Quantum Mechanics Studies of the Intrinsic Conformation of Trehalose

Alfred D. French,*,† Glenn P. Johnson,† Anne-Marie Kelterer,‡ Michael K. Dowd,† and Christopher J. Cramer§

Southern Regional Research Center, U. S. Department of Agriculture, 1100 Robert E. Lee Blvd., New Orleans, Louisiana 70124, Institut für Physikalische und Theoretische Chemie, Technische Universität Graz, Technikerstrasse 4, A-8010 Graz, Austria, and Department of Chemistry and Supercomputer Institute, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455-0431

Received: January 16, 2002

To resolve controversies over the intrinsic shape of the disaccharide α,α-trehalose and the magnitude of solvent effects, energy surfaces have been computed at different levels of molecular orbital and density functional theory. All quantum mechanical (QM) levels agree that the gauche linkage conformations observed by experimental crystallographic and solution studies are 5−7 kcal/mol lower in energy than that of the trans shape. This is quite different than the findings of the two most recent classical force field studies. In those projects, the trans shape was preferred by 3.3 kcal/mol, and it was inferred that a strong solvent effect was responsible for the gauche experimental conformations. In the QM work, a strong solvent effect is not needed to explain the preference for the gauche form because the trans shapes already have higher energy. A QM continuum model of aqueous solvation has only small effects on the torsional energy surface. The best rationalization of 24 values of the linkage torsion angles from small-molecule crystal structures is provided by HF/6-311++G**//B3LYP/6-31G* theory, a level that under-predicts the strength of hydrogen bonds.

Introduction

The chemically symmetric disaccharide α,α-trehalose (α-D-glucopyranosyl-(1,1)-α-D-glucopyranoside, Figure 1), is one of three trehalose isomers. The α,α-isomer, being dominant, is often simply called trehalose. It is the primary energy storage for some insects[1] and is found in the "honeydew" produced by insects such as aphids and white flies. While honeydew is a cause of "sticky cotton" that disrupts the spinning of cotton yarns, trehalose is often thought of in a positive way, being a natural protector against freezing and drought stress.[2,3] Originally called mushroom sugar, or mycose, it has been exploited for testing theories about sweetness.[4] It is also of fundamental theoretical interest, having seven alternating carbon and oxygen atoms in the sequence of atoms that contains the inter-ring, glycosidic linkage bonds. In that sequence, there are four different torsion angles to which the general anomeric effect should apply. Because five of the seven atoms are involved in more than one C−O−C−O torsion angle there is some interest regarding the ideal values of these torsion angles and their conformational energy profiles.[5,6]

Trehalose has been studied previously with several methods. There is agreement on its conformation in small-molecule crystals. All of the 24 available linkages in such crystals of trehalose itself, its complexes, derivatives and variants have gauche (roughly 60°) values of the ϕ and ϕ' torsion angles O-5−C-1−O-1−C-1′ and O-5′−C-1′−O-1−C-1. (See the paragraph on Nomenclature in Methods, Table 1, and Figure 1.)[4,7−24] The lowest value of either ϕ or ϕ' in crystalline trehalose moieties is 45°, in a salt complex of allo-trehalose.[7] The highest is 91° in a heavily derivatized example.[11] A crystalline complex of trehalose with maltoporin,[25] a cell-wall protein, is an exception, with reported ϕ and ϕ' values of about 123°. Estimates have also been made by NMR[26] and optical rotation.[27] Those solution-phase results are consistent with the small-molecule crystal structure results. Except for the protein complex, all of the observed ϕ and ϕ' values are consistent with ordinary exo-anomeric effects. Interestingly, the geometries from small molecule crystals do not permit interresidue, intramolecular hydrogen bonds that are often found for other crystalline disaccharides.

Theoretical calculations have also been employed. Because biological tissues that contain water benefit from the cryoprotective and antidesiccative properties of trehalose, three different groups have used solvated molecular dynamics to learn how trehalose interacts with surrounding water.[28,29,30] The first molecular dynamics studies, by Donnamaria et al., had most frequent ϕ and ϕ' values of 93°. Even though they did not contrast their solution phase results with a vacuum-phase analysis, they conjectured that the water lattice might be responsible for the simulated and observed trehalose conformations.

Two more recent studies[29,30] did include analyses of the vacuum phase. Unlike the gauche conformations (66°, 66°,[29] and 56°, 65°[30]) found in their solvated dynamics studies, a trans arrangement (ϕ and ϕ' values of about 180°) was the preferred shape for their isolated molecules. Their trans shape (Figure 1a) permits an interresidue hydrogen bond between the O-2 hydrogen and O-2′, and gave an energy 3.3 kcal/mol lower than the gauche shapes at the secondary minimum. Because the experimentally known shapes are gauche, these authors inferred

* To whom correspondence should be addressed. Southern Regional Research Center, P.O. Box 19687, New Orleans, Louisiana 70179-0687 or 1100 Robt. E. Lee Blvd., New Orleans, Louisiana 70124. Telephone: (504) 286-4410. Fax: (504)-286−4217. E-mail: afrench@nola.srrc.usda.gov.

† Southern Regional Research Center, U. S. Department of Agriculture.

‡ Institut für Physikalische und Theoretische Chemie, Technische Universität Graz.

§ Department of Chemistry and Supercomputer Institute, University of Minnesota.

10.1021/jp020126d CCC: $22.00
Published on Web 04/24/2002

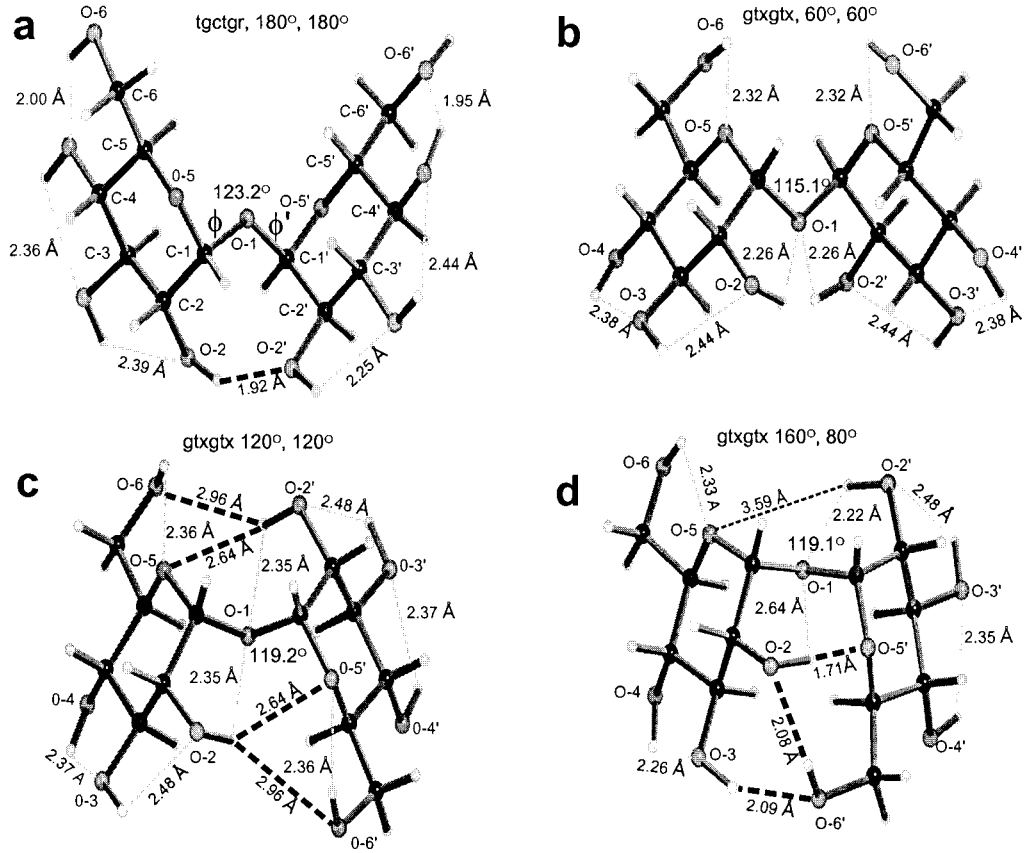

Figure 1. Drawings of various conformations discussed in the text, optimized by B3LYP/6-31G* theory at the cited grid points. The intra- and interresidue hydrogen bonds are shown as gray and heavy black dashed lines, respectively. All O−H···O angles exceed 90°. Also shown are the glycosidic angles. (a) The TGRTGC structure at ϕ, $\phi' = 180°$, 180°, proposed to be the global minimum by Liu et al.[29] and by Engelsen and Pérez.[30] Carbon atoms and some oxygen atoms are labeled. The donor−acceptor chain of hydrogen bonds is of particular interest. (b) The GTXGTX structure in the ϕ, $\phi' = 60°$, 60° shape, near the global minimum for several maps. The oxygen atoms are labeled. (c) The GTXGTX structure in the 120°, 120° shape. The proposed long range-hydrogen bonds are indicated by the heavy dashed lines. The oxygen atoms are labeled. (d) The GTXGTX, 80°, 160° structure that gave the lowest B3LYP/6-31G* energy for any gridpoint. The O-2′−H···O-5 distance (3.59 Å) is lengthened by the torsional rotations from the 120°, 120° structure. That longer distance is shown by the thin black dashed line, while its counterpart O-2-H···O-5′ distance, 1.71 Å, is indicated by a thick black dashed line.

that there is a large effect of water on the conformation of trehalose, i.e., that the intermolecular hydrogen bond of the trans form would be broken if the molecule were put into water from the gas phase, and the molecule would then assume the gauche shape. With alternative intermolecular hydrogen bonding opportunities in solution, there would be no energy penalty for breaking the O-2···O-2′ hydrogen bond, the major stabilizing factor for the trans structure. In fact, because the 3.3 kcal/mol of stabilization for the trans shape is less than that provided by an intramolecular hydrogen bond, there would be some degree of strain that would be relieved by a transition to the gauche form once the hydrogen bond was broken.

The inferred solvent effect was supported by analysis of the solvated dynamics trajectories, where a water molecule was bridging and further stabilizing the two residues of the disaccharide in the gauche form 48% of the time.[29] In some instances, the bridging water was positioned similarly to the arrangement in crystalline trehalose dihydrate.[18] One way for the gauche torsion angles to occur in crystals of even anhydrous trehalose[10] would be for chaperone water molecules to deliver the trehalose to the crystal surface in the gauche conformation, leading to crystal growth in the conformation of the otherwise secondary energy well.

A decade ago, our classical modeling studies[31] had given a global minimum for trehalose with ϕ and ϕ' values of 73°, i.e., in the midst of the observed gauche shapes. On that energy surface, the trans structure was predicted to be higher in energy by more than 10 kcal/mol, a result that is remarkably different from the newer[29,30] predictions. Although solvent was not explicitly included in our work, a dielectric constant of 4.0 had been used to approximate the effects of a condensed-phase environment, instead of a vacuum. That dielectric constant, instead of the value of 1.5 used for vacuum studies with MM3,[32] reduced the strength of a hydrogen bond to about 2 kcal/mol. This has been useful for predictions of condensed phase conformations when isolated molecules are the basis for calculations.[33] Another difference between our prior work and the two most recent studies was that we did not consider the "forbidden"[34] tg (trans to O-5, gauche to C-4) conformation of the primary alcohol group. That orientation allows O-6 to be a distant part of the hydrogen bonding network that stabilizes the trans linkage conformation.

These issues of O-6 orientation and dielectric constant are not insignificant, but more of the discrepancy between the studies resulted because different empirical potential functions were used. Liu et al., and Pérez and Engelsen, used the all-atom CHARMM-type potential for carbohydrates of Ha et al.,[35] our previous work used Allinger's (all-atom) MM3 potential, and Donnamaria et al. used the united-atom GROMOS force field.[36] Unlike the other workers, Donnamaria et al. kept the pyranose rings rigid and used no potential terms for the ϕ and ϕ' torsions. The Ha et al. potential uses the same torsional

TABLE 1: Crystal Structure Results for Trehalose

refcode	ref	ϕ	ϕ'	compd
ALTRCA	7	45.1	57.1	α-D-Allopyranosyl-α-D-allopyranoside calcium chloride pentahydrate
BABHUH	8	62.7	85.9	6-*O*-Acetyl-2-azido-3,4-di-*O*-benzyl-2-deoxy-α-D-glucopyranosyl-2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside
DBTRHA	9	77.4	80.6	6,6′-Dibromo-6,6′-dideoxy-α,α-trehalose-hexa-acetate chloroform solvate
DBTRHA		79.5	73.7	
DEKYEX	10	60.1	60.8	α,α-Trehalose
DUHXEJ	11	73.7	79.5	α,α-(2,3,4-Tri-*O*-methyl-6-methanesulfonyl)-glucopyranosyl-1-*O*-(2′,3′,4′-
DUHXEJ		80.0	91.4	Tri-*O*-methyl-6′-methanesulfonyl)-glucopyranose
HIDRIV	12	72.6	72.6	3,3′,4,4′-Tetra-*O*-benzoyl-2,2′,6,6′-tetradeoxy-α,α-ribotrehalose
HIVSEK	13	76.0	63.8	6,6′-Ditosyl-α,α-trehalose methanol solvate monohydrate
HIVSEK		77.3	67.0	
LESZAK	14	83.0	89.1	2,2′,3,3′-Tetra-*O*-acetyl-4,4′,6,6′-tetradeoxy-α,α-xylotrehalose
LETTEJ	4	64.5	75.6	3,3′-Dideoxy-α,α-arabino-trehalose monohydrate
LINNUR	15	64.9	57.3	3-O−Benzoyl-4,6:4′,6′-di-*O*-benzylidene-2,2′-dideoxy-α,α-ribotrehalose
LINPAZ	16	81.3	84.4	2,2′,3,3′-Tetra-*O*-benzoyl-6,6′-dideoxy-4,4′-di-*O*-mesyl-6,6′-dithiocyanato-α,α-trehalose
TRECAB	17	77.0	77.0	α,α-D-Trehalose-calcium bromide monohydrate
TREHAL01	18	61.8	74.9	α,α-Trehalose dihydrate
WERPEO	19	77.2	81.7	2,2′3,3′-Tetra-*O*-acetyl-6,6′-dichloro-4,4′,6,6′-tetradeoxy-α,α-trehalose
WERPEO		73.6	83.3	
YODSOZ	20	77.3	77.3	4,4′,6,6′-Tetrachloro-4,4′,6,6′-tetradeoxy-α,α-galacto-trehalose
YOXFOG	21	74.7	75.5	α,α-Allo-trehalose trihydrate
YOXFOG		73.1	74.8	
YOXFUM	22	66.3	62.4	α,α-Galacto-trehalose
8ACEAC[a]	23	79.5	77.7	Trehalose octaacetate ethyl acetate solvate
8ACETOH[a]	24	76.4	80.2	Trehalose octaacetate ethanol solvate
protein complex				
1MPQ	25	122.0	124.0	α,α-Trehalose in complex with maltoporin

[a] Refcodes for these compounds were not available.

parameters for all C−O single bonds, whereas MM3 uses distinct torsional parameters for C−O bonds in anomeric sequences.

The purpose of the present paper is to shed further light on the conformation of the global minimum structure for the isolated trehalose molecule and thus on whether the experimentally observed shape results from a large solvent effect. To accomplish this, we have resorted to ab initio and density functional quantum mechanical (QM) levels of theory. Until now, these time-consuming methods have not been used to construct energy surfaces for disaccharides. Earlier, we had studied an analogue of trehalose based on the tetrahydropyran rings of the disaccharide (THP−O−THP).[37] At the HF/6-31G* level, its trans conformation was disfavored by about 9 kcal/mol. That agreed with earlier PCILO semiempirical QM results for the same molecule by Tvaroška and Václavík.[5] Thus, adding the exo-cyclic groups to the backbone would have to lower the energy of the trans structure by about 12 kcal/mol relative to the energy of the gauche form to give a surface comparable to the vacuum surfaces based on the Ha et al. potential.[29,30] Some of our other prior work showed that the changes to the torsional energy surface were minimal when the contributions from the exo-cyclic groups were added by means of a hybrid QM/MM method.[33] However, none of the prior studies used purely QM models to address the intrinsic conformation of trehalose with all its exo-cyclic groups. Here, we use QM levels of theory to make energy surfaces for functionally complete trehalose.

Methods

Nomenclature. Carbohydrate nomenclature[38] is used, with the anomeric centers being labeled C-1 and C-1′, respectively and the ring heteroatoms labeled as O-5 and O-5′. Also, as is common in structural descriptions of carbohydrates, gauche and trans are used instead of syn and anti, respectively. Trehalose, as described in the Introduction, and the other sugars mentioned, occur naturally with their pyranosyl residues in the D-configuration. However, the results for comparable analyses in a parallel L-universe would be identical except for changes in sign of the torsion angles, so the specific configurations are not mentioned repeatedly.

Calculations. It has previously been noted that there is a substantial dependence of relative conformational energies on the level of theory employed in QM calculations on glucose.[39] Lii, Ma, and Allinger proposed that, when selecting a QM level of theory for monosaccharides, an important criterion is the correct reproduction of the energy of hydrogen bonding for the water dimer.[40] They found that geometry optimization with B3LYP/6-31G* theory gives reasonably good structures, but it overestimates the hydrogen bond stabilization. Enthalpies of water dimer formation were 7.7 kcal/mol, some 50% larger than a reasonable value. With diffuse functions and a larger basis set, e.g., the B3LYP/6-311++G** level, most of the basis set superposition that causes this error is avoided[41] but energy minimization is very time-consuming. Therefore, we first optimized the geometry with B3LYP/6-31G* theory, holding only ϕ and ϕ' constant. We then carried out a single-point calculation of the B3LYP/6-311++G** energy, a method that gives a water dimerization energy of 5.0 kcal/mol. Lii, Ma and Allinger showed that such B3LYP/6-311++G**//B3LYP/6-31G* energies for monosaccharides were reasonably similar to the fully optimized B3LYP/6-311++G** values. Momany and Willett also found that geometric differences were generally very small between the structures of the disaccharide α-maltose that were optimized at the B3LYP/6-31G* and B3LYP/6-311++G** levels of theory.[42] To test our prior observations that reduced hydrogen bonding strength in gas-phase calculations is useful for predicting condensed phase structures,[33] here we also apply a level of theory that predicts hydrogen bonding to be somewhat

too weak. HF/6-311++G** calculations give an energy of water dimerization of 3.5 kcal/mol for the B3LYP/6-31G* geometry. The QM calculations were done with Jaguar,[43] using the standard termination criteria. The torsional potential energy surfaces were generated from the individual points using the contouring program, SURFER.[44]

Starting Structures. Two characteristics of trehalose reduce the number of initial structures that must be evaluated by QM calculations to obtain a representative ϕ and ϕ' energy surface. The first trait is the small size of its accessible conformational space according to our earlier work. Therefore, we only used ϕ and ϕ' values from 20° to 200°. With 20° increments, there are 100 points to be calculated, about 30% of what is required for most disaccharides.

The second expediting trait of trehalose is its internal C_2 molecular symmetry which reduces the number of unique structures that must be studied. Large differences in the calculated energy can result from changing the orientations of the rotatable groups, which may or may not indulge in various intramolecular hydrogen bonds.[45,46] Trehalose has 10 rotatable groups, so finding the lowest-energy arrangement is a challenge, especially so because different arrangements of the exo-cyclic groups are optimal for different ϕ and ϕ' points. Brute force examination of all possibilities at the QM level is currently impossible, and some shortcuts are needed to make a reasonable approximation to the adiabatic surface. By exploiting the symmetry, the impact of the shortcuts can be diminished in two ways.

If starting geometries of both individual monosaccharide rings are identical, then for example, the energy for $\phi = 80°$ and $\phi' = 160°$, will be identical to the energy for $\phi = 160°$ and $\phi' = 80°$. Thus, it is only necessary to calculate the energy for one member of such pairs, plus the structures with $\phi = \phi'$. For our 100 point surface, only 55 calculations suffice. Alternatively, if the two monosaccharide residues have different rotating group orientations, then all 100 points must be calculated because the resulting energy surface will be asymmetric. In this case, the benefit of symmetry is that another surface can be generated by just interchanging the values of ϕ and ϕ'. Given the limited conformational space and the symmetry, a study of α,α-trehalose requires about one-sixth of the calculations needed for most other disaccharides of glucose.

To determine whether the gauche or trans form is at the global minimum in energy with a limited number of starting structures, it is essential to include the one that best stabilizes the trans conformation in the work of Liu et al.[29] and by Engelsen and Pérez.[30] That structure (Figure 1a) has its primary alcohol groups in tg orientations and the secondary hydroxyl groups on the left residue in the reverse-clockwise ("R") direction, making a crown of hydrogen bonds around that glucose ring. Thus, O-4−H points to O-3, and O-3−H points to O-2, whereas O-2−H points at O-2′ on the other residue when ϕ and ϕ' are trans. (This definition of reverse-clockwise depends on the glucose ring being in the standard viewing orientation with the C-5−O-5 bond at the back, and C-1 on the right.) The hydroxyl groups on the other residue are oriented clockwise ("C"), forming the alternative crown. This starting structure is denoted below as TGRTGC. In this model, a donor−acceptor chain of hydrogen bonds spans the entire molecule from one primary alcohol group to the other.

The symmetry-related surface for the asymmetric TGRTGC structure is for a TGCTGR structure. When the lowest energy for each ϕ and ϕ' value is taken, regardless of whether the TGCTGR or the TGRTGC structure furnished the energy value, the resulting surface is symmetric. This combined surface is one that we discuss below.

Another starting arrangement of the exo-cyclic groups (Figure 1b) was chosen based on observations of glucose models[47] and crystal structures. Both glucose residues in this second set of structures have the same arrangement of the exo-cyclic groups, with O-4−H and O-3−H oriented to point reverse-clockwise. However, the O-2−H groups are oriented trans to C-2−H, allowing donation of the O-2−H proton to the glycosidic oxygen atom, as well as reception of the proton from O-3. We call this the "X" arrangement of the hydroxyl groups. Of the crystal structures in Table 1, trans values of the H−C-2−O-2−H torsion angle occur for nine of the 16 equatorial O-2−H groups. The primary alcohol groups are positioned gauche to O-5 and trans to C-4 (the gt orientation), and the O6 hydroxyl hydrogen is pointed toward the ring oxygen atom. O-6 is frequently found in the gt conformation by experiment. After minimization of this structure (denoted GTXGTX) and the TGCTGR structure at the B3LYP/6-31G* level, each glucose residue was rigidly rotated in 20° increments to every one of the ϕ and ϕ' conformations between 20° and 200°.[48] These structures then served as the starting geometries used for the maps shown in Figures 2 and 3, below.

Six more sets of starting geometries were used. In two sets, the difference from the TGRTGC and GTXGTX sets above was that the geometry at each ϕ and ϕ' point was optimized with AMBER* [49] in MacroModel's batchmin 7.1 program[50] before carrying out the QM calculations. The remaining four sets were based on searches with MM3 calculations[51] and others using the AMBER* potential energy function in MacroModel. Similar methods to select structures for subsequent QM calculations have been used by McCarren et al., for monosaccharides,[52] and by Csonka and Sosa who did some HF/6-31G* calculations on trisaccharides.[53] However, the use of these techniques at each point on a ϕ, ϕ' map for a disaccharide is apparently novel.

These four searches used models with initial glycosidic bond angles of 150°. One set was derived by combined low-mode and Monte Carlo multiple minimum conformational searching (LMCS and MCMS) in a vacuum using MacroModel and the AMBER* potential and a dielectric constant of 1.0. The structures chosen for QM calculations had the lowest AMBER* energy after searches at each of the 55 unique ϕ, ϕ' points. Three other sets of 55 structures were taken from geometries supplied by the rules-based conformational generator in Chem-X.[54] Energies for 83 different geometries were calculated with MM3 at dielectric constants of 1.5, 3.5 and 7.5, and the structures having the lowest minimized energy at that ϕ, ϕ' point at each dielectric constant were then used as input for the QM calculations. These six individual maps are not shown as they are roughly similar to those from the first two starting structures.

The final energy surfaces (Figures 4 and 5, below) are based on the lowest energy at each ϕ, ϕ' point from *any* surface, i.e., after consideration of the eight calculated values. Each set of starting models contributed to one or more of the final maps, except for the structures selected by MM3 at a dielectric constant of 7.5.

Energies for the THP−O−THP analogues were also calculated at the same levels of theory. These surfaces (not shown) were all very similar to those from our HF/6-31G* calculations,[37] with discrepancies in their low-energy areas limited to about 0.5 kcal/mol. Solvation energies calculated with the SM5.42R model[55] were evaluated at the HF/MIDI![56] level of theory. The surface for the relative energy of solvation for the TGRTGC/TGCTGR structures is shown below in Figure 6.

Results and Discussion

The B3LYP/6-31G* energy surface for the TGRTGC/TGCTGR starting models is shown in Figure 2a. The trans structure is higher in energy than the gauche conformation by about 2.2 kcal/mol. This is in contrast to the previously reported CHARMM conformational energies, where the value for trans was lower than for gauche by 3.3 kcal/mol.[29,30] B3LYP/6-31G* theory exaggerates hydrogen bonding energy. As the hydrogen bonding strength decreases (Figure 2b and 2c) the relative energy of the trans form increases, and it is about 4.8 kcal/mol above the global minimum on the HF/6-311++G** surface (Figure 2c.) Therefore, at least with these three levels of theory and the TGRTGC and TGCTGR starting geometries, the proposed trans conformation is not the global minimum. Instead, the region (80°, 80°, a gauche conformation) of the experimental structures has the lowest energy.

On Figure 2a, the additional minima at about 140°, 90° and 90°, 140° result from interresidue hydrogen bonds from O-2 to O-5′ or O-2′ to O-5. As the strength of the hydrogen bonds decreases (Figure 2b and 2c), these minima also take on higher relative energies and become less important relative to structures such as the gauche forms that have no hydrogen bonding.

The maps for the GTXGTX exo-cyclic arrangement (Figure 3) give much higher relative energies for the trans structure, about 18 kcal/mol above the global minimum. The optimized trans GTXGTX structure has an interaction that looks like a cyclic hydrogen bond between O-2 and O-2′, but the H···O distances are long (2.49 Å). This particular, artifactual, interaction may have prevented relief of stress during energy minimization. The very large glycosidic bond angle, 128.2°, is consistent with severe stress. Other models at the 180°, 180° point had smaller, although still distorted, glycosidic angles of about 123°, and the THP−O−THP analogue had a glycosidic angle for the same conformation of 119.1°.

Unlike the TGRTGC/TGCTGR surfaces in Figure 2, there are two regions on Figure 3 that have minima of less than 1 kcal/mol for all three levels of calculation. One is the gauche region with ϕ, ϕ' values centered, in this case, at about 70°, 70°. The other is a previously undiscussed minimum, centered at about 115°, 115°. That new region is preferred on Figures 3a and 3b, which are based on the levels of theory that give higher-strength hydrogen bonding. At the 120°, 120° point (Figure 1c), the O-2−H···O-5 distances are 2.64 Å, and the O−H···O angles are 140°. Both O-6 atoms are also oriented to be acceptors for O-2−H, although the O-2−H···O-6 distances are even longer, 2.96 Å. Although such H···O distances are longer than those usually considered to correspond to hydrogen bonds, these four interactions apparently stabilize the 120°, 120° structures in the QM calculations. This extended reach of the hydrogen bonding forces in that conformation was surprising, but corroborates a conclusion from analysis of cellobiose crystal structures.[57] In that analysis, the O-3···O-5′ hydrogen bonds were always accompanied by the gt orientation of O-6′. That conformation allows an additional, bifurcated O-3···O-6′ liaison for many cellobiose structures, even though the distances involved are often quite long.

It is instructive to compare the TGRTGC/TGCTGR and GTXGTX surfaces in the 115°, 115° region. The 1 kcal/mol contour surrounding the 115°, 115° conformation in Figure 3a is very extended. On the other hand, minima at 140°, 90° and 90°, 140° are the only fragments of that region on Figure 2a. The 120°, 120° points in Figure 2 have relative energies of about 4 kcal/mol, regardless of the QM method. This is similar to the values of about 4.5 kcal at the 120°, 120° point on the surfaces

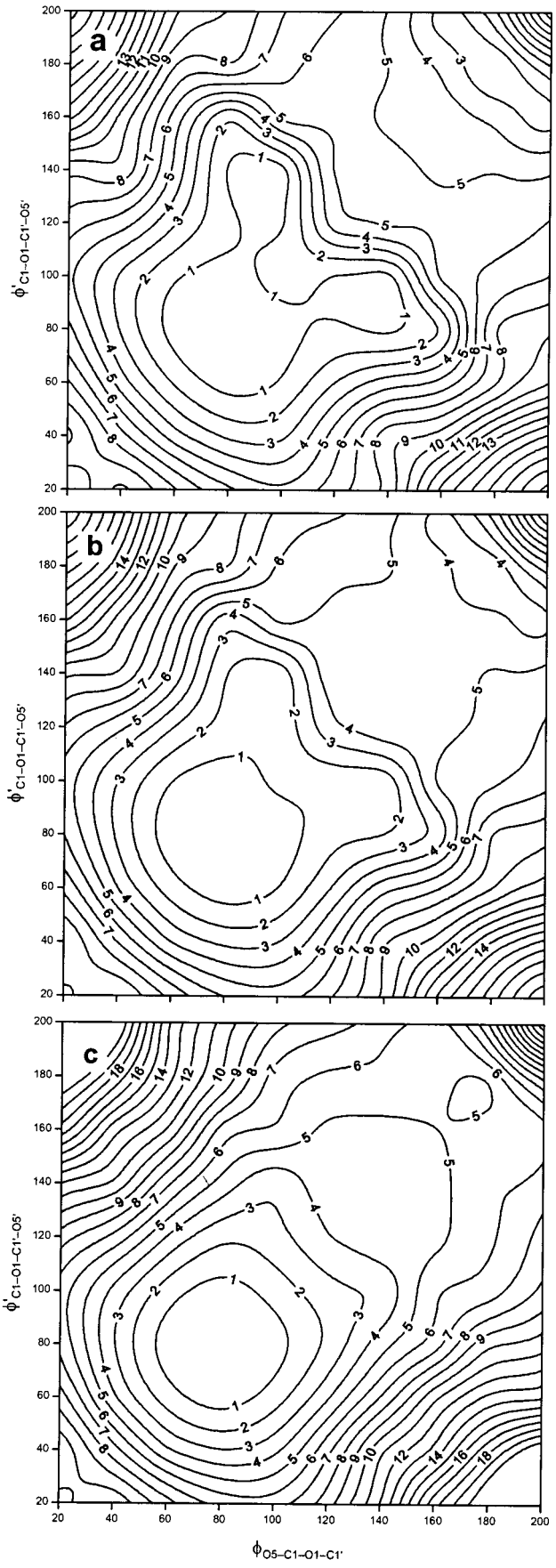

Figure 2. Energy surfaces for the TGCTGR/TGRTGC structures with contours at intervals of 1 kcal/mol. Calculated at (a) B3LYP/6-31G* theory, (b) B3LYP/6-311++G**//B3LYP/6-31G* theory, and (c) HF/6-311++G**//B3LYP/6-31G* theory.

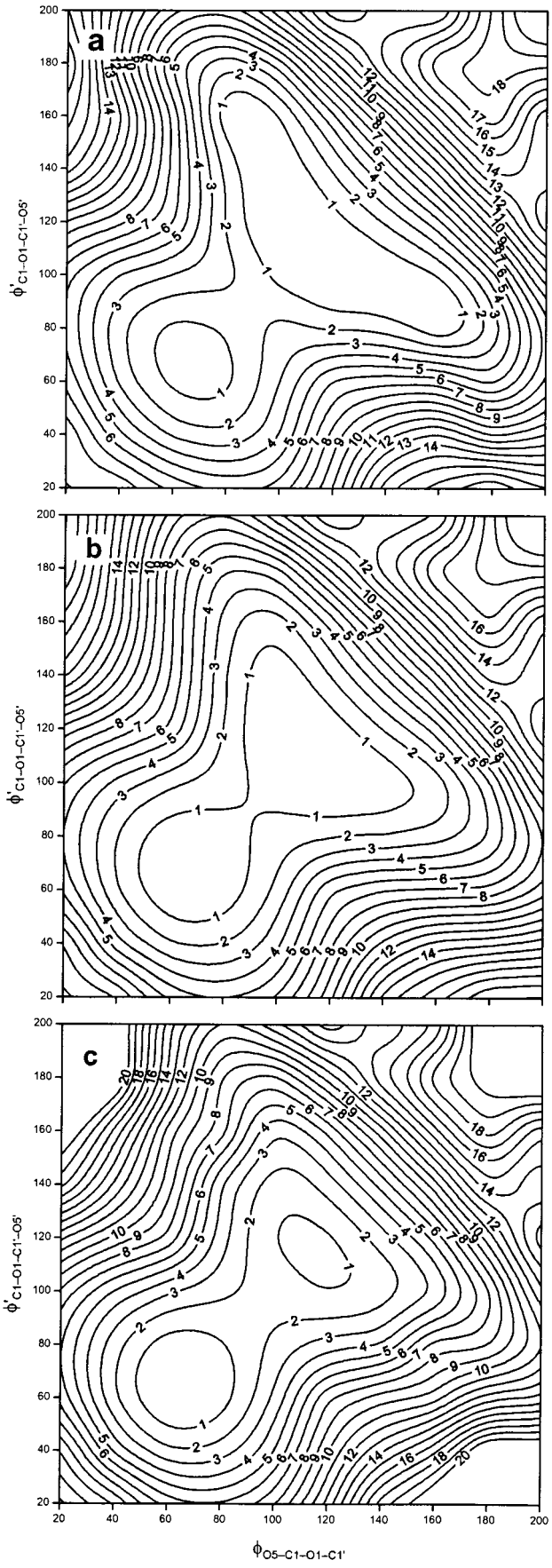

Figure 3. Energy surfaces from using the GTXGTX starting structure with contours at intervals of 1 kcal/mol. Calculated at a) B3LYP/6-31G* theory, b) B3LYP/6-311++G**//B3LYP/6-31G* theory, and c) HF/6-311++G**//B3LYP/6-31G* theory.

for the THP−O−THP analogues, which have no hydrogen bonding abilities. The "R" orientations of the O-2−H groups in the TGRTGC/TGCTGR models do not permit formation of a hydrogen bond to O-5 atoms in the other residue. If O-2−H rotated during minimization to be trans to the methine hydrogen on C-2, converting to the "X" arrangement, then it could make a hydrogen bond as found in the GTXGTX structures at 120°, 120°. Energy minimization does not cause that rotation to spontaneously occur at the 120°, 120° point, but it does occur at 140°, 100°, giving the ear-like shapes in Figure 2a. The formation of the second O-2′···O-5 bond, as found in the 120°, 120° GTXGTX models, would require the unlikely disruption of the clockwise hydroxyl arrangement on the other residue. Neither of the O-2···O-6′ interactions is possible when O-6 is in the tg position.

Some stress within the 120°, 120° conformation is indicated by the moderately large glycosidic bond angles of 119.2° for the GTXGTX disaccharide, 119.4° for the TGRTGC model, and 118.6° for the THP−O−THP analogue. In contrast, the glycosidic angle is only 115.1° for both the GTXGTX disaccharide (Figure 1b) and the analogue in the 60°, 60° conformation. Not all of the variation in glycosidic angle is necessarily due to simple steric stress. Some conformation-dependent variation in the glycosidic bond angle may result from stereoelectronic effects. However, a large glycosidic angle for the trans structures was noted earlier based on the empirical force field calculations.[29,30]

Near the extremes of the new region on Figure 3a, e.g., with $\phi, \phi' = 160°, 80°$, an improved O-2···O-5′ hydrogen bond is formed. This occurs at the expense of its counterpart O-2′···O-5 liaison, as shown in Figure 1d. In these asymmetric conformations, the O-6 or O-6′ atoms are also secondary acceptors with shorter O-2···O-6′ distances than in the 120°, 120° conformation. The $\phi, \phi' = 160°, 80°$ conformations from Figure 3a (similar to the conformations of the "ears" in Figure 2a) benefit from additional interactions provided by the gt O-6 orientation. A third interresidue hydrogen bond, O-3-H···O-6′, is also allowed with good geometry. We note that the grid points with the lowest calculated B3LYP/6-31G* energy (by 0.05 kcal/mol) were 160°, 80° and 80°, 160° (Figure 1d). The lower energies for the 70°, 70° and 115°, 115° minima result from interpolation during gridding of the data.[58] Areas encompassed by the 1 kcal/mol contours that surround the minima at about 115°, 115° decrease dramatically as the given level of theory makes hydrogen bonding less important (Figure 3).

Composite maps based on the lowest energy regardless of starting structure are shown in Figure 4, which also shows the ϕ and ϕ' values for the crystal structures in Table 1. Additional contours of 0.5 and 1.5 kcal/mol have been added. The additional starting geometries have affected the surfaces, with asymmetric structures being more important for the levels of theory with stronger hydrogen bonds. For example, in Figure 4a, the global minimum is for a 60°, 100° structure with GTRGTX geometry that makes a good O-2···O-6′ hydrogen bond. The trans structure is disfavored by 6.2 kcal/mol. The new minimum found on Figure 3 for the GTXGTX structures has corresponding graphical minima on Figure 4a with relative energies of 0.3 kcal/mol at about 100°, 120° and 120°, 100°. The crystal structures are not well-predicted by this surface because they all have corresponding energies greater than 0.5 kcal/mol.

The B3LYP/6-311++G** map in Figure 4b has a very broad region enclosed by the 1-kcal/mol contour. The lowest-energy grid point was furnished by the GTRGTX 60°, 100° structure,

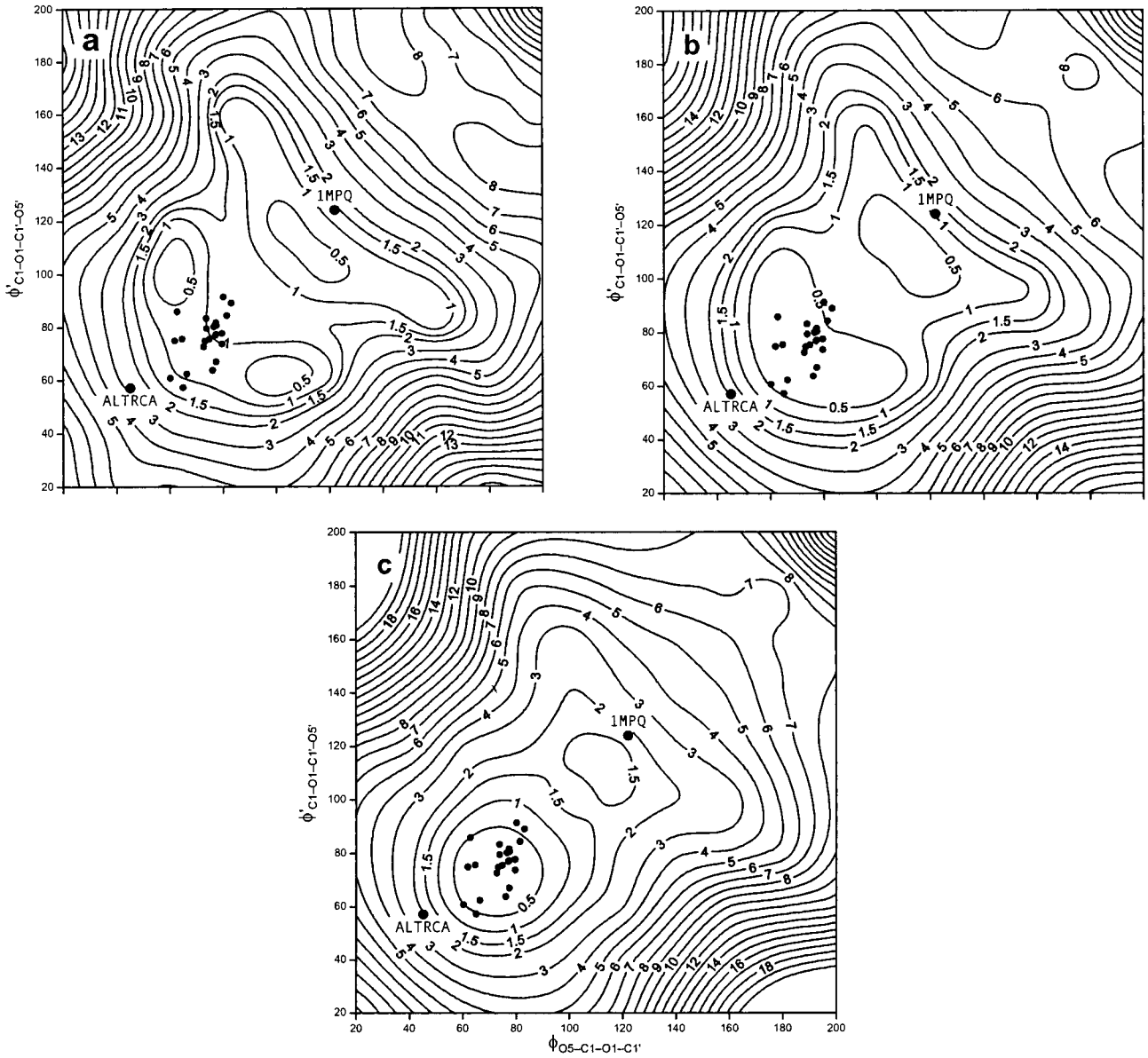

Figure 4. Energy surfaces at the various levels of theory, based on the lowest energy at each ϕ, ϕ' point from eight sets of starting geometries. Besides the contours at 1 kcal/mol intervals, contours at 0.5 and 1.5 kcal/mol are shown. The crystalline conformations from Table 1 are indicated by ●. The arbitrary choice of which torsion angle was ϕ and which would be ϕ' was made to keep the drawing as uncluttered as possible. The ALTRCA[7] and 1MPQ[25] structures are indicated. The identities of the other points are shown in Figure 5. Calculated at (a) B3LYP/6-31G* theory, (b) B3LYP/6-311++G**//B3LYP/6-31G* theory, and (c) HF/6-311++G**//B3LYP/6-31G* theory.

as in Figure 4a. However, there is a slightly lower graphical minimum close to the 60°, 80° grid point that gave an energy of 0.10 kcal/mol. It started with a TGRTGC model that was minimized with AMBER* and then B3LYP/6-31G*. During these minimizations, the rest of the structure was stable but the O-2 hydroxyl group rotated to give an H−C-2−O-2−H torsion of −166°. That rotation to the "X" conformation allows hydrogen-bonding donation to O-1 and lowers the energy, even though no interresidue hydrogen bonds are formed. The TGRTGC starting structure without the initial AMBER* minimization, with a final H−C-2−O-2−H torsion of −46°, had a much higher relative energy, 3.2 kcal/mol. In that structure, O-2−H is not donated to any oxygen. The trans structure is 5.7 kcal/mol above the global minimum on Figure 4b. The preferred structures at 180°, 180° were found by both the MM3 ϵ =1.5 and AMBER* searches. They had TGCGTR conformations, slightly different from the TGRTGC structure proposed in the studies of Liu et al. and Engelsen and Pérez. The crystal structures have lower corresponding energies on this surface than on Figure 4a, but there is substantial area inside the 0.5 kcal/mol contours that is unoccupied by observed geometries.

The gauche region dominates on the composite map based on HF/6-311++G**//B3LYP/6-31G* theory (Figure 4c.) The trans region is higher than the global minimum by 7 kcal/mol, and it no longer corresponds to a local minimum but only to a plateau. The lowest energy grid point, 80°, 80°, was found by the MM3 ϵ =1.5 search, with TGCTGC internal geometry and no interresidue hydrogen bonds. The 60°, 80° and 80°, 60° grid points described in the preceding paragraph were only 0.05 kcal/mol higher in energy on this surface. At the 60°, 60° location, the only other grid point inside the 1 kcal/mol contour, the GTXGTX structure had the lowest energy.

If the intrinsic conformation is best indicated by the surface at the B3LYP/6-311++G** level (the one that calculates the experimental enthalpy for the water dimer) then the isolated trehalose molecule would have substantial flexibility. Hydrogen-bonded structures near 115°, 115° would have almost the same

energy as the gauche structures stabilized by the exo-anomeric effect, and the barrier between them is estimated to be less than 0.75 kcal/mol. According to Figure 4b, the range of motion would easily include both minima. Hydrogen-bonded structures near 60°, 100° would also be visited often. In contrast to the large area surrounded by the 1 kcal/mol contour, the distribution of the experimental crystal structures is quite compact. Therefore, the impact of placing the molecule in a condensed phase would be a substantial contraction of the range of conformations that were readily accessible to the isolated molecule.

The 115°, 115° and 60°, 100° structures that broaden the 1-kcal/mol region in Figure 4b rely on interresidue hydrogen bonding. Such structures might have lower entropy and higher free energy than found for the gauche forms. Therefore, the structures favored intrinsically on a free energy basis might instead be represented better by a surface with reduced hydrogen bonding strength, such as Figure 4c. That surface has a small area inside the 1-kcal/mol contour. Even so, all but two of the crystal-structure conformations are found within the 1-kcal/mol contour and the majority are at 0.5 kcal/mol or less. Further, there are no *unoccupied* regions with energies of less than 1 kcal/mol. Thus, prediction of condensed phase structures is substantially improved by calculations that reduce the strength of hydrogen bonds. This is in line with our practice of using an elevated dielectric constant to reduce the strength of hydrogen bonding in MM3 and hybrid calculations, although we have been using even smaller values of the hydrogen bonding strength in those calculations.[59]

The two conformations that fall outside the 1 kcal/mol contour in Figure 4c are from the allo-trehalose calcium salt complex (ALTRCA)[7] and trehalose in the maltoporin complex (1MPQ), both near the 2 kcal/mol contour. The distortion of ALTRCA can be attributed to the interaction of five of its oxygen atoms with a Ca^{2+} ion. Not only are two hydroxyl oxygen atoms on each monosaccharide residue involved with the same ion, which could put a twist into the geometry, but so is the glycosidic oxygen atom. That could affect the ϕ and ϕ' torsional potentials.[60] The 1MPQ structure,[25] from a 3.0 Å-resolution study with $R = 0.218$, may not be precisely defined. In any case this is not a particularly high-energy distortion,[61] and it is near the previously unreported minimum.

The lowest-energy region that contains all other structures is magnified in Figure 5. Although ALTRCA (Figure 4) is somewhat of an outlier, another allo-trehalose structure (YOXFOG),[21] the trihydrate with two disaccharides per asymmetric unit, has ϕ and ϕ' values that are very close to both the calculated global minimum and to the center of the range of experimental conformations. Despite their hydration, YOXFOG molecules have almost equal values of ϕ and ϕ'. We consider the allo-trehalose molecule to be a very good analogue of trehalose. It differs from trehalose only because of its axial hydroxyl groups at the 3 and 3′ positions. An even better analogue, galacto-trehalose (YOXFUM)[22] differs from trehalose by having axial hydroxyls at its C-4 positions. Its conformation is close to that of the anhydrous trehalose structure DEKYEX,[10] on Figure 5. Although allo-trehalose was distorted in the Ca^{2+} complex, a Ca^{2+} complex of trehalose itself (TRECAB)[17] is reasonably close to the center of the distribution of observed structures.

More trehalose-type crystal structures have two molecules per asymmetric unit than have been found for other disaccharides. The degree of difference in the conformations of identical molecules in identical environments is instructive. The largest difference is found for the heavily substituted DUHXEJ[11]

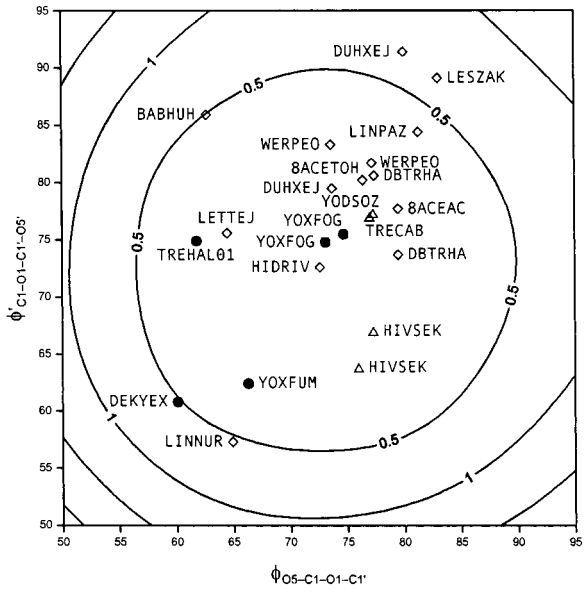

Figure 5. Lowest energy region of the HF/6-311++G**//B3LYP/6-31G* energy surface with all of the crystal structure conformations except ALTRCA and 1MPQ which are shown in Figure 4. Unsubstituted molecules that are excellent analogues of α,α-trehalose (allo-trehalose, galacto-trehalose, and trehalose itself) are indicated by ●, those with minor differences such as complexation with salt or substitution at the 6-position are indicated by △, and those with substitution or changes in configuration at C-2 are indicated by ◊.

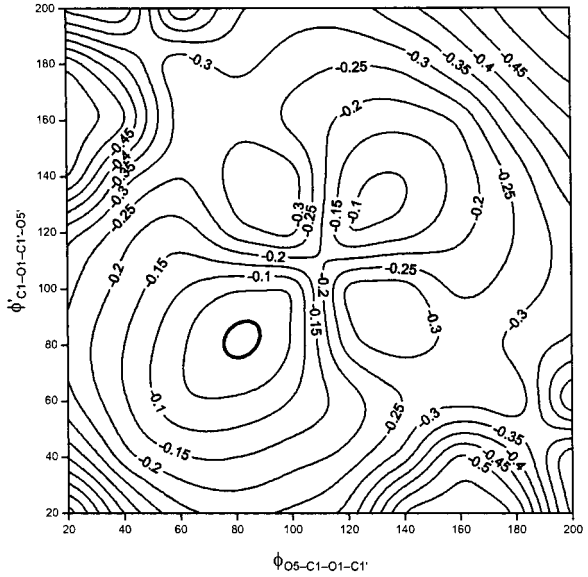

Figure 6. Difference in energy between the HF/MIDI!/SM5.42R solvent model relative energies and the HF/MIDI! relative energies for the TGRTGC/TGCTGR structures. Contours are drawn at increments of 0.05 kcal/mol, and show how much lower the relative energy is because of solvation. Surfaces based on the GTXGTX structure vary by a similar amount, but the pattern of variation is quite dissimilar.

molecules, about 6° in one torsion angle and 11° in the other. The total separation in ϕ, ϕ' space for trehalose and its dihydrate, DEKYEX and TREHAL01,[18] is similar, about 15°.

Despite the finding that a solvent effect is not needed to explain the observed gauche conformations for trehalose, it is of interest to see the effect of a continuum solvation model on the conformational energies of trehalose. Figure 6 shows the difference between the relative solvated energies and the relative energies for isolated TGRTGC and TGCTGR trehalose molecules. The largest discrepancy is about 0.55 kcal/mol, a small effect and continuum solvation stabilizes the trans structure

relative to the gauche form by about 0.45 kcal/mol. Relative solvation energies were also small for the THP−O−THP analogues calculated earlier.[5]

Conclusions

Trehalose was especially suitable for the initial conformational mapping studies of a disaccharide by QM for several reasons. It has broad importance, and a better understanding of its shape should be useful in understanding its various roles. From a practical perspective, the computational burden is lower. Lastly, previously reported relative empirical energy values had large differences. Those differences led to a controversy as to whether aqueous solution was causing the molecule to take a significantly different shape than it would if only internal forces were determining the conformation. Given the low volatility of trehalose, and consequently a great difficulty in obtaining isolated molecules experimentally, QM would be the method of choice to determine the relative energies of the isolated disaccharide conformations. Our QM results argue strongly that the proposed solvent-mediated trans-to-gauche transformation does not occur. Depending on the level of theory, the experimentally observed gauche conformations (and the 115°, 115° shapes) are favored by 5 to 7 kcal/mol over trans structures.

Acknowledgment. Gabor Csonka, Ian Carmichael and Todd Lowary offered valuable comments on a preliminary version of the paper. Magnus Färnbäck furnished the coordinates of 8ACETOH prior to publication.

References and Notes

(1) Elbein, A. D. *Adv. Carbohydrate Chem. Biochem.* **1974**, *30*, 227−256.
(2) Asahina, E.; Tanno, K. *Nature* **1964**, *204*, 1222.
(3) Franks, F. In. *Biophysics and Biochemistry at Low Temperatures*, Cambridge University Press: Cambridge, 1985, ch. 6, pp 112−124.
(4) Lee, C.-K.; Koh, L. L. I. *Carbohydr. Res.* **1994**, *254*, 281−287.
(5) Tvaroška, I.; Václavík, L. *Carbohydr. Res.* **1987**, *160*, 137−149.
(6) Cramer, C. J.; Kelterer, A.-M.; French, A. D. *J. Comput. Chem.* **2001**, *22*, 1194−1204.
(7) Ollis, J. V.; James, J.; Angyal, S. J.; Pojer, P. M. *Carbohydr. Res.* **1978**, *60*, 219−228.
(8) Luger, P.; Paulsen, H. *Acta Crystallogr., Sect. B* **1981**, *37*, 1693−1698
(9) Williams, G.; Lavallee, P.; Hanessian, S.; Brisse, F. *Acta Crystallogr., Sect. B* **1979**, *35*, 2574−2579.
(10) Jeffrey, G. A.; Nanni., R. *Carbohydr. Res.* **1985**, *137*, 21−30.
(11) Brown, J. M.; Cook, S. J.; Jones, R. H.; Khan, R. *Tetrahedron* **1986**, *42*, 5089−5096.
(12) Lee, C.-K.; Linden A. *J. Carbohydr. Chem.* **1995**, *14*, 9−16.
(13) Saviano, M.; Iacovino, R.; Benedetti, E.; Cucinotta, V.; Grasso, G.; Sciotto, D. *Z. Kristallogr.-New Crystal Structures 214* (1999) 297−299.
(14) Linden, A.; Lee, C.-K. *Acta Crystallogr., Sect. C (Cr. Str. Comm.)* **1994**, *50*, 1108−1112.
(15) Linden, A.; Lee, C.-K. *Acta Crystallogr., Sect. C (Cr. Str. Comm.)* **1995**, *51*, 747−751.
(16) Linden, A.; Lee, C.-K. *Acta Crystallogr., Sect. C (Cr. Str. Comm.)* **1995**, *51*, 751−754.
(17) Cook, W. J.; Bugg, C. E. *Carbohydr. Res.* **1973**, *31*, 265−275.
(18) Brown, G. M.; Rohrer, D. C.; Berking, B.; Beevers, C. A.; Gould, R. O.; Simpson, R. *Acta Crystallogr., Sect. B* **1972**, *28*, 3145−3158.
(19) Lee, C.-K.; Koh, L. L.; Xu, Y.; Linden, A. *Acta Crystallogr., Sect. C (Cr. Str. Comm.)* **1994**, *50*, 915−919.
(20) Lee, C.-K.; Linden, A. *Carbohydr. Res.* **1994**, *264*, 319−325.
(21) Linden, A.; Lee, C.-K. *Acta Crystallogr., Sect. C (Cr. Str. Comm.)* **1995**, *51*, 1007−1012.
(22) Linden, A.; Lee, C.-K. *Acta Crystallogr., Sect. C (Cr. Str. Comm.)* **1995**, *51*, 1012−1016.
(23) Baddeley, T. C.; Clow, S. M.; Cox, P. J.; McLaughlin, A. M.; Wardell, J. L. *Acta Crystallogr.* **2001**, E57, 456−457.
(24) Färnbäck, M.; Erikson, L.; Widmalm, G. *Acta Crystallogr. C*, in press.
(25) Wang, Y. F.; Dutzler, R.; Rizkallah, P. J.; Rosenbusch, J. P.; Schirmer, T. *J. Mol. Biol.* **1997**, *272*, 56−63.
(26) Batta, G. J.; Kövér, K. E.; Gervay, J.; Hornyäk, M.; Roberts, G. M. *J. Am. Chem. Soc.* **1997**, *119*, 1336−1345.
(27) Duda, C. A.; Stevens, E. S. *J. Am. Chem. Soc.* **1990**, *112*, 7406−7407.
(28) Donnamaria, M. C.; Howard, E. I.; Grigera, J. R. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 2731−2735.
(29) Liu, Q.; Schmidt, R. K.; Teo, B.; Karplus, P. A.; Brady, J. W. *J. Am. Chem. Soc.* **1997**, *119*, 7851−7862.
(30) Engelsen, S. B.; Pérez, S. *J. Phys. Chem. B* **2000**, *104*, 9301−9311.
(31) Dowd, M. K.; Reilly, P. J.; French, A. D. *J. Comput. Chem.* **1992**, *13*, 102−114.
(32) Allinger, N. L.; Rahman, M.; Lii, J.-H. *J. Am. Chem. Soc.* **1990**, *112*, 8293−3807.
(33) French, A. D.; Kelterer, A.-M.; Johnson, G. P.; Dowd, M. K.; Cramer, C. J. *J. Mol. Graphics Mol. Modelling* **2000**, *18*, 95−107. The trehalose surfaces were not published in this work, but the average energies of the crystalline conformations were tabulated.
(34) Marchessault, R. H.; Perez, S. *Biopolymers* **1979**, *18*, 2369−2374.
(35) Ha, S. N.; Giammons, A.; Field, M.; Brady, J. W. *Carbohydr. Res.* **1988**, *180*, 207−221.
(36) van Gunsteren, W. F. *GROMOS, Groningen Molecular Simulation Package*; University of Groningen, The Netherlands, 1987.
(37) French, A. D.; Kelterer, A.-M.; Johnson, G. P.; Dowd, M. K.; Cramer, C. J. *J. Comput. Chem.* **2001**, *22*, 65−78.
(38) *Adv. Carbohydr. Chem. Biochem.* **1997**, *52*, 43−177; *Carbohydr. Res.*, **1997**, *297*, 1−90; *J. Carbohydr. Chem.* **1997**, *16*, 1191−1280; *Pure Appl. Chem.* **1996**, *68*, 1919−2008.
(39) Barrows, S. E.; Dulles, F. J.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *Carbohydr. Res.* **1995**, *276*, 219−251.
(40) Lii, J.-H.; Ma, B.; Allinger, N. L. *J. Comput. Chem.* **1999**, *20*, 1593−1603.
(41) The importance of diffuse functions, represented by the + signs in the 6-311++G** basis set designation, to the correct calculation of the energy when correlation is included was stated in, Csonka, G. I.; Elias, K.; and Csizmadia, I. G., *Chem. Phys. Letts.* **1996**, *257* 49−60. Our work in ref 39 also reported over-estimation of hydrogen bonding strength with the 6-31G* basis set when the MP2 level of theory was used. In his comments on a draft of our manuscript, Prof. Csonka suggested that less CPU time would be required with the 6-31+G* basis set and the accuracy would be similar. In the present case, the bulk of the computer time was spent on the energy minimizations with the B3LYP/6-31G* level of theory. The succeeding single point calculations with the 6-311++G** basis set added little to the total time required.
(42) Momany, F. A.; Willett, J. L. *J. Comput. Chem.* **2000**, *21*, 1204−1219.
(43) .Jaguar 4.0, Schrödinger, Inc., Portland, Oregon, U.S.A.
(44) SURFER is available from Golden Software, Golden Colorado, U.S.A.
(45) Ha, S. N.; Madsen, L. J.; Brady, J. W. *Biopolymers* **1988**, *27*, 1927−1952.
(46) French, A. D.; Tran, V. H.; Pérez, S. *ACS Symp. Ser.* **1990**, *430*, 191−212.
(47) Barrows, S. E.; Storer, J. W.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *J. Comput. Chem.* **1998**, *19*, 1111−1129.
(48) The calculations that were redundant because of symmetry were used to test the robustness of the minimizer for these complex systems.
(49) Senderowitz, H.; Still, W. C. *J. Org. Chem.* **1997**, *62*, 1427−1438.
(50) MacroModel is available from Schrodinger, Inc., Portland, Oregon, U.S.A.
(51) Sanford, M.; Johnson, G. P.; French, A. D.; Laine, R. A. *J. Phys. Chem. A.* **2002**, *106*, 4115−4124.
(52) McCarren, P. R.; Gordon, M. T.; Lowary, T. L.; Hadad, C. M. *J. Phys. Chem. A* **2001**, *105*, 5911−5922.
(53) Csonka, G. I., Sosa, C. P. *J. Phys. Chem. A* **2000**, *104*, 7113−7122.
(54) Chem-X is no longer available.
(55) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 63.
(56) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281
(57) Peralta-Inga, Z.; Johnson, G. P.; Kelterer, A.-M.; Dowd, M. K.; Stevens, E. D.; French, A. D. *Carbohydr. Res.*, in press.
(58) Gridding is a preliminary step in the preparation of contour plots with the SURFER software. A final grid of energy data is prepared by mathematical fitting. Although it would be better if less interpolation were used, our experiences with various grid sizes and isolation of minima shows that the procedure works reasonably well.
(59) The pragmatism of reducing hydrogen bond strength so the surface can best rationalize the crystal structures should be taken in context. Although the calculated relative potential energies of an isolated molecule are being used to predict the molecular shapes in crystals, the crystalline shapes must be actually governed by the total *free* energy of the system

and kinetic effects. If the modeling method explicitly provides an opportunity to determine the free energy, then the reduced hydrogen bonding strength may not be appropriate. Furthermore, if neighboring molecules, such as solvent, are explicitly part of the model, then intramolecular hydrogen bonds may not predominate. There again, reduced hydrogen bonding strength may not be appropriate. Other explanations for the utility of reduced hydrogen bonding strength have been given in ref 33.

(60) Tvaroška, I.; Carver J. *J. Phys. Chem*. **1995**, *99*, 6234−6241.

(61) French, A. D.; Johnson, G. P.; Kelterer, A.-M.; Dowd, M. K.; Cramer, C. J. *Int. J. Quantum Chem*. 2001, *84*, 416−425.