ELSEVIER

Modeling polysaccharides: Present status and challenges

Serge Pérez,* Milou Kouwijzer,* Karim Mazeau,* and Søren Balling Engelsen†

**Centre de Recherches sur les Macromolécules Végétales, CNRS, Grenoble, France*
†The Royal Veterinary and Agricultural University, Food Technology, Frederiksberg, Denmark

The most recent tools that have been developed for modeling the three-dimensional features of polysaccharides and carbohydrate polymers are presented. The presentation starts with a description of the conformations of the monosaccharides, and of the flexible rings such as in the case of five-membered rings, and a thorough description of the conformational space that is available for a disaccharide unit, either in vacuo *or in an aqueous phase. The extension to the modeling of the parent polysaccharides is addressed, based on the assumption that owing to the size and relative rigidity of the intervening monosaccharides units, the rotations at a particular linkage can be, under some conditions, considered as independent of nearest neighbor interactions. Appropriate modeling techniques are described that can provide insights into the dimensions of the chain in a solution which is best described as a random coil accompanied by the occurrence of local "helical" regions. With the help of such descriptors such as helical parameters, the ordered state of polysaccharide strands can be readily characterized. The generation of double or triple helices can be then attempted in order to explore the occurrence of such multistranded arrangements that may be energetically stable. The final step in the determination of the structure of polysaccharides in the ordered state, is the investigation of the interactions of different helices. This may lead to either the best arrangement(s) between two polymeric chains, or to the prediction of the dimensions, and the symmetry of a three-dimensional lattice. Some of the tools which have been developed should allow automatic scarches for meaningful correlations between structures and functions, through exploratory data analysis. Structure-function or structure-property correlation could be then used to model changes arising from structural alterations. This would open the field of polysaccharide engineering.* © *Elsevier Science Inc., 1996*

Color Plates for this article are on page 361–362.

Send reprint requests to: Serge Pérez, Centre de Recherches sur les Macromolécules Végétales, CNRS, BP 53, F-38041 Grenoble Cedex 09, France.

Paper submitted as part of the Electronic Conference of the Molecular Graphics and Modelling Society, October 1996.

[1]Associated with the Université Joseph Fourier, Grenoble, France.

INTRODUCTION

Carbohydrates constitute one of the most abundant types of biomolecules occurring widely in all living matter. They function as structural or protective materials and as a form of energy storage. In addition, carbohydrates perform a much broader biological role.[1] For example, they appear to be essential in the process of infection by certain pathogenic species, they specify human blood group types and are intimately involved in the immunochemistry of blood, they determine cell–cell recognition and adhesion, they function as receptors in the antigen-stimulated lymphocyte antibody immune response, and they have an important role in cancer pathology. Their biological processes as well as their chemical and physical properties are in part determined by the conformational behavior of carbohydrates.

Carbohydrates are polyhydroxy aldehyde, ketone, acid, or alcohol compounds that have the general formula $(CH_2O)_n$, where n is between 4 and 9. Most of the carbon atoms are asymmetric, and their various stereoisomers possess distinct physical and biological properties. Figure 1 shows the Fischer projection formulas of the naturally occurring D-ose series (the sugars having the absolute configuration R at the last asymmetric carbon atom are called D); the L-series of carbohydrates occurs as well, but to a lesser extent. Other natural units are formally derived by simple modification of these sugars, such as with sulfate and acetate esters, methyl esters of carboxylic acid functions, acetals or ketals of simple carbonyl molecules. For example, in D-glucose the CH_2OH group may be replaced by COOH (e.g., glucuronic acid) CH_3 (e.g., fucose), or H (e.g., xylose); an OH group may be replaced by NH_2 (e.g., glucosamine) or $NHCOCH_3$ (e.g., *N*-acetylglucosamine). Sugar units also occur as simple derivatives. The number of monomeric carbohydrate structures is therefore gigantic.

The open-chain forms of monosaccharides are quite flexible, being capable of rotating about each of the single C–C bonds. Because the aldehyde and ketone groups of these molecules are quite reactive, this flexibility often leads to an

Journal of Molecular Graphics 14:307–321, 1996

655 Avenue of the Americas, New York, NY 10010

0263-7855/97/$15.00
PII S0263-7855(97)00011-8

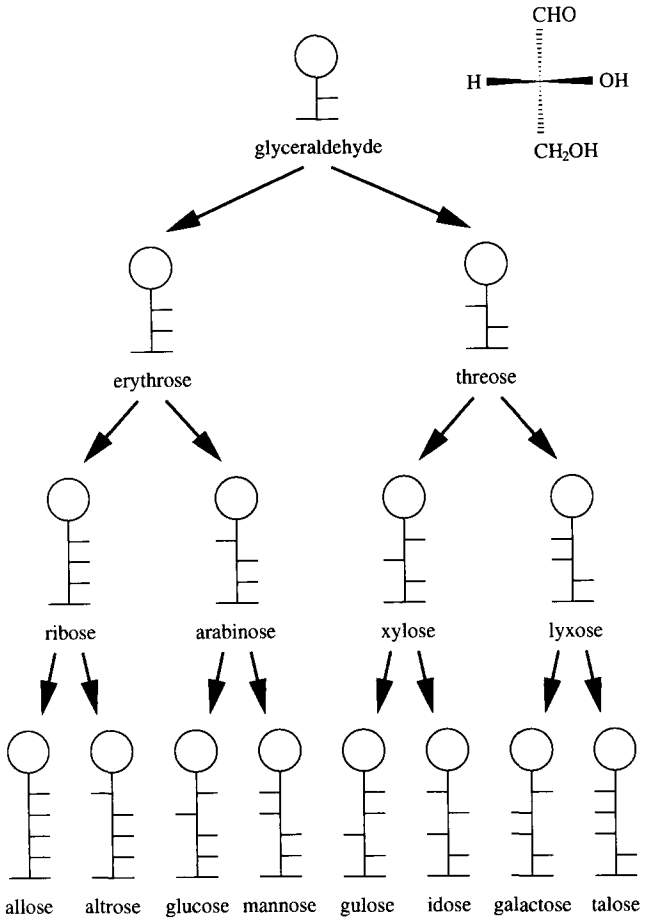

Figure 1. Fischer projection formulas of the naturally occurring D-ose series.

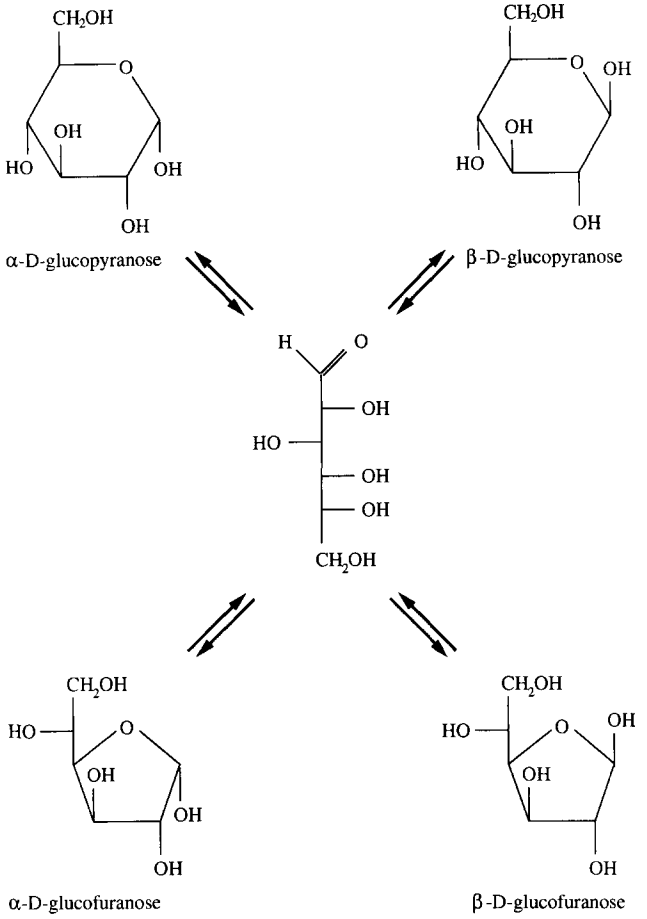

Figure 2. Mutarotation.

internal cyclization as the carbonyl group reacts with one of the hydroxyl groups from the other end of the molecule. Depending on the point of attack, the resulting rings can contain five or six atoms, one of which is oxygen; four- and seven-membered ring sugars are rather sparse. The closing of the linear molecule to make a ring creates a new chiral center at C1, called the *anomeric* carbon. There are two possibilities, which are designated with either an α or β prefix. Several forms of each of the above-mentioned monomers can exist. The cyclization is a reversible reaction, and in aqueous solution an equilibrium between the various forms will exist (see Figure 2). This is called *mutarotation.*

Monosaccharides can be linked to produce oligomeric structures through glycosidic bond formation. Water is eliminated between the anomeric hydroxyl and any one of the hydroxyls of a second monosaccharide or oligosaccharide. A disaccharide is shown in Figure 3. The glycosidic linkage consists of two bonds: the glycosidic C1–O and the aglycone O–Cx bonds. In the case of (1 → 6) linkages three bonds connect the consecutive sugar rings. Carbohydrates can exist as mono-, oligo-, and polysaccharides; however, they reach maximum complexity when they are covalently attached to other molecules in glycoproteins, glycolipids, or glycopeptidolipids. The number of possible oligo- or polysaccharide structures is enormous. There are 21 different compounds that can be formed by condensation of 2 D-glucose units. Since there are multiple points of possible linkage on a monosaccharide residue, branching of the chain is quite a common feature in carbohydrate-containing molecules. In theory, the number of all possible linear and branched isomers of a hexasaccharide was found to be more than 10^{12} (see Figure 4[2]). Not all of these actually occur in nature, but a large number of them do. The naturally occurring polysaccharide structures can be classified into different groups. Simple homopolysaccharide structures can be linear (cellulose, amylose) or branched (amylopectin, glycogen). Heteropolysaccharides with various degrees of branching can be alternating (agarose, carrageenan), block (alginate), complex linear repeat (gellan), complex

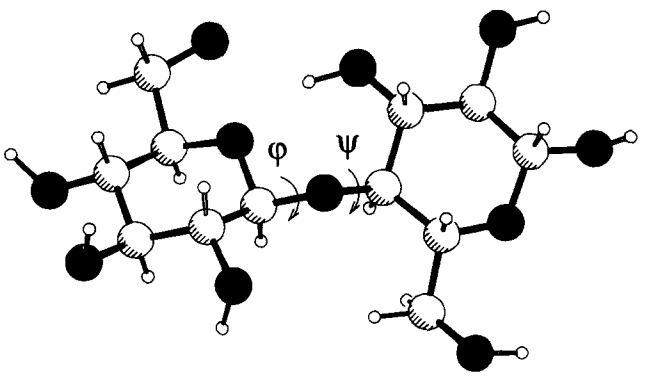

Figure 3. Cellobiose.

Number of possible isomers

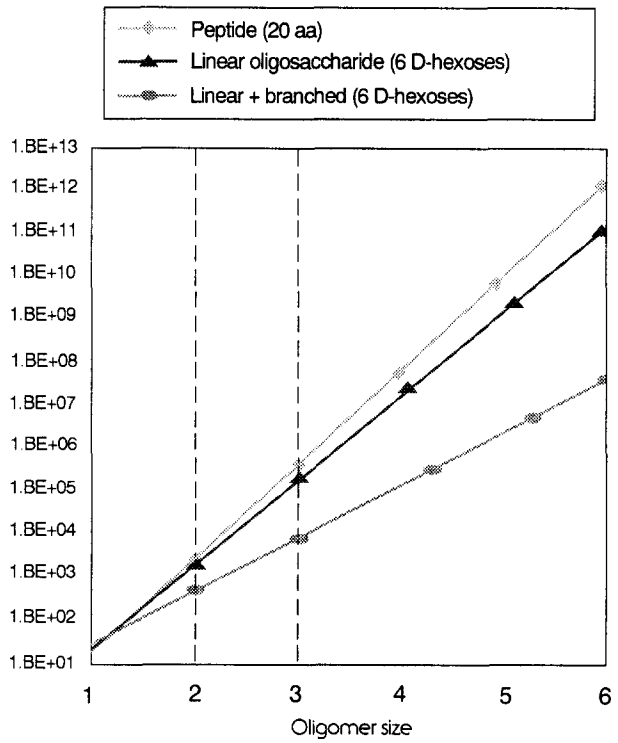

Figure 4. The isomeric barrier. [Adapted from Ref. 2.]

branched repeat (xanthan), and interrupted and branched (pectin). An illustration is given in Figure 5. The complex oligosaccharides constitute another class of carbohydrate-containing molecules. These molecules generally vary in size from disaccharides to oligosaccharides of 15 to 20 residues, but may be even larger.

The POLYS computer program has been developed to generate three-dimensional structures of polysaccharides and complex carbohydrates from their primary sequence.[3] POLYS combines a database of monosaccharide structures with a database containing population information on disaccharide fragments, using an approximation of independent neighboring glycosidic linkages. A primary structure in simple ASCII syntax is transformed by the program into Cartesian coordinates, which can be exported to frequently used formats. Also, the program can generate Boltzmann distributions of polysaccharide conformations from which configurational macroscopic properties such as persistence length and radius of gyration can be calculated and compared to experimental values derived from viscometry and light- and neutron-scattering studies. The program has been successfully applied to the study of configurational and conformational properties of heparin, and in structurally exploring pectic polysaccharides (both the homo- and the rhamnogalacturonan backbone, and the organization of arabinan and arabinogalactan side chains on the latter).

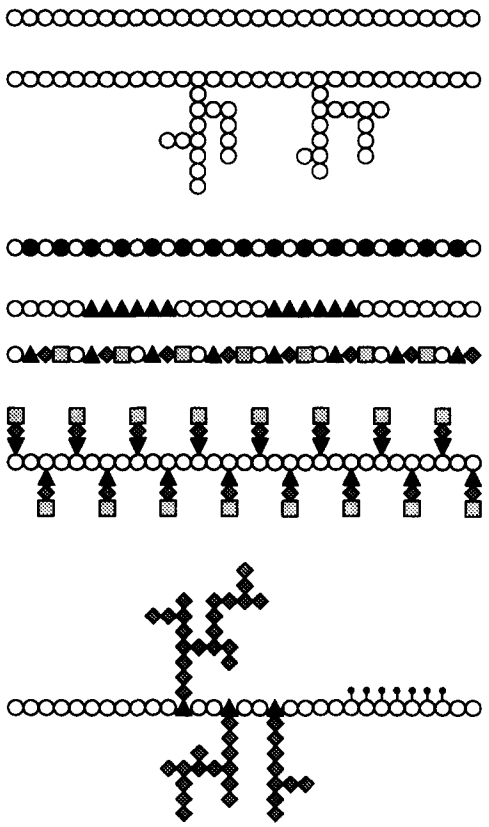

Figure 5. Examples of different types of polysaccharides. From top to bottom: linear, branched, alternating, block, complex linear repeat, complex branched repeat, and interrupted and branched.

MONOSACCHARIDES

Hemiacetal formation between OH-5 and the C1 aldehyde group in aldohexoses (glucose, for example) produces the pyranose ring, which contains an endocyclic oxygen atom

(O5) and a new asymmetric carbon center at C1. This results in a ring structure having six atoms.

The most stable conformations of six-membered ring systems (pyranoses) are the chair forms, but in the majority of naturally occurring pyranoid derivatives one of the two possible chair conformers is considerably more stable than the other. As in the case of the cyclohexane ring, the pyranoid ring can also adopt energetically less favorable conformations. Six different skew conformers separated by six different boat conformers can be identified on the pseudorotational itinerary of a pyranoid ring. The pseudorotation circle of the flexible forms of the pyranoid ring and the position of the chair conformers on the conformational sphere of the parameters Q, θ, and φ are shown in Figure 6.[4] The three puckering parameters define unambiguously the position of the individual forms of the pyranoid ring on the conformational sphere: Q is the maximum puckering amplitude, and the parameters θ and φ are angles in the range $0° < \theta \leq 180°$ and $0° < \varphi \leq 360°$, and can be thought of as polar and azimuthal angles for a sphere of radius Q. The two poles $\theta = 0°$ or 180° represent the energy wheels of the chair conformations $^1C_4$ and $^4C_1$. All 12 flexible forms are located at the equator. In unsubstituted cyclohexane the two chair forms are the prominent species. Substitution of a heteroatom in the ring and addition of hydroxyls or other exocyclic substituents further stabilize or destabilize ring conformers in relation to cyclohexane. As a general rule, the equatorial position of bulky substituents would be preferred because of a 1,3 syn-diaxial interaction that causes steric clashes. The $^4C_1$ of glucopyranose having all ring substituents in the equatorial position is preferred to the $^1C_4$ conformer, in which all the substituents are in axial orientation. However, at high temperature conformational transition to this form can arise spontaneously as demonstrated by the formation of levoglucosan (1,6-anhydro-β-D-glucopyranose). This molecule results from the $1 \rightarrow 6$ elimination of water from the unusual $^1C_4$ conformation that brings the two hydroxyl groups in close proximity. Besides the trivial case of glucopyranose, the α-L-iduronate ring (a constituent of the glycosaminoglycans heparin, heparan sulfate, and dermatan sulfate, which have potential uses as anticoagulant and antithrombotic agents) exhibits conformational mobility. Three forms, namely $^1C_4$, $^2S_0$, and $^4C_1$ of this ring, have been suggested to be responsible of the biological activities of these compounds.

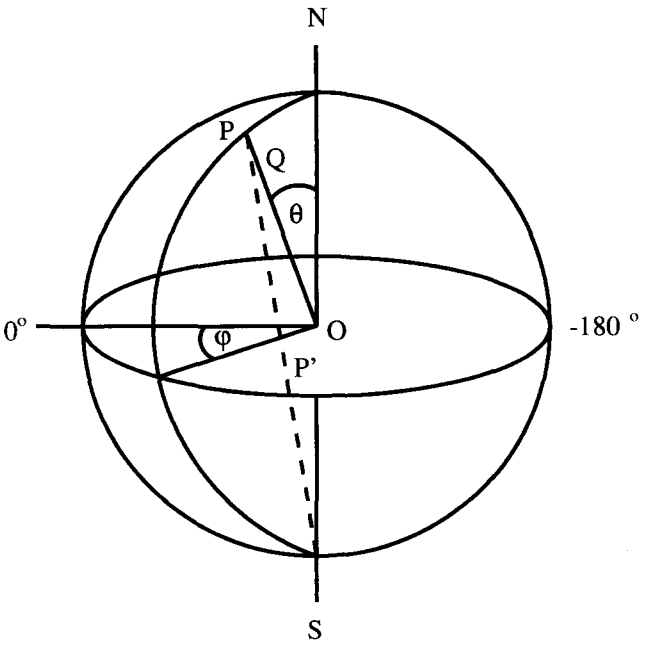

Figure 6. Puckering parameters for pyranoid rings.[4]

The five-membered ring forms (furanoses) of D-glucose are not thermodynamically stable. However, important furanose molecules are commonly found in nature. For example, D-ribose and D-deoxyribose are found as the building units of nucleic acids, and fructose is a constituent of sucrose. These rings are not planar; either one (envelope form) or two (twist form) atoms are out of the plane containing the others. The different envelope and twist conformations are of similar energy, and the barrier to their interconversions is rather small. Therefore, mixtures of different conformations might be expected in solution. Two puckering parameters are needed to define a conformation for furanoid rings: the puckering amplitude Q and the phase angle φ. Pseudorotation energy surfaces of ribofuranose along with its deoxy analog,[5] fructofuranose,[6] and arabinofuranose[7] have been reported. In general, furanose rings have two major local minima and a path of interconversion. Because eclipsing of a carbon atom and an oxygen atom requires less energy than that of two carbon atoms, the oxygen atom in the furanoid ring tends to occupy the least puckered part of the ring; hence, usually either C2 or C3, or both, will be out of the plane. In most cases, the conformation is an intermediate between the ideal envelope and twist forms; it is governed by the relative disposition of the substituents.

The conformations of monosaccharides have been determined by several methods including the formation of complexes from sugars in cuprammonia solution, X-ray crystal structure analysis, infrared spectroscopy, polarimetry, and optical rotation. In solution nuclear magnetic resonance (NMR) spectroscopy is usually the method of choice.

The energies of the different ring conformations are affected by the orientations of the hydroxymethyl group. This group usually exists in one of the three staggered positions called gauche–gauche, gauche–trans, and trans–gauche. In this terminology, the torsion angle O5–C5–C6–O6 is stated first, followed by the torsion angle C4–C5–C6–O6. It is known from crystallographic studies, NMR measurements, and theoretical calculations that the conformational equilibrium about the C5–C6 bond in aldopyranoses depends significantly on the configuration at C4. For the "gluco" configuration (O4 equatorial) the trans–gauche position is high in energy and the remaining two conformations are almost equally populated, while the trans–gauche and gauche–trans positions are preferred for those having a "galacto" configuration (O4 axial).

The hydroxyl groups have a high freedom of rotation, and they can participate in the creation of hydrogen bonds. As a result of the many possible orientations of such groups, prediction of the hydrogen-bonding network is a difficult task. Most of carbohydrates offer an exceptionally high ratio of hydroxyl groups per saccharide residues. Hydrogen bonds are usually present to either neighboring carbohydrate molecules, glycoproteins, or surrounding water molecules.

The extreme diversity of the monomeric units that can be found in carbohydrate structures stressed the need for specific methodologies to facilitate the construction of complex carbohydrates. A carbohydrate fragment library has been created.[8] This databank contains optimized geometries of many monosaccharide residues and covers most of the units that occur in polysaccharides.

The anomeric effect describes the axial preference for an electronegative substituent of the pyranose ring adjacent to the ring oxygen, whereas the exoanomeric effect describes the rotational preference of the glycosidic C1–O bonds.[9] These stereoelectronic effects are of general importance for all molecules having two heteroatoms linked to a tetrahedral center. A survey of X-ray crystallographic data reveals that these effects have geometric consequences. The most obvious feature of the experimental data on both α and β configurations is a marked difference in the molecular geometry around the acetal group. For example, in the axial configuration one observes a general shortening of the O5–C1 bond, a lengthening of the C1–O bond, and an increase in the O5–C1–O bond angle value. Molecular orbital theory accounts for these observations. The magnitude of the anomeric effect varies with the nature of the electronegative group, the polarity of the solvent, and the location of the other substituents in the molecule.

The exoanomeric effect influences the rotations around the glycosidic C1–O bond and is therefore important in determining the relative orientations of saccharide units of carbohydrate chains. The exoanomeric effect is a balance between electronic and steric effects. The three staggered orientations for rotation about the glycosidic bond are not equivalent; the exoanomeric effect causes preference for the +synclinal orientation of the aglycone group in the α series and −synclinal for the β series. A review of these effects has been published by Tvaroska and Bleha.[9]

Many force fields for molecular modeling are available. The following force fields are widely used or especially designed for carbohydrates:

- The GROMOS force field was developed for molecular dynamics simulations of proteins, nucleotides, or sugars in aqueous or apolar solutions or in crystalline form[10]
- The MM2 and MM3 force fields are molecular mechanics force fields initially meant for hydrocarbons, but now applicable to a wide range of compounds[11,12]; Tvaroska and Pérez published a modified version especially for oligosaccharides called MM2CARB[13]
- The CHARMm force field is designed for the modeling (both molecular mechanics and dynamics calculations) of macromolecular systems.[14] A revision for carbohydrates was made by Ha et al.[15] Kouwijzer and Grootenhuis redeveloped the CHEAT force field: a CHARMm-based force field for carbohydrates with which a molecule in aqueous solution is mimicked by a simulation of the isolated molecule[16,17]
- The AMBER force field was developed for simulations of proteins and nucleic acids.[18] A derivative for conformational analysis of oligosaccharides was published by Homans.[19] Glennon et al.[20] presented an AMBER-based force field especially for monosaccharides and (1 → 4)-linked polysaccharides. More recently, Woods et al. developed the GLYCAM parameter set for molecular dynamics simulations of glycoproteins and oligosaccharides that is consistent with AMBER[21]
- The consistent force field (CFF) was originally a molecular mechanics force field for cycloalkane and *n*-alkane molecules, optimized on both structural and vibrational data.[22] Later, several versions for other classes of compounds were published; among others for carbohydrates[23,24]
- The TRIPOS molecular mechanics force field is designed to simulate both biomolecules (peptides) and small organic molecules.[25] Additional parameters for conformational analysis of oligosaccharides were derived by Imberty et al.[26]
- The DREIDING force field is one of the newer force fields in this list, and it was developed for the simulation of organic, biological, and main-group inorganic molecules[27]
- The Merck Molecular Force Field (MMFF94) has been published.[28] It seeks to achieve MM3-like accuracy for small molecules in a combined "organic/protein" force field that is equally applicable to proteins and other systems of biological significance

DISACCHARIDES

The low-energy conformers of a disaccharide can be estimated using molecular mechanics. In such compounds the global shape depends mainly on rotations about the glycosidic linkages, because the flexibility of the pyranose ring is rather limited and the different orientations of the pendent groups have a limited influence on the conformational space of the disaccharide. The relative orientations of saccharide units are therefore expressed in terms of the glycosidic linkage torsional angles φ and ψ, which have the definition φ = O5–C1–O–C'_x and ψ = C1–O–C'_x–$C'_{(x-1)}$ for a $(1 \rightarrow x)$ linkage. The φ, ψ space can be explored in a systematic way. Both torsions are sequentially rotated in small increments over the full 360° range. At each point of the grid the energy according to the force field in use is calculated. It is then possible to represent the energies of all the conformations available as a contour map in the φ, ψ space. These contour maps enable graphical description of energy changes as a function of the relative orientation of the monosaccharides. They indicate the shape and position of minima, the routes for interconversion between conformers, and the heights of the transitional barriers. There are many different methods for calculating contour maps.

Calculating potential energy surfaces

In the rigid residue, or hard sphere potential surfaces approach, the constituent monosaccharides are assumed to be rigid, with pendent groups fixed. As the φ, ψ values are changed, steric interactions between the pendent groups do occur, which are unable to relax. These steric interactions cause a rapid increase in energy. This effect is especially prominent in sterically crowded molecules. In addition, surveys of a large number of known crystal structures along with and supported by semiempirical calculations reveal small but important variations in pyranoid ring geometries and orientations of pendent groups with the φ, ψ values.

These are dependent on the anomeric and exoanomeric effects and emphasize the need for a model to include bond length and angle degrees of freedom.

The strain produced by steric interactions inherent from rotation of monosaccharide residues is relieved by the inclusion of bond length and angle adjustment in the form of minimization with respect to all degrees of freedom of the system (except φ and ψ) at each grid point. During minimization pendent groups move to the nearest minimum downhill of the starting point. In the process of driving the molecule through unfavorable regions of the φ, ψ space, large steric interactions can sometimes cause pendent groups to overcome torsional barriers. This results in minimization to a different local well. This relaxed map describes a larger accessible potential energy surface than the rigid maps, a lowering of the energy barriers between minima, and a lower energy minimum far removed from the initial starting geometry.[29]

Whereas rigid residue maps represent a two-dimensional cross-section of a $3N - 6$ dimensional surface, where N is the number of atoms, relaxed maps represent a larger cross-sectional window of a given potential energy surface because they allow minimization of the internal coordinates (bond lengths, bond angles, and torsional angles) to local low energy wells. However, as minimization will lead only to conformations "downhill" from the starting structure, the torsional dimension where most conformational variation occurs is limited to only one orientational well. It is possible that rotation of pendent groups over torsional barriers could produce lower energy conformations at that point in the φ, ψ space. Ideally at each point in this space an investigation of all possible combinations of pendent group orientations is required (i.e., assuming that each pendent group can exist in each of the three idealized staggered orientation, 3^n different conformations at each point in the φ, ψ space, where n is the number of pendent torsions). This results in 3^{12} (531 441) conformations for a simple disaccharide and 3^{19} (1.16×10^9) conformations for a more complex disaccharide typical of heparin.

Adiabatic maps attempt to represent the lowest energy of all possible pendent group orientations at each point in the φ, ψ space. On comparison with the corresponding relaxed maps, adiabatic maps are flatter, allow greater freedom about the glycosidic bonds, locate additional minima, and reduce the barriers between the minima.

At present there are several different methods for calculating adiabatic conformational maps: In the most commonly used method, the energy at each point in φ, ψ space for several different starting geometries is evaluated systematically and the lowest energy for each point is used to generate the map. This can be time consuming, therefore such a systematic search is possible only for carbohydrates of limited size and flexibility.

Several procedures have been developed to scan the energy surface as a function of the two glycosidic angles in an efficient way. For example, the Random Molecular Mechanics (RAMM)[30] grid method searches the orientation of pendent groups at each point in φ, ψ space. At each point, 1 000 steps of a random walk procedure vary pendent group orientation and evaluate unrelaxed energies. Only the resultant lowest energy structure is optimized, and accepted as the energy for that point in the φ, ψ space.

The prudent ascent method moves through the φ, ψ space in a way that is dependent on previous minimizations. Large steric interactions are minimized by doing the most favorable geometries first, in a way similar to the local relaxed map. It makes use of inelastic deformations that decrease the energy by recalculating the energies of surrounding geometries using the new lower energy structure as the starting geometry. On average the energy for each point in the φ, ψ space is calculated twice.[31]

With the CICADA method (Channels In Conformational space Analyzed by Driver Approach) the potential energy surface is explored by driving separately each selected torsion angle with a concomitant full-geometry optimization at each increment (except for the driven angle).[32]

The Monte Carlo method is essentially a random search method. From a starting configuration (A) a new configuration (B) is generated by random displacement of one or more atom(s). The new configuration is either accepted or rejected, on the basis of an energy criterion. When the energy of B is lower than or equal to that of A, B, will be accepted. When it is higher, it will be accepted only if the Boltzmann factor (for the desired temperature) is greater than a random number taken from a uniform distribution between 0 and 1. When B is rejected, A is counted again before a new configuration will be generated; when B is accepted, it will serve as a new starting configuration. The process is repeated many times, and results in a large number of configurations, which should be representative for the system. The method is more efficient for atomic or simple molecular systems than for complex (macro)molecular systems, since a random displacement in the latter case will generally lead to such distortions of a molecule that the energy of a new configuration will usually be very high. Metropolis Monte Carlo methods have been applied to the conformational analysis of oligosaccharides with the aim of deriving ensemble average parameters,[33,34] or exploring the multiple conformations adopted by a complex polysaccharide such as xyloglucan.[35]

In molecular dynamics simulations an ensemble of configurations is generated by applying motion laws to the atoms of the molecular system. The two major simulation techniques are molecular dynamics, in which Newton's equations of motion are integrated over time, and stochastic dynamics, in which the Langevin equation for Brownian motion is integrated over time. Several algorithms have been developed for molecular dynamics simulations. Such simulations follow a system for a limited time. Physically observed properties are computed as the appropriate time averages through the collective behavior of individual molecules. For the results to be meaningful, the simulations must be sufficiently long so that the important motions are statistically well sampled. Experimentally accessible spectroscopic and thermodynamic quantities can be computed, compared, and related to microscopic interactions. Such modeling techniques have been applied to a wide range of oligosaccharides. The structural flexibility of these molecules has been confirmed. Highly branched oligosaccharides have also been investigated by means of molecular dynamics. The results obtained with these have been interpreted as demonstrating a more rigid behavior than that found for linear oligosaccharides.

It should be noted that molecular dynamics is severely

limited by the available computer power. With presently available computers, it is feasible to perform a simulation with several thousand explicit atoms for a total time of up to about a few nanoseconds. To explore adequately the conformational space it is necessary to perform many such simulations. In addition, it may be possible that carbohydrate molecules undergo dynamic events on longer time scales. These motions cannot be investigated with standard molecular dynamics techniques.

It is important to recognize that most quantum mechanical and molecular mechanical procedures are designed to treat molecules in the isolated state. Omission of the effect of the environment from the calculation results in a neglect of the fraction of the energy contribution that arises from these interactions. For example, a carbohydrate in an aqueous or crystalline environment will usually form hydrogen bonds only to neighboring molecules, while the simulation of the molecule *in vacuo* is dominated by conformations with energetically favorable intramolecular hydrogen bonds.[17]

Several different approaches have been proposed to treat solvation effects.[36] In the simplest approach, the effect of the solvent is achieved by increasing the dielectric constant for calculations of electrostatic interactions or by the use of a distance-dependent dielectric constant. Unfortunately, this affects all electrostatics. An alternative approach is to treat the solvent as a dielectric continuum. The conformational free energy of a given conformer in a particular solvent may be described as arising from the contribution of the energy of the isolated state and the solvation free energy.[37] A computationally efficient method is the use of the CHEAT95 force field, which is parametrized in such a way that the simulation of isolated carbohydrates mimicks the behavior of the molecule in aqueous solution.[17]

At present, the best approach is the inclusion of the environment in the simulation, namely a molecular dynamics simulation with explicit water molecules or other surrounding molecules. By applying periodic boundary conditions a true, but still small, system is simulated. Of course, this is time consuming for an oligosaccharide in water.

Probing potential energy surfaces

It is the objective of the present section to illustrate that potential energy surfaces can be put to a demanding test. Indeed, many observable properties of oligosaccharides can be calculated and have been shown to be sensitive to the details of the conformational energy surface.

More than 3 600 crystal structure determinations of carbohydrates are listed now in the Cambridge Crystallographic Data Base. X-Ray analysis gives the best data for the conformation of a carbohydrate. Precise atomic coordinates are provided, along with an explicitly defined environment. Although the crystalline state is often dismissed as irrelevant to biological processes, comparisons with crystal structures are among the most precise tests of modeling available for carbohydrate molecules, provided that packing forces are taken into account. By molecular dynamics simulations of crystal structures both force fields and methods can be validated.[38]

A more common method in which crystallographic data are used to test computer simulations is the superposition of conformations found in crystal structures on a calculated potential energy map. For example, in Color Plate 1 the potential energy surface of cellobiose as a function of the glycosidic torsion angles is given. A search in the crystallographic database for molecules with a link similar to that in sucrose results in a number of conformations, which can be plotted on the calculated surface.

It should be kept in mind that the conformations found in crystals can be influenced by packing effects, so that they differ from the preferred conformation(s) in aqueous solution and *in vacuo*. An interesting example of the problems that can arise with *in vacuo* calculations is given by sucrose. In Color Plate 2 maps calculated with MM3[39] and CHEAT95[17] are given, together with a population density map calculated from a molecular dynamics simulation in water.[40] The MM3 map predicts a high potential energy (5.5 kcal/mol higher than the global minimum) for the conformation of the sucrose link found in raffinose. The simulation in water shows that this is an artifact of the MM3 force field; the calculations with the CHEAT95 force field perform much better in this respect. The lowest energy conformation is stabilized by an intramolecular hydrogen bond, which cannot be formed in the raffinose conformation. Nevertheless, this conformation appears to be stabilized by surrounding molecules. An extensive study of the energy contributions in this glycosidic link showed that the problem was not the overlapping anomeric sequence, as was suggested,[41] but a high barrier for one of the torsions.[17]

In solution, the method of choice to study the three-dimensional structure of saccharides is nuclear magnetic resonance (NMR), through the parameters represented by chemical shifts, coupling constants, nuclear Overhauser effects (NOEs), and also relaxation time measurements. While the conformational dependence of the carbon chemical shifts is far from understood, coupling constants can be used to evaluate the magnitude of the torsion angles, NOE measurements can provide estimations of distances between protons located in rather close proximity. In addition, relaxation time measurements give information on the mobility and the behavior of molecules in solution.

A major difficulty in the determination of the conformation of an oligosaccharide from NMR data is the flexibility of the carbohydrates, especially of the glycosidic links. When multiple conformations are present in solution, NMR data will represent a time-averaged conformation. Since the geometric parameters are usually related in a nonlinear way to the experimental data, these data can be misleading. Consider, for example, an oligosaccharide in solution that occupies two distinct conformations, one with a relatively short distance between the protons at both sides of the link, and one with a rather large distance (which is preponderant). The measured NOE is an average value, so the NOE could easily lead the interpreter to a single nonexistant conformation. Even when it is known that two conformations are present, errors can easily be made since the preponderant conformation will produce only a small contribution to the resulting NOE.

There are not many experimental means, other than NMR, suitable for probing carbohydrate conformations and evaluating calculated potential energy surfaces. However, the optical activity of saccharides depends on their chemical

composition, configuration, and conformation. Models have been developed and applied to disaccharides in aqueous solution.[42,43] These studies result in the location of preferred regions in the configurational space, rather than in the location of some well-defined points. The technique for oligosaccharides is not widely applicable, but useful to complement NMR methods. Sucrose,[42] cellobiose, and maltose[43] are among the numerous disaccharides that have been investigated so far. It has been reported that the optical rotation observed in the agarose gels can be satisfactorily accounted for in terms of associated double-helix chain conformations rather than an extended simple helix.[44]

THE DISORDERED STATE OF POLYSACCHARIDES

The polysaccharide chains in solution tend to adopt a more or less coiled structure. Such a dissolved random coil would fluctuate between local and overall conformations. Polysaccharides are able to assume an enormous variety of spatial arrangements around the glycosidic linkages because these molecules have extensive conformational freedom. Theoretical polysaccharide models are based on studies of the relative abundance of the various conformations, in conjunction with the statistical theory of polymer chain configuration.[45] Possible interactions between residues of the polysaccharide chain that are not nearest neighbors in the primary sequence of the polymer are ignored. The range of conformations of polymer molecules is reflected by a Monte Carlo sample. The observable properties of dissolved polysaccharides are averaged over the entire range of conformations accessible to the chain, and they may be determined from conformational states derived from the potential energy surfaces of the consecutive disaccharide fragments. This approach yields properties corresponding to the equilibrium state of the chain. Results refer to a model for an unperturbed chain that ignores the consequences of the long-range excluded volume effect, because only nearest neighbor interactions are accounted for in the computation of the φ, ψ surfaces.

Given a sufficient Monte Carlo sample of unperturbed polysaccharide chains, it is possible to assess average properties of the polymer in question simply by computing arithmetic averages over the chains of the sample. For example, the mean square end-to-end distance, the mean square radius of gyration, the average persistence length, dipole moment, etc., are all average geometric properties readily computed from a knowledge of the coordinates of the atoms or atomic groups making up the Monte Carlo sample.[46] Models of native polysaccharides, refined to various extent, have been presented.[47–49] Monte Carlo methods have been applied to exploring the multiple conformations occurring in a complex polysaccharide such as xyloglucan.[35]

Models of polymer chain extension were first used to compare the effect of the glycosidic linkage geometry of simple polysaccharide chains, e.g., cellulose and amylose.[50] Both polymers are $1 \rightarrow 4$-linked glucans, the only difference being the anomeric configuration on the C1 atom of the monomeric unit: α and β for amylose and cellulose, respectively. The calculated data show a remarkable pseudo-helical chain trajectory of the amylosic chains, and the characteristic ratio of 5 denotes a moderately compact chain configuration. This behavior is the direct consequence of the glycosidic bond geometry because changing this geometry from the α to the β configuration has a dramatic effect on the character of the chain trajectory. Relative to amylose, the cellulosic characteristic ratio of 100 is predicted to increase by 20-fold. This reflects the extended character of the cellulosic chains. Investigation of the effect of solvent on those two representative polysaccharides has also been attempted.[51] It was found, in good concordance with the experimentally observed solvent dependence, that significant changes in the unperturbed chain dimensions occur. The characteristic ratio for amylose is larger in water than *in vacuo,* whereas for cellulose it is smaller. Here again, the incorporation of solvation remains a difficult attempt, and these conclusions should be considered with caution.

The following example provides an illustration of the an application of this procedure to the characterization of the solution behavior of pectic substances. Pectins are a family of polysaccharides that constitute a large portion of the cell wall of many higher plants, where they influence growth, development, and senescence. They are extensively used as gel formers and thickening agents by the food industry. The basic backbone of pectin polysaccharide is formed by $(1 \rightarrow 4)$-linked α-D-galacturopyranosyl residues, either free or in ester form. These homogalacturonan sequences may be interspersed at intervals with β-L-rhamnopyranosyl residues carrying the major part of neutral sugar side chains, mainly arabinans, galactans, or arabinogalactans. Three different molecular modeling studies[52–54] were carried out on theoretical polysaccharide chain models of the linear part of pectin molecules. All of them illustrate the important extended and stiff character of homogalacturonan sequences. As a result of different force fields and strategies used in the calculation of the potential energy surfaces of the parent disaccharides, the reported chain dimensions differ considerably. There is an excellent consistency between the calculated characteristic ratios of 57[52] and 47[54] that have been computed from relaxed MM3 and CHARMm maps, respectively. The characteristic ratio established from potential energy surfaces calculated with a "rigid residue approach"[53] is between 150 and 253, depending on the value of the glycosidic bond angle. The solvent polarity as well as the ionic state of the galacturonic acid residues affect the conformational behavior of the glycosidic linkage; these two lead to an increase in the unperturbed limiting chain dimensions. The insertion of rhamnose residues in the primary sequence[52] does not seriously disrupt the overall chain propagation, as shown by a small decrease (8% with 25% rhamnose) in the characteristic ratio. This result is in contrast with another study that concluded that the insertion of rhamnose units decreased the characteristic ratio by about 50%.[53] This discrepancy may be due to placement of the rhamnose units, the occurrence of which follows a defined pattern derived from experimental investigations or in a Bernoullian distribution.

In the physicochemical analysis of pectin chains, the characterizations of both their size and shape were studied by many techniques such as osmometry measurement, wide- and low-angle laser light scattering coupled or not with size-exclusion chromatography, and viscometry, low-

speed sedimentation equilibrium, and small-angle neutron scattering.[52] The heterogeneity of the primary structure along with the presence of aggregates could affect the molecular state in solution and therefore hamper the accurate determination of molecular weight. This is why the literature on solution features of pectin is full of conflicting reports. Depending on the authors and the measuring techniques used, pectin molecules have been reported to behave as rigid-rod particles or as coils of variable stiffness. Some authors have reported that the stiffness of chains could depend on the degree of esterification. The neutral sugar content has been reported to affect the conformation and it has been suggested that fractions rich in neutral sugars are responsible for the high molecular weight found in some cases. A quantitative comparison between predicted shape and measured shape is complicated because the heterogeneity of the chemical structure makes a straightforward comparison between experimental results difficult. The joint use of small-angle neutron scattering, viscosimetric, and molecular modeling studies[52] on a series of samples having well-characterized degrees of methylation and rhamnose content provided a consistent characterization of the configurational features of pectins. More elaborate characterizations are still awaiting the availability of tailor-made samples of pectins.

THE ORDERED STATE OF POLYSACCHARIDES

Like other polymers, polysaccharides form helical structures. The helix symmetry can be denoted by u_v, which means that there are u repeat units in v turns of the helix.[55] The helical arrangement can also be described in terms of a set of helical parameters (n, h); n is the number of repeating units per turn of the helix, and h is the translation along the helix axis. The chirality of the helix is described by the sign of h: a negative value designates a left-handed helix.[56]

An X-ray diffraction pattern from a highly crystalline sample gives through the positions of the reflections information about the unit cell dimensions and the space group. The atomic positions can be deduced from the intensities of the reflections. It is usually impossible to obtain highly crystalline samples of polysaccharides, which limits the quality of the diffraction pattern. The diffracted intensities are restricted to layer lines. The layer line spacings give information related to the axial advance h of the molecule. The meridian (vertical axis) of the diffraction pattern only has intensities on layer lines that are a multiple of u, so this gives information about the helical symmetry.[55]

With the information from the X-ray diffraction pattern and molecular modeling a number of possible structures can usually be calculated. These models include left- and right-handed helices, and single and coaxial multiple helices. In the case of double helices, the strands can be either parallel or antiparallel. At present, for some polysaccharides the number of strands in the helix is well known, but for others this is still under debate. The POLYS program,[3] already mentioned in the introduction, has been extended for the purpose of multiple helices. A single strand is positioned in such a way that the helix axis coincides with the z axis of a coordinate system, after which a rotation or screw operation is applied. In this way double or triple helices are easily generated.

Examples of better known single-helical polysaccharide structures include cellulose, 1 → 4-linked β-D-glucose units. Different polymorphs exist, and although the crystal structures are not yet known with atomic accuracy, it is known that cellulose in the predominant forms are single-stranded twofold helices.[57] Amylose exists as 1 → 4-linked α-D-glucose units and it crystallizes as double helices. Here, too, different forms are known, which differ mainly in the water content of the unit cells.[58,59] In both cases the amylose molecules form left-handed, sixfold helices of 2.1 nm (n = 6 and h = −0.35 nm). In the double helix the individual strands are oriented in parallel. A triple helix is formed by β(1 → 3)-glucan.[60] In this polysaccharide β-D-glucose units are linked 1 → 3. An illustration of the some of the examples given here is shown in Color Plate 3.

The combination of molecular modeling and X-ray data for a linear homopolysaccharide is illustrated by amylose. According to the experimental data, the helices have sixfold symmetry repeating every 2.1 nm. In Figure 7 the potential energy surface calculated as a function of the glycosidic torsion angles of maltose (which is the repeating unit of amylose) is shown. The helical parameters of the amylosic strand generated for each combination of these torsion angles are superimposed on the surface. The conformations that are in agreement with the experimental data are found at the intersections of the contours n = 6 and h = 0.35 nm (for a right-handed chirality) or h = −0.35 nm (for a left-handed chirality). The conformations generating a left-

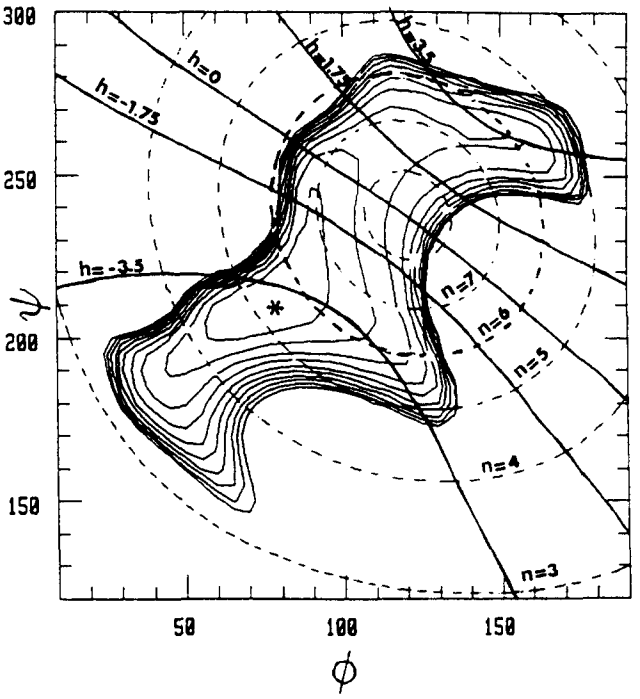

*Figure 7. Iso-*n *and iso-*h *(in angstroms) contours superimposed on the potential energy surface of maltose. The isoenergy contours are drawn by interpolation of 1 kcal/mol with respect to the calculated minimum (*). The* h = 0 *contour divides the map into right-handed (*h > 0*) and left-handed (*h < 0*) regions.*

handed helix are near the calculated energy minimum, whereas the alternative conformations appear to be unstable. Therefore, a left-handed model seems to be appropriate.[56]

For some polysaccharides diffraction patterns have been measured that are not easily interpreted. Agarose, for example, is a linear polysaccharide consisting of alternating 3-linked β-D-galactose and 4-linked 3,6-anhydro-α-L-galactose units (see Fig. 8). Three diffraction patterns have been reported that were interpreted as single helices.[61] A fourth diffraction pattern, however, is still under debate. It might result from a double helix in which the individual strands are shifted half the pitch with respect to each other, but the validity of this structure is not widely accepted.[61,62]

Carrageenans form another class of polysaccharides, the structure of which is not yet completely understood. Different forms exist, three of which are also schematically given in Figure 8. It is generally accepted that ι-carrageenan forms double helices in the ordered state. Although there is a great similarity between ι- and κ-carrageenan, the structure of κ-carrageenan is still under debate. Both single- and double-helical models are considered. For λ-carrageenan it is rather unlikely that a double helix is formed.[63]

The next step in the determination of the structure of polysaccharides in the ordered state is the investigation of the interaction of different helices: the packing. In a crystal structure of a polymer the chains can be packed parallel or antiparallel. In cellulose this packing is still one of the remaining questions for the different polymorphs; it is almost impossible to distinguish the two possibilities on the basis

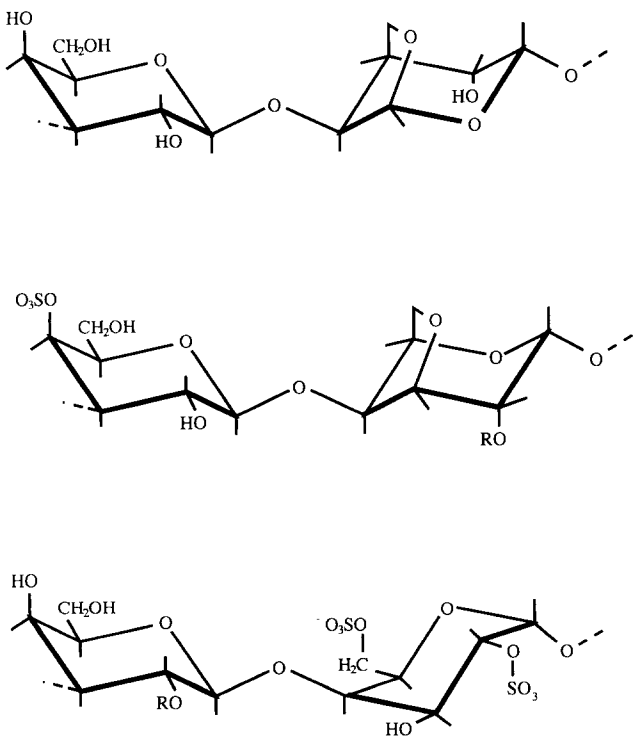

Figure 8. Schematic representation of the primary structure of agarose (top), ι- and κ-carrageenan (middle; in the ι-form R = SO_3^-, and in the κ-form R = H), and λ-carrageenan (bottom; R = H or SO_3^-).

of the few reflections that can be measured by X-ray diffraction.

Molecular modeling becomes more and more a powerful tool in the study of the packing of polysaccharides. Models can be built and energies can be calculated. A method of calculating chain–chain interactions was published by Pérez et al.[56] Given a rigid model of an isolated (single or multiple) helix, its interaction with a second helix is calculated at varied helix axis translations and mutual rotational orientations while keeping the helices in van der Waals contact. For each setting, the energy is calculated, as is the interchain distance. No energy minimizations are performed (which is in such a study hardly applicable: it is time consuming and leads only to the nearest minimum), but the energy and interchain distance are calculated on a three-dimensional grid. This is illustrated in Figure 9. For efficient packing not only is low energy of importance. Coupled values of the rotations of the individual chains are perhaps even more important (they indicate the presence of a rotation axis). When the translation between the helix axis is related to the repeat distance of the helix, the rotation axis might even be a screw axis.

This procedure has been applied successfully for amylose. Amylose is found as A type (in cereal starches) or B type (in tuber starches). In both types the same double helix is found. The chain-pairing procedure was applied to the left-handed double-helical model of this helix, and significant low-energy chain pairings were selected, both parallel and antiparallel. The most promising was a parallel packing of two helices, wherein the mutual rotations appeared to be coupled and the translation was half the fiber repeat. The interchain distance was calculated to be 1.077 nm. This is in excellent agreement with crystallographic studies,[58,59] except that the interchain distance is slightly larger. In A-type starch it is 1.062 nm and in B-type starch it is 1.068 nm. Another, looser type of interaction is seen in the crystal structure of A-type starch; this arrangement corresponds to a calculated secondary minimum that is among the low-energy chain pairings.[56]

The procedure has been applied to several other polysaccharides[64]; an interesting example of larger complexity is agarose. As mentioned above, different diffraction patterns exist. First, an extensive search was carried out to find (single and multiple, left- and right-handed) helices of low energy that were in agreement with the observed layer line spacings and helical symmetry. This is much more difficult than for a homopolymer since the helical parameters are now mainly depending on not two, but four, glycosidic torsion angles. Thus, the method illustrated in Figure 7 cannot be applied. For the models obtained the chain-pairing procedure was applied. This resulted in parallel and antiparallel packings of a left- and a right-handed helix for each diffraction pattern. For one of the diffraction patterns the energies of these four models are given in Table 1. It was reported that this pattern resulted from a crystal structure in a trigonal crystal system.[61] With this knowledge a crystal structure was built with the left-handed helical model in an antiparallel orientation. Extra symmetry appeared to be present in our model, and the space group could be assigned.[65] The asymmetric unit of the cell contains only one agarose repeat unit. Furthermore, the cell has about 30% solvent-accessible space, which is 3 or 4 water molecules

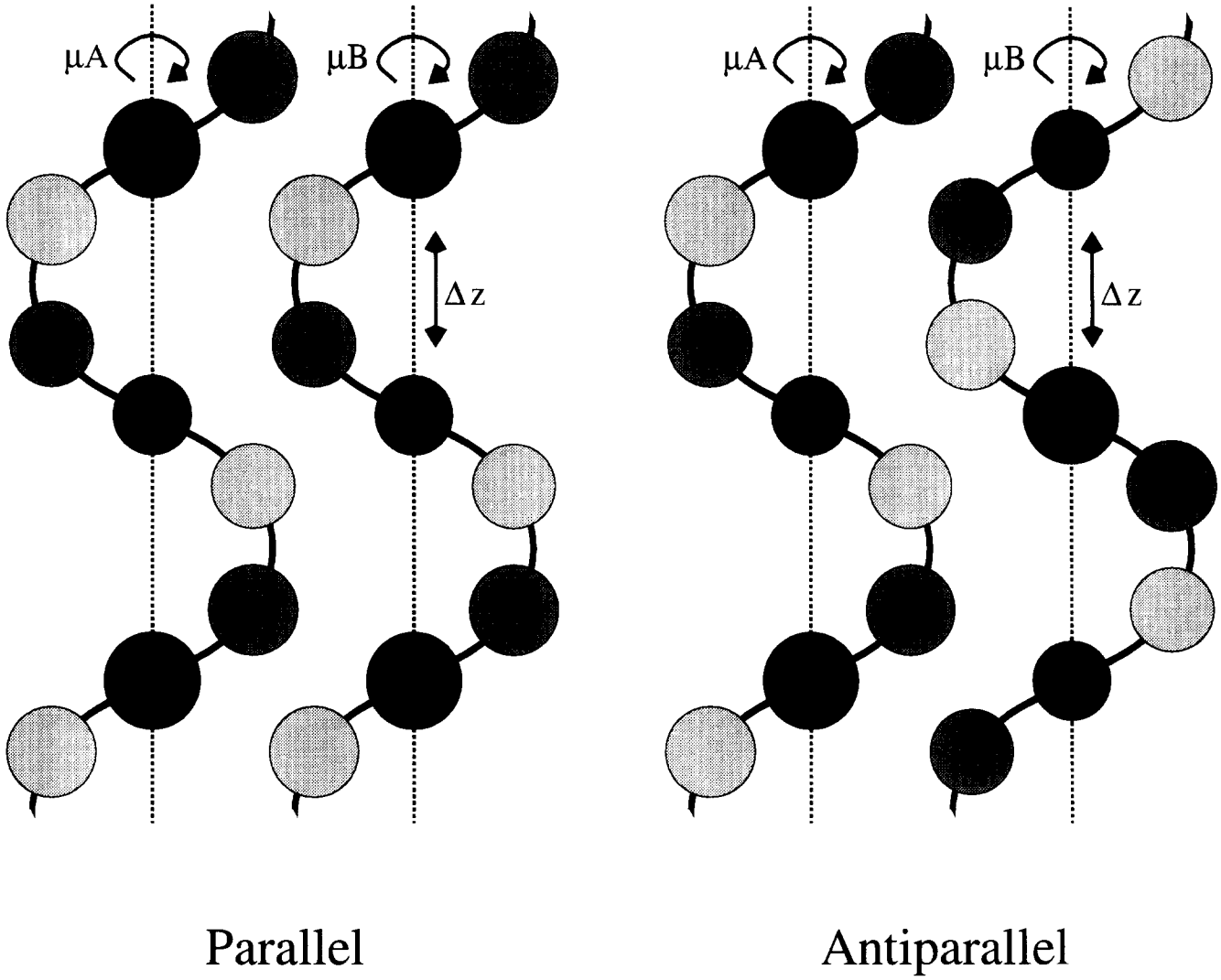

Figure 9. Schematic representation of the chain-pairing procedure.

per agarose repeat. An impression of the structure is shown in Figure 10. The verification of this model will come through the comparison of the calculated and observed diffraction patterns, which is being done at present. A preliminary test is the comparison of the unit cell dimensions, and here the agreement is excellent. The length of the c axis is related to the layer line spacing, which we had used in building the helix, but the length of the a axis (in a trigonal crystal system equal to the b axis) follows directly from our calculated interchain distance. This axis was experimentally determined to be 1.024 (0.01) nm; in our calculations it is 1.04 (0.04) nm.

With the increasing speed of computers, more complex polysaccharide studies will become possible. Xanthan, for example, and other charged polysaccharides shall soon be within the possibilities of computer modeling.

Table 1. Relative energies[a] of the different packings of the different models

		E_{rel}		
Chirality	Packing	Helix	Packing	Total
Right handed	Parallel	2.0	0.9	2.9
	Antiparallel	2.0	0.9	2.9
Left handed	Parallel	0.0	1.4	1.4
	Antiparallel	0.0	0.0	0.0

[a]In kilocalories per mole, per repeating unit.

CONCLUSIONS

It was the aim of the present work to describe the most recent tools that have been developed for modeling the three-dimensional features of polysaccharides and carbohydrate polymers. It was shown that the primary structures of polysaccharides vary in composition, sequence, molecular weight, anomeric configuration, linkage position, and charge density. As a consequence, an almost infinite array of chemical structures and conformations can be generated for polysaccharides. Additional variability also arises from environmental changes such as ionic strength and degree of hydration. To cope with such complex macromolecules, the

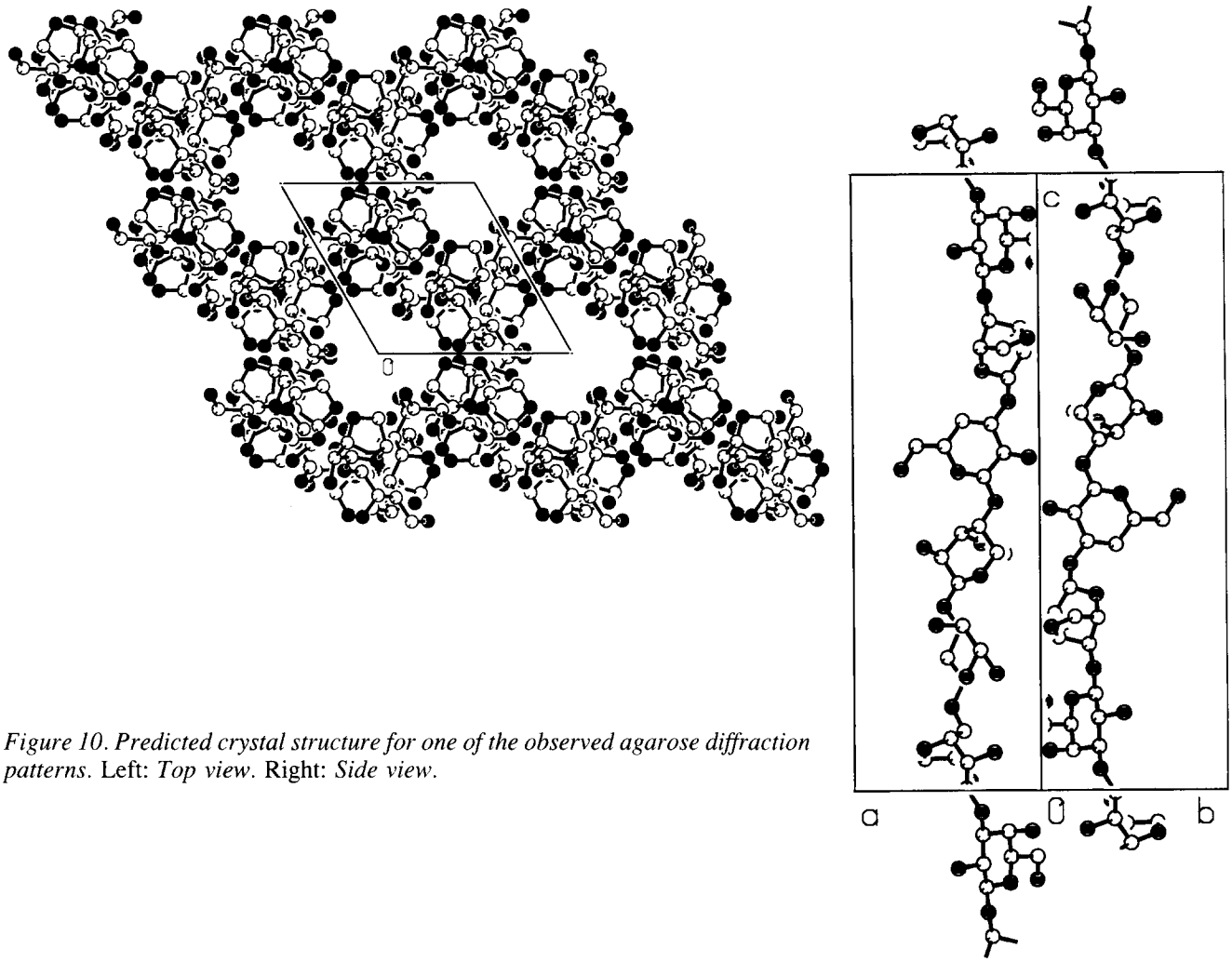

Figure 10. Predicted crystal structure for one of the observed agarose diffraction patterns. Left: *Top view.* Right: *Side view.*

integration of molecular modeling into the biophysical analysis of polysaccharides is required. Steady progress has been made, which allows a description of the conformations of flexible rings as in the case of five-membered rings, and a thorough description of the conformational space that is available for a disaccharide, either *in vacuo* or in an aqueous solution. Several force fields along with a dedicated parametrization are available. They are in principle capable of dealing with the specific stereoelectronic effects such as the anomeric and exoanomeric effects, and can cope with the enhanced complexity arising from the hydrogen-bonding capacities of the carbohydrates. It has been clearly stated that these force fields may not be used straightforwardly to probe in an unequivocal fashion. In particular the use of NMR can be envisaged only when the question of internal motion is fully understood.

One of the prerequisites for extending the modeling from disaccharides to polysaccharides implies that the rotations of a particular glycosidic linkage can be considered, under some conditions, to be independent of the nearest neighbors. Consequently, the conformations of a polysaccharide can be described conveniently by the glycosidic torsion angles from consecutive dimeric fragments. In solution, polysaccharide chains tend to adopt less ordered structures (random coil) that fluctuate among different local and overall conformations. Proper modeling can provide insight into the dimensions of these random coils and such descriptors as persistence lengths or characteristic ratio can be readily assessed. Interesting, the occurrence of local helical regions may be detected from such simulations. It has been observed that these locally ordered conformations precede the regular local helical arrangements that are found in the solid state. With the help of such descriptors as helical parameters, the ordered state of polysaccharide strands can be readily characterized. The generation of double or triple helices is then attempted in order to investigate the occurrence of such multistranded arrangements that may be energetically stable. The final step in the determination of the structure of polysaccharides in the ordered state is the investigation of the interactions of different helices, which may lead to either the best arrangement(s) between two polymeric chains or to the prediction of the dimensions and the symmetry of a three-dimensional lattice.

The characterization of secondary, tertiary, and quaternary structural levels of organizations of polysaccharides is a prerequisite for understanding the molecular basis of their properties and/or functions and for controlling and manipulating these properties and/or functions through rational changes in molecular fine structure. Therefore, all the different steps that have been described above have been integrated in a general computer program that enables the prediction of the three-dimensional arrangements of poly-

saccharide chains from the knowledge of the primary structure. Figure 11 is a synopsis of such a program. It incorporates some features such as the prediction of low-energy arrangements between different species of polymer chains that could be indicative of the formulation of blends involving polysaccharides. Further work is required for establishing quantitative structure–properties relationships in the area of polysaccharides, with a particular emphasis on the gel-forming and viscoelastic properties. The wealth of information available in present databases can certainly be rationalized. Some of the tools that have been developed should allow automatic searches for meaningful correlations between structures and functions through exploratory data analysis. Structure–function or structure–property correlations could be then used to model changes arising from structural alterations. This would open the field of polysaccharide engineering.

Other challenges will be in modeling supramolecular structure embedding complex assemblies of polysaccharides. Only realistic modeling of the microfibrillar structure of cellulose chains will provide insight in the understanding of the unique rheological properties displayed by such native arrangements, which are readily available in renewable raw materials. Other major polysaccharide architectures, such as that of the starch granule, would also be investigated. Here, the challenge lies in putting together such elementary pieces as double helices and branching points located at key points of amylose chains to construct the elementary clusters containing more than several 100 000 atoms. Only such constructions will lead to the understanding of how the ratio of amylose over amylopectin controls the establishment of unique types of granule. In turn, the control of such a ratio, via the routes of molecular biology, will create ad hoc-type granules, for which a significant range of properties can be foreseen. In the area of plant cell walls, the situation is even more intricate since several polysaccharide moieties in the form of microfibrils and various complex polysaccharides interact via proteins. These are only selected examples, which indicate some important structures and architectures as challenging candidates to model within the decade to come.

ACKNOWLEDGMENTS

This work was supported by the Institut National de la Recherche Agronomique to S.P.; M.K. benefited from a grant in the framework of CARENET network as part of the

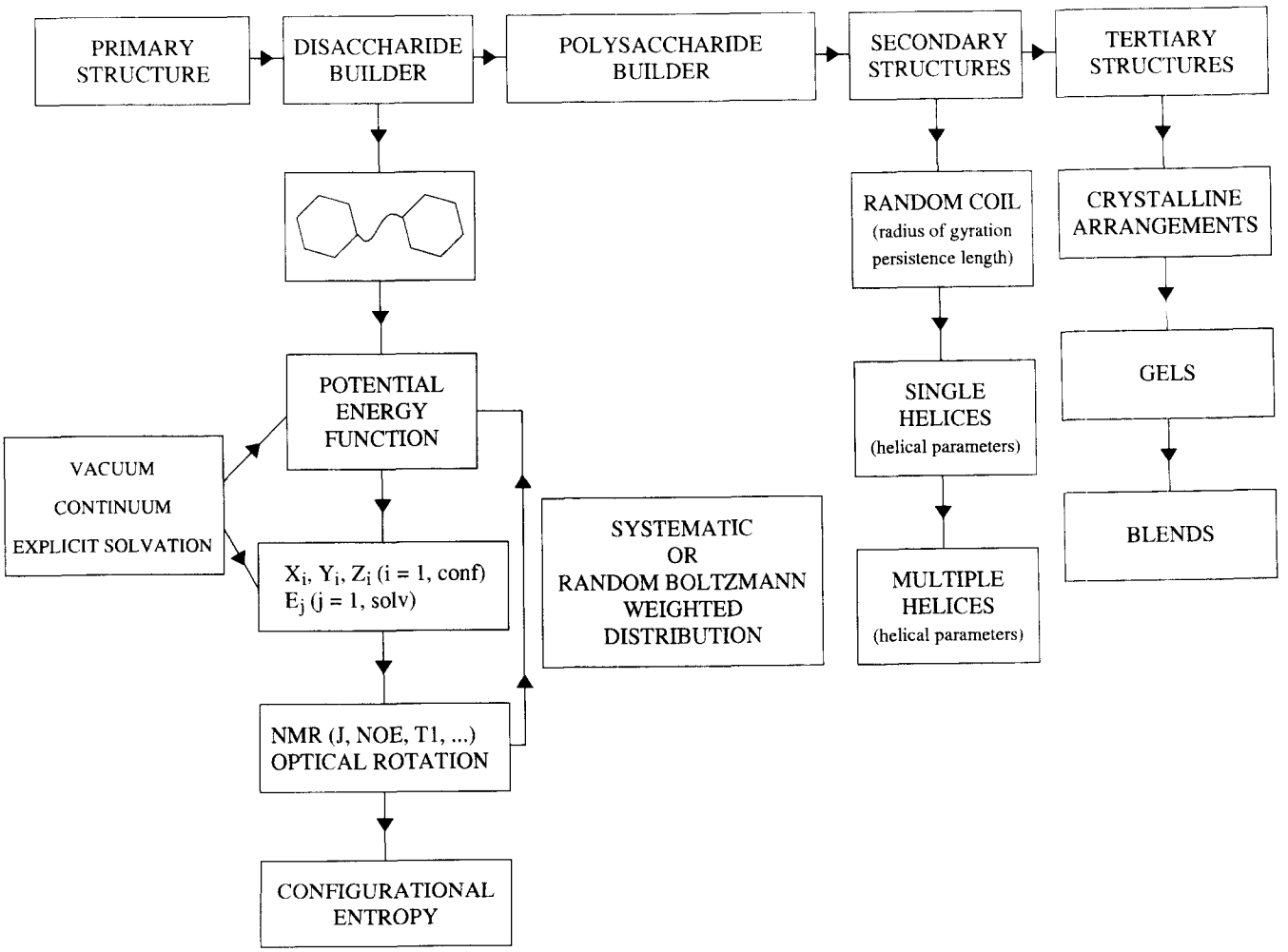

Figure 11. Flow chart describing the steps leading from the knowledge of the primary structure to the different structural levels of increasing complexity.

Training and Mobility of Research EC program. K.M. was supported by the Centre National de la Recherche Scientifique, and S.B.E. was supported by a grant in the framework of the program Conception Macromolécules Assistée par Ordinateur (ORGANIBIO).

REFERENCES

1 Dwek, R.A. Glycobiology: Toward understanding the function of sugars. *Chem. Rev.* 1996, **96,** 683–720

2 Laine, R. A calculation of all possible oligosaccharide isomers both branch and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: The isomer barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* 1994, **4,** 759–767

3 Engelsen, S.B., Cros, S., Mackie, W., and Pérez, S. A molecular builder for carbohydrates: Application to polysaccharides and complex carbohydrates. *Biopolymers* 1996, **39,** 417–433

4 Cremer, D. and Pople, J.A. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* 1975, **97,** 1354–1358

5 Levitt, M. and Warshel, A. Extreme conformational flexibility of the furanose ring in DNA and RNA. *J. Am. Chem. Soc.* 1978, **100,** 2607–2613

6 French, A.D. and Tran, V. Analysis of fructofuranose conformations by molecular mechanics. *Biopolymers* 1990, **29,** 1599–1611

7 Cros, S., Hervé du Penhoat, C., Pérez, S., and Imberty, A. Modelling of arabinofuranose and arabinan. 1. Conformational flexibility of the arabinofuranose ring. *Carbohydr. Res.* 1993, **248,** 81–93

8 Pérez, S. and Delage, M.-M. A data base of three-dimensional structures of monosaccharides from molecular-mechanics calculations. *Carbohydr. Res.* 1991, **212,** 253–259

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

10 van Gunsteren, W.F. *GROMOS: Groningen Molecular Simulation Program Package.* University of Groningen, Groningen, The Netherlands, 1987

11 Allinger, N.L. Conformational analysis. 130 MM2. A hydrocarbon force field utilizing V_1 and V_2 torsional terms. *J. Am. Chem. Soc.* 1977, **99,** 8127–8134

12 Allinger, N.L., Yuh, Y.H., and Lii, J.H. Molecular Mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* 1989, **111,** 8551–8566

13 Tvaroska, I. and Pérez, S. Conformational-energy calculations for oligosaccharides: A comparison of methods and a strategy of calculation. *Carbohydr. Res.* 1986, **149,** 389–410

14 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculation. *J. Comput. Chem.* 1983, **4,** 187–217

15 Ha, S.N., Giammona, A., Field, M., and Brady, J.W. A revised potential-energy surface for molecular mechanics studies of carbohydrates. *Carbohydr. Res.* 1988, **180,** 207–221

16 Grootenhuis, P.D.J. and Haasnoot, C.A.G. A CHARMm based force field for carbohydrates using the CHEAT approach: Carbohydrate Hydroxyl groups represented by Extended Atoms. *Mol. Simul.* 1993, **10,** 75–95

17 Kouwijzer, M.L.C.E. and Grootenhuis, P.D.J. Parametrization and application of CHEAT95, an extended atom force field for hydrated (oligo)saccharides. *J. Phys. Chem.* 1995, **99,** 13426–13436

18 Weiner, S.J., Kollman, P.A., Nguyen, D.T., and Case, D.A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* 1986, **7,** 230–252

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

20 Glennon, T.M., Zheng, Y.-J., Le Grand, S.M., Shutzberg, B.A., and Merz, K.M., Jr. *J. Comput. Chem.* 1994, **15,** 1019–1040

21 Woods, R.J., Dwek, R.A., Edge, C.J., and Fraser-Reis, B. Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. *J. Phys. Chem.* 1995, **99,** 3832–3846

22 Lifson, S. and Warshel, A. Consistent force field for calculations of conformation, vibrational spectra, and enthalpies of cycloalkane and *n*-alkane molecules. *J. Chem. Phys.* 1968, **49,** 5116–5129

23 Kildeby, K., Melberg, S., and Rasmussen, K. Conformations of α- and β-D-glucopyranose from and empirical force field. *Acta Chem. Scand.* 1977, **A31,** 1–13

24 Engelsen, S.B. and Rasmussen, K. Conformations of disaccharides by empirical force field calculations. V. Conformational maps of β-gentiobiose in an optimized consistent force field. *Int. J. Biol. Macromol.* 1993, **15,** 56–62

25 Clark, M., Cramer, R.D., III, and Van Opdenbosch, N. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* 1989, **10,** 982–1012

26 Imberty, A., Hardman, K.D., Carver, J.P., and Pérez, S. Molecular modelling of protein–carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A. *Glycobiology* 1991, **1,** 631–642

27 Mayo, S.L., Olafson, B.D., and Goddard, W.A., III. DREIDING: A generic force field for molecular simulations. *J. Phys. Chem.* 1990, **94,** 8897–8909

28 Halgren, T.A. Merck Molecular Force Field. *J. Comput. Chem.* 1996, **17,** 490–641

29 French, A.D. Rigid- and relaxed-residue conformational analyses of cellobiose using the computer program MM2. *Biopolymers* 1988, **27,** 1519–1525

30 Kozár, T., Petrák, F., and Gálová, Z. RAMM—a new procedure for theoretical conformational analysis of carbohydrates. *Carbohydr. Res.* 1990, **204,** 27–36

31 Hooft, R.W.W., Kanters, J.A., and Kroon, J. Implementation and use of the method of prudent ascent in conformational analysis using molecular mechanics. *J. Comput. Chem.* 1991, **12,** 943–947

32 Koca, J., Pérez, S., and Imberty, A. Conformational analysis and flexibility of carbohydrates using the CICADA approach with MM3. *J. Comput. Chem.* 1995, **16,** 296–310

33 Peters, T., Meyer, B., Stuike-Prill, R., Somorjai, R., and

Brisson, J.-R. A Monte Carlo method for conformational analysis of saccharides. *Carbohydr. Res.* 1993, **238,** 49–73
34 Weimar, T., Meyer, B., and Peters, T. Conformational analysis of α-D-Fuc-(1 → 4)-β-D-GlcNac-OMe. One-dimensional transient NOE experiments and Metropolis Monte Carlo simulations. *J. Biomol. NMR* 1993, **3,** 399–414
35 Levy, S., York, W.S., Stuike-Prill, R., Meyer, B., and Staehelin, L.A. Simulations of the static and dynamic conformation of xyloglucan. The role of fucosylated side chain in surface specific side chain folding. *Plant J.* 1991, **1,** 195–215
36 Orozco, M., Alhambra, C., Barril, X., Lopez, J.M., Busquets, M.A., and Luque, F.J. Theoretical methods for the representation of solvent. *J. Mol. Modeling* 1996, **1,** 1–15
37 Tvaroska, I. and Kozar, T. Theoretical studies on the conformation of saccharides. 3. Conformational properties of the glycosidic linkage in solution and their relation to the anomeric and exoanomeric effects. *J. Am. Chem. Soc.* 1980, **102,** 6929–6936
38 Kouwijzer, M.L.C.E., van Eijck, B.P., Kroes, S.J., and Kroon, J. Comparison of two force fields by molecular dynamics simulations of glucose crystals: Effect of using Ewald sums. *J. Comput. Chem.* 1993, **14,** 1281–1289
39 French, A.D. and Dowd, M.K. Exploration of disaccharide conformations by molecular mechanics. *J. Mol. Struct. (Theochem.)* 1993, **286,** 183–201
40 Engelsen, S.B., Hervé du Penhoat, C., and Pérez, S. Molecular relaxation of sucrose in aqueous solution: How a nanosecond molecular dynamics simulation helps to reconcile NMR data. *J. Phys. Chem.* 1995, **99,** 13334–13351
41 van Alsenoy, C., French, A.D., Cao, M., Newton, S.Q., and Schafer, L. Ab initio-MIA and molecular mechanics studies of the distorted sucrose linkage of raffinose. *J. Am. Chem. Soc.* 1994, **116,** 9590–9595
42 Stevens, E.S. and Duda, C.A. Solution conformation of sucrose from optical rotation. *J. Am. Chem. Soc.* 1991, **113,** 8622–8627
43 Stevens, E.S. and Sathyanarayana, B.K. Potential energy surfaces of cellobiose and maltose in aqueous solution: A new treatment of disaccharide optical rotation. *J. Am. Chem. Soc.* 1989, **111,** 4149–4154
44 Schafer, S.E. and Stevens, E.S. A reexamination of the double-helix model for agarose gels using optical rotation. *Biopolymers* 1995, **36,** 103–108
45 Brant, D.A. and Goebel, K.D. A general treatment of the configurational statistics of polysaccharides. *Macromolecules* 1975, **8,** 522–530
46 Flory, P.J. *Statistical Mechanics of Chain Molecules.* Wiley-Interscience, New York, 1969
47 Jordan, R.C., Brant, D.A., and Cesàro, A. A Monte Carlo study of the amylosic chain conformation. *Biopolymers* 1978, **17,** 2617–2632
48 Gagnaire, D., Pérez, S., and Tran, V. Configurational statistics of single chains of α-linked glucans. *Carbohydr. Polym.* 1982, **2,** 171–191
49 Brant, D.A. and Christ, M.D. Realistic conformational modeling of carbohydrates. In: *Computer Modeling of Carbohydrate Molecules* (A.D. French and J.W. Brady, eds.), ACS Symposium Series 430. American Chemical Society, Washington D.C., 1990, pp. 42–68
50 Burton, B.A. and Brant, D.A. Comparative flexibility, extension, and conformation of some simple polysaccharide chains. *Biopolymers* 1983, **22,** 1769–1792
51 Urbani, R. and Cesàro, A. Solvent effects on the unperturbed chain conformation of polysaccharides. *Polymer* 1991, **32,** 3013–3020
52 Cros, S., Garnier, C., Axelos, M.A.V., Imberty, A., and Pérez, S. Solution conformations of pectin polysaccharides: Determination of chain characteristics by small angle neutron scattering, viscometry, and molecular modeling. *Biopolymers* 1996, **39,** 339–352
53 Ruggiero, J.R., Urbani, R., and Cesàro, A. Conformational features of galacturonans. II. Configurational statistics of pectic polymers. *Int. J. Biol. Macromol.* 1995, **17,** 213–217
54 Boutherin, B., Mazeau, K., and Tvaroska, I. Conformational statistics of pectin substances in solution by a Metropolis Monte Carlo study. *Carbohydr. Polym.* 1996 (in press)
55 Millane, R.P. Polysaccharide structures: X-ray fiber diffraction studies. In: *Computer Modeling of Carbohydrate Molecules* (A.D. French and J.W. Brady, eds.), ACS Symposium Series 430. American Chemical Society, Washington D.C., 1990, pp. 315–331
56 Pérez, S., Imberty, A., and Scaringe, R.P. Modeling of interactions of polysaccharide chains. In: *Computer Modeling of Carbohydrate Molecules* (A.D. French and J.W. Brady, eds.), ACS Symposium Series 430. American Chemical Society, Washington D.C., 1990, pp. 281–299
57 Marchessault, R.H. and Sundararajan, P.R. Cellulose. In: *The Polysaccharides* (G.O. Aspinall, ed.), Vol 2. Academic Press, New York, 1983, pp. 11–95
58 Imberty, A., Chanzy, H., Pérez, S., Buléon, A., and Tran, V. New three-dimensional structures for A-type starch. *J. Mol. Biol.* 1988, **201,** 365–378
59 Imberty, A. and Pérez, S. A revisit to the three-dimensional structure of B-type starch. *Biopolymers* 1988, **27,** 1205–1221
60 Deslandes, Y., Marchessault, R.H., and Sarko, A. Triple-helical structure of (1 → 3)-β-D-glucan. *Macromolecules* 1980, **13,** 1466–1471
61 Foord, S.A. and Atkins, E.D.T. New X-ray diffraction results from agarose: Extended single helix structures and implications for gelations mechanism. *Biopolymers* 1989, **28,** 1345–1365
62 Arnott, S., Fulmer, A., Scott, W.E., Dea, I.C.M., Moorhouse, R., and Rees, D.A. The agarose double helix and its function in agarose gel structures. *J. Mol. Biol.* 1974, **90,** 269–284
63 Rees, D.A., Morris, E.R., Thom, D., and Madden, J.K. Shapes and interactions of carbohydrate chains. In: *The Polysaccharides* (G.O. Aspinall, ed.), Vol. 1. Academic Press, New York, 1982, pp. 195–290
64 Pérez, S. A priori crystal structure modeling of polymeric materials. In: *Electron Crystallography of Organic Molecules* (J.R. Fryer and D.L. Dorset, eds.). Kluwer Academic Publishers, The Netherlands, 1990, pp. 33–53
65 Spek, A.L. Platon, an integrated tool for the analysis of the results of a single crystal determination. *Acta Crystallogr.* 1990, **A46,** C34