NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

A COMBINED MOLECULAR DYNAMICS AND NMR SPECTROSCOPIC PROTOCOL FOR THE CONFORMATIONAL ANALYSIS OF OLIGOSACCHARIDES

by

Vikram Varma

B.Sc. (Hons.), Simon Fraser University (Canada), 1987

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department of Chemistry

Simon Fraser University

July 1993

ISBN 0-315-91327-4

APPROVAL

Name: Vikram Varma

Degree: Ph.D.

Title of Thesis: A Combined Molecular Dynamics and NMR Spectroscopic Protocol for the Conformational Analysis of Oligosaccharides

Examining Committee:

Chair: Dr. D. Sutton

_____

Dr. B.M. Pinto (Associate Professor)
Senior Supervisor

_____

Dr. I.D. Gay (Professor)
Committee Member

_____

Dr. A.S. Tracey (Supervisor NMR Services)
Committee Member

_____

Internal Examiner: Dr. S. Wolfe (Professor)

_____

External Examiner: Dr. J.P. Carver(Professor,U.Toronto)

Date Approved: August 31, 1993

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay:

A COMBINED MOLECULAR DYNAMICS AND NMR SPECTROSCOPIC PROTOCOL FOR THE CONFORMATIONAL ANALYSIS OF OLIGOSACCHARIDES

Author:

(signature)

VIKRAM VARMA

(name)

10/8/93

(date)

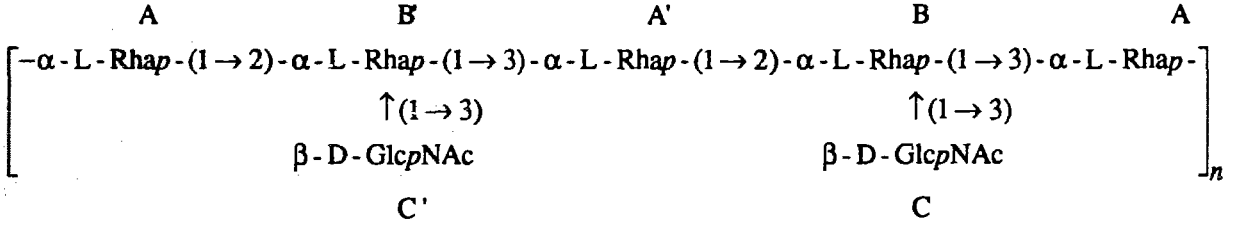
Abstract

A combined experimental and theoretical protocol for the conformational analysis of oligosaccharides is presented.

Three disaccharides, methyl α-D-mannopyranosyl-(1→3)-α-D-mannopyranoside, methyl β-D-galactopyranosyl-(1→4)-β-D-glucopyranoside, and propyl β-D-2-acetamido-2-deoxy glucopyranosyl-(1→3)-α-L-rhamnopyranoside, are used to evaluate a protocol for conformational analysis that makes use of molecular dynamics calculations with the CHARMM force field. Dynamics trajectories computed in vacuo and in water are used to calculate time-averaged NMR parameters such as spin-lattice relaxation times (T_1), Nuclear Overhauser Enhancements (NOE), and heteronuclear spin-spin coupling constants ($^3J_{CH}$). The calculated NMR parameters are then compared to experimental values and used to evaluate the computational procedure. The energetically accessible conformations are effectively sampled by the simulations.

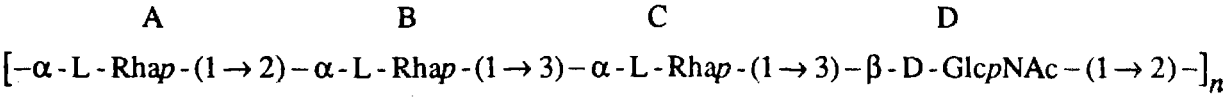
The method has been extended to the conformational analysis of higher-order oligosaccharides corresponding to the cell-wall polysaccharide of the *Streptococcus* Group A, and the *Shigella flexneri* Y O-antigen.

The *Streptococcus* Group A cell-wall polysaccharide is comprised of a backbone of rhamnopyranosyl units connected by alternating α-L-(1→3) and α-L-(1→2) linkages, to which are attached N-acetyl-β-D-glucosamine (β-D-Glc*p*NAc) residues at the 3 positions of the rhamnose backbone.



A branched trisaccharide [A'-(C)B], a tetrasaccharide [A'-(C)B-A], a pentasaccharide [C'-B'-A'-(C)B], and two hexasaccharides [C'-B'-A'-(C)B-A] and [A-(C')B'-A'-(C)B], have been chosen for study.

The *Shigella flexneri* Y O-antigen is a linear polysaccharide that is composed of rhamnose units linked α-L-(1→3) and α-L-(1→2), interspersed by N-acetyl-β-D-glucosamine (β-D-Glc*p*NAc) to form a periodic repeating unit ABCD.



A heptasaccharide corresponding to the fragment [ABCDA'B'C'] of the *Shigella flexneri* Y polysaccharide has been investigated.

The conformational properties of all of the oligosaccharides have been studied using molecular dynamics simulations. Interproton distances derived from ROESY spectra are used to determine the starting conformations of the oligosaccharides used in the dynamics calculations, and dynamics simulations are computed with proton pairs constrained to the ROESY-derived distances, as well as with the constraints removed. These dynamics trajectories are used to calculate ROESY buildup curves with CROSREL, a program that treats cross relaxation by means of a full matrix relaxation approach. The calculated buildup curves compare favorably with the experimental buildup curves.

The study demonstrates that molecular dynamics, in conjunction with NMR spectroscopy, can be a useful tool in the understanding of the conformational behavior of oligosaccharides in solution.

The results provide a model for antigen topology that can be used to infer some of the critical features of antibody-antigen interactions.

To the memory of my father, and to my family, who have sacrificed so much for my education. I hope its worth it.

Acknowledgments

Thanks to

BMP, for providing the environment, the facilities, and the motivation to carry out this research.

NMR: Alan and Marcey for everything I know about NMR, Kerry for providing a standard to shoot for, and Ian Gay for all the help on basic NMR and density matrix theory.

UNIX jocks: Chan and Noham for their combined wisdom and their willingness to share it, Tony Schoemaker (SGI) for all the cool tips, Uwe and Chris for their collaboration, and Steve Kloster for his help.

Lab Rats: Kerry, Arlette, Mariño for synthesizing all the compounds, and the entire group for just being there.

Miscellaneous: Gabe Kalmar for the loan of his computer to write a large part of this thesis and to the coffee crowd for all the wonderful bitch sessions.

And my "**wife**", who worked as hard as I did during the writing of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER II

LIST OF FIGURES

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

LIST OF ABBREVIATIONS AND SYMBOLS

2D	2 dimensional
3D	3 dimensional
Å	Ångstrom, unit 10^{-8} cm
Hz	Hertz, $sec.^{-1}$
δ	chemical shift
$\Delta\delta$	chemical shift difference
γ	Proton gyromagnetic ratio
$\hbar$	Planks constant $/2\pi$
∇	Legendrian operator
W_0	Zero quantum transition rate
W_1	Single quantum transition rate
W_2	Double quantum transition rate
ν	frequency in cm^{-1}
ρ	dipole dipole relaxation rate
σ	cross relaxation rate
τ_c	Correlation time
Φ	the angle defined as the H1-C1-O1-Cn torsion angle where Cn is the aglyconic carbon involved in the glycosidic bond
Ψ	the angle defined as the C1-O1-Cn-On torsion angle where Cn and Hn are the aglyconic carbon and proton involved in the glycosidic bond
θ	angle is defined as the H2-C2-N-H angle in Glc*p*NAc
ω	the angle defined as the torsion O5-C5-C6-O6.
ω_0	Spectrometer frequency

AMBER	Assisted Model Building and Energy Refinement
CD	cirular dichroism
CHARMM	Chemistry at HARvard Molecular Mechanics
DMM	dimethoxymethane (1,1, dimethoxymethane)
FT	Fourier transform
Gal*p*NAc	2-acetamido-2-deoxygalactopyranose
Glc*p*NAc	2-acetamido-2-deoxyglucopyranose
GESA	Geometry of Saccharides
gt	one of the three staggered rotamers of the ω angle with ω = 60º
gg	one of the three staggered rotamers of the ω angle with ω =-60º
HSEA	Hard Sphere Exoanomeric Effect
J	spin spin coupling
LPS	lipopolysaccharide
Man*p*	Mannopyranose
Me	methyl
MM2	Molecular Mechanics 2
MMC	Metropolis Monte Carlo method
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Enhancement
NOESY	Nuclear Overhauser Enhancement Spectroscopy
OMe	methoxy
ORD	Optical Rotatory Dispersion
QSAR	Quantitative Structural Activity Relationship
QSPR	Quantitative Structural Property Relationship
R_L	the leakage rate
Rha*p*	rhamnopyranose
ROESY	Rotating Frame Overhauser Enhancement Spectroscopy

ROE	Rotating Frame Overhauser Enhancement
T_1s	spin-lattice relaxation times
tg	one of the three staggered rotamers of the ω angle with ω =-180º
THP	2 methoxy tetrahydropyran.
Xyl	Xylose

Chapter I
INTRODUCTION

1.1. Cell-Surface Carbohydrates

Carbohydrates are an important class of biological macromolecules that display a diversity of functions ranging from energy metabolism to more complex processes such as targeting of lysosomal enzymes, mediation of protein folding inside the cell and providing specific carbohydrate protein interactions that dictate cellular recognition processes extracellularly [1-3]. The last of these includes biologically important functions such as viral infection of cells, cancer metastasis, cellular response to disease and inflammation, and cell differentiation, all of which are mediated by cell surface carbohydrates.

In prokaryotes, carbohydrates play a large role in the cell wall. Besides the peptidoglycan, which is composed of a linear polysaccharide cross-linked by peptides and serves to provide rigidity to the cell, gram negative bacteria also produce lipopolysaccharides (LPS) which extend outward from the cell wall. Bacterial lipopolysaccharides are complex molecules, composed of three regions[4]. The lipid portion (Lipid A region) is attached to an R core composed of unusual carbohydrates and an O side chain that is the polysaccharide portion, containing a 3-5 sugar repeating unit. The O side chain has different lengths, and along with a portion of the R core extends outward from the outer membrane. It is the O side chain that is one of the antigenic determinants of the bacteria.

The lipid bilayer of a eukaryotic cell membrane is impregnated by glycoproteins and glycolipids. Glycoproteins are classified into two groups, the N-linked and O-linked glycoproteins, by their mode of attachment. The N-linked glycoproteins have a β-N-link between 2-acetamido-2-deoxy glucose (Glc*p*NAc) and the amide side chain of

asparagine. The O-linked glycoproteins are α or β linked to a serine or threonine, generally through 2-acetamido-2-deoxy galactose (Gal*p*NAc) or Xylose (Xyl). Glycolipids such as phospatidyl inositol, cerebrosides and glycosphingolipids are also present. The carbohydrate part of these molecules extends into the extracellular space[5,6] and the cell surface is covered, therefore, by complex oligosaccharides.

1.1.1. Biological Functions

Cell surface carbohydrates have been implicated in many biological recognition processes. In their function as cell surface receptors they serve as ligands for hormones, antibodies, toxins, cell attachment proteins of viruses and a particular group of carbohydrate binding proteins known as lectins[7-11]. Extracellular carbohydrates also modulate cell-cell interactions and are now recognized as important participants in cell adhesion processes. The discovery of a class of cell adhesion molecules[12,13] that specifically target leukocytes to regions of inflammation, named selectins[14-16] due to their lectin-like N-terminal domains, has led to the identification of α-(1→3) fucosylated derivatives of polylactosamine ([-β-D-Glc*p*NAc-(1→4)-β-D-Gal*p*-]$_n$) as their ligands[17-21]. Carbohydrates have also been found to play a major role in sperm adhesion to the egg cell[22] and to inhibit cell adhesion in developing cells [1,23]. It is, however, the antigenicity of carbohydrates that has been the focus of much research over the last decade. Although it is still difficult to determine whether the antigenic determinant of a glycoprotein is the oligosaccharide or the peptide fragments, the finding that in many cases it is indeed the oligosaccharide portion has moved carbohydrate research to the forefront.

One facet of oligosaccharide research is the study of conformational properties and its relationship to the variety of interactions mediated by oligosaccharides. Conformational changes in proteins that result from their glycosylation could perhaps

explain the difference in function between glycosylated and non glycosylated peptides. Oligosaccharides display considerable antigenic diversity that most probably originates from the different surfaces that each saccharide presents to the binding lectin or antibody. Insight into the structure function relationships of carbohydrates will be gained by knowledge of their three-dimensional structure and molecular surface.

1.2. Conformational Analysis

1.2.1. Rationale

The conformation of an oligosaccharide determines its molecular surface. Conformational analysis can be used to explain the differences in binding between structurally similar compounds as well as a predictive tool to produce inhibitors and drugs with increased binding capabilities. By analysis of the interactions between the bound ligand and protein it can be determined which groups can be substituted and the manner in which to produce a better 'fit'. In the absence of the protein, analysis of substrate analogues, transition state analogues and inhibitors can also be used to generate information about the protein binding site.

1.2.1.1. Substrate Analogues

Many examples of the successes of conformational analysis can be found in the literature. Analysis of the binding sites of various antibodies and lectins has led to the synthesis of deoxy analogues of the natural substrates. These synthetic analogues exhibit increased affinities for the binding proteins [24]. On a more commercial level, studies on sucrose[25] and the sweetness receptor [26] have shown that the spatial relationship between a hydrogen bond donor and acceptor on the substrate is critical to its relative affinity. An increase in the hydrophobicity of one face of the substrate also leads to compounds with greater sweetness. A recent product of such analyses is Sucralose, the 4,1',6'-trichloro-

4,1',6'-deoxy-galacto-sucrose, which shows 650 times the sweetness of sucrose and is used as an artificial sweetener registered in Canada (Splenda®) [27].

1.2.1.2. Rational Drug Design

Rational drug design is another important application of such studies. Analysis of the charge distribution and conformation of naturally occurring, biologically active compounds has led to whole classes of drugs, antibiotics and inhibitors[28]. A recent development in this field is the formation of 3D databases of molecular structures. Quantitative Structural Activity Relationship (QSAR) and Quantitative Structural Property Relationship (QSPR) databases exist to locate compounds with similar biological and physical properties[29]. QSAR databases can be searched for molecules with common surface topology and charge distribution to known pharmacomimetics. Compounds thus located can then be synthesized and tested for biological activity.

1.2.1.3. Protein Design

Analysis of the transition state geometries has led to innovative, new technologies like catalytic antibodies [30]. Pauling had suggested that enzymes catalyze reactions by stabilization of the transition state [31], hence lowering the activation energy of the reaction and increasing the rate. In direct proof of this theory, Shultz [32] and Lerner [33] raised antibodies to negatively charged tetrahedral, transition state analogues of carbonate and ester hydrolysis, showing that these proteins could catalyze ester hydrolysis.

These examples are indicative of the power and utility of conformational analysis. Currently there are several tools available to aid in determination of molecular structure.

1.3 Methods of Conformational Analysis

Physical methods used in early studies of conformational analysis of oligosaccharides include circular dichroism (CD), optical rotatatory dispersion (ORD), and X-ray and neutron diffraction. While CD and ORD methods have been used to measure the helical content of polysaccharides[34], no quantitative data about the orientation of the individual monosaccharide rings with relation to one another could be obtained. Recently, however, a semiempirical theory of optical rotation of saccharides has been developed[35], and it is now possible to calculate the solution rotation of a saccharide. The method has been tested on a number of mono- and disaccharides[36-40], but has yet to be extended to larger oligosaccharides.

1.3.1. Diffraction (X-ray and Neutron)

X-ray and neutron diffraction, on the other hand, can provide information on the scale of Angstroms about the geometry of any crystalline compound. There are several examples of crystal structures of oligosaccharides, ranging from monosaccharides to polysaccharides. In addition, oligosaccharides have been cocrystallized with their associated binding proteins[41] or antibody fragments[42,43]. These structures have provided substantial information about the molecular interactions that occur in protein-carbohydrate binding. The crystal structures of the L-arabinose binding protein and D-galactose binding protein[41] indicate that H-bond formation is a primary impetus of the protein specificity and binding. A recurrent motif in carbohydrate-protein interactions is the "stacking" of aromatic residues of the protein on the monosaccharide rings, thus providing a stabilizing hydrophobic interaction. It is thought that these interactions also provide a steric basis for binding specificity by acting as a pocket for the pyranose rings of the saccharide.

Diffraction techniques suffer from two setbacks.

1) Crystals have to be grown and it is often difficult to crystallize higher order oligosaccharides

2) Crystal packing forces may play a role in crystal formation. Thus the conformation present in the crystal may not be representative of that present in biological systems[44].

These drawbacks make it difficult, and in some instances undesirable to use diffraction.

A method that can be used routinely, with little sample preparation, total sample recovery, and with rapid results is high resolution Fourier transform Nuclear Magnetic Resonance[45,46] (NMR) Spectroscopy. NMR spectroscopy can be used to measure conformationally dependent parameters such as chemical shift, spin spin coupling (J), spin-lattice relaxation times (T_1s) and Nuclear Overhauser Enhancements (NOEs), and information thus derived can be used to construct a model structure.

1.3.2. Nuclear Magnetic Resonance Spectroscopy

Magnetically active nuclei can be considered to be atomic dipoles that, when placed in a magnetic field, align themselves with and against the field. The two states induced by the field are commonly termed the α and β states. Transitions between the low energy β state and the α state can be brought about by application of a radio frequency pulse of a frequency exactly equal to the difference in energy between the two states. The familiar NMR spectrum is a plot of these frequencies relative to an arbitrary standard value.

1.3.2.1. Chemical Shift (δ)

Fortuitously, protons, because of their large gyromagnetic ratio and high natural abundance (99.985%), have the best sensitivity of magnetically active nuclei that are

found in common organic compounds. In addition, they are extremely sensitive to their environment. The utility of nuclear magnetic resonance in chemistry is that nuclei absorb different frequencies in response to their differing environment[47,48]. This dependence on environment is exploited through chemical shifts, which are characteristic for protons belonging to, or next to, different functional groups. Compounds can thus be deduced from their NMR spectrum.

Although chemical shifts cannot be used quantitatively, they can be utilized to approximate conformation. Proximity to electron withdrawing groups and π bonded systems have marked effects on the chemical shifts, and thus chemical shift information can be used to derive information on molecular structure.

1.3.2.2. Scalar Coupling (J)

Scalar coupling, visible as splitting of resonances in the NMR spectrum can also be used to predict conformation. The vicinal coupling constant $^3J_{HH}$ through carbon has been shown to exhibit a dependence on the torsion angle ϕ, which can be described by the Karplus equation[49]

$$J=A + B\cos^2(\phi) + C\cos(\phi) \qquad (1.1)$$

In addition, Karplus type equations have been proposed for different types of coupling, e.g., $^3J_{CCCH}$, $^3J_{CSCH}$ and $^3J_{COCH}$ [50-53]. The latter, i.e., $^3J_{COCH}$, is of interest in the conformational analysis of carbohydrates since both glycosidic torsion angles are defined by a C-O-C-H torsion (see Figure 2.2). Measurement of these coupling constants at natural abundance is now possible with inverse detected experiments[54]. Two independent studies on carbohydrates with fixed geometries and correlation with crystal structures show that the magnitude of the carbon-proton coupling constant $^3J_{COCH}$ can also be described by[52,53]

$$J=A + B \cos^2(\phi) + C \cos(\phi) \qquad (1.2)$$

The values of the constants A, B and C were slightly different for the two studies.

Average values of $^3J_{HH}$ are observed for protons attached to freely rotating carbons or in cyclic compounds that rapidly interconvert between conformations. Coupling constants are also sensitive to bond length and bond angles and therefore, cannot be forwarded as absolute proof of the geometry of a molecule. Variations of the Karplus equation have been proposed[55] to account for deviations from tetrahedral geometry and the presence of electronegative substituents. The sinusoidal nature of the Karplus equation and its analogues introduces an uncertainty in the prediction of a

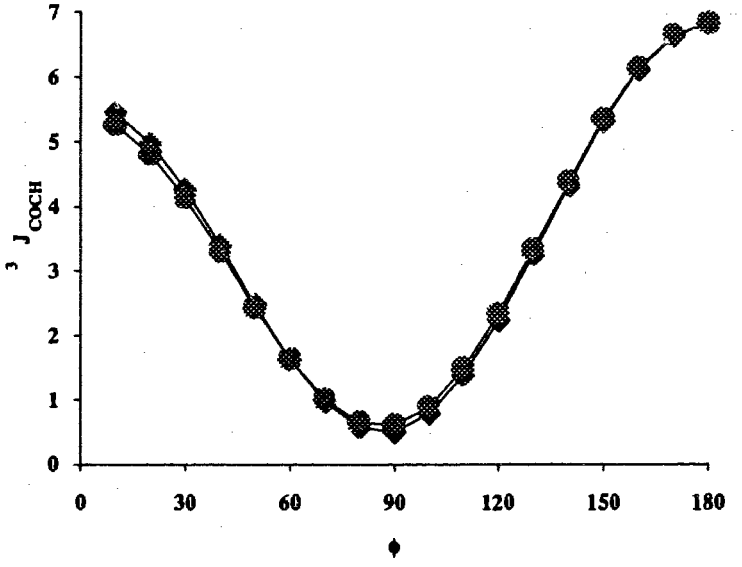


Figure 1.1. $^3J_{COCH}$ calculated from Karplus type equations.
—◆— $5.7\cos^2(\phi) - .6\cos(\phi) + .5$ (Ref. 34)
—●— $5.5\cos^2(\phi) - .7\cos(\phi) + .7$ (Ref. 35)

torsion angle from a measured coupling constant and a single J value can correspond to more than one torsion angle.

1.3.2.3. Spin Lattice Relaxation (T_1s)

In conjunction with information from the coupling constants and chemical shifts, T_1s and NOEs can be used to deduce the conformation of a molecule. With protons, both the T_1s and NOEs are modulated by dipolar interactions, and thus are highly dependent on the distance between nuclei. The spin lattice relaxation time is defined as the time constant with which the longitudinal magnetization will return to a thermal equilibrium[56]

$$\frac{dM_z}{dt} = -\frac{1}{T_1}\left(M_z(t) - M_z^o\right) \qquad (1.3)$$

where

$$\frac{1}{T_1} = \left(\frac{\mu_o}{4\pi}\right)\frac{\gamma^4\hbar^2}{10}\left[\tau_c + \frac{3\tau_c}{1 + (\omega_o\tau_c)^2} + \frac{6\tau_c}{1 + (2\omega_o\tau_c)^2}\right] K\sum_{i\neq j} r^{-6} + R_S \qquad (1.4)$$

and K is a constant that will be discussed later (see Equation 1.15).

The summation of distances in equation 1.4 is a major drawback of T_1s, i.e., that it is difficult to measure the contribution of each proton to the relaxation time of a particular nucleus. NOEs, on the other hand, allow some measure of assessment of the effect of an individual proton on another. This makes NOEs a very powerful tool for the determination of molecular structure[57].

1.3.2.4. The Nuclear Overhauser Effect

The Nuclear Overhauser effect[58-60] can simply be defined as the change in intensity observed in one resonance when another is perturbed. Since the observation by Bell and Saunders[61] that the intensities of a set of NOEs can be used to calculate relative inter proton distances, NOE spectroscopy has become a major tool in structural determination of organic compounds.

Perturbation of one resonance in a spin system can affect the intensity of another. This change in intensity is known as the Nuclear Overhauser effect, and is modulated by dipolar interactions between the nuclei. Consider a two spin system, AX. There are four states, $|\alpha\alpha\rangle$, $|\alpha\beta\rangle$, $|\beta\alpha\rangle$, $|\beta\beta\rangle$, as depicted in Figure 1.2.

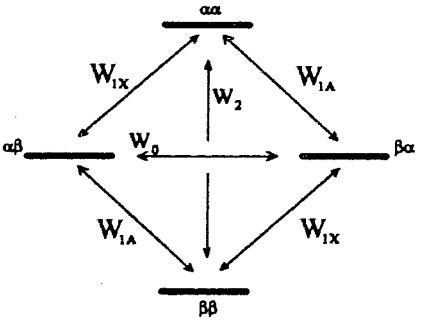


Figure 1.2. Energy levels in a two spin system.

If the rate of interconversion (**W**) between these states is defined as in Figure 1.2, and n^O and n are the equilibrium and non equilibrium poplulations of the the spin states, then the rate of change of population of these four states can be described by the equations[62]

$$\frac{dn_{\alpha\alpha}}{dt} = (W_{1A} + W_{1X} + W_2)(n^o_{\alpha\alpha} - n_{\alpha\alpha}) - (W_{1X})(n^o_{\alpha\beta} - n_{\alpha\beta}) - (W_{1A})(n^o_{\beta\alpha} - n_{\beta\alpha}) - (W_2)(n^o_{\beta\beta} - n_{\alpha\alpha}) \quad \textbf{(1.5)}$$

$$\frac{dn_{\alpha\beta}}{dt} = (W_{1A} + W_{1X} + W_o)(n^o_{\alpha\beta} - n_{\alpha\beta}) - (W_{1X})(n^o_{\alpha\alpha} - n_{\alpha\alpha}) - (W_{1A})(n^o_{\beta\beta} - n_{\beta\beta}) - (W_o)(n^o_{\beta\alpha} - n_{\beta\alpha}) \quad \textbf{(1.6)}$$

$$\frac{dn_{\beta\alpha}}{dt} = (W_{1A} + W_{1X} + W_o)(n^o_{\beta\alpha} - n_{\beta\alpha}) - (W_{1X})(n^o_{\beta\beta} - n_{\beta\beta}) - (W_{1A})(n^o_{\alpha\alpha} - n_{\alpha\alpha}) - (W_o)(n^o_{\alpha\beta} - n_{\alpha\beta}) \quad \textbf{(1.7)}$$

$$\frac{dn_{\beta\beta}}{dt} = (W_{1A} + W_{1X} + W_2)(n^o_{\beta\beta} - n_{\beta\beta}) - (W_{1X})(n^o_{\beta\alpha} - n_{\beta\alpha}) - (W_{1A})(n^o_{\alpha\beta} - n_{\alpha\beta}) - (W_2)(n^o_{\alpha\alpha} - n_{\alpha\alpha}) \quad \textbf{(1.8)}$$

The intensity of the A resonance is proportional to the quantity M_A where

$$M_A = (n_{\alpha\alpha} - n_{\beta\alpha}) + (n_{\alpha\beta} - n_{\beta\beta}) \quad \textbf{(1.9)}$$

and similarly, the intensity of the X resonance is proportional to M_X

$$M_X = (n_{\alpha\alpha} - n_{\alpha\beta}) + (n_{\beta\alpha} - n_{\beta\beta}) \qquad (1.10)$$

The equations (1.5-1.8) can combined with 1.9 and 1.10 and rewritten as

$$\begin{aligned} \frac{dM_A}{dt} &= \rho(M_A^o - M_A) + \sigma(M_X^o - M_X) \\ \frac{dM_X}{dt} &= \rho(M_X^o - M_X) + \sigma(M_A^o - M_A) \end{aligned} \qquad (1.11)$$

where it is assumed that $W_{1X} = W_{1A}$ and ρ, the direct dipole-dipole relaxation,

$$\rho = \frac{1}{T_1} = (2W_1 + W_2 + W_o) + R_s \qquad (1.12)$$

R_S is the leakage rate or contributions to relaxation from other mechanisms.
and σ, the cross relaxation,

$$\sigma = (W_2 - W_o) \qquad (1.13)$$

A. Relationship Between NOE, τ_c ω and r

Thus far, all that has been mentioned about ρ and σ is that

$$\rho = (2W_1 + W_2 + W_o) + R_S \quad \text{and} \quad \sigma = (W_2 - W_o)$$

W represents the transition rate induced by dipolar relaxation between the states. The rates W are in turn proportional to the spectral density function $J(\omega)$ which is given by

$$J(\omega) = \frac{2\tau_c}{1 + \omega^2 \tau_c^2} \qquad (1.14)$$

τ_c is the correlation time of the vector between the two dipoles and ω is the Larmor frequency of the dipoles. Therefore, the complete expressions for ρ and σ are

$$\rho = \frac{\gamma^4 \hbar^2}{20r^6}(J(\omega_o) + 3J(\omega_1) + 6J(\omega_2)) + R_S$$
$$\sigma = \frac{\gamma^4 \hbar^2}{20r^6}(6J(\omega_2) - J(\omega_o)) \qquad (1.15)$$

Since the NOE is a function of both ρ and σ it can be seen from the explicit expression for ρ and σ that the NOE has a dependence on the correlation time τ_c, the Larmor frequency ω, and the distance between the dipoles **r**. The dipolar coupling is proportional to r^{-6}, and hence an important aspect of the Nuclear Overhauser effect is that only nuclei that are spatially close will exhibit an effect. This relationship makes it possible to calculate distances between nuclei, making the NOE a powerful tool in conformational analysis. Traditionally, two types of NOE experiments are used, steady state[63] and transient NOE[62].

B. Steady State NOEs

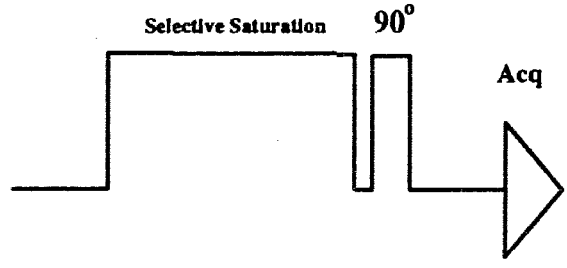


Figure 1.3. Pulse sequence of a steady state NOE experiment.

In a steady state experiment, a selective pulse saturates the proton of interest, during which time the NOE builds up. This is followed by a 90^O pulse and acquisition. Steady state experiments are usually run in the difference mode[64], which means that irradiation is done on alternate scans and by cycling the phase of the receiver, the irradiated spectra are subtracted from the spectra acquired without irradiation, yielding a

spectrum that is the difference of the two. The only signals that will be observed will be the irradiated proton and protons that show a change in intensity as a result of the irradiation. In this case, in equation 1.11

$$\frac{dM_A}{dt} = \frac{dM_X}{dt} = 0 \qquad \textbf{(1.16)}$$

If it is assumed that proton A is saturated, $M_A = 0$ and M_A^o is the equilibrium magnetization, and if these values are substituted in Equations 1.11, the steady state NOE can be written as

$$M_X = M_X^o + \frac{\sigma}{\rho} M_A^o$$

$$NOE = M_X - M_X^o = \frac{\sigma}{\rho} M_A^o \qquad \textbf{(1.17)}$$

In the extreme narrowing limit, i.e., $\omega_o \tau_c << 1$, the expression for ρ and σ reduce to

$$\rho = \frac{1}{2} \frac{\gamma^4 \hbar^2}{r^{-6}} \tau_c + R_S \text{ and } \sigma = \frac{1}{4} \frac{\gamma^4 \hbar^2}{r^{-6}} \tau_c \qquad \textbf{(1.18)}$$

and the NOE is then

$$NOE = \frac{1}{2 + \frac{1}{\tau_c \gamma^4 \hbar^2} R_S r^{-6}} M_A^o \qquad \textbf{(1.19)}$$

displaying the relationship between the NOE and the internuclear distance r

C. Transient NOEs

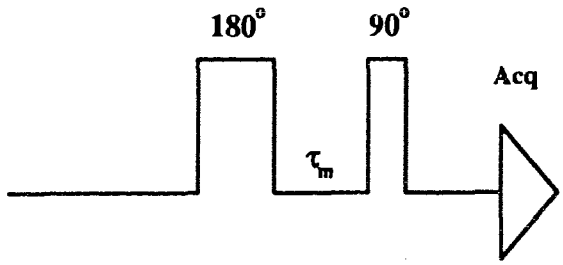


Figure 1.4. Pulse sequence of a transient NOE experiment.

The pulse sequence for a transient NOE experiment is shown in Figure 1.4. In a transient NOE experiment the first 180° pulse selectively inverts the magnetization of a particular proton, and the time dependence of the NOE to other protons is measured by varying the mixing time, τ_m.

A description of transient effects requires solution of the coupled set of differential equations (1.11). The general solution to this set is

$$\begin{aligned} M_A(t) &= c_{11}e^{-(\rho+\sigma)t} + c_{12}e^{-(\rho-\sigma)t} + k_1 \\ M_X(t) &= c_{21}e^{-(\rho+\sigma)t} + c_{22}e^{-(\rho-\sigma)t} + k_2 \end{aligned} \qquad \textbf{(1.20)}$$

The coefficients c_1, c_2 and k will vary for different boundary conditions and for a transient NOE experiment, the boundary conditions (using normalized intensities) are $M_A(\infty) = 1$ and $M_A(0) = 1$ and $M_X(\infty) = 1$ and $M_X(0) = -1$. The characteristic solution is then

$$\begin{aligned} M_A(t) &= -e^{-(\rho+\sigma)t} + e^{-(\rho-\sigma)t} + 1 \\ M_x(t) &= -e^{-(\rho+\sigma)t} - e^{-(\rho-\sigma)t} + 1 \end{aligned} \qquad \textbf{(1.21)}$$

The NOE at time t is expressed as

$$\begin{aligned} NOE(t) &= M_A(t) - M_A^0 \\ &= -e^{-(\rho+\sigma)t} + e^{-(\rho-\sigma)t} \end{aligned} \tag{1.22}$$

If the identity

$$e^{xt} = 1 + xt + (xt)^2 + (xt)^3 + \ldots$$

is used, at time $(\rho + \sigma)t << 1$ equation 1.22 reduces to

$$\begin{aligned} NOE(t) &= -(1-(\rho+\sigma)t) + (1-(\rho-\sigma)t) \\ &= -2\sigma t = -\left(\frac{1}{2}\frac{\gamma^4\hbar^2}{r^{-6}}\tau_c\right)t = R\frac{t}{r^{-6}} \end{aligned} \tag{1.23}$$

where R is a constant of proportionality. The distance between the two protons can thus be calculated from the slope of the initial NOE buildup.

D. 2D NOE Spectroscopy

As the molecules that are being studied increase in size, one dimensional difference and transient experiments become increasingly difficult due to spectral overlap, which does not allow for selective excitation. The problem can be solved to some degree by moving to two dimensions. The 2D NOESY[65,66] experiment is very similar to its 1D counterpart, the transient NOE experiment. In the 2D version of the experiment the initial 180° pulse is replaced by two 90° pulses allowing for frequency labeling in the t_1 domain (F1 dimension), followed by the mixing time and acquisition.

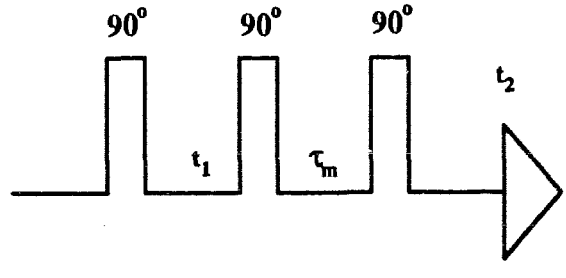


Figure 1.5. Pulse sequence of a 2D NOESY experiment.

E. Dependence of the NOE on τ_c

For a two spin system, the equation for the NOE is (Equation 1.22)

$$NOE(t) = -e^{-(\rho+\sigma)t} + e^{-(\rho-\sigma)t} \tag{1.22}$$

This function reaches a maximum value at time, t_{max}

$$t_{max} = \left(\frac{1}{2\sigma}\right)\ln\left(\frac{\rho+\sigma}{\rho-\sigma}\right) \tag{1.24}$$

The maximum NOE is then

$$NOE_{max} = -\left(\frac{\rho+\sigma}{\rho-\sigma}\right)^{-\left(\frac{\rho+\sigma}{2\sigma}\right)} + \left(\frac{\rho+\sigma}{\rho-\sigma}\right)^{-\left(\frac{\rho-\sigma}{2\sigma}\right)} \tag{1.25}$$

In an NOE experiment, the value $\omega\tau_c$ is critical. For a transient NOE experiment, it can be shown that for $\omega\tau_c << 1$ and $\omega\tau_c >> 1$ the NOE approaches limiting values of 0.385 and -1 respectively; when $\omega\tau_c \cong 1$, the function passes through zero and the observed NOE is approximately 0. The result of this dependence is that the NOESY experiment is limited to use on molecules which are large or small. For intermediate sized molecules, where the product $\omega\tau_c \cong 1$, another experiment, the ROESY[67,68] experiment has been proposed.

1.3.2.5. Rotating Frame Overhauser Spectroscopy (ROESY)

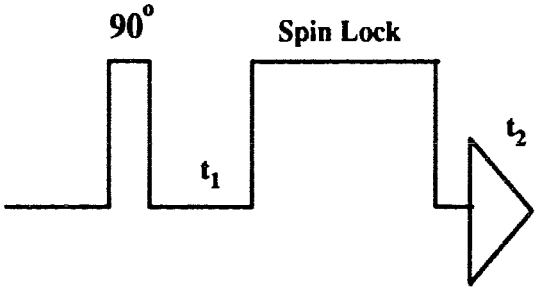


Figure 1.6. Pulse sequence of a 2D ROESY experiment..

Observation of NOEs in the rotating frame circumvents the problem arising from the dependence on $\omega\tau_c$. The theory of rotating frame relaxation is similar to that discussed for the transient NOE experiment, and the expression for the NOE_{max} is the same as in equation 1.24. In the ROESY experiment a spin lock pulse of lower power follows the initial 90° pulse, effectively locking the magnetization in the xy plane and allowing for transverse NOE buildup. The magnetization now appears to precess about the effective field, which makes an angle α as shown in Figure 1.7.

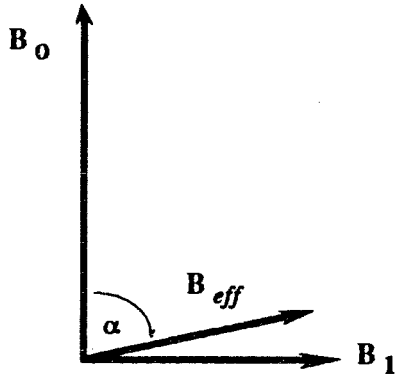


Figure 1.7.
The effective magnetic field.

The expression for the spectral density $\mathbf{J}(\omega)$ is now given by

$$J(\omega) = \frac{2\tau_C}{1+\omega_{eff}^{2}\,\tau_C^{2}} \qquad \textbf{(1.26)}$$

ω_{eff} is the frequency of the effective magnetic field and is defined by

$$\omega_{eff} = \left((\gamma B_1)^2 + \delta\omega^2\right)^{\frac{1}{2}} \qquad \textbf{(1.27)}$$

$\gamma\mathbf{B_1}$ is the strength of the spin locking field and $\delta\omega$ is the offset between the spin lock and the resonance frequency of the nucleus. The expressions for ρ and σ are different since the relaxation is modulated in the transverse plane, and in the homonuclear case,

$$\rho = \frac{\gamma^4 \hbar^2}{20 r_{ij}^6}(5J(\omega_{o\rho}) + 9J(\omega_{1\rho}) + 6J(\omega_{2\rho}))$$
$$\sigma = \frac{\gamma^4 \hbar^2}{20 r_{ij}^6}(3J(\omega_{1\rho}) + 2J(\omega_{o\rho})) \qquad (1.28)$$

Examination of the equations 1.28, indicates that the NOE_{max} in the rotating frame will remain positive for all values of the correlation time τ_c, and increases from a minimum value of 0.385 at $\omega\tau_c << 1$ to a maximum of 0.675 for $\omega\tau_c >> 1$. A plot of the NOE and ROE are shown in Figure 1.8.

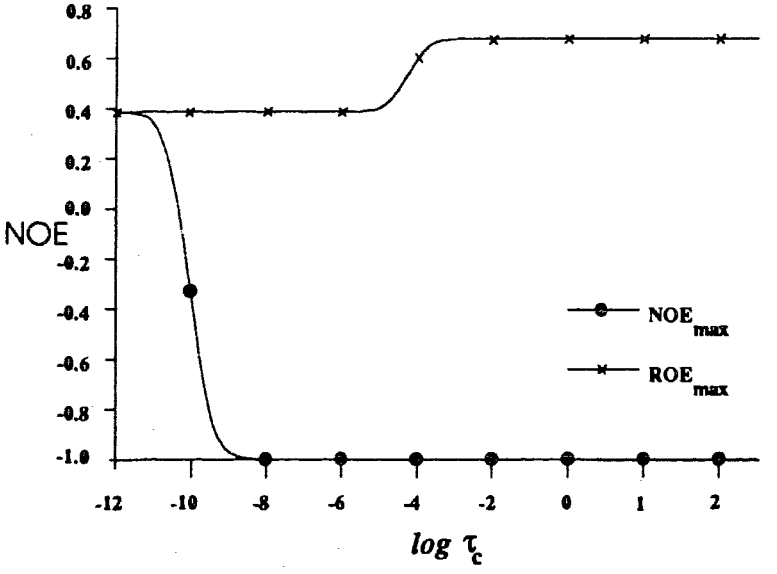


Figure 1.8. Dependence of the NOE and ROE on the correlation time τ_c.

From Figure 1.8 it can be seen that the ROESY experiment can be extremely useful for molecules in the range of $\omega\tau_c \cong 1$ where the NOE would be negligible. This is apparent from the comparison of the ROESY and NOESY spectra of the pentasaccharide (**5**), illustrated in Figure 1.9. The ROESY spectrum was acquired with a 500ms spin lock. A large number of cross peaks are visible in this spectrum that are absent from the NOESY spectrum of the same compound acquired with a mixing time of 500ms. The important cross peaks are now of very low intensity and some are no longer present.

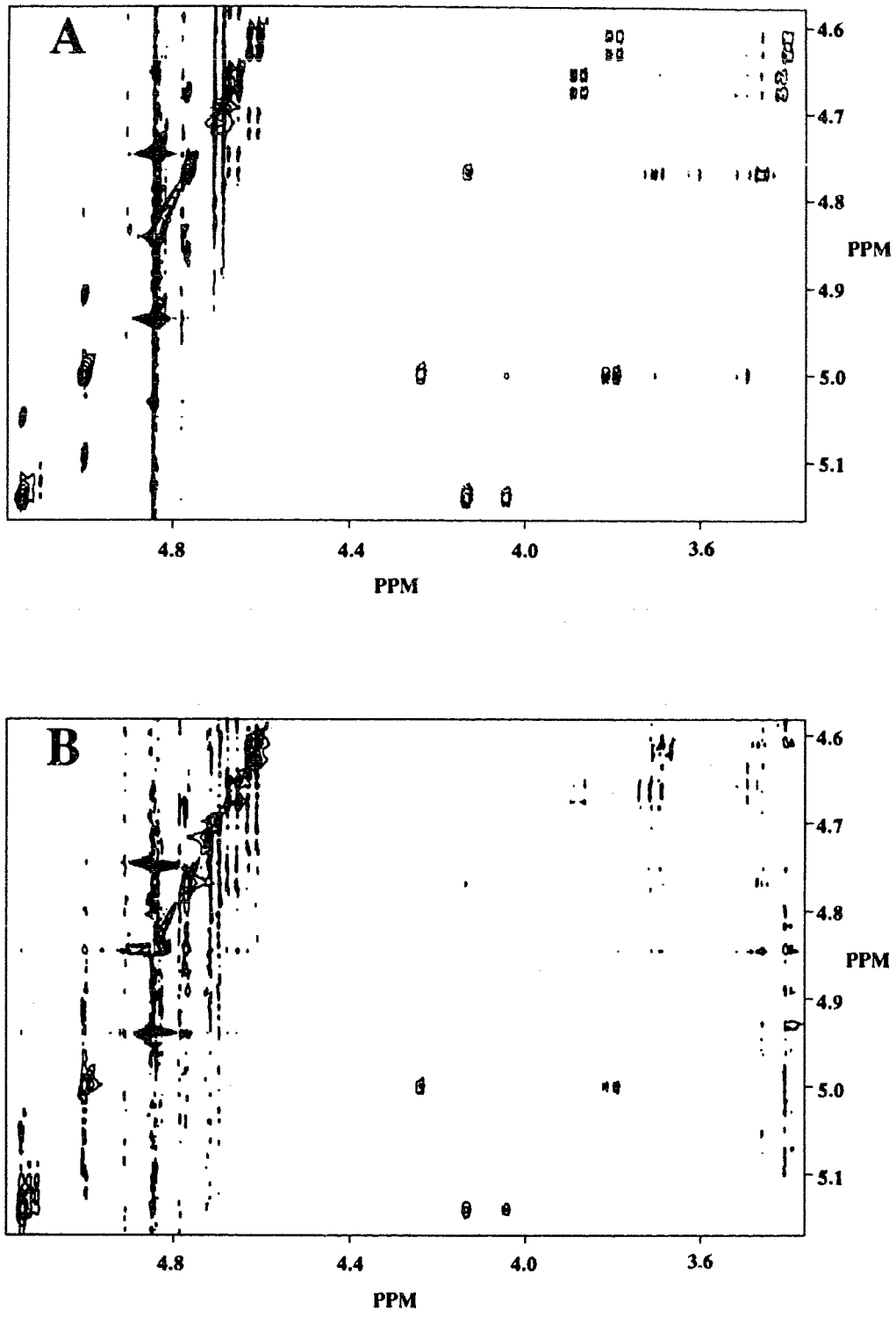


Figure 1.9. Comparison of the A) ROESY and B) NOESY spectra of the pentasaccharide (**5**).

A. Quantitation of ROESY Data

Information from a ROESY experiment is obtained in the form of cross peaks in the 2D spectrum between spatially close nuclei. The relationship between the cross peaks and internuclear distances must be established in order to gather quantitative information from the experimental data.

A.1. Relationship Between Cross Peaks a_{ij} and Internuclear Distance r_{ij}

In order to understand the relationship between cross peaks and the internuclear distance, the origins of the cross peaks must be understood. The rate of change of magnetization in an AX system was described by a set of coupled differential equations (Equations 1.1)

$$\frac{dM_A}{dt} = -\rho(M_A^o - M_A) - \sigma(M_X^o - M_X)$$
$$\frac{dM_X}{dt} = -\rho(M_X^o - M_X) - \sigma(M_A^o - M_A)$$

Extension to a system of N spins, gives N equations

$$\begin{aligned}
\frac{dM_A}{dt} &= -\rho_A(M_A^o - M_A) - \sigma_{AX}(M_X^o - M_X) - \quad \ldots \quad -\sigma_{AN}(M_N^o - M_N) \\
\frac{dM_X}{dt} &= -\sigma_{AX}(M_A^o - M_A) - \rho_X(M_X^o - M_X) - \quad \ldots \quad -\sigma_{XN}(M_N^o - M_N) \\
&\ldots \\
\frac{dM_N}{dt} &= -\sigma_{AX}(M_A^o - M_A) - \sigma_{XN}(M_X^o - M_X) - \quad \ldots \quad -\rho_X(M_N^o - M_N)
\end{aligned} \tag{1.29}$$

This can be written conveniently in matrix form as

$$\begin{bmatrix} \frac{dM_A}{dt} \\ \frac{dM_X}{dt} \\ \cdots \\ \frac{dM_N}{dt} \end{bmatrix} = - \begin{bmatrix} \rho_A & \sigma_{AX} & \cdots & \sigma_{AN} \\ \sigma_{AX} & \rho_X & \cdots & \sigma_{XN} \\ \cdots & & & \\ \sigma_{AN} & \sigma_{XN} & \cdots & \rho_N \end{bmatrix} \begin{bmatrix} M_A^0 - M_A \\ M_X^0 - M_X \\ \cdots \\ M_N^0 - M_N \end{bmatrix}$$

or

$$\frac{d\bar{M}}{dt} = -\bar{R} \cdot \bar{M} \tag{1.30}$$

where $\bar{M}$ is the matrix of magnetization, and $\bar{R}$ is the relaxation matrix. The solution of this equation is

$$\bar{A}(\tau_m) = e^{-\left(\bar{R}\tau_m\right)} \bar{M}^0 \tag{1.31}$$

where $\bar{A}$ is the matrix of intensities and $\bar{M}^0$ is the initial magnetization. In order to calculate the individual internuclear distances, the cross peak intensity a_{ij} has to be related to the cross relaxation rate. The exact, analytic solution of the equation 1.31 would involve diagonalization of the relaxation matrix $\bar{R}$, which, for a system of N spins is an N x N matrix. However, after substitution of the appropriate boundary conditions and normalization ($M_A = M_B = M_N = 1$), equation 1.31 can be written as

$$\bar{A}(\tau_m) = \bar{I} - \bar{R}\tau_m + \frac{\left(\bar{R}\tau_m\right)^2}{2!} - \frac{\left(\bar{R}\tau_m\right)^3}{3!} + \ldots \tag{1.32}$$

where $\bar{I}$ is the identity matrix and normalized intensities have been used. At time $\tau_m <$ 1 the first two terms of the equation 1.32 will be much larger than the remaining terms, and the equation 1.31 can be further reduced to

$$\bar{A}(\tau_m) = \bar{I} - \bar{R}\tau_m$$

Writing out the matrices **A** and **R** in the equation

$$\begin{bmatrix} A_{11} & a_{11} & \cdots & a_{1N} \\ a_{21} & A_{22} & \cdots & a_{2N} \\ \cdots & & & \\ a_{N1} & a_{N1} & \cdots & A_{NN} \end{bmatrix} = \bar{I} - \begin{bmatrix} \rho_A & \sigma_{AX} & \cdots & \sigma_{AN} \\ \sigma_{AX} & \rho_X & \cdots & \sigma_{XN} \\ \cdots & & & \\ \sigma_{AN} & \sigma_{XN} & \cdots & \rho_N \end{bmatrix} \cdot \tau_m$$

indicates that there is a direct relationship between the cross peak intensity a_{ij} and the internuclear distance r_{ij} , given by

$$a(t)_{ij} = \sigma_{ij}\ t + b = \frac{K}{r_{ij}^{-6}} t + b \qquad \textbf{(1.33)}$$

Thus, if several experiments are run, the initial slope of the ROESY buildup curves can be fitted to the equation 1.33, and will yield the cross relaxation rate σ_{ij}. This approach to determining σ_{ij} is called the Initial Rate analysis, and is an approximation that is only valid at short mixing times where the buildup is linear.

In most cases, especially when the ROESY spectrum of a compound is recorded at only one mixing time, it is assumed that the observed intensities, a_o and a_{ij} are proportional to the cross relaxation rates σ_o and σ_{ij}, respectively.

A.2. Calculation of Internuclear Distances

Both of the methods allow calculation of the cross relaxation rate σ_o and σ_{ij}. The cross relaxation rate can then be related to distance by Equation 1.28,

$$\sigma = \frac{\gamma^4 \hbar^2}{20 r_{ij}^{-6}} (3J(\omega_1) + 2J(\omega_0)) \qquad \textbf{(1.28)}$$

If isotropic tumbling is assumed, then the terms $J(\omega_0)$ and $J(\omega_1)$ will be constant for all proton pairs and

$$\sigma = \frac{K}{r_{ij}^{-6}} \tag{1.34}$$

Use of this relationship and the cross relaxation rates σ_0 and σ_{ij}, the distances can then be estimated by a simple approximation, i.e.

$$r_{ij} = r_o \left(\frac{\sigma_o}{\sigma_{ij}} \right)^{1/6} \tag{1.35}$$

where r_0 is a known, fixed distance such as the distance between two protons within a ring (termed the ruler distance), and σ is the cross relaxation between the two protons of interest. Once the internuclear distances of all the observed contacts are calculated and if sufficient inter-ring contacts are observed, the three dimensional structure of the oligosaccharide can be deduced.

1.3.3. Computational Methods

Theoretical calculations can be also be utilized to predict the conformations of the compounds being studied, and with computer analysis can lead to a better understanding of the three-dimensional molecular surfaces involved. The basic premise of all computational methods is that a molecule will preferentially occupy the lowest energy state available to it, and the methods thus involve searching for the lowest energy conformation. Two theoretical approaches that can be applied to the study of carbohydrates are *ab initio* molecular orbital calculations and molecular mechanics or force field calculations.

1.3.3.1. *Ab initio* Calculations

Ab initio calculations[69-71] represent a rigorous approach to structure calculations. Consider a single, stationary Hydrogen atom. The energy of the atom is the sum of the kinetic energy of the electron and the potential energy associated with the electron nuclear interaction, i. e.,

$$E = \hat{T} + \hat{U} \tag{1.36}$$

where $\hat{T}$ and $\hat{U}$ are linear operators corresponding to the kinetic and potential energy. Here, the kinetic energy is the linear differential operator

$$\hat{T} = -\frac{\hbar^2}{2m}\nabla^2 \tag{1.37}$$

and the potential energy is the coulombic energy associated with the electron interacting with the nuclear charge,

$$\hat{U} = -\frac{Ze^2}{4\pi r} \tag{1.38}$$

With atomic units, equation 1.36 can be written as

$$E = \hat{T} + \hat{U} = -\frac{1}{2}\nabla^2 - \frac{Z}{r} \tag{1.39}$$

The motion of the electron can be described by a wave function ϕ, which is the solution of the Schrödinger equation, and can be represented by

$$\hat{H}\phi = E\phi \tag{1.40}$$

where $\hat{H}$ is called the Hamiltonian operator and is defined as $\hat{H} = \{\hat{T} + \hat{U}\}$. There are several exact solutions to equation 1.40 represented by one electron atomic orbitals. Calculation of the energy requires solution of the equation

$$E = \left\langle \phi^* \left| \hat{H} \right| \phi \right\rangle \tag{1.41}$$

which involves calculating the integrals

$$E = \int \phi^* \left| -\frac{1}{2}\nabla^2 - \frac{Z}{r} \right| \phi \qquad (1.42)$$

In a molecule, there are more interactions to consider, namely electron-electron and electron-nuclear interactions. The Hamiltonian takes on a more complex form, i.e.,

$$\hat{H} = -\frac{1}{2}\sum_{i}^{electrons}\nabla^2 - \sum_{i}^{electrons}\sum_{s}\frac{1}{r_{is}} - \sum_{i}^{electrons}\sum_{k}^{nuclei}\frac{Z}{r_{ik}} - \sum_{i}^{nuclei}\sum_{s}\frac{Z_i Z_s}{r_{is}} \qquad (1.43)$$

Even for the simplest molecule, H_2^+, there is no exact solution to the equation 1.43, and the molecular orbitals, ψ, are approximated using a Linear Combination of Atomic Orbitals (LCAO), i.e,

$$\psi = \phi_1 + \phi_2 + \phi_3 + \phi_4 + \ldots \qquad (1.44)$$

This makes the solution of the Schrödinger equation difficult and time consuming. In order to simplify the integrals that have to be solved to compute the energy of the molecule, most programs today use a linear combination of gaussian functions to describe atomic wave functions. The number of gaussian functions that are used as a basis determines the accuracy of the calculation. Due to the number of electrons involved, the number of atoms in a molecule that can be handled by *ab initio* calculations is severely limited, and the application of molecular orbital calculations to carbohydrates has thus far been restricted to monosaccharides[72,73]. An alternative method of structure calculation that can manage large molecules is molecular mechanics[74].

1.3.3.2. Molecular Mechanics

The energy of a diatomic molecule as a function of the internuclear separation, $U(r)$, can be approximated by an empirical function called a Morse potential,

$$U(r) = \left(D_e\left(1 - e^{-\beta(r-r_o)}\right)\right) \qquad (1.45)$$

where D_e is the dissociation energy and β is a constant. Near the equilibrium bond length, the potential can be expanded into a Taylor series and approximated by the quadratic term, similar to the potential described by Hooke's law for a harmonic oscillator, i.e.,

$$U(r) = \frac{1}{2}\left(k(r - r_o)^2\right) \tag{1.46}$$

where k is the force constant associated with the bond and is related to De and β by the relation

$$k = \beta^2 D_e$$

A superposition of the Morse potential and the harmonic oscillator for the dissociation[75] of H_2 are shown in Figure 1.10.

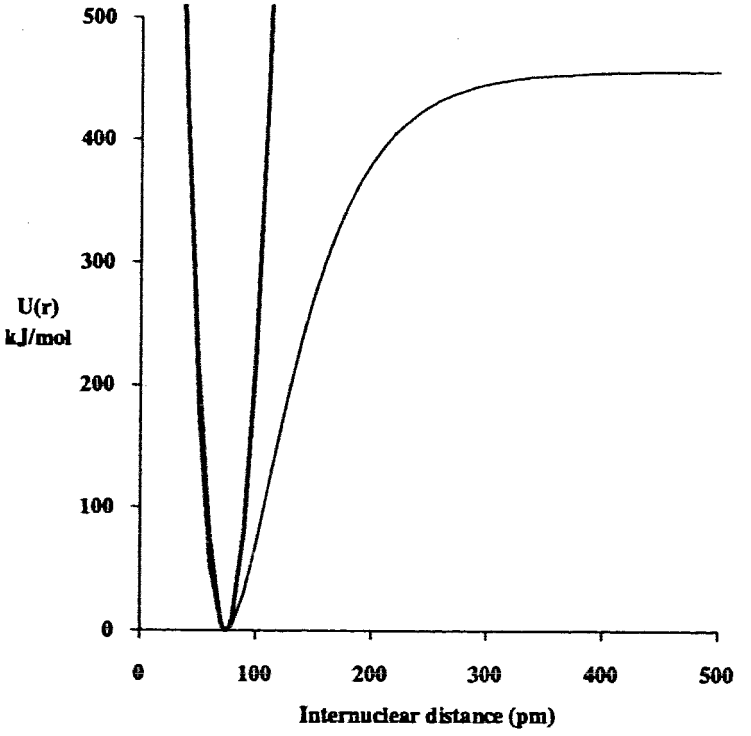


Figure 1.10. The Morse potential for H_2.

The energy of an H_2 molecule can thus be approximated as a function of its internuclear distance by a simple classical equation.

If a linear triatomic molecule (A-B-C) is considered, the energy of the system is dependent on the position of all three atoms. In internal coordinates, only the two distances corresponding to the A-B distance (r_{ab}) and the B-C (r_{bc}) distance need be considered. Two Potential Energy (PE) diagrams, similar to that drawn for H_2, can be drawn for the displacements from equilibrium of the respective atoms, but greater perspective is obtained from combining the two curves to produce a potential energy surface.

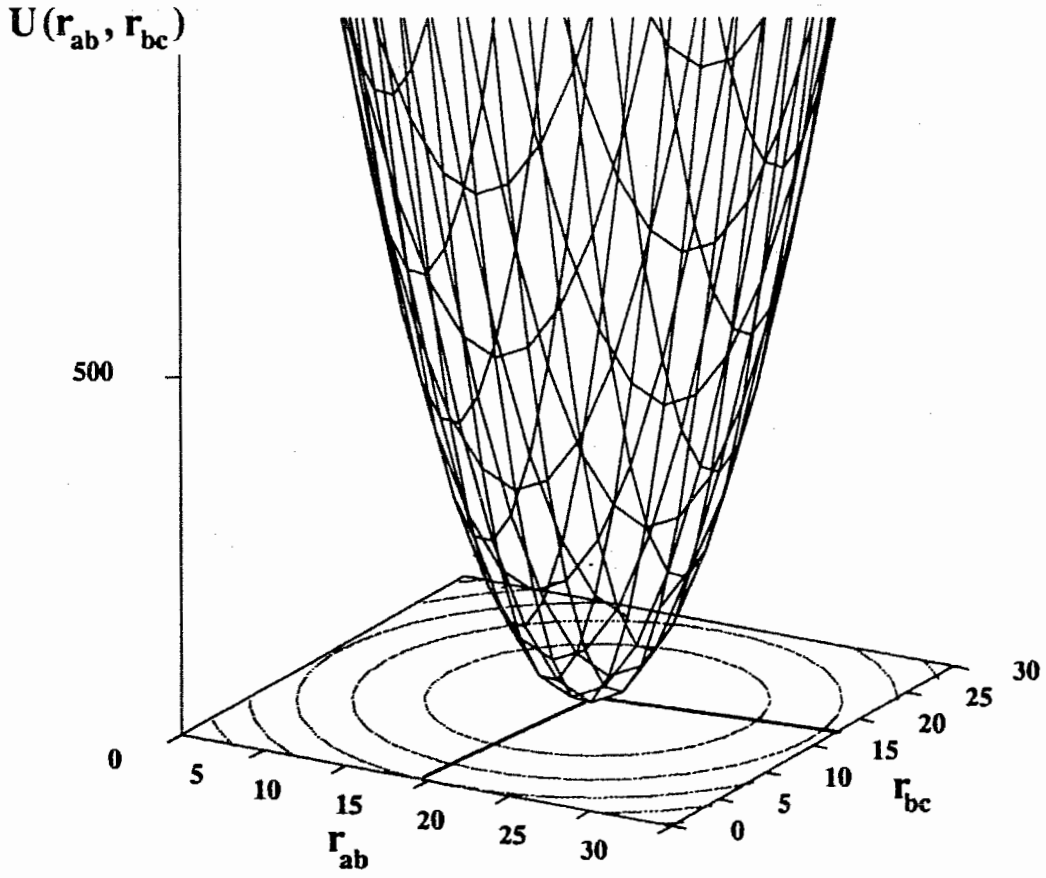


Figure 1.11. The potential energy surface for a linear triatomic molecule.

The potential energy can be written as a function of the two internuclear distances, i.e.,

$$U(r_{ab}, r_{bc}) = \frac{1}{2}\left(k_{ab}(r_{ab} - r_{ab}^{o})^2\right) + \frac{1}{2}\left(k_{bc}(r_{bc} - r_{bc}^{o})^2\right) \qquad \textbf{(1.47)}$$

The example of the linear triatomic molecule is a special case of the larger group of triatomic compounds where the angle ∠ABC is equal to 180°. When the molecule is not linear a third parameter, the energy associated with distortions of the angle $\theta_{\angle ABC}$ from its equilibrium value, contributes to the energy of the molecule and the potential energy can then be written as a function of three terms,

$$U(r_{ab}, r_{bc}, \theta_{abc}) = \frac{1}{2}\left(k_{ab}(r_{ab} - r_{ab}^{o})^2\right) + \frac{1}{2}\left(k_{bc}(r_{ab} - r_{bc}^{o})^2\right) + \frac{1}{2}\left(k_{\theta}(\theta_{abc} - \theta_{abc}^{o})^2\right) \qquad \textbf{(1.48)}$$

The angle bending is also treated like a harmonic oscillator. The PE surface is now a function of three variables and it is not possible to depict the surface of more than two variables in three dimensions. Potential energy surfaces of more that two variables are often referred to as hypersurfaces.

As larger molecules are studied, 1,4 and 1,3 interactions, i.e., interactions between atoms that are not directly bonded have to be included. Physically, these interactions correspond to van der Waals or Heitler-London forces. In general, they are represented by the sum of attractive and repulsive forces, i.e.,

$$U(r) = \left(\frac{A}{r^{12}} - \frac{B}{r^{6}}\right) \qquad \textbf{(1.49)}$$

Equation 1.49 is referred to as a Lennard-Jones or a 6/12 potential. Aside from the van der Waals energy, there is an additional variation in energy associated with the rotation about torsion angles. Consider ethane. The variation of energy as a function of the H_a-H_b torsion angle is illustrated[76] in Figure 1.12. This cannot be fully reproduced by

the Lennard-Jones potential, but can be approximated by simple functions, two of which are

$$V_{tor} = k_{\phi}(1-\cos(n\phi-\delta)) \qquad \textbf{(1.50a)}$$

or

$$V_{tor} = \sum_{j=1}^{3}\frac{1}{2}V_j(1-\cos(j\phi)) \qquad \textbf{(1.50b)}$$

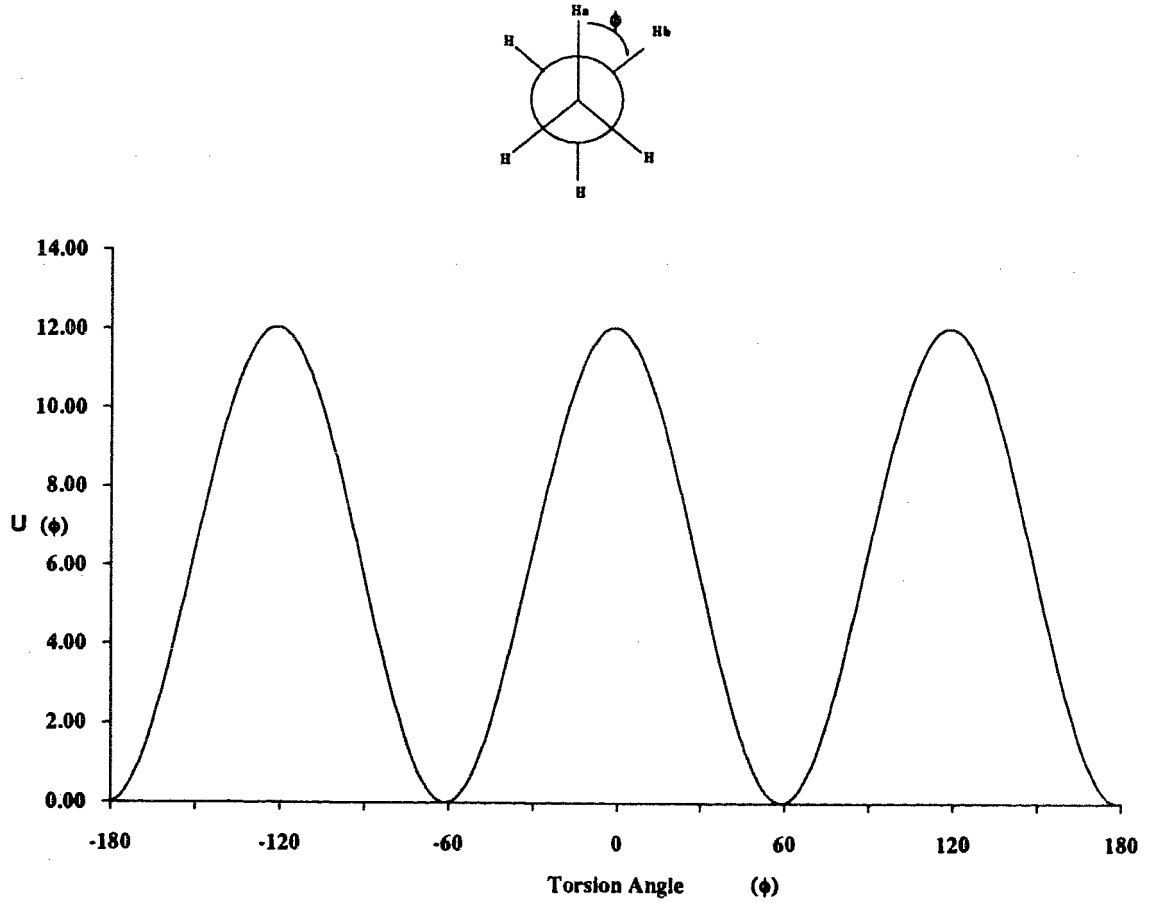


Figure 1.12. The energy associated with the HCCH torsion angle in ethane.

Description of the energy of a molecule as a function of the torsion angle are especially important when heteroatoms are included amongst the four atoms that make

up the torsion, since asymmetric variations in the energy that cannot be explained simply by van der Waals interactions occur. Two important examples of this behavior are the anomeric effect and the gauche effect, which will be discussed later.

It is thus easy to see the origins of methods that use classical functions to calculate the energy associated with the positioning of the nuclei in a molecule. Calculations of these kinds are called molecular mechanics or force field calculations.

In addition to the interactions mentioned above, molecules that contain a large number of heteroatoms and polar functionalities cannot adequately be described by steric interactions alone. Charges associated with acidic and basic groups must also be considered. In order to reproduce these interactions, most force fields include an electrostatic term that is a simple coulombic potential, i.e.,

$$\sum_{i \neq j} \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} \qquad \textbf{(1.51)}$$

where q_i is the charge on each atom.

The total potential energy of a molecule V_{TOT} can then be written as a sum of each of its components, i.e.

$$V_{TOT} = V_{BOND} + V_{ANGLE} + V_{VDW} + V_{DIHEDRAL} + V_{ELECTROSTATIC}$$

or, more specifically, the general form of a potential function is

$$V_{TOT} = k_b (r-r_o)^2 + k_\theta (\theta - \theta_o)^2 + \sum_{j \neq i} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + k_\phi (1 + \cos(n\phi - \delta)) + \sum_{i \neq j} \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} \qquad \textbf{(1.52)}$$

This simplistic approach greatly reduces the number of calculations required per atom, and the limitations of molecular mechanics on the number of atoms in a system of study are less than those imposed by *ab initio* methods. Some potential functions

include terms to account for specific effects such as H-bonding. Distortions from the Morse potential that are a result of the quadratic approximation (Equation 1.45) are partially corrected by including the cubic term from the Taylor series expansion.

It can be seen from the preceding discussion that the determination of the force constants, or parameterization, of any force field requires experimental data. Values of the various constants, or parameters, k_b, k_θ, k_ϕ, can initially be obtained from experimental data or *ab initio* calculations. The parameters are then optimized so as to give the best fit to observed structural or physical data, either by a least squares method or by actual inspection of the calculated results. The results of a molecular mechanics program will therefore depend on the experimental data to which the parameters are fit, and often a particular parameterization will be best suited to reproduction of the data to which it is fit, or the class of compounds for which the experimental data was measured.

1.4. Conformational Analysis of Oligosaccharides

The monosaccharide units that form an oligosaccharide possess a well defined geometry. Pyranoses in general adopt a chair conformation, either the $^1C_4$ or a $^4C_1$ conformation, shown in Figure 1.13. Proton coupling constants observed for

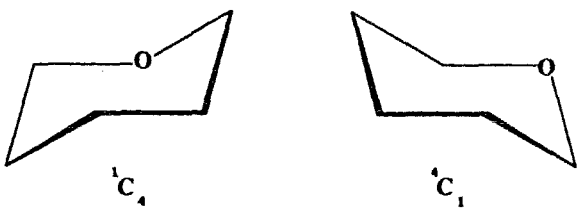


Figure 1.13. $^1C_4$ and $^4C_1$ conformations of pyranose rings.

carbohydrates in solution indicate that the pyranose rings do not undergo severe distortions from their most stable chair conformations[77]. This restriction limits the conformational mobility of an oligosaccharide to rotation about the interglycosidic bonds

and rotation of any pendant groups such as primary hydroxyls and acetamido side chains, greatly simplifying calculations. The overall conformation of an oligosaccharide can therefore be specified by the interglycosidic angles[78] Φ and Ψ. In Figure 1.14, ring B is referred to as the aglycon. The Φ angle is defined as the H1-C1-O1-Cn torsion angle and the Ψ angle is defined as the C1-O1-Cn-Hn torsion angle where Cn and Hn are the aglyconic carbon and proton involved in the glycosidic bond. The angles are defined to be positive when the distal atom in the torsion group is clockwise from the proximal atom. These angles are sometimes referred to as the Φ^H and Ψ^H angles since they refer to the proton of the glycoside and aglycon. In addition to the Φ and Ψ angles, another variable that has to be defined for most of the saccharides is the ω angle. This angle is defined as the torsion O5-C5-C6-O6, with the three staggered rotamers being referred to as the *gt* (gauche-trans, $\omega = 60°$), *gg* (gauche-gauche, $\omega = -60°$), and the *tg* (trans-gauche, $\omega = -180°$) conformations. The terminology is derived by stating the O6-C6-C5-O5 angle first, and the O6-C6-C5-C4 angle second.

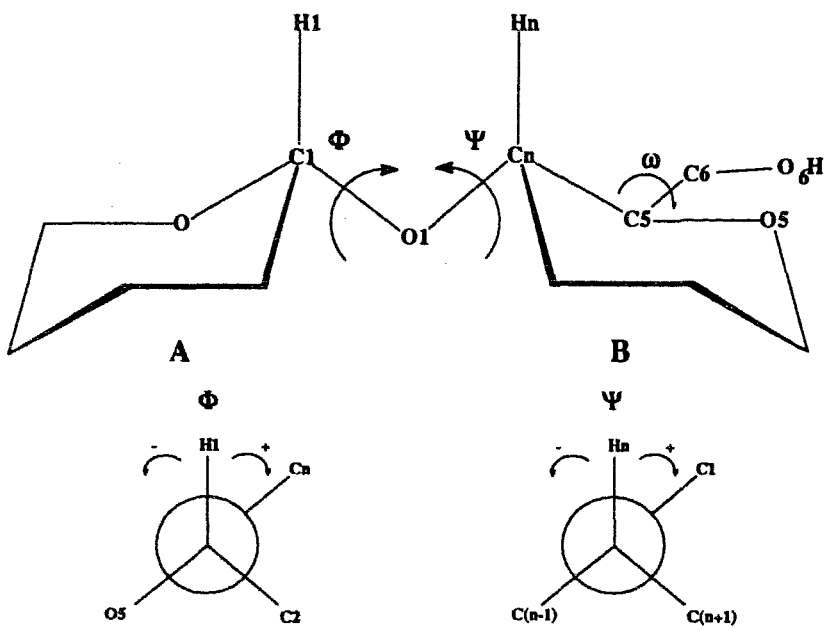


Figure 1.14. Definition of the Φ, Ψ and ω angles for a glycosidic linkage.

1.4.1. The Anomeric and Gauche Effects[78-86]

Hexopyranoses exhibit a preference for the axial configuration, a phenomenon that is termed the anomeric effect, and manifests itself in acetal fragments of the type X-C-Y where X=N, O, S or Se and Y=O, N, S, Br, Cl, F[83]. This preference has been explained in terms of an n - σ* interaction between the lone pair of the X heteroatom and the σ* orbital of the adjacent C-Y bond[80,84].

By analogy with the anomeric effect, the preference of the aglycon to adopt a gauche conformation to the neighboring X-C bond is termed the exoanomeric effect[78]. This serves to maximize n-σ* overlap, this time between the lone pair of the Y atom and the X-C σ* orbital.

The gauche effect was proposed to explain the observed preference of certain compounds to exhibit " a tendency to adopt that structure which has a maximum number of gauche interactions between the adjacent electron pairs and/or polar bonds"[81]. In carbohydrates, this effect is observed in the behavior of the primary alcohol of hexopyranoses, or in (1→6) linked oligosaccharides.

Attempts have been made to reparameterize force fields to account for such effects[87-95]. The gauche effect, in principle, should be reproduced by a sum of suitably chosen electrostatic, van der Waals and torsional potentials. The anomeric effect can also be reproduced by low periodicity torsional terms for the C-O-C-O fragment. Observed C-O bond length and C-O-C bond angle distortions can be reproduced by changing the equilibrium bond lengths and bond angles associated with the functional group. This can be accomplished by identifying the acetal fragment and calculating the equilibrium bond lengths and angles as a function of the torsion angles.

1.4.2. Force Fields

1.4.2.1. HSEA

HSEA[7,96,97] (Hard Sphere Exoanomeric Effect) calculations take advantage of the premise that monosaccharide rings behave as rigid entities. This treatment, incorporated into programs such as GESA[98] (Geometry optimization of Saccharides) considers atoms to be hard spheres of fixed Van der Waals radii. Subsequently, each monosaccharide ring, by virtue of its inflexibility, will have a fixed volume. The stability of a particular saccharide is then calculated by computing the energies that arise due to non bonded interactions between the constituent monosaccharide units in addition to the energy contributed due to the exoanomeric effect. The non bonded interactions between two atoms is calculated by the function first published by Kitaygorodski[99,100]

$$V = 3.5\left(- \left(\frac{0.04}{z^6}\right) + 8.5(10^3)\exp^{-13z} \right) \text{ kcal/mol} \qquad (1.53)$$

where

$z = \frac{r_{ij}}{r_o}$ r_o is the equilibrium inter atomic distance.

$r_{ij} = 1.11(r_i + r_j)$ r_i , r_j are the Van der Waals radii.

A correction to account for the effects of the exoanomeric effect is added to this equation[7,,97,101],

$$EA = k_1 (1-\cos\phi) + k_2 (1-\cos 2\phi) + k_3 (1-\cos 3\phi)$$

where k_1, k_2, k_3 are constants dependent on whether the sugar is an α or β anomer.

The HSEA function is the simplest molecular mechanics type of calculation that has been applied to carbohydrates. In fact, the potential function is restricted to the computation of oligosaccharides. Recently, with the introduction of the GEGOP[95] (Geometry of glycopeptides) program, this function has been modified to encompass proteins, allowing for the modeling of glycopeptides.

1.4.2.2. MM2, MM3, AMBER, CHARMM

Several more complex force fields have been utilized in the calculation of oligosaccharide structure, the most notable of these being MM2[74], MM3[102] (Molecular Mechanics 2 and 3), AMBER[103] (Assisted Model Building and Energy Refinement) and CHARMM[104] (Chemistry at HARvard Molecular Mechanics). The potential functions employed by these methods have separate terms to account for bond stretching, angle bending and torsional rotations, as well as Van der Waals and electrostatic interactions. These programs use different force constants for different atom types and are therefore, much more flexible and can be used with various classes of compounds. MM2, which has evolved from MM1, is considered the prototypical molecular mechanics program. The potential function used is shown in detail overleaf (Figure 1.15). MM2 has been applied to several classes of compounds, including carbohydrates. It has been reparameterized to mimic the anomeric effect in O-C-O-R type fragments[87-91] (a version sometimes referred to as MM2CARB) and has been used with limited success.

A major limitation of the MM2 force field is its inability to accurately predict vibrational data. In order to correct this, the MM3 force field was introduced[102], and though similar in many respects to its predecessor, the inclusion of additional terms has made MM3 extremely accurate in the reproduction of hydrocarbon geometry. The parameterization for the anomeric effect has been retained in MM3, and subsequently it has been used to model saccharides. Extensive studies have been carried out on a vast array of disaccharides, and the results show an improvement from MM2[105-107].

Both AMBER and CHARMM are representative of the multifunctional force fields that are now available, and have similar potential functions. The most notable difference between the two is the exclusion of an explicit H-bonding term in CHARMM. The CHARMM force field will be discussed in further detail in Chapter II.

$$V_{TOT} = V_{VDW} + V_C + V_{DIP} + V_{BEND} + V_{SB} + V_{TOR}$$

V_{VDW} - Van der Waals term

$$V_{VDW} = \varepsilon^*(2.9(10^5)\exp(-\frac{12.5}{P}) - 2.5P^6) \qquad P \leq 3.31$$

$$= \varepsilon^* 336.176P^2 \qquad P \geq 3.31$$

ε^* is a constant determined by the atom types

$P = \frac{r^*}{R}$ and $r^* = r_i + r_j$, the combined Van der Waal's radii and R is the effective internuclear distance. For $P \leq 3.31$, this represents a variation of the Lennard-Jones 6/12 potential.

V_C - Bond compression energy

$$V_C = 143.88 (\frac{k_s}{2})(\Delta l)^2(1 + CS\Delta l)$$

143.88 is a constant to convert mdynes/Å to kcal/mol/Å^2.
k_s is the stretching constant for the bond in mdynes/Å.
$\Delta l = l - l_{eq}$ and l_{eq} is the equilibrium bond length.
CS is the cubic stretch term.

V_{DIP} - Dipole interaction energy

$$V_{DIP} = 14.39418(\mu_A)(\mu_B)(\cos(x) - 3\cos(\alpha_A)\cos(\alpha_B)) / R^3$$

14.39418 is a constant to convert ergs/molecule to kcal/mole.
μ_A , μ_B are the bond moments of the two bonds.
x is the angle between the dipoles, and R is the distance between the midpoints of the two bonds. α are the angles between the dipoles and the vector **R**.

V_{BEND} - Bending energy

$$V_{BEND} = 0.043828 \frac{k_b}{2} (\Delta\Theta)^2 (1 + CF(\Delta\Theta)^4)$$

0.043828 = conversion factor for mdyne Å/rad^2/molecule to kcal/deg^2/mole.
k_b is the bending constant for the bond in mdyne Å/rad^2 for a specific angle type.
$\Delta\Theta = \Theta - \Theta^0$ where Θ^0 is the angle at minimum energy.
CF is the sextic bending term.

V_{SB} - Stretch bend energy

$$V_{SB} = 2.51118\, k_{sb}\, \Delta\Theta_{abc}\, (\Delta l_{ab} + \Delta l_{bc})$$

k_{sb} = stretch bend constant in mdyne/rad for the angle abc
$\Delta\Theta = \Theta - \Theta^0$ where Θ^0 is the angle at minimum energy.
$\Delta l = l - l_{eq}$ and l_{eq} is the equilibrium bond length.

V_{TOR} - Torsional energy

$$V_{TOR} = \frac{k_1}{2}(1-\cos\phi) + \frac{k_2}{2}(1-\cos 2\phi) + \frac{k_3}{2}(1-\cos 3\phi)$$

k_1, k_2, k_3 are constants which determine the position and relative magnitude of the potential energy maxima and minima.

Figure 1.15. The MM2 force field[108]

1.4.3. The Multiple Minima Problem

Once a suitable force field has been selected, the optimum geometry of the oligosaccharide has to be calculated. This is a complex procedure which involves searching the potential energy surface for the absolute or global minimum. If the two dimensional case of the H_2 molecule is considered, the minimum energy structure can be obtained from the derivative of the equation 1.46, i.e.,

$$\frac{\delta U(r)}{\delta r} = \frac{1}{2}\frac{\delta\left(k(r-r_o)^2\right)}{\delta r} = k(r-r_o) \tag{1.54}$$

A *critical point* occurs at a value of r where the derivative is equal to 0, which is at $r = r_0$. In order for this to be a minimum, the second derivative of the function must be positive, i.e.,

$$\frac{\delta^2 U(r)}{\delta r^2} > 0 \tag{1.55}$$

If the potential energy is a function of two variables, i.e., $U = U(r_1, r_2)$ as in the case of the linear triatomic molecule, the potential energy of the molecule is now described by a surface. The potential energy surface of a linear triatomic, illustrated in Figure 1.11, shows a single minimum, which greatly simplifies calculation of the optimum geometry structure. In this case, by analogy with equations 1.54 and 1.55 , the two conditions

$$\frac{\delta U(r_{ab}, r_{bc})}{\delta r_i} = 0 \tag{1.56}$$

and

$$\frac{\delta^2 U(r_{ab}, r_{bc})}{\delta r_i^2} > 0 \tag{1.57}$$

do not necessarily imply a minimum. A third condition,

$$\frac{\delta^2 U}{\delta r_{ab}{}^2}\frac{\delta^2 U}{\delta r_{bc}{}^2}-\left(\frac{\delta^2 U}{\delta r_{ab}\delta r_{bc}}\right)^2>0 \qquad \textbf{(1.58)}$$

must be met for the function to be a minimum. If equation 1.58 < 0, then the function is a maximum. In the case of the linear triatomic there is only one point where equation 1.55 is fulfilled, when $r_{ab}=r_{ab}^{o}$ and $r_{bc}=r_{bc}^{o}$, and this is a minimum. Several minimization algorithms are currently in use for determining the minima of a multivariable function. The method of Steepest Descent[109] is the most primitive of these. In a molecule with *n* atoms, partial derivatives of the energy are calculated numerically by computing the change in energy with respect to the 3*n* cartesian coordinates individually, i.e.,

$$\frac{\delta U(x_1,y_1,z_1,\ldots,z_n)}{\delta x_1}=\frac{U(x_1+\Delta x_1,y_1,z_1,\ldots,z_n)-U(x_1,y_1,z_1,\ldots,z_n)}{\Delta x_1} \qquad \textbf{(1.59)}$$

for all 3n coordinates. The coordinates are then incremented by a constant times the partial derivatives, i.e.,

$$U_{new}=U(x_1+k\frac{\delta U}{\delta x_1},y_1+k\frac{\delta U}{\delta y_1},z_1+k\frac{\delta U}{\delta z_1},\ldots,z_n+k\frac{\delta U}{\delta z_n}) \qquad \textbf{(1.60)}$$

The procedure is continued until the energy change, U_{new} - U_{old}, is positive, i.e., the energy is no longer decreasing, at which point a new set of partial derivatives is calculated and the procedure repeated until the energy no longer decreases and a minimum is achieved.

The Conjugate Gradient method[110] is a more sophisticated version of the method of Steepest Descent in which the 'memory' of the previous step is used in calculating the increment added to the coordinates. The increment δ is given by

$$\delta_i = -\nabla_i + \delta_{n-1}\left[\frac{\nabla_i}{(\nabla_i)_{n-1}}\right]^2 \qquad \textbf{(1.61)}$$

Where ∇_i is the partial derivative of U with respect to the *i*th coordinate and the subscript *n*-1 indicates the quantity from the previous step.

The Conjugate Gradient method and the method of Steepest Descent are both first derivative methods, i.e., they utilize the first derivative to determine the magnitude and direction of the increment. An example of a method that employs the second derivative is the Newton-Raphson algorithm[111]. If the gradient of the potential is expanded as a truncated Taylor series about an initial starting point $(x_1^o, y_1^o, z_1^o, \ldots z_n^o)$

$$\begin{gathered} g(\bar{X}^o) \cong g(\bar{X}^o) + g'(\bar{X}^o)(\bar{X}^o - \bar{X}) \\ \bar{X}^o = (x_1^o, y_1^o, z_1^o, \ldots z_n^o) \end{gathered} \qquad \textbf{(1.62)}$$

The vector $\bar{X}^o$ has been introduced to simplify the equation. The step size $\Delta\bar{X}$ can be written as

$$\Delta\bar{X} = \bar{X}^o - \bar{X} = \frac{g(\bar{X}^o)}{g'(\bar{X}^o)} \qquad \textbf{(1.63)}$$

or, using the relationship $g(\bar{X}^o) = U'(\bar{X}^o)$, this can be written as

$$\Delta\bar{X} = \bar{X}^o - \bar{X} = \frac{U'(\bar{X}^o)}{U''(\bar{X}^o)} \qquad \textbf{(1.64)}$$

The advantage of the Newton Raphson technique is the rapid convergence as the function approaches a minimum, and the second derivatives allow for distinction between true minima and saddle points. However, for a system of *n* atoms, $U''(\bar{X}^o)$ is a $3n$ X $3n$ matrix, and inversion of this matrix is time consuming. Most modeling programs therefore suggest the method of steepest descent or conjugate gradient for

optimization of the initial geometry, and as the structure approaches a minimum, switching to the Newton Raphson minimization.

As the potential energy surface becomes a function of more variables, several minima will occur. The problem of finding the global minimum is now complicated by the presence of other local minima. All the methods described will optimize the input structure to the closest minimum.. This is the largest drawback of computational methods in general - there is no known algorithm for determining the global minimum of a multivariable function.

Computation of the minimum energy conformation must then proceed by sampling the entire conformational surface, or by selecting a starting geometry which presents the most chemical 'sense', i.e., one that is least hindered, or in some cases the geometry of known X-ray structures is used as input. Selection of a starting geometry is prone to bias - the alternative is calculating the energy of all the possible conformations and is impossible for even a medium sized molecule. In light of this, it is easy to see why limiting the degrees of freedom of a molecule is so important, and why calculations of the Hard Sphere type are so attractive to computational chemists. These approximations reduce the dimensions of the potential surface to the number of variable torsion angles in the oligosaccharide.

A common method for sampling the potential energy surface is a grid search - the variable angles, i.e., the Φ, Ψ and ω angles are simultaneously incremented by fixed amounts to span the entire surface, and each of the resulting structures is then minimized. Figure 1.16 is a schematic illustration of a grid search of a hypothetical disaccharide with one ω angle. The potential energy is reduced to a function of three variables $U = U(\Phi,\Psi,\omega)$, and by systematically incrementing each of these, a set of structures spanning a three dimensional grid is generated. A single structure corresponding to the

global minimum can then be located by minimization of each of these structures and comparison of their energies.

This procedure is restricted to smaller structures. As the number of variables increases, the size of the increment used for each variable must also increase. but very soon an unmanageable number of conformations is produced. To overcome this problem, larger oligosaccharides are broken down into their component di- or trisaccharides and each of the optimized blocks is combined to form the final structure.

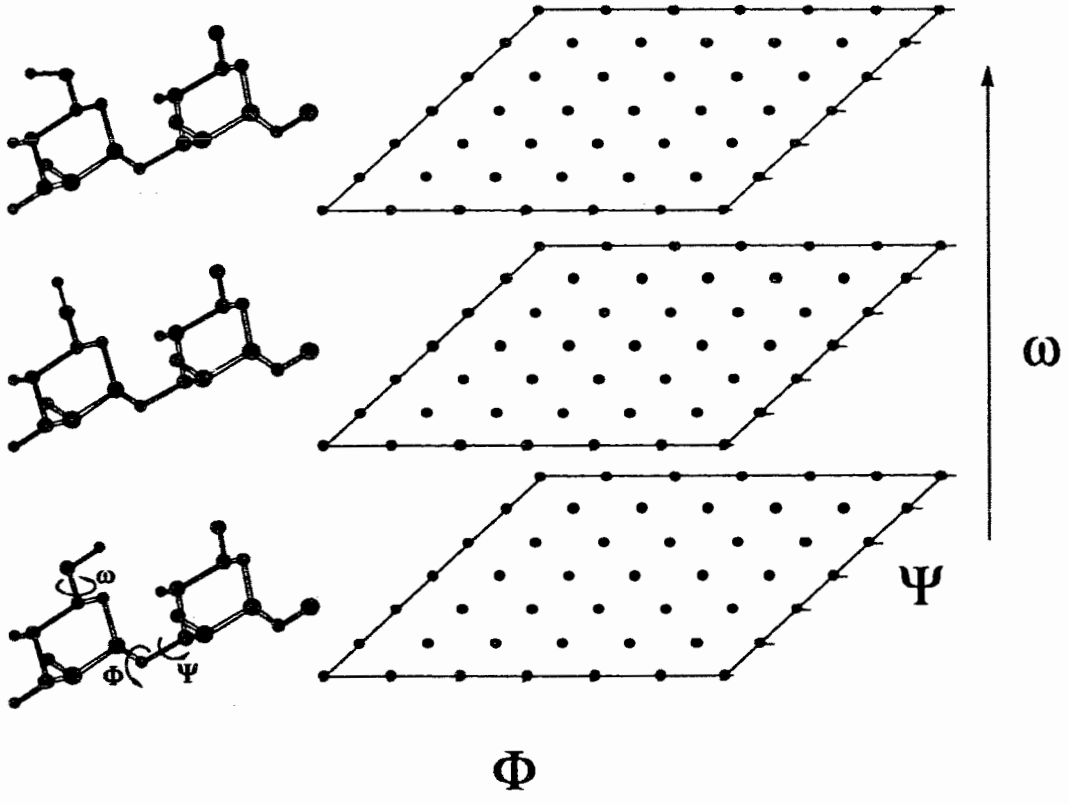


Figure 1.16. Schematic representation of a grid search.

1.5. Conformational Averaging

This approach to determining the minimum energy conformation of a compound has been applied with some success to a vast variety of oligosaccharides[7,77,96,97,112].

However, it was recognized that when multiple low energy conformations of a compound exist, most of the observables of the compound represent an ensemble average, with contributions from each conformer weighted according to its population[113].

1.5.1. Statistical Averaging

Cummings and Carver [113] proposed a statistical mechanics approach to weighting the contributions from energetically accessible conformations to a particular observable. They proposed that the relative populations of a compound follow a Boltzmann distribution. Thus, if the energy of a given conformation is E_i , the population of a given conformation P_i can be calculated by

$$P_i = \exp[-E_i/kT] / \sum_j \exp[-E_j/kT] \qquad (1.65)$$

where k is Boltzmann's constant, and T is the temperature in K. The ensemble average of the observable, denoted ‹x›, can then be calculated by the formula

$$\langle x \rangle = \sum P_i x_i \qquad (1.66)$$

This approach has been tested on several mono-, di- and trisaccharides [113-115]. The Φ and Ψ angles were incremented by 3-10° to produce between 14,400 and 1296 structures for each glycosidic link that was examined. The resulting structures were used to calculate ensemble average NMR parameters, namely proton spin lattice relaxation times (T_1) and NOE values, and a good correspondence was found with the limited experimental data. However, the approach is limited to small molecules - it is clear that for *n* glycosidic linkages, using an increment of x degrees for the grid resolution, $(360/x)^{2n}$ structures need to be calculated. Even for a pentasaccharide, with n=4, the increment would have to be quite large. Another consideration is that it is often not enough to compute the energy of the structure. Some of the rotamers will relax, showing subtle changes in their conformations to accommodate some of the non-bonded interactions. This requires that

every structure be minimized with the torsion angles constrained to points on the surface, which is computationally intensive.

In most of these studies, although the force-field is different, the approach of statistical averaging of conformers was retained. Many disaccharide and trisaccharide calculations have been performed using this method, but an indication of the limiting nature of the statistical mechanics approach is the paucity of calculations on larger oligosaccharides.

1.5.2. Monte Carlo Methods

A more sophisticated approach to conformational averaging employs the use of the Metropolis Monte Carlo[116] (MMC) algorithm to calculate an ensemble of states. Monte Carlo simulations were first used to calculate the thermodynamic properties of many body systems and have recently been applied to conformational analysis of carbohydrates[117,118]. The method is based on the change in energy (ΔE) associated with a change in position (or conformation). The extrapolation of the method to conformational analysis is straightforward. To sample conformational states, random steps in Φ and Ψ are taken. The change in energy $\Delta E = \Delta E_{new} - \Delta E_{old}$ caused by the change is calculated. If $\Delta E < 0$, the new conformation is accepted. If $\Delta E > 0$, the new conformation is accepted if

$$\exp [-\Delta E / kT] < \chi \qquad (1.67)$$

where χ is a random number between 0 and 1, k is Boltzmann's constant, and T is the temperature in K . Therefore the number of states that will be sampled is related to the number of steps, the constant χ, and the temperature at which the simulation is carried out.

Extrapolation of these studies to structure function relationships is limited by the fact that most of these studies are done in vacuum. Conformational studies done in vacuum suffer from the major setback of not being able to mimic the hydrogen bonding and solvation capabilities of a solvent. Since all bioactive polysaccharides are found and perform their biological functions in aqueous environments, the focus of conformational analysis of oligosaccharides has moved towards modeling of carbohydrates in their solvated form. This trend has been accompanied by a subtle shift in molecular modeling from the study of static, low energy structures, to more flexible, conformationally labile models. In addition, the comparison of computed structures was traditionally made with X-ray crystallographic data, whereas now NMR data are generally used to test the accuracy of the model. Since the solution conformation represents a time averaged structure, molecular dynamics [119-122] is used to simulate molecular motion over a period of time. The dynamics trajectories are used to calculate the NMR observables.

1.5.3. Molecular Dynamics

If any atom in a molecule is displaced from its equilibrium position, a force will act on it to return it to its equilibrium position. This is the basis of molecular dynamics. This force F, is related to the potential function V that is used in the dynamics calculations by the equation

$$F_i = -\frac{\delta}{\delta r_i} V(r_1, r_2, \ldots r_n) \qquad \textbf{(1.68)}$$

At time t = 0, random velocities are assigned to the individual atoms using the accessible energy at temperature T, i.e.,

$$\sum_i \frac{1}{2} m_i v_i^2 = \frac{3}{2} N k_B T \qquad \textbf{(1.69)}$$

where N is the number of degrees of freedom, m_i are the masses of the individual atoms, v_i the velocity of each atom and k_B is Boltzmann's constant. The position of the atoms can be calculated by numerical integration of Newton's equation of motion.

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} \qquad i = 1, N \tag{1.70}$$

Here a_i is the acceleration of each atom. Using the algorithm proposed by Verlet[123] the position of the atoms can be calculated by

$$\begin{aligned} r_i(t+\Delta t) &= r_i(t) + v_i(\Delta t) + \frac{1}{2} a_i(\Delta t)^2 \\ &= r_i(t) + v_i(\Delta t) + \frac{1}{2m_i} \frac{\delta V(r_1, r_2, \ldots r_n}{\delta r_i} (\Delta t)^2 \end{aligned} \tag{1.71}$$

One of the first examples of the use of molecular dynamics to model oligosaccharides was published by Homans *et. al.* [124]. The minimum energy conformation of two disaccharides was determined using semiempirical (MNDO) calculations. These structures were then used as starting points for molecular dynamics simulations in vacuo at 300K. The study showed that there were limited torsional oscillations, of the order of ± 20°, about the glycosidic torsion angles, and the average values of the Φ and Ψ angles showed an excellent correspondence to the NMR derived structure.

1.6. Thesis Overview

The primary goal of this research is to model the dynamics of oligosaccharides in solution and provide an appreciation of their 3 dimensional structure, while evaluating the CHARMM force field for use on carbohydrates. Chapter II describes a procedure for correlation of a calculated minimum energy structure with measured NMR parameters. The computational procedure uses molecular dynamics, both in vacuo and in water, to simulate the motion and range of mobility of the molecules tested. Three disaccharides,

α-D-Man*p*-(1→3)-α-D-Man-OMe (**1**), β-D-Gal*p*-(1→4)-β-D-Glc*p*-OMe (**2**), β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-OPr (**3**), are used to evaluate this method. Minima predicted by various force fields are used as starting points for dynamics calculations, and the dynamics trajectories are used to calculate time averaged T_1, NOE and $^3J_{CH}$ values which are compared to experimental data.

Chapter III describes a combined NMR and molecular dynamics approach to model oligosaccharides corresponding to the *Streptococcus* Group A cell wall polysaccharide. Two dimensional (2D) NMR techniques are used to measure NOEs and ROEs in order to obtain quantitative interproton distances. These distances are then used as distance constraints to identify starting points for dynamics simulations. Dynamics are calculated both with and without the constraints and the dynamics trajectories are then used to calculate NOE (ROE) buildup curves, and average three bond coupling constants, $^3J_{COCH}$. A comparison between the calculated and observed NOEs (ROEs) is then performed to test the validity of the method. In Chapter IV the same method is applied to a heptasaccharide corresponding to a fragment of the *Shigella flexneri* Variant Y cell surface lipopolysaccharide.

Chapter II

MOLECULAR DYNAMICS AND NMR SPECTROSCOPIC ANALYSIS OF THE SOLUTION CONFORMATION OF DISACCHARIDES

2.1. Introduction

Force field calculations have met with reasonable success in the prediction of molecular structure and properties[74]. Many molecular mechanics programs are now available, but some suffer from the setback of being best suited to reproduce a particular parameter, while not producing sufficiently accurate results for other important ones. Furthermore, some of the force fields are well parameterized for certain classes of compounds, but are unsuccessful in predicting the conformation of other classes. Obviously, the goal in the design and implementation of a force field is to provide one that will reproduce the geometry and experimentally observed parameters for all classes of compounds.

The choice of a force field to model oligosaccharides is difficult. The HSEA force field[7,96,97] was specifically developed for the modeling of oligosaccharides. However, while providing reasonable agreement with experimental data for some oligosaccharides, it is generally considered an overly simplistic treatment. Information about the behavior of the pyranose rings, for example, is not available from calculations performed using GESA[98]. Another factor influencing the choice of a force field is the applicability of the force fields to other classes of molecules, i.e., to nucleic acids and proteins. As a result, studies on oligosaccharides have employed different force fields, all with varying degrees of success.

The CHARMM force field[104] has been applied extensively to calculation of proteins and nucleic acid structure and has been successful in their simulation. It was

therefore of interest to see how the program would perform in the calculation of oligosaccharide structure. The form of the potential function is shown in Figure 2.1. The bond stretch and angle energy terms are analogous to their MM2 equivalents, but lack the cubic terms. The torsion terms of MM2 are replaced by a single potential function. Asymmetry of a particular torsion potential is therefore reproduced by a combination of van der Waals, electrostatics, and torsion potentials. CHARMM assigns explicit charges to atoms and uses a coulombic potential to describe electrostatic interactions. An additional function, not present in MM2, the improper torsion term maintains planarity of certain atoms.

A requirement of any force field that will be used to model carbohydrates is that it reproduce the anomeric and exoanomeric effects. Molecular mechanics requires experimental data to provide the force constants of the various terms in the potential function. Information about bond lengths and bond angles are easily available from crystal structure data, but experimental data corresponding to the energy barriers associated with the rotation around a torsion angle are seldom available. Thus, force fields generally rely on *ab initio* molecular orbital calculations to provide information on conformational preferences about a torsion angle. A model compound is used to calculate the energy of rotamers at fixed intervals about a torsion angle, and the resulting energy curve represents its torsional potential. The simplest, most commonly used model for the anomeric O-C-O fragment is dimethoxymethane (DMM). Wiberg and Murcko[145] have calculated DMM at the 6-31G* level. However, a better model for saccharides is 2-methoxytetrahydropyran (2-OMe-THP; see Figure 2.2). Due to the increased number of atoms and lack of symmetry, this molecule was optimized using the 3-21G basis set followed by single point energy calculations with the 6-31G* basis set[145].

CHARMM (Chemistry at **HAR**vard **M**acromolecular **M**echanics) - The total energy of the molecule is calculated by the equation

$$V_{TOT} = E_{BOND} + E_{BONDANGLE} + E_{VDW} + E_{TOR} + E_{ELEC} + E_{IMPTOR}$$

E_{BOND} - bond energy term

$$E_{BOND} = k_b (r-r_o)^2$$

$E_{BONDANGLE}$ - Bond angle energy term

$$E_{BONDANGLE} = K_\Theta (\Theta - \Theta_o)^2$$

E_{VDW} - Van der Waals' term

$$E_{VDW} = \sum_{j \neq i} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$

r_{ij} refers to the internuclear distance between atoms r_i and r_j.
Both A_{ij} and B_{ij} are constants specific to an atom type.

E_{TOR} -Torsional energy term

$$E_{TOR} = K_\Phi(1+ \cos(n\Phi - \delta))$$
$$n=1,2,3$$

E_{ELEC} - Electrostatic energy term

$$E_{ELEC} = \sum_{i \neq j} \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}}$$

q_j is the charge assigned to a particular atom.
ε_o is the dielectric constant.
r_{ij} is the internuclear distance between the atoms i and j.

E_{IMPTOR} - Improper torsion term

$$E_{IMPTOR} = K_\omega(\omega - \omega_0)^2$$

Figure 2.1. The potential function of the CHARMM force field[125].

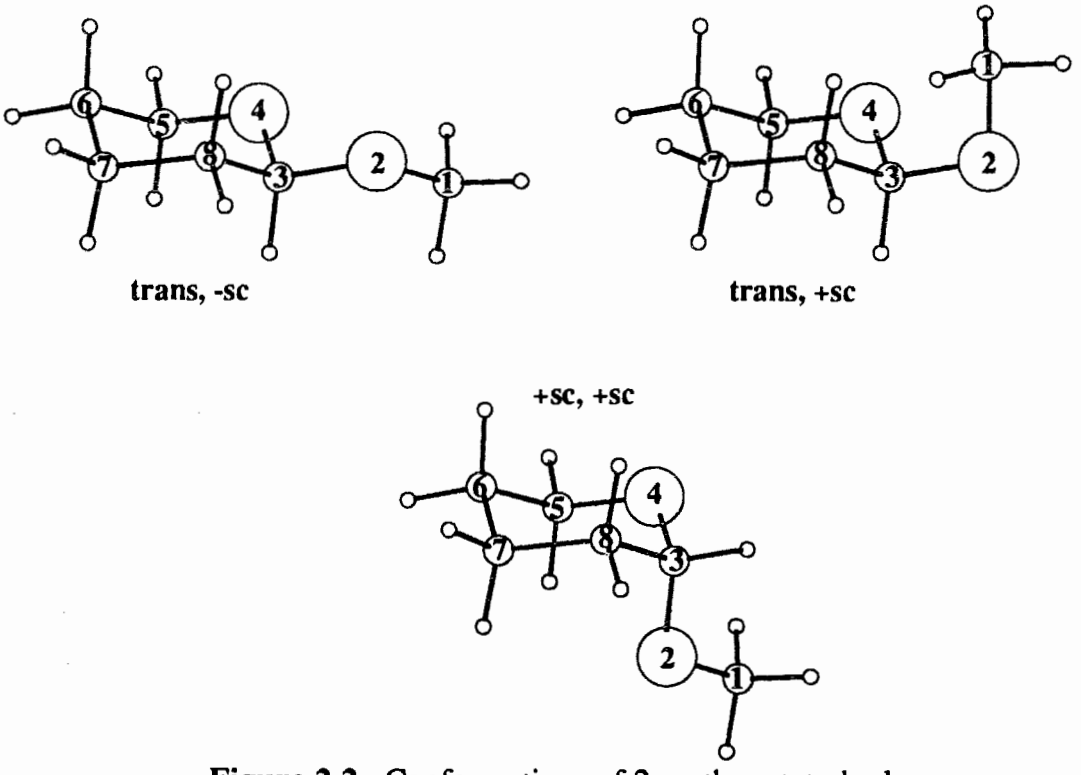


Figure 2.2. Conformations of 2-methoxytetrahydropyran.

Table II.1. Calculated energies and geometries of 2-methoxytetrahydropyran and experimental geometries of methyl pyranosides.

	CHARMM		MM3		*Ab initio*[145]		Experimental[a]	
	trans, -sc	sc, sc	trans, -sc	sc, sc	trans, -sc	sc, sc	β	α
ΔE (kcal/mol)	2.36		0.88		1.33			
Bond Lengths (Å)								
C1-O2	1.41	1.41	1.42	1.42	1.44	1.44	1.43	1.43
O2-C3	1.41	1.41	1.41	1.42	1.40	1.42	1.38	1.40
C3-O4	1.42	1.42	1.43	1.42	1.43	1.42	1.43	1.42
O4-C5	1.41	1.42	1.42	1.42	1.44	1.45	1.43	1.44
Bond Angles								
C1-O2-C3	115.2	114.8	112.5	112.5	115.6	115.2	113.4	113.1
O2-C3-O4	109.4	110.0	109.6	111.9	108.8	111.1	107.9	112.3
C3-O4-C5	114.9	115.5	112.5	114.4	113.8	114.6	111.4	113.4
Torsion Angles								
C1-O2-C3-O4	-46.8	58.9	-74.2	73.6	-56.4	63.6		
C5-O4-C3-O2	178.8	59.7	178.7	61.6	181.3	174.6		

[a]Average values for 8 X-Ray and 3 Neutron studies (β) and 5 X-Ray and 2 Neutron studies (α)[224].

Table II.2. Calculated energies and geometries of 2-methoxytetrahydropyran.

	CHARMM		MM3		*Ab initio*[145]	
	trans, -sc	trans, sc	trans, -sc	trans, sc	trans, -sc	trans, sc
ΔE (kcal/mol)	0.83		2.84		2.82	
Bond Lengths (Å)						
C1-O2	1.41	1.41	1.42	1.42	1.44	1.44
O2-C3	1.41	1.41	1.41	1.41	1.40	1.41
C3-O4	1.42	1.42	1.43	1.43	1.43	1.40
O4-C5	1.41	1.41	1.42	1.42	1.44	1.44
Bond Angles						
C1-O2-C3	115.2	115.2	112.5	116.1	115.6	117.3
O2-C3-O4	109.4	109.4	109.6	110.2	108.8	109.0
C3-O4-C5	114.9	114.9	112.5	112.3	113.8	114.1
Torsion Angles						
C1-O2-C3-O4	-46.8	25.3	-74.2	65.6	-56.4	43.4
C5-O4-C3-O2	178.8	177.9	178.7	171.7	181.3	175.0

It was found that the axial form of 2-OMe-THP was more stable than the equatorial conformation by 1.33 kcal/mol. This value can be compared to the experimental value (0.89 kcal/mol) obtained in $CFCl_3/CDCl_3$ solvent (85/15) by low temperature NMR spectroscopy [84,146].

The axial and equatorial anomers of 2-OMe-THP were used in order to test whether CHARMM would reproduce the anomeric and exoanomeric effects. Calculations were performed on the three low energy conformations of 2-OMe-THP (see Figure 2.2) using both MM3 and CHARMM. The descriptors **+sc** and **-sc** refer to the synclinal or gauche orientation about the C-O-C-O-C units. The results of these calculations are tabulated in Tables II.1 and II.2, together with the experimental values for methyl glycopyranosides, and can be summarized as follows. The bond lengths calculated by MM3 and CHARMM are shorter than the experimental values for methyl α- and β- pyranosides[224]. The maximal difference in the bond lengths between the calculated (MM3 and CHARMM) values for 2-OMe-THP and the experimental values for the methyl pyranosides[224] is 0.05 Å. The bond angles from both the MM3 and CHARMM calculations show parallel trends to the *ab initio* result, but none of these calculations matches exactly the trends observed for the methyl α and β glycoside pairs, i.e., a difference of approximately 2° in the endocyclic C5-O4-C3 angle and 4° in the O4-C3-O2 bond angle. The maximal difference in the bond angles between the calculated (MM3 and CHARMM) and the experimental values[224,225] is 4°.

Of the two molecular mechanics calculations, only MM3 reasonably reproduces the axial/equatorial experimental energy difference in solution. CHARMM predicts the conformational preferences, but the magnitude of the energy difference does not coincide with either the experimental[84,146] or the *ab initio* results[145]. The axial form is more stable than the equatorial form, but the difference in energies of the two conformers is

calculated to be ≈ 3 kcal/mol, which is high compared to the experimental value of 0.89 kcal/mol [84,146].

With regard to the exoanomeric effect, the energy difference between the two lowest energy conformers, i.e., the trans, -sc and the trans, +sc conformations of the equatorial form of 2-OMe-THP, as calculated by CHARMM, is 0.83 kcal/mol, which is low compared with the *ab initio* result of 2.82 kcal/mol, as well as the MM3 calculated energy difference of 2.84 kcal/mol. The free energy $\Delta G°$ for 2-OMe-THP has been shown to result mainly from the $\Delta S°$ component, the enthalpy, $\Delta H°$, being approximately zero[226]. Praly and Lemieux[227] have attributed this to specific interactions between the solvent and the solute. It is clear that solvent interactions play a significant role in this equilibrium, and thus, parameterization of a force field to match either gas phase experimental data or *ab initio* data may not present the best solution. It is also noteworthy that Tvaroska has shown, by means of *ab initio* calculations including the electric field effects, that the trans, +sc conformation is closer in energy to the trans, -sc conformation than predicted on the basis of the previous *ab initio* calculations[228].

Given the fact that no one force field matched the experimental data, and given that CHARMM was one of the programs available that could calculate the molecular dynamics trajectories of compounds in water solvent, it was decided to proceed with calculations based on the CHARMM force field. The compounds considered are of fixed configuration, and thus the irreproducibility of the experimental axial-equatorial energy difference will not affect the outcome of the calculations. However, the minor differences in geometry caused by torsions about the exocyclic C-O bonds will affect both intra ring and inter ring NOEs. In order to assess the severity of these effects, a monosaccharide, methyl β-D-mannopyranoside, was optimized using both CHARMM and MM3, and NOEs were calculated for both structures with a program that utilizes a complete relaxation matrix (see experimental). The endocyclic O-C-O bond angle,

which showed the greatest variation in the earlier comparison of 2-OMe-THP was 113.2° in the structure optimized with MM3 and 115.0° in the structure optimized with CHARMM. A change in this angle will have the largest effect on the distances between the H1 and H5 protons of the ring. However, the largest variation (15 %) was observed in the H5-H3 NOE, which falls within the experimental error (Table II.3). The variation in geometry will thus be undetectable by NOEs.

Table II.3. Calculated NOEs for methyl α-D-mannopyranoside optimized with CHARMM and MM3.

A. NOEs upon saturation of H1.

	H1	H2	H3	H4	H5	H6	H6'	HMe
CHARMM		0.18	0.04	0.00	0.11	0.00	0.00	0.03
MM3		0.18	0.03	0.00	0.13	0.00	0.00	0.03

B. NOEs upon saturation of H5.

	H1	H2	H3	H4	H5	H6	H6'	HMe
CHARMM	0.12	-0.04	0.13	0.09		0.05	0.02	0.00
MM3	0.14	-0.04	0.11	0.07		0.04	0.03	0.03

Molecular dynamics represents an alternative to statistical averaging. The conformation corresponding to a local or global minimum can be used as the starting structure for the dynamics calculations. Once the global minimum has been identified by scanning the conformational surface using a grid search, which can be performed using a fairly large increment, dynamics simulations can be used to model the behavior of the compound near the minimum. Ideally, the distribution of conformations should cover all the accessible states within the potential energy well, and when multiple minima exist

the dynamics should sample the different populations. This would afford the same correlation between the independent variables, i.e., the Φ, Ψ and ω angles, as statistical averaging, and is far less computationally demanding.

We proposed to use molecular dynamics to simulate accessible energy states, and compare certain NMR parameters, namely, proton NOEs, T_1s, and the interglycosidic coupling constants, $^3J_{CH}$, calculated from the dynamics trajectories to those experimentally observed. Three disaccharides - α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**), β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**) and β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**) - were chosen as molecules with which to test this method. These compounds are shown in Figure 2.2.

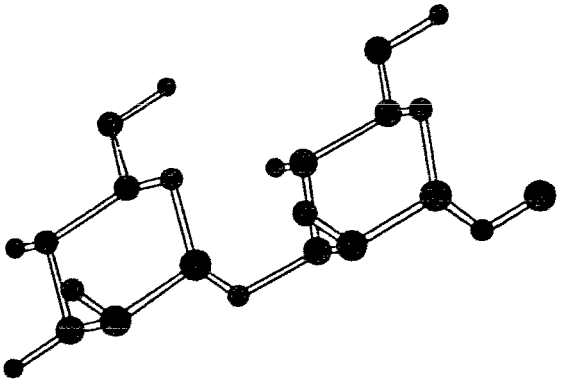
The disaccharide α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**) corresponds to an important branch point in the family of N-linked high-mannose cell surface oligosaccharides[1]. Consequently, this compound has been the subject of several studies and there is substantial experimental[126-134] and computational[92,115,124,126-135] data, available with which to compare the results obtained from any further study. Similarly, β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**), commonly known as methyl-β-D-lactoside, has also been investigated at great length[136-142]. Although the ^{1}H NOE and T_1 data are limited, studies on specifically ^{13}C-labeled methyl-β-D-lactoside[141] have yielded the $^3J_{CH}$ values, which provide another point of comparison. The disaccharide β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**) is a fragment of the *Streptococcus* Group A cell surface polysaccharide. This organism is a major focus of ongoing research in our laboratory, and thus presented an excellent starting point for the study of larger oligosaccharide fragments of the Group A polysaccharide.

2.2. Research Objectives

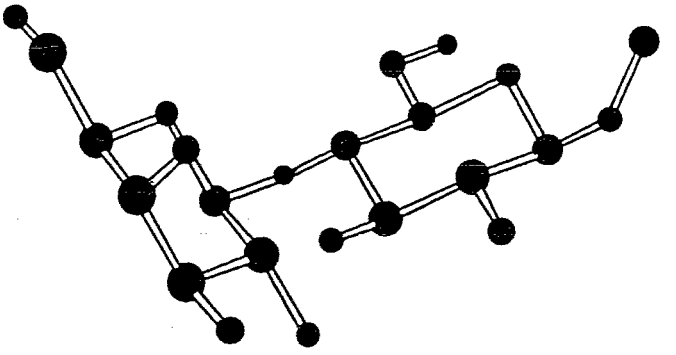
The objectives of this research project are to:

1) Test the ability of the CHARMM force field to model oligosaccharides by calculation of the three disaccharides and comparison of the results with NMR data and results obtained using other force fields.

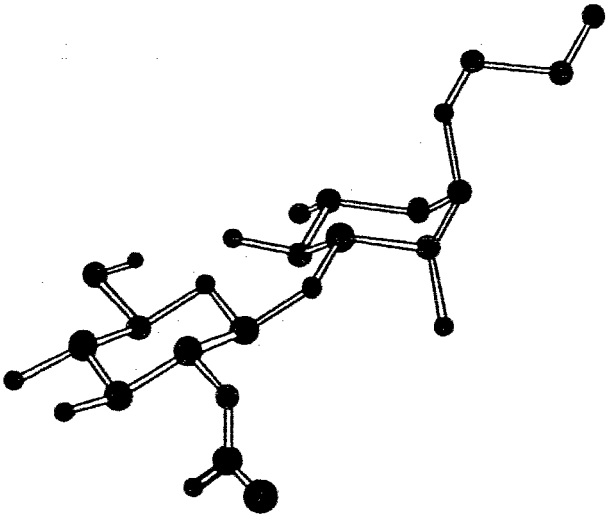
2) To develop a protocol to model oligosaccharides using molecular dynamics simulations, and to determine the validity of molecular dynamics as a procedure for conformational averaging.



α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (1)



β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (2)



β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (3)

Figure 2.3.

The three disaccharides α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**), β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**) and β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**).

2.3. Experimental

2.3.1. Computational

Computations were performed on an SGI-4D25 using QUANTA[143], a commercially available graphics interface to CHARMM. The version of CHARMM used in this study was charmm21r2, with the standard parameter set PARM30.

2.3.1.1. Molecular Mechanics Calculations

In addition to the Φ and Ψ and ω angles, it is convenient to define one other torsion, the θ angle, which corresponds to the H2-C2-N-H angle of the N-acetamido group in the Glc*p*NAc. In compounds **1** and **2**, in order to avoid ambiguity between the two ω angles, **ω1** is defined as the angle of the aglycon, and **ω2** is the torsion of the nonreducing ring. The **ω1** and **ω2** have local minima in each of the *gg*, *gt* and *tg* conformations. The conformations of the disaccharides are referred to as, for example, *gtgt*, where the first two letters correspond to the **ω1** angle, and the second two letters to the **ω2** angle.

To locate the global minimum, the following strategy was adopted for each of the disaccharides. Initially, 9 starting conformations, corresponding to the local minimum of each pendant primary hydroxyl (and in the case of the Glc*p*NAc, the N-acetamido) group were generated. Each of these rotamers was then used as a starting structure in a grid search where the Φ and Ψ angles were incremented simultaneously by 30°. This is exactly equivalent to simultaneously incrementing the Φ, Ψ angles by 30° and the ω or θ angles by 120°. In total, 144 x 9 structures were generated as starting points in a 4 dimensional grid. The methyl and propyl aglycons were placed in the respective local minima in accord with the exoanomeric effect[78,79], and were not varied.

The 144 structures in this grid were completely minimized using the Powell [110] algorithm till the gradient of the force was < 0.0001 kcal/(mol Å). The minimum energy structure was selected and a relaxed map was calculated by incrementing the Φ and Ψ angles in 10° steps to produce a total of 1296 conformations, which were minimized with the Φ and Ψ angles constrained to their initial values. Contour maps and percentage maps were calculated from the results of this calculation. Percentage maps were calculated using the relationship described in equation 1.65,

$$P_i = \exp[-\Delta E_i / kT] / \sum_j \exp[-\Delta E_j / kT] \qquad (1.65)$$

2.3.1.2. Molecular Dynamics Simulations

The minimum energy structure from the grid search was selected and used as an input structure for molecular dynamics simulations, which were calculated in vacuo and in a water box of 15 Å dimension. For the simulations in H_2O the structure was centered in the water box and reminimized. Dynamics simulations were performed using the Verlet[123] algorithm. The protocol followed for a dynamics simulation involved 4,980 steps of heating from 0 K to a final temperature of 300 K. Heating was followed by 10,000 steps of equilibration to achieve constant temperature. The dynamics simulation was then run for 10,000 to 30,000 steps with a time step of 1 fs for the numerical integration and coordinates output every 10 fs, resulting in dynamics trajectories of 10-30 ps. Dynamics simulations of **3** were carried out for longer periods of 100,000 200,000 and 300,000 steps (100, 200 and 300 ps), and no transitions were observed between the various minima. The coordinates from the dynamics simulations were then input into programs (see Appendix) to calculate the NOE, T_1 and $^3J_{CH}$ values, thus averaging the parameters.

GESA calculations were performed both without and with the DIPO option to include explicitly the Burkert torsional term for the exoanomeric effect. The minimum energy structures obtained from the GESA and MM2 calculations were used as starting structures for dynamics simulations with CHARMM after the Φ and Ψ angles were constrained to the input values. NOE, T_1 and $^3J_{CH}$ values were calculated as described previously.

2.3.1.3. Calculation of T_1s and NOEs

NOEs and T_1s were calculated using programs written by Chris Schafmeister (See Appendix). Four principle assumptions were made.

- Relaxation occurs primarily through dipole-dipole interactions and contributions of other pathways are negligible.
- A single rotational correlation time τ_c is assumed, i.e., the molecule tumbles isotropically with no preferred axis.
- The rotational correlation time τ_c is of the order of 10^{-10} s for the molecules studied.
- Exchangeable hydrogens do not participate in the relaxation process.

For each of the 1,000-3,000 frames generated by the dynamics simulation, the distance r_{ij} between protons r_i and r_j ($j \neq i$) was calculated, and the average r_{ij}^{-6} value, i.e. $\frac{1}{n}\sum_{i \neq j}\frac{1}{r^{-6}}$, where n is the number of dynamics frames, was computed. T_1s were then calculated by the formula[56]

$$\frac{1}{T_1} = \left(\frac{\mu_o}{4\pi}\right)^2 \frac{\gamma^4 \hbar^2}{10}\left[\tau_c + \frac{3\tau_c}{1 + (\omega_o \tau_c)^2} + \frac{6\tau_c}{1 + (2\omega_o \tau_c)^2}\right] \sum_{i \neq j} r_{ij}^{-6} + R^s$$

where

r_{ij}	is the internuclear distance between protons i and j.
τ_c	is the isotropic rotational correlation time.
μ_o	is the magnetic permeability of free space.

γ is the proton gyromagnetic ratio.
R^s is the relaxation rate due to other mechanisms
$\hbar$ is Planck's constant divided by 2π.

NOEs were calculated using the equations defined by Noggle and Schirmer[63]. The enhancement of proton H_i on saturation of H_s denoted as f(i, s), can be calculated by a coupled set of equations,

$$R_i f(i,s) + \sum_{k \neq i} \sigma_{ki} f(k,s) = \sigma_{is}$$

here, $R_i = \frac{1}{T_1}$ of proton i, and

σ_{ki} is the cross relaxation rate between H_k and H_i which is given by

$$\sigma_{ki} = \left(\frac{\mu_o}{4\pi}\right)^2 \frac{\gamma^4 \hbar^2}{10} \left[\frac{6\tau_c}{1 + (2\omega_o \tau_c)^2} - \tau_c\right] r_{ki}^{-6} \quad \text{for } i \neq k$$

These equations were solved using Gaussian elimination by a program which incorporated a subroutine taken from the IMSL (International Mathematics and Statistics) library.

In order to test the program, single point calculations of methyl α-D-mannopyranoside and methyl β-D-manopyranoside were carried out. The results were then compared to previously published experimental results and are presented in Table II.4. The observed values in each Table were obtained from Brisson et. al.[126]. When calculating the T_1 and NOE values, the following constants were used: τ_c=0.5x10^{-10} seconds, ω_0 = 360 MHz, R_s = 0.03 (also obtained from Brisson et. al.[126]).

Table II.4. Calculated and observed NOEs and T_1s for methyl D-mannopyranoside.

A. Calculated and observed T_1s (s^{-1}) for methyl α-D-mannopyranoside.[a] The average error for the observed values is ±10%.

	H1	H2	H3	H4	H5	H6	H6'	HMe
Obs	2.5	2.3	1.9	2.3	2.3	0.70	0.70	1.3
Calc	1.98	2.33	2.36	2.92	1.79	0.67	0.65	1.2

Table II.4. (Continued)

B. Calculated and observed absolute NOEs for methyl α-D-mannopyranoside NOEs upon saturation of H1. The average error for the observed values is ±20%.

	H1	H2	H3	H4	H5	H6	H6'	HMe
Obs		0.13						0.03
Calc		0.22	-0.03	0.01	0.01	0.00	0.00	0.03

C. Calculated and observed T_1s (s^{-1}) for methyl β-D-mannopyranoside.[a] The average error for the observed values is ±10%.

	H1	H2	H3	H4	H5	H6	H6'	HMe
Obs	1.1	1.9	1.6	2.9	1.3	0.70	0.70	1.4
Calc	1.46	2.28	1.94	2.84	1.60	0.60	0.60	1.30

D. Calculated and observed absolute NOEs for methyl β-D-mannopyranoside upon saturation of H1. The average error for the observed values is ±20%.

	H1	H2	H3	H4	H5	H6	H6'	HMe
Obs		0.15	0.08	0.02	0.15			0.05
Calc		0.18	0.04	0.00	0.13	0.00	0.00	0.03

E. Calculated and observed relative NOEs (/H1) for methyl-β-D-mannopyranoside upon saturation of H2.

	H1	H2	H3	H4	H5	H6	H6'	HMe
Obs	1		1.6	0.3	-0.6		0.4	
Calc	1		1.39	0.05	-.23	0.00	0.00	0.00

F. Calculated and observed relative NOEs (/H1) for methyl-β-D-mannopyranoside upon saturation of H5.

	H1	H2	H3	H4	H5	H6	H6'	HMe
Obs	1	-0.2	0.6	0.3		0.3	0.06	
Calc	1	-0.31	0.95	0.46		0.39	0.07	-0.01

a. For methyl β-D-mannopyranoside (Φ= 50, ω=60),
for methyl α-D-mannopyranoside (Φ= -50, ω=60).

The calculated values of both T_1 and NOE also match those that were calculated using the full-relaxation NOE program that is part of the GESA program. Similar values have been calculated by Brisson and Carver[126]. The calculated NOE's also display the "three spin effect", as indicated by the negative NOE calculated for H2 on saturation of H5 in methyl β-D-mannopyranoside (Table II.3F).

2.3.2. NMR

Sample Preparation: The samples (5 mg) were dissolved in 99.98% D_2O and lyophilized five times, followed by five freeze thaw cycles under vacuum to remove dissolved oxygen. The tubes were then sealed under vacuum. All NMR spectra were recorded on a Bruker AMX spectrometer operating at a ^1H frequency of 400 MHz. Experiments were conducted on samples of the β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**) and β-D-GlcpNAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**) prepared by Dr. Jose Marino-Albernas and Dr. Kerry Reimer[144], respectively. β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe), ^{13}C-labeled selectively at the anomeric carbon of the glucose ring was a gift from Dr. A. S. Serianni.

T_1s were measured using a standard inversion recovery (180 - τ - 90 - Acq) sequence. The data were then fit to the equation

$$Mz(t) = Mo\, e^{(-\tau / T_1)}$$

using a three parameter fit with software provided by Bruker. Steady state NOE experiments were performed in the difference mode on non-spinning samples in order to ensure good subtraction of the FIDs. The frequency of irradiation was moved alternately from the resonance to be saturated to 7,000 Hz upfield from the center of the spectrum, and the FIDs were subtracted from each other. Typically, 1,024 transients were recorded for each irradiated resonance.

2 4. Results and Discussion

2.4.1. α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe)

The Φ and Ψ angles and the two ω angles subtended by the primary hydroxyl groups of both mannose rings present four variable angles in the mannobioside α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**).

The results of the grid search for compound **1** are shown in Table II.5. The lowest energy structure from the grid search was then selected. Initially, the *gtgg* conformation was found to be the lowest in energy. Further manipulation of the torsions about all the hydroxyl groups yielded the *gtgt* conformation as the global minimum. Table II.5 lists the lowest energy structures resulting from each starting conformer.

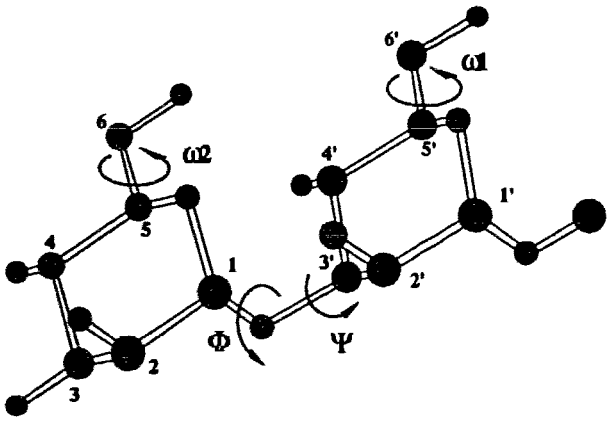


Figure 2.4. The four variable angles, Φ, Ψ, ω1 and ω2 of α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**).

Table II.5. Relative energies of the minimum energy conformations of α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**) obtained from a grid search.

Structure	Φ	Ψ	ω1	ω2	Relative Energy (kcal/mol)
gtgt	-29.2	51.4	46.1	42.2	0.00
gtgg	-30.4	53.2	44.5	-32.6	1.47
gggg	-48.5	-11.9	-49.1	-33.2	3.13
gggt	-31.5	52.8	-58.1	43.7	3.51
tggt	-30.6	52.6	154.9	42.2	4.50
tggg	-31.0	54.9	155.0	-32.5	6.19
gttg	-29.4	55.0	44.5	163.3	6.87
ggtg	-29.3	59.4	-51.7	161.6	10.94
tgtg	-30.2	56.5	155	163	11.57

Table II.6. Comparison of the minimum energy structure of α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**) calculated using CHARMM with the minima derived from other molecular mechanics methods.

Method	Φ	Ψ	ω1	ω2
HSEA[126]	-50	-10	60	60
GESA[a]	-20	44	60	60
MM2CARB[115]	-41	57	177	-63
MNDO[135]	-20	50	60	60
CHARMM	-29	51	46	42
Crystal Structure[b]	-	-	-65	
Crystal Structure[c]	-57	-19	66	81

a. Calculations performed using the Burkert dipole term to include the exo-anomeric effect.

b Methyl α-D-mannopyranoside[133].

c. α-D-Man*p*-(1-3)-β-D-Man*p*-(1-4)-α-D-Glc*p*NAc[134].

Although the *gtgg* conformation is only 1.47 kcal/mol higher in energy than the *gtgt* conformation, the Φ and Ψ angles of the two conformations are the same, and no difference would be observed in the intra-ring NOE nor the $^3J_{CH}$ values.

A comparison of the Φ, Ψ, ω1 and ω2 of the minimum energy conformation of α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**) obtained using various force fields shows some similarities in the potential surface predicted by the various force fields. The minimum energy structure calculated using the HSEA program[126] without the explicit inclusion of the Burkert dipole term to reproduce the exoanomeric effect has a Ψ angle of -10°, which differs from the results of the other force fields used.

Dynamics simulations were performed for the minimized structures as described. The contour plot of the percent distribution calculated by molecular mechanics and the scatter plot illustrating the distribution of conformations sampled during a 10 ps dynamics simulation are illustrated in Figure 2.5. The dynamics trajectories are centered around a Φ, Ψ angle of 0, -50, whereas the percent map shows a maximum at the calculated global minimum of (Φ, Ψ) -29, 51. This is due to the fact that the dynamics simulations are done in H_2O, which disrupts any stabilizing H bonds that occur in the vacuum calculations. The relaxed map indicates a small percentage centered at Φ, Ψ angles of -140°, -10°, and examination of these conformations reveals that this population results from a structure which has the non-reducing mannose ring in a half boat conformation. Since there is no experimental evidence to confirm this, this result was considered to be erroneous and was not included in the calculations.

More recently, Carver et. al.[147] have performed molecular dynamics simulations of α-D-Man*p*-(1→3)-β-D-Man*p*-(1→OMe) using CHARMM with a force field modified for carbohydrates, the PEF422 force field[148,149]. The results of dynamics simulations calculated using this set of parameters have been consistently shown to lead to more flexibility than the standard parameter sets, as evidenced by *ring flips* to the $^1C_4$

conformation during the dynamics simulation of α-D-Glucopyranose[150]. Increased internal motion is observed in the simulation of the mannobioside, and the dynamics trajectories show transitions between two distinct conformations. Although measured NOEs and T_1s at different field strengths provide some evidence for these internal motions[151], it is difficult to quantify the extent to which they occur. No such transition is observed in the dynamics simulations calculated using the standard force field. The discrepancy is evidently caused by the different magnitudes of the force constants.

Metropolis Monte Carlo (MMC) simulations of **1**[117] carried out using the HSEA force field show a distribution around the global minimum of -50°, -10°. The conformational space sampled is similar in dimensions to that observed from molecular dynamics simulation carried out with the PARM30 force field, and grows more diffuse with increasing temperature.

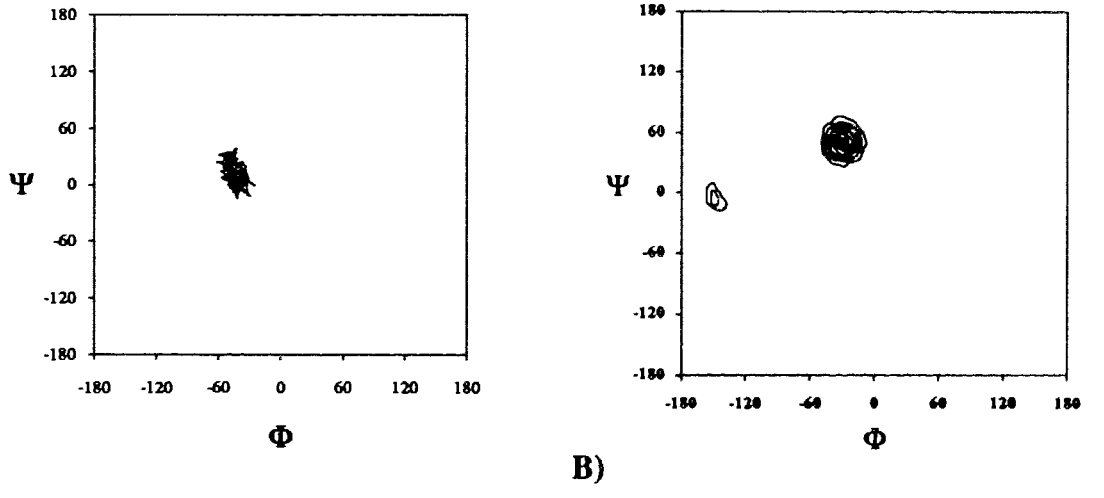


Figure 2.5. A) Plot of the molecular dynamics trajectories in H_2O and B) Percentage map illustrating the population distribution of α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**).

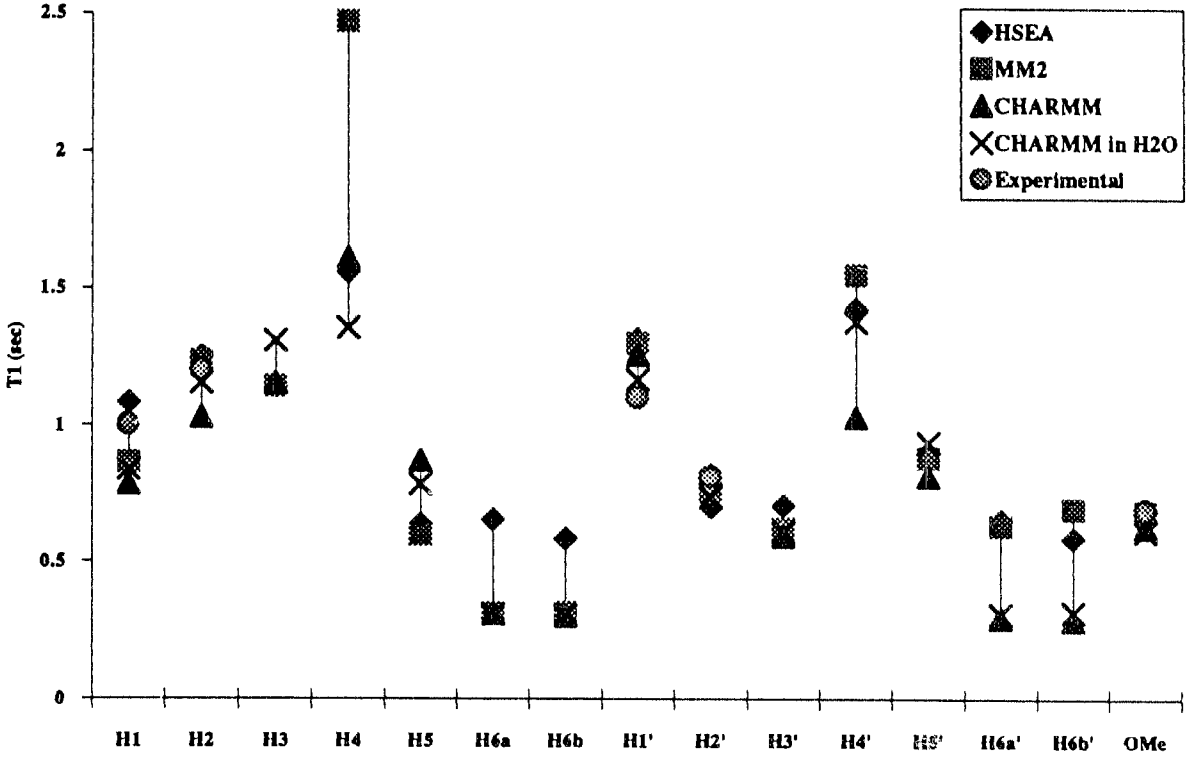


Figure 2.6. A comparison of experimental T_1s[126] and T_1s calculated for different minimum energy structures from their dynamics trajectories for α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**).

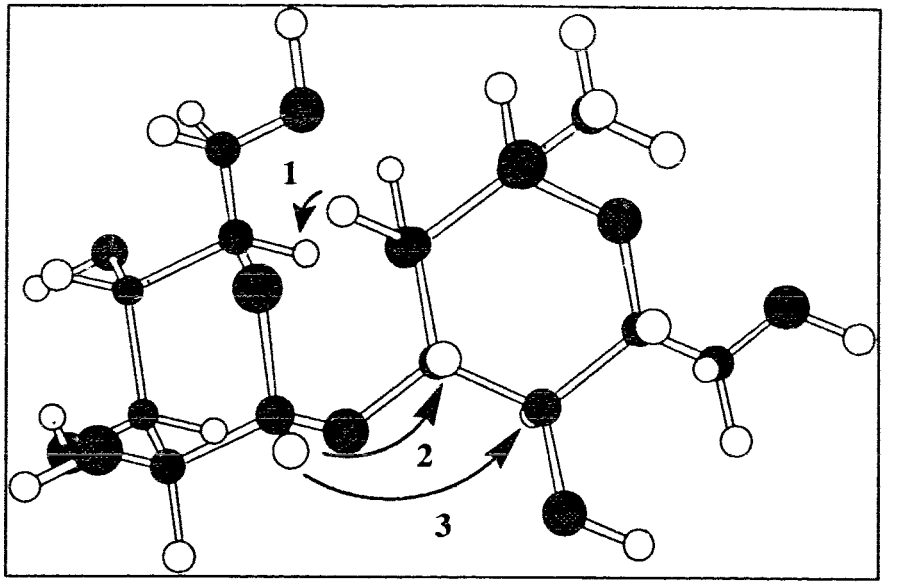


Figure 2.7. NOE contacts observed in α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**)[126]. 1) H2'-H5 2) H1-H3' 3) H1-H4'.

Table II.7. A comparison of the relative NOEs observed experimentally and calculated from the dynamics trajectories of the minimum energy structures of **1**.

Saturated Proton	Observed Proton	Experimental and Calculated NOEs				
		Exp.[126]	GESA[a]	MM2	CHARMM	CHARMM in H_2O
H1	H2(ref)	1.0	1.0	1.00	1.00	1.00
	H2'	0.0	0.03	0.29	0.11	-0.11
	H3'	1.0	0.87	0.97	1.20	1.08
	H4'	0.1	0.11	-0.06	0.08	0.00
H2'	H1'(ref)	1.0	1.00	1.00	1.00	1.00
	H3'	0.9	0.73	0.53	0.74	0.82
	H5	0.6	0.97	0.13	0.18	0.85

[a]GESA minimum energy structure calculated without the Burkert dipole term.

The high degree of overlap in the NMR spectrum of oligosaccharides limits the availability of experimental data. Figure 2.6 illustrates the problem. Only five of the fifteen protons are well separated enough from the rest of the spectrum such that their T_1s can be measured. There is a good fit between the experimental data available and the T_1s calculated using the dynamics trajectories of the various structures. This is shown in Figure 2.6. However, as the figure illustrates, the major divergence between the T_1s calculated for all the structures occurs at H4 of the nonreducing mannopyranose ring. A variation of almost 1 s between the calculated T_1s is observed at this proton, a difference which could possibly distinguish between the structures and help determine which is closer to the actual conformation. Unfortunately, since the chemical shift of the proton in the NMR spectrum was unresolved, no T_1 could be measured for this proton.

The NOEs calculated for the minimum energy structures obtained from the various potential energy calculations are compared with the experimental NOEs [126,128] in Table II.7. Relative NOEs, i.e. the ratio of two NOEs is used for comparison, as suggested by Brisson and Carver [126], since a ratio would be less affected by the errors introduced experimentally. The intra-ring H1-H2 cross peaks are used to calculate the relative NOEs, since these are the least conformationally dependent. There has been much discussion about the conformation of α-D-Man*p*-(1→3)-α-D-Man*p*-(1→OMe) (**1**). Brisson and Carver [126,128] reported observing NOEs between the proton pairs H2'-H5, H1-H3', H1-H4', and H1-H2' (here the prime refers to the aglyconic mannopyranose ring; see Figure 2.7). Controversy arose about the existence of certain critical intra-ring NOEs. The presence of the H2'-H5 peak was disputed by Homans *et. al.* [131,132] However, the observation of this peak in a selectively deuterated mannobioside[129] effectively demonstrated that the H2' of the aglycon is in close proximity to the H5 of the nonreducing ring.

This evidence seems to support the conclusions of Brisson and Carver. The existence of both the H1-H4' and the H2'-H5 cross peaks, in conjunction with the absence of an H1-H2'

cross peak would seem to indicate that both the Φ and Ψ angles are negative, as proposed. The structure calculated to be the minimum energy conformation using the HSEA program, with Φ and Ψ angles of -50°, -10° respectively, places the H1 within 3.5 Å of the H4' and the H2' 2.5 Å from H5. The minimum energy conformations calculated using CHARMM and the various other force fields place these atom pairs much further apart. In addition, in the CHARMM structure, the H1-H2' distance is <3.0 Å, which would produce a large NOE. This is not observed experimentally, a result that is consistent with the HSEA structure in which this distance is > 4.0 Å.

The calculated NOE values for the various minimum energy structures reflects this inconsistency. In most cases a non zero NOE is calculated between the H1 and H2' protons. The NOE values calculated using the HSEA structure account for the H1-H4 cross peak, with a negligible NOE (3%) between the H1-H2' pair. However, the H2'-H5 distance is greatly exaggerated, as indicated by the calculated relative NOE of 97%, compared to the experimental value of 60%.

The analysis of the $J_{5,6}$ and $J_{5,6'}$ coupling constants[127] indicate an approximately 50 : 50 ratio of the *gt* and *gg* rotamers for both mannopyranose rings. Table II.5 shows that the ratio of these conformations is not reproduced by the CHARMM calculations.

The HSEA minimum is identical to one of the minimum energy conformations predicted by CHARMM, namely the *gggg* conformation (Table II.5). This conformation is 3 kcal/mol higher in energy than the global minimum and is the only structure significantly different from the rest of the minima. A comparison of all these minima reveals one significant difference between the minimum derived from the *gggg* conformer and the other minimum energy conformers, namely the hydrogen bonding. The *gggg* minimum has only two intra-ring hydrogen bonds, one between the ring oxygen (O5) and the hydroxyl proton of O6, the primary hydroxyl group, of the nonreducing ring, and the other in the aglycon between the 2'-hydroxyl proton and the ring oxygen. Every other minimum energy

conformation of the mannobioside listed in Table II.5 reveals a hydrogen bond from the glycosidic oxygen to the 2'-hydroxyl of the aglycon, effectively fixing the Ψ angle. A positive Ψ angle causes rotation of this hydroxyl group, placing the 2'-hydroxyl proton distal from this oxygen. This is one example of inter-residue H-bonding that could account for artificially low energies for certain structures. An inter-residue H-bond between the 2 hydroxyl group of the reducing mannopyranoside and the O6 of the adjacent ring has been proposed by Homans et. al[131]. It is interesting to note that the minimum energy structure predicted by GESA with the explicit inclusion of the dipole term, is exactly equivalent to the minimum predicted by CHARMM (Table II.6), and that CHARMM, without the electrostatic term, predicts a minimum energy conformation similar to the HSEA and GESA minimum (data not shown).

Molecular dynamics simulations with the explicit inclusion of solvent may overcome this intramolecular H-bonding, as reflected in the results of the molecular dynamics simulation in H_2O, in which the Φ, Ψ angles move toward the HSEA/GESA minimum (Figure 2.5). The NOEs calculated from the trajectories of the simulations of the mannobioside performed in H_2O best reproduce the experimentally observed NOEs, except for the enhancement of H4' when H1 is saturated, which is calculated to be 0%.

These results indicate that the inclusion of H-bonding in CHARMM can lead to minimum energy structures that may not be represented in solution. Furthermore, since CHARMM occasionally produces anomalous structures, the results of any calculation performed with CHARMM must be carefully analyzed to exclude such results.

2.4.2. β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe)

The four variable angles and the numbering used as reference in β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**) are shown in Figure 2.8.

The results from the grid searches of the nine starting conformations, corresponding to the local minima of the primary hydroxyl groups, are listed in Table II.8. Several of the structures converge to the same minimum. Three of the starting structures - the *ggtg*, *tggg* and *tggt* conformations - produce minimum energy structures that have different conformations about the ω angles. The global minimum as calculated by CHARMM occurs at Φ ,Ψ angles of 163°, 6°. This structure is stabilized by two inter-ring H-bonds, one formed between the O6 of the glucose ring and the 6-hydroxyl of the galactose, and the other from the 3-hydroxyl of the glucose to the O2 of the galactose. By examination and arrangement of the hydroxyl groups, a lower energy minimum is obtained with Φ, Ψ angles of 161° and 6° and **ω1**, **ω2** angles of -47° and 55°, respectively, corresponding to the *gggt* conformation. A torsion search with simultaneous increments of the Φ, Ψ, **ω1** and **ω2** angles confirms this to be the minimum energy conformation.

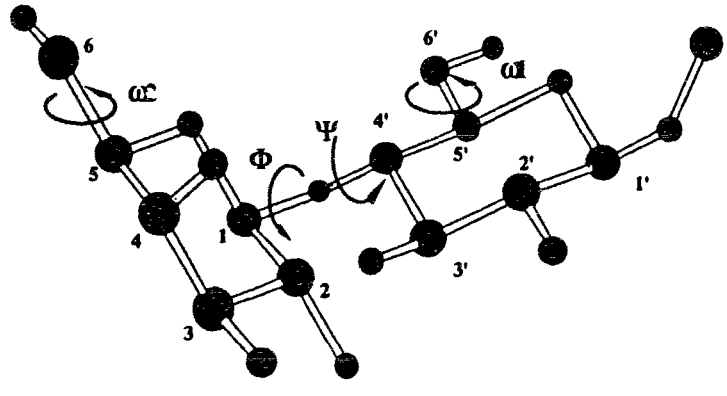


Figure 2.8. The four variable angles, Φ, Ψ, **ω1** and **ω2** of β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**).

Table II.8. Relative energies of the minimum energy conformations of β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**) obtained from a grid search.

Structure	Φ	Ψ	ω1	ω2	Relative Energy (kcal/mol)
gggg	163.8	6.0	-45.6	-47.1	0.00
gtgg	-17.9	-28.2	46.9	-41.7	0.26
gtgt	-15.7	-27.8	47.2	58.5	0.56
gggt	164.9	7.2	-48.1	49.3	1.71
gttg	32.8	171.8	67.1	63.9	3.71
tgtg	28.3	171.6	-38.3	62.4	6.71

Table II.9. Comparison of the minimum energy structure of β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**) calculated using CHARMM with the minima derived from other molecular mechanics methods.

Method	Φ	Ψ	ω1	ω2
GESA	56	-4	72	50
GESA[a]	34	2	72	50
MM2 (85)[136]	24	-59	gg	gt
MM2 (87)	172	0	gg	gt
MM3	44	-54	gg	gt
CHARMM [b]	161	6	-47	55
CHARMM [c]	48	4	64	68
Crystal Structure[d]	49	-12	72	50

a. Calculations performed using the Burkert dipole term to include the exo-anomeric effect.
b. CHARMM calculation performed using all energy terms.
c. CHARMM calculation performed excluding electrostatic terms.
d. β-D-galactopyranosyl-(1→4)-β-D-glucopyranoside [152].

Unlike the mannobioside, which seems to exhibit only one minimum energy conformation with the CHARMM force field, there are several minima obtained for the lactoside. The second lowest energy structure, with Φ and Ψ angles of -18°, -28° has no intra-ring H-bonds. Both the *gttg* and the *tgtg* conformations, with Φ, Ψ angles of 33°, 172° have inter-ring H-bonds. In order to understand the role of H-bonding in determining the minimum energy conformations, it is of interest to ascertain the minimum energy conformation without H-bonding. Unfortunately, CHARMM does not have an explicit H-bonding term, and the entire electrostatic potential has to be "switched off". This may serve to delineate the potential energy surface solely due to steric interactions. We recognize that this procedure drastically alters the net force field. However, the procedure is used only to provide a starting point for the dynamics simulations. The minimum energy structure thus derived has Φ, Ψ angles of 48° and 4°. During the dynamics simulations, this structure is placed in a water box and the dynamics trajectories are calculated utilizing the **entire force field with all potential functions**. This procedure should serve to minimize intramolecular H-bonding and allow the hydroxyl groups to interact with the solvent.

The minimum energy conformations predicted by the different methods are listed in Table II.9. The minimum energy structure calculated using GESA and that predicted by CHARMM without electrostatics are equivalent. A comparison of the several minima predicted by the various programs, i.e., GESA, MM2, CHARMM indicates that all of the methods locate common minima. Extensive research on methyl β-lactoside and its analogs has been carried out using a combination of the HSEA and MM2(85) potential functions[136,137]. A grid search was performed using HSEA, and minima thus located were then optimized using MM2. Four of the six minima in this study have structures similar to those predicted using the combined HSEA/MM2 approach. However, the global minimum predicted with the latter calculations, with Φ, Ψ angles of 24° and -59°

respectively, does not have an analogous structure predicted by CHARMM. The CHARMM global minimum has Φ, Ψ angles of 161° and 6°, respectively.

The results of the molecular dynamics simulations are compared with the percentage distributions as calculated using relaxed maps from molecular mechanics calculations (Figure 2.9). The two plots are almost superimposable, indicating that the dynamics samples the accessible conformational space.

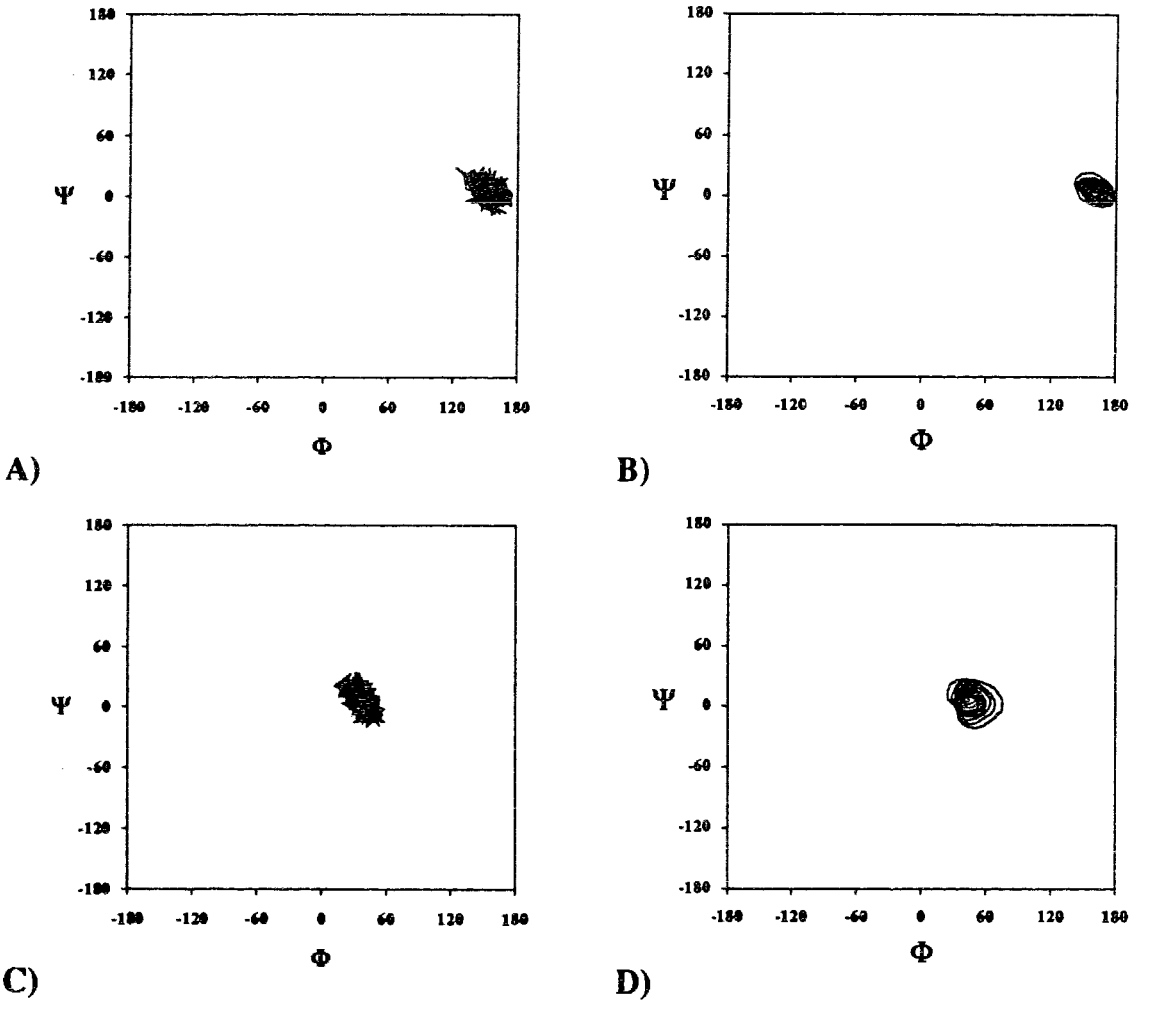


Figure 2.9. A) Scatter plot of the molecular dynamics trajectories in H_2O and B) percentage map illustrating the population distribution of β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**). C) & D) Calculations as in A) and B) without the electrostatic term. Average temperature during the dynamics simulation was 280 K.

T_1s were calculated from the dynamics trajectories of the various structures. A correlation time of 120 ps was used[139,140]. A comparison of the calculated T_1s (Figure 2.10) with those experimentally obtained shows a remarkable correspondence. Consideration of the differing geometries of all the structures indicates the relative insensitivity of the T_1s to the conformation. However, there is a major divergence in the experimental T_1 and that calculated for the H2 of the galactose in the minimum energy structure predicted by CHARMM. Examination of the various conformations shows that variation of the Φ, Ψ angles does have the largest effect on the proximity of this proton to the ring protons of the glucose residue. In the CHARMM structure (Φ, Ψ 160°, 6°) the H4'-H2 distance is < 2.5 Å, and in the GESA structure, this distance is > 4.0 Å, effectively isolating H2. Thus, cross relaxation between H2 and the other protons would be minimal, and this would account for the long T_1 calculated from the GESA structure.

The ^{1}H spectrum of **2** is shown in Figure 2.12. It can be seen that the chemical shift of several key ring protons, i.e., H3', H4', H5' and H3, overlap, making it difficult to observe and quantify NOEs. In addition, the two anomeric protons have very similar chemical shifts, complicating selective excitation. In order to overcome this problem, β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) selectively ^{13}C-labeled at the anomeric carbon of the glucose ring was used. The large J_{CH} between the labeled carbon and the anomeric proton results in the H1 of the glucose being split into a doublet, thus removing the overlap with the H1 of the galactose. The results of a difference NOE experiment are shown in Figure 2.12. The negative peaks in the spectrum are due to residual saturation. A normal spectrum was multiplied by a suitable factor and added to this spectrum to negate the effect of the presaturation on the integrals. The spectrum illustrates the obstacle presented by the overlap between the H3, H4', H5' and to a certain extent H2 chemical shifts. Efforts to deconvolute the spectrum were unsatisfactory and hence the H3', H4' and H2 NOEs had to be integrated together.

The presence of a strong NOE between H1 and H4' excludes the structure predicted by CHARMM with Φ, Ψ angles of 160°, 6° as the major conformation, as this would not place the two protons in close proximity. A second interesting feature of the spectrum is the presence of an NOE between H1 and both H6'$_R$ and H6'$_S$. Observation of these peaks rule out the MM2 minimum, since this conformation will not bring these proton pairs within 3 Å. The coupling constants $J_{5,6S}$ and $J_{5,6R}$ of the glucopyranose are 2.3 and 5.2 Hz respectively and correspond to an approximately 1:1 ratio of the *gg* and *gt* conformations. For galactopyranose the values are 4.0 Hz for $J_{5,6S}$ and 8.2 Hz for $J_{5,6R}$ which are consistent with a predominace (> 70%) of the *gt* rotamer for this residue[136,137,141]. NOEs calculated from the dynamics trajectories of the MM2 and CHARMM minimum energy conformations, and calculated using a weighted average of HSEA structures are compared with the experimental results in Table II.10. The NOEs calculated using the dynamics trajectories of the structure first minimized without the electrostatic term show the best correlation to the experimental data.

Interglycosidic CH coupling constants ($^3J_{CH}$) have been determined both by selectively labeling the C1 and C4' atoms[141], and more recently, at natural abundance by inverse detection[142]. The $^3J_{CH}$ values observed are 4.0 Hz between H1 and C4', and 5.1 Hz between C1 and H4'. Due to the sinusoidal nature of the relationship between torsion angle and coupling constant, there are four possible angles corresponding to a single J value. The observed values of 4.0 Hz and 5.1 Hz thus correspond to values of ±137°/ ±33° and ±148°/ ±17° for Φ/Ψ. Φ angles of ±137° and Ψ angles of ±148° orient the two rings with considerable steric interaction, and are thus unlikely to occur in solution. The values 33°/17° would best correspond to the GESA structure (34°/2° for Φ/Ψ).

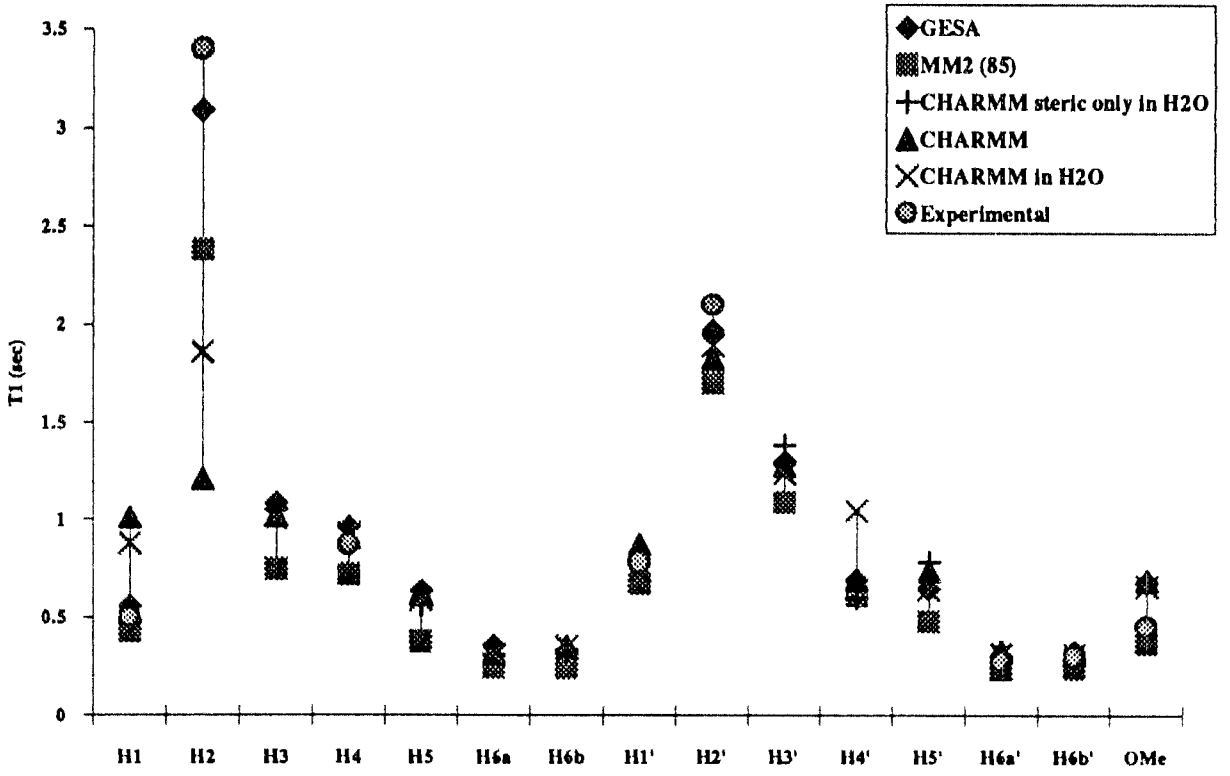


Figure 2.10. A comparison of experimental T_1s and T_1s calculated for different minimum energy structures from their dynamics trajectories for the compound β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**).

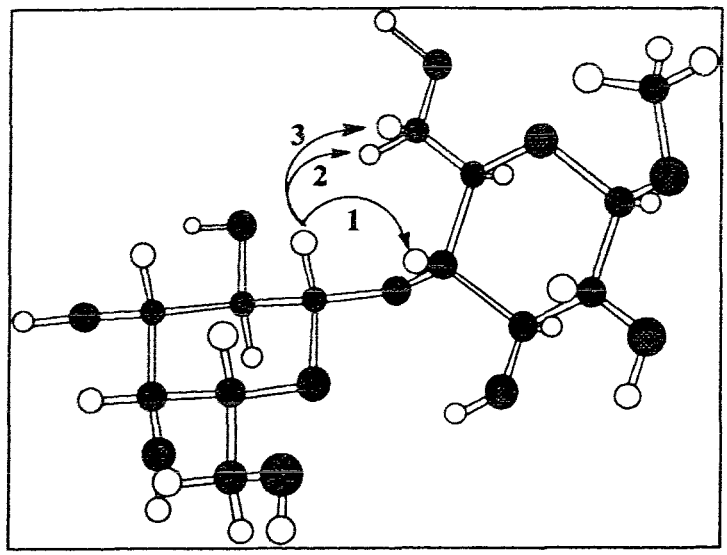


Figure 2.11. NOE contacts observed in β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**). 1) H1-H4' 2) H1-H6'$_R$ 3) H1-H6'$_S$.

Table II.10. A comparison of the observed relative NOEs (±20%) and those calculated from the dynamics trajectories of the minimum energy structures of **2**.

Saturated Proton	Observed Proton	Experimental and Calculated NOEs				
		Experimental	GESA	MM2	CHARMM in H_2O	CHARMM in H_2O[a]
H1	H2+H3+H4'	3.69	3.38	3.40	1.83	3.05
	H5(ref)	1.00	1.00	1.00	1.00	1.00
	H6'$_S$	0.20	0.33	0.00	0.00	0.18
	H6'$_R$	0.15	0.09	-0.09	0.09	0.18

[a]The starting structure for the dynamics was the minimum energy conformation obtained after minimization with the electrostatic term off.

The NOE data, in conjunction with the $^3J_{COCH}$ and T_1 values, seem to indicate that the predominant conformation in solution has interglycosidic Φ/Ψ angles of 50°/4° (±20°), corresponding to the conformation predicted by CHARMM when the

electrostatic term is switched off, which is essentially the same as the GESA minimum energy conformation. This shows some similarity to the geometry of the crystal structure of β-lactoside in which the Φ and Ψ angles are 49° and -12° respectively.

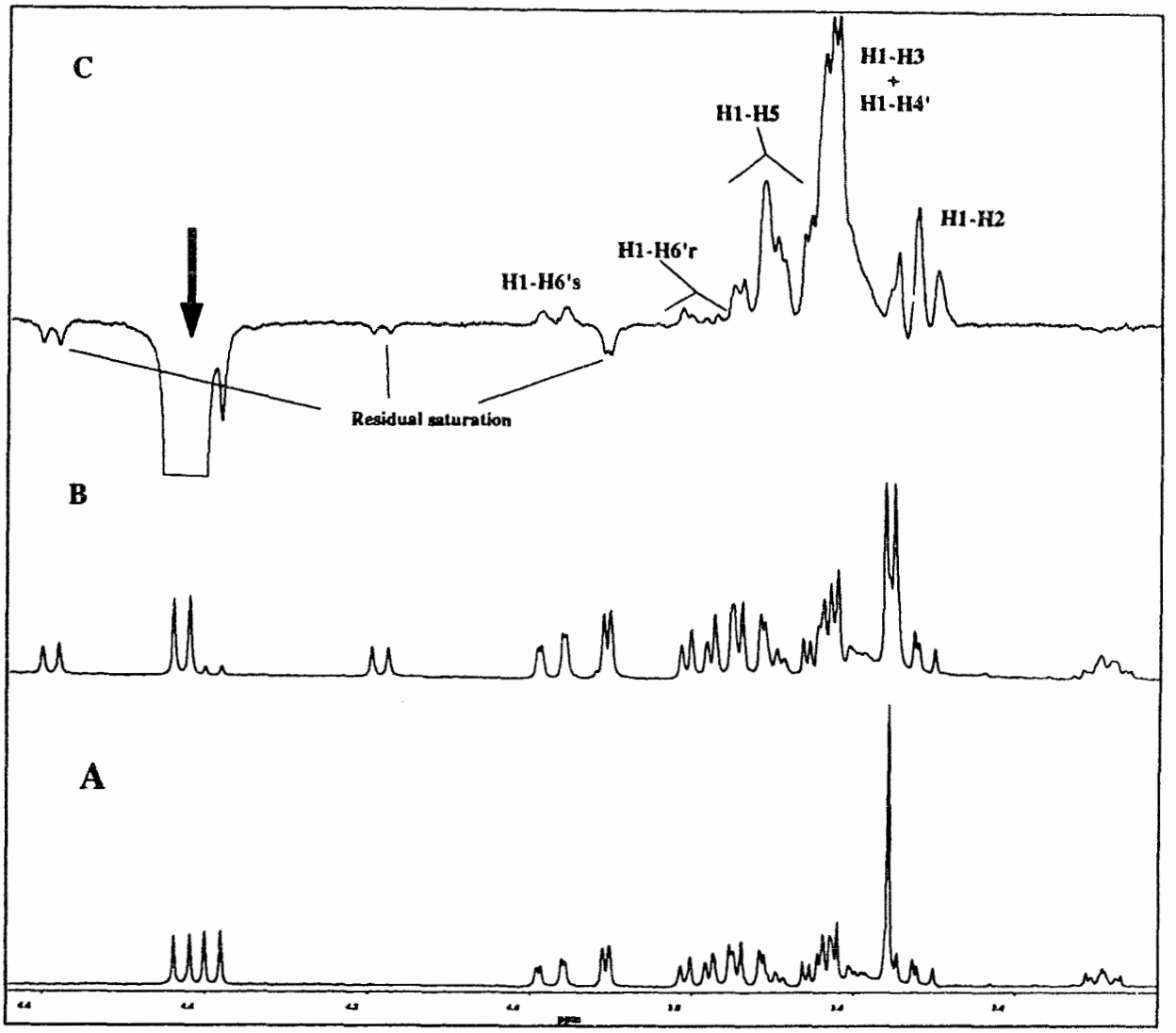


Figure 2.12. A) ^{1}H NMR spectrum of β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**). B) ^{1}H NMR spectrum of β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) ^{13}C-labeled at the anomeric carbon of the glucose. C) NOE difference spectrum of the selectively labeled compound.

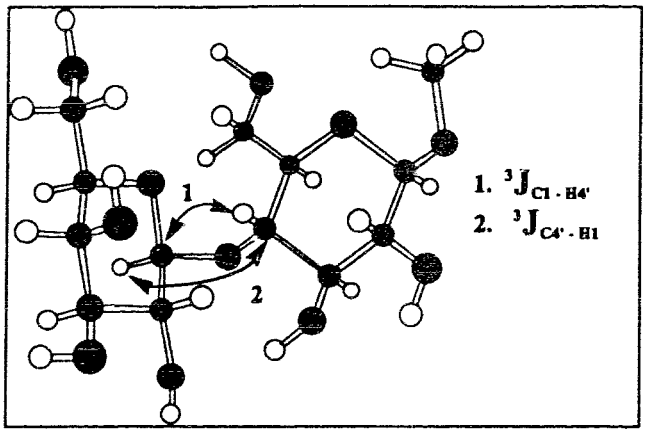


Figure 2.13. The two $^3J_{COCH}$ coupling constants observed in the compound β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→OMe) (**2**).

Table II.11. A comparison of the observed $^3J_{COCH}$ and those calculated from the dynamics trajectories of the minimum energy structures of **2**.

Method	$^3J_{COCH}$	
	C4' - H1	**C1 - H4'**
GESA[a]	3.9	5.6
MM2 (85)[136]	4.8	2.0
MM2 (87)	6.7	5.5
MM3	3.0	2.1
CHARMM[c]	5.5	5.5
CHARMM[d]	3.6	5.4
Crystal Structure[e]	2.5	5.4
Experimental[141,142]	4.0 ± 0.2	5.1 ± 0.3

a. Calculations performed using the Burkert dipole term to include the exoanomeric effect.
b. The MM3 and MM2(87) are single point calculations, i.e., no dynamics simulations were run.
c. CHARMM calculation performed using all energy terms.
d. CHARMM calculation performed excluding electrostatic terms.
e. β-D-galactopyranosyl-(1-4)-β-D-glucopyranoside[152].

2.4.3. β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr)

The disaccharide **3** differs from the compounds **1** and **2** since the aglyconic monosaccharide does not possess a primary hydroxyl group and the terminal aglycon is a propyl group. The Φ, Ψ angles, along with the ω angle of the Glc*p*NAc unit represent three variable torsions. In addition to these three angles, the orientation of the acetamido group, which can be specified by the H2-C2-N-H angle, defined as the θ angle (see Figure 2.14), can potentially influence the conformation. Thus, there are four angles which must be simultaneously varied while performing a grid search. The three rotamers of the θ angle are defined as plus gauche (*pg*) when the θ angle is between 0° and 120°, minus gauche (*mg*) when the angle is between 0° and 120° and trans (*tr*) when the angle is between 120° and -120°. Thus, the conformations corresponding to the local minima in which the ω angle is 60° (*gt*) and the θ angle is 60° (*pg*) can be referred to as the *gtpg* conformation, where the first two letters specify the ω angle and the next two specify the θ angle.

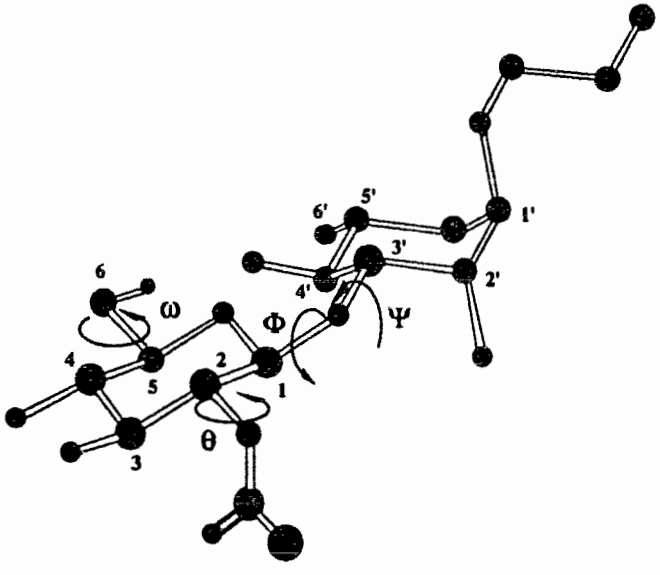


Figure 2.14. The four variable angles, Φ, Ψ, ω and θ of β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**).

The results of the grid search are tabulated in Table II.12. One of the initial conformations, corresponding to the *gtmg* conformation, converges to the same minimum as the *gtpg* conformation. None of the minimum energy conformers exhibit a negative θ angle. However, the θ angle is almost eclipsed in the minimum energy stucture predicted by CHARMM. It is noteworthy that the θ angle differs from the crystal structure of Glc*p*NAc[153] in which this angle is trans, i.e. ~180°. Since the experiments were done in D_2O, the J_{HH} the between the amide NH-proton and H2 could not be measured in order to deduce the θ angle in solution.

There are two distinct minima within 3 kcal/mol of the global minimum at Φ/ Ψ 34°/ -56°. Two additional structures, corresponding to rotamers of the ω and θ angles, but with the same Φ, Ψ angles, are also found within 3 kcal/mol of the global minimum. Examination of these five minimum energy conformations shows that the global minimum, with a θ angle of 5°, has the carbonyl oxygen of the acetamido group extremely close (< 3.0 Å) to the anomeric proton. This would result in a deshielding effect and the chemical shift change which would accompany such an effect is not observed in the ^{1}H NMR spectrum of **3**. Further analysis shows that the *gtpg* and *gttr* conformers both possess three H-bonds, two of which are inter-ring and are possible influences on the Φ and Ψ angles. The ω angle in two of the five minima is in the *gt* conformation and three structures are in the *gg* conformation.

In order to determine the effect of the H-bonds on the stability of the various conformations, once again, a grid search was performed without the electrostatic term. In this case, the θ angle is 145° but the global minimum is found to be the same as that found with all the energy terms. Analysis of this minimum energy conformation reveals only one O-H pair within H-bonding distance, the 2'-hydroxyl proton of the Rha*p* and the ring oxygen of the Glc*p*NAc. This H-bond is found in the global minimum

(Table II.12). It is possible that this orientation of the 2'-hydroxyl group is sterically favored, and this would account for the similar results obtained using both methods.

A comparison with the results from the GESA calculations (Table II.13) shows that the structure predicted by GESA, with both the Φ and Ψ angles positive, is not in the same Ψ locale as the minimum energy conformation predicted by CHARMM. The Φ angle is still small and positive, which keeps the H1 close to the H3', but the H1-H2' distance is increased considerably by the negative rotation of the Ψ angle.

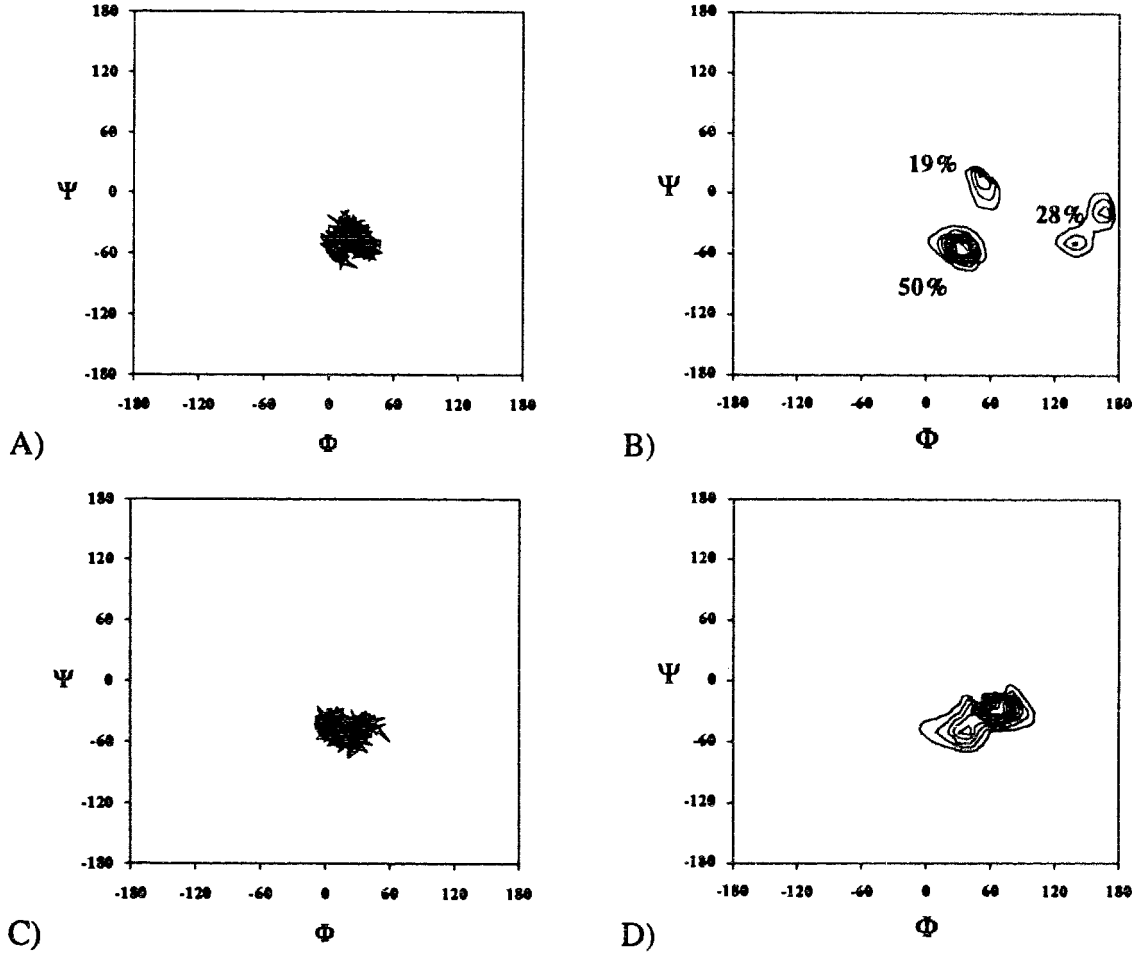


Figure 2.15. A) Plot of the molecular dynamics trajectories and B) percentage map illustrating the population distribution of β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**). C) & D) Calculations as in A) and B) without the electrostatic term. Average temperature during the dynamics was 275 K.

The global minimum obtained using GESA without the Burkert term had, in addition to a positive rotation of the Φ angle by 20°, a Ψ angle that was very close to eclipsed (-4°). This conformation, which is similar to the minimum obtained from the *ggpg* starting structure, with Φ and Ψ angles of 56°, 15°, places the H1 distant from the H2'. Similarly, the trans Φ angle (164°) observed in the minimum obtained from the *ggmg* conformation positions the H1 such that no NOE would be observed across the gycosidic bond to the H3'.

Dynamics simulations of **3** were performed for the CHARMM minimum energy conformations. The percentage map was calculated as described in the experimental section, i.e., the minimum from the grid search was selected and the Φ and Ψ angles were incremented in 10° steps and minimized with the Φ and Ψ angles constrained to their initial starting values. A contour map and percentage map were then calculated. Comparison of the dynamics trajectories and the percentage maps (Figure 2.15) shows that while the simulation effectively samples the conformational space in the region of the minimum energy structure, the trajectories do not sample two of the three populations that are predicted by the grid search. In order to include these populations, the dynamics had to be started from the corresponding local minima, namely the *ggpg* and *ggmg* conformations. NOEs and T_1s values are then weighted by the appropriate population distribution. Simulations performed in vacuo with the *ggmg* conformation as the starting structure display interconversion to the global minimum very early in the simulation.

The calculated and experimental T_1s are compared in Figure 2.16. There is little variation in the relaxation times for the different structures, with the exception of H2, which shows a divergence of about 0.5 s.

Although comparison of the T_1 values is inconclusive, more information can be derived from the NOE and interglycosidic $^3J_{CH}$ values. Steady state NOE experiments

on **3** were performed and the results are shown in Table II.14. Inter-ring NOEs to the H2' and H3' are observed on saturation of H1 of the Glc*p*NAc, and to the H1 and H5 on saturation of H2'. The H2'-H5 NOE is very weak and is overlapped by the H2'-H4' enhancement. Examination of the minimum predicted by CHARMM shows that the H2'-H5 distance is ~ 3.3 Å, which places it at the outer limit detectable by NOE, and in the GESA minimum, this distance is > 5.0 Å, well beyond the range of observable enhancement. The NOEs calculated from the minima predicted by the various force fields are listed in Table II.14. The H1-H3 NOE is overestimated in all the structures, but overall, the best fit is obtained from the average of the dynamics simulations started at the different minima predicted by CHARMM.

Inverse detected ^{13}C-^{1}H correlated experiments optimized for detection of long range (3-bond) coupling were run on **3**. The $^{3}J_{CH}$ values obtained for the C3'-H1 coupling was 4.6 Hz. The C1-H3' coupling was unresolved. The $^{3}J_{CH}$ values calculated from the dynamics trajectories are shown in Table II.15. Both the GESA and CHARMM structures show calculated values for the C3'-H1 coupling which are within 10 % of the experimental value. However, the C1-H3' coupling calculated for the GESA structure is 4.6 Hz, while the value of 2.7 Hz obtained for the CHARMM structure may not be easily observed. The average $^{3}J_{CH}$ values calculated from the averaged dynamics simulations were 5.0 Hz and 3.7 Hz for H1-C1-O1-C3' (Φ) and C1-O1-C3'-H3' (Ψ), respectively, which is inconclusive.

The results indicate that the dynamics calculations are unable to simulate transitions between different minima. This is, as discussed earlier, entirely parameter dependent.

Table II.12. Relative energies of the minimum energy of conformations of β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**) obtained from a grid search.

Structure	Φ	Ψ	ω	θ	Relative Energy (kcal/mol)
gtpg	34.0	-55.5	54.7	4.5	0.00
gttr	29.9	-48.9	55.8	129.0	0.68
ggpg	56.3	15.0	-51.5	26.7	0.80
ggtr	32.7	-46.6	-51.5	128.5	2.16
ggmg	163.9	9.4	46.2	5.0	2.83
tgpg	54.7	19.5	171.0	23.7	5.04
tgtr	26.5	-49.3	173.0	129.4	5.50
tgmg	31.5	-54.8	157.8	3.1	6.52

Table II.13. Comparison of the minimum energy structure of β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**) calculated using CHARMM starting from minima derived from other molecular mechanics methods.

Method	Φ	Ψ	ω	θ
GESA	58	-4	59	174
GESA[a]	20	26	59	174
CHARMM[b]	34	-55	55	5
CHARMM[c]	36	-48	62	145
Crystal Structure[d]	-	-	59	-61

a. Calculations performed using the Burkert dipole term to include the exoanomeric effect.
b. CHARMM calculation performed using all energy terms.
c. CHARMM calculation performed excluding electrostatic terms.
d. β-N,N' Diacetyl chitobiose Trihydrate [153].

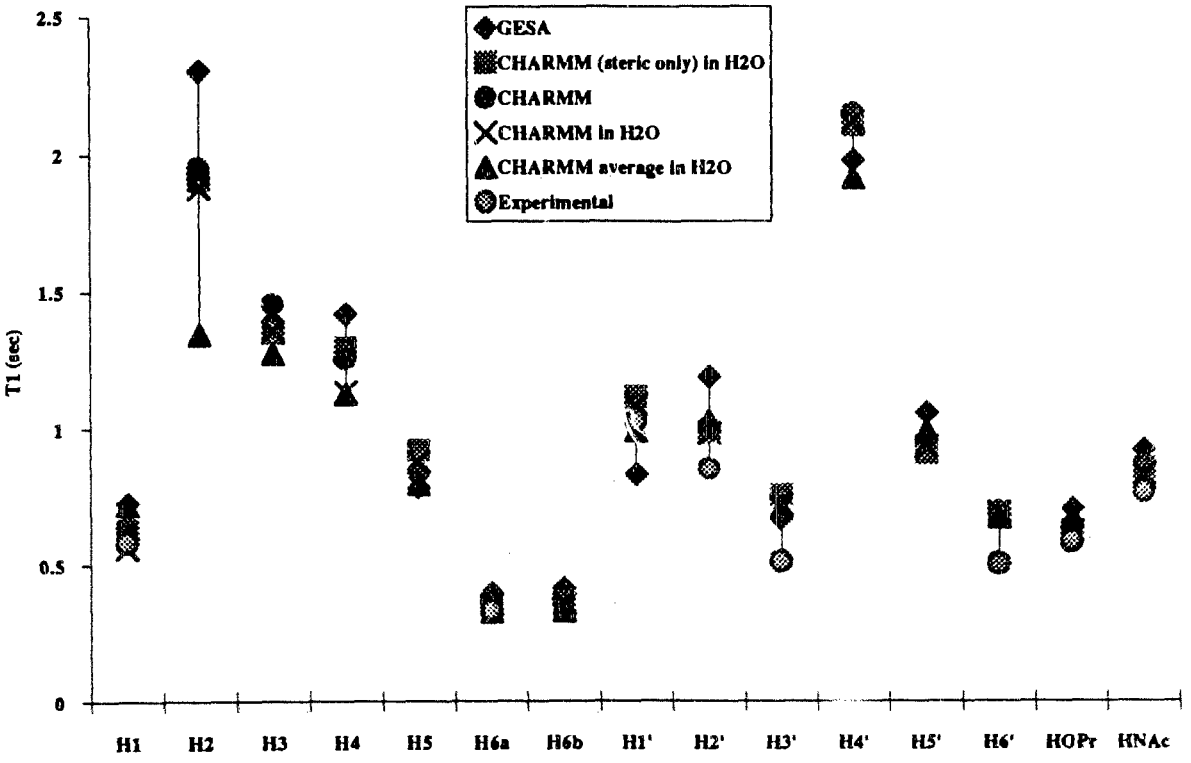


Figure 2.16. A comparison of experimental T_1s and T_1s calculated for different minimum energy structures from their dynamics trajectories for the compound β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (3).

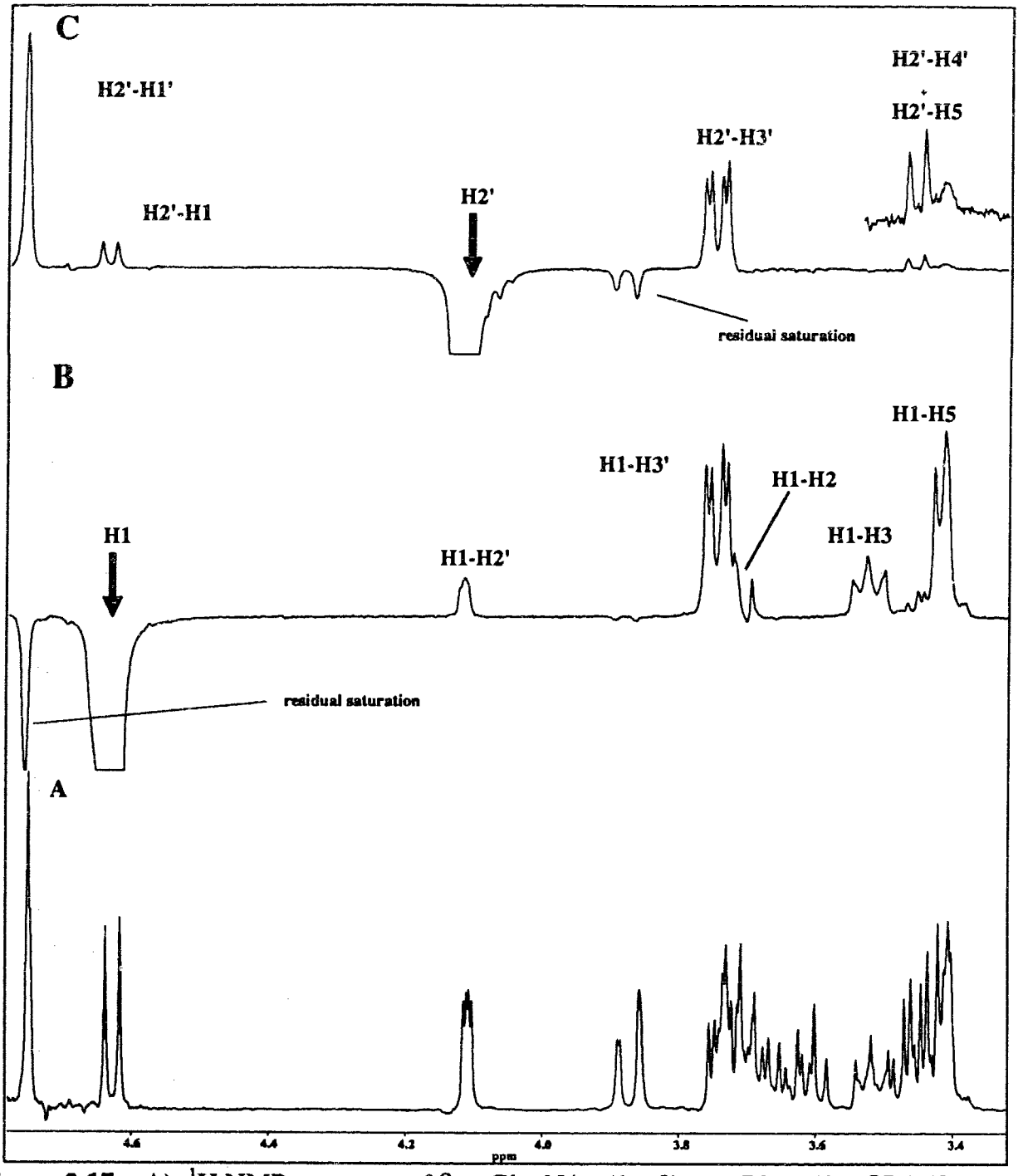


Figure 2.17. A) ^{1}H NMR spectrum of β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**). NOE difference spectrum of **3** after B) saturation of the anomeric proton of the Glc*p*NAc ring. C) saturation of H2 of the Rha*p* ring.

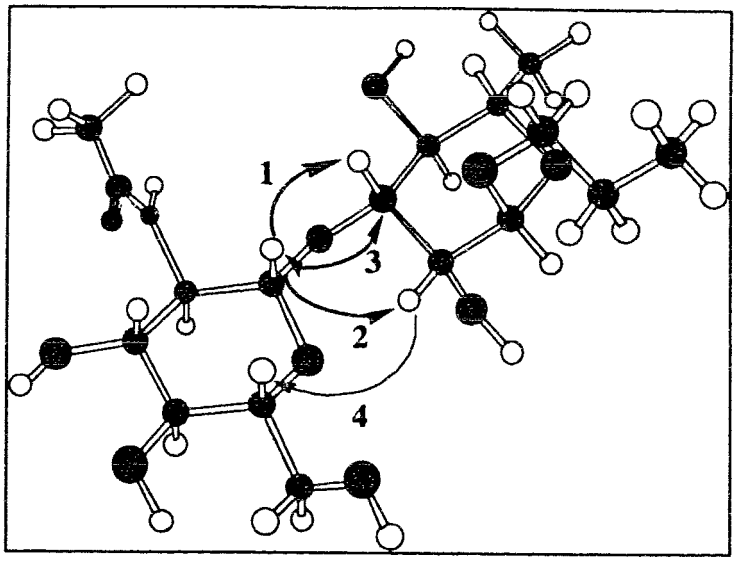


Figure 2.18. NOEs and $^{3}J_{COCH}$ observed in β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**).

Table II.14. A comparison of the relative NOEs (±20%) observed experimentally and calculated from the dynamics trajectories of the minimum energy structures of **3**.

Saturated Proton	Observed Proton	Experimental and Calculated NOEs				
		Experimental	GESA	CHARMM in H_2O Global min	CHARMM in H_2O Steric only	CHARMM in H_2O Average
H1	H3	0.40	0.93	1.45	1.22	1.50
	H5	1.00	1.00	1.00	1.00	1.00
	H2'	0.15	0.07	0.55	0.82	0.25
	H3' + H2	1.29	2.19	1.87	2.08	1.27
H2'	H1'	0.95	0.97	1.30	1.45	0.94
	H3'	1.00	1.00	1.00	1.00	1.00
	H4' + H5	0.12	0.14	0.24	0.18	0.15
	H1	0.14	0.05	0.46	0.65	0.23

Table II.15. A comparison of the observed $^3J_{COCH}$ values and those calculated from the dynamics trajectories of the minimum energy structures of β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→OPr) (**3**).

Method	$^3J_{COCH}$	
	C3' - H1	C1 - H3'
GESA[a]	5.0	4.6
CHARMM[b]	4.7	2.7
CHARMM[c]	5.0	2.6
CHARMM (average)[d]	4.8	3.7
Experimental	4.6 ± 0.2	-

a. Calculations performed using the Burkert dipole term to include the exoanomeric effect.
b. CHARMM calculation performed using all energy terms.
c. CHARMM calculation performed excluding electrostatic terms.
d. Calculated from a weighted average of the dynamics trajectories.

2.5. Conclusions

A comparison of the potential energy surface as calculated by the various force fields shows a correspondence, and a minimum identified by one force field is generally present on the surface calculated using the others. However, the magnitude of the relative energies, and hence the global minimum, is located at different points.

The **global** minimum energy structures predicted by CHARMM were, in two of the three cases studied, inconsistent with the observed NMR data. Analysis of the minimum energy conformations of the mannobioside and methyl-β-D-lactoside revealed that they were stabilized by intramolecular H-bonds. Solvation of the global minimum or removal of the H-bonds by exclusion of the electrostatic term to locate a new minimum, followed by dynamics simulations with the complete force field resulted in a different conformation which was more consistent with the NMR results.

Molecular dynamics simulations sample the energetically accessible conformational space in the vicinity of the starting conformation. The regions sampled by the dynamics calculations, although shifted in location when compared to the percentage maps, are comparable to the area included in statistical averaging. However, if multiple minima exist, the dynamics method with CHARMM is unsuccessful in simulating transitions between the minima; dynamics simulations must therefore be performed with starting conformations corresponding to each independent minimum and a weighted average of these simulations must be used to include the contribution from the different populations.

For larger oligosaccharides, dynamics simulations represent an alternative to statistical averaging as a sampling procedure. As the number of linkages increases, the grid search can be replaced by locating the starting conformations with inter proton distances derived from NMR experiments. Since the NOE is a function of (r^{-6}) it must

be realized that a visible NOE may be due to either the contribution of a vanishingly small population that brings the two protons between which the NOE is observed into extremely close proximity, or to a single conformation. It is difficult, if not impossible to distinguish between the two possibilities, expecially when the number of observed NOEs is small. Rapid internal motion about the glycosidic bond, which would support the former theory, has been demonstrated to be a factor in the relaxation behavior of oligosaccharides[151,229]. A large range of flexibility has also been observed in molecular dynamics simulations of Man*p* α-1→3)-Man*p*[151,229]. However, molecular dynamics simulations with other force fields do not show a similar range of variation[230].

The derivation of a single conformation that fits the observed NMR data remains the simplest solution, and must be considered in the initial analysis. Thus, observed NOEs can be used to determine conformations that can be used as starting structures in the molecular dynamics simulations.

CHARMM has a facility such that energy minimization and dynamics simulations will proceed with the inclusion of NMR data as pseudo energy terms. Thus, NMR data can be input directly into the minimization or dynamics calculation. Such a protocol, based on combined NMR spectroscopic analysis and molecular dynamics simulations, is developed and illustrated in Chapters III and IV.

Chapter III
APPLICATION OF 2D NMR SPECTROSCOPY AND MOLECULAR DYNAMICS CALCULATIONS TO THE CONFORMATIONAL ANALYSIS OF OLIGOSACCHARIDES CORRESPONDING TO THE CELL-WALL POLYSACCHARIDE OF *STREPTOCOCCUS* GROUP A

3.1. Introduction

3.1.1. *Streptococcus* Group A

The *Streptococcus* Group A represents a virulent subgroup of the *Streptococcus* family of bacteria. The genus *Streptococcus pyogenes* causes skin and throat infections, pneumonia and streptococcal pharyngitis, commonly known as strep throat. If untreated in children, strep throat can lead to rheumatic fever, which causes heart valve damage[154,155].

The close relationship between *Streptococcus* and rheumatic fever is thought to be brought on by an autoimmune response, a result of antibodies produced during an immune response to strep infection cross reacting with heart valve tissue. Antibodies raised to both the surface M protein and the group-specific cell -wall carbohydrate have been shown to react with human cardiac tissue [156-159].

In the search for a vaccine, the M protein would be a prime candidate for immunization against streptococcal infection. However, due to the hypervariable nature of the M protein[160,161], an antibody raised against one strain may not be successful in preventing infection by another. Further, the crossreaction exhibited by antibodies that have been raised against M protein to heart tissue indicates that inoculation with the M protein may in fact cause rheumatic fever.

3.1.2. The Group A Cell-Wall Polysaccharide

A major focus of our research effort has been to attempt to identify fragments of the Group A cell wall polysaccharide that will elicit an immune response that targets the

organism specifically. A part of this ongoing project has been the synthesis of oligosaccharides that correspond to the *Streptococcus* Group A cell-wall polysaccharide[144,162-169]. This polysaccharide is comprised of a backbone of poly-L-rhamnopyranosyl units connected by alternating α-L-(1→3) and α-L-(1→2) linkages, to which are attached N-acetyl-β-D-glucosamine (β-D-Glc*p*NAc) residues at the 3 positions of the rhamnose backbone[170,171], as shown in Figure 3.1.

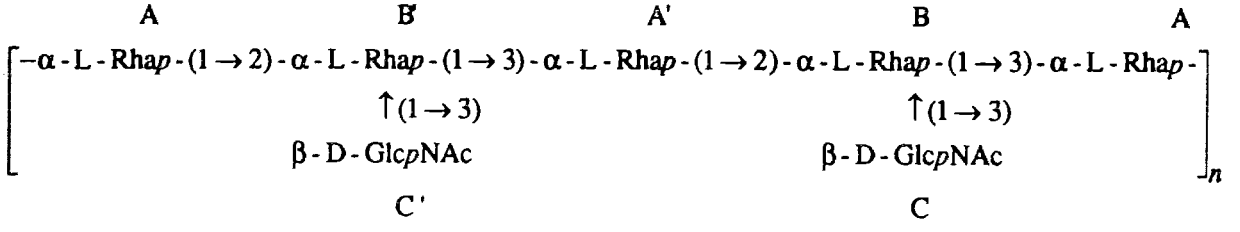


Figure 3.1 *Streptococcus* Group A cell-wall polysaccharide

These oligosaccharides were used to develop antigens to study the nature of antibody antigen interactions in inhibition binding assays with monoclonal antibodies raised against the bacterial polysaccharide and polyclonal antibodies raised against the antigens[172]. A natural extension of this work is to define the antibody hapten interactions at the molecular level. A first step in this analysis is the modeling of the topology of the hapten. A panel of suitable oligosaccharides ranging from a branched trisaccharide (**4**) corresponding to the fragment [A'-(C)B], a tetrasaccharide (**5**) [A'-(C)B-A], a pentasaccharide (**6**) [C'-B'-A'-(C)B] and two hexasaccharides (**7**) [C'-B'-A'-(C)B-A] and (**8**) [A-(C')B'-A'-(C)B], were chosen as candidates for the study. Compounds **4-9** are shown in Figure 3.2.

Prior to any analysis, the ^{1}H and ^{13}C NMR spectrum of each compound must be completely assigned. Oligosaccharides present complicated spectra with extreme overlap, and thus, assignment is carried out with two dimensional NMR techniques.

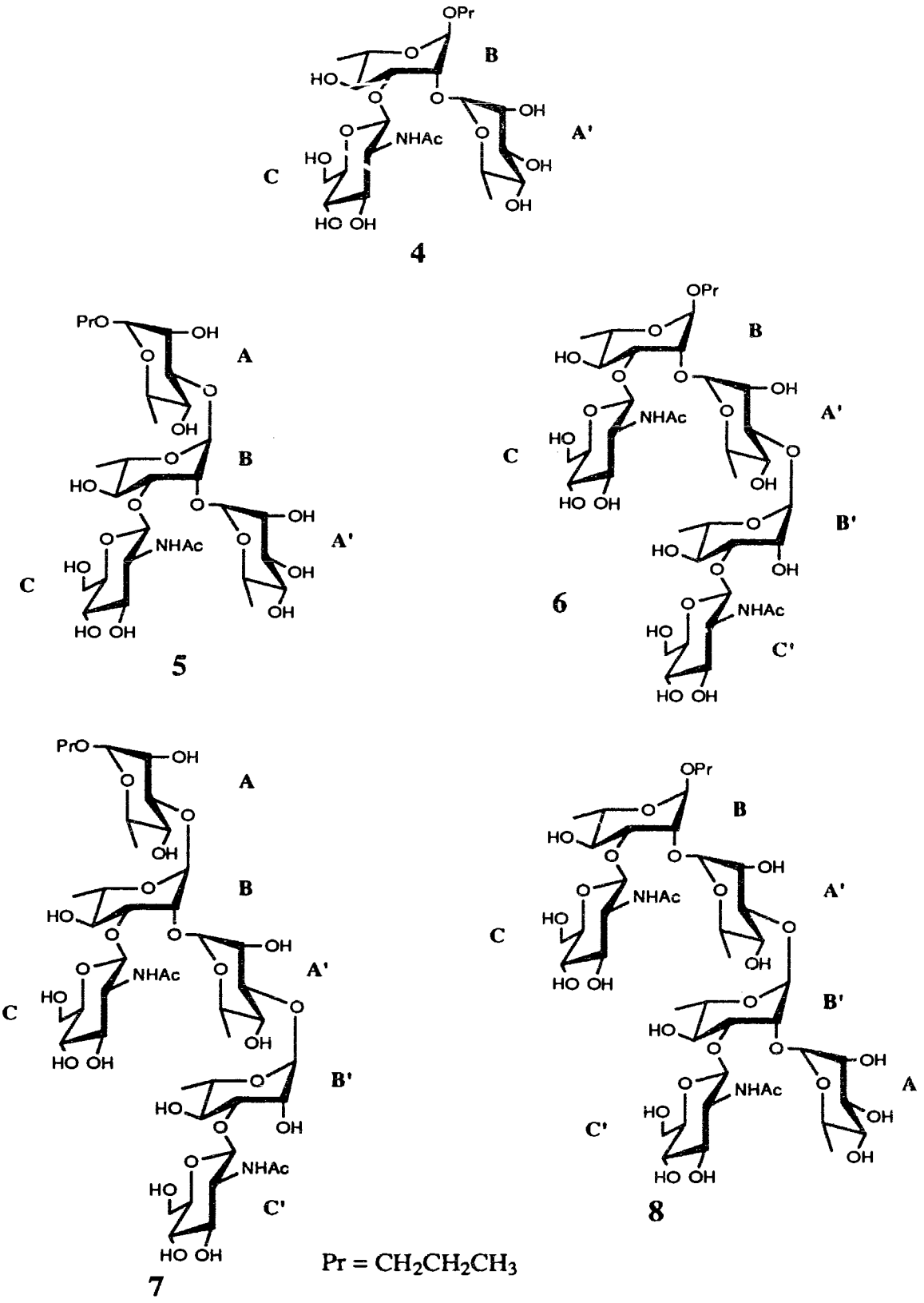


Figure 3.2. Oligosaccharides corresponding to fragments of the cell-wall polysaccharide of *Streptococcus* Group A.

3.2. Overview of 2D NMR Spectroscopy

The primary use of NMR spectroscopy is still the characterization of organic compounds, but even these methods have evolved with the widespread availability of 2D NMR techniques. The 1D ^1H spectrum of the hexasaccharide (**7**) is shown in Figure 3.3. The resonances above 2.0 ppm can be assigned to the methyl protons of the N-acetyl group, the methyl protons of the four rhamnose residues as well as the β and γ protons of the propyl aglycon. There is a large degree of overlap in the region of 3.0 - 4.0 ppm, corresponding to the resonances of 24 protons. Assignment of these peaks would require a series of selectively decoupled spectra. Even then, most resonances would be difficult, if not impossible, to assign unambiguously.

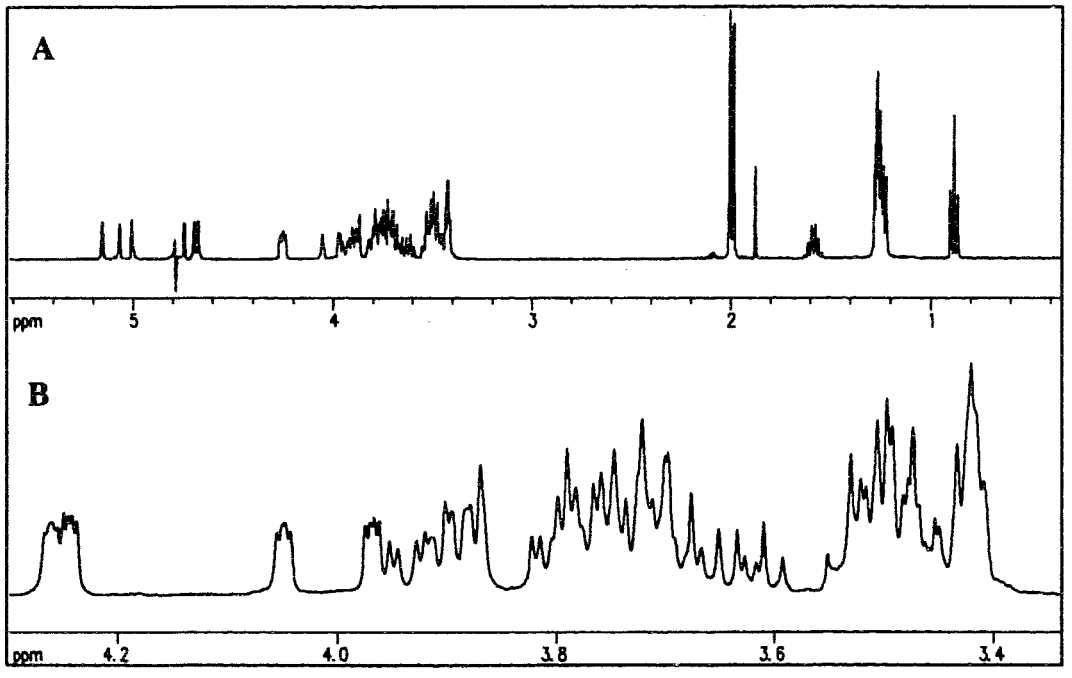


Figure 3.3. A. 1D ^1H NMR spectrum of the hexasaccharide (**7**) in D_2O, with presaturation of residual H_2O (4.77ppm). **B.** Expansion of the region between 3.3-4 ppm which includes the resonances of 24 protons.

Analysis and complete assignment of the ^{1}H and ^{13}C NMR chemical shifts is greatly facilitated by the use of 2D NMR spectroscopy. Typically, all that is needed is a TOCSY (TOtal Correlation SpectroscopY)[173] spectrum to assign the chemical shifts to a particular ring, a NOESY (Nuclear Overhauser Effect SpectroscopY)[65,66] spectrum to determine the sequence of the oligosaccharide, and a ^{1}H-^{13}C correlated spectrum to assign the ^{13}C chemical shifts. Since most of these compounds are used in 5mg/ml quantities, the ^{1}H-^{13}C correlated spectrum is run in the inverse mode[54], which utilizes the higher sensitivity of the ^{1}H nucleus.

The procedure is illustrated for the case of the hexasaccharide (7). A TOCSY spectrum of compound (7) is shown in Figure 3.4. Assignment of the resonances is straightforward as they are simply read off the chemical shift in either the row or column. Since the anomeric protons of all the residues are downfield of the ring protons, they can be used as starting points for the assignments. A projection of an F1 slice at 5.0 ppm illustrates how the chemical shifts can be assigned to entire rings. Note the high degree of overlap in the chemical shifts of the Glc*p*NAc rings. This degeneracy can be lifted by moving to three dimensions. The use of sufficient digital resolution permits the measurement of coupling constants from the spectrum. In order to determine the primary sequence, a further piece of information is required, and this can be obtained from a NOESY experiment. A NOESY spectrum is shown in Figure 3.5. Since only protons that are in close proximity to each other will contribute to off-diagonal peaks, the proton spectrum can now be assigned. For example, the H1 resonance of ring A shows cross peaks (not visible in the diagram) to the propyl aglycon, and is assigned the peak at 4.77 ppm. The B ring which is 1→3 linked to A will show a cross peak from H1 to the H3 of the A ring. This enables the assignment of the peak at 5.06 ppm to the H1 of the B ring. In a similar fashion, the H1 assignments of the A', B', C and C' rings can be made. Table III.1 lists the ^{1}H chemical shifts for compound 7 and the other

oligosaccharides. Assignments of the ^{13}C signals can be made from the ^{1}H-^{13}C correlated spectrum. An inverse correlated ^{1}H-^{13}C spectrum of compound **7** is shown in Figure 3.6. The F1 dimension represents the ^{13}C chemical shift, and since the ^{1}H is already known, ^{13}C chemical shifts can be assigned by reading the ^{13}C chemical shift off the vertical (F1) axis. Table III.2 lists the ^{13}C NMR chemical shifts for compounds **4-8**.

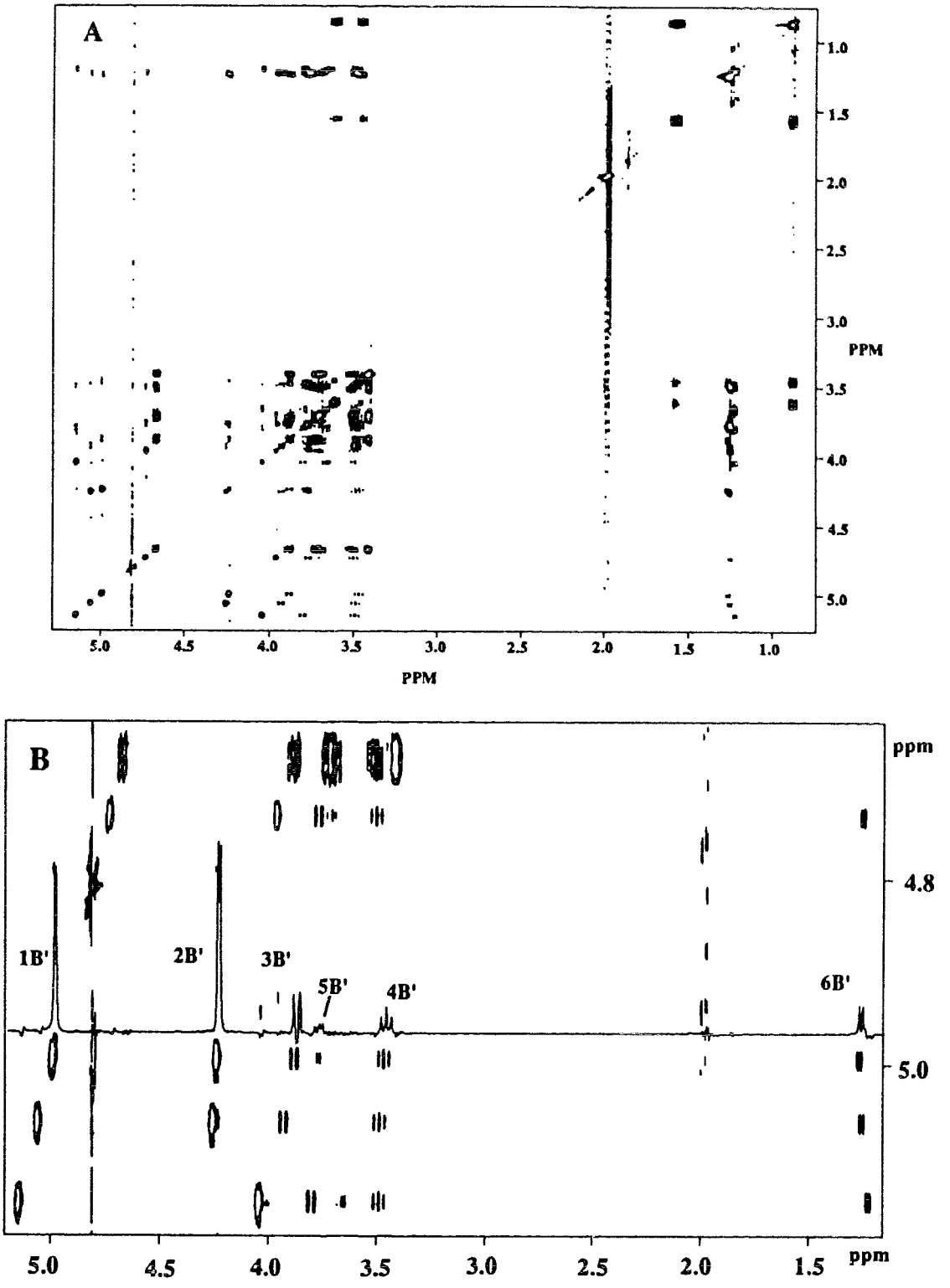


Figure 3.4. TOCSY spectrum of the hexasaccharide (7). A) Full spectrum. B) Expansion of the anomeric region (4.6-5.2ppm).

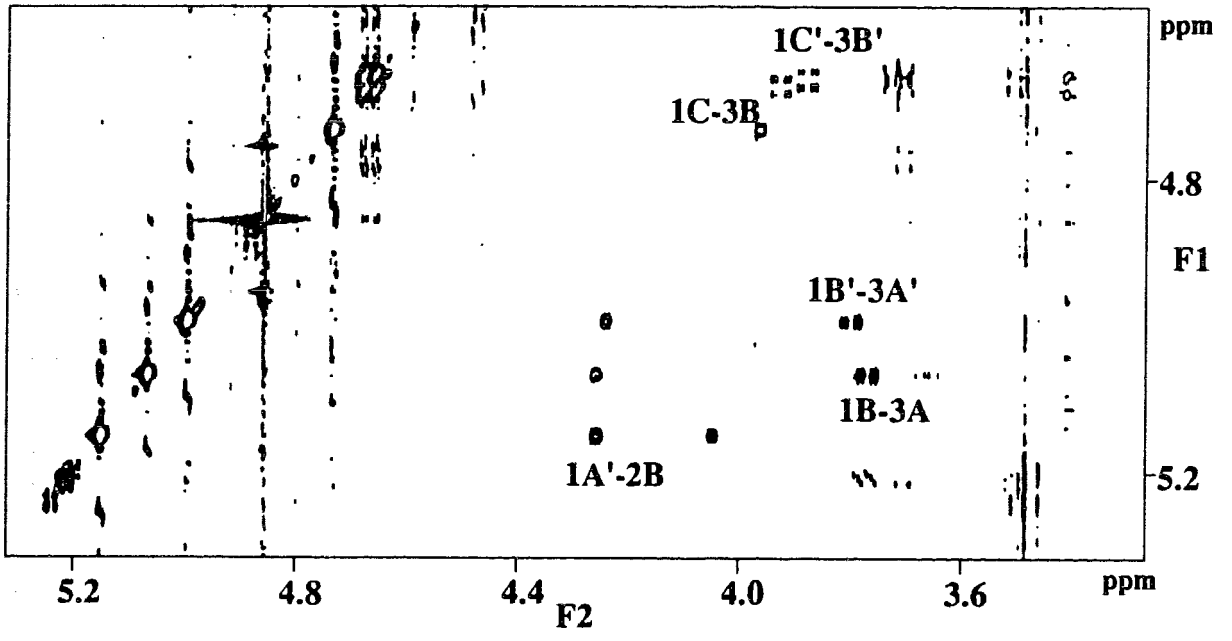


Figure 3.5. Expansion of the anomeric region from the NOESY spectrum of the hexasaccharide (7).

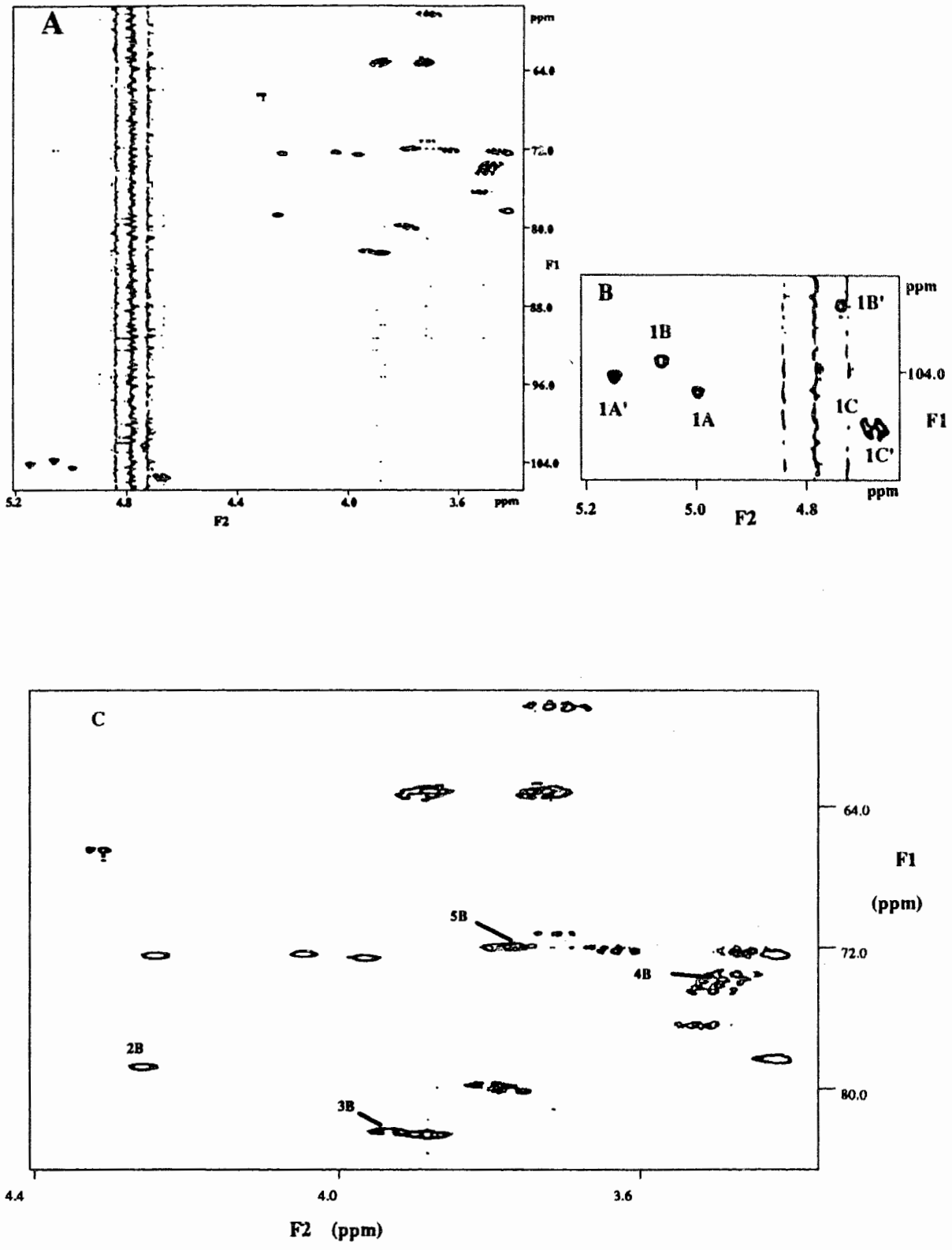


Figure 3.6. Inverse correlated ^{1}H-^{13}C spectrum of the hexasaccharide (7).
A) Expansion of the region 3.5 -5.2 ppm displaying the ring protons. B) Expansion of the anomeric region. C) Expansion of the region 4.4-3.0 ppm.
For clarity, only the cross peaks corresponding to the B ring have been labeled.

Table III.1. ^{1}H NMR data[a] (ppm) for the compounds **4-8**

	(4)[163]	(5)[164,165]	(6)[167]	(7)[168]	(8)[162]
Ring Proton					
1A		4.74		4.74	5.14
2A		3.96		3.97	4.01
3A		3.77		3.77	3.74
4A		3.50		3.51	3.39
5A		3.71		3.72	3.63
6A		1.25		1.25	1.21
1B	4.78	5.09	4.77	5.07	4.77
2B	4.10	4.25	4.13	4.26	4.13
3B	3.80	3.94	3.80	3.94	3.81
4B	3.46	3.49	3.46	3.49	3.45
5B	3.67	3.77	3.68	3.78	3.68
6B	1.22	1.25	1.25	1.26	1.25
1C	4.64	4.70	4.63	4.68	4.63
2C	3.69	3.70	3.69	3.70	3.69
3C	3.50	3.51	3.50	3.51	3.50
4C	3.41	3.42	3.40	3.41	3.40
5C	3.40	3.42	3.40	3.42	3.40
6CS	3.87	3.89	3.88	3.90	3.87
6CR	3.71	3.74	3.72	3.73	3.72
1A'	5.11	5.14	5.14	5.15	5.15
2A'	3.99	3.99	4.04	4.05	4.03
3A'	3.73	3.74	3.81	3.81	3.81
4A'	3.39	3.39	3.50	3.50	3.52
5A'	3.67	3.63	3.71	3.67	3.71
6A'	1.22	1.22	1.24	1.23	1.24
1B'			5.00	5.00	5.08
2B'			4.24	4.24	4.25
3B'			3.88	3.89	3.95
4B'			3.47	3.48	3.50
5B'			3.77	3.79	3.77
6B'			1.26	1.26	1.27
1C'			4.68	4.68	4.70
2C'			3.72	3.72	3.71
3C'			3.52	3.53	3.51
4C'			3.40	3.42	3.42
5C'			3.40	3.42	3.42
6C'S			3.88	3.88	3.87
6C'R			3.72	3.72	3.72

[a] In D_2O at room temperature.

Table III.2. ^{13}C NMR data[a] (ppm) for the compounds **4-8**.

	(4)[163]	(5)[164,165]	(6)[167]	(7)[168]	(8)[162]
Ring Carbon					
1A		102.4		102.3	104.3
2A		72.5		72.8	72.6
3A		80.3		80.4	72.7
4A		74.9		74.6	74.7
5A		71.5		71.5	71.8
6A		19.3		19.3	19.3
1B	101.2	104.4	101.2	103.8	101.1
2B	79.6	79.0	79.3	78.9	79.3
3B	82.7	82.4	82.8	82.5	82.7
4B	74.8	74.6	73.6	74.1	74.0
5B	71.7	72.1	72.0	72.2	71.6
6B	19.4	19.3	19.4	19.5	19.3
1C	105.3	105.2	105.3	105.2	105.3
2C	58.7	58.7	58.6	58.7	58.5
3C	76.6	76.7	76.6	76.7	76.5
4C	72.9	72.9	72.6	72.6	72.5
5C	78.6	78.6	78.5	78.6	78.4
6C	63.7	63.8	63.6	63.7	63.5
1A'	104.4	103.8	104.1	104.1	104.0
2A'	72.8	72.9	72.6	72.6	72.5
3A'	72.8	72.8	79.7	79.9	79.3
4A'	73.9	74.0	74.3	74.3	74.4
5A'	71.9	71.9	72.1	72.4	72.0
6A'	19.4	19.4	19.5	19.4	19.2
1B'			104.4	104.5	103.6
2B'			72.6	72.8	78.9
3B'			82.7	82.8	82.4
4B'			74.0	73.7	73.8
5B'			72.1	72.1	72.0
6B'			19.3	19.6	19.5
1C'			105.4	105.4	105.2
2C'			58.4	58.6	58.5
3C'			76.5	76.5	76.6
4C'			72.6	72.6	72.5
5C'			78.4	78.5	78.4
6C'			63.4	63.5	63.5

[a] In D_2O at room temperature.

3.2.1 Corrections to ROESY Cross Peaks

In the preceding section, a NOESY experiment was used to aid in the determination of the primary sequence of the oligosaccharide. NOE spectroscopy has further reaching aspects, as discussed in Section 1.2, and can be used to determine the three dimensional structure of a compound through calculation of internuclear distances. Following the assignment of the ^{1}H and ^{13}C resonances, the three dimensional structure of the oligosaccharides can be determined by analysis of the cross peaks in the NOESY or ROESY spectrum. Due to the fact that for most of the compounds studied, $\omega\tau_c \approx 1$, the NOE will be small. In this case, NOESY experiments cannot be used, and ROESY experiments must be performed instead. Integration of the cross peaks observed in ROESY spectra can be converted into interproton distances, but prior to computation of the interproton distances, complications arising from the ROESY technique must be corrected.

Ideally, the cross peaks in a ROESY spectrum are influenced by only the cross relaxation between the spins. However, two additional factors have been judged to be important in attenuating the integrated intensity of the cross peaks, namely offset effects and Hartmann Hahn effects[174].

(i) Offset Effects

The effective spin lock field B_{eff} will cause the magnetization to precess around it. However since the spin lock will not be on resonance for the entire sweep width, this will cause it to make an angle α with the transverse plane (see Figure 1.7). The relaxation behavior of the magnetization can then be considered to be a sum of the transverse and longitudinal relaxation. For a small chemical shift range with a fairly strong spin lock, such that the angle the effective field makes with the static field is greater than 80° , the longitudinal contribution can be neglected. The observed cross peak intensity will then be

$$a_{ij_{(obs)}} = \sigma_{ij}(\sin\alpha_i \sin\alpha_j)$$

(ii) Hartmann Hahn effects

Transfer of magnetization through scalar coupling will produce off-diagonal peaks that will be the same phase as the diagonal or opposite phase to the ROE cross peaks, hence decreasing the magnitude of the observed ROE.

Corrections for offset and Hartmann Hahn effects

Both these effects will contribute to the cross peaks. The observed peak intensities will then be[174]

$$a_{ij_{(obs)}} = \sigma_{ij}(\sin\alpha_i \sin\alpha_j) + \sum \text{HOHAHA contributions} \qquad (3.1)$$

or

$$\sigma_{ij} = \frac{a_{ij_{(obs)}} - \sum \text{HOHAHA contributions}}{(\sin\alpha_i \sin\alpha_j)} \qquad (3.2)$$

To calculate distances, the integrated intensities must first be corrected for Hartmann Hahn effects. There are two cases :

a) Protons *i* and *j* are dipolar coupled, and proton *j* is scalar (J) coupled to proton *k*. In this case, the NOE buildup of the *ij* cross peak will be modulated by HOHAHA transfer from proton *j* to *k*. The observed cross peak will then be[174]

$$a_{ij}(obs) = \sigma_{ij}(\sin\alpha_i \sin\alpha_j)\,(1 - 0.5\sum s^2) \qquad (3.3)$$

where the Hartmann Hahn factor

$$s = \sin 2\phi \quad \text{and} \quad \phi = 0.5 \arctan\left(\frac{(1+\cos\alpha_{jk})\,J_{jk}}{2(\nu_j - \nu_k)}\right) \qquad (3.4)$$

b) Protons *i* and *j* are both dipolar and scalar coupled. Since the HOHAHA peaks will be of opposite phase to the NOE peaks, here the NOE buildup of the *ij* cross

peak will be the sum of the NOE contribution and the HOHAHA contribution, as defined by the equation below[174]:

$$a_{ij}(obs) = \sigma_{ij}(\sin\alpha_i \sin\alpha_j)(1 - 0.5\sum s^2) + 0.5\sin\alpha_i \sin\alpha_j \bullet s_{ij}^2 \qquad (3.5),$$

where s is as defined in equation 3.4. The second case is important as it is representative of the H1-H2 protons in rhamnose residues. This distance is generally taken to be the ruler distance, and hence it is necessary to apply this correction to the H1-H2 cross peak so that the distances calculated are not overestimated.

In this manner, cross peaks observed in the ROESY spectrum of a compound can be translated into reliable estimates of distances between the proton pairs. These distances can be used to narrow the range of possible conformers, replacing the grid search that is utilized for smaller compounds, but is not feasible for larger compounds due to the number of linkages.

3.3. Research Objectives

The conformational properties of the compounds **4-8** will be studied using the ROESY technique to obtain interproton distances. Molecular dynamics simulations of the conformations thus derived will be carried out, and in order to test the validity of the conformations thus derived, the ROESY buildup curves will be calculated from molecular dynamics simulations and compared to the experimental ROE buildup. This comparison should provide an indication of the "correctness" of the model.

The study will provide information about the conformational surface of these compounds, which encompass the entire repeating unit, and through extrapolation, aid in the identification of salient features of the polysaccharide that determine its binding characteristics and specificity.

3.3. Experimental

3.3.1. Conformational Analysis

The general scheme for the conformational analysis is outlined in a flowchart in Figure 3.7.

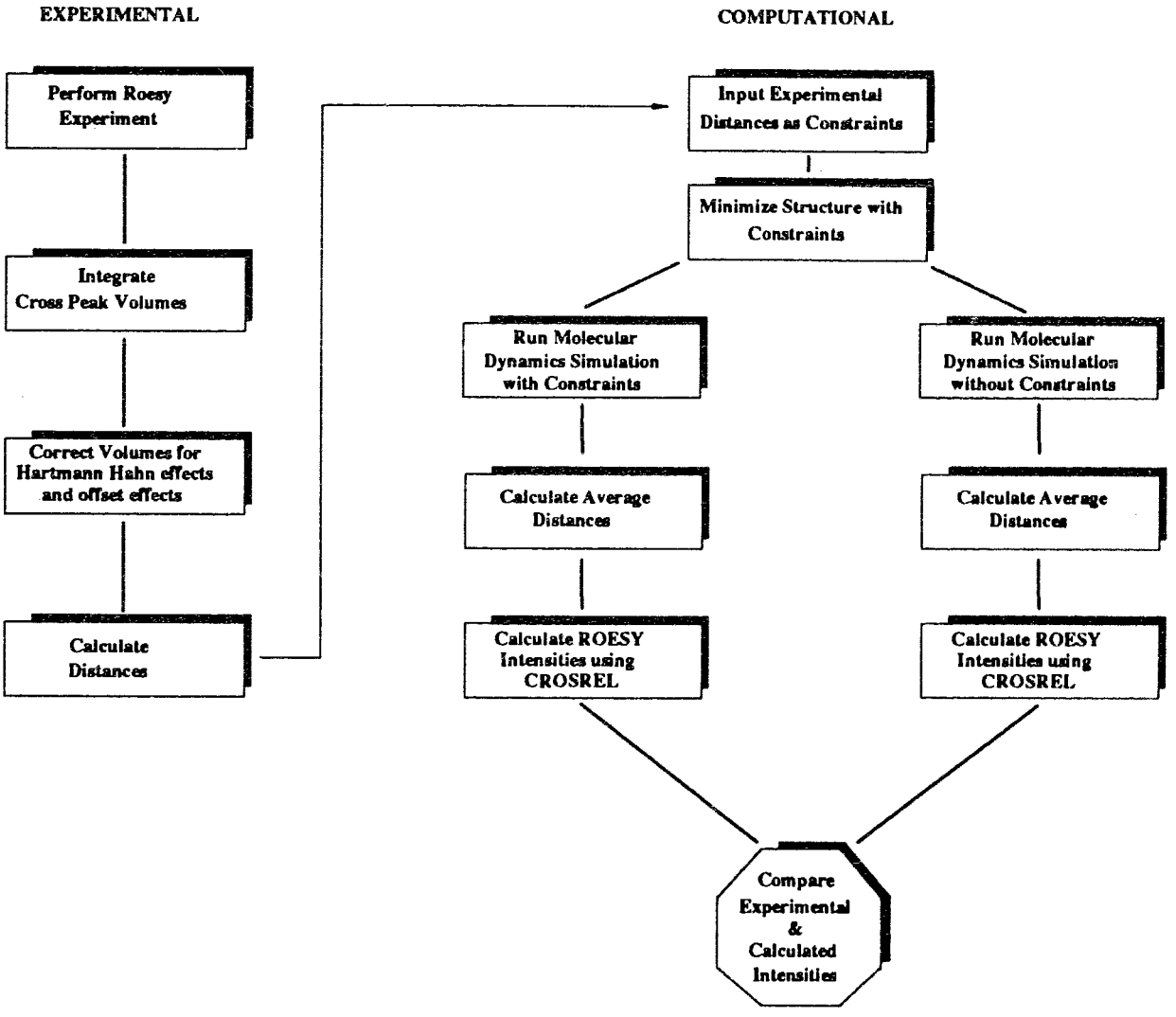


Figure 3.7. General scheme for the Conformational Analysis.

3.3.1.1. Computations

Calculations were performed on Silicon Graphics 4D25 and 4D35 workstations with the program QUANTA, interfaced to the force field CHARMM[104], commercially available software marketed by Molecular Simulations Inc.[175].

Interproton distances derived from the ROESY spectra of the compounds **4-8** were input as distance constraints into QUANTA, and the structures were minimized subject to these constraints. Minimizations were performed with 2000-5000 steps of the Powell algorithm[110]. The minimized structure was then centered in a water box of 15-30 Å and minimized further, with the constraints, until the rms force was < 0.01 (kcal/(mol Å). The resulting structure was used as a starting structure for the dynamics simulations. To save computer time, water molecules more than 7 Å away from the molecule were excluded from the calculation. Dynamics calculations, performed with the Verlet algorithm[123], were then calculated both with and without the constraints, and typically included 10,000 steps of heating over a period of 10 ps, followed by an equilibration period of the same time. The simulations were run for a period of 10 ps and 30 ps for the constrained and unconstrained dynamics simulations, respectively.

The coordinates from the dynamics calculations were output every 10 fs. These coordinates were then used to calculate the average r^{-6} (see experimental section in Chapter II) which was used as input for CROSREL[176,177] a program that uses a full relaxation matrix to calculate both NOESY and ROESY buildup curves. Multi-spin effects are accounted for[177]. CROSREL requires input of the chemical shifts and scalar couplings to allow for the calculation of the Hartmann Hahn effects. If tight coupling was present between a proton pair (as in H4 and H5 of the Glc*p*NAc) the coupling constant was estimated from a series of protected compounds. The other experimental variable required as input is the correlation time, τ_c. Since determination of τ_c requires

measurement of the ^{13}C T_1s, which was not feasible due to the limited amounts of compounds available, the correlation time was not determined experimentally. Instead, it was estimated using an option afforded by CROSREL. ROEs were calculated for a grid which systematically increments the correlation time, τ_c, and the leakage rate, R_L. The calculated ROEs are fitted to experimental ROEs for certain intraresidue crosspeaks, typically the H1-H2 cross peaks for the rhamnose rings, and the H1-H5 cross peaks for the Glc*p*NAc rings, which, due to the rigidity of the pyranose rings, are considered to be relatively invariant distances. This optimized correlation time is then used for the calculation of the ROESY buildup curves.

3.3.2. NMR Spectroscopy

Sample Preparation: The samples (5 mgs) were dissolved in 99.98% D_2O and lyophilized five times to exchange the hydroxyl protons with deuterium. They were then redissolved in 99.98% D_2O and subjected to five freeze thaw cycles to remove dissolved oxygen. The tubes were sealed under vaccum. NMR experiments were run on a Bruker AMX400 spectrometer. Samples were run spinning at temperatures ranging from 288-294 K. The temperature was chosen so as to shift the residual H_2O in the sample to minimize overlap with the peaks of interest. Typically, each 2D experiment consisted of 512 experiments, each 2K in size, using time proportional phase increments to provide quadrature detection in F1. The sweep width in the ROESY spectra was 3-3.5 ppm (1200-1300Hz) giving a maximum digital resolution of 0.6 Hz/pt in F2 and 2.5 Hz/pt in F1. In order to minimize the effects of folded peaks, the filter was set equal to the sweep width. The transmitter power was attenuated so that a high power π pulse was usually between 5-6 μs. For ROESY and TOCSY experiments the spin lock power was attenuated to give π pulses of 150 μs and 30 μs, respectively. Filtered ROESY experiments were performed using the pulse sequence described by Shaka et. al.[178]. A

($\pi/2$) pulse is followed by a composite spin lock pulse of $\pi_{-x}\pi_x$ pulses in order to eliminate unwanted TOCSY effects. The FIDs were zero filled to a 4K(F2) X 1K (F1) data set, providing a resolution of 0.3 Hz/pt in F2 and 1.2 Hz/pt in F1, and processed using an exponential line broadening of 0.3 in the F2 dimension and a sine squared apodization function with a shift of 2 in F1. Integration was performed after automatic baseline correction in both dimensions. Measurement of the three bond coupling constants ($^3J_{CH}$) of the hexasaccharide (**8**) was done by inverse 2D ^{13}C-^1H correlated spectroscopy on a 600MHz Bruker AMX spectrometer at the Carlsberg Laboratory in Copenhagen, Denmark, using a mixing time (125 ms) optimized for detection of the 3-bond coupling constants with a 4K X 512 data set (digital resolution = 1.2 Hz/pt). In cases where the T_1 noise of the residual H_2O interfered with the signal, the coupling constants were obtained from a 1D coupled ^{13}C spectrum, acquired on a Bruker AM500 spectrometer (operating at a ^1H frequency of 500MHz), also at the Carlsberg Laboratory. The 1D data sets were acquired with a digital resolution of 0.2 Hz/pt.

3.4. Results

3.4.1. Trisaccharide (4)

ROESY Spectrum

The anomeric region of the ROESY spectrum of the trisaccharide (**4**) is shown in Figure 3.8. Due to the high resolution and minimal overlap of resonances in the spectrum, the cross peaks can be identified. Intra ring contacts from the anomeric protons of the C ring to the H2 and H3 of the B ring, and from the 1H of the A ring to the H2 of the B ring as well as to the H2 of the C ring are visible. Although the chemical shift of the H2C overlaps that of the 5A' proton, analysis of the pattern and similar cross peaks in subsequent compounds **5-8** as well as analysis of NOESY spectra of these compounds are consistent with this assignment. The H1 of ring B shows an ROE to both the H5 proton of ring A' and to one of the diastereotopic protons of the propyl group (labeled HPr(a)). Since the chemical shift of both of these protons show a significant overlap (3.67 ppm and 3.69 ppm respectively) both of these cross peaks overlap. Similarly, there is an intra-ring ROE/TOCSY cross peak between the H1 and H3 of Ring A' that overlaps the ROE between the H1 of ring A' and the H2C of Ring C.

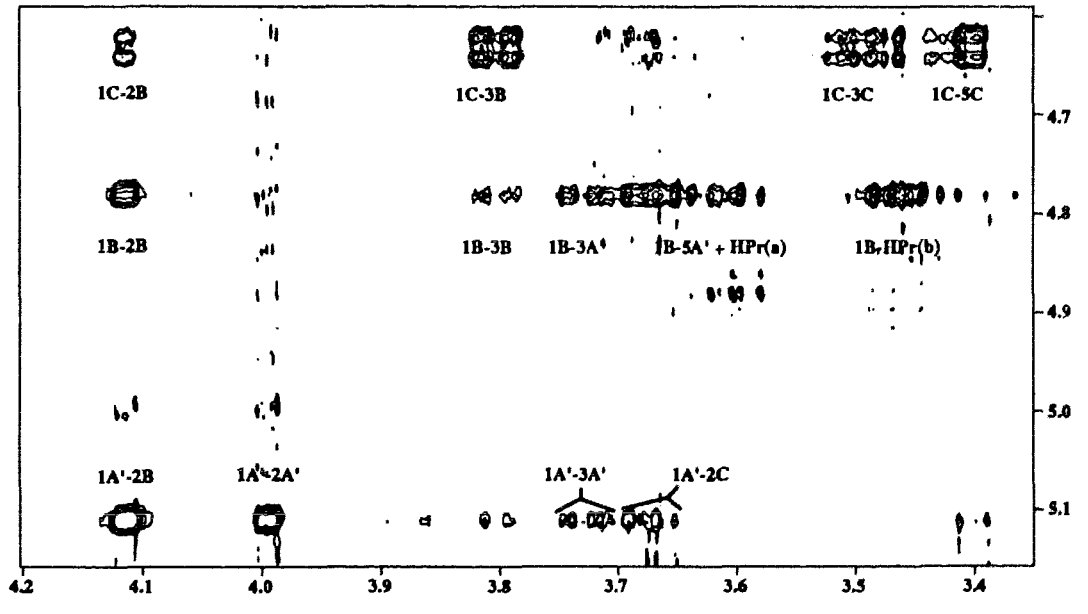


Figure 3.8. Expansion of the ROESY spectrum of the trisaccharide (**4**).

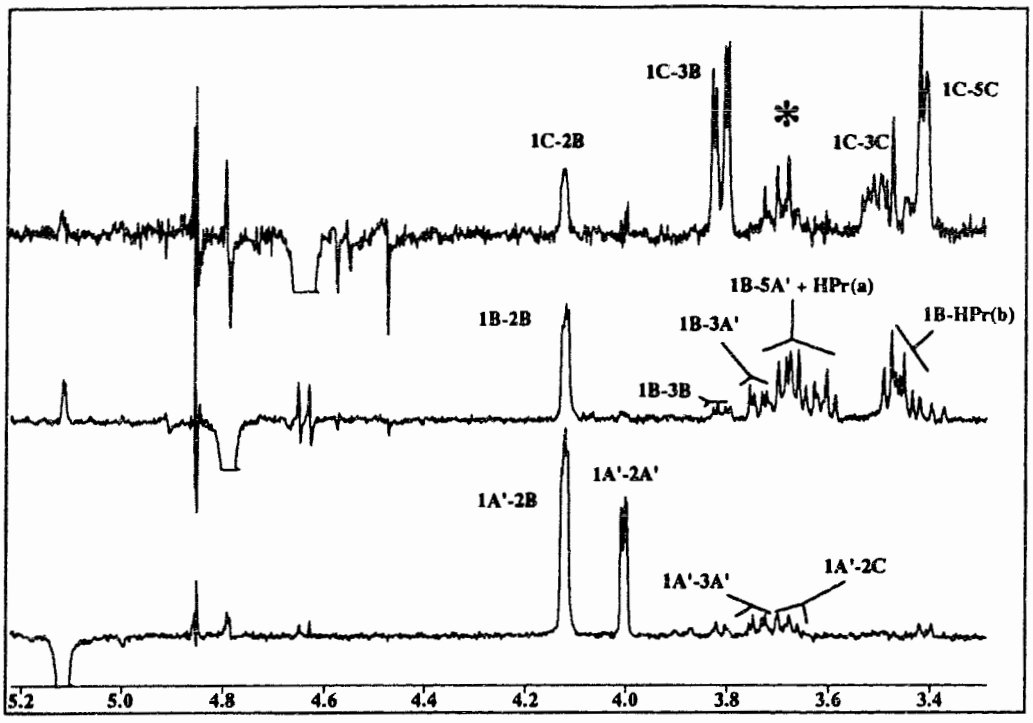


Figure 3.9. F1 slices from the ROESY spectrum of the trisaccharide (**4**).

1D Slices through the chemical shift of the anomeric protons have been extracted and plotted in Figure 3.9. The asterisk marks a cross peak that is an artefact, since it is only found in the F1 dimension. The contacts observed in the ROESY spectrum of **4** are displayed in Figure 3.10.

Offset and Hartmann Hahn corrections

Table III.3 lists the offset and Hartmann Hahn contributions to the important cross peaks of the ROESY spectrum of the branched trisaccharide (**4**). Due to the narrow spectral width of the ring protons, the offset correction is so small that it is trivial - it is less than the experimental error associated with the integration of the cross peaks. However, it has been included for the sake of completeness. The Hartmann Hahn factors (s^2) are listed in column 4 of Table III.3. In the trisaccharide (**4**), the chemical shift of a few important protons overlap. The protons H4C and H5C of the Gl*p*NA*c* ring are completely overlapped. They are also strongly coupled, and the Hartmann Hahn factor (s^2) is approximately equal to 1. As a result, the HOHAHA effects will not alter the

integrated intensity of the H5C peak, i.e. any transfer to the H4 from the H5 will be integrated along with the H5 peak. Similarly, there is an overlap between the H5 of the A' ring and one of the diastereotopic protons of the propyl aglycon, designated HPr(a), the most upfield of the two α protons. The H1 of the B ring shows an ROE to both of these protons. In this case the corrections are applied as follows. By first determining the distance to HPr(b), the distance to the HPr(a) proton can be estimated, and using this distance, the contribution of each proton to the cross peak can be calculated. In this particular case, the HPr(b) distance was calculated from the ROESY data to be 2.73 Å and the HOHAHA corrections are applied to each one.

Figure 3.11 illustrates the effect of the corrections on the ROE buildup of the trisaccharide (**4**). The experimental buildup curves are plotted in column 1, and column

Table III.3. Offset and Hartmann Hahn effects calculated for the trisaccharide (**4**).

Ring	Cross peak	$sin\alpha_I\ sin\alpha_S$	HH transfer	s^2
Ring C	1C-5C	0.987	5C-4C	1.00
			5C-6Cs	0.00
			5C-6CR	0.01
	1C-3C	0.989	3C-4C	0.73
			3C-2C	0.43
	1C-2B	0.993	2B-1B	0.04
			2B-3B	0.07
	1C-3B	0.988	3B-2B	0.07
			3B-4B	0.26
Ring B	1B-2B	0.997	2B-1B	0.04
			2B-3B	0.07
	1B-5A'	0.992	5A'-6A'	0.00
			5A'-4A'	0.40
	1B-HPr(b)	0.987	HPr(b)-HPr(a)	0.80
			HPr(b)-HPr2	0.00
Ring A'	1A'-2A'	0.995	2A'-1A'	0.04
			2A'-3A'	0.09
	1A'-2B	0.996	2B-1B	0.04
			2B-3B	0.07
	1A'-2C	0.991	2C-1C	0.10
			2C-3C	0.43

2 graphs the data corrected for offset effects and Hartmann Hahn transfer. Comparison of the corresponding sets of graphs indicates that the Hartmann Hahn contributions are most significant in cases where the sets of protons are both scalar and dipolar coupled. In this case, a percentage of the diagonal peak will contribute to the cross peak.

Calculation of Internuclear Distances

Internuclear distances calculated using both the initial rate and the two spin approximation are compared in Table III.4, and the distances calculated for each mixing time are shown graphically in Figure 3.12. Comparison of the values obtained from the two different methods shows that the two distances are within the estimated error, with the exception of the 1B-5A' distance and the 1C-2B distance. Since both methods give equivalent distances, it was decided to use the two spin approximation to calculate the distances for the following reasons:

- For some molecules, the mixing times chosen may be past the initial rate
- At short mixing times Hartmann Hahn effects may be prevalent, which will result in distorted values. This is especially true for cross peaks between distant protons where the intensity of the peak is low.
- It is difficult to distinguish in which regime the initial rate is valid, i.e. there are several possible lines that can be drawn to estimate the initial slope.
- Incorrect calculation of the initial slope for the ruler distance can lead to propagation of errors in the calculated distances.

Therefore, for all of the following structures, distances were calculated using the ratio of the integrated intensities as described. The distances obtained from the ROESY experiments were then used as constraints during molecular dynamics simulations.

Two sets of experiments were conducted on the trisaccharide (**4**). In order to effectively average the two data sets, one cross peak was referenced to the value obtained

in another set, and the remaining cross peaks were subsequently multiplied by the same factor.

The curves in Figures 3.11 and 3.12 are drawn through the average value, and the error bars shown are between the actual values of the two data sets. The variation can be summarized as follows. The uncertainty of the cross peak intensities is usually of the order of ±15 %. In some cases, i.e., for low intensity cross peaks, the uncertainty can be as high as ±25 %, whereas for other peaks the integrated intensities from the two data sets matches almost exactly.

Distances calculated from each of the data sets show a variation of ±5 %. The range of distances calculated from each mixing time showed a greater variation (between ± 5 and 10 %), and therefore the highest and lowest value of the distance was taken to be the upper and lower bound of the distance.

For all of the following compounds (**5**-**8**), a minimum of 2 data sets were acquired. The errors observed in the trisaccharide are typical of all of the compounds studied, and thus the variation of distance with mixing time, which is greater than the observed error, was taken to be the upper and lower bounds of the calculated distances. Distances calculated from each of the data sets show a variation of input as constraints Molecular dynamics simulations of the trisaccharide were carried out as described in the Experimental section. The constraints input for the trisaccharide (**4**) are listed in Table III.5.

Molecular Dynamics Simulations

The results of the molecular dynamics simulations are plotted in the graphs in Figures 3.13 and 3.14, and tabulated in Tables III.6 and III.7. It can be seen that, as expected, the constrained dynamics simulation provides limited variation in the parameters upon which constraints have been imposed. There is also little difference between the dynamics simulations calculated in H_2O and the simulation in *vacuo*, hence

the dynamics simulations of larger oligosaccharides were calculated exclusively in H_2O. Comparison between the constrained and unconstrained dynamics simulations shows that, while the distances and the glycosidic angles have a considerably greater range of freedom, the average values of the two dynamics simulations are not significantly different. However, in the A'-B linkage, even though the Φ and Ψ angles are within the same range for the constrained and unconstrained dynamics simulations ($\Phi = 37^\circ$ and 56° and $\Psi = 15^\circ$ and 31° for the constrained and unconstrained dynamics simulations, respectively), the distances between the protons across the glycosidic linkage, i.e., the 1A'-2B distance is increased from 2.14 Å to 2.73 Å .

CROSREL Calculation of the ROESY Buildup Curves

The coordinates from both dynamics simulations were used as input for the program CROSREL to simulate a series of ROESY buildup curves. A grid search was performed to estimate τ_c as described in the Experimental section. The results of these calculations are shown in the series of graphs in Figure 3.15. Inspection of these graphs reveals that most of the trends are reproduced qualitatively. One notable exception is the diagonal peak of the H6(pro S) of the Glc*p*NAc. This peak displays an oscillatory character that would indicate that there is still continued Hartmann Hahn transfer of the H6(pro S) diagonal peak, even at such long mixing times. Though this oscillatory effect is not reproduced accurately in the calculated buildup curves, the relative magnitude of the diagonal peak is. The second exception are the cross peaks calculated from the unconstrained dynamics for the anomeric proton of ring A'. In this case, the relative magnitudes of the cross peaks are actually reversed from the experimentally observed order. Experimentally, the 1A'-2B cross peak is observed to be the strongest, by far. In the calculated curves, the intra-ring 1A'-2A' cross peak is the strongest. This is a direct result of the differences observed in the dynamics. The A'-B linkage seems to be assuming a different conformation from the initial, constrained conformation, as

evidenced by the increase in the 1A'-2B distance. This results in the magnitude of the 1A'-2B cross peak decreasing, changing the order of the intensities. However, while the order of the 1A'-2B cross peaks calculated for the constrained dynamics is correct, the relative magnitudes are not. The 1A'-2B cross peak is exaggerated. This is also true of the cross peaks calculated from the anomeric proton of Ring C, i.e., the 1C-3B cross peak. Since all of the calculated distances are directly dependent on the ruler distances, any errors in assessing the ruler distances will be reflected in the calculated distances. One possibility is that differences in T_1s, caused either by the different environment of the various protons, motional anisotropy or internal motion, will modulate the ROEs differently, causing the magnitude of the cross peaks to vary considerably for different proton pairs, irrespective of the distance between them. This is evident when the H1-H2 cross peaks of the different Rha*p* rings is considered. Inspection of these peaks shows that the magnitudes vary significantly even though both rings are in the same chair conformation ($^4C_1$) and therefore should have the same H1-H2 distance. This variation in the observed intensity of the cross peaks of the ruler proton pairs, is, in all probability, one source of error.

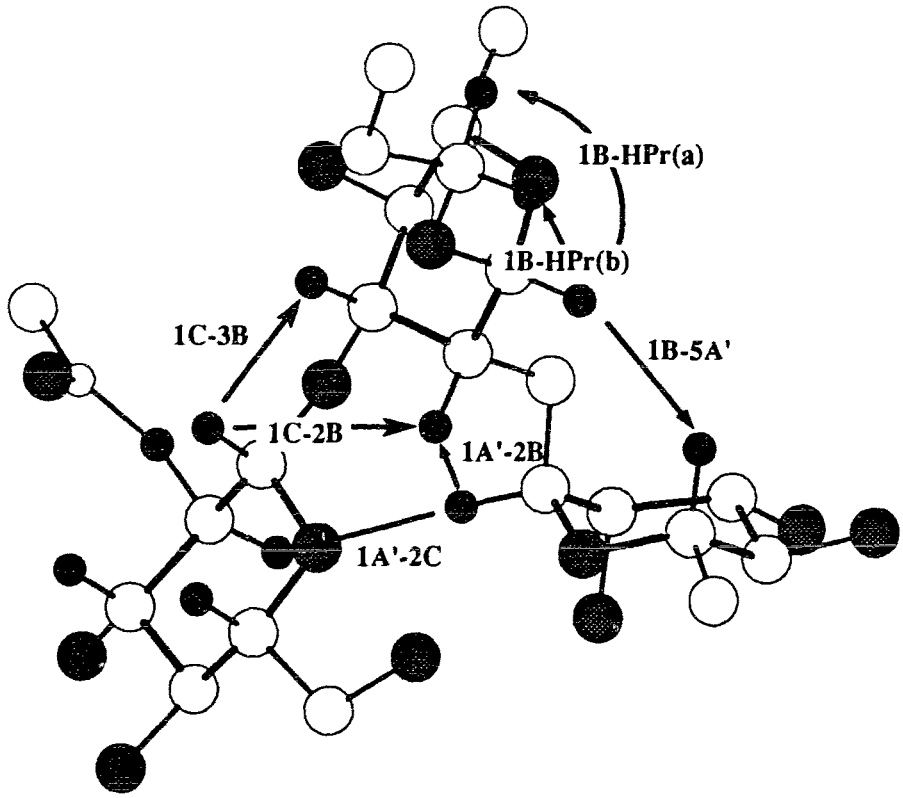


Figure 3.10. Interproton contacts derived from the ROESY spectrum of the trisaccharide (**4**).

Table III.4. A comparison of the internuclear distances [Å] obtained with the initial rate approximation, the distance obtained from a single experiment, and the average distance from the different mixing times.

Ring	Proton Pair	Distance calculated for a single mixing time (300 ms)	Average Distance calculated from all mixing times	Distance calculated with the initial rate appproximation
Ring A'	H1A'-H2B	2.17	2.10	2.09
	H1A'-H2C	2.96	2.70	2.63
Ring B	H1B-H5A'	2.60[1]	2.73[1]	2.38[1]
	H1B-HPr(a)			
	H1B-HPr(b)	2.65	2.59	2.51
Ring C	H1C-H2B	3.04	2.90	3.71
	H1C-H3B	2.23	2.25	2.21

[1]This distance was calculated by estimation of the H1B-HPr(a) distance, since the H1B-H5A' and H1B-HPr(a) cross peaks overlap.

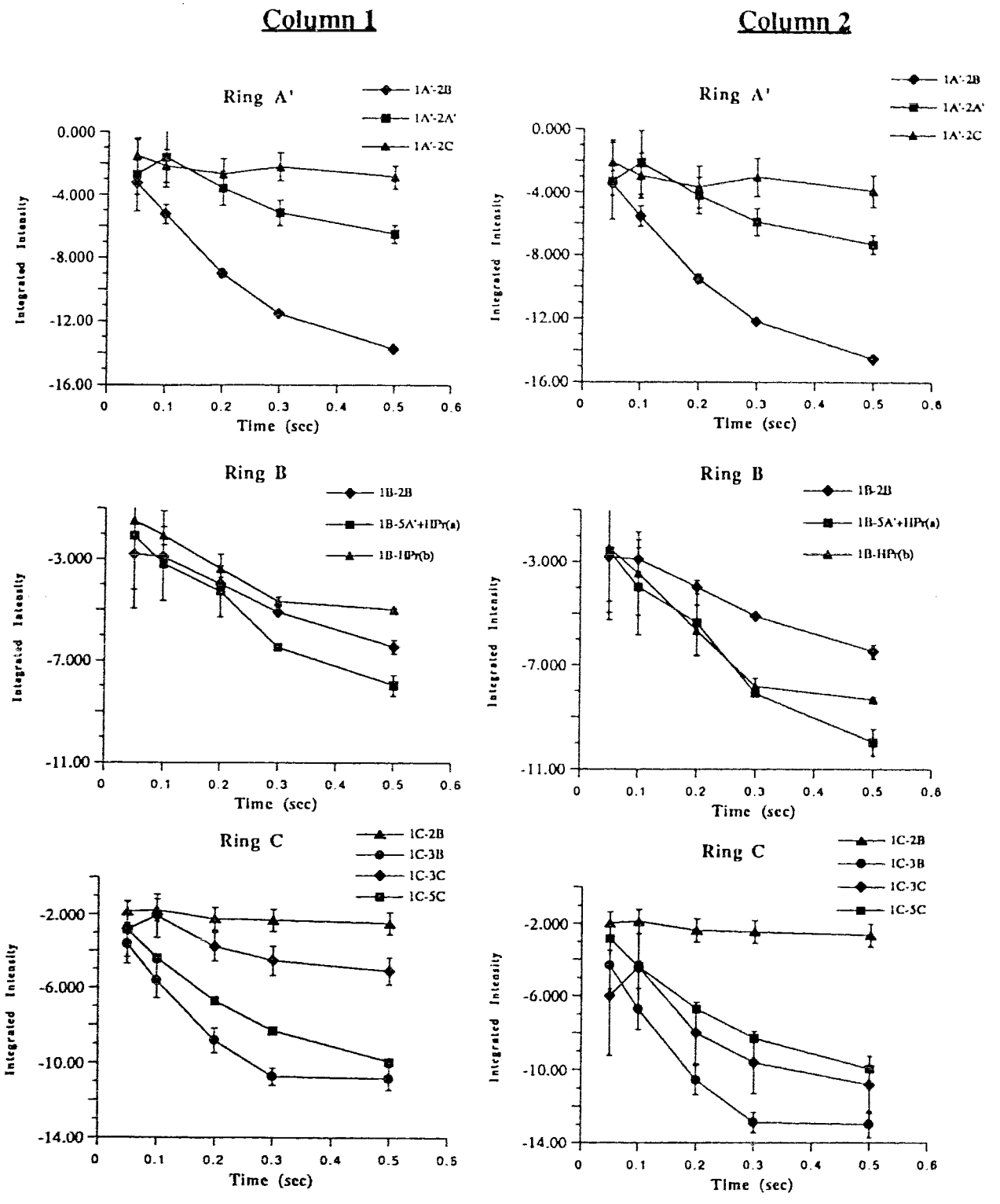


Figure 3.11. The effect of offset and Hartmann Hahn corrections on the ROE buildup of the trisaccharide (4). Column 1) Uncorrected data. Column 2) Corrected data.

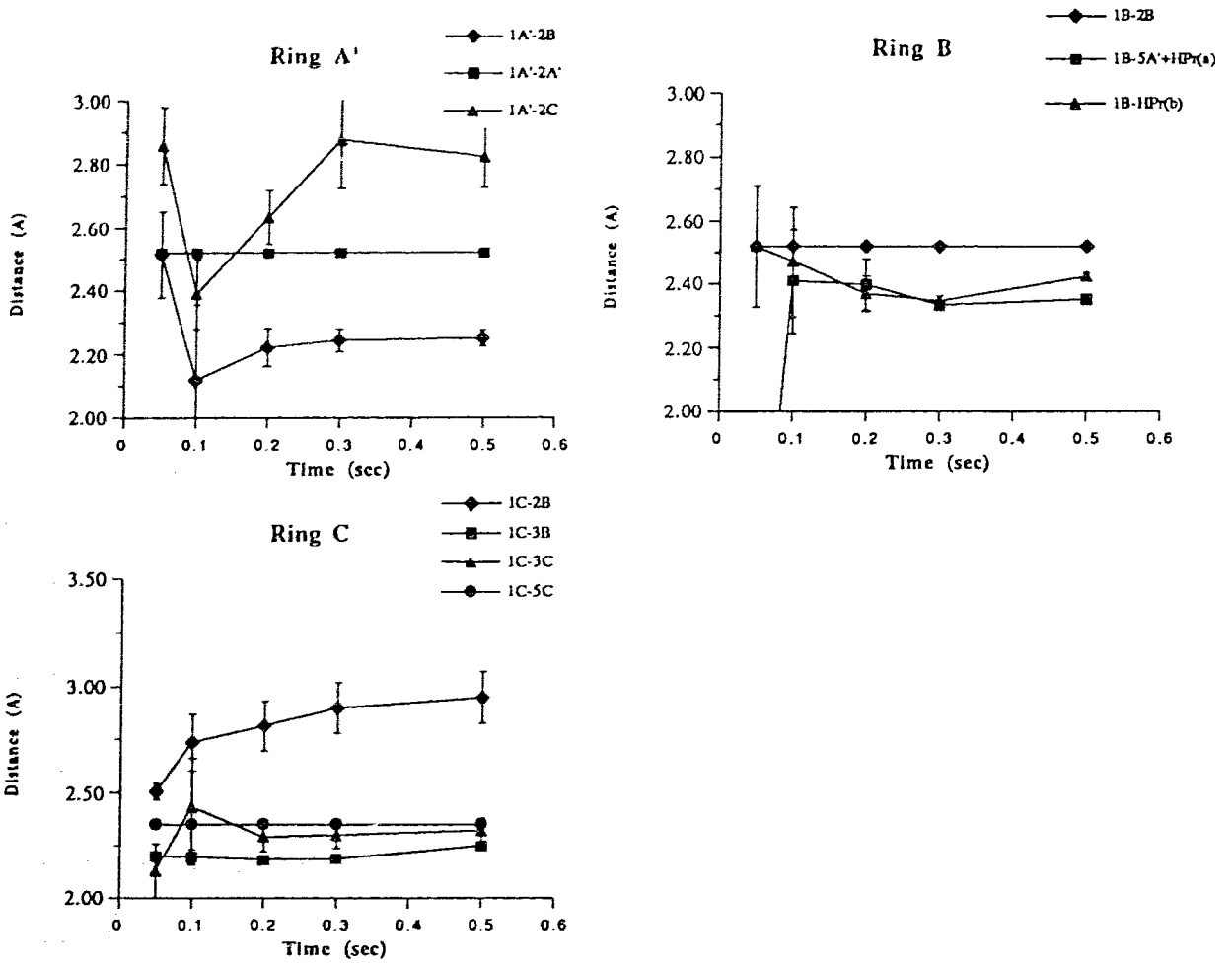


Figure 3.11. Interproton distances calculated from ROESY data of the branched trisaccharide (**4**).

Table III.5. Constraints input for the molecular dynamics simulation of the trisaccharide (**4**).

Proton Pair	Constrained distance [Å]	Upper limit[a]	Lower limit
H1A'-H2B	2.10	2.20	2.00
H1A'-H2C	2.70	2.90	2.60
H1B-H5A'	2.60	2.80	2.40
H1C-H2B	2.90	3.00	2.80
H1C-H3B	2.25	2.35	2.15

[a]The limits were derived from the highest and lowest values obtained from the experiment.

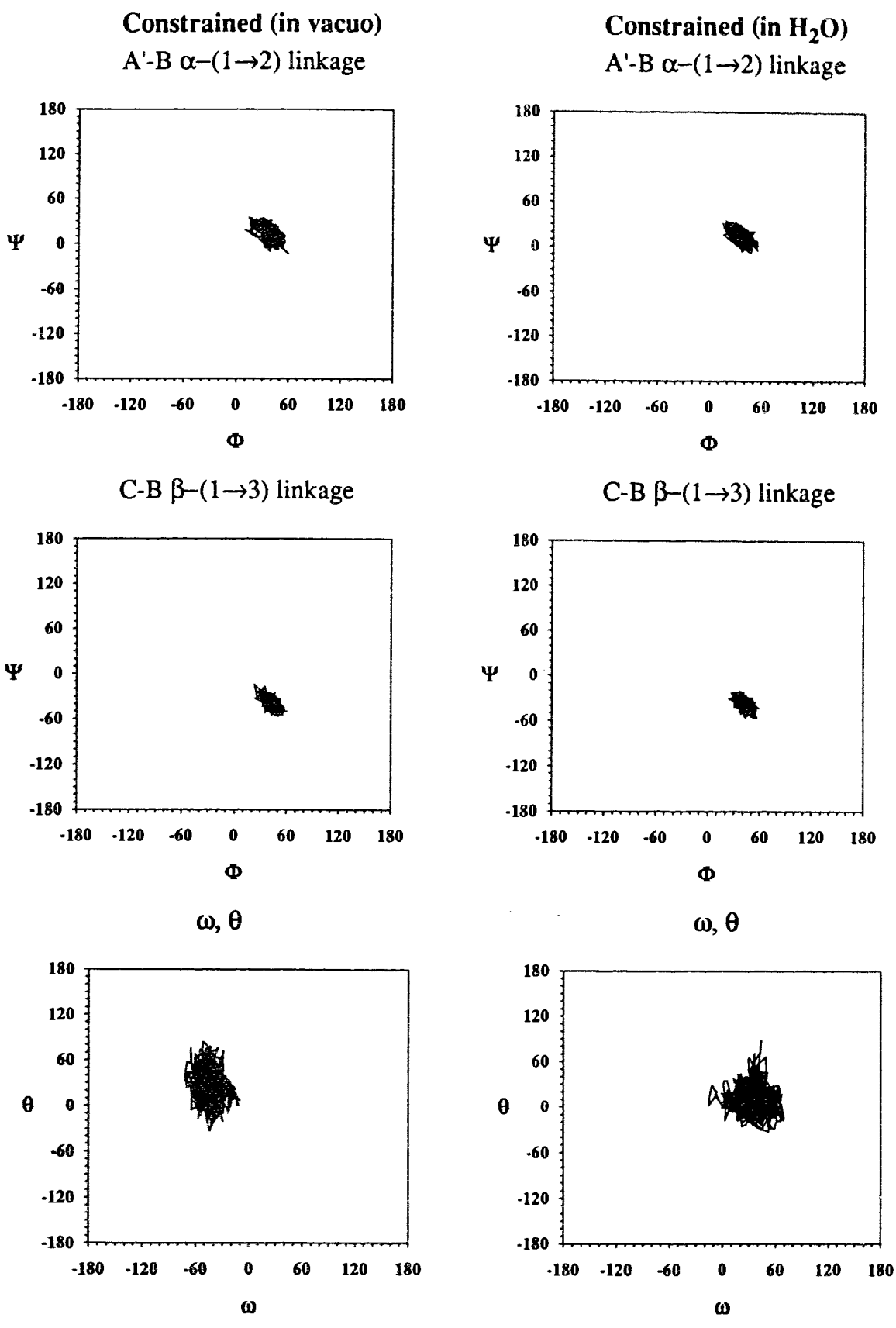


Figure 3.13. (Caption Overleaf)

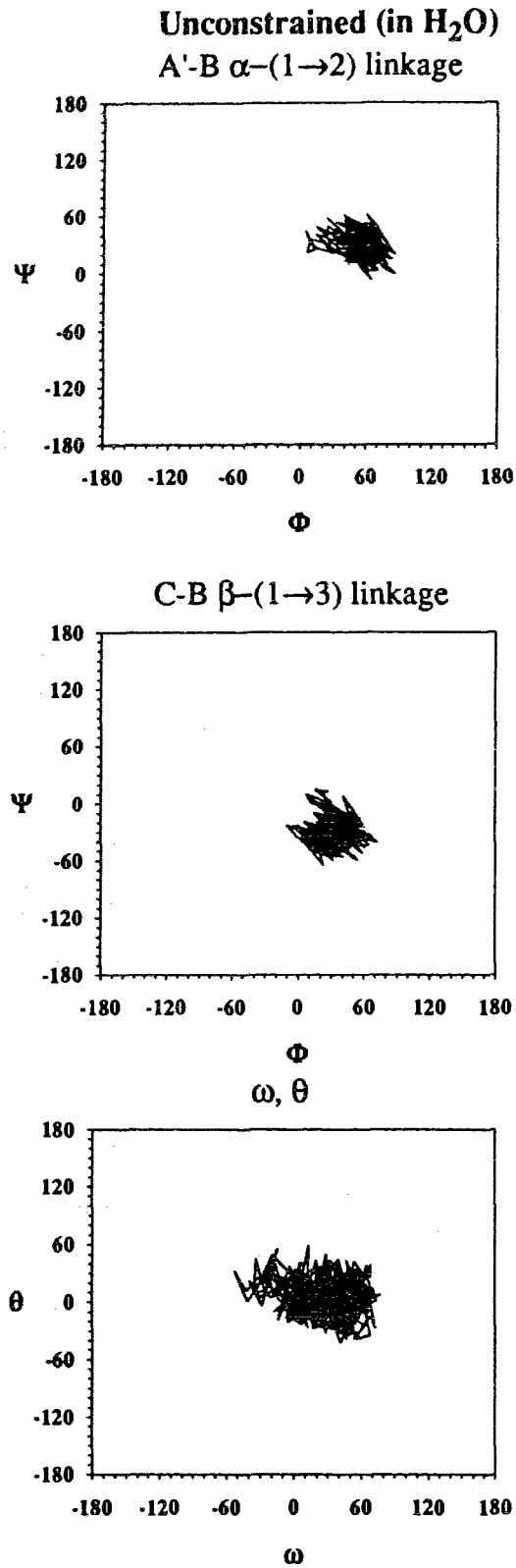


Figure 3.13. (Previous two pages). Variation of selected torsion angles during dynamics simulations of the trisaccharide (**4**). Column 1) Constrained dynamics in vacuo. Column 2) Constrained dynamics in H_2O. Column 3) Unconstrained dynamics in H_2O.

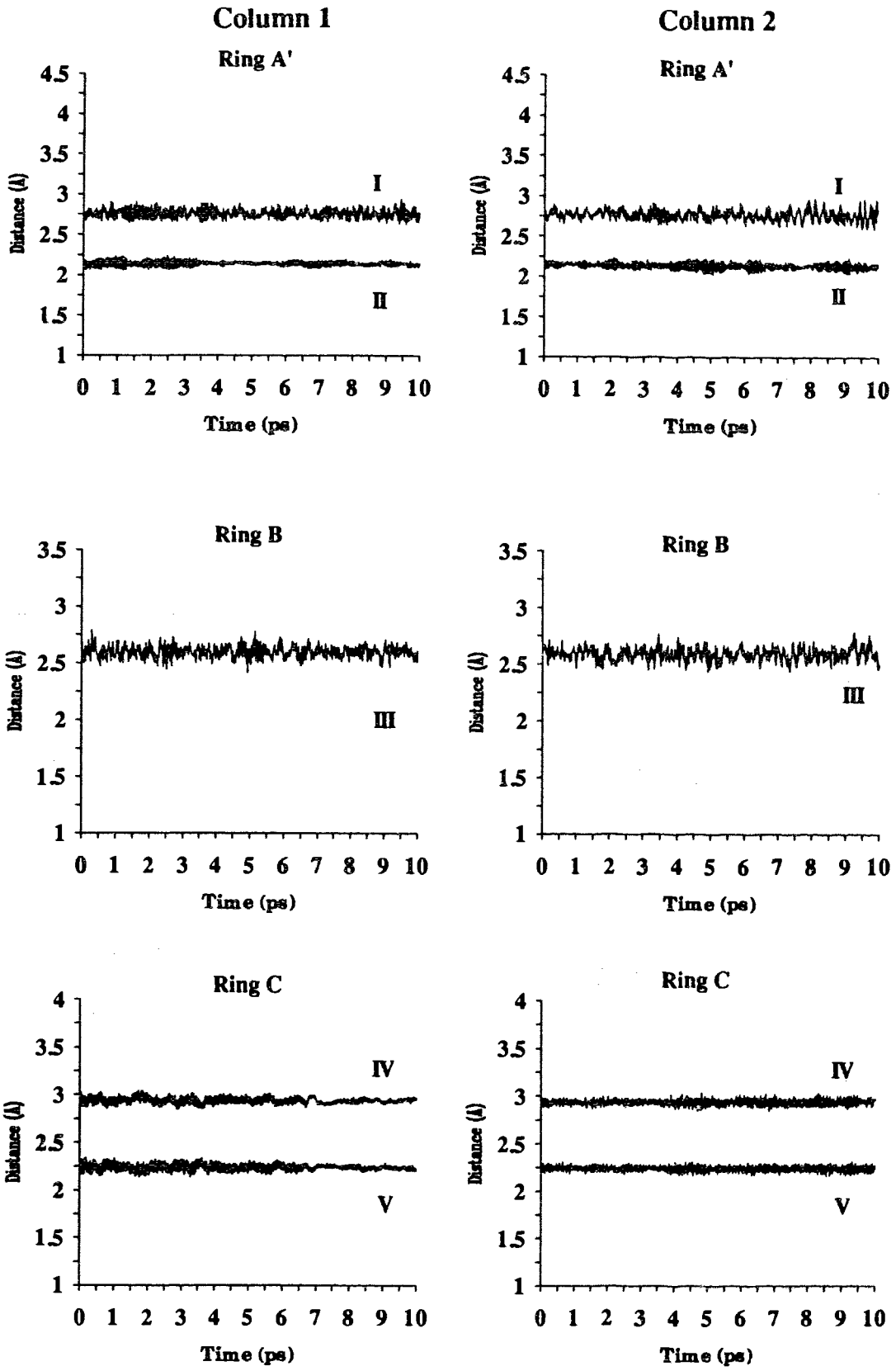


Figure 3.14. (continued overleaf).

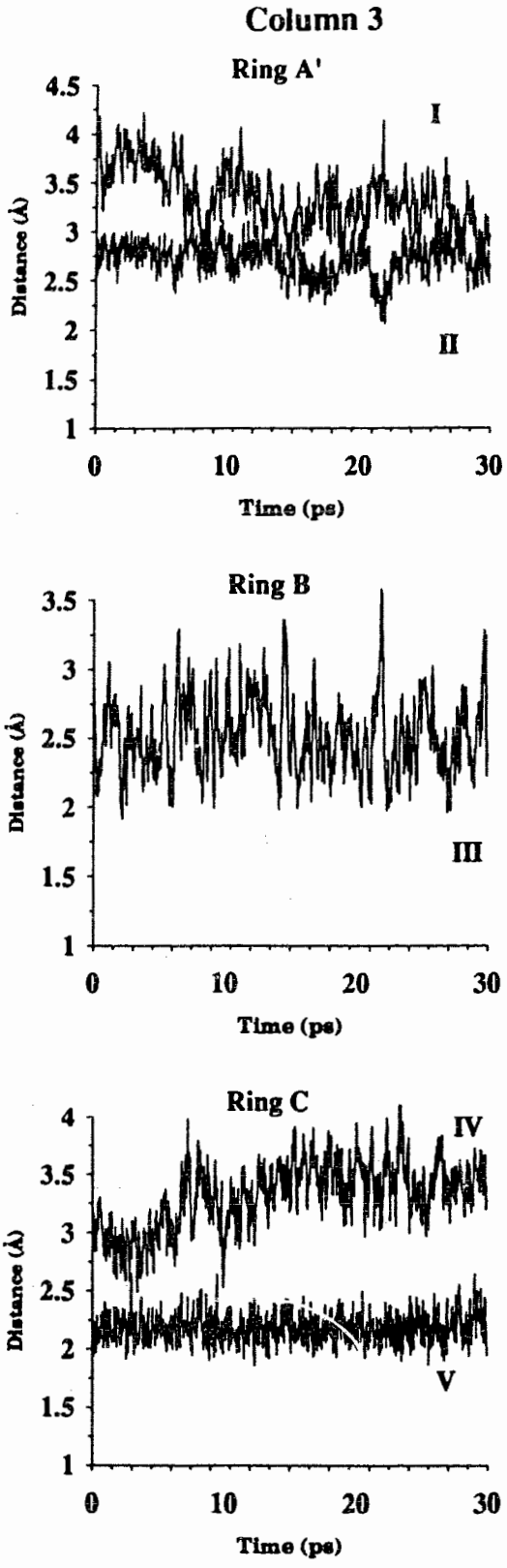


Figure 3.14. Variation of selected interproton distances during dynamics simulations of the trisaccharide (4). Column 1.) Constrained dynamics in vacuo. Column 2.) Constrained dynamics in H_2O. Column 3.) Unconstrained dynamics in H_2O.

Table III.6. Average, high, and low values of selected internuclear distances during dynamics simulationsof the trisaccharide (4)[a].

	Constrained Dynamics in vacuo			Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Proton Pair	Average	High	Low	Average	High	Low	Average	High	Low
1B-HPr(a)	2.50	3.07	2.02	2.48	3.26	1.98	2.53	3.71	1.99
1B-HPr(b)	3.00	3.67	2.07	2.99	3.73	2.24	2.99	3.84	2.03
1B-H5A'	2.60	2.79	2.42	2.59	2.79	2.44	2.52	3.9	1.89
1A'-H2B	2.14	2.23	2.05	2.14	2.23	2.03	2.71	3.24	2.06
1A'-H2C	2.76	2.94	2.64	2.77	2.97	2.60	3.34	4.28	2.22
1C-H2B	2.94	3.04	2.85	2.94	3.03	2.85	3.31	4.19	2.39
1C-H3B	2.24	2.34	2.14	2.24	2.31	2.17	2.18	2.65	1.83

[a]The average temperature during the dynamics simulation was 280 K for the constrained dynamics in vacuo, 301 K for the constrained dynamics in H_2O, and 310 K for the simulation without constraints.

Table III.7. Average, high, and low values of the Φ and Ψ angles during dynamics simulations of the trisaccharide (4)[a].

	Constrained Dynamics in vacuo			Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Angle	Average	High	Low	Average	High	Low	Average	High	Low
Link A-Propyl									
Φ	58.59	93.40	17.42	57.33	83.47	4.91	58.83	97.23	11.84
Link C-B									
Φ	43.73	60.30	22.99	42.69	59.16	23.64	34.9	72.38	-10.23
Ψ	-40.65	-14.45	-57.09	-39.18	-20.56	-57.14	-30.33	20.13	-64.35
Link A'-B									
Φ	38.59	59.43	9.63	36.79	56.01	15.22	56.6	90.61	0.61
Ψ	15.62	35.43	-12.79	14.98	34.88	-7.91	31.34	62.13	-6.37
ω and θ angles									
θ	25.87	84.28	-33.44	10.69	89.13	-32.83	5.39	61.47	-56.79
ω	-45.70	-10.28	-71.73	34.30	70.13	-15.57	31.6	77.29	-52.85

[a]The average temperature during the dynamics simulation was 280 K for the constrained dynamics in vacuo, 301 K for the constrained dynamics in H_2O, and 310 K for the simulation without constraints.

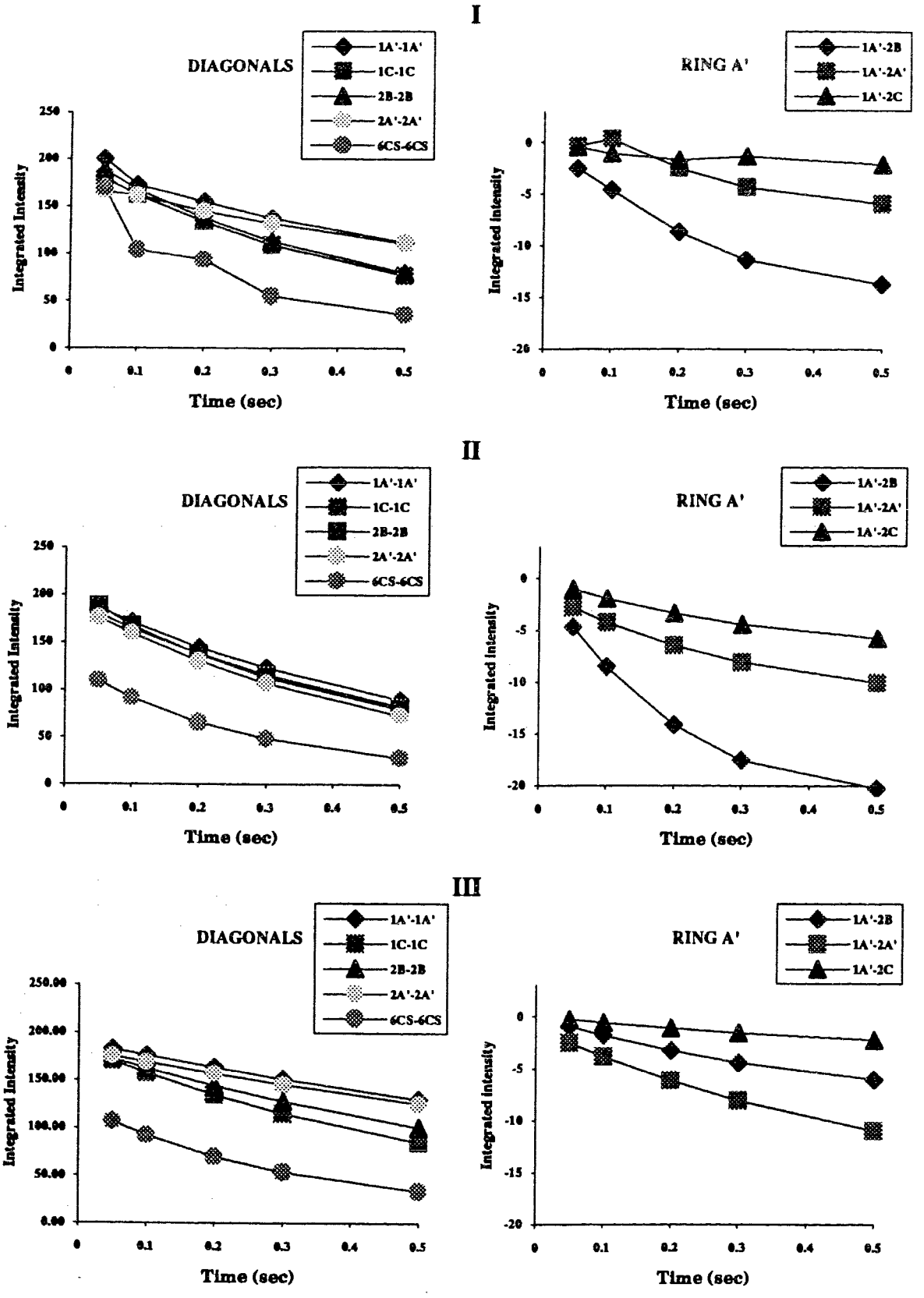


Figure 3.15. (Caption overleaf)

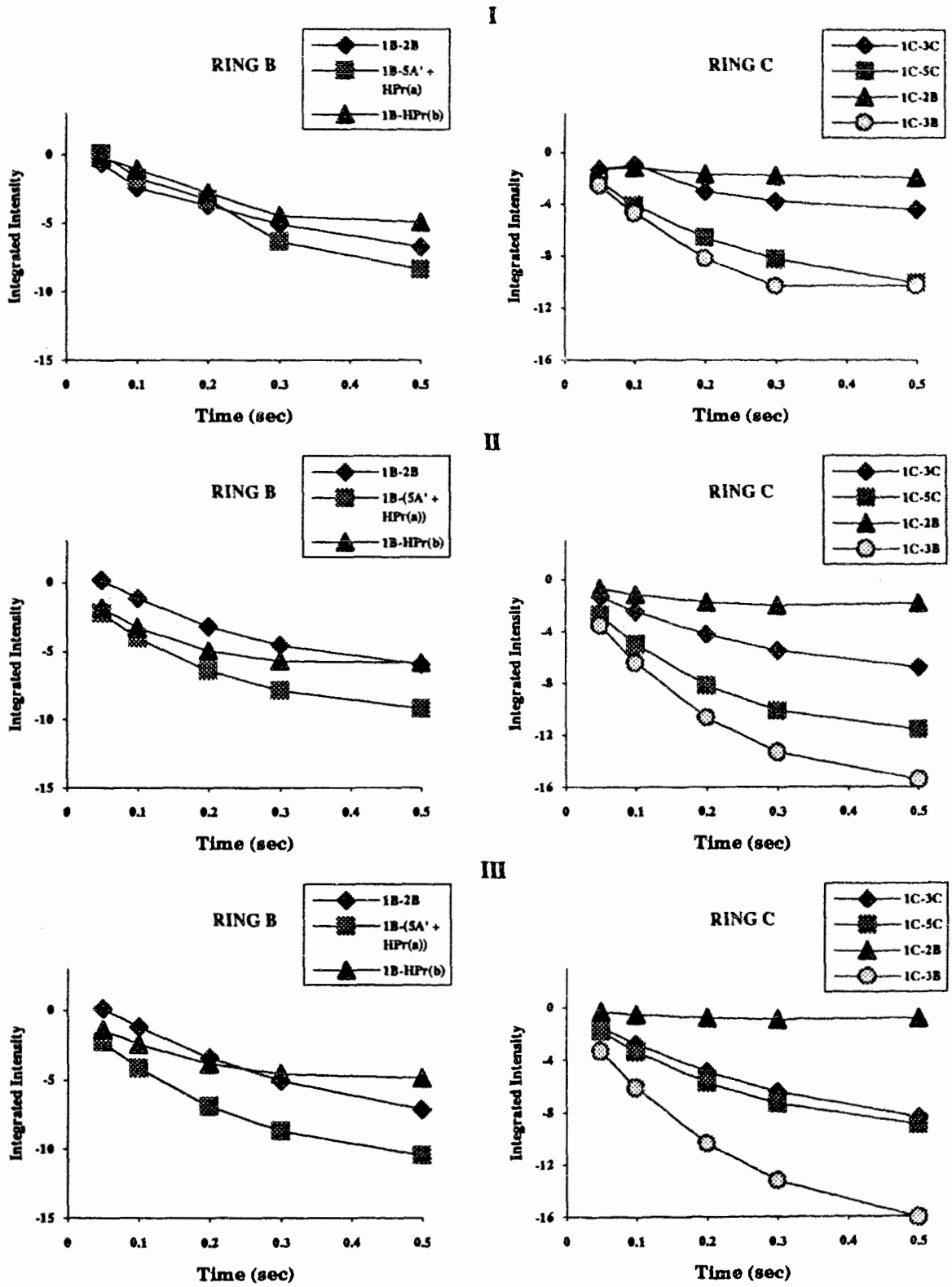


Figure 3.15. Comparison of the experimental and calculated ROESY buildup curves for **4**. I) Experimental buildup curves. II) Buildup calculated from the constrained dynamics trajectories in H_2O. III) Buildup calculated from the unconstrained dynamics trajectories in H_2O.

3.4.2. Tetrasaccharide (5)

ROESY Spectrum

The ROESY spectrum for the tetrasaccharide (**5**) acquired with a 500ms mixing time and the corresponding F1 slices are shown in Figures 3.16 and 3.17. Apart from the cross peaks from the anomeric proton of ring B to the H3 proton of Ring A across the glycosidic linkage, the similarities to the corresponding spectrum of the trisaccharide (**4**) are marked, which is not surprising since the tetrasaccharide is essentially the trisaccharide with a Rha*p* added α–(1→3) to the terminal (reducing end) residue. The A'-B linkage displays the characteristic peak from the anomeric proton of the B ring to the H5 proton of the A' ring, and the Glc*p*NAc H1 shows ROEs to the H3 and H2 protons of the B ring. Some interesting features of the spectrum appear due to the α–(1 →3) linkage. There is a very strong cross peak from the anomeric proton of ring B across the glycosidic linkage to the H3 proton of Ring A, and also to the H5 of ring A.

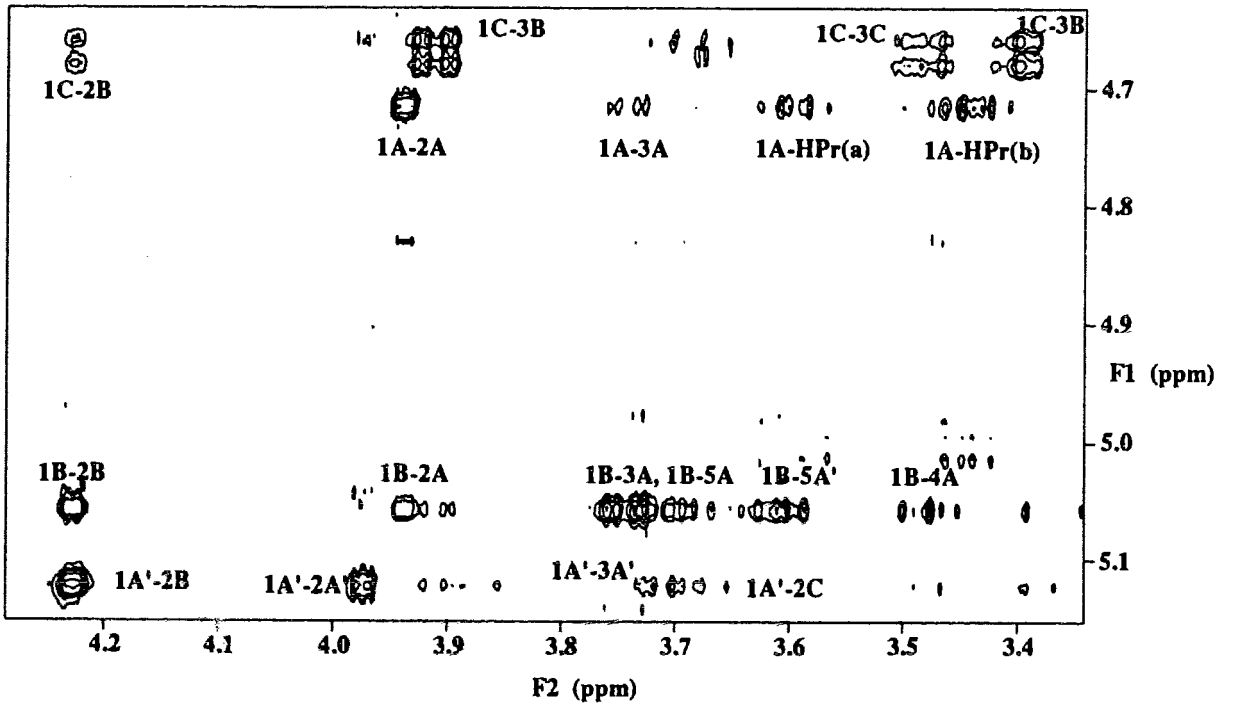


Figure 3.16. Expansion of the ROESY spectrum of the tetrasaccharide (**5**).

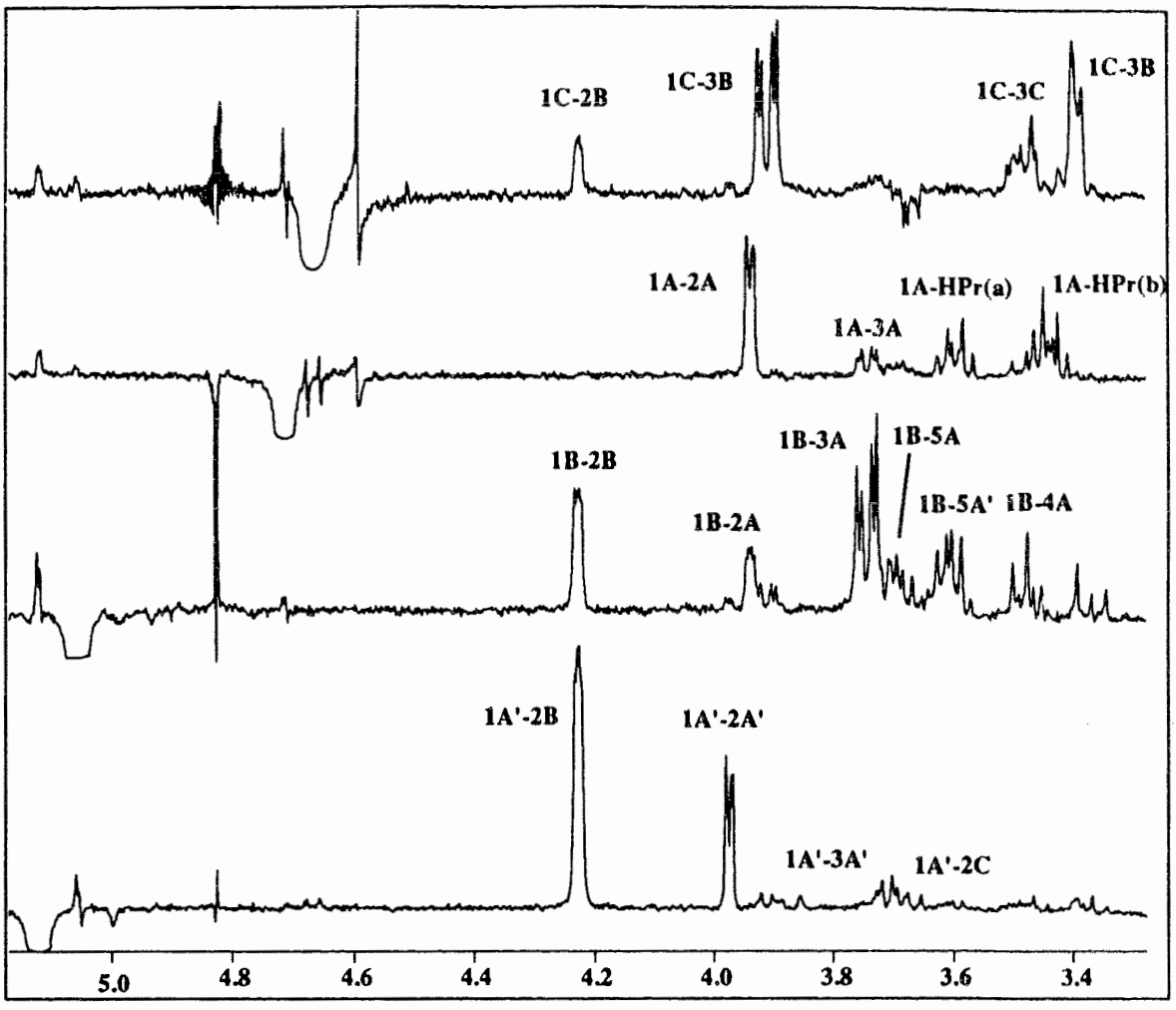


Figure 3.17. F1 slices from the ROESY spectrum of the tetrasaccharide (**5**).

Since there is no conformation in which the anomeric proton of ring B would be proximal to H5 of ring A, and H3 is not directly scalar coupled to H5, the ROE must be due to H3-H4-H5 TOCSY transfer, or relayed effects. However, a relayed ROE would have the opposite phase of the ROE cross peaks, (i.e., the same phase as the diagonal), and thus the H1B-H5A cross peak cannot be attributed to relay. This transfer does not appear in the C-B linkage which, although β, is also a (1→3) linkage, and places the protons H1 of the C ring (Glc*p*NAc), and the H3 and H5 of the B residue (Rha*p*) in the same relative positions. Therefore, the Φ and Ψ angles, which govern the relative geometry between these three protons, must play a pivotal role in the modulation of this

transfer. The interproton contacts observed in the ROESY spectrum of the tetrasaccharide (**5**) are displayed in Figure 3.18.

Offset and Hartmann Hahn Corrections

The ROE data was analyzed as described for the trisaccharide. Hartmann Hahn factors and offset corrections were computed (Table III.8) and the experimental data was corrected for both effects. The experimental data and data corrected for offset and HOHAHA effects are plotted in Figure 3.19.

Table III.8. Offset and Hartmann Hahn effects calculated for the tetrasaccharide (**5**).

Ring	Cross peak	$sin\alpha_I\ sin\alpha_S$	HH transfer	s^2
Ring A	1A-2A	0.9944	a)2A-1A	0.01
			b)2A-3A	0.11
	1A-HPr(a)	0.9918	a)HPr(a)-HPr(b)	0.04
			b)HPr(a)-H3Pr	0.57
	1A-HPr(b)	0.9897	a)HPr(a)-HPr(b)	0.04
			b)HPr(b)-H3Pr	0.57
Ring A'	1A'-2B	0.9969	a)1B-2B	0.03
			b)2B-3B	0.07
	1A'-2A'	0.9942	a)1A'-2A'	0.01
			b)2A'-3A'	0.08
	1A'-2C	0.9901	a)1C-2C	0.42
			b)2C-3C	0.07
Ring B	1B-2B	0.9971	a)2B-1B	0.03
			b)2B-3B	0.07
	1B-2A	0.9940	a)1A-2A	0.01
			b)2A-3A	0.11
	1B-3A	0.9914	a)2A-3A	0.11
			b)3A-4A	0.27
	1B-5A'	0.9892	a)5A'-6A'	0.00
			b)5A'-4A'	0.28
Ring C	1C-5C	0.9856	a)5C-4C	1.00
			b)5C-6Cs	0.02
			c)5C-6CR	0.00
	1C-3C	0.9874	a)3C-4C	0.72
			b)3C-2C	0.42
	1C-2B	0.9974	a)2B-1B	0.03
			b)2B-3B	0.07
	1C-3B	0.9941	a)3B-2B	0.07
			b)3B-4B	0.14

Calculation of Internuclear Distances

After correction for offset and HOHAHA effects , the inter proton distances were calculated for each of the mixing times and averaged (Figure 3.20). The distances obtained for the series of mixing times showed variations of ~0.3 Å. These distances (Table III.5) were then used as constraints, and molecular dynamics simulations were carried out.

Molecular Dynamics Simulations

The results of the dynamics simulations are displayed in Figures 3.21-3.24, and exhibit essentially the same behavior as the trisaccharide. The Φ, Ψ maps (Figures 3.21, 3.22) and the plots of the distances (Figures 3.23, 3.24) during the dynamics simulations show a greater variation during the unconstrained dynamics simulations. The Φ and Ψ angles for each of the linkages show an increased freedom around essentially the same region during the unconstrained dynamics simulation as the constrained dynamics simulation. A comparison of the distance trajectories does reveal an interesting fact. Although all of the dihedral angles are confined to essentially the same region, the distances obtained from the constrained and unconstrained dynamics simulations vary by as much as 0.5 Å. This is most evident in the 1A'-2B and the 1C-2B distance. The average values of the distances and Φ and Ψ angles are tabulated in Tables III.10 and III.11.

CROSREL Calculation of the ROESY Buildup Curves

The dynamics trajectories were input into CROSREL and ROESY buildups were calculated. A τ_c of 400 ps and an R_S value of 0.1 Hz were found to best fit the experimental data. Calculated and experimental ROESY curves for the tetrasaccharide are displayed in Figure 3.24.

A comparison of the calculated and experimental ROESY buildup curves reveals a better fit for those calculated from the constrained dynamics trajectories. The fit is not as good as for the trisaccharide, especially in the case of the A' and B' rings. Experimentally, the intensities of the 1A'-2B cross peaks are found to be greater than the intra ring 1A'-2A' cross peak. This is not reproduced in the calculated buildup, where the intra ring cross peak has the greatest intensity which indicates that the 1A'-2B distance is overestimated and is actually shorter than calculated. The CROSREL calculation of the ROESY buildup curves of the cross peaks to H1B displays discrepancies in the relative magnitude of the 1B-2B and 1B-5A' contacts, which indicates that the 1B-5A' and 1B-3A distances are incorrect.

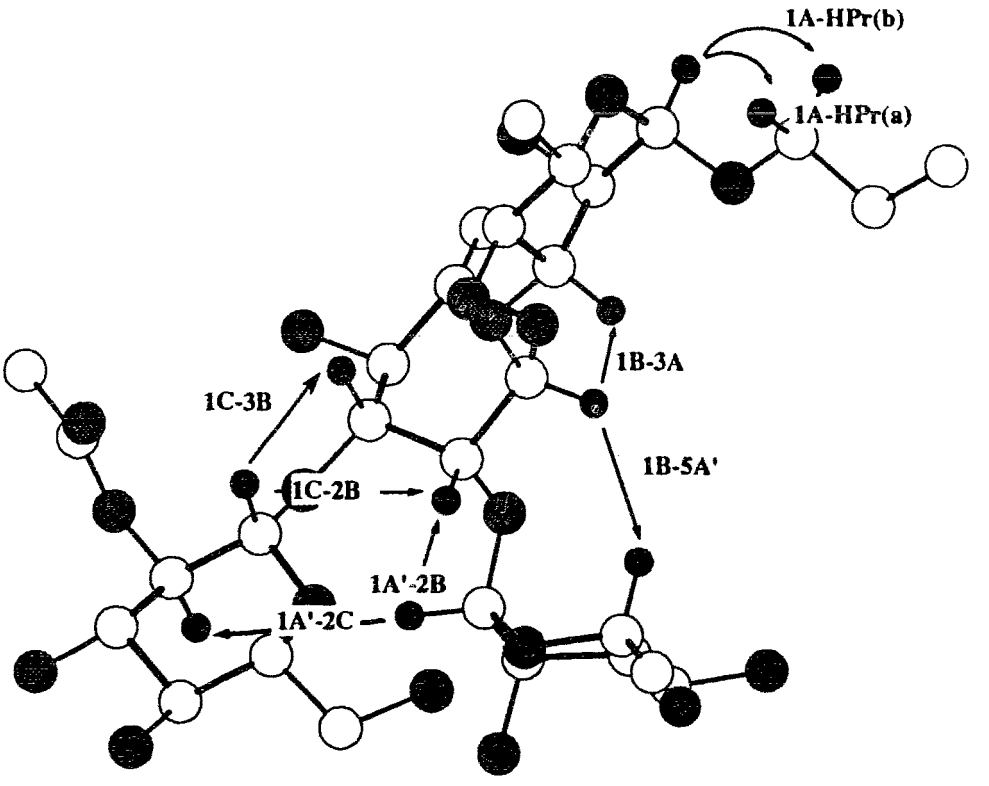


Figure 3.18. Interproton contacts derived from the ROESY spectrum of the tetrasaccharide (**5**).

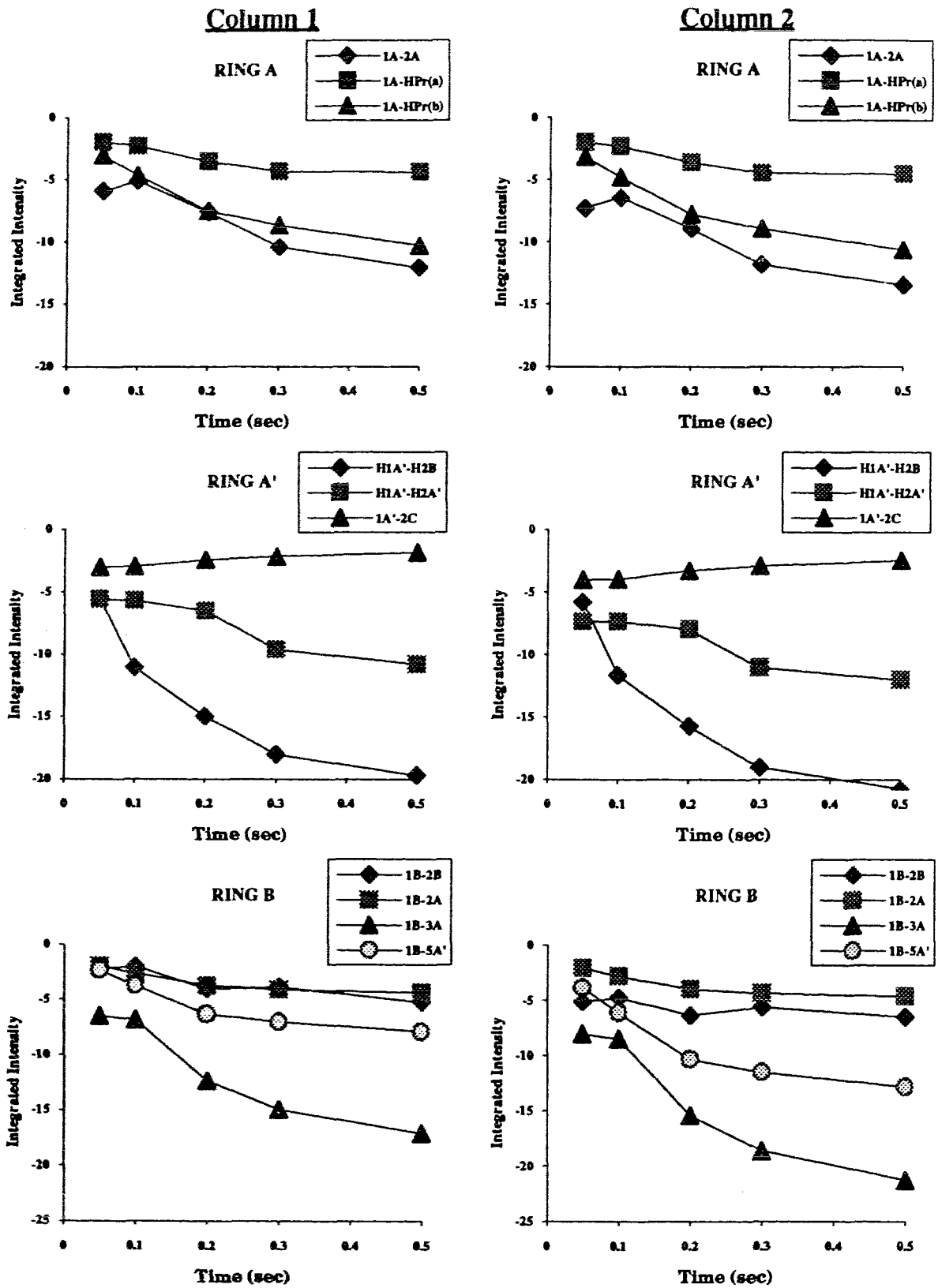


Figure 3.19. (Caption overleaf).

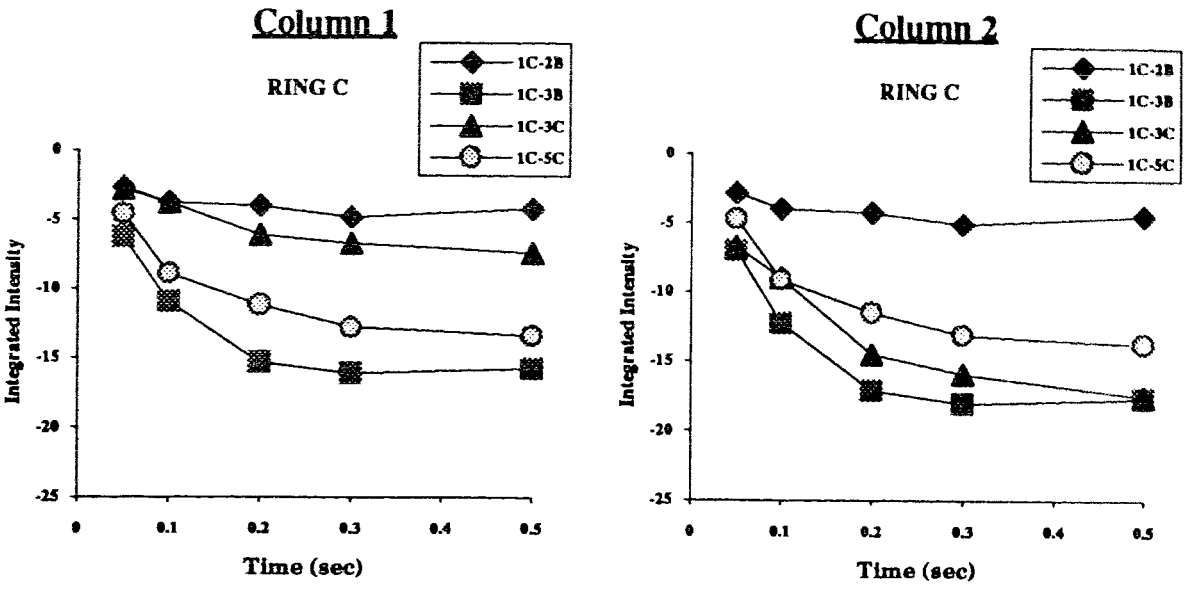


Figure 3.19. (This page and previous page). The effect of offset and Hartmann Hahn corrections on the ROE buildup of the tetrasaccharide (**5**). Column 1.) Uncorrected data. Column 2.) Corrected data.

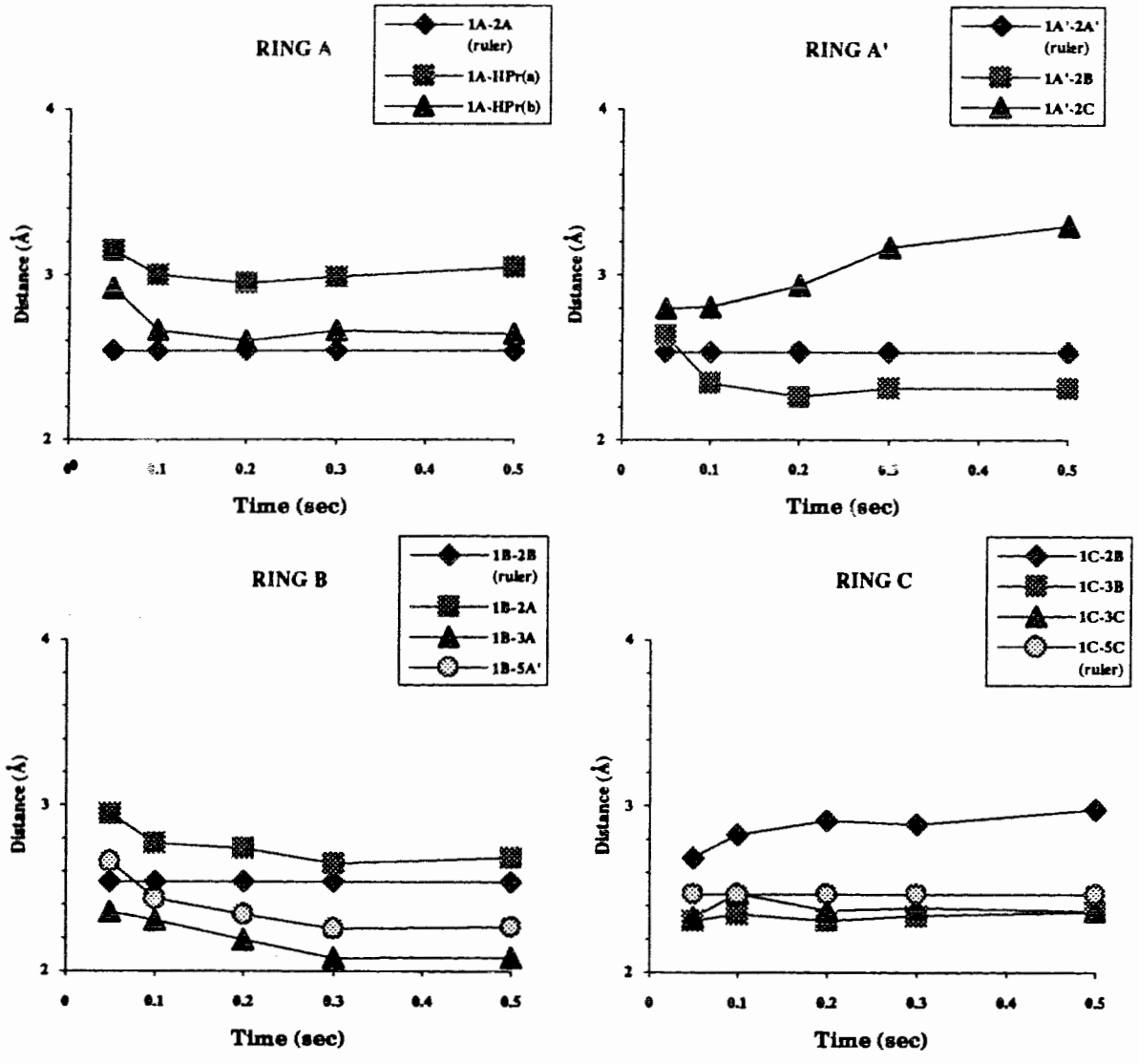
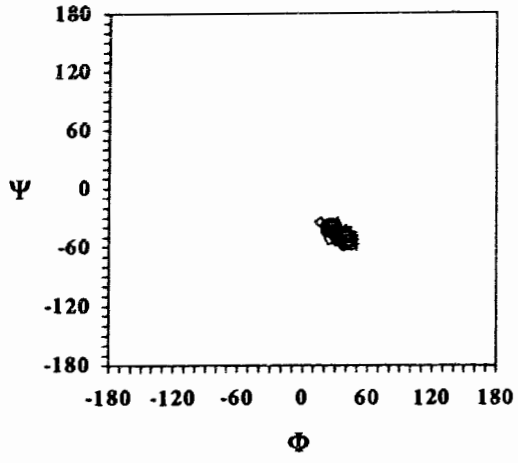


Figure 3.20. Inter proton distances calculated from ROESY data of the tetrasaccharide (**5**).

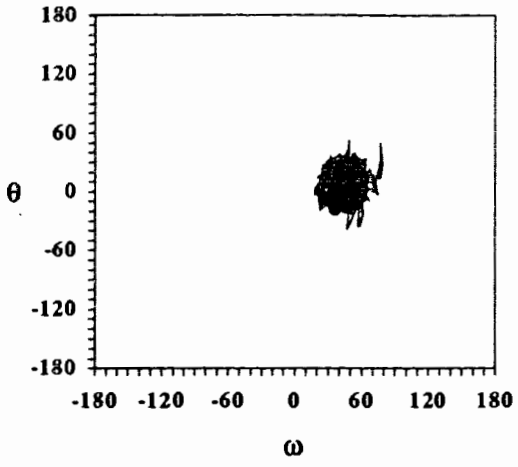
Table III.9. Constraints input for the molecular dynamics simulation of the tetrasaccharide (**5**).

Proton Pair	Constrained distance [Å]	Upper limit	Lower limit
H1A'-H2B	2.38	2.63	2.26
H1A'-H2C	3.00	3.30	2.80
H1B-H2A	2.76	2.96	2.64
H1B-H3A	2.20	2.35	2.07
H1B-H5A'	2.39	2.64	2.25
H1C-H2B	2.86	2.97	2.75
H1C-H3B	2.33	2.43	2.23

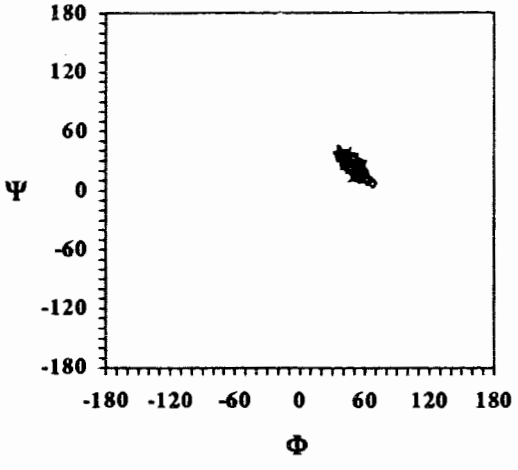
[a]The limits were derived from the highest and lowest values obtained from the experiment.



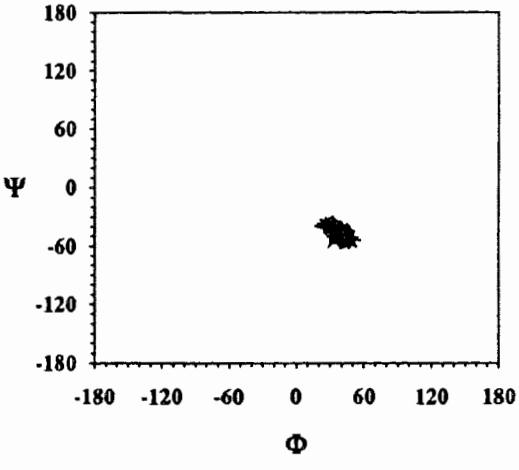
B-A α-(1→3) linkage.



ω and θ angles.

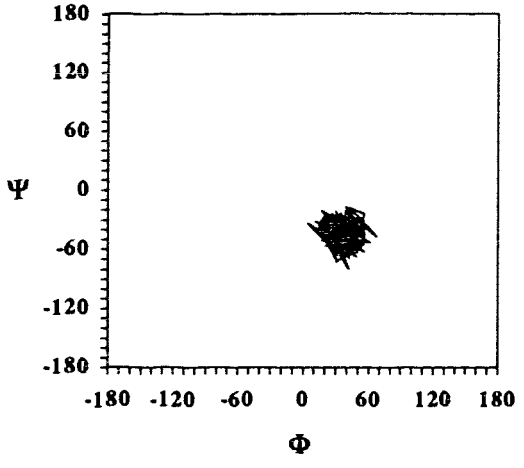


A'-B α-(1-2) linkage

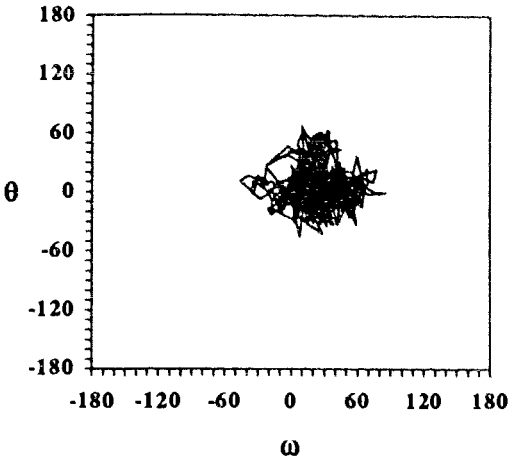


C-B β-(1→3) linkage

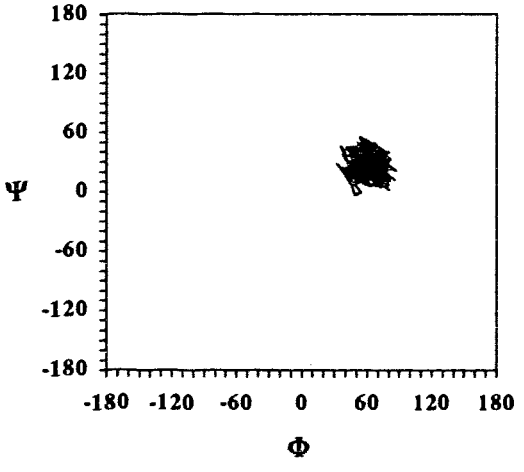
Figure 3.21. Variation of selected torsion angles during constrained dynamics simulations of the tetrasaccharide (**5**).



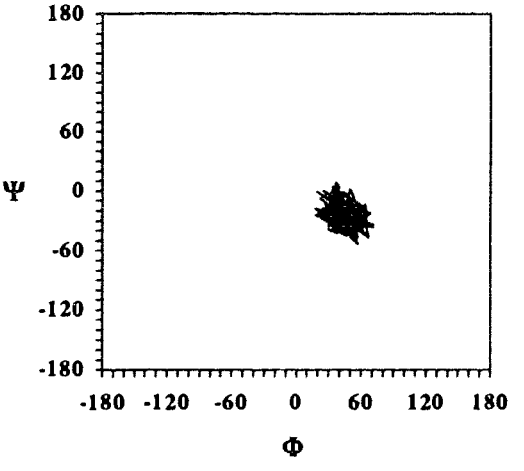
B-A α-(1→3) linkage.



ω and θ angles.

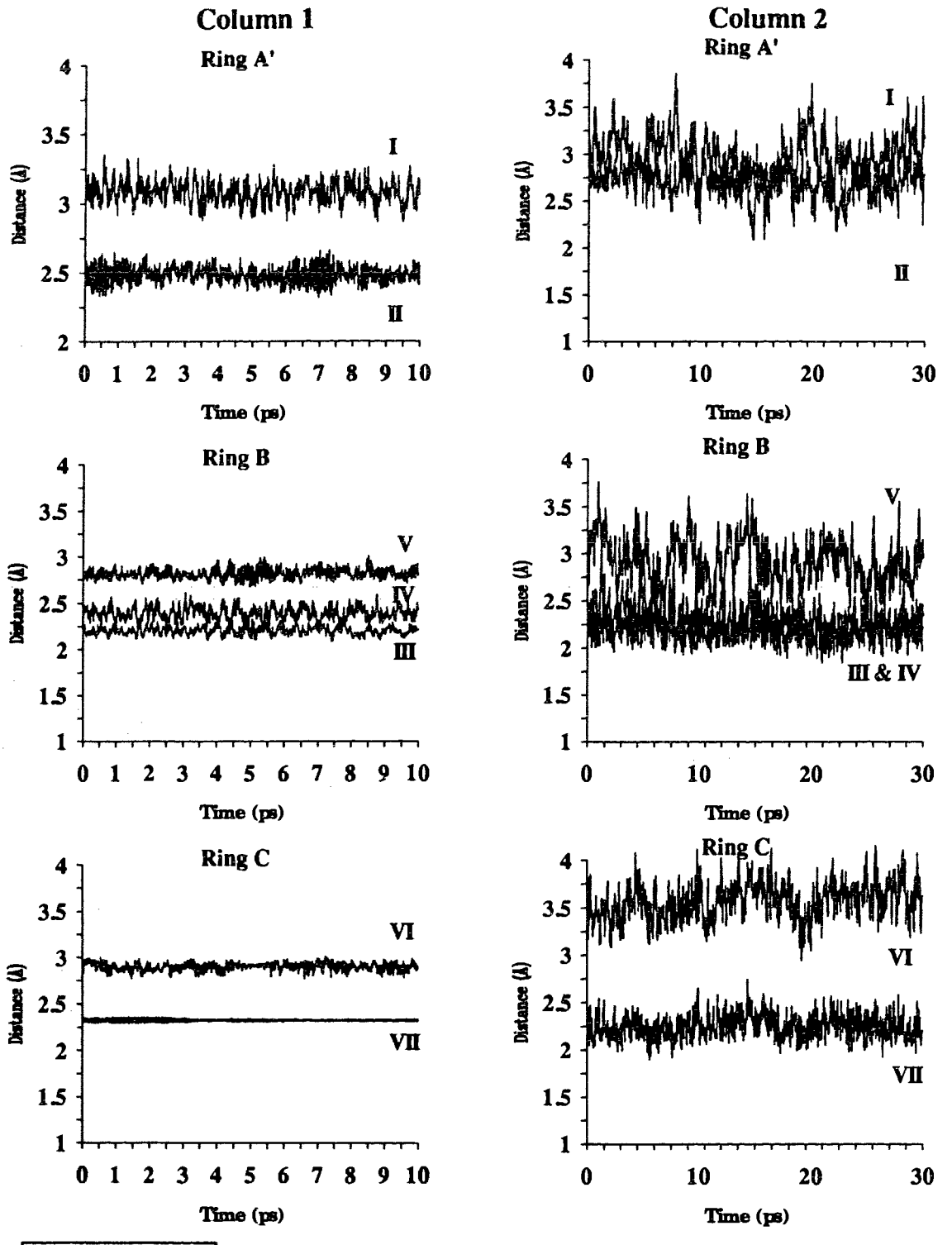


A'-B α-(1-2) linkage



C-B β-(1→3) linkage

Figure 3.22. Variation of selected torsion angles during dynamics simulations of the tetrasaccharide (**5**).



LEGEND	
I)	H1A'-H2C
II)	H1A'-H2B
III)	H1B-H3A
IV)	H1B-H5A
V)	H1B-2A
VI)	H1C-H2B
VII)	H1C-H3B

Figure 3.23. Variation of selected interproton distances during dynamics simulations of the tetrasaccharide (**5**). Column 1) Constrained dynamics in H_2O. Column 2) Unconstrained dynamics in H_2O.

Table III.10. Average, high, and low values of selected interproton distances during dynamics simulations of the tetrasaccharide (5)[a].

	Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Proton Pair	Average	High	Low	Average	High	Low
1A-HPr(a)				2.95	3.83	2.06
1A-HPr(b)				2.46	3.17	1.98
1B-H2A	2.82	3.01	2.68	2.92	3.79	2.15
1B-H3A	2.21	2.36	2.08	2.20	2.70	1.84
1B-H5A'	2.39	2.62	2.17	2.29	3.22	1.84
1A'-H2B	2.49	2.66	2.32	2.72	3.23	2.18
1A'-H2C	3.08	3.36	2.87	2.96	3.86	2.04
1C-H2B	2.90	3.02	2.78	3.57	4.20	2.90
1C-H3B	2.33	2.36	2.29	2.25	2.74	1.90

[a]Average temperature during the dynamics simulation was 284 K for the dynamics simulation with constraints and 291 K for the simulation without constraints.

Table III.11. Average, high, and low values of the glycosidic Φ and Ψ angles during dynamics simulations of the tetrasaccharide (5)[a].

	Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Angle	Average	High	Low	Average	High	Low
Link A-Propyl						
Φ	54.75	87.50	-0.39	55.66	87.02	17.97
Link B-A						
Φ	35.32	52.12	11.26	36.12	67.98	4.18
Ψ	-48.61	-28.95	-64.22	-45.67	-12.35	-79.53
Link C-B						
Φ	37.73	57.02	16.32	45.47	72.09	13.57
Ψ	-46.19	-29.44	-62.63	-24.85	8.92	-54.56
Link A'-B						
Φ	49.90	71.59	30.58	61.72	88.69	27.44
Ψ	24.35	45.18	2.64	27.42	55.27	-3.89
ω and θ angles						
Ring C						
θ	6.77	52.41	-38.53	3.59	72.20	-44.48
ω	43.30	78.90	17.64	25.23	85.52	-46.05

[a]Average temperature during the dynamics simulation was 284 K for the dynamics simulation with constraints and 291 K for the simulation without constraints.

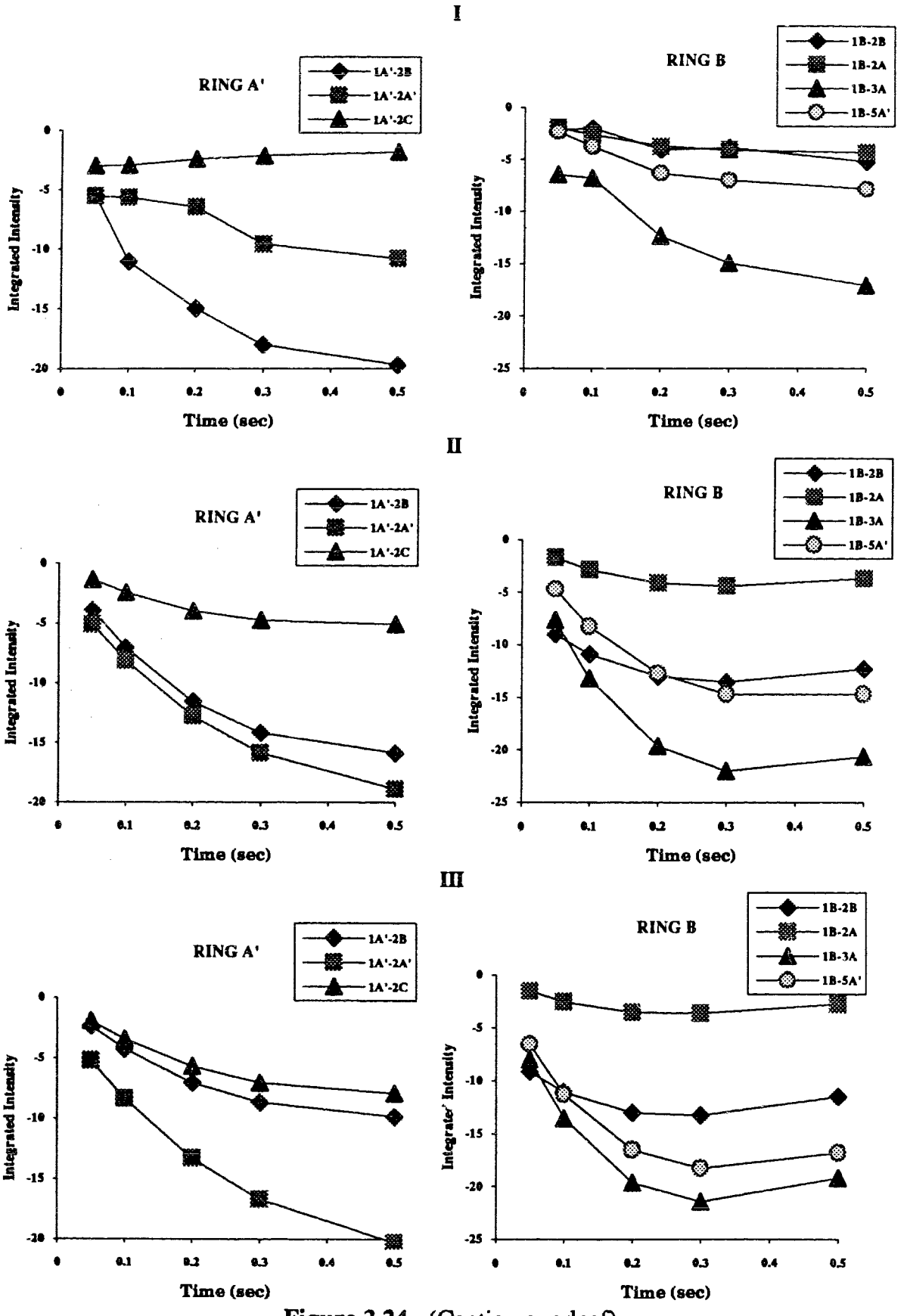


Figure 3.24. (Caption overleaf)

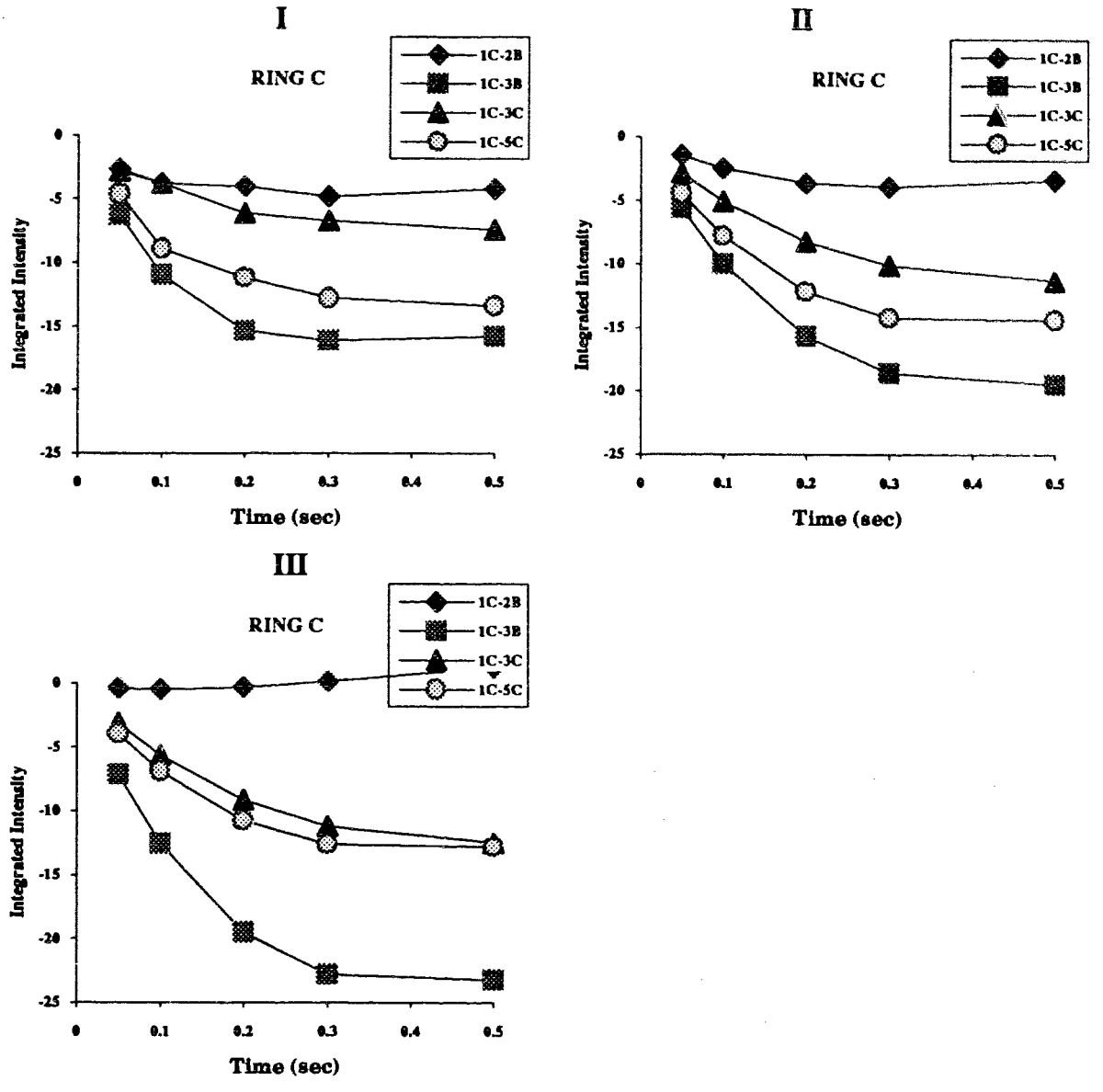


Figure 3.24. Comparison of the experimental and calculated ROESY buildup curves for the tetrasaccharide (**5**). I) Experimental buildup curves. II) Buildup curves calculated from the constrained dynamics trajectories in H_2O. III) Buildup curves calculated from the unconstrained dynamics trajectories in H_2O.

3.4.3. Pentasaccharide (6)

Roesy Spectrum

The anomeric region of the ROESY spectrum of the pentasaccharide (**6**) and the corresponding F1 slices are shown in Figures 3.26 and 3.27. In comparison to the trisaccharide (Figure 3.25), the pentasaccharide has two additional residues, the B' and C' rings, which are comprised of a Rha*p* (the B' ring) α–(1→3) linked to the A' ring and a Glc*p*NAc ring (the C' ring) linked β–(1→3) to the B' ring (Figure 3.25). The B-C linkage displays almost identical peaks as the B'-C' linkage. As is the case with the B-C linkage in the trisaccharide (**4**) and the tetrasaccharide (**5**), besides the two intra residue cross peaks to H3 and H5, cross peaks are visible to the H3 and H2 of the aglycons (i.e., the C and C' rings). The key interactions which unambiguously determine the conformation of the A'-B linkage, namely cross peaks corresponding to interactions between H1A'-H2C and H1B-H5A', are visible in the pentasaccharide. The well delineated chemical shift of the 5A' proton helps resolve any ambiguity in the assignment of the 1B'-5A' cross peak that may have been present in the tetrasaccharide.

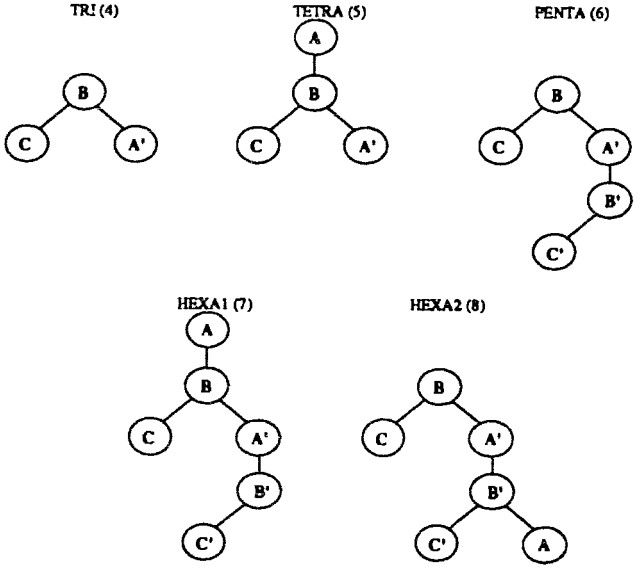


Figure 3.25. Schematic representation of the compounds **4-8.**

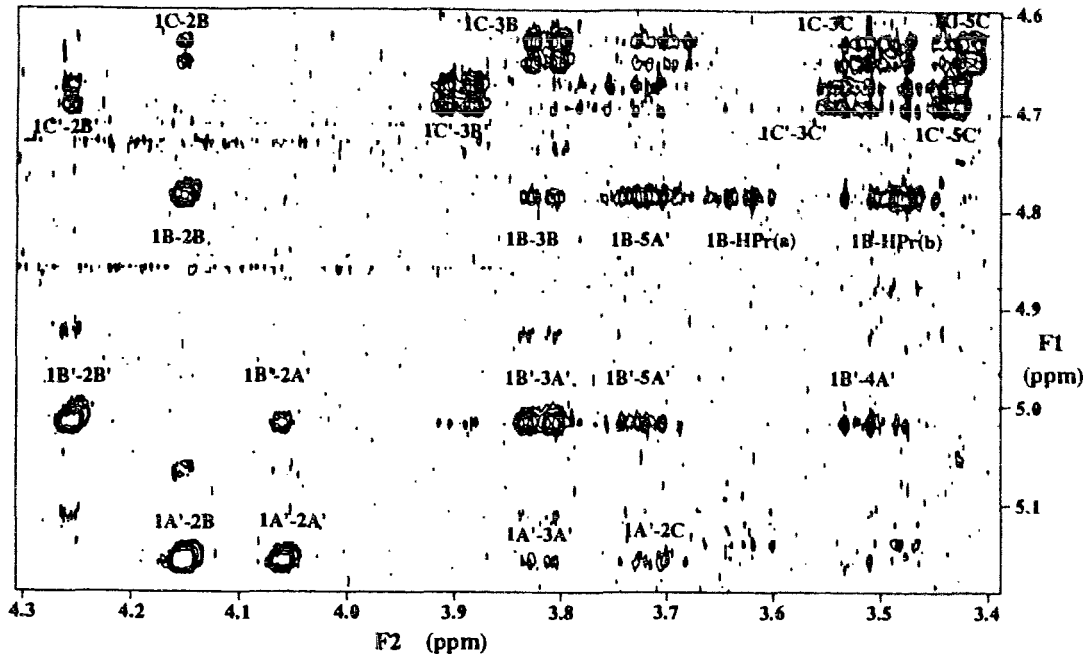


Figure 3.26. Expansion of the ROESY spectrum of the pentasaccharide (**6**).

A very weak ROE between the 1A' and 2C protons is identified in the ROESY spectrum of the pentasaccharide. The B'-A' linkage shows similar peaks to the analagous B-A linkage in the tetrasaccharide (**4**). The 1B' proton shows cross peaks to all the protons in the A' ring. It is impossible to distinguish which of these are genuine ROEs and which are due to TOCSY transfer. However, one additional piece of data was observed. The H2 proton of the A' ring shows a weak ROE to the H5 of the B' ring (see discussion). This indicates that the H1B'-H2A' cross peak may be due to TOCSY, since if H1B' is close to H2A', the H5 of the B' ring would not be in the proximity of the 2A' proton.

The F1 slice through the chemical shift of the 1C' proton shows a multiplet centered at 3.7 ppm which has not been assigned. Inspection of the 2D spectrum reveals that this peak is not symmetric, i.e., it is not present along the chemical shift of the resonance corresponding to the other transition of the H1C doublet, and therefore, it is judged to be an artifact.

The contacts derived from the ROESY spectrum of the pentasaccharide are displayed in Figure 3.28.

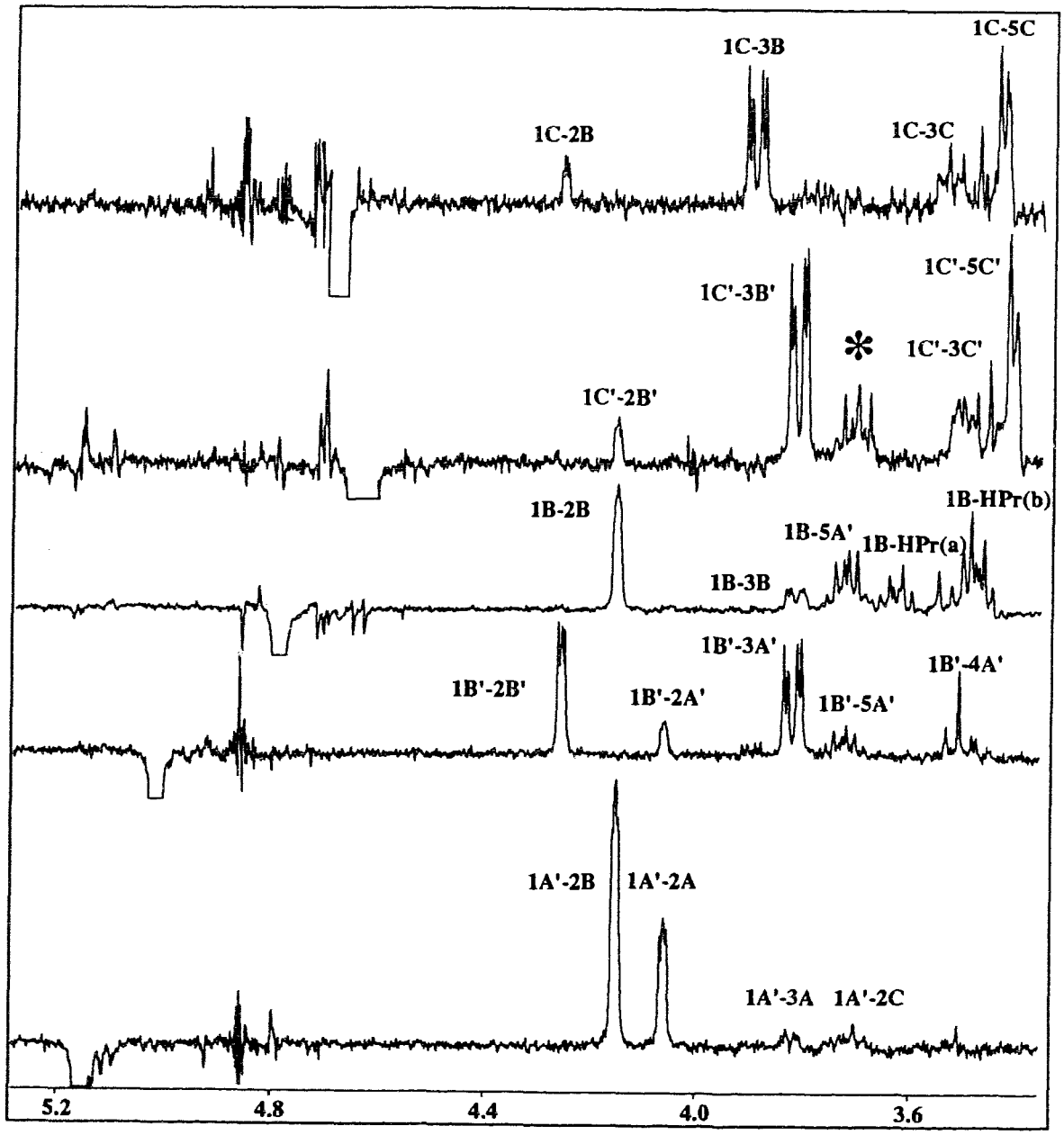


Figure 3.27. F1 slices from the ROESY spectrum of the pentasaccharide (**6**).

Table III.12. Offset and Hartmann Hahn effects calculated for the pentasaccharide (**6**)

Ring	Cross peak	$sin\alpha_1\ sin\alpha_S$	HH transfer	s^2
Ring B	1B-2B	0.9964	a)2B-1B	0.02
			b)2B-3B	0.05
	1B-HPr(a)	0.9896	a)HPr(a)-HPr(b)	0.70
			b)HPr(a)-H3Pr	0.00
	1B-HPr(b)	0.9969	a)HPr(a)-HPr(b)	0.70
			b)HPr(b)-H3Pr	0.00
	1B-5A'	0.9909	a)H5A'-H6A'	0.00
			b)H5A'-H4A'	0.37
Ring A'	1A'-2B	0.9957	a)1B-2B	0.02
			b)2B-3B	0.05
	1A'-2A'	0.9948	a)1A'-2A'	0.02
			b)2A'-3A'	0.09
	1A'-2C	0.9901	a)1C-2C	0.42
			b)2C-3C	0.07
Ring B'	1B'-2B'	0.9972	a)2B'-1B'	0.05
			b)2B'-3B'	0.05
	1B'-3A'	0.9923	a)2A'-3A'	0.09
			b)3A'-4A'	0.23
	1B-4A'	0.9871	a)4A'-5A'	0.37
			b)4A'-3A'	0.23
Ring C	1C-5C	0.9855	a)5C-4C	1.00
			b)5C-6Cs	0.00
			c)5C-6Cr	0.09
	1C-3C	0.9870	a)3C-4C	0.77
			b)3C-2C	0.42
	1C-2B	0.9975	a)2B-1B	0.02
			b)2B-3B	0.05
	1C-3B	0.9920	a)3B-2B	0.05
			b)3B-4B	0.19
Ring C'	1C'-5C'	0.9860	a)5C'-4C'	1.00
			b)5C'-6C's	0.00
			c')5C'-6C'r	0.10
	1C'-3C'	0.9875	a)3C'-4C'	0.77
			b)3C'-2C'	0.40
	1C'-2B	0.9972	a)2B-1B	0.05
			b)2B-3B	0.05
	1C'-3B	0.9932	a)3B-2B	0.05
			b)3B-4B	0.15

The Hartmann Hahn and offset correction factors are tabulated in Table III.12, and the effect of the corrections is displayed in Figure 3.29.

Calculation of Internuclear Distances

The ROE data of the pentasaccharide yielded the average interproton distances displayed in Figure 3.30 and tabulated in Table III.13. The average interproton distance calculated from all the mixing times shows a variation of the order of ± 0.5 Å. During the dynamics simulations the H1A'-H2C distance was assigned a larger range due to the fact that for one series of ROESY spectra this distance was calculated to be a larger value.

Molecular Dynamics Simulations

The results of the dynamics simulations are displayed in Figures 3.31-3.34 and tabulated in Table III.14 and III.15. The trend of increased flexibility during the dynamics simulation without the constraints is observed in the pentasaccharide (**6**). One exception is the A'-B link, in which, during the unconstrained dynamics simulation, the Ψ angle appears to have moved into a new minimum between 0° and -60° compared to the Ψ angle of between 0° and 60° in the contrained dynamics simulation. This is also reflected in a greater variation in the 1A'-2C (2.4-5.6 Å) distance(Figure 3.34, Ring A').

CROSREL Calculation of the ROESY Buildup Curves

The calculated ROESY curves are compared with experimental curves in Figure 3.35. A large discrepancy is observed between the calculated and experimental curves of the B ring. Experimentally, the cross peak to the propyl α proton proximal to the anomeric proton of the B ring (labeled HPr(b)) is observed to be the strongest cross peak. In the calculated curve, the H1B-H2B cross peak is the most intense. Further, the calculation is unsuccessful in reproducing the cross peak to the remaining α proton of the propyl group, HPr(a). The cross peaks from the H1 of ring A' are extremely well

reproduced using the constrained dynamics trajectory. As a result of the variation in the 1A'-2C distance in the unconstrained dynamics the 1A'-2C cross peak is almost non existent. Although the relative magnitudes of the cross peaks are incorrect, the overall trends calculated for the cross peaks from the 1C, 1B' and 1C anomeric protons are the same as those in the experimental spectra when the constrained dynamics trajectory is used as input. The unconstrained dynamics simulation is not as successful, and while the C and C' rings do not deviate from the experimental curves as obviously, there is a complete change in the relative ordering in the calculated curves for the B' ring. Experimentally, the interglycosidic 1B'-3A' cross peak is observed to be the strongest, but when the intensities are calculated with the unconstrained trajectory, it is one of the weaker peaks. The plots of the unconstrained dynamics trajectories of the 1B'-3A' distance (Figure 3.34, Ring B') show that this distance is almost 3 Å. While this is not much greater than the constrained distance of 2.35 Å, the effect on the ROE is a dramatic reduction in the intensity of the 1B'-3A' cross peak, which illustrates the effect of the r^{-6} dependence.

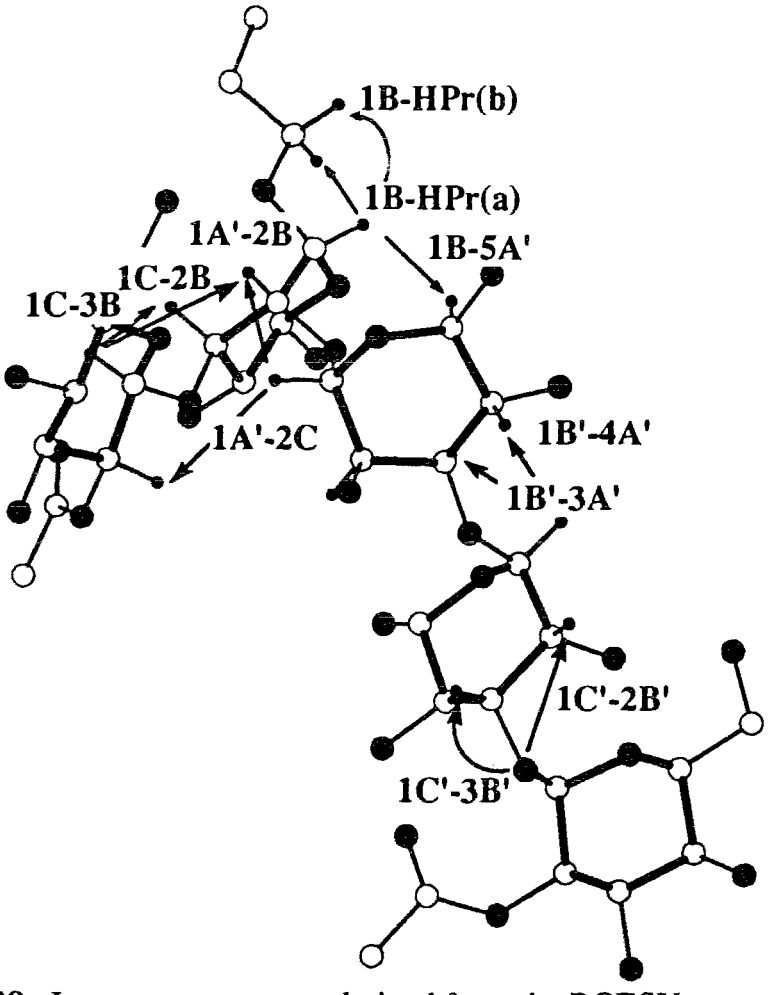


Figure 3.28. Interproton contacts derived from the ROESY spectrum of the pentasaccharide (**6**).

Table III.13. Constraints input for the molecular dynamics simulation of the pentasaccharide (**6**).

Proton Pair	Constrained distance (Å)	Upper limit[a]	Lower limit
H1B-HPr(a)	2.93	3.08	2.78
H1B-HPr(b)	2.51	2.61	2.41
H1B-H5A'	2.67	2.77	2.57
H1A'-H2B	2.28	2.38	2.18
H1A'-H2C	3.21	3.71	2.81
H1B'-H3A'	2.34	2.44	2.24
H1B'-H4A'	2.89	3.04	2.59
H1C-H2B	3.03	3.28	2.73
H1C-H3B	2.27	2.37	2.17
H1C'-H2B'	2.92	3.25	2.48
H1C'-H3B'	2.20	2.30	2.10

[a]The limits were derived from the highest and lowest values obtained from the experiment.

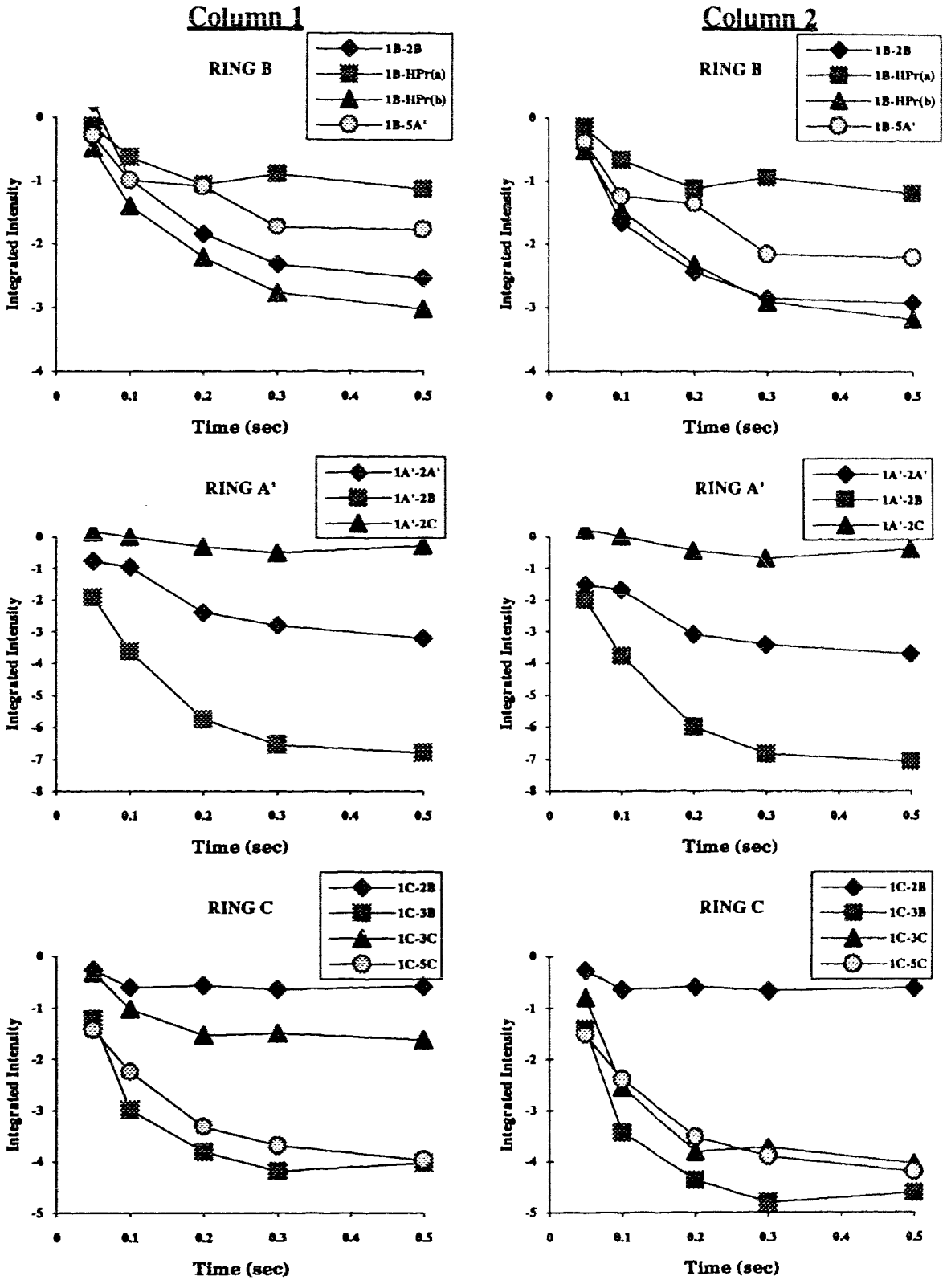


Figure 3.29. The effect of offset and Hartmann Hahn corrections on the ROE buildup of the pentasaccharide (**6**). Column 1) Uncorrected data. Column 2) Corrected data.

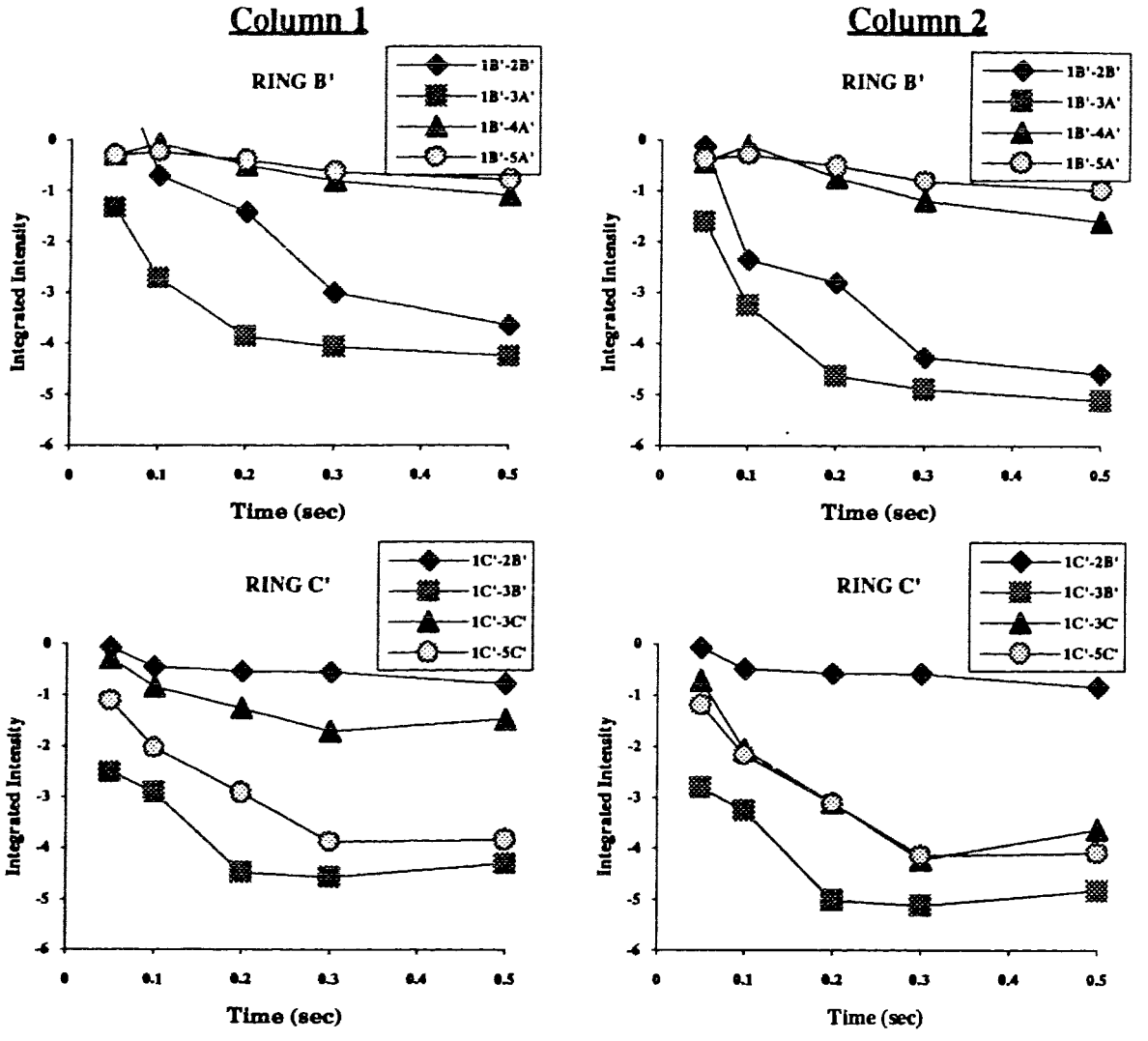


Figure 3.29. (continued from previous page)

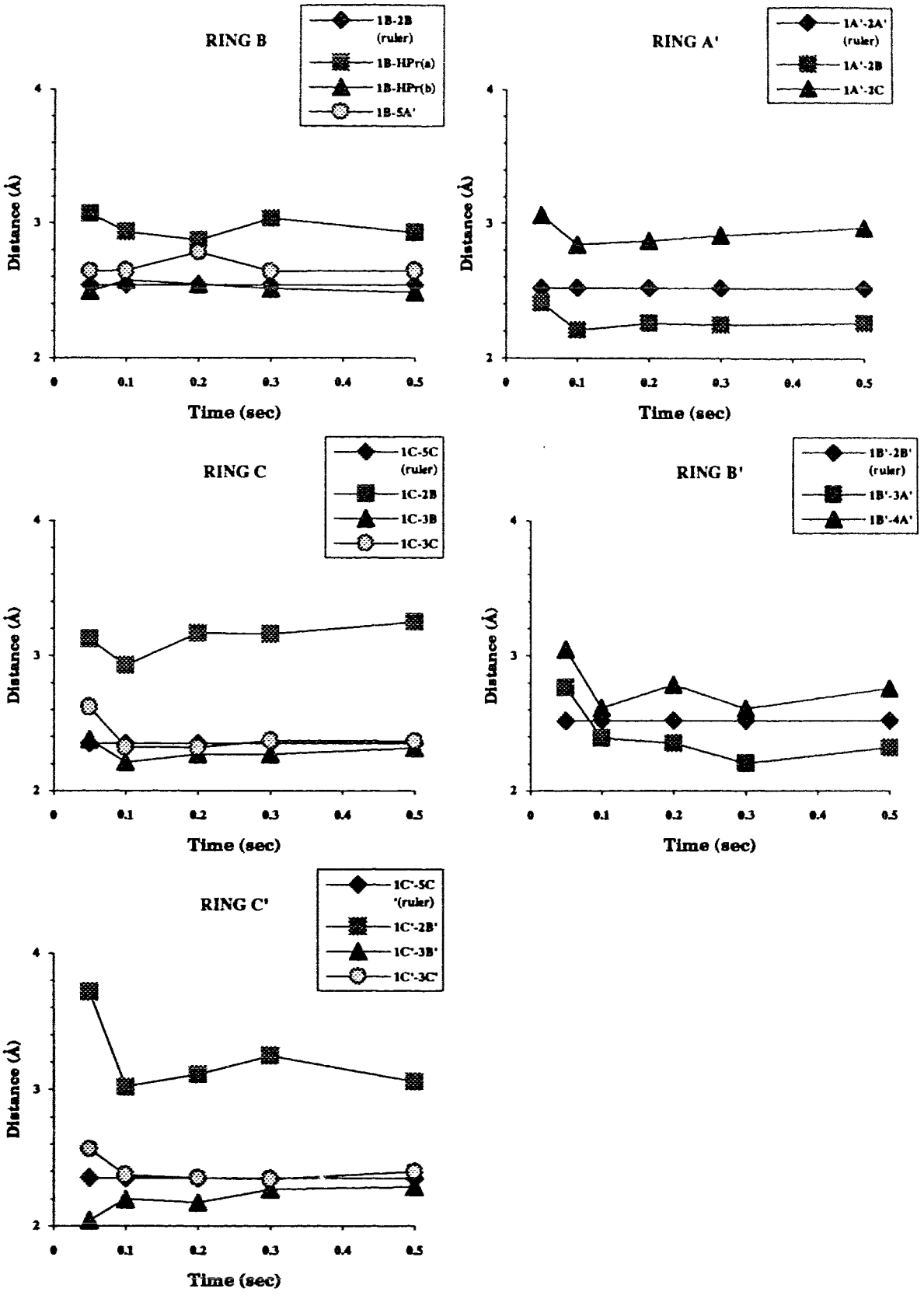
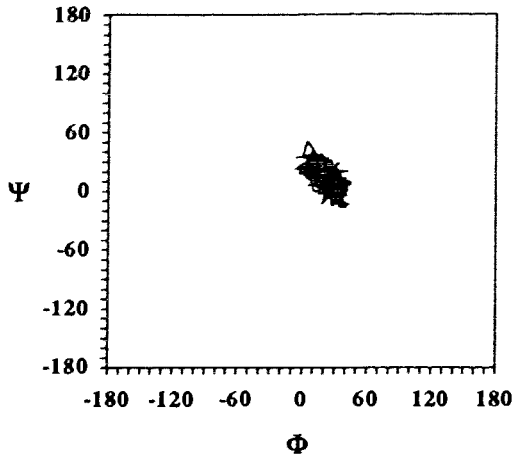
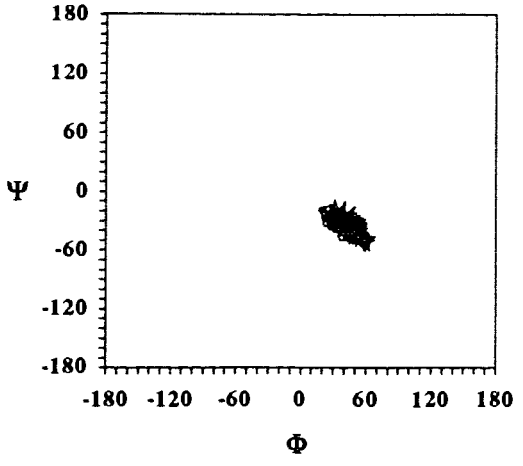


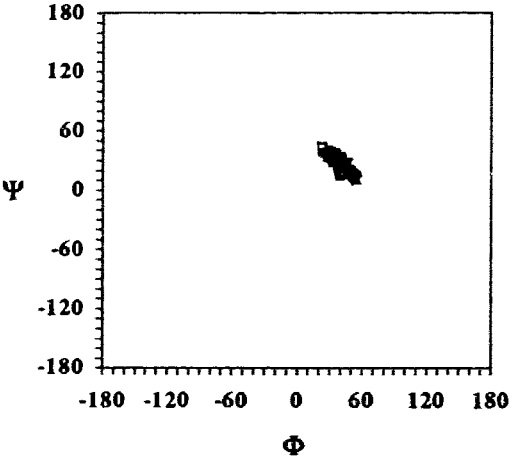
Figure 3.30. Distances calculated from ROESY data of the pentasaccharide (**6**).



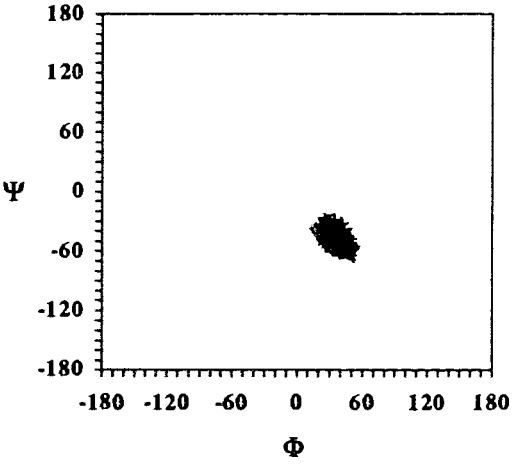
A'-B α-(1→2) linkage



C-B β-(1→3) linkage

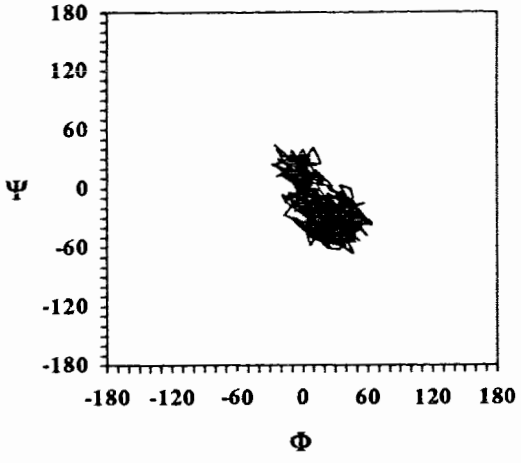


B'-A' α-(1→3) linkage

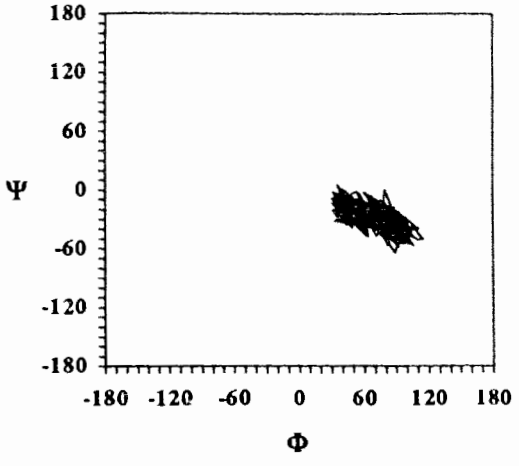


C'-B' β-(1→3) linkage

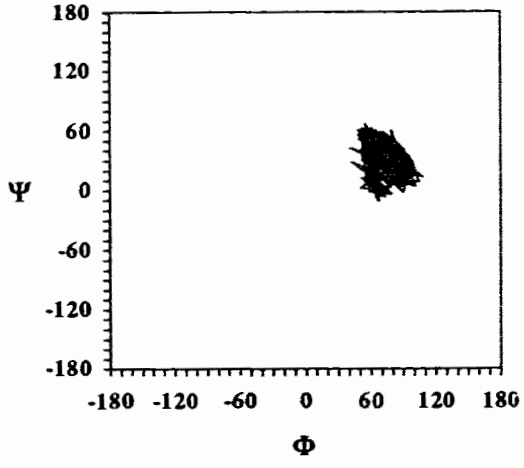
Figure 3.31. Variation of the Φ, Ψ angles during constrained dynamics simulations of the pentasaccharide (**6**) in H_2O.



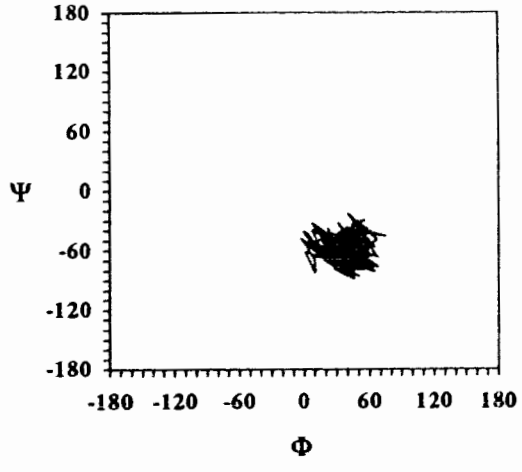
A'-B α-(1→2) linkage



C-B β-(1→3) linkage



B'-A' α-(1→3) linkage



C'-B' β-(1→3) linkage

Figure 3.32. Variation of the Φ, Ψ angles during dynamics simulations of the pentasaccharide (**6**) in H_2O.

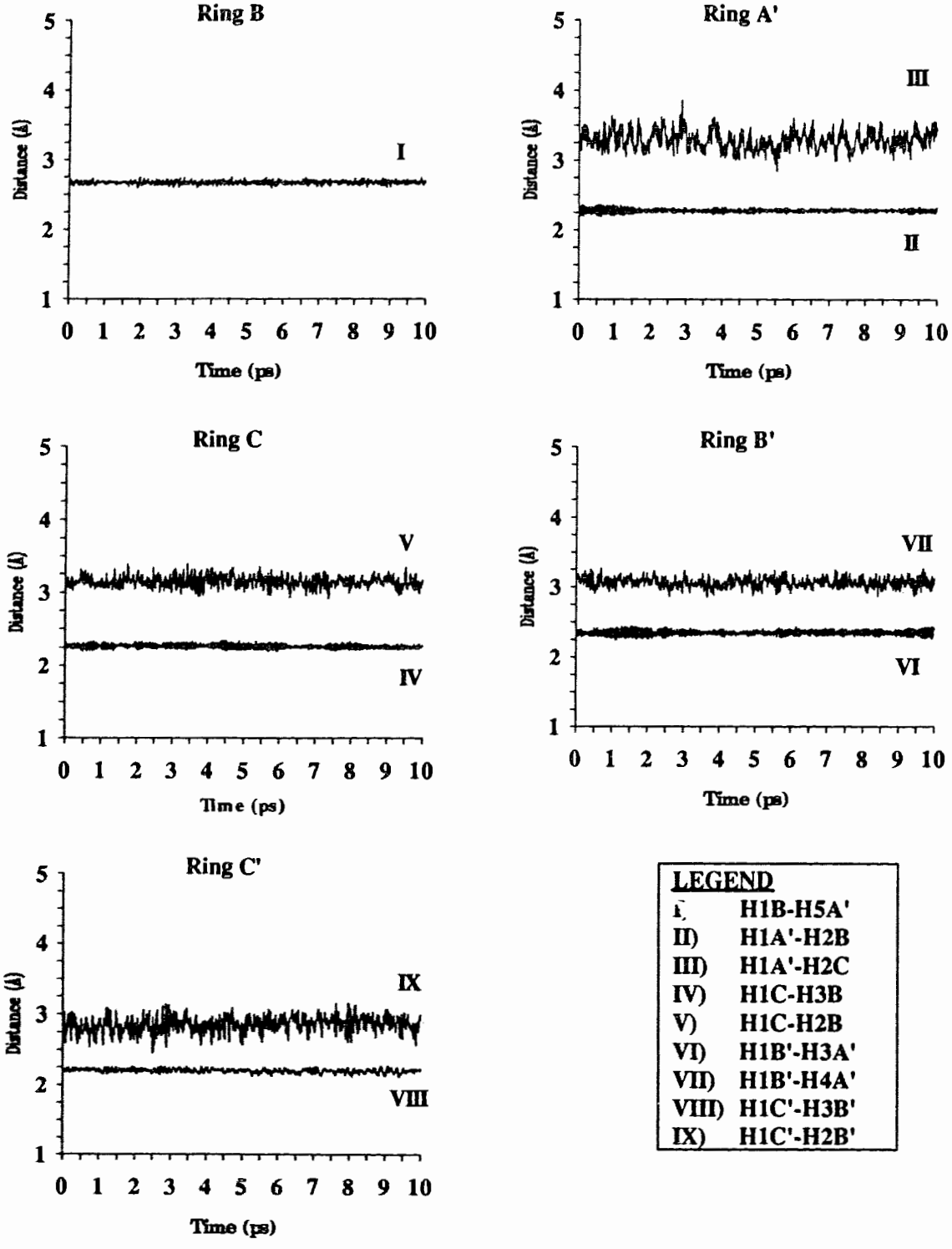


Figure 3.33. Variation of selected internuclear distances during constrained dynamics simulation of the pentasaccharide (**6**) in H_2O.

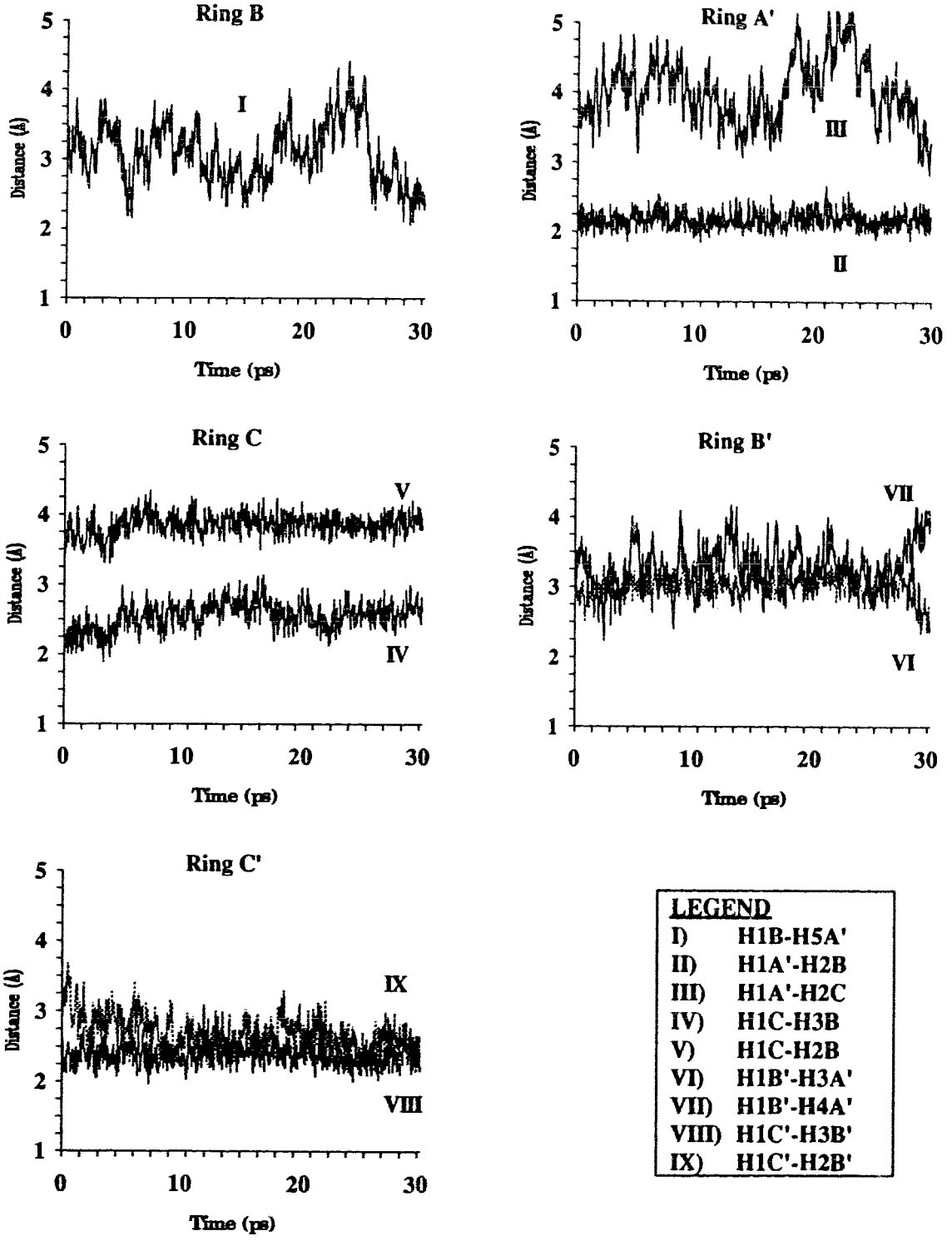


Figure 3.34. Variation of selected internuclear distances during unconstrained dynamics simulation of the pentasaccharide (**6**) in H_2O.

Table III.14. Average, high, and low values of selected internuclear distances during dynamics simulations of the pentasaccharide (**6**)[a].

	Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Proton Pair	Average	High	Low	Average	High	Low
1B-HPr(a)	2.93	3.06	2.77	3.00	3.69	1.98
1B-HPr(b)	2.51	2.59	2.41	2.55	3.77	1.91
1B-H5A'	2.67	2.75	2.58	3.04	4.41	1.96
1A'-H2B	2.28	2.36	2.18	2.18	2.67	1.85
1A'-H2C	3.29	3.86	2.85	3.97	5.61	2.43
1B'-H3A'	2.35	2.44	2.25	2.99	3.50	2.16
1B'-H4A'	3.07	3.26	2.85	3.33	4.27	2.23
1C-H2B	3.15	3.39	2.91	3.86	4.35	3.25
1C-H3B	2.26	2.35	2.19	2.56	3.20	1.86
1C'-H2B'	2.86	3.18	2.42	2.69	3.67	1.97
1C'-H3B'	2.20	2.27	2.11	2.39	2.89	1.95

Table III.15. Average, high, and low values of the glycosidic Φ and Ψ angles during dynamics simulations of the pentasaccharide (**6**)[a].

	Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Angle	Average	High	Low	Average	High	Low
Link A-Propyl						
Φ	57.53	75.77	30.40	59.91	97.32	12.80
Link C-B						
Φ	44.35	68.65	16.44	75.74	117.21	19.66
Ψ	-35.63	-9.01	-60.98	-30.26	14.05	-64.80
Link A'-B						
Φ	22.47	45.41	-6.39	22.57	65.05	-30.22
Ψ	12.68	51.50	-16.95	-26.23	44.75	-69.85
Link B'-A'						
Φ	40.52	60.39	19.32	73.56	108.60	39.16
Ψ	27.71	49.67	5.44	34.15	74.20	-13.38
Link C'-B'						
Φ	36.75	57.28	12.62	36.67	74.78	-11.60
Ψ	-48.64	-21.78	-71.25	-59.40	-12.97	-88.59
ω and θ angles						
Ring C						
θ	5.00	40.34	-55.74	8.96	60.67	-37.21
ω	56.42	100.82	-1.37	38.85	84.92	-31.25
Ring C'						
θ	132.98	179.49	-179.86	14.10	180.00	-179.91
ω	22.49	72.77	-42.66	16.00	74.00	-69.56

[a] Average temperature during the dynamics simulation was 301K for the simulation with constraints and 299K for the simulation without constraints.

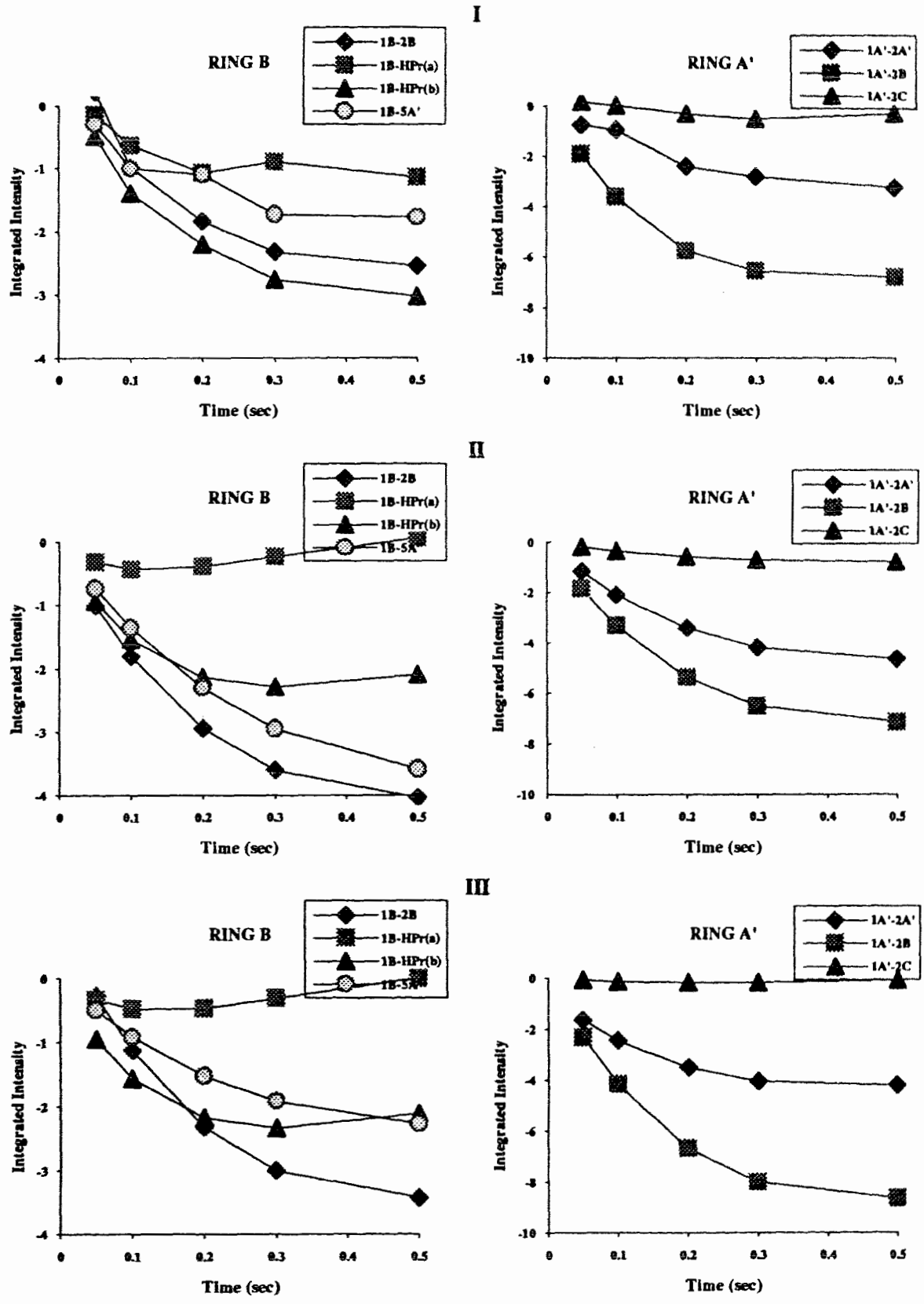


Figure 3.35. Comparison of the experimental and calculated ROESY buildup curves for the pentasaccharide (**6**). I) Experimental buildup curves. II) Buildup curves calculated from the constrained dynamics trajectories in H_2O. III) Buildup curves calculated from the unconstrained dynamics trajectories in H_2O.

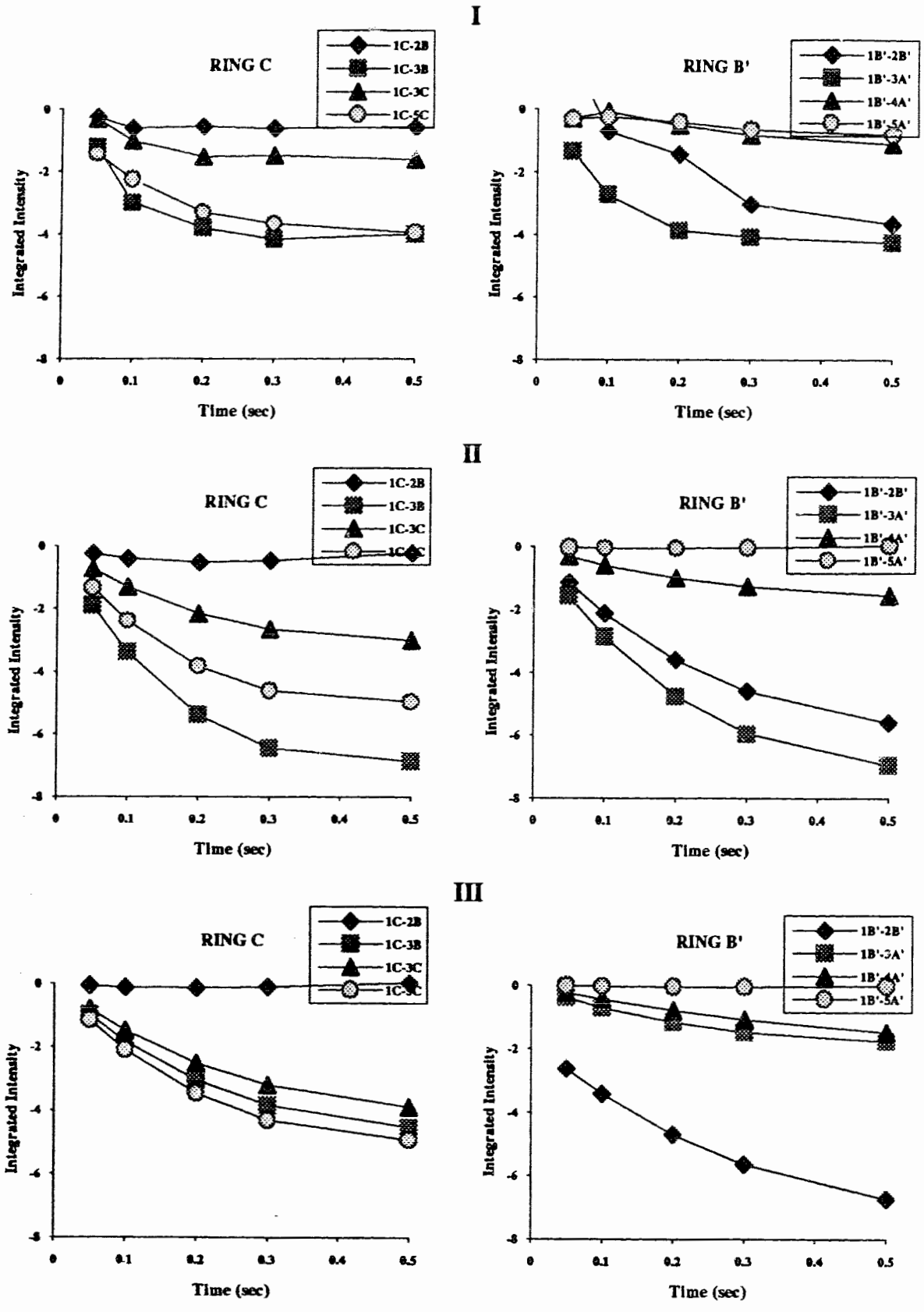


Figure 3.35. (continued from previous page)

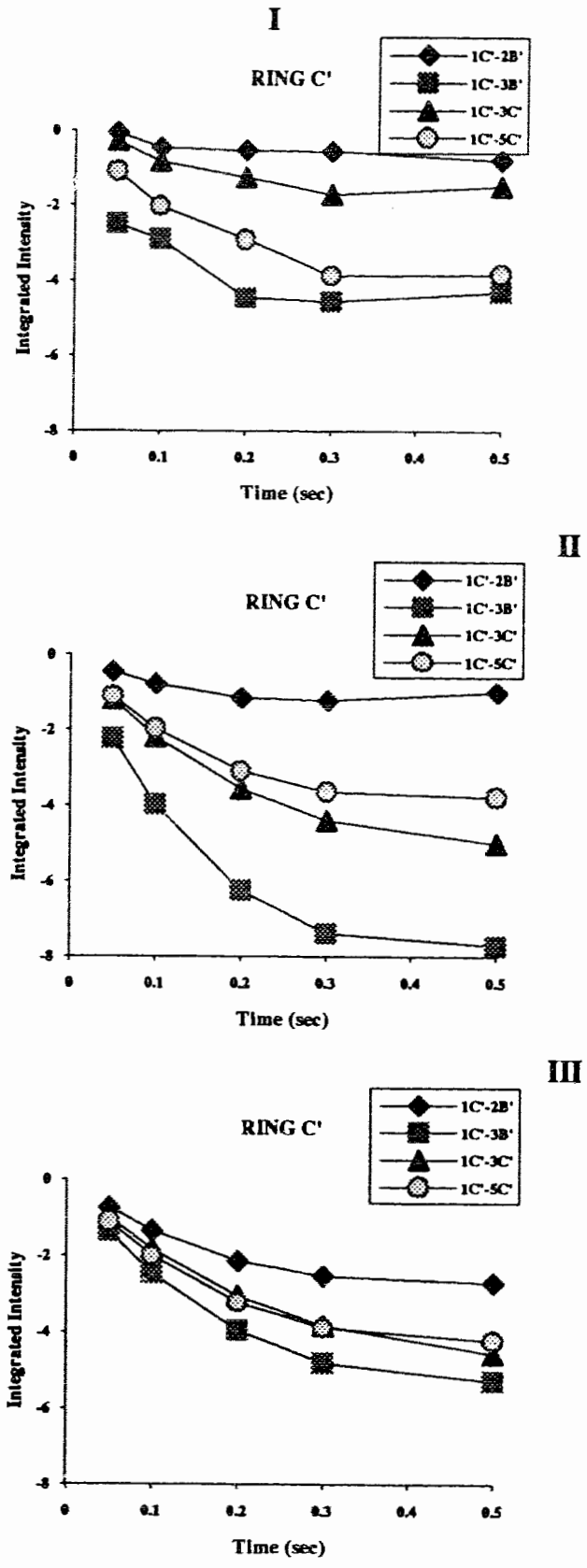


Figure 3.35. (continued from previous page)

3.4.4. Hexasaccharide (7)

Roesy Spectrum

The hexasaccharide (7) possesses components of all of the three smaller structures, the tri, tetra and pentasaccharides (Figure 3.25). The ROESY spectrum of the hexasaccharide (7) is displayed in Figure 3.36 along with the corresponding F1 slices in Figure 3.37. The cross peaks observed in the smaller structures are present in the spectrum of this compound as well. There are no novel linkages - all of the linkages present have been encountered previously in the smaller structures. The spectrum does show greater residual TOCSY cross peaks, as is visible along the chemical shift of the anomeric protons of the two Glc*p*NAc (C and C') rings. The H1B' shows distinct ROE contacts to both the H4 and the H2 of the A' ring. The H1B-H3A-H5A anomaly persists,

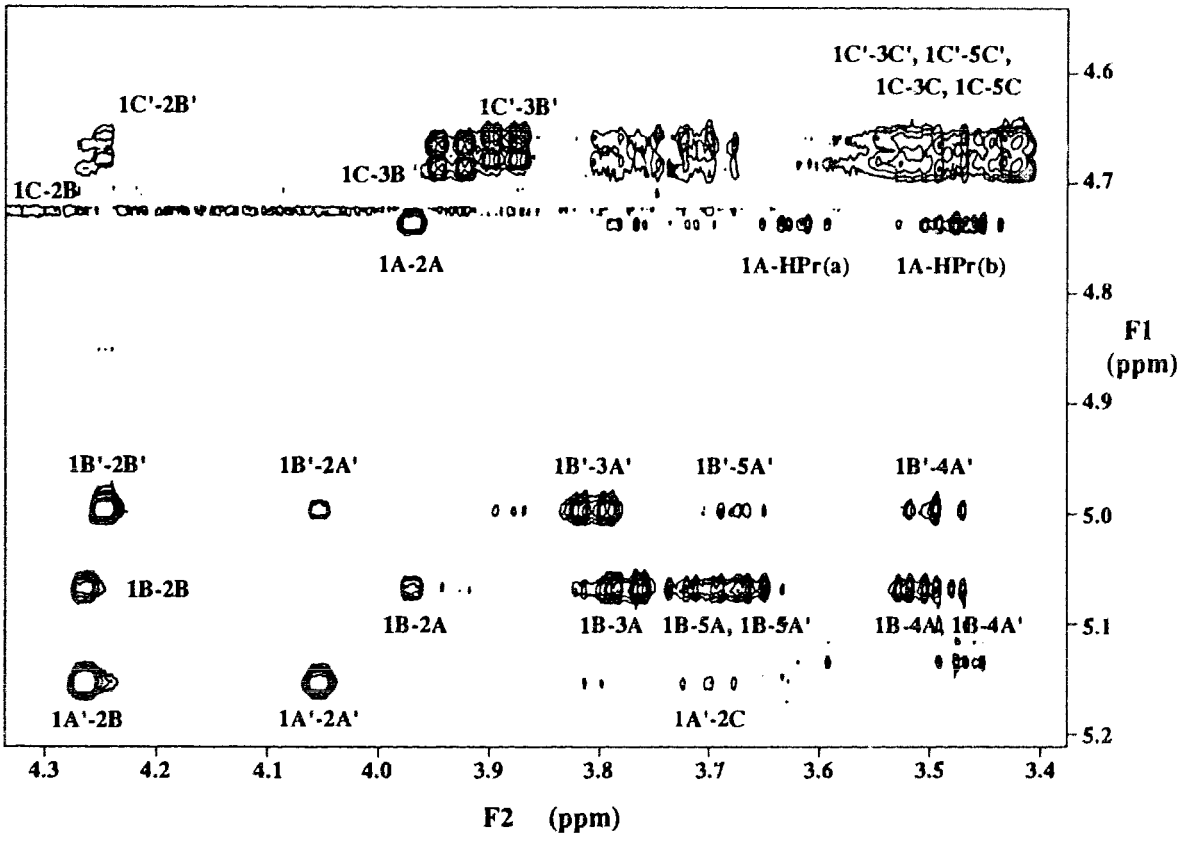


Figure 3.36. Expansion of the ROESY spectrum of the hexasaccharide (7).

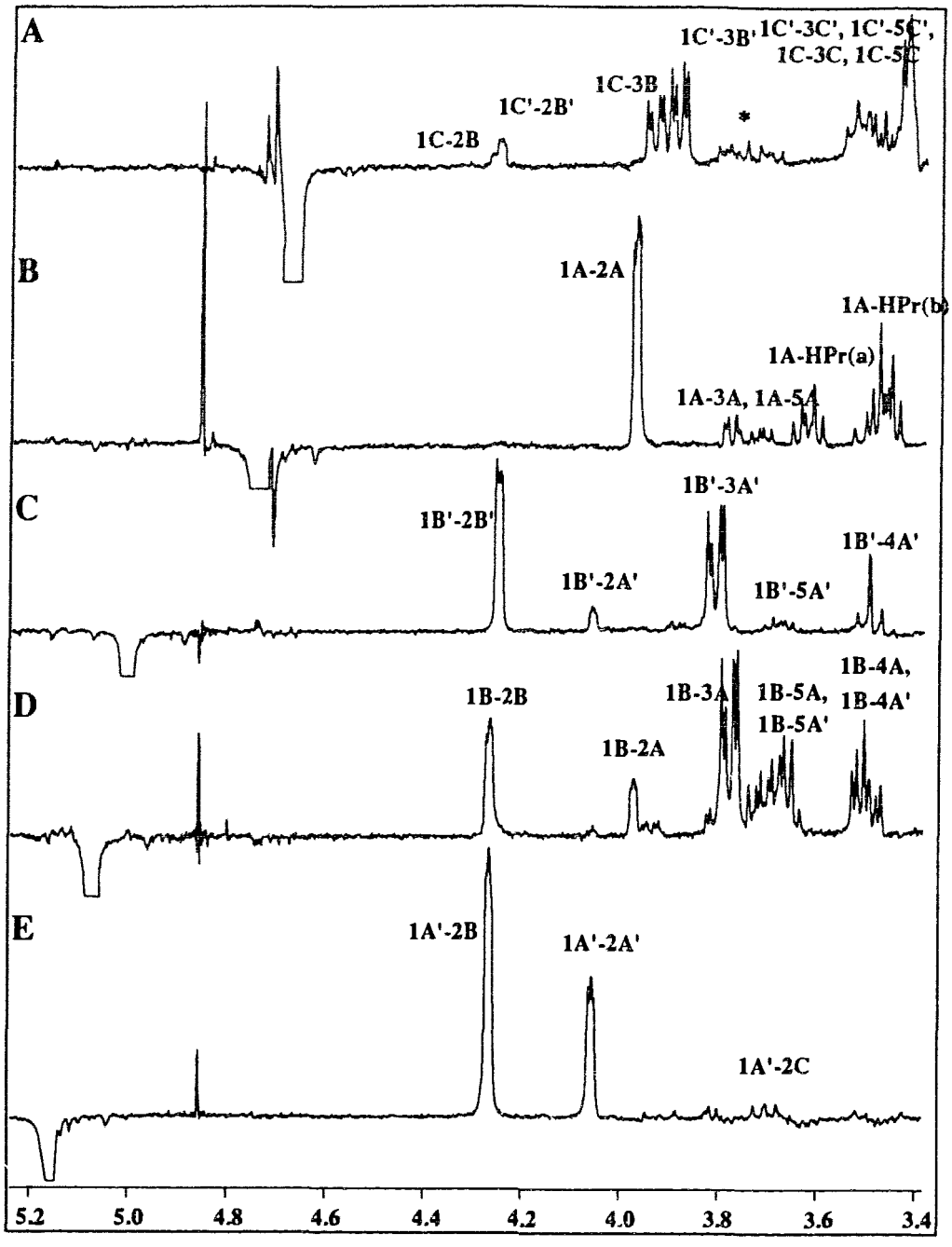


Figure 3.37. F1 slices from the ROESY spectrum of the hexasaccharide (**7**).

and this time it is visible in both the B'-A' and the B-A links.

The 1B proton shows a cross peak to both the 5A and the 5A' protons. Since the chemical shift of both these protons overlap, the 1B-5A and 1B-5A' cross peaks were integrated together. The 1B-5A cross peak probably arises as a result of relay from the 1B-3A peak, as with the cross peaks along the chemical shift of the 1B' proton, where

Offset and Hartmann Hahn Corrections

Table III.16. Offset and Hartmann Hahn effects calculated for the hexasaccharide (7).

Ring	Cross peak	$sin\alpha_I\ sin\alpha_S$	HH transfer	s^2
Ring A	1A-2A	0.9964	a)2A-1A	0.01
			b)2A-3A	0.11
	1A-HPr(a)	0.9888	a)HPr(a)-HPr(b)	0.64
			b)HPr(a)-H3Pr	0.00
	1A-HPr(b)	0.9961	a)HPr(a)-HPr(b)	0.64
			b)HPr(b)-H3Pr	0.00
Ring B	1B-2B	0.9970	a)2B-1B	0.03
			b)2B-3B	0.07
	1B-2A	0.9939	a)2A-1A	0.01
			b)2A-3A	0.11
	1B-3A	0.9939	a)2A-3A	0.11
			b)3A-4A	0.28
	1B-5A'	0.9895	a)5A'-6A'	0.00
			b)5A'-4A'	0.48
Ring C	1C-5C	0.9849	a)5C-4C	1.00
			b)5C-6Cs	0.00
			c)5C-6CR	0.07
	1C-3C	0.9867	a)3C-4C	0.67
			b)3C-2C	0.41
	1C-2B	0.9971	a)2B-1B	0.03
			b)2B-3B	0.07
	1C-3B	0.9936	a)3B-2B	0.07
			b)3B-4B	0.13
Ring A'	1A'-2B	0.9968	a)1B-2B	0.03
			b)2B-3B	0.07
	1A'-2A'	0.9946	a)1A'-2A'	0.02
			b)2A'-3A'	0.08
	1A'-2C	0.9897	a)1C-2C	0.41
			b)2C-3C	0.07
Ring B'	1B'-2B'	0.9970	a)2B'-1B'	0.03
			b)2B'-3B'	0.06
	1B'-3A'	0.9919	a)2A'-3A'	0.08
			b)3A'-4A'	0.22
	1B'-4A'	0.9866	a)4A'-5A'	0.48
			b)4A'-3A'	0.22
Ring C'	1C'-5C'	0.9849	a)5C'-4C'	1.00
			b)5C'-6C's	0.00
			c')5C'-6C'R	0.07
	1C'-3C'	0.9871	a)3C'-4C'	0.63
			b)3C'-2C'	0.42
	1C'-2B	0.9969	a)2B'-1B'	0.03
			b)2B'-3B'	0.06
	1C'-3B	0.9929	a)3B'-2B'	0.06
			b)3B'-4B'	0.15

the α-(1→3) linkage gives rise to both a 1B'-3A' and a 1B'-5A' peak. The contacts derived from the ROESY spectrum are displayed in Figure 3.8.

In order to determine the magnitude of the 1B-5A' peak, which is critical in determining the conformation about the α-(1→2) linkage, it was noted that in the case of the 1B' interactions, the 1B'-5A' peak was ~30% of the 1B'-3A' cross peak. Therefore 30% of the 1B-3A peak was subtracted from the integral of the 1B-(5A + 5A') peak and the 1B-5A' distance was calculated. This procedure is an extreme approximation, but represented the best solution since neither deconvolution of the peaks nor integration of the 1B-4A peak was possible. The cross peaks were then corrected for Hartmann Hahn and offset effects, and the results are tabulated in Table III.16 and displayed in Figure 3.38.

The problem of overlap is considerable in this structure. The chemical shifts of the ring protons of both the Glc*p*NAc rings are completely degenerate. Further, the ring protons of the corresponding residues to which cross peaks occur, the B and B' rings also have very similar chemical shifts. This makes it impossible to integrate the two sets of cross peaks belonging to the C and C' rings independently. Judging from the intensities of the cross peaks, it seems that the two rings display approximately the same behavior, and hence the same distances were assigned to the C and C' rings.

Calculation of Internuclear Distances

The distances calculated for the hexasaccharide (**7**) from the ROESY spectra and used as input for the constrained dynamics calculations are listed in Table III.17 and displayed in Figure 3.40 to illustrate the variance with respect to mixing time.

Molecular Dynamics Simulations

The **Φ, Ψ** maps and trace of selected internuclear distances during the constrained and unconstrained dynamics simulations are shown in Figures 3.41-3.44.

The various linkages display more freedom during the unconstrained dynamics simulation, but are restricted to roughly the same areas. The exception is the B'-A' linkage. The constraints place both the Φ and Ψ in the range of 0° to 60°, but during the unconstrained dynamics simulation the Ψ angle moves to the 0° to -60° range. This change in the Ψ angle is accompanied by an increase in the average distance between the 1B' and 4A' protons from 3.11 Å to 4.32 Å. The results of the dynamics simulations are tabulated in Tables III.18 and III.19.

CROSREL calculation of the ROESY buildup curves

The experimental and CROSREL calculated ROESY buildups curves are presented in the series of graphs in Figure 3.45. In most cases, the fit is satisfactory for the cross peaks from the anomeric protons. The calculated cross peaks from the A' and B' rings show a divergence from the experimental curves. In ring A', the relative ordering of the 1A'-2A' and 1A'-2B' cross peaks are reversed for the ROEs calculated from the unconstrained dynamics trajectory. The 1B'-3B' and 1B'-2B' cross peaks calculated from the constrained dynamics trajectory show approximately the same intensity. This is not observed in the experimental build up curves.

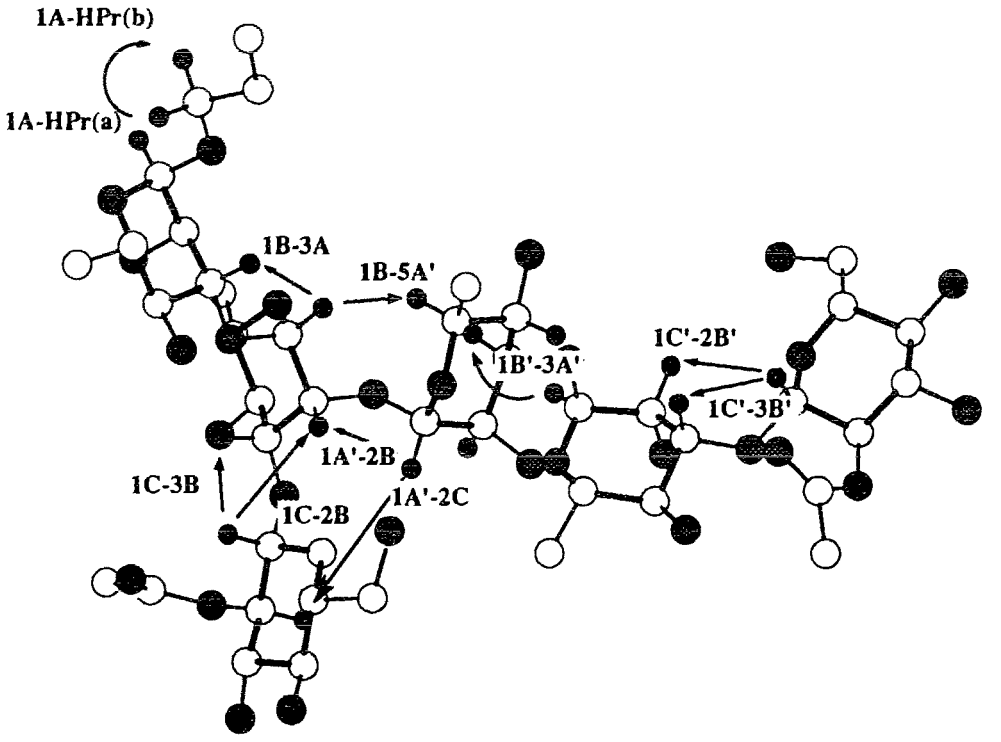


Figure 3.38. Interproton contacts derived from the ROESY spectrum of the hexasaccharide (7).

Table III.17. Constraints input for the molecular dynamics simulation of the hexasaccharide (7).

Proton Pair	Constrained distance [Å]	Upper limit	Lower limit
H1A-HPr(a)	3.04	3.48	2.94
H1A-HPr(b)	2.69	3.09	2.59
H1B-H2A	3.02	3.32	2.82
H1B-H3A	2.30	2.45	2.15
H1B-H5A'	2.44	2.64	2.54
H1A'-H2B	2.31	2.51	2.21
H1A'-H3B	3.70	4.20	3.20
H1A'-H2C	3.50	3.70	3.40
H1B'-H2A'	3.03	3.23	2.93
H1B'-H3A'	2.39	2.49	2.29
H1B'-H4A'	2.86	3.36	2.36
H1C-H2B	3.03	3.33	2.83
H1C-H3B	2.19	2.04	2.34
H1C'-H2B'	3.03	3.33	2.83
H1C'-H3B'	2.19	2.04	2.34

[a]The limits were derived from the highest and lowest values obtained from the experiment.

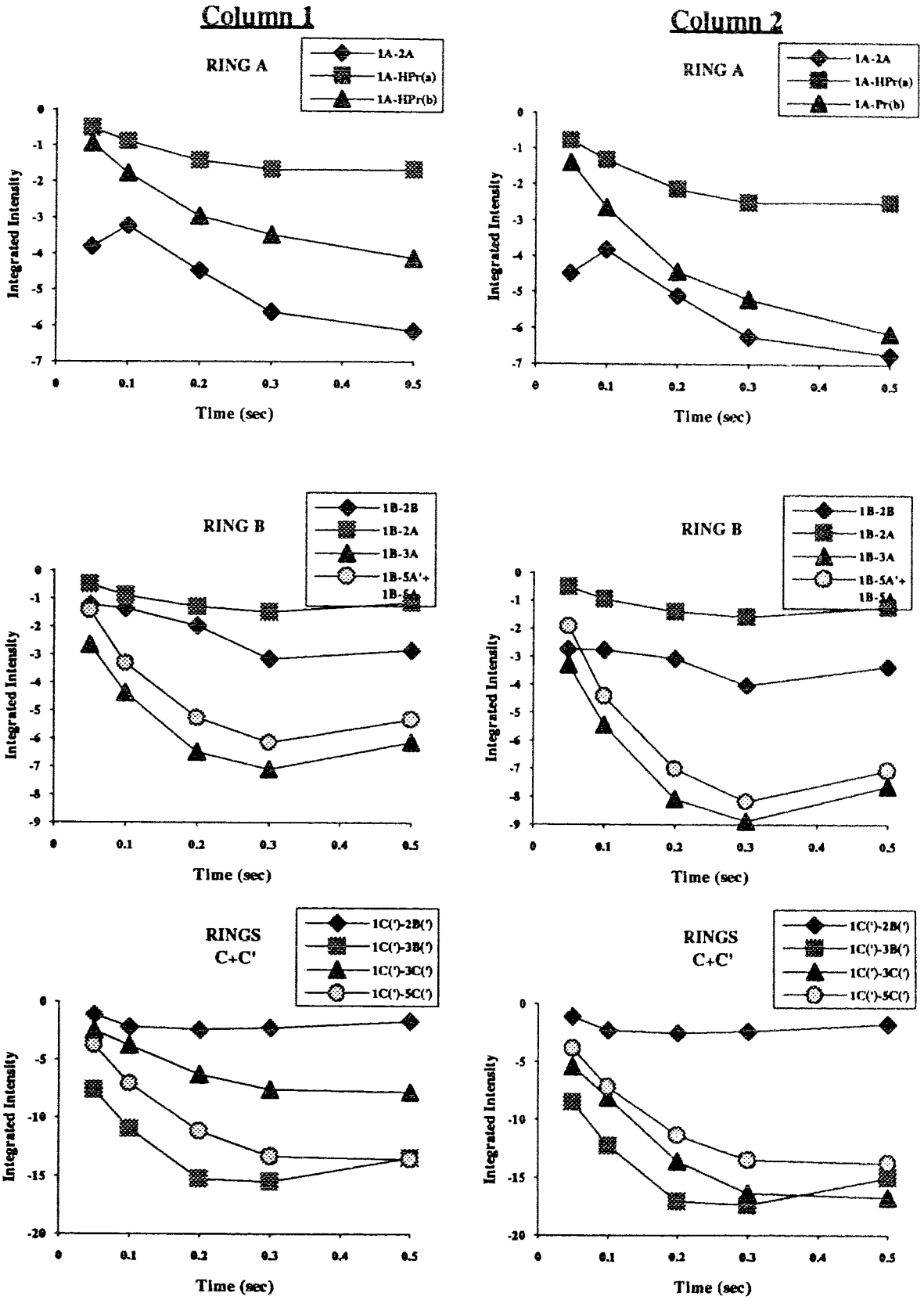


Figure 3.39. The effect of offset and Hartmann Hahn corrections on the ROE buildup of the hexasaccharide (**7**). Column 1.) Uncorrected data. Column 2.) Corrected data.

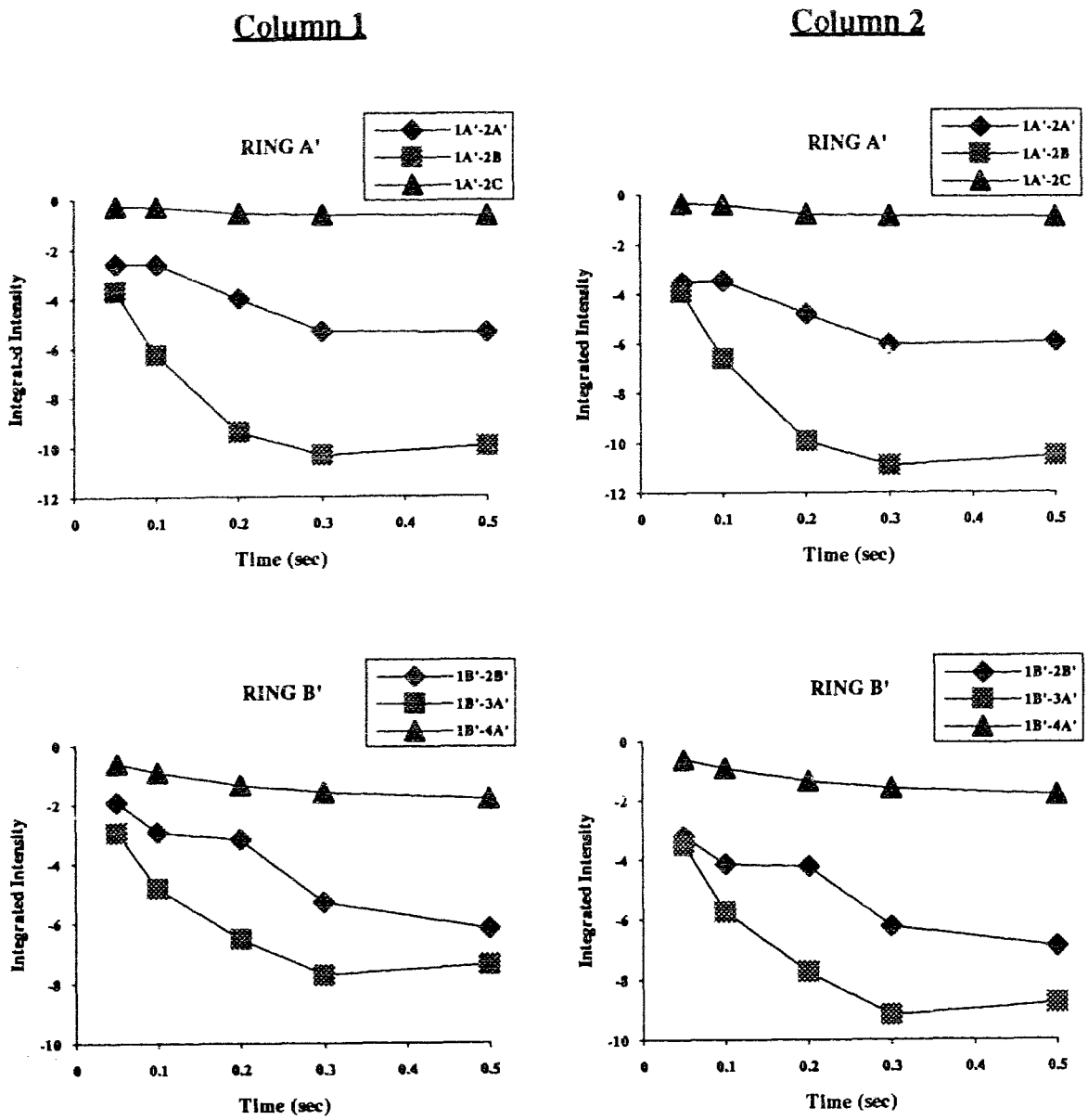


Figure 3.39. (continued from previous page).

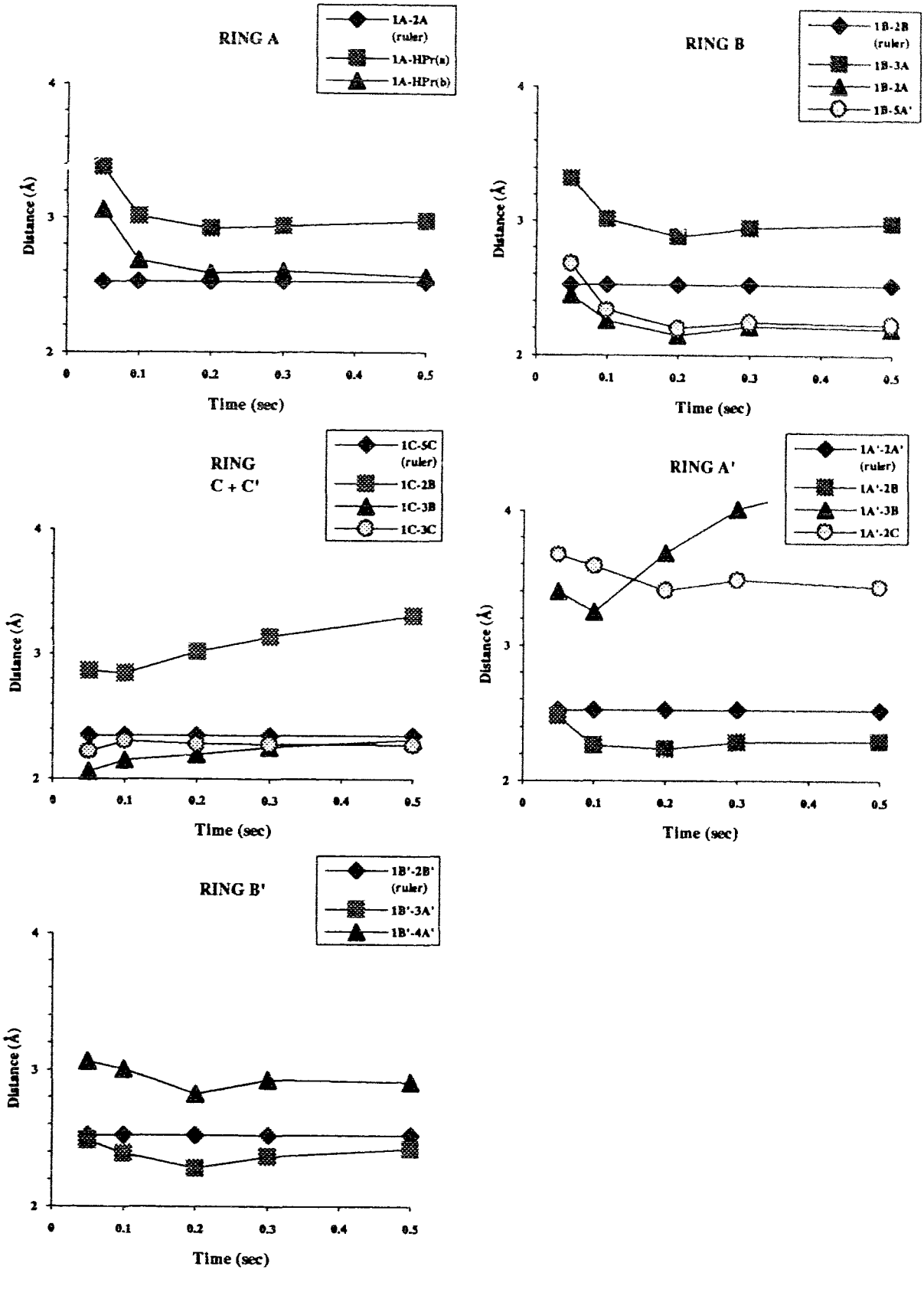
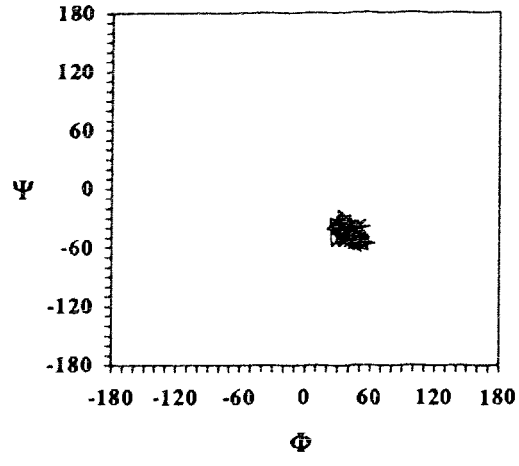
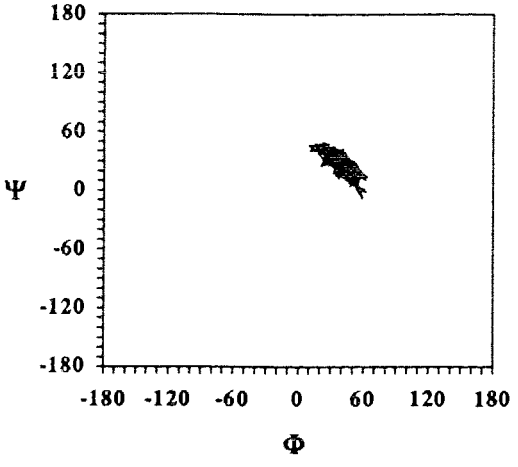


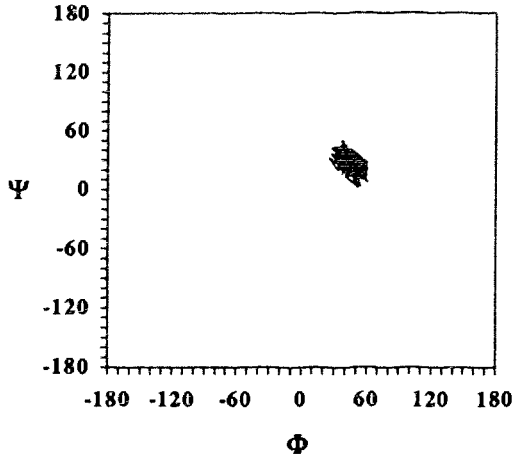
Figure 3.40. Distances calculated from ROESY data of the hexasaccharide (**7**).



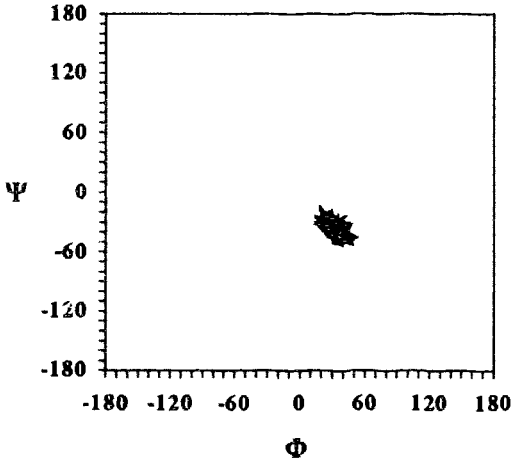
B-A α-(1→3) linkage



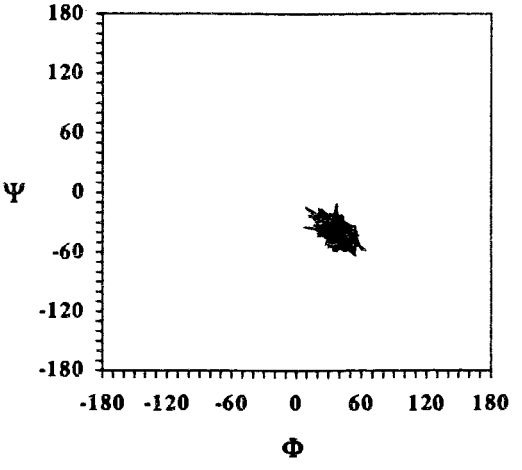
B'-A' α-(1→3) linkage



A'-B α-(1→2) linkage

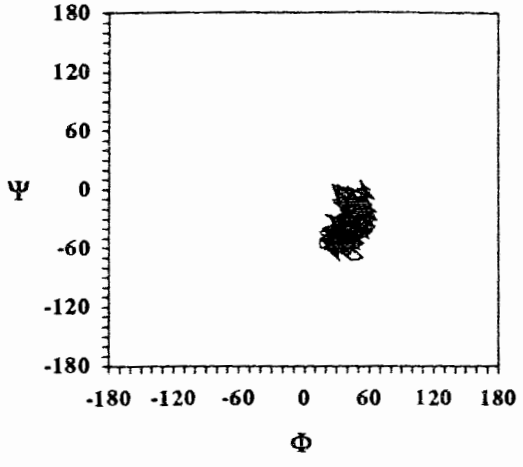


C-B β-(1→3) linkage

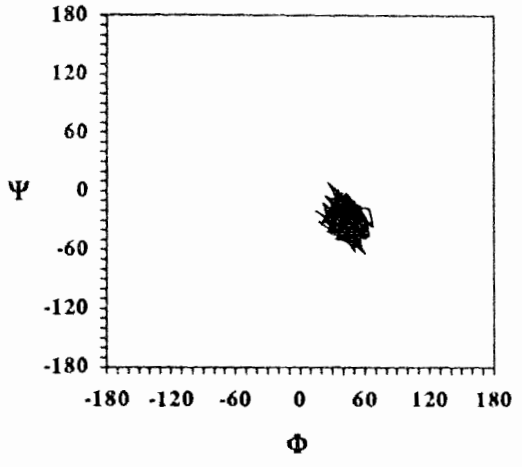


C'-B' β-(1→3) linkage

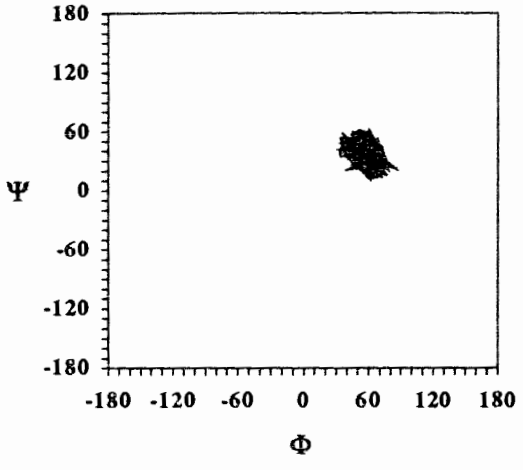
Figure 3.41. Variation of the Φ, Ψ angles during constrained dynamics simulations of the hexasaccharide (7) in H_2O.



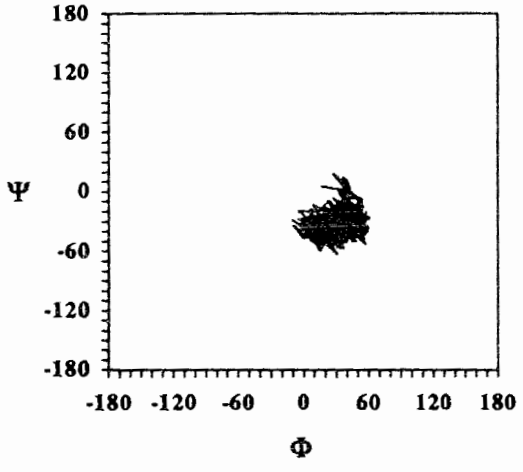
B-A α-(1→3) linkage



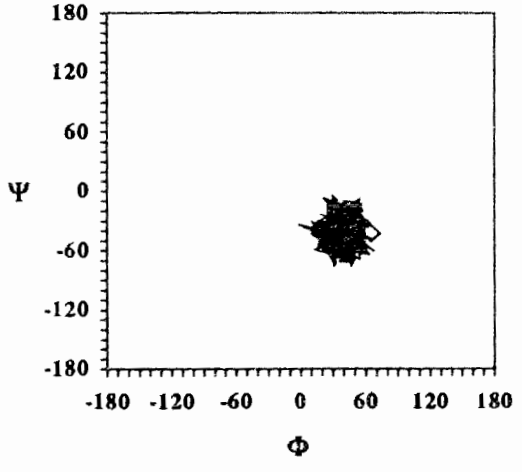
B'-A' α-(1→3) linkage



A'-B α-(1→2) linkage



C-B β-(1→3) linkage



C'-B' β-(1→3) linkage

Figure 3.42. Variation of the Φ, Ψ angles during unconstrained dynamics simulations of the hexasaccharide (7) in H_2O.

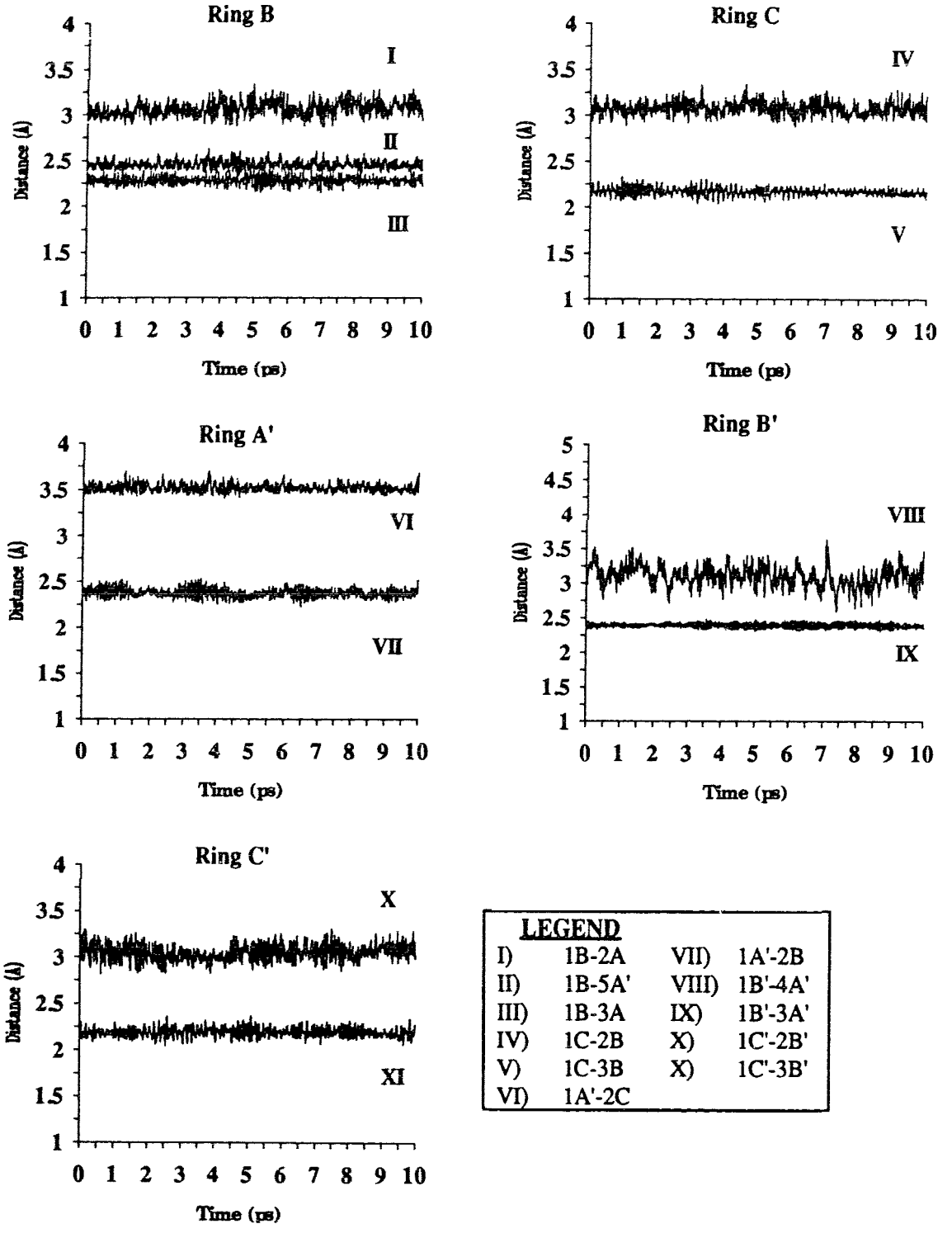


Figure 3.43. Variation of selected interproton distances during constrained dynamics simulations of the hexasaccharide (**7**) in H_2O.

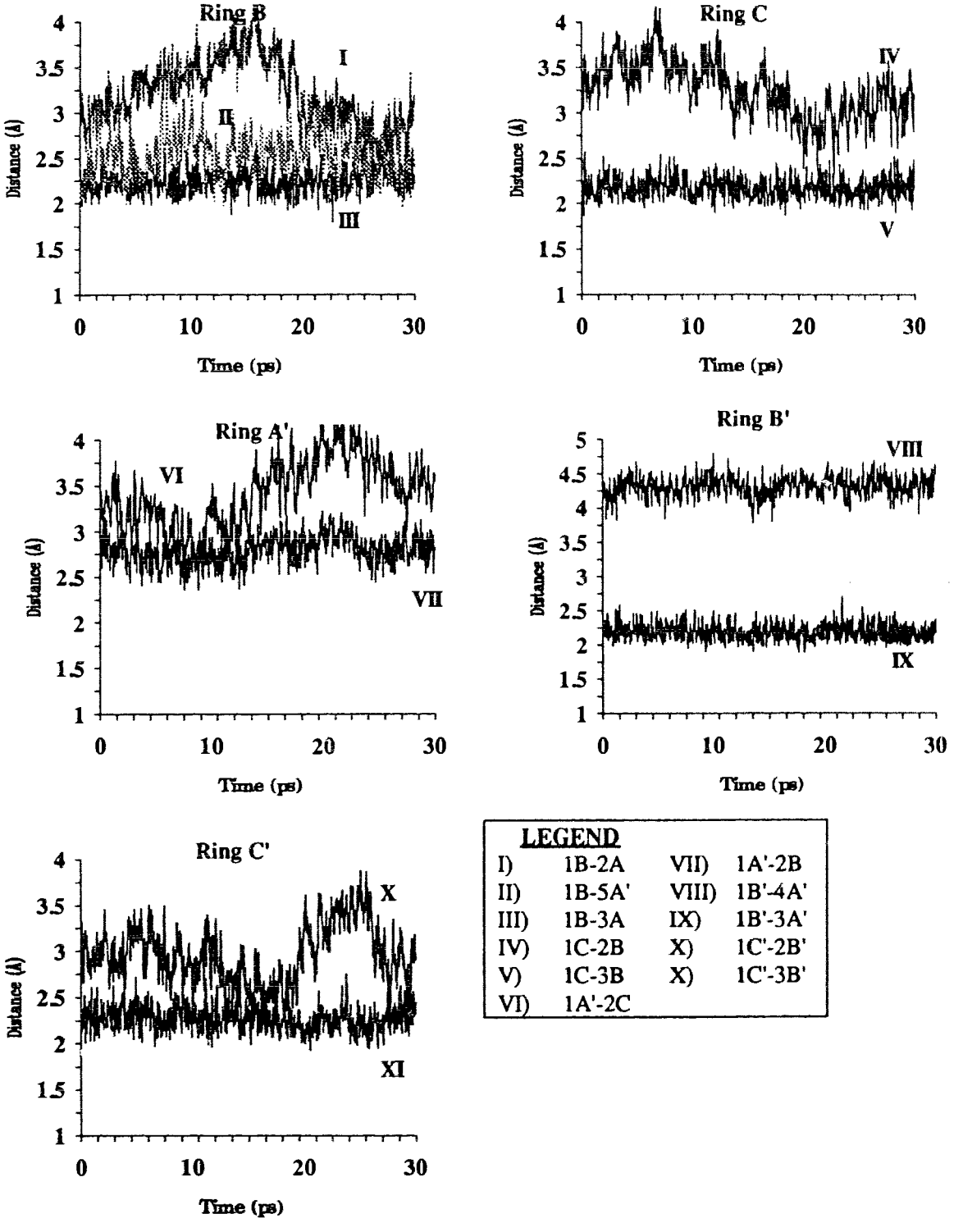


Figure 3.44. Variation of selected interproton distances during unconstrained dynamics simulations of the hexasaccharide (7) in H_2O.

Table III.18. Average, high, and low values of selected internuclear distances during dynamics simulations of the hexasaccharide (7) in H_2O[a].

	Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Proton Pair	Average	High	Low	Average	High	Low
1A-HPr(a)	3.09	3.43	2.96	3.13	4.00	2.24
1A-HPr(b)	2.72	3.01	2.59	2.56	3.49	1.93
1B-H2A	3.06	3.33	2.86	3.21	4.32	2.18
1B-H3A	2.28	2.43	2.14	2.25	2.77	1.79
1B-H5A'	2.46	2.63	2.30	2.53	3.35	1.86
1C-H2B	3.08	3.34	2.89	3.25	4.26	2.33
1C-H3B	2.18	2.33	2.04	2.17	2.57	1.85
1A'-H2B	2.38	2.53	2.23	2.79	3.32	2.28
1A'-H2C	3.52	3.70	3.41	3.42	4.35	2.29
1B'-H3A'	2.39	2.48	2.30	2.19	2.70	1.87
1B'-H4A'	3.11	3.62	2.59	4.32	4.84	3.64
1C'-H2B'	3.05	3.31	2.82	2.97	3.94	2.14
1C'-H3B'	2.19	2.37	2.05	2.27	2.77	1.91

[a]Average temperature during the dynamics simulation was 304 K for the simulation with constraints and 299 K for the simulation without constraints.

Table III.19. Average, high, and low values of the glycosidic Φ and Ψ angles during dynamics simulation of the hexasaccharide (7) in H_2O[a].

	Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Angle	Average	High	Low	Average	High	Low
Link A-Propyl						
Φ	66.23	93.11	40.80	59.86	93.47	22.81
Link B-A						
Φ	41.56	64.43	20.04	42.46	72.27	1.16
Ψ	-45.55	-21.09	-64.64	-37.26	11.82	-73.33
Link C-B						
Φ	32.91	52.87	13.00	31.63	65.47	-13.96
Ψ	-37.32	-13.24	-53.71	-32.46	20.74	-62.93
Link A'-B						
Φ	45.67	60.58	25.28	56.56	87.27	24.61
Ψ	26.59	50.33	1.75	36.48	70.06	7.73
Link B'-A'						
Φ	40.39	63.18	9.61	44.51	72.77	9.00
Ψ	27.63	50.39	-7.64	-30.06	9.33	-64.03
Link C'-B'						
Φ	37.80	62.98	6.96	38.83	74.83	-1.74
Ψ	-40.11	-10.95	-63.16	-44.27	7.50	-77.42
ω and θ angles						
Ring C						
θ	3.63	62.33	-46.02	10.09	65.89	-33.85
ω	28.55	66.81	-39.45	24.10	78.36	-34.88
Ring C'						
θ	6.87	58.14	-35.27	-2.41	47.35	-79.01
ω	31.21	64.22	-10.51	8.61	74.23	-63.67

[a]Average temperature during the dynamics simulation was 304 K for the simulation with constraints and 299 K for the simulation without constraints.

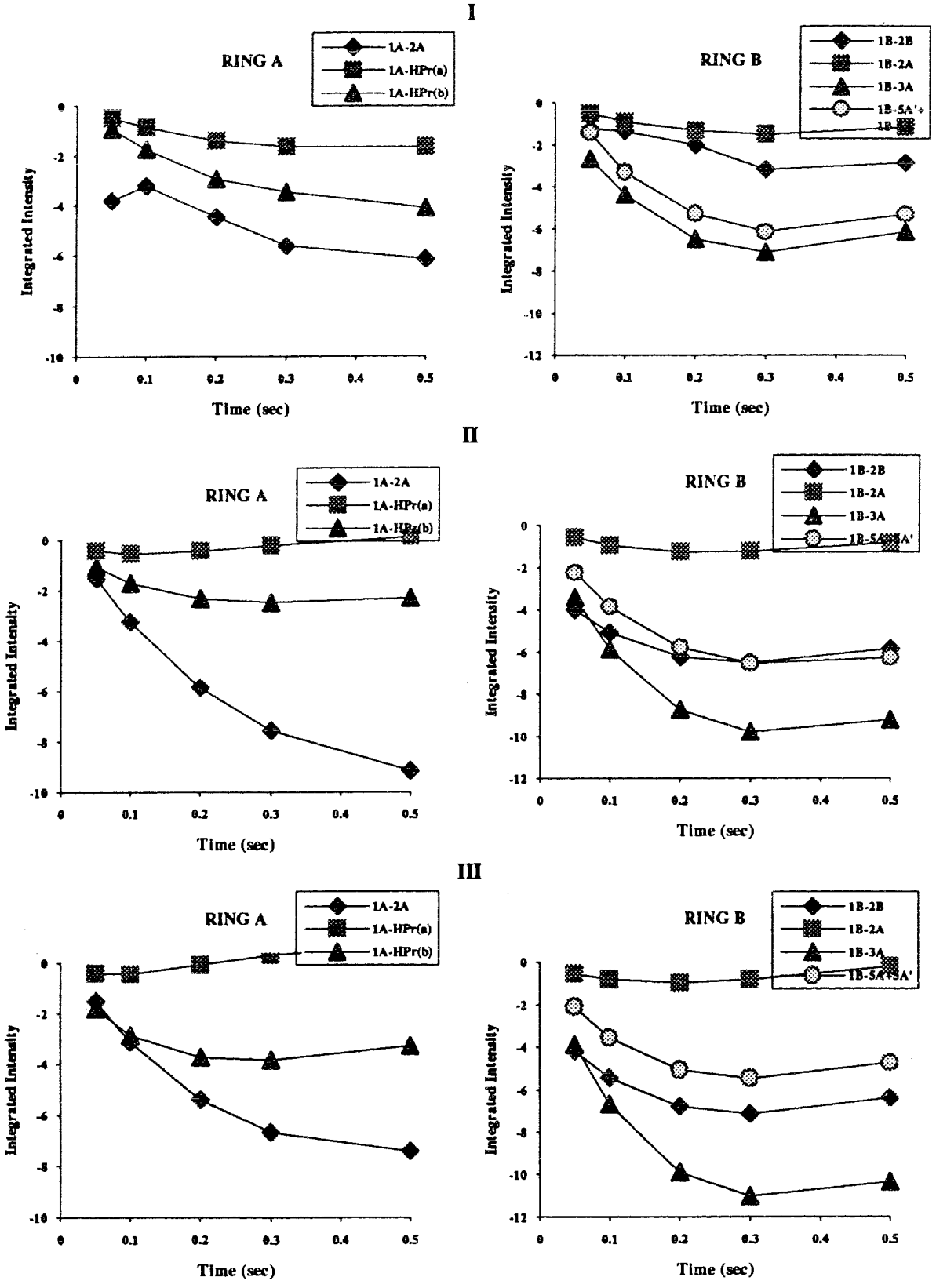


Figure 3.45. (Caption two pages overleaf).

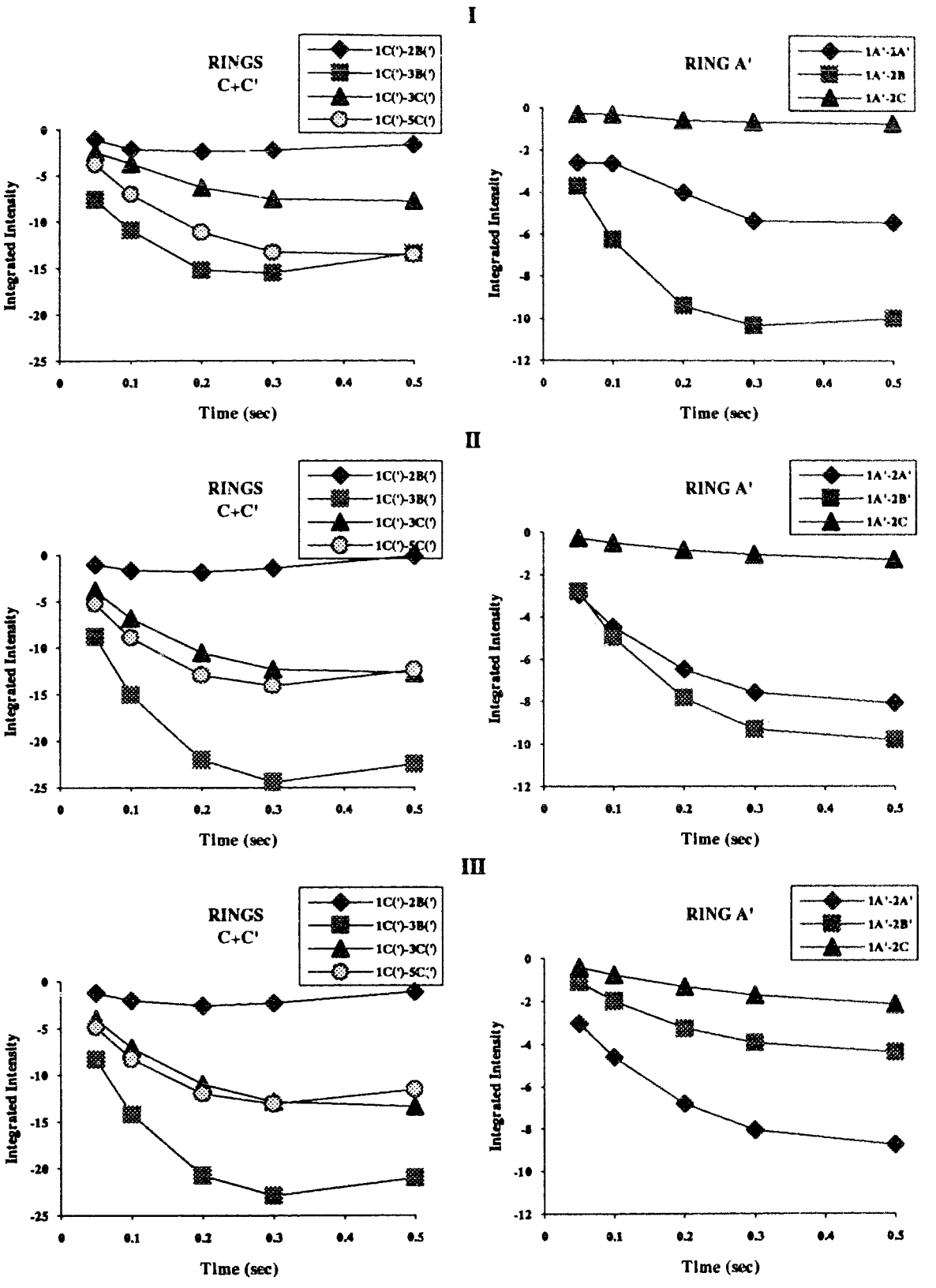


Figure 3.45. (Caption overleaf).

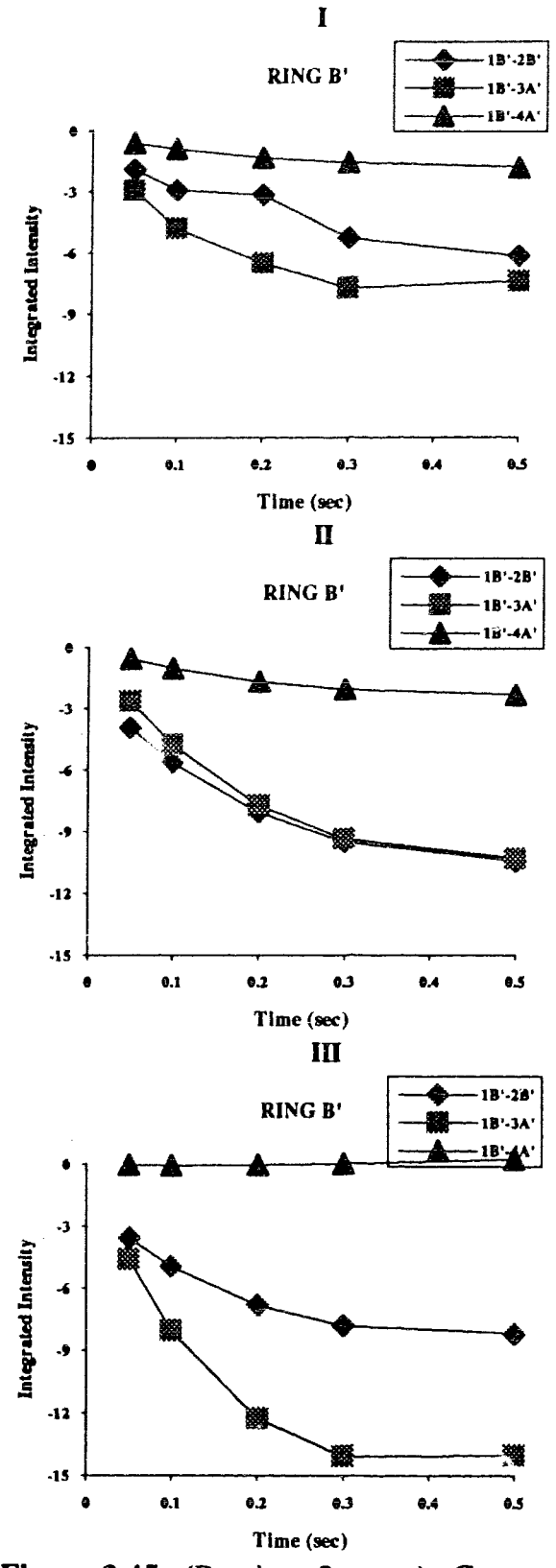


Figure 3.45. (Previous 2 pages). Comparison of the experimental and calculated ROESY buildup curves for the hexasaccharide (**7**). I) Experimental buildup curves II) Buildup curves calculated from the constrained dynamics trajectories in H_2O. III) Buildup curves calculated from the unconstrained dynamics trajectories in H_2O.

3.4.5. Hexasaccharide (8)

Roesy Spectrum

The hexasaccharide (**8**) is the final structure of the *Streptococcus* series. The ROESY spectrum of **8** and the corresponding F1 slices are shown in Figure 3.46 and Figure 3.47. The salient features are labeled on the spectrum. The spectrum appears to share the same features as the spectra of the related structures **4-7** , and although the hexasaccharide (**8**) would seem to be more sterically crowded than the other structures there do not seem to be any obvious compensations in terms of conformation. The characteristic 1B-5A' cross peak that appears in all the α-(1→2) linkages is present for both α-(1→2) linkages (the A'-B and A-B' linkages) in the hexasaccharide (**8**). The single α-(1→3) linkage joining the B' and the A' ring also displays the strong 1B'-3A' peak with its accompanying 1B'-5A' partner. As with the other compounds, the 1B'-2A' and 1B'-4A' peaks are also present, making it difficult to discern which peak is an ROE and which is due to residual TOCSY effects. An additional cross peak that is observed

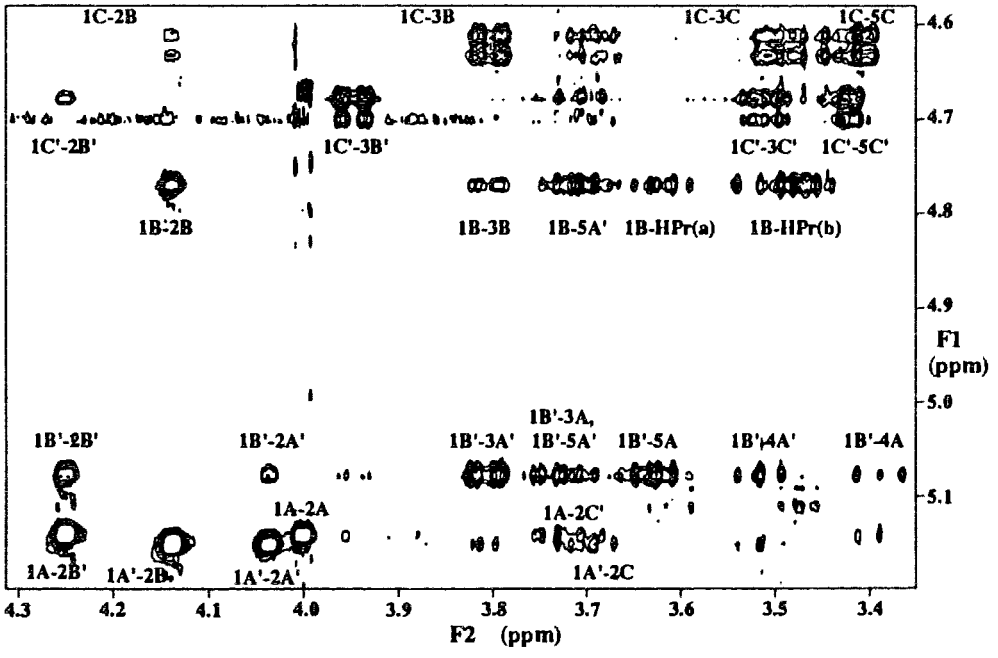


Figure 3.46. Expansion of the ROESY spectrum of the hexasaccharide (**8**).

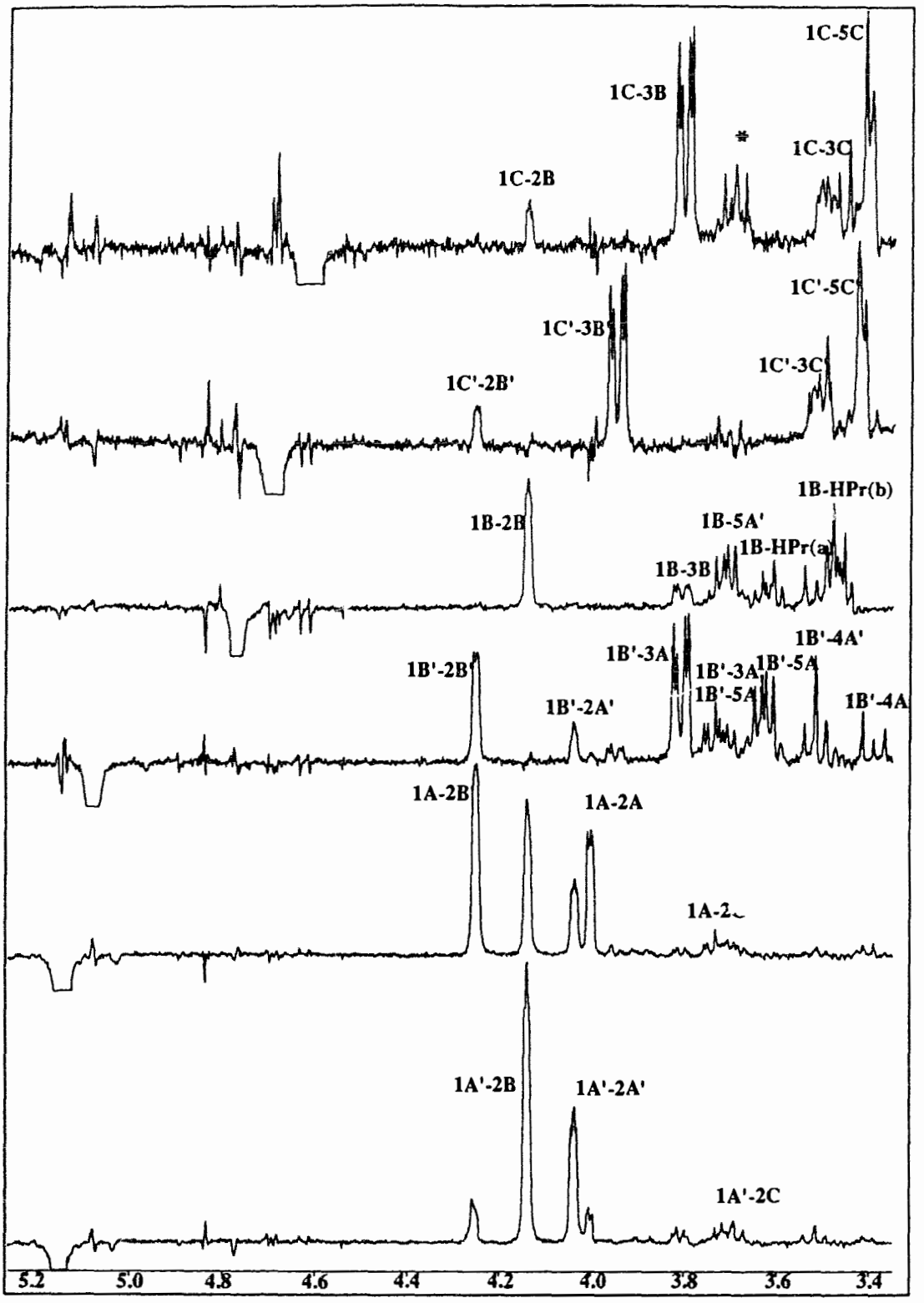


Figure 3.47. F1 slices from the ROESY spectrum of the hexasaccharide (**8**). The F1slice through the chemical shift of the 1A proton shows residual peaks from the 1A' cross peaks and vice versa due to the close chemical shift of the two protons. * Residual TOCSY peaks to the H2 and H6 of the Glc*p*NAc rings.

Table III.20. Offset and Hartmann Hahn effects calculated for the hexasaccharide (**8**).

Ring	Cross peak	$sin\alpha_I\ sin\alpha_S$	HH transfer	s^2
Ring A	1A-2A	0.9944	a)2A-1A	0.02
			b)2A-3A	0.07
	1A-2B'	0.996	a1B'-2B'	0.05
			b)2B'-3B'	0.08
	1A-2C'	0.9899	a)1C'-2C'	0.08
			b)2C'-3C'	0.40
Ring B	1B-2B	0.9964	a)2B-1B	0.02
			b)2B-3B	0.05
	1B-HPr(a)	0.9894	a)HPr(a)-HPr(b)	0.65
			b)HPr(a)-H3Pr	0.00
	1B-HPr(b)	0.9867	a)HPr(a)-HPr(b)	0.65
			b)HPr(b)-H3Pr	0.00
	1B-5A'	0.9909	a)H5A'-H6A'	0.00
			b)H5A'-H4A'	0.45
Ring C	1C-5C	0.9851	a)5C-4C	1.00
			b)5C-6Cs	0.00
			c)5C-6CR	0.07
	1C-3C	0.9870	a)3C-4C	0.68
			b)3C-2C	0.42
	1C-2B	0.9961	a)2B-1B	0.02
			b)2B-3B	0.05
	1C-3B	0.9921	a)3B-2B	0.05
			b)3B-4B	0.18
Ring A'	1A'-2B	0.9957	a)1B-2B	0.02
			b)2B-3B	0.05
	1A'-2A'	0.9946	a)1A'-2A'	0.02
			b)2A'-3A'	0.10
	1A'-2C	0.9899	a)1C-2C	0.42
			b)2C-3C	0.07
Ring B'	1B'-2B'	0.9971	a)2B'-1B'	0.05
			b)2B'-3B'	0.08
	1B'-2A'	0.9949	a)1A'-2A'	0.02
			b)2A'-3A'	0.10
	1B'-3A'	0.992	a)2A'-3A'	0.10
			b)3A'-4A'	0.26
	1B-5A	0.9892	a)4A-5A	0.30
			b)5A-6A	0.00
Ring C'	1C'-5C'	0.9856	a)5C'-4C'	1.00
			b)5C'-6C's	0.00
			c')5C'-6C'R	0.08
	1C'-3C'	0.9874	a)3C'-4C'	0.72
			b)3C'-2C'	0.40
	1C'-2B'	0.9974	a)2B'-1B'	0.05
			b)2B'-3B'	0.08
	1C'-3B'	0.9942	a)2B'-3B'	0.08
			b)3B'-4B'	0.14

along the 1B' chemical shift can be assigned unambiguously to the 4A' proton by virtue of its extreme upfield shift relative to the other H4 protons.

The degree of overlap in the spectrum of the hexasaccharide (**8**) is less than that of the hexasaccharide (**7**). Although the chemical shifts of the 1A and 1A' are very close, the cross peaks are well separated, permitting unambiguous assignment and integration. An exception is provided by the 1A-2C' and 1A'-2C cross peaks, which have been integrated together. Both peaks are of low intensity, so it is impossible to judge the relative magnitudes of each. In order to get some estimate of these distances, the integrals were divided equally. Contacts derived from the ROESY spectrum are displayed in Figure 3.48. The offset and HOHAHA corrections are tabulated in Table III.20 and displayed in Figure 3.49.

Calculation of Internuclear Distances

After the corrections, the interproton distances were calculated. These distances are displayed in Figure 3.50 and the constraints derived from them are tabulated below.

Table III.21. Constraints input for the molecular dynamics simulation of the hexasaccharide (**8**).

Proton Pair	Constrained distance [Å]	Upper limit[a]	Lower limit
H1B-HPr(a)	2.79	2.94	2.69
H1B-HPr(b)	2.44	2.69	2.34
H1B-H5A'	2.64	2.84	2.54
H1A'-H2B	2.22	2.42	2.12
H1A'-H2C	3.06	3.26	2.91
H1B'-H2A'	3.45	3.60	3.30
H1B'-H3A'	2.38	2.23	2.28
H1B'-H5A	2.59	2.74	2.44
H1A-H2B'	2.30	2.45	2.20
H1A-H2C'	3.05	3.35	2.90
H1C-H2B	2.95	3.30	2.65
H1C-H3B	2.18	2.33	2.08
H1C'-H2B'	2.85	3.05	2.65
H1C'-H3B'	2.20	2.35	2.10

[a]The limits were derived from the highest and lowest values obtained from the experiment.

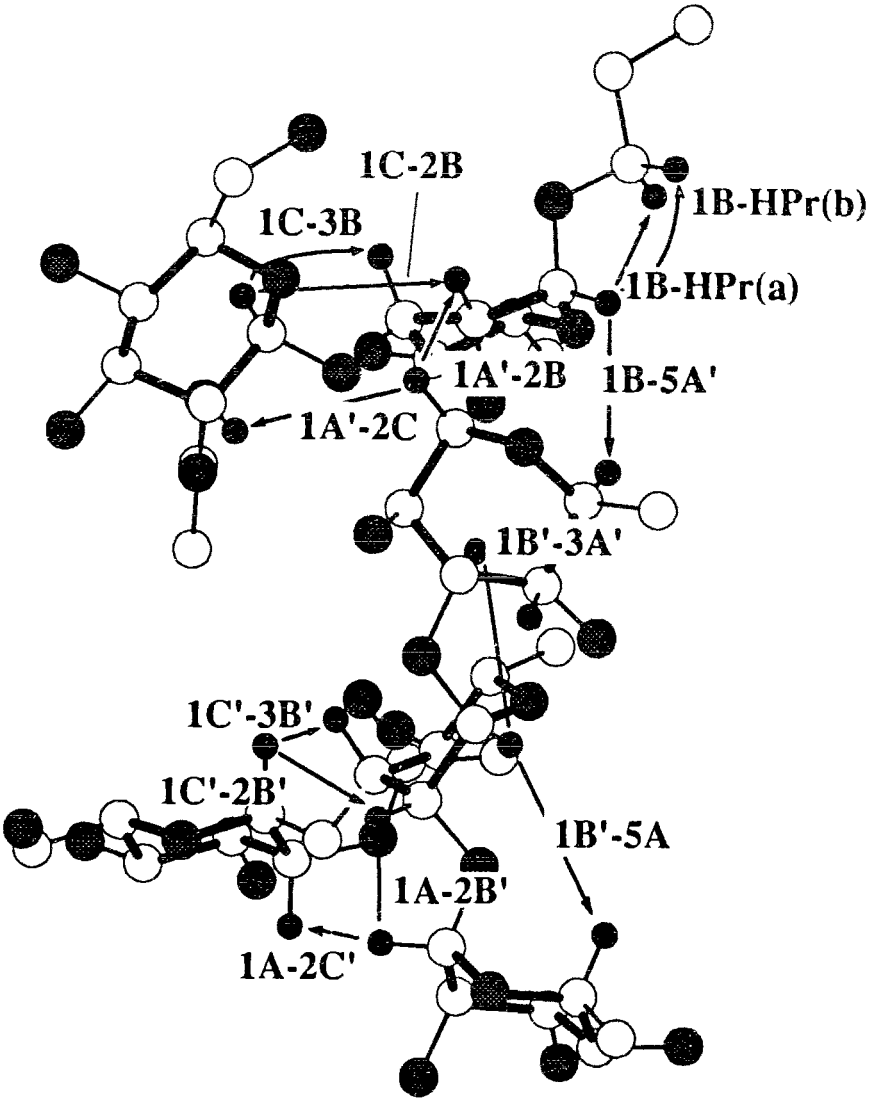


Figure 3.48. Interproton contacts derived from the ROESY spectrum of the hexasaccharide (**8**).

Molecular Dynamics Simulations

Dynamics simulations, both constrained and unconstrained, were calculated and the interglycosidic (Φ, Ψ) angles and the trajectories of selected internuclear distances are displayed in Figures 3.51 - 3.54. Following the dynamics calculations, ROESY buildups were calculated and compared to experimental ROE data. Both calculated and experimental buildup curves are displayed in Figure 3.55.

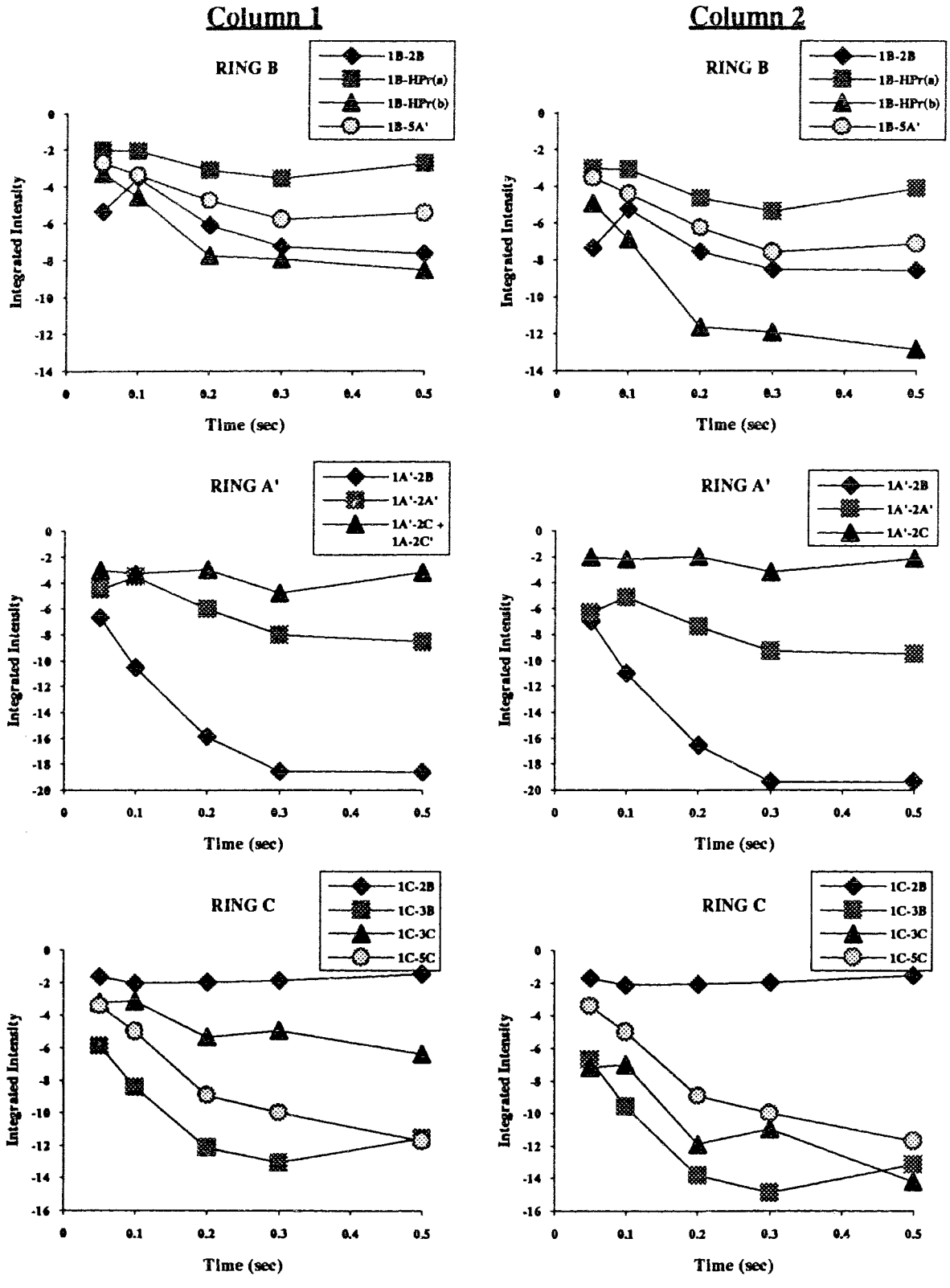


Figure 3.49. (Caption overleaf).

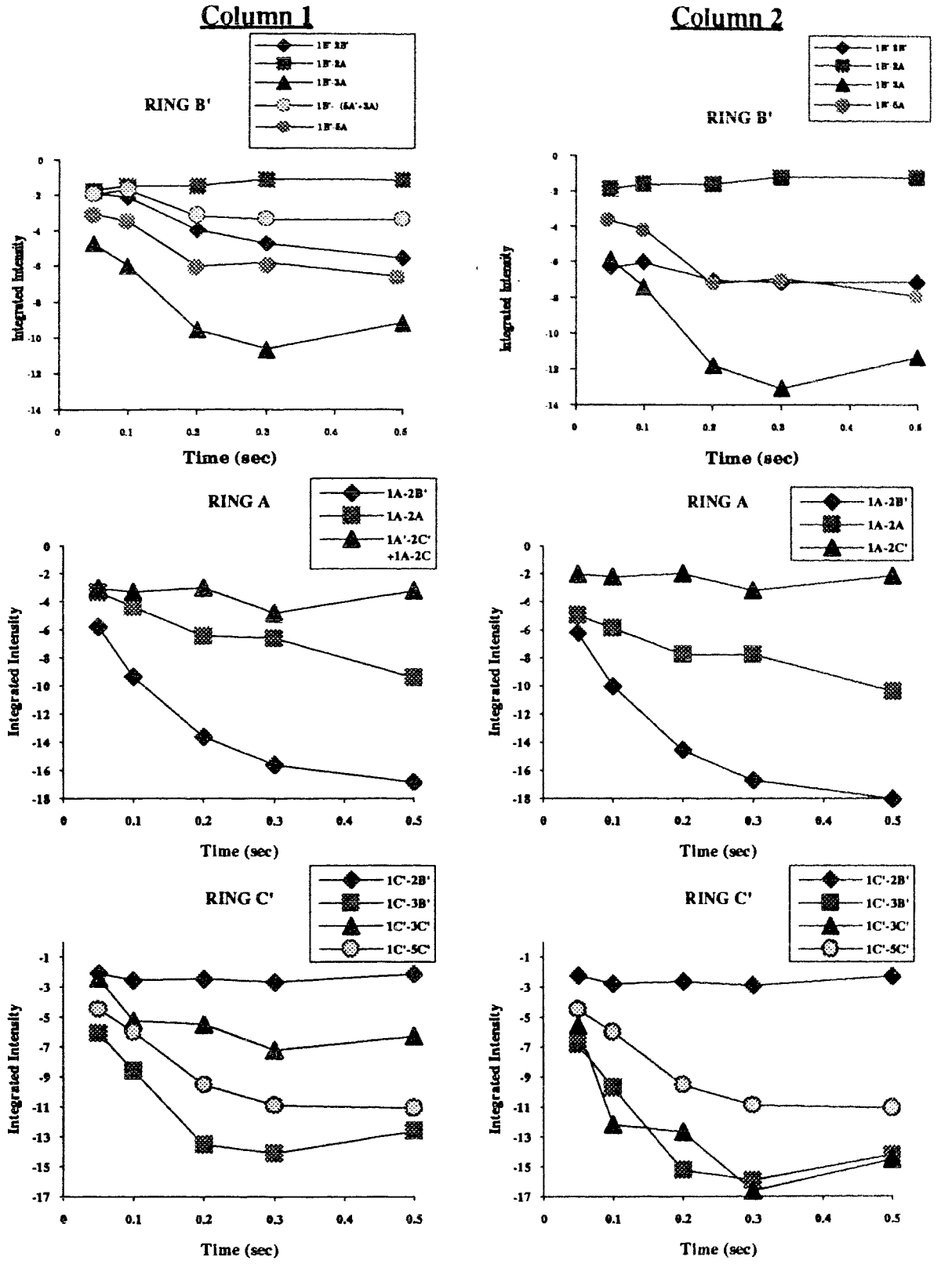


Figure 3.49. The effect of offset and Hartmann Hahn corrections on the ROE buildup of the hexasaccharide (**8**). Column 1.) Uncorrected data. Column 2.) Corrected data.

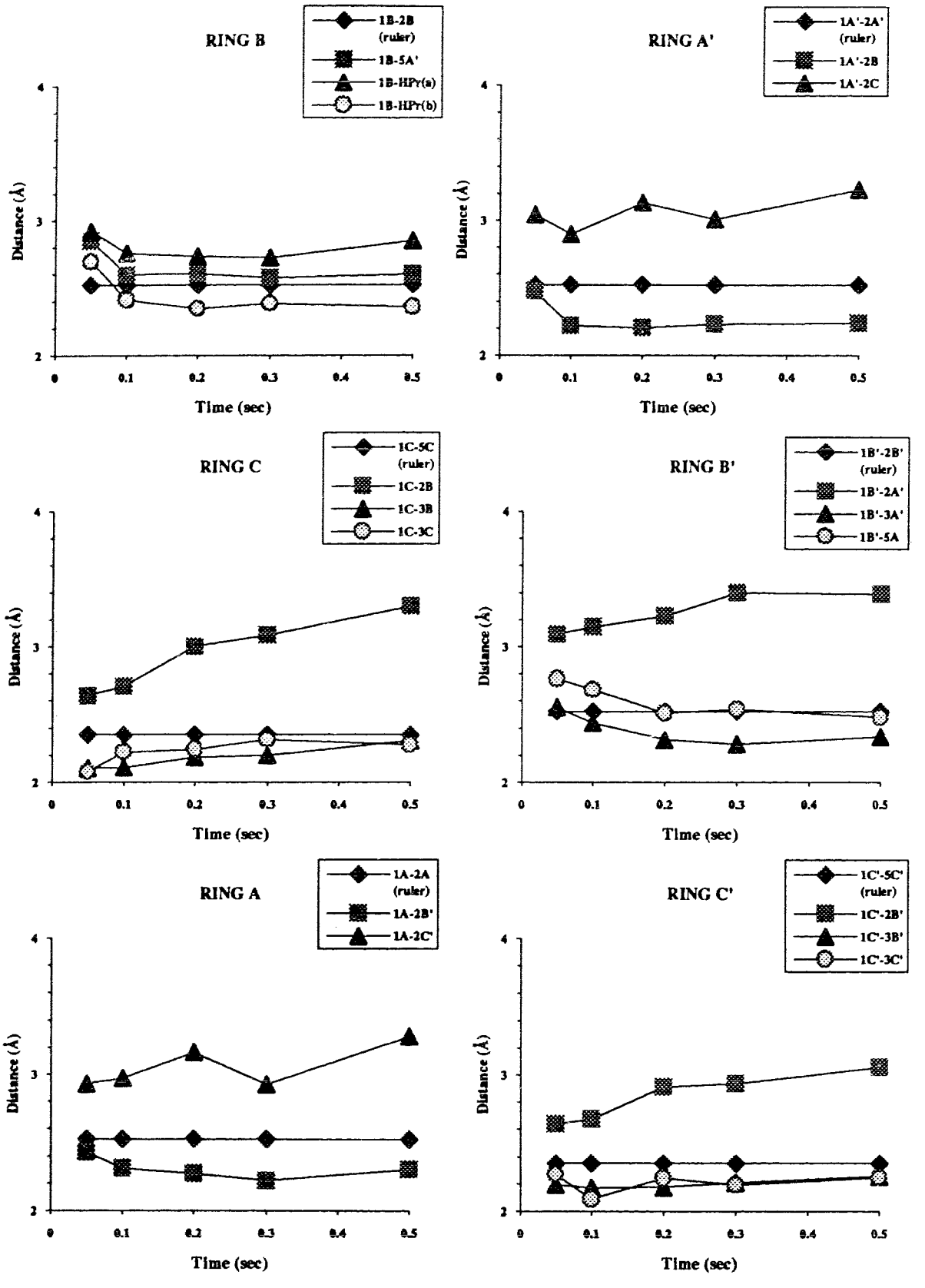
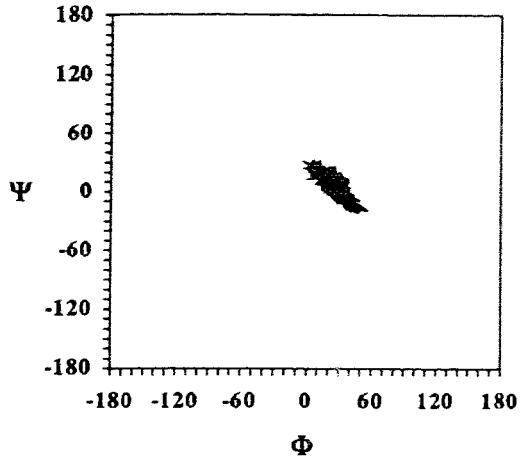
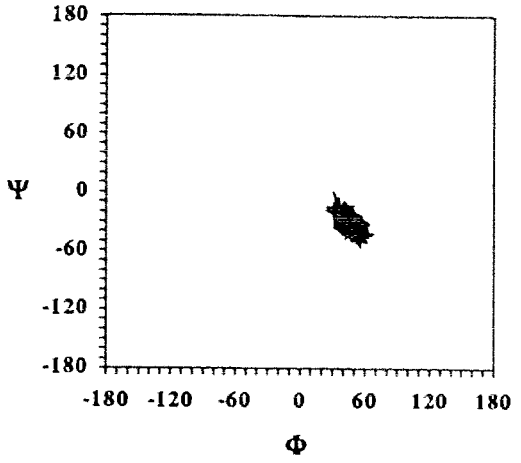


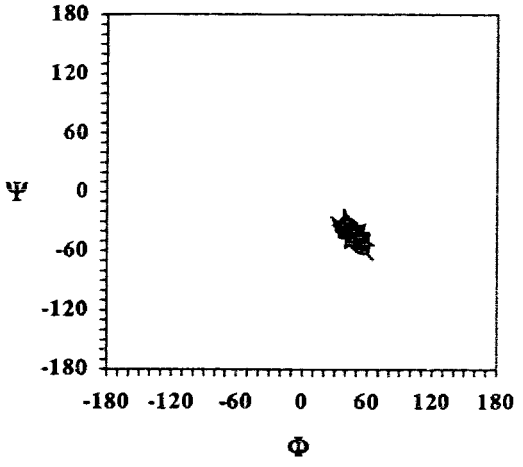
Figure 3.50. Distances calculated from the ROESY data of the hexasaccharide (8).



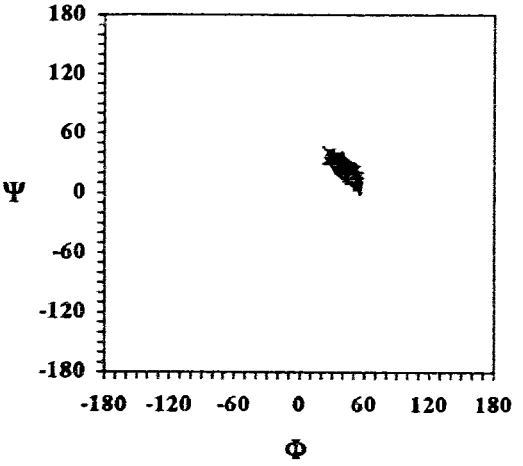
A'-B α-(1→2) linkage



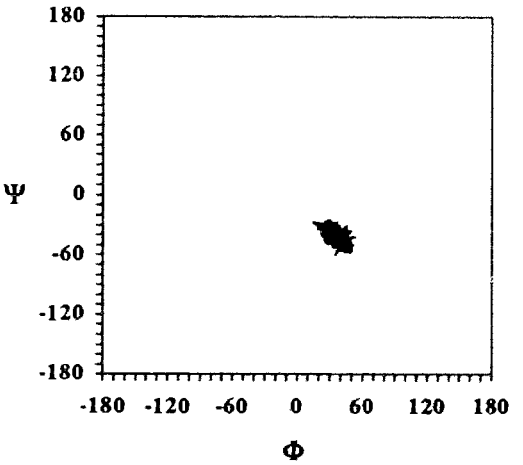
C-B β-(1→3) linkage



B'-A' α-(1→2) linkage

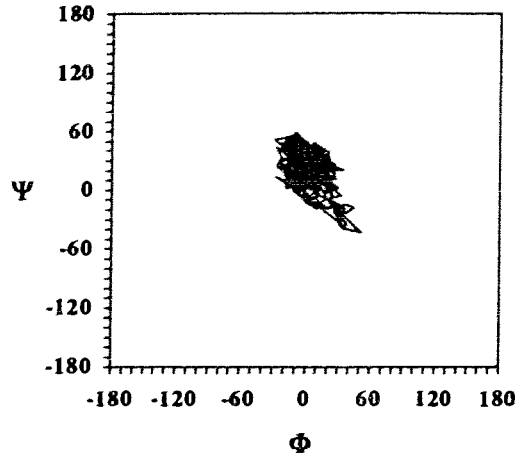


A-B' α-(1→2) linkage

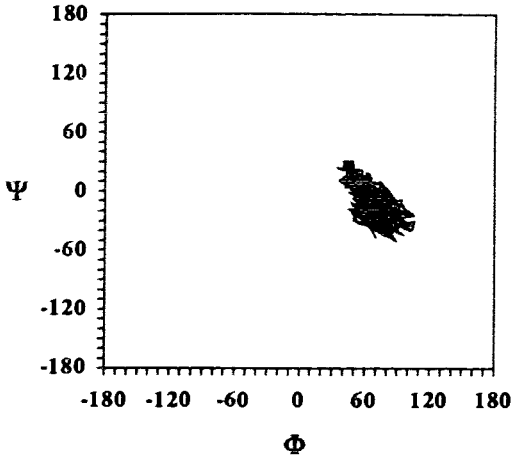


C'-B' β-(1→3) linkage

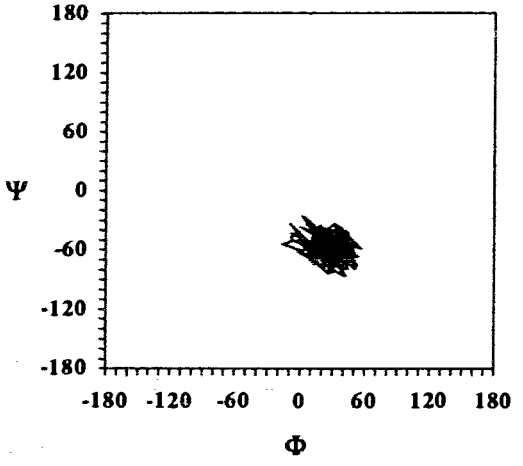
Figure 3.51. Variation of the Φ, Ψ angles during constrained dynamics simulations of the hexasaccharide (**8**) in H_2O.



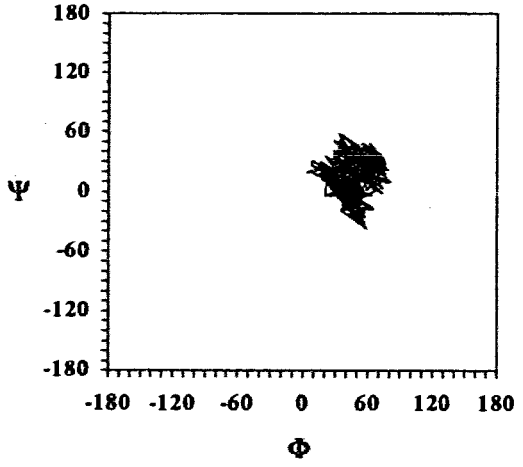
A'-B α-(1→2) linkage



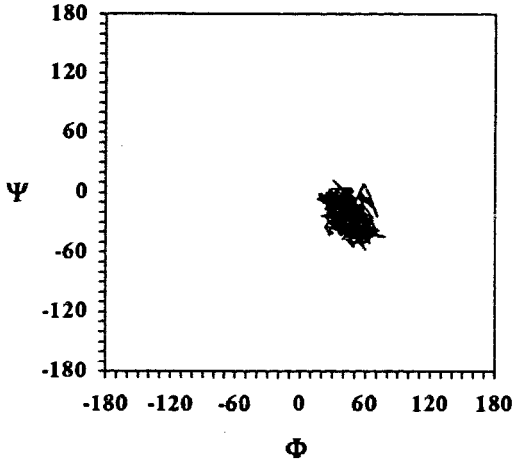
C-B β-(1→3) linkage



B'-A' α-(1→2) linkage

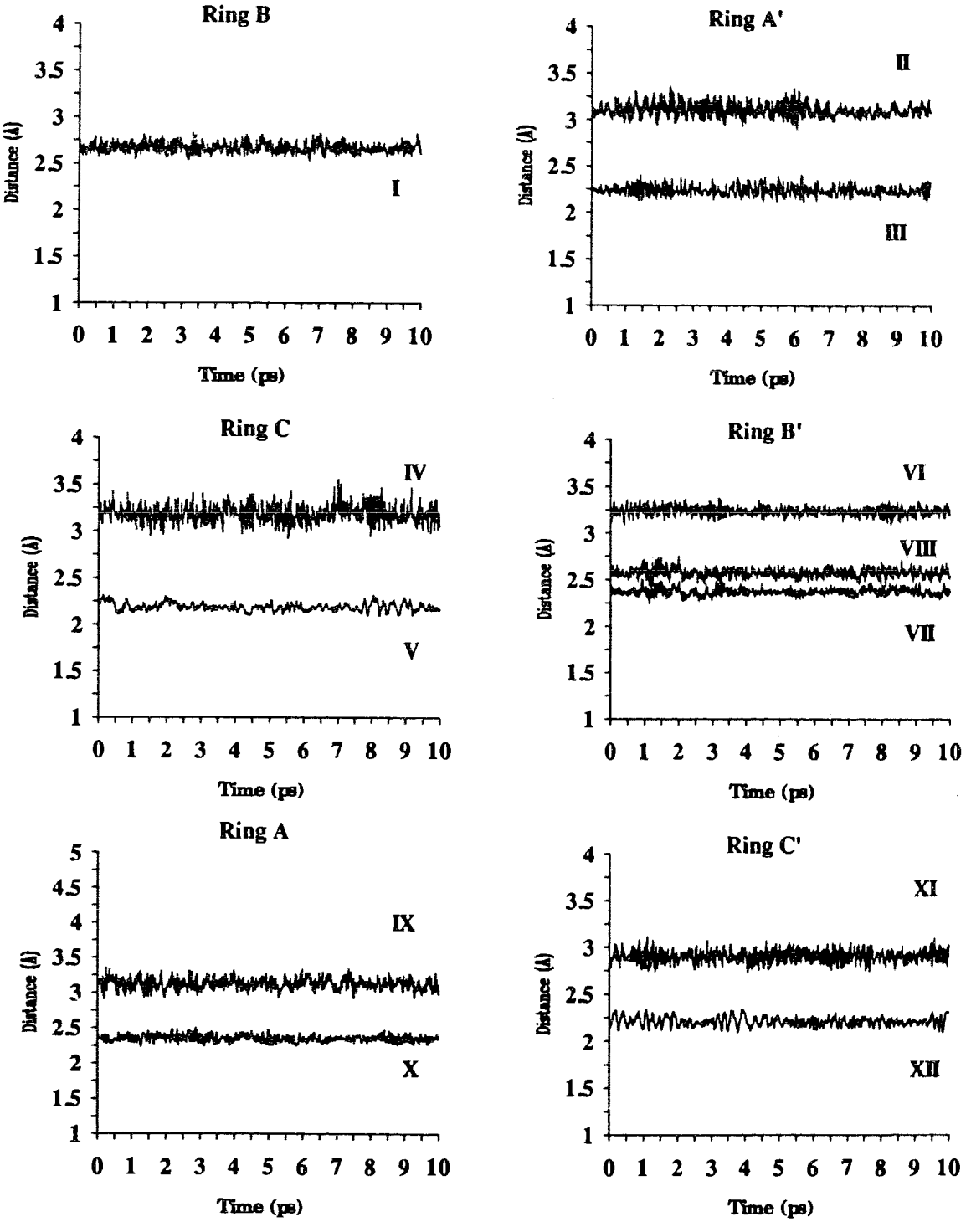


A-B' α-(1→2) linkage



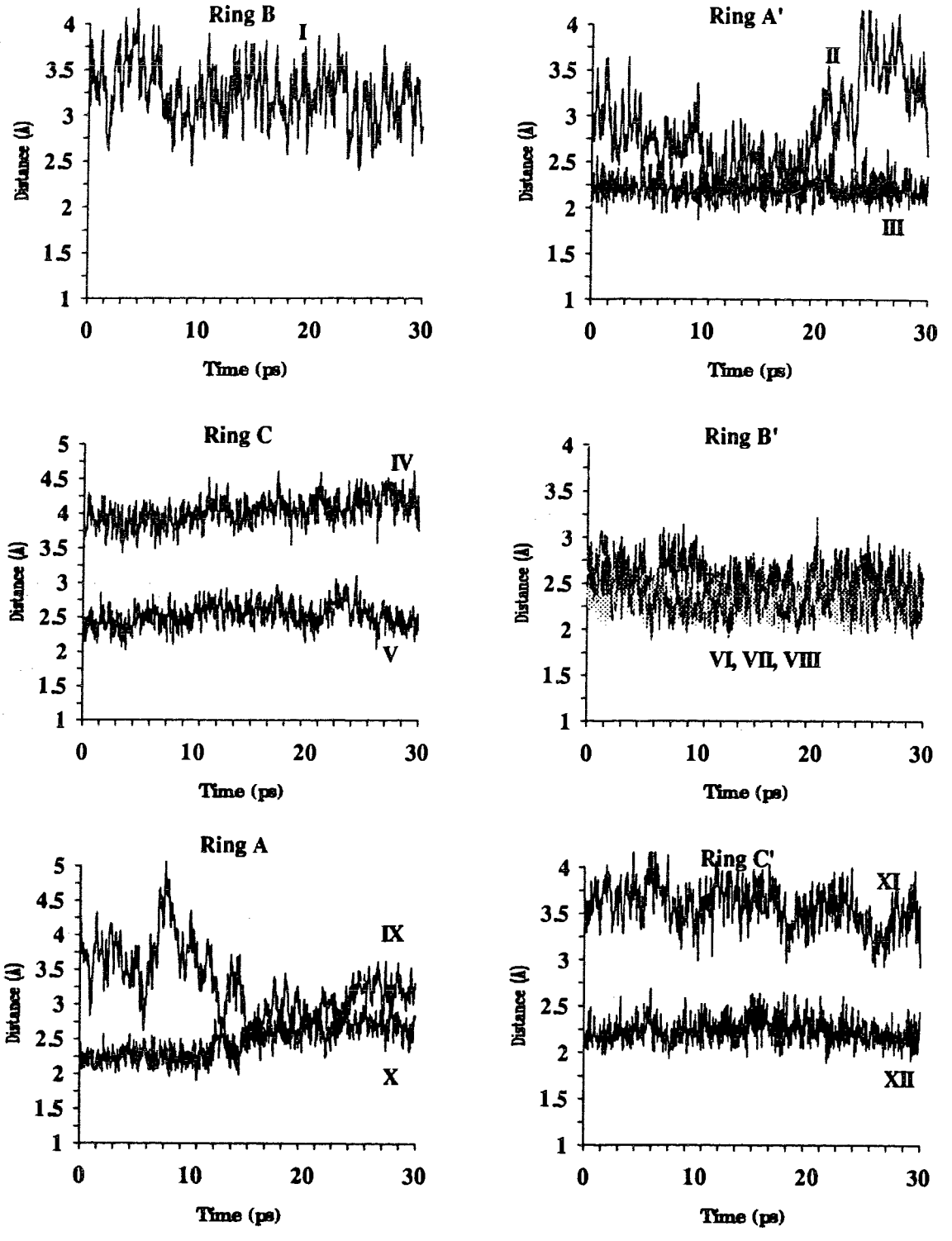
C'-B' β-(1→3) linkage

Figure 3.52. Variation of the Φ, Ψ angles during unconstrained dynamics simulations of the hexasaccharide (**8**) in H_2O.



LEGEND			
I)	1B-5A'	VII)	1B'-3A'
II)	1A'-2C	VIII)	1B'-5A
III)	1A'-2B	IX)	1A-2C'
IV)	1C-2B	X)	1A-2B'
V)	1C-3B	XI)	1C'-2B'
VI)	1B'-2A'	XII)	1C'-3B'

Figure 3.53. Variation of selected interproton distances during constrained dynamics simulations of the hexasaccharide (**8**) in H_2O.



LEGEND			
I)	1B-5A'	VII)	1B'-3A'
II)	1A'-2C	VIII)	1B'-5A
III)	1A'-2B	IX)	1A-2C'
IV)	1C-2B	X)	1A-2B'
V)	1C-3B	XI)	1C'-2B'
VI)	1B'-2A'	XII)	1C'-3B'

Figure 3.54. Variation of selected interproton distances during unconstrained dynamics simulations of the hexasaccharide (**8**) in H_2O.

Table III.22. Average, high, and low values of selected internuclear distances during dynamics simulations of the hexasaccharide (**8**)[a].

	Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Proton Pair	Average	High	Low	Average	High	Low
1B-HPr(a)	2.81	2.97	2.69	3.16	3.99	2.13
1B-HPr(b)	2.45	2.62	2.33	2.55	3.62	1.98
1B-5A'	2.67	2.83	2.53	3.23	4.39	2.33
1A'-2B	2.24	2.42	2.12	2.22	2.87	1.87
1A'-2C	3.11	3.36	2.91	2.88	4.45	1.95
1C-2B	3.19	3.55	2.93	4.03	4.70	3.39
1C-3B	2.19	2.32	2.10	2.52	3.23	1.98
1B'-2A'	3.24	3.38	3.08	2.51	3.14	1.95
1B'-3A'	2.38	2.51	2.25	2.30	2.87	1.91
1B'-5A	2.58	2.76	2.43	2.39	3.23	1.85
1A-2B'	2.35	2.50	2.23	2.47	3.12	1.90
1A-2C'	3.13	3.36	2.91	3.31	5.07	2.20
1C'-2B'	2.91	3.12	2.73	3.57	4.38	2.86
1C'-3B'	2.21	2.34	2.09	2.24	2.80	1.89

[a]Average temperature during the dynamics simulation was 299 K for the simulation with constraints and 298 K for the simulation without constraints.

Table III.23. Average, high, and low values of the glycosidic Φ and Ψ angles during dynamics simulation of the hexasaccharide (**8**)[a]..

	Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Angle	Average	High	Low	Average	High	Low
Link A-Propyl						
Φ	50.13	71.48	16.45	60.11	99.34	-16.87
Link A'-B						
Φ	25.86	56.47	-3.01	2.72	57.68	-38.31
Ψ	5.85	33.09	-21.81	28.16	63.13	-51.33
Link C-B						
Φ	46.93	67.82	23.63	68.74	113.26	34.45
Ψ	-32.58	0.43	-55.99	-15.87	31.46	-58.65
Link B'-A'						
Φ	48.11	66.15	26.00	29.87	57.09	-16.24
Ψ	-44.45	-17.45	-69.40	-57.68	-25.32	-86.60
Link A-B'						
Φ	42.04	59.16	21.34	47.06	83.89	-1.46
Ψ	24.77	46.91	-1.59	18.19	57.96	-36.73
Link C'-B'						
Φ	36.26	53.04	15.01	45.44	78.56	14.41
Ψ	-41.66	-23.88	-60.21	-25.14	12.42	-63.90
ω and θ angles						
Ring C						
θ	2.29	54.41	-39.01	11.29	89.48	-53.36
ω	50.74	84.87	12.54	-4.16	84.53	-82.35
Ring C'						
θ	5.41	49.29	-30.94	12.50	84.27	-54.36
ω	34.61	76.17	-7.20	13.11	70.90	-62.68

[a]Average temperature during the dynamics simulation was 299 K for the simulation with constraints and 298 K for the simulation without constraints.

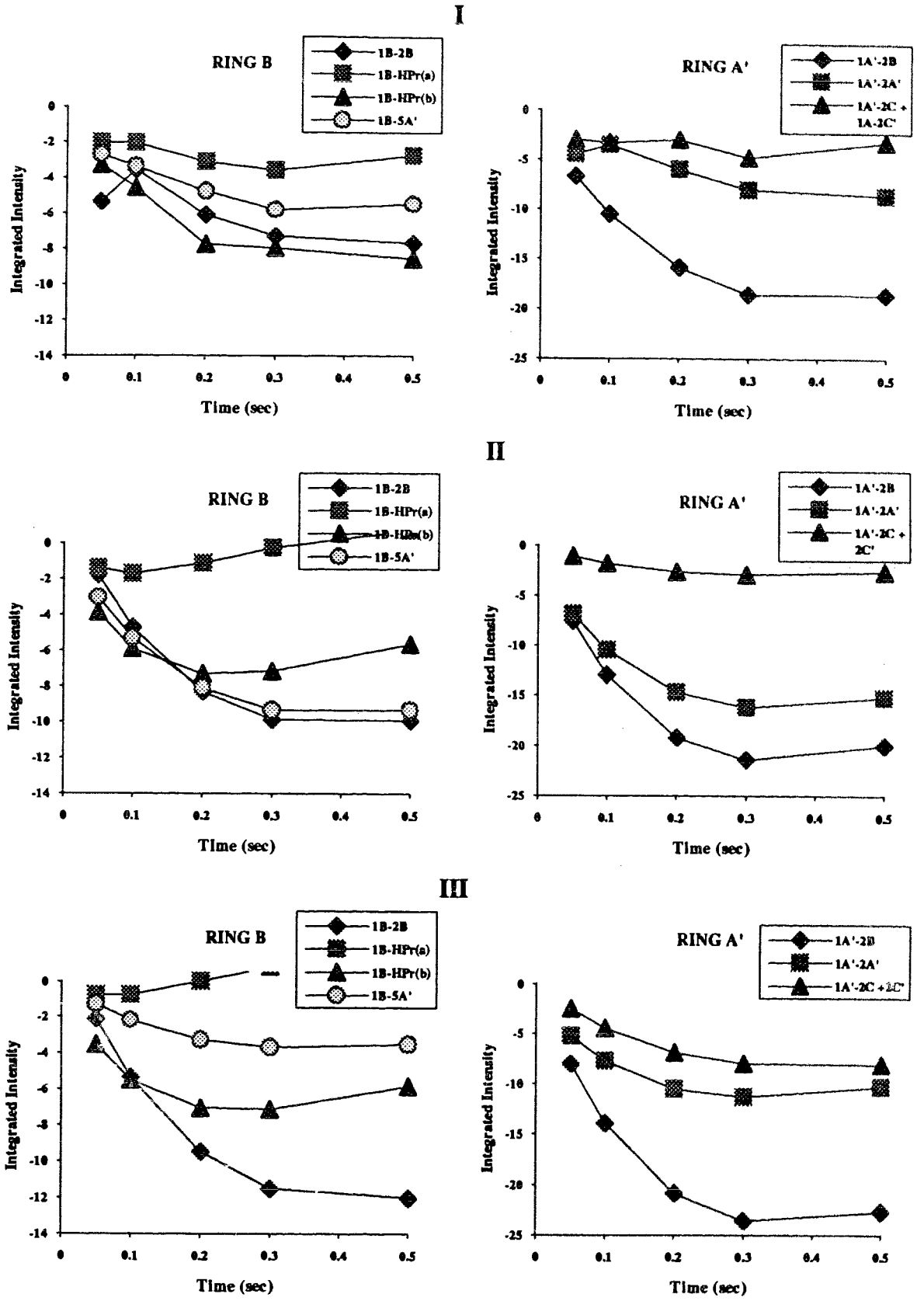


Figure 3.55. (Caption two pages overleaf)

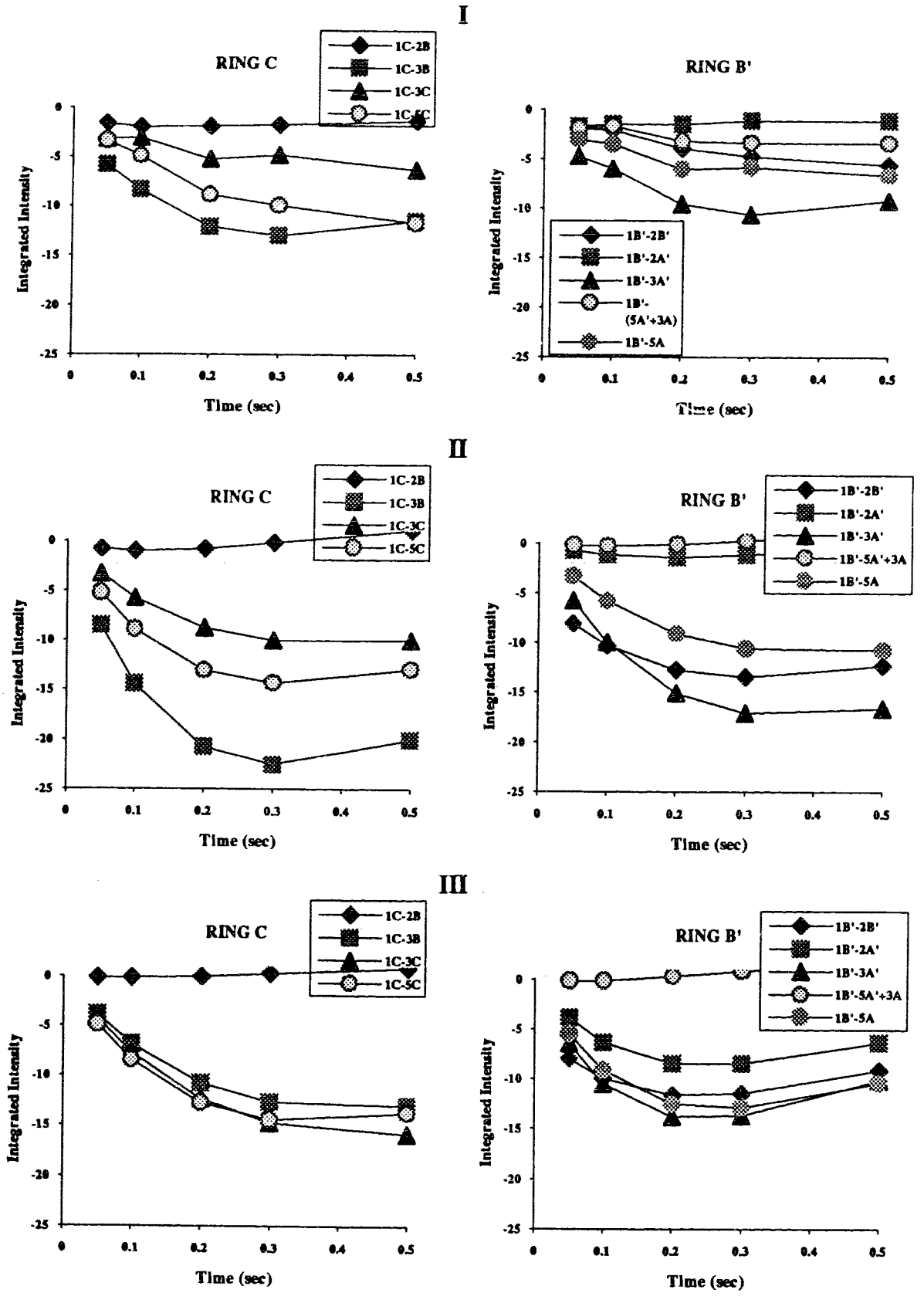


Figure 3.55. (Caption Overleaf).

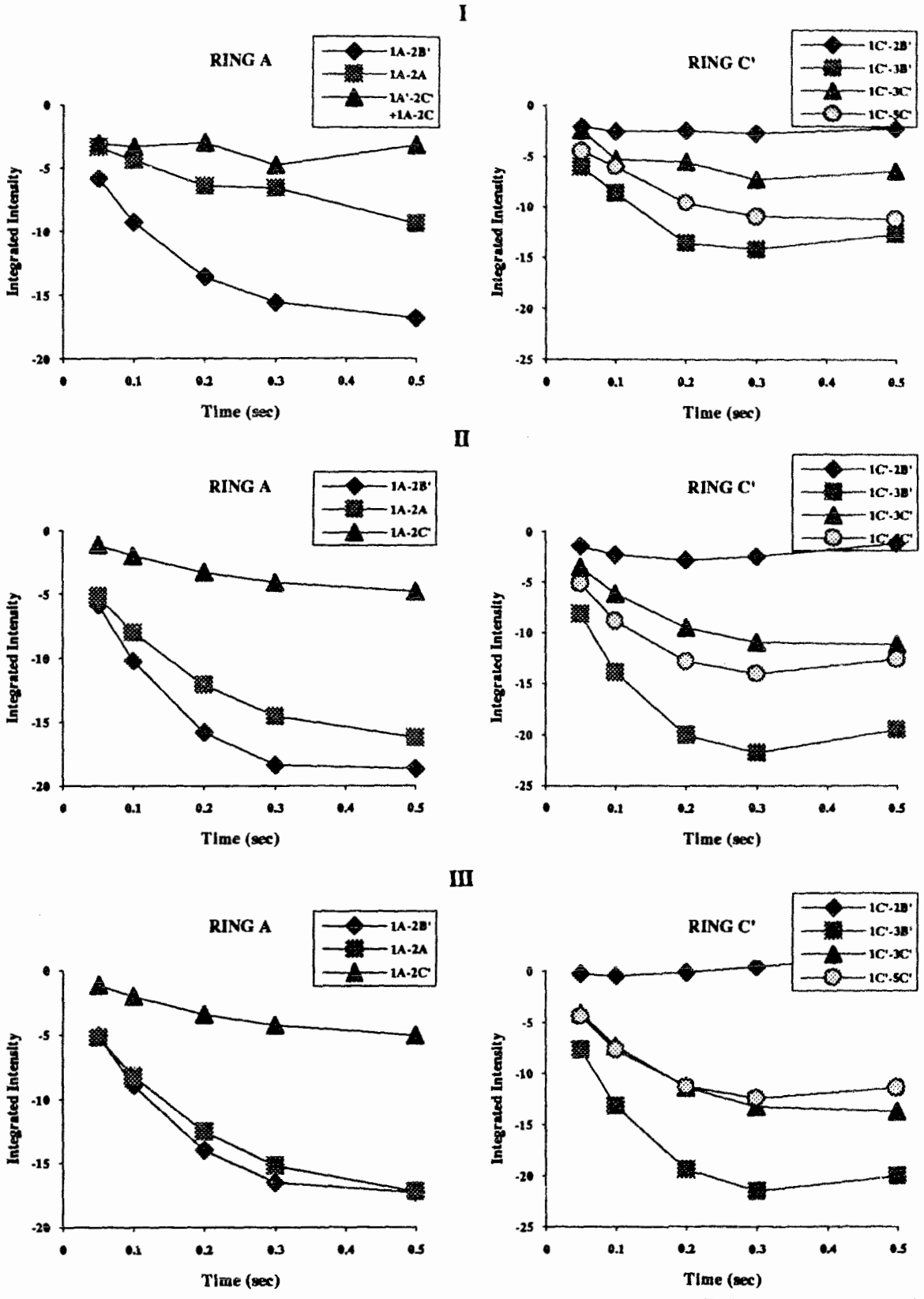


Figure 3.55. Comparison of the experimental and calculated ROESY buildup curves for the hexasaccharide (8). I) Experimental buildup curves I) Buildup curves calculated from the constrained dynamics trajectories in H_2O. III) Buildup curves calculated from the unconstrained dynamics trajectories in H_2O.

3.5. Discussion

Analysis of the ROESY spectra

The NMR spectra of the oligosaccharides (**4-8**) show many similar features, which leads to the conclusion that the compounds share similar conformational properties. The presence of characteristic cross peaks in the ROESY spectra of all the compounds examined confirms this fact. The average Φ and Ψ angles for the various linkages of the oligosaccharides (**4-8**) obtained from the constrained dynamics trajectories are listed in Table 3.24, and the structures corresponding to these 'average' values are superimposed in Figure 3.56.

There are three distinct types of linkage present in the *Streptococcus* oligosaccharides, the β-(1→3) link that joins the C ring (Glc*p*NAc) to the B (Rha*p*) ring, the α-(1→2) linkage from the A (Rha*p*) ring to the B ring and the α-(1→3) link from the B ring to the A ring. Of the five oligosaccharides examined, the compounds (**5-8**) have at least one example of each linkage, the exception being the trisaccharide (**4**), which does not possess an α-(1→3) link. The data from the ROESY experiments unambiguously defines the conformational properties of two of the three linkages.

The α-(1→2) (A-B' and A'-B) linkage

In all the α-(1→2) linkages, the 1A-2B cross peak is, without exception, the strongest peak observed from the anomeric proton of ring A. A second interaction observed between the anomeric proton of ring B and the H5 of ring A unambiguously defines the interglycosidic angles. This is in agreement with previous studies on similar compounds. A conformational study of the disaccharide α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→OMe)[77,179] yielded essentially the same results. The two cross peaks (H1A'-H2B and H1B-H5A') were observed, but in addition, a H1A'-H1B cross peak of much weaker intensity was also observed. Inspection of the anomeric region in the spectra of all the

Table III.24. Comparison of the average Φ and Ψ angles derived from constrained molecular dynamics simulations of the compounds **4-8**.

Link	A'-B		A-B'		B-A		B'-A'		C-B		C'-B'	
Compound	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ
4	39	16							44	-41		
5	50	24			35	-49			38	-46		
6	22	13					41	28	44	-36	37	-49
7	46	27			42	-46	40	28	33	-37	38	-40
8	26	6	42	25			48	-44	47	-33	36	-42

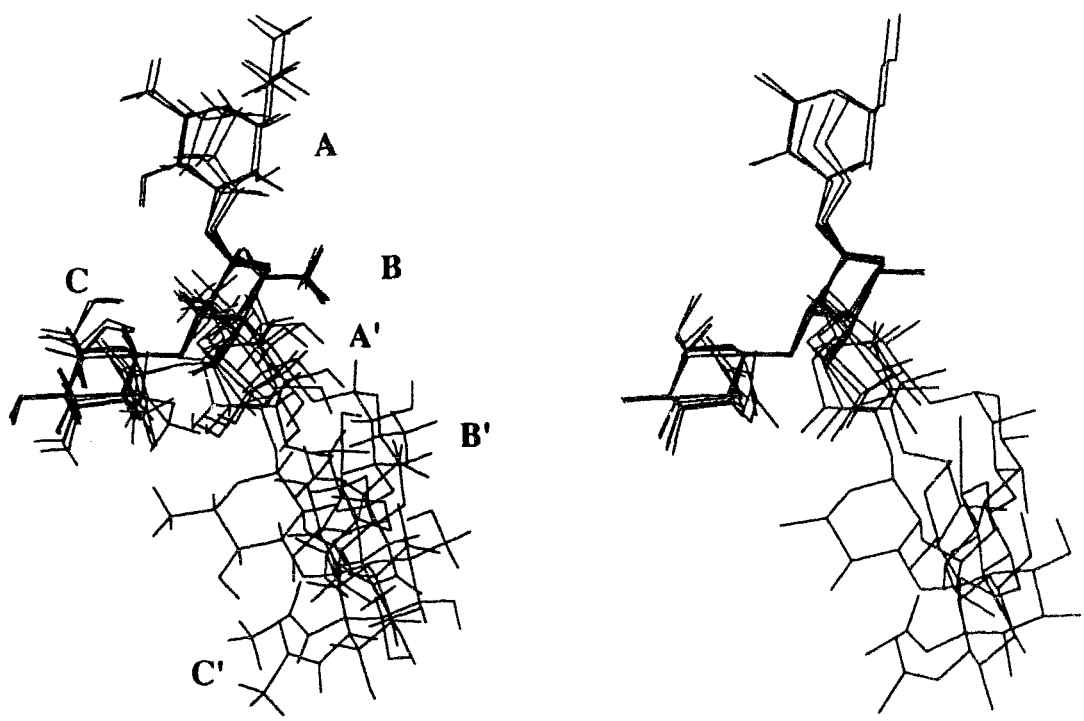


Figure 3.56. A superposition of the conformations derived from the average values of the constrained molecular dynamics simulations of compounds **4-8**. A) All atoms. B) All non hydrogen atoms.

compounds reveals no trace of an analogous 1A'-1B cross peak, which confirms further that the range of the Φ and Ψ angles lies in the positive region. The absence of this peak may be due to the additional residue in the 3-position of the B ring. However, in studies on the same disaccharide by other researchers[77,180,181] no H1'-H1 cross peak in the disaccharide α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→OMe), was reported.

The β-(1→3) (C-B and C'-B') linkage

In the C-B linkages, the 1C-3B cross peak is the strongest peak observed in the ROESY spectrum. This, in conjuction with the observation of a H1C-H2B cross peak, is sufficient to predict that both the Φ and Ψ angles are gauche. Further, the relative magnitudes of both cross peaks require that the Φ angle be positive and the Ψ angle be negative. This is the only set of angles for which the distances obtained from the experimental data are satisfied, and fits the preferences that would be predicted by the exoanomeric effect.

In all of the compounds, at "branch" points, i.e. where the fragment α-L-Rha*p*-(1→2)[β-D-Glc*p*NAc-(1→3)]-α-L-Rha*p* occurs, a third cross peak, of very low intensity, which conclusively fixes the locale of the Φ and Ψ angles of both the A'-B and the C-B linkage was identified. This is the 1A'-2C cross peak and it is present in the spectrum of every compound studied that possesses the branched type structure α-L-Rha*p*-(1→2)[β-D-Glc*p*NAc-(1→3)]-α-L-Rha*p*. This cross peak is in perfect agreement with the geometries of the A'-B and C-B link already described, and is the final piece of evidence which confirms the conformation of this fragment.

The α-(1→3) (B-A and B'-A') linkage

The α-L-(1→3) linkage, i.e., the B-A or B'-A' linkage, does not provide as many clues to its conformation as the A-B and C-B links. The B-A linkage is a point of major divergence in conformation amongst the compounds **4-8**. This can be seen from the

superposition of structures in Figure 3.56 and is due to the difference in Ψ angles between the compounds as a result of the imposed constraints.

The strong inter glycosidic 1B-3A* cross peak is present, but closer examination of the spectra of compounds **5-8** shows that there are 1B-2A as well as 1B-4A cross peaks present. Additionally, in all of the compounds which possess an α-L-(1→3) linkage, besides the cross peaks mentioned, there appears a cross peak which was identified as a H1B-H5A interaction. A brief glance at a model of the B-A fragment of the compounds (**5-8**) indicates that is impossible to bring the 1B proton in close proximity to the H5 of the A ring, which dispels any possibility of this being a direct effect. In fact the closest distance between the 1B and 5A protons, given the $^1C_4$ conformation of the A ring, is 3.64 Å. Earlier studies on the disaccharide α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→OMe) and trisaccharides containing this fragment[77,180,181] indicate only a 1B-3A cross peak, and make no mention of a 1B-5A interaction.

There are two possible explanations for the observation of both the 1B-4A and 1B-2A cross peaks. There could be two conformations about this linkage or, more simply, TOCSY effects could account for one or both of these cross peaks. Residual TOCSY effects are a drawback of the ROESY experiment. Recently, Hwang et. al.[178] have proposed a new pulse sequence that eliminates TOCSY effects in ROESY experiments. In order to distinguish between the two possibilities, i.e., multiple conformations vs TOCSY, the sequence proposed by Hwang et. al.[178], herein referred to as a filtered ROESY, was used to acquire a spectrum of the pentasaccharide (**6**). The only visible cross peaks from the anomeric proton of ring B are to the H3 and to the H4 of ring A. The additional peaks (H1B-H2A, H1B-H5A) that were seen in the ROESY spectrum must therefore be due to residual Hartmann Hahn effects. Inspection of the

* The following discussion applies to both the B-A and the B'-A' linkages, but for the sake of brevity, only the B-A linkage is specified.

Hartmann Hahn factors for the proton pairs H3-H2 and H4-H5 for the A and A' rings reveal moderately large values of the transfer efficiency s^2 (~0.1 and 0.3 for the H3-H2 and H4-H5 proton pairs), which is consistent with this hypothesis.

Further, all of the compounds studied possess another (1→3) link in the form of the [β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-] fragment (the C-B fragment). This glycosidic linkage, which displays the same strong cross peak from the anomeric proton of the C ring to the 3B proton of the aglycon, does not exhibit any cross peak to the 5B proton. This would indicate that besides the difference in configuration at the anomeric center, the two fragments, B-A and C-B, have different geometries. A final piece of evidence to support this is the presence of a 2A'-5B' cross peak in the spectrum of the pentasaccharide (**6**). The average Φ and Ψ angles of the compounds **4**, **7** and **8**, obtained from the molecular dynamics simulations calculated with the 1B-4A constraints, 40° and 28° respectively, place the H5 of ring B' in close proximity to the H2 of the A' ring. A negative Ψ angle, such as postulated for the compounds **5**, **6** and **8**, would not place the H5 in the same position while keeping the H1B-H2A distance within the constraints. Inspection of the ROESY spectra of the compounds **5-8** revealed that this cross peak was definitely present in the spectrum of all of the compounds except for the tetrasaccharide (**5**) where interference with other peaks in this region made identification of the 2A'-5B' difficult. These results indicate that both the Φ and Ψ angles for this linkage are both positive, and that three α–(1→3) linkages, i.e., the B-A linkages in compounds **5-7** and the B'-A' linkage in compound **8**, have been incorrectly assigned negative Ψ angles.

Consequently, the 1B(')-2A(') constraint was replaced by a 1B(')-4A(') constraint, and the dynamics simulations, both constrained and with the constraints removed, were recalculated for the three compounds **5**, **7** and **8**. Analysis of the trajectories shows that the Φ, Ψ angles of the other linkages are not influenced by the different conformation of the α(1→3) linkage. The results of the simulations are illustrated in Figures 3.58 -3.59.

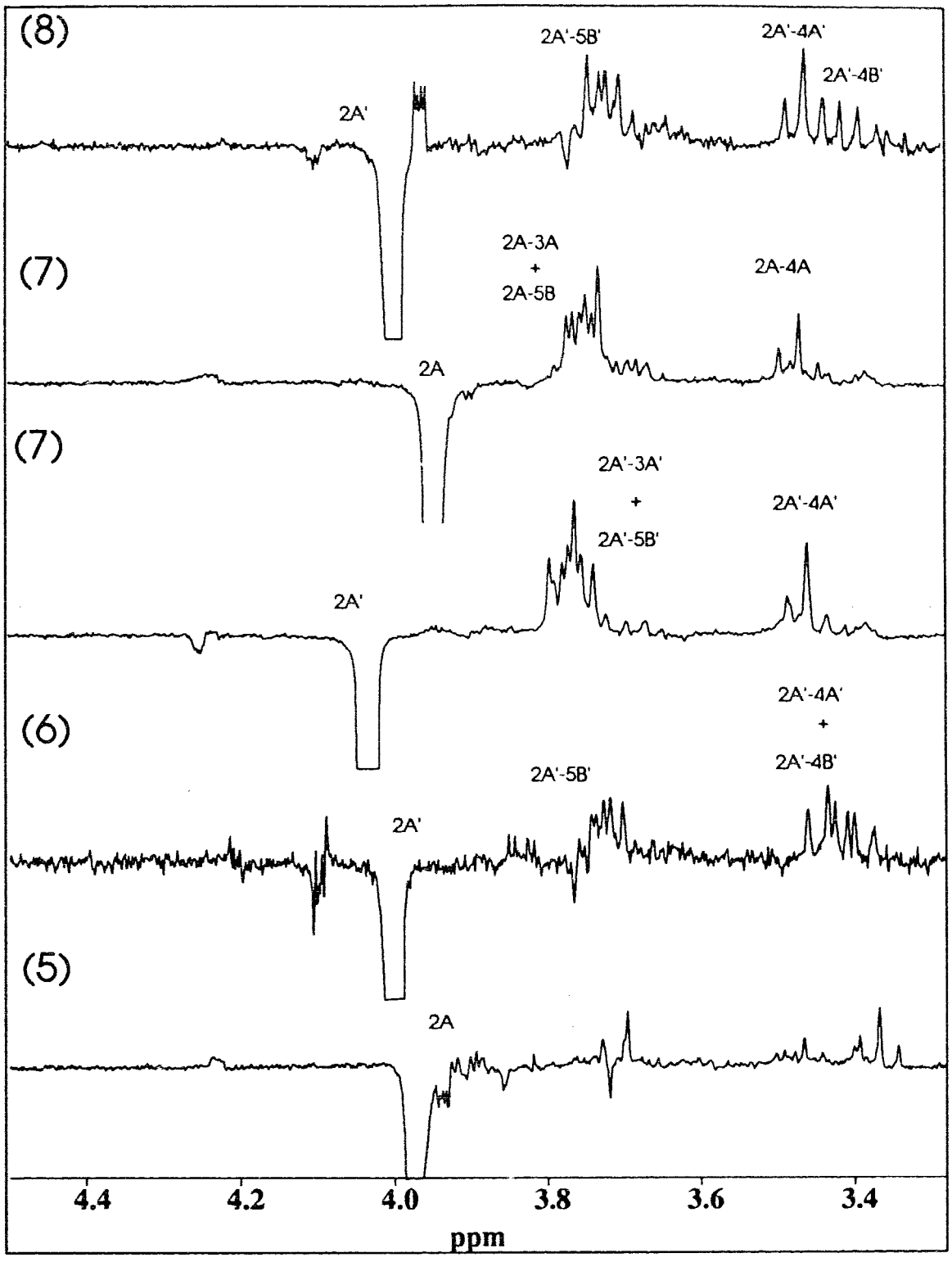


Figure 3.57. F1 slices from the ROESY spectra of compounds **4-8**, which display the presence of a H2'-H5 interaction in the α-(1→3) linkages.

As expected, the constrained dynamics simulations show a narrow range of motion. In the trajectories from the dynamics simulations with the constraints removed, the behavior of the B-A linkage of both the tetrasaccharide (**5**) and the hexasaccharide (**7**) is similar - the Ψ angles seem to be moving toward the negative quadrant, which is the minimum predicted by CHARMM. This is not observed in the B'-A' linkage of the hexasaccharide **8**. ROESY buildup curves were calculated with CROSREL from the new dynamics trajectories for the affected linkage in the three compounds, but though an improvement in the fit is observed, it is marginal.

A new set of average values of the Φ and Ψ angles as calculated from these results is tabulated in Table III.25. A superposition of these "new" structures thus constructed is displayed in Figure 3.61.

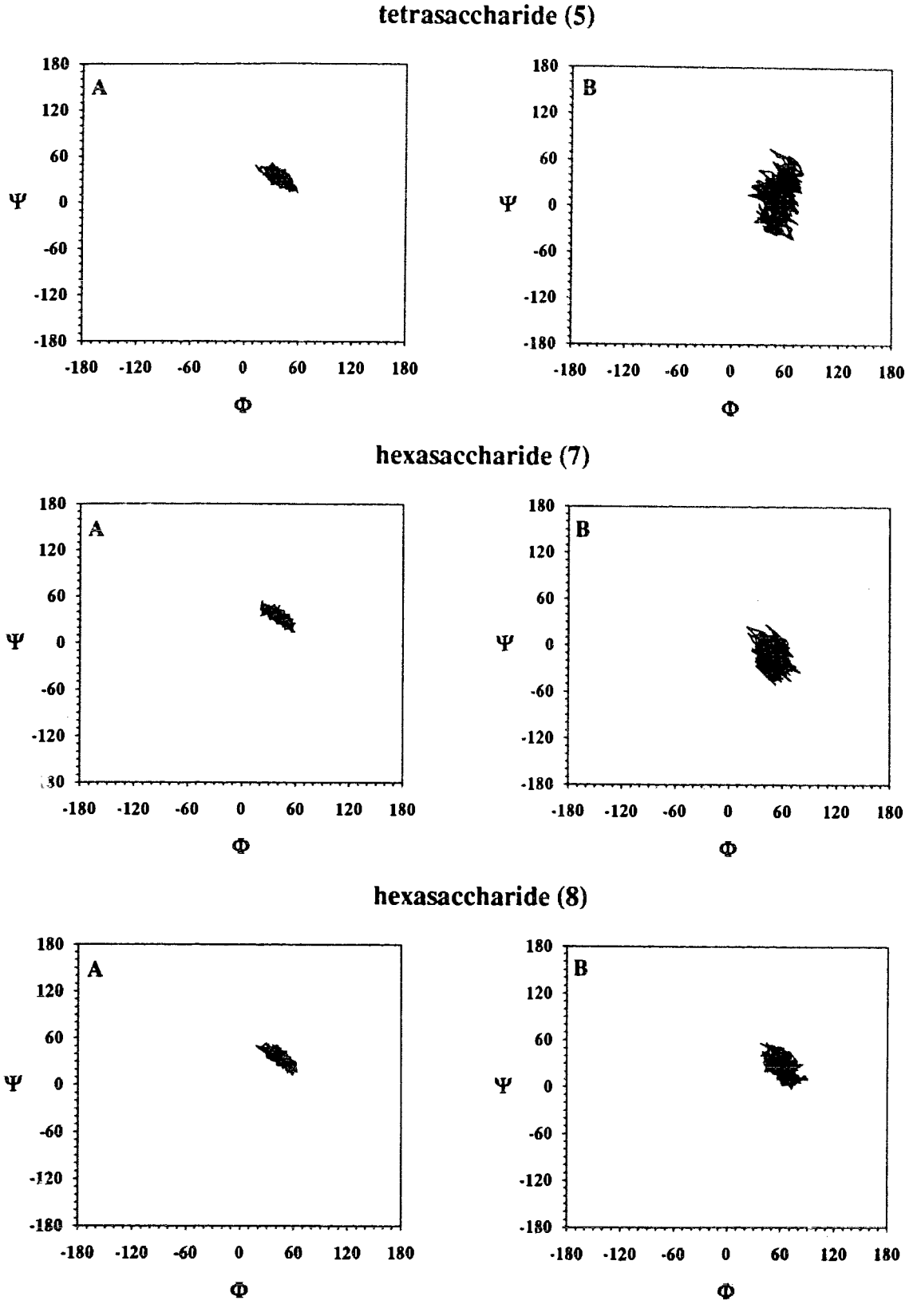
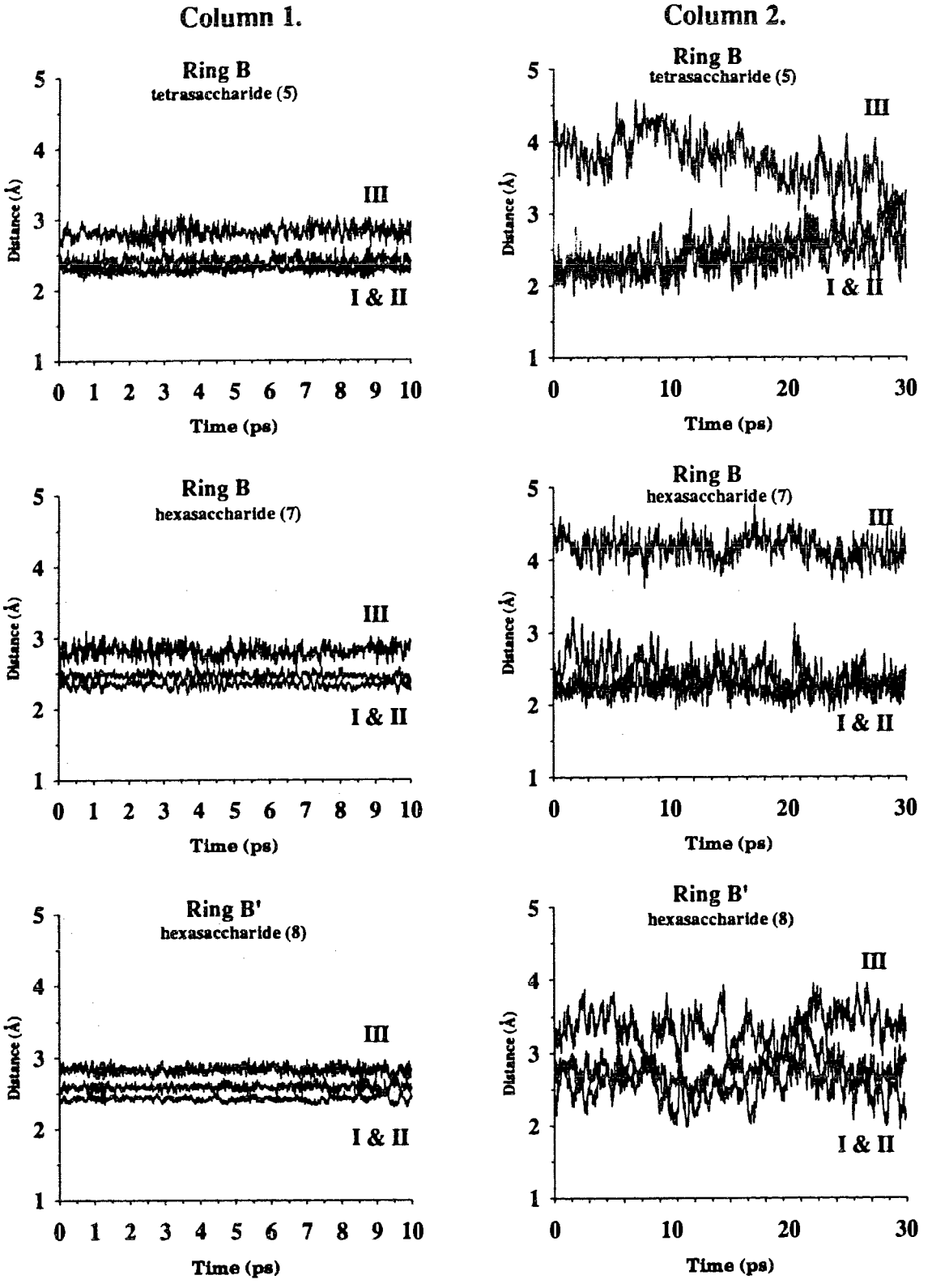


Figure 3.58. Variation of the Φ, Ψ angles of linkage B-A (B'-A' for compound **8**) during dynamics simulations of the compounds **5**, **7** and **8**. A.) Constrained dynamics using a 1B(')-4A(') constraint. B) Unconstrained dynamics.



LEGEND	
I)	H1B-H3A
II)	H1B-H5A'
III)	H1B-H4A

Figure 3.59. Variation of selected interproton distances during dynamics simulations of the compounds **5**, **7** and **8**.
Column 1. Constrained dynamics with a 1B(')-4A(') constraint.
Column 2. Unconstrained dynamics.

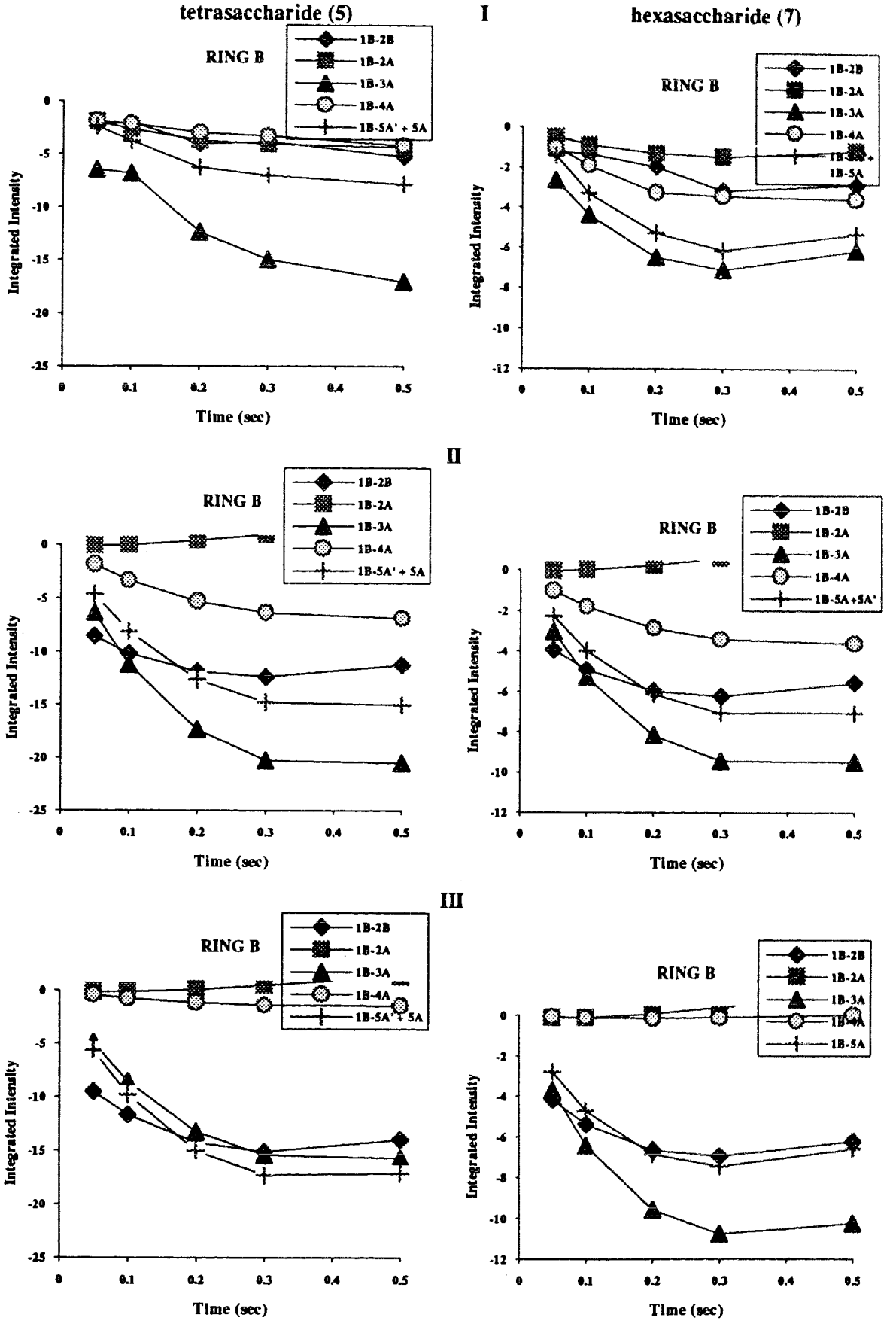


Figure 3.60. (Caption overleaf).

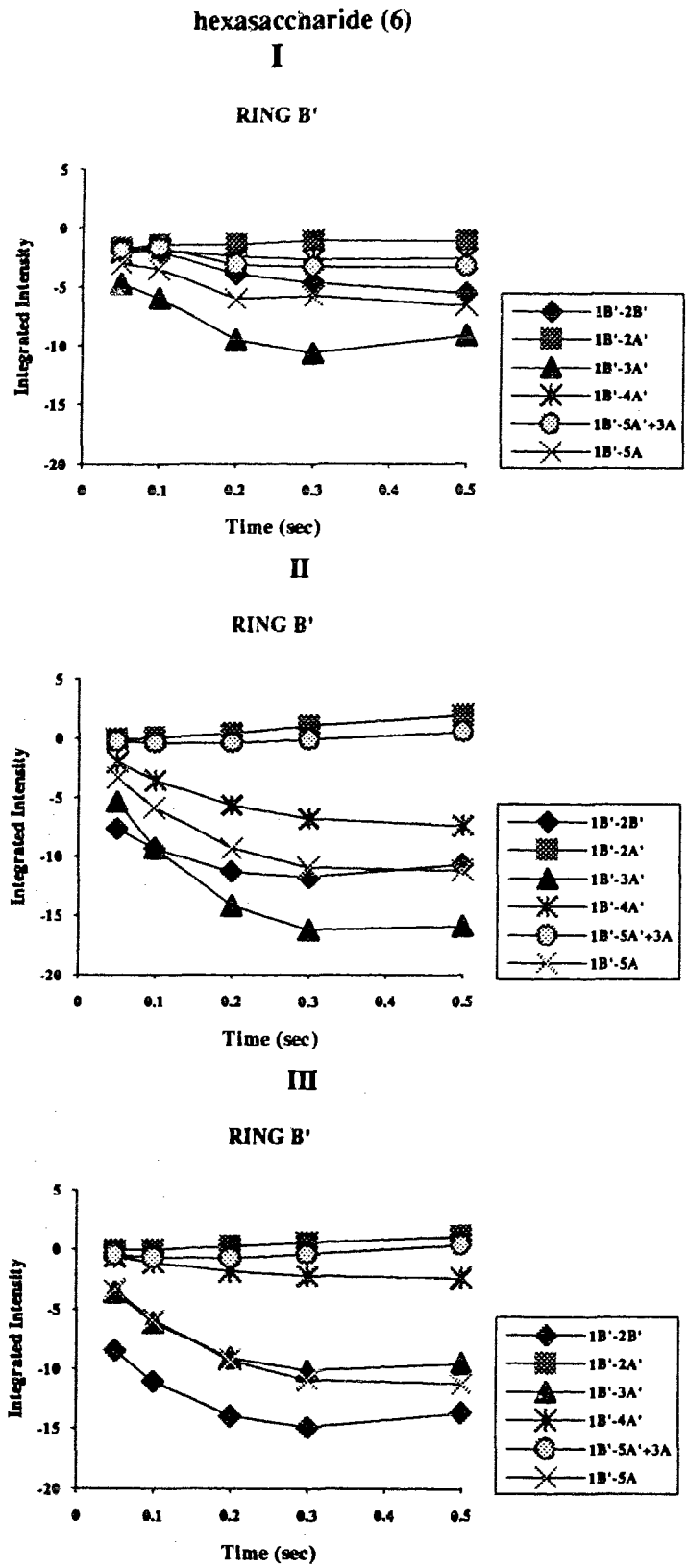


Figure 3.60. (This page and previous page) Experimental and calculated ROESY buildup curves for the B(B') ring of compounds **5**, **7** and **8**. I.) Experimental buildup curves. II) Buildup curves calculated from constrained dynamics trajectory. III) Buildup curves calculated from unconstrained dynamics trajectory.

Table III.25. Comparison of the average Φ and Ψ angles derived from constrained molecular dynamics simulations of the compounds **4-8**.

Link	A'-B		A-B'		B-A		B'-A'		C-B		C'-B'	
Compound	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ
4	39	16							44	-41		
5	51	24			39	33			38	-46		
6	22	13					41	28	44	-36	37	-49
7	47	25			41	34	39	31	33	-36	38	-45
8	28	24	45	24			47	34	40	-36	36	-43

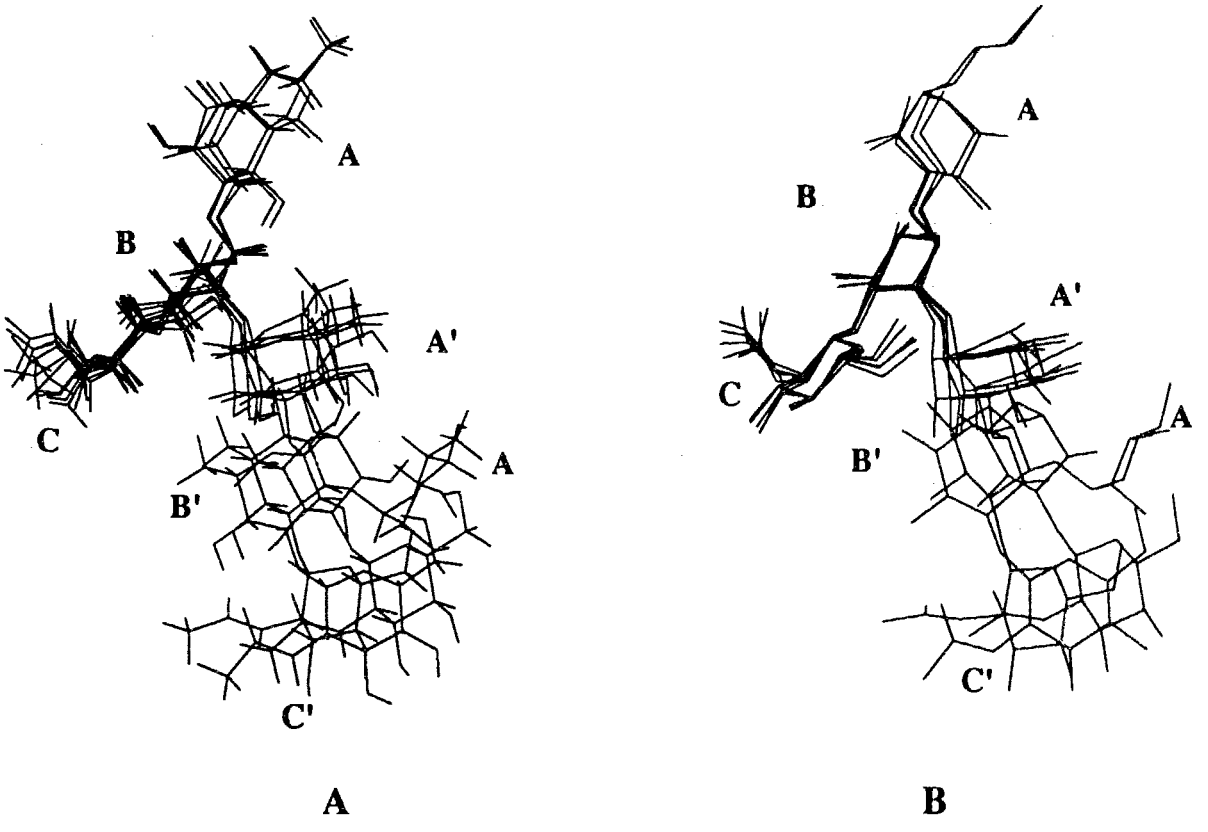


Figure 3.61. A superposition of the conformations derived from the average values of the constrained molecular dynamics simulations of compounds **4-8**. A) All atoms. B) All non hydrogen atoms.

Analysis of the CROSREL Calculations

The simulation of the ROESY buildup curves is faced with some severe limitations. Of these, the two most serious are a lack of experimentally determined correlation times, τ_c , and the associated assumption of isotropic reorientation. The obvious solution to this problem is the measurement of the ^{13}C T_1s, but given the limited quantities of sample available, this is at present not possible. Alternatively, long molecular dynamics simulations, of the order of nanoseconds, can be used to calculate τ_c. This method is faced with the limitations of computational time and disk space for long dynamics simulations, especially for large molecules. Further, since the dynamics trajectories are force field dependent, the calculation of the correlation time will also be influenced by the parameter set, and can produce varied results[176].

The calculated τ_c values and leakage rates R_L are tabulated in Table III.26. The predicted correlation time falls within the range that would be expected from the observed NOEs, i.e., that in the series of compounds studied, the product $\omega\tau_c \approx 1$ so that the NOEs range from positive to negative. The pentasaccharide (**6**), has a lower τ_c than

Table III.26. Correlation times, Leakage rates R_L, and R_w factors obtained from CROSREL calculations of the compounds **4-8**.

Compound	Dynamics (constrained)			Dynamics (no constraints)		
	τ_c (ps)	R_L (s^{-1})	R_w	τ_c (ps)	R_L (s^{-1})	R_w
trisaccharide (**4**)	192	0	.80	140	0	.60
tetrasaccharide (**5**)	400	10	.50	220	0	.60
tetrasaccharide (**5**)[a]	420	10	.45	400	0	.80
pentasaccharide (**6**)	340	0	.25	300	0	.31
hexasaccharide (**7**)	620	0	.38	630	0	.44
hexasaccharide (**7**)[a]	620	0	.38	600	0	.42
hexasaccharide (**8**)	700	0	.42	650	0	.38
hexasaccharide (**8**)[a]	650	0	.42	700	0	.44

[a]Dynamics simulations calculated after including the 1B(')-4A(') constraint.

the smaller tetrasaccharide which is an incorrect result, since **6** displays NOEs that are the same as the diagonal in a NOESY experiment, indicating a negative NOE, and hence a correlation time of > 400 ps.

It is difficult to estimate the error introduced by the assumption of a single τ_c. It is clear that the compounds **5-8** do not possess spherical symmetry. CROSREL has some provisions for the inclusion of anisotropic motion. This would also involve calculation of the τ_c from the dynamics data. The benefits from this approach are questionable, since it was shown that the anisotropic τ_c factors did not produce an appreciable improvement in the fit between the calculated and experimental buildup curves[176,177].

Restrictions imposed by overlap in the NMR spectrum of the compounds **4-8**, which prevents integration of the diagonal cross peaks, is another problem encountered in the simulation of the ROE buildup. The fit between the calculated and the experimental curves is also dependent on the diagonal intensities. CROSREL uses an **R** factor analogous to that used in X-Ray crystallography to judge the fit of the calculated and experimental data,

$$R_w = \left(\frac{\sum_{i,j,\tau_m} \left(w_{ij} * \left(A_{ij}^{obs} - A_{ij}^{calc} \right)^2 \right)}{\left(w_{ij} * \left(A_{ij}^{obs} \right)^2 \right)} \right)^{1/2} \qquad (3.5)$$

A_{ij} are the intensities of the cross peaks, w_{ij} is a weighting factor, and τ_m are the mixing times used in the experiment.

Inspection of the R_w factors calculated for the compounds **4-8** (Table III.26), reveals that the fit between the calculated and experimental buildup curves is not as good as the test case of methyl-β-cellobioside, in which the R_w factors were calculated to be of the order of 0.15[176,177]. However, in order to compare the calculated and observed intensities, CROSREL scales the calculated peak intensities before a fit is performed.

Thus, incomplete integration of the experimental data leads to incorrect scaling and hence a large R_w factor.

The ROESY buildup curves calculated with CROSREL are, at best, qualitative. Most of the experimentally observed trends are reproduced successfully. In the majority of the cases studied, the buildup curves calculated with the constrained dynamics trajectory provides a better fit to the experimental data than that calculated for the unconstrained dynamics. Where major differences are observed between the two calculated curves, i.e., the cross peaks calculated for the A' ring in the trisaccharide (**4**) and the tetrasaccharide (**5**) (Figure 3.15, Ring A'), the difference is due to the unconstrained dynamics simulation finding a new minimum.

Major discrepencies occur in the calculation of the distal diastereotopic propyl proton (designated HPr(a) in the text) for all the compounds where such a cross peak is calculated, and also in the calculation of the anomalous 1B-5A (1B'-5A') cross peak that is observed. Since both of these protons exhibit strong Hartman Hahn transfer, 5A to 4A and HPr(a) to HPr(b), this seems to suggest that the simulation of the Hartman Hahn transfer is not completely successful due to the inability to include the diagonal peaks.

CROSREL also fails to reproduce the observed ROESY-TOCSY type peaks, i.e., transfer of magnetization through a system *ijk*, where the protons *i* and *j* exhibit an ROE and protons *j* and *k* are J coupled. In system of this type, initially there is an enhancement of proton j through ROE effects, followed by transfer of magnetization through TOCSY effects, which is not easily distinguished from a true ROE since it exibits the same phase as the ROE. This effect is not reproduced in the CROSREL calculations. Simulation of this effect is crucial in the determination of the conformation of the B-A linkage. Thus, CROSREL, while representing a valuable tool for the analysis of ROESY buildup, cannot be used to distinguish between the two possible conformations in the B-A linkage.

Analysis of the Chemical Shift Differences ($\Delta\delta$)

Another aspect of the NMR spectra that can be analyzed to yield information about the conformation is the differences in chemical shift. The very basis of NMR spectroscopy is that the frequency of proton resonances differ due to their environment. ^{13}C and ^{1}H chemical shift studies [78,182-184] have been used to predict the conformation of saccharides with reasonable success. The ^{1}H and ^{13}C chemical shifts of the saccharides (**4-8**) are listed in Table III.1 and III.2, respectively. Correlation between the chemical shifts and the conformations of these compounds can best be accomplished by comparison with a parent compound. In this way, barring any chemical modification, differences in chemical shift can be interpreted directly as a difference in conformation. Since it represents the minimal repeating unit, the trisaccharide (**4**) is chosen as the parent, and the ^{1}H and ^{13}C chemical shifts of the A and A' rings of the compounds **5-8** are compared to the corresponding ^{1}H and ^{13}C shifts of the A' ring in the trisaccharide (**4**). The $\Delta\delta$ values (Figures 3.62, 3.63) represent the difference between the chemical shift of the various oligosaccharides (**5-8**) and the trisaccharide (**4**), i.e. a negative value of $\Delta\delta$ indicates that the analogous proton in the trisaccharide has a higher chemical shift, i.e., is upfield from the corresponding proton in the oligosaccharide and is therefore shielded by comparison to it, and analogously, a positive $\Delta\delta$ value implies a lower chemical shift, and hence a deshielding in the trisaccharide. A similar comparison of the shifts in the B and C rings of **4** with those of the B(B') rings and the C(C') rings of **5-8** gives the information displayed in Figures 3.62 and 3.63.

Analysis of the ^{1}H chemical shift differences reveals two extreme cases. The A rings of both the tetrasaccharide (**5**) and the hexasaccharide (**7**) differ from the A' ring of the trisaccharide (**4**) since in both of these compounds the A ring is the terminal (reducing end) moeity. Comparison of the chemical shifts of both of these compounds shows substantial differences in both the ^{1}H and ^{13}C spectra, of the order of -0.4 ppm in

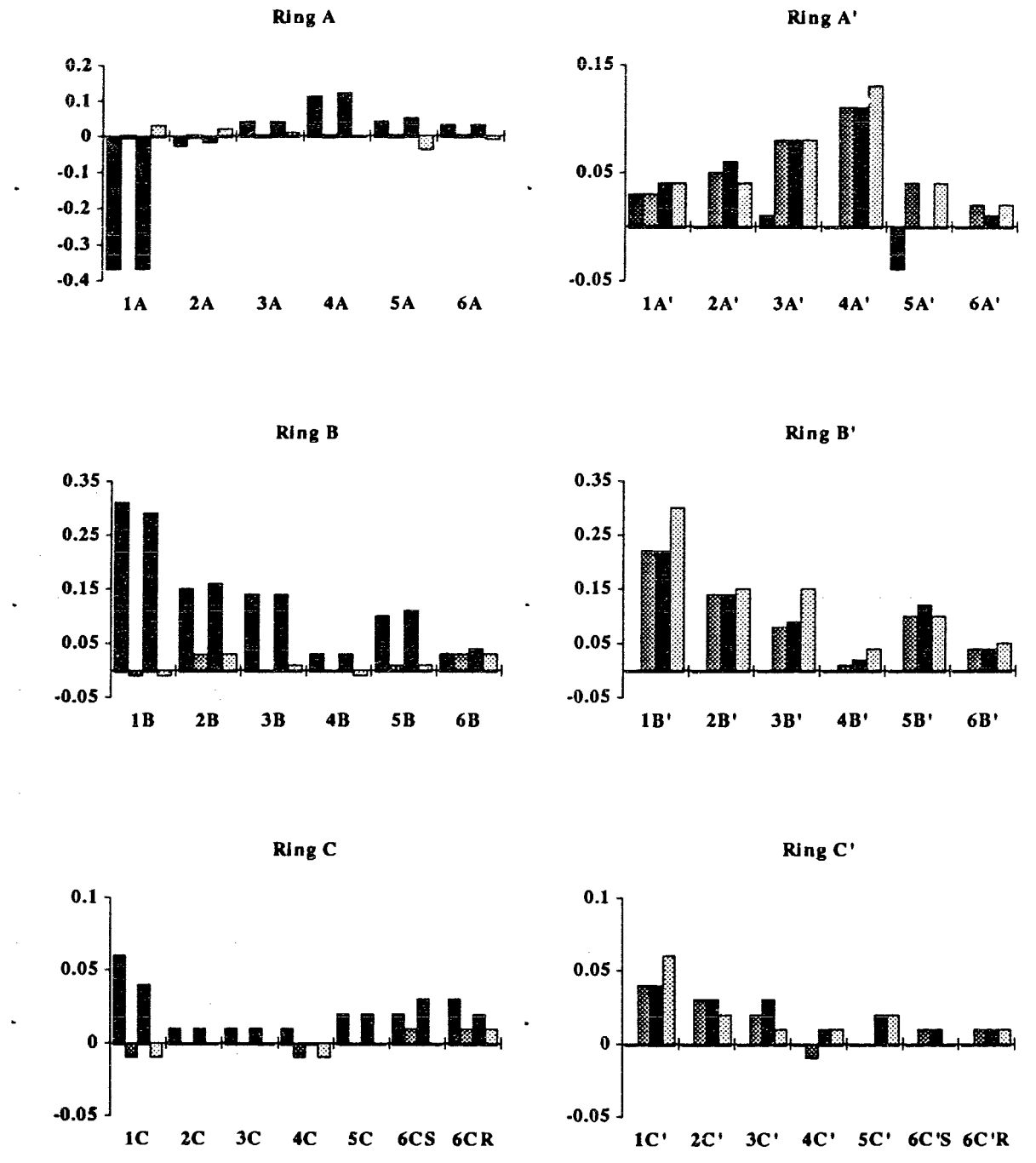


Figure 3.62. Differences in the ^{1}H chemical shift of the compounds **4-8**.

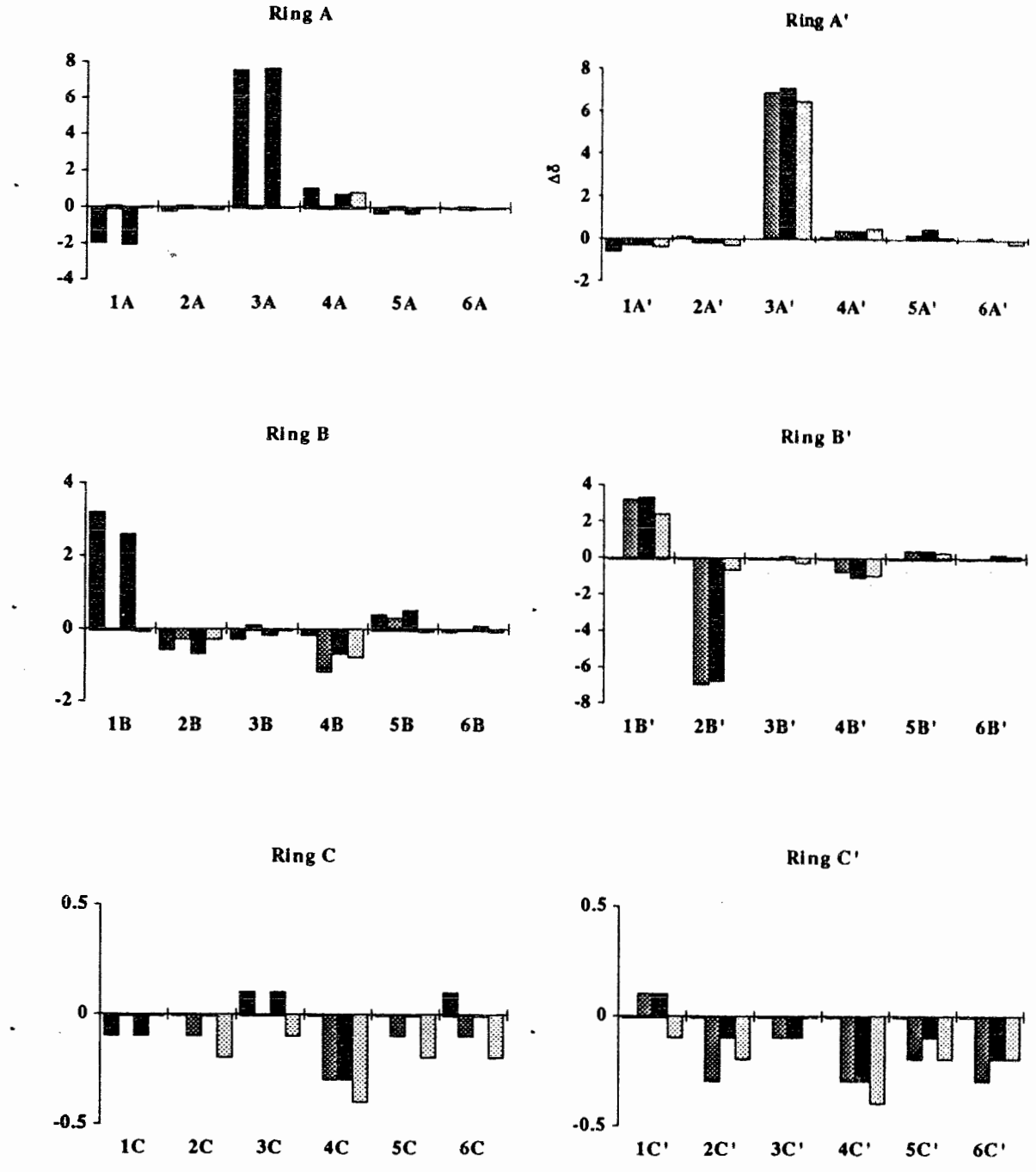


Figure 3.63. Differences in the ^{13}C chemical shift of the compounds **4-8**.

the proton shift differences and 2 ppm in the ^{13}C shift differences. The average values of the Φ and Ψ angles for the A'-B linkage listed in Table 3.25 are 38° and 17° respectively which places the anomeric proton of ring A' (and ring A in the hexasaccharide (**8**)) within a distance of < 2.5Å of the O1 and O5 of ring C (and ring C' in the hexasaccharide (**8**)). This would lead to a substantial deshielding of the 1A(') proton and would account for the large $\Delta\delta$ observed. A similar effect is also evident in the anomeric ^{1}H and ^{13}C chemical shifts of ring B in compounds **5** and **7**, in which $\Delta\delta$ is of the order of +0.3 ppm for the ^{1}H spectra and 2-3 ppm for the ^{13}C spectra, where the reversal of the sign is now due to the presence of an aglycon (rather than an absence of one, as in the previous example); similar effects are observed in the chemical shift of the H1 of the B' ring in compounds **6**, **7** and **8**. A positive Ψ angle in the B-A (and B'-A') linkage would place the H1 of ring B (B') within 2.5 Å of the O4 of ring A (A'). This would account for the deshielding observed. This is yet another piece of evidence in support of the positive Φ and Ψ angles for the B-A (and B'-A') linkage, since a negative Ψ angle would not bring the H1 of ring B (B') near any desheilding groups.

The second case where a large chemical shift difference is observed results from the effect of glycosylation at a certain position. The A' ring in the trisaccharide (**4**) and the tetrasaccharide (**5**) is unsubstituted, but in the three larger oligosaccharides **6**-**8**, it is glycosylated at the 3 position. The effects of this change can be seen in the ^{1}H $\Delta\delta$s, in which the ring protons of residues **6**, **7** and **8** are desheilded. This is consistent with the increased steric interactions that would occur upon substitution at the 2-position, and the effect is most pronounced on H4A', as would be expected if the Φ, Ψ angles were both positive. Although the ^{13}C chemical shift of C3 changes by ~8 ppm (glycosylation shift), there is no difference in the rest of the ^{13}C chemical shifts. This change in $\Delta\delta$ is also observed in the B' ring. The 2-position of the B ring of the trisaccharide, to which the chemical shifts of the B' ring are being compared, is substituted, whereas in the B'

ring the 2 position is unsubstituted. Since the comparison is now between a substituted (trisaccharide (**4**)) compound and compounds not functionalized at the 2-position, the difference is in the opposite direction for the ^{13}C shifts, i.e., of the order of -8 ppm. The ^{1}H chemical shift differences that are observed are a sum of the effects due to the A' ring and the glycosylation of the 2-position. Of these two, the effect of the A' ring can be judged to be predominant since the ^{1}H shifts of residue B' in the hexasaccharide (**8**), which is glycosylated at the 2-position, displays the same trends in $\Delta\delta$ as the B' residue in **5** and **7**.

The C and C' rings show very little variation in the chemical shifts. Most of the $\Delta\delta$ for the ^{1}H shifts are of the order of 0.02 ppm and the differences in ^{13}C shifts are ± 0.5 ppm, both within the range of experimental error. The largest deviation is the anomeric proton, which shows a difference of 0.05 ppm in residues which are attached to a B residue that is not the terminal aglycon. It is unclear why this trend is observed since the A (A') ring is not in the immediate vicinity of the C (C') ring.

In general, the comparison of the ^{1}H and ^{13}C shifts corroborates the conformations derived from the the ROESY data.

The ω Angle

The ω angle of hexapyranoses has long been a subject of discussion. The presence of a Glc*p*NAc residue in the compounds (**4-8**) necessitates both the determination of the ω angle in each Glc*p*NAc residue, and whether the overall conformation is influenced by this angle.

The ω angle of Glc*p*NAc in particular has been studied in detail. Newman projections of the the three rotamers about the ω angle are shown in Figure 3.64. In the solid state, crystal structures of D-glucose and its derivatives show a ratio of 60:40 of *gg*:*gt* conformers, while the *tg* conformer was not present[185] in any of the crystals studied. An additional parameter that can be used to determine the conformation of this

fragment is the proton-proton coupling constants, which follow a vicinal Karplus relationship[49]. In solution, the chemical shifts of the H-6proS (designated $H6_S$) and the H-6proR (designated $H6_R$) have been unambiguously assigned by selective deuteration of these protons[186,187]. Studies on D-glucose and D-mannose[188], both of which have an equatorial 4-OH, show that $H6_S$ is downfield of $H6_R$ (i.e. $\delta H6_S > \delta H6_R$) and that $J_{5,6S}$ ~2 Hz and $J_{5,6R}$ ~ 5.5 Hz. The 6_S and 6_R protons of Glc*p*NAc display the same trend in chemical shift and coupling constants, which indicates that the N-acetyl group does not influence the behaviour of the ω angle. There have been many studies aimed at the calculation of the relative populations of conformations about the ω angle and the observed coupling constants calculated with the Karplus equation[49]. Due to the nature of the ω angle, i.e. the two oxygen atoms involved in the torsion, and with consideration of the effect of electronegative substituents on the observed coupling constants, a modified Karplus equation is generally used[55,189]. An example of this type of equation is[55]

$$^3J_{HH} = P_1\cos^2\phi + P_2\cos\phi + P_3 + \sum\Delta\chi_i\left\{P_4 + P_5\cos^2(\xi_i\phi + P_6|\Delta\chi_i|)\right\} \qquad (3.7)$$

P_i are empirical parameters that are determined by a least squares fit of the equation to experimental data, χ is the difference between the electronegativity of the substituent and hydrogen, ξ is ±1, depending on the torsion angle of the substituent, and ϕ is the torsion angle between the two protons for which the coupling is being calculated.

The $J_{5,6S}$ and $J_{5,6R}$ values calculated for each conformer with equation 3.7 are displayed in Figure 3.64. The measured coupling constants are not consistent with any one conformation, and although the J values are calculated for the optimum torsion angles, i.e., *gt* = 60°, *gg* = -60° and *tg* = 180°, and may not be representative of the staggered geometry in solution (where gg ≠ exactly -60°), it can be assumed that the observed coupling constants are actually averaged over the three different conformations.

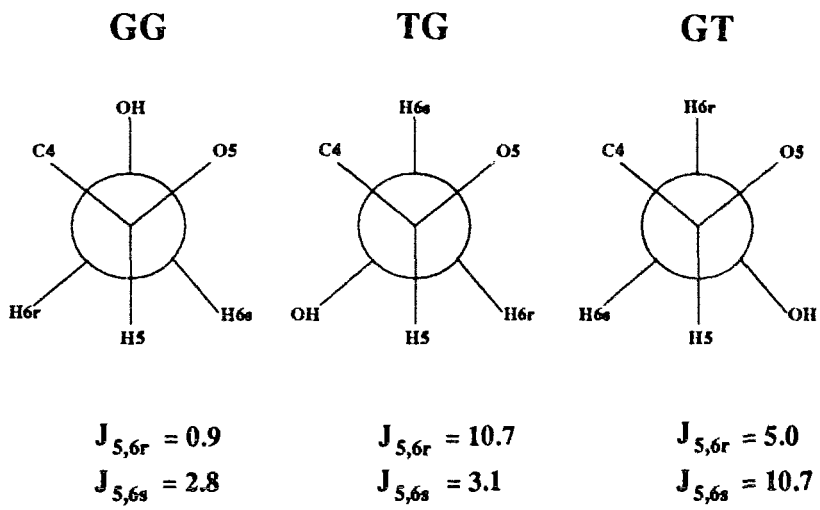


Figure 3.64. The three staggered rotamers about the C5-C6 bond and their calculated coupling constants (Hz).

Attempts to calculate the percentage distribution of the *gg*, *gt*, *tg* conformations for Glc*p*NAc that would exhibit the observed J values have met with limited success since each of these studies assigns a negative value to the population of the *tg* conformer to account for the observed coupling constants[190-192]. A successful molecular dynamics simulation should be able to reproduce the population distribution about the ω angle and hence the observed values of $^3J_{HH}$. Analysis of the dynamics trajectories of compounds **4-8** shows that ω angles of the compounds **4-8** are generally present in the *gt* conformation, with transitions to the *gg* conformation (Figure 3.65). The molecular dynamics trajectories were used to calculate the average $^3J_{HH}$ for the 5,6 proton pairs, and the results are displayed in Table III.27. The results do not match the experimentally observed $^3J_{HH}$ for any of the structures studied.

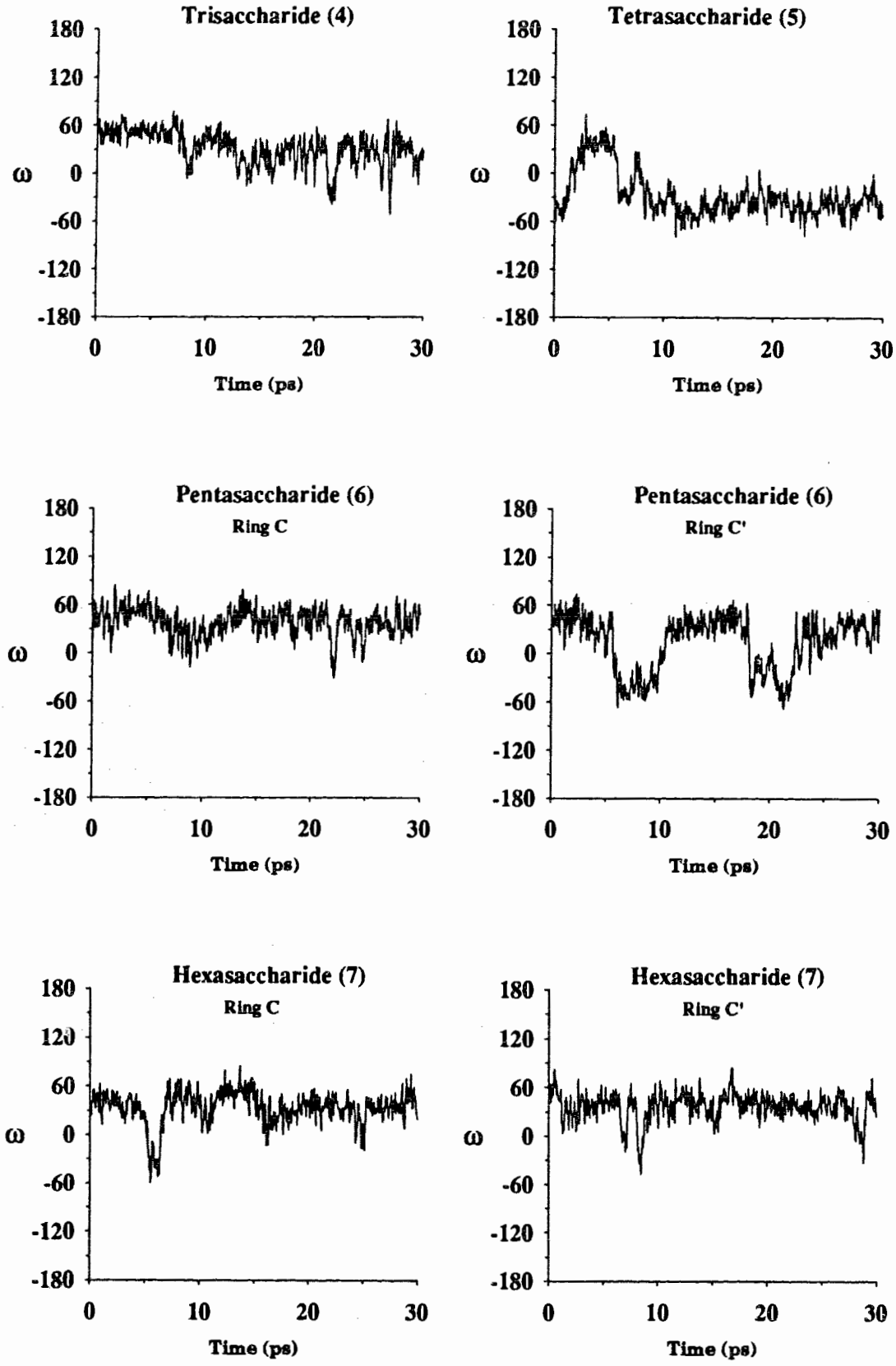


Figure 3.65. (continued overleaf).

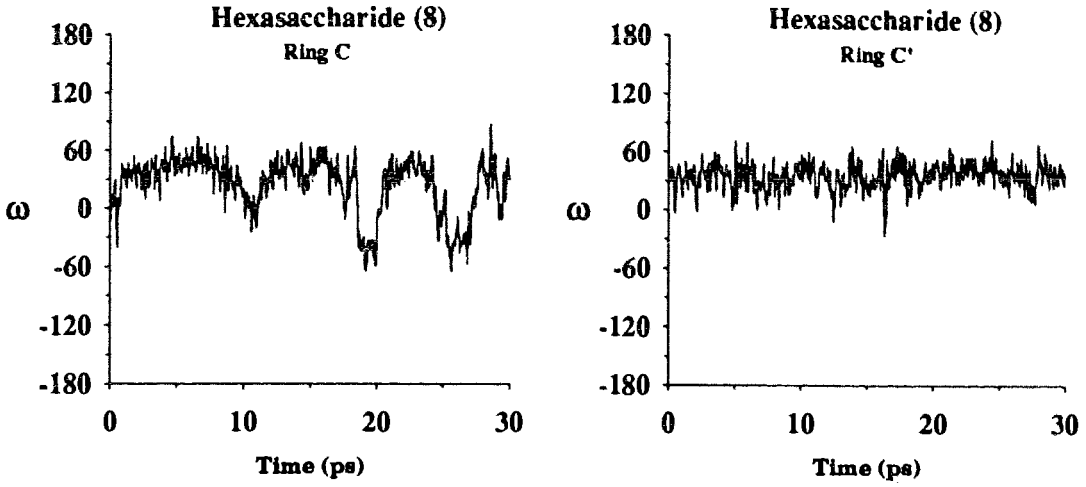


Figure 3.65. Trajectories of the ω angles of compounds **4-8** during dynamics simulations without constraints.

Table III.26. $J_{5,6}$ Values (Hz) calculated from the dynamics trajectories of compounds **4-8.**

Compound		$J_{5,6R}$	$J_{5,6S}$
trisaccharide (4)		9.96	6.09
tetrasaccharide (5)		2.57	2.32
pentasaccharide (6)	(C)	10.89	5.13
	(C')	8.27	6.12
hexasaccharide (7)	(C)	10.4	5.8
	(C')	10.7	5.5
hexasaccharide (8)	(C)	9.2	6.1
	(C')	10.9	5.9

The ROESY spectra of the compounds **4-8** display, without exception, a cross peak between the $H6_S$ and the H5 or H4 proton of the Glc*p*NAc residue. An expansion of the ROESY spectra of this region for compounds **4-8** is displayed in Figure 3.66. Due to the overlap of the H5 and H4 chemical shift, it is not possible to determine to which of these two protons the contact is observed. No cross peak is seen from the $H6_R$ proton. A NOESY spectrum of **8** also displays the same results, which indicates that this is not an artifact of the ROESY spectra due to Hartmann Hahn transfer of magnetization from $H6_R$ to $H6_S$. The ROE/NOE result points toward a predominantly *gt* conformation with the $H6_R$ almost eclipsed with the ring oxygen, i.e. a ω angle of ~90-100. This would place the $H6_S$ between the H5 and H4 protons and would account for both the ROE observed and the observed $J5,6_R$ and $J5,6_S$ coupling constants.

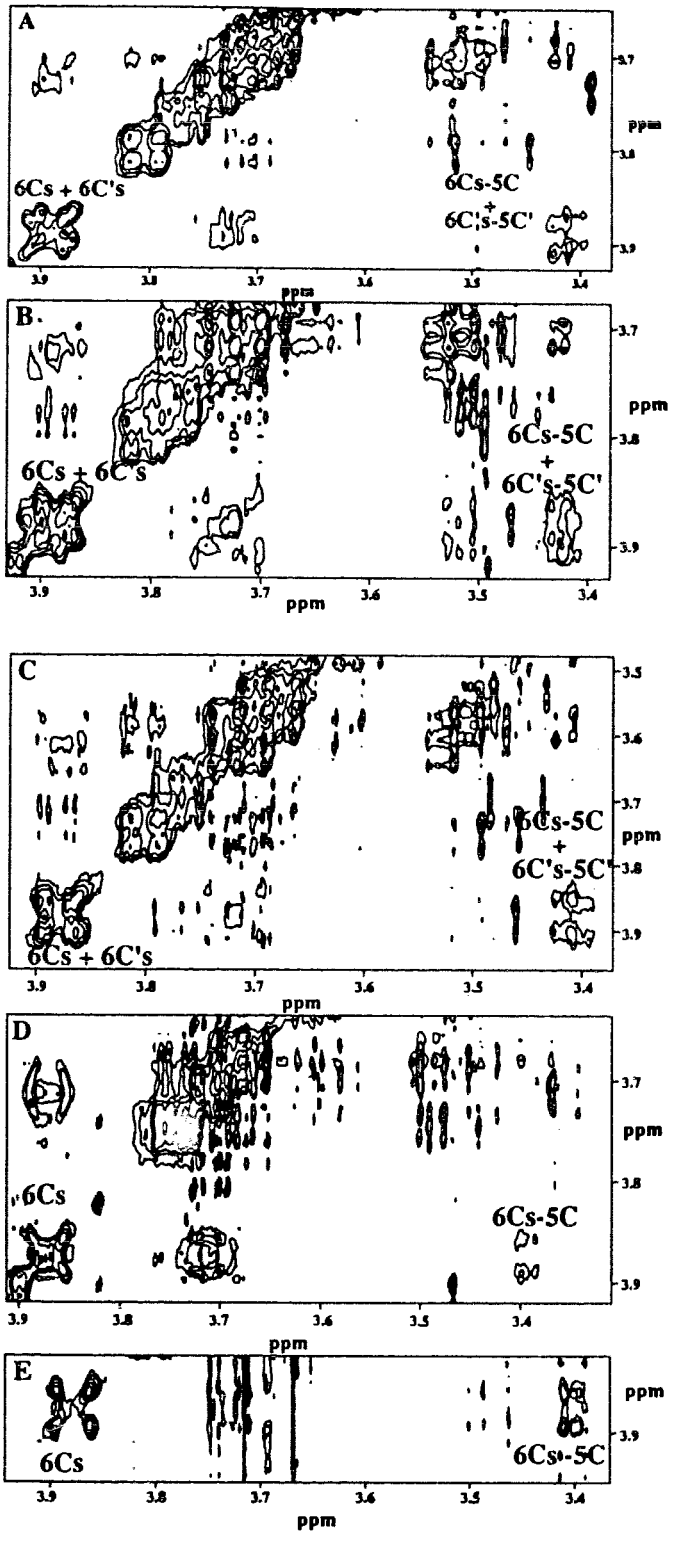


Figure 3.66. Expansion of the ROESY spectra of compounds **4-8**. A) Hexasaccharide (**8**). B) Hexasaccharide (**7**). C) Pentasaccharide (**6**). D) Tetrasaccharide (**5**). E) Trisaccharide (**4**).

Analysis of $^3J_{COCH}$ values

The utility of three bond coupling constant ($^3J_{CH}$) values across the glycosidic linkage for conformational analysis was discussed in Chapter I. As a final point of comparison, the $^3J_{CH}$ values of the hexasaccharide (**8**) were measured by both direct observation and inverse detected ^{13}C-^1H correlations, optimized for the long range coupling. The results of these experiments are tabulated in Table III.28 along with the $^3J_{CH}$ values calculated from the dynamics trajectories.

Table III.27 Experimental and calculated $^3J_{CH}$ values (Hz) for the hexasaccharide (**8**).

Linkage	Contact	Exp. ±0.2 Hz	Dynamics (constrained)	Dynamics (no constraints)
(A'-B)	H1A'-C2B (Φ)	4.5	4.3	4.9
	C1A'-H2B (Ψ)	5.1	4.6	3.8
(C-B)	H1C-C3B (Φ)	4.5	3.3	1.0
	C1C-H3B (Ψ)	-	3.7	3.1
(B'-A')	H1B'-C3A' (Φ)	5.0	2.7	1.8
	C1B'-H3A' (Ψ)	5.3	3.9	4.5
(A-B')	H1A-C2B (Φ)	4.0	2.9	1.9
	C1A-H2B' (Ψ)	5.1	4.6	3.8
(C'-B')	H1C'-C3B' (Φ)	4.5	3.7	4.3
	C1C'-H3B' (Ψ)	-	3.1	3.7

The calculated coupling constants show a good correspondence with the observed values, especially with those values calculated from the constrained dynamics trajectories. The difference in the values calculated for the A-B and A'-B' illustrates the large variation in J values associated with even a small change in the Φ, Ψ values - the average phi values for the A-B and A'-B' link are 26 and 42 respectively and their corresponding J values are calculated to be 4.3 Hz and 2.9 Hz, a difference of almost 40 percent.

Conformation of the Polysaccharide

The *Streptococcus* Group A polysaccharide has been characterized by two dimensional NMR methods that have yielded the complete assignment of the ^{1}H NMR spectrum[193] (Table III.28). The ^{1}H NMR spectrum of the polysaccharide displays three distinct anomeric protons which are assigned to the A, B and C residues on the basis of their chemical shift and connectivity as determined by COSY and NOESY experiments. It is of interest to compare the native polysaccharide to the fragment oligosaccharides **4-8** to determine if there are any common features and thus infer some conformational properties of the polysaccharide.

The NOESY spectrum of the polysaccharide (Figure 3, Reference 193) displays the A-B and B-A intra ring cross peaks, but no NOE between the 1C and 3B protons is visible. From the limited data available, it is not possible to comment on whether this is due to a lack of this NOE or that the experimental parameters (mixing time, temperature) did not allow for its observation.

Table III.28. ^{1}H NMR chemical shifts (ppm) of the *Streptococcus* Group A Polysaccharide[193] .

	Unit		
^{1}H	**B**	**A**	**C**
1	5.09	5.17	4.75
2	4.27	4.07	3.72
3	4.00	3.85	3.57
4	3.52	3.56	3.46
5	3.81	3.73	3.46
6	1.31	1.27	3.94
6R			3.77

A comparison of the ^{1}H chemical shifts, similar to the comparison performed above for the various residues, is displayed in Figure 3.66. In order to account for differences in the spectra due to the standard (sodium 4,4-di-methyl-4-sila-[2,2,3,3,$^2H_4$]-

pentanoate) and temperature (55°C) used in the study of the polysaccharide, the difference in the chemical shift between H1B' of **8** and H1B of the polysaccharide was added to the chemical shift of all the compounds **4-8**. This does not guarantee the equivalence of the spectra, since the choice of a standard is arbitrary. However, the environment of the B' residue in the compound **8** should best approximate that in the polysaccharide. Another factor to be considered in the comparison is the method of assignment. The assignments of the polysaccharide were made using standard COSY and Relayed COSY experiments which, in the case of highly overlapped regions will not provide the accuracy afforded by the TOCSY experiments employed in our study. For this reason, only chemical shift differences > 0.05 ppm are considered to be significant. The chemical shift of the compounds **4-8** is subtracted from that of the polysaccharide, and thus, a positive $\Delta\delta$ in Figure 3.36 represents an upfield shift of the proton in question, which in turn implies a specific shielding. Similarly , a negative $\Delta\delta$ represents a downfield shift, which indicates that the oligosaccharide proton is deshielded with respect to the analogous proton in the polysaccharide.

To facilitate this analysis, a 24-mer was constructed from the average Φ, Ψ angles of the constrained dynamics trajectories of compounds **4-8**, which, if the extrapolation of conformational properties is valid, should approximate a part of the surface of the polymer. The 24-mer thus constructed (Figure 3.67) forms a well defined helix with an internal diameter of ~13 Å (as measured from two atoms chosen arbitrarily across the ring) and a pitch of ~10 Å (from C5 of one Glc*p*NAc residue to the C5 of the next Glc*p*NAc residue), with the Glc*p*NAc residues almost perfectly aligned on the outside of the helix.

Analysis of the $\Delta\delta$s reveals that besides the expected shielding of the anomeric proton of specific residues due to the absence of an aglycon, i.e. the H1A in compounds **5** and **7**, and the H1B in compounds **4**, **6** and **8**, there are a few protons with a significant

(> 0.05 ppm) Δδ. These include the 2H, 3H and 5H of the B residue in the compounds **4**, **5** and **8**, i.e. those without the A ring as an aglycon. The 3H and 5H of the B residue are on the face of the ring in direct contact with the A residue, and the change in chemical shifts, i.e., a deshielding of the B residue upon addition of the A ring as an aglycon due to increased steric contacts, is consistent with the postulated relationship between Δδ and steric interactions[194-196]. This analysis can be applied to explain the observed Δδ in the H4 of the A' ring, and the observed Δδ of the 2B proton may be ascribed to the proximal 4 hydroxyl of the A ring. The model does not account for the observed Δδ between the 1C proton of the oligosaccharides, nor the observed shifts in the H3B' of the hexasaccharide (**7**) and pentasaccharide (**6**). Both of these protons are in relatively conserved regions of the oligosaccharides.

The oligosaccharides **4-8** have been used in inhibition studies to characterize monclonal and polyclonal antibodies raised against the glycoconjugates of selected synthetic oligosaccharides as well as the native polysaccharide[172]. These studies show that :

a) The branch point, i.e., the α-L-Rha*p*-(1→2)[β-D-Glc*p*NAc-(1→3)]-α-L-Rha*p* element appears to be an essential component for the recognition of the synthetic oligosaccharides by both the polyclonal and monoclonal antibodies.

b) The hexasaccharide (**8**), which is comprised of two of the branch points, appears to be the best inhibitor of the antibodies.

These findings are consistent with the model of the surface subtended by the polysaccharide as indicated by the 24-mer. This model indicates that the Glc*p*NAc residue is exposed on the periphery of the helix, the helix providing a platform. This feature is likely to have an important influence on its recognition by the immune system.

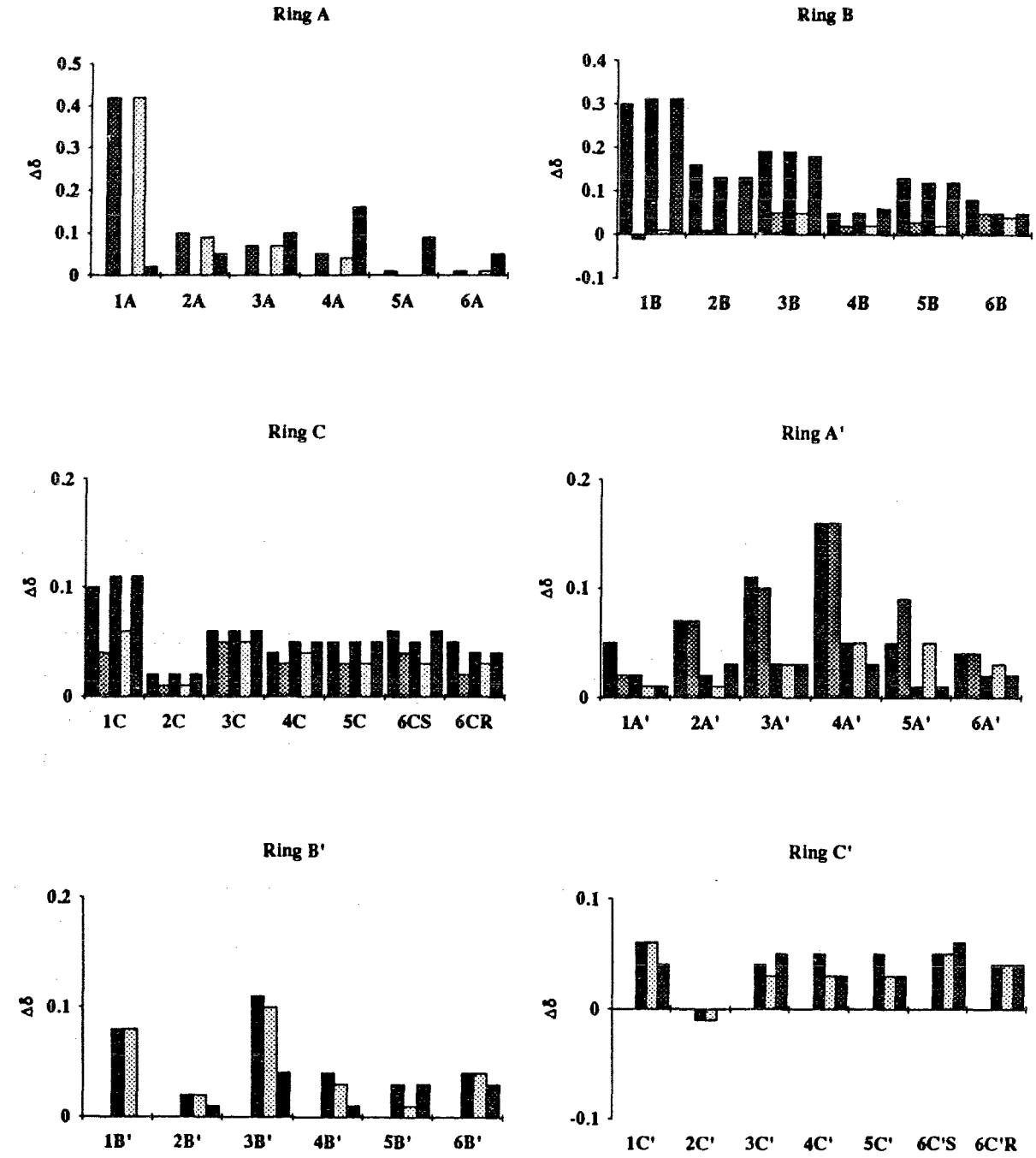


Figure 3.66. Differences in the ^{1}H chemical shift between the ring protons of the polysaccharide and the compounds **4-8**.

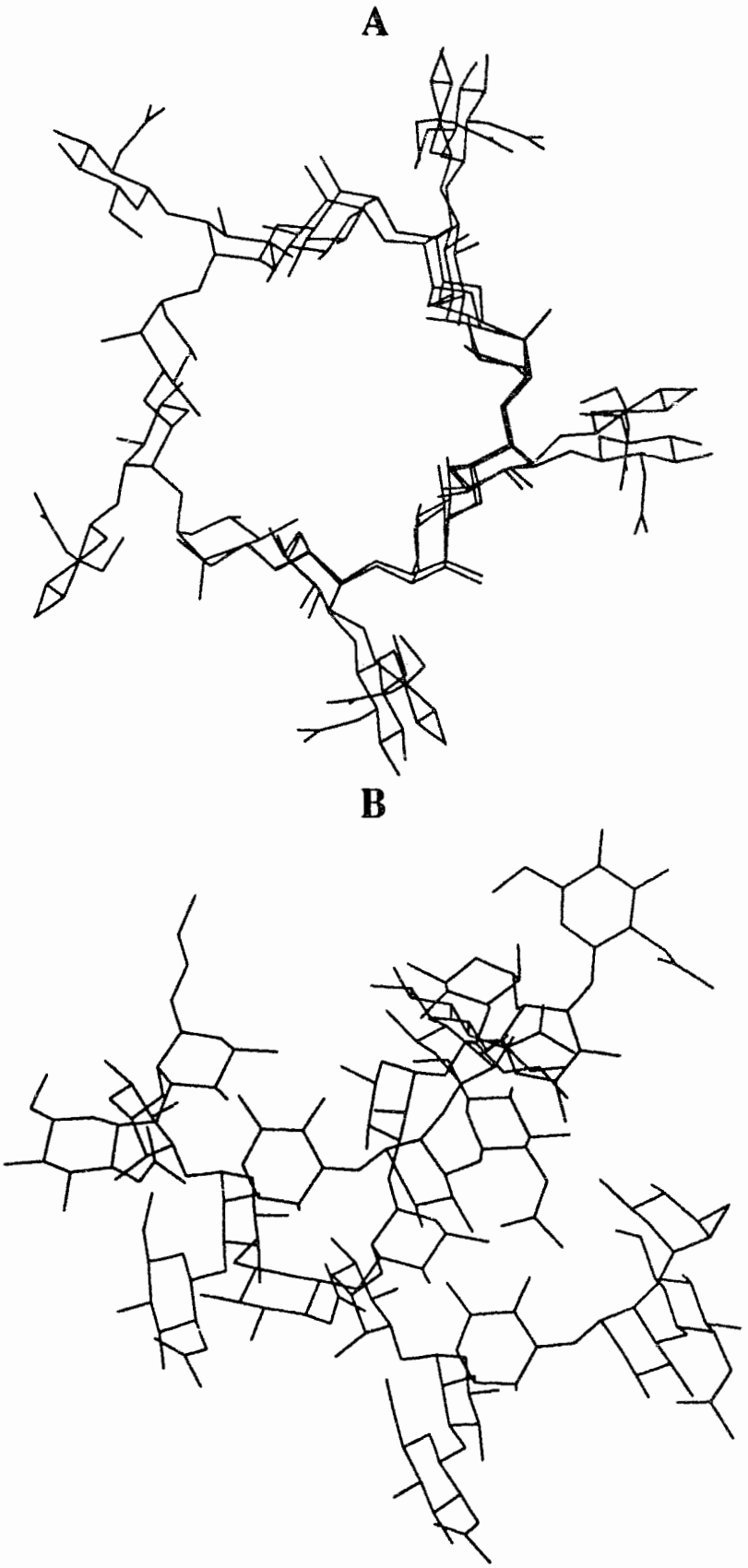


Figure 3.67. 24-mer constructed from the average Φ, Ψ angles obtained from the constrained molecular dynamics simulations of **4-8**. A) Top view. B) Side view.

Chapter IV

CONFORMATIONAL ANALYSIS OF A HEPTASACCHARIDE FRAGMENT CORRESPONDING TO THE *SHIGELLA FLEXNERI* VARIANT Y LIPOPOLYSACCHARIDE O-ANTIGEN

4.1. Introduction

In the previous Chapter it was mentioned that one aspect of the research in our group has been synthesis and characterization of oligosaccharide fragments of the *Streptococcus* Group A cell-wall polysaccharide. Similar studies are also being conducted on another virulent bacterium, *Shigella flexneri* variant Y. Strains of *Shigella* are responsible for bacillary dysentery and infection can be particularly severe in infants. The majority of *Shigella* strains show a resistance to antibiotics and hence an effective vaccine would provide a simple and cost effective preventive treatment.

The *Shigella flexneri* variant Y O-antigen is a linear polysaccharide[198,199] that is composed of rhamnose units linked α-L-(1→3) and α-L-(1→2), interspersed by N-acetyl-β-D-glucosamine (β-D-Glc*p*NAc) to form a periodic repeating unit ABCD as is illustrated in Figure 4.1.

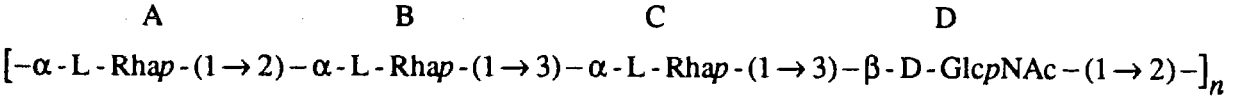


Figure 4.1. The *Shigella flexneri* variant Y lipopolysaccharide O-antigen.

The solution conformation of the *Shigella flexneri* variant Y O-antigen was first probed by Bock *et. al.* [77]. Utilizing HSEA calculations in conjunction with NMR experiments, they studied ten different di-, tri-, and tetrasaccharide fragments of the polysaccharide. A comparison of the chemical shifts of all of these compounds revealed

a pattern of shielding and deshielding effects. The minimum energy structure predicted by the HSEA calculations was used as a model to rationalize this phenomena by examination of the specific interactions that occur as a result of substitutions on the various rings. Most of the experimentally observed phenomenon could be explained using the HSEA minimum on the basis of proximity of the proton in question to neighboring hydroxyl or ring oxygens. Inter-residue NOEs were also observed and the results were used as one more point of comparison. The observed NOEs matched those that would be expected in the calculated structure.

Research efforts directed toward the synthesis of oligosaccharide fragments have yielded the pentasaccharide[200] ABCDA', as its methyl glycoside as well as hexasaccharide[201] ABCDA'B' and heptasaccharide[202] ABCDA'B'C' segments of the polysaccharide as their propyl glycosides. A few of these haptens have been used in inhibition assays with monoclonal antibodies raised against the polysaccharide in order to map the combining site of the antibodies[203].

Table IV.1. Relative free energy of binding[203] of various inhibitors to the monoclonal antibody SYA/J-6.

Oligosaccharide	Δ(ΔG) kcal/mol
ABCDA'	-0.1
ABCD	0.0
BCDA'	0.1
BCD	0.6
CDA'	1.8
ABC	does not bind

The absence of a significant difference in the binding energies between the pentasaccharide ABCDA' and the tetrasaccharide ABCD indicate that the tetrasaccharide is most likely the unit recognized by the antibody. Further studies with specifically

modified saccharides have been performed to further elucidate the structural features involved in the binding[204].

These experimental data provide a good opportunity to compare the conformational properties of oligosaccharides as determined by NMR and molecular dynamics calculations and to rationalize the observed trends in binding affinities on the basis of key conformational features. The heptasaccharide (**9**), was chosen as the structure upon which to do the modeling since it is just one unit short of twice the repeating unit and thus possesses the features of all combinations of trisaccharide and tetrasaccharide fragments. Determination of the three-dimensional structure of the heptasaccharide will realize one aspect of the objectives of the research project. i.e., elucidation of some of the characteristic topographical properties that influence the binding.

4.2. Research Objectives

At the onset of this study, preliminary analysis of the heptasaccharide (**9**) had begun, and assignment of the ^{1}H and ^{13}C NMR spectra was completed[167]. The chemical shifts are presented in Table IV.2. In order to complete the assignment of the ^{1}H and ^{13}C spectrum COSY, inverse correlated ^{1}H-^{13}C-COSY, TOCSY and ROESY experiments were performed. Besides providing the information necessary for the assignment, the ROESY spectrum also provided some preliminary information about the interresidue contacts observed.

This chapter will apply the methods discussed in the previous chapter to the conformational analysis of the heptasaccharide (**9**), taking the preliminary analysis one step further. The prominent conformational features of **9** will be rationalized in terms of the observed binding affinities of other oligosaccharide fragments. Lastly, the conformation of the heptasaccharide will be used to infer the conformation of a larger

oligosaccharide in an attempt to determine the conformational properties of the native polysaccharide.

4.3. Experimental

The experimental methods are detailed in Chapter III.

Table IV.2. ^{1}H and ^{13}C NMR data[9] for the heptasaccharide (9)[a].

	^{1}H	^{13}C		^{1}H	^{13}C
Ring			Ring		
1C'	4.74	102.2	1C	4.82	103.9
2C'	3.96	71.7	2C	3.81	72.7
3C'	3.76	80.2	3C	3.75	79.9
4C'	3.50	74.5	4C	3.50	74.3
5C'	3.71	71.0	5C	3.99	71.8
6C'	1.26	19.3	6C	1.20	19.1
1B'	5.15	103.4	1B	5.13	103.5
2B'	4.03	80.9	2B	4.02	80.7
3B'	3.91	72.6	3B	3.88	72.6
4B'	3.43	74.8	4B	3.45	74.8
5B'	3.77	71.3	5B	3.72	71.3
6B'	1.20	19.3	6B	1.29	19.3
1A'	5.11	103.7	1A	4.93	105.0
2A'	4.12	81.4	2A	4.05	72.7
3A'	3.83	72.3	3A	3.76	73.3
4A'	3.28	75.0	4A	3.40	74.7
5A'	3.64	71.8	5A	3.67	71.8
6A'	1.20	19.2	6A	1.23	19.5
1D	4.69	104.8			
2D	3.81	58.4			
3D	3.58	84.2			
4D	3.50	71.7			
5D	3.40	78.6			
6DR	3.71	60.1			
6DS	3.86				

[a]Chemical shifts were measured in D_2O at room temperature.

4.3. Results

4.3.1. ROESY Spectra

The anomeric region of the ROESY spectrum of the heptasaccharide (**9**) obtained with a 500ms spin lock is expanded in Figure 4.2, and the corresponding F1 slices are shown in Figure 4.3. There are a subtantial number of folded in peaks, which can be identified by their out of phase character. This is a result of narrowing the sweep width to include only the resonances of the ring protons and excluding the resonances of the propyl aglycone, the methyl groups of the rhamnose rings, and the methyl group of the N-acetyl group, thereby increasing the resolution of the 2D spectrum. The folded in peaks, while complicating the spectrum, do not interfere with the important cross peaks, and the enhanced resolution thus obtained greatly facilitates assignment of the cross peaks.

The ROESY spectrum of **9** shows some of the features observed in the oligosaccharides **4-8**. Both of the α–(1→2) linkages i.e., A-B and A'-B', display the characteristic 1-2 and 5-1 cross peaks, indicating a similarity in conformation. There are two α-(1→3) linkages between the residues B-C and B'-C'. Cross peaks are observed between the 1B' proton and the 2C', 3C', 4C', 5C' protons. All of these peaks have been observed for the same linkage in compounds **4-8**. The B-C linkage, however displays a much more intense cross peak to the 2C proton, and the 1B-5C cross peak is absent. A cross peak corresponding to a 1D-2A' interaction is visible, and there are two cross peaks across the C-D linkage, a 1C-3D and 1C-4D cross peak. All of these cross peaks were identified in preliminary work with the exception of the 1C-4D cross peak, which though present, was unassigned.

In addition to the series of spectra acquired with a minimal sweep width, ROESY spectra were also recorded using a sweep width of 10 ppm, which spans the entire spectrum of the heptasaccharide. Two cross peaks were observed (Figure 4.2 B) from

the methyl group of the N-acetyl moeity, and these were identified as to the H2 of the C ring and to the H4 of the A' ring.

4.3.2. Quantitation and Correction of ROESY Data

As a result of the high degree of overlap in the NMR spectrum of **9**, Hartmann Hahn and offset corrections are difficult to apply. It is been shown in Chapter III that due to the placement of the spin lock and the narrow range of sweep widths involved, offset correction is not critical and thus in the case of overlapping peaks, the average value of the correction is applied.

This sort of approximation cannot be made when corrections for Hartmann Hahn, effects are made, especially in cases where the integrated intensity of one of the overlapped peaks is to be used as the ruler distance, as in the case of the 1A-2A cross peak. The 2A resonance is overlapped with the 2B resonance, and both of these protons display a cross peak to the anomeric proton of the A ring, which results in the peaks being almost superimposed.

Improper correction of this peak can lead to errors in the calculation of the distances associated with the 1A resonance. In order to avoid this, the following procedure is adopted to deconvolute the two peaks. The corresponding F1 slice is read, and the overlapped peaks in the 1D spectrum thus obtained is deconvoluted using a lorentzian fit, and the area of each of these deconvoluted peaks was then used to calculate the relative integral of the cross peak. In the case of the 1C-2C peak, however, no correction of this type can be applied. The baseline around this peak shows traces of both a 1C-2D and a 1C-3C cross peak, both of which are attributable to TOCSY transfer. Parts of these peaks are completely enveloped by the 1C-2C peak, making it impossible to deconvolute. Since the integral of the ruler peak (1C-2C) is made greater by this overlap, the error introduced by this could lead to the distances between the C and D

rings to be larger than they actually are. The offset and Hartmann Hahn corrections are tabulated in Table IV.3. and the results of the corrections are displayed in Figure 4.4.

Table IV.3. Offset and Hartmann Hahn effects calculated for the heptasaccharide (**9**).

Ring	Cross peak	$sin\alpha_I\ sin\alpha_S$	HH transfer	s^2
Ring A	1A-2A	0.995	2A-1A	0.02
			2A-3A	0.06
	1A-2B	0.995	2B-1B	0.01
			2B-3B	0.24
Ring A'	1A'-2A'	0.996	2A'-1A'	0.02
			2A'-3A'	0.07
	1A'-2B'	0.994	2B'-1B'	0.01
			2B'-3B'	0.31
Ring C	1C-2C	0.992	2C-1C	0.04
			2C-3C	0.55
	1C-3D	0.988	2D-3D	0.07
			3D-4D	0.35
Ring C'	1C'-HPr(b)	0.986	HPr(b)-HPr(a)	0.80
			HPr(b)-HPr2	0.00
Ring B	1B-2B	0.994	2B-1B	0.01
			2B-3B	0.24
	1B-2C	0.991	1C-2C	0.00
			2C-3C	0.55
	1B-3C	0.991	2C-3C	0.55
			3C-4C	0.29
	1B-5A	0.989	5A-6A	0.00
			5A-4A	0.23
Ring B'	1B'-2B'	0.994	2B'-1B'	0.01
			2B'-3B'	0.31
	1B'-3C'	0.991	2C'-3C'	0.12
			3C'-4C'	0.27
	1B'-4C'	0.991	3C'-4C'	0.27
			4C'-5C'	0.373
	1B'-5A'	0.989	5A'-6A'	0.00
			5A'-4A'	0.13

4.3.3. Determination of Constraints and Initial Conformations for Molecular Dynamics

The problem of TOCSY transfer occurs in the ROESY spectrum of the heptasaccharide, as evidenced by the presence of cross peaks from 1B' to all of the ring protons of the C' ring, except 1C'. There is also transfer throughout the Glc*p*NAc ring, indicated by cross peaks corresponding to all the ring protons of the Glc*p*NAc along the chemical shift of 1C. To distinguish between ROESY and TOCSY peaks, NOESY and filtered ROESY experiments were run on **9**. The results from both of these experiments are almost identical. Cross peaks corresponding to 1C-2D and 1C-5D interactions, present in the ROESY spectrum, are no longer visible in either the NOESY or filtered ROESY spectrum. The F1 slice of either spectrum along the 1B' chemical shift shows no indication of a 1B'-5C' cross peak. All of the spectra are thus used to distinguish between NOE cross peaks and those due to TOCSY transfer, and thus aid in the determination of constraints.

The NOE contacts observed between the A and B rings in the ROESY, NOESY, filtered ROESY are the 1B-5A, and 1A-2B cross peaks. Examination of all the spectra does not show a 1B-1A cross peak. This implies that the Φ and Ψ angles are both in the positive quadrant, since a negative Ψ angle would place the H1B-H1A close together (~2.8Å). Even with the constraints entered, the angles seemed to move toward a +/- Φ/Ψ angle, so an additional constraint, i.e., H1B-H1A > 4.0 has to be entered. The A'-B' linkage demonstrates the same behavior. A single contact due to a 1A'-2B interaction is observed. Again, the absence of a 1A'-1B' peak indicates that the distance 1A'-1B' is greater than 4.0Å. The 1B'-5A' cross peak is also visible, conclusive proof that the two linkages are in similar conformations.

In the B-C linkage the cross peaks observed in the ROESY spectrum are between protons 1B-3C, 1B-2C, and 1B-4C. The NOESY and filtered ROESY both

show the 1B-3C and 1B-2C cross peaks (and the 1B-5A peak), but no 1B-4C cross peak. Noticably absent from the ROESY spectrum is the 1B-4C-5C TOCSY peak. The 4C-5C HH factor (s^2) is 0.12 compared to ~0.4 for most 4-5 Rhap s^2 values, and this could be the reason that the 1B-4C-5C TOCSY peak is absent. The presence of the 1B-2C cross peak in both the NOESY and the filtered ROESY spectra position the 1B proton close to the 2C proton. Though the spectral region 4.0-3.2ppm is very crowded, inspection of the spectra also shows the H2C-H5B to be absent. These results indicate a +ve Φ angle and a -ve Ψ angle.

TOCSY transfer throughout the D ring causes many contacts to be observed between the C and D rings. In the ROESY spectrum 1C-3D, 1C-4D, 1C-2D, 1C-5D cross peaks are identifiable. The NOESY and filtered ROESY show the 1C-3D and 1C-4D cross peaks, but it is difficult to discern whether the 1C-2D cross peak occurs since the 2C peak has the same chemical shift. The 3D and 4D resonances overlap, so the distances were calculated by observing that the total cross peak corresponds to a distance of 2.56Å. This distance can be considered to be the sum of two distances, i.e.,

$$\frac{1}{2.56^6} = \frac{1}{2.8^6} + \frac{1}{2.9^6}$$

and thus the 1C-3D and 1C-4D distances are constrained to 2.8Å and 2.9Å, respectively. A single cross peak,corresponding to a 1D-2A' interaction, is observed between the D and A' residues, and thus the D-A linkage cannot be assigned unequivocally.

The B'-C' linkage, like the C-D linkage, is complicated by the transfer of magnetization around the C' ring. Numerous contacts, i.e., 1B'-2C', 1B'-3C', 1B'-5C', 1B'-5A', 1B'-4C' are observed in the ROESY spectrum: . The NOESY and filtered ROESY show 1B'-3C'and 1B'-5A' cross peaks. In addition, a weak cross peak is observed to the H4C' in the filtered ROESY, but not in the NOESY, which is a similar

result to that observed for the compounds 4-8. This indicates that the Ψ angle is in the negative quadrant.

Two weak cross peaks are observed from the methyl group of the N-Acetyl moeity. These are unambiguously identified as being to the H2C and the H4A' protons. Since there is no ruler distance, these integrals were not converted to distances but the H2C and the H4A' protons are placed so as to be < 3.5Å away from the methyl group during the dynamics simulations.

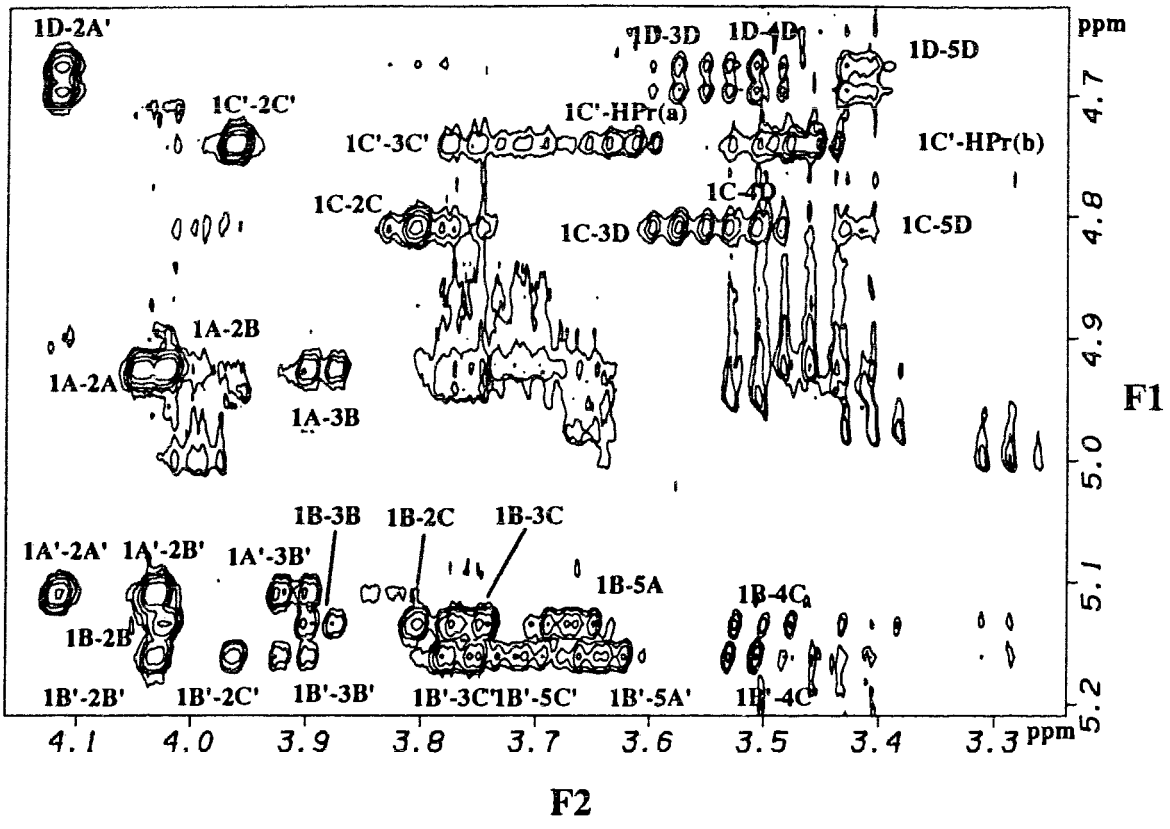
After correction for offset and Hartmann Hahn effects, the cross peaks are used to determine interproton distances as described in the previous chapter. The constraints obtained are tabulated in Table IV.4 and displayed in Figure 4.5.

Table IV.4. Constraints input for the molecular dynamics simulation of the heptasaccharide (**9**).

Proton Pair	Constrained distance	Upper limit[a]	Lower limit
H1A-H2B	2.28	2.48	2.08
H1B-H5A	2.59	2.69	2.49
H1B-H3C	2.17	2.27	2.07
H1B-H2C	2.86	3.06	2.66
H1C-H3D	2.80	3.00	2.60
H1C-H4D	2.90	3.10	2.80
H1D-H2A'	2.13	2.23	2.03
H1A'-H2B'	2.16	2.26	2.06
H1B'-H5A'	2.56	2.76	2.36
H1B'-H3C'	2.21	2.31	2.11
H1B'-H4C'	2.80	3.00	2.70

[a]The limits were derived from the highest and lowest values obtained from the experiment.

Molecular Dynamics simulations were performed as described in the previous chapter. The results are illustrated in Figures 4.6-4.9, and the average values for the constrained interproton distances and the Φ, Ψ angles are tabulated in Table IV.5 and IV.6



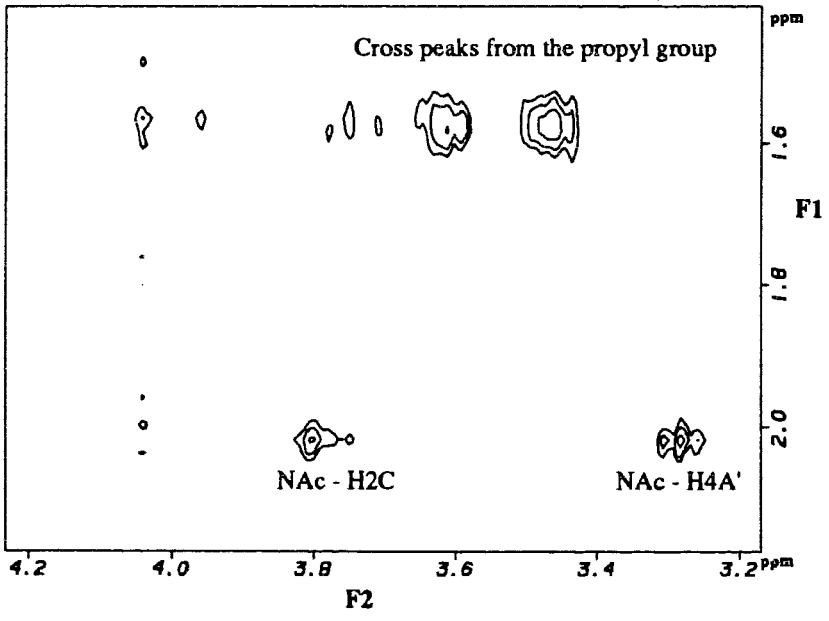


Figure 4.2. Expansion of the ROESY spectrum of the heptasaccharide (9). A. Anomeric region. B. Cross peaks from the methyl group of the N-Acetyl moeity.

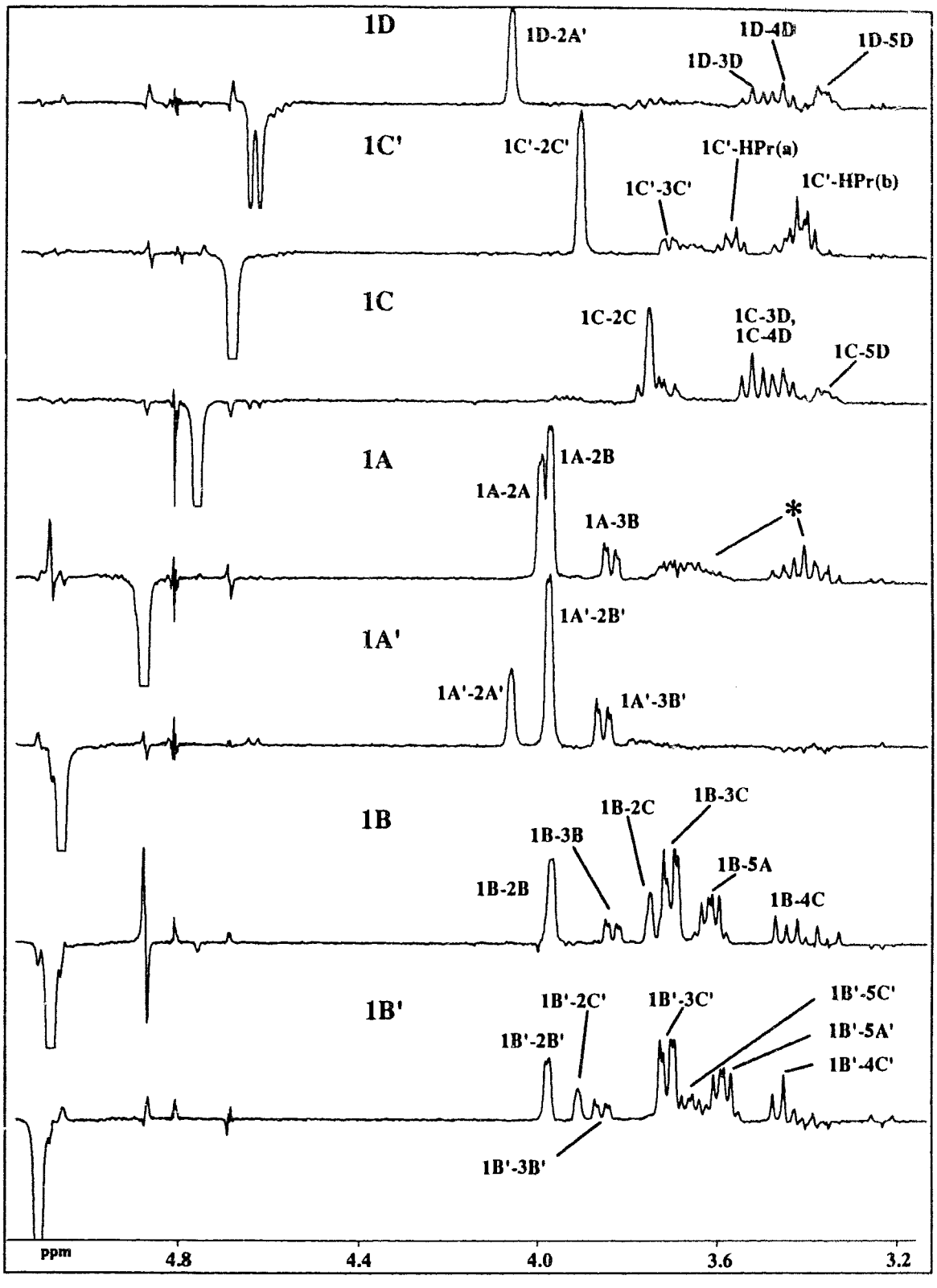


Figure 4.3. F1 slices from the ROESY spectrum of the heptasaccharide (**9**).

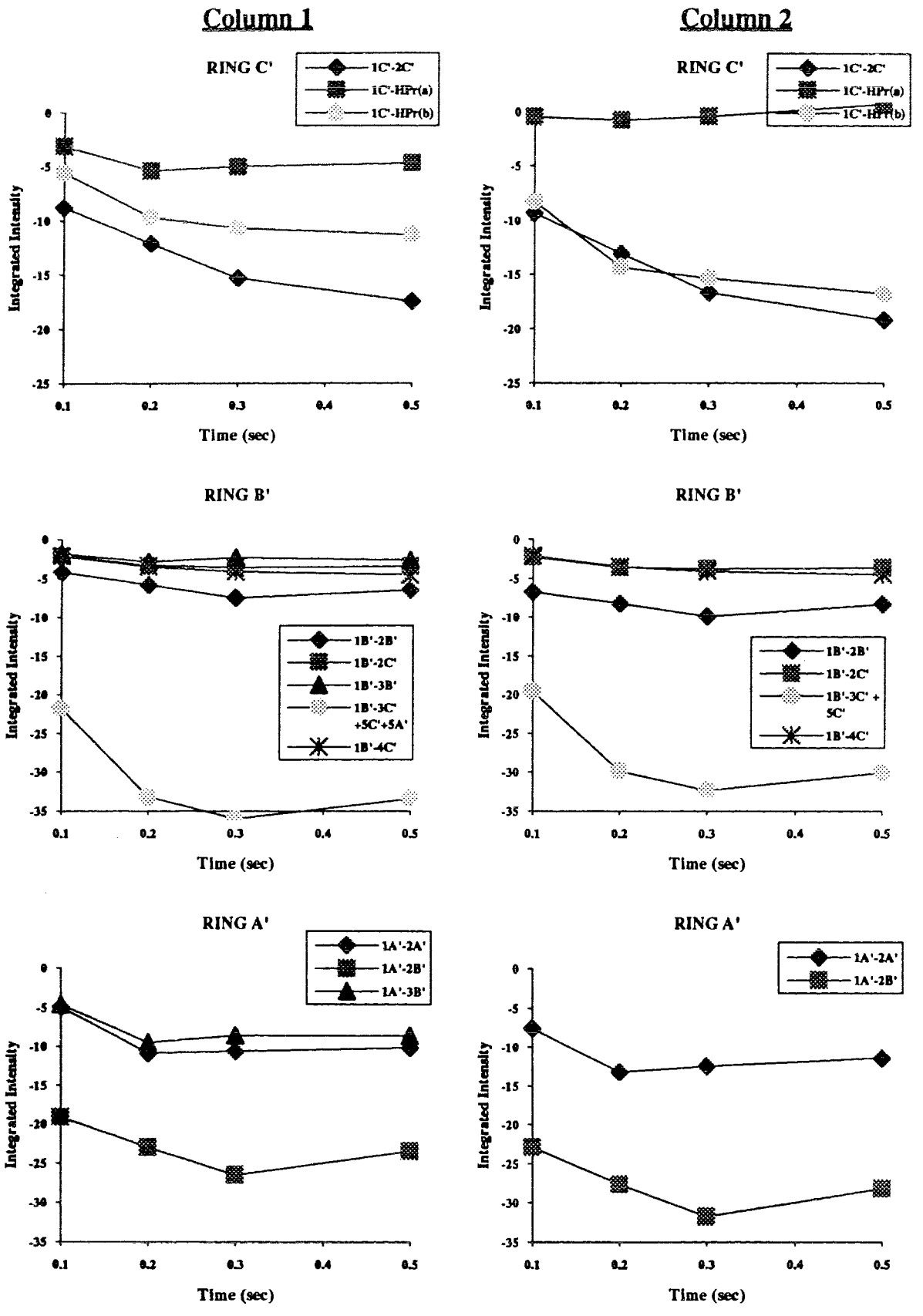


Figure 4.4. (caption two pages overleaf)

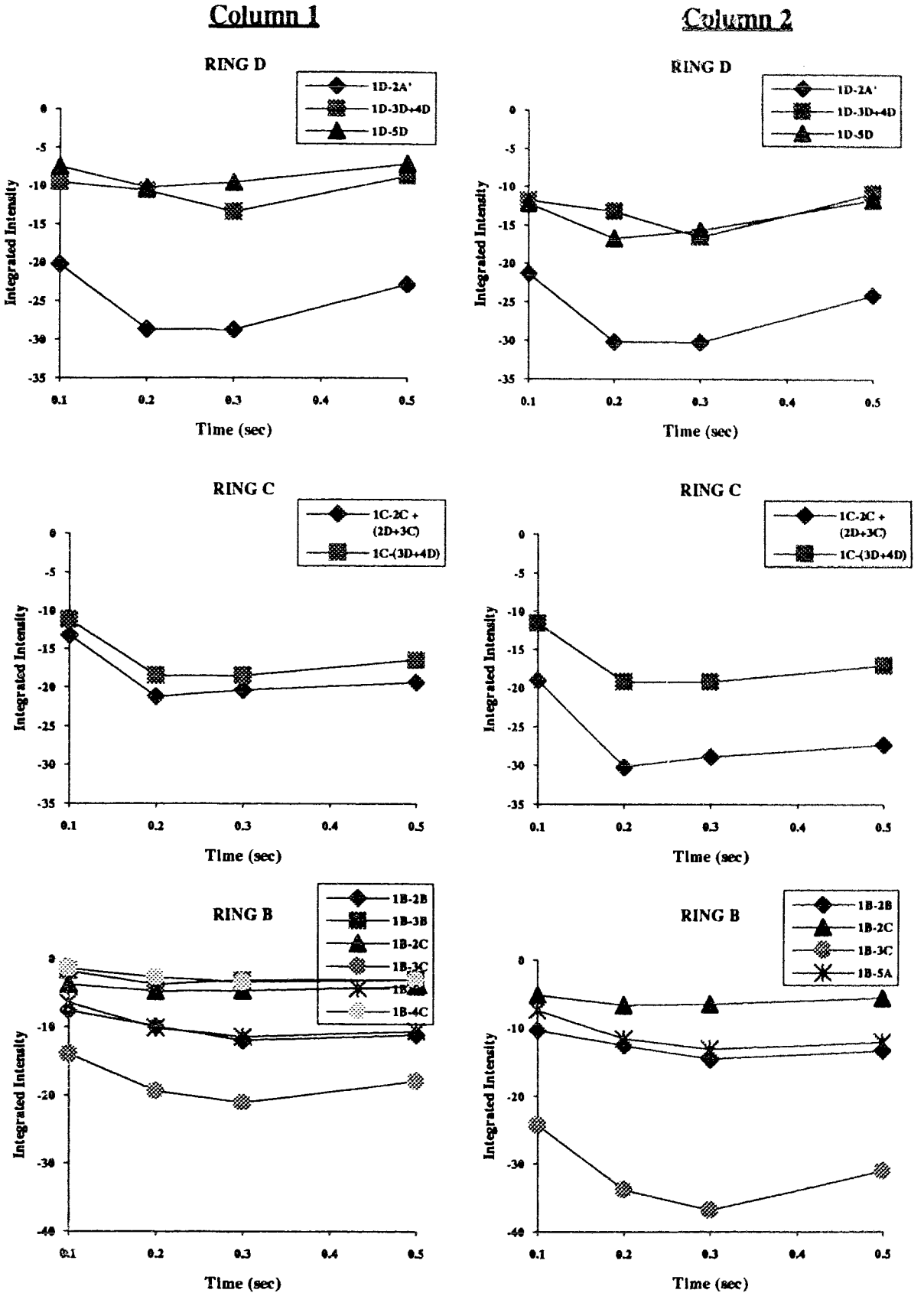


Figure 4.4. (caption overleaf)

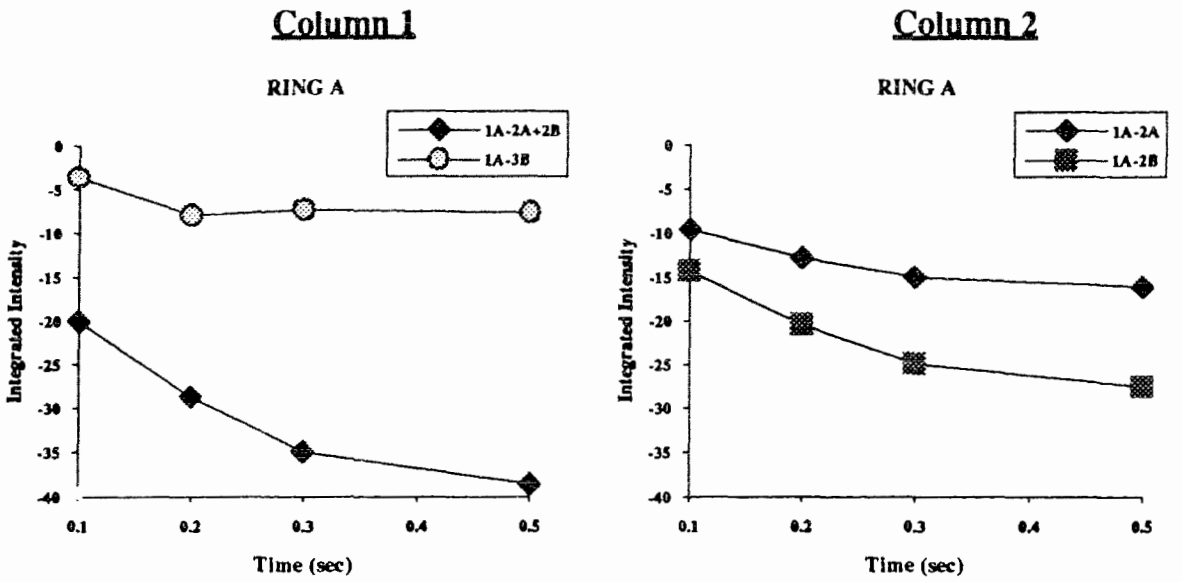


Figure 4.4. (This page and previous two pages) The effect of offset and Hartmann Hahn corrections on the ROE buildup of **9**. Column 1) Uncorrected data. Column 2) Corrected data.

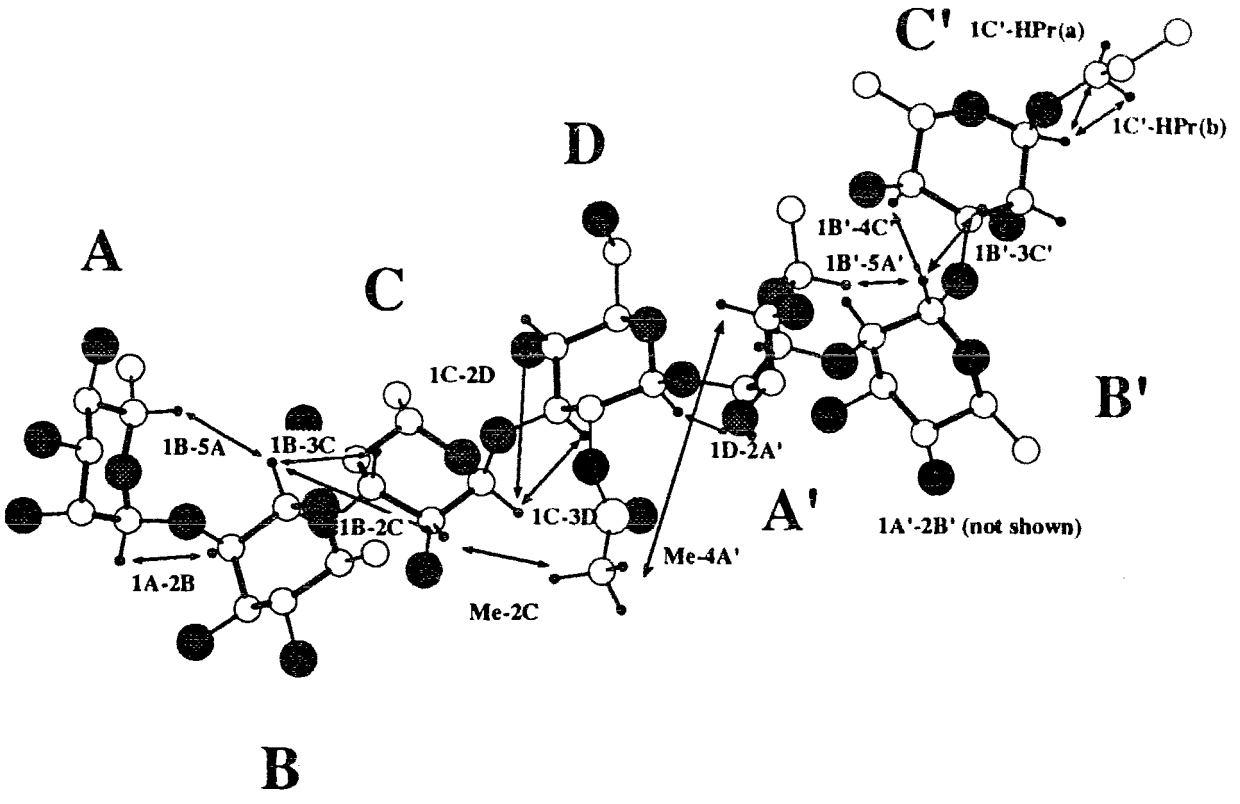


Figure 4.5. Interproton contacts derived from the ROESY spectrum of the heptasaccharide (**9**).

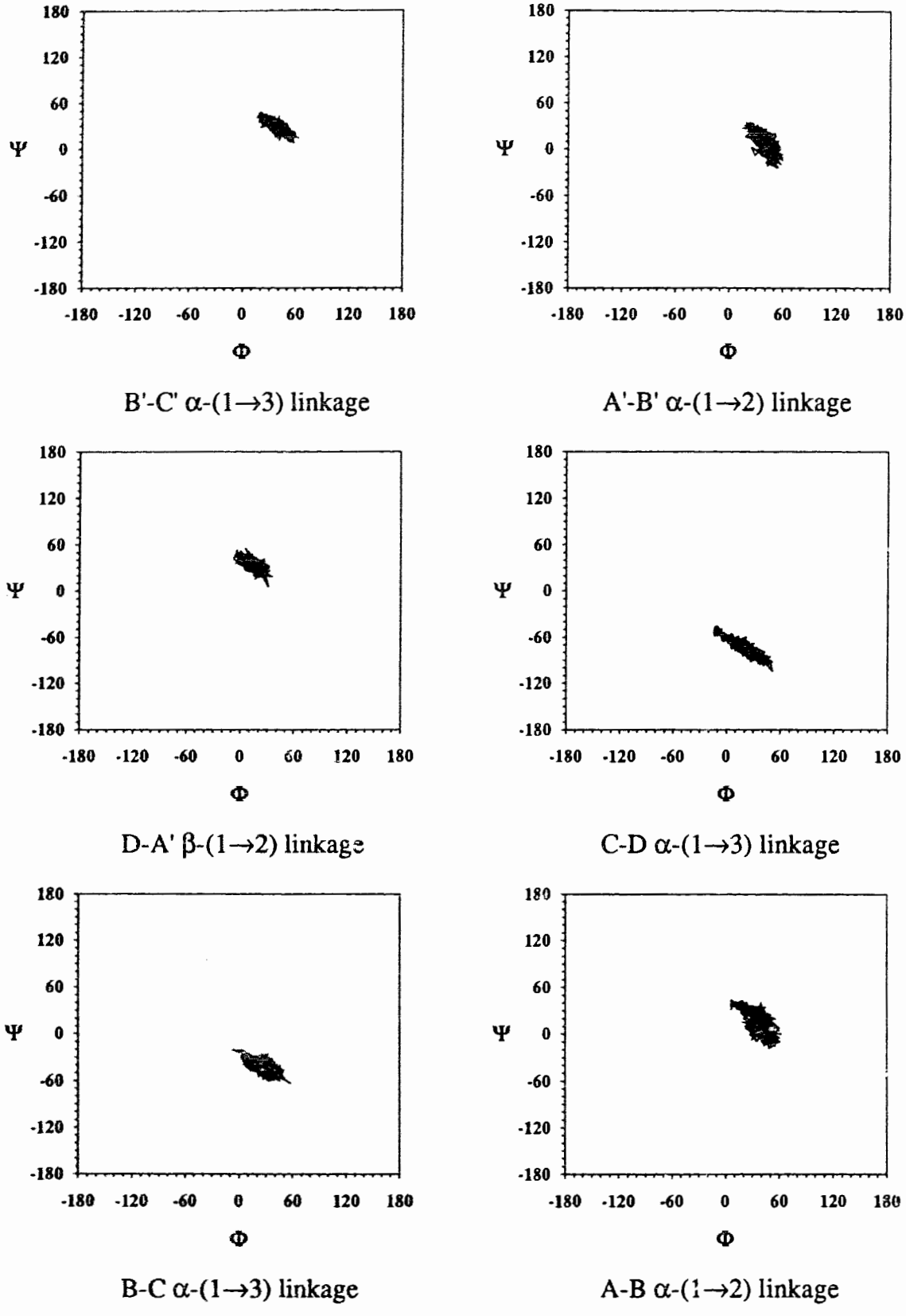
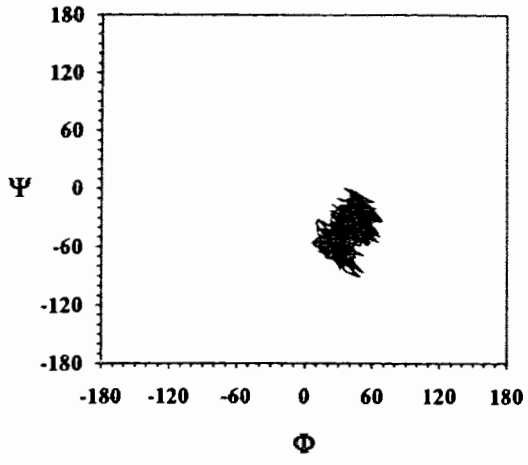
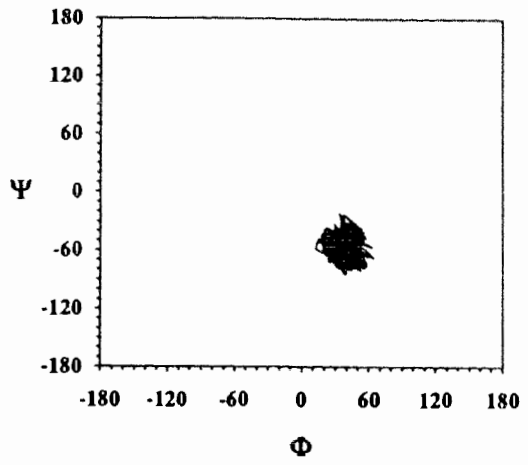


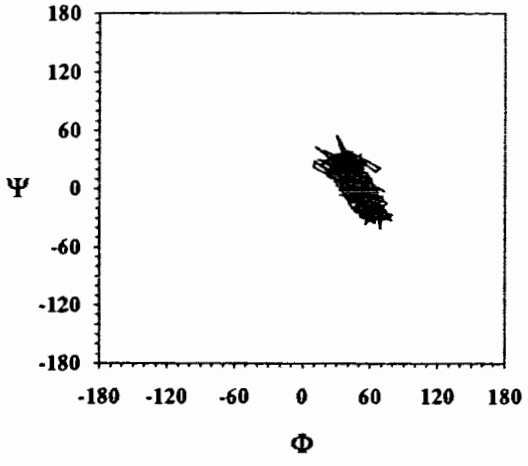
Figure 4.6. Variation of the Φ, Ψ angles during constrained dynamics simulations of the heptasaccharide (**9**) in H_2O.



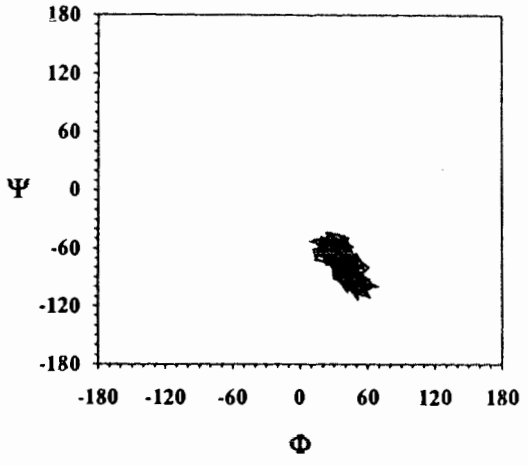
B'-C' α-(1→3) linkage



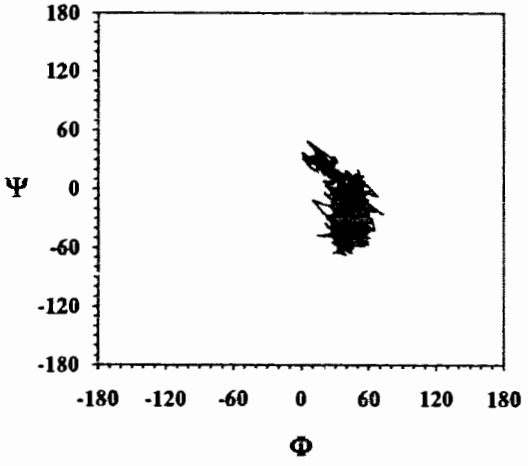
A'-B' α-(1→2) linkage



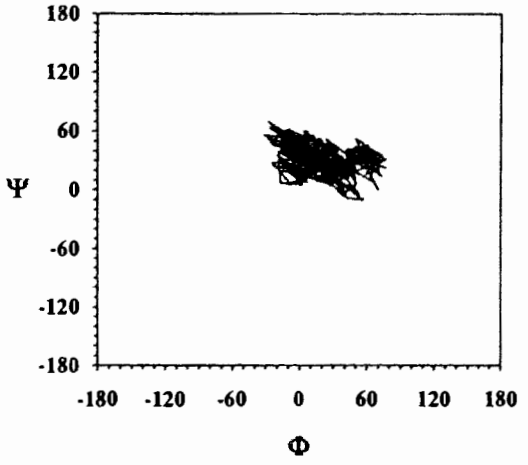
D-A' β-(1→2) linkage



C-D α-(1→3) linkage

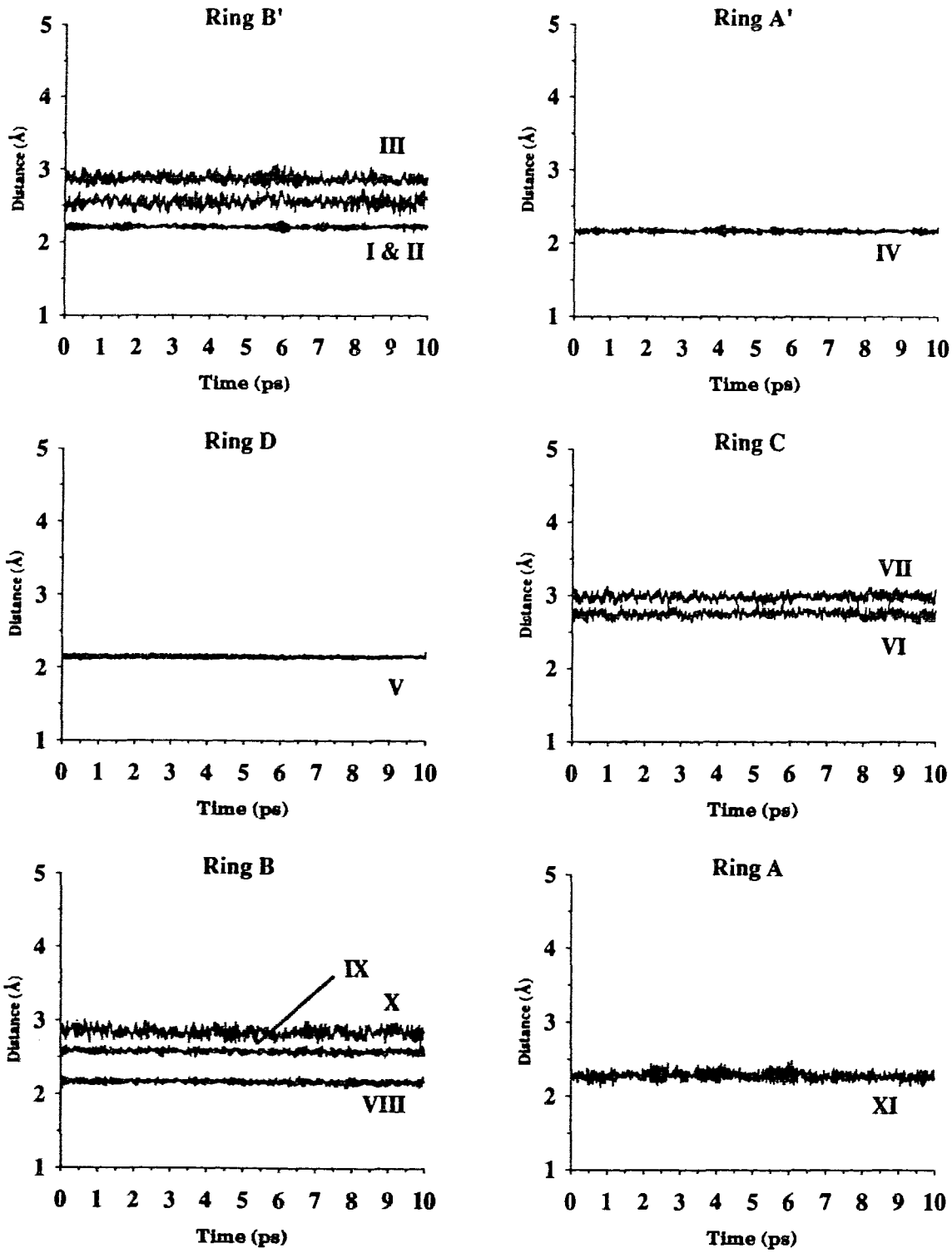


B-C α-(1→3) linkage



A-B α-(1→2) linkage

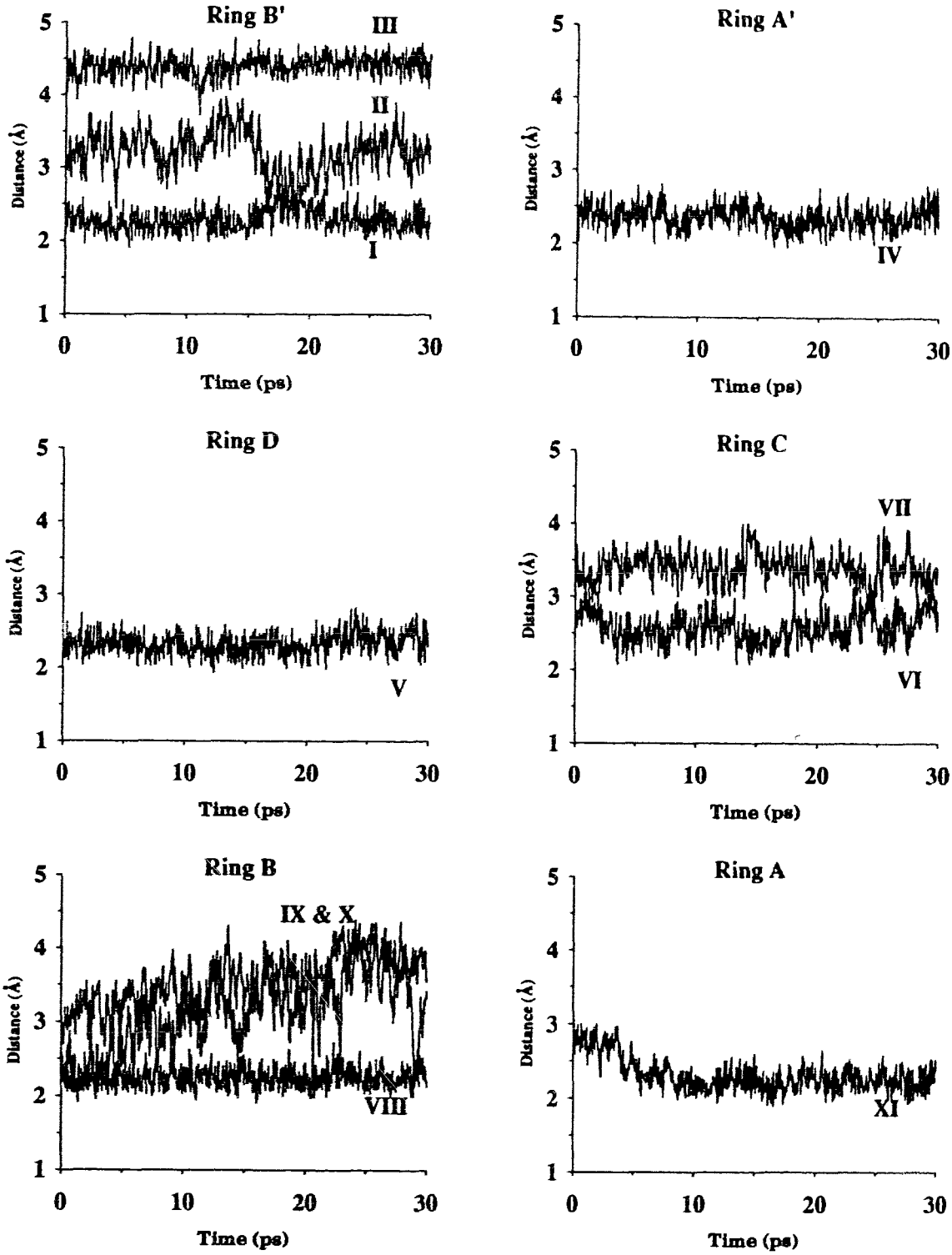
Figure 4.7. Variation of the Φ, Ψ angles during unconstrained dynamics simulations of the heptasaccharide (**9**) in H_2O.



LEGEND

I) H1B'-H3C'	IV) H1A'-H2B'	VIII) H1B-H3C
II) H1B'-H5A'	V) H1D-H3A'	IX) H1B-H5A
III) H1B'-H4C'	VI) H1C-H3D	X) H1B-H2A
	VII) H1C-H4D	XI) H1A-H2B

Figure 4.8. Variation of selected interproton distances during constrained dynamics simulations of the heptasaccharide (**9**) in H_2O.



LEGEND					
I)	H1B'-H3C'	IV)	H1A'-H2B'	VIII)	H1B-H3C
II)	H1B'-H5A'	V)	H1D-H3A'	IX)	H1B-H5A
III)	H1B'-H4C'	VI)	H1C-H3D	X)	H1B-H2A
		VII)	H1C-H4D	XI)	H1A-H2B

Figure 4.9. Variation of selected interproton distances during unconstrained dynamics simulations of the heptasaccharide (**9**) in H_2O.

Table IV.5. Average, high, and low values of selected internuclear distances during dynamics simulations of the heptasaccharide (**9**).

	Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Proton Pair	Average	High	Low	Average	High	Low
1C'-HPr(b)	2.68	2.73	2.62	2.44	3.07	2.00
1B'-4C'	2.87	3.07	2.73	1.39	4.79	3.72
1B'-3C'	2.17	2.25	2.08	2.25	2.72	1.91
1B'-5A'	2.55	2.74	2.35	3.21	4.05	2.20
1A'-2B'	2.17	2.25	2.08	2.37	2.92	1.95
1D-2A'	2.15	2.19	2.10	2.32	2.81	1.93
1C-3D	2.75	2.89	2.61	2.59	3.26	2.05
1C-4D	2.99	3.13	2.86	3.40	4.09	2.79
1B-3C	2.17	2.25	2.08	2.25	2.72	1.91
1B-2C	2.84	3.01	2.65	3.47	4.41	2.29
1B-5A	2.58	2.68	2.50	3.19	4.40	1.89
1A-2B	2.28	2.48	2.13	2.32	3.18	1.88

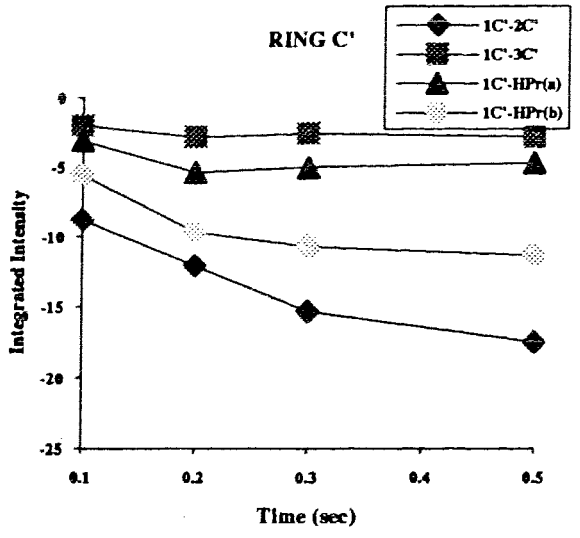
[a]Average temperature during the dynamics simulation was 295 K for the simulation with constraints and 303 K for the simulation without constraints.

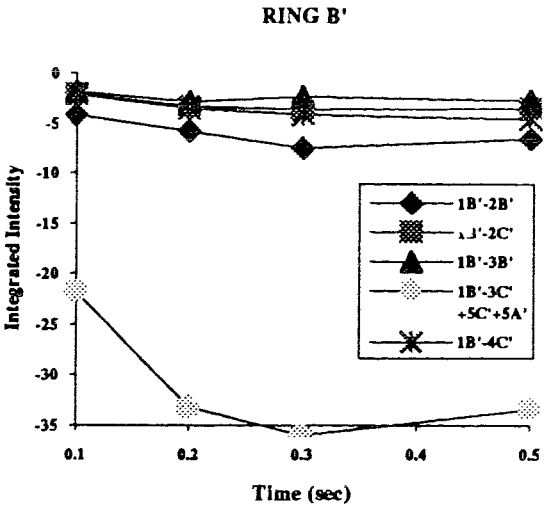
Table IV.6. Average, high, and low values of the glycosidic Φ and Ψ angles during dynamics simulations of the heptasaccharide (**9**)[a].

	Constrained Dynamics in H_2O			Unconstrained Dynamics in H_2O		
Angle	Average	High	Low	Average	High	Low
Link C'-Propyl						
Φ	50.13	71.48	16.45	60.11	99.34	-16.87
Link B'-C'						
Φ	39.58	63.03	16.05	40.09	75.88	5.19
Ψ	28.96	48.07	7.54	-42.63	8.16	-90.99
Link A'-B'						
Φ	40.98	60.03	15.62	37.72	64.13	7.48
Ψ	7.72	35.50	-23.91	-56.91	-18.62	-86.03
Link D-A'						
Φ	13.98	35.71	-7.03	48.75	80.32	4.83
Ψ	34.81	55.96	5.81	1.05	55.60	-46.82
Link C-D						
Φ	22.94	51.37	-14.61	37.92	71.28	-4.26
Ψ	-76.34	-45.88	-104.67	-74.42	-28.33	-117.88
Link B-C						
Φ	29.21	57.56	-7.48	41.49	76.76	-2.67
Ψ	-43.79	-20.16	-63.51	-24.37	48.92	-70.61
Link A-B						
Φ	39.18	61.91	6.28	17.00	79.49	-33.41
Ψ	13.21	44.30	-18.12	33.74	69.69	-11.24
ω and θ angles						
Ring D						
θ	-151.53	179.50	-179.86	113.25	179.98	-180.00
ω	33.38	72.98	-72.27	12.47	71.54	-78.98

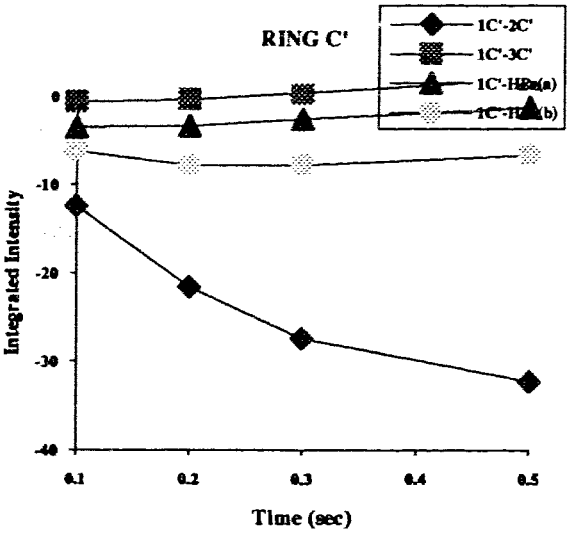
[a]Average temperature during the dynamics simulation was 295 K for the simulation with constraints and 303 K for the simulation without constraints.

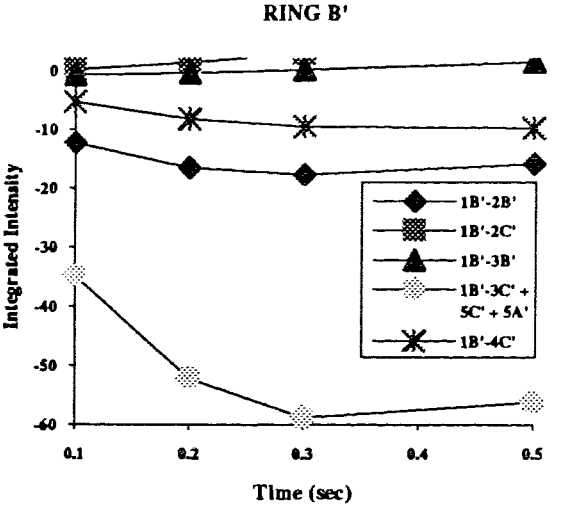
I


RING C'
1C'-2C'
1C'-3C'
1C'-HPr(a)
1C'-HPr(b)
Integrated Intensity
Time (sec)

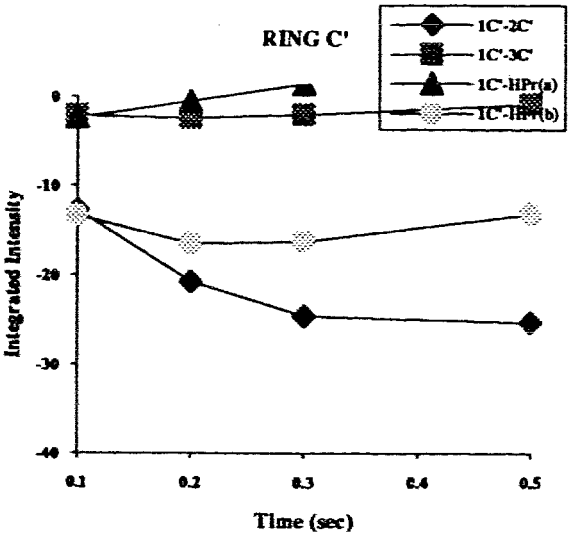

RING B'
1B'-2B'
1B'-2C'
1B'-3B'
1B'-3C' +5C'+5A'
1B'-4C'
Integrated Intensity
Time (sec)

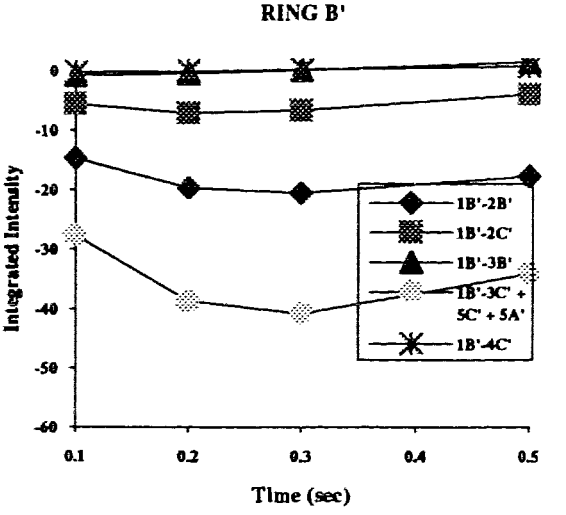
II


RING C'
1C'-2C'
1C'-3C'
Integrated Intensity
Time (sec)

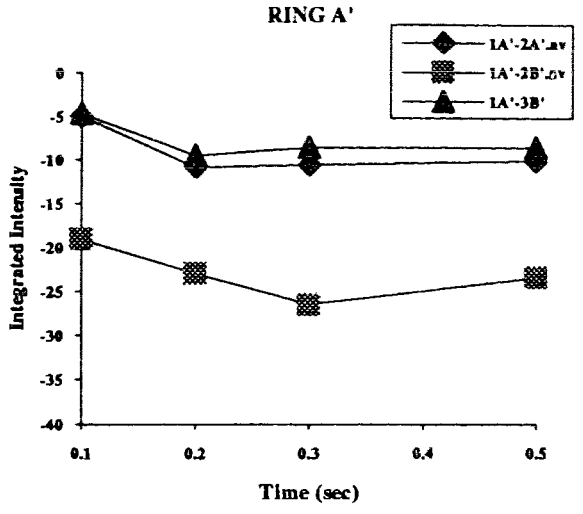

RING B'
1B'-2B'
1B'-2C'
1B'-3B'
1B'-3C' + 5C' + 5A'
1B'-4C'
Integrated Intensity
Time (sec)

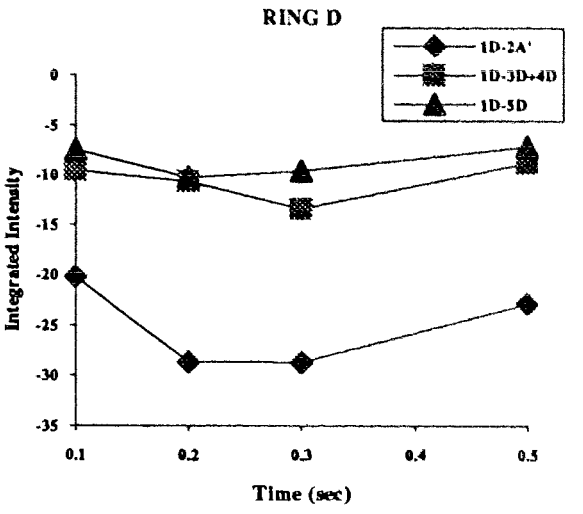
III


RING C'
1C'-2C'
1C'-3C'
1C'-HPr(a)
Integrated Intensity
Time (sec)

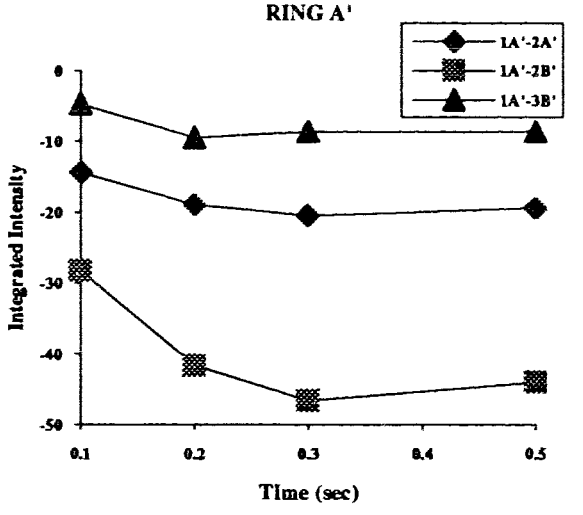

RING B'
1B'-2B'
1B'-2C'
1B'-3B'
1B'-3C' + 5C' + 5A'
1B'-4C'
Integrated Intensity
Time (sec)

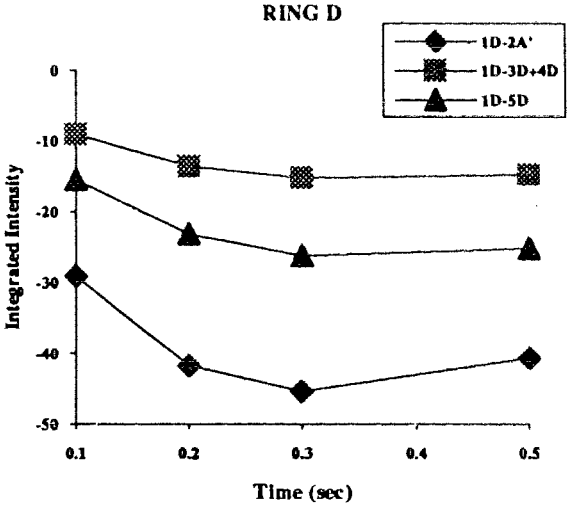
I


RING A'
1A'-2A'
1A'-2B'
1A'-3B'
Integrated Intensity
Time (sec)

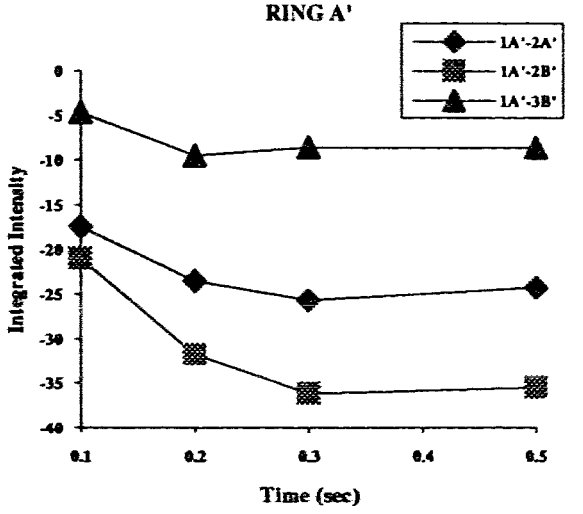

RING D
1D-2A'
1D-3D+4D
1D-5D
Integrated Intensity
Time (sec)

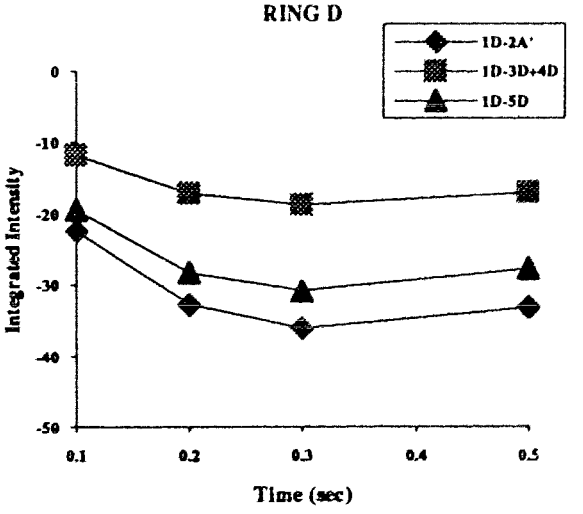
II


RING A'
1A'-2A'
1A'-2B'
1A'-3B'
Integrated Intensity
Time (sec)

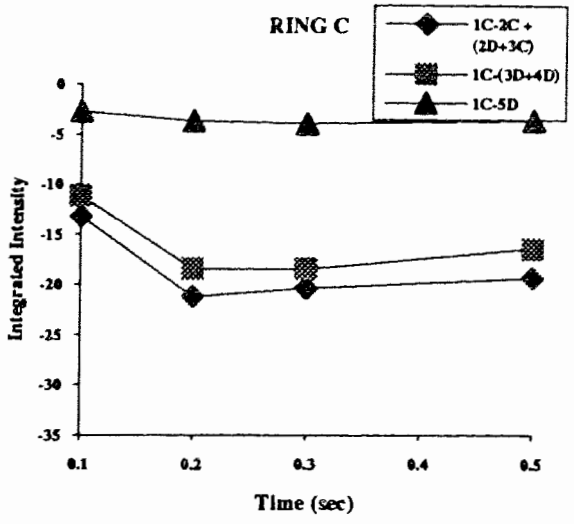

RING D
1D-2A'
1D-3D+4D
1D-5D
Integrated Intensity
Time (sec)

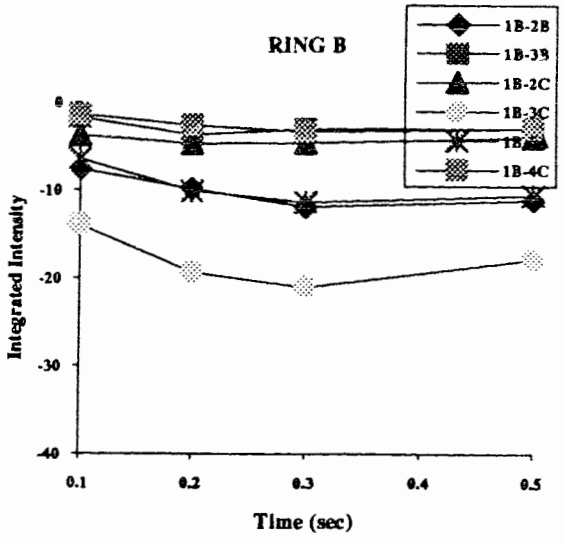
III


RING A'
1A'-2A'
1A'-2B'
1A'-3B'
Integrated Intensity
Time (sec)

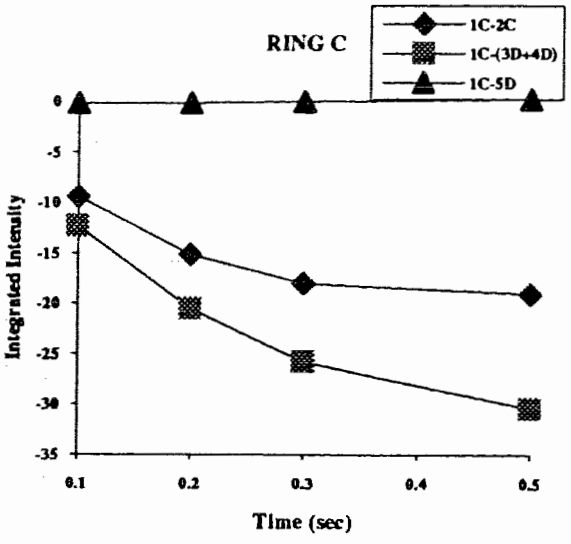

RING D
1D-2A'
1D-3D+4D
1D-5D
Integrated Intensity
Time (sec)

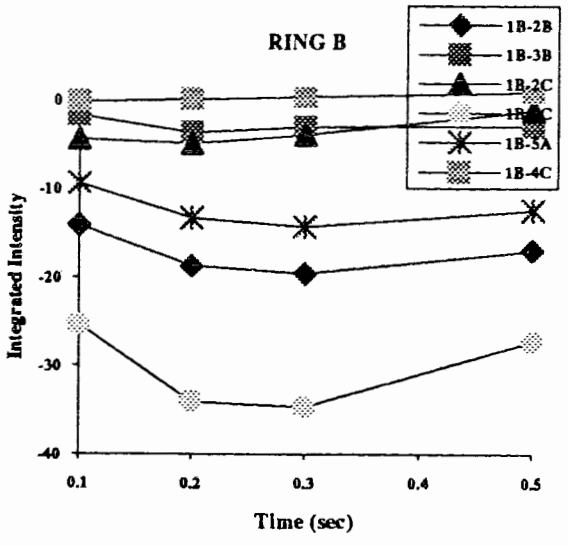
I


RING C
1C-2C + (2D+3C)
1C-(3D+4D)
1C-5D
Integrated Intensity
Time (sec)

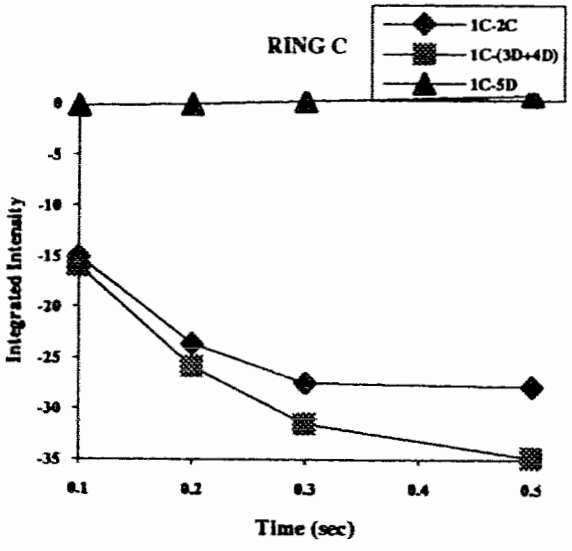

RING B
1B-2B
1B-3B
1B-2C
1B-3C
1B-4C
Integrated Intensity
Time (sec)

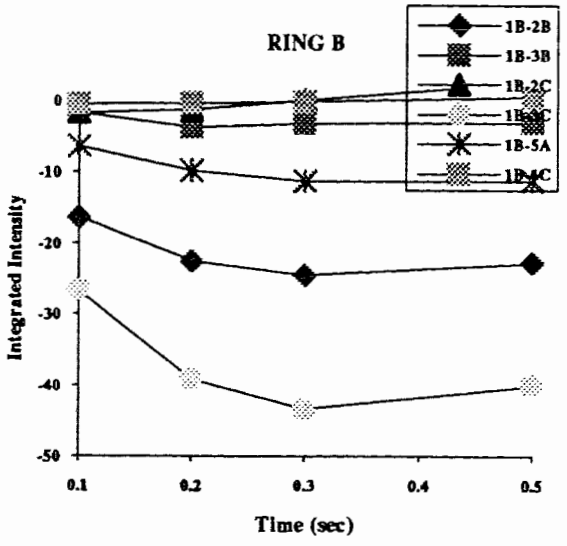
II


RING C
1C-2C
1C-(3D+4D)
1C-5D
Integrated Intensity
Time (sec)


RING B
1B-2B
1B-3B
1B-2C
1B-5A
1B-4C
Integrated Intensity
Time (sec)

III


RING C
1C-2C
1C-(3D+4D)
1C-5D
Integrated Intensity
Time (sec)


RING B
1B-2B
1B-3B
1B-2C
1B-5A
Integrated Intensity
Time (sec)

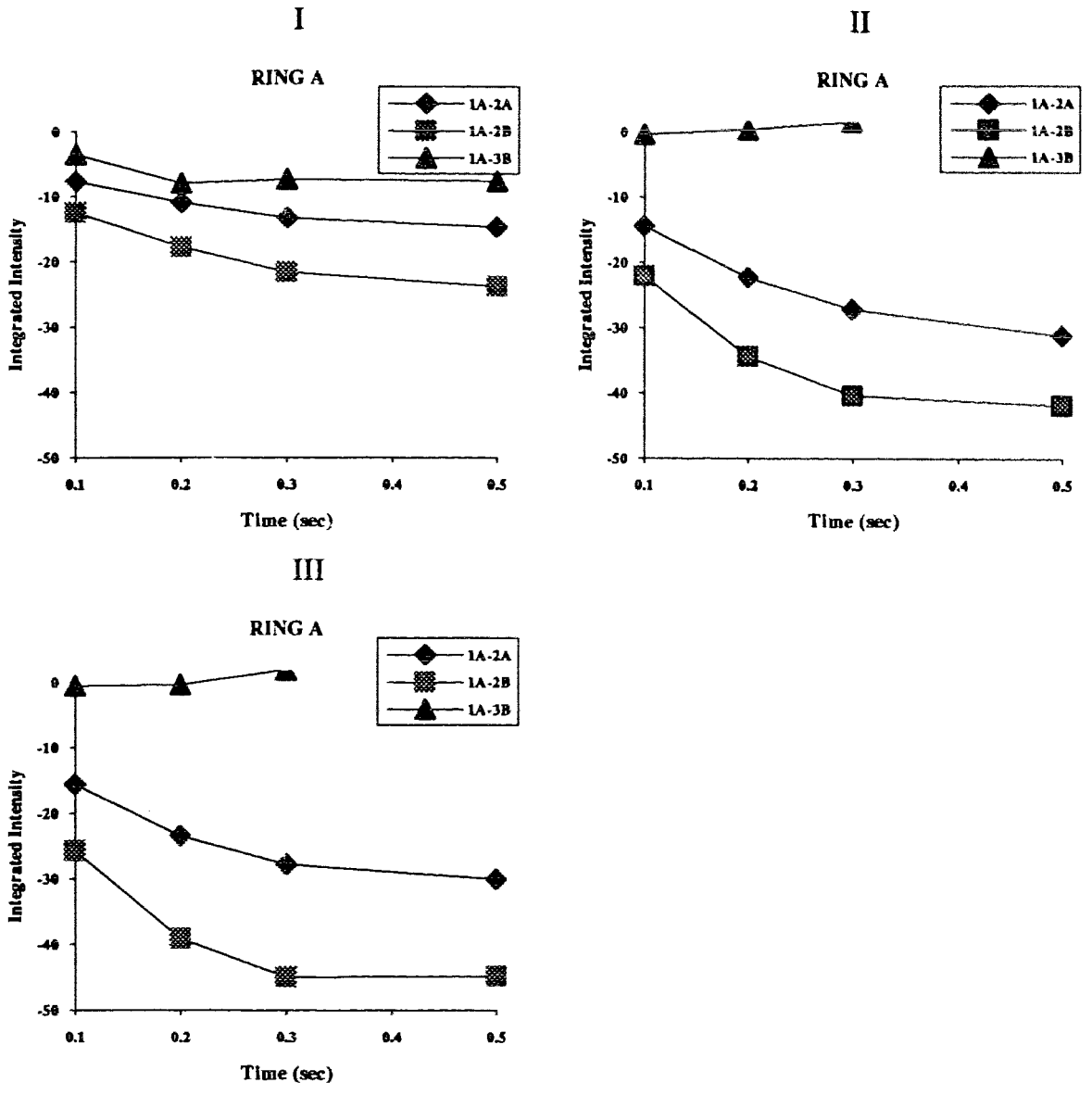


Figure 4.10. (previous 4 pages) Comparison of the experimental and calculated ROESY buildup curves for **9**. I) Experimental buildup curves. II) Buildup curves calculated from the constrained dynamics trajectories in H_2O. III) Buildup calculated from the unconstrained dynamics trajectories in H_2O.

4.4. Discussion

Conformational studies on oligosaccharide sequences of the *Shigella flexneri* variant Y lipopolysaccharide O-antigen have been focused primarily on tetrasaccharide fragments (the repeating unit), and a substantial body of NMR data has been collected. Early work utilized difference experiments to obtain NOE data, and relied heavily on chemical shift differences to rationalize proposed conformations. Utilizing steady state difference methods on tetra, tri and disaccharide fragments, as well as the polysaccharide (YPs) itself, Bock et al.[77] observed the inter-ring NOEs displayed in Figure 4.11. The NOEs observed by Bock[77] have been substantiated by later work on similar polysaccharides from different strains of *Shigella* by Jansson et. al.[205]. However, for all of the compounds studied, only one inter-ring NOE contact is observed, with the exception of the α-(1$\rightarrow$2) (the A-B) linkages. This could be due to complications associated with 1D steady state and transient NOE experiments addressed in Chapter I, specifically the poor enhancement due to problems arising from correlation times as well as the difficulty in selective saturation of desired resonances. Although it is possible to deduce the relative proximity of the two protons involved in an NOE contact, a single inter-residue NOE is insufficient to assign unambiguously the Φ, Ψ angles based solely on experimental data, and thus, molecular modelling can be employed to narrow further the range of possibilities. Thus, with HSEA calculations, Bock et. al.[77] obtained a minimum energy conformation that would account both for the observed NOE contacts as well as the chemical shift differences within the series of compounds.

The average Φ and Ψ angles from the constrained molecular dynamics calculations performed in the present study are tabulated in Table IV.7, along with the values from the minimum energy conformation as obtained from HSEA calculations[77]. The two conformations are displayed in Figure 4.12, and whereas the A'B'C' units at the reducing end exhibit a similar conformation in both models, the ABCD unit differs

significantly. The reducing end displays a linear, extended surface, while the last three residues, the A, B, C residues are almost folded back on each other. A preliminary analysis shows that the difference between these two regions is the presence of the Glc*p*NAc residue, which implies that the difference in conformation may somehow be induced by the Glc*p*NAc ring. An analysis of the ROESY spectrum and the various chemical shift changes between the C and C' residue could provide some clues as to whether this is indeed the case.

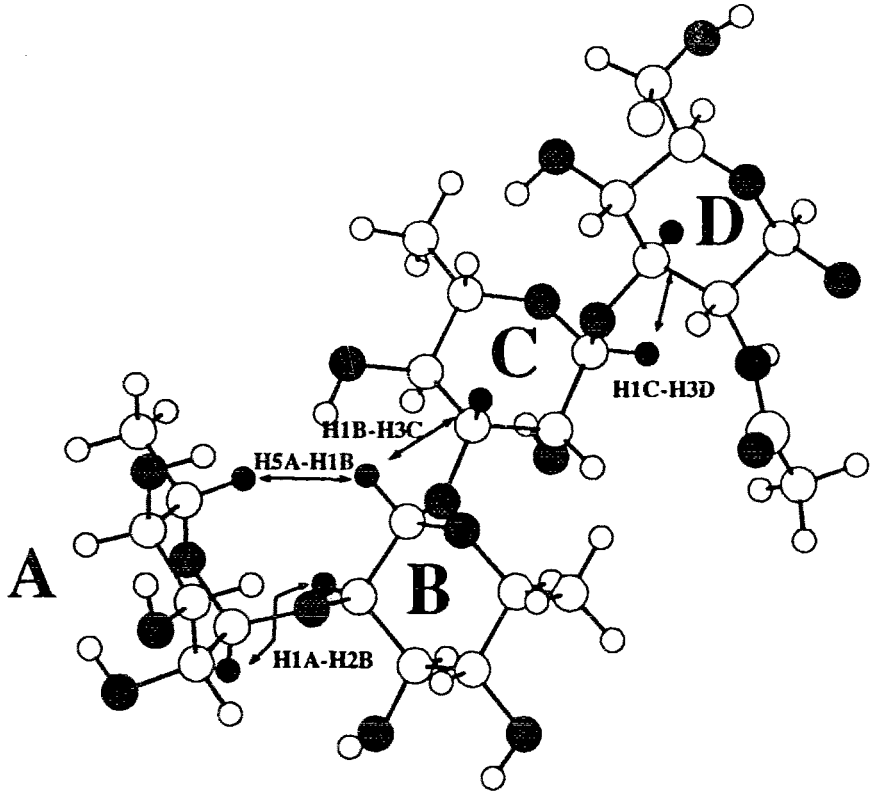


Figure 4.11. NOE contacts observed by Bock et. al.[77] in the tetrasaccharide fragment ABCD.

Table IV.7. Φ and Ψ angles calculated by HSEA calculations[77] and average values derived from constrained molecular dynamics simulations for **9**.

Linkage	A-B	B-C	C-D	D-A'	A'-B'	B'-C'
HSEA	45,15	50,15	40,15	50,10		
9	39,13	29,-44	23,-76	14,35	41,8	40,29

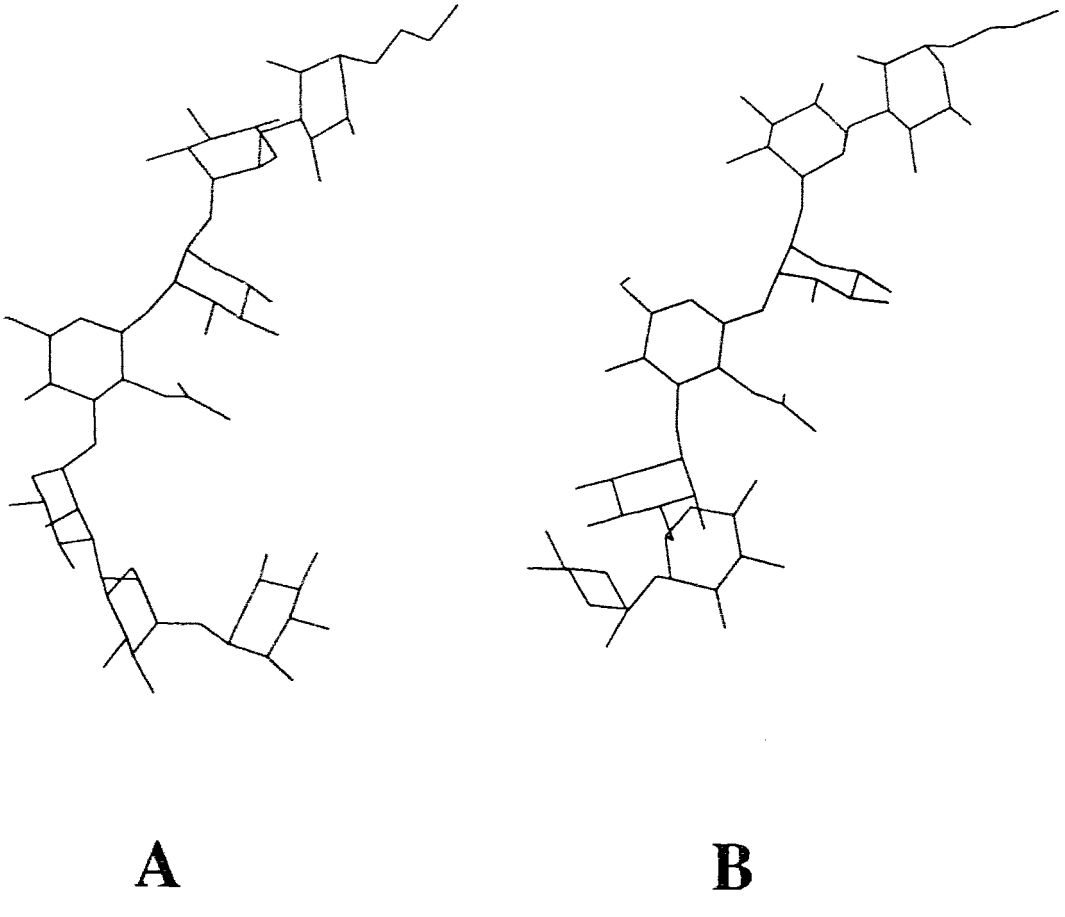


Figure 4.12. Minimum energy conformations of the heptasaccharide (**9**). A) Conformation derived from the average values of the constrained molecular dynamics simulations and B). Conformation predicted by HSEA calculations[77].

The heptasaccharide (**9**) shares a number of common linkages with the *Streptoccoccus* oligosaccharides (**4-8**), namely an α-L-Rhap-(1→2)-α-L-Rhap linkage, and an α-L-Rhap-(1→3)-α-L-Rhap linkage, and thus cross peaks characteristic of each of these linkages appear in the ROESY spectrum of the heptasaccharide.

The A-B and A'-B' linkages are both α-(1→2) links analogous to those in the compounds **4-8**. The ROESY spectrum shows a strong 1A(')-2B(') cross peak across the linkage, indicating that the two protons are adjacent and the Φ/Ψ angles are gauche rather than anti. The second cross peak observed between the rings involved in this linkage is between the 1B(') and 5A(') protons. There are two possible sets of angles that will place these proton pairs within the required distances with Φ/Ψ angles of 47/-41 and 47/29. Both of these torsions lie within minima predicted by CHARMM, the former being the global minimum, 0.1 kcal lower in energy than the latter. Distinction between the two conformations can be made on the basis that a conformation with a negative Ψ angle will place the 1A(') and 1B(') protons in close proximity. The absence of a 1A(')-1B(') interaction in the ROESY spectra precludes this conformation. The conformation of the α-(1→2) linkages is in total agreement with that predicted by HSEA calculations[77]. The 1A-2B and 1B-5A interactions are also observed by Bock et. al.[77]

The two α-(1→3) linkages, the B-C and B'-C' linkages, display markedly different properties. The 1B' proton displays the same interactions that are observed for the α-(1→3) link in the spectra of compounds **4-8**, i.e., aside from a 1B'-5A' interaction, a strong cross peak from the 1B' proton to the 3C' proton, and weaker cross peaks to the 2C', 4C' and 5C'. The 1B proton exhibits analogous interactions except for the 1B-5C cross peaks. The results from NOESY and filtered ROESY experiments indicate that the 1B-2C and 1B-3C cross peaks result from true NOE interactions, (either direct or relayed), as is the 1B'-3C' interaction. The 1B'-4C' peak is extremely weak, as is the 1B'-2C' cross peak. It is difficult therefore, to predict the location of the 1B' proton with

respect to the 2C' and 4C' protons. The 2C'-5B' cross peak, which could be used as conclusive proof that the conformation of the B'-C' linkage in the heptasaccharide is similar to that of the α-(1-3) linkages in the compounds **4-8**, is difficult to discern due to

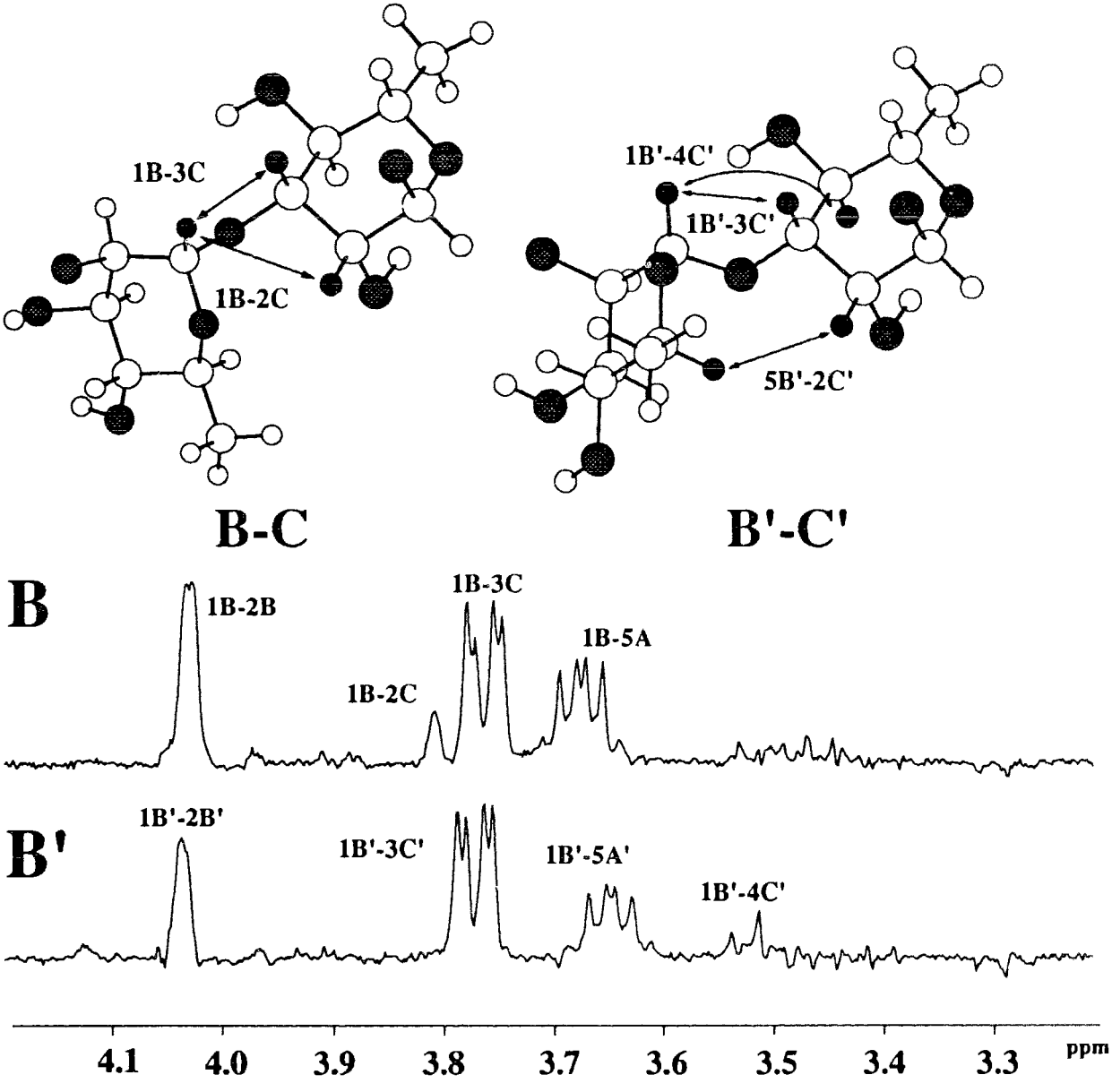


Figure 4.13. Expansion of the F1 slices from the filtered ROESY and disaccharide fragments displaying the different NOE contacts observed from the anomeric protons of the B and B' residues.

overlap between the 3C', 3C, 3A and 5B' proton chemical shifts, but in the NOESY spectrum this peak can be tentatively identified since the multiplicity i.e., the splitting pattern, is consistent with the superposition of a H3C' and H5B' multiplets. These results indicate that the B'-C' linkage occupies the same conformational space as the analogous α-(1→3) linkages in the series of *Streptococcus* Group A oligosaccharides (Chapter III).

The analysis of the B-C and the B'-C' linkages is an example of why extrapolation of conformational properties must be performed with caution. Since the linkage is the same it is tempting to use the data from the *Streptococcus* compounds and assign positive Φ/Ψ angles to the B'-C' and B-C linkages, given the similarity in cross peaks. In the latter compounds, the H1B-H4C ROE leads also to an H1B-H5C cross peak via TOCSY effects. Preliminary examination of the B-C linkage in the heptasaccharide (**9**) shows that the Hartmann Hahn factor (s^2) for the 4C-5C proton pair is much lower than that usually observed (0.121 compared to an average of ~0.3 for most H4-H5 pairs in rhamnose) due to the downfield shift of the 5C proton. This could explain the lack of transfer of magnetization from 4C to 5C and the absence of an H1B-H5C cross peak. However, the presence of a 1B-2C cross peak in the NOESY and filtered ROESY spectra is conclusive proof that the 1B proton is in fact proximal to the 2C proton, and not to the 4C proton. Additional proof that the 1B-2C peaks is not an artifact is furnished by Jansson et. al.[205]. NOESY experiments on various polysaccharides containing the same linkages display a 1B-2C NOE (for example, see Figure 4 and Figure 6 in reference 3a). The 1B-2C cross peak is not detected in the study by Bock et. al.[77], and our observation of this cross peak unambiguously determines the Φ Ψ angles of this linkage.

The Φ/Ψ angles of the B'-C' linkage are similar to those observed by Bock et. al.[77], but the B-C linkage does not display the same interglycosidic angles. There is no obvious reason why the B-C and B'-C' linkages behave differently. In fact, assignment

of the same values of Φ/Ψ to the B-C linkage would afford an extended, linear structure, which would seem to possess less unfavorable steric interactions, but the addition of the D (Glc*p*NAc) residue in some way affects the conformation of the B-C linkage.

Two linkages present in the heptasaccharide that are absent in the family of *Streptococcus* oligosaccharides previously examined are the β-(1→2) linkage between residues D and A' and the α-(1→3) linkage from the C ring to the Glc*p*NAc (D) ring.

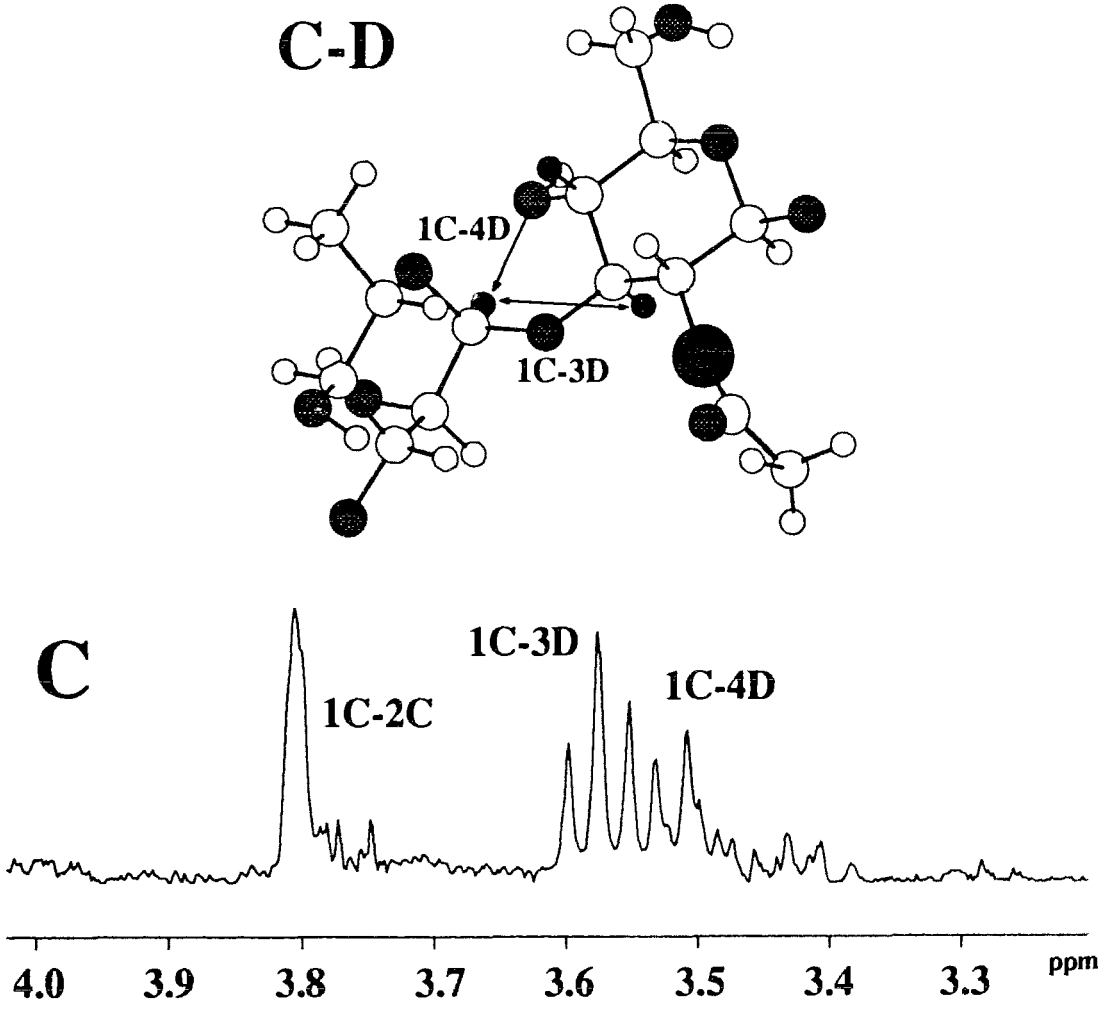


Figure 4.14. Expansion of the F1 slice from the filtered ROESY and disaccharide fragment displaying the different NOE contacts observed from the anomeric proton of the C residue.

The latter, the α-(1→3) (C-D) linkage is well defined by the cross peaks from the 1C to the 3D and 4D protons, conclusively placing the 1C proton between both of them, which would orient the Φ and Ψ angles in the +/- gauche region (Figure 4.14). The assigned Φ/Ψ angles of 23/-76 reproduces the 1C-4D and 1C-3D interactions. No H1C-H4D enhancement is observed by Bock et. al.[77]. The proposed conformation consistent with this NOE contact represents a major discrepency between our results and those of Bock et. al., and thus this cross peak was carefully scrutinized in order to determine its authenticity and make sure it was not an artifact. Examination of the ^{1}H chemical shifts shows that two proton shifts, H4C and H4C', overlap with the H4D chemical shift. The 4C' proton is too far removed to show any enhancement from the 1C proton, and in all the compounds studied, no intra-ring enhancement from a H1 to the H4 of a rhamnose residue has been observed. Since both of these possibilities can be ruled out, the peak can be identified as that arising from a 1C-4D contact. A similar peak is observed by Jansson et. al.[205], but has been assigned as a 1C-4C (intra-ring) contact (Figures 4 and 6 and Table IV; reference 3a). The calculated H1-H4 interproton distance for an α-L-rhamnopyranosyl unit is 4Å which leads to the conclusion that the latter assignment is incorrect.

In the D-A' linkage, only one significant interglycosidic contact is observed, that between the 1D and the 2A protons, and consequently the Φ and Ψ angles cannot be determined unambiguously by examination of the anomeric region of the spectra. Absence of a 1D-1A contact does indicate that this distance is > 4.0 Å, which would place these angles in the positive quadrants. This is in the same range as the Φ/Ψ angles predicted by HSEA calculations (50,10).

The observation of two other cross peaks in the ROESY spectrum of **9** conclusively establishes the conformation of the trisaccharide portion CDA'. Cross

peaks from the methyl of the N-acetyl moiety to 4A' and to 2C are observed, and although no quantitative distances were derived from these cross peaks (see Results section), this is indicative of the protons being within 4.0Å of the methyl group. This positions the Φ and Ψ angles for the D-A' linkage as indicated above and also sets the theta angle in the vicinity of 180°. The conformation proposed by Bock et. al.[77] would also account for the presence of these cross peaks.

An analysis of the differences in the ^{1}H and ^{13}C chemical shift between the heptasaccharide and the polysaccharide (Y-PS), in an analogous fashion to that performed in Chapter 3, is illustrated in the set of graphs in Figures 4.15 and 4.16. In order to account for the different standards used to calibrate the spectra, initially the ^{1}H and ^{13}C chemical shifts of the 1D proton and carbon in the heptasaccharide were set to the same values as that of the corresponding signals in the polysaccharide. The projected $\Delta\delta$ values represent the difference between the chemical shift of the Y-PS and the heptasaccharide, i.e. a negative value in the chart indicates that the heptasaccharide chemical shift is downfield with respect to the Y-PS, and similarly, a positive value of $\Delta\delta$ implies an upfield shift in the heptasaccharide. The ^{1}H shifts of the heptasaccharide were also compared to those of the decasaccharide[198] corresponding to the sequence A"B"CDABC'D'A'B'. The chemical shifts of the 1D protons were again made equivalent to negate solvent and temperature effects. Since the C and C', and the D and D' residues show identical chemical shifts, the comparison between the heptasaccharide and the decasaccharide C and D residues was straightforward. However, faced with the choice of which A and B residues in the decasaccharide were analogous to those in the heptasaccharide, the following guidelines were followed. The A and B rings of the heptasaccharide are analogous to the A" and B" rings in the decasaccharide, since they are terminal residues in both compounds. The A' and B' rings of the heptasaccharide are compared to the A and B rings of the decasaccharide. The results are displayed in Figure

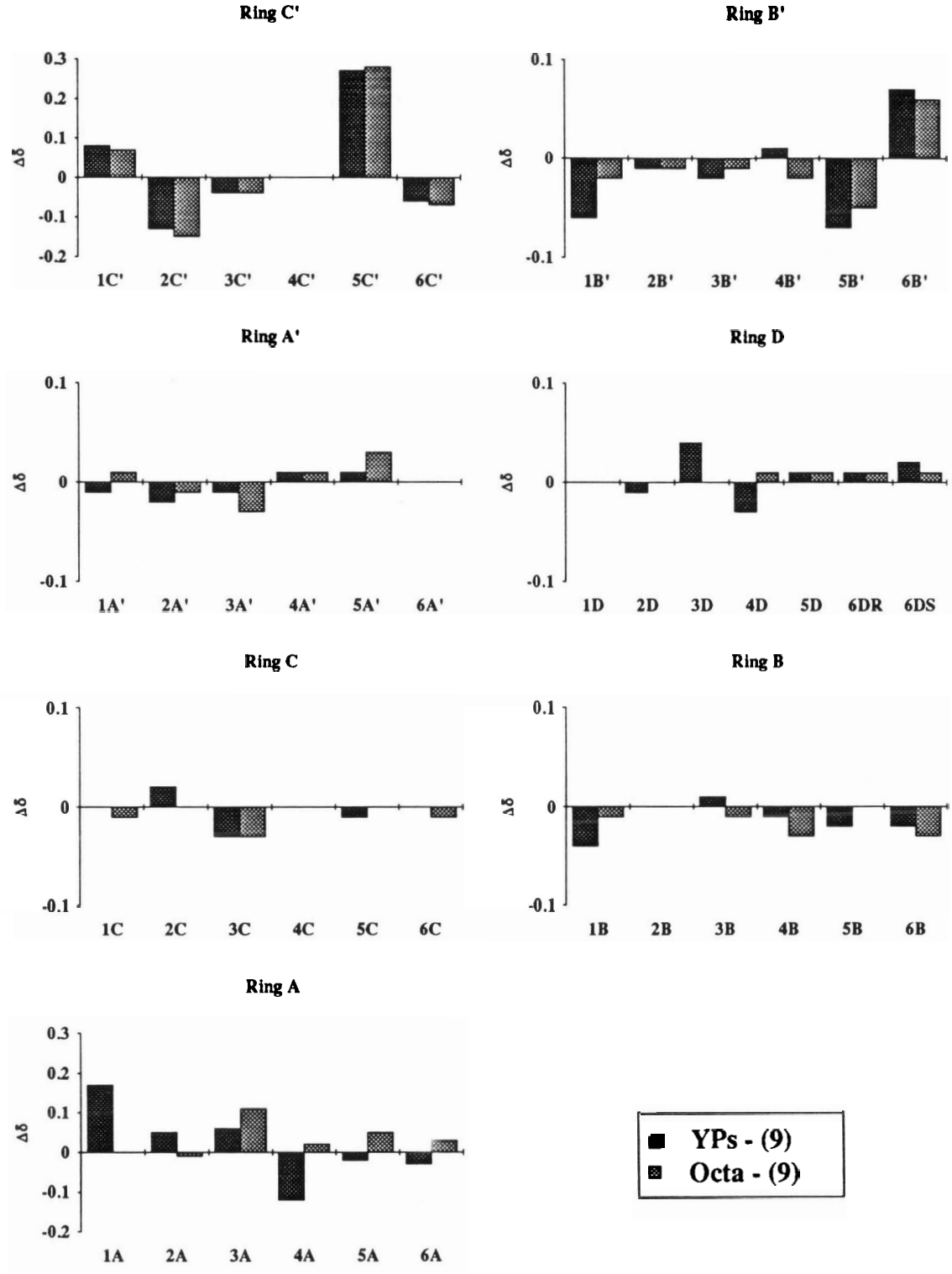


Figure 4.15. Differences in the ^{1}H chemical shift of the heptasaccharide (**9**) and the *Shigella flexneri* variant Y O-antigen polysaccharide and the decasaccharide fragment A"B"CDABC'D'A'B'.

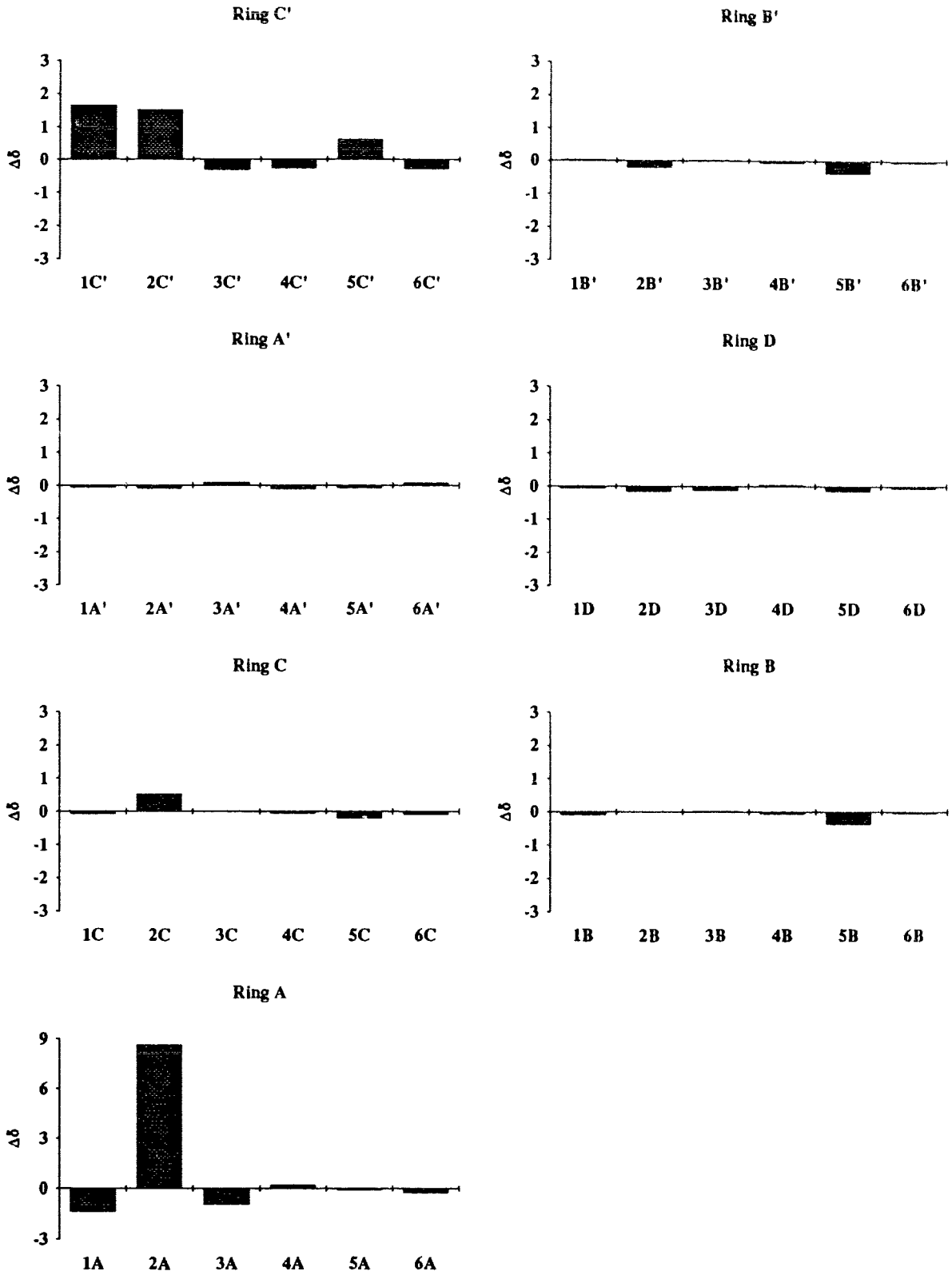


Figure 4.16. Differences in the ^{13}C chemical shift of the heptasaccharide (**9**) and the *Shigella flexneri* variant Y O-antigen polysaccharide.

4.15, and since they are exactly the same as those from the comparison of the heptasaccharide and the polysaccharide, they will not be discussed in detail.

With regard to the ^{1}H chemical shifts, the largest difference is observed between the 5C' and the corresponding 5C proton in the heptasaccharide. The downfield shift of ~0.3ppm of the 5H of rhamnose is noticed in the presence of a GlcpNAc aglycone, but a corresponding shift in the ^{13}C spectrum is not observed. Bock and coworkers[77] attribute this to a deshielding effect on the 5C proton by the O4 of the D ring, which, in the conformation proposed by them, are 2.64Å apart. Our structure obtained from the distance constraints does not show this 5C-4(O)D interaction, but the 5C proton is close to the C=O bond of the N-acetyl group, which may cause the observed deshielding. This deshielding should affect the chemical shift of the 3C proton as well, since it is closer to the carbonyl group, but no such effect is observed. The 2C' proton also shows a downfield shift (deshielding) of ~0.1 ppm, which is accompanied by an upfield shift of the 2C carbon chemical shift. This can be attributed to the compression of the 2C proton by the adjacent methyl group of the N-acetyl moiety. The differences in chemical shift of the 1C' proton in the two structures is due to the propyl aglycon.

The conformation of the B'-C' linkage of the heptasaccharide, which is distinct from that of the B-C linkage (the latter being representative of the polysaccharide) will bring the 1B' proton in the vicinity of the 4-OH group of the C' residue, as well as place the 2C' and 5B' protons closer together than in the B-C linkage. The observed deshielding of the 1B' and 5B' protons is consistent with these changes. The H6 of the B' residue of the heptasaccharide is shielded with comparison to the analogous proton in the polysaccharide. This can be attributed to the increased steric interactions of the methyl group in the polysaccharide. It is difficult to judge how significant the differences, which are of the order of 0.05ppm, are, but overall, the chemical shift differences of the

B' and C' ring support the existence of different conformations for the B-C and the B'-C' linkages.

Both the ^{13}C and ^{1}H spectrum of the residues B, C, D and A' residues in the heptasaccharide and polysaccharide show identical chemical shifts, within experimental error. The A ring, which is the terminal residue in the heptasaccharide shows an upfield shift of the carbon at the 2 position, which is the effect of glycosylation in the polysaccharide. The chemical shifts of the protons of the A ring in the heptasaccharide display an upfield shift of the 1, 2 and 3 protons. The 4A proton of the heptasaccharide is shifted downfield by 0.2 ppm, which can be explained by the absence of the interaction between the 4A proton and the methyl of the N-acetyl group that would occur in the polysaccharide.

The molecular dynamics trajectories calculated with the constraints removed samples essentially the same conformational space as the constrained dynamics, with a slightly greater range of freedom in the Φ/Ψ angles. The exceptions in this case are the B'-C' and A'-B' linkages, both of which move to the minimum energy conformations predicted by CHARMM during the heating and equilibration stages of the dynamics, and remain in these minima during the simulation.

ROESY buildup curves calculated using CROSREL display, without exception, the same trends as are observed in the experimental ROESY buildup, but once again the absolute magnitude differs considerably. The fit between buildup curves calculated from the constrained dynamics trajectories and experimental buildup is better, as evidenced by the R values, indicating that the simulations with the constraints provide a more accurate representation of the conformational space sampled.

The $^{3}J_{COCH}$ coupling constants across the glycosidic bond have also been measured by Bock et. al.[77] The values of these coupling constants are tabulated in Table IV.8 along with the average values calculated from the molecular dynamics trajectories.

The experimental data only include the coupling constant corresponding to the Ψ angle, i.e., C1-O1-C'X-H'X, where the prime refers to the aglycone. The measured value (3.6 Hz) for the B-C linkage corresponds almost exactly to the calculated value (3.7 Hz), but the other values are not in good agreement, particularly in the case of the C-D linkage, where a $^3J_{COCH}$ of 3.5Hz was measured, but the calculated coupling for both the constrained and unconstrained dynamics was ~ 1Hz.

The ω angle is another conformational variable and both the H6-H5 NOE as well as the $J_{5,6}$ can be used as additional points of comparison. A cross peak in the ROESY spectrum corresponding to a H5-$H6_S$ NOE is observed, while no contact is observed between the H5 and $H6_R$, a similar result to that observed in compounds **4-8**. This would indicate that the conformation is predominantly *gt*, which would place the $H6_S$ and the H7 protons in close proximity. However, as mentioned in the discussion of the ω angle in Chapter III, a *gt* conformation with an ω angle of 60° is not consistent with the $J_{5,6}$ values observed, and the measured values of the 5,6 coupling constants (~1.5 Hz ($J_{5,6S}$) and ~5.5 Hz ($J_{5,6R}$)) are identical to those observed in the oligosaccharides (**4-8**), consistent with an ω angle of ~90-100°. The dynamics simulations maintain the *gt* conformation with transitions to the *gg* conformation. The calculated values of the $J_{5,6}$ coupling constants (7.98 Hz ($J_{5,6R}$) 6.48 Hz ($J_{5,6S}$) for the unconstrained dynamics, and 9.54 Hz ($J_{5,6R}$) 5.34 Hz ($J_{5,6S}$) for the constrained dynamics) do not produce a good correlation with the experimental results.

More recent data on inhibition binding of various epitopes suggests that the BCD trisaccharide provides the minimum requirements for effective binding to complementary antibodies[206], which is consistent with earlier results in which the D residue (Glc*p*NAc) is judged to be essential. Single site functional group replacement of various ligands provides a "map" of the antibody site by delineating important antibody-epitope contacts. These studies show that the 4B, 4D, 4C hydroxyl groups and

Table IV.8. Experimental and calculated $^3J_{COCH}$ values (Hz) for the heptasaccharide (**9**).

Linkage		Exp.[77]	Dynamics (constrained)	Dynamics (no constraints)
B'-C'	Φ	-	3.4	3.3
	Ψ	3.6	4.2	3.1
A'-B'	Φ	-	3.4	3.5
	Ψ	3.5	4.9	1.9
D-A'	Φ	-	5.2	2.5
	Ψ	-	3.8	5.1
C-D	Φ	-	4.8	3.5
	Ψ	3.5	0.9	1.0
B-C	Φ	-	3.7	-
	Ψ	3.6	3.1	3.7
A-B	Φ	-	3.9	4.5
	Ψ	-	4.9	3.7

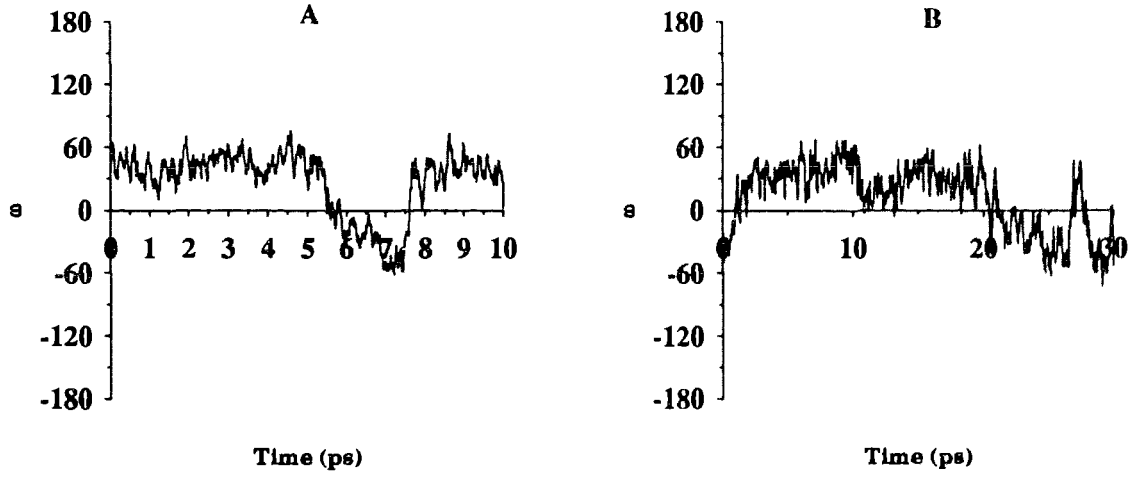


Figure 4.17. The variation of the ω angle during dynamics simulations of **9**. A) Constrained dynamics. B) Dynamics without constraints.

the 6C methyl display important interactions in the antibody binding. The overall conformation obtained from the NMR constraints is in accordance with the results from these binding studies[203,204]. If the Glc*p*NAc does in fact influence the conformation of the B-C linkage, as our results indicate, this might also explain why the antibody does not bind the ABC trisaccharide. Further inhibition studies with oligosaccharides containing 2-deoxy rhamnose has confirmed that deoxygenation at the 2 position of the C ring provides haptens with increased binding[204]. The C*D (where * denotes deoxy) shows increased binding over its oxygenated analog, CD, and the antibody displays a greater affinity for the trisaccharide BC*D than the BCD trisaccharide. The 2-OH of the C residue does in fact extend outward from the surface of the heptasaccharide in both the conformation proposed by Bock and coworkers as well as in the conformation proposed in this study and may result in unfavorable interaction with the protein.

Analysis of the chemical shift differences between the residues in the core of the heptasaccharide, i.e, the B, C, D and A' residues, and the polysaccharide suggests that there is a similarity in the conformation of these residues. An oligosaccharide with 24 residues was constructed using the average Φ/Ψ angles of each linkage from the constrained dynamics trajectories in an attempt to approximate a polysaccharide. The resulting structure is displayed in Figure 4.18. This structure forms a helix of extremely small pitch, approximately 9.2-9.5Å in which all of the N-acetyl groups from the Glc*p*NAc residues are turned in toward the center of the helix. A large number of polar groups, namely the 2 and 3 hydroxyls of the B ring, the 4 and 6 hydroxyls of the D ring, and the 2-OH of the C ring are all on the surface of the helix. Interestingly, the 2-OH of the C residues is one of the polar residues pointed outward from the surface of the helix. This hydroxyl is situated between the methyl groups of C and B residues, which are in all probability the major antigenic determinants in this region. This offers a simple explanation for the observed increased binding of oligosaccharides synthesized with

2-deoxy rhamnose substituted for the C residue - the area is made increasingly hydrophobic and thus increases the affinity of the antibody for the synthetic oligosaccharides tested. The 3C and 4C hydroxyl groups as well as the 6C methyl group are also on the outer surface of the helix, which would explain their activity in the inhibition studies[206].

A preliminary report of the crystal structure of an Fab fragment cocrystallized with a pentasaccharide ABCDA' and a trisaccharide BC*D (* denotes 2-deoxy rhamnose) has recently been published[207,208]. It represents one of the first antibody Fab-ligand complexes to be crystallized, and although no data on the conformation of either the ligand or the protein was reported, the Φ, Ψ angles of the bound trisaccharide are within ±20° of those calculated using the HSEA force field[209]. Furthermore, the structure of the BC*D-Fab complex indicates that, for this particular antibody combining site, a hydropnobic interaction of the methyl group of the N-acetyl function with a tyrosine residue is observed. Of more interest is the fact that the 2-OH group of the C residue would also be directed at the aromatic ring. The greater potency of the BC*D over the BCD trisaccharide is attributed, therefore, to the increased hydrophobic interaction with the tyrosine residue[210]. The result is at variance with our model of the B-C and C-D linkages in the free ligand.

Further comparison between bound and free oligosaccharide conformations is necessary in order to assess the validity of hypotheses of antibody-ligand interactions advanced in terms of conformations of the free ligands.

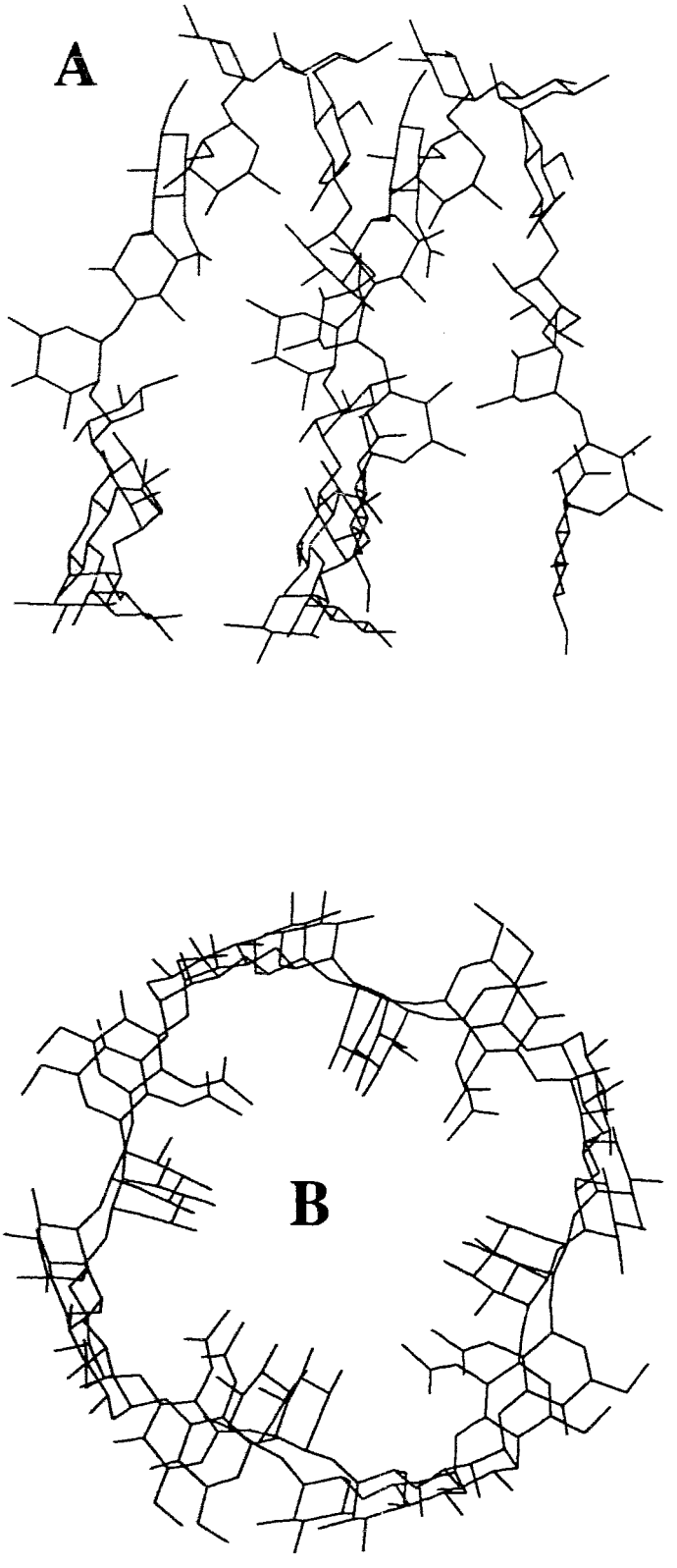


Figure 4.18. Oligosaccharide fragment constructed from the average Φ, Ψ angles obtained from the molecular dynamics simulations of **9**. A) Side on view. B) Top view.

CHAPTER V

CONCLUSIONS

In Chapter II of this thesis, the applicability of the CHARMM force field to oligosaccharide modeling and the viability of molecular dynamics as a conformational averaging procedure has been assessed. The minimum energy conformations predicted by CHARMM with the standard parameter set were not consistent with experimental data for two of the three compounds studied. Molecular dynamics simulations in H_2O samples the conformational space in the region of the starting minimum energy conformation, but if multiple minima exist, the dynamics calculations performed with CHARMM are unable to simulate transitions between the various minima even during dynamics simulations of the order of 100 ps[211]. Therefore, simulations starting from different local minima, followed by statistical averaging might give a better solution to the problem. Alternatively, dynamics simulations at higher temperatures might be necessary.

In Chapters III and IV, molecular dynamics simulations and 2D NMR techniques have been used to study the conformations of six oligosaccharides, a trisaccharide (**4**), a tetrasaccharide (**5**), a pentasaccharide (**6**), and two hexasaccharides (**7, 8**), corresponding to the cell-wall polysaccharide of *Streptococcus* Group A, and a heptasaccharide (**9**) corresponding to the *Shigella flexneri* variant Y O-lipopolysaccharide. Dynamics simulations were calculated with proton pairs constrained to the distances derived from ROESY experiments. The range of variation in the Φ, Ψ angles during the constrained dynamics simulation in H_2O is $\pm 30°$, and simulations calculated without the constraints show increased flexibility (of the order of $\pm 50°$).

The validity of the proposed conformations for compounds **4-8** is based on the NMR derived constraints. The accuracy of the distances calculated from cross peaks in 2D experiments is dependent on several factors, both experimental[212-214] and

theoretical[215-217]. Assuming that the pulses are measured to within ±0.1 μs, a data set size that provides good digital resolution is used, and the spectrum is correctly phased and baseline corrected, the signal to noise ratio is critical to the integrals. The effect of noise on the precision of a 2D integral has been examined[214], and in order to determine a distance to within 10 % accuracy, a signal to noise ratio of ~3 is required. This has been achieved for most of the cross peaks from which distances have been derived.

Two assumptions have been made in the calculation of interproton distances, the two spin approximation and the assumption of a single, isotropic correlation time. Studies have shown that any additional accuracy gained from measuring the buildup rates is insignificant for distances under 3.0 Å[217], which represents the majority of the distances obtained. Three spin models have led to the conclusion that although the distances obtained using a single correlation time are shorter than distances obtained from a more complex model, errors are roughly of the order of 20 % or less[215,216]. The effect of internal motion, however, is more pronounced[218], and has been demonstrated to be a factor in the relaxation behavior of carbohydrates[151]. The assumption that a single conformation contributes to the observed NOEs has been made, but the possibility of several minor conformations contributing to the NOE must be considered as well. Examples of oligosaccharides that exist in multiple conformations have been published[222,223], and it is entirely possible that the model here represents one of several that will account for the NMR data.

Entropy is thought to play a major role in the binding of proteins to their oligosaccharide ligands, and currently there are three theories that, while in agreement about the importance of the entropic contribution to protein-carbohydrate interactions, diverge in the origin of this contribution. These entropic contributions to the free energy of binding are thought to arise from either the solvent[219], the binding protein[220] or the carbohydrate ligand[221]. The results presented here do not specifically contradict nor

support any of the current theories on oligosaccharide protein interactions. Although it is easy to postulate the 2, 3 di-substitution of a rhamnopyranosyl (the A ring) residue as a reason for the observed limited flexibility in compounds **4-8**, the heptasaccharide **9** does not possess any branch points, and an extended conformation would provide less steric interactions than the conformation derived from the results of the ROESY experiments.

A comparison of the ^{1}H chemical shifts reveals no significant differences between the chemical shifts of the *Streptococcus* polysaccharide and the chemical shifts corresponding to internal sequences in the hexasaccharides **7** and **8**, nor between that of the heptasaccharide **9** and the parent *Shigella flexneri* Variant Y lipopolysaccharide O-chain. Based on this comparison, extrapolation of the average conformations derived from the constrained molecular dynamics simulations to larger structures appears to be justified.

In conclusion, Molecular Dynamics represents a viable method of conformational averaging, and when used in combination with NMR derived constraints, the behavior of oligosaccharides in solution can be simulated.

REFERENCES

1. Paulson, J. C., *Trends Biochem. Sci.*, **14**, 272-276, (1983).

2. Feizi, T., *Trends Biochem. Sci.*, **16**, 84-86, (1991).

3. Jentoft, N., *Trends Biochem. Sci.*, **15**, 291-294, (1990).

4. Sharon, N., Complex Carbohydrates, their Chemistry, Biosynthesis and Functions, Addison-Wesley Inc., Massachusets, 1975.

5. Hughes, R. C., *Essays in Biochemistry*, **11**, 1-36, (1975).

6. Alberts, B., Bray, D, Lewis, J. Raff, M., Roberts, K. & Watson, J. D., Molecular Biology of the Cell, Garland Publishing, Inc., New York, 1989.

7. Lemieux, R. U., IUPAC Frontiers in Chemistry, (L. J. Laidler Ed.), Pergamon Press, New York, pp 3-24 (1982).

8. Feizi, T., *Biochem. J.*, **245**, 1-11, (1987).

9. Feizi, T., *Nature*, **314**, 53-57, (1985).

10. E. J. Goldstein (Ed), Carbohydrate-Protein Interaction, ACS symposium Series, American Chemical Society, Washington D. C., (1979).

11. Clarke, A. E. & Wilson, I. A. (Eds), Carbohydrate-Protein Interaction, *Current Topics in Microbiology and Immunology*, **139**, Springer-Verlag, Heidelberg, (1988).

12. Springer, T. A., *Nature*, **346**, 425-433, (1990).

13. Stoolman, L. M., *Cell*, **56**, 907-910, (1989).

14. Bevilacqua, M. P. & Nelson, R. M., *J. Clin. Invest.*, **91**, 379-387, (1993).

15. Lasky, L. A., *Science*, **258**, 964-969, (1992).

16. Springer, T. A. & Lasky, L. A., *Nature*, **349**, 196-197, (1991).

17. Lowe, J. B., Stoolman, L. M., Nair, R. P., Larsen, R. D., Berhend, T. L. & Marks, R. M., *Cell*, **63**, 475-484, (1990).

18. Phillips, M. L., Nudelman, E., Gaeta, F. C., Perez, M., Singhal, A. K., Hakomori, S. & Paulson, J. C., *Science*, **250**, 1130-1132, (1990).

19. Walz, G., Aruffo, A., Kolanus, W., Bevilacqua, M. & Seed, B., *Science*, **250**, 1132-1135, (1990).

20. Tiemeyer, M., Swiedler, S. J., Ishihara, M., Moreland, M., Schweingruber, H., Hirtzer, P. & Brandley, B. K., *Proc. Natl. Acad. Sci. USA.*, **88**, 1138-1142, (1991).

21. Goelz, S. E., Hession, C., Goff, D., Griffiths, B., Tizard, R., Newman, B., Chi-Rosso, G. & Lobb, R., *Cell*, **63**, 1349-1356, (1990).

22. Wasserman, P. M., *Annu. Rev. Cell Biol.*, **3**, 109-142, (1987).

23. Ruoslahti, E., *J. Biol. Chem.*, **264**, 13369-13372, (1989).

24. Spohr, U., Paszkiewiecz-Hnatiw, E., Morishima, N. & Lemieux, R. U., *Can. J. Chem.*, **70**, 254-271, (1992).

25. F.W. Lichtenthaler, S. Immel, Proceedings of the European Chemoreception Research Organization (ECRO) Meeting, Reims, (1991).

26. Lee, C.K., *Adv. Carbohydr. Chem. Biochem.*, **45**, 199-351, (1987).

27. Kreis, U., Ph.D. Thesis, Technical University of Darmstadt, Darmstadt, Germany, (1992).

28. Roberts, S. M. (Ed), Molecular Recognition: Chemical and Biochemical Problems, The Royal Society of Chemistry, (1989).

29. Berger, M., *Iris universe*, **23**, 28-32, (1993).

30. Lerner, R. A., Benkovic, S. J. & Schultz, P. G., *Science*, **252**, 659-667, (1991)

31. Pauling, L., *Am. Sc.*, **36**, 51-58, (1948).

32. Pollack, S. J., Jacobs, J. W. & Schultz, P. G., *Science*, **234**, 1570-1573, (1986).

33. Tramontano, A., Janda, K. D. & Lerner, R. A., *Science*, **234**, 1566-1570, (1986).

34. Morris, E. R., Rees, D. A., Thom, D. & Welsh, E. J., Progress in Clinical and Biological Research, (Marchesi, V. T., Ginsburg, v., Robbins, P. W. & Fox, C. F. Eds.), Alan R. Liss Inc, New York, 1978. Reprinted from J. Supramol. Struc., **6**, 259-274, (1977).

35. Stevens, E. S. & Sathyanarayana, B. K., *Carbohydr. Res.*, **166**, 181-193, (1987).

36. Stevens, E. S. & Sathyanarayana, B. K., *J. Am. Chem. Soc.*, **111**, 4149-4154, (1989).

37. Duda, C. A. & Stevens, E. S., *J. Am. Chem. Soc.*, **112**, 7406, (1990).

38. Duda, C. A. & Stevens, E. S., *Carbohydr. Res.*, **206**, 347-351, (1990).

39. Stevens, E. S. & Duda, C. A., *J. Am. Chem. Soc.*, **113**, 8622-8627, (1991).

40. Stevens, E. S., *Carbohydr. Res.*, **244**, 191-195, (1993).

41. Quiocho, F. A., *Pure & Appl. Chem.*, **61**, 1293-1306, (1989).

42. Cygler, M., Rose, D. R. & Bundle, D. R., *Science*, **253**, 442-445, (1991).

43. Deisenhofer, J., *Biochemistry*, **20**, 2361-2370, (1981).

44. Paulsen, H., *Angew. Chem. (Internat.)*, **29**, 823-839, (1990).

45. Lowe, I. J. & Norberg, R. E., *Phys. Rev.*, **107**, 46-61, (1957).

46. Ernst, R. R. & Anderson, W. A., *Rev. Sci. Instrum.*, **37**, 93-102, (1966).

47. Proctor, W. G. & Yu, F. C., *Phys. Rev.*, **77**, 717, (1950).

48. Dickinson, W. C., *Phys. Rev.* **77**, 736-737, (1950).

49. Karplus, M., *J. Chem. Phys.*, **30**, 11-15, (1959).

50. Aydin, R., Loux, J-P. & Gunther, H., *Angew. Chem. (Internat.)*, **21**, 449, (1982).

51. Tvaroska, I., Mazeau, K., Blanc-Muesser, M., Lavaitte, S., Driguez, H. & Taravel, F. R., *Carbohyd. Res.*, **229**, 225-231, (1992).

52. Tvaroska, I., Hricovini, M. & Petrakova, E., *Carbohydr. Res.*, **189**, 359-362, (1989).

53. Mulloy, B., Frenkiel, T. A. & Davies, D. B., *Carbohydr. Res.*, **184**, 39-46, (1988).

54. Maudsley, A. A. & Ernst, R. R., *Chem. Phys. Lett.*, **50**, 368-372, (1977).

55. Hasnoot, C. A. G., De Leeuw, F. A. A. M. & Altona, C., *Tetrahedron*, **36**, 2783-2792, (1980).

56. Kalk, A. & Berendsen, H. J. C., *J. Magn. Reson.*, **24**, 343-366, (1976)

57. Anet, F. A. L. & Bourn, A. J. R., *J. Am. Chem. Soc.*, **87**, 5250-5251, (1965).

58. Overhauser, A. W., *Phys. Rev.*, **92**, 411-415, (1953).

59. Solomon, I. & Bloembergen, N., *J. Chem. Phys.*, **25**, 261-266, (1956).

60. Wirtz, J., Jain, P. L. & Batdorf, R. L., *Phys. Rev.*, **102**, 920-921, (1956).

61. Bell, R. A. & Saunders, J. K., *Can. J. Chem.*, **48**, 1114-1122, (1970).

62. Solomon, I., *Phys. Rev.*, **99**, 559-565, (1955).

63. Noggle, J. H. & Schirmer, R. E., The Nuclear Overhauser Effect, Academic Press, New York, (1971).

64. Richarz, R., Wuthrich, K., *J. Magn. Reson.*, **30**, 147-150, (1978).

65. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R., *J. Chem. Phys.*, **71**, 4546-4553, (1979).

66. Macura, S. & Ernst, R. R., *Mol. Phys.*, **41**, 95-117, (1980).

67. Bothner-By. A. A., Stephens, R. L., Lee, J-M., Warren, C. D. & Jeanloz, R. W., *J. Am. Chem. Soc.*, **106**, 811-813, (1984).

68. Bax, A. & Davis, D. B., *J. Magn. Reson.*, **63**, 207-213, (1985).

69. Pople, J. A. & Beveridge, D. L., Approximate Molecular Orbital Theory, McGraw-Hill Inc., USA, (1970).

70. Clark, T., A Handbook of Computational Chemistry, John Wiley & Sons Inc., Canada, (1985).

71. Hehre, W. J., Radom, L., Schleyer, P. v.R. & Pople, J. A., Ab Initio Molecular Orbital Theory, John Wiley & Sons Inc., Canada, (1986).

72. Garret, E. C. & Serriani, A. S., *Carbohydr. Res.*, **206**, 183-191, (1990).

73. Garret, E. C. & Serriani, A. S., Computer Modeling of Carbohydrate Molecules, (French A. D. & Brady, J. W., Eds.), American Chemical Society, Washington D. C., (1990).

74. Burkert, U. & Allinger, N. L., Molecular Mechanics, American Chemical Society, Washington D. C., (1982).

75. McQuarrie, D. A., Quantum Chemistry, University Science Books, California, USA, (1983).

76. Maskill, H., The Physical Basis of Organic Chemistry, Oxford University Press, New York, (1985).

77. Bock, K., Josephson, S. & Bundle, D. R., *J. Chem. Soc. Perkin Trans.* II, 59-70, (1982).

78. Lemieux, R. U. & Koto, S., *Tetrahedron*, **30**, 1933-1944, (1974).

79. Lemieux, R.U., Molecular Rearrangements (de Mayo, P., Ed.), Part 2, Wiley InterScience, New York, 709-768, (1964).

80. Romers, C., Altona, C., Buys, H. R. & Havinga, E., Topics in Stereochemistry, Vol. 4 (Eliel, E. L. & Allinger, N. L., Eds.), Wiley InterScience, New York, 39-97, (1969).

81. Wolfe, S., Whangbo, M. H. & Mitchell, D. H., *Carbohydr. Res.* **69**, 1-26, (1979).

82. Kirby, A. J., The Anomeric Effect and Related Stereoelectronic Effects at Oxygen, Springer, Berlin, New York, (1983).

83. Tvaroska, I. & Bleha, T., *Adv. Carbohydr. Chem. Biochem.*, **47**, 45-123, (1989).

84. de Hoog, A. J., Buys, H. R., Altona, C. & Havinga, E., *Tetrahedron*, **25**, 3365-3375, (1969).

85. Wiberg, K., Murcko, M. A., Laidig, K. E. & MacDougall, P. J., *J. Phys. Chem.*, **94**, 6956-6959, (1990).

86. Wolfe, S., *Acc. Chem. Res.*, **5**, 102-111, (1972).

87. Norskov-Lauritsen, L. & Allinger, N. L., *J. Comput. Chem.*, **5(4)**, 326-335, (1984).

88. Aped, P., PhD Thesis, Tel-Aviv University, Tel-Aviv, Israel, (1988).

89. Aped, P., Apelorg, Y., Ellencweig, A., Fuchs, B., Goldberg, I., Karni, M. & Tartakovsky, E., *J. Am. Chem. Soc.*, **109**, 1486-1495, (1987).

90. Tran, V., Buleon, A., Imberty, A. & Perez, S., *Biopolymers*, **28**, 679-690, (1989).

91. Jeffrey, G. A. & Taylor, R., *J. Comput. Chem.*, **1**, 99-109, (1980).

92. Homans, S. W., *Biochemistry*, **29**, 9110-9118, (1990).

93. Stuike-Prill, R., Ph. D. Thesis, Univerität Oldenburg, Oldenburg, Germany, (1991).

94. Stuike-Prill, R. & Meyer, B., unpublished results.

95. Stuike-Prill, R. & Meyer, B., *Eur. J. Biochem.*, **194**, 903-919, (1990).

96. Lemieux, R. U., Bock, K., Delbaere, L.T.J., Koto, S. & Rao, V.S., *Can. J. Chem.*, **58**, 631-652, (1980).

97. Thogersen, H., Lemieux, R. U., Bock, K. & Meyer, B., *Can. J. Chem*, **60**, 44-57, (1982).

98. Meyer, B., Int. Carbohydr. Symp., Vancouver, Canada, Abstr. II/25, (1982).

99. Kitaigorodsky, A. I., *Tetrahedron*, **14**, 230-236, (1961).

100. Kitaigoradsky, A. I., *Chem. Soc. Rev.*, **7**, 133-163, (1978).

101. Burkert, U., *Tetrahedron*, **33**, 2237, (1977).

102. (a) Allinger, N. L., Yuh, Y. H. & Lii, J-H., *J. Am. Chem. Soc.*, **111**, 8552-8566, (1989). (b) Allinger, N. L. & Lii, J-H., *J. Am. Chem. Soc.*, **111**, 8566-8575, (1989). (c) Lii, J-H. & Allinger, N. L., *J. Am. Chem. Soc.*, **111**, 8575-8582, (1989). (d) Allinger, N. L., Rahman, M. & Lii, J-H., *J. Am. Chem. Soc.*, **112**, 8293-8307, (1990).

103. Weiner P. K. & Kollman P. A., *J. Comput. Chem.*, **2**, 287-303 (1981).

104. Brooks, B. R. Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M., *J. Comput. Chem.*, **4**, 187-217, (1983).

105. Dowd, M. K., Zeng, J., French, A. D. & Reilly, P. J., *Carbohydr. Res.*, **230**, 223-244, (1992).

106. Dowd, M. K., French, A. D. & Reilly, P. J., *Carbohydr. Res.*, **233**, 15-34, (1992).

107. Dowd, M. K., Reilly, P. J. & French, A. D., *J. Comput. Chem.*, **13**, 102-114, (1992).

108. Allinger, N. L., Operating Instructions for the MM2 and MMP2 Programs, QCPE, Indiana University, (1985).

109. Wiberg, K. B., *J. Am. Chem. Soc.*, **87**, 1070-1078, (1965).

110. Fletcher, R. & Powell, M. J. D., *Compt. J.*, **6**, 163-167, (1963).

111. Boyd, R. H., *J. Chem. Phys.*, **49**, 2574-2583, (1968).

112. Bock, K., *Pure & Applied Chem.*, **5**, 605-622, (1983).

113. Cumming, D. A. & Carver, J. P., *Biochemistry*, **26**, 6664-6676, (1987).

114. Cumming, D. A. & Carver, J. P., *Biochemistry*, **26**, 6676-6683, (1987).

115. Imberty, A., Tran, V. & Perez, S., *J. Comput. Chem.*, **11**, 205-216, (1982).

116. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Augusta, H., Teller, A.& Teller, E., *J. Chem. Phys.*, **21**, 1087-1092, (1953).

117. Peters, T., Meyer, B., Stuike-Prill, R., Somorjai, R. & Brisson, J. R., *Carbohydr. Res.*, **238**, 49-73, (1993).

118. Lemieux, R. U., Delbaere, L. T. J., Beierbeck, H. & Spohr, U., Host-Guest Molecular Interactions: From Chemistry to Biology, John Wiley & Sons, Toronto, Canada, (1991).

119. Rahman, A. & Stillinger, F. H., *J. Chem. Phys.* **55**, 3336-3359, (1971).

120. Stillinger, F. H. & Rahman, A., *J. Chem. Phys.*, **57**, 1281-1292, (1972).

121. Stillinger, F. H. & Rahman, A., *J. Chem. Phys.*, **60**, 1545-1557 (1974).

122. Rahman, A., *Phys. Rev.*, **136**, A405-411, (1964).

123. Verlet., L., *Phys. Rev.*, **159**, 78-103, (1967).

124. Homans,S. W., Pastore, A., Dwek, R. A. & Rademacher, T. W., Biochemistry, **26**, 6649-6655, (1987).

125. CHARMM User's Guide, Polygen Corp., Waltham, M.A. © 1987,

126. Brisson, J. R. & Carver, J. P., *Biochemistry*, **22**, 1362-1368, (1983).

127. Brisson, J. R. & Carver, J. P., *Can. J. Biochem. Cell Biol.*, **61**, 1067-1078, (1983).

128. Brisson, J. R. & Carver, J. P., *Biochemistry*, **22**, 3671-3680, (1983).

129. Cumming, D. A., Dime, D. S., Grey, A. A., Krepinsky, J. J. & Carver, J. P., *J. Biol. Chem.*, **261**, 3208-3213, (1986).

130. Homans, S.W., Dwek, R. A. & Rademacher, T. W., *Biochemistry*, **26**, 6553-6560, (1987).

131. Homans, S. W., Dwek, R. A., Fernandes, D. L. & Rademacher, T. W., *FEBS Lett.*, **150**, 503-506, (1983).

132. Homans, S. W., Dwek, R. A., Fernandes, D. L. & Rademacher, T. W., *FEBS Lett.*, **164**, 231-235, (1983).

133. Jeffrey, G. A., McMullan, R. K. & Takagi, S., *Acta Cryst.*, **B33**, 728-737, (1977).

134. Warin, V., Baert, F., Fouret, R., Strecker, G., Spik, G., Fournet, B. & Montreuil, J., *Carbohydr. Res.*, **76**, 11-22, (1979).

135. Homans, S.W., Dwek, R. A. & Rademacher, T. W., *Biochemistry*, **26**, 6571-6578, (1987).

136. Rivera-Sagredo, A., Solis, D., Diaz-Mauriño, T. , Jiminéz-Barbero, J. & Martin-Lomas, M., *Eur. J. Biochem.*, **197**, 217-228, (1991).

137. Rivera-Sagredo, A., Jiminéz-Barbero, J. & Martin-Lomas, M., *Carbohydr. Res.*, **221**, 37-47, (1991).

138. Hirotsu, K. & Shimada, A., *Bull. Chem. Soc. Jap.*, **47**, 1872-1879, (1974).

139. Berry, J. M., Hall, L. D. & Wong, K. F., *Carbohydr. Res.*, **56**, C16-C20, (1977).

140. Kline, P. C., Serriani, S. A., Huang, S-G., Hayes, M. & Barker, R., *Can. J. Chem.*, **68**, 2171-2182, (1990).

141. Hayes, M., Serriani, S. A. & Barker, R., *Carbohydr. Res.*, **100**, 87-101, (1982).

142. Poppe, L. & van Halbeek, H., *J. Magn. Reson.*, 93, 214-217, (1991).

143. Polygen Corp., 200 Fifth Avenue, Waltham, M. A.

144. Reimer, K. B. & Pinto, B. M., *J. Chem. Soc. Perkin Trans.* I, 2103-2111, (1988).

145. Wiberg, K. B. & Murcko, M. A., *J. Am. Chem. Soc.*, **111**, 4821-4828, (1989).

146. de Hoog, A. J., *J. Org. Magn. Reson.*, **6**, 233, (1974).

147. Carver, J. P., Mandel, D., Michnick, S. W., Imberty, A. & Brady, J. W., Computer Modeling of Carbohydrate Molecules, (French A. D. & Brady, J. W., Eds.), American Chemical Society, Washington D. C., (1990).

148. Rasmussen, K., *Acta Chem. Scand.*, **36A**, 323-327, (1982).

149. Melberg, S. & Rasmussen, K., *Carbohydr. Res.*, **78**, 215-224, (1980).

150. Brady, J. W., *J. Am. Chem. Soc.*, **108**, 8153-8160, (1986).

151. Hrincovíni, M., Shah, R. N. & Carver, J. P., *Biochemistry*, **31**, 10018-10023, (1992).

152. Hirotsu, K., Shimada, A., *Bull. Chem. Soc. Jap.*, 47, 1872-1879, (1974)

153. Mo, F., *Acta Chim. Scand.*, **A33**, 207-218, (1979). (b) Mo, F. & Jensen, L. H., *Acta Cryst.*, B**31**, 2867-2873, (1975).

154. Bisno. A. L., Principles and Practice of Infectious Diseases, (Mandell, G. L., Douglas, R. G. & Bennet, E. J. (Eds)), 2nd Edn., Wiley, New York, pp 1133-1142, (1985).

155. Fischetti, V. A., *Sci. Am.*, 59-65, (1991).

156. Dale, J. B. & Beachey, E. H., *J. Exp. Med.*, **162**, 583-591, (1985).

157. Cunningham, M. W., Antone, S. M., Guliza, J. M., McManus, B. M., Fischetti, V. A. & Gauntt, C. J., *Proc. Natl. Acad. Sci. USA*, **89**, 1320-1324, (1992).

158. Goldstein, I., Rebeyrotte, P., Parlenbas, J. & Halpern, B., *Nature*, **219**, 866-868, (1968).

159. Goldstein, I., Scebat, L., Renais, J., Hadjinsky, P. & Dutartre, J., *Israel. J. Med. Sci.*, **19**, 483-490, (1983).

160. Scott, J. R., Molecular Basis of Bacterial Pathogenesis, **IX**,(Iglewski, B. H. & Clark, V. L. Eds.), Academic Press Inc., London, pp 177-199, (1990).

161. Fischetti, V. A., Jarymowycz, M., Jones, K. F. & Scott, J. R., *J. Exp. Med.*, **164**, 971-980, (1986).

162. Reimer, K. B., Harris, S. L., Varma, V. & Pinto, B. M., *Carbohydr. Res.*, **228**, 399-414, (1992).

163. Pinto, B. M., Reimer, K. B. & Tixidre, A., *Carbohydr. Res.*, **210**, 199-219, (1991).

164. Andrews, J. A. & Pinto, B. M., *J. Chem. Soc. Perkin Trans.* I, 1785-1792, (1990).

165. Andrews, J. A., M.Sc. Thesis, Simon Fraser University, Burnaby, Canada, (1987).

166. Reimer, K. B., M.Sc. Thesis, Simon Fraser University, Burnaby, Canada, (1987).

167. Reimer, K. B., Ph. D. Thesis, Simon Fraser University, Canada, (1991).

168. Marino-Albernas, J.-R., Harris, S. L., Varma V. & Pinto, B. M. *Carbohydr. Res.*, (in press).

169. Pinto, B. M., Carbohydrate Antigens, (Garegg, P. J. & Lindberg, A. A., Eds.), ACS Symposium Series, American Chemical Society, Washington D. C., pp 112-131, (1993).

170 Coligan, J. E., Kindt, T. J. & Krause, R. M., *Immunochem.*, **15**, 755-760, (1978).

171. Coligan, J. E., Schnute, W. C., Jr. & Kindt, T. J., *J. Immun.*, **114**, 16554-1658, (1975).

172. Reimer, K. B., Gidney, M. A., Bundle, D. R. & Pinto, B. M., *Carbohydr. Res.*, **232**, 131-142, (1992).

173. Braunschweiler, L. & Ernst, R. R., *J. Magn. Reson.*, **53**, 521-528, (1983).

174. Bax, A. *J. Mag, Reson.*, **77**, 134-147, (1988).

175. Molecular Simulations Inc., 200 Fifth Avenue, Waltham, M. A..

176. Leeflang, B. R., Ph.D. Thesis, Utrecht University, Netherlands, (1991).

177. Leeflang, B. R.& B. R., Kroon-Batenburg, L. M. J., *J. Biomol. NMR*, **2**, 495-518, (1992).

178. Hwang, T-L. & Shaka, A. J., *J. Am. Chem. Soc.*, **114**, 3157-3159, (1992).

179. Widmalm, G., Byrd, R. A. & Egan, W., *Carbohydr. Res.*, **229**, 195-211, (1992).

180. Kochetov, N. K., Lipkind, G. M., Shashkov, A. S. & Nifant'ev, N. E., *Carbohydr. Res.*, **221**, 145-168, (1991).

181. Lipkind, G. M., Shashkov, A. S., Mamyan, S. S. & Kochetov, N. K., *Carbohydr. Res.*, **181**, 1-12, (1988).

182. Shaskov., A. S., Lipkind, G. M., Knirel, Y. A. & Kochetkov, N. K., *Magn. Reson. Chem.*, **26**, 735, 1988.

183. Dais, P. & Perlin, A. S., *Carbohydr. Res.*, **146**, 177-191, 1986.

184. De Bruyn, A., *J. Carbohydr. Chem.*, 10(2), 159-180, 1991.

185. Marchessault R. H. & Perez, S., *Biopolymers*, **18**,2369-2374, (1979).

186. Hori, H., Nakajima, T., Nishida, Y., Ohrui, H. & Meguro, H., *J. Carbohydr. Chem.*, **5**, 585-600, (1986).

187. Gagnaire, D., Horton, D. & Taravel, F. R., *Carbohydr. Res.*, **27**, 363-372, (1973).

188. Nishida, Y., Ohrui, H. & Meguro, H., *Tetrahedron Lett.*, 1575-1578, (1974).

189. Altona, C. & Hasnoot, C. A. G., *Org. Magn. Reson.*, **417**-429, (1980).

190. Boyd, J., Porteous, R., Soffe, N. & Delepierre, M., *Carbohydr. Res.*, **139**, 35-46, (1985).

191. Nishida, Y., Hori, H., Ohrui, H. & Meguro, H., *Carbohydr. Res.*, **170**, 106-111, (1987).

192. Nishida, Y., Hori, H., Ohrui, H. & Meguro, H., *J. Carbohydr. Chem.*, **7**, 239-250, (1988).

193. Huang, D. H., Krishna, N. R. & Pritchard, D. G., *Carbohydr. Res.*, **155**, 193-199, (1986).

194. Ritchie, R. G. S., Natsuko, C. & Perlin, A. S., *Can J. Chem.*, **54**, 2301-2309, (1976).

195. Cheney, B. V., *J. Am. Chem. Soc.*, **90**, 5386-5390, (1968).

196. Koch, H. J. & Perlin, A. S., *Carbohydr. Res.*, **15**, 403-410, (1970).

197. Braun, D. G., Herbst, H. & Schalch, W., The Immune System, Vol. 2, Karger, Basel, (1981).

198. Carlin, N. I. A., Lindberg., A. A., Bock, K. & Bundle, D. R., *Eur. J. Biochem.*, **139**, 189-194, (1984).

199. Kenne, L., Lindberg, B., Petersson, K., Katzenellenbogen, E. & Romanowska, E., *Carbohydr. Res.*, **56**, 363-370, (1977).

200. Pinto, B. M., Morissette, D. G. & Bundle, D. R., *J. Chem. Soc.Perkin Trans.* I, 9-14, (1987).

201. Pinto, B. M., Reimer, K., B., Morissette, D. G. & Bundle, D. R., *J. Org. Chem.*, 54, 2650-2656, (1990).

202. Pinto, B. M., Reimer, K., B., Morissette, D. G. & Bundle, D. R., *J. Chem. Soc. Perkin Trans.* I, 293-299, (1990).

203. Bundle, D. R., *Pure & Appl. Chem.*, 61, 1171-1180, (1989).

204. Hanna, H. R. & Bundle, D. R., *Can. J. Chem.*, **71**, 125-134, (1993).

205. Jansson, P-E., Kenne, L. & Wehler, T., *Carbohydr. Res.*, **166**, 271-282, (1987).

206. Bundle, D. R., Auzanneau, F.-I., Oomen, R. & Sigurskold, B. W., XV1th International Carbohydrate Symposium, Paris, France, Abstract B115, p448, (1992).

207. Vyas, M. N., Vyas, Bundle, D. R. & Quiocho, F. A., Symposium on Conformational Studies of Oligosaccharides, Polysaccharides and Glycoconjugates, Le Croisic, France, Abstract II-4, (1992).

208. Vyas, M. N., Vyas, K. N, Meikle, P. J., Sinnott, B., Pinto. B. M., Bundle, D. R. & Quiocho, F. A., *J. Mol. Biol.*, **231**, 133-136, (1993).

209. Bundle, D. R., Personal communication.

210. Vyas, M. N., Vyas, K. N, Meikle, P. J., Sinnott, B., Pinto. B. M., Bundle, D. R. & Quiocho, F. A., manuscript in preparation.

211. Varma, V., unpublished data.

212. Derome, A. E., Modern NMR Techniques for Chemistry Research, Pergamon Press, Oxford, (1987).

213. Weiss, G. H. & Ferretti, J. A., *J. Magn. Reson.*, **55**, 397-407, (1983).

214. Nadjari, R. & Grivet, J.-PH., *J. Magn. Reson.*, **98**, 259-270, (1992).

215. Thomas, P. D., Basus, V. J. & James, T. L., *Proc. Natl. Acad. Sci. USA*, **88**, 1237-1241, (1991).

216. Keepers, J. W. & James, T. L., *J. Magn. Reson.*, **57**, 404-426, (1984).

217. Clore, G. M. & Groneneborn, A. M., *J. Magn. Reson.*, **61**, 158-164, (1985).

218. Neuhaus, D. & Williamson, M. P., The Nuclear Overhauser Effect in Structural and Conformational Analysis, VCH Publishers Inc, Great Britain, (1989).

219. Lumry, R. & Rajender, S., *Biopolymers*, **9**, 1125-1277, (1970).

220. Eftink, M. R., Anusiem, A. C. & Biltonen, *Biochemistry*, **22**, 3384-3896, (1983).

221. Carver, J. P., *Pure & Appl. Chem.*, **65**, 763-770, (1993).

222. Baumann, H., Brisson, J.-R., Michom F., Pon, R. & Jennings, H. J., *Biochemistry*, **32**, 4007-4013, (1993).

223. Brisson, J.-R., Bauman, H., Imberty, A., Perez, S. & Jennings, H. J., *Biochemistry*, **31**, 4996-5004, (1992).

224. Jeffrey, G. A., Anomeric Effect Origin and Consequences, (Szarek, W. A. & Horton, D. eds.), American Chemical Society, Washington D. C., (1979).

225. Pinto, B. M., Schlegel, H. B. & Wolfe, S., *Can. J. Chem.*, **65**, 1658-1662, (1987).

226. Booth, H., Khedhair, A. K. & Readshaw, A. S., *Tetrahedron*, **23**, 4699-4723, (1987).

227. Praly, J.-P. & Lemieux, R. U., *Can. J. Chem.*, **65**, 213-223, (1987).

228. Tvaroska, I., personal communication to Pinto, B. M.

229. Carver, J. P., 75th Canadian Chemical Conference and Exhibition, Edmonton, Canada, Abstract BP-D1, (1992).

230. Homans, S. W. & Rutherford, T. J., Symposium on Conformational Studies of Oligosaccharides, Polysaccharides and Glycoconjugates, Le Croisic, France, Abstract I-1, (1992).

APPENDIX

Source code for the programs to calculate the NOEs and T_1s. Written by Chris Schafmeister.

```
      PROGRAM NOE

      include       'general.i'

C
C
C     stitle is a comment attatched to the file
C     tc, w0, rs hold constants read from the input file
C     nproto is the number of protons in the input file
C     nsolvp is the number of protons to solve nOe's for
C     asolvp is an array containing the id's of the protons to solve for
C     nframe is the number of dynamics frames in the input file
C     ddict  is a dictionary containing the names of the protons
C     weight contains the weighting factor of this frames NOE's
C     danoes is an array containing the nOe's of each proton due
C            to irradiating another proton
C     daslvd is a 2D array containing the sum of weighted NOE's for
C            every dynamics frame read sofar
C

      double precision    tc, w0, rs
      integer             nproto, nsolvp, asolvp(maxpro), nframe
      record /dict/       ddict
      double precision    weight, danoes(maxpro)
      double precision    daslvd(maxslv,maxpro), twght

      character*80        stitle
      character*10        sname
      integer             u, i, iframe, islv

      u = 5

C           Load the header and initialize the constants for calculating
C           NOE's

      call header( u, stitle, tc, w0, rs, nproto, ddict, nsolvp, asolvp, nframe )
      call cconsts( tc, w0, rs )
      cipcnt = nproto
```

```
      if ( cipcnt.gt.maxpro ) then
            write ( 6, 300 ) maxpro
            write ( 6, 310 ) cipcnt
            goto 9999
300         format ( 'ERROR: NOECALC can only support ', i1, ' protons.' )
310         format ( '       The input file has ', i1, ' protons.' )
      endif

C           ZERO all entries of the daslvd array
      do islv= 1, nsolvp
            do i= 1, cipcnt
                  daslvd(islv,i) = 0.0
            end do
      end do

C           Read in each frame and calculate NOE's for each proton
C           to solve for

      twght = 0.0
      do iframe=1, nframe

C           Load a frame, and calculate all 1/r**6

            call frame( u, ddict, weight)
            twght = twght + weight
            call idists

C                 Solve NOE's for each requested proton

            do islv=1, nsolvp

                  call getnoe( danoes, asolvp(islv) )

C                       Add weighted NOE's to totals
                  do i=1, cipcnt
                        daslvd(islv,i) =
     +                   daslvd(islv,i)
     +                     +weight*danoes(i)
                  end do
            end do
      end do

C           Divide all NOE's by the total weight

      do islv=1, nsolvp
```

```
            do i=1, cipcnt
                  daslvd(islv,i) = daslvd(islv,i)/twght
            end do
      end do

C           Display results

      write ( 6, 90 ) stitle
90    format ( 'TITLE: ', A80 )

      do islv=1, nsolvp

            call dsname( ddict, sname, asolvp(islv) )

            write ( 6, 100 ) sname
100         format ( 'NOEs due to irradiating proton ', a )

            do i= 1, cipcnt
                  if ( i.ne.asolvp(islv) ) then
                        call dsname( ddict, sname, i )
                        write ( 6, 110 ) sname, daslvd(islv,i)
110                     format ( 'Proton: ', a, ' noe=', F10.6 )
                  end if
            end do

            write ( 6, 120 )
            write ( 6, 130 )
120         format ( '----------------------------------------' )
130         format ( ' ' )

      end do

9999  continue

      END
```

```
      PROGRAM t1calc

      include       'general.i'

C
C
C     stitle is a comment attatched to the file
C     tc, w0, rs hold constants read from the input file
C     nproto is the number of protons in the input file
C     nsolvp is the number of protons to solve nOe's for
C     asolvp is an array containing the id's of the protons to solve for
C     nframe is the number of dynamics frames in the input file
C     ddict  is a dictionary containing the names of the protons
C     weight contains the weighting factor of this frames NOE's
C     dat1s is an array containing the t1's of each proton due
C           to irradiating another proton
C     daslvd is a 1D array containing the sum of weighted T1's for
C           every dynamics frame read sofar
C

      double precision      tc, w0, rs
      integer               nproto, nsolvp, asolvp(maxpro), nframe
      record /dict/         ddict
      double precision      weight, dat1s(maxpro)
      double precision      daslvd(maxpro), twght

      character*80          stitle
      character*10          sname
      integer               u, i, iframe, islv

      u = 5

C           Load the header and initialize the constants for calculating
C           T1's

      call header( u, stitle, tc, w0, rs, nproto, ddict, nsolvp, asolvp, nframe )
      call cconsts( tc, w0, rs )
      cipcnt = nproto

      if ( cipcnt.gt.maxpro ) then
            write ( 6, 300 ) maxpro
            write ( 6, 310 ) cipcnt
            goto 9999
300         format ( 'ERROR: T1CALC can only support ', i1, ' protons.' )
310         format ( '       The input file has ', i1, ' protons.' )
      endif
```

```
C               ZERO all entries of the daslvd array
        do i= 1, cipcnt
                daslvd(i) = 0.0
        end do

C               Read in each frame and calculate NOE's for each proton
C               to solve for

        twght = 0.0
        do iframe=1, nframe

C               Load a frame, and calculate all 1/r**6

                call frame( u, ddict, weight)
                twght = twght + weight
                call idists

C                       Solve T1's for each requested proton


                call gett1( dat1s, asolvp(islv) )

C                               Add weighted T1's to totals
                do i=1, cipcnt
                        daslvd(i) = daslvd(i) + weight*dat1s(i)
                end do
        end do

C               Divide all t1's by the total weight

        do i=1, cipcnt
                daslvd(i) = daslvd(i)/twght
        end do

C               Display results

        write ( 6, 90 ) stitle
90      format ( 'TITLE: ', A80 )

        write ( 6, 100 )
100     format ( 'T1"s' )

        do i= 1, cipcnt
                call dsname( ddict, sname, i )
                write ( 6, 110 ) sname, daslvd(i)
```

```
110             format ( 'Proton: ', a, ' T1=', F30.20 )
        end do

        write ( 6, 120 )
        write ( 6, 130 )
120     format ( '----------------------------------------' )
130     format ( ' ' )


9999    continue

        END
```

Subroutines

```
C
C
C       idists
C
C
C       initialize 1/r**6 matrix
C
C       ax, ay, az are the x,y,z coordinates of all the protons
C       np is the number of protons
C       The coordinates are in angstroms
C
        subroutine idists

        include         'general.i'


        integer                 i, j
        double precision        rij, rt, ame

        do i= 1, cipcnt-1
                do j= i+1, cipcnt
           if (mflag(i).eq.0) then
                if (mflag(j).eq.0) then
                                call dist(i,j,rij)
                                rt = rij**(-6)
                                call setr6( i, j, rt )
                        else
                                call rhme(j, i, ame)
                                rt = ame
                                call setr6( i, j, rt )
                        endif
                else
                        if (mflag(j).eq.0) then
                                call rhme(i,j,ame)
                                rt = ame
                                call setr6(i,j, rt)
                        else
                                call dist(i,j,rij)
                                rt = 0.25 * (rij**(-6))
                                call setr6(i,j, rt)
                        endif
```

```
            endif
            end do
        end do

        return
        end


C
        subroutine dist(k1, k2, dij)
C
C       calculates the distance between two atoms, returns the
C       value in angstroms
C

      include       'general.i'


      integer k1 , k2
      double precision dij, dx, dy,dz


      dx = (ax(k1) - ax(k2))*(1d-10)
      dy = (ay(k1) - ay(k2))*(1d-10)
      dz = (az(k1) - az(k2))*(1d-10)
      dij = dsqrt(dx*dx + dy*dy + dz*dz)
      return
      end




C
C       setr6
C
C       Set the value of 1/r**6 in the matrix
C       j must be larger than i
C
        subroutine setr6( i, j, rt )

        include         'general.i'

        integer                 i,j
        double precision        rt

        integer                 ipos
```

```
      ipos = (j-1)*(j-2)/2 + i
      rarray(ipos) = rt

      return
      end




C
C
C     r6
C
C     return the value of 1/r**6 for the two protons
C
C
      function r6( i, j )

      include               'general.i'

      double precision      r6
      integer               i, j

      integer               ipos, ti, tj

      if ( i.eq.j ) then
            r6 = 0
            return
      endif

      ti = min( i, j )
      tj = max( i, j )

      ipos = (tj-1)*(tj-2)/2 + ti

      r6 = rarray(ipos)

      return
      end


C

      subroutine rhme(i,j,ame)
C
```

```
C     HEATLEY et al. (1980) J.C.S PERKIN II, VOL 2, 919-924.
C     I = atom # of proton in methyl group MA,MB,MC
C     J = atom # of proton outside methyl group
C

      include             'general.i'


      integer i, j, k, l
      double precision    ame, rij(3), a, b
      double precision    x1(3), y1(3), z1(3), cosb


      ame = 0.0
      a = 0.0
      b = 0.0
      l = 1
      x1(l) = ax(i)
      y1(l) = ay(i)
      z1(l) = az(i)
      call dist(i,j,rij(l))
      l = l + 1
      do 10 k = 1, cipcnt
        if ((mflag(i).eq.mflag(k)).and.(i.ne.k)) then
              call dist(k,j,rij(l))
          x1(l) = ax(k)
          y1(l) = ay(k)
          z1(l) = az(k)
          l = l + 1
        end if
10     continue

      do 15 k = 1,3
      a = a + rij(k)**(-6.0)
15     continue

      do 20 k = 1,2
C      if (rij(k).gt.8.0) return
          do 30 l = k+1, 3
          cosb = 0.0
          cosb = (x1(k) - ax(j))*(x1(l) - ax(j))*(1d-20)
          cosb = cosb + (y1(k) - ay(j)) * (y1(l) - ay(j))*(1d-20)
          cosb = cosb + (z1(k) - az(j)) * (z1(l) - az(j))*(1d-20)
          cosb = cosb / (rij(k) * rij(l))
        b = b + 0.50*(3.0*cosb*cosb - 1.0)*(rij(k)*rij(l))**(-3.0)
```

```
30    continue
20    continue
      ame = (a+2.0*b)/9.0
        return
        end



C
C
C       This module maintains the DICT object
C
C       DICT is an array of up to DICTSZ strings
C       to which strings can be added and searched
C       for.
C
C


C
C       DINIT
C
C       Initialize the dictionary
C
        SUBROUTINE DINIT( DD )

        INCLUDE         'general.i'

        RECORD /DICT/ DD

        DD.INEXT = 1

        RETURN
        END



C
C       DADD
C
C       ADD A NAME TO THE DICTIONARY
C
        SUBROUTINE DADD( DD, SNAME )

        INCLUDE         'general.i'

        RECORD /DICT/           DD
        CHARACTER*10            SNAME
```

```
      DD.SNAMES(DD.INEXT) = SNAME
      DD.INEXT = DD.INEXT + 1

      RETURN
      END



C
C     DFIND
C
C     RETURN THE INDEX OF THE NAME, OR 0 IF THERE IS NO MATCH
C
      SUBROUTINE DFIND( DD, I, SNAME )

      INCLUDE       'general.i'

      RECORD        /DICT/  DD
      CHARACTER*10          SNAME
      INTEGER               I

      DO I=1, DD.INEXT-1
            IF ( SNAME .EQ. DD.SNAMES(I) ) GOTO 9000
      END DO
      I = 0
9000  RETURN
      END



C
C     DSNAME
C
C     RETURN THE I'TH NAME IN THE DICTIONARY
C
      SUBROUTINE DSNAME( DD, SNAME, I )

      INCLUDE       'general.i'

      RECORD        /DICT/  DD
      CHARACTER*10          SNAME
      INTEGER               I

      SNAME = DD.SNAMES(I)

      RETURN
      END



C  IMSL ROUTINE NAME   - LEQIF                    LEIF0010
C                                  LEIF0020
```

```
C-------------------------------------------------------------------LEIF0030
C                                                                   LEIF0040
C  COMPUTER        - IBM/DOUBLE                                     LEIF0050
C                                                                   LEIF0060
C  LATEST REVISION  - JUNE 1, 1982                                  LEIF0070
C                                                                   LEIF0080
C  PURPOSE         - LINEAR EQUATION SOLUTION - FULL MATRICES       LEIF0090
C                    (VIRTUAL MEMORY VERSION)                       LEIF0100
C                                                                   LEIF0110
C  USAGE           - CALL LEQIF(A,IA,N,MA,B,IB,M,IJOB,WK,IER)       LEIF0120
C                                                                   LEIF0130
C  ARGUMENTS  A    - INPUT N BY N MATRIX CONTAINING THE             LEIF0140
C                    COEFFICIENT MATRIX OF THE EQUATION AX=B.       LEIF0150
C                    ON OUTPUT, A IS REPLACED BY THE LU             LEIF0160
C                    DECOMPOSITION OF A ROWWISE PERMUTATION         LEIF0170
C                    OF A.  (INPUT/OUTPUT)                          LEIF0180
C             IA     - ROW DIMENSION OF A EXACTLY AS SPECIFIED      LEIF0190
C                    IN THE DIMENSION STATEMENT OF THE CALLING      LEIF0200
C                    PROGRAM. (INPUT)                               LEIF0210
C             N      - ORDER OF MATRIX A. (INPUT)                   LEIF0220
C             MA     - NUMBER OF COLUMNS PER BLOCK (INPUT). THE     LEIF0230
C                    CHOICE OF MA WILL AFFECT THE SOLUTION          LEIF0240
C                    SPEED AS FOLLOWS.  AS MA IS INCREASED,         LEIF0250
C                    THE ALGORITHM WILL RUN FASTER, UNTIL A         LEIF0260
C                    POINT IS REACHED BEYOND WHICH 2*MA*IA          LEIF0270
C                    WORKING PRECISION WORDS CANNOT BE HELD         LEIF0280
C                    IN MAIN MEMORY, WITHOUT PAGING.  MA            LEIF0290
C                    MUST BE LESS THAN OR EQUAL TO N.               LEIF0300
C             B      - INPUT N BY M MATRIX CONTAINING THE M         LEIF0310
C                    RIGHT HAND SIDES OF THE EQUATION AX = B.       LEIF0320
C                    ON OUTPUT, THE SOLUTION MATRIX X REPLACES B.   LEIF0330
C                    (INPUT/OUTPUT)                                 LEIF0340
C             IB     - ROW DIMENSION OF B EXACTLY AS SPECIFIED      LEIF0350
C                    IN THE DIMENSION STATEMENT OF THE CALLING      LEIF0360
C                    PROGRAM. (INPUT)                               LEIF0370
C             M      - NUMBER OF RIGHT HAND SIDES (COLUMNS IN B).   LEIF0380
C                    (INPUT)                                        LEIF0390
C             IJOB   - OPTION PARAMETER.  (INPUT) IJOB=I IMPLIES WHENLEIF0400
C                    I=0, FACTOR THE MATRIX AND SOLVE THE           LEIF0410
C                      EQUATION AX=B.                               LEIF0420
C                    I=1, SOLVE THE EQUATION AX=B.  THIS            LEIF0430
C                      OPTION IMPLIES THAT LEQIF HAS ALREADY        LEIF0440
C                      BEEN CALLED USING IJOB=0 SO THAT             LEIF0450
C                      THE MATRIX A HAS ALREADY BEEN                LEIF0460
C                      FACTORED, AND THAT WK HAS NOT                LEIF0470
C                      BEEN ALTERED SINCE THAT CALL.                LEIF0480
C             WK     - REAL WORK AREA OF LENGTH 3*N.                LEIF0490
C             IER    - ERROR PARAMETER. (OUTPUT)                    LEIF0500
C                    TERMINAL ERROR                                 LEIF0510
C                    IER=129 INDICATES THAT MATRIX A IS             LEIF0520
C                      ALGORITHMICALLY SINGULAR. (SEE THE           LEIF0530
C                      CHAPTER L PRELUDE.)                          LEIF0540
C                                                                   LEIF0550
```

```
C  PRECISION/HARDWARE  - SINGLE AND DOUBLE/H32                    LEIF0560
C                     - SINGLE/H36,H48,H60                       LEIF0570
C                                                                LEIF0580
C  REQD. IMSL ROUTINES - SINGLE/VBLA=SAXPY,UERTST,UGETIO         LEIF0590
C                        DOUBLE/VBLA=DAXPY,UERTST,UGETIO         LEIF0600
C                                                                LEIF0610
C  NOTATION            - INFORMATION ON SPECIAL NOTATION AND     LEIF0620
C                        CONVENTIONS IS AVAILABLE IN THE MANUAL  LEIF0630
C                        INTRODUCTION OR THROUGH IMSL ROUTINE UHELP LEIF0640
C                                                                LEIF0650
C  COPYRIGHT           - 1982 BY IMSL, INC. ALL RIGHTS RESERVED. LEIF0660
C                                                                LEIF0670
C  WARRANTY            - IMSL WARRANTS ONLY THAT IMSL TESTING HAS BEEN LEIF0680
C                        APPLIED TO THIS CODE. NO OTHER WARRANTY,  LEIF0690
C                        EXPRESSED OR IMPLIED, IS APPLICABLE.    LEIF0700
C                                                                LEIF0710
C-----------------------------------------------------------------LEIF0720
C                                                                LEIF0730
      SUBROUTINE LEQIF (A,IA,N,MA,B,IB,M,IJOB,WK,IER)            LEIF0740
C                          SPECIFICATIONS FOR ARGUMENTS          LEIF0750
      INTEGER            IA,IB,N,MA,M,IJOB,IER                   LEIF0760
      DOUBLE PRECISION   A(IA,N),B(IB,M),WK(N,3)                 LEIF0770
C                          SPECIFICATIONS FOR LOCAL VARIABLES    LEIF0780
      INTEGER            I,IBIG,IIK,IP1,IZ,J,JJ,JJJ,JZ,KB,KBP1,  LEIF0790
     *                   KKB,LB,LIMK,LIMK0,LIMK1,LIMK2,LIML,LIML0, LEIF0800
     *                   LIML1,LIML2,NBLOCK                      LEIF0810
      DOUBLE PRECISION   ABIG,AM,REPS,TEMP                       LEIF0820
      DATA               REPS/1D-15/                             LEIF0830
C                          FIRST EXECUTABLE STATEMENT            LEIF0840
      IER = 0                                                    LEIF0850
      DO 5 I=1,N                                                 LEIF0860
    5 WK(I,1) = I                                                LEIF0870
      NBLOCK = (N-1)/MA+1                                        LEIF0880
C                          BEGIN GAUSSIAN ELIMINATION            LEIF0890
      DO 80 KB=1,NBLOCK                                          LEIF0900
        LIMK0 = (KB-1)*MA                                        LEIF0910
        LIMK1 = LIMK0+1                                          LEIF0920
        LIMK2 = MIN0(LIMK0+MA,N)                                 LEIF0930
        IF (KB.GT.1 .OR. IJOB.EQ.1) GO TO 20                     LEIF0940
        DO 15 J=LIMK1,LIMK2                                      LEIF0950
          TEMP = 0.0D0                                           LEIF0960
          DO 10 I=1,N                                            LEIF0970
            TEMP = DMAX1(TEMP,DABS(A(I,J)))                      LEIF0980
   10     CONTINUE                                               LEIF0990
          WK(J,2) = TEMP                                         LEIF1000
   15   CONTINUE                                                 LEIF1010
   20   CONTINUE                                                 LEIF1020
C                          FACTOR FIRST BUFFER                   LEIF1030
        DO 40 I=LIMK1,LIMK2                                      LEIF1040
          ABIG = 0.0D0                                           LEIF1050
          DO 25 J=I,N                                            LEIF1060
            JZ = WK(J,1)                                         LEIF1070
            IF (DABS(A(JZ,I)).LE.ABIG) GO TO 25                  LEIF1080
```

```
        ABIG = DABS(A(JZ,I))                                    LEIF1090
        IBIG = J                                                LEIF1100
 25     CONTINUE                                                LEIF1110
      IF (ABIG.EQ.0.0D0) GO TO 9000                             LEIF1120
      IF (IJOB.EQ.1) GO TO 30                                   LEIF1130
      IF (ABIG.LE.10.0D0*REPS*WK(I,2)) GO TO 9000               LEIF1140
 30     CONTINUE                                                LEIF1150
      TEMP = WK(IBIG,1)                                         LEIF1160
      WK(IBIG,1) = WK(I,1)                                      LEIF1170
      WK(I,1) = TEMP                                            LEIF1180
      IF (I.GE.N) GO TO 40                                      LEIF1190
      IZ = WK(I,1)                                              LEIF1200
      IP1 = I+1                                                 LEIF1210
      DO 35 J=IP1,N                                             LEIF1220
        JZ = WK(J,1)                                            LEIF1230
        AM = A(JZ,I)/A(IZ,I)                                    LEIF1240
        IF (AM.EQ.0.0D0) GO TO 35                               LEIF1250
        CALL DAXPY(M,-AM,B(IZ,1),IA,B(JZ,1),IA)                 LEIF1260
        IF (IJOB.EQ.1) GO TO 35                                 LEIF1270
        IF (I.GE.LIMK2) GO TO 35                                LEIF1280
        CALL DAXPY(LIMK2-I,-AM,A(IZ,I+1),IA,A(JZ,I+1),IA)       LEIF1290
 35     CONTINUE                                                LEIF1300
 40   CONTINUE                                                  LEIF1310
      IF (IJOB.EQ.1) GO TO 80                                   LEIF1320
      IF (KB.GE.NBLOCK) GO TO 75                                LEIF1330
      KBP1 = KB+1                                               LEIF1340
      DO 70 LB=KBP1,NBLOCK                                      LEIF1350
        LIML0 = (LB-1)*MA                                       LEIF1360
        LIML1 = LIML0+1                                         LEIF1370
        LIML2 = MIN0(LIML0+MA,N)                                LEIF1380
        LIML = LIML2-LIML0                                      LEIF1390
        IF (KB.GT.1) GO TO 55                                   LEIF1400
        DO 50 J=LIML1,LIML2                                     LEIF1410
          TEMP = 0.0D0                                          LEIF1420
          DO 45 I=1,N                                           LEIF1430
            TEMP = DMAX1(TEMP,DABS(A(I,J)))                     LEIF1440
 45       CONTINUE                                              LEIF1450
          WK(J,2) = TEMP                                        LEIF1460
 50     CONTINUE                                                LEIF1470
 55     CONTINUE                                                LEIF1480
C                    DO ELIMINATION ON SECOND BLOCK             LEIF1490
C                     USING FACTORS SAVED IN FIRST              LEIF1500
      DO 65 I=LIMK1,LIMK2                                       LEIF1510
        IF (I.GE.N) GO TO 65                                    LEIF1520
        IZ = WK(I,1)                                            LEIF1530
        IP1 = I+1                                               LEIF1540
        DO 60 J=IP1,N                                           LEIF1550
          JZ = WK(J,1)                                          LEIF1560
          AM = A(JZ,I)/A(IZ,I)                                  LEIF1570
          IF (AM.EQ.0.0D0) GO TO 60                             LEIF1580
          CALL DAXPY(LIML,-AM,A(IZ,LIML1),IA,A(JZ,LIML1),IA)   LEIF1590
 60       CONTINUE                                              LEIF1600
 65     CONTINUE                                                LEIF1610
```

```
   70  CONTINUE                                                    LEIF1620
   75  CONTINUE                                                    LEIF1630
   80 CONTINUE                                                     LEIF1640
C                       BACK SUBSTITUTION                          LEIF1650
     DO 105 KKB=1,NBLOCK                                           LEIF1660
       KB = NBLOCK+1-KKB                                           LEIF1670
       LIMK0 = (KB-1)*MA                                           LEIF1680
       LIMK2 = MIN0(LIMK0+MA,N)                                    LEIF1690
       LIMK = LIMK2-LIMK0                                          LEIF1700
       DO 100 IIK=1,LIMK                                           LEIF1710
         I = LIMK2+1-IIK                                           LEIF1720
         IZ = WK(I,1)                                              LEIF1730
         TEMP = A(IZ,I)                                            LEIF1740
         DO 85 JJJ=1,M                                             LEIF1750
           B(IZ,JJJ) = B(IZ,JJJ)/TEMP                              LEIF1760
  85     CONTINUE                                                  LEIF1770
         IF (I.EQ.1) GO TO 100                                     LEIF1780
         DO 95 JJ=2,I                                              LEIF1790
           J = I+1-JJ                                              LEIF1800
           JZ = WK(J,1)                                            LEIF1810
           TEMP = A(JZ,I)                                          LEIF1820
           DO 90 JJJ=1,M                                           LEIF1830
             B(JZ,JJJ) = B(JZ,JJJ)-TEMP*B(IZ,JJJ)                  LEIF1840
  90       CONTINUE                                                LEIF1850
  95     CONTINUE                                                  LEIF1860
 100   CONTINUE                                                    LEIF1870
 105 CONTINUE                                                      LEIF1880
C                       SORT SOLUTION VECTOR                       LEIF1890
     DO 120 JJJ=1,M                                                LEIF1900
       DO 110 I=1,N                                                LEIF1910
         IZ = WK(I,1)                                              LEIF1920
         WK(I,3) = B(IZ,JJJ)                                       LEIF1930
 110   CONTINUE                                                    LEIF1940
       DO 115 I=1,N                                                LEIF1950
         B(I,JJJ) = WK(I,3)                                        LEIF1960
 115   CONTINUE                                                    LEIF1970
 120 CONTINUE                                                      LEIF1980
     GO TO 9005                                                    LEIF1990
9000 IER = 129                                                     LEIF2000
     CALL UERTST(IER,6HLEQIF )                                     LEIF2010
9005 RETURN                                                        LEIF2020
     END                                                           LEIF2030
C  IMSL ROUTINE NAME   - UERTST                                    UERT0010
C                                                                  UERT0020
C-----------------------------------------------------------------UERT0030
C                                                                  UERT0040
C  COMPUTER            - IBM/SINGLE                                UERT0050
C                                                                  UERT0060
C  LATEST REVISION     - JUNE 1, 1982                              UERT0070
C                                                                  UERT0080
C  PURPOSE             - PRINT A MESSAGE REFLECTING AN ERROR CONDITION UERT0090
C                                                                  UERT0100
C  USAGE               - CALL UERTST (IER,NAME)                    UERT0110
```

```
C                                                       UERT0120
C  ARGUMENTS   IER   - ERROR PARAMETER. (INPUT)                 UERT0130
C                  IER = I+J WHERE                      UERT0140
C                   I = 128 IMPLIES TERMINAL ERROR MESSAGE,  UERT0150
C                   I =  64 IMPLIES WARNING WITH FIX MESSAGE, UERT0160
C                   I =  32 IMPLIES WARNING MESSAGE.         UERT0170
C                   J = ERROR CODE RELEVANT TO CALLING       UERT0180
C                     ROUTINE.                           UERT0190
C          NAME   - A CHARACTER STRING OF LENGTH SIX PROVIDING   UERT0200
C                THE NAME OF THE CALLING ROUTINE. (INPUT)   UERT0210
C                                                       UERT0220
C  PRECISION/HARDWARE  - SINGLE/ALL                          UERT0230
C                                                       UERT0240
C  REQD. IMSL ROUTINES - UGETIO,USPKD                        UERT0250
C                                                       UERT0260
C  NOTATION         - INFORMATION ON SPECIAL NOTATION AND         UERT0270
C                  CONVENTIONS IS AVAILABLE IN THE MANUAL     UERT0280
C                  INTRODUCTION OR THROUGH IMSL ROUTINE UHELP  UERT0290
C                                                       UERT0300
C  REMARKS     THE ERROR MESSAGE PRODUCED BY UERTST IS WRITTEN       UERT0310
C           TO THE STANDARD OUTPUT UNIT. THE OUTPUT UNIT         UERT0320
C           NUMBER CAN BE DETERMINED BY CALLING UGETIO AS        UERT0330
C           FOLLOWS..  CALL UGETIO(1,NIN,NOUT).              UERT0340
C           THE OUTPUT UNIT NUMBER CAN BE CHANGED BY CALLING     UERT0350
C           UGETIO AS FOLLOWS..                          UERT0360
C                    NIN = 0                         UERT0370
C                   NOUT = NEW OUTPUT UNIT NUMBER        UERT0380
C                   CALL UGETIO(3,NIN,NOUT)           UERT0390
C           SEE THE UGETIO DOCUMENT FOR MORE DETAILS.          UERT0400
C                                                       UERT0410
C  COPYRIGHT        - 1982 BY IMSL, INC. ALL RIGHTS RESERVED.     UERT0420
C                                                       UERT0430
C  WARRANTY         - IMSL WARRANTS ONLY THAT IMSL TESTING HAS BEEN UERT0440
C                APPLIED TO THIS CODE. NO OTHER WARRANTY,   UERT0450
C                EXPRESSED OR IMPLIED, IS APPLICABLE.       UERT0460
C                                                       UERT0470
C-----------------------------------------------------------------------UERT0480
C                                                       UERT0490
      SUBROUTINE UERTST (IER,NAME)                            UERT0500
C                     SPECIFICATIONS FOR ARGUMENTS       UERT0510
      INTEGER         IER                               UERT0520
      INTEGER         NAME(1)                             UERT0530
C                     SPECIFICATIONS FOR LOCAL VARIABLES  UERT0540
      INTEGER         I,IEQ,IEQDF,IOUNIT,LEVEL,LEVOLD,NAMEQ(6),    UERT0550
     *             NAMSET(6),NAMUPK(6),NIN,NMTB              UERT0560
      DATA           NAMSET/1HU,1HE,1HR,1HS,1HE,1HT/          UERT0570
      DATA           NAMEQ/6*1H /                          UERT0580
      DATA           LEVEL/4/,IEQDF/0/,IEQ/1H=/              UERT0590
C                     UNPACK NAME INTO NAMUPK           UERT0600
C                     FIRST EXECUTABLE STATEMENT        UERT0       CALL USPKD
(NAME,6,NAMUPK,NMTB)                              UERT0620
C                     GET OUTPUT UNIT NUMBER            UERT0630
      CALL UGETIO(1,NIN,IOUNIT)                           UERT0640
```

```
C                     CHECK IER                    UERT0650
      IF (IER.GT.999) GO TO 25                     UERT0660
      IF (IER.LT.-32) GO TO 55                     UERT0670
      IF (IER.LE.128) GO TO 5                      UERT0680
      IF (LEVEL.LT.1) GO TO 30                     UERT0690
C                   PRINT TERMINAL MESSAGE         UERT0700
      IF (IEQDF.EQ.1) WRITE(IOUNIT,35) IER,NAMEQ,IEQ,NAMUPK        UERT0710
      IF (IEQDF.EQ.0) WRITE(IOUNIT,35) IER,NAMUPK                  UERT0720
      GO TO 30                                     UERT0730
    5 IF (IER.LE.64) GO TO 10                      UERT0740
      IF (LEVEL.LT.2) GO TO 30                     UERT0750
C                   PRINT WARNING WITH FIX MESSAGE UERT0760
      IF (IEQDF.EQ.1) WRITE(IOUNIT,40) IER,NAMEQ,IEQ,NAMUPK        UERT0770
      IF (IEQDF.EQ.0) WRITE(IOUNIT,40) IER,NAMUPK                  UERT0780
      GO TO 30                                     UERT0790
   10 IF (IER.LE.32) GO TO 15                      UERT0800
C                   PRINT WARNING MESSAGE          UERT0810
      IF (LEVEL.LT.3) GO TO 30                     UERT0820
      IF (IEQDF.EQ.1) WRITE(IOUNIT,45) IER,NAMEQ,IEQ,NAMUPK        UERT0830
      IF (IEQDF.EQ.0) WRITE(IOUNIT,45) IER,NAMUPK                  UERT0840
      GO TO 30                                     UERT0850
   15 CONTINUE                                     UERT0860
C                   CHECK FOR UERSET CALL          UERT0870
      DO 20 I=1,6                                  UERT0880
        IF (NAMUPK(I).NE.NAMSET(I)) GO TO 25                UERT0890
   20 CONTINUE                                     UERT0900
      LEVOLD = LEVEL                               UERT0910
      LEVEL = IER                                  UERT0920
      IER = LEVOLD                                 UERT0930
      IF (LEVEL.LT.0) LEVEL = 4                    UERT0940
      IF (LEVEL.GT.4) LEVEL = 4                    UERT0950
      GO TO 30                                     UERT0960
   25 CONTINUE                                     UERT0970
      IF (LEVEL.LT.4) GO TO 30                     UERT0980
C                   PRINT NON-DEFINED MESSAGE      UERT0990
      IF (IEQDF.EQ.1) WRITE(IOUNIT,50) IER,NAMEQ,IEQ,NAMUPK        UERT1000
      IF (IEQDF.EQ.0) WRITE(IOUNIT,50) IER,NAMUPK                  UERT1010
   30 IEQDF = 0                                    UERT1020
      RETURN                                       UERT1030
   35 FORMAT(19H *** TERMINAL ERROR,10X,7H(IER = ,I3,          UERT1040
     1    20H) FROM IMSL ROUTINE ,6A1,A1,6A1)                  UERT1050
   40 FORMAT(27H *** WARNING WITH FIX ERROR,2X,7H(IER = ,I3,    UERT1060
     1    20H) FROM IMSL ROUTINE ,6A1,A1,6A1)                  UERT1070
   45 FORMAT(18H *** WARNING ERROR,11X,7H(IER = ,I3,           UERT1080
     1    20H) FROM IMSL ROUTINE ,6A1,A1,6A1)                  UERT1090
   50 FORMAT(20H *** UNDEFINED ERROR,9X,7H(IER = ,I5,          UERT1100
     1    20H) FROM IMSL ROUTINE ,6A1,A1,6A1)                  UERT1110
C                                                  UERT1120
C                   SAVE P FOR P = R CASE          UERT1130
C                    P IS THE PAGE NAMUPK          UERT1140
C                    R IS THE ROUTINE NAMUPK       UERT1150
   55 IEQDF = 1                                    UERT1160
      DO 60 I=1,6                                  UERT1170
```

```
   60 NAMEQ(I) = NAMUPK(I)                                    UERT1180
   65 RETURN                                                  UERT1190
      END                                                     UERT1200
C  IMSL ROUTINE NAME   - UGETIO                                UGET0010
C                                                            UGET0020
C-----------------------------------------------------------UGET0030
C                                                            UGET0040
C  COMPUTER           - IBM/SINGLE                           UGET0050
C                                                            UGET0060
C  LATEST REVISION    - JUNE 1, 1981                         UGET0070
C                                                            UGET0080
C  PURPOSE            - TO RETRIEVE CURRENT VALUES AND TO SET NEW    UGET0090
C                       VALUES FOR INPUT AND OUTPUT UNIT     UGET0100
C                       IDENTIFIERS.                         UGET0110
C                                                            UGET0120
C  USAGE              - CALL UGETIO(IOPT,NIN,NOUT)           UGET0130
C                                                            UGET0140
C  ARGUMENTS    IOPT  - OPTION PARAMETER. (INPUT)            UGET0150
C                       IF IOPT=1, THE CURRENT INPUT AND OUTPUT    UGET0160
C                       UNIT IDENTIFIER VALUES ARE RETURNED IN NIN  UGET0170
C                       AND NOUT, RESPECTIVELY.              UGET0180
C                       IF IOPT=2, THE INTERNAL VALUE OF NIN IS    UGET0190
C                       RESET FOR SUBSEQUENT USE.            UGET0200
C                       IF IOPT=3, THE INTERNAL VALUE OF NOUT IS   UGET0210
C                       RESET FOR SUBSEQUENT USE.            UGET0220
C               NIN   - INPUT UNIT IDENTIFIER.               UGET0230
C                       OUTPUT IF IOPT=1, INPUT IF IOPT=2.   UGET0240
C               NOUT  - OUTPUT UNIT IDENTIFIER.              UGET0250
C                       OUTPUT IF IOPT=1, INPUT IF IOPT=3.   UGET0260
C                                                            UGET0270
C  PRECISION/HARDWARE - SINGLE/ALL                           UGET0280
C                                                            UGET0290
C  REQD. IMSL ROUTINES - NONE REQUIRED                       UGET0300
C                                                            UGET0310
C  NOTATION           - INFORMATION ON SPECIAL NOTATION AND  UGET0320
C                       CONVENTIONS IS AVAILABLE IN THE MANUAL    UGET0330
C                       INTRODUCTION OR THROUGH IMSL ROUTINE UHELP  UGET0340
C                                                            UGET0350
C  REMARKS      EACH IMSL ROUTINE THAT PERFORMS INPUT AND/OR OUTPUT
UGET0360
C               OPERATIONS CALLS UGETIO TO OBTAIN THE CURRENT UNIT    UGET0370
C               IDENTIFIER VALUES. IF UGETIO IS CALLED WITH IOPT=2 OR  UGET0380
C               IOPT=3, NEW UNIT IDENTIFIER VALUES ARE ESTABLISHED.   UGET0390
C               SUBSEQUENT INPUT/OUTPUT IS PERFORMED ON THE NEW UNITS. UGET0400
C                                                            UGET0410
C  COPYRIGHT          - 1978 BY IMSL, INC. ALL RIGHTS RESERVED.      UGET0420
C                                                            UGET0430
C  WARRANTY           - IMSL WARRANTS ONLY THAT IMSL TESTING HAS BEEN UGET0440
C                       APPLIED TO THIS CODE. NO OTHER WARRANTY,   UGET0450
C                       EXPRESSED OR IMPLIED, IS APPLICABLE.       UGET0460
C                                                            UGET0470
C-----------------------------------------------------------UGET0480
C                                                            UGET0490
```

```
      SUBROUTINE UGETIO(IOPT,NIN,NOUT)                          UGET0500
C                                  SPECIFICATIONS FOR ARGUMENTS  UGET0510
      INTEGER                    IOPT,NIN,NOUT                  UGET0520
C                                  SPECIFICATIONS FOR LOCAL VARIABLES  UGET0530
      INTEGER                    NIND,NOUTD                     UGET0540
      DATA                       NIND/5/,NOUTD/6/               UGET0550
C                                  FIRST EXECUTABLE STATEMENT    UGET0560
      IF (IOPT.EQ.3) GO TO 10                                   UGET0570
      IF (IOPT.EQ.2) GO TO 5                                    UGET0580
      IF (IOPT.NE.1) GO TO 9005                                 UGET0590
      NIN = NIND                                                UGET0600
      NOUT = NOUTD                                              UGET0610
      GO TO 9005                                                UGET0620
    5 NIND = NIN                                                UGET0630
      GO TO 9005                                                UGET0640
   10 NOUTD = NOUT                                              UGET0650
 9005 RETURN                                                    UGET0660
      END                                                       UGET0670
C  IMSL ROUTINE NAME   - VBLA=DAXPY                             VBDB0010
C                                                               VBDB0020
C-----------------------------------------------------------------VBDB0030
C                                                               VBDB0040
C  COMPUTER            - IBM/DOUBLE                             VBDB0050
C                                                               VBDB0060
C  LATEST REVISION     - JANUARY 1, 1978                        VBDB0070
C                                                               VBDB0080
C  PURPOSE             - COMPUTE A CONSTANT TIMES A VECTOR PLUS  VBDB0090
C                        A VECTOR, ALL DOUBLE PRECISION         VBDB0100
C                                                               VBDB0110
C  USAGE               - CALL DAXPY (N,DA,DX,INCX,DY,INCY)      VBDB0120
C                                                               VBDB0130
C  ARGUMENTS    N      - LENGTH OF VECTORS X AND Y. (INPUT)     VBDB0140
C               DA     - DOUBLE PRECISION SCALAR. (INPUT)       VBDB0150
C               DX     - DOUBLE PRECISION VECTOR OF LENGTH      VBDB0160
C                        MAX(N*IABS(INCX),1). (INPUT)           VBDB0170
C               INCX   - DISPLACEMENT BETWEEN ELEMENTS OF DX. (INPUT)  VBDB0180
C                        X(I) IS DEFINED TO BE..                VBDB0190
C                        DX(1+(I-1)*INCX) IF INCX.GE.0 OR       VBDB0200
C                        DX(1+(I-N)*INCX) IF INCX.LT.0.         VBDB0210
C               DY     - DOUBLE PRECISION VECTOR OF LENGTH      VBDB0220
C                        MAX(N*IABS(INCY),1). (INPUT/OUTPUT)    VBDB0230
C                        DAXPY REPLACES Y(I) WITH DA*X(I)+Y(I) FOR  VBDB0240
C                        I=1,...,N.                             VBDB0250
C                        X(I) AND Y(I) REFER TO SPECIFIC ELEMENTS   VBDB0260
C                        OF DX AND DY, RESPECTIVELY. SEE INCX AND   VBDB0270
C                        INCY ARGUMENT DESCRIPTIONS.            VBDB0280
C               INCY   - DISPLACEMENT BETWEEN ELEMENTS OF DY. (INPUT)  VBDB0290
C                        Y(I) IS DEFINED TO BE..                VBDB0300
C                        DY(1+(I-1)*INCY) IF INCY.GE.0 OR       VBDB0310
C                        DY(1+(I-N)*INCY) IF INCY.LT.0.         VBDB0320
C                                                               VBDB0330
C  PRECISION/HARDWARE  - DOUBLE/ALL                             VBDB0340
C                                                               VBDB0350
```

```
C  REQD. IMSL ROUTINES - NONE REQUIRED                    VBDB0360
C                                                         VBDB0370
C  NOTATION       - INFORMATION ON SPECIAL NOTATION AND     VBDB0380
C                  CONVENTIONS IS AVAILABLE IN THE MANUAL   VBDB0390
C                  INTRODUCTION OR THROUGH IMSL ROUTINE UHELP VBDB0400
C                                                         VBDB0410
C  COPYRIGHT      - 1978 BY IMSL, INC. ALL RIGHTS RESERVED.  VBDB0420
C                                                         VBDB0430
C  WARRANTY       - IMSL WARRANTS ONLY THAT IMSL TESTING HAS BEEN VBDB0440
C                  APPLIED TO THIS CODE. NO OTHER WARRANTY,  VBDB0450
C                  EXPRESSED OR IMPLIED, IS APPLICABLE.     VBDB0460
C                                                         VBDB0470
C-----------------------------------------------------------------VBDB0480
C                                                         VBDB0490
      SUBROUTINE DAXPY  (N,DA,DX,INCX,DY,INCY)              VBDB0500
C                                                         VBDB0510
C                     SPECIFICATIONS FOR ARGUMENTS        VBDB0520
      DOUBLE PRECISION  DX(1),DY(1),DA                    VBDB0530
      INTEGER           N,INCX,INCY                       VBDB0540
C                     SPECIFICATIONS FOR LOCAL VARIABLES  VBDB0550
      INTEGER           I,IY,M,MP1,NS,IX                  VBDB0560
C                     FIRST EXECUTABLE STATEMENT          VBDB0570
      IF (N.LE.0.OR.DA.EQ.0.D0) RETURN                    VBDB0580
      IF (INCX.EQ.INCY) IF (INCX-1) 5,15,35               VBDB0590
    5 CONTINUE                                            VBDB0600
C                     CODE FOR NONEQUAL OR NONPOSITIVE    VBDB0610
C                       INCREMENTS.                       VBDB0620
      IX = 1                                              VBDB0630
      IY = 1                                              VBDB0640
      IF (INCX.LT.0) IX = (-N+1)*INCX+1                   VBDB0650
      IF (INCY.LT.0) IY = (-N+1)*INCY+1                   VBDB0660
      DO 10 I=1,N                                         VBDB0670
        DY(IY) = DY(IY)+DA*DX(IX)                         VBDB0680
        IX = IX+INCX                                      VBDB0690
        IY = IY+INCY                                      VBDB0700
   10 CONTINUE                                            VBDB0710
      RETURN                                              VBDB0720
C                     CODE FOR BOTH INCREMENTS EQUAL TO 1 VBDB0730
C                       CLEAN-UP LOOP SO REMAINING VECTOR VBDB0740
C                       LENGTH IS A MULTIPLE OF 4.        VBDB0750
   15 M = N-(N/4)*4                                       VBDB0760
      IF (M.EQ.0) GO TO 25                                VBDB0770
      DO 20 I=1,M                                         VBDB0780
        DY(I) = DY(I)+DA*DX(I)                            VBDB0790
   20 CONTINUE                                            VBDB0800
      IF (N.LT.4) RETURN                                  VBDB0810
   25 MP1 = M+1                                           VBDB0820
      DO 30 I=MP1,N,4                                     VBDB0830
        DY(I) = DY(I)+DA*DX(I)                            VBDB0840
        DY(I+1) = DY(I+1)+DA*DX(I+1)                      VBDB0850
        DY(I+2) = DY(I+2)+DA*DX(I+2)                      VBDB0860
        DY(I+3) = DY(I+3)+DA*DX(I+3)                      VBDB0870
   30 CONTINUE                                            VBDB0880
```

```
   RETURN                                                VBDB0890
C                    CODE FOR EQUAL, POSITIVE, NONUNIT   VBDB0900
C                     INCREMENTS.                        VBDB0910
  35 CONTINUE                                           VBDB0920
     NS = N*INCX                                        VBDB0930
     DO 40 I=1,NS,INCX                                  VBDB0940
       DY(I) = DA*DX(I)+DY(I)                           VBDB0950
  40 CONTINUE                                           VBDB0960
     RETURN                                             VBDB0970
     END                                                VBDB0980
```

```
C
C
C
C       header
C
C       Read the header, return tc, w0, #protons, a dictionary with
C       all the names, #protons to solve for, an array of the
C       ids of the protons to solve for and the number of frames in
C       the file.
C
C       on entry (u) is the unit of the file
C
        subroutine header( u, stitle, tc, w0, rs, np, dn, nsp, asp, nf )

        include             'general.i'

        character*80        stitle
        integer             u
        double precision    tc, w0, rs
        integer             np, nsp
        record /dict/       dn
        integer             asp(maxpro), nf

        integer             ip, i
        character*10        sname

90      format ( 'TITLE=', A80 )
100     format ( 'TC=', E20.3 )
110     format ( 'W0=', E20.3 )
115     format ( 'RS=', E20.3 )
120     format ( 'NUMBER OF PROTONS=', I5 )
130     format ( A10 )
135     format (a10, i3)
140     format ( 'NUMBER OF PROTONS TO SOLVE FOR=', I5 )
```

```
150   format ( A10 )
160   format ( 'NUMBER OF FRAMES=', I5 )

      call dinit( dn )

C           Read the tc,w0 constants

      read ( u, 90 ) stitle
      read ( u, 100 ) tc
      read ( u, 110 ) w0
      read ( u, 115 ) rs

C           Read all the proton names

      read ( u, 120 ) np
      do i=1, np
            read ( u, 135) sname, mflag(i)
D           write(6,*)sname, mflag(i)
            call dadd( dn, sname )
      end do

C           Read the names of the protons to solve for

      read ( u, 140 ) nsp
      do i=1, nsp
            read ( u, 150 ) sname
            call dfind( dn, ip, sname )
            asp(i) = ip
      end do

C           Read the number of frames

      read ( u, 160 ) nf

      return
      end



C
C
C     frame
C
C
C     read a dynamics frame from the file, setting the proton
C     coordinates in the ax,ay,az arrays at the same index as
```

```
C       the protons' name dictionary index.
C
C       on entry (u) is the unit of the file
C
        subroutine frame ( u, dd, weight)

        include         'general.i'

        record /dict/           dd
        integer                 u
        double precision        weight

        integer                 i, fp, ip
        character*10            sname
        double precision        x, y, z

C               read the number of protons in the frame and the weighting
C               of the frame

100     format ( 'FRAME PROTONS=', I5 )
110     format ( 'WEIGHT=', E20.3 )
120     format ( A10, E20.3, E20.3, E20.3 )

        read ( u, 100 ) fp
        read ( u, 110 ) weight

C               read the protons coordinates

        do i= 1, fp
                read ( u, 120 ) sname, x, y, z
                call dfind( dd, ip, sname )
                ax(ip) = x
                ay(ip) = y
                az(ip) = z
        end do

        return
        end

C  IMSL ROUTINE NAME   - USPKD                                  USPK0010
C                                                               USPK0020
C-------------------------------------------------------------------USPK0030
C                                                               USPK0040
C  COMPUTER            - IBM/SINGLE                             USPK0050
C                                                               USPK0060
C  LATEST REVISION     - JUNE 1, 1982                           USPK0070
```

```
C                                                       USPK0080
C  PURPOSE          - NUCLEUS CALLED BY IMSL ROUTINES THAT HAVE
USPK0090
C                  CHARACTER STRING ARGUMENTS              USPK0100
C                                                       USPK0110
C  USAGE            - CALL USPKD  (PACKED,NCHARS,UNPAKD,NCHMTB)
USPK0120
C                                                       USPK0130
C  ARGUMENTS    PACKED - CHARACTER STRING TO BE
UNPACKED.(INPUT)      USPK0140
C            NCHARS - LENGTH OF PACKED. (INPUT)  SEE REMARKS.
USPK0150
C            UNPAKD - INTEGER ARRAY TO RECEIVE THE UNPACKED
USPK0160
C                 REPRESENTATION OF THE STRING. (OUTPUT)       USPK0170
C            NCHMTB - NCHARS MINUS TRAILING BLANKS. (OUTPUT)
USPK0180
C                                                       USPK0190
C  PRECISION/HARDWARE  - SINGLE/ALL                          USPK0200
C                                                       USPK0210
C  REQD. IMSL ROUTINES - NONE                                USPK0220
C                                                       USPK0230
C  REMARKS  1.  USPKD UNPACKS A CHARACTER STRING INTO AN
INTEGER ARRAY USPK0240
C            IN (A1) FORMAT.                              USPK0250
C         2.  UP TO 129 CHARACTERS MAY BE USED.  ANY IN EXCESS OF
USPK0260
C            THAT ARE IGNORED.                            USPK0270
C                                                       USPK0280
C  COPYRIGHT        - 1982 BY IMSL, INC. ALL RIGHTS RESERVED.
USPK0290
C                                                       USPK0300
C  WARRANTY          - IMSL WARRANTS ONLY THAT IMSL TESTING HAS
BEEN USPK0310
C                  APPLIED TO THIS CODE. NO OTHER WARRANTY,
USPK0320
C                  EXPRESSED OR IMPLIED, IS APPLICABLE.       USPK0330
C                                                       USPK0340
C-----------------------------------------------------------------USPK0350
     SUBROUTINE USPKD  (PACKED,NCHARS,UNPAKD,NCHMTB)
USPK0360
C                     SPECIFICATIONS FOR ARGUMENTS       USPK0370
     INTEGER          NC,NCHARS,NCHMTB                        USPK0380
C                                                       USPK0390
     LOGICAL*1        UNPAKD(1),PACKED(1),LBYTE,LBLANK
USPK0400
```

```
      INTEGER*2       IBYTE,IBLANK                                    USPK0410
      EQUIVALENCE (LBYTE,IBYTE)                                       USPK0420
      DATA            LBLANK /1H /                                    USPK0430
      DATA            IBYTE /1H /                                     USPK0440
      DATA            IBLANK /1H /                                    USPK0450
C                       INITIALIZE NCHMTB                             USPK0460
      NCHMTB = 0                                                      USPK0470
C                       RETURN IF NCHARS IS LE ZERO                   USPK0480
      IF(NCHARS.LE.0) RETURN                                          USPK0490
C                       SET NC=NUMBER OF CHARS TO BE DECODED          USPK0500
      NC = MIN0 (129,NCHARS)                                          USPK0510
      NWORDS = NC*4                                                   USPK0520
      J = 1                                                           USPK0530
      DO 110 I = 1,NWORDS,4                                           USPK0540
      UNPAKD(I) = PACKED(J)                                           USPK0550
      UNPAKD(I+1) = LBLANK                                            USPK0560
      UNPAKD(I+2) = LBLANK                                            USPK0570
      UNPAKD(I+3) = LBLANK                                            USPK0580
  110 J = J+1                                                         USPK0590
C                       CHECK UNPAKD ARRAY AND SET NCHMTB             USPK0600
C                       BASED ON TRAILING BLANKS FOUND                USPK0610
      DO 200 N = 1,NWORDS,4                                           USPK0620
        NN = NWORDS - N - 2                                           USPK0630
        LBYTE = UNPAKD(NN)                                            USPK0640
        IF(IBYTE .NE. IBLANK) GO TO 210                               USPK0650
  200 CONTINUE                                                        USPK0660
  210 NCHMTB = (NN + 3) / 4                                           USPK0670
      RETURN                                                          USPK0680
      END                                                             USPK0690
```