

RESEARCH ARTICLE | JUNE 15 1998

Optimized density functionals from the extended G2 test set



Hartmut L. Schmider; Axel D. Becke



J. Chem. Phys. 108, 9624–9631 (1998)

<https://doi.org/10.1063/1.476438>



View
Online



Export
Citation

CrossMark

The Journal of Chemical Physics
Special Topic: Adhesion and Friction
Submit Today!

AIP Publishing

Optimized density functionals from the extended G2 test set

Hartmut L. Schmider and Axel D. Becke

Department of Chemistry, Queen's University, Kingston, Ontario, Canada K7L 3N6

(Received 21 January 1998; accepted 10 March 1998)

A recently suggested procedure for the systematic optimization of gradient-corrected exchange-correlation functionals [A. D. Becke, *J. Chem. Phys.* **107**, 8554 (1997)] has been applied to the extended G2 test set [L. A. Curtiss *et al.*, *J. Chem. Phys.* **106**, 1063 (1997)], which consists of the standard heats of formation of 148 molecules. The limit of reproduction of the experimental data in this test set is found to be 1.78 kcal/mol mean absolute error, with a maximum of 8.89 kcal/mol error for the ozone molecule. This compares rather well with previous results for G2 theory itself (1.58 and 8.2 kcal/mol, respectively). We show that fair stability can be obtained by our optimization procedure. © 1998 American Institute of Physics. [S0021-9606(98)00623-0]

I. INTRODUCTION

In the course of the last decade, there has been substantial progress in the refinement of energy density functionals. As a consequence, density functional theory (DFT) has become a viable alternative to conventional quantum mechanical techniques for the reproduction and prediction of spectroscopic and thermochemical data (e.g. Refs. 1–6). Early work in the field was dominated by the local (spin) density approximation (LSDA); however, the severe overbinding of this method makes corrections depending on density derivatives absolutely necessary.^{2,3} If only the first-order gradient is employed, the resulting functionals are known as “generalized gradient approximations” (GGAs). A great variety of such functionals is available today, and their accuracy has been tested and proven time and again.

However, the exact form of the exchange-correlation functional $E_{XC}[\rho]$ remains elusive. Another step in the direction of greater accuracy was the introduction of hybrid methods,^{4,5} which use a small proportion of “exact exchange” on the basis of the so-called adiabatic connection formula. Functionals involving the Laplacian of the density have also entered the DFT instrumentarium (e.g. Refs. 7, 8), and techniques to obtain “exact” potentials from highly accurate molecular densities are now employed to gain further insights into the nature of E_{XC} (e.g. Ref. 9). The search for better approximations continues, as density functional methods become—at least for some applications—compatible in accuracy with rather evolved post-Hartree-Fock methods.

In a recent paper,⁶ the last of a series,^{2–6} a systematic procedure to optimize approximate exchange-correlation functionals at the (hybrid-) GGA level was suggested and applied to benchmark systems. Here, we briefly review the method, and assess it further. It is based on a remapping of the dimensionless reduced gradient variable

$$s_{\sigma} = \frac{|\nabla \rho_{\sigma}|}{\rho_{\sigma}^{4/3}}, \quad (1)$$

where σ denotes spin α or β , to a new finite variable $u(s_{\sigma}^2)$. Exchange (subscript X) and correlation (subscript C) functionals are written in the form:

$$E_X = \sum_{\sigma} \int e_{X\sigma}(\rho_{\sigma}) g_{X\sigma}(s_{\sigma}^2) d\mathbf{r}, \quad (2)$$

$$E_{C\alpha\beta} = \int e_{C\alpha\beta}(\rho_{\alpha}, \rho_{\beta}) g_{C\alpha\beta}(s_{av.}^2) d\mathbf{r}, \quad (3)$$

$$E_{C\sigma\sigma} = \int e_{C\sigma\sigma}(\rho_{\sigma}) g_{C\sigma\sigma}(s_{\sigma}^2) d\mathbf{r}, \quad (4)$$

$$E_C = E_{C\alpha\beta} + \sum_{\sigma} E_{C\sigma\sigma}. \quad (5)$$

The $e(\rho)$ in Eqs. (2)–(4) are local volume energy densities of a uniform electron gas, and the $g(s^2)$ denote “gradient correction factors” which are to be determined. The actual $e(\rho)$ take different forms for each of exchange, like-spin, and unlike-spin correlation (see Ref. 6 for details). We make the simplifying assumption that the unlike-spin correction factor for correlation, $g_{C\alpha\beta}$ (Eq. (3)) depends only on an average gradient variable $s_{av.}^2 = (s_{\alpha}^2 + s_{\beta}^2)/2$. The total correlation energy is just the sum of the like- and unlike-spin contributions (Eq. (5)).

To determine each optimum correction factor $g(s^2)$, it is expanded in a power series in the remapped variable $u(s^2)$:

$$g(s^2) = g(u) = \sum_{i=0}^m c_i u^i(s^2), \quad (6)$$

where we have omitted the subscripts for brevity. Each correction factor $g(u)$ is treated separately, and each mapping of s^2 to u is specific:

$$u_{X\sigma}(s_{\sigma}^2) = \frac{\gamma_{X\sigma} s_{\sigma}^2}{1 + \gamma_{X\sigma} s_{\sigma}^2}, \quad (7)$$

$$u_{C\alpha\beta}(s_{av.}^2) = \frac{\gamma_{C\alpha\beta} s_{av.}^2}{1 + \gamma_{C\alpha\beta} s_{av.}^2}, \quad (8)$$

$$u_{C\sigma\sigma}(s_{\sigma}^2) = \frac{\gamma_{C\sigma\sigma} s_{\sigma}^2}{1 + \gamma_{C\sigma\sigma} s_{\sigma}^2}. \quad (9)$$

This functional form of the remapped finite variables $u(s^2)$ is derived from simple approximate expressions for gradient

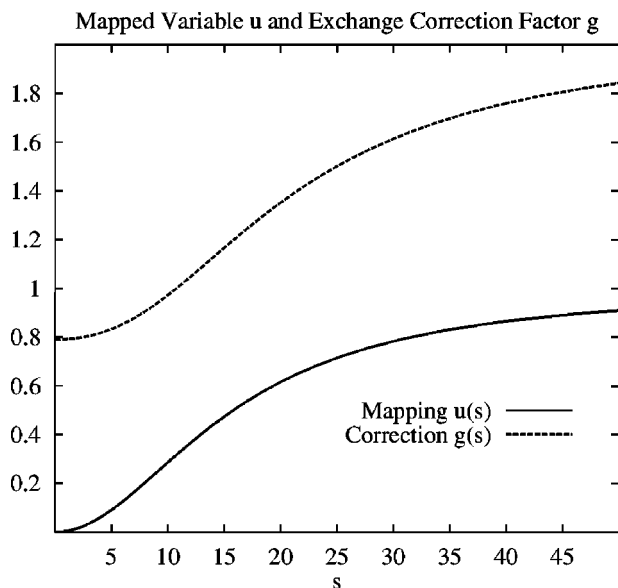


FIG. 1. Remapping of the reduced gradient $s = |\nabla\rho|/\rho^{4/3}$ onto the finite independent variable $u = \gamma s^2/(1 + \gamma s^2)$. The graph shows both a typical exchange correction factor $g(s)$ and the mapped variable $u(s)$ as a function of s . The example chosen here is the exchange correction for fit 2c in Table II, where $\gamma = 0.004$.

corrections.^{6,10} The parameters γ need to be chosen for each contribution (exchange or correlation, like or unlike spin) independently. Since the γ 's are nonlinear parameters, and hence fitting to extensive molecular data is inconvenient, atom-optimized values are chosen instead and kept fixed (see Ref. 6 for more details; the values used in this work are $\gamma_{X\sigma} = 0.004$, $\gamma_{C\alpha\beta} = 0.006$ and $\gamma_{C\sigma\sigma} = 0.2$). In contrast, the expansion coefficients c_i are linear parameters and may be determined from experimental thermochemical data directly by means of straightforward linear least-squares fitting.

To illustrate the mapping from the reduced gradient s to the finite variable u , we show $u(s)$ for exchange (Eq. (7)) in Fig. 1 (full curve). The dotted curve shows a typical exchange correction factor $g_{X\sigma}(s_\sigma)$ for one of the fits (2c in

TABLE I. Least-squares residuals Δ , root-mean-squares deviations, and GoFs for fits to 233 thermochemical data.

Order	No. par.	Δ (kcal ² /mol ²)	RMS (kcal/mol)	GoF
0	4	21160	9.53	9.6
1	7	2313	3.15	3.2
2	10	1847	2.82	2.9
3	13	1820	2.79	2.9
4	16	1774	2.76	2.9
5	19	1723	2.72	2.8
6	22	1686	2.69	2.8

Tables II and III) obtained in the course of this work. From the obvious similarity of the overall shape of $u(s)$ and $g(s)$, we see how the mapping of s onto u justifies the power series Eq. (6) of the functional relationship between g and the independent variable u . Obviously, it is beneficial to choose the value of γ such that this shape similarity is observed.

A linear exact-exchange mixing parameter c_{exact} may also be introduced. The full energy expression for E_{XC} is then

$$E_{\text{XC}} = E_X + E_C + c_{\text{exact}} E_X^{\text{exact}}, \quad (10)$$

making this a ‘‘hybrid’’ scheme. E_X and E_C are given by Eqs. (2) and (5), respectively, but E_X in Eq. (10) does not now represent the full exchange energy of the system. If the series expansions Eq. (6) for all three contributions (exchange, $\alpha\beta$ - and $\sigma\sigma$ -correlation) are truncated uniformly at order m , we have $p = 3(m + 1) + 1$ linear parameters to be determined.

Recently,¹¹ Curtiss and co-workers have published a compilation of 148 experimentally determined standard heats of formation that were used to gauge the *ab initio* procedure known as G2 theory.¹² These test data are the extension of an earlier compilation^{12,13} that was centered on zero-temperature atomization energies. The authors also compared commonly used DFT methods with their scheme. They concluded that the performance of standard DFT methods is

TABLE II. Residual mean absolute errors in several properties for a series of 10-parameter linear least-squares fitted functionals. The fit sets are: 1a = 56 HoFs (298 K) (original G2 set); 1b = 1a + 42 ionization pot. + 25 electron aff. + 8 proton aff.; 1c = 1b + 10 tot. energies; 2a = 148 HoFs (298 K) (extended G2 set); 2b and 2c like 1b and 1c, respectively. The values in brackets are the largest absolute errors.

Property	1a	1b	1c
56 HoFs (298 K, G2-1) [kcal/mol]	1.50 (4.69)	1.77 (7.10)	1.70 (5.68)
148 HoFs (298 K, G2-2) [kcal/mol]	2.15 (12.85)	2.60 (14.51)	2.15 (12.84)
42 Ionization Potentials [eV]	0.20 (0.59)	0.10 (0.54)	0.12 (0.59)
25 Electron Affinities [eV]	0.24 (0.58)	0.08 (0.26)	0.09 (0.33)
8 Proton Affinities [kcal/mol]	1.1 (3.2)	1.1 (4.3)	1.2 (4.9)
10 Total Energies [mhartree]	49 (180)	155 (379)	3 (9)
	2a	2b	2c
56 HoFs (298 K, G2-1) [kcal/mol]	1.72 (6.59)	1.76 (5.03)	1.75 (5.89)
148 HoFs (298 K, G2-2) [kcal/mol]	1.78 (8.89)	1.88 (8.35)	1.89 (9.11)
42 Ionization Potentials [eV]	0.58 (1.26)	0.12 (0.65)	0.12 (0.66)
25 Electron Affinities [eV]	0.49 (0.79)	0.09 (0.33)	0.09 (0.34)
8 Proton Affinities [kcal/mol]	1.7 (4.7)	1.2 (3.5)	1.1 (4.3)
10 Total Energies [mhartree]	90 (135)	71 (163)	4 (7)

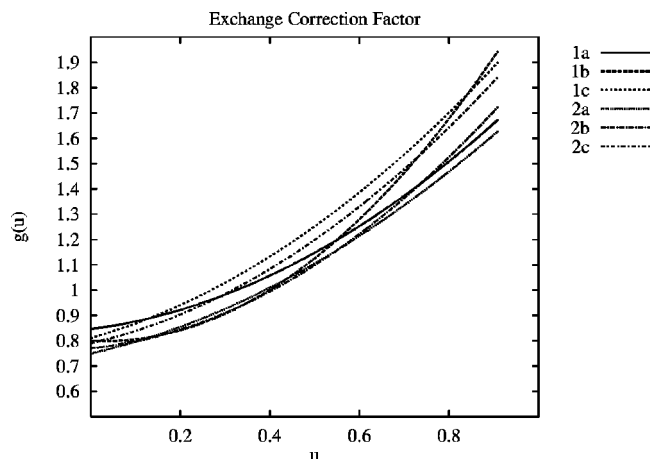


FIG. 2. Exchange correction factors $g_{X\sigma}(u_{X\sigma})$ as a function of the mapped variable $u_{X\sigma} = 0.004s_{\sigma}^2/(1 + 0.004s_{\sigma}^2)$ (also see Fig. 1). The different curves correspond to the fits in Table II. Note that due to the small size of γ , values of $u_{X\sigma}$ close to 1 are never reached.

considerably poorer than G2, showing wider distributions of error, and that hybrid schemes perform better than “pure” Kohn-Sham DFT methods.

In this paper, we will use the new G2 test set to assess the optimized DFT of Ref. 6, and to determine the limit of precision possible in this GGA framework by fitting the gradient corrections directly to the full set of available thermochemical information.

II. RESULTS AND DISCUSSION

We have used as many data points for the determination of an optimal gradient corrected density functional as possible. We have combined the 148 heats of formation (HoFs, at standard temperature) given in Ref. 11 with previously published experimental values for 42 ionization potentials (IPs), 25 electron affinities (EAs),¹⁴ 8 proton affinities (PAs, see Ref. 13), and 10 total atomic energies (H to Ne, TEs),⁶ arriving at a total of 233 thermochemical data as a test set.

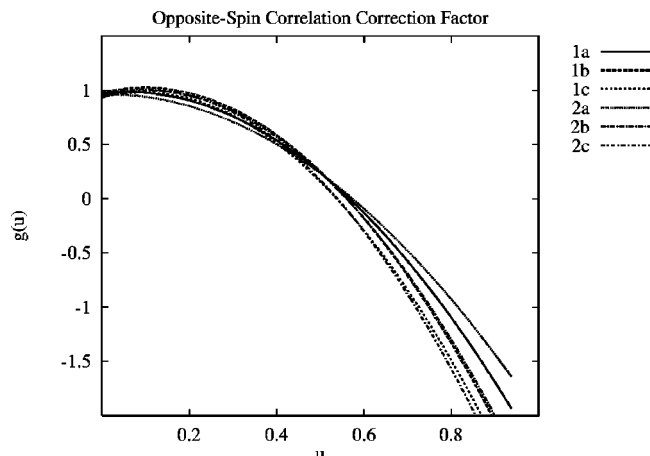


FIG. 3. Unlike-spin correlation correction factors $g_{C\alpha\beta}(u_{C\alpha\beta})$ as a function of the mapped variable $u_{C\alpha\beta} = 0.003(s_{\alpha}^2 + s_{\beta}^2)/(1 + 0.003(s_{\alpha}^2 + s_{\beta}^2))$. The different curves correspond to the fits in Table II. Note that due to the small size of γ , values of $u_{C\alpha\beta}$ close to 1 are never reached.

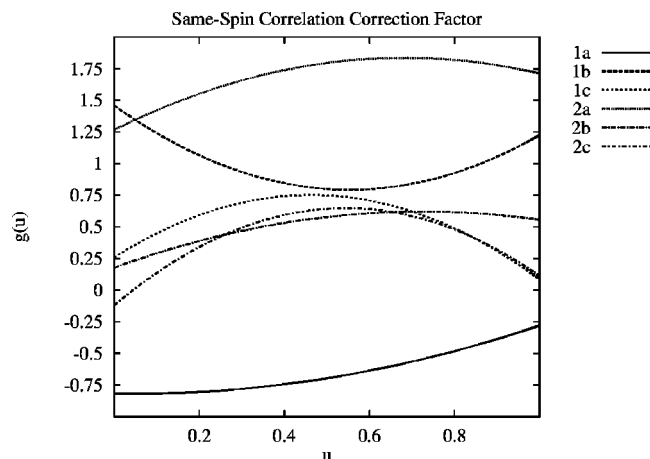


FIG. 4. Like-spin correlation correction factors $g_{C\sigma\sigma}(u_{C\sigma\sigma})$ as a function of the mapped variable $u_{C\sigma\sigma} = 0.2s_{\sigma}^2/(1 + 0.2s_{\sigma}^2)$. The different curves correspond to the fits in Table II.

The required electronic structure calculations were performed by the fully numerical DFT program NUMOL,^{15–18} employing grids with 40 radial points per principal atomic “shell” on each center, and 302 angular points at each radius (194 for hydrogen and helium). We believe this grid to be reasonably saturated, on the basis of comparisons with results obtained on the smaller grids of our previous work^{2–6} (i.e., 20 radial points per atomic shell times 194, or 110 for H and He). The overall statistics of our fits differ insignificantly with the two meshes. Computations were performed “post-LSDA,” namely, the densities computed are LSDA densities, and all gradient-corrected energies are derived from these. Atomic cores were frozen, with the exception of the alkali atoms when compared with their cations. The latter

TABLE III. Linear expansion coefficients for the 10-parameter fits in Table II. The coefficients are grouped according to exchange, unlike-spin and like-spin correlation, and exact exchange parameter.

Coefficient	1a	1b	1c
$c_{X,0}$	0.845975	0.800103	0.810936
$c_{X,1}$	0.228183	-0.084192	0.496090
$c_{X,2}$	0.749949	1.47742	0.772385
$c_{C\alpha\beta,0}$	0.975483	0.977621	0.939269
$c_{C\alpha\beta,1}$	0.398379	0.931199	0.898121
$c_{C\alpha\beta,2}$	-3.73540	-4.76973	-4.91276
$c_{C\sigma\sigma,0}$	-0.817637	1.44946	0.262077
$c_{C\sigma\sigma,1}$	-0.054676	-2.37073	2.12576
$c_{C\sigma\sigma,2}$	0.592163	2.13564	-2.30465
c_{exact}	0.229015	0.199352	0.192416
	2a	2b	2c
$c_{X,0}$	0.749200	0.770587	0.790194
$c_{X,1}$	0.402322	0.180767	0.400271
$c_{X,2}$	0.620779	0.955246	0.832857
$c_{C\alpha\beta,0}$	0.964641	0.965362	0.934715
$c_{C\alpha\beta,1}$	0.050527	0.863300	1.14105
$c_{C\alpha\beta,2}$	-3.01966	-4.61778	-5.33398
$c_{C\sigma\sigma,0}$	1.26686	0.170473	-0.120163
$c_{C\sigma\sigma,1}$	1.67146	1.24051	2.82332
$c_{C\sigma\sigma,2}$	-1.22565	-0.862711	-2.59412
c_{exact}	0.232055	0.237978	0.219847

TABLE IV. Heats of formation of the 148 molecules in the G2 extended test set (Ref. 11) in kcal/mol. The column labeled "present work" gives the values obtained with the parameters derived from a fit to the full set of 233 thermochemical data (fit 2c in Tables II and III).

Mol. No.	System	Present work	Expt.	Dev.
1	H ₂	1.71	0.00	1.71
2	LiH	34.72	33.30	1.42
3	BeH	77.87	81.70	-3.83
4	CH	142.24	142.50	-0.26
5	³ CH ₂	93.42	93.70	-0.28
6	¹ CH ₂	103.78	102.75	1.03
7	CH ₃	34.22	35.00	-0.78
8	CH ₄	-17.72	-17.90	0.18
9	³ NH	83.43	85.20	-1.77
10	NH ₂	42.18	45.10	-2.92
11	NH ₃	-11.59	-10.97	-0.62
12	OH	8.56	9.40	-0.84
13	H ₂ O	-56.14	-57.80	1.66
14	HF	-63.71	-65.14	1.43
15	¹ SiH ₂	63.87	65.20	-1.33
16	³ SiH ₂	84.46	86.20	-1.74
17	SiH ₃	45.93	47.90	-1.97
18	SiH ₄	7.32	8.20	-0.88
19	PH ₂	29.70	33.10	-3.40
20	PH ₃	0.50	1.30	-0.80
21	H ₂ S	-4.80	-4.90	0.10
22	HCl	-21.70	-22.06	0.36
23	Li ₂	53.74	51.60	2.14
24	LiF	-78.31	-80.10	1.79
25	C ₂ H ₂	56.84	54.19	2.65
26	C ₂ H ₄	12.40	12.54	-0.14
27	C ₂ H ₆	-19.81	-20.08	0.27
28	CN	108.65	104.90	3.75
29	HCN	33.18	31.50	1.68
30	CO	-24.25	-26.42	2.17
31	HCO	7.98	10.00	-2.02
32	H ₂ CO	-26.18	-25.96	-0.22
33	CH ₃ OH	-47.60	-48.00	0.40
34	N ₂	4.03	0.00	4.03
35	N ₂ H ₄	19.74	22.79	-3.05
36	NO	19.91	21.58	-1.67
37	³ O ₂	-3.43	0.00	-3.43
38	H ₂ O ₂	-31.17	-32.53	1.36
39	F ₂	1.12	0.00	1.12
40	CO ₂	-95.63	-94.05	-1.58
41	Na ₂	28.07	33.96	-5.89
42	³ Si ₂	137.16	139.87	-2.71
43	P ₂	35.41	34.31	1.10
44	³ S ₂	26.02	30.74	-4.72
45	Cl ₂	-1.69	0.00	-1.69
46	NaCl	-43.37	-43.56	0.19
47	SiO	-21.36	-24.64	3.28
48	CS	69.08	66.90	2.18
49	³ SO	-1.49	1.20	-2.69
50	ClO	21.88	24.19	-2.31
51	ClF	-14.01	-13.24	-0.77
52	Si ₂ H ₆	16.90	19.10	-2.20
53	CH ₃ Cl	-20.07	-19.56	-0.51
54	CH ₃ SH	-5.72	-5.50	-0.22
55	HOCl	-18.25	-17.80	-0.45
56	SO ₂	-66.73	-70.95	4.22
57	BF ₃	-269.61	-271.41	1.80
58	BCl ₃	-97.73	-96.30	-1.43
59	AlF ₃	-282.03	-289.03	7.00
60	AlCl ₃	-140.03	-139.72	-0.31
61	CF ₄	-223.81	-223.04	-0.77
62	CCl ₄	-20.85	-22.94	2.09

TABLE IV. (Continued.)

Mol. No.	System	Present work	Expt.	Dev.
63	OCS	-36.95	-33.08	-3.87
64	CS ₂	23.25	27.95	-4.70
65	COF ₂	-146.71	-152.70	5.99
66	SiF ₄	-376.87	-385.98	9.11
67	SiCl ₄	-157.18	-158.40	1.22
68	N ₂ O	19.15	19.61	-0.46
69	ClNO	9.84	12.36	-2.52
70	NF ₃	-35.89	-31.57	-4.32
71	PF ₃	-226.41	-229.07	2.66
72	O ₃	41.38	34.10	7.28
73	F ₂ O	4.98	5.86	-0.88
74	ClF ₃	-43.18	-37.97	-5.21
75	C ₂ F ₄	-164.53	-157.40	-7.13
76	C ₂ Cl ₄	-3.44	-2.97	-0.47
77	CF ₃ CN	-116.80	-118.40	1.60
78	CH ₃ CCH (propyne)	45.64	44.20	1.44
79	CH ₂ CCH ₂ (allene)	42.91	45.50	-2.59
80	C ₃ H ₄ (cyclopropene)	68.95	66.20	2.75
81	CH ₃ CHCH ₂ (propylene)	5.46	4.78	0.68
82	(CH ₂) ₃ (cyclopropane)	14.01	12.70	1.31
83	C ₃ H ₈ (propane)	-24.30	-25.00	0.70
84	CH ₂ CHCHCH ₂ (butadiene)	25.80	26.30	-0.50
85	CH ₃ CCCH ₃ (2-butyne)	36.19	34.80	1.39
86	C ₄ H ₆ (methylene cyclopropane)	46.20	47.90	-1.70
87	C ₄ H ₆ (bicyclo-butane)	55.75	51.90	3.85
88	C ₄ H ₆ (cyclo-butene)	39.72	37.40	2.32
89	(CH ₂) ₄ (cyclo-butane)	7.25	6.80	0.45
90	(CH ₃) ₂ CCH ₂ (iso-butene)	-1.79	-4.00	2.21
91	C ₄ H ₁₀ (<i>trans</i> -butane)	-29.16	-30.00	0.84
92	(CH ₃) ₃ CH (iso-butane)	-28.12	-32.07	3.95
93	(C ₂ H ₄) ₂ C (spiro-pentane)	46.17	44.30	1.87
94	C ₆ H ₆ (benzene)	16.67	19.74	-3.07
95	CH ₂ F ₂	-108.33	-107.71	-0.62
96	CHF ₃	-166.51	-166.60	0.09
97	CH ₂ Cl ₂	-22.93	-22.83	-0.10
98	CHCl ₃	-23.32	-24.66	1.34
99	CH ₃ NH ₂	-6.52	-5.50	-1.02
100	CH ₃ CN	18.60	18.00	0.60
101	CH ₃ NO ₂	-18.92	-17.80	-1.12
102	CH ₃ ONO	-16.90	-15.90	-1.00
103	CH ₃ SiH ₃	-6.42	-7.00	0.58
104	HCOOH	-90.94	-90.50	-0.44
105	HCOOCH ₃	-85.00	-85.00	0.00
106	CH ₃ CONH ₂	-59.28	-57.00	-2.28
107	C ₂ H ₄ NH (aziridine)	29.50	30.20	-0.70
108	NCCN	75.52	73.30	2.22
109	(CH ₃) ₂ NH	-4.34	-4.40	0.06
110	C ₂ H ₅ NH ₂	-12.90	-11.30	-1.60
111	H ₂ CCO	-14.84	-11.35	-3.49
112	C ₂ H ₄ O (oxirane)	-12.09	-12.57	0.48
113	CH ₃ CHO	-39.90	-39.70	-0.20
114	(CHO) ₂ (glyoxal)	-50.97	-50.70	-0.27
115	C ₂ H ₅ OH	-54.43	-56.21	1.78
116	CH ₃ OCH ₃	-42.44	-44.00	1.56
117	C ₂ H ₄ S (thioxirane)	18.93	19.60	-0.67
118	(CH ₃) ₂ SO	-34.12	-36.20	2.08
119	C ₂ H ₅ SH	-11.01	-11.10	0.09
120	(CH ₃) ₂ S	-9.61	-8.90	-0.71
121	CH ₂ CHF	-35.25	-33.20	-2.05
122	C ₂ H ₅ Cl	-27.30	-26.80	-0.50
123	CH ₂ CHCl	4.46	8.90	-4.44
124	CH ₂ CHCN	44.99	43.20	1.79
125	(CH ₃) ₂ CO	-52.21	-51.93	-0.28
126	CH ₃ COOH	-102.38	-103.40	1.02
127	CH ₃ COF	-105.42	-105.70	0.28

TABLE IV. (Continued.)

Mol. No.	System	Present work	Expt.	Dev.
128	CH ₃ COCl	-58.97	-58.00	-0.97
129	C ₃ H ₇ Cl	-30.37	-31.52	1.15
130	(CH ₃) ₂ CHOH	-61.25	-65.20	3.95
131	C ₂ H ₅ OCH ₃	-52.38	-51.70	-0.68
132	(CH ₃) ₃ N	-4.87	-5.70	0.83
133	C ₄ H ₄ O (furan)	-8.06	-8.30	0.24
134	C ₄ H ₄ S (thiophene)	26.02	27.50	-1.48
135	C ₄ H ₄ NH (pyrole)	23.44	25.90	-2.46
136	C ₅ H ₅ N (pyridine)	27.52	33.60	-6.08
137	HS	33.11	34.18	-1.07
138	CCH	139.23	135.10	4.13
139	CHCH ₂	69.30	71.60	-2.30
140	CH ₃ CO	-4.87	-2.40	-2.47
141	H ₂ COH	-5.67	-4.08	-1.59
142	CH ₃ O	2.81	4.10	-1.29
143	C ₂ H ₂ O	-0.10	-3.70	3.60
144	CH ₃ S	27.28	29.80	-2.52
145	C ₂ H ₅	27.83	28.90	-1.07
146	(CH ₃) ₂ CH	21.03	21.50	-0.47
147	(CH ₃) ₃ C	11.03	12.30	-1.27
148	NO ₂	3.55	7.91	-4.36

cases also served as a test for the validity of the frozen-core approximation (which they successfully passed). We use the same MP2 molecular geometries as in G2 theory itself, and scaled Hartree-Fock vibrational frequencies for zero-point and finite-temperature corrections.¹¹

In determining the coefficients of a truncated series expansion such as Eq. (6), we need to determine the point of “diminishing returns” for the optimum order m , beyond which the inclusion of further terms is no longer justified. The “goodness of fit” (GoF, see e.g. Ref. 19), is an index in the least-squares interpretation of experimental results that measures the quality of a given fit. It takes the number p of fit parameters, and the number n of data points into account, as well as the least-squares residual Δ :

$$\text{GoF} = \sqrt{\Delta/(n-p)}, \quad (11)$$

$$\Delta = \sum_{i=1}^n (x_{\text{mod},i} - x_{\text{obs},i})^2 / \sigma_i^2.$$

If the weights $1/\sigma_i^2$ are related to the actual experimental statistical error, the GoF should be close to 1 for an appropriate model. In our case, no such statement can be made since we use unit weights throughout. However, it still stands to reason that the optimum number of parameters has been reached when an increase does not yield a lower GoF.

In Table I, the least-squares residuals, root-mean-square deviations and GoFs for fits of the power series (Eq. (6)) are listed for various truncation orders m . It may be seen at a glance that the first nonconstant term (i.e. $m=1$) yields a dramatic improvement in the overall fit, but that after the 2nd-order term ($m>2$), the value of the GoF levels out and changes insignificantly. This means that the lowering of the RMS deviation is offset by an increase in the number of parameters. We therefore deem that truncation at order $m=2$ is optimal, leaving us with 10 parameters: the expansion co-

efficients c_0 , c_1 and c_2 for each of exchange, like-spin and unlike-spin correlation functionals, and the mixing parameter c_{exact} for the exact exchange. This is consistent with our earlier experience,⁶ based on visual inspection of the correction functions $g(u)$ for physical sensibility.

We performed a variety of fits based on subsets of the available data, to test the effect of parameter determination from one type of information on the reproduction of others. Table II lists mean absolute deviations and maximum errors for six 10-parameter fits (order $m=2$ plus an exact-exchange term). Fits of type 1 include the original G2 test set of 56 molecules²⁰ in the form of heats of formation at standard temperature. Fits of type 2 use the extended set of HoFs. These sets were either used alone (1a,2a), or augmented by ionization potentials, electron and proton affinities (1b,2b). 1c and 2c use total energies for the first ten atoms of the Periodic Table as additional constraints.

In general, heats of formation and proton affinities are reproduced rather well, the latter even if not included in the fitting set. On the other hand, electron affinities and ionization potentials are somewhat more problematic if not included in the fit set. It is encouraging that for each fit of the 1-series, the degradation of quality from the small G2 set (which is used in the fit) to the extended set is not bad. For fit 1c, less than 0.5 kcal/mol deterioration is observed in the average, and the quality of that fit for the whole 148-HoF set is still of the order of 2 kcal/mol mean absolute error.

Note that for the heats of formation, the smallest mean error in Table II (1.78 kcal/mol for fit 2a) is only 0.2 kcal/mol above that of G2 theory as reported in Ref. 11. It is better than the modified versions of G2 (Refs. 21–23) tested in that work, and it is a great improvement over standard hybrid methods (B3LYP, see Ref. 11), which were found to yield above 3 kcal/mol error on average. Certainly, these are encouraging results.

As can be seen from the last entry for each fit, the reproduction of total energies is generally poor if they are not included in the “training set.” If they are, however (fits 1c and 2c), their inclusion does not appreciably degrade the reproduction of the other data. In fact, some improve slightly (HoFs, proton affinities). We deem it beneficial to include total energies, since they effect a general stabilization of the fits, in the sense that the quality becomes more “even.”

The question arises: how strongly do differences in the fit data affect the functions $g(u)$ which are the central objects in these fits? To investigate this, we have plotted the $g(u)$ for exchange (Fig. 2), unlike-spin correlation (Fig. 3) and like-spin correlation (Fig. 4) for all the fits listed in Table II. While for both the exchange and the unlike-spin ($\alpha\beta$) correlation, the overall shape remains stable for all the fits, the like-spin correlation quantity $g_{C\sigma\sigma}$ is excessively sensitive towards the data that were used in the training set. This is also apparent from the values of the expansion coefficients c_i themselves. We give them in Table III for the 6 fits listed in Table II. The zero-order value c_0 for each exchange and correlation correction determines the overall position of the $g(u)$ curves, and also the value of the energy functional in the uniform electron gas limit. It can be seen from Table III that, throughout the fits, it varies only be-

tween 0.75 and 0.85 for exchange, and between 0.93 and 0.98 for unlike-spin correlation. However, for the like-spin correlation, it varies considerably, in both magnitude and sign. We can only conclude that the like-spin correlation is not well determined by the present data. This is perhaps not surprising, since $\sigma\sigma$ -correlation is by far the smallest part of the exchange-correlation energy. It is also the part most strongly affected by self-correlation error, making it the most difficult to describe appropriately. Note, however, that fits 1c and 2c (finely dashed and dash-dotted curves, respectively) do show similar behavior, especially for values of $u_{C\sigma\sigma}$ larger than 0.6. This argues strongly for the inclusion of total energies in the training set. We believe that the convex behavior of fits 1a and 1b (full and dashed curves) is unphysical, and that therefore heats of formation alone do not suffice to determine like-spin correlation functionals.

From Table III it can also be inferred that the values of the exact-exchange parameter c_{exact} are quite stable (about 0.19–0.24), i.e. well-determined by the present scheme. Inclusion of total energies tends to lower them somewhat.

Tables IV–VIII, and the bar graph Fig. 5, display deviations from experiment for fit 2c (i.e. inclusion of all available data in the fit). It is seen by inspection of Fig. 5 that the distribution of errors in the HoFs is clearly centered around zero, indicating that the mean error arises from a small number of large deviations, compensated by many cases of rather low error. This accumulation around a low mean value is almost as pronounced as in the corresponding G2 graph of Ref. 11. Of course, this is partly due to the least-squares procedure employed, which tends to create a Gaussian distribution of errors around a mean. It is further evidence that our analytical model is appropriate for the reproduction of the data.

As may be seen in Table IV, the largest errors are observed for molecules containing several fluorine atoms (such as AlF_3 and SiF_4), and for ozone. Problems with the former were noted previously by Curtiss *et al.*¹¹ for the G2 scheme. For the most serious cases, their deviations have the same sign as ours, but are usually somewhat smaller. They attempted to remedy the situation by including spin-orbit corrections, and found that while chlorine-containing molecules are improved considerably, fluorine-containing ones are practically unaffected. They concluded that “apparently, there is some inherent problem in G2 theory with some of the fluorine molecules other than the neglect of the spin-orbit effect.” Since we observe similar problems for many of the fluorine molecules, this problem might be not inherent in G2 theory, but of a more general nature. Note that the deviations are not systematic; while the heat of formation of COF_2 is overestimated by 6 kcal/mol, it is more than 7 kcal/mol too low for the similar molecule C_2F_4 . Furthermore, we do not encounter the problems with chlorine-containing systems that were observed in Ref. 11 and partly remedied by inclusion of spin-orbit corrections.

The most problematic ionization potential (Table V) is for the oxygen molecule, whose IP is strongly overestimated by 0.66 eV. We believe that better results may be obtained if different geometries for O_2 and O_2^+ are employed. The MP2-optimized geometries used in this work are in particularly

TABLE V. Ionization potentials of 42 systems (Ref. 13) in eV. The column labeled “present work” gives the values obtained with the parameters derived from a fit to the full set of 233 thermochemical data (fit 2c in Tables II and III).

	System	Present work	Expt.	Dev.
1	$\text{H}\rightarrow\text{H}^+$	13.67	13.60	0.07
2	$\text{He}\rightarrow\text{He}^+$	24.73	24.59	0.14
3	$\text{Li}\rightarrow\text{Li}^+$	5.47	5.39	0.08
4	$\text{Be}\rightarrow\text{Be}^+$	9.11	9.32	-0.21
5	$\text{B}\rightarrow\text{B}^+$	8.56	8.30	0.26
6	$\text{C}\rightarrow\text{C}^+$	11.44	11.26	0.18
7	$\text{N}\rightarrow\text{N}^+$	14.64	14.54	0.10
8	$\text{O}\rightarrow\text{O}^+$	13.87	13.61	0.26
9	$\text{F}\rightarrow\text{F}^+$	17.54	17.42	0.12
10	$\text{Ne}\rightarrow\text{Ne}^+$	21.60	21.56	0.04
11	$\text{CH}_4\rightarrow\text{CH}_4^+$	12.52	12.62	-0.10
12	$\text{NH}_3\rightarrow\text{NH}_3^+$	10.11	10.18	-0.07
13	$\text{OH}\rightarrow\text{OH}^+$	13.07	13.01	0.06
14	$\text{H}_2\text{O}\rightarrow\text{H}_2\text{O}^+$	12.54	12.62	-0.08
15	$\text{HF}\rightarrow\text{HF}^+$	15.99	16.04	-0.05
16	$\text{Na}\rightarrow\text{Na}^+$	5.11	5.14	-0.03
17	$\text{Mg}\rightarrow\text{Mg}^+$	7.79	7.65	0.14
18	$\text{Al}\rightarrow\text{Al}^+$	5.97	5.98	-0.01
19	$\text{Si}\rightarrow\text{Si}^+$	8.11	8.15	-0.04
20	$\text{P}\rightarrow\text{P}^+$	10.42	10.49	-0.07
21	$\text{S}\rightarrow\text{S}^+$	10.45	10.36	0.09
22	$\text{Cl}\rightarrow\text{Cl}^+$	13.03	12.97	0.06
23	$\text{Ar}\rightarrow\text{Ar}^+$	15.79	15.76	0.03
24	$\text{SiH}_4\rightarrow\text{SiH}_4^+$	10.91	11.00	-0.09
25	$\text{PH}\rightarrow\text{PH}^+$	10.17	10.15	0.02
26	$\text{PH}_2\rightarrow\text{PH}_2^+$	9.90	9.82	0.08
27	$\text{PH}_3\rightarrow\text{PH}_3^+$	9.81	9.87	-0.06
28	$\text{SH}\rightarrow\text{SH}^+$	10.43	10.37	0.06
29	$\text{H}_2\text{S}\rightarrow\text{H}_2\text{S}^+ (^2B_1)$	10.42	10.47	-0.05
30	$\text{H}_2\text{S}\rightarrow\text{H}_2\text{S}^+ (^2A_1)$	12.63	12.78	-0.15
31	$\text{HCl}\rightarrow\text{HCl}^+$	12.73	12.75	-0.02
32	$\text{C}_2\text{H}_2\rightarrow\text{C}_2\text{H}_2^+$	11.18	11.40	-0.22
33	$\text{C}_2\text{H}_4\rightarrow\text{C}_2\text{H}_4^+$	10.30	10.51	-0.21
34	$\text{CO}\rightarrow\text{CO}^+$	14.10	14.01	0.09
35	$\text{N}_2\rightarrow\text{N}_2^+ (^2\Sigma_g^-)$	15.83	15.58	0.25
36	$\text{N}_2\rightarrow\text{N}_2^+ (^2\Pi_u)$	16.53	16.70	-0.17
37	$\text{O}_2\rightarrow\text{O}_2^+$	12.73	12.07	0.66
38	$\text{P}_2\rightarrow\text{P}_2^+$	10.34	10.53	-0.19
39	$\text{S}_2\rightarrow\text{S}_2^+$	9.58	9.36	0.22
40	$\text{Cl}_2\rightarrow\text{Cl}_2^+$	11.36	11.50	-0.14
41	$\text{ClF}\rightarrow\text{ClF}^+$	12.55	12.66	-0.11
42	$\text{CS}\rightarrow\text{CS}^+$	11.42	11.33	0.09

poor agreement with experiment for this system. Our result for the isovalent molecule S_2 is much better. The most problematic cases for the electron affinities (Table VI) are the CH_3 and NH_2 radicals, and the chlorine molecule. The radicals are isoelectronic, and the other members of the series (OH and F) show increasingly better quality. This might be a systematic trend, and therefore indicative of a problem with the model.

Proton affinities (Table VII) and total energies (Table VIII) are both rather well reproduced by this fit. The PA for acetylene is an exception, as previously observed.^{4,6}

III. SUMMARY, CONCLUSIONS AND OUTLOOK

In this paper, we have demonstrated that a recently proposed method for the systematic optimization of energy den-

TABLE VI. Electron affinities of 25 systems (Ref. 13) in eV. The column labeled “present work” gives the values obtained with the parameters derived from a fit to the full set of 233 thermochemical data (fit 2c in Tables II and III).

	System	Present work	Expt.	Dev.
1	C←C ⁻	1.26	1.26	0.00
2	CH←CH ⁻	1.22	1.24	-0.02
3	³ CH ₂ ←CH ₂ ⁻	0.61	0.65	-0.04
4	CH ₃ ←CH ₃ ⁻	-0.26	0.08	-0.34
5	NH←NH ⁻	0.22	0.38	-0.16
6	NH ₂ ←NH ₂ ⁻	0.50	0.74	-0.24
7	O←O ⁻	1.48	1.46	0.02
8	OH←OH ⁻	1.70	1.83	-0.13
9	F←F ⁻	3.37	3.40	-0.03
10	O ₂ ←O ₂ ⁻	0.38	0.44	-0.06
11	NO←NO ⁻	0.16	0.02	0.14
12	CN←CN ⁻	3.98	3.82	0.16
13	Si←Si ⁻	1.30	1.38	-0.09
14	P←P ⁻	0.82	0.75	0.07
15	S←S ⁻	2.14	2.08	0.06
16	Cl←Cl ⁻	3.64	3.62	0.02
17	SiH←SiH ⁻	1.19	1.28	-0.09
18	¹ SiH ₂ ←SiH ₂ ⁻	1.09	1.12	-0.04
19	SiH ₃ ←SiH ₃ ⁻	1.37	1.44	-0.07
20	PH←PH ⁻	1.01	1.00	0.01
21	PH ₂ ←PH ₂ ⁻	1.19	1.26	-0.07
22	SH←SH ⁻	2.30	2.31	-0.01
23	PO←PO ⁻	1.20	1.09	0.11
24	S ₂ ←S ₂ ⁻	1.64	1.66	-0.02
25	Cl ₂ ←Cl ₂ ⁻	2.63	2.39	0.24

sity functionals at the GGA level, may be applied with success to a wide variety of thermochemical data. Functionals were optimized using ionization potentials, electron and proton affinities, total energies, and a large set of recently published heats of formation.

We have arrived at some basic conclusions:

- The optimum expansion order in Eq. (6) is $m=2$, as determined by the goodness-of-fit indices given in Table I.
- Within the present framework, the best reachable reproduction of heats of formation is 1.78 kcal/mol average error (as compared with 1.58 kcal/mol in G2 theory itself). Inclusion of all other types of data, including

TABLE VII. Proton affinities of 8 systems (Ref. 13) in kcal/mol. The column labeled “present work” gives the values obtained with the parameters derived from a fit to the full set of 233 thermochemical data (fit 2c in Tables II and III).

	System	Present work	Expt.	Dev.
1	H ₂	99.7	100.8	-1.1
2	NH ₃	204.8	202.5	2.3
3	H ₂ O	165.5	165.1	0.4
4	C ₂ H ₂	156.6	152.3	4.3
5	SiH ₄	153.9	154.0	-0.1
6	PH ₃	187.1	187.1	-0.0
7	H ₂ S	168.6	168.8	-0.2
8	HCl	134.0	133.6	0.4

TABLE VIII. Total energies of 10 atoms in hartree. The column labeled “present work” gives the values obtained with the parameters derived from a fit to the full set of 233 thermochemical data (fit 2c in Tables II and III).

	System	Present work	Expt.	Dev.
1	H	-0.502	-0.500	-0.002
2	He	-2.909	-2.904	-0.005
3	Li	-7.485	-7.478	-0.007
4	Be	-14.664	-14.667	0.003
5	B	-24.649	-24.654	0.005
6	C	-37.840	-37.845	0.005
7	N	-54.583	-54.589	0.006
8	O	-75.068	-75.067	-0.001
9	F	-99.738	-99.734	-0.004
10	Ne	-128.940	-128.938	-0.002

total energies, only mildly increases this value to 1.89 kcal/mol. This implies that a satisfactory stability of the fit has been reached.

- The most problematic cases of HoFs occur for fluorine-containing molecules, which is similar to G2 theory. The problems with chlorine compounds occurring in the latter are not observed here.
- Properties other than heats of formation (such as ionization potentials), need to be included in the fit set to obtain a good overall DFT. This is even more strongly indicated since there is no great deterioration of the HoF reproduction.
- The inclusion of total energies in the fit has a generally beneficial effect. Other properties are only slightly affected, whereas the total energies themselves improve dramatically. It also seems that total energies are required to stabilize the determination of the like-spin correlation part.
- The determination of exchange-correlation parameters directly from experimental thermochemical data can lead to functionals competitive with the G2 scheme. G2

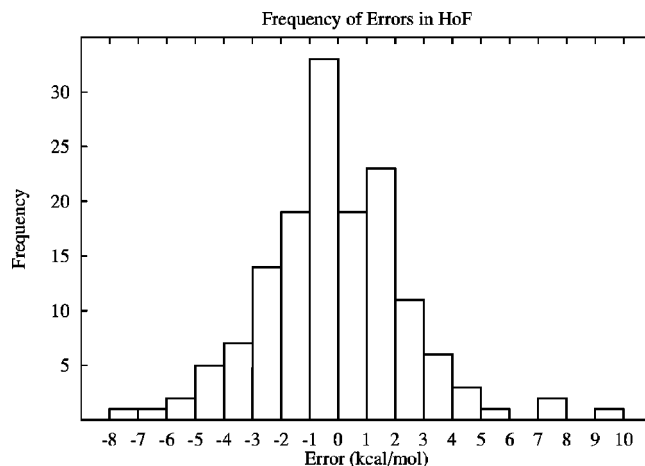


FIG. 5. Bar graph of residual deviations of heats of formation at 298 K. The underlying gradient-corrected functional is the one derived from a 10-parameter fit to the full set of 233 thermochemical data (see Table II, 2c). The graph shows the frequency with which an error $\epsilon = \text{HoF}_{\text{fit}} - \text{HoF}_{\text{exp}}$ occurs in a given interval. The bars are placed in the center of the interval they denote, i.e. the bar with the frequency of an error $-2 \text{ kcal/mol} < \epsilon \leq -1 \text{ kcal/mol}$ is centered at -1.5 .

theory is more reliable generally (due to its pure *ab initio* approach) but at considerably higher cost.

At this point, a few remarks about the scope and limitations of the fits presented here are in order. Three points should be kept in mind when interpreting or using them:

- (1) We used MP2 optimized geometries and scaled Hartree-Fock zero-point corrections. This was done to ensure maximum compatibility with the G2 data. We do not expect the statistics of the fits to change much if other schemes are used *uniformly*, in which case changes will be systematic, and the fit should compensate.
- (2) Our procedure is “post-LSDA.” Since LSDA yields fair densities, it was preferred to the more expensive gradient-corrected functionals. Self-consistency would require cycles of refitting to *all* data, an expense that is hardly necessary given the low deviations resulting in the present simple scheme.
- (3) It must be taken into consideration that the present results are obtained on a numeric grid i.e. at the *basis-set limit*. We expect that results will differ if parameters are refit in moderate basis sets, although the overall statistics are likely to be similar. Usage of the present parameter values is justified only for very large basis sets; we are presently working on reparametrizations for standard bases.

Analogous investigations are presently under way in the context of beyond-GGA DFT (i.e. higher-order gradients). Preliminary results show even smaller errors in fitting the G2 data.

ACKNOWLEDGMENTS

This work is supported by The Natural Sciences and Engineering Research Council of Canada (NSERC). We also

thank P. M. W. Gill and L. A. Curtiss for supplying MP2 geometries and scaled Hartree-Fock vibrational and finite-temperature corrections.

- ¹A. C. Schreiner, J. Baker, and J. W. Andzelm, *J. Comput. Chem.* **18**, 775 (1997).
- ²A. D. Becke, *J. Chem. Phys.* **96**, 2115 (1992).
- ³A. D. Becke, *J. Chem. Phys.* **97**, 9173 (1992).
- ⁴A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- ⁵A. D. Becke, *J. Chem. Phys.* **104**, 1040 (1996).
- ⁶A. D. Becke, *J. Chem. Phys.* **107**, 8554 (1997).
- ⁷A. D. Becke and M. R. Roussel, *Phys. Rev. A* **39**, 3761 (1989).
- ⁸E. I. Proynov, A. Vela, and D. R. Salahub, *Chem. Phys. Lett.* **230**, 419 (1995); **234**, 462(E) (1995).
- ⁹D. J. Tozer, V. E. Ingamells, and N. C. Handy, *J. Chem. Phys.* **105**, 9200 (1996), and references therein.
- ¹⁰A. D. Becke, *J. Chem. Phys.* **84**, 4524 (1986).
- ¹¹L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *J. Chem. Phys.* **106**, 1063 (1997).
- ¹²L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, *J. Chem. Phys.* **94**, 7221 (1991).
- ¹³P. M. W. Gill, B. G. Johnson, J. A. Pople, and M. J. Frisch, *Int. J. Quantum Chem., Symp.* **26**, 319 (1992).
- ¹⁴Unlike in previous work, we included electron affinities in the fits, although negative ions are (in principle) not stable within the LSDA. We deemed this justified, since in our actual calculations, the anions are described adequately, and a check showed that our fits do not change significantly on omission of EAs.
- ¹⁵A. D. Becke, *J. Chem. Phys.* **88**, 2547 (1988).
- ¹⁶A. D. Becke and R. M. Dickson, *J. Chem. Phys.* **89**, 2993 (1988).
- ¹⁷A. D. Becke, *Int. J. Quantum Chem., Symp.* **23**, 599 (1989).
- ¹⁸A. D. Becke and R. M. Dickson, *J. Chem. Phys.* **92**, 3610 (1990).
- ¹⁹E. Prince and C. H. Spiegelman, in *International Tables for Crystallography*, edited by A. J. C. Wilson (Kluwer Academic, Dordrecht, 1992), Vol. C, Chap. 8.4, p. 618.
- ²⁰We have included the H₂ molecule in the original set. This is consistent with previous work by us and other authors.
- ²¹L. A. Curtiss, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.* **98**, 1293 (1993).
- ²²B. J. Smith and L. Radom, *J. Chem. Phys.* **99**, 6468 (1995).
- ²³L. A. Curtiss, P. Redfern, B. J. Smith, and L. Radom, *J. Chem. Phys.* **104**, 5148 (1996).