

## Computational methods for studying oligo- and polysaccharide conformations

Igor Tvaroška

Institute of Chemistry, Center of Chemical Research of the  
Slovak Academy of Sciences, 842 38 Bratislava, Czechoslovakia.

**Abstract** - The paper documents the progress accomplished during the last two decades in a methodological development of the computational methods for oligo- and polysaccharides in solid state and in solution. Emphasis of analysis is placed on the quality of the method used for calculations of the structure, the energy of isolated molecule, and the influence of the environment as well as on the strategy of the investigation of the conformational properties. Finally, some examples are given to illustrate how computational methods can be used to estimate solution behaviour of oligosaccharides.

### INTRODUCTION

Saccharides are ubiquitous in nature. They occur in all life forms and, because of their unusual properties they present a unique source of chemicals. Saccharides of living organisms perform a very broad biological role (ref. 1-3) they function as structural materials, energy reserves, and adhesives; they appear to be essential in the process of infection by certain pathogenic species; they specify human blood 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 response, and they have an important role in cancer pathology. Evidence has accumulated that the shape (conformation) of saccharides determine the biological function of these compounds (ref. 3).

In the past two decades the accumulation of a large body of structural data based on powerful crystallographic methods of X-ray and neutron diffraction has revealed consistent trends in bonding and chemical connectivity of solid state structure of saccharides (ref. 4). Significant progress has been made in the development of experimental tools, needed to study the conformational complexities of saccharides in solution, in particular high-resolution NMR spectroscopy (ref. 5). A similar revolution has taken place in the computational methods for modeling the shape of saccharides (ref. 6).

This paper attempts to survey the progress in the computational methods for studying oligo- and polysaccharide conformations. It is not our intention to cover all papers which lie within scope of this review. Our aim is rather to analyze the present state of art of conformational analysis of saccharides. The first part of this paper sets out the energetic framework in which we will operate. This is followed by a description of the methods, and strategies used to calculate the energies of the isolated molecules and the effect of the environment, particularly free energy of solvation. Finally, some examples will be given to illustrate how these methods can be used to estimate dilute solution properties of saccharides.

### MOLECULAR THREE-DIMENSIONAL STRUCTURE

For three-dimensional structure the natural questions which arise are whether the relatively minor changes in the chemical structure produce meaningful changes in conformation and, if so, whether the environment (crystalline, solution or protein) plays a role in bringing about and stabilizing these changes. To address these questions, it is important to

find the amount of energy required to alter the various geometrical parameters within a molecule.

### Molecular energetics

Molecular shape is defined by three different types of molecular parameters: bond lengths, bond angles and torsion angles. Variations in the molecular geometry of a molecule are then very simply define as changes in these parameters: bond stretching or compression, bond bending or deformation and bond twisting or torsion. Typical force constant for bond stretching (in mdyn/Å) are 4.4 for the single C-C bond and 5.36 for single C-O bond (ref. 7) Assuming a Hooke's law dependence this translates to approximately 1450 kJ/molÅ<sup>2</sup>, so that the distortion of a single bond of only 0.03Å would cost about 1.2 kJ/mol. Bond angle bending is less expensive and angular distortion of about 10° involves the same amount of energy as the distortion of a single bond by about 0.05Å. Torsional changes involve the rotation about the bond axis; the barrier to rotation about the single C-C bond is about 11.6 kJ/mol, which is approximately the difference between the staggered and eclipsed conformations. The barrier to rotation of the methoxyl group in dimethoxymethane is approximately 3.8 kJ/mol (ref. 7). Thus, as a rule of thumb, bond stretching is roughly two orders of magnitude more expensive energetically than rotations about single bond, with bond angle deformations falling in the intermediate range.

### Intermolecular interactions and molecular geometry

In crystals, in solution or at receptor sites one is dealing potentially with a variety of intermolecular interactions, and most of them are quite weak compared to those involved in chemical bonding. A crystal structure corresponds to a free energy minimum which is not necessarily the global conformational minimum. In any situation of this type the minimum in the potential energy represents a balance between the attractive and repulsion interactions to be discussed subsequently. The nomenclature of these interactions is varied and terms are not always clearly defined or distinguished from one another. Generally, they fall into three classes: non-bonded, electrostatic, and hydrogen bonding. The lines of distinction between these classes are not always particularly sharp. The intermolecular interactions between a carbohydrate and solvent or protein molecules compete with intramolecular forces in the subtle equilibrium which is responsible for the dynamical equilibrium of several interconverting conformers.

The intermolecular interactions all fall on the low end of the scale of energies required to bring about distortions of molecular geometry. This already suggests that if the environment has any influence on the molecular geometric parameters then those most likely to be affected will be the torsion angles around single bonds rather than distortions in bond angles or bond lengths which require substantially larger energies to bring about significant changes. However, the analysis of a large body of structural information concerning the detailed geometry of acetal and hemiacetal moieties in saccharides established (ref. 9) that there are characteristic patterns of bond lengths and bond angles associated with particular torsion angles. This coupling represents a manifestation of the anomeric and exo-anomeric effects. Thus, we are concerned with the general problem of the interplay between intermolecular forces and molecular geometry. Amylose polymorphs (ref. 10) as well as crystal structures of linear and cyclic maltodextrins (ref. 11) may serve to demonstrate differences due to changes in a small number of geometrical parameters. Since the effects are cumulative they lead to significant changes in overall shape of the molecule.

## METHODS AND STRATEGIES FOR COMPUTING CONFORMATIONAL PROPERTIES

In studying a saccharide shape, one is seeking the answers to three basic questions:

1. What are the preferred molecular conformation for a saccharide molecule ?
2. How does the difference in energy between molecular conformations vary in the various environments ?
3. What are the differences in geometry, if any, in the molecular conformations observed in the various environments ?

To answer these questions, a typical study would consist of all or some of the following steps:

1. Determination of the crystal structure to obtain the geometrical parameters.
2. Determination of the existence of different conformations by the appropriate physical measurements.
3. Determination of differences in energy among the different molecular conformations by appropriate computational methods.
4. Determination of differences in energy among the molecular conformations from one environment to another by appropriate computational methods.

We start with a few notes on two first steps and then proceed with a short outline and discussion of computational methods to indicate very briefly the kinds of tools and techniques which can be applied to conformational studies of oligo- and polysaccharides.

Although the crystal structure analyses are generally routine, obtaining a single crystal of suitable quality may be a serious problem for more complex saccharides. In fact the diffraction analysis of any crystalline polymer is not approached in the same manner as a classical single crystal analysis. Because of a relative paucity of diffraction data, the methods where atom positions are determined directly from intensity data are not applicable. Rather, a model analysis technique must be used in which, first, a suitable initial model of the structure is established, followed by its iterative refinement in which differences between experimental and diffraction intensities and those calculated from the model are minimized by suitably varying the model (ref. 12,13). The modeling procedure used to establish the initial model is based on stereochemical constraints and a determination of the differences in energy among the different molecular conformations in the different crystal structures. Recent developments in solid state NMR spectroscopy, particularly the crossed polarization magic angle spinning technique indicate that this could be a very powerful and sensitive probe of solid state molecular conformation and environments for saccharides (ref. 14,15).

Although X-ray and neutron diffractions are powerful techniques, the high-resolution NMR spectroscopy has become the most valuable physical tool for studying conformations of saccharides (ref. 5,16-19) particularly in solution. Chemical shifts, coupling constants, NOE's and relaxation rates contain detailed information about conformational structure of saccharides in solution. However, due to the flexibility of saccharides, experimentally observed parameters might correspond to a dynamical average of several conformers. The existence of several conformers in solution of various disaccharides was already clearly seen in optical rotation experiment (ref. 20). Since the experimental values are averaged values, their interpretation requires the knowledge of all individual conformers involved in averaging process. This is equally true for polysaccharides, for which any experimentally measured property will represent an averaged quantity over all conformational states of that polymeric chain, and therefore will not necessarily reflect the property of most probable state only, provided that other states are available. From an experimental point of view, the range of techniques often play a complementary role in the elucidation of polysaccharide solution conformation (ref. 21). Thus, to interpret the experimental data in solution it is always necessary to have a model based on the methods available for the calculation of molecular energies in the different environments. Fortunately, the powerful techniques for calculating these energies have evolved over the past two decades.

#### Theoretical methods of conformational analysis

Methods of the theoretical conformational analysis can be classified in several ways. We have chosen as most illustrative the one shown in Fig. 1 (ref. 22). The direct methods are based on calculations of the total energy of the molecule  $E_{\text{tot}}$  which is minimized with respect to all or to some of the structural parameters. In indirect methods, on the other hand, conclusions are drawn from the analysis of dipole moments, different types of spectra, etc. In this part of the present paper we shall devote our attention exclusively to direct methods.

The application of direct methods to conformational analysis of saccharides implies two questions:

1. How does one account for the dependence of  $E_{tot}$  on the structural parameters ?
2. How does one carry out the minimization of the  $E_{tot}$  ?

The use of two or more theoretical schemes in estimating the total energy is typical for a non-uniform methods. In general, the total energy is split into the energy of non-bonded interaction,  $E_{nb}$ , and the remaining energy contributions,  $V$ . The contributions  $E_{nb}$  and  $V$  are estimated on a completely different basis. On the other hand, when only one theoretical scheme is required for computation of  $E_{tot}$ , the method is said to be uniform. This is what happens with the quantum mechanical methods in which all-electrons are used, or at least all-valence-electrons.

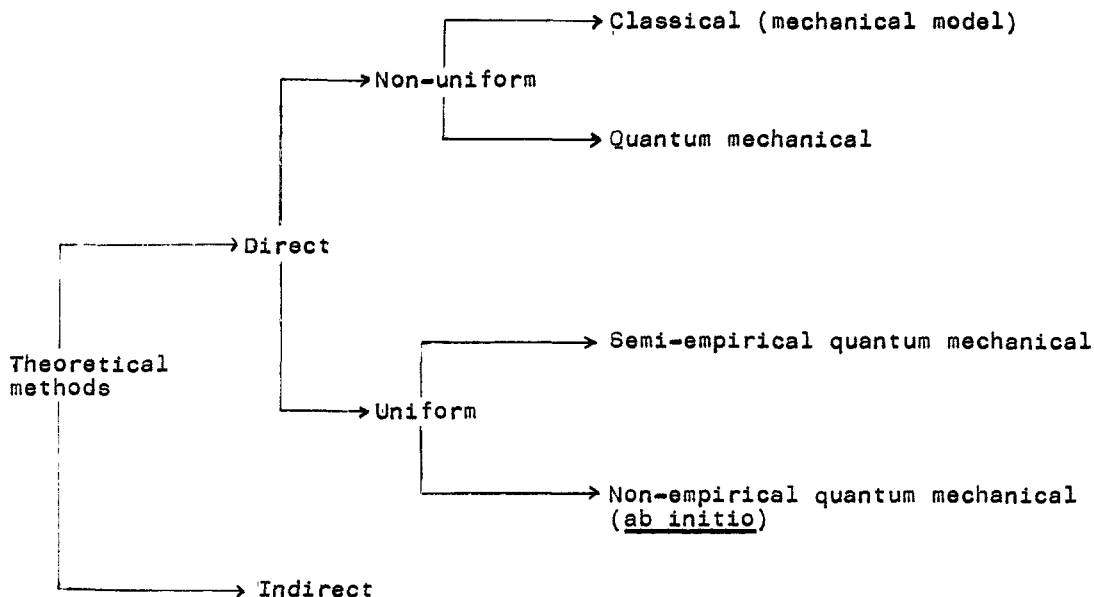


Fig. 1. Classification of theoretical methods of conformational analysis.

#### Classical methods

The non-uniform classical methods which originated from vibrational spectroscopy have now matured into force field or molecular mechanics methods (see e.g. Ref. 23-25). They have included specific terms for perturbations in molecular geometry and the total molecular energy can be written as

$$E_{tot} = E_{nb} + V \quad (1)$$

where the term  $V$  is defined in terms of the same geometrical parameters discussed at the beginning of this paper. The choice of analytical forms for the force field, the determination force field constants and the transferability of force fields and their constants among different chemical systems are subject of considerable interest and research, particularly for systems with heteroatoms. Different kinds of calculations based on these models were applied to saccharides, from a simple one assuming only non-bonded interactions term to the full force field method; and these are well documented elsewhere (ref. 21-30). Here we outline only of some of the more popular approaches in a saccharide field in order to discuss briefly their performance in the investigation of molecular conformations.

**Potential-function calculations.** A group of classical methods which are usually named potential function calculations (PF) is characterized by the incomplete estimation of  $V$  term and the incomplete minimization of the total energy. The  $E_{\text{tot}}$  of a saccharide molecules is minimized only with respect to torsion angles, usually to those defining the relative orientations of glycosidically linked aglycon relative to the monosaccharide unit. The intramolecular energy term  $V$  represents the summation of contributions from electrostatic interactions, torsion term, and hydrogen bonding. These contributions are expressed with the suitable empirical formulae. Procedures differ mainly in the way of estimating  $V$ , and the contributions to  $V$  are selected differently by different authors.

A typical expression for non-bonded interactions  $E_{\text{nb}}$  consist of two parts. One part give the repulsion energy of two atoms at small distances; the other leads to a weak attraction at large distances. Usually one of the following two formulas is used:

$$\text{the Lenard-Jones ; } E_{\text{nb}}(r_{ij}) = B_{ij}/r_{ij}^{12} - A_{ij}/r_{ij}^6 \quad (2)$$

$$\text{or the Buckingham ; } E_{\text{nb}}(r_{ij}) = B_{ij}\exp(-C_{ij}/r_{ij}) - A_{ij}/r_{ij}^6 \quad (3)$$

in which lettered constants are empirical parameters for any pair of atom  $i$  and  $j$  and  $r_{ij}$  is the distance between these atoms.

In the oldest approach, so-called hard-sphere model (ref. 28,29) was used.

$$\begin{aligned} E_{\text{nb}}(r_{ij}) &= 0 && \text{for } r_{ij} \geq r_{ij}^0 \\ &= \infty && \text{for } r_{ij} < r_{ij}^0 \end{aligned} \quad (4)$$

where  $r_{ij}^0$  is the sum of van der Waals radii of interacting atoms. In this approach, the atoms are considered as hard-spheres of a radius equal to their atomic van der Waals radius. In this way, one can construct the conformational maps indicating the sterically allowed zones, without contact between spheres, and the sterically forbidden zones, in which the spheres bump into each other. The hard-sphere method obviously takes cognizance of short-range repulsion only and disregards completely the attractive forces between atoms. Moreover, it operates at an all or nothing level, it does not give any information about the position of the energy minima, i.e. about the preferred conformation.

Various formulations of the potential-function calculations have been used successfully, (see e.g. ref. 26-38), particularly in conjunctions with X-ray diffraction studies (ref. 28-34). Much of the success of this approach is because van der Waals interactions have a dominant influence on the conformations of many oligo- and polysaccharides, and it is relatively easy to rank conformation in order of increasing or decreasing van der Waals repulsion. However, numerous attempts to calculate conformational energies more quantitatively have shown that the direct application of the potential function methods (which had been developed primarily for nonpolar molecules) to saccharides require modification of the method with particular to properly accounting for the anomeric and exo-anomeric effects (ref. 9). Quantum mechanical studies of the model compounds were instrumental in improving the PF method. It was evident from the calculations of torsional energetics of acetals (ref. 39) that the PF procedure must be modified by the proper incorporation of the lone-pair electrons into calculation of intramolecular energy. The lone-pairs bring about the anisotropy of electron distribution in a molecule such that the majority of the charge density is localized in their direction. One may regard the lone-pairs as pseudoatoms with a size and dipole moment and incorporate them into the PF scheme. A simple PF method was successfully modified (ref. 39) in this way, by incorporation of the lone-pair dipoles in the calculation of electrostatic term.

An alternative way of the adapting the PF computational scheme to molecules

exhibiting the exo-anomeric effect consists of the addition of a proper, preferably simple, potential term. In other words, an attempt is made in this approach to establish the extra energy term  $E_{\text{exo}}$  beforehand, and then to incorporate it into the PF method. In one of the first trials, non-bonded interactions (serving as an extra energy term) were combined (ref. 40) with the *ab initio* 4-31G torsional potential (ref. 41), which corresponds to the motionless state at zero absolute temperature, of dimethoxymethane. In that procedure some interactions were evidently counted twice, and the corrected approach known as HSEA (Hard Sphere Exo-Anomeric) was suggested (ref. 42) where the extra energy  $E_{\text{exo}}$  is expressed as a simple function of the torsional angle for each anomer separately. However, the adoption of the acyclic molecule as a model for derivation of  $E_{\text{exo}}$  introduces an artificial character to HSEA method from a physical viewpoint, since the model is symmetrical about  $180^\circ$ . In spite of this, the HSEA method was used in a number of studies (ref. 43). A different approximation to the exo-anomeric energy term (ref. 44) was developed from molecular orbital (MO) calculations from cyclic models as the difference between the torsional potential of 2-methoxytetrahydropyran and that of 2-ethyltetrahydropyran. The resulting different expressions for  $\alpha$ - and  $\beta$ -anomers as a function of the angle  $\tau$  (ref. 45) were subsequently used as additional energy contributions in the so-called PFOS method for studies of oligosaccharide conformations (ref. 30).

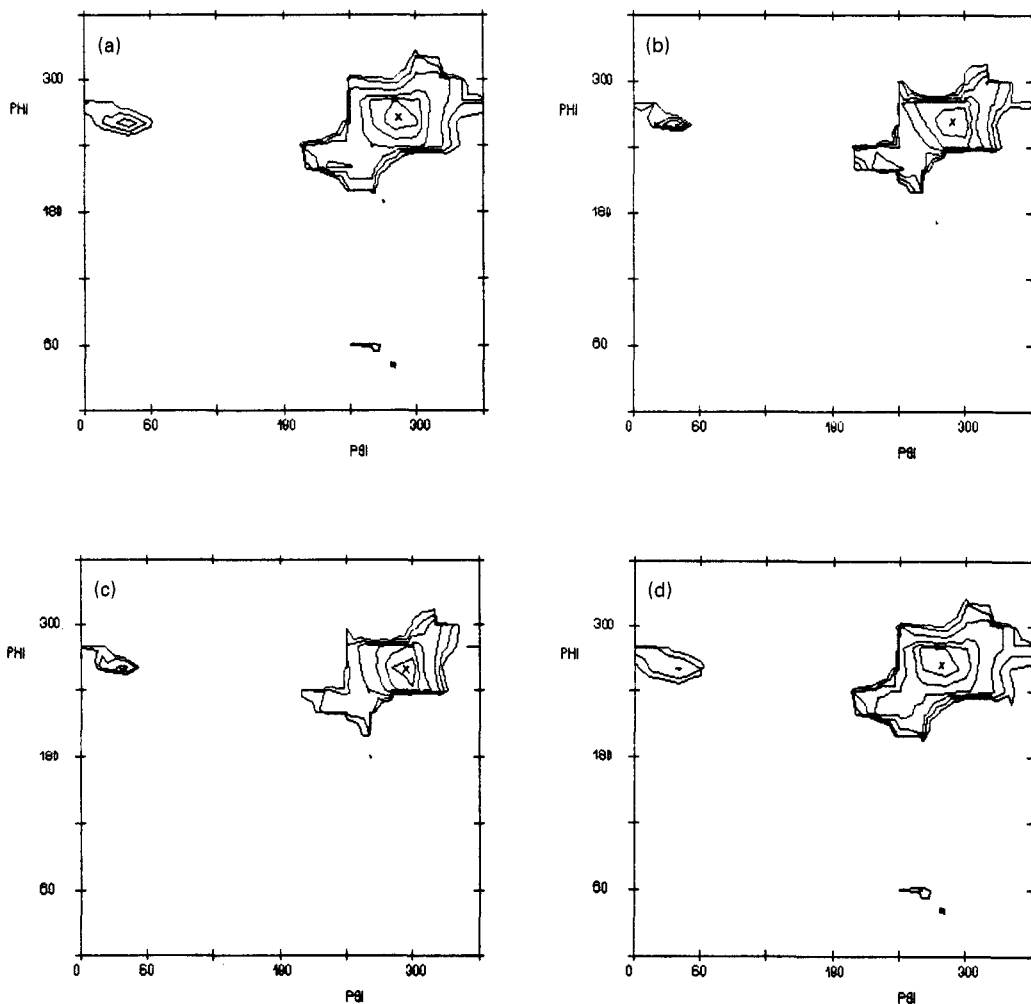


Fig. 2. Energy surface (kJ/mol) for methyl  $\beta$ -xylobioside as a function of torsion angles  $\phi$  and  $\psi$ ; x indicates the lowest calculated minimum; (a) calculated by the PFS method; (b) calculated by the PFK method; (c) calculated by the HSEA method; (d) calculated by the PFOS method.

For purpose of this paper we have calculated energy of methyl  $\beta$ -xylobioside (1) as a function of two torsion angles  $\Phi$  and  $\Psi$  by using following four PE methods:

$$1. \quad E_{\text{tot}} = E_{\text{nb}}(\text{S}) + E_{\text{tors}} \quad (5)$$

where  $E_{\text{nb}}(\text{S})$  is represented by the Lenard-Jones potential (eg. 2) with Scott-Scheraga parameters (ref. 46); torsion around glycosidic C-O bond being represented by a term  $E_{\text{tors}}$  having three-fold periodicity and a barrier 4.1 kJ/mol (ref. 8). We designate this method as PFS.

$$2. \quad E_{\text{tot}} = E_{\text{nb}}(\text{K}) + E_{\text{tors}} \quad (6)$$

where  $E_{\text{nb}}(\text{K})$  is calculated using Kitaigorodski's form (ref. 46) of Buckingham potentials (eq. 3). We used the abbreviation PFK.

$$3. \quad E_{\text{tot}} = E_{\text{nb}}(\text{K}) + E_{\text{exo}} \quad (7)$$

The  $E_{\text{exo}}$  term is based on 4-31G ab initio calculated energy of dimethoxymethane (ref. 41). This method is known as HSEA (ref. 42).

$$4. \quad E_{\text{tot}} = E_{\text{nb}}(\text{S}) + E_{\text{tors}} + E_{\text{exo}} \quad (8)$$

where  $E_{\text{exo}}$  contribution is based on cyclic model compounds (ref. 45), and three-fold barrier in  $E_{\text{tors}}$  term is 8.2 kJ/mol. This method was designated as PFOS (ref. 30).

The calculated two-dimensional energy-contour maps calculated for 1 by the four methods are given in Fig. 2. Maps calculated with the same fixed geometrical parameters but different algorithms have similar shapes but differ in detail. The calculations with PFK and PFS give very similar results. Inspection of the maps revealed the presence of six minima. The HSEA method predict more restricted flexibility than the other three methods. The HSEA method shows only two minima in the two deep narrow wells, centred at  $-60^\circ$  and  $+50^\circ$ . All other conformers have significantly higher energies. The performance of above methods for calculations of oligosaccharides was carefully tested (ref. 30) against available experimental data or previous quantum mechanical calculations.

The analysis revealed that the shape of the conformational surface and relative energies calculated by the above methods depend to great extent on the choice of molecular geometry. Moreover, the form of  $E_{\text{exo}}$  in the HSEA method strongly biases the conformational equilibrium of oligosaccharides towards the unique conformer in the vicinity of  $\Phi$  at  $60^\circ$  or  $-60^\circ$ , whereas other conformers are at much higher energies. Although the other PF methods do not suffer from the same limitation they provide only qualitative description of the conformational energy surface for oligosaccharide.

Force field or molecular mechanics methods. The development of the classical methods culminated in force field or molecular mechanics (MM) methods. The current popularity of these methods appears to rest on the fact that, the most of the inconveniences of PF calculations, like fixed bond lengths and bond angles, either free or frozen bond rotation, and unrealistically hard contact between non-bonded atoms, are removed. The MM methods are based on the same philosophy as the PF method, i.e. a molecule is considered as a collection of atoms held together by the harmonic forces. These forces can be described by the potential functions of structural features like bond lengths, bond angles and so on. The combination of these potential functions is the force field. The energy of the molecule in the force field arises from deviation from ideal structural features and can be approximated by a sum of energy contributions

$$E_{\text{tot}} = E_b + E_a + E_{\text{tor}} + E_{\text{nb}} + E_{\text{hb}} + E_{\text{elst}} \quad (9)$$

The first three terms are the intramolecular terms and describe essentially all of the important intramolecular motions of a molecule, that is bond stretching, angle bending, and internal rotation. The last three terms are used to describe the non-bonded interactions, hydrogen bonding, and electrostatic interactions that is the interactions between the atoms which are not bonded. If there were other intramolecular mechanisms affecting the energy, such as out plane bending or lone pair electron interactions, they may be also added to the force field. The energy of molecule is minimized with respect to all or some structural parameters. Generally speaking, there are no strict rules indicating how many or what types of energy functions should be used. Because of this, many different molecular mechanics force fields have been developed (ref. 23-25). However, it is important to recognize that  $E_{\text{tot}}$  is only a measure of intramolecular strain relative to a hypothetical situation. By itself  $E_{\text{tot}}$  has no physical meaning. In addition, the component terms will change depending on their form and the choice of parameters. The parametrization of the force field relies heavily on experimental data collected over years, and judicious choice must be made in sorting out good and bad data because the reliability of the method can be no better than the data used for parametrization. Force fields have been parametrized or fitted to give excellent geometries, relative conformational energies, heats of formation, crystal packing arrangements, and even reactivities for different types of molecules - except saccharides. The most probable reason for failure is the unusual conformational behaviour of saccharides caused by the interaction of the lone electron pairs of the oxygen atoms, the consequence of which are the anomeric and exo-anomeric effects. The anomeric and exo-anomeric effects are primarily quantum mechanical, and they have no corresponding expression in the classical mechanics. Most recently, however, the first force fields for saccharides have been created and applied for mono- and oligosaccharides.

The first force-field developed for saccharides (ref. 23,24) described correctly the overall saccharide structure, but failed to reproduce the changes in geometry in the vicinity of the anomeric carbon when configurations of substituents are changed. As well as, the incorporation of lone pairs (ref. 7,25) into molecular mechanics schemes MM1 and MM2 did not bring about the improvement in the description of the anomeric center geometry. The modification of MM1 scheme (ref. 48) by the separate parameters for the ring oxygen, glycosidic oxygen and anomeric carbon, which moreover differ for  $\alpha$ - and  $\beta$ -anomers in the calculation of the energy terms, represented the first successful attempt. Recently, the MM2 method was amended in a similar way (ref. 49) and used to calculate the structure of various conformers of oligosaccharides (ref. 30,49). The computations demonstrated the ability of the method to render the variations of disaccharide geometry with the changes in conformation around the glycosidic bonds. In spite of the success, the described modifications are handicapped by the necessity of an introduction of additional parameters to an already large set of parameters. To avoid this problem, the standard C-O bonds length in MM2 method has recently been expressed (ref. 50) as a function of conformation of both of the central C-O bonds in acetal segment. Further testing will assess the merits and shortcomings of this and other modifications of force fields. A full reproduction of the anomeric and exo-anomeric effects in conformational energies and valence geometry of saccharides is an ultimate goal of this effort.

#### Uniform methods

The quantum mechanical methods form the group of uniform methods. In contrast to the classical methods they are very much concerned with the three dimensional distribution of electrons around the nuclei. Today, the situation is considerably different from that in 1965, when it was possible to write in an authoritative book on conformational analysis (ref. 51): "It is fair to state that the quantum mechanical approach to conformational problems has not yet as been very useful. It is the fundamentally correct approach, however, it will probably pay off in the long run". With the help of high-speed electronic computers and advances in conceptual and computational techniques which allow quantum mechanics to be applied to many systems of chemical interest that were before thought to be inaccessible to quantum chemical studies.

With neglect of relativistic effects and within the scope of the Born-Oppenheimer approximation, the exact wave function and energy of the



system are solutions of the Schroedinger equation (ref. 52,53). Based on the approximations used in solving the Schroedinger equation the uniform methods can be classified into two groups: ab initio or non-empirical and semiempirical. Unlike the classical methods, both groups of uniform methods are without ambiguity in the underlying concepts. However, the choice of the method to some extent depends upon the size of the structure to be investigated. Therefore the utility of ab initio quantum mechanical methods is limited because of their computational complexity. They often require a reduction in the size of molecule to a smaller one of manageable proportions. Thus, several small acyclic and cyclic molecules were used as models for studies on the structural segments of saccharides. Fig. 3 shows a classification of uniform methods. The methodology of these methods and the fruits of their applications to saccharides and their model compounds has been subject of recent review (ref. 6,9,54) and therefore will not be discussed here.

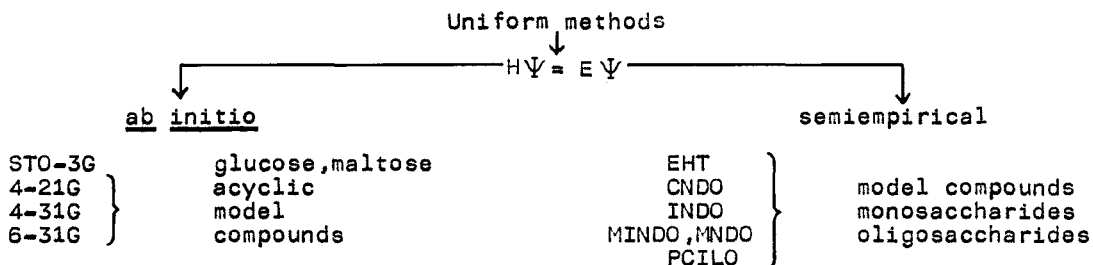


Fig. 3. Classification of uniform methods and the brief indications of their applications to saccharides.

#### Role of the environment

It is important to recognize that most quantum mechanical and classical schemes are designated to treat isolated molecules. Omitting the environment from the calculations amounts, in some cases, to dealing with only half of the energy contributions of the complete system. Fortunately many of the basic principles and models for the calculation of molecular energies are also applicable to the calculation of lattice energetics. In principle the extension of classical approach from molecules to crystal is straightforward. The basic assumption is that the intramolecular interactions may be treated as the sum of atom...atom interactions which, again are approximated by non-bonded interactions, coulombic electrostatic term etc. (ref. 58). For example, crystalline polysaccharides usually adopt a regular helical shape which is subject to quantitative description in terms of relatively small number of lattice parameters and symmetry relationships. Helical shapes are thus easy to describe, conceptually simple, and, hence, frequently adopted as model idealizations (ref. 12,13). Much more complicated is situation in solution and in what follows we shall focus more on this subject.

Solution. Several different approaches have been suggested to this difficult problem (ref. 59). One possible way to overcome part of this problem is to include a limited number of solvent molecule in the calculation. In this case the system of the solute and a small number of solvent molecules is treated as a "supermolecule". The supermolecule approach may be useful in determining the specific solvation sites (ref. 60). However, for most processes in saccharide solution the overall solvation energy rather than the optimal solvation sites is of interest. In fact, the solvation energy can be obtained only by a treatment which includes many solvent molecules (ref. 61).

An alternative approach is to treat the solvent as a dielectric continuum. This approach was based on the solvophobic model and was developed for saccharides (ref. 62,63). It will be described briefly later. The continuum models are useful in estimating conformational energies of polar molecules. However they are of limited validity for the quantitative treatment of ionic reactions and for studies of the balance of interactions between the solute-solute and solute-solvent.

A combination of supermolecule and continuum approaches has been reported recently (ref. 64). Probably ultimate approach for studying solvent effect will be the brute force method of computer simulation of the given solution. This can be based either on Monte Carlo or molecular dynamics calculations including explicitly the solute and many solvent molecules. However, there do not seem to be suitable potential functions for calculating the potential energy of saccharides in different solvent.

The method of free-energy calculation of individual conformers in dilute solution has been described in detail in our previous paper (ref. 62,63) and therefore, here we shall repeat only the basic principles of the model. The process of dissolving a molecule in a solvent involves two steps. The first is the creation, in the solvent, of a cavity of sufficient size to accommodate the solute molecule in the given conformation. The cavity formation requires a Gibbs free energy  $\Delta G_{\text{cav}}$ . The second step is the introduction into this cavity of the solute molecule which then interacts with the surrounding solvent molecules. In our model, the interaction part of the solvation free energy,  $\Delta G_{\text{solv}}$ , is composed of the free energy of dispersion,  $\Delta G_{\text{disp}}$ , and electrostatic,  $\Delta G_{\text{elst}}$ , interactions. The total free energy  $\Delta G_{\text{Tot}}$  of the individual conformers with corresponding intramolecular energy  $\Delta G_{\text{im}}$  in the given solvent can be written as

$$\Delta G_{\text{Tot}} = \Delta G_{\text{im}} + \Delta G_{\text{solv}} = \Delta G_{\text{im}} + \Delta G_{\text{cav}} + \Delta G_{\text{elst}} + \Delta G_{\text{disp}} \quad (10)$$

The calculation of the cavity term is based on an expression taken from the Scaled Particle Theory, which has been successfully used in studies of the thermodynamic properties of aqueous and nonaqueous solutions (ref. 65). The electrostatic term is calculated according to Onsager's theory of the reaction field as applied by Abraham and Bretschneider (ref. 66). The dispersion interactions take into account both attractive and repulsive nonbonded interaction, using a combination of the London dispersion equation and Born-type repulsion (ref. 62). Expressions for the individual terms of the solvation free-energy function have been given, together with all necessary solvent parameters in our previous papers (ref. 62,63,67).

### Strategies

Analysis of the performance of individual theoretical methods led to a suggestion of two possible approaches or strategies in the investigation of conformational properties of oligosaccharides in solution (ref. 30). In the first, the two-dimensional  $(\Phi, \Psi)$  relaxed energy surface is calculated by the appropriate method, optimizing the geometry for each  $(\Phi, \Psi)$  conformer. Then the solvation energy  $(\Phi, \Psi)$  map is calculated for each solvent and superimposed onto the vacuum relaxed surface to obtain different  $(\Phi, \Psi)$  maps for any given solvent. Next, these  $(\Phi, \Psi)$  free energy maps are used to determine statistical weights for the calculation of ensemble averaged NMR experimental characteristics of a given oligosaccharide.

However, the foregoing approach requires considerable effort and computer time, and is probably feasible only for simple disaccharides. In fact, this way of calculations was only used for the investigation of conformational properties of maltose (ref. 68). Therefore, as an alternative procedure the following simpler methodology was proposed. In the first step, a two-dimensional  $(\Phi, \Psi)$  energy surface is calculated with a fixed molecular geometry generated from average coordinates (ref. 69-71). The local minima on this map are used as starting points for optimization of geometry and for precise specification of minima on the  $(\Phi, \Psi)$  surface. At this point, the contribution of solvation energy is calculated and superimposed onto the vacuum energy for each optimized conformer. The average properties are calculated only with these minima and compared with experimental data. Although this approach is usually sufficient, oligosaccharide may appear in conformers which, may have minima on the solvent-specific  $(\Phi, \Psi)$  surface, that are not present on the  $(\Phi, \Psi)$  surface of the isolated oligosaccharide molecule.

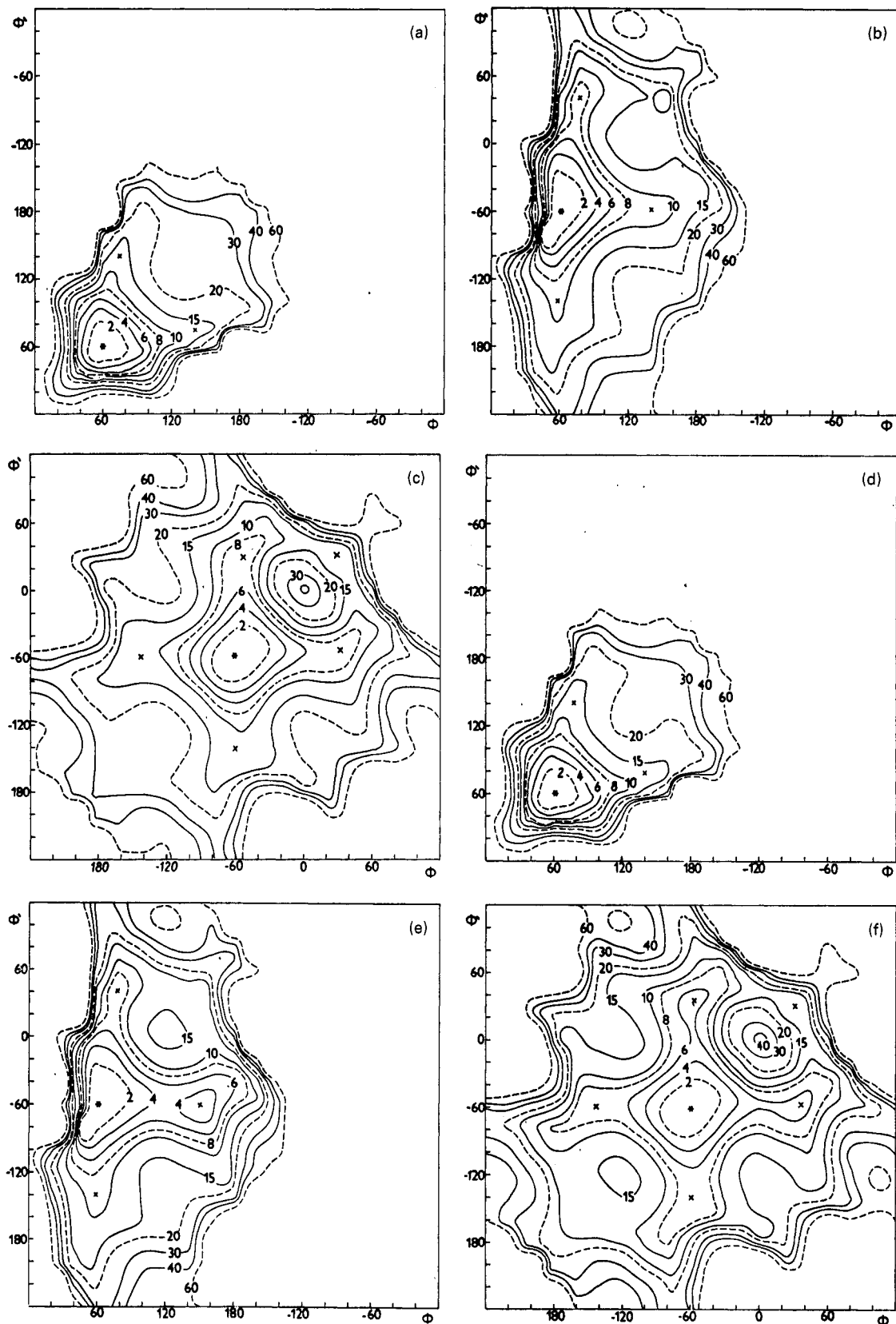


Fig. 4. Energy surface (kJ/mol) for  $\underline{2}$  as a function of torsion angles  $\Phi$  and  $\Phi'$ . Isolated molecule of the  $\alpha, \alpha$ -form (a),  $\alpha, \beta$ -form (b),  $\beta, \beta$ -form (c); aqueous solutions for three forms are on (d), (e), and (f). \* indicates the lowest calculated minimum, x indicates calculated local minima.

## EXAMPLES OF CONFORMATIONAL ANALYSIS IN OLIGOSACCHARIDES SOLUTIONS

A fundamental question which often arises with regard conformational properties of oligosaccharides in solution is that of defining a mixture of conformations whose relative populations, are suitable for interpreting the solution properties of the molecule. We have shown that the computational techniques have matured in the last decade to the point where they can provide rather reliable information on molecular energetics of saccharides, especially when one is concerned with energetic differences in molecular conformations rather than in the absolute energy of any particular conformation. In this section we cite two examples, trehalose and methyl  $\alpha$ -xylobioside, in order to demonstrate that our knowledge of expected conformations and our confidence in predicting them by using the combination of the NMR and computational techniques are now quite well-founded.

**Trehalose.** The incentive for determining the solution behaviour of the trehalose was to compare the orientational properties around a glycosidic linkage in three different orientations of the C-O-C-O-C-O-C segment, and to aid in understanding of anomeric and exo-anomeric effects. Trehalose is the general name given to a family of three D-glucosyl D-glucosides, nonreducing disaccharides namely,  $\alpha,\alpha$ -;  $\alpha,\beta$ - and  $\beta,\beta$ -. Conformational properties of the trehaloses, were studied by using 2-(tetrahydropyran-2-yloxy)tetrahydropyran (2) as a model. The details of the calculations are described in the original paper (ref. 49).

The Fig. 4a shows ( $\Phi, \Phi'$ ) conformational energy maps for an isolated molecule of the  $\alpha,\alpha$ -form of 2 (2aa) and its aqueous solution (Fig. 4d). The contour maps demonstrate how both the exo-anomeric effect and steric interactions between the pyranoid-ring atoms influence the rotation around the glycosidic C-O bonds. Only a small percentage of the conformational space lies within the 20 kJ/mol region. An inspection of the ( $\Phi, \Phi'$ ) maps reveals the existence of three minima. The lowest energy is found for conformer (sc,sc), having  $\Phi = 63.4$  and  $\Phi' = 59.1^\circ$ . The (ap,sc) conformer with  $\Phi = 145.1$ ,  $\Phi' = 67.5^\circ$  lies 9.87 kJ/mol higher in energy. The symmetrically related conformation (sc,ap) with  $\Phi = 67.5^\circ$ ,  $\Phi' = 145.1$  has the same energy. The (sc,sc) conformer corresponds to the crystal-structure conformation of  $\alpha,\alpha$ -trehalose (ref. 55-57). A comparison of the map for 2aa in solution with that for an isolated 2aa molecule shows only a minor alteration in the character of the maps as a result of the solvent effect. This is demonstrated in Fig. 4d for aqueous solution. The minima in solution and energy differences are the same as those for an isolated molecule. The calculated population of the (sc,sc) conformer is 96.5%, both in an isolated molecule and in solution.

Two-dimensional ( $\Phi, \Phi'$ ) contour maps for an isolated  $\alpha,\beta$ -form of 2 (2ab) molecule and a aqueous solution thereof are given in Fig. 4. It is evident from comparison of the 2ab maps with the 2aa maps that the change in the anomeric configuration at one anomeric carbon atom broadens the area enclosed by the contour 20 kJ/mol. The rotation around the equatorially oriented C-1'-O-1 bond is less hindered than that about the axially oriented C-1-O-1 bond. The maps show the existence of four minima (sc,-sc), (sc,ap), (sc,sc), and (ap,-sc). The lowest minimum appeared at  $\Phi = 65.4^\circ$  and  $\Phi' = -54.7^\circ$ , the next, with 5.1 kJ/mol higher energy at (151.5, -58.4). The energies of the (61.9, -147.8) and (72.5, 46.3) conformers relative to (sc,-sc) are 5.6 and 5.3 kJ/mol, respectively. The distribution of conformers for the isolated molecule is 74:8:9:9. In solution, the abundance of the (sc,-sc) conformer decreases as the polarity of the solvent increases, with the lowest value of 33% in water. In contrast, the abundance of the (ap,-sc) conformer increases to 58% in water. The populations of other two change only little. The solvent effect is most pronounced in water. Whereas, in other solvents, (sc,-sc) is the principal conformer, in aqueous solution (ap,-sc) is predicted to be the principal species.

The calculated two-dimensional maps for  $\beta,\beta$ -form of the 2 (2bb), given in Fig. 4 contain six local minima (-sc,-sc), (-sc,sc), (.sc,ap), (sc,-sc), (sc,sc), and (ap,-sc). The lowest minimum (-sc,-sc) appeared at  $\Phi = -58.6^\circ$  and  $\Phi' = -59.0^\circ$ , and it has  $C_2$  symmetry, as well as the (sc,sc) conformer (34.6, 34.6) with 6.8 kJ/mol higher energy. The energy of the (sc, -sc)

(59.0, -39.3) conformer and of the symmetrically related (-sc,sc) (-39.3, 59.0) conformer, relative to (-sc,-sc), is 4.3 kJ/mol. The last two symmetrical conformers, (-33.4, 60.8) and (60.8, -33.4), are situated at 7.8 kJ/mol above the absolute minimum. The distribution of conformers for the isolated molecule is 66:24:6:4. The molecular representations of the most stable conformers for three forms of the 2 are displayed on Fig. 5.

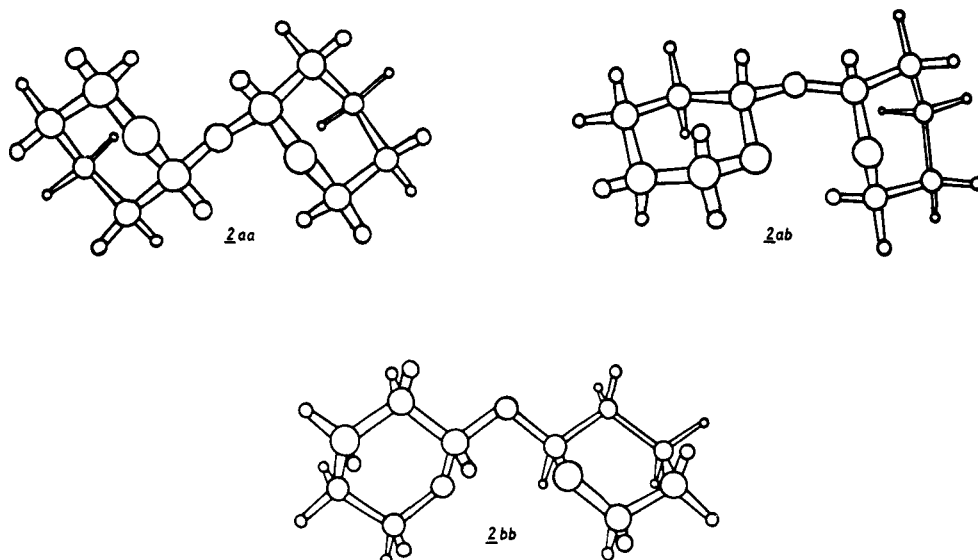


Fig. 5. Molecular representation of the most stable conformations for each form of the 2.

It follows from our results that the rotation around the C-O bond in (1-1)-linked disaccharides depends on the orientation of the adjacent C-O bonds in the C-O-C-O-C-O-C segment. The orientation of the C-1-O-5 or C-1'-O-5' bond is determined by the configuration of the anomer; for the  $\alpha$ -anomer (or in the axial form) the orientation is sc; for the  $\beta$ -anomer (or an equatorial form) the orientation is ap. The shape of the energy curve for the rotation around the C-1-O-1 or C-1'-O-1' bond resembles that in the axial and equatorial forms of 2-methoxytetrahydropyran (ref. 44).

The calculations of the 2-methoxytetrahydropyran revealed the existence of two minima for rotation around the axially oriented C-1-O-1 bond, and three minima for the equatorially oriented C-1-O-1 bond. On the other hand, in the 2, the orientation around the C-1-O-1 or C-1'-O-1 bond influences the rotation about the adjacent C-1'-O-1 or C-1-O-1 bond. Therefore, we have found only three minima for 2aa, four for 2ab, and six for 2bb. Disaccharides usually exist in a complex, conformational equilibrium in solution, and the composition of the mixture of conformers depends on the solvent. In this respect, the solution properties of the  $\alpha,\alpha$ -form of (1-1)-linked disaccharides are unique. As follows from our results, 96% of the 2aa is in the (sc,sc) conformation, and the abundance of this conformer is not solvent-dependent.

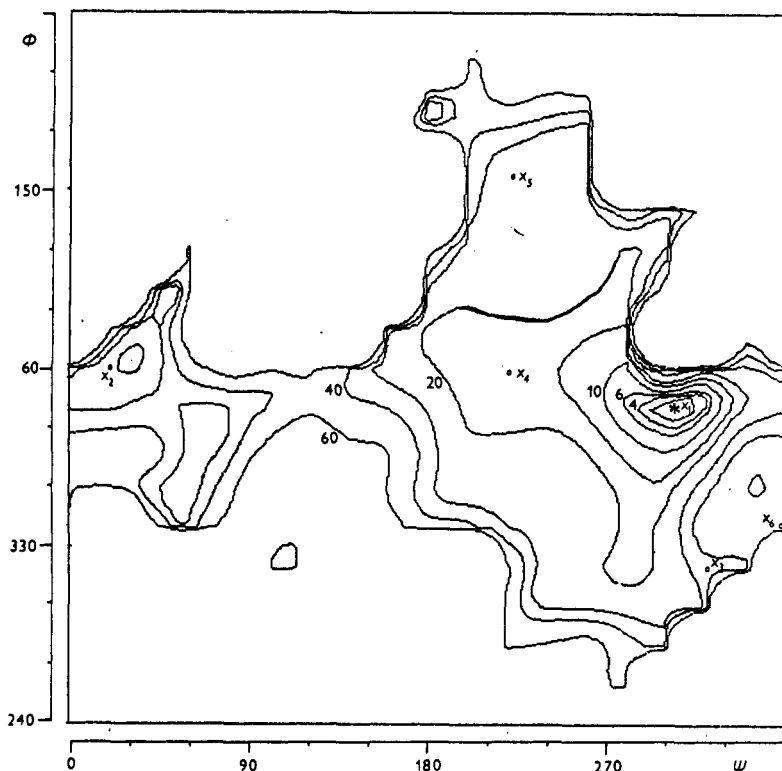
**Methyl  $\beta$ -xylobioside.** The next example is the study in which both theoretical and experimental approaches have been used in order to study solution behaviour of the methyl-O- $\beta$ -D-xylopyranosyl- $\beta$ -D-xylopyranoside (methyl  $\beta$ -xylobioside, 1). Results of semiempirical quantum chemical PCILO calculations with solvent effect evaluation has been correlated with measurements of interglycosidic vicinal coupling constants using new selective two-dimensional NMR techniques (ref. 72). Interest was focused on the study of influence of the solvent and the temperature upon the population of individual conformers.

Heteronuclear interglycosidic coupling constants  $J^\Phi$  are in water relatively constant between 278 K and 318 K ; values are within an interval 4.6 Hz - 4.9 Hz. With increasing temperature (up to 358 K)  $J^\Phi$  decreases to 4.4 Hz and 4.1 Hz respectively. More evident dependence was detected in an array of atoms corresponding to the dihedral angle  $\Psi$ ,  $J^\Psi$ . Values ranged from 5.6 Hz at 278 K to 4.2 Hz at 358 K. Similar trends were obtained in methanol ;  $J^\Psi$  was constant at low temperatures (5.3 - 5.0 Hz), smaller values were obtained at 298 K and 318 K. The dependence of  $J^\Psi$  was also monotonic. The coupling constants decrease from 5.8 Hz at 238 K to 4.5 Hz at 318 K.

Coupling constants obtained in dimethyl sulfoxide indicate that this solvent has a different effect on conformational equilibria of 1 in comparison to previous solvents. Coupling  $J^\Phi$  and  $J^\Psi$  at 298 K are 5.7 Hz and 5.6 Hz, respectively i.e. about 1 Hz larger than in water or methanol, at the same temperature. The temperature dependence is not so pronounced, a maximal change of both couplings is 0.6 Hz at 80 degrees difference. With regard to the averaged character of the measured coupling constants, we have employed PCILO calculations for the quantitative interpretation of  $J^\Phi$  and  $J^\Psi$  coupling constants in NMR measurements.

An inspection of the  $(\Phi, \Psi)$  map (Fig. 6). For the isolated molecule of 1 reveals the existence of six areas of the energy minima. These were used as a starting points for minimization. Then, based on calculated energies the statistical weights for the calculation of ensemble averaged torsion angles  $\langle \Phi \rangle$  and  $\langle \Psi \rangle$  in four solvent were determined. The ensemble averaged coupling constant  $\langle J^\Phi \rangle$  and  $\langle J^\Psi \rangle$  were calculated by using determined Karplus-type equation which describes conformational dependence of  $J_{CH}$  in C-O-C-H segment of bonded atoms on dihedral angle (ref. 73).

Fig. 6. Energy surface (kJ/mol) for methyl  $\beta$ -xylobioside as a function of torsion angles  $\Phi$  and  $\Psi$  calculated by the PCILO quantum mechanical method.



The calculated values of  $\langle J^\Phi \rangle$  and  $\langle J^\Psi \rangle$  follows the experimental trends, at 25 degrees in dioxane the values are 2.7 Hz and 4.9 Hz; in methanol 2.6 Hz and 4.8; in dimethyl sulfoxide 2.7 Hz and 4.8 Hz; and in water 2.8 Hz and

5.0 Hz, respectively. Thus the differences in conformer population of 1 in the equilibrium mixture with the change of temperature and solvent condition observed by NMR measurement might be interpreted with the help of calculations.

### CONCLUDING REMARKS

We have shown how theoretical conformational analysis of saccharides can yield a great deal of information of interest to chemist and biochemist concerned with the solid state and solution. It provides the most direct means for examining differences in conformational behaviour of saccharides which are influenced by the environment. A quantitative measure of this influence of the environment (crystal, solution or protein) may be obtained by computationally examining both the molecular and intermolecular energetics to sort out the interactions which bring about the observed changes from one environment to the next.

While computational techniques have reached now a significant level of sophistication, their applicability to a variety of problems, the transferability of potential forms, and their predictive power are problems and challenges which still remain.

The application of these techniques to problems involving conformational complexities is a symbiotic one: the techniques are required for the investigation of the problems and in their solution one can discover their power and limitations.

The combination of experimental techniques and computational methods is one that is sure to provide a great deal of information on the conformational complexities of oligo- and polysaccharides.

### REFERENCES

1. B. Sharon, Complex Carbohydrates : Their Chemistry, Biosynthesis and Functions, Addison-Wesley, Reading (1975).
2. J.F. Kennedy and C.A. White, Bioactive Carbohydrates : In Chemistry, Biochemistry and Biology, Ellis Horwood, Chichester (1983).
3. K. Olden, B.A. Bernard, M.J. Humphries, T.K. Yeo, K.T. Yeo, S.L. White, S.A. Newton, H.C. Bauer, and J.B. Parent, Trends Biochem. Sci. **10**, 78-82 (1985).
4. G.A. Jeffrey and A.D. French, Molecular Structure by Diffraction Methods, **6**, p. 183, The Chemical Society, London (1978).
5. K. Bock, Pure Appl. Chem. **55**, 605-622 (1983).
6. I. Tvaroška, in G. Náray-Szabó (Ed.), Theoretical Chemistry of Biological Systems, pp. 283-344, Elsevier, Amsterdam (1986).
7. N.L. Allinger, J. Am. Chem. Soc., **99**, 8127-8134 (1977).
8. I. Tvaroška and T. Bleha, Tetrahedron Lett., 249-252 (1975).
9. I. Tvaroška and T. Bleha, Advan. Carbohydr. Chem. Biochem., in print.
10. A. Imberty, H. Chanzy, S. Perez, A. Buleon and V. Tran, J. Mol. Biol., **201**, 365-378 (1988).
11. W. Pangborn, D. Langs and S. Perez, Int. J. Biol. Macromol., **7**, 363-369 (1985).
12. P. Zugenmaier and A. Sarko, Biopolymers **15**, 2121-2136 (1976).
13. P.J.C. Smith and S. Arnott, Acta Crystallogr., **A34**, 3-15 (1978).
14. C.A. Fyfe, R.L. Dudley, P.J. Stephenson, Y. Deslendes, G.K. Hamer and R.H. Marchessault, J. Macromol. Sci. Rev. Macromol. Chem. Phys. **C23**, 187-216 (1983).
15. F. Horii, A. Hirai and R. Kitamaru, ACS Symp. Series, **260**, 27-42 (1984).
16. S.W. Homans, R.A. Dwek and T.W. Rademacher, Biochemistry **26**, 6511-6578 (1987).
17. D.A. Cumming and J.P. Carver, Biochemistry **26**, 6664-6676 (1987).
18. R. Cherniak, R.G. Jones and E. Reiss, Carbohydr. Res. **172**, 113-138 (1988).
19. J.P. Carver and D.A. Cumming, Pure Appl. Chem. **11**, 1465-1476 (1987).
20. D. Rees and D. Thom, J. Chem. Soc. Perkin Trans. II, 191-201 (1977).
21. D.A. Brant, in J. Preiss (Ed.), The Biochemistry of Plants, **3**, Academic Press, New York, pp. 425-472 (1980).
22. A. Golebiewski and A. Parczewski, Chem. Rev. **74**, 519-530 (1976).
23. S.R. Niketic and K. Rasmussen, The Consistent Force Field : A Documentation, Lecture Notes in Chemistry, **3**, Springer-Verlag, Berlin (1977).

24. K. Rasmussen, Potential Energy Functions in Conformational Analysis, Lecture Notes in Chemistry, 37, Springer-Verlag, Berlin, (1985).
25. U. Burkert and N.L. Allinger, Molecular Mechanics, ACS Monograph, **177**, Washington, D.C., (1982).
26. D.A. Brant, Ann. Rev. Biophys. Bioeng. **1**, 369-408 (1972).
27. D.A. Brant, Quart. Rev. Biophys. **9**, 527-596 (1976).
28. G.N. Ramachandran (Ed.), Conformation of Biopolymers, Academic Press, New York, 1967.
29. D.A. Rees, MPT Int. Rev. Sci., Org. Chem. Series One, **7**, 251-283 (1973).
30. I. Tvaroška and S. Pérez, Carbohydr. Res. **149**, 389-410 (1986).
31. P.R. Sundararajan and V.S.R. Rao, Tetrahedron **24**, 289-295 (1968).
32. D.A. Rees, J. Chem. Soc. Sect. B, 217-226 (1969).
33. P. Zugenmeier and A. Sarko, Biopolymers **15**, 2121-2136 (1976).
34. I. Tvaroška, K. Ogawa, Y. Deslandes and R.H. Marchessault, Can. J. Chem. **61**, 1608-1616 (1983).
35. G.M. Lipkind, V.E. Verousky and N.K. Kochetkov, Carbohydr. Res. **133**, 1-13 (1984).
36. A.D. French and V.C. Murphy, Carbohydr. Res. **27**, 391-406 (1973).
37. B. Leps, H. Labishinski, G. Barnickel, H. Bradaczek and P. Giesbrecht, Eur. J. Biochem. **144**, 279-286 (1984).
38. S.G. Whittington, Biopolymers **10**, 1481-1489 (1971).
39. I. Tvaroška and T. Bleha, Biopolymers **18**, 2537-2547 (1979).
40. R.U. Lemieux, K. Bock, L.T.J. Delbaere, S. Koto and V.S. Rao, Can. J. Chem. **58**, 631-653 (1980).
41. G.A. Jeffrey, J.A. Pople, J.S. Binkley and S. Vishveshwara, J. Am. Chem. Soc. **100**, 373-379 (1978).
42. H. Thøgersen, R.U. Lemieux, K. Bock and B. Meyer, Can. J. Chem. **60**, 44-57 (1982).
43. R.U. Lemieux and K. Bock, Arch. Biochem. Biophys. **221**, 125-134 (1983).
44. I. Tvaroška and T. Kožár, Carbohydr. Res. **90**, 173-185 (1981).
45. I. Tvaroška, Carbohydr. Res. **125**, 155-160 (1984).
46. R.A. Scott and H.A. Scheraga, J. Chem. Phys. **42**, 2209-2215 (1965).
47. A.I. Kitaigorodski, Tetrahedron **14**, 230-236 (1961).
48. G.A. Jeffrey and R. Taylor, J. Comput. Chem. **1**, 99-109 (1980).
49. I. Tvaroška and L. Václavík, Carbohydr. Res. **160**, 137-149 (1987).
50. L. Norskov-Lauritsen and N.L. Allinger, J. Comput. Chem., **5**, 326-335 (1984).
51. E.L. Eliel, N.L. Allinger, S.J. Angyal and G.A. Morrison, Conformational Analysis, Interscience, New York (1965).
52. B. Pullman (Ed.), Quantum Mechanics of Molecular Conformations, Wiley, New York (1976).
53. G. Náray-Szabó (Ed.) Theoretical Chemistry of Biological Systems, Elsevier, Amsterdam (1986).
54. I. Tvaroška and T. Bleha, Chem. Papers, **39**, 805-849 (1985).
55. G.C. Brown, D.C. Rohrer, B. Berkin, C.A. Beevers, R.O. Gould and R. Simpson, Acta Crystallogr. **28B**, 3145-3158 (1972).
56. T. Taga, M. Senna and K. Osaki, Acta Crystallogr. **28B**, 3258-3263 (1972).
57. G.A. Jeffrey and R. Nanni, Carbohydr. Res. **137**, 21-30 (1980).
58. H. Tadokoro, Structure of Crystalline Polymers, Wiley, New York (1979).
59. B. Pullman (Ed.) Environmental Effects on Molecular Structure and Properties, Jerusalem Symp. Quantum Chem. Biochem., **8**, Reidel, Dordrecht (1976).
60. I. Tvaroška and T. Kožár, Int. J. Quantum Chem., **23**, 765-778 (1983).
61. R.H. Lemieux, Proceedings of the VIIth International Symposium on Medicinal Chemistry, Uppsala, Sweden, **1**, 329-351 (1984).
62. I. Tvaroška and T. Kožár, J. Am. Chem. Soc., **102**, 6929-6936 (1980).
63. I. Tvaroška, Biopolymers **21**, 1887-1892 (1982).
64. D.A. Beveridge and G.W. Schueler, J. Phys. Chem., **78** 2064-2068 (1975).
65. R.A. Pierotti, Chem. Rev. **76**, 717-726 (1976).
66. R.J. Abraham and E. Bretschneider, in W.J. Orwille-Thomas (Ed.), Internal Rotation in Molecules, Chapter 13, Academic Press, London (1974).
67. I. Tvaroška and T. Kožár, Theor. Chim. Acta **70**, 99-114 (1986).
68. V. Tran, A. Buleon, A. Emberty and S. Pérez, Biopolymers in print (1988).
69. S. Arnott and W.E. Scott, J. Chem. Soc. Perkin Trans. II, 324-335 (1972).
70. B. Sheldrick and D. Akrigg, Acta Crystallogr. **36B**, 1616-1621 (1980).
71. I. Tvaroška and J. Gajdoš, Chem. Zvesti **41**, 485-500 (1987).
72. M. Hricovíni, I. Tvaroška and J. Hirsch, in preparation.
73. I. Tvaroška, M. Hricovíni and E. Petráková, Carbohydr. Res. submitted for publication.