Computer Modeling
of Carbohydrate Molecules

Alfred D. French, EDITOR
U.S. Department of Agriculture

John W. Brady, EDITOR
Cornell University

Developed from a symposium sponsored
by the ACS Divisions of Carbohydrate Chemistry;
Cellulose, Paper, and Textile Chemistry;
and Computers in Chemistry
at the 197th National Meeting
Dallas, Texas,
April 9–14, 1989

The paper used in this publication meets the minimum requirements of American
National Standard for Information Sciences—Permanence of Paper for Printed Library
Materials, ANSI Z39.48–1984.

ACS Symposium Series

M. Joan Comstock, *Series Editor*

1990 ACS Books Advisory Board

Foreword

The ACS SYMPOSIUM SERIES was founded in 1974 to provide a medium for publishing symposia quickly in book form. The format of the Series parallels that of the continuing ADVANCES IN CHEMISTRY SERIES except that, in order to save time, the papers are not typeset but are reproduced as they are submitted by the authors in camera-ready form. Papers are reviewed under the supervision of the Editors with the assistance of the Series Advisory Board and are selected to maintain the integrity of the symposia; however, verbatim reproductions of previously published papers are not accepted. Both reviews and reports of research are acceptable, because symposia may embrace both types of presentation.

Preface

THE MODELING OF CARBOHYDRATES IS UNDERGOING RAPID development. For example, the first comprehensive conformational mappings of disaccharides with flexible residues and the first molecular dynamics studies of carbohydrates have only recently been published. At the same time, interest in carbohydrates has been increasing dramatically, and there is a need for a publication that gently introduces the uninitiated and provides an overview of current research in the area. We feel that *Computer Modeling of Carbohydrate Molecules* meets these needs.

Three of the chapters, *Modeling of Glucopyranose: The Flexible Monomer of Amylose*, *Conformational Analysis of a Disaccharide (Cellobiose) with the Molecular Mechanics Program (MM2)*, and *Molecular Mechanics NMR Pseudoenergy Protocol To Determine Solution Conformation of Complex Oligosaccharides*, were not presented at the symposium on which this volume is based but have been included because they are relevant to the focus of the book. Additionally, an introductory chapter was written by the editors in order to both set the stage for the material that follows and to give some practical advice for those who wish to begin modeling carbohydrates. The chapters are generally placed in order of increasing complexity of either the modeling technique or the molecules being modeled.

Our efforts were supported by three divisions of the American Chemical Society: Carbohydrate Chemistry; Cellulose, Paper, and Textiles; and Computers in Chemistry. Additional financial support was provided by Polygen Corporation, suppliers of the Quanta Modeling System, and Chemical Design, developers and distributors of CHEM-X.

The editors wish to thank the authors who tailored their chapters extensively to meet the goals of this project and the referees who were not authors for their helpful cooperation. Kevin Hicks of the ACS Carbohydrate Division and the U.S. Department of Agriculture provided superb logistical support, for which we are grateful as well.

ALFRED D. FRENCH
U.S. Department of Agriculture
New Orleans, LA 70179

JOHN W. BRADY
Cornell University
Ithaca, NY 14853–7201

May 3, 1990

Chapter 1

Computer Modeling of Carbohydrates

An Introduction

Alfred D. French[1] and J. W. Brady[2]

[1]Southern Regional Research Center, U.S. Department of Agriculture, P.O. Box 19687, New Orleans, LA 70179
[2]Department of Food Science, Cornell University, Ithaca, NY 14853–7201

Carbohydrates comprise a larger portion of the organic world than all other types of biomolecules combined. Cellulose, the primary structural polysaccharide of plant cell walls, is possibly the most abundant of all biopolymers, and monomers, oligomers and polymers of glucose and fructose serve as the energy reserves and primary foods of the biosphere. Carbohydrates are also the bases for large industries and even of entire national economies. Traditional interest in these mainstays of carbohydrate chemistry has recently been supplemented by a growing appreciation of the importance of the carbohydrate fractions of glycopeptides and glycoproteins in many diverse biological roles, such as recognition processes (1).

In both their industrial and biological functions, the 3-dimensional characteristics of carbohydrates are important. Many of these stereochemical features are described for carbohydrates in the classic text by Stoddart (2). The importance of stereochemistry is underscored by the unique chemical and physical properties of the individual sugars, many of which are configurational isomers. Stereochemistry also plays a role in determining the properties of polysaccharides. Molecular shape is as significant for the properties of an industrially modified starch as it is for the recognition of one particular blood type and the rejection of others.

Coincident with this increased interest in carbohydrates, techniques for studying molecular shape have improved. Single crystal diffraction experiments can (and often do) give a fast and precise description of molecules in the solid state. Recent advances in nmr provide increasingly detailed conformational information about the solid state as well as on solutions. However, the structural characteristics of many carbohydrate molecules remain unknown. It is often difficult to obtain the single crystals needed for crystallography, even if the required amount (100 or so milligrams) of pure material is available to attempt crystal growth. Some carbohydrates persist as syrups, and oligomers and polymers often form only microcrystalline particles or fibers that yield inadequate data for a complete structural determination by diffraction methods alone.

In 1960, D.W. Jones supplemented fiber diffraction data from cellulose with a computer model, a list of proposed atomic coordinates that was stored in a digital computer (3). He then

calculated the diffraction intensities that would arise from the
model and compared them with the observed intensities. The model was
repeatedly readjusted, with intensities calculated and compared at
each adjustment, in a trial and error study. While inconclusive,
this was one of the first reports of computer modeling of a
carbohydrate. More recently, the interpretation of coupling
constants and nuclear Overhauser effects from nmr spectra has been
expedited by computer models that provide a ready source of distances
and angles. Since these early efforts, computer modeling has become
an integral part of some procedures for structural determination.
Several papers in this volume discuss the combination of fiber
diffraction with modeling through minimization of the sum of
diffraction intensity error plus the intra- and inter-molecular
energies. Several other papers cover augmentation of nmr data with
theoretical simulations in various ways.
     In 1963, V.S.R. Rao undertook a more ambitious task: the
prediction of the likely conformations of polysaccharides from a
computerized survey of model structures (4). As a result of atomic
overlap, some model conformers had higher energies than others, a
criterion by which most models could be rejected. These predictions
were not accompanied by experimental data for the subject molecules,
leaving to experimentalists the task of corroboration or refutation.
Although many advances in computers and methods have occurred in the
intervening decades, predicting polysaccharide conformations based
upon relative conformational energies continues to be of substantial
interest.
     Theory is best combined with experiment, so that each can
support the other. However, some problems are not amenable to
experiment. Acyclic glucose, for example, occurs in such small
concentrations that experimental data is overwhelmed by data from the
pyranose forms. In contrast, a model is easily built and studied.
Tasks that are even more hypothetical are also within the
capabilities of modeling, such as a comparison of molecular
properties with and without a hydrogen bond. A good modeling study
provides a framework for integrating the experimental results from
various techniques to provide a greater overall understanding.

Theoretical Background for Computer Modeling

Molecular modeling calculations attempt to predict physical
properties for molecular systems based on the numerical solution of
the equations that embody the physical laws that govern their
behavior (5,6). At the most fundamental level, this approach
involves the direct solution of Schrödinger's equation for the
nuclear and electronic degrees of freedom. Since these studies
determine energies directly from first principles, they are referred
to as ab initio calculations. Such calculations rapidly become
impossible in the practical sense for systems containing more than a
few atoms heavier than hydrogen, and it becomes necessary to invoke
various additional approximations to extend these calculations to
systems containing more than about two dozen atoms.
     Small monosaccharides have molecular sizes at the upper limit of
the range that is currently treatable with ab initio methods. An
example of the application of ab initio calculations to carbohydrates
is given in the paper by Garrett and Serianni in this volume.
Semiempirical quantum mechanical calculations, which use simplified
molecular Hamiltonians with parameters taken from experiment, extend
quantum mechanical calculations to larger molecules. However, the
reliability is reduced compared to the best ab initio results.
Recent advances in semiempirical quantum methods (see the chapter

herein by Dieter and Stewart) make them quite viable for molecules
the size of a disaccharide, for example, which can be especially
useful if steps along the pathway of some chemical change are to be
modeled.   However, alternate approaches are needed to extend
theoretical descriptions to larger molecules and condensed-phase
systems.

Molecular Mechanics

One such approach is to treat the motions of atomic nuclei as
classical particles, since most of the quantal character of molecules
resides in their electronic motions.   It is then possible to use the
Born-Oppenheimer approximation to solve for the electronic energies
at fixed nuclear positions, and to treat these electronic energies as
the potential energy field for the nuclear motions.   After this
separation, analytic, empirical energy functions may be used to
approximate the way in which the molecular energy changes with the
coordinates of the constituent atoms.   We will call any technique a
"molecular mechanics" (mm) calculation if it uses such analytic
energy functions to predict changes in a system's energy arising from
variations in its atomic coordinates (5,6).   Of the papers in this
volume, the majority reports mm studies of one type or another.
     The empirical energy functions used in mm calculations usually
consist of sums of terms representing various, easily-conceptualized
contributions to the total energy of a molecule.   For example, such
energy functions generally contain terms to represent the energy of
stretching or compressing chemical bonds, bending bond angles, and
changing torsion angles.   These functions also generally contain
terms to represent van der Waals (non-bonded) interactions and
electrostatic interactions between the various partially charged
atoms and/or dipoles in a molecule.   An example of such an energy
function is given by equation (1) in the paper by Madsen, et al. in
this volume.   If the energy components have been carefully selected,
then these semiempirical expressions may give a useful approximation
to the functional dependence of the molecular energy.   The adjustable
parameters that appear in these functions can then be selected by
exhaustive comparison of calculated molecular properties with
experimental measurements, in order to give the most physically
realistic representation possible with the chosen functional form.
Two characteristics of carbohyhdrate structure are often given
special attention when constructing potential functions.   Although
hydrogen bonding and "anomeric effects" are certainly important in
other compounds, they are thought to be especially important for
carbohydrates.

Hydrogen Bonding.   Because of great strength (for attractions not
based on covalent interactions), short range, and strong angular
dependence, hydrogen bonding is often a powerful structuring force.
This is certainly true for carbohydrates, which have hydroxyl groups
that can simultaneously donate and accept protons of hydrogen bonds.
Rotatability of the hydroxyl groups and possible bifurcation make
prediction of hydrogen bonding structures difficult, but the energy
of hydrogen bonding is a major force in determining carbohydrate
structures.   Some published potential functions treat hydrogen bonds
in the same way as all other electrostatic interactions.   Other
programs use separate terms for hydrogen bonding energy or, in
capitulation to the complexity, simply ignore hydrogen bonding
altogether.
     To indicate the impact of hydrogen bonding on studies of
molecular shape, one can look ahead to Figure 5, a calculation of the

intramolecular energy for various conformations of maltose. Imagine
the effect of a new, additional, intermolecular, hydrogen bond with a
heat of formation of -5 kcal/mol. This hypothetical new hydrogen
bond might be possible only near a conformation of $\phi = 0^{\circ}$, $\psi = 180^{\circ}$,
a point that is more than 3 kcal/mol above the central minimum that
had been based solely on intrinsic forces. When the new, inter-
molecular bond is considered, the minimum at $\phi = 0^{\circ}$, $\psi = 180^{\circ}$ would
be favored compared to the central minimum, completely changing the
results from the modeling study. The paper by Tran and Brady herein
shows changes in the preferred conformation that depend on the
presence or absence of an intramolecular hydrogen bond.

Because of the effects of intermolecular hydrogen bonds (and
other intermolecular forces - see the chapter by Ragazzi et al.), the
study of isolated molecules may not be as helpful as desired for the
prediction of conformations in condensed phases. While it will
always be interesting to know the vacuum conformations, it may be
necessary to explicitly include hydrogen bonding partners for the
prediction of conformations in solids or solutions. Since the energy
of hydrogen bond formation is large relative to the Boltzmann
constant times room temperature, errors in the treatment of hydrogen
bonds in modeling studies can result in the prediction of structures
that would be quite improbable for a given set of experimental
conditions. Uncertainty regarding hydrogen bonds overshadows many
other types of errors in modeling studies.

Anomeric Effects. Several characteristics discovered first for
carbohydrates are associated with anomeric carbon atoms. For
example, despite the lower energy in cyclohexane or cyclopentane for
equatorial substituents, axial (α) forms are often preferred for
substituents at the anomeric center in solutions of substituted
sugars such as methyl glycosides. The axial anomers of these
glycosides also tend to be more stable toward acid hydrolysis.
Secondly, the C-O bond lengths are altered in the vicinity of the
anomeric carbon, depending on the anomeric form and the rotational
orientation of the substituent attached to the glycosidic oxygen
atom. These aglycon groups of glycosides also exhibit a marked
orientational preference for gauche conformations, a phenomenon known
as the exo-anomeric effect.

These effects, the subjects of a recent, comprehensive
review (7), are now recognized as general characteristics of
molecules that have two electronegative heteroatoms linked to a
tetrahedral carbon center. Except for the readily observed
differences of C-O bond lengths in carbohydrate crystals, there is
substantial controversy regarding the magnitude and theoretical basis
for these effects (8). For example, the apparent preference for
axial substituents may not be so much an intrinsic property of the
molecule. Instead, the preference may depend more on the relative
energies of solvation, since these preferences refer to dissolved
molecules. Several molecular mechanics potential functions have
attempted to incorporate suitable treatments of these effects (9-12).

Energy Minimization. The phrase "molecular mechanics calculation" is
perhaps associated most often with the determination of an "ideal"
structure through automated optimization of atomic positions, where
the term "ideal structure" usually is taken to mean the lowest-energy
structure. This energy minimization approach is based on the
assumption (which also underlies many quantum mechanical studies)
that physically observed properties will be those of the single
structure with the lowest mechanical energy, implicitly equating the

free energy of the system with this mechanical potential energy (or enthalpy). The structure of lowest energy can be sought in various ways, from elaborate energy minimization calculations to simply searching for conformations that avoid steric overlaps, as Ramachandran first did with his studies of allowed dipeptide conformations (13). Even modeling studies that use interactive computer graphics to visually "dock" one molecule on the surface of another on a video screen could be considered as mm energy minimizations. In this situation, an intuitive "feel" for the repulsive energy resulting from atomic overlaps guides the modeler in moving the two molecules. Other types of mm calculations, such as molecular dynamics (see the chapter by Madsen et al., this volume) and Monte Carlo simulations (14), attempt to simulate the ensemble-averaged behavior of molecular systems. These last two methods directly include entropic effects so that relative free energies can be calculated. Also, molecular properties determined by these studies are based on weighted averages over the many different possible states.

Limitations of Molecular Mechanics

Philosophical Limitations. It must be remembered that in any of these calculations using an analytic semiempirical energy function, the division of the molecular energy into arbitrary categories is a simplistic caricature of the actual physical situation. Therefore, mm has inevitable limitations on how accurately it can model the behavior of real molecules. As an example of these limitations, the force constant for bending a given bond angle in a model will depend to some extent on whether the two bonds that define the angle are allowed to stretch or compress, and, if so, by how much (as in a so-called "valence force field" in molecular spectroscopy). Because of these interrelationships, observeable structural parameters and energies cannot, in general, be used directly in parameterization. Rotations about torsion angles will also involve deformations of the bond lengths and angles, as well as van der Waals interactions. Furthermore, many empirical energy functions treat bond lengths and angles as harmonic oscillators, while in reality, there may be significant anharmonicity in these motions even for thermally accessible deformations. Harmonic functions do not provide for bond breakage at large extensions, or for bond exchange, which means that commonly used mm energy functions cannot treat chemical reactions.
     All of these problems can to some extent be addressed by making the energy functions more complex or by introducing more adjustable parameters. For example, Morse functions could be used instead of harmonic oscillators for bond deformations. Also, bond stretches and bends could be coupled. However, the basic philosophical limitation of mm methods remains. That being said, it is also true that mm methods can provide structures and energies that are as good as, or better, than those resulting from more elegant and time-consuming methods, _if_ mm is used within its limitations.

Strategic Limitations. While the actual application of mm calculations to carbohydrate molecules is in most cases straightforward, there are a number of pitfalls that may trap the unwary. The first concerns the choice of an appropriate potential energy function to be used for a particular problem. The adjustable parameters that appear in the energy function must be carefully chosen to give the closest match possible to relevant experimental data. Unfortunately, because observable atomic characteristics vary as a function of environment, atomic parameters developed for the

vapor phase may not be transferable to the same atom or group in a
solution or a crystal. For the same reason, different potential
energy functions, containing different approximations and with
different characteristics, may be needed for types of molecules that
are as different as proteins and carbohydrates. Simply mixing
parameters taken from disparate sources can lead to a very poor
approximation, regardless of how useful the individual parameters
were in the integral unit from which they came. A wide variety of
force fields for carbohydrates has been developed (9-12, 15-22), as
well as general-purpose force fields such as MM3 (23). Several
aspects of force field development are discussed in the chapter by
Rasmussen and Fabricius, and a paper herein by Scarsdale et al.
discusses modifications to a force field used for proteins to
accommodate the carbohydrate fraction of a glycopeptide. Some care
should be taken to select a combination of energy function and
parameter set that is appropriate for the problem at hand. In cases
of nonstandard structures, users should test the applicability of the
force field to the most analogous system possible for which
experimental data are available. (This is a good idea for standard
structures, too!)
     Energy minimization calculations are frequently fraught with
operational difficulties. Most arise from the complex nature of the
potential energy function for a large molecule with many internal
degrees of freedom (5,6). Typical mm programs contain one or more
energy minimization algorithms that can be used to automatically
adjust the atomic positions to reduce the molecular energy. Usually,
this procedure is applied repeatedly until a local minimum in energy
is approached closely enough that further effort would result in
insignificant improvement. New users are sometimes tempted to
terminate these optimizations prematurely because of the time
required to produce what are often small improvements in energy.
However, because of the complexity of the total energy function for a
large molecule, large structural transitions can occur surprisingly
late in an optimization. Early termination of the minimization can
miss transitions that would lower the energy substantially after a
long period of minor change.

The Multiple Minimum Problem. Attempts to find the molecular
structure with the lowest energy are difficult for molecules as
complicated as carbohydrates. The biggest obstacle is the "multiple
minimum problem" which arises because energy surfaces for complex
molecules have more than one local minimum, and because algorithms
that minimize the energy in mm (or quantum mechanics) programs will
proceed from the starting conformation to the closest local minimum
on the energy surface and stop. This is shown in Figure 1, where a
typical energy minimization starts from a structure with a high
energy on the far left part of the curve, and terminates in the local
minimum at A, since there is no general solution to the problem of
the global minimization of a multidimensional function. In general
there is no way to know that A is not the global minimum unless
additional energy minimizations are started from a structure having a
conformation beyond barrier B.
     Figure 1 also illustrates one of the problems that comes from
associating the lowest energy structure with the physically observed
structure. Point C has the lowest potential energy (the global
minimum), but beyond barrier D there is another minimum at E that is
not quite as deep but much broader than the minimum at C. As a
result, the Boltzmann-weighted integral over this conformation is
larger than that for the C well, meaning that the well at E with
slightly higher potential energy is actually the favored form due to

its lower free energy. However, it should also be clear that, except for the crystalline state, the observed equilibrium properties of such systems would be Boltzmann-weighted averages of the characteristics of all three wells (indeed, of all possible states). Several papers in this book have converted the calculated energies for the many conformations studied into populations for the calculation of average properties. The motions of the system, such as can be simulated by molecular dynamics studies (see the chapter by Madsen et al.), can allow an equilibrium partitioning between all of the possible forms. If the dynamics simulation is run long enough, the molecule will spend enough time in each conformation to be representative of the real system and the free energy can be determined.

Taking a specific example of the multiple minimum problem from carbohydrate chemistry, suppose that glucose was optimized from a starting shape with O6 of the primary alcohol group near one of the three staggered positions that result from rotation about the C5-C6 bond. Typically, the modeling software would produce a model with O6 in the "best" position close to the position in the starting model, but would not test the other two likely O6 positions. To learn which of the three positions is "best", starting models with O6 in the other two staggered positions might also be optimized and the energies of all three models compared. (We note that this simple treatment has not yet produced agreement with experiment.)

Monosaccharides have many structural variations that correspond to local minima that must be considered. Acyclic carbohydrates can rotate at each carbon, and each of the three staggered conformers is likely to correspond to a local minimum. The shapes of sugar rings also often vary. Furanose rings usually have two major local minima and a path of interconversion. Experimental evidence shows a clear preference for only one chair form for some pyranose rings, but others could exist in several conformers. For example, the $^4C_1$, $^1C_4$ and $^2S_0$ conformers must all be considered as possible structures for L-iduronate, as discussed by Ragazzi et al. in this book.

Each secondary hydroxyl group could also have one of three staggered conformations. The number of local minima arising from each type of variation must be multiplied by the number of each other type to give the total number of local minima that should be anticipated. Any analysis of oligomers or polymers must acknowledge the hopelessly large numbers of local minima in these molecules. Fortunately, some simplifying assumptions can be made (see the papers herein by Tran and Brady and by French, Tran and Pérez). The paper herein by Tvaroška, Kožár and Hricovíni describes a random walk method to sample alternate arrangements of pendant groups.

Comparison with Experiment. For carbohydrates, modeling work has often sought to produce a theoretical structure that matches the results from a highly accurate diffraction experiment. This approach is problematic, however, because crystal packing can distort a molecule from the shape that it would have as an isolated molecule. As shown in the chapter by French, Rowland and Allinger, when the same, relatively rigid glucose residue appears in numerous crystalline environments, there are substantial differences in the conformation. Of course, the errors in the experimental determination must also be considered.

These problems aside, diffraction studies still may not provide results that describe molecular structures in one of the most interesting states: aqueous solution. For solutions, diffraction experiments are less relevant, and nmr and circular dichroism are the best sources of experimental information (see the chapter by

Tvaroška, Kožár and Hricovíni in this book). One way to account for
the effect of solvent on conformation might be to represent the
molecule without environmental influences, and then explicitly
include the solvent or other environmental molecules in the
calculation. While avoiding built-in influences of environment is a
satisfying concept, it is difficult to obtain by experiment
parameters that lack those influences. Several methods have been
used to study solvation effects, including continuum descriptions
(24) and the explicit treatment of solvent molecules in Monte Carlo
and molecular dynamics simulation.

Conformational Analysis

Rather than collecting and comparing structures that correspond to
the multiple minima, it is often more useful to depict how the energy
changes during variations of one or more of the most important
structural features, generating a surface similar to Figure 1. Such
a comprehensive study of the energy surface is called conformational
analysis (CA). In CA, the energy is calculated at suitable
increments of important conformational coordinates, and plotted, at
least conceptually, on a one-, two- or multi-dimensional grid.
Such energy maps depict the heights of the barriers and the widths of
the minima, as well as showing the positions of the minima.

CA of Monosaccharides. The simplest type of conformational analysis
for glucopyranose might be a systematic rotation of its primary
alcohol group. The hydroxymethyl group could be rotated in
increments of $20°$, with the structure optimized at each increment in
all respects except for the primary alcohol position. Even this
"simple" analysis is complicated because at some points alternate
hydroxyl hydrogen orientations on the rotating oxygen atom may result
in lower energy than that given by the starting orientation.
Usually, however, these alternate hydroxyl group orientations will
not occur automatically as a result of the energy minimization
process. It will be necessary to try each possible staggered
rotation of the hydroxyl group at each rotation of the hydroxymethyl
group. That change may in turn require changes in the rotation of
the hydroxyl group on C4, and so forth.
        Variations in ring conformation are often the primary interest,
even when the general type of ring shape is known. Ring shapes are
described semi-quantitatively by their puckering, a measure of the
departure from an all-planar shape. The Cremer-Pople (C-P) system
(25,26) permits the description of N-membered rings with N-3
parameters, a very useful shorthand. Although the C-P puckering
notation applies to rings of any size, 5- and 6-membered rings are
described here because they are especially common in carbohydrates.

Furanose Conformations. Figure 2 shows the different envelope (E)
and twist (T) forms for the 5-membered rings. This "conformational
wheel" is for ketofuranoses, in which the ring carbon atoms are
numbered 2-5; the carbons of aldofuranoses are numbered 1-4. (The
conformations of some aldofuranoses are described in the chapter by
Garrett and Serianni.) The hydrocarbon analogue, cyclopentane,
readily interconverts between all adjacent E and T forms, a
phenomenon known as pseudorotation. In carbohydrates, with a ring
oxygen atom and varied substituents, some forms have lower energy
than others. Therefore, there are barriers several kcal/mol high to
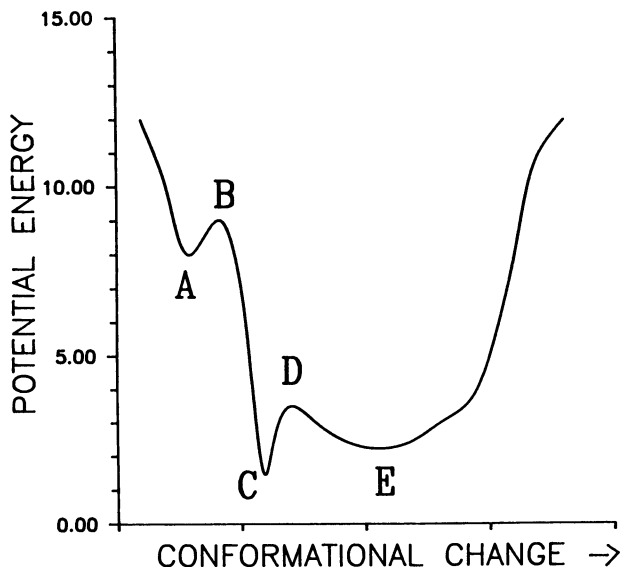facile pseudorotation.

Figure 1. Potential energy values as some (unspecified) conformational variable is changed. A represents a local (false) minimum and C represents the global minimum (assuming that all other variable parameters are also in the least energetic conformations). B and D are barriers that are not overcome during minimization, and E is a broad minimum that contains a range of structures. In solution, both C and E structures would be present in substantial amounts. Because of the small enthalpy difference between C and E, and the greater number of structures belonging to the E classification, E structures would dominate.
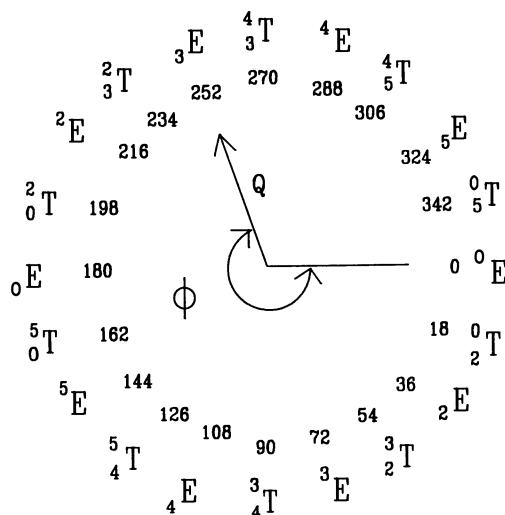

Figure 2. Puckering angles (ϕ) for perfect envelope (E) and symmetrical twist (T) forms of fructofuranose. The non-planar ring atoms in symmetrical twists are displaced equally above and below the ring. The amplitude of puckering (Q) is the radius of the circle.

For five-membered rings, there are two major C-P puckering parameters, ϕ and Q. The phase angle of puckering, ϕ, describes the position on the conformational wheel at which the puckering occurs (Figure 2). The puckering amplitude, Q, describes the extent of deviation of the ring atoms from a mean plane. These two parameters are conveniently depicted in a planar polar coordinate system. A conformational analysis of the furanose ring entails a $360°$ range of ϕ; Q values will range from 0 to 0.6 or 0.8 Å.

The energies of the different ring conformations are affected by the rotational orientations of the two primary alcohol groups of fructofuranose. Therefore, all 9 combinations of likely orientations of these groups must be considered before the energy differences inherent in different ring conformations can be understood (French, A.D.; Tran, V.H. Biopolymers, In Press).

Pyranose Conformations. Figure 3 shows the different conformations for 6-membered rings (adapted from a drawing by Jeffrey and Yates (27)). There is a θ parameter besides Q and ϕ because several types of puckering are possible for a given Q and ϕ. In addition to the E (envelope) notation used in Figure 3, six-membered rings with only one out-of-plane atom are also called sofas or half-boats. The E descriptor was selected here because S is already used to denote skewed pyranose conformations (which have two atoms on opposite sides of the plane, separated by one atom). The H label is already used for half-chairs, which have two adjacent atoms on opposite sides of the mean plane. Typically, the E and H forms are not important unless a double bond is present.

Spherical polar coordinates are used for conformational representation of pyranose rings in the C-P system. Unlike the free pseudorotation of cyclopentane, the stable conformations of cyclohexane conformers are in deeper energy wells. Even among the (less stable) equatorial ($\theta = 90°$) forms, pseudorotation is somewhat hindered. Substitutions of heteroatoms in the ring and additions of hydroxylic or other exocyclic substituents further stabilize or destabilize other conformers compared to cyclohexane. A conformational analysis of an iduronate ring has been reported based on variation of ϕ and θ (28), and a study of the glucopyranose ring (29) based on the conformational descriptors of Pickett and Strauss (30) is also available.

Most modeling software packages do not vary puckering in increments of the C-P parameters (REFINE (28) is an exception) but a utility for producing rings with specific puckerings is available (31)). Instead, most programs have "dihedral drivers" that allow for the specification of starting and ending values of torsion angles as well as increment size. Another mechanism for changing conformation is available in some programs. This allows the positions of selected atoms to be fixed in space while allowing all other atoms to relax to positions of lowest local energy. Some programs allow variation of one or two of the coordinates of selected atoms while holding the other coordinate(s) fixed. This can be used to restrain ring atoms to a specific puckering. Optimized pyranose rings with various restrained conformations have puckering amplitudes that vary substantially (Haasnoot, C.A.G., personal communication).

Alternate Conformational Representations. In some instances, the C-P formalism does not provide the most economical description, as exemplified by the six-atom ring of dihydropyran. Like cyclohexene, dihydropyran has a double bond that enforces essential coplanarity on four contiguous ring atoms. While the C-P puckering description for
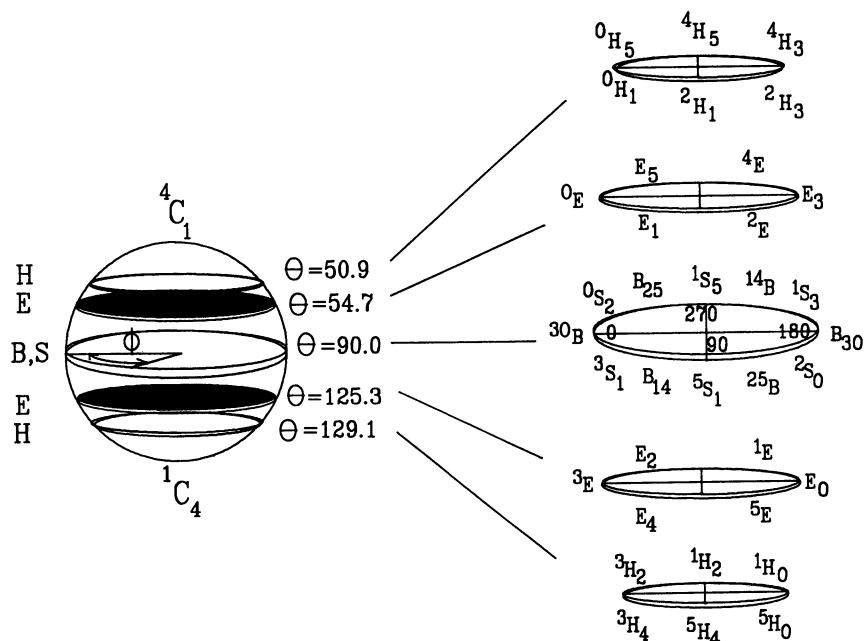
Figure 3. The conformational sphere for pyranoid rings. The perfect chairs are at the north and south poles (θ = 0 and 180°, respectively). The boat and skew (B and S designations) at the equator permit pseudorotation that is slightly hindered, at least for cyclohexane. The envelopes, E (also called sofas and half-boats), and half-chairs, H, are not observed for rings composed of saturated carbon and oxygen atoms, but are important forms for rings with unsaturated carbon atoms. The amplitude of puckering corresponds to the radius of the sphere.

dihydropyran involves all five of the disks in Figure 3, most conformations cannot be accessed without great energy penalty because of the double bond. Therefore, the conformational surface can be simplified.

The two ring atoms that are opposite the ends of the double bond in dihydropyran are allowed to deviate from the plane enforced by the double bond. The ring conformations can be adequately summarized by plotting the energy against the displacements of these two atoms (Figure 4). This provides a quick indication of the position and amplitude of puckering. Another example of economical representation is discussed in a chapter by French, Rowland and Allinger. Therein, the flexing of the glucopyranose ring within the $^4C_1$ conformation was studied by plotting the energy vs. the distance between O1 and O4. In that paper, the C-P notation is also used to advantage.

CA of Disaccharides. Because of the multiple minima problem, disaccharides are inherently more complicated to model than monosaccharides. More computer time is required to optimize disaccharide structures because the cpu time depends roughly on the number of atoms squared. A typical CA of a disaccharide determines the variation in energy for all mutual orientations of the two monosaccharide residues. These orientations are expressed by the glycosidic linkage torsion angles, ϕ and ψ, shown in Figure 5. The orientations of the secondary hydroxyl groups are important because their positions often affect the calculated energy in pyranose rings by several kcal/mol. The preferred side group orientations will change when ϕ and ψ change, so several possible arrangements must be considered. In simple work, the hydroxyl hydrogen atoms are sometimes ignored with the justification that they could usually rotate to avoid a conflict.

CA's of disaccharides give different results depending on whether or not the residues are allowed to adjust internally at each increment. If the internal bond angles and torsion angles for each residue in a disaccharide are allowed to change in response to influences from the other residue during CA, the study is called a flexible-residue analysis. If not, then the study is called a rigid-residue analysis. The CA of maltose, shown in Figure 5, is a flexible-residue map. Some disadvantages of the rigid-residue method are:

1. Rigid maps depend on the exact choice of starting model (8).
2. Important minima on the energy surface are likely to be ignored since energies are high on most of the surface except near the starting shape.
3. Barriers between minima are likely to be grossly overestimated.

Disadvantages of flexible-residue models include:

1. Much more computer time is needed.
2. The input and results are more complicated.

Several recent papers compare the two types of analysis (32-34) (see also the paper herein by Tran and Brady). Brant and Christ compare the abilities of the two approaches to predict experimental behavior in their chapter herein. Two strategies for constructing relaxed maps are discussed in the chapters by Tran and Brady and by French, Tran and Pérez.
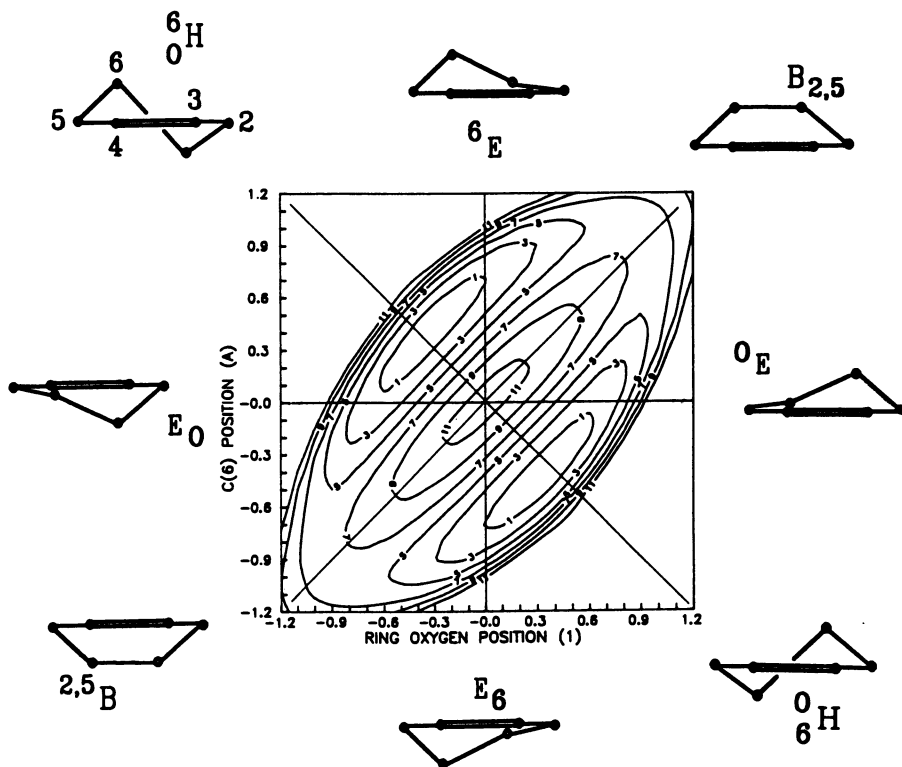
Figure 4. Conformational map for dihydropyran. Because of the
double bond, 4 atoms are always almost coplanar and a limited number
of conformations is probable. The energy contours are at 2 kcal/mol
intervals, starting 1 kcal/mol above the minima. The favored
conformations are half-chairs, and the easiest paths of transition
between the two are through the boat forms. The symmetry of this
energy map applies only to dihydropyran, and not to derivatives which
cause increases and decreases in the sizes of the allowed (low-
energy) areas. This map was calculated with MMP2(85) at increments
of 0.1 A shift of the two non-planar atoms. Three of the carbon
atoms were held in a plane while C6 and O1 were held at specific
distances above and below the plane. Otherwise, the structure was
fully relaxed at each increment. The reader may enjoy plotting the
indicated path of conformational interchange (pseudorotation) on a
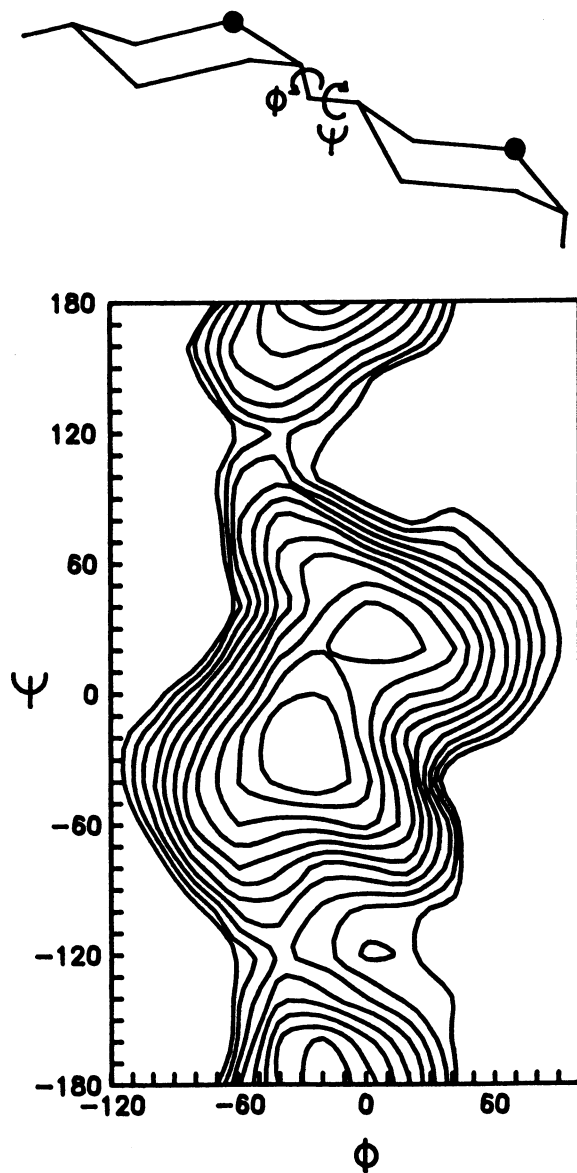copy of Figure 3.

Figure 5. Conformational map for maltose, calculated with MMP2(85), using the methods that are described in the chapter herein by French, Tran and Pérez with four starting models. ϕ and ψ were varied in steps of $20°$. Contours are at intervals of one kcal/mol.

While most CA's of disaccharides have depended only on intrinsic characteristics of the molecule, experimental results depend strongly on the environment. By experiment, Kamide and Saito (35) have shown that the degree of flexibility of cellulose and its derivatives is strongly dependent on the dielectric constant of the solvent as well as the exact type and degree of substitution. Since a substantial portion of the polymer flexibility depends on the extent of variability of the torsion angles at the intermonomer linkage, the dependence of polymer flexibility on type of solvent and substitution means that the disaccharide flexibility also should depend on those factors. Non-polar solvents allowed the molecules to have greater flexibility than did polar solvents (35).

CA of Polysaccharides. Polysaccharides adopt a wide variety of shapes that depend on their composition and their environment. In solution, polymers are almost always random coils that have local regions that might be similar to conformations that are found in the solid state. The chapter by Brant and Christ discusses conformations of polysaccharides in solutions both in terms of these local regions and by the overall shape of the random coil in terms of end-to-end distance, etc. The following discussion concerns only linear (unbranched) molecules, and refers only to regular polymers, i.e., those that have repeated sequences of monomeric residues located by screw-axis (helical) symmetry.

The parameters **n** and **h** are simple descriptors of the conformations of regular helices. **n** is the number of residues (or repeated residue sequences) per helix turn, and **h** is the rise per residue along the helix axis (Figure 6). By definition, if all residues and linkages are identical to their predecessors and successors, the polymer is a helix. A helix can be left-handed or right handed (Figure 6); a helix with **n** = 2 has both chiralities since the second residue can be generated from the first equally well by rotation about the helix-axis in either direction. Some workers designate left-handed helices by negative values of **n**, while other workers have used negative values of **h** for left-handed structures. Helices can be non-integral; a helix with **n** = 3.5 would have seven residues completing two helix turns. An alternate nomenclature (see the paper by Millane herein) describes only the symmetry of the helix and conflicts somewhat with established crystallographic nomenclature. In this system, the number of residues per crystallographic repeat is given, with a subscript of the number of helix turns per repeat. Thus, a polysaccharide might be described as an 8_3 helix. In **n-h** nomenclature, **n** would be 2.666, and **h** would be one-third of the length of the fiber repeat distance. A third alternative (see the paper by Ragazzi et al. herein) describes helices in terms of the rotation per residue about the helix axis. Thus, a ϕ_h of 180° indicates a helix with two residues per turn, and 90° indicates a four-fold helix.

To analyze conformations of polysaccharides, workers have modified the ϕ-ψ programs used for CA of disaccharides to include monitoring distances to atoms farther than the next residue along the chain. When a modeling program has been suitably modified for polymers, it gives an extremely high energy region at **h** = 0. This is because, for **n** > 1, **h** = 0 requires successive helix turns to occupy the same space, a physical impossibility.

The results of such studies are obtained in ϕ, ψ space and a routine calculates the **n** and **h** values as functions of ϕ and ψ. This conversion depends on the exact coordinates of the residue and a precise glycosidic bond angle. However, most monosaccharides are
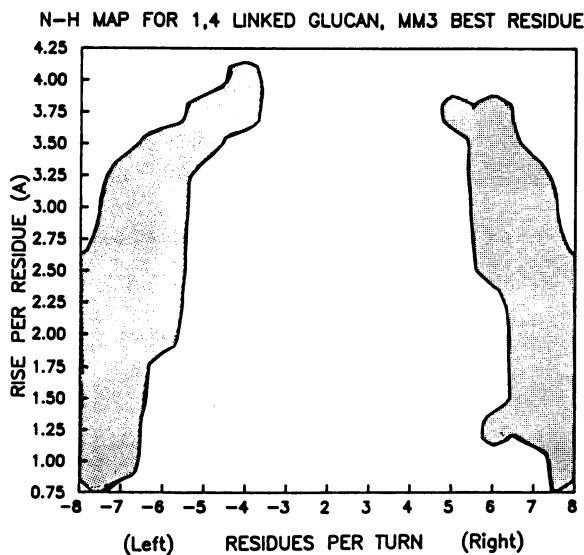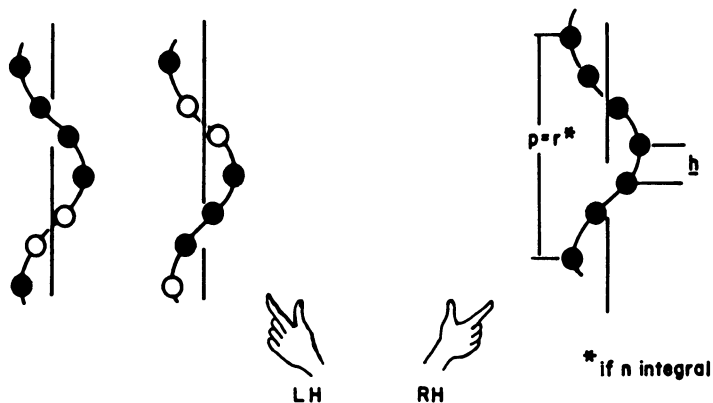
Figure 6. An **n-h** map for amylose (α-1->4 glucan) along with definitions of left- and right-handed helices and **n** and **h**. With the thumb aligned along the helix axis, the hand with the forefinger that follows the helix backbone indicates the chirality. (An inverted left-hand helix is still left-handed!) The number of residues, **n**, per helix repeat, **r**, (or pitch, **p**) shown is six, and the rise per residue, **h**, is indicated. There are two allowed zones (shaded) on this **n-h** map, for right- and left-handed helices. The glycosidic angle was allowed to have values of 110 - 122°. The residue geometry was the optimized MM3 model, taken from the chapter in these proceedings by French, Rowland and Allinger, and gives rise to collapsed helices (low **h** values) with 7 and 8 residues per turn and extended (left-handed) helices with 4 residues per turn. Helices with **n** = -6 and **h** = 3.5 A, corresponding to native starch (see the chapter by Imberty, Pérez and Scaringe) are also allowed. Other residue geometries expand the ranges of the allowed helical shapes considerably.

somewhat flexible (29) (also see the chapter herein by French, Rowland and Allinger), and polysaccharides often crystallize in various shapes that require variations in the geometry of the constituent monosaccharides. Therefore, a thorough conformational analysis for a polysaccharide requires either the use of flexible residues or parallel studies with several rigid residues that span the important range of residue variation. When flexibility of the residues and linkages is incorporated, the values of ϕ and ψ cannot specify **n** and **h**. Thus, an alternate representation (36) of conformation-space (Figure 6) is useful. It plots the results of the analysis (in this example, for inter-residue, hard-sphere contacts) of each **n-h** combination on a grid of **n** and **h**.

Practical Perspectives

Simple wire, ball and stick, and space-filling models have been useful in the practice of chemistry for many years. We do not predict that computer modeling will eliminate physical models. There will always be advantages to concrete representations of attempts to describe molecules. Instead, computer modeling offers an unprecedented degree of quantification and an avenue to analysis of dynamic behavior not possible with physical models. Also, computer models can be built very quickly.

Many of the techniques for computer modeling were developed some time ago but the slow speed and high cost of computers prevented their wide application. Now, compared to laboratory instruments, a computer suited for many types of modeling studies may seem quite inexpensive. (It even may seem inexpensive compared to a good collection of real models!) In the past decade, the price/performance ratio of computers has improved by two orders of magnitude and it appears that improvements will continue.

The prospect of widespread and lucrative markets for easy-to-use modeling software has spurred development of software packages that can be used by those who are not full-time computer specialists. The keys to use by non-specialists are the graphical user interface and the availability of "canned" software. (The chapter by Jeffrey adds perspective on the use of "canned" software.) A number of programs for modeling are available from the Quantum Chemistry Program Exchange (QCPE), Department of Chemistry, Indiana University, Bloomington, Indiana 47901, for as little as $100, and fairly comprehensive packages such as TRIBBLE cost $400. Commercially developed packages sometimes incorporate routines or even entire programs from the QCPE (with proper credit) but can cost as much as $100,000 or even more. However, academic and other not-for-profit users can often buy the same software at discounts up to 97%.

The computing time required for theoretical studies varies widely depending on the type of modeling being done. For example, the minimization of potential energies for 13 glucose residues with different O1--O4 distances by five different programs required about two cpu hours on a DEC Microvax 3100. A fairly thorough (relaxed) conformational analysis of a disaccharide might take a week of cpu time on the same computer. Solution dynamics simulations of a disaccharide might require several cpu months before reliable results could be obtained on the same machine. The latter problem is clearly a candidate for a super-computer.

Computers costing in the range of $10,000 to $20,000 can execute most of the widely available modeling software packages. This is quite inexpensive compared to the cost of developing major software, or even of continuously converting new releases of programs written for another computer. Therefore, it may be quite reasonable to buy a

specific computer just because it will run a particular program without further effort. The advice that users should select software first and then hardware has proven solid over the years. Another view is that no single software system is likely to provide all the answers and that major hardware expenditures should support as many candidate programs as possible.

One aspect of computer modeling is easy to ignore until work commences. Running even a relatively slow computer for hours, weeks or months produces a tremendous amount of information. Hundreds of megabytes of disk space are needed to contain a large modeling package and the output from two or three productive people. For large simulations, the analysis of these data sets becomes the rate-limiting step for productivity of the modeler. Some of this data will be managed by a large modeling system, but as the user requests more sophisticated information from modeling, development of unique programs may be necessary. For example, major modeling packages usually do not provide output of the C-P puckering parameters. Coordinates of structures output from a modeling study must be put into the correct form for input to a program for puckering parameters if that information is desired. For such reasons, some computer skills are needed.

Conclusions

Can a carbohydrate chemist become a computer modeler? Having started several people with diverse backgrounds on modeling studies, both of us think that is more likely than a modeler becoming a bench chemist. Admittedly, there are many pitfalls for the beginner and vexing problems for the experienced worker. The abilities of modeling are often oversold. However, the promise and honest success indicated by the collected papers in this book show that computer modeling of carbohydrates has already produced a wide variety of useful results.

Literature Cited

1. Albersheim, P.A.; Darvill, A.G. Scientific American 1985, September, 58-64.
2. Stoddart, J.F. Stereochemistry of Carbohydrates; Wiley-Interscience, New York, 1971.
3. Jones, D.W. J. Polym. Sci., 1960, 42, 173-188.
4. Rao, V.S.R.; Sundararajan, P.R.; Ramakrishnan, C.; Ramachandran, G.N. In Conformation in Biopolymers, Vol. 2, Academic Press: London, 1963.
5. Brooks, C.L.; Karplus, M.; Pettitt, B.M. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics, Advances in Chemical Physics Series, Vol. LXXI. Wiley, New York, 1988.
6. McCammon, J.A.; Harvey, S.C. Dynamics of Proteins and Nucleic Acids, Cambridge University Press, Cambridge, 1987.
7. Tvaroška, I.; Bleha, T. in Advances in Carbohydrate Chemistry and Biochemistry; Tipson, R.S.; Horton, D. Eds.; Academic: San Diego, 1989, Vol. 47, pp 45-123.
8. Tvaroska, I.; Pérez, S. Carbohydr. Res. 1986, 149, 389-410.
9. Lemieux, R.U.; Bock, K.; Delbaere, L.T.J.; Koto, S.; Rao, Can. J. Chem. 1980, 58, 631-653.
10. Tvaroška, I. Carbohydr. Res. 1984, 125, 155-160.
11. Nørskov-Lauritsen, L.; Allinger, N.L. J. Comput. Chem. 1984, 5, 326-335.
12. Marsden, A.; Robson, B.; Thompson, J.S. J. Chem. Soc. Faraday Trans. I, 1988, 84, 2519-2536.

13. Ramachandran, G.N.; Ramakrishnan, C.; Sasisekharan, V. in Aspects of Protein Structure; Ramachandran, G.N., Ed.; Academic: London, 1963, p 121.
14. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. J. Chem. Phys. 1953, 21, 1087-1092.
15. Goebel, C.V.; Dimpfl, W.L.; Brant, D.A. Macromolecules 1970, 3, 644-654.
16. Rao, V.S.R.; Vijayalakshmi, K.S.; Sundararajan, P.R. Carbohydr. Res. 1971, 17, 341-352.
17. Rees, D.A.; Smith, P.J.C. J. Chem. Soc. Perkin Trans. II 1975, 830-835.
18. Thøgerson, H.; Lemieux, R.U.; Bock, K.; Meyer, B. Can. J. Chem. 1982, 60, 44-57.
19. Rasmussen, K. Acta Chem. Scand. Ser. A 1982, 36, 323-327.
20. Pertsin, A.J.; Kitaigorodsky, A.I. The Atom-Atom Potential Method - Applications to Organic Molecular Solids; Springer-Verlag: Berlin, 1987.
21. Koehler, J.E.H.; Saenger, W.; van Gunsteren, W.F. Eur. Biophys. J. 1987, 15, 197-210.
22. Ha, S.N.; Giammona, A.; Field, M.; Brady, J.W. Carbohydr. Res. 1988, 180, 207-221.
23. Allinger, N.L.; Yuh, Y.H.; Lii, L-H. J. Am. Chem. Soc. 1989, 111, 8551-8566.
24. Tvaroška, I.; Kožár, T. J. Am. Chem. Soc. 1980, 102, 6929-6936.
25. Cremer, D.; Pople, J.A. J. Amer. Chem. Soc. 1975, 97, 1354-1358.
26. Essen, H.; Cremer, D. Acta Crystallogr. 1984, B40, 418-420.
27. Jeffrey, G.A.; Yates, J.H. Carbohydr. Res. 1979, 74, 319-322.
28. Ragazzi, M.; Ferro, D.R.; Provasoli, A. J. Comput. Chem. 1986, 7, 105-112.
29. Joshi, N.V.; Rao, V.S.R. Biopolymers 1979, 18, 2993-3004.
30. Pickett, H.M.; Strauss, H.L. J. Am. Chem. Soc. 1970, 92, 7281-7290.
31. Millane, R.; Nzewi, E.U. J. Appl. Crystallogr. 1989, 22, 138-143.
32. French, A.D. Biopolymers, 1988, 27, 1519-1525.
33. Ha, S.N.; Madsen, L.; Brady, J.W. Biopolymers, 1988, 27, 1927-1952.
34. Tran, V.; Buleon, A.; Imberty, A.; Pérez, S. Biopolymers, 1989, 28, 679-690.
35. Kamide, K.; Saito, M. Advances in Polymer Science (Biopolymers) 1987, 83, 1-56.
36. French, A.D.; French, W.A. In Fiber Diffraction Methods; French, A.D.; Gardner, K.H. Eds.; ACS Symposium Series No. 141; American Chemical Society: Washington, DC, 1980, pp 239-250.

Chapter 2

Experimental and Theoretical Bases for Accurate Modeling

An Experimentalist Looks at Modeling

G. A. Jeffrey

Department of Crystallography, University of Pittsburgh, Pittsburgh, PA 15260

Canned Science has the great advantage that it allows the experimentalist to carry out theoretical calculations on molecules of his interest. Conversely the theoretician has a vast resource of structural data available in data bases or can carry out routine crystal structure analyses to test his predictions. Both adventures have pitfalls for the inexperienced. In the molecular modeling of carbohydrates, hydrogen-bonding is particularly troublesome because of the orientational freedom and the donor and acceptor property of the many –OH groups. This leads to a multiplicity of local minima of nearly equal energy. The hydrogen bonding in peptides and proteins, in comparison, is easier to model, since the predominant $>$NH and C=O groups have neither orientational freedom nor donor and acceptor properties. Calculations on isolated molecules tend to emphasize intramolecular hydrogen bonding, which is frequently superceded by intermolecular bonding in solution or in the solid-state. Electrostatic interactions are difficult to parameterize due to polarization effects and the difficulty of defining the point-charge on an atom. To overcome these obstacles to successful modeling of oligo- and polysaccharides, the *hydrogen-bond structure* under investigation needs to be constrained to take into account existing data on such factors as (1) most probable bond lengths and bond angles, (2) three-center hydrogen bonding (about 25 percent), (3) cooperativity; favoring infinite or finite chains, homodromic loops, ribbons and nets, (4) the "Excluded Region", which places constraints on hydrogen-bond lengths versus hydrogen-bond angles.

In the last decade, there has been an explosion of "Canned Science". By Canned Science, I refer to computer programs and data bases which can be bought and used successfully simply by following the "instructions on the can". It is no longer necessary to be a theoretical chemist to carry out computations at any of the three levels of molecular modeling; *ab-initio*, semi-empirical, or empirical force field (1-3). By the same token, it is not necessary to be a crystallographer to have ready access to 70,000 organic and organo-metallic crystal structures, of which about 2000 are carbohydrates (4). A minimal knowledge of crystallography is required to carry

out a single crystal structure analysis using the modern computer-controlled X-ray diffraction equipment, which includes all the software necessary to measure the structure amplitudes, determine the structure phase angles and refine the atomic positional and thermal parameters. The results of advances in computer technology are especially valuable at a time when these aspects of science are becoming increasingly specialized. It provides the researcher with access to generations of expertise in fields other than that of his own. It permits the specialist once again to become a generalist, at least to some degree. The use of this computer software by investigators who are relatively inexperienced in the appropriate theory or experiment has its pitfalls, of course (5). There could be an increase in inaccurate data or misleading predictions. The experimentalist using a theoretical program will, if he is wise, use the general criterion; do the results make chemical sense? If possible, he will first apply the theory to a problem where he already knows the answer before extending it to a related problem where he desires an answer. Similarly, the theoretician will, or should, first test the theoretical methods against known answers. Most commonly available programs do, in fact, provide this 'evidence' of credibility. However, unlike experimental structure determination, the theoretical methods do not include an internal means of calculating standard deviations for the bond lengths, valence angles and torsion angles derived. There is no way of assessing their accuracy other than by comparison with experimental data, or with theoretical calculations carried out by other methods. It is not difficult to envision a next level of software that warns the user of possible erorrs, such as inconsistency with existing relevant data, unrealistic energies, unlikely molecular geometry, or unacceptable non-bonding distances. Editors of both experimental and theoretical journals should welcome these programs. They would both improve and make more standard the present refereeing procedures, which make relatively little systematic use of available computer data bases.

Once computers leave the bureacracy of computing centers and become laboratory instruments, there is a burst of innovative and adventurous applications. The appearance of the departmental supercomputer, with its dedicated software for molecular modeling by *ab-initio* or molecular dynamics calculations, at a price comparable to that of the departmental NMR or X-ray diffraction equipment, suggests that this time is imminent. A conservative use of theory by the experimentalist is to correlate his observations and to obtain an insight into the electronic interpretation of what he observes. The early *ab-initio* calculations on methanediol, methoxymethanol, and dimethoxymethane in connection with the anomeric and exo-anomeric effect is such an example (6-8). In that work, *ab-initio* calculations were used to correlate a chemical observation, the anomeric effect (9), with the crystallographic observation of a bond-length shortening (10) and a preferred orientation of the glycosidic bond (11). Another example is the use of a semi-empirical method, PCILO, to examine the cooperativity or non-additivity of cyclic systems of hydrogen bonds which were observed from the X-ray and neutron diffraction structure analyses of the α-cyclodextrin hexahydrate (12).

The theoretician uses these programs to predict structure, either of single molecules or of assemblages of molecules, using X-ray or NMR data, when available, to test his predictions (13-15). It has been known for a long time that even the earlier molecular mechanics programs can predict the structures of certain types of molecules with excellent reliability. For the cyclic alkanes, an accuracy comparable to that of the best X-ray crystal structure analysis can be obtained. In fact, the method is more widely applicable since neither compound nor crystals are necessary (16).

With monosaccharides, the structures of the relatively rigid pyranoses and methylpyranosides in the crystalline state can be accurately predicted if the hydroxyl groups are oriented as in the crystal, in the directions appropriate to form

intermolecular hydrogen bonds (17). Oligosaccharides present a more difficult problem since some *ad-hoc* decisions must be made concerning the intramolecular hydrogen bonding. Intra-residue intramolecular bonds have to be prevented, while certain inter-residue intramolecular bonds have to be permitted. For cellobiose, for example, there are two quite different conformations, both with two inter-residue intramolecullar hydrogen bonds, shown in Figure 1a&b. In the crystal structure (18), only one is observed, as in Figure 1c. It is difficult to predict the influence of these bonds on the conformational populations in solution without reliable solvation studies.

    With three linkage bonds as in the molecular mechanics study of the 1–6 disaccharide gentiobiose (19), the lowest 24 minima within 7 kcal mole^{-1} which were reported did not include that observed in the crystal structure (20). Had the crystal structure conformation been one of the sets of starting parameters, it would have been included, but it is reasonable to assume that the conformational population in solution was also not fully represented.

    Another example of problems arising from hydrogen bonding is provided by an interesting comparison of semi-empirical MNDO, PCILO, and molecular mechanics calculations on n-acetyl β-D-glucosamine (21). These three methods gave different energy sequences for the nineteen global minima considered. The lowest from PCILO was one of the highest from MNDO, and the lowest from a molecular mechanics method was not the lowest from the two semi-empirical methods. This could be a consequence of the different treatment of hydrogen bonding by the three methods considered. A previous paper in this symposium suggested that MNDO underestimates hydrogen bond energies, while PCILO may overestimate hydrogen bonding. The crystal structure of n-acetyl β-glucosamine has not been determined since it is the α-epimer that crystallizes from aqueous solution. The crystal structure of the α-epimer suggests that all the hydrogen bonding will be intermolecular for the β-epimer in the solid state or in solution (22).

Why is Hydrogen Bonding a Difficult Problem in the Molecular Modeling of Carbohydrates?

There are two obvious reasons. One is clearly the high level of hydrogen bond functionality that is present in carbohydrates, with four donors and six potential acceptors per monomer in an oligo- or polysaccharide. Each hydroxyl group can be both a hydrogen bond donor and acceptor, and each ring or glycoside oxygen may or may not be an acceptor. In contrast, the $\diagdown$NH, –NH$_2$, –$\overset{+}{\text{N}}$H$_3$ functional groups, common in other biological molecules, are donors but rarely acceptors, while N and O=C groups can only be acceptors.

    The second reason is that every –OH- - -O hydrogen bond involves knowledge of the -C-O-H torsion angle. This means that each hydrogen bond introduced into the molecular model involves at least one additional variable parameter and preferably more. The assumptions that hydrogen bonds are linear and that the hydrogen atom will lie on, or even close, to the O- - -O line have been shown to be invalid by the many neutron diffraction studies of the hydrogen bonding in crystals (23). In the carbohydrate, nucleoside and nucleotide crystal structures, for example, about 25 percent of the hydrogen bonds are three-centered (bifurcated donor) bonds (24,25).

For the $\diagdown$NH- - - N bonds that predominate in peptides, proteins and nucleic acids, there are no additional parameters added for each hydrogen bond, since the position of the hydrogen atom is defined by the adjacent non-hydrogen atoms.
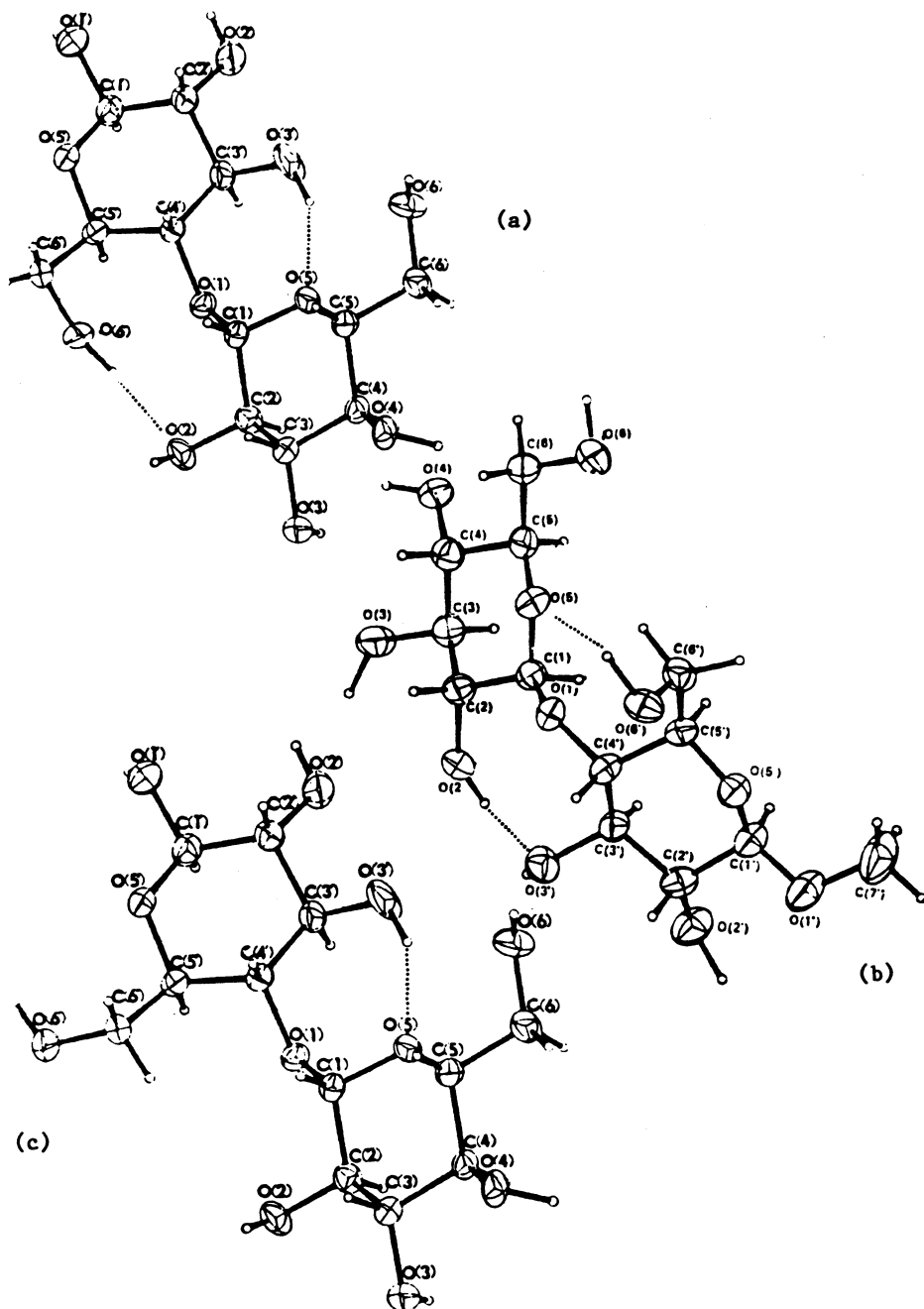
Figure 1. Conformations of cellobiose with inter-residue intramolecular hydrogen bonding. (a,b) conformations with two inter-residue bonds. (c) hydrogen bonding observed in the crystal structure (18).

How is Hydrogen Bonding Treated in Modeling Methods?

Ab-initio calculations include all hydrogen bond interactions. When these are isolated molecule calculations, the models are those with maximum intramolecular hydrogen bond energy. Even if the O–H- - -O angles are close to 90° and the H- - -O distances are greater than 3.0 Å, there will still be an attractive coulombic term which favors an intramolecular hydrogen bond orientation of the OH groups. Such a calculation for glucose, for example, would orient the hydroxyls so as to form a cooperative ring of intramolecular hydrogen bonds around the perifery of the molecule (26). Such weak bonds are pre-empted by the geometrically more favorable intermolecular bonds in the crystal and to solvent molecules in solution. This makes *ab-initio* calculations inappropriate for most biological molecules which exist in environments where the hydrogen bonding is predominantly intermolecular. This can be avoided by fixing the orientation of the hydroxyl groups so that they cannot form intramolecular hydrogen bonds. Then the calculation ceases to be *ab-initio*. It contains some experimental content. The primary use of *ab-initio* calculations for such molecules is likely to be to provide source data for parameterizing molecular mechanics and dynamics programs.

Hydrogen bonding is included in empirical force field calculations in two ways. In the MM series (27), bond dipoles are placed at the centers of the bonds and the Jeans equation is used:

$$\sum_i \sum_j C \, \mu_i \mu_j (\cos X_{ij} - 3 \cos \alpha_i \cos \alpha_j) D . r_{ij}^3 .$$

This may be supplemented with a Morse potential (28) which involves more parameters and a knowledge of the equilibrium hydrogen bond lengths.

It is not surprising therefore that force-fields directed at hydrogen-bonded molecules have favored the simpler charge-charge formula,

$$q_i q_j / \varepsilon r_{ij} \text{ or } q_i q_j / \varepsilon r_{ij}^2$$

if the dielectric constant is made distance-dependent. This term is used alone (29) or with a fine-tuning component of the form

$$\sum_{\text{H-bond}} A r_{ij}^{-12} + B r_{ij}^{-10},$$

which is parameterized for different groups of hydrogen bonds, as in AMBER (30, 31).

The early version of the molecular dynamics program CHARMM (32) also included some rather complex X–H- - -A angle-dependent terms, although there seems to be some uncertainty whether they are really advantageous.

The coulombic term is an atom-pair atomic point charge *model* for the interaction between the electrostatic potentials within a molecule or between molecules. Considerable attention has been paid to deriving "appropriate" values for the $q_i q_j$ parameters (33,34). One of the more sophisticated ways is to use *ab-initio* methods to calculate the molecular electrostatic potential and then least squares fit the atomic point charges to these potentials (35). This method is only applicable to relatively simple molecules, and the point charges derived are quite basis-set dependent, varying as much as 30% between STO-3G and 6-31G*. Unfortunately, atomic point charges are not well-defined physical properties. In the multipole method of experimental charge density analysis they correspond to the monopole populations (36). These populations depend upon the assumption made concerning the radial distribution functions. Within limits any point charge values can be derived by dividing up the electron density distribution in different ways between the atoms. Until there is a better way of distributing electron density between atoms in molecules

(37), the coulombic q_iq_j values, like the ε dielectric constants, must be regarded as adjustable parameters, chosen so as to give the best fit between theory and experiment. Even then, multi-atom and cooperative effects affect hydrogen bond energies by the order of ten percent and need to be included in the calculations (38). This is particularly true for carbohydrates in the solid state and in solution, where finite and infinite chains and three-dimensional nets of hydrogen bonds are energetically favored. Both the dipole-dipole and charge-charge models alternate slowly with distance, as compared with the Lennard-Jones non-bonding interactions. They therefore tend to over-emphasize intramolecular bonding in the same way as *ab-initio* calculations, when applied to isolated molecules.

Some Studies of Molecular Modeling for Hydration

Biochemistry and chemistry takes place mostly in solution or in the presence of large quantities of solvent, as in enzymes. As the necessary super-computing becomes available, molecular dynamics must surely be the method of choice for modeling structure and for interpreting biological interactions. Several attempts have been made to test the capability of molecular dynamics to predict the known water structure in crystalline hydrates. In one of these, three amino acid hydrates were used; serine monohydrate, arginine dihydrate and homoproline monohydrate. The first two analyses were by neutron diffraction, and in the latter X-ray analysis was chosen because there were four molecules and four waters in the asymmetric unit. The results were partially successful, but the final comments of the authors were "this may imply that methods used currently to extract potential function parameters are insufficient to allow us to handle the molecular-level subtleties that are found in aqueous solutions" (39).

In an oligonucleotide-drug hydrate complex, the appearance of a clathrate hydrate-like water structure prompted a molecular dynamics simulation (40). Again the results were only partially successful, prompting the statement, "The predictive value of simulation for use in analysis and interpretation of crystal hydrates remains to be established." However, recent molecular dynamics calculations have been more successful in simulating the water structure in the host lattice of α-cyclodextrin and β-cyclodextrin in the crystal structures of these hydrates (41,42).

Some Suggestions for Modeling Hydrogen Bonding in Carbohydrates

It seems clear that correct parameterization of the electrostatic terms is going to be the most difficult aspect of developing reliable predictions for carbohydrates in the solid state or in solution (43). What appears to be necessary to reduce the multiplicity of global minima are more constraints on the structure of the hydrogen bonding, similar to the constraints on bond lengths and valence angles used by protein crystallographers when refining their X-ray crystal structure analyses. Examples of such constraints are:

(1) Use H-bond length statistics to define the most probable bond lengths and angles as are used in parameterizing the covalent bond structure. A vast source of information on this subject is available in the Cambridge Crystallographic Data Base for all types of organic and organometallic molecules. Hydrogen bonds, like covalent bonds, can be expanded or compressed from the equilibrium value for a particular donor-acceptor pair. Since their force constants are about 15 times weaker, the range of values is much wider than for covalent bonds, of the order of 1 Å, as compared with 0.05 Å for C–C bonds. A similar spread of values is observed with the O–H- - -O angles (44).

(2) Favor cooperative systems of finite or infinite chains of O–H- - -OH- - - OH- - - or homodromic loops (45,46), as in Figures 2, 3 and 4.
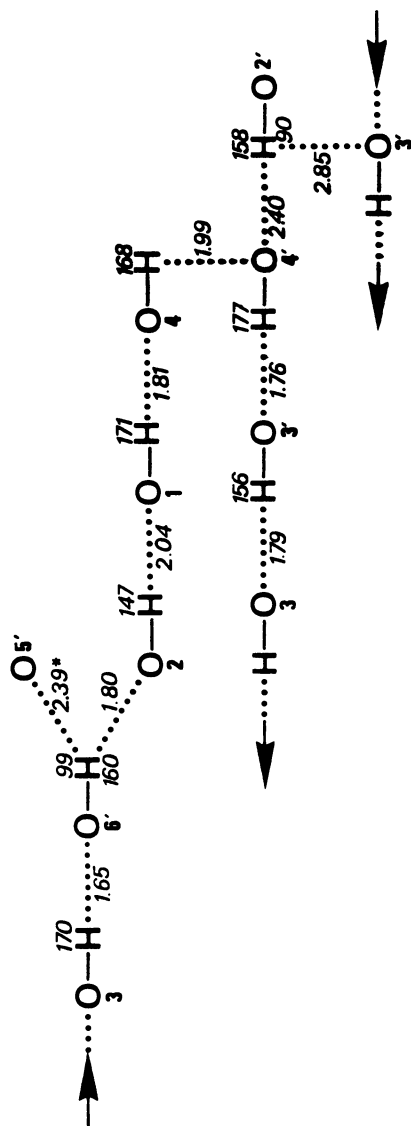
Figure 2. Schematic diagram of the hydrogen-bond structure in the crystal structure of gentiobiose (GENTOS01: REFCODE in Cambridge Crystallographic Data Base). The arrows indicate infinite chains. Distances are H--O in Å, angles are O-H--O in degrees. The covalent O-H bond lengths have been normalized to 0.97 Å.
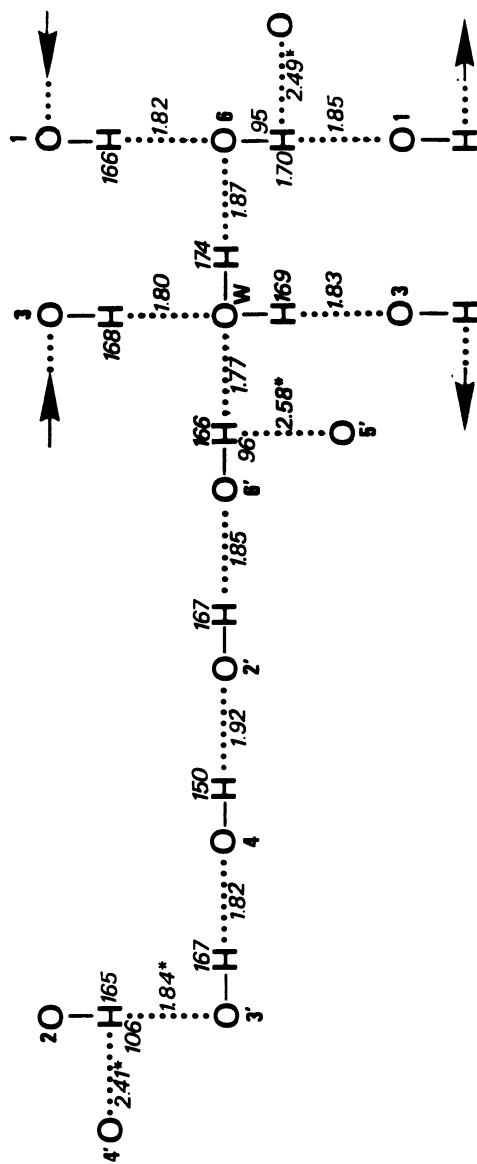
Figure 3. Schematic diagram of the hydrogen-bond structure of β-maltose monohydrate (MALTOS11). The arrows indicate infinite chains. Distances and angles are from the neutron diffraction analysis.
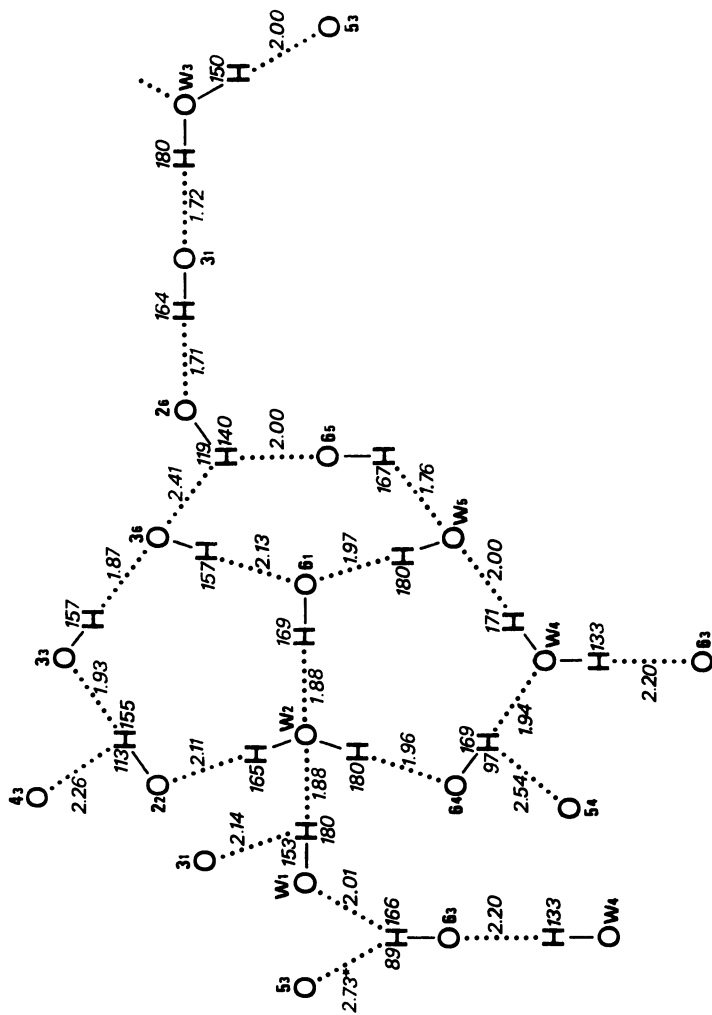
Figure 4. Schematic diagram of section of hydrogen-bond structure containing homodromic cycles in the crystal structure of cyclodextrin hexahydrate.

(3) Favor hydrogen bond structures in which three-center bonds cross-link their chains and loops into nets and thereby enhance cooperativity (47).

(4) Use of the concept of the 'Excluded Region' to limit the orientation of –OH groups and both orientation and translations of water molecules (48).

Programming these constraints will require the same type of approach to the hydrogen bond aspect of molecular modeling as is already being applied to the molecular mechanics of covalent bonding.

Literature Cited

1. Hehre, W. J.; Radom, L.; von Schleyer, P. R; Pople, J. A. Ab-initio Molecular Orbital Theory; John Wiley & Sons: New York, 1986.
2. Malrieu, J. P. In Modern Theoretical Chemistry; Schaefer, H. F., Ed.; Plenum Press: New York, 1977; Vol. 7, Chapter 3, pp 69-104.
3. Burkert, U.; Allinger, N. L. Molecular Mechanics; American Chemical Society: Washington, DC, 1982; Monograph 177.
4. Allen, F. H.; Kennard, O.; Taylor, R. Accts. Chem. Res. 1983, 16, 146-53.
5. Jones, P. J. Chem. Soc. Rev. 1984, 13, 157-72.
6. Jeffrey, G. A.; Pople, J. A.; Radom, L. Carbohydr. Res. 1972, 25, 117-31.
7. Jeffrey, G. A.; Pople, J. A.; Radom, L. Carbohydr. Res. 1974, 38, 81-95.
8. Jeffrey, G. A.; Pople, J. A.; Binkley, J. S.; Vishveshwara, S. J. Amer. Chem. Soc. 1978, 100, 373-79.
9. Lemieux, R. U.; Chu, W. J. Abstr. Papers Am. Chem. Soc. 1958, 133, S1N.
10. Berman, H. M.; Chu, S. S. C.; Jeffrey, G. A. Nature, 1967, 157, 1576-77.
11. Jeffrey, G. A. In Anomeric Effect. Origin and Consequences; Szarek, W. A.; Horton D., Eds.; Am. Chem. Soc. Symposium Series, 1979, No. 87.
12. Lesyng, B.; Saenger, W. Biochem. Biophys. Acta 1981, 678, 408-12.
13. Rees, D. A.; Smith, P. J. C. J. Chem. Soc. Perkin II, 1975, 830-35.
14. Bock, K.; Meldal, M.; Bundle, D. R.; Iversen, T.; Garegg, P. J.; Norbert, T.; Lindberg, A. A.; Svenson, S. B. Carbohydr. Res. 1984, 130, 23-34.
15. Brady, J. W. J. Am. Chem. Soc. 1986, 108, 8153-60.
16. Engler, E. M.; Andose, J. D.; von Schleyer, P. R. J. Am. Chem. Soc. 1973, 95, 8005-25.
17. Jeffrey, G. A.; Taylor, R. J. Computat. Chem. 1980, 1, 99-109.
18. Chu, S. S. C.; Jeffrey, G. A. Acta Crystallogr., Sect. B (1968) 24, 830-838.
19. Melberg, S.; Rasmussen, K. Carbohydr. Res. 1980, 78, 215-24.
20. Rohrer, D. C.; Sarko, A.; Bluhm, T. L.; Lee, Y. N. Acta Crystallogr., Sect. B 1980, 36, 650-54.
21. Yadav, J. S.; Barnickel, G.; Bradaczek, H. J. Theor. Biol. 1982, 95, 151-66.
22. Mo, F.; Jensen, L. H. Acta Crystallogr. 1975, 31, 2867-73.
23. Jeffrey, G. A.; Takagi, S. Accts. Chem. Res. 1978, 11, 264-70.
24. Ceccarelli, C.; Jeffrey, G. A.; Taylor, R. J. Molec. Struct. 1981, 70, 255-71.
25. Jeffrey, G. A.; Maluszynska, H.; Mitra, J. Int. J. Biol. Macromol. 1985, 7, 336-48.
26. Kroon-Batenburg, L. M. J.; Kanter, J. A. Acta Crystallogr., Sect. B 1983, 39, 749-54.
27. Burkert, U.; Allinger, N. L. Molecular Mechanics; Am. Chem. Soc. Monograph, 1982, No. 177.
28. Taylor, R. J. Molec. Struct. 1981, 71, 311-325.
29. Lifson, S.; Hagler, A. T.; Dauber, P. J. Am. Chem. Soc. 1979, 101, 5111-5121.
30. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Propata, S. Jr.; Weiner, P. J. Am. Chem. Soc. 1984, 106, 765-84.

31. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. J. Computat. Chem. 1986, 7, 230-252.
32. Brooks, B. R.; Briccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J. Computat. Chem. 1983, 4, 187-217.
33. Singh, U. C.; Kollman, P. A. J. Computat. Chem. 1984, 5, 129-45.
34. Cox, S. R.; Williams, D. E. J. Computat. Chem. 1981, 2, 304-23.
35. Williams, D. E.; Jan, J. Adv. Atomic Mol. Phys. 1987, 23, 87-129.
36. Spackman, M. A.; Stewart, R. F. In Methods and Applications in Crystallographic Computing; Hall, S. R.; Ashida, T., Eds.; Clarendon Press: Oxford, UK, 1984; pp. 302-20.
37. Bader, R. F. W. Accts. Chem. Res. 1985, 18, 9-15.
38. Barnes, P.; Finney, J. L.; Nicolas, J. D.; Quinn, J. E. Nature 1979, 282, 459-64.
39. Goodfellow, J. M.; Finney, J. L.; Barnes, P. Proc. Roy. Soc. London B 1982, 214, 213-28.
40. Mezei, M.; Beveridge, D. L.; Berman, H. M.; Goodfellow, J. M.; Finney, J. L.; Neidle, S. J. Biomol. Struct. Dynam. 1983, 1, 287-97.
41. Koehler, J. E. H.; Saenger, W.; van Gunsteren, W. F. Eur. Biophys. J. 1987, 15, 197-210.
42. Koehler, J. E. H.; Saenger, W.; van Gunsteren, W. F. Eur. Biophys. J. 1987, 15, 211-24.
43. Wilcox, G. L.; Quiocho, F. A.; Levinthal, C.; Harvey, S. C.; Maggiora, G. M.; McCammon, J. A. J. Computer-Aided Molecular Design 1987, 1, 271-81.
44. Jeffrey, G. A. Landolt-Bornstein. New Series, Group VII, Vol. 1b; Saenger, W., Ed.; Springer-Verlag: Berlin, 1989; Sect. 2.7, pp. 277-348.
45. Jeffrey, G. A.; Mitra, J. Acta Crystallogr., Sect. B 1983, 39, 469-80.
46. Saenger, W. Nature 1979, 279, 343-4.
47. Koehler, J. E. H.; Saenger, W.; van Gunsteren, W. F. J. Biomolec. Struct. & Dnyamics 1988, 1, 181-198.
48. Savage, H.; Finney, J. Nature 1986, 322, 717-20.

Chapter 3

"Modified Neglect of Diatomic Overlap"-type Semiempirical Methods

Kenneth M. Dieter and James J. P. Stewart

Frank J. Seiler Research Laboratory, U.S. Air Force Academy, CO 80840

The current status of the semiempirical methods pioneered by Michael J. S. Dewar is given. These methods are made available to non—theoreticians through the programs MOPAC and AMPAC. Some capabilities of MOPAC and the form of the data input to the program are outlined.

Chemists seeking to use computational chemistry to support experimental efforts now have three general theoretical tools available to them: force field or molecular mechanics models, ab initio molecular orbital (MO) models and semiempirical MO models (1). Each of these tools have strengths and weaknesses which must be evaluated to determine which is most appropriate for a given applications.

At one end of the spectrum are the molecular mechanics techniques, such as AMBER, CHARM, MODEL, and the MM programs, which use classical mechanics relationships to describe interactions between atoms in a chemical system. The associated algorithms include parameters which are optimized to reproduce experimental energies and geometries. These techniques are extremely fast and can accurately calculate energies of ground state systems. Heats of reaction and relative conformational stabilities can be derived for even very large systems such as enzymes. Without a quantum mechanical foundation, however, no information is available on electronic structure or properties dependent on electronic structure. Processes involving bond making/breaking cannot be represented, so full reaction profiles cannot be modeled.

At the other end of the spectrum are the ab initio quantum molecular models, which are rigorous within the Hartree—Fock/ Roothaan—Hall (HF/RH) formalisms. Electronic structure is calculated, and dependent properties are derivable. In theory, full reaction profiles can be modeled. In practice, however, their speed makes it impractical to apply the more accurate

methods to systems larger than, for example, acetic acid. The computing time required for these calculations is on the order of thousands to hundreds of thousands times that required for molecular mechanics calculations, depending on the basis set chosen for the ab initio calculations. There is always the temptation to use a smaller, less adequate basis set or to make simplifying assumptions about geometries or symmetry during a reaction to speed the calculations. Without proper justification, however, these procedures bias the calculations and cast doubt on the conclusions drawn from the study.

Semiempirical MO (SE–MO) calculations lie between these two extremes. Like ab initio models, SE–MO models are quantum mechanical in formalism. Like molecular mechanics, algorithms include parameters optimized to reproduce experimental values of molecular properties. The quantum mechanical foundation enables the calculation of electronic structure and derived properties, while, at the same time, the use of approximate functions for the interaction integrals reduces computing time by three or more orders of magnitude as compared to comparable ab initio calculations. While the computing time is still far too great to allow treatment of systems the size molecular mechanics can model, it is practical to look at real–life systems as opposed to being forced to modeling them with very small molecules. The concern is, of course, that the loss of rigor in the development of the SE–MO models will make them unrealistic. The forms of the parametric functions coupled with the optimization of parameters to reproduce experimental results is intended to compensate for these approximations. Additionally, the calculations are sufficiently inexpensive as to allow for extensive testing of these models to determine the validity of their application to specific studies.

Given, then, that SE–MO models are the only ones sufficiently fast to allow for the study of carbohydrate systems, while at the same time being able to provide information on electronic structure and reaction profiles, the remainder of this chapter will deal with the most commonly used models.

Historical Development

The history of general purpose SE–MO models is virtually totally the history of the work of Dewar and coworkers (2) based on the Intermediate Neglect of Differential Overlap (INDO) (3) and the Neglect of Diatomic Differential Overlap (NDDO) (3) approximations. While Pople and coworkers first developed these approximations to the full HF/RH treatment in the mid–sixties, it remained until 1975 before a model capable of calculating a variety of properties for a wide range of compounds became available. Dewar called this model MINDO/3 (4), for Modified INDO/version 3. It was eventually parametrized for 10 elements, but not in all combinations. A survey of molecules (5) containing C, H, N and O showed that the average error in heats of formation, ΔH_f, was 11 kcal/mol, in bond lengths, 0.022 Å, and in ionization potential, 0.7 eV.

MNDO. Despite its success, Dewar recognized certain weaknesses (6) in MINDO/3 due to the INDO approximation, such as the inability to model lone pair − lone pair interactions. Additionally, due to the use of diatomic parameters in MINDO/3, it was increasingly difficult to extend MINDO/3 to additional elements. Because of this, Dewar began working on a new model based on the better NDDO approximation.

This new model (6), called MNDO for Modified Neglect of Diatomic Overlap, was published by Dewar and Thiel in 1977. With MNDO the average errors (5) for the same survey of C, H, N and O molecules decreased to 6.3 kcal/mol for ΔH_f, 0.014 Å for bond lengths and 0.48 eV for ionization potentials. Since MNDO used only atomic parameters, parameterization of MNDO to include additional elements was much easier than with MINDO/3, and, over the next eight years, parameters were optimized for 16 elements in addition to C, H, N and O.

AM1. While MNDO was widely accepted and extensively used, there were still some deficiencies in the model. In particular, excessive repulsions were observed in MNDO potential energy surfaces just outside chemical bonding distances. This deficiency manifested itself (5,7) in the inability of MNDO to model hydrogen bonding, as well as in large positive errors in the ΔH_f of sterically crowded molecules and in heats of activation. Again Dewar set off to correct this deficiency.

The result of this effort was AM1, for Austin Model 1, published in 1985 (7). This model was essentially a reparameterized version of MNDO with a small change in the core—core repulsion function designed to eliminate the spurious repulsions. With this change and the more effective parameterization possible with a better optimization procedure, average errors in calculated properties again decreased. More importantly, AM1 was able to reproduce hydrogen bonding, making possible for the first time the study of biochemical reactions using these methods. Subsequently, AM1 has been, and still is being, expanded to include additional elements.

PM3. Throughout this work, parameters were optimized for a few elements at a time while fixing all previously optimized parameters. This was largely due to the combination of computer resources and optimization procedures available. Recently a new optimization algorithm was developed which allowed the simultaneous optimization of parameters for a larger number of elements (8). Using this procedure, a new model was parameterized for 12 elements (9). This model is called MNDO—PM3 to indicate it is the third parameterization of MNDO, AM1 being the second. In addition to improving performance for some classes of compounds, such as nitro compounds, which have presented problems for the other models, PM3 appears to model hypervalent compounds more accurately than ever before. The statistical distribution of errors in calculated ΔH_f is also significantly tighter than with the other models in a survey of over 500 compounds. While these results appear promising, PM3 will be validated only through extensive testing and application similar to that experienced by MINDO/3, MNDO and AM1.

Accuracy of NDDO—Based Models

Although both MNDO and AM1 have been used for several years, PM3 is still very new. A brief summary of the level of accuracy to be expected from calculations involving these methods is given in Table I. Note that average errors are given only for those elements for which parameters are available for all three methods.

Computer Programs. As Dewar has often stated (7,10), his whole intention in this work has been to develop a tool fast enough, simple enough and accurate enough for experimental chemists to use as an aid in their own research. A major contribution towards this goal in the last few years has been the combination of separate programs, written by Dewar and his coworkers to

Table I. Average errors in calculated properties

Heats of Formation (kcal/mol)

Type of compounds	No. of compounds	Average errors		
		PM3	MNDO	AM1
Saturated hydrocarbons	24	4.5	7.0	6.7
Hydrocarbons with double bonds	42	2.8	4.1	4.4
Aromatic hydrocarbons	7	4.1	2.7	4.2
Alcohols	7	1.7	5.8	6.8
Ethers	8	4.0	10.1	5.8
Aldehydes and ketones	15	4.6	4.5	4.5
Acids	10	3.1	2.8	4.9
Compounds containing H, C, and O, only	57	4.5	5.6	6.0
Organophosphorus-V compounds	15	10.9	53.9	15.5
Compounds of C, H, N, and O, only	276	5.5	11.2	7.5
Wide range of compounds	713	8.2	22.5	13.8

Molecule Bond Lengths (Å)

Type of bond	No. of bond	Average errors		
		PM3	MNDO	AM1
C-H	51	0.009	0.010	0.014
C-C	72	0.017	0.014	0.017
C-O	18	0.012	0.021	0.030
O-H	8	0.014	0.014	0.012
O-P	3	0.026	0.045	0.044

Table I. Average errors in calculated properties *(Continued)*

Molecule Angles (Degrees)

Type	No.	Average errors		
		PM3	MNDO	AM1
H–C–H	4	1.2	4.8	4.1
H–C–C	19	1.3	1.7	1.3
H–C–O	3	3.3	2.7	2.9
C–C–C	16	2.3	4.1	2.4
C–C–O	4	2.1	0.7	1.0
C–O–H	2	2.9	7.3	2.3
C–C–C–C	7	8.2	22.6	8.8

Dipole Moments (D)

Type	No.	Average errors		
		PM3	MNDO	AM1
Compounds of C, H, N, and O, only	47	0.29	0.32	0.25
Wide range of compounds	132	0.37	0.44	0.35

Ionization Potentials (eV)

Type	No.	Average errors		
		PM3	MNDO	AM1
Compounds of C, H, N, and O, only	101	0.58	0.64	0.48
Wide range of compounds	264	0.56	0.77	0.59

implement these models, into comprehensive, user–friendly programs. There are currently three such programs: MNDO88, developed by Thiel at Wuppertal in West Germany (11); AMPAC (12), under continuing development by Dewar; and MOPAC (13), under continuing development in our laboratory after having its genesis in the Dewar group. The following are some of the major capabilities available within MOPAC and AMPAC. MNDO88 has similar capabilities.

Capabilities of MOPAC

Geometry Optimization. The most commonly used function of SE–MO calculations is the optimization of molecular geometries and the calculation of properties, such as ΔH_f, corresponding to the optimized geometries. The calculation starts with an approximate geometry input by the user using either cartesian coordinates or, more commonly, internal coordinates (bond lengths, bond angles and dihedral angles) to define the geometric points at which atoms are located. The forces acting on the system are calculated and, using this information, the geometry is changed incrementally to reduce the total energy. The geometry is optimized when the energy can no longer be reduced, that is, when the norm of the first derivatives of energy with respect to geometric parameters (gradient norm) is zero. In reality it is extremely difficult to reduce the gradient norm to zero. Additionally, because of the limitations of the quantum mechanical algorithms and the optimization algorithm, as well as the limited precision of the computer, the minima of the gradient norm and the energy will not necessarily coincide. In practice, the difference in geometries is very small; as a result, the gradient norm at the energy minimum will also be small. However, in some cases this can result in an optimization taking an excessively long time. The programs monitor the changes and anticipated changes in the gradient norm and energies and stop the calculations when they fall below preset limits. These limits can be adjusted to loosen or tighten the criteria as required for specific studies. The program then outputs the optimized geometry, ΔH_f, ionization energy based on Koopmans' theorem (14), charge distribution and dipole. Many other properties, such as delocalized and localized MOs, bond orders, etc., can also be output at the user's request.

Reaction Paths. A natural extension to the optimization of geometries and calculation of corresponding properties is the calculation of a reaction profile. A particular geometric coordinate approximating the reaction coordinate is changed incrementally to observe the corresponding change in system energy. For example, in the conversion of hydroxymethylene to formaldehyde, the $O-C-H_1$ bond angle can be fixed at various values ranging from its value in the optimized geometry of hydroxymethylene to that in the optimized geometry of formaldehyde while optimizing all other geometric coordinates.
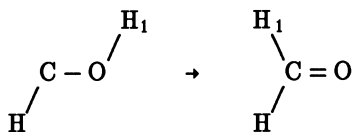
When the corresponding energies for these points are plotted against the

"reaction coordinate", an approximation to the reaction profile is obtained, from which the heat of activation can be inferred.

Transition State Optimization. While the approximate geometry of the transition state (TS) can be obtained from the geometry corresponding to the maximum in a reaction path calculation, the actual reaction coordinate will normally involve more than one internal coordinate. Consequently, further optimization is normally required to determine the actual TS geometry. This geometry will correspond to the point of minimum energy for 3N−5 coordinates and a maximum in energy in the one coordinate leading from the reactant(s) to the product(s). The gradient norm for the system is still zero but the total energy is not at a minimum. Normal geometry optimization will not work because that procedure reduces both the gradient norm and the energy. In other words, using that procedure would result in the reoptimization of the reactants or products, depending upon which side of the actual TS the approximate TS geometry lies. To solve this problem, different procedures are available to optimize the TS geometry by minimizing only the gradient norm. The difference in the ΔH_f of the reactants and the TS is, then, the heat of activation predicted by the SE−MO model.

Force Constant Calculation. The characterization of specific points on the potential energy surface (PES) is in doubt until a force constant calculation is accomplished. The optimized geometry of a stable molecule corresponds to a local minimum on a PES and will, consequently, have all positive force constants. Transition states, on the other hand, will have a single negative force constant corresponding to the movement of the atoms along the reaction coordinate, since energy is at a maximum along that coordinate. It is also possible to "optimize" to a geometry which will have more than one negative force constant. These geometries normally correspond to points of no chemical significance. Consequently, procedures for force constant calculations are provided to allow users to characterize critical points along a reaction profile to ensure proper geometries, and corresponding values of properties of interest, have been calculated. In addition, vibrational frequencies are calculated from the mass weighted force matrix for comparison with experimentally observed spectra.

Carrying out a MOPAC Calculation

The data set. The data requirements for a MOPAC job are highly standardized. The calculation is controlled mainly by keywords. For example, if an unrestricted Hartree−Fock calculation is to be done on an ammonium ion using the AM1 method, then the keywords UHF, CHARGE=1, and AM1 would be used.

Two lines are provided to allow documentation of the system. Typically, this would consist of the chemical name of the system and the reason why the calculation was being done.

This is followed by the geometry, in either cartesian or internal coordinates. Each atom in the system is entered on a separate line. If internal coordinates are used, the atom's position is defined relative to other atoms in terms of a distance and two angles. If Cartesian coordinates are used, the position of each atom is defined relative to some arbitary origin in Cartesian space.

Construction of the coordinate matrix, or Z—matrix, is the most difficult step in carrying out a MOPAC calculation. To make this task easier, special routines and programs have been written. One of these is DRAW.

Computational Requirements. With minor modifications, these programs all run on computers which support FORTRAN—77. With 1 megabyte (Mb) of storage very small molecules can be run using MOPAC. For molecules with 20 − 100 atoms, between 1.6 Mb and 16 Mb of storage are required. Estimates of computer time vary widely depending on the computer used and the nature of the calculation, however an idea of the time needed can be given by an example of a calculation on a VAX 11—780. For cyclohexane, on a VAX 11—780, using C—C distances on 1.5 Å, C—H of 1.1 Å, all angles tetrahedral, and no symmetry used, the time required to optimize the geometry is about 15 minutes. In general, the time required will rise as the cube of the number of atoms.

DRAW. Like MOPAC, DRAW is public domain software. Copies can be obtained from the Quantum Chemistry Program Exchange (QCPE) at the University of Indiana. DRAW allows the data input and output to be represented graphically. Because of the difficulty in assembling a valid Z—matrix, DRAW has been provided with an editor which allows a user to see how the geometry of the system changes as the coordinates are changed.

Other Graphics Programs. Various other programs are commercially available, such as SYBYL from Tripos and CHEM—X from Chemical Design, which, in addition to allowing MOPAC or AMPAC data—files to be easily generated, have the capability of doing a molecular mechanics energy minimization. This ability is very useful for 'cleaning up' user—written geometries prior to carrying out a semiempirical calculation, thereby reducing the time required to produce fully optimized geometries.

Accuracy of Semiempirical Methods

A knowledge of the accuracy, strong points, and weak points, of each method is necessary in order to efficiently carry out computational chemistry research. We will first look at a summary of the three most accurate methods in MOPAC, and then at their strengths and weaknesses.

Summary of MNDO, AM1, and PM3

Heats of Formation. Average errors for representative systems are given in Table I. From this, we see that the ΔH_f of simple organic compounds are predicted on average with chemically useful accuracy. Experimental values of ΔH_f apply to gas phase species at 298K. In order to see how well these methods work in general, the average error in ΔH_f for a wide range of systems, representing the chemistries of H, C, N, O, F, Si, P, S, Cl, Br, and I, are given. These systems include representative anions, cations, and radicals.

Molecular Geometries. Both MM and ab initio methods are more accurate at predicting molecular geometries than semiempirical methods. However, as stated earlier, MM methods are limited in their scope, while ab initio methods are generally too expensive. On average, semiempirical methods can predict bond lengths within about 2%, and angles within 3^0. Many angles are defined by symmetry; these are not included in the analysis.

Dipole moments and Ionization potentials. Semiempirical methods predict dipole moments within about 0.4 D of that observed experimentally. Ionization potentials are not so well predicted, errors averaging about 0.7 eV.

Strengths and Limitations of Specific Methods

Strengths of MNDO. Unlike its precedent, MINDO/3, MNDO (and the later methods) was based on atomic rather than diatomic parameters. This enabled it to be parameterized for a large number of elements. Additionally, MNDO was the first semiempirical method to be based on molecular rather than atomic data. This gave it a generality other methods lacked. Many properties, such as molecular geometry, dipole, polarizability and hyperpolarizability, activation barriers, vibrational frequencies, and first and higher ionization potentials could now be calculated using a single program and a single method.

Limitations of MNDO. From its inception, some important limitations of MNDO were apparent. Sterically crowded molecules were calculated too unstable; for example, the ΔH_f of neopentane is predicted by MNDO to be -24.6 kcal/mol, compared with the observed -40.3 kcal/mol. On the other hand, four—membered rings were predicted to be too stable, this reaching a limit in cubane, which was predicted to be 49.6 kcal/mol too stable. Later on, other limitations were discovered, the most important from a biochemical standpoint being the virtually complete lack of a hydrogen bond. Other deficiencies included the extreme instability of hypervalent molecules. This effectively precluded the application of MNDO to organophosphorus compounds of biologic interest. Finally, activation barriers were predicted to be too high.

Strengths of AM1. AM1 is the first semiempirical NDDO method to accurately reproduce the heat of dimerization of water. This was achieved by modifying the core—core interaction. This change also corrected the instability of sterically crowded molecules, and the excessive stability of four membered rings. Average errors in ΔH_f obtained using AM1 are reduced by almost 40%, relative to MNDO.
Limitations of AM1. AM1 is still a relatively new method. As such, only a few limitations have become apparent thus far. One of the more important of these are the continued inability to adequately represent hypervalent chemistry. Very recently, AM1 parameters for phosphorus and sulfur have become available. Preliminary results indicate these are a considerable improvement over the MNDO method for predicting the ΔH_f of hypervalent compounds. Unfortunately, the phosphorus, parameters include a barrier at 3.0A which complicates the interpretation of reaction mechanisms. For example, in the reaction $PF_3 + F_2 = PF_5$. there are two spurious barriers which must be crossed before the real barrier to the reaction is encountered. In some cases, for example in the reaction $P_2 + P_2 = P_4$, the spurious barrier is significantly higher than the true barrier.

Strengths of PM3. PM3 is the first NDDO method to adequately treat hypervalent systems. This is exemplified by sulfuric acid, for which the PM3 error in ΔH_f is -5.8 Kcal/mol. On the whole, PM3 is more accurate than

either AM1 or MNDO, the average error in ΔH_f being reduced by just over 40% relative to AM1, and 63% relative to MNDO.

Limitations of PM3. At the present time, PM3, the most recent NDDO method, has one severe limitation: the lack of adequate testing. At present only information about ground state systems is available.

    Other limitations include the very different charge distribution compared to that given by MNDO and AM1. Some MM methods use semiempirical charge distributions, with various parameters being based on these charge distributions. Since the PM3 charge distribution is very different from that of MNDO or AM1, MM methods based on these charges will not be compatible with the PM3 charges.

Application to Carbohydrate Chemistry

The authors, not being familiar with carbohydrate chemistry, have limited this to a general discussion of two potential applications of semiempirical methods to the study of carbohydrate chemistry.

    With the advent of PM3, biochemical reactions, for example, those involved in the Embden–Meyerhof pathway, can be studied. Until now, systems such as glucose–6–phosphate were either poorly represented, or were prohibitively slow to calculate.

    As with most other computational methods, care must be exercised in the application of these techniques. Calculations assume isolated molecules, i.e. molecules in a vacuum, at absolute zero. Consequently, although the ΔH_f applies to the system at 298K, kinetic energy is not taken into account. However, calculated activation barriers can be used to predict relative reaction rates at 298K.

    High polymer calculations can be performed on polysaccharides. Calculation of unit cell translation vectors (15), heats of polymerization (15), and elastic moduli (16) can readily be done. The accuracy of such calculations is the same as that of equivalent molecular species. A limitation of elastic moduli calculations is that the polymer is assumed to be 100% ordered, a state not commonly found in polysaccharides.

Summary.

At the present time, it appears that the applicability of semiempirical methods to the study of carbohydrate chemistry has been neglected. Methods are now available for the non–theoretician to investigate molecular systems, reaction mechanisms, and fundamental physical properties, without the need for any extensive knowledge of theoretical methods. Despite this, most computational studies appear to be limited to the use of molecular mechanics techniques. This is probably a natural consequence of the high accuracy and high speed of MM calculations. However, in recent years the ease of use, accuracy, and generality of semiempirical methods have improved considerably. While MM techniques should continue to be used for the study of ground state systems, carbohydrate chemists should be aware of the potential of semiempirical methods as a research tool, particularly for the study of reactions.

Literature Cited
1.   For a recent review of these methods, see Clark, T. <u>A Handbook of</u>
        <u>Computationa lChemistry</u>; John Wiley and Sons: New York, 1985.
2.   For a more e xtensive review of semiempirical methods, see Thiel, W.
        <u>Tetrahedron</u> 1988, <u>44</u>, 7393.
3.   For a good ii troduction to these techniques, see Pople, J. A.;
        Beveridge, D L.   <u>Approximate Molecular Orbital Theory</u>;
        McGraw—Hill: New York, 1970.
4.   Bingham, R, \C.; Dewar, M. J. S.; Lo, D. H. <u>J. Am. Chem. Soc</u>. 1975,
        <u>97</u>, 1285.
5.   Dewar, M. J. S.; Thiel, W. <u>J. Am. Chem. Soc</u>. 1977, <u>99</u>, 4907.
6.   Dewar, M. J. S.; Thiel, W. <u>J. Am. Chem. Soc</u>. 1977, <u>99</u>, 4899.
7.   Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. <u>J. Am.</u>
        <u>Chem. Soc</u>. 1985, <u>107</u>, 3902.
8.   Stewart, J. J. P. <u>J. Comp. Chem</u>. 1989, <u>10</u>, 209.
9.   Stewart, J. J. P. <u>J. Comp. Chem</u>. 1989, <u>10</u>, 221.
10.   Dewar, M. J. S. <u>J. Phys. Chem</u>. 1985, <u>89</u>, 2145.
11.   Thiel, W., Program MNDO88, version 2.1, June 1988.
12.   Available through QCPE, Department of Chemistry, Indiana
        University, Bloomington, IN 47405.
13.   Available through QCPE.
14.   Koopmans, T. <u>Physica</u> (Utrecht) 1934, <u>1</u>, 104.
15.   Stewart, J. J. P. <u>New Polymeric Materials</u> 1987, <u>1</u>, 53.
16.   Klei, H. E.; Stewart, J. J. P. <u>Int. J. Quant. Chem</u>. 1986, <u>20</u>, 529.

Chapter 4

Realistic Conformational Modeling
of Carbohydrates

Applications and Limitations in the Context
of Carbohydrate-High Polymers

David A. Brant and Marie D. Christ

Department of Chemistry, University of California, Irvine, CA 92717

Computer modeling has become a valuable component of studies of
carbohydrate structure-property relationships. When the carbohy-
drate in question is a dissolved oligo- or polysaccharide, successful
application of the technique will require that molecular flexibility be
taken properly into account, and it is likely to require a number of
significant approximations in order to deal with the high dimen-
sionality of the macromolecule-solvent system and the imperfections
in our knowledge of the relevant potential energy functions. In this
paper we examine the statistical mechanical foundations of equilib-
rium conformational modeling in flexible oligo- and polysaccharide
systems and attempt to identify the nature and implications of the
several levels of approximation that are typically invoked. We hope
as well to illustrate that conformational modeling in the carbohydrate
high polymer regime can provide insight and discipline for corre-
sponding studies of the lower molecular weight analogues.

Computer modeling of carbohydrates presents a problem of considerable complex-
ity, particularly when the molecules of interest are oligo- or polysaccharides present
in the predominantly aqueous environment where most carbohydrate macro-
molecules fulfill their native biological function or are employed in commercial
applications. Complexity arises not only from the frequently intricate topology of
the macromolecular carbohydrate structures found in nature but also from the com-
paratively strong, and sometimes long range, intra- and intermolecular interactions
that occur among these highly hydroxylated, and frequently ionic, molecules and
the molecules of the polar solvents in which they dissolve. These circumstances
dictate that practical computer modeling of carbohydrate polymers and oligomers
will depend upon a series of approximations, first, to reduce the dimensionality of
the system to manageable proportions and, second, to provide a tractable represen-
tation of the forces and/or energies of interaction among the constituents of the sys-
tem.

   The initial objective of this paper is to identify, within the context of
macromolecular solution modeling, the typical approximations, explicit and
implicit, and to discuss the possible implications of such approximations. Sec-
ondly, we hope to demonstrate that modeling of carbohydrate high polymer solu-

tion behavior provides, in addition to insights about the molecular basis of the properties of such systems, an excellent arena for highly sensitive tests of conformational potential surfaces or potential functions proposed for modeling of the solution behavior of lower molecular weight oligosaccharides. The discussion will be focused on modeling of those equilibrium observable properties of dissolved oligo- and polysaccharides that reflect the configurational average over all of the conformational states of the flexible carbohydrate subject molecules, each weighted appropriately by its probability of occurrence. Reference to the dynamic, i.e., time dependent, characteristics of such systems will occur as needed to illuminate the discussion of the equilibrium treatment, but no attempt is made here to present a systematic analysis of the methods and approximations attending molecular dynamics simulations of carbohydrates.

Carbohydrate Flexibility

A monosaccharide is flexible by virtue of the rotations of exocyclic groups, i.e., OH and CH_2OH, and its skeletal vibrations, including some relatively low frequency, large amplitude modes associated with the collective motions of the ring atoms.[1] When energy becomes concentrated in the latter modes sufficient to surmount barriers of several kcal/mol, these high amplitude motions can carry a pyranose sugar between the two alternate chair forms or into the boat-skew boat conformational regime, where the six membered ring can execute pseudorotation along a coordinate of relatively low energetic relief.[2] Motions of the latter type are known to make important contributions to the equilibrium properties of some systems, e.g., nucleic acids, containing furanose sugars.[3] The alternate chair and boat-skew boat forms of the pyranose sugars clearly correspond to a major structural departure from the least energetic chair form and, if present in a given carbohydrate, could introduce important contributions to the observable mean properties. They are, however, generally higher in energy than the preferred chair, and there are few documented examples of important contributions to equilibrium properties from these pyranoside ring conformers.[4-6] In some contexts, the exocyclic torsions of the sugar residues are not significant and are justifiably ignored. The overall mean dimensions of a high molecular weight polysaccharide, for example, are not sensitive to the exocyclic hydroxyl orientation. On the other hand, the fit of a sugar epitope to its binding site on an antibody protein may well depend on exocyclic hydroxyl orientation and even, perhaps, on the induction of a particular subgroup of sugar ring conformations accessible through excursions along the relatively soft, low frequency skeletal normal coordinates. In nearly all cases high frequency vibrational motions, corresponding primarily to bond stretching, will make only small contributions to the equilibrium properties of interest.

When consideration turns to disaccharides or higher oligomers, molecular flexibility becomes an important consideration in almost every context, because rotations about the chemical bonds of the glycosidic linkage can occur to produce large amplitude structural variations for energy expenditures of kT or less, which are readily available from the random thermal motions of the system. For high molecular weight polysaccharides molecular flexibility becomes the dominant issue; even when the local torsional mobility at the individual glycosidic linkages is highly restricted, as is often the case, this limited local variability accumulates and is amplified over the many such linkages of a high polymer to generate the host of conformationally different, but nearly equienergetic, chain geometries that determine the observable mean properties of the system. The longer the chain segment, the more important this effect. Significantly, the influence of such torsional flexibility on molecular properties does not vanish as the degree of polymerization declines, and

it is frequently of importance for the properties, observable or otherwise, of carbohydrate oligomers with as few as two sugars.

Conformational Averaging

Computer modeling of carbohydrate solution properties thus requires in most cases averaging over the available conformational states of the carbohydrate molecule in order to make a quantitative and conceptually accurate connection between carbohydrate chemical structure, molecular environment, and the physical and biological behavior of the system. Averaging over the states of the solvent is also required, as we shall shortly see. For nearly every case of conceivable interest the methods of *classical* statistical mechanics will suffice for this purpose. This implies neglect of the quantized character of certain vibrational modes, i.e., those dominated by bond stretching and stiff valence angle bending, of high enough frequency so as not to be fully excited at the temperatures of interest. This is an approximation entirely justified by the virtual independence of such modes of the macromolecular conformation or environment.[7]

   A brief discussion of the essential operations is required in order to identify and evaluate the possible approximations. If we let $\{p\} = p_1, p_2, p_3, ... p_{3N}$ and $\{q\} = q_1, q_2, q_3, ... q_{3N}$ represent the sets of conjugate cartesian momenta and coordinates, respectively, for all $N = N_c + N_s$ atoms of the system, N_c comprising a single subject carbohydrate molecule and N_s comprising the solvent and other components of the medium, then the classical phase integral (canonical partition function) for the system at fixed temperature (T) and volume (V) may be expressed by

$$Q \sim \int ... \int exp[-\beta H(\{p,q\})]d\{p\}d\{q\} \qquad (1)$$

where, as usual, $\beta = (kT)^{-1}$, $H(\{p,q\})$ is the classical Hamiltonian,

$$H(\{p,q\}) = K(\{p\}) + U(\{q\}) \qquad (2)$$

and the integral is over the 3N momenta and 3N coordinates of the N atoms referred to some laboratory based coordinate system.[8] Here $d\{p\}$ represents the product $dp_1 dp_2 dp_3...dp_{3N}$, and $d\{q\}$ is similarly defined. Only the nuclear coordinates are included as justified by the Born-Oppenheimer approximation. Factoring the kinetic energy terms, $exp[-\beta K(\{p\})]$, each dependent on a single cartesian momentum, and integrating over all 3N momenta in the usual fashion[9] yields, without approximation,

$$Q = \frac{Z_N}{\prod_{i=1}^{N}\Lambda_i^3} \qquad (3)$$

where each factor Λ_i depends on T but not on $\{q\}$. Here Z_N is the classical configuration integral

$$Z_N = \int \ldots \int_V \exp[-\beta U(\{q\})] d\{q\} \tag{4}$$

which requires for its evaluation a specification of the potential energy function $U(\{q\})$ for the system comprising the carbohydrate molecule and its solvent medium and integration over the fixed volume of the system V. Ignoring the weak temperature dependence of the Λ_i, Z_N is routinely used to compute average properties of the system (see below).[7,8]

It is convenient at this point to transform the 3N laboratory based cartesian coordinates $\{q\}$ to $3N_c-6$ internal coordinates $\{q_{int}\} = \{b,\tau,\theta\}$, which describe the *relative* positions of the carbohydrate atoms, and $3N_s + 6$ external coordinates $\{q_{ext}\}$, $3N_s$ of which locate the solvent, i.e., medium, molecules in the laboratory based system and 6 of which describe the position and general orientation of the carbohydrate molecule in the laboratory system. The internal coordinates $\{b,\tau,\theta\}$ are often conveniently chosen as a non-redundant set of spherical polar coordinates, i.e., bond lengths $\{b\}$, valence angles $\{\tau\}$, and bond torsion angles $\{\theta\}$. Following this transformation, the configuration integral is expressed by

$$Z_N = \int \ldots \int_V \exp[-\beta U(\{q_{int},q_{ext}\})] J_{int} d\{q_{int}\} J_{ext} d\{q_{ext}\} \tag{5}$$

where $U(\{q_{int},q_{ext}\})$ describes the potential energy of interactions (configuration energy) among all of the molecules present, carbohydrate and solvent. The factor J_{ext} is the Jacobian of the transformation needed to generate a proper volume element for integrating the angular components of $\{q_{ext}\}$ over the volume of the system; J_{int} serves the same purpose for the internal coordinates.

The Potential of Mean Force

Most "molecular mechanics" expressions for the potential energy U of a molecule contain no explicit terms in $\{q_{ext}\}$, and these are frequently described as "vacuum" potentials, implying rigorous applicability of the potential only in the absence of the numerous and, hence, troublesome, solvent molecules.[8,10] A typical such function is shown in eqn. 6

$$U(\{b,\tau,\theta\}) = \sum_j K_j^b (b_j - b_j^0)^2 + \sum_k K_k^\tau (\tau_k - \tau_k^0)^2 + \sum_l K_l^\theta (1 + \cos(n_l \theta_l))$$

$$+ \sum_{m,n} \left(\frac{A_{m,n}}{r_{m,n}^{12}} - \frac{C_{m,n}}{r_{m,n}^6} + \frac{q_m q_n}{D r_{m,n}} \right) \tag{6}$$

where terms expressing contributions to U from distortions of bond lengths b_j, valence angles τ_k, and torsion angles θ_l from their preferred values are readily identified from their arguments, and the additional terms, expressed explicitly as functions of interatomic separations $r_{m,n}$, reflect contributions from pairwise van der Waals and electrostatic interactions between atoms m and n. Additional terms to

account explicitly for hydrogen bonded interactions have sometimes been included in eqn. (6) but are generally regarded at present as unnecessary.[10]

Difficulties associated with specification of the dependence of U on $\{q_{ext}\}$, or of integrating over the many dimensions of $\{q_{ext}\}$ if the $\{q_{ext}\}$ dependence is explicit, can be circumvented in an alternative fashion by purely formal integration over the external coordinates for fixed $\{q_{int}\}$ to yield[8]

$$Z_N = C_{ext} \int \ldots \int_V \exp[-\beta U'(\{q_{int}\})]J_{int}d\{q_{int}\} \tag{7}$$

where C_{ext} is a constant, VN_s^{+2}, required to maintain the correct dimensions of Z_N, and the temperature dependent quantity $U'(\{q_{int}\})$ is a *free energy* defined by

$$U'(\{q_{int}\}) = -\beta^{-1}\ln\left(\frac{\int \ldots \int_V \exp[-\beta U(\{q_{int},q_{ext}\})]J_{ext}d\{q_{ext}\}}{C_{ext}}\right) \tag{8}$$

The configuration integral of equation 7 is seen to involve only the internal coordinates, or *conformation,* of the macromolecule. Contributions to Z_N from interactions of solvent molecules with one another and with the carbohydrate molecule have been taken into account formally in the *potential of mean force* $U'(\{q_{int}\})$, which describes for each conformation specified by $\{b,\tau,\theta\}$ the effective potential energy of the carbohydrate averaged over all positions, orientations, and corresponding interactions within the volume V of the solvent molecules and the carbohydrate molecule. In keeping with the usual relationship of potential energy and force, the slope of $U'(\{q_{int}\})$ at a point in conformation space specified by $\{q_{int}\}$ along any one of its independent variables q_{int} describes the mean force encountered by the molecule for displacement of the conformation along that coordinate.[8]

It is clear that the function $U'(\{q_{int}\})$ may be approximated by an expression of the form of eqn. (6). Whether a potential of this form, involving no explicit description of the solvent, is appropriate depends on the relative relaxation rates of the solvent motions and the macromolecular intramolecular coordinates. For the slow, conformationally most significant, glycosidic and exocyclic bond rotations of the carbohydrate it is apparent that averaging of solvent motions can occur easily on the time scale of these torsions. It is more difficult, however, to know how much important conformational detail is submerged by the averaging process.

Obviously, whether a molecular mechanics potential $U(\{q_{int}\})$ in the form of eqn. (6) is appropriately labeled a "vacuum" potential will depend on how the parameters governing the terms of the function are established. If data used to parameterize $U(\{q_{int}\})$ come from condensed phase experiments, as is often the case, it is probably more appropriate to regard the potential as a potential of mean force $U'(\{q_{int}\})$ which reflects, albeit in some non-specific way, the influence of the medium on the conformational energy surface. Alternatively, explicit representations of solvent effects have sometimes been incorporated through continuum treatments of the solvation.[11] Most encouraging in this regard is the prospect that potentials $U'(\{q_{int}\})$ may be calculable in the near future using molecular dynamics (or Monte Carlo) simulations to compute $U'(\{q_{int}\})$ from $U(\{q_{int},q_{ext}\})$ for systems in which the carbohydrate is constrained to fixed values of the internal coordinates. This would provide a conceptually more satisfying version of $U'(\{q_{int}\})$ for addressing the properties of large, flexible oligo- and polysaccharides.

Problems associated with the high dimensionality of many carbohydrate systems of interest do not stem solely from the large number of solvent molecules; the atoms of biologically and chemically interesting oligo- and polysaccharides can also be very numerous, and the number of variables in the set $\{b,\tau,\theta\}$ can easily exceed several thousand. Further reductions in dimensionality thus have some appeal. It has been common to reduce the dimensionality of $U'(\{b,\tau,\theta\})$ by recognizing that energetically accessible variations of bond lengths $\{b\}$ have little influence on the molecular properties of interest, and these parameters have often been fixed at appropriate mean values.[8] On the other hand, torsions, especially about the bonds of the glycosidic linkage, produce large amplitude motions for a small expenditure of energy, and these motions must be included *explicitly* in any averaging process. If exocyclic torsions are unimportant for the problem at hand, they are often ignored completely, which is equivalent, for example, to collapsing the exocyclic hydroxyl hydrogens into the corresponding hydroxyl oxygens.

Valence angle variation, and highly hindered "improper" torsions about the endocyclic bonds of the pyranose chair, are of intermediate consequence. Although the amplitudes of intramolecular displacements resulting *directly* from the energetically accessible distortions of valence angles $\{\tau\}$ and improper rotations $\{\theta_h\}$ away from their least energetic values may be small enough to ignore, the strong coupling between hindrances to glycosidic torsional variation and the valence angles and improper torsions renders the fixing of valence and improper torsional angles at some mean values a procedure of questionable merit.[12,13] Nevertheless, in order to progress without excessive computational effort, it has been common to fix at their mean values (or, in the case of exocyclic torsions, to ignore altogether) all structural parameters from the set $\{b,\tau,\theta\}$ except the "soft" torsions $\{\theta_s\}$ of the glycosidic linkages.[8]

The configuration integral and all conformational averages are then evaluated using

$$Z_N \sim \int \ldots \int \exp\,[-\beta V(\{\theta_s\})]\,d\{\theta_s\} \qquad (9)$$

Proportionality factors, including the mean value J_{int} which is accordingly moved outside the integral sign,[14] do not influence mean properties calculated with Z_N and are omitted here to leave only the essential features. Following reasoning akin to that leading to eqns. 7 and 8, the quantity $V(\{\theta_s\})$ is then properly understood as the potential of mean force for the carbohydrate molecule, averaged, with proper weightings, over all intermolecular interactions *and* likewise over all variations of the "hard" internal variables $\{b,\tau,\theta_h\}$ for each conformation specified by $\{\theta_s\}$. Coupling of $\{\tau,\theta_h\}$ with $\{\theta_s\}$ dictates that no single choice of mean values for the hard variables is likely to produce a good approximation to the $V(\{\theta_s\})$ surface. Thus, as changes in $\{\theta_s\}$ occur to yield steep increases in energy due to repulsive interactions of certain atoms, small and energetically inexpensive adjustments in $\{\tau,\theta_h\}$ can often occur to relieve the repulsions precipitated by these changes in $\{\theta_s\}$.

With the advent of molecular dynamics simulations applied to carbohydrates, one can anticipate the direct computation of more conceptually appealing surfaces of V in $\{\theta_s\}$ space from a given $U(\{q_{int},q_{ext}\})$ in the near future. Monte Carlo integration over $\{q_{ext}\}$ and $\{b,\tau,\theta_h\}$ for fixed $\{\theta_s\}$ provides an alternative procedure, but one which is probably less attractive in terms of efficiency than the molecular dynamics approach. A second alternative, known as adiabatic mapping, provides an approximation to $V(\{\theta_s\})$, and applications of this method to carbohydrates have recently begun to appear.[12,13] In this approach the conformational

energy of the carbohydrate is minimized, given some representation of $U'(\{q_{int}\})$, for each selected grid point in $\{\theta_s\}$ space; so far no explicit incorporation of the solvent molecules has been reported. The adiabatic map thus represents the minimum potential energy of the carbohydrate for each point in $\{\theta_s\}$ space rather than the temperature dependent mean or free energy required in Z_N. To the extent that the function $U'(\{q_{int}\})$ is steep for all variations of $\{b,\tau,\theta_h\}$ the adiabatic map should represent a good approximation to $V(\{\theta_s\})$.

<u>Separability of the Energy</u>

Reduction of the potential energy $U(\{q_{int},q_{ext}\})$, involving simultaneously the coordinates of all atoms of the solvent and the carbohydrate solute, to the potential of mean force or *conformational energy* $V(\{\theta_s\})$ in the lower dimensional space of the linkage torsion angles of an oligo- or polysaccharide is clearly advantageous in helping to grasp the relationship of carbohydrate chemical structure and the properties of interest. For high polymers, and even for many oligosaccharides of biological interest, e.g., the oligosaccharide moieties of glycoproteins, the dimensionality of $V(\{\theta_s\})$ remains high enough to prevent ready conceptualization of the features of the conformational energy surface.

Further simplifications can sometimes be effected by representing $V(\{\theta_s\})$ as a sum of contributions $V_i(\{\theta_s(i)\})$ from each of the several sugar residues of the carbohydrate.

$$V(\{\theta_s\}) = \sum_{i=1}^{n} V_i(\{\theta_s(i)\}) \qquad (10)$$

Here the notation $V_i(\{\theta_s(i)\})$ is meant to imply that whereas the contribution V_i of residue i to the total conformational energy of the carbohydrate *may* depend on all of the variables in the set $\{\theta_s\}$, it need not do so and may indeed depend only on some limited number of them. The partitioning in eqn. (10) yields the greatest simplification if the residue contribution V_i depends upon a set of variables $\{\theta_s(i)\}$ upon which no other residue contribution V_j depends. Stated more directly, there is in this case no intersection of the coordinate sets $\{\theta_s(i)\}$ and $\{\theta_s(j)\}$ for any pair of residues i and j, $\{\theta_s(i)\} \cap \{\theta_s(j)\} = \phi$.

In the simplest case, frequently assumed (explicitly or implicitly) to be valid for linear (unbranched) oligo- and polysaccharides, the contribution V_i of residue i depends only on the torsion angles (ϕ_i, ψ_i or ϕ_i, ψ_i, ω_i) of the glycosidic linkage with its successor in the chain. (See Figure 1 for definitions.) This is equivalent to saying that the conformational energy of residue i arises only from its interactions with residue i + 1; the interaction energy of i with i - 1 is attributed in this bookkeeping scheme, without further approximation, to residue i - 1. This is frequently a good approximation, because pyranose sugars are relatively rigid and the torsion angles of the glycosidic linkages are usually highly constrained. This circumstance often greatly reduces the possibility for interactions between second and more distant neighbors in the primary residue sequence. When this approximation of *independent*, first-neighbor interactions is justified, conceptualization of the conformational energy of a carbohydrate of any size reduces to a consideration of only those conformational energy surfaces for the several kinds of dimeric (disaccharide) segments occurring in the molecule.

An example of such a surface is shown in Figure 2, which presents the "Ramachandran map" for the disaccharide segment, cellobiose, of cellulose. In this case, since cellulose is homopolymeric, Figure 2 contains all of the energetic infor-
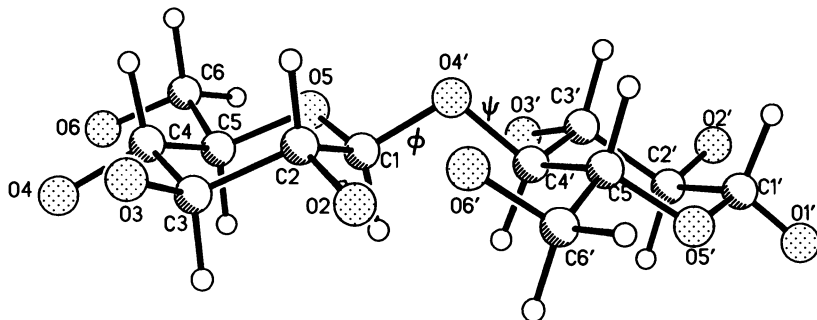
Figure 1. Ball and stick drawing of 4-O-(β-D-glucopyranosyl)-β-D-glucose ("β-cellobiose"). Glycosidic linkage torsion angles φ = θ[H1, C1, O4', C4'] and ψ = θ [C1, O4', C4', H4'] are shown in their respective reference conformations at φ = ψ = 0; the positive sense of rotation about a given chemical bond corresponds to clockwise motion of the several atoms pendant on the atom defining the remote end of that bond when the pendant atoms are viewed along the bond.[8] For (1→6)-linkages rotation about the bond C5'-C6' is described by the torsional angle ω.
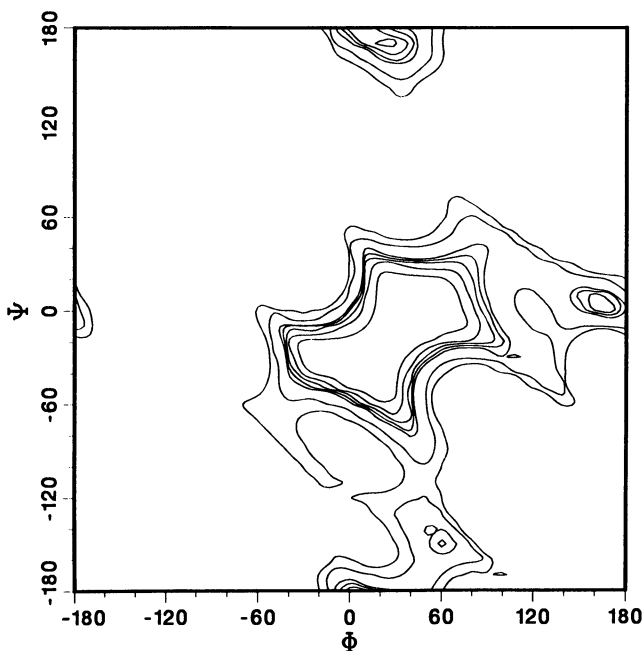
Figure 2. A contour diagram of the conformational energy of β-cellobiose computed from eqn. (6) holding constant all variables except φ, ψ; see ref. 5 for details. The rigid glucose residue geometry was taken from ref. 23, and the valence angle β at O4' was chosen as 116° in accordance with the results of pertinent crystal structure determinations. Contours are drawn at 2, 4, 6, 8, 10, 25, and 50 kcal/mol above the absolute minimum located near φ, ψ = -20°, -30°; higher energy contours are omitted.

mation needed to characterize the total conformational energy of a dissolved cellulosic chain of any length. Discussions to follow of cellulosic and amylosic chains are based on the simplifications inherent in the "independent residue" or "independent linkage" approximation. Although we shall not pursue these matters here, it should be remarked that the independent unit implicit in the statement $\{\theta_s(i)\} \cap \{\theta_s(j)\} = \phi$ need not be a single residue (linkage), and even if independence of structural units cannot be assumed, there remain effective general methods for carrying out the required conformational averaging.[7,15]

A little thought will reveal, in fact, that for flexible high polymer chains the conformational energy can never be truly separable into independent contributions. Because residue i always retains the possibility of interaction with other residues more or less remote in the primary sequence, $\{\theta_s(i)\}$ must include all of the glycosidic torsion angles intervening between residue i and sequentially remote interaction partners. This situation, which clearly precludes $\{\theta_s(i)\} \cap \{\theta_s(j)\} = \phi$ in the strict sense, can be dealt with by partitioning the interactions of residue i into short and long range categories after Flory,[16] who showed that the mean properties of the chain in a suitable "θ-solvent" will be "unperturbed" by the interactions of long range in the chain sequence. Thus, comparisons of mean properties of high polymers computed using classical statistical mechanical methods coupled with "short range" conformational energies $V_i(\{\theta_s(i)\})$, e.g., $V_i(\phi_i, \psi_i)$, may be compared with experimental results gathered under or corrected to the unperturbed state. If attention is confined to stiff and/or short chain segments, long range interactions are unimportant, and the distinction between perturbed and unperturbed molecules vanishes in practice. Explicit statistical mechanical treatment of the long range interactions remains a formidable problem.[17,18]

The independent residue approximation also may not provide an accurate representation of the short range interactions. It will certainly fail for densely branched oligosaccharides of the sort found linked through asparagine to glycoproteins. For linear chains it can be expected to fail when $(1 \rightarrow 2)$ linkages occur, because these tend to provoke interactions between second and higher neighbor residues in the sequence. Even $(1 \rightarrow 4)$ linkages can be problematic when, as in amylose, their repetition leads to a high propensity for self intersection of the chain with a periodicity of 6-7 residues (see below).

Computation of Average Properties

Given an approximation to $V(\{\theta_s\})$ acceptable for the purposes at hand, one can proceed to compute equilibrium, i.e., statistically mechanically averaged, values <P> for properties $P(\{\theta_s\})$ of interest using standard procedures which weight each conformation of the carbohydrate molecule by the Boltzmann factor of $V(\{\theta_s\})$ normalized by the configuration integral given in eqn. (9).

$$<P> = Z_N^{-1} \int \ldots \int P(\{\theta_s\}) \exp[-\beta V(\{\theta_s\})] d\{\theta_s\} \tag{11}$$

Execution of the mathematical operations specified by eqn. (11), which will not concern us here, depends on the details of the dependence of the functions $P(\{\theta_s\})$ and $V(\{\theta_s\})$ on $\{\theta_s\}$.[7] For many properties $P(\{\theta_s\})$ of interest the computation for carbohydrates is straightforward and simple, especially if $V(\{\theta_s\})$ is separable into *independent* terms $V_i(\{\theta_s(i)\})$.[15]

The points we wish to illustrate here do not require actual computations with eqn. (11) and can be made in terms of three simple geometrical properties $P(\{\theta_s\})$: These are (1) the *vector* distance **r** between two atoms of the molecule and its

magnitude r, (2) the square of this distance r^2, and (3) the inverse 6th power of this distance, r^{-6}. The vector **r** could connect the beginning and end of a linear polymer chain; in this case <**r**> is known as the end-to-end or persistence vector of the chain and its scalar magnitude |<**r**>| is the measurable persistence length. Alternatively, **r** could be the vector between any pair of carbohydrate hydrogens related by dipolar or scalar coupling in an NMR experiment, and |<**r**>| the mean separation. The mean square end-to-end length <r^2> is another measurable property of a high polymer chain, while the quantity <r^2>$^{1/2}$ could be used as an alternative measure of the mean separation of a pair of hydrogens in a carbohydrate of lower molecular weight. Interest in <r^{-6}> might stem from a desire to compute the contribution to the nuclear Overhauser effect at one hydrogen arising from a second hydrogen separated in space by the fluctuating vector **r**.

Consider first the substitution of **r** for $P(\{\theta_s\})$ in eqn. (11). The mean vector property <**r**> is computed by averaging each scalar component of the vector separately. Imagine that **r** connects atoms W and Z in Figure 3, that atoms W, X, Y, and Z are connected by bonds of fixed length joined at fixed valence angles, that atoms W, X, and Y are confined to fixed positions in the plane of the paper, and that torsional rotation θ occurs about the X-Y bond which allows Z to move on the circular path depicted. If the rotation θ is "free" such that the potential energy is constant for all values of θ, then all points on the circular locus are equally probable, and the *mean* position of Z, i.e., the terminus of <**r**>, lies at point z. The mean vector <**r**> would terminate at z for any potential function symmetric in θ; for any potential function at all, except one that allows absolutely no rotational motion, the vector <**r**> will terminate at a point that is not on the circle. Thus, the mean position of Z as seen from W is not any one of the positions that Z can actually adopt, and, while the magnitude |<**r**>| may correspond to some separation that W and Z can in fact achieve, it is incorrect to attribute the separation to any real conformation of the entity W-X-Y-Z. Mean conformations that would place Z at a position z relative to the fixed positions of W, X, and Y have been called "virtual" conformations.[19,20] It is clear that such conformations can never be identified with any conformation that the molecule can actually adopt.

One might ask whether the distance <r^2>$^{1/2}$ is subject to a more straight-forward interpretation than |<**r**>|. Clearly, for any distribution of conformations accessible to a molecule <r^2>$^{1/2} \geq$ |<**r**>|, and the equality holds only for the trivial case of a single accessible conformation. Computations of either quantity involve a process of averaging trigonometric functions that has the effect of distorting space in a non-orthogonal way.[21] It is therefore no more correct to identify the computed quantity <r^2>$^{1/2}$ with a distance generated by any real conformation than it is to do so for |<**r**>|. Conversely, any interatomic distance derived from an experiment that observes the subject molecules in an ensemble of conformations should not be identified with a distance that occurs in a particular conformation of the real molecule. Obviously, as the structural diversity of the observed range of conformers diminishes, the force of this enjoinder likewise declines.

In some cases it may be difficult to assess the structural diversity of the observed molecule. Consider, for example, replacement of $P(\{\theta_s\})$ in eqn. (11) by r^{-6} as required to compute contributions to the nuclear Overhauser effect. Because of the resulting strong dependence of the integrand on inverse separation distance, one notes that only those conformations corresponding to small values of r will contribute significantly to the integral regardless of the dependence of the weighting factor $Z^{-1}\exp\{-\beta V(\{\theta_s\})\}$ on conformation. Thus, there may be a wide range of conformations accessible to the molecule which contribute little or nothing to <P>. Attempts to reproduce measured NOE's using eqn. (11) may be quite as successful using an incorrect potential surface implying strong conformational constraint as with a more accurate one that permits conformational excursions into domains that,

by virtue of large values of r^6, make little contribution to the NOE. Stated alternatively, the NOE experiment does not probe conformation space in a uniform fashion, and hence, may provide questionable tests of a proposed conformational energy surface.

Testing Conformational Energy Surfaces

It is a principal objective of this paper to illustrate that many proposed carbohydrate conformational energy surfaces can be put to a demanding test in the high polymer context, because the observable and calculable solution properties of high polymers are particularly sensitive to the details of the conformational energy surface $V(\{\theta_s\})$. Attention will focus initially on the solution behavior of cellulosic chains for which the dimeric segment and a corresponding conformational energy surface, calculated by fixing valence lengths and angles at reasonable mean values and ignoring any possible substituents, are shown in Figures 1 and 2, respectively. Consider now the computation, using eqn. (11), of the characteristic ratio

$$C_\infty = \underset{n \to \infty}{\text{Lim}} \langle r^2 \rangle_0 / nL^2$$

for a dissolved cellulose ester or ether.[4,5,15] Here n is the number of glucose residues in the chain, L is the length of the glycosidic oxygen-to-glycosidic oxygen virtual bond vector spanning one glucose residue, and C_∞ is thus seen to be directly proportional to the unperturbed mean square end-to-end distance $\langle r^2 \rangle_0$, i.e., computed using only short range contributions to $V_i(\{\phi_s(i)\})$, namely $V_i(\phi_i, \psi_i)$. Figure 4 shows that the computed value of C_∞ is quite sensitive to the choice of mean valence angle β at the oxygen of the glycosidic bridge. The computed temperature dependence of C_∞ is also shown. For comparison, the experimental value of C_∞ for cellulose triacetate in two different Θ-solvents is also given as a function of temperature.[22] Selected numerical results are shown in Table I.

Table I.   Measured and Computed Room Temperature Characteristic Ratio and Temperature Coefficients for Cellulosic and Amylosic Chains

Cellulosic Chains	C_∞	$d\ln C_\infty / dT, K^{-1}$
experimental[22]	36	-0.0075
rigid residue model, $\beta = 116°$	107	-0.0017
rigid residue model, $\beta = 120°$	75	-0.0031
rigid residue model, $\beta = 124°$	32	-0.0034
relaxed residue model	11	-0.0026
Amylosic Chains		
experimental[21]	5	-0.0065
rigid residue model, $\beta = 115°$	4.5	-0.0065
relaxed residue model	3	-0.0010

Sensitivity of the computed C_∞ to the valence angle β can be understood by viewing the conformational energy surfaces $V_i(\phi_i, \psi_i)$ for several values of β. Figure 2 shows a cellobiose map computed with β fixed at 116°, a value near the mean glycosidic valence angle drawn from appropriate crystallographic studies.[23] The glucose rings are constrained to remain in the favored $^4C_1$ ring conformation.[4,5] This map, computed with all conformational variables held constant except for ϕ and ψ, discloses a relatively small fraction of the ϕ, ψ conformation space accessible
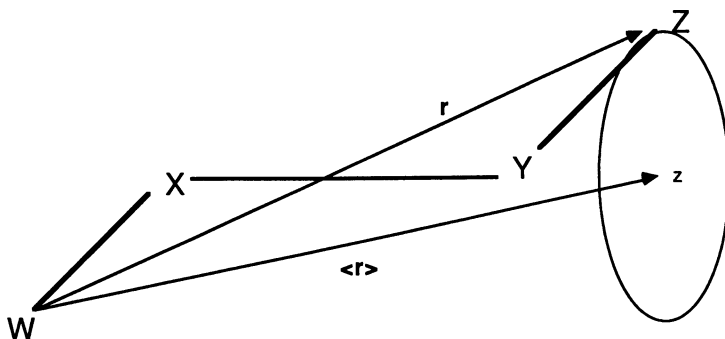
Figure 3. "Molecule" W-X-Y-Z subject to internal rotation along the torsional coordinate θ. The vector **r** connects atoms W and Z. The *mean* vector <**r**> terminates at z for rotation along θ subject to any hindrance potential symmetric in θ.
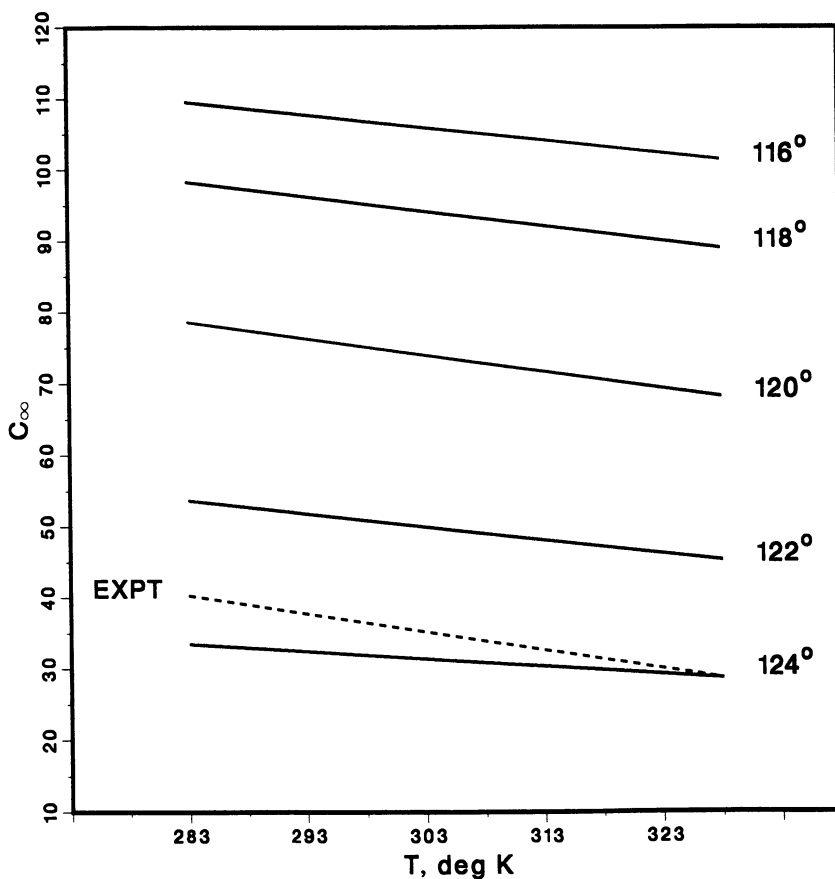
Figure 4. The characteristic ratio C_∞ calculated[5] as a function of temperature T for values of the glycosidic valence angle $116° \leq \beta \leq 124°$. The dashed curve describes experimental results for cellulose triacetate.[22]

to the dimeric segment at ordinary temperatures. This feature, coupled with the geometric characteristics of the β-(1→4)-linkage, produces a computed value of C_∞, ~100, that is too large compared to the room temperature experimental result, ~35. Opening of the valence angle to about 124° causes C_∞ to fall into the experimental range. The reduction in C_∞ is a clear consequence of the greatly increased conformational freedom of the disaccharide unit that arises from opening up the bridge angle; see Figure 5, which was computed for $\beta = 124°$. Greater local conformational freedom results in greater flexibility and tortuosity of the chain and, consequently, a smaller mean square end-to-end distance. Notice, however, that the less conformationally restrictive conformational energy surface, Figure 5, although yielding a reasonable value of C_∞, fails seriously to reproduce the observed negative temperature coefficient (Table I). The temperature coefficient provides an especially demanding test of the model, in particular, of the conformational energy surface, because it reflects the detailed shape of the surface in the vicinity of its least energetic regions.

Given the need to choose an unrealistically large value of β in order to obtain an energy surface for cellobiose open enough to reproduce C_∞ (if not $dlnC_\infty/dT$), one is immediately attracted to the possibility that the fully relaxed (or adiabatic) cellobiose surface recently published by French[12] might successfully explain cellulosic solution behavior. The adiabatic surface, by virtue of its allowance for relaxation of bond lengths and angles for each choice of ϕ, ψ, involves the relief of repulsive nonbonded interactions of the sort that tightly constrain the low energy domains of Figure 2. Hence it is much more open as can be seen in Figure 6. This surface yields computed room temperature values of C_∞ and $dlnC_\infty/dT$ that are both considerably *smaller* than those observed experimentally (Table I). A snapshot,[24] which freezes the thermal motions of a cellulosic chain based on the adiabatic surface, is shown in Figure 7, where only the positions of the glycosidic oxygens appear. These oxygens are linked in Figure 7 by virtual bonds which span the sugar residues. Mean directional correlations[25] of the successive virtual bonds with the direction of the first such bond are shown in Figure 8 to decay rather quickly to zero in keeping with the quite flexible character predicted for the chain.

It is of some interest to contrast the chain behavior predicted on the basis of the adiabatic map with that predicted by the rigid residue model for $\beta = 116°$. In the latter case the directional correlation function, also shown in Figure 8, decays much more slowly, and in contrast to the relaxed residue model, which predicts too flexible and compact a chain, the rigid residue model predicts behavior that is too extended; see the snapshot in Figure 9. It is clear, however, that relatively small, perhaps easily defensible, adjustments of either model can yield a calculated room temperature C_∞ ~ 35. For example, one might wish to incorporate explicitly into the relaxed model the acetate ester groups of the experimental reference compound, cellulose triacetate. Accurate modeling of the experimental temperature coefficient is a more difficult matter. It is our opinion that the large negative temperature coefficient can be explained only by the occasional introduction of alternate ring conformers into the chain.[4,5]

Comparisons of the implications of rigid and relaxed residue models for homoglucans need not stop with cellulose. A relaxed map for maltose, the dimeric segment of amylose, has recently been published by Brady and coworkers.[13] This is shown in Figure 10. Computed and measured values of C_∞ and $dlnC_\infty/dT$ are given in Table I. For comparison, the much more conformationally constrained conventional maltose surface is shown in Figure 11.[21,24] In this case the good agreement between the computed and measured values of C_∞ and $dlnC_\infty/dT$ (Table I) is the result of a refinement of the rigid maltose energy surface specifically to
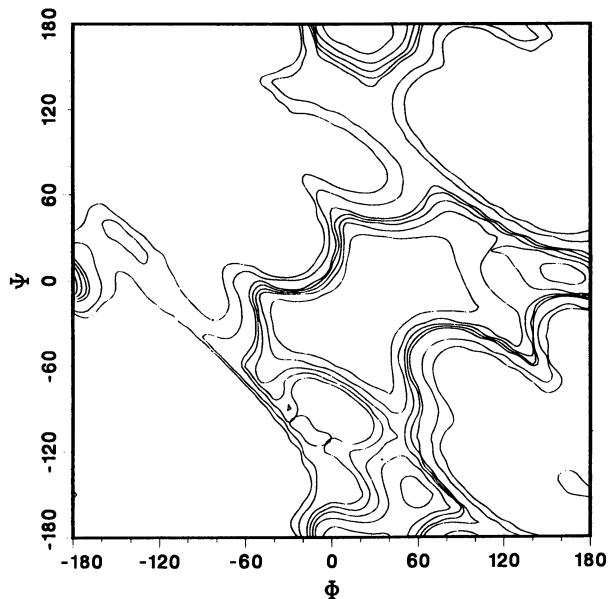
Figure 5. Contour diagram for β-cellobiose analogous to Fig. 2 for β = 124°. Contours are drawn at 2, 4, 6, 8, 10, 25, and 50 kcal/mol relative to the absolute minimum near φ, ψ = 10°, -40°.
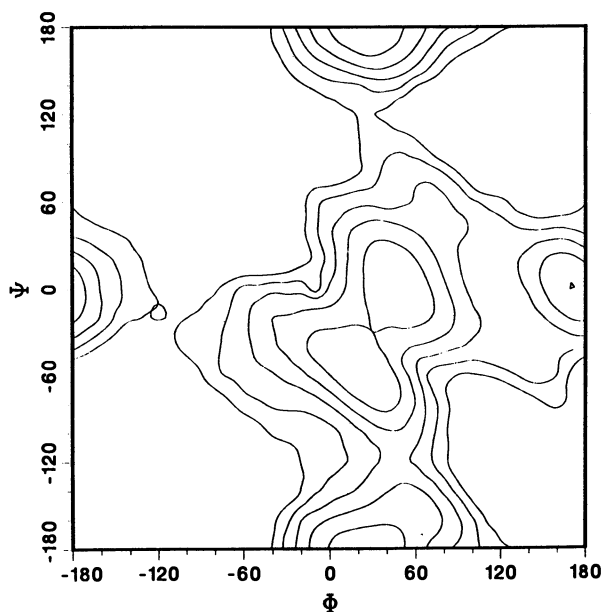
Figure 6. Relaxed or adiabatic conformational energy surface for β-cellobiose as computed by French[12]. Contours are drawn at 2, 4, 6, 8, and 10 kcal/mol above the minimum near φ, ψ = 20°, -60°.
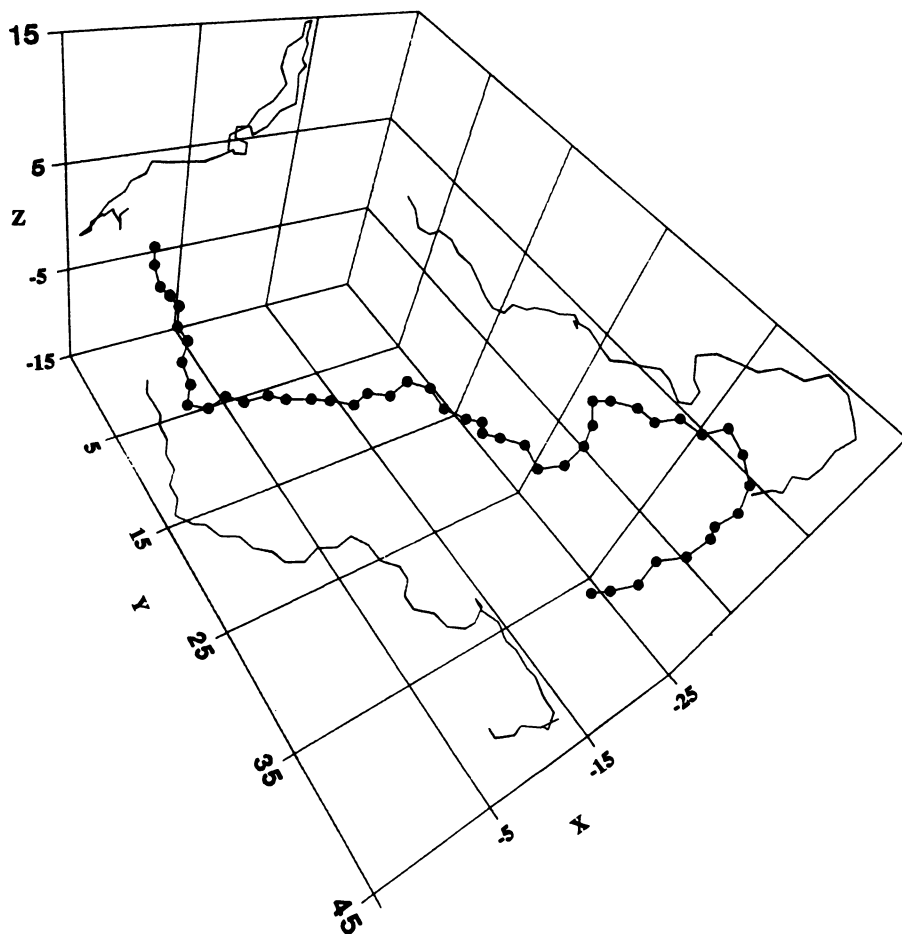
Figure 7. A "snapshot" of a typical cellulosic chain trajectory taken from a Monte Carlo sample of cellulosic chains, all based on the conformational energy map of Fig. 6. Filled circles representing glycosidic oxygens, linked by virtual bonds spanning the sugar residues (not shown), allow one to trace the instantaneous chain trajectory in a coordinate system that is rigidly fixed to the residue at one end of the chain. Projections of the chain into three mutually orthogonal planes assist in visualization of the trajectory in three dimensions.
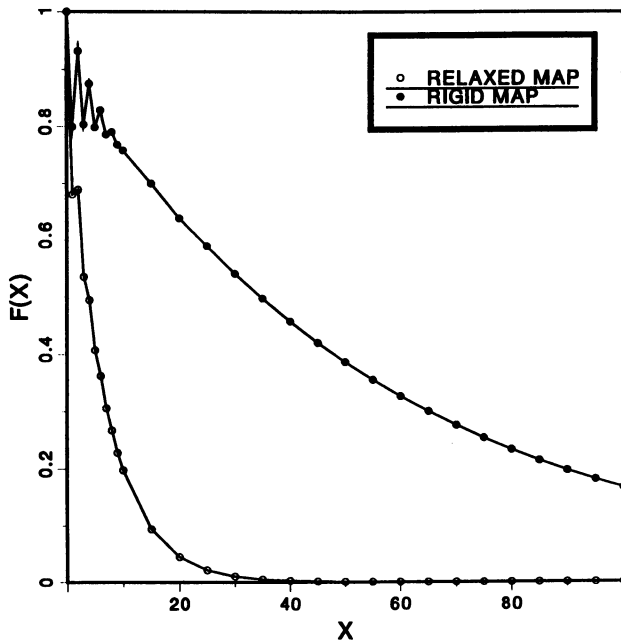
Figure 8. The mean directional correlation F(x) of virtual bond x with the initial virtual bond in the chain.[25] Closed circles correspond to a calculation based on the "rigid" cellobiose map of Fig. 2; open circles refer to the relaxed cellobiose surface of Fig. 6.
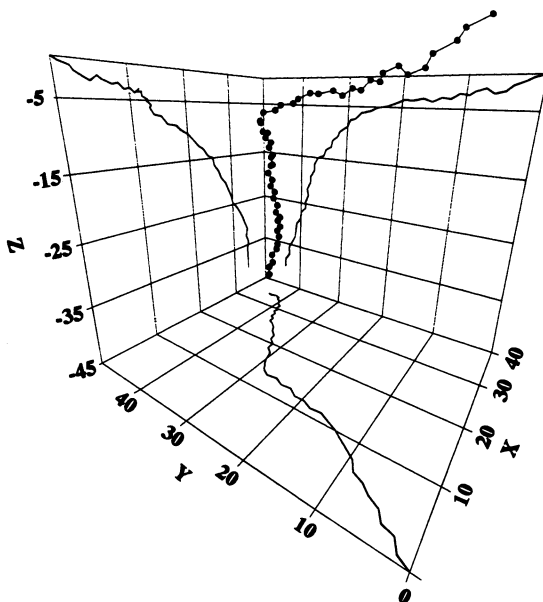
Figure 9. Same as Fig. 7 except based on the rigid β-cellobiose map of Fig. 2.
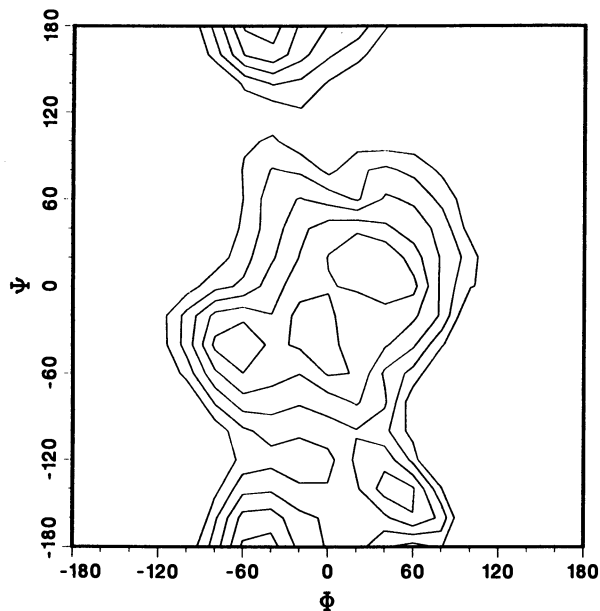
Figure 10. Relaxed (adiabatic) conformational energy map for β-maltose as computed by Brady and coworkers.[13] Contours are drawn at 2, 4, 6, 8, and 10 kcal/mol above the minimum near φ, ψ = -60°, -40°. The β-maltose structure may be derived from that of β-cellobiose in Fig. 1 by inversion of the stereochemical configuration at C1.
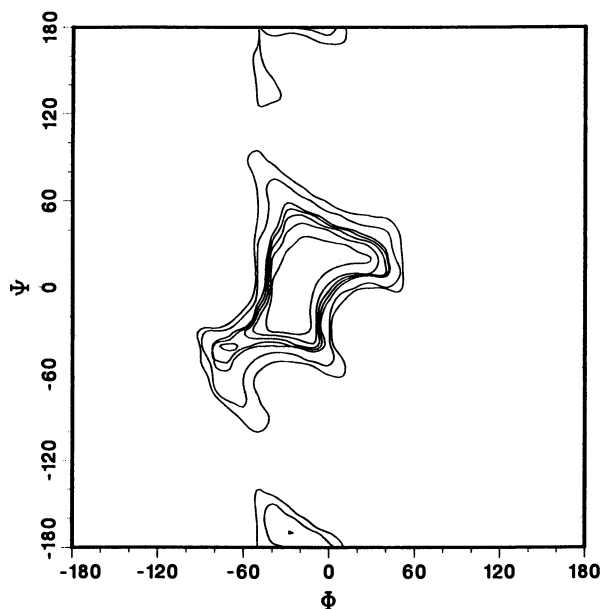
Figure 11. Conformational energy map for rigid α-maltose.[24] Contours are drawn at 2, 4, 6, 8, 10, 25, and 50 kcal/mol above the minimum near φ, ψ = -30°, -20°.

bring the model into agreement with these experimental observables. A typical chain trajectory (snapshot) generated with the rigid residue model is shown in Figure 12. It has the *ca.* six-fold pseudohelical character associated by many with the dissolved amylosic chain. The nascent annulus, which can harbor iodine, alkanols, alkanoic acids, and detergents in inclusion complexes, is clearly discernable. A snapshot based on the relaxed residue model, Figure 13, is rather different, and creation of the *ca.* six-fold helical cavity for the formation of inclusion complexes must evidently result more from induction by the included agent.

Differences in the conformational characteristics of typical amylosic chains modeled with rigid and adiabatic maltose energy surfaces appear again when the directional correlation functions are compared in Figure 14. The regular periodic decay in this function for the rigid residue model reflects the strong tendency for short segments of the chain based on this model to adopt six-fold helix-like trajectories. Inherent flexibility causes the correlations to decay, but with persistence of the periodic character. Directional correlations of the chain based on the relaxed residue model decay more rapidly and with a completely different, lower frequency, periodic character.

Another way to compare the behavior predicted by these alternative models is to consider the *mean* trajectory, or persistence vector, of the chain as viewed from a coordinate system with its origin rigidly affixed to the sugar residue at one end of the chain.[24] The mean trajectory based on the rigid residue model shown in Figure 15 has the anticipated helical character. Again, only virtual bonds and glycosidic oxygens are shown for clarity, and in this case only the projection into the XY plane is shown. Reduction in the *mean* length of the successive virtual bonds with progress toward the remote end of the chain is a consequence of loss of directional persistence as the chain grows longer and is *not* an artifact of an effort to render perspective. The mean incremental contribution to the chain extension from each additional residue added at the remote end diminishes, because its directional correlation with the first virtual bond is less than that of its predecessor. When the chain is long enough so that the correlation function for the rigid residue model in Figure 14 has decayed to zero, the successive glycosidic oxygens at the remote end of the mean trajectory in Figure 15 will have coalesced to a point, and the persistence vector will have converged to its asymptotic length. This is a manifestation of the distortion of space, described earlier, that results from averaging over the ensemble of conformations of a flexible molecule.

The mean trajectory for the amylose chain based on the model using the adiabatic energy surface is clearly very different; see Figure 15. The first virtual bond, subject to no conformational averaging in a coordinate system attached to the initial residue, nearly superimposes on that for the rigid model. (Superposition would be exact but for minor differences in the residue structural geometries inherent in the two models.) Curvature of the trajectory reflects the low frequency periodic character of the chain disclosed in Figures 13 and 14 but, as expected from the corresponding correlation function in Figure 14, the persistence vector has effectively converged after approximately 12 residues and one pseudohelical period.

Evidently the "vacuum" relaxed maltose map of Brady *et al.*[13] does not provide a more quantitatively correct model for C_∞ and $d\ln C_\infty/dT$ of dissolved amylosic chains than does the conventional rigid residue model, despite the impression that it should account more realistically for the structural characteristics of the elementary maltose disaccharide segment of the amylosic chain. Incorporation of the effects of aqueous solvation into the potential surface might serve to improve its performance in this respect. In any case these comparisons highlight the need to devise a variety of rigorous tests of the potential functions and other attendant approximations associated with computer modeling of carbohydrate conformation. The rigid maltose model,[21,24] refined specifically to account for C_∞ and $d\ln C_\infty/dT$
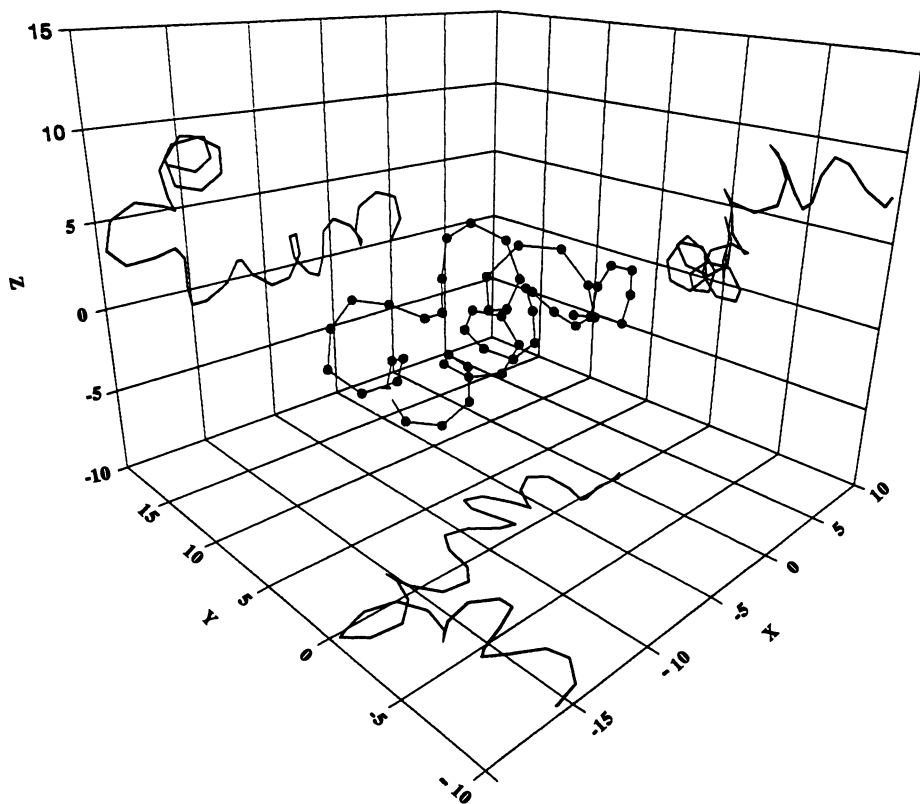
Figure 12. Snapshot as in Fig. 7 for an amylosic chain trajectory based on the rigid residue maltose map of Fig. 11.
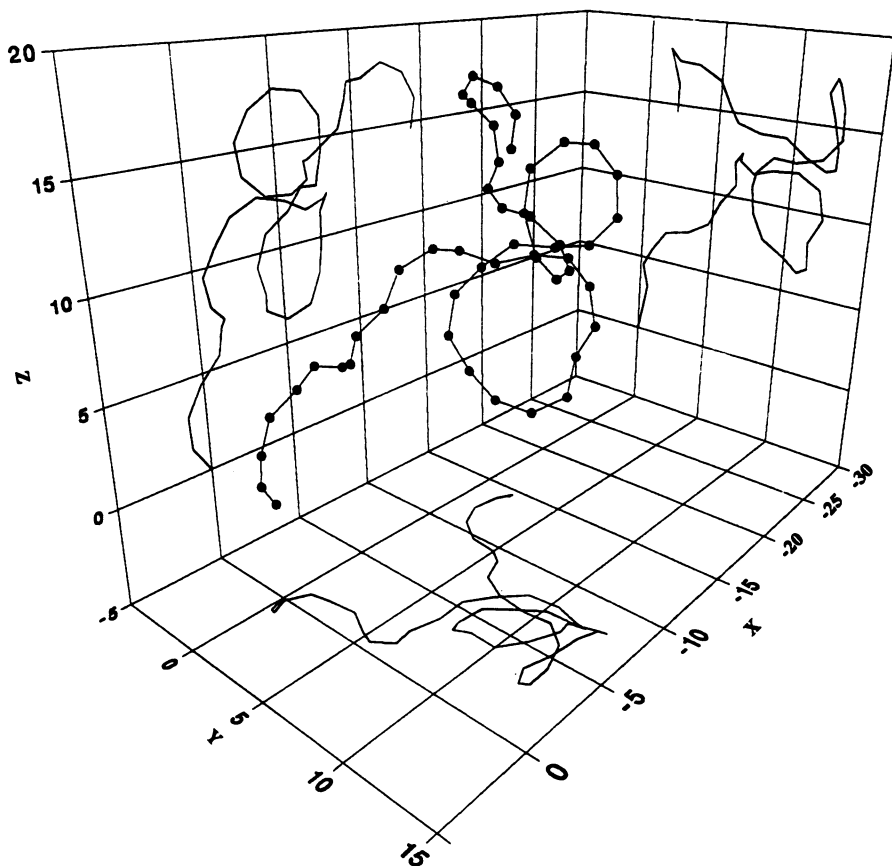
Figure 13. Same as Fig. 12 for an amylosic chain trajectory based on the relaxed residue maltose map of Fig. 10.
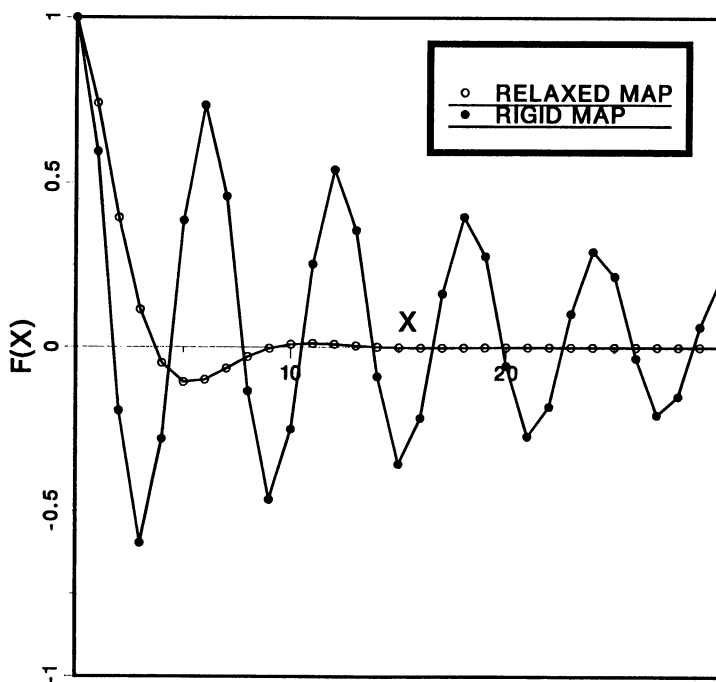
Figure 14. Directional correlation function as in Fig. 8 for amylosic chains based on the rigid residue model (filled circles) and relaxed residue model (open circles).
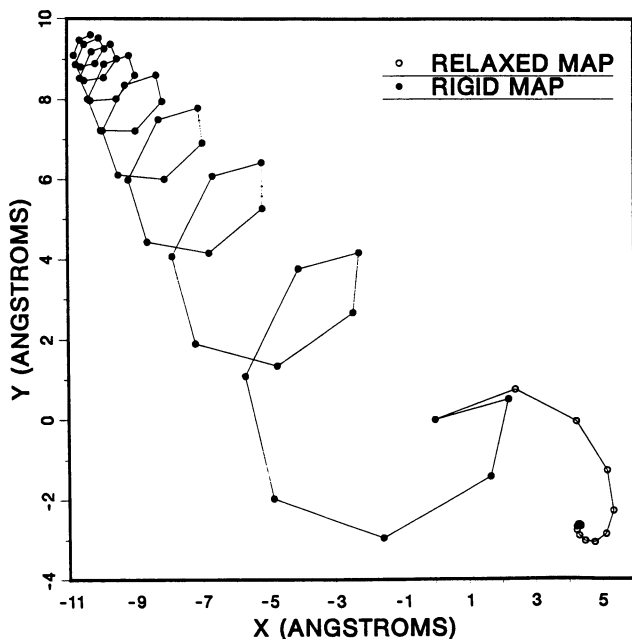
Figure 15. The *mean* trajectories of amylosic chains based on the rigid residue (filled circles) and relaxed residue (open circles) models projected into the XY plane of a coordinate system attached to a terminal residue. Circles represent the *mean* positions of successive glycosidic oxygens in the primary sequence. The persistence vector (mean end-to-end vector) for a chain of x residues is the vector (not shown) connecting the origin and the mean position of the glycosidic oxygen separated from it along the chain by x virtual bonds.

of aqueous amylosic chains, may prove completely inadequate when tested against other sorts of experimental data. It must, for example, because it is based on the independent linkage approximation, incorporate implicit features that account with quantitative success, but without physical reality, for the problem, potentially serious for any model that purports to explain aqueous amylosic solution data, of self intersection of the amylose chain at a range of 6-8 residues in the chain sequence.[24] These implicit features render the surface potentially inappropriate for describing the properties of the maltooligomers with DP $\leq$ 5. Adiabatic maltose surfaces generated in the presence of an aqueous solvation layer, or, better, potentials of mean force in ϕ, ψ space computed for aqueous maltose offer hope of improved general models, but may still suffer in their applicability to polymers and larger oligomers from the difficulties associated with non-separability of the conformational energy.

Other Manifestations of the Breakdown of the Independent Linkage Approximation

We wish to discuss briefly several other instances in which separability of the conformational energy of oligo- or polysaccharides may become an important issue. The dense packing of the oligosaccharide chains in the bi- and triantennary N-linked glycosyl groups of many glycoproteins clearly obviates any approximation as simple as $\{\theta_s(i)\} \cap \{\theta_s(j)\} = \phi$ in computer modeling of these structures.[19-20] Similar, but less severe, problems arise in considerations of comblike, branched microbial polysaccharides such as xanthan or succinoglycan.[26] Here the sidechain may in principle interact with several mainchain residues. In many cases these sidechain-mainchain interactions introduce a coupling of the glycosidic linkage torsions that vitiates the assumption of independent (groups of) linkage torsions. A mean field treatment of the influence of sidechain-mainchain interactions on the solution configurations of branched microbial polysaccharides has been developed,[27] and elementary applications have been described in two recent publications.[28-29] This method is capable of dealing approximately with the effects of second neighbor and more remote short range interactions of the residues along the backbone of linear polysaccharides. Further consideration of these effects, for example, from the interactions at the range of 6-7 residues described above for amylosic chains or those provoked by the occurrence of (1→2)-linkages, represents an area open for further methodological advances.

　　　To illustrate the consequences for polymeric properties of (1→2)-linkages we consider briefly a sulfated α-(1→2)-L-fucan isolated from the eggs of certain echinoderms and implicated in the species specificity of egg fertilization.[30] Structural details of the repeating disaccharide unit are shown in Figure 16, from which the potential steric complications of attaching a third α-L-fucose residue to the dimer at atoms O2 or O1' are evident. The rigid residue conformational energy surface for the disaccharide of Figure 16 is shown in Figure 17, where a map of helical rise h in Å units per residue is superimposed as shown by the labelled contours. A contour for h = 0 (not shown) bisects the region bounded by the contours of h = 0.5 Å and h = -0.5 Å and corresponds to the locus in ϕ, ψ space of all self-intersecting helices of zero pitch. (Here negative values of h designate right-handed helices; the row of small circles in Figure 17 in the vicinity of h = 0 is an artifact of the contouring routine.)

　　　The class of least energetic conformers for the α-(1→2)-L-fucan disaccharide is found in the low energy region of the surface near ϕ, ψ = 30°, 35°. This minimum sits directly astride the locus of self-intersecting helices and thus discloses that longer α-(1→2)-L-fucan oligosaccharides, *with every glycosidic linkage constrained to adopt ϕ, ψ values characteristic of the lowest energy region of the disaccharide map,* would suffer untenable steric difficulties. These begin to become apparent, as already suggested by cursory examination of Figure 16, in the trisac-
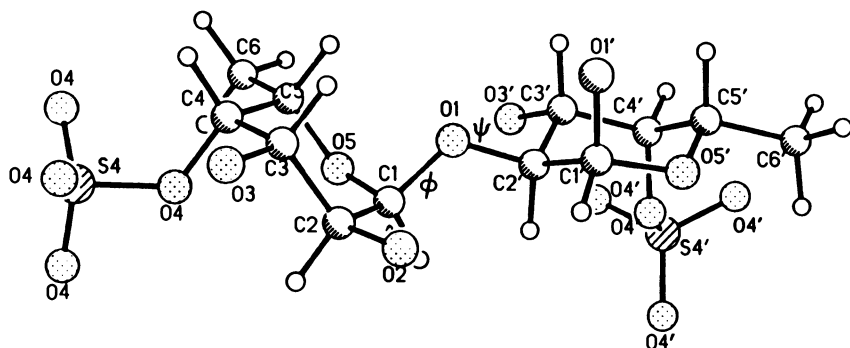
Figure 16. Ball and stick drawing of 2-O-(4-O-sulfo-α-L-fucopyranosyl)-4-O-sulfo-α-L-fucopyranose ("fucobiose") in the ϕ, $\psi = 0°$ reference conformation as in Fig. 1. Structural characteristics of rigid sulfated fucose residue deduced from energy minimization with MM2 to be reported elsewhere.
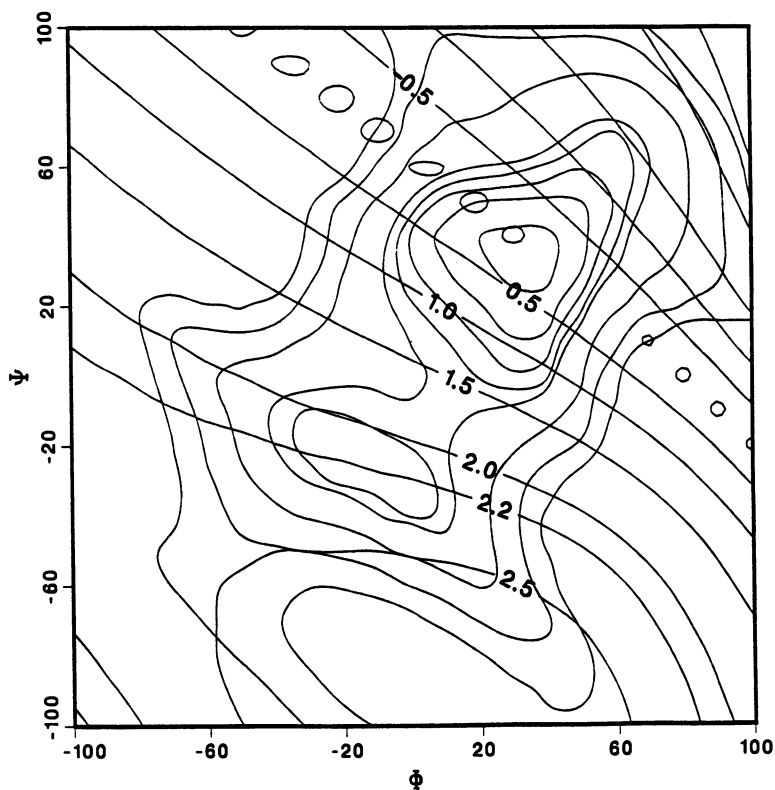
Figure 17. Contour diagram for fucobiose analogous to Fig. 2. Contours are drawn at absolute energies -1, 0, 1, 2, 3, 10, 25, and 50 kcal/mol. Labeled contours depict helical rise per residue - see text for details.

charide. This is clear from Figure 18, which shows the energy of one terminal residue of an α-(1→2)-L-fucan trisaccharide due to its interactions with the other two residues, for all conformations subject to the helical condition $\phi_i = \phi_j$, $\psi_i = \psi_j$ for all linkage pairs i and j. Here we see that steric conflicts raise the absolute minimum on the surface by approximately 3 kcal/mol in comparison with Figure 17, and the energy near ϕ, $\psi = 30°$, 35° at the minimum on the disaccharide map has risen from ca. -1 to ca. 50 kcal/mol. This result suggests that the introduction of even an isolated 1→2 linkage in an oligosaccharide sequence can precipitate interactions of second neighbor sugar residues that may vitiate the independent gly-cosidic linkage approximation.

By the time the nascent helical α-(1→2)-L-fucan reaches a length of five residues the interaction energy of a terminal residue with the other residues of the chain becomes infinite for all ϕ, ψ pairs lying on or near the locus h = 0. Figure 19 shows the energy surface for a terminal residue of the helical α-(1→2)-L-fucan hex-amer. Superimposed are not only the contours of h shown in Figure 17 but also labeled contours of the number of residues per helical turn n. The energy of a ter-minal residue has effectively converged at DP = 6, and no further information is available from consideration of longer helical segments of the α-(1→2)-L-fucan.
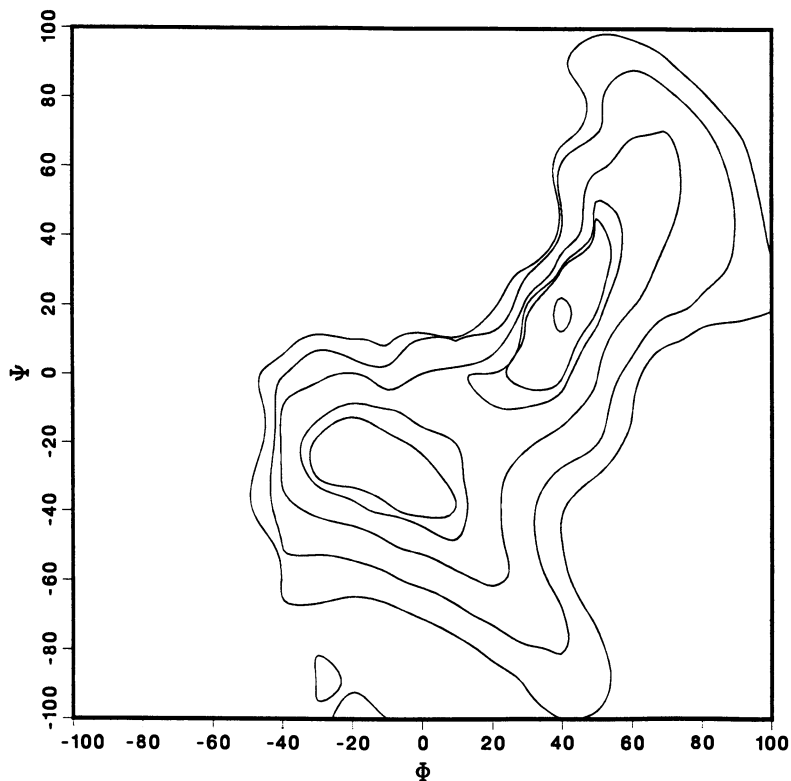
Figure 18. Contour diagram for the conformational energy of the terminal residue of a trisaccharide segment of the α-(1→2)-L-fucan constrained to satisfy the helical condition - see text for details. Contours are drawn at absolute ener-gies 2, 4, 5, 10, 25, and 50 kcal/mol.
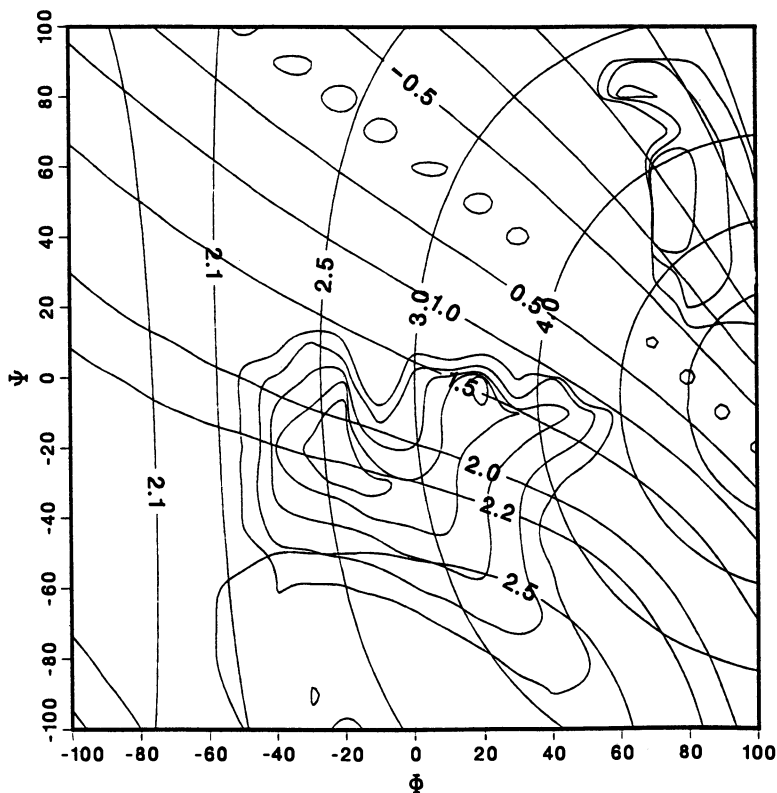
Figure 19. Contour diagram for the conformational energy of the terminal residue of a hexasaccharide segment of the α-(1→2)-L-fucan constrained to satisfy the helical condition. Contours of axial rise per residue and residues per helical turn are superimposed - see text for details. Contours are drawn at absolute energies 2, 4, 10, 25, and 50 kcal/mol.

The surface in Figure 19 thus clearly represents *half* the total energy of an interior α-L-fucose residue embedded in a long α-(1→2)-L-fucan helix due to its interactions with the other residues of the chain.

Two energetically competitive regions of *helical* conformation space appear on the hexamer energy surface, near φ, ψ = -25°, -15°, where h ≈ 2.1 Å and n ≈ 2.6, and near φ, ψ = 20°, 0°, where h ≈ 1.5 Å and n ≈ 3.5 Å. As an alternative to treating the dissolved α-(1→2)-L-fucan chain, in parallel with the above treatments of the cellulosic and amylosic chains, as a sequence of independent, or approximately independent, glycosidic linkages with the properties inherent in the disaccharide energy surface of Figure 17, it is tempting to consider the fucan as a semi-flexible chain deriving its tortuosity from fluctuations about one or the other (or both) of the low energy helical structures identified in Figure 19. An analysis following this approach will be reported elsewhere.

Acknowledgments

Literature Cited

1.   J. J. Cael, J. L. Koenig, and J. Blackwell, *Biopolymers*, **14**, 1885 (1975).
2.   D. Cremer and J. A. Pople, *J. Amer. Chem. Soc.*, **97**, 1354 (1975).
3.   C. Altona and M. Sundaralingam, *J. Amer. Chem. Soc.*, **94**, 8205, (1972).
4.   K. D. Goebel, C. E. Harvie, and D. A. Brant, *Appl. Polym. Symp.*, **28**, 671 (1976).
5.   D. A. Brant, *Carbohydr. Polym.*, **2**, 232 (1982).
6.   M. Ragazzi, D. R. Ferro, B. Perly, G. Torri, B. Casu, P. Sinay, M. Petitou, and J. Choay, *Carbohydr. Res.*, **165**, C1 (1987).
7.   P. J. Flory, *Macromolecules*, **7**, 381 (1974).
8.   D. A. Brant, *Q. Rev. Biophys.*, **9**, 527 (1976).
9.   T. L. Hill, *Statistical Mechanics*, McGraw-Hill, New York (1956).
10. S. N. Ha, A. Giammona, M. Field, and J. W. Brady, *Carbohydr. Res.*, **180**, 207 (1988).
11. I. Tvaroska, *Biopolymers*, **21**, 1887 (1982).
12. A. D. French, *Biopolymers*, **27**, 1519 (1988).
13. S. N. Ha, L. J. Madsen, and J. W. Brady, *Biopolymers*, **27**, 1927 (1988).
14. D. R. Herschbach, H. S. Johnston, and D. Rapp, *J. Chem. Phys.*, **31**, 1652 (1959).
15. D. A. Brant and K. D. Goebel, *Macromolecules*, **8**, 522 (1975).
16. P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, New York (1953).
17. K. F. Freed, *Renormalization Group Theory of Macromolecules*, Wiley-Interscience, New York (1987).
18. M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, Oxford University Press, New York (1986).
19. D. A. Cumming, R. N. Shah, J. J. Krepinsky, A. A. Grey, and J. P. Carver, *Biochemistry*, **26**, 6655 (1987).
20. D. A. Cumming and J. P. Carver, *Biochemistry*, **26**, 6664 (1987).
21. D. A. Brant and W. L. Dimpfl, *Macromolecules*, **3**, 655 (1970).
22. D. W. Tanner and G. C. Berry, *J. Polym. Sci.: Polym. Phys. Ed.*, **12**, 941 (1974).
23. S. Arnott and W. E. Scott, *J. Chem. Soc. Perkin II*, 324 (1972).
24. R. C. Jordan, D. A. Brant, and A. Cesaro, *Biopolymers*, **17**, 2617 (1978).
25. B. A. Burton and D. A. Brant, *Biopolymers*, **22**, 1769 (1983).
26. G. Gravanis, M. Milas, M. Rinaudo, and B. Tinland, *Carbohydr. Res.*, **160**, 259 (1987).
27. T. A. Talashek, Ph.D. Dissertation, University of California, Irvine, 1988.
28. T. A. Talashek and D. A. Brant, *Carbohydr. Res.*, **160**, 303 (1987).
29. D. A. Brant and T. A. Talashek, in *Industrial Polysaccharides: The Impact of Biotechnology and Advanced Methodologies*, Gordon and Breach Science Publishers, New York (1987).
30. P. L. DeAngelis and C. G. Glabe, *Biochemistry*, **27**, 8189 (1988).

Chapter 5

Molecular Dynamics Simulations of Carbohydrates and Their Solvation

L. J. Madsen[1], S. N. Ha, V. H. Tran[2], and J. W. Brady

Department of Food Science, Cornell University, Ithaca, NY 14853–7201

Molecular dynamics (MD) simulations are a class of
molecular mechanics calculation which directly model the
motions of molecular systems, often providing
considerable information which cannot be obtained by any
other technique, theoretical or experimental. MD
simulations have only recently been applied to problems
of carbohydrate conformation and motions, but it is
likely that this technique will be widely used for
modeling carbohydrates in the future. This paper
introduces the basic techniques of MD simulations and
illustrates the types of information which can be gained
from such simulations by discussing the results of
several simulations of sugars. The importance of
solvation in carbohydrate systems will also be
discussed, and procedures for including solvation in
molecular dynamics simulations will be introduced and
again illustrated from carbohydrate studies.

Molecular mechanics calculations are an attempt to understand the
physical properties of molecular systems based upon an assumed
knowledge of the way in which the energy of such systems varies as a
function of the coordinates of the component atoms. While this term
is most closely associated with the conformational energy analyses
of small organic molecules pioneered by Allinger (1), in their more
general applications molecular mechanics calculations include energy
minimization studies, normal mode calculations, molecular dynamics
(MD) and Monte Carlo simulations, reaction path analysis, and a
number of related techniques (2). Molecular mechanics

[1]Current address: Department of Physical Sciences, Mesa College, Grand Junction, CO 81502

[2]Current address: Laboratoire de Physicochimie des Macromolécules, Institut National de la
Recherche Agronomique, B.P. 527, Nantes, Cédex F–44026, France

calculations have long been used to analyze biopolymer conformations, perhaps beginning with the work by Ramachandran on the allowed conformations of polypeptides (3). The history of molecular mechanics conformational energy studies of carbohydrates is almost as old, with the first such studies of carbohydrates (4) coming very shortly after the initial studies of peptides. Conformational energy calculations have become an integral part of polysaccharide studies (5), particularly in the interpretation of fiber diffraction data (6,7).

As the theoretical study of protein and nucleic acid conformations matured, the inherently dynamical behavior of these molecules was recognized, and beginning with the simulations of Karplus and coworkers, these motions were studied theoretically by directly modeling them in molecular dynamics simulations (8). The systematic investigation of protein motions has revealed that in many cases the physical properties and biological function of these polymers cannot be understood without taking internal motions into consideration (2). Although MD simulations of these classes of molecules have now become quite commonplace, parallel studies of carbohydrates were not attempted until quite recently. Only in the last few years have MD studies of carbohydrates been reported (9-13). While the importance of flexibility in carbohydrate molecules has long been recognized (14-16), these simulations have served to directly illustrate the effects of such dynamical flexibility, and it is certain that MD simulations of carbohydrates will become routine in the near future.

It is also now generally understood that the structure and biological function of many biomolecules are affected by aqueous solvation. For this reason, theoretical models of biopolymers must include these solvent effects as well as internal flexibility. The unique structural behavior of water, particularly immediately adjacent to solutes, makes it difficult to apply continuum theories to aqueous solutions. However, with the development of high speed computers, it is now becoming feasible to directly model the behavior of aqueous systems through MD and Monte Carlo simulations which specifically include solvent water molecules. A wide variety of such calculations have now been reported, including studies of pure water (17) and solutions of nonpolar atoms (18-20), ions (21,22), alcohols (23,24), urea (25,26), peptides (27,28), and even proteins (29). MD simulations of carbohydrate solvation could be particularly useful, since there are a number of unanswered questions concerning the interaction of these molecules with water (30,31). Complex solution behavior might well be expected for carbohydrates, which contain a number of adjacent polar hydroxyl groups held in relatively fixed positions around the pyranoid rings, as well as nonpolar CH and CH_2 groups and the ether-like ring oxygen atoms. This expectation of complex solution behavior has been fulfilled in the solution simulations of carbohydrate molecules which have been reported (13,32, and Ha, S.; Gao, J.; Tidor, B.; Brady, J.W.; Karplus, M. submitted to J. Am Chem Soc.), and demonstrates the need for further studies which include the effects of aqueous solvation. In this paper, the general technique of MD simulations will be reviewed and examples of applications to

carbohydrate molecules will be discussed, as well as the inclusion
of aqueous solvation into such studies.

Molecular Dynamics Simulations

Molecular dynamics calculations are a technique for modeling
physical systems on the microscopic scale in which atomic motions
are directly simulated by numerically solving the classical Newton's
equations of motion for all of the atoms in the system subject to
the forces arising from some given force field. MD simulations were
originally developed to model simple physical systems such as
monatomic rare gases (33), but they have now been applied to a wide
variety of problems, including water and aqueous solutions, and
essentially all types of biopolymers (2). Several sophisticated MD
programs are currently generally available, including the well-known
molecular mechanics packages CHARMM (34), AMBER (35), and GROMOS
(36), and many such programs are available commercially.
        A fundamental requirement of molecular mechanics studies of any
type, including MD simulations, is a complete description of the
variation of the total potential energy of the system of interest as
a function of the molecular coordinates. Strictly speaking, this
energy is given by the complete solution of the Schrödinger equation
for the entire system, including solvent molecules, as their
positions evolve with time. Since for macromolecules and condensed
phases the accurate calculation of this quantum mechanical energy is
not possible, it is common to employ analytic, semi-empirical energy
expressions which have theoretically reasonable functional forms and
which have been parameterized to the results of experiments and
quantum mechanical calculations of simple molecules. Extensive sets
of such potentials have been developed to describe condensed phases
of small molecules, including water (17,37,38), simple organic
compounds (1,39), and various biopolymers (34-36). Such empirical
energy surfaces have also been developed for sugars (40-42), since
carbohydrates contain functional groups which do not occur in other
biological molecules and since potential parameters are sensitive to
chemical environment. Unfortunately, to date none of these
carbohydrate parameterizations have proven to be completely
satisfactory. Work on developing improved force field parameters
for various types of molecules is continuing.
        Semiempirical potential energy surfaces vary in detail and even
in form from one set to the next (2), but most represent the
intramolecular potential as a sum of electrostatic and van der Waals
interactions between non-bonded atoms and terms for hindered
rotation about molecular bonds. Bond stretching and angle bending
forces are derived from harmonic restoring potentials, and special
functions are sometimes included to account for hydrogen bonding. A
typical example of such a function of the internal coordinates $\underset{\sim}{q}$
might be

$$V(\underset{\sim}{q}) = \Sigma k_{b_i}(b_i-b_{i0})^2 + \Sigma k_{\theta_i}(\theta_i-\theta_{i0})^2 + \Sigma k_{\phi_i}[1 + \cos(n\phi_i-\delta_i)]$$

$$+ \Sigma(A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^6 + q_iq_j/r_{ij}) \tag{1}$$

where k_{b_i} and k_{θ_i} are the bond stretching and angle bending force constants, ϕ_i is a torsion angle with a force constant k_{ϕ_i}, periodicity n, and phase factor δ_i, and q_i and q_j are the atomic partial charges. A_{ij} and B_{ij} are van der Waals constants and r_{ij} in this equation represents the interatomic distance between atoms i and j. An energy function of this type allows the calculation of the variation of the system energy as a function of the positions of every atom in the system, and because it has an analytic form, also permits the direct analytic evaluation of the derivatives of this function with atomic positions, which gives the atomic forces. Of course, in order for molecular mechanics calculations to be of use, it is necessary for the potential energy functions to be sufficiently realistic as to adequately mimic physical behavior. For this reason considerable effort goes into the development of the parameters which appear in the energy function (2,34). This parameterization is generally accomplished by the matching of calculated properties to experimental measurements, as a function of the parameter set for selected small model compounds.

Representing the molecular potential energy as an analytic function of the nuclear coordinates in this fashion implicitly invokes the Born-Oppenheimer approximation in separating the very fast electronic motions from the much slower ones of the nuclei. This separation allows the electronic energy at each nuclear configuration to be represented as the potential energy for the motions of those nuclei for that configuration (2). The energy function parameterization is thus an empirical approximation to this quantum mechanical energy, and in typical classical molecular mechanics calculations, is the only place where the quantum behavior of the molecular system is included. The system is subsequently considered to be governed solely by classical mechanics, whether the simulation is an energy minimization, normal mode analysis, or MD or Monte Carlo calculation. For most structural and dynamical properties of interest this is a valid approximation, but biopolymers, including sugar, do contain hydrogen atoms, whose mass is sufficiently small that quantum effects for these atoms can be significant. Quantum effects may be particularly important in chemical reactions, electron and proton transport, and other events which involve large changes in electronic distributions.

From elementary classical mechanics (43), the force acting on an atom in a molecular mechanics system is the negative gradient, or derivative, of the potential energy function with respect to atomic position,

$$\vec{F}_i = - \vec{\nabla}_i V \tag{2}$$

These forces can be substituted into Newton's equations of motion,

$$\vec{F}_i = m_i \vec{a}_i \tag{3}$$

which can in principle be integrated numerically on a large digital computer to provide a complete description of the motion of every atom in the molecular system as a function of time, $q(q_0, v_0, t)$, for

any given set of initial positions q_0 and velocities v_0 at the initial time t_0. Because the forces are conservative (that is, derivable from a potential as in equation (2)), the total energy of the system is constant throughout the simulation, as are the linear and angular momenta. Properties calculated from such a simulation are thus microcanonical ensemble properties (constant energy, number of particles, and volume) (44). The total energy, or Hamiltonian, is made up of potential and kinetic energy terms,

$$E_T = E_K + V \tag{4}$$

which are not constant and which fluctuate as potential energy is converted into motion (kinetic energy) and then back again. The instantaneous temperature of the MD system is calculated from the velocities $\vec{v}_i$ of the atoms as

$$T = \frac{1}{3N} \sum_{i=1}^{N} m_i |\vec{v}_i|^2 \tag{5}$$

where N is the total number of atoms in the system, and this quantity will of course fluctuate as the kinetic energy fluctuates. The thermodynamic temperature of the system is the mean value of this instantaneous temperature averaged over a sufficiently long dynamics sequence such that the temperature converges to a stable value characteristic of all possible motions in the system under the given conditions.

In a typical MD simulation, a starting structure q_0 is selected, usually from crystallographic data, or in the case of sufficiently simple small carbohydrates, stereotypical chair or boat forms, and initial velocities $\vec{v}_{10}$ for every atom are selected from a Boltzmann distribution at the desired system temperature (2). The integration of the equations of motion is carried out numerically using one of several integration procedures. The most commonly used method is the Verlet algorithm (2,45,46), which calculates the position at a time $t' = t + \Delta t$ through a truncated Taylor series expansion in the time step Δt about the position at t, $x(t)$. More elaborate integration schemes, such as the Gear predictor-corrector algorithm are also sometimes used (2), but these require considerably more computer time. The integration time step size Δt in the Verlet algorithm must be small (usually in the femtosecond range) relative to the fastest motions of the system for the procedure to be reliable, which means that the simulation of a lengthy period of time will require quite a large number of time steps. Each time step of the integration requires the costly evaluation of the forces at that step, thus making the calculation expensive in terms of computer time and limiting the periods which may be practically simulated.

Since the selection of starting conditions in MD simulations can be somewhat arbitrary, and not necessarily realistic, it is necessary to "equilibrate" the system by integrating the equations of motion for some period of time during which the behavior may not

be truly representative of the physical system. During this period
the integration is carefully monitored to ensure that the energy is
being conserved and that the temperature is stable and has the
desired value. If as the result of relaxation processes the system
temperature drifts away from the desired value by more than some
small tolerance, it is customary to either scale all of the atomic
velocities in the system by an appropriate factor to bring the
temperature back to the specified value, or to again assign new
values for all of the atomic velocities by random selection from
a Boltzmann distribution. Because it may not be physically
representative, the equilibration portion of the trajectory, which
must last for a number of picoseconds, is not analyzed for the
calculation of system properties.

   After the system has stabilized sufficiently to be considered
"equilibrated" (that is, all artificial stresses produced by the
selection of initial conditions have relaxed away), the integration
is continued for a much longer period without further intervention
for the purpose of actually simulating equilibrium dynamical
behavior. In principle, when a simulation has been integrated for a
sufficiently long period of time, the mean properties computed as
time averages over the entire MD simulation will converge to the
"true" thermodynamic, canonical ensemble properties of the system.
Unfortunately, it is difficult to know when this "thermodynamic
limit" has been reached, and in systems of even moderate complexity
the thermodynamic limit may not be attainable with realistic
computer times. Reported MD simulations of carbohydrates have
typically ranged from 10-20 ps to 500 ps in length, although
supercomputers now permit simulations of modest-sized systems in the
nanosecond range.

Molecular Dynamics Simulations of Carbohydrates

The first reported molecular dynamics simulations of carbohydrates
began to appear in 1986, with the publication of studies of the
vacuum motions of α-D-glucopyranose (9), discussed below, and the
dynamics of a hexa-NAG substrate bound to lysozyme (10), which are
described in greater detail in the chapter by Post, et al. in this
volume. Since that time, simulations of the dynamics of many more
carbohydrate molecules have been undertaken. A number of these
studies are described in subsequent chapters of this volume. The
introduction of this well developed technique to problems of
carbohydrate structure and function could contribute substantially
to the understanding of this class of molecules, as has been the
case for proteins and related biopolymers.

   One of the most significant potential contributions of MD
studies to the field of carbohydrate chemistry is in exploring the
degree of flexibility of these molecules and in illuminating any
possible biological or structural roles for such flexibility.
Although polysaccharides have been usefully modeled in the past as
rigid monomer units (5), sugars are not completely rigid, and when
MD simulations are applied to carbohydrates a wide variety of
internal motions, fluctuations and conformational transitions are
found to occur. For example, in MD simulations of glucopyranose
(9), when MD trajectories were initiated in the $^4C_1$ conformation,

the molecule remained in this stable chair conformation, but the
ring actually executed numerous small-scale oscillations about this
average conformation. Figure 1 displays the history through a
particular MD simulation of one of the ring torsion angles, C3-C4-
C5-C6. As can be seen, this angle was clearly not constant in time
during the simulation, although its mean value was stable as the
ring oscillated about the average value.

Fluctuations in ring torsion angles such as seen in Figure 1
result in fluctuations in the overall shape, or pucker, of sugar
rings. It is possible to represent the puckering away from
planarity of a ring of atoms by a set of parameters developed by
Cremer and Pople, which describe the degree of puckering and the
conformational form of the resulting puckered structure (47).
Normal thermal motions of the type represented in Figure 1 result in
a continual fluctuation in this puckering. Larger fluctuations in
ring pucker can lead to actual transitions in ring conformation, as
from chair to boat forms. Figure 2 illustrates the evolution of one
of the ring torsion angles, C1-C2-C3-C4, for a glucose trajectory
which began with the molecule in the $^1C_4$ conformation (a discussion
of this conformational terminology can be found in reference 48),
but which underwent a transition to a twist boat form half-way
through the simulation. Figure 3 displays the history of the
Cremer-Pople pucker parameter θ, which specifies conformational
form, for the same trajectory. Values of θ around 0° correspond to
the $^4C_1$ conformation, and values around 90° correspond to the
various twist-boat forms. As can be seen from the figures, the
transition which occurred in this trajectory involved several
changes in the ring torsion angle, but because these oscillations
were accompanied by correlated changes in adjacent torsion angles,
they produced only one change in overall ring conformation, which
occurred directly and without returning to the higher energy $^1C_4$
form.

MD simulations have also demonstrated that the pendant alcohol
groups on carbohydrate rings are not rigid, but undergo frequent
rotations and orientational transitions. For example, in the same
series of simulations of α-D-glucopyranose discussed above (9), the
primary alcohol group was not found to be locked into the
crystallographic GT conformation, but instead frequently jumped
between the three major low energy conformers available to it, and
established an equilibrium distribution between these three forms
based upon their relative Boltzmann-weighted probabilities. In
addition to rotations about the C5-C6 bond, the hydroxyl groups
themselves also frequently rotate, changing their orientation.
Figures 4 and 5 displays the history of the torsional angle C4-C5-
C6-O6, specifying the conformation of the primary alcohol group, and
for the angle O5-C1-O1-H, specifying the orientation of the anomeric
hydroxyl group for a typical α-D-glucopyranose trajectory in vacuum
calculated from an MD simulation using the Rasmussen PEF422 energy
function (41). Such motions cannot be adequately studied by
conformational energy studies alone, and require dynamics
simulations for their characterization.

For D-glucose, NMR studies have demonstrated that as a result
of steric crowding, the $^1C_4$ form is not present at room temperature
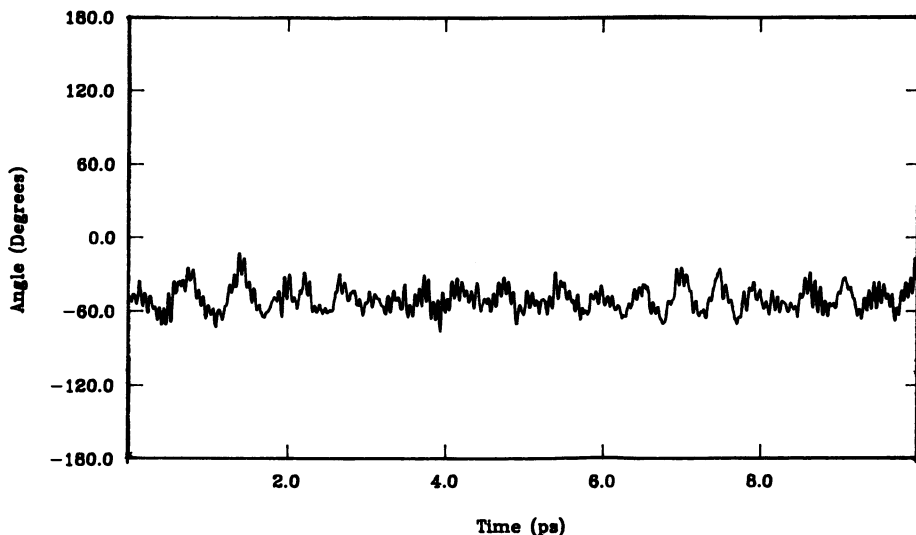
Figure 1. History of the dihedral angle C3-C4-C5-O5 calculated from a typical molecular dynamics simulation of a α-D-glucopyranose in the $^4C_1$ conformation in vacuum. (Reproduced from Ref. 9. Copyright 1986 American Chemical Society.)
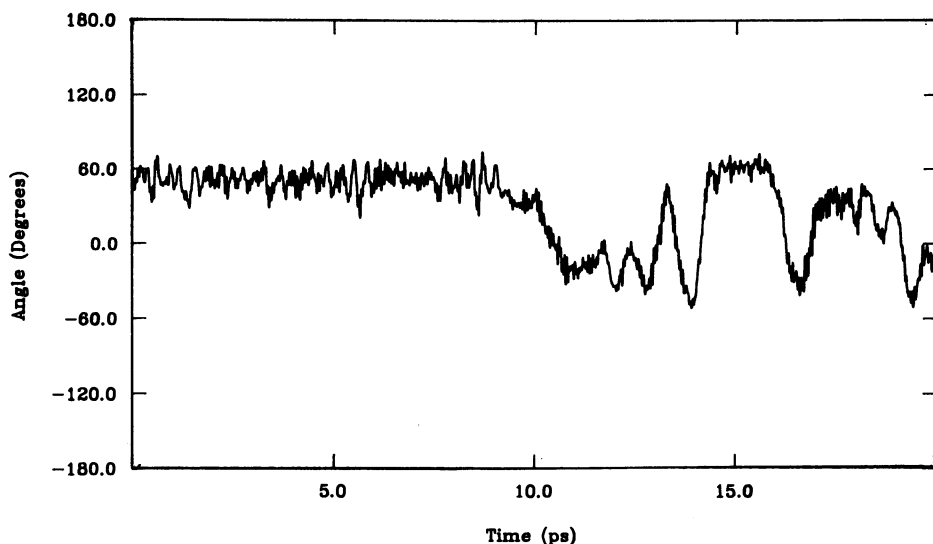
Figure 2. History of the ring torsion angle C1-C2-C3-C4 calculated from a molecular dynamics simulation of the motions of an α-D-glucopyranose molecule in vacuum which began in the $^1C_4$ conformation and which subsequently underwent a transition to a twist-boat conformation. (Reproduced from Ref. 9. Copyright 1986 American Chemical Society.)
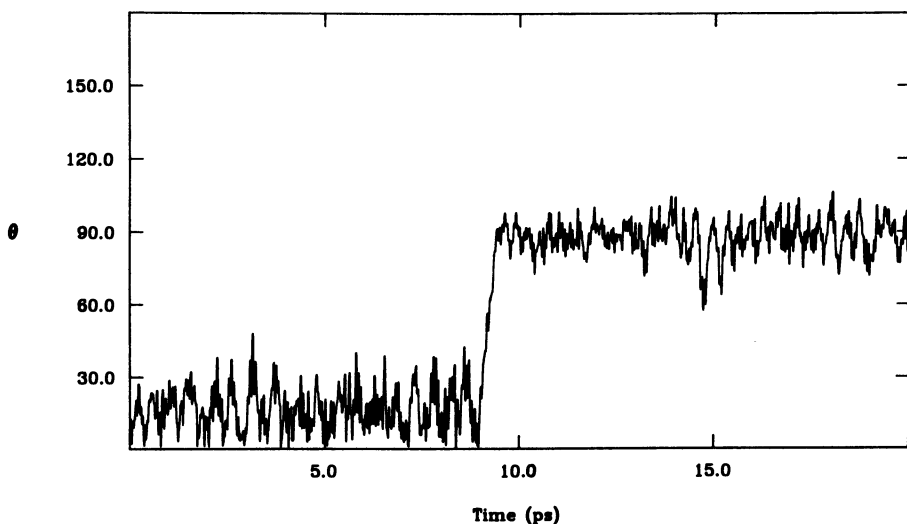
Figure 3. History of the Cremer-Pople pucker parameter θ
calculated for the same trajectory illustrated in Figure 2.
(Reproduced from Ref. 9. Copyright 1986 American Chemical
Society.)
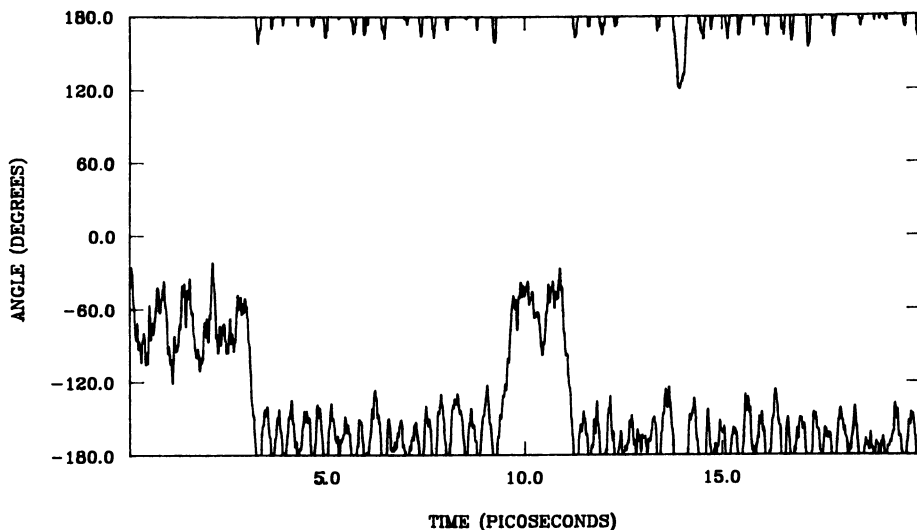
Figure 4. History of the torsion angle C4-C5-C6-O6 calculated
from a typical MD simulation of α-D-glucopyranose in vacuum
using the Rasmussen potential energy function PEF422 (41).
(Reproduced from Ref. 9. Copyright 1986 American Chemical
Society.)

to any great extent ($\underline{49}$). At higher temperatures, however, this form must occasionally arise spontaneously, as the formation of levoglucosan (1,6-anhydro-β-D-glucopyranose) demonstrates ($\underline{50}$). This molecule results from the 1,6 elimination of water from β-D-glucopyranose, which can only happen when the inversion of the ring from the usual $^4C_1$ conformation to the higher energy $^1C_4$ conformer brings these two hydroxyl groups into close proximity. Such a spontaneous ring inversion has been observed in molecular dynamics simulations of β-D-glucopyranose ($\underline{11}$) using the PEF 422 energy function ($\underline{41}$). Figure 6 illustrates the history of the Cremer-Pople pucker parameter θ from a simulation of β-D-glucopyranose at a somewhat elevated temperature (320K) which was initiated in the lowest energy $^4C_1$ conformation. After approximately 21 ps of simulation, the molecule underwent a spontaneous transition to higher energy twist-boat forms, and after approximately 60 ps, briefly converted to the still less favorable $^1C_4$ conformation, where the elimination reaction is possible, before returning to a twist-boat and finally to the original $^4C_1$ conformation. Transitions such as this in monomer ring geometries, which can arise spontaneously in MD simulations but which cannot be easily incorporated into static conformational energy calculations, could have important consequences if they occurred in polysaccharides. Such chair-to-boat transitions have been suggested to play a role in determining polysaccharide structure (end-to-end lengths) in aqueous solution ($\underline{51}$).

Solvation

Since the cost in computer time of MD simulations increases rapidly with the number of particles (atoms) in the system, many simulations of proteins and nucleic acids have studied these biopolymers in vacuum, due to the far larger cost of including solvent molecules. Many important biological processes, however, are the direct or indirect result of aqueous solvation, and simulating physical properties of this type requires the inclusion of solvent molecules ($\underline{2}$). Although some generalized solvation effects such as dielectric screening can be represented by continuum models, many of the most interesting consequences of aqueous solvation require an explicit representation of the solvent structuring immediately adjacent to the solute. For this reason, it is clearly desirable to include water molecules in MD simulations in such a way as to be physically reasonable.

In principle, it is a simple matter to include solvent water molecules directly in MD simulations, since appropriate intermolecular potential energy functions for water are available ($\underline{17},\underline{37},\underline{38}$); one would just surround the solute molecules with a sufficient number of water molecules to approximate a bulk solution. Unfortunately, a "sufficient number" of water molecules might be enormous, since many of the effects of aqueous solvation are long range or are due to entropic contributions arising from "structuring" of the solvent, which may be cooperative in nature. In any "droplet" representation of a solution, the water molecules farthest from the solvent would constitute a boundary between a bulk liquid phase and a vacuum. Since such a boundary would
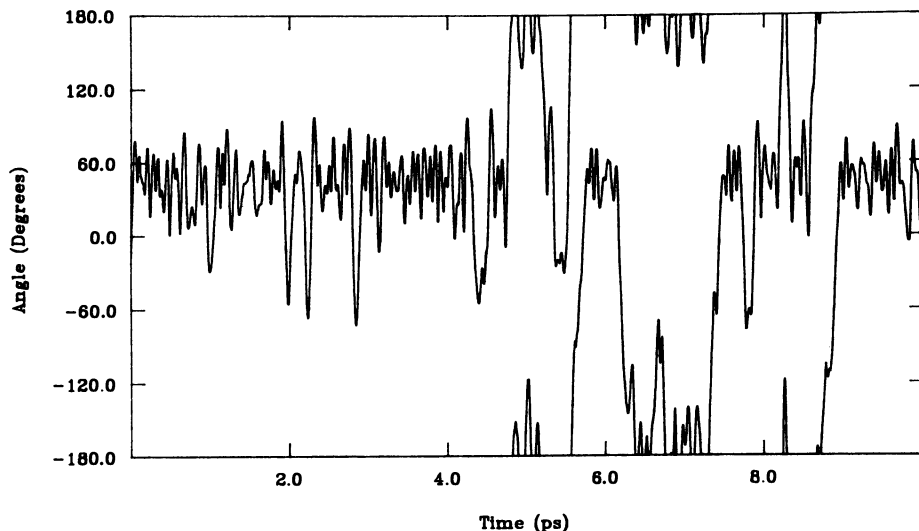
Figure 5. History of the torsion angle O5-C1-O1-H calculated from a typical MD simulation of α-D-glucopyranose in vacuum using the Rasmussen potential energy function PEF422 (41). (Reproduced from Ref. 9. Copyright 1986 American Chemical Society.)
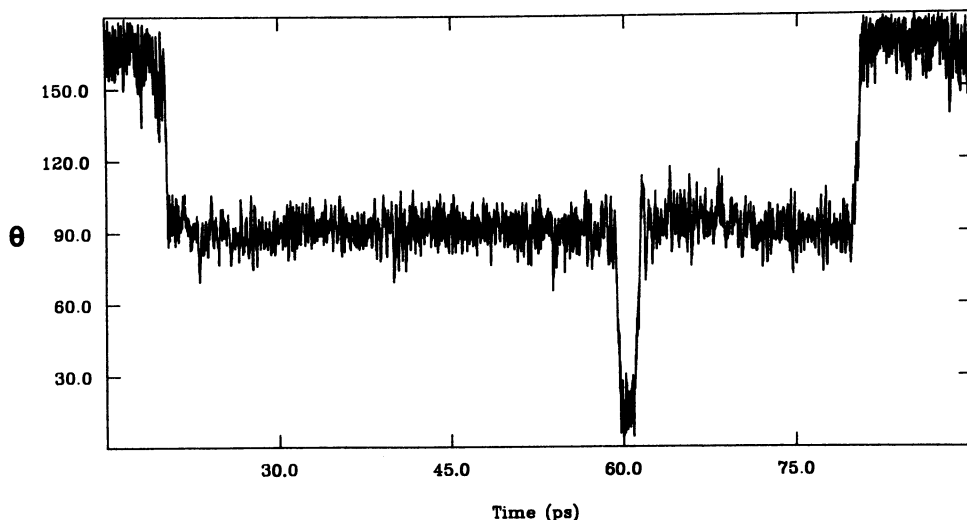
Figure 6. History of the Cremer-Pople pucker parameter θ calculated from molecular dynamics simulation of β-D-glucopyranose in the $^4C_1$ conformation in vacuum using the Rasmussen potential energy function PEF422. This molecule underwent four transitions in conformation. (Reproduced with permission from Ref. 11. Copyright 1987 Elsevier Science Publishers.)

substantially affect the properties of those solvent molecules in the boundary region, a large number of molecules would need to be included in the calculation to ensure that this perturbed boundary region was far from the solute-solvent region of interest.

In order to avoid edge effects, an approximation known as "periodic boundary conditions" (2,52) is often used, in which the solute is placed in the center of a cubic box of finite dimensions, surrounded by water molecules, and this entire box is then surrounded in every direction with exact images of itself. Atoms near to the surface of this primary box interact with image, or "ghost" water molecules in the image boxes rather than with a vacuum, eliminating direct edge effects. The equations of motion of the image particles are not integrated directly, but can be generated at each force evaluation by applying the appropriate symmetry operations on the coordinates of the atoms in the central, primary box. Usually "minimum-image" periodic boundary conditions are employed in such calculations, which means that only the closest ij interaction for two atoms i and j and all their possible images is included in the interaction forces; on the assumption that the forces decrease strongly with distance. This criterion means neglecting interactions between atoms greater than one half of the box length apart, which implies that no molecule, including the solute, interacts with its own image. The solute molecule must be covered with several layers of water molecules, such that the outer layer experiences little perturbation from bulk-like behavior due to the solute. Under these minimum-image periodic boundary conditions, the resulting system thus represents a physically unattainable state, an infinitely dilute solution at finite concentration. Such a system is a microcanonical ensemble (constant volume, constant energy, constant N).

Once the boundary conditions have been implemented, the calculation of solution molecular dynamics proceeds in essentially the same manner as do vacuum calculations. While the total energy and volume in a microcanonical ensemble calculation remain constant, the temperature and pressure need not remain fixed. A variant of the periodic boundary condition calculation method keeps the system pressure constant by adjusting the box length of the primary box at each step by the amount necessary to keep the pressure calculated from the system second virial at a fixed value (46). Such a procedure may be necessary in simulations of processes which involve large volume changes or fluctuations. Techniques are also available, by coupling the system to a Brownian heat bath, for performing simulations directly in the canonical, or constant T,N, and V, ensemble (2,46).

To date, only a few solution calculations for carbohydrates have been attempted (one such study of mannitol and sorbitol is described in the chapter by Grigera in this volume), but the results of these early studies bear out the expectation that solvation effects in carbohydrate systems can be both significant and difficult to predict. In the case of pyranoid rings, molecular solvation is further complicated by the close juxtaposition of these groups in essentially fixed relative orientations (assuming no conformational changes in the ring). Under such circumstances, molecular stereochemistry could play important physical roles, as is

illustrated by the simple hexose monosaccharides. All of these molecules are structural isomers differing only in their stereochemistries at the various ring carbon atoms, yet they exhibit different physical properties in aqueous solution, such as the characteristic equilibrium anomeric ratios (31,50), indicating that these properties are determined at least in part by aqueous solvation.

As an example of the possible importance of the special nature of aqueous solvation in determining the properties of carbohydrate solutions, it is worthwhile to consider the results of recent MD simulations of α-D-glucopyranose in aqueous solution (32). In this study, a minimum-image periodic boundary conditions MD simulation was conducted for a single α-D-glucopyranose molecule surrounded by 207 SPC water molecules (38) at a temperature of 300K. This number of water molecules in the primary box is sufficient for two complete solvation shells around the solute, and a portion of a third shell. The simulation was run for 10.2 ps of equilibration and an additional 32 ps of data collection, using an integration step size of 1 fs. Although this run was apparently not long enough to allow complete thermodynamic convergence of all of the physical properties examined, it was quite lengthy in terms of computer time, requiring more than 40 days of DEC VAX 11/750 CPU time, demonstrating one of the difficulties of solution calculations.

The presence of aqueous solvent was found to have little effect upon the mean structure of the pyranoid ring in these MD simulations, with only slight deviations in the time-averaged structure away from that observed in vacuum simulations or in the crystallographic diffraction experiments (32). However, the presence of the solute had substantial effects upon the average "structuring" of the solvent. Figure 7 displays a pair distribution function g(r), defined as (27)

$$g(r) = \frac{1}{4\pi\rho r^2} \frac{dN(r)}{dr} \tag{7}$$

which is the normalized probability of finding the oxygen atom of a solvent water molecule a given distance r from some particular atom in the sugar solute molecule, where ρ is the bulk number density. The g(r) illustrated in Figure 7 shows the probability of finding a water oxygen atom as a function of distance from the glucose C6 carbon atom as calculated from the simulations. The form of this curve is typical of hydrophobic or nonpolar hydration, with a broad peak centered around 3.4 Å, the approximate van der Waals contact distance, representing the first solvation shell water molecules. This type of nonpolar distribution function contrasts with that found in the case of hydrogen bonding, as illustrated in Figure 8, which displays the pair distribution function for water oxygen atoms around the O6 hydroxyl oxygen atom. The first peak in this curve occurs at a much closer distance, 2.7 Å, which is typical of hydrogen bonding, since the nearest neighbors of this hydroxyl group will be those water molecules which are hydrogen bonded to it. This much higher and quite narrow first peak, with a deep first minimum, indicates that these hydrogen bonded water molecules are very
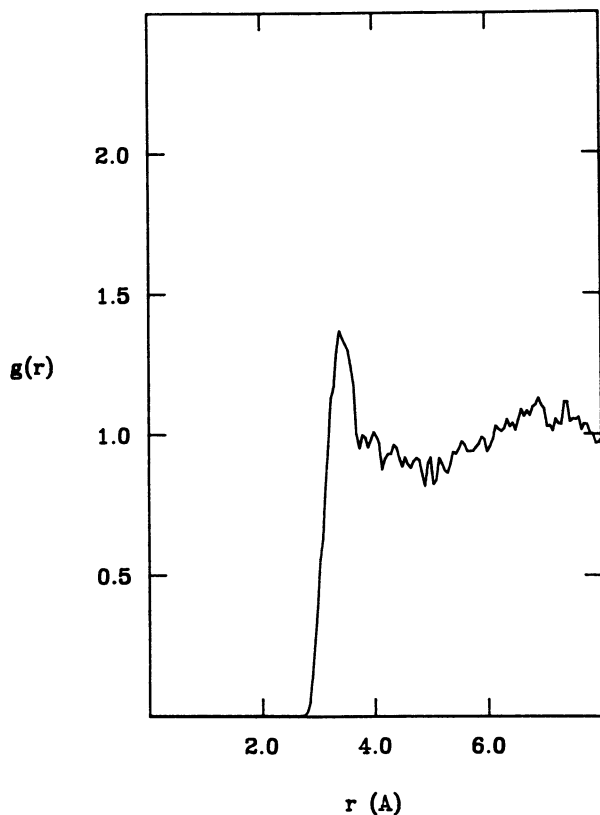
Figure 7. Water oxygen-exocyclic methylene carbon pair
distribution function, calculated from a molecular dynamics
simulation of α-D-glucopyranose in aqueous solution, giving the
normalized probability of finding a water oxygen atom a given
distance r from the C6 carbon atom. (Reproduced from Ref. 32.
Copyright 1989 American Chemical Society.)
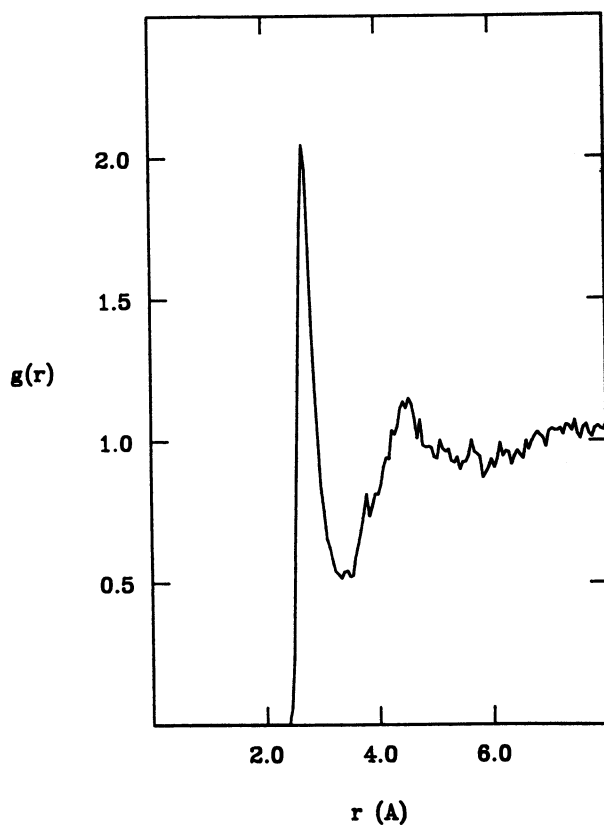
Figure 8. Water oxygen-hydroxyl oxygen O6 pair distribution function, calculated from a molecular dynamics simulation of α-D-glucopyranose in aqueous solution. (Reproduced from Ref. 32. Copyright 1989 American Chemical Society.)

localized spatially by the requirements of the hydrogen bond, which balances strong electrostatic attractions with van der Waals repulsions. Figure 9 illustrates the pair distribution function for the water molecules around the ring oxygen; the breadth of the first peak in this function, centered around 3.25 Å, and its low maximum height, indicate that this atom is not significantly hydrogen bonded to its nearest water molecule neighbors, due to its lower charge and lack of a hydrogen atom substituent.

In addition to imposing spatial restrictions on the distributions of adjacent water molecules, the various groups in the solute sugar molecule also impose orientational structuring upon these solvent molecules. Figure 10 displays the distributions of orientations for water molecules around the methylene carbon C6. This function is the integrally normalized probability for each water molecule being oriented such that it makes an angle θ between its OH bond vectors and the vector from the water oxygen to the carbon atom. This function is calculated for those molecules within 4.9 Å of the carbon atom (nearest neighbors), as this distance marks the first minimum in the pair distribution function for that atom. The curve in Figure 10 is typical for hydrophobic hydration (27). The peak at $\cos(\theta)=1$ corresponds to one hydroxyl group pointed directly away from the nonpolar group, and the value of nearly zero at -1.0 indicates virtually no probability of a hydroxyl group pointing directly at the methylene group. The broad peak around -0.33, the tetrahedral angle, is a consequence of the tetrahedral structure of the SPC water molecule; if one of the hydroxyl groups is pointing directly away from the CH_2 group, then the other must be making a tetrahedral angle with the CO vector. This type of orientational structuring, somewhat similar to a fluctuating clatherate, is adopted as the best way to solvate such a small nonpolar species, as each water molecule can still make hydrogen bonds to other water molecules and thus avoid the high energetic cost of losing a hydrogen bond (27).

Figure 11 displays the same type of orientational distribution function for water molecules adjacent to the O3 hydroxyl oxygen atom. As can be seen, in the case of normal hydrogen bonding, the inverse behavior is seen, with one of the water hydroxyl groups pointing directly at the solute hydroxyl oxygen atom. The poorly hydrogen-bonding ring oxygen atom imposes little orientational structuring, as can been seen from Figure 12. Surprisingly, however, this orientational distribution function for those water molecules around the O2 hydroxyl oxygen atom, illustrated in Figure 13, was substantially perturbed in these simulations. This perturbation was apparently due to the interference in orientational structuring caused by other adjacent groups, perhaps in particular the requirements of the two CH groups 1 and 2 which constitute an extended region of hydrophobicity, which is closer to the O2 hydroxyl group than to the axial O1 group in the alpha anomer (see Figure 14). Preliminary simulations using other water models indicate that this orientational perturbation is indeed related to the anomeric preference, and is apparently involved in the solvent-induced anomeric distributions observed in aqueous solution (31,50). The solvation requirements of the specific spatial arrangements resulting from particular stereochemistries thus give rise to the
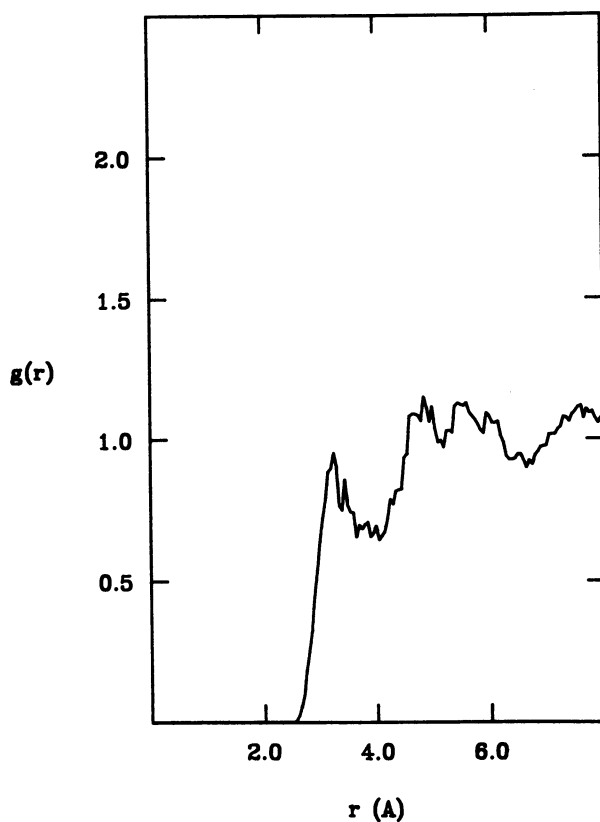
Figure 9. Water oxygen-ring oxygen pair distribution function, calculated from a molecular dynamics simulation of α-D-glucopyranose in aqueous solution. (Reproduced from Ref. 32. Copyright 1989 American Chemical Society.)
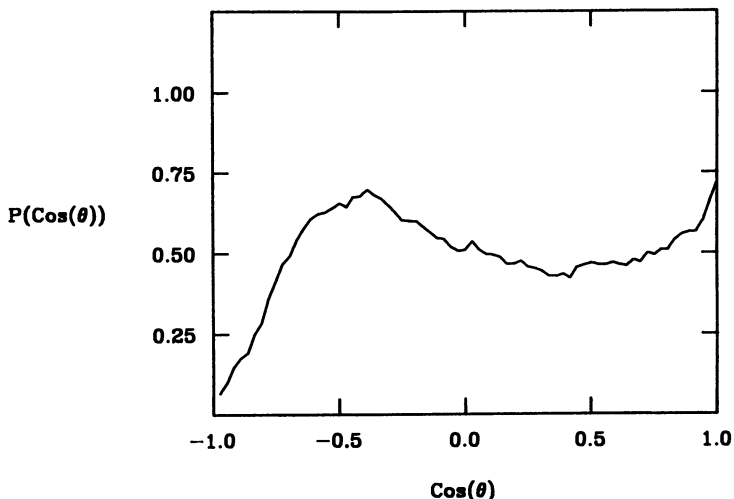
Figure 10. Distribution of orientations for water molecules
adjacent to the exocyclic methylene carbon atom C6 as calculated
from a molecular dynamics simulation of α-D-glucopyranose in
aqueous solution. The function plots the frequency of
occurrence of an angle θ between the water OH bond vectors and
the vector from the carbon atom to the water oxygen atom. A
value of cos(θ) of 1.0 corresponds to an OH bond vector pointing
directly away from the carbon atom. (Reproduced from Ref. 32.
Copyright 1989 American Chemical Society.)
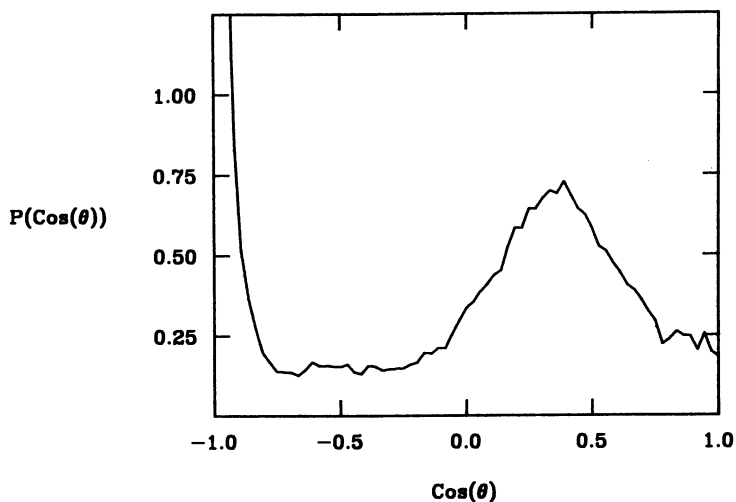
Figure 11. Distribution of orientations for water molecules
adjacent to the hydrogen-bonding O3 hydroxyl oxygen atom, as
calculated from an MD simulation of α-D-glucopyranose, as in
Figure 10. (Reproduced from Ref. 32. Copyright 1989 American
Chemical Society.)
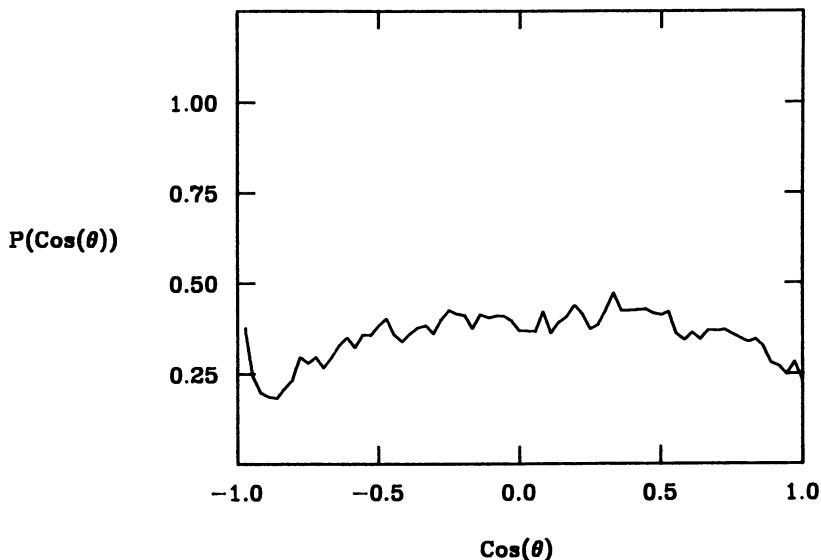
Figure 12. Distribution of orientations for water molecules adjacent to the ring oxygen atom of α-D-glucopyranose in aqueous solution, calculated as in Figure 10. (Reproduced from Ref. 32. Copyright 1989 American Chemical Society.)
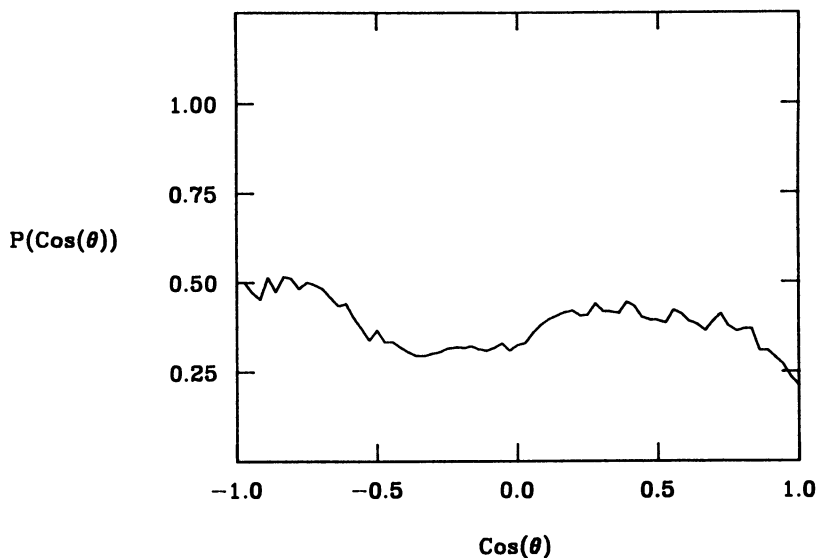
Figure 13. Distribution of orientations for water molecules adjacent to the O2 hydroxyl oxygen atom of α-D-glucopyranose in aqueous solution, calculated as in Figure 10. (Reproduced from Ref. 32. Copyright 1989 American Chemical Society.)
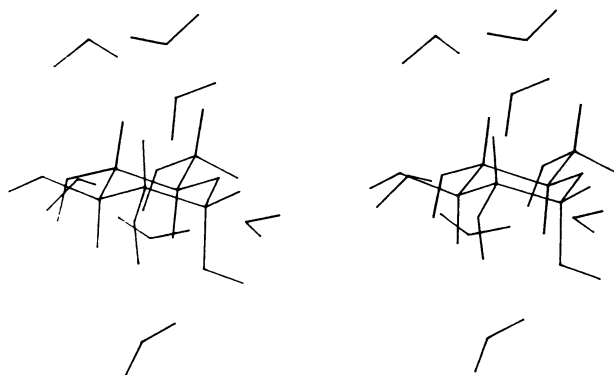
Figure 14. A stereoview of a typical "snapshot" from a
molecular dynamics simulation of α-D-glucopyranose in aqueous
solution, showing selected water molecules close to the O1 and
O2 hydroxyl groups. The hydrophobic hydration requirements of
the aliphatic hydrogen atoms on C1 and C2 impose additional
structuring restrictions of the orientations of the water
molecules hydrogen bonding to the O2 hydroxyl group.
(Reproduced from Ref. 32. Copyright 1989 American Chemical
Society.)

differences in solution properties of the various sugars. For this
reason, it may often be necessary to include solvent in theoretical
treatments of carbohydrate molecules.

Conclusions

Although they have only recently been applied to carbohydrate
problems, molecular dynamics simulations should contribute
substantially to our understanding of a variety of structural and
conformational phenomena. MD calculations can help to quantify
rates for a variety of transition processes, which other statistical
mechanical simulation techniques (Monte Carlo) cannot explore. MD
simulations may also help resolve "virtual" or time-averaged
structures in disaccharides which oscillate between one or more low-
energy conformations (described in the chapter by Carver, et al. in
this volume). More ambitious MD simulations may help answer old
questions about the conformation of polysaccharides such as amylose
and any possible role which random changes in ring conformation may
play. Because of the importance of aqueous solvation in many
biological systems, including the carbohydrates, theoretical studies
which include solvent in an attempt to understand the exact role of
solvent in these systems will be necessary. Again, MD simulations
offer an excellent method for including solvation directly and
examining solution behavior at a level of detail which is not
possible in experiment. Presumably, MD simulations of carbohydrates
will soon become as commonplace as are conformational energy studies
of these molecules.

Literature Cited

1) Burkert, U.; Allinger, N.L. Molecular Mechanics, ACS Monograph 177, American Chemical Society, Washington, 1982.
2) Brooks, C.L.; Karplus, M.; Pettitt, B.M. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. Advances in Chemical Physics, Wiley Interscience: New York, 1988, Vol. LXXI.
3) Ramachandran, G.N.; Ramakrishnan, C.; Sasisekharan, V. J. Mol. Biol. 1963,7,95.
4) Rao, V.S.R.; Sundararajan, P.R.; Ramakrishnan, C.; Ramachandran, G.N. in Conformation in Biopolymers, Vol. 2, G.N. Ramachandran, ed., Academic Press, London, 1967.
5) Brant, D.A. Ann. Rev. Biophys. Bioeng. 1972,1,369.
6) Wu, H.-C.H.; Sarko, A. Carbohydr. Res. 1978, 61, 7.
7) Imberty, A.; Perez, S. Biopolymers 1988, 27, 1205.
8) McCammon, J.A.; Gelin, B.R.; Karplus, M. Nature 1977,267,585.
9) Brady, J.W. J. Am. Chem. Soc. 1986,108,8153.
10) Post, C.B.; Brooks, B.R.; Karplus, M.; Dobson, C.M.; Artymiuk, P.J.; Cheetham, J.C.; Phillips, D.C. J. Mol. Biol. 1986,190,455.
11) Brady, J.W. Carbohyd. Res. 1987,165,306.
12) Prabhakaran, M.; Harvey, S.C. Biopolymers 1987,26,1087.
13) Grigera, J.R. J. Chem. Soc., Faraday Trans. I, 1988,84,2603.
14) Goebel, C.V.; Dimpfl, W.L.; Brant, D.A. Macromolecules 1970, 3,644.
15) Joshi, N.V.; Rao, V.S.R. Biopolymers 1979,18,2993.
16) French, A.D.; Murphy, V.G. Carbohydr. Res. 1973,27,391; Polymer 1977,18,489.
17) Stillinger, F.H.; Rahman, A. J. Chem. Phys. 1974,60,1545.
18) Dashevsky, V.G.; Sarkisov, G.N. Mol. Phys. 1974,27,1272.
19) Owicki, J.C.; Scheraga, H.A. J. Am. Chem. Soc. 1977,99,7413.
20) Swaminathan, S.; Harrison, S.W.; Beveridge, D.L. J. Am. Chem. Soc. 1978,100,5705.
21) Palinkas, G; Riede, W.O.; Heinzinger, K. Z. Naturforsch. 1977,32a,1137.
22) Dang, L.X.; Pettitt, B.M. J. Chem. Phys. 1987,86,6560.
23) Okazaki, S.; Nakanishi, K.; Touhara, H. J. Chem. Phys. 1984,81,890.
24) Tanaka, H.; Nakanishi, K.; Touhara, H. J. Chem. Phys. 1984,81,4065.
25) Kuharski, R.A.; Rossky, P.J. J. Am. Chem. Soc. 1984,101,5794.
26) Tanaka, H.; Nakanishi, K.; Touhara, H. J. Chem. Phys. 1985,82,5184.
27) Rossky, P.J.; Karplus, M. J. Am Chem. Soc. 1979,101,1913.
28) Ravishanker, G.; Mezei, M.; Beveridge, D.L. J. Comput. Chem. 1986,7,345.
29) Ahlstrom, P.; Teleman, O.; Jönsson, B. J. Am. Chem. Soc. 1988,110,4198.

30) Suggett, A. in Water: A Comprehensive Treatise, Vol. 4, F. Franks, ed., Plenum, New York, 1975, pp. 519-567.

31) Franks, F. Pure and Appl. Chem. 1987,59,1189.

32) Brady, J.W. J. Am Chem. Soc. 1989,111,5155.

33) Rahman, A. Phys. Rev. 1964,136,A405.

34) Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. J. Comput. Chem. 1983,4,187.

35) Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. J. Comp. Chem. 1986,7,230.

36) van Gunsteren, W.F.; Berendsen, H.J.C.; Hermans, J.; Hol, W.G.J.; Postma, J.P.M., Proc. Natl. Acad. USA 1983,80,4315.

37) Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. J. Chem. Phys. 1983,79,926.

38) Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; Hermans, J. in Intermolecular Forces, B. Pullman, ed., Reidel, Dordrecht, 1981,331.

39) Jorgensen, W.L., J. Am. Chem. Soc., 1981,103,335; Ibid. 1981,103,341.

40) Ha, S.N.; Giammona, A.; Field, M.; Brady, J.W. Carbohydr. Res., 1988,180,207.

41) Rasmussen, K. Acta Chem. Scand., Ser. A 1982,36,323, and references therein.

42) Koeler, J.E.H.; Saenger, W.; van Gunsteren, W.F. Eur. Biophs. J. 1987,15,197.

43) Goldstein, H. Classical Mechanics, 2nd Edition, Addison-Wesley, Reading, MA, 1980.

44) McQuarrie, D.A. Statistical Mechanics, Harper and Row, New York, 1976.

45) Verlet, L. Phys. Rev. 1967,159,98.

46) McCammon, J.A.; Harvey, S.C. Dynamics of Proteins and Nucleic Acids, Cambridge University Press, Cambridge, 1987.

47) Cremer, D.; Pople, J.A. J. Am Chem. Soc. 1975,97,1354.

48) Stoddart, J.F. Stereochemistry of Carbohydrates, Wiley-Interscience, New York, 1971.

49) Perkins, S.J.; Johnson, L.N.; Phillips, D.C.; Dwek, R.A. Carbohydr. Res. 1977,59,19.

50) Shallenberger, R.S. Advanced Sugar Chemistry, AVI Publishing, Westport, CT, 1982.

51) Goebel, K.D.; Harvie, C.E.; Brant, D.A. Appl. Polym. Symp. 1976,28,671.

52) Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. J. Chem. Phys. 1953,21,1087.

RECEIVED March 21, 1990

Chapter 6

Ab Initio Molecular Orbital Calculations on Carbohydrates

Conformational Properties of Deoxygenated Furanose Sugars

Eugenia C. Garrett and Anthony S. Serianni[1]

Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556

Ab initio molecular orbital calculations have been conducted on the four deoxytetro-furanoses, 2-deoxy-α- and β-D-*glycero*-tetro-furanoses and 3-deoxy-α- and β-D-*glycero*-tetrofuranoses, to assess the effect of furanose ring conformation on structural parameters (e.g., bond lengths, angles and torsions) and on total energies. Geometric optimizations of the planar and ten non-planar (envelope) forms of each compound were performed using the STO-3G and 3-21G basis sets, allowing a full comparison of results and a general assessment of the potential errors and limitations associated with calculations of intact carbohydrates using these basis sets. A limited inspection of more extended basis sets (e.g., 6-31G*) was also conducted. Proposed models for the conformational dynamics of the four deoxy-tetrofuranoses are evaluated in light of calculations conducted previously on the structurally-related D-aldotetrofuranoses, yielding important information on the effect of ring deoxygenation on furanose conformational behavior.

The conformational properties of furanose rings have received considerable attention in recent years because of the impact these properties may have in mediating biological processes[1-8]. Most notable in this respect are the β-D-ribofuranose **1** and 2-deoxy-β-D-*erythro*-pentose **2** (Scheme 1) components of ribo- (RNA) and deoxyribonucleic (DNA) acids. It is well known that the furanose ring adopts specific shapes depending on its local structural environment in a biopolymer. For example, in tRNA, the

---

[1]To whom correspondence should be addressed.

ribofuranose ring prefers a C3'-endo conformation in helical segments, whereas the C2'-endo conformer is commonly observed in loop regions[9]. In DNA, the deoxyribose ring assumes a C3'-endo conformation in the A-form, whereas in B-DNA, the C2'-endo or C3'-exo conformers are preferred[10]. Thus, the deoxyribose ring experiences significant conformational change in the interconversion of A-DNA and B-DNA, a process that presumably occurs in vivo. An appreciation of the factors that determine furanose ring conformational dynamics is a prerequisite to understanding the energetics of DNA and RNA conformational interconversion in solution. Furanose rings are also found as components of biologically-important polysaccharides, although their role in determining the overall conformational properties of these biopolymers has not been studied very extensively.

The conformational dynamics of furanose rings may be described by the mechanisms of pseudorotation[11-14] and inversion. The former mechanism describes a continuous pathway of interconversion between twenty idealized non-planar (envelope, twist) conformers (Figures 1 and 2) that does not involve the planar form (e.g., $^3E \rightarrow E_4 \rightarrow {}^OE$). Inversion describes interconversion between non-planar forms via the planar form (e.g., $^3E \rightarrow$ planar $\rightarrow {}^OE$). It is not clear whether one or both of these mechanisms play a role in DNA and RNA conformational dynamics in vivo, although it is generally held that barriers to conformer interconversion are low[15].

Although frequently employed to assess furanose conformation in solution, experimental NMR parameters such as chemical shifts and spin-coupling constants are not unequivocal in establishing preferred furanose geometries because of the effect of conformational averaging on these values[16]. Furthermore, NMR cannot address the issue of conformer energetics. At present, therefore, it appears that the best approach to evaluate furanose conformational dynamics is one that employs calculational and experimental components. Several empirical and semi-empirical calculational studies have been reported on the relative flexibilities of the furanose ring in DNA and RNA[1,17-19], yielding conflicting conclusions. For example, Levitt and Warshel[1] have proposed a flexible model for deoxyribose dynamics in DNA, whereas a more rigid model is preferred by Olson and Sussman[18]. It is evident that the conformational dynamics of furanose rings is not completely understood at present, and that the vast majority of the reported studies have been confined to **1** and **2** because of their obvious biological roles. It is our contention that experimental(e.g., NMR) and computational studies of other ring configurations (e.g., *arabino*, *lyxo*, *xylo*) are essential to attaining a global understanding of the structural behavior of these rings.
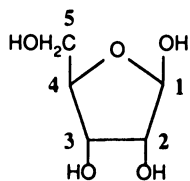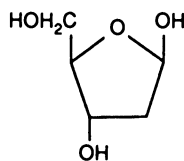
β-D-ribofuranose *1*   2-deoxy-β-D-erythropentose *2*

Scheme 1
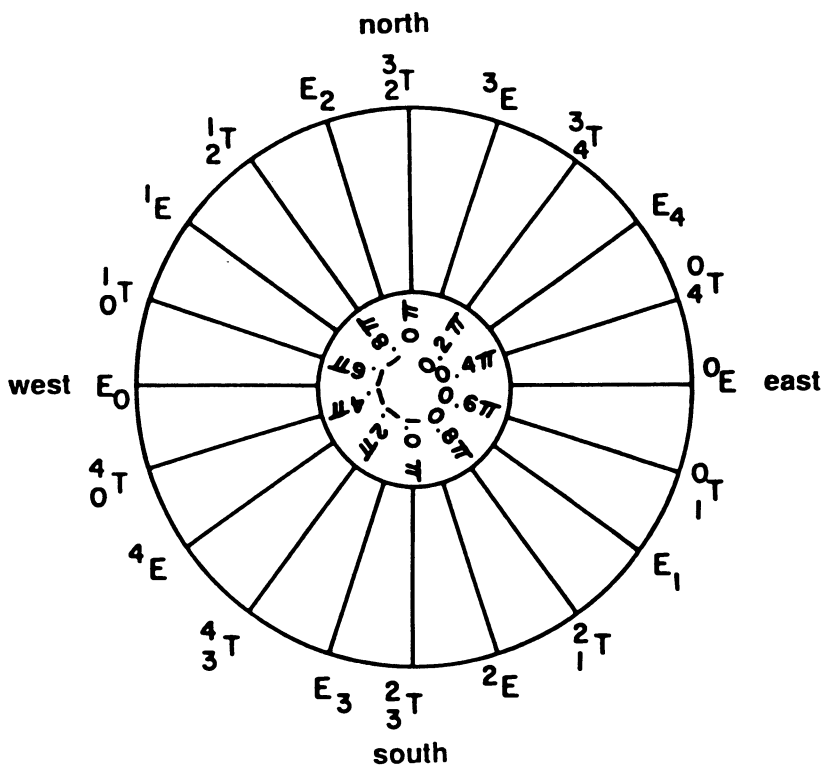
Figure 1. The pseudorotational itinerary[12-14] describing the interconversion of non-planar furanose conformers. Regions of the itinerary are denoted as north, south, east and west as indicated. Envelope and twist conformers are denoted by E and T, respectively.

We have been interested in applying *ab initio* molecular orbital calculations to furanose rings in an attempt to better define their structures and conformational properties. In a recent study[20], we examined the tetrofuranoses, α- and β-D-erythrofuranose (**3**, **4**) and α- and β-D-threofuranose (**5**, **6**) (Scheme 2) using the Gaussian 80 program developed by Pople and coworkers[21]. Complete geometric optimizations of the planar and ten envelope forms were performed on each isomer with the STO-3G basis set, and refined with single-point 3-21G calculations. While this work revealed several interesting findings, its obvious shortcoming was the choice of basis set. As a consequence, the present study was initiated to address two problems: (1) to examine the effect of basis set on calculated furanose geometries and total energies, and (2) to examine the effect of ring deoxygenation on furanose conformation and dynamics. We have chosen two deoxyfuranoses as model systems, 2-deoxy-α- and 2-deoxy-β-D-*glycero*-tetrofuranose (**7**, **8**) and 3-deoxy-α- and 3-deoxy-β-D-*glycero*-tetrofuranose (**9**, **10**) (Scheme 3), which are the monodeoxy analogues of the tetrofuranoses **3-6**. Complete geometric optimizations have been performed on eleven conformers of each compound (10 envelope, 1 planar) using the STO-3G and 3-21G basis sets for comparative purposes, and optimized geometric parameters (bond lengths, angles and torsions) and total energies are tabulated and discussed in the context of the pseudorotation and inversion models.

Experimental

The Gaussian 80 program[21], as implemented on an IBM 370/3033 mainframe computer at the Notre Dame Computing Center, was used for most of the calculations. Calculations were also conducted with the Gaussian 86 program[22a] as implemented on a Digital VaxStation 3200 computer. Geometric optimizations were performed with the minimal STO-3G basis set[23, 24] and the split-valence 3-21G basis set[25]. Computations were performed on ten envelope (E) forms (Figure 2), each with one appropriate endocyclic torsion angle fixed at 0° (to maintain a given envelope form), while all remaining molecular parameters were optimized by analytic gradient methods; for planar forms, two endocyclic torsion angles were fixed at 0°. Initial estimates of structural parameters (bond lengths, angles and torsions) were made by inspection of crystallographic data[26, 27]. Geometry optimizations required about 6 h and 12 h of cpu time per conformer for the STO-3G and 3-21G basis sets, respectively, on the IBM computer.
    Several calculations using Gaussian 86[22a] and Gaussian 88[22b] were performed at the 4-31G, 4-31G*, 6-31G and 6-31G* levels on the planar form of 2-deoxy-α-D-

^{1}E (envelope) conformation ^{2}T$_1$ (twist) conformation
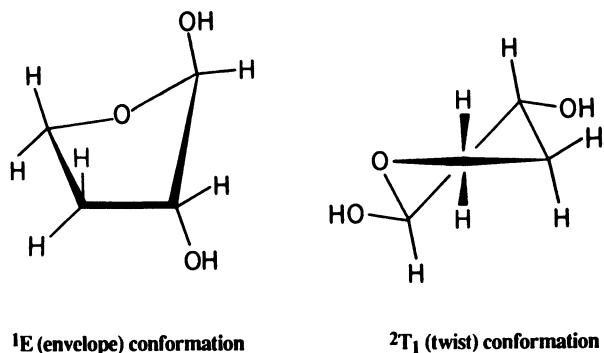
Figure 2. The two classes of non-planar furanose conformers of 3-deoxy-β-D-*glycero*-tetrofuranose **10**. The ^{1}E (envelope) conformer has C2, C3, C4 and O4 coplanar and C1 out-of-plane. The ^{2}T$_1$ (twist) conformer has C3, C4 and O4 coplanar and C1 and C2 out-of-plane.

α-D-erythrofuranose *3*   β-D-erythrofuranose *4*   α-D-threofuranose *5*   β-D-threofuranose *6*

Scheme 2

2-deoxy-α-D-glycero-tetrofuranose **7**   2-deoxy-β-D-glycero-tetrofuranose *8*

3-deoxy-α-D-glycero-tetrofuranose *9*   3-deoxy-β-D-glycero-tetrofuranose *10*
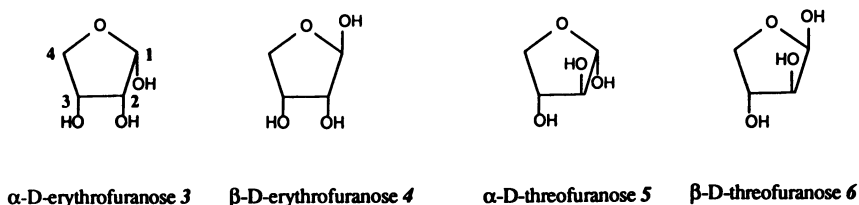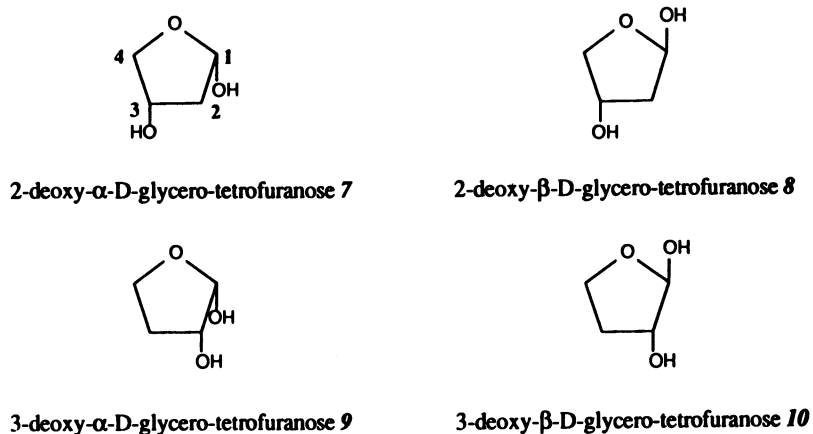
Scheme 3

glycero-tetrose **7** to estimate the effect of extended
basis sets on optimized molecular parameters.

Theoretical calculations of sugars are complicated
by the choice of C-O bond rotamers for the hydroxy groups
in the molecule. It was impractical to investigate all
rotamer combinations for each conformer of **7-10**, as this
would require nine optimizations (3^2) per conformer. The
choice of C-O rotamers, therefore, was made by model
inspection with the aim of minimizing intramolecular
hydrogen bonding and optimizing stereoelectronic effects
at C1 (Figure 3). In the latter regard, our previous *ab
initio* calculations have shown[20] that the most stable C1-
O1 rotamer is that having OH-1 *gauche* to H1 and the ring
oxygen, as expected from the "exoanomeric effect"[28, 29];
these observations were verified in the deoxyfuranoses.
The implications of this approach are discussed in more
detail in the Results section.

The nomenclature used in this paper to describe
furanose ring conformation derives from the work of
Sundaralingam and coworkers[13, 15] in which a pseudoro-
tational pathway describes the interconversion of non-
planar conformers (Figure 1). Two parameters, P (phase
angle) and τ_m (puckering amplitude), are required to
define the complete ring structure of a conformer. The
relationship between furanose ring conformation and P is
illustrated in Figure 1, where, for example, the ^3E
conformation corresponds to P=0.1π. To simplify the
presentation of data, conformers are identified by P/π,
where ^3E corresponds to a value of P/π=0.1, E$_4$ to a value
of 0.3, and so forth.

Results

A. Bond Lengths. Previous *ab initio* calculations with
the STO-3G basis set on the tetrofuranoses **3-6** showed
that endocyclic C-C and C-O bond lengths vary
systematically with ring conformation[20]. The three
endocyclic C-C bonds in the deoxytetrofuranoses **7-10** show
similar cyclic behavior (Figure 4A, 4C); for example, the
C1-C2 bond length is maximal at 0.3 and 1.3 P/π (i.e., in
conformations where substituents on C1 and C2 are
eclipsed) and minimal at 0.9 and 1.9 P/π (i.e., in
conformations where substituents on C1 and C2 are
maximally staggered). Curves observed for C2-C3 and C3-
C4 bond lengths are similar in shape but are phase-
shifted relative to that for C1-C2 by 0.2 P/π and 0.4
P/π, respectively. Similar curves are obtained with the
3-21G basis set (Figure 4B, 4D), although bond lengths
are shorter and curve amplitudes greater with this basis
set. In general, anomeric configuration does not affect
the response of endocyclic C-C bond length to ring
conformation (Figure 4B, 4D).

In contrast to C-C bond lengths, the response of
endocyclic C-O bonds in **7-10** depends on ring conformation
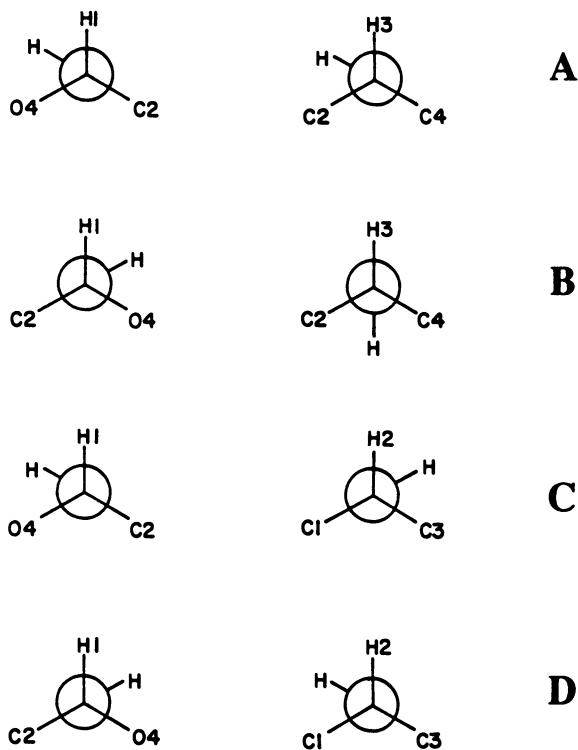
Figure 3. The initial exocyclic C-O rotamers used for conformational energy calculations on the deoxytetro-furanoses **7** (A), **8** (B), **9** (C) and **10** (D). The C1-O1 rotamers were chosen to optimize the "exoanomeric effect"[28,29].
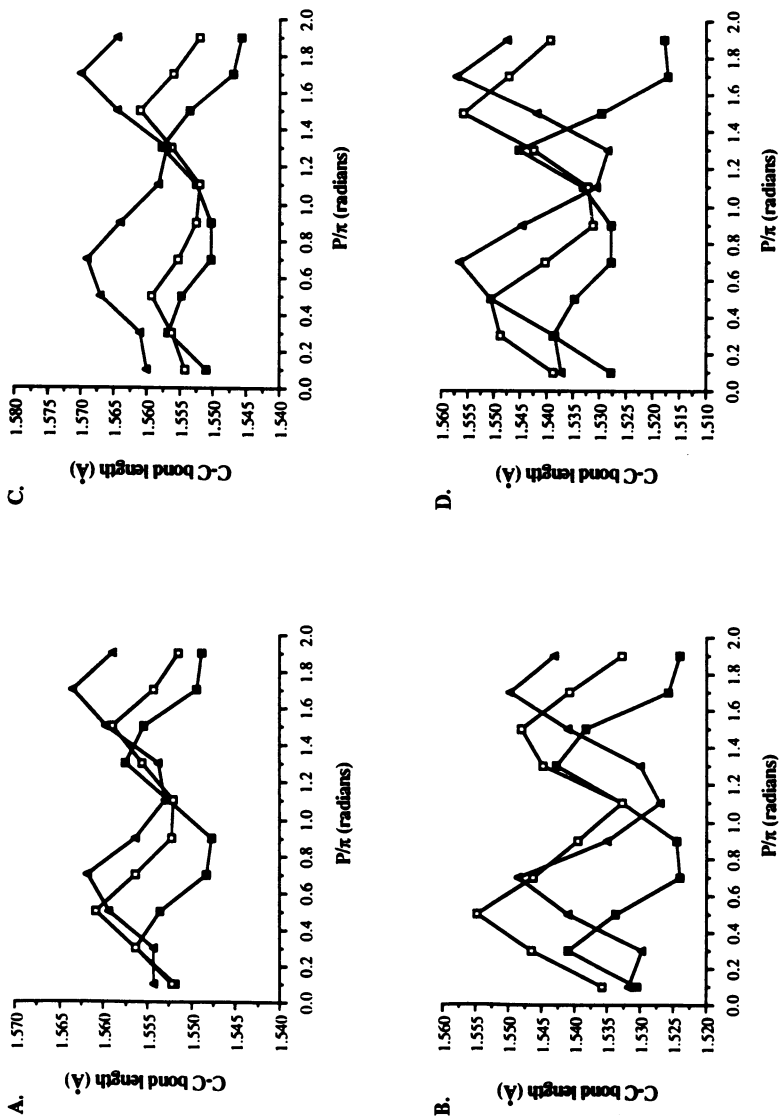
Figure 4. The effect of ring conformation on furanose ring endocyclic C-C bond lengths in **7** (A and B) and **8** (C and D) using the STO-3G (A and C) and 3-21G (B and D) basis sets. C1-C2 (■), C2-C3 (□), C3-C4 (▲).

and anomeric configuration. Computations at the 3-21G level show that, for α-anomers, the C1-O1 bond is shortest at about 1.7 P/π and longest at about 0.5 P/π (Figure 5B), whereas the opposite is found for β-anomers (Figure 5D). A similar pattern is observed for the C4-O4 bond. In contrast, the C1-O4 bond in α-anomers is shortest at 0.7 P/π and longest at 0.1 P/π, with a local minimum at 1.5 P/π; for β-anomers, the curves are similar except that the global minimum occurs at about 1.5 P/π and the local minimum at 0.7 P/π (Figure 5B, 5D). Comparison of STO-3G and 3-21G basis sets (Figure 5A-D) shows that curve shapes are conserved, but C-O bonds decrease overall in length and the magnitude of bond length change is enhanced in the 3-21G calculations. Relative bond lengths also change with basis set as shown in Figure 5A-D; the more reliable 3-21G basis set shows that C4-O4 > C1-O4 > C1-O1 for all furanose conformations.

STO-3G calculations on the tetrofuranoses **3-6** showed that C-H bonds in the vicinity of the ring oxygen (e.g., C1-H1, C4-H4*R*, C4-H4*S*) depend on ring conformation[20]. The observed dependence on conformation was explained by postulating that these C-H bonds increase in length as they become more antiperiplanar to a lone-pair orbital of the ring oxygen[20]. Similar behavior is observed in the deoxytetrofuranoses **7-10** with the STO-3G and 3-21G basis sets (Figures 6 and 7). Curves obtained with computations using the split-valence basis set, however, are shifted to shorter bond lengths and generally have greater amplitudes. It is also interesting to note that the calculated C1-H1 bond length is significantly longer than the C4-H4*R* and C4-H4*S* bonds using the STO-3G basis set, whereas all three bonds are comparable in length with the 3-21G basis set.

B. Coordinated Bond Lengths In the Vicinity of the Anomeric Center. In D-aldofuranoses, the C1-O1 bond is quasi-axial in ^{O}E-E_1 conformers (0.7 P/π) of α-anomers, and E_O-^1E conformers (1.7 P/π) of β-anomers (Figure 1). In contrast, the C1-O1 bond assumes a quasi-equatorial orientation in E_O-^1E and ^{O}E-E_1 conformers of α- and β-anomers, respectively. The orientation of the C1-O1 bond is expected to have a profound effect on ring electronic structure in the vicinity of the anomeric center. An inspection of C-O and C-H bond lengths in the vicinity of the anomeric carbon (e.g., C4-O4, O4-C1, C1-O1, C1-H1) for quasi-axial and quasi-equatorial orientations of the C1-O1 bond (Figures 5-7) reveals several interesting relationships. When the C1-O1 bond is quasi-axial, the C1-O1 and C4-O4 bonds are maximal or near maximal in length, whereas the O4-C1 and C1-H1 bond lengths are at or near their minimum values (Scheme 4A). In contrast, when the C1-O1 bond is quasi-equatorial, C1-O1 and C4-O4 bond lengths are at or near minima, and C1-O4 and C1-H1 bonds are maximal or near maximal in length (Scheme 4B).
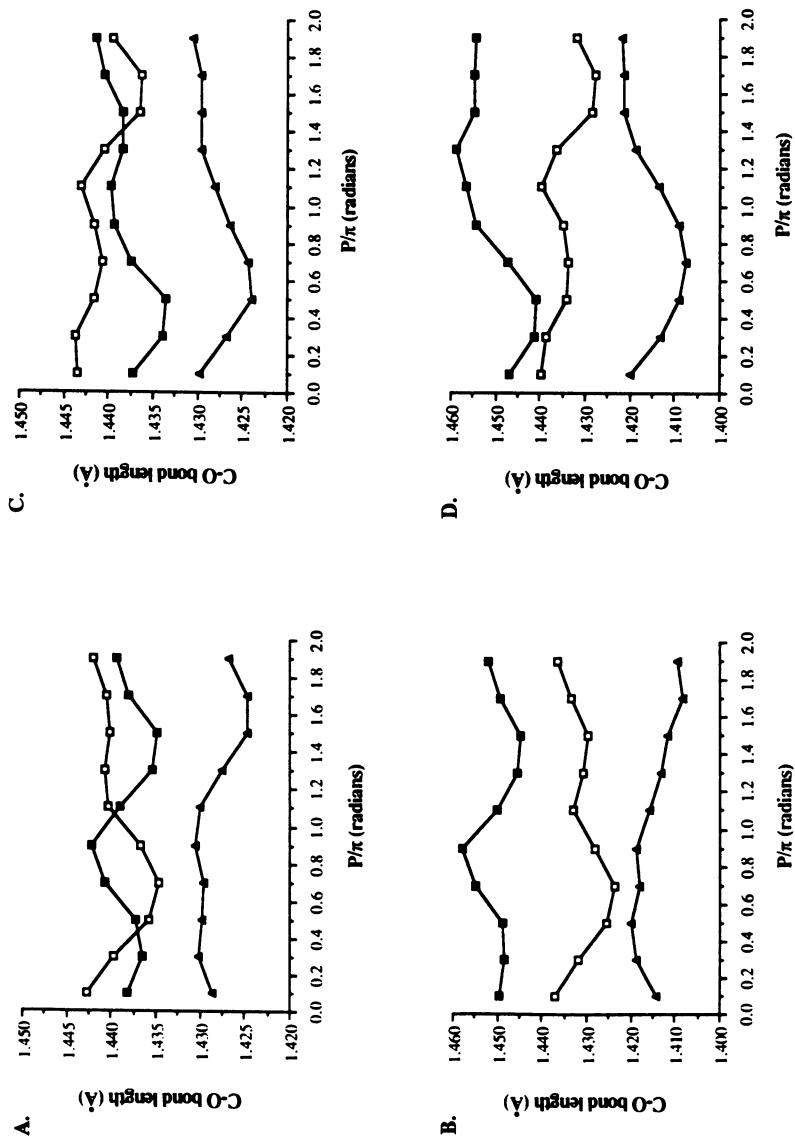
Figure 5. The effect of ring conformation on C-O bond lengths in **7** (A and B) and **8** (C and D) using the STO-3G (A and C) and 3-21G (B and D) basis sets. C4-O4 (■), C1-O4 (□), C1-O1 (▲).
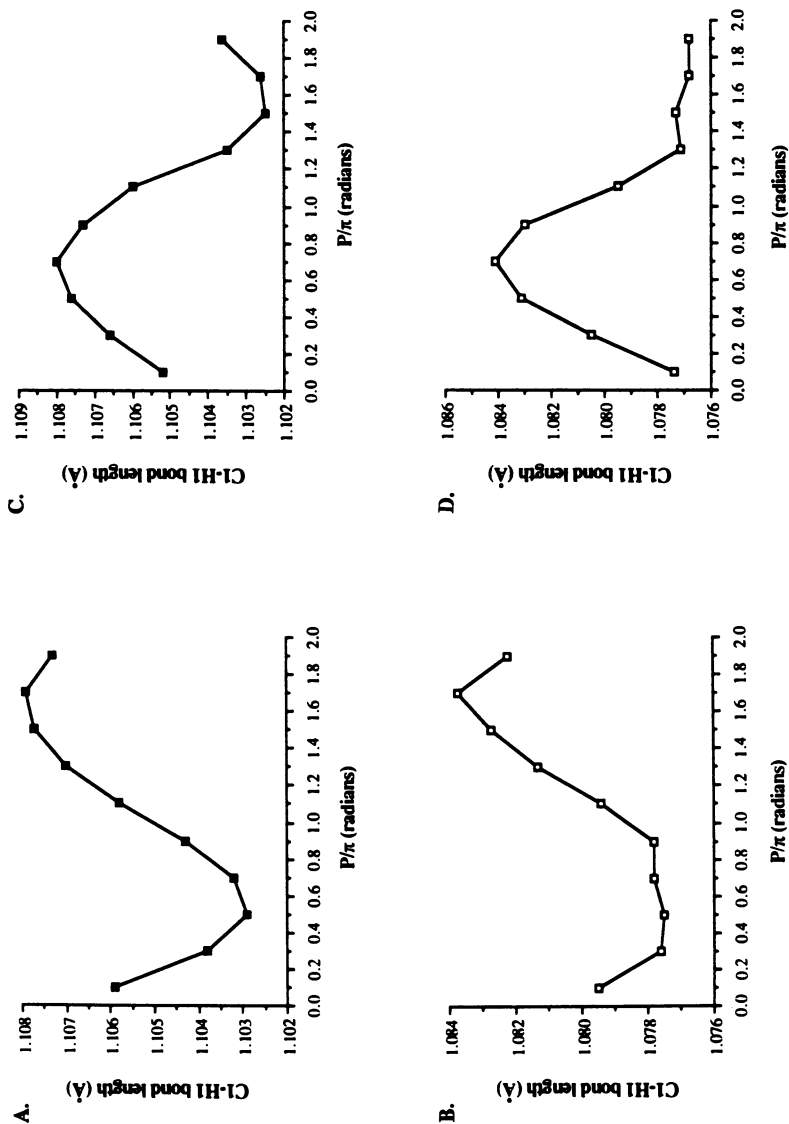
Figure 6. The effect of ring conformation on the C1-H1 bond length in **7** (A and B) and **8** (C and D) using the STO-3G (A and C) and 3-21G (B and D) basis sets.
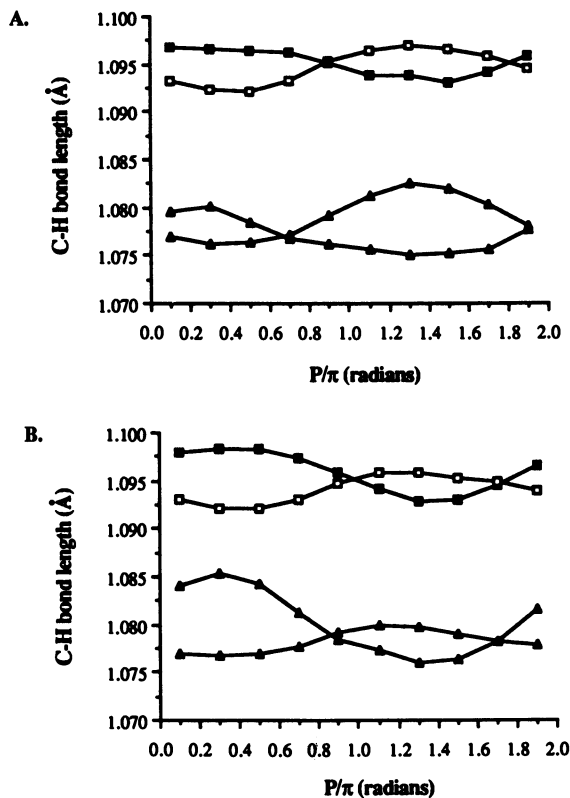
Figure 7. The effect of ring conformation of the C4-
H4R (open symbols) and C4-H4S (solid symbols) in **7** (A)
and **8** (B). STO-3G data are shown in squares; 3-21G
data are shown in triangles.
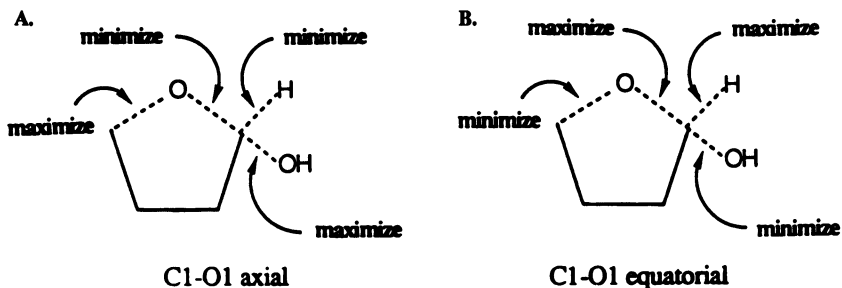
Scheme 4

These effects suggest n-σ* donation to the O4-C1 bond
from the ring oxygen when C1-O1 is quasi-axial, as
expected from the "anomeric effect"[30]. This donation
would be expected to decrease the O4-C1 bond length and
increase the C4-O4 and C1-O1 bond lengths; the
explanation of the effect of C1-O1 bond orientation on
the C1-H1 bond length is less obvious. It is also
possible that these observed trends may be affected by
the C1-O1 torsion angle; in this study only one C1-O1
torsion was studied (Figure 3) which was chosen to
optimize the "exoanomeric effect"[28, 29].

The implications of the above observations may be
important, especially if similar trends are observed in
pyranose anomers. For example, with respect to the
mechanism of acid-catalyzed hydrolysis of pyranosides,
endocyclic C-O bond cleavage (preceeded by O5 protona-
tion) may be assisted in β-anomers in which the C1-O1
bond is equatorial, since the O4-C1 bond may already be
extended in these anomers. By a similar argument,
exocyclic C-O scission (preceeded by O1 protonation) may
be assisted in the hydrolysis of α-pyranosides in which
the C1-O1 is axial and extended, thus resembling the
transition state. Post and Karplus[31] have recently
suggested that enzyme-catalyzed glycoside hydrolysis of
β-pyranosides may indeed take place by ring oxygen
protonation, followed by endocyclic C-O bond scission.

C. Bond Angles. Force-field calculations[32] have shown
that endocyclic bond angles in furanoses vary
systematically with ring conformation. Previous *ab
initio* calculations on the tetrofuranoses **3-6**[20] revealed
a similar dependency on conformation that is essentially
unaffected by anomeric configuration. The C1-C2-C3 bond
angle showed maxima at 0.4 and 1.4 P/π and minima at 0.9
and 1.9 P/π. The remaining four curves were phase-
shifted by 0.2 P/π in order around the ring. CCO and COC
angles were found to be comparable in magnitude, and
larger than CCC bond angles. Similar results are
obtained from STO-3G calculations on **7-10** (Figure 8A).
The observed similarity in CCO and COC angles, however,
is not consistent with angle bending forces[33] that predict
COC to be greater than CCO. Calculations with the 3-21G
basis set produce the same overall pattern of endocyclic
bond angle response to conformation (Figure 8B, 8C), but
appear to more accurately predict the expected trend in
their relative magnitudes, that is, COC > CCO > CCC.
Interestingly, 3-21G data suggests that the COC bond
angle is minimal at conformations in which the ring
oxygen is out-of-plane (^{0}E , E_{0}), but these minima are not
equivalent. The global minimum occurs at ^{0}E in α-
anomers, and at E_{0} in β-anomers, that is, in conformations
where C1-O1 assumes a quasi-axial orientation (Figure 9).
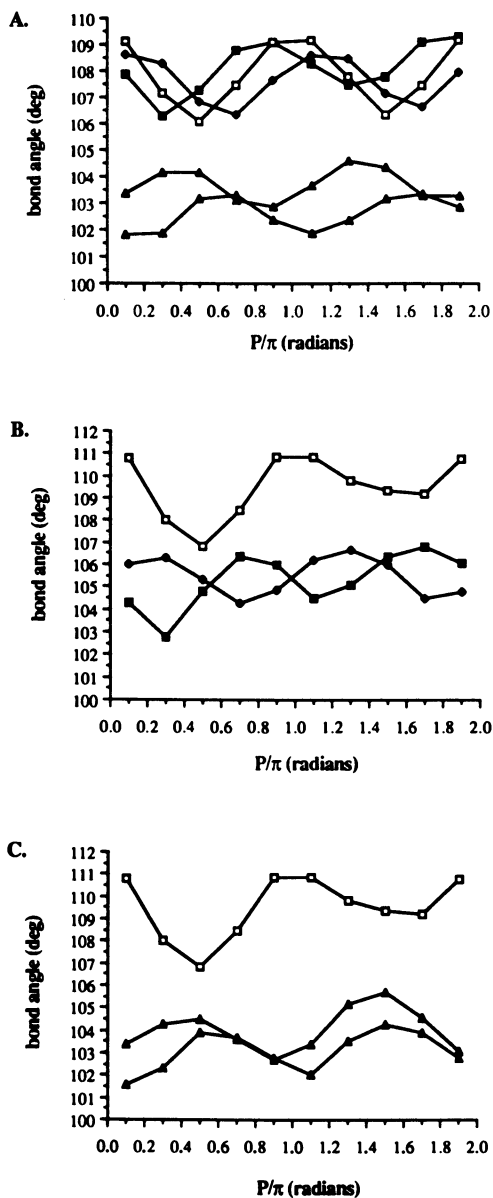
Figure 8. The effect of ring conformation on
endocyclic bond angles in 7:   C1-C2-C3 (▲), C2-C3-C4
(Δ), C3-C4-O4 (■), C4-O4-C1 (□), O4-C1-C2 (◊).   (A)
STO-3G data. (B and C) 3-21G data showing differences
between CCC, CCO and COC bond angles.
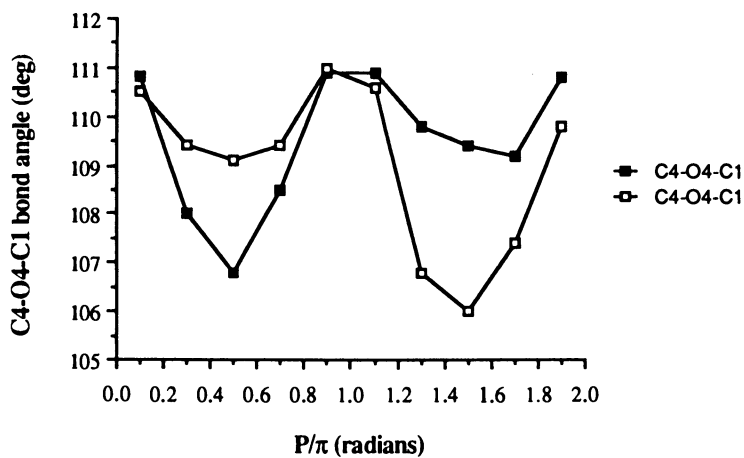
Figure 9. The effect of anomeric configuration on the C4-O4-C1 bond angle in **7** (filled symbols) and **8** (open symbols) using the 3-21G basis set.

D. Bond Torsions. The effect of furanose ring structure and configuration on puckering amplitude is not currently understood quantitatively. Recent *ab initio* studies of the tetrofuranoses **3-6** indicate that puckering amplitude depends on conformation, and ranges from 16°-24°. The puckering amplitudes of **7-10**, determined from STO-3G and 3-21G calculations, are shown in Figure 10. The 3-21G calculations predict larger puckering amplitudes than STO-3G calculations, which is consistent with similar comparisons made on non-carbohydrate furanoid ring systems[34]. Furthermore, the effect of conformation on puckering amplitude is not completely conserved between the two basis sets. Curves for **9** and **10** appear somewhat flatter than those for **7** and **8**, with **8** showing the largest variation of puckering with conformation. The pseudorotational itinerary (Figure 1), therefore, appears to be more circular for **9** and **10** than for **7** and **8**. Puckering minima near 0.0 and 1.0 P/π were commonly observed in STO-3G calculations of the tetrofuranoses **3-6**, and inspection of the STO-3G data in Figure 10 reveals a similar result. In contrast, local minima are more commonly observed at 0.5 and/or 1.5 P/π in 3-21G calculations (Figure 10), that is, in conformations having the ring oxygen out-of-plane. Presumably the puckering is reduced in ^{o}E and E conformers in order to maintain a maximal COC bond angle.

The issue of exocyclic C-O conformation in theoretical calculations of sugars is complicated by a lack of knowledge of the intrinsic and extrinsic (e.g., solvent-mediated) factors controlling C-O torsions in these molecules, especially for C-O bonds involving non-anomeric carbons. The initial C1-O1 bond torsions used for calculations on **7-10** (Figure 3) were selected to optimize the "exoanomeric effect"[28, 29], as previous *ab initio* calculations at the STO-3G level on the tetro-furanoses **3-6**[20], and experimental evidence[35, 36], indicates that the C1-O1 rotamer having the anomeric hydroxyl proton *gauche* to O4 and H1 is most stable. The exoanomeric dihedral angle (H1-C1-O1-H), however, does appear to depend somewhat on ring conformation (Figure 11).

Of the remaining C-O torsions in **7-10**, C3-O3 and C2-O2 rotamers were chosen in **7** and **9**, respectively, to minimize or prevent potential intramolecular bonding. In **9**, geometric optimization of all ten envelope forms did not induce a change in the initial C2-O2 rotamer, although the torsion angle was refined by the optimization (Figure 12). In contrast, for **7**, geometric optimization induced a significant change in the C3-O3 rotamer (from the initial rotamer in Figure 3 to that having the hydroxy proton anti to H3) (Figure 13) during optimization of the ^{2}E conformer. Presumably this rotational change was driven by hydrogen bonding between O1 and O3 in ^{2}E where both C-O bonds are quasi-axial and thus properly oriented for H-bonding. To prevent this
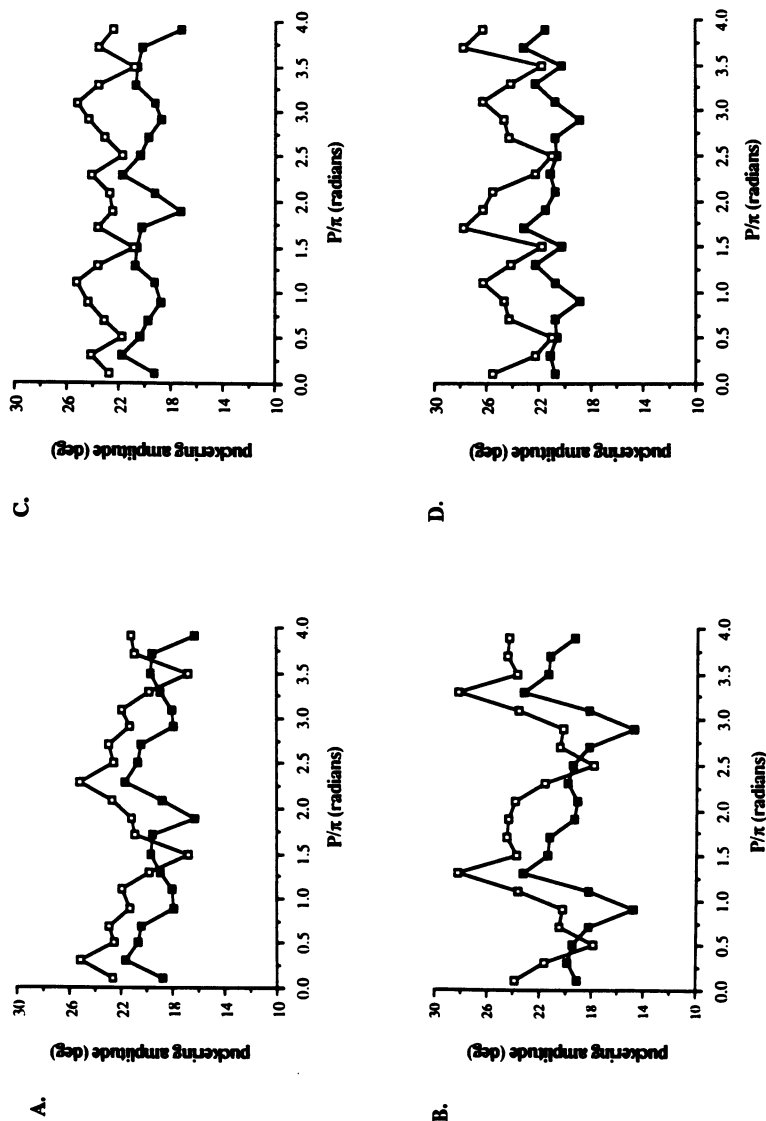
Figure 10. The effect of ring conformation on furanose ring puckering amplitude in **7** (A), **8** (B), **9** (C) and **10** (D) using the STO-3G (open symbols) and 3-21G (solid symbols) basis sets.

A.

B.
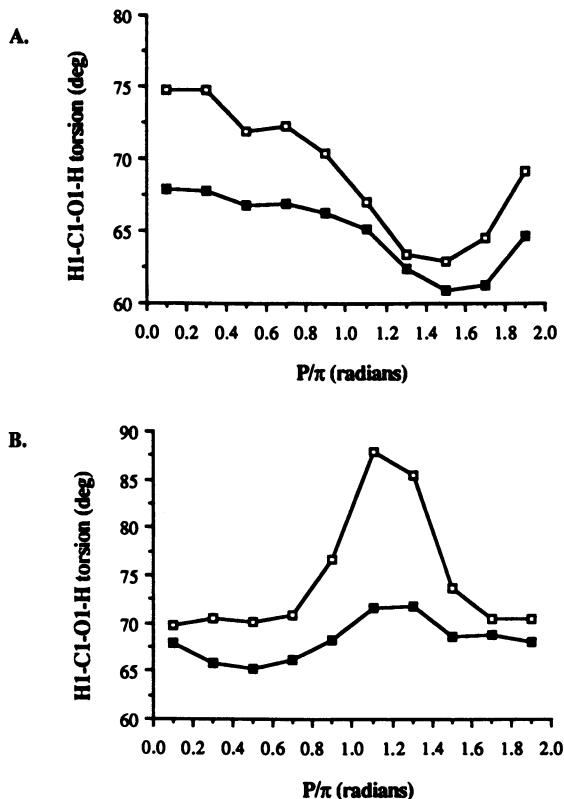
Figure 11. The effect of ring conformation on the H1-C1-O1-H exocyclic torsion angle in **8** (A) and **9** (B). STO-3G data are shown in filled symbols, and 3-21G data are shown in open symbols.
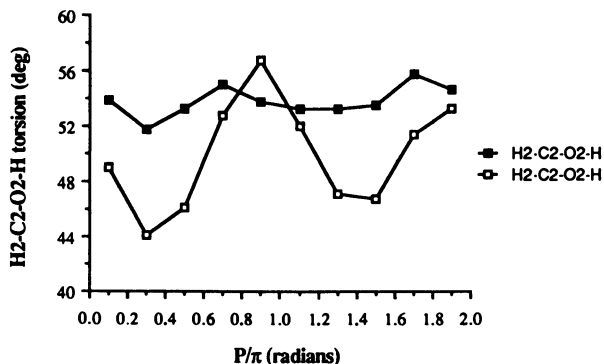
Figure 12. The effect of ring conformation on the H2-C2-O2-H torsion angle in **9** using the STO-3G (filled symbols) and 3-21G (open symbols) basis sets.

interaction, whose presence would presumably add greater stability to the ^2E conformer and thus invalidate our attempt to study intrinsic energies, the C3-O3 torsion was held constant (-60 °) (Figure 3) for 3-21G optimizations of the ^2E conformer. Due to similar rotations during optimization, the C1-O1 (-70°) and C3-O3 (-60°) were held constant during E_3 optimization, whereas the C1-O1 torsion (-70°) was held constant during ^4E optimization.

In **10**, the C1-O1 torsion was held constant at 70° (Figure 3) during 3-21G optimizations of the E_4, E_1, ^2E and E_3 conformers.

Clearly the problem of C-O rotamers is complex and adds uncertainty to the results of these calculations. Differential intramolecular hydrogen bonding in puckered conformers, if present, would notably affect the calculated dependence of ring conformation on total energy. In this study we sought to reduce this contribution to total energy and thereby study the intrinsic behavior of the molecule. In aqueous solution, potential intramolecular hydrogen bonding will compete with intermolecular hydrogen bonding, with the latter possibly dominating due to the large excess of solvent. This being the case, intramolecular hydrogen bonding may not, in general, be a major determinant of preferred furanose conformation in aqueous solution. However, the presence of a water solvent cage around the sugar with its own hydrogen bonding network may induce other presently unknown forces that preferentially act to stabilize or destabilize specific furanose conformers. Even in structures in which there is no opportunity for intramolecular hydrogen bonding (e.g., **8**), the effect of C-O orientation on calculated ring structure and energetics remains to be established. Recent *ab initio* calculations on 2-deoxy-β-D-erythrofuranosylamine[37] using the 3-21G basis set showed that the conformation of the C3-O3 bond did not sigificantly affect the energy profile of the molecule; in contrast, however, the energy profile of the corresponding 2-fluoro derivative was notably affected by C3-O3 bond conformation because of intramolecular OH--F interactions.

E. Conformational Energy Calculations. *Ab initio* calculations on the tetrofuranoses **3-6** using the STO-3G basis set[20] showed that ring configuration significantly affects preferred conformation. Total energy-conformation curves for α-D-erythrofuranose **3**, α-D-threofuranose **5** and β-D-threofuranose **6** revealed a single (global) minimum at 0.4 P/π, 0.4 P/π and 1.6 P/π, respectively. Single-point refinement of the STO-3G energies at the 3-21G level gave slightly phase-shifted curves and larger energy differences, but the presence of a single (global) energy minimum was conserved. In contrast, β-D-erythrofuranose **4** showed radically

different behavior, with the STO-3G data producing a relatively flat energy-conformation curve with a global minimum at 1.2 P/π. Single-point 3-21G refinement of these data, however, produced a curve with two well-defined minima of comparable energies at 0.0 and 1.0 P/π, a result which appears to be more consistent with experimental data[8]. This disparity between basis sets indicates that STO-3G calculations may not be reliable in predicting accurate conformational energy profiles in some furanoses.

In this study, we examined the effect of basis set more thoroughly by conducting complete geometric optimizations with the STO-3G and 3-21G basis sets to obtain and compare conformational energy profiles on **7-10** (Figure 14). In all four cases, the energy difference between the least and most stable conformers is greater in 3-21G data than in STO-3G data. In **7, 8** and **10,** the general shape of the profile is maintained, although additional "fine structure" appears to arise at the 3-21G level. In contrast, there is a notable difference between basis sets for **9.** Of the four structures studied, **9** is the only isomer having cis-1,2 hydroxyl groups, and care was taken to choose a C2-O2 torsion to prevent potential intramolecular hydrogen bonding between the adjacent hydroxy groups at O1 and O2. The cause of the disparity is presently unclear, but more confidence is placed on the result obtained from the 3-21G analysis.

Based on the more reliable 3-21G data, the conformational behavior of **7-10** is summarized in Scheme 5. In **7,** north-south interconversion (Figure 1) between the two most stable conformers, E_4 and ^2E, occurs predominantly via east conformers (E_1 and ^OE) through a comparatively low activation barrier. In contrast, the two most stable north and south conformers of **8,** ^4E and E_2, interconvert predominantly via west conformers, E_O and ^1E, and the activation barrier is higher. It is interesting to note that the interconversion of north and south conformers of the structurally-related 2-deoxy-β-D-*erythro*-pentose **2,** appears to occur predominantly through east conformers[18]; the presence of a destabilizing 1,3-interaction between O1 and C5 (not present in east conformers) destabilizes west conformers of **2.** This destabilizing interaction is absent in west conformers of **8,** thereby permitting a west interconversion pathway.

In **9,** a south conformer (^2E) is highly preferred, although a local minimum is observed at conformers near ^3E (north conformer). The pathway of interconversion is characterized by a high energy barrier through east conformers (E_1, ^OE, E_4). Two minima are observed for **10,** a global minimum at west conformers, ^4E and E_O, and a local minimum at ^OE. Interconversion between these conformers is characterized by high energy barriers which are comparable for the north and south pathways.
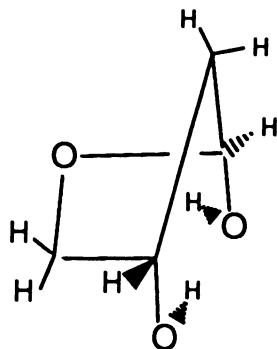
Figure 13.   The ^2E conformer of **7** showing the potential for intramolecular hydrogen bonding between the hydroxyl substituents at C1 and C3.   The C3–O3 bond torsion change induced during 3–21G geometry optimization of this structure was presumably driven by this interaction.

2-deoxy-α-D-glycero-tetrofuranose
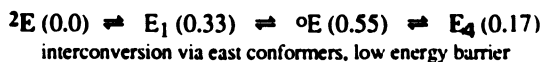
^2E (0.0) ⇌ E_1 (0.33) ⇌ ^oE (0.55) ⇌ E_4 (0.17)

interconversion via east conformers, low energy barrier

2-deoxy-β-D-glycero-tetrofuranose

^4E (0.0) ⇌ E_o (1.65) ⇌ ^1E (1.70) ⇌ E_2 (0.90)

interconversion via west conformers, high energy barrier

3-deoxy-α-D-glycero-tetrofuranose

^2E (0.0) ⇌ E_1 (2.00) ⇌ ^oE (3.03) ⇌ E_4 (2.97) ⇌ ^3E (2.93)

interconversion via east conformers, high energy barrier

3-deoxy-β-D-glycero-tetrofuranose

^4E (0.0) ⇌ E_o (0.05) ⇌ ^oE (3.00)

interconversion via north or south conformers, high energy barrier
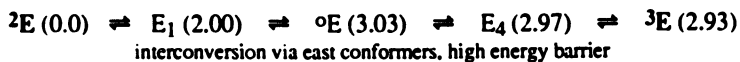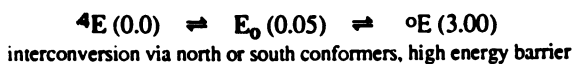
Scheme 5
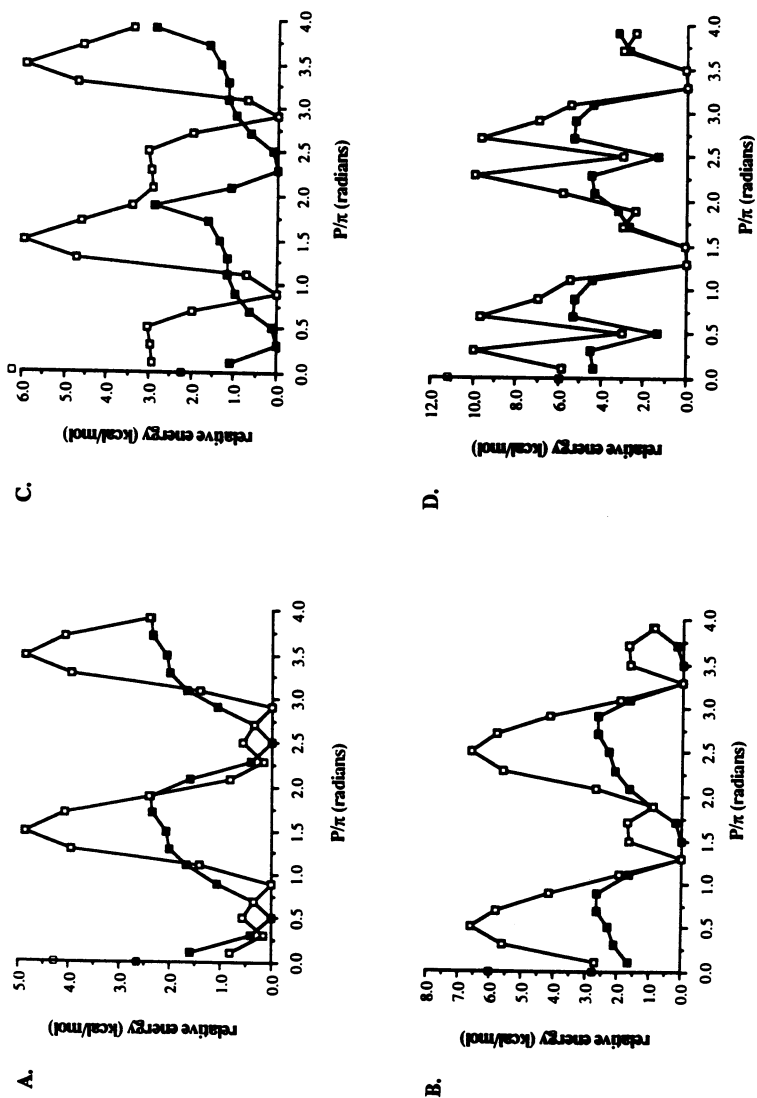
Figure 14. The effect of ring conformation on relative energy of **7** (A), **8** (B), **9** (C) and **10** (D). STO-3G data are shown in filled symbols, and 3-21G data are shown in open symbols. Energies of the planar forms are indicated on the y-axes.

The above energy profiles suggest significantly different conformational behavior in **7-10** both in terms of preferred geometries and modes of conformational interconversion in the gas phase. Inherent in the above analysis is the assumption that the 3-21G calculations provide a reliable picture of gas phase behavior. Of course, this assumption will require validation through calculations with more sophisticated basis sets, especially those employing d-orbitals (e.g., 6-31G*). In addition, while it is difficult to extrapolate the data in Figure 14 to conditions in aqueous solution, it is nevertheless clear that the nature of conformational averaging, which affects the interpretation of NMR parameters, depends on ring configuration, and that a singular approach to NMR parameter interpretation cannot be applied to all furanose structures.

Energy profiles in Figure 14 also reveal that planar furanose forms are often of lower energy than puckered conformers. For example, relative conformational energies determined for **7** with the 3-21G basis set indicate that the planar conformer is more stable than the E_0 conformer; in **8,** the planar conformer is calculated to be more stable than $^\circ E$. These observations suggest that the conformational dynamics of some furanose rings may not be completely described by pseudorotation; in these cases, conformer interconversion may occur by both inversion and pseudorotational pathways, with the latter being the more preferred route.

F. Effect of Basis Set on Optimized Molecular Parameters in Furanoses. The above results indicate that the 3-21G basis set is, in general, more reliable than the STO-3G basis set in molecular orbital calculations of furanoses. However, the 3-21G basis set lacks polarization functions which may be important in studies of furanoses that contain both endocyclic and exocyclic oxygen atoms. We conducted a limited study to assess the effect of extended basis sets on calculated geometries by optimizing the planar conformer of 2-deoxy-α-D-*glycero*-tetrose **7** using the STO-3G, 3-21G, 4-31G, 4-31G*, 6-31G and 6-31G* basis sets. These data are shown in Table 1. Results obtained with the 4-31G and 6-31G basis sets were very similar and are discussed below as a single group (4-31G/6-31G). Results obtained with the 4-31G* and 6-31G* basis sets are discussed below as a single group (4-31G*/6-31G*) for the same reason. Implicit in the following discussion is the assumption that the observed trends between basis sets will be independent of furanose ring configuration and conformation.

Bond lengths generally decrease with basis set in the order STO-3G, 3-21G, 4-31G/6-31G and 4-31G*/6-31G*. For example, the C2-C3 bond length is 1.557 Å (STO-3G), 1.546 Å (3-21G), 1.539 Å (4-31G/6-31G) and 1.536 Å (4.31G*/6-31G*), for an overall decrease in length of

1.3% (0.021 Å). The C-O bond lengths determined from STO-3G and 4-31G*/6-31G* decrease by 2.2 - 3.2%. The C-H bonds decrease in the order STO-3G → 3-21G → 4-31G/6-31G, but are slightly lengthened relative to 4-31G/6-31G in the 4-31G*/6-31G* data.

Bond angles are also notably affected by basis set (Table 1). It is interesting to note that the 3-21G basis set predicts a similar COC bond angle (113.0°) as the 4-31G* and 6-31G* basis sets (113.8°); these values are substantially larger than that predicted by the STO-3G basis set (110.4°). The largest differences between basis sets occurs for the COH bond angles (e.g., for C3-O3-H, 104.2° with STO-3G and 113.7° with 6-31G); these deviations could be significant, as these angles affect the position of hydroxyl protons and thus their ability to participate in intra- and intermolecular hydrogen bonds.

Bond torsions are not greatly affected by basis set. The largest difference occurs for the C1-O1 bond torsion which varies from 66.5° (6-31G*) to 75.4° (3-21G), with the STO-3G angle (68.3°) in closer agreement with the 6-31G* result. In contrast the C3-O3 bond torsion is only slightly affected by basis set. This observation points to the significantly different factors governing the exoanomeric C-O torsions in sugars which may not be properly treated by some basis sets. Since two oxygen atoms are involved in regulating the exoanomeric C-O torsion, use of basis sets with polarization functions may be required to more accurately evaluate its behavior.

Discussion

Previous *ab initio* molecular orbital studies of the aldotetrofuranoses[20] using the minimal STO-3G basis set showed that bond lengths, bond angles and bond torsions are affected by furanose ring configuration and conformation. Of particular interest were the changes in bond lengths, especially those bonds in the vicinity of the anomeric center. Such changes in bond length might be important in determining the structure and reactivity of furanose anomers. Furthermore, subtle changes in overall molecular dimensions as a function of ring conformation may be important in mediating molecular recognition and catalysis between enzymes and furanose substrates. However, bond length changes observed with the simple STO-3G basis set remained to be validated by calculations with more extended basis sets.

This study has compared STO-3G and 3-21G optimized geometries of furanose conformers and has shown that, in general, the overall patterns of bond length changes predicted by the STO-3G and 3-21G basis sets are similar, although the absolute changes differ with basis set. In particular, the systematic changes of C-H and C-O bond lengths near the anomeric center suggest a role in

determining chemical reactivity (i.e., glycoside hydrolysis, anomerization) and preferred conformation.

Previous studies conducted with the STO-3G basis set on aldotetrofuranoses[20] suggested that the STO-3G basis set may not be reliable in predicting total energy profiles for furanose conformers. For example, the energy profile for β-D-erythrofuranose 4 determined from STO-3G optimization was significantly different than that determined from single-point 3-21G calculations using the same STO-3G optimized molecular parameters; the latter results appeared more consistent with experimental data. The present study provides further evidence that energy profiles are not reliable when obtained with the minimal basis set. The greater reliability of 3-21G energy calculations derives from its ability - as shown in this study - to predict bond lengths and bond angles in closer agreement to those observed experimentally. Thus, we conclude that carbohydrate calculations using *ab initio* methods should be conducted with basis sets no less sophisticated than the 3-21G basis set if reasonable structures and energetics are to be obtained.

The calculated total energy profiles for the aldotetrofuranoses 3, 5 and 6 contain a well-defined global minimum[20], suggesting that these compounds, at least in the gas phase, prefer conformations found in a limited region of the pseudorotational itinerary. This behavior is notably different than that of the aldotetrofuranose 4 and the deoxytetrofuranoses 7, 8 and 10. Calculated total energy profiles of the latter compounds contain global and local minima of relatively similar energies. Thus, 4, 7, 8 and 10 appear to be more conformationally mobile in the gas phase. In some cases, these preferred conformations are similar (i.e., they are contiguous along the pseudorotational itinerary), whereas in others these conformations may be notably different (i.e., north and south geometries). Furthermore, the dynamics of conformer interconversion differs between 4, 7, 8 and 10, since energy barriers between preferred conformers differ for each structure. These calculations show that structure and configuration have a profound effect on the conformational dynamics of furanose rings, at least in the gas phase. Solution studies of 4[7] appear to support the conformational behavior predicted by *ab initio* methods; the solution behavior of 7-10 is currently under investigation.

The absolute changes in total energy with furanose ring conformation appear to be greater with the 3-21G basis set than with STO-3G calculations. It is likely that these basis sets represent the two extreme limits, and that calculations currently underway with the 6-31G* basis set will yield total energy profiles with absolute energy changes intermediate in magnitude.

Hydroxyl group orientation in computational studies of carbohydrates still remains a problem. This study has

Table 1. Comparison of Optimized Molecular Parameters
For the Planar Conformer of 2-Deoxy-α-D-*glycero*-tetrose **7**
Obtained With Different Basis Sets

Parameter[1]	STO-3G	3-21G	4-31G	4-31G*	6-31G	6-31G*
R1-2	1.557	1.546	1.539	1.536	1.540	1.537
R2-3	1.559	1.544	1.531	1.533	1.532	1.534
R3-4	1.437	1.445	1.439	1.404	1.441	1.405
R4-5	1.437	1.424	1.418	1.391	1.420	1.392
R5-6	1.106	1.079	1.079	1.085	1.080	1.085
R5-7	1.431	1.420	1.411	1.386	1.412	1.387
R7-8	0.991	0.967	0.953	0.949	0.952	0.948
R1-9	1.086	1.081	1.081	1.083	1.082	1.084
R1-10	1.087	1.079	1.078	1.081	1.080	1.082
R2-11	1.097	1.082	1.082	1.085	1.083	1.086
R2-12	1.434	1.442	1.433	1.402	1.434	1.404
R12-13	0.991	0.966	0.952	0.948	0.951	0.947
R3-14	1.095	1.079	1.078	1.083	1.079	1.084
R3-15	1.096	1.077	1.076	1.081	1.077	1.082
A1-2-3	104.4	105.3	105.7	104.5	105.8	104.5
A2-3-4	110.1	107.5	106.9	108.0	106.9	108.0
A3-4-5	110.4	113.0	113.6	113.8	113.6	113.8
A4-5-6	107.7	107.8	107.5	107.6	107.5	107.5
A4-5-7	109.6	110.2	110.2	110.6	110.3	110.7
A5-7-8	103.7	109.8	112.2	108.5	112.5	108.6
A2-1-9	111.6	112.0	111.6	111.8	111.6	111.9
A2-1-10	110.6	109.3	110.1	110.3	110.3	110.4
A3-2-11	110.0	110.7	111.0	110.4	111.0	110.5
A3-2-12	108.2	106.3	106.8	108.1	106.9	108.0
A2-12-13	104.2	110.8	113.5	109.6	113.7	109.7
A2-3-14	111.0	112.3	112.9	112.1	112.9	112.1
A2-3-15	109.1	108.2	109.5	109.5	109.7	109.5
T3-4-5-6	121.2	122.2	122.7	121.7	122.7	121.7
T3-4-5-7	119.0	117.2	117.8	118.8	117.9	118.8
T6-5-7-8	68.3	75.4	72.7	66.7	71.5	66.5
T3-2-1-9	120.8	122.1	121.6	121.1	121.5	121.2
T3-2-1-10	117.9	115.4	117.2	117.9	117.5	117.9
T4-3-2-11	118.0	120.1	120.7	119.1	120.7	119.3
T4-3-2-12	121.4	120.3	120.7	121.4	120.9	121.4
T11-2-12-13	57.8	58.9	57.8	57.1	57.9	57.3
T1-2-3-14	120.7	119.1	118.5	119.6	118.5	119.6
T1-2-3-15	120.4	118.8	118.4	119.5	118.4	119.6

[1]R, A and T refer to bond lengths (in Å), bond angles
(in °), and bond torsions (in °) used to set the Z-matrix
input file.
Atom numberings in **7** are shown in Scheme 6.
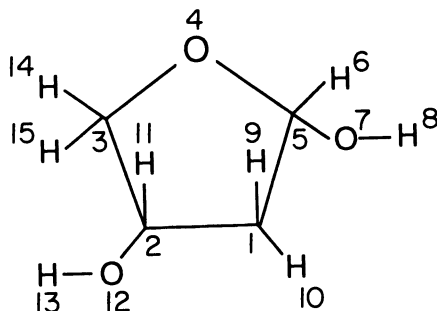
Scheme 6

confirmed that the C1-O1 bond in aldofuranoses prefers to be *gauche* to the ring oxygen and *anti* to C2, as predicted by the exoanomeric effect[28,29]. Thus, for the C1-O1 bond, a rational argument exists to limit its conformation. However, there are no rules at present to deduce preferred C-O torsions for non-anomeric ring hydroxyl groups. It is not inconceivable that hydroxyl group orientation may affect the overall energetics of conformer interconversion, especially in condensed phases. Intramolecular and intermolecular hydrogen bonding are likely to stabilize/destabilize specific conformers and thereby affect the overall energy profile. We have assessed this possibility by optimizing the planar and envelope conformers of **8** with O3-H anti to H3 (Case I) and with O3-H anti to C4 (Case II). These results are shown in Figure 15. The overall shape of the two profiles is conserved, with each profile showing one global minimum and one local minimum. However the global minimum in Case I is the local minimum : Case II.
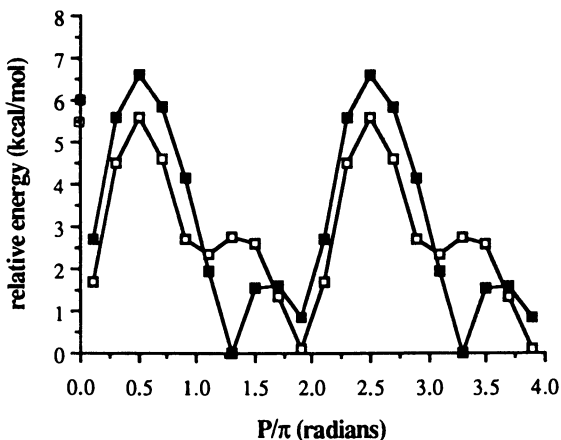
Figure 15. The effect of C3-O3 bond conformation on the energy profile of **8**. Geometries were optimized with the 3-21G basis set. Case I, O3-H bond anti to H3, closed symbols; Case II, O3-H bond anti to C4, open symbols.

Thus, following the lowest energy pathway between minima might require C3-O3 bond rotation in **8**. This cursory examination indicates that, while the location of energy minima may not be affected by non-anomeric hydroxyl conformations, the relative energies of these minima may indeed depend on these conformations.

References

1. Levitt, M.; and Warshel, A. J. Am. Chem. Soc. 1978, 100, 2607.
2. Harvey,S.C.; Prabhakaran, M. J. Am. Chem. Soc. 1986, 108, 6128.
3. Gerlt, J. A.; Youngblood, A.V. J. Am. Chem. Soc. 1980, 102, 7433.
4. Raap, J.;van Boom, J.H.; van Lieshout, H.C.; Haasnoot, C.A.G. J. Am. Chem. Soc. 1988, 110, 2736.
5. Rinkel, L. J.; Altona, C. J. Biomol. Struct. Dyn. 1987, 4, 1987.
6. Cyr, N.; Perlin, A.S. Can. J. Chem. 1979, 57, 2504.
7. Serianni, A.S.; Barker, R. J. Org. Chem. 1984, 49, 3292.
8. Angyal, S.J. Carbohydr. Res. 1979, 77, 37.
9. Saenger, W. Principles of Nucleic Acid Structure, Springer-Verlag, New York, p. 252 and p. 349.
10. Saenger, W. Principles of Nucleic Acid Structure, Springer-Verlag, New York, p. 256 and p. 261.
11. Kilpatrick, J.E.; Pitzer, K.S.; Spitzer, R. J. Am. Chem. Soc. 1947, 69, 2483.
12. Altona, C.; Geise, H.J. Tetrahedron 1968, 24, 13.
13. Altona, C.; Sundaralingam, M. J. Am. Chem. Soc. 1972, 94, 8205.
14. Altona, C.; Sundaralingam, M. J. Am. Chem. Soc. 1973, 95, 2333.
15. Westhof, E.; Sundaralingam, M. J. Am. Chem. Soc. 1983, 105, 970.
16. Jardetzky, O. Biochim. Biophys. Acta 1980, 621, 227.
17. Schlick, T.; Peskin, C.; Broyde, S.; Overton, M. J. Comp. Chem. 1987, 8, 1199.
18. Olson, W.K.; Sussman, J.L. J. Am. Chem. Soc. 1982, 104, 270.
19. Olson, W.K. J. Am. Chem. Soc. 1982, 104, 278.
20. Serianni, A.; Chipman, D.M. J. Am. Chem. Soc. 1987, 109, 5297.
21. Binkley, J.S.; Whiteside, R.A.; Krishnan, A.; Seeger, R.; Defrees, D.J.; Schlegel, H.B.; Topiol, S.; Khan, L.R.; Pople, J.A. QCPE 1981, 13, 406.

22. (a) Gaussian 86, M.J. Frisch, J.S. Binkley, H.B.
     Schlegel, K. Raghavachari, C.F. Melius, R.L. Martin,
     J.J.P. Stewart, F.W. Bobrowicz, C.M. Rohlfing, L.R.
     Kahn, D.J. Defrees, R. Seeger, R.A. Whiteside, D.J.
     Fox, E.M. Fleuder and J.A. Pople, Carnegie-Mellon
     Quantum Chemistry Publishing Unit, Pittsburgh, PA,
     1984.   (b)   Gaussian 88, M.J. Frisch, M. Head-Gordon,
     H.B. Schlegel, K. Raghavachari, J.S. Binkley, C.
     Gonzalez, D.J. Defrees, D.J. Fox, R.A. Whiteside, R.
     Seeger, C.F. Melius, J. Baker, R. Martin, L.R. Kahn,
     J.J.P. Stewart, E.M. Fluder, S. Topiol and J.A.
     Pople, Gaussian, Inc., Pittsburgh, PA, 1988.
23. Hehre, W.J.; Stewart, R.F.; Pople, J.A. J. Chem.
     Phys. 1969, 51, 2657.
24. Newton, M.D.; Latham, W.A.; Hehre, W.J.; Pople, J.A.
     J. Chem Phys. 1970, 52, 4064.
25. Binkley, J.S.; Pople, J.A.; Hehre, W.J. J. Am. Chem.
     Soc. 1980, 102, 939.
26. Barragan, I.; Lopez-Castro, A.; Marquez, R. Acta
     Crystallogr., Sect. B 1977, 33, 2244.
27. Barragan, I.; Lopez-Castro, A.; Marquez, R. Acta
     Crystallogr., Sect. B 1978, 34, 295.
28. Lemieux, R.U. Pure Appl. Chem. 1971, 25, 527.
29. Lemieux, R.U.; Koto, S.; Voisin, D. Anomeric Effect:
     Origin and Consequences, Szarek, W.A., Horton, D.
     Eds.; ACS Symposium Series 87, American Chemical
     Society, Washington, DC, 1979; p 17.
30. Lemieux, R. Molecular Rearrangements, de Mayo, P.
     Ed.; Wiley-Interscience, New York, 1963; 713.
31. Post, C.B.; Karplus, M. J. Am. Chem. Soc. 1986, 108,
     1317.
32. DeLeeuw, H.P.M.; Haasnoot, C.A.G.; Altona, C. Isr.
     J. Chem. 1980, 20, 108.
33. Westheimer, F. Steric Effects in Organic Chemistry,
     Newman, M.S. Ed.; Wiley, New York, 1956, Chapter 12,
     pp 523-555.
34. Cremer, D.; Pople, J.A. J. Am. Chem. Soc. 1975, 97,
     1358.
35. Jeffrey, G.A.; Pople, J.A.; Binkley, J.S.
     Vishveshwara, S. J. Am. Chem. Soc. 1978, 100, 373.
36. Dais, P.; Perlin, A. Can. J. Chem. 1982, 60, 1648.
37. Lesyng, B.; Marck, C.; Guschlbauer, W. Intl. J.
     Quant. Chem. 1985, 28, 517.

Chapter 7

Modeling of Glucopyranose

The Flexible Monomer of Amylose

Alfred D. French[1], R. S. Rowland[2], and Norman L. Allinger[3]

[1]Southern Regional Research Center, U.S. Department of Agriculture, P.O. Box 19687, New Orleans, LA 70179
[2]Department of Biochemistry, University of Alabama, Birmingham, AL 35294
[3]Department of Chemistry, University of Georgia, Athens, GA 30602

The variability of the distance between O1 and O4 ($\underline{D}$) in glucopyranose rings was modeled with the computer program CHARMM, three versions of MM2, and MM3. $\underline{D}$ is of interest because models of amylosic oligomers have dramatically different shapes when they are built with glucose residues that have large or small $\underline{D}$. In the Cambridge Structural Database (excepting cyclo-amyloses), $\underline{D}$ ranges between 4.05 and 4.67 Å, with a mean of 4.411 Å. Models with lowest energy from the five programs had $\underline{D}$ values of 4.55 ± 0.02 Å when the dielectric constant was set for isolated molecules (1.5). Higher dielectric constants had no effect on $\underline{D}$ of MM2 models, but $\underline{D}$ in an MM3 model was 4.47 Å when the dielectric constant was set to 4, appropriate for crystals. The residue geometry was optimized at 13 different values of $\underline{D}$, giving different bond and torsion angles. Amounts of change in these angles were similar to those in the database, as were their values at a given $\underline{D}$. The most severe differences were about 3° for O5-C1-O1 and C3-C4-O4. Predicting correct amounts of change shows that a modeling force field is suitably partitioned among the various terms for bond length stretching, torsional rotation, van der Waals interaction, etc., and validates it for other modeling studies. The energy needed to deform the residue over the observed range of $\underline{D}$ is less than 2 kcal/mol.

Goebel and Brant (1) showed that the likely shapes of computer models of amylose, a polymer of 1->4 linked α-D-glucose, depend on the exact geometry of the monomeric unit as well as on the valence-bond and torsion angles at the glycosidic linkage. Subsequently, the distance between O1 and O4 ($\underline{D}$) of the monomer (Figure 1) was found to be an indicator of residue geometry that correlates with the shapes of models of various single- and double-helices of amylose (2,3). The correlation of this indicator with the number of residues in macrocycles of crystalline cycloamyloses (3) was confirmed by Saenger (4) and this variable virtual-bond length is

explicitly incorporated in the fiber diffraction and modeling software of Zugenmaier and Sarko (5).

Variation in $\underline{D}$ affects the shape of amylose models through changes in the spatial relationship between the O1-C1 bond vector and the O4-C4 vector (3). While $\underline{D}$ of α-D-glucose residues has a wide range and works fairly well for describing the flexibilty in amylose helices, $\underline{D}$ is nearly constant in β-D-glucose residues despite similar variability in ring shape. The difference in the extent of variation of $\underline{D}$ for the two anomeric forms arises because the bond vectors are roughly perpendicular in the α ring but parallel in the β ring. Therefore, the study of other polysaccharides may require indicators of residue geometry other than the virtual bond length, which so nicely simplifies the modeling of amylose.

Variation in oligomeric, and hence, polymeric shape arising from glucose residues with different $\underline{D}$ is shown in Figure 2 by two maltotetraose models. Both models are built from residues having $^4C_1$ conformations, and their valence-bond and torsion angles at the glycosidic linkage are identical. However, the distance between the terminal O1 and O4 atoms is 13.7 Å in the upper tetramer and 5.8 Å in the lower model. The upper model is composed of the α-residues with intra-residue $\underline{D}$ of 4.570 Å from the crystal structure report on methyl-β-maltoside (6). The lower model is composed of non-reducing residues from α-maltose (7) with $\underline{D}$ of 4.052 Å. The covalent chemical environments of the two residues are similar so the geometric differences must be due mostly to different crystal packings. The 0.518 Å difference in $\underline{D}$ itself affects the tetramer shape. However, the major difference results from cumulative variations in the location of adjacent residues, which arise from differences of about 30° in the angles between the C1-O1 and C4-O4 bond vectors. Residues with intermediate values of $\underline{D}$ lead to curvatures intermediate to the above tetramers.

By superimposing the C2, C3, C5 and O5 atoms of these two residues, their structural differences can be seen (Figure 3). The residue pairs were fitted by an algorithm (8) furnished as part of the CHEM-X modeling system (CHEM-X is developed and distributed by Chemical Design Ltd, Oxford, England). The C3-C4-C5 plane in the long residue is more perpendicular to the seat of the chair, while its O5-C1-C2 plane is rotated more towards coplanarity with the seat of the chair. Motions of these three-atom planes, plus smaller variations in the bond angles such as O5-C1-O1, are amplified (by the lengths of the C1-O1 and C4-O4 bonds) to give the observed range of $\underline{D}$. However, standard modeling programs are based on bond and torsion angles, not motions of planes, so our modeling study focuses on changes in the ordinary internal coordinates.

About a decade ago, Pensak and French investigated this flexibility with the program MM1 and a limited set of crystallo-graphic results (9). Since then, the number of crystal studies has increased, and new modeling software was developed in attempts to improve accuracy. Therefore, we have reinvestigated this problem. While we focus on changing $\underline{D}$, we believe that the abilities and deficiencies disclosed in our study will apply to other modeling studies such as conformational analyses of disaccharides with flexible residues.

In this study, we assume that crystal structures will have the lowest possible total of intra- and inter-molecular potential energy. However, the partitioning of the potential energy between intra- and inter-molecular terms will vary among crystal structures, distorting the glucose residues away from the shape of lowest energy in a way that will reflect more-or-less random
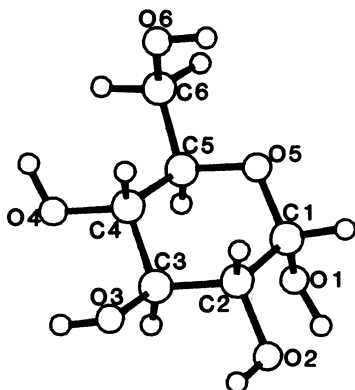
Figure 1.   α-D-Glucose in the starting conformation used herein.
Hydroxyl groups are pointed clockwise, and O6 is gauche to O5 and
gauche to C4 (the gg position).
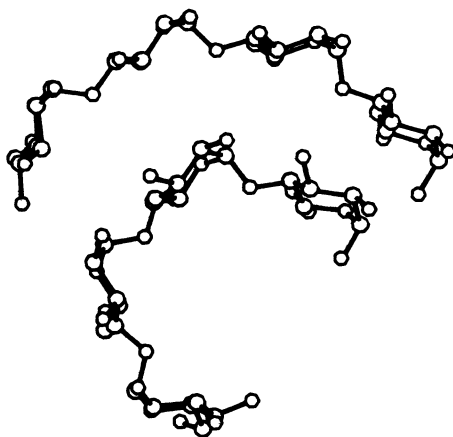
Figure 2.   Two maltotetraose models constructed from residues of
glucose having different D.   They have identical linkage bond
and torsion angles.
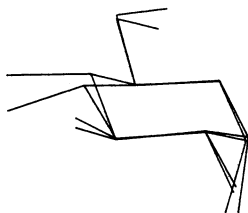
Figure 3.   Comparison of the residue geometries used to make the
tetramers in Figure 2.

differences in the partitioning of total energy. In a large
population of different crystal structures that contain one or more
glucose residues, the D observed most frequently should correspond
closely to a model of lowest energy. D values observed less
frequently would correspond to models of higher energy, and
structural details such as bond- and torsion-angles should agree
when D values of the models and real residues agree. Even though
it is hard to determine the least energetic shape of the glucose
ring by experiment, comparisons of changes in the various molecular
parameters while D changes can indicate how well the force field
works. Another use for a study of this type is establishing the
range of monomeric variation as a prelude to study of other
polymers.

Modeling Software

For our work, three versions of the program MM2 (1977, 85 and 87)
(10,11) were used as well as a pre-release version of MM3, the
successor to MM2 (12). (MM2(77) is available from the Quantum
Chemistry Program Exchange, Department of Chemistry, Indiana
University, Bloomington, Indiana 47901, as are the two recent
versions to academic users. Commercial users can get MMP2(85),
MM2(87) and MM3 from Molecular Design Ltd, 2132 Farallon Drive, San
Leandro, California. MM3 is also available, to for-profit and not-
for-profit users, from Technical Utilization Corp., Inc., 235 Glen
Village Court, Powell, Ohio 43065). MM2 and MM3 are intended to
model a wide variety of molecules. This wide applicability is
attempted through use of complicated potential energy terms. The
CHARMM program (13) (available from Polygen Corporation, 200 Fifth
Avenue, Waltham, Massachusetts 02154) was also tested. This
general-purpose (molecular mechanics, dynamics, etc.) program has a
simpler potential that is often used for proteins, but it can apply
to carbohydrates through the use of parameters developed especially
for carbohydrates (14).
    To illustrate the different complexities, CHARMM's energies
arising from torsional terms depend only on the two central atoms,
and there is only one cosine term. For MM2 and MM3, there are
three cosine terms for each four-atom sequence that defines a
torsion angle. MM2 treats lone pairs of electrons on hydroxyl and
ether oxygen atoms as separate "atoms" that also must be
parameterized. Therefore, many more parameters must be used with
the MM2 and MM3 programs than the CHARMM program. (The parameters
for all the atomic sequences in glucose, and many other molecules,
are furnished with MM2 and MM3.)
    In another example of differences in complexity, the bond-
stretching energy in CHARMM is calculated with a harmonic
oscillator function. MM3 solves the problem described by French,
Tran and Perez in this book for MM2's cubic stretching function by
using a quartic function for bond stretching. Additional
complexity in MM3 is described in Ref. 12.
    Carbohydrates have been included in the wide range of
molecules used in the parameterization of MM2 and of MM3. Alcohol
and ether parameters have usually been determined from simple
alcohols and ethers themselves. However, carbohydrates contain
some unusual features in the acetal linkages, and in the many
vicinal hydrogen-bonded hydroxyl groups. The "anomeric effect",
first discovered by Edward (15) and popularized by Lemieux (16), is
best known in carbohydrates, although, of course, it occurs in
other classes of compounds as well. One apparent result of this
effect is that an axial alkoxy substituent is often more stable
than the corresponding equatorial substituent when attached at the
C1 position of a tetrahydropyranyl ring. This effect can be

mimicked in molecular mechanics by a suitable torsional potential that is included in MM1 and all versions of MM2.

Another anomeric effect is that acetal C-O bonds, and to a lesser extent, the bonds between acetal carbons and ether oxygens, are shortened or elongated as a function of their associated torsional angles. Jeffrey and Taylor modified MM1 to account for these anomeric effects (17) and similar additions were put in the standard 1985 version of MM2 (11). The parameterization of MM3 for anomeric effects is preliminary, with recent (18-20) results being monitored.

Versions of MM2 before 1987 calculate energies for hydrogen bonds that are too high, compared to experiment. With the 1987 release, molecular energies are lowered by a variable amount when an atomic sequence that could correspond to an hydrogen bond is detected. The amount depends on the geometry of the atoms involved in the sequence. MM3 was parameterized to account for hydrogen bonding from the beginning. Since the MM3 potential function does not use lone pairs, it has a practical advantage over MM2, especially for carbohydrates. The lone pairs, required for correct use with MM2, increase the number of "atoms" in a carbohydrate molecule, often by 50%, causing calculations with MM2 to take twice as long as with MM3.

Modeling Details

The starting coordinates were from a model residue (Figure 1) with O6 in the gg position (the torsion angle O5-C5-C6-O6 is -60°). The secondary hydroxyls were arranged clockwise. This description applies when the ring is viewed from above (H4 is closer to the viewer than C4). The torsion angles between the hydroxyl hydrogens and the hydrogen atoms on the carbons are roughly +60° at C1, C2 and C4 and -60° at C3. Preliminary studies showed that this arrangement has lowest energy when using the MM2(85) force field. Initially, the default dielectric constants of 1.5 were used (1.0 for CHARMM), suited to isolated molecules. Thirteen models with values of D in the range from 3.9 to 5.1 Å were optimized with each program. D was kept at the starting values by using provisions within the programs to fix some atoms at specified coordinates while optimizing all other atomic positions. The MM2(77), MMP2(85) and MM2(87) programs were versions for Vax computers distributed by the QCPE; MM3 was a pre-release Vax version. CHARMM results were kindly provided by Professor Brady.

Selection from the Cambridge Crystallographic Database

The bond-lengths, bond-angles and torsion angles of each model were compared with information from 46 glucose residues in crystal structures in the 1989 Cambridge Structural Database (CSD) (21) (Table 1). Residues from cycloamyloses were not included since their macrocycles of 6-8 glucose residues impose additional, systematic limits on the ring geometry (3). Molecules that contained disordered oxygen atoms, such as 1-kestose, were also not included. No structures with crystallographic R factors greater than 0.10 were used. Only one R exceeded 0.07 and the mean is 0.044. D ranges over more than 0.6 Å, while the C1--C4 distance varies only one eighth as much, as shown in Table 2.

Table 1:   Refcode, Journal Codon, Volume, Year, Page, [D] and
           Compound Name

BAGZEO   CRBRAT   93   135 1981   [4.367]
   1-O-α-D-Glucopyranosyl-D-mannitol dihydrate
BAVCAC   JLACBF   *   2372 1981   [4.567]
   6-O-(α-D-Glucopyranosyl)-D-glucitol
BAXSEY01   ZKKKAJ 161   69 1982   [4.326, 4.510]
   4-Nitrophenyl-α-D-glucopyranoside
BIZHIB   CRBRAT 108   163 1982   [4.5160]
   4-O-α-D-Glucopyranosyl-D-glucitol
BOPXEJ   ZKKKAJ 160   259 1982   [4.666, 4.442]
   Phenyl-α-D-glucopyranoside
CEKLUZ   ACSCEE   40   389 1984   [4.639]
   Disodium uridine diphophoglucose dihydrate
CELGIJ   ACSCEE   40   531 1984   [4.130]
   O-α-D-Glucopyranosyl-(1-2)-O-β-D-fructofuranosyl-(6-2)-β-
   D-fructofuranoside monohydrate (6-kestose)
CIMDUX   ACSCEE   40   1338 1984   [4.495]
   Disodium glucose-1-phosphate hydrate
DECGPY10   JACSAT   98   6628 1976   [4.360]
   1-Decyl α-D-glucopyranoside
DEKYEX   CRBRAT 137   21 1985   [4.414, 4.526]
   α-D-Glucopyranosyl-α-D-glucopyranoside
DUDXOP   IJBMDR   7   363 1985   [4.480, 4.483, 4.237]
   Methyl-α-maltrotrioside tetrahydrate
FONYUC   ACSCEE   43   1809 1987   [4.218]
   4-O-α-D-Glucopyranosyl-N-methylmoranoline dihydrate
GAFVIS   CRBRAT 169   1 1987   [4.483]
   Octyl α-D-glucopyranoside monohydrate
GAFVOY   CRBRAT 169   1 1987   [4.291]
   Octyl α-D-glucopyranoside hemihydrate
GLUCMH11   ACBCAR   29   365 1973   [4.513]
   α-D-Glucose monohydrate
GLUCSA01   ACBCAR   35   656 1979   [4.486]
   α-D-Glucose
GLUCUR20   ACBCAR   27   1969 1971   [4.476]
   α-D-Glucose-urea complex
IMATUL   ACBCAR   29   514 1973   [4.370]
   Isomaltulose monohydrate
KGLUCP02   ACSCEE   40   389 1984   [4.401]
   Dipotassium glucose-1-phosphate dihydrate
LACTOS10   ACBCAR   27   994 1971   [4.455]
   α-Lactose monohydrate
MALTOS11   ACBCAR   33   2490 1977   [4.410]
   β-Maltose monohydrate
MALTOT   ACBCAR   34   213 1978   [4.052, 4.224]
   α-Maltose
MELEZT01   ACBCAR   32   2598 1976   [4.324, 4.422]
   Melezitose monohydrate
MELIBM10   ACBCAR   34   508 1978   [4.574]
   6-O-Galactopyranosyl-(α,β)-glucopyranose monohydrate
   (melibiose H_2O)
MGLUCP   ACBCAR   24   897 1968   [4.375]
   Methyl α-D-glucopyranoside

Continued on next page

Table 1. Refcode, Journal Codon, Volume, Year, Page, [D] and Compound Name *(Continued)*

```
MMALTS     ACCRA9  23   1038 1967           [4.570]
    Methyl β-maltoside monohydrate
MOGLPR     CRBRAT  80     15 1980           [4.356]
    Methyl-3-O-α-D-glucopyranosyl-α-D-glucopyranoside
PHMALT     ACBCAR  32    155 1976   [4.555, 4.338, 4.478, 4.221]
    Phenyl-α-maltoside
PLANTE10   ACBCAR  28    425 1972           [4.368]
    Planteose dihydrate
RAFINO     ACBCAR  26    290 1970           [4.427]
    Raffinose pentahydrate
STACHY10   ACSCEE  43    806 1987           [4.322]
    O-α-D-Galactopyranosyl-(1-6)-O-α-D-galactopyranosyl-
    (1-6)-O-α-D-glucopyranosyl-(1-2)-α-D-fructofuranoside
    pentahydrate (stachyose)
SUCROS11   ACBCAR  29    797 1973           [4.534]
    Sucrose
TRECAB     CRBRAT  31    265 1973           [4.402]
    α,α-D-Trehalose-calcium bromide monohydrate
TREHAL01   ACBCAR  28   3145 1972        [4.210, 4.340]
    α,α-Trehalose       dihydrate
TURANS01   ACBCAR  34   1873 1978           [4.545]
    O-α-D-Glucopyranosyl-(1-3)-β-D-fructopyranose   (turanose)
```

Table 2.   1--4 Distances (Å) for Glucose Residues

	O1--O4	C1--C4
Mean Distance	4.411	2.881
Std. Deviation of Sample	0.130	0.020
Std. Deviation of Mean	0.019	0.003
Minimum	4.052	2.845
Maximum	4.666	2.919

Systematic changes in bond angles and torsion angles were indicated by plotting the parameters against $\underline{D}$ with the program GRAPHER, available from Golden Software, P.O. Box 281, Golden, Colorado 80402. The experimental bond and torsion angle values were fitted to first order lines. The curves for the models were connections of the points by straight line segments.

Overall Modeling Results

The energy vs. $\underline{D}$ curves are shown in Figure 4 for the five force fields with default dielectric constants. (The energies from each program were normalized by subtracting the lowest energy obtained.) All force fields predict a minimum near 4.55 Å, while the old work with MM1 gave a minimum at 4.27 Å, close to the middle of the range observed at that time (4.30 Å). The mean C1--C4 distance for the five models with $\underline{D}$ of 4.5 Å is 2.869 Å while the MM1 model had a short value, 2.795 Å, that was outside of the observed range.

Although the normalized curves in Figure 4 are nearly identical, the raw values of the minimal energies are different. MM2(77) and MMP2(85) values are 13.0 and 13.3 kcal/mol, while MM2(87) gave 9.3 kcal and MM3 gave 4.0 kcal. The decrease for MM2(87) is caused by the clockwise hydrogen bonding. MM3 is a new force field. Total steric energies from MM1, MM2 and MM3 can be used to calculate enthalpies of formation by adding the steric energy to the sum of strain-free enthalpies. This is not the case for CHARMM, which gave a minimum of 70.2 kcal/mol.

MMP2(85) and MM2(87) models with the hydrogen on O1 gauche to the ring oxygen (as in crystalline glucose) gave normalized curves that were nearly identical to those in Figure 4. Dielectric constants of 4 (solid phase) and 80 (aqueous solution) for MMP2(85) models also gave the same corrected curves. Changes of the position of the hydrogen on O1 and dielectric constant were not investigated with CHARMM or MM2(77). With MM3, a dielectric constant of 4.0 shifted the minimum in the energy vs. $\underline{D}$ curve to 4.472 Å, with a C1--C4 distance of 2.882. (These values were taken from a model that was optimized without any constraint on the O1--O4 distance.) While about 3.2 standard deviations larger than the mean $\underline{D}$ in the CSD, the model value is slightly smaller than in crystals of glucose-urea complex (4.476 Å) (22), glucose (4.486 Å) (23) and glucose monohydrate (4.513 Å) (24), the only unsubstituted glucose structures in the CSD.

Based on the study with MM3 and the dielectric constant of 4, predicted and observed distributions of $\underline{D}$ are shown in Figure 5. The predicted distribution corresponds to an unskewed Gaussian curve. Distributions for the other programs and MM3 with the vacuum dielectric constant were similar, but their maxima were located at about 4.55 Å. We judged that distance to be too long, in part because the mean observed value is about 0.14 Å (7.4 standard deviations) smaller. Also, those predicted distributions based on vacuum dielectric constants called for a large fraction of
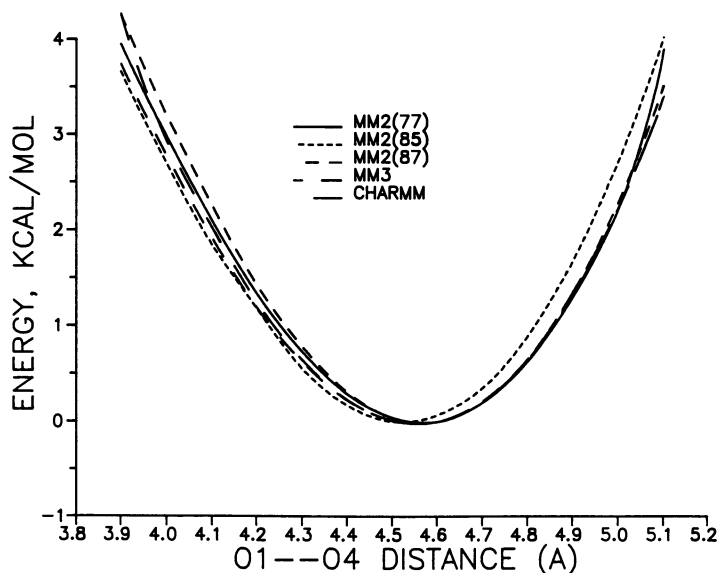
Figure 4. Energy vs. $\underline{D}$ as calculated by five different modeling programs. Default dielectric constants were used. The values graphed were normalized by subtracting the lowest value found with each program from all the other energy values.
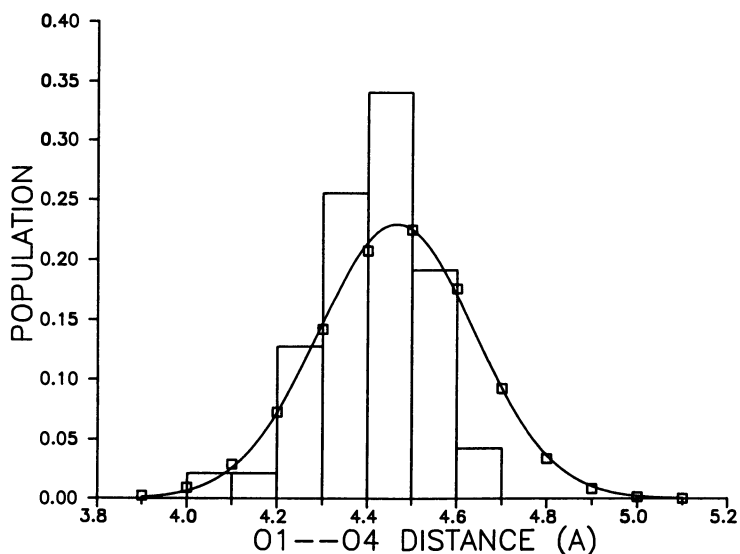
Figure 5. Frequency of $\underline{D}$, as observed (bar graph) and as predicted by MM3, using a dielectric constant of 4.0. The squares are the fractions calculated at each 0.1 A, and the curve is a normal gaussian curve that was fitted to three of the predicted points.

the observed population to have $\underline{D}$ larger than 4.7 Å, and there are no observations of $\underline{D}$ as large as 4.7 Å.

Detailed Comparisons of the Model and Experimental Residues

All five modeling programs gave similarly accurate molecular parameters. One program would fit one parameter better than another program, but overall the qualities of fit were similar. However, the MM3 work with a dielectric constant of 4 gave slightly better overall agreement and predicted a better distribution of $\underline{D}$ and so that is the only work discussed in detail below.
    The various observed parameters were examined for bi-modal distribution that might have resulted, for example, from one family with O6 in gt positions, and another family with O6 in gg postions in crystals. All parameters seemed to be randomly distributed, consistent with slight deformations due to various crystal-packing arrangements. Parameters likely to be correlated with changes in $\underline{D}$ are shown in Table 3 for the three unsubstituted crystalline residues, the Arnott-Scott (25) and CSD averaged residues and the MM3 models freely optimized and held at 4.40 Å. CSD and Arnott-Scott values are based mostly on x-ray diffraction studies done at room temperature, without subsequent corrections for thermal motion. These corrections can lengthen the bond lengths by about 0.005 Å and increase bond angles by about 0.2°. Such errors are probably not important in understanding the flexibility of glucose, but can explain some discrepancies.

Bond Lengths. Since the parameterization for anomeric effects in MM3 is preliminary, it is not surprising that the greatest bond-length discrepancy is for the C1-O1 distance. Further indications of problems related to anomeric effects are that C5-O5 and O5-C1 distances are longer in crystals than in MM3 models while the other, exocyclic C-O bonds (not shown) are modeled very well. One explanation is that the focus in parameterization has been for compounds with a carbon atom attached at O1, as in a methyl glycoside, instead of the hydrogen in this work. Also, the C1-O1 and C1-O5 lengths vary with torsion angle about C1-O1, a factor not monitored in this study of the CSD. The mean length of the crystallographic C5-C6 bond (1.512 Å) is shorter than for the models by 0.015 Å (not shown). This slight shortening (compared to average C-C distances) has been observed before (26) but has been given no attention in modeling studies as far as we know. The endocyclic C-C distances agree well.

Bond Angles. Many of the bond angles show changes with $\underline{D}$. The endocyclic angles are plotted in Figure 6. The scatter in these experimentally determined bond angles is high but the trends are close to those predicted by the models. The exocyclic angles to O1 and O4 are in Figure 7. The discrepancies between models and experiment for these angles are the most severe. The model angles bend at a rate that is consistent with the experimental results, but are offset by as much as 3 degrees.

Torsion Angles. Endocyclic torsion angles change systematically with $\underline{D}$ by as much as 29°/Å (Figure 8). The six different ring torsion angles have experimental ranges of 10.7 to 18.0°. Despite the scatter in the experimental points, the agreements between the observed and predicted slopes are encouraging.

Mean Deviation of Fit. Also shown in Table 3 are the mean distances between the individual atoms of various pairs of rings

Table 3:   Parameters for Crystalline and Model Glucose Residues
           (Distances are in Å and angles are in degrees)

Study	Urea Complx.	Gluc.	Gluc. Hydrate	A-S AVG	CSD AVG	MM3 ε=4	MM3 ε=4 Fixed at 4.40 Å
Parameter							
O1--O4	4.476	4.486	4.513	4.400	4.411	4.472	4.400
C1--C4	2.889	2.874	2.867	2.887	2.881	2.882	2.871
C1-C2	1.517	1.534	1.510	1.523	1.523	1.524	1.523
C2-C3	1.526	1.525	1.522	1.521	1.521	1.524	1.524
C3-C4	1.520	1.520	1.521	1.523	1.522	1.526	1.526
C4-C5	1.524	1.529	1.513	1.525	1.527	1.530	1.530
C5-O5	1.444	1.428	1.451	1.436	1.441	1.425	1.425
O5-C1	1.414	1.427	1.427	1.414	1.417	1.407	1.407
C1-O1	1.384	1.391	1.412	1.415	1.407	1.436	1.435
C4-O4	1.422	1.426	1.435	1.426	1.428	1.438	1.436
O5-C1-C2	110.0	110.1	110.9	109.2	110.0	110.2	110.2
C1-C2-C3	110.9	111.1	112.7	110.5	110.6	111.4	111.1
C2-C3-C4	109.7	109.9	109.0	110.4	109.9	108.6	108.7
C3-C4-C5	109.2	111.2	111.4	110.2	110.8	109.7	110.0
C4-C5-O5	109.8	108.8	108.9	119.9	110.1	109.9	109.7
C5-O5-C1	113.9	113.8	113.1	113.9	114.0	114.8	114.1
O5-C1-O1	112.0	111.5	110.2	109.2	110.7	108.6	107.9
C2-C1-O1	109.1	109.3	110.1	108.4	108.1	111.0	110.9
C3-C4-O4	111.8	108.2	108.6	110.4	110.2	107.2	107.0
C5-C4-O4	106.2	110.9	109.2	109.9	108.2	109.2	108.8
O5-C1-C2-C3	55.4	54.1	53.0	56.9	56.4	54.8	55.8
C1-C2-C3-C4	-54.5	-51.3	-50.5	-53.5	-53.6	-54.5	-54.0
C2-C3-C4-C5	55.2	53.3	53.5	52.5	52.8	55.4	54.7
C3-C4-C5-O5	-57.4	-57.5	-58.5	-54.8	-54.4	-57.3	-57.2
C4-C5-O5-C1	61.1	62.2	61.4	61.4	59.2	60.0	60.8
C5-O5-C1-C2	-59.5	-60.9	-58.7	-62.0	-60.1	-58.3	-59.8
Puckering							
Q (Å)	0.574	0.567	0.560	0.568	0.561[#]	0.568	0.571
Θ	1.9	3.5	4.9	2.4	1.2	2.7	1.5
φ**	267	323	303	24	45	233	258

Average Deviation of 6 Ring Atoms Fitted by Least Squares (A)

Urea Complex	----						
Glucose	.019	----					
Hydrate	.020	.013	----				
Arnott-Scott	.016	.019	.026	----			
Cambridge*	.014	.019	.025	.008	----		
MM3 ε=4 4.47	.010	.020	.021	.021	.017	----	
MM3 ε=4 4.40	.010	.016	.020	.016	.013	.007	----

* This ring was constructed from the average values of the
   above parameters. The ring came within 0.0096 A of
   closing. Its O1--O4 distance was 4.416 A.
[#] This number reflects the slight geometrical changes resulting
   from closing the ring based on average parameters instead of
   using the average value of the observed rings (0.564 Å).
** Values of φ do not imply significantly different
   structures when Θ is close to zero.
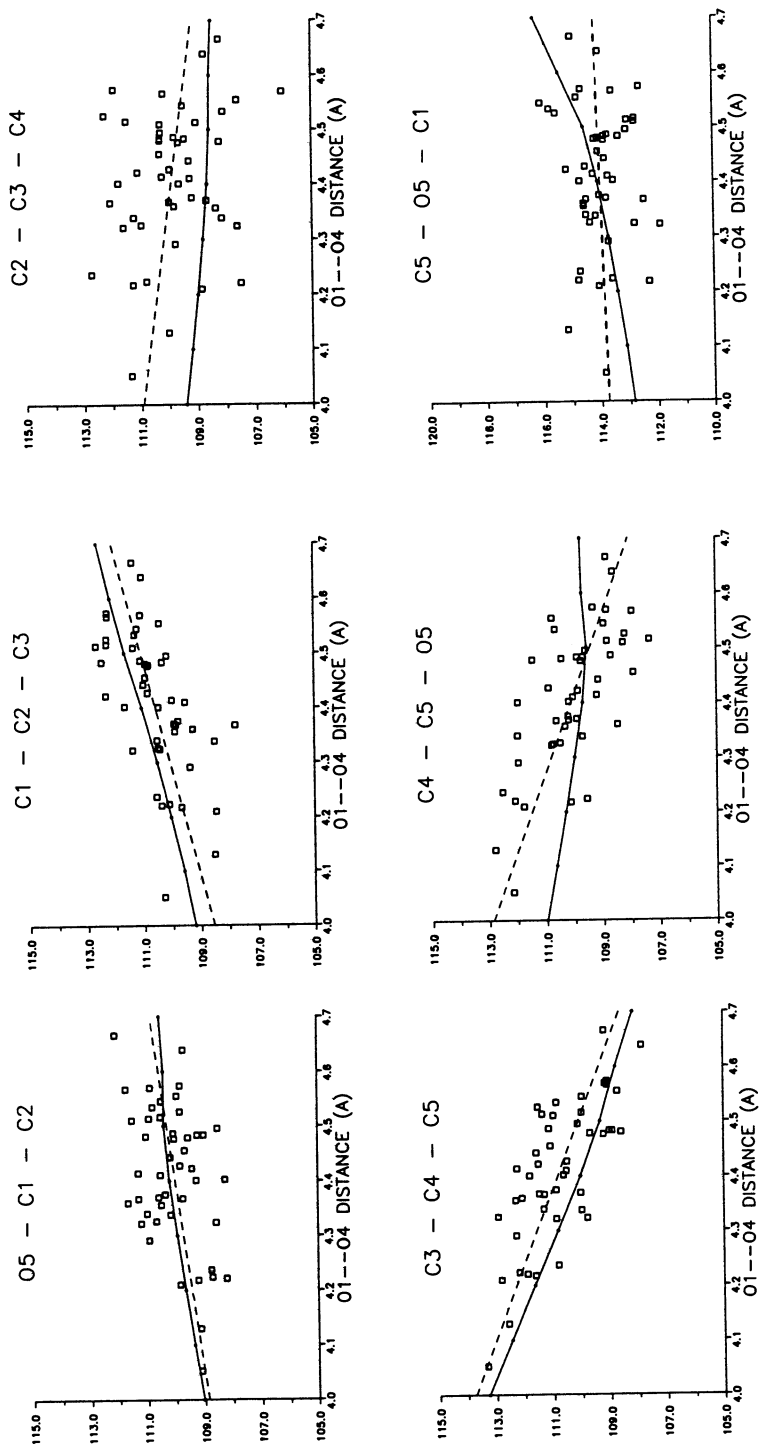
Figure 6. Endocyclic bond angles for glucopyranose plotted against D. The regression line based on the observed values is dashed, and the model values are shown by the solid line.
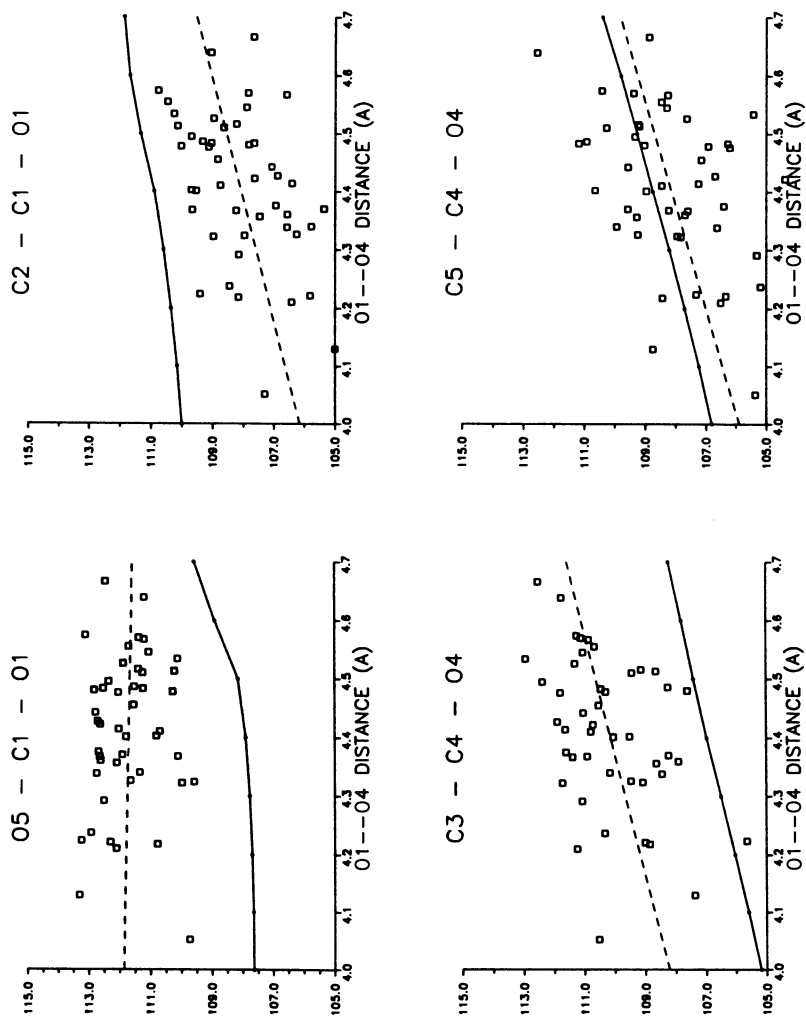
Figure 7. Exocyclic bond angles that could affect $\underline{D}$ as in Figure 6.
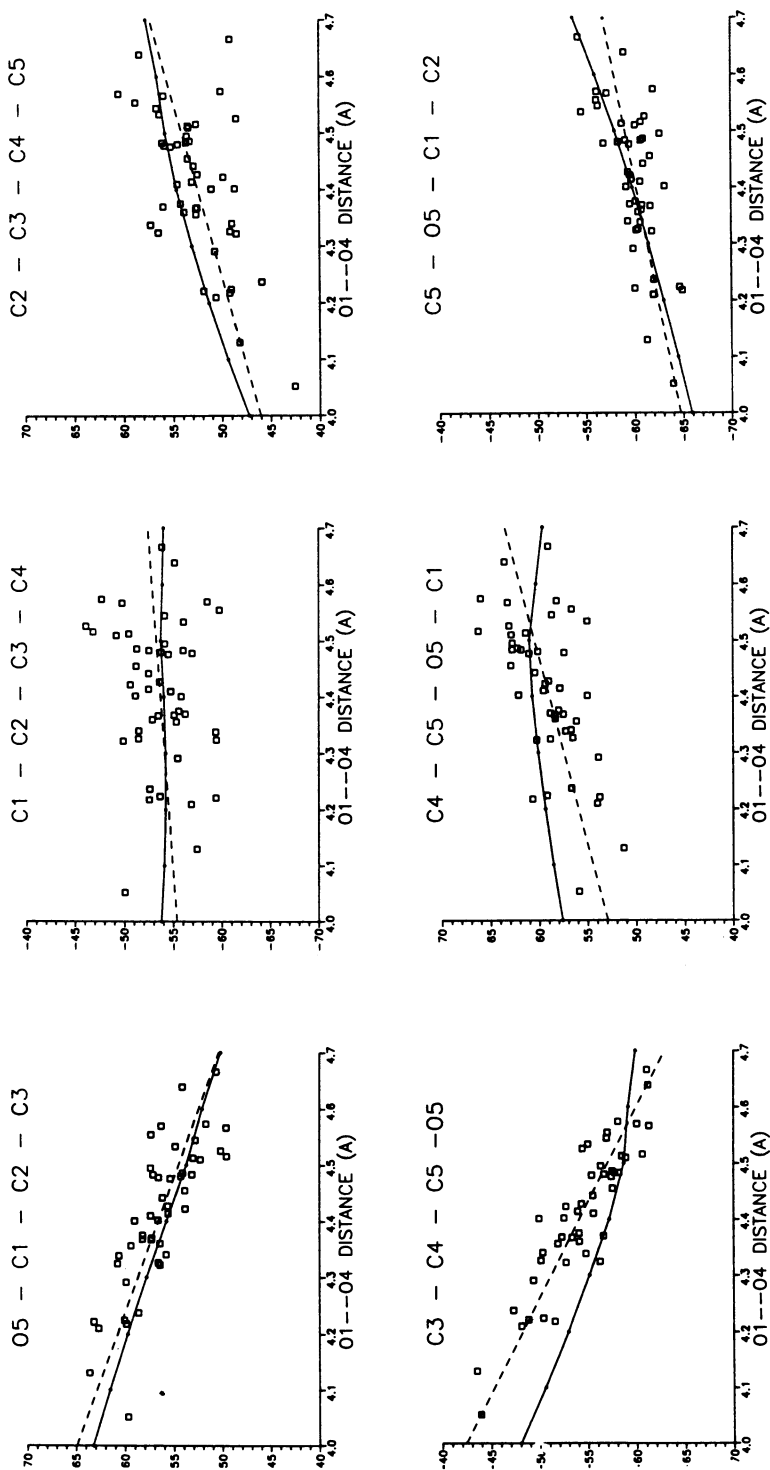
Figure 8. Endocyclic torsion angles, as in Figure 6.

after the rings were superimposed, based on all six ring atoms. In order to include a comparison with a summary of the crystallographic data, a ring was produced from the average bond lengths and angles and torsion angles in the CSD. After generating all the ring atoms, starting with C1 and progressing by increasing carbon number, the distance between O5 and C1 was 0.0096 Å shorter than the average value, but the ring was used without further ado. It fit the Arnott-Scott averaged residue (25) very closely, and was similar to the MM3 models. On a high-resolution video display, rings appear to be identical if their mean deviation of fit for six atoms is less than 0.020 Å, while slight differences can be seen in the positions of pendant atoms.

Cremer-Pople Puckering. Pyranose ring geometry is formally described by the Cremer-Pople puckering parameters Q, Θ, and ϕ (27). These parameters were calculated for the atomic coordinates in the CSD and for the models with a program written by Larry Madsen. Q is the deviation of the ring atoms from a mean plane. ϕ indicates the position of puckering (which atoms deviate most from the mean plane), and Θ indicates the extent of distortion from the perfect $^4C_1$ conformation. Figure 9a illustrates these concepts (see also Figure 3 in the introductory chapter of this book). Table 3 includes the puckering parameters for the seven experimental and model rings.

   There is a very small observed range of amplitude (Q), and Q is essentially invariant with $\underline{D}$ in both models and the CSD. The CSD mean is 0.564 Å while the best MM3 model has a Q of 0.570 Å, agreeing well. Figure 9b displays the observed Q values and the line from the MM3 models.

   Figure 9c plots Θ vs. $\underline{D}$ for the models and observed crystal structures. For this graph, the sign of Θ was changed for points with $\phi > 180^\circ$ to preserve a straight line. The CSD regression line is close to the model line. Perfect chairs ($\Theta = 0.0^\circ$) are found for a model with $\underline{D}$ of 4.35 Å and for the CSD line at 4.43 Å. The least energetic model has a Θ of 2.7°; there is no reason why the energetically optimal residue should be a perfect chair.

   Figure 9d shows the ϕ vs. $\underline{D}$ relationship, and the view in Figure 9e is down toward the top of the Cremer-Pople sphere which has been projected onto a plane. The 0,0 point corresponds to the perfect $^4C_1$ shape. In this plot, the dashed line corresponds to the $\phi = 60/240$ meridian and is not a regression line. Figure 9e shows that Θ varies away from a perfect chair in all directions. Therefore, glucose residues would be flexible when modeling glucans with other linkages as well.

   These studies of puckering support the description of the primary changes in the ring as the O1-O4 distance changes that was given in the introduction. When Θ is 90°, a ϕ of 60° indicates a $B_{1,4}$ conformer and a ϕ of 240° indicates a $^{1,4}B$. Above these equatorial points on the $\phi=60/240$ meridian, at $\Theta = 54.7^\circ$, are the E_1 and ^4E half boats. Therefore, models with short $\underline{D}$ (with ϕ near 60°) tend toward the E_1 conformers and the models with longer $\underline{D}$ and ϕ near 240° tend towards the ^4E half-boats. Q stays almost constant during changes in $\underline{D}$ because lengthening of $\underline{D}$ moves C1 towards the mean plane with C4 moving simultaneously further away from the mean plane.
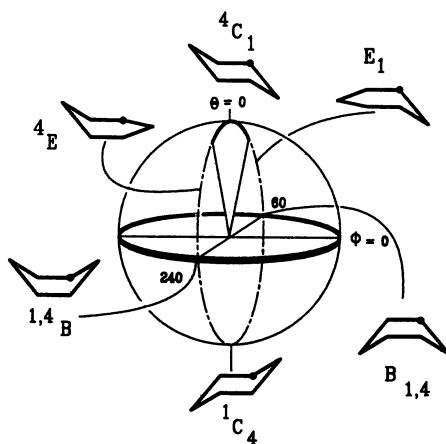
Figure 9a. Spherical polar depiction of pyranose puckering. The equatorial belt is the path of facile pseudorotation through all the Boats and Skew-boats (ϕ rotation). Perfect chairs are at the North and South poles, and the Half-boat (Envelope) forms are at Θ of 54.7°. This diagram has a reversed direction of positive pseudorotation from that shown in figure 3, Chapter 1.
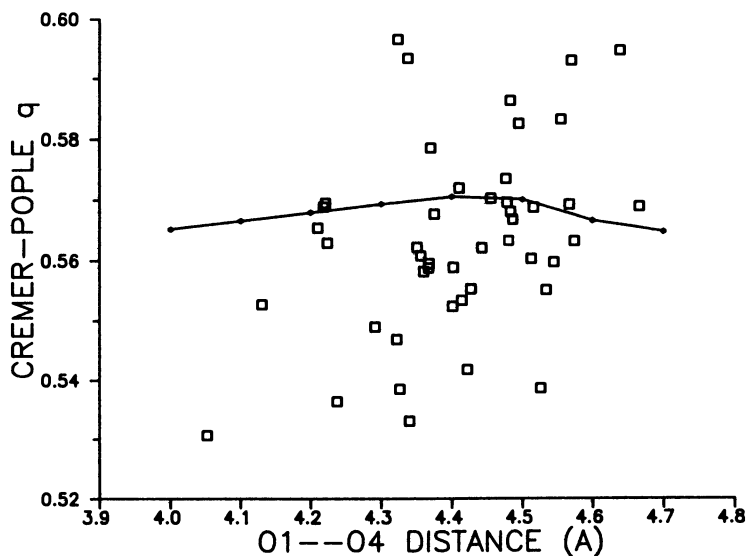
Figure 9b. Observed and predicted Q puckering parameters vs. <u>D</u>. The range of observed values is quite small.
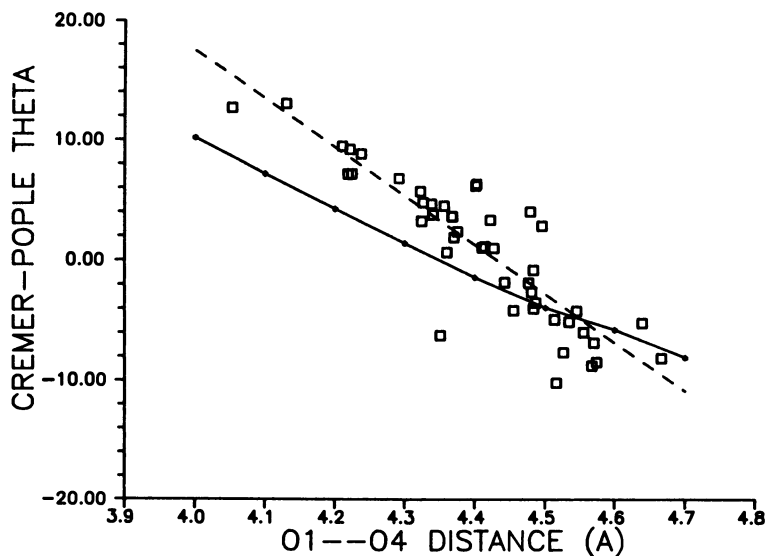
Figure 9c. Observed and predicted Θ puckering parameters vs. D.
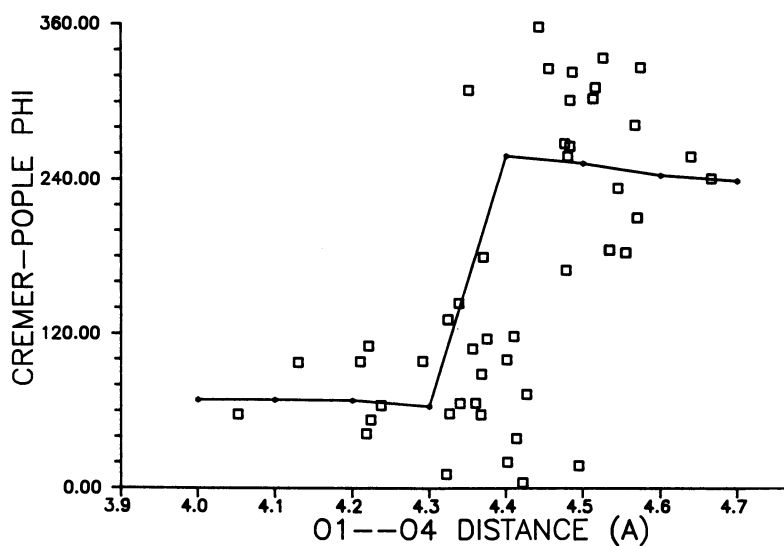
Figure 9d. The φ puckering parameter, plotted against D. At 4.35 Å, the length for a perfect chair, φ changes from 60° to 240°.
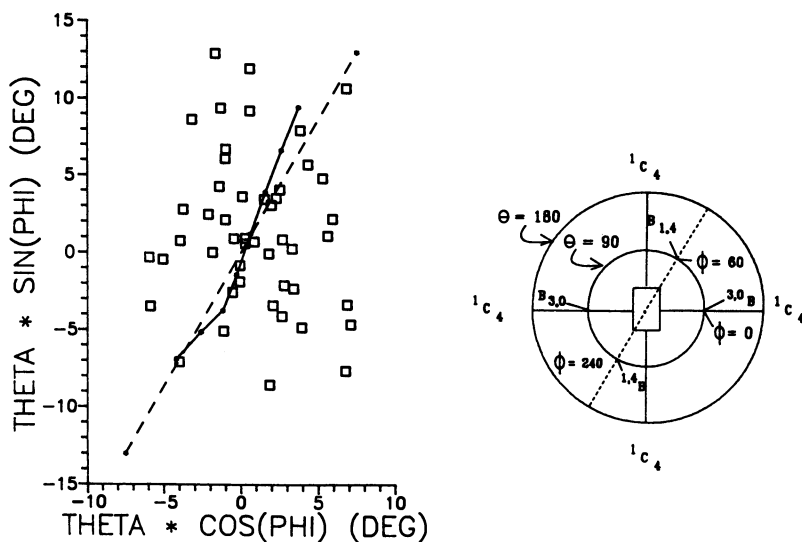
Figure 9e.   A projection of the Cremer-Pople sphere (*see* Fig. 9A) onto a plane perpendicular to the polar axis. Θ and ϕ values for experiment and models were converted to cartesian coordinates, with the model points connected by solid lines. The central (0,0) point corresponds to a perfect chair, and the dashed line follows the 60°−240° meridian.

Discussion and Conclusions

The modeling work shows that the energy to deform the pyranose ring over the observed range of $\underline{D}$ of 4.05 to 4.67 Å (less than 2 kcal/mol) can come readily from environmentally variable factors such as hydrogen bonding and van der Waals forces. Further, examination of Table 1 supports our assertion that such forces are the main cause of these deformations. Besides the example of non-reducing residues from maltoses with $\underline{D}$ of 4.052 and 4.570 Å cited in the introduction, one may compare the glucose residue in sucrose (4.534 Å) with the analogous glucose residues in the sucrose moiety in 6-kestose (4.130 Å) and 1-kestose (28) (4.575 Å). Five residues in Table 1 come from α,α-trehalose. (A sixth residue is related by symmetry to the residue from TRECAB.) While the aforementioned glucose residues in sucrose and maltose moieties are merely isolated from the locations of chemical differences, the trehalose residues are all chemically identical. Their $\underline{D}$ values range from 4.210 Å to 4.526 Å. On the other hand, the Arnott-Scott averaged ring, which was derived from a variety of pyranose sugars, is nearly identical to the average ring produced in this paper from only glucose residues. The alterations in ring geometry arising from the alternate hydroxyl group positions must be very small for the two rings to have a mean deviation of only 0.008 Å.

When determining the range of likely helical shapes from intrinsic properties of amylose, this variability in monomer shape is almost as important as hindered rotation about the bonds linking the monomers. This conclusion is supported by conformational analyses of maltose such as shown in Figure 5 of the introductory chapter of this book. There are relatively small ranges (about 40°) of allowed torsional rotation within one kcal/mol of the minimum (one must correct for the fact that there are two glucose residues in maltose when making such a comparison).

As soon as crystal structure determinations of glucose became available, the question arose as to which of the slightly different geometries would be most appropriate for modeling amylose. This question implies that the residue geometry would remain fixed in a rigid-residue type of modeling study. While not ideal, this approximation still has some utility, arising from the sheer size of polymeric molecules. If environmental effects are ignored, it seems that the Arnott-Scott average residue remains a good choice. Better still, a series of studies should be done with residues with different geometries, and the overall conclusions should include results from each residue geometry. The selection of the residues is simplified for amylose by using $\underline{D}$ as a criterion. Alternatively, the Θ and ϕ puckering parameters could be used.

The MM3 ring with lowest energy leads to collapsed (V-type) amylose helices with about eight residues per turn and would also be suited for modeling the more extended native starch double helices with six residues per turn (See Figure 6 and the material on n-h maps in the introductory chapter of this book. The chapter by Pérez, Imberty and Scaringe also discusses native starch helices). The best MM3 model is also reasonably close in structure to both of the average residues which lead to collapsed amylose helices with seven residues per turn (2). A model with $\underline{D}$ of 4.25 Å, suited for collapsed, six-residue per turn helices, has an internal energy about 0.4 kcal/mol higher than the minimum.

The extensive data on glucose permit study of the rates of systematic changes of molecular parameters that correlate with $\underline{D}$ in models and crystal structures. All five modeling programs balanced bond stretching, angle bending and torsional motion well enough that the rates of most changes were consistent with experiment.

Since these factors are well balanced, flexible-residue
conformational analyses using any of these force fields should be
reasonably correct, within limitations such as neglect of
environment.

Comparing the mean deviations of fit in Table 3, it seems that
the quality of the modeling software enables prediction of the
structure of the glucose ring by modeling to be comparable to
prediction by crystal structure determination. The structure of
the least energetic MM3 ring is especially close to the crystal
structure of the glucose-urea complex and very close to crystalline
glucose itself and to glucose monohydrate.

There are some defects in the model that we hope can be
resolved in a future release of MM3. The bond lengths and bond
angles around the anomeric center are the most pressing. Although
there was little effect of different anomeric substituents on bond
lengths with the MM2 force field, this will have to be studied at
length with MM3, as well as following the torsion angles that can
affect these bond lengths. On the other hand, the magnitude of
these errors is probably not very important when attempting to
determine the properties of a polymer. Of the errors involving
bond lengths and bond angles, those involving bond angles are
probably more important in modeling the polymer.

Since the studies with MM1 a decade ago ($\underline{9}$), the predicted
best $\underline{D}$ increased 0.2 Å and the mean experimental $\underline{D}$ grew 0.11 Å.
Improvement of the software is perhaps most easily shown by the
better C1--C4 distance which was previously predicted to be outside
the observed range. The current model predicts that a few
structures should be observed with $\underline{D}$ between 4.7 and 4.9 Å, and
perhaps some will eventually be found. More low-temperature
diffraction results would be welcome. However, this modeling study
of isolated models has not accommodated two factors that might
shorten the upper limit on $\underline{D}$ and skew the distribution. Residues
with $\underline{D}$ longer than 4.6 Å may not fit together in crystals as
compactly as shorter residues. Such a decrease in the density
would increase the packing energy. Another potential cause of a
skewed distribution is the influence of puckerings other than those
with ϕ = 60 and 240°. Figure 9e shows that the models mostly
puckered along that line, while the crystal structures have much
more random puckerings. For a given increase in potential energy,
the random puckerings could result in shorter $\underline{D}$ than those from
models that were purely puckered on the E_1--^4E path.

Distribution Version of MM3

After completion of this work, the parameters for MM3 for alcohols
and ethers were finalized, and papers describing these parameters
were submitted to J. Am. Chem. Soc. Most of the results for
glucose do not change significantly, although there are some small
differences. For example, $\underline{D}$ is 4.469 Å compared to 4.471 Å. A
notable change regards the O1-C1-O5 bond angle, which increased
1.8°, while the C-C4-O4 angles decreased by 0.5°.

Acknowledgments

U.S. Department of Agriculture, but for the purposes of
reproduction of the work.

Literature Cited

1. Goebel, C.V.; Dimpfl, W.L.; Brant, D.A. *Macromolecules*
    1970, 3, 644-654.
2. French, A.D.; Murphy, V.G. *Carbohydr. Res.* 1973, 27, 391-406.
3. French, A.D.; Murphy, V.G. *Polymer*. 1977, 18, 489-494.
4. Saenger, W. *Biochem. and Biophys. Res. Comm*. 1980, 92,
    933-938.
5. Zugenmaier, P.; Sarko, S. *Biopolymers* 1976, 15, 2121-2136.
6. Chu, S.S.C.; Jeffrey, G.A. *Acta Crystallogr*. 1967, 23,
    1038-1049.
7. Takusagawa, F.; Jacobson, R.A. *Acta Crystallogr*. 1978,
    B34, 213-218.
8. Ferro, D.R.; Hermans, J. *Acta Crystallogr*. 1977, A33,
    345-347.
9. Pensak, D.A.; French, A.D. *Carbohydr. Res*. 1980, 87, 1-10.
10. Allinger, N.L. *J. Am. Chem. Soc*. 1977, 99 8127-8134.
11. Nørskov-Lauritsen, L.; Allinger, N.L. *J. Comput. Chem*.
    1984, 5, 326-335.
12. Allinger, N.L.; Yuh, Y.H.; Lii, J-H. *J. Am. Chem. Soc*.
    1989, 111, 8551-8566.
13. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States,
    D.J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem*.
    1983, 4, 187-217.
14. Ha, S.N.; Giammona, A.; Field, M.; Brady, J.W.
    Carbohydr. Res. 1988, 180, 207-221.
15. Edward, J.T. *Chem. Ind. (London)*, 1955, 1102-1104.
16. Lemieux, R.U.; Koto, S.; Voisin, D. In *Anomeric Effect.
    Origin and Consequences*; Horton, D.; Szarak, W.A. Eds.;
    ACS Symposium Series 87; American Chemical Society:
    Washington, DC, 1979; pp 17-29.
17. Jeffrey, G.A.; Taylor, R. *J. Comput. Chem*. 1980, 1, 99-109.
18. Longchambon, F; Gillier-Pandraud, R.; Wiest, R.; Rees,
    B.; Bitschler, A.; Feld, R.; Lehman, M.S.; Becker, P.
    Acta Crystallogr. 1985, B41, 47-56.
19. Fuchs, B.; Ellencweig, A.; Tartakovsky, E.; Aped, P.
    Angew. Chem. 1986, 98, 289-90.
20. Pichon-Pesme, V.; Hansen, N.K. *J. Molec. Struct.
    (Theochem.)* 1989, 183, 151-160.
21. Allen, F.H.; Bellard, S.; Brice, M.D.; Cartwright, B.A.;
    Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-
    Peters, B.G.; Kennard, O.; Motherwell, W.D.S.; Rodgers,
    J.R.; Watson, D.G. *Acta Crystallogr*. 1979, B35, 2331.
22. Snyder, R.L.; Rosenstein, R.D. *Acta Crystallogr*. 1970, B27,
    1969-975.
23. Brown, G.M.; Levy, H.A. *Science* 1965, 147, 1038-1039.
24. Hough, E.; Niedle, S.; Rogers, D.; Troughton, P.G.H.
    Acta Crystallogr. 1973, B29, 365-367.
25. Arnott, S.; Scott, W.E. *J. Chem. Soc. Perkin II* 1972,
    324-335.
26. Rohrer, D.C. *Acta Crystallogr*. 1972, B28, 425-433.
27. Cremer, D.; Pople, J.A. *J. Am. Chem. Soc*. 1975, 97, 1354-1358.
28. Jeffrey, J.A.; Park, Y.J. *Acta Crystallogr*. 1972, B28,
    257-267.

Chapter 8

Molecular Modeling of Acyclic Carbohydrate Derivatives N,N'-Dimethyl- and N,N'-Dihexylxylaramide

Model Compounds for Synthetic Poly(hexamethylenexylaramide)

L. Chen[1], B. Haraden[1], R. W. Kane[1], D. E. Kiely[1], and R. S. Rowland[2]

[1]Department of Chemistry and [2]Department of Biochemistry, University of Alabama, Birmingham, AL 35294

Some results on the molecular modeling of N,N'-dimethylxylaramide (1) and N,N'-dihexylxylaramide (2) using *MacroModel V.2* are presented. Nine minimized conformers were considered and their populations calculated. Average $J_{2,3}$-$J_{3,4}$ couplings were then calculated and those values compared to experimental coupling values. A good fit was obtained for each compound after adjusting the van der Waals cutoff radii option to 3 Å. Two *sickle* conformations and an *extended zig-zag* conformation were calculated as the dominant conformers for the xylaramides 1 and 2. A case is made for the similarity in conformational populations of xylitol and xylaramides, both unprotected and as hydroxyl protected forms.

We have recently developed a synthetic procedure for the preparation of polyhydroxypolyamides (hydroxylated nylons) of general structure I (<u>1-2</u>). While our synthetic method has some unique

$$-[\overset{\overset{\text{O}}{\|}}{\text{C}}-(\text{CHOH})_x-\overset{\overset{\text{O}}{\|}}{\text{C}}-\underset{\text{H}}{\text{N}}-(\text{CH}_2)_y-\underset{\text{H}}{\text{N}}]-_n$$
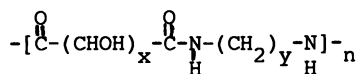
I

features, syntheses of examples of this class of polymer have been previously reported, first by Ogata and co-workers (<u>3</u>) and more recently by Hoagland (<u>4</u>). In order to study the conformational characteristics of the acyclic carbohydrate monomer components of such polymers, we have recorded the ^{1}H NMR spectra of the polymers and begun to compare experimental coupling constants with those generated using molecular modeling techniques. Results as applied to N,N'-dimethylxylaramide (1) and N,N'-dihexylxylaramide (2), models for poly(hexamethylenexylaramide) (3), are presented.

RESULTS AND DISCUSSION

The ^{1}H NMR spectra (300 MHz) of the model diamides **1** and **2**, and the polyamide **3**, were recorded in trifluroacetic acid-d (TFA-d). Since the xylaramide component of **1-3** is symmetrical, one observes a single average coupling for H_2-H_3 and H_3-H_4. Experimental values for $J_{2,3}$ = $J_{3,4}$ are: for **1**, 3.45 Hz; for **2**, 3.32 Hz, and for **3**, 3.26 Hz. (See Table I.) *Figure* 1 shows the relevant portion of the ^{1}H NMR spectra from model compound **2** and the polymer **3**. Since the vicinal proton coupling values for **1-3** are all very similar, it was concluded that **1** and **2** should be good conformational models for the carbohydrate component of the polyamide **3**. We then turned to molecular modeling to try to determine the populations of the principal (low energy) conformations of **1** and **2** that give rise to the observed average coupling values.

Measured vicinal proton coupling values have been used extensively to assign approximate dihedral angles and preferred conformations for acyclic carbohydrates (5), commonly as their per-*O*-acetyl derivatives (6,7). Based upon a number of such studies it has been concluded that conformations of acyclic carbohydrates in solution with 1,3-parallel interactions between OH or OR groups are unfavorable. To relieve these unfavorable interactions acyclic carbohydrates typically undergo 120° rotations about appropriate C-C bonds to form "bent" or "sickle" conformations. Of particular relevance to the subject of this paper, are the studies by Angyal et al. on the conformational analysis of xylitol pentaacetate (5, reference 8) and a recent report by Franks and co-workers describing a high field (620 MHz) ^{1}H NMR study on xylitol itself (4, reference 9). Xylitol and its pentaacetate, like the xylaramides of our study, are symmetrical acyclic xylose derivatives with a single average coupling from H_2-H_3 and H_3-H_4. Vicinal proton coupling constants ($J_{2,3}$ = $J_{3,4}$) for compounds **1-5** are given in Table I. It is of interest to note

TABLE I. $J_{2,3}$ COUPLING

Constants for Compounds 1 - 5

	1	2	3	4	5
$J_{2,3}$ (Hz)	3.45[a]	3.32[a]	3.26[a]	4.49[b] 3.65[c] 3.40[d]	5.2[e]

a. Trifluroacetic acid-d as solvent
b. D_2O as solvent, reference 9
c. Pyridine-d5 as solvent, reference 9
d. Acetone-d6 as solvent, reference 9
e. $CDCl_3$ as solvent, reference 8

that the backbone proton signals from xylitol pentaacetate are adequately separated at 250 MHz (8), but those of xylitol are poorly separated at the same spectrometer frequency.
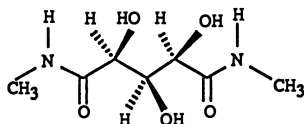
1
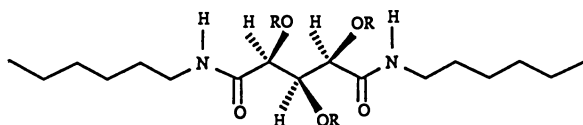
2 R = H

2a R = TMS
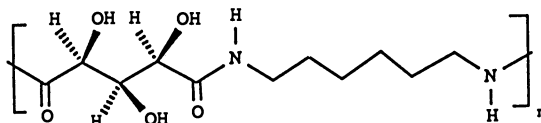
3
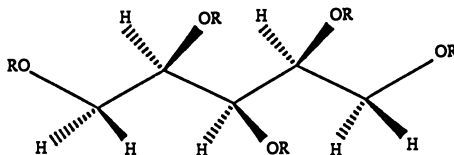
4 R = H

5 R = Ac

FIGURE 1. ^{1}H NMR SPECTRA (3.3 – 5.1 ppm) of Poly(hexamethylenenexylaramide) (**3**) and *N,N'*-Dihexylxylaramide (**2**) in TFA. Signals at **4**.99 and 4.88 are from the xylaramide moiety (H-2,4 and H-3 respectively); the signal at 3.51 ppm is from ND-CH$_2$.

Angyal and co-workers (**8**) concluded that the conformational distribution of xylitol pentaacetate in $CDCl_3$ is between two sickle conformations with the *extended zig-zag* conformation being unimportant. The $J_{2,3}$-$J_{3,4}$ average value of 5.2 Hz is from a large coupling from antiparallel H_3-H_4 (or H_2-H_3) and a small coupling from H_2-H_3 (or H_3-H_4). By comparison, Franks et al. (**9**) obtained couplings for H_2-H_3 and H_3-H_4 of xylitol in three different solvents with all values being lower than those recorded for xylitol pentaacetate (Table I). The $J_{2,3}$ values for xylitol in pyridine-d5 and acetone-d6 (3.65 and 3.40 Hz respectively) are close in value to those we observed for compounds **1-3** (3.48-3.26 Hz range) suggesting average conformational similarity of xylitol and xylaramides around the C_2-C_3-C_4 bonds. These smaller coupling values also suggest a lower "sickle" conformation contribution from xylitol and the xylaramides than is observed with xylitol pentaacetate.

In an attempt to test this latter hypothesis for the xylaramides **1-3**, we carried out a molecular modeling study using the MM2 based *MacroModel V 2.0* program (**11-12**). This study was done using an Evans and Sutherland Terminal PS 350 and a Vax 11-750 computer. As stated in reference **11**, "The *MacroModel* MM2 field differs from the standard field in that it uses the point charge electrostatic model with partial charges derived from the MM2 bond dipoles whereas the standard MM2 electostatic treatment uses a dipole-dipole model." Information concerning the *MacroModel* molecular modeling program is available from W. C. Still, Department of Chemistry, Columbia University, New York, New York 10027. For all of the calculations we used the same protocol in choosing conformations for minimization, but obtained different results by changing program parameters available on the program menu.

Protocol Used For Selection of Conformations to Be Minimized

1) The H-N and C=O of the amide (H-N-C=O) groups were placed in the more stable *anti* relationship (**13**).
2) The fully *extended zig-zag* conformation (alkyl groups and *xylo* component) was minimized and minimizations were then done on five additional conformers generated by 60° increment rotations around the C_1-C_2 bond of the *xylo* moiety. A second set of conformers was produced by rotation in 60° increments around the C_4-C_5 bond on the lowest energy conformer in the first set. This process gave a single, minimized, fully extended, *zig-zag* conformer simply designated as *Extended* (*Figure 2*).
3) Successive rotations of 120° in a clockwise or counterclockwise direction around the C_2-C_3 and C_3-C_4 bonds were carried out on the *Extended* conformation from above. Each of these conformers was minimized further by 60° rotations around the C_1-C_2 and C_4-C_5 bonds. This process gave an additional eight conformers (*Figure 2*).

4) Conformer populations were calculated on the basis of the following relationship:

$$Na/No \; = \; e^{-\Delta E/RT}$$

Where Na/No = the molar ratio of some rotamer a to the most stable rotamer o, with ΔE = energy difference between the two rotamers (14).
    The population (in percent) of any rotamer a among a total of i rotamers is:

$$Pa = \frac{Na/No}{\Sigma \, Ni/No} \; x \; 100$$

    5) Theoretical coupling constants for individual rotamers were obtained directly from *MacroModel* employing an empirical generalization of the Karplus Equation (15). Calculated average couplings came from the expression:

$$Jcalcd = \Sigma \, Xi \; x \; Ji$$

X and J are the population and average coupling (in our case the average of $J_{2,3}$ and $J_{3,4}$) respectively from a total of i rotamers (16).
    The conformers in *Figure 2* are named descriptively and given a rotational label using terminology appropriate for acyclic carbohydrate conformers [(5, 16 (pp 182-203)].

Results From Two Sets of Calculations Using Different MacroModel Options.

For the first set of calculations (**Table II**) we selected the following menu options: a) electrostatic and hydrogen bonding option **On**; b) dielectric constant set to that of TFA, 8.6 (17); c) van der Waals cutoff radius option set to 7 Å (default value). When these options were applied to N,N'-dimethylxylaramide (1), two sickle forms dominated (*Sickles-1* and *3*) with notable contributions from the *Extended, U-1*, and *S-1* conformations. $J_{2,3}$Calcd (4.43 Hz) was significantly larger than $J_{2,3}$expt (3.45 Hz).[3] From these and other data, and from close examination of components of the total energy for each conformation (particularly the van der Waals and electrostatic contributions), we decided to repeat the calculations using a van der Waals cutoff radius of 3 Å, down from the default value of 7 Å. This value was chosen to insure that interactions between parallel 1,3-hydroxyl groups (separated by about 2.5 Å) would be calculated and to correct for artificially high percentages of the U conformations (particularly U-1) due to van der Waals stablizing interactions between alkyl chains.
    Using the smaller van der Waals cutoff radius (3.0 Å, the calculations gave three conformations (*Sickles-1* and *3*, and *Extended*) totaling almost 93% of the total population (**Table II**). As expected, *Sickle-1* and *Sickle-3* were the dominant conformers, but the *Extended* conformer was also calculated to be present in
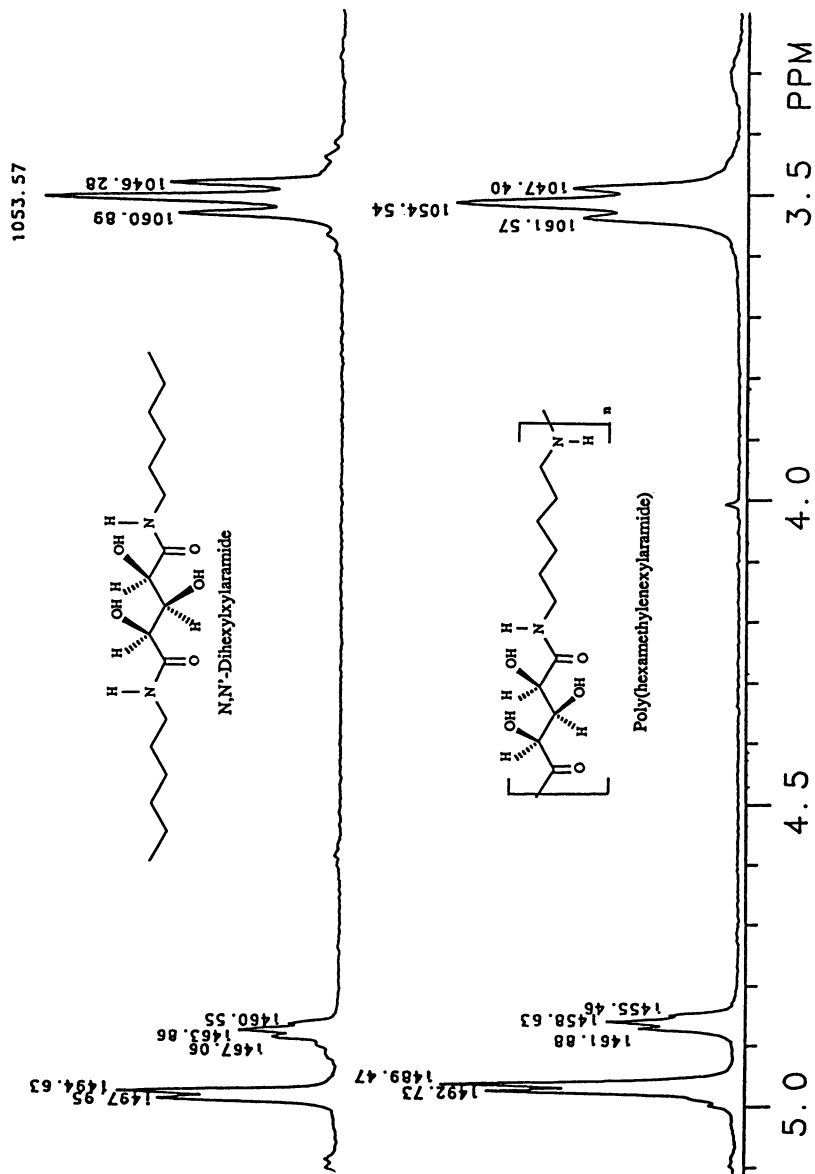
FIGURE 2. Minimized *Extended, Sickle,* and *U* conformations for *N'N'*-Dimethylxylaramide (1) and *N,N'*-Dihexylxylaramide (2).
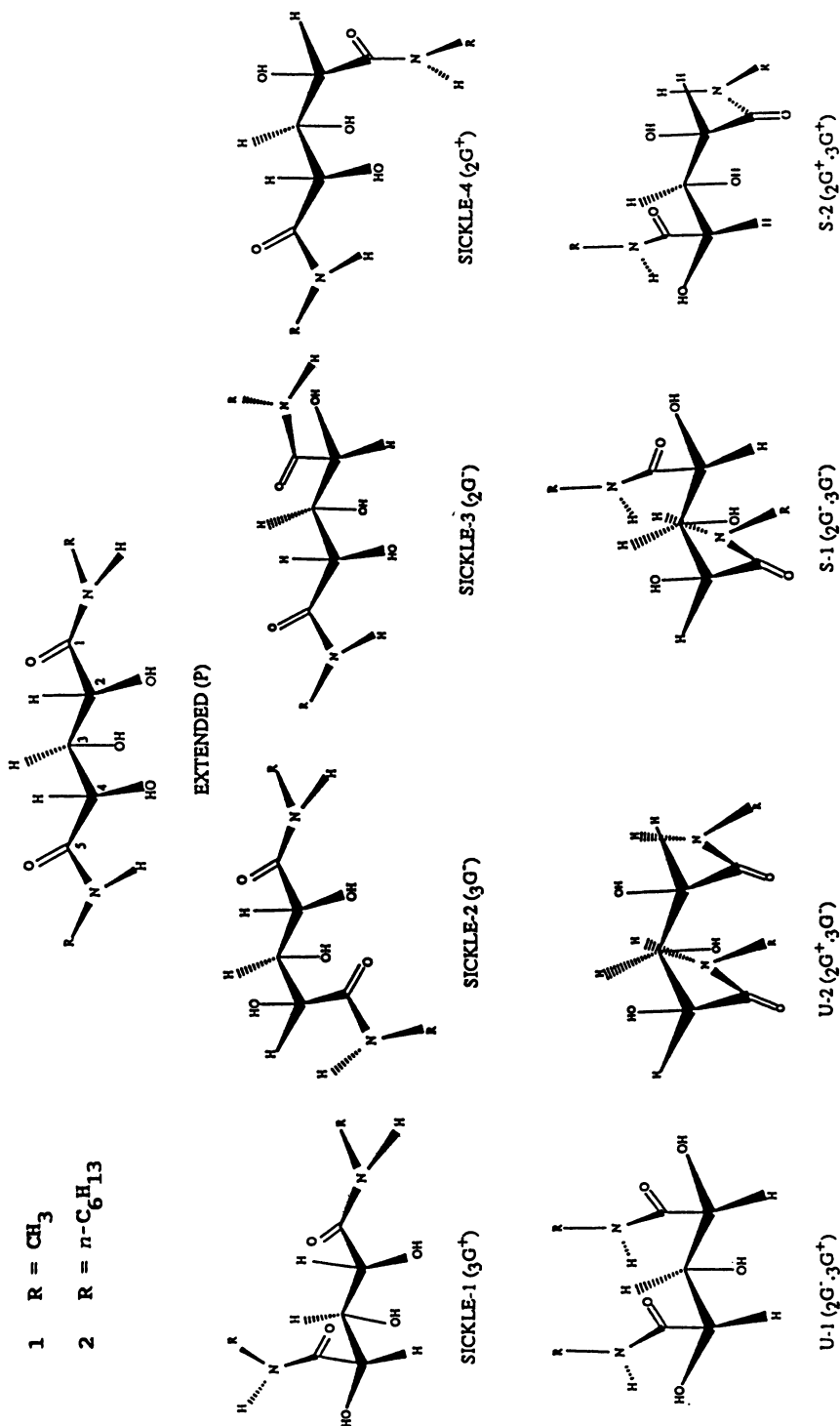
Figure 2 (Continued). *N,N*-Dihexylxylaramide.

Table II.   Calculated Energies and Conformation Populations of N,N'-Dimethylxylaramide (1) Using *MacroModel* V 2^a

Descriptive Label	Rotational Label	KJoul	KCal	Energy Population (%)
Extended	P	10.10	3.08	16.4
		14.36	3.43	26.7
Sickle-1	$_3G^+$	8.95	2.14	26.2
		13.63	3.26	36.0
Sickle-2	$_3G^-$	16.58	3.96	1.1
		22.28	5.32	1.0
Sickle-3	$_2G^-$	8.13	1.94	36.7
		14.06	3.36	30.2
Sickle-4	$_2G^+$	16.44	3.93	1.2
		22.09	5.28	1.1
U-1	$_2G^-$, $_3G^+$	11.97	2.86	7.6
		22.28	5.32	1.0
U-2	$_2G^+$, $_3G^-$	23.12	5.53	0.1
		31.34	7.49	0.0
S-1	$_2G^-$, $_3G^-$	12.29	2.94	6.7
		23.27	5.56	0.7
S-2	$_2G^+$, $_3G^{+-}$	13.49	3.22	3.8
		19.59	4.68	3.1

a.   The first set of energy and population values for each conformer was obtained using the following *menu options*:   electrostatic and hydrogen bonding set to **On**, dielectric constant set to **8.6 D**, van der Waals cutoff radii set to **7** Å.   The second set of values for each conformer was obtained using the same menu options as above except the van der Waals cutoff radii option which was set to **3** Å.

significant amount (27%). The $J_{2,3}$calcd (3.53 Hz) from these modeling calculations was in good agreement with $J_{2,3}$expt (3.45 Hz).

When a van der Waals cutoff radius of 3 Å was used to calculate the conformational populations of *N,N'*-dihexylxylaramide (2), similar results were obtained (**Table III**). *Sickles-1* and *3* and the *Extended* conformation accounted for 94% of the total conformational population. Again, good agreement was obtained for $J_{2,3}$calcd (3.66 Hz) and $J_{2,3}$expt (3.32 Hz).

An important point yet to be made concerns the shape of what we have described as the *minimized-extended* or simply the "Extended" conformation. This conformation is in fact slightly bent, in order to partially relieve the parallel 1,3 hydroxyl interactions, as shown in the calculated differences in dihedral angles between the vicinal protons of $H-C_3-C_2-H$ and $H-C_4-C_3-H$. For 1 the angles were 61.0° and 53.5°, respectively and for 2, 61.1° and 53.7°.

In conclusion, it appears that xylaramides 1 and 2 in TFA, like xylitol (**4**) in non-aqueous solvents (9), have more conformational flexibility than does xylitol in water (9) and most certainly xylitol pentaacetate (5) in chloroform. Furthermore, given the similar $J_{2,3}$expt values of xylitol, 1 and 2 (non-aqueous solvents, Table I), xylitol may be expected to have about the same conformational distribution as has been calculated for 1 and 2. The conformational similarity between xylitol and unprotected xylaramides is paralleled by the similarity between hydroxyl protected xylitol (xylitol pentaccetate, 5) and a hydroxyl protected xylaramide [tris-*O*-(trimethysilyl)*N,N'*-dihexylxylaramide, **2a**]. The average $J_{2,3}$expt values for these compounds in $CDCl_3$ respectively, are 5.2 Hz (8) and 5.3 Hz. (Details will be reported elsewhere).

The results described in this paper represent an important first step toward understanding the conformations of the polyamides I, and may provide useful information for the conformational analysis of xylitol.

EXPERIMENTAL

General Procedures – All solvents were evaporated under reduced pressure at 40 °C. Infrared (IR) spectra were recorded on a Beckman Aculab spectrometer and ^{1}H NMR spectra were recorded at 300 Mhz (Nicolet Fourier Transform Spectrometer).

N,N'-Dihexylxylaramide (2). To a 250 mL round-bottom flask equipped with a magnetic stirrer was added methanol (150 mL), and the flask then cooled to 5 °C. Acetyl chloride (5 mL) was added to the cold methanol and then xylaric acid (18 g, 0.18 mol, reference 18) was added to the methanolic HCl solution. The reaction mixture was refluxed for 16 h, concentrated to a syrup, and residual water removed from the syrup by azeotropic distillation with benzene. Esterification was complete but the product (1) contained more than one ester component as both ester and 5-membered lactone functions were observed in its IR spectrum (neat, 1745 and 1795 cm^{-1}, ester and lactone C=O respectively).

Table III. Calculated Energies and Conformation Populations
    of *N,N'*-Dihexylxylaramide (2) Using *MacroModel V 2*[a]

Descriptive Label	Rotational Label	Energy KJoul	KCal	Population (%)
Extended	P	52.81	12.62	22.5
Sickle-1	$_3G^+$	51.32	12.27	41.5
Sickle-2	$_3G^-$	59.59	14.24	1.4
Sickle-3	$_2G^-$	52.09	12.45	30.3
Sickle-4	$_2G^+$	59.87	14.31	1.2
U-1	$_2G^-, \ _3G^+$	61.20	14.63	0.7
U-2	$_2G^+, \ _3G^-$	68.76	16.43	0.0
S-1	$_2G^-, \ _3G^-$	61.10	14.60	0.7
S-2	$_2G^+, \ _3G^+$	59.33	14.18	1.5

a. The energy and population values for each conformer were obtained using
the following *menu options*: electrostatic and hydrogen bonding set to **On**,
dielectric constant set to **8.6 D**, van der Waals cutoff radii set to **3 Å**.

To a portion of the esterification product (11 mmol) in methanol
(18 mL) was added triethylamine (1 mL), then hexylamine (2.53 g,
25 mmol), and the reaction mixture was refluxed overnight. The
reaction mixture was cooled to room temperature and the solid
product (2) isolated by vacuum filtration, washed with cold
methanol, and dried: yield 2.42 g, 62%; mp 177-180 °C; IR (KBr)
1640 and 1545cm^{-1}, Amide I and Amide II C=O; ^{1}H NMR (TFA-d) δ 4.99
(d, 2H, H-2 & 4, $J_{2,3}$ = $J_{3,4}$ = 3.32 Hz), 4.88 (t, 1H, H-3, $J_{2,3}$ =
$J_{3,4}$ = 3.32 Hz), 3.51 (t, 4H, H-1', $J_{1',2'}$ = 7.2 Hz), 1.70 (m, 4H,
H-2'), 1.39 (m, 12H, H-3',4' and 5'), 0.94 ppm (t, 6H, 6', $J_{5',6'}$
= 6.65 Hz).

Anal. Calcd for $C_{17}H_{34}N_2O_5$: C, 58.93; H, 9.89; N, 8.09.
Found: C, 59.01; H, 9.94; N 8.04.

N,N'-Dimethylxylaramide 1. To a 250 mL round-bottom flask
equipped with a magnetic stirrer was added methanol (50 mL) and
the xylaric acid - methanol esterification product (6.7 mmol in 10
mL of methanol) prepared as described above. Triethylamine (6.5
mL, 46 mmol) and methylammonium chloride (1.0 g, 14.7 mmol) were
then added to the methanol solution and the reaction mixture was
refluxed with stirring overnight. The mixture was concentrated
and the residue was washed several times with methanol. The

product was not appreciably soluble in methanol and was obtained as white crystals: yield 0.52 g, 37% ; mp 203-205 °C; IR (KBr) 1660, 1630 and 1560 cm^{-1}, Amide I and Amide II C=O; ^{1}H NMR (TFA-d) δ 4.91 (d, 2, H-2 & 4, $J_{2,3} = J_{3,4} = 3.45$ Hz), 4.79 (t, 1, H-3, $J_{2,3} = J_{3,4} = 3.45$ Hz), 3.03 ppm (s, 3, NCH$_3$).
    Anal. Calcd for C$_7$H$_{12}$N$_2$O$_5$: C, 40.78; H, 6.84; N, 13.59. Found: C, 40.87; H, 6.84; N, 13.51.

Literature Cited

1. Lin, T-H Ph.D. Dissertation, University of Alabama at Birmingham, Birmingham, Alabama, 1987.
2. Kiely, D. E.; Lin, T-H U. S. Patent 4 833 230, 1989.
3. Ogata, N.; Sanui, K.; Hosoda, Y.; Nakamura, H. *J. Polym. Sci.* 1976, **14**, 783-792.
4. Hoagland, P. *Carbohydr. Res.*, 1981, **98**, 203-208.
5. Horton, D.; Wander, J. D. *J. Org. Chem.*, 1974, **39**, 1859-1863 and earlier references cited therein.
6. Blanc-Muesser, M.; Defaye, J.; Horton, D. *Carbohydr. Res.*, 1980, **87**, 71-86.
7. Sweeting, L. M.; Coxon, B.; Varma, R. *Carbohydr. Res.*, 1979, **72**, 43-55.
8. Angyal, S. J.; Le Fur, R.; Gagnaire, D. *Carbohydr. Res.*, 1971, **23**, 121-134.
9. Franks, F.; Kay, R. L.; Dadok, J. *J. Chem. Soc., Faraday Trans. 1*, 1988, **84**, 2595-2602.
10. Allinger, N. L. *J. Am. Chem. Soc.*, 1977, **99**, 8127-8134.
11. Goldsmith, D. J.; Bowen, J. P.; Qamhiyeh, E.; Still, W. C. *J. Org. Chem.*, 1987, **52**, 951-953.
12. Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.*, 1989, **111**, 4379-4386.
13. Challis, B. C.; Challis, J. A. in *Comprehensive Organic Chemistry*; Sutherland, I. O., Ed.; Pergamon Press: Oxford, 1979; Vol. 2, Section 9.9, pp 986-994.
14. Luisi, P. L.; Ciardelli, F. in *Reactivity, Mechanism, and Structure in Polymer Chemistry*; Jenkins, A. D.; Ledwith, A., Eds.; Wiley - Interscience: New York, 1974; pp 483-485.
15. Haasnott, C. A. G.; De Leeuw, F. A. A. M.; Altona, C. *Tetrahedron*, 1980, **36**, 2783-2792.
16. Bovey, F. A. High Resolution NMR of Macromolecules; Academic Press: New York, 1972; pp 182-186.
17. Carey, F. A.; Sunberg, R. J. *Advanced Organic Chemistry*, 2nd Edition; Plenum Press: New York, 1984; Part A, p. 203. See also *Organic Solvents*, Vol.II of *Techniques of Organic Chemistry*, 3rd Edition; Riddick, J. A.; Bunger, W. B., Eds.; Wiley-Interscience, New York, 1970.
18. Cantrell, C. E.; Kiely, D. E.; Abruscato, G. T.; Riordan, J. M. *J. Org Chem.*, 1977, **42**, 3562-3567.

Chapter 9

Solvent Effects on Conformation of Carbohydrates

Molecular Dynamics Simulation of Sorbitol, Mannitol, and Methoxytetrahydropyran

J. Raul Grigera

Instituto de Fisica de Liquidos y Sistemas Biologicos (IFLYSIB), University of La Plata, c.c. 565, 1900 La Plata, Argentina

Molecular dynamics (MD) simulations show that the conformations of sorbitol and mannitol depend on the type of solvent. The predicted conformations agreed well with experiment, supporting the view that MD has a good predictive value for solutions of carbohydrates. Preliminary dynamics results for methoxy-tetrahydropyran (MTHP) show that the methoxy group moves more in water than in vacuum.

Molecular conformation is highly related to functional properties. Since the conformation of the crystalline solids can be precisely determined by diffraction methods, molecular modeling is most important for interpreting molecular structures in solution. This is, however, even more difficult for theoreticians. While carbohydrates dissolve in a variety of solvents, the important solvent for biological systems is water and this solvent deserves special emphasis.

Molecular dynamics (MD) simulation have been used for several years to get information on both equilibrium and dynamical conditions of various systems, including solutions of complex molecules. However, only a few carbohydrates have been studied (1-3).

Sorbitol and mannitol represent a pair of hexytols that differ only in the configuration of one hydroxy group at C2. This slight difference in their configurations gives both compounds differing physicochemical properties. For example, sorbitol is three and one half times more soluble than mannitol in water. Previous MD simulation of these hexytols (2) pointed out some characteristics that warrant further discussion. In particular their conformations depended on the solvent system.

In this work we discuss further the previous results from simulations of sorbitol and mannitol and compare them with new calculations and recent experimental data. We also present some preliminary data for methoxy-tetrahydropyran (MTHP) in vacuo and in water.

Methods

Computer Simulation. The GROMOS package (Biomos n.v. Groningen) was used for the MD simulations. Equations of motion were integrated using a leap-frog algorithm at a time interval of 2 fms. A thermal bath and a hydrostatic pressure system kept the pressure and temperature of the main system constant. This constant-temperature, constant-pressure procedure is part of the original GROMOS package. The SHAKE procedure held constant the fixed distances in the model.

All data reported are from runs made after equilibration. The precise time to attain equilibrium was not determined but equilibration was monitored both by the drift of total energy and the stability of the system density. The simulation boxes were cubes (for both hexytols) or a truncated octahedron (for MTHP). The VAX 11/750 of the IFLYSIB performed the main calculations. Graphic displays were produced with a personal computer and ALCHEMY software (Tripos Inc., St. Louis).

Force Field. Only atom-atom interactions were considered i.e. bond lengths and bond angles were taken as rigid. No explicit torsional potential was used, so the energy changes with change in torsion angle result only from the electrostatic and the van der Waals forces of the atoms involved. We adopt this position since we consider that a predefined torsional potential may bias the conformation. The parameters for the force field are GROMOS based, although the partial charges have important differences. For sorbitol and mannitol the parameters were from Ref. 2 while those for MTHP are described in Table V below. Our parameters correspond to a set devised to be used with the explicit addition of water if water is to be considered. While we cannot exclude the possibility that some information from solution has inadvertently been included in those parameters, we believe that the isolated molecule will be reproduced if solvent is not explicitly present.

Solvent. The water molecules conformed to the Simple Point Charge Extended model (SPC/E) (4), which is summarized in Table I. The non-polar´ solvents were taken as monoatomic non-charged atomic liquids with the same Lennard-Jones (6-12) parameters as oxygen in water, making an argon-like solvent.

Table I. Principal features of SPC/E Water Model

O-H bond	0.1 nm
H-O-H angle	109
A (oxygen centered)	2.6169×10^{-3} /(nm^6.kJ/mol)
B (oxygen centered)	2.6332×10^{-6} /(nm^{12}.kJ/mol)
Oxygen charge	-0.8476 e
Hydrogen charge	0.4238 e

Sorbitol and Mannitol

Initially, the hexytols have planar zig-zag conformations with C–O bond lengths of 0.143 nm, C–C bond lengths of 0.152 nm, CCC angle of 113° and CCO angle of 110°. All of these parameters are based on crystallographic information (5).

Results

Table II shows the average end-to-end distance over 20 ps for mannitol and sorbitol in vacuuo and in solution of an argon-like (L–J) solvent and SPC/E water. The average lengths all indicate sickle shapes, except for mannitol in water which is fully extended. This points to a specific solute-solvent interaction between mannitol and water, not just an unspecific solvent effect that is not present in solvent other than water. The model non-aqueous solvent is very artificial, but it should represent the main features of the class of non-polar, spherically symmetric solvents.

Table II. Average End-to-End Distances for Sorbitol and Mannitol (nm)

	Isolated Molecule	L–J Solvent	SPC/E Water
Mannitol	0.53	0.52	0.64
Sorbitol	0.55	0.56	0.55

Figure 1 shows the trajectories for the end-to-end distances for sorbitol and mannitol in water and mannitol the L–J solvent. The smaller fluctuations in end-to-end distance of mannitol in water might be interpreted in terms of lowered mobility of the molecule. However, Figure 2 shows that there are still large fluctuations in the torsion angles during the simulation. Therefore, the internal mobility is high, although compensating changes in torsion angles keep the final distance rather constant.

Table III shows the computed proton-proton scalar (J–J) coupling, along with experimental values. Coupling constants have been computed by using the Karplus formula in the form

$$J = -1.4 \cos \phi + 9.4 \cos^2 \phi + 1.6 \tag{1}$$

They agree qualitatively, confirming the predictive capability of the method.

Hydration. Some dynamic characteristics of hydration can be obtained from these simulations. According to Samoilov (7,8) we define the ˋhydration time ratioˊ R as the ratio between the average time that a water molecule spends near the solute (ts) and the average time that water molecules spend near to other water molecule (tw); i.e. R=ts/tw. The ˋhydration numberˊ here is defined as the number of water molecules that remain, on the average, around a solute molecule at a prescribed distance (0.425 nm in this case).
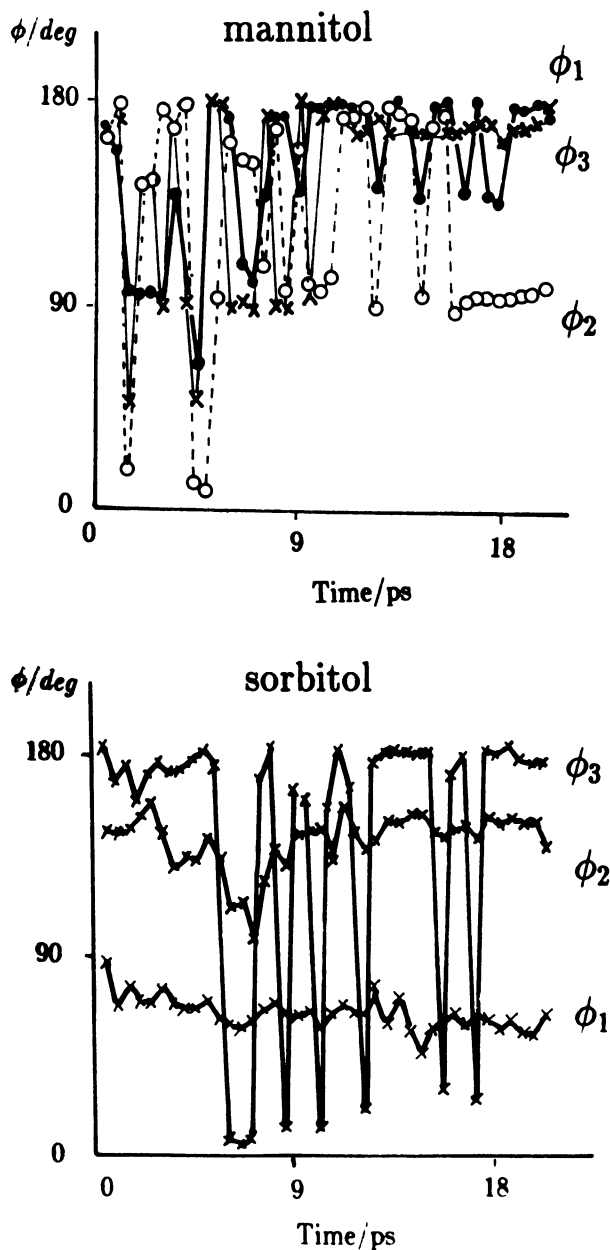
Figure 1. Trajectories of torsion angles ø1 (C1-C2-C3-C4); ø2 (C2-C3-C4-C5) and ø3 (C3-C4-C5-C6) for mannitol and sorbitol during 20 ps simulation.

Table III. NMR Proton Coupling Constants for Mannitol and
            Sorbitol in Water

J	Mannitol		Sorbitol		
	calc.	exp(a)	calc.	exp(b)	exp(c)
H 1´.2	8.543	6.426	5.078	6.55	6.55
H 1, 2	1.976	2.935	4.240	3.55	4.25
H 2, 3	11.411	8.990	4.685	5.90	6.0
H 3, 4	1.633	1.021	1.562	1.7	2.47
H 4, 5	5.382	–	9.855	8.25	7.70
H 5, 6	3.917	–	1.617	2.95	3.33
H 6, 6´	5.441	–	9.759	6.3	6.24

a) Franks et al.(9), b) Hawkes and Lewis (6),
c) D.B. Davies (quoted in Ref. 9).

Table IV shows the values for both polyols. The hydration
numbers are a consequence of molecular shape. Following
Samoilov, we classified both compunds as `negatively hydrated
since their hydration time ratios are less than one, with
sorbitol being more negatively hydrated.

Table IV. Hydration Numbers and Hydration Time
            Ratio for Sorbitol and Mannitol

	Hydration Number	Hydration Time Ratio(R)
Sorbitol	11.45	0.39
Mannitol	13.22	0.80

Although the concept of negative hydration was advanced by
Samoilov several years ago (7), the idea that an interacting group
might increase the mobility of surrounding water is not easily
accepted. When considering the individual atoms of the hexytol,
the water residence times vary from atom to atom. While the water
residence times for some atoms (e.g. O3 and O4 of mannitol) are
high, the average over the whole molecule is a ratio of less than
one. It might be suspected that a single solute, for which the
statistics are certainly poor, could have a local temperature
higher than the average, producing an artificially larger mobility
in the neighborhood of the solute. In our case we have eliminated
that possibility by using a separate temperature scaling for
solute and solvent.
    Recent 620.6 MHz nmr results on sorbitol and mannitol (9)
confirm that sorbitol rotates more freely in water than mannitol.
This suggests that there is less solute-solvent interaction in
sorbitol. Calorimetric results (10) predict that sorbitol and
mannitol should have hydration behavior similar to that described
above. Those workers, however, referred to `structure breaking´
properties, even though no structural data was obtained.
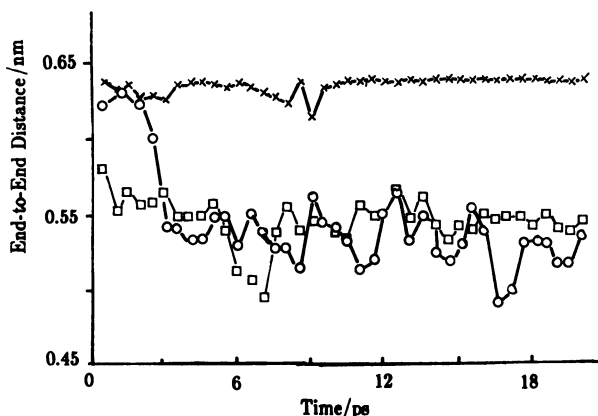
Figure 2.   Time evolution of the end-to-end distance of:(□)sorbitol
in water; (✗)mannitol in water; (○) mannitol in a LJ solvent.

MTHP

Methoxytetrahydropyran (MTHP) (Figure 3) has received considerable
attention as a simple substitute for the glycosides in hydration
studies (11-13). In our initial studies of the axial anomer, we
kept the ring rigid, as well as the bond lengths of the methoxy
group. In order to further reduce the time required for the
calculations, we used "united atoms" for CH, CH2 and CH3. Partial
charges for the united atoms were the sum of the individual
components given by Mardsen et al. (14). This lowers the dipole
moment of the model composed of united atoms, compared to the
experimental value. (*See* **Table V.**)

Table V. Interaction Parameters for MTHP

United atom name		group	Q/e	6 1/2 [A/(Kcal.nm /mol]]	12 1/2 [B/(Kcal.nm /mol]]
C1	(CG)	CH	0.207	228.98 x 10 -3	17.2265 x 10 -3
C2	(CR)	CH2	0.000	193.98 x 10 -3	12.0887 x 10 -3
C3	(C✶)	CH2	0.000	193.98 x 10 -3	12.0887 x 10 -3
C4	(C✶)	CH2	0.000	193.98 x 10 -3	12.0887 x 10 -3
C5	(CS)	CH2	0.066	193.98 x 10 -3	12.0887 x 10 -3
C6	(CM)	CH3	0.141	193.22 x 10 -3	10.4000 x 10 -3
O1	(OG)	O	-0.282	96.72 x 10 -3	1.7514 x 10 -3
O5	(OS)	O	-0.132	96.72 x 10 -3	2.2880 x 10 -3

The LJ parameters for the X-Y interaction are obtained by the
product of the parameter of the table for each atom.
     Again, no explicit torsional potentials were used.   Figure 4
shows the potential energy arising only from the atom-atom
interactions for charges in the torsion angle O5-C1-O1-C6 in the
absence of solvent.

Structure.   We have simulated MTHP in isolation and in an
infinitely diluted aqueous solution (56 water molecules in a
truncated octahedron).   The average position value of ∅ (O1-C1-
O5-C6) is 85° in isolation and 75° in water. Previous calculations

using energy minimization techniques (15) found several allowed conformers for rigid, axial MTHP.

Let us consider the results of Mardsen et al. (14) as representative. The least energetic conformer has ø of 60°, and the next two ø of 120° and 180°. In the static description of the system, as obtained by energy minimization, the relative populations of the three conformers are determined after consideration of their relative energies and the height of the barriers. This may give a clearer idea of the average conformation. In our dynamics simulation, we have a time average that includes excursions to the different conformers. If both methods are reliable, our average ø value should be closer to the least energetic conformer from the static study than to any other.

Our `solution´ value is closer to the 60° mininimum from the statics study. Forcefields such as used by Mardsen et al. are `solution equivalent´ (i.e. contain information on the interactions in aqueous solution). Therefore, their results are not for a truly isolated molecule, but might be expected to be equivalent to our solution model. To allow determinations of conformation in vacuum and other solvents, water information should not appear in the basic potentials. The presence of water information in force fields is a common problem.

Dynamics. The mobility of the methoxy segment differs for the isolated and solution states. Figure 5 shows the trajectories of the ø angle in both, with differing average values and ranges of fluctuation. While models of sorbitol and mannitol showed decreased mobility in water compared to vacuum, the methoxy group of MTHP is more mobile in water than in vacuum. Since we used only one solvent, we cannot distinguish between unspecific solvent effects or water-dependent properties.

Hydration. Since we have a detailed dynamics study with explicit water molecules we can describe the hydration of MTHP. Using the definitions developed above, we have a hydration number of 13.6 and a hydration time ratio of 1.42, if the cut-off radius is 0.425 nm, O5-C5 and O1-C6 have the highest residence times and hydration numbers. The reason that carbon atoms are apparently favored for hydration is that we check the proximity of water molecules by the distance to the water oxygen. Some hydrogen bonds to O1 and O5, for example, give very close proximity of water oxygen to the neighboring carbons atoms. Thus, the criterion for hydration of the atoms is met but there may not actually be any strong interaction between the solvent and the carbon atoms.

The hydration values for MTHP are different than those for sorbitol and mannitol. This is not surprising since most molecular properties are quite different. This result depends on the model since all were studied with the same simulation procedure.

Figure 6 shows a molecule of MTHP and some water molecules around it. This picture is a snapshot; not an average. This picture gives some gross features of the hydration but from it alone we cannot assign well-defined positions and orientations of the water molecules. Even so, the hydration structure compares, at least qualitatively, well with the one proposed by Tvaroska and Kozar (15). For easy comparison we have used the same labels as
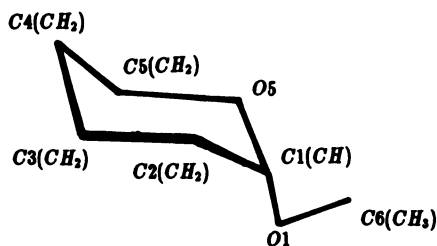
Figure 3. Molecular structure and labeling of MTHP.
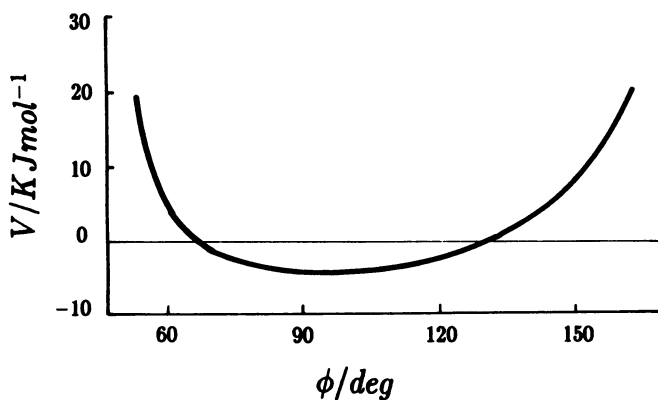
Figure 4. Potential of the torsional angle ϕ in MTHP produced by the atom-atom interaction.
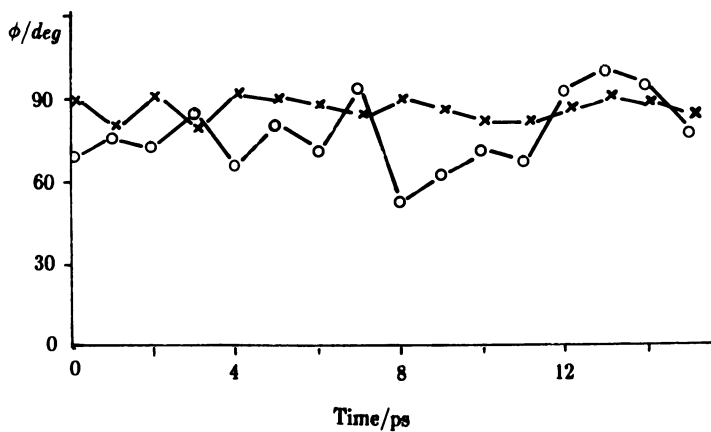
Figure 5. Trajectories of the angle ϕ during simulation of MTHP in isolation (✗) and in aqueous solution(○).
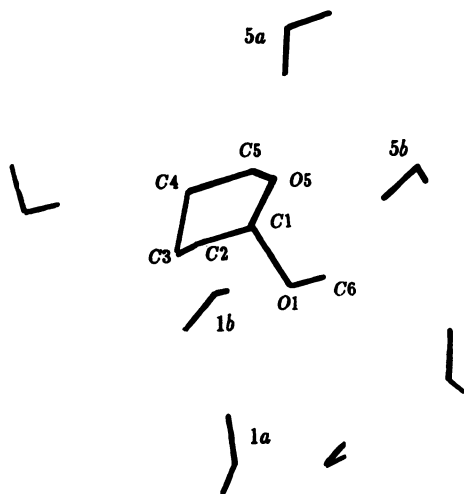
Figure 6. A MTHP molecule and the nearest water molecules. The picture corresponds to a single MD configuration and not to an average configuration. Water molecules are labeled as in Figure 2 of Ref. 15.

in Ref. 15 for the four water molecules shown there. The results for MTHP are still preliminary, and several aspects, such as molecular flexibility and other solvents, must be considered before the study is finished. These simulations are underway in our laboratory.

Conclusions

These examples of simulations of the molecular dynamics of carbohydrates show the possibility of predicting their behavior in different solvents. Experimental work has confirmed these findings. While theoretical prediction is becoming more reliable, it is only qualitative and we must consider the theoretical results within the framework of the actual capability of the methods. Current minicomputers allow simulation of large system. Polysaccharides, for instance, are being studied by this technique. However, the description of carbohydrate solutions is still poor, and simple systems can help in the understanding of the problems.

Acknowledgments

Literature Cited

1.   Brady J.W. J. Am. Chem. Soc. 1986, 108, 8153.
2.   Grigera J.R. J.Chem.Soc. Faraday 1 1988, 148, 2603.
3.   Kohler J. Disseration. Berlin 1987. Kohler J.; Saenger W; van
     Gunsteren W.F. Eur.Biophys.J. 1988, 16, 153.   J. Biomol.
     Struct. Dyn. 1988, 6, 181.
4.   Berendsen H.J.C.; Grigera J.R.; Straatsma T. J. Phys. Chem.
     1987, 91, 6269.
5.   Jeffrey G.A.; Kim H.S. Carbohydr. Res. 1970, 14, 207.
6.   Hawkes G.E.; Lewis D. J. Chem. Soc. Perkin Trans. II, 1984,
     2073.
7.   Samoilov O. Ya. Disc. Faraday Soc. 1957, 24, 141.
8.   Samoilov O. Ya. Structure of Aqueous Electrolyte Solutions
     and Hydration of Ions. Consultants Bureau, New York, 1965.
9.   Franks F.; Dadok J. ; Kay R. L. unpublished.
10. Wilson D. R.; Wen-Yang W. J. Phys. Chem. 1976, 80, 413.
11. Kozar T.; Tvaroska I. Theor. Chim. Acta 1979, 53, 9.
12. Tvaroska I.; Kozar T. J. Am. Chem. Soc. 1980, 102, 6929.
13. Tvaroska I. Carbohydr. Res. 1984, 125, 155.
14. Mardsen A.; Robson B. ; Thompson J.S. J. Chem. Soc. Faraday 1
     1988, 84, 2519.
15. Tvaroska I.; Kozar T. Internatl. J. Quantum Chem. 1983, 23, 765.

Chapter 10

Oligosaccharides in Solution

Conformational Analysis by NMR Spectroscopy and Calculation

Igor Tvaroška[1], Tibor Kožár[2], and Miloš Hricovíni[1]

[1]Institute of Chemistry, Center of Chemical Research, Slovak Academy of Sciences, 842 38 Bratislava, Czechoslovakia
[2]Institute of Experimental Physics, Slovak Academy of Sciences, 043 53 Kosice, Czechoslovakia

The ^{13}C chemical shifts of oligosaccharides depend on solvent and temperature, showing that oligosaccharides are flexible in solution. The three-bond, proton-carbon coupling constants of methyl xylobioside also indicate that the molecule is flexible. In order to interpret the conformational information from flexible oligosaccharides, a new method for computerized conformational analysis was developed in which the energy is minimized with respect to internal coordinates (bond lengths, bond angles and torsion angles). The favored orientation of pendant groups is partially determined by an algorithm based on a random walk technique. The use of this new program, RAMM, is demonstrated with a study of methyl glucobioside.

A complete understanding of the role of carbohydrates in biological systems requires knowledge of the distribution at equilibrium of the various conformers in aqueous solution. The conformational behavior of carbohydrates in solution can be examined from different vantage points (1), but the most relevant approach is, no doubt, study of dilute solutions themselves. At present, high resolution NMR spectroscopy is the primary tool for determination of three-dimensional structure of oligosaccharides in solution. Optical rotation is also very sensitive to conformation (2) and there is a new, semi-empirical theory of optical rotation of oligosaccharides (3).

Conformational analysis of oligosaccharides in solution by NMR is mainly based on hydrogen-hydrogen distance information obtained from the nuclear Overhauser effect (n.o.e.) (4). Other independent NMR parameters that provide conformational information such as heteronuclear coupling constants have been used in the past (5) and have become more accessible recently (6-8).

The internal flexibility of oligosaccharides is a major obstacle to interpretation of experimental data. To deduce three-dimensional structure, one must, therefore, be able to correctly model internal flexibility. Various methods and results for conformational energy calculations for oligosaccharides have recently been reviewed (9-13). Therefore, no attempt will be made here to describe such efforts to calculate conformational energy surfaces.

In the first part of this paper, we discuss several aspects of the internal flexibility of oligosaccharides. Then, a new methodology for conformational analysis of carbohydrates is described that attempts to solve the problem of the effect of pendant group flexibility on the location of minima on the conformational energy surfaces. This methodology, embodied in the program, RAMM, is applied to a conformational analysis of a 1->2 linked dimer of methylated glucose.   Then we discuss the applicability of three-bond coupling constants to conformational studies and show that one-bond, proton-carbon coupling constants may also be useful. Finally, we will measure ^{13}C chemical shifts and $^3J_{CH}$ coupling constants across the glycosidic linkage to demonstrate the flexibility of methyl xylobioside in solution. In that analysis, we also discuss the n.o.e. values that were observed experimentally and predicted with modeling.

Flexibility of Oligosaccharides in Solution

In a study of flexibility of oligosaccharides, there are four major questions:

1. What are the accessible molecular conformations for the molecule in question?
2. What are the differences in geometry among these molecular conformations?
3. What are the differences in energy among these conformers?
4. How does the conformational energy surface vary in different environments?

Different experimental techniques often play complementary roles in the elucidation of these questions.

Molecular conformations can be entirely defined in terms of three different molecular parameters: bond lengths, bond angles and torsion angles. Variations in the molecular geometry are then simply defined as changes in these parameters, which can be generally divided into three groups:

1. Conformations of individual pyranoid rings.
2. Relative orientations of the monosaccharide residues.
3. Relative orientations of pendant groups with respect to the monosaccharide residues.

Owing to the complexity of the internal motion of oligo-saccharides, one of the major computational approaches in use today calculates only the energies arising from changes in the torsion angles that define the relative orientations of rigid monosaccharide units. Therefore, the effects of variation in the conformations of the individual residues and the changes in relative orientations of the pendant groups are ignored. However, surveys of a large number of known crystal structures (14-19), supported by theoretical calculations, establish that there are small but fairly important differences in the geometries of pyranoid rings. Some geometrical differences appear to result from crystal-packing effects and are fairly random. Other differences in residue geometry, specifically in acetal segments, vary systematically when the conformation about the glycosidic linkage changes. A detailed discussion of this phenomenon, the exo-anomeric effect, along with other anomeric effects, can be found in Ref. 19. Here, we mention two examples that show the internal flexibility of oligosaccharides.

The first example is based on a survey of numerous crystal structures that contain different linear maltodextrins ([20-26]). Their interglycosidic linkages adopt several conformations, with the torsion angles ϕ and ψ having values between 72-121.7° and 107.7-155.0°, respectively, for ranges of about 50° each. In the second example, there is a striking interdependence between the C5-O5-C1 and O5-C1-O1 bond angles and the orientation of the anomeric bond, shown in Figure 1 for 2-methoxytetrahydropyran. The C5-O5-C1 bond angle displays the maximum variation, ranging between 105.5 and 115.5° for the equatorial form, and 101.5 to 108.5° for the axial form.

Because of these systematic changes in internal residue geometry with changes in the linkage torsion angles, coupled with a large observed range of torsion angles, a more appropriate approach was suggested ([10,27]) for conformational analyses of disaccharides. In that approach, the energy is minimized with respect to all internal coordinates at each increment of conformational change. Recently, such calculations have been carried out ([28-31]). However, uncertainties in the optimal orientations of pendant hydroxyl and hydroxymethyl groups remain. In optimized structures, the pendant groups usually remain in the local energy wells nearest to their starting orientations. Thus, starting orientations may influence the results. This is documented for maltose ([10,28]) where the orientation of the hydroxymethyl groups strongly influence the character of the ϕ, ψ energy maps. To avoid the influence of starting orientations of pendant groups on the energy maps, we have developed a new procedure for conformational analyses.

A New Procedure for Theoretical Conformational Analysis of Oligosaccharides

Our procedure depends on a new computer program, RAMM (RAndom Molecular Mechanics), which is applicable to any kind of biomolecule. It is described in detail elsewhere (Kožár, T.; Petrak, F.; Galova, Z.; Tvaroška, I. Carbohydr. Res., in Press). Only the basic characteristics of RAMM and its application to conformational analysis of disaccharides are discussed here, concentrating on the effect of the orientations of pendant groups on the energy values at the various ϕ and ψ torsion angles.

The orientation of the pendant groups in a disaccharide composed of two hexapyranose residues can be described by 10 torsion angles. To simplify the problem for this demonstration, we assume that all bond lengths, bond angles and the other torsion angles are fixed. With only three staggered minima for each of the pendant groups, the number of possible conformations is still 3^{10} (59,049) for each ϕ and ψ. It is therefore almost impossible to analyze all possible conformations, and a major objective of our methodology was to surmount this difficulty.

Program RAMM. The input starting molecular geometry is described in terms of internal coordinates (bond lengths, angles and torsion angles) instead of cartesian coordinates for individual atoms. We represent the molecule as a branched structure because it facilitates conformational analyses in terms of the most interesting parameters. The current version of the program is dimensioned for simultaneous variation of ten torsion angles by a grid method. Thus, the conformational analysis in terms of ϕ and ψ can be carried out for hexasaccharides. Rather than explicitly testing each of the

staggered pendant group conformations (3^{10} for a disaccharide, 3^{26} for a hexasaccharide) RAMM can, as an option, apply a random walk in hopes of finding the energetically most favorable combinations of the pendant group orientations for each φ and ψ combination. Any or all of the pendant groups can be varied with the random walk option, with each additional group increasing the required computer time.

The minimization of energy with respect to all (or a selected number of) parameters may be applied either to each randomly generated geometry or to the geometry with lowest energy found after a long random walk. In the latter case, perhaps 1000 steps would be taken at each φ and ψ combination. At each step, the random structure generator would create a new combination of pendant group orientations and the energy would be calculated without optimization. After the 1000 steps, the structure with the lowest energy would be optimized. This procedure does not guarantee that the structure at each φ and ψ combination has the lowest possible energy, but it certainly can avoid problematic conformations and will usually have an energy close to the lowest possible energy.

Since the energy calculation is the most time-consuming step in our procedure, empirical molecular mechanics is the only reasonable choice at present for energy analysis. Currently, the MM2CARB (10) molecular mechanics force field is used. However, the structure of the program code allows this force field to be replaced by any other, or even by quantum chemical methods for estimation of the molecular energy. An optional calculation of the effect of solvents on conformational energies, based on the continuum model (32), is included in the program.

The performance of RAMM and the influence of pendant groups on the conformational energy is demonstrated with a study of the 1->2 linked disaccharide, methyl 2-O-(methyl-β-D-glucopyranosyl-β-D-glucopyranoside (methyl glucobioside, 1). The numbering of the atoms, shown in Figure 2, proceeds from the non-reducing end (primed) to the reducing end (unprimed). The carbon atoms of the methyl groups have been labeled CM1 and CM2', respectively.

The geometrical parameters of 1 were based on the standard structures of hexopyranoses (33). The glycosidic linkage is described by the two torsion angles

φ = C2'-C1'-O2-C2 and ψ = C1'-O2-C2-C1.

Sections through the two-dimensional conformational energy map for 1 are shown in Figures 3 and 4 for fixed, random and relaxed orientations of pendant groups. The curves determined by the random walk procedure resulted from 1000 iterations where the orientation was allowed to have torsion angles of n * 20°, with n having values between 0 and 17. The energy displayed in the random walk curve for each conformation was the lowest obtained during the 1000 iterations.

Figure 3 shows the dependence of the conformational energy on φ when ψ = 120°. The energy curves for rigid (squares) and random walk (stars) geometries are very similar except that the random walk geometries have lower absolute energy values. The third curve (rhombuses) shows the energies of structures from the random walk procedure that were subsequently optimized with respect to the pendant group orientations (the residues remained rigid). This curve has still lower energies at all points, and the barrier at 80° actually became a local minimum after optimization.

More dramatic differences among these three types of energy analyses are shown in Figure 4. This section of the two-dimensional
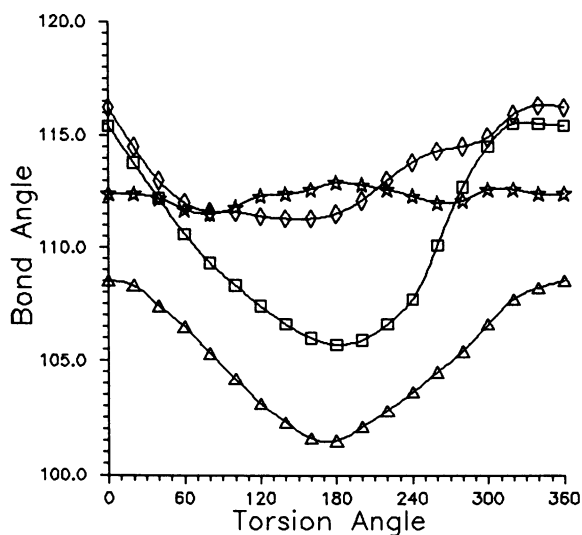
Figure 1. The variation of bond angles C5-O5-C1 and O5-C1-O1 with the torsion angle φ for 2-methoxytetrahydropyran. The curves with squares (C5-O5-C1) and triangles (O5-C1-O1) are for the axial form and the rhombuses (C5-O5-C1) and stars (O5-C1-O1) are for the equatorial form. These curves were calculated with PCILO, with full optimization of geometry at each increment.
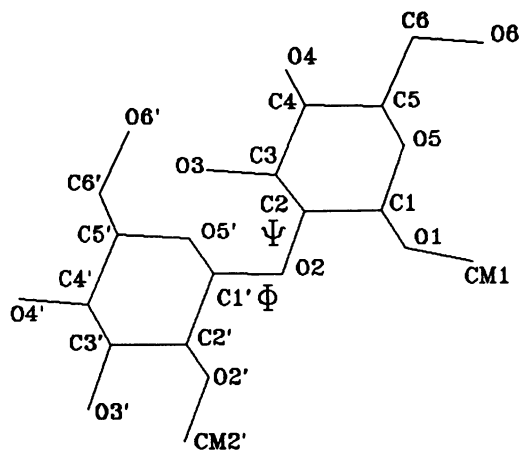
Figure 2. Numbering of atoms in **1**.
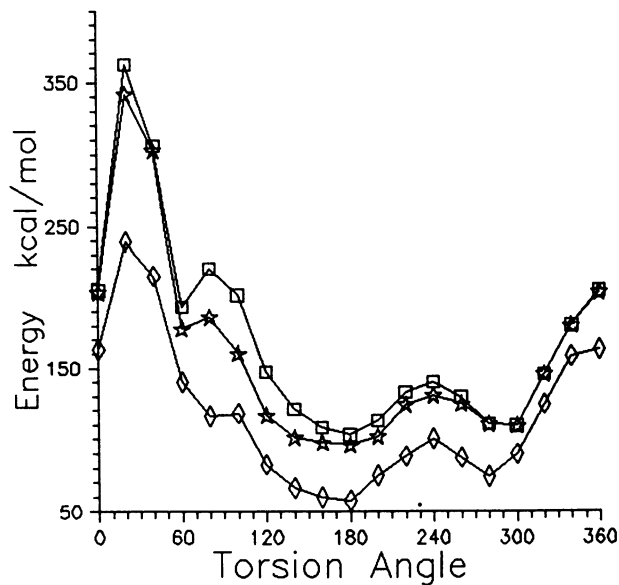
Figure 3. Potential energy arising from rotation in **1** about the anomeric C-O bond, calculated by the RAMM program with fixed (squares), random (stars) and relaxed (rhombuses) orientations of pendant groups.
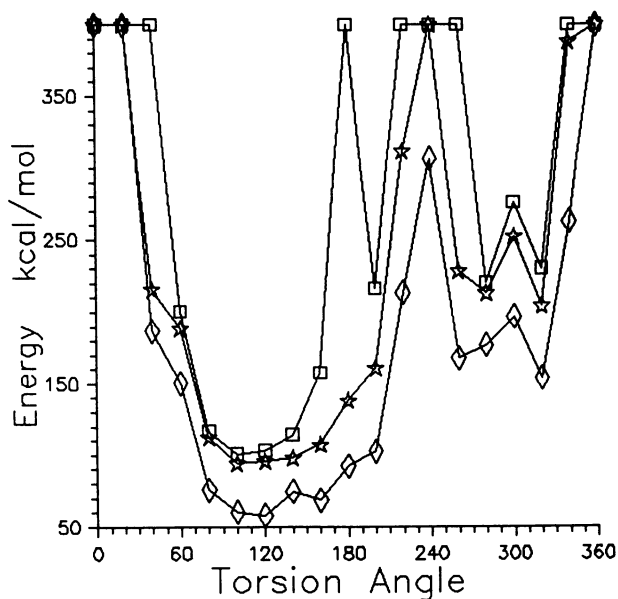
Figure 4. The same as in Figure 3, but for rotation about the aglycon C-O bond.

map has a fixed φ value of 180°. Changes in the ψ torsion angle are
very hindered in the range from 180 to 360°, with the highest barrier
at about 200°. In this region, the repulsive terms of van der Waals
interactions dominate because of the conflicts between pendant groups
on the two residues. Since most of these steric conflicts are
relieved through variations of the orientations of the pendant
groups, after the random walk treatment, the minimum at 120° spread
to 220°. Again, minimization, starting from the results of the
random walk orientations, lowered the energies even more.

The orientations of side groups for two conformers are shown in
Table 1, including the starting orientation, the best orientation
resulting from the random walk (RW), and the optimized (OPT)
orientation.

Table 1. Comparisons of Orientations for Pendant Groups in Xylobiose

			Conformation			
Torsion Angle	Start		φ = 80° ψ = 120°		φ = 180° ψ = 160°	
			RW	OPT	RW	OPT
C2-C1-O1-CM1	-167		-119	-167	-162	-167
C4-C5-C6-O6	59 (gg)		306	300 (tg)	49	43 (gg)
C2-C3-O3-HO3	-60		215	194	119	158
C3-C4-O4-HO4	84		280	150	228	160
C5-C6-O6-HO6	180		299	312	159	166
C1'-C2'-O2'-CM2'	-80		265	273	117	105
C4'-C5'-C6'-O6'	59 (gg)		120	-177 (gt)	61	65 (gg)
C2'-C3'-O3'-HO3'	-60		206	183	293	288
C3'C4'O4'-HO4'	64		234	154	254	42
C5'-C6'-O6'-HO6'	180		253	296	333	63

This table shows that there are substantial changes in the
orientations of the pendant groups to accompany the changes in
energy. Also, the optimal orientations are not generally similar for
the two linkage conformations.

These results indicate that the random walk procedure is an
efficient tool to improve the performance of the molecular mechanics
methods and to provide a better description of oligosaccharide
conformations. While Figures 3 and 4 illustrate the effects of
changes in the pendant group orientations, in normal use the entire
structure would be optimized after the random walk procedure had
determined low-energy positions for the pendant groups.

A consequence of the random walk method is that transitions among
different orientations of pendant groups occur. While a few of these
transitions occur passively during analyses based only on
optimizations of rigidly rotated structures, the random walk method
actively generates the alternate positions to find the preferred
alternate orientations. In the papers in this book by Tran and Brady
and by French, Tran and Pérez, the hydroxyl groups were assumed to
have either clockwise or counter-clockwise orientations, maximizing
the intra-residue hydrogen bonding. Our procedure is more general,
giving a way to study other types of pendant groups such as the
methyl groups of this example.

<u>Conformational Analysis of Oligosaccharides by NMR Spectroscopy</u>

Conformational analysis of oligosaccharides in solution by NMR spectroscopy is based on the study of chemical shifts, n.o.e.'s, and three-bond, proton-carbon coupling constants. Generally, the experimental NMR parameters $P_{experimental}$ (such as n.o.e.'s or coupling constants) of the N discrete conformations (i) are:

$$P_{experimental} = \sum_i x_i P_i \qquad (1)$$

where

$$x_i = \exp\,(-\Delta G_i/kT)/\sum_i \exp(-\Delta G_i/kT) \qquad (2)$$

In (2), G, k and T are the free energy, the Boltzmann constant, and the absolute temperature, respectively. Individual NMR parameters P_i usually depend in a nonlinear way on their geometrical parameter Gp_i. As examples, n.o.e. = $f(r) \sim r^{-6}$ where r denotes proton-proton distance, and $^3J_{CH} = f(\phi)\cos(\phi), \cos^2(\phi)$, where ϕ is still the torsion angle at the glycosidic linkage. In those cases, <P> cannot be obtained from the average <Gp> of the geometrical parameters Gp_i of the individual conformations involved in the averaging process (34)

$$<P> = \sum_i x_i P_i = \sum_i x_i f(Gp_i) \neq f(\sum_i x_i Gp_i) = f(Gp) \qquad (3)$$

This may be misleading in the determination of molecular conformation when a conformational equilibrium exists with rates that are fast on the NMR time scale. Short proton-proton distances are overemphasized when distances determined by n.o.e.'s are used as a tool for modeling the conformation in solution. On the other hand, the values of coupling constants change rapidly with torsion angles, so a slight change in the torsion angle (more precisely, in the distribution of values for the torsion angle) may change the coupling constant by a few units, especially for intermediate values.
Until now, the determination of three-dimensional structures of oligosaccharides in solution was based primarily on proton-proton distance information obtained from n.o.e. data. Here, we discuss the application of three-bond proton-carbon coupling constants.

<u>Three-Bond Proton-Carbon Coupling Constants</u>. Routine applications of the three-bond C-O-C-H proton-carbon coupling constants ($^3J_{CH}$) have been complicated by experimental difficulties involved in their measurement using classical ^1H-coupled ^{13}C NMR spectroscopy and a limited knowledge of the angular dependence of $^3J_{CH}$ for the C-O-C-H sequence of bonded atoms (5). However, recently developed semi-selective, two-dimensional NMR techniques (35,36) enable the measurement of long-range couplings with high accuracy in reasonable time and thus $^3J_{CH}$ are useful for conformational analysis. (Since the technique requires a selective pulse on the proton, and other protons must not be irradiated, the method does fail when the proton signals are very close.)
Vicinal proton-carbon coupling constants in a series of conformationally rigid monosaccharide derivatives showed a satisfactory dependence of $^3J_{CH}$ upon the torsion angles (7). The

geometries of the C-O-C-H segments in these molecules were known from
crystal structure analysis. Using modified two-dimensional, semi-
selective INEPT ($\underline{8}$) we obtained 14 values of $^3J_{CH}$ (precision 0.2 Hz)
for different torsion angles covering the range of 80 - 280°. Three
other coupling constants for dihedral angles from 0 to 60° were taken
from the literature ($\underline{37,38}$). All 17 $^3J_{CH}$ values were then used to
set up Karplus-type equations of the form A $\cos^2(\phi)$ + B $\cos(\phi)$ + C.
Solution of this equation yielded the final expression (4) for the
dependence of $^3J_{CH}$ on torsion angle ($\underline{7}$)

$$^3J_{CH} = 5.7 \cos^2(\phi) - 0.6 \cos(\phi) + 0.5 \qquad\qquad (4)$$

To confirm equation (4), we used the FPT (Finite Perturbation Theory)
INDO (Intermediate Neglect of Differential Overlap) method ($\underline{39}$) to
calculate the $^3J_{CH}$ for various values of torsion angles. A
comparison of the experimental and calculated values is plotted in
Figure 5.
      As another test of (4), calculated $^3J_{CH}$ couplings for 15
conformers of isomaltose were used (Tvaroška, I.; Pérez, S.; Imberty,
A. $\underline{Biopolymers}$, in press). Having five different proton-carbon
couplings for each of the conformers, 75 couplings were calculated
and those are plotted together with the calculated angular dependence
in Figure 6.
      Comparison of the calculated and experimental values shows that
the FPT-INDO method reproduces in a satisfactory way the coupling
constants measured on rigid model compounds ($\underline{7}$). The coefficients A,
B, and C in equation 4 obtained from the calculated $^3J_{CH}$ values are
A = 5.5 Hz, B = -0.5 Hz and C = 0.8 Hz. This same dependence on
torsion angle also justifies equation 4 and can be used for
interpretation of experimental interglycosidic coupling constants in
oligosaccharides.

One-bond Proton-Carbon Coupling Constants. In spite of the utility
of the observed three-bond, proton-carbon coupling constants, their
small magnitude, combined with difficulties of measurement, partially
limits their applicability. Therefore, any reliable information on
the magnitude of one-bond coupling constants as a function of
dihedral angles around glycosidic bonds should be of great value for
the determination of conformations of oligosaccharides. Recently, as
a first step in this direction, we calculated one-bond ($^1J_{CH}$)
coupling constants for the model compounds, dimethoxymethane and
2-methoxytetrahydropyran (Tvaroška, I. $\underline{Carbohydr.\ Res.}$, in Press).
The results from using the FPT-INDO method show that the effect of a
change in configuration at the anomeric carbon is to shift the one-
bond C-H coupling constant and that the one-bond coupling constant
has characteristic variations with the dihedral angle about the
carbon-oxygen bond. Distinctive features of the dependence of ($^1J_{CH}$)
on the glycosidic torsion angle ψ are shown in Figure 7, where the
values are plotted for both axial and equatorial forms of 2-methoxy-
tetrahydropyran. These results indicate that one-bond coupling
constants could complement other NMR parameters as probes of
oligosaccharide conformation in solution, provided that the form of
the angular dependence is known. Whether this dependence can be
determined for oligosaccharides is under investigation in our
laboratory ($\underline{40}$).

$$J(C-H) = 5.7 \cos^2(PHI) - 0.6 \cos(PHI) + 0.5$$
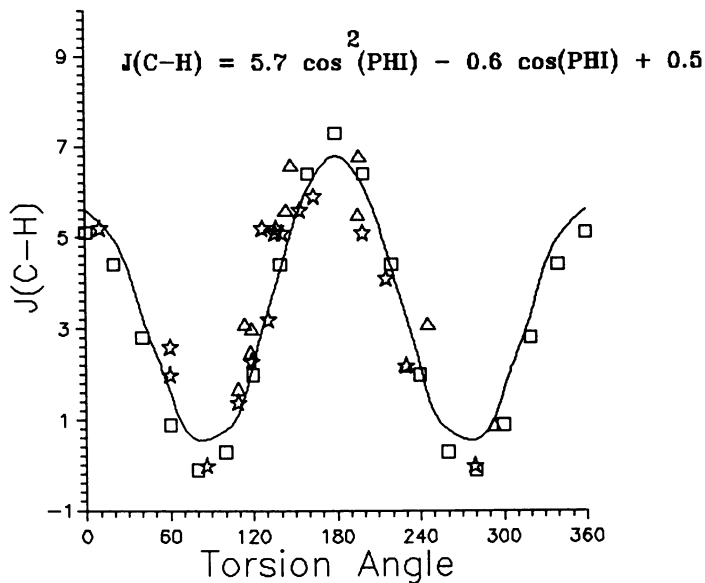
Figure 5.   Relationship between $^3J_{CH}$ and the C-O-C-H dihedral angle represented by equation (4) (full line).   Squares represent FPT-INDO calculations, stars represent experimental values from Ref. 5.

$$J(C-H) = 5.7 \cos^2(PHI) - 0.6 \cos(PHI) + 0.5$$
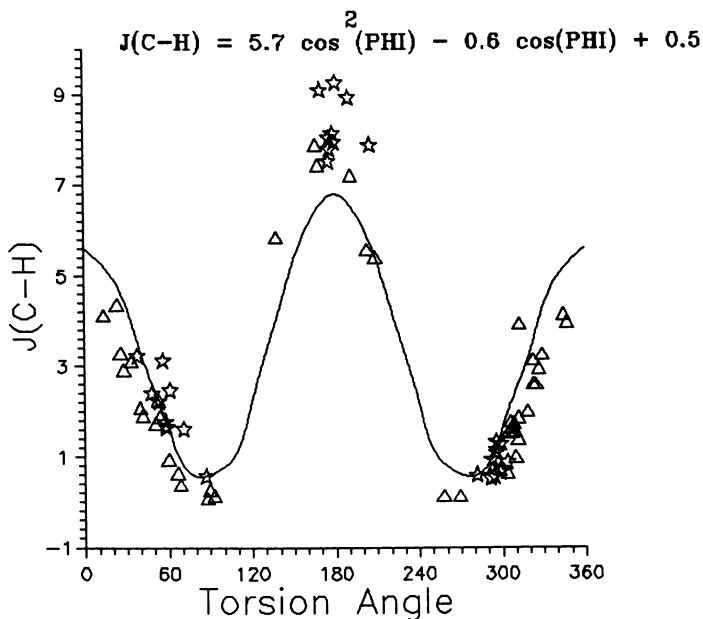
Figure 6.   The same as Figure 5, but stars and triangles are for isomaltose conformers taken from Ref. 39.

Conformational Flexibility Detected by NMR Chemical Shifts and
Coupling Constants

We briefly present here a conformational analysis of the
disaccharide, methyl (4-O-β-D-xylopyranosyl-β-D-xylopyranoside)
(methyl xylobioside, 2). The chemical shifts and three-bond C-H
coupling constants cannot be explained by the presence of a single
conformer, as described more fully elsewhere (Hricovíni, M.;
Tvaroška, I.; Hirsch, J. Carbohydr. Res., in press).
     The temperature dependencies of the chemical shift values for
both C1' and C4 were determined in four different solvents (water,
dimethyl sulfoxide, methanol and dioxane) and are shown in Figures 8
and 9. The resonance for C1' at 298° C varied from 101.6 ppm in D_2O
to 104.0 ppm in methanol. The resonance for C4 at the same
temperature varied from 75.3 ppm in dimethyl sulfoxide to 78.3 ppm in
methanol. The most pronounced temperature dependence is observed in
water and dioxane, where C1' and C4 signals varied from 101.4 ppm to
101.9 ppm (C1', water, 278-358° K) and from 75.7 ppm to 76.5 ppm (C4,
dioxane, 288-360° K), respectively. Thus, both temperature and
solvent dependence of ^{13}C shifts indicate different conformational
behavior of the molecule at various physico-chemico conditions. This
feature is manifested even more clearly by the dependencies of the
three-bond proton-carbon J^ϕ and J^ψ coupling constants (φ = H1'-C1'-
O4-C4 and ψ = H4-C4-O4-C1') which are plotted against temperature in
Figures 10 and 11.
     Values of J^ϕ in D_2O (Figure 3) changed from 4.8 Hz at 278° to 4.1
Hz at 358° K; the J^ψ values ranged from 5.6 Hz to 4.2 Hz. A similar
trend was observed in methanol. The differences between the changes
in J^ϕ and J^ψ are also considerable, with J^ψ having greater dependence
on temperature. Thus, these experiments suggest that, in accordance
with observations in the solid state, the rotations about the O1-C4
bond are less hindered than those about the C1'-O1' bond. Higher
values of both the J^ϕ and J^ψ couplings were obtained in dimethyl
sulfoxide at 298° K. These couplings were about 1 Hz larger than in
methanol or D_2O, however, unlike in the former solvents, temperature
dependence is less pronounced. Constant values of J^ϕ and J^ψ were
observed in dioxane at various temperatures. Since there is a strong
dependence of the chemical shift for C4 on temperature, some
structural variability must be present. However, the coupling
constants do not vary with temperature for any of the tested
concentrations. This must result from a special coincidence, wherein
the conformations present happen to give rise to coupling constants
that remain constant. The converse could not be true, wherein the
chemical shift could change while the conformation did not vary.
     We think that the constancy of the n.o.e. values can be explained
in this case by the averaging process during the NMR experiment (34).
Since n.L.e. values and coupling constants depend in different, non-
linear ways on proton-proton distances and torsion angles, the time-
averaged values of the geometrical parameters derived from n.o.e.
values can differ from parameters indicated by coupling constants.
In the present case, we think that the conformational dynamics of the
n.o.e. values and of coupling constants are different. Therefore,
the constant values of the n.o.e.'s in this case probably do not
indicate conformational rigidity of the molecule.
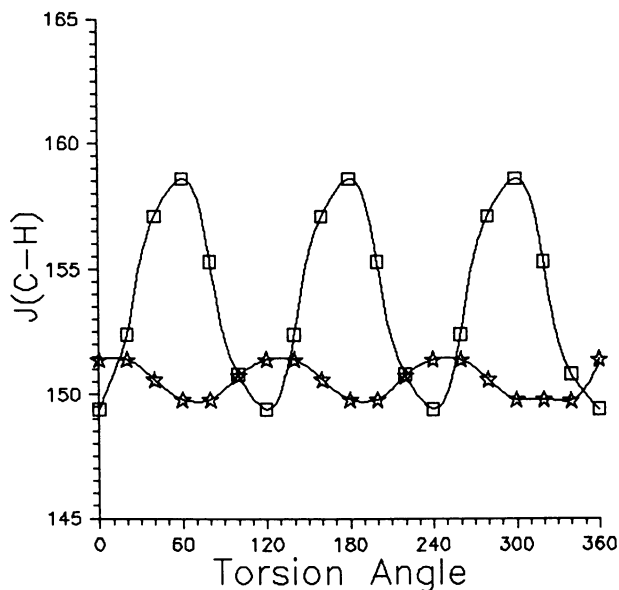
Figure 7. A plot of the calculated FPT-INDO results for the one-bond coupling constant $^1J_{CH}$ in 2-methoxytetrahydropyran as a function of torsion angle around the aglycon C-O bond with the methoxy group in axial (curve labeled with squares) and equatorial (stars) positions.
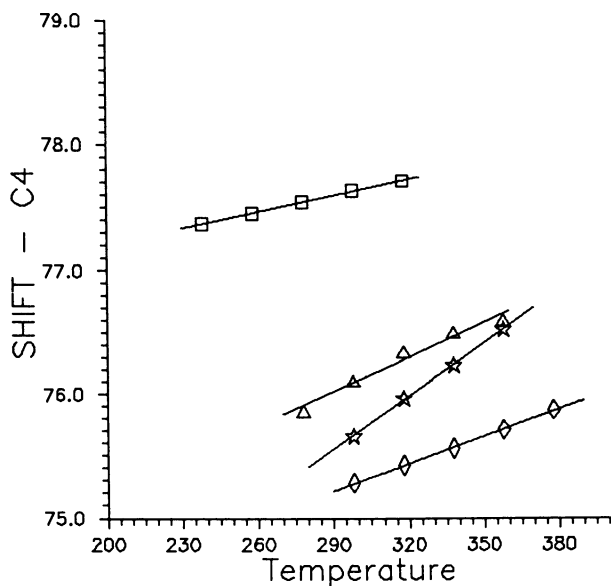
Figure 8. A plot of temperature dependence of the chemical shift for C4 of **2** for dioxane (curve labeled with stars), methanol (squares), dimethyl sulfoxide (rhombuses) and water (triangles).
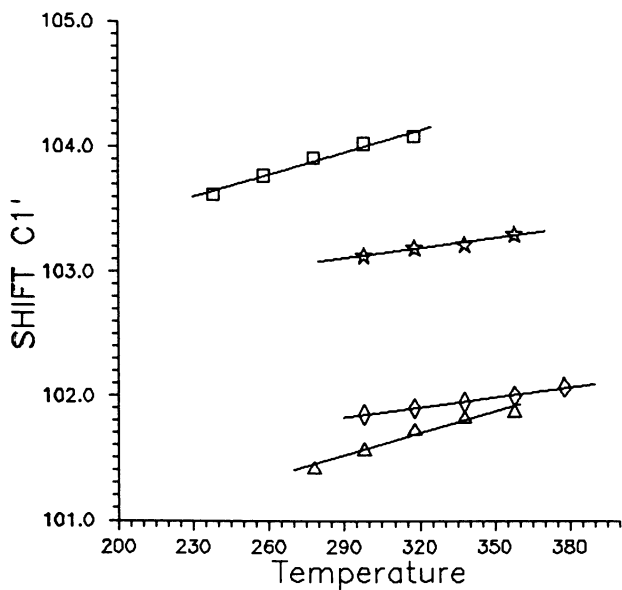
Figure 9. The same as Figure 8, but for the chemical shift of C1'.
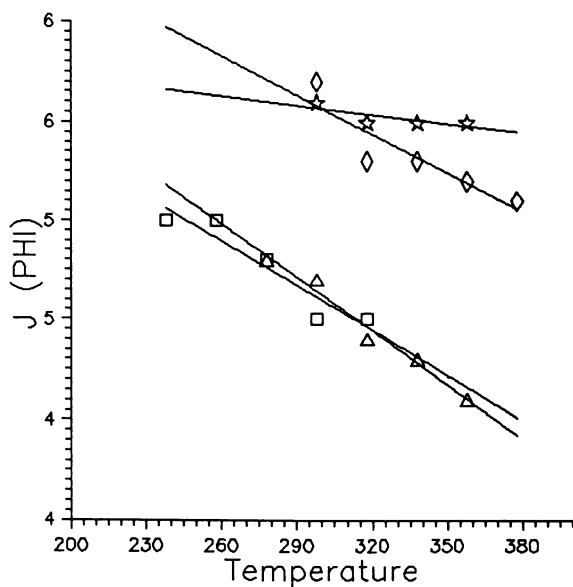
Figure 10. A plot of temperature dependence of the J^{ϕ} coupling constant of **2** for dioxane (curve labeled with stars), methanol (squares), dimethyl sulfoxide (rhombuses) and water (triangles).
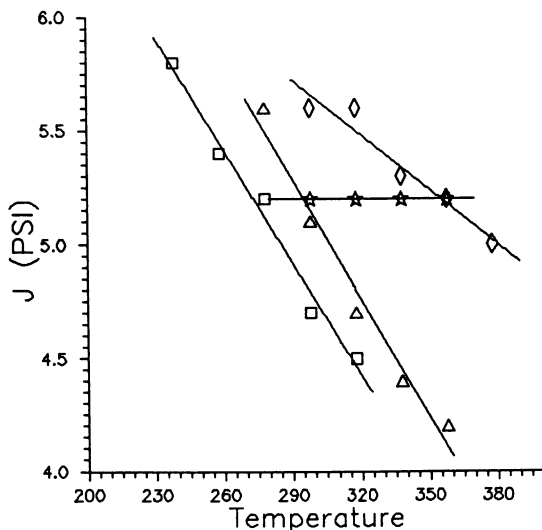
Figure 11. The same as in Figure 10, but for the J$^\psi$ coupling constant.

Conclusions

These results indicate that NMR techniques can be used effectively for conformational analysis of oligosaccharides in solution. Even if n.o.e. values are constant, molecular flexibility can be indicated by heteronuclear coupling constants. The new computational procedure for conformational analysis of oligosaccharides described here allows the structural refinement and geometrical optimization to be carried out conveniently and rapidly on a wide variety of carbohydrates. Our program may changed to include any other desired force field or other method to calculate energy. Preliminary results with this program stress the importance of the orientations of the pendant groups for locating minima on an energy surface.

Acknowledgment

Literature Cited

1. Franks, F. In Polysaccharides in Foods, Blanshard, J.M.V., Ed.; Butterworth, London,1979; pp. 33-49.
2. Rees, D.A.; Thom, D. J. Chem. Soc., Perkin Trans II 1977, 191-201.
3. Stevens, E.S.; Sathyanarayana, B.K. J. Am. Chem. Soc. 1989, 111, 4149-4154.
4. Carver, J.P.; Brisson, J.-R. In Biology of Carbohydrates Ginsburg, V.; Robbins, P.W.; Eds., John Wiley: New York, 1984; Vol.2.
5. Marshall, J.L. Carbon-Carbon and Carbon-Hydrogen NMR Couplings VCH: Deerfield Beach; 1984.
6. Cano, F.H.; Foces-foces, C. J. Org. Chem. 1987, 52, 3367.

7. Tvaroška, I.; Hricovíni M.; Petrakova E. <u>Carbohydr.Res</u>. 1989, <u>189</u>, 359-362.

8. Hricovíni M.; Tvaroška I.; Uhrin D.; Batta, G. <u>J. Carbohydr. Chem</u>. 1989, 389-394.

9. Brant, D.A. In <u>The Biochemistry of Plants</u>, Preiss, J., Ed.; Academic: New York, 1980; Vol.3, pp. 425-472.

10. Tvaroška, I.; Pérez S. <u>Carbohydr. Res</u>. 1986, <u>149</u>, 389-410.

11. Tvaroška, I. In <u>Theoretical Chemistry of Biological Systems</u>, Naray-Szabo, G., Ed.; Elsevier, Amsterdam, 1986; pp. 283-348.

12. Tvaroška, I. <u>Int. J. Quantum. Chem</u>. 1989, <u>XXXV</u>, 141-151.

13. Tvaroška, I. <u>Pure Appl. Chem</u>. 1989, <u>61</u>, 1201-1216.

14. Pensak, D.A.; French, A.D. <u>Carbohydr. Res</u>. 1980, <u>87</u>, 1-10.

15. Jeffrey, G.A.; French, A.D. In <u>Molecular Structure by Diffraction Methods</u>, Sutton, L.E.; Truter, M.R., Eds.; Chemical Society, Specialist Periodical Reports, London, 1978; Vol.6, pp. 183-221.

16. Tvaroška, I.; Kožár, T. <u>Chem. Zvesti</u> 1981, <u>35</u>, 425-440.

17. Fuchs, B.; Schleifer, L; Tartakovsky, E.; <u>Nouv. J.Chim</u>., 1984, <u>8</u>, 275-278.

18. Longchambon, F. PhD. Thesis, University of Paris-Nord, Bobigny, France, 1984.

19. Tvaroška, I.; Bleha, T. <u>Adv. Carbohydr. Chem. Biochem</u>. 1989, 47, 45-123.

20. Gress, M.E.; Jeffrey, G.A. <u>Acta Crystallogr</u>. 1977, <u>B33</u>, 2490-2495.

21. Chu, S.L.C.; Jeffrey, G.A., <u>Acta Crystallogr</u>. 1967, <u>23</u>, 1038-1049.

22. Takusagawa, F.; Jacobson, R.A. <u>Acta Crystallogr</u>. 1978, <u>B34</u>, 213-218.

23. Brisse, F.; Marchessault, R.H.; Pérez, S.; Zugenmaier, P. <u>J. Am. Chem. Soc</u>., 1982, <u>104</u>, 7470-7476.

24. Tanaka, I.; Tanaka, N.; Ashida, T.; Kakudo, M. <u>Acta Crystallogr</u>. 1976, <u>B32</u>, 155-160.

25. Goldsmith, E.; Sprang, S.; Fletterick, R.; <u>J. Mol.Biol</u>., 1982, <u>156</u>, 411-427.

26. Pangborn, W.; Lang, D.; Pérez, S. <u>Int. J. Biol. Macromol</u>. 1985, <u>7</u>, 363-369.

27. Tvaroška, I.; Kožár T. <u>Carbohydr. Res</u>., 1981, <u>90</u>, 173-185.

28. Ha, S.N.; Madsen, L.J.; Brady, J.W. <u>Biopolymers</u> 1988, <u>27</u>, 1927-1952.

29. Tran, V.; Buleon, A.; Imberty, A.; Pérez, S. <u>Biopolymers</u> 1989, <u>28</u>, 679-690.

30. French, A. D. <u>Carbohydr. Res</u>., 1989, <u>188</u>, 206-211.

31. French, A. D. <u>Biopolymers</u> 1988, <u>27</u>, 1519-1523.

32. Tvaroška, I.; Kožár, T. <u>J. Am. Chem. Soc</u>. 1980, <u>102</u>, 6929-6936.

33. Tvaroška, I.; Gajdoš, J. <u>Chem. Papers</u> 1987, <u>41</u>, 485-500.

34. Kessler, H.; Griesinger, C.; Lautz, J.; Muller, A.; van Gunsteren, W.F.; Berendsen, J.C. <u>J. Am. Chem. Soc</u>. 1988, <u>110</u>, 3393-3396.

35. Bax, A.; Freeman R. <u>J. Am. Chem. Soc</u>. 1982, <u>104</u>, 1099-1100.

36. Jippo, T.; Kamo, O.; Nagayama, K. <u>J. Magn. Resonance</u> 1986, <u>66</u>, 344-348.

37. Hamer, G. K.; Balza, F.; Cyr, N.; Perlin, A.S. <u>Can. J. Chem</u>. 1978, <u>56</u>, 3109-3116.

38. Thogersen, H., Ph. D. Thesis, The Technical University of Denmark, Lyngby, 1977.

39. Pople, J.A.; McIver, Jr., J.W.; Ostlund, N.S. <u>J. Chem. Phys</u>., 1968, <u>49</u>, 2960-2965.

40. Hricovíni, M. and Tvaroška, I., Vth European Symposium on Carbohydrates, Prague. Czechoslovakia, August 21-25, 1989.

41. Brisson, J.-R.; Carver, J.P. <u>Biochemistry</u> 1983, <u>22</u> 1362-1368.

RECEIVED March 9, 1990

Chapter 11

Optimized Potential Energy Functions in Conformational Analysis of Saccharides

Kjeld Rasmussen and Jesper Fabricius

Chemistry Department A, Technical University of Denmark, DK–2800 Lyngby, Denmark

A short presentation of the Consistent Force Field is given, with emphasis on parametrization and optimization of energy function parameters. For best possible calculation of structure, potential energy functions with parameter values optimized on both structural and other properties must be used. Results from optimization with the Consistent Force Field on alkanes and ethers are applied to glucose, gentiobiose, maltose and cellobiose. Comparison is made with earlier and with parallel work. The meaning and use of conformational maps is discussed shortly.

This paper presents a few examples of applications of the program package called the Consistent Force Field (CFF). The program has been extensively described in the literature (1-2), as has the strategy of its use (3-4), but a short overview may be pertinent here.

The CFF system

The concepts. All interatomic interactions are modeled with a set of mathematical functions which, when summed over all interactions, gives the potential energy of a molecule. The potential energy functions, the PEFs, contain adjustable parameters which, for a start, are taken from similar work or are merely guessed.

The way to start a CFF parametrization is: Select a set of PEFs, with associated parameters. Choose a set of molecules, closely related to the problem in hand (for carbohydrates: alkanes, cycloalkanes, ethers, alcohols); their structures should be determined and their vibrational spectra assigned to a reasonable precision. Put in their structures by specifying atomic coordinates; they need not be accurate.

The methods. As shown in Figure 1, the potential energy of each molecule is minimized, giving the equilibrium conformations pertaining to the chosen energy functions with the associated initial parameters. In practice, all energy gradients should be zero to a high precision. The result is the set of conformations at equilibrium, and one can now calculate the second order derivatives or force constants, individual values for each pair of coordinates, which will give the vibrational spectra in the form of normal frequencies and normal coordinates.

From these static and dynamic properties, thermodynamic functions and other properties may be calculated. All calculated values of structural, vibrational and other properties may then be compared with the corresponding observed values.

The reason for the claim to consistency is that measured values of observables such as bond lengths, angles, torsions, frequencies, dipole moments and, for crystals, unit cell dimensions, are put in. The program will compare them with the corresponding calculated values, and will optimize the energy function parameters so that, on the next run-through, a better reproduction of the measured values is obtained. When one is satisfied to some criterion that the model cannot do better, one has a consistent set of parameters for the chosen set of energy functions; see Figure 1. It is possible to optimize on molecular structure in the gaseous phase and in crystals, and on molecular vibrational frequencies and dipole moments, in one and the same calculation.

Developing one's own PEFs is much more time-consuming than applying them, and some problems arise. Here just two points are to be emphasized.

Molecular Structure. One is the question of which experimentally determined type of molecular structure to use, as many structure types are available in the literature.

They are derived from x-ray and neutron diffraction of crystals, and from electron diffraction and spectroscopic measurements with microwave, infrared and Raman techniques on the gaseous phase. For optimization of PEFs on small molecules, gas-phase structures are used. They are rather numerous, and they are all calculated
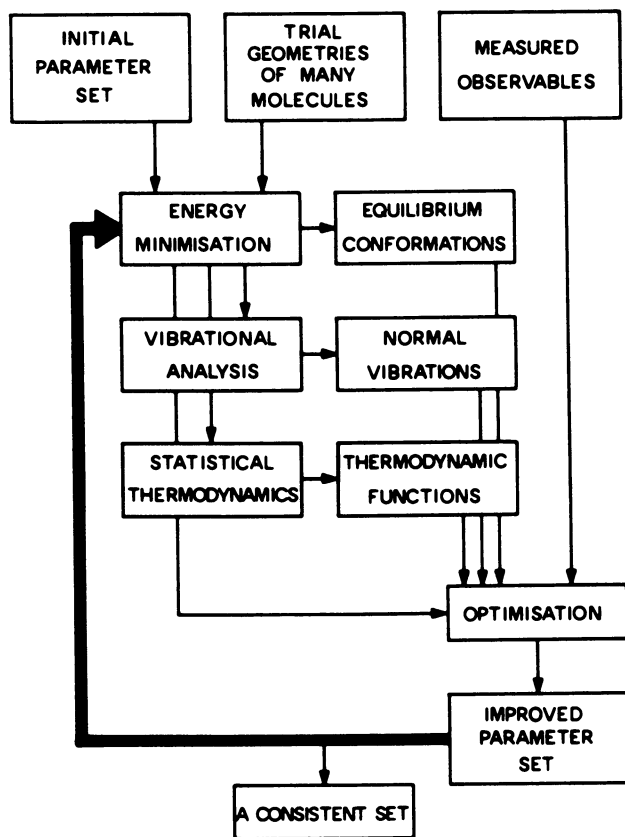
Figure 1. The CFF Cycle
(Reproduced with permission from Ref. 4.
Copyright 1989 Kluwer Academic Publishers.)

from molecular parameters fitted to reproduce experimental data. Thorough presentations are not easy to find (5-6); a short summary is given in a monograph on the Consistent Force Field (2).

In principle, the CFF should aim at reproducing equilibrium structures r_e, but those are known only for a few small compounds. The r_g structures are very often presented, but they are temperature-dependent, being the thermal average values of the inter-nuclear distances, and should therefore not be used. We use the r^0_α or the r_z structures which in principle are identical. They are derived from spectroscopic and electron diffraction measurements and represent the distances between average nuclear positions in the vibrational ground state at 0 K (5). They are temperature-independent, and they are fairly easy to calculate from the most frequently published structures r_a and r_g. According to structural chemists (K. Kuchitsu, personal communication) the r^0_α or r_z is the molecular structure which most closely resembles low-temperature neutron-diffraction results. Therefore it is well suited for deriving a set of parameters that should be reliable for larger molecules whose structures, if they are known, stem from low-temperature X-ray or preferably neutron diffraction.

Potential Energy Functions and Parameters. The second point is the importance of non-bonded interactions. The program was developed to optimize also on unit cell dimensions in addition to the usual conformational properties, because this gives the possibility of optimizing on properties that are very sensitive to non-bonded interactions.

By far the most difficult interactions to model are the non-bonded, because of near-cancellation of strongly distance-dependent forces of opposite signs. Only proper handling of non-bonded interactions will give sensible results in the calculation of structures of molecules as flexible as saccharides. Yet very few observables of small molecules depend strongly on non-bonded interactions (the -C-C- torsion in n-butane is an exception), wherefore optimization on crystals is needed as argued above.

The terms used in CFF are very simple; simpler than most other functional forms used by similar programs; see Figure 2. As we deal with the modeling of chemical systems, we divide the most important interatomic interactions into two parts: bonded and non-bonded.

The bonded interactions are almost always modeled with harmonic (parabolic) functions which practice is acceptable close to equilibrium. For non-bonded interactions, the van der Waals part is modeled with inverse power terms in the interatomic distances, 12, or occa-

sionally 9, for the overlap repulsion, 6 for the London attraction. In addition, electrostatic terms are essential when effects of polar groups are to be considered. In these Coulomb terms, atomic monopoles are used; this concept has hitherto proved to be acceptable. Atomic charges are usually taken from Mulliken population analysis of ab initio calculations with basis sets preferably larger than minimal. They are reproduced in CFF by a rather intricate algorithm from one charge parameter per atom type. In the energy calculations a dielectric constant is used. The built-in value is 2.0, but it can be changed in the input. The choice is purely pragmatic: if it is 1, simulating vacuum, the electrostatic energy dominates entirely; if it is 10, its effect can hardly be seen. Values of 2.0 to 3.5 are most practical. (The value 1 is used in calculations on purely ionic inorganic crystals.)

Because of the simple functions this model is too crude to be of practical use, and we must add a number of secondary terms; they depend explicitly on valence, torsional, and out-of-plane angles where appropriate. Figure 2 shows the terms relevant to the work reported here. The designations primary and secondary are conceptually significant: the secondary terms are necessary because the present formulations of the primary terms is not sufficently accurate. In cases where carboxyl, amido, imino and other groups occur, out-of-plane angles are usually included.

Please note the meaning of the word "parameter". In the CFF context, K_b is not a force constant of any bond in any molecule, and θ_o is not the equilibrium value of any valence angle. They are energy function parameters with units of force constant and angle. In the actual case, $kJmol^{-1}\text{\AA}^{-2}$ and rad.

Saccharides

The two main fields of application in the CFF group in Lyngby are saccharides and coordination compounds. Here we shall mention only the saccharide work. The first attempts to calculate saccharide equilibrium structures were made by use of two PEFs developed by trial-and-error, PEF300 (7-8) without and PEF400 (9-10) with charges. In spite of this, good results were obtained, both for structures of glucose and for the thermodynamic equilibrium between the anomers.

In the present work we introduce two PEFs containing parameters optimized on straight-chain and cyclic ethers, some of them containing anomeric carbon atoms. In these optimizations an anomeric carbon atom was given its own symbol and parameter attributes. The functions are named PEFAC1 and PEFAC2; the latter has Coulomb terms included. Trial-and-error parameters from PEF400

$$\text{primary terms} \begin{cases} V_b = \sum_{\text{bonds}} \tfrac{1}{2} K_b (b-b_o)^2 \\[2em] V_{n-b} = \sum_{\substack{\text{non-bonded}\\\text{interactions}}} \tfrac{1}{2} \left(\dfrac{A}{r^{12}} - \dfrac{B}{r^6} + \dfrac{e_i e_j}{r} \right) \end{cases} \text{two-body interatomic interactions}$$

$$\text{secondary terms} \begin{cases} V_\theta = \sum_{\text{angles}} \tfrac{1}{2} K_\theta (\theta-\theta_o)^2 \\[2em] V_\phi = \sum_{\substack{\text{single}\\\text{bonds}}} \tfrac{1}{2} K_\phi (1+\cos k\phi) \end{cases} \text{correction terms}$$

$$V_{total} = V_b + V_{n-b} + V_\theta + V_\phi$$

Figure 2. Potential Energy Functions
(Reproduced with permission from Ref. 4.
Copyright 1989 Kluwer Academic Publishers.)

for hydroxo groups were appended for this application.
Optimization on alcohols was not done, as available
experimental data are insufficient for our purpose. The
work on PEFAC1 and FEFAC2 is not yet published. Table I
shows the parameter values in PEFAC1 and PEFAC2.

Glucose. The improvement in the calculated structure
which is obtained by use of the new PEFs is indicated in
Table II which includes comparison with previous results
(8,11). The measured values pertain to neutron-diffrac-
tion data for α-glucose (12) and x-ray diffraction data
for β-glucose (13). The separate treatment of anomer
carbon implies that the largest deviation in bond length
is no longer found for the anomeric C-O.
    Two particular details are the oxygen ring angle
and the angle at the anomeric carbon. The comparatively
open angles are better reproduced than the more closed.
At present, no explanation can be given. On the other
hand, the equilibrium ratio of the anomers has changed
to 0.48:0.52 rather than the value of 0.36:0.64 which

Table I. Two partially Optimized PEFs.
Units are chosen so as to give energy
in kcal mol^{-1}

	PEFAC1		PEFAC2	
	K_b	b_o	K_b	b_o
O–H	1070.	0.955	1070.	0.955
C–C	563.077	1.5157	563.077	1.5134
K–C	563.077	1.4824	563.077	1.4853
C–H	670.000	1.0990	670.000	1.0866
K–H	670.000	1.0990	670.000	1.0866
O–C	863.000	1.4007	862.231	1.3990
O–K	863.000	1.3945	863.000	1.3948
	K_θ	θ_o	K_θ	θ_o
K–O–H	80.	1.80	80.	1.80
C–O–H	80.	1.80	80.	1.80
C–C–C	142.447	109.	142.447	109.
K–C–C	142.447	109.	142.447	109.
C–K–C	142.447	109.	142.447	109.
O–C–C	143.837	109.	143.838	109.
O–C–K	143.837	109.	143.838	109.
O–K–C	143.900	109.	143.900	109.
C–C–H	93.500	109.	93.500	109.
K–C–H	93.500	109.	93.500	109.
C–K–H	93.500	109.	93.500	109.
C–O–C	143.353	1.8418	143.336	1.8462
K–O–C	143.897	1.8816	143.897	1.8807
K–O–K	143.900	1.8463	143.900	1.8458
O–K–O	143.882	109.	143.882	109.
O–C–H	93.498	109.	92.477	109.
O–K–H	93.498	109.	92.477	109.
H–C–H	74.800	109.	74.800	109.
H–K–H	74.800	109.	74.800	109.
	K_ϕ	n	K_ϕ	n
H–C–C–H	1.2809	3.	1.2809	3.
H–C–K–H	1.2809	3.	1.2809	3.
H–O–C–H	2.7849	3.	2.7575	3.
H–O–K–H	1.6491	3.	1.6537	3.
	A	B	A	B
C––	559.123	18.865	559.123	18.856
K––	559.128	18.818	559.128	18.815
O––	292.392	12.599	292.394	12.580
H––	160.137	7.746	160.139	7.705
			e	
C.			0.00001	
K.			-0.002	
O.			-0.108	
H.			0.140	

agrees with the experimental value found in aqueous
solution. The rotamer ratio for the hydroxymethyl group
is still the same as calculated before: gt to gg = 0.78-
:0.22, whereas a compilation of crystal structures gave
a ratio of 0.40:0.60 (14). This discrepancy may be due
to the method: calculation on an isolated molecule can-
not account for intermolecular interactions in crystals.
     The inclusion of electrostatic terms in PEF400
gives an only marginally different situation, with no
differences within the precision of the data given in
Table II. Therefore only one column (for PEFAC2) is
listed. Somewhat better structural details were obtai-
ned for the monosaccharides by the use of optimized PEFs
such as PEFAC2, but the thermodynamic equilibria became
less well fitted.

Disaccharides. The most flexible disaccharide calcula-
ted before is gentiobiose (3,9). The conformation as
found in the crystal (15) was minimized in the new func-
tion PEFAC1 without electrostatic terms; the difference
in geometry is noticeable essentially in one torsion. If
charges are included, PEFAC2, almost the same picture
obtains.

Table II. Glucose in an Optimized
Potential Energy Function PEFAC2
and Comparisons with Non-Optimized
Functions (7-8, 10)

	α		β	
	calc	meas	calc	meas
C5O5C1	113.8	113.8	113.8	112.7
O5C1O1	111.4	111.6	109.0	107.0

	Deviations meas(12-13) − calc		
	PEF300	PEF400	PEFAC2
Bonds/Å			
max	−0.039	−0.034	0.029
rms	0.014	0.014	0.004
Angles/°			
max	−4.5	−4.8	4.0
rms	1.8	1.8	0.1
Torsions/° endocyclic			
max	4.5	5.3	3.0
rms	3.2	3.3	0.1
hybrid			
max	4.5	4.4	2.4
rms	3.2	2.8	0.8

Table III. Gentiobiose Conformations: Crystal Conformation Minimized in Four Potential Energy Functions

PEF	PEF300	PEF400	PEFAC1	PEFAC2	Crystal
ϕ	47.	85.	60.	63.	63.
ψ	-177.	-155.	-178.	-176.	-156.
ω	-178.	152.	173.	173.	-178.
C1O6C6	114.	115.	114.	114.	113.
H1'---H$_R$	2.28	2.66	2.42	2.43	2.40
H1'---H$_S$	3.50	3.56	3.07	3.10	3.12

Some conformational details are compared in Table III. The very open COC angle is well reproduced, and so are the three torsions along the glycosidic linkage. Two H---H distances are of special interest because they can be estimated by NMR techniques and can therefore give a clue to the solution conformation; they are reproduced quite well.

From a similar comparison with non-optimized functions, PEF300 (8) without and PEF400 (10) with charges, we see that the conformational details are not nearly so well reproduced. In the case of gentiobiose the optimized PEF has therefore made a real improvement.

First and foremost it is the optimization of the non-bonded interactions which has brought about the improvement. The consistency of the approach is emphasized by the fact that the subset of parameters for C and H were optimized on data for small alkanes, cycloalkanes and alkane crystals, and was used unchanged during optimization on ethers; alcohols were not yet included. These results prompted a reexamination of maltose (16) and cellobiose (17).

French has recently presented comparisons of rigid and relaxed conformational maps for cellobiose and maltose obtained with the MMP2(1985), which includes anomeric effects. The fully relaxed maps show interesting details.

The results for maltose and cellobiose are shown in Figures 3 and 4. They were obtained with the non-optimized parameter sets PEF300 without charges (16-17), MMP2 which uses dipole-dipole interactions instead of explicit charges (18-19), PEF400 with charges (3,11), and the optimized set PEFAC1 without charges (this work).

For maltose, Figure 3 shows that there are no significant differences between the geometric results found with the four rather different potential energy functions. The only discrepancy is the absence of minimum 1 in the map of French (18). Most crystal structure data fall within the valley joining the three upper points.
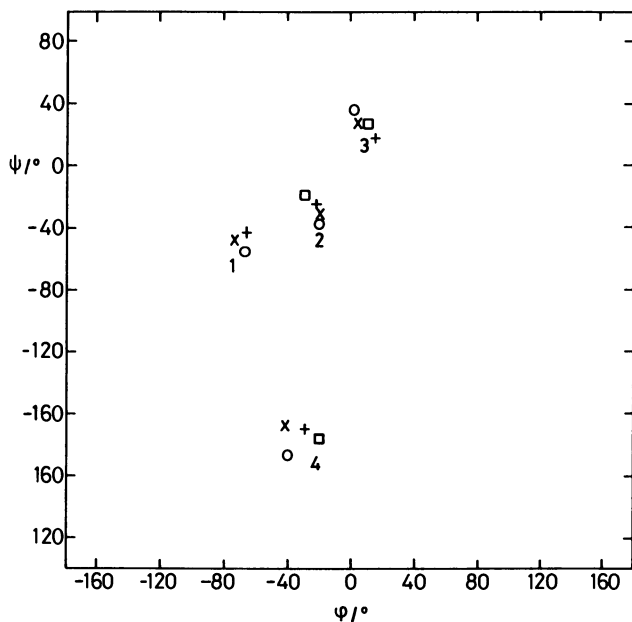
Figure 3. Conformational Map of Maltose.
+ PEF300, × PEF400, o PEFAC1, □ MMP2

This valley is investigated in some detail in the origi-
nal paper on maltose (10) where also references to ex-
perimental work can be found. Some details are given in
Table IV; it is noteworthy that in PEFAC1 one conformer
is dominant, which is in agreement with the results of
PEF400   (3,10).

Table IV.   Maltose Conformers in PEFAC1

Conformer	1	2	3	4
$\phi/°$	-67.2	-19.5	2.5	-40.0
$\psi/°$	-56.2	-35.9	37.1	173.3
C1'O4C4/°	117.5	116.3	118.1	118.6
H1'---H4/Å	3.149	2.366	2.322	3.642
$\Delta G/kJmol^{-1}$	0.000	6.336	16.939	13.714
n_i	0.924	0.072	0.001	0.003

For cellobiose, the situation is slightly diffe-
rent, as seen in Table V and Figure 4. The most obvious
difference is that only five minima are found with MMP2
and PEFAC1. In PEFAC1, essentially two conformers are
populated, and almost equally so; they span the diffrac-
tion results, as summarized by French (19).
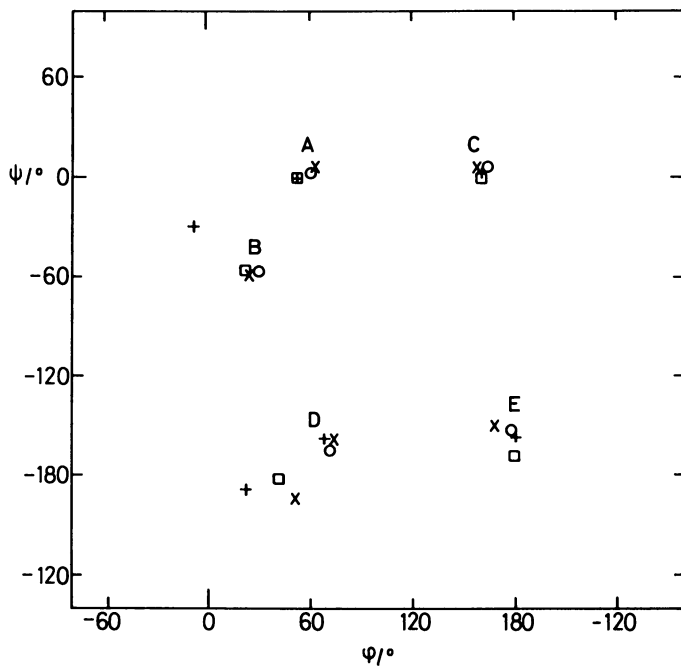
Figure 4. Conformational Map of Cellobiose.
+ PEF300, × PEF400, o PEFAC1, ▢ MMP2

Table V.   Cellobiose conformers in PEFAC1.
           Units as in Table IV

Conformer	1	2	3	4	5	6
ϕ	57.6	163.9	70.1	28.2	merges	177.5
ψ	4.6	4.4	-165.0	-57.8	into	-150.7
C1'O4C4	114.0	115.5	116.5	115.5	conf.	121.2
H1'---H4	2.451	3.543	3.575	2.380	3	3.924
ΔG	0.359	13.845	3.994	0.000		54.788
n_i	0.418	0.002	0.097	0.483		0.000

Conformational Maps and Surfaces. The conformational
maps in Figures 3 and 4 are very small sections of the
true conformational surfaces of 136 dimensions (3 coor-
dinates per atom plus one for the energy). The special
choice of ϕ and ψ as the coordinates of the section is
the traditional one, which is sensible because the gross
conformational features are described well by just those
two. One should just not forget that, on moving from
one point to another, many coordinates may change appre-
ciable, within a small energy interval. In consequence,
a point in the two-dimensional map represents an entire
"family" of points in 135--dimensional space.
     Only minima are shown, because they always inter-
ested us most, as they correspond to structures which in
principle can exist in solution and in crystals. As
argued above, a point is not unique; a change in, say, a
CCOH torsion might cause a slight change in ϕ or ψ or
both. Therefore it would be correct to say that a mini-
mum in conformational space is represented by a small
but unspecified area around a point in the conformatio-
nal map, which was earlier (7) termed a manifold. The-
refore, a difference in (ϕ,ψ) of (10,10) is really no
difference at all.
     Many people calculate various variants of "fully
relaxed" conformational surfaces. A fully relaxed sur-
face is just a set of points, namely the conformations
of minimum energy. One might ask the questions: what is
the significance of those contour plots? -and what is
their use? If the answer is that they may guide us in
modeling intermediate conformations which might be taken
up in crystals, in aqueous solution, or near the active
site of en enzyme, a more rational use of computer faci-
lities would probably be to chart valleys of the confor-
mational map (16-17). A better approach than this,
though more costly, is to let the molecule deform along
low-frequency normal coordinates and follow the confor-
mational evolvement in time with Molecular Dynamics. A
procedure for selecting the normal coordinates most

relevant to conformational interchange was worked out
for the case of a coordination compound (<u>20</u>).

Conclusion

We can state, in conclusion, that optimization of the
potential energy function parameters on experimental
data of small model compounds has led to a parameter set
that gives an overall improvement of the accuracy of
postdiction and, by implication, of the validity of
prediction. The improvement is most marked in the most
flexible substance.
     The phrase postdiction is used to emphasize that
properties are calculated which were not used in the
development of the PEF and that we are not dealing with
just reproduction. Postdiction is therefore "predic-
tion" of known properties while prediction deals with so
far unknown properties.

Technical Matters

The CFF program is available, free of charge, and can
most easily come on EARN or BITNET, from KEAKJR at VM.-
UNI-C.DK or at NEUVM1 or from UNIJF at VM.UNI-C.DK or at
NEUVM1. No responsibility for problem solving and tech-
nical updatings can be accepted; the manpower available
forbids service of any kind. At the time of writing,
distribution through a software house is under consi-
deration.
     The CFF is known to run or have run on CRAY XMP,
Amdahl VP1100, many IBMs, Siemens, UNISYS, CDC, many
VAXes, Ardent Titan. The program is a patchwork pre-
pared over 20 years, written in IBM FORTRAN IV and later
cleaned to conform to FORTRAN 77; new routines are writ-
ten in FORTRAN 77. Development is now done on an Amdahl
VP1100, and vectorization is used where appropriate.

Literature Cited

1.   Niketić, S. R.; Rasmussen, K. <u>The Consistent
      Force Field: A Documentation</u>; Lecture Notes in
      Chemistry, Vol. 3; Springer-Verlag: Berlin,
      Heidelberg, New York, 1977.
2.   Rasmussen, K. <u>Potential Energy Functions in
      Conformational Analysis</u>; Lecture Notes in Chemi-
      stry, Vol. 37; Springer-Verlag: Berlin, Heidel-
      berg, New York, Tokyo, 1985.
3.   Rasmussen, K. In <u>Molecular Structure and Dyna-
      mics</u>; Balaban, M., Ed.; Balaban: Jerusalem, 1980;
      pp 171-210.
4.   Rasmussen, K. In <u>Strategies for Computer Chemi-
      stry</u>; Tosi, C., Ed.; Kluwer: Dordrecht, 1989; pp
      13-29.

5.   Kuchitsu, K.; In Phys. Chem. Ser. 1, Vol. 2, MTP
     Int. Rev. Sci; Butterworths: 1972; pp 203-239.
6.   Kuchitsu, K.; Cyvin, S. J. In Molecular Struc-
     tures and Vibrations; Cyvin, S. J., Ed.; Elsevi-
     er: Amsterdam, 1972; Chapter 12; pp 183-211.
7.   Kildeby, K.; Melberg, S.; Rasmussen, K. Acta
     Chem. Scand. 1977, A31, 1-13.
8.   Melberg, S.; Rasmussen, K. Acta Chem. Scand.
     1978, A32, 187-188.
9.   Melberg, S.; Rasmussen, K. Carbohydr. Res. 1980,
     78, 215-224.
10. Melberg, S.; Rasmussen, K. J. Mol. Struct. 1979,
     57, 215-239.
11. Rasmussen, K. Acta Chem. Scand. 1982, A36, 323-
     327.
12. Brown, G. M.; Levy, H. A. Science 1965, 147,
     1038-1039.
13. Chu, S. S. C.; Jeffrey, G. A. Acta Cryst. 1968,
     B24, 830-838.
14. Lemieux, R. V.; Brewer, J. T. Adv. Chem. Ser.
     1973, 117, 121-146.
15. Rohrer, D. C.; Sarko, A.; Bluhm, T. L.; Lee, Y.
     N. Acta Cryst. 1980, B36, 650-654.
16. Melberg, S.; Rasmussen, K. Carbohydr. Res. 1979,
     69, 27-38.
17. Melberg, S.; Rasmussen, K. Carbohydr. Res. 1979,
     71, 25-34.
18. French, A. D. Carbohydr. Res. 1989, 188, 206-211.
19. French, A. D. In Cellulose and Wood - Chemistry
     and Technology; Schuerch, C., Ed.; Wiley: New
     York 1989; pp 103-118.
20. Niketić, S. R.; Rasmussen, K. Acta Chem. Scand.
     1981, A35, 213-218.

Chapter 12

Conformational Analysis of a Disaccharide (Cellobiose) with the Molecular Mechanics Program (MM2)

Alfred D. French[1], V. H. Tran[2], and Serge Pérez[2]

[1]Southern Regional Research Center, U.S. Department of Agriculture, P.O. Box 19687, New Orleans, LA 70179
[2]Institut National de la Recherche Agronomique, B.P. 527, 44026, Nantes, France

A strategy for automated, flexible-residue conformational analysis of disaccharides is presented with examples from a study of cellobiose. The strategy includes modifications of the MM2 program to give a rigid dihedral driver option that starts with the same intra-residue geometry at each increment of the driven torsion angles. This avoids the propagation of residue distortions from one conformation to the next. In analyzing cellobiose, the use of four starting models with different combinations of side group orientations provided at least one satisfactory optimization for each linkage conformation. Each starting model contributed to a table of lowest energy values but the low-energy region of the resulting map was similar to earlier work based on a single starting model.

Many monosaccharides have a single, well-established, preferred ring conformation, such as $^4C_1$. Therefore, the objective of a typical conformational analysis (CA) of disaccharides is the understanding of the varying energetic relationship between the two residues as they are rotated about their bonds to the oxygen atom of the glycosidic linkage. These rotations are described by the torsion angles ϕ and Ψ, shown in Figure 1.

One might (naively) employ CA to answer the question, "What is the most likely shape of a molecule?" However, crystallographic and other experimental evidence shows that the conformations of individual residues ([1,2]), disaccharides ([3]) and polysaccharides ([4]) vary, often substantially. Perhaps then, it is more appropriate to think of CA as a tool for predicting the range or ranges of attainable conformations. Of these attainable conformations, observed values of ϕ and Ψ will vary, depending on crystal packing in the solid state or the type of solvent in solutions.

Although the main variables of disaccharide CA are ϕ and Ψ, an objective treatment requires finding the least energetic combination of all the other conformational variables at each ϕ,Ψ point. In a practical sense, this requires computer models of sugar residues that are flexible. All bond lengths, bond angles and torsion angles other than ϕ and Ψ must be adjusted at each increment of ϕ and Ψ in order to obtain the lowest possible potential energy.
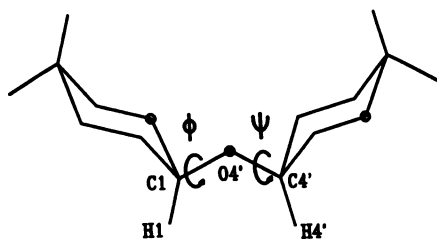
Figure 1. A (1 -> 4) disaccharide showing Ψ and φ, based on the
torsion angles H1-C1-O4'-C4' and C1-O4'-C4'-H4', respectively.

While pioneering work with flexible carbohydrate residues was done a decade ago (5), CA with flexible residues over all of ϕ,Ψ space is a recent development (6-9).

The molecular mechanics program used in the present work was the 1985 version of MMP2 (10,11). Despite some successes with MM2, and its predecessor, MM1, on carbohydrates (12-14), its application to CA of disaccharides is not straight-forward. The major difficulties with CA of disaccharides using MM2, or any other program, arise from the multiple minimum problem. A strategy for surmounting this classic obstacle is presented in the following paper by Tran and Brady. That laborious strategy depends on the availability of a flexible definition of the pattern for conformational searching in the CHARMM (15) program. MM2 has no such facility for a semi-automated, pseudo-radial conformational search, resulting in an additional challenge.

This paper gives an alternative to the strategy described in the Tran-Brady paper for performing CA of disaccharides. The method herein is not as elegant, but is perhaps better suited to automation. Another advantage is that it is easier to describe the construction of a given map of conformational energy over ϕ,Ψ space so that other workers could reproduce it. In order to automate this simpler approach, it was still necessary to modify the MM2 program, and the modifications are described. Preliminary modeling work on cellobiose (7,16) is confirmed by examples that use the more complete treatment permitted by the modified program.

Both the Tran-Brady paper and this one describe initial attempts to develop methods and the underlying philosophy for CA through models of complicated structures that can deform inelastically. (Here, an inelastic deformation means that an alternate conformation for one or more structural features was adopted during energy minimization. Examples include the rotation of an hydroxyl group through an energy barrier to an alternate staggered position or the changing of a pyranoid ring from the $^4C_1$ shape.) Although we want the molecular model to deform during CA, we must cope with the inelastic deformations that occur when analyzing combinations of ϕ and Ψ that have high energies. This is a problem when using the standard facilities for CA within MM2 because the starting geometry for each optimization is the previously optimized structure. Any inelastic deformation is thus likely to be transmitted to the next structure and the conformation and energy will not, in general, be the same before and after $360°$ of rotation. This difficulty is in addition to the more classic aspect of the multiple minima problem where an overwhelming number of possible structures must be tested to ascertain the least energetic structure. The strategy presented thus must overcome both types of problem.

The problem of inelastic deformations is in addition to other problems associated with trying to assess the potential energies at various rotations about bonds. Burkert and Allinger (17) have discussed several aspects of these calculations, including the problem that the rotations are usually defined by only one of several torsion angles associated with a given bond. Typically, there is an artifactual "lag" in the torsion angles that are not used by the modeling program to define the rotation about the bond.

Flexible-Residue Justification

In light of the difficulties just discussed, one might wonder whether the incorporation of residue flexibility is worthwhile. "Rigid-residue" methods such as HSEA (18) require far less computer time than flexible-residue methods. We cite two practical advantages of

allowing internal adjustments besides the basic appeal of incorporating a known aspect of the molecule in the model:

1.   Since the residue can flex, detailed aspects of the starting geometry of the residue are not critical. With rigid-residue analysis, starting geometries taken from various crystal structures give minima in different positions (19). Rigid-residue analyses starting from disaccharide crystal structures will almost inevitably favor the starting conformation if the potential functions are reasonable.

2.   If the various ϕ,Ψ combinations found in single-crystal diffraction studies are plotted on CA maps, the energies corresponding to these combinations are often lower on maps prepared with flexible residues than on maps made with rigid residues (3,20,21). The energies calculated with flexible-residue methods for experimentally determined conformations are in accord with energies that could be expected from hydrogen bonding and van der Waals forces.

While it is difficult to verify experimentally the calculated heights of conformational barriers, it seems that flexible-residue methods can give better results. Energies based on rigid residues increase to artificially high values at large distances from the starting ϕ,Ψ conformation (22).

The MM2 Program

The computer program used herein, MM2, is one of many (23) that adjust ("optimize") the atomic coordinates of a molecule to produce a structure at a local minimum on a multidimensional hypersurface of potential energy. Such programs require predefined equations and constants for the calculation of the energy of every type of interaction, i.e., bond stretching, bond angle bending, torsions and non-bonded van der Waals forces. In its academic versions, MM2 (and MMP2 versions that include delocalized pi electrons) does not provide graphic display and is best considered a tool for structure optimization (energy minimization) and for CA. Neither version includes facilities for molecular dynamics or Monte Carlo techniques.

Attractive Attributes. MM2 is attractive for several reasons:

1.   It is a general-purpose program that is carefully parameterized for a wide variety of molecular types.

2.   Two recent versions of MM2, MMP2(85) and MM2(87), automatically compensate for the anomeric effects that are important for sugars. Accomodations for carbohydrates are discussed further in the chapter in this book by French, Rowland and Allinger.

3.   MM2 is available (except to Communist countries) for a copying fee through the Quantum Chemistry Program Exchange (QCPE), Department of Chemistry, Indiana University, Bloomington, IN 47901. There are several versions for several kinds of computers. Only academic workers can obtain the newest version, MM2(87) (or MMP2(85) on which this work is based) through the QCPE. Other users may get those versions from Molecular Design, Ltd, San Leandro, California, 2132 Farallon Drive 94577. The commercial versions use the same methods for energy and structure calculations, but are enhanced for easier preparation of input files, etc.

The manual provided by QCPE for MM2 is useful as are journal articles (10,11); two books are recommended to prospective users of MM2 (24,25). Also, the QCPE sponsors training courses.

Limitations. Experience shows that there are some limitations when working with MM2.

1. The task of creating input files is tedious for molecules as large as disaccharides and additional support is advisable for users of the academic versions. Several programs from the QCPE provide this capability, as do a number of commercial programs. The best of such programs create a standard MM2 input file after the user draws the structure on a terminal screen.
2. Like other programs for determining least energetic conformations, MM2 only finds local minima. Alternate structures separated by energy barriers must be explicitly tested and their energies compared. It is especially difficult to cover all possible alternate structures for carbohydrates. This is due both to the nature of carbohydrates and to a limitation in MM2. Only two torsion angles can be varied systematically in the standard program.
3. MM2 is slow compared to programs with simpler potential functions, although it is rapid compared to quantum mechanical methods. MM2 requires lone pairs of electrons on all ether and hydroxyl oxygen atoms and nitrogen atoms. These lone pairs are treated as if they are atoms and thus the number of "atoms" is increased by as much as 50% for carbohydrates. This can double the required computer time compared to calculations not using lone pairs.
4. The complexity of the potential functions inhibits the extent of parameterization, although many structures can be modeled.
5. The modifications described below are necessary for automated CA of molecules that can deform inelastically. New releases of MM3, the successor to MM2, should incorporate some of these changes. (See the chapter by French, Rowland and Allinger for information on MM3.)

Problems with Modeling Carbohydrates

Two aspects of carbohydrate structure are especially problematic for modeling because of the multiple minimum problem:

Ring Geometry. The number of possible ring conformers ($^4C_1$, $^1C_4$, $^1S_5$, etc.) is potentially large. That number is squared to give the number of starting models that might require consideration for a disaccharide, since the two rings in disaccharides could possibly have two different forms. In some cases (26), one must test several ring forms, increasing the complexity of the study. During the optimization of very flexible rings such as fructofuranoses (French, A. D.; Tran, V. H. *Biopolymers*, In press), several different conformations can be visited enroute to the least energetic structure.

Rotating Side Groups. The positions of rotating side groups on sugars affect the calculated energy values. Primary alcohol groups usually exist in staggered positions (gg, gt, and tg) (27) that correspond to local minima. Primary alcohol groups of pyranoses occur mostly in one of two positions, avoiding interactions such as between O4 and O6 in glucose if O6 has a tg position. In both solids

and solutions, gt and gg positions are preferred for glucose, while the tg and gt positions are preferred for galactose (28).

Even hydrogen atoms in secondary hydroxyl groups are problematic, with different arrangements giving a range of energy values. However, one need not usually consider all three staggered positions for each hydroxyl group. The lowest energies for models of pyranose rings occur when the secondary hydroxyl groups all have similar relative orientations. This enables the formation of cooperative rings of intramolecular hydrogen-bonds. These similar orientations are described as clockwise (C) or anticlockwise (R) (8). A paper by Tvaroška, Kožár and Hricovíni in this book describes an alternate procedure for coping with variable side group positions.

In the present case (cellobiose), four different models were tested. They were gtgtRR, gtgtCC, ggggRR and ggggRC, shown in Figure 2. More combinations were not used as starting models because the number of changes in the energy map seemed to diminish with each successive trial. Unless all possibilities are tried, of course, there is no way to know with certainty that the lowest energy has been attained at each ϕ, Ψ point. While more structures can be tested, it is not reasonable to test all possibilities. About one week is required to test each starting model on a MicroVax II and there are about 3^{10} possibilities. Instead, we seek a result that will have an error less greater than 1 kcal/mol at each ϕ, Ψ point, at least in the interesting, low-energy zones. This error is in addition to the overall deficiencies in the force field, such as the underestimation of hydrogen bonding energy in MMP2(85) (29) and neglect of any environmental interactions.

Most molecular mechanics studies do not indicate that one position of the primary alcohol group has an energy prohibitively higher than the others. This is not consistent with the experimental data so we conclude that the model is not complete. Also, inter-residue hydrogen bonds are often observed under experimental conditions but intra-molecular hydrogen bonds are favored in our models because the molecule is isolated. Therefore, the purpose in using a variety of different starting models is not to determine the preferred side group orientations. Instead, alternate starting arrangements were used to assure attainment of low energies for $\phi-\Psi$ values that otherwise might have higher energy values caused by positions of side groups that cause interference.

Problems with Flexible-Residue Analysis

Because the internal geometry of each residue responds to forces arising from the proximity of the other half of the disaccharide, an apparent conflict arises between two desirable goals of CA. On one hand, we hope that model residues deform during changes in ϕ and Ψ in a manner similar to real molecules that undergo similar motions. One might expect that the structure and energy values of real molecules would be different before and immediately after 360° rotations about ϕ and Ψ. On the other hand, a ϕ, Ψ map must have the same energies at +180 and −180° in order to show the minimal energy at each ϕ, Ψ conformation. This conflict is a typical example of the difference between kinetically determined results and thermodynamically determined ones. Energy minimization algorithms however, cannot generally overcome false minima, so inelastically deformed models are not brought to the thermodynamically best structure during CA.

A modeling study can avoid inelastic deformations by only searching conformation space close to the minima as in the pseudo-radial search method described in the preceding paper by Tran and Brady. That type of search mimics the thermal motion of a molecule,

which mostly stays within the low-energy areas. Only after the low-energy regions are established does one attempt to determine the energies of linkage conformations that might deform the model inelastically. The approach used in the present paper simply tries several different starting models at each point.

Dihedral Drivers

During CA of a disaccharide, the two residues are rotated about their bonds to the linking oxygen. MM2 has a "dihedral driver" facility that accepts the initial, final and increment size values of two torsion angles. At each increment of these torsion angles, the energy is minimized, providing a value for a point on the energy map. The two torsion angles of the molecular model are held at the specified values by assigning a large potential energy to changes of the two torsion angles. This approach allows optimization of all other structural characteristics for all atoms including those that define the torsion angles. After optimization is complete, the energy is recalculated with the usual torsional potential. Two types of dihedral drivers are available in standard MM2. One option provides for changes of torsion angles <u>within</u> rings. It functions slowly according to the program manual and will not be discussed further. The other available option is for use with side groups, and therefore would be better suited for changing the values of ϕ and Ψ.
     With this option (the -1 option in the MM2 manual), the residues of the starting model are rotated rigidly (without internal change) to the first ϕ,Ψ combination to be considered. After the first optimization finishes, the first torsion angle specified (eg. ϕ) is changed by its increment, rigidly rotating one of the newly optimized residues. This new structure is optimized, and the process continues until ϕ has undergone all the specified increments. Then, the second torsion angle, Ψ, is changed by its specified increment and all values of ϕ are again tested. This scheme is shown in Figure 3 (Option -1), with each arrowhead representing a point where the structure would be optimized.

<u>The Problem with the Standard Driver</u>. For ease of use, it is desirable to step both ϕ and Ψ through $360°$ in an automated procedure. However, this will cause the model to pass through some conformations that result in inelastic deformations. Since the standard driver begins each optimization with the internal residue geometries of the preceding conformation, reorientations of side groups and other deformations are often carried forward. Although it is possible that optimizations at subsequent conformations would "repair" the residue geometry, it does not happen often.
     The effects of propagated distortions of the residue are shown in Figure 4, a CA map without contouring that was prepared with the standard driver. The gtgtRR starting model of cellobiose had an energy of 31.4 kcal/mol (its conformation was $\phi = 20$, $\Psi = -60$). After rigidly rotating to $\phi = -180$, $\Psi = -180$ and optimizing at increments of $20°$ over $360°$, the smallest energy found was 32.8 kcal/mol. The secondary hydroxyl group orientations were changed at an early ϕ,Ψ conformation and not restored. Another manifestation of the deformation is that the energy values at $\phi = -180$, $\Psi = 140$ and at $\phi = +180$, $\Psi = 140$, differ by 5 kcal/mol. As the conformational search proceeded between these two points, a side group changed to a different (but not the initial) position.
     Figure 4 might represent well the energies that would be found immediately after a real molecule was forced to change conformations

g t g t C C   g g g g R R

g t g t C C   g g g g R C
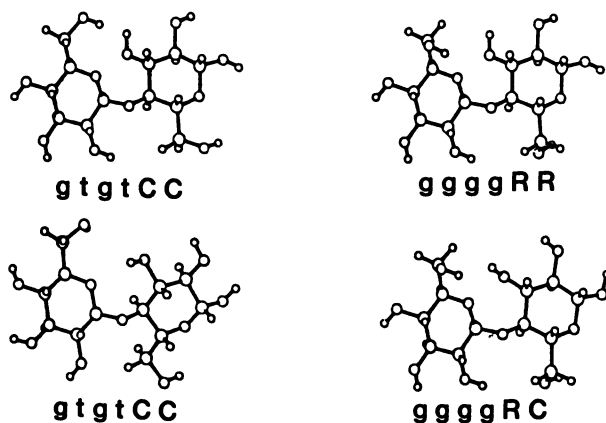
Figure 2. The four starting models used for the study of cellobiose
(lone pairs of electrons are not shown). Convention defines the R
and C notation when the residue is in a conventional orientation and
is viewed from above. The least energetic structure observed in this
study is gtgtRR. This Figure and Figure 5 were drawn with CHEMX,
developed and distributed by Chemical Design Ltd, Oxford, England.

ψ

φ   φ

Standard   New Option −2
Option −1
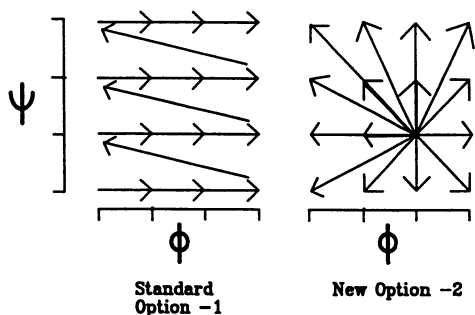
Figure 3. A comparison of two methods of producing starting
conformations. With standard option −1, the conformations are
generated from the preceding structure. With our −2 option, all
conformations within a run are generated from the same, single
starting point.

along the path given by the standard driver. However, such a path of
conformational change is improbable. Real molecules would avoid
high-energy conformations and deformed species would eventually
revert to lower-energy conformations regardless of how inelastically
deformed a model might be. The faults in this map (failure to attain
the energy value of the starting conformation and the differences in
energies at each side) result from the continuous application of the
standard dihedral driver in MM2.

A New Driver. In our strategy, we analyze each ϕ, Ψ conformation
independently. Each optimization starts with the same residue
geometries, which are rotated rigidly from the initial conformation
directly to the ϕ, Ψ point in question. MM2 was modified so this
task can be automated through a new dihedral driver option that we
have designated as -2. The relations of starting models to the
optimized points are also shown in Figure 3 for the new driver
option. This approach maintains control over the starting geometry,
and directly overcomes the two faults described for results from the
standard driver.

A New Problem. While our new driver solves some important problems,
it creates a new one, i.e., structures at several ϕ, Ψ points fail
to optimize properly. An example is shown in Figure 5 for cellobiose
with ϕ of -100 and Ψ of -80. Initially, this conformation, when
imposed on a gtgtRR model, places the centers of the O2 and O3' atoms
only 0.488 Å apart (Figure 5a). (In a rigid-residue analysis, this
conflict would cause a very high energy to be calculated.) Some of
the bonds to the lone pairs overlap and a contact of 0.119 Å occurs
between one of the lone electron pairs and the other oxygen atom.
      Severe distortions occured when MM2 moved the atoms to try to
reduce the energy of the tangled model in Figure 5a. The
optimization did not proceed correctly because movement to resolve
the inter-residue conflicts would have initially increased the
severity of the van der Waals repulsions. Instead, some of the bond
lengths and other features assumed highly improbable values. The
resulting structure (Figure 5b) has a reported energy of -6469
kcal/mol. (A suitable warning was issued by MM2 that non-standard
bond lengths had occurred and that optimization was terminated.)
This wrong value of the energy results from the cubic term in the
bond-stretching component of the calculated energy. As noted in Ref.
11, "When energy minimization is done with a very poor starting
geometry, [the cubic function] may lead to disaster--with the
molecule flying apart." Since bond lengths were initially in the
correct range, the cubic contribution to bond stretching was not
suppressed and the large negative energy was obtained.
      The third structure (Figure 5c) is an optimized result with
the same values of ϕ and Ψ and an energy of 54.2 kcal/mol, taken from
the work depicted in Figure 4 that used the standard driver. The
starting geometry was gtgtRR, but earlier optimizations had
reoriented the hydroxyl groups on the non-reducing residue and
adjusted the residue geometries. This preconditioning eliminated the
tangling and allowed MM2 to successfully optimize the structure.
However, optimizations at following ϕ,Ψ values failed to return the
hydroxyl groups to the gtgtRR position. This is why, when the CA
comes to the point with ϕ = 20 and Ψ = -60 (as in the starting
geometry), the calculated energy was 32.8 kcal/mol instead of 31.4
kcal/mol. Although the individual residues are distorted in
Figure 5c they still would be classed as $^4C_1$ shapes.
      Together, Figures 4 and 5b contrast the problems of the two
types of automated analysis. The gradual approach to high-energy

```
45 43 46 51 46 45 45 42 38 36 35 36 37 39 41 43 44 44 45   180
49 51 52 53 50 45 44 44 42 40 39 37 39 43 45 46 45 44 47   160
55 62 62 61 60 53 49 48 49 48 46 45 48 52 54 50 51 49 50   140
57 64 73 64 60 57 50 46 45 44 44 45 48 55 56 56 54 53 57   120
49 53 61 69 70 59 57 49 46 44 43 42 43 48 55 62 53 51 51   100
45 47 51 59 64 65 53 45 42 42 43 41 39 41 47 53 58 52 46    80
47 43 45 50 54 56 57 56 41 39 39 40 39 39 42 46 49 50 47    60
44 43 43 45 48 47 48 49 42 38 36 36 37 39 40 42 43 43 44    40
39 41 43 45 48 42 42 42 42 39 35 34 35 38 40 41 40 38 40    20
37 40 43 42 42 40 38 37 38 37 35 34 34 37 40 41 40 37 37     0  Ψ
37 40 44 44 41 40 37 35 34 35 35 34 34 36 39 40 41 39 38   -20
41 42 46 46 43 39 37 35 34 33 34 36 37 38 39 39 39 42 41   -40
45 46 49 51 49 42 38 38 36 34 33 34 39 43 43 41 39 40 45   -60
44 50 53 54 54 49 43 39 40 39 35 34 38 44 48 47 43 41 44   -80
45 48 55 62 56 55 48 45 42 42 40 38 39 43 50 51 49 45 45  -100
48 48 52 57 62 55 52 48 45 42 41 39 40 43 48 52 50 47 48  -120
43 46 50 52 56 54 49 46 44 42 40 38 37 40 46 49 49 45 43  -140
42 42 47 49 49 50 47 43 40 39 37 37 36 38 41 45 46 45 42  -160
44 44 47 52 46 45 45 42 38 36 35 36 37 39 41 43 44 44 45  -180

 1  1  1  1  1                               1  1  1  1  1
-8 -6 -4 -2 -0 -8 -6 -4 -2     2  4  6  8  0  2  4  6  8
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0

                        φ
```

Figure 4. Energy values (Kcal/mol) for a gtgtRR starting model produced with MM2 and its standard option -1 dihedral driver.
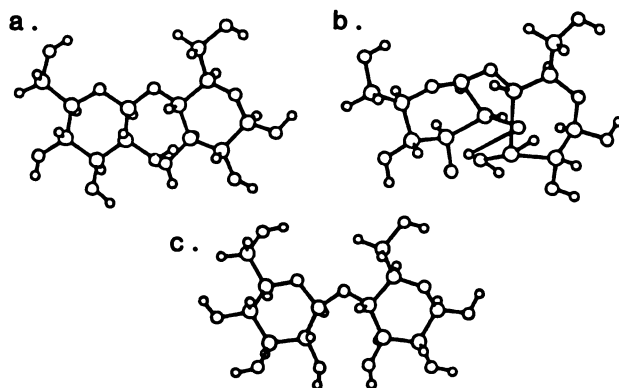
Figure 5. a) The starting model of cellobiose (gtgtRR) after rigid rotation to Ψ = -80, φ = -100. b) The result of attempted optimization by MM2. c) The same linkage conformation, but the structure was taken from the study that produced the map in Figure 4.

conformations with the standard driver is more likely to provide
successful optimizations. On the other hand, energy values and
residue geometries depend on which conformations preceded the ϕ,Ψ
point in question. With the new driver, in which the starting
residue geometries are rigidly rotated to the desired ϕ and Ψ values,
bad starting geometries are more likely and structures may not
optimize properly. The extent of the problems inherent in conducting
CA of disaccharides with the standard driver option is, if anything,
understated in this demonstration because of the equatorial linkages
in cellobiose. During such automated CA, models with axial linkages
encounter more severe inter-residue contacts and, hence, residue
deformations.

Working Around the New Problem. Since structures such as the one
shown in Figure 5b are computational artifacts, their energies should
be discarded. There are at least three ways to minimize the impact
of the missing energy values. If there is no energy value for a ϕ,Ψ
point, one can be extrapolated from neighboring values. The SURFER
program (Golden Software, Golden, Colorado) produces contour plots
from grids with missing data through extrapolation. Conformations
affected by this problem have energies so high that the conformations
are improbable, and a reasonable error in the extrapolated value will
have little effect on the important, low-energy regions of the ϕ,Ψ
map.
        A second way depends on the use of several starting models
with different hydroxyl and primary alcohol group orientations for
calculation of the energy at each ϕ,Ψ point. Since our goal is to
determine the lowest energy value at each point, the energies that
are clearly in error can be discarded and the best remaining energy
values can be used. If several starting structures are used, it will
be rare if none of them produces a reasonable value. At the ϕ and Ψ
values of the models in Figure 5, three of the four starting models
failed to optimize properly. The fourth, however, gave an energy of
52.4 kcal/mol, 1.8 kcal/mol lower than the value obtained with the
standard driver option.
        A third approach is to use a satisfactorily optimized geometry
from a neighboring point as a starting geometry. If that is done, one
will probably find that conformations and energies depend on the
direction of approach.
        The best remedy is to prevent the entanglement that results in
the incorrect structures. As shown in the chapter by Brant and
Christ, one way to minimize inter-residue contacts is to increase the
bond angle at the oxygen atom that links the two residues together to
about 125°. While the optimization routine will return the value of
the glycosidic angle to about 117 degrees, the residue geometries
will simultaneously adjust to avoid tangling. Such a modification to
the above strategy has been fairly successful in preliminary testing.
After the structures to be used as starting geometries have been
initially optimized, their linkage bond angles are increased to the
larger value. These new structures are then used as the starting
models with the new driver option. The MM3 program may often avoid
this problem because of two changes. Explicit lone pairs will not be
used, and the cubic bond stretching function of MM2 will be replaced
by a quartic equation (30).

Clues from Optimization Reports

Testing for Valid Optimizations by Energy Value. With hundreds or
thousands of data points to examine, detailed inspection for
successfully optimized structures is tedious. Limited experience has

shown that successfully optimized disaccharides can be detected by
their energy values. Their "Final Steric Energies" should be between
about 25 and 75 kcal/mol with MMP2(85). Other software, including
MM2(87), will have rather different ranges of energies, as will other
molecular structures. The least energetic MMP2(85) values for
permethylated cellobiose, for example, are about 80 kcal/mol (French,
A. D. Unpublished data). Values outside this range indicate that
the structure has not been properly optimized, as discussed above.
Problems may exist even when MM2's energy is within the above range.
Large, unreasonable values for individual terms may fortuitously
balance each other in a way that their total appears to be
reasonable.

Evidence for Transitions. Standard MM2 gives a record of the energy
values and the average atomic movement as the structure adjusts to
provide lower energy values. Initially, the movement is often large
(several hundredths of an Ångstrom). It becomes progressively
smaller as the energy approaches the final value. The rate of change
in these values indicates the extent and type of difference between
the initial and final structures. Two clues can be gained from
examining the average atomic movement values.
    While the large atomic movements often cease almost
immediately, they may stay at a nearly constant, moderate value for
an extended time before dropping off further. Initial large
movements correspond to changes in initial atomic positions for most
atoms. Movement values that remain nearly constant indicate a change
(probably a rotation) of one group relative to another while the
relative positions of the atoms within the groups are nearly
unchanged. In the special case where an energy barrier to rotation
is overcome, the average atomic movement may increase temporarily and
then resume its downward trend.

Modifications to MM2

We changed the MM2 program for more efficient use on disaccharides.
The modified version gives the same results as the original, unless
the new options are selected. Our version reports that a transition
may have occurred if the average atomic movement increases. This
eliminates the need to report the history of the average atomic
movement during CA.
    Megabytes of disk space per CA run are saved by omitting
redundant information and the reports of average atomic movement. We
have implemented IPRINT options 5 and 6 in addition to the options
1-4 of the standard program. Both 5 and 6 eliminate the same
information from the standard output, but option 6 does not produce
the secondary output files (FOR009.DAT) that roughly correspond to
MM2 input files, further saving disk space. Besides conserving disk
space, the briefer output files can be more quickly scanned for the
important results. Another change was to place the energy result in
the FOR009.DAT files as well as in the main system output. This
stores the final energy values on disk even if the main output is
sent to the video display when using options 1-5.
    A greater understanding of the MM2 program is needed to
implement the new -2 option for the dihedral driver. In MM2, a
temporary file stores the coordinates at the end of each optimization
for use as starting positions for the next optimization. The
procedures that create these files had to be changed. An alternative
to modification of MM2, used previously (3), was to create separate
input files for each ϕ,Ψ conformation of interest. This allowed use
of the standard driver with a 0° increment size. Special programs
could be used to prepare all of the input files. The check of

whether the new driver has been properly implemented is whether it gives the same results at a variety of conformations as the standard driver, used with an increment size of zero.

<u>Application to Cellobiose</u>

A preliminary energy map for cellobiose has been published (<u>7,16</u>) but it was based on only one combination of rotating group positions. Similar maps based on other starting models were needed to confirm the initial work. The optimal gtgtRR structure from the earlier work was altered to give three additional starting structures (all shown in Figure 2). φ and Ψ were stepped in $20°$ increments from -180 to $+160°$. In the previous work, an irregular grid was used, with $10°$ increments in the low-energy regions. The default dielectric constant of 1.5 was used, appropriate for an isolated molecule. The MM2 calculations were carried out on VAX computers. The energy values were managed with a program given in the Appendix that was written in GWBASIC for IBM-PC compatibles.

<u>Results</u>

Energy values for the gtgtRR model were the same as computed earlier for the same φ and Ψ values, confirming that the program modifications had not altered the calculated energy values. From the results for all four starting models, the utility program in the Appendix selected the 324 lowest energy values shown in Figure 6a. Figure 6b shows that 220 of those 324 energies arose from the gtgtRR starting structure used in the earlier work. Another 57 preferred conformations started as ggggRR, 31 points arose from ggggRC, and 16 came from gtgtCC structures.

  Figure 6c shows the ranges of energy values at each φ,Ψ point. These ranges are based only on structures that optimized properly. Therefore, some ranges are based on less than four energies. The magnitudes of the ranges show the importance of the rotating groups, although the difference between the gtgtRR model and the one with the lowest energy at the conformation in question was usually small, shown in Figure 6d.

  In the four sets of 324 points calculated, structures failed to optimize properly 37 times. Figure 6e shows the locations and the numbers of those models. All φ,Ψ points that corresponded to improperly optimized conformations, when tested with other starting models, gave optimized energies at least 10 kcal above the minimum.

  Figure 7 is a contour plot of the data in Figure 6a. It is almost identical to the plot published earlier that embodied 497 points based on one starting model instead of 324 points for each of 4 starting models. Implications of the various minima and barriers are discussed in Refs. 7 and 16. Figure 8a shows the differences between maps made with driver options -1 and -2 for the gtgtRR starting model. Figure 8b is a SURFER contour plot of the aperiodic data in Figure 4. These plots exhibit the various options of the utility program in the Appendix, the effects of the new driver, and the addition of extra starting geometries.

<u>Conclusions</u>

The strategy conveyed in this paper permits coherent results from an automated CA while using flexible residues. By testing all the starting models over the entire range of φ and Ψ, parallel sets of data were obtained that were submitted to a simple program for final analysis. This minimizes the personal time required to produce a

```
47 49 51 53 48 43 42 43 39 36 34 34 36 40 42 44 42 43   160
49 56 57 57 54 47 44 43 42 40 40 39 42 44 47 48 48 46   140
53 56 63 63 57 54 49 45 44 43 41 42 45 49 50 51 50 50   120
47 52 60 66 66 55 53 48 46 44 42 40 41 45 51 52 50 47   100
44 45 51 58 65 58 50 46 43 42 42 39 38 40 44 49 50 45    80
42 42 45 51 53 57 53 46 41 39 38 38 37 37 39 42 45 44    60
40 41 42 45 47 49 50 51 42 38 35 34 35 37 37 39 39 39    40
36 38 40 43 44 43 43 44 43 38 34 32 33 36 38 38 36 35    20
34 37 40 42 42 40 39 38 38 37 34 32 32 35 38 39 37 34     0  Ψ
35 37 42 41 41 39 36 35 34 34 34 33 33 35 37 38 39 36   -20
39 41 45 44 41 39 37 35 33 32 32 34 36 37 38 38 38 41   -40
45 46 49 49 43 40 38 37 35 33 31 32 37 41 42 40 38 39   -60
44 50 53 56 52 43 40 39 39 37 34 33 36 40 46 46 42 40   -80
45 49 56 59 58 56 44 41 41 40 38 36 37 41 46 49 47 45  -100
47 49 54 60 60 55 51 44 41 40 39 37 38 41 45 50 48 46  -120
43 46 49 52 59 54 50 45 41 39 37 36 35 38 44 47 47 44  -140
42 43 49 49 49 51 47 42 37 35 34 34 34 36 39 43 44 44  -160
44 44 47 51 46 45 45 41 37 33 32 32 34 37 39 41 42 43  -180

 1  1  1  1  1                            1  1  1  1
-8 -6 -4 -2 -0 -8 -6 -4 -2    2  4  6  8  0  2  4  6
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0

                       φ
```

Figure 6a.   MM2 "Final Steric Energy" values for all tested values of
Phi and Psi.   The value at each point is the lowest of the energies
calculated for the four starting models.   (The largest φ and Ψ
values are 160°).

```
1  4  2  2  1  1  3  1  1  1  1  1  1  1  1  1  1  1   160
1  2  4  2  4  1  3  1  1  1  1  1  2  1  1  1  1  1   140
4  1  2  2  4  4  3  4  1  1  1  3  1  3  1  1  1  3   120
3  3  3  2  1  4  4  4  1  1  1  1  1  3  3  1  1  1   100
3  3  1  4  2  4  4  4  4  3  1  1  1  1  1  3  2  1    80
1  3  1  4  3  2  1  4  1  4  4  3  1  1  1  1  3  1    60
3  4  1  4  4  4  2  4  1  1  1  1  3  1  1  3  3  1    40
1  3  3  1  3  3  4  2  2  1  1  1  1  3  1  1  3  3    20
1  3  3  3  3  3  3  1  1  1  1  1  1  1  1  1  1  4     0  Ψ
4  4  3  1  1  3  3  3  3  1  1  1  1  1  3  1  1  1   -20
1  1  3  1  1  3  3  3  3  3  1  1  1  1  1  1  1  1   -40
1  1  1  1  1  1  1  3  3  1  1  1  1  1  1  1  1  1   -60
1  1  3  1  4  1  1  1  1  1  1  1  1  1  1  1  1  1   -80
1  1  1  3  1  4  1  1  1  1  1  1  1  1  1  1  1  1  -100
1  1  1  1  4  1  1  1  1  1  1  1  1  1  1  1  1  1  -120
1  1  3  3  1  1  1  1  1  1  1  1  1  1  1  1  1  1  -140
1  1  1  1  1  3  1  1  1  1  1  1  1  1  1  1  1  1  -160
1  2  2  3  1  1  3  1  1  1  1  1  1  1  1  1  1  1  -180

 1  1  1  1  1                            1  1  1  1
-8 -6 -4 -2 -0 -8 -6 -4 -2    2  4  6  8  0  2  4  6
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0

                       φ
```

Figure 6b.   The starting model at each point that gave the lowest
energy (1 = gtgtRR, 2 = gtgtCC, 3 = ggggRC and 4 = ggggRR).

```
1  3  2  2  5  3  2  4  4  3  5  5  6  5  3  3  3  2   160
1  1  2  2  2  2  2  2  2  3  4  4  2  4  3  3  3  2   140
3  1  1  2  2  3  2  2  2  2  3  3  2  2  5  3  3  3   120
3  4  1  1  0  3  3  3  2  2  3  3  2  5  1  5  3  3   100
3  2  1  2  1  5  3  1  2  3  3  3  3  2  3  5  2  3    80
3  2  1  3  5  3  4  1  2  3  3  3  3  3  3  3  3  5    60
4  3  2  1  2  1  3  1  2  2  3  3  3  3  3  4  4       40
4  4  3  2  3  1  1  2  1  3  3  4  4  4  4  4  4  4    20
4  3  4  3  3  3  2  2  1  2  3  4  4  4  4  4  4  5     0  Ψ
5  4  4  3  3  4  4  3  3  2  2  3  5  4  5  6  5  5   -20
5  4  0  4  4  4  5  5  5  3  3  2  4  5  6  7  5      -40
5  5  5  5  5  4  5  5  6  5  3  4  3  4  5  6  7  5   -60
5  1  4  3  0  5  5  4  6  6  4  3  3  5  5  5  6  7   -80
6  4  3  5  4  0  4  6  5  6  6  7  3  3  5  5  5  6  -100
4  3  3  5  4  5  8  6  6  6  6  6  6  3  5  3  5  4  -120
3  2  4  4  4  6  4 10  6  6  6  6  6  6  5  3  3  3  -140
2  2  2  3  4  4  4  3  5  5  6  6  5  5  4  4  3  2  -160
3  3  3  2  3  3  3  4  3  4  5  6  5  4  4  4  3  2  -180

1  1  1  1  1                          1  1  1  1
-8 -6 -4 -2 -0 -8 -6 -4 -2     2  4  6  8  0  2  4  6
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0

                        φ
```

Figure 6c.   The ranges of energy values resulting from the 4 starting geometries.

```
 .  .  2  1  .  .  1  .  .  .  .  .  .  XX  .  .  .  .  .  160
 .  .  2  1  1  .  1  .  .  .  .  . XX  .  .  .  .  .  140
XX  .  .  2  2  3  .  1  .  .  .  1  . XX  .  .  .  1  120
 1  4  .  .  .  2  .  2  .  .  .  .  . XX  .  .  .  100
 .  .  .  2  1  .  2  .  .  2  .  .  .  .  . XX XX  .   80
 .  .  .  3  3  1  .  .  .  .  1  .  .  .  . XX  .   60
 1  .  .  .  2  1  2  1  .  .  .  .  .  .  .  .  1  .   40
 .  1  .  .  1  1  1  2  .  .  .  .  .  .  .  .  .   20
 .  .  1  .  1  2  1  .  .  .  .  .  .  .  .  .  .  .    0  Ψ
 .  .  .  .  .  1  1  1  .  .  .  .  .  .  .  .  .  .  -20
 .  . XX  .  .  .  1  1  1  .  .  .  .  .  .  .  .  .  -40
 .  .  .  .  .  .  .  1  .  .  .  .  .  .  .  .  .  -60
 .  .  2  . XX  .  .  .  .  .  .  .  .  .  .  .  .  .  -80
 .  .  .  1  . XX  .  .  .  .  .  .  .  .  .  .  . -100
 .  .  .  1  .  .  .  .  .  .  .  .  .  .  .  .  . -120
 .  .  1  .  .  .  .  .  .  .  .  .  .  .  .  .  . -140
 .  .  .  .  .  4  .  .  .  .  .  .  .  .  .  .  . -160
 .  .  1  .  .  .  1  .  .  .  .  .  .  .  .  .  . -180

1  1  1  1  1                          1  1  1  1
-8 -6 -4 -2 -0 -8 -6 -4 -2     2  4  6  8  0  2  4  6
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0

                        φ
```

Figure 6d.   The difference between gtgtRR and the best value of energy.   If the gtgtRR model was best or the difference was less than 0.5 kcal/mol, a "." is shown.   If a structure failed to optimize correctly (energy outside the range 25 to 75 kcal/mol), XX is shown.

```
.   .   .   1   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   160
.   .   .   .   .   .   .   .   .   .   .   .   1   1   1   .   .   .   140
1   .   1   1   .   .   .   .   .   .   .   .   .   1   .   .   .   .   120
.   .   .   .   3   .   .   .   .   .   .   .   .   .   2   .   .   .   100
.   .   .   .   2   .   .   .   .   .   .   .   .   .   1   1   .   80
.   .   .   .   .   .   1   .   .   .   .   .   .   .   .   1   .   60
.   .   .   .   .   .   1   1   .   .   .   .   .   .   .   .   .   40
.   .   .   .   .   .   .   1   .   .   .   .   .   .   .   .   .   20
.   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   0    Ψ
.   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .  -20
.   .   2   .   .   .   .   .   .   .   .   .   .   .   .   .   .  -40
.   .   1   .   .   .   .   .   .   .   .   .   .   .   .   .   .  -60
.   .   1   1   3   .   .   .   .   .   .   .   .   .   .   .   .  -80
.   .   .   .   3   .   .   .   .   .   .   .   .   .   .   .   . -100
.   .   .   1   .   1   .   .   .   .   .   .   .   .   .   .   . -120
.   .   .   1   .   .   .   .   .   .   .   .   .   .   .   .   . -140
.   .   .   .   1   .   .   .   .   .   .   .   .   .   .   .   . -160
.   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   . -180

1   1   1   1   1                           1   1   1   1
-8  -6  -4  -2  -0  -8  -6  -4  -2      2   4   6   8   0   2   4   6
 0   0   0   0   0   0   0   0   0      0   0   0   0   0   0   0   0

                            φ
```

Figure 6e.   Location and number of models that failed to optimize
(energy values were outside the range of 25 to 75 kcal/mol).

Figure 7.   The contoured map equivalent to the energy grid in
Figure 6a.   Contours are drawn at 1 kcal/mol intervals.

```
  0 -1 -2 -0   0   0 -1   1   2   3   3   3   3   2   2   2   2   1   0  180
  1  2 -0 -1   2   2   1   1   3   4   5   3   3   3   2   2   2   1   0  160
  6  7  3  3   5   6   5   5   7   8   7   7  XX   8   7   2   3   2   2  140
 XX  8 10 -1   0  -0   1   1   1   2   2   1   4  XX   6   5   4   3  XX  120
  1 -2  1  3   4   2   4  -1  -0   0   1   2   2   2  XX  11   4   3   3  100
  2  1  0 -1  -2   6   1  -1  -1  -2   1   2   1   2   4  XX  XX   7   2   80
  6  1  0 -3  -2  -3   4  10  -0   1   1   1   2   2   2   4  XX   6   5   60
  3  2  1 -1  -1  -3  -4  -4  -0   1   1   2   2   2   2   3   3   4   4   40
  3  3  2  2   3  -2  -3  -4  -2   1   1   1   2   2   2   3   3   4   4   20
  3  2  3  0  -1  -1  -2  -1  -0   0   1   2   2   2   2   3   3   3        0  Ψ
  3  2  2 -2  -0  -1  -1  -1  -0   0   1   1   2   2   2   2   3   3      -20
  2  1 XX  2   2  -1  -1  -0   0   1   1   2   1   1   1   1   1   1   2  -40
  0  0 -1  2   6   3  -0  -0   0   1   1   2   2   2   2   1   1   1   0  -60
  0 -1 -3 -2  XX   6   3   0   1   1   1   2   2   4   2   2   1   1   0  -80
  0 -1 -2  2  -2  XX   3   4   1   1   1   2   2   2   4   2   1   1   0 -100
  1 -1 -2 -3   2  -1   1   3   4   2   2   2   2   3   2   2   1   1     -120
  0 -0  0 -1  -4  -1  -1   2   3   3   3   2   2   2   2   2   2   1   0 -140
 -0 -1 -1  0  -0  -4  -0   1   2   4   3   3   2   2   2   2   2   1   0 -160
  0 -1 -1  1   0   0  -1   1   2   3   3   3   3   2   2   2   2   1   0 -180

  1  1  1  1  1                                   1   1   1   1   1
 -8 -6 -4 -2 -0  -8  -6  -4  -2       2   4   6   8   0   2   4   6   8
  0  0  0  0  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

φ

Figure 8a. Grid of energy differences between results from the
standard MM2 option -1 driver and the modified option -2 driver.
Positive values indicate that the -2 driver gave a lower value
(gtgtRR starting structures only).
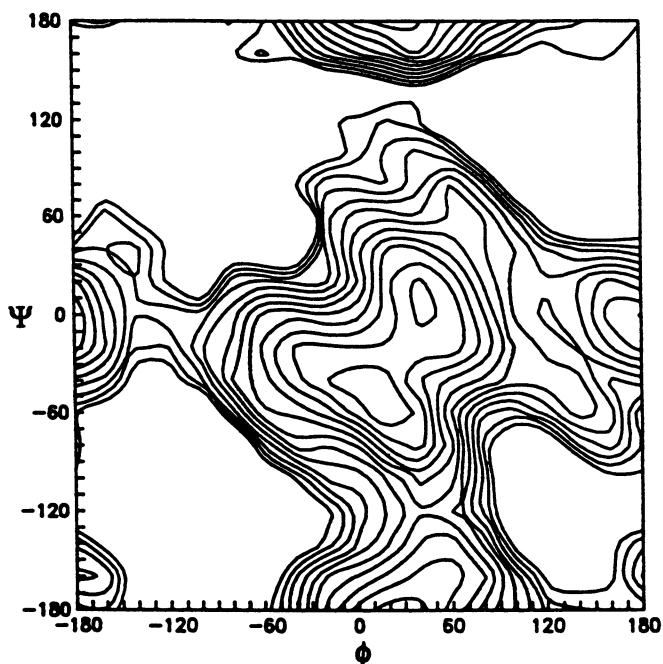
Figure 8b. Contour map based on the standard option -1 driver.
Contours are drawn at 1 kcal/mol levels.

flexible-residue analysis. Although still computationally expensive
(each starting model required about 2.5 cpu days on a MicroVax III),
that factor will diminish to relative insignificance as economical
computers become faster. An advantage to the new driver option is
that the calculated energy for any given ϕ,Ψ point depends only on
the starting geometry and not the preceding points. This not only
avoids the faults discussed above, but it permits combination of the
results with other results. For example, the energies at even values
of ϕ and Ψ could be interspersed with energies at odd values in
limited areas to produce a higher-resolution analysis.

        Overall, the new driver option is a step forward in CA of
molecules that can a deform inelastically. Calculations of energy
take longer because most starting geometries are not as close to the
final result as they are with the standard -1 driver option.
However, it is not necessary to calculate energies at both -180° and
$+180^{\circ}$, saving some time, for a net loss in speed of about 10%. The
problems solved by the new driver option are critical, while the new
problem of occasional improper optimization can readily be worked
around.

Acknowledgments

APPENDIX

The procedure used to prepare the data for the various tables
(Figures 3, 6, and 8) follows. After completing the MM2 runs, the
main output files were each processed with the VMS editor, ED. The
command to write all lines with "FINAL STERIC ENERGY" to a file was
given (WR FILENAME.NRG ALL "FINAL S"). After QUITting the editor,
the NRG file was loaded into the editor and the text was stripped
off, using the command string, S /FINAL STERIC ENERGY // whole. That
was followed with S /KCAL.// whole. Using a communications program,
the 4 NRG files were transfered to an IBM-PC/AT compatible computer
and merged with each other, in such a way that there were 4 columns
of energy values. The resulting file was then input to the following
program available from the author on disk.

        The following program is written in GWBASIC for IBM-PC
compatibles. It assumes that energy values are in a single column
for all Ψ,ϕ points for a given starting model. Energies for
additional starting models must be in additional columns, with all
values in each row corresponding to the same Ψ and ϕ. Energy values
outside the range of 20 to 75 kcal/mol are discarded by the program.
They should not be discarded manually. Besides the input files for
SURFER, the program produced the uncontoured energy maps in Figures
4, 6 and 8.

```
10    'PROGRAM TO TAKE ONE OR MORE LISTS OF MM2 ENERGIES, PREPARE FILE
      FOR
20    'SURFER MAPS, ETC.
30    'WRITTEN BY A. D. FRENCH - VERSION 2.0 MARCH 2, 1989
40    DIM Z(10),M(30,30)
50    UPLIM=75:' Values above this generally result from malformed
      structures
60    LO=25:' Values below this are erroneous for MMP2(85), cellobiose
70    PRINT "Energy analysis utility for MM2 output, SURFER input."
80    PRINT "MAPREP V. 2.0  - March 2, 1989":PRINT:PRINT
90    '
100   'End of preliminaries, start of file handling
110   '
120   PRINT "Current allowed energy range is ";LO;" - ";UPLIM
130   INPUT "NAME OF INPUT FILE";FI$
140   OPEN "I",1,FI$
150   INPUT "NAME OF OUTPUT FILE FOR SURFER INPUT";FO$
160   OPEN "O",2,FO$
170   '
180   'Set up ranges and increments for Phi, Psi
190   '
200   INPUT "IS THIS A STANDARD -180 TO +160 STEP 20 MAP? (Y/N)[Y]";A$
210   IF LEFT$(A$,1)="Y" OR LEFT$(A$,1)="y" OR LEFT$(A$,1)="" GOTO 290
220   INPUT "STARTING PHI VALUE";PHBEG
230   INPUT "ENDING PHI VALUE";PHEND
240   INPUT "INCREMENT OF PHI";PHDEL
250   INPUT "STARTING PSI VALUE";PSBEG
260   INPUT "ENDING PSI VALUE";PSEND
270   INPUT "INCREMENT OF PSI";PSDEL
280   GOTO 340
290   PHBEG=-180:PHEND=160:PHDEL=20
300   PSBEG=-180:PSEND=160:PSDEL=20
310   '
320   'Rest of input, set up type of map desired.
340   INPUT "HOW MANY COLUMNS OF ENERGIES ARE THERE";NCOL
350   PRINT "WHAT TYPE OF ANALYSIS IS DESIRED?"
360   PRINT "    1.  Usual Phi, Psi and Lowest Energy Values"
370   PRINT "    2.  Phi, Psi and Number of Column with Lowest Energy
      Value"
380   PRINT "    3.  Range of Energy Values at Each Phi, Psi"
390   PRINT "    4.  Bad Values on Phi, Psi Grid"
400   PRINT "    5.  Difference Between any Column and Best Value"
410   PRINT "    6.  Phi, Psi and Energy from only One of Several
      Columns"
420   INPUT "YOUR CHOICE (1-6)";OUTVAL
430   IF OUTVAL <1 OR OUTVAL>6 THEN GOTO 350
440   IF OUTVAL=1 THEN OUTVAL$="E":GOTO 720
450   IF OUTVAL=6 THEN OUTVAL$="E":INPUT "Column Number ";ICOL:GOTO
      540
460   IF OUTVAL=2 AND NCOL>1 THEN OUTVAL$="CC": GOTO 720
470   IF OUTVAL=2 THEN PRINT CHR$(7);"1 Column, no choice!": GOTO 360
480   IF OUTVAL=3 AND NCOL>1 THEN OUTVAL$="R": GOTO 720
490   IF OUTVAL=3 THEN PRINT CHR$(7);"1 Column, no range!": GOTO 360
500   IF OUTVAL=4 THEN OUTVAL$="B":GOTO 720
510   IF OUTVAL=5 THEN OUTVAL$="D"
520   IF OUTVAL=5 THEN INPUT "Number of Column for Comparison";ICOL
530   GOTO 720
540   'Routine to extract just one column of energy values'
550   FOR Y=PSBEG TO PSEND STEP PSDEL
560   ROW=ROW+1:COL=0
```

```
570   FOR X=PHBEG TO PHEND STEP PHDEL
580   COL=COL+1
590   FOR K=1 TO NCOL
600   IF K=ICOL THEN INPUT #1,ENG ELSE INPUT #1,DUM
610   NEXT K
620   IF ENG<LO OR ENG>UPLIM GOTO 680
630   M(ROW,COL)=ENG
640   PRINT #2,X,Y,ENG
650   IF X=-180 THEN PRINT #2,-X,Y,ENG
660   IF Y=-180 THEN PRINT #2,X,-Y,ENG
670   IF X=-180 AND Y=-180 THEN PRINT #2,-X, -Y,ENG
680   NEXT X
690   NEXT Y
700   CLOSE:GOTO 1140
710   '
720   'Start of big loop to read and process data
730   '
740   FOR Y=PSBEG TO PSEND STEP PSDEL
750   ROW=ROW+1
760   COL=0
770   FOR X=PHBEG TO PHEND STEP PHDEL
780   COL=COL+1
790   FOR K=1 TO NCOL
800   INPUT #1,Z(K)
810   NEXT K
820   ENG=9999:ZMAX=-9999:ZMIN=9999:BAD=0
830   FOR K=1 TO NCOL
840   IF Z(K)>UPLIM OR Z(K)<LO THEN
Z(K)=UPLIM:BADCOUNT=BADCOUNT+1:BAD=BAD+1
850   IF ENG>Z(K) THEN ENG=Z(K):CHOICE=K
860   IF Z(K)=UPLIM GOTO 890
870   IF Z(K)>ZMAX THEN ZMAX=Z(K)
880   IF Z(K)<ZMIN THEN ZMIN=Z(K)
890   NEXT K
900   RANGE=ZMAX-ZMIN
910   IF Z(ICOL)<>UPLIM THEN DRANGE=Z(ICOL)-ENG ELSE DRANGE=99
920   COL(CHOICE)=COL(CHOICE)+1
930   IF OUTVAL$="CC" THEN ENG=CHOICE
940   IF OUTVAL$="R" THEN ENG=RANGE
950   IF OUTVAL$="B" THEN ENG=BAD
960   IF OUTVAL$="D" THEN ENG=DRANGE
970   IF ENG=UPLIM AND OUTVAL$="E" GOTO 1030
980   PRINT #2,X,Y,ENG
990   IF X=-180 AND PHEND <> 180 THEN PRINT #2,-X,Y,ENG
1000  IF Y=-180 AND PSEND <> 180 THEN PRINT #2,X,-Y,ENG
1010  IF X=-180 AND Y=-180 AND PHEND<>180 AND PSEND<>180 THEN PRINT
      #2,-X, -Y,ENG
1020  M(ROW,COL)=ENG
1030  NEXT X
1040  NEXT Y
1050  CLOSE #1:CLOSE #2
1060  '
1070  'End of big loop.  Start to report
1080  '
1090  PRINT BADCOUNT; " ENERGIES WERE OUTSIDE THE RANGE OF ";LO;" TO
      ";UPLIM
1100  FOR K=1 TO NCOL
1110  PRINT COL(K);" ENERGIES TAKEN FROM COLUMN ";K
1120  NEXT K
1130  '
1140  'Section to make simple plots
```

```
1150 '
1160 INPUT "Do you wish to send Phi,Psi grids to disk (Y/N)[N]";A$
1170 IF LEFT$(A$,1)="Y" OR LEFT$(A$,1)="y" THEN A$="Y" ELSE A$="N"
1180 IF A$="N" GOTO 1220
1190 INPUT "Filename ";FO1$
1200 IF FO1$=FO$ THEN PRINT CHR$(7);"Would Overwrite 1st file!":GOTO
     1190
1210 OPEN "O",1,FO1$
1220 FOR Y=ROW TO 1 STEP -1
1230 FOR X=1 TO COL
1240 PRINT USING "###";M(Y,X);
1250 IF A$="Y" THEN PRINT #1, USING "###";M(Y,X);
1260 NEXT X
1270 PRINT USING "#####";PSBEG+PSDEL*(Y-1)
1280 IF A$="Y" THEN PRINT#1, USING "#####";PSBEG+PSDEL*(Y-1)
1290 NEXT Y
1300 FOR X=PHBEG TO PHEND STEP PHDEL
1310 IF X<0 THEN FLAG$="-" ELSE FLAG$=" "
1320 PRINT "  ";FLAG$;
1330 IF A$="Y" THEN PRINT#1,"  ";FLAG$;
1340 NEXT X
1350 PRINT
1360 IF A$="Y" THEN PRINT#1,""
1370 FOR K=3 TO 1 STEP -1
1380 FOR X=PHBEG TO PHEND STEP PHDEL
1390 S$=STR$(ABS(X))
1400 PRINT "  ";LEFT$(RIGHT$(S$,K),1);
1410 IF A$="Y" THEN PRINT#1, "  ";LEFT$(RIGHT$(S$,K),1);
1420 NEXT X
1430 PRINT
1440 IF A$="Y" THEN PRINT#1,""
1450 NEXT K
1460 CLOSE
1470 END
```

Literature Cited

1. Joshi, N.; Rao, V.S.R. *Biopolymers* 1979, 18, 2993-3004.
2. Pensak, D. A.; French, A. D. *Carbohydr. Res.* 1979, 89, 1-10.
3. French, A. D. *Carbohydr. Res.* 1989, 188, 206-211.
4. French, A. D.; Murphy, V. G. *Polymer* 1977, 18, 489-494.
5. Melberg, S.; Rasmussen, K. *Carbohydr. Res.* 1979, 71, 25-34.
6. Homans, S. W.; Dwek, R. A.; Rademacher, T. W. *Biochemistry*, 1987, 26, 6571-6578.
7. French, A. D. *Biopolymers* 1988, 27, 1519-1525.
8. Ha, S. N.; Madsen, L. J.; Brady, J. W. *Biopolymers* 1988, 27, 1927-1952.
9. Tran, V.; Buleon, A.; Imberty, A.; Pérez, S. *Biopolymers* 1989, 28, 679-690.
10. Allinger, N. L. *J. Am. Chem. Soc.* 1977, 99, 8127-34.
11. Nørskov-Lauritsen, L.; Allinger, N. L. *J. Computational Chem.* 1984, 326-35.
12. Jeffrey, G. A.; Taylor, R. *J. Computational Chem.* 1980, 1, 99-109.
13. Jeffrey, G. A.; Park, Y. J. *Carbohydr. Res.* 1979, 74, 1-5.
14. Kroon-Batenburg, L. M. J.; Kroon, J.; Northolt, M. G. *Polymer*, 1986, 27, 290-292.
15. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Saminathan, S.; Karplus, M. *J. Computational Chem.* 1983, 4, 187-217.

16. French, A. D. In <u>Cellulose and Wood - Chemistry and Technology</u>; Schuerch, C., Ed.; Wiley: New York, 1989, p. 103-118.
17. Burkert, U.; Allinger, N.L. <u>J. Computational Chem</u>. 1982, <u>3</u>, 40-46.
18. Thogersen, H.; Lemieux, R. U.; Bock, K.; Mayer, B. <u>Can. J.Chem</u>. 1982, <u>60</u> 44-57.
19. Tvaroška, I.; Pérez, S. <u>Carbohydr. Res</u>. 1986, <u>149</u>, 389-410.
20. Gibson, K. D.; Scheraga, H. A. <u>Biopolymers</u> 1966, <u>4</u>, 709-712.
21. Gelin, B. R.; Karplus, M. <u>J. Am. Chem. Soc</u>. 1975, <u>97</u>, 6996-7006.
22. Ferretti, V.; Bertolasi, B.; Gilli, G.; Accorsi, C. A. <u>Acta Crystallogr</u>. 1984, <u>C40</u>, 531-535.
23. Osawa, E. <u>QCPE Bull</u>. 1986, <u>6</u>, 76-79.
24. Clarke, T. <u>A Handbook of Computational Chemistry</u>; John Wiley & Sons: New York, 1985.
25. Burkert, U.; Allinger, N. L. <u>Molecular Mechanics</u>, ACS Monograph 177, American Chemical Society: Washington, DC, 1982.
26. Ragazzi, M.; Ferro, D. R.; Perly, B.; Torri, G.; Casu, B.; Sinay, P.; Petitou, M.; Choay, J. <u>Carbohydr. Res</u>. 1987, <u>165</u>, c1-c5.
27. Sundaralingam, M. <u>Biopolymers</u> 1968, <u>6</u>, 189-213.
28. Nishida, Y.; Hori, H.; Ohrui, H.; Meguro, H. <u>Carbohydr. Res</u>. 1987, <u>170</u>, 106-111.
29. MMP2(85) Users Manual. Quantum Chemistry Program Exchange, Bloomington, Indiana (1987) p 12.
30. Allinger, N. L.; Yuh, Y. H.; Lii, J-H. <u>J. Am. Chem. Soc</u>. 1989, <u>111</u>, 8551-8566.

Chapter 13

Conformational Flexibility of Sucrose

Static and Dynamical Modeling

V. H. Tran[1] and J. W. Brady[2]

[1]**Institut National de la Recherche Agronomique, B.P. 527, 44026, Nantes, France**
[2]**Department of Food Science, Cornell University, Ithaca, NY 14853–7201**

In analyzing the conformational properties of carbohydrates, it has often been the practice to consider individual monomer rings to be rigid units, with the only molecular flexibility lying in torsional rotations about the linkage bonds. However, conformational energy maps prepared using this assumption can be quite misleading in deciding which conformations are energetically allowed, and almost useless in understanding the likely motions of disaccharides, such as might be computed in molecular dynamics simulations. Molecular flexibility can be incorporated into conformational energy maps by preparing relaxed or adiabatic energy maps. However, a number of operational difficulties complicate the calculation of relaxed energy maps. This paper discusses some of these practical problems which arise in the preparation of relaxed conformational energy maps for complex dimers like disaccharides, and how these problems complicate the physical interpretation of such maps. The importance of molecular flexibility is illustrated through the application of flexible conformational energy mapping and molecular dynamics simulations to sucrose.

A particularly common application of computer modeling to carbohydrate molecules is the use of calculated conformational energy maps for disaccharide glycosidic linkages as a tool in understanding oligosaccharide conformational structures (1). Although past studies have often treated the internal degrees of

freedom of disaccharides as rigid, except for the glycosidic
torsion angles themselves, it is now widely accepted that the ring
forms of sugars are not completely rigid, and that the internal
flexibility of these rings can strongly couple with glycosidic
rotations to substantially affect calculated conformational energy
maps (2-4). For this reason, it can be worthwhile to calculate
relaxed or adiabatic energy maps in which the molecular structure
is relaxed by energy minimization at each point on the glycosidic
map in order to relieve internal strains (5). It is also possible
to use molecular dynamics simulations (6) to directly model the
motions that a disaccharide molecule might undergo, and to relate
these motions to the calculated conformational energy map.
Recently, we performed such a series of calculations for the mixed
disaccharide molecule sucrose using the general molecular
mechanics program CHARMM (7). A complete description of this work
has been presented elsewhere (8,9), and we shall describe here
only the general strategy for such studies, using the sucrose work
as an example. In particular, we will focus on technical
difficulties which arise in these calculations, including possible
nonphysical artifacts, and examine the distinction between
adiabatic and relaxed maps and their physical interpretation.

Adiabatic and Relaxed Energy Maps

Ramachandran-like conformational energy maps (1) represent
disaccharide energies as depending only on two degrees of freedom,
ϕ and ψ, when actually there are a very large number of possible
combinations of values for the many internal degrees of freedom on
which the total energy depends, such as the various bond lengths
and angles and the other internal torsion angles. Flexible
monomer conformational energy maps are obtained when the molecular
geometry is relaxed by energy minimization at each value of ϕ and
ψ on a regular grid. Any such conformational energy map in which
energy minimization has been used to relieve internal strains can
be referred to as a "relaxed" energy map. However, because of the
high dimensionality of the structural problem (due to the large
number of bonds, angles, hydroxyl orientations, etc.) there will
in general be a very large number of structures which are local
minima for the empirical energy function. This situation can be
visualized by noting that for any given value of ϕ and ψ there are
many possible arrangements of all of the hydroxyl groups and
primary alcohols in the monomers. For any of these local minimum-
energy structures, any small change in geometry initially produces
an increase in energy. However, only one of these structures will
have the lowest possible energy consistent with the specified
values of the glycosidic torsion angles. A conformational energy
map in which only these lowest energy values are plotted as a
function of the glycosidic angles will be referred to in this
paper as an "adiabatic" energy map. The single lowest energy
structure possible for the molecule, including allowing ϕ and ψ to
vary, is referred to as the global minimum energy structure. The
difficult problem of finding these lowest energy structures is
often referred to as the "multiple-minimum" problem, which arises
in any minimization of a function of many variables, and which has

contributed to frustrating efforts to predict biopolymer
structures from sequence using molecular modeling techniques (10).

Calculating a meaningful relaxed energy map for a disaccharide
molecule is not necessarily straightforward due to the
complications which arise from the multiple-minimum problem. No
general mathematical solution to this problem exists. As a
result, energy minimization from any given starting structure for
a given molecule will in general produce only the nearest local
minimum-energy structure not separated from the starting point by
any significant barriers. The conformation and its energy thus
produced will therefore depend strongly upon the starting
structure. In practice, the results may also depend on the
minimization algorithm. These difficulties imply that a relaxed
map produced by simply rotating an arbitrary disaccharide
structure about the glycosidic torsions to various values of ϕ and
ψ, followed by minimization, may produce unusual and physically
unlikely structures, even if the rotations are done in a
systematic fashion. This difficulty arises because some of the
starting structures in such a regular procedure will themselves
almost certainly be physically unlikely.

Overcoming this problem would require performing minimizations
for an extremely large number of starting structures at each value
of ϕ and ψ to find the conformation with the lowest possible
energy. In most cases it is not practical to make such a thorough
search due to the limitations of available computer time.
Furthermore, for reasons discussed below, an adiabatic map may in
some cases not be the most useful of the possible relaxed energy
maps. The most physically meaningful map would be one in which
the relaxation by minimization occurred in a manner as close as
possible to the actual physical process being represented. If one
started from the global minimum-energy structure (assuming that it
is possible to identify that structure at the beginning of the
study!), then perturbing the conformation to values of ϕ and ψ
that are close by in configuration space should produce only small
changes in the other conformational variables, thus perhaps
reducing the number of possibilities that would have to be
investigated by separate minimizations from different starting
structures. As one moved further away from this minimum-energy
structure however, the other internal degrees of freedom would
presumably change by varying amounts, and at some point an
arbitrary distance away in the (ϕ,ψ) map, some other combination
of the various internal coordinates than that found in the global
minimum-energy structure may produce the lowest energy at that
(ϕ,ψ) gridpoint when minimized. A map that ignored this shift and
continued to plot the energies of those minimized structures which
had for the most part the same general conformation for these
other internal variables as the starting minimum, regardless of
whether they were the absolute lowest values at the grid point,
would then be a "local" relaxed surface, or "partial" relaxed
surface. Such a surface might represent the energy surface over
which the molecule would move during a dynamical structural
fluctuation if there is no relaxation to any possible lower energy
forms.

Calculation of Relaxed and Adiabatic Maps for Sucrose

Very few calculations of adiabatic energy maps for disaccharides
coupled with molecular dynamics simulations have been undertaken,
partially as a result of the great expense and labor involved in
such studies. One molecule which has been studied by both of
these types of molecular mechanics calculations is sucrose.
Sucrose was selected for study for several reasons. This
molecule, which is the dimer of a pyranoid glucose ring and a β-
furanoid fructose ring, connected by an $\alpha[1\rightarrow2]$ linkage, offers the
possibility of studying the interplay of flexibility between the
furanoid ring and the more rigid six-membered ring. It is a
particularly good molecule for theoretical modeling since it has
been extensively studied by a variety of experimental techniques.
The crystal structure of this molecule has been determined by both
x-ray (11) and neutron diffraction (12) analysis. Several NMR
studies of the conformation of sucrose have been reported in the
literature for molecules in both aqueous (13,14) and DMSO (15)
solution. Mathlouthi et al. (16) have studied sucrose using Raman
spectroscopy, and some previous molecular mechanics work is also
available for comparison (13,17).
      Preliminary investigation revealed that sucrose is
significantly restricted in its conformational fluctuations by
steric crowding, which is consistent with previous molecular
mechanics studies (13,17). If the internal geometry of the rings
and sidechains is not allowed to relax as ϕ and ψ are changed,
then the crystal geometry is essentially the only possible
conformation, due to steric overlaps. Having thus used available
experimental information in this particular case to identify a
reasonable low energy conformation, this conformation could then
be used to map out a family of related low energy conformations at
nearby values of the angles ϕ and ψ by using this crystal
structure as a starting geometry, and rotating the structure to
various adjacent values of ϕ and ψ ($\phi\pm\Delta,\psi\pm\Delta$) and minimizing. This
procedure was then extended to the next nearest (ϕ,ψ) grid points,
using the newly minimized structures found in the previous round
as starting structures, and extended throughout the entire
configuration space if desired. As one moves farther away from
the original minimum structure in quasi-radial directions in (ϕ,ψ)
space, it becomes increasing more likely that some other
arrangement of the various internal coordinates might actually
lead to the lowest possible energy value at that (ϕ,ψ) grid point.
Because rearrangement to any such lower energy structure could
involve an initial increase in energy, the minimization searches
might thus fail to find these alternate conformations.
      After mapping out a sufficient region of (ϕ,ψ) space using
this quasi-radial approach of exploration ("sufficient" being
defined as including all nearby conformers within some specified
high-energy contour), alternate structures with lower minimized
energy might be sought by arbitrarily making large changes in
various other internal degrees of freedom followed by a new cycle
of exhaustive minimization. Such essentially random searching
must be attempted not only in the vicinity of the already located
minima but also in other regions of configuration space where

stable minima might be possible, and particularly in the region of possible barriers between conformations. Any stable minima in (ϕ,ψ) space that are identified must then also be used as the seed structures for mapping out the relaxed energy of structures nearby in (ϕ,ψ) space. If these newly identified minimum-energy structures have basically the same arrangement of all of their internal degrees of freedom as the previously identified low energy conformer, then their local energy surface would simply be an extension of that surface into the new region of configuration space; under ideal or fortunate circumstances, the extension of the former partial or local surface throughout the full configuration space would have identified these other local minima as well.

As a concrete example of this procedure, consider the case of sucrose. With the CHARMM-like potential energy function of Ha et al. (18), which probably overemphasizes intramolecular hydrogen bonding in vacuum calculations, the hydroxyl groups around pyranoid rings all tend to line up in circuits around the perimeter of the ring as each hydroxyl dipole attempts to hydrogen bond to its neighbor. These circuits can "point" in one of two directions around the ring (Figure 1), either in the same direction as the numbering of the ring atoms (referred to here as clockwise), or in the opposite direction (which we will call counterclockwise). While in general these two conformational patterns are close in energy, they are structurally distinct and thus lead to two separate "families" of possible conformational arrangements for each value ϕ and ψ for sucrose in vacuum. Transitions between these two families do not occur in vacuum since the activation energy is large. However, intuition and the results of preliminary MD simulations of both glucose (19) and sucrose (unpublished work) in aqueous solution indicate that these networks become disrupted by hydrogen bonding to aqueous solvent.

Three low-energy (ϕ,ψ) regions were identified for sucrose (called A, B, and C; see Figure 2). For two of these regions, complete and separate partial or local maps could be developed which primarily differed only in the direction of the hydrogen bond circuits in the glucose residue. For the third general conformational region, only one of these arrangements was of low energy, since for that family of (ϕ,ψ) combinations, the opposite hydrogen bond circuit had one fewer hydrogen bond. The global minimum energy structure, which has the counterclockwise hydrogen bond pattern, and its associated local energy map, were labeled S1. The minimum energy conformation in the same region (A) of (ϕ,ψ) space but with the opposite hydrogen bond orientation was called S2, along with the local relaxed energy surface associated with it. The crystal structure has a hydrogen bond pattern most closely resembling the higher energy S2 geometry due to intermolecular crystal contacts. The clockwise geometry in the "B" region was called S3, and the conformation in this same region with the opposite, counterclockwise pattern was called S4. S5 was used to designate the sole low energy conformation in the C region, which is counterclockwise, along with its relaxed energy surface. Five separate local energy maps were developed, three of which (S1, S4, and S5) could be combined to give a complete

A   **B**
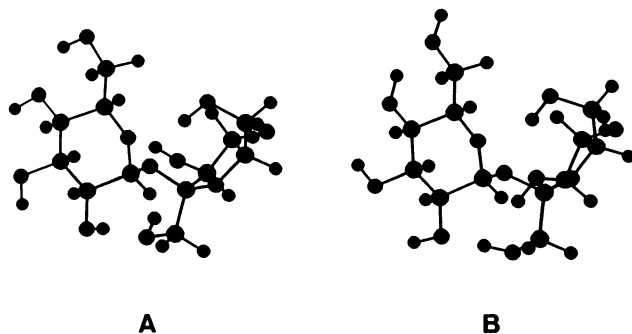
Figure 1. Two different general orientations of the hydroxyl groups around the six-membered ring in sucrose. (A) The counterclockwise orientation of the global minimum S1. (B) The clockwise orientation for the local minimum S2.
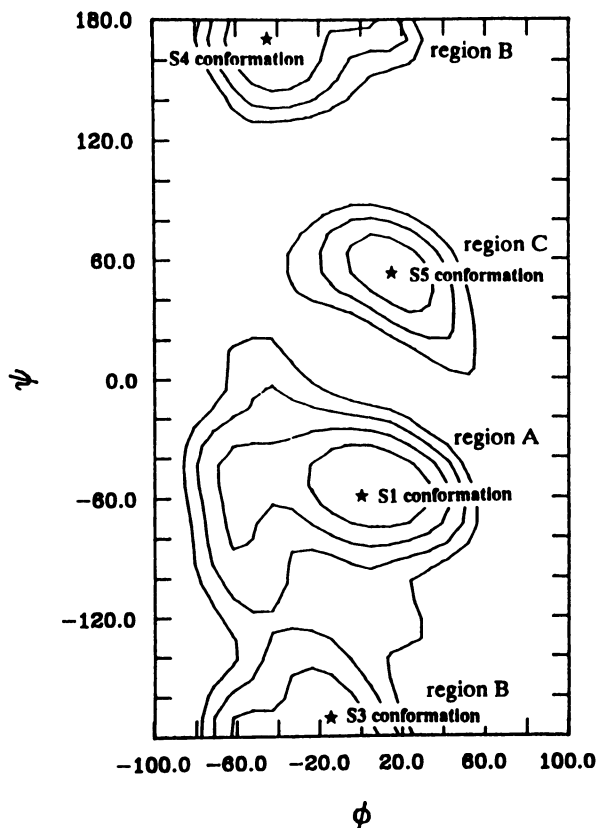
Figure 2. The calculated adiabatic energy map for sucrose. Contours are indicated at 2, 4, 6, and 8 kcal/mol above the global S1 minimum. The stars refer to the various minima calculated with the present potential energy function.

relaxed surface, as could be done with the remaining two, which would produce a surface with only two major low-energy regions. A complete adiabatic map for sucrose with this potential energy function could then be synthesized by taking the lowest energy structure for each (ϕ,ψ) combination, regardless of from which surface it comes, and plotting these together, as in Figure 2.

Any such 2-dimensional energy map is a cross section through the complex, multidimensional configurational hyperspace of a moderately large molecule like sucrose, and obscures many important structural details. Several competing and sometimes mutually exclusive inter-ring hydrogen bonds are particularly important in determining the relative stabilities of the local minima. As an example, Figure 5a illustrates those regions on the local ("partial") map around the S4 minimum where several hydrogen bonds are possible. These patterns would be different for the S3 map in the same (ϕ,ψ) region (8).

Because it represents the lowest energy structure possible for each combination of ϕ and ψ, the adiabatic surface should ideally be the surface over which the disaccharide moves as it undergoes conformational fluctuations and transitions. As such, the adiabatic surface would be essential to the interpretation of molecular dynamics simulations of disaccharides. However, while thermodynamics guarantees that given sufficient time, a system will inevitably relax to the lowest (free) energy state, during the course of a conformational fluctuation by a disaccharide, the molecule may pass through a given region of (ϕ,ψ) space in a very brief time (on the order of a few tens of femtoseconds in some cases). If the time required for the system to relax to the adiabatic surface is long compared to this timescale for dynamical fluctuations, then the molecule will probably not relax to the adiabatic surface, but instead will follow the local relaxed surface. Such a case exists for disaccharides in vacuum using the Ha et al. energy function (18), since the transition from one hydrogen-bonded pattern to the other (such as from clockwise to counterclockwise) initially involves the simultaneous breaking of several hydrogen bonds before the opposite hydrogen-bonding pattern can be established. This situation is an artificial or nonphysical one, however, since it would not arise in aqueous solution, where no hydrogen bonds would need to be broken during a hydroxyl rotation, as alternate hydrogen bond partners would be present in the water molecules themselves. (In fact, in aqueous solution, such circuits of hydrogen bonds might not form at all, since the hydroxyl groups could satisfy their hydrogen bonding requirements by bonding with the solvent, without the entropic constraints of the intramolecularly bonded ring (19)). In the case of sucrose, the partial, local surfaces are topologically very similar, so that there is little difference between them and the actual adiabatic surface.

Molecular Dynamics Simulations of Sucrose

Preparation of the adiabatic energy map for sucrose revealed that while only one conformation is allowed for this molecule if it is kept rigid, when flexibility is taken into account, several

relatively low energy structures are possible. In order to
examine the actual dynamical behavior of these conformers, and to
study how transitions between these structures might occur, a
series of molecular dynamics simulations were also calculated for
sucrose using the CHARMM program (7), and were interpreted in
terms of the relaxed energy surfaces.

Trajectories with different initial velocity assignments were
computed for each of the local minimum energy structures located
in the three low energy regions of (ϕ, ψ) space (A, B, and C; see
Figure 2). The calculations were done in the microcanonical
ensemble at a temperature of 300K $\pm$ 5K. Energy was well conserved
throughout the trajectories, and no overall drifts in molecular
temperature were observed. Small ensembles of trajectories (12
for S1 and 6 each for the other minima) were calculated for the
averaging of system properties. Each trajectory was equilibrated
by velocity reassignments during an initial period of 20ps,
followed by another 20ps of dynamics used for data collection.
All possible non-bonded interactions in these trajectories were
included in the energy calculations; that is, no long-range
truncations were applied. The dielectric constant for these
calculations was taken to be unity, as was the case in the energy
mapping study.

In general, these trajectories were very well described by the
relaxed maps. Figure 3 shows an example of a typical trajectory
superimposed onto the corresponding partial or local relaxed
energy map; in this example, the adiabatic surface would also
provide an equally good representation. As can be seen, the
asymmetries in the motions and the magnitudes of the structural
fluctuations are easily understood in terms of the energy map, and
would be incomprehensible using a rigid rotation energy analysis.

From a direct examination of the trajectories it is possible
to make estimates for the timescales of the various dynamical
processes which occur in these molecules in the course of normal
thermal fluctuations. Figure 4 illustrates the variation with
time of the dihedral angle C5g-C6g-O6g-H'6g as calculated from two
different molecular dynamics simulations. As can be seen from
this figure, the dynamical behavior can vary somewhat with the
simulation conditions; in one simulation this internal coordinate
exhibited frequent and very fast changes in conformation with
durations of less than 0.1 ps, while in the other case two of
these conformers were more metastable, with lifetimes of about 1
ps, and with a steady oscillation between the two states and
longer times required to accomplish the transitions. Figure 5
illustrates the competition between two possible hydrogen bonds
consistent with the S4 family of conformations. Figure 5a is a
superposition upon the local relaxed energy map of the regions
where the H'2g...O3f and H'2g...O6f hydrogen bonds are possible.
As can be seen from this figure, these two hydrogen bonds are
mutually exclusive, and as the H'2g hydrogen atom makes a hydrogen
bond with one of these acceptors, it must break its hydrogen bond
with the other. Figures 5b and 5c show that the exchange expected
from the local energy map does indeed occur. Three exchanges were
observed in this simulation; from an ensemble of such simulations,
or from a single simulation of much greater duration, it would be
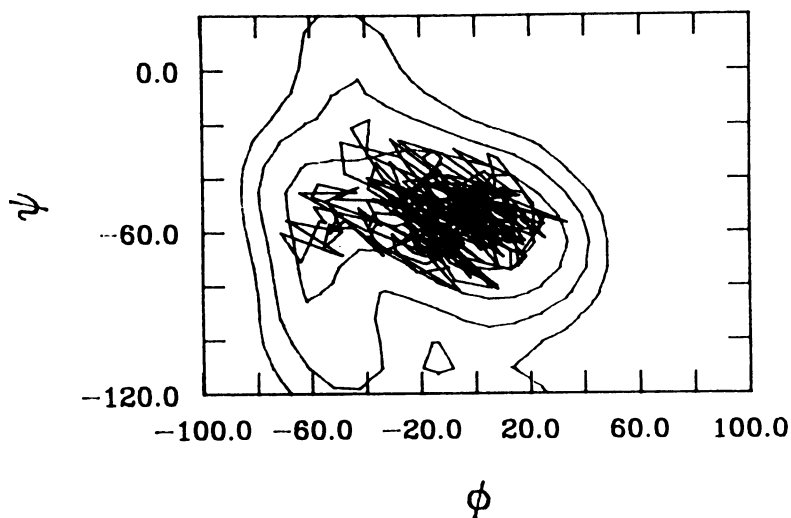
Figure 3.   An example of a sucrose trajectory in (ϕ, ψ) space superimposed upon the local relaxed or partial conformational energy map (S2).
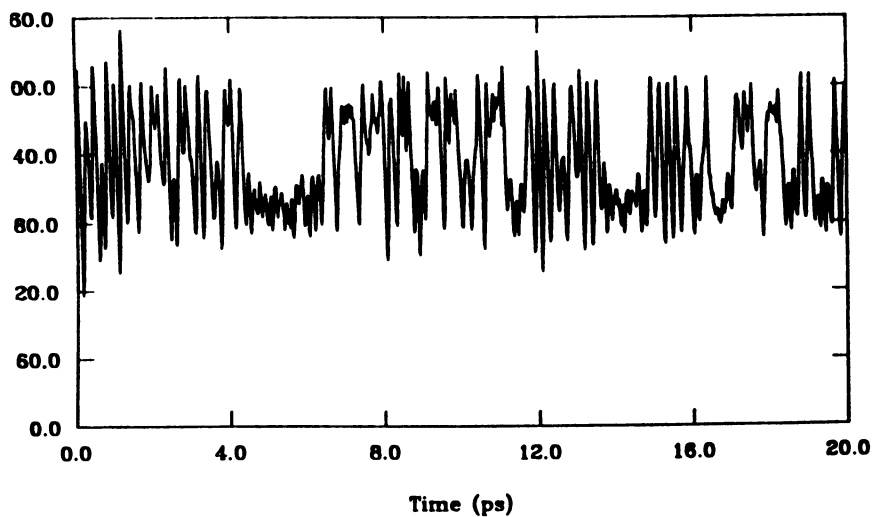
Figure 4a. Fast oscillations between the three stable orientations $(60°, -60°, 180°)$.
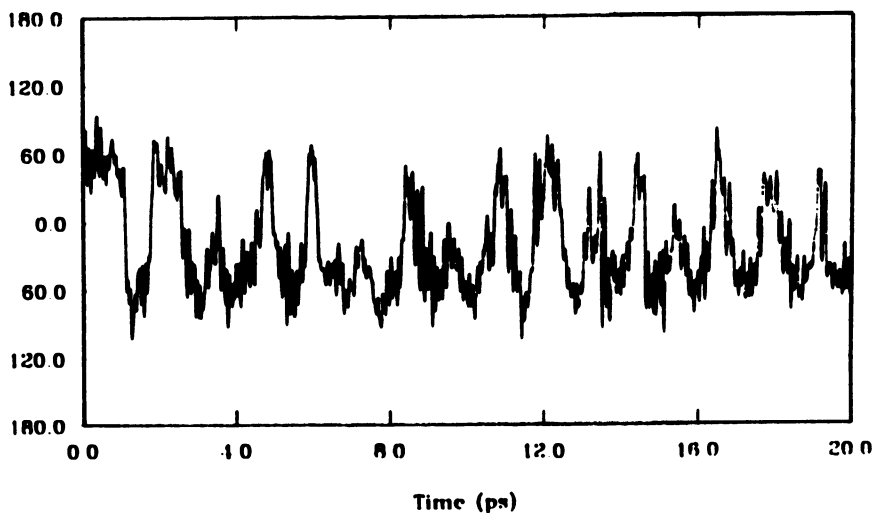
Figure 4b. Slower oscillation between longer-lived metastable conformers at 60° and –60°.
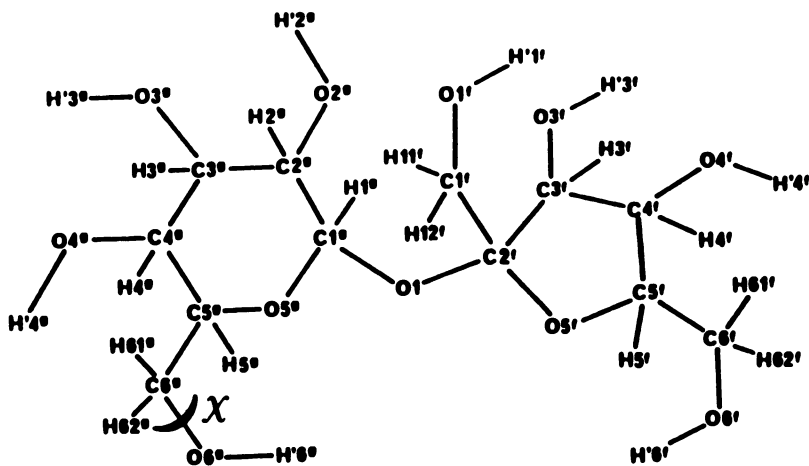
Figure 4c. An example of different timescales for motions in the dihedral angle C5g-C6g-O6g-H′6g (indicated in the molecular sketch) as calculated in two separate molecular dynamics simulations.

S4 map
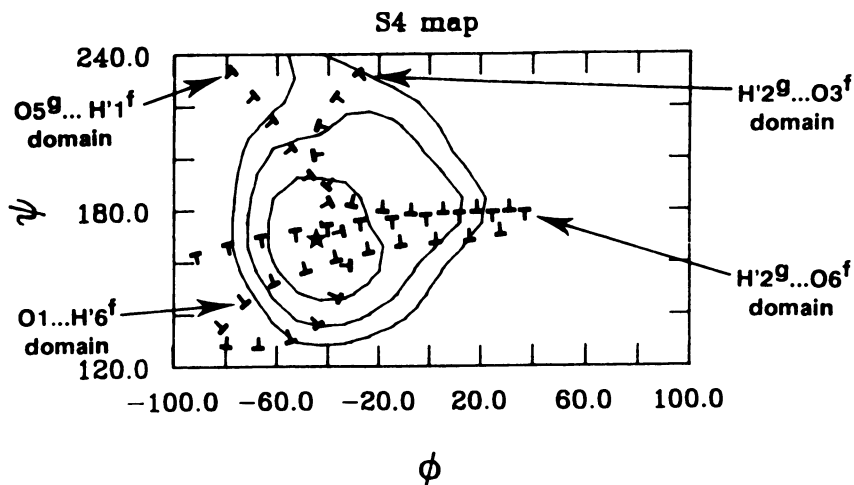
Figure 5a. An example of a partial energy map, the local relaxed map for the S4 family of conformations. Contours are indicated at 4, 6, and 8 kcal/mol above the global S1 minimum, which does not appear on this map. The dashed lines surround the different inter-residue hydrogen bond domains (with a cutoff criterion of 2.05Å for the O . . . H distance), with the tic marks on the dashes pointing toward the region where the given hydrogen bond is allowed.
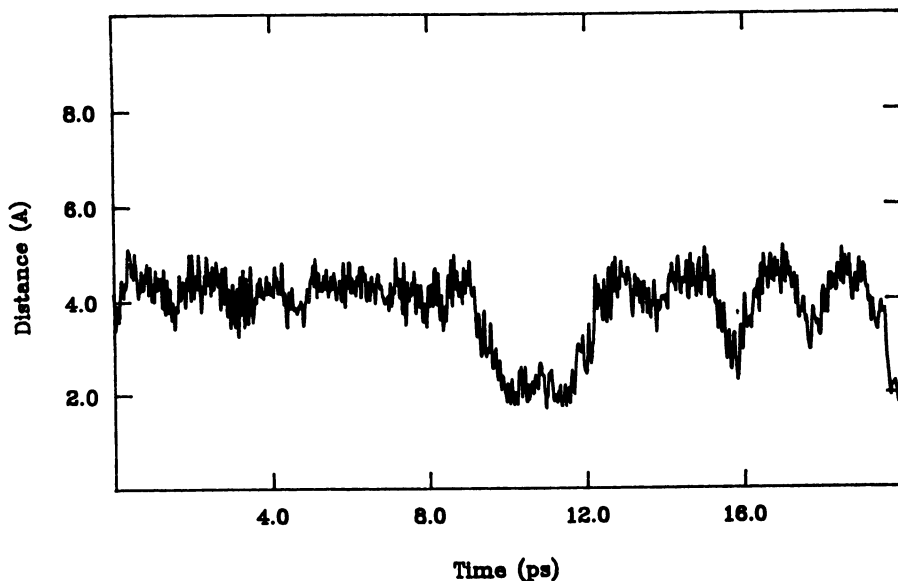
Figure 5b. History of the fluctuations in the hydrogen bond distance H'2g . . . O3f as calculated from a molecular dynamics trajectory.
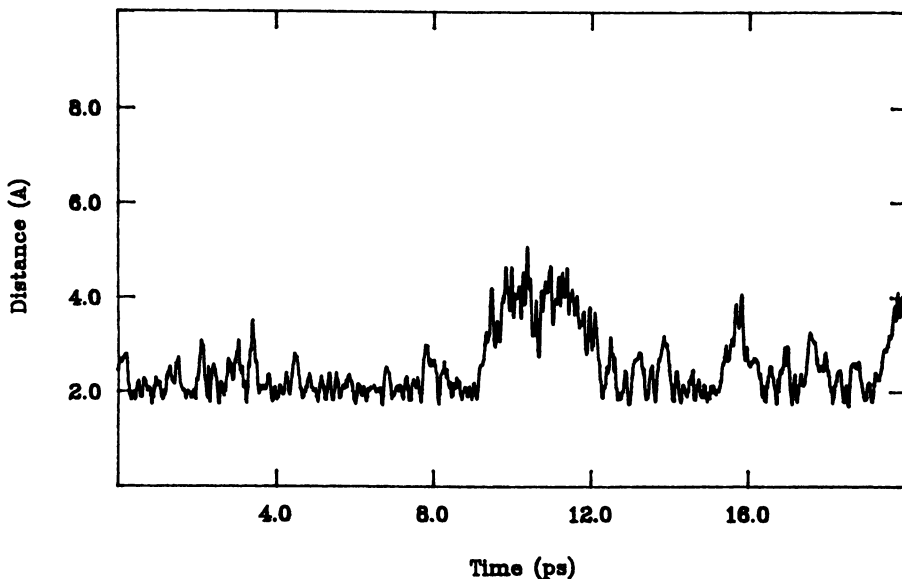
Figure 5c. History of the H'2g . . . O6f hydrogen bond distance, as calculated from the same trajectory.

possible to estimate a transition frequency. From Figure 5 it can be seen that the actual transition takes place rather slowly, taking as much as a picosecond to complete each time it occurs.

Transitions between major conformational wells for the sucrose molecule are rare in vacuum on the picosecond timescale, although one transition was observed in these simulations. Figure 6 illustrates a trajectory which began in the S3 conformation and underwent a transition to the lower energy S2 conformation midway through the simulation. This transition, the only one observed in a total of 720 ps of dynamics in 36 separate trajectories, occurred between the two forms predicted by Christofides and Davies (15) to occur in DMSO. As can be seen, the course of the transition is well described by the topology of the energy surface, and the approximate path could be easily predicted in advance from this map. In the figure, the transition is shown superimposed on the combination of the S2 and S3 local maps, rather than the adiabatic surface. While the topological differences between these two surfaces are again small, it is worthwhile to note that the S2 structure is not the lowest energy form in the central well, and that the molecule did not at any time during or after this transition relax to this S1 conformer, with its opposite pattern for the pyranoid hydrogen bond circuit. Neither did any of the trajectories started in the S2 conformation make a transition to the S1 form. This stability indicates that the timescale for such a relaxation to the true adiabatic surface is indeed much longer than the timescale for conformational fluctuations and transitions for this molecule in vacuum.
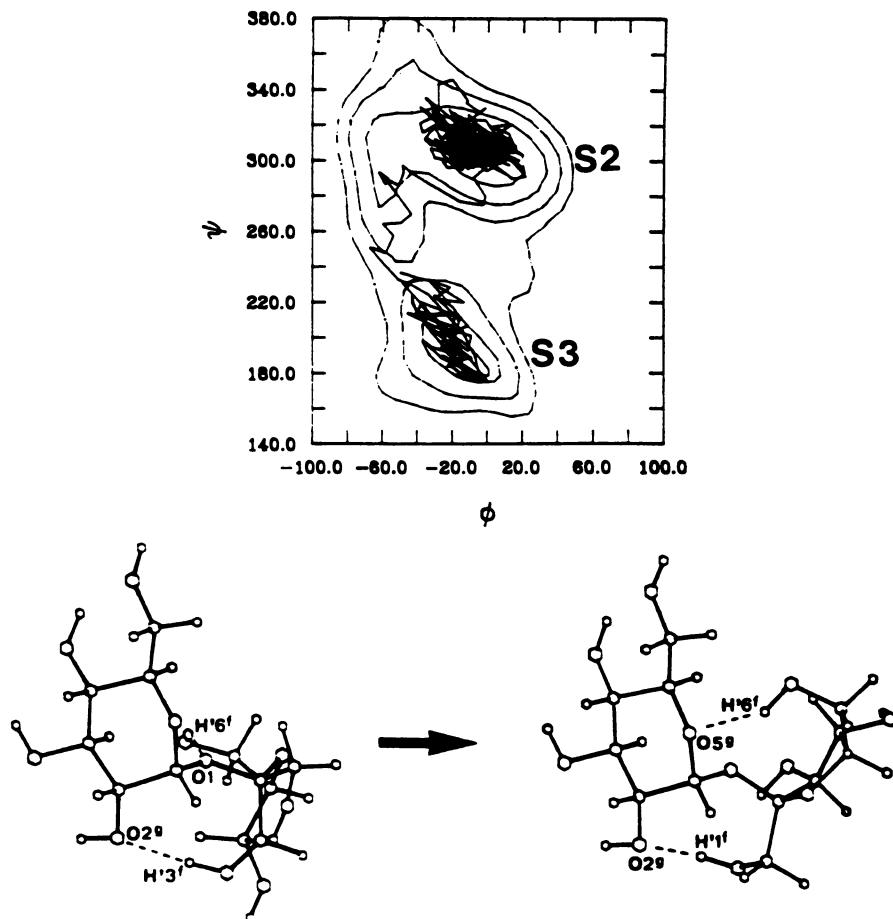
Figure 6. A molecular dynamics trajectory which underwent a transition from the S3 to the S2 conformation, superimposed on the relaxed surface created by combining the S2 and S3 surfaces. At the bottom are illustrations of the molecular geometry in the beginning and final conformations.

Conclusions

A complete understanding of the structures and motions of carbohydrates will require including flexibility in the conformational energy descriptions of disaccharides which are frequently used to model polymers built from sugar monomers. In particular, understanding the dynamics of disaccharides will necessitate the production of relaxed energy maps of one form or another which take into account the internal structural readjustments in monomer units as the glycosidic angles change. Because intramolecular reorganization rates are frequently slow compared to the dynamics timescale, the most useful energy surface

for describing disaccharide conformations might not be the adiabatic map but a relaxed surface in which the minimizations progress outward from identified local minima in a quasi-radial manner, proceeding from structure to structure in roughly the same sequence as might occur during a real conformational fluctuation. It should also be remembered, as illustrated by the artifacts in the sucrose studies, that vacuum calculations could potentially be inadequate. Most carbohydrates are found in aqueous environments, and the influence of solvent upon conformation might be significant. Since physical properties are determined by free energies rather than mechanical energies, it may be necessary in some cases to compute (ϕ,ψ) maps of the potential of mean force from solution MD studies, which would include the solvent contribution.

Literature Cited

1.   Brant, D.A. Ann. Rev. Biophys. Bioeng. 1972,1,369.
2.   French, A.D.; Murphy, V.G. Carbohydr. Res. 1973,27,391; Polymer 1977,18,489.
3.   Ha, S.N.; Madsen, L.J.; Brady, J.W. Biopolymers 1988,27,1927.
4.   Tran, V.; Buleon, A.; Imberty, A.; Perez, S. Biopolymers 1989,28,679.
5.   Gelin, B.R.; Karplus, M. J. Am. Chem. Soc. 1975,97,6996.
6.   Brooks, C.L.; Karplus, M.; Pettitt, B.M. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics; Advances in Chemical Physics, Wiley-Interscience: New York, 1988, Vol. LXXI.
7.   Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. J. Comput. Chem. 1983,4,187.
8.   Tran, V.H.; Brady, J.W. Biopolymers in press.
9.   Tran, V.H.; Brady, J.W. Biopolymers in press.
10. McCammon, J.A.; Harvey, S.C. Dynamics of Proteins and Nucleic Acids; Cambridge University Press, Cambridge, 1987.
11. Brown, G.M.; Levy, H.A. Acta Crys. 1973,B29,790.
12. Hanson, J.C.; Sieker, L.C.; Jensen, L.H. Acta Crys. 1973,B29,797.
13. Bock, K.; Lemieux, R.U. Carbohydr. Res. 1982,100,63.
14. McCain, D.C.; Markley, J.C. J. Am. Chem. Soc. 1986,108,4259.
15. Davies, D.B.; Christofides, J.C. Carbohydr. Res. 1987,163,269.
16. Mathlouthi, M.; Luu, C.; Meffroy-Biget, A.M.; Luu, D.V. Carbohydr. Res. 1980,81,213.
17. Ferretti, V.; Bertolasi, V.; Gilli, G. Acta Cryst. 1984,C40,531.
18. Ha, S.N.; Giammona, A.; Field, M.; Brady, J.W. Carbohydr. Res. 1988,180,207.
19. Brady, J.W. J. Am. Chem. Soc. 1989,111,5155.

Chapter 14

Molecular Modeling Studies on Unbranched Complex Carbohydrates

Application to a Linkage Region Fragment from Connective Tissue Proteoglycans

N. Rama Krishna, Bo-Young Choe, and Stephen C. Harvey

Departments of Biochemistry and Physics, and the Comprehensive Cancer Center, University of Alabama, Birmingham, AL 35294

A general molecular modeling methodology particularly suitable for unbranched complex carbohydrates was described. This methodology employed molecular dynamics (MD) and energy minimization (EM) procedures together with inter-residue spatial constraints across the linkages derived from 2D-NOESY spectroscopy. The first step in this methodology is the generation of a wide variety of starting conformations that span the (ϕ,ψ) space for each linkage. In the present study, for each linkage, nine starting conformations that span the (ϕ,ψ) space in a resonable manner were constructed using the torsion angles ϕ and ψ corresponding to the gauche+, gauche-, and trans configurations across each of the two bonds constituting the linkage. These conformations were subjected to a combined MD/EM refinement using the NOESY derived constraints as pseudoenergy functions. Families of conformations for the whole molecule were then constructed from the structures derived for each linkage. This procedure was demonstrated on a fragment from the carbohydrate-protein linkage region of connective tissue proteoglycans.

Connective tissue proteoglycans are predominantly composed of glycosaminoglycans, some oligosaccharides and a core protein (1). The glycosaminoglycans are covalently attached to the core protein through a unique linkage region composed of a short oligosaccharide. During the course of our investigation on the various oligosaccharides derived from the proteoglycans, we have developed a general molecular modeling methodology based on NMR and molecular dynamics (MD) and energy minimization (EM). While the methodology can be adapted for any complex carbohydrate, unbranched oligosaccharides particularly lend themselves to a simplified analysis by this methodology since the conformation for the whole oligosaccha-

ride can be constructed from the conformations of the individual
modules composed of two residues linked together (a disaccharide or
a monosaccharide linked to an amino acid). Implicit in this modular
analysis approach is the assumption that interactions between non-
neighbouring residues are negligible. Such a situation is commonly
realized at the oligosaccharide level in connective tissue proteo-
glycans where 1->3 and 1->4 linkages are frequently encountered.
We will illustrate our methodology using O-β-D-galactopyranosyl
-(1->4)-O-β-D-xylopyranosyl-(1->O)-L-serine (GXS), a fragment from
the carbohydrate-protein linkage region of xylose/serine linked
proteoglycans.

Methodology:

The combined use of MD, EM and constraints derived from nuclear
Overhauser effects (NOEs) to refine model structures for biological
molecules is now well established (2-7). The protocol used in our
methodology is shown in Figure 1. Because of the lack of a general
solution to treat multiple minima in the (ϕ,ψ) space (8,9), it is
essential to construct several starting structures so that they can
span the available conformational space in a reasonable manner (8).
Hence Stage 1 of the protocol involves the construction of several
starting structures that satisfy this requirement. Our initial
attempts to generate such structures by subjecting an arbitrary
conformation of GXS to a high-temperature (1000 K) MD simulation (a
50 ps MD simulation without constraints, where random atomic velo-
cities corresponding to 1000 K were repeatedly assigned every 5 ps)
showed that this MD simulation could not overcome the barriers
across the interresidue linkages, and as a result the conformations
tended to localize around three points in the (ϕ,ψ) space rather
than sample the available space. Ha et al have chosen, in their
simulation of maltose (8), starting conformations defined on a 20°
grid in the (ϕ,ψ) space and these were subsequently energy mini-
mized. In the present study, we chose a total of nine starting con-
formations over the (ϕ,ψ) space for each interresidue linkage
(i.e., 81 starting conformations for GXS, and in general 9^n start-
ing conformations for an oligosaccharide with " n " linkages) and
subjected each of these to high-temperature MD evolution (1000 K, 5
ps). The starting conformations were defined by assigning the three
torsional angle values corresponding to gauche+, gauche- and trans
configurations across the C-O and O-C' bonds defining the linkage.
Our choice of 9 starting conformations for each linkage was dic-
tated by our desire to limit the computations to a tractable num-
ber, while assuring that enough conformations have been selected to
span the (ϕ,ψ) space in a reasonable manner. Clearly, the protocol
shown in Figure 1 can be used for any arbitrary number of starting
conformations spanning the (ϕ,ψ) space.
     Our protocol and the adiabatic mapping procedure described by
Ha et al (8) have two features in common. They are both based on
molecular mechanics approach, and they both start with several
structures scattered at regular intervals over the (ϕ,ψ) space. But
the objectives of the approaches are fundamentally different.
Whereas the adiabatic mapping procedure is intended to fully char-
acterize the conformational energy surface of disaccharides, our
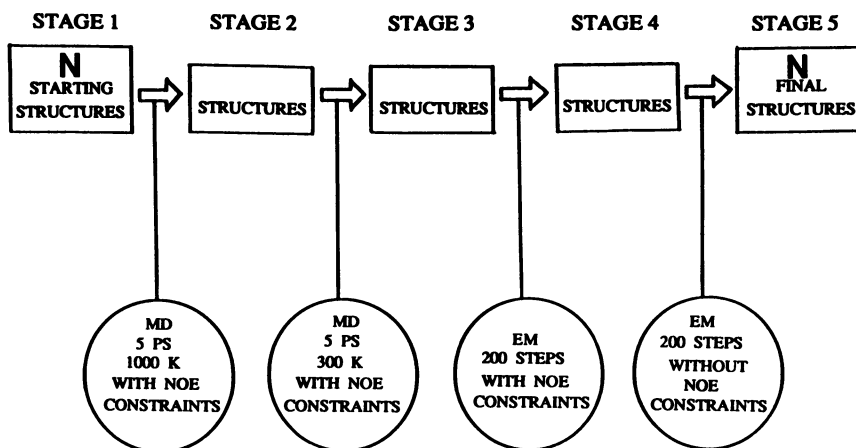
Figure 1: The protocol used in the molecular modeling
studies on complex carbohydrates. For unbranched carbohy-
drates, this protocol is applied to each interresidue lin-
kage at a time and uses N starting conformations (Stage 1)
and arrives at N final structures (Stage 5). In the cur-
rent investigation, we chose 9 starting conformations as
described in the text. The Stage 5 structures generated
for each linkage serve as modules for the construction of
the structures for the entire molecule. These
are subjected to a final step of energy minimization.

procedure is aimed at finding all conformations that are consistent with a set of experimentally determined distance constraints.

In order to relate the modeled structures to experimental observables during this high-temperature MD simulation, spatial constraints between hydrogen atoms across the linkage were included in the form of a pseudoenergy function (vide infra). The resulting structures (Stage 2) were annealed by an additional 5 ps MD simulation at 300 K with NOE constraints to arrive at the set of structures in "Stage 3". These structures were subjected to 200 steps of energy minimization using the conjugate gradient method. The NOE constraints were retained during this step to produce models compatible with NOE data. The resulting set of nine structures (Stage 4) were then subjected to 200 steps of energy minimization without NOE constraints. This last step relieves the strains in the structures introduced by the NOE pseudoenergy function, and produces models (Stage 5) with acceptable stereochemistry in terms of bond lengths and bond angles.

The parameters of the potential energy function used in our calculations on GXS were the same as those used in an earlier study of cyclodextrin (10), which were derived from GROMOS (11,12) with slight modification, as follows. The partial charges for a single sugar unit including the hydrogens were calculated (10) using the Gaussian 80 (UCSF) program with minimal basis set and without geometrical optimization. The atomic charges for the serine residue were those of GROMOS (11). The amino group of serine was assumed to be deprotonated and the carboxyl was assumed to be protonated to remove the positive and negative charges so as to make the charge distribution compatible with that in the native core protein. Table I contains a list of partial charges for the GXS molecule used in the computer modeling studies. A distant dependent dielectric constant was used in the calculations (6).

In addition to the usual terms for the covalent and noncovalent interactions, a constraint energy penalty term was added to the potential function for some of the MD/EM runs in the protocol as described above. A semiharmonic form was used for this term:

$$E(NOE) = K/2 \ (r - r_0)^2 \quad \text{for } r > r_0$$
$$= 0 \qquad\qquad \text{for } r \leq r_0 \qquad\qquad (1)$$

where K is the force constant (3 kcal/mol $^{-1}$ A $^{-2}$ in the present calculations), r is the interproton distance, and r_0 is the NOE cutoff distance. A value of 3.5 Å has been used for r_0 in our calculations.

NMR Spectroscopy:

The test oligosaccharide, GXS, was characterized extensively on a Bruker WH-400 NMR spectrometer (operating frequency 400 MHz) by 1D- and 2D-NMR spectrtoscopy. The synthesis (13) and the NMR spectroscopic characterization (13,14) of GXS has been reported elsewhere.

To generate the distance constraints to be used in the constraint energy function, E(NOE), the spatial contacts were established on the basis of a 2D-NOESY experiment with a 400 ms mixing time performed on a Bruker WH-400 NMR spectrometer. The following

Table I . List of Partial Atomic Charges for GXS

Residue Name	Atom Name	Atom Form	Atom Type	Sequence Number	Atom Partial Charge
Gal	C1'	C1	CS1	1	0.199
Gal	H1'	H1	HC	2	0.050
Gal	C2'	C2	CS1	3	0.026
Gal	O	O2	OA	4	-0.386
Gal	H	H9	HO	5	0.268
Gal	H2'	H2	HC	6	0.093
Gal	C3'	C3	CS1	7	0.065
Gal	O	O3	OA	8	-0.356
Gal	H	H10	HO	9	0.232
Gal	H3'	H3	HC	10	0.062
Gal	C4'	C4	CS1	11	0.056
Gal	O	O4	OA	12	-0.360
Gal	H	H11	HO	13	0.241
Gal	H4'	H4	HC	14	0.069
Gal	C5'	C5	CS1	15	0.061
Gal	H5'	H5	HC	16	0.059
Gal	C6'	C6	CS2	17	0.020
Gal	O	O6	OA	18	-0.529
Gal	H	H12	HO	19	0.385
Gal	H6'	H6	HC	20	0.081
Gal	H7'	H7	HC	21	0.051
Gal	O5'	O5	OS	22	-0.270
Xyl	O	O40	OS	23	-0.268
Xyl	C1'	C10	CS1	24	0.212
Xyl	H1'	H8	HC	25	0.045
Xyl	C2'	C20	CS1	26	0.027
Xyl	O	O20	OA	27	-0.333
Xyl	H	H22	HO	28	0.223
Xyl	H2'	H21	HC	29	0.100
Xyl	C3'	C30	CS1	30	0.168
Xyl	O	O30	OA	31	-0.301
Xyl	H	H32	HO	32	0.215
Xyl	H3'	H31	HC	33	-0.053
Xyl	C4'	C40	CS1	34	0.101
Xyl	H4'	H41	HC	35	0.008
Xyl	C5'	C50	CS1	36	0.002
Xyl	H5'	H51	HC	37	0.088
Xyl	H5"	H52	HC	38	0.043
Xyl	O5'	O10	OS	39	-0.267
Ser	O	O60	OS	40	-0.257
Ser	Cβ	C60	CH2	41	-0.095
Ser	Hβ'	H61	HC	42	0.110
Ser	Hβ"	H62	HC	43	0.115
Ser	Cα	C70	CH1	44	-0.005
Ser	Hα	H71	HC	45	0.107
Ser	N	N1	NT	46	-0.465
Ser	H	H72	H	47	0.131
Ser	H	H73	H	48	0.230
Ser	C	C80	C	49	0.381
Ser	O	O80	O	50	-0.294
Ser	O	O90	OA	51	-0.400
Ser	H	H90	HO	52	0.317

interresidue NOE contacts were observed: Gal-H1 to Xyl-H4; Gal-H1 to Xyl-H5$_{eq}$; and Xyl-H1 to Ser-H$_\beta$'. In addition, intraresidue contacts Gal-H1 to Gal-H3, Gal-H1 to Gal-H5, Xyl-H1 to Xyl-H3, and Xyl-H1 to Xyl-H5$_{ax}$ were also observed. All these inter- and intraresidue spatial contacts were used in the calculations.

Results:

Generation of Starting Conformations (Stage 1):

On the basis of the ^{1}H vicinal coupling constant analysis (13,14), both the sugars were assumed to adopt the $^4C_1$ chair conformations as starting structures in Stage 1. No sugar repuckering was observed in any of the simulations. The exocyclic torsion angle θ (O5'-C5'-C6'-O6') for the galactose residue has been assigned a starting value of 178.2° as observed in the crystal structure for β-D-galactose (15). An analysis of the vicinal coupling constant data (J$_{\alpha\beta}$', J$_{\alpha\beta}$") for the side chain of the serine residue indicated that the sidechain existed predominantly in the "C" rotamer. Hence a value of $\chi = 60°$ was used for all the Stage 1 conformations. In defining the Stage 1 conformations, the torsion angles for each linkage (i.e., ϕ_1,ψ_1 for X-S and ϕ_2,ψ_2 for G-X) were given values that correspond to the gauche+, gauche- and trans configurations. This procedure generated 9 conformations for each linkage (see Tables II and III).

MD/EM Calculations:

The interresidue linkages X-S and G-X were subjected separately to the protocol shown in Figure 1 and the variations in the torsion angles due to refinement at different stages are shown in Tables II and and III, respectively. The side chain orientation of serine defined by the torsion angle χ remains relatively invariant at the various stages of the refinement (see Table II). The exocyclic torsion angle, θ (O5'-C5'-C6'-O6') for the galactose residue groups into two values centered around 177.2 and 60.7°in the final stage. On the other hand, the linkage torsion angles ϕ_1 and ψ_1 and ϕ_2 and ψ_2 experience considerable variations at various stages of the MD/EM refinement. These variations at different stages are also plotted in Figures 2 and 3 to emphasize the convergence of nine starting conformations into distinct families. The conformations for X-S converge into three distinct families; A$_1$, B$_1$, and C$_1$. For the other linkage, G-X, the conformations converge into two distinct families, A$_2$ and B$_2$.

In each case, there was only one set of conformations in the final stage that correctly reproduced the observed NOESY contacts (see Table IV). These are the A$_1$ conformations for X-S and the A$_2$ conformations for G-X. Apparently the other conformations represent models that are trapped in local energy minima. Since they do not satisfy the observed NOE contacts, these other conformations (B$_1$, C$_1$ and B$_2$) are ignored in the remainder of the analysis.

From these results, the conformation for the GXS molecule was constructed as A$_2$A$_1$. Here the average linkage torsion angles for the A$_2$ (G-X) and A$_1$ (X-S) families have been used. A value of

Table II ; Results of MD/EM Calculations on the XS Linkage

Torsion Angles in X-S*	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Final Family
ϕ_1	+60.0	-62.0	-63.4	-61.6	-59.9	
ψ_1	+60.0	+179.8	+175.5	+178.6	+178.7	B_1
χ	+60.0	+55.1	+63.5	+61.1	+61.3	
ϕ_1	+60.0	+165.1(-194.9)	+165.6(-194.4)	-177.7	-178.3	
ψ_1	-60.0	+92.4	+94.7	+68.7	+67.9	A_1
χ	+60.0	+69.8	+58.9	+57.9	+58.0	
ϕ_1	+60.0	+5.8	-54.5	-60.5	-58.7	
ψ_1	±180.0	+160.3	+174.8	+179.2	+179.6	B_1
χ	+60.0	+73.5	+55.8	+62.2	+63.1	
ϕ_1	-60.0	+144.5(-215.5)	+167.7(-192.3)	-175.2	-175.8	
ψ_1	+60.0	+95.2	+80.8	+73.1	+73.2	A_1
χ	+60.0	+68.7	+61.6	+61.2	+60.6	
ϕ_1	-60.0	-68.4	-63.3	-70.8	-67.9	
ψ_1	-60.0	-140.8	-92.2	-76.8	-75.7	C_1
χ	+60.0	+70.0	+64.1	+67.3	+66.7	
ϕ_1	-60.0	-49.3	-64.5	-61.4	-61.4	
ψ_1	±180.0	-175.8	+176.9	±180.0	±180.0	B_1
χ	+60.	+62.7	+58.1	+62.1	+62.1	
ϕ_1	±180.0	-120.5	+167.4(-192.6)	-177.5	-177.9	
ψ_1	+60.0	+63.7	+85.5	+70.9	+69.3	A_1
χ	+60.0	+50.9	+61.7	+58.7	+57.8	
ϕ_1	±180.0	-110.6	-118.8	-172.2	-171.7	
ψ_1	-60.0	+50.4	+59.4	+81.8	+81.4	A_1
χ	+60.0	+64.7	+52.0	+59.8	+59.8	
ϕ_1	±180.0	+175.9(-184.1)	+163.9(-196.1)	+179.8(-180.2)	-177.5	
ψ_1	±180.0	+95.9	+82.4	+71.7	+69.7	A_1
χ	+60.0	+58.0	+60.2	+58.9	+58.5	

* ϕ_1(O5'-C1'-O-Cβ) and ψ_1(C1'-O-Cβ-Cα) define the linkage torsion angles between Xyl and Ser. χ(O-Cβ-Cα-N) defines the side chain orientation of Ser. The angles are expressed in degrees.

Table III ; Results of MD/EM Calculations on the GX linkage

Linkage Torsion Angles in G-X*		Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Final Family
ϕ_2		+60.0	-46.2	-56.9	-64.9	-67.1	A_2
ψ_2		+60.0	+99.5	+115.0	+127.2	+129.0	
ϕ_2		+60.0	-51.2	-80.9	-69.0	-75.2	B_2
ψ_2		-60.0	-40.2	-63.3	-51.9	-54.8	
ϕ_2		+60.0	-1.2	-60.9	-57.8	-62.1	A_2
ψ_2		±180.0	+132.5	+123.2	+123.7	+127.6	
ϕ_2		-60.0	-49.8	-60.9	-57.4	-59.4	A_2
ψ_2		+60.0	+118.2	+126.9	+124.9	+126.8	
ϕ_2		-60.0	-77.6	-71.8	-77.8	-76.0	B_2
ψ_2		-60.0	-67.6	-61.4	-55.9	-55.1	
ϕ_2		-60.0	-55.8	-58.9	-57.1	-60.7	A_2
ψ_2		±180.0	+140.6	+120.9	+124.4	+126.4	
ϕ_2		±180.0	-56.9	-60.3	-64.2	-64.7	A_2
ψ_2		+60.0	+134.5	+119.5	+129.3	+128.9	
ϕ_2		±180.0	-65.7	-59.5	-63.2	-65.9	A_2
ψ_2		-60.0	+158.2	+123.8	+131.9	+132.2	
ϕ_2		±180.0	-78.9	-57.3	-56.9	-59.6	A_2
ψ_2		±180.0	+138.4	+127.6	+124.7	+127.0	

* ϕ_2(O5'-C1'-O-C4') and ψ_2(C1'-O-C4'-C3') define the linkage torsion angles between Gal and Xyl. The angles are expressed in degrees.
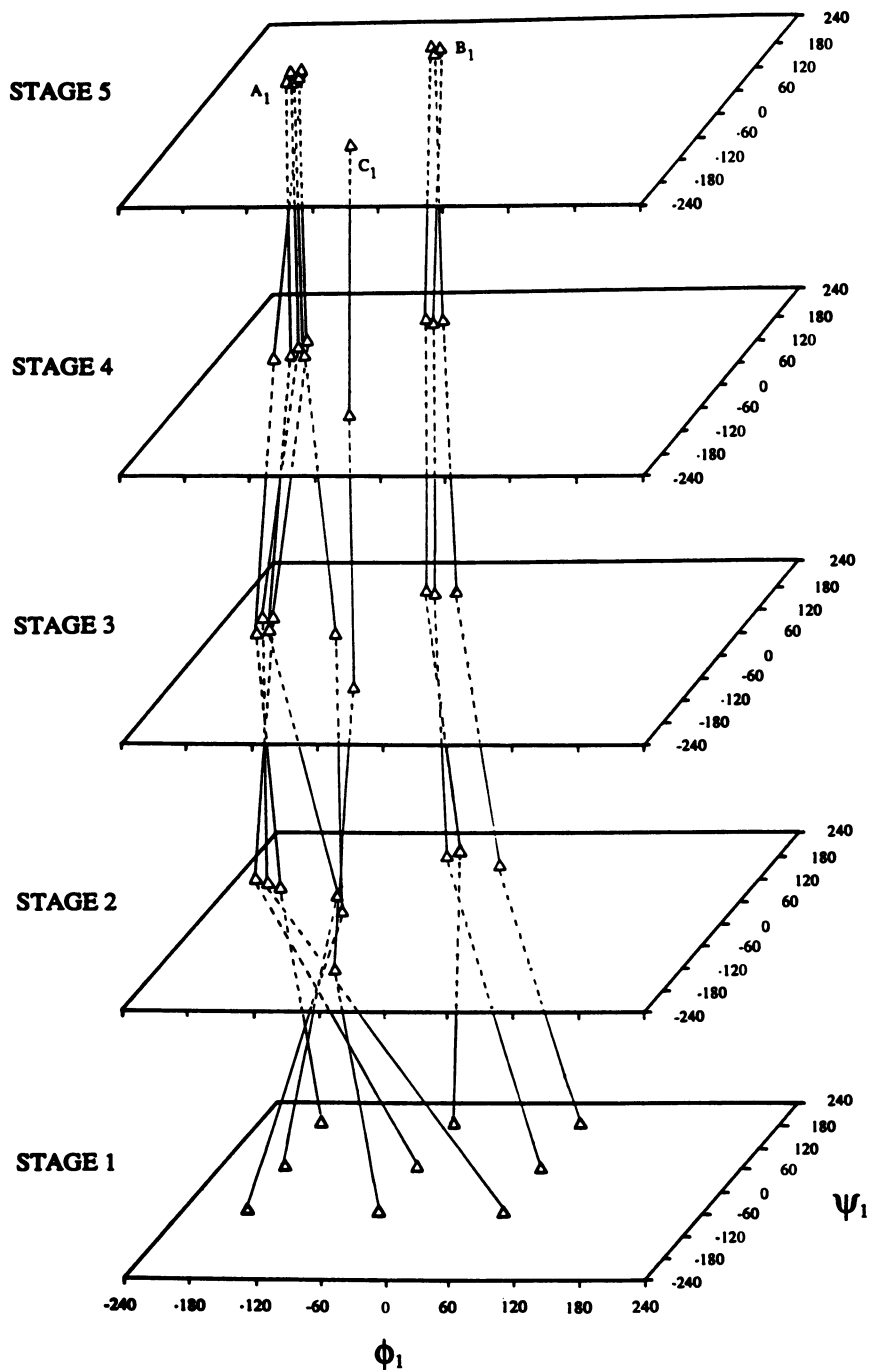
Figure 2: Variation in the linkage torsional angles (ϕ_1, ψ_1) for the XS linkage of GXS at different stages of the MD/EM protocol in figure 1.
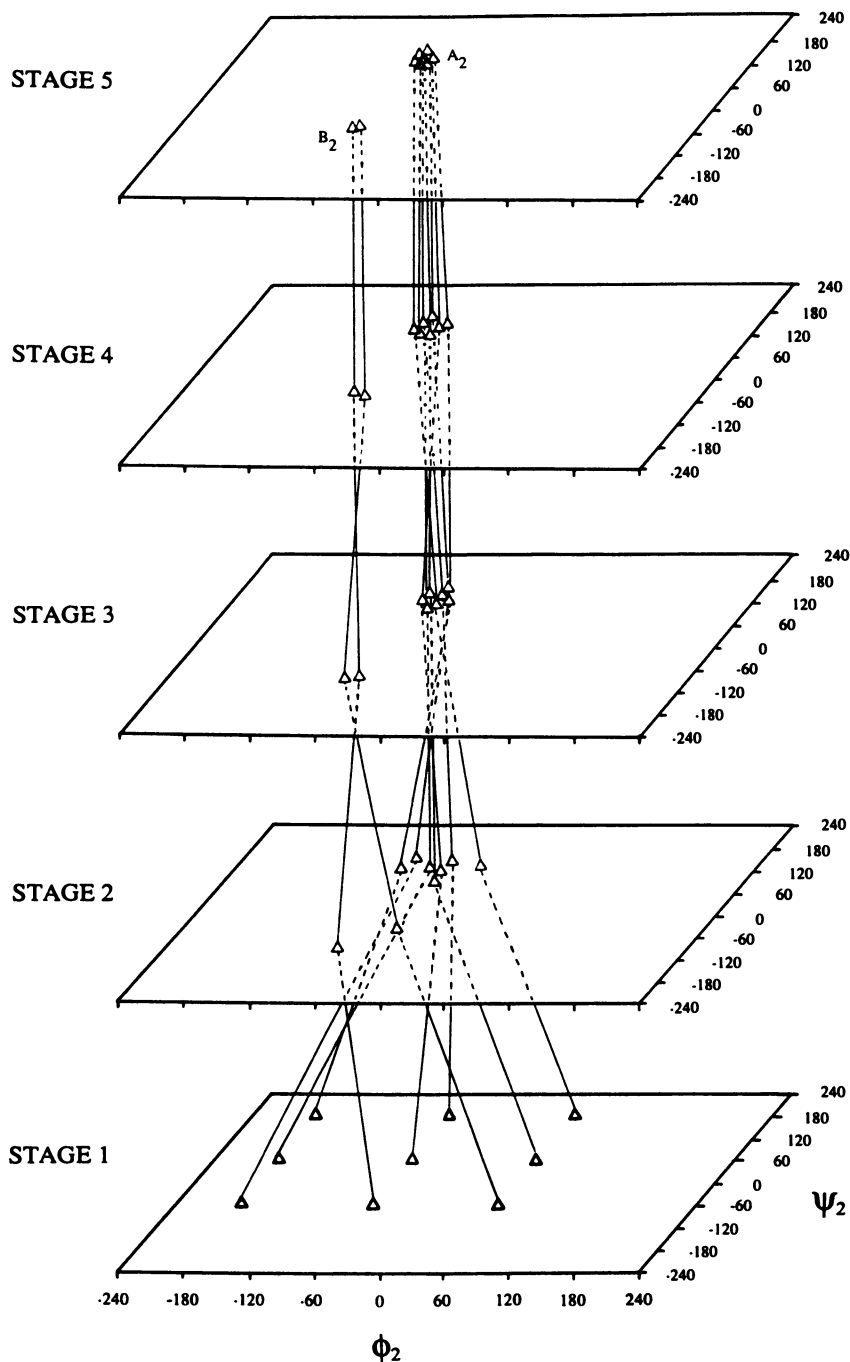
Figure 3: Variation in the linkage torsional angles (ϕ_2, ψ_2) for the GX linkage of GXS at different stages of the MD/EM protocol in figure 1.

177.2° was used for the exocylioc torsion angle of galactose in the A_2 conformations. To account for any long range effects on charge distribution and to relieve any steric conflicts that might arise due to the modular construction of the conformation of the whole molecule, the family of conformations A_2A_1 was subjected to additional energy minimization without NOE constraints. The resulting family, $A_2'A_1'$, is representative of the conformations of GXS compatible with the experimental data. In these conformations ($A_2'A_1'$), one weak interresidue hydrogen bond between Xyl 3'OH and Gal O5' has been detected (O...H 2.6 Å , O-H...O 114°). This was an artifact stemming from the <u>in vacuo</u> calculations, and could be overcome by choosing an appropriate dielectric constant. Figure 4 shows a typical example from the final A_2' A_1' set of structures for GXS.

Even though the $A_2'A_1'$ family explains the observed NOE contacts in GXS, we have also examined whether a conformational exchange involving different families could also explain the observed interresidue NOE contacts. Table IV shows a comparison of the calculated interresidue proton distances for the families A_1, B_1, C_1, A_2 and B_2, together with the average torsion angles. The B_1 family predicts that both β' and β" protons will experience a significant NOE contact with Xyl-H1' proton, whereas the C_1 family predicts a NOE contact between the Xyl-H1' and Ser Hβ". These contacts were not observed in the experiment. The B_2 family for the GX linkage predicts that both the H5' and H5" protons of xylose will experience strong NOE contact with Gal-H1' proton. This too is at variance with the experiment. Thus a conformational exchange with families such as $A_2'B_1'$, $B_2'A_1'$, $B_2'C_1'$ etc would have resulted in NOE contacts that were not observed in the experiment on GXS. This suggests that these families do not make a major contribution to the conformational equilibrium of GXS. However, for other types of oligosaccharides, such a conformational exchange should also be considered in the modeling calculations.

In conclusion, we have developed a general methodology particularly suitable for modeling of unbranched complex carbohydrates, since it lends itself to a modular analysis of a long carbohydrate chain. This methodology is based on MD/EM calculations with NMR derived constraints introduced into the calculations to generate conformations compatible with the experimental data.

Our approach differs from that of Scarsdale et al (7) in some essential details. The form of the pseudoenergy function chosen in our work to represent NOE constraints is similar to that used by Kaptein et al (5) in their protein modeling studies, but differs from that of Scarsdale et al. The latter authors have used a potential function that has a negative minimum at $r = r_0$, approaches zero for $r \gg r_0$ and is positive for $r \ll r_0$. This function leads to rapid convergence, but the distance r_0 has maximum weight because of the form of the function. The semiharmonic potential chosen in our work allows us to set an upper limit on r_0 for the distance between any two protons that produce an NOE. We have selected a conservative value of 3.5 Å for r_0. Further, Eq.(1) gives equal weight to all distances less than r_0, and hence the computed distances will be relatively free of any bias induced by the pseudoenergy function as long as the NOE distance constraint

Table IV ; Average linkage torsion angles (in degrees) and
          interresidue distances (in Å) for Stage 5 families

Linkage	Gal - Xyl		Xyl - Ser		
Family	A_2	B_2	A_1	B_1	C_1
Average Linkage Torsion Angles ϕ_2	-62.8	-75.6	ϕ_1	-176.2	-60.0	-67.9
ψ_2	128.3	-55.0	ψ_1	72.3	179.4	-75.7
Average Inter-Residue Distances GH1-XH4	2.4	3.5	XH1-SHβ'	2.9	3.1	3.6
GH1-XH5'	2.8	2.4	XH1-SHβ"	3.7	2.6	3.0
GH1-XH5"	3.7	2.9	XH1-SHα	4.1	4.9	4.0
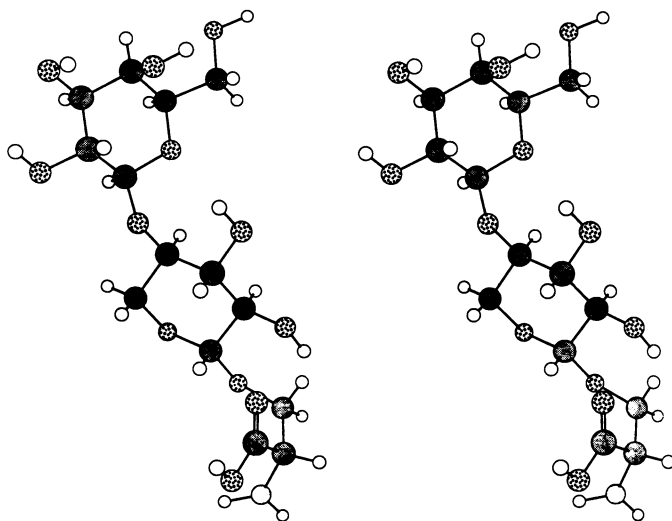
Figure 4: A typical conformation from the $A_2'A_1'$ family of
final structures for GXS.

is satisfied. A second difference involves our final step in the protocol (Figure 1) in which the NOE constraints were lifted and the conformations were further subjected to EM to relax any unreasonable steric and bond angle distortions that result from the previous steps (MD/EM) that employed distance constraint penalty term (E(NOE)).

Acknowledgments:

Literature Cited:

1. Roden, L., in The Biochemistry of Glycoproteins and Proteoglycans, Ed. Lennarz, W.J., Plenum, New York, p. 267.
2. Levitt, M., J. Mol. Biol. 1983, 170, 723.
3. Clore, G.M., Gronenborn, A.M., Brunger, A.T., and Karplus, M., J. Mol. Biol., 1985, 186, 435.
4. Folkers, P.J.M., Clore, G.M., Driscoll, P.C., Dodt, J., Khler, S., and Gronenborn, A.M., Biochemistry, 1989, 28, 2601.
5. Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R., and van Gunsteren, W.R., J. Mol. Biol., 1985, 182, 179.
6. McCammon, J.C., and Harvey, S.C., Dynamics of Proteins and Nucleic Acids, 1987, Cambridge, London.
7. Scarsdale, J.N., Ram, P., Prestegard, J.H., and Yu, R.K., J. Comput. Chem., 1988, 9, 133.
8. Ha, S.N., Madsen, L.J., and Brady, J.W., Biopolymers, 1988, 27, 1927.
9. Scheraga, H.A., Biopolymers, 1983 22, 1.
10. Prabhakaran, M., and Harvey, S.C., Biopolymers, 1987, 26, 1087.
11. Van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hoe, W.G.J., and Postua, J.P.M., Proc. Natl. Acad. Sci. USA, 1983, 80, 4315.
12. Aqvist, J., van Gunsteren, W.F., Leijonmarck, M., and Tapia, O., J. Mol. Biol. 1985, 183, 471.
13. Ekborg, G., Klonger, M., Roden, L., Jensen, J.W., Schutzbach, J.S., Huang, D.H., Krishna, N.R., and Anantharamaiah, G.M., Glycoconj. J. 1987. 4, 255.
14. Van Halbeek, H., Dorland, L., Veldink, G.A., Vliegenthart, F.G., Garegg, P.J., Norberg, T., and Lindberg, B., Eur. J. Biochem., 1982, 127, 1.
15. Sheldrick, B., Acta Cryst 1976, B32, 1016.

Chapter 15

Molecular Mechanics NMR Pseudoenergy Protocol To Determine Solution Conformation of Complex Oligosaccharides

J. Neel Scarsdale[1], Preetha Ram[2], James H. Prestegard[2], and Robert K. Yu[1]

[1]Department of Biochemistry and Molecular Biophysics, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA 23298–0614
[2]Department of Chemistry, Yale University, New Haven, CT 06511

Here we present a protocol for the determination of the solution conformation of complex oligosaccharides which relies on a systematic combination of NMR distance constraints, derived from the ratio of cross peak intensities in two dimensional cross-relaxation correlated (NOESY) experiments, and molecular mechanics calculations. In general, it is not possible to determine a sufficient number of distance constraints from cross-peak intensity data to permit the unambiguous determination of solution conformation from NMR distance constraint data alone. Molecular mechanics calculations, on the other hand give rise to a number of minimum energy structures which cannot be distinguished on the basis of potential energy alone. In combination, however, an accurate structural definition may arise. NMR distance constraint data, on the one hand serves to select between energetically similar conformers. Molecular mechanics calculations, on the other hand, serve to exclude energetically unreasonable NMR structural solutions. In addition, our protocol allows us to relax the assumption that the observed cross relaxation rates result from a single rigid conformer by assuming that the observed cross relaxation rates result from a weighted average over multiple conformers. As an illustration, we have applied this protocol to the problem of the determination of the solution conformation of the neutral tetrasaccharide headgroup of the glycolipid globoside. Comparison of the results from the one and two-state models suggests that only a narrow range of conformers is present in solution—a result which is consistent with the antigenic determinant and receptor functions which have been proposed for globoside.

It has long been recognized that NMR cross relaxation data can provide information on the solution conformation of biological macromolecules, mainly through the inverse sixth power dependence of the cross-relaxation rates on internuclear distance. Only recently with the advent of sophisticated two-dimensional acquisition techniques has it become possible to attempt the *a-priori* determination of the solution conformation

of a biological macromolecule. Noteworthy examples include the studies by Wüthrich and coworkers (1–4) and by Clore, Gronenborn, et al. (5–9) on a number of small proteins and polypeptides. Applications to other types of biological macromolecules are certainly feasible (10–24). Those we shall present here involve oligosaccharides occurring as the hydrophilic portion of glycolipids and of glycoproteins.

The basis for the determination of solution conformation from NMR data lies in the determination of cross relaxation rates between pairs of protons from cross peak intensities in two-dimensional nuclear Overhauser effect (NOE) experiments. In the event that pairs of protons may be assumed to be rigidly fixed in an isotopically tumbling sphere, a simple inverse sixth power relationship between interproton distances and cross relaxation rates permits the accurate determination of distances. Determination of a sufficient number of interproton distance constraints can lead to the unambiguous determination of solution conformation, as illustrated in the early work of Kuntz, et al. (25). While distance geometry algorithms remain the basis of much structural work done today (1–4), other approaches exist. For instance, those we intend to apply here represent NMR constraints as pseudoenergies for use in molecular dynamics or molecular mechanics programs (5–9).

Molecular mechanics and molecular dynamics programs use empirical energy functions to represent molecular properties and provide an efficient way of representing our accumulated knowledge of preferred bond geometries. Furthermore, they allow for the placement of portions of biomolecules, such as the exocyclic CH_2OH, OH and N-Acetyl groups of carbohydrates in reasonable orientations, even when there are insufficient NMR constraints to fix their orientations. NMR distance constraints may be incorporated into these calculations through the use of a distance dependent error function which minimizes when a preferred geometry is found, much as the energy functions representing the bonding and nonbonding interactions among atoms.

Combining NMR data with theoretical molecular structure calculations through the use of such functions is particularly attractive given their complementary natures. In general it is not possible to determine a sufficient number of constraints to permit the unequivocal determination of the three-dimensional solution structure of a biological macromolecule, while potential energy calculations give rise to a number of energetically similar structures which cannot be distinguished given the limited accuracy of the empirical energy functions used in these calculations. In combination, an accurate structural definition may arise. NMR distance constraints, on the one hand, serve to exclude energetically similar conformers which are not consistent with the NMR data. Potential energy calculations, on the other hand, serve to exclude energetically unreasonable NMR structural solutions. Furthermore, the use of NMR data treated as pseudoenergies is appealing because it would seem possible to tailor the pseudoenergy function to accurately represent represent the precision of NMR distance constraint measurements. It should also be possible to allow for a realistic interplay of NMR and chemical bonding constraints by the choice of an appropriate weighting function for the pseudo potential.

We present here applications using the molecular mechanics program contained in the model building package AMBER (AMBER, copyright 1986, University of California, San Francisco, is obtained via a licensing agreement with the regents of the University of California.) Even with the proper representation of NMR data, the use of such programs suffers from at least two limitations. One is the tendency of search

algorithms to identify structures which correspond to local rather than global minima. Another is the possibility that NMR data represent an average over multiple conformers rather than the simple rigid structures identified in a conventional energy search. We shall explore the local minimum problem through the use of multiple starting structures well displaced from one another in conformational space. We shall approach the problems associated with assuming the cross relaxation data result from a single rigid conformer by comparing the results of a one-state calculation with results from a two-state model which assumes the observed cross-relaxation data pertain to an average over two discrete conformers.

As an illustration of our methodology, we have chosen the determination of the solution conformation of the oligosaccharide headgroup of globoside which is the glycosphingolipid shown in Figure 1. It has a neutral tetrasaccharide headgroup and is the major glycolipid present in the erythrocytes from all but a small minority of individuals. Globoside has been implicated in a number of biological functions including serving as the P-blood group antigen (26,27) and serving as a receptor for strains of *Escherichia Coli* responsible for polynephritis (28). Besides being of importance with respect to antigenic and receptor functions of globoside, the conformation of the oligosaccharide headgroup of globoside provides a fair test of our methodology in that there are insufficient distance constraints to accurately determine the conformation from cross relaxation data alone, and there has been considerable speculation concerning the conformational flexibility of the oligosaccharide headgroup of cell surface glycolipids.

The Choice of a Pseudoenergy Function

In order to incorporate distance constraints derived from two-dimensional cross-relaxation data in a molecular mechanics program, we have chosen to treat the constraints as a pseudoenergy function. This function should ideally reflect the distance dependence of cross relaxation rates. Previously, we had proposed a function of the form (40):

$$E_{ab}^{NOE} = W \left[ABS \left(\frac{1}{r_{ab}^6} - \frac{1}{r_{0ab}^6} \right) - \frac{1}{r_{0ab}^6} \right]. \qquad (1)$$

In expression 1, r_{0ab} is the distance between protons a and b determined on the basis of cross relaxation rates; r_{ab} is the distance between them at a given stage of the calculation, and W is a weighting factor chosen to properly balance errors in theoretical energies and NMR pseudoenergies $\approx$ 700 kcal/Å^6 which corresponds to a well depth of $\approx$ 10 kcal/mol, when $r_{0ab} = 3.0$Å. This function has the correct $1/r_{ab}^6$ distance dependence for all values of r_{ab}. It has a minimum at $r_{ab} = r_{0ab}$, and the depth of the minimum is proportional to $1/r_{0ab}^6$. These properties accurately reflect the distance dependent precision of the NMR data. This function, however, is not suitable for integration with molecular mechanics algorithms which require that the first and sometimes second derivatives be continuous for all values of r_{ab}.

In order to integrate the NMR distance constraint data into a molecular mechanics minimization routine, we have chosen a smooth approximation of the above function

with the form:

$$E_{ab}^{NOE} = W \left[\left(\frac{1}{r_{ab}^3} - \frac{1}{r_{0ab}^3} \right)^2 - \frac{1}{r_{0ab}^6} \right].$$

(2)

This function has a minimum at the appropriate place and a well depth proportional to $1/r_{0ab}^6$. For distances $r_{ab} \ll r_{0ab}$, energies become positive with a $1/r_{ab}^6$ dependence as they should. The $1/r_{ab}^3$ dependence on distance for $r_{ab} \gg r_{0ab}$ is not strictly correct. This long range dependence may actually assist in a rapid convergence of the calculation when trial structures have interproton distances much larger than the value observed via cross relaxation studies. This function represents a reasonable choice of a pseudoenergy term given the additional constraints imposed on the derivatives of the function.

Functions such as 1 and 2 permit the inclusion of a pseudoenergy term for all pairs of protons in a molecule, whether or not cross relaxation rates are sufficiently large to be observed. This is made possible by the fact that for sufficiently long distances, the energy contributions from these functions are negligible. For cases where no connectivity is observed between a pair of protons r_{0ab} is set equal to some distance, in our case 4.0Å, beyond which a connectivity between protons would be lost in the noise. Inclusion of a pseudoenergy term in the absence of an observed connectivity is important since it serves to exclude conformers with interproton distances short enough to produce a connectivity when none is observed. We shall make limited use of this fact in our structural determinations.

It is common practice to assume that observed cross relaxation rates result from a single rigid conformer. In these cases the pseudoenergy function described above is adequate. Violation of this assumption would not, however, be at all unusual. Violation would give rise to the possibility that the observed cross relaxation rates represent an average over multiple conformers, each significantly different than the single state structural solution which best fits the observed cross relaxation data. In certain cases, it is possible to relax this assumption. If two or more conformers were to exist and interconvert on a time scale which was short compared to the cross relaxation time, but long compared to the correlation times which were important for spin relaxation ($10^{-8} \ll 10^{-2}$ sec), the observed cross relaxation rate can be written as a weighted average over cross relaxation rates for the individual conformers. If we assume the energies for the various conformers to be approximated by the molecular mechanics energies, we can model the behavior without introducing additional variables. The average cross relaxation rate is given by:

$$\sigma_{ab} = \chi_1 \left(\sigma_{1ab} - \sigma_{2ab} \right) + \sigma_{2ab})$$

(3)

where:

$$\chi_1 = \frac{e^{-(E_2 - E_1)/kT}}{1 + e^{-(E_2 - E_1)/kT}}$$

(4)

In expression 3, σ_{ab} is the average cross relaxation rate between protons a and b; χ_1 is the mole fraction of one of the conformers, and σ_{1ab} and σ_{2ab} are the cross relaxation rates for the individual conformers. In expression 4, E_1 and E_2 are the molecular mechanics energies for the conformers, k is the Boltzmann constant and T is the temperature at which the data were obtained. The pseudoenergy term may be

written in terms of the average cross relaxation rate σ_{ab}:

$$E_{NOE} = W \left[(\sqrt{\sigma_{ab}} - \sqrt{\sigma_{0ab}})^2 - \sigma_{0ab} \right].$$ (5)

In expression 5, W is a weighting factor; as before, σ_{0ab} is the measured cross relaxation rate and σ_{ab} is the calculated cross relaxation rate at any stage in the calculation as given by expression 3. Since σ is directly proportional to $1/r^6$, this expression reduces to expression 2 in the limit of a single rigid conformer, if W is adjusted to absorb the proportionality constants.

The two-state model was implemented by treating the two conformers as a pair of non-interacting molecules contributing to the spin relaxation properties in proportion to their Boltzmann factors. The overall energy to be minimized for the two-state calculation is defined as:

$$E = E_1 + E_2 + \sum_{a>b} E_{NOE}(\sigma_{ab}).$$ (6)

In expression 6, E_1 and E_2 are the molecular mechanics energies for the two conformers and $E_{NOE}(\sigma_{ab})$ is the distance constraint energy for a pair of protons. From equations 5 and 6, it is clear that E_{NOE} couples the conformation of the two molecules through the explicit dependence of the cross relaxation rates on the geometry of each conformer.

In expression 6, the molecular mechanics energies E_1 and E_2 are not weighted by the populations expected on the basis of Boltzmann factors. Unlike the calculation of real energies, factors weighting various contributions are meant only to maintain a balance of errors. There is no difference in the precision of molecular mechanics energies calculated for the major and minor conformers; however the importance of the NMR data for each conformer varies according to the population of that conformational state.

Experimental Methodology

Globoside (2 mg), obtained from human erythrocytes and kindly provided by Dr. S. Ando of the Tokyo Metropolitan Institute of Gerontology, was prepared for high-resolution NMR analysis as described previously (29). The sample was dissolved in .4 ml of d_6-Me$_2$SO-D$_2$O (98:2 v./v.). Perdeuterated solvents were obtained from Merck and Company (St. Louis MO) or Aldrich, Inc., (Milwaukee, WI).

Figure 2 shows a one-dimensional spectrum for globoside. This spectrum was in large part assigned through a spin echo correlated (SECSY) experiment. Additional assignments were made through a coupling correlated (COSY) experiment which had somewhat better resolution than the experiment used to make the initial assignments. A complete list of resonance assignments is given in Table I.

We shall focus our attention on the anomeric resonances between 4.1 and 4.9 ppm. There are four such resonances, each corresponding to one of the sugar rings in globoside. The resonance at 4.82 ppm with the small coupling constant is clearly the resonance for the anomeric proton of the α-D-galactose residue. The others at 4.54, 4.24 and 4.14 ppm correspond to the β-D-N-Acetylgalactosamine, the β-D-galactose and the β-D-glucose residues, respectively.
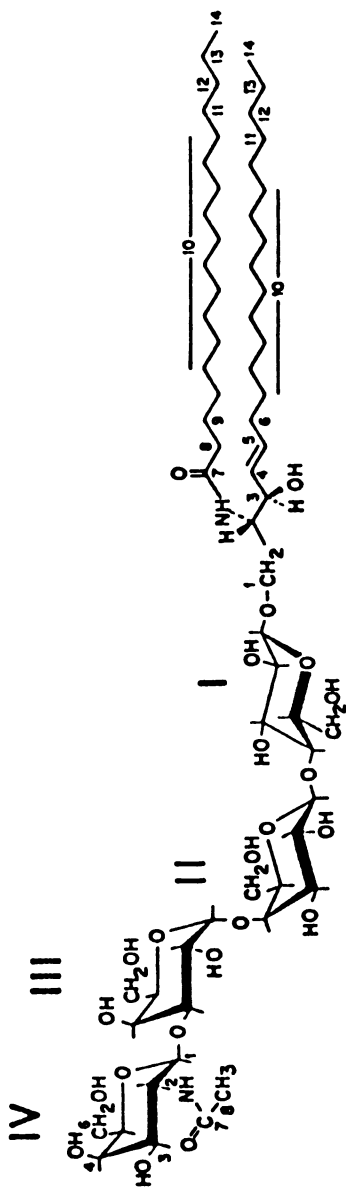
Figure 1. Primary structure for globoside.
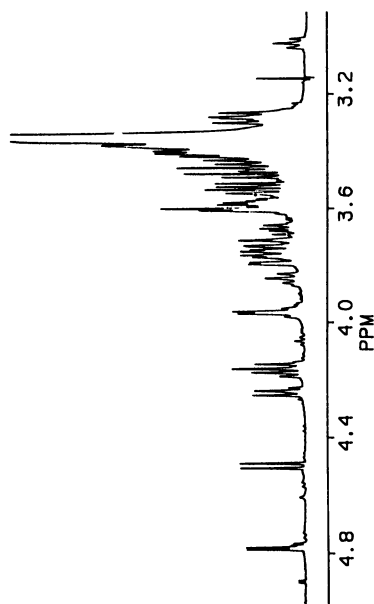
Figure 2. One-dimensional ^{1}H-NMR spectrum for globoside at 323 K. (Reproduced from ref. 40. Copyright 1986 American Chemical Society.)

Table I: Chemical Shifts for Globoside in Me$_2$SO-d$_6$, D$_2$O at 323°K[1]

residue	1	2	3	4	5	6a	6b
IV (β-D-GalNAc)	4.54	3.77	3.46	3.67	3.42	3.60	3.57
III (α-D-Gal)	4.82	3.82	3.65	4.00	4.15	3.49	
II (β-D-Gal)	4.24	3.33	3.43	3.84	3.58	3.65	3.77
I (β-D-Glc)	4.14	3.08	3.34	3.29	3.29	3.63	3.76

[1]SOURCE: Reproduced from ref. 40. Copyright 1986 American Chemical Society.

Pure Absorbtion NOESY Experiments

In Figure 3, we present a contour plot of the region from a pure absorbtion NOESY experiment on globoside which contains all of the connectivities between the oligosaccharide protons. The NOESY experiment was obtained at 303°K with a mixing time of 250 ms on a Bruker WM500 spectrometer and consisted of 192 real and 192 imaginary t_1 points which were acquired and stored separately to achieve quadrature in the t_1 dimension (30). Each t_1 point consisted of 1K complex points over a 3205 Hz sweep width averaged for 64 transients with a 1.2 sec. repetition rate. Data were processed using the FTNMR (FTNMR, copyright 1988 Hare Research, is obtained via a licensing agreement with Hare Research.) program.

In this experiment, there are two classes of cross peaks connecting anomeric resonances to other resonances in the spectrum, those due to interactions with protons on the same residue, and those due to interactions with protons on adjacent residues. In each case, intensities are proportional to the inverse sixth power of the distance of each proton from the anomeric proton. Given the signal to noise ratio in these experiments, only cross peaks for protons within 3.5Å are observed. For α-D-galactosides and α-D-glucosides, one expects to see only a 1-2 equatorial ring connectivity at 2.5Å. For β-D-galactosides and β-D-glucosides, one expects to see 1-3 and 1-5 axial connectivities, also at 2.5Å. Given the previous assignments, these are easily found. For example, in the column through the β-D-galactose anomeric resonance at 4.24 ppm, the cross peak at 3.43 is to the H3 resonance from the same residue, while the cross peak at 3.58 ppm is to the H5 resonance from the same residue. These cross peaks can be integrated to provide distance standards for the determination of distances between interresidue proton pairs.

Examination of Figure 3 shows one or more cross peaks for each anomeric resonance which correspond to interresidue cross peaks. The strongest interresidue connectivities are in fact assigned to the anomeric-aglyconic proton connectivities. For example, in the case of the β-D-galactose residue, the interresidue cross peak at 3.29 ppm corresponds to the H4 resonance of glucose. Additional weaker interresidue connectivities are observed for all anomeric resonances. For example in the case of the β-D-galactose anomeric resonance, one sees connectivities to the H6a and H6b resonances of glucose at 3.63 and 3.76 ppm, respectively. These weaker cross peaks provide additional distance constraints which serve to improve conformational definition.
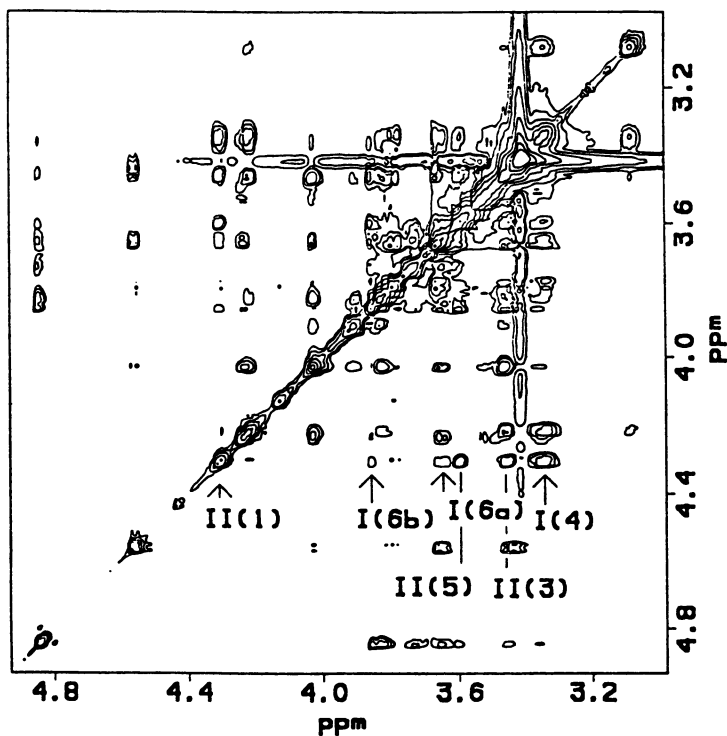
Figure 3. Contour plot of 500 MHz pure absorption NOESY spectrum of globoside at 303 K. The labeled cross peaks correspond to the interresidue and intraresidue connectivities for globoside. (Reproduced from ref. 40. Copyright 1986 American Chemical Society.)

Table II: Distance Constraint Data for Globoside Dissolved in Me$_2$SO-d$_6$, D$_2$O[1]

	r_{ij}
IV(1)–III(3)	2.3 ± 0.03
IV(1)–III(4)	2.9 ± 0.09
IV(1)–IV(3)	2.5^2
IV(1)–IV(5)	2.5^2
III(1)–II(4)	2.4 ± 0.03
III(1)–II(5)	3.5 ± 0.2
III(1)–II(6a)	2.6 ± 0.04
III(1)–II(6b)	2.6 ± 0.04
III(1)–III(2)	2.5^2
III(5)–III(3)	2.5^2
II(1)–I(4)	2.1 ± 0.02
II(1)–I(6a)	2.8 ± 0.05
II(1)–I(6b)	2.6 ± 0.03
II(1)–II(5)	2.5^2

[1]SOURCE: Reproduced from ref. 40. Copyright 1986 American Chemical Society.

[2]Rigidly fixed proton pair used as a standard in computing interresidue distances.

The Determination of Distance Constraints for Globoside

In order to quantitate these cross peak intensities from two-dimensional data sets, it is necessary to use cross peak volumes. These volumes were obtained with a two dimensional trapezoidal rule approximation. Only the volume above a base plane at an average level of a rectangular region surrounding the cross-peak was used. The use of such a base plane offers an advantage in that it serves to remove bias introduced into cross peak volumes by changes in base plane levels. Volumes were computed for several different choices of base planes in order to obtain error limits on cross peak volumes. For cross peaks at 3.0Å, volume errors were typically on the order of 20%. An average over all cross peaks along with standard error propagation methods was used to estimate errors in the various distance constraints.

Cross peak volumes were converted to distance constraints via the use of an expression of the form:

$$r_{ij} = \left(\frac{a_0}{a_{ij}} \right)^{1/6} r_0. \qquad (7)$$

In expression 7, r_{ij} is the distance to be determined; a_{ij} is the corresponding cross peak intensity, a_0 is the cross peak intensity corresponding to a rigidly fixed pair of protons on the same saccharide residue, and r_0 is the distance between them. In Table II, we present a list of distance constraints and their associated errors which were determined from the ratio of cross peak intensities in the NOESY data set shown in Figure 3.

Calculational Methodology

In choosing a molecular mechanics program in which to incorporate distance constraint pseudoenergies and test protocols for the determination of solution structure, one would seek a program with a well tested molecular mechanics basis, a convenient means of generating starting structures, some flexibility in the way in which energy contributions are incorporated and adequate means for the comparison of output structures. We have chosen the AMBER (Associated Model Building and Energy Refinement) package introduced by Weiner and Kollman (31). This program meets all of the above requirements and has been used in a variety of structural studies on biomolecules (32–34).

Modification of the AMBER Force Field for Simulations on Complex Carbohydrates

The original AMBER force field parameters were based on experimental data and subsequently refined with molecular mechanical studies on the structure and energies of protein and nucleic-acid like compounds, leaving sugars inadequately simulated. Therefore, we have modified the all atom force field of AMBER using a small number of additions to improve simulations of sugars. We have avoided extensive reparameterization which would have likely increased computation time and perhaps decreased the compatability of this force field with proteins and nucleic acids. At an absolute minimum, bond and dihedral angles unique to glycosidic linkages would have to be defined. These include the C–O–C, O–C–O, and C–O–C–O. Initially, parameters for these segments were adapted from analogous parameters present in the AMBER force field for a united atom representation of ribosides and nucleic acids. We also introduced lone pairs (LP) on the glycosidic and ring oxygens (OL) of each saccharide residue. This requires specification of additional parameters. Allinger, et al. (35) had previously modified the MM2 force field to simulate oxygen-containing sugar-like molecules such as dimethoxy methane, tetrahydrofuran and dimethyl ether. A common feature of these modifications is the explicit inclusion of lone pairs on oxygen atoms. Lone pairs were treated as pseudo atoms with well defined atomic radii, Van der Waal's constants and bonding parameters. Parameters for lone pairs were taken from Allinger's force field (35–37). Such an explicit definition of lone pairs in oxygen atoms requires a redistribution of partial atomic charges, especially of the lone pair bearing oxygen atoms. The charges on the terminal hydrogens of the glycosidic fragment, C–OL–C–OL–C were chosen to be consistent with charges on analogous hydrogens of compounds already in the AMBER data base. The remaining charges were apportioned among the anomeric carbon, the lone pairs and the lone pair bearing oxygen atoms using Allinger's dipole moment data (37).

Major adjustments of parameters included force constants and equilibrium angles for C–OL–C and C–C–OL. Parameters were optimized to produce agreement with the crystal structure of lactose (38). The glycosidic angle, C1–On′–Cn′, is around 117° in most glycosides. The equilibrium value for this angle was therefore changed to 117° in the force field. This increases the torsional flexibility of the molecule at the glycosidic linkage. We note that since the glycosidic oxygen is a special lone pair bearing type, this change in the equilibrium bond angle value does not affect values assigned to chemical linkages normally defined in the AMBER data base.

In our early minimized structures of lactose, the angles around the glycosidic oxygen C3'–C4'–O4' and C5'–C4'–O4' showed the largest deviations from the crystal geometry. We therefore restricted the tendency towards large C–C–O (OL) angles by decreasing the equilibrium value for this type of angle and increasing the corresponding force constant to effect a certain measure of rigidity in the system. These changes caused a decided improvement in the lactose test case. Again since this oxygen is a special atom type, these changes do not affect the original data base. All added and modified parameters are summarized in tables III and IV.

Selection of Starting Structures

In order to assure the location of the global minimum rather than local minimum energy structures, we have started calculations using three structures well displaced from one another in conformational space. These structures differ from one another in the relative orientations of the carbohydrate residues. The orientations of these residues were defined in terms of two dihedral angles about the glycosidic linkage. The dihedral angle determined by the anomeric proton, anomeric carbon, glycosidic oxygen and aglyconic carbon was designated ϕ, and the dihedral angle determined by the aglyconic proton, aglyconic carbon, glycosidic oxygen and anomeric carbon was designated ψ. The first structure, designated A, had initial ϕ, ψ values of 60°, 0° for the IV,III linkage, 300°, 0° for the III,II linkage and 60°, 0° for the II,I linkage (Figure 1). The second structure, designated B, had initial ϕ, ψ values of 300°, 0°, 60°, 0° and 300°, 0° for the respective linkages, while the third structure, designated C, had initial ϕ, ψ values of 180°, 0° for each pair of carbohydrate residues.

Structure A was chosen to have ϕ, ψ values near those for the single minimum energy structure determined for globoside using Bock and Lemiuex's HSEA program (39,40). Structure B was chosen because it also has ϕ, ψ values near those commonly observed in previous studies of oligosaccharide conformation. Structure C was chosen to have ϕ, ψ values displaced 120° from the other initial structures.

In Table V, we present rms deviations for the NMR distance constraints for structures A, B and C for globoside. Here we present four sets of deviations. The first set, which we have designated rms_{obs}, was obtained by including only those distance constraints which corresponded to observed NOE connectivities in the calculation. The second set, which we have designated rms_{intra}, was obtained by including only the subset of observed connectivities which corresponded to intraresidue connectivities in the calculation. The third set of deviations, which we have designated rms_{inter}, was obtained by including only the subset of observed connectivities which corresponded to interresidue connectivities in the calculation, and the final set of deviations, which we have designated rms_{all}, was obtained by including in addition to the constraints corresponding to observed NOE connectivities, the constraints which corresponded to the lack of observed NOE connectivities. In Table V, we also present a list of significant violations (deviations ≥0.5Å) of the distance constraints corresponding to observed connectivities for these initial structures for globoside. From these data, it is clear that none of these initial structures represents a reasonable fit of the experimental distance constraints.

Table III: Force field parameters for Glycolipids

Bond Parameters[1]

Bond	K_r	r_eq
CT−OL	536.0	1.407
HO−OL	460.0	0.960
LP−OL	460.0	0.600

Angle Parameters[2]

Angle(θ)	K_θ	θ_{eq}
CT−OH−HO	55.0	108.5
CT−OL−HO	55.0	108.5
CT−OS−CT	60.0	109.5
CT−OL−CT	60.0	117.0
OL−CT−OS	46.0	102.2
OH−CT−OS	46.0	102.2
OH−CT−OL	46.0	102.2
OL−CT−OL	46.0	102.0
LP−OL−HO	24.0	101.0
LP−OL−LP	24.0	131.0
CT−OL−LP	35.0	105.2
HO−CT−LP	24.0	101.0
CT−CT−OL	101.0	107.5

Torsional Parameters[3]

Angle(ϕ_i)	V_1	γ_1	V_2	γ_2	V_3	γ_3
CT−CT−CT−OL	0.05	0.0	0.05	180.0	0.09	0.0
HC−CT−CT−OL	0.0	0.0	0.0	0.0	0.09	0.0
OL−CT−CT−OL	0.0	0.0	−0.03	180.0	0.15	0.0
CT−CT−OL−LP	0.0	0.0	0.0	0.0	0.0	0.0
CT−CT−OL−HO	0.4	0.0	0.0	0.0	0.045	0.0
HC−CT−OL−CT	0.0	0.0	0.0	0.0	0.275	0.0
HC−CT−OL−LP	0.0	0.0	0.0	0.0	0.0	0.0
HC−CT−OL−HO	0.0	0.0	0.0	0.0	0.1	0.0
OL−CT−OL−CT	−0.085	0.0	−0.6	180.0	0.0	0.0
OL−CT−OL−LP	0.0	0.0	0.25	180.0	−0.1	0.0
OH−CT−OL−LP	0.0	0.0	0.25	180.0	−0.1	0.0
OL−CT−OL−HO	0.0	0.0	0.0	0.0	0.0	0.0
CT−CT−OL−CT	0.0	0.0	0.0	0.0	0.65	0.0
OH−CT−OL−CT	−0.085	0.0	−0.6	180.0	0.0	0.0

[1] $E_{bond} = \sum_{bonds} K_r (r - r_{eq})^2$

[2] $E_{angle} = \sum_\theta K_\theta (\theta - \theta_{eq})^2$

[3] $E_{torsions} = \sum_\Phi \sum_n \frac{V_n}{2} [1 + cos(n\Phi - \gamma_n)]$

Table IV Electrostatic and Nonbonded parameters for Glycolipids

Electrostatic Charges	Atoms	Charge	
	H	0.083	
	CT(terminal)	−0.118	
	CT(anomeric)	0.460	
	OL	0.298	
	LP	−0.328	
	OH	−0.500	
	HO	0.310	
Van der Waals Parameters[1]	Atom	R_e	ϵ
	OL	1.65	0.002
	LP	0.80	0.025

$$^{1}E_{VanderWaals} = \sum_{i<j} \epsilon \left[(R_e/R_{ij})^{12} - 2(R_e/R_{ij})^6 \right]$$

Calculational Protocol for NMR Pseudoenergy-Molecular Mechanics Calculations

All minimizations were executed on a Vax 11-750 computer equipped with a floating point accelerator. Structures were minimized both in the presence and the absence of distance constraint pseudoenergies. In all cases a steepest descent minimizer was used for the first 500 cycles of the calculation, and a conjugate gradient minimizer was used for the remainder of the calculation where possible. An initial stepsize of 0.05 was used, and convergence for one-state calculations was defined as the rms of the norm of the gradient being ≤ 0.05 kcal/Å. NMR distance constraints were input via the normal parameter input routine of AMBER and converted to pseudoenergies via a modified bond constraint energy calculation routine.

In the initial stages of the calculation, a group minimizer was used in which the geometry of the individual residues were held constant. Each additional residue was allowed to move only via a combination of translations and rotations about the center of mass for that residue. In this stage of the calculation, a simplified potential surface consisting of bond, angle and distance constraint pseudoenergies was used. These simplifications proved to be important in minimizing the tendency of our calculations to become trapped in local minima. In the next stage of our calculation, all atoms were allowed to move independently of one another and torsional and 1-4 nonbonded energy terms were added. Finally after 100 such cycles, all additional nonbonded interactions were added, and the calculation was allowed to run to convergence. During these stages of the calculation, distance constraints involving only ring protons were assigned a weight corresponding to a well depth of 100 kcal which reflects the greater uncertainty introduced by the fact that we do not attempt to model the rapid motional averaging and/or second order character which could exist for this group (41). These weights were similar to the weights assigned to the harmonic potentials which were used in the AMBER force field to represent bond energies and represent a compromise which allows NMR pseudoenergies to dominate during these initial stages of the calculation, yet which prevents the occurrence of structures with unreasonable bond geometries

Table V: Deviations from Experimental Distance Constraints for Various Structural Solutions for Globoside

	rms_{obs}[1]	rms_{intra}[2]	rms_{inter}[3]	rms_{all}[4]	Significant Violations[5]		
A	0.06	0.04	0.18	0.06	III(1)–II(6b)	0.6	$(2.6)^6$
					III(1)–II(5)	0.8	$(3.5)^6$
					IV(1)–III(4)	1.0	$(2.9)^6$
B	0.12	0.04	0.34	0.12	III(1)–II(6a)	1.5	$(2.6)^6$
					III(1)–II(6b)	1.7	$(2.6)^6$
					III(1)–II(5)	1.2	$(3.5)^6$
					II(1)–I(6a)	0.9	$(2.8)^6$
					II(1)–I(6b)	0.9	$(2.6)^6$
C	0.15	0.04	0.44	0.12	IV(1)–III(3)	1.1	$(2.3)^6$
					IV(1)–III(4)	1.0	$(2.9)^6$
					III(1)–II(4)	1.0	$(2.4)^6$
					III(1)–II(6a)	1.1	$(2.6)^6$
					III(1)–II(6b)	1.8	$(2.6)^6$
					III(1)–II(5)	1.6	$(3.5)^6$
					II(1)–I(4)	1.4	$(2.1)^6$
					II(1)–I(6a)	0.6	$(2.8)^6$
					II(1)–I(6b)	1.7	$(2.6)^6$
A′	0.05	0.02	0.16	0.06	IV(1)–III(3)	1.0	$(2.3)^6$
					IV(1)–III(4)	1.7	$(2.9)^6$
					II(1)–I(4)	1.0	$(2.9)^6$
					III(1)–II(5)	0.5	$(3.5)^6$
					III(1)–II(6a)	−0.5	$(2.9)^6$
					II(1)–I(6a)	−0.6	$(2.8)^6$
B′	0.15	0.02	0.45	0.10	III(1)–II(4)	0.9	$(2.4)^6$
					III(1)–II(5)	1.5	$(3.5)^6$
					III(1)–II(6a)	1.9	$(2.6)^6$
					III(1)–II(6b)	2.2	$(2.6)^6$
					II(1)–I(6a)	1.5	$(2.8)^6$
					II(1)–I(6b)	1.6	$(2.6)^6$
C′	0.16	0.03	0.46	0.12	IV(1)–III(3)	1.4	$(2.3)^6$
					IV(1)–III(4)	1.5	$(2.9)^6$
					III(1)–II(4)	1.3	$(2.4)^6$
					III(1)–II(5)	1.6	$(3.5)^6$
					III(1)–II(6a)	0.5	$(2.6)^6$
					III(1)–II(6b)	1.7	$(2.6)^6$
					II(1)–I(4)	1.5	$(2.1)^6$
					II(1)–I(6a)	0.7	$(2.8)^6$
					II(1)–I(6b)	1.7	$(2.6)^6$
D′	0.03	0.02	0.10	0.04	II(1)–I(6a)	0.6	$(2.8)^6$
					II(1)–I(6b)	0.7	$(2.6)^6$
E′	0.05	0.02	0.15	0.08	IV(1)–III(4)	−0.5	$(2.9)^6$
					II(1)–I(6a)	0.7	$(2.8)^6$
					II(1)–I(6b)	0.7	$(2.6)^6$

Table V: Continued

	rms_{obs}[1]	rms_{intra}[2]	rms_{inter}[3]	rms_{all}[4]	Significant Violations[5]		
A″	0.02	0.0	0.05	0.02			
B″	0.02	0.0	0.05	0.02			
C″	0.02	0.0	0.05	0.02			
HSEA[7]	0.04	0.03[8]	0.10	0.04	III(1)–II(5)	0.5	(3.5)[6]
					II(1)–I(6b)	0.7	(2.6)[6]
D′A′	0.02[9]	0.0 [9]	0.05[9]	0.03[9]			
D′B′	0.02[9]	0.0 [9]	0.05[9]	0.02[9]			
A′B′	0.02[9]	0.0 [9]	0.07[9]	0.02[9]	III(1)–II(5)	0.6	(3.5)[6]

[1]RMS deviations (in Å) for twenty-seven distance constraints corresponding to observed connectivities in NOESY data set on globoside

[2]RMS deviations (in Å) for eighteen distance constraints corresponding to observed intraresidue connectivities in NOESY data set on globoside

[3]RMS deviations (in Å) for nine distance constraints corresponding to observed interresidue connectivities in NOESY data set on globoside

[4]RMS deviations (in Å) for entire set of forty-two distance constraints included in structural calculation on globoside

[5]Deviations ≥ 0.5Å for observed constraints

[6]Experimental value in Å

[7]Distance obtained from calculation in reference 40

[8]Intraresidue distances were held fixed at values observed in crystal structures in calculation described in reference 40

[9]Interproton distances were calculated from the formula $r_{ij} = r_{ij1}r_{ij2}/(\chi_1 r_{ij1}^6 + \chi_2 r_{ij2}^6)^{1/6}$ where r_{ij1} and r_{ij2} are the distances in the model conformers and χ_1 and χ_2 are the mole fractions of the model conformers. These values are in no sense real distances and are included merely for comparison with other data.

upon convergence of these calculations with heavy weightings of the NMR distance constraint pseudoenergies.

When nonbonded energies were initially included, we frequently noted a tendency to invert configuration at the various ring carbons, especially in structures obtained from starting points B and C. This was prevented through the use of additional constraints added during the initial stages of the calculation, when this tendency was most pronounced due to a combination of large nonbonded energies and heavy weightings of the distance constraint pseudoenergies. For α-D-galactose III, constraints were added between the glycosidic oxygen and 3 and 5 protons at 2.6Å using a heavy weight. These constraints are analogous to constraints included between the anomeric proton and H3 and H5 protons in β-linked D-glucosides and D-galactosides.

In the final stages of the calculations, distance constraints were assigned a weight corresponding to 25 kcal/mol for constraints involving only ring protons and 12.5 kcal/mol for constraints involving exocyclic protons. These weights were chosen to allow for a realistic interplay between distance constraint pseudoenergies and molecular mechanics energies. Also in the final stages of the calculations, fifteen additional constraints were added for cases where a short interresidue distance was predicted in the structures obtained upon convergence of the calculations with heavy weightings of

the distance constraint pseudoenergy; yet no connectivity was observed in the NOESY data set. For these constraints the interproton distance was set equal to 4.0Å and was assigned the same weighting as the distances corresponding to observed connectivities. Inclusion of these additional constraints led to significant changes in the structures obtained upon convergence of the calculations with heavy weighting of the distance constraints.

Two-state calculations based on expression 6 rely on the validity of molecular mechanics energies to a much higher degree than the one-state calculation. Therefore, it is reasonable to start calculations with initial structures near the molecular mechanics minima. Several starting structures were generated for this purpose. A', B' and C' are conformers obtained by minimizing starting structures A, B and C without NMR constraints. Two additional minimum energy structures, designated D' and E', were generated by refining in the absence of distance constraints the structures obtained upon convergence of the one-state calculations with heavy weightings of the distance constraint pseudoenergies. Starting points for these calculations were chosen to be combinations of the three lowest energy structures. The first starting point, designated $D'A'$, was a pair of structures corresponding to the two lowest energy minima, namely D' and A'. The second starting point, designated $D'B'$, was a pair of structures corresponding to the lowest and third lowest minima, namely D' and B', and the third, designated $A'B'$, was a pair of structures corresponding to the second and third lowest minima, namely A' and B'. Furthermore, two of these starting structures, namely $D'B'$ and $A'B'$, are combinations of structures well displaced from one another in conformational space, as they should be if a two-state model with intermediate interconversion rates is to be a good approximation of the actual structure.

Two-state calculations were carried out using the same constraints as one-state calculations. All energies and pseudoenergies were included from the first stage. The weights assigned the distance constraints varied from 12.5 to 200 kcal/mol in an effort to force the calculations out of local minima. In the final stages of the calculation, constraints were assigned weights which were identical to those used in the one-state calculations. The calculation was allowed to run to convergence, with convergence being defined as the rms of the norm of the gradient being ≤ 0.1 kcal/Å.

In order to quantitate the differences between structures obtained using various protocols, rms deviations between the cartesian coordinates for the atoms in the oligosaccharide moieties for various pairs of structures were calculated using the orthogonal transform routine of the MOULD program (MOULD, 1988, is a program written at Yale University by Dr. Simon K. Kearsley.), in which structures were superimposed by first moving the centroids of the structures to the origin and then formulating a rotation matrix for the structures that superimposed the structures for a minimum rms deviation (42).

Structures Obtained in the Absence of Distance Constraints

Even when initial structures were refined in the absence of distance constraint data, geometries change significantly. ϕ, ψ values for these minimum energy structures along with the additional minimum energy structures are presented in Table VI. Ball and

Table VI: Conformational Properties of Various Structural Solutions for Globoside[1]

	$\phi_{IV,III}, \psi_{IV,III}$	$\phi_{III,II}, \psi_{III,II}$	$\phi_{II,I}, \psi_{II,I}$	E
A' [2]	73 , 59	314 , 322	80 , 35	−2.6
B' [2]	43 , 296	66 , 58	38 , 304	4.2
C' [2]	159 , 38	192 , 0	169 , 2	7.2
D' [2]	44 , 298	310 , 337	68 , 336	−5.4
E' [2]	67 , 310	321 , 324	60 , 334	8.0
A'' [3]	22 , 317	321 , 340	53 , 347	6.4
B''[3]	336 , 333	318 , 340	52 , 347	17.9
C''[3]	18 , 320	329 , 329	47 , 348	9.6
D'A' [4]				
$\chi_1 = 1.0$	37 , 311	317 , 342	55 , 345	−1.4
$\chi_2 = 0.0$	72 , 58	316 , 323	79 , 35	2.4
D'B' [4]				
$\chi_1 = 1.0$	29 , 315	317 , 342	54 , 347	−0.65
$\chi_2 = 0.0$	42 , 296	66 , 58	39 , 305	10.2
A'B' [4]				
$\chi_1 = .38$	78 , 55	327 , 330	73 , 36	9.0
$\chi_2 = .62$	37 , 305	65 , 57	32 , 315	8.7

[1]SOURCE: Reproduced with permission from ref. 24. Copyright 1988 Wiley.

[2]Structural solution obtained by minimization in the absence of distance constraint pseudoenergies.

[3]One state structural solution obtained by minimization with distance constraint pseudoenergies.

[4]Two state structural solution obtained by minimization with distance constraint pseudoenergies.

stick drawings of structures **A'**, **B'** and **C'** are presented in Figure 4. It is significant that these calculations do not converge to a single minimum. The fact that significantly different starting structures are found from different starting points supports the idea that molecular mechanics calculations suffer from a rather severe tendency to become trapped in local minima. The apparent global minimum energy structure corresponds to **D'** and has an energy of −5.4 kcal/mol. It proves to be quite similar to the structures obtained upon refining initial structures **A** and **C** in the presence of distance constraints. The structure is also quite similar to the minimum energy structure obtained for globoside using Bock and Lemieux's HSEA program (39,40). The ϕ values in structure **D'** typically agree within 15° with the corresponding ϕ values in the HSEA structure, but there is less agreement in the values of ψ. These departures, however, are less significant, given that previous studies on the conformational energy surface for the oligosaccharide moiety of globoside indicate that rather large variations in ψ correspond to small differences in energy (39,40).

The other minimum energy structures are also interesting. Two of these minima, namely **A'** at −2.6 kcal/mol and **E'** at 8.0 kcal/mol have ϕ, ψ values for the **III,II** and **II,I** linkages similar to those in the apparent global minimum. They share an overall "L" shape with the apparent global minimum but differ in the orientation of
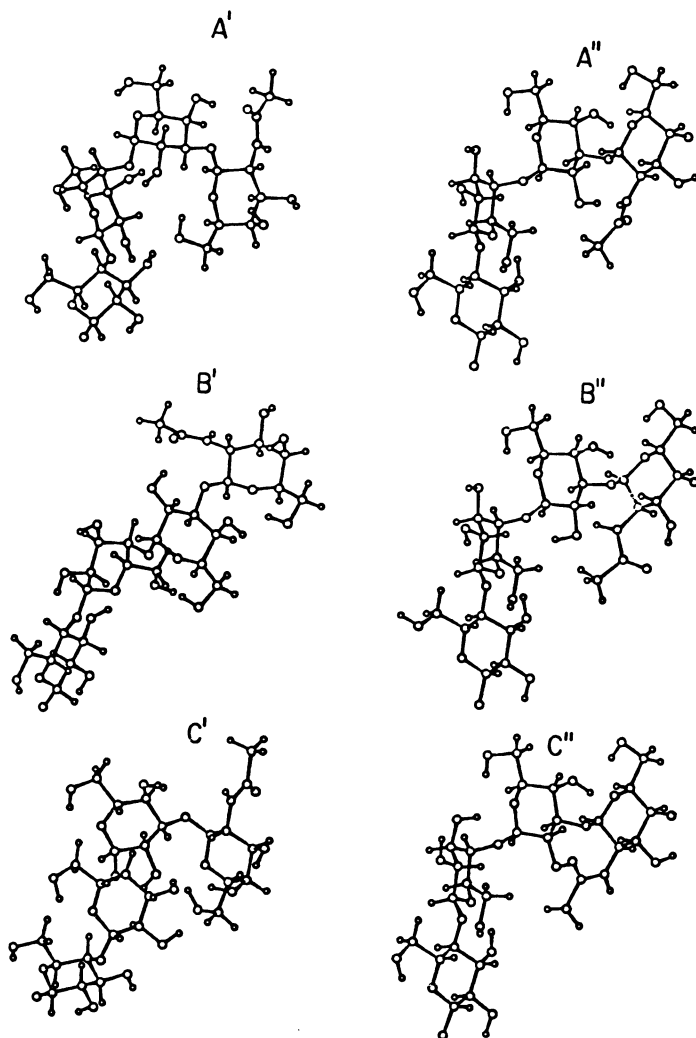
Figure 4. Comparison of energy refined structures A', B', and C' with NMR structures A'', B'', and C''. (Reproduced with permission from ref. 24. Copyright 1988 Wiley.)

the terminal GalNAc IV residue. More severe departures are observed in minima **B'** at 4.2 kcal/mol and **C'** at 7.2 kcal/mol, which have ϕ, ψ values for the **III,II** linkage which are quite different from the apparent global minimum.

The presence of energetically similar minima well displaced from one another in conformational space is consistent with studies on the conformational energy surface of disaccharides with the MMCARB2 program (43,44). This contrasts with previous studies on the solution conformation of globoside using the HSEA program, which predict only a narrow range of energetically similar conformers. This difference illustrates that by assuming rigidly fixed residue geometries, the HSEA program may artificially exclude regions of conformational space due to unfavorable contacts involving exocyclic groups. The limitations with assuming rigid residue geometries have been pointed out by Tvaroska and Perez (44).

In Table V, we present rms deviations for the distance constraints for each of these energy refined structures. In Table V, we also present a list of significant violations (deviations ≥ 0.5Å) for the distance constraints corresponding to observed connectivities for each of these energy refined structures. While the agreement for intra-residue distances has improved somewhat over the starting structures, there is little, if any improvement over the starting structures in fitting the interresidue constraints. The lack of agreement between the observed interresidue distances and the corresponding distances in the minimum energy structures indicates that molecular mechanics calculations have not lead to the correct solution conformation or that NOE distances cannot simply be equated with interproton distances in any single conformer.

NMR Refined One-State Structural Solutions

Among the possible problems with the one-state molecular mechanics description are that we have not located the proper global minimum and/or that there are energy contributions which are poorly represented which influence solution conformation. The NMR refined single conformer model allows distance constraint data to compensate for these potential deficiencies. The protocol described previously also provides a means of getting out of certain local minima by providing a significantly different potential surface in the early stages of the calculation. A summary of the ϕ, ψ values for the various one-state structural solutions, **A''**, **B''** and **C''** is presented in Table VI. Ball and stick drawings for these conformers are presented in Figure 4.

In order to quantitate the differences between these NMR refined structures, we have obtained three sets of RMS deviations by comparing structures **B''** and **C''** to structure **A''**. The first, designated rms_{all} corresponds to the deviations obtained when all atoms in the oligosaccharide moiety were included in the comparison. These deviations were typically on the order of 0.6Å. The second set, designated rms_{main}, corresponds to the deviations obtained when exocyclic groups whose orientations were unconstrained by NMR distance constraints were excluded from the comparison. These deviations were typically on the order of 0.3Å. Finally, the third set of deviations, designated rms_{I-III}, was obtained by excluding both unconstrained exocyclic groups and the terminal β-D-GalNAc residue from the comparison. These deviations were on the order of 0.2Å.

In Figure 5, we present a stereoview of the superposition of structures **A″**, **B″** and **C″** obtained when all atoms were included in the structural comparison. In Figure 5, we also present a stereoview of the superpositions of these three structures when unconstrained exocyclic groups were excluded from the comparison.

From the data in Table VI and Figure 5, it is clear that a combination of molecular mechanics calculations and distance constraint pseudoenergies converge to essentially a single minimum for the III,II and II,I linkages. In other words for two of these linkages, a nearly unique structural definition arises when a combination of distance constraint pseudoenergies and molecular mechanics calculations is employed. For the third linkage, namely the terminal GalNAc($\beta 1 - 3$)Gal linkage, there are two minima satisfying both bonding and NMR constraints.

The glycosidic torsion angles of structures **A″** and **C″** are very similar to those in the one-state conformational solutions determined in a previous study (40). Structure **B″** differs significantly in the orientation of the terminal GalNAc residue. All of these structures, however, display the overall "L" shape hypothesized to be significant in receptor function (39). The energies for these structures are 6.4 kcal/mol for structure **A″**, 18 kcal/mol for structure **B″** and 9.6 kcal/mol for structure **C″**. The energy difference between structures **A″** and **C″** arises largely because of differences in the orientations of unconstrained exocyclic groups whose orientations are not constrained by NMR data.

The presence of two general structures (**A″**, **C″**) and **B″** which differ in the orientation of the their terminal residues should not by itself be taken as evidence of conformational flexibility at the terminal linkage as was suggested in a previous study (40). It can arise simply because our data are not of sufficient quality to uniquely define the orientation of the terminal residues.

In Table V we present rms deviations for the distance constraints for each of these NMR refined structures. In Table V, we also present a list of significant violations (deviations ≥ 0.5Å) for these structures. With the exception of the III(1)–II(5) constraint, which may be an artifact due to strong coupling between the II(5) and II(6) protons (41), all of the constraints corresponding to experimental constraints are satisfied within experimental error. It is also worth noting that the structures obtained via a combination of distance constraint pseudoenergies and molecular mechanics calculations show significant improvements in fitting distance constraints involving exocyclic methylene groups over structures obtained using the distance constraint pseudoenergies with Bock and Lemieux's HSEA program (40). This improvement also arises from relaxing the assumptions of rigid crystal structure geometries for the individual residues.

NMR Refined Two-State Conformational Solutions

Although distance constraints are reasonably well satisfied in the one-state calculation, it is only done at a sacrifice of molecular mechanics bonding energies. All of the one state NMR refined structural solutions are higher in energy than the energy minimized structures **A′** and **B′**. It is therefore desirable to explore other means of fitting the NMR data. Another possible explanation for the lack of agreement between

the proton distances in structures determined in the absence of distance constraints and experimental distances is that multiple conformers exist in solution. The observed cross relaxation data would then be an average of the cross relaxation rates for the individual conformers. In order to test this hypothesis, we have tried fitting the experimental distance constraint data using the two-state model described previously and a set of starting points which corresponds to combinations of the various minimum energy structures in the absence of distance constraint pseudoenergies. None of the individual structures satisfies the observed data within experimental error. The two-state model allows for the possibility that combinations of structures near these minima represent an adequate fit of the experimental data. A summary of ϕ, ψ values and fractional populations for the individual conformers from the various two-state structural solutions is presented in Table VI. A ball and stick drawing of the lowest two-state structural solution is presented in Figure 6.

In examining the data in Table VI for the various structural solutions, several facts are worth noting. The lowest energy two-state structural solutions both involve significant occupation of only one conformational state. In each of these structural solutions, the dominant conformer is similar to structures A″ and C″ which were obtained as one-state structural solutions in the presence of distance constraint pseudopotentials. The third two-state structural solution involves significant occupation of two rather different conformational states similar to structures A′ and B′. This structural solution suggests conformational flexibility at the terminal IV,III linkage, which is consistent with our earlier work (40). Some additional flexibility is predicted at the III,II linkage. This additional flexibility was not predicted in our previous work, probably because the assumption of rigid residue geometry excluded some conformers which should have been allowed.

In Table V we present rms deviations for the distance constraints for each of the various two-state NMR structural solutions. In Table V, we also present a list of significant violations (deviations $\geq 0.5\text{Å}$) for these structural solutions. With the exception of the III(1)–II(5) constraint, all distance constraints corresponding to observed connectivities were satisfied within experimental error.

Discussion

Since both the one and two-state structural solutions provide an adequate fit of the experimental data, we must rely on an additional criterion to favor one approach over the other. Such a criterion is provided by the molecular mechanics energy. The lowest energy one-state structural solution has a molecular mechanics energy of 6.4 kcal/mol, while the lowest energy two-state structural solution had a molecular mechanics energy of -1.4 kcal/mol—significantly lower than lowest energy one-state structural solution. Fractional occupation of the second conformational state is small (.01) increasing to ≈ 0.1 when a constant dielectric of 10 is used instead of a distance dependent dielectric.

In Figure 7, we present a stereoview of the superposition of the dominant conformers from the lowest energy two-state structural solutions, D′A′ and D′B′ and the one-state structural solutions A″, B″ and C″ when all atoms in the oligosaccharide moiety of globoside were included in the comparison. In Figure 7, we also present

a

b
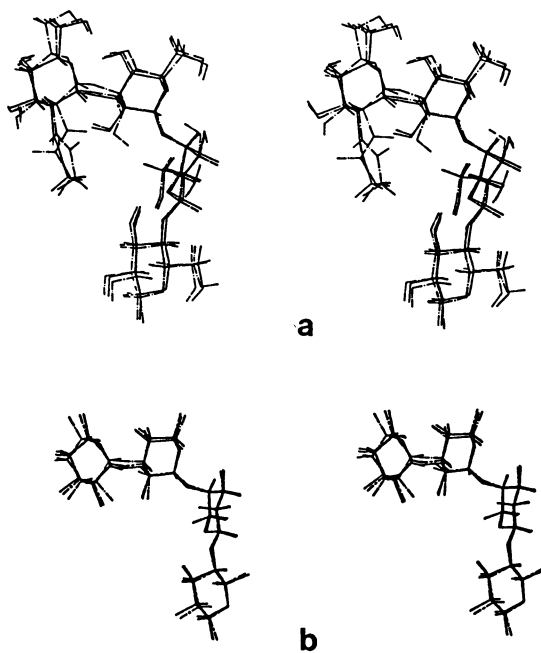
Figure 5. Stereoview of the superposition of the NMR refined one-state structural solution, A″ (–), B″ (–·), and C″ (–-) (a) with unconstrained exocyclic groups included in the comparison and (b) excluded from the comparison.

1

2
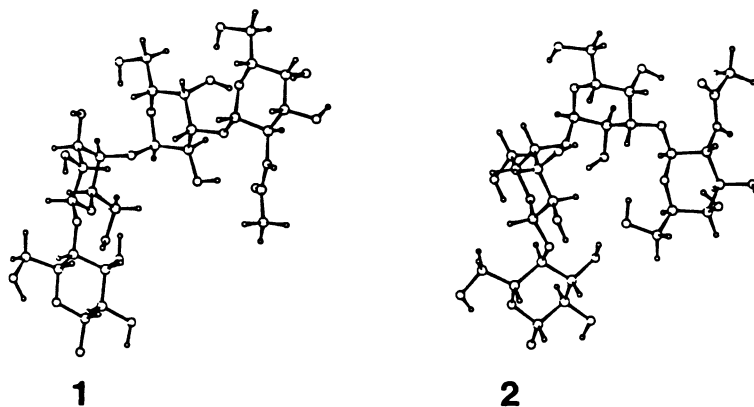
Figure 6. Ball and stick drawing of the dominant (1) and minor (2) conformers from the lowest energy two-state structural solution D′A′. (Reproduced from ref. 40. Copyright 1986 American Chemical Society.)

a stereoview of the superposition of these structures obtained when unconstrained exocyclic groups were excluded from the comparison.

From the data in Table VI and Figure 7, it is clear that the dominant conformer from the lowest energy two-state structural solution is quite similar to the lowest energy one-state structural solution, the chief difference being in the orientation of the terminal β-D-GalNAc residue. We conclude, therefore, that structures which are similar to the lowest energy one-state structural solution A'', represent the predominant conformer present in solution. The one-state procedure thus provides a reasonable and time efficient approach to structural analysis for this molecule.

This does not mean that minor conformers and/or motional averaging are unimportant. The presence of even small amounts of minor conformers or limited conformational averaging as is suggested by the lowest energy two state structural solution could significantly affect the observed cross relaxation rate. The inverse sixth power dependence of the cross relaxation rate on interproton distance serves to strongly weight contributions from conformers with short interproton distances. Therefore the presence of even small amounts of conformers with short interproton distances can exert a disproportionate amount of influence on the observed cross relaxation rate.

The heavy weighting of conformers with short interproton distances could explain the significant decrease in the molecular mechanics energies obtained for the predominant conformer when we relax the assumption that NMR cross relaxation data are satisfied by a single rigid conformer. By allowing the presence of an additional conformational state, we no longer require that a single conformer satisfy all of the constraints imposed by NMR cross-relaxation data. Those constraints which were satisfied via structural distortions which exacted a considerable penalty in terms of the molecular mechanics energy in the single state structural solution could now be satisfied via a combination of conformational states, a predominant conformer with a geometry which is closer to the minimum energy geometry and the presence of minor conformers with appropriately short interproton distances.

This is shown graphically in Figure 8 where we present a stereoview of the superposition of the dominant conformer from the lowest energy two-state structural solution, $D'A'$, the lowest energy one-state structural solution, A'' and the apparent global minimum energy structure D'. From these data, it is apparent that the predominant conformer from the lowest energy two-state structural solution exhibits smaller deviations from the apparent global minimum energy structure than the lowest energy one-state structural solution.

Conclusion

We have developed a protocol which relies on a combination of molecular mechanics calculations and distance constraint pseudoenergies to predict the solution conformation of biomolecules. When a simplified potential surface is used during the initial stages of the calculation, the final structures obtained upon convergence of the calculation are remarkably similar, even though the starting structures were grossly different. In other words, our protocol seems rather immune to some of the local minimum problems which plague molecular mechanics calculations. This protocol has enabled us to

a

b
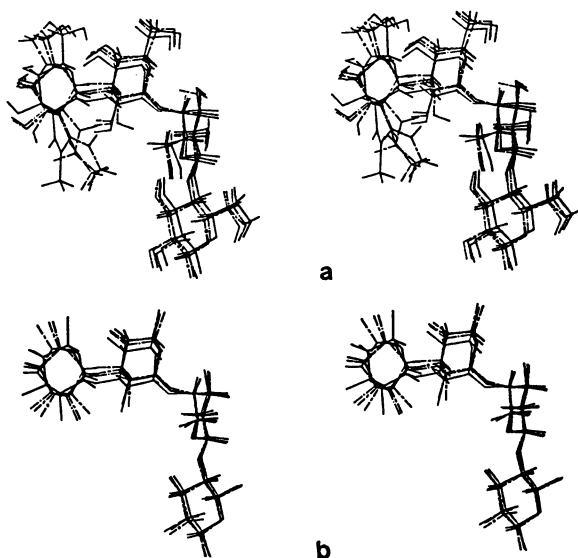
Figure 7. Stereoview of the superposition of the NMR refined structural solutions A″ (—·), B″ (—), and C″ (—·) and dominant conformer from the lowest energy two-state structural solution D′A′ (—) (a) with unconstrained exocyclic groups included in the comparison and (b) excluded from the comparison.

a

b
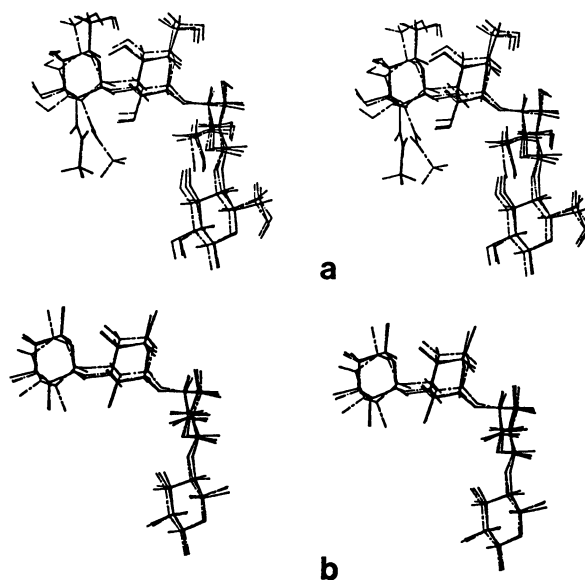
Figure 8. Stereoview of the superposition of apparent global minimum for globoside, structure D′ (—), the dominant conformer from the lowest energy two-state structural solution, D′A′ (—·) and the lowest energy one-state structural solution A″ (—·) (a) with unconstrained exocyclic groups included in the comparison and (b) excluded from the comparison.

generate viable one state descriptions of the globoside oligosaccharide headgroup that agree with the general "L" shape postulated to be important in receptor function.

We have explored the possibility that NMR data might be better represented through use of a model which permits interpretation of NMR data in terms of an average over discrete conformational states. Although this model led to structural solutions with very small occupation of minor conformational states and dominant conformers that were visually very similar to the lowest energy one-state structural solution, the molecular mechanics energies for these dominant conformers were significantly lower. These dominant conformers were also similar to the lowest energy structural solutions obtained in the absence of NMR distance constraints. These data indicate that only a narrow range of structures are likely to be present in solution and are consistent with the receptor and antigenic determinant functions proposed for globoside in that a certain amount of structural rigidity is likely to be necessary for these proposed biological functions.

Literature Cited

1. Braun, W.; Wider, G.; Lee, K. H. and Wüthrich, K. J. Mol. Biol. **1983**, 169, 921–948.
2. Havel, T. F. and Wüthrich, K. J. Mol. Biol. **1985**, 182, 281–294.
3. Williamson, M. P.; Havel, T. F. and Wüthrich, K. J. Mol. Biol. **1986**, 189, 377–382.
4. Wüthrich, K. Science **1989**, 264, 1516–1521.
5. Clore, G. M.; Gronenborn, A. M.; Brünger, A. T. and Karplus, M. J. Mol. Biol. **1986**, 186, 433–455.
6. Nilges, M.; Clore, G. M. and Gronenborn, A. M. FEBS Lett. **1988**, 229, 317–324.
7. Nilges, M.; Gronenborn, A. M.; Brünger, A. T. and Clore, G. M. Protein Eng. **1988** 2, 27–38.
8. Nilges, M.; Clore, G. M. and Gronenborn, A. M. FEBS Lett. **1988**, 239, 129–1336.
9. Folkers, P. J. M.; Clore, G. M.; Driscoll, P. C.; Dodt, T.; Köhler, S. and Gronenborn, A. M. Biochemistry **1989**, 28, 2601–2617.
10. Banks, K. M.; Hare, D. R. and Reid, B. R. Biochemistry **1989**, 28, 6996–7010.
11. Hare, D. R.; Shapiro, L. and Patel, D. J. Biochemistry **1986**, 25, 7445–7456.
12. Hare, D. R.; Shapiro, L. and Patel, D. J. Biochemistry **1986**, 25, 7456–7464.
13. Rao, B. N. N.; Dua, V. K. and Bush, C. A. Biopolymers **1985**, 24, 2207.
14. Bush, C. A.; Yan, Z.-Y. and Rao, B. J. Am. Chem. Soc. **1986**, 108, 6168–6173.
15. Yan, Z.-Y.; Rao, B. N. N. and Bush, C. A. J. Am. Chem. Soc. **1987**, 109, 7663–7669.
16. Brisson, J.-R. and Carver, J. P. Biochemistry, **1983**, 22, 3671–3680.
17. Brisson, J.-R. and Carver, J. P. Biochemistry, **1983**, 22, 3680–3686.
18. Cumming, D. A.; Dime, D. S.; Grey A. A.; Krepinsky, J. J. and Carver, J. P. J. Biol. Chem. **1986**, 261, 3208–3213.
19. Cumming, D. A. and Carver, J. P. Biochemistry **1987**, 26, 6664–6675.

20. Dabrowski, J.; Davrowski, U.; Bremel, W; Kordowicz, M. and Hanfland, P. Biochemistry **1988**, 27, 5149–5155.
21. Homans, S. W.; Dwek, R. A. and Rademacher, T. W. Biochemistry **1987**, 26, 6553–6560.
22. Homans, S. W.; Pastore, A.; Dwek, R. A. and Rademacher, T. W. Biochemistry **1987**, 26, 6649–6654.
23. Homans, S. W.; Dwek, R. A. and Rademacher, T. W. Biochemistry **1987**, 26, 6571–6578.
24. Scarsdale, J. N.; Ram, P.; Prestegard, J. H. and Yu, R. K. J. Comput. Chem. **1988**, 9, 133–147.
25. Kuntz, I. D.; Crippen, G. M. and Kollman, P. A. Biopolymers **1979**, 18, 939–957.
26. Marcus, D. M.; Nakai, M. A. and Kundu, S. K. Proc. Natl. Acad. Sci. U. S. A. **1981**, 78, 5406–5410.
27. Marcus, D. M.; Nakai, M. A.; Kundu, S. K. and Suzuki, A. Semin. Hematol. **1981**, 18, 63–71.
28. Leffler, H. and Svanburg-Eden, C. Infect. Immun. **1981**, 34, 920–924.
29. Prestegard, J. H.; Koerner, T. A. W.; Demou, P. C. and Yu, R. K. J. Am. Chem. Soc. **1982**, 104, 4993–4995.
30. States, D. J.; Haberkorn, R. A. and Reuben, D. J. J. Magn. Reson. **1982**, 48, 286–292.
31. Weiner, P. K. and Kollman, P. A. J. Comput. Chem. **1981**, 2, 287–303.
32. Broido, M. S.; Zon, G. and James, T. L. Eur. J. Biochem. **1985**, 150, 117–128.
33. Suzuki, E.; Pattabiraman, N; Zon, G. and James, T. L. Biochemistry **1986**, 25, 6854–6865.
34. Feurstein, B. G.; Pattabiraman, N. and Marton, L. J. Proc. Natl. Acad. Sci. U. S. A. **1986**, 83, 5948–5952.
35. Allinger, N. L.; Chang, S. H. M.; Glaser, D. H. and Hönig, H. Isr. J. Chem. **1980**, 20, 51–56.
36. Allinger, N. L. and Chung, D. Y. J. Am. Chem. Soc. **1976**, 98, 6798–6803.
37. Nørskov-Lauritsen, N. and Allinger, N. L. J. Comput. Chem. **1984**, 5, 326–336.
38. Fries, D. C.; Rao, S. T. and Sundaralingam, M. Acta. Crystallogr., **1971**, B27, 994–1005.
39. Bock, K.; Breimer, M. E.; Brignole, G. C.; Hannson, G. C.; Karlsson, K. A.; Larson, G.; Leffler, H.; Sammuelson, B. E.; Strömborg, N.; Eden, C. S. and Thurin J. J. Biol. Chem. **1985**, 260, 8545–8551.
40. Scarsdale, J. N.; Yu, R. K. and Prestegard, J. H. J. Am. Chem. Soc. **1986**, 108, 6778–6784.
41. Kay, L. E.; Holak. T. A.; Johnson, B. A.; Armitage, I. M. and Prestegard, J. H. J. Am. Chem. Soc. **1986**, 108, 4242–4244.
42. Mackay, A. L. Acta. Crystallogr. Sect. A: Found. Crystallogr. **1983**, A40, 165.
43. Jeffery, G. A. and Taylor, R. J. J. Comput. Chem. **1980**, 1, 99–109.
44. Tvaroska, I. and Perez, S. Carbohydr. Res. **1986**, 169, 389–403.

Chapter 16

Conformational Analysis of Oligosaccharides
Reconciliation of Theory with Experiment

J. P. Carver[1], D. Mandel[1], S. W. Michnick[1], A. Imberty[2,4], and J. W. Brady[3]

[1]Departments of Medical Genetics and Medical Biophysics, University of Toronto, Ontario, Canada M5S 1A8
[2]Laboratoire de Physicochimie des Macromolécules, Institut National de la Recherche Agronomique, B.P. 527, 44072, Nantes, France
[3]Department of Food Science, Cornell University, Ithaca, NY 14853-7201

The primary tools for the determination of three-dimensional structure for oligosaccharides are X-ray diffraction and NMR. The latter experimental technique makes use of the nuclear Overhauser effect (NOE) which yields information on the distances between hydrogens in the molecule. When these distances span a glycosidic linkage, information regarding the torsional angles about that linkage can be deduced. However, a major problem with this otherwise ideal approach is internal flexibility. Because the NOE builds up over hundreds of milliseconds, any flexibility on this time scale will result in fluctuations in trans-glycosidic H-H distances and influence the final NOE value. To deduce three-dimensional structure from NOE measurements, one must, therefore, be able to model the internal flexibility of the oligosaccharide. It is straightforward to calculate the ensemble average steady state NOE and NOESY intensities once one has generated the ensemble. Thus by a careful comparison of quantitative NOE measurements with calculated values, it is possible to evaluate the adequacy of the potential energy functions used to generate the ensemble. To date we have been unable to find a set of potential energy functions that allows us to predict adequately experimental NOE values.

One of the more intriguing current questions in biology is the quest for the biological role of the carbohydrate components of glycoproteins and glycolipids. Cells of different lineages invest a considerable fraction of their metabolic energy into the complex biosynthetic pathways which generate these compounds with high specificity. Clearly these energetically expensive processes have been conserved and even elaborated during evolution - but why? Our laboratories have taken the approach that the information content of

[4]On postdoctoral leave from CERMAV, Grenoble, France

the oligosaccharide moieties lies in their three-dimensional structures. Thus, alterations in three-dimensional structure constitute alterations in the signals encoded by these molecules. Clues to their functions, therefore, should be available through a correlation of biological status with modifications in primary structure that lead to alterations in three-dimensional structure. Our focus in this article is on methods for the derivation of three-dimensional structure information from experimental NMR measurements.

The "experiment-of-choice" for the exploration of the three-dimensional structures of oligosaccharides is the one-dimensional NOE. This NMR experiment is easily performed with modern instruments and can be readily quantified. However, the deduction of three-dimensional structure directly from measured NOE's is frustrated at two levels. First, although the geometrical dependence of the NOE effect is well understood (1), derivation of inter-hydrogen distances from measured NOE's requires that a complete set of NOE's between all hydrogens in the molecule be measured. This is rarely, if ever, possible, particularly for oligosaccharides where the vast majority of resonances in the ^{1}H spectrum are crowded into 0.5 ppm. The errors associated with the use of a partial set of NOE's vary greatly and therefore it is difficult to justify such approximations a priori. Fortunately it is relatively simple to calculate all the NOE's (observable or not) from any particular molecular geometry (2). Thus it is straightforward to explore torsion angle space for angles at which the predicted NOE's are in agreement with those observed (2). However, such an approach assumes that the molecule adopts only one fixed three-dimensional structure. This is where the second complication comes in.

Considerable internal flexibility about the glycosidic linkage is clearly evident from calculations of the Boltzman distribution of molecular structures using the potential energy surfaces currently employed for oligosaccharides (2-6). Similarly the force fields of Brady (7) and of Rasmussen (8) when used in molecular dynamics calculations (9, 10 and calculations below), reveal considerable flexibility about the glycosidic linkage. When trajectories are extended over time periods longer than 10 ps, even the force field used by Homans et al. (11, 12) has been found to generate large torsional angle fluctuations (Dwek, R.A., personal communication).

These results suggest that we must model glycosidic torsion angle flexibility and incorporate it into the interpretation of trans-glycosidic NOE measurements. In our earlier work (2,13,14) we were aware of this flexibility but, since we were able to find single, low-energy, conformations consistent with the NOE data, we assumed that the solvation of the oligosaccharides (neglected in the calculations referred to above) in reality restricted the flexibility of the molecules. However, in several subsequent cases these single conformations had potential energies which were quite high relative to the global minimum on the surface (15). We therefore became suspicious that these were what Jardetzky has termed "virtual" conformations (16) and these suspicions were confirmed when we found several examples where no single conformations compatible with the data existed (5,15; in preparation).

The question of how to include internal flexibility into the interpretation of NOE's has been discussed by Noggle and Schirmer (1) in their classic book. Provided that the internal motions are fast compared to the longitudinal relaxation rates and slow with respect to the correlation time, then the NOE effect will be a function of the ensemble average values of $(1/r_{ij})^6$, where r_{ij} is the distance between the ith and jth hydrogens. Since the longitudinal relaxation times for hydrogens in oligosaccharides are on the order of hundreds of milliseconds to seconds and the rotational correlation times are usually about 100 ps, there are at least six orders of magnitude between these limits and it seemed reasonable to assume initially that internal motions lie between these limits.

Thus, in recent studies (5,6), the ensemble average NOE's have been calculated by using an "ensemble average relaxation matrix". This matrix was generated by replacing the inverse sixth power of the inter-hydrogen distances, appearing in the expressions for the bulk- and cross-relaxation terms, by their ensemble averages. The latter were calculated as the Boltzman weighted sums over all the states of the values of $(1/r_{ij})^6$. When these ensemble average NOE's were compared to observed values, they were found to give closer agreement than single geometries corresponding to "preferred" three-dimensional structures (Table I) (5,6,15). These studies used statistical mechanics methods based on potential energy surfaces derived from the rotation of fixed hexose rings about the glycosidic torsion angles. Quite different NOE's are predicted if different potential energy surfaces are used (Table I).

Molecular mechanics calculations for disaccharides (7,18; Imberty, A.; Tran, V.; Perez S. J. Comp. Chem., in press; also see below), not surprisingly, have shown that the assumption of rigid geometry leads to artificially steep potential energy surfaces. Such calculations demonstrate that permitting flexibility in bond lengths and angles further increases the magnitude of torsional angle fluctuations and thus has an important impact on calculated ensemble average properties. Because molecular dynamics force fields generally include this bond length and bond angle flexibility and also since dynamics calculations are expected to efficiently sample the statistically significant regions of conformational space, we decided to examine the properties of NOE's calculated from molecular dynamics generated ensembles. This article represents a "progress report" describing results to date. More complete exploration of this approach will be reported elsewhere.

Recently, molecular dynamics calculations have been performed with explicit inclusion of water molecules. Although the frequencies and residence times are altered, the amplitudes of conformational transitions about the glycosidic bond appear not to be significantly changed in these preliminary studies (Brady J.W., unpublished results; 26). Thus solvation does not appear to restrict glycosidic angle flexibility but does have a strong damping effect and influences population distributions. The direct acknowledgment of flexibility in the interpretation of NOE's is therefore absolutely essential.

Methods

Calculations were performed for the disaccharide Manα(1–3)Manβ (see Figure 1 for the structure) which is one of the key linkages in the N-linked oligosaccharide family of structures.

Relaxed Map. A relaxed or adiabatic potential energy map was calculated for the Man(α1–3)Manβ linkage using the MMP2(85) version of the original molecular mechanics programme MM2 (19). The procedure was as previously described for Man(α1–3)Manα (Imberty, A.; Tran, V.; Perez S. J. Comp. Chem., in press) except that maps were calculated for only four combinations of the C5C6 rotamers: GT-GT, GT-GG, GG-GT and GG-GG. These four maps were merged into a single map (Figure 2) by choosing the lowest energy found for each phi/psi pair (phi ≡ H1C1O1C'$_x$ and psi ≡ C1O1C'$_x$H'$_x$). On the latter map four minima were found which are labeled MA, MB, MC and MD in Figure 2. The initial geometry used was taken from the crystal structure of Man(α1–3)Man(β1–4)GlcNAc determined by Warin et al. (20).

Molecular dynamics. Calculations were performed using the programme CHARMM (21) with a force field for saccharides based on the PEF422 force field of Rasmussen (8). Newton's equations of motion were integrated using a Verlet algorithm with time steps of 1 fs. Over a period of 5 ps, the temperature was raised to 300 K in 15 degree increments with velocity rescaling every 250 fs. This was followed by a 15 ps equilibration period during which the velocities were periodically rescaled. A 20 ps dynamics trajectory was then generated, during which average (over 250 fs) temperature fluctuations of less than 6 degree were observed. All CH bond lengths were constrained to within an error tolerance of 1×10^{-10} using the SHAKE (25) algorithm of CHARMM (21). Four trajectories of 20 ps were started from each of the four minima found in the relaxed MMP2(85) map (MA,MB,MC and MD in Figure 2), using a different random number seed for each run. The starting geometry for each calculation was derived from either the carbohydrate topology and parameter files (Brady, J.W., unpublished results) or from the final coordinates of runs which ended up in the appropriate minimum. These geometries were then refined by minimizing the energy using the steepest descent and conjugate gradient algorithms of CHARMM (21). In total sixteen independent 20 ps trajectories were thus accumulated. Transitions from minima corresponding to the initial conformation to other minima on the surface were observed during the equilibration period in eleven cases so that the starting geometries for the sixteen trajectories corresponded to minima MA,MB,MC and MD in seven, seven, zero, and two cases, respectively.

NOE Calculations. Coordinates were recorded every 10 fs. From the time series of molecular geometries the running averages of $(1/r_{ij})^6$ for all inter-hydrogen distances were calculated. These running averages were output every picosecond and used to calculate the ensemble average NOE's (5) using the programme DYNAMO, developed in Toronto. For the relaxed maps, a 20°x20° grid was used to calculate the ensemble average NOE's from the four maps corresponding to GT-GT, GT-GG, GG-GT and GG-GG. These calculations were performed in

Nantes following the general methodology of Cumming and Carver (5). All the NOE calculations were performed for a previously described (22,23) hexa-deuterio version of Manα1–3Manβ because the observed NOE's do not overlap and are much larger for this disaccharide than those for the undeuterated version. Thus the NOE's for the deuterated compound can be much more accurately measured.

NOE Measurements. The one-dimensional NOE data were collected at 300 MHz on a Bruker AM-300 NMR spectrometer operating at 300 K. Because the relaxation times for the protons of the hexa-deuterated compound ranged from 0.4 s to 3.8 s, delays of 20 s. were used between scans. Values for the ^{13}C T_1's were also measured and found to range from 0.32 s to 1.5 s; all values were consistent with a rotational correlation time of 1.1×10^{-10} s.

Results and Discussion

Relaxed map. The relaxed map (Figure 2) for Man(α1–3)Manβ shows little difference from that calculated for Man(α1–3)Manα by Imberty et al. (Imberty, A.; Tran, V.; Perez S. J. Comp. Chem., in press) but is strikingly different from those previously obtained with the HSEA and HEAH potentials (6,17) in that two new low energy regions (MC and MD) have appeared. The MM2(85) potentials allow more conformational states to be reached at lower energies; thus the surface is enlarged and low energy pathways appear between minima. The minimum at MB corresponds to that found with the HSEA potential and is close to that found in the crystal structure of Warin et al. (20). The minimum at MA is in the region of the hydrogen bonded structure originally suggested by Dwek's group (24) and is the global minimum obtained with the HEAH potential (6,17). The potential energy surface calculated using the PFOS potentials (as described in Imberty, A.; Tran, V.; Perez S. J. Comp. Chem., in press) also showed all four minima but the energy barriers between minima are much higher than in the MM2(85) maps.

Molecular Dynamics. As has been found by others using the Rasmussen PEF422 force field with CHARMM (9,10 and Yan, Z. Y.; Bush, C.A. Biopolymers, in press), the hexose ring geometries were stable in the chair form over all the trajectories. Some trajectories displayed many transitions in both phi and psi (Figure 3A) while others were restricted to different regions of the phi/psi map (Figures 3B&C). The superposition of all sixteen trajectories is shown in Figure 3D. A considerable portion of the phi/psi map has been explored during the total of 320 ps corresponding roughly to the 8 Kcal contour on the relaxed potential energy map (Figure 2). Thus the two force fields show good agreement. Plots showing the time evolution of phi, psi and both omega values are found in Figures 4A–D for a case with many transitions (trajectory shown in Figure 3A). Transitions are more frequently seen in psi than in phi; however large variations in magnitude ($\pm70°$ in phi and $\pm100°$ in psi) are found for both angles. These results appear to be in disagreement with the conclusions of Homans et al. (11) from one 10 ps trajectory for the same linkage using a different force field [it should be noted that alternate definitions of phi and psi are used in ref 11]. However, the last two ps of that trajectory showed a
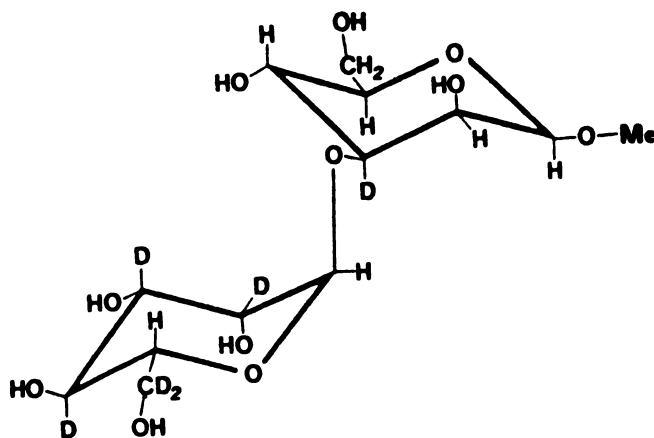
Figure 1: The structure of the synthetic deuterated disaccharide Man(α1-3)Manβ showing the sites of deuteration by the letter "D".
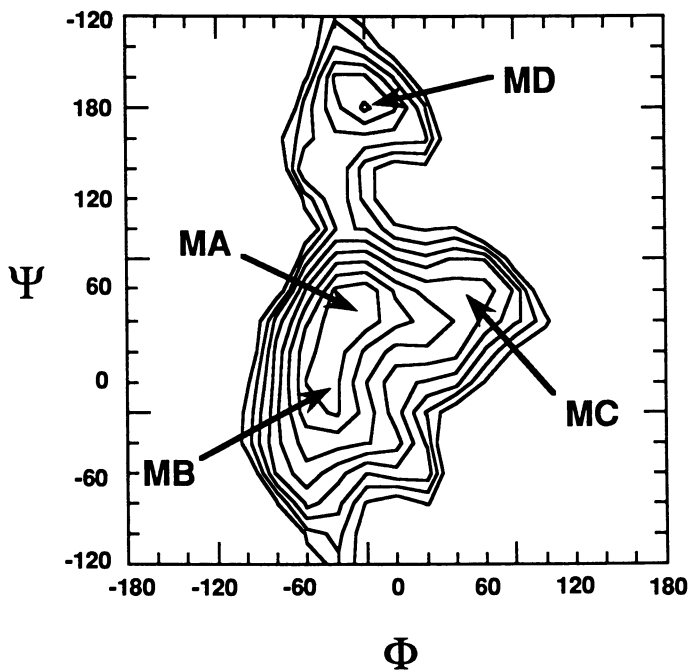
Figure 2: The composite "relaxed" potential energy surface for Man(α1-3)Manβ calculated by using the molecular mechanics programme MM2(85).

A.

B.
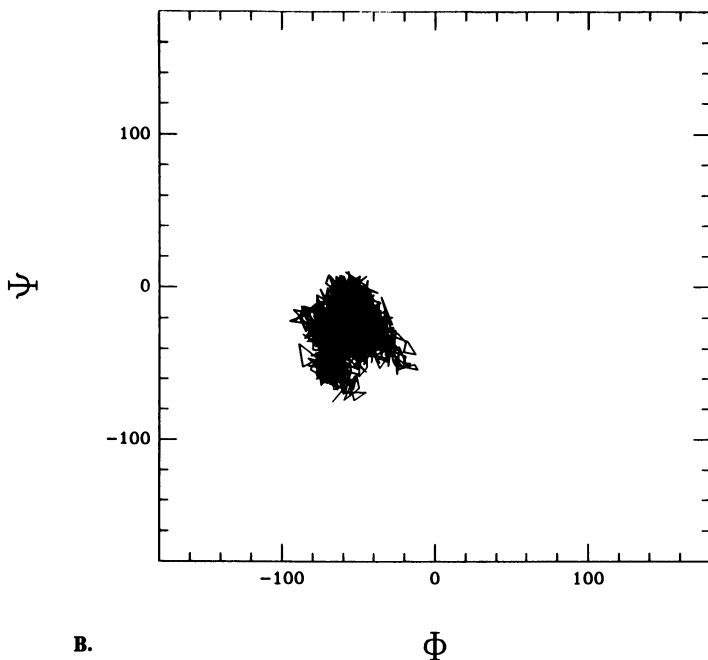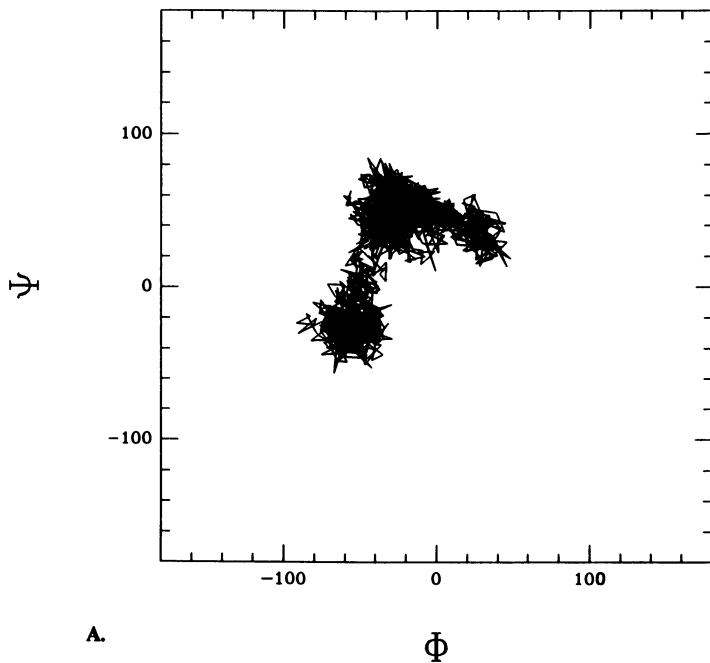
Figure 3: Three examples out of the sixteen independent 20 ps trajectories are shown in A-C. Part D shows the superposition of all sixteen trajectories. The time axis is in units of picoseconds.

C.
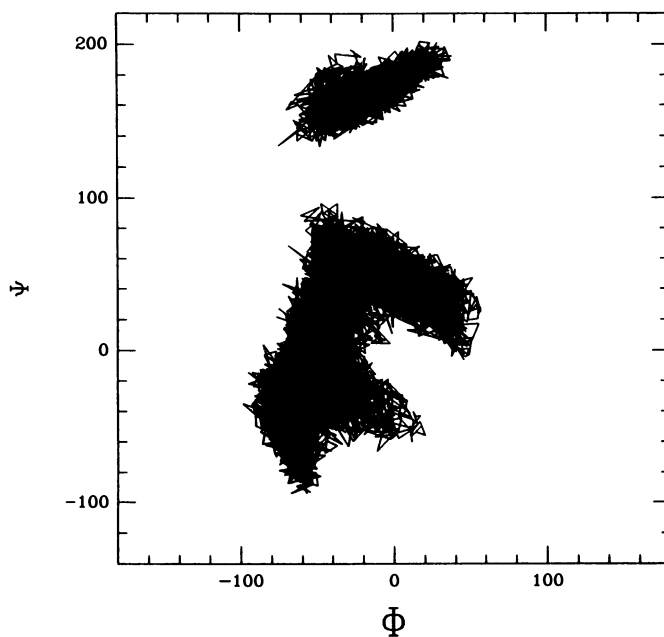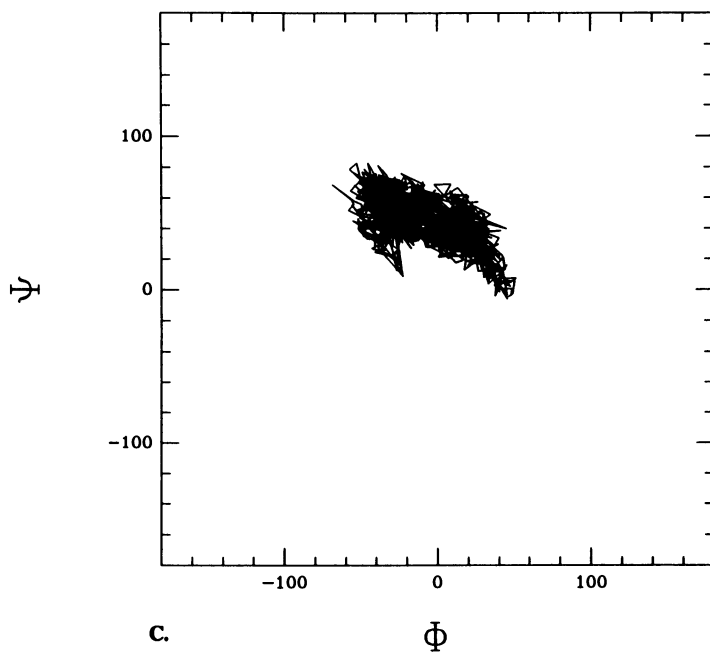
Figure 3 continued.
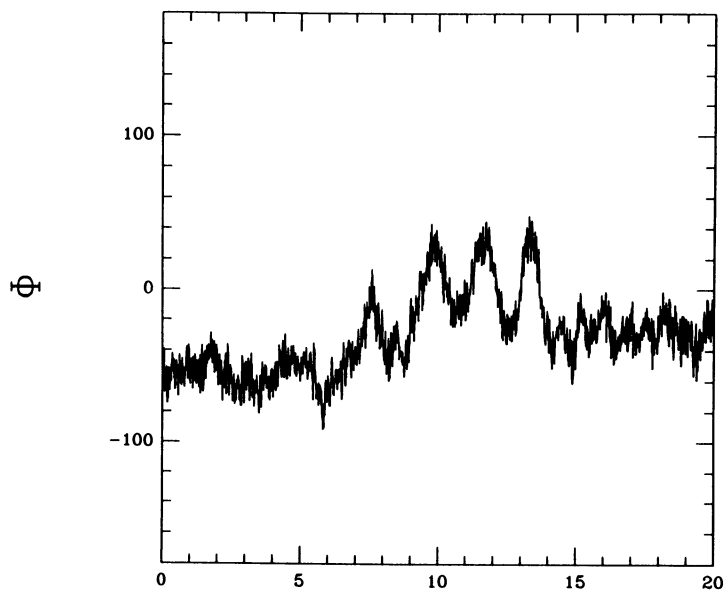
A.
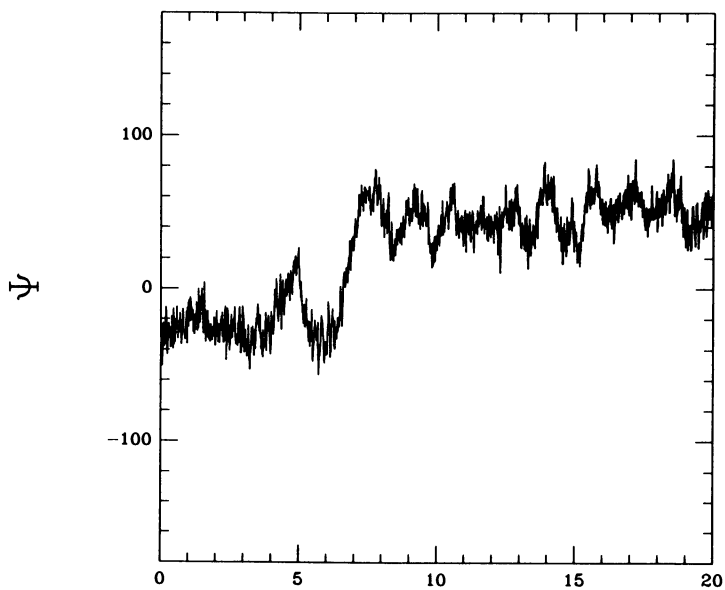
B.   TIME

Figure 4: The time evolution of phi, psi and omega (for the α- and
the β-linked mannose residues) are shown in panels A-D,
respectively, for the trajectory shown in Figure 2A. The time
axis is in units of picoseconds.
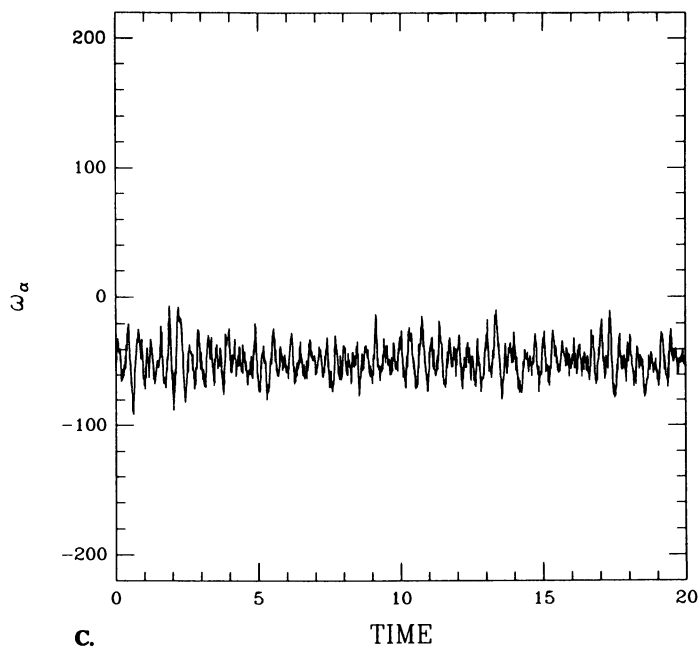
C.
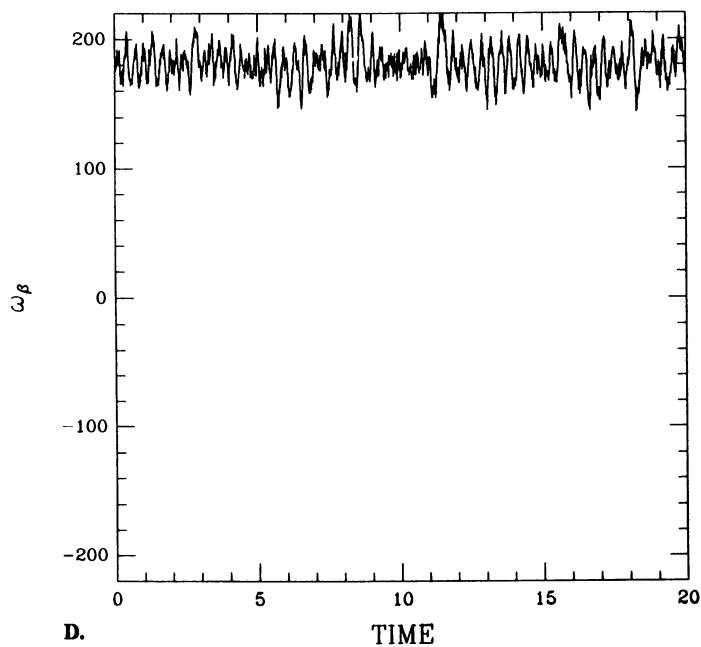
D.

Figure 4 continued.

65° shift in phi (from -33° to +32°, in our notation) accompanied by
a 20° shift in psi (from +21° to +41°, in our notation). These
shifts are consistent with a transition to MC from somewhere between
MA and MB. In fact, recent calculations which have extended the
trajectory to much longer times, result in the observation of
multiple transitions (R.A. Dwek, personal communication). Thus
there is now general agreement that multiple transitions occur for
the Man(α1-3)Manβ linkage over the time periods needed for NOE
measurements.

Nuclear Overhauser Effects. The time evolution of the NOE's
calculated for hexa-deuterated Man(α1-3)Manβ from each of the
sixteen trajectories are shown in Figure 5. In each case the
ensemble average NOE value derived from the complete 320 ps ensemble
is shown as a horizontal dashed line.   In Figure 5A are shown the
calculated NOE's for βMan H2 upon irradiation of αMan H1; whereas,
in Figure 5B, the calculated NOE's for αMan H5 upon irradiation of
βMan H2 are depicted. The NOE values for each of the trajectories
start out at very different values because the initial geometries
differed. For those trajectories that displayed several transitions
between minima, the calculated NOE's drift slowly towards the 320 ps
average. For those trajectories that remained localized to one
minimum during the 20 ps period, the NOE's are essentially constant.
Since the NOE values associated with different trajectories are
barely converging towards the 320 ps ensemble average, it is clear
that 20 ps is not nearly long enough for statistically significant
sampling of the conformational ensemble. It should be emphasized
that although the composite 320 ps trajectory shown in Figure 3D
appears to have sampled a large region of torsion angle space, in
order for this to be a statistically significant sample the
trajectory must revisit the low energy regions a sufficient number
of times to give a true Boltzman distribution. Thus times even
longer than 320 ps may be needed for proper statistical sampling of
torsional angle space.
      It is also worth pointing out that some NOE's are more
sensitive to conformational flexibility than others. When linkage
flexibility results in motions which cause a particular
inter-hydrogen distance to fluctuate widely, then the NOE associated
with that pair of hydrogens will be very sensitive to the nature of
the potential energy surface used to simulate that motion. In
contrast when the motions result in very little alteration in an
inter-hydrogen distance, then the associated NOE's will be
insensitive to the potential functions used.

Comparison of Molecular Dynamics with other Methods. In Table I are
shown the results of the use of a variety of methods for the
calculation of relative and absolute NOE's for the Man(α1-3)Manβ
linkage. The values of the relative NOE's derived from the dynamics
calculations are in better agreement with experiment than those
derived from any other method (for example, the NOE to βM-H4 is 1.2
s. which is less than two standard deviations from the observed
value of 0.7±0.2 s., whereas the next closest value is that from the
HSEA surface at 2.0, more than six standard deviations from the
observed value). However, the absolute NOE's are overestimated
considerably. There are several possible explanations for this

A.

B.
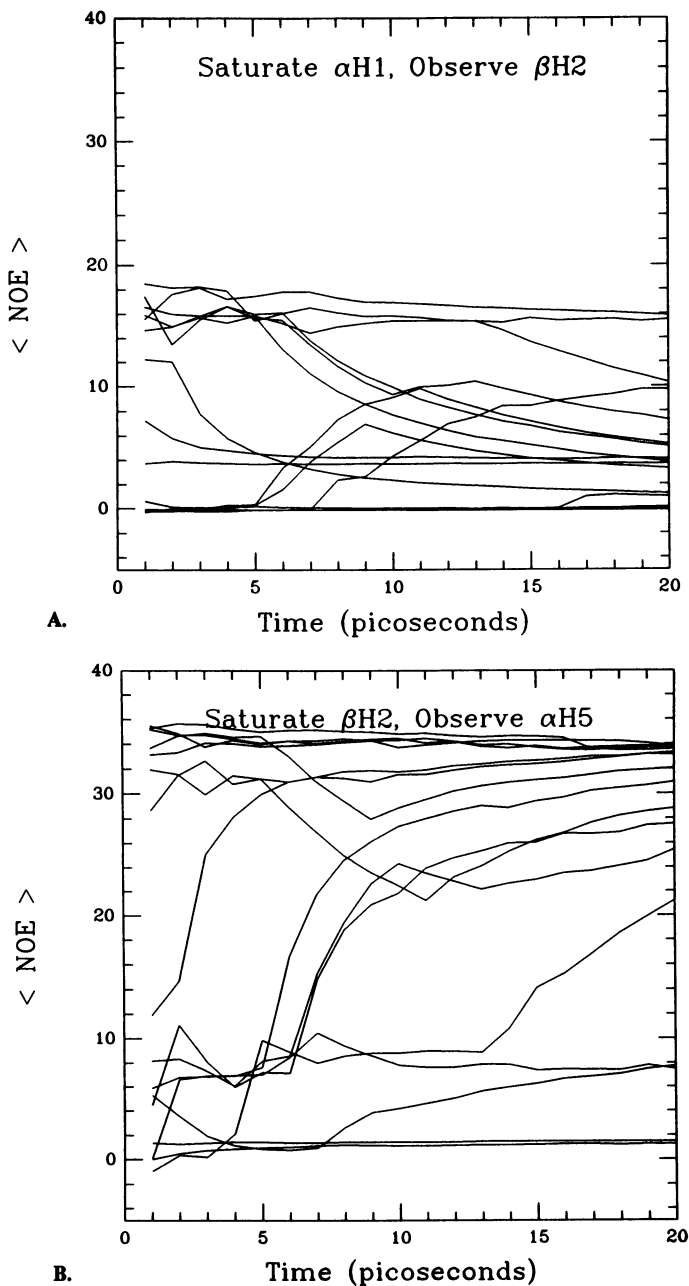
Figure 5: The time evolution for all sixteen trajectories (calculated as described in "Methods") of two different calculated NOE's. A: the NOE on the resonance of the H2 of Manβ when the resonance of the H1 of Manα(1-3) is irradiated. B: the NOE on the resonance of the H5 of Manα(1-3) when the resonance of the H2 of Manβ is irradiated.

Table I

Comparison of Observed and Calculated NOE's for Man(α1-3)Manβ

Irradiation of H1 of Man(α1-3)

	Relative NOE[a]		Absolute NOE	
	[βM-H2]	[βM-H4]	[βM-H2]	[βM-H4]
Method				
HSEA	1.2	2.0	1.2	2.0
HEAH	11	3.2	5.7	1.6
PFOS	13	1.7	7.8	1.0
PFOS-H	18	0.36	20.3	0.4
MMP2(85)	8.2	1.1	7.4	1.0
MD(320ps)	1.1	1.2	5.9	6.5
Obs'd	1.8±0.4	0.7±0.2	1.8±0.4	0.7±0.2

[a]Relative to the NOE on αM-H5.

discrepancy which will be the subject of a future communication. Briefly, however, one possible source for this discrepancy is that we have not sufficiently sampled torsion angle space; we are currently exploring this possibility by extending the trajectories to longer times. Possible problems with the calculated NOE's could arise from the neglect of the presence of the isopropyl group and of a small residual of hydrogens at positions 6 and 6' of the αMan residue. However, if these were important effects then one would not expect both NOE's to be equally affected. The agreement between the observed and calculated relative NOE's suggested that the discrepancy arises from a constant factor affecting all NOE's equally. A possible cause for a systematic error in the calculated absolute NOE's would be an incorrect choice for the spectral density function used to calculate the relaxation matrix. Classical spectral density functions ($\underline{1}$) were used in these calculations together with the rotational correlation time derived from the ^{13}C T_1's. These are reasonable choices if the assumption that internal motions are slow with respect to the rotational correlation time is valid. However, the dynamics trajectories suggest that this is not so. Transitions between local minima on the potential energy surface appear to occur with a frequency of about one every 10 ps. If this observation corresponds to the true situation in solution then the internal motions are an order of magnitude faster than the rotational correlation time. Under such circumstances, the spectral density function used in these calculations is incorrect. This aspect requires further investigation, particularly once the data from dynamics calculations specifically including water become available.

Conclusion

Molecular dynamics using the Rasmussen force field PEF422 in vacuo predicts that the Manα(1-3)Manβ linkage is highly flexible, exhibiting excursions of $\pm 70°$ in phi and $\pm 100°$ in psi. The ensemble average relative NOE values calculated from the 320 ps of combined trajectories are in close agreement with those observed and are in better agreement than those obtained from a previous statistical mechanics approach ($\underline{15},\underline{16}$). Clearly, an insufficient length of time was examined in previous dynamics studies ($\underline{11},\underline{12}$) which concluded that the linkage is fixed. Furthermore, trajectories of several hundred picoseconds will be needed before a statistically significant exploration of conformational space has occurred for this force field and this linkage. Thus extensive dynamics calculations will be required to predict adequately the ensemble NOE's by this approach.

Acknowledgments

Literature Cited

1. Noggle, J.H.; Schirmer, R.E. The Nuclear Overhauser Effect;
    Academic Press: New York, 1971.
2. Brisson, J.-R.; Carver, J.P. Biochemistry 1983, 22, 1362.
3. Lemieux, R.U.; Bock, K. Arch. Biochem. Biophys. 1983, 221, 125
4. Tvaroska, I.; Perez, S. Carbohydr. Res. 1986, 149, 389.
5. Cumming, D.A.; Carver, J.P. Biochemistry 1987, 26, 6664.
6. Carver, J.P.; Cumming D.A. Pure & Appl. Chem. 1987, 11, 1465.
7. Ha, S.N.; Madsen, L.J.; Brady, J.W. Biopolymers 1988, 27, 1927.
8. Rasmussen, K. Acta Chem. Scand. 1982, A 36, 323.
9. Brady, J.W. J. Am. Chem. Soc. 1986, 108, 8153.
10. Brady, J.W. Carbohydr. Res. 1987, 165, 306.
11. Homans, S.W.; Pastore, A.; Dwek, R.A.; Rademacher, T.W.
    Biochemistry 1987, 26, 6649.
12. Homans, S.W.; Edge, C.J.; Ferguson, M.A.J.; Dwek, R.A.;
    Rademacher, T.W. Biochemistry 1989, 28, 2881
13. Brisson, J.-R.; Carver, J.P. Biochemistry 1983, 22, 3671.
14. Brisson, J.-R.; Carver, J.P. Biochemistry 1983, 22, 3680.
15. Cumming, D.A.; Shah, R.N.; Krepinsky, J.J.; Grey, A.A.; Carver,
    J.P. Biochemistry 1987, 26, 6655.
16. Jardetzky, O. Biochim. Biophys. Acta 1980, 621, 227.
17. Carver, J.P.; Michnick, S.W.; Imberty, A.; Cumming, D.A. In
    Carbohydrate Recognition in Cellular Function (Ciba Foundation
    Symposium 145); Wiley: Chichester, UK, 1989; p 6.
18. French, A.D. Biopolymers 1988, 27, 1519.
19. Burkert, U.; Allinger, N.L. In Molecular Mechanics 1982, ACS
    Monograph 177, American Chemical Society, Washington D.C.
    MMP2(85) is available from the Quantum Chemistry Program
    Exchange, Department of Chemistry, Indiana University,
    Bloomington Indiana 47401.
20. Warin, V.; Baert, F.; Fouret, R.; Strecker, G.; Fournet,B.;
    Montreuil, J. Carbohydr. Res. 1979, 76, 11.
21. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.;
    Swaminathan, S.; Karplus, M. J. Comput. Chem. 1983, 4, 187.
22. Dime, D.S.; Rachaman, E.; Dime, C.E., Grey, A.A., Carver, J.P.;
    Krepinsky, J.J. J. Labelled Cpds. Radiopharm. 1986, 24, 725.
23. Cumming, D.A.; Dime, D.S.; Grey, A.A.; Krepinsky, J.J.; Carver,
    J.P. J. Biol. Chem. 1986, 261, 3208.
24. Homans, S.W.; Dwek, R.A.; Fernandes, D.L.; Rademacher, T.W.
    FEBS Lett. 1982, 150, 503.
25. van Gunsteren, W.F.; Berendsen, H.J.C. Molec. Phys. 1977, 34,
    1311.
26. Edge, C.J.; Singh, U.C.; Bazzo, R.; Taylor, G.L.; Dwek, R.A.;
    Rademacher, T.W. Biochemistry 1990, 29, 1971.

Chapter 17

Modeling of Interactions of Polysaccharide Chains

Application to Crystalline Polymorphism of Starch Granules

Serge Pérez[1], A. Imberty[1], and Raymond P. Scaringe[2]

[1]**Laboratoire de Physicochimie des Macromolécules, Institut National de la Recherche Agronomique, B.P. 527, 44026, Nantes, Cédex 03, France**
[2]**Research Laboratory, Eastman Kodak Company, Rochester, NY 14650**

This paper describes a method for predicting the packing relationship of two polysaccharide chains and its use in studying polymorphism in starch. Given a rigid model of an isolated double helix, its interaction with a second double helix is studied at varied helix-axis translations and mutual rotational orientations while keeping the helices in van der Waals contact. The stability of each structure is evaluated by an energy calculated using atom-atom potentials that includes compensation for hydrogen bonding. Models for starch were based on the fiber repeat distance on fiber diffraction patterns and are double-helices composed of left-handed single strands related by two-fold rotational symmetry. Two stable relationships were found for both the parallel and antiparallel models. The structure predicted to be most stable corresponds to a duplex of parallel double-helices as found in both the crystalline A and B allomorphs. From this result, an explanation of the transition from B to A is proposed.

Over the years, modeling of carbohydrates has emphasized intramolecular rather than intermolecular structures. The same holds true in the study of synthetic polymers and polypeptides. Only one such study for carbohydrates comes to mind (1) where the unit cell dimensions and symmetry were not used. Even there, a volume constraint was used, limiting the possible structures. When such constraints are used, one does not obtain an explanation for why the crystal structure is the stable form.

We hope to understand and develop general rules for the stability of some inter-helix arrangements. Methods for investigating the inter-helix structure and energy through non-bonded forces have been suggested by a number of workers (2-7). Those procedures minimize the inter-helix energy. Our method (8,9) moves the helices as close to each other as is possible without causing inter-penetration of the van der Waals radii of atoms of the two different helices. After the helices are positioned to the shortest interhelical distance for a given rotation and helix-helix translation, the energy is calculated. This technique takes considerably less computer time than methods involving energy minimization.

In the present work, we extend the method to compensate for the hydrogen bonds present in carbohydrates. The hydroxylated character of carbohydrate polymers influences between-chain interactions through networks of hydrogen bonds that occur during crystallization. Frequently, several possible attractive interactions exist that lead to different packing arrangements, and several allomorphic crystalline forms have been observed for polysaccharides such as cellulose, chitin, mannan and amylose. The situation is even more complex when water or other guest molecules are present in the crystalline domains. Another complication is that polysaccharide polymorphism includes different helix shapes as well.

For the present work, we studied the polymorphism of starch with our extended method. Starch, an energy reserve for green plants and a major food carbohydrate, has many practical applications. Native starch exhibits two different diffraction patterns that depend on the botanical origin: A-type in cereal starches and B-type in tuber starches (10). In both, diffraction is thought to arise mainly from the short chains that are connected at branch points of the amylopectin component of starch (11). The short chains have 12 to 20 D-glucose residues linked α(1-4). Fiber diffraction studies (12,13) demonstrated that both forms have the same 1.05 nm repeat distance along the helix axis. Recent reinvestigation of their crystal structures (14,15) established that the same conformation of the individual amylosic strands (a nearly perfect left-handed, six-fold helix repeating in 2.1 nm) exists in both allomorphs. Through the rotational symmetry of the double-helix, the repeat distance is halved to 1.05 nm. Such double-helices were first proposed in 1972 by Kainuma and French (16).

The recent reinvestigations showed that the double-helices of A and B starch are packed in a parallel fashion. They are stabilized mainly by numerous van der Waals interactions and by hydrogen bonding. The important differences between the two structures lie in the amount of water present, and positioning of the helices to accomodate the different amounts of water. Under certain

conditions of heat and moisture, the B form can be converted to the A form. This solid-state conversion has been observed in fibers (12) and in intact starch granules (17).

Methods

Our stategy is to first construct models of single-stranded helices that repeat in 2.1 nm and decide whether they are left- or right-handed. This is done with a traditional Ramachandran plot of energy vs. Φ and Ψ torsion angles, overlaid whith contours of iso-n and iso-h. A single strand is then used to generate the second strand of a double-helix through two-fold rotation (for every atom with coordinates of x, y and z there is a new one at -x, -y and z). The resulting double-helix is then placed in our new program which generates a second double-helix and tests the interactions between the two double-helices.

Nomenclature. A fragment of amylosic chain (maltose) is shown in Figure 1, with labels on the atoms and torsion angles of interest. The relative orientation of two contiguous (1-4) linked α-D-glucose residues is given by torsional angles Φ and Ψ which are defined by the 4 atom sequences O-5 - C-1 - O-1 - C-4' and C-1 - O-1 - C-4' - C-5', respectively. Other conformational parameters are the orientations of the primary hydroxyl groups around C-5 - C-6 bonds. This orientation is referred to as either gauche-trans, gauche-gauche or trans-gauche. In this terminology, the torsion angle O-5 - C-5 - C-6 - O-6 is stated first, followed by the torsion angle C-4 - C-5 - C-6 - O-6 (18). The sign of the torsion angles agrees with the rules recommended by the IUPAC-IUB Commission of Biochemical Nomenclature (19,20). Helical arrangements are described in terms of a set of helical parameters (n,h); n is the number of residues (i.e. backbone glucose 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. Arbitrarily, a right-handed helix will have positive h value; conversely, negative values of h will designate left-handed helices. Whenever the values h = 0 or n = 2 are intercrossed, the screw sense of the helix changes to the opposite sign.

Modeling the Single-Strand Helix. Starting geometry for the glucose residue in its $^4C_1$ conformation was taken from the crystal structure of α-D-Glucose monohydrate (21). The reported positions for hydrogen atoms were not used since they are known to be poorly determined by x-ray diffraction studies. Instead, the positions of hydrogen atoms that are attached to carbons were generated using a C-H bond length of 0.1 nm and a bond vector related appropriately to the C-C and C-O bond vectors. Hydroxylic hydrogen atoms were not considered.

Program PFOS (22) calculated the energy of the maltose disaccharide at the values of Φ and Ψ, using the force-field described in reference 23. The value of the C-1 - O-1 - C-4' valence angle was 117° and Φ and Ψ were stepped in 5° increments. Iso-energy contours are drawn at 1 kcal/mol spacings with respect to the minimum energy. The contours of iso-n and iso-h were derived with the algorithm reported by Gagnaire et al. (24). This map, based only on a disaccharide, is not accurate near h=0 for a polysaccharide because helices with small values of h would suffer steric conflicts between adjacent turns. Since the helices in the native forms of starch are very extended, this is not a problem in this case. Because of the exact geometry of the selected glucose residue, helices with large values of n and h are favored. Other residue geometries must be used to construct satisfactory models of known allomorphs such as V_6 amylose.

Chain Pair Modeling. In the following analysis, we assume that the chains are regular helices, i.e. that they have screw symmetry, with a repeat distance, t. In a perfect crystal, such chains must either be parallel or antiparallel. Four interhelical parameters are required to define the geometric orientation of chain A relative to chain B (see Figure 2). The parameters and their ranges are :

    μA : rotation of A about its axis (0° to 360°)
    μB : rotation of B about its axis (0° to 360°)
    Δx : normal distance between the helix axes of A
         and B (no limit)
    Δz : translation along the helix axis of one chain
         relative to the other (0 to t, nm)

Such a set of interhelical parameters relates directly to the symmetry operations which are found in crystal structures.

    : μA ≠ μB, represents the case where chain A and chain B are not related by any symmetry operation. Both independent chains would be needed to define the asymmetric unit of a crystal.
    : μA = μB, represents the case where chain B is derived from chain A by a pure translational symmetry element.
    : μA = μB + 180° and Δz = 0, represents the situation where the two chains are parallel and related by a two-fold operation. A two-fold screw-axis would be described by μA = μB + 180° and Δz = t/2.
    : μA = -μB + 180° and Δz = 0, represents the situation where the two chains are antiparallel and related by a two-fold operation. For a two-fold screw axis, the relationship is μA = -μB + 180° and Δz = t/2.
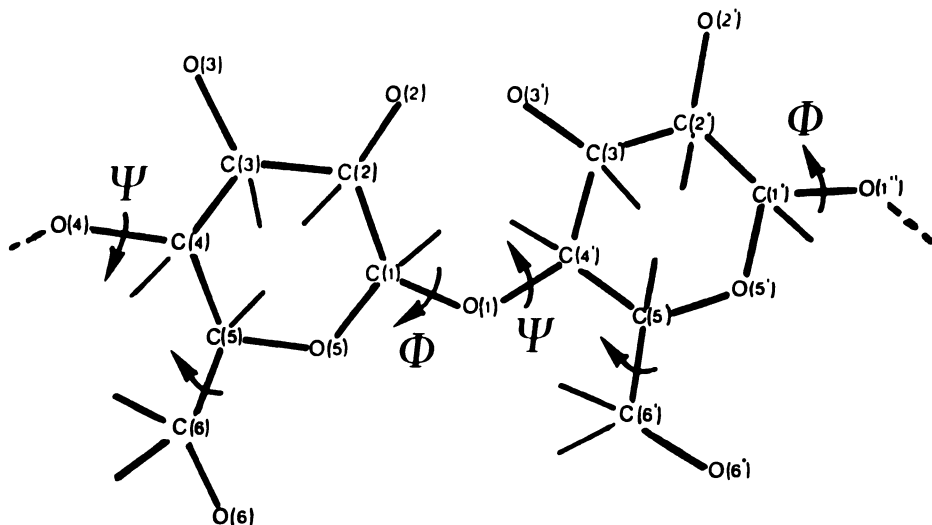
Figure 1. Schematic representation of two contiguous α-(1-4) linked glucopyranose residues (maltose), along with the labelling of the atoms and the torsion angles of interest.
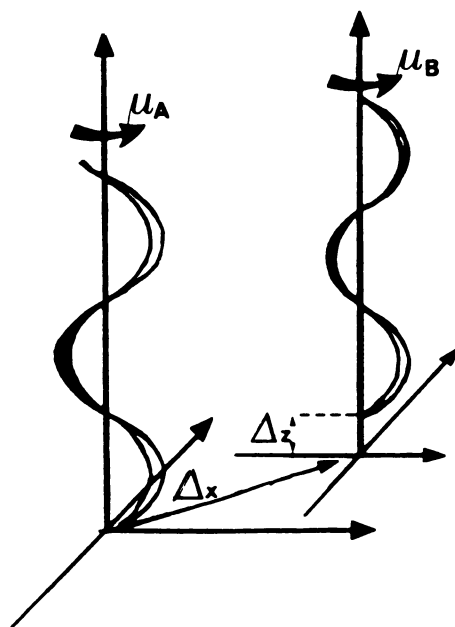
Figure 2. Interhelical parameters required to define the geometric orientation of chain A relative to chain B.

Other relationships between crystallographic elements of symmetry and interhelical parameters can be easily derived.

Contacting Procedure. As shown in Figure 3, for a given values of μA, μB and Δz, our program determines a normal translation Δx that places one or more atoms in van der Waals contact without any interpenetration as described earlier (8). The surface of the chain is defined by circumscribing a hard-sphere of the appropriate van der Waals radius Ri, around each constituent atom. In general, the final position of the two polymeric chains is characterized by the following :
     1/ For at least one atom pair (i,j), the ith atom of the chain A is separated from the jth atom of the chain B by the sum of Ri and Rj. The atom pair i, j which satisfies this condition is referred to as the determining contact.
     2/ There is no atom pairing between the two chains that has a distance closer than the appropriate van der Waals raddii sum.
Obviously, condition (2) cannot be fulfilled for an atom pair involved in an hydrogen bond. Since the hydroxylic hydrogens are not explicitly considered here, hydrogen bonds are defined in terms of the distance between the hydrogen donor oxygen atom and the oxygen which accepts the hydrogen atom. All potential couples of atoms eligible to participate in an interchain hydrogen bond are identified and ommited from the contacting procedure. This implicitely means that hydrogen bonding will not violate principle (1) for the van der Waals bonded atoms.

Interchain Energy Calculations. If a contacting procedure is used, chain-chain construction requires only geometric information, and in principle, one can subsequently calculate the energy of the resulting interactions (E_{AB}) to any degree of approximation.
     For a formally infinite chain, the expression for the interchain interaction energy is :

$$E_{AB} = \sum_{i=1}^{Na} \sum_{j=1}^{Nb} \sum_{n=1}^{\omega} E_{ijn}$$

where Na is the number of atoms per identity period of chain A, Nb is the number of atoms per identity period of chain B, and ω is the number of repeating units.
     Atom-atom potentials have been used extensively for the study of molecular crystals, and many useful empirical parameters sets have been designed. The interaction energy of the two chains is considered to be the sum over all pairwise interactions. In the present work, such interaction is considered according to the 6-12 potential functions. The energy of an atom pair is given by an expression of the form :

$$E_{ijn} = A/r_{ijn}^{12} - C/r_{ijn}^{6}$$

where r_{ijn} is the distance between the ith atom in the reference identity period of chain A and the jth atom in the nth identity period of chain B. These terms incorporate a short-range repulsive interaction and a short-range attractive interaction, respectively. To these may be added coulombic interactions. As for the energy stabilization arising from hydrogen bonding, an extra term has to be included. In the present work, we limit ourselves to investigating whether the information provided by short-range interactions alone is of utility for identifying structural assemblies of polymer chains. In performing the interchain energy calculation, we have used a cutoff distance such that dij<1.5 x Rij, where Rij is the appropriate van der Waals radii sum. The number of interatomic contacts satisfying this particular condition may be also computed, and referred to as the number of "close contacts". It represents an indicator of the complementarity of the shapes of the interacting chains, in a particular given orientation. The energy contribution arising from hydrogen bonding (VHB:kcal/mol) is calculated as described previously (25):

$$V_{HB} = 331.4 \; (\; R - 0.255 \;) \; (\; R - 0.305)$$

where R is the distance between oxygen atoms which should lie between 0.255 and 0.305 nm.

<u>Practical Treatment and Rules of Selection</u>. In practical calculations we have adopted the procedure of mapping Δx and E_{AB} as a function of the structural variables, μA, μB and Δz. The two orientational angles μA and μB are strictly bound between 0 and 360°, whereas Δz is bound between 0 and the fiber repeat : <u>t</u>. The investigation is performed by rotating μA and μB over the whole angular range by selected increments (typically from 10° to 1°); the relative translation (Δz) between the two chains is investigated over the length of a whole fiber repeat, typically by 0.05 to 0.01 nm increments. For each setting of the chains, as generated by μA, μB and Δz, the magnitude of the perpendicular off-set (Δx) is derived according to the contacting procedure outlined above. Then the values of the energy E_{AB} corresponding to each set of chain orientations are computed. Several routes can be envisaged to search for the energy minima. We found that the mapping procedure was quite adequate, particularly in the first steps of the investigations, since it provides a complete overview of the symmetry (or lack of symmetry) to the chain-chain interactions. In order to compare the structures of model chain interactions to those determined experimentally, a procedure for extracting the geometric parameters from

the experimental chains was designed. A suite of computer programs to perform all these treatments was written (9).

Since the present procedure allows for a thorough examination of all possible chain-pairings, many stable interactions are expected to be found. Whereas all these arrangements may represent situations which are likely to occur in the amorphous state, only a few of them would correspond to chain pairing capable of generating efficient packing. Therefore, for the selection of models of chain interactions the following rules were considered. Priority should be given to chain pairing occurring for :

    1/ coupled values of μA and μB.
    2/ $\Delta z = \underline{t}$ / n, where n is a simple integer or $\Delta z = 0$.
    3/ low values of the interchain energy.

<u>Miscellaneous</u>. All the calculations were performed on Vax workstations of the I.N.R.A. Research Centre in Nantes. The molecular drawings were realized with the program PITMOS (26) and the space-filling representations with the program SPACEFIL (27). The iso-energy and iso-contour plots were drawn with the aid of the program MONGO (28).

<u>Results and Discussion</u>

<u>Chain Building</u>. Both crystalline polymorphs of starch have the same fiber repeat of 1.05 nm and are built with the same unit : a parallel stranded double-helix. Each strand has six glucose residues per turn in 2.1 nm and the two strands are related by a two-fold axis of symmetry; this creates the apparent 1.05 nm fiber repeat. The chirality of the helices has been postulated to be either right-handed (<u>12,13</u>) or left-handed (<u>14,15</u>).

The choice of the left-handed chirality has been made from the following conformational analysis. Model building was performed starting from the conformational analysis of the $\alpha(1-4)$ linkage of maltose. The potential energy surface calculated as a function of the glycosidic torsion angles Φ and Ψ is represented in Figure 4. The helical parameters of the amylosic strand generated for each combination of Φ and Ψ are superimposed on the potential energy surface, which is divided in regions of left-handed and right-handed chirality by the contour iso-<u>h</u> = 0. The conformations which generate six-fold helices repeating in 2.1 nm are located at the intersection of <u>n</u> = 6 and <u>h</u> = 0.35 nm (for a right-handed chirality) or <u>h</u> = -0.35 nm (for a left-handed chirality). The conformation generating a left-handed helix (Figure 5a) has a lower energy than the right-handed one; besides it corresponds to the conformations observed for some maltodextrins in the crystalline state (<u>25</u>). The value of the torsion angles at the glycosidic linkages are $\Psi = 86.7°$ and $\Phi = -146.4$; the hydroxymethyl groups are all in a <u>gauche-gauche</u> orientation. Assembling two such parallel single strands through a two-fold operation
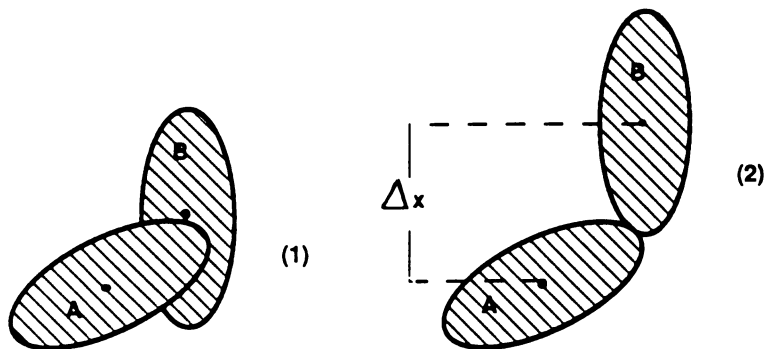
Figure 3. Contacting procedure. (1) Projection along the chain axis of two interpenetrating helices. (2) Helix B has been moved such that the two are in contact, but no longer interpenetrating.
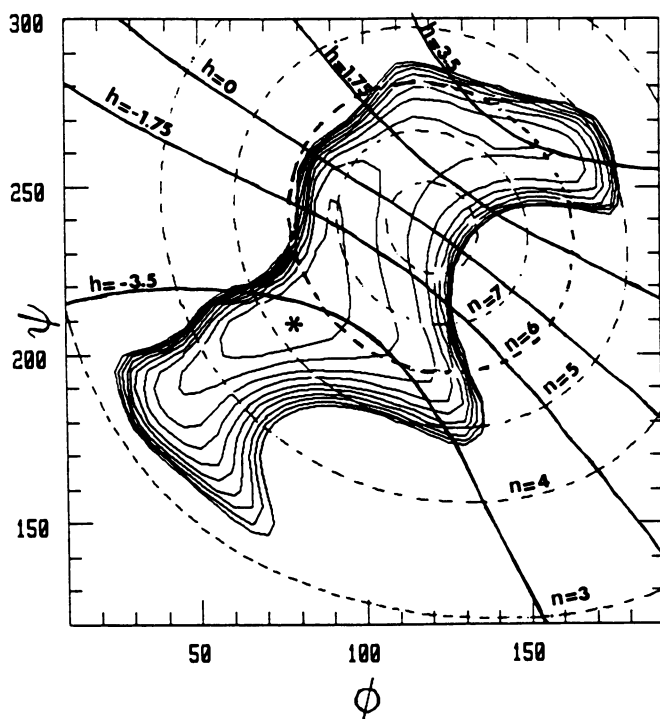
Figure 4. Selected iso-n and and iso-h contours superimposed on the potential energy surface for maltose. The iso-energy contours are drawn by interpolation of 1 kcal/mol with respect to the relative energy minimum (*). The iso-h = 0 contour divides the map into two regions, corresponding to right-handed (h>0) and left-handed (h<0) chirality.

Table 1 : Cartesian coordinates (Ang.) of the trisaccharide used to build the double-helix (oriented along the c axis). The three other glucose units of the same strand are generated by the symmetry operation (-x, -y,z+10.5) and then the other strand is generated by the symmetry operation (-x,-y,z)

atom	x	y	z
C-1	-3.24970	-2.20880	-2.00830
C-2	-2.58490	-3.56300	-1.88590
C-3	-1.24730	-3.57710	-2.70770
C-4	-1.42450	-2.95960	-4.08700
C-5	-2.05990	-1.59110	-4.00040
C-6	-2.32050	-0.94950	-5.34180
O-2	-2.40470	-4.02490	-0.65960
O-3	-0.85650	-4.94170	-2.82040
O-4	-0.13340	-2.84400	-4.70390
O-5	-3.34810	-1.71300	3.34260
O-6	-3.04640	-1.81830	6.22730
H-1	-4.23680	-2.24000	-1.57130
H-2	-3.31640	-4.17360	-2.39440
H-3	-0.52240	-2.94980	-2.21040
H-4	-2.08630	-3.58370	-4.66930
H-5	-1.41660	-0.96430	-3.40040
H-61	-2.88560	-0.04130	-5.19320
H-62	-1.37590	-0.75330	-5.82710
C-1	-3.53690	1.71250	1.49450
C-2	-4.37720	0.45960	1.61670
C-3	-3.72040	-0.70570	0.79480
C-4	-3.27420	-0.24330	-0.58440
C-5	-2.40690	0.99140	-0.49760
C-6	-1.98150	1.53800	-1.83890
O-2	-4.68710	0.07240	2.84290
O-3	-4.70670	-1.72650	0.68200
O-4	-2.52840	-1.30340	-1.20130
O-5	-3.15670	2.04580	0.16030
O-6	-3.09690	1.73220	-2.72450
H-1	-4.05760	2.55170	1.93160
H-2	-5.27170	0.78780	1.10820
H-3	-2.81470	-1.01980	1.29220
H-4	-4.14560	0.01780	-1.16670
H-5	-1.54240	0.74770	0.10240
H-61	-1.47770	2.48160	-1.69020
H-62	-1.33920	0.81820	-2.32430
C-1	-0.28490	3.92190	4.99770
C-2	-1.79010	4.02290	5.11990
C-3	-2.47080	2.87150	4.29780
C-4	-1.84710	2.71660	2.91860
C-5	-0.34420	2.58280	3.00550
C-6	0.34200	2.48800	1.66420
O-2	-2.28050	4.09750	6.34610
O-3	-3.84800	3.21510	4.18500
O-4	-2.39220	1.54060	2.30150
O-5	0.19390	3.75940	3.66350
O-6	-0.04760	3.55110	0.77870
H-1	0.18130	4.79240	5.43500
H-2	-1.95330	4.96170	4.61150
H-3	-2.28990	1.93010	4.79510
H-4	-2.05680	3.60170	2.33640
H-5	-0.12300	1.71240	3.60540
H-61	1.41100	2.52360	1.81290
H-62	0.03980	1.57190	1.17870

permits formation of a double helix having no short
contacts (Figure 5b). Space-filling representations of
this model are shown in Figure 5c and 5d. The atomic
coordinates from which the double-helical chain having
$\underline{n}$ = 6 and $\underline{h}$ = -0.35 nm can be generated, are listed in
Table 1.

Modeling of the Chain-Chain Interactions.   All the
possible arrangements occurring between parallel and
antiparallel double-helices were examined. This was
performed by rotating μA and μB over the whole angular
range from 0 to 360° by 10° increments. The relative
displacement of the two chains was investigated over the
whole length of the fiber repeat ($\underline{t}$ = 1.05 nm) by 0.05 nm
increments. For each setting of the chains (i.e. parallel
and antiparallel), as a function of μA, μB and Δz, the
magnitude of the perpendicular off-set, Δx was computed
according to the contacting procedure outlined above. The
values of the energy corresponding to each set of chain
orientations were evaluated. Within the three-dimensional
(μA, μB, Δz) space, the search for the energy minima was
performed.

Table 2 : Description of the Geometrical and Energetical
Features of the Significant Low-Energy Chain-Pairings

	PARA 1	PARA 2	ANTI 1	ANTI 2
μA(°)	11.5	26.0	78.5	47.0
μB(°)	11.5	167.5	41.5	13.0
Δz(nm)	0.525	0.322	0.786	0.739
Δx(nm)	1.077	1.120	1.077	1.116
E(kcal/mol)	-26.6	-19.8	-23.7	-20.9
ΔE	0.0	6.8	2.9	5.7

     The calculated interchain parameters for the best
energy minima for parallel or antiparallel arrangements
are collected in Table 2. The best pairing in terms of
energy is found for a parallel arrangement of the double-
helices (PARA 1); it corresponds to an inter-chain vector
of 1.077 nm. Another stable parallel chain-pairing
(PARA 2) is found which has an interaction energy
6.8 kcal/mol above the previous one. As for the anti-
parallel case, two stable chain-pairings are found. The
more stable one (ANTI 1) is characterized by a distance
between the center of mass of the two double-helices of
1.077 nm and for coupled values of μA and μB
($\mu A = -\mu B + 180°$). The associated inter chain energy is
about 3 kcal/mol higher than the lowest energy; it
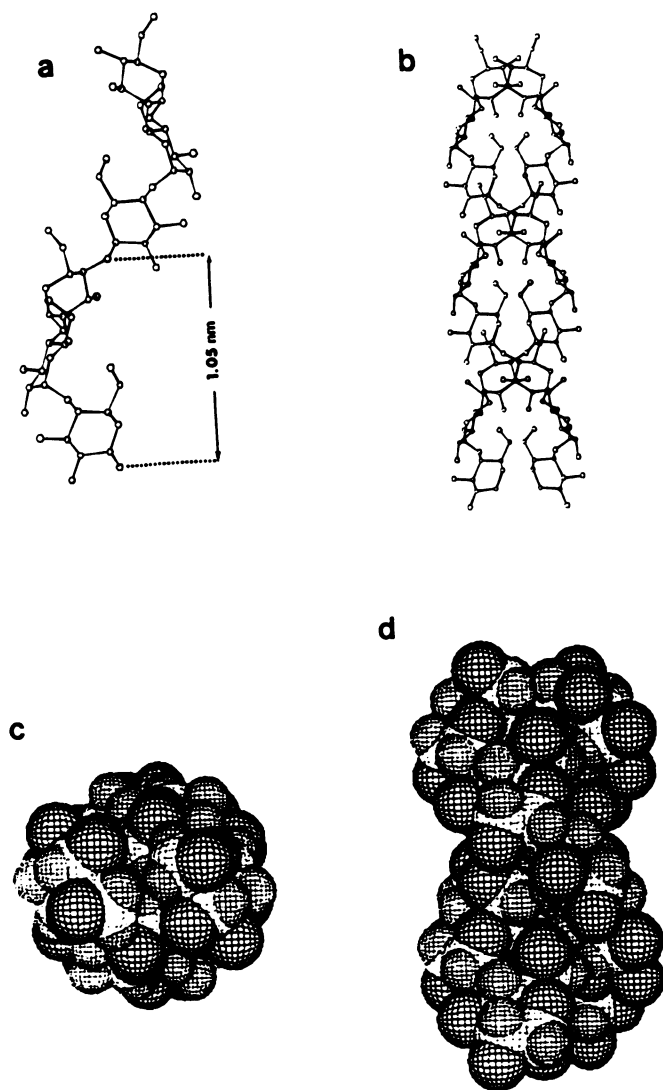
Figure 5. Molecular drawings of : a) one single strand of an amylosic chain in the left-handed conformation, having a six-fold symmetry, repeating in 2.1 nm. b) the double-helix generated by the association of two single strands, through two-fold symmetry operation. c) and d) Space-filling plots of the double helix, projected along and perpendicular to the fiber axis, respectively.

corresponds to the establishement of hydrogen bonds of the type O-2...O-2, O-3...O-3 and O-3...O-5. This arrangement is depicted in Figure 6. Another chain-pairing (ANTI 2) has an interchain energy 5.7 kcal/mol higher than the global minimum. Based upon the criteria presented in the Methodology Section, (Δx, energy, and coupling of μA and μB), the PARA 1 model is considered to be the best candidate for generating crystalline arrangements having high density.

The best pairing of parallel chains (PARA 1) is depicted in Figure 7. The coupling relation between the two rotational angles is such that $\mu A = \mu B$. The existence of a coupling between the rotation angles implies the occurrence of a symmetry element between the two double-helices; it implies that the asymmetric unit contains at most one double-helix (in fact, the building of the double-helices implies some other symmetry elements and thus the content of the asymmetric unit is reduced even more). The particular relationship $\mu A = \mu B$ indicates that the pairing is achieved through pure translational symmetry. Since in that case, the three-dimensional space ($\mu A, \mu B, \Delta z$) is reduced to a two-dimensional one ($\mu A = \mu B, \Delta z$) a straigthforward representation can be done. The result of this study is shown in Figure 8. Figure 8a is a representation of the variations of the optimum interchain perpendicular off-set, Δx, as a function of Δz, translation along the chain direction, and coupled rotations of μA and μB. This map exhibits obvious six-fold symmetry, resulting from the symmetry elements of the double-helix. The energy values corresponding to each pairing are represented in Figure 8b. For both maps it is clear that the minimum occurring for $\mu A = \mu B = 11.5°$, $\Delta z = 0.525$ nm (indicated by an arrow) corresponds to very favorable pairing, not only in terms of energy (-26.5 kcal/mol, in the selected scheme of calculation), but also in terms of density, since the distance between the centers of mass of the two double-helices is 1.077 nm. It has to be emphasized that such a pairing occurs for a value of 0.525 nm of Δz which corresponds to exactly one half of the fiber repeat. The duplex of double-helices corresponding to this minimum is represented in Figure 7. It appears that this very favored interaction is due to two factors. The translation Δz of half a fiber repeat generates a very good "nesting" of the two double-helices and thus this structure is stabilized by numereous van der Waals interactions. The occurrence of hydrogen bonds of the type O-3...O-3 and O-2...O-6 stabilizes the pairing further. An evaluation of the different energy components demonstrates that the hydrogen bonding and the van der Waals contributions contribute to the energy of interaction in the ratio 60/40.

a

b
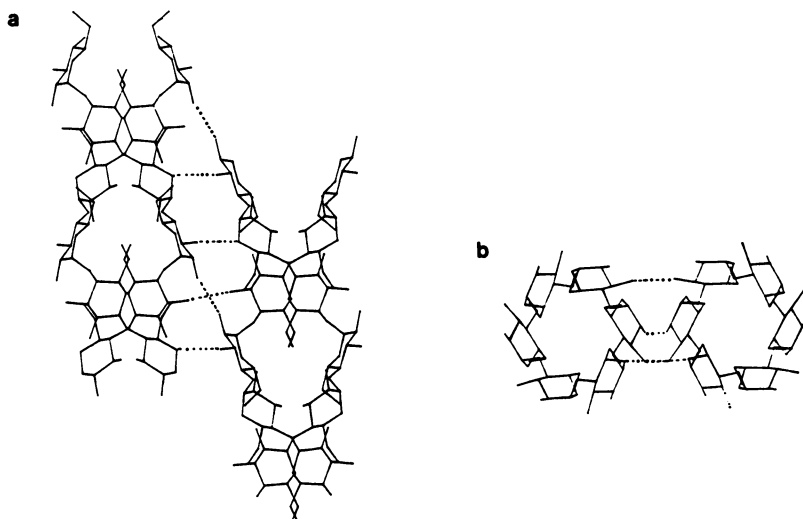
Figure 6. Molecular drawings of the best arrangement of
two anti-parallel double-helices. a) Projection
perpendicular to the chain axis. b) Projection along
the chain axis. Inter double-helix hydrogen bonds are
shown as dotted lines.

a

b
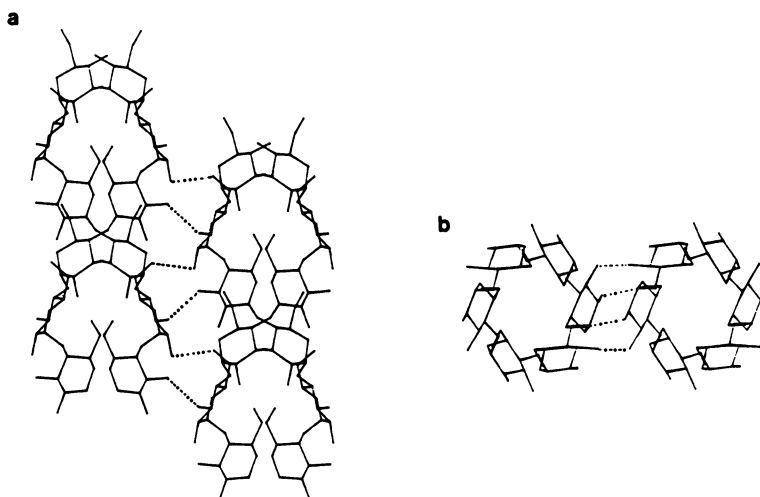
Figure 7. Molecular drawings of the best arrangement of
two parallel double-helices. a) Projection
perpendicular to the chain axis. b) Projection along
the chain axis. Inter double-helix hydrogen bonds are
shown as dotted lines.
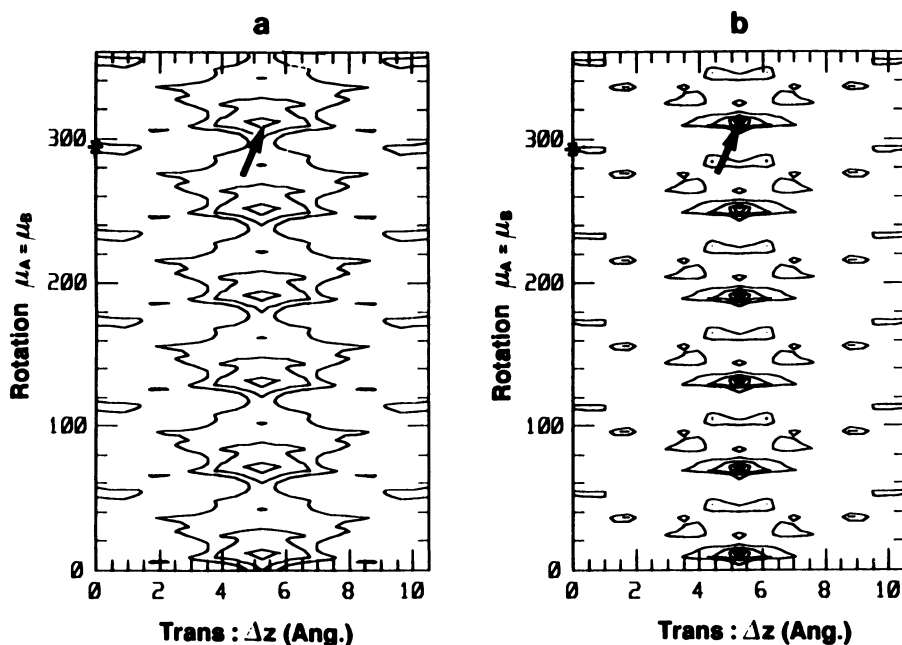
Figure 8. Contour maps calculated for a parallel arrangement of double helices, as a function of the translation: Δz, along the fiber axis and the coupled rotation angles μA = μB. a) Variations of the perpendicular off-set (Δx). Contours correspond to 1.1, 1.15 and 1.2 nm; the arrow indicates the lowest value, and generates the PARA 1 model. b) Interchain energies calculated for the corresponding Δx. Contours correspond to -25, -20, -15, -10 and -7 kcal/mol; the arrow indicates the lowest value. N.B. The * indicates the loose interaction found in the A-type crystal structure.

Comparison with the Crystalline Structure of A- and B-
Type Starch.   In the most recent crystallographic studies
on the crystalline part of starch (14,15) the structure
of both polymorphs are based on a parallel arrangement of
left-handed double-helices. In the two observed
structures the double-helices are slightly different
since small variations away from the perfect six-fold
symmetry are found. Nevertheless, they correspond closely
to the model studied here. The essential result is that
in these two structures the closest interactions between
two neighboring double-helices correspond closely to the
duplex described as PARA 1.
       In the crystal structures, neighboring double-
helices have the same rotational orientation and the same
translation of half a fiber repeat as in the PARA 1
model. Only the Δx vector is slightly larger in the
calculated interaction (1.077 nm) than in the observed
ones : 1.062 nm and 1.068 nm in the A type and B type,
respectively. This may be due to the fact that in the
crystal structures the helices depart slightly from
perfect 6-fold symmetry. Also, no interpenetation of the
van der Waals surfaces is allowed in the calculations,
whereas some of them may occur in the cristallographic
structure. It is quite interesting to note that the
network of inter double-helices hydrogen bonds found in
the calculated PARA 1 model reproduces those found in the
crystalline structures.
       The differences between the two polymorphs occur
from other effects. In fact the B-type polymorph has an
hexagonal symmetry; each double-helix has only three
neighbors (corresponding to the interaction described
above). The channel created by six double-helices packed
in the hexagonal fashion is occupied by a column of water
molecules. On the contrary, in the less hydrated A-type
structure, each double-helix is surrounded by six
neigbouring ones. The chain pairing described by the
PARA 1 model corresponds to four out of six of these
interactions; the two other ones being looser. It is very
interesting to see that this type of loose interaction
which is generated by a translation along the b axis
(Δx = 1.172nm, Δz = 0) can also be detected on the maps
shown in Figure 8. This arrangement does not correspond
to the lowest energy minimum but is among the low energy
chain pairings. It occurs for μA = μB = 55°, Δz = 0 nm,
with a perpendicular off-set Δx = 1.176 nm and an energy
of -7 kcal/mol), this minimum being located by an asterix
in Figure 8. Therefore, the A-type structure has a nearly
close-packed arrangement of double-helices, each double-
helix being surrounded by six neighbors. The B-type
structure consists of a more open packing of double-
helices, with a greater amount of inter-helical water.

Understanding the Crystalline Polymorphism in Native
Starch.   From this study, it is clear that the A and B
forms have in common not only a double-helix but a

particular duplex of double-helices which appears to correspond to the best calculated chain-pairing. From this knowledge we envision a model for the transition between the two allomorphs in the solid state. This transition has only been observed from B to A starch, i.e. from the more hydrated structure to the less hydrated one. The proposed model is shown in Figure 9. Upon progressive removal of the water molecules, and unidirectional slipping of the duplexes, it is possible to convert B-type packing to A-type packing while maintaining the integrity of the duplex.

Conclusions

We have presented a computational method for investigating the structure and stability of pairs of polymer helices. The chain pairs were positioned by a strictly geometric procedure and characterized by energies based on atom-atom potentials that included interchain hydrogen bonding. The structures derived from this process are surprisingly good approximations to the experimental ones. These results indicate that isolated chain-pair models can provide useful information about possible modes of interactions between the chains. Since no water was used in the modeling, it appears that these water molecules do not play any significant role in the establishement of interhelical relationships.

Because the proposed method uses rigid model helices, the entire variation space may be readily surveyed. In both the parallel and antiparallel cases, a favorable chain packing occurs with identical distances between double helices. Nevertheless, the parallel arrangement is favored by a lower energy. Consideration of electrostatic interactions could reverse or confirm these findings. Since the double-helices are fairly cylindrical, the finding of favorable energies for both parallel and antiparallel packings is not surprising.

The observed crystal structures of both polymorph A and B of crystalline starch have been rationalized. Chief among these findings is a duplex of parallel double-helices which constitutes a common structural feature of the two forms. It corresponds to the most favored way to pack double-helices that have "crests" and "troughs". Since the same duplex was found in both forms, it was proposed that this duplex is maintained during transition from the B to the A form. The lattice energy of the A form is apparently lower than the B form, omitting the contribution of water. This agrees with the observation that the A-type polymorph has a higher melting temperature (29) and hence is more stable than the B-type. This is despite the fact that the A-type has a more negative entropy of formation than the less tightly packed B-polymorph (30).
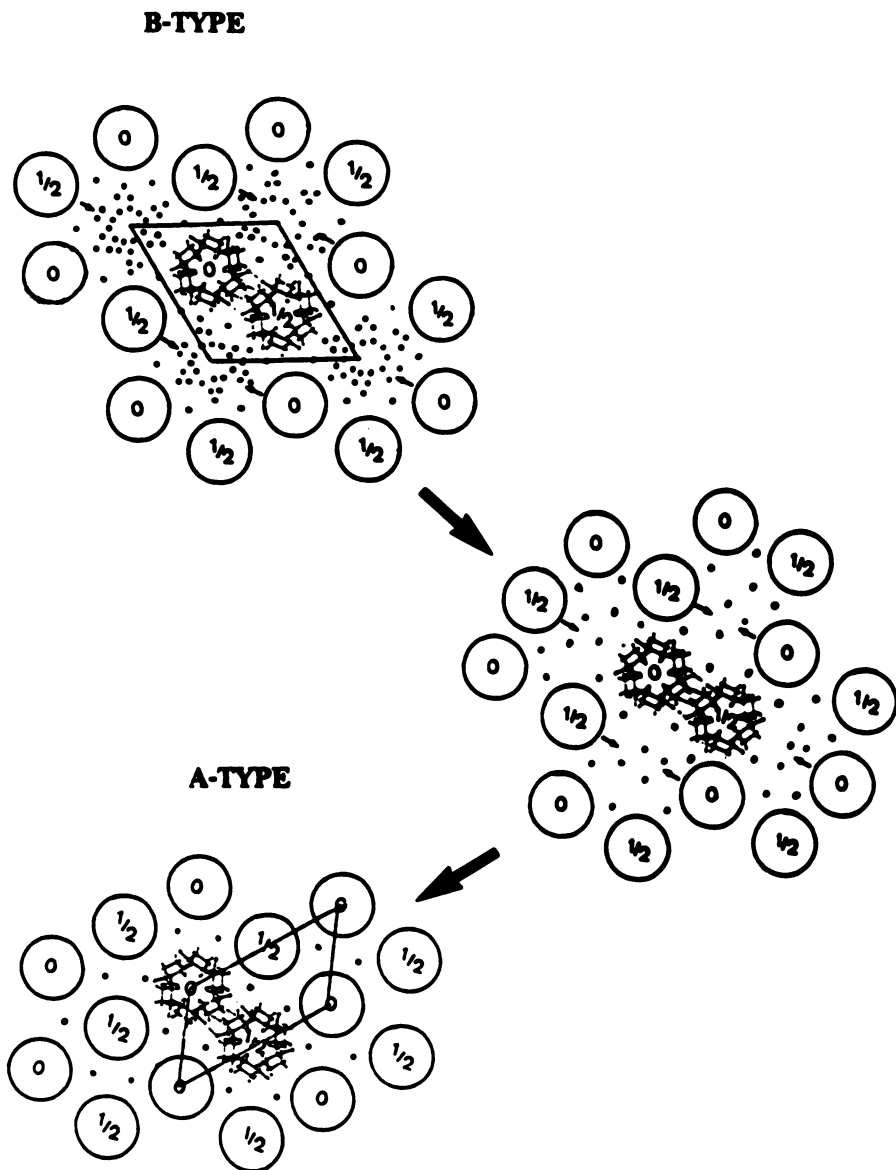
B-TYPE

A-TYPE

Figure 9. Model of the polymorphic transition from the B-type to the A-type starch, in the solid state. The parallel double-helices which form the duplex are labelled 0 and 1/2 (this indicates their relative translation along the c axis). The water molecules are shown as dots.

Literature Cited

1.   Sarko, A. <u>TAPPI</u> 1978, <u>61</u>, 130-135.
2.   Hopfinger, A.J.; Walton, A.G. <u>J. Macromol. Sci. Phys</u>. 1969, <u>B3</u>, 195-208.
3.   Hopfinger, A.J.; Walton, A.G. <u>J. Macromol. Sci. Phys</u>. . 1970, <u>B4</u>, 185-199.
4.   Hopfinger, A.J.; <u>Biopolymers</u> 1971, <u>10</u>, 1299-1315.
5.   Tai, K.; Kobayashi, M.; Tadokoro, H. <u>J. Polym. Sci., Polym. Phys. Ed</u>. 1976, <u>14</u>, 783-797.
6.   Chou, K.C.; Nemethy, G.; Scheraga, H.A. <u>J. Phys. Chem</u>. 1983, <u>87</u>, 2869.
7.   Chou, K.C.; Nemethy, G.; Scheraga, H.A. <u>J. Am. Chem. Soc</u>. 1984, <u>106</u>, 3161.
8.   Scaringe, R.P.; Pérez, S. <u>J. Phys. Chem</u>. 1987, <u>91</u>, 2394-2403.
9.   Scaringe, R.P.; Pérez, S. to be published
10. Katz, K.R.; van Itallie, T.B. Z. <u>Phys. Chem. Abt</u>. 1930, <u>A150</u>, 90-100.
11. French, D. In <u>Starch, Chemistry and Technology, 2nd Ed</u>.; Whistler, R.L.; BeMiller, J.N.; Parschall, E.F., Eds; Academic Press : New York, 1984; p 183-247.
12. Wu, H.C.H.; Sarko, A. <u>Carbohydr. Res</u>. 1978, <u>61</u>, 7-25.
13. Wu, H.C.H.; Sarko, A. <u>Carbohydr. Res</u>. 1978, <u>61</u>, 27 40.
14. Imberty, A.; Chanzy, H.; Pérez, S.; Buléon, A.; Tran, V. <u>J. Mol. Biol</u>. 1988, <u>201</u>, 365-378.
15. Imberty, A.; Pérez, S. <u>Biopolymers</u> 1988, <u>27</u>, 1205-1221.
16. Kainuma, K.; French, D. <u>Biopolymers</u> 1972, <u>11</u>, 2241-2250.
17. Sair, L. <u>Cereal Chem</u>. 1967, <u>44</u>, 8-26.
18. Marchessault, R.H.; Pérez, S. <u>Biopolymers</u> 1979, <u>18</u>, 2369-2374.
19. IUPAC-IUB Commission on Biological Nomenclature <u>J. Mol. Biol</u>. 1970, <u>52</u>, 1-17.
20. IUPAC-IUB Commission on Biological Nomenclature <u>Arch. Biochem. Biophys</u>. 1971, <u>145</u>, 405-421.
21. Hough, E.; Neidle, S.; Rogers, D.; Troughton, P.G.H. <u>Acta Cryst</u>. 1973, <u>B29</u>, 365-367.
22. Pérez, S. DSc Thesis, Grenoble, 1978.
23. Tvaroska, I.; Pérez, S. <u>Carbohydr. Res</u>. 1980, <u>149</u>, 389-410
24. Gagnaire, D.; Pérez, S.; Tran, V. <u>Carbohydr. Res</u>. 1980, <u>78</u>, 89-109.
25. Pérez, S.; Vergelati, C. <u>Polymer Bull</u>. 1987, <u>17</u>, 141-148.
26. Pérez, S.; Scaringe, R.P. <u>J. Appl. Cryst</u>. 1986, <u>19</u>, 65-66.
27. Henry, D.R. <u>Comput. Chem</u>. 1983, <u>7</u>, 119-135.
28. Pogge, R. <u>MONGO program</u>, Lick Observatory, CA, USA; Vax Graphic Station version by Tom Lew, University of Toronto, CANADA.
29. Donovan, J.W.; Mapes, C.J. <u>Staerke</u> 1980, <u>32</u>, 190-193.
30. Gidley, M.J. <u>Carbohydr. Res</u>. 1987, <u>161</u>, 301-304.

Chapter 18

A New Generation of Gel-Forming Polysaccharides

An X-ray Study

R. Chandrasekaran and V. G. Thailambal

Whistler Center for Carbohydrate Research, Purdue University, West Lafayette, IN 47907

Using computer modeling, jointly with x-ray fiber diffraction data, the molecular architectures of two different gel-forming polysaccharides have been examined. Preliminary results indicate that the neutral and doubly branched capsular polysaccharide from *Rhizobium trifolii* can form a 2-fold single helix of pitch 1.96 nm or a half-staggered, 4-fold double-helix of pitch 3.92 nm. The molecules are likely to be stabilized by main chain -- side chain interactions. Detailed structure analysis reveals that the monovalent salt forms of gellan, an anionic, linear extracellular polysaccharide, exist as half-staggered, parallel double-helices containing 3-fold, left-handed polysaccharide chains of pitch 5.63 nm. The double-helix is stabilized by interchain hydrogen bonds involving the carboxylate groups. The crystal structure of the potassium salt shows that double-helix -- potassium -- water -- potassium -- double-helix interactions promote the aggregation of molecules and subsequent gelation. Extrapolation of these results by computer model building swiftly reveals the ability of calcium ions to establish direct and strong double-helix -- calcium -- double-helix interactions. This explains the good gelling behavior of gellan at a very low calcium concentration. Further, modeling calculations enable us to understand that the poor gelling properties of native gellan are due to the presence of the glycerate, rather than acetate, groups.

Computer modeling is an exciting and powerful technology for obtaining the three-dimensional structures of molecules. The process begins with a knowledge of the chemical structure. Based on the fundamental principles of stereochemistry, the shapes of molecules, ranging from simple oligosaccharides to complex

carbohydrates, can be investigated. Computer modeling can further be used to examine, for example, protein-polysaccharide, and nucleic acid-drug, interactions. A detailed study of the molecular architectures is necessary in order to understand the structure-function-property relationships. This, in turn, is crucial to the development of many industrial polymers and new biotechnological products. Aqueous solutions of the recently discovered bacterial polysaccharides such as gellan and related polymers, and the capsular polysaccharide from *Rhizobium trifolii* (RTPS), either display extremely high viscosity over a wide range of temperature or form gels of excellent strength and texture at low concentrations. They are particularly useful to the food industry. Interactions involving the polymer chains, water molecules and cations in the case of anionic polymers, together with those between the crystalline (ordered) and non-crystalline (disordered) regions of the polymer specimens are responsible for the rheological behavior of the polymer solutions (1).

X-RAY FIBER DIFFRACTION

Structural information at the molecular level can be extracted using a number of experimental techniques which include, but are not restricted to, optical rotation, infra-red and ultra-violet spectroscopy, nuclear magnetic resonance in the solid state and in solution, diffraction using electrons, neutrons or x-rays. Not all of them, however, are capable of yielding structural details to the same desirable extent. By far, experience shows that x-ray fiber diffraction (2), in conjunction with computer model building, is the most powerful tool which enables to establish the spatial arrangement of atoms in polymer molecules.

All three major classes of biopolymers -- polysaccharides, polypeptides and polynucleotides, as well as most synthetic polymers, tend to adopt long helical structures. Under these circumstances, it is impossible to produce single crystals suitable for traditional x-ray crystallographic investigations, but often possible to prepare polycrystalline and oriented specimens with the long molecular axes aligned nearly parallel to each other. A typical specimen diffracts the incident x-ray beam to give a series of Bragg reflections sampled at reciprocal lattice points on layer lines. The positions of the spots depend on the unit cell dimensions. The distribution of intensities and the layer line spacings reveal the symmetry and pitch of the helix respectively. Together with the actual intensities of the reflections, the spatial arrangement of atoms in the unit cell can be delineated using computer model building procedures.

Sampling of x-ray diffraction does not occur with uniaxially oriented (but non-crystalline) specimens which only produce continuous intensity distribution on layer lines. Information on intermolecular association cannot be inferred in these cases. Between the uniaxially oriented, non-crystalline specimens and the oriented polycrystalline materials, there is a whole range of "ordering" which yields different types of diffraction patterns in terms of detailed structures (3).

MODEL BUILDING AND STRUCTURE REFINEMENT

The first step in the structure analysis is to build a stereochemically plausible model of a molecular asymmetric unit that fits into a helix. It must have the pitch and symmetry characteristics determined from the layer line spacings and from the systematic absences and overall intensity distribution in the diffraction pattern. The linked-atom tree geometry, based on the primary structure, is set up to describe the building block of the helix. All bond lengths and most bond angles are usually fixed at their expected standard values which are derived from a survey of the crystal structures of appropriate monomers. The main variables are the relevant conformation angles. If necessary, the endocyclic bond angles in the case of flexible pyranose rings, and the glycosidic bond angle at each bridge oxygen atom, are also varied. The Linked-Atom Least-Squares (LALS) procedure (4) developed in our laboratory is used for generating helical structures. It is further used to refine the helical model against the x-ray intensities. The Bragg data permit the refinement of the packing parameters of helices in the unit cell. The LALS program allows us to achieve 1) generation of compression-free helical model which is designed to have the major conformation angles in preferred domains; 2) maximization of the agreement between observed and calculated x-ray structure amplitudes; and 3) optimization of nonbonded and hydrogen bond interactions within, and between molecules. If there are competing models, their relative merits can be assessed on the basis of any, or all, of the above three criteria using Hamilton's test (5). In addition, the traditional crystallographic R-factor can be used to judge the correctness of a model.

In favorable cases, where good quality Bragg x-ray data are available as in gellan, periodic water molecules and cations in the unit cell can be located from successive difference electron density maps. Their positions can be co-refined with those of the polymer chains until the guest molecules are properly coordinated and/or hydrogen bonded with the polymer chains and complete convergence is reached. The LALS procedure has been successfully employed to solve the three-dimensional structures of several polymers, using Bragg diffraction data from well oriented polycrystalline fibers. Recently, a vectorized version (6) of the LALS program for the supercomputer CYBER 205 has been implemented. This has reduced the execution time by more than an order of magnitude and has the capacity to handle large asymmetric units containing up to 1000 atoms or more. The program is coded to include continuous diffraction amplitudes as input data. With the availability of automated intensity measurement using the optronics film scanner (7,8), it is now possible to use amplitudes of continuous diffraction and sampled, Bragg reflections either separately or jointly in the refinement of helical polymers. Using these new procedures, we have already determined the structural features of two DNA•RNA hybrids (9), three DNA•DNA duplexes (10) and the gel-forming polysaccharide kappa-carrageenan (11), all of which involve continuous intensities as data sets. Further details of the procedure are given elsewhere (12).

MOLECULAR ARCHITECTURES

It is worth pointing out the potentials of computer modeling vs.
x-ray fiber diffraction analysis. Computer model building enables
to generate a number of plausible models, but it cannot readily
distinguish one from the other. X-ray fiber diffraction analysis
involves computer modeling, followed by least-squares refinement of
the structue against the experimentally measured x-ray data.
Therefore, the merging of modeling with diffraction data helps to
discriminate between viable alternatives and, in favorable cases,
also provides the best solution. The purpose of this paper is to
illustrate this with a few recent examples. Here we summarize the
important structural features and their relationship to the observed
physical properties of gellan and RTPS. The determination of the
potassium gellan crystal structure was facilitated by the
excellent x-ray data to a final R-value of 0.18. The atomic
coordinates and structure factors have been listed by Chandrasekaran
et al (13). This structure has been instrumental in producing a
speculative crystal structure of calcium gellan and the molecular
structure of native gellan (14). This is an excellent example of
the accomplishment of the model building technique. The paucity of
x-ray data in the case of RTPS has not allowed us beyond the
generation of single and double-helical models consistent with the
helical parameters discernible in the diffraction pattern. Since
the details of the calculations have already been published, they
are not repeated in this paper.

RHIZOBIUM TRIFOLII CAPSULAR POLYSACCHARIDE. This polysaccharide is
very unusual in that it has a doubly branched, neutral
hexasaccharide repeating unit (15) shown below.

```
            β-D-Gal-(4←1)β-D-Gal
                 1
                 ↓
                 6
      →4)α-D-Glc(1→3)α-D-Man(1→3)β-D-Gal(1→
                 2
                 ↑
                 1
            α-D-Gal
```

Its aqueous solutions are not only viscous but become gels over the
temperature range 20 - 90°C. The course of sol-gel transition as a
function of temperature monitored by optical rotation (15) suggests
the conversion of a conformationally mobile random coil in the sol
state to a rigid ordered conformation in the gel state. The
midpoint of transition is approximately around 45°C. Only oriented,
noncrystalline fibers could be prepared, the best of them at 45°C.
A typical diffraction pattern shown in Figure 1 indicates that the
molecules in the specimen are aligned nearly parallel to each other,
but are not organized laterally beyond a short range. The layer
line spacings are consistent with a c-repeat of 19.6Å. The first
meridional spot occurs on the second layer line. The strong
reflection on the equator corresponds to 18.0Å spacing which mirrors
the likely diameter of the helix.

Figure 1.   An x-ray diffraction pattern from an oriented, noncrystalline fiber of the *Rhizobium trifolii* capsular polysaccharide using CuKα radiation. (Reproduced with permission from Ref. 16.   Copyright 1987 Gordon and Breach.)

There are two distinct stereochemical possibilities for the helix which are consistent with the intensity distribution. One of them is a 2-fold single-helix of pitch 19.6Å and the other a 4-fold, half-staggered, parallel, double-helix of pitch 39.2Å. The double-helix could be right- or left-handed. In all cases, there is considerable conformational mobility about the (1→6) linkage of the disaccharide side chain. Preliminary models have been built for each possibility and, due to insufficient diffraction data, detailed x-ray refinements have not been conducted for any of them.

The single-helix follows a right-handed sinuous path around the molecular axis as shown in Figure 2. The disaccharide side chain sticks out (Figure 2a) or folds back on the main chain (Figure 2b) corresponding to the conformation θ[C4-C5-C6-O6] being *gauche plus* or *trans* (16). A representative structure of the double-helix (17) is shown in Figure 3. In this model, the side chain adopts a folded conformation. Both single- and double-helices indicate a considerable amount of side chain -- backbone interactions and they may have a structural role in the gelation process. These tentative structures cannot be discriminated at this stage.

GELLAN. The extracellular, anionic, linear polysaccharide from *Pseudomonas elodea*, is called gellan in its deacylated form. It is industrially important because of its ability to form cation-dependent gels. Its tetrasaccharide repeating motif (A-B-C-D) is given below:

→3)β-D-Glc(1→4)β-D-GlcA(1→4)β-D-Glc(1→4)α-L-Rha(1→.
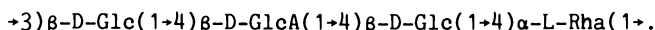
Strong and brittle gels can be formed with gellan in the presence of potassium or at a drastically reduced concentration of calcium (18,19). Native gellan, as secreted by the bacterium, is substituted by L-glycerate groups at C2 on all the O-3 substituted glucose residues and acetate groups at C6 on half of them (20) and it forms only weak and rubbery gels (21). Samples suitable for detailed structure analysis could be readily prepared from a number of monovalent salt forms of gellan. A typical diffraction pattern from a well oriented, polycrystalline fiber of the potassium salt of gellan is shown in Figure 4. Similar patterns can be obtained from the lithium form (22) and other monovalent salts. The intensity distributions and unit cell dimensions (12) are very similar indicating structural similarity in all these cases.

Crystal structures of the lithium (22) and potassium (13) forms have so far been established. The latter has not only confirmed the molecular features of the gellan double-helix first proposed (22), but also provided precise details on the locations of cations and water molecules and their interactions with the double-helices. Figure 5 shows the morphology of the double-helix which is made up of two half-staggered, parallel, 3-fold left-handed polysaccharide chains of pitch 56.3Å. The major stabilizing force is the interchain hydrogen bond and the potassium coordination, both involving the carboxylate group in every tetrasaccharide repeating unit. These interactions, as well as the interchain hydrogen bonds, are illustrated in Figure 6. Two double-helices are packed in an antiparallel fashion in the trigonal unit cell of dimensions *a* = *b* = 15.8Å and *c* = 28.2Å with a lateral separation of 9.0Å
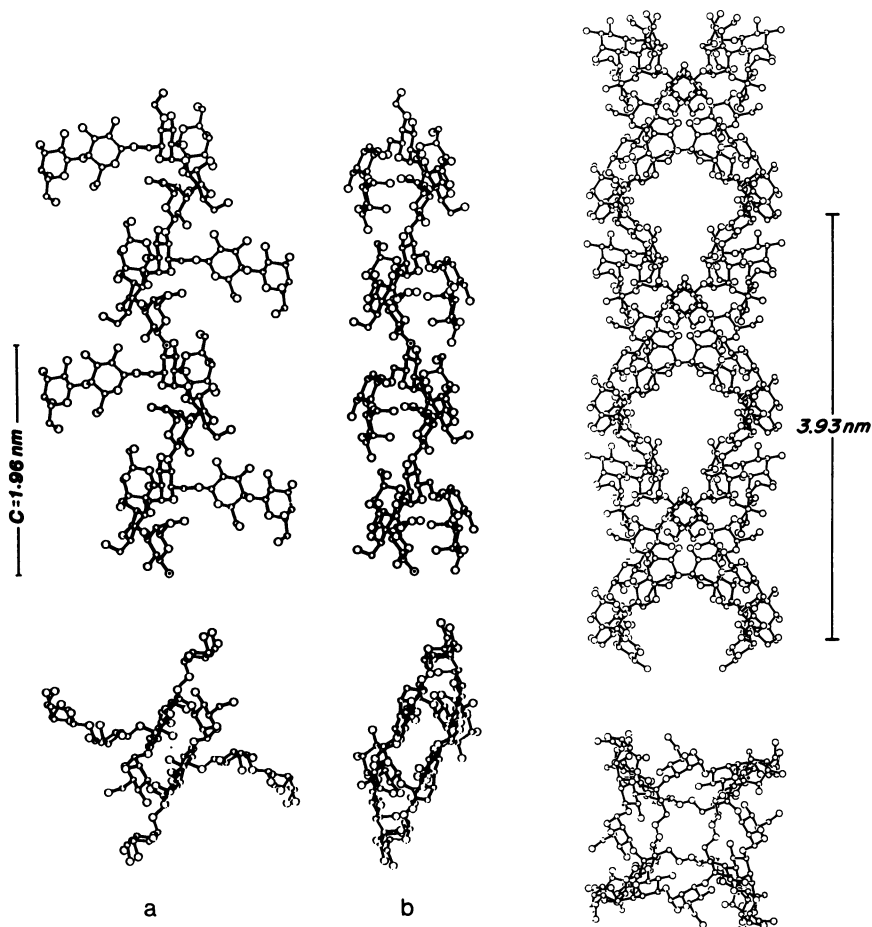
Figure 2. Two mutually perpendicular views of the 2-fold single-helix of RTPS. The sidechain conformation is *gauche plus* in (a) and *trans* in (b). (Reproduced with permission from Ref. 16. Copyright 1987 Gordon and Breach.)

Figure 3. Two mutually perpendicular views of the 4-fold double-helix of RTPS.
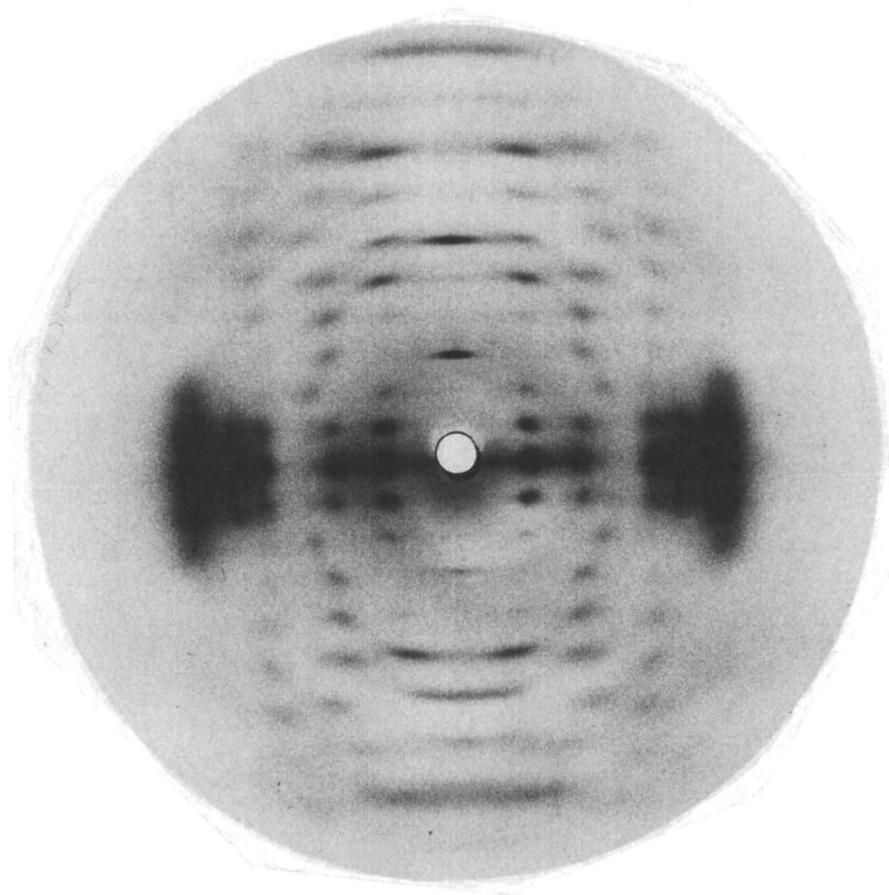
Figure 4. X-ray diffraction pattern from a well oriented and polycrystalline fiber of potassium gellan using CuKα radiation. (Reproduced with permission from Ref. 13. Copyright 1988 Elsevier.)
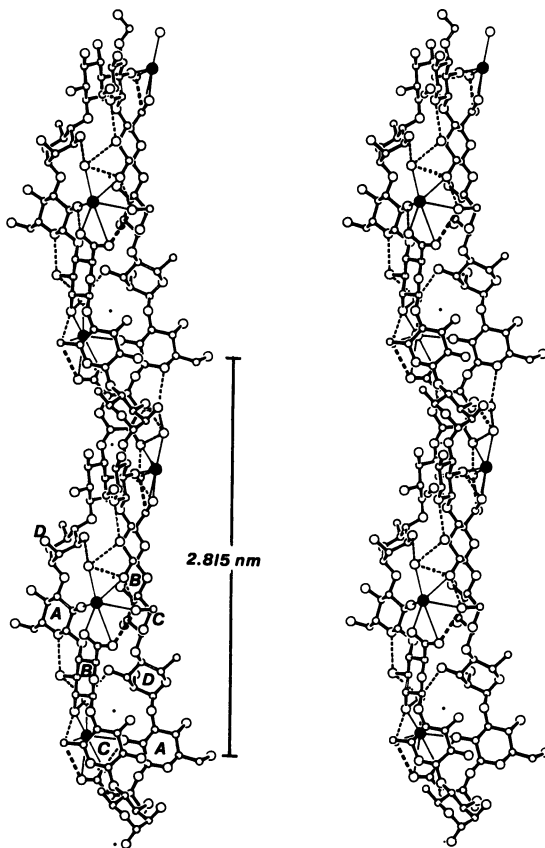
Figure 5. A stereo view of the gellan double-helix featuring the intrachain hydrogen bonds (thin dashed lines), interchain hydrogen bonds (thick dashed lines), potassium ions (filled circles), and water molecules (open circles) and the six ligands attached to each potassium ion (thin lines).
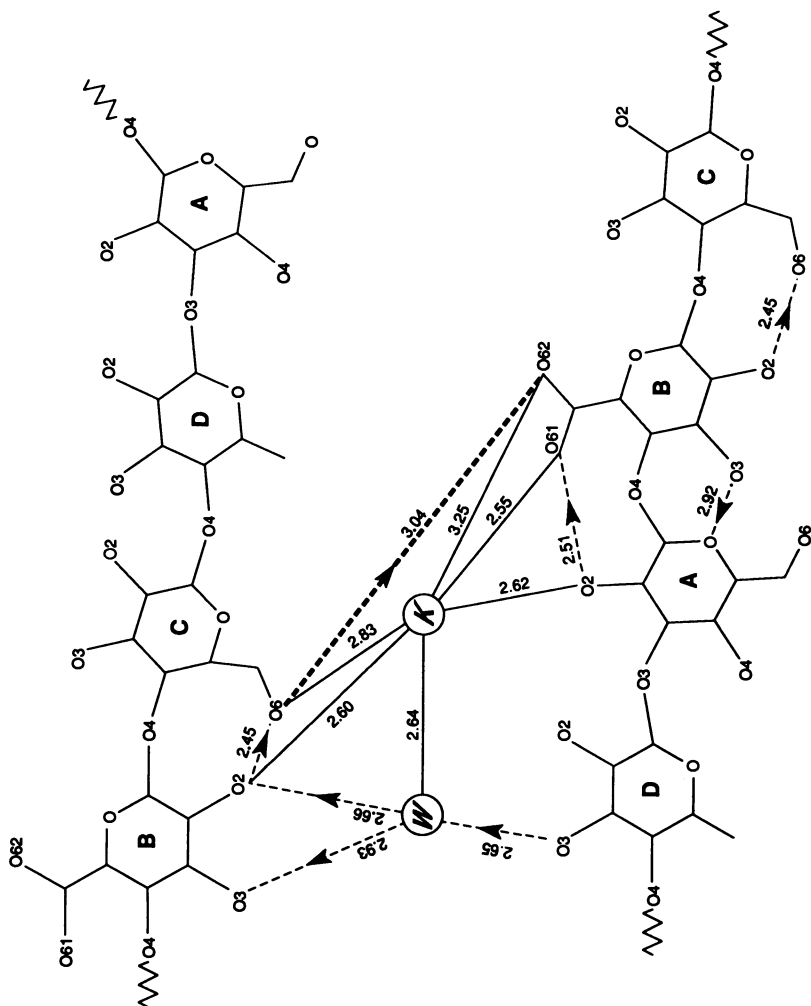
Figure 6. Schematic illustration of the hydrogen-bonding and potassium ion (K) coordination interactions in the gellan double-helix, including the first shell water molecule (W). The distances are given in Å units.

(Figure 7). The gellan molecules are crosslinked by double-helix --
potassium -- water -- potassium -- double-helix interactions which
are responsible for subsequent gelation process.

The packing arrangement further indicates that pairs of
carboxylate oxygen atoms belonging to adjacent double-helices are
5.1Å apart and that the two potassium ions coordinated to these
carboxylates are 4.3Å from each other. As speculated in our
previous study (13), using computer model building, we have been
able to substitute a single calcium ion, for the two potassium ions,
which makes a direct crosslink between the two double-helices (14)
and this is shown in Figure 8. Such direct crosslinks would be much
stronger than those observed in the case of potassium gellan. This
would explain the strong gelation behavior of calcium at a very
small fraction of the concentration required of potassium or other
monovalent ions (18,19). A view of the proposed crystal
structure model for calcium gellan in the *ab*-plane is shown in
Figure 9.

NATIVE GELLAN. The computer model building approach has further
enabled us to produce a stereochemically satisfactory molecular
structure of native gellan (14) which incorporates the same gellan
double-helix backbone as found in the potassium gellan crystal
structure (13). The acetate group could be readily attached to the
peripheral hydroxymethyl group of the glucose residue A in every
repeating motif and this addition does not perturb the packing
arrangement. On the other hand, the L-glycerate group could be
properly fixed only after a mild rotation of 30° for the carboxylate
group about its C5-C6 bond, along with a concerted 2Å movement of
the potassium ion and its first shell water molecule. Contrary to
previous expectations, the glycerate group is not only accommodated
in the double-helix, but its secondary hydroxyl group also
contributes additional stability by making an interchain hydrogen
bond with the oxygen atom O2 of the glucuronate residue. The
resulting morphology of the native gellan double-helix with
potassium coordination is illustrated in Figure 10.

We found it impossible to mimic the potassium gellan packing
arrangement for the native gellan double-helices due to severe
intermolecular short contacts involving the glycerate groups and
backbone atoms. These could not be relieved without a substantial
lateral expansion of the unit cell which particularly weakens the
intermolecular interactions mediated by carboxylate groups, cations
and water molecules that are crucial for molecular aggregation.
This might explain the observed weak and rubbery gels of this
material.

CONCLUSIONS

Computer model building and x-ray fiber diffraction studies on the
Rhizobium trifolii capsular polysaccharide and gellan have provided
the molecular geometries of some newly developed gel-forming
polysaccharides useful to the food industry. Preliminary single-
and double-helical models of RTPS would be stabilized primarily by
side chain -- backbone interactions. The crystal structure of
potassium gellan shows that the molecule is a half-staggered,
parallel double-helix stabilized by interchain hydrogen bonds and

2.815nm

2.815 nm
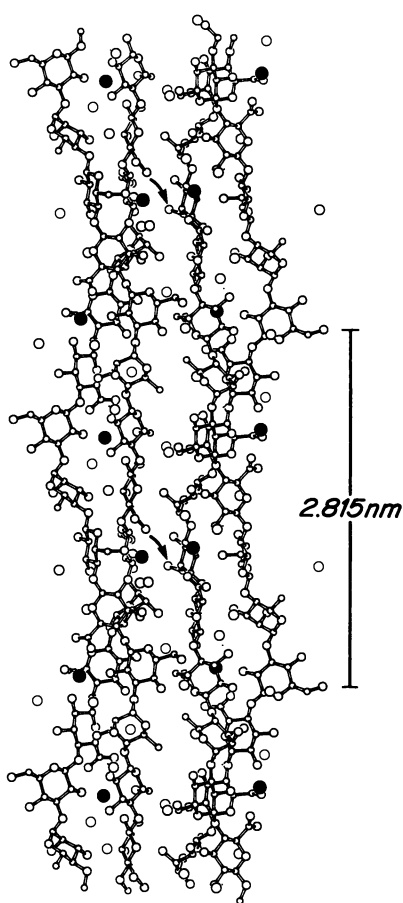
Figure 7. Packing of two adjacent, up and down, gellan molecules, the potassium ions (filled circles) and the water molecules (open circles) coordinated to them in the unit cell, viewed along the [110] direction. (Reproduced with permission from Ref. 13. Copyright 1988 Elsevier.)
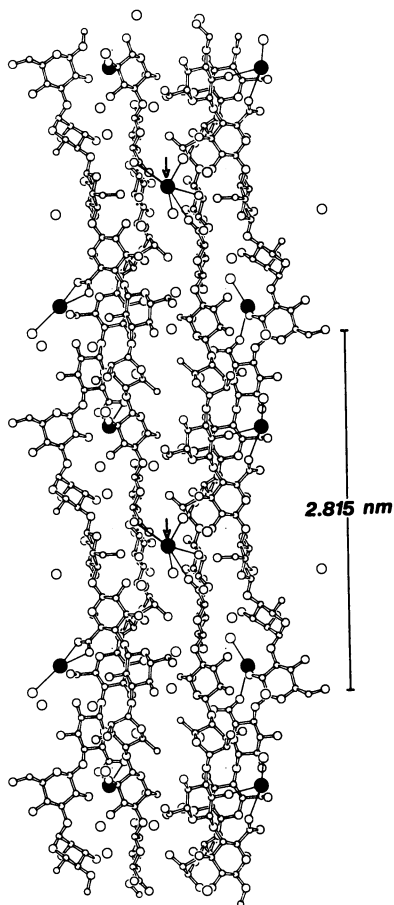
Figure 8. Proposed packing arrangement of calcium gellan, showing two, up and down, double-helices crosslinked at arrows by calcium ions (filled circles) and associated water molecules (open circles) in the unit cell, viewed normal to the molecular axes. (Reproduced with permission from Ref. 14. Copyright 1989 Elsevier.)
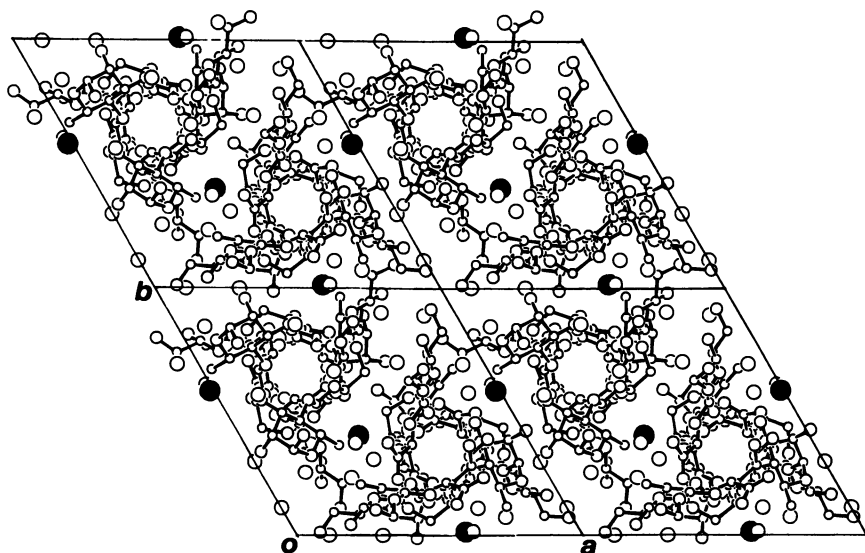
Figure 9. The crystal structure model for calcium gellan
viewed along the *c*-axis. The calcium ions are shown by filled
circles and water molecules by open circles. (Reproduced with
permission from Ref. 14. Copyright 1989 Elsevier.)
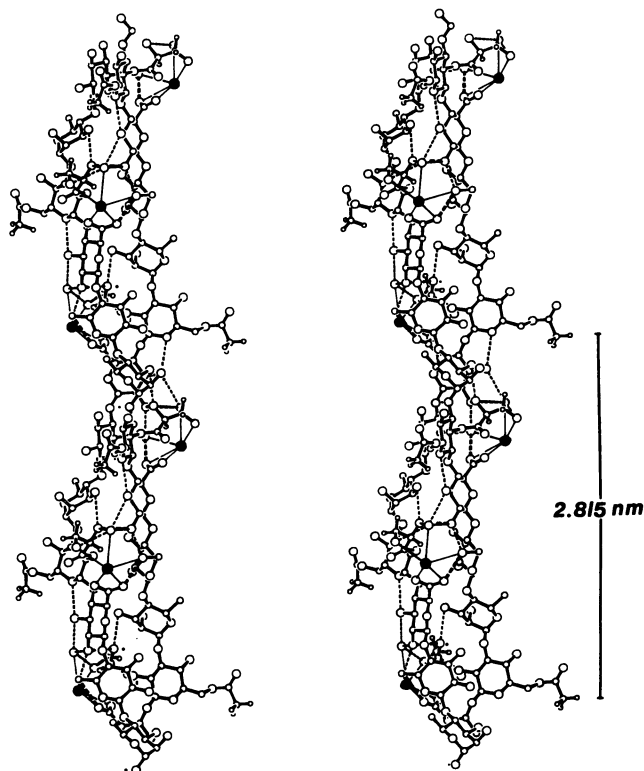
Figure 10. A stereo view of the native gellan double-helix and the coordinating potassium ions (filled circles) and water molecules (open circles). The intrachain (thin dashed lines), interchain (thick dashed lines) hydrogen bonds and ligands to potassium ions (thin lines) stabilize the double-helix. Both acetate and glycerate groups are shown along with their hydrogen atoms. (Reproduced with permission from Ref. 14. Copyright 1989 Elsevier.)

via coordination of potassium ions with the carboxylate groups. Antiparallel packing arrangement of double-helices leads to the aggregation of molecules and subsequent gelation process. Extrapolation of the potassium crystal structure using model building techniques has enabled us to visualize the most probable calcium gellan crystal structure which explains its strong gelling behavior at very low concentrations. Native gellan which has L-glycerate and acetate substitutions can also exist as a double-helix. Its weak gelling behavior apparently is due to its inability to retain the tight packing arrangement observed for the deacetylated material.

ACKNOWLEDGMENTS

Literature Cited

1. Arnott, S. In Developments in Food Carbohydrate; Birch, G.G.; Shallenberger, R.S., Eds.; Applied Science Publishers: England, 1977; pp 43-60.

2. Arnott, S.; Chandrasekaran, R. In Biopolymeri; Proceedings of the Italian Association for the Sicence and Technology of Macromolecules Summer School, Cargnano di Garcia, Italy, 1984; XXIII 1-12.

3. Arnott, S. In Fiber Diffraction Methods; French, A.D.; Gardner, K.H., Eds.; ACS Symposium Series No. 14; American Chemical Society: Washington, D.C., 1980; pp 1-30.

4. Smith, P.J.C; Arnott, S. Acta Crystallogr. 1978, A34, 3-11.

5. Hamilton, W.C. Acta Crystallogr. 1965, 18, 502-10.

6. Millane, R.P.; Byler, M.A.; Arnott, S. In Supercomputer Applications; Numrich; R.W., Ed.; Plenum Publishers: New York, 1985; pp 289-302.

7. Millane, R.P.; Arnott, S. J. Appl. Crystallogr. 1985, 18, 419-23.

8. Millane, R.P.; Arnott, S. J. Macromol. Sci., Physics 1985, B24, 193-227.

9. Arnott, S.; Chandrasekaran, R.; Millane, R.P.; Park, H.-S. J. Mol. Biol. 1986, 188, 631-40.

10. Park, H.-S.; Arnott, S.; Chandrasekaran, R.; Millane, R.P.; Campagnari, F. J. Mol. Biol. 1987, 197, 513-23.

11. Millane, R.P.; Chandrasekaran, R.; Arnott, S.; Dea, I.C.M. Carbohydr. Res. 1988, 182, 1-17.

12. Chandrasekaran, R. In Frontiers in Carbohydrate Research: Food Applications; Millane, R. P.; BeMiller, J.N.; Chandrasekaran, R., Eds.; Elsevier: New York, 1989, in press.

13. Chandrasekaran, R.; Puigjaner, L.C.; Joyce, K.L.; Arnott, S. Carbohydr. Res. 1988, 181, 23-40.

14. Chandrasekaran, R.; Thailambal, V.G. Carbohydr. Polymers 1989, in press.

15. Gidley, M.J.; Dea, I.C.M.; Eggleston, G.; Morris, E.R. Carbohydr. Res. 1987, 160, 381-96.

16. Chandrasekaran, R.; Millane, R.P.; Walker, J.K.; Arnott, S.; Dea, I.C.M. In Industrial Polysaccharides: The Impact of Biotechnology and Advanced Methodologies; Stivala, S.S.; Crescenzi, V.; Dea, I.C.M., Eds.; Gordon and Breach: New York, 1987; pp 111-8.

17. Chandrasekaran, R.; Millane, R.P.; Arnott, S. In Gums and Stabilisers for the Food Industry 4; Phillips, G.O.; Wedlock, D.J.; Williams, P.A., Eds.; IRL Press: Oxford, 1988; pp 183-91.

18. Crescenzi, V.; Dentini, M.; Coviello, T. Carbohydr. Res. 1986, 149, 425-32.

19. Grasdalen, H.; Smidsrod, O. Carbohydr. Polymers 1987, 7, 371-94.

20. Kuo, M.-S.; Dell, A.; Mort, A.J. Carbohydr. Res. 1986, 156,173-87.

21. Moorhouse, R. In Industrial Polysaccharides: Genetic Engineering, Structure/Property Relations and Applications; Yalpani, M., Ed.; Elsevier: Amsterdam, 1987; pp 187-206.

22. Chandrasekaran, R.; Millane, R.P.; Arnott, S.; Atkins, E.D.T. Carbohydr. Res. 1988, 175, 1-15.

Chapter 19

Polysaccharide Structures

X-ray Fiber Diffraction Studies

R. P. Millane

Whistler Center for Carbohydrate Research, Purdue University, West Lafayette, IN 47907

X-ray fiber diffraction can be used to visualize ordered structures of polysaccharides at atomic resolution. The structures provide information on structure/property relationships in these systems. X-ray fiber diffraction techniques are outlined and are illustrated with applications to typical polysaccharides.

X-ray fiber diffraction is used to determine the three-dimensional structures of polymeric (or polymer-like) materials (1,2). These molecules usually do not form three dimensional crystals and are therefore not amenable to the powerful methods of conventional x-ray crystallography. Fiber diffraction is used if the molecules can be prepared as fibers in which the molecular axes are approximately parallel. The molecules or crystallites, however, are randomly rotated about their long axes. The numbers and types of such materials are large, and fiber diffraction studies have played an important role in determinations of the structures of DNA, polysaccharides and other linear biopolymers. Random rotation of the molecules or crystallites results in cylindrical averaging of the diffraction pattern. This reduces the amount of diffraction data available and poses unique problems in the analysis compared to single crystal crystallography. I describe in this chapter the practical application of fiber diffraction to determining polysaccharide structures, and the critical role played by modeling.

Many water soluble polysaccharides control rheology in biological and commercial systems by thickening solutions or forming gels (3,4). In some cases, ordered, but often non-interacting, polymer segments contribute to solution viscosity as a result of the increased hydrodynamic volume of the molecules or aggregates. Polysaccharide gels are formed by cooperative associations of structurally regular polymer segments into interchain junction zones that are the tie-points of an infinite network. The chains in the junction zones pack together in an ordered and geometrically regular

manner, analogous to chain packing in the solid state. The network is solubilized by interconnecting chain segments that are disordered and conformationally mobile, as in solution. Hence, knowledge of the secondary (molecular) and tertiary (intermolecular) structures of polysaccharides is crucial to understanding the properties of these systems, particularly where ordered structures are involved in junction zones. For thickening polysaccharides, the conformations derived can provide a starting point for estimating chain flexibility and thence hydrodynamic parameters.

X-ray fiber diffraction analysis is the only means by which one can accurately visualize a highly hydrated polymer system at atomic resolution ([1],[2]). Although the organization in a fiber specimen is usually artificial, its details can help illuminate the ordered states of polysaccharides that occur in solutions and gels. Fiber diffraction studies, therefore, can prompt, facilitate and constrain modeling of complex systems by providing physically plausible, tangible, and precise components. The number of measured diffraction data is usually less than the number of independent atoms in the repeating unit of the molecule, and they rarely extend beyond 3 Å resolution. They are therefore insufficient, on their own, for direct structure determination. However, these data can be systematically augmented with reliable stereochemical information, including the primary polymer structure, helix pitch and symmetry, probable sugar ring geometries, characteristic hydrogen bond and polar interaction geometries, and the requirement that the distances between non-bonded atoms be longer than minimum acceptable values. Structure determination therefore involves stereochemical modeling of the polymer and its packing to supplement the diffraction data. The melding together of these rather different kinds of information can lead to a very detailed structure in which most of the atomic positions are defined to within a few tenths of an Angstrom, which is a precision adequate for identifying the critical interactions within and between molecules.

Fiber diffraction methods are outlined in the next section. In the following section, examples of the application of these techniques to some polysaccharide structures are described. The examples have been chosen to illustrate a range of situations that can arise in a fiber diffraction analysis.

X-Ray Fiber Diffraction

Polymer specimens suitable for fiber diffraction analysis are prepared by stretching hydrated films or fibers (while optimizing conditions such as humidity and salt concentration) to induce molecular orientation and crystallinity. In a non-crystalline fiber specimen, the molecules are approximately parallel but are randomly rotated relative to each other about their long axes, and are not organized laterally. In a polycrystalline specimen, the molecules are parallel too, but are also organized laterally into microcrystallites (each of which is a very small three-dimensional crystal) that are randomly rotated relative to each other. An x-ray diffraction pattern is obtained by exposing the fiber specimen to a colimated, monochromatic x-ray beam, and the diffracted x-rays detected using x-ray sensitive photographic film.

Polymers invariably form helical structures, and the helix symmetry is denoted by u_v, indicating that there are u repeat units in v turns of the helix. The helix pitch is denoted by P and the molecular repeat distance is c = vP. X-ray diffraction patterns from non-crystalline specimens contain diffracted intensities restricted to <u>layer lines</u> that are spaced by 1/c. On a diffraction pattern from a polycrystalline specimen, diffraction signals, or <u>Bragg reflections</u>, occur only at discrete positions on the layer lines, the positions being related to the lateral dimensions of the unit cell of the crystal. The meridian (vertical axis) of the diffraction pattern is devoid of diffracted intensity unless the layer line number ℓ is a multiple of u, so that u can be determined straightforwardly. The diffracted intensities can be calculated using standard expressions (<u>2</u>), for model structures (i.e. given the atomic coordinates).

The x-ray diffraction data consist of samples of the diffracted intensity along the layer lines in the case of patterns from noncrystalline specimens, and intensities of the Bragg reflections for patterns from polycrystalline specimens. The diffraction pattern is digitized using a scanning microdensitomer and the diffracted intensities extracted using special purpose software (<u>5-10</u>). Fiber diffraction patterns contain a significant amount of background scatter due to amorphous material in the specimen that must be estimated and removed (<u>5,9</u>). Intensities along layer lines are most accurately measured using deconvolution or profile fitting techniques (<u>6,8</u>). Intensities of Bragg reflections are determined by integrating the film optical densities over the region of the spot (<u>5,7,8,10</u>).

Using the pitch, symmetry, monomer geometries and other stereochemical constraints, a number of <u>types</u> of molecular model can be constructed. Typical dilemmas are whether the molecular helix is left- or right-handed, whether the molecule is a single helix or co-axial double-helix (and in the later case whether the two chains in the duplex are parallel or antiparallel), or whether, if there are two or more molecules in the unit cell, the molecules are parallel or antiparallel. Solution of a structure therefore involves refinement and adjudication: All candidate models are refined until the fit with the measured x-ray amplitudes or steric factors allows one model to be declared significantly superior to the others by some standard statistical test.

The two programs most widely used for analysis and refinement of polymer models are the linked-atom least-squares (LALS) system (<u>11,12,13</u>) and the variable virtual bond (PS79) method (<u>14</u>). The principles of these two systems are the same and comparisons show that they give similar results (<u>15</u>). The LALS method has been used more extensively and is the one described here. The system has been used to refine a wide range of polysaccharides, polynucleotides, polyesters and polypeptides (<u>1,16</u>). The bond lengths and angles in polymers have very nearly the same values as in the corresponding monomers and so are known rather precisely (from single crystal analyses). This reduces the solution of the polymer structure to that of determining conformation angles about single bonds. Also, the conformations of some structural entities such as rigid rings may also be assumed to persist from monomer to polymer. It is

therefore possible to prepare a linked-atom description of a molecule in which interatomic relationships are described in terms of bond lengths, bond angles and conformation angles. A further source of stereochemical data is the requirement that a model exhibit no over-short nonbonded interatomic distances.

In the LALS system, the quantity Ω given by

$$\Omega = \sum_m e_m \Delta\theta_m^2 + \sum_m \omega_m \Delta F_m^2 + \sum_m k_m \Delta d_m^2 + \sum_m \lambda_m G_m \qquad (1)$$

$$= \quad E \quad + \quad X \quad + \quad C \quad + \quad L$$

is minimized by varying a set of chosen parameters consisting of conformation angles, packing parameters, and x-ray scale and attenuation factors. Preferred conformational domains can be achieved by restraining conformation angles (θ_m) to their respective expected values, and they constitute the first term E. The term X involves the differences ΔF_m between the model and experimental x-ray amplitudes - Bragg and/or continuous. The term C involves restraints, used to ensure that over-short non-bonded interatomic distances are driven beyond acceptable minimum values, that hydrogen bond and coordination geometries are close to the expected configurations, as well as a variety of other geometric relationships (12). The e_m, ω_m and k_m are weights that are inversely proportional to the estimated variances of the data. The term L involves constraints that are relationships to be satisfied exactly ($G_m = 0$) and the λ_m are Lagrange multipliers. Constraints are used, for example, to ensure connectivity from one helix pitch to the next and to ensure that chemical ring systems are closed. The ratio $(\Omega_P/\Omega_Q)^{\frac{1}{2}}$ can be used in Hamilton's test (17) to assess the significance of differences between models P and Q. The ratio $(X_P/X_Q)^{\frac{1}{2}}$ can be used to decide whether the x-ray data alone are sensitive to the differences between the two models. These tests are used to reject putative models, which is a necessary part of structure determination. In the final stages of refinement, bond angles and additional conformation angles can be allowed to vary in a "stiffly elastic" fashion from their mean values if there are sufficient data to justify the increased number of parameters. If sufficient x-ray data are available, it is sometimes possible to locate ordered guest molecules, such as counterions and water molecules, in the crystal structure by difference Fourier synthesis using phases derived from the polymer model (18). Location of such molecules can provide useful information on the mechanisms by which these guest molecules stabilize cooperative interactions between ordered polymer segments.

The overall scheme for determining polymer structures using this technique is outlined in Figure 1. The exact sequence of steps depends on the particular problem at hand, but the overall strategy is generally as follows: A linked-atom description of the molecule is prepared using the geometry of the fixed groups in the structure. The model is optimized using LALS, but excluding the term X in (1), by varying parameters (for polysaccharides these would correspond to conformation angles at the glycosidic linkages and about single bonds in side groups) so that it has the correct

```
┌─────────────────────────────────────┐
│ FIBER DIFFRACTION PATTERN            │
└─────────────────────────────────────┘

              ┌───────────────────────────────────┐
              │ POLYMER PRIMARY STRUCTURE          │
              └───────────────────────────────────┘

                        ┌─────────────────────────────────────┐
                        │ STANDARD BOND LENGTHS AND ANGLES     │
                        └─────────────────────────────────────┘

┌──────────────────────┐  ┌──────────────────────────────────┐
│ If specimen is       │  │ Produce MOLECULAR MODELS constrained │
│ oriented, determine  │  │ to have appropriate pitch, symmetry │
│ HELIX PITCH and      │  │ and restrained to have minimum steric │
│ possible MOLECULAR   │  │ compression, optimum hydrogen      │
│ SYMMETRIES.          │  │ bonding, etc.  Attempt decision among │
└──────────────────────┘  │ symmetry choices.                  │
                          └──────────────────────────────────┘

┌──────────────────────┐  ┌──────────────────────────────────┐
│ If specimen is also  │  │ Determine possible modes           │
│ crystalline determine│  │ of CHAIN PACKING                   │
│ DENSITY, UNIT CELL and│ └──────────────────────────────────┘
│ possible SPACE GROUPS.│
└──────────────────────┘  ┌──────────────────────────────────┐
                          │ Optimise models to fit            │
                          │ X-RAY INTENSITIES while           │
                          │ maintaining CONSTRAINTS           │
                          │ and STERIC RESTRAINTS.            │
                          │ Attempt decision among            │
                          │ symmetry and packing              │
                          │ choices.                          │
                          └──────────────────────────────────┘

┌───────────────────────────────────────────────────────┐
│ Use ELECTRON DENSITY DIFFERENCE SYNTHESIS to           │
│ determine possible ION and/or WATER sites.             │
└───────────────────────────────────────────────────────┘

┌───────────────────────────────────────────────────────┐
│ Refine augmented CRYSTAL MODEL until complete.         │
└───────────────────────────────────────────────────────┘
```

Figure 1. Scheme for defining and refining polymer
structures. (Reproduced from ref. 1. Copyright
1980 American Chemical Society.)

pitch and helix symmetry while minimizing the steric compression. At this stage it may be possible to reject some of the models (particular symmetries for example) because of excessive steric compression (over-short non-bonded interatomic distances). If the specimen is polycrystalline, the surviving molecular models are packed in the unit cell and both the molecular and packing parameters are refined to minimize the steric compression. There are usually a number of possible packing arrangements and some of these may be eliminated on steric grounds as a result of these refinements. The surviving crystal structure models (or molecular models if the specimen is non-crystalline) are then co-refined against the steric compression and the diffraction data. In a final comparison of these optimized models, one hopes to be able to eliminate all but one model, on the basis of Ω, X and/or C in (1), leaving the correct structure. This structure may be further refined by varing additional parameters and using difference Fourier synthesis as described above.

Structures

The goal in a fiber diffraction analysis is to use the diffraction data to the greatest extent possible as they represent the primary experimental data. The stereochemical information is used to make up for deficiencies in the diffraction data. I have therefore chosen to illustrate these techniques with a number of structures that yield different qualities of diffraction data. Further details are given in the articles referenced. Chondroitin 4-sulfate is a glycosaminoglycan that can be prepared as well oriented polycrystalline fibers that give good diffraction patterns and hence much structural information can be derived. The carrageenan/agar family of algal polysaccharides give diffraction patterns with a variety of resolutions and thus lead to structures of various precisions.

<u>Chondroitin 4-sulfate</u>. Chondroitin 4-sulfate is a glycosaminoglycan that is prominent in mammalian connective tissue, and is covalently bound to core proteins as a component of proteoglycans (<u>19</u>). Proteoglycans are extended structures occupying a large hydrodynamic volume, exhibit high viscosities, are reversibly compressible in solution, and play a variety of roles in tissue function. A number of glycosaminoglycan structures have been determined by fiber diffraction (see reference (<u>16</u>) for a review) including the calcium (<u>20</u>), sodium (<u>21</u>) and potassium (<u>18</u>) salts of chondroitin 4-sulfate. Determination of the structure of the potassium salt is outlined here.

The primary structure of chondroitin 4-sulfate is the repeated disaccharide

$$\rightarrow 3)-\beta-D-Glc-(1\rightarrow 4)-\beta-D-GalNac-4-sulfate-(1\rightarrow$$
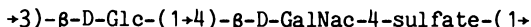
A fiber diffraction pattern of the potassium salt is shown in Figure 2 (<u>18</u>). Sharp Bragg reflections extend to approximately 3.0 Å resolution with meridional intensities on the 6th and 9th layer lines. The diffraction pattern can be indexed on the basis of a
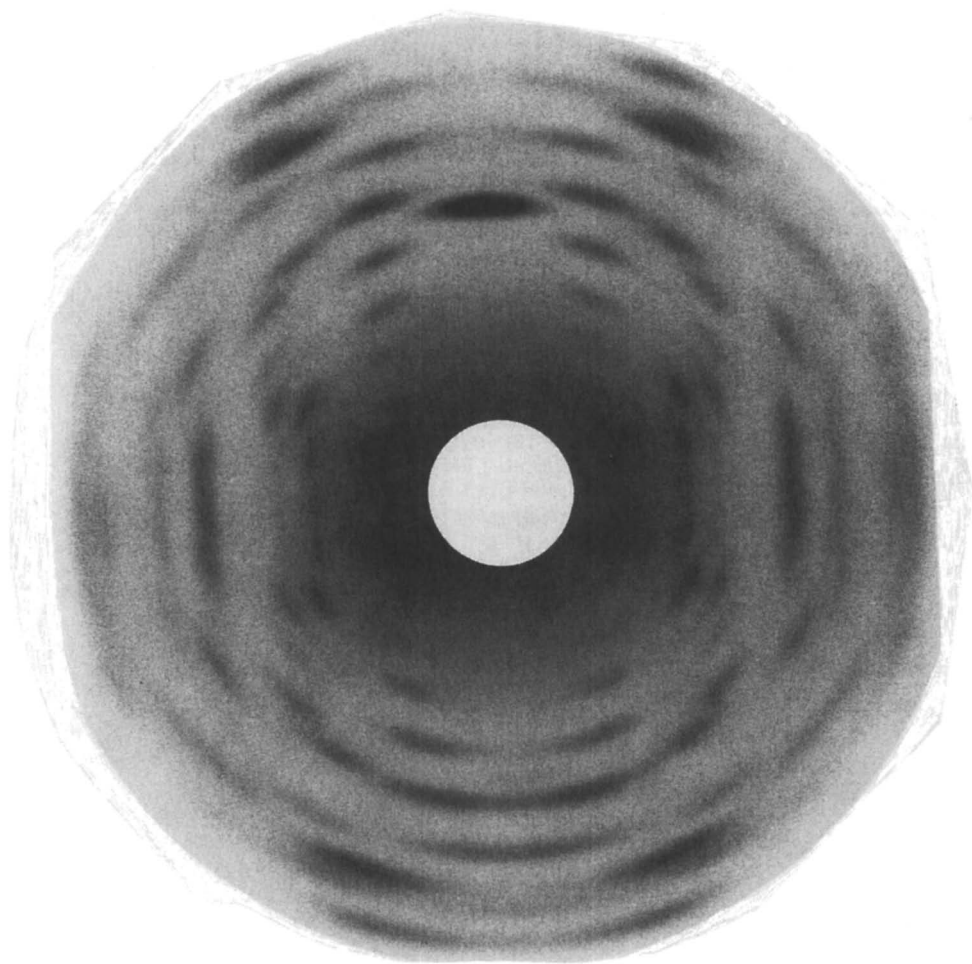
Figure 2.   X-ray fiber diffraction pattern from K$^+$ chondroitin 4-sulfate.   (Reproduced with permission from ref. 18.   Copyright 1983 Academic Press Inc.)

trigonal lattice with unit cell dimensions a = b = 13.85 Å and c = 27.76 Å. The systematic absences on the meridian indicate 3-fold helix symmetry (absence of meridional intensity on the third layer-line is accidental rather than systematic). The unit cell volume is consistent with two chains passing through the unit cell. Hence there are four possible crystal structure models containing left- or right-handed chains and parallel or antiparallel pairs of chains in the unit cell.

Isolated molecular models of appropriate pitch and symmetry of either chirality could be generated free of steric compression using standard (22) pyranose rings in the $^4C_1$ conformation. Neither parallel nor antiparallel double-helical models of either chirality are sterically feasible (18). The 4 conformation angles at the glycosidic oxygens and the 6 defining the side-group orientations were variables during these refinements. Two additional parameters are required to define the relative orientation (μ) and axial translation (w) of the two chains in the unit cell. To ensure that only two polyanion chains intersect the unit cell, the molecular screw axes must coincide with the space group screw axes at coordinates (2/3, 1/3) and (1/3, 2/3) in the unit cell base plane. Possible packings for antiparallel molecules were explored by calculating the contact term C in (1) while treating the refined molecular models as rigid bodies and packing them into the unit cell over the full range of μ and w. This gave two minima in (μ,w) space for each chirality, or four models. For the parallel chain models, $w = 0$ and the variation of C with μ was calculated in the same way, giving one minimum for each chirality (18). These six packing models were each optimized against the intra- and intermolecular steric compression and the x-ray data while varying the ten conformation angles and the packing parameters. Analysis of the optimized models showed that four of them could be rejected on the basis of poor agreement between the calculated and measured x-ray amplitudes (18). Both surviving models are packed antiparallel, one containing left-handed chains and the other right-handed chains.

Fourier difference maps calculated for the model with the right-handed chains showed no plausible cation sites near charged groups. For the models containing left-handed chains however, difference Fourier synthesis indicated a fully co-ordinated cation site in the vicinity of sulfate and carboxylate groups. This constitutes good evidence that the left-handed model is correct and further development of the crystal structure was undertaken on this basis. A potassium ion was positioned as indicated by the difference map and the whole structure (polyanion plus potassium ion) re-refined. Four periodic water molecules were similarly located in a succession of difference Fourier maps. Final structure refinement involved releasing the 12 conformation angles and 12 endocyclic bond angles of the pyranose rings, the two glycosidic bridge bond angles and the acetamido conformation angle, allowing each to vary in a stiffly elastic manner. The final model contains no overshort interatomic distances and all potential hydrogen bond donors are involved in hydrogen bonds. The crystallographic residual $R = \sum |F_o - F_c| / \sum F_o$ (where F_o and F_c are the observed and calculated x-ray amplitudes, respectively) is equal to 0.20.

The final endocyclic and glycosidic bridge bond angles deviated by a maximum of 3° and an average of 1° from their standard values. The ring conformation angles changed by a maximum of 5° and an average of 2°. The conformation of the polyanion is shown in Figure 3 and the (1→4) linkage is stabilized by an O3---O5 hydrogen bond. The crystal packing (Figure 4) is stabilized by hydrogen bonds between a nitrogen atom and a sulfate oxygen atom of adjacent antiparallel chains (18). The potassium ion lies on a dyad axis and coordinates to sulfate, carboxylate and ring oxygens in two antiparallel chains (18). Three of the four water molecules located provide intramolecular stabilization and one bridges between adjacent antiparallel chains (18).

Carrageenans. The carrageenan/agar family of polysaccharides are extracted from the marine algae Rhodophyceae and used in the food industry as gelling, thickening and stabilizing agents (23). All carrageenans are linear polysaccharides of alternating (1→3) and (1→4) linked galactose units, different fractions having different sulfate and 3,6-anhydride contents. The two principal gelling fractions, kappa- and iota-carrageenan, contain β-D-galactose and 3,6-anhydro-α-D-galactose units. In both, most D-galactose units carry a half-ester sulfate group at O4; the difference between them being that most 3,6-anhydro-galactose units are sulfated at O2 in iota- but not in kappa-carrageenan (Figure 5). Agarose differs from the carrageenans only in having the anhydrogalactose units replaced by the L-enantiomer, and being totally non-sulfated (Figure 5). Kappa- and iota-carrageenan, and agarose form thermally-reversible gels upon heating and cooling of aqueous solutions.

The structures of agarose, iota- and kappa-carrageenan have been determined by x-ray fiber diffraction (24-27). The quality of the diffraction data obtained from each of these three specimens varies considerably and the way in which these data are used in structure determination is outlined here. Diffraction patterns from oriented specimens of agarose, and kappa- and iota-carrageenan are shown in Figure 6 (28). The molecular repeat distances derived from these patterns are listed in Table I.

Table I. Molecular Repeat Distances and Pitches
of Agarose and Carrageenan Structures

Molecule	c (Å)	Pitch of a single chain (Å)
Agarose	9.5	19.0
Iota-carrageenan	13.3	26.6
Kappa-carrageenan	25.0	25.0

The agarose diffraction pattern (Figure 6a) contains intensity distributed continuously along layer lines indicating that the
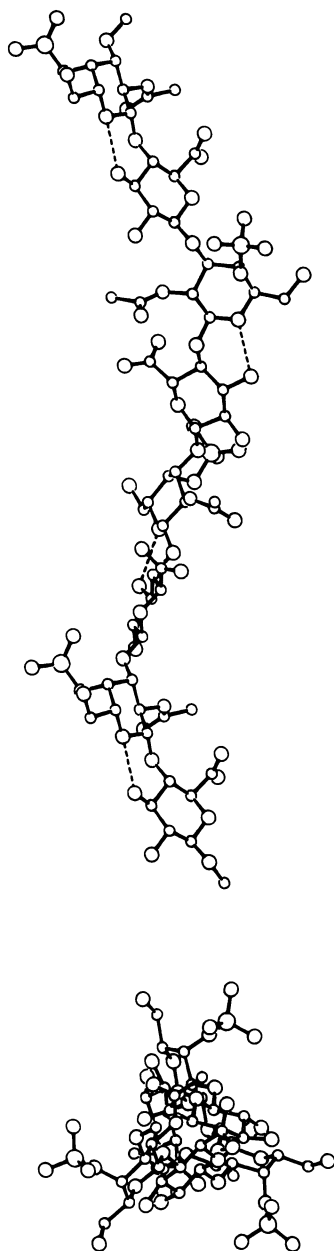
Figure 3. Conformation of the K$^+$ chondroitin 4-sulfate structure viewed normal to and along the helix axis. The O3···O5 hydrogen bonds are shown by broken lines. (Reproduced with permission from ref. 18. Copyright 1983 Academic Press Inc.)
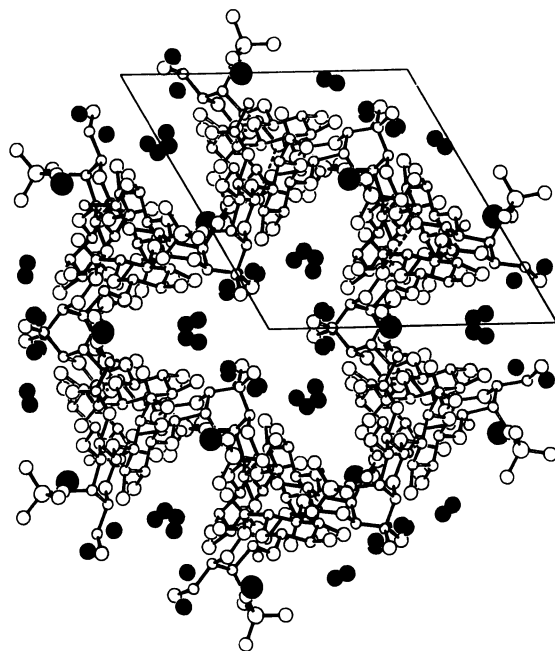
Figure 4. Helix axis projection of the K^+ chondroitin 4-sulfate crystal structure. Molecules with full and open bonds indicate up- and down-pointing chains. The large and small filled circles represent potassium ions and water molecules respectively. (Reproduced with permission from ref. 18. Copyright 1983 Academic Press Inc.)
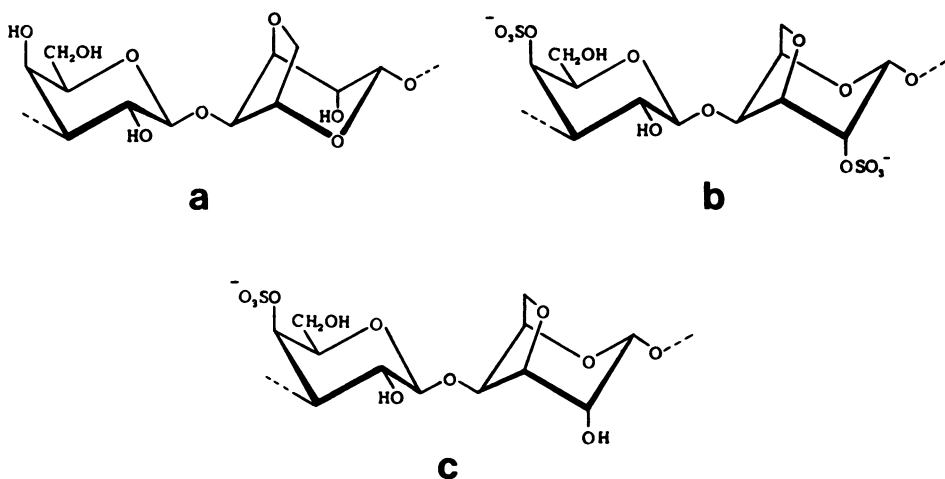
Figure 5. Chemical repeating units of (a) agarose, (b) iota-carrageenan, and (c) kappa-carrageenan.
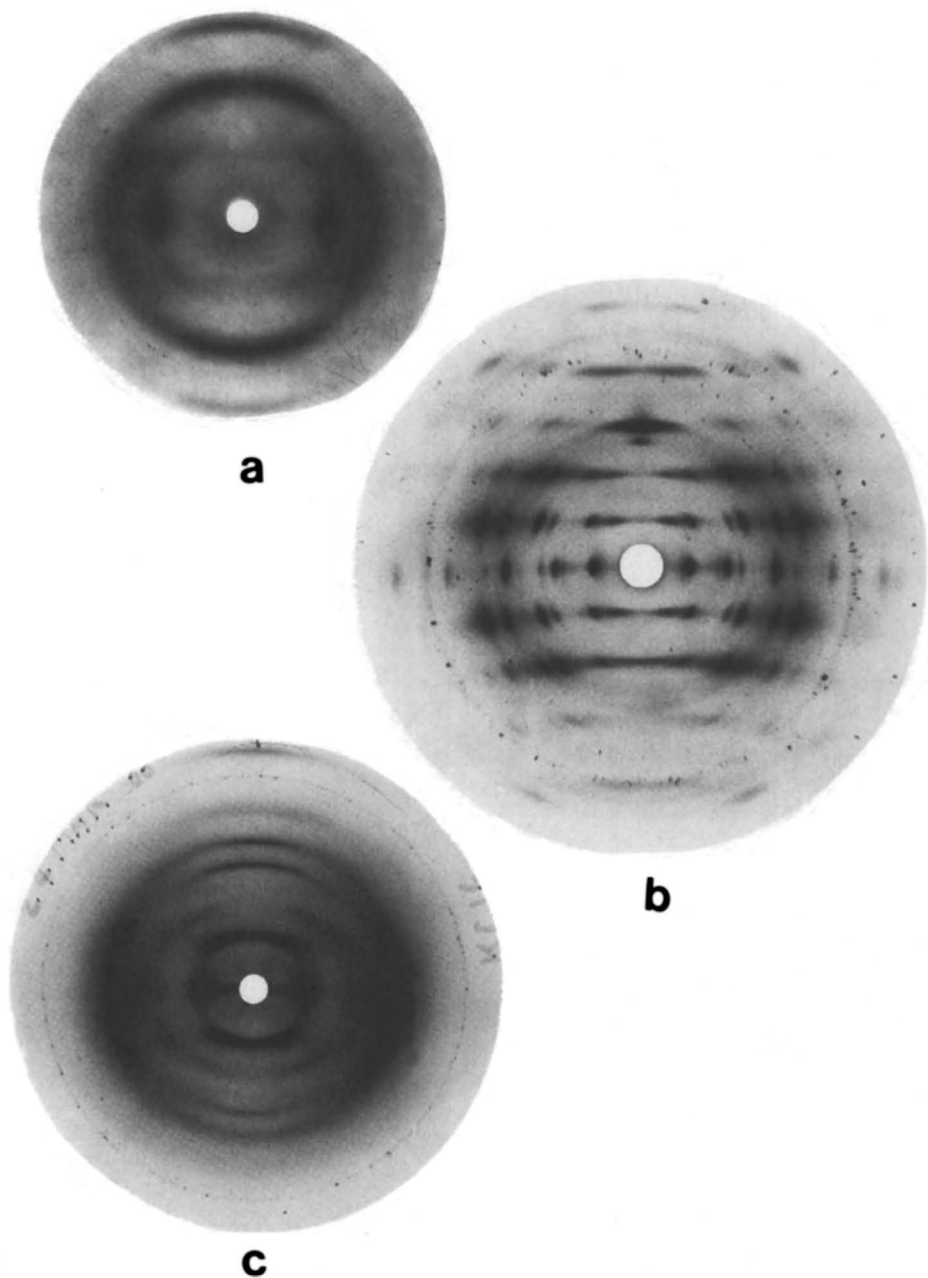
Figure 6. X-ray fiber diffraction patterns from (a) agarose, (b) Ca^{2+} iota-carrageenan, and (c) K^+ kappa-carrageenan. (Reproduced with permission from ref. 28. Copyright 1989 Elsevier.)

specimen is oriented but noncrystalline. The layer line spacing
corresponds to a repeat distance of 9.50 Å and the pattern of
meridional intensities indicates 3-fold helix symmetry.
Construction of molecular models consistent with the observed repeat
distance and symmetry shows that only double-helices containing
coaxial 3-fold chains of pitch 19.0 Å are sterically acceptable
(24). The two chains are translated relative to each other along
their common axis by exactly half the pitch, which reduces the
repeat distance to 9.50 Å. Double-helices containing either left or
right-handed chains are sterically acceptable. Agreement between
the measured and calculated diffracted amplitudes is better for the
double-helix model containing left-handed chains than for the one
containing right-handed chains. Left-handed chirality is also
supported by optical rotation measurements (24). It is concluded
therefore, that the double-helix contains left-handed chains. The
agarose double-helix is shown in Figure 7a, and a notable feature is
an interior cavity that extends along the helix axis. The atoms
forming the lining of this cavity include O2 of galactose and O5 of
anhydro-galactose, suggesting that it is occupied by water molecules
that participate in a hydrogen bonding system that contributes to
the stability of the double-helix. No other hydrogen bonding scheme
within the structure is suggested by the model. Three of the four
hydroxyl groups are on the outer periphery of the structure and
presumably engage in hydrogen bonding with neighboring helices or
solvent. Consistency between optical rotations calculated from the
double-helix and those observed during the sol-gel transition,
indicates that this double-helix and its aggregates are
representative of the junction zones in the gel network.

A diffraction pattern from the Ca^{2+} salt of iota-carrageenan
(Figure 6b) contains Bragg reflections showing that the molecules
pack on a trigonal lattice with unit cell dimensions a = b = 13.7 Å
and c = 13.3 Å (26). On each layer line there is also some
continuous intensity that is consistent with an array of molecules
packed in a regular lattice in which the sites are occupied randomly
by up- and down-pointing molecules (1,26). Diffraction patterns
from the Mg^{2+} salt prepared by rapid drying include weak additional
layer lines that interleave those in other patterns (26). This
indicates an underlying periodicity of 26.6 Å and extinction of
every second layer line is most simply explained by a coaxial
double-helix structure containing two parallel chains of pitch 26.6
Å that are translated by exactly half the pitch. Right-handed
chains are sterically less compressed than left-handed chains, and
are also supported by optical rotation measurements (29). This
double-helix model is stabilized by an inter-chain hydrogen bond
between O2 and O6 of the galactose residues and is shown in Figure
7b. Incorporating the double helix in the crystal lattice and
refining the molecule as well as varying the relative positions of
the up- and down-pointing molecules in the "statistically
crystalline" lattice, indicates an average symmetry of the molecular
array of $P3_2 12$ and gives good agreement between the calculated and
measured diffraction amplitudes (26). Difference Fourier synthesis
indicates the presence of ordered calcium ions in the structure and
incorporation of these results in a model with acceptable
stereochemistry and improved x-ray agreement following further

refinement (26). The iota-carrageenan double-helix has a compact
carbohydrate core with sulfates protruding in pairs away from the
center of the helix (Figure 7b). Each calcium ion is coordinated to
two sulfate groups, each from a different double-helix, in such a
way as to produce a $--Ca^{2+}---SO_4^---Ca^{2+}---SO_4^---$ chain running
through the lattice parallel to the fiber axis (26). Each galactose
sulfate contributes one oxygen atom to the coordination sphere of
each of two cations. Thus each double helix is linked to its
neighbors by cations, forming a three-dimensional network.

A diffraction pattern from the K^+ salt of kappa-carrageenan
(Figure 6c) contains continuous intensities on layer lines so that,
as with agarose, the molecules are approximately parallel but there
is no lateral organization (27). The pattern indicates 3-fold helix
symmetry and the layer line spacing (25.0 Å) is approximately twice
that of iota-carrageenan (Table I). It follows that the structure
of kappa-carrageenan is not a half-staggered double-helix of the
iota-carrageenan type and a larger number of sterically plausible
models must be considered. Obvious candidate models are coaxial
duplexes with parallel but not half-staggered chains, or single-
stranded molecules with conformations similar to one chain of an
iota duplex. Coaxial duplexes with antiparallel chains would also
preserve the 25.0 Å layer line spacings. Both left- and right-
handed single chain models free of steric compression could be
constructed. Possible juxtapositions of single chains to form
coaxial double-helices were explored by calculating maps of the
steric compression C as a function of the relative orientation (μ)
and axial displacement (w) of the two chains. Parallel and
antiparallel double-helices containing left-handed chains contained
serious steric anomolies and were rejected. Surveys of the steric
compression for double-helices containing right-handed, parallel and
antiparallel chains showed that a wide variety of such models are
sterically feasible. The first structural question is whether the
diffraction data supports single helices or minor variations on an
iota-like structure. Left- and right-handed single helix models,
and double-helix models containing parallel right-handed chains
(starting at a variety of juxtapositions of the two chains), were
refined against the continuous x-ray data. One model for each
single helix, and four double-helix models resulted from these
refinements (27). Both single helix models are quite inferior to
all the parallel double-helix models due to poor x-ray agreement,
and were therefore rejected. The best parallel double-helix model
is offset, by a rotation of 28° and a translation of 1.0 Å, from the
half-staggered position and is shown in Figure 7c. One of the two
O2---O6 interchain hydrogen bonds per disaccharide in iota-
carrageenan is retained (Figure 7c). The conformation angles at the
glycosidic linkages are somwehat different to those of iota-
carrageenan (26) which could be due to the different juxtapositions
of the two chains, or to the difference in pitch (ΔP = 1.6 Å)
between the two structures. Double-helix models containing
antiparallel chains were refined starting at various juxtaposition
of the chains and this led to seven sterically acceptable models
(27). Each of these models is inferior to the parallel model
described above, although in some cases the discrimination is not
compelling.
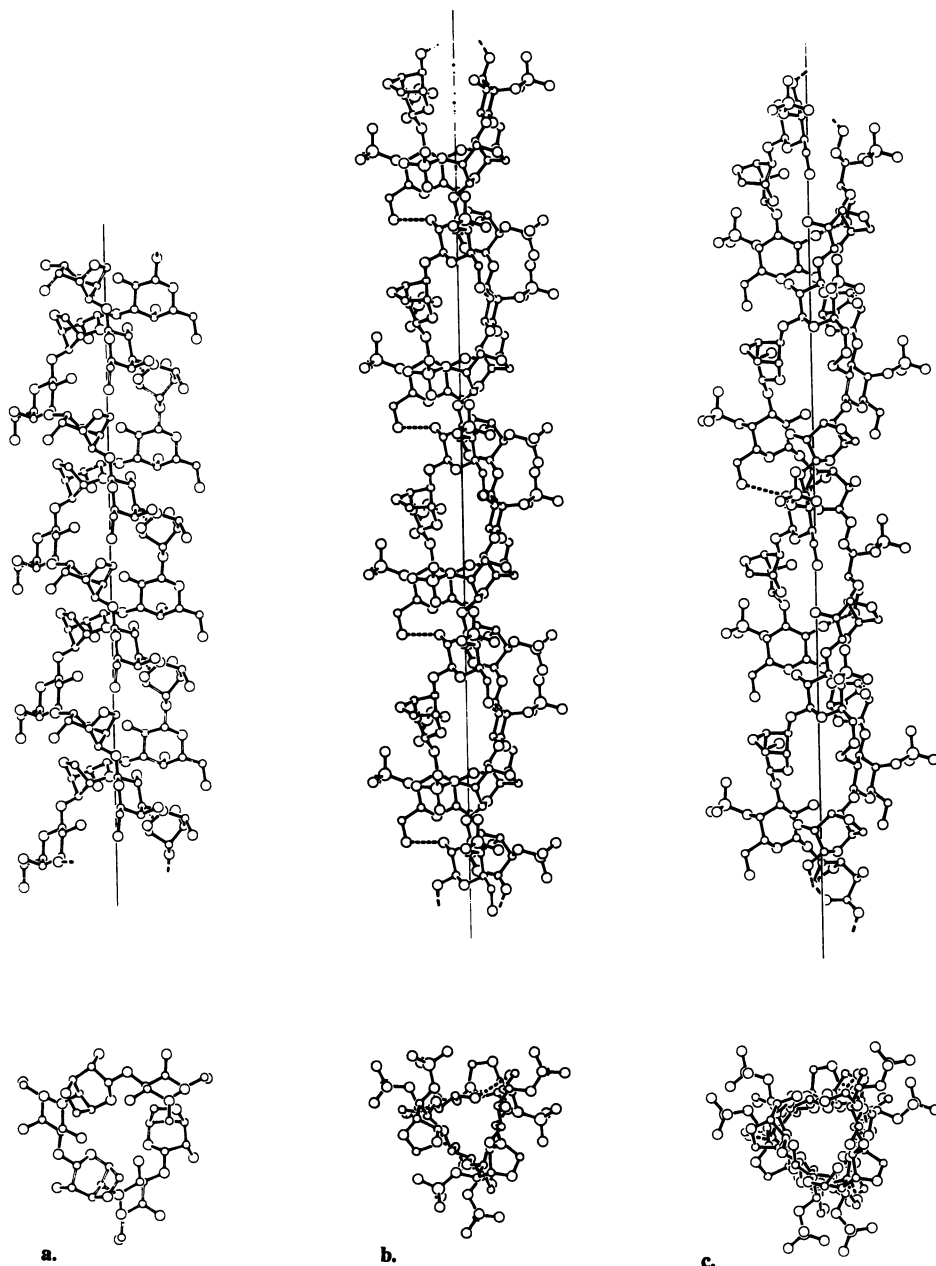
Figure 7. Mutually perpendicular views of the (a) agarose, (b) iota-carrageenan, and (c) kappa-carrageenan double-helix structures. The two chains are shown with open and full bonds, and the O6---O2 hydrogen bonds by broken lines. (Reproduced with permission from ref. 28. Copyright 1989 Elsevier.)

Conclusions

X-ray fiber diffraction can be used to visualize highly hydrated
polymer specimens at atomic resolution. An essential part of such
an analysis is the inclusion of reliable stereochemical information
to supplement the diffraction data. Structure determination
involves modelling and refinement of putative structures, and
adjudication amongst the optimized models. This technique has been
successfully applied to a number of polysaccharides. The precision
of resulting structures is often sufficient to identify the critical
interactions within and between molecules, that are responsible for
the unique properties of these materials.

Acknowledgments

Literature Cited

1. Arnott, S. In Fiber Diffraction Methods; French A. D.;
    Gardner, K. H., Eds.; ACS Symposium Series No. 141; American
    Chemical Society: Washington, DC, 1980; pp 1-30.
2. Millane, R. P. In Computing in Crystallography 4: Techniques
    and New Technologies; Isaacs, N. W.; Taylor, M. R., Eds.;
    Oxford Univ. Press, Oxford; 1988, pp 169-186.
3. Morris, E. R. In Gums and Stabilizers for the Food Industry 3;
    Phillips, G. O.; Wedlock, D. J.; Williams, P. A., Eds.;
    Elsevier: New York, 1986; pp 3-16.
4. Aspinall, G. O., Ed.; The Polysaccharides; Academic Press: New
    York, 1982.
5. Fraser, R. D. B.; MacRae, T. P.; Miller, A.; Rowlands, R. J.
    J. Appl. Crystallogr. 1976, 9, 81-94.
6. Makowski, L. J. Appl. Crystallogr. 1978, 11, 273-283.
7. Fraser, R. D. B.; Suzuki, E.; MacRae, T. P. In Structure of
    Crystalline Polymers; Hall, I. H., Ed.; Elsevier: New York,
    1984, pp 1-37.
8. Millane, R. P.; Arnott, S. J. Macromol. Sci. - Phys. 1985,
    B24, 193-227.
9. Millane, R. P.; Arnott, S. J. Appl. Crystallogr. 1985, 18,
    419-423.
10. Millane, R. P. J. Macromolecular Sci. - Phys. 1989, B28, 149-
    166.
11. Arnott, S.; Wonacott, A.J. Polymer, 1966, 7, 157-166.
12. Smith, P. J. C.; Arnott, S. Acta Crystallogr. 1978, A34, 3-11.
13. Millane, R. P.; Byler, M. A.; Arnott, S. In Supercomputer
    Applications; Numrich, R. W., Ed.; Plenum: New York, 1985; pp
    289-302.
14. Zugenmaier, P.; Sarko, A. In Fiber Diffraction Methods;
    French, A.D.; Gardner, K.H., Eds.; ACS Symposium Series Vol.
    141, American Chemical Society: Washington, DC, 1980; pp. 225-
    237.

15. Millane, R. P.; Narasaiah, T. V. Polymer 1989, 30, 1763-1767.
16. Arnott, S.; Mitra, A. K. In Molecular Biophysics of the Extracellular Matrix; Arnott, S.; Rees, D. A.; Morris, E. R., Eds.; Humana Press: Clifton, New Jersey, 1984, pp. 41-67.
17. Hamilton, W. C. Acta Crystallogr. 1965, 18, 502-510.
18. Millane, R. P.; Mitra, A. K.; Arnott, S. J. Mol. Biol. 1983, 169, 903-920.
19. Hascall, V. C.; Hascall, G. K. In Cell Biology of the Extracellular Matrix; Hay, E. D., Ed.; Plenum: New York; 1981, pp. 39-63.
20. Cael, J. J.; Winter, W. T.; Arnott, S. J. Mol. Biol. 1978, 125, 21-42.
21. Winter, W. T.; Arnott, S.; Issac, D. H.; Atkins, E. D. T. J. Mol. Biol. 1978, 125, 1-19.
22. Arnott, S.; Scott, W. E. J. Chem. Soc. Perkins Trans. II 1972, 324-335.
23. Glicksman, M. In Food Hydrocolloids; Glicksman, M., Ed.; CRC Press: Boca Raton, 1983, Vol. 2, pp 73-113.
24. Arnott, S.; Fulmer, A.; Scott, A. E.; Dea, I. C. M.; Moorhouse, R.; Rees, D. A. J. Molecular Biology 1975, 90, 269-284.
25. Anderson, N. S.; Campbell, J. W.; Harding, M. M.; Rees, D. A.; Samuel, J. W. B. J. Mol. Biol., 1969, 45, 85-99.
26. Arnott, S.; Scott, W. E.; Rees, D. A.; McNab, G. C. A. J. Mol. Biol. 1974, 90, 253-267.
27. Millane, R. P.; Chandrasekaran, R.; Arnott, S.; Dea, I. C. M. Carbohydrate Res. 1988, 182, 1-17.
28. Millane, R. P.; Nzewi, E. U.; Arnott, S. In Frontiers in Carbohydrate Research 1: Food Applications; Millane, R. P.; BeMiller, J. N.; Chandrasekaran, R., Eds.; Elsevier: London, 1989, pp 104-131.
29. Rees, D. A.; Scott, W. E.; Williamson, F. B. Nature 1970, 227, 390-392.

Chapter 20

Molecular Mechanics and the Structure of Iduronate-Containing Carbohydrates

M. Ragazzi, A. Provasoli, and D. R. Ferro

Instituto di Chimica delle Macromolecole del CNR, via E. Bassini, 15—1-20133, Milano, Italy

Previously, our molecular mechanics studies showed that
the iduronate ring can adopt the ^2S_O (skew boat)
conformation. This finding rationalized n.m.r. results
for heparin (Hep), dermatan sulfate (DeS) and other
compounds in solution. In other work, simultaneous
minimization of both intra- and inter-chain energies
produced well-resolved structures of synthetic
polymers. Here, we apply this minimization method (in
the REFINE program) to investigate further the solid-
state structures of the above glycosaminoglycans. The
results favor the $^1C_4$ form of iduronate ring for the
isolated, two-fold polyanion helix of Hep, while all
three ring forms ($^4C_1$, ^2S_O, and $^1C_4$) are compatible
with the unit axial rise observed by Mitra et al. in
three DeS allomorphs. In particular, models based on
$^1C_4$ or the ^2S_O forms of iduronate rings agreed better
with observed structure factors from x-ray fiber
diffraction studies of the tetragonal form than did
models containing $^4C_1$ rings. The simultaneous presence
of all three ring shapes in crystals of DeS was shown
to be possible, even after including solvated sodium
ions. While we failed to obtain a clear preference
among these structures, we can warn against simplistic
interpretation of experimental data.

Molecular mechanics methods based on force-field representations of
the energy have been a powerful help in solving and refining a wide
variety of molecular structures and have become a common tool for
conformational analysis of simple and complex compounds. For the
past 15 years, these methods have been applied more to proteins than
other biomolecules. Apparently, the numerous single crystal studies
of proteins have been an important factor in this orientation of
modeling. The 3-dimensional structures of oligo- and poly-
saccharides often have complicated composition, sequences and
branching. Worse, these molecules have usually failed to provide
single crystals for diffraction studies and data, if any, arises from
powders or fibers. Analyses of such data are often much less
definitive. From the theoretical point of view, the stereo-
electronic effects in carbohydrates and the charged groups that are
often present make difficult the development of a precise potential

set. Therefore, analysis of even small carbohydrate molecules is difficult. Nonetheless, there has been some progress, mostly using simple methods, possibly supplemented for the exo-anomeric effect but based on fixed bond lengths and angles.

Iduronate-Containing Saccharides in Solution

While data from n.m.r. indicated that the $^1C_4$ prevails in solutions of the iduronate-containing saccharides, periodate data was interpreted to mean that the $^4C_1$ is correct. The possibility that other ring forms might be present was seldom considered. In one case, a "distortion" of the chair was invoked to explain the data from iduronate-containing compounds in solution that could not be rationalized with a mixture of the two forms (1-3).
     Cremer and Pople have described the puckering of sugar rings in terms of pseudo-angles (4). When computations are done in cartesian coordinates, it is possible to "drive" or step through values of the Cremer-Pople angles and minimize the energy for all other conformational variables. In a force-field study on the α-L-iduronic acid 2-sulfate monomer (5) using the computer program REFINE (6), we drove the ring shape, exploring the whole surface of the Cremer-Pople sphere in 5° increments. The energies, computed with modified potentials from the program, MM2 (7), showed the possible coexistence of the ^2S_O (skew-boat) form with the two chairs. This suggests a plasticity of the ring rather than a flexibility.
     This finding of three likely forms for iduronic acid (IdoA) was applied to the interpretation of vicinal proton coupling constants ($^3J_{HH}$) measured for dermatan sulfate (DeS), Heparin (Hep), and their oligomers in solution (8). Other ring forms could be disregarded in a first-order analysis. A thorough examination of these data on 28 compounds led us to propose qualitative rules regarding the relative populations of the IdoA ring forms, depending on sulfation and on the position along the saccharidic chain (9) (a more extended paper by Ferro, D. R. et al., Carbohydr. Res., is in press). In particular (see Table I) $^4C_1$ appears to be the least populated state in almost every case except for a significant 59% figure for the non-sulfated monomer and some 30-40% at the non-reducing end of most compounds.

Table I. Qualitative Conformer Populations of idoA and Sulfated idoA
          ($idoA_S$) in Different Compounds (+ Corresponds to about 20%)

	$^1C_4$	^2S_O	$^4C_1$
IdoA	++ (41)	–	+++ (59)
$IdoA_S$	++++(+) (90)	(+) (10)	–
$IdoA_S^-...$	+ (+)	+ (+)	+ (+)
$IdoA_S^-...$	+++ (+)	–	+ (+)
...-$IdoA_S^-...$	++ (+)	++ (+)	–
...-$IdoA_S^-...$	++ (+)	++ (+)	–
...-A*-$IdoA_S^-...$	+++ (+)	+	(+)
...-A*-$IdoA_S^-...$	++	+++	–
...-A*-$IdoA_S^-$A*-...	(+)	++++	(+)
Heparin	+++ (59)	++ (41)	–
Dermatan Sulfate	+++ (57)	++ (43)	–

A* represents 2-deoxy-2-sulfamido-α-D-glucose 3,6-disulfate

With one exception, the $^1C_4$ form prevails over the ^2S_O shape in all other sequences. The exception is for sulfated iduronate in peculiar sequences, i.e. flanking the trisulfated glucosamine (A**) found in heparin's active site for bonding to AT-III.

These conformer populations were confirmed by a number of n.o.e. measurements (the H2--H5 distance can act as a marker in identifying the ring form). The case of a hexasaccharide (compound 23 in Ref. 9) where the three IdoA residues yield different n.o.e. enhancements, is striking. Moreover, models computed with molecular mechanics have been put forward for the peculiar pentasaccharide of heparin. The models are in acceptable agreement with the observed vicinal coupling constants and n.o.e. values (10) (a full paper by Ragazzi, M. et al. Carbohydr. Res., is in press).

Solid-State Conformation of DeS and Hep

While the conformations of the IdoA ring in solution seem to be understood (11) including the presence of the ^2S_O shape (12-13), such is not the case for the solid state. Here we will discuss just two of the glycosaminoglycans, DeS and Hep. Circular Dichroism (CD) measurements (14) performed on both solutions and films of DeS suggested that the ring conformations do not change even though the phase changes. This conclusion is confirmed by ^{13}C chemical shifts (15) and is therefore at odds with the interpretation of the x-ray data for DeS given by Mitra et al., where only the $^4C_1$ form was considered (16). As for Hep, the $^1S_3$ and $^1C_4$ forms were proposed to occur together (17). Thus, we wanted to learn whether models of Hep and DeS structures containing IdoA in the ^2S_O form could be reconciled with the x-ray and other experimental data.

Methods of Calculation

Modeling studies are most useful when experimentally determined structures are of modest quality and suitable force-field potentials and modeling software are available. Although statistical methods such as Monte Carlo and molecular dynamics would be preferred in solution or other disordered states, we feel that energy minimization criteria are valid for static, ordered structures such as crystals. In the case of crystals, both intramolecular (conformational) and packing energies should be taken into account simultaneously. Such a total energy minimization method, with suitable crystallographic constraints, has been applied in different steps of the analysis of crystalline structures of three different synthetic polymers. Structures of these molecules, namely, isotactic trans-1,4-poly-penta-1,3-diene (ITPP), poly-pivalolactone (PPVL), and isotactic cis-1,4-poly(2-methyl-penta-1,3-diene)(PMPD), do not have troublesome features such as charged groups, counterions, and solvent molecules.
    In the case of ITPP, Ferro and Brückner (18) showed that unrestrained minimization of the total energy for a microcrystal corresponding to unstretched fibers yields a structure in very close agreement to the crystallographically refined one. This was in contrast to the earlier results with less accurate calculations. Furthermore, their calculations, which used slightly modified MM2 potentials and a modest restraint on the cartesian coordinates, provided a stereochemically acceptable model that reproduces the powder diffraction profile as accurately as the least-squares fitted model.

The need for simultaneous minimization of intra- and inter-molecular interactions was even clearer in the succeeding work on two crystalline forms of PPVL (Ferro, D. R. et al., <u>Macromolecules</u>, in press). An early model for the α structure, proposed by Perego et al. (<u>19</u>), showed a significant deviation from the planarity of the ester group. The values of its dihedral angles were criticized by Cornibert et al. (<u>20</u>) on the basis of energy calculations in the isolated helical chain. The powder diffraction results by Brückner (<u>21</u>) gave a model in substantial agreement with the model of Perego.

Our computations on the isolated chain seemed to confirm the planarity of the ester group, with a deviation of approximately $10°$ from planarity on the next dihedral angle. However, when this refined model was introduced in the α crystal, the conformation switched to the one derived from the x-ray experiments. Although Brückner's refined coordinates were not used for the modeling calculations, the minimum-energy model reproduced the powder diffraction profile very closely. It was also found that the chain conformation in the γ form is similar to the most stable conformation for the isolated chain. By changing the crystal packing, the calculation was able to reproduce the conformational change in the individual chains.

These results are quite promising for using computer models as starting points in crystallographic analyses of unknown polymeric structures. The powder diffraction pattern for the β form of PMPD yielded too few peaks to solve the structure by the Rietveld method. Energy calculations, based on observed cell dimensions and space group, have given a model that reproduces the powder profile with striking accuracy (Ferro, D. R.; Brückner, S. unpublished results).

We conclude that: a) all degrees of freedom should be unconstrained (including those related to the shape of the sugar rings); b) the packing forces can significantly influence the conformation even of a polymer main chain; and c) a suitable potential set (with all atoms), used in total energy minimization, can account for these effects. With these considerations in mind, we attempted to model the crystal structures of DeS and an isolated chain of Hep.

Results and Discussion

The procedures used for these two polymers were different because a full description of DeS had been published by Mitra et al. (<u>16</u>) but no three-dimensional structure was available for Hep. In both cases, helical symmetry was imposed so that each residue "sees" the same environment. The final DeS model included sodium ions and water as well as adjacent chains while only an isolated model of Hep was constructed.

<u>Isolated Helical Chain of Heparin</u>. We only considered models of Hep with IdoA in the $^1C_4$ and $^2S_0$ forms since the $^4C_1$ form had been ruled out by Nieduszynski et al. (<u>17</u>) on crystallographic grounds and the $^1S_3$ form is 4-5 kcal/mol less stable than other forms based on previous work (<u>5</u>). Our first tasks concerned the conformations of the two disaccharide repeating units which had been methylated at both ends to roughly simulate a continuing chain. In a shorthand notation, we use A_S for 2-deoxy-2-sulfamido-α-D-glucose and I_S for α-L-iduronic acid 2-sulfate. The glycosidic torsion angles have the usual definitions, $φ = (H1-C1-O-C4')$ and $ψ = (C1-O-C4'-H4')$.

The relaxed conformational energy map for A_S-I_S with the IdoA ring in the $^1C_4$ form is shown in Figure 1. The map for the ^2S_O form is quite similar. The I_S-A_S disaccharide has two distinct minima (marked I and II on Figure 2) whose relative stability depends on the I_S form and on the rotation angle about the C5-C6 bond in A_S. Conformers of minimal energy as determined by these and similar maps were used as starting points in building the 2_1 helical structural models of Hep. Minimum I, for both forms of the IdoA ring, yields values of the axial rise per disaccharide that are too low (h=0.7 nm) and was discarded. The other minimum leads to values of h somewhat higher than observed (0.84 nm)

As shown in Table II, the optimal value of the rotation angle per disaccharide unit around the helix axis, ϕ_h, for a free ^2S_O helix (structure a) is about 135°. When a two-fold axis is imposed (structure b), the energy (E) rises to 3.4 kcal/mol (with h=0.88 nm). If h is subsequently reduced to 0.84 nm (structure c), E increases to 4.8 kcal/mol. Continued search for the least energetic ^2S_O model that conforms to the symmetry requirements led to structure d, with E=3.2 kcal/mol.

The optimal value of ϕ_h for Hep helices based on the $^1C_4$ form is very close to 180°. Imposing an exact ϕ_h of 180° gave an h=0.86 nm and E of 1.4 kcal/mol (structure e). By shortening h to 0.84 nm, the value observed by diffraction experiment, E increases to 1.8 kcal/mol (structure f). Again, further searching located a lower minimum (structure g) that has an energy of 1.3 kcal/mol, 1.9 kcal lower than the best ^2S_O model. (If the Coulomb term is neglected, the energy is 3.6 kcal/mol lower.) Since these calculations only pertain to the isolated polyanion, the results may not be conclusive but they clearly favor the $^1C_4$ form rather than the ^2S_O form for the two-fold Hep helix.

Table II. Helical Models of Heparin, Corresponding to the $^1C_4$ and ^2S_O Forms of IdoA Rings, as Described in the Text (Observed Values are: ϕ_h = 180°, h = .84 nm)

| IdoA form | | ϕ_h (degrees) | h (nm) | E | $E^\#$ | A_S - I_S - A_S | | | |
				kcal/mol		ϕ	ψ	ϕ	ψ
(a)	(^2S_O)	134	.882	0.0*	(0.0)*	−41	−43	45	9
(b)		180	.878	3.4	(2.9)	−28	−26	60	17
(c)		180	.840	4.8	(3.7)	−29	−25	59	13
(d)		180	.840	3.2	(2.8)	−14	−24	48	9
(e)	($^1C_4$)	180	.882	1.4	(−0.4)	−39	−31	36	26
(f)		180	.840	1.8	(−0.3)	−41	−30	37	25
(g)		180	.840	1.3	(−0.8)	−38	−31	40	19

\# Values in parentheses do not include the electrostatic term (otherwise, ε=3)
* Taken as reference.

The Isolated Chain and the Crystalline Form of DeS. The molecular conformation given by Mitra et al. (16) for the tetragonal structure of DeS was used as the starting point for building all helical models of the isolated polyanion, assuming the tetrasaccharide I-N-I-N to be
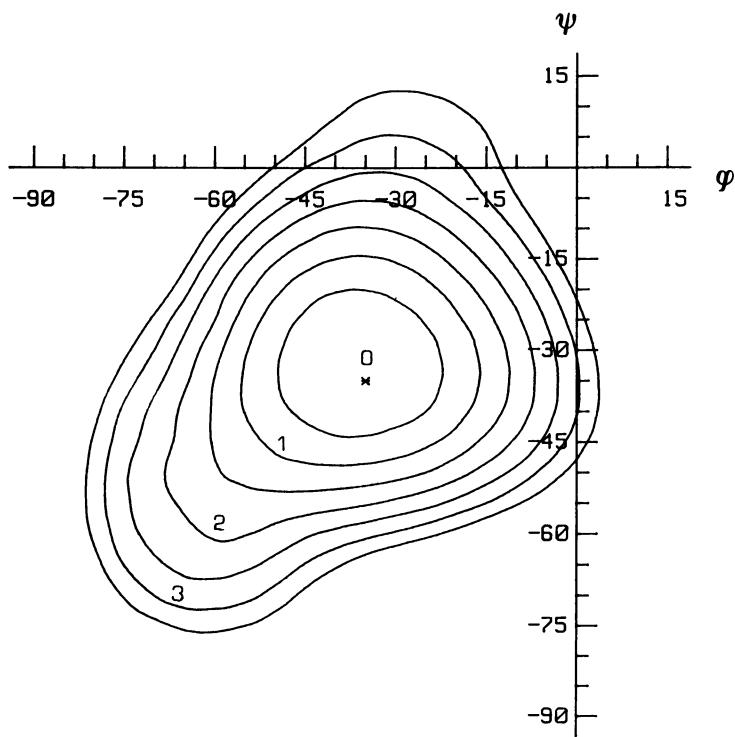
Figure 1. Map of the minimized energy of the dimer A_S-I_S as a function of the glycosidic torsion angles, with IdoA ring in $^1C_4$ form. Energy levels differ by .5 kcal mol^{-1}.
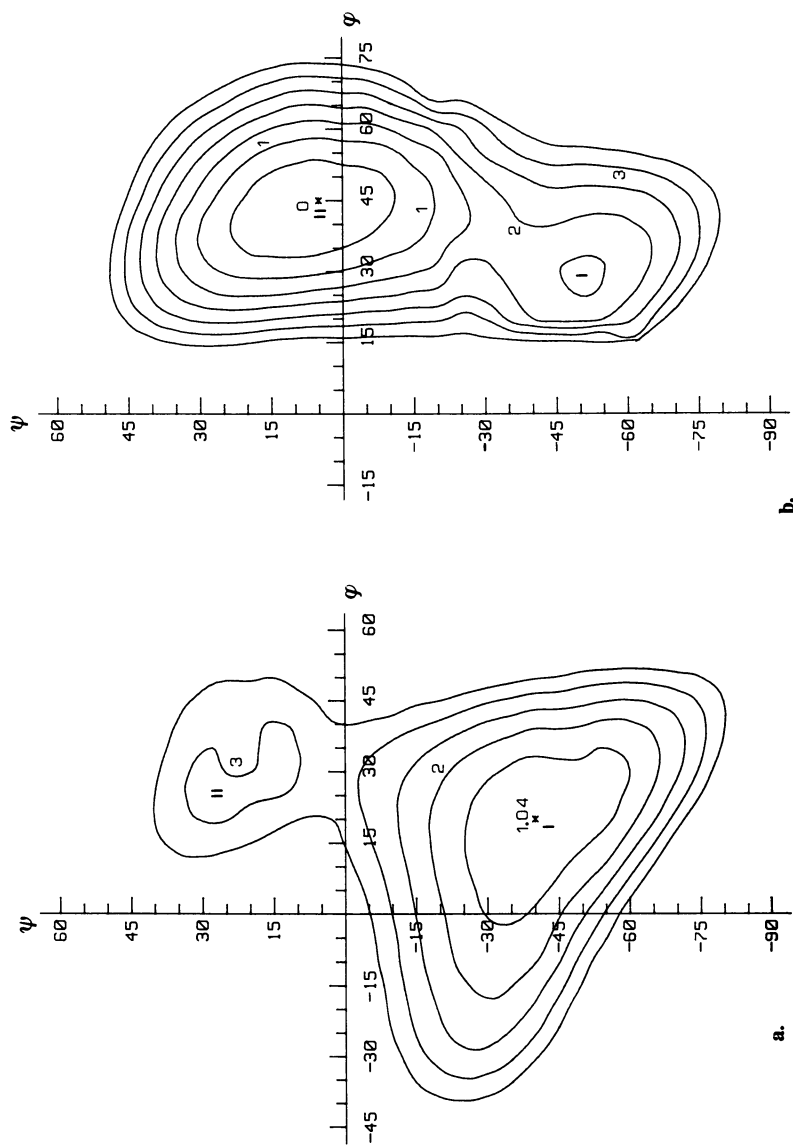
Figure 2. Maps of the minimized energy (levels differ by .5 kcal mol^{-1}) of the I_S-A_S dimer as a function of the glycosidic torsion angles with the IdoA ring in $^1C_4$ (a) and ^2S_O (b) form, respectively.

the asymmetric unit (I represents α-L-iduronic acid and N represents 2-deoxy-2-acetamido-β-D-galactose 4-sulphate). We first obtained the minimum for 8_3 models based on the $^4C_1$ form, a rise per disaccharide of 0.919 nm, and a ϕ_h of 135°. Further calculations explored the changes in energy as the helix conformation was changed. By changing the values of h and ϕ_h, we obtained models of the 8_5, 3_1, 3_2 and 2_1 helices for each form of the IdoA ring that yield the observed diffraction repeat distance. Their energies are reported in Table III. While the $^2S_0$ helices have energies nearly as low as the $^4C_1$ models, the $^1C_4$ models are not favored. The latter models lack the hydrogen bonds O5(N) to HO3(I) and O2(I) to HN2(N). Energies were also computed without the electrostatic contribution to gain a rough idea of the effect of solvent. These values have smaller range, but the $^2S_0$ form is still preferred.

Table III.   Energies (kcal mol^{-1}) per Disaccharide Unit of Various
Helical Structures of the Isolated Chain of DeS. Values in
Parentheses do not Include the Electrostatic Contribution
Otherwise, ε=3

Symmetry	8_3	8_5	3_1	3_2	2_1
ϕ_h (deg)	135	-135	120	-120	180
h (nm)	.919	.919	.941	.941	.939
$^1C_4$	4.33 (0.65)	5.43 (2.71)	5.19 (1.95)	5.74 (1.79)	4.69 (1.31)
$^2S_0$	3.02 (2.79)	0.00* (0.00)*	3.14 (2.41)	0.96 (1.15)	0.05 (0.01)
$^4C_1$	2.97 (3.38)	0.36 (0.38)	2.67 (2.56)	0.80 (0.71)	2.25 (1.84)

* Taken as reference

Further work dealt with the tetragonal form. A 4_3 helix was assumed based on the crystallographic result (16). Agreement with the experimental data was better for the $^2S_0$ and $^1C_4$ polyanions than for the $^4C_1$ models (Ferro, D. R. et al., 3rd International Workshop "Recent Developments in Industrial Polysaccharides", Trieste (Italy) 1988,; Crescenzi, V. ed., in press).
    When modeling the complete structure, many problems arose. Among them were the proper treatment of water molecules and sodium ions, two of which were assigned by Mitra to special crystallographic positions. The asymmetric unit in the energy minimization contains four saccharide units (bearing a total of four charges), four sodium ions (three plus two halves) and 15.5 water molecules. Another problem was the choice of the interaction potentials for sodium ions, derived from Potenzone and Hopfinger (22), and for water and hydroxyl groups, fitted from ab initio computations of Matsuoka et al. (23) (those potentials reproduced the oxygen-oxygen distances of ice).

The asymmetric unit mentioned above, with the coordinates given by Mitra et al. (based on $^4C_1$ iduronate rings) was then surrounded, according to the crystallographic symmetry, by all the atoms up to a cutoff distance of 0.6 nm. All of the partial residues arising from the above selection were completed, and the sodium atoms needed to ensure neutrality were added. The resulting microcrystal contains about 5200 atoms, but due to symmetry, only 215 atoms have independent coordinates for a total of 645 variables in the minimization (see Figure 3).

A preliminary energy refinement was used to adjust the polar hydrogen atoms and lone pairs of electrons. Then the complete structure was optimized with only constraints imposed by the crystallographic symmetry and unit cell geometry. Keeping these constraints, we then drove the conformation of the IdoA rings from the $^4C_1$ toward the ^2S_O conformer. Care was taken to ensure continuity (and reversibility) of this process by changing the Cremer-Pople puckering angles in steps as small as 5 to 10° and by minimizing the energy with respect to all other degrees of freedom. An increase in the energy of 10 kcal/mol was found for an intermediate ring structure, in agreement with the results obtained for the monomeric iduronate (5). A structure with minimal energy for the DeS models based on the ^2S_O form was obtained, and the drive continued toward the $^1C_4$ models. Each of the structures was minimized extensively to obtain final structures. A part of the central helix for each model is shown in Figure 4.

Table IV summarizes a few features of the models. Their total energies are within about 4 kcal/mol of each other, with the ^2S_O form being least likely. However, the individual components of the total energy vary widely, due to different patterns of intra- and inter- chain hydrogen bonds that result from competition with water bridges. In the $^1C_4$ models, the polyanion has no intrachain hydrogen bonds.

Table IV. Energies of Tetragonal (4_3) Models Computed for the
        Complex DeS + 4 Na$^+$ + 15.5 H_2O (see Figures 3 and 4)

IdoA ring form	$^4C_1$	^2S_O	$^1C_4$
Total energy	-177.4	-173.2	-177.2
- bonded	67.0	61.9	61.5
- van der Waals	-77.8	-68.7	-74.9
- el. static (ε=3.)	-166.7	-166.3	-163.8
water	-183.1	-172.5	-181.4
- average per molec.	-11.8	-11.1	-11.7

A dominant role seems to be played by constraining the Na$^+$ ions to special crystallographic positions. In all cases, the water molecules near them have high energies (as high as -8 kcal/mol). The unrestricted water molecules (those nearer the Na$^+$ ions that are two sugar units down the chain) have the lowest energy (as low as -15 kcal/mol). Thus, we feel that the assumption of all of the characteristics (except ring shape) of the model proposed by Mitra et al. has been detrimental to our attempts to determine the correct ring shape. Further minimizations, in order to escape from possibly false minima, and simultaneous crystallographic refinement are in progress.

diagonal 2-axis

-c/8

O

effective
cut-off ∼8Å

P4₃2₁2

a=b=12.67Å
c=73.53Å
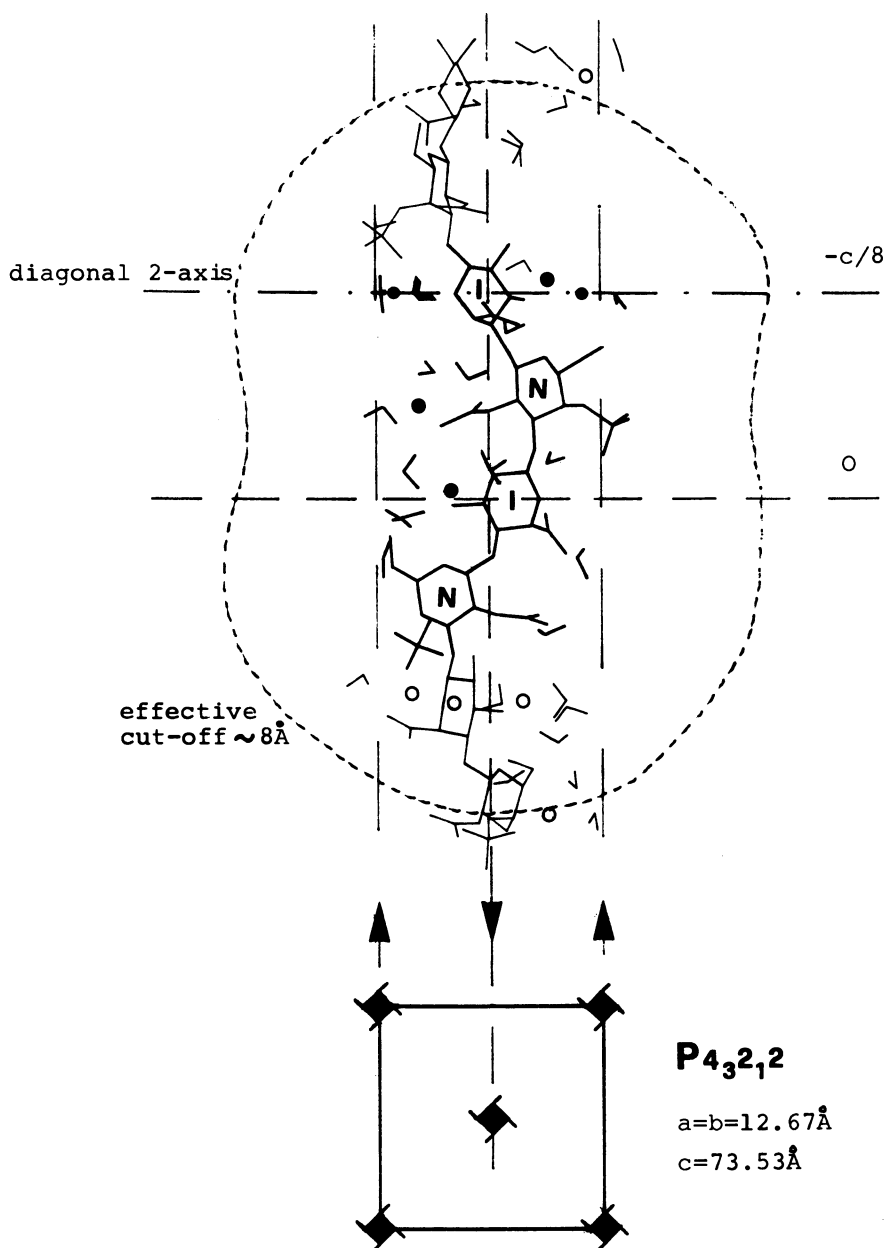
Figure 3.   Schematic representation of the microcrystal considered in
the computations; the asymmetric unit (4 DeS residues, 5 Na[+] ions and
16 water molecules, evidenced by heavy lines) is reproduced according
to the crystallographic symmetry: note the two ions and the water
molecule in special position (on the diagonal 2-axis at -c/8).
Hydrogen atoms in the polymer and lone-pairs are omitted in the
drawing.

a.

b.
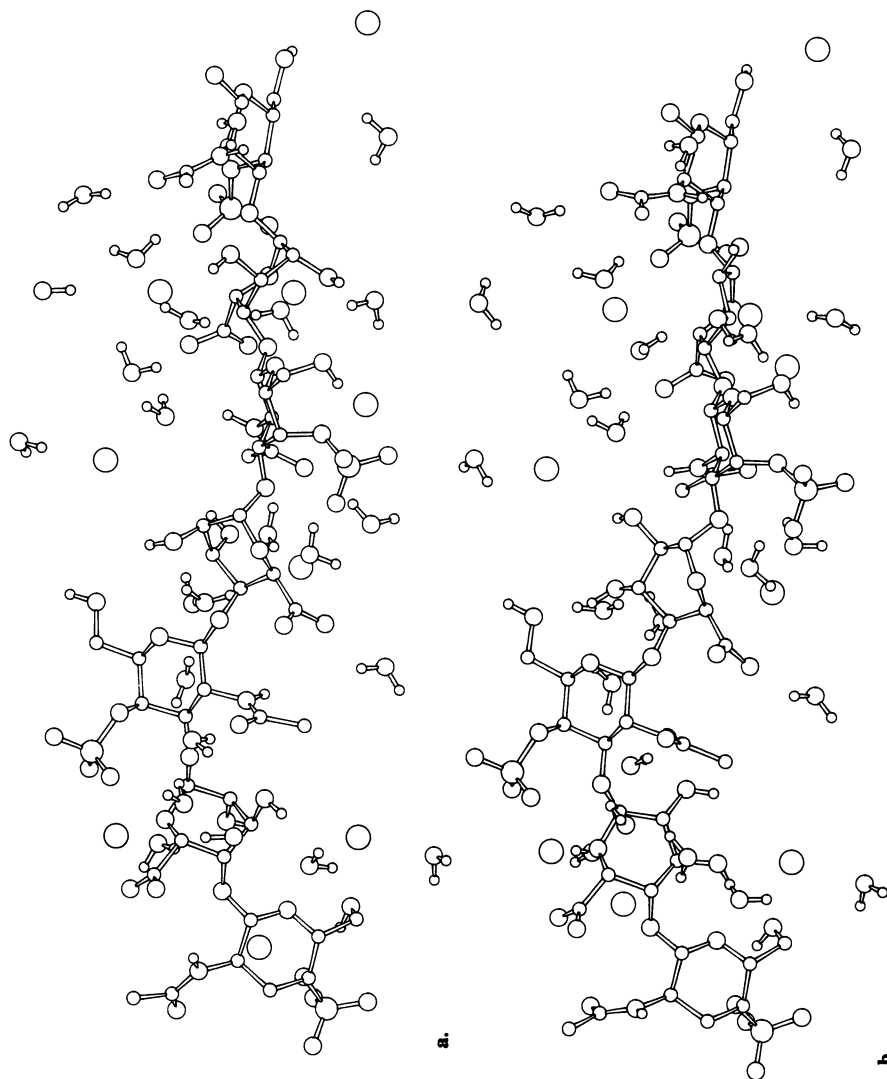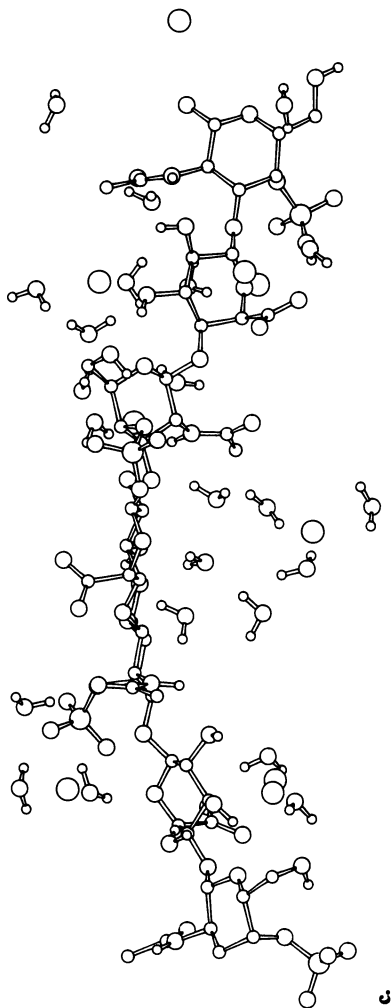
Figure 4. Portion of the central helix in the energy-minimized structures of the crystalline DeS with the IdoA ring in $^1C_4$ (a), $^2S_0$ (b) and $^4C_1$ (c) forms, respectively.

Conclusions

Our analysis warns against simplistic interpretations of the data, but fails, at this stage, to give a clear solution to these structures. A more complete set of crystallographic data should be used with the computational methods presented here for further understanding.

Acknowledgments

Molecular structures of Figure 4 were drawn by using the SCHAKAL program (24), as adapted (by M.R.) for the local HP-9000 computer.

Literature cited

1. Rees, D. A.; Morris, E. R.; Stoddart, J. F.; Stevens, E. S. Nature (London) 1985, 317, 480.
2. Auge, J.; David, S. Tetrahedron 1984, 40, 2101-2106.
3. Snyder, J. R.; Serianni, A. J. Org. Chem. 1986, 51, 2694-2702.
4. Cremer, D.; Pople, J. A. J. Am. Chem. Soc. 1975, 97, 1354-1358.
5. Ragazzi, M.; Ferro, D. R.; Provasoli, A. J. Comput. Chem. 1986, 7, 105-112.
6. Ragazzi, M.; Ferro, D. R. "REFINE", Program and user's manual available on request.
7. Allinger, N. L.; Yuh, Y. H. Q.C.P.E. 1980, 12, No. 395.
8. Ferro, D. R.; Provasoli, A.; Ragazzi, M.; Torri, G.; Casu, B.; Gatti, G.; Jacquinet, J.-C.; Sinay, P.; Petitou, M.; Choay, J. J. Am. Chem. Soc. 1986, 108, 6773-6778.
9. Casu, B.; Petitou, M.; Provasoli, A.; Sinay, P. Trends Biol. Sci. 1988, 13, 221-225.
10. Ragazzi, M.; Ferro, D. R.; Perly, B.; Torri, G.; Casu, B.; Sinay, P.; Petitou, M.; Choay, J. Carbohydr. Res. 1987, 165, C1-C5.
11. Casu, B.; Choay, J.; Ferro, D. R.; Gatti, G.; Jacquinet, J.-C.; Petitou, M.; Provasoli, A.; Ragazzi, M.; Sinay, P.; Torri, G. Nature (London) 1986, 322, 215-216.
12. Sanderson, P. N.; Huckerby, T. N.; Nieduszynski, I. A. Biochem. J. 1987, 243, 175-181.
13. Van Boeckel, C. A. A.; van Aelst, S. F.; Wagenaars, G. N.; Mellema, J.- R. Recl. Trav. Chim. Pays-Bas 1987, 106, 19-29.
14. Cziner, D. G.; Stevens, E. S.; Morris, E. R.; Rees, D. A. J. Am. Chem. Soc. 1986, 108, 3790-3795.
15. Winter, W. T.; Taylor, M. G.; Stevens, E. S.; Morris, E. R.; Rees, D. A. Biochem. Biophys. Res. Commun. 1986, 137, 87-93.
16. Mitra, A. K.; Arnott, S.; Atkins, E. D. T.; Isaac, D. H. J. Mol. Biol. 1983, 169, 873-901.
17. Nieduszynski, I. A.; Gardner, K. H.; Atkins, E. D. T. ACS Symp. Ser. 1977, 48, 73-90.
18. Ferro, D. R.; Brückner, S. Macromolecules 1989, 22 2359-2363.
19. Perego, G.; Melis, A.; Cesari, M. Makromol. Chem. 1972, 157, 269-278.
20. Cornibert, J.; Hien, N. V.; Brisse, F.; Marchessault, R. H. Can. J. Chem. 1974, 52, 3742-3747.
21. Bruckner, S.; Meille, S.V.; Porzio, W. Polymer 1988, 29, 1586-1589.
22. Potenzone, R. Jr.; Hopfinger, A. J. Polymer J. 1978, 10, 181-199.
23. Matsuoka, O.; Clementi, E.; Yoshimine, M. J. Phys. Chem. 1976, 64, 1351-1361.
24. Keller, E. Chem. Unserer Zeit 1980, 14, 56-60.

Chapter 21

Modeling in Crystal Structure Analysis of Polysaccharides

A. Sarko, C.-H. Chen, B. J. Hardy, and F. Tanaka[1]

Department of Chemistry and Cellulose Research Institute, College of Environmental Science and Forestry, State University of New York, Syracuse, NY 13210

The utility of modeling in diffraction analysis of polysaccharides is demonstrated by a 20-year accumulated record of successful structure determinations. However, as the results obtained with some polysaccharides—notably celluloses—have shown, some structural details can remain unresolved. The ambiguities may be traced to varying modeling methods, particularly when applied to differing data sets. When structurally different but superficially similar models are possible, different methods may produce results that are inconsistent, not allowing the identification of the most probable structure. Nonetheless, modeling methods are steadily extending the capabilities of structure analysis of polysaccharides, both in crystalline and other states. Especially, simulations of molecular dynamics may offer the most promise in dealing with the amorphous and gel states, as well as with crystalline phase transformations. Of the latter, mercerization of cellulose is of particular interest.

Diffraction analysis—whether employing x-rays, electrons, or neutrons—is the method of choice for obtaining structural information on crystalline substances. The application of the well understood principles and methods of diffraction analysis to single crystals of sufficient size and perfection can lead to a detailed determination of the crystal structure, without recourse to any auxiliary methodology. Hundreds of mono- and oligosaccharide molecules have been characterized by these means *(1)*, yielding not only an increased understanding of their structures in the solid state, but also a data base useful for extrapolation to other states and molecular interactions.

The same is not true of polysaccharides. When crystalline, these substances are usually *polycrystalline*, that is, their crystalline phase is composed of a multitude of individual, sub-microscopic crystallites that are dispersed in an amorphous phase. The latter may constitute the major phase. In contrast to

[1]Current address: Wood Research Institute, Kyoto University, Uji, Kyoto 611, Japan

carbohydrate single crystals, the diffraction data obtainable from such materials are not sufficient to permit crystal structure determination based on the data alone. Recourse must be sought in trial-and-error procedures where diffraction intensities are calculated for various models for comparison with the observed intensities—or, in other words, to *modeling*.

It is thus not surprising that until computer-based molecular modeling became feasible, the detailed crystal structure analysis of polysaccharides was generally not possible. It has been little more than 20 years since the first, computer-assisted, and a reasonably detailed structure analysis of a polysaccharide was completed (2). Importantly, it was based on *exhaustive* model analysis, thus leaving less doubt about the correctness of the general features of the structure. Useful principles, not only from the structural point of view but also pertaining to methodology, were established that have helped formulate further modeling procedures. As computer power increased and our molecular modeling sophistication grew, such modeling has become an inseparable component of the diffraction analysis of polysaccharides. Now the number of apparently successful structure solutions of polysaccharides exceeds 100, counting all polymorphs, variants, derivatives, and complexes. The use of this methodology has also extended into areas of structure analysis where diffraction methods yield almost no information, for example, the amorphous state and the gel state, and the dynamics of molecules.

How modeling has been useful in the crystal structure analysis of polysaccharides—and how it could lead to a better understanding of other condensed-phase states—can be illustrated with structural work done on cellulose. It is one of the world's most important and widely used raw materials whose structure, properties, derivatives, and transformations remain under continuous study. Some of the results, problems, and indications of future directions resulting from the study of its crystalline structure—and the attendant roles for molecular modeling—are briefly described in the following.

Modeling in Fiber Crystallography

Cellulose Structures. Native cellulose—or *cellulose I*—is usually found in a fibrous, well oriented, highly crystalline form, thus being a suitable material for diffraction analysis. A typical x-ray fiber diagram of cellulose I of ramie is shown in Fig. 1. In our hands, 50 intensities (with 115 diffraction planes contributing) could be obtained from this diagram, which has allowed a detailed, modeling-assisted structure analysis to be performed (3). The importance of modeling in this analysis may be appreciated by comparing the availability of 50 observations (intensities) to the need to locate 44 non-hydrogen atoms—represented by 132 spatial parameters—in the unit cell.

The major structural characteristics of cellulose I, determined in this and other studies, include extended chains stabilized along their lengths by two intramolecular hydrogen bonds per glucose residue ($O(3)\cdots O(5)'$ and $O(6)\cdots O(2)'$), the arrangement of the chains into sheets stabilized by one intermolecular hydrogen bond per residue ($O(3)\cdots O(6)$), and the packing of the sheets into a three-dimensional structure marked by a *parallel-chain* arrangement. These structural characteristics are illustrated in Fig. 2.

The question of *parallel vs. antiparallel* chain packing in cellulose I has been a controversial one practically since the first cellulose structure was proposed. A consensus appears to be forming, however, based on both diffraction analysis and other experimental evidence, that *one* of two possible
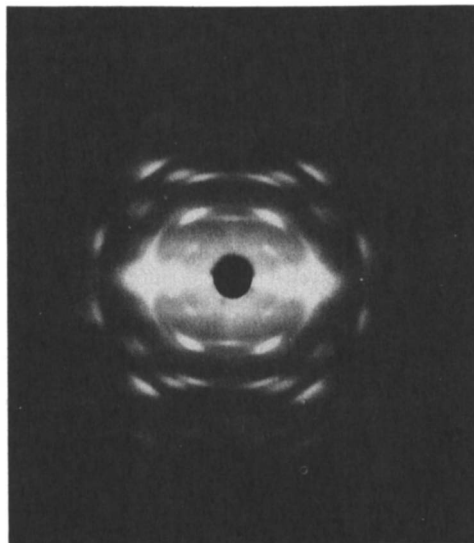
Figure 1. X-ray diffraction pattern of ramie cellulose I.
Fiber axis is vertical.
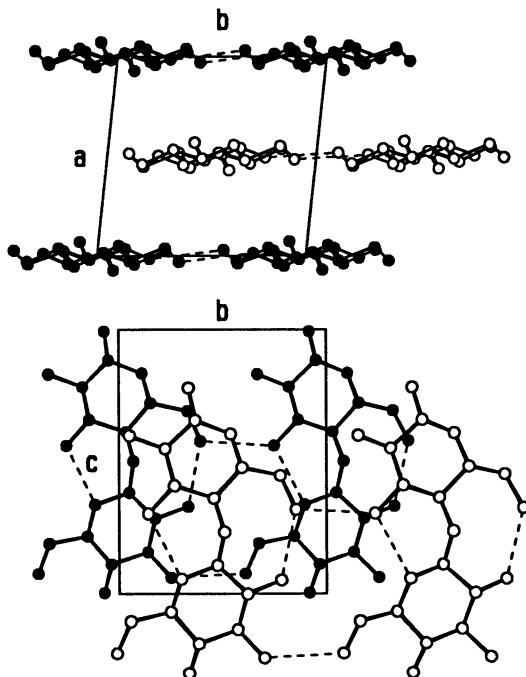
Figure 2. Two projections of the ramie cellulose I crystal structure:
on a-b plane (top), on b-c plane (bottom). Hydrogen atoms are
omitted; hydrogen bonds are shown by dashed lines.

parallel-chain structures is indeed correct. The question of chain polarity is an important one for at least two reasons: understanding biosynthesis of cellulose where the mechanisms of the forming of the chains and their crystallization are at issue *(4)*, and the mechanism of alkali mercerization of cellulose I. Mercerization results in *cellulose II*, which fiber diffraction studies have shown to be based on an *antiparallel-chain* structure *(5)*. Of further interest is the fact that cellulose can crystallize in four polymorphic forms which fall into parallel- and antiparallel-chain families, with well-defined pathways marking their transitions *(6)*. These differences extend to cellulose derivatives (*e.g.*, acetates *(7,8)*) and complexes (*e.g.*, with alkali), where packings with both chain polarities have been observed *(5)*. Despite the knowledge of these general structural features, some nagging uncertainties remain. For example, which of the two very similar parallel-chain structures—the *up-* or the *down-chain*—is the correct structure for cellulose I? How does *Valonia* cellulose I differ in structure from other native celluloses? When during mercerization does the conversion of a parallel-chain structure to an antiparallel one occur? Although diffraction analysis can be expected to provide the data that are necessary to answer these questions, the interpretation of these data is proving to be difficult.

Can Modeling Help Solve The Remaining Problems? In fiber x-ray diffraction analysis the convergence to the most probable crystal structure generally proceeds via the minimization of x-ray residuals, such as:

$$R = \frac{\Sigma \, ||F_o| - |F_c||}{\Sigma \, |F_o|} \qquad\qquad R'' = \left\{ \frac{\Sigma w \, ||F_o| - |F_c||^2}{\Sigma w \, |F_o|^2} \right\}^{\frac{1}{2}} \qquad (1)$$

where $|F_o|$ and $|F_c|$ are the observed and calculated *structure factor amplitudes*, respectively, and w's are weights. (The $|F_o|$ are obtained from the measured intensities of the x-ray reflections, whereas the $|F_c|$ are calculated from the atomic positions of the crystal structure, thus being a function of the model under refinement.) Because of the aforementioned polycrystalline characteristics of fibrous polysaccharides and the attendant paucity of diffraction data, these functions are usually modulated by the inclusion of stereochemical terms, for example:

$$\Phi = f \, R'' + (1-f) \, E \qquad (2)$$

where f is a fractional weight and E may be given by:

$$E = \Sigma w_i (a_i - a_i^o)^2 + W \Sigma w_{ij}(d_{ij} - d_{ij}^o)^2 \qquad (3)$$

In eq. 3, the first term measures the deviations in the bond lengths, bond angles, and conformation angles (a_i) from the corresponding accepted norms (a_i^o), whereas the second term measures the deviations of *nonbonded* interatomic distances (d_{ij}) from the corresponding equilibrium values (d_{ij}^o). (The terms and summations are appropriately weighted by w_i, w_{ij}, and W). Many variations in the stereochemical function for E are possible. For example, the bonded terms in eq. 3 may be replaced by different bond length, bond angle, and torsion angle distortion terms (and cross-terms) as are available from such popular modeling programs as MM2 *(9)*, AMBER *(10)*, CHARMm *(11)*, etc. Likewise, the *nonbonded* term may be replaced by, for example, Lennard-Jones, Buckingham, Kitaigorodskii, or other energy

functions, amplified as desired by electrostatic energy and hydrogen-bond energy terms. It is thus clear that the course of the structure refinement—via the minimization of the function Φ (or similar)—could be greatly influenced by the choice of the functions for E, as well as the weight f. When the x-ray data are well resolved and plentiful, it is very probable that the choice of almost any stereochemical function would not influence the outcome of the structure determination. On the other hand, in cases where the reverse is true or, as in the case of cellulose I where the parallel and antiparallel structures are similar on atomic scale, unpredictable problems caused by the choice of stereochemical functions may be expected to arise.

Another important function of the stereochemical modeling is the *prediction* of suitable initial models for refinement by the function of eq. 2. Although this goal may not be completely attainable, modeling can be expected to predict the crystal structures correctly, on the assumption that the latter lie in energy minima. The problem here is two-fold: 1) the most probable chain conformation model must be predicted first, 2) to be followed by the prediction of its three-dimensional packing within the crystal lattice. Typical molecular mechanics functions have been optimized with small-molecule crystal structures, and such optimization is generally attempted with a variety of molecular structures in order to generate functions that are as widely applicable as possible. These functions may be used to predict the structures or conformations of *isolated* polysaccharide molecules, but there is as yet no assurance that they could accurately predict both the conformations and the crystal structures of polysaccharides. Their applicability and validity could be assessed by attempting to predict the known crystal structures of carbohydrates, but this is a formidable and computing-intensive task in view of the possibility of many local energy minima.

An example of the differences in predicting behavior by different methods is shown in Fig. 3. The conformation of an isolated cello-octaose molecule was predicted using both the program DISCOVER *(12)* into which the MM2(85) potential function set *(9b)* was incorporated, and the modeling-only component of the PS79 program which utilizes the energy function of eq. 3 *(13)*. The structure predicted with MM2 functions was not constrained in length, whereas the PS79 structure was constrained to a length of 41.46Å, the four-fold multiple of the 10.34Å fiber repeat of cellulose I. Conformational refinement as part of diffraction analysis is normally done in the latter fashion to assure that the molecule fits properly on the crystal lattice. However, as is typical of the use of PS79, refinement started with two independent halves of the molecule which met in the middle, but without constraining the molecule to any particular conformation. It is clear from a comparison of the two predicted structures that the MM2/DISCOVER conformation twists ~180° along its length, while PS79 maintains a flat, ribbon-like conformation that very nearly conforms to two-fold helix symmetry. The lengths of both conformations are virtually identical. The PS79-predicted conformation is very close to the structure based on the crystallographic data, shown for comparison on the right in Fig. 3. In additional tests using the modeling-only part of LALS, another popular refinement method used in fiber diffraction analysis *(14)* and one that is based on a completely different methodology compared to PS79, the conformation predicted by it for an isolated chain was very similar to that from PS79.

The differences in these structures are further delineated in terms of φ,ψ-angles, shown in Table I. The φ,ψ-angles of individual residues are remarkably consistent along each model, resulting in regular conformations.

The MM2 model resides very near the minimum 2 in the cellobiose energy map (cf. Fig. 9). (Among others, the crystal structure of methyl cellobioside-methanol complex is found in that minimum (15)). On the other hand, the PS79 model resides on the shallow saddle point between minima 2 and 3.

The reasons for the different outcomes in these conformational predictions are not clear at this time, but are being investigated further. Hydrogen bond energies were excluded in all predictions, thus strong attractive forces were not a factor in the observed differences.

Table I. The ϕ,ψ-angles (deg.) of the conformations shown in Fig. 3

Residues	MM2		PS79		Crystal	
	ϕ	ψ	ϕ	ψ	ϕ	ψ
1−2	-7.0	-36.7	24.4	-26.3	29.5	-26.5
2−3	-2.7	-36.6	24.1	-26.2	25.1	-29.4
3−4	-3.8	-35.8	25.9	-27.0		
4−5	-4.4	-35.5	27.9	-24.1		
5−6	-4.2	-35.6	24.8	-25.6		
6−7	-3.4	-36.4	25.1	-26.6		
7−8	-3.9	-36.4	25.3	-27.0		

Examples of the influence of stereochemical modeling on three-dimensional packing analysis are shown in Fig. 4. The unit cell structure based on one of the probable chain models of cellulose I was obtained by three refinement methods: 1) based on x-ray data only (minimizing R'', eq. 1), 2) pure stereochemical prediction (minimizing E, eq. 3), and 3) combination of the two by minimizing Φ (eq. 2), with $f = 0.95$. The values of the residuals for the three structures are shown in Table II.

As the above results show, the gross features of the cellulose I crystal structure predicted by various methods do not differ appreciably, but the accompanying deviations in the R''-factors are significant. When these predictions are used to assess, for example, whether the cellulose I crystal structure is based on parallel- or antiparallel-chains, the range in the R''-factors seen for the parallel models (cf. Table II) is comparable to that between the two different polarity models. As shown in Fig. 5, the most probable parallel- and antiparallel-chain structures of cellulose I, refined by minimizing the function Φ, differ in R''-factors by approximately the same extent as the three predictions for the parallel model shown in Fig. 4 and Table II.

The variations resulting from the above methods are not complete without mentioning one added complication—the differences in x-ray intensity data sets obtained by different investigators. As shown in Fig. 1, a typical x-ray fiber diagram contains many overlapping diffraction maxima whose intensities should be resolved into as many individual components as is possible, for increased precision of analysis. Because of varying methods used by different investigators to measure and resolve intensities, including the indexing of difficult-to-resolve reflections, considerable variability can be
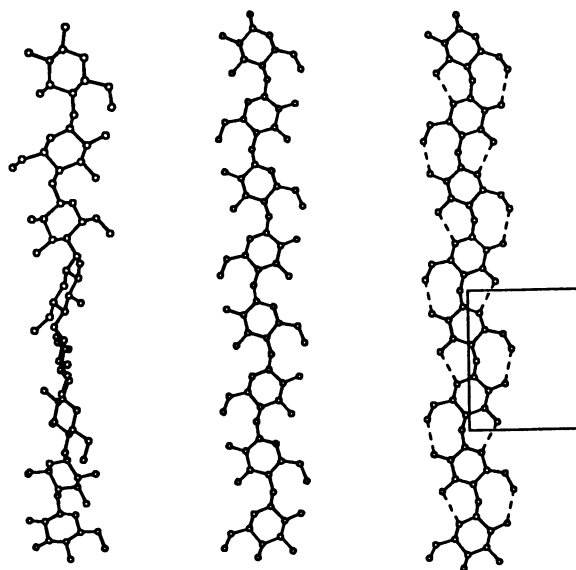
Figure 3. Isolated chain conformations of cellulose predicted by: MM2(85) (left) and PS79 (middle). The conformation on the right is that of the crystal structure of cellulose I *(3)*. Hydrogen bonds are shown by dashed lines.
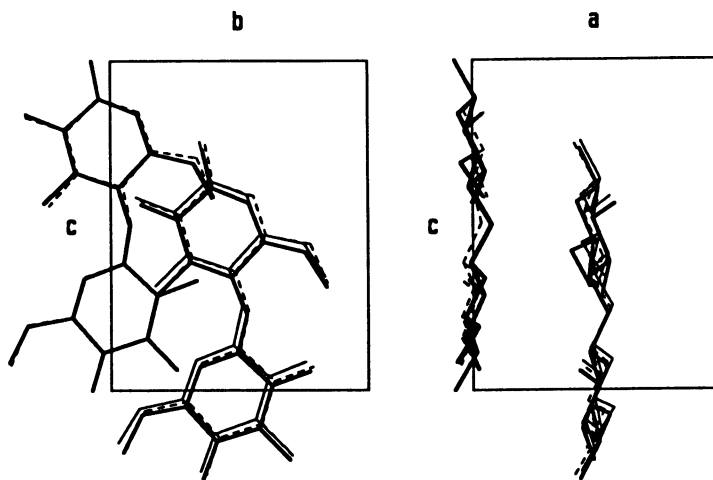
Figure 4. Two projections of the cellulose I structure predicted by three methods: based on *R"* (dashed line); based on *E* (bold line); based on Φ (thin line).
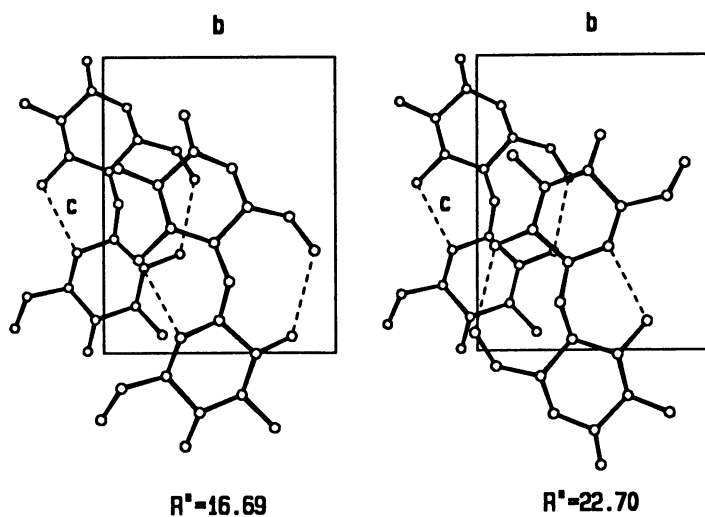
Figure 5. Most probable *parallel* (left) and *antiparallel* (right)
cellulose I models.

expected on these accounts, particularly in the weaker reflections. At least five different data sets for cellulose I have been published, collected from *Valonia* and ramie celluloses, using various methods *(3,16-19)*. In a previously published analysis of some of these data sets a range of variability was indeed observed, although the preference for parallel over antiparallel packing was generally upheld *(20)*. A thorough evaluation of the available cellulose I data sets is currently in progress by us.

Table II. The x-ray residuals for the three packing models shown in Fig. 4

Model	R	R"
Φ refinement	0.179	0.167
X-ray refinement	0.174	0.155
Stereochemical refinement	0.277	0.233

Finally, the question of the ability of the modeling methods to predict the crystal lattice (*i.e.*, the *unit cell*) from the conformation of the chain should be addressed, despite the expected computational difficulties. Based on previous work, in which the prediction of the unit cells of all four cellulose polymorphs in both parallel and antiparallel chain packing polarities was attempted *(21)*, it appears that at least some crystallographic data are necessary for such prediction. As shown in Fig. 6, the locations of the energy minima predicted for both parallel and antiparallel packing polarities coincided reasonably well with the known unit cells. The fiber repeat (*i.e.*, the c axis of the unit cell) was held constant at 10.34Å in these predictions, while varying the other parameters (*a, b, γ*) in a manner consistent with the known crystalline density of cellulose. It should be noted that this modeling was done in "relaxed" manner in today's terminology (except for the constant 10.34Å repeat), using the function given in eq. 3. The questions concerning further refinement of this process as well as the impact of reducing the amount of crystallographic constraints imposed on it, still remain unanswered.

Modeling of Molecular Dynamics of Cellulose

When crystalline cellulose I is treated with aqueous alkali solutions of sufficient strength, a process known as *mercerization* takes place. As a result of it, cellulose I is converted to cellulose II, the most stable of the four crystalline cellulose polymorphs. The conversion proceeds in the solid state, without apparent destruction or change in the fibrous morphology of the cellulose. As our diffraction analysis indicates, however, it is accompanied by a reversal of the chain packing polarity—from the parallel-chain cellulose I to the antiparallel-chain cellulose II. Crystallographic studies of an intentionally slowed-down process of mercerization have shown that the cellulose I to cellulose II conversion takes place in steps, involving as many as five crystalline alkali-cellulose or hydrated cellulose complexes *(22-25)*. Each step in this sequence is a crystal-to-crystal phase transformation, occurring without an intervening stage of decrystallization. At the same time, the
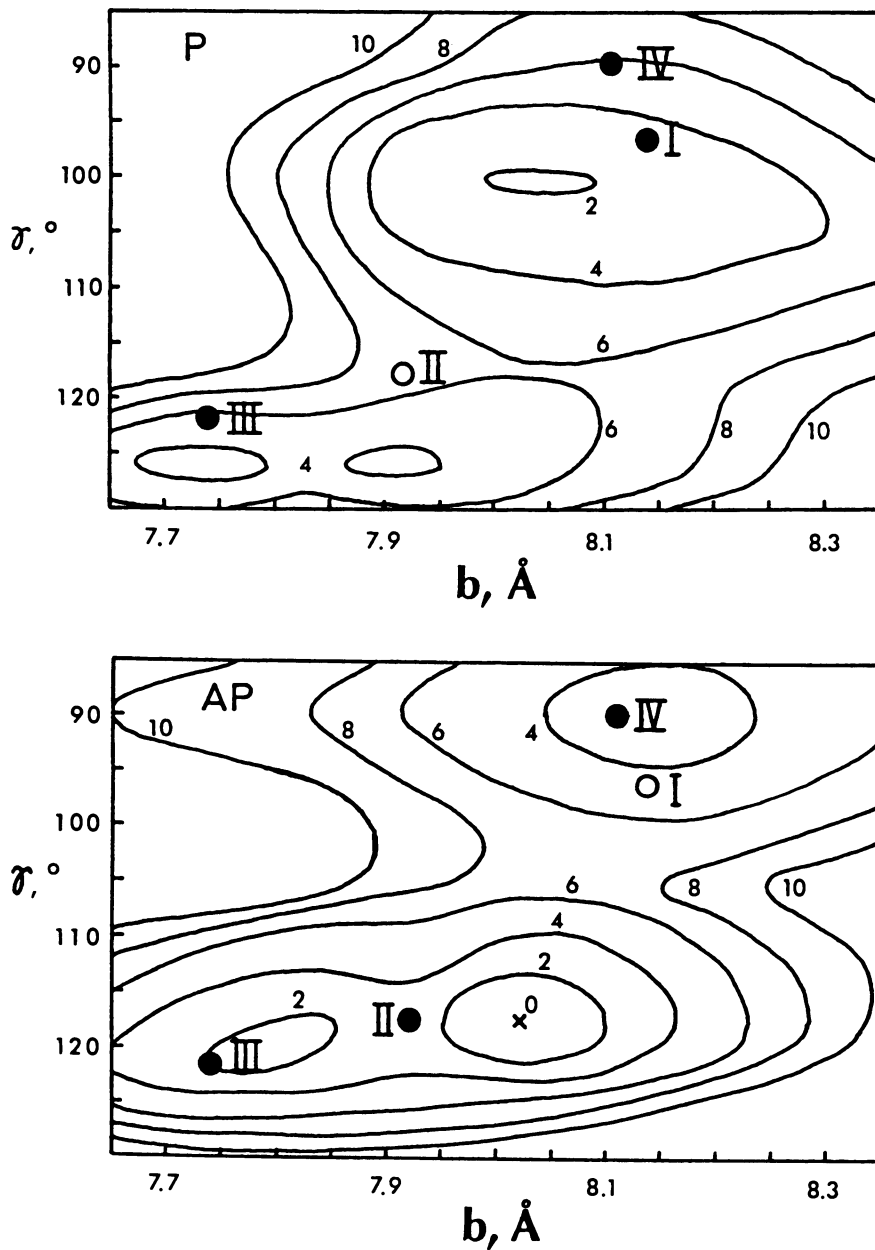
Figure 6. Prediction of unit cells for the four cellulose polymorphs in *parallel* (top) and *antiparallel* (bottom) polarities. Reproduced with permission from ref. 22 (Copyright John Wiley & Sons, 1984).

amounts of alkali and water present in these structures suggest that the cellulose chains have considerable mobility during the transformations.

Based on the known morphology of the cellulose fiber and the crystalline characteristics of the alkali-cellulose complexes, a mechanism of mercerization has been proposed which accounts for the accompanying change in the chain packing polarity (23-25). As shown schematically in Fig. 7, alkali first penetrates the amorphous regions which separate the crystallites of cellulose I, swelling and converting them to the first alkali-complex ("Na-cellulose I")—which even early during conversion appears to be substantially antiparallel in structure (26). This new phase then slowly "peels" chains from the surfaces of adjoining unconverted crystallites until the transformation is completed. In the process, the incompletely antiparallel Na-cellulose I phase anneals to a completely antiparallel-chain structure. (Subsequent steps, which involve more alkali followed by washing and drying, ultimately result in a conversion to the final product—cellulose II).

For these steps to proceed as outlined above, individual cellulose chains must possess sufficient segmental mobility in order to detach from one crystalline phase and to transfer to another one. To test the extent of this mobility, we turned to molecular dynamics—a technique for simulating molecular motion over time. In such modeling, a molecule at rest in a low energy state is imparted sufficient energy to set its atoms in random motion. Through torsional rotations, bond angle bending, and bond stretching vibrations, the molecule will undergo continual changes in shape. These changes can be followed quantitatively by numeric integration of Newton's equation of motion. A record of these changes—a trajectory—thus becomes available, which, if the time span is sufficiently long, allows the calculation of statistical and other properties of the molecule. As in static molecular modeling, molecular dynamics simulations can be expected to be significantly influenced by the potential functions used.

Our initial, small models of an isolated cellulose chain ranged from the dimer (cellobiose) to the octamer. The dynamics of these fragments have thus far been simulated only in vacuum, using different potential energy functions such as those of MM2(85) (9) and AMBER (10), with and without contributions from electrostatic terms and hydrogen bonds, etc. (The program DISCOVER, customized for carbohydrates and for operation on the Alliant FX/80 computer, has been used (12).) Generally, the time span for the simulations has been of the order of several hundred picoseconds to 1 nanosecond.

Although much additional work remains to be done, some interesting results have already emerged. For example, as illustrated in Figs. 8 and 9, the trajectories of the ϕ and ψ angles of cellobiose show a wide range of fluctuations in these angles over a time span of 500 ps, and varied residence times in the principal potential energy minima. The conformation changes considerably while resident in a minimum, before crossing over the energy barrier separating two minima. Even a greater degree of flexibility is manifested by the rotation of the C(6) hydroxymethyl group. As shown in Fig. 10, the O(6) hydroxyl spends considerable time in each of the three staggered rotational positions, commonly described as *gt*, *gg*, and *tg*.

The freedom to vary ϕ and ψ so extensively—even if exaggerated here owing to the absence of other molecules—renders cellulose fragments such as an octamer, surprisingly flexible. Cellulose is generally considered to be a stiff molecule, but at least single-molecule simulations in vacuum at 400 K show it to have considerable internal mobility. The range of conformations that result from this mobility are quite apparent when the molecular shape is
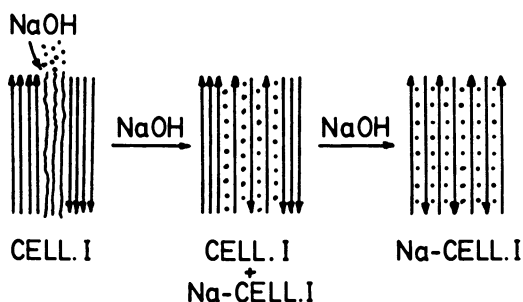
Figure 7. Conversion of cellulose I crystallites (left) to those of Na-cellulose I (right), by the action of NaOH. Reproduced with permission from ref. 24 (Copyright John Wiley & Sons, 1987).
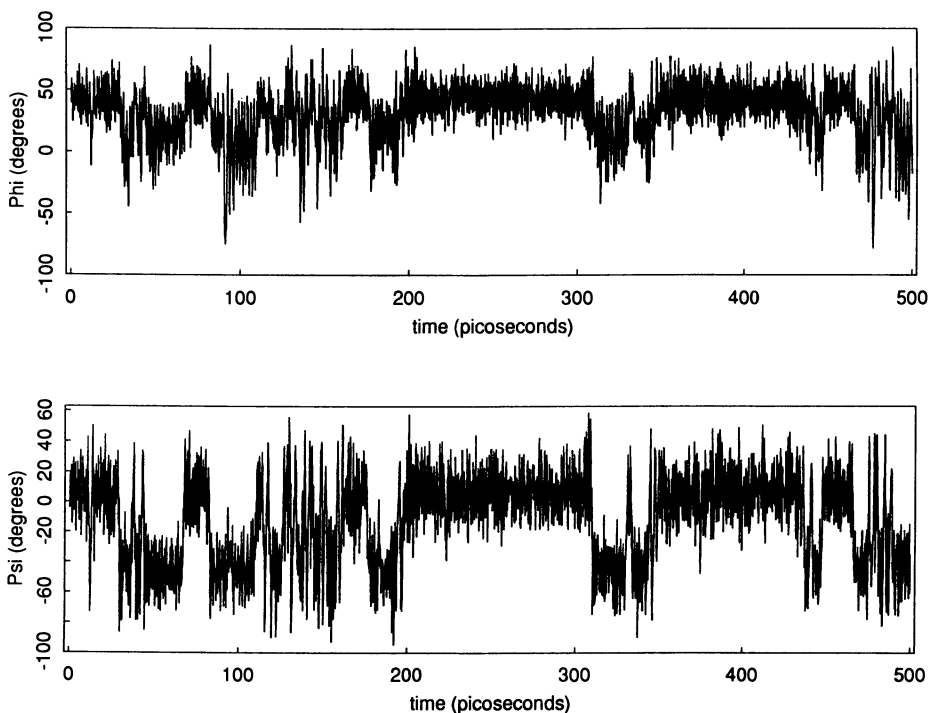
Figure 8. Trajectory of the ϕ (top) and ψ (bottom) angles of cellobiose during 500 ps simulation, at 400 K. (MM2(85) functions).
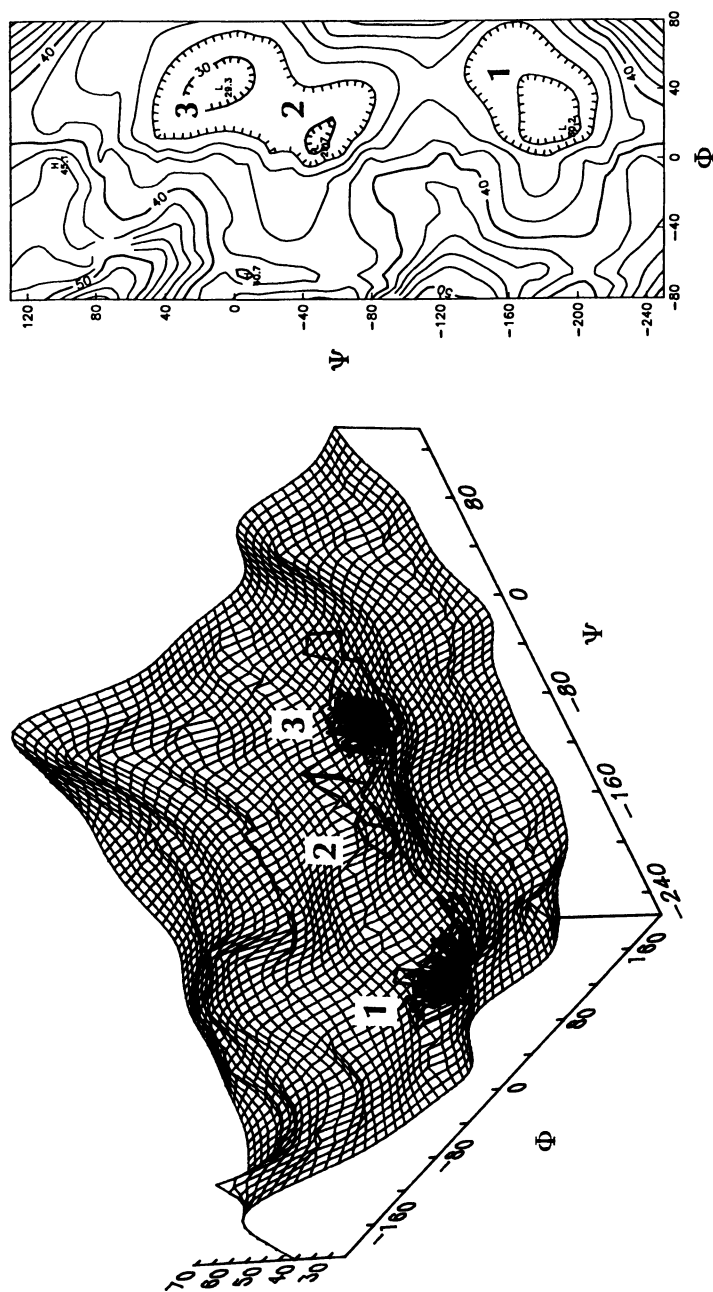
Figure 9. Potential energy surface for cellobiose at 400 K. The trajectory of conformational changes during a portion of the simulation are shown on the left. Energy contours in the vicinity of minima 1-3 are shown on the right. Barrier heights: 5.3 Kcal/mol between minima 1 and 2, 1.3 Kcal/mol between 2 and 3. (MM2(85) functions).
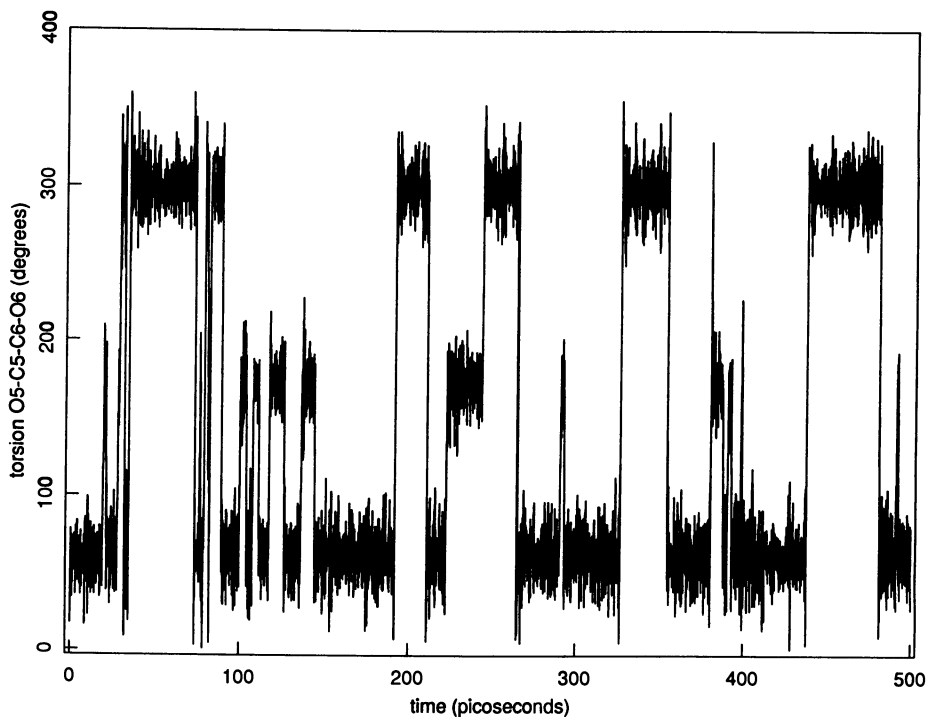
Figure 10. Trajectory of the rotation of one of the C(6) hydroxymethyl
groups in cellobiose, 400 K. (*gg*: 300°, *tg*: 180°,
gt: 60°. MM2(85) functions.)
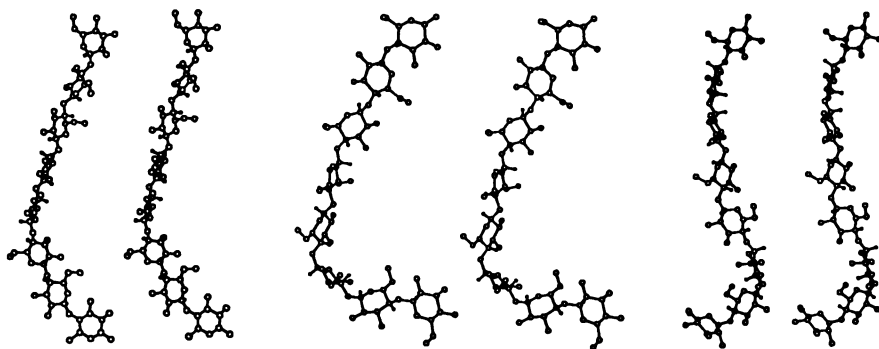
Figure 11. Snapshots of three conformations of an octamer
fragment at 400 K. (Stereo views.)

viewed on the computer screen during the molecular dynamics simulation. An octamer still appears to preserve an extended overall shape much of the time, but while it moves from one extended conformation to another, it "wriggles" through many varied and convoluted conformations. In Fig. 11 are shown snapshots of three typical conformations observed during a 200 ps simulation run at 400 K.

These studies are continuing to evaluate the effects of different potential functions, contributions from electrostatic and hydrogen bond attractions, temperature, etc. The effects of solvents, complexing agents, and adjacent surfaces of crystalline regions will be investigated in due course.

Acknowledgments

Literature Cited and Notes

1. G.A. Jeffrey and M. Sundaralingam, *Adv. Carbohydr. Chem. Biochem.*, **30**, 445 (1974); **31**, 347 (1975); **32**, 353 (1976); **34**, 345 (1977); **37**, 373 (1980); **38**, 417 (1981); **48**, 203 (1985).
2. A. Sarko and R.H. Marchessault, *J. Am. Chem. Soc.*, **89**, 6454 (1967).
3. C. Woodcock and A. Sarko, *Macromolecules*, **13**, 1183 (1980).
4. C.H. Haigler and M. Benziman, in *Cellulose and Other Natural Polymer Systems*, R.M. Brown, Jr., ed., Plenum Press, New York, NY, 1982, p. 273.
5. A. Sarko, H. Nishimura, and T. Okano, in *The Structures of Cellulose: Characterization of the Solid States*, R.H. Atalla, ed., ACS Symp. Ser., 340, American Chemical Society, Washington, DC, 1987, p. 169.
6. A. Sarko, in *Cellulose: Structure, Modification, and Hydrolysis*, R.A. Young and R.M. Rowell, eds., John Wiley, New York, NY, 1986, p. 29.
7. A.J. Stipanovic and A. Sarko, *Polymer*, **19**, 3 (1978).
8. E. Roche, H. Chanzy, M. Boudeulle, R.H. Marchessault, and P. Sundararajan, *Macromolecules*, **11**, 86 (1978).
9. (a) N.L. Allinger, *J. Am. Chem. Soc.*, **99**, 8127 (1977); (b) *Quantum Chemistry Program Exchange*, Chemistry Department, Indiana University, Bloomington, IN 47405; Molecular Design, Ltd., Hayward, CA.
10. S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta Jr., and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984).
11. B.R. Brooks, R.E. Broccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187 (1983).
12. DISCOVER, a molecular modeling program by Biosym Technologies, Inc., 10065 Barnes Canyon Road, San Diego, CA 92121.
13. P. Zugenmaier and A. Sarko, in *Fiber Diffraction Methods*, A.D. French and K.H. Gardner, eds., ACS Symp. Ser., 141, American Chemical Society, Washington, DC, 1980, p. 225.
14. P.J.C. Smith and S. Arnott, *Acta Cryst.*, **A34**, 3 (1978).
15. A.D. French, in *Cellulose and Wood—Chemistry and Technology*, C. Schuerch, ed., John Wiley and Sons, 1989, p. 103.
16. A. Sarko and R. Muggli, *Macromolecules*, **7**, 486 (1974).

17. K.H. Gardner and J. Blackwell, *Biopolymers*, **13**, 1975 (1974).
18. J. Mann, L. Roldan-Gonzales, and H.J. Wellard, *J. Polym. Sci.*, **42**, 165 (1960).
19. D.P. Miller, W.A. Roughead, and A.D. French, *Proc. 1983 Dissolving and Specialty Pulps Conf.*, TAPPI Press, Atlanta, GA, p. 233.
20. A.D. French, W.A. Roughead, and D.P. Miller, in *The Structures of Cellulose: Characterization of the Solid States*, R.H. Atalla, ed., ACS Symp. Ser., 340, American Chemical Society, Washington, DC, 1987, p. 15.
21. A. Sarko, *Appl. Polym. Symp.*, **28**, 729 (1976).
22. T. Okano and A. Sarko, *J. Appl. Polym. Sci.*, **29**, 4175 (1984).
23. T. Okano and A. Sarko, *J. Appl. Polym. Sci.*, **30**, 325 (1985).
24. H. Nishimura and A. Sarko, *J. Appl. Polym. Sci.*, **33**, 855 (1987).
25. H. Nishimura and A. Sarko, *J. Appl. Polym. Sci.*, **33**, 867 (1987).
26. H. Nishimura, T. Okano, and A. Sarko, to be published.

Chapter 22

Computer Simulation of Protein–Carbohydrate Complexes

Application to Concanavalin A and L-Arabinose-Binding Protein

V. S. R. Rao, B. V. S. Reddy, C. Mukhopadhyay, and M. Biswas

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

The CCEM (Contact Criteria and Energy Minimization) method has been developed and applied to the study of protein-carbohydrate interactions. The method uses available X-ray data on the native protein to generate realistic models of a variety of proteins with various ligands. The examples discussed in this paper are concanavalin A (ConA) and L-arabinose-binding protein (ABP). Complexes of ConA with alpha-methyl-mannoside (αMeMan) and a trimannoside, 3,6-di-O-(α-D-mannopyranosyl)-D-mannoside, which is present as a structural moiety in all complex carbohydrates and glycopeptides, were generated using the 2.4Å-resolution X-ray data of native ConA. It is suggested that the higher affinity of ConA for the trimannoside may be due to extended interactions rather than to an extended sugar binding site of ConA. The X-ray crystal structure data reported on the complex of ABP with L-arabinose were used to generate complexes of ABP with other sugars. The results are in good agreement with the experimental data.

There is no doubt that the three-dimensional structures of proteins and their complexes with various ligands are important in understanding many biological processes. The X-ray diffraction technique has been widely used to study such systems and much of our understanding of how proteins interact with their ligands are based on this technique. Although this is the only method that gives information about the precise positions of atoms, it has been noted that unless the structure is solved at high resolution (below 2.0Å) the details of protein-ligand interactions derived from this technique are not always reliable (1).

The large amount of X-ray data on protein structures that is available in the literature has not been fully exploited in

understanding biomolecular interactions. Model-building has
played an important role in the proper interpretation of X-ray
crystallographic data. Although the actual methods vary widely,
most model-building studies aim at interpreting existing X-ray
data rather than utilizing the available data to generate unknown
complexes. The CCEM method, developed in our laboratory (2), uses
available X-ray data to generate realistic models of a variety of
proteins with various ligands. Protein-ligand complexes can be
generated even when X-ray crystallographic data is available only
on the native protein. In cases where, perhaps, one complex has
been solved the CCEM method allows one to generate a large variety
of complexes with different ligands. This method, therefore,
provides a means for the comparison of the conformations of
different complexes of a particular protein with a variety of
ligands. In this paper, some examples are discussed with
reference to concanavalin A (ConA) and arabinose-binding protein
(ABP).
     ConA is specific for D-mannose and D-glucose and exhibits a
higher affinity for complex carbohydrates with a trimannosidic
core compared to the monosaccharides (3,4). These carbohydrates
are potential receptors on the outer surface of cells. The
structure of native ConA has been reported from X-ray
crystallographic studies at 2.4Å resolution (5). The only complex
of ConA which has been reported is with Methyl-α-D-mannoside
(αMeMan) which was solved at 6Å resolution (6). In this study the
approximate location of the sugar binding site was identified. No
details of protein-ligand interactions could be discerned at such
low resolution. In earlier work from our laboratory, using
contact criteria alone, we were able to identify possible binding
orientations for αMeMan (7) and other inhibitors of ConA (8,9).
These monosaccharide ligands occupy what we will refer to as the
primary sugar binding site. These simple fitting experiments
were extended to build complexes of ConA with larger ligands,
particularly with the trimannosidic core structure (10-12), the
common structural moiety present in all complex carbohydrates and
crucial for their binding to ConA. These studies indicated that
either of the two terminal mannose residues, the α(1→3) or α(1→6)
linked mannose, could fit in the primary sugar binding site in a
similar orientation as the monosaccharide αMeMan, but extended
interactions were possible only when the α(1→6) terminal mannose
residue was placed in the primary binding site. The middle
mannose could not bind in the primary site for stearic reasons.
Brewer and Bhattacharya (3) have proposed from NMR studies that
the two non-reducing terminal mannose residues in the
trimannosidic core bind simultaneously to an extended binding site
on each monomer of ConA. These workers give no indication as to
which of the two mannose residues occupy the primary binding site.
However, the interaction of the two terminal mannose residues with
the protein was suggested to be the reason for the better
inhibitory potency of the trimannosidic moiety compared to αMeMan.
Carver et al. (13) using model building and NMR experiments
suggested that the α(1→6)terminal mannose is likely to occupy the
primary binding site. In this paper the CCEM method has been used
to extend our earlier work to identify 1) the energetically

preferred mode(s) of binding of αMeMan to ConA and 2) which of the mannose residues in the trimannosidic core binds to ConA in the primary binding site and also to describe the possible interactions between the other two mannose residues and the protein.

L-arabinose-binding protein (ABP) is a periplasmic protein which binds specifically to L-arabinose (L-Ara) although other monosaccharides can act as inhibitors of this protein. Complexes of ABP with L-arabinose have been studied by X-ray crystallography at 2.8, 2.4 and 1.7Å resolutions (14-16). The predicted modes of binding of L-Ara to ABP from high resolution (1.7Å) data shows that some of the crucial protein-ligand interactions were either missed or misinterpreted at lower resolutions (2.8Å, 2.4Å). We have previously shown that starting from the 2.4Å resolution data of ABP, the CCEM method enables us to generate a three-dimensional model of the ABP-L-Ara complex which successfully predicts most of the hydrogen bonds reported at 1.7Å resolution (17). We now show that, starting from 2.4Å resolution data, the CCEM method can also be used to generate complexes of ABP with other inhibitors. This data is compared with the available experimental data to provide a theoretical explanation for the observed differences in the binding affinities of the various inhibitors chosen and also to throw light on the importance of noncovalent interactions in complex formation.

Method of Calculation

The CCEM method (contact criteria and energy minimization) as has been described previously (2) involves two steps: i) fitting of the ligand in the protein binding site using contact criteria and ii) discrimination of the allowed conformations by minimizing the conformational energy of the protein-ligand complex.

The coordinates of the proteins are taken from available X-ray data in the Protein Data Bank. The sugar ligands are generated in the binding site of the protein using a standard geometry (18) or using the available X-ray crystal structure of the particular ligand. The rigid-body rotation method is used to move the sugar in the binding site using rotational and translational parameters.

In the first step, those orientations for which there are no severe stearic contacts between the ligand and the protein are taken as the allowed orientations. In the second step, the energy of the complex is calculated using the allowed points generated in the first step as starting points. Empirical potential energy functions which include the van der Waals, electrostatic, torsional and hydrogen bond functions are used. A term which estimates the contribution to the energy of the exo-anomeric effect is also included for carbohydrate ligands (17). The potential energy functions and constants used in the calculations reported for ABP have been described earlier (17). An alternative version of the programme using the functions and constants given by Jorgensen et al. (19) is used for the calculations described for ConA. In the force field given by Jorgensen et al., no separate hydrogen bond function is used since this interaction is considered implicitly in the electrostatic energy component. For

complexes with ConA the charges on the ligand atoms are calculated using the CNDO/2 method (20).

During energy minimization, the side chains of the protein in the binding site are treated as flexible while backbone and other side chains are treated as rigid. The ligand is also treated as flexible and the energy of the ligand is included in the calculations. Also allowed to vary are the three rotational parameters defined by the Eulerian angles (phi,theta and psi), and the three translational parameters (x, y, and z) that define the orientation of the ligand in the binding site. Energy minimization is then carried out using the search method of Rosenbrock (21) (for ABP) or using the gradient-based method of Dennis and Mei (22) (for ConA). Some of the calculations were repeated using both the methods. We find that there is no significant difference in the final conclusions.

Results and Discussion

(I) Modelling Studies on Concanavalin A

(i) ConA-αMeMan: αMeMan was generated using the neutron diffraction data of Jeffrey et al. (23). The co-ordinates of ConA reported by Hardman and Ainsworth at 2.4Å (5) were taken from the Brookhaven Protein Data Bank (entry 3CNA). The protein residues considered for modelling the sugar binding site are given in Table I. Starting points for the force field calculations and optimization of interactions were taken from the stearic maps of the allowed orientations for αMeMan in the ConA binding site reported by Sekharudu and Rao (7). Of all the minimized orientations for the ConA-αMeMan complex the two binding modes with lowest energy are shown in Figures 1 and 2. The orientation of the ligand in the binding site of ConA, given by the three Eulerian angles and three translational parameters, for the two binding modes are (319°, 95°, 173°, 24.03Å, 24.16Å, 24.75Å) and (330°, 112°, 356°, 23.53Å, 23.92Å, 25.35Å), respectively.

Table I. Amino acid residues considered in modelling the sugar binding site of ConA (including the two metal ions)

8	9	10*	11*	12*	13	14*	15*	16*	17*	18
19*	20	21*	22*	23	24	25	26	28	31	32
33	34	37	38	39	40	41	42	43	44	94
95	96*	97*	98	99*	100*	101*	102	103	104	145
156	157	158	164	165	166	167	168	169	170	171
199	200	201	202	203	204	205*	206	207	208*	209
210	211	221	222	223	224	225	226*	227	228*	229*
230*	231	232	233	234	235	236	237		Mn++ Ca++	

* amino acid residues which are considered as flexible in the present study.
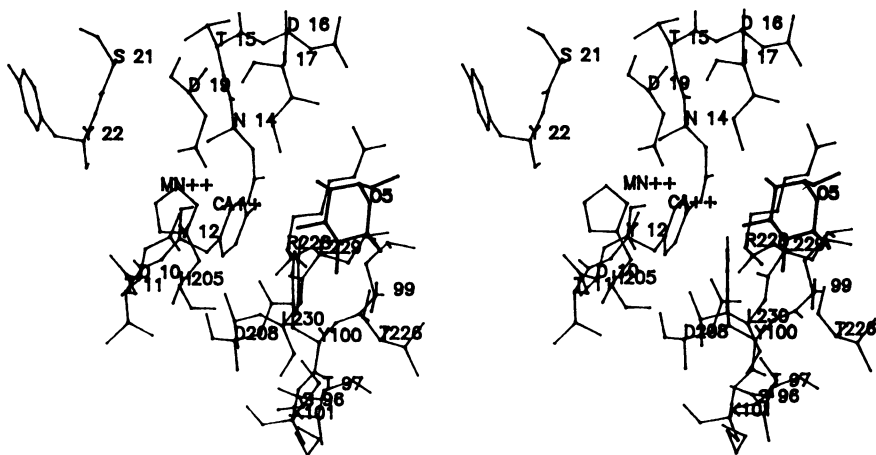
Figure 1. Stereoscopic projection of the complex between αMeMan and ConA : Binding mode (i). Details of the possible hydrogen bonds are given in Table III.
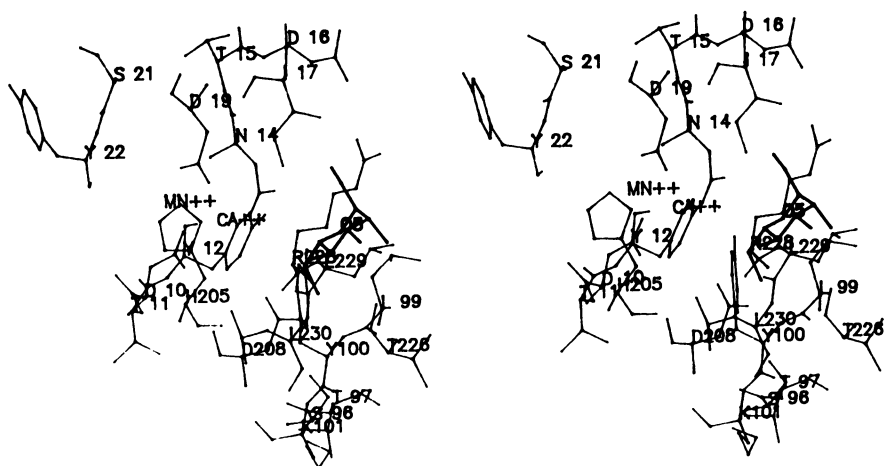
Figure 2. Stereoscopic projection of the complex between αMeMan and ConA : Binding mode (ii). Details of the possible hydrogen bonds are given in Table III.

Details of the conformational energies of the ConA-ligand complexes are given in Table II. The difference in energy between the two binding modes for αMeMan in the binding site of ConA is due mainly to the difference in the interaction energy component.

Table II. Conformational energies of ConA-ligand complexes

	Ligand[a] Energy	Protein[b] Energy	Interaction[c] Energy	Total[d] Energy	Normalized[e] Energy
1) ConA-αMeMan					
i 7.145	-1037.826	-39.736	-1070.416	-1076.116
ii 8.997	-1038.995	-34.714	-1064.711	-1070.411
2) ConA - Trimannoside[*], α(1→6) linked mannose residue in the primary binding site					
i 25.256	-1036.706	-61.484	-1072.934	-1088.434
ii 25.764	-1043.048	-53.847	-1071.100	-1086.600
3) ConA - Trimannoside, α(1→3) linked mannose residue in the primary binding site					
i 21.636	-1037.018	-46.959	-1062.340	-1077.840
ii 18.092	-1039.074	-41.823	-1062.805	-1078.305

[a]Conformational energy of the ligand in the end bound form.
[b]Conformational energy of the binding site residues of the protein.
[c]Interaction energy between the protein and the ligand.
[d]Sum of the above three components.
[e]Total Energy - Global minimum energy of the ligand.

```
*
            Man α(1→6)
         A            \
                        Man -OCH3    Trimannoside
                      /     B
            Man α(1→3)
              C
```
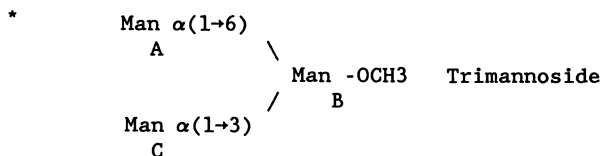
This difference occurs due to the smaller number of possible hydrogen bonds in the second binding mode compared to the first (Table III). The donor-acceptor relationship of the sugar hydroxyl groups in forming hydrogen bonds with the protein proposed by the present study are compared with that discussed by Poretz and Goldstein (24). In the first favoured binding mode of αMeMan, the C3 hydroxyl oxygen accepts a hydrogen and both the C4 and C6 hydroxyl groups donate a hydrogen in forming the hydrogen bonds with the protein, in agreement with the Goldstein criteria

(24). In addition the C4 and C6 hydroxyl groups also form two good hydrogen bonds by accepting a hydrogen each (Table III). The

Table III. Hydrogen bonds between ConA and αMeMan

Hydroxyl Groups	d/a[1]	Protein Residue	Goldstein criteria
Binding mode (i)			
OH1	--	---	a
OH2	d	ND2 ASN 14 (2.02[2],143.[3])	a
OH3	a	ND2 ASN 14 (2.29,140.)	a
OH4	a	N LEU 99 (2.24,154.)	
	d	OD1 ASP 208 (1.88,167.)	d
OH6	a	NE ARG 228 (2.14,157.)	
	d	O. THR 226 (1.86,152.)	d
O5	a	NH1 ARG 228 (2.50,130.)	-
Binding mode (ii)			
OH1	--	---	a
OH2	--	---	a
OH3	a	NE ARG 228 (1.99,142.)	a
OH4	d	OD1 ASP 208 (1.89,163.)	d
OH6	a	N TYR 100 (2.26,145.)	
	d	OD1 ASP 208 (1.89,170.)	d
O5	--	---	-

[1] d for donor, a for acceptor of a hydrogen in the hydrogen bond.
[2] hydrogen bond distance, Y-H<---->X, in Angstroms
[3] hydrogen bond angle, YHX, in degrees

C2 hydroxyl group forms a hydrogen bond by donating a hydrogen to the side chain nitrogen of Asn 14 rather than acting as an acceptor as suggested by Poretz and Goldstein. The anomeric oxygen does not form a hydrogen bond, while a weak hydrogen bond is possible between the ring oxygen, O5, and the side chain of Arg 228. In the second binding mode the C3,C4 and C6 hydroxyl groups also follow the Goldstein criteria. In this binding mode the C6 hydroxyl group also forms an additional hydrogen bond by accepting a hydrogen while the C1 and C2 hydroxyl groups and the ring oxygen, O5, are not involved in hydrogen bonding.

In a recent study, Hamodrakas et al. (25) using proton NMR measurements and a sophisticated interactive graphics facility proposed two possible modes of binding for 4'-nitrophenyl α-D-mannopyranoside to ConA. These two orientations differ from the two favoured orientations for αMeMan proposed in the present study. These authors fitted the ligand in the sugar binding site

with the help of an interactive graphics system where visual
adjustments were used to optimize the favourable hydrophobic
interactions and maximize the number of hydrogen bonds between the
protein and ligand. No attempt was made by these workers to
estimate the energy of the proposed binding modes. In contrast,
the CCEM method optimizes all interactions (non-bonded,
electrostatic and hydrogen bond) simultaneously allowing one to
choose the energetically most favourable conformations for the
complex. In fact we find that the orientations for the sugar
suggested by Hamodrakas et al. have high initial energies which
converge to some of the previously identified local minima.

(ii) <u>ConA-Trimannoside</u>: The trimannosidic core (structure given
at the foot of Table II) was generated using a standard geometry
and the preferred conformations were taken from those proposed by
Biswas et al. (<u>12</u>). Modelling was carried out placing each of the
terminal mannose residues, in turn, in the primary binding site
and allowing the resulting complex to minimize as described above.
Different possible conformations for the $\alpha(1{\rightarrow}3)$ linked glycosidic
bond of the ligand were tried. The conformational angles of the
$\alpha(1{\rightarrow}6)$ glycosidic bond are rather restricted as has been shown
earlier (<u>12</u>). Flexibility is possible only in the chi angle
(C4B-C5B-C6B-O6B) which can take all three staggered orientations
in the isolated ligand. When the $\alpha(1{\rightarrow}6)$ linked mannose residue of
the trimannosidic core is placed in the primary binding site of
ConA, only a chi angle close to $+60°$ places the other terminal
mannose residue in close proximity to the protein surface
facilitating additional interactions between the ligand and the
protein (<u>3</u>). Thus, a chi angle of $+60°$ has been used in the
present study.

      When the $\alpha(1{\rightarrow}6)$ linked terminal mannose residue is placed at
the primary binding site in the first favoured orientation of
αMeMan (Table III), the hydrogen bonds reported for αMeMan remain
essentially unchanged. In this orientation the $\alpha(1{\rightarrow}3)$ terminal
mannose comes close to the protein surface and its C4 hydroxyl
group forms a good hydrogen bond with the hydroxyl group of Tyr
100. The extra interactions from the second terminal mannose
residue contribute significantly to the interaction energy. When
the $\alpha(1{\rightarrow}6)$ terminal residue is placed in the second possible
orientation of αMeMan in the ConA binding site, it moves slightly
away from the monosaccharide orientation resulting in the loss of
some of the hydrogen bonds reported for the monosaccharide.
However, in this orientation the other terminal mannose residue
also forms a hydrogen bond with NE2 His205 (Table IV).
Stereoscopic projections for the complex with the $\alpha(1{\rightarrow}6)$terminal
mannose residue in the primary site of ConA in the two possible
orientations are shown in Figures 3 and 4. These orientations are
slightly different from the one proposed by Carver et al. (<u>13</u>),
using NMR and docking studies.
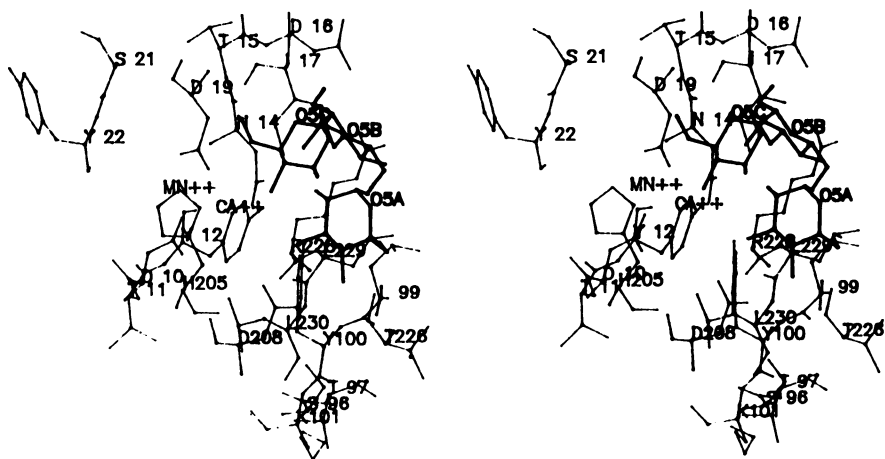
Figure 3. Stereoscopic projections of the complex between the trimannoside and ConA, with the α(1→6)-linked terminal mannose in the primary binding site : Binding mode (i). Details of the possible hydrogen bonds are given in Table IV.
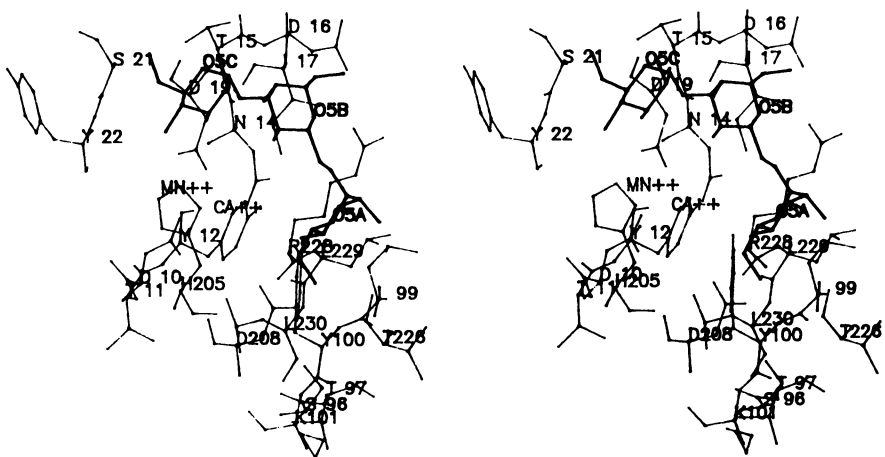
Figure 4. Stereoscopic projection of the complex between the trimannoside and ConA, with the α(1→3)-linked terminal mannose in the primary binding site : Binding mode (ii). Details of the possible hydrogen bonds are given in Table V.

Table IV. Hydrogen bonds between ConA and the trimannoside,
α(1→6)-linked mannose residue in primary binding site

Hydroxyl groups	Protein Residue
Binding mode (i)	
OH2A	ND2 ASN 14 (2.03,133.)
OH3A	OD1 ASP 208 (2.12,167.)
OH4A	N LEU 99 (2.15,156.)
	OD1 ASP 208 (1.76,178.)
OH6A	NE ARG 228 (2.19,153.)
	O THR 226 (1.91,147.)
O5A	NH1 ARG 228 (2.46,136.)
OH4C	OH TYR 100 (1.95,170.)
Binding mode (ii)	
OH4A	OH TYR 12 (2.11,154.)
OH6A	N TYR 100 (2.23,143.)
OH2B	O PRO 13 (2.49,144.)
OH4C	NE2 HIS 205 (2.34,140.)

Carver et al. propose that the middle mannose residue in the
trimannosidic core forms hydrogen bonds with the protein whereas
in the present study it is the α(1→3) terminal mannose residue
which forms additional stabilizing interactions with the protein.
    When the α(1→3)-linked terminal mannose residue is placed in
the binding site in the first orientation all the hydrogen bonds
obtained for the monosaccharide are retained except for the one
involving the C2 hydroxyl group (Table V). In this mode of
binding, however, the α(1→6)linked mannose interacts only weakly
with the protein. This is reflected in the poor interaction energy
(Table II) for the ConA-trimannoside complex having the α(1→3)
terminal mannose residue in the sugar binding site. In the second
orientation the α(1→3) mannose residue in the primary binding
site, as mentioned for the α(1→6) linked mannose residue, moves
slightly away from the monosaccharide orientation resulting in
fewer hydrogen bonds when compared to αMeMan in the same
orientation. In this orientation, although the other terminal
residue does interact with the protein, the weak interaction of
the residue in the primary binding site leads to a weak complex as
reflected in the poor interaction energy (Table II).
    Thus the present study shows that ConA interacts with the
trimannosidic core by binding preferentially to the α(1→6)-linked
terminal mannose residue at the primary binding site. Depending
on the orientation of the α(1→6) terminal residue in the primary
binding site the α(1→3) terminal mannose residue interacts either
with Tyr 100 or His 205 on the surface of ConA (Figures 3 and 4).
This indicates that there is no specific extended binding site

for the trimannoside; rather the sugar residues outwards from the primary binding site interact non-specifically with the protein. Hence the increased affinity of ConA for the larger ligands may be due to extended interactions rather than to the existence of an extended binding site.

Table V. Hydrogen bonds between ConA and the trimannoside, α(1→3) linked mannose residue in primary binding site

Hydroxyl groups	Protein Residue
Binding mode (i)	
OH3C	ND2 ASN 14 (2.30,139.)
OH4C	N LEU 99 (2.16,153.)
	OD1 ASP 208 (1.83,171.)
OH6C	NE ARG 228 (2.25,159.)
	O THR 226 (1.82,153.)
Binding mode (ii)	
OH3C	N ARG 228 (2.48,156.)
OH4C	OD1 ASP 208 (1.94,159.)
OH3B	OD1 ASP 16 (2.41,157.)
OH4A	NH1 ARG 228 (2.26,152.)

II) <u>Modelling Studies on L-arabinose-binding Protein (ABP)</u>

Solution studies have shown (<u>26</u>) that D-Gal can bind to ABP with an affinity comparable to that of L-Ara. D-Fuc and D-Xyl, on the other hand, have been found to act only as moderate inhibitors (<u>26</u>,<u>27</u>). To explain these observations the CCEM method was used to generate three-dimensional complexes of ABP with the α and β-anomers of D-Gal, D-Fuc and D-Xyl. We report here the results obtained with the α-anomers only, since complexes with the β-anomers of these sugars have been found to be energetically less favoured.

The low resolution difference electron density maps indicated that the ligands L-Ara and D-Gal bind to ABP in the same manner (15). Based on this observation the α-anomers of D-Gal, D-Fuc and D-Xyl were initially placed in the binding site in the same position as that of β-L-Ara (<u>17</u>). The stereochemically allowed orientations of the sugars in the binding site having minimum stearic overlap and a good number of hydrogen bonds with the protein were chosen as the starting points for the energy minimization. The conformational energies of the protein-inhibitor complexes generated using this method (Table VI) suggest that of the three inhibitors α-D-Gal will have the highest inhibitory property, while α-D-Xyl will be least active. This is in agreement with solution studies (<u>27</u>).

TABLE VI. Comparison of the most probable modes of binding
of various inhibitors to ABP

LIGAND	ORIENTATION OF LIGAND						CONFOR-MATIONAL ENERGY (kcal/mol)	HYDROGEN BONDING SCHEME
	Phi	The	Psi	X	Y	Z		
	(degrees)			(Å)				
α-D-Gal	196	86	90	13.8	56.3	53.9	-59.6	OH(1) - Lys10 NZ
								OH(1) - Asp89 OD2
								OH(2) - Lys10 NZ
								OH(3) - Asn205 ND2
								OH(3) - Asn232 ND2
								OH(4) - Arg151 NH2
								OH(4) - Asn232 OD1
								OH(6) - Arg151 NH1
α-D-Fuc	192	88	99	13.8	56.3	53.9	-57.7	OH(1) - Asp89 OD2
								OH(2) - Lys10 NZ
								OH(3) - Asn205 ND2
								OH(4) - Arg151 NH2
								OH(4) - Asn232 OD1
α-D-Xyl	200	49	140	14.0	56.9	53.5	-54.2	OH(1) - Asp89 OD2
								OH(2) - Lys10 NZ
								OH(3) - Asn232 ND2
								OH(4) - Asn232 OD1
								O(5) - Arg151 NH1
β-L-Ara[*]	196	88	96	14.2	56.6	53.5	-59.2	OH(1) - Asp89 OD2
								OH(2) - Lys10 NZ
								OH(3) - Asn205 ND2
								OH(3) - Asn232 ND2
								OH(4) - Arg151 NH2
								OH(4) - Asn232 OD1
								O5 - Arg151 NH1

* from previous studies of Mukhopadhyay and Rao ([17]).

The most probable mode of binding of α-D-Gal was found to be quite similar to that of β-L-Ara (Table VI) although a few differences do exist in the hydrogen bonding scheme. In the ABP-α-D-Gal complex OH(1) forms an extra hydrogen bond with Lys 10 (Table VI) (Figure 5). In this complex Arg 151 forms hydrogen bonds with OH(4) and OH(6), whereas in the ABP-β-L-Ara complex Arg 151 forms hydrogen bonds with OH(4) and O5. Thus in the complex of ABP with α-D-Gal the hydrogen bond between O6 and Arg 151 essentially replaces that between O5 and Arg 151 present in the ABP-β-L-Ara complex. In the ABP-α-D-Gal complex the ring oxygen O5 is not involved in hydrogen bonding with any of the protein residues. It is interesting to note that the 2.4Å resolution X-ray study of ABP-β-L-Ara complex also failed to predict any specific interaction between the ring oxygen O5 of β-L-Ara and the protein residues. It can be recalled here that the binding site of ABP was located originally using the heavy atom analog of D-Gal (6-bromo-6-deoxy-galactose). Thus it seems possible that the use of a heavy atom analog of D-Gal might have biased the interpretation of the data obtained for the ABP-β-L-Ara complex at lower resolution (2.4 Å).

For α-D-Fuc and α-D-Xyl the higher conformational energy of the complexes suggest that α-D-Fuc and α-D-Xyl will be weaker inhibitors than α-D-Gal. α-D-Fuc will, however, be better than α-D-Xyl. In the case of the ABP-α-D-fucose complex (Figure 6), OH(1) forms a hydrogen bond with Asp 89, OH(2) with Lys 10 and OH(3) with Asn 205. OH(4) can form two hydrogen bonds with Arg 151 and Asn 232. O5 in this complex may not form any hydrogen bond. In the ABP-α-D-xylose complex (Figure 7) the presence of an equatorial hydroxyl group at the C4 atom leads to a small change in the orientation of the pyranose ring in the binding site. In this complex OH(1) forms a hydrogen bond with Asp 89, OH(2) with Lys 10. OH(3) and OH(4) both can form hydrogen bonds with Asn 232, and O5 can hydrogen bond with Arg 151.

Comparison of the modes of binding of these inhibitors (Table VI) shows that the bidentate hydrogen bonds between Arg 151, Asn 232 and the sugar hydroxyls are possible only in ABP-β-L-Ara and ABP-α-D-Gal complexes, while in the ABP-α-D-Fuc and ABP-α-D-Xyl complexes either or both of the bidentate hydrogen bonds are lost. The present study thus reveals that the bidentate hydrogen bonds involving Arg 151, Asn 232 and the sugar hydroxyls are crucial in governing the binding specificity of ABP. This study, therefore, provides valuable information about the modes of binding of some of the inhibitors to ABP and also provides a theoretical explanation for the relative binding affinities of these sugars t the protein.
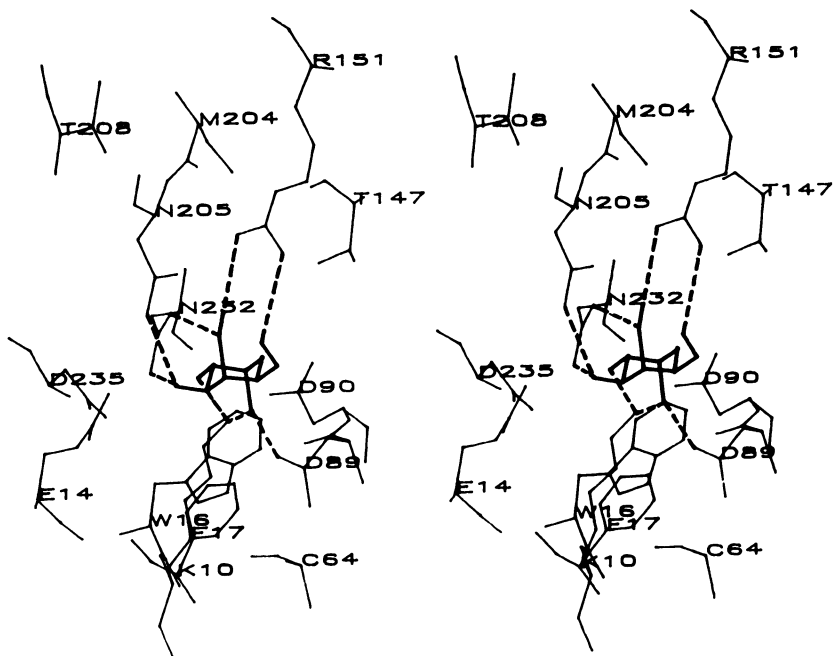
Figure 5. Stereoscopic projection of the complex between α-D-Gal and ABP. Details of the hydrogen bonds are given in Table VI.
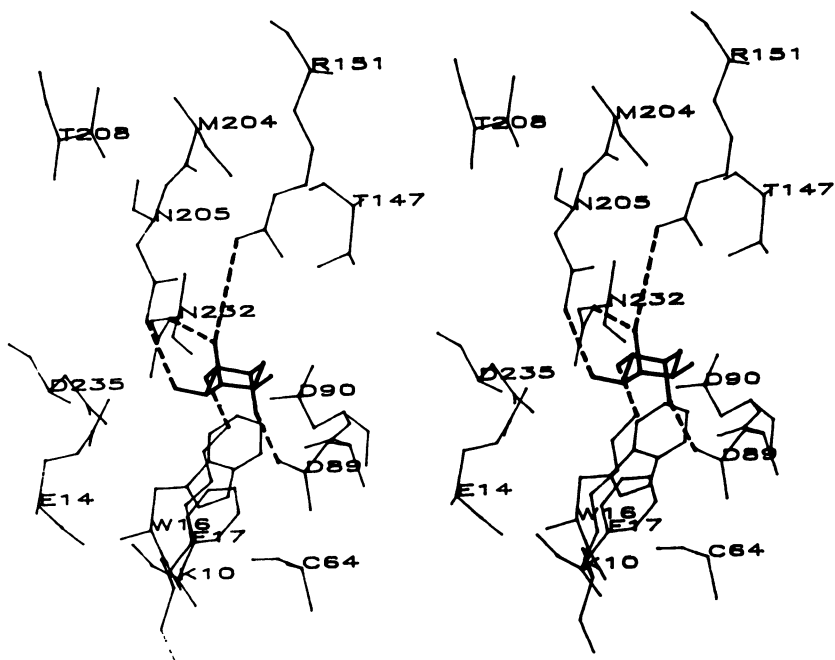
Figure 6. Stereoscopic projection of the complex between α-D-Fuc and ABP.
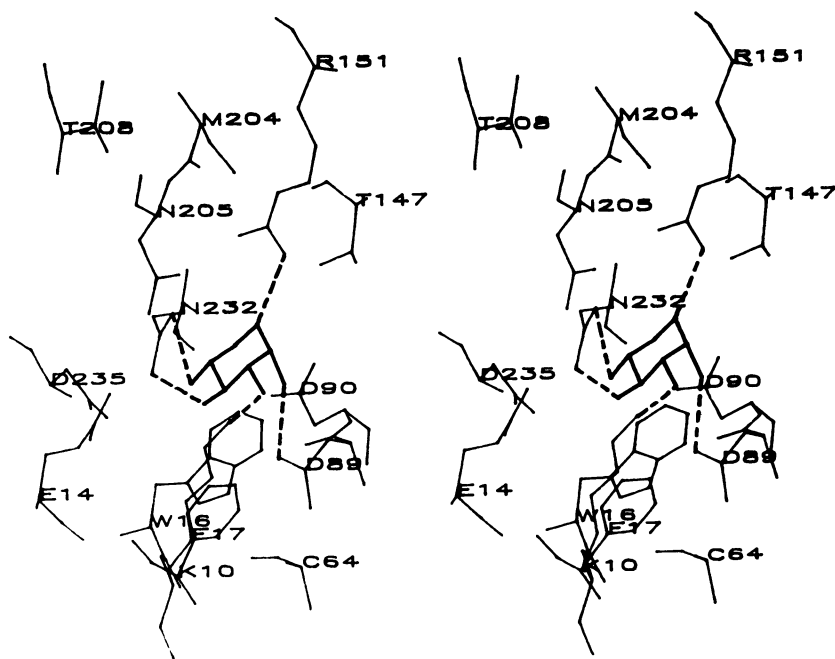
Figure 7. Stereoscopic projection of the complex between α-D-Xyl and ABP.

REFERENCES

1.   Quiocho, F.A. Ann. Rev. Biochem. 1986, 55, 287.
2.   Rao, V.S.R., Biswas, M., Mukhopadhyay, C. and Balaji, P.V. J. Mol. Struct. 1989, 194, 203.
3.   Brewer, C.F and Bhattacharya, L. J. Biol. Chem. 1986, 261, 7306.
4.   Bhattacharya, L., Haraldsson, M. and Brewer, C.F. J. Biol. Chem. 1987, 262, 1294.
5.   Hardman, K.D. and Ainsworth, C.F. Biochemistry 1972, 11, 4910.
6.   Hardman, K.D. and Ainsworth, C.F. Biochemistry 1976, 15, 1120.
7.   Sekharudu, Y.C. and Rao, V.S.R. Int. J. Biol. Macromol. 1984, 6, 337.
8.   Sekharudu, Y.C. and Rao, V.S.R. J. Biomol. Struct. Dyn. 1984, 2, 41.
9.   Sekharudu, Y.C. and Rao, V.S.R. Curr. Sci. (India) 1984, 53, 403.
10. Biswas, M., Sekharudu, Y.C. and Rao, V.S.R. Int. J. Biol. Macromol. 1986, 8, 2.
11. Sekharudu, Y.C., Biswas, M. and Rao. V.S.R. Int. J. Biol. Macromol. 1986, 8, 9.
12. Biswas, M., Sekharudu, Y.C. and Rao, V.S.R. Carbohydr. Res. 1987, 160, 151.
13. Carver, J.P., Mackenzie, A.E. and Hardman, K.D. Biopolymers 1985, 24, 49.
14. Quiocho, F.A., Gilliland, G.L. and Phillips, G.N., Jr. J. Biol. Chem. 1977, 252, 5142.
15. Newcomer, M.E., Gilliland, G.L. and Quiocho, F.A. J. Biol. Chem. 1981, 256, 13213.
16   Quiocho, F.A. and Vyas, N.K. Nature 1984, 310, 381.
17. Mukhopadhyay, C. and Rao, V.S.R. Int. J. Biol. Macromol. 1988, 10, 217.
18. Arnott, S. and Scott, W.E. J. Chem. Soc., Perkins Trans. 1972, 2, 324.
19. Jorgensen, W.L. and Tirado-Rives, J. J. Am. Chem. Soc. 1988, 110, 1657.
20. Pople, J.A. and Segal, G.A. J. Chem. Phys. 1966, 44, 3289.
21. Rosenbrock, H.H. Comput. J. 1960, 3, 175.
22. Dennis, J.E., Jr. and Mei, H.H.W. J. Optimization Theory and Applications 1979, 28, 453.
23. Jeffery, G.A., McMullan, R.K. and Takagi, S. Acta Cryst. 1977, B33, 728.
24. Poretz, R.D. and Goldstein, I.J. Biochemistry 1970, 9, 2890.
25. Hamodrakas, S.J., Alexandraki, E., Troganis, A. and Stassinopoulou, C.I. Int. J. Biol. Macromol. 1989, 11, 17.
26. Brown, C.E. and Hogg, R.W. J.Bacteriol. 1972, 111, 606.
27. Miller, D.M., III, Olson, J.S., Pflugrath, J.W. and Quiocho, F.A. J. Biol. Chem. 1983, 258, 13665.

Chapter 23

Lysozyme Hydrolysis of β-Glycosides

A Consensus Between Binding Interactions and Mechanism

Carol Beth Post[1,3,4], Christopher M. Dobson[2], and Martin Karplus[3]

[1]**Department of Biological Sciences, Purdue University, West Lafayette, IN 47907**
[2]**Inorganic Chemistry Laboratory, Oxford University, OX1 3QR, England**
[3]**Chemistry Department, Harvard University, Cambridge, MA 02138**

Analysis of a molecular dynamics trajectory of the
enzyme-substrate complex chicken lysozyme and (GlcNAc)$_6$
provides insight into the mechanism of polysaccharide
hydrolysis by lysozyme. An alternative mechanism, which
involves *endo*cyclic bond cleavage and no ring distortion
was formulated. Binding interactions stabilize a
conformation of the glycosidic linkage to be hydrolyzed
which is optimum for catalysis by the alternative
mechanism. In contrast, the other linkages of (GlcNAc)$_6$
have a cellulose-like conformation. The energy of the
average structure and the dynamically averaged energy
calculated for residues in each of the six (GlcNAc)
binding sites vary due to differences in intermolecular
nonbond contributions; the configurational and
intramolecular nonbond energies are similar for all
sites. The relative energies are consistent with the
experimental observation that the E-F dimer product
dissociates more rapidly than the A-D tetramer product.
Atomic fluctuation cross-correlations between enzyme and
substrate reveal that correlations are not uniform
throughout the binding cleft.

In this paper we describe energetic and dynamic properties of the
enzyme-substrate interactions in the complex of chicken lysozyme and
hexa-(N-acetylglucosamine), (GlcNAc)$_6$, as obtained from a molecular
dynamics (MD) simulation (1). Lysozyme was the first enzyme to have
its three-dimensional structure determined by X-ray crystallography
(2), yet the details of the catalytic mechanism remain elusive.
Although experiments have played an essential role in determining
certain features of the enzymic reaction, the information that they
provide is limited. The kinetics and thermodynamics of species
along the reaction pathway have been measured (3 and references
cited therein), evidence for certain chemical transformations has
been obtained from isotope effects (4,5) and the importance of
functionalities has been demonstrated by studying different

[4]Current address: Department of Medicinal Chemistry, Purdue University, West Lafayette, IN
  47907

substrates or modified enzymes (6,7). However, in this case, as in
others, supplementary information supplied by MD or related
techniques is essential for a full understanding of the mechanism at
the atomic level.

One result from the analysis of the MD simulation was the
proposal of a new enzymic pathway for hydrolysis by lysozyme. We
begin with a description of the alternative mechanism, and the basis
on which it was proposed. The energetics of the individual GlcNAc
units in the lysozyme cleft are then presented, followed by a
graphical representation of the correlation between the atomic
fluctuations of the substrate and those of the enzyme. Of
particular interest is the fact that the binding interactions
stabilize a bound state conformation for the two glycosides involved
in hydrolysis that is optimum for catalysis by the alternative
mechanism and which differs from the conformations of the other
glycosides. These conformational features are described in the
final two sections.

Mechanism

A pathway (Scheme I) (8,9) for the hydrolysis of oligoglycosides by
lysozyme that differs from the previously accepted mechanism
(Scheme II) (3,10-12) is described in this section. The alternative
pathway, suggested by results of a 55-ps MD simulation of the
lysozyme·(GlcNAc)$_6$ complex (1), is consistent with the available
experimental data and with stereoelectronic considerations.
Experimental data have demonstrated that Glu 35 and Asp 52 are
essential, as shown by recent site-directed mutagenesis results (13)
which corroborate chemical modification studies (3,14 and references
cited therein), and that the reaction proceeds with retention of
configuration at C$_1$ (3 and references cited therein). A fundamental
feature of the alternative pathway is that an *endocyclic* bond is
broken in the initial step, in contrast to the *exocyclic* bond
cleavage in the accepted mechanism.

The MD simulation employed an initial structure with (GlcNAc)$_6$
built into the active site by use of the crystallographic
coordinates of a lysozyme-(GlcNAc)$_3$ complex; details have been
reported (1). In the crystallographic structure, (GlcNAc)$_3$ occupies
sites A, B and C in the active site cleft. To determine coordinates
for the sugar molecules in sites D, E and F, a GlcNAc monomer was
built into each site using a computer graphics system. Starting in
site D, a GlcNAc monomer in a regular chain conformation was added
with a β-linkage to the terminal oxygen atom of the sugar in site C.
The bonds of the glycosidic linkage were rotated until the sugar
fitted the site without unreasonably close contacts with the
protein. Further fitting was done by rotating the hydroxyl and
acetamide side-groups of the sugar to optimize hydrogen-bond
formation. The sugars in sites E and F were built sequentially from
site D in a similar fashion. Removal of bad contacts in some cases
involved rotation of amino acid side-chains; no rotations of
backbone dihedral angles were required. Favorable binding of the
N-acetylglucosamine residue to site D was found without distorting
the ring. The model was constructed to maximize intermolecular
contact and to avoid short interatomic distances. During the
simulation, the chair form of the pyranose ring in site D remained
unperturbed. The motions of the carboxyl group of Glu 35 led to
hydrogen bonds with the endocyclic oxygen O$_5$ and the hydroxymethyl
oxygen O$_6$ but *not* with the exocyclic oxygen O$_4'$.
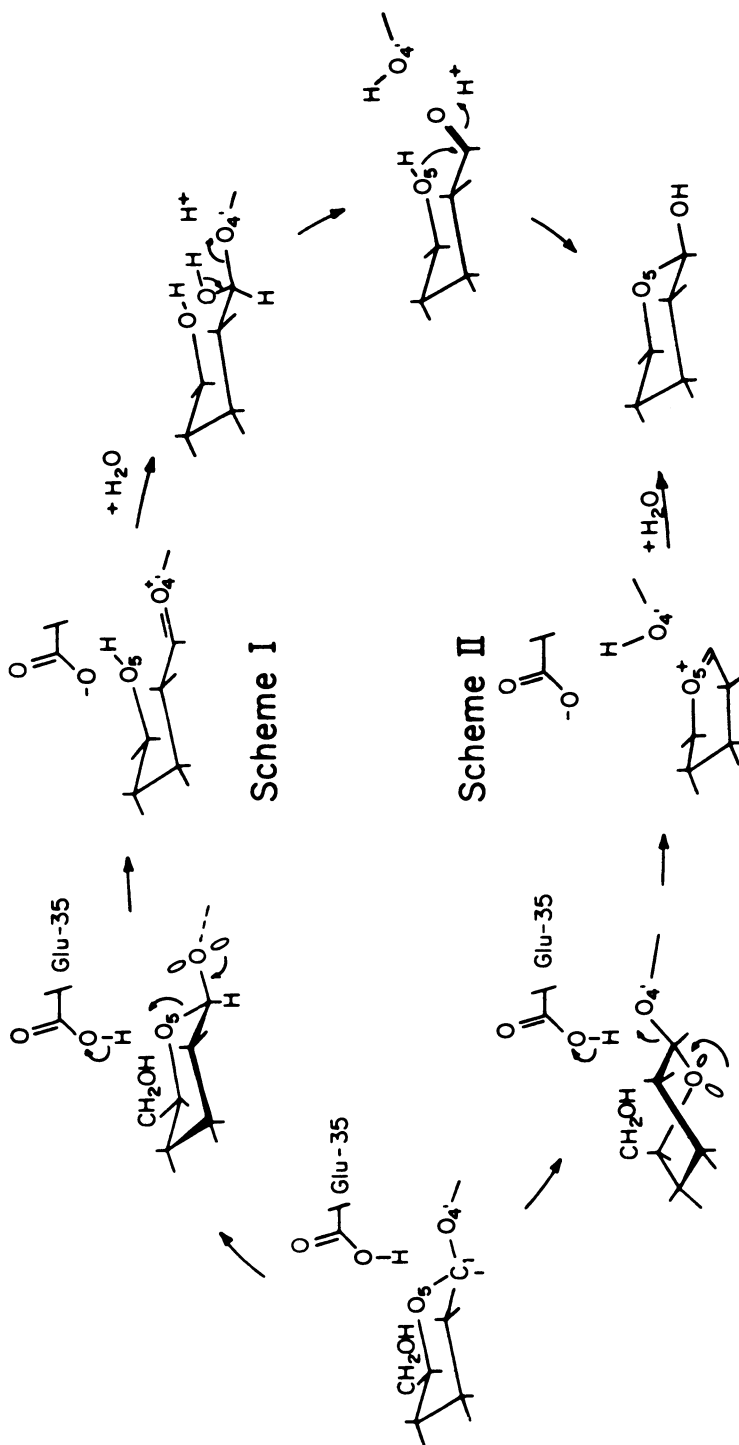
Scheme I

Scheme II

The standard mechanism for lysozyme (Scheme II) was proposed by Phillips and co-workers (10) on the basis of model building and data for the nonenzymatic hydrolysis of glycosides. An essential element of this proposal is the distortion of the GlcNAc residue in site D. The resulting twist-boat conformation makes it possible to take advantage of stereoelectronic assistance (15-17) from the ring oxygen O_5 in the transition state leading to cleavage of the exocyclic C_1–O_4' bond. Scheme II also involves protonation of O_4' by Glu 35 and yields the cyclic oxocarbonium ion which can be stabilized by the carboxylate group of Asp 52.

The initial step in the alternative hydrolysis mechanism is protonation of the ring O_1 by Glu 35 (Scheme I). Cleavage of the endocyclic C_1–O_5 bond forms the acyclic oxocarbonium ion intermediate, which is stabilized by Asp 52. Attack by water, cleavage of the C_1–O_4' bond, and ring closure then lead to the observed products. Existing experimental data on lysozyme hydrolysis are consistent with Scheme I (see references in Post and Karplus (9)). Moreover, distortion of the ring in site D is not required and the antiperiplanar orientation of an exocyclic O_4' lone pair orbital relative to the cleaved C_1–O_5 bond found in the simulation (see section on "Enhancement of a Substrate Conformation Optimum for Catalysis") is in accord with stereoelectronic requirements (16).

In Scheme I, a role of the enzyme is to catalyze the reaction by means of orientational (entropic) contributions, instead of the distortional (enthalpic) stabilization assumed in the standard mechanism (Scheme II). In particular, the intermolecular interactions could serve to restrict oscillation about the dihedral angle ϕ (O_5–C_1–O_4'–C_4') and to aid in maintaining the proper geometry for reclosing the ring. The MD results are only suggestive, and nothing in the present analysis would require that the same mechanism be found in all β-glycosidases.

Energy of GlcNAc Sites

An analysis of the potential energy of the individual GlcNAc units was performed to obtain information concerning the degree of interaction at each site of the binding cleft of lysozyme, as well as to assess the fit achieved by model building. The potential energy comprises the configurational terms for bonds, angles and dihedral angles and nonbond terms for van der Waals, electrostatics, and hydrogen bonds (18). Coordinates from the initial crystallographic/model–built structure and the average dynamics structure, both optimized by energy minimization, were used to evaluate the energies. The energy correspondong to individual coordinate sets from the trajectory was also calculated to obtain the average energy at each site over the 55 ps of the simulation. This procedure provides the average potential energy, $<E(r)>$, the physically relevant quantity, which can be compared to the energy of the average structure, $E(<r>)_{dyn}$. (The potential energy function is expressed in terms of atomic positions, r.) Evaluating $E(<r>)_{dyn}$ is computationally more efficient and corresponds most closely to the use of an X-ray structure to estimate interactions. However, the average structure may have artifacts introduced by averaging atomic positions which could lead to significant deviations from the average energy $<E(r)>$. Such dynamical averaging effects are evident when $<E(r)>$ deviates from $E(<r>)_{dyn}$. In contrast, deviations between $<E(r)>$ and $E(<r>)_{init}$, the energy of the X-ray/model built structure,

can arise from either dynamic effects or differences in the average structure (19).

$E(<r>)_{init}$, $E(<r>)_{dyn}$ and $<E(r)>$ are plotted in Figure 1a for residues in sites A through F of the binding cleft as labeled along the abcissa. The contributions from the configurational, the substrate-substrate nonbond and substrate-protein nonbond terms to $E(<r>)_{init}$, $E(<r>)_{dyn}$ and $<E(r)>$ are plotted in Figures 1b through 1d, respectively, and listed in Table I. Decomposition of the total potential energy per residue provides insight into the nature of the enzyme-substrate interaction. From Figure 1a it is seen that relative to $E(<r>)_{init}$ (▲) and $E(<r>)_{dyn}$ (■), the average energy $<E(r)>$ (●) is higher overall due to the presence of kinetic energy in the 304 K simulation, which allows deviation from the energy minimum. The increased energy resides primarily in the configurational terms, as shown by the upward shift in the dotted curve of Figure 1d.

The middle GlcNAc residues have a lower potential energy than the terminal residues for all curves in Figure 1a. Indeed sites C and D have the lowest energies of the hexasaccharide. This aspect contrasts with the poor steric contacts expected in site D if there were strain in the ring bound at this site (20).

The variation of the potential energy among the sites results primarily from substrate-protein interactions; examination of Figures 1b, 1c and 1d shows that the intermolecular nonbond energy varies along the cleft (dot-dash curves) while the intramolecular nonbond (dash curves) and configurational (dot curves) terms are nearly equal at all sites. Thus the geometries of the monomers are energetically equivalent, with no structural strain induced in favor of intermolecular nonbond interactions.

$E(<r>)_{init}$ and $E(<r>)_{dyn}$ are very similar except for sites A and B. In the initial structure, the energy $E(<r>)_{init}$ for B is higher than that of any other site, while the average structure energy $E(<r>)_{dyn}$ of site B is lower than that of A, E and F. $E(<r>)_{dyn}$ of site B is decreased because of the more favorable intermolecular nonbond energy obtained in the dynamic calculation (compare Figures 1b and 1c).

The deviation in the substrate energy at site A between $E(<r>)_{init}$ and $E(<r>)_{dyn}$ is the result of dynamic averaging. To demonstrate this point, the relative site energies for the averages $E(<r>)_{dyn}$ and $<E(r)>$ are compared. The site dependence for the average energy $<E(r)>$ and the energy of the average dynamics structure $E(<r>)_{dyn}$ is similar except at site A; site A has the highest potential energy when $E(<r>)_{dyn}$ is evaluated, yet its energy is lower than that of E and F when $<E(r)>$ is evaluated. The large $E(<r>)_{dyn}$ value is due to unfavorable intramolecular nonbond energy (see Figure 1c and Table I); close van der Waals contacts result when the coordinates are averaged over the trajectory and are not removed by energy minimization. In the individual dynamics structures, the contacts are longer so that the relative energy of site A is lower when $<E(r)>$ is evaluated.

There is an asymmetry with respect to the middle of $(GlcNAc)_6$ in the energy profile; site E has higher energy $<E(r)>$ than site B (Figures 1a and 1d). Compared with the other GlcNAc sites, there are fewer contacts between the substrate and the enzyme at site E. In particular, the intermolecular hydrogen bond energy is 1/3 to 1/2 that for the other residues. This lack of interaction is consistent with results of transglycosylation experiments which indicate that site E has a low substrate specificity (3). The acetamido group of
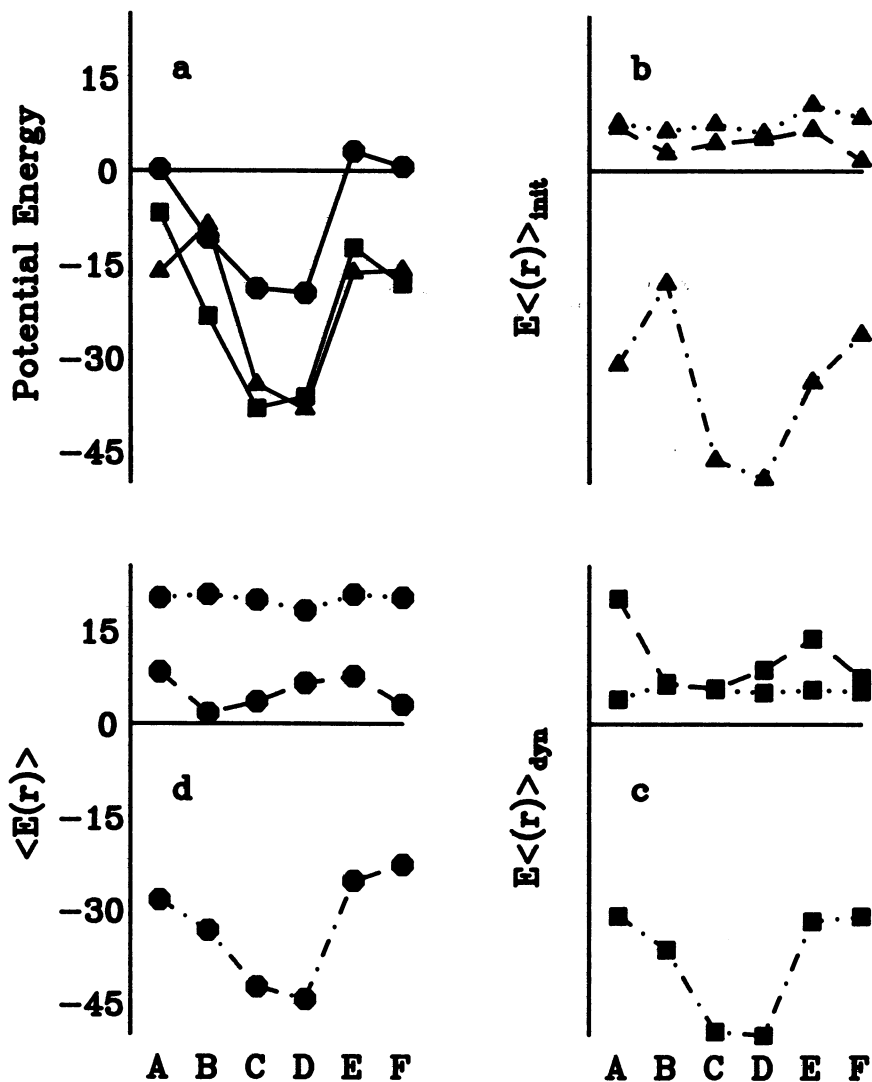
Figure 1. Site energies (kcal/mole) of (GlcNAc)₆ bound to lysozyme. The total potential energy per site for residues A through F of (GlcNAc)₆ was calculated for the initial structure from X-ray and model building, $E(<r>)_{init}$, (▲) for the average dynamic structure, $E(<r>)_{dyn}$ (■), and the energy averaged over 55 ps of dynamics structures, $<E(r)>$ (●) (a). The contributions from configurational terms (bonds, angles, dihedrals) (dot), intramolecular (dash) and intermolecular (dot-dash) nonbond terms (van der Waals, electrostatic, hydrogen bond) to $E(<r>)_{init}$, $E(<r>)_{dyn}$ and $<E(r)>$ are plotted in panel b, c and d, respectively.

GlcNAc is one functionality which does confer some specificity, and this sidechain makes one of the two important hydrogen bonds for this residue in the simulation. During catalysis, the small intermolecular energy at site E could enhance product release of the E,F dimer moiety, which is known to be fast relative to release of the tetramer A-D (21).

Table I. Energy of GlcNAc Residues in the Lysozyme-Substrate Complex

Site	Configurational[a]	Nonbond[b]		Total
		Intermolecular	Intramolecular	
	optimized initial coordinates, $E(<r>)_{init}$			
A	7.8	-31.0	7.0	-16.2
B	6.4	-18.0	2.9	-8.7
C	7.6	-46.4	4.5	-34.3
D	6.1	-49.4	5.2	-38.1
E	10.7	-33.8	6.7	-16.4
F	8.6	-26.2	1.7	-15.9
	individual dynamics coordinates, $<E(r)>$			
A	20.3	-28.4	8.4	0.3
B	20.8	-33.3	1.7	-10.8
C	19.9	-42.4	3.6	-18.9
D	18.2	-44.4	6.6	-19.6
E	20.8	-25.3	7.7	3.2
F	20.3	-22.7	3.1	0.7
	optimized dynamics coordinates, $E(<r>)_{dyn}$			
A	4.0	-30.9	20.2	-6.7
B	6.7	-36.3	6.3	-23.3
C	5.6	-49.5	5.8	-38.1
D	5.1	-50.1	8.8	-36.2
E	5.6	-31.7	13.8	-12.3
F	5.3	-30.9	7.5	-18.1

[a]Sum of bond, angle and dihedral angle energy terms.
[b]Sum of van der Waals, electrostatic and hydrogen bond energy terms.

Substrate-Enzyme Fluctuation Correlations

Normalized cross-correlations in the atomic fluctuations between substrate and lysozyme atoms were calculated from

$$\chi = \frac{< \Delta r_i \cdot \Delta r_j >}{< \Delta r_i^2 >^{1/2} < \Delta r_j^2 >^{1/2}}$$

where Δr stands for $(r - <r>)$. Cross-correlation coefficients are a measure of the interdependence of the motions of atoms, and can reveal interactions which are not apparent from the static picture provided by an averaged structure. The residues with atoms having motions correlated with a substrate atom are shown in Figure 2. Correlations with values greater than 0.3 are shown in thick lines. As is evident from the figure, correlated motions of the substrate and enzyme involve protein atoms in the binding cleft and primarily on the right-side as shown in the lower view of Figure 2. There is no correspondence between high correlation in fluctuations and hydrogen bond; i.e., not all residues which hydrogen bond to (GlcNAc)$_6$ have high correlation coefficients, and some highly

correlated residues do not hydrogen bond to $(GlcNAc)_6$. Although it has been demonstrated that solvent alters correlations calculated from MD trajectories (22-25), the results for lysozyme·$(GlcNAc)_6$ should not be greatly affected by the lack of solvent because there are no waters mediating the intermolecular contacts.

Hydrogen bond of Glu 35

The nature of the interactions of the catalytically essential residue Glu 35 (13,14) was investigated by an analysis of the sidechain hydrogen bond. One-ps average structures were used to find the time dependence of the energy for the hydrogen bonds involving the two terminal carboxylate atoms $H_{\varepsilon 1}$ and $O_{\varepsilon 2}$. The hydrogen bond potential and parameters include both a radial dependence between the acceptor and donor as well as angular dependences as previously described (1).

$$E_{HBond} = \left(\frac{A}{r_{ad}^6} - \frac{B}{r_{ad}^4} \right) \left(\cos^2_{aa-a-H} \theta \right) \left(\cos^2_{a-H-d} \theta' \right)$$

The hydrogen bonds of the Glu 35 sidechain are shown in Figure 3. Two aspects of the time dependence of the hydrogen bond are of interest. In the case of the protonated carboxylic oxygen, motion of the hydroxyl proton allows two hydrogen bond acceptors such that $H_{\varepsilon 1}$ switches between the endocyclic oxygen O_5 and the hydroxymethyl oxygen O_6 of the residue in site D. A hydrogen bond is always present involving one or the other acceptor (Figure 4a). The second pattern is from the unprotonated oxygen, $O_{\varepsilon 2}$, and is a bifurcated hydrogen bond in which the oxygen simultaneously interacts with two donors, the mainchain amide H of lysozyme residues 109 and 110 (Figure 4b). Some modulation of the $O_{\varepsilon 2}$ hydrogen bonds can be seen; the bond to 110 is stronger initially and then the two energies become nearly equal from 15 to 35 ps, possibly correlated with a switch in the $H_{\varepsilon 1}$ bond toward O_6.

    The three hydrogen bonds to Glu 35 stabilize the observed conformation in which $H_{\varepsilon 1}$ interacts with the endocyclic oxygen and has little contact with the glycosidic oxygen, O_4'.

Enhancement of a Substrate Conformation Optimum for Catalysis

In the bound state, the conformation of the carboxymethyl sidechain in site D and the glycosidic dihedral angles linking residues in sites D and E differ from those of the other $(GlcNAc)_6$ residues and of cellulose (26). As illustrated with the residues from sites A and B in Figure 5, the cellulose-like conformation involves interresidue hydrogen bonds (27) between the endocyclic oxygen O_5 of one residue and H_3 of the next residue, and between H_6 and O_3 of the same two residues, respectively. These hydrogen bonds produce a helical twist to the chain such that the value for the glycosidic dihedral ϕ ($O_5-C_1-O_4'-C_4'$) is $-86°$, similar to that of cellulose, $-98°$ (26). The unbound state of $(GlcNAc)_6$ would be presumed to have the cellulose-like dihedral angles. While the linkages between GlcNAc residues in sites A, B, C and D maintain the cellulose conformation, the linkage between D and E differs in a way which promotes catalysis by Scheme I.
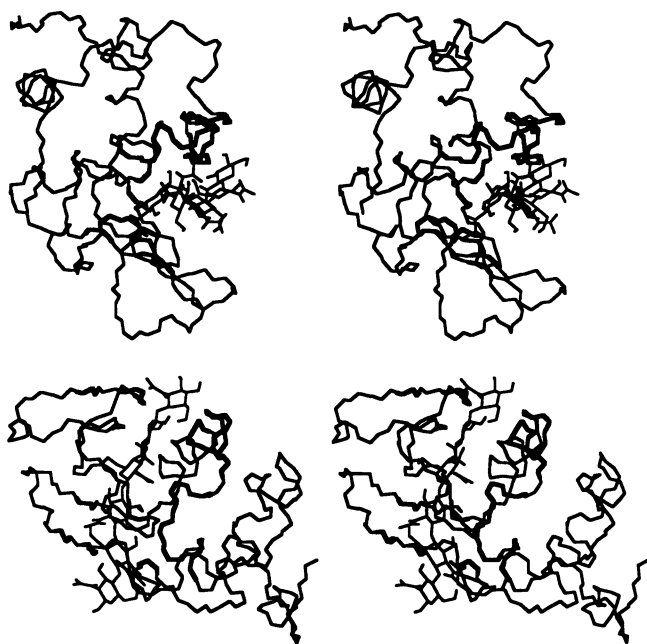
Figure 2. Cross-correlations in lysozyme and (GlcNAc)$_6$ fluctuations. Protein mainchain atoms drawn in thick lines correspond to positive correlations > 0.3. These lysozyme residues are 35, 42, 44, 52, 57, 63, 73-76, 94-104, 106-110, 112 and 113. All (GlcNAc)$_6$ atoms are drawn in thin lines. The two stereo views are related by two approximately 90° rotations.
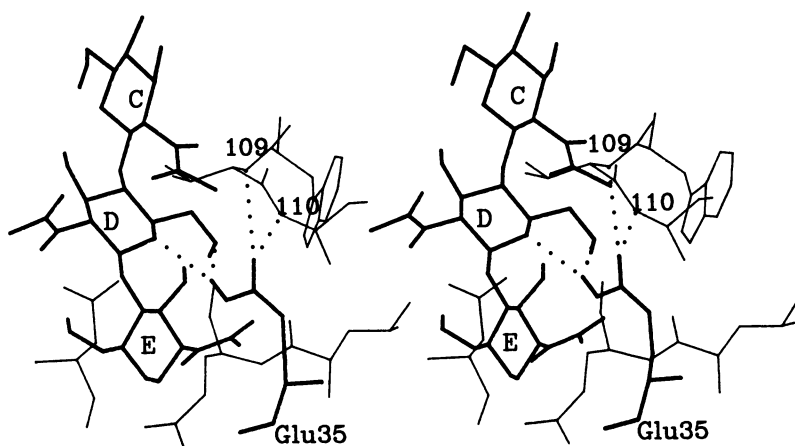
Figure 3. Stereo view of the active site cleft of lysozyme near site D. Hydrogen bonds of the Glu 35 sidechain are shown in dotted lines. The sidechain atom Hε_1 of Glu 35 is shown.
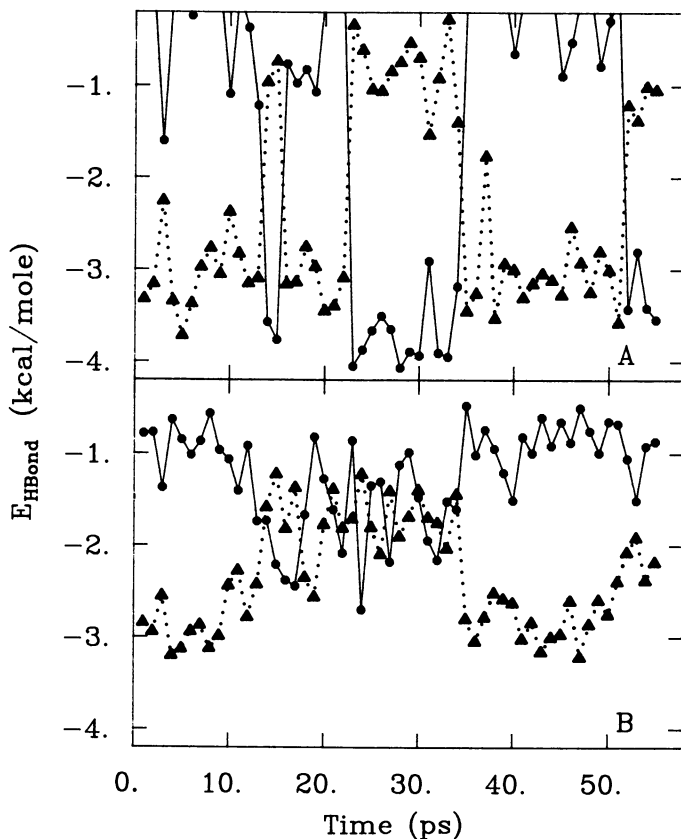
Figure 4. Time series of the hydrogen bond energy for the four hydrogen bonds of Glu 35:   (A) $H_{\varepsilon 1}$ to the ring O_5 (▲) and O_6 (●) of site D; (B) $O_{\varepsilon 2}$ to HN of residue 110 (▲) and 109 (●).
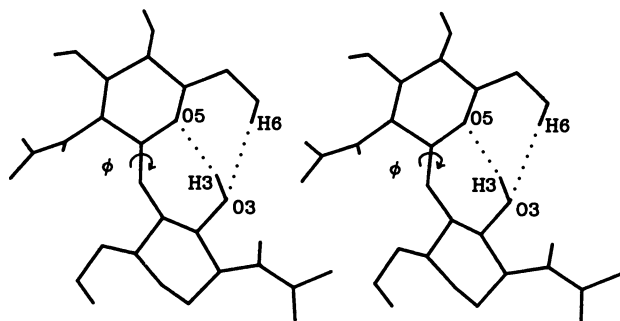
Figure 5. Intersaccharide hydrogen bonds (dotted lines) and the glycosidic angle ϕ (O_5-C_1-O_4'-C_4') between site A and B. These hydrogen bonds and ϕ (= -86°) are similar to those found for cellulose.

        By the alternative mechanism, Glu 35 protonates the ring oxygen
of D. Furthermore, optimum stereoelectric assistance for ring
opening is achieved when ϕ is -60°. These two aspects are not
consistent with a cellulose-like conformation: (i) the
intersaccharide hydrogen bond of H_6 to O_3 would sterically interfere
with Glu 35 protonation of the ring oxygen and (ii) ϕ would differ
by 38°. (For the other intersaccharide linkages, the $H_6 \rightarrow O_3$
hydrogen bond does exist and ϕ varies from approximately -86 to
-75°.)
        Binding to lysozyme stabilizes a conformation in site D
compatible with the alternative mechanism.   (i) H_6 of residue D does
not interact with residue E but forms a strong hydrogen bond to the
mainchain O of residue 57, a residue involved with the unusual
buried β turn in lysozyme.   (ii) The value of ϕ between D and E,
which was -54° in the initial structure, stabilized at -62°, near
the optimum of -60° for stereoelectronic assistance.

Conclusions

MD simulations can aid in the understanding of enzymic reactions by
providing new insights into the structures and intermolecular
interactions fundamental to the chemical catalysis. By studying the
structures from the simulation of the lysozyme-(GlcNAc)$_6$ complex, we
have proposed an alternative to the accepted mechanism which
accounts for the available experimental observations. The proposal
of this lysozyme mechanism illustrates one way in which simulations
can serve to generate new ideas which can be explored by experiment
and computation.
        The information obtained from the simulation included an
explanation from the energetics for the lack of specificity in site
E and a possible driving force for product release, as well as a
description of how the hydrogen bond interactions and glycosidic
dihedral angle of the GlcNAc residue in site D could promote
catalysis via the alternative mechanism. Since the alternative
mechanism was suggested by examining the results of the simulation,
there was no bias in the initial model building of the substrate
(see above). As such, the support of the alternative mechanism by
the nature of the hydrogen bond pairs, the relative site energies
and the glycosidic ϕ angle is a consequence of the simulation. The
dynamics also improved the initial model built complex in that the
substrate-enzyme interaction energy was lowered in site B.
Additional studies of the enzyme-substrate complementarity in other
complexes along the reaction path are under way. Since the initial
report of an alternative pathway for lysozyme hydrolysis (8,9,28)
work on the solution hydrolysis of glucosides has demonstrated the
existence of a ring opening mechanism (29,30).
        It is hoped that the analysis and results reported here will
stimulate new experiments on the lysozyme mechanism.

Literature Cited

  1. Post, C. B.; Brooks, B. R.; Karplus, M.; Dobson, C. M.;
      Artymiuk, P. J.; Cheetham, J. C.; Phillips, D. C. J. Mol. Biol.
      1986, 190, 455.
  2. Blake, C. C. F.; Koenig, D. F.; Mair, G. A.; North, A. C. T.;.
      Phillips, D. C.; Sarma, V. R. Nature 1965, 206, 757.
  3. Imoto, T.; Johnson, L. M.; North, A. C. T.; Phillips, D. C.;
      Rupley, J. A. In The Enzymes; Boyer, P. D., Ed.; Academic: New
      York, 1972; p 665.

4. Smith, L. E. H.; Mohr, L. H.; Raftery, M. A. J. Am. Chem. Soc.
    1973, 95, 7497.
5. Rosenberg, S.; Kirsch, J. F. Biochemistry 1981, 20, 3196.
6. Pollock, J. J.; Sharon, N. Biochemistry 1970, 9, 3913.
7. Rupley, J. A.; Gates, V. Proc. Natl. Acad. Sci. 1967, 57, 496.
8. Post, C. B.; Karplus, M. In Mechanisms of Enzymatic Reactions:
    Stereochemistry, Steenbock Symp. 1985; Frey, P. A., Ed.;
    Elsevier: New York, 1985; p 345.
9. Post, C. B.; Karplus, M. J. Am. Chem. Soc. 1986, 108, 1317.
10. Blake, C. C. F.; Mair, G. A.; North, A. C. T.; Phillips, D. C.;
    Sarma, V. R. Proc. Roy. Soc. London, Series No. B167; 1967;
    p 365.
11. Vernon, C. A. Proc. Roy. Soc. London, Series No. B67; 1967;
    p 378.
12. Walsh, C. Enzymatic Reaction Mechanisms, 1979; W. H. Freeman:
    San Francisco.
13. Malcolm, B. A.; Rosenberg, S.; Corey, M. J.; Allen, J. S.;
    Baetselier, A.; Kirsch, J. F. Proc. Natl. Acad. Sci. 1989, 86,
    133.
14. Kuroki, R.; Yamada, H.; Moriyama, T.; Imoto, T. J. Biol. Chem.
    1986 261, 13571.
15. Gorenstein, D. G.; Findley, J. N.; Luxon, B. A.; Kar, D. J. Am.
    Chem. Soc. 1977, 99, 3473.
16. Kirby, A. Acc. Chem. Res. 1984, 17, 305.
17. Kirby, A. CRC Crit. Rev. Biochem. 1987, 22, 282.
18. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.;
    Swaminathan, S.; Karplus, M. J. Comput. Chem. 1983, 4, 187.
19. Dobson, C. M.; Karplus, M. Meth. Enzym. 1986, 131, 362.
20. The experimental data which have been used to suggest ring
    distortion in site D may not be relevant. The interpretation of
    data on the energetics of binding is complicated when it is
    unknown what group on the ring gives rise to poor contacts
    (Schindler, M.; Assaf, Y.; Sharon, N.; Chipman, D. M.
    Biochemistry 1977, 16, 423), whether the bound conformation is
    analogous to that of the natural substrate, and what is the
    contribution from the release of bound water (9). In addition,
    structural results on complexes with synthetic substrates are
    not important if the complex is not homologous to a reaction
    intermediate (Ford, L. O.; Johnson, L. N.; Machin, P. A.;
    Phillips, D. C.; Tjian, R. J. Mol. Biol. 1974, 88, 349).
21. Chipman, D. M.; Pollock, J. J.; Sharon, N. J. Biol. Chem. 1968,
    243, 487.
22. Ahlstrom, P.; Teleman, O.; Jonsson, B.; Forsen, S. J. Am. Chem.
    Soc. 1987, 109, 1541.
23. Brooks, C. L., III; Karplus, M.; Pettitt, B. M. Adv. Chem. Phys.
    1988, 71, p 1.
24. Chen, L. X. Q.; Engh, R. A.; Brunger, A. T.; Nguyen, D. T.;
    Karplus, M.; Fleming, G. R. Biochemistry 1988, 27, 6908.
25. van Gunsteren, W. F.; Karplus, M. Biochemistry 1982, 21, 2259.
26. Gardner, K. H.; Blackwell, J. Biopolymers 1974, 13, 1975.
27. Ham, J. T.; Williams, D. G. Acta Crystallogr. 1977, B26, 1373.
28. Fleet, G. W. J. Tetrahedron Lett. 1985, 26, 5073.
29. Guindon, Y.; Anderson, P. C. Tetrahedron Lett. 1987, 28, 2485.
30. Guindon, Y.; Bernstein, M. A.; Anderson, P. C. Tetrahedron Lett.
    1987, 28, 2225.

Author Index

Affiliation Index

Subject Index