Multi-coefficient Correlation Method: Comparison of Specific-Range Reaction Parameters to General Parameters for $C_nH_xO_y$ Compounds

Patton L. Fast, Nathan E. Schultz, and Donald G. Truhlar*

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431

We optimized the coefficients for 11 multi-coefficient correlation methods (MCCMs) against 54 atomization energies for molecules composed of C, H, and O atoms and containing 343 bonds. The methods included two scaling-all-correlation (SAC) methods, a multi-coefficients SAC (MCSAC) method, three Utah methods, two Colorado methods, MC-QCISD, MCG3, and G3S. The mean unsigned errors are (on average) 36% lower than for calculations with general parameters. The mean unsigned error per bond for the most highly recommended methods is only 0.1−0.2 kcal/mol.

1. Introduction

Scaling of the correlation energy and the use of multi-coefficient correlation methods (MCCMs) can provide high-accuracy thermochemistry and molecular geometries at less cost than methods using brute force.[1−11] These methods correspond to extrapolating electronic structure calculations toward the limit of complete configuration interaction. So far, all MCCMs have been general parametrizations based on compounds composed of first- and second-row elements (B, C, N, O, F, Al, Si, P, S, and Cl) plus H[6,8,9,11] or, in one case,[7] on first-row elements (Li, Be, C, N, O, and F) plus H and, in another,[10] on compounds containing all atoms from H to Ar. For simpler semiempirical methods, one can often obtain more accurate results by parametrizing to a specific reaction or a delimited range of molecules.[12−18] [The resulting parameters are called specific reaction parameters or specific-range reaction parameters (SRP) to distinguish them from the general parameters.] In the present work we consider applying this SRP approach with the scaling-all-correlation (SAC) method and MCCM; however, because we are referring to a training set of $C_nH_xO_y$ atomization reactions instead of a particular reaction, we refer to this type of parametrization as specific-range reaction parameters (SRP). The range of molecules we focus attention on is actually reasonably broad, namely all 27 molecules composed entirely of C, H, and O atoms from a previously reported[9] data set of 82 zero-point-exclusive atomization energies as well as 27 additional $C_nH_xO_y$ compounds. This class of compounds includes a large subset of molecules important for applications in combustion chemistry, atmospheric chemistry, biochemistry (e.g., sugars), and environmental chemistry. If we can obtain more accurate results in such critical application areas by using SRPs instead of general parameters, the extra effort to obtain such parameters is worthwhile, and the lack of theoretical purity in adding a new set of semiempirical parameters will be tolerable.

Two of the questions we address are: (i) Do we still obtain physical parameters if we limit the training set to the subset of original training molecules containing only C, H, and O? (ii) How much do the mean errors decrease if we use new parameters adjusted to $C_nH_xO_y$ atomization energies instead of the broader atomization energy set?

2. Theory

We will obtain SRPs for 11 MCCMs, including two proposed originally by us[5,6,8,9,11] and the G3S scheme of Curtiss et al.[10] These are the methods recommended most highly[9,11] as a result of our previous investigations[6−9,11] of more than 40 SAC methods and MCCMs. SAC methods[1−5] are actually special cases of MCCM with only one semiempirical coefficient, or, especially considering that SAC came first, it is probably more appropriate to say that some MCCMs (MC−SAC methods) are multi-coefficient generalizations of the SAC approach, whereas other MCCMs are combinations of the generalized SAC approach with infinite-basis[19,20] (IB) extrapolation methods. A third class of MCCM, namely the empirical-infinite-basis (EIB) methods, are empirical versions of the IB approach without any SAC element, but EIB methods seem to be less efficient than other MCCMs, and they will not be considered further in this article.

We have developed a useful shorthand notation for writing MCCM energy expressions. In particular, if E denotes energy, and M/B denotes a calculation at many-electron level M with one-electron basis set B, we define

$$\Delta E(M2|M1/B) \equiv E(M2/B) - E(M1/B) \quad (1)$$

$$\Delta E(M/B2|B1) \equiv E(M/B2) - E(M/B1) \quad (2)$$

and

$$\Delta E(M2|M1/B2|B1) \equiv \Delta E(M2|M1/B2) - \Delta E(M2|M1/B1) \quad (3)$$

With this notation the 11 SAC and MCCM energy expressions considered in this article are

$$E(\text{SAC-MP2/pDZ}) = c_0 E(\text{HF/pDZ}) + c_1 \Delta E(\text{MP2}|\text{HF/pDZ}) + E(\text{SO}) \quad (4)$$

$$E(\text{SAC-MP4SDQ/pDZ}) = c_0 E(\text{HF/pDZ}) + c_1 \Delta E(\text{MP4SDQ}|\text{HF/pDZ}) + E(\text{SO}) \quad (5)$$

$$E(\text{MCSAC-QCISD}/6\text{-}31\text{G(d)}) =$$
$$c_0 E(\text{HF}/6\text{-}31\text{G(d)}) + c_1 \Delta E(\text{MP2}|\text{HF}/6\text{-}31\text{G(d)} +$$
$$c_2 \Delta E(\text{QCISD}|\text{MP2}/6\text{-}31\text{G(d)}) + E(\text{SO}) \quad (6)$$

$$E(\text{MCCM-UT-QCISD;}6\text{-}31\text{G(2df,p);}\ 6\text{-}31\text{G(d)}) =$$
$$c_0 E(\text{HF}/6\text{-}31\text{G(d)}) + c_1 \Delta E(\text{HF}/6\text{-}31\text{G(2df,p)}|6\text{-}31\text{G(d)} +$$
$$c_2 \Delta E(\text{MP2}|\text{HF}/6\text{-}31\text{G(d)}) +$$
$$c_3 \Delta E(\text{MP2}/6\text{-}31\text{G(2df,p)}|6\text{-}31\text{G(d)}) +$$
$$c_4 \Delta E(\text{QCISD}|\text{MP2}/6\text{-}31\text{G(d)}) + E(\text{SO}) \quad (7)$$

$$E(\text{MCCM-UT-MP4SDQ}) = c_0 E(\text{HF}/\text{pDZ}) +$$
$$c_1 \Delta E(\text{HF}/\text{pTZ}|\text{pDZ}) + c_2 \Delta E(\text{MP2}|\text{HF}/\text{pDZ}) +$$
$$c_3 \Delta E(\text{MP2}|\text{HF}/\text{pTZ}|\text{pDZ}) + c_4 \Delta E(\text{MP4SDQ}|\text{MP2}/\text{pDZ}) +$$
$$E(\text{SO}) \quad (8)$$

$$E(\text{MCCM-CO-MP2; MG3;}\ 6\text{-}31{+}\text{G(d)}) =$$
$$c_0 E(\text{HF}/6\text{-}31{+}\text{G(d)}) + c_1 \Delta E(\text{HF}/\text{MG3}|6\text{-}31\text{G(d)}) +$$
$$c_2 \Delta E(\text{MP2}|\text{HF}/6\text{-}31\text{G(d)}) + c_3 \Delta E(\text{MP2}/\text{MG3}|6\text{-}31\text{G(d)}) +$$
$$E(\text{SO}) \quad (9)$$

$$E(\text{MC-QCISD}) = c_0 E[\text{HF}/6\text{-}31\text{G(d)}] +$$
$$c_1 \Delta E[\text{MP2}|\text{HF}/6\text{-}31\text{G(d)}] + c_2 \Delta E[\text{MP2}/\text{MG3}|6\text{-}31\text{G(d)}] +$$
$$c_3 \Delta E[\text{QCISD}|\text{MP2}/6\text{-}31\text{G(d)}] + E(\text{SO}) \quad (10)$$

$$E(\text{MCCM-UT-CCSD}) = c_0 E(\text{HF}/\text{pDZ}) +$$
$$c_1 \Delta E(\text{HF}/\text{pTZ}|\text{pDZ}) + c_2 \Delta E(\text{MP2}|\text{HF}/\text{pDZ}) +$$
$$c_3 \Delta E(\text{MP2}|\text{HF}/\text{pTZ}|\text{pDZ}) + c_4 \Delta E(\text{CCSD}|\text{MP2}/\text{pDZ}) +$$
$$E(\text{SO}) \quad (11)$$

$$E(\text{MCCM-CO-MP2}) = c_0 E(\text{HF}/\text{pDZ}) +$$
$$c_1 \Delta E(\text{HF}/\text{pTZ}|\text{pDZ}) + c_2 \Delta E(\text{MP2}|\text{HF}/\text{pDZ}) +$$
$$c_3 \Delta E(\text{MP2}|\text{HF}/\text{pTZ}|\text{pDZ}) + E(\text{SO}) \quad (12)$$

$$E(\text{MCG3}) = c_0 E[\text{HF}/6\text{-}31\text{G(d)}] +$$
$$c_1 \Delta E[\text{HF}/\text{MG3}|6\text{-}31\text{G(d)}] + c_2 \Delta E[\text{MP2}|\text{HF}/6\text{-}31\text{G(d)}] +$$
$$c_3 \Delta E[\text{MP2}|\text{HF}/\text{MG3}|6\text{-}31\text{G(d)}] +$$
$$c_4 \Delta E[\text{MP4SDQ}|\text{MP2}/6\text{-}31\text{G(d)}] +$$
$$c_5 \Delta E[\text{MP4SDQ}|\text{MP2}/6\text{-}31\text{G(2df,p)}|6\text{-}31\text{G(d)}] +$$
$$c_6 \Delta E[\text{MP4}|\text{MP4SDQ}/6\text{-}31\text{G(d)}] +$$
$$c_7 \Delta E[\text{QCISD(T)}|\text{MP4}/6\text{-}31\text{G(d)}] + E(\text{SO}) \quad (13)$$

$$E(\text{G3S}) = E[\text{HF}/6\text{-}31\text{G(d)}] + c_0\{\Delta E[\text{MP2}|\text{HF}/6\text{-}31\text{G(d)}] +$$
$$\Delta E[\text{MP3}|\text{MP2}/6\text{-}31\text{G(d)}] + \Delta E[\text{MP4}|\text{MP3}/6\text{-}31\text{G(d)}]\} +$$
$$c_1 \Delta E[\text{QCISD(T)}|\text{MP4}/6\text{-}31\text{G(d)}] +$$
$$c_2 \Delta E[\text{HF}/\text{G3large}|6\text{-}31\text{G(d)}] +$$
$$c_3 \Delta E[\text{MP2(full)}|\text{HF}/\text{G3large}|6\text{-}31\text{G(d)}] +$$
$$c_4\{\Delta E[\text{MP3}|\text{MP2}/6\text{-}31{+}\text{G(d)}|6\text{-}31\text{G(d)}] +$$
$$\Delta E[\text{MP3}|\text{MP2}/6\text{-}31\text{G(2df,p)}|6\text{-}31\text{G(d)}]\} +$$
$$c_5\{\Delta E[\text{MP4}|\text{MP3}/6\text{-}31{+}\text{G(d)}|6\text{-}31\text{G(d)}] +$$
$$\Delta E[\text{MP4}|\text{MP3}/6\text{-}31\text{G(2df,p)}|6\text{-}31\text{G(d)}]\} + E(\text{SO}) \quad (14)$$
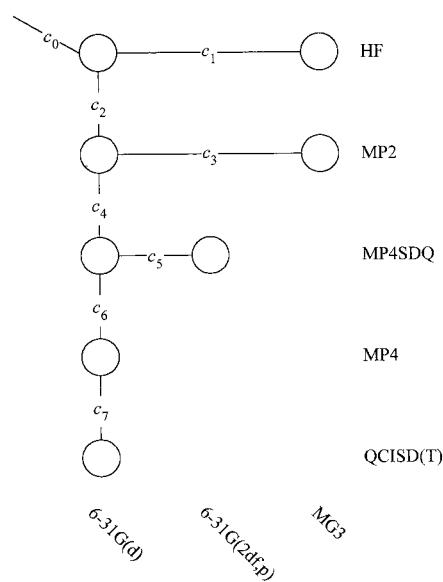
where the many-electron levels [HF,[21] MP2,[21] MP3,[21] MP4SDQ,[21] MP4,[21] CCSD,[22] QCISD,[23] and QCISD(T)[23]] and one-electron basis sets [pDZ ≡ cc-pVDZ,[24,25] pTZ ≡ cc-pVTZ,[24,25] 6-31G-(d),[21] 6-31G(2df,p),[21] 6-311G(d,p),[21] 6-311+G(d,p),[21] 6-311G-
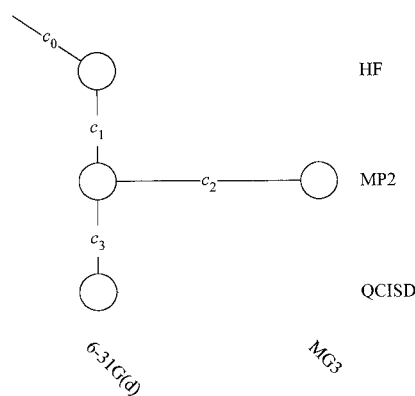
Figure 1. Coefficient tree for MC-QCISD.
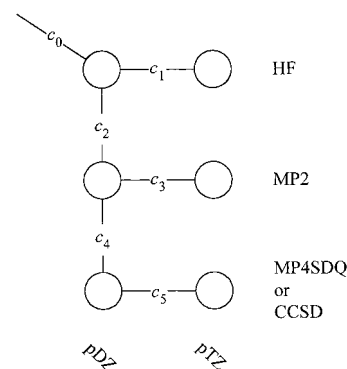
Figure 2. Coefficient tree for SAC-MP2/pDZ, SAC-MP4SDQ/pDZ, MCCM-CO-MP2, MCCM-UT-MP4SDQ, and MCCM-UT-CCSD.

Figure 3. Coefficient tree for MCG3.

(2df,p),[21] 6-311+G(3df,2p),[21] MG3,[8] and G3large[26]] are standard and explained elsewhere. The quantity $E(\text{SO})$ is explained below. Except when "full" is indicated, only valence electrons are correlated. These energy expressions are easier to understand when represented visually as in Figures 1−3. In these figures, a diagonal line is the coefficient of the smallest-basis-set uncorrelated energy, a horizontal line corresponds to eq 1 or 3, and a vertical line corresponds to eq 2.

Note that the methods of eqs 4 and 5 were introduced in ref 5, the methods of eqs 11 and 12 were introduced in ref 6, the method of eq 13 was introduced in ref 8, the method of eq 8

MCCM-SRP Study of $C_nH_xO_y$ Compounds

J. Phys. Chem. A, Vol. 105, No. 16, 2001 **4145**

TABLE 1: Coefficients Optimized in This Work over a 54-molecule Data Set Containing Only H, C, and O Atoms and over the Original 82-molecule Data Set[a]

method	version	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7
SAC-MP2/pDZ	SRP-HCO-s	1.0000	1.2660						
SAC-MP4SDQ/pDZ	SRP-HCO-s	1.0000	1.3980						
MCSAC-QCISD/6-31G(d)	SRP-HCO-s	1.0000	1.6169	2.4248					
	2s	1.0000	1.6024	2.4677					
MCCM-UT-QCISD;6-31G (2df,p);6-31G(d)	SRP-HCO-s	0.9808	0.8948	1.1178	1.6591	1.0175			
	2s	0.9776	1.1221	1.1256	1.7296	1.2099			
MC-QCISD	SRP-HCO-s	1.0173	0.9304	1.3959	0.4421				
	2s	1.0014	1.0807	1.2584	0.9625				
MCCM-CO-MP2;MG3;6- 31+G(d)	SRP-HCO-s	0.9976	1.9518	0.8006	1.8793				
	2s	0.9704	1.3625	0.8607	2.0276				
MCCM-CO-MP2	SRP-HCO-s	1.0349	2.0168	0.7502	1.6960				
MCCM-UT-MP4SDQ	SRP-HCO-s	1.0273	1.6167	0.9268	1.1904	0.4652			
MCCM-UT-CCSD	SRP-HCO-s	1.0325	1.8885	0.7867	1.6133	0.0687			
MCG3	SRP-HCO-s	1.0047	0.8935	1.0585	1.2488	0.9434	0.5458	1.0677	1.5834
G3S	SRP-HCO-s	1.0774	1.4276	1.0930	1.1529	1.3163	0.2742		

[a] Spin−orbit contributions are explicit for these versions of the coefficients. For the first 10 methods, they are added for atoms and open-shell molecules as discussed in ref 5, whereas for the G3S theory they are added only for atoms as discussed in ref 10.

was introduced in ref 9, the method of eq 14 was introduced in ref 10, and the methods of eqs 6, 7, 9, and 10 were introduced in ref 11.

In addition, we compare our results with the G3[26] method. The G3 energy expression is

$$
\begin{aligned}
E(\text{G3}) = {} & E[\text{QCISD(T)/6-31G(d)}] + \\
& \Delta E[\text{MP4/6-31+G(d)}|\text{6-31G(d)}] + \\
& \Delta E[\text{MP4/6-31G(2df,p)}|\text{6-31G(d)}] + \\
& \Delta E[\text{MP2(full)/G3large}|\text{6-31G(2df,p)}] - \\
& \Delta E[\text{MP2/6-31+G(d)}|\text{6-31G(d)}] + E(\text{SO}) + \\
& \qquad\qquad\qquad\qquad\qquad E(\text{HLC-G3}) \quad (15)
\end{aligned}
$$

where

$$
E(\text{HLC-G3}) = \sum_{i=1}^{4} c_i h^{(\text{G3})}{}_i \quad (16)
$$

where $h^{(\text{G3})}{}_1$ is $(n_\alpha - n_\beta)$ and $h^{(\text{G3})}{}_2$ is $(n_\alpha + n_\beta)$ for atoms, and $h^{(\text{G3})}{}_3$ is $(n_\alpha - n_\beta)$ and $h^{(\text{G3})}{}_4$ is $(n_\alpha + n_\beta)$ for molecules.

We note that the G3 and G3S methods are defined to use MP2(full)/6-31G(d) geometries. In contrast, the SAC methods and MCCMs may be used with any reasonable geometries. In this article, for all multilevel methods except G3 and G3S, we use MP2/pDZ geometries for the methods based on pDZ and pTZ basis sets and MP2(full)/6-31G(d) geometries for methods involving the other basis sets.

Note that we have not included core correlation or scalar relativistic effects in any of the SAC or MCCM energy expressions given above; thus, they are included implicitly in the coefficients, except for G3 and G3S where core correlation is explicit because of the use of MP2(full) calculations. We treat spin−orbit effects in two possible ways, denoted "s" for explicit spin−orbit and "m" for "minimal." In the former $E(\text{SO})$ is nonzero for atoms and open-shell molecules, as discussed previously,[5] except for G3 and G3S, where $E(\text{SO})$ is nonzero only for atoms, to follow the precise prescription of the original G3[26] and G3S[10] methods. Spin−orbit effects are implicit in the minimal versions of the other methods. In previous articles we have presented coefficients both for the minimal approach and for approaches where core correlation, spin−orbit effects, or both are included explicitly; the results are very similar on

average. This article considers both the minimal approach and the explicit spin−orbit approach, but because the results are very similar we present only the type-s results in the article, and the type-m results in Supporting Information.

3. Parametrizations

The parameters for all 11 SRP methods were first optimized on the 27 molecules in the previous[9] 82-molecule data set that contains only C, H, and O atoms. We concluded that the 27-molecule subset was too small to yield physical parameters for all 11 methods. Therefore, we added 27 new C, H, and O molecules from the G3/99[27] test suite, bringing the final size of the C, H, O data set up to 54. These results appear quite physical, and all discussion of C, H, O SRPs is based on results with this 54-molecule set.

The zero-point-exclusive atomization energies of the 54 test molecules were obtained from experimental heats of formation at 298 K combined with theoretical calculations of vibrational and rotational contributions; details are provided in Appendix A. All electronic structure calculations were performed with the electronic structure package Gaussian98.[28]

The SRP coefficients are labeled SRP-HCO-s when treating spin−orbit effects explicitly and are given in Table 1. The coefficients obtained with the minimal approach are labeled SRP-HCO-m and are given in Table B-1 in the Supporting Information. The coefficients in both cases were obtained by a least-squares fit to the 54 zero-point-exclusive atomization energies. For those methods where version 2s coefficients were not available, we calculated them and added them to Table 1. Version 2s coefficients are those obtained over the original 82-molecule data set (including B, N, F, Al, P, S, and Cl and well as H, C, and O) including spin−orbit effects explicitly.

For completeness, Table B-4 of the Supporting Information contains a complete list of the 54 molecules in the CHO test set along with their experimental atomization energies. It also contains the values calculated by both the *s* and *m* versions of three of the methods parametrized in this article.

4. Results

Table 2 gives the mean errors. In all cases the mean errors are computed for the 54-molecule test set. For each method, we give two sets of mean errors: the top row is obtained with the present SRP coefficients, and the bottom row is obtained

TABLE 2: Mean Errors[a]

method	n	N	MSE[a]	MUE[a]	RMSE[a]	cost energy[b]	energy[c]	gradient[d]
SAC-MP2/pDZ	5	1	−1.07	9.02	10.92	8	8	3
			−5.31	10.38	11.88			
SAC-MP4SDQ/pDZ	6	1	−0.17	3.96	5.33	20	32	13
			5.90	7.13	8.53			
MCSAC-QCISD/6-31G(d)	6	2	0.03	5.59	9.05	26	52	14
			−2.48	6.41	9.49			
MCCM-UT-QCISD;6-31G(2df,p);6-31G(d)	6	5	−0.44	1.68	2.28	56	90	42
			0.15	2.16	3.06			
MC-QCISD	6	4	−0.20	1.37	2.25	153	209	71
			−0.55	2.01	3.33			
MCCM-CO-MP2;MG3;6-31+G(d)	5	4	−0.20	1.83	2.59	132	163	83
			−3.37	4.23	5.23			
MCCM-CO-MP2	5	4	−0.41	2.33	3.28	191	193	98
			−0.86	3.76	4.59			
MCCM-UT-MP4SDQ	6	5	−0.20	1.58	2.55	202	218	110
			2.15	3.25	4.42			
MCCM-UT-CCSD	6	5	−0.35	2.27	3.15	247	409	938
			3.42	4.23	10.51			
MCG3	7	8	−0.06	0.65	0.84	271	414	1265
			0.02	1.00	1.24			
G3	7	4	−1.00	1.32	1.63	1011	1908	6374
G3S	7	6	0.05	0.70	0.90	1011	1908	6516
			−0.94	1.23	1.64			

[a] All results in this table are based on explicit inclusion of spin−orbit effects. The first row for each method gives the mean errors for the coefficients optimized over the 54-molecule data set. The second row for each method, except G3 (which does not have a separate parametrization for data sets; therefore, does not have a second row entry) and G3S (the second row entry for G3S is for the original coefficients of ref 10 that were obtained from a data set containing 299 pieces of data), gives the mean errors for the coefficients optimized over the 82-molecule data sets. In both cases the mean errors refer to the 54-molecule C, H, O set. [b] CPU time for one isobutane single-point energy calculation on an Origin 2000 with R12000 processors and normalized to the time (14 s) for one HF/6-31G(d) energy calculation. [c] CPU time for one *tert*-butyl radical single-point energy calculation on an Origin 2000 with R12000 processors and normalized to the time (12 s) for one HF/6-31G(d) energy calculation. [d] CPU time for furan (C_4H_4O) gradient calculation on an Origin 2000 with R12000 processors and normalized to the time (83 s) for one HF/6-31G(d) gradient.

with the general parameters optimized for the previous broad 82-molecule data set (or, for G3S, for an even broader data set as explained in a footnote to the table). Table 2 also gives the number N of semiempirical coefficients in each method plus four columns related to computer time, which is a measure of cost. For large systems the computer time for an energy calculation scales as the size of the system to the power n, where n is given in the table. To illustrate the cost more concretely, we also give computer times for two energy calculations and one gradient calculation; each of the three rows corresponding to these calculations is separately normalized to the cost for an HF/6-31G(d) calculation. The computer times give a rough measure of the relative costs of the methods, and they illustrate how this measure depends somewhat on which calculation is used to measure the cost.

For comparison with the MCCMs, Table 3 gives mean errors and costs for traditional single-level calculations, in particular for all the levels that are used as components in any of the MCCMs considered here.

5. Discussion

We follow three conventions to simplify the discussion. (1) In the discussion, when we mention the cost of any method or when we mention a ratio of costs, we will simply use for each method the median of the three cost values shown in Tables 2 and 3. One could use more complicated cost measures, but this seems good enough for discussion purposes. (2) Furthermore, we will use the mean unsigned error for discussion purposes. Those interested in mean signed error or root-mean-squared error can consult the tables. (3) Because all energies are molar, we say kilocalories rather than kilocalories per mole.

Before discussing the specific MCCMs, it is constructive to make an interpretative remark about Table 3. Note that the mean errors on the 54-molecule HCO test set tend to be considerably larger than the mean errors on the original 82-molecule test set. This happens because the 27 new molecules we added in this article tend to be larger and hence more difficult than the average molecule in the broad test set. In particular, there are 236 bonds (counting all types of bonds, even double bonds and triple bonds, as one bond) in the 27 new C, H, O molecules. This may be compared with 107 bonds in the original 27 C, H, O molecules. Thus it is always important to keep in mind the test set when one quotes errors. In the rest of the discussion all errors that we mention are for the 54-molecule CHO test set.

Consider the extent to which SRP parametrization improves the accuracy when compared with using general parameters. One might naively expect that less would be gained with more accurate models, at least if accuracy and robustness go hand in hand. We did not find, however, a good correlation of percentage gain with initial mean error of the method. For the 11 SRP models in Table 2, the decrease in mean unsigned error for the C, H, O compounds ranges from 13 to 57%, with a mean of 36%.

The least expensive method in Table 2 is SAC-MP2/pDZ. The SRP version gives a mean unsigned error of 9.0 kcal with a cost of 8, only slightly better than the error of the general parameters of 10.4 kcal. Nevertheless, the error is better than (smaller than) that for 24 of the 26 single-level methods in Table 3, and 16 of those 24 methods have costs greater than 8. The second least expensive method in Table 2 is SAC-MP4SDQ/pDZ, with a cost of 20 and an error for the SRP version of 4.0 kcal. The error in the SRP version is 44% lower than for the

MCCM-SRP Study of $C_nH_xO_y$ Compounds

J. Phys. Chem. A, Vol. 105, No. 16, 2001 **4147**

TABLE 3: Mean Errors for All of the Components, with Spin−Orbit Included in Each Case

method	MSE[a]	MUE[a]	RMSE[a]	cost		
				energy[b]	energy[c]	gradient[d]
HF/6-31G(d)	−182.10	182.10	199.61	1	1	1
	−123.79	123.79	143.13			
HF/6-31+G(d)	−184.01	184.01	202.15	2	2	1
	−124.64	124.64	144.46			
HF/pDZ	−188.12	188.12	206.79	3	2	1
	−128.90	128.90	149.62			
MP2/6-31G(d)	−49.04	49.04	55.09	2	3	2
	−28.37	28.54	33.94			
MP2/6-31+G(d)	−50.52	50.52	57.10	4	5	2
	−28.71	28.71	34.01			
MP2/pDZ	−40.37	40.37	44.74	8	8	2
	−27.78	27.78	31.33			
MP3/6-31G(d)	−60.22	60.22	65.57	4	9	5
	−41.02	41.02	45.76			
HF/6-31G(2df,p)	−174.98	174.98	192.48	9	6	6
	−116.58	116.58	136.04			
MP4SDQ/6-31G(d)	−61.44	61.44	67.38	40	85	7
	−40.52	40.52	45.88			
MP3/6-31+G(d)	−62.36	62.36	68.22	9	17	10
	−41.99	41.99	46.99			
QCISD/6-31G(d)	−62.65	62.65	69.06	26	52	11
	−40.97	40.97	46.64			
MP4SDQ/pDZ	−53.68	53.68	59.09	20	32	12
	−40.76	40.76	47.17			
HF/pTZ	−178.86	178.86	197.20	47	28	21
	−120.00	120.00	140.02			
MP2/6-31G(2df,p)	−11.90	13.15	15.22	30	38	22
	−4.57	10.17	12.59			
HF/MG3	−178.51	178.51	196.62	41	25	23
	−117.93	117.93	137.89			
HF/G3large	−177.86	177.86	196.16	50	38	29
	−117.35	117.35	137.21			
MP2/MG3	−9.17	10.82	12.94	127	157	57
	−2.55	8.17	10.17			
MP2/pTZ	−6.75	8.30	9.88	183	185	74
	−3.87	6.97	8.70			
MP3/6-31G(2df,p)	−23.13	23.13	24.86	69	139	97
	−18.38	18.38	20.82			
MP2(full)/G3large	−6.27	9.04	10.71	217	250	111
	−0.65	8.55	10.52			
MP4SDQ/6-31G(2df,p)	−26.32	26.32	28.67	79	153	154
	−19.22	19.22	22.11			
MP4/6-31G(d)	−54.11	54.11	59.65	40	85	329
	−34.17	34.17	38.89			
QCISD(T)/6-31G(d)	−57.05	57.05	62.87	65	103	670
	−36.25	36.25	41.26			
MP4/6-31+G(d)	−55.77	55.77	61.84	78	182	785
	−34.55	34.55	39.42			
CCSD/pDZ	−54.47	56.25	62.01	65	224	840
	−42.23	42.23	49.19			
MP4/6-31G(2df,p)	−16.40	16.40	18.02	651	1373	4453
	−10.59	10.89	12.70			

[a] All results in this table are based on explicit spin−orbit. First row, over 54-molecule test suite; second row, over 82-molecule test suite. [b] CPU time for one isobutane single-point energy calculation on an Origin 2000 with R12000 processors and normalized to the time (14 s) for one HF/6-31G(d) calculation. [c] CPU time for one *tert*-butyl radical single-point energy calculation on an Origin 2000 with R12000 processors and normalized to the time (12 s) for one HF/6-31G(d) calculation. [d] CPU time for one C_4H_4O gradient calculation on an Origin 2000 with R12000 processors and normalized to the time (83 s) for one HF/6-31G(d) gradient.

general parametrization version and is more than a factor of 2 lower than *every* single-level method in Table 3, even though eight of these have costs over 100.

For the other nine methods, the SRP parameters lower the errors compared with the general parameters by factors of 1.1, 1.3, 1.5, 2.3, 1.6, 2.1, 1.9, 1.1, and 1.8, respectively, which are significant improvements in most cases. We therefore judge the SRP reparametrization effort to be a success, but we are also encouraged that the errors do not improve by exorbitant amounts; that speaks well for the robustness of the general parametrizations.

The most successful $n = 5$ method in Table 2 is MCCM-CO-MP2;MG3;6-31+G(d). The error is 1.8 kcal at a cost of 132. The most successful $n = 6$ method is MC-QCISD with an error of 1.4 kcal and a cost of 153. The most successful $n = 7$ method is MCG3 with an error of 0.65 kcal and a cost of 414. With 6.35 bonds per molecule (computed from the numbers given above), these three values correspond, respectively, to 0.29, 0.22, and 0.10 kcal/mol of bonds, which is a dramatic confirmation of the fact that one obtains "chemical accuracy" (usually defined, for a bond energy, as 1 kcal/mol or better). If the methods were to be applied to larger molecules, we would

hope that the mean error per bond would not increase significantly because the methods are size-consistent.

6. Concluding Remarks

Quantum chemical methods with empirical and/or extrapolatory elements such as Austin Model 1[29] (AM1), the SAC extrapolation method,[1,5,30] Gaussian-3 [e.g., G3[26] and G3-(MP2)[31]] theory, the complete-basis-set extrapolation methods (e.g., CBS-4[32] and CBS-QB3[33]), hybrid Hartree−Fock density-functional theory (e.g., B3LYP[34] and MPW1PW91[35]), and Weizmann methods (e.g., W2[36]) have demonstrated tremendous power for computational thermochemistry and the calculation of potential energy surfaces; recent reviews from various points of view are available.[37−40] A widely recognized strength of these methods is their general model chemistry character by which they provide a well-defined, unique answer for any chemical problems that are posed. A corresponding weakness is that the accuracy is also fixed, sometimes gratifyingly high and other times not high enough.

There are various ways one can try to increase the accuracy of quantum chemistry predictions. One approach, which increases the cost, is to raise the level of theory on which the method is based, for example, from G3(MP2) to G3, from CBS-4 to CBS-QB3, or from MC-QCISD to MCG3. A second approach, which does not raise the cost, is to parametrize against more apposite data. For example, a dynamicist who needs a complete potential surface for the reaction OH + CH$_4$ → H$_2$O + CH$_3$ might fit the parameters of an affordable method to limited experimental data or high-level theoretical data on reactants or products and perhaps on a few additional structures such as the saddle point. Another practitioner, who is interested in a wide range of reactions involving O, OH, HO$_2$, and hydrocarbons, oxohydrocarbons, and hydroxyhydrocarbons and their fragments, might fit the parameters of a selected level of theory to data on compounds containing only H, C, and O. By leaving nitrogen-, sulfur-, metal-, and halogen-containing compounds out of the training set, the model becomes less robust in general but also less compromised for H, C, O compounds. The former approach is called specific reaction parameters (SRP), and the latter approach, which is tested in this article, is called specific-range reaction parameters or specific range parameters (also SRP).

The SRP models presented here were parametrized against the energies of 54 atomization energies, but the method is more general. One could use molecular geometries, dipole moments, solvation energies, ionization potentials, or whatever data are available and relevant for the compounds of interest. The border between general and SRP models is sometimes ambiguous. For example, one might intend a model to be general but later find that it has an SRP character due to deficiencies in the original collection of training data. Our emphasis here, though, is the designed SRP models, and the advantages of well-designed SRP models seem to be underappreciated. In the present case we find that the gain in accuracy in using an SRP model depends on the model. Although SRP methods were tested here against general parametrizations, one can also build SRP models in cases where general parametrizations do not exist. In addition, although this article involves single-point energy calculations, the MCCMs can also be used for geometry optimizations.[43]

MCCM calculations with either general or specific reaction parameters may be performed by the MULTILEVEL code,[44] which is available over the Internet.[45]

Appendix A. Experimental Zero-Point-Exclusive Atomization Energies

The experimental zero-point-exclusive atomization energies for the 27 new molecules in the data set were derived from experimental $\Delta_f H^0_{298}$ data.[41] The experimental data for the 27 molecules from the original 82-molecule data set are described in a previous work.[9]

The experimental zero-point-exclusive atomization energy for molecule C$_x$H$_y$O$_z$ was calculated from its experimental enthalpy of formation at 298 K[41] by the following:

$$\Sigma D_e \, (C_xH_yO_z) = x\Delta_f H° \, (C, 0 \, K) - x[H° \, (C, 298 \, K) - $$
$$H° \, (C, 0 \, K)] + y\Delta_f H° \, (H, 0 \, K) - y[H° \, (H, 298 \, K) - $$
$$H° \, (H, 0 \, K)] + z\Delta_f H° \, (O, 0 \, K) - z[H° \, (O, 298 \, K) - $$
$$H° \, (O, 0 \, K)] \Delta_f H° \, (C_xH_yO_z, 298 \, K) + $$
$$[H° \, (C_xH_yO_z, 298 \, K) - H° \, (C_xH_yO_z, 0 \, K)] + ZPE$$

The experimental values, given in Table 4, for atomic $\Delta_f H°$ (0 K) and [$H°$ (298 K) − $H°$ (0 K)] are taken from ref 42 and the experimental values for $\Delta_f H°$ (C$_x$H$_y$O$_z$, 298 K) are taken from ref 41. The molecular thermal contribution to the enthalpy [$H°$ (C$_x$H$_y$O$_z$, 298 K) − $H°$ (C$_x$H$_y$O$_z$, 0 K)] and the zero-point energy (ZPE) were obtained from MP2/cc-pVDZ geometry and frequency calculations, with the frequencies scaled by 0.9790 as described previously.[5]

TABLE 4

atom	$\Delta_f H^0$ (0 K)	$[H^0(298 \, K) - H^0(0 \, K)]_{st}$
H	51.63	1.01
C	169.98	0.25
O	58.99	1.04

Supporting Information Available: Appendix B is in the Supporting Information for this journal. Appendix B contains coefficients for the SRP minimal methods, mean errors for the minimal methods and their components without spin−orbit contributions, experimental atomization energies for all 54 C, H, O compounds, and calculated atomization energies for these compounds by the *s* and *m* versions of the following methods: MCCM-CO-MP2; MG3,6-31+G(d), MC-QCISD, and MCG3. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Gordon, M. S.; Truhlar, D. G. *J. Am. Chem. Soc.* **1986**, *108*, 5412.
(2) Gordon, M. S.; Truhlar, D. G. *Int. J. Quantum Chem.* **1987**, *31*, 81.
(3) Gordon, M. S.; Nguyen, K. A.; Truhlar, D. G. *J. Phys. Chem.* **1989**, *93*, 7356.
(4) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *234*, 64.
(5) Fast, P. L.; Corchado, J.; Sánchez, M. L.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 3139.
(6) Fast, P. L.; Corchado, J. C.; Sánchez, M. L.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 5129.
(7) Fast, P. L.; Sánchez, M. L.; Corchado, J. C.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 11679.
(8) Fast, P. L.; Sánchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett.* **1999**, *306*, 407.
(9) Tratz, C. M.; Fast, P. L.; Truhlar, D. G. *Phys. Chem. Commun.* **1999**, *2(14)*, 1.
(10) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 1125.
(11) Fast, P. L.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 6111.

MCCM-SRP Study of $C_nH_xO_y$ Compounds

J. Phys. Chem. A, Vol. 105, No. 16, 2001 **4149**

(12) Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618.

(13) Liu, Y.-P.; Lu, D.-h.; Gonzalez-Lafont, A.; Truhlar, D. G.; Garrett, B. C. *J. Am. Chem. Soc.* **1993**, *115*, 7806.

(14) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231.

(15) Barrows, S. E.; Dulles, F. J.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *Carbohydr. Res*. **1995**, *276*, 219.

(16) Bash, P. A.; Ho, L. L.; MacKerell, A. D., Jr.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3698.

(17) Sekusak, S.; Cory, M. G.; Bartlett, R. J.; Sabljic, A. *J. Phys. Chem. A* **1999**, *103*, 17394.

(18) Sekusak, S.; Piecuch, P.; Bartlett, R. J.; Cory, M. G. *J. Phys. Chem. A* **2000**, *104*, 8779.

(19) Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *294*, 45.

(20) Fast, P. L.; Sánchez, M. L.; Truhlar, D. G. *J. Chem. Phys*. **1999**, *111*, 2921.

(21) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. In *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(22) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910.

(23) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(24) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(25) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358.

(26) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rasslov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.

(27) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374.

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

(29) Dewar, M. J. S.; Zoebisch, E.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(30) Irikura, K. K.; Frurip, D. J. *ACS Symp. Ser.* **1998**, *677*, 419.

(31) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *1109*, 4703.

(32) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A., Jr. *J. Chem. Phys*. **1996**, *104*, 2598.

(33) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys*. **1999**, *110*, 2822.

(34) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(35) Adamo, C.; Barone, V. *J. Chem. Phys*. **1998**, *108*, 664.

(36) Martin, J. M. L.; DeOliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843.

(37) Corchado, J. C.; Truhlar, D. G. *ACS Symp. Ser.* **1998**, *712*, 106.

(38) Klopper, W.; Bak, K. L.; Jorgensen, P.; Olsen, J.; Helgaker, T. *J. Phys. B* **1999**, *32*, R103.

(39) Curtiss, L. A.; Redfern, P. C.; Frurip, D. J. *Rev. Comput. Chem.* **2000**, *15*, 147.

(40) Dunning, T. H., Jr. *J. Phys. Chem. A* **2000**, *104*, 9062.

(41) Petersson, G. A.; Malik, D. K.; Wilson, W. G.; Ochterski, J. W.; Montgomery, J. A., Jr.; Frisch, M. J. *J. Chem. Phys*. **1998**, *109*, 10570.

(42) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106* 1063.

(43) Rodgers, J. M.; Fast, P. L.; Truhlar, D. G. *J. Chem. Phys*. **2000**, *112*, 3141.

(44) Rodgers, J. M.; Lynch, B. J.; Fast, P. L.; Pu, J.; Truhlar, D. G. MULTILEVEL-version 2.2; University of Minnesota, Minneapolis, 2000.

(45) http://comp.chem.umn.edu/multilevel.