

## Formation Enthalpies of Ions: Routine Prediction Using Atom Equivalents

Didier Mathieu\* and Yohann Pipeau

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

Received January 13, 2010

**Abstract:** In view of identifying routine procedures to estimate formation enthalpies of ionic systems such as energetic salts or ionic liquids on the basis of density functional theory (DFT), various combinations of atom equivalent (AE) schemes, functionals, and basis sets are compared, using a specially designed training set to parametrize the models. After correction, none of the functionals considered proves significantly more reliable than B3LYP. A small but systematic improvement is noted as AE values are allowed to depend on the atomic environment. However, AE parameters fail to make up for basis set limitations, in contrast to previous observations for neutrals. Finally, a good trade-off between reliability and cost is obtained for ions using B3LYP/6-31++G\*\* energies.

### 1. Introduction

In recent years, molecular ionic species have received much interest as constituents of ionic liquids or high energy density materials.<sup>1–3</sup> Depending on the application in view, ions with high stability and/or high energy content are needed. In this context, procedures to predict the formation enthalpies of ionic compounds are needed to focus new syntheses on the most promising candidates. While a plethora of methods are available for neutral molecules, such as group additivity schemes,<sup>4</sup> molecular mechanics,<sup>5,6</sup> or quantitative structure property relationships,<sup>7,8</sup> evaluation of the standard formation enthalpy  $\Delta_f H^\circ$  of ions in gas-phase must resort to quantum mechanics for most cases of practical interest.<sup>9–12</sup> Provided a suitable basis set is used, high level *ab initio* methods perform well for charged species without any specific treatment.<sup>13,14</sup>

However, more efficient methods are needed to predict  $\Delta_f H^\circ$  for arbitrary ions on a routine basis. Unfortunately, detailed investigations of the performance of such procedures when applied to charged species are still lacking. Available data indicate that semiempirical methods are not suitable for such systems,<sup>15,16</sup> while simple approaches based on density functional theory (DFT) yield reasonable estimates of the energy content of energetic salts.<sup>17,18</sup> Therefore, this work compares simple quantum procedures, with a focus on

efficient DFT-based methods involving empirical local corrections. Such procedures have been extensively investigated for neutrals.<sup>19,20</sup> However, they might be less effective for ions in view of the delocalization arising from the fact that their wave function usually involves several canonical structures.

More specifically, the present work examines how the performance of standard correction schemes is affected on going from neutrals to ions. Experimental gas phase enthalpies of ions, taken from the literature, are compared to the results of extensive calculations using various levels of theory for the electronic energy, different procedures to convert this energy into formation enthalpies, and a newly introduced approach to define the training set employed to fit the adjustable parameters involved. A clear picture of the accuracy that may be obtained according to the specific procedure employed is thus obtained. Cost effective procedures are identified, with a reliability consistent with the uncertainties associated with the evaluation of intermolecular interactions.<sup>9,18,21</sup> They are now implemented in a user-friendly software package<sup>22</sup> and allow bench chemists to perform routine calculations of  $\Delta_f H^\circ$  for materials of potential interest. On the other hand, in contrast to many recent papers in the field,<sup>23–25</sup> no attempt is made to improve the accuracy of state-of-the-art DFT methodologies for gas-phase enthalpies.

\* To whom correspondence should be addressed. E-mail: didier.mathieu@cea.fr.

## 2. Computational Methods

**2.1. Theoretical Framework.** By definition, the formation enthalpy of a given compound made of different atomic elements H, C, N, ... with  $Z = 1, 6, 7, \dots$  is obtained as the difference between the theoretical enthalpy  $H^\circ$  of the compound and the corresponding enthalpies  $H^\circ(Z, \text{stp})$  of every atom  $Z$  in its standard reference state (i.e., standard temperature and pressure):

$$\Delta_f H^\circ = H^\circ - \sum_Z n_Z H^\circ(Z, \text{stp}) \quad (1)$$

Here,  $n_Z$  is the number of atoms with atomic number  $Z$  in the compound. For most light elements, the theoretical evaluation of  $H^\circ(Z, \text{stp})$  is straightforward. For instance, for the nitrogen atom, it is simply  $H^\circ(\text{N}_2, \text{stp})/2$  since its standard reference state is the  $\text{N}_2$  molecule. Nevertheless, in some cases,  $H^\circ(Z, \text{stp})$  may be more difficult to calculate. Even for carbon, whose reference state (graphite) is well-known, a problem arises because many theoretical methods are only implemented for finite systems and cannot be used to compute  $H^\circ(\text{C}_{\text{graphite}}, \text{stp})$ .

This problem is avoided by the use of molecular or atomic reference states, i.e., gaseous species  $i$  with well-established formation enthalpies  $\Delta_f H^\circ(i)$ . The formation enthalpy is then obtained as

$$\Delta_f H^\circ = \sum_i N_i \Delta_f H^\circ(i) + (H^\circ - \sum_i N_i H^\circ(i)) \quad (2)$$

where the numbers  $N_i$  of the reference species must be consistent with the empirical formula of the compound studied. In other words, if  $n_Z^i$  stands for the number of atoms  $Z$  in species  $i$ , the values of  $N_i$  must satisfy for every element  $Z$ :

$$\sum_i n_Z^i N_i = n_Z \quad (3)$$

In principle, any set of reference compounds  $i$  can be used as long as the linear system defined by eq 3 is regular and exhibits a solution  $\{N_i\}$ . Notwithstanding the role of the level of theory used to compute  $H^\circ$  data, it is important to keep in mind that calculated  $\Delta_f H^\circ$  values also depend in practice on the reference compounds selected.

In some cases, reference compounds  $i$  with accurate  $\Delta_f H^\circ$  data available may be found in such a way that the number of each kind of chemical bond is the same for the reference compounds and for the molecule studied. In other words, the compound studied is obtained from species  $i$  through an isodesmic reaction.<sup>4</sup> This situation is especially favorable as the calculation of the enthalpy difference between the molecule under study and reference species benefits from an effective cancellation of errors. In principle, a computer algorithm coupled with a database for accurate  $\Delta_f H^\circ$  data could be used to identify suitable reference species for every new compound studied. In the lack of such a program, the use of isodesmic reactions is not convenient for routine calculations. In fact, it is not always possible to identify a suitable isodesmic scheme. For ions, this approach is

problematic as bond orders cannot always be assigned unambiguously.

Relaxing the constraint to rely on isodesmic reactions allows more practical approaches. For instance, a molecular reference species  $Z_m$  may be introduced for every atomic element  $Z$  in the compound studied.<sup>26</sup> Possible values of  $m$  are  $m = 2, 60, 2, 2, 8, \dots$  for elements H, C, N, O, S, ..., respectively. In this case, the linear system defined by eq 3 is diagonal. As a result, eq 2 becomes simply

$$\Delta_f H^\circ = H^\circ + \sum_Z \frac{n_Z}{m} (\Delta_f H^\circ(Z_m) - H^\circ(Z_m)) \quad (4)$$

It is clear from eq 4 that the role of the differences  $\Delta_f H^\circ(Z_m) - H^\circ(Z_m)$  between experimental and theoretical enthalpies is simply to shift the zero of enthalpies in order to make theoretical data consistent with the conventional thermodynamic reference state. On the other hand, the second term in this equation depends in practice on the reference compounds used because of the uncertainties associated with experimental  $\Delta_f H^\circ(Z_m)$  and theoretical  $H^\circ(Z_m)$  data.

As the formation enthalpies of gaseous atoms are well-known for the main group elements, their use as reference states yields small uncertainties associated with experimental data. Formation enthalpies are then obtained as the difference between the formation enthalpies of the gaseous atoms and the atomization enthalpy of the compound:

$$\Delta_f H^\circ = \sum_Z n_Z \Delta_f H^\circ(Z) - (\sum_Z n_Z H^\circ(Z) - H^\circ) \quad (5)$$

Unfortunately, this approach leads to much larger errors than isodesmic reaction schemes since the errors on calculated enthalpies do not effectively cancel for atomization reactions. A simple approach to remove these systematic errors consists in introducing empirical parameters  $X_Z$ :

$$\Delta_f H^\circ = H^\circ + \sum_Z n_Z C_Z = H^\circ + \sum_Z n_Z (\Delta_f H^\circ(Z) - H^\circ(Z)) + \sum_Z n_Z X_Z \quad (6)$$

In further attempts to improve the results,  $X_Z$  may be assumed to depend not only on the atomic number  $Z$  of the atom but also on its environment within the molecule under study:

$$\Delta_f H^\circ = H^\circ + \sum_Z n_Z (\Delta_f H^\circ(Z) - H^\circ(Z)) + \sum_A X_A \quad (7)$$

The last sum in eq 7 runs over every atom in the compound studied. The present work compares the performance of procedures based on this equation, according to the level of theory used to compute  $H^\circ$  and the actual definition adopted for the  $X_A$  parameters, hereafter referred to as atom equivalents (AEs).

Beyond AE schemes, a number of alternatives have been introduced to convert HF or DFT energies into formation enthalpies while correcting their main deficiencies.<sup>27,28</sup> These methods focus on neutral molecules and rely on empirical relationships. However, simpler procedures based on eq 7 are not necessarily less reliable.<sup>29,39</sup> In favorable situations,

**Table 1.** Definition of the Correction Procedures P3, P5, P8, and P10 Used in This Work

|     |  |
|-----|--|
| P3  | specific AE for H, $X_A = a + bZ$ for $A = \text{C, N, O, F}$                  |
| P5  | one AE for every element: H, C, N, O, F  |
| P8  | additional AEs $\text{C}', \text{N}', \text{O}'$ for atoms with multiple bonds |
| P10 | H1, C4, C3, C2, N3, N2, N1, O2, O1, F1   |

they yield average absolute deviations (AAD) from experimental results close to 2.5 kcal/mol.<sup>30,40</sup> Therefore, all models considered in the present work rely on eq 7. They differ only in the definition adopted for the  $X_A$  parameters.

**2.2. Definition of Atom Equivalents.** Five procedures are considered in this work for predicting formation enthalpies of ions from theoretical total energies and eq 7. They are named  $P_m$  where  $m$  is the number of adjustable parameters required to fit  $\Delta_f H^\circ$  values of compounds made of HCNOF atoms. In procedure P5,  $X_A$  is assumed to be independent of the environment of the atom. It depends only on the atomic number  $Z$  of the atom. Therefore, only five adjustable parameters are required for HCNOF. Procedure P8 introduces three additional AEs  $\text{C}'$ ,  $\text{N}'$ , and  $\text{O}'$  for atoms involved in multiple bonds, as done by Rice and co-workers.<sup>39,41</sup>

In procedure P10, an atom equivalent depends not only on the atomic symbol  $X$  but also on the atom coordination number  $n$ .<sup>42</sup> It is thus denoted  $Xn$ . In principle, the 10 P10 parameters listed in Table 1 do not allow for handling of ammonium and hydronium cations, due to the lack of N4 and O3 parameters. In fact, it was previously noted that using the N3 value instead of N4 for ammonium salts yields satisfactory results.<sup>17</sup> In this work, using N3 and O2 parameters for  $\text{NH}_4^+$  and  $\text{H}_3\text{O}^+$  does not lead to specially large deviations. This indicates that N4 and O3 values derived from these two ions are close to the N3 and O2 parameters, which refer to atoms in different bonding environments but with the same hybridizations. The derivation of optimal N4 and O3 values is beyond the scope of this work, as such values should be averaged over typical atomic environments. Moreover, for reasons detailed in section 3, present parameters are derived from gas-phase data for neutrals.

Finally, procedure P3 aims at reducing the number of adjustable parameters by taking advantage of the linear correlation often observed between the values of the equivalents for CNOF atoms and the corresponding atomic numbers. In other words, while a constant value is attributed to the H equivalent, the others are assumed to vary according to  $X_A = a + bZ$  where  $Z$  is the atomic number of atom  $A$ . This leaves only three adjustable parameters:  $a$ ,  $b$ , and  $H$ . Such a procedure to reduce the number of empirical parameters is especially interesting in view of extending AE methods beyond first-row atoms.

These procedures are summarized in Table 1. In earlier studies, specific AEs or group equivalents are sometimes introduced for some special chemical groups, such as nitros or azides.<sup>41,42</sup> However, such group specific parameters hamper the generality of the procedure. In fact, they are often unsuitable for ions because of ill-defined bond orders. Accordingly, no attempt is made here to introduce such group parameters. On the other hand, alternative approaches based

on bond equivalents (BEs), charge-dependent AEs, or BEs depending on Mulliken bond populations have also been investigated. However, for the present compounds, they prove significantly worse than the  $P_n$  procedures described above, despite encouraging results sometimes reported for other data sets or theoretical levels.<sup>30,31,43</sup>

**2.3. Computational Procedures.** Having defined the atom equivalents  $X_A$  involved in eq 7, computational procedures remain to be selected to derive theoretical enthalpies  $H^\circ(Z)$  for gaseous atoms and  $H^\circ$  for the compound under study. In this work,  $H^\circ(Z)$  is computed at the G3MP2B3 level,<sup>44</sup> a specially efficient version of the well-known G3 composite method.<sup>45</sup> This choice is irrelevant as any deficiency in  $H^\circ(Z)$  will be absorbed into the AE values  $X_A$ . However, an explicit evaluation of  $H^\circ(Z)$  with reasonable accuracy makes it possible to interpret  $X_A$  as approximate corrections to  $H^\circ$ .

The evaluation of molecular enthalpies  $H^\circ$  from total electronic energies  $E_0$  (frozen atoms at 0 K) and vibrational frequencies is straightforward within the ideal gas and harmonic approximations. Because frequencies add a significant computational overhead compared with single-point energy calculations, simple additive schemes have been developed to estimate  $H^\circ - E_0$ , which includes the zero-point energy as well as thermal contributions.<sup>30,43,46</sup> In this work, it is obtained from standard enthalpic corrections  $H_{\text{corr}}(Z)$  introduced by Winget and Clark:<sup>46</sup>

$$H^\circ = E_0 + \sum_Z n_Z H_{\text{corr}}(Z) \quad (8)$$

Finally, the approach based on eqs 7 and 8 amounts to adding atomic parameters to the total quantum chemical energy  $E_0$  in order to obtain the formation enthalpy:

$$\Delta_f H^\circ = E_0 + \sum_Z n_Z (\Delta_f H^\circ(Z) - H^\circ(Z) + H_{\text{corr}}(Z)) + \sum_A X_A = E_0 + \sum_A Y_A \quad (9)$$

The  $Y_A$  parameters introduced previously<sup>42</sup> and also referred to as  $-\epsilon_A$ <sup>18</sup> are commonly used for straightforward conversion of  $E_0$  data into  $\Delta_f H^\circ$  values. In this work, it was decided to make their various contributions explicit. This facilitates the interpretation of the  $X_A$  parameters. Since the latter should ideally be zero, their magnitude provides an estimate of the errors in the other contributions. Previous studies show that these errors are dominated by the uncertainties associated with  $E_0$  values.<sup>42,46</sup>

The derivation of formation enthalpies from eq 7 involves the calculation of total energies  $E_0$ . They are computed using various levels of theory: the nonlocal exchange HF functional;<sup>47</sup> the self-consistent-charge density functional tight binding scheme (SCC-DFTB);<sup>48</sup> functionals based on the local density approximation (LDA):  $X\alpha$ <sup>49</sup> and SVWN;<sup>49,50</sup> functionals based on the generalized gradient approximation (GGA): BP86,<sup>51,52</sup> BLYP,<sup>51,53,54</sup> PW91,<sup>55,56</sup> mPW91,<sup>57</sup> PBE,<sup>58,59</sup> and HCTH;<sup>60</sup> hybrid GGA functionals (H-GGA): B3LYP,<sup>51,53,61</sup> B3P86,<sup>51-53</sup> B3PW91,<sup>51,53,55</sup> PBE1PBE,<sup>62</sup> B1LYP,<sup>63</sup> B98,<sup>64</sup> and the “half and half” functionals



BHandH and BHandHLYP implemented in Gaussian<sup>65</sup> following those introduced by Becke;<sup>66</sup> meta GGA functionals (M-GGA): TPSS<sup>67</sup> and VSXC;<sup>68</sup> and hybrid meta GGA functionals (HM-GGA): B1B95<sup>61</sup> and BMK.<sup>69</sup>

The large number of functionals presently considered stems from the fact that AE-based correction procedures have been so far applied only to a few popular functionals, especially BP86 and B3LYP. A systematic application of such procedures is of interest because functionals discarded as inaccurate on the basis of raw  $\Delta_f H^\circ$  predictions might prove valuable if the errors lend themselves to effective corrections. For instance, earlier attempts to obtain good thermochemistry from X $\alpha$  calculations rely on atom-dependent values of the Slater exchange parameter  $\alpha$  rather than AE-based corrections.<sup>70</sup> Although so far unsuccessful, the search for effective procedures to estimate  $\Delta_f H^\circ$  from X $\alpha$  calculations is of special interest as the X $\alpha$  functional lends itself to a fully analytic calculation of the Hamiltonian matrix, in contrast to the others.<sup>71</sup> In this work, the standard value of 0.7 is used for the Slater coefficient in order to make up for the lack of an explicit correlation model. With regard to the correlation part of SVWN, the default version in Gaussian<sup>65</sup> is used, namely, the one numbered III in the original paper.<sup>50</sup> Present nomenclature for other DFT functionals follows the one adopted in a recent review of their performance, in which a more comprehensive list of references may be found.<sup>72</sup> More recent HM-GGA functionals, including TPSSh<sup>67</sup> or the M06 family,<sup>25,73</sup> are not yet available in our group and lie beyond the scope of the present paper. Although they are relatively costly, these functionals perform remarkably well without *a posteriori* corrections. It will be of interest to investigate whether their predictions can be further enhanced by correction procedures such as those considered here.

In addition to the new AEs introduced in the present paper, alternative approaches to  $\Delta_f H^\circ$  are used for comparison: G3MP2B3,<sup>44</sup> and popular semiempirical methods based on the NDDO approximation, namely, AM1,<sup>74</sup> PM3,<sup>75,76</sup> and RM1.<sup>15</sup> The latter is a recent reparametrization of the AM1 Hamiltonian which provides a remarkable improvement for the prediction of formation enthalpies of organic and biological molecules. In order to compare the predictive power of the models for ions, the root-mean-square deviation (RMSD) between calculated and observed values is used as the main criterion. However, the average absolute deviation (AAD) is also reported to make comparison with previous work easier. On the other hand, although this study focuses on ions, RMSD and AAD values derived from a leave-one-out cross validation of the training set data are also considered as rough indicators of the reliability of these procedures for neutral systems. The following software is used for all present calculations: MOPAC7 for semiempirical methods,<sup>77</sup> the original DFTB code for SCC-DFTB,<sup>48</sup> and Gaussian 03W for *ab initio* and DFT methods.<sup>65</sup>

### 3. Database

The selection of a suitable database to assess or parametrize computational procedures is no trivial task. The scope of the method, its expected accuracy, or the number of

adjustable parameters to be fitted must be considered. Over the years, highly accurate thermochemical data have been collected to assess the performance of high-level theories, especially composite *ab initio* models.<sup>78–82</sup> Since they exhibit only relatively small species, such data sets are not optimal to parametrize more approximate procedures applicable to large organic compounds.

Such procedures, based on either molecular mechanics, semiempirical Hamiltonians, HF, or DFT are developed using larger data sets obtained by including somewhat less reliable data, often without error bars.<sup>15,16,83,84</sup> This is acceptable, as the corresponding procedures yield typical errors significantly larger than experimental uncertainties. Extended data sets used to develop general schemes invariably exhibit a significant proportion of hydrocarbons or other simple, monofunctionalized organic compounds. As a result, they might provide too optimistic views of the reliability of a given procedure when applied to unusual compounds, such as molecular ions.<sup>46</sup>

An alternative approach consists in developing specific parameters on the basis of a restricted family of compounds. For instance, training sets focused on nitro compounds have been used to derive AEs specially optimized for energetic materials.<sup>39,41</sup> Specialized equivalents have also been published for hydrocarbons,<sup>85,86</sup> propellanes,<sup>87</sup> and some monofunctionalized compounds.<sup>88,89</sup> For molecular ions, this strategy is not well suited owing to the scarcity and relative lack of reliability of available gas phase data. Furthermore, while AE values depend on local atom environments, they should not depend on the total charge of the compound. Indeed, besides a charged group, ions may exhibit the same functional groups as neutral molecules. Therefore, it would make no sense to develop specific AEs for ionic systems.

Accordingly, present procedures are parametrized exclusively against data for neutral CHNOF molecules compiled in Table 2, while  $\Delta_f H^\circ$  data for ions (Tables 3 and 4) are used only for validation purposes. This provides a stringent test of the transferability of the parameters. For most compounds in the training set,  $\Delta_f H^\circ$  is reported to within <1 kcal/mol (Table 2). In contrast, error bars are often unavailable for ions. Therefore,  $\Delta_f H^\circ$  data reported in Tables 3 and 4 are prone to large uncertainties, and one should not attach too much significance to individual values. Nevertheless, they are significantly more reliable on average than present DFT-based values, as confirmed by their overall good agreement with G3MP3B3 data (Tables 3 and 4). Therefore, the corresponding RMSD and AAD values provide suitable comparison criteria to assess the performance of the present procedures.

The present training set is specially designed to ensure a balanced coverage of the many possible chemical environments for an atom in polyatomic species, thus avoiding a bias of the parametrization due to the prevalence of specific moieties such as alkyl groups. First, all possible bonds between CNOF atoms are listed, considering only 1, 2, and 3 as possible formal values for the bond orders. Then, dangling bonds are saturated with H atoms. This yields 24 simple compounds for which experimental  $\Delta_f H^\circ$  values are available. For each of these compounds, and whenever

**Table 2.** Experimental Formation Enthalpies (with Error Bars when Available) and Corresponding Deviations for Theoretical Values Calculated Using G3MP2B3, PM3, and DFT, More Specifically the P5 Procedure Applied to B3LYP/6-31++G\*\* Total Energies<sup>a</sup>

| compound  | CAS number | exptl.       | G3MP2B3 | PM3   | DFT  |
|---|------------|--------------|---------|-------|------|
| HN=NH ( <i>trans</i> )  | 15626-43-4 | 50.9 ± 2     | -2.9    | -13.1 | -3.3 |
| C <sub>10</sub> H <sub>8</sub> (naphthalene)  | 91-20-3    | 35.9 ± 2     | -2.6    | 4.8   | 1.0  |
| H <sub>5</sub> C <sub>6</sub> -NO <sub>2</sub>  | 98-95-3    | 16.4 ± 0.2   | -2.2    | -1.7  | -1.2 |
| C <sub>4</sub> H <sub>4</sub> N <sub>2</sub>  | 289-95-2   | 46.8 ± 0.4   | -2.2    | -8.6  | -4.8 |
| F-CH=CH <sub>2</sub>  | 75-02-5    | -32.5        | -1.9    | 3.8   | -3.1 |
| C <sub>9</sub> H <sub>7</sub> N (quinoline)   | 91-22-5    | 47.9         | -1.9    | -0.2  | -0.2 |
| H <sub>5</sub> C <sub>6</sub> -N=N-C <sub>6</sub> H <sub>5</sub>                          | 17082-12-1 | 96.9 ± 0.3   | -1.9    | -6.0  | -0.7 |
| HCN   | 74-90-8    | 32.3         | -1.7    | 0.5   | -1.9 |
| (CH <sub>3</sub> ) <sub>2</sub> C=O   | 67-64-1    | -49.8 ± 0.1  | -1.4    | -3.3  | -2.6 |
| H <sub>2</sub>  | 1333-74-0  | 0.0          | -1.0    | -13.4 | 6.0  |
| C <sub>6</sub> H <sub>6</sub>   | 71-43-2    | 19.8 ± 0.2   | -0.7    | 3.6   | -0.5 |
| C <sub>5</sub> H <sub>5</sub> N   | 110-86-1   | 33.6         | -0.7    | -3.1  | -1.9 |
| H <sub>2</sub> C=O  | 50-00-0    | -26.0 ± 0.1  | -0.7    | -8.1  | 0.7  |
| C <sub>2</sub> H <sub>4</sub> N <sub>4</sub> (1-methyl-1H-tetrazole)                      | 16681-77-9 | 77.2 ± 0.5   | -0.5    | 6.9   | -1.0 |
| H <sub>2</sub> C=CH <sub>2</sub>  | 74-85-1    | 12.5 ± 0.1   | -0.5    | 4.1   | 0.7  |
| HCCCH <sub>3</sub>  | 74-99-7    | 44.3 ± 0.2   | -0.5    | -4.1  | -1.0 |
| HCCH  | 74-86-2    | 54.3 ± 0.2   | -0.2    | -3.6  | 1.4  |
| H <sub>5</sub> C <sub>6</sub> -NH <sub>2</sub>  | 62-53-3    | 20.8 ± 0.2   | -0.2    | 0.5   | 0.5  |
| H <sub>3</sub> CF   | 593-53-3   | -56.0        | -0.2    | 2.2   | -3.1 |
| HF  | 7664-39-3  | -65.3 ± 0.2  | 0.0     | 2.4   | 1.4  |
| C <sub>3</sub> H <sub>3</sub> NO (isoxazole)  | 288-14-2   | 19.6 ± 0.1   | 0.0     | 15.3  | -0.2 |
| H <sub>2</sub> C=CH-CH <sub>2</sub> -CH <sub>3</sub>                                      | 106-98-9   | -0.1 ± 0.2   | 0.0     | 1.4   | 0.2  |
| H <sub>2</sub> C=C(CH <sub>3</sub> ) <sub>2</sub>   | 115-11-7   | -4.3 ± 0.3   | 0.2     | 0.7   | 1.0  |
| C <sub>4</sub> H <sub>4</sub> N <sub>2</sub> (pyridazine)                                 | 289-80-5   | 66.5 ± 0.3   | 0.2     | -10.5 | -1.4 |
| H <sub>3</sub> C-CH <sub>2</sub> -OH  | 64-17-5    | -55.9 ± 0.5  | 0.2     | -0.7  | -1.7 |
| H <sub>2</sub> O  | 7732-18-5  | -57.8 ± 0.01 | 0.2     | 4.3   | 2.4  |
| H <sub>3</sub> C-CH <sub>3</sub>  | 74-84-0    | -20.0 ± 0.1  | 0.2     | 1.7   | -1.4 |
| CH <sub>4</sub>   | 74-82-8    | -17.9        | 0.2     | 4.8   | 0.2  |
| H <sub>3</sub> C-OH   | 67-56-1    | -48.1 ± 3    | 0.2     | -3.6  | -1.0 |
| NF <sub>3</sub>   | 7783-54-2  | -31.6        | 0.2     | 7.2   | -2.6 |
| H <sub>3</sub> C-CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>3</sub>                        | 106-97-8   | -30.0 ± 0.2  | 0.5     | 1.0   | -1.4 |
| H <sub>3</sub> C-NO <sub>2</sub>  | 75-52-5    | -17.9        | 0.5     | 1.9   | -0.5 |
| (CH <sub>3</sub> ) <sub>3</sub> CONO  | 540-80-7   | -41.1 ± 1    | 0.5     | 16.0  | 2.6  |
| HO-N=O  | 7782-77-6  | -18.3        | 0.5     | 4.8   | 0.7  |
| N <sub>2</sub>  | 7727-37-9  | 0.0          | 0.5     | 17.4  | 0.5  |
| F <sub>3</sub> C-NF <sub>2</sub>  | 335-01-3   | -169.0 ± 0.6 | 0.5     | 1.0   | -1.0 |
| NH <sub>3</sub>   | 7664-41-7  | -11.0        | 0.7     | -1.9  | 0.2  |
| C <sub>8</sub> H <sub>6</sub> N <sub>2</sub> (phthalazine)                                | 253-52-1   | 78.8 ± 0.8   | 0.7     | -6.5  | 1.9  |
| H <sub>2</sub> N-C <sub>6</sub> H <sub>4</sub> -NO <sub>2</sub> ( <i>p</i> -nitroaniline) | 100-01-6   | 13.2 ± 0.4   | 0.7     | -2.4  | 0.7  |
| C <sub>4</sub> H <sub>8</sub> N <sub>2</sub> O <sub>3</sub> (4-nitromorpholine)           | 4164-32-3  | -31.3 ± 0.4  | 0.7     | -1.7  | -1.0 |
| H <sub>3</sub> C-ONO <sub>2</sub>   | 598-58-3   | -29.1 ± 0.3  | 0.7     | -3.1  | -1.9 |
| H <sub>3</sub> C-CH <sub>2</sub> -CH <sub>2</sub> -NH <sub>2</sub>                        | 107-10-8   | -16.7 ± 0.2  | 0.7     | 0.0   | -0.2 |
| (CH <sub>3</sub> ) <sub>3</sub> C-NO <sub>2</sub>   | 594-70-7   | -42.3 ± 0.8  | 1.0     | 10.0  | 3.6  |
| H <sub>3</sub> C-COOH   | 64-19-7    | -103.5 ± 0.6 | 1.0     | 1.4   | -0.7 |
| C <sub>2</sub> H <sub>3</sub> N <sub>3</sub> (1,2,4-triazole)                             | 288-88-0   | 46.1 ± 0.2   | 1.0     | 5.7   | -0.2 |
| HOOH  | 7722-84-1  | -32.5        | 1.2     | -8.1  | 0.5  |
| H <sub>3</sub> C-NH <sub>2</sub>  | 74-89-5    | -5.5         | 1.2     | 0.2   | 0.2  |
| HO-NO <sub>2</sub>  | 7697-37-2  | -32.1        | 1.2     | -5.7  | -0.5 |
| F <sub>2</sub>  | 7782-41-4  | 0.0          | 1.4     | -21.5 | -0.0 |
| (CH <sub>3</sub> ) <sub>2</sub> N-NO <sub>2</sub>   | 4164-28-7  | -1.2 ± 0.3   | 1.4     | 2.4   | -1.4 |
| OF <sub>2</sub>   | 7783-41-7  | 5.9          | 1.4     | -10.5 | -2.4 |
| H <sub>3</sub> C-OOH  | 3031-73-0  | -31.3        | 1.7     | -5.7  | -1.0 |
| HN=O  | 14332-28-6 | 23.8         | 1.7     | -9.8  | 4.8  |
| H <sub>2</sub> N-OH   | 7803-49-8  | -11.4        | 1.7     | -1.9  | 0.7  |
| FO-NO <sub>2</sub>  | 7789-26-6  | 2.5          | 1.7     | -8.4  | -2.6 |
| H <sub>2</sub> N-NH <sub>2</sub>  | 302-01-2   | 22.8         | 1.7     | -1.9  | 1.0  |
| C <sub>4</sub> H <sub>4</sub> N <sub>2</sub> (pyrazine)                                   | 290-37-9   | 46.9 ± 0.4   | 1.9     | -7.4  | -0.7 |
| H <sub>2</sub> C=CH-CN  | 107-13-1   | 41.3         | 2.4     | 8.8   | 1.0  |
| C <sub>6</sub> H <sub>10</sub> N <sub>2</sub> O <sub>2</sub> (1-nitropiperidine)          | 7119-94-0  | -10.6 ± 0.6  | 2.9     | 3.1   | 2.2  |
| C <sub>8</sub> H <sub>6</sub> N <sub>2</sub> (quinoxaline)                                | 91-19-0    | 57.4 ± 0.8   | 3.3     | -1.2  | 3.3  |
| HOF   | 14034-79-8 | -23.5        | 3.6     | -5.5  | 2.2  |
| H <sub>2</sub> C=NH   | 2053-29-4  | 16.0 ± 2     | 4.8     | 5.0   | 5.0  |
| FNH <sub>2</sub>  | 15861-05-9 | -11.5 ± 0.6  | 5.7     | 6.7   | 3.8  |
| F <sub>3</sub> C-OF   | 373-91-1   | -182.8 ± 2.4 | 6.5     | -4.3  | 4.8  |

<sup>a</sup> Experimental values taken from the NIST Webbook,<sup>90</sup> except for HN=NH whose formation enthalpy is taken from ref 96. The reader is referred to references therein for further details. Unit: kcal/mol.

possible, a new molecule is obtained by substitution of a hydrogen atom, in such a way that  $\Delta_f H^\circ$  is available in the

NIST Webbook for the new derivative.<sup>90</sup> At this stage, aromatic systems are not represented in the database since

**Table 3.** Same Data As in Table 2 for Cations<sup>a</sup>

| compound  | exptl.  | G3MP2B3 | PM3   | DFT   |
|---|---------|---------|-------|-------|
| C <sub>3</sub> H <sub>5</sub> <sup>+</sup> (cyclopropyl)          | 235.0 a | -6.2    | -2.2  | -9.8  |
| NO <sub>2</sub> <sup>+</sup> (nitrogen dioxide)                   | 233.0 a | -4.8    | -24.4 | 1.9   |
| C <sub>4</sub> H <sub>9</sub> <sup>+</sup> (isobutyl)             | 176.0 a | -3.3    | 2.6   | -7.6  |
| OH <sup>+</sup> (triplet)   | 309.1 b | -2.9    | -19.4 | 6.0   |
| NH <sub>4</sub> <sup>+</sup> (ammonium)                           | 155.0 a | -2.2    | -1.4  | -2.4  |
| CH <sup>+</sup>   | 387.8 b | -2.2    | -21.7 | 9.3   |
| CHO <sup>+</sup> (formyl)   | 199.0 a | -1.4    | -22.0 | 4.5   |
| C <sub>4</sub> H <sub>7</sub> <sup>+</sup> (methyl allyl)         | 207.9 b | -1.4    | 3.8   | -7.4  |
| NO <sup>+</sup> (nitric oxide)                                    | 237.0 a | -1.2    | 1.2   | 8.4   |
| NH <sub>2</sub> <sup>+</sup> (triplet)                            | 302.0 b | -1.2    | -41.8 | 4.3   |
| C <sub>2</sub> H <sub>4</sub> <sup>+</sup> (ethylene)             | 257.0 a | -0.7    | -8.1  | -5.3  |
| C <sub>3</sub> H <sub>5</sub> <sup>+</sup> (propenyl)             | 237.0 a | -0.7    | 1.2   | -4.5  |
| C <sub>4</sub> H <sub>5</sub> O <sup>+</sup> (C-protonated furan) | 165.0 b | -0.5    | 10.3  | -3.3  |
| C <sub>5</sub> H <sub>6</sub> N <sup>+</sup> (pyridinium)         | 178.0 b | 0.0     | 9.1   | -3.1  |
| CH <sub>3</sub> O <sup>+</sup> (H <sub>2</sub> COH)               | 169.3 b | 0.7     | -2.9  | 1.2   |
| CH <sub>3</sub> <sup>+</sup> (methyl)                             | 261.0 a | 1.0     | -4.3  | 3.3   |
| C <sub>2</sub> H <sub>5</sub> <sup>+</sup> (ethyl)                | 216.0 a | 1.2     | 6.5   | 0.0   |
| CH <sub>4</sub> N <sup>+</sup>                                    | 179.4 b | 1.2     | 5.7   | 0.7   |
| C <sub>3</sub> H <sub>3</sub> <sup>+</sup> (cyclopropenyl)        | 257.0 a | 1.2     | 12.7  | 0.2   |
| C <sub>7</sub> H <sub>7</sub> <sup>+</sup> (tropolium)            | 209.0 a | 1.2     | 12.0  | -4.8  |
| HCNH <sup>+</sup>   | 225.8 b | 1.7     | -12.2 | 4.5   |
| OCOH <sup>+</sup>   | 141.0 b | 1.7     | -1.4  | 3.6   |
| CH <sub>3</sub> CO <sup>+</sup>                                   | 156.0 b | 1.9     | 2.9   | 2.9   |
| C <sub>6</sub> H <sub>5</sub> <sup>+</sup> (phenyl)               | 269.3 b | 1.9     | 19.4  | 0.7   |
| C <sub>3</sub> H <sub>7</sub> <sup>+</sup> (1-propyl)             | 211.0 b | 2.2     | 3.3   | -21.5 |
| CH <sub>4</sub> N <sup>+</sup> (methaniminium)                    | 178.0 a | 2.6     | 7.2   | 2.2   |
| CH <sub>3</sub> CNH <sup>+</sup>                                  | 195.0 b | 2.6     | 2.6   | 1.4   |
| C <sub>3</sub> H <sub>5</sub> <sup>+</sup> (allyl)                | 226.0 a | 2.6     | 6.7   | -1.0  |
| HO-CH-OH <sup>+</sup>   | 96.0 b  | 3.1     | -0.7  | 1.7   |
| CH <sub>3</sub> OH <sub>2</sub> <sup>+</sup>                      | 136.0 b | 3.1     | 20.6  | 1.2   |
| C <sub>4</sub> H <sub>9</sub> <sup>+</sup> ( <i>n</i> -butyl)     | 183.0 b | 3.1     | 7.6   | -1.9  |
| C <sub>3</sub> H <sub>7</sub> <sup>+</sup> (2-propyl)             | 190.9 b | 3.3     | 6.2   | -1.4  |
| C <sub>3</sub> H <sub>3</sub> <sup>+</sup> (propynyl)             | 281.0 a | 3.6     | -5.5  | 0.2   |
| CH <sub>3</sub> -OH-CH <sub>3</sub> <sup>+</sup>                  | 130.0 b | 4.3     | 27.0  | 0.5   |
| CH <sub>3</sub> CHOH <sup>+</sup>                                 | 139.0 b | 4.5     | 5.5   | 1.9   |
| C <sub>2</sub> H <sub>3</sub> <sup>+</sup> (vinyl)                | 266.0 a | 5.0     | -1.9  | 5.7   |
| C <sub>4</sub> H <sub>9</sub> <sup>+</sup> ( <i>tert</i> -Butyl)  | 165.8 b | 5.3     | 12.0  | 1.0   |
| C <sub>7</sub> H <sub>7</sub> <sup>+</sup> (benzyl)               | 212.0 a | 5.5     | 15.3  | 1.9   |
| C <sub>5</sub> H <sub>9</sub> <sup>+</sup> (cyclopentyl)          | 188.0 a | 5.7     | 5.3   | 2.4   |
| H <sub>3</sub> O <sup>+</sup> (hydronium)                         | 138.9 a | 6.0     | 20.1  | 7.6   |
| C <sub>4</sub> H <sub>7</sub> <sup>+</sup> (2-butenyl)            | 200.0 a | 6.2     | 11.7  | 0.5   |
| C <sub>6</sub> H <sub>11</sub> <sup>+</sup> (cyclohexyl)          | 177.0 a | 7.2     | 9.1   | 3.1   |
| C <sub>4</sub> H <sub>7</sub> <sup>+</sup> (cyclobutyl)           | 213.0 a | 17.9    | 12.4  | 13.4  |

<sup>a</sup> Sources: *a* = ref 15, *b* = ref 98.

only integer bond orders have been considered. Therefore, additional aromatic and nitro compounds are subsequently included in order to introduce fractional formal bond orders.

Finally, this work relies on a training set of 64 neutral molecules and a validation set of 73 ions, including 43 cations and 30 anions. All of these compounds are listed in Tables 2, 3, and 4. To obtain statistical data involving SCC-DFTB, three compounds are discarded. F<sub>2</sub> is removed from the training set, as SCC-DFTB yields an unrealistic F-F bond length of 1.112 Å, to be compared with the B3LYP/6-31G\* value of 1.404 Å. The OH<sup>+</sup> and NH<sub>2</sub><sup>+</sup> cations are not considered either since the data available correspond to their triplet state, which cannot be handled by the tight-binding formalism at the basis of SCC-DFTB.

## 4. Equilibrium Geometries

**4.1. Assessment against B3LYP/6-31G\* Structures.** The reliability of different levels of theory for the determination of equilibrium geometries is extensively documented in the literature.<sup>91,92</sup> B3LYP/6-31G\* geometries are known to be

**Table 4.** Same Data as in Table 2 for Anions<sup>a</sup>

| compound  | exptl.   | G3MP2B3 | PM3   | DFT  |
|---|----------|---------|-------|------|
| CN <sup>-</sup> (cyanide)   | 17.7 b   | -3.1    | 9.8   | -5.3 |
| C <sub>6</sub> H <sub>5</sub> <sup>-</sup> (phenyl)                 | 54.7 b   | -2.9    | -2.9  | -0.2 |
| C <sub>5</sub> H <sub>5</sub> <sup>-</sup> (cyclopentadienyl)       | 21.3 a   | -2.6    | -5.3  | 2.4  |
| HCO <sub>2</sub> <sup>-</sup> (formate)                             | -110.9 a | -1.7    | 0.0   | -3.8 |
| H <sub>3</sub> C-CH <sub>2</sub> <sup>-</sup>                       | 35.1 b   | -1.4    | -3.3  | 0.7  |
| C <sub>6</sub> H <sub>5</sub> CO <sub>2</sub> <sup>-</sup>          | -97.3 b  | -1.2    | 7.4   | 0.7  |
| C <sub>6</sub> H <sub>5</sub> O <sup>-</sup>                        | -39.4 b  | -1.2    | -4.5  | -0.5 |
| H <sub>3</sub> C-N-CH <sub>3</sub> <sup>-</sup>                     | 26.1 b   | -1.0    | -18.2 | -2.2 |
| C <sub>4</sub> H <sub>4</sub> N <sup>-</sup>                        | 18.9 b   | -1.0    | -7.4  | 1.4  |
| C <sub>5</sub> H <sub>5</sub> <sup>-</sup>                          | 19.6 b   | -0.7    | -3.6  | 4.1  |
| HCC <sup>-</sup>  | 65.5 b   | -0.5    | 10.8  | 0.5  |
| C <sub>6</sub> H <sub>5</sub> O <sup>-</sup> (phenoxy)              | -40.5 a  | 0.0     | -3.6  | -0.5 |
| CHO <sup>-</sup>  | 1.9 b    | 0.0     | -9.3  | 0.0  |
| NH <sub>2</sub> <sup>-</sup>  | 27.0 b   | 0.0     | 11.2  | 3.1  |
| OH <sup>-</sup> (hydroxide)   | -33.2 a  | 0.0     | 15.5  | 2.2  |
| CH <sub>2</sub> CN <sup>-</sup>                                     | 25.1 b   | 0.2     | 3.3   | -3.1 |
| CH <sub>3</sub> NH <sup>-</sup>                                     | 32.0 b   | 0.2     | -10.3 | 0.5  |
| C <sub>2</sub> H <sub>6</sub> N <sup>-</sup> (dimethyl nitrogen)    | 24.7 a   | 0.2     | -16.7 | -0.7 |
| H <sub>2</sub> CCH <sup>-</sup>                                     | 52.8 b   | 0.5     | 8.8   | 2.6  |
| C <sub>2</sub> H <sub>3</sub> O <sub>2</sub> <sup>-</sup> (acetate) | -122.5 a | 0.7     | 2.9   | -0.5 |
| NO <sub>2</sub> <sup>-</sup>  | -45.2 b  | 1.0     | 2.2   | -1.9 |
| CH <sub>2</sub> NO <sub>2</sub> <sup>-</sup>                        | -27.2 b  | 1.2     | -16.0 | -2.2 |
| HOO <sup>-</sup>  | -22.5 b  | 1.4     | -1.4  | -0.7 |
| CH <sub>4</sub> N <sup>-</sup> (methylamine)                        | 30.5 a   | 1.7     | -8.6  | 0.5  |
| C <sub>5</sub> H <sub>11</sub> <sup>-</sup> (neopentyl)             | 3.2 a    | 1.9     | 7.9   | 9.6  |
| NO <sub>3</sub> <sup>-</sup> (nitrate)                              | -74.7 a  | 2.2     | -18.4 | -2.6 |
| CH <sub>3</sub> <sup>-</sup>  | 33.2 b   | 2.4     | 18.2  | 4.8  |
| C <sub>2</sub> H <sub>5</sub> O <sup>-</sup> (ethoxy)               | -47.5 a  | 4.5     | 2.6   | 2.9  |
| CH <sub>3</sub> O <sup>-</sup> (methoxy)                            | -36.0 a  | 4.5     | -1.9  | 3.6  |
| H <sup>-</sup> (hydruere)   | 33.2 b   | 6.2     | 58.6  | 4.8  |

<sup>a</sup> Sources: *a* = ref 15, *b* = Supporting Information from ref 94.**Table 5.** Summary of Root Mean Square Deviations of Bond Lengths and Angles Calculated Using Efficient Methods (AM1 and SCC-DFTB) from Corresponding Values Calculated at the B3LYP/6-31G\* Level

|          | lengths (Å) |          | angles (deg) |          |
|----------|-------------|----------|--------------|----------|
|          | AM1         | SCC-DFTB | AM1          | SCC-DFTB |
| neutrals | 0.015       | 0.025    | 1.327        | 2.103    |
| cations  | 0.043       | 0.046    | 7.855        | 5.720    |
| anions   | 0.016       | 0.039    | 1.606        | 3.422    |

quite accurate. In fact, the composite methods G3B3 and G3MP2B3 rely on B3LYP/6-31G\* structures.<sup>44</sup> However, for large ions or when a large number of structures is to be considered, more efficient methods are of interest. Presently available data suggest SCC-DFTB as the method of choice for fast optimization of molecular geometries, except in the presence of NO bonds for which huge errors are noted.<sup>93,94</sup> However, this conclusion emerges from investigations mainly focused on common CHNO neutrals. Therefore, further assessment of geometries obtained with such methods is of interest, especially for ions.

Experimental gas-phase geometries are clearly not available for most species considered in this work. Therefore, the quality of SCC-DFTB and NDDO structures is assessed against B3LYP/6-31G\* geometries. Among present NDDO methods, the best agreement with B3LYP/6-31G\* geometries is obtained for AM1 structures. Root mean square deviations from reference B3LYP/6-31G\* data for bond lengths and angles are reported in Table 5. Deviations from B3LYP/6-31G\* prove systematically larger for ions than for neutral species, and larger for cations than for anions. This latter

**Table 6.** RMS Deviations from the Experiment of P10-B3LYP/6-311++G\*\* Enthalpies for Neutrals (N), Cations (C), Anions (A), and the Whole Set of Ions (AI)<sup>a</sup>

| geometry     | N               | C   | A   | AI                |
|--------------|-----------------|-----|-----|-------------------|
| B3LYP/6-31G* | 2.1 (−4.8/+6.0) | 5.7 | 3.3 | 4.8 (−21.5/+13.4) |
| AM1          | 2.3 (−4.8/+7.4) | 6.4 | 5.3 | 6.0 (−10.0/+25.3) |
| SCC-DFTB     | 3.0 (−6.7/+6.0) | 5.0 | 4.0 | 4.6 (−9.8/+15.1)  |

<sup>a</sup> In addition, minimum and maximum deviations are given in parentheses for neutrals and ions. All data are in kcal/mol. The performance of SCC-DFTB for cations is overestimated by the fact that cations in triplet states are not considered at this level.

finding might be unexpected, since the minimal valence basis set used at the AM1 level might be expected to be a more significant problem for anions than for cations, owing to the more diffuse character of anionic electron clouds. However, it is understandable if the limitations of AM1 for ions arise primarily as a result of the simplifications regarding the Hamiltonian matrix, rather than as a consequence of the lack of flexibility of the basis set.

SCC-DFTB appears significantly less reliable than AM1 for anions and neutral species, with RMSD values for bond lengths and angles about twice as large. This result is unexpected since AM1 is parametrized primarily against data for neutral compounds. Moreover, it contrasts with previous investigations focused on CHNO compounds.<sup>93,94</sup> However, a detailed examination of present data reveals that the relatively poor performance of SCC-DFTB is due to fluorinated compounds. In particular, C–F bond lengths in F<sub>3</sub>C–NF<sub>2</sub> are overestimated by as much as 0.177 Å. Moreover, while N–O bonds are known to be poorly described by SCC-DFTB,<sup>93,94</sup> it is especially true for the O<sub>2</sub>N–OF bond length which is too long by 0.131 Å.

**4.2. Influence on Energies and Enthalpies.** Total energy increases by up to 6 kcal/mol are observed on substituting B3LYP/6-31G\* geometries with either SCC-DFTB or AM1 structures, with many increases in the range 2.5–5 kcal/mol. This is consistent with the root-mean-square increase of about 5 kcal/mol observed previously on substituting DFT geometries with molecular mechanics structures.<sup>42</sup> Because such energy variations have the same magnitude as typical deviations from experimental values of  $\Delta_f H^\circ$  derived from present AE schemes, as detailed below, applying such schemes to approximate geometries would lead to a dramatic loss of accuracy.

However, it is interesting to note that this is not the case provided that the AEs are specifically optimized for these more approximate geometries. Indeed, Table 6 reports the performance of the P10 procedure applied to B3LYP/6-311++G(2df,2p) energies calculated on geometries optimized at lower levels of theory, using AE specifically optimized for the corresponding structures. Only a moderate increase of the RMSD is observed on going from B3LYP/6-31G\* structures to either AM1 or SCC-DFTB structures. This shows that reoptimizing the AE parameters is an efficient way to make up for systematic errors affecting the underlying geometries. As a result, the use of single-point calculations on AM1 or SCC-DFTB geometries is reasonable, provided that specific AEs are used. On the other hand, as discussed in the sequel, the best results for ions, and more

**Table 7.** RMS Deviations from Experiment of B3LYP Enthalpies Calculated Using the P3, P5, P8, and P10 Correction Schemes, with the 6-311++G(2df,2p) and 6-31++G\*\* Bases<sup>a</sup>

|                            | N               | C   | A   | AI                |
|----------------------------|-----------------|-----|-----|-------------------|
| 6-311++G(2df,2p) basis set |                 |     |     |                   |
| P3                         | 2.2 (−5.5/+7.4) | 6.2 | 3.6 | 5.3 (−22.0/+14.8) |
| P5                         | 2.6 (−4.3/+6.2) | 6.3 | 3.2 | 5.2 (−22.0/+15.1) |
| P8                         | 2.3 (−5.3/+5.7) | 6.0 | 3.2 | 5.0 (−22.2/+13.4) |
| P10                        | 2.2 (−4.5/+5.7) | 5.9 | 3.2 | 4.9 (−22.2/+13.4) |
| 6-31++G** basis set        |                 |     |     |                   |
| P3                         | 2.9 (−4.1/+8.8) | 6.6 | 3.7 | 5.6 (−22.5/+14.3) |
| P5                         | 2.8 (−4.8/+7.9) | 6.7 | 3.5 | 5.6 (−22.5/+14.3) |
| P8                         | 2.7 (−5.5/+6.7) | 6.5 | 3.4 | 5.4 (−22.7/+13.6) |
| P10                        | 2.3 (−4.5/+5.5) | 6.1 | 3.4 | 5.2 (−22.7/+12.7) |

<sup>a</sup> Minimum and maximum deviations for neutrals and ions are also given in parentheses. All values are reported in kcal/mol.

specifically for anions, are obtained with relatively flexible bases including diffuse functions. Therefore, the use of B3LYP/6-31G\* geometries is no dramatic overhead. Unless mentioned otherwise, the following results refer to such structures.

## 5. Comparison of Present Correction Schemes

While many different correction schemes are used in the literature, systematic comparisons of their relative merits are still lacking. Some results suggest that more flexible schemes yield more reliable  $\Delta_f H^\circ$  values.<sup>29</sup> However, with a lack of cross-validation or application of the models to external test sets, it is not possible to determine whether the improvement observed on increasing the flexibility of the empirical correction scheme reflects a true enhancement of the predictive power of the method.

For all theoretical levels employed in the present work to compute total energies, the procedures P3, P5, P8, and P10 have been applied to convert total energies into formation enthalpies. Whatever the specific procedure used, quite similar results are obtained. Nevertheless, the RMSD values reported in Table 7 indicate a small but mostly systematic improvement with the number of adjustable parameters, with the P10 and P8 procedures yielding the smallest RMSD values. On the other hand, no improvement is noted concerning the minimum and maximum deviations from experimental values. Similar observations can be made for other basis sets and functionals. The only exceptions concern the functionals for which the assumption at the basis of the P3 correction scheme breaks down, such as PBE1PBE and to some extent PBE (c.f. section 8). In such cases, P3 naturally leads to very poor predictions. The fact that going from P5 to P10 provides only a few improvements indicates that the P8/P10 definitions for the AEs account only for a small fraction of their environment dependence. Alternatively, one might consider introducing even more specific AEs, or group equivalents, which appear to provide some improvement for neutrals.<sup>41,42</sup> However, their determination requires a training set larger than the present one. Moreover, this approach does not address the fundamental limitations of equivalents associated with their local character.



**Table 8.** RMS Deviations (kcal/mol) from Experimental Results of B3LYP Enthalpies Calculated Using the P5 Procedure and a Variety of Basis Sets, for Neutrals Compounds of the Training Set (N), Cations (C), Anions (A) and All Ions (AI) of the Test Set<sup>a</sup>

|                  | N   | C   | A    | AI   | CPU  |
|------------------|-----|-----|------|------|------|
| 6-31G*           | 4.9 | 7.7 | 26.1 | 17.9 | 273  |
| 6-31G**          | 3.9 | 7.5 | 25.8 | 17.7 | 338  |
| 6-311G*          | 4.0 | 6.5 | 16.7 | 11.9 | 477  |
| 6-31+G*          | 3.6 | 7.3 | 9.7  | 8.4  | 731  |
| 6-31+G**         | 2.9 | 6.7 | 9.6  | 8.1  | 861  |
| 6-31++G**        | 2.8 | 6.7 | 3.5  | 5.6  | 1022 |
| 6-311+G*         | 3.3 | 6.7 | 5.3  | 6.1  | 1211 |
| 6-311+G**        | 2.5 | 6.1 | 4.9  | 5.6  | 1418 |
| 6-311++G**       | 2.5 | 6.0 | 3.4  | 5.1  | 1663 |
| 6-311++G(2df,2p) | 2.6 | 6.3 | 3.2  | 5.2  | 7803 |

<sup>a</sup> The last column provides a rough indication of the relative cost of the different bases as implemented in Gaussian. Each CPU number represents the relative CPU time of a single point calculation for a typically energetic salt, namely, a bicyclic azolium with empirical formula  $C_7H_9N_6O_2$  and no symmetry.<sup>95</sup> An AM1 calculation using Gaussian corresponds roughly to CPU = 1.

## 6. Influence of the Basis Set

It is well-known that relatively extended basis sets including diffuse functions are required for an accurate description of the electronic structure and properties of anions. The present work shows that this conclusion remains valid for  $\Delta_f H^\circ$  after application of the present correction schemes. In other words, AEs are not efficient to make up for basis set deficiencies. This observation is illustrated in Table 8 by the results obtained with the B3LYP functional associated with the P10 procedure.

For neutrals, cations, and anions,  $\Delta_f H^\circ$  predictions steadily improve as more flexible bases are used. For neutrals and cations, the basis set limit appears to be obtained using 6-311+G\*, with 6-31+G\*\* already providing quite good results. For anions, diffuse functions on hydrogen atoms, usually considered to play a marginal role,<sup>92</sup> prove necessary to obtain RMSD values below 3.5 kcal/mol. Increasing the flexibility of the basis beyond 6-311++G\*\* still yields some improvement for anions. However, it is not very significant, since the RMSD decreases only by 0.2 kcal/mol if 6-311++G(2df,2p) is used instead. Therefore, one might prefer the former basis which is about 5 times more efficient.

Not surprisingly, a flexible basis set including diffuse functions is especially mandatory for anions. RMSD values > 15 kcal/mol are obtained otherwise. With the most flexible bases, comparable RMSDs (2.6–3.2 kcal/mol) are obtained for neutrals and anions, while the corresponding value (6.3 kcal/mol) for cations is significantly larger, as discussed in further detail in section 9.

While diffuse functions are especially important for anions, they prove significant for neutral compounds as well, with the RMSD decreasing from 3.9 to 2.9 kcal/mol on going from 6-31G\*\* to 6-31+G\*\*. Not surprisingly, their role is not significant for cations which exhibit more compact electron clouds.

Focusing on neutrals, the deviations from experimental results increase 2-fold on going from the most flexible to the smallest basis. The RMSD of 3.9 kcal/mol obtained using

**Table 9.** RMS Deviations (kcal/mol) from Experimental Results of Enthalpies Calculated Using the P10 Procedure and the 6-31++G\*\* Basis Set<sup>a</sup>

| functional | N   | C    | A    | AI   |
|------------|-----|------|------|------|
| Xalpha     | 3.0 | 22.7 | 20.6 | 21.8 |
| SVWN       | 3.0 | 15.6 | 9.6  | 13.4 |
| BLYP       | 3.4 | 8.6  | 4.8  | 7.3  |
| BP86       | 3.3 | 7.2  | 3.4  | 6.0  |
| PBEPBE     | 3.3 | 7.5  | 4.8  | 6.5  |
| mPWPW91    | 3.3 | 7.3  | 4.1  | 6.2  |
| PW91PW91   | 3.2 | 7.3  | 3.8  | 6.1  |
| HCTH       | 3.3 | 7.0  | 3.8  | 5.9  |
| HF         | 4.8 | 26.1 | 19.2 | 24.2 |
| TPSS       | 3.7 | 7.6  | 5.5  | 6.8  |
| VSXC       | 2.7 | 7.1  | 5.5  | 6.5  |
| BHandH     | 3.0 | 9.8  | 9.8  | 9.8  |
| BHandHLYP  | 2.6 | 6.9  | 7.4  | 7.1  |
| B1LYP      | 2.2 | 7.0  | 5.5  | 6.4  |
| B3P86      | 2.2 | 13.9 | 10.7 | 12.6 |
| PBE1PBE    | 2.1 | 6.0  | 4.8  | 5.5  |
| B3LYP      | 2.3 | 6.1  | 3.4  | 5.2  |
| B3PW91     | 2.2 | 5.8  | 3.9  | 5.1  |
| B98        | 2.2 | 6.1  | 4.3  | 5.4  |
| B1B95      | 2.1 | 6.3  | 6.4  | 6.3  |
| BMK        | 2.8 | 6.4  | 6.6  | 6.5  |

<sup>a</sup> The functionals are listed from top to bottom according to the underlying approximation, starting with LDA and followed by GGA, HF, M-GGA, H-GGA, and HM-GGA.

the popular 6-31G\*\* basis is significantly larger than the value of 2.5 kcal/mol obtained with the most flexible bases. On the other hand, a numerical basis set such as DN\*\* proves superior to a Gaussian basis of similar flexibility, such as 6-31G\*\*, in line with previous results.<sup>20</sup> This finding is not surprising as a numerical basis better spans the space of the s and p atomic orbitals.

Finally, 6-31++G\*\* appears to provide a reasonable trade-off between efficiency and accuracy if anions are to be considered, while 6-31+G\*\* should be sufficient otherwise. Indeed, it yields RMSD values within 0.5 kcal/mol of those obtained using 6-311++G(2df,2p) which is almost 8 times more costly. This good performance is observed for neutral compounds, for cations, and for anions as well.

## 7. Influence of the Functional

A comparison of the performance of various functionals for  $\Delta_f H^\circ$  prediction is provided in Table 9, using the P10 procedure and the 6-31++G\*\* basis set. The best results are obtained using popular H-GGA functionals, especially B3LYP and B3PW91, with RMSD < 5.3 kcal/mol for ions and < 2.4 kcal/mol for neutrals. Although PBE1PBE results are better for neutrals, they appear less reliable for ions. Other H-GGA functionals, especially B3P86, BHandH, and to a lesser extent BHandHLYP do not perform so well. The relatively poor results obtained using BHandHLYP may be attributed to the overestimated contribution of HF exchange in this inappropriately constructed functional. Similarly, the results obtained using the more complicated HM-GGA approach are somewhat less satisfactory. All in all, among the DFT functionals considered in this paper, B3LYP, B3PW91, B98, and PBE1PBE emerge as the most reliable for  $\Delta_f H^\circ$  predictions on ionic systems.



**Table 10.** Atom Equivalents for the P5 Procedure with the 6-31++G\*\* Basis Set (kcal/mol)<sup>a</sup>

| functional | H     | C      | N      | O      | F      | $R^2$ |
|------------|-------|--------|--------|--------|--------|-------|
| Xalpha     | -30.6 | -274.4 | -332.1 | -390.9 | -462.2 | 1.00  |
| SVWN       | 2.1   | -98.1  | -133.2 | -171.2 | -225.7 | 0.98  |
| BLYP       | -2.0  | 33.9   | 42.7   | 50.9   | 51.7   | 1.00  |
| BP86       | 1.4   | 41.4   | 50.0   | 57.1   | 55.0   | 1.00  |
| PBEPBE     | -1.8  | 13.1   | 12.9   | 10.8   | -1.9   | 0.73  |
| mPWPW91    | 0.4   | 39.4   | 46.5   | 52.7   | 50.0   | 1.00  |
| PW91PW91   | -0.4  | 33.7   | 39.2   | 44.1   | 39.6   | 1.00  |
| HCTH       | 5.7   | 35.8   | 39.3   | 41.7   | 37.8   | 0.99  |
| HF         | -10.2 | -107.5 | -145.0 | -183.3 | -203.9 | 0.98  |
| TPSSTPSS   | 1.6   | 46.5   | 53.3   | 59.5   | 58.9   | 1.00  |
| VSXC       | 1.9   | 52.9   | 57.6   | 64.7   | 67.8   | 0.99  |
| BHandH     | -11.0 | -134.5 | -176.1 | -219.5 | -263.6 | 0.99  |
| BHandHLYP  | 1.1   | 27.5   | 26.1   | 25.0   | 29.3   | 0.84  |
| B1LYP      | -0.4  | 30.4   | 34.0   | 37.3   | 39.9   | 0.99  |
| B3P86      | 14.6  | 108.4  | 124.4  | 139.6  | 150.9  | 1.00  |
| PBE1PBE    | 0.2   | 14.1   | 10.0   | 4.4    | -4.8   | 0.02  |
| B3LYP      | 2.7   | 41.5   | 46.7   | 51.4   | 53.8   | 0.99  |
| B3PW91     | 2.7   | 31.8   | 33.2   | 33.4   | 30.4   | 0.95  |
| B98        | 2.1   | 31.5   | 33.6   | 34.5   | 33.2   | 0.96  |
| B1B95      | -1.1  | 32.8   | 36.1   | 38.3   | 40.1   | 0.98  |
| BMK        | 0.2   | 25.8   | 29.7   | 30.4   | 32.2   | 0.98  |

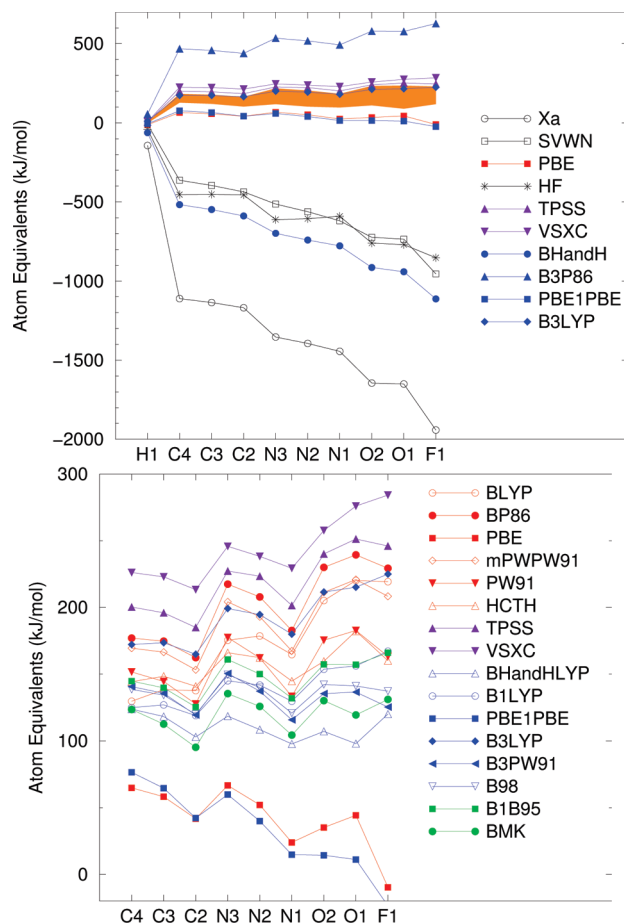
<sup>a</sup> The functionals are listed from top to bottom according to the underlying approximation, starting with LDA and followed by GGA, HF, M-GGA, H-GGA, and HM-GGA. The last column reports the squared correlation coefficients between AE values and atomic numbers for CNOF atoms.

Turning our attention to more efficient models, the best results are obtained within GGA. BP86 and HCTH, in particular, yield quite satisfactory results, comparable to those obtained with the BMK functional, which is the most costly considered in this paper (about 3 times more time-consuming than SVWN using Gaussian). In fact, GGA appears as valuable as HM-GGA as far as ions are concerned, although BLYP proves somewhat less reliable than other GGA functionals. This result contrasts with what is observed for neutral compounds from the training set, where HM-GGA proves more reliable than GGA, especially if the B1B95 functional is used. Besides such general trends, significant conclusions regarding the relative performance of various functionals within the same family can hardly be drawn from present data in view of the small differences between corresponding RMSD data.

## 8. Atom Equivalents

To discuss the main features of atom equivalents as defined in the present paper, values obtained for the P5 procedure, using the 6-31++G\*\* basis set and various functionals, are listed in Table 10. In addition, the variation of environmentally dependent atom equivalents for the P10 procedure, using the same basis set, is illustrated in Figure 1 for the functionals considered in this paper. All AEs derived in this work are statistically well-defined. Indeed, the associated standard deviations derived from a singular value decomposition are systematically  $\leq 1\%$  of their actual values.

As stated in section 2.1, the explicit evaluation of all contributions to  $\Delta_f H^\circ$  allows a straightforward interpretation of these values in terms of additive corrections to  $E_0$ . Their magnitude is much smaller than additive contributions to  $\Delta_f H^\circ - E_0$  often used in practical schemes.<sup>42</sup> On the other hand, it is much larger than additive corrections to theoretical



**Figure 1.** Dependence of P10 atom equivalents for use with 6-31++G\*\* on the energy functional. Color code for the functionals: white = LDA, red = GGA, indigo = M-GGA, blue = H-GGA, and green = HM-GGA. In the upper graph, Xa stands for the X $\alpha$  functional, and the shaded area is where data symbols for most GGA, H-GGA, and HM-GGA functionals are to be found, as detailed in the lower graph.

formation enthalpies derived from molecular reference states.<sup>40</sup> This is understandable as errors on total energies partially cancel out for such reference states.

While the reliability of calculated enthalpies depends mostly on the basis set, AE values exhibit a more significant dependence on the functional. Focusing on 6-31++G\*\* data, they increase according to the following order: X $\alpha$   $\ll$  BHandH < HF  $\approx$  SVWN  $\ll$  PBE  $\leq$  PBE1PBE < most DFT functionals < TPSS < VSXC  $\ll$  B3P86, with negative values for the former four methods. This significant dependence indicates that AE corrections are suitable to make up for systematic errors associated with the functional but not so efficient to correct for basis set effects.

Considering all parameters as a whole, the upper plot in Figure 1 clearly shows that the main difference is between the AE for hydrogen and those for heavier elements. With regard to the latter, their different values turn out to depend on both the number of electrons and the coordination number of the atom. For some models requiring relatively large corrections to  $E_0$ , such as X $\alpha$ , SVWN, or B3P86, the atomic number have the most significant influence on AE values. However, for most functionals, the role of the coordination number is equally significant, as is clear from the bottom

graph in Figure 1. This confirms the actual significance of the slight improvement observed on going from the P5 to the P10 correction scheme.

An interesting feature is the linear correlation observed between AE values and atomic numbers, at the basis of the present P3 procedure. Considering AEs derived for P5 and listed in Table 10, this correlation is usually very high, with squared correlation coefficients  $R^2 \geq 0.99$  in most cases. This explains the similar performances of P3 and P5 procedures for most functionals. However, no such correlation is observed for PBE1PBE, while it is relatively poor ( $R^2 = 0.71$ ) for PBE.

Considering the influence of the atomic environment, it is clear from Figure 1 that, in most cases, AE values increase with the atom coordination number. The reverse dependence is observed only in few cases, for instance, for oxygen using HCTH or VSXC and for nitrogen at the HF level. This indicates that for most DFT functionals, namely those associated with positive equivalents, errors on total energies are less significant for atoms with lower coordination numbers, while the opposite is true for  $X\alpha$ , SVWN, and BHandH, which require negative equivalents.

At the HF level, all atom equivalents are negative, as expected in view of the variational character of the Hartree–Fock theory, which implies overestimated  $E_0$  values. Interestingly, notwithstanding the role of coordination numbers, the corrections to SVWN energies are quite similar. This might appear paradoxical in view of the very different behaviors of HF and LDA approximations, for instance, the fact that HF tends to underestimate bond energies, while the opposite is true for LDA functionals.<sup>91</sup> However, it must be kept in mind that present corrections apply to total energies, not to binding energies. Therefore, the similar magnitudes of HF and LDA equivalents is related to the fact that both theories yield similar errors on molecular total energies, while total energies of isolated atoms are more severely overestimated at the LDA level.

Compared with HF or SVWN functionals, the more approximate  $X\alpha$  method requires corrections roughly twice as negative, indicating that  $E_0$  values are even more overestimated. On the other hand, the fact that  $X\alpha$  bond dissociation energies are not significantly worse than SVWN values<sup>91</sup> indicates that the overestimation of  $E_0$  with respect to SVWN is observed for isolated atoms as well. The last functional for which significantly negative atom equivalents are obtained is BHandH. This functional is unique among advanced DFT functionals with regard to its overestimation of total energies. All other functionals yield essentially positive AEs, i.e., underestimated  $E_0$  values. PBE1PBE and, to a lesser extent, PBE emerge as the functionals requiring the least significant corrections. For energies calculated using the 6-31++G\*\* basis, this is clear considering the AEs listed in Table 10 for P5 corrections and those shown in Figure 1 for P10 corrections. Of course, since these functionals underestimate  $E_0$ , increasing the size of the basis set decreases  $E_0$  and therefore calls for larger corrections.

Because the magnitude of AEs reflects errors of the functionals on calculated *total energies*, while they are more often assessed on the basis of a comparison of *energy*

**Table 11.** Summary of RMS Deviations (kcal/mol) between Theoretical and Experimental Enthalpies for Sets Made of Neutral Compounds (N), Cations (C), Anions (A), and All Ions (AI)

|                      | N    | C    | A    | AI   |
|----------------------|------|------|------|------|
| G3MP2B3              | 1.9  | 4.4  | 2.4  | 3.7  |
| P10–B3LYP/6-311++G** | 2.1  | 5.7  | 3.3  | 4.8  |
| P5–B3LYP/6-311++G**  | 2.5  | 6.0  | 3.4  | 5.1  |
| P10–B3LYP/6-31++G**  | 2.3  | 6.1  | 3.4  | 5.2  |
| P5–B3LYP/6-31++G**   | 2.8  | 6.7  | 3.5  | 5.6  |
| AM1                  | 11.0 | 9.6  | 15.2 | 12.3 |
| PM3                  | 7.0  | 13.3 | 14.2 | 13.7 |
| RM1                  | 12.3 | 25.6 | 26.9 | 26.2 |
| P5-SCC-DFTB          | 14.5 | 14.9 | 28.8 | 16.2 |
| P10-SCC-DFTB         | 11.5 | 11.8 | 31.1 | 22.3 |

differences with experimental data, the present study provides some new insight into the relative performances of different functionals. Present results should also be useful in view of extending available AE schemes to new elements while minimizing the number of adjustable parameters.

## 9. Calculated Enthalpies

This final section examines in more detail the enthalpies calculated using the various procedures considered in this paper. For this purpose, PM3, G3MP2B3, and P5-B3LYP/6-31++G\*\* enthalpies are compared to experimental values in Tables 2, 3, and 4 for neutrals, cations, and anions, respectively. An overview of the relative performance of some of the most interesting AE schemes presently introduced is provided in Table 11, where they are compared to G3MP2B3 and NDDO methods.

**9.1. G3MP2B3 Enthalpies.** In order not to spoil present AE values with the use of spurious data, only compounds with experimental  $\Delta_f H^\circ$  values in reasonable agreement with G3MP2B3 values are included in the training set. As a result, all G3MP2B3 values reported in Table 2 are within 7 kcal/mol from experimental ones, in line with the usual performance of the G3MP2B3 method.<sup>97</sup> In fact, the largest deviations arise for compounds for which the NIST values are reported with significant error bars (up to 2.5 kcal/mol for instance for  $F_3C-OF$ ).

While an accurate description of the electronic structure of anions might be expected to prove more challenging owing to its diffuse character,  $\Delta_f H^\circ$  values for anions are quite satisfactory, with RMSD = 2.4 kcal/mol and all deviations between –6 kcal/mol (for  $HCOO^-$ ) and +6.3 kcal/mol (for  $H^-$ ), as shown in Table 4. However, somewhat larger deviations are observed for cations, as reported in Table 3. In particular, the experimental value found for the cyclobutyl cation is 18 kcal/mol above experimental results. This disagreement might stem from the fact that this cation can easily undergo transition to isomers about 9 kcal/mol lower in energy.<sup>99</sup> In fact, the RMSD for cations, reported as 4.4 kcal/mol in Table 11, drops to 3.5 kcal/mol if the cyclobutyl cation is not considered. This value remains almost 50% larger than the corresponding value for anions. The larger RMSD obtained for cations cannot be explained by the occurrence of triplet species ( $OH^+$  and  $NH_2^+$ ) in the cation data set, as deviations from the experiment for these open

**Table 12.** Atom Equivalents for the P10 Procedure (kcal/mol)

| model for $E_0$                      | H   | C4   | C3   | C2   | N3   | N2   | N1   | O2   | O1   | F    |
|--------------------------------------|-----|------|------|------|------|------|------|------|------|------|
| For Use with B3LYP/6-31G* Geometries |     |      |      |      |      |      |      |      |      |      |
| B3LYP/6-31+G**                       | 2.7 | 41.2 | 41.5 | 39.4 | 47.6 | 46.5 | 43.0 | 50.5 | 51.4 | 53.8 |
| B3LYP/6-31++G**                      | 2.7 | 41.2 | 41.5 | 39.4 | 47.6 | 46.5 | 43.0 | 50.5 | 51.4 | 53.8 |
| B3LYP/6-311++G**                     | 3.3 | 44.3 | 45.3 | 44.7 | 55.3 | 54.1 | 52.2 | 63.6 | 65.0 | 71.3 |
| B3LYP/6-311++G(2df,2p)               | 3.3 | 45.8 | 47.0 | 46.3 | 57.6 | 56.2 | 54.1 | 66.5 | 67.8 | 74.5 |
| For Use with AM1 Geometries          |     |      |      |      |      |      |      |      |      |      |
| B3LYP/6-311++G(2df,2p)               | 3.0 | 45.8 | 47.1 | 46.5 | 56.1 | 55.0 | 54.1 | 64.1 | 67.9 | 73.6 |
| B3LYP/6-31+G**                       | 2.6 | 41.5 | 41.6 | 39.5 | 47.9 | 46.3 | 43.0 | 50.5 | 51.1 | 53.6 |
| For Use with SCC-DFTB Geometries:    |     |      |      |      |      |      |      |      |      |      |
| B3LYP/6-311++G(2df,2p)               | 2.6 | 47.3 | 47.5 | 45.1 | 58.4 | 55.0 | 52.5 | 66.6 | 66.6 | 72.6 |
| B3LYP/6-31+G**                       | 1.8 | 43.2 | 42.2 | 38.0 | 48.9 | 45.3 | 39.9 | 51.0 | 49.7 | 51.7 |

shell systems are not especially large. According to the data in Tables 3 and 4, it arises because of relatively large deviations ( $>5$  kcal/mol) for some cyclic species (cyclopropyl, cyclopentyl, cyclohexyl). All in all, the RMSD increases by ca. 60% on going from neutrals to ions. This increase might stem to some extent from larger uncertainties associated with experimental data for gas-phase ions.

For neutrals, the RMSD between G3MP2B3 and experimental enthalpies can be further decreased by the application of additive corrections. In the present case, bond equivalents<sup>97</sup> do not perform significantly better than P10 corrections, in view of the respective RMSD values of 1.5 and 1.6 kcal/mol. In fact, this specific case of G3MP2B3 calculations for neutrals is the only one where BE corrections are found to provide better results than AE corrections. However, no correction scheme was found to improve G3MP2B3 enthalpies for ions. Therefore, one should consider raw G3MP2B3 values rather than bond-corrected values for studies involving ionic systems.

**9.2. Semiempirical Enthalpies.** PM3 and RM1 were developed in order to overcome the limitations of AM1 to predict  $\Delta_f H^\circ$ . In fact, Table 11 clearly shows that none of these semiempirical methods is suitable for ions. In fact, PM3 and RM1 predictions for ions are even worse than AM1 values. This is especially true for RM1, probably as a result of its more empirical character. Furthermore, although this recent method is reported to yield a significant improvement over previous NDDO schemes with regard to the prediction of formation enthalpies for neutral organic compounds,<sup>15</sup> it proves even worse for neutral compounds in the present training set. This may be explained by the fact that the large data sets employed to fit RM1 and earlier NDDO methods are not representative of the variety of bonding patterns spanned by the present training set. While none of these methods are parametrized for ions, the fact that deviations from experimental results for ions are twice as large with RM1 than with AM1 or PM3 confirms the idea that the superiority of RM1 for standard organic compounds is obtained at the expense of its reliability for less commonly encountered structures. PM3 appears to provide a reasonable trade-off between reliability for common structures and applicability to less common moieties.

All in all, present NDDO results for ions are not significantly better than those obtained using obsolete schemes such as MINDO/3 or MNDO.<sup>98</sup> Indeed, accordingly to the RMSD data in Table 11, typical errors are about 12

kcal/mol, or somewhat lower for neutrals using PM3. They are at best about 4–6 times larger than G3MP2B3 errors. This confirms the potential interest of procedures more reliable than available NDDO approaches for ions, while at the same time being more efficient than G3MP2B3 for complex systems.

**9.3. Enthalpies Derived from Present AE Schemes.** The combination of DFT energies with AEs provides such procedures. For instance, it is clear from Tables 2, 3, and 4 that P10 equivalents applied to B3LYP/6-31++G\*\* energies provide  $\Delta_f H^\circ$  values much more reliable than PM3 values. The improvement is dramatic for small molecules including  $\text{HN}=\text{HN}$ ,  $\text{N}_2$ , and  $\text{F}_2$ ; some medium-size molecules such as isoxazole; and small ions such as  $\text{NO}_2^+$ ,  $\text{OH}^-$  and  $\text{H}^-$ , including some triplet species,  $\text{OH}^+$  and  $\text{NH}_2^+$ . With respect to G3MP2B3, such a DFT/AE approach yields errors typically 20% larger for neutrals and 40% larger for ions, as clear from Table 11. Using 6-311++G\*\* instead of 6-31++G\*\*, the corresponding increases are respectively 10% and 30%. Therefore, notwithstanding the possibility of larger uncertainties associated with data for ions, their enthalpies prove more difficult to predict than values for neutrals with DFT/AE procedures.

The lower accuracy of present DFT/AE approaches compared with G3MP2B3 is consistent with recent findings based on group equivalents optimized for nitro compounds.<sup>18,41</sup> Applied to present ions, the latter yield respectively 6.8 and 5.2 kcal/mol for RMSD data associated with cations and anions. These values are larger than present ones reported in Table 11, as expected from their specialization toward energetic materials. However, these data confirm present conclusions regarding application of DFT/AE to ions, such as the largest deviations observed for cations and the clear superiority of such approaches over NDDO schemes.

Many combinations of the functional/basis set/correction scheme yield very similar results, with the best ones obtained using popular H-GGA functionals: B3LYP, B3PW91, B98, and PBE1PBE, followed by GGA functionals BP86 and HCTH. Whatever the correction scheme employed, SCC-DFTB yields especially poor results, especially for anions. This is probably related to the localized basis set specific to this method.<sup>48</sup> For practical purposes, application of the P10 equivalents listed in Table 12 for B3LYP/6-31++G\*\* or B3LYP/6-311++G\*\* energies appears to be a valuable procedure. P10-B3LYP/6-31++G\*\* is now the default procedure for ions in the MATEO program.<sup>22</sup>



## 10. Conclusion

The present work is the most comprehensive investigation to date of the relative performance of various approaches to estimate the formation enthalpies of ions. In particular, it provides a clear overview of the accuracy to be expected from AE schemes, and new insight into their relative merits.

The most reliable DFT/AE procedures yield a root-mean-square deviation (RMSD) from experimental values < 2.5 kcal/mol for neutrals, but close to 5 kcal/mol for a database of 73 ions. This loss of accuracy is mainly attributed to the fact that present local corrections cannot capture the electron delocalization in ions, associated with the fact that their electron cloud typically involves several canonical structures. Allowing a dependence of AE parameters on the coordination number of the atoms provides a small but systematic improvement. Therefore, such coordination-dependent parameters should preferably be used, unless nonequilibrium structures such as transition states are to be considered.

On the other hand, approximate geometries derived from relatively low theoretical levels, such as AM1 or SCC-DFTB, may be used with no significant loss of accuracy with respect to calculation on B3LYP/6-31G\* geometries provided that the AEs employed are specifically optimized for these more approximate geometries. Although AE corrections prove quite efficient, and have been recently shown to make up for errors associated with the use of small basis sets for many systems,<sup>19,20</sup> present results indicate that this is not really the case for anions, with the associated RMSD steadily increasing as less flexible bases are considered. This is bad news for practical applications. Indeed, unless more sophisticated correction schemes are introduced, this result implies that routine calculations of formation enthalpies for ionic liquids or energetic salts require relatively costly basis sets including diffuse functions in order to get the most from DFT.

Nevertheless, although more efficient procedures are desirable for routine calculations, present DFT/AE procedures fill the gap between costly composite methods and unreliable semiempirical schemes. In particular, the use of B3LYP/6-31++G\*\* energies provides a good trade-off between reliability and cost for anions and cations. In view of predicting the performance of energetic salts, such procedures appear especially suitable. Indeed, as long as the reliability of such predictions depends on the uncertainties associated with the evaluation of lattice energies, using more sophisticated procedures to compute gas phase enthalpies will not necessarily lead to significant improvement.

**Acknowledgment.** The authors thank Dr. M. Elstner for the SCC-DFTB code used in this work. The authors also acknowledge fruitful comments from anonymous reviewers.

**Supporting Information Available:** Molecular files containing B3LYP/6-31G\* optimized geometries in .xyz format. Theoretical G3MP2B3 thermochemical data are included as comments in these files (in atomic units). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Freemantle, M.; Welton, T.; Rogers, R. D. *An Introduction to Ionic Liquids*; Royal Society of Chemistry: London, 2009.
- (2) Guo, Y.; Tao, G.-H.; Joo, Y.-H.; Wang, R.; Twamley, B.; Parrish, D. A.; Shreeve, J. M. *Energy Fuels* **2009**, *23*, 4567–4574.
- (3) Klapoetke, T. M.; Sabate, C. M. Z. *Anorg. Allg. Chem.* **2009**, *635*, 1812–1822.
- (4) Irikura, K. K.; Frurip, D. J. *Computational thermochemistry: prediction and estimation of molecular thermodynamics*; American Chemical Society: Washington, DC, 1998.
- (5) Allinger, N. L.; Yan, L. *J. Am. Chem. Soc.* **1993**, *115*, 11918–11925.
- (6) Mathieu, D.; Simonetti, P. *Mol. Eng.* **1999**, *8*, 121–134.
- (7) Sukhachev, D. V.; Pivina, T. S. *Propellants Explos. Pyrotech.* **1994**, *19*, 159–164.
- (8) Marino, D. J. G.; Peruzzo, P. J.; Krenkel, G.; Castro, E. A. *Chem. Phys. Lett.* **2003**, *369*, 325–334.
- (9) Gutowski, K.; Holbrey, J.; Rogers, R.; Dixon, D. *J. Phys. Chem. B* **2005**, *109*, 23196–23208.
- (10) Gutowski, K.; Rogers, R.; Dixon, D. *J. Phys. Chem. B* **2006**, *110*, 11890–11897.
- (11) Gao, H.; Ye, C.; Piekarski, C. M.; Shreeve, J. M. *J. Phys. Chem. C* **2007**, *111*, 10718–10731.
- (12) Wang, L.; He, Y.-L. *Int. J. Mass Spectrom.* **2008**, *276*, 56–76.
- (13) Curtiss, L.; Raghavachari, K. *Theor. Chem. Acc.* **2002**, *108*, 61–70.
- (14) DeYonker, N. J.; Grimes, T.; Yockel, S.; Dinescu, A.; Mintz, B.; Cundari, T. R.; Wilson, A. K. *J. Chem. Phys.* **2006**, *125*, 104111.
- (15) Rocha, G. B.; Freire, R. O.; Simos, A. M.; Stewart, J. J. P. *J. Comput. Chem.* **2006**, *27*, 1101–1111.
- (16) Repasky, M.; Chandrasekhar, J.; Jorgensen, W. *J. Comput. Chem.* **2002**, *23*, 1601–1622.
- (17) Beaucamp, S.; Bernand-Mantel, A.; Mathieu, D.; Agafonov, V. *Mol. Phys.* **2004**, *102*, 253–258.
- (18) Byrd, E. F. C.; Rice, B. M. *J. Phys. Chem. A* **2009**, *113*, 345–352.
- (19) Brothers, E. N.; Scuseria, G. E. *J. Chem. Theory Comput.* **2006**, *2*, 1045–1049.
- (20) Delley, B. *J. Phys. Chem. A* **2006**, *110*, 13632–13639.
- (21) Beaucamp, S.; Mathieu, D.; Agafonov, V. *J. Phys. Chem. B* **2005**, *109*, 16469–16473.
- (22) Mathieu, D. *J. Hazard. Mater.* **2010**, *176*, 313–322.
- (23) Friesner, R. A.; Knoll, E. H. *J. Chem. Phys.* **2006**, *125*, 124107.
- (24) Schwabe, T.; Grimme, S. *Acc. Chem. Res.* **2008**, *41*, 569–579.
- (25) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1849–1868.
- (26) Brothers, E. N.; Scuseria, G. E. *J. Phys. Chem. A* **2008**, *112*, 13706–13711.
- (27) Cioslowski, J.; Liu, G.; Piskorz, P. *J. Phys. Chem. A* **1998**, *102*, 9890–9900.



- (28) Duchowicz, P.; Castro, E. *Arkivoc* **2001**, 2, 227–241.
- (29) Guthrie, J. J. *Phys. Chem. A* **2001**, 105, 9196–9202.
- (30) Ruzsinszky, A.; Van Alsenoy, C.; Csonka, G. *J. Phys. Chem. A* **2003**, 107, 736–744.
- (31) Ruzsinszky, A.; Csonka, G. *J. Phys. Chem. A* **2003**, 107, 8687–8695.
- (32) Wang, X.; Wong, L.; Hu, L.; Chan, C.; Su, Z.; Chen, G. *J. Phys. Chem. A* **2004**, 108, 8514–8525.
- (33) Long, D. A.; Anderson, J. B. *Chem. Phys. Lett.* **2005**, 402, 524–528.
- (34) Friesner, R. A.; Knoll, E. H.; Cao, Y. *J. Chem. Phys.* **2006**, 125, 124107.
- (35) Yan, G.-K.; Li, J.-J.; Li, B.-R.; Hu, J.; Guo, W.-P. *J. Theor. Comput. Chem.* **2007**, 6, 495–509.
- (36) Goldfeld, D. A.; Bochevarov, A. D.; Friesner, R. A. *J. Chem. Phys.* **2008**, 129, 214105.
- (37) Wu, J.; Xu, X. *J. Chem. Phys.* **2007**, 127, 214105.
- (38) Wu, J.; Xu, X. *J. Comput. Chem.* **2009**, 30, 1424–1444.
- (39) Rice, B.; Pai, S.; Hare, J. *Combust. Flame* **1999**, 118, 445–458.
- (40) Politzer, P.; Ma, Y.; Lane, P.; Concha, M. C. *Int. J. Quantum Chem.* **2005**, 105, 341–347.
- (41) Byrd, E. F. C.; Rice, B. M. *J. Phys. Chem. A* **2006**, 110, 1005–1013.
- (42) Rousseau, E.; Mathieu, D. *J. Comput. Chem.* **2000**, 21, 367–379.
- (43) Ruzsinszky, A.; Kristyan, S.; Margitfalvi, J. L.; Csonka, G. I. *J. Phys. Chem. A* **2003**, 107, 1833–1839.
- (44) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, 110, 7650–7657.
- (45) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, 109, 7764–7776.
- (46) Winget, P.; Clark, T. *J. Comput. Chem.* **2004**, 25, 725–733.
- (47) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; McGraw-Hill: New York, 1989.
- (48) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, 58, 7260–7268.
- (49) Slater, J. C. *The Self-Consistent Field for Molecules and Solids*; McGraw-Hill: New York, 1974; Vol. 4.
- (50) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, 58, 1200–1211.
- (51) Becke, A. D. *Phys. Rev. A* **1988**, 38, 3098–3100.
- (52) Perdew, J. P. *Phys. Rev. B* **1986**, 33, 8822–8824.
- (53) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785–789.
- (54) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, 157, 200–206.
- (55) Perdew, J. P. *Unified Theory of Exchange and Correlation Beyond the Local Density Approximation*; Akademie Verlag: Berlin, Germany, 1991.
- (56) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, 46, 6671–6687.
- (57) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, 108, 664–675.
- (58) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, 77, 3865–3868.
- (59) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, 78, 1396–1396.
- (60) Boese, A. D.; Handy, N. C. *J. Chem. Phys.* **2001**, 114, 5497–503.
- (61) Becke, A. D. *J. Chem. Phys.* **1996**, 104, 1040–1046.
- (62) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, 110, 6158–6159.
- (63) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1997**, 274, 242–250.
- (64) Schmider, H. L.; Becke, A. D. *J. Chem. Phys.* **1998**, 108, 9624–9631.
- (65) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (66) Becke, A. D. *J. Chem. Phys.* **1992**, 98, 1372–1377.
- (67) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, 91, 146401.
- (68) Voorhis, T. V.; Scuseria, G. E. *J. Chem. Phys.* **1998**, 109, 400–410.
- (69) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, 121, 3405–3416.
- (70) Vauthier, E.; Blain, M.; Odier, S.; Barone, V.; Comeau, M.; Fliszar, S. *THEOCHEM* **1995**, 340, 63–70.
- (71) Zope, R. R.; Dunlap, B. I. *J. Chem. Theory Comput.* **2005**, 1, 1193–1200.
- (72) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, 111, 10439–10452.
- (73) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, 120, 215–241.
- (74) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, 107, 3902–3909.
- (75) Stewart, J. J. P. *J. Comput. Chem.* **1989**, 10, 109–220.
- (76) Stewart, J. J. P. *J. Comput. Chem.* **1989**, 10, 221–264.
- (77) mopac7. <http://sourceforge.net/projects/mopac7> (accessed Apr 21, 2010).
- (78) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, 94, 7221–7230.
- (79) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, 106, 1063–1079.

- (80) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374–7383.
- (81) Cioslowski, J.; Schimeczek, M.; Liu, G.; Stoyanov, V. *J. Chem. Phys.* **2000**, *113*, 9377–9389.
- (82) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123*, 124107.
- (83) Stewart, J. J. P. *J. Comput. Phys.* **1989**, *10*, 221.
- (84) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.
- (85) Herndon, W. C. *Chem. Phys. Lett.* **1995**, *234*, 82–86.
- (86) Mole, S.; Zhou, X.; Liu, R. *J. Phys. Chem.* **1996**, *100*, 14665–14671.
- (87) Osmont, A.; Catoire, L.; Goekalp, I. *Energy Fuels* **2008**, *22*, 2241–2257.
- (88) Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K.; Allinger, N. L. *J. Phys. Org. Chem.* **1992**, *5*, 225–229.
- (89) Schmitz, L.; Chen, K.; Labanowski, J.; Allinger, N. *J. Phys. Org. Chem.* **2001**, *14*, 90–96.
- (90) NIST Chemistry Web Book. <http://webbook.nist.gov/chemistry> (accessed Apr 21, 2010).
- (91) Ziegler, T. *Chem. Rev.* **1991**, *91*, 651–667.
- (92) Foresman, J. B.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*; Gaussian, Inc.: Pittsburg, PA, 1993.
- (93) Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. A* **2006**, *110*, 13551–13559.
- (94) Otte, N.; Scholten, M.; Thiel, W. *J. Phys. Chem. A* **2007**, *111*, 5751–5755.
- (95) This is 4-Methyl-1-(1-methyl-4-nitroimidazol-2-yl)-1,2,4-triazolium.
- (96) Biehl, H.; Stuhl, F. *J. Chem. Phys.* **1994**, *100*, 141–145.
- (97) Bharthwaj, B.; Melius, C. F. *J. Phys. Chem. A* **2005**, *109*, 1734–1747.
- (98) Halim, H.; Heinrich, N.; Koch, W.; Schmidt, J.; Frenking, G. *J. Comput. Chem.* **1986**, *7*, 93–104.
- (99) Koch, W.; Liu, B.; Defrees, B. J. *J. Am. Chem. Soc.* **1988**, *110*, 7325–7328.

CT100024R