Shermo: A general code for calculating molecular thermochemistry properties

Tian Lu,[a],* Qinxue Chen,[a]

[a] Beijing Kein Research Center for Natural Sciences, Beijing 100022, P. R. China (http://www.keinsci.com)

* Correspondence author. E-mail: sobereva@sina.com

ORCiD
Tian Lu: 0000-0002-1822-1229
Qinxue Chen: 0000-0003-0155-2387

Abstract: Calculation of molecular thermodynamic quantities is one of the most frequently involved task in daily quantum chemistry studies. In this article, we present a general, stand-alone, powerful and flexible code named Shermo for calculating various common thermochemistry data. This code is compatible with Gaussian, ORCA, GAMESS-US and NWChem and has many unique advantages: the output information is very easy to comprehend; thermodynamic quantities can be fully decomposed to contributions of various sources; temperature and pressure can be conveniently scanned; two quasi-rigid-rotor harmonic oscillator (quasi-RRHO) models are supported to properly deal with low frequencies; different frequency scale factors can be simultaneously specified for calculating different thermodynamic quantities; conformation weighted thermodynamic data can be directly evaluated; the code can be easily run and embedded into shell script. We hope the Shermo program will bring great convenience to quantum chemists. This code can be freely obtained at http://sobereva.com/soft/shermo.

1 Introduction

Calculation of molecular thermodynamic quantities, such as enthalpy, Gibbs free energy and heat capacity, is the most frequently involved problem in daily quantum chemistry research. Due to its importance, almost all popular quantum chemistry packages have their own code to compute thermodynamic data. However, these codes often have various shortcomings and limitations, for example: (1) No better model beyond harmonic approximation for low frequency modes is supported (2) Frequency scale factors for different purposes cannot be separately specified at the same time (3) Data is inconvenient to recalculate at other temperatures and pressures (4) Temperature and pressure cannot be easily scanned (5) Contribution to thermodynamic data from various vibrational modes cannot be individually printed (6) Conformation weighting cannot be directly taken into account (7) Output information of some programs is complicated and thus difficult for beginners to understand. Since these problems have caused great inconveniences to researchers in the field of quantum chemistry, we believe it is quite necessary to develop a powerful, flexible and easy-to-use code for calculating molecular thermodynamic quantities to address aforementioned issues. Due to this consideration, a code named Shermo was recently developed by us and will be introduced in this article. Briefly speaking, Shermo is a stand-alone and versatile molecular thermodynamic calculation code based on output file of frequency analysis task of mainstream quantum chemistry programs, including Gaussian,[1] ORCA,[2] GAMESS-US[3] and NWChem.[4] Since Shermo makes molecular thermochemistry analysis unprecedentedly convenient, we believe this code will become a new important member in many quantum chemists' toolbox.

In Section 2, we will describe the model employed by Shermo for calculating thermodynamic data, in Section 3 various features of Shermo will be introduced. A few application examples of using Shermo to study practical chemical problems will be given in Section 4. Finally, in Section 5 we summarize this article.

2 Calculation model

Shermo calculates molecular thermodynamic data mostly based on rigid-rotor harmonic oscillator model under ideal gas assumption, which is also employed in thermochemistry analysis of almost all quantum chemistry codes. This model has been very detailedly documented in manual of Shermo and also introduced in many books,[5, 6] therefore full introduction of this model will not be presented here, only key points will be mentioned.

Shermo is able to calculate all common molecular thermodynamic data, including internal energy (U), enthalpy (H), Gibbs free energy (G), entropy (S), heat capacity of constant volume (C_V) and constant pressure (C_P), as well as molecular partition function (q). All of them can be decomposed as contributions from translation, rotation, vibration and electronic excitation by ignoring their couplings, which usually is a reasonable assumption. In Shermo, these contributions can be printed separately.

The rotation contribution is dependent of molecular principal moments of inertia (I) and rotational symmetry number (σ). The I is computed based on atomic masses and current geometry, while σ is determined by point group of current molecule. Obviously, correct identification of point group is crucial. In Shermo, point group is automatically detected by our modified version of SYVA

code,[7] which is fairly robust and fast.

The description of the vibration part has key influence on calculation accuracy of thermodynamic quantities. Usually, the frequencies involved in the calculation are obtained based on harmonic approximation, which is the simplest model of representing molecular vibrations. In this case, the vibrational contribution to various thermodynamic data can be calculated as follows

$$U(0)_{\text{vib}} = \text{ZPE} = \frac{1}{2}\sum_i h\nu_i$$

$$U_{0 \to T}^{\text{vib}} \equiv U(T)_{\text{vib}} - U(0)_{\text{vib}} = RT\sum_i \left(\frac{h\nu_i}{kT}\right)\frac{e^{-h\nu_i/(kT)}}{1 - e^{-h\nu_i/(kT)}}$$

$$C_{\text{V,vib}}(T) = R\sum_i \left(\frac{h\nu_i}{kT}\right)^2 \frac{e^{-h\nu_i/(kT)}}{[1 - e^{-h\nu_i/(kT)}]^2} \tag{1}$$

$$S_{\text{vib}}(T) = R\sum_i \left\{\frac{h\nu_i}{kT}\frac{e^{-h\nu_i/(kT)}}{1 - e^{-h\nu_i/(kT)}} - \ln[1 - e^{-h\nu_i/(kT)}]\right\}$$

where R is ideal gas constant, k is Boltzmann constant, T stands for temperature, ν_i denotes vibrational frequency of mode i. As can be clearly seen, all of them can be viewed as summation of contribution of various vibrational modes. In order to make analysis of the nature of the quantities feasible, Shermo allows printing individual contribution coming from different vibrational modes.

It is well known that neglect of anharmonic effect and inaccurate representation of potential energy surface must lead to significant error in thermodynamic quantities. Fortunately, this kind of error is relatively systematical and thus can be largely eliminated by applying pre-fitted scale factor for harmonic frequencies prior to calculation of various thermochemistry data. Frequency scale factors (λ) corresponding to various calculation levels have been fitted in many literatures, some of them provided λ for zero-point energy (λ_{ZPE}),[8-13] while some other ones also presented λ for entropy (λ_{S}) and heating contribution to internal energy or enthalpy (λ_{heat}).[14-17] Usually both λ_{S} and λ_{heat} are close to 1.0, while λ_{ZPE} deviates from 1.0 notably, therefore applying a global frequency scale factor in all thermodynamic data calculations is evidently inappropriate. In Shermo, the λ_{ZPE}, λ_{S}, λ_{heat} and λ for heat capacity can be simultaneously specified and properly considered in the calculation of various thermodynamic data.

The harmonic approximation treatment for vibrations, which is often referred to as rigid-rotor harmonic oscillator (RRHO) model, breaks down for very low frequency modes. This point can be easily understood from below equations, which show contribution to entropy and heating contribution to internal energy due to vibration mode i.

$$S^i(T) = R\left\{\frac{h\nu_i}{kT}\frac{1}{e^{h\nu_i/(kT)} - 1} - \ln[1 - e^{-h\nu_i/(kT)}]\right\}$$

$$U_{0 \to T}^i(T) = \left(\frac{h\nu_i}{kT}\right)\frac{RT}{e^{h\nu_i/(kT)} - 1} \tag{2}$$

Clearly, as the ν_i approaches zero, the above two quantities diverge because their denominators tend to vanish. In order to make calculation result for a system containing low frequencies more reasonable, Shermo supports two popular methods to tackle this issue and they are usually referred to as quasi-RRHO (QRRHO). The first QRRHO treatment is the one originally suggested by Truhlar and coworkers, see refs. [18, 19] for example. This method artificially raises all frequencies that

lower than 100 cm^{-1} to 100 cm^{-1}, the cutoff can be customized in Shermo. Recently, in an article of studying aqueous pK_a prediction,[20] it was shown that for small molecules, the accuracy of calculation result is insensitive to the choice of the cutoff, while for molecular complexes, the original cutoff (100 cm^{-1}) performs best.

Another QRRHO treatment supported by Shermo was proposed by Grimme,[21] and it only applies to entropy. Essentially, this method performs interpolation between the vibrational entropy calculated under harmonic approximation (S^{RRHO}) and the one calculated based on free-rotor model (S^{FR}), the latter can be expressed as

$$S_{\mathrm{FR}} = R\left(\frac{1}{2} + \ln\sqrt{\frac{8\pi^3 \mu' kT}{h^2}} \right)$$

(3)

with

$$\mu' = \frac{\mu B_{\mathrm{av}}}{\mu + B_{\mathrm{av}}} \qquad \mu = \frac{h}{8\pi^2 \nu}$$

(4)

where μ is moment of inertia of a free-rotor with same frequency as the vibrational mode, μ' corresponds to effective moment of inertia. The B_{av} denotes average molecular moment of inertia and was empirically chosen to be 10^{-44} kg·m^2. According to our experience, the result of the two aforementioned QRRHO models do not differ with each other notably, while both of them may improve accuracy of free energy computed by RRHO model considerably for highly flexible systems, which are usually accompanied by large number of low frequencies. As indicated in ref. [21], the error of RRHO on large organic systems containing 300~400 atoms can easily reach as high as 3~4 kcal/mol at standard condition. Therefore, we always suggest using one of the QRRHO models in Shermo instead of RRHO. Relatively speaking, we prefer to employ the Grimme's method since it has more rigorous physical background than the way of artificially raising low frequencies.

Electronic excitation also contributes to thermodynamic quantities and can be considered in the calculation as long as user specifies energy levels in Shermo; for example, its contribution to entropy is expressed as

$$S_{\mathrm{ele}} = R\ln\left(\sum_i g_i e^{-\varepsilon_i/(kT)} \right) + R\frac{\sum_i g_i[\varepsilon_i/(kT)]e^{-\varepsilon_i/(kT)}}{\sum_i g_i e^{-\varepsilon_i/(kT)}}$$

(5)

where ε_i is electronic energy level i with respect to ground state, g_i is degeneracy of level i. Note that contribution of electronic excitation is not considered in thermodynamic analysis code of almost all quantum chemistry programs, since additional calculation is needed to evaluate excitation energies, and energy separation between lowest excited state and ground state is usually as large as several eV, in this case contribution due to electron excitations is fully negligible. However, extremely low-lying excited states, for example, the ones caused by spin-orbit coupling splitting of ground state,[5] should be taken into account in thermodynamic data calculation. When ground state is degenerate solely due to electron spin, Eq. 5 reduces to

$$S_{\mathrm{ele}} = R\ln g_0$$

(6)

where g_0 in this context corresponds to spin multiplicity. Since Shermo automatically identifies spin multiplicity from output file of quantum chemistry program, this kind of electron contribution to

entropy is always automatically considered.

Flexible molecules usually have many thermal accessible conformations, while molecular complexes usually have many thermal accessible configurations. Appropriate consideration of these conformations or configurations is crucial for obtaining reliable and meaningful thermodynamic data. The conformation (or configuration) weighted thermodynamic data can be obtained via the following way

$$U^{\mathrm{w}} = \sum_i^{\mathrm{conf}} p_i U_i \qquad C_V^{\mathrm{w}} = \sum_i^{\mathrm{conf}} p_i C_{V,i} \qquad S^{\mathrm{w}} = S^{\mathrm{conf}} + \sum_i^{\mathrm{conf}} p_i S_i \qquad (7)$$

where p stands for Boltzmann weight, which can be determined based on relative Gibbs free energies between the conformers.[22] The quantities with subscript i denote the data calculated for conformer i in usual way. The S^{conf} is known as conformation entropy and can be evaluated as

$$S^{\mathrm{conf}} = -R \sum_i^{\mathrm{conf}} p_i \ln p_i \qquad (8)$$

The conformation weighted Gibbs free energy can be evaluated as $G^{\mathrm{w}} = H^{\mathrm{w}} - TS^{\mathrm{w}}$. In Shermo, thermochemistry quantities and Boltzmann weights for all conformers as well as conformation weighted thermodynamic data can be directly outputted, users just need to provide a list containing path of output files of frequency analysis task for all thermal accessible conformers.

3 Characteristics of Shermo

3.1 Basic characteristics

Shermo is a code written in Fortran, its executable file of Windows and Linux platforms as well as a detailed manual can be freely downloaded at http://sobereva.com/soft/shermo. Source code can also be freely obtained upon request from us.

Shermo can be run very easily without any installation step or preparation of running environment. After booting up Shermo, running parameters will be loaded from "settings.ini" file, and user will be asked to input path of the input file, then molecular information and various thermodynamic quantities will be printed on screen.

Shermo can also run purely via command line, for example:

```
Shermo /nico/maki.out -P 5 -sclZPE 0.96 -ilowfreq 2 > ll.txt
```

This command invokes Shermo to deal with "/nico/maki.out", the output information is redirected to "ll.txt", the pressure is set to be 5 atm, the frequency scale factor for ZPE is set to be 0.96, and Grimme's QRRHO treatment for entropy is employed. All other parameters that not explicitly specified will employ the values defined in the "settings.ini" file. Clearly, Shermo can be conveniently embedded into shell script to automatically process large amount of files.

3.2 Input file

Output file of harmonic frequency analysis task of Gaussian, ORCA, GAMESS-US and NWChem can be directly used as input file of Shermo. Geometry optimization can also be simultaneously conducted, for example, output file of "opt freq" task of Gaussian and ORCA is also acceptable. Shermo loads all necessary information that needed by calculating thermochemistry

quantities from input file, including electronic energy, spin multiplicity, elements and coordinates of all atoms, and harmonic frequencies.

For flexibility, Shermo also defined a private format with "shm" as suffix. This file contains everything that needed by Shermo calculation, thus it can also be employed as input file. The shm format is concise and clear, thus one can easily manually create or modify it. The shm file for a typical molecule HCN is illustrated in Figure 1. The shm file can also be directly generated by Shermo after loading output file of quantum chemistry program.

```
*E   Electronic energy (a.u.)
      -91.6752090000
*wavenum  Wavenumbers (cm-1). Negative value means imaginary frequency
  952.5433
  952.5433
 2541.2397
 3915.9410
*atoms  Information of all atoms: Name, mass (amu), X, Y, Z (Angstrom)
H    1.007830   0.000000   0.000000   -1.569030
C   12.000000   0.000000   0.000000   -0.498995
N   14.003070   0.000000   0.000000    0.654010
*elevel  Relative energy w.r.t. ground state (eV) and level degeneracy
      0.000000   1
```

Figure 1 shm file of HCN molecule. Blue texts are comments

Due to the generality of the shm format, users of other quantum chemistry programs may also make use of Shermo to compute thermodynamic data, what they need to do is extracting data from output file of the programs manually or through their own code and then write the data into a file with the shm format.

To compute conformation weighted data, a list file should be prepared and used as input file, each of its lines records path of a frequency analysis output file of a conformer. For example:

```
D:\yohane\conf1.out;233.463126
D:\yohane\conf2.out;233.476224
D:\yohane\conf3.out;233.458291
```

The values after semicolon are electronic energies in a.u., which usually should be calculated at a level better than frequency analysis, since electronic energy dominates internal energy, enthalpy and Gibbs free energy and meantime it is highly sensitive to calculation level. If the electronic energies are not explicitly provided, they will be directly loaded from the output files of frequency analysis task.

3.3 Running parameters

Running parameters of Shermo are loaded from the "settings.ini" file or directly specified via arguments in command line. The parameters are summarized in Table 1, detailed introduction can be found in manual.

Table 1 Running parameters in Shermo program.

E	Electronic energy. If not explicitly specified, it will be loaded from input file
prtvib	If printing contribution of each vibrational mode
T	Temperature used for calculation
P	Pressure used for calculation
sclZPE	Frequency scale factor for ZPE
sclheat	Frequency scale factor for heating contribution to U, namely $U(T)$-$U(0)$
sclS	Frequency scale factor for entropy
sclCV	Frequency scale factor for heat capacity
ilowfreq	Model for dealing with low frequencies
ravib	Cutoff of raising low frequencies
outshm	If exporting .shm file after loading output file of quantum chemistry program
defmass	Type of default atomic masses
modmass	Modification on default atomic masses

By default, atomic masses employed in the calculation are loaded from output file of quantum chemistry program, user can also use the "defmass" parameter to request Shermo to employ average atomic masses or masses of principal isotopes. Atomic masses for specific atoms can be further modified by the "modmass" parameter.

Using Shermo to scan temperature and pressure is straightforward, users simply need to specify lower limit, upper limit and stepsize. For example, specifying "T= 50,200,10" in the "settings.ini" file or adding "-T 50,200,10" argument in running command will request Shermo to output various thermodynamic data every 10 K from 50 to 200 K. Two-dimension scan for both temperature and pressure is also supported.

4 Examples

In this section, we present two simple examples to briefly illustrate use and practical value of the Shermo program. Shermo 2.0, Gaussian 16. A.03[1] and ORCA 4.2.1[2] were utilized in the examples. Molecular structure was plotted by ChimeraX program.[23]

4.1 Formaldehyde

This example studies thermodynamic data of formaldehyde. Optimization and harmonic frequency analysis were finished at the popular B3LYP/6-31G* level,[24, 25] after that electronic energy was evaluated at the significantly more accurate CCSD(T)/cc-pVQZ level,[26, 27] the result was found to be -114.368896 a.u. In order to calculate thermodynamic quantities for the formaldehyde at 350 K, one can simply run this command:

```
Shermo H2CO_optfreq.out -sclZPE 0.9806 -T 350 -E -114.368896
```

where 0.9806 is the frequency scale factor fitted for ZPE at B3LYP/6-31G* level.[17] Ref. [17] also presented scale factor for entropy and heating contribution to enthalpy (1.0015 and 0.9989, respectively), they are not explicitly specified here because they are quite close to the default value (1.0). Note that the default pressure is 1 atm and the default calculation model is RRHO.

After running the above command, detailed molecular information will be printed on screen,

then contribution from various sources to thermodynamic data will be shown. For example, the vibration contribution:

```
Vibrational q(V=0):    1.014697E+000
Vibrational q(bot):    3.145066E-011
Vibrational U(T)-U(0):    0.226 kJ/mol    0.054 kcal/mol    =H(T)-H(0)
Vibrational U:    69.275 kJ/mol    16.557 kcal/mol    =H
Vibrational S:    0.767 J/mol/K    0.183 cal/mol/K    -TS:    0.064 kcal/mol
Vibrational CV:    3.500 J/mol/K    0.837 cal/mol/K    =CP
Zero-point energy (ZPE):    69.05 kJ/mol,    16.50 kcal/mol    0.026300 a.u.
```

Finally, the total result is given:

```
Total q(V=0):    5.316493E+033
Total q(bot):    1.647853E+023
Total q(V=0)/NA:    8.828242E+009
Total q(bot)/NA:    2.736323E-001
Total CV:    28.443 J/mol/K    6.798 cal/mol/K
Total CP:    36.758 J/mol/K    8.785 cal/mol/K
Total S:    224.315 J/mol/K    53.613 cal/mol/K    -TS:    18.764 kcal/mol
Thermal correction to U:    78.006 kJ/mol    18.644 kcal/mol    0.029711 a.u.
Thermal correction to H:    80.916 kJ/mol    19.339 kcal/mol    0.030819 a.u.
Thermal correction to G:    2.405 kJ/mol    0.575 kcal/mol    0.000916 a.u.
Electronic energy:    -114.3688960 a.u.
Sum of electronic energy and ZPE, namely U/H/G at 0 K:    -114.3425965 a.u.
Sum of electronic energy and thermal correction to U:    -114.3391852 a.u.
Sum of electronic energy and thermal correction to H:    -114.3380768 a.u.
Sum of electronic energy and thermal correction to G:    -114.3679799 a.u.
```

It can be seen that all information that involved in daily thermochemistry study has been shown in a very compact and clear format.

The important vibration contribution to thermodynamic quantities can be decomposed to better study its nature. If "-prtvib 1" argument is added to the running command or "prtvib" in the "settings.ini" is set to 1, then contributions from every vibrational mode will be shown as follows

Mode	Wavenumber	Freq	Vib. Temp.	q(V=0)	q(bot)
	cm^-1	GHz	K		
1	1198.95	0.35944E+05	1725.02	1.00728919	0.08568734
2	1280.05	0.38375E+05	1841.71	1.00521179	0.07238064
3	1563.19	0.46863E+05	2249.07	1.00162166	0.04030246
4	1850.19	0.55467E+05	2662.00	1.00049784	0.02231794
5	2914.57	0.87377E+05	4193.41	1.00000626	0.00250221
6	2965.59	0.88906E+05	4266.81	1.00000508	0.00225311

Mode	Wavenumber	ZPE	U(T)-U(0)	U(T)	CV(T)	S(T)
	cm^-1	kcal/mol	kcal/mol	kcal/mol	cal/mol/K	cal/mol/K
1	1198.95	1.68073	0.02499	1.70572	0.35443	0.08585
2	1280.05	1.79443	0.01907	1.81350	0.28827	0.06484
3	1563.19	2.19133	0.00725	2.19858	0.13328	0.02393

4	1850.19	2.59366	0.00263	2.59630	0.05726	0.00851
5	2914.57	4.08576	0.00005	4.08581	0.00179	0.00016
6	2965.59	4.15727	0.00004	4.15732	0.00150	0.00013

From the data a tendency can be observed, that is the lower the frequency of a mode, the larger its contribution to $U(T)$-$U(0)$, C_V and S. In contrast, the higher the frequency of a mode, the larger its contribution to ZPE and $U(T)$.

In order to study variation of various thermodynamic quantities with respect to change of temperature, for example, from 50 to 1000 K with stepsize of 5 K, one can add "-T 50,1000,10" to the running command. After calculation, two files "scan_SCq.txt" and "scan_UHG.txt" will be generated in current folder, which contain thermodynamic data at various temperatures and can be easily plotted via data visualization softwares such as gnuplot[28] and Origin programs.[29] Fig. 2 exhibits variation of thermal corrections to internal energy, enthalpy and Gibbs free energy with respect to temperature.
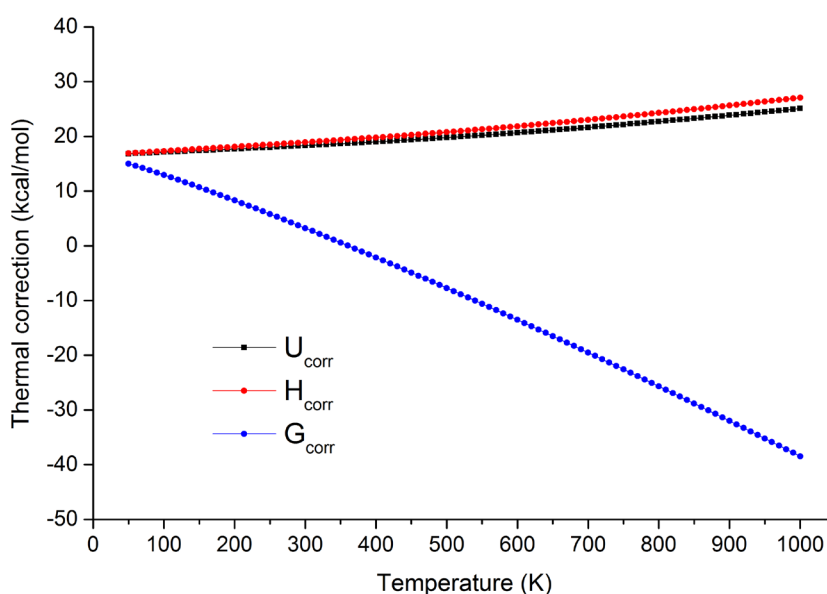

Figure 2 Variation of thermal corrections to internal energy (U_{corr}), enthalpy (H_{corr}) and Gibbs free energy (G_{corr}) with respect to temperature for formaldehyde.

4.2 Remdesivir

Remdesivir is a broad-spectrum antiviral medication and recently shown to be the most promising treatment for COVID-19 virus.[30, 31] Our recent conformation search study of remdesivir by our Molclus program[32] determined a batch of its low-energy conformations in aqueous environment.[33] The structure of the most stable conformer of the remdesivir is shown in Fig. 3.
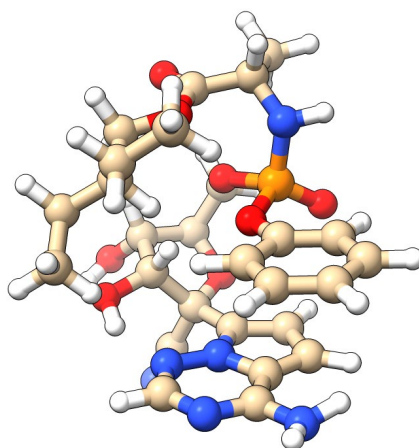
Figure 3 Structure of the most stable conformation of remdesivir in water.

The geometry optimization and harmonic frequency analysis of remdesivir were conducted at B3LYP-D3(BJ)/6-31G* level[24, 25, 34] by Gaussian. Since remdesivir is a highly flexible molecule with huge number of rotatable bonds, there are as many as 15 and 21 vibrational modes with frequencies lower than 100 and 150 cm^{-1}, respectively. As mentioned earlier, the common RRHO treatment is incapable of reasonably representing contributions to thermodynamic quantities from very low frequency modes like these. In this example, several thermodynamic data were computed at standard condition for remdesivir using RRHO and the two QRRHO models supported by Shermo, the results are summarized in Table 2. To activate the QRRHO models, users only need to add "-ilowfreq 1" or "-ilowfreq 2" argument to command-line or make corresponding modification on the "settings.ini" file.

Table 2 Several thermodynamic data of remdesivir calculated by various models of representing vibrational contribution

Model	$U(T)$-$U(0)$ (kcal/mol)	-TS (kcal/mol)	C_V (cal/mol/K)
RRHO	24.1	49.1	151.3
QRRHO(1) [a]	23.2	42.4	150.9
QRRHO(2) [b]	24.1	43.4	151.3

[a] Raising all vibrational frequencies lower than 100 cm^{-1} to 100 cm^{-1}
[b] Grimme's entropy interpolation between harmonic and free-rotor approximations. Note that this treatment does not alter $U(T)$-$U(0)$ and C_V of RRHO.

From Table 2 it can be seen that the $U(T)$-$U(0)$ and C_V calculated under RRHO and QRRHO(1) model do not differ with each other markedly, however, both the two QRRHO models changed the -TS term of RRHO by more than 10%, rendering the importance of applying QRRHO model when calculating Gibbs free energy for very flexible systems like remdesivir. Since the two QRRHO models gave similar results, both of them could be regarded as reasonable and can thus be safely used in practice.

Since flexible systems usually have multiple thermal accessible conformations, the effect of conformation weighting should be properly taken into account in the calculation of their thermodynamic data. Gibbs free energies for a batch of low-energy conformers of remdesivir were previously determined by us using thermal correction at B3LYP-D3(BJ)/6-31G* level and electronic energies evaluated using the robust double-hybrid functional PWPB95-D3(BJ)[35] in combination with high quality def2-QZVPP basis set.[36] For simplicity, in the calculation of conformation weighted

thermochemistry quantities in this example, only the three conformers with lowest Gibbs free energies are taken into account. After manually creating a conformation list file as mentioned in Section 3.2 and taking it as input file to run Shermo, the thermodynamic quantities and Boltzmann weights of all conformers, as well as conformation weighted data are immediately printed out, as shown below

```
    #System       U              H              G            S          CV
               a.u.           a.u.           a.u.        J/mol/K     J/mol/K
        1   -2320.713764   -2320.712819   -2320.821265    954.971     657.837
        2   -2320.709745   -2320.708801   -2320.817499    957.191     657.295
        3   -2320.715999   -2320.715054   -2320.822180    943.341     654.902


    System   1    Relative G=    2.401 kJ/mol    Boltzmann weight=  27.380 %
    System   2    Relative G=   12.289 kJ/mol    Boltzmann weight=   0.507 %
    System   3    Relative G=    0.000 kJ/mol    Boltzmann weight=  72.113 %


    Conformation weighted data:
    Electronic energy:     -2321.369269 a.u.
    U:      -2320.715355 a.u.
    H:      -2320.714411 a.u.
    G:      -2320.822488 a.u.
    S:        951.727 J/mol/K    Conformation entropy:      5.132 J/mol/K
    CV:       655.718 J/mol/K
    CP:       664.032 J/mol/K
```

From the data it can be seen that the conformer 3 plays a dominant role, however, the conformer 1 can never be ignored, since its population (27.38%) is also prominent. The conformer 2 in fact can be ignored, because its Boltzmann weight is close to zero at the current condition (298.15 K and 1 atm) due to its relatively high free energy.

5 Summary

In this article, we introduced our newly developed code named Shermo, which is designed as a general tool for calculating common molecular thermodynamic quantities. In addition, the theoretical backgrounds involved in Shermo are briefly described, and practical application examples are given. This very user-friendly program makes thermochemistry analysis at different temperatures and pressures as well as with various frequency scale factors much more convenient than ever. Shermo also brings many additional advantages, for example, blackbox QRRHO model can be easily applied to low frequencies, conformation or configuration weighted quantities can be directly obtained, various components of thermodynamic quantities can be studied individually. Due to its usefulness, it is expected that the Shermo program will become a popular code for thermodynamic analysis of molecular systems.

Finally, we would like to mention that although using Shermo in combination with common quantum chemistry programs is able to reasonably study thermodynamic quantities for most majority of molecular systems, if there are very strict requirements on the accuracy of the data, sometimes more advanced models may be needed to further improve representation of the system, which are beyond current scope of the Shermo program. For example, anharmonic effect of certain

vibrational modes may need to be explicitly represented,[37] torsional motions of some groups or fragments should be represented as hindered rotors,[38] strongly coupled torsions may be better represented via multistructural torsional method.[39]

References

1. M. J. Frisch; G. W. Trucks; H. B. Schlegel; G. E. Scuseria; M. A. Robb; J. R. Cheeseman; G. Scalmani; V. Barone; G. A. Petersson; H. Nakatsuji; X. Li; M. Caricato; A. V. Marenich; J. Bloino; B. G. Janesko; R. Gomperts; B. Mennucci; H. P. Hratchian; J. V. Ortiz; A. F. Izmaylov; J. L. Sonnenberg; Williams; F. Ding; F. Lipparini; F. Egidi; J. Goings; B. Peng; A. Petrone; T. Henderson; D. Ranasinghe; V. G. Zakrzewski; J. Gao; N. Rega; G. Zheng; W. Liang; M. Hada; M. Ehara; K. Toyota; R. Fukuda; J. Hasegawa; M. Ishida; T. Nakajima; Y. Honda; O. Kitao; H. Nakai; T. Vreven; K. Throssell; J. A. Montgomery Jr.; J. E. Peralta; F. Ogliaro; M. J. Bearpark; J. J. Heyd; E. N. Brothers; K. N. Kudin; V. N. Staroverov; T. A. Keith; R. Kobayashi; J. Normand; K. Raghavachari; A. P. Rendell; J. C. Burant; S. S. Iyengar; J. Tomasi; M. Cossi; J. M. Millam; M. Klene; C. Adamo; R. Cammi; J. W. Ochterski; R. L. Martin; K. Morokuma; O. Farkas; J. B. Foresman; D. J. Fox. *Gaussian 16 A.03*, Wallingford, CT, 2016.
2. F. Neese. Software update: the ORCA program system, version 4.0. *WIREs Comput. Mol. Sci.*, **8**, e1327 (2018). DOI: 10.1002/wcms.1327.
3. M. W. Schmidt; K. K. Baldridge; J. A. Boatz; S. T. Elbert; M. S. Gordon; J. H. Jensen; S. Koseki; N. Matsunaga; K. A. Nguyen; S. Su; T. L. Windus; M. Dupuis; J. A. Montgomery. General atomic and molecular electronic structure system. *J. Comput. Chem.*, **14**, 1347 (1993). DOI: 10.1002/jcc.540141112.
4. H. J. J. van Dam; W. A. de Jong; E. Bylaska; N. Govind; K. Kowalski; T. P. Straatsma; M. Valiev. NWChem: scalable parallel computational chemistry. *WIREs: Comp. Mol. Sci.*, **1**, 888 (2011). DOI: 10.1002/wcms.62.
5. K. K. Irikura. Appendix B: Essential Statistical Thermodynamics. In *Computational Thermochemistry: Prediction and Estimation of Molecular Thermodynamics (ACS Symposium Series 677)*, American Chemical Society: Washington, DC, 1998.
6. F. Jensen. *Introduction to Computational Chemistry*. 3rd ed.; John Wiley & Sons: West Sussex, 2017; pp 454-459.
7. L. Gyevi-Nagy; G. Tasi. SYVA: A program to analyze symmetry of molecules based on vector algebra. *Comput. Phys. Commun.*, **215**, 156 (2017). DOI: 10.1016/j.cpc.2017.01.019.
8. I. M. Alecu; J. Zheng; Y. Zhao; D. G. Truhlar. Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries. *J. Chem. Theory Comput.*, **6**, 2872 (2010). DOI: 10.1021/ct100326h.
9. Y. Tantirungrotechai; K. Phanasant; S. Roddecha; P. Surawatanawong; V. Sutthikhum; J. Limtrakul. Scaling factors for vibrational frequencies and zero-point vibrational energies of some recently developed exchange-correlation functionals. *J. Mol. Struct. (THEOCHEM)*, **760**, 189 (2006). DOI: 10.1016/j.theochem.2005.12.007.
10. M. K. Kesharwani; B. Brauer; J. M. L. Martin. Frequency and Zero-Point Vibrational Energy Scale Factors for Double-Hybrid Density Functionals (and Other Selected Methods): Can Anharmonic Force Fields Be Avoided? *J. Phys. Chem. A*, **119**, 1701 (2015). DOI: 10.1021/jp508422u.
11. M. P. Andersson; P. Uvdal. New Scale Factors for Harmonic Vibrational Frequencies Using the B3LYP Density Functional Method with the Triple-ζ Basis Set 6-311+G(d,p). *J. Phys. Chem. A*, **109**, 2937 (2005). DOI: 10.1021/jp045733a.

12. J. A. Pople; A. P. Scott; M. W. Wong; L. Radom. Scaling Factors for Obtaining Fundamental Vibrational Frequencies and Zero-Point Energies from HF/6–31G* and MP2/6–31G* Harmonic Frequencies. *Isr. J. Chem.*, **33**, 345 (1993). DOI: 10.1002/ijch.199300041.

13. R. S. Grev; C. L. Janssen; H. F. Schaefer. Concerning zero‑point vibrational energy corrections to electronic energies. *J. Chem. Phys.*, **95**, 5128 (1991). DOI: 10.1063/1.461680.

14. B. Chan; L. Radom. Frequency Scale Factors for Some Double-Hybrid Density Functional Theory Procedures: Accurate Thermochemical Components for High-Level Composite Protocols. *J. Chem. Theory Comput.*, **12**, 3774 (2016). DOI: 10.1021/acs.jctc.6b00554.

15. J. P. Merrick; D. Moran; L. Radom. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A*, **111**, 11683 (2007). DOI: 10.1021/jp073974n.

16. P. Sinha; S. E. Boesch; C. Gu; R. A. Wheeler; A. K. Wilson. Harmonic Vibrational Frequencies: Scaling Factors for HF, B3LYP, and MP2 Methods in Combination with Correlation Consistent Basis Sets. *J. Phys. Chem. A*, **108**, 9213 (2004). DOI: 10.1021/jp048233q.

17. A. P. Scott; L. Radom. Harmonic Vibrational Frequencies: An Evaluation of Hartree−Fock, Møller−Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.*, **100**, 16502 (1996). DOI: 10.1021/jp960976r.

18. R. F. Ribeiro; A. V. Marenich; C. J. Cramer; D. G. Truhlar. Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *J. Phys. Chem. B*, **115**, 14556 (2011). DOI: 10.1021/jp205508z.

19. A. V. Marenich; W. Ding; C. J. Cramer; D. G. Truhlar. Resolution of a Challenge for Solvation Modeling: Calculation of Dicarboxylic Acid Dissociation Constants Using Mixed Discrete–Continuum Solvation Models. *J. Phys. Chem. Lett.*, **3**, 1437 (2012). DOI: 10.1021/jz300416r.

20. L. Xu; M. L. Coote. Methods To Improve the Calculations of Solvation Model Density Solvation Free Energies and Associated Aqueous pKa Values: Comparison between Choosing an Optimal Theoretical Level, Solute Cavity Scaling, and Using Explicit Solvent Molecules. *J. Phys. Chem. A*, **123**, 7430 (2019). DOI: 10.1021/acs.jpca.9b04920.

21. S. Grimme. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem.-Eur. J.*, **18**, 9955 (2012). DOI: 10.1002/chem.201200497.

22. T. Lu; S. Manzetti. Wavefunction and reactivity study of benzo[a]pyrene diol epoxide and its enantiomeric forms. *Struct. Chem.*, **25**, 1521 (2014). DOI: 10.1007/s11224-014-0430-6.

23. T. D. Goddard; C. C. Huang; E. C. Meng; E. F. Pettersen; G. S. Couch; J. H. Morris; T. E. Ferrin. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.*, **27**, 14 (2018). DOI: 10.1002/pro.3235.

24. P. J. Stephens; F. J. Devlin; C. F. Chabalowski; M. J. Frisch. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.*, **98**, 11623 (1994). DOI: 10.1021/j100096a001.

25. P. C. Hariharan; J. A. Pople. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chem. Acc.*, **28**, 213 (1973).

26. J. T. H. Dunning. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.*, **90**, 1007 (1989).

27. A. Szabo; N. S. Ostlund. *Modern Quantum Chemistry*. Dover Publications: New York, 1989.

28. gnuplot program, http://www.gnuplot.info (accessed on May 15, 2020).

29. Origin program, https://www.originlab.com (accessed on May 15, 2020).

30. Remdesivir, https://en.wikipedia.org/wiki/Remdesivir (accessed on May 15, 2020).

31. C. Scavone; S. Brusco; M. Bertini; L. Sportiello; C. Rafaniello; A. Zoccoli; L. Berrino; G. Racagni; F. Rossi; A. Capuano. Current pharmacological treatments for COVID-19: what's next? *Br. J. Pharmacol.* (2020). DOI: 10.1111/bph.15072.

32. Tian Lu, molclus program, http://www.keinsci.com/research/molclus.html (accessed on May 15, 2020).

33. Tian Lu, Conformation search for remdesivir using Molclus program in combination with molecular dynamics simulation by xtb program, http://bbs.keinsci.com/thread-16255-1-1.html (accessed on May 15, 2020).

34. S. Grimme; J. Antony; S. Ehrlich; H. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, **132**, 154104 (2010).

35. L. Goerigk; S. Grimme. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.*, **7**, 291 (2011). DOI: 10.1021/ct100466k.

36. F. Weigend; R. Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, **7**, 3297 (2005). DOI: 10.1039/b508541a.

37. V. Barone. Anharmonic vibrational properties by a fully automated second-order perturbative approach, **122**, 014108 (2005). DOI: 10.1063/1.1824881.

38. P. Y. Ayala; H. B. Schlegel. Identification and treatment of internal rotation in normal mode vibrational analysis. *J. Chem. Phys.*, **108**, 2314 (1998). DOI: 10.1063/1.475616.

39. D. Ferro-Costas; M. N. D. S. Cordeiro; D. G. Truhlar; A. Fernández-Ramos. Q2DTor: A program to treat torsional anharmonicity through coupled pair torsions in flexible molecules. *Comput. Phys. Commun.*, **232**, 190 (2018). DOI: https://doi.org/10.1016/j.cpc.2018.05.025.