F1000Research
F1000Research 2020, 9(Chem Inf Sci):291 Last updated: 26 MAY 2020

SOFTWARE TOOL ARTICLE

GoodVibes: automated thermochemistry for heterogeneous computational chemistry data [version 1; peer review: 2 approved with reservations]

Guilian Luchini [1*], Juan V. Alegre-Requena [1*], Ignacio Funes-Ardoiz [2], Robert S. Paton [1]

[1]Department of Chemistry, Colorado State University, Fort Collins, Colorado, 80523, USA
[2]Institute of Organic Chemistry, RWTH Aachen, Aachen, 52074, Germany

[*] Equal contributors

v1 **First published:** 24 Apr 2020, **9**(Chem Inf Sci):291
https://doi.org/10.12688/f1000research.22758.1

Latest published: 24 Apr 2020, **9**(Chem Inf Sci):291
https://doi.org/10.12688/f1000research.22758.1

Abstract

GoodVibes is an open-source Python toolkit for processing the results of quantum chemical calculations. Thermochemical data are not simply parsed, but evaluated by evaluation of translational, rotational, vibrational and electronic partition functions. Changes in concentration, pressure, and temperature can be applied, and deficiencies in the rigid rotor harmonic oscillator treatment can be corrected. Vibrational scaling factors can also be applied by automatic detection of the level of theory and basis set. Absolute and relative thermochemical values are output to text and graphical plots in seconds. GoodVibes provides a transparent and reproducible way to process raw computational data into publication-quality tables and figures without the use of spreadsheets.

Keywords

GoodVibes, thermochemistry, quasi-harmonic, quasi-RRHO, automated thermodynamic calculations

Open Peer Review

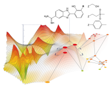
Reviewer Status ? ?

	Invited Reviewers **1**	 **2**
version 1 24 Apr 2020	? report	? report

1 **Robert Pollice**, University of Toronto, Toronto, Canada

2 **Kjell Jorner**, AstraZeneca, Macclesfield, UK

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the Chemical Information Science gateway.

python

This article is included in the Python collection.

Corresponding author: Robert S. Paton (robert.paton@colostate.edu)

Author roles: Luchini G: Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Alegre-Requena JV**: Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Review & Editing; **Funes-Ardoiz I**: Investigation, Methodology, Software, Writing – Review & Editing; **Paton RS**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This material is based upon work supported by the National Science Foundation under the Center for Computer Aided Synthesis (C-CAS), grant number CHE-1925607.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Luchini G, Alegre-Requena JV, Funes-Ardoiz I and Paton RS. **GoodVibes: automated thermochemistry for heterogeneous computational chemistry data [version 1; peer review: 2 approved with reservations]** F1000Research 2020, **9**(Chem Inf Sci):291 https://doi.org/10.12688/f1000research.22758.1

First published: 24 Apr 2020, **9**(Chem Inf Sci):291 https://doi.org/10.12688/f1000research.22758.1

Introduction

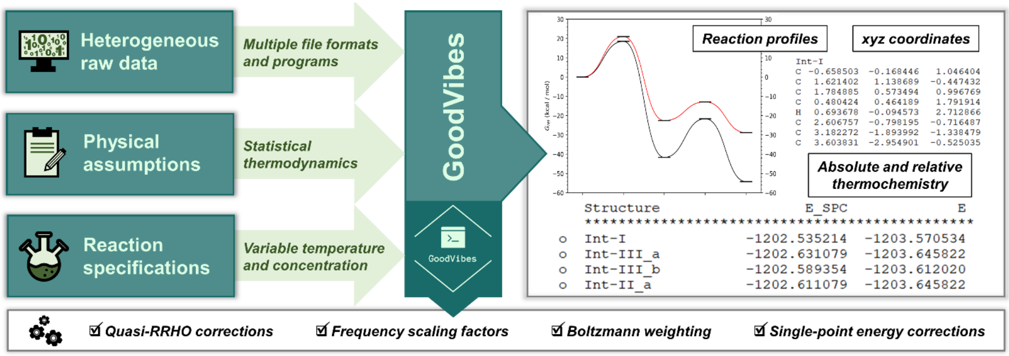
Quantum chemistry software packages implement various levels of theory and basis sets, so-called *model chemistries*, that can be used to optimize molecular geometries and compute ground state vibrational modes. Statistical mechanical expressions using these vibrational frequencies, along with other contributing terms to the partition function, are used to obtain the enthalpy, entropy and Gibbs energy values required to understand, validate or predict experimental observations. GoodVibes was developed to address several challenges faced by practitioners of computational thermochemistry.

Firstly, the rigid-rotor harmonic oscillator (RRHO) model is routinely used to obtain vibrational entropic contributions and is the default for most electronic structure packages. However, the harmonic approximation fails to accurately describe low frequency modes and alternative models may be more appropriate[1–3]. Corrections to the RRHO model may be theoretically desirable, but the absence of practical and accessible tools to implement such corrections has limited their widespread adoption. Here we present and detail the use of the program GoodVibes[4], a Python-based project used for obtaining thermochemical values while applying corrections using quasi-harmonic approximations, alongside other corrections relevant to the overestimation of the zero-point energy, multi-conformer ensembles, and standard concentrations.

Secondly, computational chemistry projects often combine results from different software. For example, geometries may be optimized with one program and the energies evaluated with another. Complex spreadsheets are frequently used to process these results and to prepare figures and manuscripts. However, errors in spreadsheets are commonplace and may be hard to detect. For this reason, GoodVibes allows the combination of data from multiple program outputs as part of the same project. Our group has utilized this program in informatics and organic mechanistic studies[5], however, GoodVibes is not limited to a specific area of chemistry and has been used in published studies by more than 30 different research groups. As an example, previous studies in organic catalysis and mechanism, photocatalysis, and inorganic structure characterization have made use of this program package to process and make corrections to thermochemical data[6]. Figure 1 details an overview of the GoodVibes workflow from input data to the various outputs. Input options supplied via the command line enable chemists with basic experience with Python to process a large number of computational output files and to generate publication-ready data. These data (e.g. figures, tables) can be quickly reproduced by any other user with access to the raw data and the GoodVibes code. With recent and coming changes to standards of supplying full computational outputs through open repositories (such as Zenodo and ioChem-BD[7]) alongside publications, GoodVibes allows for complete transparency in how reported thermochemical values were obtained.

Methods

By recomputing translational, rotational, vibrational and electronic partition functions from the data generated by quantum chemistry programs (such as vibrational frequencies, molecular mass, etc.), thermochemistry can be calculated in GoodVibes at any specified temperature or concentration/pressure. By default, these are set to 298.15 K and 1 atmosphere. A notable automated correction to the RRHO model is applied to low frequency vibrational modes. These low frequency modes (typically less than 100 cm^{-1}) are not well approximated as harmonic and their entropy contributions tend to be overestimated. Methods to avoid this include nontrivial hindered-rotor calculations and computationally expensive anharmonic calculations, both of which become more infeasible



```
Int-I
C  -0.658503  -0.168446   1.046404
C   1.621402   1.138689  -0.447432
C   1.784885   0.573494   0.996769
C   0.480424   0.464189   1.791914
H   0.693678  -0.094573   2.712866
C   2.606757  -0.798195  -0.716487
C   3.182272  -1.893992  -1.338479
C   3.603831  -2.954901  -0.525035
```

```
   Structure                         E_SPC             E
   ***********************************************
o  Int-I                        -1202.535214  -1203.570534
o  Int-III_a                    -1202.631079  -1203.645822
o  Int-III_b                    -1202.589354  -1203.612020
o  Int-II_a                     -1202.611079  -1203.645822
```

Figure 1. The GoodVibes workflow. Computational chemistry output data files are parsed, and thermochemistry data is generated based on a series of user-defined physical assumptions and reaction conditions. The program outputs absolute and relative values to text and reaction profiles as plots. The execution time is seconds.

with higher atom counts[1]. Both Cramer/Truhlar[2] and Grimme[3] have proposed simple, more widely adopted corrections, so-called quasi-harmonic approximations, formulated specifically to obtain vibrational entropies for these low frequency modes. GoodVibes will also automatically apply empirical scaling factors to computed frequencies. These corrections arise from the tendency of electronic structure calculations (e.g. with density functional theory) to overestimate vibrational frequencies relative to experiment, and hence zero-point energies. Linear scaling factors have been collated for a number of functional and basis set combinations for sets of small organic molecules. GoodVibes accesses the scaling factors compiled by the Truhlar group across several studies[8], automatically detects the level of theory and basis set and scales the frequencies if there is a match.

Single point energy calculations performed at more expensive levels of theory and with larger basis sets are commonly used in combination with the thermal corrections obtained from separate calculations, often using different software packages[9]. A multitude of output files from different program packages along with correction applied by GoodVibes can be combined to construct a potential energy surface by using an easily interpretable YAML file that defines the elementary steps of the reaction, file definitions and formatting options.

Implementation

GoodVibes is a module implemented in Python, currently supported by 2.6-7 and all 3.x versions. The module requires the NumPy library (version 1.14.2 or greater), with optional importing of Matplotlib (version 2.2.4 or greater) to graph potential energy surfaces. To test the accuracy and upkeep of this coding package, a series of tests have been implemented in TravisCI, checking functionality and accuracy of the code when the master branch is updated on GitHub against target Python versions on Linux, macOS and Windows operating systems.

Operation

GoodVibes is compatible with Windows 10, Linux and macOS operating systems. This software is appropriate for use with output files produced by a wide range of calculation types, including density functional theory, wave function theory, molecular mechanics, COSMO-RS solvation calculations, and semi-empirical methods. Current supported programs include Gaussian 09[10], Gaussian 16[11], ORCA 4[12] single point energy calculation files, and COSMOtherm[13] COSMO-RS solvation free energy output files. GoodVibes provides an output file with tabulated thermochemical data, optionally exported as a CSV file. Cartesian coordinates of processed files can also optionally be exported in an XYZ file. Plots of energy profiles are optionally generated.

Use case

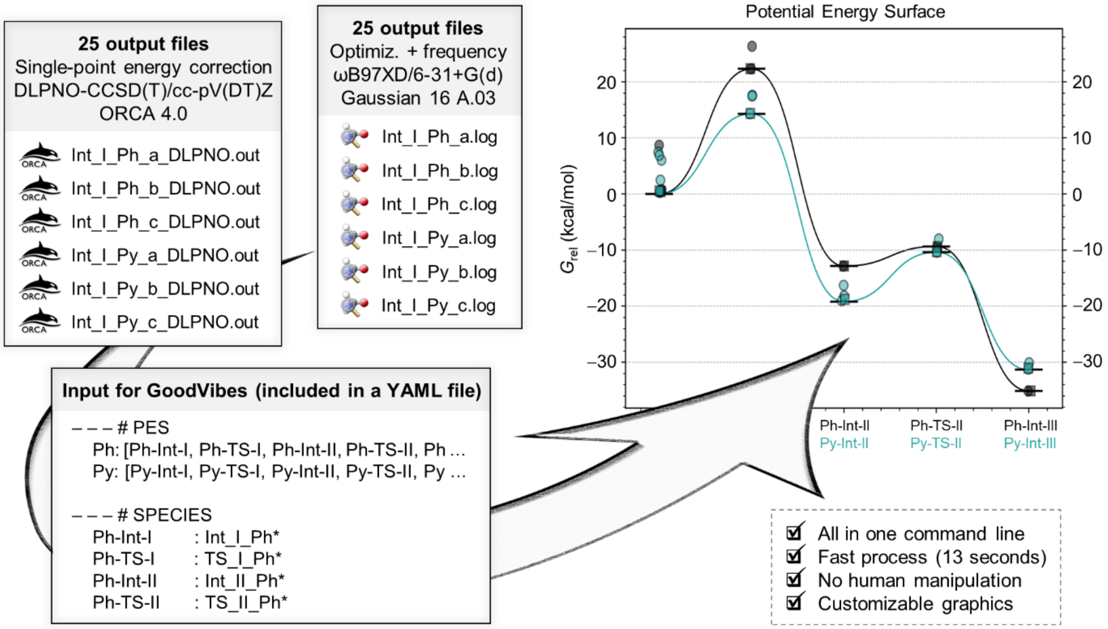
In this section we show how GoodVibes can be used with a variety of input options and files to transform heterogeneous computational chemistry data into human-readable tabulated values and figures. In this example, 50 calculation output files (25 Gaussian geometry optimizations and vibrational frequency calculations with 25 corresponding ORCA single point calculations) are used to create the data in Table 1 and graphed in Figure 2. All raw data was taken from a 2018 study[14], and is freely accessible through a Zenodo repository[15]. Each point in Figure 2 represents a unique conformer's Gibbs energy. The Boltzmann-weighted values are shown as dashes connected by the curved profiles. Optimizations were done with ωB97X-D/6-31+G(d)[16] implemented in Gaussian using an "ultrafine" pruned (99,590) integration grid. Considering solvent effects of the reaction, calculations were run with the SMD solvation model[17] using ethanol, and the concentration of each substance was set to 1.0 M for further thermochemical calculation. GoodVibes allows for solvent media corrections to entropy based on select solvent standard state concentration[18], which was applied to ethanol in this case.

Thermochemistry is evaluated at a temperature of 80°C (353.15 K) in accordance with experimental conditions (the temperature assumed in the original calculations does not influence the results from GoodVibes). These values are corrected using the quasi-harmonic approximation proposed by Grimme. One conformer of intermediate II in the 'Py' pathway has a small imaginary frequency of 2.9 cm^{-1}, which was "inverted" to a real value 2.9 cm^{-1}, as done in previous works to small imaginary frequencies, typically under >i50 cm^{-1}[14,19].

Separate single point energies are extracted from ORCA calculations performed at a coupled cluster level of theory with a (DZ/TZ) basis set extrapolation (DLPNO-CCSD(T), cc-pVDZ/cc-pVTZ)[20], then used for calculations and added to the GoodVibes output for comparison. A potential energy surface is constructed by using Boltzmann weighted averaging of all conformers at each step in the pathway. A multi-structural correction is then applied to the resulting Gibbs free energy based on the number and energy of distinguishable conformers present for each species[21]. The Gibbs energy profile is constructed from options specified in the YAML file containing the reaction pathway steps, file definitions and plot formatting options. All of the output files, correction and formatting

Table 1. Tabulated relative Boltzmann weighted thermochemical values (shown in kcal·mol^{-1}) from the "Ph" and "Py" pathways. Including: single point calculations (ΔE_SPC), energy (ΔE), zero-point energy (ΔZPE), Gibbs free energy (ΔG) and quasi-harmonic corrected Gibbs free energy (qh-ΔG).

Ph pathway	**ΔE_SPC**	**ΔE**	**ΔZPE**	**ΔG**	**qh-ΔG**
Ph-Int-I	0.00	0.00	0.00	0.00	0.00
Ph-TS-I	23.56	23.54	-1.55	21.29	22.31
Ph-Int-II	-12.29	-14.14	-0.57	-13.49	-12.86
Ph-TS-II	-8.13	-12.17	-1.34	-9.97	-9.35
Ph-Int-III	-33.59	-37.61	-0.57	-36.61	-35.11
Py pathway	**ΔE_SPC**	**ΔE**	**ΔZPE**	**ΔG**	**qh-ΔG**
Py-Int-I	0.00	0.00	0.00	0.00	0.00
Py-TS-I	15.08	16.52	-1.20	14.27	14.29
Py-Int-II	-17.45	-18.19	-0.55	-20.97	-19.21
Py-TS-II	-9.59	-13.17	-0.97	-10.96	-10.39
Py-Int-III	-29.39	-33.24	-0.70	-33.04	-31.35



Input for GoodVibes (included in a YAML file)

```
– – – # PES
  Ph: [Ph-Int-I, Ph-TS-I, Ph-Int-II, Ph-TS-II, Ph ...
  Py: [Py-Int-I, Py-TS-I, Py-Int-II, Py-TS-II, Py ...

– – – # SPECIES
  Ph-Int-I     : Int_I_Ph*
  Ph-TS-I      : TS_I_Ph*
  Ph-Int-II    : Int_II_Ph*
  Ph-TS-II     : TS_II_Ph*
```

- ☑ All in one command line
- ☑ Fast process (13 seconds)
- ☑ No human manipulation
- ☑ Customizable graphics

Figure 2. A reaction Gibbs energy profile is produced directly from the command above and saved to PNG file. Here, two reaction pathways, "Ph" and "Py", are displayed.

options are supplied to GoodVibes to output tabulated data and a graph of the reaction pathway from a single command:

```
python -m goodvibes *.log --spc DLPNO --pes science.yaml --graph
science.yaml -t 353.15 --imag --invertifreq -5 --media ethanol -c 1
```

Additional usage examples are described at GoodVibes GitHub repository, where several features have been added in response to requests from the community of users.

Conclusion

GoodVibes is a Python-based tool that calculates thermochemical data from quantum mechanical calculations in a transparent and reproducible way. GoodVibes may be employed with any type of chemical structure, including organic and inorganic molecules of varying sizes as well as with single point calculations performed by differing programs. Additionally, GoodVibes contains many additional automated features that are designed to save time for researchers, allowing for the calculation of thermochemical data at any temperature or concentration, incorporating valuable and overlooked corrections to the RRHO model through quasi-harmonic and vibrational scaling factor corrections and construction of potential energy surfaces with applied corrections accounting for the accessibility of multiple conformations. For projects involving the analysis of a large number of computational chemistry output files, GoodVibes helps to prevent human errors associated with spreadsheets, and can be used to reproduce any table or figure from the raw data.

Data availability

Zenodo: Data Supporting GoodVibes: Automating and applying thermodynamic corrections to harmonic frequency calculations, https://doi.org/10.5281/zenodo.3662845[15].

This project contains data referenced in the use case.

Software availability

This code has been made accessible for chemists of various levels of computational experience and is easily installed. The most recent version of our open-sourced Python package GoodVibes v3.0.1 is freely available on GitHub at https://github.com/bobbypaton/GoodVibes. GoodVibes may be installed as a Python module from the command line using either PyPI (https://pypi.org/project/goodvibes/) or Conda (https://anaconda.org/patonlab/goodvibes) using the commands:

```
pip install goodvibes
```

or:

```
conda install -c patonlab goodvibes
```

or, by downloading the repository from GitHub and running the following command from the extracted directory:

```
python setup.py install
```

Archived source code at the time of publication: https://doi.org/10.5281/zenodo.595246[22]

License: MIT

Acknowledgements

Yanfei Guan (MIT), Jaime Rodríguez-Guerra (Charité Universitätsmedizin Berlin Institut für Physiologie).

References

1. [(a)] Ayala PY, Schlegel HB: **Identification and Treatment of Internal Rotation in Normal Mode Vibrational Analysis.** *J Chem Phys.* 1998; **108**(6): 2314–2325.
   Publisher Full Text
   [(b)] McClurg RB, Flagan RC, Goddard WA: **The hindered rotor density-of-states interpolation function.** *J Chem Phys.* 1997; **106**(16): 6675–6680.
   Publisher Full Text
2. Ribeiro RF, Marenich AV, Cramer CJ, *et al.*: **Use of solution-phase vibrational frequencies in continuum models for the free energy of solvation.** *J Phys Chem B.* 2011; **115**(49): 14556–14562.
   PubMed Abstract | Publisher Full Text
3. Grimme S: **Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory.** *Chem - A Eur J.* 2012; **18**(32): 9955–9964.
   Publisher Full Text
4. Luchini G, Alegre-Requena JV, Funes-Ardoiz I, *et al.*: **GoodVibes, Version 3.0.1**. 2020.
   Reference Source
5. [(a)] Liu B, Alegre-Requena JV, Paton RS, *et al.*: **Unconventional Reactivity of Ethynylbenziodoxolone Reagents and Thiols: Scope and Mechanism.** *Chem Eur J.* 2020; **26**(11): 2386–2394.
   Publisher Full Text
   [(b)] Koniarczyk JL, Greenwood JW, Alegre-Requena JV, *et al.*: **A Pyridine–Pyridine Cross-Coupling Reaction via Dearomatized Radical Intermediates.** *Angew Chemie.* 2019; **131**(42): 15024–15028.
   Publisher Full Text
   [(c)] Luchini G, Ascough DMH, Alegre-Requena JV, *et al.*: **Data-mining the diaryl(thio)urea conformational landscape: Understanding the contrasting behavior of ureas and thioureas with quantum chemistry.** *Tetrahedron.* 2019; **75**(6): 697–702.
   Publisher Full Text
6. [(a)] Lewis RD, Garcia-Borràs M, Chalkley MJ, *et al.*: **Catalytic iron-carbene intermediate revealed in a cytochrome c carbene transferase.** *Proc Natl Acad Sci.* 2018; **115**(28): 7308–7313.
   Publisher Full Text
   [(b)] Svatunek D, Houszka N, Hamlin TA, *et al.*: **Chemoselectivity of Tertiary Azides in Strain-Promoted Alkyne-Azide Cycloadditions.** *Chem Eur J.* 2019; **25**(3): 754–758.
   Publisher Full Text
   [(c)] Gomes GDP, Loginova Y, Vatsadze SZ, *et al.*: **Isonitriles as Stereoelectronic Chameleons: The Donor-Acceptor Dichotomy in Radical Additions.** *J Am Chem Soc.* 2018; **140**(43): 14272–14288.
   PubMed Abstract | Publisher Full Text
   [(d)] Wodrich MD, Busch M, Corminboeuf C: **Expedited Screening of Active and Regioselective Catalysts for the Hydroformylation Reaction.** *Helv Chim Acta.* 2018; **101**(9): e1800107.
   Publisher Full Text
   [(e)] Ye J, Kalvet I, Schoenebeck F, *et al.*: **Direct α-alkylation of primary aliphatic amines enabled by CO_2 and electrostatics.** *Nature Chem.* 2018; **10**(10): 1037–1041.
   Publisher Full Text
   [(f)] Grayson MN: **Mechanism and Origins of Stereoselectivity in the Cinchona Thiourea- and Squaramide-Catalyzed Asymmetric Michael Addition of Nitroalkanes to Enones.** *J Org Chem.* 2017; **82**(8): 4396–4401.
   PubMed Abstract | Publisher Full Text
   [(g)] Besora M, Vidossich P, Lledós A, *et al.*: **Calculation of Reaction Free Energies in Solution: A Comparison of Current Approaches.** *J Phys Chem A.* 2018; **122**(5): 1392–1399.
   PubMed Abstract | Publisher Full Text
7. Álvarez-Moreno M, de Graaf C, López N, *et al.*: **Managing the computational chemistry big data problem: the ioChem-BD platform.** *J Chem Inf Model.* 2015; **55**(1): 95–103.
   PubMed Abstract | Publisher Full Text
8. Alecu IM, Zheng J, Zhao Y, *et al.*: **Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries.** *J Chem Theory Comput.* 2010; **6**(9): 2872–2887.
   PubMed Abstract | Publisher Full Text
9. Feller D, Peterson KA, Hill JG: **On the effectiveness of CCSD(T) complete basis set extrapolations for atomization energies.** *J Chem Phys.* 2011; **135**(4): 044102.
   PubMed Abstract | Publisher Full Text
10. **Gaussian 09**. Gaussian Inc., Wallingford CT, 2016.
    Reference Source
11. **Gaussian 16**. Gaussian Inc., Wallingford CT, 2016.
    Reference Source
12. Neese F: **Software update: the ORCA program system, version 4.0.** *Wiley Interdiscip Rev Comput Mol Sci.* 2017; **8**(1): e1372.
    Publisher Full Text
13. Klamt A: **Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena.** *J Phys Chem.* 1995; **99**: 2224–2235.
    Publisher Full Text
14. Hilton MC, Zhang X, Boyle BT, *et al.*: **Heterobiaryl Synthesis by Contractive C – C Coupling via P(V) Intermediates.** *Science.* 2018; **362**(6416): 799–804.
    Publisher Full Text
15. Luchini G, Alegre-Requena JV, Paton RS: **Data Supporting GoodVibes: Automating and applying thermodynamic corrections to harmonic frequency calculations (Version 1.0) .** [Data set]. *Zenodo.* 2020.
    http://www.doi.org/10.5281/zenodo.3674160
16. [(a)] Becke AD: **Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals.** *J Chem Phys.* 1997; **107**: 8554–8560.
    Publisher Full Text
    [(b)] Chai JD, Head-Gordon M: **Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections.** *Phys Chem Chem Phys.* 2008; **10**(44): 6615–6620.
    PubMed Abstract | Publisher Full Text
    [(c)] Hehre WJ, Ditchfield R, Pople JA: **Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules.** *J Chem Phys.* 1972; **56**: 2257–2261.
    Publisher Full Text
    [(d)] Hariharan PC, Pople JA: **The influence of polarization functions on molecular orbital hydrogenation energies.** *Theoret chim Acta.* 1973; **28**: 213–222.
    Publisher Full Text
    [(e)] Krishnan R, Binkley JS, Seeger R, *et al.*: **Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions.** *J Chem Phys.* 1980; **72**: 650–654.
    Publisher Full Text
    [(f)] McLean AD, Chandler GS: **Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18.** *J Chem Phys.* 1980; **72**: 5639–5648.
    Publisher Full Text
    [(g)] Francl MM, Pietro WJ, Hehre WJ, *et al.*: **Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements.** *J Chem Phys.* 1982; **77**: 3654–3665.
    Publisher Full Text
17. Marenich AV, Cramer CJ, Truhlar DG: **Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions.** *J Phys Chem B.* 2009; **113**(18): 6378–6396.
    PubMed Abstract | Publisher Full Text
18. Harvey JN, Himo F, Maseras F, *et al.*: **Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis.** *ACS Catal.* 2019; **9**(8): 6803–6813.
    Publisher Full Text
19. [(a)] Sure R, Grimme S: **Comprehensive Benchmark of Association (Free) Energies of Realistic Host-Guest Complexes.** *J Chem Theory Comput.* 2015; **11**(8): 3785–3801.
    PubMed Abstract | Publisher Full Text
    [(b)] Liu Z, Patel C, Harvey JN, *et al.*: **Mechanism and reactivity in the Morita-Baylis-Hillman reaction: the challenge of accurate computations.** *Phys Chem Chem Phys.* 2017; **19**(45): 30647–30657.
    PubMed Abstract | Publisher Full Text
20. [(a)] Purvis GD III, Bartlett RJ: **A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples.** *J Chem Phys.* 1982; **76**(4): 1910–1918.
    Publisher Full Text
    [(b)] Pople JA, Head-Gordon M, Raghavachari K: **Quadratic Configuration Interaction. A General Technique for Determining Electron Correlation Energies.** *J Chem Phys.* 1987; **87**: 5968–5975.
    Publisher Full Text
    [(c)] Riplinger C, Neese F: **An efficient and near linear scaling pair natural orbital based local coupled cluster method.** *J Chem Phys.* 2013; **138**(3): 034106.
    PubMed Abstract | Publisher Full Text
    [(d)] Riplinger C, Sandhoefer B, Hansen A, *et al.*: **Natural triple excitations in local coupled cluster calculations with pair natural**

orbitals. *J Chem Phys.* 2013; **139**(13): 134101.
PubMed Abstract | Publisher Full Text
[(e)] Riplinger C, Pinski P, Becker U, *et al.*: **Sparse maps-A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory.** *J Chem Phys.* 2016; **144**(2): 024109.
PubMed Abstract | Publisher Full Text

21. Plata RE, Singleton DA: **A Case Study of the Mechanism of Alcohol-Mediated Morita Baylis-Hillman Reactions. the Importance of Experimental Observations.** *J Am Chem Soc.* 2015; **137**(11): 3811–3826.
Publisher Full Text

22. Luchin G, Alegre-Requena J, Rodríguez-Guerra J, *et al.*: **bobbypaton/GoodVibes: GoodVibes v3.0.1 (Version v3.0.1).** *Zenodo.* 2019.
http://www.doi.org/10.5281/zenodo.3346166

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 26 May 2020

https://doi.org/10.5256/f1000research.25128.r63185

? **Kjell Jorner**
Early Chemical Development, Pharmaceutical Sciences, R&D, AstraZeneca, Macclesfield, UK

Paton and co-workers describe an open source Python program, GoodVibes, for the calculation of thermochemistry from quantum-chemical calculations. The program automates a lot of the tedious work needed to process the output files and to add correction terms. Importantly, it presents a reproducible workflow for extracting and visualizing energies of reaction paths. GoodVibes is now in its third version and has over the last few years become a standard tool for computational chemists both in academia and industry. Additional functionality in the latest version includes, for example, automated calculation of diastereo- and enantioselectivity, potential energy surface plotting, and including separate single point energies.

This software tool article is, therefore, most welcome not only as a reference for the program but by providing more detail and references to the underlying theory as well as examples of how to use the program. In particular, the program implements many elements of good scientific software practice, such as using a Git repository with version control, a version release system, and continuous integration testing. The program is also easy to install via pip or conda.

It would be good if the authors could be more explicit regarding the “multi-structural correction” that is “applied based on the number and energy of distinguishable conformers present for each species". Is this on by default, and what is the basis for distinguishing between unique conformers? Also, could you be more explicit regarding the symmetry detection and correction (“ssym” argument)? How does it detect the internal symmetries of the molecule? And why is it off by default, when it seems that most standard quantum chemistry codes (*e.g.* Gaussian) include at least external symmetry by default? Also, the use of “solvent media corrections” in conjunction with an implicit solvent method is strongly discouraged in ref 18 (Harvey, Himo, Maseras), and it would be good to add a note on this caveat with the inexperienced user in mind.

I think the program is very useful as is, but here are some suggested improvements:
- Make the program work as an importable Python module and not only as a standalone script. This is especially helpful when integrating it into automated (Python-based) workflows.

- Storing internal data in separate files (*e.g.* csv) rather than in the Python source code may be more maintainable.
- Separate documentation on, *e.g.*, Read the Docs may be more accessible.
- Document Python code with docstrings in NumPy or Google style.
- Output of SVG or other vector graphics files for lossless figures.
- A quick way to support more quantum chemistry codes (at least on Linux or Mac) would be to support output from the UniMoVib program (https://github.com/zorkzou/UniMoVib).

Minor correction:

- “Gibbs energy” -> “Gibbs free energy” (page 3/9 and other examples)
- “a Python-based project” -> “a Python-based program/tool” (page 3/9)
- “standard concentrations” -> “standard state concentrations” (page 3/9)

Addition: After reading the review by Dr. Pollice, I agree with him that the handling of the moment of inertia for Grimme's method is not correct. There is a reference implementation of his method which can be found in the xtb Github repository: https://github.com/grimme-lab/xtb. The "subroutine axis2" handles the calculation of the average moment of inertia. After correcting for this in the code, I think the authors should do an analysis of how large the error is for using the erroneous moment of inertia, e.g., as a function of molecular size as Dr. Pollice points out.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: physical organic chemistry, reaction modelling, machine learning

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 18 May 2020

https://doi.org/10.5256/f1000research.25128.r63186

? **Robert Pollice**
University of Toronto, Toronto, ON, Canada

The paper “GoodVibes: automated thermochemistry for heterogeneous computational chemistry data” by Paton *et al.* describes GoodVibes, an open-source program to compute thermochemical parameters from the output of popular computational chemistry software accounting for multiple well-established thermodynamic corrections developed and used for chemistry in the gas phase and in solutions. The aim is to provide a general tool to apply these corrections straightforwardly without much effort saving time for practitioners and making the corresponding results more reproducible for the entire scientific community.

Overall, the paper is clear, well-written, and explains the rationale for developing the program. All the theoretical models implemented are described and cited appropriately. I believe that this program is a very important contribution to the field as it simplifies the process of applying thermochemical corrections greatly, which otherwise can be quite time-consuming and is, as the authors point out appropriately, error-prone due to the considerable number of calculation steps that need to be carried out. I have tested the current version of the program and both installation and usage are well-documented and straightforward for anyone with minimal python experience.

However, having inspected the source code of the current version, I am concerned about the implementation, or lack thereof, of a few important thermochemical corrections as I will outline in the section “Major Issues” in more detail. Moreover, I believe that there are a few minor issues that require correction and I will outline them in the section “Minor Issues.”

Major issues:

1. My main concern with the current version of GoodVibes is that I believe that Grimme’s quasi-harmonic correction is not implemented correctly. In line 1922 of GoodVibes.py, the authors define the average moment of inertia as $10 \cdot 10^{-44}$ kg m^2 and use this value as a global parameter for every molecule. However, I believe that this is a misinterpretation of the method by Grimme as described in reference 3 of the paper. In that paper, Grimme uses this specific value in Figure 2 merely as an illustration of a typical example to show how the mathematical shape of this correction looks. It is never stated that this value is to be used as a global parameter for all molecules. In fact, the average moments of inertia need to be calculated for every molecule at hand from the respective Cartesian coordinates of the conformer used and the masses of all the atoms. In my opinion, this issue needs to be addressed properly by implementing the calculation of the average moment of inertia of a conformer based on its Cartesian coordinates in GoodVibes.

This is particularly important for large molecules with hundreds of atoms because their average moment of inertia will deviate significantly from the used global parameter.

2. After close inspection of the code, I believe that, currently, GoodVibes applies the solvent correction as obtained from the output of COSMOtherm as is without any adjustments. However, it has been pointed out recently that this is (unfortunately) a common mistake as COSMOtherm uses the mole fraction reference state for pure solvent rather than the concentration reference state. Consequently, this leads to inconsistencies in the standard states used and to a systematic offset in predicted Gibbs free energies of solvation. The details of this problem and how to correct for it are described in the literature.[1]

3. Moreover, I inspected the code of GoodVibes in order to find out whether a correction for the entropy of mixing of multiple species is applied. I could not find such a correction. I believe that this is an important correction, especially for reactions with changing molecularity. This correction is outlined in reference 21 of the paper, in particular in the SI of reference 21.

Minor issues:

1. Page 4, first paragraph: The authors mention that electronic structure calculations have a tendency to overestimate vibrational frequencies relative to experiment. In my opinion, this is misleading as the tendency to overestimate vibrational frequencies is not inherent to the electronic structure methods applied but rather inherent to the harmonic approximation of the vibrational potential energy surface. Hence, I suggest that the authors mention that specifically, possibly with an appropriate citation. For that purpose, citation 8 of the paper could be used as it addresses this issue properly.

2. Page 4, “Operation” paragraph: In my opinion, appropriate citations should be added for the COSMO-RS method.

3. Page 4, “Use case” first paragraph: The authors mention that in the output plots, “dashes [are] connected by the curved profiles.” In my opinion, the authors should mention explicitly what interpolation method is used to generate these curved profiles.

4. Page 4, “Use case” second paragraph: The authors state that “in previous works” small imaginary frequencies “typically under >i50 cm^{-1}” were inverted. From a mathematical point of view, I believe this to be imprecise notation. Natural linear ordering on the sets of neither complex nor imaginary numbers is possible. Hence, the notation “>i50 cm^{-1}”, even though it has been used in Chemistry, should be avoided. To resolve this issue, it should be stated that the absolute of the imaginary frequency should be below 50 cm^{-1}. In addition, the imaginary units should be added to the “small imaginary frequency of 2.9 cm^{-1}” in the same sentence.

5. This is just a suggestion rather than an issue: It would be great to have the option to save the plots generated by GoodVibes in a number of formats including, in particular, formats supporting vector graphics like PDF as this would make it easier to use the graphics directly for generating high-resolution illustrations.

References

1. Assaf K, Florea M, Antony J, Henriksen N, et al.: HYDROPHOBE Challenge: A Joint Experimental and

Computational Study on the Host–Guest Binding of Hydrocarbons to Cucurbiturils, Allowing Explicit Evaluation of Guest Hydration Free-Energy Contributions. *The Journal of Physical Chemistry B*. 2017; **121** (49): 11144-11162 Publisher Full Text

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: homogeneous catalysis, molecular interactions, noncovalent interactions, reaction mechanisms, chemical kinetics, organic electronics, machine learning

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.