Theor Chem Acc (2001) 106: 259–263
DOI 10.1007/s002140100274

Theoretical
Chemistry Accounts

Regular article

A proposed modification of CBS-4M model chemistry for application to molecules of increasing molecular size

Rois Benassi

Chemistry Department, University of Modena and Reggio E., Via Campi 183, 41100 Modena, Italy
e-mail: benassi.rois@unimo.it

Received: 7 February 2001 / Accepted: 5 April 2001 / Published online: 13 June 2001

Abstract. In order to calculate more accurately the enthalpies of formation, $\Delta H_f^\circ(298\ \mathrm{K})$, for large molecules using the CBS-4M method, a new formulation of the empirical higher-level correction to the energy is proposed: $\Delta E = a|S|_{ii}^2 I_{ii} + b(n_\alpha + n_\beta) + c\Delta\langle S^2\rangle + \Sigma n_i d_i$. The new methodology (CBS-4MB) applied to a set of 114 molecules of different size significantly decreases the mean absolute deviation from 3.78 to 2.06 kcal/mol.

Key words: Model chemistry – CBS-4M theory – Thermochemistry – Sulfur compounds – Ab initio molecular orbital methods

1 Introduction

Theoretical model chemistry is extensively used for estimating the structure, energy and thermochemical properties of molecules. The most extensively used are the G2 [1] and G3 [2, 3, 4, 5] methods developed by Pople and coworkers and the complete basis set (CBS) methods of Petersson and coworkers [6, 7, 8, 9]. In the calculation of molecular energies for small and medium-sized molecules chemical accuracy amounts to 1–2 kcal/mol or better with the G2 and G3 methods, but for molecules of larger size (more than six heavy atoms) these methodologies demand prohibitively high computational resources. Of the CBS models, CBS-4 is by far the fastest and most readily applicable to larger molecules, at the cost of a loss in accuracy relative to the other models. However, in molecules of increasing complexity, the CBS-4 method leads to a systematic increase in error parameters, and we recently proposed [10] a modification to the method that works with reasonable accuracy on both small medium-sized and large molecules. The increase in the error parameters on the calculated thermochemical properties seems to depend on the accumulation of errors [11] (the greater the size of the molecule, the greater the accumulation of errors), and a corrective term for the single atom was introduced [10] in the calculation of the empirical correction to the CBS energy.

A new version of the CBS-4 scheme, CBS-4M, has recently become available [12]. In this version, a new procedure for the localization of orbitals has been introduced in order to correct problems deriving from the use of extended basis set wave functions. This modification improves the results with respect to the CBS-4 method when applied to the molecules of the test set [6, 7, 8, 9, 11, 12].

Comparison between the results calculated for molecules of increasing size at the CBS-4 and CBS-4M levels and at the level proposed by us [10] shows that improvements are achieved with respect to the original CBS-4 approach. CBS-4M corrects several anomalous results of CBS-4 but does not eliminate completely the problem of accumulation of errors. On the other hand, when the corrective scheme proposed by us [10] is applied to CBS-4, the defects removed by CBS-4M are present.

In order to improve the degree of reliability in the calculation of thermochemical properties of large molecules, the corrections proposed previously [10] were introduced in the CBS-4M model. This step required a recalibration of the empirical contribution to the CBS energy. For this purpose, 114 molecules of different size, with or without conjugative pathways, were tested. Several of them are common to the test set previously employed [6, 7, 8, 9, 11, 12] for the calibration of the different schemes of model chemistry, and various new compounds with known thermochemical properties were also included.

2 Method

All the calculations were performed with the CBS-4M method included in the Gaussian 98 (revision A.7) [13] program. The empirical correction to CBS energy has the form

$$\Delta E = a|S|_{ii}^2 I_{ii} + b(n_\alpha + n_\beta) + c\Delta\langle S^2\rangle + \Sigma n_i d_i, \qquad (1)$$

where $|S|_{\mathrm{ii}} = \int|{}^\alpha\phi_{\mathbf{i}}{}^\beta\phi_{\mathrm{i}}|\mathrm{d}\tau$ is the absolute overlap integral between the most similar α- and β-spin orbitals and $I_{\mathrm{ii}} = (\Sigma c\mu)^2$ is the intraorbital interference factor. $\Delta\langle S^2\rangle$, the error in the spin-squared operator, $\langle S^2\rangle$ takes into account spin contamination corrections. In the last term *i* identifies the atom type (H, C, O, S, N, Cl), *n* is the total number of atoms of this type, and d_i is a constant characterizing atom *i*. The summation spans over the different atoms composing the molecule. The empirical constants *a*, *b*, *c*, and d_i were obtained by fitting experimental and calculated enthalpies of formation, $\Delta H_f^\circ(298\ \mathrm{K})$, according to the procedure previously employed [10]. The new values for the empirical constants are reported in Table 1. The new parameterized version of CBS-4M will be referred to as CBS-4MB.

3 Results and discussion

Deviations of CBS-4M and CBS-4MB enthalpies of formation from experimental values are listed in Table 2. The parameter errors, also reported in Table 2, highlight the good performance of the CBS-4MB scheme. The mean absolute deviation (MAD) of 3.26 kcal/mol obtained with the CBS-4M model [12] applied to the original test set rises to 3.78 for the set of molecules employed here. For the same set of molecules, the MAD decreases to 2.06 in the CBS-4MB scheme. The MAD obtained previously [10] for the reparameterized CBS-4 was 2.52 kcal/mol for a set of 42 molecules. For the same set of molecules, the MAD from CBS-4M is 5.12 (6.96 kcal/mol for original version of CBS-4 [10]), while in CBS-4MB it is 1.68 kcal/mol. These results, reported in Table 3, underline the importance of the correction introduced with the new localization method of the CBS-4M model as well as the importance of the empirical correction given by Eq. (1) for the evaluation of the CBS energy. The latter correction is particularly important in preventing the accumulation of small systematic errors when large molecules are examined [10, 11].

The accumulation of errors (deviations from experimental values) when the CBS-4M is employed, is exemplified in the series of linear alkanes, where the deviation increases from methane (1.13 kcal/mol) to butane (5.19 kcal/mol), with a nearly constant contribution from the carbon atoms of the molecules. This is also true of 2-methylpropane, which exhibits a value of 5.28, similar to that of the linear butane. The saturated cyclic compounds [12] behave similarly. The CBS-4MB model overcomes this deficiency, the deviations being reduced from 4–5 to 0.5–1.1 kcal/mol. A similar additive trend of errors is found in halomethanes, particularly in chlorine derivatives: the CBS-4MB scheme eliminates this defect, although for the tetrahalo derivatives the deviations remain considerable. The same occurs for alkene derivatives, although the deviation for ethylene increases slightly with respect to the CBS-4M result. For acetylene and butadiene the deviation exceeds the mean value but this occurs, though to a lesser extent, in the CBS-4M approach as well.

Table 1. Empirical constants entering in the CBS-4MB approach

Constant	Value
a	−0.002332
c	−0.041827
d_H	−0.001018
d_C	0.000321
d_S	−0.000798
d_N	−0.007745
d_O	−0.005112
d_{Cl}	0.000441
d_F	−0.003360

The introduction of the new correction in the empirical energy, recommended here, markedly improves the results for larger molecules. The importance of the corrections introduced in the CBS-4MB scheme is also emphasized by the parameter values derived from linear regression analysis:

$$\Delta H_f^\circ(298\ \mathrm{K})_{\mathrm{expt}} = a + b\Delta H_f^\circ(298\ \mathrm{K})_{\mathrm{calc}}. \qquad (2)$$

The correlation parameter R^2 increases from 0.99 to 1.00 for CBS-4MB, while *b*, dropping from 1.00 to 0.99, is still close to 1. The new expression for the empirical term including atomic correction parameters brings the value of the intercept very close to zero (1.49 for CBS-4M and −0.31 for CBS-4MB). These results would appear to demonstrate that random, not systematic, errors affect the values calculated with the CBS-4MB approach.

4 Conclusions

The results reported here assess the effect of introducing a new formulation (Eq. 1) of the empirical contribution to CBS energy with a view to obtaining reliable results for molecules of increasing complexity. The reoptimization of parameters *a* and *c* in Eq. (1) and the inclusion of a new set of atomic parameters d_i leads to improvement in the $\Delta H_f^\circ(298\ \mathrm{K})$ values, especially for larger molecules. This approach seems to reduce the effect of accumulation of small errors [11] in the calculation of the enthalpies of formation. The CBS-4M scheme, an improved version of the original CBS-4, failed to resolve the problem of error accumulation, a shortcoming overcome by the CBS-4MB scheme proposed here. Furthermore, this CBS-4MB scheme corrects the accumulating errors in the heats of formation for large molecules without altering calculated potential-energy surfaces and allows calculations of barrier heights for chemical reactions.

It has been recommended [12] that the choice of a method of model chemistry should be based on the comparison of the results obtained with small prototypes; however, this can be misleading, as is evident when calculated values for small and large molecules of the same family are obtained with CBS-4M. Although energies close to experimental values are obtained using the sophisticated G3 [2, 3, 4, 5] model, it is almost impossible to test the results for large molecules owing to the prohibitive computational requirements. In view of the relatively modest computational requirements of the CBS-4MB scheme, it would seem to be virtually the only feasible choice for studying the thermochemical properties of large molecules (up to 20–30 heavy atoms).

Table 2. Experimental and calculated enthalpies of formation ΔH_f°(298 K) (kcal/mol) and parameter errors

Compounds	Experimental[a]	CBS-4 M	Δ[b]	CBS-4MB	Δ[b]
H_2S	−4.90	−5.80	0.90	−5.07	0.17
H_2S_2	3.80	4.86	−1.06	6.17	−2.37
CH_3SH	−5.50	−7.52	2.02	−5.25	−0.25
CH_3CH_2SH	−11.10	−14.62	3.52	−10.82	−0.28
t-C_4H_9SH	−26.14	−34.25	8.11	−27.42	1.28
C_6H_5SH	26.70	16.11	10.59	25.50	1.20
CH_3SCH_3	−8.90	−12.42	3.52	−8.62	−0.28
Thirane	19.60	13.23	6.37	16.88	2.72
CH_3SSH	−1.13	−1.79	0.66	1.03	−2.16
CH_3SSCH_3	−5.80	−8.27	2.47	−3.94	−1.86
$C_6H_5SSCH_3$	24.10	13.80	10.30	25.22	−1.16
$C_6H_5SSC_6H_5$	58.40	35.46	22.94	53.96	4.44
t-$C_4H_9SS$$t$-$C_4H_9$	−47.10	−61.75	14.65	−48.29	1.19
CH_3SSCl	−5.10	−6.01	0.91	−2.52	−2.58
CH_4	−17.80	−18.93	1.13	−17.22	−0.58
CH_3CH_3	−20.00	−22.65	2.65	−19.37	−0.63
$CH_3CH_2CH_3$	−25.00	−29.00	4.00	−24.16	−0.84
$CH_3CH_2CH_2CH_3$	−30.20	−35.39	5.19	−29.00	−1.20
t-C_4H_{10}	−32.10	−37.38	5.28	−31.00	−1.10
CH_3Cl	−19.60	−21.11	1.51	−18.66	−0.94
CH_2Cl_2	−22.80	−25.40	2.60	−22.30	−0.50
$CHCl_3$	−24.70	−29.10	4.40	−25.40	0.70
CCl_4	−22.90	−30.30	7.40	−25.75	2.85
CH_3F	−56.00	−56.62	0.62	−56.40	0.40
CH_2F_2	−107.70	−107.22	−0.48	−108.51	0.81
CHF_3	−166.60	−165.09	−1.51	−167.88	1.28
CF_4	−223.00	−221.46	−1.54	−225.75	2.75
C_6H_{12} (cyclohexane)	−29.50	−37.64	8.14	−28.23	−1.27
C_6H_6	19.82	11.87	7.95	20.77	−0.95
$C_6H_5CH_3$	11.99	2.29	9.70	12.74	−0.75
C_6H_5Cl	12.39	1.44	10.95	11.00	1.39
Furan	−8.30	−11.22	2.92	−7.23	−1.07
Pyrrole	25.90	19.71	6.19	22.77	3.13
Pyridine	33.60	26.62	6.98	31.06	2.54
Thiofene	27.50	20.20	7.30	26.57	0.93
$CH_2{=}CH_2$	12.50	11.57	0.93	14.68	−2.18
$CH_2{=}CHCH_3$	4.80	2.29	2.51	6.96	−2.16
(Z)-$CH{=}CHCH_2CH_3$	−1.90	−5.23	3.33	1.00	−2.90
(E)-$CH{=}CHCH_2CH_3$	−2.90	−6.59	3.69	−0.36	−2.54
$CH_2{=}CHF$	−33.20	−34.10	0.90	−32.49	−0.71
$CH_2{=}CHCl$	5.00	2.95	2.05	6.75	−1.75
C_2F_4	−160.60	−160.15	−0.45	−163.12	2.52
C_2Cl_4	−3.00	−13.81	10.81	−8.11	5.11
trans C_4H_6	26.30	24.19	2.11	30.25	−3.95
$CH{\equiv}CH$	54.50	57.53	−3.03	60.56	−6.06
HCN	30.80	34.70	−3.90	33.25	−2.45
CH_3CN	18.00	19.14	−1.14	19.29	−1.29
CF_3CN	−118.40	−115.20	−3.20	−119.64	1.24
$CH_2{=}CHCN$	43.20	46.29	−3.09	47.82	−4.62
$HCCCN$	84.00	94.17	−10.17	95.59	−11.59
$NCCN$	73.30	79.71	−6.41	76.63	−3.33
$CH_2(CN)_2$	63.50	64.82	−1.32	63.39	0.11
$C(CN)_2{=}C(CN)_2$	169.00	175.12	−6.12	171.73	−2.73
H_2O	−57.80	−57.62	−0.18	−59.38	1.58
H_2O_2	−32.50	−28.86	−3.64	−32.75	0.25
CH_3OH	−48.20	−48.61	0.41	−48.80	0.60
$(CH3)_3COH$	−74.72	−79.59	4.87	−75.14	0.42
CH_3OCH_3	−44.00	−44.99	0.99	−43.64	−0.36
$H_2C{=}O$	−26.00	−26.82	0.82	−27.19	1.19
CH_3COCH_3	−51.90	−55.04	3.14	−52.25	0.35
Oxirane	−12.60	−15.01	2.41	−13.85	1.25
$HCOOH$	−90.50	−89.82	−0.68	−92.05	1.55
CH_3COOH	−103.40	−104.00	0.60	−104.64	1.24
CH_3COF	−105.70	−106.24	0.54	−106.50	0.80
CH_3COCl	−58.00	−61.47	3.47	−59.61	1.61
CH_3CONH_2	−57.00	−58.17	1.17	−59.74	2.74

Table 2. (*Contd.*)

Compounds	Experimental[a]	CBS-4 M	Δ[b]	CBS-4MB	Δ[b]
Ketene	−11.40	−11.64	0.24	−10.56	−0.84
NH_3	−11.00	−10.36	−0.64	−13.06	2.06
CH_3NH_2	−5.50	−6.34	0.84	−7.50	2.00
$(CH3)_2NH$	−4.40	−6.52	2.12	−6.13	1.73
$(CH3)_3N$	−5.70	−9.99	4.29	−8.10	2.40
Aziridine	30.20	27.13	3.07	27.35	2.85
NH_2NH_2	22.80	23.59	−0.79	17.96	4.84
HF	−65.10	−65.52	0.42	−66.90	1.80
HCl	−22.10	−22.65	0.55	−21.74	−0.36
CO	−26.40	−25.93	−0.47	−26.44	−0.04
CO_2	−94.10	−92.60	−1.50	−94.89	0.79
SC	66.90	65.63	1.27	67.42	−0.52
SO	1.20	0.51	0.69	−2.08	3.28
ClO	24.20	33.91	−9.71	31.94	−7.74
ClF	−13.20	−8.91	−4.29	−9.70	−3.50
HOCl	−17.80	−15.36	−2.44	−16.52	−1.28
SO_2	−70.90	−66.56	−4.34	−69.99	−0.91
COS	−33.10	−34.19	1.09	−34.17	1.07
CS_2	28.00	24.60	3.40	26.92	1.08
CF_2O	−145.60	−144.01	−1.59	−147.35	1.75
O_3	34.10	46.62	−12.52	39.77	−5.67
ClF_3	−38.00	−24.27	−13.73	−28.39	−9.61
H_2	0.00	−0.54	0.54	−0.43	0.43
F_2	0.00	5.14	−5.14	1.88	−1.88
N_2	0.00	4.42	−4.42	−1.49	1.49
Cl_2	0.00	2.50	−2.50	4.06	−4.06
O_2	0.00	1.29	−1.29	−4.00	4.00
S_2	30.70	29.52	1.18	29.53	1.17
$NH^{\cdot}$	85.20	86.02	−0.82	81.80	3.40
$NH_2^{\cdot}$	45.10	45.22	−0.12	41.76	3.34
$OH^{\cdot}$	9.40	9.51	−0.11	7.00	2.40
$HS^{\cdot}$	34.20	34.08	0.12	34.18	−0.02
$HSS^{\cdot}$	23.60	27.18	−3.58	27.89	−4.29
$CH_3^{\cdot}$	35.00	34.35	0.65	35.40	−0.40
$CH_3S^{\cdot}$	29.80	28.11	1.69	29.77	−0.03
$CH_3SS^{\cdot}$	16.40	17.43	−1.03	19.64	−3.24
t-$C_4H_9^{\cdot}$	12.30	7.88	4.42	13.64	−1.34
t-$C_4H_9S^{\cdot}$	10.10	2.60	7.50	8.83	1.27
t-$C_4H_9SS^{\cdot}$	−4.60	−10.98	6.38	−4.21	−0.39
$C_6H_5^{\cdot}$	77.00	75.85	1.15	81.62	−4.62
$C_6H_5S^{\cdot}$	54.90	44.39	10.51	50.93	3.97
$C_6H_5SS^{\cdot}$	54.70	42.12	12.58	51.43	3.27
$C_6H_5CH_2^{\cdot}$	50.31	39.90	10.41	47.45	2.86
$CH_3CO^{\cdot}$	−2.40	−3.79	1.39	−3.27	0.87
$CH_3O^{\cdot}$	4.10	7.21	−3.11	6.29	−2.19
$(CH3)_3CO^{\cdot}$	−22.80	−22.44	−0.36	−18.70	−4.10
HCO	10.00	10.32	−0.32	9.22	0.78
CN	104.90	111.01	−6.11	107.61	−2.71
Average			1.53		−0.21
Mean absolute deviation			3.78		2.06
Root mean square			5.44		2.77
Maximum			22.94		11.59
Intercept			1.49		−0.31
Slope			1.00		0.99
R^2			0.99		1.00

[a] Refs. [10, 12, 14]

[b] Error (experiment−theory)

Table 3. Mean absolute deviation obtained at different levels for a set of 42 molecules showing the effect of increasing molecular size

Method	Original values	Reparameterized values
CBS-4	6.96	2.52
CBS-4M	5.12	1.68

Acknowledgements. Financial support from the Ministero per l'Università e le Ricerca Scientifica and computing facilities from the Centro di Calcolo Interuniversitario dell'Italia Nord Orientale and from the Centro di Calcolo dell'Università di Modena are warmly acknowledged.

References

1. Curtiss LA, Raghavachari K, Trucks GW, Pople JA (1991) J Chem Phys 94: 7221
2. Curtiss LA, Raghavachari K, Redfren PC, Rassolov V, Pople JA (1998) J Chem Phys 109: 7764
3. Curtiss LA, Redfren PC, Raghavachari K, Rassolov V, Pople JA (1998) J Chem Phys 110: 7764
4. Baboul AG, Curtiss LA, Redfren PC, Raghavachari KJ (1999) Chem Phys 110: 7650
5. Curtiss LA, Redfren PC, Raghavachari K, Pople JA (1999) Chem Phys Lett 313: 600
6. Montgomery JA Jr, Ochterski JW, Petersson GA (1994) J Chem Phys 101: 5900
7. Ochterski JW, Petersson GA, Montgomery JA Jr (1996) J Chem Phys 104: 2598
8. Montgomery JA Jr, Frisch MJ, Ochterski JW, Petersson GAJ (1999) Chem Phys 110: 2822
9. Petersson GA, Malick DK, Wilson WG, Ochterski JW. Montgomery JA Jr, Frisch MJ (1998) J Chem Phys 109: 10570
10. Benassi R, Taddei F (2000) J Comput Chem 21: 1405
11. Nicolaides A, Radom L (1996) Mol Phys 88: 759
12. Montgomery JA Jr, Frisch MJ, Ochterski JW, Petersson GA (2000) J Chem Phys 112: 6532
13. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain M, Farkas CO, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala P, Cui YQ, Morokuma KD, Malick K, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin LR, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98, revision A.7. Gaussian, Pittsburgh, Pa
14. Chase MW Jr, Davies CA, Doweney JR Jr, Frurip DJ, Mc Donald RA, Syverd NJ (1985) J Phys Chem Ref Data 14 Suppl 1