AIP | The Journal of Chemical Physics

Large-scale calculations of gas phase thermochemistry: Enthalpy of formation, standard entropy, and heat capacity

Mohammad M. Ghahremanpour, Paul J. van Maaren, Jonas C. Ditz, Roland Lindh, and David van der Spoel

Citation: The Journal of Chemical Physics **145**, 114305 (2016); doi: 10.1063/1.4962627
View online: http://dx.doi.org/10.1063/1.4962627
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/145/11?ver=pdfcov
Published by the AIP Publishing

Articles you may be interested in

Improved methods for Feynman path integral calculations and their application to calculate converged vibrational–rotational partition functions, free energies, enthalpies, entropies, and heat capacities for methane
J. Chem. Phys. **142**, 044105 (2015); 10.1063/1.4905526

Heat capacity, enthalpy fluctuations, and configurational entropy in broken ergodic systems
J. Chem. Phys. **133**, 164503 (2010); 10.1063/1.3499326

Structures and heats of formation of the neutral and ionic PNO, NOP, and NPO systems from electronic structure calculations
J. Chem. Phys. **128**, 164305 (2008); 10.1063/1.2902983

Thermochemistry of Oxabicyclo-Heptanes, Oxabicyclo-Heptene: Enthalpy of Formation, Entropy, Heat Capacity, and Group Additivity
J. Phys. Chem. Ref. Data **36**, 663 (2007); 10.1063/1.2734558

Electron attachment to SF 5 CF 3 (296–563 K) and calculations of the neutral and anion thermochemistry
J. Chem. Phys. **116**, 6021 (2002); 10.1063/1.1457442

NEW Special Topic Sections

NOW ONLINE
Lithium Niobate Properties and Applications: Reviews of Emerging Trends

AIP Applied Physics Reviews

Reuse of AIP Publishing content is subject to the terms: https://publishing.aip.org/authors/rights-and-permissions. Downloaded to IP: 130.238.171.188 On: Fri, 14 Oct 2016 12:30:47

THE JOURNAL OF CHEMICAL PHYSICS **145**, 114305 (2016)

Large-scale calculations of gas phase thermochemistry: Enthalpy of formation, standard entropy, and heat capacity

Mohammad M. Ghahremanpour,[1] Paul J. van Maaren,[1] Jonas C. Ditz,[1] Roland Lindh,[2] and David van der Spoel[1,a)]

[1]*Uppsala Centre for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, SE-75124 Uppsala, Sweden*
[2]*Uppsala Centre for Computational Chemistry, Department of Chemistry at Ångström, Uppsala University, Box 538, SE-75121 Uppsala, Sweden*

(Received 7 June 2016; accepted 31 August 2016; published online 19 September 2016)

Large scale quantum calculations for molar enthalpy of formation ($\Delta_f H^0$), standard entropy (S^0), and heat capacity (C_V) are presented. A large data set may help to evaluate quantum thermochemistry tools in order to uncover possible hidden shortcomings and also to find experimental data that might need to be reinvestigated, indeed we list and annotate approximately 200 problematic thermochemistry measurements. Quantum methods systematically underestimate S^0 for flexible molecules in the gas phase if only a single (minimum energy) conformation is taken into account. This problem can be tackled in principle by performing thermochemistry calculations for all stable conformations [Zheng *et al.*, Phys. Chem. Chem. Phys. **13**, 10885–10907 (2011)], but this is not practical for large molecules. We observe that the deviation of composite quantum thermochemistry recipes from experimental S^0 corresponds roughly to the Boltzmann equation ($S = \mathrm{R} \ln \Omega$), where R is the gas constant and Ω the number of possible conformations. This allows an empirical correction of the calculated entropy for molecules with multiple conformations. With the correction we find an RMSD from experiment of ≈13 J/mol K for 1273 compounds. This paper also provides predictions of $\Delta_f H^0$, S^0, and C_V for well over 700 compounds for which no experimental data could be found in the literature. Finally, in order to facilitate the analysis of thermodynamics properties by others we have implemented a new tool `obthermo` in the OpenBabel program suite [O'Boyle *et al.*, J. Cheminf. **3**, 33 (2011)] including a table of reference atomization energy values for popular thermochemistry methods. [http://dx.doi.org/10.1063/1.4962627]

I. INTRODUCTION

Prediction of thermochemistry is crucial for designing chemicals with new functionality since fundamental properties such as Gibbs free energy, enthalpy, heat capacity, and standard entropy are needed to understand stability and reaction energies of compounds.[1–5] Therefore, a large amount of effort has gone into the development of quantum chemical methods to predict thermochemistry, especially enthalpy of formation, based on a theoretical description of molecular electronic structure and nuclear motion.[2] Methods such as Gaussian-n,[6–10] Weizman-n,[11–15] and Petersson-style complete basis set (CBS) models[16,17] have improved accuracy in *ab initio* thermochemistry by combining calculations at different levels of theory and basis sets with empirical corrections in most methods. The empirical corrections limit their predictive capability to the datasets against which they are benchmarked.[18] Moreover, calculations of absolute thermodynamics such as standard entropy and heat capacity are reported much less often than the enthalpy of formation despite their importance.

The rigid rotator-harmonic oscillator approximation to describe the motion of the nuclei in molecules is likely the weakest part of quantum methods for calculating entropy and heat capacity.[2,19] In this model, the vibrations of nuclei in a molecule are treated as independent harmonic oscillators. Under this assumption, the high frequency and low amplitude vibrations in which the nuclei remain close to the equilibrium position are described relatively accurately. Problems arise when there are low barrier torsion potentials, large amplitude motions, or anharmonic vibrations, all of which are difficult to describe harmonically and as a result their contribution to the thermodynamics functions is difficult to evaluate.[19–21] The errors associated with anharmonicity become significant at temperatures where the anharmonic modes become excited when the molecule leaves the harmonic potential surface.[20,21]

Several improvements have been suggested to alleviate these shortcomings. For instance, Katzer *et al.* treated nuclear motions by taking partial asymmetrical internal rotations into account for a number of small carbon and silicon compounds.[20] They assumed that anharmonicity only needs to be considered for some selected degrees of freedom and described the molecular vibration of silicon hydrides by a set of independent harmonic and anharmonic modes.[20,21]

[a)] Electronic mail: david.vanderspoel@icm.uu.se

They found that the anharmonic correction mainly affects the entropy and isochorous heat capacity thermodynamics functions, while the anharmonicity related contributions to the enthalpy of formation only amount to a few percent of the total vibrational contribution. Other methods have used experimentally obtained anharmonic constants,[22] second-order rotovibrational perturbation theory,[23] or quadratic correction terms[24] to take the effect of anharmonicity into account for prediction of total atomization energy and enthalpy of formation. These methods have improved thermochemistry predictions but they were applied to special cases only and the methods are not practical for complex molecules. More recently, Zheng *et al.* have developed a method called multistructural approximation (MS-AS) which allows for taking a Boltzmann average on conformational space and for considering the change in rotational partition function from structure to structure; hence, rotation is coupled to conformational change.[25–29] MS-AS has demonstrated the importance of the multistructural anharmonicity in determining absolute thermodynamic quantities by reproducing the experimental standard entropy for some small molecules such as ethanol and 1-butanol[25] with an uncertainty of about 4 J/mol K. However, it is not practical for large flexible molecules since the number of accessible conformations increases exponentially with the number of rotatable bonds.

To evaluate the predictive power for absolute quantities such as standard entropy and to highlight possible problems which might not show up in case studies, we have performed quantum calculation on over 2000 molecules up to 47 atoms using a single optimized geometry for each molecule. In the remainder of this paper, we first explain the underlying theories and computational methods and describe the experimental data provided by different resources. Second, we compare the chemical accuracy of the methods including a discussion of the problems involved in predicting each thermochemistry quantity. The methods were also compared for about 30 different chemical categories which are frequently available in biomolecules and drug-like compounds. This may help to identify chemical categories that need more attention for future studies.

II. THEORY

Here, we briefly describe the thermodynamics principles used to approximate the contribution of conformational entropy to the molecular standard entropy and how to derive experimental heat capacity at constant volume from the equation of state of an imperfect gas.

A. Conformational entropy

In order to estimate the conformational entropy for a molecule with Ω conformers, the Boltzmann equation may be used

$$S_{conf} = \mathrm{R}\ln(\Omega), \tag{1}$$

where R is the gas constant. Ω is related to the number of rotatable bonds (α) by

$$\Omega \approx 3^{\alpha} \tag{2}$$

assuming that each rotatable bond corresponds to exactly 3 conformations. Eq. (2) overestimates Ω because rotatable bonds are usually hindered by a potential barrier[30] and not all possible conformations are thermally accessible at room temperature and hence do not contribute to the partition function.[31] Therefore, we approximate $\ln(\Omega)$ by α to alleviate the overestimation. The empirical approximation for the conformational entropy then becomes

$$S_{conf} \approx \mathrm{R}\alpha. \tag{3}$$

With this, the entropy calculated by thermochemistry methods can be corrected empirically using Eq. (3) as follows:

$$S^{0}_{corrected} \approx S^{0}_{QM} + S_{conf}. \tag{4}$$

In this work the number of rotatable bonds was determined by the OpenBabel program `obconformer`[32] and the numbers were compared to the PubChem database[33] in order to check for consistency. Differences were curated manually. The number of rotatable bonds is available for all compounds from http://virtualchemistry.org.

B. Heat capacity at constant volume

We have derived experimental heat capacity at constant volume (C_V) from the isentropic expansion factor (γ)—the ratio of the heat capacity at constant pressure (C_P) to C_V—as well as from the imperfect gas equation of state for the molecules where the temperature dependence of the second virial coefficient was available. Starting from the virial expansion for an imperfect gas truncated after the second term, we have[30]

$$PV = RT + PB(T), \tag{5}$$

where P is the pressure, V is the volume, R is the gas constant, T is the temperature, and B(T) is the second virial coefficient at constant volume. Using Eq. (5), we can derive the difference between the heat capacity at constant pressure and at constant volume

$$C_P - C_V = T\left(\frac{\partial P}{\partial T}\right)_V \left(\frac{\partial V}{\partial T}\right)_P. \tag{6}$$

With this, the experimental C_V in the gas phase can be approximated using[30]

$$C_V = C_P - R - 2P\left(\frac{dB(T)}{dT}\right) - \frac{P^2}{R}\left(\frac{dB(T)}{dT}\right)^2. \tag{7}$$

The second virial coefficient is usually measured as a function of temperature and analytical parameterizations of B(T) are available,[34] therefore it is straightforward to compute the temperature derivative of B. A full list of over 1800 heat capacities C_V determined by either or both methods is given in Table S1 of the supplementary material.

The error propagation of a function of two variables $f(a,b)$ is approximated by[35]

$$\sigma_f^2 \approx \left(\frac{\partial f}{\partial a}\right)_b^2 \sigma_a^2 + \left(\frac{\partial f}{\partial b}\right)_a^2 \sigma_b^2 + 2\left(\frac{\partial f}{\partial a}\right)\left(\frac{\partial f}{\partial b}\right)\sigma, \tag{8}$$

where the σ_x^2 and σ_x are the variance and standard deviation in variable x, respectively, and σ_{ab} is the covariance between the two variables. Therefore, we can estimate the experimental uncertainty in $C_V(C_P, B'(T))$, where $B'(T)$ abbreviates $dB(T)/dT$ from

$$\sigma^2_{C_V} \approx \left(\frac{\partial C_V}{\partial C_p}\right)^2 \sigma^2_{C_P} + \left(\frac{\partial C_V}{\partial B'(T)}\right)^2 \sigma^2_{B'(T)} \qquad (9)$$

by realizing that the covariance term in Eq. (8) approaches zero in our case as the random errors in the measurements of C_P and $B'(T)$ are independent. Finally, Eq. (8) can be written as

$$\sigma^2_{C_V} \approx \sigma^2_{C_P} + \left(2P + \frac{2P^2}{R} B'(T)\right)^2 \sigma^2_{B'(T)}. \qquad (10)$$

III. METHODS

The standard G2, G3, G4,[6–10] and CBS-QB3[16,17] methods were used for about 2000 molecules up to 47 atoms, and W1U and W1BD[13] were used for about 650 molecules up to 16 atoms. Calculations at the same levels of theory were performed for isolated atoms, which are used as reference for extracting the enthalpy of formation. For these calculations the lowest energy state of the atoms had to be used. This means that for all atoms with an even number of electrons a spin multiplicity of 1 (singlet state) was used, except for carbon, oxygen, silicon and sulfur, germanium and selenium (triplet state). For all atoms with an odd number of electrons the doublet state (spin multiplicity 2) was used, except for nitrogen, phosphor, and arsenic for which the quadruplet state has the lowest energy. Our calculations reproduce the published G4 values exactly.[10] All quantum calculations presented here were performed using the Gaussian 09 software package.[36] Note that all calculations predict molar quantities, but the word molar has been left out in the text for brevity (except in the units).

The results of all calculations are tabulated in the supplementary material to this paper and in a database which is freely accessible on the Virtual Chemistry website[37–39] (http://virtualchemistry.org).

The machinery to calculate enthalpy of formation, standard entropy, and heat capacity at constant volume is implemented in a new OpenBabel[32] tool called `obthermo`. The OpenBabel program also includes a data file with the computed and experimental atomization energies described above. The `obthermo` program includes a flag to modify the symmetry number of the molecule (which affects the standard entropy calculation). This is needed since the Gaussian package does not always extract the correct symmetry from the molecular structure. For this reason we were forced to tabulate the symmetries manually for all molecules. Both the symmetry number and the number of rotatable bonds used in this study are available from the Virtual Chemistry website.[37–39] Furthermore, the `obthermo` program can be employed to compute heat capacity at constant pressure from the calculated heat capacity at constant volume ($C_{V,QM}$) and the temperature derivative of the second virial coefficient (dB/dT, see Eq. (7)), which must then be specified by the user.

IV. EXPERIMENTAL DATA

The number of data points in the analyses below is bound by the availability of experimental data. A number of databases were used to provide enthalpy of formation,[34,40,41] standard entropy,[34,40,41] heat capacity at constant pressure,[34,40–42] isentropic expansion factor,[43] and second virial coefficients.[34] Most of the data for enthalpy of formation and entropy is old and the original sources are not readily accessible. This makes it difficult to find uncertainties in experimental data. For compounds where more than one value was found, we have used the average and standard deviation of the values to be the reference value and error, respectively.

We found over 200 suspected problems with the data points which are excluded from the statistics presented in this paper (see Table S2 for details). Moreover, about 30% of the standard entropies listed in the databases are reported to be estimates. For these numbers we have assumed an error of 2%, which is shown as an error bar in Fig. 2. The uncertainty in experimental heat capacity at constant pressure in gas phase and the second virial coefficient are not generally available either. Hence, we assume an error of 2% in γ, C_P, and $B(T)$ quantities to estimate the uncertainty in heat capacity at constant volume. This gives an approximate error of 2.3% and 2% for C_V calculated from Eq. (10) and from the isentropic expansion factor, respectively. Approximately 250 out of the 270 compounds for which the enthalpy of formation is part of the G3/05 test set[44,45] were included.

V. RESULTS AND DISCUSSION

In Section V A, we evaluate all methods used by comparing the root-mean square deviation (RMSD) from experiment for each property computed based on the compounds to which all the methods were applied in order to make a fair comparison. The results obtained for enthalpy of formation (Section V B) are used as the positive control since all the methods are originally optimized to reproduce molecular energetics. Finally, in Sections V C and V D, we present the results obtained for standard entropy and heat capacity, respectively, and discuss possible solutions to improve thermochemistry predictions of these quantities.

A. Comparison of methods

Table I lists the RMSD from experiment for all properties and all six methods on those compounds up 16 atoms to which all methods were applied (Table I). For enthalpy of formation G3 and G4 have slightly lower RMSD than W1BD and W1U, which in turn are somewhat more accurate than G2 and CBS-QB3, but note that the error bar is of the same order of magnitude as the difference between methods. Nevertheless the order of accuracies is in agreement with previous studies.[18] There are small differences for compound classes, e.g., G3 shows a better performance for aromatic- and alcohol compounds as well as for radicals (Table I).

TABLE I. Root mean square deviation (RMSD) from experiment for six quantum chemical methods for the subset of compounds up to 16 atoms where calculations were done using all methods. Error bars in RMSD obtained by bootstrapping with 100 iterations. N is the number of compounds.

	N	G2	G3	G4	CBS-QB3	W1BD	W1U
$\Delta_f H^0$ (kJ/mol)	399	16.8(0.1)	14.3(0.2)	14.0(0.1)	16.7(0.1)	15.8(0.2)	16.0(0.2)
Nonhydrogens	98	23.0(0.3)	20.7(0.5)	19.1(0.5)	23.6(0.5)	21.9(0.8)	21.6(0.5)
Inorganic	109	19.5(0.7)	18.8(0.9)	16.7(1.1)	17.0(0.4)	16.7(0.4)	17.4(0.5)
Halogenated compound	75	18.5(0.1)	15.1(0.3)	15.2(0.3)	21.6(0.2)	19.3(0.4)	20.4(0.2)
Aromatic	12	7.6(0.2)	4.4(0.2)	5.9(0.2)	4.9(0.2)	10.5(0.1)	11.2(0.2)
Alcohol	16	9.8(0.3)	6.9(0.2)	7.2(0.2)	8.8(0.3)	8.2(0.2)	8.8(0.1)
Radical	22	4.9(0.1)	3.9(0.1)	4.5(0.1)	5.0(0.2)	4.2(0.1)	4.5(0.2)
S^0 (J/mol K)	374	13.7(0.1)	13.7(0.2)	11.4(0.1)	11.4(0.1)	11.2(0.1)	11.4(0.1)
Nonhydrogens	97	6.9(0.2)	7.0(0.1)	7.1(0.1)	7.3(0.1)	6.4(0.1)	6.6(0.1)
Inorganic	112	6.4(0.1)	6.2(0.2)	6.4(0.2)	6.5(0.1)	5.5(0.1)	5.4(0.2)
Halogenated compound	73	13.2(0.3)	13.5(0.2)	10.8(0.3)	11.1(0.4)	11.3(0.4)	11.1(0.2)
Aromatic	10	16.2(0.9)	14.9(1.1)	14.1(0.7)	12.7(0.5)	14.0(1.5)	14.2(0.6)
Alcohol	13	25.4(0.3)	25.8(0.8)	19.8(0.4)	20.8(0.5)	19.7(0.5)	19.8(0.3)
Radical	14	4.3(0.1)	4.1(0.1)	4.2(0.1)	4.1(0.1)	4.2(0.1)	4.2(0.1)
C_V (J/mol K)	372	9.4(0.1)	9.3(0.1)	6.0(0.1)	6.0(0.1)	6.1(0.1)	6.1(0.2)
Nonhydrogens	101	3.9(0.1)	3.8(0.0)	1.9(0.1)	2.2(0.1)	1.8(0.0)	1.9(0.1)
Inorganic	124	3.5(0.1)	3.5(0.0)	2.1(0.0)	2.4(0.1)	2.1(0.1)	2.2(0.0)
Halogenated compound	82	9.9(0.3)	10.1(0.6)	7.2(0.4)	7.4(0.4)	6.3(0.6)	6.8(0.6)
Aromatic	13	12.2(0.5)	12.5(0.4)	6.7(0.2)	7.3(0.3)	7.0(0.5)	7.4(0.4)
Alcohol	13	10.4(0.5)	9.9(0.8)	6.7(0.5)	6.5(0.4)	6.8(0.2)	6.6(0.7)
Radical	2	2.0(0.0)	2.1(0.1)	1.7(0.2)	1.8(0.2)	1.9(0.1)	1.8(0.1)

The somewhat better performance of W1BD on open shell systems in comparison to W1U is in agreement with other studies, which recommended W1BD for these compounds.[13] The largest RMSD for $\Delta_f H^0$ is found for nonhydrogens and inorganic compounds (Table I). In addition, the G4, CBS-QB3, W1BD, and W1U perform somewhat better than G2 and G3 methods for predicting S^0 and C_V for about 370 molecules, respectively (Table I).

Table II compares the G*n* family- and the CBS-QB3 methods on compounds from 16 to 47 atoms. Due to the more flexible compounds, predictions for S^0 and C_V are less accurate than for the small compounds in Table I. The RMSD values for $\Delta_f H^0$ are lower, however, which can be explained by the absence of nonhydrogen and inorganic compounds in the set of larger molecules (Table II).

Computational cost also plays a key role in evaluating computational methods in addition to the accuracy and reliability. It has been recently shown, in a study of timings of $\Delta_f H^0$ calculations, that the G4 method is 8 and 24 times slower than the G3 and CBS-QB3 methods respectively; however, it is 28 times faster than the W1BD method.[18] Considering both chemical accuracy and computational cost, the G4 method is a good compromise for thermochemistry

TABLE II. Root mean square deviation (RMSD) from experiment for the used Gn/CBS quantum chemical methods for the subset of compounds with more than 16 atoms where calculations were done using all methods. Error bars in RMSD obtained by bootstrapping with 100 iterations. N is the number of compounds.

	N	G2	G3	G4	CBS-QB3
$\Delta_f H^0$ (kJ/mol)	600	14.2(0.1)	11.7(0.1)	11.2(0.1)	14.4(0.1)
Halogenated compound	32	20.0(0.1)	16.2(0.3)	11.1(0.1)	18.6(0.3)
Aromatic	85	18.8(0.1)	8.5(0.2)	9.1(0.2)	10.3(0.1)
Alcohol	46	15.7(0.5)	13.9(0.4)	14.7(0.4)	13.6(0.6)
S^0 (J/mol K)	543	37.0(0.3)	37.1(0.2)	29.3(0.3)	28.8(0.1)
Halogenated compound	30	49.5(0.9)	49.2(1.5)	40.4(0.9)	39.6(0.9)
Aromatic	78	24.6(0.2)	23.5(0.4)	18.6(0.4)	19.3(0.2)
Alcohol	46	47.1(0.4)	47.2(0.5)	38.5(0.5)	38.2(0.4)
C_V (J/mol K)	612	18.9(0.1)	18.9(0.0)	9.7(0.0)	9.8(0.1)
Halogenated compound	32	26.1(0.2)	26.1(0.3)	14.9(0.2)	14.7(0.3)
Aromatic	88	16.9(0.2)	16.9(0.2)	10.1(0.2)	10.5(0.1)
Alcohol	58	17.8(0.2)	18.1(0.3)	9.1(0.2)	9.5(0.1)

calculations. Therefore, and for clarity of presentation, in the remainder of this paper we only show G4 results but the results for all the methods are tabulated in the supplementary material.

B. Enthalpy of formation

The residual plot showing the deviation of G4 enthalpies of formation from experiment (Fig. 1) is homogeneous, meaning there is no systematic error. We obtained equivalent variance homogenicity for the other methods (not shown). However, all the methods benchmarked are found to yield a systematic error for compounds with a high inner polarization effect such as perchloric- and phosphoric acids and for partly ionic compounds such as zinc sulfide (Table S10). Such cases have been recognized to be problematic for standard thermochemistry tools and have been improved by modifications of the methods by other authors;[46,47]

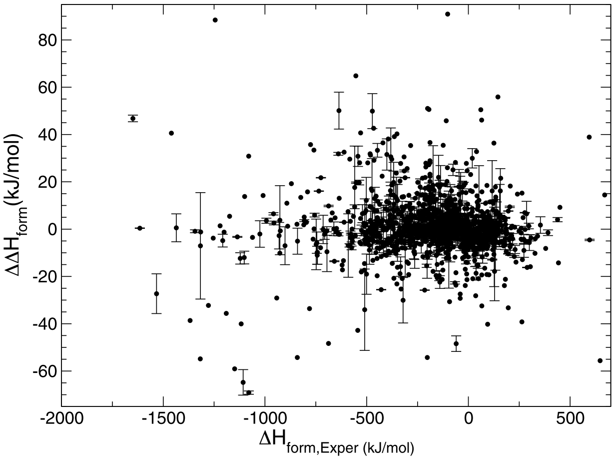

FIG. 1. Residual plot for enthalpy of formation in the gas phase using the G4 quantum chemistry method. Error bars represent the uncertainty in the experimental data.

TABLE III. Statistics of performance of the G4 method for the prediction of $\Delta_f H^0$ per compound category. G3/05 refers to compounds from the G3/05 test set.[45] Number of compounds N, Root Mean Square Deviation (RMSD, kJ/mol), Mean Signed Error (MSE, kJ/mol), slope a of a linear regression analysis (y = ax), and coefficient of determination R^2 (%). Error bars in RMSD and MSE determined by bootstrapping with 100 iterations.

Category	N	RMSD	MSE	a	R^2
Alcohol	98	13 (1)	4 (1)	0.989 (0.001)	99.7 (0.1)
Aldehyde	15	10 (1)	2 (1)	0.996 (0.003)	99.1 (0.2)
Alkane	138	15 (1)	3 (1)	0.988 (0.003)	99.8 (0.1)
Alkene	318	9 (1)	1 (1)	0.989 (0.002)	99.8 (0.1)
Alkylbromide	34	11 (1)	−0 (1)	1.014 (0.002)	99.7 (0.1)
Alkylchloride	56	14 (2)	3 (1)	0.996 (0.002)	99.8 (0.1)
Alkylfluoride	47	15 (1)	−2 (1)	1.003 (0.001)	99.8 (0.1)
Alkyne	50	7 (1)	1 (1)	1.005 (0.002)	99.8 (0.1)
Amide	14	11 (1)	5 (1)	0.976 (0.003)	98.8 (0.2)
Amine	58	13 (1)	1 (1)	0.993 (0.005)	99.8 (0.1)
Amino acid	9	18 (1)	4 (2)	0.989 (0.003)	99.1 (0.2)
Aromatic	164	8 (1)	−1 (1)	0.988 (0.002)	99.9 (0.1)
Arylbromide	5	13 (1)	−11 (1)	0.915 (0.004)	98.0 (0.2)
Arylchloride	19	10 (1)	−2 (1)	0.894 (0.003)	99.3 (0.2)
Arylfluoride	15	4 (1)	2 (1)	0.996 (0.001)	100.0 (0.1)
Carboxylic acid	25	13 (1)	4 (1)	0.993 (0.001)	99.7 (0.1)
Carboxylic ester	43	14 (1)	2 (1)	0.997 (0.002)	99.4 (0.1)
Cycloalkane	100	13 (1)	5 (1)	0.962 (0.002)	99.2 (0.1)
Cycloalkene	27	18 (3)	−1 (1)	0.999 (0.009)	98.1 (0.5)
Fluoroalkene	9	11 (1)	−1 (1)	1.008 (0.002)	99.7 (0.1)
G3/05	218	6 (1)	0 (1)	0.999 (0.001)	100.0 (0.1)
Halogenated compound	193	14 (1)	−1 (1)	1.004 (0.002)	99.9 (0.1)
Heterocyclic	75	13 (1)	−1 (1)	0.984 (0.003)	99.8 (0.1)
Inorganic	152	16 (1)	−0 (1)	1.002 (0.002)	99.9 (0.1)
Ketone	33	8 (1)	2 (1)	0.995 (0.002)	99.8 (0.1)
Nitro	15	9 (1)	−5 (1)	1.024 (0.004)	99.6 (0.1)
Nonhydrogens	157	20 (1)	−1 (1)	1.005 (0.002)	99.9 (0.1)
Phenol	16	13 (1)	5 (1)	0.964 (0.003)	99.3 (0.2)
Primary alcohol	49	12 (1)	3 (1)	0.993 (0.002)	99.8 (0.1)
Primary amine	28	5 (1)	1 (1)	0.980 (0.001)	100.0 (0.1)
Radical	21	5 (1)	−1 (1)	1.000 (0.002)	100.0 (0.1)
Secondary alcohol	30	16 (1)	7 (1)	0.984 (0.002)	99.3 (0.1)
Secondary amine	18	17 (1)	1 (1)	1.000 (0.007)	99.1 (0.2)
Thioether	10	6 (1)	−3 (1)	1.006 (0.003)	100.0 (0.1)
Thiol	16	3 (1)	−2 (1)	1.018 (0.002)	100.0 (0.1)

here, such compounds were excluded from the statistics and listed in Table S2 alongside suspected experimental errors.

Table III compares the performance of G4 theory for different functional groups, indicating that the accuracy and reliability of the calculated enthalpies of formation differ somewhat for different chemical categories. The RMSD is found to be equal to or bigger than 15 kJ/mol for seven categories (Table III), alkyl chlorides, amino acids, cycloalkenes, inorganic compounds, nonhydrogens, secondary alcohols, and secondary amines.

Among the halogenated aryls, the RMSD increases from 4 kJ/mol for arylfluoride to 10 kJ/mol for arylchloride and 13 kJ/mol for arylbromides. Similarly, the RMSD increases from 14 kJ/mol for alkylfluoride to 15 kJ/mol for alkylchloride while for alkylbromide the number is somewhat smaller again, 11 kJ/mol. These results are in agreement with other studies showing that energetics for molecules containing heavy or electron-withdrawing elements are not predicted accurately.[48,49] For instance, it has been shown that CCSD(T) with extrapolation to the complete basis-set limit, combined with the core valence correlation and relativistic effects, is needed to improve the accuracy of thermochemistry for chlorine containing molecules.[49] Moreover, it has been observed that halogen-containing molecules are severely affected by a nondynamical electron correlation; thus, a post-CCSD(T) correlation treatment may be needed.[48] For large molecules such as fatty acids and esters, a parametric empirical correction equation—based on the number of bonding, core and unpaired electrons in the ground state—has been derived in order to correct the quantum enthalpy of formation.[50] Finally for 218 compounds of the G3/05 test set[45] we find an RMSD of 6 kJ/mol considerably lower than the overall RMSD of 13 kJ/mol (Table S9) but somewhat higher than the RMSD reported by Curtiss *et al.* of 4.6 kJ/mol for the 270 enthalpies of formation in the test set using the G4 method[10] (Table S11 lists all compounds from the test set studied here).

C. Standard entropy

Fig. 2(a) representing the deviation of the calculated absolute entropies from experiment shows that G4 theory provides accurate predictions of the measured entropies for small rigid molecules. The results obtained for small nonrigid molecules whose conformational flexibility is caused by Berry pseudorotations[51–53] such as pentavalent species (PF_5 and PCl_5) are in agreement with experiment (Table S4). The calculated entropies are also in agreement with experiment for small ring compounds such as cyclopentane and methylcyclopentane (Table S4). This suggests that the pseudorotation inherent in five-membered rings[54,55] is described relatively well by the harmonic approximation. However, we find a systematic underestimation in entropy calculation for large flexible molecules (Fig. 2(a)). This stems from the poor description of large-amplitude modes in the quantum harmonic partition functions, modes that contribute significantly to the dynamics of flexible molecules having multiple low-energy conformers.[2,19–21,24,56–60] The G4 theory employs the B3LYP functional to optimize geometry and to calculate zero-point vibrational energy (ZPVE).[10] Moreover, a global factor of 0.9854 is used to mitigate the overestimation of frequencies in the calculation of ZPVE.[10] However, our results show that there is still a systematic error in the predicted entropies. Dannenfelser and Yalkowsky showed that a correction term representing the molecular flexibility, in addition to molecular rotational symmetry (σ), is needed to predict the melting entropy for flexible molecules.[61,62] Similarly, Zheng *et al.* improved the predicted value of standard entropy for some molecules by defining the $q_{con-rovib}$ partition function which combines the contribution of all stable conformers (flexibility) of the molecule with the contribution of rotations (including symmetry) and vibrations.[25,26] However, the ro-vibrational partition function is laborious to solve for complex molecules due to the potentially huge number of degrees of freedom.

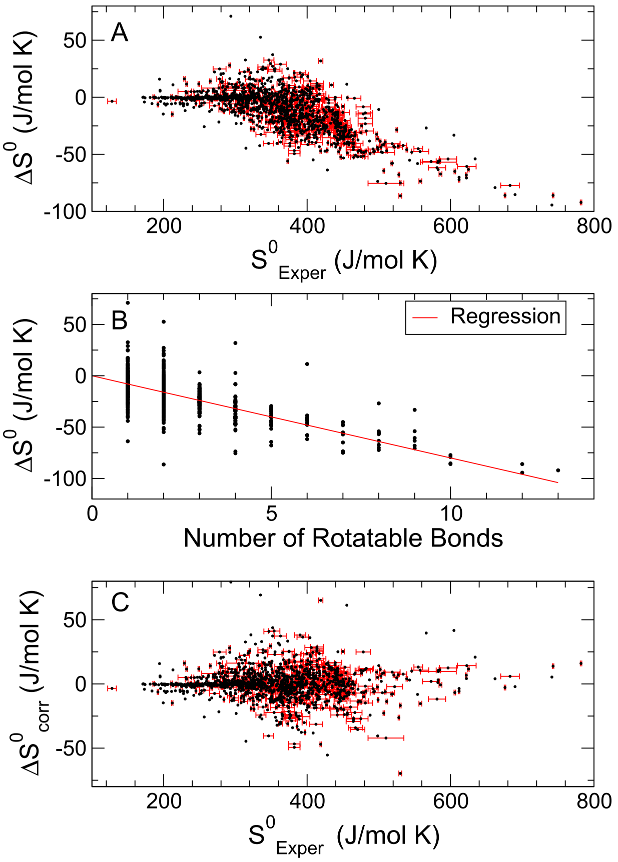

FIG. 2. G4 standard entropy. (a) This shows the deviation of G4 entropies from experiment. Estimated experimental error in red. (b) Regression analysis of the deviation of G4 entropies from experiment versus the number of rotatable bonds. The slope of the regression line (shown in red) is given in Table IV. (c) Residual plot of G4 entropies corrected using Eq. (4). Error bars represent the uncertainty in the experimental data.

Consistent with the role of molecular flexibility in the prediction of melting entropy,[62] we observe that the deviation of quantum standard entropy from experiment (Fig. 2(b)) is roughly proportional to the number of conformers as described

TABLE IV. Number of data points N and slope (J/mol K) of regression analysis of ΔS^0 against number of rotatable bonds for each method. A bootstrapping procedure with 100 iterations was used to obtain the uncertainty in the slope for each method.

Method	N	Slope
CBS-QB3	704	−7.9(0.1)
G2	716	−10.8(0.1)
G3	719	−10.7(0.1)
G4	704	−8.0(0.1)
W1BD	148	−9.5(0.2)
W1U	112	−10.1(0.3)
Theoretical		8.314 15

TABLE V. Standard entropy S^0 (J/mol K) calculated by G4 and $\mathrm{G4}_{\mathrm{Corrected}}$ in this study and the corresponding values calculated by MS-AS method. α refers to the number of rotatable bonds. All values are reported at 298.15 K.

Compound	α	MS-AS[25]	G4	$\mathrm{G4}_{\mathrm{Corrected}}$	Reference data[41]
Ethanol	1	282.3	270.3	278.3	280.4
1-butanol	3	364.7	334.9	359.1	361.7

by Eq. (1). The regression slopes presented in Table IV show that the deviation of G4 entropy of over ≈700 molecules corresponds to about 8 J/mol K per rotatable bond. The slopes are not the same for all methods used because each method is based on different levels of theory and approximations which means there may be different systematic, as well as random, errors.

Eq. (4) was used to add the conformational entropy to the calculated entropies (Fig. 2(c)), but we assumed that the contribution of conformational entropy per rotatable bond is about the ideal gas constant (R = 8.314 15 J/mol K) in order to use a uniform correction term for all the methods. As a result,

TABLE VI. Statistics of performance of the G4 method for the prediction of S^0 per compound category. G3/05 refers to compounds from the G3/05 test set.[45] Number of compounds N, Root Mean Square Deviation (RMSD, J/mol K), Mean Signed Error (MSE, J/mol K), slope a of a linear regression analysis (y = ax), and coefficient of determination R^2 (%). Error bars in RMSD and MSE determined by bootstrapping with 100 iterations.

Category	N	RMSD	MSE	a	R^2
Alcohol	92	14 (1)	−9 (1)	0.977 (0.001)	98.9 (0.1)
Aldehyde	13	12 (1)	−5 (1)	0.981 (0.004)	96.4 (0.2)
Alkane	133	13 (1)	3 (1)	1.009 (0.001)	99.3 (0.2)
Alkene	294	12 (1)	−2 (1)	0.996 (0.001)	97.8 (0.1)
Alkylbromide	33	6 (1)	1 (1)	1.004 (0.002)	99.9 (0.1)
Alkylchloride	53	11 (1)	2 (1)	1.009 (0.003)	98.2 (0.2)
Alkylfluoride	45	6 (1)	3 (1)	1.008 (0.001)	99.9 (0.1)
Alkyne	50	12 (1)	−0 (1)	0.998 (0.001)	98.9 (0.2)
Amide	10	15 (1)	8 (2)	1.021 (0.005)	93.6 (1.1)
Amine	48	15 (1)	−6 (1)	0.983 (0.003)	98.0 (0.2)
Aromatic	144	14 (1)	−0 (1)	0.998 (0.002)	93.8 (0.2)
Arylbromide	5	7 (1)	−1 (1)	0.996 (0.002)	94.6 (0.6)
Arylchloride	19	12 (1)	4 (1)	1.011 (0.002)	82.2 (0.7)
Arylfluoride	8	11 (1)	5 (1)	1.015 (0.003)	96.6 (0.9)
Carboxylic acid	20	16 (1)	−2 (1)	0.992 (0.002)	98.4 (0.2)
Carboxylic ester	41	13 (1)	−1 (1)	0.996 (0.002)	98.4 (0.2)
Cycloalkane	78	17 (1)	−0 (1)	1.000 (0.002)	93.6 (0.5)
Cycloalkene	19	8 (1)	−2 (1)	0.992 (0.002)	98.5 (0.2)
Fluoroalkene	8	8 (1)	−5 (1)	0.983 (0.002)	94.7 (0.6)
G3/05	203	7 (1)	1 (1)	1.004 (0.001)	99.3 (0.1)
Halogenated compound	180	10 (1)	1 (1)	1.005 (0.002)	99.4 (0.1)
Heterocyclic	59	13 (1)	−2 (1)	0.993 (0.002)	93.0 (0.3)
Inorganic	156	8 (1)	0 (1)	1.002 (0.001)	98.6 (0.2)
Ketone	29	12 (1)	3 (1)	1.008 (0.002)	98.4 (0.2)
Nitro	15	11 (1)	8 (1)	1.022 (0.002)	98.3 (0.2)
Nonhydrogens	153	9 (1)	1 (1)	1.007 (0.002)	99.1 (0.1)
Phenol	16	18 (1)	−11 (2)	0.969 (0.004)	89.9 (1.9)
Primary alcohol	46	10 (1)	−7 (1)	0.984 (0.001)	99.5 (0.2)
Primary amine	26	15 (1)	−10 (1)	0.971 (0.003)	98.5 (0.2)
Radical	14	6 (1)	−1 (1)	0.996 (0.002)	97.4 (0.2)
Secondary alcohol	27	17 (1)	−12 (1)	0.970 (0.002)	96.9 (0.2)
Secondary amine	17	16 (1)	−4 (1)	0.989 (0.002)	96.6 (0.3)
Thioether	8	26 (2)	−8 (1)	0.975 (0.004)	96.3 (0.5)
Thiol	15	9 (1)	−8 (1)	0.983 (0.003)	99.9 (0.1)

the systematic underestimation of S^0 for large molecules is remediated and the total RMSD is reduced significantly from 22 to 13 J/mol K for G4 (compare Tables S3 and S5). The root-mean square deviation and the relative deviation of the corrected S^0 from experiment are reported in Table S5 and all corrected values for all molecules are given in Table S6. In Table V entropies calculated here using G4 and G4$_{\text{Corrected}}$ are compared to the values previously reported by other methods. This shows that the G4$_{\text{Corrected}}$ agrees well with experimental data, indicating that S_{conf} approximates the multi-structural anharmonic effect (torsional anharmonicity) denoted by the conformational-rovibrational partition function in the MS-AS method.[56] By adding S_{conf} to S^0_{QM}, the RMSD for the G*n* theories is reduced by 9-12 J/mol K versus 3 J/mol K for the W*n* theories (Tables S3 and S5), the latter methods were however applied to small and medium-sized molecules only.

Table VI gives the statistics of G4 predictions of S^0 (with correction) per functional group. A number of systematic problems can be detected in this manner. Hydroxyl-containing compounds, in particular secondary alcohols and phenols, as well as thioether and primary amine compounds are underestimated. The large MSE values for these chemicals show that the observed deviations were systematic. A similar systematic underestimation is found for carboxylic acids, which cannot be attributed to specific outliers (Table VI). On the other hand, amides and nitro compounds are systematically overestimated.

D. Heat capacity at constant volume

Fig. 3(a) compares the experimental isentropic expansion factor (γ)[43] to the one derived from the second virial coefficient (Section II A). Although the calculated γ values are in agreement with experiment in most cases, there are quite some outliers. Therefore, the C_V values calculated using Eq. (7) were deemed not to be accurate enough to be used as the reference data to evaluate quantum methods. This may happen due to the uncertainties in the experimental second virial coefficient, the lack of higher order terms in the viral Taylor expansion of an imperfect gas (Eq. (5)), or both.

Since the thermodynamic partition functions used in standard quantum tools are based on an ideal-gas model, it is interesting to consider the deviation from ideal-gas behavior for real gases. It can be estimated from the difference between the heat capacities at constant pressure and at constant volume (Fig. 3(a)). Although, for most compounds the difference is close to the gas constant R, there are quite some, in particular smaller, compounds for which the difference implies deviation from ideal-gas behavior. These are typically small polar molecules such as ammonia; hence, a significant interaction in the gas phase can be expected. For these and other small compounds molecular flexibility might not contribute to thermodynamics functions significantly and predictions are relatively accurate (Fig. 4). For larger molecules containing more internal rotations the deviations are significant. In principle this might be because internal rotations at temperatures for which $kT \ll V_0$, where V_0 is the barrier height to rotation, contribute to the heat capacity.[63] Even for barrierless rotation, a contribution to C_V of $\frac{R}{2}$ would be expected.[30] Regression analysis between the deviation from experimental C_V and the number of rotatable bonds implies a contribution of 2 J/mol K to C_V per rotatable bond which is lower than even the barrierless rotation. Although the enthalpy of formation is reproduced accurately due in part to favorable cancellation of errors, the heat capacity is more difficult to reproduce apparently.

Table VII lists the performance per functional group of the G4 method for predicting the C_V. No large mean-signed errors or root mean-square deviations are found for any compound class.

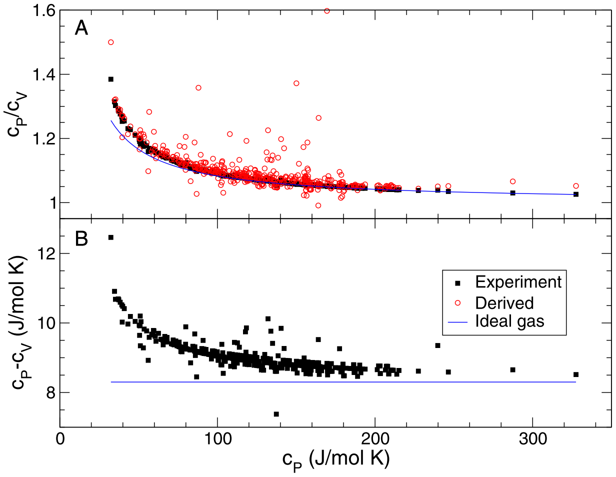

FIG. 3. (a) Isentropic expansion factor derived from the equation of state of an imperfect gas, using the heat capacity at constant pressure and the second virial coefficient, is compared to the experimental data (b) Deviation from ideal gas for each molecule shown by the difference between the heat capacity at constant pressure and at constant volume.

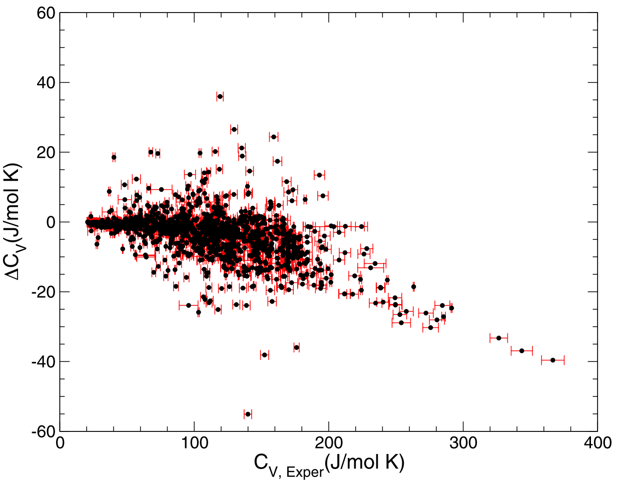

FIG. 4. Residual plot for heat capacity at constant volume in the gas phase using the G4 quantum chemistry method. Error bars represent the uncertainty in the experimental data.

TABLE VII. Statistics of performance of the G4 method for the prediction of C_V per compound category. G3/05 refers to compounds from the G3/05 test set.[45] Number of compounds N, Root Mean Square Deviation (RMSD, J/mol K), Mean Signed Error (MSE, J/mol K), slope a of a linear regression analysis (y = ax), and coefficient of determination R^2 (%). Error bars in RMSD and MSE determined by bootstrapping with 100 iterations.

Category	N	RMSD	MSE	a	R^2
Alcohol	101	8 (1)	−4 (1)	0.966 (0.002)	98.8 (0.2)
Aldehyde	14	6 (1)	−5 (1)	0.940 (0.004)	98.8 (0.2)
Alkane	144	10 (1)	−6 (1)	0.941 (0.002)	99.5 (0.2)
Alkene	348	7 (1)	−5 (1)	0.962 (0.001)	99.0 (0.1)
Alkylbromide	36	9 (1)	−6 (1)	0.929 (0.002)	99.9 (0.1)
Alkylchloride	56	9 (1)	−3 (1)	0.953 (0.004)	96.4 (0.7)
Alkylfluoride	49	10 (1)	−5 (1)	0.929 (0.003)	99.8 (0.1)
Alkyne	52	8 (1)	−5 (1)	0.950 (0.001)	99.3 (0.1)
Amide	10	10 (1)	3 (1)	1.018 (0.005)	93.1 (0.9)
Amine	59	10 (1)	−4 (1)	0.971 (0.003)	96.9 (0.2)
Aromatic	170	9 (1)	−2 (1)	0.982 (0.002)	95.5 (0.2)
Arylbromide	8	9 (1)	−7 (1)	0.937 (0.004)	95.8 (0.2)
Arylchloride	20	5 (1)	−3 (1)	0.974 (0.002)	96.7 (0.5)
Arylfluoride	16	4 (1)	0 (1)	1.001 (0.002)	97.9 (0.3)
Carboxylic acid	23	14 (1)	−6 (1)	0.947 (0.005)	96.4 (0.5)
Carboxylic ester	43	9 (1)	−1 (1)	0.993 (0.003)	97.0 (0.2)
Cycloalkane	97	6 (1)	−4 (1)	0.967 (0.002)	99.1 (0.1)
Cycloalkene	29	5 (1)	−3 (1)	0.968 (0.003)	99.2 (0.2)
Fluoroalkene	9	2 (1)	−0 (1)	0.993 (0.004)	98.2 (0.2)
G3/05	189	4 (1)	−2 (1)	0.969 (0.002)	99.2 (0.1)
Halogenated compound	202	9 (1)	−4 (1)	0.944 (0.002)	99.3 (0.1)
Heterocyclic	71	9 (1)	−4 (1)	0.951 (0.002)	96.6 (0.2)
Inorganic	125	2 (1)	−0 (1)	1.002 (0.001)	99.6 (0.1)
Ketone	34	10 (1)	−1 (1)	0.985 (0.003)	97.6 (0.2)
Nitro	14	7 (1)	−4 (1)	0.956 (0.004)	99.2 (0.2)
Nonhydrogens	122	6 (1)	0 (1)	1.007 (0.002)	98.4 (0.2)
Phenol	16	9 (1)	−4 (1)	0.968 (0.004)	86.8 (1.4)
Primary alcohol	50	9 (1)	−4 (1)	0.958 (0.003)	98.9 (0.2)
Primary amine	33	7 (1)	−5 (1)	0.960 (0.003)	98.7 (0.2)
Secondary alcohol	33	6 (1)	−2 (1)	0.978 (0.002)	98.4 (0.3)
Secondary amine	19	9 (1)	−4 (1)	0.970 (0.004)	97.6 (0.2)
Thioether	6	7 (1)	−5 (1)	0.966 (0.004)	99.5 (0.1)
Thiol	19	11 (1)	−8 (1)	0.931 (0.002)	99.8 (0.1)

VI. CONCLUSION

With the improvements of quantum theories for performing electronic structure calculations, it has become straightforward to accurately determine molecular energetics for small molecules in gas phase.[6–10,13–15,64] However, it has remained challenging to theoretically describe complex and flexible molecules. In this study, we have evaluated the performance of six popular methods on over 2000 molecules up to 47 atoms in predicting thermochemistry, particularly standard entropy and heat capacity which are not addressed often. We provide predictions of energetics for well over 700 compounds where no experimental results are available in the databases: S^0 values in Table S6, C_V values in Table S8, and $\Delta_f H^0$ value in Table S10, all at the G4 level of theory. Moreover, we have listed 215 experimental thermochemistry values for compounds that may need to be reinvestigated (Table S2).

Four out of the six methods benchmarked here use B3LYP geometries and frequencies. The zero-point energy is scaled by an empirical factor to alleviate the overestimation of the vibrational frequencies in most methods. Our results show that there is still systematic underestimation in predicting absolute thermodynamic quantities, in particular standard entropy. Therefore, applying scaling factors on the zero-point energy is insufficient to compensate for large-amplitude motions contributing to the vibrational partition functions. This indicates that the B3LYP functional may not be the best choice to perform geometry optimization on medium and large flexible compounds. Rather, dispersion-corrected functionals should be employed for thermochemistry calculations as recommended in a benchmark study of different density functionals by Goerigk and Grimme.[65] We have shown that the magnitude of anharmonic effects, mainly caused by internal rotations, to the entropy function is roughly proportional to the logarithm of the number of conformers through the Boltzmann equation (Eq. (1)). Expressing the number of conformers in terms of the number of rotatable bonds suggests that the contribution of conformational change as a result of large-amplitude motions is about the ideal gas constant per rotatable bond. The statistics show that taking the conformational entropy (S_{conf}) into account as an empirical correction to the quantum entropy decreases the root mean square deviation from the reference data by about 9-12 J/mol K for G*n* theories applied to molecules up to 47 atoms and about 3 J/mol K for W*n* methods performed on molecules up to 16 atoms. Thus, Eq. (4) allows an improved estimate of the standard entropy by performing quantum calculations on a single minimum-energy structure without increasing the computational complexity beyond the harmonic oscillator model. However, the precision of the S_{conf} highly depends on the accuracy of enumerating the molecular conformers which itself is an active research area.[31,66,67]

Comparison of different chemical categories showed that the RMSD and MSE obtained for the predicted standard entropies of compounds containing electron-withdrawing amine- or hydroxyl (secondary alcohols, phenols, and primary amines) moieties as well as compounds with thioether group are higher than other chemicals (Table VI). Similarly, the deviation from experimental enthalpy of formation obtained for nonhydrogens, amino acids, and halogenated compounds is high in comparison to other compound classes (Table III). These results suggest that it may not be straightforward to predict stability of the designed biomolecules and drug-like compounds, which are generally enriched by amine, hydroxyl, and halogenated moieties, based on their Gibbs free energy of formation since both entropy and enthalpy must be determined accurately in this context. This is problematic since, e.g., the usage of halogens and halogen bonding in rational design of drugs and materials has increased over the years.[68–71]

The sheer amount of thermochemistry data provided in this paper strongly suggests that there is a need to reach a higher level of accuracy and transparency in the existing databases to which the computational methods are benchmarked. On the other hand, further improvements in theory are still needed despite long development of quantum thermochemistry theories, in particular for large flexible molecules and molecules containing electron-withdrawing and heavy atoms. Higher order correlation corrections may be needed to improve accuracy of frequency calculations. Furthermore, relativistic effects, which play a substantial role for compounds with heavier atoms such as bromine and iodine, are ignored. Finally, since the chemistry of living systems occurs predominantly in the condensed phase rather than the gas phase, methods will need to be developed to predict thermochemistry in all phases.

SUPPLEMENTARY MATERIAL

See the supplementary material for additional methods and examples for using the `obthermo` tool. Detailed statistics benchmarking the performance of six quantum theories in the prediction of thermochemical properties as well as tables representing all the predicted and the corresponding experimental data are provided. Tables of suspected experimental errors as well as with new experimental data for heat capacity at constant volume in the gas phase are provided.

ACKNOWLEDGMENTS

The Swedish Research Council is acknowledged for financial support to D.v.d.S. (Grant No. 2013-5947) and R.L. (Grant No. 2012-3910), and for a grant of computer time (Grant No. SNIC2013-26-6) through the High Performance Computing Center North in Umeå, Sweden.

[1] K. K. Irikura and D. J. Frurip, *Computational Thermochemistry*, ACS Symposium Series (ACS, 1998), Vol. 677.

[2] W. M. F. Fabian, "Accurate thermochemistry from quantum chemical calculations?," Monatsh. Chem. **139**, 309–318 (2008).

[3] S. Manzetti, H. Behzadi, O. Andersen, and D. van der Spoel, "Fullerenes toxicity and electronic properties," Environ. Chem. Lett. **11**, 105–118 (2013).

[4] S. Manzetti, E. R. van der Spoel, and D. van der Spoel, "Chemical properties, environmental fate, and degradation of seven classes of pollutants," Chem. Res. Toxicol. **27**, 713–737 (2014).

[5] A. Walsh, "Inorganic materials: The quest for new functionality," Nat. Chem. **7**, 274–275 (2015).

[6] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, "Gaussian-1 theory: A general procedure for prediction of molecular energies," J. Chem. Phys. **90**, 5622–5629 (1989).

[7] L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, "Gaussian-1 theory of molecular energies for second-row compounds," J. Chem. Phys. **93**, 2537–2545 (1990).

[8] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, "Gaussian-2 theory for molecular energies of first- and second-row compounds," J. Chem. Phys. **94**, 7221–7230 (1991).

[9] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, "Gaussian-3 (G3) theory for molecules containing first and second-row atoms," J. Chem. Phys. **109**, 7764–7776 (1998).

[10] L. A. Curtiss, P. C. Redfern, and K. Raghavachari, "Gaussian-4 theory," J. Chem. Phys. **126**, 84108 (2007).

[11] J. M. L. Martin and G. de Oliveira, "Towards standard methods for benchmark quality *ab initio* thermochemistry—W1 and W2 theory," J. Chem. Phys. **111**, 1843–1856 (1999).

[12] S. Parthiban and J. M. L. Martin, "Assessment of W1 and W2 theories for the computation of electron affinities, ionization potentials, heats of formation, and proton affinities," J. Chem. Phys. **114**, 6014–6029 (2001).

[13] E. C. Barnes, G. A. Petersson, J. A. Montgomery, M. J. Frisch, and J. M. L. Martin, "Unrestricted coupled cluster and Brueckner doubles variations of W1 theory," J. Chem. Theory Comput. **5**, 2687–2693 (2009).

[14] A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, "W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions," J. Chem. Phys. **125**, 144108 (2006).

[15] A. Karton, S. Daon, and J. M. Martin, "W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data," Chem. Phys. Lett. **510**, 165–178 (2011).

[16] J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, and G. A. Petersson, "A complete basis set model chemistry. VI. Use of density functional geometries and frequencies," J. Chem. Phys. **110**, 2822–2827 (1999).

[17] J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, and G. A. Petersson, "A complete basis set model chemistry. VII. Use of the minimum population localization method," J. Chem. Phys. **112**, 6532–6542 (2000).

[18] J. M. Simmie and K. P. Somers, "Benchmarking compound methods (CBS-QB3, CBS-APNO, G3, G4, W1BD) against the active thermochemical tables: A litmus test for cost-effective molecular formation enthalpies," J. Phys. Chem. A **119**, 7235–7246 (2015).

[19] B. Njegic and M. S. Gordon, "Exploring the effect of anharmonicity of molecular vibrations on thermodynamic properties," J. Chem. Phys. **125**, 224102 (2006).

[20] G. Katzer and A. F. Sax, "Beyond the harmonic approximation: Impact of anharmonic molecular vibrations on the thermochemistry of silicon hydrides," J. Phys. Chem. A **106**, 7204–7215 (2002).

[21] G. Katzer and A. F. Sax, "Identification and thermodynamic treatment of several types of large-amplitude motions," J. Comput. Chem. **26**, 1438–1451 (2005).

[22]J. M. L. Martin, “Heat of atomization of sulfur trioxide, SO_3: A benchmark for computational thermochemistry,” Chem. Phys. Lett. **310**, 271–276 (1999).

[23]J. M. L. Martin and P. R. Taylor, “Benchmark *ab initio* thermochemistry of the isomers of diimide, N_2H_2, using accurate computed structures and anharmonic force fields,” Mol. Phys. **96**, 681–692 (1999).

[24]S. Parthiban and J. M. L. Martin, “Fully *ab initio* atomization energy of benzene via Weizmann-2 theory,” J. Chem. Phys. **115**, 2051–2054 (2001).

[25]J. Zheng, T. Yu, E. Papajak, I. M. Alecu, S. L. Mielke, and D. G. Truhlar, “Practical methods for including torsional anharmonicity in thermochemical calculations on complex molecules: The internal-coordinate multi-structural approximation,” Phys. Chem. Chem. Phys. **13**, 10885–10907 (2011).

[26]J. Zheng, S. L. Mielke, K. L. Clarkson, and D. G. Truhlar, “MSTor: A program for calculating partition functions, free energies, enthalpies, entropies, and heat capacities of complex molecules including torsional anharmonicity,” Comput. Phys. Commun. **183**, 1803–1812 (2012).

[27]J. Zheng and D. G. Truhlar, “Quantum thermochemistry: Multistructural method with torsional anharmonicity based on a coupled torsional potential,” J. Chem. Theory Comput. **9**, 1356–1367 (2013).

[28]J. Zheng and D. G. Truhlar, “Including torsional anharmonicity in canonical and microcanonical reaction path calculations,” J. Chem. Theory Comput. **9**, 2875–2881 (2013).

[29]J. Zheng, R. Meana-Pañeda, and D. G. Truhlar, “Prediction of experimentally unavailable product branching ratios for biofuel combustion: The role of anharmonicity in the reaction of isobutanol with OH,” J. Am. Chem. Soc. **136**, 5150–5160 (2014).

[30]D. A. McQuarrie and J. D. Simon, *Molecular Thermodynamics* (University Science books, Sausalito, CA, 1999).

[31]N. O’Boyle, T. Vandermeersch, C. Flynn, A. Maguire, and G. Hutchison, “Confab—Systematic generation of diverse low-energy conformers,” J. Cheminf. **3**, 8 (2011).

[32]N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, “Open Babel: An open chemical toolbox,” J. Cheminf. **3**, 33 (2011).

[33]S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, and S. H. Bryant, “PubChem substance and compound databases,” Nucleic Acids Res. **44**, D1202–1213 (2016).

[34]R. L. Rowley, W. V. Wilding, J. L. Oscarson, Y. Yang, and N. F. Giles, *Data Compilation of Pure Chemical Properties* (Design Institute for Physical Properties, American Institute for Chemical Engineering, New York, 2012).

[35]H. H. Ku, “Notes on the use of propagation of error formulas,” J. Res. Natl. Bur. Stand., Sect. C **70C**, 263–273 (1966).

[36]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, GAUSSIAN 09, Revision D.01, Gaussian, Inc., Wallingford, CT, 2009.

[37]C. Caleman, P. J. van Maaren, M. Hong, J. S. Hub, L. T. Costa, and D. van der Spoel, “Force field benchmark of organic liquids: Density, enthalpy of vaporization, heat capacities, surface tension, compressibility, expansion coefficient and dielectric constant,” J. Chem. Theory Comput. **8**, 61–74 (2012).

[38]D. van der Spoel, P. J. van Maaren, and C. Caleman, “GROMACS molecule & liquid database,” Bioinformatics **28**, 752–753 (2012).

[39]N. M. Fischer, P. J. van Maaren, J. C. Ditz, A. Yildirim, and D. van der Spoel, “Properties of liquids in molecular dynamics simulations with explicit long-range Lennard Jones interactions,” J. Chem. Theory Comput. **11**, 2938–2944 (2015).

[40]D. R. Lide, *CRC Handbook of Chemistry and Physics*, 90th ed. (CRC Press, Cleveland, Ohio, 2009).

[41]C. L. Yaws, *Yaws’ Handbook of Thermodynamic Properties for Hydrocarbons and Chemicals* (Knovel, 2009), http://www.knovel.com.

[42]Thermophysical Properties of Pure Substances & Mixtures, DECHEMA Gesellschaft für Chemische Technik und Biotechnologie e.V., 2011, http://i-systems.dechema.de/detherm.

[43]C. L. Yaws, *Yaws’ Critical Property Data for Chemical Engineers and Chemists* (Knovel, 2012), http://www.knovel.com.

[44]L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, “Assessment of Gaussian-3 and density functional theories for a larger experimental test set,” J. Chem. Phys. **112**, 7374–7383 (2000).

[45]L. A. Curtiss, P. C. Redfern, and K. Raghavachari, “Assessment of Gaussian-3 and density-functional theories on the G3/05 test set of experimental energies,” J. Chem. Phys. **123**, 124107 (2005).

[46]J. M. Martin, “Heats of formation of perchloric acid, $HClO_4$, and perchloric anhydride, Cl_2O_7. Probing the limits of W1 and W2 theory,” J. Mol. Struct. **771**, 19–26 (2006).

[47]D. Raymand, A. C. van Duin, M. Baudin, and K. Hermansson, “A reactive force field (reaxff) for zinc oxide,” Surf. Sci. **602**, 1020–1031 (2008).

[48]A. Karton, S. Parthiban, and J. M. L. Martin, “Post-CCSD(T) *ab initio* thermochemistry of halogen oxides and related hydrides XOX, XOOX, HOX, XOn, and HXOn (X = F, Cl), and evaluation of DFT methods for these systems,” J. Phys. Chem. A **113**, 4802–4816 (2009).

[49]D. Trogolo and J. S. Arey, “Benchmark thermochemistry of chloramines, bromamines, and bromochloramines: Halogen oxidants stabilized by electron correlation,” Phys. Chem. Chem. Phys. **17**, 3584–3598 (2015).

[50]M. V. D. Silva, R. Custodio, and M. H. M. Reis, “Determination of enthalpies of formation of fatty acids and esters by density functional theory calculations with an empirical correction,” Ind. Eng. Chem. Res. **54**, 9545–9549 (2015).

[51]R. S. Berry, “Correlation of rates of intramolecular tunneling processes, with application to some Group V compounds,” J. Chem. Phys. **32**, 933–938 (1960).

[52]P. Russegger and J. Brickmann, “Quantum states of intramolecular nuclear motion with large amplitudes: Pseudorotation of trigonal bipyramidal molecules,” J. Chem. Phys. **62**, 1086–1093 (1975).

[53]A. Caligiana, V. Aquilanti, R. Burcl, N. C. Handy, and D. P. Tew, “Anharmonic frequencies and Berry pseudorotation motion in PF_5,” Chem. Phys. Lett. **369**, 335–344 (2003).

[54]J. E. Kilpatrick, K. S. Pitzer, and R. Spitzer, “The thermodynamics and molecular structure of cyclopentane,” J. Am. Chem. Soc. **69**, 2483–2488 (1947).

[55]D. O. Harris, G. G. Engerholm, C. A. Tolman, A. C. Luntz, R. A. Keller, H. Kim, and W. D. Gwinn, “Ring puckering in five-membered rings. I. General theory,” J. Chem. Phys. **50**, 2438–2445 (1969).

[56]J. Zheng, T. Yu, and D. G. Truhlar, “Multi-structural thermodynamics of C–H bond dissociation in hexane and isohexane yielding seven isomeric hexyl radicals,” Phys. Chem. Chem. Phys. **13**, 19318–19324 (2011).

[57]K. S. Pitzer and W. D. Gwinn, “Energy levels and thermodynamic functions for molecules with internal rotation. I. Rigid frame with attached tops,” J. Chem. Phys. **10**, 428–440 (1942).

[58]K. S. Pitzer, “Energy levels and thermodynamic functions for molecules with internal rotation: II. Unsymmetrical tops attached to a rigid frame,” J. Chem. Phys. **14**, 239–243 (1946).

[59]A. P. Scott and 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**, 16502–16513 (1996).

[60]J. E. Mayer, S. Brunauer, and M. G. Mayer, “The entropy of polyatomic molecules and the symmetry number,” J. Am. Chem. Soc. **55**, 37–53 (1933).

[61]R. M. Dannenfelser, N. Surendran, and S. H. Yalkowsky, “Molecular symmetry and related properties,” SAR QSAR Environ. Res. **1**, 273–292 (1993).

[62]R.-M. Dannenfelser and S. H. Yalkowsky, “Estimation of entropy of melting from molecular structure: A non-group contribution method,” Ind. Eng. Chem. Res. **35**, 1483–1486 (1996).

[63]C. A. Wulff, “Determination of barrier heights from low-temperature heat-capacity data,” J. Chem. Phys. **39**, 1227–1234 (1963).

[64]A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kallay, and J. Gauss, “W3 theory: Robust computational thermochemistry in the kJ/mol accuracy range,” J. Chem. Phys. **120**, 4129–4141 (2004).

[65]L. Goerigk and S. Grimme, “A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions,” Phys. Chem. Chem. Phys. **13**, 6670–6688 (2011).

[66] E. E. Bolton, S. Kim, and S. H. Bryant, "PubChem3D: Conformer generation," J. Cheminf. **3**, 4 (2011).

[67] A. Supady, V. Blum, and C. Baldauf, "First-principles molecular structure search with a genetic algorithm," J. Chem. Inf. Model. **55**, 2338–2348 (2015).

[68] Y. Lu, Y. Liu, Z. Xu, H. Li, H. Liu, and W. Zhu, "Halogen bonding for rational drug design and new drug discovery," Expert Opin. Drug Discovery **7**, 375–383 (2012).

[69] P. Metrangolo, H. Neukirch, T. Pilati, and G. Resnati, "Halogen bonding based recognition processes: A world parallel to hydrogen bonding," Acc. Chem. Res. **38**, 386–395 (2005).

[70] P. Politzer, P. Lane, M. C. Concha, Y. Ma, and J. S. Murray, "An overview of halogen bonding," J. Mol. Model. **13**, 305–311 (2007).

[71] S.-Y. Lu and I. Hamerton, "Recent developments in the chemistry of halogen-free flame retardant polymers," Prog. Polym. Sci. **27**, 1661–1712 (2002).