JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

# A Hierarchy of Methods for the Energetically Accurate Modeling of Isomerism in Monosaccharides

W. M. C. Sameera*,[†] and Dimitrios A. Pantazis*,[‡]

[†]Institut Català d'Investigació Química, Av. Països Catalans 16, 43007 Tarragona, Spain

[‡]Max-Planck-Institut für Bioanorganische Chemie, Stiftstrasse 34-36, 45470 Mülheim an der Ruhr, Germany

Supporting Information

**ABSTRACT:** The performance of different wave-function-based and density functional theory (DFT) methods was evaluated with respect to the prediction of relative energies for gas-phase monosaccharide isomers. A test set of 58 structures was employed, representing all forms of isomerism encountered in D-aldohexoses. The set was built from eight hexopyranose epimers by deriving subsets of isomers that include hydroxymethyl rotamers, anomers, ring conformers, furanose, and open-chain forms. Each subset of isomers spans a different energy range and involves various stereoelectronic effects. Reference energy values were obtained with coupled-cluster calculations extrapolated to the complete basis set limit, CCSD(T)/CBS. The tested CBS-extrapolated ab initio methods include various types of Møller−Plesset (MP) perturbation theory and the localized paired natural orbital coupled electron pair approach (LPNO-CEPA). Extensive benchmarking of DFT methods was carried out with 31 functionals. The results allow us to establish a hierarchy of methods that forms a reference guide for further computational studies. Among wave-function-based methods, LPNO-CEPA proved indistinguishable from CCSD(T), offering a promising alternative for a reference method that can be applied to larger systems. MP2 and SCS-MP2 follow closely, surpassing SOS-MP2 and MP3. The *m*PW2PLYP-D double hybrid and the Minnesota M06-2X hybrid meta-GGA are the best performing density functionals and are directly competitive with wave-function-based ab initio methods. Among the remaining functionals, B3PW91, TPSSh, *m*PW1PW91, and PBE0 yield the best results on average, while PBE is the best general-purpose GGA functional, surpassing meta-GGAs and several hybrids such as B3LYP. The choice of method strongly depends on the type of isomerism that needs to be considered, since many DFT methods perform well for purely conformational isomerism, but most of them fail to describe ring versus open-chain isomerism, where LYP-based GGA functionals perform particularly poorly.

## 1. INTRODUCTION

Carbohydrates are one of the most abundant classes of biomolecules on the earth and are involved in a wide spectrum of chemical and biological processes such as energy production and storage, structural maintenance, molecular recognition, cell growth, and defense.[1,2] Experimental and theoretical characterization of carbohydrate structures is critical as their conformational space correlates intimately with their properties and functions. Knowledge of structure−function relationships of these versatile biopolymers is also essential in order to modify their physical and chemical properties and improve their functional characteristics for numerous technological applications in the food, textile, paper, and cosmetics industries.[3]

A fundamental problem in the study of carbohydrates is the extent of conformational complexity, arising from the multitude of bonding possibilities for the primary and secondary hydroxyl groups, the ring puckering modes, and the rotational flexibility of the glycosidic linkages.[4−6] As a result, it is not straightforward to get a good spectroscopic handle on structural details of carbohydrate conformations. Over the past few years, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have provided significant insight into the carbohydrate conformational space,[5,7,8] but with drawbacks including possible mismatches between the static solid-state structures obtained by X-ray crystallography and the conformations in solution. On the other hand, the observed solution properties obtained by NMR methods are representative of all populated conformational states present in solution. Other spectroscopic methods like electron microscopy, light or neutron diffraction, and infrared spectroscopy have also provided useful but less accurate information for carbohydrate conformations.[8]

Owing to the complex structure−function paradigm of carbohydrates, the primary step of experimental or theoretical characterization of these biopolymers is to describe the detailed conformations of their monomer or dimer units.[9] Even for a simple monosaccharide such as glucose, the conformational space can be immense,[10] while solvent interactions pose additional challenges.[11] The stability of particular conformations depends on a complex balance of several interactions that are often of comparable magnitude, such as hydrogen bonds between the hydroxyl groups, noncovalent interactions, and a number of other effects that are collectively termed stereoelectronic.[12] The energetic effects of individual interactions within a monosaccharide may be small in absolute terms, but they are additive and eventually determine the properties of larger carbohydrate molecules. Analogously, the errors of a computational method are cumulative, and thus any application should rely on careful benchmarking to ensure reliability of the results.

**Received:** March 20, 2012
**Published:** July 20, 2012

  

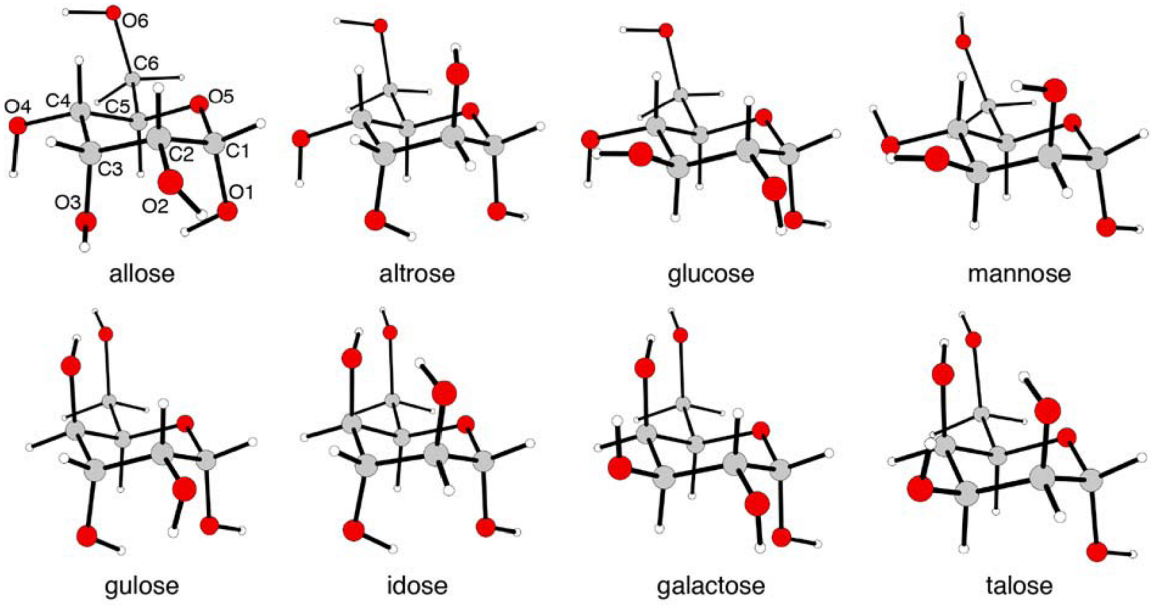


**Figure 1.** The eight $\alpha$-pyranose aldohexoses that form the basis of this work.

Computational chemistry offers a wide range of methods for the study of carbohydrate conformations, prediction of spectroscopic parameters, and calculation of energetics. Several commonly used semiempirical methods such as MNDO,[13] AM1,[14,15] and PM3[16,17] often fail to describe dispersion interactions and therefore are not satisfactory for obtaining accurate relative energies of carbohydrate conformations.[18−20] On the other hand, force field (FF) methods that incorporate van der Waals terms can provide a better description of noncovalent interactions in carbohydrates. Intensive progress has been made over the past few years with FF methods,[21] such as MM3,[22−25] CHARMM,[26−31] AMBER,[32,33] GLYCAM,[34−36] OPLS,[37,38] and GROMOS[39,40] in the study of carbohydrate conformations, with GROMOS, GLYCAM06, and MM3 being confirmed as good choices.[41] Although FF methods may not always be sufficiently accurate, in particular for anomeric effects,[42] they are still the method of choice for exploring the conformational landscape of large carbohydrate structures with thousands of atoms, especially through molecular dynamics (MD) studies.

Considerable effort is also focused on applying quantum chemical methods to the study of the relative stability of mono- and disaccharide conformations,[43−56] because they offer detailed insight into the stereoelectronic phenomena that determine carbohydrate structure. It is known that both Hartree−Fock (HF) and standard density functional theory (DFT) can describe hydrogen bonding with reasonable accuracy,[57] but they fail for interactions that are strongly dependent on dispersion forces. Several theoretical studies have been carried out for the prediction of noncovalent interactions in various systems such as DNA/RNA base pairs and trimers,[58−60] amino acid pairs,[61,62] oligopeptides,[63,64] and dimer configurations of small molecules.[65] These studies suggested that an accurate description of noncovalent interactions typically encountered in biomolecules[57,66,67] can be obtained with first-principles wave-function-based methods such as the coupled cluster theory using single, double, and perturbative triple excitations, CCSD(T), or methods based on Møller−Plesset perturbation theory. These calculations become quickly unfeasible as the computational cost rapidly increases with the size of the system. However, a number of approximations and new implementations such as dispersion corrections for density functionals[68−75] and symmetry adapted perturbation theory (DFT-SAPT),[76−78] for instance, have emerged as viable alternatives to wave-function-based ab initio techniques. Nevertheless, there is a growing area of overlap between DFT and wave-function-based methods. For even larger systems, quantum mechanics/molecular mechanics (QM/MM) approaches can be effective in describing nonbonded interactions, which are usually accounted for by the FF parametrization in the MM region, without relying on dispersion corrections to the QM energy.[79]

Aldohexose systems have been studied extensively over the years because they are prototype units for larger carbohydrates. Importantly, these monosaccharides incorporate almost all possible stereoelectronic effects and types of isomerism encountered in carbohydrates, with the exception of the glycosidic linkage. Recent studies have already provided significant information about the performance of modern computational approaches for the prediction of relative gas-phase energies of small saccharides.[55,56] In the present work, we attempt to provide a more extensive and comprehensive benchmark in terms of both the types of isomerism considered and the evaluated computational methods. To this end, we construct a test set of 58 structures starting from the eight aldohexose epimers (Figure 1) and present a detailed survey on the performance of 10 wave-function-based and 31 DFT approaches.

This work encompasses a number of generalized gradient approximation (GGA), meta-GGA, hybrid, hybrid-meta, and double-hybrid functionals with the goal of establishing a methodological hierarchy for the accurate prediction of relative energies in aldohexose conformers. At the same time, owing to both theoretical and algorithmic advances, highly accurate correlated ab initio calculations for small saccharides are also within reach, a trend that will continue in the near future. Therefore, special attention is devoted in this work to widely available and readily applicable correlated methods such as a


dx.doi.org/10.1021/ct3002305 | J. Chem. Theory Comput. 2012, 8, 2630−2645

number of Møller−Plesset perturbation theory levels and variants and local pair natural orbital (LPNO) techniques that can either provide benchmark-quality results for the present molecules or even serve as practical alternatives to DFT when higher accuracy is demanded. All methods were benchmarked against reference values obtained by CCSD(T) calculations extrapolated to the complete basis set limit (CBS). The present results provide detailed insight into the performance of ab initio and DFT methods for monosaccharides and form a useful guide for future studies of molecular carbohydrate systems.

## 2. METHODOLOGY

**2.1. Geometries.** Starting structures for the eight $\alpha$-D-aldohexoses (Figure 1) were obtained with MM geometry optimizations, using the MM3 force field as implemented in the Tinker 5.1 program package.[80] These starting structures were fully optimized with first-principles methods (*vide infra*) and used for the generation of a reference set of 58 geometries at the spin-component-scaled MP2 level (SCS-MP2).[81] No attempt is made to locate the absolute global minimum for each conformer with respect to rotations of the hydroxyl groups, since our interest is in comparisons between methods rather than in exploring the entire conformational space of aldohexoses. Of the 58 geometries, 24 structures form a subset of hydroxymethyl rotamers (three for each of the eight initially generated $\alpha$-D-aldohexose epimers), eight are $\beta$-anomers of the most stable $\alpha$-anomers included in the hydroxymethyl rotamers subset, eight are furanose forms of the aldohexoses, eight are open-chain forms, and 10 are different chair conformations of glucopyranose. The rotational isomers of each $\alpha$-pyranose were obtained as staggered forms of the hydroxymethyl group with respect to the O5−C5−C6−O6 torsion angle ($\omega$), where gas phase relaxed surface scans over the $\omega$ angle of the eight aldohexose sugars were carried out to find approximate stationary points for three important staggered rotamer conformations. Starting from these approximate stationary points, full optimizations were performed without any constraint. This leads to three conformations defined as gauche−gauche (*gg*), gauche−trans (*gt*), and trans−gauche (*tg*), depending on the position of the pendant hydroxyl group with respect to atoms O5 and C4 of the ring. Figure 2 provides a schematic description of hydroxymethyl rotational isomerism and an example of each of the other types of isomers for the case of the $\alpha$-$^4C_1$-glucopyranose of Figure 1.

For a given pyranose anomer, there are five hydroxyl and one hydroxymethyl rotational dihedral angle; the potential for three staggered conformations for each leads to $3^6$ = 729 different stereoisomers. This number grows to 2916 if the two anomeric and the two main pyranose chair forms are additionally considered,[10] still neglecting other forms of isomerism. Quantum mechanical treatment of all these isomers would involve excessive computational cost. Since the focus of the present work is on relative methodological comparisons, no conformational resampling of the 58 reference structures was carried out. Thus, the hydroxymethyl group is assumed to retain the same configuration when the stability of two anomers are compared, so that the relative energies used in the discussion better reflect the performance of the method regarding the quantification of the anomeric effect itself. This approach additionally ensures a sufficient scatter in the number of hydrogen bonds for each epimer that results in a better sampling of higher-energy conformations.

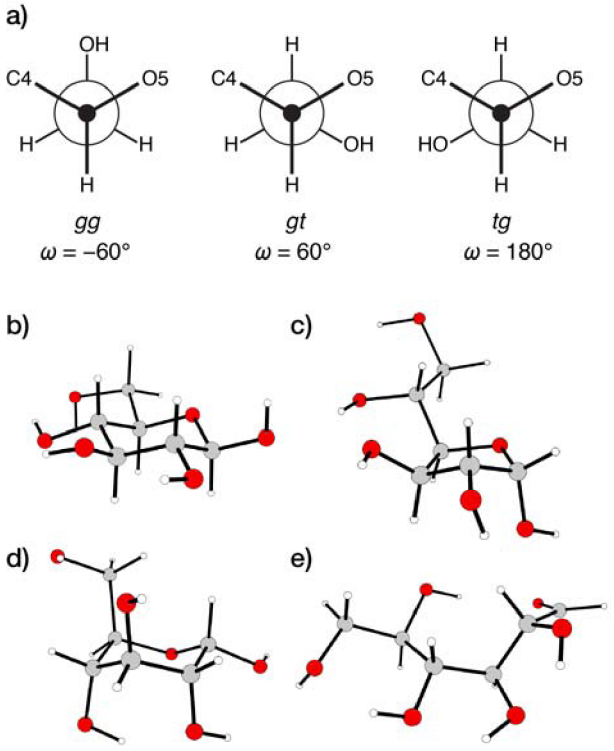


**Figure 2.** (a) Definition of rotational isomer designations and ideal $\omega$ angles, viewed along the C5−C6 bond. (b−e) Representative isomers of $\alpha$-$^4C_1$-glucopyranose: $\beta$-anomer, furanose form, $^1C_4$ chair conformer, and open-chain isomer.

The def2-TZVPP basis sets[82] were employed for all atoms, which in the present case have the contracted pattern 3s2p1d for H and 5s3p2d1f for C and O. Associated auxiliary basis sets (def2-TZVPP/C) were used for the resolution of the identity approximation to the evaluation of electron repulsion integrals (RI-MP2).[83−86] The geometries thus obtained were confirmed to be converged with respect to basis set expansion. These optimized structures were subsequently used for single-point calculations with other ab initio and DFT methods. All structures were also individually reoptimized with each density functional using the same basis sets as described above. Thus, the performance of DFT methods was assessed both with the reference and with functional-consistent geometries.

**2.2. Wave-Function-Based ab Initio Methods.** The current "gold standard" for organic molecules such as those treated here is undoubtedly the coupled cluster method with singles, doubles, and perturbative triples, extrapolated at the complete basis set limit, CCSD(T)/CBS. Therefore, single-point calculations were performed on the 58 optimized structures at the CCSD(T)/CBS level to derive the reference energies for the evaluation of all other methods. Most of the other methods tested in this study build upon the second order Møller−Plesset perturbation theory (MP2), a cost-effective and widely applied approach for incorporating correlation effects in wave function theory methods. Modifications of MP2 evaluated in the present study include the spin component scaled (SCS) approach of Grimme[81] and the scaled opposite spin (SOS) approach of Jung et al.[87] These MP2 variants distinguish between the correlation contributions from parallel and antiparallel pairs of electrons: in the case of SCS-MP2 different scaling factors are used for the parallel and antiparallel spin pairs (0.333 and 1.2, respectively), whereas in SOS-MP2 only

the antiparallel components are evaluated and scaled by a factor of 1.3. Extrapolated MP3 and SCS-MP3 calculations were also carried out,[88] but the cost and slow convergence of the MP series would hardly justify the effort of extending the series to higher orders for applications such as the present study. Pitoňák et al. have proposed an empirical variation that combines energy contributions from different levels of perturbation theory.[89] This approach, termed MP2.5, corresponds to scaled MP3 energies that are calculated as a sum of MP2/CBS and third-order energy contributions obtained with a smaller basis set with a third-order scaling factor of 0.5.

Going beyond these approaches, a modern implementation of the coupled-electron pair approximation (CEPA) was also evaluated. The recent interest[90] in CEPA methods stems from the high accuracy reported in benchmark studies,[91,92] which indicate that these methods surpass MP2 and popular DFT approaches, yielding results that are intermediate between CCSD and CCSD(T). A critical driving force for the application of CEPA methods nowadays has been the so-called local pair natural orbital approximation to single-reference correlation methods, which in the case of LPNO-CEPA produced results indistinguishable from the canonical calculations but with the computational cost reduced by up to 3 orders of magnitude.[93,94] In the present work, we employ the variant known as LPNO-CEPA/1, with the three default cutoff parameters that control the number of PNOs per electron pair ($T_{\mathrm{cutPNO}} = 3.33 \times 10^{-7}$), the perturbative selection of significant pairs ($T_{\mathrm{cutPairs}} = 10^{-4}$), and the number of contributing basis functions per PNO ($T_{\mathrm{cutMKN}} = 10^{-3}$), as implemented in the ORCA program system.[95]

Dunning's correlation-consistent cc-pV*n*Z basis sets were employed for all wave-function-based calculations. Two-point extrapolations to the CBS limit were performed individually for each energy component. Hartree−Fock energies were assumed to converge exponentially with increasing cardinal number; thus, the HF/CBS energies were obtained as

$$E_{\mathrm{HF}}^{\mathrm{CBS}} = \frac{E_{\mathrm{HF}}^{X} - E_{\mathrm{HF}}^{Y} c}{1 - c} \tag{1}$$

where $X$ and $Y$ are the highest angular momentum numbers in sequential correlation-consistent basis sets and $c = \exp \alpha(\sqrt{Y} - \sqrt{X})$. For the basis sets cc-pVTZ and cc-pVQZ that were employed in the present case, the value of 5.46 was used for the basis set dependent constant $\alpha$. Extrapolation of the Hartree−Fock energy is a minor effect since the energy obtained with the cc-pVQZ basis set is already close to the basis set limit, whereas extrapolation of the correlation energy is the dominant correction for correlated calculations. A cubic formula proposed by Helgaker and co-workers[96] was used in that case:

$$E_{\mathrm{corr}}^{\mathrm{CBS}} = \frac{X^3 E_{\mathrm{corr}}^{X} - Y^3 E_{\mathrm{corr}}^{Y}}{X^3 - Y^3} \tag{2}$$

This type of extrapolation was used for methods where the size of the problem allowed explicit calculations with cc-pV*n*Z basis sets with $n$ = T and Q (i.e., for LPNO-based and all MP2 methods). CBS energies for coupled-cluster and MP3 methods were obtained as projections over MP2/CBS values according to the procedure introduced by Hobza and co-workers,[97,98] with additional higher-order correlation corrections between CCSD(T) or MP3 and MP2 computed using the more tractable cc-pVDZ basis set. Similar to the SCS-MP2 optimizations, the resolution of the identity technique was used for all MP2 variants.[83−86]

**2.3. DFT Methods.** It is expected that the choice of method for relatively larger carbohydrates than the monosaccharides included in the present benchmark study will be guided by computational cost as much as by accuracy. For this reason, it is important to identify the best performing methods within each category of density functionals. Therefore, we attempted to evaluate representatives of widely available families of density functionals that typically correspond to different cost/performance ratios. GGA functionals include BP86,[99,100] PBE,[101] the modified version of PBE, revPBE, proposed by Zhang and Yang,[102] as well as the *m*PWPW91[103] functional incorporating the modification of the Perdew−Wang exchange by Adamo and Barone with the PW91 gradient-corrected correlation functional.[104,105] Grimme's B97D[71] functional that includes a semiempirical dispersion correction is also considered. Tested GGA functionals that use the Lee−Yang−Parr (LYP)[106] correlation are BLYP,[99] GLYP with Gill's exchange functional,[107] OLYP with Handy's two-parameter OPTX functional,[108] XLYP using the nonhybrid version of the Xu and Goddard exchange,[109] and mPWLYP.[103] TPSS[110] and M06-L[111] were used as representative meta-GGA functionals. Hybrid functionals include B3LYP,[112] the one-parameter B1LYP,[113] the X3LYP functional of Xu and Goddard,[109] as well as the B3PW91, PBE0,[114] *m*PW1PW91, and *m*PW1LYP functionals.[103] Three hybrid meta-GGA functionals were also evaluated, the TPSSh[115] and two functionals from the Minnesota family,[116] M06[117] and M06-2X.[117] Finally, the assessment included the double hybrid functionals B2PLYP[118] and *m*PW2PLYP[119] proposed by Grimme, which combine exact exchange with a second-order perturbation correction for nonlocal correlation effects.

Multiple approaches have been proposed to address the issue of dispersion corrections to DFT, ranging from explicit reconstruction of the functional form to more or less simple corrections of existing functionals based either on the density or on empirical parameters.[68−75,120,121] Because of its simplicity, in the present work we employed the pragmatic approach proposed by Grimme, which is based on atomic pairwise additive corrections and has been shown to be an efficient and cost-free way to include dispersion on top of standard DFT for large molecules.[122] Several functionals were tested for the effect of these corrections using the latest variant of the method,[72] designated here with the suffix "-D" after the name of the functional. Functionals tested in this way are the GGAs BP86-D, BLYP-D, and PBE-D; the meta-GGA TPSS-D; the hybrid B3LYP-D; and the double hybrids B2PLYP-D and *m*PW2PLYP-D.

The def2-TZVPP basis sets[82] were used for all DFT calculations, since they are big enough to ensure converged results and to exclude the possibility of BSSE and/or error compensation. Even though DFT calculations tend to converge fast with basis set size, using smaller basis sets should only be done after careful testing to ensure that the computed property does not deteriorate upon basis set enlargement, a clear indication of unpredictable error cancellations.[123] Since our goal in this work is to evaluate the functionals in terms of their intrinsic performance, we prefer to avoid potential complications arising from the use of smaller basis sets. The RI and RIJCOSX approximations (where COSX stands for the "chain of spheres" approximation to exact exchange)[124] as implemented in ORCA[95] were used for GGA and hybrid DFT

calculations, respectively, with auxiliary def2-TZVPP/J basis sets. The Gaussian09 program[125] was used for the B97D and Minnesota density functionals.

## 3. RESULTS AND DISCUSSION

**3.1. Reference Results.** In the following, the CCSD(T)/CBS relative energies for the SCS-MP2 optimized structures are presented and grouped according to the type of isomerism probed in each subset. Optimized Cartesian coordinates of all structures are provided as Supporting Information. Table 1 summarizes the relative energies obtained for the 24 isomers of the $\alpha$-pyranose aldohexoses (Figure S1) that result from the rotation of the hydroxymethyl group around the O5−C5−C6−O6 torsion angle $\omega$. As noted above, the hydroxyl rotational angles have not been scanned to find the absolute gas-phase minimum in each case, and therefore adopted the orientations shown in Figure 1. Reported energies are relative to the most stable conformer of the present subset, *gt*-allose.

**Table 1. CCSD(T)/CBS Relative Energies (kJ mol$^{-1}$) of Hydroxymethyl Rotational Isomers of the $\alpha$-Pyranose Aldohexoses**

| rotamer | $\Delta E$ | rotamer | $\Delta E$ |
|---|---|---|---|
| *gg*-allose | 8.16 | *gg*-gulose | 11.50 |
| *gt*-allose | 0.00 | *gt*-gulose | 13.23 |
| *tg*-allose | 2.06 | *tg*-gulose | 13.14 |
| *gg*-altrose | 8.78 | *gg*-idose | 5.60 |
| *gt*-altrose | 3.55 | *gt*-idose | 11.23 |
| *tg*-altrose | 3.71 | *tg*-idose | 11.57 |
| *gg*-glucose | 23.43 | *gg*-galactose | 14.84 |
| *gt*-glucose | 14.85 | *gt*-galactose | 17.01 |
| *tg*-glucose | 11.40 | *tg*-galactose | 17.33 |
| *gg*-mannose | 10.00 | *gg*-talose | 1.84 |
| *gt*-mannose | 6.32 | *gt*-talose | 27.36 |
| *tg*-mannose | 2.61 | *tg*-talose | 6.47 |

The relative populations of the rotamer conformations, *gg*, *gt*, and *tg*, are controlled by the tendency of a molecule to adopt the structure that has the maximum number of *gauche* interactions between adjacent electron pairs and polar bonds, the so-called *gauche* effect.[126] Experiment and theory provide a range of rotamer ratios for each monosaccharide that depend on the physical state and the solvent effects.[127,128] Relative energies of different rotamers are mostly sensitive to the distribution of intramolecular hydrogen bonds. Given that the individual OH group orientations were not explicitly scanned to maximize intramolecular hydrogen bonding in each epimer, the results are taken as internal reference and not interpreted in terms of experimentally comparable rotamer populations.

The anomeric effect is the best-known stereoelectronic effect in carbohydrates and refers to the preference for the axial ($\alpha$-anomer) instead of the sterically less hindered equatorial ($\beta$-anomer) orientation of an electronegative substituent at the anomeric carbon C1 of a pyranose ring.[129] Explanations for this phenomenon include repulsive dipole−dipole interactions in the equatorial orientation[130] and stabilizing hyperconjugative delocalization of the endocyclic O lone pair into the $\sigma^*$ C1−O1 orbital in the axial orientation,[131] but other interpretations have been advanced.[132−134] In Table 2, we report the relative energies of the $\beta$-anomers (Figure 3) with respect to the most stable $\alpha$-rotamer of each epimer reported in Table 1. The orientation of the hydroxymethyl group was the same as that for the respective $\alpha$-anomer. The OH groups on carbons 2−4 were optimized starting from the $\alpha$-anomer conformations without explicit rotational scans. It is important to note that the actual prediction of anomeric ratios is not merely a question of energy differences between static minima but a problem of population distributions within large conformational ensembles. This presents significant challenges for sampling of all the low-energy conformations and rotamers that may contribute to the ensemble energy. The recent molecular dynamics studies of Momany and co-workers are relevant in this respect.[53,135−137]

**Table 2. CCSD(T) Relative Energies (kJ mol$^{-1}$) of $\beta$-Anomers, Furanose Rings, and Open Chain Isomers with Respect to the Most Stable $\alpha$-Pyranose Rotamer**

| | $\beta$-anomers | furanoses | open chains |
|---|---|---|---|
| allose | −1.38 | 15.58 | 49.81 |
| altrose | 15.74 | 25.77 | 34.09 |
| glucose | 2.58 | 5.59 | 41.32 |
| mannose | 1.04 | 15.85 | 50.83 |
| gulose | 24.03 | 15.26 | 38.01 |
| idose | 9.84 | 39.70 | 38.87 |
| galactose | 7.52 | 13.26 | 37.39 |
| talose | −1.07 | 28.89 | 46.50 |

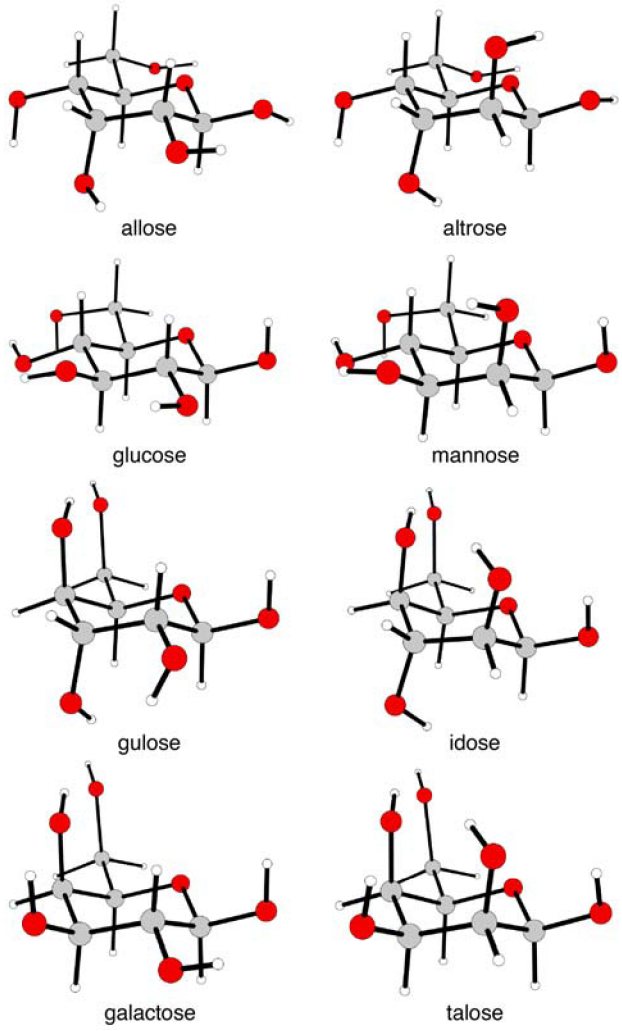


**Figure 3.** Optimized structures of the $\beta$-anomers of the most stable $\alpha$-pyranose hydroxymethyl rotamers.

Five-membered furanose rings are in general disfavored energetically compared to six-membered pyranoses,[12] and this is reproduced by the present CCSD(T)/CBS results for the $\alpha$-furanose structures included in the test set (Table 2). The eight optimized structures shown in Figure 4, which are not

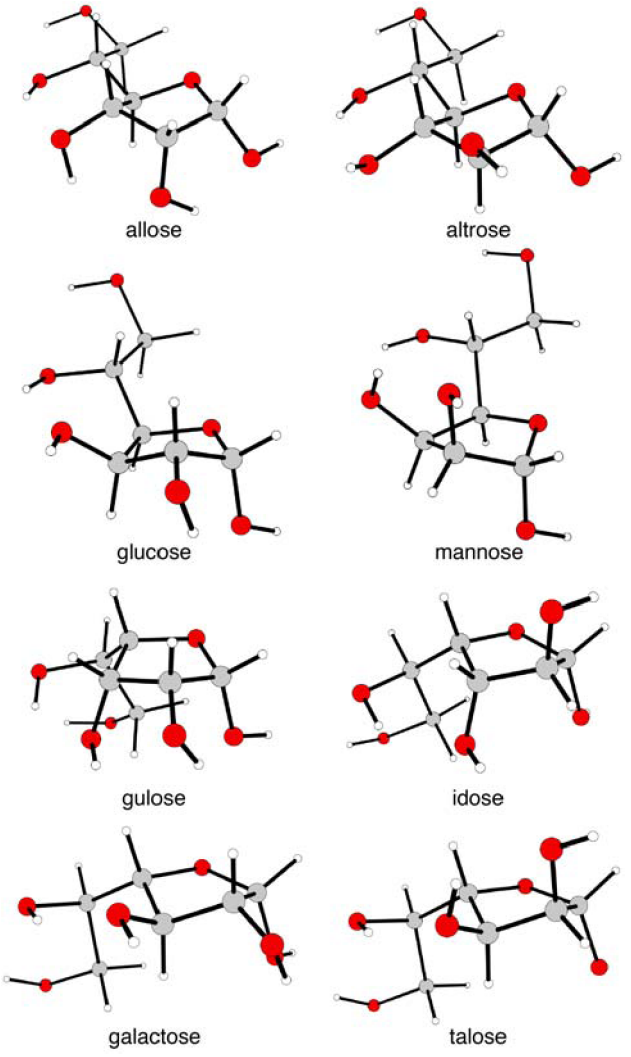


**Figure 4.** Optimized structures of eight examples of $\alpha$-furanose forms of the eight aldohexose epimers.

necessarily representative of low-lying global minima, adopt a nonplanar geometry for the tetrahydrofuran ring, with a twisted configuration of carbons C2 and C3. The least energetically destabilized five-membered ring corresponds to glucose, only 6 kJ mol$^{-1}$ higher than the most stable $\alpha$-glucopyranose rotamer included in the test set. By contrast, the particular conformer of $\alpha$-idofuranose included in our set is almost 40 kJ mol$^{-1}$ higher in energy than the most stable $\alpha$-*gg*-idopyranose, owing to the complete loss of all intramolecular hydrogen bonding in the furanose isomer and the most unfavorable quasi-staggered conformation of the O2 and O3 hydroxyl groups.

The largest energy differences of the present test set are observed for the open chain forms of the aldohexoses. Compared with the energetically most stable $\alpha$-pyranose rotamers of Table 1, the examples of acyclic aldehyde forms of all epimers shown in Figure 5 are computed to be more than 30 kJ mol$^{-1}$ higher in energy, with half of them exceeding 40 kJ

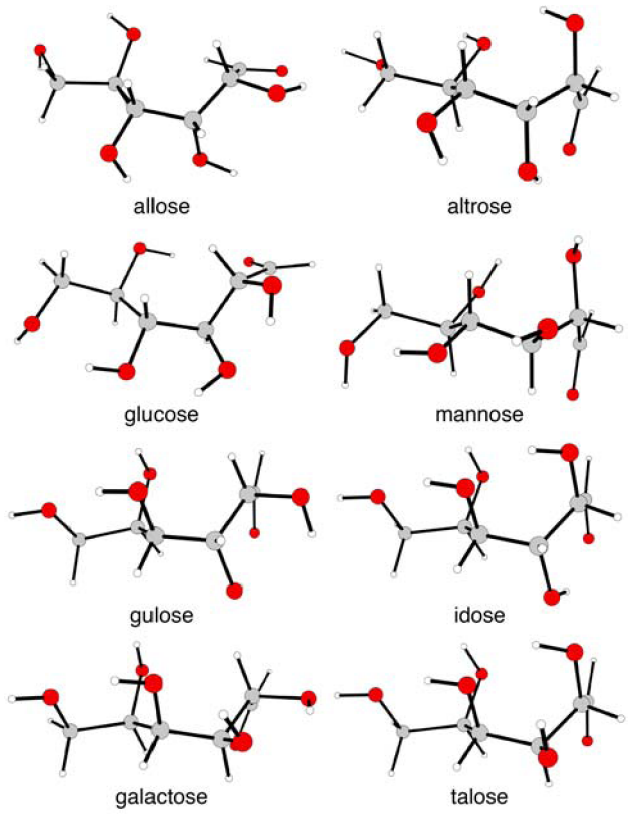


**Figure 5.** Optimized structures of open-chain forms of the eight aldohexose epimers.

mol$^{-1}$. This is the most demanding subset of our test set, not simply because the energy differences are larger but because accurate prediction of energetics in this case depends on the ability of the method to deal with a substantially different balance of stereoelectronic effects in terms of medium-range correlation (*vide infra*).

The final type of isomerism considered in this study arises from the different conformations of the pyranose ring. In principle, 38 distinct conformations are possible: two chair (C), six boat (B), six skew-boat (S), 12 half-chair (H), and 12 envelope (E) conformations.[138] The interconversion pathways between these ring topologies are visualized in terms of a planar Stoddart diagram or by a Cremer–Pople spherical mapping[139,140] where the ring-flipped $^4C_1$ and $^1C_4$ conformations occupy the two poles (superscript and subscript numerals denote the two atoms that lie above and below a four-atom reference plane). In an idealized system, the C and S conformers are minima, the B and H conformers are transition states, and the E conformers correspond to second order saddle points. However, in practice these assignments depend on the particular system. Without the intention to locate absolute global minima in each case, the present study focused on optimized sample $^4C_1$, $^1C_4$, $^1S_5$, $^OS_2$, and $^1S_3$ conformers of $\alpha$-glucopyranose and $^4C_1$, $^1C_4$, $^1S_3$, $^5S_1$, and $^1S_5$ conformers of its $\beta$-anomer (Figure 6), assigning each S minimum to the nearest canonical conformation based on the ring dihedral that is closest to zero degrees. It proved convenient for conformational searching if all the hydrogen bonds in the $^4C_1$ ring were arranged in a counter-clockwise manner, so for this subset the energy differences are referenced to the so-called *r* orientation of the OH groups in $\alpha$-$^4C_1$-glucopyranose. The relative energies of the ring conformers are compared in Table 3. As anticipated,


dx.doi.org/10.1021/ct3002305 | J. Chem. Theory Comput. 2012, 8, 2630–2645

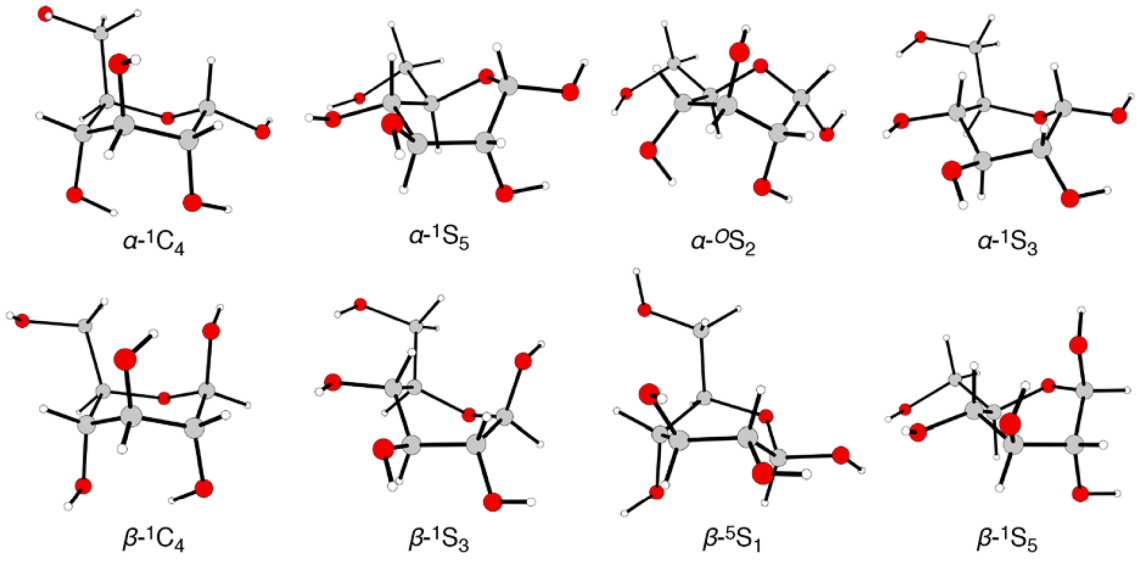


**Figure 6.** The $^{1}C_{4}$ chairs and the six skew-boat glucopyranose conformers included in this study.

**Table 3. CCSD(T) Relative Energies (kJ mol$^{-1}$) of 10 Ring Conformers of Glucopyranose**

| conformer | $\Delta E$ | conformer | $\Delta E$ |
|---|---|---|---|
| $\alpha$-$^{4}C_{1}$ | 0.0 | $\beta$-$^{4}C_{1}$ | 4.52 |
| $\alpha$-$^{1}C_{4}$ | 28.41 | $\beta$-$^{1}C_{4}$ | 32.00 |
| $\alpha$-$^{1}S_{5}$ | 35.64 | $\beta$-$^{1}S_{3}$ | 33.55 |
| $\alpha$-$^{O}S_{2}$ | 22.70 | $\beta$-$^{5}S_{1}$ | 50.27 |
| $\alpha$-$^{1}S_{3}$ | 37.89 | $\beta$-$^{1}S_{5}$ | 34.48 |

the two anomeric $^{4}C_{1}$ rings are the most stable, with the other conformations lying well above 20 kJ mol$^{-1}$ and reaching over 50 kJ mol$^{-1}$ in the case of $\beta$-$^{5}S_{1}$. We note that the $\alpha$-$^{1}S_{5}$, $\alpha$-$^{O}S_{2}$, and $\beta$-$^{1}S_{5}$ minima were located on or close to the trajectories for $^{1}C_{4}/^{4}C_{1}$ ring inversion of the two penta-*O*-methyl-substituted D-glucopyranose anomers in an ab initio molecular dynamics study by Ionescu et al.[141]

The CCSD(T)/CBS relative energies of the 58 different structures indicate that the $\alpha$-pyranose rotamers and their $\beta$ anomeric counterparts are within the most stable forms and include some of the smallest energy differences observed between isomers. Furanose forms are uniformly higher in energy than pyranose forms and in general sample higher energy differences than those observed in the anomers subset. Different ring conformations, in particular the skew-boat conformers considered here, are significantly higher in energy than the chair forms, while the open-chain isomers are the highest in energy. Furthermore, open-chain isomers prove to be critical for the present study because their relative energies with respect to the pyranose forms are the most challenging to reproduce. As will be emphasized in the following sections, the maximum errors of all methods are associated with the prediction of the relative stability of open chain forms, while several DFT methods face unexpected problems even with the isomerism between the pyranose and furanose forms of the aldohexoses.

**3.2. Wave-Function-Based ab Initio Methods.** Given the different balance of stereoelectronic effects that enter into the computation of relative energies between each type of isomer, the results for all evaluated methods are presented with respect to the five isomer types that constitute the test set: hydroxymethyl rotamers, anomers, ring conformers, furanose isomers, and open-chain forms. For clarity of presentation, only root-mean-squared (RMS) data obtained for each subset are

**Table 4. Average Root-Mean-Squared (RMS) Errors (kJ mol$^{-1}$) of Wave-Function-Based ab Initio Methods for Each Type of Isomerism with Respect to the CCSD(T)/CBS Reference Values[a]**

| method | rotamers | $\beta$-anomers | ring conformers | furanoses | open chains | avg. |
|---|---|---|---|---|---|---|
| LPNO-CEPA | 0.72 | 0.43 | 0.53 | 0.75 | 1.10 | 0.71 |
| MP2 | 0.36 | 0.48 | 0.85 | 1.14 | 3.50 | 1.27 |
| SCS-MP2 | 1.20 | 0.94 | 1.31 | 1.80 | 2.52 | 1.55 |
| CCSD | 1.29 | 1.12 | 1.28 | 1.50 | 2.92 | 1.62 |
| LPNO-CCSD | 1.31 | 0.96 | 1.07 | 1.33 | 3.56 | 1.65 |
| SCS-MP3 | 1.61 | 1.28 | 1.57 | 2.41 | 1.45 | 1.66 |
| MP2.5 | 0.79 | 0.83 | 1.11 | 0.93 | 6.66 | 2.06 |
| SOS-MP2 | 1.84 | 1.37 | 1.68 | 3.04 | 5.41 | 2.67 |
| MP3 | 1.58 | 1.57 | 1.71 | 1.88 | 9.85 | 3.32 |
| HF | 5.69 | 5.25 | 7.00 | 8.74 | 6.75 | 6.69 |
| HF/cc-pVDZ | 2.91 | 2.89 | 3.97 | 2.70 | 11.09 | 4.71 |
| MP2/cc-pVDZ | 4.16 | 4.94 | 6.38 | 7.48 | 12.15 | 7.02 |
| CCSD(T)/cc-pVDZ | 4.09 | 4.79 | 6.42 | 6.86 | 15.48 | 7.53 |

[a]All methods are CBS-extrapolated, except where the basis set is explicitly given.

discussed, along with the total average RMS error over all isomer subsets. From the results summarized in Table 4, it is seen that the errors and energy differences among all correlated CBS-extrapolated methods are quite small for most types of isomerism. In general, there are few "poor choices" among these methods for the study of conformational isomerism, and these only arise in the treatment of acyclic forms.

An important result of this benchmark study relates to what is evidently the best performing approach, LPNO-CEPA, with an average error of only 0.7 kJ $mol^{-1}$ over all types of isomerism and a particularly low error for the challenging case of acyclic isomers. Previous studies have demonstrated that CEPA is more accurate than CCSD and slightly less accurate than CCSD(T).[90−93] It has been suggested that the systematic overshooting of the CCSD correlation energy by CEPA can be interpreted as partially including the effects of connected triples excitations.[90] The above observations are in line with the results of the present test set, where no other method comes as close to the CCSD(T) reference.

Regarding the coupled cluster methods, the results indicate that the missing perturbative triples correction in CCSD introduces deviations of 1 kJ $mol^{-1}$ with respect to the CCSD(T) reference for all isomerism types studied. It should also be noted that QCISD(T), not shown in Table 1, is practically indistinguishable from CCSD(T) for the present test set. The use of the LPNO technique in combination with CCSD does not cause any deterioration of the canonical CCSD results, in line with previous observations in benchmark studies.[142] In fact, LPNO-CCSD shows slightly smaller errors than CCSD for all types of isomerism except for the relative energies of the open-chain forms, a result that is particularly encouraging for the application of the LPNO approximation.

Focusing on the Møller−Plesset methods and variants, it is obvious that, in terms of average errors over all forms of isomerism considered, the best performance is obtained with MP2. SCS-MP2 appears only marginally less accurate than MP2 but shows improved performance over MP2 for the relative stabilities of open-chain forms. To investigate whether this particular case is related to the geometry of the open-chain forms, which were optimized with SCS-MP2, the relative energies for the case of open-chain structures were recalculated with MP2-optimized structures. This did not improve the MP2 RMS error with respect to SCS-MP2 (3.36 and 2.24 kJ $mol^{-1}$, respectively), confirming that the superior SCS-MP2 results in this case are intrinsic to the treatment of correlation. It is noted that the two methods exhibit errors of different signs for the relative energies of the acyclic species, where MP2 systematically overestimates and SCS-MP2 systematically underestimates the energy difference with respect to the pyranose forms. SOS-MP2 is less accurate than MP2 or SCS-MP2 for all types of isomerism and shows from larger errors in the cases of furanose and open-chain vs pyranose relative energies. In the latter case, SOS-MP2 uniformly underestimates the energy difference by as much as 7.8 kJ $mol^{-1}$. A probable explanation for this behavior is that SOS-MP2 does not fully cover long-range correlation, a weakness that has been recently addressed with the introduction of a distance-dependent scaling factor.[143] We note that similar conclusions regarding the relative performance of SOS-MP2 compared to SCS-MP2 were recently reached by Goerigk and Grimme on a more extended data set.[144]

A rather unexpected behavior for the complexes under consideration was observed with the MP3 method, where the deterioration observed on going from the second to the third MP order of perturbation suggests that the extra computational cost (*vide infra*) does not translate into greater accuracy. The most critical problem appears again in the ring vs open-chain isomerism, where the MP3 RMS error is more than twice as large as the MP2 RMS error, an effect that is presumably related to the oscillatory behavior of the Møller−Plesset series. Inspection of the computed energies shows that this arises from overestimation of the stability of the cyclic forms by as much as 11 kJ $mol^{-1}$ compared with the CCSD(T) reference. The above deficiency is corrected in the SCS variant of MP3,[88] which reduces this particular error by roughly an order of magnitude. The poor performance of MP3 for the relative energies of open-chain species carries over into the composite MP2.5 method. In a study of interaction energies in noncovalently bound nucleic acid base pairs, the performance of the MP2.5 method was found to be superior to SCS-MP2, achieving a more balanced treatment of hydrogen-bonded compared to stacked complexes,[89] but in the present case it appears that MP2.5 benefits more from the good performance of MP2 than from the additional scaled contribution of MP3.

A note concerning the dependence of accuracy on the size of the basis set is in order at this point. It should be kept in mind that all energies obtained by ab initio methods in this work are extrapolated to the CBS limit. The evaluation of these correlated methods clearly does not stand when small basis sets (e.g., smaller than polarized triple-$\zeta$) are employed without extrapolation. To demonstrate this point, Table 4 includes the RMS errors of HF, CCSD(T), and MP2 obtained with polarized double-$\zeta$ (cc-pVDZ) basis sets. With the exception of the open-chain isomers, the HF/cc-pVDZ errors are smaller in this case than the HF/CBS values, in line with earlier observations[19,55,56] that a polarized double-$\zeta$ basis set can yield good results at an uncorrelated level. On the other hand, for correlated methods the errors are significantly larger than those of the CBS results, confirming that the use of such methods with small basis sets is of limited value.[145] The small basis set also results in outliers, for example the relative stabilities of *gg*/*tg* hydroxymethyl rotamers of allose and altrose are reversed at the CCSD(T) and MP2 levels. The inconsistent treatment of correlation with small basis sets presumably introduces larger errors in relative energies than the correlation energy recovered in absolute terms. Improved results can be obtained only by considering the effect of a larger basis set within an extrapolation scheme, and thus the use of methods such as CCSD(T) exclusively with a small basis set can lead to misleading results. Furthermore, as will be shown in the following section, the errors of these otherwise expensive methods when combined with small basis sets are larger than those of DFT approaches, negating most practical uses for such nominally correlated results.

A crucial consideration in the application of all these methods is the associated computational cost. To give an idea of the time requirements and the scaling of each method, in Table 5 we compare representative timings for selected methods included in the present study. Three basis sets have been used for single-point calculations on a glucopyranose structure: cc-pVDZ (228 basis functions), cc-pVTZ (528 basis functions), and cc-pVQZ (1020 basis functions). To facilitate comparisons, all calculations were performed on a single processor (see Supporting Information for technical details). The timings of all post-Hartree−Fock methods include the Hartree−Fock SCF component. From Table 5, it is obvious that the overhead for the correlation part of the calculation is

**Table 5. Timings (min) of Selected Methods with Three Different Basis Sets for a Single-Point Calculation on α-D-Glucopyranose**

| method | cc-pVDZ | cc-pVTZ | cc-pVQZ |
|---|---|---|---|
| HF | 5 | 61 | 613 |
| MP2 | 7 | 92 | 958 |
| LPNO-CEPA | 10 | 105 | 1036 |
| LPNO-CCSD | 15 | 148 | 1515 |
| MP3 | 22 | 372 | 4452 |
| CCSD | 155 | 2846 | |
| CCSD(T) | 428 | | |

still not significant for systems of the size considered here in the case of MP2 methods. MP3 is more expensive and scales more steeply than MP2; as noted above, given the performance noticed for the present test set, the additional cost cannot be justified. Coupled cluster methods are significantly more expensive than other methods for a given basis set, while the extremely unfavorable scaling not only in terms of time but also in memory requirements prohibits calculations after a certain problem size. LPNO-CEPA is more expensive but still competitive with MP2 in terms of both absolute computational cost and scaling. The benefit of LPNO in terms of cost is apparent in the comparison between CCSD and LPNO-CCSD. Given that systems with a few thousand basis functions can currently be treated with the LPNO approach, there is clear potential for methods such as LPNO-CEPA to be used as benchmark-quality references for larger carbohydrates or even as routine target-level theory in high-accuracy applications.

In summary, among the methods included in the study, LPNO-CEPA provides the most accurate relative energies at a reasonable cost, and therefore it is strongly recommended for highly accurate results on carbohydrate systems. LPNO-CEPA closely matches CCSD(T) for all types of isomerism at a significantly smaller computational cost. It is worth exploring whether LPNO-CEPA could replace CCSD(T) as a reference method in future studies of larger systems. MP2 and SCS-MP2 are the most obvious alternative choices, supporting the use of CBS-extrapolated MP2 in benchmark studies.[56] Significant errors with ab initio methods only appear in the comparison of pyranose vs acyclic forms, whereas pyranose−furanose isomerism presents no significant problem for most of the methods except SOS-MP2. Therefore, special care is to be taken in the study of reactions that involve ring-opening or ring formation. Regardless of the choice of method, the use of complete basis set extrapolation is indispensable for obtaining meaningful results; the only foreseeable development that could allow this absolute requirement to be circumvented would be the advent of efficient and practical explicitly correlated methods.

**Table 6. Root-Mean-Squared (RMS) Errors (kJ mol$^{-1}$) of DFT Methods for Each Type of Isomerism with Respect to the CCSD(T)/CBS Reference Values Using the Reference SCS-MP2 Geometries[a]**

| functional | rotamers | β-anomers | ring conformers | furanoses | open-chains | Avg1 | Avg2 |
|---|---|---|---|---|---|---|---|
| *m*PW2PLYP-D | 0.77 | 0.67 | 1.26 | 1.63 | 5.75 | 2.02 | 1.08 |
| M06-2X | 0.98 | 0.75 | 1.13 | 2.39 | 4.89 | 2.03 | 1.31 |
| *m*PW2PLYP | 0.61 | 0.70 | 1.39 | 2.07 | 5.72 | 2.10 | 1.19 |
| M06 | 0.94 | 2.12 | 1.46 | 1.54 | 6.24 | 2.46 | 1.52 |
| B2PLYP | 0.65 | 0.75 | 1.29 | 2.78 | 8.41 | 2.78 | 1.37 |
| B3PW91 | 1.60 | 2.32 | 2.10 | 5.71 | 2.50 | 2.85 | 2.93 |
| B2PLYP-D | 1.13 | 1.07 | 1.73 | 2.18 | 8.52 | 2.93 | 1.53 |
| TPSSh | 2.23 | 1.85 | 1.16 | 6.53 | 4.29 | 3.22 | 2.95 |
| *m*PW1PW91 | 1.48 | 2.25 | 1.89 | 4.76 | 6.04 | 3.29 | 2.60 |
| PBE0 | 1.57 | 2.38 | 1.71 | 4.17 | 8.24 | 3.62 | 2.46 |
| PBE | 2.98 | 3.21 | 1.27 | 5.90 | 7.30 | 4.13 | 3.34 |
| PBE-D | 4.07 | 3.72 | 3.18 | 2.98 | 7.75 | 4.34 | 3.49 |
| X3LYP | 1.25 | 1.75 | 2.27 | 5.13 | 11.50 | 4.38 | 2.60 |
| BP86 | 2.78 | 2.90 | 1.33 | 5.70 | 9.80 | 4.50 | 3.18 |
| *m*PW1LYP | 1.37 | 1.69 | 2.34 | 4.98 | 12.28 | 4.53 | 2.60 |
| *m*PWPW91 | 2.79 | 3.00 | 1.42 | 6.56 | 9.58 | 4.67 | 3.45 |
| TPSS | 2.93 | 2.29 | 1.60 | 7.26 | 9.85 | 4.78 | 3.52 |
| B3LYP-D | 1.84 | 1.83 | 2.89 | 3.63 | 13.73 | 4.79 | 2.55 |
| B3LYP | 1.43 | 1.91 | 2.46 | 6.02 | 13.50 | 5.06 | 2.96 |
| B1LYP | 1.62 | 1.94 | 2.66 | 5.96 | 13.27 | 5.09 | 3.04 |
| BP86-D | 4.48 | 3.96 | 4.37 | 4.46 | 10.40 | 5.53 | 4.32 |
| TPSS-D | 4.74 | 3.89 | 5.18 | 3.86 | 10.48 | 5.63 | 4.42 |
| M06-L | 1.45 | 2.06 | 2.95 | 3.36 | 20.49 | 6.06 | 2.46 |
| revPBE | 3.10 | 3.87 | 2.58 | 11.23 | 16.43 | 7.44 | 5.20 |
| B97D | 3.20 | 3.41 | 3.49 | 4.37 | 23.82 | 7.66 | 3.62 |
| *m*PWLYP | 1.67 | 2.11 | 1.93 | 7.07 | 26.75 | 7.90 | 3.19 |
| GLYP | 2.23 | 2.71 | 2.69 | 9.48 | 23.75 | 8.17 | 4.28 |
| BLYP-D | 2.92 | 3.09 | 4.22 | 3.84 | 28.30 | 8.47 | 3.52 |
| BLYP | 1.88 | 2.35 | 2.28 | 8.31 | 28.07 | 8.58 | 3.71 |
| XLYP | 1.93 | 2.33 | 2.43 | 8.76 | 31.36 | 9.36 | 3.86 |
| OLYP | 4.48 | 5.46 | 5.20 | 17.10 | 14.74 | 9.39 | 8.06 |

[a]The functionals are sorted according to the total average error, Avg1. Avg2 excludes open-chain isomerism.

**3.3. DFT Methods.** Following the same distinction between the five different types of isomerism used in the discussion of the wave-function-based ab initio results, the RMS errors of the 31 density functionals tested are reported in Table 6. The corresponding errors for functional-specific geometries (Table S1) follow essentially the same trends, so the discussion will mostly focus on the results of single-point energies on the common set of reference geometries. A quick inspection of Table 5 reveals some gross features regarding the performance of DFT methods in general. RMS errors related to the relative stabilities of hydroxymethyl rotamers, $\beta$-anomers, and gluco-pyranose ring conformers range approximately from 1 to 5 kJ mol$^{-1}$, with average values a little over 2 kJ mol$^{-1}$. In the case of pyranose−furanose isomerism, the errors are somewhat larger, 5.3 kJ mol$^{-1}$ on average. However, the most dramatic deterioration is observed for the open-chain isomers as a consequence of the bond breaking, where the RMS errors range from 5 kJ mol$^{-1}$ in the best case and reach over 30 kJ mol$^{-1}$ for the most problematic methods (12.3 kJ mol$^{-1}$ on average).

This marked difference in performance for the relative energetics of cyclic and open-chain forms is one of the most striking results of the present study and leads us to report two total average errors, one that includes all isomer subsets (Avg1) and one that excludes acyclic isomers (Avg2). At the same time, consideration should be given to the maximum errors produced by each functional for each type of isomerism, and these are provided in Table 7. The ordering of functionals in Tables 6 and 7 according to total average error roughly corresponds to Perdew's "Jacob's ladder" of density functional approximations.[146] However, there are exceptions from this idealized ordering, and these are important for choosing the most appropriate method for a given desirable cost/performance ratio, as will be discussed in the following.

A comparison of Tables 4 and 6 makes it immediately apparent that in general the accuracy expected from the majority of density functionals does not match the accuracy of wave-function-based methods, with the exception of the double hybrid *m*PW2PLYP and B2PLYP functionals, as well as of the hybrid meta-GGA M06-2X and M06 functionals. These functionals deviate by less than 2.9 kJ mol$^{-1}$ from the CCSD(T)/CBS reference results in terms of total average error (Avg1) and by less than 1.5 kJ mol$^{-1}$ on average if the open-chain isomerism is neglected (Avg2). Therefore, they can be considered competitive with wave-function-based correlated methods such as MP2, SCS-MP2, and CCSD.

The defining feature of the double hybrid functionals is that their correlation part combines the contribution of a semilocal correlation functional with an MP2 correction obtained from Kohn−Sham orbitals.[118] The exchange part is described as in a typical hybrid functional, although a higher than usual percentage of exact exchange is used: for the double hybrids tested in this work, it amounts to 53% for B2PLYP and 55% for *m*PW2PLYP. The accuracy of these functionals for thermodynamic properties has been well-established,[119,122] and it comes as no surprise that they are found at the top positions also for the present test set of isomers. Among the double hybrids tested in this work, *m*PW2PLYP and its dispersion corrected variant are to be preferred over B2PLYP on account of the smaller RMS errors for the acyclic forms. It is difficult to discern a definitive improvement upon inclusion of empirical dispersion corrections for the present test set, but inclusion of dispersion corrections appears to increase the maximum errors for all subsets except the furanose isomers. This point has been discussed previously by Csonka and Kaminsky,[56] who have ascribed the lack of improvement for GGA and meta-GGA functionals upon inclusion of empirical dispersion corrections to the predominance of the exchange part of these functionals in overestimation of relative conformational energies; global hybrid functionals proved instead more amenable to such corrections.[56]

The Minnesota density functionals also performed well, and the hybrid meta-GGA M06 and M06-2X functionals are on par with the double hybrids described above. These two functionals differ in the percentage of the exact exchange (27% for M06 and 54% for M06-2X), with the better performing M06-2X being specifically parametrized only for nonmetals.[117] It should be pointed out that the functionals that come at the top of Table 6 seem to benefit from high fractions of exact exchange, although overall this is not a key factor for the performance of a functional, judging for example from the position of TPSSh (10% exact exchange). With respect to the other member of the Minnesota family, the local meta-GGA variant M06-L, it can be seen that despite appearing low in Table 6 owing to the large total average error, it is in fact a good performer within its class for all cases of isomerism except for the acyclic forms, where it produces errors in excess of 20 kJ mol$^{-1}$. Thus, when the focus

**Table 7. Maximum Absolute Errors (kJ mol$^{-1}$) of DFT Methods for Each Type of Isomerism with Respect to the CCSD(T)/CBS Reference Values Using the Reference SCS-MP2 Geometries[a]**

| functional | rotamers | $\beta$-anomers | ring conformers | furanoses | open chain |
|---|---|---|---|---|---|
| *m*PW2PLYP-D | 2.17 | 1.57 | 2.86 | 2.31 | 7.20 |
| M06-2X | 1.83 | 1.38 | 2.38 | 4.87 | 7.31 |
| *m*PW2PLYP | 1.14 | 0.96 | 2.98 | 3.73 | 6.99 |
| M06 | 2.24 | 2.97 | 3.48 | 2.82 | 8.04 |
| B2PLYP | 1.40 | 1.26 | 2.75 | 4.63 | 9.79 |
| B3PW91 | 3.90 | 3.90 | 4.21 | 8.44 | 4.65 |
| B2PLYP-D | 3.02 | 2.71 | 4.51 | 3.52 | 10.66 |
| TPSSh | 4.57 | 2.93 | 2.32 | 7.96 | 8.72 |
| *m*PW1PW91 | 3.76 | 3.61 | 3.54 | 7.29 | 9.44 |
| PBE0 | 4.00 | 3.58 | 2.99 | 6.12 | 11.74 |
| PBE | 6.47 | 5.15 | 2.58 | 8.33 | 11.90 |
| PBE-D | 8.83 | 7.93 | 8.48 | 5.25 | 12.19 |
| X3LYP | 2.68 | 3.00 | 5.00 | 8.95 | 14.30 |
| BP86 | 6.12 | 4.81 | 2.89 | 7.48 | 15.05 |
| *m*PW1LYP | 3.02 | 2.75 | 5.36 | 8.92 | 14.77 |
| *m*PWPW91 | 6.21 | 5.19 | 3.26 | 8.29 | 14.46 |
| TPSS | 5.56 | 4.04 | 2.44 | 9.59 | 15.06 |
| B3LYP-D | 5.08 | 4.54 | 7.83 | 6.52 | 17.12 |
| B3LYP | 3.20 | 3.45 | 5.27 | 10.26 | 16.24 |
| B1LYP | 3.69 | 3.22 | 5.76 | 10.53 | 16.36 |
| BP86-D | 10.08 | 9.44 | 11.41 | 7.42 | 15.70 |
| TPSS-D | 9.78 | 9.80 | 12.06 | 6.85 | 15.49 |
| M06-L | 3.11 | 3.15 | 5.72 | 6.57 | 23.04 |
| revPBE | 6.07 | 7.81 | 6.05 | 15.07 | 19.21 |
| B97D | 7.57 | 6.85 | 9.83 | 7.80 | 28.86 |
| *m*PWLYP | 3.76 | 4.49 | 3.94 | 10.20 | 30.03 |
| GLYP | 4.51 | 5.76 | 6.18 | 14.40 | 28.14 |
| BLYP-D | 7.35 | 7.55 | 11.52 | 7.25 | 33.18 |
| BLYP | 3.78 | 5.12 | 4.83 | 12.33 | 31.26 |
| XLYP | 4.01 | 5.23 | 4.92 | 12.98 | 33.95 |
| OLYP | 8.47 | 10.62 | 11.49 | 25.48 | 20.50 |

[a]The ordering of the functionals is the same as in Table 5.

is on purely conformational isomerism, M06-L is competitive with the TPSS functional, although the present results give practically no reason to prefer either of them over conventional and computationally simpler GGA functionals such as PBE or BP86.

A technical point concerning the performance of M06-2X is the choice of integration grid, since it has been shown that small grids can introduce non-negligible errors in computed energies,[147] including in carbohydrate systems.[55] The standard grid used in the present study, designated as (75, 302), has 75 radial shells and 302 angular points. Convergence of M06-2X energies was tested using two even larger grids, defined as (99, 590) and (250, 590). The computational cost increased by approximately a factor of 3 for the (99, 590) grid and a factor of 5 for the largest (250, 590) grid. Compared with the (75, 302) grid, the energies obtained with the larger grids over the whole set of conformers show average deviations of $4.5 \times 10^{-6}$ Hartrees and $5.5 \times 10^{-6}$ Hartrees, respectively, thus confirming that M06-2X energies are practically converged. The RMS metrics for each class of isomers are also unaffected. By contrast, the use of a smaller SG1 (50,194) grid leads to more significant errors that can reach up to $6.8 \times 10^{-4}$ Hartrees, in agreement with previous reports based on such a small grid size.[55,147] Since the magnitude of these nonsystematic errors is comparable to the magnitude of energy differences considered in the present case, we suggest the (75, 302) grid as the minimum for M06-2X calculations.

Of the more "traditional" hybrid functionals, B3PW91 outperforms the hybrid meta-GGA TPSSh and all other hybrids. Compared with *m*PW2PLYP and M06-2X, B3PW91 produces larger RMS errors for hydroxymethyl rotamers, anomers, and ring conformers, as well as for the furanose isomers, but it excels in the prediction of the relative energies of open-chain isomers, for which it yields the lowest RMS error (2.5 kJ $mol^{-1}$) over all density functionals.

Immediately following the double hybrid, M06, and B3PW91 functionals is the other hybrid meta-GGA included in our evaluation, TPSSh, and the two hybrids *m*PW1PW91 and PBE0. Examining the RMS errors for each type of isomerism separately, it becomes apparent that TPSSh benefits from its superior performance on the prediction of relative energies of the acyclic forms with respect to the "traditional" hybrid-GGA functionals, while at the same time it is worse than *m*PW1PW91 and PBE0 for the pyranose−furanose isomerism. If the pyranose−open-chain isomerism is excluded from the comparison (see Avg2 in Table 6), the order of performance for hybrid GGA functionals is reversed, with PBE0 now apparently being the best choice, although the differences are rather marginal. Therefore, the discrepancy in the performance of density functionals with respect to the different types of isomerism included in the test set creates some complications, because the ordering of functionals according to the two average errors, Avg1 and Avg2, of Table 6 often differs substantially.

It is important to note that the popular B3LYP functional showed relatively low errors for the first three types of isomerism, but more pronounced errors for the furanose and open-chain species, making it a poor choice for the present test set in total (Avg1). However, the dispersion corrected version of the functional improves significantly on the prediction of the pyranose−furanose isomerism, so that based on Avg2 values, B3LYP-D appears as the second best choice among hybrid functionals after PBE0, even though, remarkably, it is the worst performing hybrid functional for the subset of open-chain isomerism. Therefore, the choice of a hybrid GGA method depends to a large extent on the type of isomerism under investigation. In general terms, B3PW91, *m*PWPW91, and PBE0 are good choices among the hybrid functionals evaluated in this work. Of these, B3PW91 is the best option over all functionals when energy differences between ring and acyclic forms are of interest. We note that these results are in line with those of Csonka and co-workers, who have also recently established that B3PW91 and PBE0 are good choices for conformational energy differences in carbohydrates.[55,56] Nevertheless, they are still inferior to *m*PW2PLYP and M06-2X for the study of conformational isomerism in view of the largest maximum errors they yield for the relevant isomer subsets. An increase of the percentage of exact exchange in the PBE hybrid to 32% has been shown to improve performance for thermochemical properties.[148] This variant was tested here and led to RMS errors of 1.29, 2.23, 1.97, 3.62, and 12.08 kJ $mol^{-1}$ for the five types of isomerism, respectively. Thus, on average the increased weight of exact exchange indeed improves performance over PBE0 for the cyclic forms (Avg2 = 2.28 kJ $mol^{-1}$), but not as a total average (Avg1 = 4.24 kJ $mol^{-1}$) because it leads to deterioration of the relative errors for the open-chain isomers.

PBE emerges as the best performing GGA functional and surpasses the meta-GGAs TPSS and M06-L in terms of accuracy over all types of isomerism. It is followed by BP86 and *m*PWPW91, which showed rather similar RMS error trends with a slight deterioration in the errors related to the open-chain forms. Since the performance of several functionals proved quite contrasting in this respect, it is again informative to differentiate between the conformational isomerism represented by the hydroxymethyl rotamers, the anomers, and the ring conformers, and the isomerism that involves changes in the number of bond types. The most dramatic example of this behavior is found among functionals that occupy the bottom rows of Table 6. For example, *m*PWLYP shows well-controlled RMS errors for the first three types of isomerism. On the basis of only these errors, there is no reason to prefer the higher-ranked GGAs for studies that mainly involve rotations of hydroxyl groups and redistribution of intramolecular hydrogen bonding. However, the errors of *m*PWLYP and other GGA functionals that perform adequately for purely conformational isomerism, such as BLYP and XLYP, seem inappropriate for the pyranose−furanose isomerism and become unacceptable for applications involving energetic comparisons between ring and open-chain isomers.

The effect of the empirical vdW corrections can be easily judged by comparing the RMS and maximum errors obtained by the PBE, BP86, BLYP, and TPSS functionals and their dispersion corrected versions. Perusal of Tables 6 and 7 indicates that the empirical dispersion correction reduces the errors for the pyranose−furanose isomerism in all cases, whereas the effect on the relative energies of open-chain species is minimal and not beneficial. By contrast, inclusion of dispersion corrections results in a more pronounced errors for the subset of ring conformers, by almost a factor of 4 for the maximum absolute errors of PBE-D and BP86-D versus PBE and BP86, respectively. Furthermore, the B97D functional does not appear to have an advantage over functionals that do not include vdW corrections. Therefore, as commented above for the case of the double hybrid functionals, inclusion of these empirical dispersion corrections does not appear to be

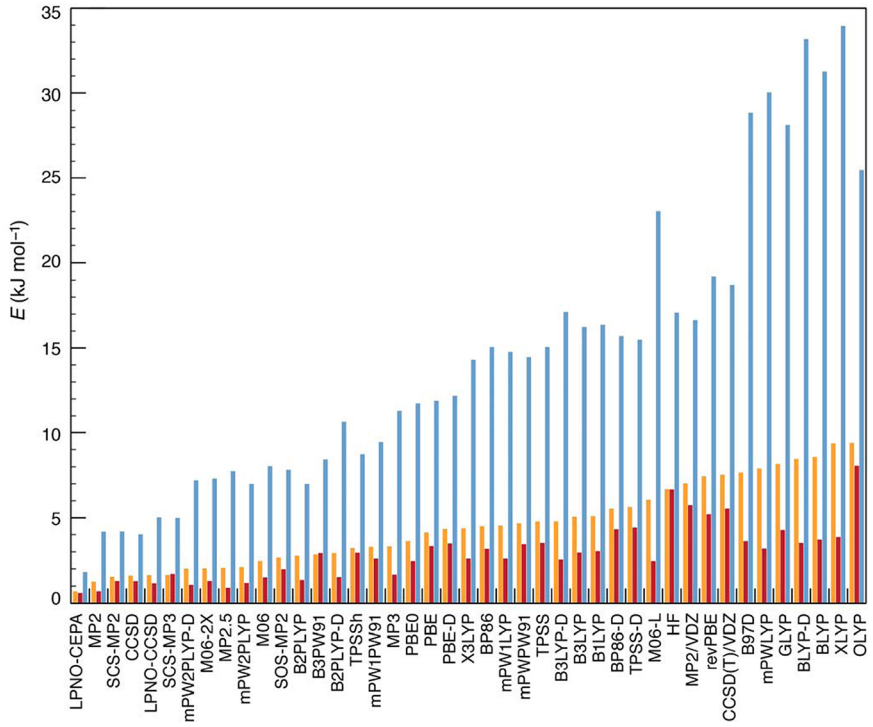


**Figure 7.** Graphical summary of the performance of all methods included in the present study in terms of total average error (orange bars), average errors excluding open-chain isomerism (red bars), and maximum absolute errors (blue bars).

uniformly beneficial for the present test set. Given that ample evidence exists from several studies that dispersion correction-corrected functionals in general outperform their uncorrected counterparts,[122] the results observed here are tentatively attributed to the fact that the conformational space and the corresponding energetics of monosaccharides are dominated by intramolecular interactions and not by strong steric factors.[56] Despite this, we consider it probable that vdW corrections will be important and beneficial in larger carbohydrate systems such as oligosaccharides, where rotations related to glycosidic bonds are more sensitive to steric repulsions and dispersion interactions between the constituent rings.

The point of greatest concern for the performance of DFT methods is the prediction of relative energies of the open-chain isomers. Most functionals typically underestimate the energy difference between the ring and the open chain forms. The only functionals that overestimate the energy difference between pyranose and acyclic forms with respect to the CCSD(T) reference are M06-2X, *m*PW1PW91, PBE0, and to a much smaller extent B3PW91. The most disappointing performance is observed for the GGA functionals occupying the bottom rows of Table 6. Since these functionals incorporate the LYP correlation, it would seem that the correlation part is implicated in the poor performance. The LYP functional is based on the Colle−Salvetti formula for the correlation energy of the helium atom and was not constructed as a formal correction to LDA.[106] As a result, LYP does not obey the uniform electron gas (UEG) limit, a feature that was found to correlate with failures of LYP-based functionals in several cases.[149−153] This is true for B97D, which also performs poorly for open-chain isomerism. It is beyond the purposes of the present study to trace the precise origins of the poor behavior of these functionals and disentangle the various factors (fraction of exact exchange, treatment of correlation, self-interaction,[154] delocalization errors,[155] etc.) that might contribute to aggravating or mitigating the problem. We note that in a wider perspective, the failure for the ring versus open-chain isomerism is relevant to the well-known problematic treatment of stereoelectronic effects in hydrocarbons by DFT.[156−159] Medium-range correlation, involved for example in 1,3 interactions, was shown to be decisive in this class of molecules and is revealed most conspicuously in the inability of several functionals to predict the relative stabilities of linear and branched alkane isomers. A relevant point in this respect is that short-range and long-range molecular interactions are not typically covered equally well by an empirical dispersion correction term such as the one used here, although unified treatments of intra- and intermolecular dispersion interactions have been recently proposed.[160]

All observations regarding the performance of the density functionals evaluated in the present study with respect to the types of isomerism included in our test set remain practically unaltered if we consider relative energies obtained with structures individually optimized with each functional, instead of single-point energies at the SCS-MP2 reference geometries. This is important for the consistency of the present results, given that in practice DFT would be the preferred choice for geometry optimizations in large-scale computational applications. It is important to note that no conformational changes are observed upon these optimizations. Table S1 and Figure S2 summarize the RMS errors of the 31 density functionals for each type of isomerism with respect to the CCSD(T)//SCS-MP2 reference values. In terms of the total average errors (Avg1), the trends in Tables 5 and S1 remain essentially the same. The overall performance of the best performing density functionals in increasing total average errors follows the order M06-2X < *m*PW2PLYP(-D) < M06 < B2PLYP < B3PW91 < *m*PW1PW91 < TPSSh < PBE0 < PBE, to mention only those functionals that yield average errors less than 4 kJ mol$^{-1}$.


dx.doi.org/10.1021/ct3002305 | *J. Chem. Theory Comput.* 2012, 8, 2630−2645

On the whole, the picture obtained from the evaluation of the 31 DFT methods in this study is characterized by the inconsistent behavior of most functionals for the five types of isomerism considered. Modeling rotational or conformational isomerism appears accessible to many standard DFT methods, with several modern functionals comparing favorably to expensive ab initio methods. Although the OH rotational conformational space of some furanoses is reasonably well described,[55] the accuracy of many common functionals is lower than wave-function-based methods for the pyranose−furanose isomerism. This is relevant not only to the more general issue of six-membered vs five-membered ring populations in hexoses but also to the study of pyranose−furanose glycosidic linkages and associated oligosaccharide conformations. When processes that involve ring-opening or formation are considered, the choice of DFT method becomes even more restricted. Overall, *m*PW2PLYP and M06-2X can be singled out as the most consistent functionals for the study of all forms of isomerism in monosaccharides. B3PW91 performs particularly well for the relative energies of the acyclic forms, while in terms of total average errors it is followed by TPSSh, *m*PW1PW91, PBE0, and PBE. It is suggested that even if practical reasons necessitate the use of a density functional other than the best-performing ones, the results should be validated by comparison with a more robust method to confirm that they are not compromised by functional deficiencies. Unless a method such as LPNO-CEPA is applicable, this validation can be against results obtained with *m*PW2PLYP and M06-2X with a sufficiently fine grid.

## 4. CONCLUSIONS

An overview of the results obtained in the present study is given in Figure 7, which summarizes the average errors of all tested wave-function-based and DFT approaches. The methods are ranked according to the total average error (orange bars). Also shown are the average errors excluding the open-chain isomerism (red bars) and the maximum errors (blue bars) of each method for the whole test set used in the present study. The maximum errors for all methods correspond to prediction of the relative stabilities of open-chain species with respect to the pyranose forms of the aldohexoses, with the exception of B3PW91 and OLYP, where maximum errors correspond to furanose isomers.

The best agreement with the CCSD(T) reference results is obtained with wave-function-based ab initio methods. LPNO-CEPA is the best performer and yields results that can be considered equivalent to the CCSD(T) reference for most purposes. In view of this excellent performance over all isomeric forms considered in this work and of the significantly reduced cost of such local correlation methods, it is suggested that LPNO-CEPA can be used as a reference method for studies of isomerism in monosaccharides and related systems. The importance of this conclusion is emphasized by the ability of LPNO methods to efficiently handle larger systems with a few thousand basis functions, potentially enabling routine use of this approach in place of methods such as MP2 and SCS-MP2.

Another important outcome of the present study with respect to DFT methods is the very different behavior of most functionals for the five different types of isomerism: while reasonably small errors are to be expected by most DFT methods for hydroxymethyl rotamers, anomers, and ring conformers, errors can become large for the pyranose−furanose isomerism and unacceptable for the pyranose−open-chain isomerism. Therefore, it is difficult to distill the variable performance of DFT functionals into a unique statistical average figure. Thus, the choice of DFT method for a particular application should combine a consideration of both the general performance of the functional and the specific accuracy expected for the problem at hand, along with a cross-checking of the performance of a few functionals belonging to different families. Ideally, when the system size is not extremely large or when a representative medium-sized model can be constructed, the LPNO-CEPA method should be applied to confirm that relative DFT energetics are reliable enough to describe the problem at hand.

Overall, the best functionals are double hybrids and the two hybrid meta-GGA Minnesota functionals M06-2X and M06. From the double hybrid functionals, mPW2PLYP-D yields the smallest average and maximum errors compared to all other DFT methods. All of these best performing functionals are directly competitive with ab initio methods and surpass several of them, such as the SOS-MP2 and even the more costly MP3. This becomes more obvious if the cyclic vs open-chain isomerism, a particularly problematic point with DFT, is excluded from the comparison. However, even though competitive on average, these functionals are still liable to produce larger maximum errors than methods like SCS-MP2 or LPNO-CCSD.

The best choice among the more conventional hybrid functionals is B3PW91, which also yields the lowest errors among all functionals for the difficult case of pyranose vs open-chain isomerism. This is followed in terms of accuracy by *m*PW1PW and PBE0. Among the pure GGA functionals the best choice is PBE, which outperforms many hybrids for the test set of the present study. Note that B3LYP benefits from the addition of dispersion corrections mainly for errors related to open-chain forms, but B3LYP-D is still found to be worse than the pure GGA BP86 functional. It is also surpassed by the meta-GGA functional TPSS, which in turn is outperformed by PBE. Thus, meta-GGA functionals offer no discernible improvement over the simpler GGA functionals, while GGA functionals that include LYP correlation should be avoided because of the severely imbalanced treatment of acyclic isomers. Empirical dispersion corrections are not found to be beneficial for the types of monosaccharide isomerism considered in this study, with the exception of pyranose−furanose isomerism.

The hierarchy of methods obtained in this work is expected to facilitate the choice of an appropriate method for applications dealing with the treatment of isomerism in small saccharides. It remains to be confirmed whether the results obtained with hexoses hold upon extension to larger oligosaccharides. Although the conclusions drawn from analysis of isomer energy differences do not necessarily carry over to reaction barriers or other physical observables such as those obtained from spectroscopy, they form a rigorously benchmarked and reliable starting point for any such investigation as well.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

Figures S1 and S2, Tables S1 and S2, Cartesian coordinates of optimized structures, and technical notes. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION


### Corresponding Author

*E-mail: wsameera@iciq.es, dimitrios.pantazis@mpi-mail.mpg.de.


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the ICIQ foundation and from the Max Planck Society.

## REFERENCES


(1) Stern, R.; Jedrzejas, M. J. *Chem. Rev.* **2008**, *108*, 5061−5085.
(2) Ramesh, H. P.; Tharanathan, R. N. *Crit. Rev. Biotechnol.* **2003**, *23*, 149−173.
(3) Robyt, J. F. *Essentials of Carbohydrate Chemistry*; Springer: Dordrecht, The Netherlands, 1997; p 414.
(4) Rao, V. S. R.; Qasba, P. K.; Balaji, P. V.; Chandrasekaran, R. *Conformation of Carbohydrates*; Harwood Academic Publishers: Amsterdam, 1998; p 409.
(5) *Carbohydrates: Structures, Syntheses and Dynamics*; Finch, P., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1999, p 352.
(6) Davis, A. P.; Wareham, R. S. *Angew. Chem., Int. Ed.* **1999**, *38*, 2978−2996.
(7) Wormald, M. R.; Petrescu, A. J.; Pao, Y. L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102*, 371−386.
(8) Pérez, S.; Kouwijzer, M. In *Carbohydrates: Structures, Syntheses and Dynamics*; Finch, P., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1999; pp 258−293.
(9) da Silva, C. O. *Theor. Chem. Acc.* **2006**, *116*, 137−147.
(10) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1993**, *115*, 5745−5753.
(11) Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci.* **2001**, *98*, 10541−10545.
(12) Miljković, M. *Carbohydrates: Synthesis, Mechanisms, and Stereoelectronic Effects*; Springer: Dordrecht, The Netherlands, 2010; p 460.
(13) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(14) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(15) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1993**, *115*, 5348−5348.
(16) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209−220.
(17) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221−264.
(18) Woods, R. J.; Szarek, W. A.; Smith, V. H. *J. Chem. Soc., Chem. Commun.* **1991**, 334−337.
(19) Barrows, S. E.; Dulles, F. J.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *Carbohydr. Res.* **1995**, *276*, 219−251.
(20) Appell, M.; Strati, G.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2004**, *339*, 537−551.
(21) Foley, B. L.; Tessier, M. B.; Woods, R. J. *WIREs Comput. Mol. Sci.* **2012**, *2*, 652−697.
(22) Allinger, N. L.; Li, F.; Yan, L. *J. Comput. Chem.* **1990**, *11*, 848−867.
(23) Allinger, N. L.; Li, F.; Yan, L.; Tai, J. C. *J. Comput. Chem.* **1990**, *11*, 868−895.
(24) Lii, J.-H.; Allinger, N. L. *J. Phys. Org. Chem.* **1994**, *7*, 591−609.
(25) Lii, J.-H.; Allinger, N. L. *J. Comput. Chem.* **1998**, *19*, 1001−1016.
(26) Kuttel, M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236−1243.
(27) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *J. Comput. Chem.* **2008**, *29*, 2543−2564.
(28) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. *J. Chem. Theory Comput.* **2009**, *5*, 2353−2370.
(29) Hatcher, E.; Guvench, O.; MacKerell, A. D. *J. Phys. Chem. B* **2009**, *113*, 12466−12476.
(30) Raman, E. P.; Guvench, O.; MacKerell, A. D. *J. Phys. Chem. B* **2010**, *114*, 12981−12994.
(31) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. *J. Chem. Theory Comput.* **2011**, *7*, 3162−3180.
(32) Glennon, T. M.; Zheng, Y. J.; Legrand, S. M.; Shutzberg, B. A.; Merz, K. M. *J. Comput. Chem.* **1994**, *15*, 1019−1040.
(33) Momany, F. A.; Willett, J. L. *Carbohydr. Res.* **2000**, *326*, 194−209.
(34) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraserreid, B. *J. Phys. Chem.* **1995**, *99*, 3832−3846.
(35) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622−655.
(36) Salisburg, A. M.; Deline, A. L.; Lexa, K. W.; Shields, G. C.; Kirschner, K. N. *J. Comput. Chem.* **2009**, *30*, 910−921.
(37) Damm, W.; Frontera, A.; TiradoRives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1997**, *18*, 1955−1970.
(38) Kony, D.; Damm, W.; Stoll, S.; van Gunsteren, W. F. *J. Comput. Chem.* **2002**, *23*, 1416−1429.
(39) Lins, R. D.; Hunenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400−1412.
(40) Hansen, H. S.; Hünenberger, P. H. *J. Comput. Chem.* **2011**, *32*, 998−1032.
(41) Stortz, C. A.; Johnson, G. P.; French, A. D.; Csonka, G. I. *Carbohydr. Res.* **2009**, *344*, 2217−2228.
(42) French, A. D.; Shaäfer, L.; Newton, S. Q. *Carbohydr. Res.* **1993**, *239*, 51−60.
(43) Wladkowski, B. D.; Chenoweth, S. A.; Jones, K. E.; Brown, J. W. *J. Phys. Chem. A* **1998**, *102*, 5086−5092.
(44) Momany, F. A.; Willett, J. L. *J. Comput. Chem.* **2000**, *21*, 1204−1219.
(45) Strati, G. L.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2002**, *337*, 1833−1849.
(46) Strati, G. L.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2002**, *337*, 1851−1859.
(47) French, A. D.; Johnson, G. P. *Can. J. Chem.* **2006**, *84*, 603−612.
(48) Momany, F. A.; Schnupf, U.; Willett, J. L.; Bosma, W. B. *Struct. Chem.* **2007**, *18*, 611−632.
(49) Schnupf, U.; Willett, J. L.; Bosma, W. B.; Momany, F. A. *Carbohydr. Res.* **2007**, *342*, 196−216.
(50) Schnupf, U.; Willett, J. L.; Bosma, W. B.; Momany, F. A. *Carbohydr. Res.* **2007**, *342*, 2270−2285.
(51) Schnupf, U.; Willett, J. L.; Bosma, W. B.; Momany, F. A. *J. Comput. Chem.* **2008**, *29*, 1103−1112.
(52) Schnupf, U.; Willett, J. L.; Bosma, W. B.; Momany, F. A. *Carbohydr. Res.* **2009**, *344*, 362−373.
(53) Schnupf, U.; Willett, J. L.; Momany, F. *Carbohydr. Res.* **2010**, *345*, 503−511.
(54) French, A. D.; Johnson, G. P.; Cramer, C. J.; Csonka, G. I. *Carbohydr. Res.* **2012**, *350*, 68−76.
(55) Csonka, G. I.; French, A. D.; Johnson, G. P.; Stortz, C. A. *J. Chem. Theory Comput.* **2009**, *5*, 679−692.
(56) Csonka, G. I.; Kaminsky, J. *J. Chem. Theory Comput.* **2011**, *7*, 988−997.
(57) Riley, K. E.; Pitoňák, M.; Černý, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 66−80.
(58) Šponer, J.; Riley, K. E.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595−2610.
(59) Pitoňák, M.; Riley, K. E.; Neogrády, P.; Hobza, P. *ChemPhysChem* **2008**, *9*, 1636−1644.
(60) Kabeláč, M.; Sherer, E. C.; Cramer, C. J.; Hobza, P. *Chem.—Eur. J.* **2007**, *13*, 2067−2077.
(61) Vondrášek, J.; Bendová, L.; Klusák, V.; Hobza, P. *J. Am. Chem. Soc.* **2005**, *127*, 2615−2619.
(62) Morgado, C. A.; Jurečka, P.; Svozil, D.; Hobza, P.; Šponer, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3522−3534.

(63) Řeha, D.; Valdés, H.; Vondrášek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S. *Chem.—Eur. J.* **2005**, *11*, 6803−6817.
(64) Valdes, H.; Pluháčková, K.; Pitonák, M.; Řezáč, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747−2757.
(65) Tekin, A.; Jansen, G. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1680−1687.
(66) Riley, K. E.; Hobza, P. *WIREs Comput. Mol. Sci.* **2011**, *1*, 3−17.
(67) Hohenstein, E. G.; Sherrill, C. D. *WIREs Comput. Mol. Sci.* **2011**, *2*, 304−326.
(68) Foster, M. E.; Sohlberg, K. *Phys. Chem. Chem. Phys.* **2010**, *12*, 307−322.
(69) Johnson, E. R.; Mackie, I. D.; DiLabio, G. A. *J. Phys. Org. Chem.* **2009**, *22*, 1127−1135.
(70) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463−1473.
(71) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(72) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104−19.
(73) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555−569.
(74) Grafenstein, J.; Cremer, D. *J. Chem. Phys.* **2009**, *130*, 124105.
(75) Steinmann, S. N.; Corminboeuf, C. *J. Chem. Theory Comput.* **2010**, *6*, 1990−2001.
(76) Williams, H. L.; Chabalowski, C. F. *J. Phys. Chem. A* **2001**, *105*, 646−659.
(77) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2005**, *123*, 214103.
(78) Hesselmann, A.; Jansen, G.; Schutz, M. *J. Chem. Phys.* **2005**, *122*, 014103.
(79) Sameera, W. M. C.; Maseras, F. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10520−10526.
(80) Ponder, J. W.; Ren, P.; Pappu, R. V.; Hart, R. K.; Hodgson, M. E.; Cistola, D. P.; Kundrot, C. E.; Richards, F. M. *Tinker*, version 5.1; Washington University School of Medicine: St. Louis, MO, 2010.
(81) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095−9102.
(82) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(83) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359−363.
(84) Bernholdt, D. E.; Harrison, R. J. *Chem. Phys. Lett.* **1996**, *250*, 477−484.
(85) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331−340.
(86) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143−152.
(87) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793−9802.
(88) Grimme, S. *J. Comput. Chem.* **2003**, *24*, 1529−1537.
(89) Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. *ChemPhysChem* **2009**, *10*, 282−289.
(90) Neese, F.; Hansen, A.; Wennmohs, F.; Grimme, S. *Acc. Chem. Res.* **2009**, *42*, 641−648.
(91) Antony, J.; Grimme, S.; Liakos, D. G.; Neese, F. *J. Phys. Chem. A* **2011**, *115*, 11210−11220.
(92) Liakos, D. G.; Hansen, A.; Neese, F. *J. Chem. Theory Comput.* **2011**, *7*, 76−87.
(93) Neese, F.; Wennmohs, F.; Hansen, A. *J. Chem. Phys.* **2009**, *130*, 114108.
(94) Hansen, A.; Liakos, D. G.; Neese, F. *J. Chem. Phys.* **2011**, *135*, 214102.
(95) Neese, F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73−78.
(96) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243−252.
(97) Jurečka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89−94.
(98) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.
(99) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(100) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824.
(101) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(102) Zhang, Y.; Yang, W. *Phys. Rev. Lett.* **1998**, *80*, 890−890.
(103) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664−675.
(104) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671−6687.
(105) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244−13249.
(106) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(107) Gill, P. M. W. *Mol. Phys.* **1996**, *89*, 433−445.
(108) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403−412.
(109) Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673−2677.
(110) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(111) Schultz, N. E.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 11127−11143.
(112) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(113) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1997**, *274*, 242−250.
(114) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158−6170.
(115) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129−12137.
(116) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157−167.
(117) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(118) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
(119) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4398−4401.
(120) Vydrov, O. A.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *133*, 244103.
(121) Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. B* **2010**, *82*, 081101.
(122) Schwabe, T.; Grimme, S. *Acc. Chem. Res.* **2008**, *41*, 569−579.
(123) Grimme, S.; Steinmetz, M.; Korth, M. *J. Org. Chem.* **2007**, *72*, 2118−2126.
(124) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. *Chem. Phys.* **2009**, *356*, 98−109.
(125) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M. T., K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, rev. A.02; Gaussian, Inc.: Wallingford, CT, 2009.
(126) Wolfe, S. *Acc. Chem. Res.* **1972**, *5*, 102−111.
(127) Marchessault, R. H.; Perez, S. *Biopolymers* **1979**, *18*, 2369−2374.
(128) Nishida, Y.; Ohrui, H.; Meguro, H. *Tetrahedron Lett.* **1984**, *25*, 1575−1578.
(129) Juaristi, E.; Cuevas, G. *The Anomeric Effect*; CRC Press: Boca Raton, FL, 1995; p 256.
(130) Edward, J. T. *Chem. Ind.* **1955**, 1102−1104.
(131) Fuchs, B.; Ellencweig, A.; Tartakovsky, E.; Aped, P. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 287−289.
(132) Vila, A.; Mosquera, R. A. *J. Comput. Chem.* **2007**, *28*, 1516−1530.
(133) Huang, Y.; Zhong, A.-G.; Yang, Q.; Liu, S. *J. Chem. Phys.* **2011**, *134*, 084103.
(134) Mo, Y. *Nat. Chem.* **2010**, *2*, 666−671.
(135) Momany, F. A.; Willett, J. L.; Schnupf, U. *THEOCHEM* **2010**, *953*, 61−82.
(136) Schnupf, U.; Willett, J. L.; Momany, F. A. *J. Comput. Chem.* **2010**, *31*, 2087−2097.

(137) Schnupf, U.; Momany, F. A. *J. Phys. Chem. B* **2012**, *116*, 6618−6627.
(138) Stoddart, J. F. *Stereochemistry of Carbohydrates*; Wiley-International: New York, 1971; p 264.
(139) Hendrickson, J. B. *J. Am. Chem. Soc.* **1967**, *89*, 7047−7061.
(140) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354−1358.
(141) Ionescu, A. R.; Bérces, A.; Zgierski, M. Z.; Whitfield, D. M.; Nukada, T. *J. Phys. Chem. A* **2005**, *109*, 8096−8105.
(142) Neese, F.; Hansen, A.; Liakos, D. G. *J. Chem. Phys.* **2009**, *131*, 064103.
(143) Lochan, R. C.; Jung, Y.; Head-Gordon, M. *J. Phys. Chem. A* **2005**, *109*, 7598−7605.
(144) Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670−6688.
(145) Csonka, G. I.; Éliás, K.; Csizmadia, I. G. *Chem. Phys. Lett.* **1996**, *257*, 49−60.
(146) Perdew, J. P.; Schmidt, K. In *Density Functional Theory and Its Application to Materials*; VanDoren, V., VanAlsenoy, C., Geerlings, P., Eds.; AIP: Melville, NY, 2001; Vol. *577*, pp 1−20.
(147) Wheeler, S. E.; Houk, K. N. *J. Chem. Theory Comput.* **2010**, *6*, 395−404.
(148) Csonka, G. I.; Perdew, J. P.; Ruzsinszky, A. *J. Chem. Theory Comput.* **2010**, *6*, 3688−3703.
(149) Zhao, S.; Li, Z.-H.; Wang, W.-N.; Liu, Z.-P.; Fan, K.-N.; Xie, Y.; Schaefer, H. F. *J. Chem. Phys.* **2006**, *124*, 184102.
(150) Paier, J.; Marsman, M.; Kresse, G. *J. Chem. Phys.* **2007**, *127*, 024103.
(151) Pantazis, D. A.; McGrady, J. E.; Maseras, F.; Etienne, M. *J. Chem. Theory Comput.* **2007**, *3*, 1329−1336.
(152) Shamov, G. A.; Schreckenbach, G.; Budzelaar, P. H. M. *J. Chem. Theory Comput.* **2010**, *6*, 3442−3455.
(153) Chéron, N.; Jacquemin, D.; Fleurat-Lessard, P. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7170−7175.
(154) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 194112.
(155) Johnson, E. R.; Mori-Sanchez, P.; Cohen, A. J.; Yang, W. *J. Chem. Phys.* **2008**, *129*, 204112.
(156) Wodrich, M. D.; Corminboeuf, C.; Schleyer, P. v. R. *Org. Lett.* **2006**, *8*, 3631−3634.
(157) Schreiner, P. R.; Fokin, A. A.; Pascal, R. A.; de Meijere, A. *Org. Lett.* **2006**, *8*, 3635−3638.
(158) Grimme, S. *Angew. Chem., Int. Ed.* **2006**, *45*, 4460−4464.
(159) Schreiner, P. R. *Angew. Chem., Int. Ed.* **2007**, *46*, 4217−4219.
(160) Steinmann, S. N.; Csonka, G. I.; Corminboeuf, C. *J. Chem. Theory Comput.* **2009**, *5*, 2950−2958.