Theor Chem Acc (2001) 106:319–328
DOI 10.1007/s002140100282

Theoretical
Chemistry Accounts

Regular article

Accurate thermochemistry from corrected Hartree–Fock results: rapid estimation of nearly experimental quality total energy using the small 6-31G(d) basis set

Sándor Kristyán[1], **Adrienn Ruzsinszky**[2], **Gábor I. Csonka**[2]

[1] Chemical Research Center, Institute of Chemistry, Hungarian Academy of Sciences, Pusztaszeri út 59–67, 1025 Budapest, Hungary
[2] Department of Inorganic Chemistry, Budapest University of Technology, 1521 Budapest Hungary

Received: 11 December 2000 / Accepted: 6 February 2001/

Abstract. Gaussian-3 ground-state total electronic energies have been approximated using single point 6-31G(d) basis set Harteee–Fock self-consistent-field (HF-SCF) total energies and partial charges based on our earlier rapid estimation of correlation energy from partial charges method. Sixty-five closed-shell neutral molecules (composed of H, C, N, O, and F atoms) of the G2/97 thermochemistry database were selected for the present study. The main feature in this work is that the basis set error has been treated by the least squares fit of rapid estimation of basis set error and correlation energy from partial charges (REBECEP) atomic parameters. With these parameters a rather accurate closed-shell ground-state electronic total energy can be obtained from a small basis set HF-SCF calculation in the vicinity of stationary points. The average absolute deviation of the best REBECEP enthalpies of formation from the experimental enthalpies of formation is 1.39 kcal/mol for the test set of 65 enthalpies of neutral molecules.

Key words: Ab initio total energy – Basis set error – Correlation energy – Partial charges – Enthalpies of formation

1 Introduction

Accurate techniques for prediction of the thermochemistry of molecules using ab initio calculations are emerging at a rapid pace. The primary goal is to reach the so-called chemical accuracy (±1 kcal/mol) reliably. One successful example is Gaussian-3 (G3) theory [1], a good quality composite technique that approximates the QCISD(T,FU)/G3large level results, making assumptions about the additivity of the basis set and correlation energy corrections. It also includes atomic spin–orbit correction and empirical higher-level correction. First, the G3 theory was assessed for a total of 299 energies (enthalpies of formation, ionization energies, electron affinities, and proton affinities) from the G2/97 test set [2]. The average absolute deviation from experiment of G3 theory for these energies is 1.01 kcal/mol. For the subset of 148 neutral enthalpies of formation the average absolute deviation is 0.94 kcal/mol [1]. The test set was extended recently and is called the G3/99 test set [3]. Actually G3 theory is tested against 376 test energies (222 neutral enthalpies of formation, 88 ionization potentials, 58 electron affinities, and eight proton affinities). Several similar techniques were developed recently [4, 5]. The common aspect in these techniques is that the correlation energy is approximated in the classical way by very expensive methods [e.g., CCSD(T) or QCISD(T)] and empirical corrections are used to reach chemical accuracy. It has been shown that the CCSD(T) total energies extrapolated to an infinite basis set can yield reliable thermochemistry without empirical corrections at a very high cost [6].

Alternatively, density functional theory (DFT) methods use considerably faster algorithms for the estimation of the correlation energy [7, 8]. Popular functionals (B3PW91 and B3LYP) [9–12] usually provide rather good results with moderately large basis sets; however, the reliability of the methods is far from that of the previously mentioned classical methods, and the systematic improvement of the results is not currently available because of the empirical functionals involved. It should also be noted that current implementations of the DFT methods also introduce new problems, for example, numerical instabilities or inaccuracy in numerical integrals.

In our earlier work [13–15] we analyzed the applicability of a radically different approach to calculate the dynamic correlation energy for closed-shell, ground-state and near-equilibrium geometry systems (composed of H, C, N, O, and F atoms) very rapidly and effectively. This procedure is called rapid estimation of correlation energy from partial charges (RECEP). In general the

---

Correspondence to: kristyan@web.inc.bme.hu, csonka@web.inc.bme.hu

correlation energy (E_{corr}) for the electronic ground state is defined [16] as the difference between the exact nonrelativistic complete configuration interactions (CI) electronic ground-state basis set limit total energy, E_{T}(CI), and the single-determinant electronic ground-state Hartree–Fock self-consistent-field (HF-SCF) basis set limit total energy, E_{T}(HF-SCF), of a system:

$$E_{\mathrm{corr}} = E_{\mathrm{T}}(\mathrm{CI}) - E_{\mathrm{T}}(\mathrm{HF\text{-}SCF}) \ . \tag{1}$$

Because the calculation of E_{T}(CI) is currently not feasible for most molecules (i.e., the number of electrons must be less than, say, about 20), in our earlier work [15] we defined the total molecular correlation energy at the G3 level of theory as the difference of the G3 and the HF-SCF/6-311 + G(2d,p) total energies:

$$E_{\mathrm{corr}}(\mathrm{G3}) = E_{\mathrm{T}}(\mathrm{G3}) - E_{\mathrm{T}}[\mathrm{HF\text{-}SCF/6\text{-}311 + G(2d,p)}] \ . \tag{2}$$

Our earlier RECEP-3 method [15] requires only a single point HF-SCF/6-311 + G(2d,p) calculation in the MP2(FU)/6-31G(d) equilibrium geometry. The quality of E_{corr}(G3) was compared to correlation energy obtained as the difference of the CCSD(T) and HF-SCF total energies extrapolated to an infinite basis set in our previous work [15]. The question of the geometry optimization was not addressed in our previous articles [14, 15]. The MP2(FU)/6-31G(d) equilibrium geometry was used for the sake of consistency, because we approximated the G3 total energies and these energies were calculated at this geometry.

In the present article the use of a considerably smaller, namely 6-31G(d), basis set is analyzed. This smaller basis set increases the speed of HF-SCF calculations considerably. For example the HF-SCF/6-31G(d) calculation of phenol and similar-sized molecules is more than 1 order of magnitude faster than the corresponding HF-SCF/6-311 + G(2d,p) calculation. The increase in speed was measured with Gaussian 98 [17]. This program was used for all the calculations described in this article. This increase in speed is critical for larger molecules and extends considerably the use of RECEP-type methods. We note that the 6-311+G(2d,p) basis set has some inconvenient features (vide infra); thus, even the G3 method uses basis sets derived from the uniformly defined 6-31G(d) basis set [1]; however, the use of such a small basis set introduces considerable basis set error into the total energy. After a brief summary of the RECEP method and the earlier results we discuss a method to obtain experimental total energy and suggest a new method to handle the basis set error. Finally we present the results obtained with the new parameter set that handles basis set error and correlation energy.

2 The RECEP method

The RECEP formula [13–15] for the E_{corr} value of ground-state covalent molecules in the vicinity of stationary points is as follows:

$$E_{\mathrm{corr}}(\mathrm{RECEP, G3, charge\ def.}) \equiv \sum_{A=1}^{M} E_{\mathrm{corr}}(N_{\mathrm{A}}, Z_{\mathrm{A}}, \mathrm{G3, charge\ def.}) \ , \tag{3}$$

where E_{corr}(RECEP, G3, charge def.) is the RECEP molecular correlation energy that approximates E_{corr}(G3) in Eq. (2) using one of the various available partial charges calculated at the HF-SCF/6-311+G(2d,p) level of theory [15]. (We chose the G3 method, but any other good-quality method can be used for this purpose.) We tested [14, 15] four partial charge definitions: the natural population analysis (NPA) [18], ChelpG [19], Merz–Kollman (MK) [20], and Mulliken. In $E_{\mathrm{corr}}(N_{\mathrm{A}}, Z_{\mathrm{A}}$, G3,charge def.), $\mathrm{N_A}$ is the “electron content” on atom A, which is generally noninteger, and is defined as (Z_{A} – partial charge), where Z_{A} is the nuclear charge of atom A. The summation in Eq. (3) runs for all M atoms in the molecule. The two basic assumptions of Eq. (3) are that the correlation energy is the sum of the RECEP atomic correlation energies and that the value of these atomic correlation energies can be estimated from the atomic electron contents (N_{A}) in the molecule in the vicinity of stationary points.

The $E_{\mathrm{corr}}(N_{\mathrm{A}}, Z_{\mathrm{A}}$, method, charge def.) atomic energy terms in Eq. (3) are interpolated as follows:

$$\begin{aligned} &E_{\mathrm{corr}}(N_{\mathrm{A}}, Z_{\mathrm{A}}, \mathrm{method,\ charge\ def.}) \\ &\quad = (N_{\mathrm{A}} - \mathrm{N1})E_{\mathrm{fitpar}}(\mathrm{N2}, Z_{\mathrm{A}}, \mathrm{method,\ charge\ def.}) \\ &\qquad + (N2 - N_{\mathrm{A}})E_{\mathrm{fitpar}}(\mathrm{N1}, Z_{\mathrm{A}}, \mathrm{method,\ charge\ def.}) \ , \end{aligned} \tag{4}$$

where N1 and N2 are integer numbers of electrons, with $\mathrm{N1} \le N_{\mathrm{A}} \le \mathrm{N2} = \mathrm{N1} + 1$, and N_{A} is the electron content around atom A. $E_{\mathrm{fitpar}}(\mathrm{N2}, Z_{\mathrm{A}}$, method, charge def.) and $E_{\mathrm{fitpar}}(\mathrm{N1}, Z_{\mathrm{A}}$, method, charge def.) in Eq. (4) are the so-called RECEP atomic parameters that transform the partial charge into energy correction. The use of Eq. (4) is obvious; however, for hydrogen atoms we suggest using $E_{\mathrm{corr}}(N_{\mathrm{A}}$, 1, method, charge def.) = N_{A} E_{fitpar}(2, 1, method, charge def.)/2. (The reason for this choice is that although one electron has no correlation effect, but a hydrogen atom in a molecule with $0 < N_{\mathrm{A}} \le 1$ still has correlation contribution. For example, in the H_2 molecule $N_{\mathrm{A}} = 1$ and the correlation energy is well defined. These effects are discussed in Refs. [13–15].)

The RECEP atomic parameters play a crucial role in our estimation and it is a very delicate question how to choose them. In our earlier work [14, 15] we tested two alternative approaches.

A priori parameters can be obtained from the CI solution of the Schrödinger equation for atoms and from (basis set limit) HF-SCF energies [13]. Alternatively CCSD(T) or DFT correlation energies can also be used; however, the atoms in molecular environments lose their original high-spin states and the error of the HF-SCF method is very sensitive to the change in spin states. The results also depend on the partial charge definition. Different N_{A} values for the same molecule require different correlation energy parameters (we report NPA and Mulliken charges later). Our experience shows that using a priori $E_{\mathrm{par}}(\mathrm{N1\ or\ N2}, Z_{\mathrm{A}})$ parameters in Eq. (4)

combined with any partial charge definition for N_A in Eq. (3) the required chemical accuracy cannot be reached [14, 15].

A posteriori parameters can be obtained from a fitting procedure that finds the minimum of $Y = \sum_{(i=1,\ldots,L)} [E_{\mathrm{corr}}(\mathrm{G3})_i - E_{\mathrm{corr}}(\mathrm{RECEP, G3, charge\ def.})_i]^2$ in a set of L molecules [14, 15]. Here $E_{\mathrm{corr}}(\mathrm{G3})_i$ is calculated according to Eq. (2) and $E_{\mathrm{corr}}(\mathrm{RECEP, G3, charge\ def.})_i$ is calculated according to Eqs. (3) and (4). Let us denote the kth $E_{\mathrm{par}}(\mathrm{N1}, Z_A, \mathrm{G3}, \mathrm{charge\ def.})$ parameter of Eq. (4) by a_k and regroup the terms to yield $E_{\mathrm{corr}}(\mathrm{RECEP, G3, charge\ def.})_i = \sum_{(k=1,\ldots,L)} x_{ki} a_k$, where the x_{ki} are the factors ($N_A - \mathrm{N1}$ and $\mathrm{N2} - N_A$, cf. Eq. 4) arising from the summation of fractional electron numbers requiring the same a_k parameter in the molecule. The $\partial Y/\partial a_k = 0$ minimization yields a system of linear equations. The solution of this system of linear equations yields the desired a_k, i.e., the $E_{\mathrm{fitpar}}(\mathrm{N1}, Z_A, \mathrm{G3}, \mathrm{charge\ def.})$, fitted values (listed in Table 1 for NPA and Mulliken partial charges). The chemical accuracy (1–2 kcal/mol) can be reached if these fitted RECEP atomic parameters are used [14, 15]. The expensive G2 or G3 total energies were reproduced from the considerably less expensive HF-SCF/6-311+G(2d,p) single-point total energies and NPA atomic partial charges using these fitted atomic correlation parameters [14, 15]. One must keep in mind that the L molecules used for the fit are in the vicinity of their stationary points. Consequently the RECEP method (Eq. 3) is recommended only for similar geometries. We note that open-shell systems require a different parameter set; in this work we work with closed shells only. The set of L molecules also defines the (N, Z) range of the resulting parameters (cf. Table 1).

Table 1. Fitted $E_{\mathrm{fitpar}}(N, Z, \mathrm{method, charge\ definition})$ rapid estimation of basis set error and correlation energy from partial charges (*REBECEP*) atomic parameters (hartree) for calculating $E_{\mathrm{corr}}(\mathrm{REBECEP, method, charge\ definition})$ by Eqs. (3) and (4) for closed-shell neutral molecules in the vicinity of stationary points from HF/6-31G(d) results. N is the number of electrons and Z is the nuclear charge. The energy corrections of 65 molecules, listed in Table 2, were used to obtain these atomic parameters. *n.a.*: not available, it was not necessary to calculate for the 65 molecules, i.e., no such partial charge value occurred in the molecules at HF/6-31G(d) level. The (N = 9, Z = 7) nitrogen is not stable in the gas phase (double anion), but one must recognize that this table lists REBECEP atomic energy parameters for atoms in a molecular environment (i.e., in a bond)

Atom	N	Z	G3, Mulliken	G3, NPA	Expt., Mulliken	Expt., NPA
H	2	1	−0.0358	−0.0330	−0.0344	−0.0315
C	4	6	−0.1013	−0.1092	−0.1003	−0.1129
C	5	6	−0.1692	−0.1697	−0.1651	−0.1666
C	6	6	−0.2190	−0.2200	−0.2190	−0.2203
C	7	6	−0.2603	−0.2588	−0.2638	−0.2613
N	6	7	−0.2651	−0.2650	−0.2625	−0.2662
N	7	7	−0.2901	−0.2866	−0.2910	−0.2866
N	8	7	−0.3160	−0.3173	−0.3191	−0.3206
N	9	7	n.a.	−0.3847	n.a.	−0.3823
O	8	8	−0.3354	−0.3402	−0.3362	−0.3396
O	9	8	−0.3740	−0.3742	−0.3758	−0.3763
F	9	9	−0.3658	−0.3672	−0.3677	−0.3655
F	10	9	−0.4234	−0.4203	−0.4210	−0.4226

Fitting the correlation parameters eliminates most of the problems arising from the atomic electron reorganization in molecular environment, the basis set imperfections, and the quantitative differences between the various charge definitions (i.e., a different parameter set is fitted to different partial charges to yield the same energy correction). At this point we recall that partial charges are essentially mathematical constructions that serve to reflect the electron content around the selected atom of the molecule to a certain point of view. These partial charges are not physically measurable quantities; however, partial charges in an ideal case have a relation to the electron distribution in a molecule. Our previous experience, gained on 65 closed-shell neutral molecules (constructed of H, C, N, O, and F atoms) of the G2/97 database, shows that NPA [18], Mulliken, and charges derived from electrostatic potential field (ChelpG [19] or MK [20]) yield almost equivalent results for the RECEP-*n* type correlation energy calculation (if, of course, the proper partial-charge-dependent set of RECEP atomic parameters is used). However, the NPA charges yield slightly superior results and we recall only these [14, 15]. The overall root-mean-square (rms) deviation of the so-called RECEP-3 total energies from G3 total energies for 65 molecules was 1.76 kcal/mol (the average absolute deviation was 1.43 kcal/mol) [15]. The rms deviation of the fitted RECEP-3 enthalpies of formation from experimental ΔH_f^0 values at 298.15 K for the same 65 molecules was 2.17 kcal/mol (the average absolute deviation was 1.75 kcal/mol) [15]. Our analysis has shown that elimination of the errors arising from the G3 method decreases the average absolute deviation of the RECEP ΔH_f^0 values from the experimental values to 1.4–1.6 kcal/mol [15].

3 Obtaining the total energy from experimental results

For a given molecule, M, the so-called "experimental" ground-state total energy, $E_T(M, \mathrm{Expt.})$, can be obtained as

$$\begin{aligned} E_T&(M, \mathrm{Expt.}) \\ &= \Delta H_f^0(M, \mathrm{Expt.}) - E_{\mathrm{ZP}}(M, \mathrm{G3}) - E_{\mathrm{therm}}(M, \mathrm{G3}) \\ &\quad - \sum_{A=1}^{M} \left[\Delta H_f^0(A, \mathrm{Expt.}) - E_T(A, \mathrm{G3}) - E_{\mathrm{therm}}(A, \mathrm{G3})\right] , \end{aligned} \tag{5}$$

where $\Delta H_f^0(M, \mathrm{Expt.})$ and ΔH_f^0 (A, Expt.) are the experimental standard enthalpy of formation of molecule M and its constituent atoms, A. $E_{\mathrm{ZP}}(M, \mathrm{G3})$ is the zero-point energy, $E_{\mathrm{therm}}(M, \mathrm{G3})$ and $E_{\mathrm{therm}}(A, \mathrm{G3})$ are the differences between the enthalpy at 298.15 K and the energy at 0 K as used in the G3 calculations [1]. The sum in Eq. (5) runs for all atoms in the molecule M and $E_T(A, \mathrm{G3})$ is the G3 total energy of these atoms. In this way the actual value of $E_T(M, \mathrm{Expt.})$ depends on the method used for the calculation of E_{ZP}, E_{therm}, and $E_T(A)$. In Eq. (5) we use the values from the G3 method [1]. A different choice of these terms, for example, complete CI values for $E_T(A)$ or more precise $E_{\mathrm{ZP}}(M)$, would influence the value of $E_T(M, \text{Expt.})$, and in this sense $E_T(M, \text{Expt.})$ can be improved; however, it is evident that $E_T(M, \text{Expt.})$ will provide perfect agreement with the experimental enthalpy of formation after the necessary corrections (cf. Eq. 5) and this does not hold for $E_T(\text{G3})$ in Eq. (2). Hereafter, the experimental correlation energy is derived analogously to the so-called G3 correlation energy (cf. Eq. 2):

$$E_{\text{corr}}(\text{Expt., basis}) = E_T(\text{Expt.}) - E_T(\text{HF-SCF/basis}) \ , \qquad (6)$$

where basis notes the actual basis set used for the HF-SCF calculation. The use of small basis sets requires specific considerations (vide infra).

Next we use these energy differences defined in Eqs. (2) and (6) in the fitting procedure. The RECEP enthalpy of formation is calculated as

$$\begin{aligned}\Delta H_f^0&(M, \text{RECEP, Expt., charge def.})\\ &= E_T(M, \text{RECEP, Expt., charge def.})\\ &\quad + E_{\text{ZP}}(M, \text{G3}) + E_{\text{therm}}(M, \text{G3})\\ &\quad + \sum_{A=1}^{M}\left[\Delta H_f^0(A, \text{Expt.}) - E_T(A, \text{G3})\right.\\ &\qquad\qquad \left. - \ E_{\text{therm}}(A, \text{G3})\right] \ . \qquad (7)\end{aligned}$$

The RECEP enthalpy of formation calculated from $E_T(M, \text{RECEP, Expt., charge def.})$ instead of $E_T(M, \text{RECEP, G3, charge def.})$ approximates the experimental enthalpy of formation better. This is because the error of the G3 method is missing from the $E_T(M, \text{RECEP, Expt., charge def.})$ values.

Comparison of Eqs. (5) and (7) shows that using the same correction factors consistently the simple difference of $E_T(M, \text{RECEP, Expt., charge def.})$ and $E_T(M, \text{Expt.})$ is equal to the difference of the corresponding ΔH_f^0 values.

4 Basis set error

The question is how the considerable simplification of the basis would influence the quality of the results. In this article we present the results obtained by the use of the smaller 6-31G(d) basis set instead of 6-311+G(2d,p). We note that problems were noticed in earlier work because the 6-311G(d) basis set was defined and used in a somewhat unsatisfactory manner [1]. For Li–Ne, the 6-311G(d) basis set has been criticized by Grev and Shaefer [21] as being too “close-in” to be of triple-zeta quality. For Na–Ar the 6-311G(d) basis was defined in a nonuniform manner across the row. To avoid these difficulties the uniformly defined 6-31G(d) basis was used for the simplest MP4 and QCISD(T) single-point calculations instead of the 6-311G(d) basis in the G3 method.

For our purposes, the 6-31G(d) basis set is advantageous because of the considerably reduced computation time. Because in Eq. (6) the E_T[HF-SCF/6-311+G(2d,p)] term is replaced by the E_T[HF-SCF/6-31G(d)] term the $E_{\text{corr}}(\text{Expt., basis})$ of Eq. (6) contains considerable basis set error. Consequently, the atomic parameters fitted to this energy will lose their original (i.e., RECEP) meaning, and they serve simply as atomic parameters that correct the HF-SCF/6-31G(d) total energy to yield near-“experimental”-quality total energy. Although the basis set error energy is usually treated at molecular level, in the present work it is implicitly partitioned among the atoms and the atomic partitions are merged with the atomic correlation energy parameters, yielding the fitted parameters shown in Table 1. For this reason we call the current parameter sets rapid estimation of basis set error and correlation energy from partial charges (REBECEP) atomic parameters.

The use of HF-SCF partial charges, derived from the one-electron density, $\rho_{\text{HF-SCF}}(\mathbf{r})$, for electron correlation effects can be justified readily. It is known that one-electron properties, such as $\rho(\mathbf{r})$, are relatively well reproduced using $\rho_{\text{HF-SCF}}(\mathbf{r})$. The partial charges are integrated quantities; thus, subtle differences in $\rho(\mathbf{r})$ are usually averaged out, leading to even smaller differences in partial charges. Moreover, the NPA partial charges that usually provide the best RECEP results are not very sensitive to the basis set used (Mulliken charges do not behave so well). The analysis of Eq. (4) shows that small variations in partial charges do not cause large changes in the RECEP energy. For example, if a more negative partial charge occurs on one of the atoms in the molecule owing to the basis set effect, an equal positive partial charge must be distributed on the other atoms (the total number of electrons does not change). According to Table 1 and Eq. (4) more negative partial charges will result in more negative atomic energy corrections and more positive partial charges will result in fewer negative atomic energy corrections. The summation of the atomic energies will average out the differences to some degree. Since we have a proportionality with the charges one can expect that the RECEP energy is less sensitive to the basis set extension effects than the correlation energy calculated in a classical way, for example, CCSD(T).

In order to study the influence of the basis set on the partial charges we performed an extensive analysis. The results show that the NPA partial charges are quite similar for the two basis sets used in this study, and the Mulliken partial charges show considerable differences as expected. Our preliminary results show that for 104 C atoms in different molecules the average deviation of the 6-311+G(2d,p) Mulliken partial charges from the 6-31G(d) Mulliken partial charges is −0.098 au (with 0.192 rms deviation). The corresponding value for NPA charges is 0.043 au (with 0.087 rms deviation). For oxygen and nitrogen atoms the similarity of the NPA charges is even better, 0.010 or 0.015, respectively. Details of this lengthy analysis will be published elsewhere.

5 Results and discussion

Here we show the results obtained with the use of the small 6-31G(d) basis set and atomic parameters (cf. Table 1) in the vicinity of stationary points for closed-shell molecules. Table 1 shows the fitted atomic parameter sets to reproduce G3 and experimental quality total energies by Eqs. (3) and (4), where the term "method" in the table head stands for G3 or experimental. These parameters were obtained by the linear fit described previously using the energies and partial charges of 65 molecules composed of H, C, N, O, and F atoms selected from the G2/97 thermochemistry database (cf. Table 2). The program used to obtain the fitted parameters is available from the authors. These parameters provide near experimental quality molecular total energy (vide infra). It is expected that good-quality total energy can be obtained for any other molecule provided that the parameter set in Table 1 is applicable for it, i.e., it consists of the atoms listed and the necessary HF/6-31G(d) results are available, as shown in our earlier work [15].

By adding the REBECEP energy correction to E_T[HF-SCF/6-31G(d)] the corresponding REBECEP total energy can be obtained (cf. Eq. 6). The only ab initio calculation required is a HF-SCF/6-31G(d) total energy calculation followed by an almost instant NPA or Mulliken partial charge analysis. A detailed numerical example was published in Ref. [14]. The corresponding ΔH_f^0(M, RECEP, Expt., charge def.) values can be obtained according to Eq. (7). Inspection of the values of the energy parameters in Table 1 shows the following. The values of the fitted REBECEP energy parameters (and the quality of the E_T and ΔH_f^0 results, vide infra) depend only slightly on the charge definition. Thus, the method is not very sensitive to the partial charges used in the fitting procedure; however, the parameter set must be used with the partial charge it belongs to. We note that our earlier experience has shown that the partial charges that approximate the electrostatic potential around the molecule (ChelpG and MK) also provide very similar parameters and quality of results [14, 15].

The experimental enthalpies of formation, ΔH_f^0 (Expt.), HF/6-31G(d) total ground-state electronic energies, E_T(HF/6-31G(d), and the corrections necessary to obtain G3 total energies, E_{corr}[G3, 6-31G(d)], defined as E_T(G3) − E_T[HF/6-31G(d)], as well as the corrections necessary to obtain "experimental-quality total energies", E_{corr}[Expt., 6-31G(d)] = E_T(Expt.) − E_T[HF/6-31G(d)] [cf. Eq. 5 for E_T(Expt.)], are shown in Table 2. All these energies are listed for the set of 65 closed-shell molecules mentioned previously. These two types of correction energies in Table 2 and the corresponding NPA and Mulliken charges were the input for the fitting procedure that resulted in the fitted REBECEP atomic energy parameters in Table 1. We do not report here the calculated partial charges in order to save space; however, these charges are available from the authors (or can be calculated easily, for axample, using the GAUSSIAN program package [17]). Because at this stage we want to reproduce the G3 results and the experimental enthalpies of formation with a much faster method, the choice of the MP2(FU)/6-31G(d) level G3 geometries is evident. We note that the speed increase using the REBECEP method compared to the G3 method is several orders of magnitude, and the required computation time scales as the HF-SCF/6-31G(d) method scales with the increase in the size of the molecule. Thus, for new molecules the use of very expensive calculated geometry, such as MP2 (FU)/6-31G(d), decreases considerably with the use (and the feasibility) of the proposed method. Considerably less expensive alternatives than MP2 (such as DFT, HF or even less expensive molecular mechanics geometries) will be investigated in the future. At the current level we limit our study to the use of the available geometries used in the G3 method. The REBECEP method can be applied to any molecule if the geometry is known and a HF/6-31G(d) calculation can be performed and the required partial charges can be calculated. For molecules not included in the G2/97 database we suggest using the experimental geometries, if available. If the experimental geometry is not available probably the best choice would be to use the B3LYP/6-31G(d) equilibrium geometries because it was found earlier that such geometries are useful alternatives to the considerably more expensive MP2 geometries [22]. We note that B3LYP/6-31G(d) total energies are not sufficiently good for reliable thermochemistry; thus, the 6-311+G(3df,2p) basis set was used in the earlier studies to obtain 3.43 kcal/mol [2] or 3.11 kcal/mol [23] average absolute deviations for various test sets. The use of B3LYP/6-31G(d) equilibrium geometries for our purposes will be tested in future work.

The energy deviations of calculated G3 enthalpies of formation from experiment (for comparison) and deviations of E_{corr}(G3) or E_{corr}(Expt.) from corresponding E_{corr}(REBECEP) values are shown in Table 3. We note again that the energy deviations from the experimental values provide directly the deviations of the ΔH_f^0(M, REBECEP, Expt., charge def.) from the experimental enthalpies of formation shown in Table 2 (cf. Eqs. 5, 7). These results show that the fitted REBECEP parameter set is capable of providing ground-state total electronic energies approaching chemical accuracy for the selected 65 molecules.

The average absolute deviation of ΔH_f^0(M, REBECEP, Expt., NPA) enthalpies of formation from the experimental enthalpies of formation (Table 2) is 1.39 kcal/mol for the test set of 65 neutral enthalpies (calculated from the results in Table 3). The average absolute deviation of E_T(M, REBECEP, G3, NPA) total energies from the G3 total energies is 1.32 kcal/mol. Thus, the G3 total energies are approximated slightly better; however, for the prediction of experimental enthalpies of formation the (REBECEP, Expt., NPA) parameter set must be used (Table 1). The enthalpies of formation calculated using the (REBECEP, Expt., NPA) parameter set are not influenced by the errors of the G3 method. The average absolute deviation of the G3 enthalpies of formation from the experimental enthalpies of formation is 0.74 kcal/mol for the selected 65 molecules. The corresponding REBECEP results obtained using Mulliken charges provide 1.7 kcal/mol average absolute deviation.

The statistical distribution of the results is more precisely characterized by the center of the distribution and the rms deviation than by the average absolute deviation. The latter is shown only for comparison purposes with the earlier results. The rms deviation of the REBECEP total energies from the G3 and the

Table 2. Experimental enthalpies of formation, HF/6-31G(d) level ground-state total electronic energies, E_T(HF-SCF), as well as G3 and experimental energy corrections, E_{corr}(G3) and E_{corr}(Expt.), respectively, for the 65 molecules selected from the G2/97 database. E_T(G3) = E_T(HF-SCF) + E_{corr}(G3) for the ground electronic state (analogous to Eq. 2) and a similar equation holds for the experimental "total energy" (analogous to Eq. 6). For comparison purposes, we used the original G3 MP2(FU)/6-31G(d) equilibrium geometries of the 65 selected molecules. The energy corrections include the correlation energies and the 6-31G(d) basis set errors

No.	Molecule	ΔH_f^0(Expt.) (kcal/mol)[a]	$-E_T$(HF-SCF) (hartree)	$-E_{corr}$(G3) (hartree)	$-E_{corr}$(Expt.) (hartree)
1	Methane (CH_4)	−17.9	40.1951	0.3052	0.3048
2	Ammonia (NH_3)	−11.0	56.1838	0.3563	0.3575
3	Water (H_2O)	−57.8	76.0098	0.3928	0.3933
4	Hydrogen fluoride (HF)	−65.1	100.0023	0.4077	0.4073
5	Acetylene (C_2H_2)	54.2	76.8156	0.4867	0.4878
6	Ethylene ($H_2C{=}CH_2$)	12.5	78.0311	0.5252	0.5249
7	Ethane ($H_3C{-}CH_3$)	−20.1	79.2286	0.5661	0.5656
8	Hydrogen cyanide (HCN)	31.5	92.8702	0.5213	0.5210
9	Formaldehyde ($H_2C{=}O$)	−26.0	113.8638	0.5934	0.5924
10	Methanol (CH_3OH)	−48.0	115.0342	0.6444	0.6442
11	Hydrazine ($H_2N{-}NH_2$)	22.8	111.1680	0.6591	0.6624
12	Hydrogen peroxide (HO–OH)	−32.5	150.7601	0.7302	0.7321
13	Carbon dioxide (CO_2)	−94.1	187.6284	0.8833	0.8813
14	CF_4	−223.0	435.6416	1.6834	1.6821
15	COF_2	−149.1	311.6104	1.2857	1.2911
16	N_2O	19.6	183.6631	0.9249	0.9276
17	NF_3	−31.6	352.5320	1.4122	1.4121
18	C_2F_4 ($F_2C{=}CF_2$)	−157.4	473.4117	1.9189	1.9111
19	CF_3CN	−118.4	428.4722	1.8224	1.8197
20	Propyne (C_3H_4)	44.2	115.8619	0.7471	0.7475
21	Allene (C_3H_4)	45.5	115.8602	0.7472	0.7465
22	Cyclopropene (C_3H_4)	66.2	115.8218	0.7487	0.7523
23	Propylene (C_3H_6)	4.8	117.0707	0.7878	0.7877
24	Cyclopropane (C_3H_6)	12.7	117.0585	0.7870	0.7882
25	Propane (C_3H_8)	−25.0	118.2633	0.8293	0.8288
26	*trans*-1,3-Butadiene	26.3	154.9181	1.0096	1.0103
27	Dimethylacetylene (2-butyne)	34.8	154.9066	1.0077	1.0084
28	Methylenecyclopropane (C_4H_6)	47.9	154.8866	1.0095	1.0072
29	Bicyclobutane	51.9	154.8708	1.0130	1.0172
30	Cyclobutane (C_4H_8)	6.8	156.0964	1.0501	1.0501
31	Isobutene (C_4H_8)	−4.0	156.1098	1.0524	1.0524
32	*trans*-Butane (C_4H_{10})	−30.0	157.2979	1.0930	1.0923
33	Isobutane (C_4H_{10})	−32.1	157.2984	1.0951	1.0947
34	Spiropentane (C_5H_8)	44.3	193.9172	1.2724	1.2730
35	Benzene (C_6H_6)	19.7	230.7020	1.4464	1.4474
36	Difluoromethane (H_2CF_2)	−107.7	237.8948	0.9996	0.9985
37	Trifluoromethane (HCF_3)	−166.6	336.7692	1.3425	1.3416
38	Methylamine ($H_3C{-}NH_2$)	−5.5	95.2091	0.6126	0.6141
39	Acetonitrile ($CH_3{-}CN$)	18.0	131.9225	0.7808	0.7805
40	Nitromethane ($CH_3{-}NO_2$)	−17.8	243.6538	1.2606	1.2606
41	Methylnitrite ($CH_3{-}O{-}N{=}O$)	−15.9	243.6596	1.2505	1.2508
42	Formic acid (HCOOH)	−90.5	188.7586	0.9312	0.9311
43	Methyl formate ($HCOOCH_3$)	−85.0	227.7857	1.1879	1.1854
44	Acetamide (CH_3CONH_2)	−57.0	207.9735	1.1625	1.1642
45	Aziridine (C_2H_4NH)	30.2	133.0373	0.8314	0.8333
46	Cyanogen (NCCN)	73.3	184.5778	1.0062	1.0067
47	Dimethylamine [$(CH_3)_2NH$]	−4.4	134.2380	0.8719	0.8733
48	*trans*-Ethylamine ($CH_3CH_2NH_2$)	−11.3	134.2468	0.8756	0.8756
49	Ketene (CH_2CO)	−11.4	151.7218	0.8157	0.8145
50	Oxirane (C_2H_4O)	−12.6	152.8654	0.8608	0.8608
51	Acetaldehyde (CH_3CHO)	−39.7	152.9135	0.8548	0.8546
52	Glyoxal (HCOCOH)	−50.7	226.5864	1.1439	1.1424
53	Ethanol (CH_3CH_2OH)	−56.2	154.0743	0.9072	0.9070
54	Dimethyl ether (CH_3OCH_3)	−44.0	154.0634	0.8994	0.8987
55	Vinyl fluoride ($CH_2{=}CHF$)	−33.2	176.8807	0.8744	0.8725
56	Acrylonitrile ($CH_2{=}CHCN$)	43.2	169.7620	1.0024	1.0050
57	Acetone (CH_3COCH_3)	−51.9	191.9599	1.1176	1.1175
58	Acetic acid (CH_3COOH)	−103.4	227.8071	1.1926	1.1928
59	Acetyl fluoride (CH_3COF)	−105.7	251.7949	1.2031	1.2028
60	2-Propanol [$(CH_3)_2CHOH$]	−65.2	193.1139	1.1721	1.1714
61	Methyl ethyl ether ($C_2H_5OCH_3$)	−51.7	193.1033	1.1627	1.1609

Table 2. (Contd.)

No.	Molecule	ΔH_f^0(Expt.) (kcal/mol)[a]	$-E_T$(HF-SCF) (hartree)	$-E_{corr}$(G3) (hartree)	$-E_{corr}$(Expt.) (hartree)
62	Trimethylamine [$(CH_3)_3N$]	−5.7	173.2682	1.1350	1.1346
63	Furan (C_4H_4O)	−8.3	228.6224	1.3052	1.3060
64	Pyrrole (C_4H_5N)	25.9	208.8059	1.2784	1.2803
65	Pyridine (C_5H_5N)	33.6	246.6938	1.4844	1.4847

[a] Ref. [1]

experimental total energies for the selected 65 molecules is in the range 1.7–2.2 kcal/mol for both, NPA and Mulliken charges respectively; the smaller value belongs to the NPA charges. The corresponding average deviations are in the range −0.08–0.05 kcal/mol. The corresponding rms deviation of the G3 total energies from the experimental total energies is 1.15 kcal/mol, and the average deviation is −0.06 kcal/mol.

Analysis of the results in Table 3 reveals that the REBECEP method has 4–5 kcal/mol deviation in the case of HCN, COF_2, cyclopropene, bicyclobutane, acetonitrile, cyanogen, pyrrole, and pyridine (nos. 8, 15, 22, 29, 39, 46, 64, 65 in Table 3) if Mulliken charge is used to reproduce experimental total energy. However, compared to the G3 values, only HCN, C_2F_4, acetonitrile, and cyanogen (nos. 8, 18, 39, 46 in Table 3) show such large (4.2, −4.7, 4.6, −4.2 kcal/mol, respectively) deviations; pyrrole and pyridine (nos. 64, 65 in Table 3) have smaller deviations (−3.8, 3.5 kcal/mol, respectively). By using NPA charges in the REBECEP estimation to reproduce the experimental total energy, COF_2, N_2O, cyclopropene, byciclobutane, and acetonitrile (nos. 15, 16, 22, 29, 39 in Table 3) deviate considerably (−5.3, −4.3, −4.4, −4.7, 4.5 kcal/mol). On the other hand, when estimating the G3 total energy, acetonitrile (no. 39) remains the only molecule with a large deviation (4.0 kcal/mol). It should be noted that recent experimental results for COF_2 have been reported [24] that cast doubt on the value in the JANAF tables used in the G2/97 test set. As a consequence the enthalpy of formation COF_2 was deleted from the recent G3/99 test set [3].

As noted previously, summing the deviation of the G3 enthalpy of formation from experiment and the E_{corr}(G3) − E_{corr}(REBECEP, G3, NPA) values in Table 3 yields directly the deviation of the (REBECEP, G3, NPA) enthalpy of formation from experiment. The experimental enthalpies of formation of bicyclobutane and spiropentane (nos. 29, 34 in Table 3) are relatively well reproduced by G3 theory, while our RECEP-3 parameter set [15] developed to correct the HF-SCF/6-311+G(2d,p) energies could not reproduce these G3 values. The results in Table 3 show a great improvement for spiropentane (no. 34) using the current REBECEP parameter set. One might reason that these strained cyclic hydrocarbons might require a different parameterization and these molecules should be treated separately; however, this does not seem necessary using the current REBECEP parameter set and partial charges. It is also worth noting that these problematic molecules have larger or middle-range enthalpies of formation (Table 2) and molecules with low absolute values of the enthalpy of formation do not have problems. Another observation is that the NPA charges provide systematically better results than the Mulliken charges with any REBECEP parameterization, and this is in agreement with our earlier G2 and G3 results [14]. We think this can be explained by the fact that Mulliken charges might show rather unphysical behavior with respect to the choice of basis set. The NPA charges are free of such problems.

It is fundamental to recognize that the parameters in Table 1 are applicable only for closed-shell molecules in the vicinity of their stationary points because all types of RECEP parameters should converge toward the corresponding $E_{corr}(N, Z)$ high-spin atomic correlation energies in free space [now including the 6-31G(d) basis set error in REBECEP] as the molecule separates into atoms in free space and the partial charges converge toward zero (or $Z - N$). This is a basic property of the physics of RECEP parameters [13–15]. If the basis goes to the infinite basis, the REBECEP parameters converge to a parameter set which contains only correlation correction (what we called RECEP parameters in our previous work). The current REBECEP parameters are suggested for closed-shell neutral molecules near to their stationary points containing H, C, N, O, and F atoms. These restrictions (atom types, vicinity of stationary points, and closed shells) come from the training set of 65 molecules used to obtain the parameters.

Further extensions are straightforward: Inclusion of new atom types (our primary goal is to include at least sulfur and phosphorus because of their biochemical importance; thus, all parameters for the 20 essential amino acids and DNA or RNA molecules would be available), charged molecules (e.g., effects of protonation and deprotonation, etc.), and radicals. (This latter case probably needs a different parameter set from the closed-shell parameters).

The true value of such a parameterized method can be judged from its capacity to predict total energies of molecules which were not included in the training set. The performance of our RECEP-3 and REBECEP parameter sets have been tested on 51 closed-shell neutral molecules (50 molecules from the G3/99 thermochemistry database [3] plus urea) not included in the training set (cf. Table 2). The results will be published elsewhere [25].

6 Conclusions

With our new REBECEP atomic parameter set it is possible to reproduce the G3 and experimental total

Table 3. Energy deviations (kcal/mol)

No.	Molecule	Deviation[a] Expt–G3	E_T(method) − E_T(REBECEP, method, charge def.)[b]			
			G3, NPA	G3, Mulliken	Expt., NPA	Expt., Mulliken
1	Methane (CH_4)	0.3	0.1	0.5	0.2	0.8
2	Ammonia (NH_3)	−0.8	0.0	−2.9	0.0	−2.5
3	Water (H_2O)	−0.3	−1.7	−2.3	−1.3	−2.1
4	Hydrogen fluoride (HF)	0.2	−2.2	−2.2	−1.8	−2.3
5	Acetylene (C_2H_2)	−0.7	−1.7	0.2	−2.1	0.1
6	Ethylene ($H_2C{=}CH_2$)	0.2	−0.4	0.6	−0.3	0.9
7	Ethane ($H_3C–CH_3$)	0.3	0.9	1.0	1.0	1.2
8	Hydrogen cyanide (HCN)	0.2	1.9	4.2	2.4	5.1
9	Formaldehyde ($H_2C{=}O$)	0.6	−0.4	1.0	−0.5	1.3
10	Methanol ($CH_3–OH$)	0.1	0.5	0.4	0.7	0.5
11	Hydrazine ($H_2N–NH_2$)	−2.1	0.5	2.2	0.2	2.0
12	Hydrogen peroxide (HO–OH)	−1.2	0.3	−2.7	−0.4	−2.7
13	Carbon dioxide (CO_2)	1.2	−3.8	−2.4	−2.0	−1.9
14	CF_4	0.9	2.3	1.8	3.6	1.8
15	COF_2	−3.4	−2.3	−1.0	−5.3	−5.2
16	N_2O	−1.7	−3.5	−1.2	−4.3	−2.1
17	NF_3	0.1	0.8	1.9	0.2	1.9
18	C_2F_4 ($F_2C{=}CF_2$)	4.9	−3.7	−4.7	−2.2	−2.0
19	CF_3CN	1.8	1.1	−2.1	1.9	−1.0
20	Propyne (C_3H_4)	−0.2	0.6	1.0	0.7	1.3
21	Allene (C_3H_4)	0.5	−0.7	−1.0	0.0	−0.2
22	Cyclopropene (C_3H_4)	−2.2	−2.3	−2.0	−4.4	−4.1
23	Propylene (C_3H_6)	0.0	0.1	0.5	0.2	0.6
24	Cyclopropane (C_3H_6)	−0.7	1.3	1.6	0.7	1.1
25	Propane (C_3H_8)	0.3	0.5	0.1	0.6	0.2
26	*trans*-1,3-Butadiene	−0.4	−1.3	−0.3	−1.5	−0.4
27	Dimethylacetylene (2-butyne)	−0.4	2.4	2.0	2.5	2.0
28	Methylenecyclopropane (C_4H_6)	1.5	0.0	−0.5	1.8	1.2
29	Bicyclobutane	−2.6	−2.4	−2.5	−4.7	−4.9
30	Cyclobutane (C_4H_8)	0.0	0.1	−0.4	0.2	−0.5
31	Isobutene (C_4H_8)	0.0	−0.8	−1.6	−0.7	−1.6
32	*trans*-Butane (C_4H_{10})	0.4	−0.3	−1.1	−0.1	−0.9
33	Isobutane (C_4H_{10})	0.2	−1.4	−2.2	−1.3	−2.2
34	Spiropentane (C_5H_8)	−0.4	−0.1	−0.7	−0.1	−0.8
35	Benzene (C_6H_6)	−0.6	2.1	2.0	2.3	1.9
36	Difluoromethane (H_2CF_2)	0.7	0.5	1.8	−0.6	0.9
37	Trifluoromethane (HCF_3)	0.5	2.0	3.5	0.6	2.0
38	Methylamine ($H_3C–NH_2$)	−1.0	−0.8	−0.2	−0.8	−0.5
39	Acetonitrile ($CH_3–CN$)	0.2	4.0	4.6	4.5	5.3
40	Nitromethane ($CH_3–NO_2$)	0.0	−0.7	−2.7	0.4	−1.9
41	Methylnitrite ($CH_3–O–N{=}O$)	−0.2	3.0	1.1	3.4	1.5
42	Formic acid (HCOOH)	0.1	0.3	1.2	−0.1	1.1
43	Methyl formate ($HCOOCH_3$)	1.6	0.0	1.6	0.5	2.4
44	Acetamide (CH_3CONH_2)	−1.1	0.6	0.6	0.2	0.1
45	Aziridine (C_2H_4NH)	−1.2	−0.4	1.0	−0.9	0.5
46	Cyanogen (NCCN)	−0.3	−0.8	−4.2	−0.8	−4.3
47	Dimethylamine [$(CH_3)_2NH$]	−0.9	−0.6	0.4	−1.1	−0.2
48	*trans*-Ethylamine ($CH_3CH_2NH_2$)	0.0	−0.7	−0.7	0.3	−0.1
49	Ketene (CH_2CO)	0.8	−1.7	−0.9	−1.5	−0.2
50	Oxirane (C_2H_4O)	0.0	0.2	−0.1	0.0	−0.3
51	Acetaldehyde (CH_3CHO)	0.1	1.1	2.0	0.7	2.0
52	Glyoxal (HCOCOH)	0.9	1.7	2.5	1.6	3.0
53	Ethanol (CH_3CH_2OH)	0.1	0.9	0.0	1.2	0.1
54	Dimethyl ether (CH_3OCH_3)	0.4	0.9	0.9	0.7	0.7
55	Vinyl fluoride ($CH_2{=}CHF$)	1.2	−0.6	0.6	−0.1	1.3
56	Acrylonitrile ($CH_2{=}CHCN$)	−1.6	2.1	3.0	0.9	2.0
57	Acetone (CH_3COCH_3)	0.0	1.7	2.1	1.4	2.0
58	Acetic acid (CH_3COOH)	−0.1	2.2	2.3	1.8	2.0
59	Acetyl fluoride (CH_3COF)	0.1	0.6	1.9	0.0	1.3
60	2-Propanol [$(CH_3)_2CHOH$]	0.5	−0.9	−2.2	−0.3	−1.9
61	Methyl ethyl ether ($C_2H_5OCH_3$)	1.1	0.9	0.1	1.5	0.5
62	Trimethylamine [$(CH_3)_3N$]	0.2	−2.8	−1.4	−2.6	−1.2
63	Furan (C_4H_4O)	−0.5	−2.9	−3.2	−3.3	−3.6

Table 3. (Contd.)

No.	Molecule	Deviation[a] Expt–G3	E_T(method) − E_T(REBECEP, method, charge def.)[b]			
			G3, NPA	G3, Mulliken	Expt., NPA	Expt., Mulliken
64	Pyrrole (C_4H_5N)	−1.2	−2.7	−3.8	−2.7	−4.0
65	Pyridine (C_5H_5N)	−0.1	3.0	3.5	3.7	4.0

[a] Deviation of calculated G3 enthalpies of formation from experiment

[b] Deviation of total energy calculated by a chosen method (G3 or experimental) from corresponding REBECEP results in the case of two partial charges (NPA or Mulliken). Recall that these deviations are not only total energy deviations but ΔH_f^0 deviations as well because E_{ZP} and E_{therm} of the molecules are taken from G3 calculations. Summing ΔH_f^0(Expt.) − ΔH_f^0(G3); i.e. column Expt-G3, and E_T (G3) − E_T(REBECEP, G3, charge def.) yields directly ΔH_f^0(Expt.) − ΔH_f^0(REBECEP, G3, charge def.)

energy from single-point small basis set ab initio HF/6-31G(d) total energy and partial atomic charge calculations. The only time-consuming step in the REBECEP method is the HF/6-31G(d) calculation, which determines the computation time requirement and limits the size of the molecule to be calculated. The REBECEP atomic parameter set was optimized for 65 molecules from the G2/97 thermochemistry database composed of H, C, N, O, and F atoms. The optimization was done by a simple multilinear least-squares fit to approximate the G3 and experimental total electronic energies for this set of molecules using two kinds of partial charge definitions (Mulliken and NPA). The instant REBECEP calculation can be used with this relatively small basis set; thus, we conclude that the basis set errors can be absorbed by the parameter set (i.e., it can be treated at atomic level). The REBECEP total energies calculated with the previously mentioned parameters and with Mulliken or NPA charges can approach the required chemical accuracy (cf. Tables 1 and 3). The average absolute deviation of $E_T(M$, REBECEP, G3, NPA) total energies from the G3 total energies is 1.32 kcal/mol (where M symbolizes the molecule and G3 indicates the type of calculation in the set used for the fit). The average absolute deviation of the $\Delta H_f^0(M$, REBECEP, Expt., NPA) enthalpies of formation from the experimental enthalpies of formation is 1.39 kcal/mol for the test set of 65 neutral enthalpies. The most problematic molecules for the REBECEP parameter set were bicyclobutane and acetonytrile (the largest deviations remain within ± 5 kcal/mol). The REBECEP atomic parameters are listed in Table 1 and are recommended for calculating molecular ground-state total energies of neutral closed-shell molecules in the vicinity of stationary points containing the atoms listed. The user has the choice of what accuracy level (G3 or experimental) is desirable or which partial charge is to be used (Mulliken or NPA). Enlarging the molecule set as well as setting up different RECEP and REBECEP parameters for open-shell molecules are among our future plans.

7 Supplementary material

The molecular geometries and the partial charges of the 65 molecules are available from the authors. The FORTRAN program that calculates the REBECEP parameters is also available.

Acknowledgements. S.K. is thankful to the Domus Hungarica Scientarium et Artium 2000 summer grant. The work was supported by an OTKA Grant (T 031767, Hungary).

References

1. Curtiss LA, Raghavachari K, Redfern PC, Rassolov V, Pople JA (1998) J Chem Phys 109: 7764
2. (a) Curtiss LA, Raghavachari K, Redfern PC, Pople JA (1997) J Chem Phys 106: 1063; (b) Curtiss LA, Redfern PC, Raghavachari K, Pople JA (1998) J Chem Phys 109: 42
3. Curtiss LA, Raghavachari K, Redfern PC, Pople JA (2000) J Chem Phys 112: 7374
4. Ochtersky JW, Petersson GA, Montgomery JA Jr (1996) J Chem Phys 104: 2598
5. Fast PL, Sanchez ML, Corchado JC, Truhlar DG (1999) J Chem Phys 110: 11679
6. Martin JML, Oliveira de G (1999) J Chem Phys 111: 1843
7. Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Oxford University Press, New York
8. Dreizler RM, Gross EKV (1990) Density functional theory. Springer, Berlin Heidelberg New York
9. Becke AD (1993) J Chem Phys 98: 5648
10. Perdew JP (1991) In: Ziesche P, Eschrig H (eds) Electronic structure of solids. Akademie, Berlin, p 11
11. Lee C, Yang W, Parr RG (1988) Phys Rev B 37: 785
12. Frisch MJ, Trucks GW, Head-Gordon M, Gill PMW, Wong MW, Foresman JB, Johnson BG, Schlegel HB, Robb MA, Replogle ES, Gomperts R, Andres JL, Raghavachari K, Binkley JS, Gonzalez C, Martin RL, Fox DJ, DeFrees DJ, Baker J, Stewart JJP, Pople JA (1993) Gaussian 92/DFT. Gaussian, Pittsburgh, Pa
13. Kristyan S, Csonka GI (1999) Chem Phys Lett 307: 469
14. Kristyán S, Csonka GI (2001) J Comput Chem 22: 241. DOI 10.1002/1096-987X(20010130)22:2<241::AID-JCC11>3.0.CO;2-C
15. Kristyán S, Ruzsinszky A, Csonka GI (2001) J Phys Chem A 105: 1926
16. Löwdin PO (1959) Adv Chem Phys 2: 207
17. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98. Gaussian, Pittsburgh, Pa
18. Reed AE, Weinstock RB, Weinhold FJ (1985) J Chem Phys 83: 735

19. Breneman CM, Wiberg KB (1990) J Comput Chem 11: 361
20. Besler BH, Merz KM Jr, Kollman PA (1990) J Comput Chem 11: 431
21. Grev RS, Schaefer HF (1989) J Chem Phys 91: 7305
22. Baboul AG, Curtiss LA, Redfern PC, Raghavachari K (1999) J Chem Phys 110: 7650
23. Petersson GA, Malick DK, Wilson WG, Ochterski JW, Montgomery JA Jr, Frisch MJ (1998) J Chem Phys 109: 10570
24. Asher RL, Appelman EH, Ruscic B (1996) J Chem Phys 105: 9781
25. Kristyán S, Ruzsinszky A, Csonka GI (2001) Theor Chem Acc (accepted)