ELSEVIER

Journal of Molecular Structure (Theochem) 584 (2002) 1–4

THEO
CHEM

www.elsevier.com/locate/theochem

Viewpoint

Proper basis set for quantum mechanical studies of potential energy surfaces of carbohydrates

G.I. Csonka[1]

Department of Inorganic Chemistry, Budapest University of Technology and Economics, H-1521 Budapest, Hungary

Received 8 January 2002; accepted 26 February 2002

Abstract

In the current paper we address the optimal selection of basis set for carbohydrates. The following basis sets are used with B3LYP method: 6-31G(d), 6-31G(d,p) 5d, 6-31+G(d,p), 6-311+G(d,p), 6-311++G(d,p), 6-311+G(2d,2p), and 6-311++G(2d,2p). The following molecules were used for the comparison: water, conformers of α-L-fucose, and β-D-glucose. The 6-31G(d) or 6-31G(d,p) basis set in combination with B3LYP functional provide unacceptably poor results for carbohydrates. (Although the HF/6-31G(d) results are quite good.) The introduction of the diffuse functions on heavy atoms is necessary for good results if B3LYP functional is used (suggested methods are B3LYP/6-31+G(d,p) or 6-311+G(d,p), the latter is closer to basis set limit). The introduction of the diffuse functions (++) on hydrogen atoms is *not* necessary if B3LYP or other density functional method is used for carbohydrates

Keywords: B3LYP method; Basis sets; Total energy; α-L-fucose conformers; β-D-glucose conformers

Quantum mechanical studies of potential energy surfaces of carbohydrates play a rather important role in the understanding of the role of various factors in stabilizing specific carbohydrate conformers. This is because carbohydrates are rather difficult tests for the considerably less expensive MM methods as they have densely packed highly polar functional groups, and the conformational energies depend on stereoelectronic effects. Our earlier results show that the MM2* method provides good qualitative results for the lowest energy rotamers of monosaccharides, but an energetically compressed conformational space with incorrectly ordered rotamers in the higher energy region [1,2]. Damm et al. [3] showed that using all the 144 carbohydrate conformers, optimized at HF/6-31G(d) level, for the OPLS-AA torsional parameters optimization gave poor results for the low energy structures. Thus only the 44 lowest energy HF/6-31G(d) structures were used to fit the torsional parameters in order to resolve the problem. This supports the observation that because of the inherent problems the MM methods are unable to yield correct energy differences for a larger energy window for carbohydrates. Barrows et al. [4] summarized the performance of the best MM methods for calculating the relative energies of D-glucose and they have found that HF/6-31G(d) method is superior compared to any MM parameterization.

Ab initio or density functional theoretical studies are very demanding of cpu time. Calculating the conformational space or the energy hypersurface of mono-, di-, or trisaccharides [5,6] easily requires months of current supercomputer time, thus selection

---

E-mail address: csonka@web.inc.bme.hu (G.I. Csonka).

[1] http://web.inc.bme.hu.

PII: S0166-1280(02)00096-9

Table 1
Total and relative energy (hartree) of water calculated with B3LYP method supplemented with various basis sets (B3LYP/6-31G(d,p) 5d geometry was used for the calculations)

Basis set	6-31G(d,p) 5d	6-311+G(d,p)	6-311++G(d,p)	6-311+G(2d,2p)	6-311++G(2d,2p)
Total energy	−76.41816	−76.45843	−76.45849	−76.46188	−76.46197
Relative energy		−0.04027	−0.00006	−0.00339	−0.00009

of the appropriate method and basis set is crucial for the *feasibility* and *quality* of a given study.

It was observed that there is a monotonic change in the relative energies of various conformers of monosaccharides (e.g. D-glucose) at HF level of theory as the basis set quality increases from 3-21G to cc-pVQZ [7,8]. The HF/3-21G results erroneously yield that the rotamers of $^1C_4$ β-D-glucose are about 7–8 kcal/mol more stable than the $^4C_1$ rotamers. This result contradicts the experimental observations and it is in error by 14 kcal/mol compared to the results of the most expensive MP2 based composite calculations [7]. The HF/6-31G(d) and cc-pVDZ results provide quite good relative energies that are close to the results of the most expensive large basis set (extrapolated) MP2 calculations. Further increase of the basis set at HF level of theory up to cc-pVTZ or cc-pV^{T}QZ causes an opposite error, i.e. it destabilizes the $^1C_4$ conformers (up to 6–10 kcal/mol). These discrepancies are due to the fact that intramolecular hydrogen bonding energies are overestimated in the HF/3-21G level calculations and they are underestimated in HF/cc-pV^{T}QZ level calculations. The optimal level of theory for monosaccharides is HF/6-31G(d) [7,9].

However, using the 6-31G(d) or 6-31G(d,p) basis sets in MP2, CCSD or density functional theory studies yield poor results [7,10]. This is because MP2, or B3LYP method supplemented with 6-31G(d) basis set overestimates the intramolecular hydrogen bonding energies and that increases the stability of the $^1C_4$ rotamers. For example, the MP2/6-31G(d) calculations erroneously stabilize this rotamer by about 6–7 kcal/mol [7]. Considerably larger basis sets (cc-pVTZ or larger) are required at the MP2 level of theory to approach the HF/6-31G(d) results or the theoretical limit for the relative energies. This extreme high cost limits the feasibility and utility of MP2 or CCSD studies for carbohydrates.

The less expensive DFT methods (B3LYP or B3P) lead to poor results similar to those of the MP2 method if 6-31G(d) basis set is used [8]. Using this inappropriate basis set with B3LYP method (B3LYP/6-31G(d,p) 5d) led to an incorrect result for the relative stability of furanose and pyranose rings [10]. However, addition of the diffuse functions (e.g. 6-31 + G(d) or aug-cc-pVDZ) improved the DFT relative energies considerably for monosaccharides [8]. The diffuse functions on heavy atoms weaken the hydrogen bonding thus effectively compensate the errors of the B3LYP/6-31G(d,p) results. Similar behavior was found for weak interactions with the B3LYP functional [11,12]. An extensive quantum mechanical study the potential energy surfaces of selected D-aldo- and D-ketohexoses (a total of 82 conformers) reiterates that the introduction of a diffuse function into the basis set is the key to better calculations of the potential energy surfaces of carbohydrates [10]. The errors of the results with a smaller basis set (B3LYP/6-31G(d,p) 5d) disappear at the higher level calculation (B3LYP/6-311++G(2d,2p)//B3LYP/6-31G(d,p) 5d).

In the current paper, we address the optimal selection of this latter basis set for carbohydrates. GAUSSIAN 98 [13] program was used for the calculations. Our experience and theoretical knowledge suggest that the use of 6-311++G(2d,2p) basis set is not necessary and considerably smaller basis sets is sufficient. First we address the use of diffuse functions on hydrogen atoms. The results in Table 1 for water shows only minor change occurs in the total energy if diffuse functions on hydrogen atoms applied. This supports the view that the use of '++' basis sets instead '+' basis sets does not provide any improvement and it is a simple waste of computer time. We note that diffuse functions on hydrogen atoms are necessary only for hydride ions. Table 2 shows our results for the conformational space of L-fucose. The most stable rotamers were identified earlier [2,14]. We reoptimized the geometries for the current study with B3LYP/6-31G(d) method, and performed a series of single

Table 2
Total (hartree) and relative (kcal/mol) energies, and relative energy differences (kcal/mol) of the eight most stable rotamers of L-fucose calculated with B3LYP method supplemented with various basis sets. (B3LYP/6-31G(d) geometry was used for the calculations)

	Basis set				
	6-31G(d)	6-31 + G(d)	6-31 + G(d,p)	6-311 + G(d,p)	6-311++G(d,p)
Total energy					
1	−611.94684	−611.98182	−612.01465	−612.16440	−612.16487
Relative energy					
2	0.72	1.58	1.56	1.60	1.61
3	2.54	2.61	2.53	2.44	2.43
4	3.97	3.65	3.52	3.44	3.43
5	3.79	3.81	3.70	3.62	3.61
6	3.62	3.58	3.46	3.38	3.36
7	5.04	4.78	4.64	4.53	4.51
8	2.81	3.35	3.29	3.31	3.31
Difference[a]					
2	−0.89	−0.04	−0.05	−0.01	
3	0.10	0.18	0.10	0.01	
4	0.55	0.23	0.10	0.02	
5	0.18	0.21	0.10	0.02	
6	0.26	0.22	0.10	0.02	
7	0.53	0.27	0.13	0.02	
8	−0.50	0.04	−0.01	0.00	
Average	0.03	0.16	0.07	0.01	
R.m.s.	0.50	0.11	0.07	0.01	

[a] The difference between the relative energies. The 6-311++G(d,p) results are used as reference.

point calculations (SCF = tight) with basis sets of increasing size up to 6-311++G(d,p). Analysis of the results show that the difference between the B3LYP/6-31G(d) and 6-311++G(d,p) results is significant (root-mean-square difference is 0.5 kcal/mol). The results show that adding diffuse functions on the heavy atoms decreases this difference considerably (by about 80%). The difference

Table 3
Total (hartree), and relative (kcal/mol) energies of the four selected conformers of β-D-glucose calculated with various methods and basis sets

Method	Basis set					Composite[a]
	B3LYP				HF/6-31G(d)[a]	
	6-31G(d)[b]	6-31 + G(d)[c]	6-311 + G(d,p)[b]	6-311++G(d,p)[b]		
Total energy						
1	−687.14920	−687.19449	−687.40271	−687.40324	−683.33194	−686.16144
Relative energy						
2, ${}^4C_1$	−0.70	−0.12	0.15	0.17	−0.15	0.27
3, ${}^1C_4$	1.65	6.92	7.71	7.76	6.73	6.41
4, ${}^1C_4$	−0.60	4.93	8.18	8.23	6.76	6.99

[a] Ref. [4,14].
[b] B3LYP/6-31G(d) geometry was used for the calculations.
[c] B3LYP/6-31 + G(d) geometry was used for the calculations.

between the 6-311 + G(d,p) and 6-311++G(d,p) results is negligible (root-mean-square difference is 0.01 kcal/mol well below the error of B3LYP method), thus using the diffuse functions on hydrogen atoms does not change the results, but it increases considerably the computational time (this depend on the size of the molecule and the number of hydrogen atoms concerned—typical time increase is about 25% or more). Table 3 shows the results for two selected $^4C_1$ (1 and 2 in Table 3) and two $^1C_4$ conformers (3 and 4 in Table 3) of β-D-glucose. As it was mentioned, the calculated energy difference between the two chair forms is a rather sensitive measure of hydrogen bonding strength. Analysis of the results clearly show the large error of B3LYP/6-31G(d) results [8,15] that erroneously stabilize the $^1C_4$ rotamers. The difference between the 6-311 + G(d,p) and 6-311++G(d,p) results is again negligible. The HF/6-31G(d) results are closer to composite results than the B3LYP/6-311 + G(d,p) results (in the latter, the intramolecular hydrogen bonding is somewhat too weak).

The main results can be summarized as follows.

1. The 6-31G(d) or 6-31G(d,p) basis set in combination with B3LYP functional provide unacceptably poor results for carbohydrates. However, the HF/6-31G(d) results are good.
2. The introduction of the diffuse functions on heavy atoms is necessary if B3LYP functional is used (suggested methods B3LYP/6-31 + G(d,p) or 6-311 + G(d,p), the latter is closer to basis set limit).
3. The introduction of the diffuse functions on hydrogen atoms is *not* necessary if B3LYP or other density functional method is used for carbohydrates (avoid B3LYP/6-31++G(d,p) method).

Acknowledgement

The financial support of the Hungarian Scientific Research foundation is gratefully acknowledged (OTKA T034764).

References

[1] G.I. Csonka, K. Éliás, I. Kolossváry, C.P. Sosa, I.G. Csizmadia, J. Phys. Chem. A 102 (1998) 1219.
[2] G.I. Csonka, K. Éliás, I.G. Csizmadia, J. Comput. Chem. 18 (1997) 330.
[3] W. Damm, A. Frontera, J. Tirado-Rives, W.L. Jorgensen, J. Comput. Chem. 18 (1997) 1955.
[4] S.E. Barrows, J.W. Strorer, C.J. Cramer, A.D. French, D.G. Truhlar, J. Comput. Chem. 19 (1998) 1111.
[5] G.I. Csonka, C.P. Sosa, I.G. Csizmadia, J. Phys. Chem. A 104 (2000) 3381.
[6] G.I. Csonka, C.P. Sosa, J. Phys. Chem. A 104 (2000) 7113.
[7] S.E. Barrows, F.J. Dulles, C.J. Cramer, A.D. French, D.G. Truhlar, Carbohyd. Res. 276 (1995) 219.
[8] G.I. Csonka, K. Éliás, I.G. Csizmadia, Chem. Phys. Lett. 257 (1996) 49.
[9] G.I. Csonka, I. Kolossváry, P. Császár, K. Éliás, I.G. Csizmadia, J. Mol. Struct. (Theochem) 395–396 (1997) 29.
[10] J.H. Lii, B.Y. Ma, N.L. Allinger, J. Comput. Chem. 20 (1999) 1593.
[11] J.E. Del Bene, W.B. Person, K. Szczepaniak, J. Chem. Phys. 99 (1995) 10705.
[12] J.J. Novoa, C. Sosa, J. Phys. Chem. 99 (1995) 15837.
[13] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, Mennucci, B., Pomelli, C., Adamo, C., S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B.G. Johnson, W. Chen, M.W. Wong, J.L. Andres, M. Head-Gordon, E.S. Replogle, J.A. Pople, GAUSSIAN 98, Gaussian Inc, Pittsburgh, PA, 1998.
[14] http://web.inc.bme.hu/mols/lfucose.
[15] http://web.inc.bme.hu/mols/glucose.