ELSEVIER

Constructing and evaluating energy surfaces of crystalline disaccharides

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

**Southern Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, New Orleans, Louisiana, USA*
†Institut für Physikalische und Theoretische Chemie, Technische Universität Graz, Graz, Austria
‡Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota, USA

This paper focuses on the methods used to construct Ramachandran plots for disaccharides. Our recent work based on a hybrid of molecular mechanics and quantum mechanics energies pointed to the need to take extra care when making these maps. Care is also important in the quantitative validation of these energy surfaces with linkage conformations that were determined by crystallography. To successfully predict conformations that have been observed experimentally, the calculation of the energy should include stereoelectronic effects and correctly weight the hydrogen bonding. Technical concerns include the method used to scan the range of conformations, starting geometries, and finding the zero of relative potential energy on a surface where the values were collected at regular intervals. The distributions of observed conformations on energy maps of sucrose, maltose, and laminarabiose at dielectric constants of 1.5 and 7.5 illustrate the effects of an elevated dielectric constant for the MM3 component of the hybrid energy calculations. At dielectric constants of 3.5 and 7.5, the overall average energies of observed conformations of sucrose and seven disaccharides of glucose were less than 1.0 kcal mol^{-1}. The distribution of corresponding energies of the various crystalline conformations conformed well to a Boltzmann-like equation. © 2000 by Elsevier Science Inc.

Keywords: carbohydrate, conformation, laminarabiose, maltose, molecular mechanics, quantum mechanics, sucrose, tetrahydropyran, trehalose

Corresponding author: Alfred D. French, Southern Regional Research Center, P.O. Box 19687, New Orleans, LA 70179-0687, USA. Tel.: 504-286-4410; fax: 504-286-4217

E-mail address: afrench@nola.srrc.usda.gov (A.D. French)

INTRODUCTION

Conformations of disaccharides are studied because their shapes are key to understanding their properties. Also, oligosaccharides and polysaccharides have the same linkages. The same factors that influence the shapes of disaccharides should apply to these larger but experimentally and computationally less tractable molecules. The conformations usually are defined by the amounts of rotation of the monosaccharide residues about their interconnecting bonds. These torsion angles, labeled ϕ and ψ, are shown in Figure 1 for three disaccharides in this report.

Ramachandran[1] plots are intended to show where in ϕ and ψ space the molecules are most frequently found or to predict where they will be found. These plots, often used to explain the conformations of various disaccharides,[2] can be based just on the molecular shapes observed by crystallography,[3] or they can show how calculated energies vary with changes in both ϕ and ψ. The energy maps are predictive because the structures should occur in the regions of lowest energy. The energy surfaces have evolved from the early allowed-disallowed calculations to contours based on minimized molecular mechanics (MM) energies calculated by programs such as MM3.[4] Further evolution of these energy calculations is based on potentials of mean force calculated by molecular dynamics simulations, often carried out in water.[5–7] These maps can be used to validate the method used to calculate the energies by comparison of experimentally determined conformations with the low-energy regions. Validated maps can be used to judge the distortion induced by various environments. The maps also show the heights of barriers and the likelihood of alternative conformations. They can augment information from experiments that are not definitive to give more plausible answers.

In the present work, the focus is on methods used to produce energy surfaces that quantitatively depict the ranges of conformations that are observed by crystallography. These predictions are ambitious (and controversial) because neighboring molecules are not explicitly considered. The same molecule

Journal of Molecular Graphics and Modelling 18, 95–107, 2000

655 Avenue of the Americas, New York, NY 10010

1093-3263/00/$–see front matter
PII S1093-3263(00)00041-3

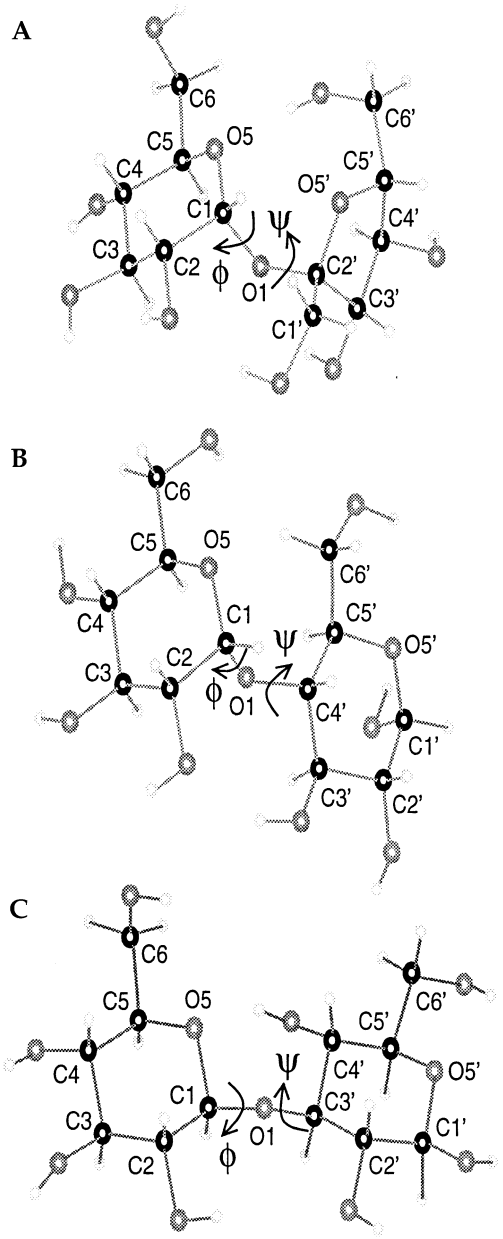

Figure 1. The three disaccharide molecules for which energy surfaces are presented. (a) Sucrose. (b) Maltose. (c) Laminarabiose. ϕ and ψ are indicated with arrows.

will have different conformations in different environments, so it is clear that the neighbors affect conformations to some degree. However, if the net forces from neighbors are small relative to the intrinsic forces and the different crystalline environments constitute random perturbations of the ideal geometry, then conformations should be predictable in a statistical sense. Further, we imagine that the same surfaces that can predict crystalline conformations will represent fairly well the shapes accessible in solution.

Several factors influence computed energy surfaces. The most important is the method used to calculate the energy. Although there is an overall similarity in results, different force fields find minima at different locations and produce varying calculated relative energies.[8] Methods such as semiempirical quantum mechanics also could be used for calculating the energy, although we do not recommend that approach. Other, more technical considerations also affect the results, such as the minimization method and the torsion driver method.

Recently, we calculated the energies for sucrose using a hybrid of both MM and ab initio quantum mechanics (QM) calculations.[9] This method adds the calculated MM energies for the full disaccharide to the differences between the QM and MM energies for a smaller but critical part of the molecule. The critical part of the molecule included the two rings and the linkage oxygen atom of the disaccharide, with all exocyclic groups replaced by hydrogen atoms. The result is a simple analog of sucrose, composed of tetrahydropyran (THP) and tetrahydrofuran (THF) rings. We also used an otherwise identical analog of sucrose that had a methyl group attached to the "glycosidic" carbon atom of the THF ring. Sucrose has been particularly difficult to model with MM methods alone,[10] and crystalline shapes of sucrose are accommodated better on our hybrid energy surfaces. We have also successfully applied the hybrid method to cellobiose, laminarabiose, maltose, nigerose, and all three trehaloses, and the QM surfaces are being published elsewhere.[11]

While making these surfaces, we gained new appreciation for their careful construction and analysis. Our hybrid maps are based on the difference between QM and MM energies for the THP-O-THF or THP-O-THP analogs. For that difference to be meaningful, the structures of the analogs at all ϕ, ψ points must be comparable. The purpose of the present paper is to describe the considerations, decisions, and obstacles that we encountered when constructing and analyzing those maps. These ideas and details should apply regardless of the method of calculating energy, and some of the questions and difficulties will present themselves to users of any software system. Some of the ideas presented herein will be incorporated in our future work. Other ideas have been advocated by our colleagues for some time, but until we saw clearly a need for high precision, we had continued with our "good enough" methods. We also discuss the comparisons of the maps with conformations determined by crystallography.

METHODS

Selection of Theoretical Level

MM force fields for modeling carbohydrates have recently been reviewed.[12] They are designed to compute structures and energies based on knowledge of familiar descriptors of molecular structure: bond lengths, bond angles, and nonbonded interactions such as van der Waals and electrostatic forces. These descriptors are given ideal values for any given set of atoms in a structure, as well as an energetic cost for deviations from the ideal values. Knowledge of these ideal values and the cost (in energy) of deviation is reduced to equations that, along with their constants, or parameters, constitute an empirical force field. Another factor is torsional energy. When rotation about a bond occurs, torsional terms are needed in most force fields to augment the contributions to the total energy from van der Waals and angle-bending forces. The torsional terms are thought to arise from interactions of orbitals or from orientation-dependent distributions of electrons. Because of the extensive stereoelectronic effects arising from the lone pairs of electrons on oxygen atoms, suitable equations for torsional

motion within alternating sequences of carbon and oxygen atoms have been especially difficult to design.

For example, consider the C-O bonds in the sequence that involves the anomeric carbon, C1, in molecules that contain glucose moieties (in Figure 1b, the atoms C5-O5-C1-O1-C4′). The unexpected degree of preference for gauche (±60°) conformations for the C5-O5-C1-O1 and C4′-O1-C1-O5 torsions is a manifestation of the general anomeric effect.[13] Energy values computed for different conformations of this sequence directly impact both the predicted likelihood of the axial or equatorial dispositions of the C1-O1 bond (whether the α- or β-anomer is preferred) as well as the preferred location of the adjacent monomeric residue that incorporates C4′. While Allinger was one of the first software developers to recognize the importance of these anomeric effects, MM3 gives energy surfaces that are only roughly similar to those from B3LYP/6-31G* and MP2/6-31G* calculations for dimethoxymethane, CH_3-O-CH_2-O-CH_3.[14] The more recent force field, united-atom AMBER*, mimics the relative QM energies for different conformations of axial and equatorial 2-methoxytetrahydropyran fairly well,[15] as does the PEF95SAC force field of Rasmussen.[16] Some force fields use one set of parameters for the axial (α-gluco) configuration and another for the equatorial (β-gluco) configuration.[17] This is expedient when chair-form rings are retained but not a good overall solution, especially when considering distorted ring shapes, five-member rings, or acyclic molecules. Other linkages are not as well researched, but their torsional properties are likely to be critical to the molecular model.[18]

QM calculations avoid the issues of detailed parameterization, but the time required to carry them out can be a problem. For example, we can calculate an energy surface for one of our analogs by MM3 in about 10 minutes. On the same computer, the comparable calculation at the HF/6-31G* level takes about 1 year. Still, we thought our problem with the conformational energies at adjacent anomeric centers of sucrose was based on failure to correctly account for the conformationally variable electronic structure. Therefore, a QM approach was attractive because calculation of the electronic structure is the raison d'être of QM.

There are several ways to use QM results. First, the disaccharide itself could be studied with QM. Compared to our analogs, the already long computations would lengthen substantially. Disaccharides have nearly twice the number of nonhydrogen atoms as the analogs, and the calculation time increases exponentially by the number of atoms. Worse yet, various combinations of orientations of the exocyclic groups must be considered at each point on the ϕ, ψ surface. Computation of the energy of thousands of conformations to develop an energy surface for a molecule as large as a disaccharide is currently impractical by normal QM methods. A second approach would be to use QM to develop new parameters for MM. The third choice is to use a hybrid method, where QM is used for the critical parts of the structure and MM is used for the rest. An example of an integrated hybrid method is available in Gaussian98,[19] although it has not been shown to be useful for carbohydrates. Our method, which is not integrated, is mostly applicable to generating energy surfaces, where a fully integrated method should have wider applicability.

If QM theory is used, the level of theory should be appropriate. The computed energies of the two glucose chairs ($^4C_1$ and $^1C_4$) vary substantially[20] depending on the level of theory. In that work, energies for two different models of each chair were calculated at a number of different levels of theory. Included were advanced electronic structure theory calculations that should provide energies that are close to those that would account for complete basis sets and full electron correlation. Semiempirical methods did not work well, but HF/6-31G* calculations gave relative energies for various glucose structures that are very similar to those from calculations using advanced electronic structure theory.[20,21] The HF/6-31G* level only works because much of the error from an incomplete basis set is canceled by errors caused by omission of electron correlation. Leaving out correlation, as in HF/6-31G* calculations, can be dangerous because correlation accounts for a huge amount of the atom–atom interaction energy. Consideration of correlation is particularly important for congested molecules, such as tris-*t*-butylmethane.[23] On the other hand, the use of correlation, either through perturbation theory, such as MP2, or density functional theory, such as B3LYP, with moderately small basis sets such as 6-31G*, overestimates energies for hydrogen bonding.[24–27]

One consequence of this overestimation was that the never-observed $^1C_4$ conformation of glucose gave slightly lower energy and is therefore preferred compared to the $^4C_1$ shape when MP2/6-31G* theory is used.[20] This apparently happens because of the much better geometry possible for intramolecular hydrogen bonds for the $^1C_4$ shape and the overly strong hydrogen bonds calculated with MP2/6-31G* theory. Findings on the glucose chairs were similar for B3LYP/6-31G* calculations.[25] For the relative energies from the anomeric effects, errors from the poor accounting for correlation in HF/6-31G* are apparently of small importance and the errors mostly cancel. The preferences for gauche torsions of the C-O-C-O sequence around the anomeric centers were fractions of a kcal mol^{-1} stronger for the B3LYP/6-31G* method than for the HF/6-31G* method.[22] If no hydrogen bonding is possible for the structures at hand, such as for our analogs, then it may be attractive to select a correlated method. To get both reasonably accurate energies and detailed structures where hydrogen bonds are involved, it probably is necessary to use correlated theory and basis sets that include diffuse functions.[25]

Selected Software

We used Chem-X,[28] RASMOL,[29] and MOLDEN[30] to sketch and display structures, although many other programs can be used. MM energy surfaces of the analog and disaccharide molecules were prepared using MM3-96.[31] The H···O hydrogen bonding equilibrium distance (1.82 Å) and force constant (3.30 kcal) from MM3-92 were used. Those parameters reproduce the geometries found in crystal structures better than the constants in more recent versions. The default dielectric constant (ϵ) of 1.5 is used during the parameterization of MM3 with gas phase experiment and QM results so it was used for comparisons of the MM3 and QM results for the analogs. We had earlier chosen a value of 3.5 as reasonable for modeling disaccharide crystal structures with MM3.[32] Here, we tried ϵ = 1.5, 3.5, and 7.5 to see their effects on the hybrid energy maps. Increased dielectric constants mimic solvent effects without having to include explicit solvent.[33] We note here that the use of elevated dielectric constants is controversial. This is mostly because the choice of dielectric constant during parameterization affects the resulting parameters, and deviation from that

chosen value will affect the overall balance of parameters. In MM3, the strengths of hydrogen bonds are inversely proportional to the dielectric constant. MM3 did not reproduce well the advanced theory QM energies of the various exocyclic variants of glucose,[21] but the hybrid method in the present work should diminish the component of those discrepancies related to anomeric effects.

QM energies for the sucrose and trehalose analogs were computed with GAMESS,[34] and Gaussian94[35] was used for the remaining analogs. We used the Restricted Hartree-Fock method with the 6-31G* basis set for the model chemistry (HF/6-31G*).

Disaccharide Starting Models

One could use initial structures obtained by crystallography or by computer sketching. The experimentally determined structures have a few disadvantages. Intermolecular hydrogen bonds often are formed in crystals, giving certain orientations to the exocyclic groups. Figure 1a shows the structure of sucrose, as determined by an X-ray diffraction study.[36] Several of the hydroxyl groups are oriented to form hydrogen bonds with other sucrose molecules in the crystal. Energy minimization of molecules with those orientations will give considerably higher energies than molecules with their exocyclic groups oriented to give the most complete networks of intramolecular hydrogen bonds.

Optimal orientations of the exocyclic groups will be different for different ϕ, ψ values, so the analysis must allow for a variety of these group orientations. Starting conformations for molecules that contain glucose should include varied positions of O6 relative to the ring oxygen and C4, as well as various orientations of the hydroxyl groups. The stable orientations of O6 to the ring oxygen and to C4 can be trans-gauche (tg), gauche-gauche (gg), or gauche-trans (gt).[37,38] The first letter refers to the orientation of O6 relative to O5 and the second to C4. A similar notation often is used for the methyl glycosides and the additional primary alcohol group on molecules such as fructofuranose. The hydroxyl orientations can be clockwise, counterclockwise, or counterclockwise, with the hydroxyl hydrogen of C2 trans to C2-H. The latter counterclockwise orientation is especially important for α-gluco but not β-gluco configurations. When in these orientations, the hydroxyl groups form donor-acceptor–donor-acceptor partial rings of intra residue hydrogen bonds. The hydrogen bonds in these rings are of relatively poor geometry and not very likely to be found experimentally, but they still affect the calculated energy of the models.[21] We view the use of a variety of these low-energy starting structures mostly as a way to give each ϕ, ψ conformation the lowest possible calculated energy rather than as a means of predicting what arrangements will be found by experiment. Although partial rings sometimes are found in crystal structures, recent nuclear magnetic resonance studies of sucrose and trehalose in water show that the hydroxyl groups are rotating freely.[39]

Figure 1b shows hydroxyl groups on the glucose residues of a maltose model in clockwise orientations. In Figure 1c, the hydroxyl groups of the laminarabiose model have counterclockwise orientations. Whether the hydroxyl groups appear clockwise depends on the orientation of the molecule, with the residues in Figures 1b and 1c being in the standard orientation with the anomeric carbon atoms on the right side of their residues and the ring oxygens to the rear. Both O6 groups on this maltose molecule are gauche to O5 and trans to C4 (the gt conformation). On the laminarabiose molecule, the O6 on the left (nonreducing) residue is gauche to both O5 and C4 (the gg conformation), whereas O6 on the reducing ring has the gt conformation. The forms of these three molecules shown here permit intramolecular hydrogen bonds. In sucrose, major bonds are between the hydroxyls on C1′ and C2, and between the hydroxyl on C6′ and O5. Molecules with maltose linkages often have hydrogen bonds between the hydroxyl groups on C3′ and C2. The crystalline molecules with laminarabiose linkages often have a hydrogen bond between the hydroxyl on C4′ and O5.

At present, it is not reasonable to exhaustively test all possible exocyclic group orientations of disaccharides at each point in ϕ, ψ space. Our approach is to rely on the minimizer to attain the optimal orientations of the exocyclic groups on disaccharides when starting from various arrangements that give low energy for the monosaccharides. Determining a variety of low-energy forms for monosaccharides is a tractable problem.[20] For example, if each of the six rotatable groups is given three different staggered orientations, there are $3^6 = 729$ combinations of structures for glucose in the usual chair form. Energy minimization with MM3 of those 729 starting structures led to roughly 100 different stable structures.[21] Of these, a smaller number will have an energy within 3 kcal mol^{-1} of the lowest energy. For use in a disaccharide, the number of structures to consider is reduced further because the position of O1-H is irrelevant for what will become the nonreducing residue and the O4′-H orientation is irrelevant for the reducing residue. Other work on glucose, with six increments of rotation for each exocyclic group and the AMBER potential, led to more than 700 stable structures,[40] so this approach may not always work. Stortz[41] pointed out that 16 starting geometries are not adequate to produce an energy surface with the lowest possible energy at each point when a very high ($\epsilon = 80$) dielectric constant is used. He demonstrated this by finding lower-energy structures with an admittedly tedious, impractical manual method. The usual clockwise and counterclockwise arrangements are not likely to be so favored at a high dielectric constant, and other arrangements will be important.

The number of starting geometries for the different disaccharides in this work varied. For cellobiose, 36 starting conformations were used; for maltose, 24; for sucrose, 48; for laminarabiose, nigerose, and the trehalose molecules, 81 were used. All of these starting structures are considered at each ϕ, ψ point. The energy surface then is constructed based on the lowest energy that was obtained at that point. The various numbers of starting structures reflected our degree of concern regarding the potential importance of the primary alcohol group orientations, as well as for comparison with previously published maps. For example, we did not include the tg conformation of O6 atoms in maltose. However, if an energy surface is to correctly reflect the force field, the number of starting models must be sufficient to provide the lowest possible energy at each ϕ, ψ point, and our future analyses will expand the number of starting models. Those somewhat larger numbers will not be as large as $3^{10} = 59{,}049$, however, the number of combinations of all possible staggered orientations.

Because the rings are not held rigid, their conformations will change during minimization. With our hybrid method, it is important for the energies calculated by MM and QM to be for

structures that are as similar as feasible. Consider the situation when the QM minimization causes one of the tetrahydropyran rings of the analog to lose its chair form to relieve a strain. If the chair form and the high-energy short contact are retained during the MM minimization, the difference surface will have a discontinuity. Rather than examine each of the 324 ϕ, ψ points on our analog surfaces to learn whether the chair form had been lost by one method or the other, we used an easily automated, rapid plan. The sketched structures for the analog were first optimized with MM3 and then used as starting geometries for the QM calculation. The structures from the QM calculations then were reoptimized with MM3. This substantially reduced the discontinuities on the MM-QM difference surfaces.

A similar situation can occur with just one method of energy calculation for points that are adjacent on maps. Points where a nonchair conformation would have a lower energy than the chair form of the starting model may not achieve it because the barrier is too high. At the neighboring point there will be just enough more contact that the barrier to the nonchair will be crossed, and a discontinuity on the calculated energy surface will result. This shows that an inadequate number of starting structures was employed. If the ring structures are easily deformed, then it is necessary to include some structures having alternative ring forms in the starting set. This was done in the case of our sucrose analog structures. The furanoid ring is fairly flexible, but there is a barrier between its two major shapes. In some conformational analyses, this is not a severe problem because areas where the ring changes shape correspond to structures that do not occur often enough to be important. However, other computational models give lower energy barriers to the loss of the chair form of the ring, and molecules other than glucose have different properties.[42]

Scanning ϕ, ψ Space

During the scan of ϕ, ψ space, the linkage torsion angles are caused to take the desired values while all other bond, angle, and torsion values are allowed to optimize. Several decisions are required regarding the method to cover ϕ, ψ space. One is the question of the increment size. We used 20° for both ϕ and ψ, over the range of −180° to +180°. Step size was tested by varying ϕ and ψ for a simple model (the THP-O-THP analog of cellobiose) in increments of 5, 15, 20, and 30°. The contour lines from plots of the 5, 15, and 20° grids were nearly superimposed in the low-energy regions, while there was serious discrepancy between the 5 and the 30° lines. This indicates that the 20° surface is fairly good, but a more thorough analysis of the impact of step size on the minimum would be useful. That analysis would necessarily include the method to interpolate the data for the contouring program (see Map Preparation).

The values of ϕ and ψ could be defined by any of three atomic sequences. For example, ϕ in Figure 1 could be defined by H1-C1-O1-C3′, O5-C1-O1-C3′, or C2-C1-O1-C3′. Although we used the definitions based on the hydrogen atoms for some of these molecules, we recommend in retrospect that the driven torsion angles always be defined in terms of nonhydrogen atoms. This is because of the relative inaccuracy of the hydrogen atom positions in diffraction studies. Another reason for using the heavy atoms to drive the rotation about the bonds is that the motion differs for the three atoms during driven rotation. For example, if ϕ is defined by H1-C1-O1-C4′, the values of O5-C1-O1-C4′ and C2-C1-O1-C4′ will differ by about ±120°. These relationships change by several degrees during rotation in the modeling study.[43] Variation is less when a nonhydrogen atom is driven.

Drivers

Not all modeling software provides the same facilities to systematically vary the torsion angles and compute the energy. Some programs allow the initial and final torsion angles and increment size to be specified and then carry out the scan of conformational space. Other programs require writing a simple external control program (batch file, script, etc.). MM3 now has four different internal "dihedral drivers." Option 4 carries out rigid rotations from the same starting structure to give the particular ϕ, ψ conformation and then starts energy minimization. This driver should produce a surface whose energies at −180° and +180° are the same.[44] Sometimes (rarely) it does not, however. Round-off or other errors can still lead to differences in the minimization pathway when starting structures have high energy. The resulting energies and structures can differ by several kcal mol^{-1}. If such a facility as the option 4 driver is not provided in the chosen software, the same effect can be obtained by preparing separate input files that already have the correct ϕ and ψ values for each ϕ, ψ point. Then, the linkage torsions are just held at the desired values by whatever facility is available.

Like option 4, MM3's option 2 driver also attains the next point on the ϕ, ψ surface by initial rigid rotations about the bonds. However, with option 2, the rotation is from the previous ϕ, ψ point rather than always from the same initial structure. There are some advantages and disadvantages to this. The main advantages are that it usually avoids interpenetrated structures (Figure 2) that cannot be minimized, and it can sometimes find lower-energy structures. With option 4, a rigid rotation from the starting structure may cause bonds from one monomeric residue to penetrate the other residue. Minimizers are

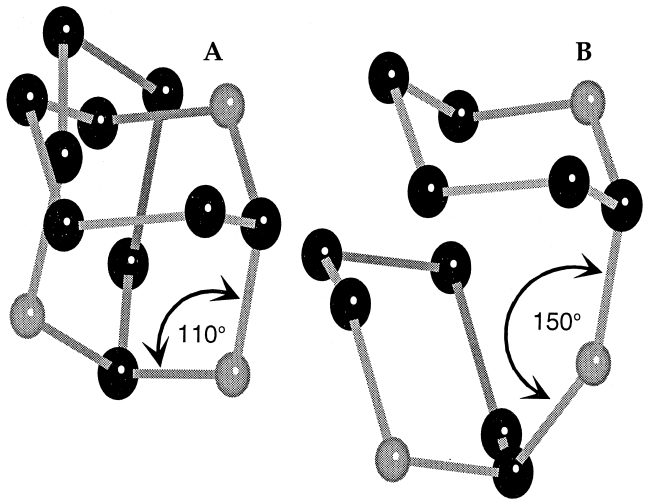

Figure 2. Ring structures in α,α-trehalose with $\phi = -60°$ and $\psi = -40°$, showing just the ring atoms for clarity. (a) Interpenetrated rings, with a glycosidic angle of 110°. (b) Relief of interpenetration, with glycosidic angle increased to 150°. (Drawing produced by Chem-X.[28])

unable to resolve this problem. With option 2, the structure gradually deforms at successive grid points before the high-energy conformation occurs so that interpenetration does not usually occur. The glycosidic bond angle often increases, and the rings will flatten out as well. Exocyclic group orientations or ring shapes may change (e.g., from chair to skew) during such deformations. If the deformations are elastic, this is not a problem, but if the structure is pushed over an energy barrier, the resulting deformation usually will be carried through the rest of ϕ, ψ space. Because of the retained deformation, energy surfaces made with option 2 drivers can give one value for the energy at $-180°$ and another at $+180°$. Somewhere in between, the structure will have undergone a plastic deformation so the structures will be different.

MM3's option 1, the "general" method, is not well suited to making disaccharide energy surfaces. Successive ϕ, ψ points are attained from the initial structure by the minimizer "dragging" the starting structure toward the next desired conformation. When this process entails no large changes in the structure, the method works, and it is suited to studies of ring shape, for example. However, the method takes longer under the best circumstances, and it has the disadvantage of the option 2 method, discussed earlier. Option 3 does not provide energy minimization.

Figure 2 shows an extreme example of an interpenetrated structure that can result from the option 4 driver. It also shows how opening up the glycosidic angle in the starting models to 150° can often avoid interpenetration. Points where interpenetration occurs will have egregiously high, calculated energies that badly distort contouring and interpolation of the energy surface. Normal ranges of relative energies are 30 to 50 kcal mol^{-1}, but interpenetrated structures may have energies much higher than their neighbors have. Besides opening up the glycosidic angle, other strategies can be used. Such points may simply be left out; letting the contouring programs make up for missing points. Also, it is normal for only some of the initial starting geometries for disaccharides to give interpenetrated structures at a given ϕ, ψ point. Others will not, leaving reasonable values to use in map construction. (In those cases, the interpenetration only involves the exocyclic groups, not the ring atoms.)

The strategy of increasing the glycosidic angle was not used for the QM calculations because of the extra time needed. Using MM-optimized structures for input to the QM programs solves this problem, however. Sometimes, a neighboring, successfully optimized structure was adjusted to the desired ϕ and ψ values for the QM calculations. Either approach provides a starting structure close to the final one, useful for saving computer time during QM calculations.

If the option 2 driver gives lower energy values than the option 4 driver, then the set of starting geometries used in the option 4 runs was incomplete. Structures that gave lower energies during additional runs with option 2 then could be added to the set of starting geometries and used with the option 4 driver to learn where else they give lower energy values.

Minimization Method

Many different minimizers are used in MM software. However, MM3 only offers Newton-Raphson schemes with either block-diagonal or full-matrix implementations. The following pertains especially to MM3. The block-diagonal minimizer is rapid and moves the structure towards the closest local minimum. The full-matrix method moves the structure more slowly towards the closest stationary point, be it either a minimum or a maximum. (Another problem is that the full-matrix minimizer in versions of MM3 before MM3-94 sometimes inverts chiral molecules.) Although the full-matrix minimizer has numerous additional capabilities, such as the calculation of vibrational frequencies, the block-diagonal method is preferred for Ramachandran maps. However, the block-diagonal minimizer sometimes can yield different structures with different orderings of the atoms in the input file. It happens because this type of minimization moves one atom at a time. Atom-order–dependent results can arise when using structures from X-ray diffraction studies that have short O-H and C-H bond lengths. This also can happen when the molecule is put into a strained conformation during the scan of the full range of ϕ and ψ. This would not normally be apparent. However, the chemically symmetric α,α- and β,β-trehalose molecules should give the same energies for structures with either values of ϕ = x, ψ = y or ϕ = y, ψ = x. Therefore, a full map for α,α-trehalose, which has a very small, low-energy region, is a good test case of how robust the methods are for getting correct values of the energy. In order to work around this, additional input files of each starting structure were prepared with different atom orders. The lowest energy was used for the maps.

Termination Criteria

MM3 offers two different criteria for terminating the minimization. One, referred to as the geometry optimization method, is based on the atomic movement. The other is based on energy change. Because relatively large changes in hydroxyl hydrogen atom positions can occur with little change in the energy, the movement often does not drop to the small value that is used by the movement criterion. Low-frequency vibration also can occur for rotation about the O-C′ bond. If the movement does not diminish to the small value specified in the program, MM3 switches to the energy-based criterion. Other than a slightly slower speed, we have seen no disadvantage from simply starting with the energy-based criterion (n $\times$ 0.00008 kcal mol^{-1}, where n is the number of atoms.)

The amount of atomic movement during minimization must be controlled to assure a smooth minimization and to prevent disruption of the structure. There is a maximum movement size in the software for this purpose. Versions of MM3 before MM3-96 sometimes will fail to complete energy minimization and give a message that the energy has gone up. The accompanying maximum atomic movement will be large.[45] Now, the movement per step of minimization is restricted to 0.10 Å and the problem is avoided.[46] With Gaussian94, a similar situation arose. After restricting the movement in some auxiliary B3LYP/6-31G* calculations on our analogs with the keyword IOP(1/8=5) along with SCF=Tight, the minimization terminated without repetitious cycling that wasted considerable time.[22]

Symmetry

As mentioned earlier, the molecular symmetry of α,α-trehalose can be used to test the robustness of the minimization. Symmetry of trehalose also is valuable for QM studies of its analog. Only structures along the diagonal of the energy surface and

those to one side of it need to be computed. This saves nearly half of the required computer time.

Map Preparation

Files containing ϕ, ψ and energy from the MM or QM programs were gridded by the program Surfer[47] using its implementation of multiquadric radial basis functions.[48] The gridding function interpolates between the data points to provide missing data and can increase the number of points used in the contouring to make a smoother plot. A more important capability provided by gridding is that it allows finer grids in the regions of the minima and coarser grids elsewhere. This should be very useful with expensive QM data, although it was not employed in the present work. It is very important to obtain the minimum energy for the gridded data, and it was expedient to use grid increments of 1°. The minimum of the resulting grid was subtracted to ensure that the minimum anywhere on the surface is zero. That gave a final grid that was used to produce the contour map.

When interpolating data to the edge of a surface, there will be a difference in curvature from what would be calculated if the surface were periodic. Although we did not use this refinement in the present calculations, using data from a range of $-200°$ to $+200°$ to calculate the interpolated values should reduce edge effects on contoured surfaces. (The energies from the $+160°$ and $-160°$ points can be used for the $-200°$ and $+200°$ values, respectively.)

Our simple hybrid energy maps were constructed from the 1° grids with the aid of the grid math functions of Surfer. First, a difference grid for the analog was constructed by subtracting the QM analog grid from the MM3 analog grid. This difference grid was subtracted from the MM3 disaccharide grid and the resulting minimum grid point energy value determined. That minimum value was subtracted from the grid data of this hybrid grid to produce the final grid file that was used in turn to make the hybrid QM/MM contour map. It is important both that there be a value of zero and that there be no interpolated value less than zero. Otherwise, the corresponding relative energies of the conformations observed in the crystal structures will be erroneous for avoidable technical reasons. This is why we used the fine, 1° grid. Errors as large as 0.5 kcal arose unless we took care of these details, and the error applies to all corresponding energies for the crystal structures. Energy surfaces such as these that show the lowest attained energy at each point sometimes are called adiabatic maps.

Crystal Structure Conformations

Consideration also must be given to the experimental data if they are to be used for quantitative evaluations of energy surfaces. Whereas a given crystal may yield only a single ϕ, ψ pair, there are often numerous different crystals that contain the linkage in question. For example, maltose linkages are found in α-maltose, β-maltose monohydrate, methyl-β-maltoside, methyl-α-maltotrioside, and so forth. In those compounds, all chemical variations are distant from the linkage atoms, and little intrinsic effect on the ϕ, ψ conformation is expected. On the other hand, heavily substituted compounds such as peracetylated maltose might or might not take the same conformations as maltose itself. It is our experience that a wide variety of compounds can be considered. Chemically altered disaccharides, as long as they have the same backbone as the disaccharide in question, have conformations that are similar to those of molecules whose substitution is not a concern. Ultimately, the best way to test the potential energy function may be to construct energy surfaces for each compound. It is interesting, however, to see molecules with varying extents of modification on the same surface, so the present approach, which takes much less effort, is valuable as well. In our paper on sucrose, we segregated the crystalline molecules into two categories. In one, differences, if any, were distant from the linkage. The other category included molecules that had substitution near the linkage or ions in the structure.[9] The second category had higher corresponding energies although there was substantial overlap between the two sets.

Although we included many different compounds that had the same backbone structure, we systematically excluded those having additional covalent linkages between the monomeric units. That criterion also excludes the cyclic compounds such as the cycloamyloses for the present paper. Overall, we assume that the external forces that cause variations from the conformation of lowest energy in different crystal structures are random. In the case of cellobiose linkages in crystals of the cellodextrins, there appears to be a preferred packing mode. Implications of the preferred packing mode will be explored in a forthcoming paper on that molecule.[11]

Experimental values of ϕ and ψ were taken from the Cambridge Structural Database (CSD).[49] We used the atomic coordinates without any additional tests, such as upper limits on the crystallographic R factor. This unusual approach is based on two ideas. One is that the modern determinations are pretty good, and second, the most distorted structures are apt to yield the poorest crystallographic determinations. This is not to say that accuracy is unimportant. The gradients on the energy surfaces are as high as 0.25 kcal degree^{-1} not far from the minimum, so an error of a few degrees is important when making a quantitative analysis. To avoid skewing the data, we used only the most recent determination for a given crystal structure. Most structures in the CSD have been checked for the correct chirality, but we have encountered several structures in the CSD for which the coordinates are for the mirror image of the correct form. These structures can be used, but only after converting to the modeled enantiomer. We also collected conformations from the Protein Data Bank,[50] but the complexes of carbohydrates with proteins require extra discussion and will be treated in the separate reports on the specific molecules.[11]

Where we had used hydrogen-based definitions of the driven torsion angles, we simulated the experimental torsion angles by adding or subtracting, as appropriate, 120° to the torsion for the heavy atoms. Tvaroška (personal communication) recommended that 117° would be better because C-C-H and O-C-H bond angles are smaller than O-C-C bond angles. This agrees with our varying calculated difference between the hydrogen- and heavy-atom–based torsion angles of 116–119° for the sucrose analog structures calculated with QM over the range of ϕ and ψ values. For the experimental structures with the same backbone as cellobiose, regression between the torsion angles based on the reported coordinates of hydrogen and the torsion based on the experimental coordinates of O5 or C4′ with the 120° correction was satisfactory. The slope was 1.01, and R = 0.985. That work emphasized that there was substantial discrepancy in the experimental hydrogen positions, however, as

much as 28°, compared to the values obtained from the heavy atoms. The root mean square discrepancy was 5.7°, even though the average signed discrepancy was only −0.04°. All in all, it would be better to use the heavy atoms to drive the torsional rotations and characterize the crystal conformations.

The pairs of experimental ϕ and ψ values are placed on the energy maps by Surfer, which also estimates the corresponding energy of each experimental conformation with its residuals calculation feature. From these energy values, an average relative energy for the crystal structures was determined for each compound.

Distribution of Energies

The general idea of computed Ramachandran plots is that the observed conformations will be found in regions with low energies, and that, when enough observations are available, the lowest-energy regions will be fully occupied. Crystal structures could possibly give exceptions. Some molecules may have a preferred packing mode, concentrating the structures in just one region. Alternatively, a low-energy conformation may not fit into a crystal lattice with appropriately high density, so that region of ϕ, ψ space will not be populated by crystal structures. Recent work on lactose proposed that the relative propensity toward crystal growth might be related to the radii of gyration of the different conformations.[51] In particular, a "folded" conformation of lactose, so far unobserved, which has a small radius of gyration, was considered less likely to crystallize despite a lower calculated potential energy and a calculated free energy only 0.2 kcal mol^{-1} higher than the global minimum.

Opinions are diverse in this area. On one hand, some workers have derived force fields for aiding the crystallographic refinement of proteins from how frequently the various conformations are observed in crystal structures.[52] In only partial concurrence, Bürgi and Dunitz[53] and Bye et al.[54] agree that the number of experimental observations should decrease exponentially with increases in energy. This is a characteristic of a Boltzmann distribution, which should apply to isolated molecules. Although Dunitz et al. agree that a Boltzmann-type equation may apply, they argue that the value corresponding to the temperature in the actual Boltzmann equation cannot be known for this type of relationship. Therefore, force fields predicted from a distribution of crystalline conformations are arbitrary. Other workers would argue that the forces of crystallization are potentially so strong and nonrandom that a detailed quantitative analysis is inappropriate. In this case, Ramachandran surfaces could only be a crude guide. Here, we describe our analysis at the level of the hypothesis described by Dunitz et al.

Because the relative energies of the crystal structure conformations for all of our molecules were calculated in the same way, they were pooled to enhance the statistical significance of the results. The calculated energies of the crystal structure observations were sorted and ranked lowest to highest. The 153 sorted energy values were plotted along the x-axis with the y-axis corresponding to y = 1 − (rank/no. of observations). Our first y value is 1 − 1/153, and the last is 1 − 153/153 (= 0). A curve having the following equation then can be fitted to the resulting data points by Equation 1:

$$p = p_0 + Ae^{(-\Delta E/\beta)}, \quad (1)$$

where values of p_0, A, and β are constants. If the data fit the curve well, the value of β is approximately the average energy. With p_0 held at 0 and A held at 1, this function reduces to a Boltzmann-type distribution as given in Equation 2:

$$p_i = e^{(-\Delta G°_i/RT)}, \quad (2)$$

where p_i is the probability of the i^{th} value, R is the universal gas constant, T is the temperature, and $\Delta G°$ is the calculated standard state free energy. Detailed data are given in French et al.[9] so that a practical understanding can be attained with minimal effort.

In the present work, we have taken the relative potential energy as the relative free energy, even though it has been shown that the relative vibrational free energies of the various minima differ from the corresponding potential energies for lactose by about 0.5 kcal.[51] (Because the linkage torsional angles are restrained in Ramachandran plots, MM3 does not provide for an energy surface that would show the remainder of the vibrational free energy. Surfaces derived from molecular dynamics generally incorporate at least some of the vibrational free energy.[5–7]) Hydrogen bonding free energies are affected substantially because of the substantial loss in entropy on formation of the hydrogen bond. Another shortcoming of the adiabatic map is that it makes no allowance for the possibility that some ϕ, ψ conformations may have greater entropy because more of their exocyclic group orientations have low energy.

RESULTS AND DISCUSSION

The average energies for the conformations from crystal structures of the various molecules are listed in Table 1 for the three different dielectric constants. Given the limited number of observed structures, the low energies for the α,β-, and β,β-trehaloses may be just coincidental. In the case of α,α-trehalose, which has 19 observations, the average relative energy is almost invariant with dielectric constant. This is because no interresidue hydrogen bonding is possible when the molecular conformation is in the neighborhood of the global minimum and the various dipole–dipole interactions are small and do not change much over the limited range of ϕ and ψ

Table 1. Average energies (kcal mol^{-1}) of the disaccharides at three dielectric constants (ϵ) from the respective hybrid surfaces

		ϵ		
Disaccharide molecule	N	1.5	3.5	7.5
Cellobiose	50	1.30	0.97	1.19
Laminarabiose	20	1.77	0.85	0.68
Maltose	30	2.29	0.89	0.47
Nigerose	7	3.83	1.32	1.03
Sucrose (THP-O-THF analog)	23	3.06	1.16	0.93
Sucrose (THP-O-Me-THF analog)	23	3.17	1.24	0.99
α,α-trehalose	19	0.52	0.47	0.48
α,β-trehalose	1	0.01	0.29	0.01
β,β-trehalose	3	0.35	0.14	0.16

Crystal structures are listed in the Supplementary Material of Reference 11.

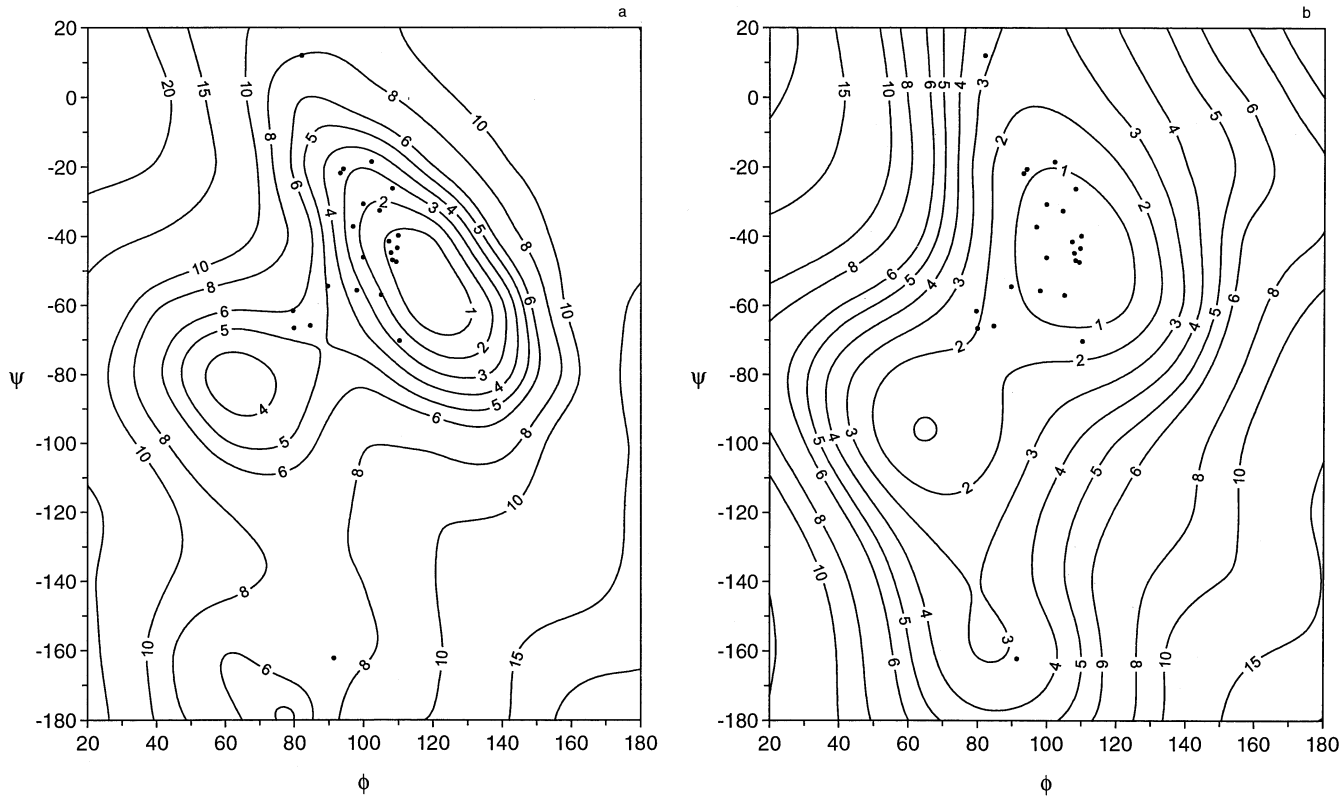

Figure 3. Hybrid ϕ, ψ energy maps of sucrose. (a) ϵ = 1.5. (b) ϵ = 7.5. The ϕ, ψ values of crystal structures from the CSD are posted on the maps. Only the region of the global minimum is shown.

values. Studies of sucrose were based on two different analogs. The second analog, used for the maps in this paper, has a methyl group attached to the anomeric carbon on the fructose side of the linkage instead of just a hydrogen atom on the analog used in the sucrose paper.[9] The hybrid energies (Table 1) and surfaces based on both analogs were similar, even

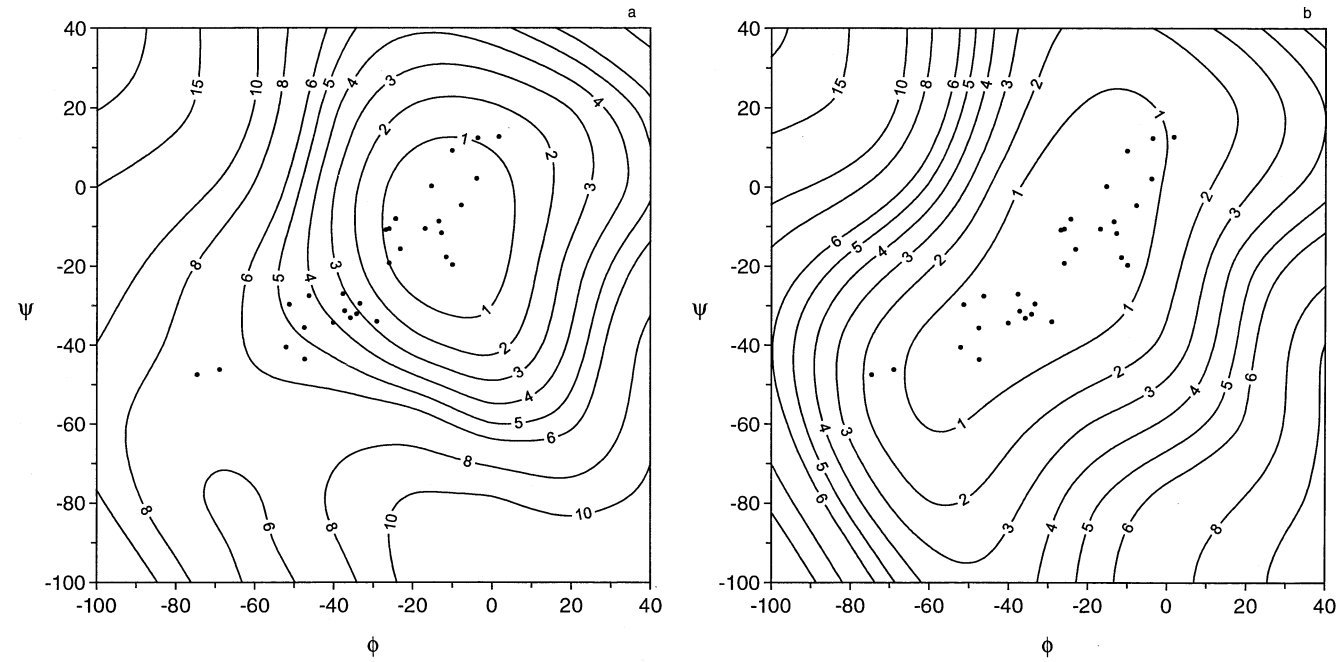

Figure 4. Hybrid ϕ, ψ energy maps of maltose. (a) ϵ = 1.5. (b) ϵ = 7.5. The ϕ, ψ values of crystal structures from the CSD are posted on the maps. Only the region of the global minimum is shown.

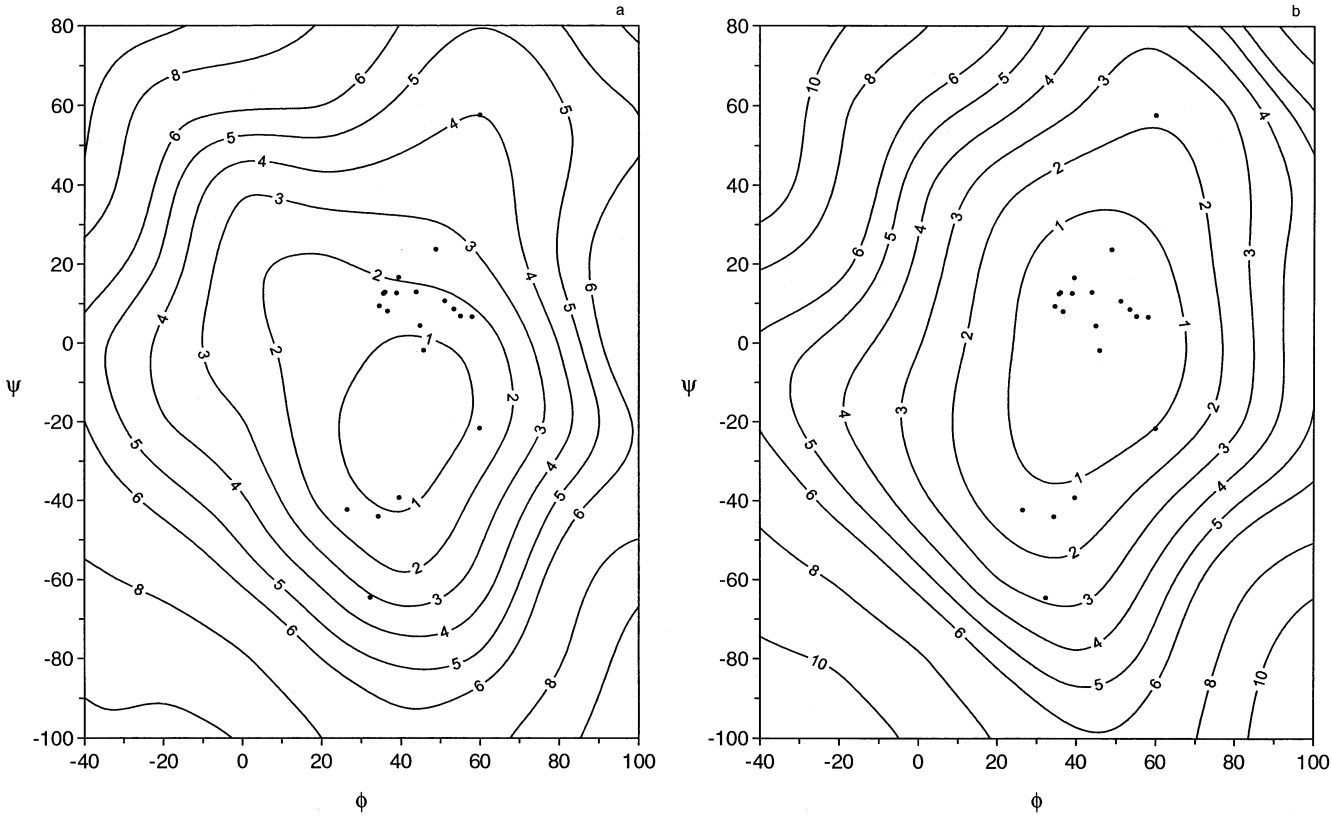

Figure 5. Hybrid ϕ, ψ energy maps of laminarabiose. (a) $\epsilon = 1.5$. (b) $\epsilon = 7.5$. The ϕ, ψ values of crystal structures from the CSD are posted on the maps. Only the region of the global minimum is shown.

though the methyl group causes a substantial difference between the analog surfaces.

Figures 3–5 show the crystallographically occupied portions of the energy surfaces that were calculated with our hybrid method and dielectric constants of 1.5 and 7.5. The minima on these surfaces are the global minima. On the sucrose map (Fig. 3a), the experimental conformations are mostly "west" (to the left of) or "northwest" (left and above) of the global minimum. The highest corresponding energy is 8 kcal mol^{-1}, an unlikely value for a conformation in solution. If such a surface does accurately reflect the conformational energy of crystalline sucrose moieties then it has little predictive utility. The fit of crystal structures on the flatter Figure 3b contrasts sharply with that in Figure 3a. The larger range of energies at the lower dielectric constant results because the global minima on both maps are stabilized by two inter residue hydrogen bonds. However, they only lower the energy by 2 kcal in Figure 3b, but they lower the energy by about 10 kcal mol^{-1} in Figure 3a. In Figure 3a, the conformations that cannot have intramolecular hydrogen bonds are penalized much more for their inability to form the hydrogen bonds. The reduced strength of the hydrogen bonds in the model has two effects. The energy well is not as deep, causing enlargement of the contours at a particular energy value. That leads to a lower average energy. Also, the reduced strength allows other contributions, such as the torsional and bending energies to dominate the energy surface, and this accounts for the relocation and reshaping of the energy contours near the minimum.

Figure 4a shows that the maltose conformations fall on nearly a diagonal line, with seemingly little regard for the energy calculated at a $\epsilon = 1.5$. While the area within the 1 kcal mol^{-1} contour is more populated than its counterpart in Figure 3a, the map has only modest predictive utility. Much of the space with energies in the 2- to 6-kcal range is not occupied, even though a number of the observed structures have corresponding energies that high. On the other hand, 27 (90%) of the 30 crystal structures are within the 1 kcal mol^{-1} contour in Figure 4b, and the contour line fits the observed structures well. There is one interresidue hydrogen bond possible near the global minimum (either O2-H$\cdots$O3′ or O3′-H$\cdots$O2), so the $\epsilon = 7.5$ surface is flattened somewhat. More dramatic here, however, is the elongation of the 1 kcal mol^{-1} contour at $\epsilon = 7.5$. The other minimum on the maltose surface, not shown in Figures 4a and 4b, is populated by a few linkages from larger cycloamyloses and will be discussed in a planned paper on maltose.[11]

Figure 5a shows the laminarabiose surface at $\epsilon = 1.5$ with observed conformations. Here, it appears that there could be insufficient crystal structures to form definite conclusions. For example, the observed structures seem to fall into three different clusters, and there might be two or three preferred packing modes. On the other hand, Figure 5b provides a much better accommodation of the observed structures. In this case, $\epsilon = 7.5$ did little to flatten the energy surface, despite the availability of an interresidue hydrogen bond in the region of the global minimum.

Figure 6 shows the cumulative frequency of observed structures *vs* their energies calculated at the three dielectric con-

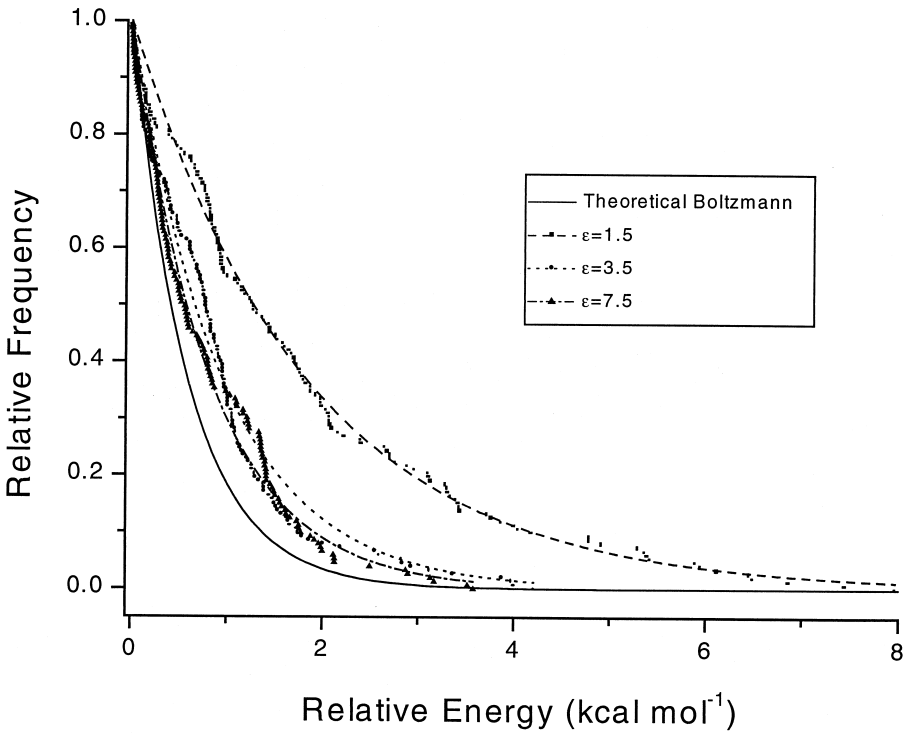

Figure 6. Boltzmann-type exponential decay curves for the pooled crystal structure energy values at various dielectric constant values, ϵ. The theoretical Boltzmann curve is that of Equation 2. The calculated values of β, from Equation 1, are as follows: $\epsilon = 1.5$, $\beta = 1.83$; $\epsilon = 3.5$, $\beta = 0.95$; and $\epsilon = 7.5$, $\beta = 0.83$. For sucrose, only the THF-O-THP energies were used.

stants, as well as the curves fit to those data. These graphs are similar to histograms, except that there are no bins. In an ordinary histogram, the number of observed structures between 0 and 0.5 kcal mol^{-1} might be used to make one bar, and the number of observed structures with corresponding energies between 0.5 and 1.0 kcal be used to make a second bar. A curve then could be fitted to the tops of the bars, similar to the curve for the cumultative frequency values at any given dielectric constant in Figure 6. The solid line corresponds to the ideal theoretical Boltzmann line, and its β-coefficient has a value of 0.59 kcal mol^{-1} (RT at 298K.) On this graph, half of the observed structures are above a horizontal line (not shown) intersecting the y-axis at the relative frequency of 0.5, and half are below. A vertical 1 kcal mol^{-1} line (not shown) intersects the Boltzmann curve at a cumulative frequency (y-axis value) of 18%, showing that 18% of the structures have higher energies. On our curve for the energy surfaces for $\epsilon = 7.5$, more than 30% of the structures have energies >1.0 kcal. More than 60% have energies greater than 1.0 kcal when they are computed at $\epsilon = 1.5$. All three fitted curves meet the criterion of Bürgi and Dunitz[53] and Bye et al.[54] in that the energies drop off in an exponential manner. The corresponding "temperatures" ($T = \beta/R$) for the fitted curves for the dielectric constants of 1.5, 3.5, and 7.5 are 924K, 480K, and 420K, respectively. Because the data for the 3.5 and 7.5 dielectric constants cross over each other, it is not clear that one is significantly closer to the ideal line than the other, but both are closer than the 1.5 data. Because all three dielectric constants give exponential decay, that criterion is not sufficient to choose which is the best in the case of our surfaces. On the other hand, when energy surfaces do not yield corresponding energies that conform to any exponential decay curve (see Figure 3b in French et al.[9]), the predictive utility of those energies is in doubt.

A consequence of this type of analysis is that the information on the distribution on the individual energy surfaces is lost. By reviewing Figures 3–5, we see that the two-dimensional distributions are generally superior for the $\epsilon = 7.5$ surfaces. On the surfaces with $\epsilon = 1.5$, there are many unpopulated regions that have energies equal to or lower than the energies corresponding to the crystal structures. The superior distribution on the surfaces with the raised dielectric constant, plus the closer conformity to the theoretical distribution curve in Figure 6, make a compelling case for the use of increased dielectric constant when predicting conformations that will be found in the solid state with our hybrid energy calculation for isolated molecules. The conclusions about dielectric constant with nonhybrid, MM3-only calculations are very similar.

CONCLUSIONS

When carefully constructed, Ramachandran surfaces can have substantial predictive utility for conformations that will be found in the solid state without explicit consideration of neighboring molecules. This claim is based on the finding that the average distortion energy was less than 1 kcal mol^{-1} at either of two dielectric constants for disaccharides that have various axial and equatorial configurations of the two bonds in their inter residue linkages. Also, the observed structures were distributed more or less randomly about the global minimum in the energy, and the frequency of observation of experimental structures diminished exponentially as their calculated energy increased. Although it would be more physically realistic to explicitly include neighboring molecules, that time-consuming consideration was not a practical necessity at this level of prediction.

Besides the need to have a good method to calculate the stereoelectronic energy, this paper described some of the technical considerations that affect energy surfaces and a number of

ways to avoid errors when testing them against crystallographic data. Perhaps most important were the use of a good technique for scanning ϕ, ψ space and getting a value of zero for the interpolated minimum energy on the contour plots. Also, the use of torsion angles based on hydrogen atom positions should be avoided, either for driving the linkage torsion angles or for determining conformations in X-ray diffraction studies. The other theme in the present paper is that the predictive power of the energy surfaces is enhanced when the strength of hydrogen bonding in the model is reduced.

ACKNOWLEDGMENT

The authors are indebted to Professor Kjeld Rasmussen for extensive and challenging comments on a preprint of this paper.

REFERENCES

1 Sasisekharaan, V. Stereochemical criteria for polypeptides and proteins. *Collagen Proc. Symp.* Madras, India, 1962, Vol. 1060, pp. 39–78. Rao, V.S.R., Sundararajan, P. R., Ramakrishnan, C., and Ramachandran, G. N. *Conformation in Biopolymers.* Academic Press, London, 1963, Vol. 2

2 Rao, V.S.R., Qasba, P.K., Baslaji, P.V., and Chandrasekaran, R. *Conformation of carbohydrates.* Harwood Academic, Amsterdam, 1998

3 Kleywegt, G.J., and Jones, T.A. Phi/Psi-chology: Ramachandran revisited. *Structure* 1996, 4, 1395–1400

4 Allinger, N.L., Rahman, M., and Lii, J.-H. A molecular mechanics force field (MM3) for alcohols and ethers. *J. Am. Chem. Soc.* 1990, **112,** 8293–8307

5 Schmidt, R.K., Teo, B., and Brady, J.W. Use of umbrella sampling in the calculation of the potential of mean force for maltose in vacuum from molecular dynamics simulations. *J. Phys. Chem.* 1995, **99,** 11339–11343

6 Liu, Q., Schmidt, R.K., Teo, B., Karplus, P.A., and Brady, J.W. Molecular dynamics studies of the hydration of α,α-trehalose. *J. Am. Chem. Soc.* 1997, **119,** 7851–7862

7 Immel, S., and Lichtenthaler, F.W. The conformation of sucrose in water: A molecular dynamics approach. *Liebigs Ann.* 1995, 1925–1937

8 Pérez, S., Imberty, A., Engelsen, S.B., Gruza, J., Mazeau, K., Jimenez-Barbero, J., Poveda, A., Espinosa, J.F., van Eyck, B.P., Johnson, G.P., French, A.D., Kouwijzer, M.L.C.E., Grootenhuix, P.D.J., Bernardi, A., Raimondi, L., Senderowitz, H., Durier, V., Vergoten, G., and Rasmussen, Kj. A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates. *Carbohydr. Res.* 1998, **314,** 141–155

9 French, A.D., Kelterer, A.-M., Cramer, C.J., Johnson, G.P., and Dowd, M.K. A QM/MM conformational analysis of crystalline sucrose moieties. *Carbohydr. Res.* 2000, **326,** 305–322.

10 French, A.D., Schäfer, L., and Newton, S.Q. Overlapping anomeric effects in a sucrose analogue. *Carbohydr. Res.* 1993, **239,** 51–60. Van Alsenoy, C., French, A.D., Cao, M., Newton, S.Q., and Schäfer, L. Ab initio-MIA and molecular mechanics studies of the distorted sucrose linkage of raffinose. *J. Am. Chem. Soc.* 1994, **116,** 9590–9595. French, A.D., and Dowd, M.K. Exploration of disaccharide conformations by molecular mechanics. *J Mol. Struct. (Theochem)* 1993, **286,** 183–201

11 French, A.D., Kelterer, A.-M., Johnson, G.P., Dowd, M.K., and Cramer, C.J. HF/6–31G* energy surfaces for disaccharide analogs. *J. Comput. Chem.* (In press)

12 Woods, R. Carbohydrate force fields. In: *Encyclopedia of Computational Chemistry,* Schleyer, P.V.R., Allinger, N.L., Clark, T., Gasteiger, J., Kollman, P.A., Schaefer, H.F. III, and Schreiner, P.R., Eds., John Wiley & Sons, Chicester, 1998 , Vol. 1, pp. 220–233

13 Tvaroška, I., and Bleha, T. Anomeric and exo-anomeric effects in carbohydrate chemistry. *Adv. Carbohydr. Chem. Biochem.* 1989, **47,** 45–123

14 Kneisler, J.R., and Allinger, N.L. Ab initio and density functional theory study of structures and energies for dimehtoxymethane as a model for the anomeric effect. *J. Comput. Chem.* 1996, **17,** 757–766

15 Senderowitz, H., Parish, C., and Still, W.C. Carbohydrates: United atom AMBER* parameterization of pyranoses and simulations yielding anomeric free energies. *J. Am. Chem. Soc.* 1996, **118,** 2078–2086

16 Fabricius, J., Engelsen, S.B., and Rasmussen, Kj. The consistent force field. 4. An optimized set of potential energy functions for aliphatic and alicyclic ethers and anomeric carbon atoms. *New J. Chem.* 1995, **19,** 1123–1137

17 Homans, S.W. A molecular mechanical force field for the conformational analysis of oligosaccharides: Comparison of theoretical and crystal structures of Manα1-3Manβ1-4GlcNAc. *Biochemistry* 1990, **29,** 9110–9118

18 Hwang, M.-J, Ni, X., Ewig, C.S., and Hagler, A.T. Derivation of class II force fields. VI. Carbohydrate compounds and anomeric effects. *Biopolymers* 1998, **45,** 435–468

19 Dapprich, S., Komáromi, I., Byun, K.S., Morokuma, K., and Frisch, M.J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct. (Theochem)* 1999, **461–462,** 1–21

20 Barrows, S.E., Dulles, F.J., Cramer, C.J., French, A.D., and Truhlar, D.G. Relative stability of alternative chair forms and hydroxymethyl conformations of β-D-glucopyranose. *Carbohydr. Res.* 1995, **276,** 219–251

21 Barrows, S.E., Storer, J.W., Cramer, C.J., French, A.D., and Truhlar, D.G. Factors controlling the relative stability of anomers and hydroxymethyl conformers of glucopyranose. *J. Comput. Chem.* 1998, **19,** 1111–1129

22 French, A.D., Kelterer, A.-M., Johnson, G.P., and Dowd, M. K. B3LYP/6-31G*, HF/6-31G* and MM3 heats of formation for disaccharide analogs. *J. Mol. Struct.* (In press)

23 Labanowski, J., Schmitz, L., Chen, K.-H., and Allinger, N.L. Heats of formation of organic molecules calculated by density functional theory: II. Alkanes. *J. Comput. Chem.* 1998, **19,** 1421–1430

24 Lii, J. -H., Ma, B., and Allinger, N.L. The importance of selecting the proper basis set in the quantum mechanical studies of the potential energy surfaces of carbohydrates. *J. Comput. Chem.* 1999, **20,** 1593–1603

25 Csonka, G.I.; Éliás, K., and Csizmadia, I.G. Relative stability of $^1C_4$ and $^4C_1$ chair forms of β-D-glucose: A density functional study. *Chem. Phys. Lett.* 1996, **257,** 49–60

26 Csonka, G.I., Éliás, K., and Csizmadia, I.G., Ab initio

and density functional study of the conformational space of $^1C_4$ α-L-fucose, *J. Comput. Chem.* 1997 **18,** 330–342
27 Csonka, G.I., Éliás, K., Kolossváry, I., Sosa, C.P., and Csizmadia, I.G. Theoretical study of alternative ring forms of α-L-fucopyranose. *J. Phys. Chem. A* 1998, **102,** 1219–1229
28 Chem-X, version 1999.1, 1999, Oxford Molecular Ltd., http://www.oxmol.com/prods/chem-x/
29 RASMOL may be downloaded from http://www.umass.edu/microbio/rasmol
30 MOLDEN may be downloaded from http://www.caos.kun.nl/~schaft/molden/molden.html
31 MM3 is available to academic users from the Quantum Chemistry Program Exchange, Department of Chemistry, Indiana University, Bloomington, IN. http://qcpe.chem.indiana.edu. Commercial users may obtain MM3 from Tripos Associates, St. Louis, MO. http://www.tripos.com/software/mm3.html
32 French, A.D., Tran, V.H., and Pérez, S. Conformational analysis of a disaccharide (cellobiose) with the molecular mechanics program (MM2). *ACS Symp. Ser.* 1989, **430,** 120–140
33 Leach, A.R. *Molecular modelling: Principles and applications.* Addison Wesley Longman Limited, Essex, England, 1996, pp. 169–171
34 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., and Montgomery, J.A. General atomic and molecular electronic structure system. *J. Comput. Chem.* 1993, **14,** 1347–1363
35 Gaussian 94, Revision E.2, Frisch, M.J., Trucks, G.W., Schlegel, H.B., Gill, P.M.W., Johnson, B.G., Robb, M.A., Cheeseman, J.R., Keith, T., Petersson, G.A., Montgomery, J.A., Raghavachari, K., Al-Laham, M.A., Zakrzewski, V.G., Ortiz, J.V., Foresman, J.B., Cioslowski, J., Stefanov, B.B., Nanayakkara, A., Challacombe, M., Peng, C.Y., Ayala, P.Y., Chen, W., Wong, M.W., Andres, J.L., Replogle, E.S., Gomperts, R., Martin, R.L., Fox, D.J., Binkley, J.S., Defrees, D.J., Baker, J., Stewart, J.P., Head-Gordon, M., Gonzalez, C., and Pople, J.A. Gaussian, Inc., Pittsburgh, PA, 1995
36 Hanson, J.C., Sieker, L C., and Jensen, L.H. Sucrose. X-ray refinement and comparison with neutron refinement. *Acta Crystallogr.Sect. B* 1973, **29,** 797–808
37 Marchessault, R.H., and Pérez, S. Conformations of hydroxymethyl group in crystalline aldopyranoses. *Biopolymers* 1979, 18, 2369–2374
38 French, A.D., and Finch, P. Monosachcarides: Geometry and dynamics. In: *Carbohydrates. Structures, syntheses and dynamics,* Finch, P. Ed., Kluwer Academic Publishers, Dordrecht, 1999, pp. 1–46
39 Batta, G., and Kövér, K.E. Heteronuclear coupling constants of hydroxyl protons in a water solution of oligosaccharides: Trehalose and sucrose. *Carbohydr. Res.* 1999, **320,** 267–272
40 Simmerling, C., Fox, T., and Kollman, P.A. Use of locally enhanced sampling in free energy calculations: Testing and application to the αβ anomerization of glucose. *J. Am. Chem. Soc.* 1998, **120,** 5771–5782
41 Stortz, C. Disaccharide conformational maps: How adiabatic is an adiabatic map? *Carbohydr. Res.* 1999, **322,** 77–86
42 Dowd, M.K., French, A.D., and Reilly, P.J. Modeling of alddopyranosyl ring puckering with MM3(92). *Carbohydr. Res.* 1994, **264,** 1–19
43 Burkert, U., and Allinger, N.L. Pitfalls in the use of the torsion angle driving method for the calculation of conformational interconversions. *J. Comput. Chem.* 1982, **3,** 40–46
44 French, A.D., Tran, V., and Pérez, S. Conformational analysis of a disaccharide (cellobiose) with the molecular mechanics program (MM2). *ACS Symp. Ser.* 1990, **430** 191–212
45 French, A.D., Miller, D.P., and Aabloo, A. Miniature crystal models of cellulose polymorphs and other carbohydrates. *Int. J. Biol. Macromol.* 1993, **15,** 30–36
46 Users with MM3 source code can replace the maximum movement of 0.25 Å in earlier versions
47 Surfer, version 7.0.0, August 25, 1999, Golden Software, Inc., Golden, CO, http://www.goldensoftware.com/frames/surferframe.htm
48 Carlson, R.E., and Foley, T.A., *Radial Basis Interpolation Methods on Track Data,* 1991, Lawrence Livermore National Laboratory, UCRL-JC-1074238. Carlson, R.E., and Foley, T.A. *Computers Math. Applic.* 1991, **21,** 29–42
49 The Cambridge Structural Database is available from Cambridge Crystallographic Data Centre, http://www.ccdc.cam.ac.uk/prods/csd.html
50 We accessed the Protein Data Bank at http://www.rcsb.org/pdb/cgi/queryForm.cgi
51 Engelsen, S.B., and Rasmussen, K.J. The consistent force field. 5. PEF95SAC: Optimized potential energy function for alcohols and carbohydrates. *J. Carbohydr. Chem.* 1997, **16,** 773–788
52 Engh, R.A., and Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. Sect. A.* 1991, **47,** 392–400
53 Bürgi, H.-B., and Dunitz, J.D. Can statistical analysis of structural parameters from different crystal environments lead to quantitative energy relationships? *Acta Crystallogr. Sect. B.* 1988, **44,** 445–448
54 Bye, E., Schweizer, B., and Dunitz, J.D. Chemical reaction paths. 8. Stereoisomerization path for triphenylphosphine oxide and related molecules: Indirect observation of the structure of the transition state. *J. Am. Chem. Soc.* 1982, **104,** 5893–5898