

Benchmark quality total atomization energies of small polyatomic molecules

Jan M. L. Martin

Department of Organic Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Peter R. Taylor

San Diego Supercomputer Center and Department of Chemistry and Biochemistry, University of California, San Diego, P.O. Box 85608, San Diego, California 92186-9784

(Received 2 January 1997; accepted 18 March 1997)

Successive coupled-cluster [CCSD(T)] calculations in basis sets of *spdf*, *spdfg*, and *spdfgh* quality, combined with separate Schwartz-type extrapolations $A + B/(l + 1/2)^\alpha$ of the self-consistent field (SCF) and correlation energies, permit the calculations of molecular total atomization energies (TAEs) with a mean absolute error of as low as 0.12 kcal/mol. For the largest molecule treated, C_2H_4 , we find $\Sigma D_0 = 532.0$ kcal/mol, in perfect agreement with experiment. The aug'-cc-pV5Z basis set recovers on average about 99% of the valence correlation contribution to the TAE, and essentially the entire SCF contribution. © 1997 American Institute of Physics.
[S0021-9606(97)02420-3]

I. INTRODUCTION

The importance of accurate thermodynamic properties can hardly be overestimated. As discussed at length in an upcoming book,¹ the available experimental data are insufficient in both quality and quantity for many purposes, and *ab initio* calculations constitute the obvious alternative.

From the *ab initio* point of view, the most fundamental thermodynamic properties are molecular total atomization energies (TAEs or ΣD_e values). While there are several ancillary issues such as core correlation, anharmonicity of the zero-point energy, and the like, the two aspects of the *ab initio* calculation that determine the quality of computed TAEs are the valence *n*-particle space or correlation treatment and the 1-particle basis set.

Since the advent of CCSD(T) method² (coupled cluster with all single and double substitutions³ and a quasiperturbative treatment of connected triple excitations²) the correlation problem appears to be largely solved,⁴ at least for systems where nondynamical correlation is not appreciably important (i.e., the \mathcal{S}_1 diagnostic⁵ is small). This leaves basis set convergence as the principal issue to contend with. Recently, one of us reported⁶ the very successful use of an extrapolation method based on the asymptotic convergence behavior^{7,8} of the interelectron cusp: $A/(l+1/2)^4 + B/(l+1/2)^5 + \dots$, where *l* is the largest angular momentum present in the basis set. Applying this extrapolation to the CCSD(T) computed TAEs using correlation consistent⁹ basis sets of *spdf*, *spdfg*, and *spdfgh* quality, mean absolute errors over a sample of 13 experimentally, very precisely known TAEs of as low as 0.30 kcal/mol could be reached.

Of course the asymptotic *l*-convergence behavior of the self-consistent field (SCF) and correlation energies is quite different, and therefore the use of separate extrapolations might improve results. We will show in the present contribution that this is indeed the case, and that results of benchmark quality (mean absolute error of about 0.1 kcal/mol) can be obtained for the same set of molecules. In addition, we

will demonstrate the same accuracy for the C_2H_4 molecule, using what rank among the largest basis set conventional coupled-cluster calculations reported to date.

II. METHODS

The largest calculations reported in the present paper were carried out using the SEWARD/SWEDEN/TITAN programs¹⁰⁻¹² running on the Cray C90 at San Diego Supercomputer Center; the other ones were run using MOLPRO96¹³ running on DEC Alpha computers at the University of Antwerp (Belgium) and at the Hebrew University of Jerusalem.

The CCSD(T) method was used throughout. All basis sets employed belong to the cc-pVnZ (correlation consistent⁹ polarized valence *n*-tuple zeta) and aug-cc-pVnZ (augmented¹⁴ cc-pVnZ) families of Dunning and co-workers. Following the suggestion of Del Bene¹⁵ and as in the earlier work, no diffuse-function "augmented" basis functions were used on H; this is denoted by aug'-cc-pVnZ. The values of *n* considered are *T* (triple), *Q* (quadruple), and 5 (quintuple); the largest basis set, aug'-cc-pV5Z, has a [7s6p5d4f3g2h/5s4p3d2f1g] contracted size.

All calculations are carried out at the CCSD(T)/cc-pVQZ optimum geometries,¹⁶ except for those on C_2H_4 where an extrapolated¹⁷ CCSD(T)/cc-pV5Z geometry was used for the cc-pV5Z calculations and optimum geometries¹⁷ for the other calculations.

III. RESULTS AND DISCUSSION

A. SCF binding energy

Three possibilities were considered here: a Feller-type 3-point geometric extrapolation,¹⁸ a 3-point Schwartz-type fit of $A + B/(l + 1/2)^\alpha$, or a 2-point extrapolation of $A + B/(l + 1/2)^5$ (a variation on the SCF energy extrapolation suggested by Petersson and co-workers¹⁹).

TABLE I. Computed [CCSD(T)], extrapolated, and observed total atomization energies and auxiliary quantities. All units are kcal/mol except α , which is dimensionless.

	Observed				Extrapolated				Calculated			
	ΣD_0^c	ZPE ^{a,c}	S-O ^{b,c}	ΣD_e^c	Extrap.	ΣD_e^d	SCF	Valence correlation	α	aug'-cc-pV5Z		
										SCF	Val corr	Core corr ^e
HNO	196.85(6)	8.56	0.22	205.64(6)	205.30	205.67	85.44 ^f	119.38	3.89	85.42	118.08	0.48
CO ₂	381.91(6)	7.24	0.43	389.68(6)	389.75	389.75	258.08	129.89	3.91	258.08	128.23	1.78
CO	256.16(12)	3.11	0.31	259.58(12)	259.56	259.56	181.59	77.01	3.69	181.56	75.93	0.96
F ₂	36.94(10)	1.30	0.77	39.01(10)	38.29	38.29	-31.07	69.43	4.26	-31.08	69.14	-0.07
N ₂	225.06(3)	3.36	0	228.42(3)	228.16	228.53	119.71	107.59	3.52	119.66	105.83	0.85
N ₂ O	263.61(10)	6.77	0.22	270.60(10)	269.73	270.23	95.15	173.32	3.93	95.09	171.35	1.26
C ₂ H ₂	388.90(24)	16.46	0.17	405.53(24)	405.04	405.04	299.93	102.67	4.37	299.84	101.62	2.44
CH ₄	392.51(14)	27.6	0.08	420.23(14)	420.18	420.18	331.60	87.33	4.55	331.52	86.70	1.25
H ₂ CO	357.25(16)	16.53	0.31	374.09(16)	374.33	374.33	264.86	108.15	4.13	264.81	107.04	1.32
H ₂ O	219.35(2)	13.25	0.23	232.83(2)	232.83	232.83	160.03	72.41	4.66	160.01	71.87	0.38
H ₂	103.27(0)	6.21	0	109.48(0)	109.48	109.48	83.86	25.62	4.31	83.84	25.49	0.00
HCN	303.10(25)	10.09	0.08	313.27(25)	312.96	313.33	204.42	106.86	3.94	204.37	105.50	1.67
HF	135.33(17)	5.85	0.39	141.57(17)	141.54	141.54	100.04	41.32	5.38	100.03	41.12	0.18
NH ₃	276.73(10)	21.33	0	298.06(10)	297.77	298.15	203.31	93.79	4.44	203.26	92.96	0.66
Mean abs. err. w/o F ₂ ^g					0.23	0.12						
Mean abs. err. w/o F ₂ ^g					0.88	0.49						
C ₂ H ₄	531.91(29) ^k	31.60 ⁱ	0.17	563.68(29)	563.77	563.77	435.11	126.30	4.25	434.94 ^j	125.00 ^j	2.36 ^k

^aAnharmonic zero-point energy.

^bSpin-orbit contribution to the atomic energy.

^cTaken from the compilation in Ref. 26 unless indicated otherwise.

^dUsing the additional correction $\Sigma_i(\text{bond order})_i \times 0.126$ kcal/mol, where i runs over all bonds with at least one N atom.

^eTaken from Ref. 27 unless indicated otherwise.

^fReference 6.

^gIncluding F₂: mean {0.26,0.16}, maximum {0.88,0.72} kcal/mol.

^hReference 28.

ⁱReference 29.

^jUsing cc-pV5Z basis set (see text).

^kReference 17.

Since the SCF TAE is essentially converged with the aug'-cc-pV5Z basis set, all methods yield almost identical results, the geometric extrapolation consistently yielding slightly lower results than the two others. We have arbitrarily chosen the two-point $A + B/(l + 1/2)^5$ extrapolation results. These are tabulated in Table I, where the computed SCF/aug'-cc-pV5Z values are also given for comparison. In all cases, the extrapolation increment beyond this basis set is less than 0.1 kcal/mol.

B. Correlation energy

The difference between the experimental nonrelativistic TAE, our SCF limit value, and the core correlation contribution should give us the valence correlation component of TAE to high accuracy. What percentage thereof is recovered by the basis sets used here? Over the sample of all molecules considered here except C₂H₄, CCSD(T)/aug'-cc-pVnZ on average recovers 83.1% for $n=D$, 93% for $n=T$, 97.7% for $n=Q$, and 98.9% for $n=5$. This means that any extrapolation method will have to cover approximately the last percent of the valence correlation part of TAE. (The latter itself constitutes a component of the total ranging from 20.8% in CH₄ over 34.2% in HCN to 64.4% in N₂O, with the 179.8% of F₂, which is unbound at the SCF level, as a pathological outlier.)

We have fitted the expression $A + B/(l + 1/2)^\alpha$ here to the valence correlation components of TAE at the CCSD(T)/

aug'-cc-pVlZ level, for $l = 3, 4$, and 5. The average optimum exponent over the sample is 4.21, with individual values ranging from 3.52 for N₂ over 3.91 for CO₂, and 4.55 for CH₄, to 5.38 for HF. As can be seen in Table I, the extrapolation beyond CCSD(T)/aug'-cc-pV5Z accounts for amounts ranging from 0.13 kcal/mol for H₂ over 0.63 kcal/mol for CH₄ and 1.11 kcal/mol for H₂CO to 1.97 kcal/mol for N₂O.

The quality of these results is best illustrated by comparing the sum of extrapolated SCF and valence correlation components, plus the core correlation term, with the experimental TAEs. This comparison is again given in Table I. The results are found to be in excellent agreement with experiment, with a mean absolute error of 0.26 kcal/mol; if the pathological F₂ molecule is eliminated, this drops to 0.23 kcal/mol. Note that no empirical correction has been applied so far, while a small empirical correction for triple bonds or cumulated π bonds was required in the previous work⁶ to achieve this accuracy. The largest remaining error, 0.88 kcal/mol, is for N₂O.

In order to illustrate typical convergence behavior for the SCF and correlation energies, we consider the basis set increments for a representative example, H₂CO. At the SCF level, enlarging the basis set from aug'-cc-pVDZ to aug'-cc-pVTZ increases TAE by 4.63 kcal/mol; for the next step up, to aug'-cc-pVQZ, the increment drops to 0.96 kcal/mol, while further enhancing the basis set to aug'-cc-pV5Z

adds only 0.06 kcal/mol. The extrapolation beyond aug'-cc-pV5Z accounts for only 0.05 kcal/mol. By contrast, the corresponding increments to the CCSD(T) valence correlation part of TAE are 6.94, 3.35, and 1.14 kcal/mol, with our extrapolation accounting for a further 1.11 kcal/mol.

Interestingly, with the exception of C₂H₂, all remaining significant errors involve nitrogen compounds. Now it has been known for some time (see e.g., Refs. 20–22) that basis set convergence in nitrogen compounds is especially slow. It was previously noted⁶ that the extrapolated value for N₂ increases another 0.3 kcal/mol when the basis sets are taken as cc-pVnZ ($n=Q,5,6$) instead, and the cause was suggested there⁶ to be the triple bond. However, the errors for such molecules as NH₃ suggest that the problem may in fact be the involvement of nitrogen in the bond. If we take the error bars of all the nitrogen-containing molecules in the table, divide them by the sum of the respective bond orders involved, and take the average, we find a suggested correction term of 0.126 kcal/mol per bond order involving at least one N atom. Adding that in to the extrapolated results leads to the numbers in the seventh column of Table I.

With a mean absolute error of 0.12 kcal/mol (F₂ not included), these numbers can be considered of “benchmark quality” by any standard; indeed, the average of the experimental uncertainties over the sample is 0.11 kcal/mol. The largest remaining error (0.49 kcal/mol) is for C₂H₂; critical evaluation of the experimental data²³ suggests that the cause for the difference there may not be entirely due to the remaining error in the calculation. The next two largest errors are 0.37 kcal/mol for N₂O and 0.24 kcal/mol for H₂CO: all remaining ones are below or equal to 0.12 kcal/mol, the calculated value often lying within the experimental error bar.

It is hard to imagine how these results could be further improved. The limiting factor for obtaining such accuracy for other molecules, using this approach, would appear to be the presence of appreciable nondynamical correlation effects, which would compromise the performance of the CCSD(T) electron correlation method. It is noteworthy in this context that N₂O is the only molecule in the series for which the \mathcal{F}_1 diagnostic⁵ reaches 0.02, which is usually taken⁴ as the criterion for important nondynamical correlation. While F₂ has a fairly low $\mathcal{F}_1=0.013$, the $3\sigma_g \rightarrow 3\sigma_u^*$ excitation has a cluster amplitude of about 0.20 in the converged CCSD wave function, which suggests that a more sophisticated electron correlation method than CCSD(T) may be in order for the accuracy being pursued here.

C. Application to C₂H₄

A CCSD(T)/aug'-cc-pV5Z calculation on ethylene, at 474 basis functions, is beyond our computational resources. However, since C₂H₄ is not a polar molecule, the difference between this basis and a CCSD(T)/cc-pV5Z calculation will be negligible, as was found⁶ for CH₄ and C₂H₂. At 402 basis functions with all 12 valence electrons correlated, this ranks among the largest CCSD(T) calculations performed to date using a conventional program (that is, with disk storage of integrals).

Successive CCSD(T)/cc-pVnZ ($n=D,T,Q,5$) calculations yield 78.9%, 93.0%, 97.5%, and 99.0%, respectively, of the valence correlation component of TAE, which is in line with the trends noted above. Carrying out a 3-point extrapolation directly to the CCSD(T)/cc-pV mZ ($m=T,Q,5$) valence correlation energies yields a CCSD(T)/cc-pV ∞ Z valence correlation energy of -0.40245 Hartree (the optimum $\alpha=4.112$), of which the cc-pV nZ basis sets ($n=D,T,Q,5$) recover 78.2, 93.2, 97.6, and 98.9%, respectively. The cc-pVQZ figure corroborates the assertion in a very recent review²⁴ that the cc-pVQZ basis set recovers about 98% of the valence correlation energy of ethylene.

In the present case, a 2-point $A + B/(l + 1/2)^5$ extrapolation to the SCF/cc-pVQZ and SCF/cc-pV5Z TAEs leads to an SCF limit TAE=435.11 kcal/mol, while a 3-point $A + B/(l + 1/2)^\alpha$ extrapolation to the valence correlation component of the CCSD(T)/cc-pVnZ ($n=3,4,5$) TAEs yields a basis set limit of 126.30 kcal/mol, the optimum α being 4.25. Together with the core correlation contribution¹⁷ of 2.36 kcal/mol, this leads to 563.77 kcal/mol or, after subtracting the best anharmonic zero point energy¹⁷ of 31.60 kcal/mol and the spin-orbit contribution of 0.17 kcal/mol, to a calculated $\Sigma D_0=532.00$ kcal/mol, in perfect agreement with the experimental value without spin-orbit contributions of 531.91 ± 0.29 kcal/mol. This is slightly better agreement than that the extrapolated result of Machado and Davidson²⁴ although the difference is only tenths of a kcal/mol.

D. Conclusion

We have shown here that CCSD(T) calculations in very large basis sets, together with separate extrapolation of the SCF and correlation energies, permit the calculation of total atomization energies of a number of small polyatomic molecules to within 0.12 kcal/mol, on average, of experiment. The largest molecule we were able to treat in this manner is C₂H₄.

The aug'-cc-pV5Z basis set recovers essentially the entire SCF binding energy and about 99% of the valence correlation component. The importance of nondynamical correlation appears to be the prime limiting factor for the accuracy at the present stage.

At present, CCSD(T)/aug'-cc-pV5Z calculations are not a realistic option for larger systems. If a mean absolute error of 0.23 kcal/mol is acceptable, CCSD(T)/aug'-cc-pVQZ combined with the 3-parameter Martin correction^{16,25,26} and explicit calculation of the core correlation contribution appears to be the most efficient alternative.²⁶ (As an example, for ethylene we thus obtain $\Sigma D_e=563.57$ and $\Sigma D_0=531.98$ kcal/mol, in excellent agreement with experiment. The 3-parameter correction accounts for 2.82 kcal/mol of these values.) In order to investigate whether the present accuracy could be obtained for larger systems using less computationally demanding correlation methods for the final basis set extrapolation steps, research is currently under way concerning 1-particle basis- n -particle space coupling.

ACKNOWLEDGMENTS

J. M. is an Honorary Research Associate of the NFWO/FNRS (National Science Foundation of Belgium) and a Yigal Allon Fellow. This research was supported by the National Science Foundation (USA) through Cooperative Agreement DASC-8902825 and by Grant No. CHE-9320718 (P.R.T.), and by a grant of computer time from SDSC. This research was further subsidized by the University of Antwerp "Geconcerteerde Actie" program and by the Institute of Chemistry computing center (Hebrew University). The authors wish to thank Professor E. R. Davidson for helpful discussions and a preprint of Ref. 24, and M. Vildibill and V. Hazelwood for assistance with the ethylene CCSD(T) calculations.

¹ *Computational Thermochemistry: Prediction and Estimation of Molecular Thermodynamics*, edited by K. K. Irikura and D. J. Frurip (American Chemical Society, Washington, DC, 1997).

² K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, *Chem. Phys. Lett.* **157**, 479 (1989).

³ G. D. Purvis III and R. J. Bartlett, *J. Chem. Phys.* **76**, 1910 (1982).

⁴ T. J. Lee and G. E. Scuseria, in *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*, edited by S. R. Langhoff (Kluwer, Dordrecht, The Netherlands, 1995).

⁵ T. J. Lee and P. R. Taylor, *Int. J. Quantum Chem. Symp.* **23**, 199 (1989).

⁶ J. M. L. Martin, *Chem. Phys. Lett.* **259**, 669 (1996).

⁷ C. Schwartz, in *Methods in Computational Physics 2*, edited by B. J. Alder (Academic, New York, 1963).

⁸ W. Kutzelnigg, *Theor. Chim. Acta* **68**, 445 (1985), and references therein; R. N. Hill, *J. Chem. Phys.* **83**, 1173 (1985); W. Kutzelnigg and J. D. Morgan III, *J. Chem. Phys.* **96**, 4484 (1992); **97**, 8821 (E) (1992).

⁹ T. H. Dunning, Jr., *J. Chem. Phys.* **90**, 1007 (1989).

¹⁰ R. Lindh, U. Ryu, and B. Liu, *J. Chem. Phys.* **95**, 5889 (1991).

¹¹ MOLECULE/SWEDEN is an electronic structure program system written by J. Almlöf, C. W. Bauschlicher, Jr., M. R. A. Blomberg, D. P. Chong, A. Heiberg, S. R. Langhoff, P.-Å. Malmqvist, A. P. Rendell, B. O. Roos, P. E. M. Siegbahn, and P. R. Taylor.

¹² TITAN is a vectorized coupled-cluster program written by T. J. Lee, A. P. Rendell, and J. E. Rice.

¹³ MOLPRO 96.2 is an *ab initio* MO package by H. J. Werner and P. J. Knowles, with contributions from J. Almlöf, R. D. Amos, M. J. O. Deegan, S. T. Elbert, C. Hampel, R. Lindh, W. Meyer, K. Peterson, R. Pitzer, A. J. Stone, and P. R. Taylor.

¹⁴ R. A. Kendall, T. H. Dunning, and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).

¹⁵ J. E. Del Bene, *J. Phys. Chem.* **97**, 107 (1993).

¹⁶ J. M. L. Martin, *J. Chem. Phys.* **100**, 8186 (1994).

¹⁷ J. M. L. Martin and P. R. Taylor, *Chem. Phys. Lett.* **248**, 336 (1995).

¹⁸ D. Feller, *J. Chem. Phys.* **96**, 6104 (1992).

¹⁹ J. A. Montgomery, Jr., J. W. Ochterski, and G. A. Petersson, *J. Chem. Phys.* **101**, 5900 (1994), and references therein.

²⁰ M. J. Frisch, J. A. Pople, and J. S. Binkley, *J. Chem. Phys.* **80**, 3265 (1984).

²¹ J. Almlöf, B. J. DeLeeuw, P. R. Taylor, C. W. Bauschlicher, Jr., and P. E. M. Siegbahn, *Int. J. Quantum Chem. Symp.* **23**, 345 (1989).

²² C. W. Bauschlicher, Jr. and H. Partridge, *J. Chem. Phys.* **100**, 4329 (1994).

²³ Joel M. Liebman (personal communication).

²⁴ F. B. C. Machado and E. R. Davidson, *J. Mol. Struct. (Theochem)* (to be published).

²⁵ J. M. L. Martin, *J. Chem. Phys.* **97**, 5012 (1992).

²⁶ J. M. L. Martin, *J. Mol. Struct. (Theochem)* (to be published) (WATOC'96 special issue).

²⁷ J. M. L. Martin, *Chem. Phys. Lett.* **232**, 343 (1995).

²⁸ M. W. Chase, Jr., C. A. Davies, J. R. Downey, Jr., D. J. Frurip, R. A. McDonald, and A. N. Syverud, *JANAF Thermochemical Tables, 3rd edition*, *J. Phys. Chem. Ref. Data* **14**, Supplement 1 (1985).

²⁹ J. M. L. Martin, T. J. Lee, P. R. Taylor, and J. P. François, *J. Chem. Phys.* **103**, 2589 (1995).