JCTC Journal of Chemical Theory and Computation

Evaluation of Density Functionals and Basis Sets for Carbohydrates

Gábor I. Csonka,*[†] Alfred D. French,[‡] Glenn P. Johnson,[‡] and Carlos A. Stortz[§]

Department of Inorganic and Analytical Chemistry, Budapest University of Technology, Szent Gellért tér 4, Budapest, H-1521 Hungary, Southern Regional Research Center, U.S. Department of Agriculture, 1100 Robert E. Lee Boulevard, New Orleans, Louisiana 70124, and Departamento de Química Orgánica-CIHIDECAR, Universidad de Buenos Aires, Ciudad Universitaria, 1428 Buenos Aires, Argentina

Abstract: Correlated ab initio wave function calculations using MP2/aug-cc-pVTZ model chemistry have been performed for three test sets of gas phase saccharide conformations to provide reference values for their relative energies. The test sets consist of 15 conformers of α- and β-D-allopyranose, 15 of 3,6-anhydro-4-*O*-methyl-D-galactitol, and four of β-D-glucopyranose. For each set, conformational energies varied by about 7 kcal/mol. Results obtained with the Hartree−Fock method, with pure density functional approximations (DFAs) like LSDA, PBEsol, PBE, and TPSS and with hybrid DFAs like B3PW91, B3LYP, PBEh, and M05-2X, were then compared to the reference and local MP2 relative energies. Basis sets included 6−31G*, 6−31G**, 6−31+G*, 6−31+G**, 6−311+G**, 6−311++G**, cc-pVTZ(-*f*), cc-pVTZ, and aug-cc-pVTZ(-*f*). The smallest basis set that gives good DFA relative energies is 6−31+G**, and more converged results can be obtained with 6−311+G**. The optimized geometries obtained from a smaller basis set, 6−31+G*, were useful for subsequent single point energy calculations with larger basis sets. The best agreement with MP2 was shown by M05-2X, but only when using a dense DFT grid. The popular B3LYP functional is not the best for saccharide conformational studies. The B3PW91 functional gives systematically better results, but other hybrid functionals like PBEh or TPSSh are even better. Overall, the nonempirical PBE GGA and TPSS meta-GGA functionals also performed better than B3LYP.

1. Introduction

Nature combines sugars in an almost infinite array of compounds that are each involved in very specific interactions. These interactions do not depend on elemental composition alone but on the shape or conformation of the carbohydrate as well. A major difficulty in understanding such interactions is that the conformational space of carbohydrates is, at least potentially, very large. Even for monosaccharides, variations in ring shape and exocyclic group orientations can combine in many different ways.[1] Combining monosaccharides into oligo- or polysaccharides results in exponentially larger numbers of potential geometries to consider. Thus, computational studies of the conformational space of carbohydrates must be fast enough to evaluate many different structures. At the same time, reliable selection of the important conformers requires highly accurate methods that have a firm theoretical foundation.

For the past 20 years, energies of carbohydrate conformers have mostly been calculated with empirical force fields, but there has been little agreement among the proponents of the various molecular mechanics methods.[2−4] In addition, most empirical methods do not handle chemical reactivity, and

* To whom correspondence should be addressed. E-mail: csonkagi@gmail.com.
† Budapest University of Technology.
‡ U.S. Department of Agriculture.
§ Universidad de Buenos Aires.

subtle dependencies on variable electronic structure are usually lost. Therefore, a number of workers have used electronic structure theory despite its much slower (and far more computationally expensive) arrival at an optimized geometry and minimized energy.

Three major categories of methods are based on electronic structure theory. The first category, semiempirical quantum mechanics, is easily the fastest but should not be used for work that demands accuracy. The MNDO, AM1, and PM3 semiempirical methods fail to provide accurate relative energies for carbohydrates,[5,6] despite attempts to improve their performance.[7] The second category, ab initio quantum mechanics or wave function (WFN) methods, such as Hartree−Fock (HF) theory, shows systematic errors caused by the lack of Coulomb-type electron correlation. For typical problems with carbohydrates, errors from the lack of correlation are often canceled by basis set errors. For example, split-valence 6−31G* and cc-pVDZ basis sets give fairly good relative energies for carbohydrate conformers with HF theory.[8] However, those good results are based on unreliable error compensation. Attempts to improve the results by using the larger cc-pVTZ or cc-pVQZ basis sets, under-stabilize the $^1C_4$ β-D-glucose conformers resulting from HF error.[8,9] Post-HF or correlated methods like MP2, MP3, MP4, and CCSD(T) with noniterative triples explicitly treat Coulomb correlation. These methods might give reliable equilibrium geometries and relative energies by MP2 complete basis set (CBS) extrapolations using aug-cc-pVTZ, aug-cc-pVQZ, or aug-cc-pV5Z basis sets (noted as MP2/CBS[4,5]). It was found that relative energies of molecules with localized bonds can be reasonably approximated (typical accuracy is around 0.2 kcal/mol) by the MP2/aug-cc-pVTZ model chemistry.[10] Thus, the conventional canonical MP2 method can be suitable for calculating reference energies. However, these calculations would be very expensive for thorough surveys of the conformational space of monosaccharides (cf., $O(N^5)$ scaling of computer time with the size N). The pseudospectral local MP2 (LMP2) approximation[11,12] is considerably faster than canonical MP2 and approaches linear scaling for large systems but neglects dispersion-relevant terms. In addition, a consistent domain selection is critically important for different conformers. There is little knowledge about the effect of this LMP2 approximation for carbohydrates, and so we shall compare the MP2 and LMP2 results.

The third category comprises methods based on Kohn−Sham (KS) Density Functional Theory (DFT).[13] These methods treat electron correlation much more efficiently than post-HF methods. The simplest approximation is the nonempirical local spin density approximation (LSDA), the second level is the generalized gradient approximation (GGA, e.g., PBE[14]), and the third level is the meta-GGA (e.g., TPSS[15]). The hybrid functionals at the fourth level are fully nonlocal and semiempirical. (Although called semiempirical, these functionals are entirely different from the unreliable classical semiempirical WFN methods such as MNDO, AM1 and PM3.)

Because of its efficiency, DFT is widely used, having been applied to many types of chemical problems, as well as to condensed matter physics. The most popular GGA functionals are PBE,[14] BLYP,[16,17] and PW91,[18] and their hybrid variants: PBEh,[14,19,20] B3LYP,[21] and B3PW91.[18,22] The functional used most often for studies of carbohydrate conformations[5,9,23−42] has been B3LYP. Its wide acceptance is based, in part, on excellent calculated molecular geometries, molecular atomization energies, and enthalpies of formation of the molecules in the G2/97 test set.[43] More recent studies,[44] however, showed that B3LYP fails seriously for larger molecules of the G3/99 test set. The increase in B3LYP error with molecular size is evident from comparison of the errors for the series of *n*-alkanes: methane, ethane, etc., up to octane.[44] In Grimme's work,[45] B3LYP gave an error only slightly smaller than HF theory (10.3 and 13.4 kcal/mol, respectively) for the difference in energy between *n*-octane and its isomer, 2,2,3,3-tetramethylbutane. Depending on the test set, B3PW91 performs better than B3LYP or vice versa.[43,46,47] One important difference between the B3LYP and B3PW91 functionals is that the former does not respect the uniform electron gas limit while the latter does (see discussion in the Methods section). It is suspected that this error contributes to the deteriorating performance of B3LYP for large molecules and metals,[48] while other functionals perform considerably better for larger molecules or even for bulk solids.[48] There is growing evidence that B3LYP is not the best hybrid GGA method for conformational space studies of larger molecules.[13,49−51]

Some authors have reported that a large basis set (6−311++G**) with diffuse functions also on the hydrogen atoms is necessary for B3LYP relative energies of carbohydrate conformers.[25,26,30,31,33,36,38,40,42] This contradicts an earlier paper of Csonka,[9] who suggested that the 6−31+G* basis set is sufficient for geometry calculations and 6−31+G** or 6−311+G** suffice for relative energy calculations, whereas the 6−311++G** basis set is unnecessarily large.

In this paper, the performance of hybrid and nonempirical pure functionals (see section 2, Methods) is tested on a set of 15 allopyranose conformers, on a set of 15 conformers of the five-membered ring 3,6-anhydro-4-*O*-methyl-D-galactitol,[52] a conformationally flexible system, and on a set of four conformers of D-glucose[8,9,23] using various basis sets[40] (Figure 1). The MP2 method is used for reference because it includes to some extent the long-range correlation effects that are missing from the popular DFT functionals. Recent work shows that medium range correlation[53−55] can be reproduced by nonempirical functionals like PBE and PBEsol.[54] Our focus here is on the relative energies calculated from the total energies. If those values are good, then reliable enthalpy and free energy values can be calculated through established methods.

2. Methods

We used the following methods for the present study: canonical MP2, pseudospectral LMP2,[11,12] HF, KS-DFT approximations like hybrid HF/DFT: B3LYP, B3PW91, PBEh, and M05-2X,[56,57] and pure nonempirical DFT of increasing complexity: LSDA, PBE, PBEsol, and TPSS. The exchange-correlation (xc) energy E_{xc}^{hybrid} of recent hybrid functionals like PBEh is defined as

$$E_{xc}^{hybrid} = aE_x^{HF} + (1 - a)E_x^{DFA} + E_c^{DFA} \qquad (1)$$

where a is a mixing factor of E_x^{HF} Hatree-Fock and E_x^{DFA} density functional approximation (local functionals like GGA or meta-GGA) exchange energies, and E_c^{DFA} is a density functional approximation correlation energy. The value of a cannot be fixed universally. Good thermochemistry can be obtained[15] by setting $0.1 < a < 0.25$. To obtain good reaction energy barriers, values of $a > 0.5$ are usually required (cf., M05-2X).[56,57]

The first successful and popular hybrid B3PW91 xc energy was defined by Becke as

$$E_{xc}^{B3PW91} = aE_x^{HF} + (1 - a)E_x^{LDA} + \\ b\Delta E_x^{B} + E_c^{LDA} + c\Delta E_c^{PW91} \qquad (2)$$

where E_x^{LDA} and ΔE_x^{B} are the LDA exchange energy and Becke 88 exchange energy correction to the LDA, respectively. E_c^{LDA} and ΔE_c^{PW91} are the LDA correlation energy and Perdew−Wang correlation energy correction to the LDA, respectively. The a, b, and c constants were optimized for good thermochemistry on the G2 test set ($a = 0.2$, $b = 0.72$, $c = 0.81$). It was noted that the b and c parameters may cancel each other's effect (e.g., the effect of increasing the value of b can be compensated by the effect of decreasing c and vice versa); thus different parameter sets can give practically the same result.[58]

The popular hybrid B3LYP xc energy was defined somewhat differently

$$E_{xc}^{B3LYP} = aE_x^{HF} + (1 - a)E_x^{LDA} + b\Delta E_x^{B} + \\ (1 - c)E_c^{VWN3} + cE_c^{LYP} \qquad (3)$$

where E_c^{VWN3} and E_c^{LYP} are the Vosko−Wilk−Nussair parametrization[59] to random phase approximation (thus not to the exact LDA correlation) and the Lee−Yang−Parr correlation energies, respectively. Becke's values for a, b, and c were used. Neither E_c^{VWN3} nor E_c^{LYP} reproduce the known exact correlation energy of the homogeneous electron gas. It has been shown that the short-range correlations are described well by the LYP functional, but important long-range correlations are missing.[60]

The methods were combined with standard Gaussian basis sets from 6−31G* up to 6−311++G** and some correlation consistent basis sets: cc-pVTZ(-*f*) (without *f* functions on heavy atoms and *d* functions on hydrogen atoms), cc-pVTZ, and aug-cc-pVTZ.[61] Because the diffuse functions on the hydrogen atoms have negligible effect on the relative energies (vide infra), we suggest eliminating these functions from the aug-cc-pVTZ basis set. The resulting basis set is denoted as a-cc-pVTZ.

The present calculations were carried out with Gaussian 03,[62] Jaguar 6.0107, or 7.0213[63] on a 32-node Linux cluster, or on a Windows-based computer. We observed numerical instabilities for pseudospectral LMP2/cc-pVQZ calculations and thus did not use this model chemistry. Note that the domains generated by the procedure applied in the Jaguar computer program are sensitive to the basis set and this might lead to erratic basis set convergence, which we did experience in this work. A similar observation was made by Kaminsky et al.[10]

When using the Jaguar implementation of M05-2X with the standard DFT grid, we found that many conformer minimizations had large numerical errors in the gradient calculations. This would lead the minimizations out of the proper minimum to a completely different geometry. Indicators of this situation were a large number of optimization steps and increasing energy. In most of these cases, the final geometry was fairly different from the geometry obtained with the higher density DFT grid. It was therefore essential to use the dense DFT grid with M05-2X in Jaguar (the computer time increased 54−75%). No similar problem was observed for the B3LYP, B3PW91, or PBEh functionals. For the B3LYP functional, the difference of the total energies calculated with standard and dense numerical integration grids is in the range of −0.07 and +0.04 kcal/mol for the conformers studied in this paper. For relative energies (vide infra), this translates to 0.02 kcal/mol mean absolute deviation (MAD) in the range of −0.06 and +0.06 kcal/mol. This is the inherent error of the numerical integration grid.
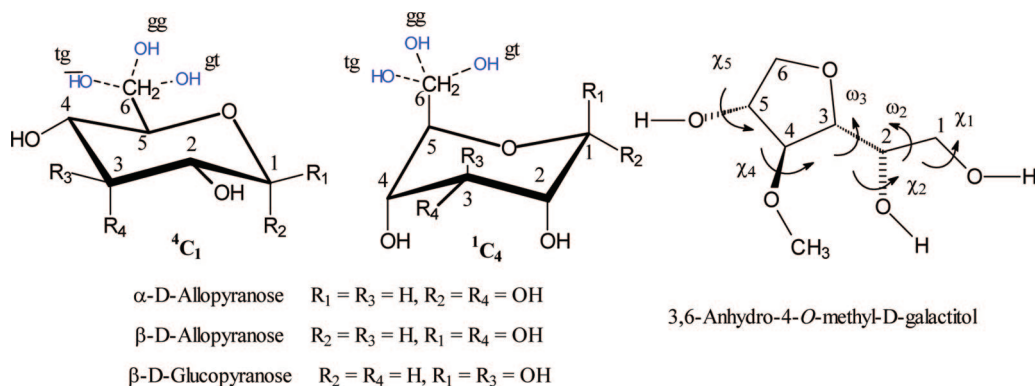

Figure 1. Schematic representations of the α,β-D-allopyranose and glucopyranose rings and the torsion angles in the 3,6-anhydro-4-*O*-methyl-D-galactitol. The ring carbon atoms are numbered according to the standard. For the O(5)−C(5)−C(6)−O(6) torsion angle of pyranoses tg, gg, and gt denotes that O(6) is in the anti position to O(5) and the gauche position to C(4), O(6) is in the gauche position to O(5) and to C(4), and O(6) is in the gauche position to O(5) and the anti position to C(4), respectively.

The geometries were fully relaxed and optimized unless single point calculations were mentioned. The puckering parameters were obtained according to Cremer and Pole.[64] The Boltzmann population of a particular conformer j within the set of n conformers was calculated as ($T = 298$ K)

$$P_j = \frac{e^{-E_j/RT}}{\sum\limits_{i=1}^{n} e^{-E_i/RT}} \quad (4)$$

where P_j is the ratio of the conformer j, E_j is the energy of the conformer j, R is the gas constant, and T is the temperature in K. P_j is expressed as % or ppm, by multiplying by 100 or 10^6, respectively. The P_j ratios are useful for characterizing the various conformer distributions and show the practical implications for the predicted equilibria that are caused by using the various model chemistries (method and basis set).

We use the following definitions of the relative energies:

$$\Delta E_{model}(conf_i, conf_{ref}) = E_{model}(conf_i) - E_{model}(conf_{ref}) \quad (5)$$

is the difference between the energy of the ith conformer, $conf_i$ and the energy of a reference conformer, $conf_{ref}$ using the given model chemistry.

$$\Delta\Delta E_{model\ A-model\ B}(conf_i, conf_{ref}) = \Delta E_{model\ A}(conf_i, conf_{ref}) - \Delta E_{model\ B}(conf_i, conf_{ref}) \quad (6)$$

is the difference between the relative energies of the ith conformer calculated with two different model chemistries.

$$\Delta\Delta E_{corrMP2}(conf_i, conf_{ref}) = \Delta E_{MP2}(conf_i, conf_{ref}) - \Delta E_{HF}(conf_i, conf_{ref}) \quad (7)$$

is the relative (MP2) correlation energy difference of the ith conformer. Note that $E_{corrMP2}(conf_i) = E_{MP2}(conf_i) - E_{HF}(conf_i)$, and thus the relative (MP2) correlation energy difference equals $\Delta E_{corrMP2}(conf_i, conf_{ref})$.

The model and reference conformer dependent mean deviation (MD) is defined as

$$MD_{model\ A-model\ B}(conf_{ref}) = \frac{1}{n-1} \sum_{i=1}^{n} \Delta\Delta E_{model\ A-model\ B}(conf_i, conf_{ref}) \quad (8)$$

and the model and reference conformer dependent mean absolute deviation (MAD) is defined as

$$MAD_{model\ A-model\ B}(conf_{ref}) = \frac{1}{n-1} \sum_{i=1}^{n} |\Delta\Delta E_{model\ A-model\ B}(conf_i, conf_{ref})| \quad (9)$$

Note that within a given test set of conformers the MD and MAD between the two compared models (model A and model B) depend on the choice of the reference conformers. If we compare several tested model chemistries to the MP2 reference using eqs 6 and 9, the evaluation of the performance of the tested models might depend on the choice of the reference conformer. One way to eliminate this problem

is to compare the range of $\Delta\Delta E_{modelA-modelB}(conf_i, conf_{ref})$ values of eq 7. We note this as the range of the relative difference, $RRD = \max \Delta\Delta E - \min \Delta\Delta E$.

3. Results and Discussion

3.1. D-Allopyranose Conformers. Schnupf et al.[40] recently performed B3LYP/6−311++G** geometry optimizations on α- and β-D-allopyranose, with chair, boat, and skew ring shapes, and various exocyclic group orientations. We selected fifteen of their 102 structures for the current work, including 13 $^4C_1$ conformers (8 α- and 5 β-anomers, with gg, gt, and tg hydroxymethyl rotamers, see Figure 1), one α- and one β- $^1C_4$ conformer (the ALL15 test set; six of the selected conformers are shown in Figure 2). The range of the B3LYP/6−31+G** relative energies is 5.51 kcal/mol. The Cartesian coordinates (Å), total energies (a.u.), and $\Delta E_{model}(conf_i, 1)$ (kcal/mol) of the tested conformers are in the Supporting Information. The first low-energy reference conformer is the $^4C_1$ α-D-allopyranose with torsion angles O(5)−C(5)−C(6)−O(6) $\cong 180°$ and C(5)−C(6)−O(6)−H $\cong -60°$, noted as tg and g^-, respectively (see, conformer 1 in Figure 2).

First we compare the MP2/cc-pVTZ(-f), MP2/a-cc-pVTZ-(-f) and LMP2/cc-pVTZ(-f) relative energies from eq 5 using the optimized B3LYP/6−31+G* geometries (to exclude geometry effects). The MP2 and LMP2 relative energy differences calculated from eq 6 and the relative correlation energy differences calculated from eq 7 show considerable deviations with the same cc-pVTZ(-f) basis set

$$\Delta\Delta E_{corrLMP2}(conf_i, 1) = 0.8\Delta\Delta E_{corrMP2}(conf_i, 1) - 0.21(kcal/mol) \quad (10)$$

with $R^2 = 0.989$. This shows that the missing correlation energy has a considerable influence on the relative energy differences calculated using eq 6. Note that the LMP2 approximation also neglects part of the dispersion-relevant double excitations and might show slower basis set convergence[10] than MP2. The intramolecular effect of this LMP2 approximation is not clear. Detailed analysis for the ALL15 test set shows that the LMP2 correlation energy is less negative, thus part of the MP2 correlation energy (about 2.4% or 32 kcal/mol) is lost. However, LMP2/cc-pVTZ(-f) and MP2/a-cc-pVTZ(-f) relative energies show a good agreement (RRD = 0.48, and MAD =0.10 kcal/mol). We have also performed MP2/cc-pVQZ(-g) calculations and CBS[3,4] extrapolation for the first five conformers and obtained good agreement with the MP2/a-cc-pVTZ(-f) results (RRD = 0.2 kcal/mol). This shows that the use of the diffuse s, p, and d functions is advantageous for MP2 relative energy calculations. At the MP2 level we found that the a-cc-pVTZ(-f) basis set represents a good compromise between accuracy and speed in agreement with Kaminsky et al.[10] Table 1 shows the good agreement between our reference MP2 and LMP2 results for β and $^1C_4$ ratios.

Smaller groups of conformers can be found in the test set for which the $\Delta\Delta E_{corrLMP2}(conf_i)$ and $\Delta\Delta E_{corrMP2}(conf_i)$ (eq 7) are within ± 0.2 kcal/mol (like conformers {1, 8, 15} shown in Figure 3a or {2, 6, 13} or {5, 7, 9, 10, 12}). Within
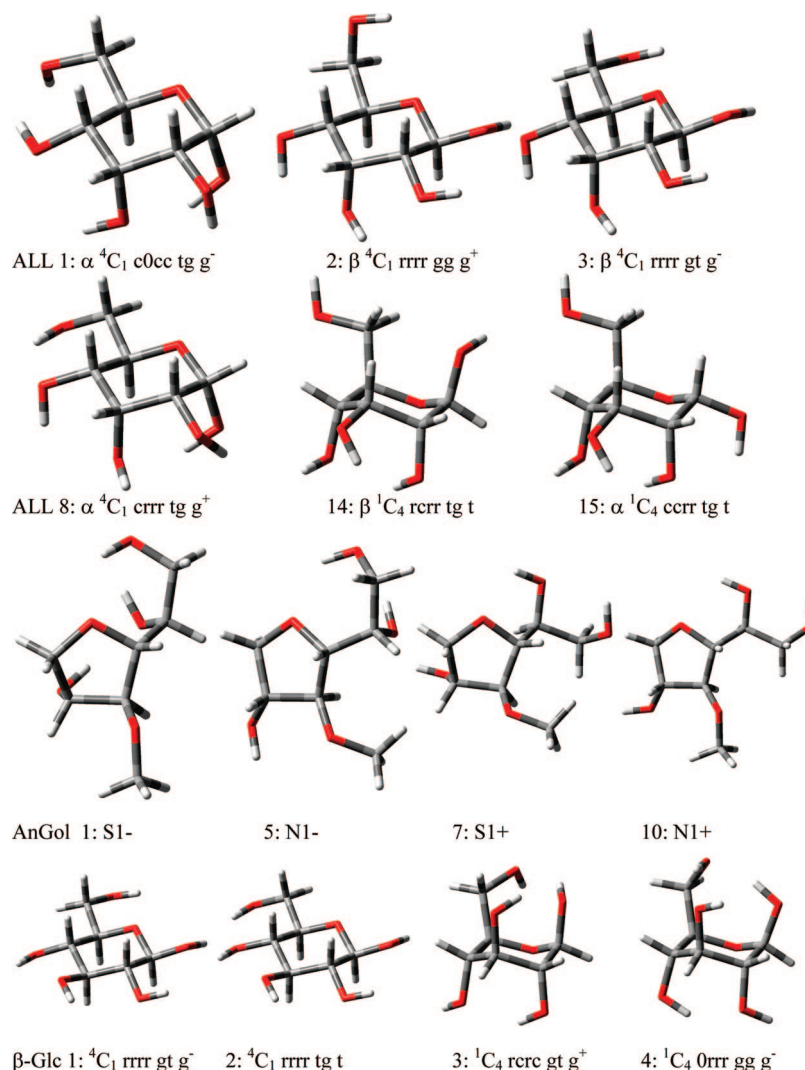
Figure 2. Geometries of some of the selected α,β-D-allopyranose, 3,6-anhydro-4-O-methyl-D-galactitol, and β-D-glucopyranose conformers of the ALL15, AnGol15 and Glc4 test sets are shown, respectively. α and β denote the anomeric configurations, $^4C_1$ and $^1C_4$ denote the ring shapes. r denotes the counterclockwise, c the clockwise, 0 the neutral OH directions. tg, gt, and gg show the hydroxymethyl orientation (cf. Figure 1), and g-, g+ and t the O(6)H orientation.

these groups, the HF, MP2, and LMP2 relative energies (eq 5) agree with MAD < 0.2 and RRD < 0.35 kcal/mol. These relative energies are almost converged because the HF energies converge considerably faster to the basis set limit than the MP2 energies and the correlation contributions to the relative energies are negligible. Thus, within these groups of conformers the HF method gives reliable relative energies at a much lower cost than the very expensive MP2/CBS extrapolations. These reliable relative energies are suitable for testing density functional methods. Note that $\Delta\Delta E_{\mathrm{corrMP2}}(\mathrm{conf}_i, 1)$s between α and β anomers are large positive values, 2.9−3.9 kcal/mol (inclusion of the MP2 correlation energy destabilizes the β anomers), and these relative energies are less converged with respect to the basis set.

Figure 3b shows the performance of the hybrid functionals and the effect of using B3PW91 (eq 2) instead of B3LYP (eq 3).[16−18] The conventional hybrid functionals show similar agreement with the MP2 results ($\mathrm{RRD}_{\mathrm{B3LYP\text{-}MP2}} = 1.19$, $\mathrm{RRD}_{\mathrm{PBEh\text{-}MP2}} = 1.12$ $\mathrm{RRD}_{\mathrm{B3PW91\text{-}MP2}} = 1.09$ kcal/mol). The M05-2X hybrid results agree best with the MP2 results

($\mathrm{RRD}_{\mathrm{M05\text{-}2X\text{-}MP2}} = 0.90$ kcal/mol). M05-2X has shown considerably better performance than B3LYP for main-group thermochemistry, kinetics, and noncovalent interactions.[56,57] The M05-2X functional does model some of the medium range correlation effect that is missing from the B3LYP functional. This could be the origin of better agreement between M05-2X and MP2 results.

Figure 3c shows quite good agreement between the nonempirical PBE, TPSS and MP2 results. PBE and TPSS functionals show the second and third best performances after the M05-2X functional ($\mathrm{RRD}_{\mathrm{PBE\text{-}MP2}} = 0.97$ and $\mathrm{RRD}_{\mathrm{TPSS\text{-}MP2}} = 1.09$ kcal/mol). As M05-2X, PBE and TPSS functionals treat medium range correlation correctly,[53] this could be the origin of the good agreement with the MP2 results. LDA shows a large error compared to MP2, and this error is the opposite of the HF error (cf., Figures 3a and c and $\mathrm{RRD}_{\mathrm{LSDA\text{-}MP2}} = 3.39$ kcal/mol).

Table 1 shows how HF results deteriorate (with serious overestimation of the β-anomer and underestimation of the $^1C_4$ conformer ratios) as the basis set quality increases to near the basis set limit, in agreement with earlier observa-

684 *J. Chem. Theory Comput., Vol. 5, No. 4, 2009*

Csonka et al.

Table 1. Boltzmann Populations Calculated (at 25 °C) for the ALL15 Test Set, for β-Anomers (of $^4C_1$ Conformers), and of Total $^1C_4$ Conformers, Calculated by Different Methods[a]

method	β-anomers (%)	$^1C_4$ (ppm)
reference method: MP2/a-cc-pVTZ($-f$)[b]	**23.0**	**280**
LMP2/cc-pVTZ($-f$)[b]	**27.6**	**339**
MP2/cc-pVTZ($-f$)[b]	11.1	**254**
B3LYP/6−311++G**	40.4	88
B3LYP/6−311+G**	40.5	86
B3LYP/6−311+G**//B3LYP/6−31+G**	41.1	83
B3LYP/6−311+G**[b]	41.2	82
B3LYP/6−31+G**	39.4	95
B3LYP/6−31+G**[b]	39.3	94
B3LYP/6−31+G*	30.7	101
B3PW91/6−311+G**	29.2	96
B3PW91/6−311+G**//B3PW91/6−31+G*	28.1	93
B3PW91/6−31+G**	28.7	99
B3PW91/6−31+G//B3PW91/6−31+G***	**26.7**	104
B3PW91/6−31+G*	19.8	110
PBEh/6−311+G*	**27.1**	92
M05-2X/6−311+G*	**18.6**	92
M05-2X/6−311+G//M05-2X/6−31+G***	**18.7**	92
M05-2X/6−311+G**//M05-2X/6−31+G*	18.8	90
M05-2X/6−31+G**	18.0	135
M05-2X/6−31+G**//M05-2X/6−31+G*	18.2	134
M05-2X/6−31+G*	11.6	132
LSDA/6−311+G**	1.0	31
PBE/6−311+G*	12.1	**138**
TPSS/6−311+G*	7.9	**141**
HF/6−31G*	68.5	41
HF/cc-pVTZ(-f)//B3LYP/6−31+G*	92.1	19

[a] The best agreements with the MP2 reference are shown in bold. [b] Geometry optimized at the B3LYP/6−31+G* level.

tions.[9] The large basis set HF results give reliable relative energies only within a few groups of conformers as discussed earlier. Figure 3a shows the effect of the electron correlation at the MP2 level (cf., MP2 and HF curves, RRD$_{HF-MP2}$ = 3.42 kcal/mol).

The Boltzmann populations in Table 1 show that B3LYP gives too large of a β-anomer ratio (cf., conformers 2, 3, 6, 11, and 13, in Figure 3) and too small of a $^1C_4$ conformer ratio. The $^1C_4$ ratios obtained by PBE and TPSS are closer to the MP2 results than any hybrid functional result (cf., Table 1). Note that B3LYP/6−31+G* results deviate considerably from the results obtained with the larger basis sets, as 6−31+G* basis set is poor for relative energy calculations. However, the single-point calculations that use this geometry and a better basis set (e.g., 6−31+G** or larger) give sufficiently converged relative energies (see discussion below).

Next we focus on determining the optimal basis set, that is, one that gives converged results for density functionals. Schnupf et al.[40] claimed that the minimal and optimal basis set for B3LYP conformational energy studies of the allopyranose conformations is 6−311++G**, contradicting the recommendations of one of the present authors (GIC).[9] However, Schnupf et al. did not present results for the recommended 6−31+G** and 6−311+G** basis sets. The relative energy differences (eq 6) in Figure 4 show that the B3LYP/6−311+G** relative energies are indistinguishable from the B3LYP/6−311++G** relative energies at the scale of the figure. Thus, eliminating the diffuse functions on the H-atoms (using + instead of ++) has a negligible

effect on relative energies,[9] although there is a very slight influence on the total energies (not shown). (Diffuse functions are important for H-atoms if they have negative partial charges.) The model dependent MAD (cf., eq 7, where model B is B3LYP/6−311++G** and model A is B3LYP/6−311+G**) is equal to 0.03 kcal/mol (comparable to the 0.02 kcal/mol grid error discussed in the Methods section). Even the B3LYP/6−31+G** results show quite small deviations from the more expensive B3LYP/6−311++G** results (MAD = 0.15 kcal/mol). Removing the p polarization functions from the H-atoms (cf. 6−31+G* basis set, results in Figure 3a), more than doubles the MAD (0.32 kcal/mol). Consequently, the 6−31+G* basis set should not be used for relative energy calculations, in agreement with previous suggestions.[9,40] Table 1 also shows that simplifying the basis set systematically decreases the proportion of β-anomers.

Finally, we consider whether single point calculations based on geometries determined with smaller basis sets can be used safely. Table 1 and Figure 4 show that the most expensive, fully optimized B3LYP/6−311++G** results and single point B3LYP/6−31+G**//B3LYP/6−31+G* or B3LYP/6−311+G**//B3LYP/6−31+G* results are very similar (MAD =0.15 and 0.02 kcal/mol, cf. eq 7). Calculations carried out with B3PW91 yield a similar conclusion (Table 1, Figure 4b). Thus, the basis set dependence of B3LYP and B3PW91 geometries is small, and these small geometry changes do not influence the relative energies for the allose test set. For geometry optimization, the 6−31+G* basis set is sufficient. The basis set and geometry errors are considerably smaller than the error of the B3LYP functional. For allose, these model chemistries require about 50% and 20% less computer time, respectively, compared to full geometry optimization with the 6−311++G** basis set. Note that a large number of diffuse functions makes the convergence of the SCF and the geometry optimization cycles more difficult, so using more diffuse functions than necessary can lead to convergence problems or no convergence at all in some cases.

In summary, for the ALL15 test set the nonempirical GGA and meta-GGA and empirical M05-2X functionals show good agreement with the reference MP2/a-cc-pVTZ($-f$) results. The order of the methods against this reference is the following: LMP2 ≫ M05-2X > PBE > TPSS > B3PW91 > PBEh > B3LYP ≫LSDA ≅ HF (basis set limit).

3.2. 3,6-Anhydro-4-O-methyl-D-galactitol Conformers. Navarro and Stortz recently studied[52] the conformational space of 3,6-anhydro-4-*O*-methyl-D-galactitol (AnGol) by molecular mechanics and quantum mechanical methods (B3LYP and MP2). AnGol has a flexible five-membered ring and six exocyclic torsion angles (cf., Figures 1 and 2), all of them with a large influence on the energy. The two main stable conformations of the five-membered ring were identified as North (N), around E_4, ^5E, and E_6 (Cremer−Pople puckering parameter $\phi \approx 250-320°$), and South (S), around $^6T_5$ ($\phi \approx 120°$). Fifteen conformers derived from the four most stable conformations in each of the ring regions and of its main side chain were used.[52] The conformers 1−15 were designated as S1−, S2−, S4−, S3−, N1−, N2−, S1+, S2+, N3−, N1+, S4+, S3+, N3+, N2+, and N4+, respectively,

Density Functionals and Basis Sets
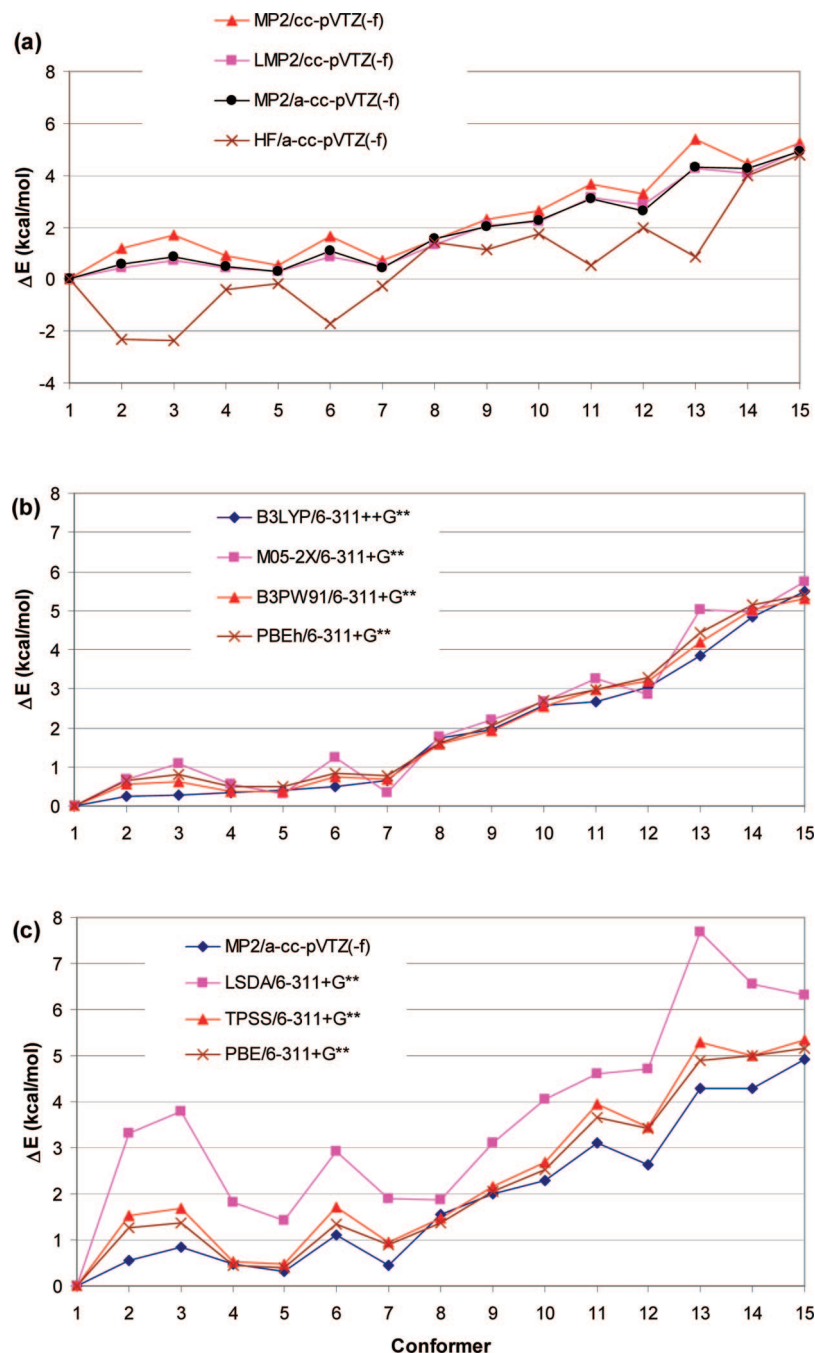
J. Chem. Theory Comput., Vol. 5, No. 4, 2009 **685**

Figure 3. Relative energies of the ALL15 conformers (with respect to conformer 1), $\Delta E_{model}(\text{conf}_i,1)$ (see, eq. 5) optimized by various model chemistries as shown in the legends. The single point MP2, LMP2, and HF energies were calculated with B3LYP/6−31+G* geometry.

in ref 52 and constitute the current AnGol15 test set (conformers 1, 5, 7, and 10 are shown in Figure 2). The range of the B3LYP/6−31+G** relative energies is 6.18 kcal/mol.[52] The MP2/6−311+G**//B3LYP/6−31+G** results for the 15 conformers in Figure 5a and Table 2 were obtained from the previous study.[52]

The conformational space of the 3,6-anhydro-4-*O*-methyl-D-galactitol is quite different from the conformational space of the D-allopyranose (cf., Figure 1). It is an interesting question whether the observations on model chemistries for allose are valid for this more flexible compound.

Earlier studies for compounds like 2-hydroxytetrahydro-pyran and 2-methoxytetrahydropyran showed[62] that MP2/6−311++G** overestimates the absolute values of relative energies compared to MP2/CBS extrapolation results.[65] It was also shown that CCSD(T) contributions are quite small (0.02−0.1 kcal/mol), so MP2/CBS results are quite well converged for correlation energy differences. Gould et al. found a considerable difference between the relative energies of the MP2 and LMP2 methods, and proposed the latter method for benchmark calculations for di- and trisaccharides because of the smaller basis set error.[66]

Our results in Table 2 show that the single point LMP2/aug-cc-pVTZ and cc-pVTZ(−*f*) results agree very well (MAD = 0.18 kcal/mol) for the proportion of N (0.07−0.09%) and $\omega_3 = \text{g}+$ (0.22%) conformers. The LMP2/aug-cc-pVTZ
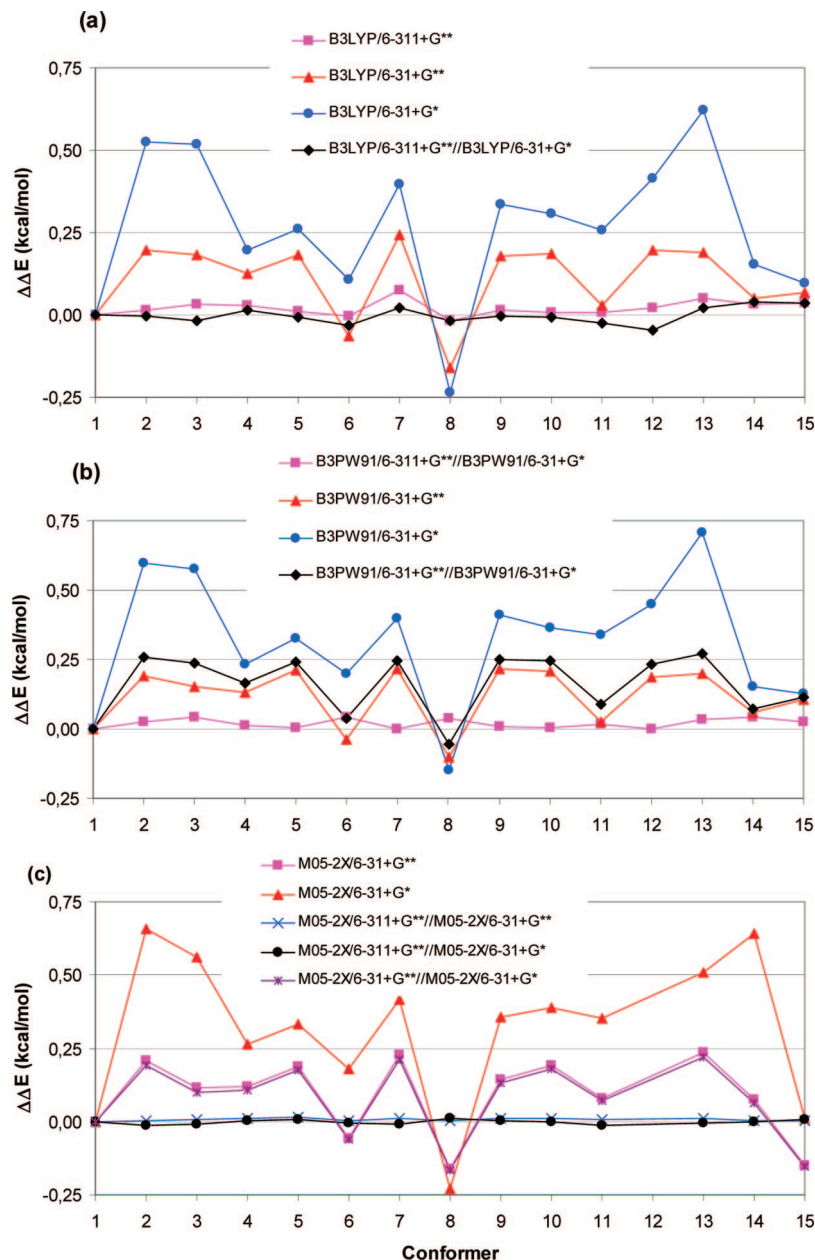
Figure 4. Differences in relative energies for the ALL15 allose conformers (with respect to conformer 1), $\Delta\Delta E_{\text{modelA}-\text{modelB}}(\text{conf}_i, 1)$ of eq 7. Model A is shown in the legends and (a) model B = B3LYP/6−311++G**, (b) model B =B3PW91/6−311+G**, and (c) model B = M05-2X/6−311+G**.

relative energies are somewhat different (MAD = 0.30 kcal/mol) and give lower ratios for the minor conformers. The MP2/6−311+G** and MP2/a-cc-pVTZ(-*f*) results agree quite well (MAD = 0.15, RRD = 0.75 kcal/mol). This is considerably better agreement than the agreement between MP2/cc-pVTZ(-*f*) and MP2/a-cc-pVTZ(-*f*) results (MAD = 0.52, RRD = 0.99 kcal/mol). This shows the importance of the inclusion of the diffuse functions into the basis set for MP2 calculations. The agreement between LMP2 and MP2 results is considerably worse for the AnGol15 test set than for the ALL15 test set (cf., Figure 5a). The differences between the LMP2 and MP2 limits can be attributed to the intramolecular effect of the missing dispersion-relevant double excitations from the LMP2 approximation. The CBS extrapolation is beyond the scope of the current study as it requires very expensive calculations with quadruple-ζ basis

sets. We can select again smaller groups of conformers for which the $\Delta\Delta E_{\text{corrMP2}}(\text{conf}_i)$ (eq 7) are within ± 0.2 kcal/mol (like conformers {1, 2}, {3, 4}, {5,6,11,12,15}, {7,8}, and {9,10}). Within these groups the HF, MP2 and LMP2 relative energies agree with MAD < 0.2 kcal/mol. The effect of the electron correlation is shown in Figure 5a (cf., HF and MP2 curves). It is larger for the Angol15 test set than for the ALL15 test set (RRD = 4.83 and MAD = 3.18 kcal/mol).

Figure 5b shows the very good agreement between M05-2X and MP2/a-cc-pVTZ(-*f*) results (MD = 0.2 and MAD = 0.3 kcal/mol), while other functionals show considerably poorer performance. The M05-2X functional shows particularly good performance for the smaller group of conformers where the relative energies are known with higher precision (MAD < 0.1 and RRD < 0.1−0.4 kcal/mol). Similar
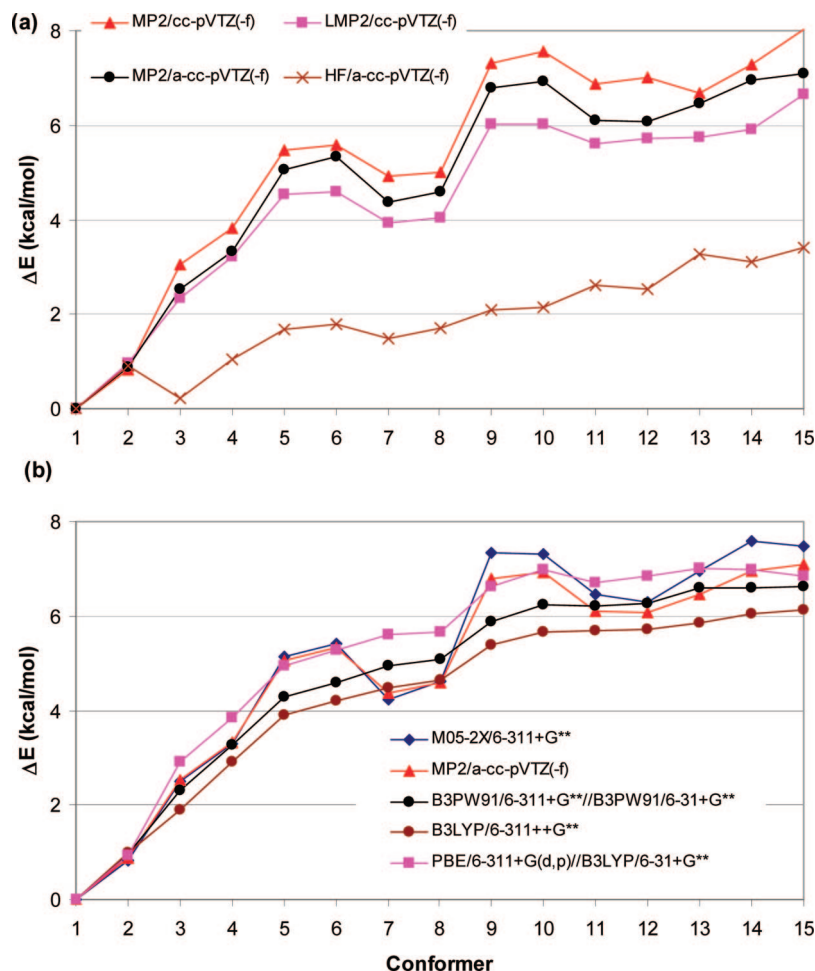
Density Functionals and Basis Sets

J. Chem. Theory Comput., Vol. 5, No. 4, 2009 **687**

Figure 5. Relative energies of the AnGol15 conformers (with respect to conformer 1), $\Delta E_{model}(conf_i, 1)$ (cf., eq 5) optimized by various model chemistries as shown in the legends. The single point MP2, LMP2, and HF energies were calculated with B3LYP/$6-31+G^*$ geometry.

agreement between MP2 and M05-2X results was observed earlier.[55] The MP2 and B3LYP results are similar for Southern conformers (1−4, 7, 8 11, 12), but the B3LYP functional underestimates the relative energies for Northern conformers (5, 6, 9, 10, and 13−15). The B3PW91 functional shows considerably better performance than B3LYP (cf. MAD = 0.42 and 0.68 kcal/mol, respectively). The PBE and TPSS (not shown) functionals show an opposite error, that is, they give correct Northern conformer relative energies (eq 5), but systematically overestimate the relative energies for Southern conformers.

The LSDA method shows very serious underestimation of the ratio of N and $\omega_3 = g+$ conformers (cf., Table 2). The HF/6−31G* model chemistry performs relatively well, as usual, but increasing the basis set quality to the basis set limit leads to serious errors and gives erroneously high populations of the minor conformers (cf., Table 2 and Figure 4b). Similar to the ALL15 test set, the HF and the LSDA functional yields the opposite error. The results in Table 2 show that the relative energies and geometries calculated by M05-2X are even less basis set dependent (0.11 kcal/mol < MAD) than those of B3LYP or B3PW91 (MAD < 0.16 kcal/mol), when comparing 6−311+G** or 6−31+G** calculations. All calculations using the 6−31+G* basis set give poor relative energies (cf. MAD = 0.54−0.62 kcal/mol).

Next, we investigate the consistency of the optimized geometries. We noted for the ALL15 and AnGol15 test sets that B3CF/6−311+G**//B3CF/6−31+G* (CF = LYP or PW91) relative energies approximate very well the relative energies of the B3CF/6−311++G** fully optimized structures. The results in Table 3 show that the geometries given by the B3LYP method with the 6−31+G*, 6−31+G**, and 6−311++G** basis sets agree very well. The small differences in Table 3 do not result in substantial changes in the relative energies (note that around the geometry minimum the energy surface is flat as the gradients are zero). Consequently, the 6−31+G* geometry can be used for single point energy calculations, and this gives a considerable efficiency compared to the geometry optimization with the 6−311++G** basis set suggested by many papers (approximately one-third of the computational time is needed). Table 3 also shows that the B3LYP optimized geometries are quite similar to those of B3PW91 (slightly different exocyclic torsions, and somewhat larger puckering), while those obtained by LSDA are very different. Minimization with the HF/6−31G* model chemistry leads to very similar geometries to those obtained by B3LYP calculations (cf. Table 3): puckering is almost identical, and small differences in the exocyclic angles can be observed. The M05-2X geometries are different from the B3LYP and HF/6−31G*

Table 2. Boltzmann Populations Calculated (at 25 °C) for the AnGol15 Test Set, for the Northern (N) Conformers, and of Conformers Carrying the Larger Side Chain Torsion, ω_3 (Defined by Atoms H3−C3−C2−C1) with a g+ Value Calculated by Different Methods[a]

Method	N conformers (%)	$\omega_3 = g+$ (ppm)
reference method: MP2/a-cc-pVTZ(-*f*)[b]	**0.090**	**2200**
LMP2/cc-pVTZ(−*f*)[c]	0.071	**2208**
LMP2/cc-pVTZ[c]	**0.092**	2196
LMP2/aug-cc-pVTZ[c]	0.078	**2200**
MP2/6−311+G**[b]	0.027	1135
MP2/cc-pVTZ(-*f*)[b]	0.017	388
B3LYP/6−311++G**	0.197	**992**
B3LYP/6−311+G**//B3LYP/6−31+G*	0.199	824
B3LYP/6−311+G**	0.143	721
B3LYP/6−31+G*	0.073	265
B3PW91/6−311+G**//B3PW91/6−31+G**	**0.103**	444
B3PW91/6−31+G**	0.076	325
M05-2X/6−311+G**	0.025	**983**
M05-2X/6−311+G**//M05-2X/6−31+G**	0.025	958
M05-2X/6−311+G**//M05-2X/6−31+G*	0.025	958
M05-2X/6−31+G**	0.023	752
M05-2X/6−31+G**//M05-2X/6−31+G*	0.021	805
M05-2X/6−31+G*	0.010	233
LSDA/6−311+G**//LSDA/6−31+G**	0.000	0
LSDA/6−31+G**	0.000	0
PBE/6−311+G**[b]	0.034	166
TPSS/6−311+G**[b]	0.031	132
HF/6−31G*	0.432	3818
HF/aug-cc-pVTZ[c]	7.275	96256
HF/cc-pVTZ[c]	3.962	48594
HF/cc-pVTZ(-*f*)[c]	4.486	46311
HF/6−311+G**[b]	3.965	49896

[a] The best agreements with the MP2 reference are shown in bold. [b] Geometry optimized at the B3LYP/6−31+G** level. [c] Geometry optimized at the B3LYP/6−311+G** level.

geometries (cf. the difference in the puckering parameters in Table 3). Thus, the differences between the M05-2X and the B3LYP geometries can be attributed to the medium range correlation effects being better described by the M05-2X functional.

The relative time scale for single point calculations LSDA/6−311+G**, B3LYP/6−311+G**, LMP2/cc-pVTZ(-*f*), LMP2/cc-pVTZ, and LMP2/aug-cc-pVTZ is 1, 1.6, 4.2, 10, and 50, respectively. Note that the time requirement of pure GGA functionals like PBE is similar to that of the LSDA.

In summary, for the AnGol15 test the M05-2X functional shows a good agreement with the reference MP2/a-cc-pVTZ(-*f*) results. The order of the functionals, LMP2 and HF is the following: M05-2X > LMP2 > TPSS ≅ PBE > B3PW91 > B3LYP ≫ HF (basis set limit) > LSDA.

3.3. β-D-Glucopyranose Conformers. The GLC4 test set is composed of two low-energy $^4C_1$ conformers (1 and 2) and two $^1C_4$ conformers (3 and 4) of β-D-glucopyranose. These four conformers were used for testing KS-DFT functionals before[9,23] and a composite energy is available[8] for the relative energies of these structures (cf., Table 4). This composite energy is based on the MP2/cc-pVTZ//MP2/cc-pVDZ model chemistry, with basis set correction at the HF level up to cc-pVQZ and correlation correction using MP2 and CCSD/6−31G*. Such estimations of the basis set limit and the correlation contribution are possibly unreliable, and the composite energies might be imprecise for conformers 3 and 4. Note that for conformers 3 and 4 the relative

energies obtained with the MP2/cc-pVTZ//MP2/cc-pVDZ model chemistry are far from the composite energies;[8,23] thus the basis set and correlation corrections are large, 2.3 and 3.3 kcal/mol, respectively.

In the present work, we use the MP2/a-cc-pVTZ(-*f*)//B3LYP/6−31+G* model chemistry as a new alternative reference (cf., Table 4). These, and the LMP2/cc-pVTZ-(-*f*)//B3LYP/6−31+G* results agree within 0.5 kcal/mol. We have also performed an MP2/CBS[3,4] estimation for the relative energies[67] using separate HF extrapolation and MP2 correlation contribution extrapolations of the cc-pVTZ and cc-pVQZ relative energies.[68] These results agree again within 0.5 kcal/mol error with the MP2/a-cc-pVTZ(-*f*) results. However, as Table 1 shows, the previous composite calculation[8] gave poor relative energy for conformer **4** (∼1.5 kcal/mol error). More precise determination of the relative energy of conformer **4** of the GLC4 set requires a considerably larger computational effort, one that is beyond the scope of the current paper.

The current results in Table 4 show the good performance of the PBE, TPSSh, PBEh, and B3PW91 functionals (MAD = 0.12, 0.28, 0.32, 0.41 kcal/mol, respectively). In addition, the single point calculations deliver quite good results showing that even the geometries obtained with the smallest basis set applied in this paper (6−31G*) are useful. The B3LYP functional again performs slightly worse, MAD = 0.69 kcal/mol) than the B3PW91, PBEh or M05-2X functionals (MAD = 0.64 kcal/mol, because of overestimation of the stability of the conformers **3** and **4**, see Table 4). The HF/cc-pVDZ model chemistry performs reasonably well but the HF/cc-pVTZ//HF/cc-pVDZ model chemistry performs poorly (with serious underestimation of the hydrogen bond (H-bond strength), cf. results for the ALL15 and AnGol15 test sets). The PBEsol/6−311+G** model chemistry, using the new pure GGA functional, performs similar to TPSS, and M05-2X, with MAD = 0.65 kcal/mol despite the fact that PBEsol was successfully applied to isomerization problems.[55] The smaller enhancement factor for the large reduced gradient gives too strong H-bonds (like the LSDA). A change to the large gradient behavior of the exchange functional leads to improved results for saccharides as our preliminary results show. The details will be published elsewhere.

Notice the poor performance of the MP2/cc-pVTZ//MP2/cc-pVDZ, MP2/cc-pVDZ, MP2/6−31-G*, and CCSD/6−31G*//MP2/6−31G* model chemistries (MAD =1.12, 3.88, and 2.61 kcal/mol, respectively) used for the composite energies. The relative energies calculated by the DFT/6−31G* model chemistry are also very poor (MD < −5 kcal/mol; MAD > 5 kcal/mol), and show serious overestimation of intramolecular H-bond strength because of the basis set error. The minimal basis set required for reasonable energy is 6−31+G**. The 6−311+G** basis set gives converged results as observed for the previous test sets.

These results show that the GLC4 test set is particularly sensitive to the correct description of the intramolecular H-bond. So this leads to a somewhat different conclusion than the results from the larger ALL15 and AnGol15 test sets. These results shows clearly that the good HF/DZ results

Table 3. Average of the Absolute Values of Geometric Deviations (Six exo-Cyclic Torsions and Puckering parameters Q and ϕ) between Optimized Geometries (1 and 2) Obtained by Different Methods and Basis Sets for the 15 Selected 3,6-Anhydro-4-*O*-methylgalactose Conformers (AnGol15 Test Set)

optimized geometry 1	optimized geometry 2	exocyclic torsions (deg)	Q (Å)	ϕ (deg)
B3LYP/6−31+G**	B3LYP/6−311++G**	0.4	0.002	0.7
B3LYP/6−31+G*	B3LYP/6−311++G**	0.5	0.002	0.7
B3LYP/6−31+G*	B3LYP/6−31+G**	0.3	0.001	0.5
B3PW91/6−31+G**	B3LYP/6−311++G**	0.8	0.007	1.6
LSDA/6−31+G**	B3LYP/6−311++G**	5.0	0.026	5.4
M05-2X/6−311+G**	B3LYP/6−311++G**	1.3	0.024	5.4
M05-2X/6−31+G**	B3LYP/6−311++G**	1.3	0.025	5.8
M05-2X/6−31+G**	M05-2X/6−311+G**	0.6	0.002	1.2
M05-2X/6−31+G*	M05-2X/6−31+G**	0.3	0.001	0.6
HF/6−31G*	B3LYP/6−311++G**	2.0	0.002	0.9
HF/6−31G*	M05-2X/6−311+G**	2.7	0.022	9.7

Table 4. Relative Energies of Conformers 2, 3, and 4 of the GLC4 Test Set with Respect to Conformer 1 (Eq *5*) and Boltzmann Populations Calculated (at 25 °C) for the $^1C_4$ Conformers, Calculated by Different Methods[a]

method	ref	**2**	**3**	**4**	% $^1C_4$
MP2/a-cc-pVTZ(-f)//B3LYP/6−31+G*	this work	**0.38**	**5.76**	**5.10**	**0.0158**
MP2/CBS[3,4]//B3LYP/6−31+G*	67		**6.25**	**5.43**	
LMP2/cc-pVTZ(-f)//B3LYP/6−31+G*	this work	**0.29**	**5.92**	**5.63**	**0.0074**
composite	8	0.27	6.41	6.99	0.0017
MP2/cc-pVTZ//MP2/cc-pVDZ	8	0.07	4.13	3.65	0.1607
MP2/6−31G*	8	−0.45	0.66	−0.60	31.4
CCSD/6−31G*//MP2/6−31G*	8	−0.34	2.07	1.68	3.1
HF/6−31G*	8	−0.15	6.73	6.76	0.0010
HF/cc-pVDZ	8	−0.09	**6.23**	**6.34**	0.0023
HF/cc-pVTZ// B3LYP/6−31+G*	this work	0.49	11.11	10.96	0.0000
HF/cc-pVQZ// B3LYP/6−31+G*	this work	0.66	12.15	11.95	0.0000
HF/CBS[3,4]//B3LYP/6−31+G*	67	0.70	12.43	12.22	0.0000
B3LYP/6−31+G*	9	−0.12	6.92	4.93	0.0113
B3LYP/6−31+G**	this work	0.06	7.22	5.40	0.0060
B3LYP/6−31+G**//B3LYP/6−31+G*	this work	0.07	7.27	5.41	0.0059
B3LYP/6−311+G**	this work	0.16	7.24	5.59	0.0048
B3LYP/6−311+G**//B3LYP/6−31+G*	this work	0.16	7.36	5.64	0.0044
B3PW91/6−31+G**	this work	0.09	6.63	5.02	0.0120
B3PW91/6−311+G**	this work	**0.19**	**6.67**	**5.22**	**0.0093**
PBEh/6−31+G**	this work	−0.02	5.98	4.48	0.0271
PBEh/6−311+G**	this work	**0.09**	**6.14**	**4.80**	**0.0178**
TPSSh/6−311+G//TPSSh/6−31G***	this work	**0.18**	**5.24**	**5.23**	**0.0166**
M05-2X/6−31+G**	this work	0.03	4.69	4.37	0.0508
M05-2X/6−311+G**	this work	0.10	4.64	4.58	0.0449
PBE/6−311+G//PBE/6−31G***	this work	**0.16**	**5.78**	**4.98**	**0.0158**
PBEsol/6−31+G**	this work	0.03	5.46	3.39	0.1708
PBEsol/6−311+G**	this work	0.13	5.54	3.62	0.1274
TPSS/6−311+G**//TPSS/6−31G*	this work	0.18	4.74	4.36	0.0558
PBE/6−31G*	this work	−0.73	−1.05	−4.83	99.9
TPSS/6−31G*	this work	−0.60	−1.23	−4.18	99.7
TPSSh/6−31G*	this work	−0.59	−0.29	−2.64	96.0

[a] All energies are indicated in kcal/mol. The best agreements with the MP2 reference are shown in bold.

are coming from basis set error, and basis set limit HF results are very poor due to serious underestimation of the intramolecular H-bond. Small basis sets overestimate the intramolecular H-bond strength. The success of the nonempirical PBE functional shows that it is particularly good for intramolecular H-bonds. Adding some exact exchange (PBEh) weakens the H-bond, and reducing the exchange-correlation gradient enhancement factor (PBEsol), strengthens the H-bond. The TPSS meta-GGA functional that was derived to mimic the large gradient behavior of PBE GGA also shows good performance, but for this special test set it is not as good as PBE.

In summary, for the GLC4 test the PBE functional shows a good agreement with the reference MP2/a-cc-pVTZ(-*f*) results. The order of the functionals, LMP2 and HF is the

following: PBE > LMP2 > TPSSh ≅ PBEh > B3PW91 > B3LYP ≅ TPSS ≅ PBEsol ≅ M05-2X ≫ HF (basis set limit).

4. Conclusions

The aim of this paper is to establish a protocol that can be used successfully and efficiently for modeling carbohydrates. In this paper, we have tested various DFT model chemistries on three sets of carbohydrate conformations. The ALL15 test set contains 13 $^4C_1$ structures (8 α- and 5 β-anomers, with gg, gt, and tg hydroxymethyl rotamers), and two $^1C_4$ conformers of α- and β-D-allopyranose. The AnGol15 test set included 15 conformers of 3,6-anhydro-4-*O*-methyl-D-galactitol, whereas the GLC4 contains two low energy $^4C_1$
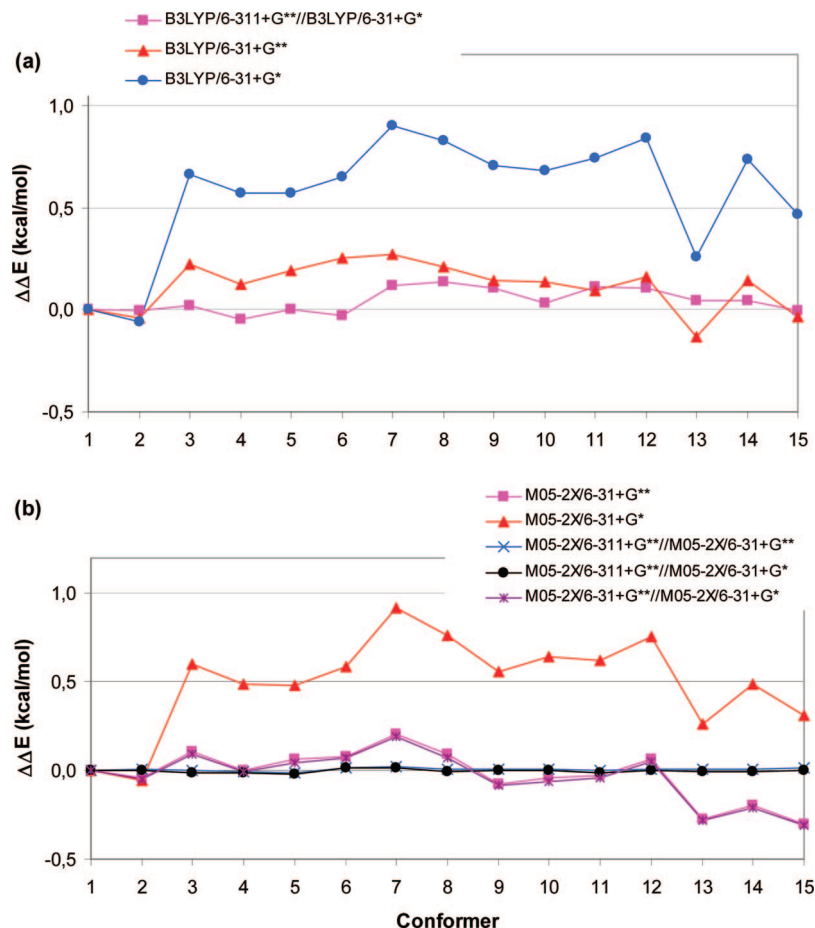
Figure 6. Differences in relative energies for the AnGol15 conformers (with respect to conformer 1), $\Delta\Delta E_{modelA-modelB}(conf_i, 1)$ of eq 7 optimized by various model chemistries. Model A is shown in the legends and (a) model B = B3LYP/6−311++G** and (b) model B = M05-2X/6−311+G**.

chair and two higher energy $^1C_4$ chair forms of β-D-glucopyranose. These test sets sample the lowest 6−7 kcal/mol energy range of conformation space. We have tested nonempirical functionals such as local spin density approximation (LSDA), generalized gradient approximation GGA (e.g., PBE, PBEsol), and meta-GGA (e.g., TPSS). We have also tested semiempirical hybrid functionals like PBEh, B3LYP, B3PW91, and a many fit-parameter empirical hybrid functional, M05-2X. Empirical functionals often perform very well on the test set used for fitting, and even outside the fitting set. However, the fitting set for M05-2X contains many small molecules, and there is no guarantee of good results on larger sugar molecules. We have chosen the MP2/a-cc-pVTZ(-f) model chemistry as reference and checked that by several MP2/CBS[3,4] calculations. The following conclusions can be drawn from our results:

1. The LMP2/cc-pVTZ(-f) and LMP2/aug-cc-pVTZ model chemistries give similar relative energies for the conformers in our test sets and those relative energies are in a reasonable (0.3 kcal/mol) agreement with MP2/a-cc-pVTZ(−f) and several MP2/CBS[3,4] results. The composite energies suggested earlier for the GLC4 test set show large 1.5 kcal/mol error compared to our new reference.

2. For B3CF (where CF = LYP or PW91) type hybrid functionals, the diffuse functions on hydrogen atoms

are not needed, but diffuse functions on heavy atoms and polarization functions on all atoms are essential for good relative energies. For geometry optimization, the 6−31+G* basis set is sufficient; thus, the B3CF/6−311+G**//B3CF/6−31+G* model chemistry gives converged relative energies with a mean absolute deviation of about 0.3 kcal/mol compared to MP2 reference results.

3. The B3CF/6−31+G**//B3CF/6−31+G* model chemistry gives relative energies with an error that is about half of the error of the functional. The quickest model chemistry that delivers useful results is B3CF/6−31+G**//B3CF/6−31G*.

4. Calculations with B3PW91 or PBEh give results similar to those with B3LYP, although β-anomers appear to be less stabilized.

5. For the ALL15 test set, the nonempirical GGA and meta-GGA and empirical M05-2X functionals show good agreement with the reference MP2 results. The order of the quality of the functionals, LMP2 and HF methods is the following: LMP2 ≫ M05-2X > PBE > TPSS > B3PW91 > PBEh > B3LYP ≫LSDA ≅ HF (basis set limit).

6. For the AnGol15 test set, M05-2X functionals shows a good agreement with the reference MP2 results. The order of the quality of the functionals, LMP2, and HF

Density Functionals and Basis Sets

J. Chem. Theory Comput., Vol. 5, No. 4, 2009 **691**

methods is the following: M05-2X > LMP2 > TPSS $\cong$ PBE > B3PW91 > B3LYP $\gg$ HF (basis set limit) > LSDA.

7. For the GLC4 test the order of the quality of the functionals, LMP2 and HF methods is the following: PBE > LMP2 > TPSSh $\cong$ PBEh > B3PW91 > B3LYP $\cong$ TPSS $\cong$ PBEsol $\cong$ M05-2X $\gg$ HF (basis set limit).

8. Calculations with HF give reasonable results with double-ζ basis sets (6−31G* and cc-pVDZ), but the HF limit (aug-cc-pVTZ) relative energies are clearly wrong, leading to serious underestimation of the stabilities of the β-anomers, $^1C_4$ chair conformations, and intramolecular H-bond strength. This result is in agreement with earlier observations.

9. Calculations with the M05-2X hybrid meta-GGA require a dense DFT grid for reliable results. The M05-2X results are better for the ALL15 and AnGol15 test sets than the results obtained by popular DFT functionals (B3LYP, B3PW91, or PBEh). For the GLC4 test set the PBE functional gives better results. Our preliminary results show that the M06-2X functional performs similarly to M05-2X for sugars, and it also requires a dense numerical integration grid for reliable results.

10. The GLC4 test set is particularly suitable for testing DFT methods because the relative energies are very sensitive to the correct description of the intramolecular H-bonds. Interestingly, non empirical functionals (particularly PBE) perform better on this test set than the empirical hybrid ones. The H-bond overestimation is decreasing in the following order: TPSS $\cong$ PBEsol $\cong$ M05-2X > TPSSh > PBE where PBE is just right. The underestimation is increasing in the following order: PBEh < B3PW91 < B3LYP $\ll$ HF (basis set limit). The results on this test set also show that small 6−31G* basis set overestimates intramolecular H-bonds with any method.

11. The popular B3LYP functional is not the best for saccharide conformational studies. The B3PW91 functional gives systematically better results, but all other hybrid functionals are even better. The M05-2X functional gives the best results.

12. We suggest to use M05-2X/6−311+G**//M05-2X/ 6−31+G* model chemistry for carbohydrate conformational space studies except for energy differences between $^1C_4$ and $^4C_1$ conformers where PBE/ 6−31+G** model chemistry performs better.

13. The nonempirical PBE GGA and TPSS meta-GGA functionals show a better overall performance than B3LYP.

Because the LMP2/cc-pVTZ and MP2/a-cc-pVTZ(-*f*) results do not always give consensus relative energies for the conformational space of the monosaccharides studied here, higher level calculations are needed for benchmarking DFT for the conformational space of carbohydrates. These infinite basis set extrapolation calculations are in progress.

Supporting Information Available: Geometries and/ or total energies of ALL15, AnGol15, and Glc4 test sets. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) French, A. D.; Brady, J. W. *ACS Symp. Ser.* **1990**, *430*, 1– 19.

(2) Pérez, S.; Imberty, A.; Engelsen, S. B.; Gruza, J.; Mazeau, K.; Jiménez-Barbero, J.; Poveda, A.; Espinosa, J. F.; van Eyck, B. P.; Johnson, G.; French, A. D.; Kouwijzer, M. L. C. E.; Grootenius, P. D. J.; Bernardi, A.; Raimondi, L.; Senderowitz, H.; Durier, V.; Vergoten, G.; Rasmussen, K. *Carbohydr. Res.* **1998**, *314*, 141–155.

(3) Imberty, A.; Pérez, S. *Chem. Rev.* **2000**, *100*, 4567–4588.

(4) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.

(5) Csonka, G. I.; Éliás, K.; Csizmadia, I. G. *J. Comput. Chem.* **1997**, *18*, 330–342.

(6) Csonka, G. I.; Angyán, J. G. *J. Mol. Struct. (THEOCHEM)* **1997**, *393*, 31–38.

(7) McNamara, J. P.; Muslim, A-M.; Abdel-Aal, H.; Wang, H.; Mohr, M.; Hillier, I. H.; Bryce, R. A. *Chem. Phys. Lett.* **2004**, *394*, 429–436.

(8) Barrows, S. E.; Dulles, F. J.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *Carbohydr. Res.* **1995**, *276*, 219–251.

(9) Csonka, G. I. *J. Mol. Struct. (THEOCHEM)* **2002**, *584*, 1– 4.

(10) Kaminsky, J.; Mata, R. A.; Werner, H. J.; Jensen, F. *Mol. Phys.* **2008**, *106*, 1899–1906.

(11) Pulay, P.; Saebo, S. *Theor. Chim. Acta* **1986**, *69*, 357–368.

(12) Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. *J. Chem. Phys.* **1995**, *103*, 1481–1490.

(13) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974–12980.

(14) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(15) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401/1–146401/4.

(16) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(17) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785– 789.

(18) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Kackson, K. A.; Pederson, M. A.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671–6687.

(19) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158– 6170.

(20) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029–5036.

(21) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(22) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(23) Csonka, G. I.; Elias, K.; Csizmadia, I. G. *Chem. Phys. Lett.* **1996**, *257*, 49–60.

(24) Lii, J.-H.; Ma, B.; Allinger, N. L. *J. Comput. Chem.* **1999**, *20*, 1593–1603.

(25) Momany, F. A.; Willett, J. L. *J. Comput. Chem.* **2000**, *21*, 1204–1219.

(26) Strati, G. L.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2002**, *337*, 1833–1849.

(27) Appell, M.; Strati, G.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2004**, *339*, 537–551.

(28) French, A. D.; Jonhson, G. P. *Cellulose* **2004**, *11*, 449–462.

(29) Hemmingsen, L.; Madsen, D. E.; Esbensen, A. L.; Olsen, L.; Engelsen, S. B. *Carbohydr. Res.* **2004**, *339*, 937–948.

(30) Momany, F. A.; Appell, M.; Strati, G.; Willett, J. L. *Carbohydr. Res.* **2004**, *339*, 553–567.

(31) Appell, M.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2005**, *340*, 459–468.

(32) French, A. D.; Jonhson, G. P.; Kelterer, A-M.; Csonka, G. I. *Tetrahedron: Asymm.* **2005**, *16*, 577–586.

(33) Momany, F. A.; Appell, M.; Willett, J. L.; Bosma, W. B. *Carbohydr. Res.* **2005**, *340*, 1638–1655.

(34) Stortz, C. A. *J. Comput. Chem.* **2005**, *26*, 471–483.

(35) French, A. D.; Jonhson, G. P. *Can. J. Chem.* **2006**, *84*, 603–612.

(36) Hricovini, M. *Carbohydr. Res.* **2006**, *341*, 2575–2580.

(37) Kurihara, Y.; Ueda, K. *Carbohydr. Res.* **2006**, *341*, 2565–2574.

(38) Momany, F. A.; Appell, M.; Willett, J. L.; Schnupf, U.; Bosma, W. B. *Carbohydr. Res.* **2006**, *341*, 525–537.

(39) Bagno, A.; Rastrelli, F.; Saielli, F. *J. Org. Chem.* **2007**, *72*, 7373–7381.

(40) Schnupf, U.; Willett, J. L.; Bosma, W. B.; Momany, F. A. *Carbohydr. Res.* **2007**, *342*, 196–216.

(41) Bohn, M. L.; Colombo, M. I.; Rúveda, E. A.; Stortz, C. A. *Org. Biomol. Chem.* **2008**, *6*, 554–561.

(42) Schnupf, U.; Willett, J. L.; Bosma, W. B.; Momany, F. A. *J. Comput. Chem.* **2008**, *29*, 1103–1112.

(43) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063–1079.

(44) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374–7383.

(45) Grimme, S. *Angew. Chem., Int. Ed.* **2006**, *45*, 4460–4464.

(46) Staroverov, V. N.; Scuseria, G. E.; Tao, J. M.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129–12137.

(47) Csonka, G. I.; Ruzsinszky, A.; Tao, J. M.; Perdew, J. P. *Int. J. Quantum Chem.* **2005**, *101*, 506–511.

(48) Paier, J.; Marsman, M.; Kresse, G. *J. Chem. Phys.* **2007**, *127*, 024103.

(49) Wodrich, M. D.; Corminboeuf, C.; Schleyer, P. V. *Org. Lett.* **2006**, *8*, 3631–3634.

(50) Grimme, S.; Steinmetz, M.; Korth, M. *J. Chem. Theory Comput.* **2007**, *3*, 42–45.

(51) Schreiner, P. R.; Fokin, A. A.; Pascal, R. A.; de Meijere, A. *Org. Lett.* **2006**, *8*, 3635–3638.

(52) Navarro, D. A.; Stortz, C. A. *Carbohydr. Res.* **2008**, *343*, 2292–2298.

(53) Ruzsinszky, A.; Csonka, G. I.; Perdew, J. P. *J. Phys. Chem. A.* **2005**, *109*, 11015–11021.

(54) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L.; Zhou, X.; Burke, K. *Phys. Rev. Lett.* **2008**, *100*, 136406.

(55) Csonka, G. I.; Ruzsinszky, A.; Perdew, J. P.; Grimme, S. *J. Chem. Theory Comput.* **2008**, *4*, 888–891.

(56) Zhao, Y.; Truhlar, D. G. *Org. Lett.* **2006**, *8*, 5753–5755.

(57) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.

(58) Csonka, G. I.; Nguyen, N. A.; Kolossváry, I. *J. Comput. Chem.* **1997**, *18*, 1534–1545.

(59) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(60) Tao, J.; Gori-Giorgi, P.; Perdew, J. P.; McWeeny, R. *Phys. Rev. A* **2001**, *63*, 325131–325135.

(61) For standard basis sets see http://www.emsl.pnl.gov/forms/basisform.html (accessed Aug 19, 2008).

(62) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; *Gaussian 03*, revision C.02, Gaussian, Inc., Wallingford CT, 2004.

(63) *Jaguar 6.0*, release 107; Schrodinger, LLC: Portland, OR, 2005.

(64) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

(65) Weldon, A. J.; Vickrey, T. L.; Tschumper, G. S. *J. Phys. Chem. A* **2005**, *109*, 11073–11079.

(66) Gould, I. R.; Bettley, H. A-A.; Bryce, R. A. *J. Comput. Chem.* **2007**, *28*, 1965–1973.

(67) Grimme, S.; Csonka, G. I. For CBS[3,4] extrapolation of HF relative energies, we used the HF/cc-pVnZ// B3LYP/6−31+G* ΔE values from Table 4, where n = T,Q. For MP2 correlation correction we used −6.81 and −6.45 kcal/mol for conformer 3, and −7.08 and −6.79 kcal/mol for conformer 4. Then we used $\Delta E[\infty] \approx \Delta E[Q] + c_{ext}(\Delta E[Q] - \Delta E[T])$, where for HF c_{ext} = 0.27; for MP2 c_{ext} = 0.730 = $1/((4/3)^3-1)$.

(68) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639–9646.

CT8004479