

Available online at www.sciencedirect.com

Chemical Physics Letters 417 (2005) 185–189

www.elsevier.com/locate/cplett

The effect of stoichiometry on ab initio-based thermochemistry predictions

Ilie Fishtik, David Urban, Jennifer Wilcox *

Department of Chemical Engineering, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, United States

Received 29 June 2005; in final form 26 July 2005

Abstract

The effect of various reaction schemes used to convert total reaction enthalpies obtained via ab initio methods into enthalpies of formation has been studied employing a new optimization technique. The reaction schemes, conventional, isostoichiometric, and isodesmic have been defined and generated for 47 species for which highly accurate experimental enthalpies of formation in addition to total reaction enthalpies at 298 K at G2 and G3 levels of theory are known. The main finding is that the effect of reaction schemes in ab initiobased thermochemistry predictions is small, especially, for enthalpy predictions involving high-level ab initio methods. 2005 Elsevier B.V. All rights reserved.

1. Introduction

Specific reaction schemes based on a set of reference species are required to convert ab initio-based total enthalpies of reaction into conventional thermochemical data, such as enthalpies of formation, where the accuracy of these theoretical predictions is governed by the chosen reaction scheme. Thus, various reaction schemes such as atomization [\[1\],](#page-4-0) isodesmic [\[2\],](#page-4-0) homodesmic [\[3\]](#page-4-0), bond separation [\[4\]](#page-4-0), group equivalent [\[5\]](#page-4-0), group additivity [\[6\]](#page-4-0), ring conserved isodesmic reactions [\[7\]](#page-4-0), etc., are currently in use. It is believed that more sophisticated reaction schemes, in particular, reactions that preserve different fragments of chemical species, i.e., bonds or groups of atoms, can cancel the errors that exist in electron correlation energies more completely, resulting in more accurate theoretical thermochemistry predictions. The effect of reaction schemes on the accuracy of ab initio thermochemistry has been intensively studied. These studies, however, were based on a direct comparison of the per-

E-mail address: jwilcox@wpi.edu (J. Wilcox).

formance of various reaction schemes for a limited number of species and individual reactions. As a result it is difficult to comprehend the effect of reaction schemes on the accuracy of ab initio-based thermochemistry predictions.

The purpose of this Letter is to evaluate the error produced by various reaction schemes when converting the total ab initio-based reaction enthalpy predictions into enthalpies of formation by employing a recently proposed optimization procedure [\[8\].](#page-4-0)

2. Notation and definitions

We consider a set of *n* chemical species B_1, B_2, \ldots, B_n for which accurate experimental enthalpies of formation $\Delta H_{f,i}^{\text{exp}}$ $(i = 1, 2, \dots, n)$ as well as ab initio-based reaction enthalpies at 298 K, H_i^{ai} $(i = 1, 2, ..., n)$, are known. If ΔH_i^{ai} $(i = 1, 2, ..., n)$ are the enthalpies of formations of species evaluated from the ab initio-based reaction enthalpies via a certain reaction scheme, then $e_i =$ $\Delta H_{f,i}^{\text{exp}} - \Delta H_{f,i}^{ai}$ $(i = 1, 2, ..., n)$ is the error associated with the ab initio method. It is, therefore, convenient to define the vectors:

Corresponding author. Fax: $+1$ 508 831 5853.

^{0009-2614/\$ -} see front matter © 2005 Elsevier B.V. All rights reserved. doi:10.1016/j.cplett.2005.10.003

ARTICLE IN PRESS

$$
\Delta H_{\mathbf{f}}^{\text{exp}} = (\Delta H_{f,1}^{\text{exp}}, \Delta H_{f,2}^{\text{exp}}, \dots, \Delta H_{f,n}^{\text{exp}})^{\text{T}},
$$
\n(1)

$$
\Delta H_{\mathbf{f}}^{\mathbf{ai}} = (\Delta H_{f,1}^{ai}, \Delta H_{f,2}^{ai}, \dots, \Delta H_{f,n}^{ai})^{\mathrm{T}},\tag{2}
$$

$$
\mathbf{H}^{\mathbf{ai}} = (H_1^{ai}, H_2^{ai}, \dots, H_n^{ai})^{\mathrm{T}},\tag{3}
$$

$$
\mathbf{e} = (e_1, e_2, \dots, e_n)^{\mathrm{T}} = (\Delta H_{f,1}^{\mathrm{exp}} - \Delta H_{f,1}^{\mathrm{eig}}, \Delta H_{f,2}^{\mathrm{exp}} - \Delta H_{f,2}^{\mathrm{eig}}, \dots, \Delta H_{f,n}^{\mathrm{exp}} - \Delta H_{f,n}^{\mathrm{eig}})^{\mathrm{T}}.
$$
\n(4)

In what follows the vector e is referred to as error vector.

A reaction scheme is defined as a set of chemical reactions that preserve a specified number and type of molecular quantities such as elements, bonds, groups, etc. Let a set of *m* linearly independent reactions ρ_i ($j = 1,2,...,m$) be

$$
\rho = v\mathbf{B} = 0,\tag{5}
$$

where

$$
\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_m)^{\mathrm{T}},\tag{6}
$$

$$
\mathbf{B} = (B_1, B_2, \dots, B_n)^{\mathrm{T}},
$$

\n
$$
\begin{bmatrix} v_{11} & v_{12} & \dots & v_{1n} \end{bmatrix}
$$
 (7)

$$
\mathbf{v} = \begin{bmatrix} v_{21} & v_{22} & \dots & v_{2n} \\ \dots & \dots & \dots & \dots \\ v_{m1} & v_{m2} & \dots & v_{mn} \end{bmatrix} . \tag{8}
$$

Here v_{ii} ($i = 1,2,...,n$; $j = 1,2,...,m$) are the stoichiometric coefficients assumed to take negative values for reactants and positive values for products. It should be noticed that the two equality signs in Eq. (5) have different meanings, i.e., the first equality sign expresses the numbering of the reactions while the second equality sign expresses the fact that reactions are balanced. Each reaction is characterized by an experimental $\Delta H_{\rho_j}^{\text{exp}}$ $(j = 1, 2, ..., m)$ and ab initio $\Delta H_{\rho_j}^{ai}$ $(j = 1, 2, ..., m)$ enthalpy change defined traditionally as in [\[9\]](#page-4-0):

$$
\nu \Delta H_f^{\text{exp}} = \Delta H_\rho^{\text{exp}},\tag{9}
$$

$$
v\Delta H_{\rho}^{ai} = \Delta H_{\rho}^{ai},\tag{10}
$$

where

$$
\Delta H_{\rho}^{\text{exp}} = \left(\Delta H_{\rho_1}^{\text{exp}}, \Delta H_{\rho_2}^{\text{exp}}, \dots, \Delta H_{\rho_n}^{\text{exp}}\right)^{\text{T}},\tag{11}
$$

$$
\Delta H_{\rho}^{\text{ai}} = (\Delta H_{\rho_1}^{ai}, \Delta H_{\rho_2}^{ai}, \dots, \Delta H_{\rho_n}^{ai})^{\text{T}}.
$$
\n(12)

3. Theory

According to the approach briefly outlined in the preceding section, the minimization of the error vector is subject to a set of constraints expressing the stoichiometry and thermochemistry of a certain reaction scheme. Explicitly, these constraints are given by Eqs. (9) and (10) and involve the stoichiometric matrix v of a set of linearly independent reactions ρ as well as their experimental, ΔH_{ρ}^{exp} , and, ab initio, ΔH^{ai}_{ρ} , enthalpy changes. The goal is to evaluate the error vector e, as expressed in Eq. (4), for various reaction schemes. Most often the methodology is carried out based upon the following considerations. A subset of q species, referred to as reference species, are selected. These are normally the species for which the experimental enthalpies of formation are known with high accuracy. Without loss of generality, we assume that these are the first q species B_1, B_2, \ldots, B_q . The value of q is such that a uniquely balanced reaction involving the reference species B_1, B_2, \ldots, B_q and one of the remaining, training subset of species $B_{q+1}, \overline{B}_{q+2}, \ldots, \overline{B}_n$ may be generated. Let these $n - q$ reactions be

$$
\rho_j = \sum_{i=1}^q v_{ji} B_i + v_{j,q+j} B_{q+j} = 0; \quad j = 1, 2, \dots, n-q.
$$
 (13)

The ab initio-based enthalpies of formation predictions, $\Delta H_{f,q+1}^{ai}$, $\Delta H_{f,q+2}^{ai}$, ..., $\Delta H_{f,n}^{ai}$ of the species B_{q+1} , $B_{q+2},...,B_n$ are then determined from

$$
\sum_{i=1}^{q} v_{ji} \Delta H_{f,i}^{\text{exp}} + v_{j,q+j} \Delta H_{f,q+j}^{ai} = \sum_{i=1}^{q} v_{ji} H_{i}^{ai} + v_{j,q+j} H_{q+j}^{ai};
$$

\n
$$
j = 1, 2, ..., n-q.
$$
\n(14)

Clearly, within this approach only the errors in the enthalpies of formation of the training set of species B_{q+1}, B_{q+2} , $..., B_n$ are evaluated.

These errors, although physicochemically meaningful, are not representative since they depend on the selection of the reference species. This approach was also later criticized by Petersson et al. [\[10\]](#page-4-0) on the basis that the stoichiometric coefficients in the reaction schemes may be very high so that even an accuracy of 0.1 kcal/mol for reference species may still result in large errors in the thermochemistry of larger molecules. Instead, Petersson et al. [\[10\]](#page-4-0) proposed to enlarge the number of reference species substantially, thus allowing the introduction of the socalled bond additivity corrections. To minimize the effect of the selection of reference species on the error associated with an ab initio method we propose a different approach. Namely, for a given reaction scheme we evaluate simultaneously the errors of all species, i.e., without the separation of the species into reference and training subsets. In doing this we, first, subtract Eq. (10) from Eq. (9), which yields

$$
ve = \Delta \Delta H_{\rho}, \tag{15}
$$

where

$$
\Delta \Delta H_{\rho} = \Delta H_{\rho}^{\text{exp}} - \Delta H_{\rho}^{\text{ai}} \n= (\Delta H_{\rho_1}^{\text{exp}} - \Delta H_{\rho_1}^{ai}, \Delta H_{\rho_2}^{\text{exp}} - \Delta H_{\rho_2}^{ai}, \dots, \Delta H_{\rho_m}^{\text{exp}} - \Delta H_{\rho_m}^{ai})^{\text{T}}.
$$
\n(16)

Now, the error vector e may be determined by minimizing the product $e^{\mathsf{T}}e$ subject to the constraint given by Eq. (15). A formally (mathematically) identical problem has been considered previously in connection with the effect of stoichiometry in quantitative structure–property relationships [\[8\]](#page-4-0), and employing these results we have

$$
\mathbf{e} = \mathbf{v}^{\mathrm{T}} (\mathbf{v} \mathbf{v}^{\mathrm{T}})^{-1} \Delta \Delta \mathbf{H}_{\rho}.
$$
 (17)

It should be mentioned that although the set of linearly independent reactions, i.e., the stoichiometric matrix v , are generated arbitrarily, the error vector is independent of the choice of v . A proof of this statement may be found in [\[8\].](#page-4-0)

4. Reaction schemes

In this Letter, we consider three different types of reaction schemes. Employing the usual stoichiometric formalism a set of linearly independent reactions, or, the stoichiometric matrix v is generated, for each of the following types of reactions:

(a) Conventional chemical reactions. A reaction that preserves the number and type of chemical elements is referred to as a conventional chemical reaction. Let ε_{ik} ($i = 1, 2, \ldots, n$; $k = 1,2,\ldots,s$) be the number of elements E_k ($k = 1,2,\ldots,s$) in the species B_i ($i = 1, 2, ..., n$). The $n \times s$ matrix

$$
\mathbf{E}_{1} \quad \mathbf{E}_{2} \quad \dots \quad \mathbf{E}_{s}
$$
\n
$$
\mathbf{\varepsilon} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \dots & \varepsilon_{1s} \\ \varepsilon_{21} & \varepsilon_{22} & \dots & \varepsilon_{2s} \\ \dots & \dots & \dots & \dots \\ \varepsilon_{n1} & \varepsilon_{n2} & \dots & \varepsilon_{ns} \end{bmatrix} \quad \mathbf{B}_{1}
$$
\n
$$
\mathbf{B}_{2}
$$
\n(18)

is called the *formula matrix*. As is well known from chemical stoichiometry, the number of linearly independent conventional reactions is equal to $m = n - \text{rank}\varepsilon$. These may be generated as follows. Let, for instance

Then a set of linearly independent conventional reactions ρ_i $(i = 1,2,...,m)$ is given by

$$
\rho_{j} = \begin{vmatrix} \varepsilon_{11} & \varepsilon_{12} & \dots & \varepsilon_{1s} & \mathbf{B}_{1} \\ \varepsilon_{21} & \varepsilon_{22} & \dots & \varepsilon_{2s} & \mathbf{B}_{2} \\ \dots & \dots & \dots & \dots & \dots \\ \varepsilon_{s1} & \varepsilon_{2s} & \dots & \varepsilon_{ss} & \mathbf{B}_{s} \\ \varepsilon_{s+j,1} & \varepsilon_{s+j,2} & \dots & \varepsilon_{s+j,s} & \mathbf{B}_{s+j} \end{vmatrix} = 0; \quad j = 1, 2, \dots, n-s
$$
\n(19)

(b) Isostoichiometric chemical reactions. A chemical reaction that, in addition to preserving the number and type of chemical elements, also preserves the number of stoichiometric coefficients is called an isostoichiometric reaction. In other words, an isostoichiometric reaction is a balanced reaction in which the algebraic sum of stoichiometric coefficients is equal to zero. The number of linearly independent isostoichiometric reaction is equal to $m = n - \text{rank } \varepsilon - 1$. Assume, for instance, that

Then, a set of linearly independent isostoichiometric reactions may be generated following the methodology outline in [\[11\]](#page-4-0).

$$
\rho_j = \begin{vmatrix}\n\varepsilon_{11} & \varepsilon_{12} & \dots & \varepsilon_{1s} & 1 & \mathbf{B}_1 \\
\varepsilon_{21} & \varepsilon_{22} & \dots & \varepsilon_{2s} & 1 & \mathbf{B}_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\varepsilon_{s1} & \varepsilon_{s2} & \dots & \varepsilon_{ss} & 1 & \mathbf{B}_s \\
\varepsilon_{s+1,1} & \varepsilon_{s+1,2} & \dots & \varepsilon_{s+1,s} & 1 & \mathbf{B}_{s+1} \\
\varepsilon_{s+j+1,1} & \varepsilon_{s+j+1,2} & \dots & \varepsilon_{s+j+1,s} & 1 & \mathbf{B}_{s+j+1}\n\end{vmatrix} = 0;
$$
\n(20)

It may be noticed that isostoichiometric reactions were deduced from chemical thermodynamics [11].

(c) Isodesmic chemical reactions. These reactions were introduced by Hehre et al. in 1970 [\[2\]](#page-4-0) and since then have been widely used in ab initio thermochemistry studies. By definition, an isodesmic reaction preserves the type and number of bonds in chemical species. Here, the type of bond is mainly understood to be the classical covalent bond (i.e., single bond, double bond, etc.). Let the structure of species be described by a specified type, P_k $(k = 1,2,...,s)$ and π_{ki} $(k = 1,2,...,l; i = 1,2,...,n)$ represent the number of bonds between the elements. Thus, we can define the matrix

$$
\boldsymbol{\pi} = \begin{bmatrix} \pi_1 & \mathbf{P}_2 & \dots & \mathbf{P}_l \\ \pi_{21} & \pi_{12} & \dots & \pi_{1l} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{n1} & \pi_{n2} & \dots & \pi_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ \pi_{n1} & \pi_{n2} & \dots & \pi_{nl} \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_n \end{bmatrix},
$$
 (21)

that may be referred to as the bond matrix. The number of linearly independent isodesmic reactions is equal to $m = n - \text{rank } \pi = n - l$. If, for instance

$$
\begin{vmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1s} \\ \pi_{21} & \pi_{22} & \dots & \pi_{2s} \\ \dots & \dots & \dots & \dots \\ \pi_{l1} & \pi_{l3} & \dots & \pi_{ll} \end{vmatrix} \neq 0,
$$

these may be generated as outlined more explicitly in [\[12\].](#page-4-0)

$$
\rho_j = \begin{vmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1s} & \mathbf{B}_1 \\ \pi_{21} & \pi_{22} & \dots & \pi_{2s} & \mathbf{B}_2 \\ \dots & \dots & \dots & \dots & \dots \\ \pi_{l1} & \pi_{1s} & \dots & \pi_{ll} & \mathbf{B}_l \\ \pi_{l+j,1} & \pi_{l+j,2} & \dots & \pi_{l+j,s} & \mathbf{B}_{l+j} \end{vmatrix} = 0; \quad j = 1, 2, \dots, n-l.
$$
\n(22)

5. Results and discussion

The optimization procedure and the three classes of reactions described above were used to investigate the effect of reaction schemes on the accuracy of ab initio

ARTICLE IN PRESS

188 I. Fishtik et al. / Chemical Physics Letters 417 (2005) 185–189

Table 1 The errors (kcal/mol) in the enthalpy of formations for various reaction schemes at G2 and G3 levels

models. This approach was applied to a set of 47 species, listed in Table 1, that have been previously used to validate ab initio-based energetic predictions at G2 and G3 levels of theory [\[4,13\].](#page-4-0) The experimental enthalpies of formation of these species are known with high accuracy, normally, with an error not exceeding ± 0.1 kcal/mol [\[14\]](#page-4-0). Employing Eqs. (18) – (22) we first generated a set of linearly independent reactions for every class of reactions as

well as their experimental and ab initio enthalpy changes. From Table 1 it follows that the species involve four elements, namely, C, N, H and O and nine types of bonds, namely, C–H, C–C, C=C, C≡C, C–O, C=O, C–N, O– H and N–H. Thus, rank $\varepsilon = 4$ and rank $\pi = 9$. This means that conventional, isostoichiometric, and isodesmic reactions do not involve more than five, six, and ten species, respectively. The stoichiometric matrices and the enthalpy change vectors of these sets of reactions were next generated and substituted into Eq. [\(17\)](#page-1-0). The results are presented in [Table 1.](#page-3-0)

An inspection of [Table 1](#page-3-0) reveals a high level of accuracy of ab initio methods at the G2 and G3 levels of theory. It is also evident that, in accordance with earlier findings [4,13], the accuracy of the G3 method is higher than that of the G2 method. A somewhat surprising result is the weak effect of the type of reaction schemes on the accuracy of G2 and G3 methods. Thus, while the accuracy of isodesmic chemical reactions is slightly higher than the accuracy of the conventional and isostoichiometric reactions, on the whole, the difference between these schemes is not as dramatic as previously thought [4,4,13]. This conclusion is especially true for the higher level of theory, G3. In other words, at high levels of theory the accuracy is almost independent of the reaction schemes.

6. Conclusion

In the literature it is often tacitly assumed that more sophisticated types of reaction schemes can substantially improve the accuracy of theoretical thermochemistry predictions. Thus, various types of reactions have been designed aiming to minimize the errors in converting the total ab initio-based reaction enthalpies into enthalpies of formation. In this Letter, we have employed a new optimization technique to investigate the effect of reaction schemes on the accuracy of ab initio thermochemistry. In doing this we used two well-known types of reactions, namely, conventional and isodesmic reactions as well as a new type of reactions referred to as isostoichiometric. Our calculations suggest that the effect of various reaction schemes on the accuracy of ab initio methods is insignificant, especially for higher level ab initio methods.

References

- [1] L.A. Curtiss, P.C. Redfern, D.J. Frurip, Rev. Comput. Chem. 15 (2000) 147.
- [2] W.J. Hehre, R. Ditchfield, L. Radom, J.A. Pople, J. Am. Chem. Soc. 92 (1970) 4796.
- [3] W.J. Hehre, R.T. McIver, J.A. Pople, P.V.R. Schleyer, J. Am. Chem. Soc. 96 (1974) 7162.
- [4] K. Raghavachari, B.B. Stefanov, L.A. Curtiss, J. Chem. Phys. 106 (1997) 6764.
- [5] S.M. Bachrach, J. Chem. Educ. 67 (1990) 907.
- [6] I. Fishtik, R. Datta, J. Phys. Chem. A 107 (2003) 6698.
- [7] R. Sivaramakrishnan, R. Tranter, K. Brezinsky, J. Phys. Chem. A 109 (2005) 1621.
- [8] I. Fishtik, R. Datta, J. Chem. Inf. Comput. Sci. 43 (2003) 1259.
- [9] S.I. Sandler, Chemical and Engineering Thermodynamics, Wiley, New York, 1999.
- [10] G.A. Petersson, D.K. Malick, W.G. Wilson, J.W. Ochterski, J.A. Montgomery, M.J. Frisch, J. Chem. Phys. 109 (1998) 109.
- [11] I. Fishtik, I. Nagypal, I. Gutman, J. Chem. Phys. 103 (1995) 7545.
- [12] I. Fishtik, R. Datta, J.F. Liebman, J. Phys. Chem. A 107 (2003) 695.
- [13] L.A. Curtiss, K. Raghavachari, P.C. Redfern, J.A. Pople, J. Chem. Phys. 112 (2000) 7374.
- [14] J.B. Pedley, R.D. Naylor, S.P. Kirby, Thermochemical Data of Organic Compounds, second edn., Chapman and Hall, New York, 1986, p. #.