BRCORP
TECHNOLOGY | INNOVATION | INDUSTRY

5th World Conference on
Applied Sciences, Engineering & Technology
02-04 June 2016, HCMUT, Vietnam

MSMC-GUI - An Automatic Setup Tool for Hindered Internal Rotation Treatment for Ab Initio Thermodynamic Property Calculations

TRIET H. M. LE[1], SON T. DO[1] AND LAM K. HUYNH[2,*]

[1]*School of Computer Science and Engg, International University – Vietnam N U, Ho Chi Minh City, Vietnam*
[2]*School of Biotechnology, International University – Vietnam National University, Ho Chi Minh City, Vietnam*
Email: ITITIU13086@student.hcmiu.edu.vn, ITITIU13100@student.hcmiu.edu.vn, hklam@hcmiu.edu.vn

Abstract: The thermodynamic properties calculation is an essential step to help understand more thoroughly the behaviour considered species within a (bio-) chemical system. As a matter of fact, in order to determine these thermodynamic parameters, a number of configurations of the species have to be prepared (e.g., the species charge, multiplicity, energy level, coordinates as well as its frequencies for Hindered Internal Rotation treatment[1]) from electronic structure output file. Moreover, some rigorous treatment of the hindrance potential from the user's hindrance file also needs attention. However, preparing such information is a tedious, prone-to-error process which requires a considerable amount of effort and time from the user with many underlying uncertainties. In this work, we present a JAVA-based automatic scheme named MSMC-GUI (an extension from our state-of-art home-made *ab initio* program MSMC[2]) to help overcome this challenge. Specifically, this tool first examines the electronic structure output and the corresponding data (e.g., hindrance potential) provided by the user and then automatically extracts the prerequisite pieces of information needed for the thermodynamic calculation step. The robustness of the tool has been tested intensively with many different types of chemical compounds ranging from simple hydrocarbons to complex ones such as the ones containing heteroatoms or even the transition states. In brief, the GUI has demonstrated to be a reliable tool for even non-experts to obtain some of the basic configuration information of the chemical species with high confidence on-the-fly.

Keywords: *Thermodynamic Properties, Hindered Internal Rotation, Automatic Scheme, MSMC, Ab Initio*

Introduction:

Computational chemistry has matured to a point that its predictions nowadays often match the quality of experiments, whenever they are available, and it offers the opportunity to perform calculations for conditions that are difficult to explore experimentally. It also plays an important role in education. However, there is a steep learning curve for such powerful tools, especially for non-experts such as experimentalists and educators/students, that limits its contributions.[3, 4]In an attempt to make computational chemistry a routine procedure which can be used by either experts or non-experts to calculate thermodynamic properties and kinetic behaviors of a chemical system, we presented a novel process to determine the important configuration parameters of the species as well as the hindrance potential energy from the user's input file. The significant feature of our approach is that it requires very little effort from the user from the start to the end to produce a complete output file for further computation. Last but not least, our tool offers user capability to submit their jobs to High-Performance Computers (HPC) to perform thermodynamic / kinetic property calculations. Especially, our code is compatible with most of the current popular platforms; hence, it gives the users more freedom to study and investigate different types of chemical systems and species at their own disposal. In this way, our code is both flexible and reliable to serve as the scientists' favourite tool for their routine research.

Materials and Methods:

First and foremost, JAVA was chosen as our main framework due to its flexibility and platform independence. Moreover, JAVA gave us a powerful tool to design and implement the Graphical User Interface of our in-house state of art MSMC engine.[2] MSMC-GUI contains many unique functions; one of the major ones is to provide the most important parameters for the thermodynamic property calculations. Details about the MSMC code can be found elsewhere.[5]

The overall architecture of the tool contains three main phases (cf. ***Figure 1***) upon which our paper would be developed. Firstly, the data provided from the user's electronic output file (e.g., the Gaussian output file) is processed to extract some important information such as the species charge, multiplicity, geometry, the normal vibrational modes of the species, etc. After this step, our tool will analyze the hindrance potential embedded in the user's hindrance input file. Then the complete configurations of the species are produced and stored in a single file defined by ourselves, namely the .minp file. Finally, the data is organized in a formatted output file and in case the user requests for thermodynamic calculations, the file would be sent to the server for further processing. The details of each process will be discussed in-depth later.

Molecular parameters analysis:

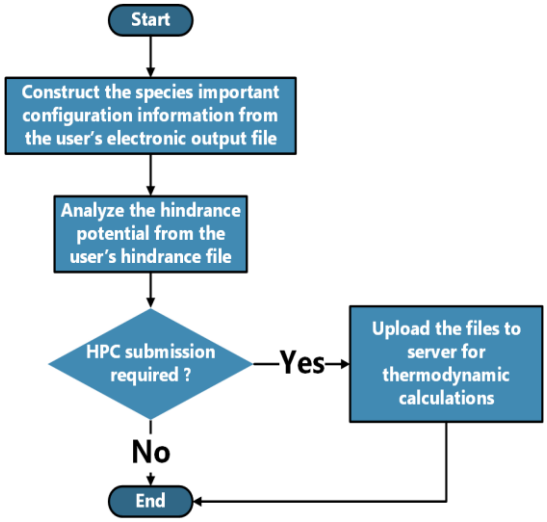


Figure 1. *The flowchart to determine the species input parameters for thermodynamic calculations*

The molecular configuration detection process is an integral process to give a big picture into the basic characteristics of the species. Among the parameters in this step, the species coordinate detection is of highest importance since it gives an insight into the arrangement of the atoms in space (cf. ***Figure 2***). To be more specific, the tool takes the coordinates of the molecules from the electronic structure output files (e.g. Gaussian output file) and produces the species structure in the Cartesian coordinate system. This process contains several steps: the coordinate extraction, the atom list creation from the raw coordinates data, and the atom connectivity setting. When this current process is done, the full geometry of the species is obtained.

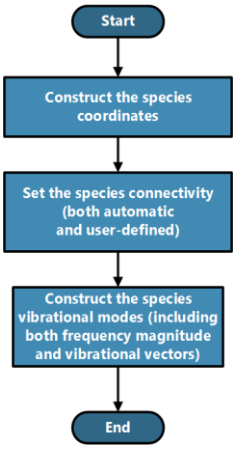


Figure 2. *Define the species geometry and vibrational modes (including frequency magnitudes and vibrational vectors).*

After the coordinates have been imported from the file, the position of each atom is placed into a list with their corresponding coordinates. In the next step, the atom connectivity is defined in the case the distance between two atoms is close enough (e.g. within a tolerance of 20% of the distance between the radii of the two atoms). Otherwise, the user has to define the connection manually by using our feature *"Active Connectivity"* to guarantee to obtain the complete geometry of the species (cf. ***Figure 3)***.

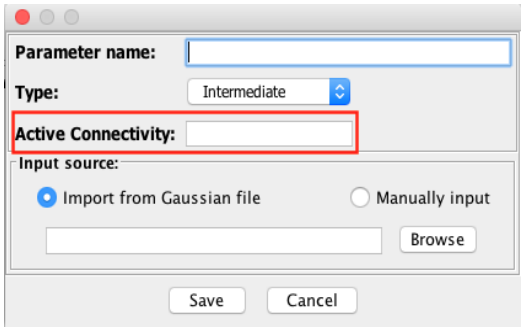


Figure 3. *Manual setting of species connectivity for complex case to aid the automatic detection*

The combination of both automatic detection process from our approach and the user defined improves the reliability in the species geometry detection. Some previous works only focused on well-defined chemical bonds,[1] but our approach can confront with much more complex cases (e.g. a transition state of species C_2H_5O in ***Figure 4***) thanks to the cooperation of users (e.g. defining the $C_1 - C_2$ bond), which results in a better approximation for later procedures.

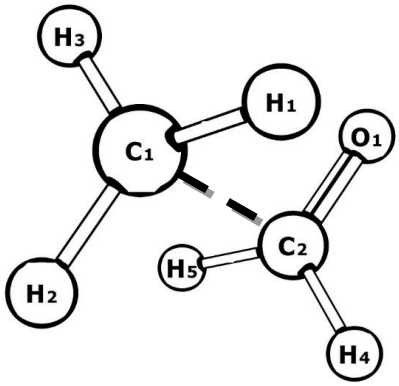


Figure 4. *The transition state of the reaction $C_2H_5O \rightarrow CH_3+CH_2O$, where the chemical bond between $C_1 - C_2$ is defined by user since the boundary value is apprx. 1.52 Å (i.e. radius of C is 0.76) but this bond is too distant (d = 2.21 Å) to be detected automatically*

The next procedure is to analyze the normal vibrational modes. Each mode generally contains two piece of information including its frequency magnitude and the vibrational vectors of each atom (cf. ***Figure 5***). In addition, the used modes would be indexed to remove the redundancy as well as complexity in this process. In practice, the elimination of used modes can reduce the complexity of the process since there are many existing cases where some modes are mixed and confusing due to symmetry.[1] These detected modes combining with the species geometry offer the users with sufficient data to prepare input for the determination of the Hindered Internal Rotations.

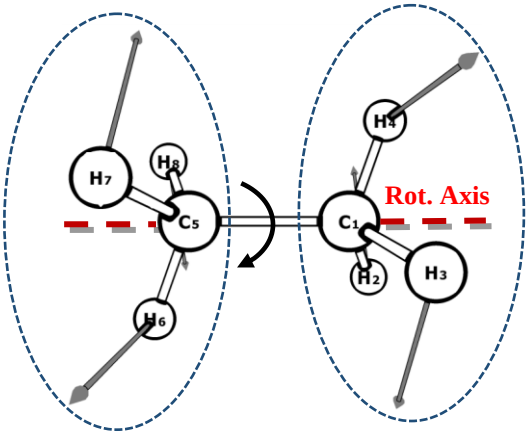


Figure 5. *The geometry and the rotational group, axis in the case of frequency = 308 cm^{-1} for C_2H_6*

After this first process, the important configuration information of the species is presented to the user. (e.g., the case of C_2H_6 is illustrated in ***Text 1***). With these data, it is sufficient to run Multi-Species Multi-Chanel (MSMC)[2] engine afterwards.

```
Charge = 0
Multiplicity = 1
EnergyLevel = -79.704198 hartree
```

Text 1. *The charge, multiplicity along with the energy level of C_2H_6 included in the .minp result file*

A. Hindrance Potential Extraction and HPC (High-Performance Computing) Submission Process

In this process, the full hindrance potential energies must first be extracted from the user's hindrance file**.** The series of hindrance potential energies is described in ***Figure 6.*** These results are summarized in our formatted output file by MSMC-GUI and then send to a HPC (e.g. MSMC engine[2]) to obtain the ultimate thermodynamics properties. In fact, our tool also provides the user the functionality to set up all the authentication information for their access to the HPC. Therefore, each time they submit their job, it will automatically remember the last log-in information to save user's considerable amount of

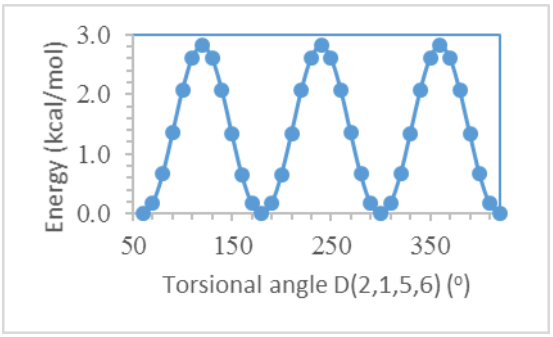


Figure 6. *The hindrance potential series corresponding to the internal rotational of the CH_3 group (frequency = 309 cm^{-1}) for ethane (C_2H_6) (cf.* ***Figure 5****)*

time. Moreover, MSMC-GUI also provides the user the ability to run their submitted file on the head node (i.e. immediate execution) or in queue (i.e. waiting for other job's termination before processing the current job). When the job is finished, a message will notify the user of the current status and all the files would be ready for user to download and use at any time (cf. ***Figure 7***). Therefore, it offers the user great opportunity to do almost everything with just several clicks using our powerful tool.

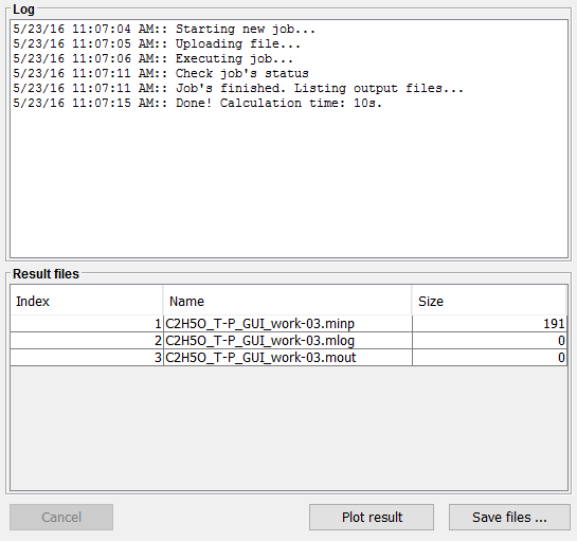


Figure 7. *The result files return from the thermodynamic calculation of MSMC engine for the reaction $CH_3 + H_2CO \rightarrow C_2H_5O \rightarrow CH_3CHO + H$*

Results and Discussion:

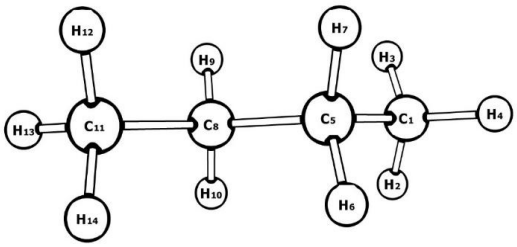
In this part, our proposed approach has been put through 100 test cases in all (e.g. 50 hydrocarbons and 50 nitrogen species) in which combinations of ring-like, multiple bonding as well as not-well defined bonding species are involved. Furthermore, the users can also take part in the correction at any time in the results produced by our method. Hence, the final results are guaranteed to be both highly accurate and flexible. To facilitate and highlight the mentioned functionalities of the tools, we only present in details n-Butane as our representative test-

case. We will conduct two following steps: (1) construct the molecular structure and analyze the vibrational modes; (2) extract the hindrance potential energies output them in our own defined .minp file (included in the supplementary materials). (3) submit the job to HPC for thermodynamic calculation. The complete list of test cases and their details can be found at the following url: https://sites.google.com/site/msmccode/test-cases/others/gui-test-cases

Test case: n-Butane (C_4H_{10})

Firstly, the geometry and the vibrational modes are manipulated and stored temporarily in our program (c.f. ***Figure 8***). To be more specific, the coordinates of each atom are calculated in term of Å while the unit of frequency is cm^{-1}. Theoretically, since n-Butane is a nonlinear species, its number of frequencies is equal to 3N – 6 where N is the number of atoms. Therefore, in this case, n-Butane has 36 frequencies. However, among these we only consider the low-frequency ones (i.e. apprx. 1000 cm^{-1} in ***Figure 8***) to reduce the calculation complexity. The MSMC-GUI shows three detected rotational groups (e.g. axis $C_1 - C_5$, $C_5 - C_8$, $C_8 - C_{11}$) which can be verified manually in ***Figure 8***.

In the last sub-process, the full Hindrance Potential Energies for each detected hindered rotor are also



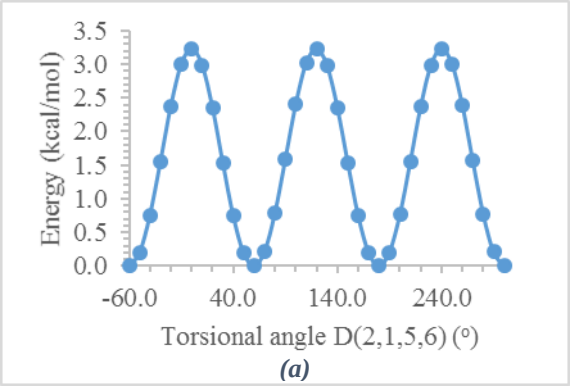
Frequencies (cm^{-1})

121	233	262	265
425	737	815	842
965	982	1016	1063

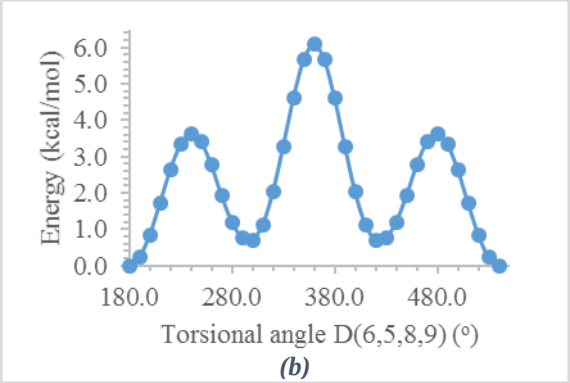
Figure 8. *Molecular representation of n-butane (C_4H_{10}) and the first 12 corresponding vibrational low-frequencies (cm^{-1}) calculated at CBS-QB3 level of theory*

indicated (cf. ***Figure 9***), then all the results are combined in an XML-based output file generated by MSMC-GUI and serve as input for thermodynamics property calculation (e.g. MSMC-engine[2] in this case). This step also marks the end of the whole parsing process of our program.

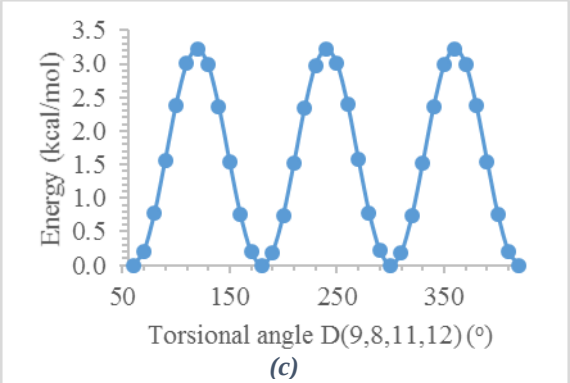
MSMC-GUI has been demonstrated to correctly provide the input for thermodynamic computation for nearly 100 test cases. Due to the highly accurate results of the parameters, the quality of the corresponding thermodynamic properties is also undeniably enhanced. Please refer to the URL (https://sites.google.com/site/msmccode/test-cases/others/gui-test-cases) for the updated and complete test-case list. It is noted that the structure of the file must follow a pre-defined format (e.g., Gaussian output file) to be successfully parsed by our tool; thus it is recommended that the users should pay close attention to the user's manual of our tool at this URL (https://sites.google.com/site/msmccode/) to avoid encounter with such problem during execution.



(a)



(b)



(c)

Figure 9*. The hindrance potential corresponding to the internal rotations* ***(a)*** *frequency = 265 cm^{-1}; (b) frequency = 121 cm^{-1}; (c) frequency = 233 cm^{-1} for n-butane (C_4H_{10}).*

Conclusion:

In this paper, a new approach has been outlined to aid the users in the determination process for the basic configuration parameters of any chemical species within a (bio-) chemical system. Besides a variety of powerful functionalities provided, our tool is remarkably simple to use and flexible enough to be editable at any time by the user. To be more specific, not only that MSMC-GUI offers the user a way to prepare input but also allow them deliver it to perform the thermodynamic calculations on the HPC and then get the final results without need of any other external program. In brief, MSMC-GUI definitely added considerable value to the computation of thermodynamics properties. In fact, more thorough study can be made to improve the accuracy and extend the real-life applications of this current study.

In the future, new features will be implemented, including (1) extend the format to other types of files not just Gaussian output file; (2) plot the results returned from the HPC thermodynamic calculations. These features are expected to be included in the next version of MSMC-GUI to make the code become more intelligent and robust tool to provide the user with all-in-one tool for their routine research on the thermodynamic properties.

Acknowledgements:

We would like to thank Ngoc N. Ho and Tuan Q. Mach from the International University, VNU-HCM for helping with the test cases. Facility and computing resources provided by the Institute for Computational Science and Technology – Ho Chi Minh City and International University, VNU-HCM are gratefully acknowledged.

References:

[1] Ayala, P.Y. and H.B. Schlegel, (1998) "Identification and treatment of internal rotation in normal mode vibrational analysis" Journal of Chemical Physics, ISSN 0021-9606, Vol. 108, No. 6, Feb 8**,** pp 2314-2325

[2] Duong, M.V., et al., (2015) "Multi-Species Multi-Channel (MSMC): An Ab Initio-based Parallel Thermodynamic and Kinetic Code for Complex Chemical Systems" International Journal of Chemical Kinetics, ISSN 05388066, Vol. 47, No. 9**,** pp 564-575

[3] Glowacki, D.R., et al., (2012) "MESMER: an open-source master equation solver for multi-energy well reactions" J Phys Chem A, ISSN 1520-5215 (Electronic)
1089-5639 (Linking), Vol. 116, No. 38, Sep 27**,** pp 9545-60

[4] Largent, R.J., W.F. Polik, and J.R. Schmidt, (2012) "Symmetrizer: Algorithmic determination of point groups in nearly symmetric molecules" J Comput Chem, ISSN 1096-987X (Electronic) 0192-8651 (Linking), Vol. 33, No. 19, Jul 15**,** pp 1637-42

[5] *https://sites.google.com/site/msmccode/*.