

# AutoSolvate-Agent: An Autonomous Agent for GPU-accelerated Solution-phase Quantum Chemistry

Xu Chen\*  
Emory University  
Atlanta, Georgia, USA  
xu.chen@emory.edu

Fangning Ren\*  
Emory University  
Atlanta, Georgia, USA  
fangning.ren@emory.edu

Dazhou Yu  
Emory University  
Atlanta, Georgia, USA  
dazhou.yu@emory.edu

Lechen Dong  
Emory University  
Atlanta, Georgia, USA  
lechen.dong@emory.edu

Fang Liu  
Emory University  
Atlanta, Georgia, USA  
fang.liu@emory.edu

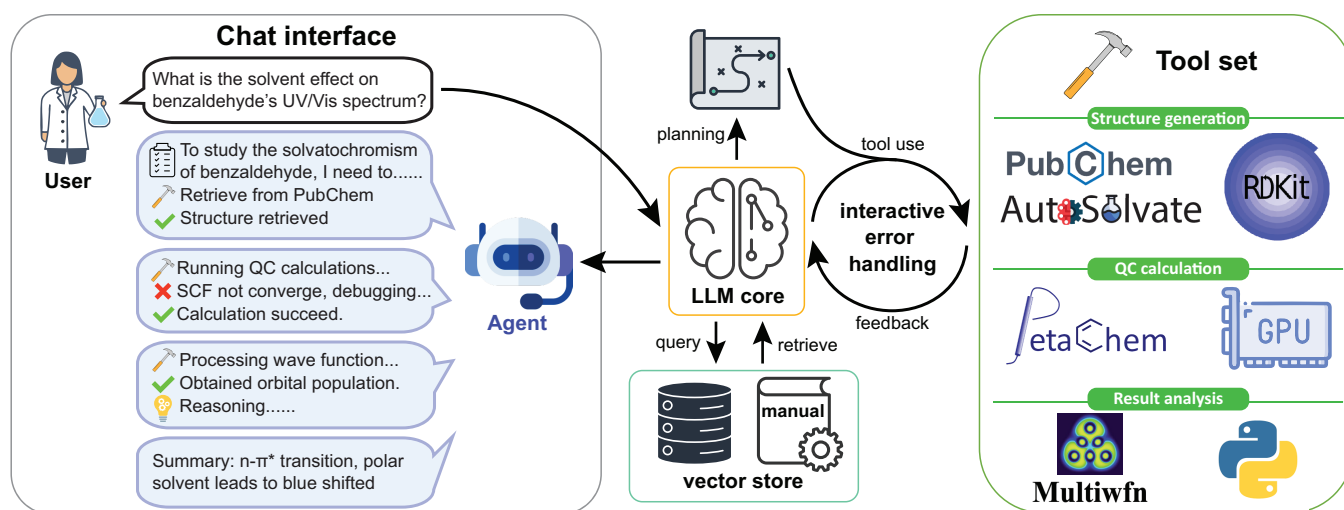


Figure 1: Schematic illustration of the agentic system for GPU-accelerated quantum chemistry calculation

## ABSTRACT

Many chemically relevant phenomena occur in solution or at complex interfaces, yet quantum chemical calculations of such systems involve intricate workflows and extensive parameter tuning, posing challenges for non-specialists and even experts. Interpreting the results of these simulations requires deep knowledge of computational chemistry, from model setup to post-processing. In this work, we introduce an autonomous agentic system specifically designed for solution-phase quantum chemistry. Built on an agentic architecture, the system supports flexible tool integration, resource management, I/O handling, and post-wavefunction analysis. Through

representative case studies, we show that this agent significantly reduces the complexity and expertise required to perform and interpret advanced simulations, making them more accessible to both theorists and experimentalists.

## CCS CONCEPTS

• **Applied computing** → **Chemistry**; *Physics*; • **Computing methodologies** → **Molecular simulation**; **Quantum mechanic simulation**; **Massively parallel and high-performance simulations**; *Natural language processing*; *Intelligent agents*.

## KEYWORDS

LLM-Based Agent, Quantum Chemistry, Computational Chemistry Automation, Density Functional Theory (DFT), Electronic Structure

### ACM Reference Format:

Xu Chen, Fangning Ren, Dazhou Yu, Lechen Dong, and Fang Liu. 2025. AutoSolvate-Agent: An Autonomous Agent for GPU-accelerated Solution-phase Quantum Chemistry. In *Proceedings of KDD'25 SciSoc LLM Workshop Large Language Models for Scientific and Societal Advances (KDD'25 SciSoc LLM Workshop)*. ACM, New York, NY, USA, 15 pages. <https://doi.org/XXXXXXX.XXXXXXX>

\*These authors contributed equally to this work.

## 1 INTRODUCTION

Quantum chemistry plays a central role in understanding molecular behavior, enabling insights into electronic structure, reactivity, and spectroscopy that are essential for fields ranging from drug design to materials science [4, 16]. With the advent of GPU-accelerated computing and increasingly accurate electronic structure methods, the computational capabilities of quantum chemistry have advanced rapidly [13, 17]. However, the practical utility of these tools remains largely confined to experts due to their steep learning curve and the complexity of setting up and interpreting calculations. Bridging this gap is crucial for democratizing access to high-accuracy molecular simulations in experimental and interdisciplinary research settings.

Several automated workflow systems have been developed to simplify quantum chemistry tasks such as geometry optimization and property prediction [7, 18]. These tools effectively reduce the user burden for routine applications, but are typically rigid and difficult to adapt to novel scientific questions. Tasks like excited-state modeling or Quantum Mechanics/Molecular Mechanics (QM/MM) simulations often require expert intervention to adjust parameters or recover from failures. More fundamentally, we identify three persistent challenges that limit the broader usability of quantum chemistry: (1) the complexity of protocol design, as selecting an appropriate model chemistry that balances physical accuracy and computational cost, along with configuring relevant options such as density fitting, convergence thresholds, or active space definitions, requires specialized expertise; (2) the lack of adaptive control, since most workflows cannot recover from convergence failures or dynamically adjust strategies based on intermediate outcomes; and (3) the opacity of computational outputs, as interpreting simulation results and diagnosing numerical issues remains difficult without a strong foundation in computational molecular science.

To address these challenges, we propose an **agentic system** powered by large language models (LLMs), capable of autonomously assisting users throughout the quantum chemistry workflow. To mitigate the complexity of protocol design, the system interprets high-level user intent and translates it into detailed, tool-specific inputs using embedded chemical knowledge. To overcome the rigidity of conventional workflows, the agent dynamically adapts to runtime feedback. This includes recovering from SCF failures, adjusting parameters, or switching methods when necessary. Finally, to improve the interpretability of results, the system offers interactive explanations, facilitates postprocessing, and integrates visualization and analysis tools, all through a natural language interface.

Our contributions are summarized as follows:

- **Agentic Framework for Quantum Chemistry:** We develop a domain-specialized agentic system built on large language models, designed to orchestrate complex quantum chemical workflows. This framework enables flexible task composition, decision making, and interaction with external simulation tools through natural language.
- **Protocol Reasoning and Generation Module:** We implement a reasoning engine that maps high-level user queries to executable simulation protocols, drawing on chemical knowledge embedded in the LLM and augmented with external knowledge bases. This empowers non-expert users to configure sophisticated simulations without requiring deep theoretical expertise.
- **Dynamic Workflow Control:** The agent supports adaptive execution by monitoring runtime states, detecting failure modes (e.g., SCF convergence issues), and autonomously modifying the workflow. This improves robustness in challenging scenarios and reduces the need for manual intervention.
- **Post-processing and Interpretation Interface:** We provide a natural language interface for interpreting results through both text and visualizations, enabling tasks such as population analysis, orbital visualization, and energy decomposition. This lowers the barrier to understanding quantum chemical output, especially for experimentalists and interdisciplinary researchers.
- **End-to-End Evaluation Across Diverse Tasks:** We validate our system on representative quantum chemistry applications, including spectral analysis and solvation modeling. The results demonstrate that our agent achieves expert-level outcomes while significantly reducing setup time and improving result accessibility.

In summary, our work aims to bridge the gap between powerful quantum chemical tools and real-world usability through an autonomous, LLM-driven agentic system. By embedding domain knowledge and enabling adaptive reasoning, we make complex simulations more accessible, interpretable, and robust for a wider range of users.

## 2 RELATED WORK

Recent advances in large-language models (LLMs) have spurred the development of intelligent agents for automating computational chemistry and molecular simulation workflows. Campbell et al. [2] proposed *MDCrow*, an LLM-based assistant that unifies molecular dynamics tasks, including tool selection, simulation setup, analysis, and literature retrieval, within a single interactive system. Zou et al. [20] introduced *El Agente Q*, a hierarchical multi-agent framework capable of autonomously planning and executing quantum chemistry tasks across geometry optimization, electronic structure calculations, and thermodynamic evaluations. In parallel, Hu et al. [8] developed *Aitomia*, a platform that combines LLMs, rule-based agents, and the MLatom ecosystem to streamline the setup and interpretation of atomistic simulations.

Several systems also focus on lowering the barrier for non-specialists through conversational interfaces. Liu et al. [6] presented *AutoSolvateWeb*, a chatbot-guided quantum chemistry platform that helps users perform multistep calculations involving explicit solvation models via natural language prompts and cloud infrastructure. Jacobs and Pollice [9] showed that prompt engineering, retrieval-augmented generation, and lightweight fine-tuning significantly improve LLM performance on input generation for simulation packages such as ORCA [15]. Yu et al. [19] proposed *Spatial-RAG*, a framework that connects LLMs with spatial databases to enable natural language-based geospatial reasoning and querying, allowing non-experts to interact with complex spatial data through tool-augmented agents.

Our work builds on these prior efforts by focusing specifically on quantum chemistry in complex environments. Unlike previous systems that target general-purpose automation or isolated task types, we introduce an LLM-based agentic system that integrates protocol generation, dynamic workflow control, and post-processing interpretation into a unified, autonomous assistant tailored to domain-specific simulation challenges.

### 3 METHODOLOGY

AutoSolvate-Agent is an autonomous assistant designed to execute and interpret complex quantum chemistry workflows through interactive, tool-augmented reasoning (Figure 1). At the center of AutoSolvate-Agent is an LLM-based reasoning core built on GPT [1] models from OpenAI, which orchestrates task planning, external tool invocation, and adaptive error handling. User queries, such as investigating solvent effects on UV/Vis spectra, are interpreted through multi-step reasoning that combines structured tool use and chemical knowledge. Leveraging the tool-calling framework from LangChain [3], the agent decomposes high-level tasks into subtasks spanning structure generation, quantum chemical computation, wave function analysis, and result summarization. To address the **complexity of protocol design**, the system translates natural language queries into valid computational workflows by retrieving relevant tool documentation from a domain-specific vector store and selecting appropriate methods and parameters (see Section 3.1). **Adaptive control** is achieved through interactive monitoring and real-time error recovery; for example, SCF convergence failures in quantum chemistry calculations trigger automatic input adjustment and re-execution (see Section 3.2). **Result interpretation** is achieved through integration with analysis tools such as Multiwfn [14] and RDKit [11], allowing the system to explain computed results and generate both textual and visual summaries (see Section 3.3). Together, these components enable AutoSolvate-Agent to support both end-to-end simulation pipelines and modular tasks (e.g., analyzing precomputed data), thereby enhancing usability, robustness, and accessibility for non-specialist users.

#### 3.1 Protocol Reasoning and Workflow Construction

To translate natural language queries into valid quantum chemistry workflows, AutoSolvate-Agent performs protocol reasoning and input generation using a tool-augmented LLM. When a user poses a query, the agent first identifies the task type and required molecular entities.

The agent then proceeds to acquire the desired object, e.g., input parameters needed for QM calculations. If a structure name/SMILES is given, the system retrieves relevant molecular information (e.g., charge, multiplicity) from PubChem [10]. If the molecular information is not accessible, the agent can leverage RDKit [11] to obtain the relevant information. For simulations involving explicit solvent effects consideration, the agent can also invoke the AutoSolvate package [7] to generate explicit solvent configurations.

Upon obtaining the basic molecular information, the appropriate QM calculation settings—such as method, basis set, and solvation model—are selected through context-aware reasoning, guided by

the TeraChem (a GPU-accelerated quantum chemistry engine)[17] Manual book [17] embedded in a vector store (via Chroma [5]) and retrieved using Retrieval-augmented generation (RAG) [12] technique. The LLM uses this knowledge to configure valid input files via validated tool wrappers.

#### 3.2 Adaptive Execution and Error Recovery

During quantum chemistry calculations, input errors and convergence issues are common, particularly in complex systems or solvent environments. To address this, AutoSolvate-Agent incorporates a feedback loop for runtime monitoring and recovery. Upon receiving a tool output, the agent checks for errors (e.g., SCF convergence failure). If encountered, the LLM parses diagnostic messages and modifies parameters—such as changing the SCF convergence threshold or updating the used method—before resubmitting the task. This enables automatic self-correction without requiring user intervention. Once successful, the agent extracts and reports key results, such as total energy, HOMO-LUMO gap, and optimized geometry.

#### 3.3 Result Analysis and Summarization

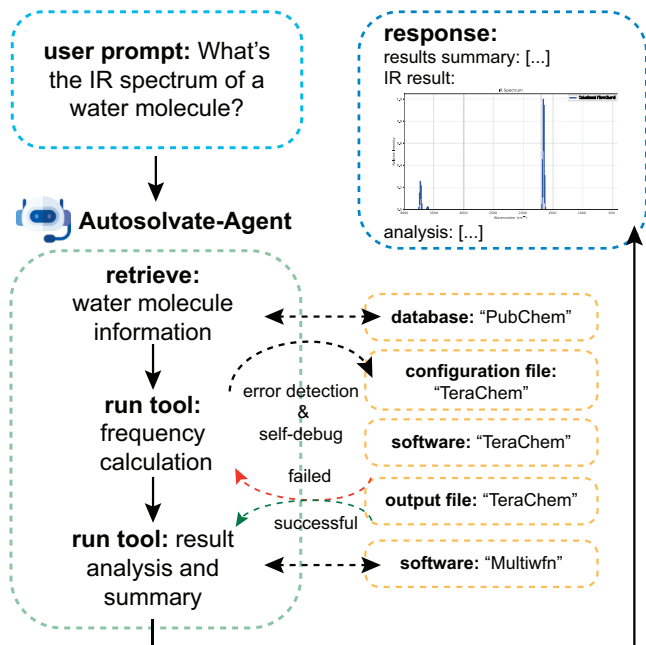
To support result interpretation and downstream reasoning, AutoSolvate-Agent integrates with Multiwfn [14] for wavefunction analysis. Multiwfn’s extensive functionality is accessible via LLM-based control, guided by embedded Multiwfn’s documentation indexed in the vector store. The agent issues analysis commands, parses output, and iteratively adjusts its plan if expected outputs are not obtained—retrieving additional documentation as needed.

For summarization, the system renders molecular structures as labeled images using RDKit, facilitating structure-aware interpretation. These images, together with numerical results, are passed to a vision-language model (VLM) such as GPT-4.1-mini [1] to produce a unified, natural language report. The final output combines chemical reasoning, quantitative results, and visual representations, improving accessibility for users with limited quantum chemistry expertise.

### 4 EXPERIMENTAL RESULTS

In the following sections, we demonstrate the capabilities of our autonomous agent system in performing quantum chemistry tasks based on the user’s request. Each task requires the agent to complete a series of interdependent sub-tasks—ranging from molecular setup and input file generation to job execution and post-analysis. To evaluate performance, we tested 10 common computational chemistry tasks, assessing the probability of producing a correct analysis across 10 independent runs for each. The correctness of each analysis is judged by QC domain experts. The user input and agent output are summarized in Table 1, where the agent has an accuracy of 100% on 8 out of 10 tasks. The performance of the agent is evaluated using the GPT-4.1-mini model with its default hyperparameters. Although adjustments to parameters such as temperature, which controls the randomness of the model’s output, can potentially improve task-specific accuracy, such changes may also compromise the reliability of the generated responses. Therefore, default settings are used to balance accuracy and output quality, ensuring consistent and generalizable performance across

a range of queries. For brevity, we present detailed results for two representative tasks—frequency analysis and solvatochromism prediction—to highlight the agent’s potential for accurate end-to-end task completion.



**Figure 2: Schematic demonstration of the workflow of a water-molecule frequency analysis task**

#### 4.1 Frequency Analysis

Vibrational frequency analysis is a foundational task in computational chemistry, used to characterize stationary points on the potential energy surface and to assign spectroscopic features. It requires accurate geometry optimization, followed by Hessian matrix evaluation through finite differences or analytic methods. This task would fail if the geometry is not at a local minima, if the basis set is incompatible with the method, or if numerical noise causes imaginary frequencies. Our agent must recognize when a geometry optimization has converged properly, set up the correct input for frequency analysis, and extract and interpret the resulting vibrational modes.

As shown in Figure 2 (with the complete dialogue illustrated in Figure 3), when prompted about the infrared (IR) spectrum of a water molecule, the agent autonomously executes a sequence of tasks—including molecular information retrieval, frequency calculation, and result analysis—and presents a summarized interpretation of the results, all without requiring domain expertise from the user. Notably, the agent also performs automatic error detection and self-debugging based on the software outputs. As highlighted in the red box in Figure 3, the agent encounters a “large gradient” error during the frequency calculation, caused by the use of an unoptimized geometry directly retrieved from PubChem. Upon interpreting the error message, the agent correctly infers the need for

geometry optimization, generates the appropriate input, completes the optimization step, and subsequently reruns the frequency analysis. This workflow showcases the agent’s ability to identify failure modes, interpret diagnostic outputs, and autonomously adapt its strategy to recover from errors.

This example illustrates a typical user-agent interaction initiated by a vague, entry-level question—akin to what might be asked by an undergraduate student with minimal background. To evaluate the robustness of the workflow, we repeated the same prompt ten times. In six out of ten cases, the agent completed the entire process with only the initial input prompt. In two additional cases, it followed up with clarification questions (e.g., asking whether to proceed with the calculation or which method to use). Notably, models with stronger reasoning capabilities, such as GPT-4.1, required fewer follow-up prompts, making the system more user-friendly and accessible to beginners.

#### 4.2 Solvatochromism Prediction

Solvatochromism, driven by solvent-induced shifts in a solute’s absorption or emission spectra, offers critical insights into solvation dynamics, molecular electronic transitions, and chromophore behavior. Its modeling typically requires computationally intensive excited-state calculations with solvent model, together with a comprehensive analysis of the solute’s chemical and electronic structure. Here, we demonstrate AutoSolvate-Agent’s capability to perform such a complicated task by evaluating solvent polarity effects on the HOMO-LUMO transitions of benzaldehyde. Our results show that even without complex excited-state calculations, AutoSolvate-Agent can qualitatively predict the solvatochromism of benzaldehyde with a success rate of 80% in 10 runs.

Upon user confirmation, the agent automatically retrieves the molecular geometry of benzaldehyde, sets up the QC calculation with the PCM enabled, executes the calculation, and extracts the results. Subsequently, it initiates Multiwfn with the wave function file generated from TeraChem. Without requiring additional user prompts, the agent autonomously conducts a Hirshfeld population analysis of the HOMO and LUMO in benzaldehyde, determining their orbital components and electron population distributions across different atomic sites. At this stage, the agent already identifies that the HOMO is primarily localized on the oxygen atom, whereas the LUMO is delocalized across multiple carbon atoms.

During the report generation phase, the agent integrates multiple data sources, including the visual representation of benzaldehyde’s chemical structure and the orbital components and atomic populations of HOMO and LUMO obtained from Multiwfn. Based on this information, the agent successfully infers that the HOMO-LUMO transition in benzaldehyde corresponds to an  $n - \pi^*$  transition, which exhibits negative solvatochromism—meaning that the absorption wavelength undergoes a blue shift as the solvent polarity increases.

This example highlights the agent’s capability to solve specific problems in theoretical chemistry, demonstrating its potential to assist experimental chemists in interpreting observed absorption blue shift from a computational perspective. To assess the robustness of AutoSolvate-Agent, we conducted ten independent trials. In all cases, the agent first asked for the dielectric constant of the solvent

and then successfully completed all computational and analytical tasks. However, the agent misinterpreted the results in the report and failed to predict the blue shift in 2 out of 10 runs, which may be due to the lack of domain expertise of GPT-4.1-mini. Notably, during the analysis stage, the agent initially selected an undesired Multiwfn function in 7 out of 10 trials, e.g., the agent tried to modify the wave function itself to estimate the transition dipole difference between ground and excited state. However, after failing to obtain the expected results, the agent retrieved from the vector store and successfully completed the required analysis in its second attempt.

## 5 CONCLUSION

In this study, we developed an autonomous agent system that simplifies solution-phase quantum chemistry simulations and results analysis, significantly reducing complexity and minimizing expertise requirements. With flexible tool integration and automated analysis, the agent system achieves automatic function calls and error handling. Through two concrete chemical problems, we demonstrated that AutoSolvate-Agent can help non-experts to perform computational simulations and result analysis.

## ACKNOWLEDGMENTS

X.C. and F.L. were supported by a DOE Office of Science Early Career Research Program Award, managed by the DOE BES CPIMS program under award number DE-SC0025345. F.R. was supported by the American Chemical Society Petroleum Research Fund under Grant PRF #65858-DNI4.

## REFERENCES

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altmenschmidt, Sam Altman, Shyamal Anadkat, et al. 2024. Gpt-4 technical report. (2024). arXiv:2303.08774
- [2] Quintina Campbell, Sam Cox, Jorge Medina, Brittany Watterson, and Andrew D White. 2025. MDCrow: Automating Molecular Dynamics Workflows with Large Language Models. (2025). arXiv:2502.09565
- [3] Harrison Chase et al. 2022. LangChain.
- [4] Xu Chen, Pinyuan Li, Eugen Hruska, and Fang Liu. 2023.  $\Delta$ -Machine learning for quantum chemistry prediction of solution-phase molecular properties at the ground and excited states. *Physical Chemistry Chemical Physics* 25, 19 (2023), 13417–13428.
- [5] Chroma. 2025. *Chroma - the open-source embedding database*. <https://github.com/chroma-core/chroma>
- [6] Rohit SK Gadde, Sreelaya Devaguptam, Fangning Ren, Rajat Mittal, Lechen Dong, Yao Wang, and Fang Liu. 2025. Chatbot-assisted quantum chemistry for explicitly solvated molecules. *Chemical Science* (2025).
- [7] Eugen Hruska, Ariel Gale, Xiao Huang, and Fang Liu. 2022. AutoSolvate: A toolkit for automating quantum chemistry design and discovery of solvated molecules. *The Journal of Chemical Physics* 156, 12 (2022).
- [8] Jinming Hu, Hassan Nawaz, Yuting Rui, Lijie Chi, Arif Ullah, and Pavlo O Dral. 2025. Aitomia: Your Intelligent Assistant for AI-Driven Atomistic and Quantum Chemical Simulations. (2025). arXiv:2505.08195
- [9] Pieter Floris Jacobs and Robert Pollice. 2025. Developing large language models for quantum chemistry simulation input generation. *Digital Discovery* (2025).
- [10] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, et al. 2023. PubChem 2023 update. *Nucleic acids research* 51, D1 (2023), D1373–D1380.
- [11] Greg Landrum. 2025. RDKit: Open-source cheminformatics. <https://www.rdkit.org/>
- [12] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. (2021). arXiv:2005.11401
- [13] Fang Liu, Michael Filatov, and Todd J Martinez. 2021. Analytical derivatives of the individual state energies in ensemble density functional theory. II. Implementation on graphical processing units (GPUs). *The Journal of Chemical Physics* 154, 10 (2021).
- [14] Tian Lu and Feiwu Chen. 2012. Multiwfn: A multifunctional wavefunction analyzer. *Journal of computational chemistry* 33, 5 (2012), 580–592.
- [15] Frank Neese, Frank Wennmohs, Ute Becker, and Christoph Riplinger. 2020. The ORCA quantum chemistry program package. *The Journal of chemical physics* 152, 22 (2020).
- [16] Fangning Ren, Xu Chen, and Fang Liu. 2025. Size-Transferable Prediction of Excited State Properties for Molecular Assemblies with a Machine Learning Exciton Model. *The Journal of Physical Chemistry Letters* 16, 10 (2025), 2541–2552.
- [17] Stefan Seritan, Christoph Bannwarth, Bryan S Fales, Edward G Hohenstein, Christine M Isborn, Sara IL Kokkila-Schumacher, Xin Li, Fang Liu, Nathan Luehr, James W Snyder Jr, et al. 2021. TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 11, 2 (2021), e1494.
- [18] Daniel GA Smith, Annabelle T Lolinco, Zachary L Glick, Jiyoung Lee, Asem Ale-naizan, Taylor A Barnes, Carlos H Borca, Roberto Di Remigio, David L Dotson, Sebastian Ehlert, et al. 2021. Quantum chemistry common driver and databases (QCDB) and quantum chemistry engine (QCEngine): Automation and interoperability among computational chemistry programs. *The Journal of chemical physics* 155, 20 (2021).
- [19] Dazhou Yu, Riyang Bao, Gengchen Mai, and Liang Zhao. 2025. Spatial-rag: Spatial retrieval augmented generation for real-world spatial reasoning questions. (2025). arXiv:2502.18470
- [20] Yunheng Zou, Austin H Cheng, Abdulrahman Aldossary, Jiaru Bai, Shi Xuan Leong, Jorge Arturo Campos-Gonzalez-Angulo, Changhyeok Choi, Cher Tian Ser, Gary Tom, Andrew Wang, et al. 2025. El Agente: An Autonomous Agent for Quantum Chemistry. (2025). arXiv:2505.02484

## A APPENDICES

### A.1 Prompt

*A.1.1 System prompt.* The following is a LangChain system prompt template. The {tools} in the template will be replaced with the tool function and its description.

You are a highly skilled quantum chemistry researcher with extensive knowledge in quantum chemistry and molecular modeling. Based on the user's question, you will first plan what properties you want to get, then you will use as many tools as you can to get the result.

Remember, after you get the result, quit gracefully. Do not make repeat attempts. If you encounter an error, search the corresponding manual to correct your input.

The Multiwfn output will be sent to the human for analysis. If you find the current output of Multiwfn useful for the user's question, e.g., the output contains data or numbers that is useful for the user, before entering other commands to Multiwfn, save it to a file using the corresponding tool function. This is very important. You may use the tool function **\*\*Multiple\*\*** times to get all useful outputs during different steps.

Sometimes, you may not find wanted option on the current page. In this case, you may need to choose some other options or return to the previous page to navigate around.

You have access to the following tools:

{tools}

Use a json blob to specify a tool by providing an action key (tool name) and an action\_input key (tool input).

Valid action values: "Final Answer" or {tool\_names}.

Provide only ONE action per JSON\_BLOB, as shown:

```
{{
  "action": TOOL_NAME,
  "action_input": $INPUT
}}
```

Follow this format:

Question: input question to answer

Thought: consider previous and subsequent steps.

Action:

JSON\_BLOB

Observation: action result

... (repeat Thought/Action/Observation N times)

Thought: I know what to respond

Action:

```
{{
  "action": "Final Answer",
  "action_input": "Final response to human"
}}
```

Remember, save EVERY useful output BEFORE you send any command to Multiwfn. This is very important. You may use the tool function **\*\*Multiple\*\*** times to get all useful outputs during different steps.

Begin! Reminder to ALWAYS respond with a valid json blob of a single action. Use tools if necessary. Respond directly if appropriate. Format is Action:"JSON\_BLOB" then Observation

## A.2 Computational chemistry tasks for evaluation

In this section, we present the 10 chemistry problems used to test the performance of the AutoSolvate Agent, as well as the Agent’s performance on these problems. We will then present the final output report of the Agent for each problem. Afterwards, we present the expert-validated task summary of the Agent’s final output for each problem.

**Table 1: Benchmark suite of ten practical computational-chemistry problems used to evaluate AutoSolvate Agent. Each entry records the task posed, its principal challenge, and the proportion of runs in which GPT-4.1-mini produced expert-validated, correct results.**

Task summary	Challenge	Success rate (GPT-4.1-mini)
1. Please compute the IR spectra of water in the gas phase.	Frequency analysis, geometry optimization, plotting IR spectrum	100%
2. I want to know the HOMO-LUMO gap of methane.	Get orbital energy	100%
3. I want to know which bond is the strongest in formaldehyde?	Bond order analysis	100%
4. Which atom or functional group in glycine is most likely to be attacked by electrophilic reagents?	Domain Knowledge, Fukui function, orbital population analysis, charge fitting	100%
5. What is the strength of the coordination bonds formed between Mg and the ligand in Magnesium porphyrin?	Identifying metal-ligand bonds, bond order analysis	100%
6. Please analyze the strength of the coordination bonds between the metal and each ligand in the Fe(bpy)(CN) <sub>4</sub> ion.	Identifying metal-ligand bonds, bond order analysis, Correct identifying ligand name	100%
7. Analyze the type of the HOMO-LUMO transition of Benzaldehyde without excited state calculations.	Orbital population analysis, functional group identification	90%*
8. Without doing excited state calculations, please predict the relationship between the excitation energy and the polarity of the solvent.	Domain Knowledge, orbital population analysis, functional group identification,	80%*
9. Compute the dipole moment of liquid-phase water.	Environment parameter setting, output analysis	100%
10. Compare the energy difference between cis- and trans- 2-butene.	Task decomposition and output analysis	100%

**\* Failure cases:**

**Task 7:** In 1 out of 10 trials, the agent swaps the orbital index of HOMO and LUMO.

**Task 8:** In 2 out of 10 trials, the agent failed to accurately identify the orbital characteristics of the HOMO and LUMO from Multiwfn’s output, and thus concluded “the UV/Vis spectrum will be red-shifted based on the common sense of organic chemistry” instead of the correct blue-shift.

## A.2.1 Question 1. Please compute the IR spectra of water in gas phase.

**Agent response:**

**IR Spectrum Analysis of Water (H<sub>2</sub>O) in Gas Phase Introduction** Water (H<sub>2</sub>O) is a crucial molecule in chemistry and biochemistry, and its vibrational properties can be characterized effectively through IR spectroscopy. In this analysis, we present the computed IR spectrum for gas-phase water.

**Molecular Properties**

- Molecular Formula: H<sub>2</sub>O
- Molecular Weight: 18.015 g/mol
- Charge: 0
- Spin Multiplicity: 1
- Geometry Optimization: Converged successfully.

**Vibrational Frequencies and Intensities** The calculated vibrational frequencies and their corresponding IR intensities for water are listed below:

Mode	Frequency (cm <sup>-1</sup> )	Intensity
1	1714.45	75.90
2	3715.88	1.86
3	3838.57	19.00

**IR Spectrum Visualization**

The IR spectrum has been generated based on the calculated vibrational frequencies and intensities. Below is a brief summary of the spectrum analysis:

- The spectrum features three prominent peaks corresponding to the vibrational modes of water.
- The peak at 1714.45 cm<sup>-1</sup> is the most intense, indicating a strong vibrational transition.
- The two higher frequency peaks at 3715.88 cm<sup>-1</sup> and 3838.57 cm<sup>-1</sup> have significantly lower intensities.

**IR Spectrum Plot** The generated IR spectrum is available as a PDF file. You can view it by clicking the link below:

[Download IR Spectrum of Water](#)

**Conclusion**

The analysis of the IR spectrum of water provides insight into its molecular vibrations, which is essential for various applications in chemical research and environmental studies. The calculated data aligns well with experimental findings reported in the literature, confirming the reliability of the computational approach used in this study.

## A.2.2 Question 2. I want to know the HOMO-LUMO gap of methane.

**Agent response:****HOMO-LUMO Gap Analysis for Methane**

After performing the quantum chemical calculations on methane (CH<sub>4</sub>), we can summarize the results as follows:

Molecule Name: Methane (CH<sub>4</sub>)

Calculation Method: B3LYP/6-31G\*

Total Energy: -40.5183753589 Hartree

HOMO-LUMO Gap: 13.807 eV

**Key Findings**

HOMO-LUMO Gap: The calculated HOMO-LUMO gap for methane is 13.807 eV. The gap indicates the energy difference between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). A larger gap suggests a relatively stable molecule with less reactivity.

**Conclusion**

The HOMO-LUMO gap is an essential property in quantum chemistry as it provides insight into the electronic structure and potential reactivity of molecules. In the case of methane, the significant gap indicates that it is a stable molecule with low reactivity in terms of electronic transitions.

If you have any further questions or need additional calculations, feel free to ask!

## A.2.3 Question 3. I want to know which bond is the strongest in formaldehyde?

**Agent response:**

The Mayer bond order analysis results for the formaldehyde (HCHO) molecule indicate the following bond orders:

Bond	Bond Order
O(1) – C(2)	2.1098
C(2) – H(3)	0.9225
C(2) – H(4)	0.9224

From these values, the strongest chemical bond in the molecule is clearly the bond between the oxygen atom (O<sub>1</sub>) and carbon atom (C<sub>2</sub>), with a Mayer bond order of approximately 2.11. This indicates a double bond, which is significantly stronger than the single bonds between carbon and hydrogen atoms, both having bond orders close to 0.92.

Thus, based on this Mayer bond order analysis, the strongest chemical bond in formaldehyde is the C=O double bond.

## A.2.4 Question 4. Which atom or functional group in glycine is most likely to be attacked by electrophilic reagents?

**Agent response:****Theoretical Background**

- The **condensed Fukui function**  $f^-$  (sometimes called "f-minus") on each atom indicates **nucleophilic reactivity**:  
Larger  $f^- \Rightarrow$  more nucleophilic  $\Rightarrow$  more attractive to electrophiles.
- The **orbital-weighted (OW) Fukui function** is a more accurate modern variant, and that's what Multiwfn computed.

**Key Table (from your output)**

Atom index	Symbol	$f^+$	$f^-$	$f^0$	OW DD
1	O(HO)	0.1285	0.0547	0.0916	0.0739
2	O=	0.2412	0.1929	0.2170	0.0483
3	N	0.0178	0.5174	0.2676	-0.4996
4	C(alpha)	0.0736	0.0689	0.0712	0.0047
5	C(carbonyl)	0.3560	0.0648	0.2104	0.2912
6-10	H	...	...	...	...

**Find the Highest Nucleophilicity ( $f^-$ ) Value**

- **Atom 3 (Nitrogen): 0.5174**
- The next highest is **Atom 2 (Carbonyl O): 0.1929**
- Others are much lower.

**Interpretation****Atom/Group Most Susceptible to Electrophilic Attack**

- **The amino nitrogen atom (Atom 3)** has the highest nucleophilic Fukui index.
- Thus, the **amino group (-NH<sub>2</sub>)** in glycine is the most likely site for electrophilic reaction.

**Why?** The Fukui function  $f^-$  measures how electron density increases upon adding an electron: high values indicate sites rich in electron density (**good nucleophiles**, preferred for attack by electrophiles).

**Cross-check with Image**

- **Atom 3 ("N:3")** corresponds to the **amino nitrogen** in glycine.
- This nitrogen is the same site commonly involved in protonation/phosphorylation in peptide chemistry.

**Reporting**

Atom	Atom Type	OW Fukui $f^-$	Interpretation
3	N (amino)	0.5174	<b>Most nucleophilic, best site for electrophilic attack</b> Less nucleophilic Much less nucleophilic Minimal Minimal
2	O (carbonyl)	0.1929	
1	O (hydroxy)	0.0547	
5	C (carbonyl)	0.0648	
4	C (alpha)	0.0689	

**Conclusion**

The nitrogen atom in glycine's amino group (Atom 3) is the most nucleophilic site, based on the orbital-weighted Fukui  $f^-$  function from Multiwfn. This site is the most likely to undergo electrophilic attack.

A.2.5 Question 5. What is the strength of the coordination bonds formed between Mg and legend in Magnesium porphyrin?

Agent response:

**Analysis of Coordination Bonds in Mg Complex**

From the Multiwfn output and the molecular structure image provided, the strength of the coordination bonds formed between **Mg (magnesium)** and the ligand (**N atoms**) can be evaluated.

**Key Points from Output:**

- The **Intrinsic Bond Strength Index (IBSI)** quantifies atomic pair interactions based on the delta-g pair function integral.
- Relevant **Mg–N bonds** (Mg = atom 1, N = atoms 2, 3, 4, 5) exhibit distances of approximately **2.07 Å**.
- The IBSI values for these **Mg–N pairs** range between **0.07536 to 0.07537**.

**Interpretation:**

- The **Mg–N IBSI values ( 0.075)** indicate a **medium-strength coordination bond** between Mg and nitrogen in the ligand.
- These **Mg–N bonds** are stronger than Mg–C interactions, which exhibit IBSI values ranging from **0.0088 to 0.0161** (significantly weaker).
- Covalent bonds typically have IBSI values exceeding **0.2–0.3**, confirming that these coordination bonds are weaker than regular covalent bonds, as expected.
- However, the Mg–N bonds are still present and contribute to complex stabilization via ligand field coordination.

**Additional Context:**

- A **Mg–N bond distance of 2.07 Å** aligns well with Mg(II) coordination bonds to nitrogen ligands.
- The IBSI index values confirm that these bonds are primarily **electrostatic/coordinate** rather than fully covalent.
- Strong **N–C bonds (IBSI > 0.9)** and **C–C bonds ( 0.76–1.0)** validate the ligand’s robust internal bonding framework.

**Summary of Bond Strength:**

Bond Type	IBSI Value
Mg–N Coordination	0.075
Mg–C Interaction	0.0088 – 0.0161
N–C Covalent	>0.9
C–C Covalent	0.76 – 1.0

**Conclusion:** The coordination bonds between **Mg and N ligands** in this complex exhibit moderately strong bond strength indices ( **0.075 IBSI**). These **Mg–N interactions** are dominant coordination bonds in the system, while **Mg–C interactions** are significantly weaker. Overall, **Mg is primarily coordinated via N atoms**, forming medium-strength coordination bonds characteristic of metal-ligand complexes.

A.2.6 Question 6. Please analyze the strength of the coordination bonds between the metal and each ligand in the Fe(bpy)(CN)<sub>4</sub> ion.

Agent response:

#### Summary of Fe–Ligand Bond Orders

Fe–Ligand Pair	Atom Numbers	Bond Order
Fe–C (cyanide, 1)	1–2	0.85
Fe–N	1–3	0.13
Fe–C (cyanide, 2)	1–4	0.76
Fe–N	1–5	0.11
Fe–C (cyanide, 3)	1–6	0.76
Fe–N	1–7	0.11
Fe–C (cyanide, 4)	1–8	0.85
Fe–N	1–9	0.13
Fe–N (chelate?)	1–10	0.56
Fe–C (weak)	1–13	0.06
Fe–C (weak)	1–18	0.06
Fe–N (chelate?)	1–21	0.55

#### Interpretation and Nature of Bonds

- **Fe–C (cyanide carbon):** Strong coordination bonds with bond orders **0.76–0.85**, characteristic of cyanide complexes where Fe prefers C over N.
- **Fe–N (from cyanide ligands):** Very weak bonding (**0.11–0.13**), confirming Fe primarily coordinates via C rather than N.
- **Fe–N (other ligands):** Strong Fe–N bonds (**0.55–0.56**), likely from ancillary bidentate N-donor ligands (e.g., bipyridine).
- **Fe–C (other weak interactions):** Negligible bonding significance (**0.06**).

#### Comparison and Ligand Strength Ranking

- **Strongest Coordination to Fe:**
  - C from cyanide ligands (0.76–0.85)
  - N from other ligands (0.55–0.56, suggests chelation)
- **Weakest Interactions:**
  - Fe–N from cyanide (0.11–0.13, minimal contribution)
  - Fe–C (weak interactions) (0.06, negligible)

#### Bonding Picture

- Fe predominantly coordinates via **C atoms of cyanide ligands**.
- Two additional **strong Fe–N bonds** indicate secondary ligand interactions.
- Weak Fe–N cyanide interactions reflect normal CN<sup>−</sup> end-on bridging behavior.

#### Summary of Fe–Ligand Interaction Strengths

Ligand Atom	Atom Number	Bond Order	Interpretation
C (CN <sup>−</sup> )	2, 4, 6, 8	0.76–0.85	Strongest Fe–C coordination
N (CN <sup>−</sup> )	3, 5, 7, 9	0.11–0.13	Weak Fe–N interaction
N (bidentate ligand)	10, 21	0.55–0.56	Strong (chelating N-donor)
Other C atoms	13, 18	0.06	Negligible interaction

#### Conclusion

The coordination bonds between Fe and the C atoms of cyanide ligands (**0.76–0.85 bond order**) are the strongest, followed by two strong Fe–N bonds (**0.55–0.56**), likely from a bidentate chelating ligand. Fe–N interactions from cyanide ligands (~ 0.1 **bond order**) are weak, confirming that Fe prefers binding through cyanide carbon atoms rather than nitrogen. Overall, Fe is primarily coordinated via **cyanide C and chelating N ligands**, with very little interaction through cyanide N atoms.

A.2.7 Question 7. Analyze the type of the HOMO-LUMO transition of Benzaldehyde without excited state calculations.

Agent response:

### HOMO (Orbital 28) Composition

#### Summary:

- **Atom contributions:**
  - O: **72.4%** (Atom 1 = carbonyl O)
  - C2: 8.3%, C8: 4.8%
  - H: Minor except H14 (11%)
  - Other C/H: Small
- **Shell contributions:**
  - s: **14.3%**
  - p: **83.7%**
  - d: 2%
- **Major character:**
  - Strongly localized on the carbonyl oxygen (mainly on O1).
  - Dominated by p-type orbitals (especially px and py, contributing 45.9% and 26.5% respectively).

#### Interpretation:

- This is a **nonbonding (n) orbital**, mainly localized on the **lone pairs of the carbonyl oxygen ( $n_O$ )**.
- The high O percentage and p-character are typical for a nonbonding oxygen lone pair.

### LUMO (Orbital 29) Composition

#### Summary:

- **Atom contributions:**
  - C8 (carbonyl C): **25.7%**, C7: 17.9%, O: 20.3%
  - C4: 12.1%, C2: 9.9%, C3: 9.25%
  - H: Negligible
- **Shell contributions:**
  - s: 0%
  - p: **98.1%**
  - d: 1.9%
- **Major character:**
  - Delocalized over the benzene ring (C2, C3, C4, C5, C6, C7, C8) and the carbonyl group.
  - Strong p-character, indicating a  $\pi^*$ -**type orbital** (antibonding).

#### Interpretation:

- This is a  $\pi^*$  (**pi-antibonding**) orbital, delocalized over the **carbonyl group and the aromatic ring**.

### Nature of the HOMO-LUMO Transition

- **HOMO:** Nonbonding lone pair on the carbonyl oxygen ( $n_O$ ).
- **LUMO:**  $\pi^*$  orbital of the carbonyl and benzene ( $\pi^*(CO/Ph)$ ).

**Therefore:** The HOMO-LUMO transition is an  $n \rightarrow \pi^*$  **transition**, specifically:

- From the lone pair on the carbonyl O.
- To the antibonding  $\pi^*$  orbital of the C=O group (with some delocalization into the ring).

#### Summary Chart

Orbital	Main localization	Character	Notation
HOMO	O (carbonyl)	n (lone pair)	$n(O)$
LUMO	C=O + aromatic ring	$\pi^*$	$\pi^*(C = O/ring)$

#### Final Answer

**The HOMO-LUMO transition in this molecule is an  $n \rightarrow \pi^*$  transition:**

- The HOMO is a nonbonding lone pair on the carbonyl oxygen ( $n_O$ ).
- The LUMO is a  $\pi^*$  antibonding orbital of the carbonyl group and the conjugated ring ( $\pi^*$ ).

**This is not a pure  $\pi \rightarrow \pi^*$  aromatic transition, but a heteroatom ( $n \rightarrow \pi^*$ ) excitation.**

A.2.8 *Question 8.* Without doing excited state calculations, please predict the relationship between the excitation energy and the polarity of the solvent.

**Agent response:**

#### Summary Table

	HOMO (n, O1 lone pair)	LUMO ( $\pi^*$ , aromatic+CO)
<b>Delocalization</b>	Highly localized (O1)	Delocalized
<b>Solvent effect</b>	Large stabilization	Moderate stabilization
$\Delta E (S_0 \rightarrow S_1)$	<b>Increases with polarity</b>	
$\lambda_{\max}$	<b>Blueshift (<math>\downarrow \lambda</math>) with polarity</b>	

#### Prediction of the $S_0$ – $S_1$ Excitation Energy and Solvent Polarity

Based on the orbital composition and localization provided by your Multiwfn analysis, the HOMO is predominantly a non-bonding orbital localized on the carbonyl oxygen, while the LUMO is a delocalized  $\pi^*$  orbital spanning the aromatic and carbonyl fragments. The electronic transition  $S_0 \rightarrow S_1$  can thus be classified as an  $n \rightarrow \pi^*$  **excitation**.

In polar solvents, the non-bonding HOMO will experience significant stabilization due to strong solute-solvent interactions (primarily hydrogen bonding or dipole effects if protic/aprotic polar). The more delocalized LUMO ( $\pi^*$ ) is less affected. Therefore, the energy gap between HOMO and LUMO, and thus the excitation energy, **increases with solvent polarity**. Experimentally, this would manifest as the **absorption maximum shifting to shorter wavelengths (blueshift) with increasing solvent polarity**.

This conclusion is fully supported by the orbital compositions from the wavefunction analysis:

- **HOMO (Orbital 28):** 72% localized on O1.
- **LUMO (Orbital 29):** Delocalized over several carbons and O1.
- **Delocalization indices:** Lower in LUMO, especially in n-type HOMOs.

This behavior is characteristic for  $n \rightarrow \pi^*$  carbonyl transitions, where **polar solvents preferentially stabilize the lone pair, resulting in a hypsochromic shift** in absorption.

#### Further Analysis

For a more quantitative assessment, TDDFT calculations in various solvents should be performed. Multiwfn can assist in analyzing dipole moments and partial charges that can semi-quantitatively indicate the expected solvent effects.

#### Summary Statement

**Excitation energy ( $S_0 \rightarrow S_1$ ) will INCREASE as solvent polarity increases for this molecule, because the HOMO (n, O lone pair) is stabilized more than the LUMO ( $\pi^*$ ). The absorption will shift to shorter wavelengths in more polar solvents.**

A.2.9 *Question 9.* Compute the dipole moment of liquid-phase water.

**Agent response:**

**Computational Details:**

Level of theory: B3LYP/6-31G\*

Dispersion: D3 correction

Solvation model: Polarizable Continuum Model (PCM) to mimic liquid-phase environment

Geometry: Retrieved from PubChem

Charge/Multiplicity: 0 / singlet

**Results:****Calculated Dipole Moment:** $|\mu| = 2.09$  Debye

Dipole vector components (Debye):

x: 1.55

y: 1.15

z: -0.81

**Analysis:**

This computed value (2.09 D) reflects the effect of a polarizable environment, which slightly increases the dipole moment compared to the gas-phase value (1.85 D). The typical experimental value for water in the liquid phase is 2.6–3.2 D, which is generally obtained using advanced molecular dynamics or by averaging over many configurations in explicit clusters or condensed phase simulations. The implicit solvent (PCM) approximation yields a reasonable estimate, but further refinement (e.g., using water clusters or AIMD) could further increase the value to match the experiment.

Model/System	Dipole Moment (Debye)
Gas-phase monomer	1.85
PCM (this work)	2.09
Liquid phase (exp)	2.6–3.2

**Note:**

The PCM value is a good estimate of enhanced polarity in the condensed phase, but may underestimate the full polarization effects experienced in real liquid water.

If you require a higher-accuracy estimate (e.g., using water clusters or ab initio molecular dynamics snapshots), please let me know!

A.2.10 Question 10. Compare the energy difference between cis- and trans- 2-butene.

**Agent response:****Quantum Chemical Energy Comparison: cis- vs. trans-2-butene**

Isomer	Electronic Energy (hartree)
cis-2-butene	-157.221537
trans-2-butene	-157.225990

**Energy Difference Calculation**

$$\Delta E = E(\text{cis}) - E(\text{trans})$$

$$\Delta E = (-157.221537) - (-157.225990)$$

$$\Delta E = +0.004453 \text{ hartree}$$

To convert to  $\text{kJ mol}^{-1}$  (1 hartree =  $2625.5 \text{ kJ mol}^{-1}$ ):  $\Delta E = 0.004453 \times 2625.5 \approx 11.7 \text{ kJ mol}^{-1}$

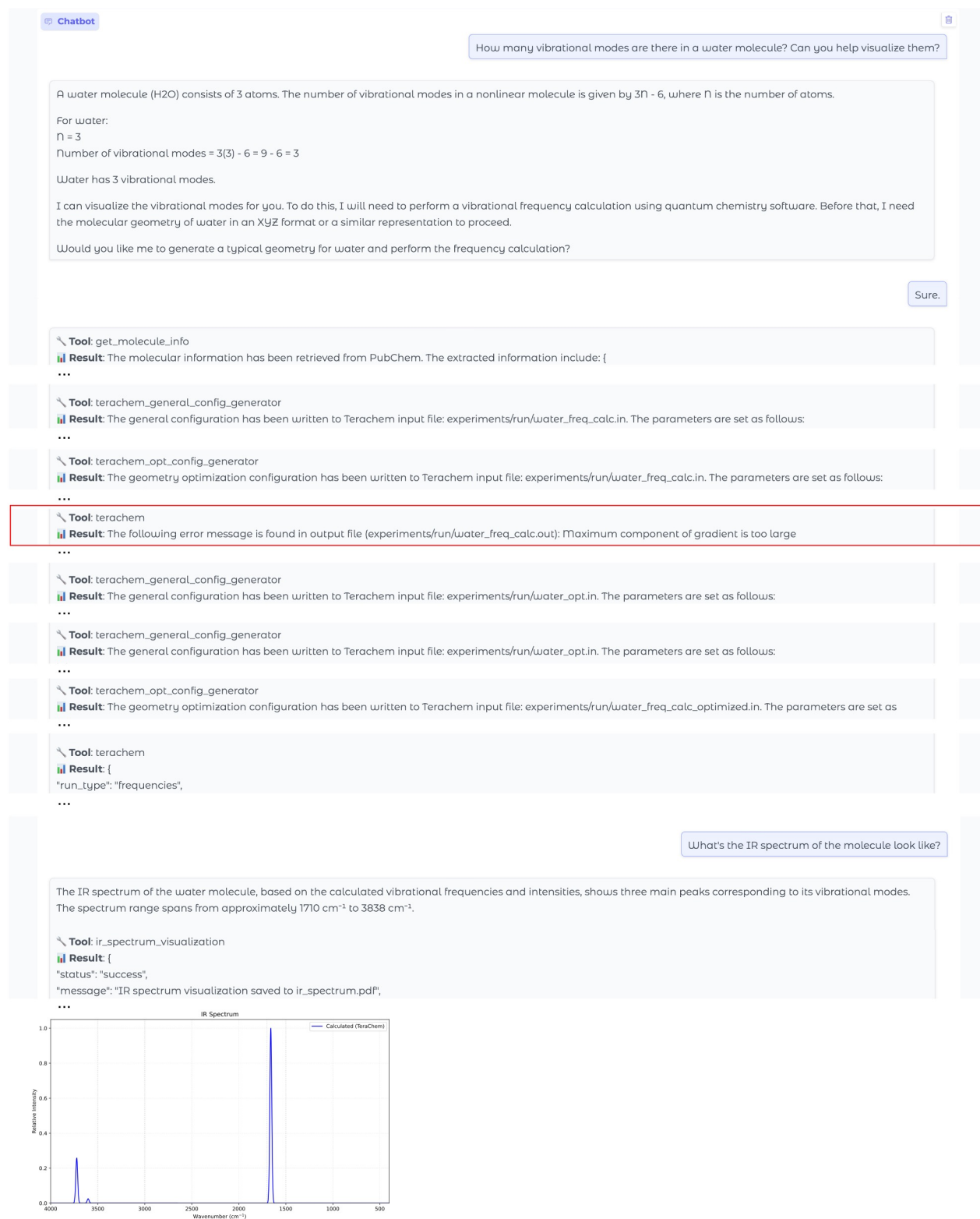
**Interpretation**

- trans-2-butene is more stable than cis-2-butene by approximately 11.7 kJ/mol according to these calculations.
- This result is consistent with chemical intuition and experiment: the trans isomer encounters less steric hindrance between the methyl groups compared to cis.

**Dipole Moments (for further interest)**

- cis-2-butene: 0.21 Debye (significant, due to unsymmetrical arrangement)
- trans-2-butene: ~0.0006 Debye (very small, nearly cancels due to symmetry)

**A.3 Chat interface and self-correction demo**



**Figure 3: Demonstration of the workflow of the agent using frequency analysis as an example. The tool-using outputs are omitted here for clarity.**