

JCTC

Journal of Chemical Theory and Computation

G4(MP2)-6X: A Cost-Effective Improvement to G4(MP2)

Bun Chan,* Jia Deng,[†] and Leo Radom*

School of Chemistry and ARC Center of Excellence for Free Radical Chemistry and Biotechnology, University of Sydney, Sydney, NSW 2006, Australia

Received September 21, 2010

Abstract: G4(MP2)-6X is developed as a composite procedure with a cost comparable to that of G4(MP2) but performance approaching that of G4. The new procedure is a variant of G4(MP2) that employs BMK/6-31+G(2df,p) geometries and has six additional scaling factors for the correlation energy components. The scaling factors and HLC parameters are optimized using the new E2 set of 526 energies, representing thermochemical properties, reaction energies and barriers, and weak interactions. G4(MP2)-6X achieves a mean absolute deviation (MAD) from benchmark values of 3.64 kJ mol⁻¹ for the E2 set, compared with 4.42 kJ mol⁻¹ for G4(MP2). For the E0 set of 148 energies, G4(MP2)-6X gives an MAD of 3.43 kJ mol⁻¹, compared with 3.22 kJ mol⁻¹ for G4 and 4.03 kJ mol⁻¹ for G4(MP2). The new G4(MP2)-6X procedure thus uses extra parametrization to provide a G4-type performance without incurring G4-type computational costs.

1. Introduction

Composite quantum chemistry methods¹ enable the accurate prediction of thermochemical properties at reduced cost in terms of computing resources. A variety of such procedures have been developed, including the Gaussian (Gn) procedures,² the complete-basis-set (CBS) methods,³ and the Weizmann (Wn) procedures.⁴ These methods range from highly accurate at modest computational cost to exceptionally accurate at much higher expense. For instance, G4 achieves a mean absolute deviation (MAD) from experimental values of 3.47 kJ mol⁻¹ (0.83 kcal mol⁻¹) for the G3/05 set⁵ of 454 energies, using MP2, MP4, and CCSD(T) component calculations of moderate individual cost. At the other end of the spectrum, by employing coupled cluster calculations up to CCSDTQ5, W4^{4c} gives an MAD of just 0.42 kJ mol⁻¹ (0.1 kcal mol⁻¹) for the W3 set^{4b} of 36 energies.

For many chemical systems of medium size, the Wn procedures, and to some extent the Gn procedures, are still prohibitively expensive. To address this problem, the Gn-(MP2) procedures have been formulated.⁶ These variants of Gn are less computationally demanding and are therefore

applicable to a much wider range of systems. The trade-off for the reduced computational requirements is that the Gn(MP2) procedures are somewhat less accurate than the Gn methods. For example, G4(MP2)^{6c} has an MAD of 4.35 kJ mol⁻¹ (1.04 kcal mol⁻¹) for the G3/05 set, compared with 3.47 kJ mol⁻¹ for G4.

The broader-ranging applicability of G4(MP2) makes it attractive to seek further improvement in its performance. Is it feasible to achieve G4-type performance while retaining G4(MP2)-type computational cost, perhaps through additional parametrization? In the present study, we examine various modifications of G4(MP2) with this end in mind and propose G4(MP2)-6X as a cost-effective improved procedure.

2. Computational Details

Standard ab initio molecular orbital theory and density functional theory (DFT) calculations⁷ were carried out with the Gaussian 03,⁸ Gaussian 09,⁹ Molpro 2006,¹⁰ and Psi 3¹¹ programs. Unless otherwise noted, geometries were optimized with the BMK¹² DFT procedure using the 6-31+G(2df,p) basis set. In specific cases, zero-point vibrational energies (ZPVEs) and thermal corrections to enthalpy (ΔH) at 298 K, derived from scaled BMK/6-31+G(2df,p) frequencies, were incorporated into the total energies.^{13,14} Single-point energies were obtained at the HF, MP2, and CCSD(T) levels for the composite procedures. Unless otherwise noted,

* Corresponding authors e-mail: chan_b@chem.usyd.edu.au; radom@chem.usyd.edu.au.

[†] Current address: Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia.

energies are reported in kilojoules per mole. In order to maximize both the accuracy and precision of the new procedures, i.e., to minimize the deviations (maximize accuracy) and the variations in the deviations (maximize precision), we simultaneously minimize the mean absolute deviation (MAD) from the benchmark values, and the standard deviation (SD) of the deviations. Preliminary calculations indicate that this is advantageous compared with the typical approach of minimizing the MAD alone. Thus, the MAD/SD procedure leads to a lower SD and yields fewer outliers (deviations $> 8.37 \text{ kJ mol}^{-1}$ (2 kcal mol^{-1})) without changing the MAD by more than 0.01 kJ mol^{-1} . In the course of these investigations, we have briefly examined various approaches for the simultaneous minimization of the MAD and SD, including minimizing the mean of the MAD and SD values [$1/2 \times (\text{MAD} + \text{SD})$], their root-mean-square value [$2^{-1/2} \times (\text{MAD}^2 + \text{SD}^2)^{1/2}$], and their harmonic mean [$2 \times (\text{MAD}^{-1} + \text{SD}^{-1})^{-1}$]. We find that the three averaging procedures give very similar results, and we have chosen for simplicity to use the mean of MAD and SD in our parametrization processes.

3. Theory

3.1. Description of G4(MP2). The G4(MP2) procedure seeks to approximate a high-level energy (CCSD(T) with a large basis set) through the use of a series of lower-level energies and an additivity scheme. While the details of the theory have been described in the original paper,^{6c} we provide here a brief summary for comparison purposes. The G4(MP2) energy is obtained in the following manner:

$$\text{G4(MP2)} = \text{HF/CBS} + E_{\text{MP2}}^{\text{corr}}/\text{G3MP2LargeXP} + \Delta E_{\text{CCSD(T)}/6-31\text{G(d)}} + \text{HLC} + \text{ZPVE} + E_{\text{SO}} \quad (1)$$

• The geometry is obtained at the B3-LYP/6-31G(2df,p) level.

• HF/CBS is an estimate of the Hartree–Fock-limit energy and is obtained by extrapolation to the complete-basis-set limit with aug-cc-pV(n+d)Z ($n = \text{T, Q}$) basis sets modified by reducing the number of diffuse and polarization functions, using the formula $E_{\text{CBS}} = [E_{\text{Q}} - E_{\text{T}} \exp(-1.63)]/[1 - \exp(-1.63)]$.

• The frozen-core approximation is employed for all correlation calculations. More specifically, the largest noble-gas core is frozen, except that the following are treated as valence orbitals:

- 3d orbitals on third-row main-group elements (Ga–Kr)
- 2s and 2p orbitals on Na and Mg and 3s and 3p orbitals on K and Ca

• $E_{\text{MP2}}^{\text{corr}}$ is the MP2 correlation energy (i.e., $E_{\text{MP2}} - E_{\text{HF}}$), calculated with the G3MP2LargeXP basis set.

• $\Delta E_{\text{CCSD(T)}}$ is the CCSD(T) correlation energy beyond MP2 (i.e., $E_{\text{CCSD(T)}}^{\text{corr}} - E_{\text{MP2}}^{\text{corr}}$), calculated with the 6-31G(d) basis set.

• HLC is a higher-level-correction term that depends on the number of α and β valence electrons (n_{α} and n_{β}). It is obtained with parameters (A, A', B, C, D , and E , mHartree) optimized using the G3/05 set:

- $-A n_{\beta}$ for closed-shell molecules
- $-A' n_{\beta} - B (n_{\alpha} - n_{\beta})$ for open-shell molecules
- $-C n_{\beta} - D (n_{\alpha} - n_{\beta})$ for atomic species
- $-E n_{\beta}$ for “single electron pair” species (e.g., Li_2)
- $A = 9.472, A' = 9.769, B = 3.179, C = 9.741, D = 2.115, E = 2.379$

• The zero-point vibrational energy (ZPVE) and thermal corrections to enthalpy are obtained with scaled (0.9854) B3-LYP/6-31G(2df,p) frequencies.

• A spin–orbit correction (E_{SO}), where available from experimental results or from accurate calculations, is included.

The sum of the first three terms in eq 1, HF/CBS + $E_{\text{MP2}}^{\text{corr}}/\text{G3MP2LargeXP} + \Delta E_{\text{CCSD(T)}/6-31\text{G(d)}}$, is designed to approximate the electronic energy at the CCSD(T)/CBS level. The HLC, through parametrization to experimental data, is used to compensate for the remaining deficiencies in the theory and in the additivity scheme.

The form of G4(MP2) indicates several components that determine its accuracy. Needless to say, an important factor is the quality of the electronic structure methods that are used for geometry optimization and in the approximation of CCSD(T)/CBS. The HLC has no effect on the energy of reactions that involve closed-shell reactants and closed-shell products. However, for radical reactions and some properties such as heats of formation, ionization energies, and electron affinities, the HLC can substantially improve the result. Thus, both the form of the HLC and the experimental data employed in the fitting are important for many properties and reactions.

3.2. G4(MP2)-6X: An Alternative Composite Procedure.

While the use of the HLC in G4(MP2) is generally very successful in improving the prediction of thermochemical properties, other approaches have also been employed to improve the performance of a simple additivity scheme. For instance, the G3S¹⁵ and MCCM/3¹⁶ procedures scale the energy of the various components in the additivity scheme, in place of using HLC corrections. In the present study, we introduce the G4(MP2)-6X procedure, which includes such scaling as well as other modifications designed to provide an improvement to G4(MP2) but without significantly affecting the computational cost.

G4(MP2)-6X has six extra parameters that are independent of the number of electrons, in addition to the six HLC parameters in the original G4(MP2). These are introduced in the hope that they will not only improve the general accuracy of the procedure but also provide an improvement in the cases where the HLC does not contribute to the relative energies as, for example, for barriers for closed-shell reactions.

The form of the six additional parameters is loosely based on the G3S procedure,¹⁵ where the correlation contributions for MP2, CCSD, and the perturbative triples correction (T) are separately scaled. In addition, we make use of a modified MP2 method, namely, SCS-MP2, which has been found by Grimme to show improved performance over standard MP2.¹⁷ This is achieved by splitting the MP2 correlation energy ($E_{\text{MP2}}^{\text{corr}}$) into opposite-spin (OS) and same-spin (SS) contributions and applying individual scaling to the two components.¹⁷ In G4(MP2)-6X, we also scale the OS and

SS contributions to $E_{\text{MP2}}^{\text{corr}}$ separately. With a similar philosophy to that for SCS-MP2, Sherrill and co-workers have introduced the SCS-CCSD method.¹⁸ However, our preliminary investigations show that separate scaling of the OS and SS contributions to the CCSD correlation energy, in the context of G4(MP2)-type procedures, does not lead to a noticeable improvement, and we therefore do not apply SCS-CCSD to G4(MP2)-6X. Finally, we use an improved procedure for geometry optimization (BMK/6-31+G(2df,p)). The description of the G4(MP2)-6X procedure is then as follows:

$$\begin{aligned} \text{G4(MP2)-6X} = & \text{HF/CBS} + E_{\text{SCS-MP2}}^{\text{corr}}/\text{G3MP2LargeXP} + \\ & \Delta E_{\text{S-CCSD}}/6\text{-31G(d)} + E_{\text{S-(T)}}^{\text{corr}}/6\text{-31G(d)} + \\ & \text{HLC} + \text{ZPVE} + E_{\text{SO}} \end{aligned} \quad (2)$$

- The geometry is obtained at the BMK/6-31+G(2df,p) level.

- HF/CBS is obtained in the same manner as for G4(MP2).¹⁹

- The following correlation energies are obtained with the G4 frozen core:

- $c1 \cdot E_{\text{MP2OS}}^{\text{corr}}$ and $c2 \cdot E_{\text{MP2SS}}^{\text{corr}}$ are the scaled OS and scaled SS contributions to the MP2/6-31G(d) correlation energy, respectively.

- $c3 \cdot E_{\text{MP2OS}}^{\text{corr}}$ and $c4 \cdot E_{\text{MP2SS}}^{\text{corr}}$ are the scaled OS and scaled SS contributions to the MP2/G3MP2LargeXP correlation energy, respectively.

- $c5 \cdot E_{\text{CCSD}}^{\text{corr}}$ is the scaled CCSD contribution to the CCSD(T)/6-31G(d) correlation energy.

- $c6 \cdot E_{\text{(T)}}^{\text{corr}}$ is the scaled perturbative triples contribution to the CCSD(T)/6-31G(d) correlation energy.

- $E_{\text{SCS-MP2}}^{\text{corr}}/\text{G3MP2LargeXP} = c3 \cdot E_{\text{MP2OS}}^{\text{corr}} + c4 \cdot E_{\text{MP2SS}}^{\text{corr}}$

- $\Delta E_{\text{S-CCSD}}/6\text{-31G(d)} = c5 \cdot E_{\text{CCSD}}^{\text{corr}} - (c1 \cdot E_{\text{MP2OS}}^{\text{corr}} + c2 \cdot E_{\text{MP2SS}}^{\text{corr}})$

- $E_{\text{S-(T)}}^{\text{corr}}/6\text{-31G(d)} = c6 \cdot E_{\text{(T)}}^{\text{corr}}$

- $c1, c2, c3, c4, c5,$ and $c6$ are parameters optimized using the E2 training set; see section 3.3 for details of E2.

- The HLC has the same form as that for G4(MP2), but the HLC parameters are reoptimized.

- Scaled BMK/6-31+G(2df,p) frequencies are used to obtain the ZPVE (0.9770) and thermal corrections to enthalpy (0.9627).^{13,14}

- The same spin-orbit contribution as in G4(MP2) is included.

A script for running G4(MP2)-6X calculations using Gaussian 09 is included in the Supporting Information.

We note that several singlet species (C_2 , Be_2 , and CN^+) among our training sets exhibit RHF to UHF instabilities. This suggests that they have substantial biradical character and are not adequately described by the RHF wave functions. We find that in these cases, the use of an unrestricted wave function leads to improved agreement with benchmark thermochemistry. However, the broken-spin-symmetry UHF wave functions in these cases are found to be highly spin-contaminated. While the improvements with UHF are encouraging, a multireference treatment is likely to help further. Nonetheless, it is advisable that an initial stability

test on the reference wave function be conducted to identify potential problems of this type, and to provide some improvement.

3.3. Training and Test Sets. In developing new composite procedures, training sets are employed to optimize the empirical parameters, while test sets are used to evaluate their performance. There are numerous training and test sets that have been previously used. For example, the G3/05 set⁵ employed for the parametrization of the HLC in G4(MP2) comprises 270 heats of formation, 105 ionization energies, 63 electron affinities, 10 proton affinities, and six hydrogen-bond energies. On the other hand, the performance of W4 was assessed with the W3 set^{4b} of 36 atomization energies.

While the above and many other training and test sets consist of data for general thermochemistry, there have also been a number of specialized sets developed for specific energy properties, such as barriers and weak interactions. Thus, the DBH24 set²⁰ consists of 24 barriers, obtained at the W1–W4 levels of theory, for a diverse range of reaction types. These include hydrogen abstractions, $\text{S}_{\text{N}}2$ reactions, hydrogen transfers, and unimolecular decompositions. Using W2 calculations, Boese et al. have compiled a hydrogen-bonding set of 16 complexes²¹ (referred to as the HB16 set hereafter), while the WI9/04 set²² comprises complexation energies for nine weakly bound complexes, calculated at the W1 level. Very recently, the GMTKN24 data set has been proposed,²³ which comprises subsets of a diverse range of 24 chemical test sets, encompassing thermochemistry, kinetics, and noncovalent interactions. The GMTKN24 set contains 731 energies.

In this study, we propose two new training and test sets that include a variety of thermochemical properties. They are termed the “energy sets” E0 and E2 and contain 148 and 526 energies, respectively. These sets contain experimental and theoretical energies of chemical accuracy. We use the E0 set for evaluating methods for geometry optimization and the E2 set for training the G4(MP2)-6X procedure. We also examine the use of the “traditional” and moderately sized G2/97 set (302 energies) for parametrization.²⁴ The details of the training and test sets are given below.

The E0 set is a selected subset of highly accurate theoretical data presented recently by Karton et al.²⁵ It contains atomization energies (AE, W4/08), barriers (ΔH^\ddagger , DBH24), hydrogen-bond energies (HB, HB16), and energies for weakly bound complexes (BE, WI9/04):

$$\begin{aligned} \text{E0(148)} = & \text{W4/08 (99, 0 K, W4+)} \\ & + \text{DBH24 (24, vibrationless, W3+)} \\ & + \text{HB16 (16, vibrationless, W2)} \\ & + \text{WI9/04 (9, vibrationless, W1)} \end{aligned}$$

Shown in parentheses are the number and the specification of the energies, and the methods by which they are obtained. W3+ and W4+ signify theoretical values obtained with procedures of at least W3 and W4 quality, respectively.

While the details of the G2/97 set have been given elsewhere,²⁴ we provide here a brief summary for comparison purposes. It consists of experimental data for heats of formation (ΔH_f), ionization energies (IEs), electron affinities (EAs), and proton affinities (PAs):

$$\begin{aligned} \text{G2/97(302)} = & \text{G2/97}\Delta H_f (148, 298 \text{ K, expt}) \\ & + \text{G2/97 IE (88, 0 K, expt)} \\ & + \text{G2/97 EA (58, 0 K, expt)} \\ & + \text{G2/97 PA (8, 0 K, expt)} \end{aligned}$$

For the IE of CN in this set, an alternative experimental value²⁶ (14.03 ± 0.02 eV, 1353 kJ mol^{-1}) has been recommended on the basis of W1 and W2 calculations.²⁷ We have adopted this value in the present study.

Our largest set is the E2 set, which is comprised of the E0 set, a modified G2/97 set termed G2/97', and the new E1 set:

$$\text{E2 (526)} = \text{E0 (148)} + \text{G2/97'} (248) + \text{E1 (130)}$$

The G2/97' set reduces the number of heats of formation from 148 (in G2/97) to 94. The new heats of formation set in G2/97' is simply termed G2/97' ΔH_f . The reduction in the number of data is done to remove redundancies in the final E2 set, as 54 of the 148 molecules in G2/97 are also present in the W4/08 set. Thus, the G2/97' set is defined as

$$\begin{aligned} \text{G2/97'}(248) = & \text{G2/97'}\Delta H_f (94, 298 \text{ K, expt}) \\ & + \text{G2/97 IE (88, 0 K, expt)} \\ & + \text{G2/97 EA (58, 0 K, expt)} \\ & + \text{G2/97 PA (8, 0 K, expt)} \end{aligned}$$

The E1 set consists of 73 heats of formation from the G3/99 set²⁸ that are not present in G2/97, calculated enthalpies for 49 radical reactions of Coote et al.²⁹ (referred to hereafter as RR49), and a subset (referred to as PR8) of the pericyclic reaction set of Houk et al.³⁰

G3/99 introduces 75 additional ΔH_f values to G2/97, of which two are already present in the W4/08 set. Removal of this duplication leads to the G3/99' ΔH_f set. The RR49 set comprises enthalpies calculated at the W1 level for 21 radical-addition (ADD) and 28 hydrogen-abstraction (ABS) reactions. Houk et al.'s pericyclic reaction set consists of experimental values for 11 barriers and 7 enthalpies for 11 pericyclic reactions. However, due to uncertainties in the experimental values, we choose to employ a subset of eight barriers, for which W1 vibrationless values are available.²⁵ The E1 set is thus defined as

$$\begin{aligned} \text{E1(130)} = & \text{G3/99'}\Delta H_f (73, 298 \text{ K, expt}) \\ & + \text{ADD (21, 0 K, W1, part of RR49)} \\ & + \text{ABS (28, 0 K, W1, part of RR49)} \\ & + \text{PR8 (8, vibrationless, W1)} \end{aligned}$$

4. Results and Discussion

We have optimized the parameters for G4(MP2)-6X using our largest E2 set (Table 1). We note that most of the HLC parameters for G4(MP2)-6X are smaller than their G4(MP2) counterparts. This is advantageous because it means that the new method is less reliant on the HLC than the original implementation. It has been pointed out that the form of HLC in G4(MP2), specifically the use of different values for *A* and *A'* for closed- and open-shell species, can lead to a diverging error for radical reactions.^{29,31} However, the difference between *A* and *A'* for G4(MP2)-6X (0.091) is substantially smaller than that for G4(MP2) (0.297). As a

Table 1. Optimized HLC (*A–E*, mHartree) and Scaling (*c1–c6*) Parameters for G4(MP2)-6X^a

	G4(MP2) ^b	G4(MP2)-6X	G4(MP2)-6X	G4(MP2)-6X
<i>A</i>	9.472	7.173	<i>c1</i>	1.327
<i>A'</i>	9.769	7.264	<i>c2</i>	0.403
<i>B</i>	3.102	3.677	<i>c3</i>	1.249
<i>C</i>	9.741	7.239	<i>c4</i>	0.486
<i>D</i>	2.115	2.404	<i>c5</i>	1.077
<i>E</i>	2.379	1.021	<i>c6</i>	0.824

^a Parameterized with the E2 set. See section 3.2 for the definition of the parameters. ^b Reference 6.

result, the problem of a diverging HLC contribution to radical reactions is significantly reduced for the new procedure.

In the following sections, we will first examine the performance of various methods for geometry optimization. Next, we will evaluate the new parameters in G4(MP2)-6X and assess their importance in improving the new procedure. We will then look at the performance of G4(MP2)-6X in comparison with G4(MP2), as well as discuss the portability of the popular G2/97 training set by testing the G2/97-trained procedure on the larger and more diverse E2 set. We will conclude our discussion by providing a more detailed analysis of the strengths and shortcomings of G4(MP2)-6X.

4.1. The Choice of Geometry. In the formulation of G4, Curtiss et al. have identified cases where closer agreement with benchmark energies can be achieved by using procedures for optimizing the geometry other than the default B3-LYP/6-31G(2df,p) method.^{2d} For instance, the use of MP2(Full)/6-31G(d) geometries improves the IEs for CH₄, BF₃, and BCl₃, while hydrogen-bond energies are improved by using B3-LYP/6-31+G(2df,p) geometries.

While the use of MP2 for geometry optimization may be desirable in some cases, it can become too costly to be employed for routine computations for larger systems, and it can also fail for open-shell systems for which there is significant spin contamination. We therefore focus on the use of alternative DFT procedures with various basis sets for obtaining geometries. In particular, we examine the performance of the BMK¹² and M06-2X³² procedures in combination with the 6-31+G(d,p), 6-31G(2df,p), 6-31+G(2df,p), and 6-311+G(2df,p) basis sets for geometry optimization. We then calculate single-point energies at the G4(MP2) level on these geometries. We use the E0 test set for our assessment, as it contains not only standard molecules, but also transition structures and hydrogen-bonded and weak complexes, which can be expected to be challenging in terms of obtaining reliable geometries. The MAD values for the E0 set and its subsets, for the various procedures used for geometry optimization, are shown in Table 2.

We find that, for the full E0 set and with the largest 6-311+G(2df,p) basis set, both BMK (4.00 kJ mol^{-1}) and M06-2X (4.24 kJ mol^{-1}) lead to lower MADs than B3-LYP (4.34 kJ mol^{-1}). We can see that this is also true for the subsets of E0, with the exception of the W4/08 set, for which the MADs for BMK and M06-2X are 0.29 and 0.63 kJ mol^{-1} , respectively, higher than that for B3-LYP. For BMK, we find that the MAD generally becomes smaller as the size of the basis set is increased. On the other hand, the basis set

Table 2. Mean Absolute Deviations (MADs, kJ mol⁻¹) from Benchmark Energies As a Function of Methods Employed for Geometry Optimization for G4(MP2)^a

		6-31+G (d,p)	6-31G (2df,p)	6-31+G (2df,p) ^b	6-311+G (2df,p)
W4/08	B3-LYP		4.36		4.47
	BMK	4.98	4.90	4.72	4.76
	M06-2X	4.95	4.99	5.09	5.10
DBH24	B3-LYP		5.71		5.75
	BMK	3.51	3.05	2.91	2.72
	M06-2X	2.65	2.93	2.69	2.60
HB16	B3-LYP		3.17		1.92
	BMK	1.91	2.85	1.76	1.74
	M06-2X	2.15	3.47	2.08	2.03
WI9/04	B3-LYP		4.09		3.47
	BMK	3.88	3.26	3.35	3.20
	M06-2X	2.89	4.32	2.90	3.08
E0	B3-LYP		4.43		4.34
	BMK	4.35	4.28	4.03	4.00
	M06-2X	4.15	4.45	4.24	4.24

^a Literature scale factors were employed for obtaining ZPVEs and ΔH . ^b The appropriate scaling factors for the 6-31+G(2df,p) basis set are taken as the average of the values for 6-31G(2df,p) and 6-311+G(2df,p); see ref 14 for more details.

dependence for M06-2X is more erratic, with the lowest MAD for the E0 set being achieved with the smallest 6-31+G(d,p) basis set, and the largest MAD for the intermediate-sized 6-31G(2df,p) basis set. For the subsets of E0, we can also see a more systematic basis set effect for BMK than for M06-2X.

While both BMK and M06-2X appear to provide superior geometries compared with those of B3-LYP, the somewhat lower MAD for BMK, and the seemingly more predictable basis set effect, have led us to favor BMK as our method of choice. In addition, it has previously been found that the use of BMK geometries improves the agreement with experimental thermochemical data for G3.³³ Turning our attention to the choice of basis set, the use of 6-31+G(2df,p) gives an overall MAD (4.03 kJ mol⁻¹) that is only slightly greater than that for the larger 6-311+G(2df,p) basis set (4.00 kJ mol⁻¹). However, further reduction in the basis set size to 6-31G(2df,p) or 6-31+G(d,p) leads to larger increases in the MAD (to 4.28 and 4.35 kJ mol⁻¹, respectively). Thus, we have chosen BMK/6-31+G(2df,p) as the method for

geometry optimization in the G4(MP2)-6X procedure. We note that the use of 6-31+G(2df,p), compared with the smaller 6-31G(2df,p) basis set employed in G4(MP2), will lead to a slightly more expensive procedure. However, the geometry optimization generally contributes only a small part to the total computational cost in a G4(MP2)-type calculation, and this is therefore a cost-effective improvement.

4.2. Components of G4(MP2)-6X and Their Contributions. Having determined the appropriate procedure for geometry optimization, we now proceed to examine the various approaches for scaling the correlation energies used in G4(MP2), and how they affect the performance of the modified procedures. The results are summarized in Table 3.

We can see that, when compared with G4(MP2) with the default B3-LYP geometries (column I), the use of BMK for geometry optimization (column II) gives a comparable MAD, a reduced SD, and three less outliers.³⁴ Reoptimization of the HLC parameters leads to a lower MAD and SD (column III). When one adds a single set of scaling parameters (c1 and c2) for the OS and SS components for the MP2 correlation energy ($E_{\text{MP2}}^{\text{corr}}$), the MAD and SD are further lowered (column IV). However, the biggest improvement accompanying this change can be seen for NO, where the number of outliers is substantially reduced from 68 to 59. A slight improvement in performance can be achieved when one scales c5 and c6 (instead of c1 and c2) for the CCSD and (T) correlation energies, respectively (column VI). Introducing different scaling parameters for the two sets of $E_{\text{MP2}}^{\text{corr}}$, i.e., c1 and c2 for $E_{\text{MP2}}^{\text{corr}}/6\text{-}31\text{G(d)}$ and c3 and c4 for $E_{\text{MP2}}^{\text{corr}}/\text{G3MP2LargeXP}$, leads to a larger improvement (column V). Finally, when one applies scaling to all six components, one arrives at the full G4(MP2)-6X procedure (column VII), with the lowest values for the MAD (3.64 kJ mol⁻¹) and SD (5.14 kJ mol⁻¹), and with a significant reduction also in the number of outliers to 44.

4.3. Performance of G4(MP2)-6X. Table 4 summarizes the statistics for the E2 test set for G4(MP2), and for G4(MP2)-6X when parametrized with the G2/97 and E2 training sets. We can see that G4(MP2) gives a respectable MAD of 4.42 kJ mol⁻¹, and an SD of 6.38 kJ mol⁻¹. The

Table 3. Performance of Various Modifications for G4(MP2) on the E2 Set^a

	I	II	III	IV	V	VI	VII
Geometry ^b	B3-LYP ^c	BMK	BMK	BMK	BMK	BMK	BMK ^d
HLC	default	default	fitted	fitted	fitted	fitted	fitted
c1	1	1	1	fitted	fitted	1	fitted
c2	1	1	1	fitted	fitted	1	fitted
c3	1	1	1	= c1	fitted	1	fitted
c4	1	1	1	= c2	fitted	1	fitted
c5	1	1	1	1	1	fitted	fitted
c6	1	1	1	1	1	fitted	fitted
MD ^e	-0.49	-0.39	-0.37	0.08	-0.15	-0.40	-0.26
MAD ^f	4.42	4.39	4.18	3.99	3.77	3.82	3.64
LD ^g	-41.61	-42.17	-38.14	-37.79	-39.41	-37.45	-36.66
SD ^h	6.38	6.05	5.88	5.61	5.41	5.47	5.14
NO ⁱ	70	67	68	59	55	56	44

^a Parameterized with the E2 set. See section 3.2 for the definition of the parameters. All energies in kJ mol⁻¹. ^b The 6-31G(2df,p) basis set was used for B3-LYP, while BMK employs 6-31+G(2df,p). ^c This set of parameters represents the original G4(MP2). ^d G4(MP2)-6X. ^e Mean deviation. ^f Mean absolute deviation. ^g Largest deviation. ^h Standard deviation. ⁱ Number of outliers (absolute deviation > 8.37 kJ mol⁻¹).

Table 4. Comparison of Deviations (kJ mol^{-1}) for the E2 Test Set for G4(MP2), and for G4(MP2)-6X When Parameterized with the G2/97 and E2 Training Sets^a

G4(MP2)					
test set	MD	MAD	LD	SD	NO
E0	-0.57	4.58	-29.64	6.78	20
G2/97'	-0.78	4.42	-41.61	6.53	37
E1	0.16	4.22	-20.15	5.55	13
E2	-0.49	4.42	-41.61	6.38	70

G4(MP2)-6X (parametrized with G2/97)					
test set	MD	MAD	LD	SD	NO
E0	-1.34	3.81	-19.40	5.05	15
G2/97'	-0.28	4.09	-38.53	5.85	31
E1	0.76	3.44	15.28	4.38	9
E2	-0.32	3.85	-38.53	5.34	55

G4(MP2)-6X (parametrized with E2)					
test set	MD	MAD	LD	SD	NO
E0	-0.42	3.43	-19.47	4.67	8
G2/97'	-0.49	4.14	-36.66	5.87	29
E1	0.37	2.92	15.17	4.04	7
E2	-0.26	3.64	-36.66	5.14	44

^a See section 3.3 for the definitions of the training and test sets.

largest deviation is $-41.61 \text{ kJ mol}^{-1}$, and there are 70 outliers (i.e., where the error is larger than 8.37 kJ mol^{-1} (2 kcal mol^{-1})) among the 526 entries. In addition, we find that G4(MP2) performs well for all three subsets of E2.

When G4(MP2) is compared with G4(MP2)-6X (parametrized with E2), we can see that G4(MP2)-6X gives lower MAD and SD values overall, as well as for the individual subsets. The new procedure also produces fewer outliers for E2 and for the subsets. We find that G4(MP2)-6X performs only slightly better for the G2/97' set, but there are major improvements for the E0 and E1 sets. For example, the MAD for the E0 set is improved by 1.15 kJ mol^{-1} , the SD is lowered by 2.11 kJ mol^{-1} , and the NO is reduced by 12. Thus, the inclusion of the six additional parameters improves the accuracy and precision for a wide range of systems, and these improvements come at effectively no additional computational cost. Notably, when trained by either the G2/97 or the E2 sets, G4(MP2)-6X achieves chemical accuracy with MADs of 3.85 kJ mol^{-1} ($0.92 \text{ kcal mol}^{-1}$) and 3.64 kJ mol^{-1} ($0.87 \text{ kcal mol}^{-1}$), respectively.

A comparison of the performance of the G2/97-parametrized G4(MP2)-6X with that parametrized using E2 reveals that the E2-parametrized method gives lower overall MAD, LD, SD, and NO values. When the statistics for the three component sets of E2 are examined, namely, E0, G2/97', and E1, we find that the E1 set is the prime beneficiary of the parametrization with the E2 set, with notably reduced MAD and SD values. The E2-parametrization also gives better statistics for the E0 set, while the performance for the G2/97' set is slightly worse, presumably due to the reduced weight of G2/97 properties in the E2 set.

While parametrization of G4(MP2)-6X with the E2 set leads to superior performance compared with fitting to G2/97, parametrization with the latter is not far behind, with an MAD and SD that are less than 0.5 kJ mol^{-1} higher than

Table 5. Deviations (kJ mol^{-1}) for the E2-Parameterized G4(MP2)-6X Procedure for the Various Test Sets^a

	MD	MAD	LD	SD	NO
E0 (148)	-0.42	3.43	-19.47	4.67	8
W4/08 (99)	-0.59	4.00	-19.47	5.32	8
DBH24 (24)	1.39	2.97	7.30	3.47	0
HB16 (16)	-1.69	1.78	-5.17	1.59	0
WI9/04 (9)	-1.23	1.34	-2.31	0.88	0
G2/97' (248)	-0.49	4.14	-36.66	5.87	29
ΔH_f (94)	-0.65	3.04	-16.62	4.61	6
IE (88)	-0.39	4.68	-36.66	6.75	12
EA (58)	-0.30	5.29	-15.53	6.56	11
PA (8)	-1.12	2.79	-7.99	3.74	0
E1 (130)	0.37	2.92	15.17	4.04	7
G3/99' (73)	0.11	3.36	15.17	4.65	6
ADD (21)	1.80	2.47	7.21	2.95	0
ABS (28)	1.31	1.68	3.72	1.53	0
PR8 (8)	-4.29	4.35	-8.44	3.02	1
E2 (526)	-0.26	3.64	-36.66	5.14	44

^a See section 3.3 for the definitions for the training and test sets.

the E2-parametrized values. Importantly, the G2/97 set appears to be very portable: using it for parametrization gives uniformly reasonable results for E0 (MAD = 3.81 kJ mol^{-1}) and E1 (3.44 kJ mol^{-1}) as well as for the G2/97' set (4.09 kJ mol^{-1}). This demonstrates that G2/97 can be used as a cost-effective training set for preliminary screening in the formulation of composite procedures.

When a procedure is developed by fitting a sizable number of parameters over a particular training set, one potential pitfall is that the resulting method may not be applicable to properties and systems that are not included in the training set. In this connection, our finding of comparable performance for the G4(MP2)-6X procedure when parametrized with the G2/97 and E2 training sets demonstrates to some extent the robustness of the form of parametrization in this method.

How does G4(MP2)-6X compare with G4? We have carried out such a comparison using the smaller E0 test set of 148 energies. The same BMK/6-31+G(2df,p) geometries were used for both G4(MP2)-6X and G4 for the sake of consistency.³⁵ We find that the G4(MP2)-6X procedure gives an MAD (3.43 kJ mol^{-1}) that is only slightly higher than that for G4 (3.22 kJ mol^{-1}) but considerably lower than that for G4(MP2) with BMK geometries (4.03 kJ mol^{-1}).

We now look at the performance of the E2-parametrized G4(MP2)-6X procedure in more detail (Table 5). We find that G4(MP2)-6X performs well for AEs and ΔH_f s (W4/08, G2/97' ΔH_f , and G3/99'), PAs, radical reaction enthalpies (ADD and ABS), intermolecular interactions (HB16 and WI9/04), and barriers in the DBH24 set, with MADs that are in all cases smaller than 4.18 kJ mol^{-1} (1 kcal mol^{-1}). On the other hand, pericyclic reaction barriers (PR8), IEs, and EAs show larger MAD values.

4.4. Outliers in the E2 Set. Table 6 lists the outliers in the E2 set, i.e., situations for which the G4(MP2)-6X energy deviates from the corresponding benchmark value by more than 8.37 kJ mol^{-1} (2 kcal mol^{-1}). Out of the 44 outliers, there are four cases where the deviation is larger than $16.74 \text{ kJ mol}^{-1}$ (4 kcal mol^{-1}). These are the AEs for ClOO• and P₄, and the IEs for B₂F₄ and B₂H₄. In addition, there are 10

Table 6. Outliers for the E2 Set for the G4(MP2)-6X Procedure Where Deviations (D) Are Larger than 8.37 kJ mol^{-1} (2 kcal mol^{-1})^a

		Range of absolute deviations (D , kcal mol^{-1})		
		$2 < D < 3$	$3 < D < 4$	$4 < D$
W4/08	AE	BHF ₂ , F ₂ O ₂ , CS ₂ , S ₄ , CS	B ₂ H ₆	CIOO, P ₄
DBH24	ΔH^\ddagger		nil	
HB16	HB		nil	
W19/04	BE		nil	
G2/97'	ΔH_f	SiH ₂ singlet, thiophene	CF ₂ O, C ₂ F ₄ , C ₂ Cl ₄ , CH ₂ CHCl	
G2/97	IE	S, CH ₄ , BF ₃ , BCl ₃ , C ₆ H ₆ , C ₆ H ₅ CH ₃ , CH ₃ F, C ₆ H ₅ NH ₂ , Si ₂ H ₆	C ₃ H ₇	B ₂ F ₄ , B ₂ H ₄
G2/97	EA	Li, F, Si, P, CH, SiH, PH, CCH, CH ₂ NC	B, C	
G2/97	PA		nil	
G3/99'	ΔH_f	benzoquinone, pyrimidine, 2-methylthiophene, SF ₆	acetylacetylene, tetramethylsilane	
RR49	ΔH_r		nil	
PR8	ΔH^\ddagger	1,3-cyclopentadiene + C ₂ H ₄ → norbornene		

^a See section 3.3 for the definitions for the training and test sets; AE = atomization energy, ΔH^\ddagger = barrier, HB = hydrogen-bond energy, BE = binding energy, ΔH_f = heat of formation, IE = ionization energy, EA = electron affinity, PA = proton affinity, ΔH_r = reaction energy.

Table 7. Comparison of Deviations (kJ mol^{-1}) for Various Alternative Procedures for the Outliers Listed in Table 6 Where the Deviations Are Larger than $12.55 \text{ kJ mol}^{-1}$ (3 kcal mol^{-1})^a

	G4(MP2)-6X	G4(MP2)	G4	W1U	W1Usc
AE CIOO	-19.47	-18.51	-16.86	-18.53	-2.99
AE P ₄	-16.83	9.84	-2.66	-4.30	-4.34
IE B ₂ F ₄	-36.66	-41.61	-37.27	-32.52	-32.59
IE B ₂ H ₄	-16.81	-9.29	-9.50	-12.28	-12.37
AE B ₂ H ₆	-15.98	-17.51	-3.62	3.42	3.42
ΔH_f CF ₂ O	16.53	17.67	16.39	8.76	9.22
ΔH_f C ₂ F ₄	-12.63	-12.51	-13.50	-28.49	-27.88
ΔH_f C ₂ Cl ₄	-13.89	-23.39	-13.12	-22.94	-21.88
ΔH_f CH ₂ CHCl	-16.62	-19.11	-15.17	-22.08	-21.56
IE C ₃ H ₇	13.95	14.02	12.17	7.48	7.72
EA B	-15.53	-21.44	-9.15	-7.96	-7.80
EA C	-14.78	-18.85	-9.86	-6.56	-6.62
ΔH_f acetylacetylene	13.45	11.04	10.68	6.72	7.56
ΔH_f tetramethylsilane	15.17	15.13	14.36	0.72	1.66

^a AE = atomization energy, ΔH_f = heat of formation, IE = ionization energy, EA = electron affinity.

cases with a deviation that falls between $12.55 \text{ kJ mol}^{-1}$ (3 kcal mol^{-1}) and $16.74 \text{ kJ mol}^{-1}$. We have briefly examined these cases using higher-level procedures. The results are shown in Table 7.

When comparing G4(MP2) and G4(MP2)-6X with G4, we generally find better agreement between G4(MP2)-6X and G4 than between G4(MP2) and G4. However, there are a number of cases where G4(MP2) agrees better with G4, notably for the IE of B₂H₄.

We find that G4 significantly reduces the deviations from benchmark values for P₄ (AE), B₂H₄ (IE), B₂H₆ (AE), B (EA), and C (EA). Presumably, the use of MP4 in G4 (instead of MP2 in G4(MP2) and G4(MP2)-6X) to account for the effect of adding diffuse and polarization functions to the 6-31G(d) basis set is beneficial in these cases. The use of the W1U procedure³⁶ further reduces the number of outliers with deviations that are larger than $12.55 \text{ kJ mol}^{-1}$ from seven to five. However, we notice that the use of W1U leads to notably larger deviations than with G4 in a number of cases, particularly for the ΔH_f of C₂F₄, C₂Cl₄, and CH₂CHCl. The use of the spin correction (sc) term in the W1Usc procedure reduces the deviation for the AE of CIOO substantially compared with W1U (from $-18.53 \text{ kJ mol}^{-1}$ to $-2.99 \text{ kJ mol}^{-1}$), but in general the two procedures perform comparably.

For the four cases where the deviations at the W1Usc level remain larger than $12.55 \text{ kJ mol}^{-1}$, namely, the IE of B₂F₄ and the ΔH_f of C₂F₄, C₂Cl₄, and CH₂CHCl, it has previously been suggested, on the basis of calculations of isodesmic reactions and G3 and G4 calculations, that the accuracy of the experimental values is questionable.²⁸ The large deviations for W1U and W1Usc in these cases support the desirability of experimental re-examination.

5. Concluding Remarks

We have developed the G4(MP2)-6X procedure as a variant of G4(MP2). It employs BMK/6-31+G(2df,p) for geometry optimization and has six additional scaling factors for the correlation energy components compared with G4(MP2), as well as reoptimized HLC parameters.

We find it to be an improvement ($\text{MAD} = 3.64 \text{ kJ mol}^{-1}$) over G4(MP2) ($\text{MAD} = 4.42 \text{ kJ mol}^{-1}$) for a wide range of properties for the E2 set of 526 energies. A comparison of G4(MP2)-6X with G4 on the E0 set shows that the performance of the new procedure (3.43 kJ mol^{-1}) approaches that for G4 (3.22 kJ mol^{-1}) for this series of energies, for which G4(MP2) has an MAD of 4.03 kJ mol^{-1} . Thus, the G4(MP2)-6X procedure bridges the gap between G4(MP2) and G4, with an accuracy that is closer to G4 but a computational cost that is comparable to that for the considerably more economical G4(MP2). While G4(MP2)-6X represents a noticeable improvement over G4(MP2), it also has cases of dramatic failures. These can generally be attributed to inadequate treatment of electron correlation.

We find that, when the G2/97 training set is used (in place of E2) for parametrization, G4(MP2)-6X still gives uniformly good results, not only for the types of energies represented in G2/97 but also for the wider range of energies represented in the larger E2 set. This reflects to some extent the robustness of G4(MP2)-6X, despite the additional parameters. In addition, it suggests the more general use of G2/97 as a cost-effective training set for the screening of theoretical procedures.

Acknowledgment. We gratefully acknowledge the award of an Australian Professorial Fellowship and funding from the ARC Centre of Excellence for Free Radical Chemistry and Biotechnology (to L.R.) and generous alloca-

tions of computer time from the National Computational Infrastructure (NCI) National Facility and Intersect.

Supporting Information Available: Script for running G4(MP2)-6X calculations with Gaussian 09; reactions of the DBH24, RR49, and PR8 sets (Table S1); zero-point vibrational energies and thermal corrections from scaled BMK/6-31+G(2df,p) frequencies and G4(MP2)-6X total electronic energies (Table S2); and deviations from experimental and benchmark values (Table S3). This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) For general overviews, see: (a) Helgaker, T.; Klopper, W.; Tew, D. P. *Mol. Phys.* **2008**, *106*, 2107. (b) Martin, J. M. L. *Annu. Rep. Comput. Chem.* **2005**, *1*, 31. (c) Raghavachari, K.; Curtiss, L. A. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005.
- (2) (a) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622. (b) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221. (c) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764. (d) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *126*, 084108.
- (3) (a) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A., Jr. *J. Chem. Phys.* **1996**, *104*, 2598. (b) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822. (c) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532. (d) Wood, G. P. F.; Radom, L.; Petersson, G. A.; Barnes, E. C.; Frisch, M. J.; Montgomery, J. A., Jr. *J. Chem. Phys.* **2006**, *125*, 094106.
- (4) (a) Martin, J. M. L.; de Oliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843. (b) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 4129. (c) Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. *J. Chem. Phys.* **2006**, *125*, 144108.
- (5) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123*, 124107.
- (6) (a) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 1293. (b) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703. (c) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127*, 124105.
- (7) See, for example: (a) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986. (b) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley: New York, 2001. (c) Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; Wiley: Chichester, U. K., 2007.
- (8) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
- (9) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
- (10) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Mitrushenkov, A.; Rauhut, G.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hetzer, G.; Hrenar, T.; Knizia, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklaß, A.; Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. *MOLPRO 2006.1*; University of Birmingham: Birmingham, U.K., 2006.
- (11) Crawford, T. D.; Sherrill, C. D.; Valeev, E. F.; Fermann, J. T.; King, R. A.; Leininger, M. L.; Brown, S. T.; Janssen, C. L.; Seidl, E. T.; Kenny, J. P.; Allen, W. D. *J. Comput. Chem.* **2007**, *28*, 1610.
- (12) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.
- (13) Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683.
- (14) The scale factors for the 6-31+G(2df,p) basis set are taken as the average of the 6-31G(2df,p) and 6-311+G(2df,p) values obtained from ref 13, namely, 0.9770 for the ZPVE and 0.9627 for ΔH . For example, the BMK/6-31+G(2df,p) scale factor for the ZPVE at 298 K is calculated as $1/2 \times [0.9752 \text{ (BMK/6-31G(2df,p))} + 0.9787 \text{ (BMK/6-311+G(2df,p))}] = 0.9770$.
- (15) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 1125.
- (16) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 3898.
- (17) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
- (18) Takatani, T.; Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2008**, *128*, 124111.

- (19) We have explored the effect of re-optimizing the exponent for the extrapolation procedure and found that this does not lead to notable improvements.
- (20) (a) Zheng, J.; Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 569. (b) Zheng, J.; Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 808.
- (21) Boese, A. D.; Martin, J. M. L.; Klopper, W. *J. Phys. Chem. A* **2007**, *111*, 11122.
- (22) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.
- (23) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2010**, *6*, 107.
- (24) (a) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063. (b) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42.
- (25) Karton, A.; Tarnopolsky, A.; Lamère, J.-F.; Schatz, G. C.; Martin, J. M. L. *J. Phys. Chem. A* **2008**, *112*, 12868.
- (26) Berkowitz, J.; Chupka, W. A.; Walter, T. A. *J. Chem. Phys.* **1969**, *50*, 1497.
- (27) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014.
- (28) Curtiss, L.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374.
- (29) Lin, C. Y.; Hodgson, J. L.; Namazian, M.; Coote, M. L. *J. Phys. Chem. A* **2009**, *113*, 3690.
- (30) (a) Guner, V.; Khuong, K. S.; Leach, A. G.; Lee, P. S.; Bartberger, M. D.; Houk, K. N. *J. Phys. Chem. A* **2003**, *107*, 11445. (b) Ess, D. H.; Houk, K. N. *J. Phys. Chem. A* **2005**, *109*, 9542.
- (31) Chan, B.; Coote, M. L.; Radom, L. *J. Chem. Theory Comput.* **2010**, *6*, 2647.
- (32) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
- (33) Zheng, W.-R.; Fu, Y.; Guo, Q.-X. *J. Chem. Theory Comput.* **2008**, *4*, 1324.
- (34) We also find that, by optimizing both G4(MP2)//B3-LYP/6-31G(2df,p) and G4(MP2)//BMK/6-31+G(2df,p) with the E2 set, the BMK-optimized procedure leads to a comparable MAD (4.18 kJ mol⁻¹) to that for the B3-LYP-optimized procedure (4.25 kJ mol⁻¹). The BMK-optimized procedure has a lower SD (5.88 kJ mol⁻¹) than that for the B3-LYP-optimized procedure (6.20 kJ mol⁻¹), as well as a lower NO (68, compared with 71 for G4(MP2)//B3-LYP/6-31G(2df,p)).
- (35) When the default B3-LYP geometries in G4 and G4(MP2) were used, the MADs for the E0 set were 3.36 and 4.58 kJ mol⁻¹, respectively, for G4 and G4(MP2).
- (36) Barnes, E. C.; Petersson, G. A.; Montgomery, J. A.; Frisch, M. J.; Martin, J. M. L. *J. Chem. Theory Comput.* **2009**, *5*, 2687.

CT100542X