

Optimal Selection of Partial Charge Calculation Method for Rapid Estimation of Enthalpies of Formation from Hartree–Fock Total Energy

Adrienn Ruzsinszky*,[†] Christian Van Alsenoy,[‡] and Gábor I. Csonka*,[†]

Department of Inorganic Chemistry, Budapest University of Technology and Economics, H-1521 Budapest, Hungary, and Department of Chemistry, University of Antwerp UIA, Universiteitsplein 1, B-2610 Antwerpen, Belgium

Received: September 5, 2002; In Final Form: October 21, 2002

The rapid estimation of basis set error and correlation energy from partial charges (REBECEP) method is improved further in several ways in the current paper. We introduce the total energies derived from experimental enthalpies of formation instead of G3 energies in the fitting procedure. This increases the precision because the error of the G3 method is eliminated. We test the use of B3LYP/6-31G(*d*) equilibrium geometries instead of MP2/6-31G(*d*) geometries. This provides a considerable speed up. We also test the application of stockholder charges in the fitting procedure. New REBECEP parameters were obtained from HF/6-311+G(2*d*,*p*) and HF/6-31G(*d*) energies and atomic charges (natural population analysis, Mulliken, and stockholder). A total of 117 closed shell neutral molecules from the G3/99 database composed of H, C, N, O, and F atoms were selected for the present study. While the correlation between various charges is reasonable for H, C, and N atoms, there is a rather poor correlation between natural population analysis and stockholder charges for O and F atoms. The best REBECEP results were obtained using the HF/6-31G(*d*) natural population analysis charges. The root mean square and the average absolute deviations from the experimental enthalpies of formation for the selected 117 molecules are 2.16 and 1.65 kcal/mol, respectively. This is a considerable improvement as compared to our previous results (root mean square and average absolute deviations were 2.92 and 2.27 kcal/mol, respectively). The results are compared to Gaussian-3 and B3LYP/6-311+G(3*df*,2*p*) enthalpies of formation (the corresponding average absolute deviations for 51 large molecules are 0.94 and 7.09 kcal/mol, respectively). The REBECEP method performs considerably better for the 117 molecules with a moderate 6-31G(*d*) basis set than the B3LYP method with large 6-311+G(3*df*,2*p*) basis set.

1. Introduction

Using very expensive methods (e.g., CCSD(T) method with a very large basis set) the enthalpy of formation for a compound could be calculated reliably by a purely quantum chemical method within chemical accuracy (1–2 kcal/mol).¹ For many large molecules, these calculations are beyond feasibility, so it is of interest to adopt an empirical parametrization that can assist reasonably inexpensive calculations in obtaining the correct answers. (Even the quite expensive Gaussian-3 method² requires empirical “higher-level correction” parameters to reach chemical accuracy). We will not review the literature in this respect here; however, we mention several successful methods that use HF-SCF/6-31G(*d*) total electronic energy results. A bond energy scheme was applied by Allinger³ to HF-SCF/6-31G(*d*) non-relativistic total electronic energies on alcohols and ethers, and it fits the experimental data on 28 compounds to a root mean square (rms) error of 0.52 kcal/mol, while the fit to a group of anomeric compounds is poorer. Wiberg⁴ proposed a method to estimate the enthalpy of formation of any hydrocarbon as the difference of its calculated HF-SCF/6-31G(*d*)/HF-SCF/6-31G(*d*) total electronic energy and the sum of the appropriate empirical energy parameters of $-\text{CH}_3$, $-\text{CH}_2$, $-\text{CH}$, and $-\text{C}$ groups. Later on, an extension of the isodesmic reaction scheme and group equivalents was proposed for “classical” molecules

using 56 atomic energy parameters for H, C, N, O, and F atoms in various molecular environments to correct HF-SCF/6-31G(*d*) total energies leading to average absolute deviations of 2 kcal/mol from experimental enthalpies of formation (the relativistic corrections are included implicitly in the atomic parameters).⁵ It was shown that the number of the atomic parameters can be reduced to 14.⁶ However, for more general use, the definition of the group is not always straightforward (various definitions might lead to some ambiguity, arbitrariness, and a large number of parameters). Anyway, the definition of groups definitely falls outside the HF-SCF/6-31G(*d*) method and requires extra empirical knowledge.

Our goal is to predict the molecular enthalpies of formation of classical molecules within the chemical accuracy using the HF-SCF total energies and the calculated charge distribution without any extra bonding, group, or structural parameters. The origin of the idea of the rapid estimation of the correlation energy is the observation that the correlation energy (the error of the HF-SCF method) is roughly proportional to the number of electrons (*N*) in a system. It is known that the HF-SCF method poorly represents the Coulomb hole correlation (in which electrons of opposite spin are kept apart). This means that the conditional probability density of electron 1 and electron 2 with opposite spin being at the same point in the space is in error. This error is well-studied for atoms and the precise nonrelativistic correlation energies are published.⁷ As is known,

[†] Budapest University of Technology and Economics.

[‡] University of Antwerp UIA.

electrons with parallel spin are correlated via the antisymmetry principle (the exchange term in the HF-SCF method), thus the spin system of the atom influences also the HF-SCF error. This behavior is described in detail in refs 8–10. Beside these two major effects, a slight dependence of the correlation energy on the nuclear charge (Z) can be observed.^{9–11}

Although the N dependence (or more precisely the $N-1$ dependence)¹¹ of the correlation energy provides a quick way to estimate the correlation energy for atoms, the spin dependence and the nuclear frame dependence make it difficult to transfer this knowledge to molecules and obtain a priori energy estimations with the required so-called chemical precision (1–2 kcal/mol). However, the atomic partial charges in a molecule can be used to estimate the number of electrons around the atoms (N_A) and the quasi-linear dependence of correlation energy on N_A provides a way to estimate the atomic correlation energy in the given molecular environment. Thus, the two fundamental assumptions of the proposed method are (i) the correlation energy is the sum of the atomic correlation energies and (ii) the atomic correlation energies can be estimated from the atomic charges.

The validity of the first assumption can be proved by the gradient vector field analysis of the electron density ($\rho(\mathbf{r})$) as the molecules can be cut apart to virial atoms in the real space.¹² If we introduce the $\rho(\mathbf{r})$ from HF-SCF and full-CI calculations and calculate the virial atomic energies respectively, the difference will provide the atomic correlation energies; thus, the existence of such energies can be proved for those molecules for which such partitioning is feasible. Difficulties might arise due to lack of full-CI electron density or domains without nuclei. However, the method is applicable for most of the known organic molecules.

The next question is how to estimate correlation energy from the atomic charges in a molecular environment? At this point, we should recall that atomic partial charges in a molecular environment are essentially mathematical constructions that serve to reflect the electron content around the selected atom of the molecule. These partial charges are not physically measurable quantities, and the partitioning of the molecular electron density can be done either in the Hilbert space spanned by the atomic orbital basis of the molecular wave function or in the real space. The general problem of any population analysis method consists of the partitioning of the molecular electron density and its assignment to the atoms that compose the molecule. However, partial charges can be defined to reproduce the measurable dipole moment and electrostatic potential of the molecule. We note that atomic multipole moments are also required for correct dipole moment predictions (due to non-spherical electron distribution about an atom in the molecule). The primary use of partial charges is to help chemists to establish empirical rules (e.g., polarity of bonds). The partial charges were successfully applied to identify the electron rich (nucleophile) and electron poor (electrophile) functional groups of molecules. Thus, partial charges in the ideal case are able to represent (in a simplified manner) the electron distribution in a molecule.

The most difficult part of the work is to find the connection between the correlation energy and the N_A . These energy factors can be obtained from the study of the exact correlation energy of the atoms in their various ionized states. For example, the C^+ , C , and C^- have -0.13873 , -0.15636 , and -0.18274 au exact correlation energies, respectively.⁷ Although the direct use of these a priori parameters to estimate the correlation energy of carbon atom in a closed shell molecular environment would not yield energy of the required precision (due to different spin

systems),⁹ it hints how the correlation energy changes with the N_A . As the HF-SCF energy and the partial charges depend on the applied basis set, the energy parameters must correct the basis set error too.^{13,14} Naturally, the energy parameters obtained for an infinite basis set (or numerical) HF-SCF result would provide the ultimate solution; however, fitting the required energy parameters (vide infra) to obtain the best results in a least squares sense can be a useful and affordable alternative.

In our previous work,¹⁴ the predictive value of the REBECEP method was tested in the following way: We used the energy parameters obtained for 65 small molecules (e.g., methane, water, ammonia, etc.) of G2-1¹⁵ and G2-2¹⁶ thermochemistry test sets without any change for 51 larger molecules (e.g., naphthalene, octane, aniline) selected from the G3-3 test set.¹⁷ The G2-1 thermochemistry test set (original G2 test set)¹⁵ includes the energies for only very small molecules containing 1–3 heavy atoms (systems such as H_2O , C_2H_6 , and CO_2), whereas the G2-2 test set includes medium-sized molecules containing 3–6 heavy atoms (systems such as C_3H_6 and C_6H_6). The two sets, G2-1 and G2-2, are together referred to as G2/97 and contain 301 test energies.¹⁶ The G3-3 test set contains 75 new enthalpies of formation for molecules that are, on average, larger (containing 3–10 heavy atoms). The largest molecules in the G3-3 test set contain 10 nonhydrogen atoms (naphthalene and azulene). The two sets, G2/97 and G3-3, are together referred to as G3/99 and contain 222 enthalpies of formation. The G3/99 test set contains 47 nonhydrogen-containing molecules, 38 hydrocarbons, 91 substituted hydrocarbons, 15 inorganic hydrides, and 31 open shell radicals. We have selected 116 closed shell neutral molecules from the G3/99 test set composed of H, C, N, O, and F atoms plus urea for testing the performance of the REBECEP method.¹⁴ The best results were obtained using the natural population analysis (NPA).¹⁸ The rms deviation from the experimental enthalpies of formation for 51 molecules selected from the G3-3 test set was 1.15, 3.96, and 2.92 kcal/mol for Gaussian-3 (G3), B3LYP/6-311+G(3df,2p), and REBECEP (NPA) enthalpies of formation, respectively (the corresponding average absolute deviations were 0.94, 7.09, and 2.27 kcal/mol, respectively).¹³ (The G3 method should be clearly distinguished from the test sets, G3/99 or G3-3.) We stress that for these calculations we used the REBECEP parameters obtained from the set of 65 molecules¹³ of the G2/97 test set. The REBECEP method performed considerably better for these 51 test molecules with a moderate 6-31G(d) basis set than the B3LYP method with a large 6-311+G(3df,2p) basis set.¹³

In the earlier papers, we investigated RECEP results obtained with Mulliken and NPA partial atomic charges.^{10,19} The earlier RECEP-3¹⁹ and REBECEP [HF/6-31(d)]¹³ parameters were optimized to reproduce the Gaussian-3 energies. The REBECEP method offers an extremely rapid estimation of a good quality total energy after HF-SCF calculation in the equilibrium molecular geometry. Energy correction parameters correct the HF-SCF energy and basis set errors.^{10,13,19}

To reproduce experimental quality enthalpies of formation, it is preferable to use an accurate molecular geometry. Because the REBECEP method offers an extremely rapid estimation of correlation energy and basis set error, geometry optimization should be as fast and precise as possible. Because in our earlier papers the results were compared to the Gaussian-3 results, for comparison reasons, we used the same geometries optimized with the rather expensive MP2/full/6-31G(d) method. To apply our method, in general, we have to find a faster and reliable geometry optimization method.

In the current paper, we introduce the following changes in the method in order to improve the applicability (make the calculations faster) and increase the precision. (i) We use a total energy derived from experimental enthalpies of formation instead of G3 energies in our fitting procedure. This increases the precision because the error of the G3 method is eliminated. (ii) We employ the use of B3LYP/6-31G(*d*) equilibrium geometries instead of MP2/6-31G(*d*) geometries. This provides a considerable speed up, and the equilibrium geometries are usually closer to the experimental results. (iii) We use the application of stockholder²⁰ and NPA charges in the fitting procedure for a larger set of 117 molecules.

We use a database of 117 closed shell molecules²¹ selected from the G2/97^{15,16} and G3-3¹⁷ molecular geometry database for this study. These molecules contain H, C, N, O, and F atoms. We excluded radicals and molecules with large relativistic effects. The radicals were not included in the study because correlation energy of radicals is different from the correlation energy of closed shell systems.⁹ To obtain stockholder charges, we use the methodology described in ref 22.

2. REBECEP Method

For a given molecule *M*, the so-called REBECEP enthalpy of formation, $\Delta H_f^0(M, \text{REBECEP, expt, basis set, charge def})$, where expt means that the REBECEP parameters were obtained from the fitting to the experimental results using a specific basis set and charge definition (charge def), can be obtained from the following equation:

$$\Delta H_f^0(M, \text{REBECEP, expt, basis set, charge def}) = E_T(M, \text{REBECEP, expt, basis set, charge def}) + E_{ZP}(M, \text{G3}) + E_{\text{therm}}(M, \text{G3}) + \sum_{A=1}^M [\Delta H_f^0(A, \text{expt}) - E_T(A, \text{G3}) - E_{\text{therm}}(A, \text{expt})] \quad (1)$$

where $E_T(M, \text{REBECEP, expt, basis set, charge def})$ is the REBECEP total energy calculated from the parameter set and partial charges using a specific basis set, $E_{ZP}(M, \text{G3})$ is zero-point vibration energy of the molecule *M* (scaled HF/6-31G(*d*) ZPE),²³ and $E_{\text{therm}}(M, \text{G3})$ is the difference between the enthalpy of the molecule at $T = 298.15$ and 0 K (calculated from the molecular heat capacity). The summation runs over all atoms (*A*) of the system: $\Delta H_f^0(A, \text{expt})$ values are the experimental standard enthalpies of formation of the constituent atoms of molecule *M*, $E_T(A, \text{G3})$ values are the G3 total energy of these atoms, and $E_{\text{therm}}(A, \text{expt})$ values are the differences between the enthalpies at $T = 298.15$ and 0 K (calculated from the elemental heat capacities). The actual values of $E_{ZP}(M, \text{G3})$, $E_{\text{therm}}(M, \text{G3})$, and $E_T(A, \text{G3})$ are method-dependent; for the present study, we use G3 values. The so-called “experimental” total energy is obtained from the experimental enthalpy of formation, $\Delta H_f^0(M, \text{expt})$, in the following way:

$$E_T(M, \text{G3, expt}) = \Delta H_f^0(M, \text{expt}) - E_{ZP}(M, \text{G3}) - E_{\text{therm}}(M, \text{G3}) - \sum_{A=1}^M [\Delta H_f^0(A, \text{expt}) - E_T(A, \text{G3}) - E_{\text{therm}}(A, \text{G3})] \quad (2)$$

Using such $E_T(M, \text{G3, expt})$ energies would provide perfect agreement with the experimental $\Delta H_f^0(M, \text{expt})$ value.

The $E_T(M, \text{REBECEP, expt, basis set, charge def})$ total energy can be obtained as a sum of an $E_T(\text{HF-SCF/basis set})$ and the REBECEP energy correction:

$$E_T(M, \text{REBECEP, expt, basis set, charge def}) = E_T(\text{HF-SCF/basis set}) + E_{\text{corr}}(\text{REBECEP, expt, basis set, charge def}) \quad (3)$$

Because $E_T(\text{HF-SCF/basis set})$ contains basis set error, the REBECEP energy correction is thus basis set-dependent. In an ideal case, the $E_T(M, \text{REBECEP, expt, basis set, charge def})$ is basis set-independent; however, small basis set dependence can be observed in practice (vide infra). The basis set dependence of the $E_T(M, \text{REBECEP, expt, basis set, charge def})$ is considerably smaller than the basis set dependence of its components (vide infra). The energy correction is the sum of the atomic corrections:

$$E_{\text{corr}}(\text{REBECEP, expt, basis set, charge def}) \equiv \sum_{A \in M} E_{\text{corr}}(Z_A, N_A, \text{expt, basis set, charge def}) \quad (4)$$

where Z_A is the nuclear charge of atom *A* and N_A is the “electron content” on atom *A*, noninteger, it can be calculated as ($Z_A - \text{partial charge}$). The $E_{\text{corr}}(Z_A, N_A, \text{expt, basis set, charge def})$ atomic energy terms of eq 4 are interpolated:

$$E_{\text{corr}}(Z_A, N_A, \text{expt, basis set, charge def}) = (N_A - N1) E_{\text{par}}(Z_A, N2, \text{expt, basis set, charge def}) + (N2 - N_A) E_{\text{par}}(Z_A, N1, \text{expt, basis set, charge def}) \quad (5)$$

where *N1* and *N2* are integer numbers of electrons, with $N1 \leq N_A \leq N2 = N1 + 1$. $E_{\text{par}}(Z_A, N2, \text{expt, basis set, charge def})$ and $E_{\text{par}}(Z_A, N1, \text{expt, basis set, charge def})$ in eq 5 are the so-called REBECEP atomic parameters that transform the partial charge into energy correction. For hydrogen atoms, we suggest to use a single parameter, $E_{\text{corr}}(1, N_A, \text{expt, basis set, charge def}) = N_A E_{\text{par}}(1, 2, \text{expt, basis set, charge def})/2$. The REBECEP atomic parameters can be obtained from the fitting procedure that finds the minimum of $Y = \sum_{(i=1,L)} [E_T(M, \text{G3, expt})_i - E_T(\text{REBECEP, expt, basis set, charge def})_i]^2$ for a selected set of *L* molecules. The details of the fitting procedure are described elsewhere.^{10,19} Using these parameters in eqs 3–5, $E_T(M, \text{REBECEP, expt, basis set, charge def})$ can be obtained. Introducing $E_T(M, \text{REBECEP, expt, basis set, charge def})$ into eq 1 leads to the desired $\Delta H_f^0(M, \text{REBECEP, expt, basis set, charge def})$.

3. Results and Discussion

3.1. Correlation between Partial Charges. In this section, we analyze the results obtained with the fitted correlation energy parameters to the experimental total energy, the correlation between various partial charges. Figure 1 shows the correlations between NPA and Mulliken vs stockholder partial charges for F, O, N, C, and H atoms calculated with the HF/6-311+G-(2d,p) method. Figure 2 shows correlations between the same type of partial charges calculated with the HF/6-31G(*d*) method. Inspection of the figures reveals that the NPA and Mulliken charges are similar in magnitude; however, the stockholder charges are considerably smaller (half or one-third). The correlation between various partial charges depends on the type of the atom and the basis set.

The correlations between various HF-SCF/6-31G(*d*) charges for hydrogen, carbon, and nitrogen atoms are quite reasonable

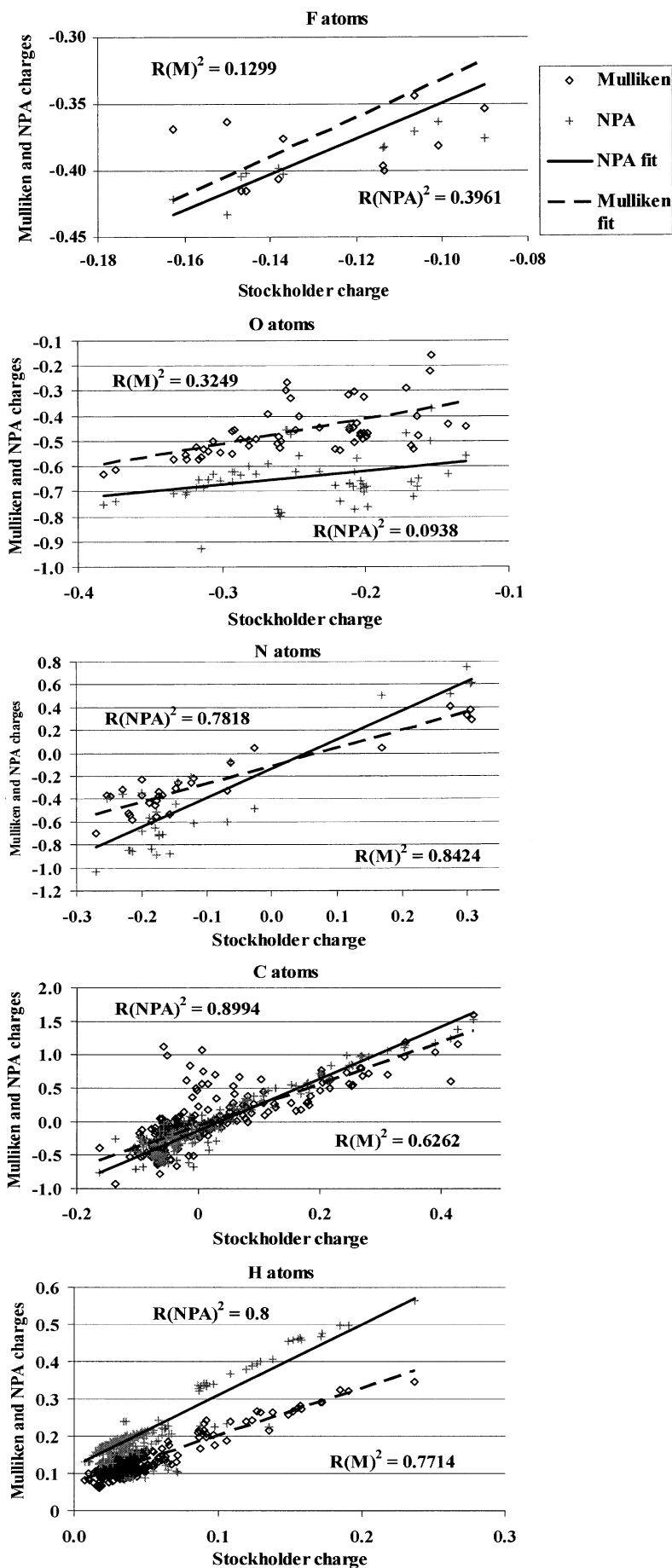


Figure 1. Correlations between NPA and Mulliken vs stockholder partial charges for F, O, N, C, and H atoms calculated with the HF/6-311+G-(2d,p) method.

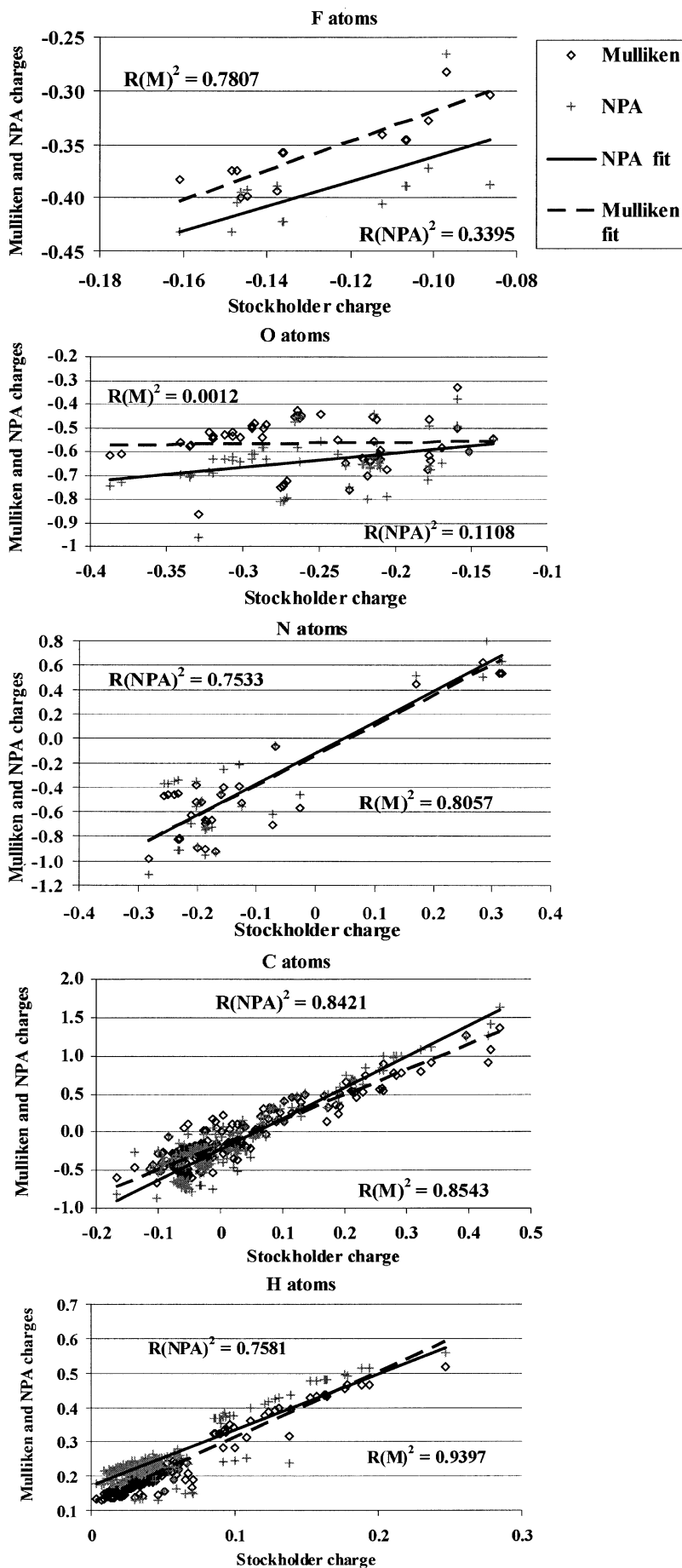


Figure 2. Correlations between NPA and Mulliken vs stockholder partial charges for F, O, N, C, and H atoms calculated with the HF/6-31G(d) method.

TABLE 1: R^2 Values of the Correlations for Mulliken Charges vs Stockholder Charges (Mulliken), NPA Charges vs Stockholder Charges (NPA), and NPA Charges vs Mulliken Charges (NPA-M) for H, C, N, O, and F Atoms of Selected Molecules^a

atoms	HF-SCF/6-31G(d)			HF-SCF/6-311+G(2d,p)		
	Mulliken	NPA	NPA-M	Mulliken	NPA	NPA-M
H	0.940	0.758	0.888	0.771	0.800	0.798
C	0.854	0.842	0.975	0.626	0.899	0.697
N	0.806	0.753	0.975	0.842	0.782	0.870
O	0.001	0.111	0.712	0.325	0.094	0.593
F	0.781	0.340	0.671	0.130	0.396	0.497

^a Calculated with the HF-SCF method using two different basis sets. Partial charges were calculated using the B3LYP/6-31G(d) optimized geometry.

(cf. the R^2 values about 0.75–0.98 in Table 1). For the more electronegative oxygen and fluorine atoms, the NPA-stockholder charge correlations are very poor, practically nonexistent (R^2 : 0.0–0.3, cf. Table 1). The Mulliken-stockholder charge correlation is nonexistent for oxygen atom.

The correlations between various HF-SCF/6-311+G(2d,p) charges for hydrogen, carbon, and nitrogen atoms change slightly (cf. the R^2 values about 0.63–0.90 in Table 1). While the Mulliken-stockholder and NPA-Mulliken charge correlation worsens, the NPA-stockholder charge correlation improves slightly with the increase of the basis set. For oxygen and fluorine atoms, the correlations improve slightly as compared to the smaller basis set; however, they remain very poor (R^2 : 0.1–0.5, cf. Table 1).

3.2. New REBECEP Parameters and Results. Table 2 shows the fitted atomic correction parameters to be used in eq 5. These parameters were obtained from a fitting procedure using the energies of 117 molecules listed in Table 3. Several parameters are missing from Table 2, e.g., E_{par} (6, 4, expt, basis set, stockholder), because no molecule in the database requires these parameters. Comparison of the current REBECEP parameter set with those obtained earlier^{13,19} shows that the parameters are quite stable with respect to the size of the set of molecules. The values of the parameters decrease with the increase of the number of the electrons around the atom in most of the cases. This is expected because the correlation energy is roughly proportional to the number of electrons.⁹ The “normal” behavior of these parameters is encouraging. The only exceptions are the parameters of the nitrogen atom calculated for the 6-311+G(2d,p) basis set (cf. parameters for stockholder charges in Table 2). The stockholder charges calculated with the 6-311+G(2d,p)

basis set are clustered tightly around -0.2 and that might disorient the fitting procedure. It can also be observed that due to the basis set error most of the atomic correction parameters are more negative for the smaller basis set (cf. Table 2). This is necessary because the sum of the E (HF) and E_{corr} values in Table 3 yields the E_{T} (M , G3, expt) of eq 2. The E_{corr} values in Table 3 are systematically more negative for the 6-31G(d) basis set than for the 6-311+G(2d,p) basis set. This can be attributed to the larger basis set error in the smaller basis set. The more negative atomic correction parameters mostly compensate the basis set error; however, this compensation is not perfect.

The values of the new parameter sets in Table 2 show a considerable stability with respect to our earlier parameter sets.¹³ The differences are rather small for C, O, and F atoms (smaller than 0.001 hartrees, usually they are about 0.0005 hartrees). The inclusion of the 51 larger molecules from the G3-3 database has the largest influence on the H and N atomic parameters. The former increases by 0.0067 while the latter decreases by 0.0032 hartrees (the latter is the average of the four energy parameters for N atoms). These small changes improve the current results. We tested the stability of the parameters and the results with respect to different choices of the set of molecules. Our current results confirm the stability observed earlier.¹⁴

Next, we analyze the influence of the choice of the basis set (6-31G(d) or 6-311+G(2d,p)), the choice of the partial charge method (stockholder, Mulliken, or NPA), and the size of the molecules (small or large molecules of the G3/99 test set) on the quality of the calculated REBECEP enthalpies of formation in Table 4. The results calculated using Mulliken charges are not shown in Table 4 to save space. Those results are consistently worse than those calculated with the use of NPA or stockholder charges (cf. statistics in Table 5). Our earlier results show that ΔH_f^0 (M , REBECEP, expt, 6-31G(d), NPA) agrees best with the experimental enthalpies of formation.^{13,19} Inspection of the statistical results in Table 5 yields the same result. Increasing the basis set to 311+G(2d,p) slightly worsens the agreement between the calculated and the experimental results. The results in Table 5 show that for the 6-31G(d) basis set the performance of the REBECEP method decreases in the following order: NPA, stockholder, and Mulliken. A slightly different conclusion can be obtained for the 6-311+G(2d,p) basis set—the use of stockholder charges in the REBECEP procedure provides slightly better results than the NPA charges. We also tested briefly the use of the HF-SCF/6-31G(d) geometries instead of B3LYP/6-31G(d) geometries. The preliminary results

TABLE 2: Fitted Atomic Correction Parameters, $E_{\text{corr}}(N1, Z_A)$ in Hartree for eq 5 to Obtain E_{corr} (REBECEP) for HF-SCF/6-31G(d) or 6-311+G(2d,p) Results and NPA, Stockholder, and Mulliken Charges

atomic no. Z_A	$N1$	basis: 6-31G(d) charge			basis: 6-311+G(2d,p) charge		
		NPA	stockholder	Mulliken	NPA	stockholder	Mulliken
1	2	-0.0382	-0.0418	-0.0359	-0.0382	-0.0404	-0.0383
6	4	-0.1137		-0.1125	-0.1541		-0.1590
6	5	-0.1658	-0.1194	-0.1657	-0.1756	-0.1546	-0.1781
6	6	-0.2207	-0.2197	-0.2198	-0.2119	-0.2119	-0.2118
6	7	-0.2613	-0.2847	-0.2610	-0.2372	-0.2297	-0.2416
6	8						-0.2787
7	6	-0.2698	-0.2694	-0.2734	-0.2718	-0.2919	-0.2949
7	7	-0.2786	-0.2861	-0.2786	-0.2631	-0.2715	-0.2632
7	8	-0.3219	-0.3157	-0.3208	-0.2903	-0.2565	-0.2906
7	9	-0.3725			-0.2438		
8	8	-0.3373	-0.3373	-0.3337	-0.3128	-0.3130	-0.3161
8	9	-0.3759	-0.3880	-0.3742	-0.3321	-0.3206	-0.3217
9	9	-0.3650	-0.3721	-0.3653	-0.3366	-0.3403	-0.3396
9	10	-0.4231	-0.4247	-0.4228	-0.3623	-0.3374	-0.3548

TABLE 3: Species, Number of Atoms and Electrons, the HF-SCF, and Correction Energies (Hartree) for Experimental Quality Total Energy for 117 Selected Molecules

no.	species	no. atoms	no. electrons	6-31G(d)		6-311+G(2d,p)	
				E (HF)	E_{corr}	E (HF)	E_{corr}
1	methane (CH ₄)	5	10	-40.1949	-0.3048	-40.2101	-0.2896
2	ammonia (NH ₃)	4	10	-56.1836	-0.3577	-56.2148	-0.3266
3	water (H ₂ O)	3	10	-76.0098	-0.3933	-76.0527	-0.3503
4	hydrogenfluoride (HF)	2	10	-100.0023	-0.4072	-100.0526	-0.3569
5	acetylene (C ₂ H ₂)	4	14	-76.8168	-0.4863	-76.8436	-0.4596
6	ethylene (H ₂ C=CH ₂)	6	16	-78.0311	-0.5249	-78.0583	-0.4977
7	ethane (H ₃ C-CH ₃)	8	18	-79.2283	-0.5658	-79.2538	-0.5403
8	hydrogencyanide (HCN)	3	14	-92.8735	-0.5177	-92.9018	-0.4894
9	formaldehyde (H ₂ C=O)	4	16	-113.8648	-0.5913	-113.9047	-0.5515
10	methanol (CH ₃ -OH)	6	18	-115.0341	-0.6443	-115.0814	-0.5971
11	hydrazine (H ₂ N-NH ₂)	6	18	-111.1678	-0.6627	-111.2173	-0.6133
12	hydrogenperoxide (HO-OH)	4	18	-150.7613	-0.7309	-150.8250	-0.6673
13	carbon dioxide (CO ₂)	3	22	-187.6310	-0.8787	-187.6923	-0.8174
14	carbon tetrafluoride (CF ₄)	5	42	-435.6415	-1.6820	-435.7779	-1.5456
15	carbonic difluoride (COF ₂)	4	32	-311.6116	-1.2898	-311.7115	-1.1900
16	dinitrogen monoxide (N ₂ O)	3	22	-183.6745	-0.9163	-183.7333	-0.8575
17	nitrogen trifluoride (NF ₃)	4	34	-352.5318	-1.4123	-352.6473	-1.2969
18	ethene, tetrafluoro- (F ₂ C=CF ₂)	6	48	-473.4123	-1.9105	-473.5679	-1.7549
19	acetonitrile, trifluoro- (CF ₃ CN)	6	46	-428.4764	-1.8154	-428.6057	-1.6861
20	propyne (C ₃ H ₄)	7	22	-115.8630	-0.7464	-115.8996	-0.7097
21	allene (C ₃ H ₄)	7	22	-115.8603	-0.7463	-115.8971	-0.7095
22	cyclopropene (C ₃ H ₄)	7	22	-115.8219	-0.7521	-115.8555	-0.7186
23	propylene (C ₃ H ₆)	9	24	-117.0706	-0.7877	-117.1081	-0.7503
24	cyclopropane (C ₃ H ₆)	9	24	-117.0581	-0.7885	-117.0912	-0.7554
25	propane (C ₃ H ₈)	11	26	-118.2630	-0.8291	-118.2990	-0.7932
26	<i>trans</i> -1,3-butadiene (C ₄ H ₆)	10	30	-154.9182	-1.0102	-154.9667	-0.9617
27	dimethylacetylene (C ₄ H ₆)	10	30	-154.9075	-1.0074	-154.9535	-0.9614
28	methylenecyclopropane (C ₄ H ₆)	10	30	-154.8863	-1.0074	-154.9300	-0.9636
29	bicyclobutane (C ₄ H ₆)	10	30	-154.8706	-1.0174	-154.9116	-0.9763
30	cyclobutene (C ₄ H ₆)	10	30	-154.8986	-1.0127	-154.9421	-0.9692
31	cyclobutane (C ₄ H ₈)	12	32	-156.0964	-1.0502	-156.1389	-1.0077
32	isobutene (C ₄ H ₈)	12	32	-156.1097	-1.0525	-156.1574	-1.0048
33	<i>trans</i> -butane (C ₄ H ₁₀)	14	34	-157.2976	-1.0927	-157.3442	-1.0461
34	isobutane (C ₄ H ₁₀)	14	34	-157.2982	-1.0950	-157.3447	-1.0484
35	spiropentane (C ₅ H ₈)	13	38	-193.9166	-1.2736	-193.9667	-1.2236
36	benzene (C ₆ H ₆)	12	42	-230.7019	-1.4476	-230.7631	-1.3864
37	difluoromethane (H ₂ CF ₂)	5	26	-237.8948	-0.9991	-237.9780	-0.9160
38	trifluoromethane (HCF ₃)	5	34	-336.7692	-1.3417	-336.8797	-1.2311
39	methylamine (H ₃ C-NH ₂)	7	18	-95.2089	-0.6144	-95.2470	-0.5762
40	acetonitrile (CH ₃ -CN)	6	22	-131.9255	-0.7775	-131.9639	-0.7391
41	nitromethane (CH ₃ -NO ₂)	7	32	-243.6570	-1.2574	-243.7395	-1.1749
42	methylnitrite (CH ₃ -O-N=O)	7	32	-243.6624	-1.2481	-243.7396	-1.1708
43	formic acid (HCOOH)	5	24	-188.7597	-0.9299	-188.8278	-0.8619
44	methyl formate (HCO-O-CH ₃)	8	32	-227.7867	-1.1844	-227.8599	-1.1111
45	acetamide (CH ₃ -CO-NH ₂)	9	32	-207.9738	-1.1639	-208.0457	-1.0920
46	aziridine (C ₂ H ₄ NH)	8	24	-133.0370	-0.8336	-133.0797	-0.7910
47	cyanogen (NCCN)	4	26	-184.5863	-0.9982	-184.6340	-0.9505
48	dimethylamine ((CH ₃) ₂ NH)	10	26	-134.2378	-0.8736	-134.2831	-0.8283
49	<i>trans</i> -ethylamine (CH ₃ CH ₂ NH ₂)	10	26	-134.2465	-0.8759	-134.2950	-0.8275
50	ketene (CH ₂ CO)	5	22	-151.7228	-0.8135	-151.7725	-0.7638
51	oxirane (C ₂ H ₄ O)	7	24	-152.8655	-0.8607	-152.9138	-0.8124
52	acetaldehyde (CH ₃ CHO)	7	24	-152.9144	-0.8538	-152.9648	-0.8033
53	glyoxal (HCO-COH)	6	30	-226.5891	-1.1397	-226.6620	-1.0669
54	ethanol (CH ₃ CH ₂ OH)	9	26	-154.0743	-0.9071	-154.1320	-0.8494
55	dimethyl ether (CH ₃ OCH ₃)	9	26	-154.0634	-0.8988	-154.1147	-0.8475
56	vinyl fluoride (CH ₂ =CHF)	6	24	-176.8809	-0.8723	-176.9135	-0.8475
57	acrylonitrile (CH ₂ =CHCN)	7	28	-169.7653	-1.0017	-169.8135	-0.9535
58	acetone (CH ₃ COCH ₃)	10	32	-191.9606	-1.1167	-192.0215	-1.0558
59	acetic acid (CH ₃ COOH)	8	32	-227.8079	-1.1919	-227.8868	-1.1130
60	acetyl fluoride (CH ₃ COF)	7	32	-251.7959	-1.2020	-251.8779	-1.1200
61	2-propanol ((CH ₃) ₂ CHOH)	12	34	-193.1138	-1.1714	-193.1819	-1.1034
62	methyl ethyl ether (C ₂ H ₅ OCH ₃)	12	34	-193.1033	-1.1610	-193.1653	-1.0990
63	trimethylamine ((CH ₃) ₃ N)	13	34	-173.2681	-1.1348	-173.3209	-1.0821
64	furan (C ₄ H ₄ O)	9	36	-228.6230	-1.3054	-228.6893	-1.2391
65	pyrrole (C ₄ H ₅ N)	10	36	-208.8061	-1.2801	-208.8677	-1.2185
66	pyridine (C ₅ H ₅ N)	11	42	-246.6940	-1.4843	-246.7599	-1.4184
67	methyl allene (C ₄ H ₆)	10	30	-154.8984	-1.0094	-154.9453	-0.9625
68	isoprene (C ₅ H ₈)	13	38	-193.9557	-1.2758	-194.0145	-1.2169
69	cyclopentane (C ₅ H ₁₀)	15	40	-195.1627	-1.3151	-195.2155	-1.2623
70	<i>n</i> -pentane (C ₅ H ₁₂)	17	42	-196.3321	-1.3566	-196.3892	-1.2994
71	neo-pentane (C ₅ H ₁₂)	17	42	-196.3328	-1.3626	-196.3897	-1.3058

TABLE 3 (Continued)

no.	species	no. atoms	no. electrons	6-31G(d)		6-311+G(2d,p)	
				E (HF)	E_{corr}	E (HF)	E_{corr}
72	1,3-cyclohexadiene (C ₆ H ₈)	14	44	-231.8302	-1.4958	-231.8943	-1.4317
73	1,4-cyclohexadiene (C ₆ H ₈)	14	44	-231.8321	-1.4944	-231.8971	-1.4293
74	cyclohexane (C ₆ H ₁₂)	18	48	-234.2069	-1.5792	-234.2700	-1.5161
75	<i>n</i> -hexane (C ₆ H ₁₄)	20	50	-235.3666	-1.6198	-235.4342	-1.5522
76	3-methyl pentane (C ₆ H ₁₄)	20	50	-235.3631	-1.6250	-235.4306	-1.5574
77	toluene (C ₆ H ₅ CH ₃)	15	50	-269.7387	-1.7134	-269.8101	-1.6420
78	<i>n</i> -heptane (C ₇ H ₁₆)	23	58	-274.4011	-1.8833	-274.4793	-1.8051
79	cyclooctatetraene (C ₈ H ₈)	16	56	-307.5215	-1.9405	-307.6058	-1.8563
80	<i>n</i> -octane (C ₈ H ₁₈)	26	66	-313.4356	-2.1470	-313.5243	-2.0582
81	naphthalene (C ₁₀ H ₈)	18	68	-383.3526	-2.3748	-383.4489	-2.2785
82	acetic acid methyl ester (CH ₃ COOCH ₃)	11	40	-266.8340	-1.4479	-266.9181	-1.3638
83	<i>t</i> -butanol (CH ₃) ₃ COH	15	42	-232.1517	-1.4380	-232.2299	-1.3598
84	aniline (C ₆ H ₅ NH ₂)	14	50	-285.7291	-1.7607	-285.8124	-1.6774
85	phenol (C ₆ H ₅ OH)	13	50	-305.5560	-1.7912	-305.6472	-1.7000
86	divinyl ether (C ₄ H ₆ O)	11	38	-229.7589	-1.3452	-229.8327	-1.2715
87	tetrahydrofuran (C ₄ H ₈ O)	13	40	-230.9748	-1.3833	-231.0431	-1.3150
88	cyclopentanone (C ₅ H ₈ O)	14	46	-268.8645	-1.5997	-268.9417	-1.5224
89	benzoquinone (C ₆ H ₄ O ₂)	12	56	-379.2309	-2.0684	-379.3404	-1.9589
90	urea (NH ₂ -CO-NH ₂)	8	32	-223.9822	-1.2069	-224.0646	-1.1245
91	pyrimidine (C ₄ H ₄ N ₂)	10	42	-262.6910	-1.5174	-262.7621	-1.4464
92	butanedinitrile (N≡C-CH ₂ -CH ₂ -C≡N)	10	42	-262.6861	-1.5159	-262.7580	-1.4440
93	pyrazine (C ₄ H ₄ N ₂)	10	42	-262.6805	-1.5275	-262.7513	-1.4567
94	acetyl acetylene (CH ₃ -CO-C≡CH)	9	36	-228.5861	-1.3017	-228.6555	-1.2323
95	crotonaldehyde (CH ₃ -CH=CH-CHO)	11	38	-229.8016	-1.3362	-229.8727	-1.2650
96	acetic anhydride (CH ₃ -CO-O-CO-CH ₃)	13	54	-379.5790	-1.9961	-379.6956	-1.8795
97	isobutane nitrile ((CH ₃) ₂ CH-C≡N)	12	38	-209.9956	-1.3074	-210.0552	-1.2478
98	methyl ethyl ketone (CH ₃ -CO-CH ₂ -CH ₃)	13	40	-230.9962	-1.3797	-231.0671	-1.3088
99	isobutanal ((CH ₃) ₂ CH-CHO)	13	40	-230.9839	-1.3834	-231.0550	-1.3123
100	1,4-dioxane (C ₄ H ₈ O ₂)	14	48	-305.8227	-1.7153	-305.9163	-1.6217
101	tetrahydropyrrole (C ₄ H ₈ NH)	14	40	-211.1440	-1.3578	-211.2065	-1.2952
102	nitro- <i>s</i> -butane (CH ₃ -CH ₂ -CH(CH ₃)-NO ₂)	16	56	-360.7691	-2.0494	-360.8821	-1.9364
103	diethyl ether (CH ₃ -CH ₂ -O-CH ₂ -CH ₃)	15	42	-232.1432	-1.4248	-232.2158	-1.3521
104	dimethyl acetal (CH ₃ -CH(OCH ₃) ₂)	16	50	-306.9912	-1.7588	-307.0873	-1.6626
105	<i>tert</i> -butylamine ((CH ₃) ₃ C-NH ₂)	16	42	-212.3203	-1.4092	-212.3895	-1.3401
106	<i>N</i> -methyl pyrrole (cyc-CH=CH-N(CH ₃)CH=CH)	13	44	-247.8369	-1.5411	-247.9061	-1.4719
107	tetrahydropyran (C ₅ H ₁₀ O)	16	48	-270.0161	-1.6473	-270.0943	-1.5691
108	diethyl ketone (CH ₃ -CH ₂ -CO-CH ₂ -CH ₃)	16	48	-270.0316	-1.6419	-270.1124	-1.5611
109	isopropyl acetate (CH ₃ -CO-O-CH(CH ₃) ₂)	17	56	-344.9123	-1.9758	-345.0173	-1.8708
110	piperidine (cyc-C ₅ H ₁₀ NH)	17	48	-250.1859	-1.6228	-250.2582	-1.5505
111	<i>tert</i> -butyl methyl ether ((CH ₃) ₃ C-O-CH ₃)	18	50	-271.1737	-1.6952	-271.2563	-1.6126
112	1,3-difluorobenzene (C ₆ H ₄ F ₂)	12	58	-428.4033	-2.1423	-428.5294	-2.0162
113	1,4-difluorobenzene (C ₆ H ₄ F ₂)	12	58	-428.4017	-2.1428	-428.5278	-2.0167
114	fluorobenzene (C ₆ H ₅ F)	12	50	-329.5531	-1.7955	-329.6467	-1.7018
115	di-isopropyl ether ((CH ₃) ₂ CH-O-CH(CH ₃) ₂)	21	58	-310.2167	-1.9558	-310.3098	-1.8627
116	ethane, -hexafluoro- (C ₂ F ₆)	8	66	-672.3791	-2.6383	-672.5913	-2.4262
117	azulene (C ₁₀ H ₈)	18	68	-383.2806	-2.3927	-383.3792	-2.2941

show that the less accurate HF-SCF geometries provide satisfactory results for the 117 molecules: the average absolute deviation from the experiment is 1.68 kcal/mol if HF-SCF/6-31G(d) NPA charges are used (cf. 1.65 kcal/mol for B3LYP geometries in Table 5).

Analysis of the details in Table 4 reveals that for several molecules the agreement between the REBECEP and the experimental enthalpies of formation is exceptionally poor and independent of basis set and charge calculation method. The REBECEP enthalpy of formation for azulene (molecule 117 in Table 4) is in error about from -8 to -10 kcal/mol independent of basis set or partial charge calculation method (the use of Mulliken charges, not shown in Table 4, yields slightly larger deviations). Leaving out the azulene from the fitting procedure improves the agreement between ΔH_f^0 (M , REBECEP, expt, 6-31G(d), NPA) and experimental results considerably, the rms and the average absolute deviations decrease to 2.00 and 1.57 kcal/mol, respectively (cf. 2.16 and 1.65 kcal/mol in Table 5, respectively). We analyzed the problem with azulene and concluded that no REBECEP parametrization using the current charges can provide good results for naphthalene and azulene in the same time. Thus, the specific charge distribution of

azulene requires different parametrization. For hydrocarbons alone (36 molecules) without azulene, we obtained 1.20 kcal/mol average absolute deviation from the experimental results (inclusion of azulene yields 1.37 kcal/mol average absolute deviation).

The next biggest difference, about half the discrepancy observed for azulene, is for carbonic difluoride (COF₂). The deviations are about -4.9 kcal/mol (cf. molecule 15 in Table 4). We note that the G3 and G3SX enthalpies of formation show considerable -3.5 kcal/mol deviation from experiment for this molecule.^{2,24} The experimental error is large for this molecule (cf. 1.4 kcal/mol error in Table 4 and in ref 25). The enthalpy of formation of carbonic difluoride has been deleted recently from the G2/97 test set; however, we kept it for comparison with earlier G3 results.

The results in Table 4 show that considerable errors can be observed for *tert*-butyl-containing molecules (e.g., *tert*-butyl methyl ether, *tert*-butylamine, *tert*-butyl alcohol, and neopentane) and bicyclobutane, C≡N, and C≡C group-containing molecules (acetonitrile, butanedinitrile, cyanogen, and dimethylacetylene). The absolute values of the errors are in the range of 3-6 kcal/mol and show only moderate dependence on the

TABLE 4: Experimental Enthalpies of Formation and Deviations of REBECEP Enthalpies of Formation from Experimental Result (kcal/mol) Calculated with the 6-311G(d) and 6-311+G(2d,p) Basis Sets Using NPA and Stockholder Population Analysis

no.	species	ΔH_f^0 (expt) ^a		deviation (expt – REBECEP)			
		value	error	6-31G(d)		6-311+G(2d,p)	
				NPA	stockholder	NPA	stockholder
1	methane (CH ₄)	-17.9	0.1	0.65	1.91	1.88	1.76
2	ammonia (NH ₃)	-11.0	0.1	0.00	-4.03	0.00	-2.52
3	water (H ₂ O)	-57.8	0.01	-1.47	-2.75	0.53	-0.56
4	hydrogenfluoride (HF)	-65.1	0.2	-1.69	-4.01	1.56	-1.21
5	acetylene (C ₂ H ₂)	54.2	0.2	-0.68	2.31	3.19	2.55
6	ethylene (H ₂ C=CH ₂)	12.5	0.1	0.44	2.76	4.15	4.10
7	ethane (H ₃ C-CH ₃)	-20.1	0.1	1.68	2.63	2.58	2.70
8	hydrogencyanide (HCN)	31.5	1.0	1.70	3.60	3.30	2.96
9	formaldehyde (H ₂ C=O)	-26.0	0.1	-0.49	1.09	2.73	2.56
10	methanol (CH ₃ -OH)	-48.0	0.1	0.64	1.24	2.93	2.81
11	hydrazine (H ₂ N-NH ₂)	22.8	0.2	-1.26	-2.24	-0.05	-1.33
12	hydrogenperoxide (HO-OH)	-32.5	0.04	-1.31	-2.46	-2.05	-3.25
13	carbon dioxide (CO ₂)	-94.1	0.01	-1.97	-3.65	1.64	-0.12
14	carbon tetrafluoride (CF ₄)	-223.0	0.3	3.50	2.88	2.70	0.02
15	carbonic difluoride (COF ₂)	-149.1	1.4	-5.44	-6.24	-3.19	-4.86
16	dinitrogen monoxide (N ₂ O)	19.6	0.1	-4.13	0.84	-2.20	2.96
17	nitrogen trifluoride (NF ₃)	-31.6	0.3	0.41	0.30	1.18	0.43
18	ethene, tetrafluoro- (F ₂ C=CF ₂)	-157.4	0.7	-2.79	-1.34	-0.07	3.52
19	acetonitrile, trifluoro- (CF ₃ CN)	-118.4	0.7	0.59	0.12	-0.68	0.10
20	propyne (C ₃ H ₄)	44.2	0.2	2.28	3.79	4.67	3.85
21	allene (C ₃ H ₄)	45.5	0.3	0.97	2.80	4.00	3.92
22	cyclopropene (C ₃ H ₄)	66.2	0.6	-3.36	-1.56	-1.39	-1.61
23	propylene (C ₃ H ₆)	4.8	0.2	1.20	2.74	3.99	3.84
24	cyclopropane (C ₃ H ₆)	12.7	0.1	1.52	2.46	0.95	0.58
25	propane (C ₃ H ₈)	-25.0	0.1	1.52	2.22	2.09	2.29
26	trans-1,3-butadiene (C ₄ H ₆)	26.3	0.3	-0.25	2.09	4.23	4.01
27	dimethylacetylene (C ₄ H ₆)	34.8	0.3	4.25	4.46		4.16
28	methylenecyclopropane (C ₄ H ₆)	47.9	0.4	2.88	3.89	3.39	2.82
29	bicyclobutane (C ₄ H ₆)	51.9	0.2	-3.66	-2.86	-4.68	-5.12
30	cyclobutene (C ₄ H ₆)	37.4	0.4	-1.45	-1.14	-0.44	-0.57
31	cyclobutane (C ₄ H ₈)	6.8	0.1	1.57	1.19	0.45	0.60
32	isobutene (C ₄ H ₈)	-4.0	0.2	0.52	0.95	2.57	1.96
33	trans-butane (C ₄ H ₁₀)	-30.0	0.2	1.17	1.68	1.54	1.84
34	isobutane (C ₄ H ₁₀)	-32.1	0.2	-0.07	0.04	0.19	0.34
35	spiropentane (C ₅ H ₈)	44.3	0.2	1.09	1.65	-1.49	-2.04
36	benzene (C ₆ H ₆)	19.7	0.2	3.78	3.61	4.05	3.41
37	difluoromethane (H ₂ CF ₂)	-107.7	0.4	-1.14	-0.35	1.28	0.83
38	trifluoromethane (HCF ₃)	-166.6	0.8	0.15	0.80	0.09	-0.21
39	methylamine (H ₃ C-NH ₂)	-5.5	0.1	-0.13	-0.86	1.71	0.21
40	acetonitrile (CH ₃ -CN)	18.0	0.1	3.90	4.17	3.92	3.10
41	nitromethane (CH ₃ -NO ₂)	-17.8	0.1	0.58	0.66	1.30	0.31
42	methylnitrite (CH ₃ -O-N=O)	-15.9	0.2	2.40	1.89	1.25	0.83
43	formic acid (HCOOH)	-90.5	0.1	-0.72	0.60	-0.22	0.29
44	methyl formate (HCOOCH ₃)	-85.0	0.2	0.04	2.55	0.59	1.73
45	acetamide (CH ₃ CONH ₂)	-57.0	0.2	0.34	0.12	-0.17	-0.81
46	aziridine (C ₂ H ₄ NH)	30.2	0.2	-1.20	-0.28	-0.72	-1.20
47	cyanogen (NCCN)	73.3	0.2	-2.92	-3.07	-2.76	-1.25
48	dimethylamine ((CH ₃) ₂ NH)	-4.4	0.2	-1.17	0.78	0.74	1.43
49	trans-ethylamine (CH ₃ CH ₂ NH ₂)	-11.3	0.2	1.33	-0.47	2.34	0.49
50	ketene (CH ₂ CO)	-11.4	0.4	-1.78	-1.33	0.72	0.65
51	oxirane (C ₂ H ₄ O)	-12.6	0.1	0.16	0.89	0.38	0.05
52	acetaldehyde (CH ₃ CHO)	-39.7	0.1	0.75	1.57	2.46	2.22
53	glyoxal (HCOCOH)	-50.7	0.2	1.57	0.50	3.61	2.20
54	ethanol (CH ₃ CH ₂ OH)	-56.2	0.1	1.54	0.58	2.46	2.04
55	dimethyl ether (CH ₃ OCH ₃)	-44.0	0.1	0.85	3.12	3.00	3.88
56	vinyl fluoride (*CH ₂ =CHF)	-33.2	0.4	0.27	2.09		
57	acrylonitrile (CH ₂ =CHCN)	43.2	0.4	0.58	1.77	2.18	1.94
58	acetone (CH ₃ COCH ₃)	-51.9	0.2	1.55	1.65	1.93	1.41
59	acetic acid (CH ₃ COOH)	-103.4	0.4	1.13	1.28	0.07	0.39
60	acetyl fluoride (CH ₃ COF)	-105.7	0.8	-0.31	-0.53	-0.02	-0.64
61	2-propanol (CH ₃) ₂ CHOH)	-65.2	0.1	0.26	-0.77	0.69	0.22
62	methyl ethyl ether (C ₂ H ₅ OCH ₃)	-51.7	0.2	2.03	2.82	2.99	3.64
63	trimethylamine ((CH ₃) ₃ N)	-5.7	0.2	-3.28	1.20	-1.18	1.49
64	furan (C ₄ H ₄ O)	-8.3	0.2	-2.87	-0.48	-3.13	-1.60
65	pyrrole (C ₄ H ₅ N)	25.9	0.1	-2.76	0.09	-2.48	-1.00
66	pyridine (C ₅ H ₅ N)	33.6	0.2	3.19	3.04	2.38	2.60
67	methyl allene (C ₄ H ₆)	38.8	0.1	1.26	2.52	3.58	3.56
68	isoprene (C ₅ H ₈)	18.0	0.3	-1.27	-0.01	2.47	1.92
69	cyclopentane (C ₅ H ₁₀)	-18.3	0.2	0.97	-0.43	-1.08	-0.89

TABLE 4 (Continued)

no.	species	ΔH_f^0 (expt) ^a		deviation (expt - REBECEP)			
		value	error	6-31G(d)		6-311+G(2d,p)	
				NPA	stockholder	NPA	stockholder
70	<i>n</i> -pentane (C ₅ H ₁₂)	-35.1	0.2	0.53	0.88	0.68	1.10
71	neo-pentane (C ₅ H ₁₂)	-40.2	0.2	-2.73	-3.52	-2.93	-3.63
72	1,3-cyclohexadiene (C ₆ H ₈)	25.4	0.2	0.14	-0.65	0.70	0.29
73	1,4-cyclohexadiene (C ₆ H ₈)	25.0	0.1	1.11	0.21	2.20	1.80
74	cyclohexane (C ₆ H ₁₂)	-29.5	0.2	-0.05	-1.60	-2.34	-1.92
75	<i>n</i> -hexane (C ₆ H ₁₄)	-39.9	0.2	0.29	0.47	0.21	0.74
76	3-methyl pentane (C ₆ H ₁₄)	-41.1	0.2	-2.53	-2.67	-2.89	-2.57
77	toluene (C ₆ H ₅ CH ₃)	12.0	0.1	2.26	1.48	1.88	1.04
78	<i>n</i> -heptane (C ₇ H ₁₆)	-44.9	0.3	-0.07	-0.06	-0.39	0.25
79	cyclooctatetraene (C ₈ H ₈)	70.7	0.4	-2.24	-2.67	0.28	-0.29
80	<i>n</i> -octane (C ₈ H ₁₈)	-49.9	0.3	-0.53	-0.70	-1.08	-0.34
81	naphthalene (C ₁₀ H ₈)	35.9	0.4	2.93	1.38	1.47	0.58
82	acetic acid methyl ester (CH ₃ COOCH ₃)	-98.4	0.4	0.58	2.15	-0.18	0.91
83	<i>t</i> -butanol (CH ₃) ₃ COH	-74.7	0.2	-2.78	-3.48	-2.49	-3.12
84	aniline (C ₆ H ₅ NH ₂)	20.8	0.2	0.12	-1.44	0.00	-1.67
85	phenol (C ₆ H ₅ OH)	-23.0	0.2	-0.26	-0.61	-0.71	-0.85
86	divinyl ether (C ₄ H ₆ O)	-3.3	0.2	-3.14	0.34	0.67	2.15
87	tetrahydrofuran (C ₄ H ₈ O)	-44.0	0.2	2.69	0.52	0.22	0.32
88	cyclopentanone (C ₅ H ₈ O)	-45.9	0.4	2.58	1.09	0.04	-0.20
89	benzoquinone (C ₆ H ₄ O ₂)	-29.4	0.8	0.03	-2.02	-0.19	-0.92
90	urea (CH ₄ ON ₂)	-56.3	0.3	1.50	-0.06	0.25	-1.43
91	pyrimidine (C ₄ H ₄ N ₂)	46.8	0.3	2.16	2.13	1.37	1.64
92	butanedinitrile (N≡C-CH ₂ -CH ₂ -C≡N)	50.1	0.2	5.12	3.35	3.78	2.29
93	pyrazine (C ₄ H ₄ N ₂)	46.9	0.3	0.45	0.19	-1.50	0.12
94	acetyl acetylene (CH ₃ -CO-C≡CH)	15.6	0.2	-2.34	-2.38	-0.47	-1.92
95	crotonaldehyde (CH ₃ -CH=CH-CHO)	-24.0	0.3	1.90	2.55	4.14	3.68
96	acetic anhydride (CH ₃ -CO-O-CO-CH ₃)	-136.8	0.4	-0.65	0.52	-3.03	-1.34
97	isobutane nitrile ((CH ₃) ₂ CH-C≡N)	5.6	0.3	1.39	0.31	0.93	-0.43
98	methyl ethyl ketone (CH ₃ -CO-CH ₂ -CH ₃)	-57.1	0.2	1.57	1.42	1.21	0.93
99	isobutanal ((CH ₃) ₂ CH-CHO)	-51.6	0.2	-1.37	-1.02	-0.65	-0.55
100	1,4-dioxane (C ₄ H ₈ O ₂)	-75.5	0.2	4.04	1.10	1.26	1.00
101	tetrahydropyrrrole (C ₄ H ₈ NH)	-0.8	0.2	-0.06	-1.48	-0.94	-0.75
102	nitro- <i>s</i> -butane (CH ₃ -CH ₂ -CH(CH ₃)NO ₂)	-39.1	0.4	0.24	-2.75	-1.94	-3.86
103	diethyl ether (CH ₃ -CH ₂ -O-CH ₂ -CH ₃)	-60.3	0.2	2.15	1.46	1.89	2.32
104	dimethyl acetal (CH ₃ -CH(OCH ₃) ₂)	-93.1	0.2	-1.37	1.23	-2.02	0.12
105	<i>tert</i> -butylamine ((CH ₃) ₃ C-NH ₂)	-28.9	0.2	-3.75	-6.41	-3.43	-5.97
106	<i>N</i> -methyl pyrrole (cyc-CH=CH-N(CH ₃)CH=CH)	24.6	0.1	-5.12	0.19	-4.47	-0.86
107	tetrahydropyran (C ₅ H ₁₀ O)	-15.2	0.2	1.99	-0.41	-1.14	-0.81
108	diethyl ketone (CH ₃ -CH ₂ -CO-CH ₂ -CH ₃)	-61.6	0.2	2.04	1.64	0.99	0.97
109	isopropyl acetate (CH ₃ -CO-O-CH(CH ₃) ₂)	-115.1	0.2	-0.26	-0.44	-2.85	-2.06
110	piperidine (cyc-C ₅ H ₁₀ NH)	-11.3	0.1	-1.13	-2.79	-2.99	-2.80
111	<i>tert</i> -butyl methyl ether ((CH ₃) ₃ C-O-CH ₃)	-67.8	0.3	-4.15	-3.06	-4.11	-3.55
112	1,3-difluorobenzene (C ₆ H ₄ F ₂)	-73.9	0.2	-0.34	-0.11	-0.51	-0.02
113	1,4-difluorobenzene (C ₆ H ₄ F ₂)	-73.3	0.2	0.34	0.53	-0.09	0.36
114	fluorobenzene (C ₆ H ₅ F)	-27.7	0.3	1.78	1.79	1.65	1.52
115	di-isopropyl ether ((CH ₃) ₂ CH-O-CH(CH ₃) ₂)	-76.3	0.4	-2.01	-2.62	-3.38	-3.06
116	ethane, -hexafluoro- (C ₂ F ₆)	-321.3	0.8	1.14	0.53	-1.74	-1.23
117	azulene (C ₁₀ H ₈)	69.1	0.8	-8.14	-10.06	-8.32	-9.72

^a From the G3/99 thermochemistry database; refs 11 and 12.

TABLE 5: Statistics for the Deviation between Calculated REBECEP and Experimental Enthalpies of Formation (kcal/mol) Depending on the Basis Sets [6-31G(d) or 6-311+G(2d,p)] and Partial Charge Calculation Methods [NPA, Stockholder, or Mulliken]

	6-31G(d)			6-311+G(2d,p)		
	NPA	stock	Mulliken	NPA	stock	Mulliken
no. of molecules	117	117	117	115	116	117
rms deviation	2.16	2.34	2.46	2.35	2.35	2.72
avg deviation	0.03	0.18	0.15	0.38	0.30	0.31
avg absolute deviation	1.65	1.79	1.93	1.86	1.80	2.10

selection of the basis set or partial charge calculation method. According to our previous study, the two most problematic molecules were azulene and butanedinitrile (deviations: -10.7 and +6.5 kcal/mol, respectively).¹³ The current REBECEP NPA 6-31G(d) parameter set in Table 2 provides some improvements

(deviations: -8.1 and +5.1 kcal/mol, respectively); however, these molecules remain at the extremes of the error range. Leaving out the three most problematic molecules, the azulene, the butanedinitrile, and the carbonic difluoride would yield 0.10, 1.90, and 1.50 kcal/mol average rms and average absolute deviations, respectively. The rms and the average absolute deviations of the Gaussian-3 and the experimental enthalpies of formation for the selected 117 molecules are 1.14 and 0.95 kcal/mol, respectively. The average absolute deviation of the B3LYP/6-311+G(3df,2p) from the experimental enthalpies of formation for the 51 larger molecules is 7.1 kcal/mol. Comparison of these results shows the advantage of the use of REBECEP method.

For small molecules, using HF/6-31G(d) stockholder charges yields poor REBECEP results, while using NPA charges yields reasonable REBECEP results (cf. NH₃, H₂O, HF, HCN, and CO₂ in Table 4). More than 4 kcal/mol deviations between ΔH_f^0 -

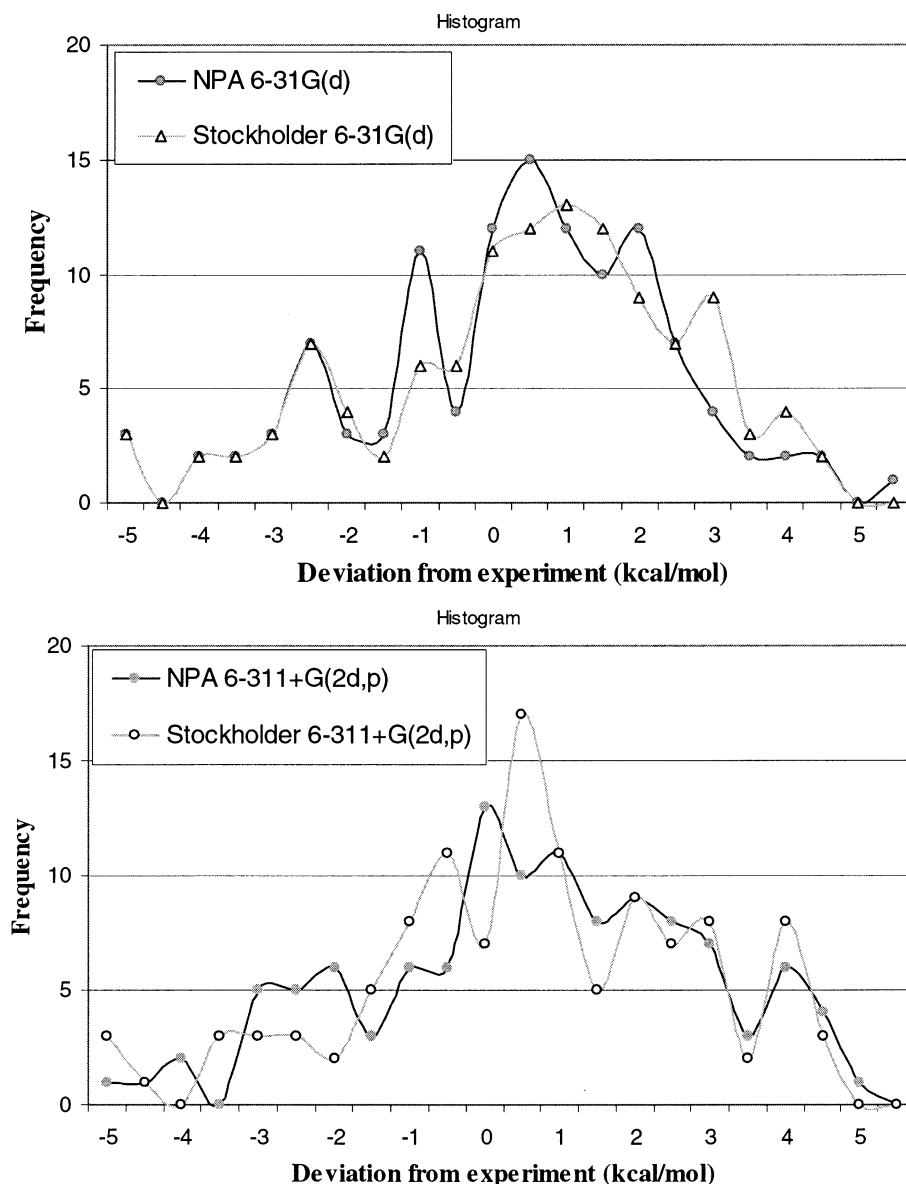


Figure 3. Histograms of REBECEP deviations (expt theory) for a test set of 117 molecules. Each point represents the frequency of a given deviation in a 0.5 kcal/mol range.

(M , REBECEP, expt, 6-31G(d), stockholder) and experimental enthalpies of formation can be observed for ammonia and hydrogen fluoride. Considerable (3–4 kcal/mol) deviations can be observed for hydrogen cyanide and carbon dioxide molecules. These small molecules contain highly electronegative fluorine, oxygen, and nitrogen atoms.

For larger molecules of the G3/99 database, the HF/6-31G(d) stockholder partial charges usually provide good results. However, using the NPA charges causes larger deviations to occur more frequently for such molecules. For example, the REBECEP NPA results for *N*-methyl-pyrrole show a considerable (–5 kcal/mol) deviation, nearly independent of basis set, while the REBECEP stockholder results provide quite good agreement with the experiment (cf. 0.2 kcal/mol for molecule 106 in Table 4).

The $\Delta H_f^0(M, \text{REBECEP, expt, 6-31G}(d), \text{NPA})$ of N_2O , 2-butyne, butanedinitrile, and dioxane deviate by more than 4 kcal/mol from the experimental results. The extent of the problem with dioxane depends on the chosen basis set. Using the larger basis set improves the agreement considerably (cf. molecule 100 in Table 4). A 3–4 kcal/mol deviation can be

observed for cyclopropene, CF_4 , NCCN, bicyclobutane, acetonitrile, benzene, trimethylamine, and divinyl-ether.

The $\Delta H_f^0(M, \text{REBECEP, expt, 6-311+G}(2d,p), \text{stockholder})$ shows reasonable agreement with the experimental results; however, for several unsaturated hydrocarbons (e.g., *trans*-1,3-butadiene, molecule 26 in Table 4), more than 4 kcal/mol deviation occurs. This deviation is basis set-dependent and the REBECEP results calculated with HF/6-31G(d) charges show better agreement with the experimental results.

Figure 3a,b graphically summarize the distributions of the deviations of $\Delta H_f^0(M, \text{REBECEP, expt, basis set, charge def})$ from the experimental results (cf. Table 4) for the basis sets and partial charge calculation methods used in this paper. To show better the properties of the distribution, we connect the frequency values by a curved line. It can be seen that the distribution of errors is not a Gaussian-like distribution. Several peaks occur in Figure 3 according to the various classes of molecules discussed above. Figure 3 shows the superior quality of $\Delta H_f^0(M, \text{REBECEP, expt, 6-31G}(d), \text{NPA})$ and $\Delta H_f^0(M, \text{REBECEP, expt, 6-311+G}(2d,p), \text{stockholder})$ results. For these methods, there is a bigger peak at the middle, 0–1 kcal/mol

range, thus smaller deviations from the experiments are more frequent.

4. Conclusion

We have selected 117 molecules from the G2/97 and G3/99 test set. All of the selected molecules are closed shell neutrals composed of H, C, N, O, and F atoms. New REBECEP atomic energy parameters were obtained using HF/6-311+G(2d,p) and HF/6-31G(d) total energies, B3LYP/6-31G(d) equilibrium geometries, and corresponding NPA, stockholder, and Mulliken atomic charges.

The NPA and Mulliken charges are similar in magnitude while the stockholder charges are considerably smaller (half or one-third). The correlations among the various HF-SCF/6-31G(*d*) charges for hydrogen, carbon, and nitrogen atoms are quite reasonable (R^2 values about 0.75–0.98). For the more electronegative oxygen and fluorine atoms, the NPA–stockholder charge correlations are nonexistent (R^2 : 0.0–0.3). The correlations between various HF-SCF/6-311+G(2d,p) charges for hydrogen, carbon, and nitrogen atoms are a little different. The Mulliken–stockholder and NPA–Mulliken charge correlations worsen, and the NPA–stockholder charge correlations improve slightly with the larger basis set. For oxygen and fluorine atoms, the charge correlations improve slightly, as compared to the smaller basis set; however, they remain very poor (R^2 : 0.1–0.5).

Comparison of the values of the new REBECEP parameter set with the earlier parameter sets shows that these parameters are stable, and increasing the molecular set provides a minor change in the parameters. From the theoretical point of view, it is expected that the increasing electron number would require more negative correction parameters. This condition is fulfilled by most of the parameters. The largest deviation from this rule was found for new parameters derived from HF-SCF/6-311+G(2d,p) stockholder charges.

We compared the various REBECEP enthalpies of formation to the experimental enthalpies of formation of 117 molecules. The performance of the REBECEP method combined with 6-31G(*d*) basis set decreases in the following order: NPA, stockholder, and Mulliken. A slightly different conclusion can be obtained for the 6-311+G(2d,p) basis set—the use of stockholder charges in the REBECEP procedure provides slightly better results than the NPA charges. The rms and the average absolute deviations of ΔH_f^0 (*M*, REBECEP, expt, 6-31G(*d*), NPA) and experimental results for 117 molecules are 2.16 and 1.65 kcal/mol, respectively. Azulene and butanedinitrile are at the extremes of the error range (deviations: –8.1 and +5.1 kcal/mol, respectively). Leaving out the azulene from the fitting procedure decreases the rms and the average absolute deviations to 2.00 and 1.57 kcal/mol, respectively. Leaving out the three most problematic molecules, the azulene, the butanedinitrile, and the carbonic difluoride, would decrease the rms and the average absolute deviations to 1.90 and 1.50 kcal/mol, respectively. For hydrocarbons alone (36 molecules) without azulene, we obtained 1.20 kcal/mol average absolute deviation

from the experimental results. Considerable REBECEP errors can be observed for *tert*-butyl-containing molecules, bicyclobutane, C≡N, and C≡C group containing molecules. These molecules probably require special REBECEP parameters. However, even in the current state, the latest parameters for the REBECEP method show a considerably better performance than the B3LYP/6-311+G(3df,2p) method, and it is much faster (it requires only a HF/6-31G(*d*) calculation). The proposed method does not reach the precision of the Gaussian-3 method; however, it is applicable to large molecules far beyond the reach of that method.

Acknowledgment. We acknowledge W. C. Herndon for reading and correcting the manuscript and J. Olah for the calculation of the stockholder charges used in this study. This work was supported by the OTKA Grant (T 031767 Hungary). A part of this work was realized within the framework of a Bilateral Cooperation Agreement between the Governments of Hungary and the Flemish community of Belgium under Project BIL98/03.

References and Notes

- (1) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014–6029.
- (2) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764; <http://chemistry.anl.gov/compmat/g3energies/g3neut.htm>, <http://chemistry.anl.gov/compmat/g399/g3en.htm>.
- (3) Allinger, N. L.; Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K. *J. Am. Chem. Soc.* **1992**, *114*, 2280–2283.
- (4) Wiberg, K. B. *J. Comput. Chem.* **1984**, *5*, 197–199.
- (5) Ibrahim, M. R.; Schleyer, P. v. R. *J. Comput. Chem.* **1985**, *6*, 157.
- (6) Castro, E. A. *J. Mol. Struct. THEOCHEM* **1994**, *304*, 93–99.
- (7) Davidson, E. R.; Chakravorty, S. J. *J. Phys. Chem.* **1996**, *100*, 6167–6172.
- (8) Chakravorty, S. J.; Clementi, E. *Phys. Rev. A* **1989**, *39*, 2290–2296.
- (9) Kristyan, S.; Csonka, G. I. *Chem. Phys. Lett.* **1999**, *307*, 469.
- (10) Kristyán, S.; Csonka, G. I. *J. Comput. Chem.* **2001**, *22*, 241.
- (11) Kristyán, S. *Chem. Phys.* **1997**, *224*, 33–51.
- (12) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990.
- (13) Kristyán, S.; Ruzsinszky, A.; Csonka, G. I. *Theor. Chem. Acc.* **2001**, *106*, 319.
- (14) Kristyán, S.; Ruzsinszky, A.; Csonka, G. I. *Theor. Chem. Acc.* **2001**, *106*, 404.
- (15) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1995**, *103*, 4192 and references therein.
- (16) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42; <http://chemistry.anl.gov/compmat/g2-97.htm>.
- (17) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374; <http://chemistry.anl.gov/compmat/g3-99.htm>.
- (18) Reed, A. E.; Weinstock, R. B.; Weinhold, F. J. *J. Chem. Phys.* **1985**, *83*, 735.
- (19) Kristyán, S.; Ruzsinszky, A.; Csonka, G. I. *J. Phys. Chem. A* **2001**, *105*, 1926; <http://pubs.acs.org/reprint-request?jp0018192/V6K6>.
- (20) Hirshfeld, F. L. *Theor. Chim. Acta Berl.* **1977**, *44*, 129.
- (21) Csonka, G. I.; Ruzsinszky, A.; Oláh, J.; Van Alsenoy, C. *J. Mol. Struct. THEOCHEM* **2002**, *589–590*, 1.
- (22) Rousseau, B.; Peeters, A.; Van Alsenoy, C. *Chem. Phys. Lett.* **2000**, *324*, 189.
- (23) Curtiss, L. A. <http://chemistry.anl.gov/compmat/g399/zpe.htm>.
- (24) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 108–117.
- (25) CCCBDB, <http://srdata.nist.gov/cccbdb/>.