Language Models Can Understand Spectra: A Multimodal Model for Molecular Structure Elucidation

Yunyue Su[1†], Jiahui Chen[1†], Zao Jiang[3], Zhenyi Zhong[2], Liang Wang[1,4,5], Qiang Liu[1*]

[1*]NLPR, Institute of Automation, Chinese Academy of Sciences, 95 Zhongguancun East Road, Haidian, 100190, Beijing, China.
[2]College of Intelligence and Computing, Tianjin University, No. 135 Yaguan Road, Haihe Education Park, Peiyangyuan, 300354, Tianjin, China.
[3]Department of Computing, The Hong Kong Polytechnic University, 11 Yucai Road, Hung Hom, 000000, Hong Kong, China.
[4]School of Artificial Intelligence, University of Chinese Academy of Sciences, No.1 Yanqihu East Rd, Huairou, 101408, Beijing, China.
[5]National University of Singapore, 21 Lower Kent Ridge Rd, 119077, Singapore, Singapore.

*Corresponding author(s). E-mail(s): qiang.liu@nlpr.ia.ac.cn;
Contributing authors: yunyue.su@ia.ac.cn; jiahui.chen@ia.ac.cn;
zao311.jiang@connect.polyu.hk; zhenyi_zhong@tju.edu.cn;
liang.wang@cripac.ia.ac.cn;
[†]These authors contributed equally to this work.

Abstract

Structure elucidation is a fundamental technique for understanding the microscopic composition of matter and is widely applied across various disciplines in the natural sciences and engineering. However, existing methods often rely heavily on prior databases or known structural information, making it difficult to resolve unknown structures. In addition, complex structures typically require the joint analysis of multiple spectroscopic modalities. This process heavily depends on expert domain knowledge and is often accompanied by high costs in terms of both time and instrumentation. To address these challenges, we propose **SpectraLLM**, the first large language model designed to support multi-modal spectroscopic joint reasoning. SpectraLLM is capable of processing either single or multiple spectroscopic inputs and performing end-to-end structure elucidation. By integrating continuous and discrete spectroscopic modalities into a shared semantic space, SpectraLLM learns to uncover substructural patterns that are consistent and complementary across spectra, enabling precise molecular structure elucidation. We pretrain and fine-tune SpectraLLM in the domain of small molecules, and evaluate it on six standardized, publicly available chemical datasets. The model achieves state-of-the-art performance, significantly outperforming existing approaches trained on single modalities. Notably, SpectraLLM demonstrates strong robustness and generalization even for single-spectrum inference, while its multi-modal reasoning capability further improves the accuracy of structural prediction.

Keywords: Structure Elucidation, Spectra, Large Language Model, Multimodal

1 Introduction

Structure elucidation is fundamental across modern chemistry, biology, and materials science, enabling the determination of molecular and crystal structures [1–6]. It supports mechanistic studies and functional analysis in biomolecules [7, 8], guides material design and defect analysis [9, 10], and optimizes product performance [11, 12]. Powered by techniques such as Cryo-EM [13], NMR [14], IR [15], and MS [16], it involves interpreting spectral features—e.g., peak shifts, intensities, fragmentation patterns—to identify functional groups and connectivity patterns [17, 18]. These are matched against spectral databases (e.g., NIST [19], MoNA [20], SDBS [21], GNPS [22]) to infer candidate structures [23, 24].

IR, Raman, UV-Vis, NMR, and MS are among the most widely used techniques for molecular structure elucidation, each probing distinct physicochemical properties through interactions with electromagnetic radiation or ionization [11, 25–31]. As each arises from a fundamentally different physicochemical principle and captures orthogonal aspects of molecular structure, we refer to them as distinct spectral modalities, analogous in spirit to multimodal signals in domains such as vision and language. Table 1 summarizes the key characteristics of these modalities, including their detection mechanisms, structural sensitivities, and characteristic outputs. Together, they provide complementary molecular insights that are difficult to obtain from any single technique alone. IR and Raman identify functional groups and vibrational modes, while UV-Vis reveals electronic transitions [32–35]. However, these methods often suffer from peak overlap and background interference, requiring expert interpretation [36]. NMR offers rich atomic-level information, especially with 2D techniques that reveal inter-nuclear correlations, but remains limited by spectral complexity and reliance on prior knowledge [37–39]. MS provides high sensitivity and throughput, yet fragmentation variability and structural ambiguity across ionization modes and instrumentation remain major challenges [40–42].

In recent years, machine learning (ML) and deep learning (DL) have emerged as powerful tools for tackling the "Automated Spectrum-to-Structure" problem, thanks to their capabilities in feature extraction and pattern recognition [43]. Early work focused on shallow models such as SVMs, random forests, MLPs, and k-nearest neighbors [44–49], which showed effectiveness in functional group recognition and structure classification [50–54]. Traditional ML methods have also been used to regress molecular properties from spectral data [55, 56], but often lack generalization across different chemical domains. For NMR-based analysis, ML has been applied to substructure identification and candidate selection [57–59], though scalability remains a concern due to combinatorial explosion in complex scenarios [60, 61].

As a result, increasing attention has been given to end-to-end deep learning approaches that aim to bypass intermediate fragment reasoning and directly predict the complete molecular structure from spectra. These methods typically rely on powerful architectures such as convolutional neural networks (CNNs)[62–64] and Transformers[65–67]. Early efforts mainly employed CNNs for automated spectral feature extraction and classification [68–70]. For instance, Kuhn et al. [71] used CNNs to analyze substructures in mixtures, while Zhao et al. [72] combined wavelet transforms to improve Raman-based mixture component identification. In Spec2Mol[40], a CNN encoder was used to extract MS/MS spectral features, followed by an RNN decoder that directly generated SMILES expressions. In recent years, Transformers have become the dominant architecture due to their powerful sequence modeling capabilities [36, 65, 73, 74]. Many researchers now model spectral analysis as a sequence-to-sequence (Seq2Seq) task, employing encoder-decoder structures to directly generate SMILES [75–83]. Hu et al. [84] and Kanakala et al. [85] both used CNNs to extract spectral features before applying Transformer-based molecular generation. Alberts et al. [36] trained models under the condition of simultaneously inputting the molecular formula and spectrum, while French et al. [86] was the first to achieve small-molecule reconstruction from infrared spectra without relying on structure libraries or expert rules. Although Transformer-based models have made significant progress, their autoregressive decoding and limited expressiveness still pose challenges in structure generation. Recently, DiffMS [87] introduced a novel framework based on diffusion models, framing the "inverse mass spectrometry" problem as a conditional molecular generation task. By incorporating the DiGress graph diffusion generator [88] and Transformer encoders, together with molecular formula priors and pretrained molecule embeddings, DiffMS significantly improves the diversity and synthesizability of generated molecules, while also enhancing generalization in low-resource settings. Despite the impressive progress of AI in structure elucidation, most existing models still rely on single-modality spectra. In contrast, human experts routinely integrate infrared, NMR, and mass

Table 1: Overview of spectroscopic modalities.

Modality	Detection Principle	Structural Sensitivity	Common Output Features
IR	Molecular vibration absorption	Functional groups with dipole moment changes	Peak positions, intensities
Raman	Inelastic light scattering	Symmetric bonds and polarizability changes	Shifted peak patterns
UV-Vis	Electronic transitions	Conjugated systems, $\pi - \pi^*$ and $n - \pi^*$ transitions	Absorbance wavelengths, spectra
^{13}C NMR	Carbon-13 spin resonance	Carbon backbone structure, hybridization	Chemical shifts of ^{13}C nuclei
^{1}H NMR	Proton spin resonance (1H)	Proton environments, local chemical shifts	Peak multiplicity, integration
HSQC NMR	^{13}C-^{1}H correlation spectroscopy	Coupling between proton and carbon atoms	2D cross-peaks (^{13}C-^{1}H pairs)
Mass (MS)	Mass-to-charge ratio (m/z)	Molecular weight and fragmentation patterns	m/z peaks, isotopic distributions

spectrometry data to perform complementary analysis, reducing ambiguity and improving predictive accuracy.

In this study, we introduce SpectraLLM, a multimodal model for molecular structure elucidation from spectroscopic data. Unlike previous modality-specific approaches, SpectraLLM employs a unified language-based architecture that accepts structured descriptions of one or more spectral modalities—including infrared, Raman, ultraviolet, and mass spectrometry—and directly infers the molecular structure via natural language reasoning. By transforming spectral peaks into textual prompts that capture key physical attributes such as wavenumber, intensity, and peak shape, our model enables fine-grained, interpretable interactions across disparate spectroscopic channels. In contrast to specialized neural architectures, the language-centric formulation further allows flexible encoding of experimental conditions and instrumentation metadata, which are often difficult to represent numerically but readily expressed in natural language—enabling broader data integration and context-aware inference. SpectraLLM achieves state-of-the-art performance across four public benchmark datasets, outperforming both traditional spectrum-to-structure pipelines and recent multimodal transformer-based baselines. Notably, the model exhibits strong generalization across modality combinations: it supports both single-spectrum inputs (e.g., MS-only or IR-only) and joint inference from multiple spectra, with performance gains increasing with spectral diversity. These results position SpectraLLM as a scalable and robust framework for automated spectroscopic analysis, and demonstrate the feasibility of a language-centric paradigm for solving complex molecular reasoning tasks.

The main contributions of this work are as follows:

- We propose a language-based model for molecular structure elucidation that performs symbolic reasoning over multimodal spectral data.
- By expressing spectral peaks and experimental conditions in natural language, the proposed model supports both single- and multi-spectrum inputs, and generalizes effectively across modality combinations, benefiting from spectral complementarity.
- SpectraLLM achieves state-of-the-art performance on four public chemical datasets. The model shows strong robustness and generalization in single-spectrum settings, and further improves accuracy when jointly reasoning over multiple spectra.

2 Results

2.1 Joint fine-tuning of multiple spectra: aligns multimodal information and molecular structure

To enable effective structure elucidation from spectral data, we fine-tune a large language model (Qwen3–32B [89]) using instruction-style prompts constructed from peak-level features. As illustrated in Fig. 1, we begin by extracting characteristic peaks from raw spectroscopic vectors—such as infrared, Raman, ultraviolet, nuclear magnetic resonance and mass spectra. These peaks capture key local features including position, intensity, peak width, and relative prominence, and are used to compose descriptive prompts that encode either single-modality or multi-modality spectral inputs. All prompts are expressed in natural language, ensuring uniform processing by the language model.
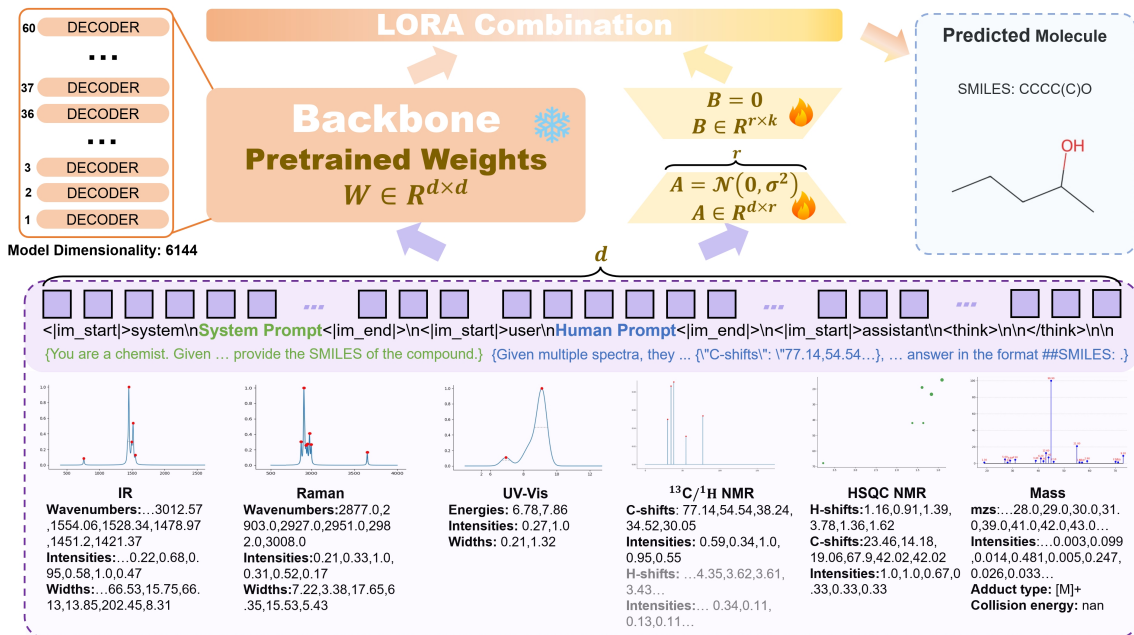
Fig. 1: Overview of the training pipeline for structure elucidation. Characteristic spectral peaks are extracted from raw IR, Raman, UV, NMR, or MS data and used to construct natural language prompts. These are input to a frozen large language model fine-tuned via LoRA. The model is trained to autoregressively generate molecular structures in SMILES format, supervised by the ground-truth sequence.

During supervised training, the model receives a peak-derived prompt and is tasked with generating the corresponding molecular structure in SMILES format. The ground-truth SMILES, treated as target sequence, enables sequence-level supervision. The training objective minimizes the standard autoregressive cross-entropy loss between predicted and ground-truth tokens. To adapt the base model to the spectral reasoning task while preserving its general language capabilities, we apply parameter-efficient fine-tuning via Low-Rank Adaptation (LoRA) [90], freezing all backbone parameters and updating only a small number of task-specific projection layers.

This training pipeline ensures consistency between the prompt encoding and symbolic generation stages, both framed entirely in the linguistic domain. At inference time, the procedure mirrors training: peak-level features from one or more spectra are converted into a unified textual prompt, which is directly decoded by the model into a candidate SMILES structure. No additional retrievers, image encoders, or external constraints are applied at test time, ensuring end-to-end generalizability.

2.2 Synergistic interaction among multiple spectra

In line with our central hypothesis that integrating complementary spectroscopic modalities enhances molecular structure prediction, we systematically evaluated model performance under controlled combinations of input spectra. Using the Multimodal Spectroscopic dataset, which includes IR, NMR (^{1}H, ^{13}C, HSQC), and MS spectra, we tested SpectraLLM under several input configurations ranging from unimodal to fully fused inputs. As shown in Table 2, a clear performance gradient emerges as more modalities are introduced.

When spectra are used in isolation, prediction performance remains limited across most evaluation metrics. MS alone yields a structure accuracy of 0.0036 and a Tanimoto similarity of 0.1844, while IR performs slightly worse in accuracy (0.0000) and substructure similarity (Tanimoto: 0.1720; MCES: 15.3234). Among NMR modalities, HSQC—encoding heteronuclear correlations—shows the strongest performance (Tanimoto: 0.2058), outperforming both ^{13}C NMR (0.1016) and ^{1}H NMR (0.0720), whose individual accuracies remain at zero. When all three NMR modalities are fused (Jointly NMR), the model achieves a substantially higher accuracy (0.1345) and stronger alignment across structural metrics (e.g., Tanimoto: 0.4151; MCES: 8.3091), confirming the complementary role of scalar shifts and correlation signals in constraining molecular connectivity. These trends persist across continuous spectral modalities as well. On the QM9s dataset, we evaluated IR, Raman, and

4

Table 2: Enhanced structure elucidation through fusion of complementary spectroscopic modalities.

Inputs	trans_rate ↑	Accuracy ↑	Tanimoto ↑	Cosine ↑	MCES ↓	Functional Group ↑	Tanimoto (MACCS) ↑	Fraggle ↑
					QM9S			
IR	99.82%	0.0055	0.1921	0.3120	7.5651	0.6599	0.4330	0.3194
Raman	99.08%	0.0314	0.2500	0.3786	6.4076	0.7317	0.5071	0.2500
UV-Vis	100.00%	0.0000	0.0790	0.1426	10.6374	0.3713	0.2026	0.2100
Jointly	98.72%	**0.1169**	**0.3355**	**0.4560**	**4.9647**	**0.7934**	**0.5785**	**0.4117**
					Multimodal Spectroscopic			
IR	99.63%	0.0000	0.1720	0.2868	15.3234	0.6023	0.4031	0.3906
MS	99.64%	0.0036	0.1844	0.2993	11.3243	0.4929	0.4254	0.4282
IR+MS	99.25%	0.0113	0.2300	0.3519	10.4164	0.6345	0.4887	0.4566
^{13}C NMR	99.64%	0.0000	0.1016	0.1801	14.8865	0.4249	0.2952	0.3607
^{1}H NMR	99.10%	0.0000	0.0720	0.1341	18.6141	0.3329	0.2203	0.2572
HSQC NMR	99.64%	0.0108	0.2058	0.3221	13.4919	0.5495	0.4392	0.4274
Jointly NMR	98.92%	0.1345	0.4151	0.5322	8.3091	0.7209	0.6367	0.5862
Jointly NMR+IR	99.60%	0.1235	0.4121	0.5341	8.3855	0.7764	0.6575	0.5809
Jointly NMR+MS	98.37%	0.1875	0.4518	0.5601	8.1682	0.7618	0.6760	0.6063
Jointly NMR+IR+MS	**99.79%**	**0.1983**	**0.4875**	**0.5973**	**8.1151**	**0.8103**	**0.7099**	**0.6222**

UV-Vis spectra in isolation and in combination. Raman spectroscopy yields the strongest individual performance (Tanimoto: 0.2500; Cosine: 0.3786), followed by IR (Tanimoto: 0.1921). UV-Vis spectroscopy alone proves least informative, with the lowest top-1 accuracy and highest MCES, likely due to the delocalized nature of electronic transitions. However, when all three modalities are jointly input, performance improves consistently across all metrics (e.g., Tanimoto: 0.3355; MCES: 4.9647), underscoring the additive value of fusing orthogonal spectral cues.

Crucially, modality fusion across the Multimodal Spectroscopic dataset yields further and more substantial improvements. The combination of MS and NMR boosts accuracy to 0.1875 and improves substructure recovery (MCES: 8.1682), highlighting their synergistic strengths—mass-based elemental composition from MS and local bonding environments from NMR. The addition of IR spectra (Jointly NMR+IR+MS) leads to the strongest overall performance: top-1 accuracy reaches 0.1983, and all other metrics—including functional group recovery (0.8103), Tanimoto (MACCS: 0.7099), and Fraggle similarity (0.6222)—reach their peak values. Notably, MCES is minimized (8.1151), indicating enhanced substructure fidelity. These findings suggest that although IR spectra offer limited discriminative power in isolation, they encode orthogonal vibrational information—particularly about bond types and functional groups—that meaningfully complements MS and NMR signals when integrated.

Taken together, these results demonstrate that diverse spectroscopic modalities contribute distinct and complementary constraints on molecular structure. SpectraLLM not only leverages these heterogeneous signals effectively, but also exhibits smooth and consistent performance gains as input diversity increases—evidence of robust multimodal reasoning grounded in chemical principles.

To gain mechanistic insight into the synergistic effects of multi-modal spectroscopic integration, we conducted a qualitative case analysis focusing on representative molecular examples successfully predicted only when specific combinations of modalities were present (Fig. 2-3). In the first panel (Fig. 2), we observed multiple instances where Raman spectroscopy critically compensated for ambiguities in IR or UV-Vis spectra alone. For example, the compound COCc1cc[nH]c1 could not be correctly reconstructed using IR or UV-Vis spectra individually, both yielding incorrect aliphatic or polycyclic skeletons. However, the inclusion of Raman spectra restored the correct heteroaromatic scaffold, suggesting that Raman's sensitivity to polarizability changes may have contributed distinct information about C-N and C=C stretches within the imidazole ring system. Similarly, the accurate prediction of C#CC(=O)C#CC was enabled only when Raman data was present, either alone or in combination, indicating its superior capacity to resolve symmetric triple bonds and cumulative ketone functionalities that are poorly distinguished by UV-Vis transitions. Conversely, the second panel (Fig. 3) illustrates cases where IR spectra were indispensable in resolving structural ambiguities left unaddressed by Raman or UV-Vis. For instance, the molecule CC(C)(C)C(=O)C was mispredicted using Raman and UV-Vis alone, both failing to capture the precise branching and carbonyl placement. Only upon incorporating IR—which directly encodes carbonyl stretching vibrations and tertiary alkyl deformations—did the model converge to the correct isomer. This trend held consistently across multiple cases, where IR served as a key modality to disambiguate carbonyl-containing alicyclic structures and hydroxyl-substituted alkynes. Together, these examples reveal that Raman and IR offer distinct yet complementary views of molecular structure, with Raman excelling at resolving symmetric and π-rich environments, and IR dominating in carbonyl and X-H bond detection.

These findings reinforce the value of multi-modal fusion in overcoming the intrinsic limitations of any single spectroscopic modality.
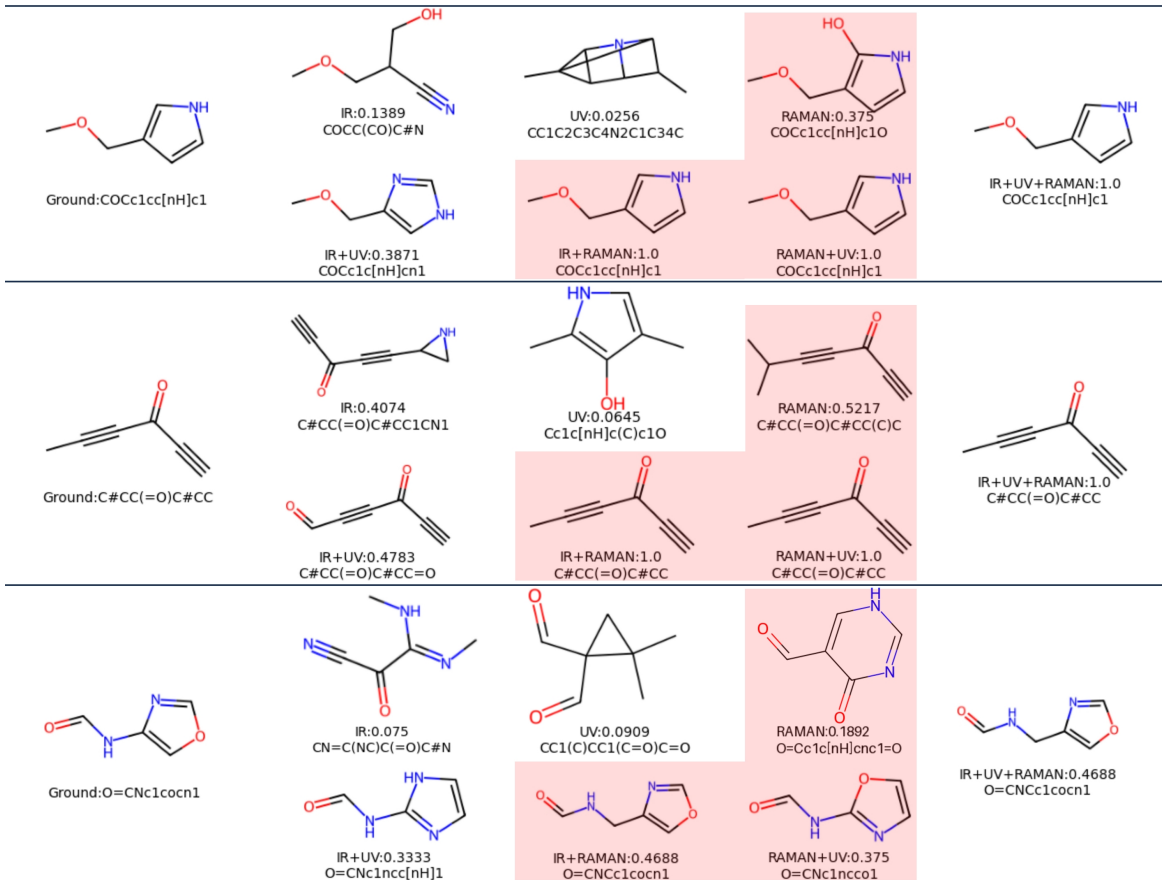

Fig. 2: **Effect of Raman spectra on structural prediction accuracy.** Three representative examples where incorporating Raman spectra corrects wrong predictions made using only IR or UV-Vis inputs. This highlights Raman's complementary role in resolving molecular substructures sensitive to polarizability.

2.3 State-of-the-art performance of SpectraLLM

While our primary focus centers on the synergistic value of multimodal spectral integration, we additionally assessed model performance under individual spectrum types in direct comparison with existing methods. Across multiple datasets and spectroscopic modalities—including infrared (IR), Raman, UV-Vis, nuclear magnetic resonance (NMR), and mass spectrometry (MS)—SpectraLLM achieves consistently superior results over established baselines (Table 3 and Table 4).

On the QM9S dataset, SpectraLLM surpasses traditional spectrum-to-structure pipelines under all unimodal conditions. For instance, using only IR spectra, SpectraLLM achieves a top-1 accuracy of 0.0055 and a Tanimoto similarity of 0.1921—more than double the score of prior neural baselines. With Raman spectra, which provide complementary vibrational cues, the model's advantage becomes even more pronounced: it attains a Tanimoto similarity of 0.2500 and a cosine similarity of 0.3786, outperforming the best previous result by over 30%. Although UV-Vis spectra are intrinsically less structurally specific, SpectraLLM still extracts modest predictive signal (Tanimoto: 0.0790), suggesting robustness to sparse or noisy modalities.

The performance gain is even more striking for NMR data. On the Multimodal Spectroscopic dataset, SpectraLLM achieves a top-1 accuracy of 0.1345 and a Tanimoto similarity of 0.4151—nearly an order of magnitude better than NMR2Struct, which lacks the capacity to model chemical reasoning from descriptive inputs. Furthermore, SpectraLLM exhibits substantially improved recovery of
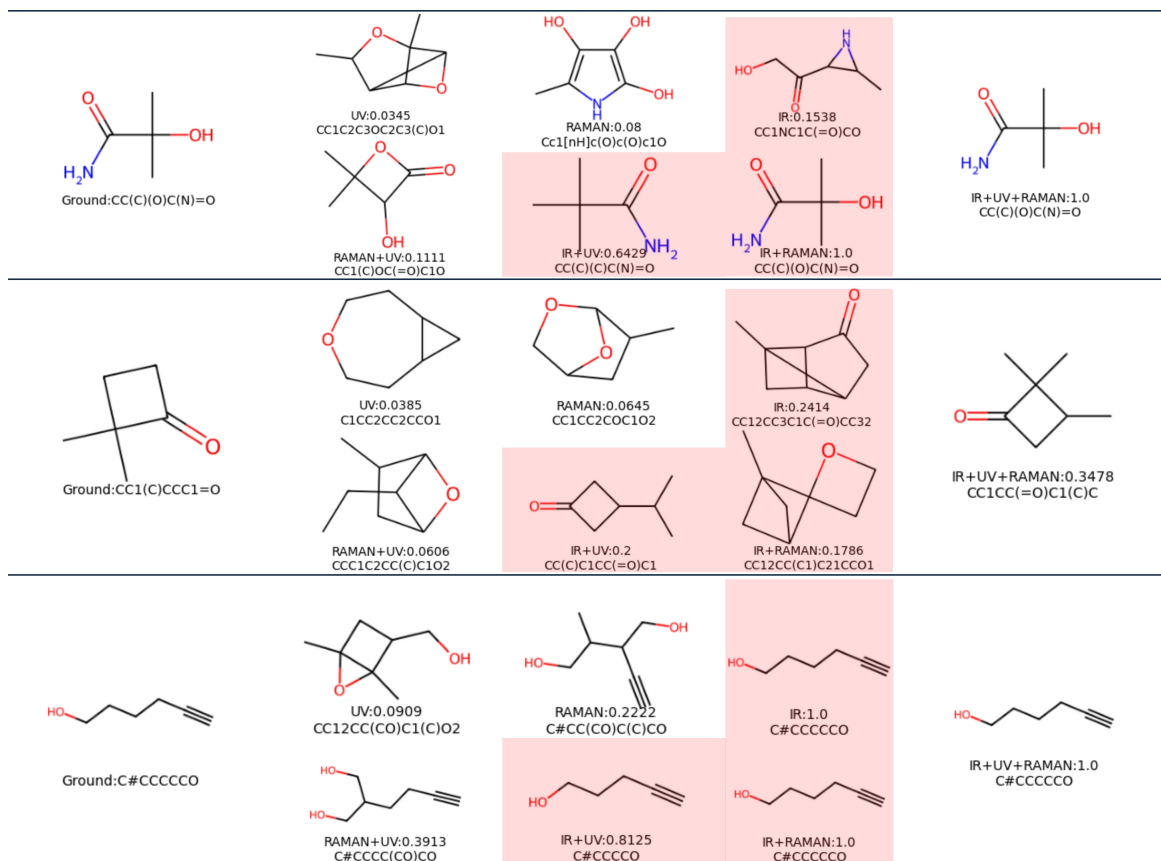
Fig. 3: **Importance of IR spectra for identifying functional groups.** Three representative examples where IR spectra are essential to correctly identify carbonyl groups and distinguish branched chain configurations. Without IR input, predictions based on Raman and UV-Vis remain ambiguous or incorrect.

functional groups (0.7209 vs. 0.1718), highlighting its ability to learn interpretable mappings between spectral patterns and chemical substructures.

Mass spectrometry-based inference presents a particularly challenging case due to its fragmented and indirect nature. Nonetheless, across three benchmark datasets—MassSpecGym, Multimodal Spectroscopic, and MassBank—SpectraLLM either matches or exceeds state-of-the-art results (Table 4). For example, on MassSpecGym, it achieves the highest cosine similarity (0.2558) and functional group recovery (0.5003), while maintaining near-perfect output validity (99.74%). Notably, unlike many previous models which are restricted to canonical collision-energy settings or rely on MS/MS-specific architectures, SpectraLLM generalizes well even when experimental details are incomplete or variable, due in part to its language-based encoding of such context.

Together, these results demonstrate the breadth and flexibility of SpectraLLM across spectrum types, chemical domains, and benchmark datasets. Even under information-limited unimodal conditions, the model extracts meaningful structure-level signals and achieves performance previously unattainable with conventional deep learning models. This robustness forms the basis for its even stronger performance in multimodal settings.

3 Discussion

In this study, we present SpectraLLM, a multimodal large language model framework that enables automated molecular structure elucidation through joint reasoning over heterogeneous spectroscopic data. Unlike conventional spectrum-to-structure approaches that are typically confined to isolated modalities, SpectraLLM directly integrates continuous spectral profiles—including IR, Raman, UV-Vis, and NMR—in their language forms, alongside discrete mass spectrometry data. This unified framework allows the model to simultaneously exploit distinct spectral signatures reflecting diverse structural and functional aspects of the molecule.

Table 3: Comparative evaluation of SpectralLLM and conventional approaches under individual spectral inputs.

Spectrum	Method	trans_rate ↑	Accuracy ↑	Tanimoto ↑	Cosine ↑	MCES ↓	Functional Group ↑	Tanimoto (MACCS) ↑	Fraggle ↑
				QM9S					
	IR-to-Structure	100.00%	0	0.0718	0.1311	11.3187	0.3151	0.1585	0.1747
IR	Spectra2Structure	100.00%	0.0019	0.0965	0.1695	10.1081	0.4383	0.2162	0.2308
	SpectralLLM	99.82%	**0.0055**	**0.1921**	**0.3120**	**7.5651**	**0.6599**	**0.4330**	**0.3194**
	IR-to-Structure	100.00%	0	0.0766	0.1395	11.3516	0.3525	0.1639	0.1959
Raman	Spectra2Structure	100.00%	0	0.1089	0.1901	9.4164	0.4419	0.2388	0.2504
	SpectralLLM	99.08%	**0.0314**	**0.2500**	**0.3786**	**6.4076**	**0.7317**	**0.5071**	**0.2500**
	IR-to-Structure	100.00%	0	0.0728	0.1326	11.424	0.3151	0.1512	0.1837
UV-Vis	Spectra2Structure	100.00%	0	0.0716	0.1313	11.1222	0.3901	0.1418	0.2092
	SpectralLLM	100.00%	0.0000	**0.0790**	**0.1426**	**10.6374**	**0.3713**	**0.2026**	**0.2100**
				Multimodal Spectroscopic					
	NMR2Struct	47.62%	0	0.0433	0.1029	30.6938	0.1718	0.1294	0.0962
NMR	SpectralLLM	98.92%	**0.1345**	**0.4151**	**0.5322**	**8.3091**	**0.7209**	**0.6367**	**0.5862**

Table 4: Benchmarking SpectralLLM against established mass spectrometry-based inference models.

Method	Validity ↑	Tanimoto ↑	Cosine ↑	Functional Group ↑	Tanimoto (MACCS) ↑	Fraggle ↑
		MassSpecGym				
Spec2Mol	62.86%	0.0849	0.1511	0.3111	0.2709	0.2065
Diffms	57.16%	**0.1597**	0.2422	0.4890	0.4305	0.3539
SpectralLLM	99.74%	0.1533	**0.2558**	**0.5003**	**0.4723**	**0.3610**
		Multimodal Spectroscopic				
Spec2Mol	75.39%	0.0988	0.1739	0.3042	0.2440	0.2587
Diffms	78.77%	0.1535	0.2351	0.4248	0.3730	0.3635
SpectralLLM	99.64%	**0.1844**	**0.2993**	**0.4929**	**0.4254**	**0.4282**
		MassBank				
Spec2Mol	71.63%	0.0857	0.0006	0.2999	0.1539	0.1102
Diffms	23.63%	0.0742	0.2088	0.1795	0.1007	0.0238
SpectralLLM	98.44%	**0.1286**	**0.2229**	**0.4539**	**0.3787**	**0.3150**

Our systematic evaluation across four benchmark datasets reveals several key insights. First, when restricted to individual spectral modalities, predictive performance remains limited, underscoring the inherent ambiguity and partial information encoded in any single type of spectrum. For example, mass spectra provide compositional and fragmentation information, while vibrational spectra capture functional group environments; neither alone suffices to fully resolve molecular architecture. However, as we incrementally increase the diversity of accessible spectral inputs, we observe consistent and substantial improvements in structural prediction accuracy. This performance gain reflects the complementary nature of the information embedded across modalities, whereby integrating multiple spectra allows the model to resolve structural ambiguities that would otherwise persist under single-modality analysis. Such a trend not only validates our central hypothesis, but mirrors the established expert practice of multimodal spectroscopic reasoning in chemical structure elucidation.

Beyond achieving state-of-the-art performance across varied spectrum-to-structure prediction scenarios, SpectraLLM demonstrates a scalable paradigm for leveraging foundation models in scientific discovery. The model's modular architecture, leveraging a domain-adapted language reasoner, offers flexibility for incorporating additional modalities and expert knowledge sources. Moving forward, the integration of curated textual annotations, reaction context, and domain-specific ontologies may further enhance both accuracy and interpretability. Moreover, the successful application of large-scale foundation models to analytical spectroscopy underscores their broader potential for addressing long-standing challenges in metabolomics, environmental monitoring, and complex mixture analysis — scientific domains where rich, yet heterogeneous, experimental data converge.

References

[1] Shklover, V.Y., Kazanskii, P., Artemov, N., Maryasev, I.: Electron microscopy and electron diffraction studies of morphology and crystal structure of natural silicas. Crystallography Reports **66**(4), 663–672 (2021)

[2] Ali, A., Chiang, Y.W., Santos, R.M.: X-ray diffraction techniques for mineral characterization: A review for engineers of the fundamentals, applications, and research directions. Minerals **12**(2),

205 (2022)

[3] Beermann, T., Brockamp, O.: Structure analysis of montmorillonite crystallites by convergent-beam electron diffraction. Clay Minerals **40**(1), 1–13 (2005)

[4] Als-Nielsen, J., Materlik, G.: Recent applications of x rays in condensed matter physics. Physics today **48**(11), 34–40 (1995)

[5] Ishchenko, A., Bagratashvili, V., Avilov, A.: Methods for studying the coherent 4d structural dynamics of free molecules and condensed state of matter. Crystallography Reports **56**, 751–773 (2011)

[6] Filipponi, A., Di Cicco, A., Natoli, C.R.: X-ray-absorption spectroscopy and n-body distribution functions in condensed matter. i. theory. Physical Review B **52**(21), 15122 (1995)

[7] Krishnan, V., Rupp, B.: Macromolecular structure determination: comparison of x-ray crystallography and nmr spectroscopy. eLS **10**, 0002716 (2012)

[8] Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.*: Crystallography & nmr system: A new software suite for macromolecular structure determination. Biological Crystallography **54**(5), 905–921 (1998)

[9] Hemath, M., Mavinkere Rangappa, S., Kushvaha, V., Dhakal, H.N., Siengchin, S.: A comprehensive review on mechanical, electromagnetic radiation shielding, and thermal conductivity of fibers/inorganic fillers reinforced hybrid polymer composites. Polymer Composites **41**(10), 3940–3965 (2020)

[10] Gu, J., Duan, F., Liu, S., Cha, W., Lu, J.: Phase engineering of nanostructural metallic materials: Classification, structures, and applications. Chemical Reviews **124**(3), 1247–1287 (2024)

[11] Stuart, B.H.: Infrared Spectroscopy: Fundamentals and Applications. John Wiley & Sons, ??? (2004)

[12] Rantanen, J., Khinast, J.: The future of pharmaceutical manufacturing sciences. Journal of pharmaceutical sciences **104**(11), 3612–3638 (2015)

[13] Adrian, M., Dubochet, J., Lepault, J., McDowall, A.W.: Cryo-electron microscopy of viruses. Nature **308**(5954), 32–36 (1984)

[14] Hall, L.: Nuclear magnetic resonance. Advances in carbohydrate chemistry **19**, 51–93 (1964)

[15] Ng, L.M., Simmons, R.: Infrared spectroscopy. Analytical chemistry **71**(12), 343–350 (1999)

[16] De Hoffmann, E., Stroobant, V.: Mass Spectrometry: Principles and Applications. John Wiley & Sons, ??? (2007)

[17] Okada, K., Kotani, A.: Interatomic and intra-atomic configuration interactions in core-level x-ray photoemission spectra of late transition-metal compounds. Journal of the Physical Society of Japan **61**(12), 4619–4637 (1992)

[18] Coates, J., *et al.*: Interpretation of infrared spectra, a practical approach. Encyclopedia of analytical chemistry **12**, 10815–10837 (2000)

[19] Linstrom, P.J., Mallard, W.G.: The nist chemistry webbook: A chemical data resource on the internet. Journal of Chemical & Engineering Data **46**(5), 1059–1063 (2001)

[20] Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., *et al.*: Massbank: a public repository for sharing mass spectral data for life sciences. Journal of mass spectrometry **45**(7), 703–714 (2010)

[21] Punjabi, D., Huang, Y.-C., Holzhauer, L., Tremouilhac, P., Friederich, P., Jung, N., Bräse, S.: Infrared spectrum analysis of organic molecules with neural networks using standard reference data sets in combination with real-world data. Journal of Cheminformatics **17**(1), 1–13 (2025)

[22] Aron, A.T., Gentry, E.C., McPhail, K.L., Nothias, L.-F., Nothias-Esposito, M., Bouslimani, A., Petras, D., Gauglitz, J.M., Sikora, N., Vargas, F., *et al.*: Reproducible molecular networking of untargeted mass spectrometry data using gnps. Nature protocols **15**(6), 1954–1991 (2020)

[23] Shiferaw, G.A., Vandermarliere, E., Hulstaert, N., Gabriels, R., Martens, L., Volders, P.-J.: Coss: A fast and user-friendly tool for spectral library searching. Journal of proteome research **19**(7), 2786–2793 (2020)

[24] Platte, F., Heise, H.M.: Substance identification based on transmission thz spectra using library search. Journal of Molecular Structure **1073**, 3–9 (2014)

[25] Smith, B.C.: Infrared Spectral Interpretation: a Systematic Approach. CRC press, ??? (2018)

[26] Butler, H.J., Ashton, L., Bird, B., Cinque, G., Curtis, K., Dorney, J., Esmonde-White, K., Fullwood, N.J., Gardner, B., Martin-Hirsch, P.L., *et al.*: Using raman spectroscopy to characterize biological materials. Nature protocols **11**(4), 664–687 (2016)

[27] Perkampus, H.-H.: UV-VIS Spectroscopy and Its Applications. Springer, ??? (2013)

[28] Picollo, M., Aceto, M., Vitorino, T.: Uv-vis spectroscopy. Physical sciences reviews **4**(4), 20180008 (2019)

[29] Bovey, F.A., Mirau, P.A., Gutowsky, H.: Nuclear Magnetic Resonance Spectroscopy. Elsevier, ??? (1988)

[30] James, T.: Nuclear Magnetic Resonance in Biochemistry. Elsevier, ??? (2012)

[31] Quinn, R.A., Melnik, A.V., Vrbanac, A., Fu, T., Patras, K.A., Christy, M.P., Bodai, Z., Belda-Ferre, P., Tripathi, A., Chung, L.K., *et al.*: Global chemical effects of the microbiome include new bile-acid conjugations. Nature **579**(7797), 123–129 (2020)

[32] Wilson, E.B., Decius, J.C., Cross, P.C.: Molecular Vibrations: the Theory of Infrared and Raman Vibrational Spectra. Courier Corporation, ??? (1980)

[33] Zhang, H., Li, L., Quan, S., Tian, W., Zhang, K., Nie, L., Zang, H.: Novel similarity methods evaluation and feasible application for pharmaceutical raw material identification with near-infrared spectroscopy. ACS omega **5**(46), 29864–29871 (2020)

[34] Kim, S., Lee, D., Liu, X., Van Neste, C., Jeon, S., Thundat, T.: Molecular recognition using receptor-free nanomechanical infrared spectroscopy based on a quantum cascade laser. Scientific reports **3**(1), 1111 (2013)

[35] Skinnider, M.A., Wang, F., Pasin, D., Greiner, R., Foster, L.J., Dalsgaard, P.W., Wishart, D.S.: A deep generative model enables automated structure elucidation of novel psychoactive substances. Nature Machine Intelligence **3**(11), 973–984 (2021)

[36] Alberts, M., Laino, T., Vaucher, A.C.: Leveraging infrared spectroscopy for automated structure elucidation. Communications Chemistry **7**(1), 268 (2024)

[37] Binev, Y., Marques, M.M., Aires-de-Sousa, J.: Prediction of 1h nmr coupling constants with associative neural networks trained for chemical shifts. Journal of chemical information and modeling **47**(6), 2089–2097 (2007)

[38] Mueller, L.: Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence. Journal of the American Chemical Society **101**(16), 4481–4484 (1979)

[39] Hu, F., Chen, M.S., Rotskoff, G.M., Kanan, M.W., Markland, T.E.: Accurate and efficient structure elucidation from routine one-dimensional nmr spectra using multitask machine learning. ACS Central Science **10**(11), 2162–2170 (2024)

[40] Litsa, E.E., Chenthamarakshan, V., Das, P., Kavraki, L.E.: An end-to-end deep learning framework for translating mass spectra to de-novo molecules. Communications Chemistry **6**(1), 132 (2023)

[41] Bittremieux, W., Wang, M., Dorrestein, P.C.: The critical role that spectral libraries play in capturing the metabolomics community knowledge. Metabolomics **18**(12), 94 (2022)

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

[43] Nguyen, D.H., Nguyen, C.H., Mamitsuka, H.: Recent advances and prospects of computational methods for metabolite identification: a review with emphasis on machine learning approaches. Briefings in bioinformatics **20**(6), 2028–2043 (2019)

[44] Wang, Z., Feng, X., Liu, J., Lu, M., Li, M.: Functional groups prediction from infrared spectra based on computer-assist approaches. Microchemical Journal **159**, 105395 (2020)

[45] Yang, J., Xu, P., Wu, S., Chen, Z., Fang, S., Xiao, H., Hu, F., Jiang, L., Wang, L., Mo, B., *et al.*: Raman spectroscopy for esophageal tumor diagnosis and delineation using machine learning and the portable raman spectrometer. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy **317**, 124461 (2024)

[46] Nalla, R., Pinge, R., Narwaria, M., Chaudhury, B.: Priority based functional group identification of organic molecules using machine learning. In: Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, pp. 201–209 (2018)

[47] Ye, S., Zhong, K., Zhang, J., Hu, W., Hirst, J.D., Zhang, G., Mukamel, S., Jiang, J.: A machine learning protocol for predicting protein infrared spectra. Journal of the American Chemical Society **142**(45), 19071–19077 (2020)

[48] Zou, Z., Zhang, Y., Liang, L., Wei, M., Leng, J., Jiang, J., Luo, Y., Hu, W.: A deep learning model for predicting selected organic molecular spectra. Nature Computational Science **3**(11), 957–964 (2023)

[49] Al, S.A., Allouche, A.-R.: Neural network approach for predicting infrared spectra from 3d molecular structure. Chemical Physics Letters **856**, 141603 (2024)

[50] Fine, J.A., Rajasekar, A.A., Jethava, K.P., Chopra, G.: Spectral deep learning for prediction and prospective validation of functional groups. Chemical science **11**(18), 4618–4630 (2020)

[51] Judge, K., Brown, C.W., Hamel, L.: Sensitivity of infrared spectra to chemical functional groups. Analytical chemistry **80**(11), 4186–4192 (2008)

[52] Klawun, C., Wilkins, C.L.: Optimization of functional group prediction from infrared spectra using neural networks. Journal of chemical information and computer sciences **36**(1), 69–81 (1996)

[53] Fessenden, R.J., Györgyi, L.: Identifying functional groups in ir spectra using an artificial neural network. Journal of the Chemical Society, Perkin Transactions 2 (11), 1755–1762 (1991)

[54] Hemmer, M.C., Gasteiger, J.: Prediction of three-dimensional molecular structures using information from infrared spectra. Analytica chimica acta **420**(2), 145–154 (2000)

[55] Wang, X., Jiang, S., Hu, W., Ye, S., Wang, T., Wu, F., Yang, L., Li, X., Zhang, G., Chen, X., *et al.*: Quantitatively determining surface–adsorbate properties from vibrational spectroscopy with

interpretable machine learning. Journal of the American Chemical Society **144**(35), 16069–16076 (2022)

[56] Chen, P.-Y., Shibata, K., Hagita, K., Miyata, T., Mizoguchi, T.: Prediction of the ground-state electronic structure from core-loss spectra of organic molecules by machine learning. The Journal of Physical Chemistry Letters **14**(20), 4858–4865 (2023)

[57] Li, C., Cong, Y., Deng, W.: Identifying molecular functional groups of organic compounds by deep learning of nmr data. Magnetic Resonance in Chemistry **60**(11), 1061–1069 (2022)

[58] Specht, T., Arweiler, J., Stüber, J., Münnemann, K., Hasse, H., Jirasek, F.: Automated nuclear magnetic resonance fingerprinting of mixtures. Magnetic Resonance in Chemistry **62**(4), 286–297 (2024)

[59] Sridharan, B., Mehta, S., Pathak, Y., Priyakumar, U.D.: Deep reinforcement learning for molecular inverse problem of nuclear magnetic resonance spectra to molecular structure. The Journal of Physical Chemistry Letters **13**(22), 4924–4933 (2022)

[60] Huang, Z., Chen, M.S., Woroch, C.P., Markland, T.E., Kanan, M.W.: A framework for automated structure elucidation from routine nmr spectra. Chemical Science **12**(46), 15329–15338 (2021)

[61] Devata, S., Sridharan, B., Mehta, S., Pathak, Y., Laghuvarapu, S., Varma, G., Priyakumar, U.D.: Deepspinn–deep reinforcement learning for molecular structure prediction from infrared and 13 c nmr spectra. Digital Discovery **3**(4), 818–829 (2024)

[62] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (2002)

[63] Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J.: A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems **33**(12), 6999–7019 (2021)

[64] O'shea, K., Nash, R.: An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458 (2015)

[65] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

[66] Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., *et al.*: A survey on vision transformer. IEEE transactions on pattern analysis and machine intelligence **45**(1), 87–110 (2022)

[67] Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., Wang, Y.: Transformer in transformer. Advances in neural information processing systems **34**, 15908–15919 (2021)

[68] Sapegin, D.A., Bear, J.C.: Structure seer–a machine learning model for chemical structure elucidation from node labelling of a molecular graph. Digital Discovery **3**(1), 186–200 (2024)

[69] Rippel, O., Snoek, J., Adams, R.P.: Spectral representations for convolutional neural networks. Advances in neural information processing systems **28** (2015)

[70] Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. Advances in neural information processing systems **29** (2016)

[71] Kuhn, S., Tumer, E., Colreavy-Donnelly, S., Moreira Borges, R.: A pilot study for fragment identification using 2d nmr and deep learning. Magnetic Resonance in Chemistry **60**(11), 1052–1060 (2022)

[72] Zhao, Z., Liu, Z., Ji, M., Zhao, X., Zhu, Q., Huang, M.: Conincedeep: A novel deep learning

method for component identification of mixture based on raman spectroscopy. Chemometrics and Intelligent Laboratory Systems **234**, 104757 (2023)

[73] Yao, L., Yang, M., Song, J., Yang, Z., Sun, H., Shi, H., Liu, X., Ji, X., Deng, Y., Wang, X.: Conditional molecular generation net enables automated structure elucidation based on 13c nmr spectra and prior knowledge. Analytical chemistry **95**(12), 5393–5401 (2023)

[74] Alberts, M., Zipoli, F., Vaucher, A.C.: Learning the language of nmr: Structure elucidation from nmr spectra using transformer models (2023)

[75] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)

[76] Cho, K., Van Merriënboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259 (2014)

[77] Liu, B., Ramsundar, B., Kawthekar, P., Shi, J., Gomes, J., Luu Nguyen, Q., Ho, S., Sloane, J., Wender, P., Pande, V.: Retrosynthetic reaction prediction using neural sequence-to-sequence models. ACS central science **3**(10), 1103–1113 (2017)

[78] Tang, Y.-J., Pang, Y.-H., Liu, B.: Idp-seq2seq: identification of intrinsically disordered regions based on sequence to sequence learning. Bioinformatics **36**(21), 5177–5186 (2020)

[79] Berman, D.S., Howser, C., Mehoke, T., Ernlund, A.W., Evans, J.D.: Mutagan: A sequence-to-sequence gan framework to predict mutations of evolving protein populations. Virus Evolution **9**(1), 022 (2023)

[80] Zhang, R., Lin, Y., Wu, Y., Deng, L., Zhang, H., Liao, M., Peng, Y.: Mvmrl: a multi-view molecular representation learning method for molecular property prediction. Briefings in Bioinformatics **25**(4), 298 (2024)

[81] Xu, Z., Wang, S., Zhu, F., Huang, J.: Seq2seq fingerprint: An unsupervised deep molecular embedding for drug discovery. In: Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, pp. 285–294 (2017)

[82] He, J., You, H., Sandström, E., Nittinger, E., Bjerrum, E.J., Tyrchan, C., Czechtizky, W., Engkvist, O.: Molecular optimization by capturing chemist's intuition using deep neural networks. Journal of cheminformatics **13**, 1–17 (2021)

[83] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)

[84] Hu, W., Hu, T., Li, B., Zou, Z., Zhu, T., Zhang, Y., Jiang, J., Luo, Y.: A machine learning protocol to directly translate molecular infrared or raman spectra into molecular formula (2023)

[85] Kanakala, G.C., Sridharan, B., Priyakumar, U.D.: Spectra to structure: contrastive learning framework for library ranking and generating molecular structures for infrared spectra. Digital Discovery **3**(12), 2417–2423 (2024)

[86] French, E., Deng, X., Chen, S., Ju, C.-W., Cheng, X., Zhang, L., Liu, X., Guan, H., Lin, Z.: Revolutionizing spectroscopic analysis using sequence-to-sequence models i: From infrared spectra to molecular structures (2025)

[87] Bohde, M., Manjrekar, M., Wang, R., Ji, S., Coley, C.W.: Diffms: Diffusion generation of molecules conditioned on mass spectra. arXiv preprint arXiv:2502.09571 (2025)

[88] Vignac, C., Krawczuk, I., Siraudin, A., Wang, B., Cevher, V., Frossard, P.: Digress: Discrete denoising diffusion for graph generation. arXiv preprint arXiv:2209.14734 (2022)

[89] Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., et al.: Qwen3 technical report. arXiv preprint arXiv:2505.09388 (2025)

[90] Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., *et al.*: Lora: Low-rank adaptation of large language models. ICLR **1**(2), 3 (2022)

[91] Alberts, M., Schilter, O., Zipoli, F., Hartrampf, N., Laino, T.: Unraveling molecular structure: A multimodal spectroscopic dataset for chemistry. Advances in Neural Information Processing Systems **37**, 125780–125808 (2024)

[92] Lowe, D.M.: Extraction of chemical structures and reactions from the literature. PhD thesis (2012)

[93] Bushuiev, R., Bushuiev, A., Jonge, N., Young, A., Kretschmer, F., Samusevich, R., Heirman, J., Wang, F., Zhang, L., Dührkop, K., *et al.*: Massspecgym: A benchmark for the discovery and identification of molecules. Advances in Neural Information Processing Systems **37**, 110010–110027 (2024)

[94] Landrum, G., et al.: Rdkit: Open-source cheminformatics software (2016)

Table **5**: Overview of four spectroscopic datasets and their modality coverage.

Dataset	IR	Raman	UV	NMR	Mass	Molecule	Spectral
QM9S	√	√	√	-	-	129,817	389,451
Multimodal Spectroscopic dataset	√	-	-	√ ^{1}H,^{13}C,HSQC	√ Positive, Negative	794,403	4,766,418
MassSpecGym	-	-	-	-	√	231,104	231,104
MassBank	-	-	-	-	√	122,746	122,746
ALL	√	√	√	√	√	943,732	5,510,655

Table **6**: Data distribution across spectral modalities and splits.

	Spectrum	Train	Val	Test	ALL
Single	Mass	301,133	15,849	14,247	331,229
	MS/MS Positive	643,605	71,511	79,287	794,403
	^{13}C NMR	643,604	71,511	79,287	794,402
	^{1}H NMR	643,586	71,509	79,285	794,380
	HSQC NMR	643,605	71,511	79,286	794,402
	IR	743,352	82,594	87,688	913,634
	Raman	105,080	11,675	13,062	129,817
	UV-Vis	105,078	11,675	13,062	129,815
Multi	IR+Raman+UV-Vis	104,948	11,660	12,416	129,024
	^{13}C+^{1}H+HSQC NMR	494,173	54,908	78,882	627,963
	IR+Raman+UV-Vis+NMR+Mass	383,870	42,652	30,937	457,459
ALL		4,812,034	517,055	567,439	5,896,528

4 Methods

4.1 Large-scale datasets of spectra for pre-training

Comprehensive and high-quality datasets are essential for enabling effective generalization across diverse domains. In the field of chemistry, abundant paired data linking molecular structures with multiple spectroscopic modalities are readily available, owing to decades of research and the routine use of spectroscopy in molecular analysis. To equip SpectraLLM with broad and multimodal spectroscopic knowledge, we pre-trained the model using diverse data compiled from multiple publicly available datasets, collectively covering five spectroscopic modalities: infrared (IR), Raman, ultraviolet-visible (UV-Vis), nuclear magnetic resonance (NMR), and mass spectrometry (MS).

For vibrational and electronic spectroscopy, we used the QM9Spectra (QM9s) dataset [48], which contains simulated IR, Raman, and UV-Vis spectra for approximately 134,000 small organic molecules generated using frequency analysis and time-dependent density functional theory (TD-DFT) at the B3LYP/def2-TZVP level. Additional vibrational and NMR data were drawn from the Multimodal Spectroscopic dataset [91], comprising simulated ^{1}H NMR, ^{13}C NMR, HSQC NMR, IR, and MS spectra (in both positive and negative ion modes) for over 790,000 molecules extracted from the USPTO chemical reaction dataset [92]. This dataset enables learning from complex, multi-spectrum representations and mirrors the integrative strategy employed by expert chemists.

For mass spectrometry, in addition to the existing data in Multimodal Spectroscopic dataset, we utilized the MassSpecGym dataset [93], which provides 231,000 curated high-resolution MS spectra across 29,000 unique compounds, representing the largest standardized public collection to date. Since MassSpecGym dataset originated in October 2024, we also incorporated MassBank [20], a comprehensive open-access repository of MS and tandem MS (MS/MS) spectra, to supplement our pretraining with recent added experimentally acquired data.

Table 5 presents the basic information of all the datasets we collected.

4.2 Data processing and representation

Our data preprocessing pipeline consists of four key steps: SMILES standardization and molecular alignment, dataset splitting, spectral feature extraction, and prompt formulation for language modeling . First, all SMILES strings from the constituent datasets were standardized, deduplicated, and aligned to ensure a unified molecular indexing scheme. This yielded a final corpus of 943,730 unique molecules (Table 5), each associated with one to five types of spectra, and containing up to 5 total spectral instances per molecule. The dataset was then split into training, validation, and test sets in

an 8:1:1 ratio based on molecular identity, while ensuring adequate coverage of each spectral modality across all subsets (Table 6). All reported results were obtained from the held-out test set. Next, we applied modality-specific preprocessing procedures to facilitate language model training.

Vibrational and electronic Spectra. For continuous spectra, we retained effective signals through modality-specific truncation and normalization procedures. In the IR modality, wavenumbers were restricted to 500–4000 cm^{-1} for the QM9s dataset and 400–4000 cm^{-1} for the multimodal spectroscopic dataset, uniformly resampled at 1 cm^{-1} intervals (3501 and 3601 points, respectively). Raman spectra adopted the same 500–4000 cm^{-1} range and sampling. UV-Vis spectra were truncated to the 1.0–15.0 eV range, with 0.02 eV sampling intervals (700 points). All spectral intensities were normalized by their maximum value to ensure comparability across molecules and suppress variation introduced by experimental conditions. To further support model interpretability and reduce input complexity, we extracted discrete peak-based representations from the preprocessed spectra. Each spectrum was encoded in structured JSON format, recording only the coordinates and relative intensities of characteristic peaks, such as: `"Wavenumbers"`: [`"wavenumber`$_1$`"`, `"wavenumber`$_2$`"`], `"Intensities"`: [`"intensity`$_1$`"`, `"intensity`$_2$`"`].

Nuclear magnetic resonance spectrum. For nuclear magnetic resonance (NMR) spectra, we applied modality-specific preprocessing and discrete feature extraction. For ^{13}C NMR and ^{1}H NMR, chemical shift ranges were set to 220–0 ppm and 12–0 ppm, respectively. A dynamic threshold—set at 1% of the spectrum's maximum intensity—was used to filter background noise, and the spectra were cropped to the minimal region containing significant peaks. The intensity values have all been normalization. For HSQC NMR, which contain 2D correlation peaks between hydrogen and carbon atoms, only peak-level information was retained. Each peak was characterized by its ^{1}H and ^{13}C chemical shifts and associated proton count (nH), defaulted to 1 if unspecified. To support downstream prompt generation, we represented NMR signals in structured formats. For ^{13}C and ^{1}H spectra, data were stored as key–value pairs of shift and intensity values, e.g., `"C-shifts"`: [`"shift`$_1$`"`, `"shift`$_2$`"`], `"Intensities"`: [`"intensity`$_1$`"`, `"intensity`$_2$`"`]. For HSQC, the 2D correlations were encoded as: `"H-shifts"`: [`"h`$_1$`"`, `"h`$_2$`"`], `"C-shifts"`: [`"c`$_1$`"`, `"c`$_2$`"`], `"Intensities"`: [`"intensity`$_1$`"`, `"intensity`$_2$`"`].

Mass Spectrometry. For mass spectrometry (MS), we retained spectra containing between 2 and 1280 peaks. All mass-to-charge ratio (m/z) values were rounded to two decimal places, and peak intensities were converted to relative abundances within a [0, 100] range, also to two decimal places. To preserve experimental context and enable condition-aware modeling, auxiliary metadata—such as instrument type, collision energy, and adduct ion species—was retained when available, distinguishing spectra acquired under different experimental settings for the same molecule. For downstream prompt construction, MS spectra were represented as structured arrays of m/z–intensity pairs in JSON format, for example: `"mzs"`: [`"m/z`$_1$`"`, `"m/z`$_2$`"`], `"intensities"`: [`"intensity`$_1$`"`, `"intensity`$_2$`"`].

Finally, we designed a diverse collection of natural language instructions for the language model, supporting both single-spectrum and multi-spectrum input formats. The detailed prompt examples will be introduced in the next subsection.

4.3 Prompt Design

Characteristic peaks from all continuous spectra were extracted using built-in functions from the SciPy library, based on local maxima and prominence thresholds. The resulting structured arrays—recorded in JSON format as described in the previous section—were subsequently translated into natural language descriptions, enabling flexible integration of spectral information from multiple modalities into a shared textual context.

Given that the prediction target is the SMILES representation of molecular structures, we naturally formulated the structure elucidation task as a sequence-to-sequence generation problem. As illustrated in Table 9 and Table 10, training data were organized in a dialogue-style format, where each example consists of a `"Human"` prompt and a corresponding `"GPT"` response. The `"Human"` component encodes the query, including the spectrum modality, extracted features, and a task directive to infer molecular structure. The `"GPT"` component provides the expected model output, i.e., the target SMILES string.

To promote generalization across varying data configurations, we designed a diverse set of input formats:

1. Single-spectrum setting (Table. 9): molecular structure prediction based solely on one spectrum type, including individual IR, Raman, UV-Vis, ^{1}H NMR, ^{13}C NMR, HSQC NMR, or MS spectra.

Table 7: Effect of Decoding Temperature on Molecular Structure Prediction Performance.

Temperature ↓	Accuracy ↑	Tanimoto ↑	Cosine ↑	MCES ↓	Functional Group ↑	Tanimoto(MACCS) ↑	Fraggle ↑
0.2	0.1919	0.4872	**0.5985**	**6.9840**	0.8084	0.7070	**0.6302**
0.4	**0.1983**	**0.4875**	0.5973	8.1151	**0.8103**	**0.7099**	0.6222
0.6	0.1859	0.4740	0.5846	6.6784	0.8039	0.7022	0.6218
0.8	0.1795	0.4555	0.5660	7.0897	0.7980	0.6819	0.6075
1.0	0.1688	0.4367	0.5518	7.1902	0.7867	0.6752	0.6019
1.2	0.1389	0.4161	0.5352	7.3024	0.7840	0.6592	0.5805

2. Multi-spectrum setting (Table. 10): molecular structure prediction using two or more spectra from the same molecule. This includes:

   - Joint IR, Raman, and UV spectra from the QM9s dataset;
   - Joint IR, NMR, and MS spectra from the Multimodal Spectroscopic dataset;
   - Combined ^{1}H NMR, ^{13}C NMR, and HSQC NMR spectra;

This formulation allows the model to reason across heterogeneous spectral evidence while remaining robust to missing modalities and variations in experimental settings.

4.4 Supervised Fine-Tuning based on LoRa

Based on the prompt templates designed in the previous section, we construct a large-scale multi-spectral question-answering dataset, whose composition is summarized in Table 6. To ensure compatibility with the input length constraints of the base language model, we exclude examples with tokenized prompts exceeding 1024 tokens. Leveraging this dataset, we perform supervised fine-tuning of the foundational Qwen3-32B language model to enable spectrum-informed molecular reasoning. Instead of updating all model parameters during training, we adopt Low-Rank Adaptation (LoRA) [90] to improve efficiency and stability. LoRA freezes the original model weights and injects trainable rank-decomposed matrices into each transformer layer, significantly reducing the number of trainable parameters while retaining performance.

Given a tokenized input sequence $x = \{x_1, x_2, ..., x_n\}$ and the corresponding target output sequence $y = \{y_1, y_2, ..., y_m\}$, the fine-tuning objective is to minimize the token-wise cross-entropy loss:

$$L_{CE} = -\sum_{t=1}^{m} log P_\theta(y_t | x, y_{<t})$$

where $P_\theta(y_t \mid x, y_{<t})$ denotes the probability of generating the t-th token in the output sequence, conditioned on the input and previous target tokens, and θ represents the trainable LoRA parameters injected into the frozen base model.

4.5 Structure Generation

To generate molecular structures from spectroscopic inputs, we use the fine-tuned SpectraLLM model in a greedy decoding setup with fixed temperature and nucleus sampling parameters. Specifically, we adopt a temperature-controlled sampling strategy with a nucleus sampling threshold of $p = 0.7$ (default setting), and vary the temperature τ to investigate its influence on generation diversity and accuracy. Lower temperatures bias the model toward more deterministic predictions, while higher temperatures encourage exploration by flattening the probability distribution. As shown in Table 7, a moderate temperature of $\tau = 0.4$ achieves the best overall performance across multiple metrics, including exact match accuracy, Tanimoto similarity, functional group recovery, and substructure alignment scores (e.g., Fraggle). Extremely low temperatures (e.g., $\tau = 0.2$) slightly improve structure similarity metrics such as cosine and MCES, but at the cost of reduced diversity. Conversely, high temperatures lead to degraded accuracy and increased structural inconsistency.

We do not employ beam search or advanced decoding control mechanisms; instead, decoding is performed greedily (beam size = 1), with syntax validity implicitly learned from training data. Although this setup is relatively simple, it proves effective in generating syntactically valid and chemically plausible SMILES strings under constrained sampling.

4.6 Metrics for structural similarity

Our evaluation goes beyond assessing the exact recovery of molecular structures, and also probes whether the model has internalized the fundamental chemical cues embedded within the spectra—such as generating chemically plausible candidates, reconstructing key substructures, and correctly identifying functional groups. To quantify this, we not only evaluate the accuracy of functional group prediction as a coarse-grained measure of structural fidelity, but also assess molecular similarity at a finer granularity using fingerprint-based Tanimoto metrics and maximum common substructure (MCES) analysis. The subsequent section provides a detailed account of each evaluation criterion.

4.6.1 Exact Match Accuracy

Exact match accuracy quantifies the proportion of model predictions that perfectly recover the reference molecular structures. A prediction is considered correct if its canonical SMILES representation exactly matches that of the ground truth. This metric provides a stringent measure of structure generation fidelity and is formally defined as:

$$\text{Exact Match Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\text{SMILES}_{\text{pred}}^{i} = \text{SMILES}_{\text{true}}^{i})$$

where $\mathbf{1}(\cdot)$ is the indicator function, and N is the total number of evaluation instances.

To further account for the uncertainty and diversity in model outputs, we also report the Top-K accuracy, which measures whether the correct structure appears among the top K ranked candidates generated by the model. This metric relaxes the exact match criterion and better reflects the model's capacity to prioritize chemically valid hypotheses. It is defined as:

$$\text{Top-K Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\text{SMILES}_{\text{true}}^{i} \in \text{SMILES}_{\text{pred}}^{i,1}, \ldots, \text{SMILES}_{\text{pred}}^{i,K}),$$

where $\text{SMILES}_{\text{pred}}^{i,k}$ denotes the k-th ranked prediction for the i-th sample. In our experiments, we report Top-1 accuracies to provide a comprehensive view of generation performance under varying tolerance thresholds.

4.6.2 Functional Group Similarity Metric

To quantify the similarity between two molecules in terms of their functional group composition, we define a functional group similarity score based on SMARTS pattern matching. Using a curated dictionary of 17 common functional groups (e.g., alcohols, ketones, ethers; see Table 11), we identify the presence of each group via substructure search implemented with RDKit [94]. Let G_1 and G_2 denote the sets of functional group types identified in molecule 1 and molecule 2, respectively. The functional group similarity S_{FG} is then computed as the Jaccard index over these sets:

$$S_{\text{FG}}(mol_1, mol_2) = \frac{G_1 \cap G_2}{G_1 \cup G_2}$$

This metric reflects the qualitative overlap in functional group composition, regardless of group count or spatial arrangement. Ranging from 0 (no shared functional group types) to 1 (complete overlap), it offers a chemically interpretable measure that complements structure-level similarity metrics.

4.6.3 Molecular Fingerprint Similarity Metrics

Molecular fingerprints are vector representations that encode the presence of specific substructures within a molecule, and are widely used to assess molecular similarity. In this work, we compute fingerprint-based similarities between predicted and reference molecules using RDKit and several established metrics:

Tanimoto Similarity. Given two binary fingerprints A and B, with $a = |A|$, $b = |B|$, and $c = |A \cap B|$, the Tanimoto coefficient is defined as:

$$\text{Tanimoto}(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{c}{a + b - c}$$

This metric is computed using ECFP4 circular fingerprints, which capture atom-centered substructures within a radius of 2 bonds.

MACCS Tanimoto Similarity. Identical in form to the standard Tanimoto coefficient, but computed over 166-bit MACCS keys, which encode the presence of predefined structural fragments rather than circular environments.

Cosine Similarity. For continuous-valued fingerprints (e.g., neural embeddings), similarity is computed as the cosine of the angle between vectors:

$$\text{Cosine}(A, B) = \frac{A \cdot B}{\| A \| \| B \|}$$

Fraggle Similarity. Based on fragment matching rather than fixed-length vectors, this metric decomposes both molecules into substructures and computes alignment-based similarity. It is implemented via RDKit's `FraggleSim` module, and reflects partial structural overlap beyond atom-level matching.

4.6.4 Derived Statistical Metrics Based on Fingerprint Similarity

In addition to raw similarity scores, we report threshold-based classification metrics that stratify prediction quality according to Tanimoto similarity levels. These derived measures capture not only exact matches but also structurally plausible alternatives that retain key substructures of the target molecule.

Approximate Match Rate. The proportion of predicted molecules with Tanimoto similarity ≥ 0.675 to the reference structure. This threshold is widely adopted to indicate moderate-to-high structural similarity in cheminformatics.

Acceptable Match Rate. The proportion of predictions with Tanimoto similarity ≥ 0.4, often associated with weak but potentially chemically relevant similarity.

4.6.5 Maximum Common Substructure Similarity

The Maximum Common Edge Subgraph (MCES) similarity quantifies the extent of structural overlap between two molecules by identifying their largest common subgraph under graph isomorphism. Given two molecular graphs G_1 and G_2, the MCES is defined as the largest subgraph G_{common} such that $G_{\text{common}} \subseteq G_1$ and $G_{\text{common}} \subseteq G_2$. The similarity is computed as the ratio of shared bonds to the maximum number of bonds in either molecule:

$$\text{MCES Similarity} = \frac{E(G_{\text{common}})}{\max(|E(G_1)|, |E(G_2)|)}$$

where $E(G)$ denotes the set of edges (chemical bonds) in graph G. This metric captures partial structural correctness and is particularly informative when global fingerprint-based measures fail to reflect local substructure similarity.

4.7 Related works and baseline models

To benchmark the performance of SpectraLLM, we compare it with state-of-the-art spectrum-to-structure models spanning various spectroscopic modalities, including mass spectrometry, vibrational spectroscopy, and NMR.

For mass spectral data, we compare against DiffMS [87], a diffusion-based molecular generation framework tailored for MS-driven structure elucidation. DiffMS employs a Transformer-based encoder to process peak formulas, neutral losses, and other spectral features, followed by a discrete graph diffusion decoder conditioned on the molecular formula. Peak attributes are embedded using MISTr, and the model is evaluated on the MassSpecGym and MultimodalSpectro datasets. As DiffMS permits controlling the number of SMILES outputs during decoding, we fix this number to 1 to ensure comparability.

For vibrational and electronic spectra, we consider Spectra-to-Structure[85] and IR-to-Structure[36] as baselines. Spectra-to-Structure uses a dual-encoder architecture comprising a Transformer for spectral segmentation and an EGNN for molecular graph embedding, jointly optimized via a contrastive loss on their latent spaces. IR-to-Structure adopts a sequence-to-sequence Transformer that generates SMILES from IR spectra and molecular formulas. To avoid information leakage, we follow the authors' "spectrum-only" protocol for both training and evaluation. As both

Table 8: LLM-only Models Surpass Vision–Language Pipelines in Spectral Structure Elucidation.

Architecture	trans_rate ↑	Accuracy ↑	Tanimoto ↑	Cosine ↑	MCES ↓	Functional Group ↑	Tanimoto(MACCS) ↑	Fraggle ↑
VLM–LLM	98.74%	0.0291	0.2455	0.3681	9.9964	0.6045	0.5098	0.4785
LLM	99.79%	**0.1983**	**0.4875**	**0.5973**	**8.1151**	**0.8103**	**0.7099**	**0.6222**

models were originally trained on IR data, we re-trained them on Raman and UV spectra using the Multimodal and QM9s datasets, adopting the same training configurations and preprocessing procedures described in their respective works.

For NMR-based prediction, we include NMR2Struct [39] as a baseline. This model integrates a convolutional encoder to extract spectral features and a Transformer decoder for SMILES generation and substructure prediction. Pre-training is conducted on a substructure assembly task to improve compositional generalization. Unlike many prior models, NMR2Struct performs direct structure generation from spectral input, without relying on the molecular formula.

4.8 Selection of model architecture

Designing an effective model architecture for structure elucidation from spectroscopic data requires careful consideration of the modality-specific characteristics and their interaction with large-scale language modeling. We explored two principal approaches for integrating multimodal spectral inputs with large language models: (1) a vision-language-centric pipeline in which continuous spectra are first interpreted via a visual encoder, and their latent representation is subsequently merged with discrete mass spectra for structure generation; and (2) a unified language-based architecture that encodes both continuous and discrete spectra as structured textual descriptions, directly processed by a pretrained LLM.

Inspired by recent advances in vision–language models (VLMs), we initially explored a two-stage architecture for multimodal spectral inference: IR , Raman, and UV-Vis spectra were rendered into 2D images and processed by a vision encoder (e.g., Qwen2.5-VL). The encoder output was decoded via beam sampling into a preliminary SMILES candidate. This intermediate representation, alongside the mass spectrum (encoded as m/z–intensity pairs or peak descriptors), was subsequently input into a large language model (LLM, e.g., Qwen3–32B) to generate the final molecular structure.

While this hybrid approach leveraged the representational flexibility of VLMs for interpreting unstructured plots, it introduced several critical limitations:

- The intermediate SMILES representation was lossy and error-prone, potentially propagating inaccuracies from the vision stage to the language model.
- The handoff between modalities disrupted end-to-end optimization, as the system lacked a shared latent space and consistent gradient flow.
- When input spectra were incomplete or noisy, the intermediate SMILES failed to preserve chemically meaningful cues, leading to reduced interpretability and robustness.

By contrast, the language-centric architecture presented in this study employs discrete peak features from all modalities as structured input to a unified LLM, eliminating the need for intermediate symbolic forms. As shown in Table 8, we benchmarked both architectures on the QM9s and MultimodalSpectro datasets, reporting structure reconstruction performance across multiple criteria, including exact match rate, top-K accuracy, and Tanimoto similarity.

Across all metrics and datasets, the unified LLM paradigm consistently outperformed the VLM–LLM cascade. Notably, improvements were most pronounced under stricter evaluation thresholds, underscoring the LLM's superior reasoning capabilities in a purely linguistic input space. We attribute this performance gain to the model's ability to jointly process all modalities in a chemically coherent and interpretable manner, without cross-modal encoding artifacts.

These findings support the adoption of a language-based modeling framework for multimodal spectral analysis. While VLMs remain valuable for raw spectrum visualization and downstream interpretability, symbolic generation is best achieved via a unified language model trained on curated, peak-based prompts.

Table 9: Representative prompts for single-modality spectroscopic inference.

Single-spectrum Prompts

Human: Given **Infrared** Spectrum {Wavenumbers: 3596.8,3549.49,3314.97,3202.86,3141.14, 1498.51,1444.0,724.0, Intensities: 1.0,0.16,0.24,0.12,0.38,0.51,0.65,0.76, Widths: 21.23, 9.14,15.3,11.94,15.47,17.52,25.88,15.47}, the spectra data includes the wavenumber positions in reciprocal centimeters as Wavenumbers and corresponding intensities as Intensities, corresponding width as Widths. Based on the information provided, predict which compound the spectra correspond to and give the SMILES of that compound. Please answer strictly in the format **##**SMILES:

GPT: ##SMILES: ON=C1CCCC1

Human: Given **Raman** spectroscopy {Wavenumbers: 2913.0,2926.0,2981.0,3685.0, Intensities: 0.75, 1.0,1.0,0.61, Widths: 4.39,22.34,24.45,10.08}, the spectra data includes the wavenumber positions in reciprocal centimeters as Wavenumbers and corresponding intensities as Intensities, corresponding width as Widths. Based on the information provided, predict which compound the spectra correspond to and give the SMILES of that compound. Please answer strictly in the format **##**SMILES:

GPT: ##SMILES: ON=C1CCCC1

Human: Given **Ultraviolet-visible** spectroscopy {Energies: 6.54,7.52, Intensities: 1.0,0.34, Widths: 0.75,0.37}, the spectra data includes the energy positions in eV as Energies and corresponding intensities as Intensities, corresponding width as Widths. Based on the information provided, predict which compound the spectra correspond to and give the SMILES of that compound. Please answer strictly in the format **##**SMILES:

GPT: ##SMILES: ON=C1CCCC1

Human: Given **Carbon-13 Nuclear Magnetic Resonance** {C-shifts: 146.12,105.1,77.93, 42.24, Intensities: 1.0,0.37,0.92,0.78}, the spectra data includes the Chemical Shift positions in ppm as C-shifts and corresponding intensities as Intensities. Based on the information provided, predict which compound the spectra correspond to and give the SMILES of that compound. Please answer strictly in the format **##**SMILES:

GPT: ##SMILES: C1=COCC1

Human: Given **Proton Nuclear Magnetic Resonance** {H-shifts: 8.32,8.31,6.96,6.95,6.95, 6.95,6.94,6.94,6.93,6.31,6.31,6.3,4.59,4.58,4.57,4.56, Intensities: 0.5,0.47,0.14,0.26,0.26, 0.2,0.19,0.32,0.16,0.51,1.0,0.53,0.33,0.89,0.91,0.34}, the spectra data includes the Chemical Shift positions in ppm as H-shifts and corresponding intensities as Intensities. Based on the information provided, predict which compound the spectra correspond to and give the SMILES of that compound. Please answer strictly in the format **##**SMILES:

GPT: ##SMILES: C1=COCC1

Human: Given **Heteronuclear Single Quantum Coherence** {H-shifts: 4.3,2.57,4.94,6.32, C-shifts: 68.39,27.85,99.64,146.03, Intensities: 1.0,1.0,1.0,1.0}, the spectra data includes the Chemical Shift positions in ppm as C-shifts,H-shifts and corresponding intensities as Intensities. Based on the information provided, predict which compound the spectra correspond to and give the SMILES of that compound. Please answer strictly in the format **##**SMILES:

GPT: ##SMILES: C1=COCC1

Human: **Mass** spectrum data: {mzs: 134.1,202.08,216.14,244.13,266.11,284.12,384.18,402.19, 545.26,645.31,806.4, intensities: 0.056,0.039,0.036,0.05,0.204,0.111,0.999,0.06,0.119, 0.351,0.63} Adduct type: [M+Na]+ Collision energy: 35(NCE) Please predict the compound's SMILES representation with LESS THAN 1000 characters thinking. The final output must strictly begin with **##**SMILES:

GPT: ##SMILES: CC(C)[C@@H]1C(=O)N([C@H](C(=O)O[C@@H(C(=O)N([C@H](C(=O)O[C@@H] (C(=O)N([C@H](C(=O)O1)CC2=CC=CC=C2)C)C(C)C)CC3=CC=CC=C3)C)C(C)CCC4=CC=C C=C4)C

Table 10: Representative prompts for multimodal spectroscopic reasoning.

Multi-spectrum Prompts

Human: Given multiple spectra, they are Infrared Spectrum {Wavenumbers:3596.8,3549.49, 3314.97,3202.86,3141.14,1498.51,1444.0,724.0,Intensities:1.0,0.16,0.24,0.12,0.38,0.51, 0.65,0.76, Widths:98.69,16.5,77.31,40.6,114.88,42.94,142.13,67.04}, Raman spectroscopy {Wavenumbers:2913.0,2926.0,2981.0,3685.0, Intensities: 0.75,1.0,1.0,0.61, Widths:4.39, 22.34,24.45,10.08}, Ultraviolet-visible spectroscopy {Energies:6.54,7.52, Intensities:1.0, 0.34, Widths:0.75,0.37}. All of these spectra are determined by the same compound, with the wavenumber postions in reciprocal centimeters as Wavenumbers, the energy postions in eV as Energies and corresponding intensities as Intensities. Based on the information provided by these spectra, predict which compound the spectra correspond to and give the SMILES of that compound. Please answer strictly in the format ##SMILES:

GPT: ##SMILES: ON=C1CCCC1

Human: Given multiple spectra, they are Carbon-13 Nuclear Magnetic Resonance {C-shifts: 146.12,105.1,77.93,42.24, Intensities:1.0,0.37,0.92,0.78}, Proton Nuclear Magnetic Resonance {H-shifts:8.32,8.31,6.96,6.95,6.95,6.95,6.94,6.94,6.93,6.31,6.31,6.3,4.59,4.58, 4.57,4.56, Intensities:0.5,0.47,0.14,0.26,0.26,0.2,0.19,0.32,0.16,0.51,1.0,0.53, 0.33,0.89, 0.91,0.34}, Heteronuclear Single Quantum Coherence {H-shifts:4.3,2.57,4.94,6.32, C-shifts:68.39,27.85,99.64,146.03, Intensities: 1.0,1.0,1.0,1.0}, Infrared Spectrum {Wave-numbers:3961.98,3931.96,3903.95,3761.87,3731.85,3703.84,3491.72, 3477.71,3449.69, 3041.47,2999.44,2767.32,2637.24,2623.24,2609.23,2595.22,1440.58, 1434.57,1420.57, 1414.56,1404.56,1392.55, Intensities:0.21,0.21,0.38,0.92,0.64,1.0,0.13,0.11,0.16,0.11,0.14, 0.11,0.13,0.11,0.11,0.18,0.12,0.22,0.16,0.1,0.16,0.18, Widths:3.79,3.15,2.35,3.7,4.26,4.22, 3.72,2.37,2.39,4.23,2.78,2.49,3.85,3.15,3.03,2.76,6.27,4.73,3.96,3.23,2.65,4.28}, Mass spectrum data {mzs:39.02,41.04,45.03,53.04, Intensities:0.28, 1.0,0.22,0.23}. All of these spectra are determined by the same compound, with the wavenumber postions in reciprocal centimeters as Wavenumbers, the energy postions in eV as Energies and corresponding intensities as Intensities. Based on the information provided by these spectra, predict which compound the spectra correspond to and give the SMILES of that compound. Please answer strictly in the format ##SMILES:

GPT: ##SMILES: C1=COCC1

Table 11: Functional group definitions used.

Index	Group	Definition
1	Alkane	[CX4]
2	Alkene	[CX3]=[CX3]
3	Alkyne	[CX2]#C
4	Arene	[$([cX3](:*):*),$([cX2+](:*):*)]
5	Alcohol	[#6][OX2H]
6	Ether	[OD2]([#6])[#6]
7	Aldehyde	[CX3H1](=O)[#6]
8	Ketone	[#6][CX3](=O)[#6]
9	Carboxylic acid	[CX3](=O)[OX2H1]
10	Ester	[#6][CX3](=O)[OX2H0][#6]
11	haloalkane	[#6][F,Cl,Br,I]
12	Alkyl halide	[CX3](=[OX1])[F,Cl,Br,I]
13	Amine	[NX3;$(NC=O)]
14	Amide	[NX3][CX3](=[OX1])[#6]
15	Nitrile	[NX1]#[CX2]
16	Sulfide	[#16X2H0]
17	Thiol	[#16X2H]