



7 July 1995

**CHEMICAL
PHYSICS
LETTERS**

Chemical Physics Letters 240 (1995) 533–540

The sensitivity of B3LYP atomization energies to the basis set and a comparison of basis set requirements for CCSD(T) and B3LYP

Charles W. Bauschlicher Jr., Harry Partridge

NASA Ames Research Center, Moffett Field, CA 94035, USA

Received 3 April 1995; in final form 8 May 1995

Abstract

The atomization energies of the 55 G2 molecules are computed using the B3LYP approach with a variety of basis sets. The 6-311+G(3df) basis set is found to yield superior results to those obtained using the augmented-correlation-consistent valence-polarized triple-zeta set. The atomization energy of SO₂ is found to be the most sensitive to basis set and is studied in detail. Including tight d functions is found to be important for obtaining good atomization energies. The results for SO₂ are compared with those obtained using the coupled-cluster singles and doubles approach including a perturbational estimate of the triple excitations.

1. Introduction

It is generally assumed that the basis set requirements for density functional theory (DFT) approaches are less severe than for more traditional *ab initio* approaches of including electron correlation. The more modest basis set requirements for DFT naturally means that larger systems can be treated. However, because of limitations of the functionals, the results of DFT calculations were, in general, not as accurate as the highest levels of theory. Recently Becke [1] proposed a hybrid functional, which was significantly more accurate for the atomization energies of the 55 G2 molecules than earlier functionals. The hybrid functional consisted of several previously proposed [2–4] functionals, where the three coefficients used to weight the terms were determined by minimizing the average absolute deviation for the difference between theory and experiment for 116 atomic and molecular properties. While the DFT atomization energies were still not as accurate as those obtained with the highest

levels of theory, they were sufficiently accurate to suggest that the hybrid functional could be used to study many classes of molecules.

More recently, Stevens et al. [5] proposed a slightly different form for the hybrid functional, which is commonly denoted B3LYP. They showed that this functional yields good results for the calculation of harmonic frequencies. Recently Langhoff [6] used this method to study infrared intensities of polyaromatic hydrocarbons and their positive ions. He found that the computed frequencies were in good agreement with experiment even for basis sets as small as 4-31G [7]. In addition, we have found [8] that the B3LYP approach yields very reasonable C–H bond energies for some hydrocarbons using basis sets as small as 6-31G*.

On the basis of these results, we proposed [9] replacing the calculation of the SCF zero-point energy and MP2 geometry optimization in the G2(MP2) method [10] with the geometry optimization and frequency calculation using the B3LYP approach in

a 6-31G* basis set. We showed that this simplification slightly reduced the average absolute deviation (1.32 kcal/mol) and maximum error (3.7 kcal/mol) for the atomization energies of the 55 G2 molecules. (Replacing the quadratic configuration interaction approach [11] including a perturbational estimate of the triple excitations [12] (denoted QCI(T)) with coupled-cluster singles and doubles approach [13] including a perturbational estimate of the triple excitations [12,14] (denoted CCSD(T)) also reduces the error; the average absolute deviation and maximum error being 1.29 and 3.1 kcal/mol, respectively.) As an aside, we also computed the atomization energies using the 6-31G* B3LYP approach. The average error was 5.18 kcal/mol with a maximum error of 31.5 kcal/mol. Replacing the 6-31G* basis set [7] by the 6-311+G(3df, 2p) set reduced the errors to 2.22 and 8.1 kcal/mol for the average and maximum errors; these values are very similar to the 2.4 and 7.6 kcal/mol for the average and maximum errors reported by Becke [1] using his functional and a numerical basis set.

In this Letter, we report on some additional basis set tests for the 55 G2 atomization energies [15] using the B3LYP functional. The atomization energy of SO₂ is found to be quite sensitive to the basis set and this molecule is studied in more detail at both the B3LYP and CCSD(T) levels.

2. Methods

Unless otherwise noted, we use the B3LYP hybrid functional as implemented in the GAUSSIAN 92/DFT [16] program. To test the effect of g functions on the atomization energy of SO₂, a series of calculations were performed using the program MULLIKEN¹, where a slightly modified version of the B3LYP functional is implemented; in Mulliken the local correla-

¹ MULLIKEN: A computational quantum chemistry program developed by J.E. Rice, H. Horn, B.H. Lengsfeld, A.D. McLean, J.T. Carter, E.S. Replogle, L.A. Barnes, S.A. Maluendes, G.C. Lie, M. Gutowski, W.E. Rudge, Stephan P.A. Sauer, R. Lindh, K. Andersson, T.S. Chevalier, P.-O. Widmark, Djamal Bouzida, G. Pacansky, K. Singh, C.J. Gillan, P. Carnevali, William C. Swope and B. Liu, Almaden Research Center, IBM Research Division, 650 Harry Road, San Jose, CA 95120-6099.

tion functional of Vosko, Wilk and Nusair [4] is replaced by the functional of Perdew and Wang [17].

We optimize the geometry and compute the vibrational frequencies using the B3LYP approach using the 6-31G* basis set. We compute the zero-point energy by scaling the harmonic frequencies by 0.98. This factor was determined by comparison with experiment, as discussed previously [9]. At this geometry, the atomization energies were determined for the augmented correlation consistent polarized valence triple zeta (aug-cc-pVTZ), the aug-cc-pVTZ with a tight d function added to all atoms except H (aug-cc-pVTZ+d), and the 6-311+G(3df, 2p) sets. For comparison, we also give the results where the geometry was optimized and zero-point energy was determined using the 6-311+G(3df, 2p) sets. The scaling factor (0.989) for the zero-point energy for this basis set was determined [9] using the same experimental values as for the 6-31G* basis set.

As shown below, the atomization energy of SO₂ is very sensitive to basis set and is considered in detail. Additional basis sets are used in this study and they are described below. The calculations are performed using computer programs GAUSSIAN 92/DFT [16] and MULLIKEN (see footnote 1). For SO₂, the B3LYP calculations are compared with those obtained using the CCSD(T) approach. Only the valence electrons are correlated in the CCSD(T) calculations. Some of CCSD(T) calculations are performed using GAUSSIAN 92/DFT [16]; for the open-shell calculations, i.e. the atoms, the UCCSD(T) approach is used. To consider larger basis sets, CCSD(T) calculations were performed using MOLPRO 94²; the open-shell calculations were performed using the restricted open-shell CCSD(T) approach [13,14].

3. Results and discussion

The B3LYP results for the atomization energies of the 55 G2 molecules are summarized in Table 1. As noted in Section 1, the 6-31G* basis set yields an average error of 5.18 kcal/mol and a maximum error of

² MOLPRO 94 is a package of ab initio programs written 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, W. Meyer, K. Peterson, R. Pitzer, A.J. Stone and P.R. Taylor.

Table 1

Summary of the B3LYP computed D_0 values, in kcal/mol. The second column is experiment (taken from Ref. [10]) and the remaining columns correspond to the experimental results minus the theoretical results

	Exp.	Geometry+Zpt ^a				
		6-31G*			6-311+G(3df, 2p)	
		6-31G*	aug-cc-pVTZ	aug-cc-pVTZ+d	6-311+G(3df, 2p)	6-311+G(3df, 2p)
BeH	46.9	-8.1	-8.0	-8.0	-8.1	-8.1
CH	79.9	0.6	-1.5	-1.6	-1.6	-1.6
CH ₂ (¹ A ₁)	170.6	4.2	0.2	0.1	0.0	0.0
CH ₂ (³ B ₁)	179.6	0.1	-1.6	-1.7	-1.8	-1.9
CH ₃	289.2	-0.7	-2.1	-2.2	-2.3	-2.3
CH ₃ Cl	371.0	2.1	2.5	2.0	1.3	1.2
CH ₄	392.5	-0.7	-0.2	-0.3	-0.5	-0.7
CH ₃ OH	480.8	6.4	1.5	1.4	0.5	0.4
CH ₃ SH	445.1	5.6	3.3	2.2	2.0	1.8
CN	176.6	3.3	1.0	0.9	0.1	-0.1
CO	256.2	6.7	5.0	4.8	3.3	3.1
CO ₂	381.9	4.1	2.7	2.5	-0.4	-0.6
CS	169.5	9.4	6.7	5.6	4.9	4.8
C ₂ H ₂	388.9	8.4	2.6	2.5	2.1	2.3
C ₂ H ₄	531.9	2.5	0.7	0.5	0.2	0.2
C ₂ H ₆	666.3	-0.4	1.6	1.5	1.2	1.0
ClF	60.3	4.5	2.0	1.2	0.9	0.8
ClO	63.3	4.6	0.5	-0.8	-1.3	-1.8
Cl ₂	57.2	9.5	3.9	2.8	3.0	2.8
F ₂	36.9	-3.8	1.3	1.3	2.2	2.1
HCN	301.8	3.5	-0.7	-0.8	-1.7	-1.9
HCO	270.3	0.4	-1.3	-1.5	-2.8	-3.0
HCl	102.2	6.5	1.8	1.3	1.1	1.1
HF	135.2	12.5	1.7	1.8	1.6	1.7
HOCl	156.3	9.8	2.6	1.9	1.4	1.4
H ₂ CO	357.3	1.7	0.9	0.7	-0.5	-0.7
H ₂ O	219.3	14.1	1.8	1.8	1.4	1.6
H ₂ O ₂	252.3	10.3	1.8	1.8	1.4	1.6
H ₂ S	173.2	7.0	1.7	0.7	0.7	0.6
LiF	137.6	6.6	2.8	2.5	1.9	1.7
LiH	56.0	1.1	-0.5	-0.5	-0.4	-0.4
Li ₂	24.0	4.2	3.7	3.6	3.5	3.5
NH	79.0	-1.0	-4.3	-4.4	-4.5	-4.4
NH ₂	170.0	1.2	-6.0	-6.1	-6.3	-6.2
NH ₃	276.7	7.6	-2.7	-2.7	-3.0	-3.1
NO	150.1	0.4	-1.4	-1.4	-2.7	-2.9
N ₂	225.1	5.7	0.0	-0.1	-0.7	-1.1
N ₂ H ₄	405.4	7.6	-5.0	-5.1	-5.8	-5.8
NaCl	97.5	7.7	5.9	5.4	4.4	4.4
Na ₂	16.6	-0.2	0.0	-0.4	-0.6	-0.5
OH	101.3	4.4	-1.6	-1.6	-1.8	-1.7
O ₂	118.0	-4.4	-2.8	-2.8	-3.6	-3.7
PH ₂	144.7	-1.2	-4.7	-5.6	-6.0	-5.9
PH ₃	227.4	3.0	-0.9	-2.3	-2.9	-2.9
P ₂	116.1	8.8	3.2	1.7	0.6	0.6

^a Indicates the basis set used to optimize the geometry and compute the zero-point energy.

Table 1
Continued

	Exp.	Geometry+Zpt ^a					
		6-31G*				6-311+G(3df, 2p)	
		6-31G*	aug-cc-pVTZ	aug-cc-pVTZ+d	6-311+G(3df, 2p)	6-311+G(3df, 2p)	
SO	123.5	7.8	1.6	-0.7	-0.9	-1.4	
SO ₂	254.0	31.5	18.2	9.7	8.4	7.6	
S ₂	100.7	5.8	0.8	-0.9	-1.2	-1.4	
SiH ₂ (¹ A ₁)	144.4	0.5	-1.2	-2.0	-1.9	-2.0	
SiH ₂ (³ B ₁)	123.4	-0.3	-1.4	-2.1	-2.1	-2.1	
SiH ₃	214.0	1.4	-0.1	-1.2	-1.2	-1.2	
SiH ₄	302.8	1.4	0.0	-1.6	-1.5	-1.4	
SiO	190.5	11.7	7.3	4.9	5.0	4.9	
Si ₂	74.0	3.7	0.5	-0.1	0.1	0.1	
Si ₂ H ₆	500.1	4.6	2.6	-0.2	0.2	0.3	
avg. error		5.18	2.59	2.29	2.20	2.22	
max. error		31.5	18.2	9.7	8.4	8.1	

31.5 kcal/mol for SO₂. Four other systems, HF, H₂O, H₂O₂ and SiO, have errors larger than 10 kcal/mol. Thus, while the B3LYP method in the 6-31G* basis yields good results for some systems, it is clear that before this level of theory is used, it should be calibrated for the systems of interest.

Using the aug-cc-pVTZ basis set [18–20] significantly reduces the error; however the maximum (18.1 kcal/mol, for SO₂) error is far from acceptable. This is significantly worse than the result obtained using the 6-311+G(3df, 2p) basis set, where the maximum error is 8.4 kcal/mol or almost 10 kcal/mol smaller than for the aug-cc-pVTZ basis set. From inspection of the aug-cc-pVTZ and 6-311+G(3df, 2p) basis sets and from the SO₂ tests described below, it is clear that the biggest difference between these two sets is the tight d function in the 6-311+G(3df, 2p) basis set. Adding an even-tempered ($\beta=3.0$) tight d function to the aug-cc-pVTZ set reduces the error, but it is still larger than for the 6-311+G(3df, 2p) basis set. The difference between the last two columns shows that optimizing the geometry and computing the zero-point energy using the 6-311+G(3df, 2p) basis set does not significantly affect the accuracy of the results.

From the results presented in Table 1 it is clear that the atomization energy of SO₂ is the most sensitive to the quality of the basis set of the 55 molecules studied. Thus we study the atomization energy of SO₂ in

more detail – see Table 2. We first consider the results obtained with the GAUSSIAN 92/DFT implementation of the B3LYP functional in the top of the table. Entry (A) is the 6-31G* result from Table 1 and is in error by 31.53 kcal/mol. Adding a compact d function to O and S, entry (B), reduces the error by more than 10 kcal/mol and the atomization energy of this set is slightly better the (2d) set (C). Keeping the (2d) polarization set and improving the valence basis (D) decreases the error slightly, but adding diffuse functions (E) increases the error by 1.37 kcal/mol. That is, the diffuse functions help the atoms more than the molecule, which suggest that the 6-311G basis set does not contain sufficiently diffuse functions. Replacing the (2d) set with the (3d) set (F) and adding an f function (G) reduces the error significantly, about 6 kcal/mol for each improvement. Removing the diffuse function from the 6-311G+(3df) set (H) decreases the error by 2.80 kcal/mol. This is even larger than for the smaller (2d) polarization set; thus the conclusion that the 6-311G basis set is not sufficiently diffuse is true for both the (2d) and (3df) polarization sets. Adding additional diffuse functions to the 6-311+G(3df) basis set ((I) versus (G)) does not significantly (0.39 kcal/mol) affect the result, suggesting that while the 6-311G set is lacking diffuse functions, the 6-311+G basis is a good compromise between size and accuracy. Using the geometry and zero-point en-

ergy from the 6-311G+(3df) basis set instead of the 6-31G* set reduces the error by only 0.82 kcal/mol ((J) versus (G)). Adding an additional set of tight d functions to the 6-311G+(3df) set increase the binding energy by 1.29 kcal/mol ((K) versus (G)). If a set of tight f functions is added as well as the tight d functions (L), the error is reduced by 0.62 kcal/mol more.

The aug-cc-pVTZ set (M) has a sizable error, as noted above. By replacing the O (entry (N)) and S (entry (O)) aug-cc-pVTZ basis sets by the 6-311+G(3df) sets, it is clear that most of the error in the aug-cc-pVTZ set arises from the oxygen basis set. Adding a tight d (P) to the aug-cc-pVTZ set dramatically reduces the error, but it is still 1.30 kcal/mol larger than the 6-311+G(3df) set. If the valence part of the aug-cc-pVTZ set is very flexibly contracted and several sets of d and f functions are added (Q) the error is only 6.84 kcal/mol. If the 6-311G(3df)+tight d and f functions (set L) is corrected with the effect of diffuse functions ((I)–(G)), the error would be 6.89 kcal/mol, which is very similar to that found with set (Q). The error for the cc-pVQZ set [20,22] with the g function deleted (R) is much larger than the 6-311+G(3df) set, indicating that the polarization set for the cc-pVQZ set is still lacking the required tight polarization functions.

The next series of calculations uses the Mulliken implementation of the B3LYP functional. The results for the 6-31G* basis set ((S) versus (A)) shows that the error is 1.28 kcal/mol larger using the Mulliken implementation. The error for the cc-pVTZ set (T) is smaller than for the 6-31G* set, but still very large. Replacing the cc-pVTZ polarization set with the 6-311G(3df) polarization set (U) reduces the error as does adding tight d and f functions (V) to the cc-pVTZ basis set. Uncontracting an additional s and p function (W) in the cc-pVTZ+tight d and f functions basis set reduces the error slightly (0.23 kcal/mol). Adding diffuse functions to the cc-pVTZ set (X) decreases the error by 0.22 kcal/mol ((X) versus (T)). The difference (1.38 kcal/mol) between the Gaussian and Mulliken implementations for the aug-cc-pVTZ set ((X) versus (M)) is very similar to that found for the 6-31G* set. Adding tight d and f functions (Y) to the aug-cc-pVTZ set reduces the error, but the difference between sets (V) and (Y) shows that diffuse functions increase the error by 0.11 kcal/mol when

tight d and f functions are also added to the aug-cc-pVTZ set, this is the opposite effect adding diffuse functions to the cc-pVTZ set; however both effects are relatively small, and much smaller than the difference between the 6-311G and 6-311+G basis sets. Thus, while it is important to add diffuse functions to the 6-311G basis set, it is much less important to add diffuse functions to the cc-pVTZ set. However, for more ionic systems such as NaCl, the diffuse functions are important for the cc-pVTZ set as well. The error for the cc-pVQZ set (Z) is 12.35 kcal/mol, which is reduced to 8.03 kcal/mol when tight d, f, and g functions are added (AA). Deleting the g functions (BB) shows that they contribute 1 kcal/mol to the binding energy. Finally we note that using the average atomic natural orbital basis set of Widmark and co-workers [21,22] (CC) also yields a sizeable error.

The study summarized in Table 2 shows that tight polarization functions are very important for accurately computing the binding energy of SO₂. These functions are in the 6-311+G(3df) basis set. Adding tight polarization functions to the cc-pVTZ basis set yields results of about the same accuracy. Adding the diffuse functions to the 6-311G basis set is found to be important, but somewhat unexpectedly, they lower the atom more than the molecule. Overall the tests suggest that the 6-311+G(3df, 2p) basis sets are a good choice for accurate B3LYP calculations. However, as shown by the SO₂ calculations, this basis set can be several kcal/mol from the B3LYP basis set limit for some cases.

In light of the success of the G2 and G2(MP2) methods one might be tempted to try to develop a higher level correction to improve the B3LYP results. However we note that BeH is one of the systems that has a sizable error in all basis sets and has the same number of α and β electrons in the molecule and atoms. Thus it is clear that a simple higher level correction that is so successful for the G2 and G2(MP2) approaches cannot be applied to the B3LYP approach.

We consider the atomization of SO₂ using the CCSD(T) approach as a function of basis set in Table 3. The error in the 6-31G* basis set (A) is almost 50.0 kcal/mol; this is larger than for the B3LYP in the same basis set. Improving the valence basis (B) increases the error slightly, but adding diffuse functions (C) reduces the error; this is the opposite effect found at the B3LYP level. The error in the 6-311+G(3df)

Table 2

B3LYP basis set calibration study of SO₂. The geometry and zero-point energy are taken from the 6-31G* B3LYP calculation, except for J, where the 6-311+G(3df, 2p) results are used

Basis	D_0	Δ
GAUSSIAN 92/DFT ^a		
(A) 6-31G*	222.47	31.53
(B) 6-31G*+tight d O(α = 2.4) S(α = 1.95)	234.05	19.95
(C) 6-31G(2d)	232.75	21.25
(D) 6-311G(2d)	235.30	18.70
(E) 6-311+G(2d)	233.93	20.07
(F) 6-311+G(3d)	239.81	14.19
(G) 6-311+G(3df)	245.59	8.41
(H) 6-311G(3df)	248.00	6.00
(I) 6-311+G(3df)+diffuse sp O(α =0.0338) S(α =0.0405)	245.20	8.80
(J) 6-311+G(3df) (using 6-311+G(3df) geometry and zero point)	246.41	7.59
(K) 6-311+G(3df)+tight d O(α = 15.0) S(α = 7.8)	246.88	7.12
(L) 6-311+G(3df)+tight df O(α d= 15.0, α f= 4.2) S(α d= 7.8, α f= 1.65)	247.50	6.50
(M) aug-cc-pVTZ	236.44	17.56
(N) aug-cc-pVTZ S 6-311+G(3df) O	245.03	8.97
(O) aug-cc-pVTZ O 6-311+G(3df) S	237.06	16.94
(P) aug-cc-pVTZ+tight d O(α =6.942) S(α =2.457)	244.29	9.71
(Q) O (11s6p5d3f)/[8s5p5d3f] S (16s9p4d3f)/[10s7p4d3f] ^b	247.16	6.84
(R) cc-pVQZ without the g	241.43	12.57
MULLIKEN ^c		
(S) 6-31G*	221.19	32.81
(T) cc-pVTZ	234.84	19.16
(U) cc-pVTZ(sp)+6-311G(3df) polarization set	244.05	9.95
(V) cc-pVTZ+tight df O(α d= 6.942, α f= 4.284) S(α d= 2.457, α f= 1.671)	243.82	10.18
(W) cc-pVTZ+tight df(same exponents as V) + uncontract 1 extra s and p	244.05	9.95
(X) aug-cc-pVTZ	235.06	18.94
(Y) aug-cc-pVTZ+ tight df(same exponents as V)	243.71	10.29
(Z) cc-pVQZ	241.65	12.35
(AA) cc-pVQZ+tight dfg O(α d= 10.193 α f= 7.198 α g= 4.984) S(α d= 3.375, α f= 2.346, α g= 1.844)	245.97	8.03
(BB) cc-pVQZ+tight df (same as AA) without g functions	244.97	9.03
(CC) O (14s9p4d3f)/[5s4p3d1f] S (17s12p5d4f)/[6s5p3d1f] Lund ANOs	239.02	14.98

^a Indicates that the GAUSSIAN 92/DFT implementation of the B3LYP approach is used.

^b The s and p primitive sets are those from the cc-pVTZ basis sets, which are contracted in a segmented manner for the 1s and 2p functions only. The polarization sets are O(α d= 8.0, 3.2, 1.28, 0.51, 0.205, α f= 6.25, 2.50, 1.0) and S(α d= 7.81, 3.125, 1.25, 0.5, α f= 3.125, 1.25, 0.5).

^c Indicates that the Mulliken implementation of the B3LYP approach is used.

basis (D) is much smaller, but still larger than found for the B3LYP approach. The aug-cc-pVTZ basis set (E) has a larger error than that found for the 6-311+G(3df) basis, and the addition of tight d functions (F) reduces the error significantly. In fact the aug-cc-pVTZ+d set has a smaller error than the 6-311+G(3df) basis set; this is different from the B3LYP results where the 6-311+G(3df) set has a smaller error. However, for both methods, these two basis sets yield similar results. The cc-pVQZ set mi-

nus the g functions (G) has a larger error than either the aug-cc-pVTZ+d or 6-311+G(3df) set, but becomes superior with the addition of tight d functions (H).

The next series of calculations uses the RCCSD(T) approach for the atoms. We first note that the results for the aug-cc-pVTZ set ((E) versus (K)) show that this is only a 0.7 kcal/mol effect. The error in the cc-pVDZ set (I) is significantly larger than the 6-31G* or 6-311G** basis sets. The difference between the cc-

Table 3

CCSD(T) basis set calibration study of SO₂. The geometry and zero-point energy are taken from the 6-31G* B3LYP calculation

Basis	D_0	Δ
UCCSD(T)		
(A) 6-31G*	204.1	49.9
(B) 6-311G**	199.1	54.9
(C) 6-311+G**	201.1	52.9
(D) 6-311+G(3df)	239.6	14.4
(E) aug-cc-pVTZ	232.2	21.8
(F) aug-cc-pVTZ+d O($\alpha=6.942$) S($\alpha=2.457$)	240.1	13.9
(G) cc-pVQZ without g function	237.4	16.6
(H) cc-pVQZ+tight d without g O($\alpha=17.637$) S($\alpha=9.609$)	241.9	12.1
RCCSD(T)		
(I) cc-pVDZ	183.9	70.1
(J) cc-pVTZ	228.4	25.6
(K) aug-cc-pVTZ	232.9	21.1
(L) cc-pVQZ	242.2	11.8
(M) cc-pV5Z	250.0	4.0
(O) cc-pV ∞ Z	260.2	-6.2
(P) cc-pVTZ+d O($\alpha=8.302$) S($\alpha=2.494$)	236.2	17.8
(Q) cc-pVQZ+d O($\alpha=10.962$) S($\alpha=3.10$)	246.3	7.3
(R) cc-pV5Z+d O($\alpha=14.982$) S($\alpha=8.009$)	251.2	2.8
(S) cc-pV ∞ Z	255.9	-1.9

pVTZ and aug-cc-pVTZ sets ((J) versus (K)) shows that diffuse functions reduce the error by 4.5 kcal/mol, which is consistent with the results for the 6-311G** set, and different from the B3LYP results. The error is dramatically reduced in the series cc-pVDZ (I), cc-pVTZ (J), cc-pVQZ (L), and cc-pV5Z [23] (M), where the CCSD(T) approach in the cc-pV5Z basis set yields an error smaller than any obtained using the B3LYP approach. We also note that for the cc-pVQZ basis set, the error in the CCSD(T) approach is smaller than for the B3LYP approach. If following Woon [24] we extrapolate to the basis set limit (O), we obtain a binding energy that is 6 kcal/mol too large. We repeat the sequence of basis sets, but with the addition of tight d functions, i.e. the series cc-pVTZ+d (P), cc-pVQZ+d (Q), and cc-pV5Z+d (R), the errors are reduced and the extrapolated value is only 1.9 kcal/mol too large. We suspect that if this series was repeated with additional tight functions, the extrapolated value would be even smaller. This is the first failure of the basis set extrapolation procedure that we have encountered, and shows that if the family of basis sets is deficient in some manner the extrapolation cannot overcome it. While this is not unexpected, it is

surprising that the cc-pV basis sets have such a problem for SO₂. From Table 1, tight polarization functions are also important for SiO, but to a much smaller extent than for SO₂. The second largest difference between the 6-31G* and 6-311+G(3df, 2p) basis sets that we have been able to find is 10 kcal/mol for PF₅. Thus while the problem is the most severe for SO₂ it is probably not unique to SO₂, and may be a problem for many hypervalent compounds.

Overall the SO₂ calculations show that the B3LYP method does perform better than the CCSD(T) approach for the small basis sets, but by expanding the basis set the CCSD(T) results approach experiment, while those for the B3LYP converge to a value that is several kcal/mol too small. It is also interesting that while tight polarization functions are important for both approaches, the addition of diffuse functions is more important for the CCSD(T) than for the B3LYP approach.

While the primary focus of this work is the atomization energies it is of some interest to comment on how the B3LYP approach performs for the geometries. For the 55 G2 molecules, experimental geometries appear to be available for all molecules except SiH₃ and the

3B_1 state of SiH_2 . We take the geometry of the 1A_1 and 3B_1 states of CH_2 from Jensen and Bunker [25], PH_2 from Herzberg [26], the diatomics from Huber and Herzberg [27], and the remainder from the tabulation of Hehre et al. [28]. (Note the results for the inner and outer H atoms in N_2H_4 are reversed in Ref. [28] – see Kohata et al. [29].) The average absolute and maximum error in the 71 bond lengths is 0.013 Å (0.008 Å) and 0.055 Å (0.039 Å) for the 6-31G* (6-311+G(3df, 2p)) basis set, respectively. For the smaller basis set the maximum error occurs for Cl_2 , while for the larger basis set Na_2 has the largest error. The average absolute and maximum error in the 26 bond angles is 0.62° (0.61°) and 1.69° (1.85°) for the 6-31G* (6-311+G(3df, 2p)) basis set, respectively. For the small basis set the maximum error is for the 1A_1 state of CH_2 , while for the larger basis set N_2H_4 has the largest error. Unfortunately there are only two dihedral angles in this set; the maximum error is 0.6° for the 6-31G* basis and 6.6° for the 6-311+G(3df, 2p) basis set. Thus overall the B3LYP approach does a reasonable job on the geometries.

4. Conclusions

The 6-31G* B3LYP calculations are accurate for some systems, but it is not safe to use a basis set of this size without performing calibration calculations. The 6-31G* set does appear to be acceptable for the optimization of the geometry and calculation of the zero-point energy. The 6-311+G(3df, 2p) set yields the most reliable results of the basis sets tested. The aug-cc-pVTZ set is found to perform significantly less well; the calculations show that the aug-cc-pVTZ set is lacking a tight d function which is important for describing some systems. Even after an even-tempered d is added, this set, on the average, does not perform as well as the 6-311+G(3df, 2p) set. Thus we conclude that currently, the 6-311+G(3df, 2p) set is a good compromise between expense and accuracy for the B3LYP approach. The CCSD(T) calculations for SO_2 also show tight polarization functions to be important in describing this system. For small basis sets the B3LYP approach yields atomization energies that are superior to those obtained using the CCSD(T) approach. However, as the basis sets become more complete, the CCSD(T) results are superior to those obtained using the B3LYP method.

References

- [1] A.D. Becke, *J. Chem. Phys.* 98 (1993) 5648.
- [2] A.D. Becke, *Phys. Rev. A* 38 (1988) 3098.
- [3] C. Lee, W. Yang and R.G. Parr, *Phys. Rev. B* 37 (1988) 785.
- [4] S.H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.* 58 (1980) 1200.
- [5] P.J. Stevens, F.J. Devlin, C.F. Chabrowski and M.J. Frisch, *J. Phys. Chem.* 98 (1994) 11623.
- [6] S.R. Langhoff, submitted for publication.
- [7] M.J. Frisch, J.A. Pople and J.S. Binkley, *J. Chem. Phys.* 80 (1984) 3265, and references therein.
- [8] H. Partridge and C.W. Bauschlicher Jr., unpublished.
- [9] C.W. Bauschlicher Jr. and H. Partridge, *J. Chem. Phys.*, in press.
- [10] L.A. Curtiss, K. Raghavachari and J.A. Pople, *J. Chem. Phys.* 98 (1993) 1293.
- [11] J.A. Pople, M. Head-Gordon and K. Raghavachari, *J. Chem. Phys.* 87 (1987) 5968.
- [12] K. Raghavachari, G.W. Trucks, J.A. Pople and M. Head-Gordon, *Chem. Phys. Letters* 157 (1989) 479.
- [13] P.J. Knowles, C. Hampel and H.-J. Werner, *J. Chem. Phys.* 99 (1993) 5219, and references therein.
- [14] J.D. Watts, J. Gauss and R.J. Bartlett, *J. Chem. Phys.* 98 (1993) 8718.
- [15] L.A. Curtiss, K. Raghavachari, G.W. Trucks and J.A. Pople, *J. Chem. Phys.* 94 (1991) 17221.
- [16] M.J. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.W. Wong, J.B. Foresman, M.A. Robb, M. Head-Gordon, E.S. Replogle, R. Gomperts, J.L. Andres, K. Raghavachari, J.S. Binkley, C. Gonzalez, R.L. Martin, D.J. Fox, D.J. DeFrees, J. Baker, J.J.P. Stewart and J.A. Pople, *GAUSSIAN 92/DFT, Revision G.2* (Gaussian, Pittsburgh, 1993).
- [17] J. P. Perdew and Y. Wang, *Phys. Rev. B* 45 (1992) 13244.
- [18] T.H. Dunning, *J. Chem. Phys.* 90 (1989) 1007.
- [19] R.A. Kendall, T.H. Dunning and R.J. Harrison, *J. Chem. Phys.* 96 (1992) 6796.
- [20] D.E. Woon and T.H. Dunning, *J. Chem. Phys.* 98 (1993) 1358.
- [21] P.-O. Widmark, J.B. Persson and B.O. Roos, *Theoret. Chim. Acta* 79 (1991) 419.
- [22] P.-O. Widmark, P.A. Malmqvist and B.O. Roos, *Theoret. Chim. Acta* 77 (1990) 291.
- [23] D.E. Woon, K.A. Peterson and T.H. Dunning, unpublished.
- [24] D.E. Woon, *Chem. Phys. Letters* 204 (1993) 29.
- [25] P. Jensen and P.R. Bunker, *J. Chem. Phys.* 89 (1988) 1327.
- [26] G. Herzberg, *Molecular spectra and molecular structure*, Vol. 3 (Van Nostrand Reinhold, New York, 1966).
- [27] K.P. Huber and G. Herzberg, *Constants of diatomic molecules* (Van Nostrand Reinhold, New York, 1979).
- [28] W.J. Hehre, L. Radom, P. von R. Schleyer and J.A. Pople, *Ab initio molecular orbital theory* (Wiley-Interscience, New York, 1986).
- [29] K. Kohata, T. Fukuyama and K. Kuchitsu, *J. Phys. Chem.* 86 (1982) 602.