Gaussian-3 theory using scaled energies

Larry A. Curtiss[a)]
Materials Science and Chemistry Divisions, Argonne National Laboratory, Argonne, Illinois 60439

Krishnan Raghavachari
Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974

Paul C. Redfern
Materials Science and Chemistry Divisions, Argonne National Laboratory, Argonne, Illinois 60439

John A. Pople
Department of Chemistry, Northwestern University, Evanston, Illinois 60208

A modification of Guassian-3 (G3) theory using multiplicative scale factors, instead of the additive higher level correction, is presented. In this method, referred to as G3S, the correlation energy is scaled by five parameters and the Hartree–Fock energy by one parameter. The six parameters are fitted to the G2/97 test set of 299 energies and the resulting mean absolute deviation from experiment is 0.99 kcal/mol compared to 1.01 kcal/mol for G3 theory. The G3S method has the advantage compared to G3 theory in that it can be used for studying potential energy surfaces where the products and reactants have a different number of paired electrons. In addition, versions of the computationally less intensive G3(MP3) and G3(MP2) methods that use scaled energies are also presented. These methods, referred to as G3S(MP3) and G3S(MP2), have mean absolute deviations of 1.16 and 1.35 kcal/mol, respectively. © *2000 American Institute of Physics.*
[S0021-9606(00)30902-3]

I. INTRODUCTION

In principle it is known how to compute the thermochemical properties of most molecules to very high accuracy (0.5 kcal/mol) using quantum chemical calculations.[1–7] This can be achieved by using very high levels of correlation, such as that obtained with coupled cluster [CCSD(T)] or quadratic configuration [QCISD(T)] methods, and very large basis sets containing high angular momentum functions. The results of these calculations are then extrapolated to the complete basis set limit and corrected for some smaller effects such as core–valence and relativistic effects. Unfortunately, this approach is limited to small molecules because of the $\sim n^7$ scaling (with respect to the number of basis functions) of the correlation methods and the need for very large basis sets.

An alternative approach applicable for larger molecules is to use a series of high level correlation calculations [e.g., QCISD(T), MP4, CCSD(T)] with moderate sized basis sets to approximate the result of a more expensive calculation. The Gaussian-n series[8] exploits this idea to predict thermochemical data. In addition, molecule-independent empirical parameters are used in these methods to estimate the remaining deficiencies in the calculations. This will work if the remaining deficiencies are systematic and scale as the number of electrons. Such an approach using ''higher level corrections'' (additive parameters that depend on the number of paired and unpaired electrons in the system) has been quite successful and the latest version, Gaussian-3 (G3) theory,[9]

achieves an overall accuracy of 1 kcal/mol for the G2/97 test set.[10,11] Petersson *et al.*[12] have developed a related series of methods, referred to as complete basis set (CBS) methods, for the evaluation of accurate energies of molecular systems. The central idea in the CBS methods is an extrapolation procedure to determine the projected second-order (MP2) energy in the limit of a complete basis set. Several empirical corrections, similar in spirit to the higher level correction used in the Gn series, are added to the resulting energies in the CBS methods to remove systematic errors in the calculations

Another approach for calculating thermochemical data that has been proposed is a scaling of the calculated energy using multiplicative parameters determined by fitting to experimental data. In some early work with this approach, Truhlar, Gordon, and coworkers[13–17] suggested scaling the correlation energy to improve reaction energies and barriers. They initially introduced a procedure based on scaling external correlation effects (SEC) from multi-configuration self-consistent field (MCSCF) calculations.[13] In a related procedure they determined optimal parameters for scaling the total correlation energy from a many-body perturbation theory calculation.[14–17] The parameters were optimized for different bond breaking reactions and bond types, and the authors also presented ''standard'' values based on averaging over several different bonds. This procedure, referred to as scaling all correlation energy (SAC), proved useful for studying the potential energy surfaces of specific systems. Subsequently, Seigbahn *et al.*[18–21] presented work based on the principle

a)Electronic mail: curtiss@cmt.anl.gov

that, in a balanced treatment, roughly the same percentage of the correlation energy is obtained for every system. This method, referred to as the parametrized correlation method (PCI-X), has a single adjustable parameter X that scales the total correlation energy obtained with a modest basis set. In all of these methods the parameters are obtained from fitting to a set of well known experimental data.

Recently, Truhlar and coworkers[22–25] have suggested more elaborate schemes that combine scaling, extrapolation to infinite basis set, and fitting to a set of experimental data. The motivation in this work was to find accurate methods for calculating continuous potential energy surfaces. These methods are based on multi-coefficient (MC) fits in which the total energy is written as a linear combination of energy terms with different basis sets. The multi-coefficient correlation method (MCCM)[22] is based on coupled cluster energies using correlation consistent basis sets, the multi-coefficient G2 method (MCG2)[23] uses G2 energies, and the multi-coefficient G3 method (MCG3)[24] uses G3 energies. Up to 10 multiplicative coefficients are used in the fits. The test set of experimental data included 49 molecules for the MCCM method, 31 for the MCG2 method, and 49 for the MCG3 method. The coefficients were obtained by least squares fitting to the training set. The resulting mean absolute deviations are very good. For example, a nine parameter fit for MCG3 gives an mean absolute deviation of 0.89 kcal/mol for the 49 molecules. For comparison, G3 theory gives a mean absolute deviation of 1.01 kcal/mol for the G2/97 test set of 299 energies. Although the MCG3 results are based on a much smaller set and are based on a combination of experimental and theoretical zero-point energies, the results are intriguing and suggest that these methods are promising.

In this paper, we report on an investigation of the replacement of the (additive) higher level correction (HLC) of G3 theory by a (multiplicative) scaling of the correlation and Hartree–Fock parts of the G3 energy. The new methods are related to several other previously reported methods, including the scaling all correlation (SAC) method of Gordon and Truhlar,[14,25] the parametrized correlation (PCI-X) method of Seigbahn *et al.*,[18] and the multi-coefficient correlation method (MCCM) of Truhlar *et al.*[22–24] However, the new methods use a different parametrization and a much larger test set for optimizing the parameters. In all of this work we have used the G2/97 test set of 299 energies that was used in the assessment of G2 theory and the development of G3 theory.[9–11] This large test set should provide for a reliable assessment of the use of a scaling approach to computational thermochemistry. In Sec. II we present some theoretical background on G3 theory. In Sec. III we present a systematic investigation of scaling of the correlation and Hartree–Fock (HF) energies in G3 theory and a detailed analysis of a six-parameter method referred to as G3S. In addition, we also consider scaling approaches for two variations of G3 theory: G3(MP2) and G3(MP3), that require less computational resources. In Sec. IV conclusions are presented.

II. THEORETICAL METHODS

Gaussian-3 (G3) theory[9] is a procedure for calculating energies of molecules containing atoms of the first- and

second-row of the periodic chart. It is based on *ab initio* molecular orbital theory and is in the spirit of G2 theory.[26] G3 theory uses geometries from second-order perturbation theory [MP2(FU)/6-31G(d)] and scaled zero-point energies from Hartree–Fock theory [HF/6-31G(d)] followed by a series of single-point energy calculations at the second-order Moller–Plesset (MP2), fourth-order Moller–Plesset (MP4), and quadratic configuration interaction [QCISD(T)] levels of theory. G3 theory was assessed on a total of 299 energies (enthalpies of formation, ionization energies, electron affinities, and proton affinities) from the G2/97 test set.[10,11] The mean absolute deviation from experiment of G3 theory for the 299 energies is 1.01 kcal/mol. G3(MP3) and G3(MP2) theories[27,28] are modifications of G3 theory, that use reduced perturbation order and take less computational time than G3 theory.

Starting from an MP4/d energy, the equation for the G3 energy is

$$
\begin{aligned}
E_0(\text{G3}) = \text{MP4}/d &+ [\text{QCISD(T)}/d - \text{MP4}/d] \\
&+ [\text{MP4/plus} - \text{MP4}/d] \\
&+ [\text{MP4}/2df,p - \text{MP4}/d] \\
&+ [\text{MP2(FU)/G3L} - \text{MP2}/2df,p \\
&- \text{MP2/plus} + \text{MP2}/d] \\
&+ E(\text{SO}) + E(\text{HLC}) + E(\text{ZPE}),
\end{aligned}
\tag{1}
$$

where $d = 6\text{-}31\text{G}(d)$, plus$=6\text{-}31+\text{G}(d)$, $2df,p = 6\text{-}31\text{G}(2df,p)$, G3L=G3Large basis set,[9] $E(\text{SO})$ =spin–orbit correction for atoms only, $E(\text{HLC})$=higher level correction, and $E(\text{ZPE})$=zero-point energy correction. The effects of larger basis sets are obtained at the MP4 and MP2 levels of theory in Eq. (1) with some additivity approximations. All correlation calculations are done with a frozen core, except the MP2 calculation with the G3Large basis set that treats all electrons, i.e., it includes core correlation.

In Eq. (1) the G3 energy is written in terms of corrections (basis set extensions and correlation energy contributions) to the MP4/d energy. Alternatively, the G3 energy in Eq. (1) can be specified in terms of HF and perturbation energy components as follows:

$$
\begin{aligned}
E_0[\text{G3}] = \text{HF}/d &+ [\text{HF/G3L} - \text{HF}/d] + [\text{E2}/d + \text{E3}/d \\
&+ \text{E4}/d + \Delta\text{QCI}/d] + [\text{E2(FU)/G3L} - \text{E2}/d] \\
&+ [\text{E3/plus} - \text{E3}/d] + [\text{E3}/2df,p - \text{E3}/d] \\
&+ [\text{E4/plus} - \text{E4}/d] + [\text{E4}/2df,p - \text{E4}/d] \\
&+ E(\text{SO}) + E(\text{HLC}) + E(\text{ZPE}).
\end{aligned}
\tag{2}
$$

where E2, E3, and E4 refer to the second-, third-, and fourth-order contributions from perturbation theory, and ΔQCI refers to the contributions beyond fourth order in a QCISDT(T) calculation. Complete expressions for the different terms in Eq. (2) are given in the Appendix.

In this study we have set the HLC term to zero and optimized parameters that scale the different terms in the G3 energy expression. There are several different ways of writing and scaling the G3 energy and in the next section we

TABLE I. Results for scaling of the G3 energy.

Method	No. of parameters	Optimized scale factors[a]	MAD, kcal/mol	RMSD, kcal/mol
G3(scaled)[b]	1	$S_C = 1.0755$	1.43	1.91
G3(scaled)[c]	2	$S_C = S_{C'} = 1.0752$, $S_{HF'} = 1.0126$	1.42	1.90
G3(scaled)[c]	3	$S_C = 1.0536$, $S_{HF'} = 1.0795$, $S_{C'} = 1.1403$	1.20	1.59
G3(scaled)[d]	4	$S_{E234} = S_{QCI} = 1.0560$, $S_{HF'} = 1.0710$, $S_{E2'} = 1.1326$, $S_{E3'} = S_{E4'} = 1.3390$	1.19	1.57
G3S(scaled)[d]	5	$S_{E234} = S_{QCI} = 1.0565$, $S_{HF'} = 1.0929$ $S_{E2'} = 1.1492$, $S_{E3'} = 1.4102$, $S_{E4'} = 0.9571$	1.02	1.41
G3S[d]	6	$S_{E234} = 1.0596$, $S_{QCI} = 1.1504$, $S_{HF'} = 1.0868$ $S_{E2'} = 1.1477$, $S_{E3'} = 1.3780$, $S_{E4'} = 0.9529$	0.99	1.39

[a]Based on minimizing the root mean square deviation (RMSD). The mean absolute deviation (MAD) is also given.
[b]Equation (3).
[c]Equation (4).
[d]Equation (5).

present a systematic study of the performance of the resulting schemes as the number of parameters is increased. The parameters are optimized to give the smallest root mean square deviation from experiment for 299 energies in the G2/97 set. This is a change from our previous optimizations of parameters where the fitting has generally been based on minimizing the mean absolute deviation. However, the difference between the two fitting schemes was found to be very small (about 0.01 kcal/mol) for our test set. The full test set includes 148 enthalpies of formation of neutral molecules (at 298 K), 88 ionization potentials, 58 electron affinities, and 8 proton affinities, for a total of 302 reaction energies. In the fits reported in this paper we have used the G2/97 test set less three ionization potentials ($C_6H_5CH_3 \rightarrow C_6H_5CH_3^+$, $C_6H_5NH_2 \rightarrow C_6H_5NH_2^+$, $C_6H_5OH \rightarrow C_6H_5OH^+$) resulting in a total of 299 energies. A similar procedure is used for the G3(MP3) and G3(MP2) energies.

All calculations in this paper were done with the GAUSSIAN94 computer program.[29] Total energies and deviations from experiment for the new methods can be obtained from the internet.[30]

III. RESULTS AND DISCUSSIONS

A. G3 theory

In this section we systematically investigate the performance of scaled G3 models as we increase the number of scaling parameters. Use of a single parameter to scale the correlation energy in G3 theory can be done by modifying Eq. (2) as follows:

$$E_0[G3(scaled)]$$

$$= HF/d + [HF/G3L - HF/d]$$

$$+ S_C\{[E2/d + E3/d + E4/d + \Delta QCI/d]$$

$$+ [E2(FU)/G3L - E2/d]$$

$$+ [E3/plus - E3/d] + [E3/2df, p - E3/d]$$

$$+ [E4/plus - E4/d] + [E4/2df, p - E4/d]\}$$

$$+ E(SO) + E(ZPE). \tag{3}$$

In this equation we set E(HLC) equal to zero. The factor S_C scales all of the correlation terms. Without scaling of the correlation energy the mean absolute deviation is 6.38 kcal/mol. The results of fitting to the G2/97 test set are given in Table I and indicate that scaling the correlation energy in G3 theory in this manner gives a mean absolute deviation of 1.43 kcal/mol for the 299 energies in the G2/97 test set. Thus, scaling with a single parameter improves the accuracy significantly. The scale factor for the one parameter fit is 1.0755. This is reasonable since one expects an underestimation of the correlation energy due to slow convergence with basis set size. This single-parameter scaling of G3 is similar to the SAC method of Truhlar et al.[25] and the PCI-X method of Seigbahn et al.,[9] although their correlation energies are based on a single basis set calculation and ours is composed of additive contributions from basis set extensions. Truhlar et al.[25] have reported tests (electronic dissociation energies, D_e) of SAC for 14 different levels of theory using a 49 molecule set. The smallest mean absolute deviation (1.99 kcal/mol) is obtained for the CCSD(T)/pTZ level of theory (polarized triple zeta basis). The scaling parameter is 1.110, which is very close to the one that we find above for the 299 energy test set. The one-parameter fit of the G3 energy for the 0 K dissociation energies (D_0) of the same 49 molecules gives a mean absolute deviation of 1.37 kcal/mol. Seigbahn et al.[19] have reported a mean absolute deviation of 1.2 kcal/mol from their PCI-X method based on CCSD(T) energies with their largest basis set ($6s5p3d2f$ contracted functions on nonhydrogens and $4s3p2d$ contracted functions on hydrogens) for the dissociation energies (D_0) of a set of 28 small first-row molecules. The one-parameter fit of the G3 energy for these 28 molecules gives a mean absolute deviation of 0.97 kcal/mol. Hence, the composite G3 basis set appears to perform somewhat better than the best previous one-parameter fits.

As a second step in scaling of G3 theory we add parameters to scale the basis set extensions at the Hartree–Fock level and at the correlation level. The is done by adding two scale factors to Eq. (3) as follows:

$E_0[\text{G3(scaled)}]$

$$= \text{HF}/d + S_{\text{HF}'}[\text{HF/G3L} - \text{HF}/d] + S_C[\text{E2}/d + \text{E3}/d$$
$$+ \text{E4}/d + \Delta\text{QCI}/d] + S_{C'}\{[\text{E2(FU)/G3L} - \text{E2}/d]$$
$$+ [\text{E3/plus} - \text{E3}/d] + [\text{E3}/2df,p - \text{E3}/d]$$
$$+ [\text{E4/plus} - \text{E4}/d] + [\text{E4}/2df,p - \text{E4}/d]\}$$
$$+ \text{E(SO)} + \text{E(ZPE)}. \tag{4}$$

$E_0[\text{G3S}]$

$$= \text{HF}/d + S_{\text{E234}}[\text{E2}/d + \text{E3}/d + \text{E4}/d] + S_{\text{QCI}}[\Delta\text{QCI}/d]$$
$$+ S_{\text{HF}'}[\text{HF/G3L} - \text{HF}/d] + S_{\text{E2}'}[\text{E2(FU)G3L} - \text{E2}/d]$$
$$+ S_{\text{E3}'}\{[\text{E3/plus} - \text{E3}/d] + [\text{E3}/2df,p - \text{E3}/d]\}$$
$$+ S_{\text{E4}'}\{[\text{E4/plus} - \text{E4}/d] + [\text{E4}/2df,p - \text{E4}/d]\}$$
$$+ \text{E(SO)} + \text{E(ZPE)}. \tag{5}$$

The parameter $S_{\text{HF}'}$ scales the basis set extension [i.e., effects beyond 6-31G(d)] at the Hartree–Fock level while the parameter $S_{C'}$ scales all such basis set extensions at the correlated level. Since the most significant basis set extensions are due to higher angular momentum functions, these parameters take into account the underestimation of their contributions at the HF and correlated levels, respectively. The results in Table I indicate that adding the $S_{\text{HF}'}$ parameter reduces the mean absolute deviation to 1.42 kcal/mol, while a further addition of the $S_{C'}$ parameter reduces it to 1.20 kcal/mol. The substantial improvement from the $S_{C'}$ parameter is consistent with the large effect of higher angular momentum functions on electron correlation. The values for $S_{\text{HF}'}$ and $S_{C'}$ are 1.0795 and 1.1403, respectively, when optimized together. Overall, the three-parameter fit gives a mean absolute deviation that is significantly better than given by G2 theory for the same test set (1.48 kcal/mol), but not as good as G3 theory with the HLC parametrization (1.01 kcal/mol).

As a third step in scaling of G3 theory we add parameters that scale the basis set extensions at different orders of perturbation theory individually. In addition, we scale the $\Delta\text{QCI}/d$ part of the 6-31G(d) correlation separately. These additional parameters are designed to reflect the different convergence behavior of the different correlation terms. Eq. (4) is modified as follows:

The scale factors for the basis set extension terms ($S_{\text{HF}'}$, $S_{\text{E2}'}$, $S_{\text{E3}'}$, $S_{\text{E4}'}$) are denoted by primes, the scale factor for the second-, third-, and fourth-order perturbation terms at the 6-31G(d) level is denoted by S_{E234}, and the scale factor for the QCI correction beyond MP4 at the 6-31G(d) level is denoted by S_{QCI}.

Optimization of all six parameters in Eq. (5) gives a mean absolute deviation of 0.99 kcal/mol, which is slightly better than standard G3 theory with the HLC correction (mean absolute deviation of 1.01 kcal/mol). The splitting of the basis set extensions for the E3 and E4 terms gives the largest reduction in the mean absolute deviation when the parameters are increased from four to six (see Table I). The optimized values for the parameters in the six-parameter fit are all of reasonable magnitude and range from 0.95 to 1.38. The largest scale factor occurs for the basis set extensions at the third order of perturbation theory. Only one scale factor is less than unity—the scale factor for the basis set extensions at fourth-order perturbation theory (0.95). Thus, it is possible to obtain a very accurate version of G3 theory with scaling of energies when the basis set extensions are included in the fitting procedure. We refer to the energy given by Eq. (5) with all six parameters optimized as the G3S energy. We have investigated the dependence of these results on the number of parameters and have found little improvement on a further increase of the number of parameters. The addition of scale factors to all 11 terms in Eq. (5) reduces the

TABLE II. Summary of results for the G3 methods with scaling and with the higher level correction for the different types of energies in the G2/97 test set.

Type	Mean absolute deviation (in kcal/mol) from experiment					
	G3[a]	G3S	G3(MP3)[b]	G3S(MP3)	G3(MP2)[c]	G3S(MP2)
Enthalpies of formation (148)	0.94	0.97	1.20	1.07	1.18	1.23
Nonhydrogen (35)	1.72	1.65	2.13	1.89	2.12	2.02
Hydrocarbons (22)	0.68	0.74	0.86	0.95	0.70	0.80
Subst. hydrocarbons (47)	0.56	0.75	0.78	0.64	0.74	0.90
Inorganic hydrides (15)	0.87	0.63	1.18	0.79	1.03	1.06
Radical (29)	0.84	0.86	1.05	1.01	1.23	1.21
Ionization energies (85)	1.13	1.08	1.22	1.26	1.41	1.49
Electron affinities (58)	0.98	0.90	1.24	1.24	1.46	1.56
Proton affinities (8)	1.34	1.17	1.25	1.10	1.02	0.74
All (299)[d]	1.01	0.99	1.22	1.16	1.30	1.35

[a]Reference 9.
[b]Reference 27.
[c]Reference 28
[d]Root mean square deviations (in kcal/mol) are G3=1.45, G3S=1.39, G3 (MP3)=1.71, G3S (MP3)=1.61, G3 (MP2)=1.81, G3S (MP2)=1.87.

TABLE III. A comparison of results for larger molecules that are not in the G2/97 test set.

Species	Expt—Theory (ΔH_f^0), kcal/mol					
	G3	G3S	G3(MP3)	G3S(MP3)	G3(MP2)	G3S(MP2)
Naphthalene	0.52	0.11	0.77	2.52	3.22	3.15
Azulene	−1.59	−1.78	−1.69	0.40	1.44	1.63
Benzoquinone	−1.11	−1.00	−1.57	−0.18	−0.78	−0.46
Chlorobenzene	0.24	−0.55	0.63	0.92	2.11	2.18

mean absolute deviation only slightly to 0.97 kcal/mol.

The results for the G3S method broken up into different types of energies (enthalpies of formation, ionization energies, etc.) as well as different types of molecules (hydrocarbons, radicals, etc.) are given in Table II. The performance of G3S is similar to that of G3 in all cases. The G3S method was also assessed on the enthalpies of formation of four larger molecules, naphthalene, azulene, benzoquinone, and chlorobenzene, as a check on problems that may occur upon extending the method to molecules not included in the training set. The results are given in Table III and indicate that G3S does as well as G3. The G3S method gives deviations from the experiment of 0.11, −1.78, −1.00, and −0.55 kcal/mol, respectively, for the four molecules. This compares well with the G3 deviations of 0.52, −1.59, −1.11, and 0.24 kcal/mol, respectively.

The energy expression in Eq. (5) is somewhat similar to the multi-coefficient energy expression proposed by Truhlar and coworkers[24] for G3 theory (MCG3) in that they also parameterize the basis set extensions. However, they use nine parameters in their fit. In addition they do not divide the higher-order correlation energies into E3, E4, and ΔQCI parts, but rather into a term that includes E3 plus the sdq (singles, doubles, and quadruples) part of E4, a term that contains only the triples part of E4, and a ΔQCI term. They also do not include the diffuse function basis set extension [6-31+G(d)] or the explicit calculation of the core-correlation. They replace the core-correlation in G3 theory by a simple estimate. Truhlar and coworkers[24] reported a mean absolute deviation of 0.89 kcal/mol for their MCG3 method. This method is based on fitting to the dissociation energies (D_e) of 49 small molecules. We have used their energy expression for the 299 energies of the G2/97 test set and find a mean absolute deviation of 2.30 kcal/mol. In this assessment we have applied their core-correlation estimate technique to ions in the test set without modification and neglected spin–orbit corrections for molecules. The best results are found for neutral enthalpies of formation (1.49 kcal/mol) and the worst for electron affinities (4.93 kcal/mol). If the nine parameters in the MCG3 expression are re-optimized the mean absolute deviation reduces to 1.14 kcal/mol, including 1.10 kcal/mol for enthalpies of formation and 1.22 kcal/mol for electron affinities. These results are substantially improved, although not as good as those from G3S. Our study also shows that use of parameters derived from a small training set may lead to large errors when applied to other systems.

B. G3(MP3) theory

The G3(MP3) energy[27] is a modification of G3 theory that eliminates the expensive MP4/2df,p calculation by evaluating the larger basis set effects at the MP3 and MP2 levels of theory. It also eliminates the MP4/+ calculation:

$$E_0[\text{G3(MP3)}] = \text{MP4}/d + [\text{QCISD(T)}/d - \text{MP4}/d]$$
$$+ [\text{MP3}/2df,p - \text{MP3}/d]$$
$$+ [\text{MP2(FU)/G3L} - \text{MP2}/2df,p]$$
$$+ E(\text{SO}) + E(\text{HLC}) + E(\text{ZPE}). \quad (6)$$

The G3(MP3) energy can be derived using scaling factors analogous to the G3S energy in Eq. (5) by setting $S_{E4'}$ to zero and the [E3/plus−E3/d] term to zero. The resulting energy equation is

TABLE IV. Results for scaling of the G3(MP3) energy.

Method	No. of parameters	Optimized scale factors[a,b]	MAD, kcal/mol	RMSD, kcal/mol
G3(MP3, scaled)	1	$S_{E234} = S_{QCI} = S_{E2'} = S_{E3'} = 1.0814$ ($S_{HF'} = 1.0$)	1.53	2.00
G3(MP3, scaled)	2	$S_{E234} = S_{QCI} = S_{E2'} = S_{E3'} = 1.0808$, $S_{HF'} = 1.0200$	1.50	1.97
G3(MP3, scaled)	3	$S_{E234} = S_{QCI} = 1.0592$, $S_{HF'} = 1.0889$, $S_{E2'} = S_{E3'}$ $= 1.1477$	1.20	1.66
G3S(MP3, scaled)	4	$S_{E234} = S_{QCI} = 1.0588$, $S_{HF'} = 1.0895$, $S_{E2'} = 1.1483$, $S_{E3'} = 1.1310$	1.20	1.66
G3S(MP3)	5	$S_{E234} = 1.0631$, $S_{QCI} = 1.1916$, $S_{HF'} = 1.0823$ $S_{E2'} = 1.1471$, $S_{E3'} = 1.0972$	1.16	1.61

[a]Based on minimizing the root mean square deviation (RMSD). The mean absolute deviation (MAD) is also given. Values in parentheses are held fixed.
[b]Equation (7).

TABLE V. Results for scaling of the G3(MP2) energy.

Method	No. of parameters	Optimized scale factors[a,b]	MAD, kcal/mol	RMSD, kcal/mol
G3(MP2,scaled)	1	$S_{E2}=S_{E34}=S_{QCI}=S_{E2'}=1.0952$, $(S_{HF}=S_{HF'}=1.0)$	2.07	2.67
G3(MP2,scaled)	2	$S_{E2}=S_{E34}=S_{QCI}=S_{E2'}=1.0948$, $S_{HF'}=1.0154$ $(S_{HF}=1.0)$	2.07	2.66
G3(MP2,scaled)	3	$S_{E2}=S_{E34}=S_{QCI}=1.0650$, $S_{E2'}=1.1926$, $S_{HF'}=1.1143$ $(S_{HF}=1.0)$	1.66	2.20
G3(MP2,scaled)	4	$S_{E2}=S_{E34}=1.0738$, $S_{QCI}=1.3205$, $S_{E2'}=1.1890$, $S_{HF'}=1.0992$ $(S_{HF}=1.0)$	1.56	2.06
G3S(MP2)	6	$S_{HF}=1.0049$, $S_{E2}=1.0694$, $S_{E34}=1.1694$, $S_{QCI}=1.2320$, $S_{E2'}=1.1553$, $S_{HF'}=1.0880$	1.35	1.87

[a]Based on minimizing the root mean square deviation (RMSD). The mean absolute deviation (MAD) is also given. Parameters in parentheses are held fixed.
[b]Equation (9).

$$E_0[\text{G3S(MP3)}]$$
$$= \text{HF}/d + S_{E234}[\text{E2}/d + \text{E3}/d + \text{E4}/d] + S_{QCI}[\Delta\text{QCI}/d]$$
$$+ S_{HF'}[\text{HF}/\text{G3L} - \text{HF}/d] + S_{E2'}[\text{E2(FU)}/\text{G3L} - \text{E2}/d]$$
$$+ S_{E3'}\{[\text{E3}/2df,p - \text{E3}/d]\} + \text{E(SO)} + \text{E(ZPE)}. \tag{7}$$

The optimized results for systematically increasing the number of parameters from 1 to 5 are given in Table IV. The five-parameter fit has a mean absolute deviation of 1.16 kcal/mol compared to 1.22 kcal/mol for G3(MP3). We refer to the five parameter fit as G3S(MP3). All of the scale factors for G3S(MP3) are close to unity. A summary of the G3S(MP3) results for the G2/97 test set is given in Table II and results for the four larger molecules are given in Table III.

C. G3(MP2) theory

The G3(MP2) energy[28] is a modification of G3 theory, similar to G2(MP2) theory,[29] that evaluates the larger basis set effects at the MP2 level:

$$E_0[\text{G3(MP2)}] = \text{E(MP4}/d) + [\text{QCISD(T)}/d - \text{MP4}/d]$$
$$+ [\text{MP2(FC)}/\text{G3MP2L} - \text{MP2}/d]$$
$$+ \text{E(SO)} + \text{E(HLC)} + \text{E(ZPE)}. \tag{8}$$

In addition, in G3(MP2) theory the MP2(FU) calculation is replaced by a frozen core calculation (FC) with the G3MP2Large basis set.[28]

It was not possible to scale the G3(MP2) energy expression analogous to G3S and G3S(MP3) and obtain an accuracy similar to that obtained by use of the higher level correction. This is shown by the results in Table V. A four-parameter fit [two parameters for correlation at the 6-31G(d) level and two parameters for the basis set extensions] gives a mean absolute deviation of 1.56 kcal/mol, which is much worse than G3(MP2) theory (1.30 kcal/mol). However, significantly better results were obtained by adding scale factors to the HF/d and E2 terms as follows

$$E_0[\text{G3S(MP2)}] = S_{HF}[\text{HF}/d] + S_{E2}[\text{E2}/d]$$
$$+ S_{E34}[\text{E3}/d + \text{E4}/d] + S_{QCI}[\Delta\text{QCI}/d]$$
$$+ S_{HF'}[\text{HF}/\text{G3L} - \text{HF}/d]$$
$$+ S_{E2'}[\text{E2(FC)}/\text{G3MP2L} - \text{E2}/d]$$
$$+ \text{E(SO)} + \text{E(ZPE)}. \tag{9}$$

The six-parameter fit for G3S(MP2) gives a mean absolute deviation of 1.35 kcal/mol, only slightly larger than the 1.30 kcal/mol obtained using the higher level correction. A summary of the G3S(MP2) results for the G2/97 test set is given in Table II and results for the four larger molecules are given in Table III.

Finally, in Table VI we present the mean deviations for the three G3S methods presented in this study as well as the three corresponding G3 methods. The mean deviation is a

TABLE VI. A summary of mean deviations for G3 and G3S methods for the different types of energies in the G2/97 test set.

Type	Mean deviation (in kcal/mol) from experiment					
	G3	G3S	G3(MP3)	G3S(MP3)	G3(MP2)	G3S(MP2)
Enthalpies of formation (148)	0.04	−0.15	0.02	−0.19	0.05	−0.12
Nonhydrogen (35)	0.22	−0.13	0.20	−0.32	0.11	0.58
Hydrocarbons (22)	−0.26	−0.09	0.19	0.48	0.08	0.02
Subst. hydrocarbons (47)	0.06	0.15	−0.14	0.05	−0.05	−0.08
Inorganic hydrides (15)	−0.09	−0.20	−0.16	−0.40	0.04	−0.21
Radical (29)	0.08	−0.67	0.01	−0.83	0.11	−1.10
Ionization energies (85)	−0.25	−0.03	−0.39	0.22	−0.17	−0.19
Electron affinities (58)	0.25	−0.39	0.40	−0.35	0.07	−0.25
Proton affinities (8)	1.06	1.08	0.97	1.08	0.78	−0.04

measure of errors that give systematic overbinding or underbinding. The only subgroups that have mean deviations of significant magnitude are the enthalpies of formation of the radicals and the proton affinities. The calculated proton affinities tend to be too small for all methods [except G3(MP2)] and the radicals tend to be underbound by the G3S methods, but not the G3 methods.

IV. CONCLUSIONS

In this paper we have presented G3 theory using multiplicative scale factors instead of an additive higher level correction. The HF basis set extension contribution is scaled by one parameter along with a five-parameter scaling of the correlation energy terms of G3 theory. This method, G3S, is based on a six-parameter fit to the G2/97 test set and has a mean absolute deviation of 0.99 kcal/mol, slightly better than G3 theory. An assessment on four molecules not in the G2/97 test set (naphthalene, azulene, benzoquinone, and chlorobenzene) indicates that these scaling methods give accurate results for these larger species as well. In addition, we have presented versions of the computationally less intensive G3(MP3) and G3(MP2) methods that use scaled energies. These methods, referred to as G3S(MP3) and G3S(MP2), have mean absolute deviations of 1.16 and 1.35 kcal/mol, respectively.

The G3S method has the advantage compared to G3 theory in that it can be used for studying potential energy surfaces where the products and reactants have a different number of paired electrons. G3 theory cannot be used in this case because of the use of the higher level correction, which depends on the number of electron pairs. We note that the spin–orbit correction in G3 theory is only for atoms and therefore this term will be discontinuous in a potential energy surface that involves breaking a molecule into one or more atoms. For systems where this correction is significant it would be necessary to calculate (or estimate) the spin–orbit correction. In addition, we have used MP2 equilibrium geometries in the determination of the parameters, although ideally one should use G3 geometries. Overall, the scaling approach to G3 theory is promising and it may lead to improved methods for calculating thermochemical data and for the exploration of potential energy surfaces.

APPENDIX: DEFINITIONS OF SCALED ENERGY TERMS

Listed below are the energies used to obtain the G3 terms in Eqs. (2)–(5), the G3(MP3) terms in Eq. (7), and the G3(MP2) terms in Eq. (9). All of the G3 terms can be obtained from four single-point calculations, QCISD(T,E4T)/6-31G(d), MP2(FU)/G3Large, MP4/6-31+G(d), and MP4/6-31G($2df,p$):

$$HF/d = E[HF/6\text{-}31G(d)],$$

$$E2/d = E[MP2/6\text{-}31G(d)] - E[HF/6\text{-}31G(d)],$$

$$E3/d = E[MP3/6\text{-}31G(d)] - E[MP2/6\text{-}31G(d)],$$

$$E4/d = E[MP4/6\text{-}31G(d)] - E[MP3/6\text{-}31G(d)],$$

$$\Delta QCI/d = E[QCISD(T)/6\text{-}31G(d)] - E[MP4/6\text{-}31G(d)],$$

$$HF/G3L = E[HF/G3Large],$$

$$E2(FU)/G3L = E[MP2(FU)/G3Large] - E[HF/G3Large],$$

$$E2(FC)/G3MP2L = E[MP2(FC)/G3MP2Large]$$
$$- E[HF/G3MP2Large],$$

$$E3/plus = E[MP3/6\text{-}31+G(d)] - E[MP2/6\text{-}31+G(d)],$$

$$E3/2df,p = E[MP3/6\text{-}31G(2df,p)]$$
$$- E[MP2/6\text{-}31G(2df,p)],$$

$$E4/plus = E[MP4/6\text{-}31+G(d)] - E[MP3/6\text{-}31+G(d)],$$

$$E4/2df,p = E[MP4/6\text{-}31G(2df,p)]$$
$$- E[MP3/6\text{-}31G(2df,p)].$$

[1] K. A. Peterson and T. H. Dunning, Jr., J. Chem. Phys. **106**, 4119 (1997).
[2] D. Feller and K. A. Peterson, J. Chem. Phys. **108**, 154 (1998).
[3] R. S. Grev, C. L. Janssen, and H. F. Schaefer, III, J. Chem. Phys. **97**, 8389 (1992).
[4] J. A. Montgomery, Jr., J. W. Ochterski, and G. A. Petersson, J. Chem. Phys. **101**, 5900 (1994).
[5] C. W. Bauschlicher, Jr. and S. R. Langhoff, Science **254**, 394 (1991).
[6] J. M. L. Martin, in *Computational Thermochemistry*, ACS Symposium Series, edited by K. K. Irikura and D. J. Frurip (American Chemical Society, Washington, DC, 1998), Vol. 677, pp. 212–236.
[7] P. L. Fast, M. L. Sanchez, and D. G. Truhlar, J. Chem. Phys. **111**, 2921 (1999).
[8] For recent reviews see L. A. Curtiss and K. Raghavacharin, in Ref. 6, pp. 176–197; L. A. Curtiss and K. Raghavachari, in *Encyclopedia of Computational Chemistry*, edited by P. v. R. Schleyer (Wiley, New York, 1998).
[9] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, J. Chem. Phys. **109**, 7764 (1998).
[10] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **106**, 1063 (1997).
[11] L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, J. Chem. Phys. **109**, 42 (1998).
[12] G. A. Petersson, in Ref. 6, pp. 237–266.
[13] F. B. Brown and D. G. Truhlar, Chem. Phys. Lett. **117**, 307 (1985).
[14] M. S. Gordon and D. G. Truhlar, J. Am. Chem. Soc. **108**, 5412 (1986).
[15] M. S. Gordon and D. G. Truhlar, Int. J. Quantum Chem. **XXXI**, 81 (1987).
[16] M. S. Gordon, K. A. Nguyen, and D. G. Truhlar, J. Phys. Chem. **93**, 7356 (1989).
[17] I. Rossi and D. G. Truhlar, Chem. Phys. Lett. **234**, 64 (1995).
[18] P. E. M. Siegbahn, R. A. M. Blomberg, and M. Svensson, Chem. Phys. Lett. **223**, 35 (1994).
[19] P. E. M. Siegbahn, R. A. M. Blomberg, and M. Svensson, J. Chem. Phys. **102**, 5377 (1995).
[20] R. A. M. Blomberg, P. E. M. Siegbahn, and M. Svensson, J. Chem. Phys. **104**, 9546 (1996).
[21] P. E. M. Siegbahn, *Advances in Chemical Physics* (Wiley, New York, 1996), Vol. XCII, p. 333.
[22] P. L. Fast, J. C. Corchado, M. L. Sanchez, and D. G. Truhlar, J. Phys. Chem. A **103**, 5129 (1999).

[23] P. L. Fast, M. L. Sanchez, J. C. Corchado, and D. G. Truhlar, J. Chem. Phys. **110**, 11679 (1999).

[24] P. L. Fast, M. L. Sanchez, and D. G. Truhlar, Chem. Phys. Lett. **306**, 407 (1999).

[25] P. L. Fast, J. C. Corchado, M. L. Sanchez, and D. G. Truhlar, J. Phys. Chem. A **103**, 3139 (1999).

[26] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).

[27] L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, Chem. Phys. Lett. **313**, 600 (1999).

[28] L. A. Curtiss, P. C. Redfern, K. Raghavachari, V. Rassolov, and J. A. Pople, J. Chem. Phys. **110**, 4703 (1999).

[29] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 94, Gaussian, Inc. Pittsburgh, PA, 1995.

[30] The total energies of the molecules and a complete tabulation of deviations from experiment can be found at the website http://chemistry.anl.gov/compmat/g3theory.htm.