*J. Phys. Chem. A* **2003,** *107,* 6698−6707

# Group Additivity vs Ab Initio

**Ilie Fishtik*** **and Ravindra Datta**

*Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609-2280*

*Received: April 22, 2003; In Final Form: June 20, 2003*

An extension of the concept of response reactions (RERs) is presented that allows the generation of thermochemical data from group additivity (GA) and ab initio methods using precisely the same stoichiometric and mathematical procedures. This goal is achieved by (i) reformulating the ordinary least squares (OLS) analysis in terms of GA reactions, (ii) realizing that the modified OLS analysis is equally applied to both GA and ab initio methods, and (iii) proving that the errors (residuals) may be uniquely partitioned into a sum of contributions associated with GA RERs. These new findings provide a remarkable interpretation and direct comparison of the accuracies and capabilities of the GA and ab initio methods.

## 1. Introduction

With the advance of experimental techniques and computational chemistry, substantial progress has been achieved over the past 10−15 years in the development of a reliable thermodynamic database for a large set of molecules.[1] Yet, there are numerous molecular species for which reliable thermodynamic data are lacking or they are incorrect. Because experimental measurements of thermochemical data are often difficult to perform and are expensive, much emphasis is currently been put on theoretical methods. The latter may be roughly divided into two categories: empirical (group additivity (GA) and molecular mechanics (MM3), etc.) and quantum chemical (density functional theory and ab initio).[2]

Despite their empirical character, the GA methods[3−7] continue to remain a powerful and relatively accurate technique for the estimation the thermodynamic properties of the chemical species, even in the era of supercomputers. Apparently, the ab initio methods[2] are more accurate. A closer examination of the problem, however, reveals that ab initio methods face essentially the same problems as the GA methods. The point is that the conversion of the accurate ab initio total energies into conventional thermochemical data, e.g., enthalpy of formation of the species, requires the adoption of certain reaction schemes involving a set of reference species. For these purposes there have been proposed several procedures for converting the total ab initio energies into the enthalpy of formation of the species based on formation,[2] isodesmic,[8] homodesmic,[9] bond separation,[10] and group equivalent[11] reactions. The use of these reaction schemes, however, requires accurate experimental data for the reference species. In many cases these data are not available. In addition, this approach introduces specific difficulties such as the stoichiometric arbitrariness of various reaction schemes.[12] Substantial improvements in the accuracy of ab initio methods may be achieved using empirical corrections, e.g., atom, bond, and group additivity corrections.[13−16] Clearly, empirical methods also require accurate experimental thermochemical data for the reference species. In recent years, ab initio methods have been used to extend the concept of GA to transition state structures.[17−19]

Because at this stage the GA and ab initio methods are the most reliable theoretical methods for estimating the thermochemical data, it is of certain interest to compare more closely their accuracy and capabilities. Although the theoretical background of the GA may be deduced from quantum mechanical considerations,[20] a direct comparison between the GA and ab initio methods is still problematic. Our reasoning in formulating the problem in this way is the following. In our previous publications we extended the conventional theory of response reactions (RERs) to both GA[21] and ab initio[12] methods by defining new types of RERs, namely, GA and isodesmic RERs. It was shown that these new types of RERs are intimately related to the calculation of the thermochemical properties of the species from both GA and ab initio methods. More recently, the RERs formalism has been shown to be productive in a more general context, namely, in the analysis and interpretation of the quantitative structure−property relationships.[22]

In this paper, we present several new developments along this line according to which the GA and ab initio methods are unexpectedly interrelated. More specifically, we prove that the generation of thermochemical data from both methods may be performed using precisely the same stoichiometric and optimization techniques.

## 2. Conventional GA Analysis

We consider a general chemical system comprising a set of $n$ species $B_1$, $B_2$, ..., $B_n$

$$\mathbf{B} = (B_1,B_2,...,B_n)^T \tag{1}$$

Each of the species in this system is characterized by a certain *structure*. By “structure” of the species it is meant a specified type $g_j$ ($j = 1,2,...,p$) and number $g_{ij}$($i = 1,2,...,n$; $j = 1,2,...,p$) of groups in a chemical species $B_i$ ($i = 1,2,...,n$). Here the term “group” is used in the same sense as in the conventional GA methods.[3] Thus, we can define the matrix

$$\mathbf{g} = \begin{bmatrix} g_{11} & g_{12} & \dots & g_{1q} \\ g_{21} & g_{22} & \dots & g_{2q} \\ \dots & \dots & \dots & \dots \\ g_{n1} & g_{n2} & \dots & g_{nq} \end{bmatrix} \tag{2}$$

that may be referred to as the *group matrix*. It is assumed that *rank* $\mathbf{g} = q$. If *rank* $\mathbf{g} = p > q$, then the $p - q$ linearly dependent

---

* Corresponding author: ifishtik@wpi.edu.

10.1021/jp035080i CCC: $25.00

columns in the group matrix are arbitrarily disregarded. For simplicity, we consider only one property, namely, the standard ideal gas-phase enthalpy of formation of the species $B_i$ ($i = 1,2,...,n$). Define the following vectors

$$\mathbf{\Delta_f H^{exp}} = (\Delta_f H_1^{\mathrm{exp}},\Delta_f H_2^{\mathrm{exp}},...,\Delta_f H_n^{\mathrm{exp}})^{\mathrm{T}} \tag{3}$$

$$\mathbf{\Delta_f H^{GA}} = (\Delta_f H_1^{\mathrm{GA}},\Delta_f H_2^{\mathrm{GA}},...,\Delta_f H_n^{\mathrm{GA}})^{\mathrm{T}} \tag{4}$$

$$\mathbf{H^{GA}} = (H_1^{\mathrm{GA}},H_2^{\mathrm{GA}},...,H_q^{\mathrm{GA}})^{\mathrm{T}} \tag{5}$$

where $\Delta_f H_i^{\mathrm{exp}}$ is the experimental standard enthalpy of formation of the species $B_i$, $\Delta_f H_i^{\mathrm{GA}}$ is the standard enthalpy of formation of the species $B_i$ calculated via GA methods, and $H_j^{\mathrm{GA}}$ is the enthalpy group value of $g_j$ ($j = 1,2,...,q$). Define also a ($n \times 1$) vector of residuals

$$\mathbf{e^{GA}} = \mathbf{\Delta_f H^{exp}} - \mathbf{\Delta_f H^{GA}} \tag{6}$$

As is well-known,[3] the conventional GA method implies that

$$\mathbf{\Delta_f H^{GA}} = \mathbf{g H^{GA}} \tag{7}$$

Normally, the vector of group values $\mathbf{H^{GA}}$ is determined from

$$\mathbf{\Delta_f H^{exp}} = \mathbf{g H^{GA}} + \mathbf{e^{GA}} \tag{8}$$

by minimizing the product $(\mathbf{e^{GA}})^{\mathrm{T}}\mathbf{e^{GA}}$. This gives[23]

$$\mathbf{H^{GA}} = (\mathbf{g^T g})^{-\mathbf{1}}\mathbf{g^T \Delta_f H^{exp}} \tag{9}$$

It is to be noted that according to the conventional ordinary least squares (OLS) analysis, the vector of residuals $\mathbf{e^{GA}}$ may be evaluated only after $\mathbf{\Delta_f H^{GA}}$ has been calculated.

## 3. GA Reactions

A common problem in chemical stoichiometry[24] is to generate a set of $m$ linearly independent reactions $\rho_j$ ($j = 1,2,...,m$) among a given set of species $B_i$

$$\boldsymbol{\rho} = \boldsymbol{\nu}\mathbf{B} = 0 \tag{10}$$

where

$$\boldsymbol{\rho} = (\rho_1,\rho_2,...,\rho_m)^{\mathrm{T}} \tag{11}$$

and

$$\boldsymbol{\nu} = \begin{bmatrix} \nu_{11} & \nu_{12} & ... & \nu_{1n} \\ \nu_{21} & \nu_{22} & ... & \nu_{2n} \\ ... & ... & ... & ... \\ \nu_{m1} & \nu_{m2} & ... & \nu_{mn} \end{bmatrix}; \mathit{rank}\ \boldsymbol{\nu} = m \tag{12}$$

where $\boldsymbol{\nu}$ is normally referred to as *stoichiometric* matrix.[24] By analogy with the conventional stoichiometry, we define a GA reaction [21] as one that satisfies the relation

$$\boldsymbol{\nu}\mathbf{g} = 0 \tag{13}$$

or

$$\left.\begin{array}{l} \nu_{j1}g_{11} + \nu_{j2}g_{21} + ... + \nu_{jn}g_{n1} = 0 \\ \nu_{j1}g_{12} + \nu_{j2}g_{22} + ... + \nu_{jn}g_{n2} = 0 \\ ... \\ \nu_{j1}g_{1q} + \nu_{j2}g_{2q} + ... + \nu_{jn}g_{nq} = 0 \end{array}\right\}; j = 1,2,...,m \tag{14}$$

Notice, to simplify the mathematical treatment, a GA reaction is not explicitly required to satisfy the mass-balance conditions. The latter, however, are always satisfied because the groups may be partitioned into chemical elements. From the dimension of eqs 14 it is clear that the number of linearly independent GA reactions is equal to $m = n - \mathit{rank}\ \mathbf{g} = n - q$. Let, for instance

$$g = \begin{vmatrix} g_{11} & g_{12} & ... & g_{1q} \\ g_{21} & g_{22} & ... & g_{2q} \\ ... & ... & ... & ... \\ g_{q1} & g_{q2} & ... & g_{qq} \end{vmatrix} \neq 0 \tag{15}$$

Then, a set of $m$ linearly independent GA reactions $\rho_j$ ($j = 1,2,...,m$)

$$\rho_j = \nu_{j1}B_1 + \nu_{j2}B_2 + ... + \nu_{jn}B_n = 0;\ (j = 1,2,...,m) \tag{16}$$

may be generated by solving the system of linear homogeneous eqs 14. A particular solution is[21]

$$\nu_{jk} = \begin{vmatrix} g_{11} & g_{12} & ... & g_{1q} & 0 \\ g_{21} & g_{22} & ... & g_{2q} & 0 \\ ... & ... & ... & ... & ... \\ g_{k-1,1} & g_{k-1,2} & ... & g_{k-1,q} & 0 \\ g_{k1} & g_{k2} & ... & g_{kq} & 1 \\ g_{k+1,1} & g_{k+1,2} & ... & g_{k+1,q} & 0 \\ ... & ... & ... & ... & ... \\ g_{q1} & g_{q2} & ... & g_{q} & 0 \\ g_{q+j,1} & g_{q+j,2} & ... & g_{q+j,q} & 0 \end{vmatrix}; k = 1,2,...,q$$

$$\nu_{j,q+h} = \delta_{j,q+h}\begin{vmatrix} g_{11} & g_{12} & ... & g_{1q} & 0 \\ g_{21} & g_{22} & ... & g_{2q} & 0 \\ ... & ... & ... & ... & ... \\ g_{q1} & g_{q2} & ... & g_{qq} & 0 \\ g_{q+h,1} & g_{q+h,2} & ... & g_{q+h,q} & 1 \end{vmatrix}$$

$$= \delta_{j,q+h}\begin{vmatrix} g_{11} & g_{12} & ... & g_{1q} \\ g_{21} & g_{22} & ... & g_{2q} \\ ... & ... & ... & ... \\ g_{q1} & g_{q2} & ... & g_{qq} \end{vmatrix} = \delta_{j,q+h}g;\ h = 1,2,...,m \tag{17}$$

where

$$\delta_{j,q+h} = \begin{cases} 1 \text{ if } j = h \\ 0 \text{ if } j \neq h \end{cases}$$

Using the properties of the determinants, the set of $m$ linearly independent GA reactions obtained above may be presented in a more compact form as

$$\rho_j = \begin{vmatrix} g_{11} & g_{12} & ... & g_{1q} & B_1 \\ g_{21} & g_{22} & ... & g_{2q} & B_2 \\ ... & ... & ... & ... & ... \\ g_{q,1} & g_{q,2} & ... & g_{q,q} & B_q \\ g_{q+j,1} & g_{q+j,2} & ... & g_{q+j,q} & B_{q+j} \end{vmatrix} = 0;\ j = 1,2,...,m \tag{18}$$

Let

$$\mathbf{\Delta H_{\boldsymbol{\rho}}^{exp}} = (\Delta H_1^{\mathrm{exp}},\Delta H_2^{\mathrm{exp}},\ ...,\ \Delta H_m^{\mathrm{exp}})^{\mathrm{T}} \tag{19}$$

$$\mathbf{\Delta H_{\boldsymbol{\rho}}^{GA}} = (\Delta H_1^{\mathrm{GA}},\Delta H_2^{\mathrm{GA}},\ ...,\ \Delta H_m^{\mathrm{GA}})^{\mathrm{T}} \tag{20}$$

where $\Delta H_j^{\mathrm{exp}}$ ($j = 1,2,...,m$) are the experimental enthalpy changes and $\Delta H_j^{\mathrm{GA}}(j = 1,2,...,m)$ are the calculated via GA enthalpy changes of the GA reactions $\rho_j(j = 1,2,...,m)$. These vectors may be determined by employing the conventional thermodynamic formalism

$$\mathbf{\Delta H}_{\rho}{}^{\mathbf{exp}} = \boldsymbol{\nu}\mathbf{\Delta_f H^{exp}} \tag{21}$$

$$\mathbf{\Delta H}_{\rho}{}^{\mathbf{GA}} = \boldsymbol{\nu}\mathbf{\Delta_f H^{GA}} \tag{22}$$

or

$$\Delta H_j^{\mathrm{exp}} = \nu_{j1}\Delta_f H_1^{\mathrm{exp}} + \nu_{j2}\Delta_f H_2^{\mathrm{exp}} + ... + \nu_{jn}\Delta_f H_n^{\mathrm{exp}};\quad j = 1,2,...,m \tag{23}$$

$$\Delta H_j^{\mathrm{GA}} = \nu_{j1}\Delta_f H_1^{\mathrm{GA}} + \nu_{j2}\Delta_f H_2^{\mathrm{GA}} + ... + \nu_{jn}\Delta_f H_n^{\mathrm{GA}};\quad j = 1,2,...,m \tag{24}$$

Inserting here the stoichiometric coefficients of the GA reactions, eq 17, we obtain

$$\Delta H_j^{\mathrm{exp}} = \begin{vmatrix} g_{11} & g_{12} & ... & g_{1q} & \Delta_f H_1^{\mathrm{exp}} \\ g_{21} & g_{22} & ... & g_{2q} & \Delta_f H_2^{\mathrm{exp}} \\ ... & ... & ... & ... & ... \\ g_{q,1} & g_{q,2} & ... & g_{q,q} & \Delta_f H_q^{\mathrm{exp}} \\ g_{q+j,1} & g_{q+j,2} & ... & g_{q+j,q} & \Delta_f H_{q+j}^{\mathrm{exp}} \end{vmatrix};\ j = 1,2,...,m \tag{25}$$

$$\Delta H_j^{\mathrm{GA}} = \begin{vmatrix} g_{11} & g_{12} & ... & g_{1q} & \Delta_f H_1^{\mathrm{GA}} \\ g_{21} & g_{22} & ... & g_{2q} & \Delta_f H_2^{\mathrm{GA}} \\ ... & ... & ... & ... & ... \\ g_{q,1} & g_{q,2} & ... & g_{q,q} & \Delta_f H_q^{\mathrm{GA}} \\ g_{q+j,1} & g_{q+j,2} & ... & g_{q+j,q} & \Delta_f H_{q+j}^{\mathrm{GA}} \end{vmatrix};\ j = 1,2,...,m \tag{26}$$

The GA reactions defined above have a very important property. Namely,

$$\boldsymbol{\nu}\mathbf{\Delta_f H^{GA}} = \mathbf{\Delta H}_{\rho}{}^{\mathbf{GA}} = 0 \tag{27}$$

This result follows from the combination of eqs 7 and 13. Alternatively, combining eqs 6, 21, and 27 gives

$$\boldsymbol{\nu}\mathbf{e^{GA}} = \mathbf{\Delta H}_{\rho}^{\mathbf{exp}} \tag{28}$$

Similar stoichiometric relations are valid for ab initio methods. Let

$$\mathbf{\Delta H}_{\rho}{}^{\mathbf{AI}} = (\Delta H_1^{\mathrm{AI}},\Delta H_2^{\mathrm{AI}},...,\Delta H_m^{\mathrm{AI}})^{\mathrm{T}} \tag{29}$$

where $\Delta H_j^{\mathrm{AI}}(j = 1,2,...,m)$ are the enthalpy changes of the GA reactions $\rho_j(j = 1,2,...,m)$ evaluated via ab initio methods. These quantities may be evaluated using either ab initio enthalpies of formation of the species or ab initio total enthalpies of the species. Define the following vectors:

$$\mathbf{\Delta_f H^{AI}} = (\Delta_f H_1^{\mathrm{AI}},\Delta_f H_2^{\mathrm{AI}},...,\Delta_f H_n^{\mathrm{AI}})^{\mathrm{T}} \tag{30}$$

$$\mathbf{H^{AI}} = (H_1^{\mathrm{AI}},H_2^{\mathrm{AI}},...,H_n^{\mathrm{AI}})^{\mathrm{T}} \tag{31}$$

where $\Delta_f H_i^{\mathrm{AI}}$ is the standard enthalpy of formation of the species $B_i$ calculated via ab initio methods and $H_i^{\mathrm{AI}}$ is the total ab initio enthalpy of the species $B_i$. Clearly, the enthalpy changes of the GA reactions calculated either via $\mathbf{\Delta_f H^{AI}}$ or $\mathbf{H^{AI}}$ should coincide independently of the method used to evaluate $\mathbf{\Delta_f H^{AI}}$. That is

$$\mathbf{\Delta H}_{\rho}{}^{\mathbf{AI}} = \boldsymbol{\nu}\mathbf{\Delta_f H^{AI}} = \boldsymbol{\nu}\mathbf{H^{AI}} \tag{32}$$

or

$$\Delta H_j^{\mathrm{AI}} = \nu_{j1}\Delta_f H_1^{\mathrm{AI}} + \nu_{j2}\Delta_f H_2^{\mathrm{AI}} + ... + \nu_{jn}\Delta_f H_n^{\mathrm{AI}} = \nu_{j1}H_1^{\mathrm{AI}} + \nu_{j2}H_2^{\mathrm{AI}} + ... + \nu_{jn}H_n^{\mathrm{AI}}\ j = 1,2,...,m \tag{33}$$

or, using the stoichiometric coefficients of the GA reactions, eq 17

$$\Delta H_j^{\mathrm{AI}} = \begin{vmatrix} g_{11} & g_{12} & ... & g_{1q} & \Delta_f H_1^{\mathrm{AI}} \\ g_{21} & g_{22} & ... & g_{2q} & \Delta_f H_2^{\mathrm{AI}} \\ ... & ... & ... & ... & ... \\ g_{q,1} & g_{q,2} & ... & g_{q,q} & \Delta_f H_q^{\mathrm{AI}} \\ g_{q+j,1} & g_{q+j,2} & ... & g_{q+j,q} & \Delta_f H_{q+j}^{\mathrm{AI}} \end{vmatrix} = \begin{vmatrix} g_{11} & g_{12} & ... & g_{1q} & H_1^{\mathrm{AI}} \\ g_{21} & g_{22} & ... & g_{2q} & H_2^{\mathrm{AI}} \\ ... & ... & ... & ... & ... \\ g_{q,1} & g_{q,2} & ... & g_{q,q} & H_q^{\mathrm{AI}} \\ g_{q+j,1} & g_{q+j,2} & ... & g_{q+j,q} & H_{q+j}^{\mathrm{AI}} \end{vmatrix};\ j = 1,2,...,m \tag{34}$$

Define further an additional characteristic of the GA reactions, namely

$$\boldsymbol{\delta}\mathbf{\Delta H}_{\rho} = \mathbf{\Delta H}_{\rho}^{\mathbf{exp}} - \mathbf{\Delta H}_{\rho}^{\mathbf{AI}} \tag{35}$$

and a vector of residuals

$$\mathbf{e^{AI}} = \mathbf{\Delta_f H^{exp}} - \mathbf{\Delta_f H^{AI}} \tag{36}$$

Subtracting eq 32 from eq 21 we finally obtain

$$\boldsymbol{\nu}\mathbf{e^{AI}} = \boldsymbol{\delta}\mathbf{\Delta H}_{\rho} \tag{37}$$

As can be seen, eq 37 totally resembles eq 28.

## 4. An Alternative Approach to GA and Ab Initio Methods

Next, we reformulate the GA OLS analysis in terms of GA reactions. The idea is to calculate the vectors of residuals $\mathbf{e^{GA}}$ without any preliminary evaluations of the vector of group values $\mathbf{H^{GA}}$. Thus, we can evaluate directly the vector of residuals $\mathbf{e^{GA}}$ by minimizing $(\mathbf{e^{GA}})^{\mathrm{T}}\mathbf{e^{GA}}$ subject to the linear constraints given by eq 28. For this purpose we employ the method of Lagrange's undetermined multipliers and minimize the Lagrangean function

$$F = (\mathbf{e^{GA}})^{\mathrm{T}}\mathbf{e^{GA}} + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{\nu}\mathbf{e^{GA}} - \mathbf{\Delta H}_{\rho}^{\mathbf{exp}}) \tag{38}$$

with respect to $\mathbf{e^{GA}}$ and $\boldsymbol{\lambda}$. The procedure results in a system of linear equations

$$2\mathbf{e^{GA}} + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\nu} = 0 \tag{39}$$

$$\boldsymbol{\nu}\mathbf{e^{GA}} = \mathbf{\Delta H}_{\rho}^{\mathbf{exp}} \tag{40}$$

It is seen that within this approach the residuals $\mathbf{e^{GA}}$ may be directly evaluated without any need to evaluate the vector of group values! The solution of eqs 39 and 40 is[22]

$$\mathbf{e}^{\mathbf{GA}} = \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\pi}^{-1}\boldsymbol{\Delta}\mathbf{H}_{\rho}^{\mathbf{exp}} \tag{41}$$

where $\boldsymbol{\pi}$ is a square matrix of order $m$ defined as

$$\boldsymbol{\pi} = \boldsymbol{\nu}\,\boldsymbol{\nu}^{\mathrm{T}} = [\pi_{rs}]$$

$$\pi_{rs} = \pi_{sr} = \sum_{i=1}^{n}\nu_{ri}\nu_{si}$$

$$r,s = 1,2,...,m \tag{42}$$

Notice that the determinant of the matrix $\boldsymbol{\pi}$ and denoted by $\Delta$

$$\Delta = \mathrm{Det}\,\boldsymbol{\pi} = \begin{vmatrix} \pi_{11} & \pi_{12} & ... & \pi_{1m} \\ \pi_{21} & \pi_{22} & ... & \pi_{2m} \\ ... & ... & ... & ... \\ \pi_{m1} & \pi_{m2} & ... & \pi_{mm} \end{vmatrix} \tag{43}$$

is necessarily a positive value and, therefore, the inverse of $\boldsymbol{\pi}$ always exists.

Although the stoichiometric matrix $\boldsymbol{\nu}$ is generated arbitrarily, the solution for $\mathbf{e}^{\mathbf{GA}}$ is unique. That is, $\mathbf{e}^{\mathbf{GA}}$ is independent of the choice of $\boldsymbol{\nu}$. A proof of this statement is given below. We also stress that there are no mathematical advantages in the above modification of the GA OLS analysis. Moreover, it may be shown that the vector of residuals $\mathbf{e}^{\mathbf{GA}}$ given by eq 41 is *equivalent* to the vector of residuals obtained by applying the conventional OLS GA analysis, i.e., given by eqs 6−9. The importance of this approach, as shown next, lies in its ability to provide a remarkable interpretation of both GA and ab initio methods.

Because the GA analysis can be performed without generating the group values, a natural question that arises in this respect is how to estimate the properties of the species from the test set, say, species $B_{n+1}$? The enthalpy of formation of this species $\Delta_f H_{n+1}^{\mathrm{GA}}$ may be evaluated from any conceivable GA reaction involving the species $B_{n+1}$. Let an arbitrary GA reaction involving the species $B_{n+1}$ be

$$\rho = \sum_{i=1}^{n}\nu_i B_i + \nu_{n+1}B_{n+1} = 0 \tag{44}$$

Because for any GA reaction, according to eq 27, we have

$$\sum_{i=1}^{n}\nu_i\Delta_f H_i^{\mathrm{GA}} + \nu_{n+1}\Delta_f H_{n+1}^{\mathrm{GA}} = 0 \tag{45}$$

then

$$\Delta_f H_{n+1}^{\mathrm{GA}} = -\frac{1}{\nu_{n+1}}\sum_{i=1}^{n}\nu_i\Delta_f H_i^{\mathrm{GA}} \tag{46}$$

Of course, this procedure is valid provided eq 46 is independent of the choice of the GA reaction involving the species $B_{n+1}$. A proof of the independence of $\Delta_f H_{n+1}^{\mathrm{GA}}$ on the choice of the GA reaction is presented in ref 22.

Because of the mathematical similarity between eqs 28 and 37, the procedure outlined above for GA methods may be extended to ab initio methods. Thus, the vector of residuals $\mathbf{e}^{\mathbf{AI}}$ may be directly obtained by minimizing $(\mathbf{e}^{\mathbf{AI}})^{\mathrm{T}}\mathbf{e}^{\mathbf{AI}}$ subject to the linear constraints given by eq 37, resulting in

$$2\mathbf{e}^{\mathbf{AI}} + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\nu} = 0 \tag{47}$$

$$\boldsymbol{\nu}\mathbf{e}^{\mathbf{AI}} = \boldsymbol{\delta}\boldsymbol{\Delta}\mathbf{H}_{\rho} \tag{48}$$

The solution of this system of linear equations is

$$\mathbf{e}^{\mathbf{AI}} = \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\pi}^{-1}\boldsymbol{\delta}\boldsymbol{\Delta}\mathbf{H}_{\rho} \tag{49}$$

Similarly, the enthalpy of formation $\Delta_f H_{n+1}^{\mathrm{AI}}$ of the species $B_{n+1}$ from a test set may be evaluated from a stoichiometrically arbitrary GA reaction, eq 44, according to

$$\sum_{i=1}^{n}\nu_i\Delta_f H_i^{\mathrm{AI}} + \nu_{n+1}\Delta_f H_{n+1}^{\mathrm{AI}} = \Delta H_{\rho}^{\mathrm{AI}} \tag{50}$$

and, hence

$$\Delta_f H_{n+1}^{\mathrm{AI}} = \Delta H_{\rho}^{\mathrm{AI}} - \frac{1}{\nu_{n+1}}\sum_{i=1}^{n}\nu_i\Delta_f H_i^{\mathrm{AI}} \tag{51}$$

## 5. GA and Ab Initio Methods in Terms of GA RERs

According to our previous development, a GA RER[21] is defined as *a reaction that is subject to group preserving conditions and involves no more than rank* $\mathbf{g} + 1 = q + 1$ species. Let $B_{i_1},B_{i_2},...,B_{i_q},B_{i_{q+1}}(1 \le i_1 < i_2 < ... < i_q < i_{q+1} \le n)$ be the $q + 1$ species involved in a GA RER. Such an RER is denoted by $g(B_{i_1},B_{i_2},...,B_{i_q},B_{i_{q+1}})$ and its general equation is

$$g(B_{i_1},B_{i_2},...,B_{i_q},B_{i_{q+1}}) = \sum_{k=1}^{q+1}\nu_{i_k}(g)B_{i_k} = 0 \tag{52}$$

where

$$\nu_{i_k}(g) = \begin{vmatrix} g_{i_1 1} & g_{i_1 2} & ... & g_{i_1 q} & 0 \\ g_{i_2 1} & g_{i_2 2} & ... & g_{i_2 q} & 0 \\ ... & ... & ... & ... & ... \\ g_{i_{k-1} 1} & g_{i_{k-1} 2} & ... & g_{i_{k-1} q} & 0 \\ g_{i_k 1} & g_{i_k 2} & ... & g_{i_k q} & 1 \\ g_{i_{k+1} 1} & g_{i_{k+1} 2} & ... & g_{i_{k+1} q} & 0 \\ ... & ... & ... & ... & ... \\ g_{i_q 1} & g_{i_q 2} & ... & \mathrm{g}_{i_q q} & 0 \\ g_{i_{q+1} 1} & g_{i_{q+1} 2} & ... & g_{i_{q+1} q} & 0 \end{vmatrix} \tag{53}$$

A complete set of GA RERs can be generated by considering all of the possible choices of $q + 1$ species from a total of $n$.

The experimental enthalpy changes of the GA RERs denoted as $\Delta H^{\exp}(g) = \Delta H^{\exp}(B_{i_1},B_{i_2},...,B_{i_q},B_{i_{q+1}})$ are interrelated with the standard enthalpies of formation $\Delta_f H_{i_k}^{\exp}$ of the species via

$$\Delta H^{\exp}(g) = \begin{vmatrix} g_{i_1,1} & g_{i_1,2} & ... & g_{i_1,q} & \Delta_f H_{i_1}^{\exp} \\ g_{i_2,1} & g_{i_2,2} & ... & g_{i_2,q} & \Delta_f H_{i_2}^{\exp} \\ ... & ... & ... & ... & ... \\ g_{i_q,1} & g_{i_q,2} & ... & g_{i_q,q} & \Delta_f H_{i_q}^{\exp} \\ g_{i_{q+1},1} & g_{i_{q+1},2} & ... & g_{i_{q+1},q} & \Delta_f H_{i_{q+1}}^{\exp} \end{vmatrix} \tag{54}$$

Similar relations hold for the enthalpy changes of the GA RERs $\Delta H^{\mathrm{AI}}(g)$ obtained from ab initio calculations:

$$\Delta H^{\mathrm{AI}}(g) = \begin{vmatrix} g_{i_1,1} & g_{i_1,2} & \ldots & g_{i_1,q} & H^{\mathrm{AI}}_{i_1} \\ g_{i_2,1} & g_{i_2,2} & \ldots & g_{i_2,q} & H^{\mathrm{AI}}_{i_2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ g_{i_q,1} & g_{i_q,2} & \ldots & g_{i_q,q} & H^{\mathrm{AI}}_{i_q} \\ g_{i_{q+1},1} & g_{i_{q+1},2} & \ldots & g_{i_{q+1},q} & H^{\mathrm{AI}}_{i_{q+1}} \end{vmatrix} \qquad (55)$$

We also introduce the quantity

$$\delta\Delta H(g) = \Delta H^{\exp}(g) - \Delta H^{AI}(g) \qquad (56)$$

We are now in a position to formulate the following results:

$$e_i^{\mathrm{GA}} = \frac{1}{\Delta}\sum_g \nu_i(g)\Delta H^{\exp}(g);\ i = 1,2,...,n \qquad (57)$$

$$e_i^{\mathrm{AI}} = \frac{1}{\Delta}\sum_g \nu_i(g)\delta\Delta H(g);\ i = 1,2,...,n \qquad (58)$$

where

$$\Delta = \frac{1}{m}\sum_g f(g) \qquad (59)$$

$$f(g) = \sum_{i=1}^{n} \nu_i^{\,2}(g) \qquad (60)$$

The proof of these results follows exactly the same lines as those considered in ref 22, and therefore, is not given here.

As can be seen from eqs 57 and 58, the residuals of the species obtained from both GA and ab initio methods may be uniquely partitioned into a sum of contributions associated with GA RERs. It may be remembered that a GA RER involves no more than $q + 1$ species where $q$ is the number of groups. Consequently, the partition of the residuals into contributions associated with GA RERs is in fact a partition into contributions coming from all possible subsets of $q + 1$ species from a total of $n$. Each of these contributions has a very simple form. Thus, the contribution of every GA RER is given by a product of two terms. One of them is the stoichiometric coefficient $\nu_i(g)$ of the species $B_i$ while the other, is the change of the experimental enthalpy change $\Delta H^{\exp}(g)$ or the difference between the experimental and ab initio enthalpy change $\delta\Delta H(g) = \Delta H^{\exp}(g) - \Delta H^{\mathrm{AI}}(g)$ in a particular GA RER. Notice that the stoichiometric coefficients of the GA RERs are solely functions of the number and type of groups, i.e., structure, while the enthalpy changes of RERs are functions of both the structure and thermochemistry. Because the GA RERs are stoichiometrically unique, the independence of eqs 41 and 49 of the choice of the stoichiometric matrix $\boldsymbol{\nu}$ becomes obvious.

Eqs 57 and 58 are also a powerful tool for the rationalization, comprehension, and interpretation of the GA and ab initio calculations. A complete list of GA RERs, i.e., their stoichiometric coefficients $\nu_i(g)$, along with the changes in their enthalpies $\Delta H^{\exp}(g)$ and $\delta\Delta H(g)$, provides detailed information about the structure−thermochemistry relationships. For instance, with a complete list of GA RERs, one can easily determine the GA RERs that have the smallest or highest contributions to the residuals, or, equivalently, a subset of species whose structure is highly or poorly correlated with the thermochemistry. In particular, the GA RERs approach may be of real use in detecting the outliers.

## 6. Examples

We consider several applications of the described above approach. To save space, we illustrate first the technical details of the algorithm with the help of a simple example comprising only five species. The capabilities of the algorithm are next briefly illustrated by applying it to two more complex systems.

**TABLE 1: Experimental Enthalpies of Formation (kJ/mol) and Total Enthalpies (hartree) at the G3 Level of Fluoro- and Chloromethanes**

| | | | $\Delta_f H_i^{\exp\,a}$ | $H_i^{\mathrm{G3}\,b}$ |
|---|---|---|---|---|
| 1 | $B_1$ | $CH_4$ | −74.6 | −40.453813 |
| 2 | $B_2$ | $CF_4$ | −933.2 | −437.302900 |
| 3 | $B_3$ | $CCl_4$ | −96.0 | −1,878.276568 |
| 4 | $B_4$ | $CH_3F$ | −234.3 | −139.645784 |
| 5 | $B_5$ | $CH_2F_2$ | −450.7 | −238.858175 |
| 6 | $B_6$ | $CH_3Cl$ | −83.7 | −499.909048 |
| 7 | $B_7$ | $CH_2Cl_2$ | −95.5 | −959.366688 |
| 8 | $B_8$ | $CF_3Cl$ | −709.2 | −797.540914 |
| 9 | $B_9$ | $CF_2Cl_2$ | −491.6 | −1,157.782200 |
| 10 | $B_{10}$ | $CHF_3$ | −692.9 | −338.082135 |
| 11 | $B_{11}$ | $CHCl_3$ | −103.2 | −1,418.823436 |
| 12 | $B_{12}$ | $CFCl_3$ | −288.7 | −1,518.027193 |
| 13 | $B_{13}$ | $CH_2FCl$ | −261.9 | −599.109188 |
| 14 | $B_{14}$ | $CHF_2Cl$ | −481.6 | −698.323412 |
| 15 | $B_{15}$ | $CHFCl_2$ | −283.3 | −1058.570575 |

[a] Reference 25. [b] Reference 26.

**6.1. Chloromethanes.** Consider the complete set of chloromethanes $CH_xCl_{4-x}$ where $x = 0, 1, 2, 3, 4$. The experimental enthalpies of formation[25] along with the total enthalpies obtained at the G3 level[26] are presented in Table 1. For this system there are three possible GA approximations depending on how the groups are defined. We consider all of them in an increasing order of accuracy.

*6.1.1. First Approximation.* In a first approximation, we consider only two types of groups, namely, $g_1 = \mathrm{C{-}H}$ and $g_2 = \mathrm{C{-}Cl}$. The group matrix **g** in this case is

$$\mathbf{g} = \begin{array}{c} \begin{array}{cc} g_1 & g_2 \end{array} \\ \begin{bmatrix} 4 & 0 \\ 3 & 1 \\ 2 & 2 \\ 1 & 3 \\ 0 & 4 \end{bmatrix} \end{array} \begin{array}{l} \\ CH_4 \\ CH_3Cl \\ CH_2Cl_2 \\ CHCl_3 \\ CCl_4 \end{array}$$

First, we derive a set of linearly independent GA reactions. Since *rank* $\mathbf{g} = 2$, the number of linearly independent GA reactions is equal to 3. According to eqs 15−17 an appropriate set of GA reactions may be generated as

$$\rho_1 = \begin{vmatrix} 4 & 0 & CH_4 \\ 3 & 1 & CH_3Cl \\ 2 & 2 & CH_2Cl_2 \end{vmatrix} = 4CH_4 - 8CH_3Cl + 4CH_2Cl_2 = 0$$

$$\rho_2 = \begin{vmatrix} 4 & 0 & CH_4 \\ 3 & 1 & CH_3Cl \\ 1 & 3 & CHCl_3 \end{vmatrix} = -8CH_4 + 12CH_3Cl - 4CHCl_3 = 0$$

$$\rho_3 = \begin{vmatrix} 4 & 0 & CH_4 \\ 3 & 1 & CH_3Cl \\ 0 & 4 & CCl_4 \end{vmatrix} = -12CH_4 + 16CH_3Cl - 4CCl_4 = 0$$

or, dividing all of the reactions by 4,

$$\rho_1 = -CH_4 + 2CH_3Cl - CH_2Cl_2 = 0$$

$$\rho_2 = -2CH_4 + 3CH_3Cl - CHCl_3 = 0$$

$$\rho_3 = -3CH_4 + 4CH_3Cl - CCl_4 = 0$$

Hence, the stoichiometric matrix is

$$\boldsymbol{\nu} = \begin{array}{c} \begin{array}{ccccc} CH_4 & CH_3Cl & CH_2Cl_2 & CHCl_3 & CCl_4 \end{array} \\ \begin{bmatrix} -1 & 2 & -1 & 0 & 0 \\ -2 & 3 & 0 & -1 & 0 \\ -3 & 4 & 0 & 0 & -1 \end{bmatrix} \end{array}$$

Now, the matrix $\boldsymbol{\pi}$, eq 42, and its inverse can be evaluated

$$\boldsymbol{\pi} = \boldsymbol{\nu}\,\boldsymbol{\nu}^{\mathrm{T}} = \begin{bmatrix} 6 & 8 & 11 \\ 8 & 14 & 18 \\ 11 & 18 & 26 \end{bmatrix}$$

$$\boldsymbol{\pi}^{-1} = \begin{bmatrix} 4/5 & -1/5 & -1/5 \\ -1/5 & 7/10 & -2/5 \\ -1/5 & -2/5 & 2/5 \end{bmatrix}$$

The experimental enthalpy changes of the above linearly independent GA reactions are

$$\Delta H_1^{\mathrm{exp}} = \frac{1}{4}\begin{vmatrix} 4 & 0 & -74.6 \\ 3 & 1 & -83.7 \\ 2 & 2 & -95.5 \end{vmatrix} = 2.7\ \mathrm{kJ/mol}$$

$$\Delta H_2^{\mathrm{exp}} = \frac{1}{4}\begin{vmatrix} 4 & 0 & -74.6 \\ 3 & 1 & -83.7 \\ 1 & 3 & -103.2 \end{vmatrix} = 1.3\ \mathrm{kJ/mol}$$

$$\Delta H_3^{\mathrm{exp}} = \frac{1}{4}\begin{vmatrix} 4 & 0 & -74.6 \\ 3 & 1 & -83.7 \\ 0 & 4 & -96.0 \end{vmatrix} = -15.0\ \mathrm{kJ/mol}$$

Similarly, the ab initio enthalpy changes at the G3 level are

$$\Delta H_1^{\mathrm{G3}} = \frac{2625.5}{4}\begin{vmatrix} 4 & 0 & -40.453813 \\ 3 & 1 & -499.909048 \\ 2 & 2 & -959.366688 \end{vmatrix} = 6.31\ \mathrm{kJ/mol}$$

$$\Delta H_2^{\mathrm{G3}} = \frac{2625.5}{4}\begin{vmatrix} 4 & 0 & -40.453813 \\ 3 & 1 & -499.909048 \\ 1 & 3 & -1418.823436 \end{vmatrix} = 10.29\ \mathrm{kJ/mol}$$

$$\Delta H_3^{\mathrm{G3}} = \frac{2625.5}{4}\begin{vmatrix} 4 & 0 & -40.453813 \\ 3 & 1 & -499.909048 \\ 0 & 4 & -1878.276568 \end{vmatrix} = 4.77\ \mathrm{kJ/mol}$$

and

$$\delta\Delta H_1 = \Delta H_1^{\mathrm{exp}} - \Delta \mathrm{H}_1^{\mathrm{G3}} = 2.7 - 6.31 = -3.61\ \mathrm{kJ/mol}$$

$$\delta\Delta H_2 = \Delta H_2^{\mathrm{exp}} - \Delta H_2^{\mathrm{G3}} = 2.3 - 10.29 = -8.99\ \mathrm{kJ/mol}$$

$$\delta\Delta H_3 = \Delta H_3^{\mathrm{exp}} - \Delta H_3^{\mathrm{G3}} = -15.0 - 4.77 = -19.77\ \mathrm{kJ/mol}$$

Thus, according to eq 41, the residuals for the GA method are equal to

$$\mathbf{e}^{\mathbf{GA}} = \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\pi}^{-1}\boldsymbol{\Delta}\mathbf{H}_{\rho}^{\mathbf{exp}} = \begin{bmatrix} -1 & -2 & -3 \\ 2 & 3 & 4 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} 4/5 & -1/5 & -1/5 \\ -1/5 & 7/10 & -2/5 \\ -1/5 & -2/5 & 2/5 \end{bmatrix}\begin{bmatrix} 2.7 \\ 1.3 \\ -15.0 \end{bmatrix} = (3.54, 0.67, -4.90, -6.37, 7.06)^{\mathrm{T}}$$

Similarly, according to eq 49, the residuals of the species at G3 level are

$$\mathbf{e}^{\mathbf{AI}} = \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{\pi}^{-1}\boldsymbol{\delta}\boldsymbol{\Delta}\mathbf{H}_{\rho} = \begin{bmatrix} -1 & -2 & -3 \\ 2 & 3 & 4 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} 4/5 & -1/5 & -1/5 \\ -1/5 & 7/10 & -2/5 \\ -1/5 & -2/5 & 2/5 \end{bmatrix}\begin{bmatrix} -3.61 \\ -8.99 \\ -19.77 \end{bmatrix} = (3.23, -1.62, -2.86, -2.34, 3.59)^{\mathrm{T}}$$

As can be seen, within the first approximation the accuracy of the G3 method is higher than the accuracy of GA method. For instance, the mean absolute deviation of the residuals is equal to 4.51 kJ/mol for GA method while the same quantity for G3 is equal to 2.73 kJ/mol.

*6.1.2. Second Approximation.* At a higher level of approximation we consider three types of groups, namely, $g_3 = CH_2$, $g_4 = CHCl$, and $g_5 = CCl_2$. The group matrix in this case is

$$\mathbf{g} = \begin{array}{cl} \begin{array}{ccc} g_3 & g_4 & g_5 \end{array} & \\ \begin{bmatrix} 6 & 0 & 0 \\ 3 & 3 & 0 \\ 1 & 4 & 1 \\ 0 & 3 & 3 \\ 0 & 0 & 6 \end{bmatrix} & \begin{array}{l} CH_4 \\ CH_3Cl \\ CH_2Cl_2 \\ CHCl_3 \\ CCl_4 \end{array} \end{array}$$

For illustration purposes we analyze this system using the GA RERs approach. Since *rank* $\boldsymbol{g} = 3$, a GA RER involves no more than $3 + 1 = 4$ species. Hence, the total number of GA RERs is equal to the number of ways four species may be selected from a total of five, i.e., is equal to 5. For example, the first four species define the following GA RER:

$$g(CH_4, CH_3Cl, CH_2Cl_2, CHCl_3) = \begin{vmatrix} 6 & 0 & 0 & CH_4 \\ 3 & 3 & 0 & CH_3Cl \\ 1 & 4 & 1 & CH_2Cl_2 \\ 0 & 3 & 3 & CHCl_3 \end{vmatrix} = -18CH_4 + 54CH_3Cl - 54CH_2Cl_2 + 18CHCl_3 = 0$$

or, in a more conventional form

$$g(CH_4, \mathrm{CH3Cl}, CH_2Cl_2, CHCl_3) = -CH_4 + 3CH_3Cl - 3CH_2Cl_2 + CHCl_3 = 0$$

A complete list of GA RERs along with their experimental and G3 enthalpy changes are given in Table 2. Based on these data, the residuals of the species may be partitioned into a sum of contributions associated with GA RERs according to eqs 57–60 (Table 3). The overall residuals may be obtained by summing over the GA RERs. Again, within the second approximation the accuracy of the G3 method exceeds that of GA method. For instance, the mean absolute deviation is 0.31 kJ/mol for G3 vs 2.18 kJ/mol for GA method.

*6.1.2. Third Approximation.* The third and the highest possible approximation in this system is the following selection of groups: $g_6 = CH_3$, $g_7 = CH_2Cl$, $g_8 = CHCl_2$, and $g_9 = CCl_3$. This selection results in the group matrix

**TABLE 2: Complete List of GA RERs and Their Enthalpy Changes (kJ/mol) for Second Approximation in Example 1**

| | $\Delta H^{\mathrm{exp}}(g)$ | $\Delta H^{\mathrm{G3}}(g)$ |
|---|---|---|
| 1. $g(CH_4,CH_3Cl,CH_2Cl_2,CHCl_3) = -CH_4+3CH_3Cl\text{-}3CH_2Cl_2+CHCl_3 = 0$ | 6.8 | 8.66 |
| 2. $g(CH_4,CH_3Cl,CH_2Cl_2,CCl_4) = -3CH_4+8CH_3Cl\text{-}6CH_2Cl_2+CCl_4 = 0$ | 31.2 | 33.12 |
| 3. $g(CH_4,CH_3Cl,CHCl_3,CCl_4) = -3CH_4+6CH_3Cl\text{-}6CHCl_3+3CCl_4 = 0$ | 52.8 | 47.42 |
| 4. $g(CH_4,CH_3Cl_2,CHCl_3,CCl_4) = -CH_4+6CH_2Cl_2\text{-}8CHCl_3+3CCl_4 = 0$ | 39.2 | 30.11 |
| 5. $g(CH_3Cl,CH_3Cl_2,CHCl_3,CCl_4) = -CH_3Cl+3CH_2Cl_2\text{-}3CHCl_3+CCl_4 = 0$ | 10.8 | 7.15 |

$$\mathbf{g} = \begin{array}{c} \\ \\ \\ \\ \\ \end{array}\begin{array}{cccc|l} g_6 & g_7 & g_8 & g_9 & \\ 4 & 0 & 0 & 0 & CH_4 \\ 1 & 3 & 0 & 0 & CH_3Cl \\ 0 & 2 & 2 & 0 & CH_2Cl_2 \\ 0 & 0 & 3 & 1 & CHCl_3 \\ 0 & 0 & 0 & 4 & CCl_4 \end{array}$$

The rank of the group matrix is equal to 4 and, consequently, the system is described only by one GA reaction. Obviously, this reaction is also a GA RER and its equation is

$$\rho = g(CH_4, CH_3Cl, CH_2Cl_2, CHCl_3)$$

$$= \begin{vmatrix} 4 & 0 & 0 & 0 & CH_4 \\ 1 & 3 & 0 & 0 & CH_3Cl \\ 0 & 2 & 2 & 0 & CH_2Cl_2 \\ 0 & 0 & 3 & 1 & CHCl_3 \\ 0 & 0 & 0 & 4 & CCl_4 \end{vmatrix}$$

$$= 24CH_4 - 96CH_3Cl + 144CH_2Cl_2 - 96CHCl_3 + 24CCl_4 = 0$$

or

$$\rho = CH_4 - 4CH_3Cl + 6CH_2Cl_2 - 4CHCl_3 + CCl_4 = 0$$

The experimental and G3 enthalpy changes of this GA reaction are $\Delta H^{\mathrm{exp}} = 4.0$ kJ/mol and $\Delta H^{G3} = -1.5$ kJ/mol. Respectively, $\delta\Delta H = \Delta H^{\mathrm{exp}} - \Delta H^{G3} = 5.5$ kJ/mol. From the above general results it can be deduced that for systems with one GA reaction the residuals are given by

$$e_i^{\mathrm{GA}} = \frac{\nu_i \Delta H^{\mathrm{exp}}}{\sum_{k=1}^{n} \nu_k^{\,2}};\ i = 1,2,...,n$$

$$e_i^{\mathrm{G3}} = \frac{\nu_i \delta\Delta H}{\sum_{k=1}^{n} \nu_k^{\,2}};\ i = 1,2,...,n$$

From these equations it follows immediately that because the difference $\delta\Delta H = \Delta H^{\mathrm{exp}} - \Delta H^{\mathrm{G3}}$ slightly exceeds $\Delta H^{\mathrm{exp}}$, the accuracy of the GA method within this approximation is (slightly) higher than the accuracy of the G3 method. The numerical values of the vector of residuals are

$$\mathbf{e}^{\mathbf{GA}} = (0.06, -0.23, 0.34, -0.23, 0.06)^{\mathrm{T}}$$

$$\mathbf{e}^{\mathbf{G3}} = (0.08, -0.31, 0.47, -0.31, 0.08)^{\mathrm{T}}$$

**TABLE 3: Partition of the Residuals into Contributions Coming from GA RERs (Second Approximation in Example 1)**

| GA RERs | $e(CH_4)$ | $e(CH_3Cl)$ | $e(CH_2Cl_2)$ | $e(CHCl_3)$ | $e(CCl_4)$ |
|---|---|---|---|---|---|
| | | | GA Method | | |
| 1 | −0.04 | 0.12 | −0.12 | 0.04 | 0 |
| 2 | −0.53 | 1.43 | −1.07 | 0.00 | 0.18 |
| 3 | −0.91 | 1.81 | 0.00 | −1.81 | 0.91 |
| 4 | −0.22 | 0.00 | 1.34 | −1.79 | 0.67 |
| 5 | 0.00 | −0.06 | 0.19 | −0.19 | 0.06 |
| overall | −1.70 | 3.29 | 0.34 | −3.75 | 1.82 |
| | | | G3 Method | | |
| 1 | 0.01 | −0.03 | 0.03 | −0.01 | 0.00 |
| 2 | 0.03 | −0.09 | 0.07 | 0.00 | −0.01 |
| 3 | −0.09 | 0.18 | 0.00 | −0.18 | 0.09 |
| 4 | −0.05 | 0.00 | 0.31 | −0.42 | 0.16 |
| 5 | 0.00 | −0.02 | 0.06 | −0.06 | 0.02 |
| overall | −0.10 | 0.04 | 0.47 | −0.67 | 0.26 |

**TABLE 4: GA vs G3 for Fluoro- and Chloromethanes at Different Levels of GA Approximations**

| | 1st | | 2nd | | 3rd | |
|---|---|---|---|---|---|---|
| | $e_i^{\mathrm{GA}}$ | $e_i^{\mathrm{G3}}$ | $e_i^{\mathrm{GA}}$ | $e_i^{\mathrm{G3}}$ | $e_i^{\mathrm{GA}}$ | $e_i^{\mathrm{G3}}$ |
| $B_1$ | −19.86 | 2.46 | −7.59 | 0.03 | 0.10 | −0.08 |
| $B_2$ | −30.72 | −2.90 | 5.82 | −0.10 | 0.03 | −0.26 |
| $B_3$ | 4.40 | 2.52 | 1.19 | 0.41 | 0.25 | 0.34 |
| $B_4$ | 32.38 | 2.77 | 11.61 | 0.26 | −0.02 | 0.78 |
| $B_5$ | 27.91 | 0.00 | 0.18 | −1.67 | −0.29 | −1.57 |
| $B_6$ | −17.54 | −2.46 | 1.44 | −0.17 | −0.37 | −0.48 |
| $B_7$ | −17.93 | −3.78 | 0.72 | 0.15 | 0.24 | 0.32 |
| $B_8$ | −7.24 | −1.03 | 0.73 | −1.60 | −0.55 | −0.44 |
| $B_9$ | 9.84 | 3.03 | 1.66 | 1.11 | 1.18 | 1.27 |
| $B_{10}$ | −2.35 | 1.80 | −10.99 | 1.90 | 0.42 | 1.46 |
| $B_{11}$ | −14.21 | −3.33 | −2.96 | −0.87 | −0.18 | −0.36 |
| $B_{12}$ | 12.22 | 2.16 | 0.31 | −0.46 | −0.81 | −0.99 |
| $B_{13}$ | 16.19 | 0.79 | 5.44 | 1.36 | 0.63 | 0.80 |
| $B_{14}$ | 8.43 | −0.47 | −6.00 | −0.72 | −0.70 | −1.24 |
| $B_{15}$ | 6.21 | −0.28 | −1.60 | 0.26 | 0.07 | 0.43 |
| MAD | 15.16 | 1.99 | 3.88 | 0.74 | 0.39 | 0.72 |

**6.2. Fluoro- and Chloromethanes.** The experimental enthalpies of formation[25] along with the total enthalpies at the G3 level[26] for all of the 15 species in this system are presented in Table 1. We have considered three levels of GA approximations that are summarized in Appendix A. The final results for all three approximations are given in Table 4. It is seen that the accuracy of the G3 method slowly increases as we move to higher levels of GA approximations. At the same time, the accuracy of the GA method strongly depends on the GA approximation. As expected, an increase in the complexity of groups results in a higher accuracy of the GA method. Interestingly, at the highest possible level of approximation, the accuracy of the GA methods exceeds the accuracy of the G3 method.

**6.3. Chloroethanes.** The experimental enthalpies of formation,[27] total enthalpies of the species at the G2 level,[28] as well as the definition of group and a brief summary of the analysis using the algorithm discussed above are presented in Appendix B and Tables 5 and 6. Again, it is seen that the accuracy of the GA method significantly depends on the definition of groups. That is, for structurally simple groups the accuracy is very low. As we move to more complex groups, however, the accuracy of the GA method rapidly increases. On the other hand, the accuracy of G2 method is less dependent on the type of GA reactions, although an increase in the complexity of groups also

**TABLE 5: Experimental Enthalpies of Formation (kJ/mol) and Total Enthalpies (hartree) at G2 Level of Substituted Chloroethanes**

| | | $\Delta_f H_i^{\mathrm{exp}\ a}$ | $H_i^{\mathrm{G2}\ b}$ |
|---|---|---|---|
| $B_1$ | $CH_3CH_3$ | −84.0 ± 0.4 | −79.626396 |
| $B_2$ | $CH_3CH_2Cl$ | −112.1 ± 0.7 | −538.777329 |
| $B_3$ | $CH_3CHCl_2$ | −132.5 ± 3.5 | −997.927129 |
| $B_4$ | $CH_3CCl_3$ | −144.6 ± 2.0 | −1457.073227 |
| $B_5$ | $CH_2ClCH_2Cl$ | −132.0 ± 3.5 | −997.926212 |
| $B_6$ | $CH_2ClCHCl_2$ | −148.0 ± 4.0 | −1457.073162 |
| $B_7$ | $CH_2ClCCl_3$ | −152.3 ± 2.4 | −1916.217376 |
| $B_8$ | $CHCl_2CHCl_2$ | −156.7 ± 3.5 | −1916.217902 |
| $B_9$ | $CHCl_2CCl_3$ | −155.9 ± 4.3 | −2375.360623 |
| $B_{10}$ | $CCl_3CCl_3$ | −148.2 ± 5.7 | −2834.501485 |

[a] Reference 27. [b] Reference 28.

increases the accuracy of the G2 method. Finally, the accuracy of the GA methods at high levels of approximation exceeds that of the G2 method.

## 7. Discussion and Concluding Remarks

The key elements of this work can be summarized as follows. The conventional formalism provided by the theory of RERs can be naturally extended so as to include additional constraints. Thus, one can define and generate a stoichiometrically unique and finite set of RERs that additionally preserves the number and types of groups. This new type of RERs, referred to as the GA RERs, has several important properties. First, the thermochemical characteristics of the GA RERs are related to the accuracy of the GA and ab initio methods. In particular, assuming that the main assumptions of the GA methods are exactly valid should result in GA RERs that have the remarkable property of being thermoneutral. Second, the GA RERs are intimately related to the conventional OLS analysis. Thus, the residuals of the thermochemical properties of the species obtained from both GA and ab initio methods may be naturally partitioned into a linear sum of contributions associated with GA RERs. Third, the GA RERs provide an unusual interrelationship between the GA and ab initio methods. Namely, the generation of thermochemical data in both methods may be done using exactly the same mathematical and stoichiometric procedures. As a result, the performance and accuracy of these methods may be directly compared.

The GA RERs approach is also a powerful interpretative concept. That is, the partitioning of the residuals into contributions coming from GA RERs provides detailed "insider" information that is unavailable within the conventional OLS GA analysis. Clearly, this information may be used to get a deeper understanding of the mathematically complex and, often, musky interrelations between the structure and thermodynamic properties of the species and, ultimately, to improve the accuracy of the GA and ab initio methods. One of the outputs of the GA RERs analysis is a subset of dominant GA RERs. Thus, one may determine a small subset of species that are mutually poorly correlated and, hence, are responsible for a low performance of the GA and ab initio methods. These species are not always outliers and, hence, their deletion may not improve the accuracy. Rather, one should concentrate on these species in order to determine the particularities of their structure that caused the poor correlation and/or make an effort to find other reference species or groups that may improve their mutual correlation.

It may be noted also that the newly defined GA RERs may be fruitfully applied to solve a wide range of other problems in physical organic chemistry. Examples include the stabilization/destabilization effects in cyclic molecules and aromaticity. Work along this line is in progress.

**Acknowledgment.** We are grateful to Prof. G. Bacskay for providing the total G3 enthalpies of fluoro- and chloromethanes.

## Appendix A. GA Approximations for Fluoro- and Chloromethanes

**1st Approximation.** For groups $g_1$ = CH, $g_2$ = CF, and $g_3$ = CCl, the group matrix (*rank* $\mathbf{g}$ = 3) is

$$\mathbf{g} = \begin{array}{cccc} g_1 & g_2 & g_3 & \\ 4 & 0 & 0 & B_1 \\ 0 & 4 & 0 & B_2 \\ 0 & 0 & 4 & B_3 \\ 3 & 1 & 0 & B_4 \\ 2 & 2 & 0 & B_5 \\ 3 & 0 & 1 & B_6 \\ 2 & 0 & 2 & B_7 \\ 0 & 3 & 1 & B_8 \\ 0 & 2 & 2 & B_9 \\ 1 & 3 & 0 & B_{10} \\ 1 & 0 & 3 & B_{11} \\ 0 & 1 & 3 & B_{12} \\ 2 & 1 & 1 & B_{13} \\ 1 & 2 & 1 & B_{14} \\ 1 & 1 & 2 & B_{15} \end{array}$$

The linearly independent GA reactions are as follows:

| | kJ/mol | |
|---|---|---|
| | $\Delta H_\rho^{\mathrm{exp}}$ | $\Delta H_\rho^{\mathrm{G3}}$ |
| $\rho_1 = -3B_1 - B_2 + 0B_3 + 4B_4 = 0$ | 219.8 | 213.20 |
| $\rho_2 = -2B_1 - 2B_2 + 0B_3 + 4B_5 = 0$ | 212.8 | 211.95 |
| $\rho_3 = -3B_1 + 0B_2 - B_3 + 4B_6 = 0$ | −15.0 | 4.77 |
| $\rho_4 = -2B_1 + 0B_2 - 2B_3 + 4B_7 = 0$ | −40.8 | −15.73 |
| $\rho_5 = 0B_1 - 3B_2 - B_3 + 4B_8 = 0$ | 58.8 | 56.74 |
| $\rho_6 = 0B_1 - 2B_2 - 2B_3 + 4B_9 = 0$ | 92.0 | 79.12 |
| $\rho_7 = -B_1 - 3B_2 + 0B_3 + 4B_{10} = 0$ | 102.6 | 89.20 |
| $\rho_8 = -B_1 + 0B_2 - 3B_3 + 4B_{11} = 0$ | −50.2 | −26.85 |
| $\rho_9 = 0B_1 - B_2 - 3B_3 + 4B_{12} = 0$ | 66.4 | 62.57 |
| $\rho_{10} = -2B_1 - B_2 - B_3 + 4B_{13} = 0$ | 130.8 | 132.17 |
| $\rho_{11} = -B_1 - 2B_2 - B_3 + 4B_{14} = 0$ | 110.6 | 111.67 |
| $\rho_{12} = -B1 - B_2 - 2B_3 + 4B_{15} = 0$ | 66.6 | 72.33 |

**2nd Approximation.** For groups $g_4$ = $CH_2$, $g_5$ = CHF, $g_6$ = CHCl, $g_7$ = $CF_2$, $g_8$ = CFCl, and $g_9$ = $CCl_2$, the group matrix (*rank* $\mathbf{g}$ = 6) is

$$\mathbf{g} = \begin{array}{ccccccc} g_4 & g_5 & g_6 & g_7 & g_8 & g_9 & \\ 6 & 0 & 0 & 0 & 0 & 0 & B_1 \\ 0 & 0 & 0 & 6 & 0 & 0 & B_2 \\ 0 & 0 & 0 & 0 & 0 & 6 & B_3 \\ 3 & 3 & 0 & 0 & 0 & 0 & B_4 \\ 1 & 4 & 0 & 1 & 0 & 0 & B_5 \\ 3 & 0 & 3 & 0 & 0 & 0 & B_6 \\ 1 & 0 & 4 & 0 & 0 & 1 & B_7 \\ 0 & 0 & 0 & 3 & 3 & 0 & B_8 \\ 0 & 0 & 0 & 1 & 4 & 1 & B_9 \\ 0 & 3 & 0 & 3 & 0 & 0 & B_{10} \\ 0 & 0 & 3 & 0 & 0 & 3 & B_{11} \\ 0 & 0 & 0 & 0 & 3 & 3 & B_{12} \\ 1 & 2 & 2 & 0 & 1 & 0 & B_{13} \\ 0 & 2 & 1 & 1 & 2 & 0 & B_{14} \\ 0 & 1 & 2 & 0 & 2 & 1 & B_{15} \end{array}$$

**TABLE 6: GA vs G3 for Chloroethanes at Different Levels of GA Approximations**

| | 1st | | 2nd | | 3rd | | 4th | | 5th | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $e_i^{GA}$ | $e_i^{G2}$ | $e_i^{GA}$ | $e_i^{G2}$ | $e_i^{GA}$ | $e_i^{G2}$ | $e_i^{GA}$ | $e_i^{G2}$ | $e_i^{G}$ | $Ae_i^{G2}$ |
| $B_1$ | 20.16 | 3.96 | 10.18 | 0.69 | 10.65 | 0.91 | 0.09 | 0.19 | 0.03 | 0.04 |
| $B_2$ | 2.88 | −0.19 | 2.88 | −0.19 | 2.41 | −0.41 | −0.40 | −0.84 | −0.21 | −0.23 |
| $B_3$ | −6.69 | 0.76 | −6.69 | 0.76 | −5.75 | 1.21 | 0.28 | 0.55 | 0.21 | 0.23 |
| $B_4$ | −7.97 | 0.28 | −17.95 | −2.62 | −17.95 | −2.63 | −0.06 | −0.08 | −0.07 | −0.08 |
| $B_5$ | −6.19 | −1.15 | 3.79 | 1.76 | 2.38 | 1.08 | 0.45 | 0.80 | 0.31 | 0.34 |
| $B_6$ | −11.37 | −3.29 | −1.39 | −0.38 | −1.39 | −0.38 | −0.61 | −0.69 | −0.62 | −0.68 |
| $B_7$ | −4.85 | −0.91 | −4.85 | −0.92 | −5.79 | −1.36 | 0.10 | −0.08 | 0.21 | 0.23 |
| $B_8$ | −9.25 | −3.93 | 0.74 | −1.02 | 2.15 | −0.34 | 0.18 | −0.13 | 0.31 | 0.34 |
| $B_9$ | 2.38 | −0.37 | 2.38 | −0.37 | 2.85 | −0.15 | −0.04 | 0.40 | −0.21 | −0.23 |
| $B_{10}$ | 20.90 | 5.21 | 10.92 | 2.30 | 10.45 | 2.07 | 0.00 | −0.12 | 0.03 | 0.04 |
| MAD | 9.26 | 2.01 | 6.18 | 1.10 | 6.18 | 1.05 | 0.22 | 0.39 | 0.22 | 0.24 |

The linearly independent GA reactions are as follows:

| | kJ/mol | |
|---|---|---|
| | $\Delta H_\rho^{exp}$ | $\Delta H_\rho^{G3}$ |
| $\rho_1 = 3B_1 - B_2 + 0B_3 \text{-} 8B_4 + 0B_6 + 0B_8 + 6B_5 = 0$ | −20.4 | −108.48 |
| $\rho_2 = 3B_1 + 0B_2 - B_3 + 0B_4 \text{-} 8B_6 + 0B_8 + 6B_7 = 0$ | −31.2 | −33.12 |
| $\rho_3 = 0B_1 + 3B_2 - B_3 + 0B_4 + 0B_6 \text{-} 8B_8 + 6B_9 = 0$ | 20.4 | 5.20 |
| $\rho_4 = 3B_1 \text{-} 3B_2 + 0B_3 \text{-} 6B_4 + 0B_6 + 0B_8 + 6B_{10} = 0$ | −175.8 | −186.00 |
| $\rho_5 = 3B_1 + 0B_2 \text{-} 3B_3 + 0B_4 \text{-} 6B_6 + 0B_8 + 6B_{11} = 0$ | −52.8 | −47.42 |
| $\rho_6 = 0B_1 + 3B_2 \text{-} 3B_3 + 0B_4 + 0B_6 \text{-} 6B_8 + 6B_{12} = 0$ | 11.4 | 8.74 |
| $\rho_7 = 3B_1 + B_2 + 0B_3 \text{-} 4B_4 \text{-} 4B_6 \text{-} 2B_8 + 6B_{13} = 0$ | −38.0 | −48.08 |
| $\rho_8 = 3B_1 + B_2 + 0B_3 \text{-} 4B_4 \text{-} 2B_6 \text{-} 4B_8 + 6B_{14} = 0$ | −105.2 | −104.82 |
| $\rho_9 = 3B_1 + 2B_2 - B_3 \text{-} 2B_4 \text{-} 4B_6 \text{-} 4B_8 + 6B_{15} = 0$ | −53.8 | −59.61 |

**3rd Approximation.** For groups $g_{10} = CH_3$, $g_{11} = CH_2F$, $g_{12} = CH_2Cl$, $g_{13} = CHF_2$, $g_{14} = CHFCl$, $g_{15} = CHCl_2$, $g_{16} = CF_3$, $g_{17} = CF_2Cl$, $g_{18} = CFCl_2$, and $g_{19} = CCl_3$, the group matrix (*rank* $\mathbf{g} = 10$) is

$$\mathbf{g} = \begin{array}{c} \\ \left[\begin{array}{cccccccccc} g_{10} & g_{11} & g_{12} & g_{13} & g_{14} & g_{15} & g_{16} & g_{17} & g_{18} & g_{19} \\ 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \\ 1 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 1 & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 1 & 0 \end{array}\right] \end{array} \begin{array}{l} \\ B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \\ B_6 \\ B_7 \\ B_8 \\ B_9 \\ B_{10} \\ B_{11} \\ B_{12} \\ B_{13} \\ B_{14} \\ B_{15} \end{array}$$

The linearly independent GA reactions are as follows:

| | kJ/mol | |
|---|---|---|
| | $\Delta H_\rho^{exp}$ | $\Delta H_\rho^{G3}$ |
| $\rho_1 = -3B_1 - 3B_2 + 0B_3 + 12B_4 - 18B_5 + 0B_6 + 0B_7 + 0B_8 + 0B_9 + 0B_{13} + 12B_{10} = 0$ | 9.6 | −46.57 |
| $\rho_2 = -3B_1 + 0B_2 - 3B_3 + 0B_4 + 0B_5 + 12B_6 - 18B_7 + 0B_8 + 0B_9 + 0B_{13} + 12B_{11} = 0$ | −12.0 | 4.51 |
| $\rho_3 = 0B_1 - 3B_2 \text{-} 3B_3 + 0B_4 + 0B_5 + 0B_6 + 0B_7 + 12B_8 - 18B_9 + 0B_{13} + 12B_{12} = 0$ | −38.4 | 1.89 |
| $\rho_4 = \text{-}3B_1 + B_2 + 0B_3 + 8B_4 - 6B_5 + 4B_6 + 0B_7 - 4B_8 + 0B_9 - 12B_{13} + 12B_{14} = 0$ | −14.0 | −5.01 |
| $\rho_5 = -3B_1 - B_2 + 0B_3 + 4B_4 + 0B_5 + 8B_6 - 6B_7 + 4B_8 - 6B_9 - 12B_{13} + 12B_{15} = 0$ | −20.8 | −4.85 |

## Appendix B. GA Approximations for Chloroethanes

**1st Approximation.** For groups $g_1 = CH$ and $g_2 = CCl$, the group matrix (*rank* $\mathbf{g} = 2$) is

$$\mathbf{g} = \begin{array}{c} \\ \left[\begin{array}{cc} g_1 & g_2 \\ 6 & 0 \\ 5 & 1 \\ 4 & 2 \\ 3 & 3 \\ 4 & 2 \\ 3 & 3 \\ 2 & 4 \\ 2 & 4 \\ 1 & 5 \\ 0 & 6 \end{array}\right] \end{array} \begin{array}{l} \\ B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \\ B_6 \\ B_7 \\ B_8 \\ B_9 \\ B_{10} \end{array}$$

The linearly independent GA reactions are as follows:

| | kJ/mol | |
|---|---|---|
| | $\Delta H_\rho^{exp}$ | $\Delta H_\rho^{G2}$ |
| $\rho_1 = B_1 - 2B_2 + B_3 = 0$ | 7.7 | 2.97 |
| $\rho_2 = 2B_1 - 3B_2 + B_4 = 0$ | 23.7 | 15.67 |
| $\rho_3 = B_1 - 2B_2 + B_5 = 0$ | 8.2 | 5.38 |
| $\rho_4 = 2B_1 - 3B_2 + B_6 = 0$ | 20.3 | 15.84 |
| $\rho_5 = 3B_1 - 4B_2 + B_7 = 0$ | 44.1 | 33.48 |
| $\rho_6 = 3B_1 - 4B_2 + B_8 = 0$ | 39.7 | 32.10 |
| $\rho_7 = 4B_1 - 5B_2 + B_9 = 0$ | 68.6 | 53.66 |
| $\rho_8 = 5B_1 - 6B_2 + B_{10} = 0$ | 104.4 | 80.10 |

**2nd Approximation.** For groups $g_3 = CH_2$, $g_4 = CHCl$, and $g_5 = CCl_2$, the group matrix (*rank* $\mathbf{g} = 3$) is

$$\mathbf{g} = \begin{array}{c} \\ \left[\begin{array}{ccc} g_1 & g_2 & g_3 \\ 6 & 0 & 0 \\ 4 & 2 & 0 \\ 3 & 2 & 1 \\ 3 & 0 & 3 \\ 2 & 4 & 0 \\ 1 & 4 & 1 \\ 1 & 2 & 3 \\ 0 & 4 & 2 \\ 0 & 2 & 4 \\ 0 & 0 & 6 \end{array}\right] \end{array} \begin{array}{l} \\ B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \\ B_6 \\ B_7 \\ B_8 \\ B_9 \\ B_{10} \end{array}$$

The linearly independent GA RERs are as follows:

| | kJ/mol | |
|---|---|---|
| | $\Delta H_\rho^{\mathrm{exp}}$ | $\Delta H_\rho^{\mathrm{G2}}$ |
| $\rho_1 = -B_1+3B_2-3B_3+B_4 = 0$ | 0.6 | 6.74 |
| $\rho_2 = B_1-2B_2+0B_3+B_5 = 0$ | 8.2 | 5.38 |
| $\rho_3 = B_1-B_2-B_3+B_6 = 0$ | 12.6 | 12.87 |
| $\rho_4 = 0B_1+2B_2-3B_3+B_7 = 0$ | 21.0 | 24.56 |
| $\rho_5 = B_1+0B_2-2B_3+B_8 = 0$ | 24.3 | 26.15 |
| $\rho_6 = 0B_1+3B_2-4B_3+B_9 = 0$ | 37.8 | 41.76 |
| $\rho_7 = -B_1+6B_2-6B_3+B_{10} = 0$ | 58.2 | 62.25 |

**3rd Approximation.** For groups $g_6 = CH_3$, $g_7 = CH_2Cl$, $g_8 = CHCl_2$, and $g_9 = CCl_3$, the group matrix (*rank* **g** = 4) is

$$\mathbf{g} = \begin{array}{cccc|c} g_6 & g_7 & g_8 & g_9 & \\ 2 & 0 & 0 & 0 & B_1 \\ 1 & 1 & 0 & 0 & B_2 \\ 1 & 0 & 1 & 0 & B_3 \\ 1 & 0 & 0 & 1 & B_4 \\ 0 & 2 & 0 & 0 & B_5 \\ 0 & 1 & 1 & 0 & B_6 \\ 0 & 1 & 0 & 1 & B_7 \\ 0 & 0 & 2 & 0 & B_8 \\ 0 & 0 & 1 & 1 & B_9 \\ 0 & 0 & 0 & 2 & B_{10} \end{array}$$

The linearly independent GA RERs are as follows:

| | kJ/mol | |
|---|---|---|
| | $\Delta H_\rho^{\mathrm{exp}}$ | $\Delta H_\rho^{\mathrm{G2}}$ |
| $\rho_1 = B_1-2B_2+0B_3+0B_4+B_5 = 0$ | 8.2 | 5.38 |
| $\rho_2 = B_1-B_2-B_3+0B_4+B_6 = 0$ | 12.6 | 12.87 |
| $\rho_3 = B_1-B_2+0B_3-B_4+B_7 = 0$ | 20.4 | 17.81 |
| $\rho_4 = B_1+0B_2-2B_3+0B_4+B_8 = 0$ | 24.3 | 26.15 |
| $\rho_5 = B_1+0B_2-B_3-B_4+B_9 = 0$ | 37.2 | 35.02 |
| $\rho_6 = B_1+0B_2+0B_3\text{-}2B_4+B_{10} = 0$ | 57.0 | 48.76 |

**4th Approximation.** For groups $g_{10} = HCCH_3$, $g_{11} = ClCCH_3$, $g_{12} = HCCH_2Cl$, $g_{13} = ClCCH_2Cl$, $g_{14} = HCCHCl_2$, $g_{15} = ClCCHCl_2$, $g_{16} = HCCCl_3$, $g_{17} = ClCCCl_3$, the group matrix (*rank* **g** = 7) is

$$\mathbf{g} = \begin{array}{cccccccc|c} g_{10} & g_{11} & g_{12} & g_{13} & g_{14} & g_{15} & g_{16} & g_{17} & \\ 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & B_1 \\ 2 & 1 & 3 & 0 & 0 & 0 & 0 & 0 & B_2 \\ 1 & 2 & 0 & 0 & 3 & 0 & 0 & 0 & B_3 \\ 0 & 3 & 0 & 0 & 0 & 0 & 3 & 0 & B_4 \\ 0 & 0 & 4 & 2 & 0 & 0 & 0 & 0 & B_5 \\ 0 & 0 & 1 & 2 & 2 & 1 & 0 & 0 & B_6 \\ 0 & 0 & 0 & 3 & 0 & 0 & 2 & 1 & B_7 \\ 0 & 0 & 0 & 0 & 2 & 4 & 0 & 0 & B_8 \\ 0 & 0 & 0 & 0 & 0 & 3 & 1 & 2 & B_9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & B_{10} \end{array}$$

The linearly independent GA reactions are as follows:

| | kJ/mol | |
|---|---|---|
| | $\Delta H_\rho^{\mathrm{exp}}$ | $\Delta H_\rho^{\mathrm{G2}}$ |
| $\rho_1 = B_1-4B_2+2B_3+0B_4+4B_5-4B_6+0B_7+B_8 = 0$ | 6.7 | −3.78 |
| $\rho_2 = 2B_1-7B_2+2B_3+B_4+6B_5-3B_6-2B_7+B_9 = 0$ | 7.8 | −6.91 |
| $\rho_3 = 4B_1-12B_2+0B_3+4B_4+9B_5+0B_6-6B_7+B_{10} = 0$ | 8.4 | −9.66 |

**5th Approximation.** For groups $g_{18} = H_2CCH_3$, $g_{19} = HClCCH_3$, $g_{20} = Cl_2CCH_3$, $g_{21} = H_2CCH_2Cl$, $g_{22} = HClCCH_2Cl$, $g_{23} = Cl_2CCH_2Cl$, $g_{24} = H_2CCHCl_2$, $g_{25} = HClCCHCl_2$, $g_{26} = Cl_2CCHCl_2$, $g_{27} = H_2CCCl_3$, $g_{28} = HClCCCl_3$, and $g_{29} = Cl_2CCCl_3$, the group matrix (*rank* **g** = 9) is

$$\mathbf{g} = \begin{array}{cccccccccccc|c} g_{18} & g_{19} & g_{20} & g_{21} & g_{22} & g_{23} & g_{24} & g_{25} & g_{26} & g_{27} & g_{28} & g_{29} & \\ 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & B_1 \\ 1 & 2 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & B_2 \\ 0 & 2 & 1 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & B_3 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & B_4 \\ 0 & 0 & 0 & 2 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & B_5 \\ 0 & 0 & 0 & 0 & 2 & 1 & 1 & 2 & 0 & 0 & 0 & 0 & B_6 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 1 & 2 & 0 & B_7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 2 & 0 & 0 & 0 & B_8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 2 & 1 & B_9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & B_{10} \end{array}$$

The linearly independent GA RERs are as follows:

| | kJ/mol | |
|---|---|---|
| | $\Delta H_\rho^{\mathrm{exp}}$ | $\Delta H_\rho^{\mathrm{G2}}$ |
| $\rho_1 = B_1-6B_2+6B_3-2B_4+9B_5-18B_6+6B_7+9B_8-6B_9+B_{10} = 0$ | 21.9 | −2.24 |

## References and Notes

(1) http://webbook.nist.gov/chemistry.

(2) Curtiss, L. A.; Redfern, P. C.; Frurip, D. J. *Rev. Comput. Chem.* **2000**, *15*, 147.

(3) Benson, S. W.; Buss, J. H. *J. Chem. Phys.* **1958**, *29*, 546.

(4) Benson, S. W. *Thermochemical Kinetics*; Wiley-Interscience: New York, 1976.

(5) Tatevskii, V. M. *The Structure of Molecules*; Khimiya: Leningrad, 1977.

(6) Cohen, N.; Benson, S. W. *Chem. Rev.* **1993**, *93*, 2419.

(7) Benson, S. W.; Cruickshank, F. R.; Golden, D. M.; Haugen, G. R.; O'Neal, H. E.; Rodgers, A. S.; Shaw, R.; Walsh, R. *Chem. Rev.* **1969**, *69*, 269.

(8) Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. *J. Am. Chem. Soc.* **1970**, *92*, 4796.

(9) Hehre, W. J.; McIver, R. T.; Pople, J. A.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1974**, *96*, 7162.

(10) Raghavachari, K.; Stefanov, B, B.; Curtiss, L. A. J. Chem. Phys. **1997**, *106*, 6764.

(11) Bachrach, S. M. *J. Chem. Educ.* **1990**, *67*, 907.

(12) Fishtik, I.; Datta, R.; Liebamn, J. F. *J. Phys. Chem. A* **2003**, *107*, 695.

(13) Petersson, G. A.; Malick, D. K.; Wilson, W. G.; Ochterski J. W.; Montgomery, J. A.; Frisch, M. J. *J. Chem. Phys.* **1998**, *109*, 10570.

(14) Melius, C. F.; Allendorf, M. D. *J. Phys. Chem. A* **2000**, *104*, 2168.

(15) Guthrie, J. P. *J. Phys. Chem. A* **2001**, *105*, 9196.

(16) Irikura, K. K. *J. Phys. Chem. A* **2002**, *106*, 9910.

(17) Sumathi, R.; Carstensen, H.-H.; Green, W. H., Jr. *J. Phys. Chem. A* **2001**, *105*, 6910.

(18) Sumathi, R.; Carstensen, H.-H.; Green, W. H. Jr. *J. Phys. Chem. A* **2001**, *105*, 8969.

(19) Sumathi, R.; Carstensen, H.-H.; Green, W. H., Jr. *J. Phys. Chem. A* **2002**, *106*, 5474.

(20) Bader, R. F. W.; Bayles, D. *J. Phys. Chem. A* **2000**, *104*, 5579.

(21) Fishtik, I.; Datta, R.; Liebamn, J. F. *J. Phys. Chem. A* **2003**, *107*, 2334.

(22) Fishtik, I.; Datta, R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1259.

(23) Draper, N. R.; Smith, H. *Applied Regression Analysis*; John Wiley & Sons: New York, 1998.

(24) Smith, W. R.; Missen, R. W. *Chemical Reaction Equilibrium Analysis: Theory and Algorithms*; John Wiley & Sons: New York, 1982.

(25) Novak, I. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 358.

(26) Bacskay, G., private communication.

(27) Manion, J. A. *J. Phys, Chem. Ref. Data* **2002**, *31*, 123.

(28) Cioslowski, J.; Liu, G.; Moncrieff, D. *J. Am. Chem. Soc.* **1997**, *119*, 11452.