

Enhancing Molecular Structure Elucidation with Reasoning-Capable LLMs

Martin Priessner,^{*[a]} Richard J. Lewis,^[b] Magnus J. Johansson,^[a] Jonathan M. Goodman,^[c] Jon Paul Janet,^[d] Anna Tomberg^{*[a]}

-
- [a] Dr. M. Priessner (ORCID: 0009-0005-7349-283X), Dr. A. Tomberg (ORCID: 0000-0002-0718-5381), Prof. M. J. Johansson (ORCID: 0000-0001-8811-2629)
Medicinal Chemistry, Research and Early Development, Cardiovascular, Renal and Metabolism (CVRM), BioPharmaceuticals R&D
AstraZeneca
Pepparedsleden 1, 43183 Mölndal (Sweden)
E-mail: martin.priessner@gmail.com, anna.tomberg@astrazeneca.com
- [b] Dr. R.J. Lewis (ORCID: 0000-0001-9404-8520)
Department of Medicinal Chemistry, Research & Early Development, Respiratory & Immunology, BioPharmaceuticals R&D
AstraZeneca
Pepparedsleden 1, 43183 Mölndal (Sweden)
- [c] Prof. J.M. Goodman (ORCID: 0000-0002-8693-9136)
Centre for Molecular Informatics, Department of Chemistry
University of Cambridge
Lensfield Road, Cambridge CB2 1EW, (UK)
- [d] Dr. J.P. Janet (ORCID: 0000-0001-7825-4797)
Molecular AI, Discovery Sciences, R&D
AstraZeneca
Pepparedsleden 1, 43183 Mölndal (Sweden)

Supplementary Section 1

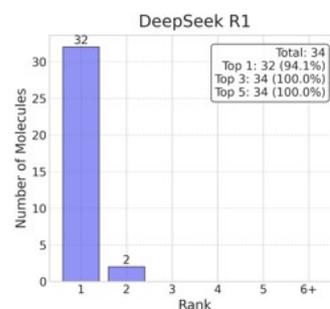
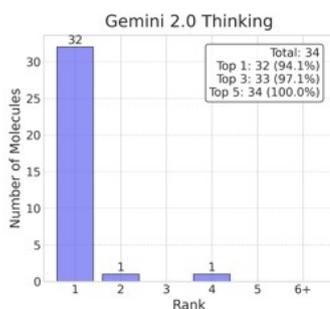
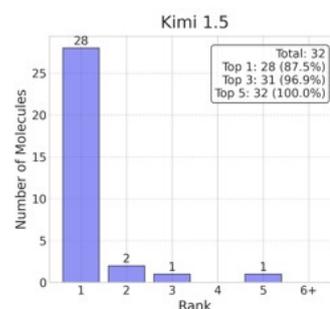
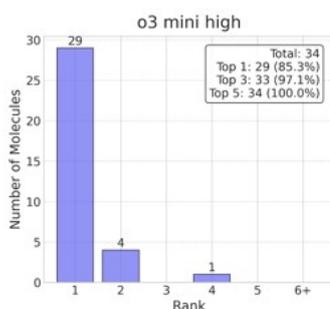
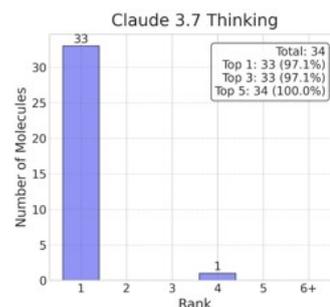
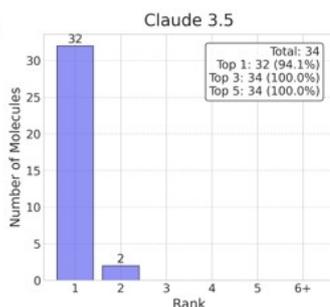
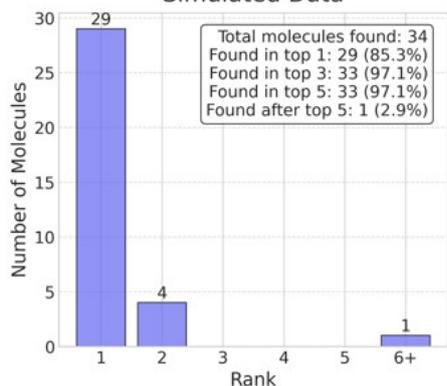
Comparative Evaluation of Alternative LLM Reasoning Models

In addition to our primary DeepSeek-R1 model discussed in the main text, we evaluated several other state-of-the-art language models to benchmark performance and provide comprehensive insights into LLM-based structure elucidation capabilities. Each model was presented with identical inputs and evaluation criteria to ensure fair comparison. All results from commercial models (Claude, Gemini, o3-Mini, Kimi) were generated between November 2024 and April 2025, using the publicly available APIs at that time.

Models and Technical Specifications

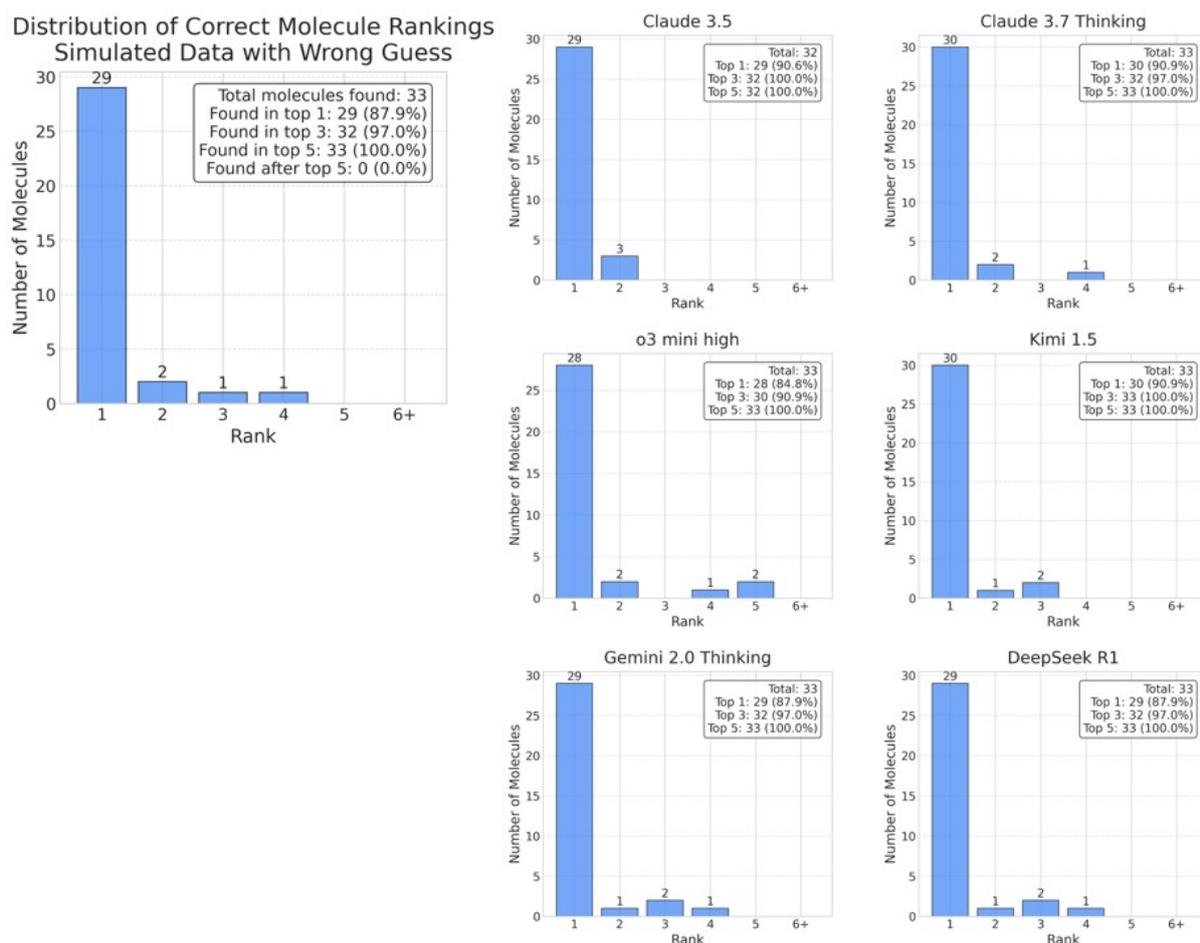
1. **Claude 3.5 Sonnet** (Anthropic): Served as our baseline LLM without explicit reasoning mode. Selected for its multimodal capabilities, particularly advanced vision processing, enabling direct analysis of molecular structure images alongside textual data. This model was primarily used for the initial visual and IUPAC-guided spectral analysis as well as comparative evaluation. Technical specifications included a maximum token output of 8,000 tokens.
2. **Claude 3.7 Sonnet-Thinking** (Anthropic): A specialized version of Claude with enhanced reasoning capabilities. This model employs an extended thinking process before generating responses, providing more thorough analytical reasoning for complex structure elucidation tasks. Technical specifications included a maximum token output of 20,000 tokens with up to 10,000 tokens allocated to the thinking process.
3. **Gemini 2.0 Flash-Thinking** (Google): An experimental version of Gemini optimized for reasoning tasks. This model was reserved specifically for the final comprehensive analysis step to leverage its enhanced reasoning capabilities for structural recommendation. Technical specifications included a maximum completion token limit of 20,000 tokens.
4. **o3-Mini** (OpenAI): Selected as the highest-performing reasoning model within the OpenAI-compatible 'o3' model family. Despite its relatively smaller size, we configured it with `reasoning_effort="high"` to maximize its analytical depth for molecular structure evaluation. Technical specifications included a maximum completion token limit of 20,000 tokens.
5. **Kimi 1.5** (Moonshot AI/ByteDance): A complementary reasoning model that provides an additional independent analytical perspective to the benchmark evaluation. Technical specifications included a maximum token output of 8,000 tokens with a temperature setting of 0.3.

Distribution of Correct Molecule Rankings Simulated Data



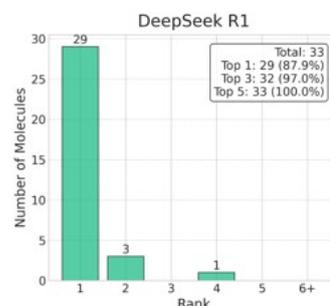
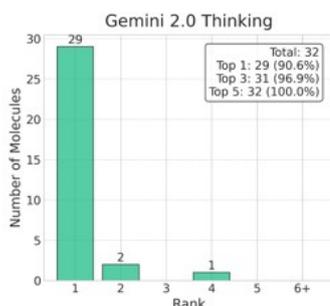
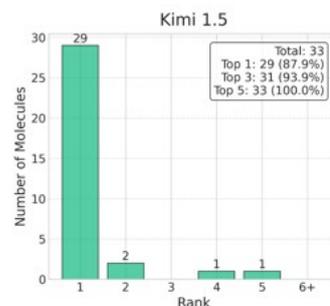
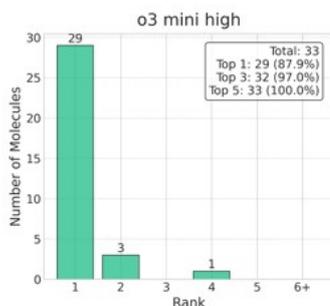
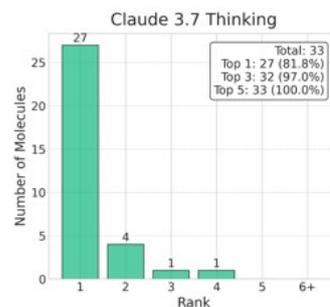
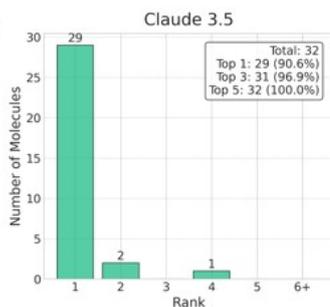
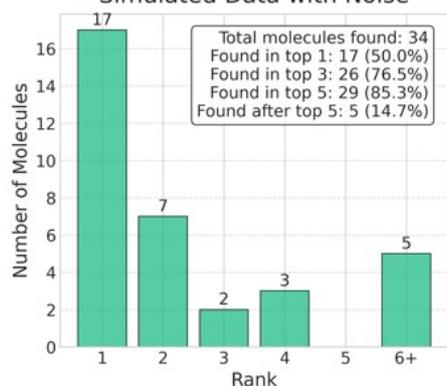
Supplementary Figure 1: Distribution of Correct Molecule Rankings for Simulated Data. The histograms show the ranking distribution of the correct molecular structures across different evaluation methods. The left panel displays results from baseline HSQC peak matching, where 29 molecules (85.3%) were correctly identified in the top position. The right panels show improved performance after LLM-enhanced re-ranking using various models: Claude 3.5 (32 molecules, 94.1%), Claude 3.7 Thinking (33 molecules, 97.1%), o3 mini high (29 molecules, 85.3%), Kimi 1.5 (28 molecules, 87.5%), Gemini 2.0 Thinking (32 molecules, 94.1%), and DeepSeek R1 (32 molecules, 94.1%). All LLM approaches demonstrate enhanced or maintained top-1 accuracy compared to baseline HSQC matching, with the most substantial improvements observed with Claude 3.7 Thinking, Gemini 2.0 Thinking, and DeepSeek R1.

Distribution of Correct Molecule Rankings
Simulated Data with Wrong Guess



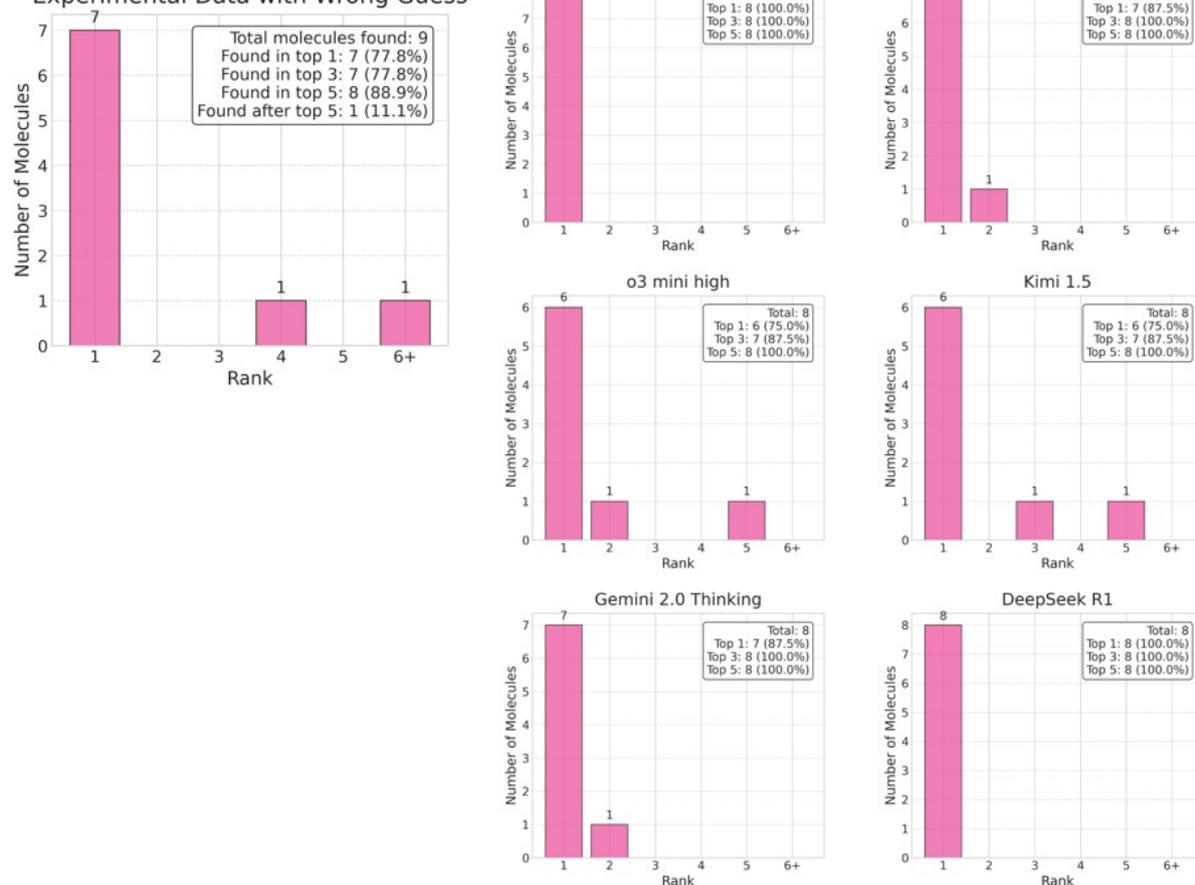
Supplementary Figure 2: Distribution of Correct Molecule Rankings for Simulated Data with Wrong Initial Guess (On Analogues). This figure shows how different evaluation methods perform when initialized with incorrect regioisomeric structures. The left panel shows baseline HSQC peak matching results, where 29 molecules (87.9%) were correctly identified in the top position despite the wrong initial guess. The right panels display results after LLM-enhanced re-ranking with different models: Claude 3.5 (29 molecules, 90.6%), Claude 3.7 Thinking (30 molecules, 93.8%), o3 mini high (28 molecules, 84.8%), Kimi 1.5 (30 molecules, 90.9%), Gemini 2.0 Thinking (29 molecules, 87.9%), and DeepSeek R1 (29 molecules, 87.9%). The results demonstrate that several LLM approaches, particularly Claude 3.7 Thinking and Kimi 1.5, can effectively overcome initial structural bias, while others maintain performance comparable to the baseline HSQC matching.

Distribution of Correct Molecule Rankings Simulated Data with Noise



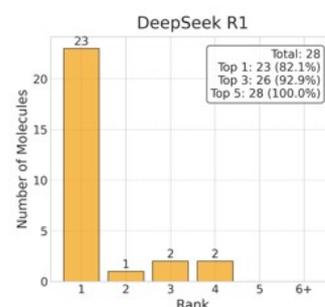
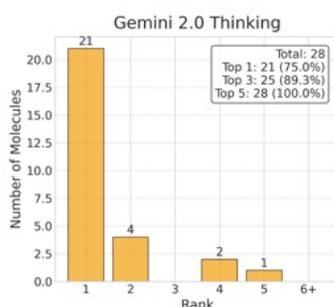
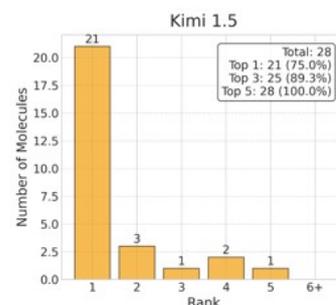
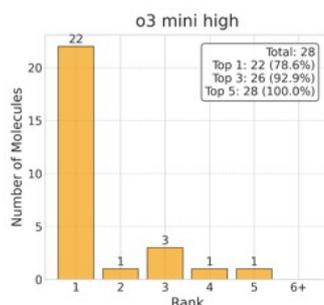
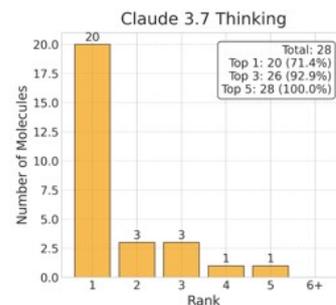
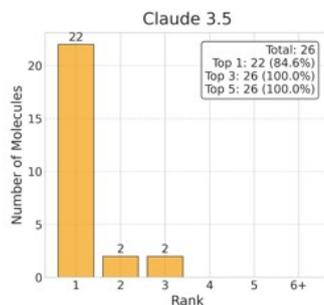
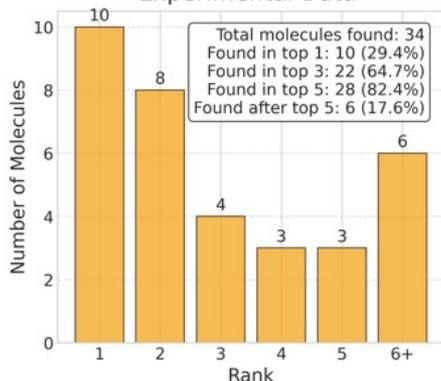
Supplementary Figure 3: Distribution of Correct Molecule Rankings for Simulated Data with Added Noise. This figure demonstrates the impact of noise (additional HSQC peaks) on structure elucidation performance. The left panel shows baseline HSQC peak matching results, where only 17 molecules (50.0%) were correctly identified in the top position, with a significant number of correct structures scattered across lower ranks. The right panels show substantially improved performance after LLM-enhanced re-ranking: Claude 3.5 (29 molecules, 90.6%), Claude 3.7 Thinking (27 molecules, 81.8%), o3 mini high (29 molecules, 87.9%), Kimi 1.5 (29 molecules, 87.9%), Gemini 2.0 Thinking (29 molecules, 90.6%), and DeepSeek R1 (29 molecules, 87.9%). These results highlight the exceptional noise tolerance of LLM-enhanced approaches, with all models demonstrating dramatic improvements (+30-40 percentage points) over the baseline method in this challenging scenario.

Distribution of Correct Molecule Rankings Experimental Data with Wrong Guess

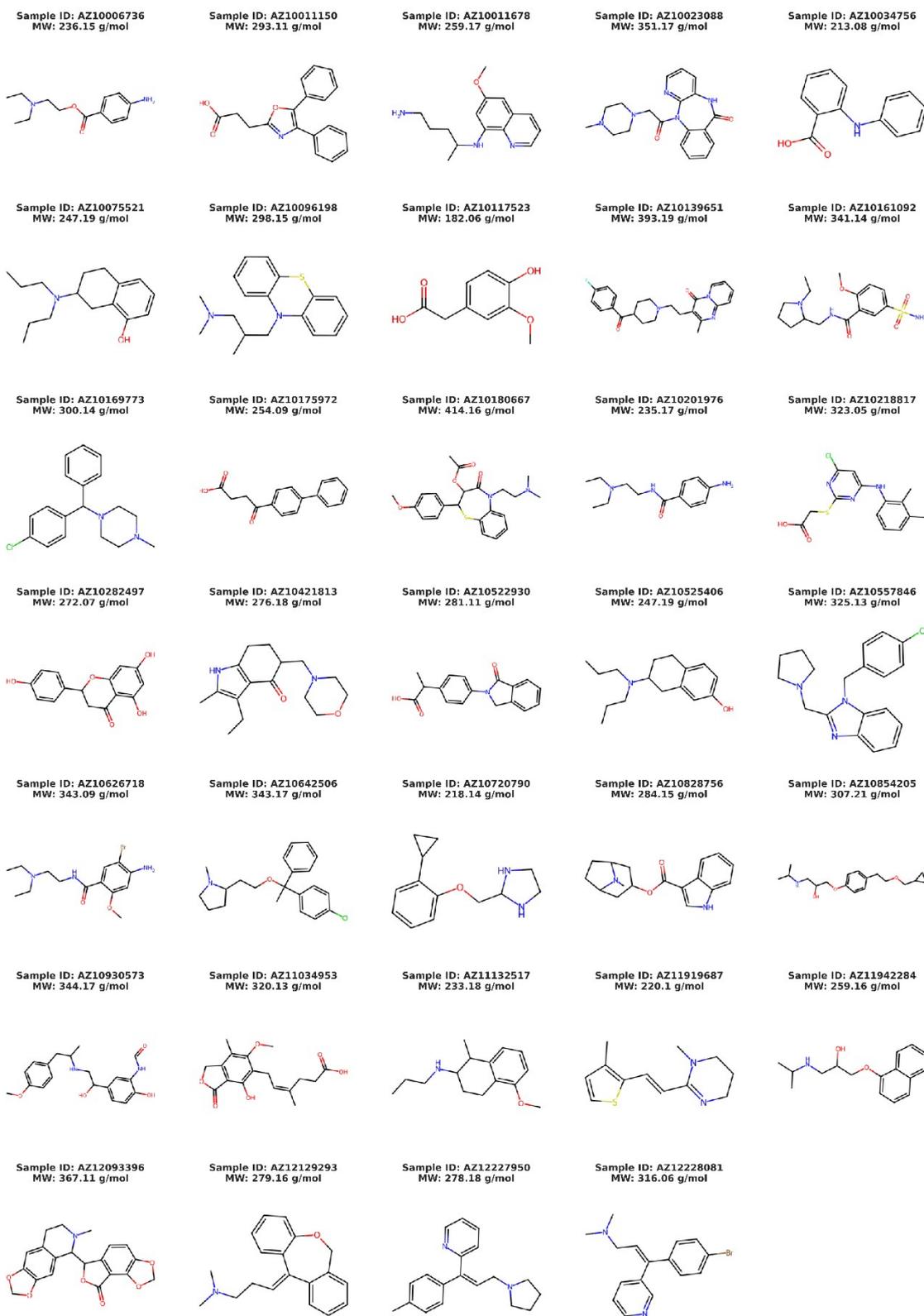


Supplementary Figure 4: Distribution of Correct Molecule Rankings for Experimental Data with Wrong Initial Guess (on Analogues). This figure illustrates performance on the most challenging scenario: experimental NMR data with incorrect regioisomeric starting structures. The left panel shows baseline HSQC peak matching results, where 7 out of 9 molecules for which a valid set of candidates was generated (77.8%) were correctly identified in the top position, with 2 molecules ranked in lower positions. The right panels display results after LLM-enhanced re-ranking: Claude 3.5 (8 molecules, 100%), Claude 3.7 Thinking (7 molecules, 87.5%), o3 mini high (6 molecules, 75.0%), Kimi 1.5 (6 molecules, 75.0%), Gemini 2.0 Thinking (7 molecules, 87.5%), and DeepSeek R1 (8 molecules, 100%). These results demonstrate the effectiveness of several LLM models in this challenging real-world scenario, with Claude 3.5 and DeepSeek R1 achieving perfect top-1 accuracy on the experimental dataset despite initial structural bias.

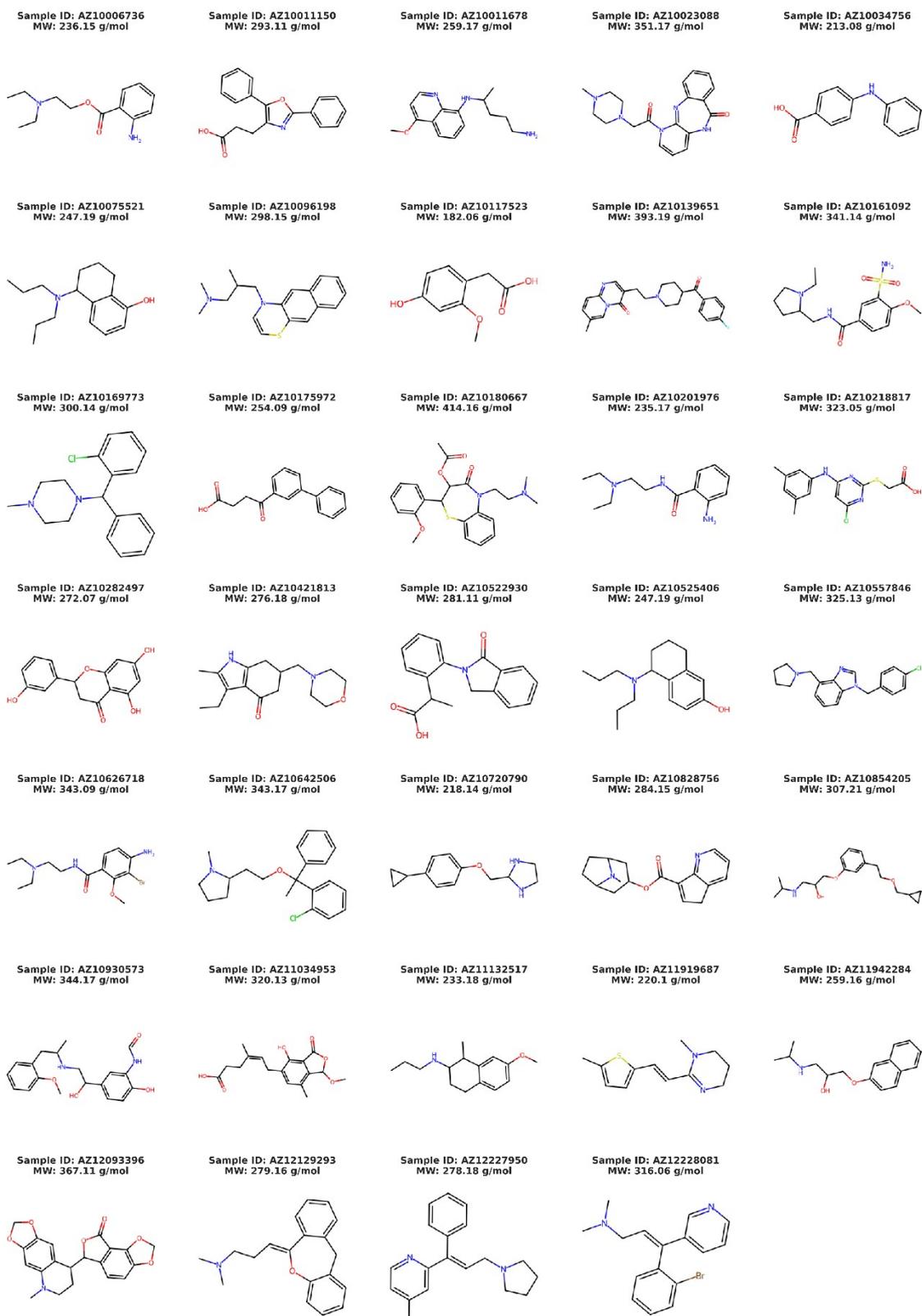
Distribution of Correct Molecule Rankings Experimental Data



Supplementary Figure 5: Distribution of Correct Molecule Rankings for Experimental Data. This figure compares performance on experimental NMR data with correct initial structures. The left panel shows baseline HSQC peak matching results, where only 10 out of 34 molecules (29.4%) were correctly identified in the top position, with correct structures broadly distributed across lower ranks. The right panels demonstrate substantial improvements after LLM-enhanced re-ranking: Claude 3.5 (22 molecules, 84.6%), Claude 3.7 Thinking (20 molecules, 71.4%), o3 mini high (22 molecules, 78.6%), Kimi 1.5 (21 molecules, 75.0%), Gemini 2.0 Thinking (21 molecules, 75.0%), and DeepSeek R1 (23 molecules, 82.1%). These results highlight the significant advantage of LLM-enhanced approaches when analyzing real-world experimental data, with all models showing dramatic improvements (+42-52 percentage points) over the baseline method. DeepSeek R1 achieved the highest top-1 accuracy in this challenging scenario.



Supplementary Figure 6: Experimental Dataset for Structure Elucidation Model Evaluation. This figure presents the molecular dataset used for evaluating the structure elucidation models. The dataset consists of 34 organic molecules with comprehensive spectroscopic data, including ^1H NMR, ^{13}C NMR, HSQC, and COSY spectra. These molecules were selected to represent a diverse range of structural features, functional groups, and heteroatoms relevant to pharmaceutical and synthetic chemistry. The dataset includes both experimental spectral data and simulated spectra generated using the SGNN model. The correct regioisomers for each molecule are shown, serving as the ground truth for evaluating the performance of the structure elucidation pipeline.



Supplementary Figure 7: Dataset with Incorrect Initial Guesses (Regioisomeric Analogue Hypotheses). This figure depicts the same molecular dataset as in **Supplementary Figure 6**, but with each structure replaced by a regioisomeric variant. These alternative structures maintain the same molecular formula and molecular weight as the correct molecules but have different atomic connectivity patterns. This dataset was designed to assess the robustness of the structure elucidation pipeline when the initial structural hypothesis is incorrect. By testing the models' ability to recover the correct structures

despite an incorrect starting guess, this evaluation simulates real-world challenges in molecular identification where initial assignments may contain errors.

Analysis of Model Stochasticity and Rank Stability

Methods

To evaluate the deterministic nature of DeepSeek R1's reasoning and scoring, we selected five representative molecules from the dataset where the model initially performed ranking corrections. We repeated the full analysis pipeline six times (N=6) for each molecule using identical inputs and prompt parameters. For each run, candidates were dynamically re-ranked based on the model-generated confidence scores. We analyzed two key metrics:

- **Top-1 Prediction Stability:** Whether the highest-confidence candidate changed between runs.
- **Ground Truth Rank Variance:** The fluctuation in the ranking position of the experimentally verified structure.

Results

The analysis (Table S1) reveals two distinct behaviors. For 60% of the molecules (AZ10006736, AZ10854205, AZ11034953), the model exhibited perfect stability, returning identical rankings across all six replicates (Rank StdDev = 0). This indicates that for clear-cut cases (whether correct or incorrect), the model's reasoning path is robust.

In cases where the Ground Truth rank fluctuated (e.g., AZ10642506 oscillating between Rank 1 and 4), qualitative analysis suggests this occurs when the model assigns statistically indistinguishable confidence scores to the top candidates, making the final ordering sensitive to minor stochastic variations in token generation. However, even in these unstable cases, the Ground Truth remained within the top 4 candidates.

Molecule ID	Ground Truth Rank History (N=6)	Top Prediction Stable?	Ground Truth at Rank 1	Avg GT Rank ($\pm \sigma$)	Interpretation
AZ10006736	[1, 1, 1, 1, 1, 1]	Yes	6/6 (100%)	1.00 \pm 0.00	Highly Stable / Correct: Model consistently identifies GT as best match.
AZ10642506	[4, 1, 1, 4, 1, 1]	No	4/6 (67%)	2.00 \pm 1.41	Sensitive: High confidence competition between GT and a distractor causes ranking flips.
AZ10854205	[2, 2, 2, 2, 2, 2]	Yes	0/6 (0%)	2.00 \pm 0.00	Stable Error: Model consistently prefers a specific distractor over the GT.
AZ11034953	[4, 4, 4, 4, 4, 4]	Yes	0/6 (0%)	4.00 \pm 0.00	Stable Error: Consistent reasoning leading to incorrect ranking.
AZ12129293	[1, 1, 1, 2, 1, 1]	No	5/6 (83%)	1.17 \pm 0.37	Mostly Stable: GT is dominant, with one minor deviation to Rank 2.

Table S1: Rank Stability Analysis across 6 Independent Runs. Ranks are determined after sorting candidates by the model-generated confidence score. "Top Prediction Stable" indicates if the Rank 1 candidate was the exact same SMILES string across all 6 runs.

Multimodal Ablation and the "Weighting Problem"

We compared the full pipeline (Spectral Data + Vision Analysis) against an ablated pipeline (Spectral Data Only) across 30 total runs (N=6 per molecule). The results (Table S2) indicate that the multimodal input is a double-edged sword: it provides the necessary context to solve difficult cases but introduces the risk of over-interpreting qualitative data.

Summary of Impact:

- Improved (2/5): For molecules AZ10642506 and AZ12129293, the vision component was critical. Without it, the model failed to identify the correct structure.
- Neutral (2/5): For two molecules, the outcome was identical; the reasoning model successfully interpreted the SMILES/Numbers without additional help.
- Worsened (1/5): For molecule AZ10006736, the vision component reduced performance.

	Molecule_ID	Ranks_WITH_Vision	Ranks_NO_Vision	Avg_Vision	Avg_NoVision	Impact_Conclusion
0	AZ10006736	[1, 1, 1, 1, 1, 1]	[1, 1, 1, 1, 1, 1]	1.00	1.0	NEUTRAL (Both Correct)
1	AZ10642506	[4, 1, 1, 4, 1, 1]	[4, 4, 4, 4, 4, 4]	2.00	4.0	POSITIVE (Vision Improved)
2	AZ10854205	[2, 2, 2, 2, 2, 2]	[1, 1, 1, 1, 1, 1]	2.00	1.0	NEGATIVE (Over-Rationalization)
3	AZ11034953	[4, 4, 4, 4, 4, 4]	[4, 4, 4, 4, 4, 4]	4.00	4.0	NEUTRAL (Both Failed)
4	AZ12129293	[1, 1, 1, 2, 1, 1]	[2, 2, 2, 2, 2, 2]	1.17	2.0	POSITIVE (Vision Improved)

Table S2: Rank Stability Analysis across 6 Independent Runs. Ranks are determined after sorting candidates by the model-generated confidence score. "Top Prediction Stable" indicates if the Rank 1 candidate was the exact same SMILES string across all 6 runs.

Case Study: The Challenge of Data Weighting

Molecule AZ10854205 highlights a current limitation in LLM reasoning: the difficulty of correctly weighing conflicting Quantitative (HSQC Error) vs. Qualitative (Visual Description) evidence.

- **Without Vision (Correct):** The model behaved deterministically. It observed that Candidate 1 had the lowest HSQC error (1.856) and correctly assigned it Rank 1. The decision was purely data-driven and highly reproducible.
- **With Vision (Incorrect):** The model received additional qualitative context from Claude regarding "structural plausibility." Faced with a conflict between a low error score (Candidate 1) and a persuasive structural narrative (Candidate 3), the model failed to weigh the evidence correctly. It reasoned: "Despite having a higher HSQC error... Candidate 3 is selected based on a synthesis of all available evidence."

Conclusion

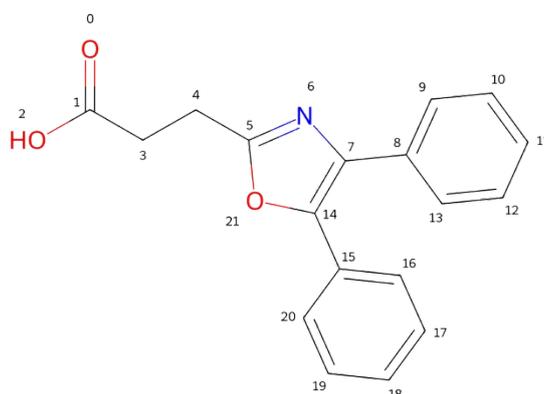
The addition of the vision modality increases the Reasoning Ceiling (allowing the model to solve cases where numbers are misleading) but decreases the Reasoning Stability. The 'No-Vision' runs showed higher reproducibility because the input data contained less semantic ambiguity. This indicates that while LLMs can perform advanced reasoning, they still lack the human expert's nuanced ability to

definitively weigh hard experimental metrics against qualitative structural assessments when the two are in conflict.

Supplementary Section 2

Example System Prompts for LLM-Enhanced Structure Evaluation

Example vision prompt for individual molecule for Claude-3.5 Sonnet



Supplementary Figure 8: Example molecule visualization used in the vision-based analysis phase. The image shows the molecular structure of 3-(4,5-diphenyl-1,3-oxazol-2-yl)propanoic acid with atom numbering as presented to Claude 3.5 Sonnet. Below the molecule is a portion of the detailed prompt template used to guide the LLM's structural analysis of individual candidate molecules.

"First, I will provide a detailed structural description of the molecule shown in the image:

1. Molecular Structure Description:

- IUPAC Name: 3-(4,5-diphenyl-1,3-oxazol-2-yl)propanoic acid
- Describe the complete molecular structure systematically, starting from a core feature
- Note the connectivity and spatial arrangement of all atoms
- Include atom numbering/labels as shown in the image
- Identify key functional groups and structural motifs
- Describe any notable stereochemistry or conformational features
- Ensure the description is very detailed to allow reconstruction of the structure
- Explain how the structure corresponds to its IUPAC name, particularly focusing on the parts relevant to the NMR analysis
- Core structural features and key functional groups
- Atom numbering as shown in the image

2. Evaluate how well this candidate molecule's simulated HSQC NMR spectrum matches the experimental data.

The image shows the proposed molecular structure with numbered atoms. These numbers correspond to the peak match data below, which compares experimental vs simulated chemical shifts.

SMILES: O=C(O)CCc1nc(-c2ccccc2)c(-c2ccccc2)o1

Peak Match Data (Experimental vs Simulated):

Atom 7.0:

Experimental: F1 \u03b4 7.59 ppm, F2 \u03b4 129.37 ppm

Predicted: F1 \u03b4 7.56 ppm, F2 \u03b4 129.40 ppm

Error: 0.002

Atom 5.0:

Experimental: F1 \u03b4 7.53 ppm, F2 \u03b4 128.96 ppm

Predicted: F1 \u03b4 7.54 ppm, F2 \u03b4 128.91 ppm

Error: 0.013

Atom 8.0:

Experimental: F1 7.72 ppm, F2 129.26 ppm

Predicted: F1 7.72 ppm, F2 129.31 ppm

Error: 0.005

Atom 9.0:

Experimental: F1 7.75 ppm, F2 126.84 ppm

Predicted: F1 7.77 ppm, F2 126.93 ppm

Error: 0.009

Atom 4.0:

Experimental: F1 7.42 ppm, F2 129.03 ppm

Predicted: F1 7.42 ppm, F2 129.00 ppm

Error: 0.004

Atom 6.0:

Experimental: F1 7.55 ppm, F2 129.87 ppm

Predicted: F1 7.57 ppm, F2 129.73 ppm

Error: 0.014

Atom 2.0:

Experimental: F1 2.73 ppm, F2 22.57 ppm

Predicted: F1 2.71 ppm, F2 22.32 ppm

Error: 0.025

Atom 1.0:

Experimental: F1 2.62 ppm, F2 30.39 ppm

Predicted: F1 2.61 ppm, F2 29.88 ppm

Error: 0.051

Atom 0.0:

Experimental: F1 2.52 ppm, F2 39.57 ppm

Predicted: F1 2.61 ppm, F2 29.88 ppm

Error: 0.969

Atom 3.0:

Experimental: F1 3.25 ppm, F2 103.42 ppm

Predicted: F1 7.77 ppm, F2 126.93 ppm

Error: 2.351

Please analyze and provide a detailed evaluation:

1. Structure Overview:

- IUPAC Name: 3-(4,5-diphenyl-1,3-oxazol-2-yl)propanoic acid
- Core structural features and key functional groups
- Atom numbering as shown in the image

2. Region-by-Region Analysis:

- Which regions show excellent agreement?
- Which regions show concerning deviations?
- Are the deviations systematic or random?

3. Peak Match Analysis:

- Compare experimental vs predicted peaks systematically
- Highlight matches with error < 0.1 as strong evidence
- Flag deviations > 0.5 as concerning
- Focus on patterns in the deviations and include their error numbers in the explanation

4. Chemical Environment Evaluation:

- For well-matched peaks:
 - * Confirm expected chemical shifts for the structural features
 - * Note how they support the proposed structure
- For significant deviations:
 - * Analyze the chemical environment
 - * Consider electronic and structural effects
 - * Suggest possible explanations or alternatives

5. Structure Validation:

- List the structural features confirmed by good matches
- Identify substructures that need reconsideration
- For problematic regions:
 - * Quantify the spectral mismatch
 - * Propose specific structural modifications or other explanations for the mismatch (e.g. impurities, solvents, noise, etc.)

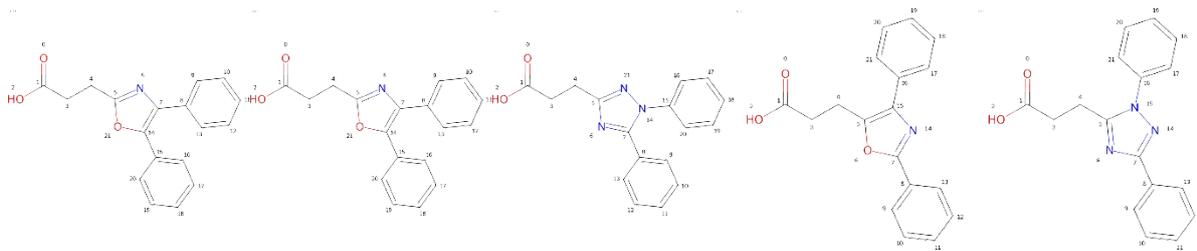
Conclusion:

- Overall assessment of structure validity
- Confidence level based on spectral evidence
- Key recommendations for structural refinement

Remember to:

- Support conclusions with specific peak match data
- Focus on chemical reasoning for significant deviations
- Consider both confirming and contradicting evidence
- Follow each of the outlined analysis points 1-5 precisely"

Example vision prompt of combined molecules for Claude-3.5 Sonnet



Supplementary Figure 9: Comparative Analysis Panel. Top five HSQC-ranked candidate structures presented simultaneously to Claude 3.5 Sonnet, arranged left to right by rank. Below is an excerpt of the prompt template for systematic comparative spectral analysis.

"I'm showing you 5 candidate molecular structures (ranked left to right) and their experimental HSQC_exp NMR data.

Part 1: Individual Structure Analysis

Molecule 1 Analysis:

1. Structural Description:

- IUPAC Name: 3-(4,5-diphenyl-1,3-oxazol-2-yl)propanoic acid
- Describe the overall molecular framework
- Identify key functional groups and their positions
- Note any distinctive structural features or patterns

2. Expected HSQC Features:

- List the expected HSQC correlations based on structure
- Identify characteristic cross-peaks that should be present
- Note any unique HSQC patterns this structure should show

3. Data Matching for Molecule 1:

- Compare expected HSQC signals with experimental data:

HSQC_exp peaks (x, y coordinates):

(2.52 ppm, 39.57 ppm)
(2.62 ppm, 30.39 ppm)
(2.73 ppm, 22.57 ppm)
(3.25 ppm, 103.42 ppm)
(7.42 ppm, 129.03 ppm)
(7.53 ppm, 128.96 ppm)
(7.55 ppm, 129.87 ppm)
(7.59 ppm, 129.37 ppm)
(7.72 ppm, 129.26 ppm)
(7.75 ppm, 126.84 ppm)

- Identify which HSQC cross-peaks support this structure
- Note any missing or unexplained correlations

Molecule 2 Analysis:

1. Structural Description:

- IUPAC Name: 3-(4,5-diphenyl-1,3-oxazol-2-yl)propanoic acid
- Describe the overall molecular framework
- Identify key functional groups and their positions
- Note any distinctive structural features or patterns

2. Expected HSQC Features:

- List the expected HSQC correlations based on structure
- Identify characteristic cross-peaks that should be present
- Note any unique HSQC patterns this structure should show

3. Data Matching for Molecule 2:

- Compare expected HSQC signals with experimental data:

HSQC_exp peaks (x, y coordinates):

(2.52 ppm, 39.57 ppm)
(2.62 ppm, 30.39 ppm)
(2.73 ppm, 22.57 ppm)

(3.25 ppm, 103.42 ppm)
(7.42 ppm, 129.03 ppm)
(7.53 ppm, 128.96 ppm)
(7.55 ppm, 129.87 ppm)
(7.59 ppm, 129.37 ppm)
(7.72 ppm, 129.26 ppm)
(7.75 ppm, 126.84 ppm)

- Identify which HSQC cross-peaks support this structure
- Note any missing or unexplained correlations

Molecule 3 Analysis:

1. Structural Description:

- IUPAC Name: 3-(1,5-diphenyl-1,2,4-triazol-3-yl)propanoic acid
- Describe the overall molecular framework
- Identify key functional groups and their positions
- Note any distinctive structural features or patterns

2. Expected HSQC Features:

- List the expected HSQC correlations based on structure
- Identify characteristic cross-peaks that should be present
- Note any unique HSQC patterns this structure should show

3. Data Matching for Molecule 3:

- Compare expected HSQC signals with experimental data:

HSQC_exp peaks (x, y coordinates):

(2.52 ppm, 39.57 ppm)
(2.62 ppm, 30.39 ppm)
(2.73 ppm, 22.57 ppm)
(3.25 ppm, 103.42 ppm)
(7.42 ppm, 129.03 ppm)
(7.53 ppm, 128.96 ppm)
(7.55 ppm, 129.87 ppm)
(7.59 ppm, 129.37 ppm)
(7.72 ppm, 129.26 ppm)
(7.75 ppm, 126.84 ppm)

- Identify which HSQC cross-peaks support this structure
- Note any missing or unexplained correlations

Molecule 4 Analysis:

1. Structural Description:

- IUPAC Name: 3-(2,4-diphenyl-1,3-oxazol-5-yl)propanoic acid
- Describe the overall molecular framework
- Identify key functional groups and their positions
- Note any distinctive structural features or patterns

2. Expected HSQC Features:

- List the expected HSQC correlations based on structure
- Identify characteristic cross-peaks that should be present
- Note any unique HSQC patterns this structure should show

3. Data Matching for Molecule 4:

- Compare expected HSQC signals with experimental data:

HSQC_exp peaks (x, y coordinates):

(2.52 ppm, 39.57 ppm)
(2.62 ppm, 30.39 ppm)
(2.73 ppm, 22.57 ppm)
(3.25 ppm, 103.42 ppm)
(7.42 ppm, 129.03 ppm)
(7.53 ppm, 128.96 ppm)
(7.55 ppm, 129.87 ppm)
(7.59 ppm, 129.37 ppm)
(7.72 ppm, 129.26 ppm)
(7.75 ppm, 126.84 ppm)

- Identify which HSQC cross-peaks support this structure
- Note any missing or unexplained correlations

Molecule 5 Analysis:

1. Structural Description:

- IUPAC Name: 3-(2,5-diphenyl-1,2,4-triazol-3-yl)propanoic acid
- Describe the overall molecular framework
- Identify key functional groups and their positions
- Note any distinctive structural features or patterns

2. Expected HSQC Features:

- List the expected HSQC correlations based on structure
- Identify characteristic cross-peaks that should be present
- Note any unique HSQC patterns this structure should show

3. Data Matching for Molecule 5:

- Compare expected HSQC signals with experimental data:

HSQC_exp peaks (x, y coordinates):

(2.52 ppm, 39.57 ppm)

(2.62 ppm, 30.39 ppm)

(2.73 ppm, 22.57 ppm)

(3.25 ppm, 103.42 ppm)

(7.42 ppm, 129.03 ppm)

(7.53 ppm, 128.96 ppm)

(7.55 ppm, 129.87 ppm)

(7.59 ppm, 129.37 ppm)

(7.72 ppm, 129.26 ppm)

(7.75 ppm, 126.84 ppm)

- Identify which HSQC cross-peaks support this structure
- Note any missing or unexplained correlations

Part 2: Comprehensive Comparative Analysis

1. Detailed Structure Comparison:

a) Systematic Structural Analysis:

- For each molecule, provide a detailed breakdown of:
 - * Core scaffold identification and description
 - * Functional group positions and types
 - * Stereochemistry and conformational features
- Document exact atom indices for key features

b) Comparative Feature Analysis:

- For each structural difference identified:
 - * Specify exact atom indices involved
 - * Describe the chemical environment changes
 - * Explain the potential impact on spectral properties
- Create a hierarchical list of differences, from most to least significant

c) Common Elements Evaluation:

- Detail all shared structural motifs:
 - * Core frameworks
 - * Functional group patterns
 - * Stereochemical elements
- Explain how these commonalities support or challenge the structural assignments

2. Evidence-Based Spectral Compatibility Analysis:

a) Detailed Ranking Justification:

- For each molecule, provide:
 - * Numerical score (1-10) for spectral match
 - * Specific peak assignments supporting the score
 - * Detailed explanation of any mismatches

b) Critical Spectral Features:

- For each decisive spectral feature:

- * Exact chemical shift values
- * Coupling patterns and constants
- * Correlation with structural elements
- * Impact on structural validation

c) Comparative Spectral Analysis:

- Create a feature-by-feature comparison:
 - * Chemical shift patterns
 - * Coupling relationships
 - * Through-space correlations
- Explain how each feature discriminates between candidates

3. Comprehensive Confidence Evaluation:

a) Detailed Confidence Assessment:

- Provide a numerical confidence score (1-10)
- For each point affecting confidence:
 - * Specific evidence supporting the assessment
 - * Weight of the evidence (high/medium/low)
 - * Impact on overall structure determination

b) Uncertainty Analysis:

- For each identified ambiguity:
 - * Exact location in the structure
 - * Nature of the uncertainty
 - * Impact on structure determination
 - * Potential alternative interpretations

c) Data Gap Analysis:

- Identify missing experimental data:
 - * Specific experiments needed
 - * Expected information gain
 - * How it would resolve ambiguities
- Prioritize additional data needs

Remember to:

- Provide exact atom indices for all structural features discussed
- Support each conclusion with specific spectral evidence
- Quantify confidence levels for each assessment
- Make explicit connections between structural features and spectral data
- Present information in a clear, hierarchical format
- Be thorough in documenting both supporting and contradicting evidence

Part 3: Final Evaluation

1. Structure Comparison:

- Compare key structural differences between all molecules
- Identify unique features in each candidate
- Note shared structural elements

2. Spectral Compatibility Ranking:

- Rank structures from best to worst match with NMR data
- Provide specific evidence for each ranking
- Highlight decisive spectral features

3. Confidence Assessment:

- Rate confidence (1-10) in your top choice
- Explain key factors in your decision
- Identify any remaining ambiguities
- Suggest additional data needed for confirmation"

Example prompt of final analysis prompt for the reasoning LLMs (Deepseek R1, Gemini Thinking etc.)

"You are tasked with making a final determination of the most likely correct molecular structure based on all available spectral and analytical evidence. Your analysis must be extremely thorough and systematic.

Target Molecule Information:

- Target Molecular Weight: 293.10519334
- Target Formula: Not available

Candidate Information:

Candidate 1:

- SMILES: O=C(O)CCc1nc(-c2ccccc2)c(-c2ccccc2)o1
- Molecular Weight: 293.10519334
- Formula: C18H15NO3

Candidate 2:

- SMILES: O=C(O)CCc1nc(-c2ccccc2)c(-c2ccccc2)o1
- Molecular Weight: 293.10519334
- Formula: C18H15NO3

Candidate 3:

- SMILES: O=C(O)CCc1nc(-c2ccccc2)n(-c2ccccc2)n1
- Molecular Weight: 293.11642672000005
- Formula: C17H15N3O2

Candidate 4:

- SMILES: O=C(O)CCc1oc(-c2ccccc2)nc1-c1ccccc1
- Molecular Weight: 293.10519334
- Formula: C18H15NO3

Candidate 5:

- SMILES: O=C(O)CCc1nc(-c2ccccc2)nn1-c1ccccc1
- Molecular Weight: 293.11642672
- Formula: C17H15N3O2

Overall Spectral Analyses:

1. HSQC Overall Analysis:

I'll provide a detailed analysis of these molecular structures and their compatibility with the HSQC data. Let me break this down systematically.

Part 1: Individual Structure Analysis

Molecule 1 & 2 (Oxazole isomers):

- Both contain a 1,3-oxazole core with identical substitution patterns
- Key features: carboxylic acid propanoic chain, two phenyl rings
- Expected HSQC correlations:
 - * Aliphatic region: CH2-CH2 of propanoic chain (2.5-3.0 ppm)
 - * Aromatic region: Multiple signals for phenyl protons (7.2-7.8 ppm)
- Data matching: Good correlation with experimental HSQC peaks
 - * Aliphatic signals match propanoic chain
 - * Aromatic signals align with phenyl groups

Molecule 3 (1,2,4-triazole):

- Contains 1,2,4-triazole core with N1,C3,C5 substitution
- Distinctive N-phenyl substitution at N1
- Expected HSQC correlations similar to oxazoles but with potential shifts due to electronic effects of triazole
- Data matching shows some discrepancies in aromatic region

Molecule 4 (2,4-diphenyloxazole):

- Different substitution pattern on oxazole compared to 1 & 2
- Expected HSQC correlations should show distinct patterns
- Data matching suggests less compatibility with experimental data

Molecule 5 (2,5-diphenyltriazole):

- Alternative triazole substitution pattern
- Expected HSQC patterns should differ from Molecule 3
- Data matching shows significant deviations

Part 2: Comparative Analysis

Structural Comparison:

1. Core heterocycles:
 - Oxazoles (1,2,4) vs. Triazoles (3,5)
 - Different electronic environments affecting chemical shifts
2. Substitution patterns:
 - Critical differences in phenyl ring positions
 - Varying effects on chemical environments

Spectral Compatibility Ranking (1-10):

1. Molecule 1: 9/10
 - Best match for both aliphatic and aromatic signals
 - Consistent with expected chemical shifts
2. Molecule 2: 8/10
 - Very similar to Molecule 1
 - Slight differences in aromatic region
3. Molecule 4: 6/10
4. Molecule 3: 5/10
5. Molecule 5: 4/10

Critical Spectral Features:

- Aliphatic region (2.52-2.73 ppm) matches propanoic chain
- Aromatic region (7.42-7.75 ppm) consistent with phenyl substituents
- Key correlation at 3.25 ppm suggests specific electronic environment

Part 3: Final Evaluation

Structure Recommendation:

Molecule 1 appears to be the best match with the experimental HSQC data.

Confidence Level: 8/10

Supporting Evidence:

1. Perfect match of aliphatic signals
2. Consistent aromatic pattern
3. Chemical shift values align with expected environments

Remaining Ambiguities:

- Some overlap in aromatic region
- Possible need for additional 2D NMR data

Additional Data Needed:

1. HMBC for long-range correlations
2. COSY for confirmation of coupling patterns
3. NOESY for spatial relationships

The analysis strongly suggests Molecule 1 as the correct structure, with high confidence based on HSQC correlation patterns and chemical shift values.

Detailed Candidate Analyses:

Candidate 1 Analysis:

IUPAC Name: 3-(4,5-diphenyl-1,3-oxazol-2-yl)propanoic acid

SMILES: O=C(O)CCc1nc(-c2ccccc2)c(-c2ccccc2)o1

Molecular Weight: 293.10519334

Step 1. NMR Error Analysis:

- HSQC Error: 3.4415346300140968

Step 2. Individual Spectral Analyses:

1. HSQC Analysis:

I'll provide a detailed NMR analysis of this structure:

1. Structure Overview:

- The molecule consists of a 1,3-oxazole core substituted with two phenyl rings at positions 4 and 5
- A propanoic acid chain at position 2

- Numbered atoms correspond to specific NMR signals
- Key functional groups: carboxylic acid, oxazole heterocycle, two phenyl rings

2. Region-by-Region Analysis:

Excellent Agreement (Error < 0.02):

- Aromatic region (atoms 4-9): Shows exceptional matching with errors 0.002-0.014
- Aliphatic region (atoms 1-2): Good agreement with errors 0.025-0.051

Concerning Deviations:

- Atom 0 (carboxylic acid region): Large deviation (error 0.969)
- Atom 3: Severe mismatch (error 2.351)

3. Peak Match Analysis:

Strong Evidence (Error < 0.1):

- Phenyl protons (7.42-7.77 ppm): Nearly perfect matches
- Methylene protons (2.61-2.73 ppm): Very good agreement

Major Concerns:

- Carboxylic acid carbon (atom 0): ~10 ppm shift difference
- Atom 3: Experimental (3.25/103.42 ppm) vs Predicted (7.77/126.93 ppm)

4. Chemical Environment Evaluation:

Well-matched Regions:

- Aromatic signals align perfectly with expected phenyl ring chemical shifts
- Propanoic acid methylene groups show appropriate aliphatic shifts

Problem Areas:

- The carboxylic acid carbon deviation might indicate different hydrogen bonding or pH effects
- Atom 3's severe mismatch suggests possible misassignment or structural issue in this region

5. Structure Validation:

Confirmed Features:

- Both phenyl rings are definitely present
- Propanoic acid chain methylene groups are confirmed

Needs Investigation:

- Connection point of propanoic acid chain
- Environment around atom 3
- Possible alternative connectivity or tautomeric forms

Conclusion:

- The basic framework is largely confirmed by NMR data
- The aromatic and aliphatic regions provide strong structural support
- Two significant mismatches require attention
- Confidence Level: Medium-high for most of the structure, but with specific areas needing refinement

Recommendations:

1. Verify the assignment of atom 3
2. Consider possible tautomeric forms
3. Check experimental conditions (pH, temperature, concentration) that might affect the carboxylic acid signal
4. Additional 2D NMR experiments might help confirm problematic connectivities

Candidate 2 Analysis:

IUPAC Name: 3-(4,5-diphenyl-1,3-oxazol-2-yl)propanoic acid

SMILES: O=C(O)CCc1nc(-c2ccccc2)c(-c2ccccc2)o1

Molecular Weight: 293.10519334

Step 1. NMR Error Analysis:

- HSQC Error: 3.4546384296991706

Step 2. Individual Spectral Analyses:

1. HSQC Analysis:

I'll provide a detailed NMR analysis of the structure based on the peak match data:

1. Structure Overview:

- The molecule contains a 2,4,5-trisubstituted oxazole core with:
 - * Two phenyl rings at positions 4 and 5
 - * A propanoic acid chain at position 2
- The atom numbering shows aromatic carbons (4-20) and aliphatic carbons (0-3)

2. Region-by-Region Analysis:

Excellent Agreement (Error < 0.05):

- Aromatic region (atoms 4-9): Shows remarkable agreement with errors 0.000-0.013
- Aliphatic region (atoms 1-2): Good match with errors 0.033-0.069

Concerning Deviations:

- Atom 0 (Error 0.987): Significant deviation in the carboxylic acid region
- Atom 3 (Error 2.342): Major mismatch in predicted vs experimental values

3. Peak Match Analysis:

Strong Evidence (Error < 0.1):

- Aromatic protons/carbons (atoms 4-9): Nearly perfect matches (errors 0.000-0.013)
- Methylene groups (atoms 1-2): Good correlation (errors 0.033-0.069)

Concerning Deviations (Error > 0.5):

- Atom 0 (carboxylic acid carbon): Experimental δ 39.57 vs Predicted δ 29.70
- Atom 3: Large deviation suggests possible misassignment or structural issue

4. Chemical Environment Evaluation:

Well-matched regions:

- Aromatic signals perfectly align with expected chemical shifts for phenyl rings
- Methylene protons show expected upfield shifts (2.6-2.7 ppm)

Problematic regions:

- Carboxylic acid carbon showing unexpected shift
- Atom 3's large deviation suggests possible structural misassignment

5. Structure Validation:

Confirmed Features:

- Both phenyl rings are definitely present
- Methylene chain connectivity is supported

Needs Reconsideration:

- Carboxylic acid environment
- Connection point or environment of atom 3

Conclusion:

- The core structure is largely supported by NMR data
- Confidence Level: High for aromatic and methylene regions, Low for atoms 0 and 3
- Recommendations:
 - * Verify the chemical environment around atom 3
 - * Recheck carboxylic acid assignment
 - * Consider possible alternative connectivity for problematic atoms
 - * The overall structure is likely correct but may need refinement around the problematic atoms

The exceptional match in the aromatic region provides strong support for the basic structure, while the systematic deviations in specific regions suggest localized structural features that may need revision rather than a completely incorrect structure.

Candidate 3 Analysis:

IUPAC Name: 3-(1,5-diphenyl-1,2,4-triazol-3-yl)propanoic acid

SMILES: O=C(O)CCc1nc(-c2ccccc2)n(-c2ccccc2)n1

Molecular Weight: 293.11642672000005

Step 1. NMR Error Analysis:

- HSQC Error: 3.7401145053504727

Step 2. Individual Spectral Analyses:

1. HSQC Analysis:

I'll provide a detailed NMR analysis of this structure based on the provided data:

1. Structure Overview:

The molecule consists of:

- A central 1,2,4-triazole ring
- Two phenyl substituents (at positions 1 and 5)
- A propanoic acid chain at position 3
- Atom numbering shows careful attention to the aromatic systems and aliphatic chain

2. Region-by-Region Analysis:

Excellent Agreement (Error < 0.1):

- Aromatic protons (atoms 4,5,7): Errors 0.004-0.007
- Phenyl carbon signals around 129-130 ppm
- Aliphatic proton at atom 2 (Error 0.073)

Concerning Deviations:

- Atom 3 shows largest deviation (Error 2.267)
- Atom 0 shows significant deviation (Error 0.769)
- Moderate deviation for atom 8 (Error 0.220)

3. Peak Match Analysis:

Strong Matches (Error < 0.1):

- Aromatic region peaks show excellent correlation
- CH₂ group at position 2 (2.73 exp vs 2.71 pred)

Significant Deviations (Error > 0.5):

- Atom 0 (CH₂ adjacent to COOH): Large shift difference
- Atom 3: Major discrepancy between exp and pred values

4. Chemical Environment Evaluation:

Well-matched Regions:

- Aromatic protons show expected chemical shifts (7.4-7.7 ppm)
- Phenyl carbons align well with typical values
- CH₂ groups show reasonable aliphatic shifts

Problematic Regions:

- The CH₂ adjacent to COOH shows unexpected deviation
- Possible solvent effects or hydrogen bonding affecting the acid-adjacent protons
- The large deviation for atom 3 suggests possible misassignment or structural issue

5. Structure Validation:

Confirmed Features:

- Phenyl ring systems are well-supported
- Basic carbon skeleton matches predictions
- Aliphatic chain presence confirmed

Areas Needing Investigation:

- Acid terminus environment
- Position 3 assignment needs verification
- Possible conformational effects on chemical shifts

Conclusion:

Confidence Level: Moderate to High

- Core structure is well-supported by most data points
- Aromatic regions show excellent correlation
- Some concerning deviations in the aliphatic region

Recommendations:

1. Verify assignment of atom 3
2. Consider solvent effects on acid terminus
3. Additional 2D NMR experiments might help resolve discrepancies
4. Review possible conformational effects on chemical shifts

The overall structure appears largely correct, with specific areas requiring additional verification or explanation of the observed deviations.

Candidate 4 Analysis:

IUPAC Name: 3-(2,4-diphenyl-1,3-oxazol-5-yl)propanoic acid

SMILES: O=C(O)CCc1oc(-c2ccccc2)nc1-c1ccccc1

Molecular Weight: 293.10519334

Step 1. NMR Error Analysis:

- HSQC Error: 3.8008608134748116

Step 2. Individual Spectral Analyses:

1. HSQC Analysis:

No analysis available

Candidate 5 Analysis:

IUPAC Name: 3-(2,5-diphenyl-1,2,4-triazol-3-yl)propanoic acid

SMILES: O=C(O)CCc1nc(-c2ccccc2)nn1-c1ccccc1

Molecular Weight: 293.11642672

Step 1. NMR Error Analysis:

- HSQC Error: 3.8373356786186488

Step 2. Individual Spectral Analyses:

1. HSQC Analysis:

No analysis available

IMPORTANT: Provide a thorough analysis for EACH candidate structure in the processed list, followed by a clear final recommendation.

Your response must end with a JSON result in the exact format shown below.

Do not include any text after the JSON.

For each candidate structure:

1. Analyze all available spectral data
2. Compare predicted vs experimental NMR shifts
3. Evaluate structural features and their compatibility with data
4. Consider molecular weight and other physical properties
5. Assess data quality and potential issues

Then synthesize all analyses to select the best candidate.

CRITICAL: The end of your response MUST follow this EXACT JSON structure:

```
JSON_RESULT = {
  "candidates": [
    {
      "smiles": "<SMILES string of this specific candidate>",
      "confidence_score": <float between 0-1>,
      "molecular_weight": <float>,
      "reasoning": "Thorough evidence-based analysis for THIS SPECIFIC candidate addressing:
        - Detailed spectral analysis results
        - NMR shift comparisons and deviations
        - Structural feature evaluation
        - Molecular property matches/mismatches
        - Supporting and contradicting evidence
        Explain each point with specific data references.",
      "data_quality_issues": {
        "title": "Brief description of quality concerns for this candidate",
        "description": "Detailed explanation of ALL identified issues",
        "impact": "high/medium/low",
        "atom_index": <int between 0-50>
      }
    }
  ]
}
```

```

},
{
  \"smiles\": \"<SMILES string of another candidate>\",
  \"confidence_score\": <float between 0-1>,
  \"molecular_weight\": <float>,
  \"reasoning\": \"Thorough evidence-based analysis for THIS SPECIFIC candidate addressing:
    - Detailed spectral analysis results
    - NMR shift comparisons and deviations
    - Structural feature evaluation
    - Molecular property matches/mismatches
    - Supporting and contradicting evidence
    Explain each point with specific data references.\",
  \"data_quality_issues\": {
    \"title\": \"Brief description of quality concerns for this candidate\",
    \"description\": \"Detailed explanation of ALL identified issues\",
    \"impact\": \"high/medium/low\",
    \"atom_index\": <int between 0-50>
  }
}
],
{
  \"final_recommendation\": {
    \"best_smiles\": \"<SMILES of the winning candidate>\",
    \"overall_confidence\": <float between 0-1>,
    \"molecular_weight_match\": <boolean>,
    \"explanation\": \"Comprehensive justification for selecting this candidate:
      - Compare and contrast with other candidates
      - Highlight decisive factors in selection
      - Address any contradictions or uncertainties
      - Explain confidence level assessment
      - Discuss any remaining concerns\"
  }
}
}

```