Available online at www.sciencedirect.com
SCIENCE DIRECT®
Tetrahedron: Asymmetry
Tetrahedron: *Asymmetry* 16 (2005) 577–586

Fluorinated cellobiose and maltose as stand-ins for energy surface calculations

Alfred D. French,[a,*] Glenn P. Johnson,[a] Anne-Marie Kelterer[b] and Gábor I. Csonka[c,*]

[a]*Southern Regional Research Center, US Department of Agriculture, 1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA*
[b]*Institute of Physical and Theoretical Chemistry, Graz University of Technology, Technikerstrasse 4, A-8010 Graz, Austria*
[c]*Inorganic Chemistry Department of the Budapest University of Technology, Szent Gellert ter 4, Budapest H-1521, Hungary*

Received 22 November 2004; accepted 8 December 2004

Abstract—To better understand computational predictions of disaccharide conformations, ϕ,ψ maps were constructed for two analogs in which all hydroxyl groups were replaced with fluorine atoms (F-cellobiose and F-maltose). These molecules do not permit hydrogen bonding but should give better steric representation than analogs in which hydrogen atoms replaced *exo*-cyclic groups. Hartree Fock and B3LYP density functional quantum mechanics (QM) theory were used. The preferred ring shape for fluorinated glucose depends on the level of QM theory, but over the limited ϕ,ψ space that was studied, the rings remained in the $^4C_1$ form. Also, fluorine atoms are remote enough that they do not affect the torsional energies for the glycosidic bonds. F-Cellobiose maps were predictive of the conformations in crystals, but F-maltose maps were less so. The QM F-cellobiose map and an MM4::QM hybrid map for cellobiose itself were similar. However, the hybrid maltose map had many more experimental conformations within its 2-kcal/mol contour than did the QM F-maltose map. The apparent mean strength of an intra-molecular, inter-residue hydrogen bond is about 3 kcal/mol, based on the energy for many of the hydrogen bonded maltose structures on the F-maltose map. The F-maltose map was similar to a new QM map for an analog of maltose in which all hydroxyl groups were replaced with hydrogen atoms.
Published by Elsevier Ltd.

1. Introduction

Individual disaccharides have extensive ranges of conformations, mostly from variable geometries of the linkages between the monosaccharides. Understanding this flexibility is important not only for disaccharides themselves, but also for larger carbohydrates. For example, the disaccharides cellobiose and maltose are the shortest molecules that have the same inter-residue linkages as the polysaccharides cellulose and amylose, respectively. Likely shapes of the larger molecule will correspond to likely shapes for the corresponding disaccharides. Besides various experimental methods, a classic approach to understanding the shape properties of disaccharides is Ramachandran, or ϕ,ψ, mapping of the energy hypersurface. In such studies, the two monosaccharide rings are systematically rotated in increments around their bonds to the oxygen atom that links the two rings. At each increment of rotation, the energy is calculated. Ramachandran maps plot the energy against axes of ϕ and ψ, the torsion angles that describe the amount of rotation of the monosaccharide rings (Fig. 1). The regions of energy minima indicate probable structures. A map is predictive if all experimentally observed structures are in the lowest-energy regions and there are no unoccupied regions of low energy.

Computational conformational analysis of disaccharides has seen numerous advances since the initial work in the 1960's.[1] These include consideration of stereoelectronic effects at the glycosidic linkage in the HSEA[2] and PFOS[3] methods, work with the semi-empirical molecular orbital PCILO[4] method, and relaxed-residue maps based on energy minimization with general purpose molecular mechanics (MM) software.[5] More recently, molecular dynamics calculations have included explicit solvent.[6] Other approaches used a hybrid[7] of MM and electronic structure theory (often called quantum mechanics or QM) or, in special circumstances, pure QM.[8] Even today, the methods employed are still limited by the available software and hardware resources.

QM methods are, in principle, able to calculate accurately the stereoelectronic effects that arise from the lone

* Corresponding authors. Tel.: +1 504 286 4410; fax: +1 504 286 4217 (A.D.F.); tel.: +361 463 1835; fax: +361 463 3633 (G.I.C.); e-mail addresses: afrench@srrc.ars.usda.gov; csonka@web.inc.bme.hu

0957-4166/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.tetasy.2004.12.007

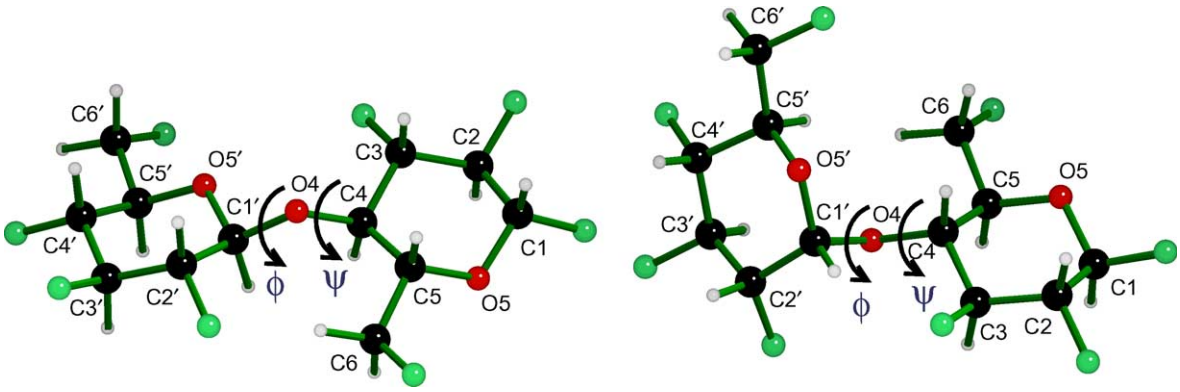

Figure 1. Drawings of F-cellobiose and F-maltose. Also indicated are the atom numbering and the ϕ and ψ torsion angles. Fluorine atoms are the unlabeled large, green spheres.

pairs of electrons in carbohydrates, a task that requires extra effort in MM. On the other hand, QM demands substantial amounts of computer time, and more accurate levels of theory demand even more time. The calculated energies of sugars depend very much on the orientations of their *exo*-cyclic groups. Therefore, finding the lowest-energy structure for each ϕ, ψ location on an energy surface requires consideration of a large number of combinations of the *exo*-cyclic group orientations.[9] For example, cellobiose has 10 *exo*-cyclic bonds about which rotation can occur, besides the ϕ and ψ torsions. Consideration of just the three staggered conformations for each of these gives 3^{10} or 59,049 combinations. Even though most of those combinations are unstable, it is possible that a different one would be optimal at each grid point on a ϕ,ψ map. Calculations at respectable levels of QM are, at present, too time-consuming to do a thorough, straight-forward test of all possibilities.

Interestingly, we have found that intra-molecular hydrogen bonding is of limited importance to successful prediction of observed structures with isolated model molecules. This is in spite of the strong effects of the *exo*-cyclic group orientations and consequent hydrogen bonding schemes on the calculated energy. Our calculations have been most predictive when the strength of hydrogen bonding is reduced substantially by an elevated dielectric constant.[8,10] Also, our QM calculations on analogs of the disaccharides that have had all of their *exo*-cyclic groups replaced by hydrogen are surprisingly predictive of crystalline structures[11] despite obvious shortcomings from a steric viewpoint.

In this study, we use models of fluorinated cellobiose and maltose (Fig. 1) to calculate energy maps for comparison with experimental structures. Although these compounds do not exist, it is simple to create their computational counterparts. They are sterically better approximations of disaccharides than our models[11] that were based on two *O*-linked tetrahydropyran (THP) rings. Substitution with fluorine is practiced in experimental studies of the effects of removal of a hydrogen bonding capability. According to Taylor,[12] fluorinated carbohydrates are used in biochemical studies to replace hydroxyl groups with minimal disturbance of molecular structure and conformation. Fluorine replaces hydroxyls on a one-to-one basis, allowing an increase in molecular bulk[13,14] compared to the hydrogen atoms in THP–O–THP models. The replacement of OH by F is not a large steric change. The van der Waals radius is only slightly smaller for fluorine than for oxygen (1.35 Å vs 1.4 Å). The distance from the carbon atom of CH_3F to the van der Waals surface of the fluorine is 3.35 Å versus 3.45 Å for the distance from the carbon of CH_3OH to the surface of the hydroxyl, including the hydrogen, and 3.22 Å to the surface of just the oxygen.

Such simplification has precedent. Historically, some modeling systems united the hydroxyl hydrogen and oxygen atoms, a recent example being the empirical CHEAT force field.[15] For either our substitution with fluorine or the united hydroxyl groups, the only rotatable groups for cellobiose and maltose are the two (former) primary alcohol groups. Consideration of all staggered conformations results in $3^2 = 9$ 'starting' structures, any one of which could minimize to the structure that has the lowest energy at a given point on the energy surface. Thus, the fluorinated molecules offer an expensive (at the QM level) but feasible system that has no hydrogen bonding capability but sterically approximates the analogous disaccharides.

Using models of the fluorinated disaccharides as stand-ins, or substitutes, for models of the native disaccharides gives a new means to predict the likely conformations of disaccharides. In the present work, we also have calculated new MM::QM hybrid maps for the native disaccharides. We have also calculated a QM map for a maltose analog with all hydroxyl groups replaced by hydrogen. That analog retains *exo*-cyclic methyl groups and gives a better steric representation than our THP–O–THP analogs, although perhaps not as good as F-maltose. Consideration of these additional maps gives additional insight into the factors that control carbohydrate conformation.

2. Methods

We refer to the fluorinated analog molecules as F-cellobiose and F-maltose, and the original, nonfluorinated

molecules as cellobiose and maltose. The linkage torsion angles (Fig. 1) are defined as ϕ = C4–O4–C1′–O5′, and ψ = C1′–O4–C4–C5, based on carbohydrate nomenclature.[16]

QM calculations were done with Jaguar,[17] Gaussian,[18] and GAMESS.[19] Calculations were carried out for the cellobiose analog at the HF/6-31G(d) and HF/6-311+G(d)//HF/6-31G(d)[†] levels of theory, and the B3LYP/6-31G(d) density functional method. Following completion of the cellobiose studies, preliminary studies were carried out for 28 *gg*,*gg*[‡] maltose structures with HF/6-31G(d) theory and with B3LYP theory and the 6-31G(d), 6-31+G(d), 6-311G(d), and 6-311+G(d) basis sets. There was nearly perfect (R^2 = 0.999) correlation between the B3LYP/6-31+G(d) level and the best level, B3LYP/6-311+G(d). The next best correlation with the best level, R^2 = 0.973, (Fig. 2) was provided by the least expensive level that we tried, HF/6-31G(d). The slope from that regression, 0.82, indicates the effect of neglected electron correlation in HF calculations. Thus discrepancies of just tenths of a kcal/mol would result from use of HF/6-31G(d) instead of the largest calculation if the relative energies were multiplied by 0.82. However, we did not scale these energies.

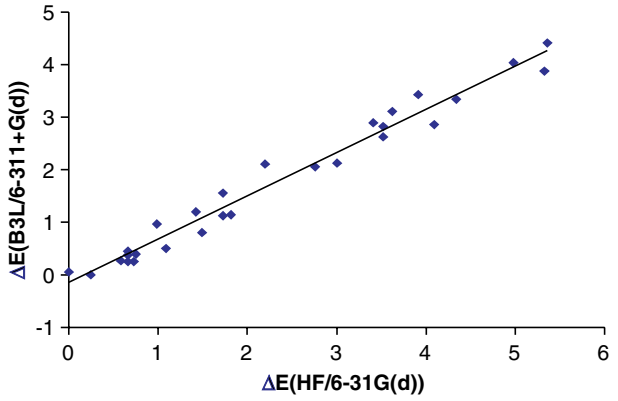

Figure 2. Correlation between HF/6-31G(d) and B3LYP/6-311+G(d) relative energies in kcal/mol for *gg*,*gg* F-maltose with different ϕ and ψ torsion angles. The slope is 0.823, the intercept is −0.15 kcal/mol, and R^2 = 0.973.

Because of the expense of these calculations, two compromises were made. The full ranges of ϕ and ψ were not covered for the F-disaccharides. Instead, only the main regions of the crystal structures were covered by grid search, accompanied by unconstrained energy minimizations in the known major minima. Increments of ϕ and ψ were 20°. We used 81 grid points for the F-cellobiose maps and 49 for the F-maltose structures. The F-cellobiose map at the B3LYP/6-31G(d) level and the F-maltose map at the HF/6-31G(d) level were also expedited by examining the minimized energies of all nine starting geometries at nine grid points in the lowest-energy region. To save computer time, those starting geometries already at high relative energies were not considered at the remaining grid points. Ultimately, 351 B3LYP/6-31G(d) minimizations at grid points were carried out for F-cellobiose instead of the $9 \times 81 = 729$ possibilities due to this reduction, and 161 HF/6-31G(d) minimizations for maltose instead of $9 \times 49 = 441$.

Our nonintegral hybrid method[7] involves calculation of MM energy maps for the complete disaccharide and MM and QM maps for the THP–O–THP analog. The MM analog map is subtracted from the MM disaccharide map, and the QM analog map is added to the difference to give the hybrid surface. Hybrid maps for cellobiose itself were constructed with 107 combinations of the hydroxyl and primary alcohol group orientations at each ϕ,ψ point. The hybrid was constructed using the B3LYP/6-311++G(d,p)//B3LYP/6-31G(d) map[20] for the THP–O–THP backbone and MM4[21] for the remainder of the molecule. The dielectric constant for the cellobiose calculation was set to 7.5, and it was 1.5 for the THP–O–THP analog. The maltose surface used 58 starting geometries and was constructed with HF/6-31G(d) theory for the THP–O–THP backbone[22] and MM4 for the remainder, again with the same dielectric constant strategy. We used the MM3(1992) hydrogen bonding parameters instead of the standard MM4 values. The hydrogen bonding parameters were important for the predictiveness of the resulting maltose hybrid surface but made little difference for the cellobiose hybrid map. The increased dielectric constant was important for both hybrid maps.

Molecules that had the backbone structure of THP rings and the appropriate linkage geometry were taken from the Cambridge Structural Database (CSD)[23] and the Protein Databank (PDB)[24] to validate the energy maps. Thus, even though the maps are created with the hypothetical fluorinated molecules, the crystal structures are from real molecules that are chemically more similar to the actual disaccharide (or the actual disaccharide itself). A detailed discussion of the selected small, cellobiose-like molecules is published.[25] The other plotted structures are discussed briefly below. Small molecule structures from the CSD are quite accurate but there is more uncertainty in the geometries of carbohydrates that are complexed with proteins in the PDB. That may contribute to the wider range of observed geometries but there could be other factors, such as distortion for some biochemical purpose. That said, even though most of the observed conformations of lactose moieties in the PDB are similar to the cellodextrins in the PDB, some lactose geometries in the PDB deviate considerably (see Ref. 26). The reason for this is unclear, and the PDB lactose data are not included in the present discussion.

[†] The // notation indicates that the molecular geometry was fully optimized at the level to the right of the //, while the reported energy is for that geometry at the theoretical level on the left of the //.

[‡] The *gg*,*gg* notation refers to the O5–C5–C6–F6 and C4–C5–C6–F6 torsion angles on the nonreducing and reducing ends, respectively. The *gg* shapes place F6-*gauche* (−60°) to O5 and +*gauche* (+60°) to C4. The *gt* designates that F6 is *gauche* to O5 and *trans* to C4, while *tg* indicates that F6 is *trans* to O5 and *gauche* to C4.

3. Results and discussion

3.1. Stability of ring shapes

Almost all experimentally observed structures of glucopyranose rings are in the $^4C_1$ shape, in agreement with our calculations.[27,28] We were unsure of the effect of fluorine substitution, so we checked on the preferred conformation of F-β-glucose. Both chairs and all six of the skew form rings[29] were generated. Each ring form had its C6–F group in all three staggered orientations. Table 1 shows the energies for the two chairs and the lowest-energy skew forms with various levels of theory. At the two HF levels of theory, the $^4C_1$ chair is not the preferred form. The correlated B3LYP method with the larger 6-31+G(d) basis set does indicate a preference for the $^4C_1$ form over the $^1C_4$ ring, but the skew form $^1S_5$ is even more stable. The results on the two chairs are in line with previous results on glucopyranose that show increased calculated stability for the $^4C_1$ shape with better basis set quality at the HF, MP2, and various DFT levels of theory.[28] The dependence on the level of theory is known, based on the *gauche* effect that is stronger at the better levels of theory, compared to the dipole–dipole interactions that favor the less frequently observed $^1C_4$ chair at the lower level theoretical levels.[30] No explanation is available for the absolute preference for the skew form. From this exercise we concluded that it is necessary to monitor the ring shapes in the disaccharide studies to make sure that our model was not spontaneously changing to its theoretically preferred forms. For all $^4C_1$ F-glucose calculations, the favored C6–F orientation was *gg*. The energy depended on the level of theory, with the relative energies for the *gt* structures being 1.03, 1.03, and 0.40 kcal/mol at the HF/6-31G(d), B3LYP/6-31G(d), and B3LYP/6-31+G(d) levels. The *tg* structures had relative energies of 1.89, 2.44, and 0.87 kcal/mol, respectively, at the three levels.

Table 1. Relative energies for both chair forms and the best nonchair form for fluorinated β-glucopyranose

Conformer	Level of theory	$E_{relative}$ (kcal/mol)
$^4C_1$	HF/6-31G(d)	0.00
$^1S_5$	HF/6-31G(d)	0.05
$^1C_4$	HF/6-31G(d)	−1.61
$^4C_1$	B3LYP/6-31G(d)	0.00
$^1S_5$	B3LYP/6-31G(d)	−0.06
$^1C_4$	B3LYP/6-31G(d)	−1.00
$^4C_1$	B3LYP/6-31+G(d)	0.00
$^1S_5$	B3LYP/6-31+G(d)	−0.03
$^1C_4$	B3LYP/6-31+G(d)	0.44

3.2. Effect of fluorine on glycosidic torsion angles

Conceivably, the highly electronegative F atoms could alter the computed ϕ,ψ torsional energies, even though there is a C–C bond between the F atom and the C–O bonds of the linkage. The assumption of an insignificant effect over this distance is inherent in most, if not all, empirical force fields, but this was still a concern. To address this, energy profiles for the molecule F–CH_2–CH_2–O–CH_3 were compared to those for HO–CH_2–CH_2–O–CH_3, in which the C–C–O–C sequence was driven through 180° with O–C–C–O and F–C–C–O held at 180°. These model fragments contain the inter-ring linkage C–O and O–C bonds as well as the C1–C2 and C2–O2H or C2–F groups. Figure 3 shows the B3LYP/6-31+G(d) energy profiles, which have negligible dependence on whether the terminal group is F or OH.

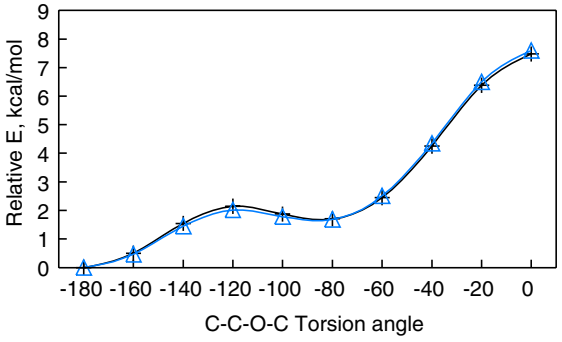

Figure 3. Comparison between the relative energy curves for CH_3–O–CH_2–CH_2–F and CH_3–O–CH_2–CH_2–OH at the B3LYP/6-31G(d) level. Points from CH_3–O–CH_2–CH_2–F are marked with △, and points from CH_3–O–CH_2–CH_2–OH are marked with +.

In addition, full ϕ,ψ maps were made with B3LYP/6-31G(d) for the three molecules X–CH_2–CH_2–O–CH_2–CH_2–Y, where X and Y = OH, X and Y = F and X = OH, Y = F. In these calculations, the X–C–C–O and O–C–C–Y sequences were held at torsion angles of −79° and −70°, and the H–O–C–C sequences were held at 180°. All of ϕ,ψ space was considered for all three molecules at the B3LYP/6-31G(d) level. All showed the global minimum in the same location and most of the differences in relative energy were less than 1 kcal/mol. In the highest energy region, the map with X and Y = OH was higher in energy than the X and Y = F map by 5.0 kcal/mol. In the important low-energy region, the differences were less than 0.5 kcal/mol but it could be said that the minima on the X and Y = F map were slightly deeper than those on the X and Y = OH map. From these calculations we conclude that the electronic effects of F are likely to be negligible in our main calculations for the F-disaccharides.

Further support for the negligible electronic effect of fluorine atoms on the conformation came from inspection of the Mulliken charges on CH_3F and CH_3OH as well as the molecule X–CH_2–CH_2–O–CH_2–CH_2–Y. The numeric value of charges for oxygen represents the sum of charges on both the oxygen and hydrogen, for comparison with the F atoms. The O in CH_3OH has a charge of −0.196, while the F in CH_3F has a charge of −0.325. In X–CH_2–CH_2–O–CH_2–CH_2–Y, the carbon atoms adjacent to the central oxygen (which represents the glycosidic oxygen in the disaccharide) have a charge of −0.24 when X and Y = F, −0.26 when X = F, Y = OH, and −0.27 when X and Y = OH. The differences among these values are small.

3.3. F-Cellobiose

Figure 4a–c show the HF/6-31G(d), HF/6-311+G(d)//HF/6-31G(d), and B3LYP/6-31G(d) maps, respectively,

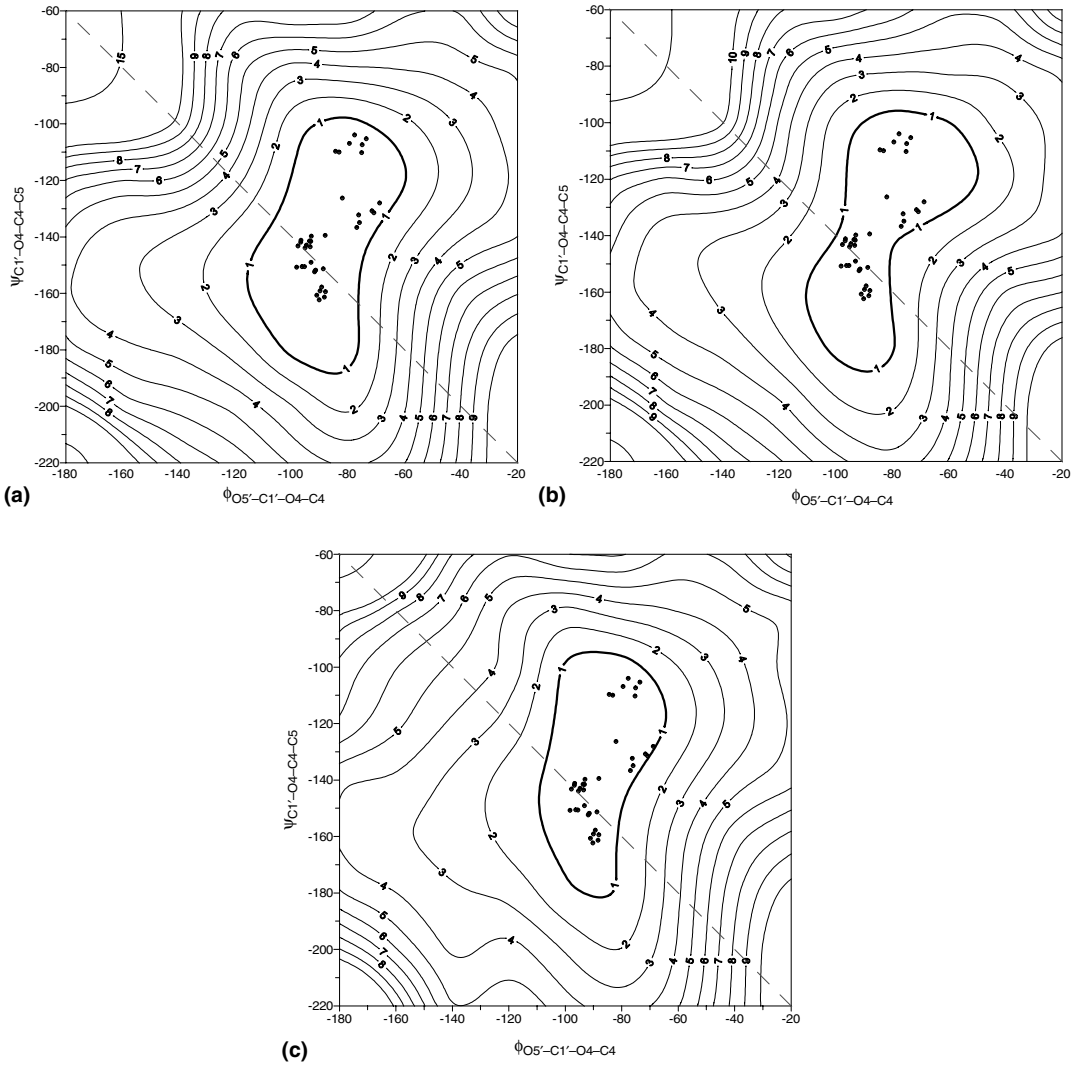

Figure 4. (a) HF/6-31G(d) energy surface for F-cellobiose with experimentally observed small molecule crystal structure geometries shown as points. (b) HF/6-311+G(d)//HF/6-31G(d) energy surface for F-cellobiose with small molecule structures. (c) B3LYP/6-31G(d) energy surface for F-cellobiose with small molecule structures.

for F-cellobiose and the conformations from the CSD. There is a small effect of the level of theory, not as large as those seen for different theory for QM maps of trehalose.[8] At the HF/6-31G(d) level, 66 of the 81 lowest-energy structures for each map point had *gg*,*gg* C6–F positions, with 9 *gg*,*tg* and 4 *gt*,*tg* pairings. That dominance of the *gg*,*gg* structures agrees with the above findings for F-glucose. The HF/6-311+G(d)//HF/6-31G(d) map had 15 *gg*,*gg*, 12 *gg*,*tg*, and 8 *gt*,*tg* structures, but 44 *gt*,*gg* structures. A few points failed to yield converged results on both maps. The B3LYP/6-31G(d) surface was composed of 64 *gg*,*gg*, 8 *gt*,*gg*, 4 *gt*,*gt*, 4 *gt*,*tg*, and 1 *gt*,*tg* structures.

On all of three F-cellobiose maps, the crystallographically observed small molecule conformations fit within the 1-kcal/mol contour. One structure, a very heavily substituted cellobiose, that is in a secondary minimum that is not on these maps (see Fig. 8, below). The dashed diagonal line indicates structures that would have internal twofold screw-axis symmetry. Because of their essentially infinite length, crystalline cellulose molecules often have this symmetry, but the small molecules cannot. In cellobiose, for example, the reducing and nonreducing residues are chemically different and cannot be related by symmetry. However, a consequence of this geometry, regardless of exact symmetry, is that the dimers, oligomers or polymers of β-1,4-linked glucopyranose residues have the shape of a flat ribbon that can pack very efficiently in crystals, like bricks in a wall. Conformations falling near this line, therefore, may be the result of distortions from crystal packing forces. Of the several surfaces that we have produced recently with QM, hybrid, and MM theory for molecules related to cellobiose,[20] the fluorinated maps show the lowest energy for the twofold screw structures. Except for the F-cellobiose maps, and the new hybrid map below (Fig. 8), the twofold lines pass through areas that have minimum energies of more than 0.8 kcal/mol. Somewhat older cellobiose maps, also made with elevated dielectric constants but with fewer starting structures, showed low energies for the twofold structures.[31]

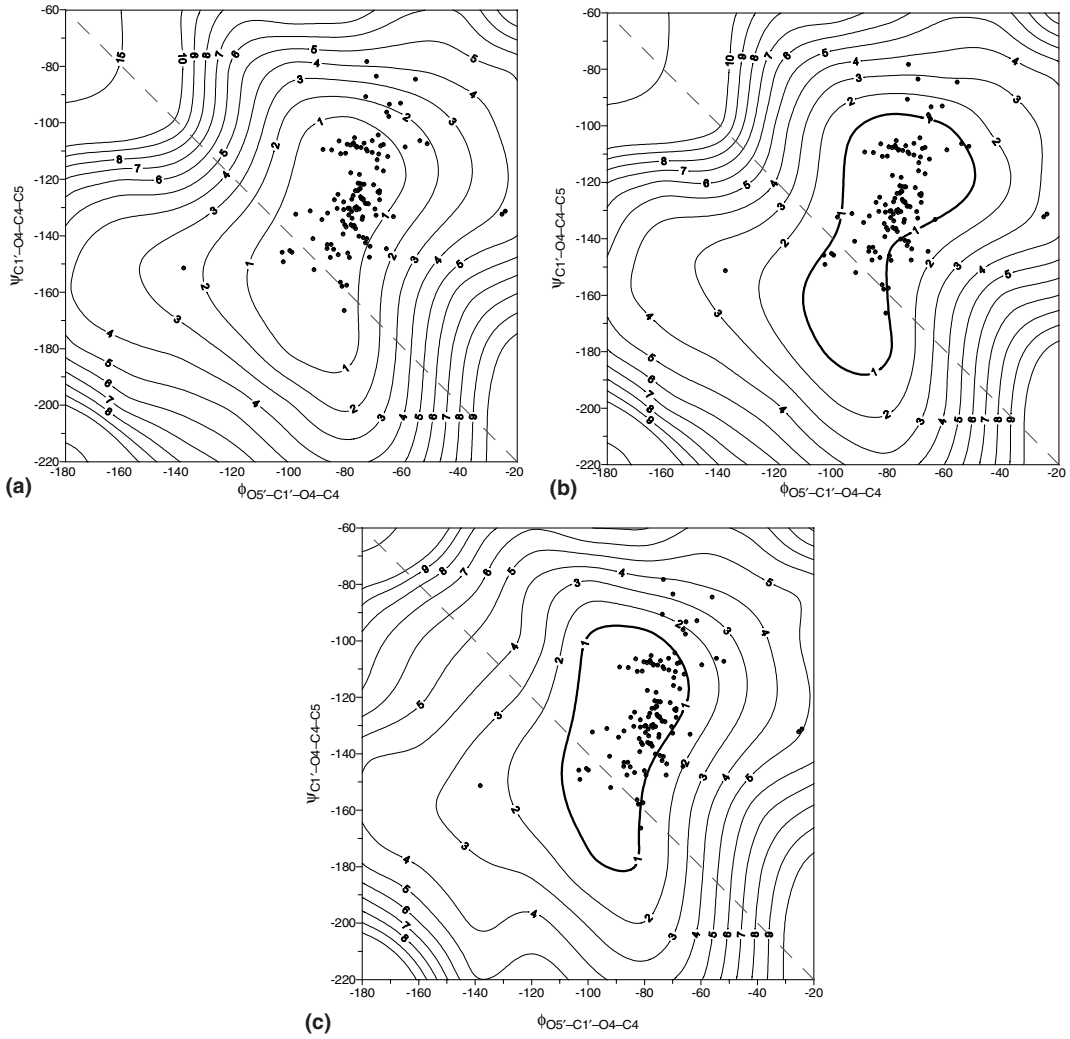

Figure 5. (a) HF/6-31G(d) energy surface for F-cellobiose with cellodextrin protein complexes. (b) HF/6-311+G(d)//HF/6-31G(d) energy surface for F-cellobiose with cellodextrin protein complexes. (c) B3LYP/6-31G(d) energy surface for F-cellobiose with cellodextrin protein complexes.

Figure 5a–c show the same energy surfaces with the conformations from cellodextrin oligosaccharides that are complexed with proteins, taken from the PDB. Some of the conformations have rather high energies on this surface but most corresponding energies are less than 1 kcal/mol. What can be concluded from these predictions of the geometry of cellobiose using QM calculations on F-cellobiose? The experimentally observed conformations of cellobiose coincide with predictions based on high-quality QM calculations for a molecule that does not permit hydrogen bonding. The various levels of theory are all reasonably predictive. The most expensive of these three calculations, B3LYP/6-31G(d), is not demonstrably superior to the others.

Table 2 shows the energies for the three main minima[31] for F-cellobiose in ϕ,ψ space (see also Fig. 8, below). There is only one observed structure in an alternative minimum but the energies for these conformations are of interest for molecules in solution and possibly, folded chain structures. The glycosidic bond angles are larger for the alternative minima, as are the energies. There is a modest dependence for both these properties on the level of theory. Presumably, B3LYP theory provides

Table 2. Relative energies (kcal/mol) at minima[a] and glycosidic angles (°) for F-cellobiose

Level of theory	Rel. *E* center of map	τ (C1′–O4–C4)	Rel. *E* right edge	τ (C1′–O4–C4)	Rel. *E* top edge	τ (C1′–O4–C4)
HF/6-31G(d)	0	117.73	2.95	120.33	3.56	120.89
HF/6-311+G(d)	0		2.90		3.39	
B3LYP/6-31G(d)	0	116.31	2.48	118.14	3.41	118.87
B3LYP/6-311+G(d)	0		2.67		3.10	
B3LYP/6-311++G(d,p)	0		2.58		3.04	

[a] The central minima are at $\phi,\psi \approx -80°, -120°$; the right edge minima at $\phi,\psi \approx 60°, -120°$; and the top (or bottom) minima at $\phi,\psi \approx -80°, -300°$. The latter regions are not shown on the F-cellobiose maps.

lower contributions to the van der Waals energies for the strained structures in the alternative minima. We note that the alternative minimum that is split by the sides of the map (not populated by experimental structures) is the location of the global QM minimum for cellobiose itself, according to Strati et al.[32,33]

3.4. F-Maltose

Based on the above results (see the paragraph in Methods on preliminary maltose studies and the low sensitivity to level of theory for F-cellobiose in Figs. 4 and 5), HF/6-31G(d) theory was deemed sufficient to make a predictive surface for F-maltose. Figure 6a–c show the energy surface for F-maltose with three sets of observed conformations from the CSD. Figure 6a contains the acyclic, small-molecule structures, fewer than half of which are inside the 1-kcal/mol contour. While the maximum energy for any crystal structure is less than 4 kcal/mol, the distribution of the observed structures is not random, but instead more or less along a diagonal line. Most of the structures at the upper right, such as crystalline maltose itself, benefit from intra-molecular, inter-residue hydrogen bonding, while those within the 1-kcal/mol contour and to the lower left do not. Thus, the latter structures either cannot form hydrogen bonds because they have no hydroxyl groups, such as maltose octaacetate, or the hydroxyl groups form inter-molecular hydrogen bonds. The map can be interpreted to indicate that observed structures in one area of ϕ,ψ space are favored by factors inherent in the F-maltose model, but that intra-molecular, inter-residue hydrogen bonding, and other effects can distort the linkage by more than 3 kcal/mol.

Figure 6b shows the locations of the very large number of conformations of the α- and β-cyclodextrins (six and seven glucose residues per macrocycle). Although there is a range of observed conformations, the center of the population is shifted from that of the acyclic structures. From this we conclude that the closure of the macrocycle restricts the flexibility of the glycosidic linkages.

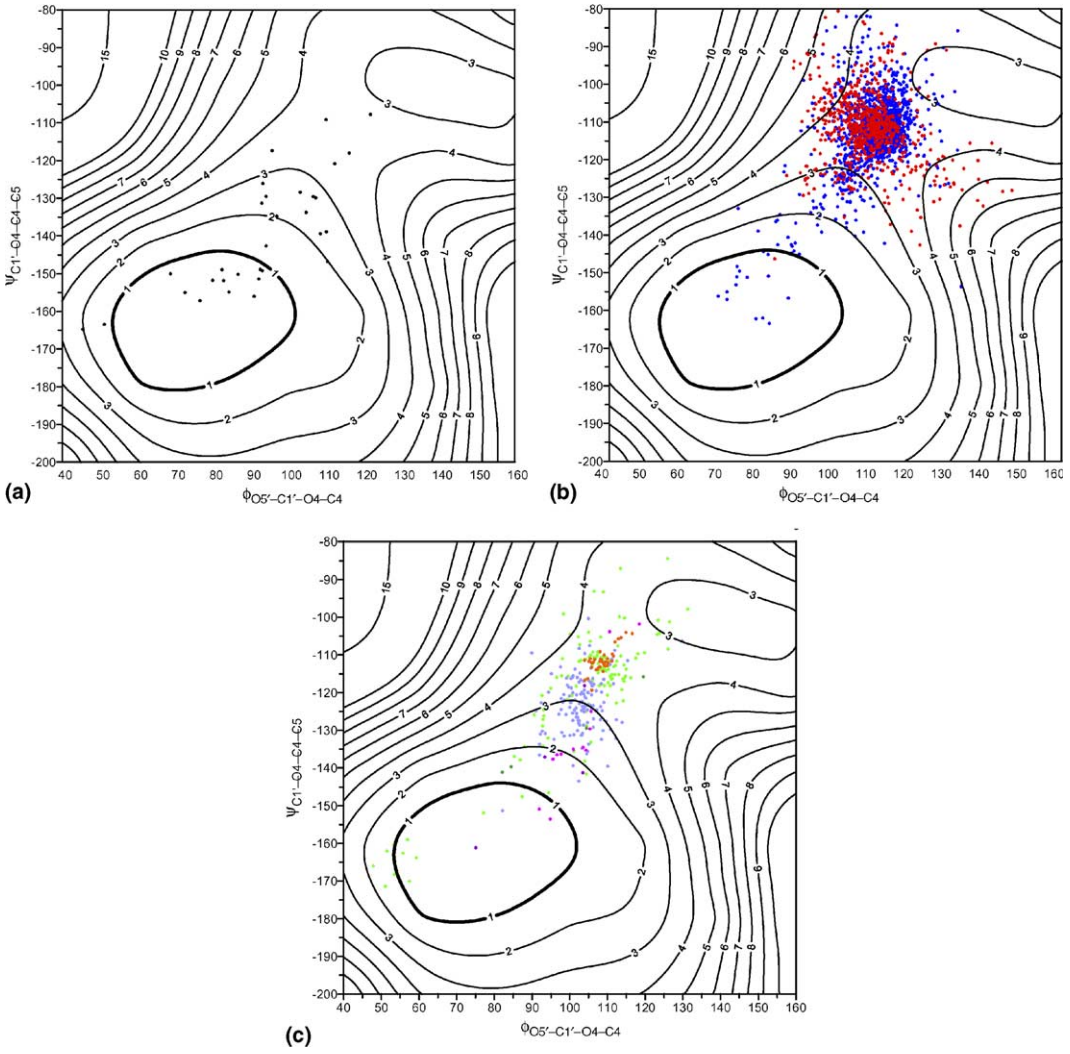

Figure 6. F-Maltose computed at the HF/6-31G(d) level: (a) with crystal structure conformations of nonmacrocyclic molecules plotted; (b) with crystal structure conformations from α- and β-cyclodextrins (red and blue, respectively); (c) with crystal structure conformations from cyclodextrins with 8 (green), 9 (gold), 10 (purple), 12 (light green), 14 (red), and 26 (light blue) glucose residues.

Many of these cyclodextrins have inter-residue hydrogen bonds between O-3 and O-2′, so it is not possible to distinguish between the effects of hydrogen bonding and closure of the macrocycle on the ϕ and ψ values.

Figure 6c shows the conformations from the larger cyclodextrins plotted on the HF/6-31G(d) energy surface. These larger molecules have different restrictions on accessible values for ϕ and ψ. In the α- and β-cyclodextrins, the C-6 groups are all on the same side of the macroring. The macrorings with 10 and 14 residues have a 'flip' in two of the linkages between rings, so that half of the C6 groups are on one side and half on the other.[34] The macrorings with 26 glucose residues have two helical segments connected with flipped geometries, so that there is an 'up' helix as well as a 'down' helix in the same molecule.[35] The flipping linkage geometries fall into a secondary minimum that appears on the full space ϕ,ψ energy maps. See Ref. 31 for an example of a full ϕ,ψ map constructed with the empirical force field, MM3, and an MM3::HF/6-31G(d) hybrid map for maltose is also available.[22] Energy values for all of ϕ,ψ space are also plotted below in Figure 8.

Figure 7 shows the minimum energy conformations for each of the nine starting geometries at two levels of theory, HF/6-31G(d) and B3LYP/6-31G(d). Also shown are the final ϕ,ψ values for the nine structures started in the secondary minimum, calculated with both levels of theory. The effect of C6–F orientation on the ϕ values is substantial, yielding minima that varied in location over a range of about 40°, similar to the 44° range of ϕ values reported by Strati et al.[32] for various starting geometries of cellobiose.

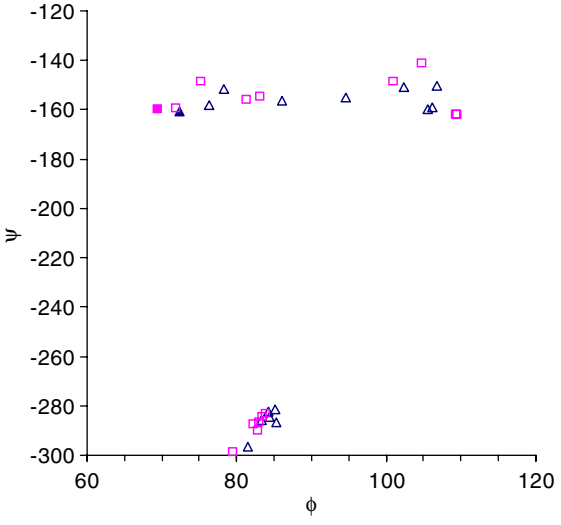

Figure 7. Locations of the minimum energy conformations for the F-maltose structures. There are 18 points in each of the two groups. The upper group aligns with the experimental structures and the main energy minimum. It depicts the ϕ,ψ values from the nine starting structures, minimized without constraints with the 6-31(d) basis set and HF and B3LYP theory. The lower group corresponds to the alternative minimum on the hybrid maltose map (Fig. 9) below. The HF/6-31G(d) data are represented by □ and the B3LYP/6-31G(d) data are marked with △. The global minima are represented with filled symbols.

3.5. Hybrid disaccharide and methylated tetrahydropyran analog QM maps

MM4::QM hybrid maps for actual cellobiose and maltose are shown in Figures 8 and 9. Both of these maps seem to be very predictive. The more rapid generation of the hybrid surfaces, compared to the F-disaccharide calculations, along with the high quality of prediction, suggests that the hybrid method is preferred for prediction of the real disaccharides. However, hybrid maps prepared at lower dielectric constants, that is, with stronger hydrogen bonding, were less predictive than these surfaces. On those maps (not shown), only the hydrogen bonded molecules have low energies. More cellobiose linkage conformations are within the 1-kcal/mol contour on its present hybrid surface than are maltose linkages. The effects of ring closure on linkage con-

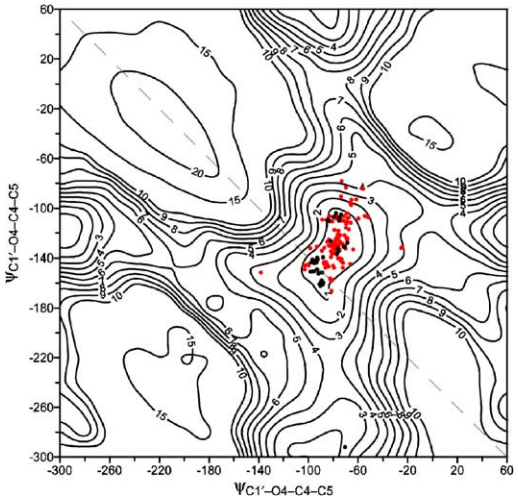

Figure 8. MM4::B3LYP/6-311++G(d,p)//B3LYP/6-31G(d) hybrid map for cellobiose with experimental conformations from Figures 4 and 5. Black and red dots represent structures from the CSD and PDB, respectively.

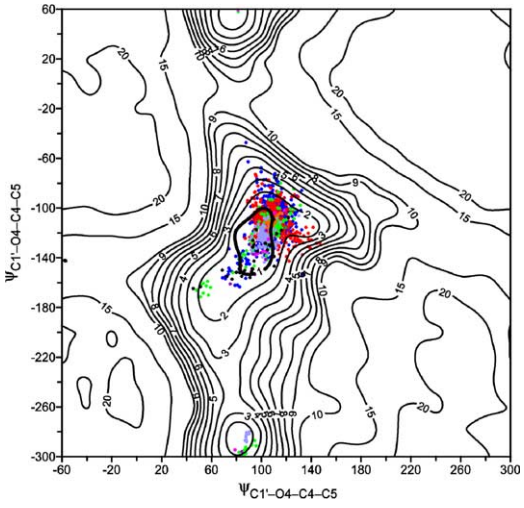

Figure 9. MM4::HF/6-31G(d) hybrid map for maltose with crystal structures from Figure 6 plotted.

formations of the cyclodextrins may play a substantial role in the explanation of this situation.

The inherent differences between di-equatorial and axial–equatorial linkages are clear in Figures 8 and 9. On the di-equatorial cellobiose map, a secondary minimum nearly as low in energy as the central minimum is split by the left and right edges. On the axial–equatorial maltose surface, that same region corresponds to energies of about 15 kcal/mol. Both surfaces have secondary minima of less than 2 kcal/mol that are split by the top and bottom edges of the maps. The (top/bottom) minima are populated with experimental structures. For cellobiose, there is only one, heavily substituted, observed structure.[36] The maltose structures in the top/bottom region are all from the larger cyclodextrins. Those linkages relieve the strain that would occur in the glucose rings[37] if all linkages in the larger molecules had geometries similar to those in the small cyclodextrins.

The map in Figure 10 is from a model that has methyl groups attached to tetrahydropyran (THP) rings in the positions of primary alcohol groups. In another description, the model molecule is the same as maltose with all OH groups replaced by hydrogen. Thus, this maltose analog is sterically more complete than the simple THP–O–THP analogs studied earlier.[11] The fit of the crystal structures on this surface is comparable to that on the F-maltose surface, although the structures in the secondary minimum of the Me,Me–THP–O–THP surface are slightly higher in energy than reported above for the F-disaccharide surface. This overall comparability supports the validity of our assumptions that the fluorines would not affect the torsional energies for the glycosidic linkage bonds.

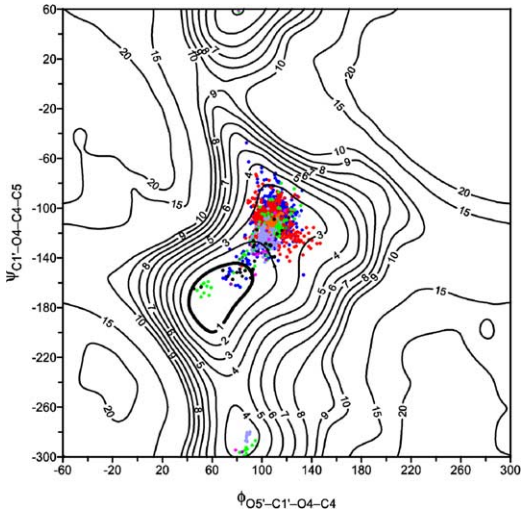

Figure 10. HF/6-31G(d) full map for 5,5′-dimethyl THP–O–THP maltose analog with crystal structures from Figure 6.

4. Conclusions

In this work, we have explored the possibility of using F-disaccharides for predictive modeling of disaccharides and larger, more complex carbohydrates. This was undertaken because of the substantial multiple minimum problem inherent in sugars that have many hydroxyl groups. There are so many combinations of *exo*-cyclic group orientations possible, compared to just three per monomeric unit in F-disaccharides. Although our earlier work had shown that the torsional energies for the glycosidic bonds were very important in determining the observed conformations, that work was based on simple THP–O–THP analogs. Those analogs had obvious steric shortcomings that could be partly relieved with an analog based on substitution of all hydroxyl groups with fluorine atoms.

Whether these analogs are suitable was tested by studying the ring shape stability and the influence of the fluorine atoms on the glycosidic torsional bond energies. These issues were not obstacles, and all of the F-cellobiose energy surfaces were very predictive of the observed conformations. On the other hand, the F-maltose map was only predictive of structures that lack inter-residue hydrogen bonds. Structures having hydrogen bonds were outside of the 2-kcal/mol contour. From this observation, it appears that there is a strong (2–3 kcal/mol) hydrogen bonding effect on the linkage geometry if a hydrogen bond is formed. The cyclodextrin and cycloamylose structures added a very large number of linkage geometries to the study, including some that populate an alternative minimum. By far, most of their linkage geometries were outside of the 2-kcal/mol contour on the F-maltose surface. Although there could be a strong effect on the linkage geometry from closure of the macrocycle, the geometries in the main minimum are not so different from those of the nonmacrocyclic maltodextrins. Hydrogen bonding is also a likely factor in the conformational preference for maltose linkages in macrocycles.

Additional energy surfaces were constructed for the disaccharides themselves with our MM4::QM hybrid method. They took less computer time than for the F-disaccharides, and gave a substantially more predictive result for maltose molecules, although there was little difference among the various cellobiose-type maps herein. Finally, a map for another maltose analog that lacked hydrogen bonding ability was quite similar to the F-maltose map. This confirmed that the major deficiency of the F-maltose model was its total lack of hydrogen bonding ability, and not an effect of the fluorine atoms on the glycosidic torsional energies or some new, nonbonded interactions of the fluorine atoms.

Acknowledgements

We thank Professor E. D. Stevens, University of New Orleans, for suggesting the study of fluorinated disaccharides. Professor N. L. Allinger, University of Georgia, Athens, participated in helpful discussions. This work was supported by the US Department of Agriculture, OTKA Hungary (T034764), and the Graz University of Technology.

References

1. Rao, V. S. R.; Sundararajan, P. R.; Ramakrishnan, C.; Ramachandran, G. N. In *Conformation in Biopolymers*; Academic: London, 1963; Vol. 2.
2. Lemieux, R. U.; Bock, K.; Delbaere, L. T. J.; Koto, S.; Rao, V. S. R. *Can. J. Chem.* **1980**, *58*, 631–653.
3. Tvaroska, I.; Perez, S. *Carbohydr. Res.* **1986**, *149*, 389–410.
4. Tvaroška, I.; Václavík, L. *Carbohydr. Res.* **1987**, *160*, 137–149.
5. French, A. D. *Biopolymers* **1988**, *27*, 1519–1525.
6. Naidoo, K. J.; Brady, J. W. *J. Am. Chem. Soc.* **1999**, *121*, 2244–2252.
7. French, A. D.; Kelterer, A.-M.; Cramer, C. J.; Johnson, G. P.; Dowd, M. K. *Carbohydr. Res.* **2000**, *326*, 305–322.
8. French, A. D.; Johnson, G. P.; Kelterer, A.-M.; Dowd, M. K.; Cramer, C. J. *J. Phys. Chem. A* **2002**, *106*, 4988–4997.
9. Stortz, C. A. *Carbohydr. Res.* **1999**, *322*, 77–86.
10. French, A. D.; Kelterer, A.-M.; Johnson, G. P.; Dowd, M. K.; Cramer, C. J. *J. Mol. Graph. Modell.* **2000**, *18*, 95–107.
11. French, A. D.; Kelterer, A.-M.; Johnson, G. P.; Dowd, M. K.; Cramer, C. J. *J. Comput. Chem.* **2001**, *22*, 65–78.
12. Taylor, N. F. *ACS Symp. Ser.* **1987**, *374*, ix–x.
13. Withers, S. G.; Street, I. P.; Percival, M. D. *ACS Symp. Ser.* **1987**, *374*, 59–77.
14. Tsuchiya, T. *Adv. Carbohydr. Chem. Biochem.* **1990**, *48*, 91–277.
15. Kouwizjer, M. L. C. E.; Grootenhuis, P. D. J. *J. Phys. Chem.* **1995**, *99*, 13426–13436.
16. McNaught, A. D. *Carbohydr. Res.* **1997**, *297*, 1–90; Also published in other places, including. *Pure Appl. Chem.* **1996**, *68*, 1919–2008.
17. Jaguar 5.5, Schrödinger, L. L. C., Portland, OR, 1991–2003.
18. Gaussian 98 (Revision A.1x), Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski,V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian, Inc., Pittsburgh, PA, 2001.
19. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
20. French, A. D.; Johnson, G. P. *Cellulose* **2004**, *11*, 449–462.
21. Allinger, N. L.; Chen, K.-H.; Lii, J.-H.; Durkin, K. A. *J. Comput. Chem.* **2003**, *24*, 1447–1472.
22. French, A. D.; Johnson, G. P.; Kelterer, A.-M.; Dowd, M. K.; Cramer, C. J. *Int. J. Quantum Chem.* **2001**, *84*, 416–425.
23. Allen, F. H. *Acta Crystallogr. Sect. B* **2002**, *58*, 380–388.
24. Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. *Acta Crystallogr. Sect. D* **2002**, *58*, 899–907.
25. French, A. D.; Johnson, G. P. *Cellulose* **2004**, *11*, 5–22.
26. French, A. D.; Johnson, G. P. The Shapes of Cellulose. In *Cellulose: Biosynthesis, Structure, and Applications*; Brown, R. M., Jr., Saxena, I., Eds., in preparation.
27. Barrows, S. E.; Dulles, F. J.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *Carbohydr. Res.* **1995**, *276*, 219–251.
28. Csonka, G. I.; Éliás, K.; Csizmadia, I. G. *Chem. Phys. Lett.* **1996**, *257*, 49–60.
29. Dowd, M. K.; French, A. D.; Reilly, P. J. *Carbohydr. Res.* **1994**, *264*, 1–19.
30. Allinger, N. L. Conversation of August 11, 2004. The *gauche* effect has been studied but specific references to the F–C–C–F case are not covered in those papers. See for instance Wolfe, S. *Acc. Chem. Res.* **1972**, *5*, 102–111.
31. Mendonca, S.; Johnson, G. P.; French, A. D.; Laine, R. A. *J. Phys. Chem. A* **2002**, *106*, 4115–4124.
32. Strati, G. L.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2002**, *337*, 1833–1849.
33. Strati, G. L.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2002**, *337*, 1851–1859.
34. Saenger, W.; Jacob, J.; Gessler, K.; Steiner, T.; Hoffmann, D.; Sanbe, H.; Koizumi, K.; Smith, S. M.; Takaha, T. *Chem. Rev.* **1998**, *98*, 1787–1802.
35. Nimz, O.; Gessler, K.; Usón, I.; Sheldrick, G. M.; Saenger, W. *Carbohydr. Res.* **2004**, *339*, 1427–1437.
36. Ernst, A.; Vasella, A. *Helv. Chim. Acta* **1996**, *79*, 1279–1294.
37. French, A. D.; Murphy, V. G. *Polymer* **1977**, *18*, 489–494.