**AI ECD：一键完成对复杂天然产物NMR和ECD计算**

**前言**

复杂天然产物往往需要进行NMR和ECD的计算，由于多数天然产物专业人员量子化学计算机操作能力有限，所以NMR和ECD的计算困难重重。为了便于计算，尽量减少计算操作，笔者在chatGPT的帮助下，设计了一套计算复杂天然产物NMR和ECD的脚本，脚本经历多次修改和测试，终于实现了一键计算操作，此脚本能自动处理计算结果，计算完成后直接获得NMR和ECD结果，因此拿出来与诸位先进朋友分享交流，还请多多指教，批评指正。

本脚本及其配套的相关软硬件的安装请参考《[Rocky Linux 9.3系统安装及Gromacs,Gaussian, xtb, ORCA, SVM等程序安装](http://bbs.keinsci.com/forum.php?mod=viewthread&tid=43105&page=1&extra=" \l "pid280774" \t "_blank)》

Molclus程序: [《使用molclus程序做团簇构型搜索和分子构象搜索》](http://bbs.keinsci.com/thread-577-1-1.html)

Molclus配套的PBS作业提交脚本：**《**[分享Molclus配套的PBS作业提交脚本](http://bbs.keinsci.com/thread-15598-1-1.html)**》**

SVM-M复杂天然产物NMR计算：[https://github.com/Anan-Wu-XMU/SVM-M](https://github.com/Anan-Wu-XMU/SVM-M" \t "_blank)

ChatGPT: <https://chat.openai.com/>

本脚本及其配套的相关输入文件主要是在前人基础进行再次修改和创作，在使用本脚本进行科学研究，研究结果发表时，如若能正确引用，笔者将万分感激！  
  
**1. molclus懒人脚本简介**

本文附件的压缩包中有一个00autorunall.sh 文件，就是一个简单的shell命令的组合，根据自己的研究需要，在前人的基础上，结合chatGPT，将自己的诉求输入chatGPT，就会产生所需要的shell命令，将不同的诉求命令组合起来就产生了本脚本。脚本B站视频请移步《[chatGPT辅助生成molclus懒人脚本：一键完成对复杂天然产物NMR和ECD计算](https://www.bilibili.com/video/BV1SN411g7NH/?vd_source=e38be56554d121c442bfbd0ddcadcc8c" \l "reply493288204" \t "_blank)》

**2. 脚本的依赖程序**

脚本依赖Molclus 1.12，xtb -191025及crest功能，Gaussian 16 A03，ORCA 5.04 ，Shermo 2.4，Multiwfn 3.8 (dev) 等程序，在运行脚本前确保服务器中能使用这些程序。这些程序的安装及使用可参考社长大人Sobereva老师的相关博文：

Molclus程序安装和使用参考《[使用molclus程序做团簇构型搜索和分子构象搜索](http://bbs.keinsci.com/thread-577-1-1.html" \t "_blank)》

xtb的安装参考《[将Gaussian与Grimme的xtb程序联用搜索过渡态、产生IRC、做振动分析](http://sobereva.com/421" \t "_blank)》

Gaussian安装参考《[Gaussian的安装方法及运行时的相关问题](http://bbs.keinsci.com/thread-10814-1-1.html" \t "_blank)》

ORCA 的安装参考《[量子化学程序ORCA的安装方法](http://bbs.keinsci.com/forum.php?mod=viewthread&tid=11697&extra=page%3D1%26filter%3Dtypeid%26typeid%3D11%26digest%3D1" \t "_blank)》

Shermo的安装参考《[使用Shermo结合量子化学程序方便地计算分子的各种热力学数据](http://bbs.keinsci.com/thread-17494-1-1.html" \t "_blank)》

及《[Multiwfn在Linux下安装的中文说明](http://bbs.keinsci.com/thread-40529-1-1.html" \t "_blank)》

Python的安装可Google查询，还需要如下数据库；

python -m pip install numpy pandas  
python -m pip install xlsxwriter  
python -m pip install  Workbook  
python -m pip install numpy  
python -m pip install  pandas  
python -m pip install  glob  
python -m pip install openpyxl

以上软件可正常使用，最好将软件的路径加入系统环境变量，以达到输入软件名就可直接启动软件，准备完毕，就可启动脚本了，该脚本即是压缩包里00autorunall.sh，即是一个简单的Shell命令的组合，在CentOS 7.6和CentOS Stream 9中均可正常运行。

**3. 首次使用需要的设定**

脚本依赖的程序可正常使用后。

* 将本压缩包解压，进入该目录，赋予执行权限

unzip molclus112.zip  
cd molclus112  
chmod +x  \*

* 给解压之后的目录文件中，**提供输入文件**；**如果未提供输入文件将会出现错误提示coord file does not exit，输入文件为需要有多个构象的 “\*\_traj.xyz”和只有一个构象的“\*\_traj2.xyz”，比如511rr\_traj.xyz, 511rr\_traj2.xyz, 511sr\_traj.xyz, 511sr\_traj2.xyz.**其中多个构象的**“\*\_traj.xyz”文件将被moclus进行使用计算，只有一个构象的“\*\_traj2.xyz”将被python脚本09\_H\_nmrstat.py提取相应的甲基氢的编号，为Multiwfn给出氢谱数据提供输入编号**
* 修改00autorunall.sh中第27行，37行中xtb程序crest在本机中的实际绝对路径，例子中为：**/root/soft/xtb/crest**
* 修改05\_set.ini中第15行，orca在本机中的实际路径，例子为：orca\_path= "/root/soft/orca504/orca"
* 脚本增加了记录脚本运行时长，运行结束会将运行时间发送到邮箱进行提醒，这个功能是笔者的最爱，非常的方便。如果系统有发邮件功能，将脚本最后一行“echo " Normal termination of " | mail -s "MASTER complete " [tttt@163.com](mailto:tttt@163.com)”中邮箱部分，换成自己的邮箱，计算完成会发邮件提示。如何给系统安装发邮件功能，感兴趣者可自行Google。
* 准备就已经完成了。接着就是一行命令./00autorunall.sh即可获得NMR和ECD结果。如需要放进后台运行，输入命令nohup ./00autorunall.sh  &>  output.txt &，计算结束还会有邮件提示。
* 在当前目录下，即在这个文件夹下，如果只想获得NMR结果，运行./01\_mol\_nmr.sh，或者nohup ./01\_mol\_nmr.sh &

**4. 脚本的计算流程**

为了便于管理，笔者将脚本的计算流程分为10步:

**4.1 构象搜索**

使用Gentor, Spartn, XTB, Gromacs等均可，主要为脚本开始计算提供traj.xyz文件。压缩包中含一个例子的traj.xyz，请替换为自己的构象搜索结果文件。笔者在Linux系统中《[脚本分享：sobtop联合Gromacs对天然产物进行分子动力学构象搜索](http://bbs.keinsci.com/thread-41881-1-1.html" \t "_blank)》或Window系统中《[Windows系统Gromacs进行分子动力学构象搜索](http://bbs.keinsci.com/thread-41215-1-1.html" \t "_blank)》有详细介绍两种文件的产生，将这两个相应的构象文件“\*\_traj.xyz”和“\*\_traj2.xyz”放入Linux系统6molclus112文件夹中，文件就准备完成了。

有三个原因推荐用Gromacs的分子动力学的方法进行构象搜索。1，在NMR和ECD的计算中，能量最低的构象，或者接近能量最低的构象非常关键，如果没有找到或者找到的构象不接近实际实验中分子的构象或在自然界中可能存在的构象，那么之后步骤再高明的计算，再高精度的基组和泛函的计算，都会存在不容忽视的误差，因此，强烈推荐用分子动力学方法进行构象搜索，特别是含义环状体系的分子，如果连普通6元环船椅式构象都不能搜索到，怎么指望多元环更复杂的环状构象？2，分子动力学构象搜索可以用XTB，Gromacs，Amber等；3，笔者推荐Gromacs，易安装，易用，关键一个字“快”。

**4.2 XTB 快速优化**

产生初始能量由低到高排列稳定的构象，为Gaussian进一步优化提供可靠的输入文件。注意，首次使用请修改脚本中**crest**的路径为本机的实际路径。crest为XTB中的一个辅助插件，当安装好XTB程序后，还需要下载**相应版本的crest.tgz**插件，解压后，赋予权限，可以放到XTB文件夹中。在使用中需要输入crest的实际绝对路径，比如脚本的路径在/root/soft/xtb/crest。

**4.3 对上步中前15个能量最低的构象调佣Gaussian进行精确耗时的优化**

如果不想做任何修改，本脚本默认设置数量为15，15基本不会漏掉搜索到的能量最低的构像，熟悉Molclus程序者可以自由调节03\_set.ini中参数。计算结束后，并会输出能量由低到高排列的文件0003\_output.xyz。压缩包中含有以下几个本步计算所需的文件，如果需要可自行调节相关参数。

03\_set.ini

03\_tem1.gjf

03\_tem2.gjf

03\_tem3.gjf

**4.4 对0003\_output.xyz进行振动分析**

 每次运行molclus程序时，该程序都会检查当前文件的.out，并删除，每一步的运行结果都有需要保留，所以脚本多了一步将上一步产生的out转变log的命令，以避开molclus的清理.out作业，相应的Multiwfn需要识别当前目录的.log文件。

molclus开始振动分析，排除虚频的结构，得到0004\_input.xyz 和能量由低到高排列的0004\_output.xyz。压缩包中含有以下几个本步计算所需的文件：

04\_set.ini

04\_tem.gjf

**4.5 调用ORCA**

用较高级别的基组和泛函计算单点能。压缩包中含有以下几个本步计算所需的文件：

05\_set.ini

05\_tem.inp

注意，首次使用，需要修改05\_set.ini中orca的路径

**4.6 计算0004\_input.xyz文件中所有构象的NMR**

由于NMR相对不耗时，而复杂天然产物的NMR往往也不是一种泛函和基组就能确定的，所以选用三组基组和泛函对所有构象NMR进行计算，以便选取与实验结果比较接近的结果。以下文件是本步计算所需：

06\_set\_nmr.ini

06\_tem01\_nmr.gjf

06\_tem02\_nmr.gjf

06\_tem03\_nmr.gjf

**4.7 计算0004\_input.xyz文件中前5个构象的ECD**

由于ECD较为耗时，在以往计算经验中发现第6之后的构象占比在1%以下，对整体ECD图谱影响甚微，所以选择前五个。确保我们选择了主要的构象之后，我们又选用三套基组和泛函对前5个构象的ECD进行计算，以便选取与实验结果比较接近的结果。以下文件是本步计算所需：

07\_set\_ecd.ini

07\_tem01\_ecd.gjf

07\_tem02\_ecd.gjf

07\_tem03\_ecd.gjf

**4.8 Shermo计算Bloztmann比例**

首先脚本会自动创建Shermo的输入文件0008\_sher01\_input.txt，文件第一列是寻找当前目录振动分析结果文件004\_freq\_gau\*.out，第二例是orca计算结果的单点能。0008\_sher02\_input.txt是前五个构象的信息。结果输出为，前15个构象的Bloztmann比例0008\_re\_sher01rate.txt和前5个构象的比例0008\_sher02\_input.txt

**4.9 Multiwfn处理得到权重之后的NMR结果**

首先脚本会自动生成0009\_nmr01\_rate.txt文件，第一列是NMR的Gaussian计算结果文件，第二列是由Shermo计算的Bloztmann比例。

此步计算涉及以下文件，

09\_C01\_nmr.txt

09\_C02\_nmr.txt

09\_C03\_nmr.txt

09\_H01\_nmr.txt

09\_H02\_nmr.txt

09\_H03\_nmr.txt

熟悉Multiwfn的人会知道，以上文件就是Multiwfn的输入命令，由于链上C-C单键可以自由旋转，一般实验所得的甲基或有的次甲基的氢化学位移是一样的，而计算中对于一个特定构象上的三个氢化学位移不一样，为了与实验接近，需要对相应的氢进行平均化，因此需要设定09\_H01\_nmr.txt，09\_H02\_nmr.txt，09\_H03\_nmr.txt文件中相应的氢的编号。脚本09\_H\_nmrstat.py根据**只有一个构象的“\*\_traj2.xyz”文件提取相应的**氢的编号，制作09\_H01\_nmr.txt，09\_H02\_nmr.txt，09\_H03\_nmr.txt文件。

由三套基组和泛函因此生成三个Multiwfn输入文件0009\_nmr01\_rate.txt，0009\_nmr02\_rate.txt，0009\_nmr03\_rate.txt。

此步计算结束，会有相应的三组权重之后H NMR和C NMR结果。

0009\_HNMRdata01.txt  0009\_CNMRdata01.txt

0009\_HNMRdata02.txt  0009\_CNMRdata02.txt

0009\_HNMRdata03.txt  0009\_CNMRdata03.txt

**4.10 Multiwfn处理得到权重之后的ECD结果**

首先脚本会自动生成00010\_ecd01\_rate.txt文件，第一列是Gaussian计算ECD结果文件，第二列是由Shermo计算的前五个构象的Bloztmann比例。如有需要可适当调节如下文件的相关参数：

10\_ecd01.txt

10\_ecd02.txt

10\_ecd03.txt

由三套基组和泛函因此生成三个Multiwfn输入文件00010\_ecd01\_rate.txt，00010\_ecd01\_rate.txt，00010\_ecd01\_rate.txt。

此步计算结束，会有相应的三组权重之后ECD结果。

00010\_dislin01.png  00010\_spectrum01\_curve.txt 00010\_curveall01.txt

00010\_dislin02.png  00010\_spectrum02\_curve.txt 00010\_curveall02.txt  
00010\_dislin03.png  00010\_spectrum03\_curve.txt 00010\_curveall03.txt

最后，在molclus112目录中还有一个01.sh脚本，还可将每步命令放入其中，可以分别对每一步进行运算，分步测试。

脚本还保留了.pbs文件的部分命令行，如果使用openpbs作业提交管理功能，只需将00autorunall.sh 修改为00autorunall.pbs，并修改其中的节点就可使用。

**6. 注意**

对于该脚本00autorunall.sh  中每一行的含义，感兴趣者可以输入到chatGPT中去了解命令的具体含义，同时也可以按照自己需要进行适当修改和改进。

6.1 python脚本09\_H\_nmrstat.py，能够自动提取提取分子中甲基氢编号，因此，**不需要**提供09\_H01\_nmr.txt，09\_H02\_nmr.txt，09\_H03\_nmr.txt，**需要**有一个只含有一个分子构象的名字为\*\_traj2.xyz的文件，比如2R4S\_traj2.xyz，后缀名“\_traj2.xyz”是必须的。并且更改09\_H\_nmrstat.py脚本中第一行为自己系统的python路径，比如笔者python3.7路径如下#!/usr/local/python37/bin/python3.7，如何安装并行系统自带python版本和python3.7，请自行谷歌搜索安装。

6.2 脚本是for循环，天然产物中往往好几个相对构型需要计算，比如2S3R4R, 2S3S4S, 2R3R4S等**（温馨提示，几个构型的原子编号最好一致，以供与实验对比，若编号不一致，提取化学位移与实验对比将会非常麻烦）**，加入for循环后，可将这些分子统统放进一个文件，脚本将进行逐个计算。因此需要有多个构象的 “\*\_traj.xyz”和只有一个构象的“\*\_traj2.xyz”，比如

5112R3R4S \_traj.xyz

5112R3R4S \_traj2.xyz

5112S3R4R \_traj.xyz

5112S3R4R \_traj2.xyz

……

6.6 注意脚本./01\_mol\_nmr.sh   中第222行：

head -n 3 "${base\_name}\_08\_sher00\_input.txt" > "${base\_name}\_08\_sher01\_input.txt"

数字3需要与文件06\_set\_nmr.ini中第二行计算NMR构象ngeom= 3的数字一致，请根据计算机的繁忙程度自行调节，如果需要快速获得结果的话，可以选3，如果想保险一些，计算资源非常充足，可用7, 10等。此外，如果NMR和ECD计算完成后，若发现Gaussian正常完成，而分布比例有问题，或未能成功产生权重的结果，检查错误出现的位置，比如./01\_mol\_nmr.sh   中第222行数字与实际的06\_set\_nmr.ini数字不一样，相应修改后，可单独运行./002\_shermo.sh。

6.7 增加了一个python脚本dp4.py，提取\*\_09\_\*NMRdata\*.txt中权重之后的最终结果到excel, 这样可以一次获得每个构型的最终的化学位移及屏蔽常数。使用python脚本还需要安装一些数据库，如下：

python -m pip install numpy pandas  
python -m pip install xlsxwriter  
python -m pip install  Workbook  
python -m pip install numpy  
python -m pip install  pandas  
python -m pip install  glob  
python -m pip install openpyxl

6.8 python脚本nmrstat.py和shell脚本004\_motosvm.sh，提供一个实验获得的C谱数据exp.txt，运行./004\_motosvm.sh将会给出每个构型与实验的差异，判断是哪个构型。该nmrstat.py来自于SVM-M计算NMR方法。

6.9 贴一个计算成功的例子链接：[https://pan.baidu.com/s/18mwUEOtLENnRKP-YttBJVQ?pwd=qcl0](https://pan.baidu.com/s/18mwUEOtLENnRKP-YttBJVQ?pwd=qcl0" \t "_blank)

提取码：qcl0

6.10

最后，感谢中科院第一海洋研究所崔博士好友设计完成的09\_H\_nmrstat.py脚本，此外该脚本还得到其他很多人的帮助和支持，在此表示真诚的感谢！

本文中相关脚本及其配套的输入文件主要是在前人基础进行再次修改和创作，在使用它进行科学研究，研究结果发表时，如若能正确引用，笔者将万分感激！

由于本人非计算机出身，才疏学浅，脚本中难免有错误，请各位多多指教，谢谢大家！

\*\*AI ECD: One-Click Calculation of Complex Natural Product NMR and ECD\*\*

\*\*Introduction\*\*

Complex natural products often require NMR and ECD calculations. Due to the limited quantum chemistry computing skills among many natural product professionals, these calculations can be quite challenging. To facilitate these computations and minimize operational steps, I designed a script with the help of ChatGPT that allows one-click NMR and ECD calculations for complex natural products. After numerous modifications and tests, this script can now automatically process calculation results, directly providing NMR and ECD outcomes upon completion. I am sharing this with fellow researchers and welcome your feedback and corrections.

For the installation of the script and its associated software and hardware, please refer to "Rocky Linux 9.3 System Installation and Gromacs, Gaussian, xtb, ORCA, SVM, and other programs installation."

Molclus Program: "Using Molclus for Cluster Configuration and Molecular Conformation Search"

Molclus PBS Job Submission Script: "Molclus PBS Job Submission Script"

SVM-M Complex Natural Product NMR Calculation: [GitHub Link](https://github.com/Anan-Wu-XMU/SVM-M)

ChatGPT: [ChatGPT Link](https://chat.openai.com/)

The script and associated input files are modified and created based on previous work. If you use this script for scientific research and publish your findings, proper citation would be greatly appreciated!

### 1. Introduction to the Molclus Lazy Script

The compressed file attached contains a 00autorunall.sh file, which is a simple combination of shell commands. Based on personal research needs and using ChatGPT, I input specific requirements to generate the necessary shell commands. By combining different command demands, this script was created. For a video guide, please visit "ChatGPT-Assisted Generation of Molclus Lazy Script: One-Click Calculation of Complex Natural Product NMR and ECD."

### 2. Script Dependencies

The script relies on programs such as Molclus 1.12, xtb -191025 and crest, Gaussian 16 A03, ORCA 5.04, Shermo 2.4, Multiwfn 3.8 (dev), etc. Ensure these programs are available on the server before running the script. Installation and usage references can be found in the relevant blogs by Sobereva:

- Molclus installation and usage: "Using Molclus for Cluster Configuration and Molecular Conformation Search"

- xtb installation: "Using Gaussian with Grimme's xtb for Transition State Search, IRC Generation, and Vibration Analysis"

- Gaussian installation: "Gaussian Installation Methods and Related Issues"

- ORCA installation: "Quantum Chemistry Program ORCA Installation Method"

- Shermo installation: "Using Shermo with Quantum Chemistry Programs for Thermodynamic Data Calculation"

- Multiwfn installation: "Multiwfn Installation Guide in Linux"

For Python, install the following libraries:

```sh

python -m pip install numpy pandas xlsxwriter Workbook glob openpyxl

```

Ensure these programs are in the system environment path for easy execution. After preparation, run the script (00autorunall.sh), which is a combination of simple shell commands, functioning in both CentOS 7.6 and CentOS Stream 9.

### 3. Initial Setup

After ensuring the dependent programs are operational:

- Unzip the package and grant execution permissions:

```sh

unzip molclus112.zip

cd molclus112

chmod +x \*

```

- Provide input files in the unzipped directory; missing input files will prompt an error message ("coord file does not exist"). Input files should include multiple conformations ("\*\_traj.xyz") and a single conformation ("\*\_traj2.xyz").

- Modify paths in 00autorunall.sh (lines 27, 37) for the xtb program crest and in 05\_set.ini (line 15) for ORCA as per your system's setup.

- The script records execution time and sends an email notification upon completion. Set up mail functionality in your system and replace the email in the script with your own.

- Start the script with:

```sh

./00autorunall.sh

```

For background execution:

```sh

nohup ./00autorunall.sh &> output.txt &

```

For NMR results only, run:

```sh

./01\_mol\_nmr.sh

```

or for background execution:

```sh

nohup ./01\_mol\_nmr.sh &

```

### 4. Script Calculation Workflow

The script divides the calculation workflow into ten steps for management:

1. \*\*Conformation Search\*\*: Generate traj.xyz files using Gentor, Spartan, XTB, Gromacs, etc. Replace example traj.xyz files in the package with your own. Recommended to use Gromacs for conformation search due to its speed and ease of use.

2. \*\*XTB Quick Optimization\*\*: Sort initial conformations by energy for further Gaussian optimization. Modify crest path for first-time use.

3. \*\*Gaussian Optimization\*\*: Optimize the lowest energy conformations. Adjust parameters in 03\_set.ini as needed.

4. \*\*Vibration Analysis\*\*: Perform vibration analysis to exclude structures with imaginary frequencies.

5. \*\*ORCA Single Point Energy Calculation\*\*: Use higher-level basis sets and functionals.

6. \*\*NMR Calculation\*\*: Calculate NMR for all conformations using three sets of basis sets and functionals.

7. \*\*ECD Calculation\*\*: Calculate ECD for the top 5 conformations using three sets of basis sets and functionals.

8. \*\*Shermo Boltzmann Proportions\*\*: Generate Boltzmann proportions for the top 15 and top 5 conformations.

9. \*\*Multiwfn NMR Results\*\*: Generate weighted NMR results using Multiwfn.

10. \*\*Multiwfn ECD Results\*\*: Generate weighted ECD results using Multiwfn.

An additional script, 01.sh, allows running each step separately for testing.

### 5. Notes

- For the detailed meaning of each command in 00autorunall.sh, consult ChatGPT and make necessary adjustments.

- Python script 09\_H\_nmrstat.py extracts methyl hydrogen numbers for NMR calculations. Modify the first line with your system's Python path.

- The script supports multiple conformations in a loop; ensure appropriate input files (e.g., 5112R3R4S\_traj.xyz and 5112R3R4S\_traj2.xyz).

- Adjust the number of conformations for NMR calculations based on computational resources.

- Use dp4.py to extract final NMR results into Excel.

- Use nmrstat.py and 004\_motosvm.sh to compare calculated NMR data with experimental results.

- Example of a successful calculation: [Baidu Link](https://pan.baidu.com/s/18mwUEOtLENnRKP-YttBJVQ?pwd=qcl0) (Password: qcl0)

### Acknowledgments

Thanks to Dr. Cui from the First Institute of Oceanography, CAS, for the 09\_H\_nmrstat.py script and to all others who have contributed. If this script aids your research, proper citation is appreciated.

Thank you for your feedback and suggestions!