A systematic method to estimate and validate enthalpies of formation using error-cancelling balanced reactions

Philipp Buerger[a], Jethro Akroyd[a], Sebastian Mosbach[a], Markus Kraft[a,b,*]

[a] Department of Chemical Engineering and Biotechnology, University of Cambridge, West Cambridge Site, Philippa Fawcett Drive, Cambridge CB3 0AS, United Kingdom
[b] School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive 637459, Singapore

ARTICLE INFO

ABSTRACT

This paper presents an automated framework that uses overlapping subsets of reference data to systematically derive an informed estimate of the standard enthalpy of formation of chemical species and assess the consistency of the reference data. The theory of error-cancelling balanced reactions (EBRs) is used to calculate estimates of the standard enthalpy of formation. Individual EBRs are identified using linear programming. The first part of the framework recursively identifies multiple EBRs for specified target species. A distribution of estimates can then be determined for each species from which an informed estimate of the enthalpy is derived. The second part of the framework iteratively isolates inconsistent reference data and improves the prediction accuracy by excluding such data. The application of the framework is demonstrated for test cases from organic and inorganic chemistry, including transition metal complexes. Its application to a set of 920 carbon, hydrogen and oxygen containing species resulted in a rapid decrease of the mean absolute error for estimates of the enthalpy of formation of each species due to the identification and exclusion of inconsistent reference data. Its application to titanium-containing species identified that the available reference values of $TiOCl$ and $TiO(OH)_2$ are inconsistent and need further attention. Revised values are calculated for both species. A comparison with popular high-level quantum chemistry methods shows that the framework is able to use affordable density functional theory (DFT) calculations to deliver highly accurate estimates of the standard enthalpy of formation, comparable to high-level quantum chemistry methods for both hydrocarbons and transition metal complexes.

© 2017 The Combustion Institute. Published by Elsevier Inc. All rights reserved.

1. Introduction

The development of automated procedures, for analysing chemical species and mechanisms [1–6], facilitate the investigation of progressively complex reaction systems. The availability of large sets of consistent chemical data is of key importance. Many data sets used by such tools are collated literature data [7,8] that are held in repositories [9–19] and are used for benchmarking computational methods [20–25]. Alongside these opportunities there remain challenges. One question is concerned with the consistency of chemical data (see e.g., [26–31]). Using accurate single-point level calculations to validate one species at a time is computationally demanding and quickly becomes intractable for large systems. This paper therefore presents a solution to this problem and provides a framework to systematically evaluate the consistency of thermochemical data for chemical species.

Consistent and accurate thermochemical data, such as species enthalpies, heat capacities, and Gibbs free energies, are an essential part of any detailed chemical model. The Active Thermochemical Tables (ATcT) [12,13,32,33] for example define a network of known thermochemical data, which relies on multiple reported experimental measurements and accurate theoretical high-level estimates for each species. A final set of reference values is proposed by processing the thermochemical network in a self-consistent fashion, until internal consistency is achieved. The uncertainties of the reference values are used to weight the contribution of each reference value within the thermochemical network. The systematic combination of multiple reference values allows for accurate and internally consistent estimates of thermochemical data.

The standard enthalpy of formation is a fundamental parameter required to calculate accurate values of the enthalpy and Gibbs free energy changes of the reactions in a chemical model. Inconsistencies in the standard enthalpy of formation could lead to significant

errors affecting the accuracy, predictive performance and quality of any model using such data.

In the past, different methods have been introduced to estimate the enthalpies of formation. The simplest are *additive* or *group contribution methods* [34–37]. They rely on the regularity of molecular and structural groups. They are computationally cheap and predictive in nature. To achieve qualitatively accurate results, their application is limited to well studied systems with precisely and accurately defined functional groups [38,39]. Among others these include the Benson group additivity method [34,35]. *Molecular mechanics methods* are computationally less demanding than other methods but are not universally applicable because they rely on empirical parameters and correction terms [39]. Electronic structure calculations at a high level of theory are used by *quantum chemistry methods* to estimate the enthalpies of formation. This type of single-point calculation is computationally demanding. The errors scale with the size of the molecule [40,41], and the calculations become intractable for large molecules [42,43]. In addition, care must be taken to choose the right level of theory [42,44,45] and various correction terms are needed to achieve consistent and accurate estimates. HEAT [46–48] is an example of a purely theoretical method, which considers various correction terms, used to estimate highly accurate standard enthalpies of formation for very small molecules.

Fortunately, the errors incurred in electronic structure calculations are systematic. Different methods have been developed to reduce and cancel the impact of these errors on estimates of the enthalpy of formation. Among these are the bond additivity correction (BAC) [49–52] and the atom additivity correction (AAC) [38,53]. Both rely on predefined parameters associated with the level of theory used in the calculation.

Error-cancelling balanced reactions (EBRs) exploit structural and electronic similarities between the species in a reaction to reduce the impact of the inherited systematic errors. The standard enthalpy of formation from an EBR is calculated based on the application of Hess' Law to the reaction. This method has been applied to a variety of different systems (see e.g., [5,54–58]. The absence of any empirical parameters makes this method suitable for automation. The total electronic energies for all species in the reaction and the enthalpies of formation need to be known (experimentally or theoretically) for all except one species, for which the unknown enthalpy of formation can be estimated. This is achieved by re-organising the equation defining Hess' Law

$$\Delta_r H^{\circ}_{298.15\,\text{K}} = \sum_{s\in S^P} \nu(s)\Delta_f H^{\circ}_{298.15\,\text{K}}(s) - \sum_{s\in S^R} \nu(s)\Delta_f H^{\circ}_{298.15\,\text{K}}(s), \quad (1)$$

to solve for the unknown enthalpy of formation

$$\nu(s_T)\Delta_f H^{\circ}_{298.15\,\text{K}}(s_T) = \sum_{s\in S^P} \nu(s)\Delta_f H^{\circ}_{298.15\,\text{K}}(s)$$
$$-\sum_{s\in S^R\setminus\{s_T\}} \nu(s)\Delta_f H^{\circ}_{298.15\,\text{K}}(s) - \Delta_r H^{\circ}_{298.15\,\text{K}}, \quad (2)$$

where $\nu(s)$ is the stoichiometric coefficient and $\Delta_f H^{\circ}_{298.15\,\text{K}}(s)$ is the standard enthalpy of formation of species s, S^P and S^R are the set of products and reactants in the reaction and $\Delta_r H^{\circ}_{298.15\,\text{K}}$ is the reaction enthalpy. The standard enthalpies of formation on the right-hand side of Eq. (2) are known. The standard enthalpy of formation of the target species s_T on the left-hand side of Eq. (2) is to be calculated. The method requires the identification of suitable EBRs fulfilling a set of constraints defined by the type of EBR.

Since the introduction of EBRs by Pople and co-workers [59,60], several types of EBRs have been proposed [57,59–66]. For example, isogyric, isodesmic, hypohomodesmotic, homodesmotic and hyperhomodesmotic reactions. The use of EBRs has been shown to en-

able the calculation of accurate estimates of the enthalpy of formation on the back of affordable electronic structure calculations. Generally, the more structural and electronic similarity that is preserved by the reaction, the more accurate the resulting estimate of the enthalpy of formation.

In our previous work [67] a high-level description was presented of an abstract and systematic framework to validate thermochemical data for chemical species and recommend what future experiments or calculations would be required to improve the data. The purpose of the current work is to give a detailed description of the algorithms used by the framework to identify error-cancelling balanced reactions (EBRs) and to calculate informed estimates of the standard enthalpies of formation. The framework facilitates the assessment of the consistency of chemical data, in this case for the standard enthalpy of formation. This is achieved by iteratively isolating potentially inconsistent reference data. By excluding such data, an improved prediction accuracy is achieved. The performance of the framework is demonstrated using four different reaction classes and test cases from organic and inorganic chemistry, including transition metal complexes. The required ground state energies and vibrational frequencies of reference species were calculated using affordable DFT. Calculated estimates of the standard enthalpy of formation for hydrocarbons and titanium-containing species are compared against calculated values using popular high-level quantum chemistry methods.

2. Methodology

This section presents a detailed description of the algorithms used to identify a set of EBRs, calculate informed estimates of the standard enthalpy of formation for a species and assess the consistency of the required reference data. The algorithms are implemented as part of an automated and systematic framework to estimate the standard enthalpy of formation for a set of target species. A set of reference species, each consisting of the total electronic energy, the molecular connectivity, the spin multiplicity and a known enthalpy of formation, is required.

The definitions of the EBRs considered in this work are given in Section 2.1, followed by a description of the electronic structure calculations in Section 2.2. The reference data are described in Section 2.3, followed by a detailed algorithmic description of the framework in Section 2.4.

2.1. Types of error-cancelling balanced reactions

Many different reaction classes of EBRs have been proposed [57,59–66]. The following types of EBRs are used in this work.

Isogyric reactions (reaction class RC1) are the least restrictive and only conserve the number of spin pairs on either side of the reaction. An example for an isogyric reaction is given in Fig. 1.

Isodesmic reactions (reaction class RC2) conserve the number of each type of bond on either side of the reaction. No constraint is placed on the chemical environment near each bond. An example for an isodesmic reaction is given in Fig. 2.

Reaction class RC3 extends the concept of isodesmic reactions. The identity and the total number of additional atoms bonded to the atom on either side of the bond is conserved in addition to the number of each type of bond on either side of the reaction. Figure 3 presents an example of such a reaction.

Reaction class RC4 extends the concept of RC3. The bond type and identity of each neighbouring atom, including the total number of additional atoms bonded to the neighbouring atoms, is conserved in addition to the constraints imposed by RC3. Possible constraints for propylene glycol ($C_3H_8O_2$) are presented in Fig. 4 and an example reaction is given in Fig. 5.