



Budapest University of Technology and Economics

Faculty of Chemical Technology and Biotechnology

George A. Oláh Doctoral School of Chemistry and Chemical Technology

Development and application of accurate thermochemical methods

Thesis Statements

Written by: Ádám Ganyecz

Advisor: Dr. József Csontos

Department of Physical Chemistry and Materials Science
MTA-BME Lendület Quantum Chemistry Research Group

2019

1 Introduction

Accurate thermochemical data are indispensable to provide meaningful forecast about any reaction and process. Design of chemical plants, modeling of climate change, or understanding biological systems, just to mention a few examples, would be unimaginable without these data available in orderly fashion in databases.

However, in recent decades, theoretical methods have been gaining traction in the field of thermochemistry. Since the pioneering work of Pople and associates dramatic progress has been made regarding the calculations of thermochemical quantities. Numerous composite model chemistries have been defined with diverse design goals including the subsequent G_n theories¹, the CBS² (complete basis set) models, the ccCA³ (correlation consistent composite approach), the Weizmann-n⁴, and the HEAT⁵ protocols. Also, the focal point analysis of Allen and Császár⁶ and the procedure of Feller, Peterson, and Dixon⁷ (FPD) should be mentioned here, and note that although they do not work with fixed recipes they are closely related to the aforementioned model chemistries.

Today it is well understood how to design a composite model chemistry which is capable of achieving high accuracy in calculating thermochemical quantities. Because the treatment of electron correlation is of the utmost importance, accurate models rely on the coupled-cluster (CC) expansion of the wavefunction. Unfortunately, high accuracy CC calculations require large basis sets and the steep scaling of such approaches limits the thermochemical applications to small systems. It is also known that the slow convergence of the CC methods in one-electron basis sets is due to the correlation

¹J. Chem. Phys. 94(11), 7221 (1991); J. Chem. Phys. 109(18), 7764 (1998); J. Chem. Phys. 126(8), 084108 (2007)

²J. Chem. Phys. 89(4), 2193 (1988); J. Chem. Phys. 94(9), 6081 (1991); J. Chem. Phys. 94(9), 6091 (1991); J. Chem. Phys. 101(7), 5900 (1994); J. Chem. Phys. 104(7), 2598 (1996)

³J. Chem. Phys. 124(11), 114104 (2006); J. Chem. Phys. 125(10), 104111 (2006); J. Chem. Phys. 127(15), 154117 (2007)

⁴J. Chem. Phys. 111(5), 1843 (1999); J. Chem. Phys. 120(9), 4129 (2004)

⁵J. Chem. Phys. 121(23), 11599 (2004); J. Chem. Phys. 125(6), 064108 (2006); J. Chem. Phys. 128(11), 114111 (2008)

⁶J. Chem. Phys. 99(6), 4638 (1993); J. Chem. Phys. 108(23), 9751 (1998)

⁷Annu. Rep. Comput. Chem. 8, 1 (2012); Theor. Chem. Acc. 131(1), 1079 (2012)

cusp and it can be alleviated using explicitly correlated methods⁸. Recently, several implementations of explicitly correlated CC methods have become available. This triggered a series of investigations about the incorporation of explicitly correlated methods into some of the aforementioned model chemistries. In some cases, these sophisticated model chemistries can provide as accurate data as that obtained from cutting edge experimental methods. Event hough, the size of the studied systems limits the applicable level of theory, for experimentally challenging species with short lifetime or very reactive nature, theoretical models can provide benchmark quality data.

Several compilation exist for thermochemical data (JPL⁹, NIST-JANAF¹⁰, Burcat¹¹, CODATA¹², etc.) however, as the amount of data increases the burden of maintaining, critically evaluating, and keeping up-to-date them is getting more and more cumbersome. Usually, the most accurate value or the simple weighted average of the best ones are recommended for use by these databases, and, these selections have their own drawbacks. For instance, the error is propagated through all the subsequent steps, and there is no feedback to the preceding determinations; only one, or a small group, of the results is used for the determination; in order to take account of new findings, every step should be done again.

Another approach is the thermochemical network (TN), when all available data are used simultaneously to determine the thermochemical values in question. The Active Thermochemical Tables (ATcT) of Ruscic¹³ and the NEAT of Császár and Furtenbacher¹⁴ are representative examples for this approach. However, the results provided by TNs are only as good as the quality of the sources, therefore accurate thermochemical data is still needed by either experiments or theoretical calculations.

One field, where these compilations are useful is atmospheric chemistry. Climate

⁸Z. Phys. 54(5), 347 (1929); Theor. Chim. Acta 68(6), 445 (1985); Chem. Phys. Lett. 199(5), 497 (1992); J. Chem. Phys. 116(15), 6397 (2002); Chem. Phys. Lett. 398(1), 56 (2004)

⁹Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies

¹⁰NIST-JANAF Thermochemical Tables

¹¹<http://garfield.chem.elte.hu/Burcat/burcat.html>

¹²CODATA Key Values for Thermodynamics

¹³<http://atct.anl.gov/>

¹⁴Chem. Eur. J. 16(16), 4826 (2010)

change, which is probably caused by greenhouse gases (GHG) and ozone depleting agents as a byproduct of human activities, is one of the most paramount challenge to overcome in the 21st century. Because chlorofluorocarbons (CFCs) and hydrochlorofluorocarbons (HCFCs) are ozone depleting agents, they have been covered by the Montreal Protocol, and consequently, they have been replaced with ozone friendly alternatives, such as hydrofluorocarbons (HFCs) and perfluorocarbons (PFCs) in the chemical industry. However, these substances have considerably larger contribution to global warming than CFCs and HCFCs. HFCs and other GHGs are subject of the Kyoto Protocol. To prolongate its impacts, because the second commitment period for it ends in 2020, the Paris Agreement was negotiated and adopted by the representatives of 196 parties in 2015.

To study and understand the effects of anthropogenic activities using the so-called chemistry-climate models, accurate kinetic and thermodynamic data for atmospherically important reactions and corresponding species are inevitable. Several databases already contain substantial information, however, some data is still inaccurate and discrepancies also exist among these databases. These variations are due to the fact that the chemistry of the troposphere and stratosphere is dominated by free radical reactions, and these open-shell systems pose a challenge to the experimental and theoretical investigations. Nevertheless, high accuracy calculations have become the method of choice for these species because they can yield competitive, and even superior data to experiments.

Surprisingly, another family of molecules without well-established formation enthalpies consists of nucleobases. To have accurate energetics of the processes involving nucleobases, and to improve our understanding of structure-property relationships, high-quality thermochemical data are crucial. Despite of these needs, the available experimental data are in disarray, especially the sublimation enthalpies. For example, in the case of adenine the sublimation enthalpy values are scattered in a 30 kJ/mol range. Nevertheless, high-level *ab initio* calculations in the gas phase can help sorting out these discrepancies if the heat of formation values are well-established in the solid

phase. However, there is also some uncertainty surrounding the combustion enthalpy values, which have been used to derive the heat of formation of the nucleobases.

On the one hand, from a practical, thermochemical point of view, the main issue is the low vapor pressure of the nucleobases, which makes the measurement of the heat of sublimation difficult. On the other hand, the temperature correction to the sublimation enthalpy can also be problematic. The currently utilized method is based on an empirical correction. Meanwhile, high-quality heat capacities can be easily computed with quantum chemical methods nowadays.

During my PhD work I have had two goals kept in my mind. On the one hand, I wanted to develop an accurate thermochemical protocol, as well as apply previously defined ones, to obtain reliable theoretical data for the aforementioned species. On the other hand, I also wanted to combine my theoretical results with critically evaluated literature data in various TNs, thereby providing the best possible thermochemical values based on our current knowledge.

2 Technical details

For the calculations of fluoroethyl radicals, a HEAT-like model chemistry, the diet-HEAT protocol¹⁵ was utilized. It contains iterative triple and perturbative quadruple excitations in CC theory, as well as scalar relativistic and diagonal Born–Oppenheimer corrections with correlation-consistent basis sets relying on the following scheme $E_{\text{TOT}} = E_{\text{HF}}^{\infty} + \Delta E_{\text{CCSD(T)}}^{\infty} + \Delta E_{\text{CCS DT}} + \Delta E_{\text{CCS DT(Q)}} + \Delta E_{\text{core}}^{\infty} + \Delta E_{\text{DBOC}} + \Delta E_{\text{Rel}} + \Delta E_{\text{ZPE}}$.

An explicitly correlated variant of the above protocol, called diet-HEAT-F12 was also developed during this work. The methods and basis sets used for the calculation of the individual components are selected to reproduce, as close as possible, without using any fitted parameters, the benchmark original HEAT contributions. This protocol was applied to fluorinated and chlorinated methanes and ethanes, as well as to nucleobases.

To get the desired thermodynamic functions the standard equations of statistical thermodynamics were used within the rigid rotor-harmonic oscillator model. To correct the errors of the harmonic oscillator model for internal coordinates with large amplitude motions the one-dimensional hindered rotor model (1D-HR) was applied¹⁶.

Then, the quality of thermochemical data was further improved by using the TN approach. In the framework of which, one can write the standard reaction enthalpy at 298 K, $\Delta_r H_{298}^{\circ}$, for a given reaction as

$$\Delta_r H_{298}^{\circ}(i) = \sum_j^n \nu_{ij} \Delta_f H_{298}^{\circ}(j) + \epsilon(i).$$

In the above $\epsilon(i)$ is the uncertainty of $\Delta_r H_{298}^{\circ}(i)$, $\Delta_f H_{298}^{\circ}(j)$ is the heat of formation of the species j at 298 K, and ν_{ij} is its corresponding stoichiometric number. This is an overdetermined system of linear equations, and can be regarded as a multiple linear regression problem. Its solution was located using the iteratively reweighted least squares method.

¹⁵J. Phys. Chem. A 114(50), 13093 (2010)

¹⁶J. Chem. Edu. 77(11), 1495 (2000); Theor. Chem. Acc. 118(5), 881 (2007)

3 Results

At first, fluoroethyl radicals were studied with the diet-HEAT model chemistry, which provided the most reliable theoretical values at the time for the 8 species studied. This model chemistry was more accurate than those had been utilized by previous studies, and in contrast with prior calculations, it also had a well-defined uncertainty measure. The importance of the excitations beyond the perturbative triples was noteworthy. The contribution to the heats of formation obtained from CCSDT(Q) calculations increased with the size of the radical, i.e., with the number of fluorine atoms, and it was as large as 5 kJ/mol for $\text{CF}_3\text{--CF}_2$. Relativistic corrections showed a similar trend reaching their maximum, 5 kJ/mol, for $\text{CF}_3\text{--CF}_2$. Therefore, they seem to be unavoidable in thermochemical protocols seeking chemical accuracy. In those cases where conformer isomerism occurred for the radicals conformer, specific values along with Boltzmann averaged quantities were also reported. Although previous investigators had noted the presence of conformers and predicted the energy order for the conformers in line with my findings, thermodynamic functions had solely been reported for the most stable isomer. It was also notable that an arrangement of $\text{CH}_2\text{F--CHF}$ that had been found to be a conformer with the HF/6-31G(d) method was not a minimum anymore, either at the MP2/cc-pVTZ or CCSD(T)/cc-pVTZ level of theory. Although, originally, in the case of $\text{CF}_3\text{--CH}_2$ and $\text{CF}_3\text{--CHF}$ there were large discrepancies between my calculations and the experimental values, these were eliminated by updating the auxiliary data used in the original experimental report with currently recommended ones. Thus, it was observed that my results not only agreed well with the best experimental data but even had smaller uncertainties attached.

The next work introduced an *ab initio* composite model chemistry with moderate-cost, baptized as diet-HEAT-F12. It was developed to calculate accurate thermochemical data for species containing first- and second-row atoms, and was designed to be parameter free, i.e., nor experimental, neither theoretical datasets were used to determine scaling factors or extrapolation coefficients. In addition, no experimental data

was utilized during the development, i.e., exclusively high accuracy theoretical data was used to guide the selection of the levels of theory for calculating the various components of the protocol. The model proposed has a 0.5 kcal/mol 95% confidence interval for the benchmark HEAT values. It was also shown by using a test set of accurate TAEs that the diet-HEAT-F12 protocol yielded the most accurate data among the conceptually similar model chemistries, W3-F12, W3X, and W3X-L, investigated.

In the next study, diet-HEAT-F12 and thermochemical network approach were applied to get the heat of formation values for 50 atmospherically relevant species including various chlorocarbons, CFCs, HFCs, HCFCs, and PFCs. In 28 out of the 50 cases, this method presented the most reliable estimates for the heats of formation. For the other species, the results agreed well with the most accurate data from ATcT or JPL, except CF_3-CF_3 , CF_3 , CHF_3 , and CH_2F_2 , where small discrepancies occurred, i.e., the error bars did not overlap. The largest non-overlapping region, 0.5 kJ/mol, obtained for $\Delta_f H_{298}^\circ(\text{CHF}_3)$. To analyze the results of the TN, well-known statistical metrics were used, namely the DFBETA and SDFBETA values, as well as the residuals. According to these metrics a few additional species, CF_4 , CH_3Cl , CHCl_3 , CCl_4 , CFCl_3 , CF_2Cl_2 , existed in the diet-HEAT-F12-network, for which further investigation is still needed to collect more established heats of formation. This study also showed that conventional CCSD(T) calculations, even with the 5Z basis sets, could suffer from BSSE, while explicitly correlated methods are hardly affected by this error. Therefore, the usage of the diet-HEAT-F12 model chemistry generally is recommended over the conventional diet-HEAT. Furthermore, the work highlighted the importance of the quality and reliability of the data provided by the theoretical model and supplied into the TN. It was observed that, low-quality theoretical results undermine the benefits of the TNs. Obviously, the most vulnerable cases are those where there is no accurate experimental data available to challenge/control the influence of inaccurate calculations.

During the last part of my PhD work, reliable and consistent heats of formation and sublimation enthalpies had been established for uracil, thymine, cytosine, and adenine. On the one hand, the diet-HEAT-F12 protocol was utilized to provide accurate

theoretical $\Delta_f H_{298}^\circ$ (g) data, by taking into account the various tautomers of cytosine and the large amplitude motion for thymine. On the other hand, the available combustion and sublimation data were revised and recalculated using accurate theoretical heat capacities. These data and additional well-established experimental values along with my diet-HEAT-F12 results were arranged into a TN. The solution of the TN yielded the best estimates for the studied thermochemical quantities. Although, in case of uracil, thymine, and cytosine, the results were in accord with the values considered to be the most accurate at the time, my analysis showed that previous error bars were too optimistic. In the case of adenine, the available data was not consistent with my results. This could be attributed to the fact that different preparations of the sample lead to different degree of crystallinity. Nevertheless, this concern required further investigation, which was out of the scope of the study. Solvation free energies were also calculated with various methods and compared to the data derived from experiments. For thymine, cytosine, and adenine our study provided the best estimates for solvation free energies.

4 Applicability

The diet-HEAT-F12 model chemistry developed during my PhD work has already been used for the determination of heat of formation of various halogenated species, and even nucleobases, and it can be applied to similar medium-sized molecules. The results provided in this work can also be used as benchmark data for the development of other methods.

5 New scientific results

- I It was found that, the magnitude of the CCSDT(Q) and relativistic contributions can be as large as, or even more, than 5 kJ/mol each – even with systems containing 7 atoms. Because this observation was confirmed for both fluoroethyl radicals and nucleobases, it might be regarded as a generally valid phenomenon, and when chemical accuracy is sought, calculation of these terms is recommended.[1,4]
- II An *ab initio* composite model chemistry with moderate-cost, baptized as diet-HEAT-F12, was developed to calculate accurate thermochemical data for species containing first- and second-row atoms, and was designed to be parameter free. The model proposed has better performance compared to similar non-parameter free explicitly correlated model chemistries. Formula for its 95% confidence interval was also worked out.[2]
- III Studying the importance of the quality and reliability of the data provided by the theoretical model and supplied into the thermochemical network, it was observed, that the TNA can not compensate for theoretical results with uncertainty of 8 kJ/mol or more and give unreliable solution.[3]
- IV Diet-HEAT-F12 protocol was utilized along with thermochemical network approach and provided the most reliable heats of formation for 28 out of 50 cases for fluorinated and chlorinated methanes and ethanes. Additionally, according to the statistical analysis 6 species require further investigation for more established data. Furthermore, in case of 46 species out of 50 the diet-HEAT-F12 results are consistent with the results from the network, i.e. the diet-HEAT-F12 results are themselves provide a parameter free consistent set of data.[3]
- V Based on diet-HEAT-F12 calculations and the TN approach for uracil, thymine, and cytosine, the most reliable heat of formation values were found. In the case of adenine, due to the discrepancy regarding the previous literature data, my diet-HEAT-F12 result is recommended.[4]

6 Publications and presentations

6.1 Publications related to the thesis

- 1 Accurate Theoretical Thermochemistry for Fluoroethyl Radicals,
Á. Ganyecz, M. Kállay, and J. Csontos,
J. Phys. Chem. A, **2017**, 121:1153
DOI: 10.1021/acs.jpca.6b12404, IF: 2.836, citations: 3
- 2 Moderate-Cost *Ab Initio* Thermochemistry with Chemical Accuracy,
Á. Ganyecz, M. Kállay, and J. Csontos,
J. Chem. Theory Comput., **2017**, 13:4193
DOI: 10.1021/acs.jctc.7b00607, IF: 5.399 , citations: 6
- 3 High Accuracy Quantum Chemical and Thermochemical Network Data for the Heats of Formation of Fluorinated and Chlorinated Methanes and Ethanes,
Á. Ganyecz, M. Kállay, and J. Csontos,
J. Phys. Chem. A, **2018**, 122:5993
DOI: 10.1021/acs.jpca.8b00614, IF: 2.836, citations: 5
- 4 Thermochemistry of Uracil, Thymine, Cytosine, and Adenine,
Á. Ganyecz, M. Kállay, and J. Csontos,
J. Phys. Chem. A, **2019**, 123:4057
DOI: 10.1021/acs.jpca.9b02061, IF: 2.836, citations: 0

6.2 Other publications

- Theoretical and Thermochemical Network Approaches To Determine the Heats of Formation for HO₂ and Its Ionic Counterparts,
Á. Ganyecz, J. Csontos, B. Nagy, and M. Kállay
J. Phys. Chem. A, **2015**, 119:1164
DOI: 10.1021/jp5104643, IF: 2.836, citations: 6

- Oxygen reduction reaction on TiO₂ rutile (110) surface in the presence of bridging hydroxyl groups,
 Á. Ganyecz, P. D. Mezei, and M. Kállay
Comput. Theor. Chem., **2019**, 1168:112607
 DOI: 10.1016/j.comptc.2019.112607, IF: 1.344 , citations: 0
- Synthesis and characterization of isophorondiamine-based oligoamides: catalytic effect of amides during the curing of epoxy resins,
 L. Kárpáti, Á. Ganyecz, T. Nagy, G. Hamar, E. Banka, M. Kállay, and V. Vargha
Polym. Bull., **2019**
 DOI: 10.1007/s00289-019-02987-6, IF: 1.858, citations: 0

6.3 Presentations

- Oral presentation: **XXXVIII. Kémiai Előadó Napok**, 2015, Szeged, Hungary
 Pontos termokémiai számítások fluorozott etilgyökökre, Á. Ganyecz, J. Csontos, and M. Kállay
- Oral presentation: **Innováció a Természettudományban - Doktorandusz konferencia**, 2015, Szeged, Hungary
 Aktív termokémiai táblázatok- a kvantumkémiai számítások szerepe, Á. Ganyecz, J. Csontos, and M. Kállay
- Poster presentation: **Magical Mystery Tour of Electron Correlation in Honour of Jeppe Olsen on the Occasion of his 60th Birthday**, 2016, Oslo, Norway
 Thermochemistry of fluoroethyl radicals, Á. Ganyecz, J. Csontos, and M. Kállay
- Oral presentation: **Graduate Conference on Theoretical Chemistry 2016**, 2016, Keszthely, Hungary
 High-Accuracy Thermochemistry of Fluoroethane Radicals, Á. Ganyecz, J. Csontos, and M. Kállay

- Poster presentation: **16th International Congress of Quantum Chemistry**, 2018, Menton, France
Accurate Theoretical Thermochemistry of Nucleobases, Á. Ganyecz, J. Csontos, and M. Kállay
- Poster presentation: **9th Molecular Quantum Mechanics Conference**, 2019, Heidelberg, Germany
Thermochemistry of Uracil, Thymine, Cytosine, and Adenine, Á. Ganyecz, J. Csontos, and M. Kállay