Molclus Automation Script: One-Click NMR and ECD Calculations for Complex Natural Products

Introduction

NMR and ECD calculations for complex natural products often present a high computational barrier for researchers, requiring extensive knowledge of quantum chemistry software, script writing, and data processing. To address this issue, we have developed a one-click automation script that integrates Molclus, Gaussian, ORCA, Multiwfn, and Shermo, enabling a fully automated workflow from conformational search to final result analysis.

This script has undergone multiple optimizations and tests to generate NMR and ECD calculations automatically while performing data analysis, significantly reducing complexity and improving efficiency. This document provides a comprehensive guide to using the script, making it accessible for beginners while offering optimization tips for experienced researchers.

Software Installation Guide

For detailed installation instructions, please refer to "Rocky Linux 9.3 System Installation and Setup of Gromacs, Gaussian, XTB, ORCA, SVM, and Other Programs."

---

1. Script Overview

The script ./11cal_nmr_ecd.sh is a Shell-based automation tool that integrates multiple quantum chemistry calculation programs to streamline NMR and ECD calculations.

1.1 Key Features

- **Conformational Search**: Utilizes Gromacs or XTB for molecular dynamics simulations to generate initial conformers.

- **Conformational Optimization**: Employs XTB and Gaussian for conformer optimization, ensuring the lowest-energy stable structures.

- **Vibrational Analysis**: Eliminates imaginary frequency conformers to confirm structural stability.

- **Single-Point Energy Calculation**: Uses ORCA for high-accuracy single-point energy calculations.

- **NMR Calculations**: Computes NMR chemical shifts using Gaussian.

- **ECD Calculations**: Performs time-dependent DFT (TD-DFT) calculations for ECD spectra using Gaussian.

- **Data Processing**: Uses Multiwfn and Shermo to weight-averaged calculated results and compare them with experimental data.

- **Data Visualization**: Automatically generates NMR and ECD plots for publication and presentation purposes.

1.2 Required Software

Before running the script, ensure the following programs are installed and environment variables are correctly configured:

- **Molclus** 1.12

- **XTB** -191025 (including crest functionality)

- **Gaussian** 16 A03

- **ORCA** 5.04

- **Shermo** 2.4

- **Multiwfn** 3.8 (dev)

- **Python** (Anaconda, Python 3.7+ with numpy, pandas, xlsxwriter, openpyxl, glob)

- **Gromacs** 2024.2

- **VMD** 1.9.4

- **ImageMagick**

- **sobtop**

---

2. Installation and Configuration

2.1 System Requirements

- **Operating System**: CentOS 7.6 or CentOS Stream 9 (other Linux distributions may work but require compatibility testing).

- **Hardware Requirements**: Multi-core CPU and large-memory servers are recommended for faster calculations.

2.2 Software Installation

Ensure the following software is correctly installed and accessible via the command line:

- **Molclus**: Refer to Sobereva's blog for installation.

- **XTB**: Refer to Sobereva's blog for installation.

- **Gaussian**: Refer to Sobereva's blog for installation.

- **ORCA**: Refer to Sobereva's blog for installation.

- **Shermo**: Refer to Sobereva's blog for installation.

- **Multiwfn**: Refer to Sobereva's blog for installation.

2.3 Environment Configuration

unzip 6molclus112.zip

cd 6molclus112

chmod +x *

Place the .pdb structure file from ChemDraw 3D (e.g., 004rs.pdb) in the directory, along with experimental data files (.csv for ECD and .txt for NMR carbon shifts), such as 004rs_nmr.txt and 004rs.csv.

---

3. Calculation Workflow

3.1 Running the Calculation

./11cal_nmr_ecd.sh

3.2 Running in Background

nohup ./11cal_nmr_ecd.sh &> output.txt &

3.3 Result Analysis

- **NMR Results**: Run ./12ana_nmr.sh to extract NMR data from three basis sets and perform DP4+ carbon shift comparisons.

- **ECD Results**: Run ./13ana_ecd.sh to compare experimental and calculated ECD spectra and generate high-resolution 3D structure plots.

- **SVM Analysis**: Run ./14ana_svm.sh to apply SVM-based classification for determining the relative configuration of compounds.

---

4. Calculation Steps

4.1 Conformational Search

Perform molecular dynamics-based conformational searches using Gromacs, generating a traj.xyz file.

4.2 XTB Quick Optimization

Use crest for rapid conformational optimization to provide reasonable input for Gaussian calculations.

4.3 Gaussian Precise Optimization

By default, the lowest 15 energy conformers are optimized. Modify 03_set.ini to adjust the number of conformers.

4.4 Vibrational Analysis

Perform vibrational analysis in Gaussian to eliminate imaginary frequency structures.

4.5 ORCA Single-Point Energy Calculation

Compute single-point energy for conformers to provide data for Boltzmann weighting.

4.6 NMR and ECD Calculations

- **NMR Calculations**: Compute NMR chemical shifts using three different functionals and basis sets.

- **ECD Calculations**: Compute ECD spectra for the top five major conformers to improve computational efficiency.

4.7 Result Analysis and Visualization

- Calculate Boltzmann distributions to determine major conformer contributions.

- Compute weighted NMR and ECD results.

- Generate experimental vs. calculated ECD spectra comparison plots.

- Automatically output high-quality 3D molecular structures.

---

5. Feature Updates (March 2025)

- **Enhanced Automation**: Supports simultaneous conformer calculations and automatic identification of methyl hydrogen numbering.

- **Optimized Calculation Workflow**: Added sobtop charge calculations, improved conformer search accuracy.

- **Data Visualization**: High-quality 3D structure rendering and experimental vs. calculated ECD comparison plots.

- **Python-Based Data Processing**: Added dp4.py for extracting NMR data and supporting SVM-based structural analysis.

---

6. Conclusion

This script fully automates the NMR and ECD calculation workflow for natural products, significantly improving computational efficiency while providing comprehensive result analysis and visualization, making it ideal for scientific publications and presentations.

If you use this script in your research and publish results, please provide appropriate citations to the contributors.

The script is continuously being optimized, and feedback or suggestions are always welcome!