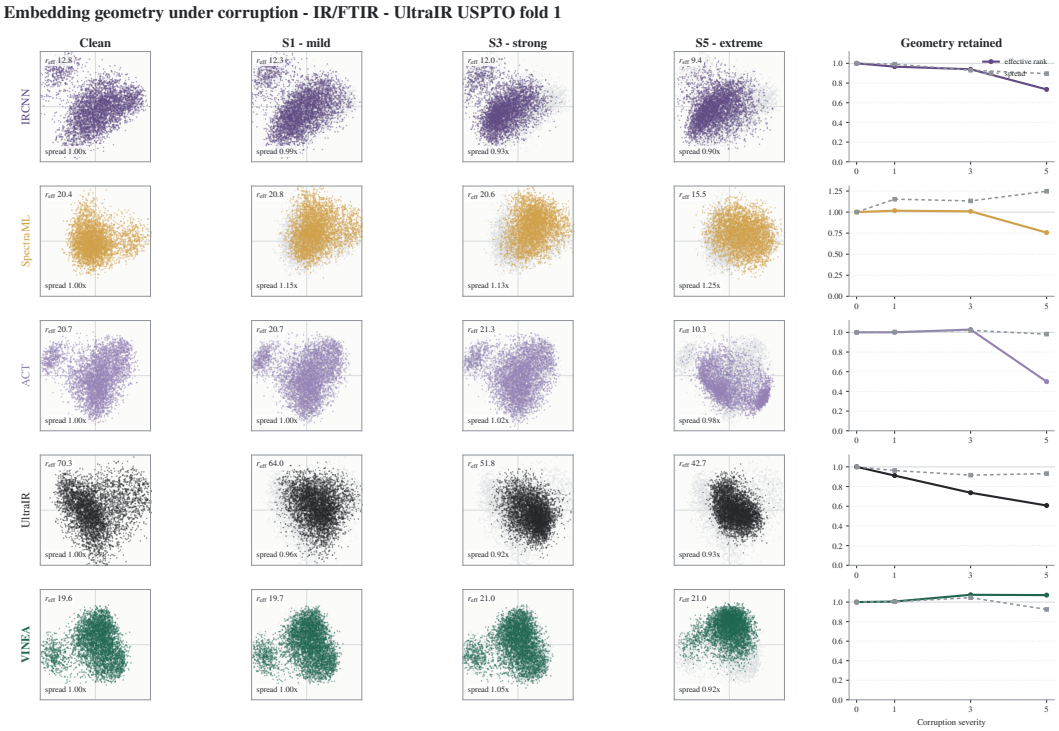
Supplementary Figure 14: **Embedding geometry under NMR corruption on MSD.** Each row shows one encoder trained with the S1 to S3 corruption ramp and evaluated on clean, S1, S3, and S5 spectra. The two-dimensional PCA basis is fitted to the clean embeddings for that model and reused at every severity. Coloured points show the indicated condition and grey points reproduce the clean reference. The effective rank r_{eff} and centred RMS spread are calculated in the original embedding space. The right column reports both quantities relative to clean. VINEA preserves a broad embedding through S3 and shows substantially less contraction than the dedicated NMR encoders at S5. A representative training seed is shown.



Supplementary Figure 15: **Embedding geometry under HSQC corruption on MSD.** Rows correspond to the corruption-ramp-trained encoders and columns show clean, S1, S3, and S5 spectra. Each model is displayed using its own fixed clean-PCA basis and shared row limits. Grey points indicate the clean reference. Effective rank and centred RMS spread are computed before projection and normalized to their clean values in the right column. The large expansion observed for some baselines reflects geometric instability rather than improved information content. VINEA changes gradually through S3 and avoids this extreme expansion before the severe S5 degradation. A representative training seed is shown.



Supplementary Figure 16: **Embedding geometry under Raman corruption on ChEMBL.** The fixed clean-PCA views compare clean, S1, S3, and S5 embeddings for each corruption-ramp-trained encoder. Coloured points show corrupted embeddings and grey points show the clean reference in the same coordinate system. Effective rank and centred RMS spread are measured in the original embedding space and summarized relative to clean in the right column. VINEA maintains its effective dimensionality across the full corruption range, including S5, while its spread decreases only under the strongest corruption. The corrected native-frequency Raman embeddings are used. A representative training seed is shown.



Supplementary Figure 17: **Embedding geometry under IR/FTIR corruption on UltraIR USPTO fold 1.** Each row shows clean, S1, S3, and S5 embeddings in a PCA basis fitted once to that model's clean embeddings. Grey points provide the clean reference. Effective rank and centred RMS spread are evaluated in the original embedding space and reported relative to clean in the right column. VINEA retains both its effective dimensionality and overall spread through S5, indicating that severe corruption reorganizes the representation without causing global dimensional collapse. A representative training seed is shown.