JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Formation Enthalpies Derived from Pairwise Interactions: A Step toward More Transferable Reactive Potentials for Organic Compounds

Didier Mathieu

CEA, DAM, Le Ripault, F-37260 Monts, France

Supporting Information

ABSTRACT: A new approach to the development and parametrization of reactive potentials for organic compounds is put forward. As a byproduct of preliminary efforts in this direction, the performance of a simple representation of the energy of equilibrium structures in term of pairwise atom−atom and bond−bond contributions is investigated. For now, each contribution is assumed constant, given the multiplicity of covalent bonds, rather than computed on-the-fly from geometries and bond orders. In spite of this rough approximation, the approach performs remarkably well by comparison with semiempirical quantum chemical methods. Nevertheless, further refinement proves necessary for some unstable species involved in chemical reactions. As it stands, the present model appears as a promising basis in view of less empirical and more versatile alternatives to group contribution methods for the fast prediction of heats of formation, although much work remains to be done to demonstrate its value as a starting point toward better reactive potentials.

1. INTRODUCTION

Progress in the field of molecular simulation depends on the availability of reliable and versatile interatomic potentials eliminating the need to derive forces between atoms from quantum mechanics. A large variety of such potentials have been put forward to describe different kinds of bonding situations, including ionic crystals,[1] metals,[2−4] or simple covalent materials.[5,6] Interatomic forces are especially complex in the case of molecular systems, such as organic compounds made of light elements including H, C, N and O, owing to the directionality of covalent bonds, the occurrence of delocalization effects in conjugated systems, the long-range Coulomb interactions arising from interatomic charge transfer in polar compounds, and the need to monitor the changes affecting the charge distribution upon chemical reactions. In fact, most potentials (or force fields) presently available for molecular systems do not allow for reactivity as they assume a fixed topology for the bond network. The so-called reactive molecular dynamics model introduced in 2003 overcomes this limitation through an extension of classical molecular mechanics force fields.[7−9] This is achieved through the use of modified analytic expressions parametrized on a system specific basis for nonequilibrium configurations along selected reaction pathways.

More transferable reactive potentials are difficult to develop as they must satisfy stringent conditions. First, they must be consistent with all regions of interest on the potential energy surface (PES). Ensuring this condition requires in principle an extensive sampling of the PES along its many dimensions. Second, spurious low-energy minima and saddle points must not be accessible, in order to avoid unrealistic configurations to be obtained. This condition is all the more difficult to ensure as the expression for the potential is complex. Another major obstacle stems from the tedious parametrization process inherent to the large number of adjustable parameters.

In this context, reactive potentials for organic systems are mostly restricted to C and/or H atoms.[10−14] A notable exception is the so-called ReaxFF reactive force field, parametrized for a significant number of elements and thus applicable to a wide range of systems, including not only hydrocarbons[15] and carbon nanotubes[16] but also silicium,[17] polymers,[18] explosives,[19−21] amino acids in water[22] or metals, oxides, and hydrides.[23−25] Nevertheless, current parametrizations of this potential suffer from a lack of transferability.[26] This is illustrated by various artifacts, such as spurious minima observed on the PES of nitric acid[27] or an excessive amount of N_2 dimers and undercoordinated Si atoms in β-Si_3N_4.[28] For the time being, such problems are circumvented through an ad hoc reparametrization against ab initio data relevant for the specific system under study.[21,22] On the other hand, the large number of disposable parameters employed in reactive force fields makes it difficult to assess the relative merits of various formulations. Deviations from experiment may be attributed either to deficiencies inherent to the formalism employed or to a failure to identify suitable parameters.

Reactive potentials are typically fitted against ab initio PESs. The only exception concerns the charge model which may be fitted against atomic charges derived from electronic wave functions, as done for instance in the ReaxFF parametrization.[29] Splitting the total energy into contributions that can be fitted independently is clearly advantageous, as it leads to a reduction of the number of parameters that have to be fitted simultaneously. This strategy is used in modern intermolecular potentials resolved into separate contributions for core repulsion, dispersion, polarization, or hydrogen bonding. However, it might be difficult to apply to reactive potentials in the lack of a suitable decomposition scheme.

Received: August 29, 2011
Published: March 6, 2012

dx.doi.org/10.1021/ct2006083 | J. Chem. Theory Comput. 2012, 8, 1295−1303

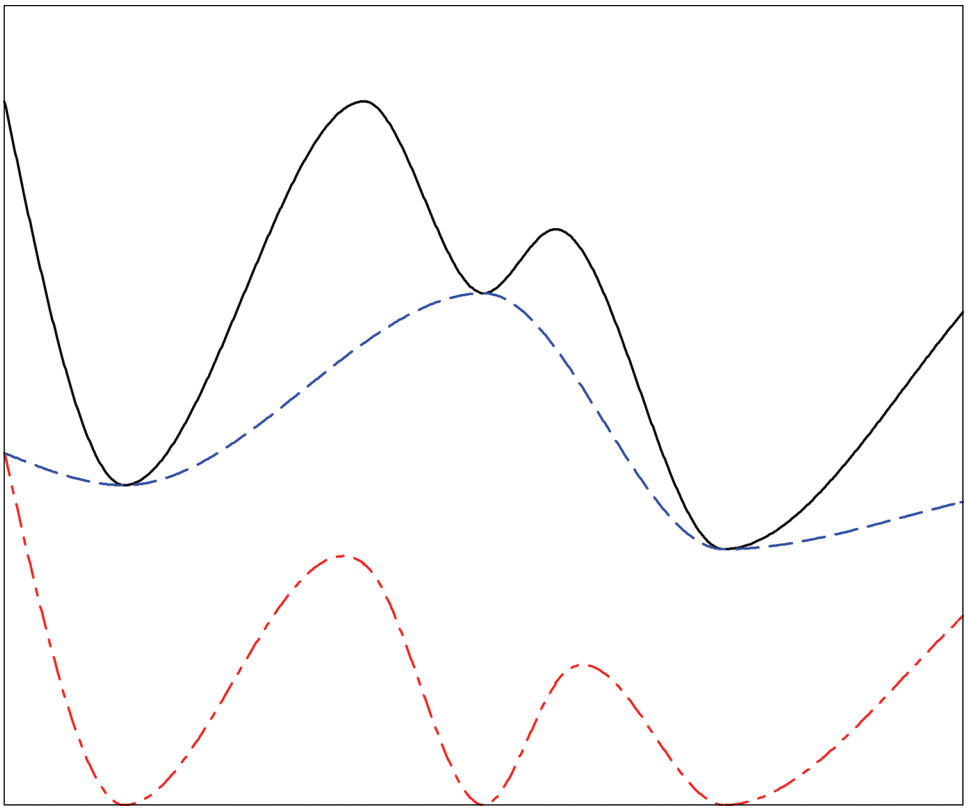

Figure 1. Parametrization strategy presently considered. The total potential energy (black solid line) is split into two contributions. The first contribution E_0 (blue dashed line) is obtained from the interpolation of molecular energies calculated using simple schemes, similar to present group contribution or molecular mechanics approaches to $\Delta_f H^0$. The second contribution E_1 (red dashed−dotted line) is the difference between the total potential and the first contribution.

Another approach to reduce the empirical character of reactive potentials consists in introducing explicit electronic variables, in addition to the atomic coordinates. This does not introduce any significant computational overhead as long as they can be easily derived from atomic coordinates, like bond orders commonly employed in reactive potentials, or treated as dynamic variables and propagated together with atom positions, as for the partial atomic charges describing Coulomb interactions in ReaxFF. Similarly, supplementary electronic variables might be introduced into the short-range component of reactive potentials, either in terms of hybrid orbitals[30] or bond order elements.[26]

In this work, it is suggested that the potential might be split into two contributions explicitly dependent on electronic variables, as detailed in Section 2. As a first step into this direction, we then report in Section 3 a preliminary investigation of the performance of a simple model describing the energy of equilibrium structures in term of atom−atom and bond order dependent electron−electron interactions. Finally, the performance of this preliminary model, compared to well-established approaches, is discussed in Section 4.

2. PRESENT APPROACH

2.1. Derivation. The performance of the ReaxFF reactive potential when it comes to the prediction of molecular formation enthalpies $\Delta_f H^0$ is disappointing, compared with that of much simpler additivity schemes. This is clearly a consequence of ReaxFF aiming at providing a fair representation of whole regions of the PES, in contrast to additivity schemes restricted to equilibrium structures. This suggests that although the energies of nonequilibrium structures are clearly not easy to describe in term of simple analytic expressions, this might be the case for equilibrium structures. Therefore, it could be fruitful to split the total energy into a simple contribution E_0 correctly describing the energies of equilibrium geometries and a more complex contribution E_1 describing the energy cost associated with distorted structures.

In addition to the advantages inherent to splitting the potential into independent contributions, further attractive features of this strategy, illustrated in Figure 1, include: (1) the possibility to ensure that the PES does not exhibit spurious low-energy regions due to the relative simplicity of E_0; and (2) the fact that E_1 is consistent with the positive energy contributions used in current potentials and associated with bond elongation, angle bending, or other geometry distortions, whose values are zero for undistorted equilibrium structures.

While the success of additivity schemes for $\Delta_f H^0$ supports the possibility that E_0 may be well represented by a simple expression, it must not be oversimplified. In particular, a simple atom−atom pairwise potential (or central force field) is not suitable for molecular systems, as demonstrated by its obvious inability to account for strain in small rings, such as cyclopropane. Accordingly, present force fields for molecules rely on a cluster expansion of the total potential energy E as a sum of multibody terms:

$$E = \sum_i E_i + \sum_{i,j} E_{i,j} + \sum_{i,j,k} E_{i-j-k} + \sum_{i,j,k,l} E_{i-j-k-l} \quad (1)$$

dx.doi.org/10.1021/ct2006083 | J. Chem. Theory Comput. 2012, 8, 1295−1303

Table 1. Bond Parameters h_{i-j} and Corresponding Numbers of Occurrence (N_{occ}) in Parentheses for Single, Aromatic, Double, and Triple Bonds[a]

$i-j$	h_{i-j}	N_{occ}	$i-j$	h_{i-j}	N_{occ}	$i-j$	h_{i-j}	N_{occ}	$i-j$	h_{i-j}	N_{occ}
H−H	104.2	(1)									
C−H	128.5	(602)									
N−H	112.2	(59)									
O−H	119.4	(88)									
C−C	155.9	(536)	C∼C	159.6	(93)	C═C	153.3	(132)	C≡C	179.4	(31)
C−N	145.5	(69)	C∼N	148.9	(35)	C═N	149.5	(6)	C≡N	200.5	(13)
C−O	153.0	(180)	C∼O	147.6	(9)	C═O	188.1	(151)	C≡O	257.3	(1)
N−N	110.4	(7)	N∼N	119.7	(5)	N═N	126.7	(9)	N≡N	225.9	(5)
N−O	95.5	(7)	N∼O	114.5	(4)	N═O	134.7	(13)			
O−O	79.5	(13)	O∼O	85.3	(1)						

[a]Bond parameters in kcal/mol. Single, aromatic, double, and triples bonds, H−H, O∼O, C═O, and C≡O, respectively, are associated with unique bond types, encountered respectively in dihydrogen, ozone, and carbon monoxide. In contrast, as mentioned in the text, the N≡N parameter introduced for dinitrogen is also used for the terminal $N^+{=}N^-$ bond in azide groups.

where the sums run over individual atoms and atom pairs, triples, and quadruples, respectively. Unfortunately, the need to define potential energy terms for every type of atom triples and quadruples that may be encountered inevitably leads to very large numbers of disposable parameters and therefore to significant parametrization issues. This problem might be solved by sticking as much as possible to a pairwise picture and introducing a minimum number of multibody terms as required to obtain satisfactory results.

Electronic variables, such as dynamic atomic charges or bond orders, afford the possibility to introduce some multibody effects without the need to introduce complex explicit functions of numerous atom positions. In this work, bond orders are used to define electron-rich regions at positions r_a, r_b, ... and with corresponding electron populations a, b, ... In addition to atom−atom interactions, this approach considers electron−electron interactions between the electron charges located at r_a, r_b, ... As a result, eq 1 is replaced by the following expression:

$$E_0 = \sum_i E_i + \sum_{i,j} E_{ij} + \sum_{a,b} E_{ab} \qquad (2)$$

where the energy terms E_{ij} and E_{ab} depend only on the atom−atom and electron−electron distances, denoted, respectively, as r_{ij} and r_{ab}. It should be kept in mind that E_0 coincides with the total potential only for equilibrium geometries and that another contribution E_1 must be added for nonequilibrium configurations.

A major challenge in the development of reactive potentials stems from the fact that the PES may exhibit spurious low-energy pockets that limit transferability. Potentials involving superposition of numerous and complex analytical energy terms are especially prone to this problem, as illustrated by low-energy artifacts observed on the ReaxFF PES of nitric acid.[27] The present approach provides no definite solution to this challenging problem. Nevertheless, the present emphasis on pairwise interactions should make in easier to get some insight into the conditions associated with the occurrence of artificial low-energy basins.

2.2. Preliminary Implementation. A literal implementation of the above-mentioned approach involves setting up a database of compounds with corresponding values of total energies E available from either high-level ab initio computations or experiment, computing equilibrium geometries and wave functions, localizing the electron densities in order to obtain the values a, b, ... and positions r_a,r_b, ... of the charge populations involved in eq 2, fitting a charge equilibration model against these charge populations, and finally, an analytic expression depending on atom positions, bond orders, and atomic charges against the reference energies E.

This will clearly require a significant work. Therefore, before embarking on this long-term project, a simplified model which does not account for three-dimensional (3D) molecular geometries is investigated here. In the lack of atomic coordinates, only short-range interactions between atoms maintained at specific distances may be taken into account. Atom−atom interactions are considered not only for bonded atom pairs but also for geminal atom pairs. With regard to bond−bond interactions, only those between adjacent bonds in strained m-membered rings with $m < 6$ are considered, since such bonds are somewhat closer to each other than in usual molecular environments.

It should be emphasized that every atom is characterized only by its atomic symbol. In other words, no distinct types are introduced to account for its environment, in line with future developments toward reactive potentials. With regard to covalent bonds, fractional bond orders cannot be calculated in the lack of a 3D geometry. Therefore, bond multiplicities (i.e., formal bond orders) are used instead. In the context of reactive simulations, it should be relatively easy to replace them with dynamic bond orders derived from a charge equilibration method somewhat similar to the atom-bond electronegativity equalization method.[31]

Standard values are introduced for the interaction between two atoms X and Y, depending on their being in either geminal positions (X···Y) or bonded by a single (X−Y), aromatic (X∼Y), double (X═Y), or triple (X≡Y) bond. As indicated in Table 1, the terminal $N^+{=}N^-$ bond in azide groups is assimilated to a triple N≡N bond (as in N_2) in order to account for the attraction between the opposite formal charges on the adjacent nitrogen atoms. Finally, only discrete values of atom−atom potentials for five distances specific to the X,Y pair under consideration are needed, rather than the full pairwise potential function. This is clearly a rough approximation. In particular, the distance between two geminal atoms X and Y depends on the hybridization of the central atom, being especially large for sp hybridization, as the atoms lie in opposite directions with respect to their common neighbor. To account for this fact, additional correction terms denoted X···Y(sp) are

introduced in this case to make up for the otherwise overestimated X···Y interaction.

The interaction between two adjacent bonds in a strained m-membered ring is denoted after the multiplicity of the bonds and the value of m. For instance, (−5∼) stands for the interaction between a single bond in a 5-membered ring, adjacent to an aromatic bond in the same ring. Two such interactions are present for instance in indane. As a further example, (−3═) denotes the interaction between a double bond and a single bond within a 3-membered ring, as encountered in diazirine and cyclopropene derivatives.

The main energy contributions are the negative bond enthalpies $-h_{i-j}$, which are obviously binding (negative) contributions. In contrast, geminal X···Y interactions are positive as bonds maintain atoms within van der Waals distances. For instance, the C−C distance between geminal carbon atoms in propane is d_{CC} = 2.5 Å, while the corresponding van der Waals distance is d_w = 3.4 Å. Similarly, bond−bond interactions are expected to be positive, as they describe the increased repulsion between adjacent bonds made closer to each other owing to ring constraints. Finally, the sp corrections must clearly be negative to be consistent with their interpretation. The present model thus relies on four types of energy contributions, two positive and two negative, according to

$$E_0 = \sum_{i\cdots j} E_{i\cdots j} + \sum_{a\vee b} E_{a\vee b} - \sum_{i-j} h_{i-j} - \sum_{i\cdots j/\mathrm{sp}} E_{i\cdots j/\mathrm{sp}} \quad (3)$$

where the four summations run respectively other pairs $i\cdots j$ of geminal atoms, adjacent bonds $a\vee b$ in strained rings, covalent bonds $i-j$, and geminal atoms bracketing a central atom with sp hybridization, denoted $i\cdots j/\mathrm{sp}$. This scheme is hereafter referred to by the GR-BH acronym, which keeps track of the signs of the four contributions to the energy and uses letters reflecting their physical origin, namely geminal interactions (G), ring strains (R), bonds (B), and hybridization (H).

It should be kept in mind that in view of the many approximations introduced and the lack of an explicit dependence of the energy on 3D geometries, this preliminary implementation of the present approach is only a small step on the way to a full reactive potential. In particular, although its pairwise character should make its extension to 3D equilibrium structures rather straightforward, much work will be needed to definitely establish its value as a starting point toward modeling out-of-equilibrium geometries.

2.3. Parametrization. The aim of this paper is to assess the viability of the above-mentioned approach. For this purpose, it is shown that it is consistent with experimental formation enthalpies $\Delta_f H^0$. In other words, we show that reasonable estimates of $\Delta_f H^0$ may be obtained as:

$$\Delta_f H^0 = \sum_i \Delta_f H^0(i) + E_0 \quad (4)$$

where $\Delta_f H^0(i)$ is the formation enthalpy of atom i in the gas phase, presently taken from the NIST Chemistry Web book,[32] and E_0 is the energy obtained using eq 3. The parameters involved in the evaluation of E_0 according to eq 3 are obtained through a linear regression against a database of 622 formation enthalpies for $C_xH_yN_zO_w$ organic compounds listed as Supporting Information. This data set is especially well-suited to this study as it was previously used to assess state-of-the-art semiempirical Hamiltonians[33] and empirical corrections to density functional theory (DFT),[34] hence allowing for straightforward comparisons. Furthermore, to better estimate the predictive value of the present approach, an additional regression is carried out using only a subset of 269 compounds as training set. The new parameters thus obtained are then used to predict the formation enthalpies of the 353 remaining molecules.

3. MODEL PARAMETERS

Before reporting on the performance of this model, the values of the parameters fitted against the 622 compounds in the data set are compiled in this section. It must be kept in mind that such parameters may be somewhat ill-defined, owing to strong linear dependencies between the number of occurences of the corresponding interactions. In particular, the number of bonded and geminal interactions involving C and H atoms is strongly correlated and may affect the significance of the corresponding parameters. A database of ab initio data not biased toward the most commonly encountered compounds should be useful to eliminate such linear dependencies. Meanwhile, values of individual parameters must be taken with care.

3.1. Bond Parameters. Present values of the h_{i-j} parameters involved in eq 3 are listed in Table 1. They might be expected to match standard bond enthalpies reported in established compilations.[35] A comparison between both sets of bond enthalpies is provided in Figure 2. A perfect match is obviously obtained for the H−H, C≡O, and N≡N bond enthalpies, since these values are unambiguously determined by the formation enthalpies of the corresponding diatomic species H_2, CO and N_2.

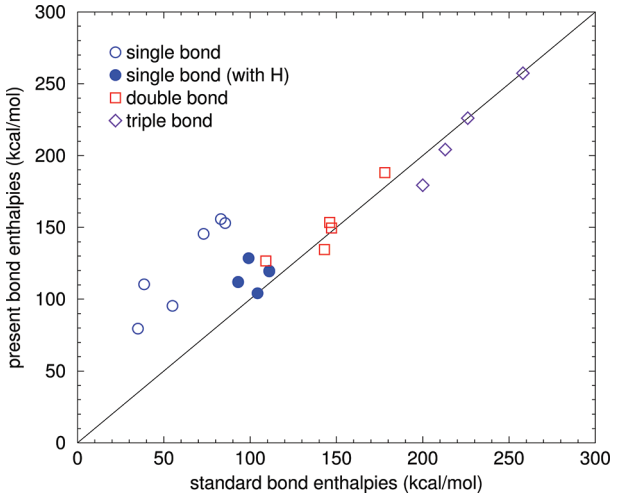

Figure 2. Comparison of present bond parameters with standard bond enthalpies.

However, it should be kept in mind that tabulated bond enthalpies derived from experiment implicitly include the destabilizing interactions between geminal atoms, in contrast to present values. Therefore, the h_{i-j} parameters are in principle larger (i.e., more stabilizing) than standard bond enthalpies. This is actually clearly the case only for single bonds between non-hydrogen atoms, as illustrated in Figure 2. As a matter of fact, the number and magnitude of destabilizing interactions is

dx.doi.org/10.1021/ct2006083 | J. Chem. Theory Comput. 2012, 8, 1295−1303

largest for such bonds. For X−H single bonds, only geminal interactions H···Y between the hydrogen atom and every other neighbor Y of the X atom are to be considered. For multiple bonds, the number of geminal interactions is similarly reduced, owing to the lower coordination number of the bonded atoms.

3.2. Geminal Interactions. The 10 geminal interactions $E_{i\cdots j}$ involved in eq 3 are listed in Table 2. Their values are remarkably similar, in the 20−30 kcal/mol range, a fact which might lead to some simplifications in future extensions of the present scheme to additional elements beyond C−H−N−O. A steady increase from about 20 to 22 and finally 24 kcal/mol is observed on going from H···H to H···C to C···C interactions. The repulsions tend to be more significant (closer to 30 kcal/mol) as soon as atoms with lone pairs (namely N and O) are considered. The only exception concerns the N···O repulsion. The latter arises mainly for amide compounds in the present database. Therefore, the specially low value (21.6 kcal/mol) observed for this repulsion is no surprise, as it absorbs the enhanced stabilization associated with the amide group. In fact, introducing a specific amide contribution leads to a stabilization energy close to 5 kcal/mol associated with every amide moiety, while the N···O repulsion increases from 21.6 to 24.4 kcal/mol.

Table 2. Geminal Interaction Parameters $E_{i\cdots j}$ Involved in eq 3 and Corresponding N_{occ} in Parentheses in the Data Set

	H	C	N	O
H	19.6(554)	–	–	–
C	22.2(586)	24.1(566)	–	–
N	27.1(87)	27.4(108)	27.9(13)	–
O	29.4(162)	28.5(257)	21.6(20)	26.0(99)

3.3. Ring Strains. The values of the bond−bond repulsion energies in strained rings are reported in Table 3. As expected, they are very small for five-membered rings, where valence angles are only slightly distorted and steadily increase on going from five- to four-membered and especially to three-membered rings. In addition, these repulsion energies tend to be more significant as multiple bonds are considered, again confirming general trends.

Table 3. Repulsions between Adjacent Bonds in Strained Rings (kcal/mol) and Corresponding N_{occ} in Parentheses

(−3−)	8.9(25)	(−4−)	6.3(16)	(−5−)	1.2(51)
(−3=)	18.6 (5)	(−4~)	11.6 (1)	(−5~)	0.8 (4)
		(−4=)	8.0 (3)	(−5=)	1.6(10)
				(~5~)	1.9(19)

4. RESULTS

4.1. Performance of the Present Model. The performance of models at predicting $\Delta_f H^0$ for a set of N compounds is most often characterized by the average absolute deviation (AAD) or root-mean-square deviation (rmsd) from experiment. The minimal (MIN) and maximal (MAX) deviations are also of interest. These data derived from individual enthalpies listed as Supporting Information are provided in Table 4 for both the present method and the well-established semiempirical quantum chemical schemes. In addition to GR-BH results, RM1 data are obtained as part of the present work, using the MOPAC7 program.[36] In this table, B3LYP denotes $\Delta_f H^0$ calculations based on simple atom equivalents to convert total energies calculated using the B3LYP density functional to formation enthalpies.[34] B3LYP+dispersion stands for similar calculations based on dispersion-corrected B3LYP/6-31+G(d,p) energies.[34]

Table 4. Performance at Predicting the Present 622 Enthalpies[a]

	N	AAD	rmsd	MIN	MAX
training set	269	2.5	4.0	−24.0	+17.6
test set	353	3.2	5.1	−31.5	+10.5
whole set	622	2.9	4.6	−31.5	+17.6
whole set[b]	622	2.7	4.3	−25.6	+19.5
B3LYP/6-31G(d)[b,d]	622	3.1	n/a	n/a	n/a
B3LYP+dispersion[b,d]	622	2.4	n/a	n/a	n/a
PDDG/PM3[c]	622	3.2	4.7	−39.1	+29.1
PDDG/PM3[b,d]	622	2.8	n/a	n/a	n/a
PDDG/MNDO[c]	622	5.2	7.4	−51.4	+41.6
RM1	622	4.2	6.7	−48.5	+55.9
PM5[c]	622	4.8	8.4	−100.5	+53.6
MNDO[c]	622	8.4	12.6	−49.6	+94.6
AM1[c]	622	6.7	9.0	−30.8	+38.2
PM3[c]	622	4.4	6.0	−34.9	+35.8
bond additivity[b]	622	9.8	15.1	−118.5	+72.5

[a]AAD, rmsd, MIN, and MAX are calculated from the N molecules in the corresponding set and reported in kcal/mol. GR-BH results are reported in the top four rows, with the top three being obtained with the model fitted against the 269 compounds in the training set. [b]Fit against all 622 compounds in the data set. [c]Ref 33. [d]Ref 34.

Since the performance criteria depend to some extent on the data sets employed to fit and assess the predictive schemes, they have been computed not only for the present model fitted against the 622 compound in the data set but also for other parameter values fitted against the 269 compounds in the training set. On going from the training to the test set, AAD and rmsd increase respectively from 2.5 to 3.2 kcal/mol and from 4.0 to 5.1 kcal/mol, as shown in Table 4. This confirms the good predictive values of GR-BH compared with the much more costly quantum chemical procedures.

4.2. Comparison with Quantum Models. According to present results, the predictive value of the GR-BH procedure, as estimated from the predictions carried out for the 353 molecules in the test set, compares favorably to corresponding AAD and rmsd data for the best semiempirical methods, including PDDG/PM3. In fact, the performance of GR-BH is similar to that of B3LYP/6-31G(d), although clearly not as good as B3LYP+dispersion. Focusing on models fitted against all 622 compounds, the lowest AAD are B3LYP+dispersion < GR-BH < PDDG/PM3 < B3LYP/6-31G(d).

GR-BH involves 49 parameters to fit the 622 formation enthalpies in the present data set. This number of disposable parameters is similar to the numbers of parameters in semiempirical methods, namely 45 and 69 for the MNDO and PM3 Hamiltonians, respectively, and 61 to 85 for the corresponding PDDG versions.[33] The reason why these quantum chemical procedures do not perform as well as GR-BH, in spite of stronger physical grounds and larger numbers of disposable parameters, clearly lies in the fact they attempt to describe many molecular features in addition to $\Delta_f H^0$, including equilibrium geometries and electronic structures.

However, even if parameters are specifically optimized for the prediction of $\Delta_f H^0$, the formalism provided by semiempirical Hamiltonians does not prove significantly superior to

dx.doi.org/10.1021/ct2006083 | J. Chem. Theory Comput. 2012, 8, 1295−1303

Table 5. Theoretical $\Delta_f H^0$ Values Calculated Using Various Methods for the $C_{40}H_{56}$ Carotenes Studied in Ref 44 (kcal/mol)[a]

	B3LYP	AM1	PM3	Leal	GR-BH
α-carotene	91.3	82.1 (−9.2)	78.1 (−13.2)	87.3 (−4.0)	88.2 (−3.1)
β-carotene	89.8	86.4 (−3.4)	79.6 (−10.2)	n/a	85.6 (−4.2)
lycopene	102.1	58.4 (−43.7)	53.0 (−49.1)	57.3 (−44.8)	122.2 (+20.1)
prolycopene	114.3	62.2 (−52.1)	56.8 (−57.5)	61.4 (−52.9)	122.2 (+7.9)

[a]Deviations from B3LYP values in parentheses.

the GR-BH procedure, according to the AAD and rmsd values of 3.4 and 4.9 kcal/mol obtained for a fit against an extended data set of 1217 formation enthalpies, including additional elements beyond C, H, N, and O.[37] This is understandable as this formalism arises from an original emphasis on the electron wave function, with subsequent attempts to describe geometries and enthalpies concerned mostly with reoptimizations of the parameters.

Finally, the performance of the GR-BH method is especially striking with regard to the minimum and maximum deviations, whose magnitudes are significantly lower than corresponding deviations obtained using the semiempirical methods, as clear from Table 4. More accurate predictions for the whole data set can only be obtained using much more costly calculations, based on either advanced density functional schemes or ab initio techniques. All in all, the performance of GR-BH is remarkable considering the simplicity of the method, the relatively small number of parameters involved, and the room still open for improvement.

4.3. Comparison with Bond and Group Additivity Schemes. The present model exhibits significant differences when compared to previously reported additive schemes, based on either bond[38−40] or group[41−43] contributions. In its simplest approximation, the bond additivity method consists of splitting $\Delta_f H^0$ into standard bond contributions, or equivalently, in considering only the bond parameters in the GR-BH procedure. The performance of this naive approach when applied to the present compounds is reported in the last row of Table 4. A 3-fold reduction of typical errors is achieved on going from this oversimplified picture to the present model.

In principle, even more accurate results can be obtained using more sophisticated bond addivity schemes, such as the Laidler[38,39] or Pedley[40] methods, which introduce additional parameters to account for the bond environments. Group contribution methods, as introduced by Benson et al., also perform quite well.[43] However, these good performances are obtained at the cost of extensive parametrization. For the sake of illustration, let us consider a simple group contribution method where a group is defined as a nonterminal atom characterized by its atomic symbol, the symbols of its neighbors and the multiplicity of its bonds. For instance, $H_3C-COOH$ is viewed as the collection of three groups, denoted $C(-C)(-H)_3$, $C(-C)(=O)(-O)$, and $O(-C)(-H)$. Such a group contribution scheme requires as many as 96 groups to describe all 622 compounds in the data set. Nevertheless, it is not suitable for highly strained rings, as it would predict exactly the same energy content (reported on a per mass unit) for cyclopropane and cyclohexane. In contrast, the present approach requires only 40 parameters if ring strains are neglected and 49 parameters otherwise. As a result, the determination of these parameters from the present data set is straightforward. In contrast, these data are too scarce for developing a conventional group contribution method applicable to all compounds. As pointed out by Benson and co-workers,[43] the large number of groups required can be a deterrent to their routine application, although the associated parameters can now be obtained from ab initio calculations. However, their main drawback in the present context stems from their heavy dependence on the bond topology, which makes them inconvenient in view of dynamic descriptions of reactive processes.

5. FURTHER ASSESSMENT OF THE METHOD

5.1. Solids and Extended Conjugated Systems. In contrast to quantum models, it is in principle straightforward to apply the GR-BH method to covalent solids. In practice, the reliability of the results is unpredictable, considering the parametrization of the present method against standard organic molecules. Nonetheless, the value $\Delta_f H^0$ = 4.2 kcal/mol predicted for both graphite and diamond is only about 4 kcal/mol too large with respect to experiment. In fact, GR-BH predicts diamond to be slightly more stable (by 0.01 kcal/mol) than graphite, in contrast to experiment.

The fairly good prediction obtained for graphite is rather unexpected as a number of significant effects are clearly lacking in GR-BH, including conjugation. Thus, one might expect formation enthalpies of extended conjugated structures, like graphite, to be overestimated. In fact, this model yields $\Delta_f H^0$ = 735 kcal/mol for fullerene C_{60}, a value significantly larger than the experimental estimate of 610 ± 30 kcal/mol,[32] although not as large as corresponding values of 973, 812, and 779 kcal/mol obtained using AM1, PM3, and RM1. Assuming that the experimental value is correct, the overestimated values obtained from calculations are especially puzzling for semiempirical methods, since they do in principle account for conjugation.

Beyond C_{60} and graphite, experimental $\Delta_f H^0$ values for extended conjugated molecules are scarse. Furthermore, although $\Delta_f H^0$ can be routinely obtained from high-level ab initio calculations for small molecules, its theoretical determination for large compounds remains challenging. This is clearly illustrated by a recent study of four $C_{40}H_{56}$ carotenes.[44] The most reliable approach applicable to such large conjugated compounds consists in DFT calculations combined with isodesmic reaction schemes. It was pointed out in this earlier study that for a given level of theory, differences up to about 10 kcal/mol may be observed between $\Delta_f H^0$ predictions depending on the isodesmic reactions considered. In addition, it was shown that significantly smaller values are predicted by various more approximate approaches, including a recent additivity scheme and semiempirical Hamiltonians, with or without empirical corrections (see Supporting Information). In contrast to these approximate procedures, the even simpler GR-BH model is unable to account for the small enthalpy difference between lycopene and prolycopene. Nevertheless, on average, it provides numerical values in much better agreement with DFT, despite a significant overestimation for lycopene, consistent with the lack of mechanism to account for the stabilization

associated with the all-trans conjugated backbone. These results are summarized in Table 5.

5.2. Strained Cage Structures. Another expected limitation of the present method stems from its simple description of ring strain energy based only on the nine parameters listed in Table 3. Since only 41 molecules out of 622 exhibit either 3- or 4-membered rings, further validation of the approach is needed. In particular, because of the additive character of present strain corrections, the GR-BH method might perform poorly for cage compounds, despite the excellent result (within 1 kcal/mol from experiment) obtained for cubane, the only cage compound in the present data set. To further investigate the performance of GR-BH for such molecules, by comparison to AM1, PM3, and RM1 methods, these models are applied to 20 homocubanes. The results are reported in Table 6 and compared to accurate $\Delta_f H^0$ data

Table 6. Theoretical $\Delta_f H^0$ Values Calculated Using Various Methods for the Homocubanes Studied in Ref 45 and Corresponding AAD, MIN, and MAX Deviations from G3MP2B3 data (kcal/mol)

	G3MP2B3	AM1	PM3	RM1	GR-BH
1	94.7	103.9	75.4	61.5	105.1
2	87.3	91.4	67.9	53.8	89.8
3	66.1	69.2	48.6	34.3	76.3
4	47.5	61.3	41.7	27.3	60.6
5	62.8	59.9	40.4	25.7	76.3
6	46.3	65.5	45.2	31.6	60.6
7	18.3	22.9	12.3	−2.7	31.8
8	8.4	28.1	17.0	1.9	16.0
9	22.9	34.1	22.1	7.2	31.8
10	35.8	32.1	20.5	5.5	46.4
11	40.3	47.6	33.6	18.5	45.2
12	42.9	59.4	44.2	30.6	45.2
13	17.8	25.7	14.5	−0.7	31.8
14	58.3	49.3	33.9	20.7	65.6
15	59.5	57.8	41.9	27.8	65.6
16	36.8	27.4	14.7	3.5	52.2
17	31.2	26.1	14.6	−0.5	46.4
18	45.3	46.1	32.5	18.4	46.4
19	84.7	94.4	72.8	59.5	84.8
20	49.3	47.6	32.5	17.6	59.8
AAD	-	8.0	12.5	25.7	9.1
MIN	-	−9.4	−24.4	−37.6	+0.1
MAX	-	+19.6	+8.6	−6.5	+15.3

obtained at the G3MP2B3 level of theory.[45] Average absolute deviations from the G3MP2B3 values are, respectively, 8.0 kcal/mol for AM1, 9.1 kcal/mol for GR−BH, 12.5 kcal/mol for PM3, and 25.7 kcal/mol for RM1. They suggest that the low average deviations previously reported using RM1 might be obtained at the expense of the reliability of the method for unusual structures. In addition, the relatively poor performance of PM3 and RM1 may be attributed to the lack of an explicit recipe to account for strain in such structures.

While AM1 and GR-BH yield the lowest deviations from G3MP2B3 values, it may be noted that GR-BH values are systematically overestimated, which is gratifying in view of further improvement of the method, taking advantage either of 3D geometries or additional corrections. In contrast, AM1 deviations prove less systematic.

5.3. Reaction Mechanisms. Because highly efficient schemes such as GR-BH are especially needed for reactivity problems whose mechanism is a priori unknown and in view of future investigations of reactive potentials based on this simple picture, a preliminary assessment of this procedure against decomposition reactions is carried out in this section. As it stands, applying GR-BH to such problems is quite awkward and probably untimely. However, it might point to some deficiencies and help focus further investigations.

To date, the worst deviations from high-level reference data have been observed for the unimolecular decomposition of 5-aminotetrazole and 5-iminotetrazole recently studied by Paul et al.[46] The enthalpy changes associated with reactions R1−R9, shown in Figure 3, have been calculated at the GR-BH level.

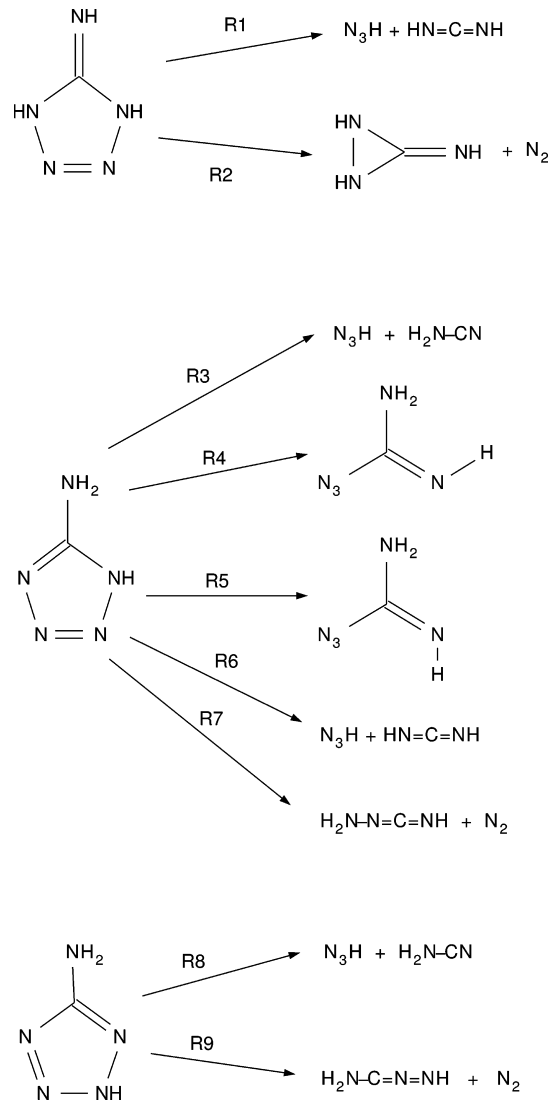

Figure 3. Some reactions involved in the decomposition of tetrazole derivatives and studied in ref 46.

The results are summarized in Table 7. Most of them are reasonable, with deviations from reference ab initio data no larger than corresponding deviations calculated using semi-empirical methods. However, a dramatic breakdown is observed for R1, R6, and R7, with reaction enthalpies too large by about 100 kcal/mol. An examination of $\Delta_f H^0$ data reveals that these errors reflect a systematic overestimation of the enthalpies of

dx.doi.org/10.1021/ct2006083 | J. Chem. Theory Comput. 2012, 8, 1295−1303

Table 7. Reaction Enthalpies (kcal/mol) for the Decomposition of 5-Iminotetrazole, 1*H*-5-Aminotetrazole, and 2*H*-5-Aminotetrazole Compared to Reference Data Computed in Ref 46 at the CCSD(T)/aug-cc-pVTZ//B3LYP/6-311++G(3df,3pd) Level of Theory

	CCSD(T)	AM1	PM3	RM1	GR-BH
R1	15.9	−4.3	28.6	26.0	127.3 (+111.4)
R2	−5.5	−24.7	12.0	38.4	−8.9 (−3.3)
R3	24.2	−4.1	27.6	29.2	18.6 (−5.6)
R4	7.5	−6.4	2.5	13.6	2.3 (−5.2)
R5	9.9	−6.0	5.3	15.3	2.3 (−7.6)
R6	27.0	7.7	38.8	43.3	130.8 (+103.8)
R7	−8.5	−35.5	1.2	23.9	86.4 (+94.9)
R8	26.6	−8.4	25.8	26.6	13.7 (−12.9)
R9	15.9	−29.9	10.2	28.0	−20.5 (−36.5)

the two carbodiimide products $H_2N-N=C=NH$ and $H-N=C=NH$. Among the 622 compounds primarily considered in this work, there is no such compound, and only 6 molecules exhibit the C=N double bond. Another significant error concerns the reaction enthalpy of R9, too low by 36 kcal/mol, primarily owing to the fact that GR-BH underestimates the enthalpy of $H_2N-C^+=N^-=NH$ by about 20 kcal/mol with respect to the other methods.

It is clear that the present approach, as implemented in GR-BH, is not yet ready to predict reliable thermochemical data for unstable species involved in decomposition reactions. This is no surprise in view of its underlying assumptions and parametrization against standard close-shell neutral organic molecules. For instance, it is unlikely to satisfactorily describe the enthalpy of zwitterionic species, such as $H_2N-C^+=N^-=NH$. Nevertheless, according to the results obtained for carbodiimide, the errors appear to be systematic and could therefore be accounted for through additional corrections. For instance, introducing an additive correction for this moiety immediately decreases the AAD deviation for R1, R6, and R7 to only 6 kcal/mol. Further development of the present scheme into a full-blown potential energy function of atom coordinates and bond orders will provide additional opportunities to remedy to such deficiencies.

6. CONCLUSION

This work demonstrates that the introduction of pairwise interactions between geminal atoms and adjacent bonds in strained rings dramatically improves the performance of simple bond additivity schemes for $\Delta_f H^0$ prediction, as reflected by a 3-fold decrease of the typical deviations from experiment. The new predictive scheme thus obtained exhibits unique features. While not as reliable as group additivity methods and ab initio procedures, it is more versatile than the former and much more efficient than the latter. It may be viewed as a highly efficient alternative to semiempirical methods for standard organic compounds.

In fact, it appears especially promising as a basis toward further developments, the most obvious one being the design of a versatile and reliable predictive scheme for organic compounds. To date, the addition of group equivalents to semiempirical enthalpies provides such a method, and a value as low as 2.2 kcal/mol was reported for the AAD obtained using this approach along with PM3.[33] Combining such group equivalents with GR-BH appears even more attractive as the computational overhead associated with quantum aspects would be completely avoided.

Another potential application of the present scheme is to the detailed study of unknown complex reaction mechanisms, where it could contribute to focus high-level computations onto the most plausible reaction pathways. However, it is first necessary to make up for some limitations of the method outlined in this paper.

This should be possible through the introduction of empirical corrections and/or the development of the method into a full-blown reactive potential, with the availability of atomic positions and dynamic bond orders and charges providing additional opportunities for improvement. Much work is still needed in order to assess the viability of this approach. Nevertheless, given the fact that it is still in its infancy, present results are quite encouraging.

■ ASSOCIATED CONTENT

Supporting Information

A table with GR-BH heats of formation for the 622 compounds in the database, compared with the values calculated using semiempirical Hamiltonians and experiment. The values in the columns, from left to right, are: compound name, data set (Fit = training set; Tst = test set), experimental heat of formation, and deviations from experiment computed using various theoretical levels (kcal/mol). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

E-mail: didier.mathieu@cea.fr.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Matsuia, M.; Akaogib, M. *Mol. Simul.* **1991**, *6*, 239−244.
(2) Finnis, M. W.; Sinclair, J. E. *Philos. Mag. A* **1984**, *50*, 45−55.
(3) Daw, M. S.; Baskes, M. I. *Phys. Rev. B* **1984**, *29*, 6443−6453.
(4) Baskes, M. I. *Phys. Rev. B* **1992**, *46*, 2727−2742.
(5) Stillinger, F. H.; Weber, T. A. *Phys. Rev. B* **1985**, *31*, 5262−5271.
(6) Tersoff, J. *Phys. Rev. B* **1988**, *37*, 6991−7000.
(7) Stoliarov, S. I.; Westmoreland, P. R.; Nyden, M. R.; Forney, G. P. *Polymer* **2003**, *44*, 883−894.
(8) Nyden, M. R.; Stoliarov, S. I.; Westmoreland, P. R.; Guo, Z. X.; Jee, C. *Mater. Sci. Eng., A* **2004**, *365*, 114−121.
(9) Keffer, D.; Jiang, B.; Selvan, M. E.; Edwards, B. J. *J. Phys. Chem. B* **2009**, *113*, 13670−13677.
(10) Brenner, D. W. *Phys. Rev. B* **1990**, *42*, 9458−9471.
(11) Brenner, D. W.; Shenderova, O. A.; Harrison, J. A.; Stuart, S. J.; Ni, B.; Sinnott, S. B. *J. Phys.: Condens. Matter* **2002**, *14*, 783−802.
(12) Los, J.; Ghiringhelli, L.; Meijer, E.; Fasolino, A. *Phys. Rev. B* **2005**, *72*, 214102.
(13) Mrovec, M.; Moseler, M.; Elsässer, C.; Gumbsh, P. *Prog. Mater. Sci.* **2007**, *52*, 230−254.
(14) Pastewka, L.; Pou, P.; Pérez, R.; Gumbsch, P.; Moseler, M. *Phys. Rev. B* **2008**, *78*, 161402R.
(15) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. III *J. Phys. Chem. A* **2001**, *105*, 9396−9409.
(16) Nielson, K. D.; van Duin, A. C. T.; Oxgaard, J.; Deng, W. W.; Goddard, W. A. III *J. Phys. Chem. A* **2005**, *109*, 493−499.
(17) Buehler, M. J.; van Duin, A. C. T.; Goddard, W. A. *Phys. Rev. Lett.* **2006**, *96*, 095505.
(18) Desai, T. G.; Lawson, J. W.; Keblinski, P. *Polymer* **2011**, *52*, 577−585.

(19) Strachan, A.; van Duin, A. C. T.; Chakraborty, D.; Dasgupta, S.; Goddard, W. A. III *Phys. Rev. Lett.* **2003**, *91*, 098301.
(20) Strachan, A.; Kober, E. M.; van Duin, A. C. T.; Oxgaard, J.; Goddard, W. A. III *Phys. Rev. Lett.* **2005**, *122*, 054502.
(21) Zhang, L.; Zybin, S. V.; van Duin, A. C. T.; Dasgupta, S.; Goddard, W. A. III *J. Phys. Chem. A* **2009**, *113*, 10619−10640.
(22) Rahaman, O.; van Duin, A. C. T.; Goddard, W. A. III; Doren, D. J. *J. Phys. Chem. B* **2011**, *115*, 249−261.
(23) Zhang, Q.; Cagin, T.; van Duin, A. C. T. III; A. G., W. *Phys. Rev. B* **2004**, *69*, 045423.
(24) Ojwang, J. G. O.; van Santen, R. A.; Kramer, G. J.; van Duin, A. C. T.; Goddard, W. A. III *J. Chem. Phys.* **2009**, *131*, 044501.
(25) Mayernick, A. D.; Batzill, M.; van Duin, A. C. T.; Janik, M. J. *Surf. Sci.* **2010**, *604*, 1438−1444.
(26) Watanabe, T. *J. Comput. Electron.* **2011**, *10*, 2−20.
(27) Mathieu, D.; La Hargue, J.-P. *CHOCS* **2011**, *39*, 36−46.
(28) Hudson, T. S.; Nguyen-Manh, D.; van Duin, A. C. T.; Sutton, A. P. *Mater. Sci. Eng., A* **2006**, *422*, 123−135.
(29) Chenoweth, K.; van Duin, A. C. T. III; A. G., W. *J. Phys. Chem. A* **2008**, *112*, 1040−1053.
(30) Mathieu, D.; Lucas, A. *Comput. Mater. Sci.* **2007**, *38*, 514−521.
(31) Cong, Y.; Yang, Z.-Z. *Chem. Phys. Lett.* **2000**, *316*, 324−329.
(32) Afeefy, H.; Liebman, J.; Stein, S. Neutral Thermochemical Data. In *NIST Chemistry Webbook*; Linstrom, P., Mallard, W., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, 2011; http://webbook.nist.gov/chemistry/.
(33) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622.
(34) Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2008**, *4*, 227−308.
(35) Anderson, R. T. *Polar Covalence*; Academic Press: New York, 1983.
(36) *MOPAC*, version 7; http://sourceforge.net/projects/mopac7 (accessed February 07, 2011).
(37) Stewart, J. J. P. *J. Mol. Model.* **2004**, *10*, 6−12.
(38) Laidler, K. J. *Can. J. Chem.* **1956**, *34*, 626−648.
(39) Santos, R. C.; Leal, J. P.; Simoes, J. A. M. *J. Chem. Thermodyn.* **2009**, *41*, 1356−1373.
(40) Pedley, J. B. *Thermochemical Data and Structures of Organic Compounds*; TRC: College Station, TX, 1994; Vol. *I*.
(41) Benson, S. W.; Buss, J. H. *J. Chem. Phys.* **1958**, *29*, 546−572.
(42) Benson, S. W. *Thermochemical Kinetics*, 2nd ed.; Wiley: New York, 1976.
(43) Benson, S. W.; Cohen, S. Current status of group additivity. In *Computational thermochemistry*; Irikura, K. K., Frurip, D. J., Eds.; American Chemical Society: Washington DC, 1998; Vol. *677*, pp 20−46.
(44) Tu, C.-Y.; Guo, W.-H.; Hu, C.-H. *J. Phys. Chem. A* **2008**, *112*, 117−124.
(45) Novak, I. *J. Chem. Inf. Model.* **2004**, *44*, 903−906.
(46) Paul, K. W.; Hurley, M. M.; Irikura, K. K. *J. Phys. Chem. A* **2009**, *113*, 2483−2490.