SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

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

[1]New Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA)

[2]The Hong Kong Polytechnic University

[3]Tianjin University Peiyangyuan Campus

*Yunyue Su and Jiahui Chen contributed equally to this work.

†Corresponding author: qiang.liu@nlpr.ia.ac.cn

SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

Background

- Molecular structure elucidation is a **fundamental task** across chemistry, biology, and materials science, supporting molecular/crystal structure determination for mechanistic research, material design, and product optimization.
- Spectroscopic techniques (IR, Raman, UV-Vis, NMR, MS) are **indispensable tools** for this task: each probes distinct physicochemical properties of molecules, and provides complementary structural information.
- Human experts rely on **joint interpretation of multi-spectral data** to resolve structural ambiguity, which cannot be achieved by single-spectrum analysis alone.
- Machine learning has emerged as a promising approach to **automate spectrum-to-structure inference**, with mainstream methods built on CNN and Transformer architectures.

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

Overview of spectroscopic modalities

Each provides unique and complementary structural constraints for molecular identification

SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

Current Limitations

- Confined to **single spectral modality**, unable to integrate complementary multi-spectral evidence.
- **Fixed numerical feature** design restricts experimental context integration, leading to poor generalization.
- Heavily dependent on **pre-compiled spectral libraries**, with weak performance on unknown molecules.
- Modality-specific models **lack a unified framework** for cross-modal collaborative reasoning.

Our Proposal

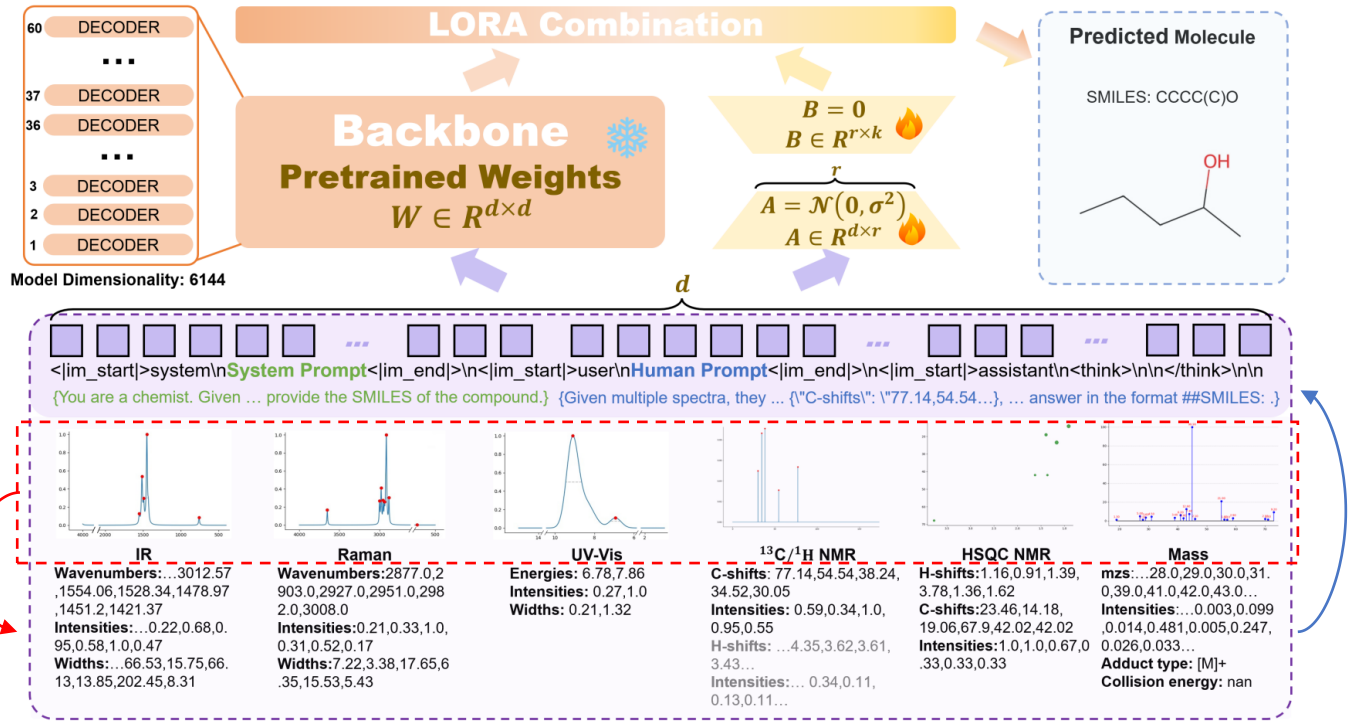
- Build a **unified spectral processing framework** through natural language, and flexibly add contextual experimental metadata.
- Enable **joint analysis** of one or more spectral modalities, leveraging complementary information across different spectra.
- Realize **end-to-end structure prediction** without library matching, handcrafted rules, or modality-specific encoders.
- Guarantee **robust performance** in unimodal settings, with monotonically improved accuracy as spectral diversity increases.

SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

SpectraLLM

Discretizing Continuous and Discrete Spectra into Sequence Tokens:

- Extract **significant feature signals** from each spectrum, including peak positions, intensities, and widths, to construct a unified prompt.
- Embed relevant **experimental metadata** (e.g., instrument type, collision energy, ionization mode) into prompts for better contextual modeling.
- Enable **joint modeling** of all five spectra (IR, Raman, UV-Vis, NMR, MS), unifying both continuous and discrete modalities in a single shared language space.



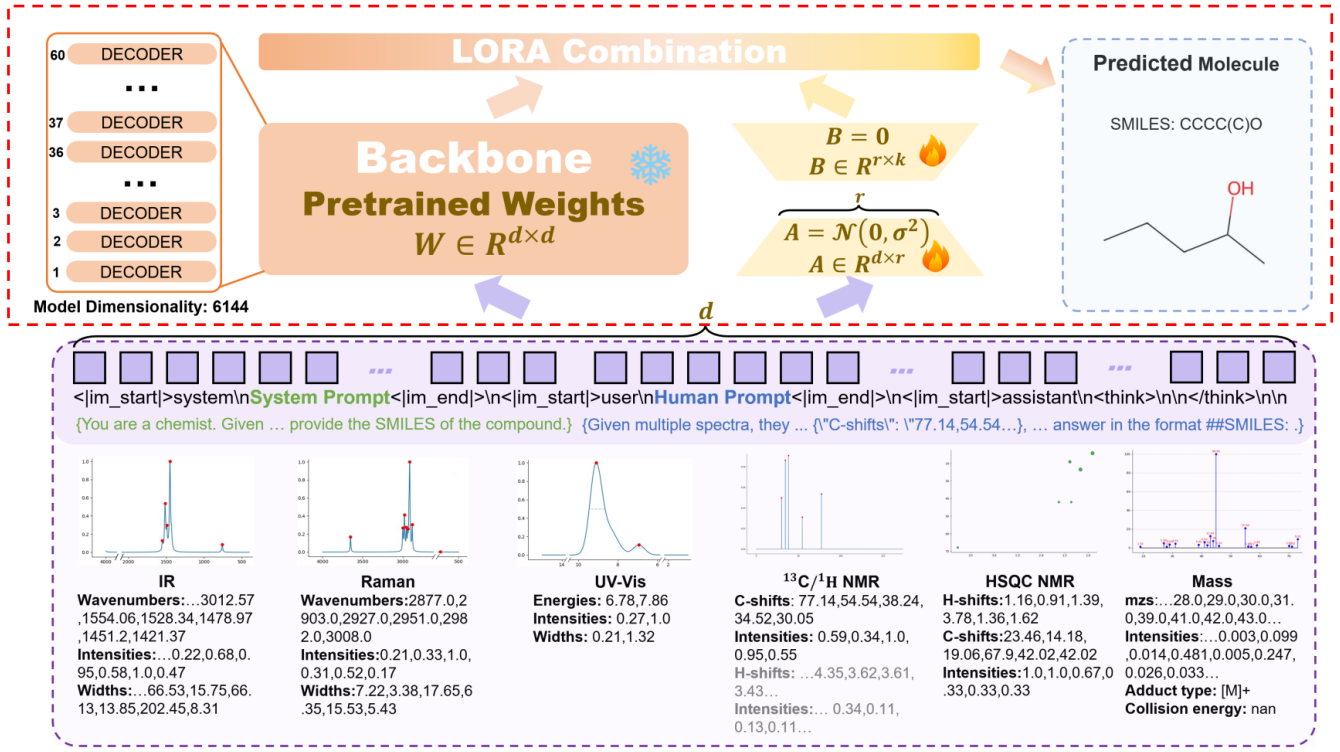
Overview of the training pipeline for structure elucidation

SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

SpectraLLM

Sequence-to-SMILES Generation via LLM with LoRA:

- Utilizing a large-scale pre-trained language model (Qwen3-32B) as the frozen backbone, with parameter-efficient fine-tuning via LoRA on massive spectrum-structure paired data.
- Preserves the general linguistic reasoning ability of the pre-trained LLM, while efficiently aligning spectral features with linguistic priors.
- Achieves direct, autoregressive generation of chemically valid SMILES sequences from tokenized spectroscopic inputs.



Overview of the training pipeline for structure elucidation

SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

SpectraLLM

Core Prompt Design Principles:

- Each spectral segment is clearly marked with its modality type, for the model to accurately distinguish the physical meaning of different spectra.
- Key spectral features are encoded in a consistent key-value format, ensuring format alignment across all modalities.
- Experimental metadata is naturally embedded into prompts to enhance the model’ s generalization across different acquisition settings.
- Adopt the "Human-GPT" instruction-response format, which aligns with the pre-trained LLM’ s conversational paradigm to reduce fine-tuning costs.

Multi-spectrum Prompts	
Case1	
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`
Case2	
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 {Wavenumbers: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`

Representative prompts for multimodal spectroscopic reasoning

SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

Settings

Four datasets

Dataset	Modality Coverage	Data Type	Total Molecules	Total Spectra
QM9s	IR, Raman, UV-Vis	Simulated	129,817	389,451
Multimodal Spectroscopic	IR, $^{1}H/^{13}C$/HSQC NMR, MS	Simulated	794,403	4,766,418
MassSpecGym	MS	Experimental	29,000	231,104
MassBank	MS	Experimental	-	122,746
ALL	IR, Raman, UV-Vis, NMR, MS	Simulated + Experimental	**943,732**	**5,510,655**

Six Metric Categories

- **Chemical Validity:** Validity of generated SMILES sequences
- **Structural-level correctness:** Functional Group overlap, MCES, Fraggle similarity
- **Global structure similarity:** Tanimoto (ECFP4), Tanimoto (MACCS), Cosine similarity

State-of-the-Art Baselines

- **Vibrational spectra:** IR-to-Structure, Spectra2Structure
- **NMR:** NMR2Struct
- **Mass spectrometry:** Spec2Mol, Diffms

SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

Main Results

1. Unimodal Comparative Analysis

- SpectraLLM achieves state-of-the-art performance across all unimodal settings, with near-perfect chemical validity and significantly higher structural similarity than modality-specific baselines.

Spectrum	Method	Validity ↑	Tanimoto ↑	Cosine ↑	MCES ↓	Functional Group ↑	Tanimoto (MACCS) ↑	Fraggle ↑
					QM9s			
IR	IR-to-Structure	100.00%	0.0718	0.1311	11.3187	0.3151	0.1585	0.1747
	Spectra2Structure	100.00%	0.0965	0.1695	10.1081	0.4383	0.2162	0.2308
	SpectraLLM	99.82%	**0.1921**	**0.3120**	**7.5651**	**0.6599**	**0.4330**	**0.3194**
Raman	IR-to-Structure	100.00%	0.0766	0.1395	11.3516	0.3525	0.1639	0.1959
	Spectra2Structure	100.00%	0.1089	0.1901	9.4164	0.4419	0.2388	0.2504
	SpectraLLM	99.08%	**0.2500**	**0.3786**	**6.4076**	**0.7317**	**0.5071**	**0.2500**
UV-Vis	IR-to-Structure	100.00%	0.0728	0.1326	11.424	0.3151	0.1512	0.1837
	Spectra2Structure	100.00%	0.0716	0.1313	11.1222	0.3901	0.1418	0.2092
	SpectraLLM	100.00%	**0.0790**	**0.1426**	**10.6374**	**0.3713**	**0.2026**	**0.2100**
					Multimodal Spectroscopic			
NMR	NMR2Struct	47.62%	0.0433	0.1029	30.6938	0.1718	0.1294	0.0962
	SpectraLLM	98.92%	**0.4151**	**0.5322**	**8.3091**	**0.7209**	**0.6367**	**0.5862**

Comparative evaluation of SpectraLLM and conventional approaches under individual spectral inputs

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
SpectraLLM	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
SpectraLLM	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
SpectraLLM	98.44%	**0.1286**	**0.2229**	**0.4539**	**0.3787**	**0.3150**

Benchmarking SpectraLLM against established mass spectrometry-based inference models

SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

Main Results

2. Multimodal Fusion & Synergy Analysis

Enhanced structure elucidation through fusion of complementary spectroscopic modalities

Inputs	Validity ↑	Tanimoto ↑	Cosine ↑	MCES ↓	Functional Group ↑	Tanimoto (MACCS) ↑	Fraggle ↑
QM9s							
IR	99.82%	0.1921	0.3120	7.5651	0.6599	0.4330	0.3194
Raman	99.08%	0.2500	0.3786	6.4076	0.7317	0.5071	0.2500
UV-Vis	100.00%	0.0790	0.1426	10.6374	0.3713	0.2026	0.2100
Jointly	98.72%	**0.3355**	**0.4560**	**4.9647**	**0.7934**	**0.5785**	**0.4117**
Multimodal Spectroscopic							
IR	99.63%	0.1720	0.2868	15.3234	0.6023	0.4031	0.3906
MS	99.64%	0.1844	0.2993	11.3243	0.4929	0.4254	0.4282
IR+MS	99.25%	0.2300	0.3519	10.4164	0.6345	0.4887	0.4566
^{13}C NMR	99.64%	0.1016	0.1801	14.8865	0.4249	0.2952	0.3607
^{1}H NMR	99.10%	0.0720	0.1341	18.6141	0.3329	0.2203	0.2572
HSQC NMR	99.64%	0.2058	0.3221	13.4919	0.5495	0.4392	0.4274
Jointly NMR	98.92%	0.4151	0.5322	8.3091	0.7209	0.6367	0.5862
Jointly NMR+IR	99.60%	0.4121	0.5341	8.3855	0.7764	0.6575	0.5809
Jointly NMR+MS	98.37%	0.4518	0.5601	8.1682	0.7618	0.6760	0.6063
Jointly NMR+IR+MS	99.79%	**0.4875**	**0.5973**	**8.1151**	**0.8103**	**0.7099**	**0.6222**

- IR+Raman achieves the strongest synergy (Tanimoto=0.5320), outperforming even the 3-modality combination, highlighting the complementary nature of vibrational spectra.

Train-test modality transfer matrix on QM9s (Tanimoto similarity)

test\train	IR	Raman	UV-Vis	IR+Raman	IR+UV-Vis	Raman+UV-Vis	IR+Raman+UV-Vis
IR	0.1474	0.0674	0.0605	0.2012	0.2012	0.0629	0.1921
Raman	0.0669	0.3251	0.0721	0.3951	0.0639	0.4112	0.2977
UV-Vis	0.0592	0.0632	0.0777	0.0579	0.086	0.0877	0.0812
IR+Raman	0.1372	0.1209	0.0671	0.532	0.1468	0.1319	0.4231
IR+UV-Vis	0.1379	0.0642	0.0704	0.1803	0.2088	0.0645	0.1962
Raman+UV-Vis	0.0671	0.2746	0.0741	0.3687	0.0786	0.4594	0.3353
IR+Raman+UV-Vis	0.1373	0.1161	0.0706	0.4922	0.17	0.1305	0.4213

- Prediction accuracy improves monotonically with increasing spectral diversity, as complementary structural constraints from different modalities reduce inference ambiguity.

Effect of IR, Raman, and UV-Vis modality combinations on spectrum-to-structure performance on QM9s

SpectraLLM Q9MS retrained	Validity ↑	Tanimoto ↑	Tanimoto (MACCS) ↑	Cosine ↑	MCES ↓	Functional Group ↑	Fraggle ↑
IR-Only	100.00%	0.1474	0.3436	0.2462	7.1245	0.5309	0.2686
Raman-Only	98.90%	0.3251	0.5748	0.4490	5.4009	0.7986	0.3958
UV-Vis-Only	100.00%	0.0777	0.2081	0.1407	10.3874	0.3600	0.2137
IR+Raman	99.63%	0.5320	0.7363	0.6288	3.0312	0.8834	0.4806
IR+UV-Vis	99.63%	0.2088	0.4471	0.3280	6.0551	0.6791	0.3333
Raman+UV-Vis	99.45%	0.4594	0.6718	0.5699	4.0405	0.8431	0.4561
IR+Raman+UV-Vis	98.53%	0.4053	0.6501	0.5202	3.9926	0.8370	0.4479

- Models trained on multi-spectral inputs show significantly better cross-modality generalization than unimodal-only models.

SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

Main Results

3. Robustness & Generalization Validation

Performance on Real-World MS Datasets (MassSpecGym / MassBank)

Setting	State	Model	Validity ↑	Tanimoto ↑	Cosine ↑	Functional Group ↑	Tanimoto (MACCS) ↑	Fraggle ↑
MassSpecGym	trained	Diffms	57.16%	0.159705	0.242187	0.489004	0.430529	0.353878
		SpectraLLM	99.74%	0.1537	0.2565	0.5016	0.4735	0.362
	untrained	Spec2Mol	62.86%	0.084861	0.151115	0.311157	0.270927	0.206558
		SpectraLLM	92.97%	0.1139	0.2039	0.4354	0.4005	0.2988
MassBank	trained	Diffms	23.63%	0.074269	0.100735	0.179588	0.156596	0.153949
		SpectraLLM	98.46%	0.1306	0.2264	0.454	0.3848	0.3201
	untrained	Spec2Mol	71.63%	0.085741	0.153861	0.299915	0.241751	0.214962
		SpectraLLM	92.79%	0.1125	0.2019	0.187	0.3612	0.2922

Zero-Shot Transfer to Unseen NIST IR/UV Experimental Spectra

Datasets	Model	Validity ↑	Tanimoto ↑	Cosine ↑	MCES ↓	Functional Group ↑	Tanimoto (MACCS) ↑	Fraggle ↑
NIST-ir	IR-to-Structure	98.02%	0.0095	0.0657	33.4585	0.1309	0.0699	0.0720
	Spectra2Structure	99.31%	0.0128	0.0769	29.5700	0.2104	0.0923	0.1267
	SpectraLLM	99.43%	0.0727	0.1368	22.3229	0.3191	0.1964	0.2176
NIST-uv	IR-to-Structure	99.12%	0.0673	0.1147	25.9537	0.2799	0.1529	0.1101
	Spectra2Structure	99.37%	0.0726	0.1285	23.3650	0.3058	0.1989	0.1407
	SpectraLLM	99.56%	0.0744	0.1381	23.0612	0.3231	0.2084	0.1411

- SpectraLLM achieves strong zero-shot transfer to real-world experimental spectra, outperforming baselines even without training on these datasets.

Effect of modality-specific simulated noise on spectrum-to-structure performance on QM9s Jointly

*Note: I=IR, R=Raman, U=UV-Vis; * denotes the modality with added simulated noise*

SpectraLLM QM9s Jointly	Noisy Strong	Validity ↑	Tanimoto ↑	Cosine ↑	MCES ↓	Functional Group ↑	Tanimoto (MACCS) ↑	Fraggle ↑
I+R+U	-	98.72%	0.3355	0.4560	4.9647	0.7934	0.5785	0.4117
I*+R+U	mild	96.89%	0.3365	0.4659	5.3299	0.7819	0.5955	0.4011
	moderate	97.99%	0.2688	0.3963	6.2112	0.6851	0.5212	0.3598
	severe	87.18%	0.1712	0.2799	8.4433	0.5010	0.3774	0.2556
I+R*+U	mild	98.17%	0.2487	0.3785	6.4272	0.6669	0.5073	0.3498
	moderate	95.79%	0.1608	0.2691	7.8891	0.5168	0.3735	0.2811
	severe	97.44%	0.0827	0.1495	9.5714	0.2058	0.2611	0.3719
I+R+U*	mild	98.90%	0.4076	0.5302	4.4417	0.8744	0.6746	0.4624
	moderate	99.27%	0.3972	0.5223	4.4511	0.8608	0.6611	0.4554
	severe	99.08%	0.3901	0.5153	4.5018	0.8547	0.6530	0.4478
I*+R*+U	mild	96.34%	0.2163	0.3398	7.0894	0.6063	0.4592	0.3115
	moderate	89.56%	0.1546	0.2607	8.4029	0.4828	0.3653	0.2655
	severe	57.88%	0.0749	0.1372	10.1472	0.3348	0.2466	0.1663
I+R*+U*	mild	98.53%	0.2550	0.3824	6.4526	0.6704	0.5060	0.3494
	moderate	96.15%	0.1602	0.2678	7.9476	0.5319	0.3795	0.2787
	severe	95.97%	0.0850	0.1535	9.5544	0.2074	0.2612	0.3704
I*+R+U*	mild	97.07%	0.3431	0.4719	5.2047	0.8098	0.6091	0.4124
	moderate	98.35%	0.2796	0.4048	6.1825	0.6789	0.5240	0.3682
	severe	80.40%	0.1694	0.2784	8.5068	0.4890	0.3712	0.2488
I*+R*+U*	mild	96.70%	0.2205	0.3473	6.9820	0.6272	0.4676	0.3154
	moderate	87.91%	0.1527	0.2556	8.4740	0.4659	0.3505	0.2548
	severe	39.01%	0.0792	0.1427	10.2981	0.3342	0.2508	0.1718

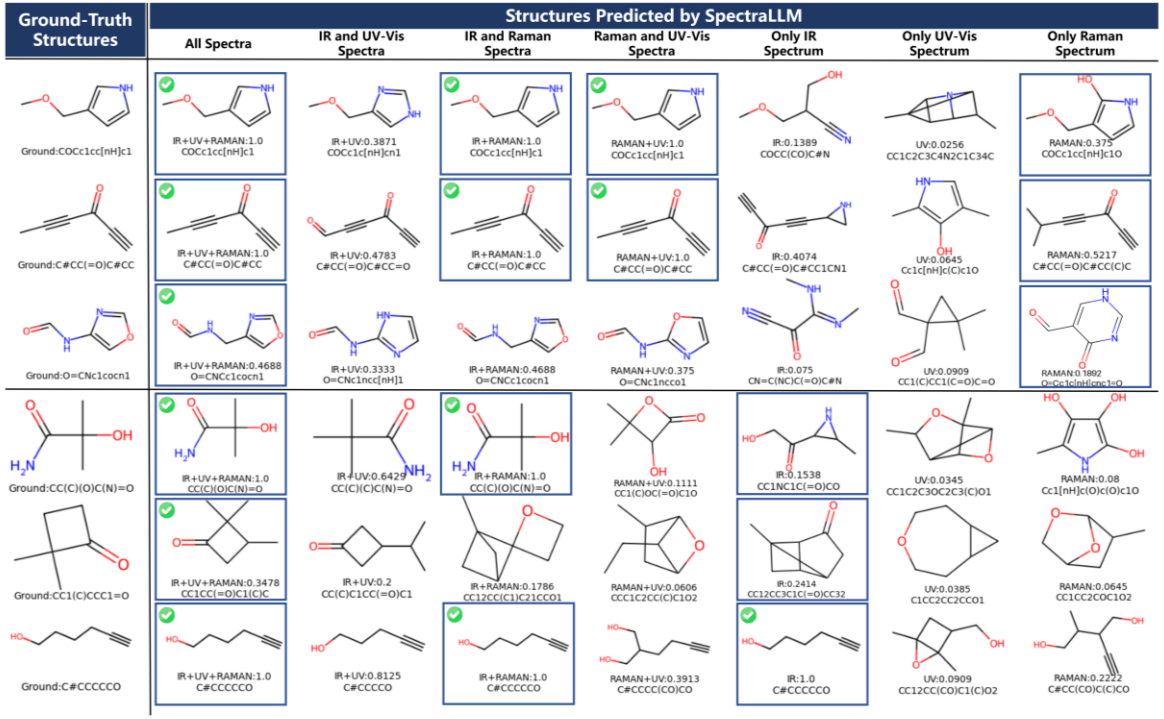
- SpectraLLM degrades gracefully under spectral noise, and leverages clean spectra to compensate for corrupted modalities.
- Even with severe noise on a single modality, the model retains meaningful predictive power.

SpectraLLM: Uncovering The Ability Of LLMs For Molecule Structure Elucidation From Multi-Spectral

Case Study

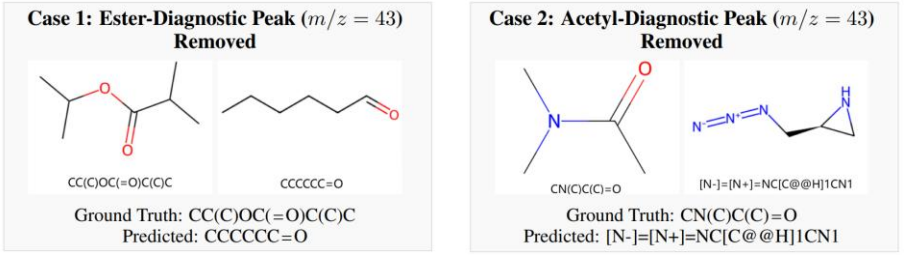
Positive Case: Multimodal Complementarity

**Green checkmark = exact match to ground truth*

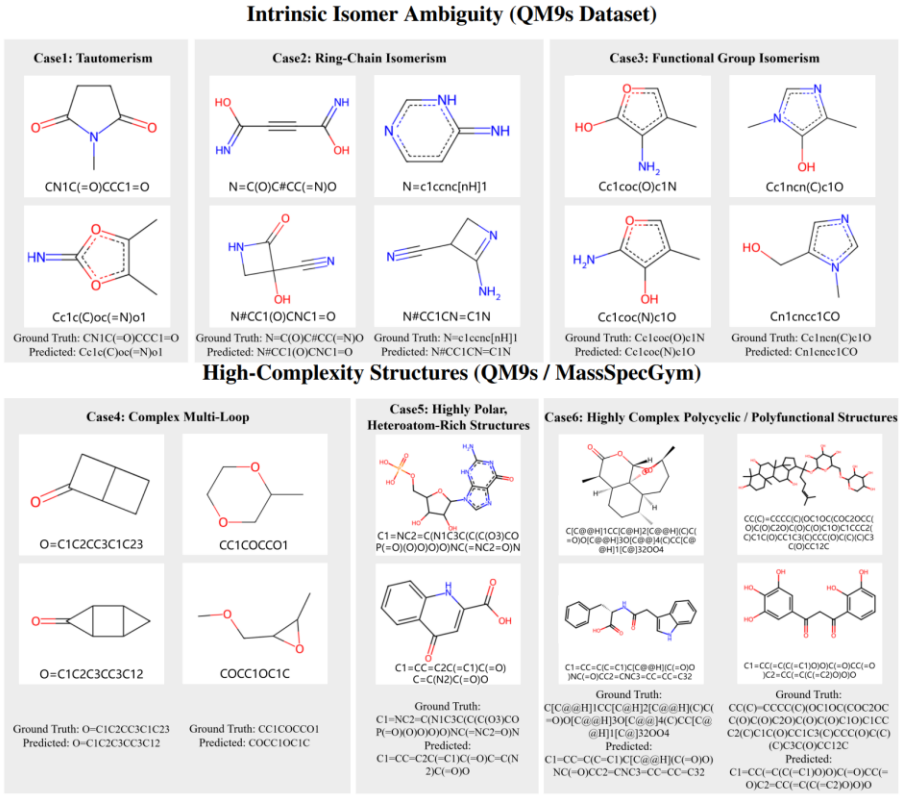


- Integrating complementary spectroscopic modalities eliminates structural ambiguity from unimodal inputs, and significantly improves the accuracy of molecular structure reconstruction.
- Main failure modes originate from the intrinsic spectral ambiguity of isomers (an information-theoretic limit of the inverse problem), rather than model defects.

Controlled Test: Diagnostic Peak Removal



Limitation Analysis: Failure Modes



中国科学院
CHINESE ACADEMY OF SCIENCES

ICLR

THANK YOU