

ACS SYMPOSIUM SERIES 930

NMR Spectroscopy and Computer Modeling of Carbohydrates

Recent Advances



EDITED BY
Johannes F. G. Vliegthart
and **Robert J. Woods**

NMR Spectroscopy and Computer Modeling of Carbohydrates

ACS SYMPOSIUM SERIES **930**

NMR Spectroscopy and Computer Modeling of Carbohydrates

Recent Advances

Johannes F. G. Vliegthart, Editor
Utrecht University

Robert J. Woods, Editor
University of Georgia

**Sponsored by the
ACS Division of Carbohydrate Chemistry**



American Chemical Society, Washington, DC



Library of Congress Cataloging-in-Publication Data

NMR spectroscopy and computer modeling of carbohydrates / Johannes F. G. Vliegthart, editor ; Robert J. Woods, editor.

p. cm.—(ACS symposium series ; 930)

“Developed from a symposium sponsored by the Division of Carbohydrate Chemistry at the 226th National Meeting of the American Chemical Society, New York, September 7–11, 2003”—T.p. verso.

Includes bibliographical references and indexes.

ISBN 13: 978-0-8412-3953-1 (alk. paper)

1. Carbohydrate—Computer simulation. 2. Nuclear magnetic resonance spectroscopy.

I. Vliegthart, J. F. G. II. Woods, Robert James, 1962- III. American Chemical Society. Division of Carbohydrate Chemistry. IV. Series.

QP519 9.N83N694 2006
612'.015780113—dc22

2005044892

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

Copyright © 2006 American Chemical Society

Distributed by Oxford University Press

ISBN 10: 0-8412-3953-3

All Rights Reserved. Reprographic copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Act is allowed for internal use only, provided that a per-chapter fee of \$33.00 plus \$0.75 per page is paid to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. Republication or reproduction for sale of pages in this book is permitted only under license from ACS. Direct these and other permission requests to ACS Copyright Office, Publications Division, 1155 16th Street, N.W., Washington, DC 20036.

The citation of trade names and/or names of manufacturers in this publication is not to be construed as an endorsement or as approval by ACS of the commercial products or services referenced herein; nor should the mere reference herein to any drawing, specification, chemical process, or other data be regarded as a license or as a conveyance of any right or permission to the holder, reader, or any other person or corporation, to manufacture, reproduce, use, or sell any patented invention or copyrighted work that may in any way be related thereto. Registered names, trademarks, etc., used in this publication, even without specific indication thereof, are not to be considered unprotected by law.

PRINTED IN THE UNITED STATES OF AMERICA

Foreword

The ACS Symposium Series was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of the series is to publish timely, comprehensive books developed from ACS sponsored symposia based on current scientific research. Occasionally, books are developed from symposia sponsored by other organizations when the topic is of keen interest to the chemistry audience.

Before agreeing to publish a book, the proposed table of contents is reviewed for appropriate and comprehensive coverage and for interest to the audience. Some papers may be excluded to better focus the book; others may be added to provide comprehensiveness. When appropriate, overview or introductory chapters are added. Drafts of chapters are peer-reviewed prior to final acceptance or rejection, and manuscripts are prepared in camera-ready format.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previously published papers are not accepted.

ACS Books Department

Table of Contents

| | |
|--|-----|
| Preface | xi |
| 1 Introduction to NMR of carbohydrates <i>Johannes F. G. Vliegthart</i> | 1 |
| 2 Conformational dynamics of oligosaccharides: techniques and computer simulations <i>Jennie L. M. Jansson, Arnold Maliniak, Goran Widmalm</i> | 20 |
| 3 Structure and dynamics of carbohydrates using residual dipolar couplings <i>James H. Prestegard Xiaobing Yi</i> | 40 |
| 4 Protein-carbohydrate interactions: a combined theoretical and NMR experimental approach on carbohydrate-aromatic interactions and on pyranose ring distortion <i>Jesús Jiménez-Barbero, F. Javier Cañada, Gabriel Cuevas, Juan L. Asensio, Nuria Aboitiz, Angeles Canales, M. Isabel Chávez, M. Carmen Fernández-Alonso, Alicia García-Herrero, Silvia Mari, and Paloma Vidal</i> | 60 |
| 5 Laser photo CIDNP technique as a versatile tool for structural analysis of inter- and intramolecular protein-carbohydrate interactions <i>Hans-Christian Siebert, Emad Tajkhorshid, Johannes F. G. Vliegthart, Claus-Wilhelm von der Lieth, Sabine André, and Hans-Joachim Gabiu</i> | 81 |
| 6 Hydroxy protons in structural studies of carbohydrates by NMR spectroscopy <i>Corine Sandström and Lennart Kenne</i> | 114 |
| 7 Oligosaccharides and cellulose crystal surfaces : computer simulations <i>Bas R. Leeftang, J. Albert van Kuik, and Loes M. J. Kroon-Bratenburg</i> | 133 |
| 8 Biomolecular dynamics: testing microscopic predictions against macroscopic experiments <i>Andrew Almond</i> | 156 |
| 9 Structural insights: towards a molecular comprehension of the GH13 amylase specificity <i>Gwénaëlle André-Leroux</i> | 170 |
| 10 QM/MM simulations in carbohydrates <i>Abdul-Mueed Muslim, Jonathan P. McNamara, Hoda Abdel-Aal, Ian H. Hillier, and Richard A. Bryce</i> | 186 |
| 11 Predicting the three-dimensional structures of anti-carbohydrate antibodies: combining comparative modeling and MD simulations <i>Jane Dannow Dyekjaer and Robert J. Woods</i> | 203 |
| 12 Carbohydrate structure from NMR residual dipolar couplings: is there a correlation between lactose's anomeric configuration and its three-dimensional structure? <i>Darón I. Freedberg, Susan O. Ano, Scott E. Norris, and Richard M. Venable</i> | 220 |
| 13 Incorporation of carbohydrates into macromolecular force fields: development and validation <i>Sarah M. Tschampel, Karl N. Kirschner, and Robert J. Woods</i> | 235 |
| 14 Docking carbohydrates to proteins: empirical free energy functions <i>Alain Laederach</i> | 258 |
| 15 Molecular modeling of carbohydrates with no charges, no hydrogen bonds, and no atoms <i>Valeria Molinero and William A. Goddard III</i> | 271 |
| 16 Molecular modeling of retaining glycosyltransferases <i>Igor Tvaroška</i> | 285 |
| 17 Two unexpected effects found with 2, 3, 4, 6-tetra-O-methyl-D-gluco- and mannopyranosyl oxacarbenium ions : an O-2 pseudo-equatorial preference and a large H-2-C-2--O-2-CH ₃ syn preference <i>Andrei Ionescu, LiJie Wang, Marek Z. Zgierski, Tomoo Nukada, and Dennis M. Whitfield</i> | 302 |
| Author Index | 323 |
| Subject Index | 325 |

Preface

In the past ten years, the interest in the functional and structural aspects of carbohydrates and glycoconjugates has increased enormously. Great insights have been made into the mode of action of glycans in cells, tissues, organs, and organisms. As more and more information becomes known implicating carbohydrates in biological function, increasing pressure is placed on all aspects of biophysical methods to adapt and advance in order to provide structural insight into the biology of carbohydrates and conjugated glycans. Many existing tools that have served admirably to probe structure–function relationships in proteins are difficult or impossible to apply to complex glycans and glycoproteins. Established structural methods, such as X-ray diffraction or NMR spectroscopy also face enormous challenges when applied to the study of biologically relevant glycans in glycoproteins.

Although significant improvements in analytical techniques and instrumentation have reduced the amounts of material needed for unambiguous determination of the primary structure, still difficulties exist in obtaining sufficient material of acceptable purity, due to the presence of microheterogeneity in the glycans present in glycoproteins. In principle, NMR spectroscopy is well suited for work with less homogeneous samples, but limitations associated with molecular weight and intermolecular kinetics attenuate its application to large glycoproteins or protein–polysaccharide complexes. In addition, the flexibility of oligosaccharides, especially those containing 1-6 linkages, introduces unique challenges in interpreting NMR data in terms of three-dimensional structures. However, the developments in high-resolution NMR spectroscopy render it possible to deduce at least the primary sequences of glycoprotein-derived glycans. NMR parameters also allow relevant conclusions to be drawn concerning the confirmation and dynamics of glycans in free or covalently bound form, this data alone is rarely sufficient to completely

determine the three-dimensional structure. Combination with other approaches such as molecular mechanics calculations and molecular dynamics simulations are often indispensable.

To focus the attention of glycoscientists on the scope of present day NMR spectroscopy for the structural studies of carbohydrates and glycoconjugates, Chapters 1–7 of this volume are collected, mainly based on contributions at the American Chemical Society's (ACS) *Tutorial Symposium on NMR Spectroscopy of Carbohydrates* in 2004. Various NMR techniques and applications are presented as advanced introductions to the type of problems that can be studied by these methods. For the functioning of glycans in biological systems, recognition and interaction phenomena with complementary molecules are key features. The investigation of such processes is also referred to in several chapters.

Because of the challenges and limitations of traditional experimental methods, computational techniques can contribute unique insight into the relationship between oligosaccharide structure and biological function. Perhaps in no other field can biomolecular calculations make as major a contribution; whether they serve to provide models to assist in the interpretation of otherwise insufficient experimental data or to provide *a priori* models for the structure of oligosaccharides or insight into the mechanisms of carbohydrate recognition. From the application of quantum mechanics in the study of carbohydrate-processing enzymes to the classical dynamics simulation of oligosaccharides and their protein complexes, the range of systems being examined theoretically is rapidly expanding.

While their promise is great, the potential for computational methods to provide misleading information must be recognized. When applied to experimentally underdetermined systems, as is frequently the case when working with NMR data, the onus is clearly on the theorist to establish the accuracy and suitability of the chosen computational method. Biomolecules are dynamic and none more so than oligosaccharides. Thus, it is imperative that carbohydrate modeling and methodological development address the temporal properties of these molecules as well as the spatial properties. Oligosaccharides frequently populate multiple conformational families, arising from rotation about the glycosidic linkages, that is, they do not generally exhibit well-defined tertiary structures. To adequately sample these conformational properties, which often have lifetimes in excess of 5 ns, requires

extremely long MD simulations to be performed. When combined with the need to include water molecules explicitly, these simulations become some of the most demanding of any biomolecular system.

Given their structural diversity and flexibility, it is not surprising that carbohydrate modeling faces many unique challenges. Examples of many of these issues were discussed at the ACS *Computational Carbohydrate Chemistry Symposium* in 2004 and are highlighted in Chapters 8–16.

The editors hope that the readers will appreciate the various contributions and will profit from them in solving structural problems of glycans. It is a pleasure to thank the Symposium lecturers and the writers of the chapters for their contributions and efforts.

Johannes F. G. Vliegthart

Bijvoet Center
Utrecht University
Padualaan 8
NL–3584CH Utrecht
The Netherlands
j.f.g.vliegthart@chem.uu.nl

Robert J. Woods

Complex Carbohydrate Research Center
University of Georgia
315 Riverbend Road
Athens, GA 30602

Chapter 1

Introduction to NMR Spectroscopy of Carbohydrates

Johannes F. G. Vliegthart

Bijvoet Center, Department of Bioorganic Chemistry, Utrecht University,
Padualaan 8, 3584 CH Utrecht, The Netherlands

In this chapter an introductory overview is presented of advances in NMR spectroscopy of carbohydrates. The main emphasis is on the application of ^1H -NMR spectroscopy for identification and structural studies of glycans.

Introduction

The application of NMR spectroscopy to carbohydrates has a relatively long history, but its suitability for structural analysis has increased enormously in recent years (1).

The NMR spectroscopy of biomolecules in general has undergone an almost complete revolution in the past 25 years. Spectacular developments in the instrumentation, pulse sequences, spectral interpretation, isotope labeling of compounds and molecular modeling techniques have led to new possibilities to determine the primary structure and the three-dimensional structure of biomolecules in solution (2). The high resolutions that can be obtained with the most advanced spectrometers allow the unraveling of details of the structure and

render possible the study of the molecular dynamics in solution. Even for large molecules significant information can be derived. The most impressive progress has been made for proteins and nucleic acids. For these compounds the main chain, the side-chains of the constituting residues and the homo- and heterotypic interactions can be established with a high degree of accuracy. The advances in isotope labeling through cloning techniques and organic chemistry have stimulated the further development of the direct and indirect spectral detection of various nuclei.

For carbohydrates and glycoconjugates ^1H and ^{13}C have proved to be extremely valuable to determine primary structures. In fact the characterization of (partial) structures of glycoprotein-derived N-glycans has greatly facilitated the unraveling of biosynthetic routes and studying the functional roles of these glycans in complex biological systems. Another important aspect concerns the confirmation of the identity of glycan structures that are supposed to be identical to known compounds.

Owing to the inherent flexibility of carbohydrate chains, the characterization of the three-dimensional structure in solution is rarely feasible into the same detail as for proteins and nucleic acids. Nevertheless, interesting results have been obtained. For the study of the interaction of carbohydrates with complementary compounds NMR spectroscopy can be a valuable tool (1). The labeling of carbohydrates and glycan chains with isotopes is a bit more cumbersome for carbohydrates than for other bio-macromolecules. Glycan-labeling is eagerly waiting for further innovations. In addition to ^1H and ^{13}C spectra also those of the nuclei ^{15}N , ^{17}O , ^{19}F , ^{31}P (whether or not fully isotopically enriched) and of several metal ions in carbohydrates have been recorded. Obviously, resolution and sensitivity are different for the various nuclei and thereby decisive for the type of information that can be extracted from the spectra.

This introductory chapter will mainly be focused on ^1H -NMR spectra carbohydrates and glycoconjugates.

^1H NMR spectra of glycans and the reporter group concept

Usually, the spectra of unprotected glycans are recorded in $^2\text{H}_2\text{O}$ after full exchange of the exchangeable protons (3.4). Spectra recorded at NMR machines operating at 500 MHz or at higher frequencies contain sufficient details to be used as identity card (5). For the (partial) assignment of the resonances in novel compounds additional NMR experiments are needed. For the characterization of compounds described in literature, mostly comparison of the spectral data to reference data is sufficient. Two groups of signals can be distinguished. First,

the so-called bulk signal containing mainly the non-anomeric protons, present in a rather narrow spectral range between 3.2 and 3.9 ppm. Secondly, the structural-reporter-group signals that are found outside the bulk region (6-8). The chemical shift patterns of the structural reporter groups comprising chemical shifts and couplings are translated into structural information, based on a comparison to patterns in a library of relevant reference compounds. The comparison of NMR data for many closely related glycans resulted in empirical rules to correlate chemical shift values with carbohydrate structures. Successful application of this approach requires accurate calibration of the experimental conditions like sample temperature, solvent and pH (3,4).

The structural reporter group signals can be subdivided into the following categories:

Anomeric protons: shifted downfield, due to their relative unshielding by the ring oxygen atom.

Protons that can be discerned outside of the bulk region, as a result of glycosylation shifts, or under influence of substituents such as sulfate, phosphate and acyl groups.

Deoxysugar protons.

Alkyl and acyl substituents like methyl and acetyl, glycolyl, pyruvate, respectively.

The NMR database 'sugabase' to identify glycan structures has been founded on such assignments (8-10). It should be emphasized that definitive conclusions on the identity of novel compounds invariably require validation by experimental data from independent approaches. Since the reporter group signals are relatively insensitive for alterations in the structural elements remote from the corresponding locus, the structural-reporter group concept has proved its usefulness for the identification of numerous compounds. In particular, the concept is invaluable for the analysis of glycoconjugate-glycans that form an ensemble of closely related compounds.

2D-spectra

To assign resonances in the region of the bulk signal and of coinciding structure reporter group signals, 2-D homonuclear correlation type of spectra, such as various COSY or TOCSY experiments are needed. In this way spin systems corresponding to monosaccharide constituents can be traced. In general the interpretation of such spectra can start from an anomeric signal or from any other well-resolved signal. For compound I (see Fig. 1), a heptasaccharide methyl β -glycoside, corresponding to a low molecular mass glycan of α -D-hemocyanin of the snail *Helix Pomatia* (11), the TOCSY spectrum is depicted in Fig.2 (12).

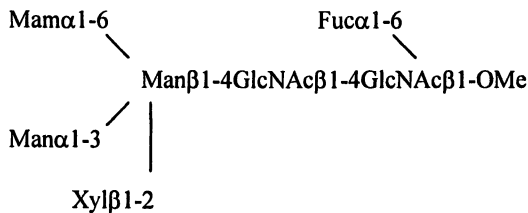


Figure 1. Structure of compound 1. The monosaccharide constituents are in the text and spectra abbreviated as: 4' = Man α 1-6, 4 = Man α 1-3, 3 = Man β 1-4, 2 = GlcNAc β 1-4, 1 = GlcNAc β 1-, X = Xyl β 1-2, F = Fucal-6, OMe = O-Methyl.

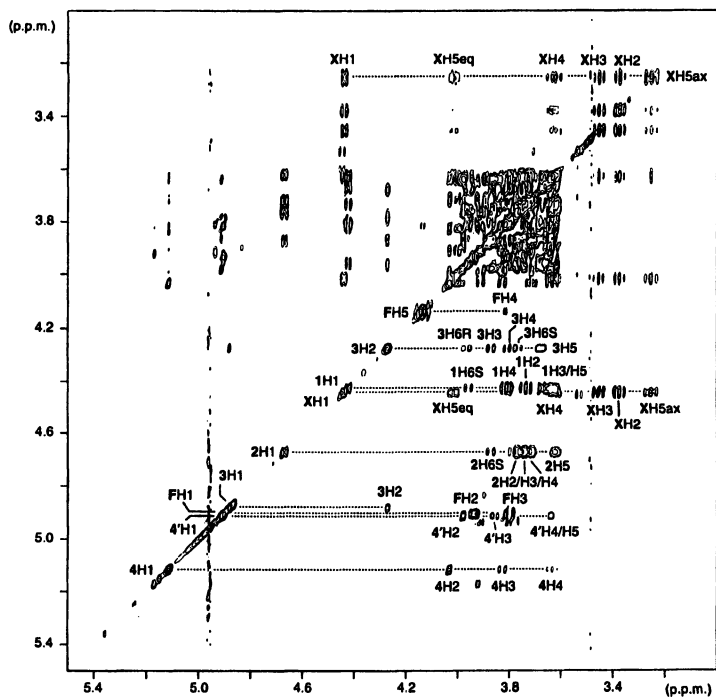


Figure 2. $^1\text{H}-^1\text{H}$ TOCSY spectrum at 500 MHz, 281.5 K, mixing time 100ms.

Homonuclear COSY and TOCSY spectra do not provide monosaccharide sequence information, due to the absence of coupling over the glycosidic linkage. Often NOESY or ROESY spectra are used for this purpose. In many cases the most intense NOESY peak identifies the linkage, but not always. In Fig. 3 the ROESY spectrum of compound 1 is presented (12).

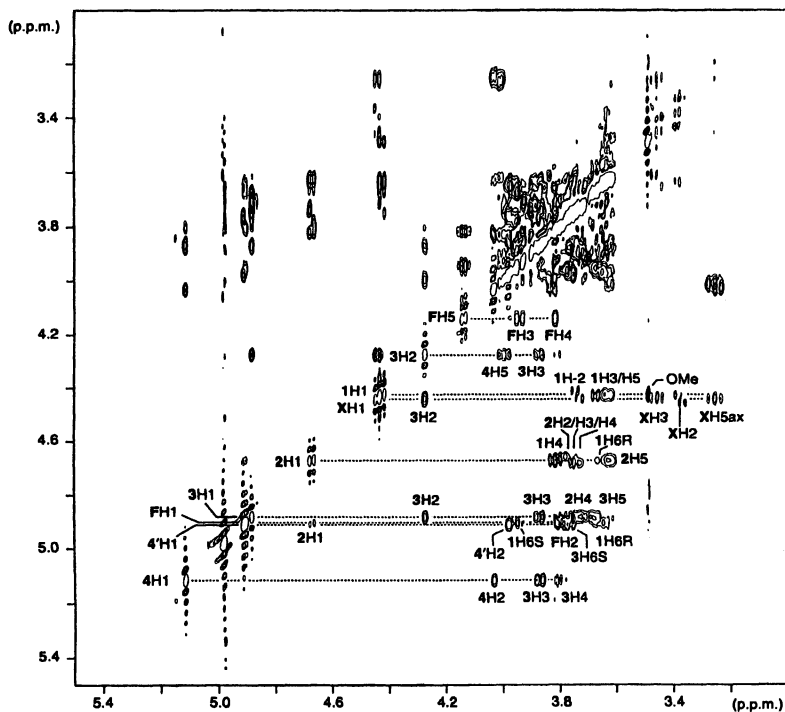


Figure 3. ROESY spectrum of compound I at 500 Mz, 281.5 K, mixing-time 155 ms.

^1H - ^{13}C Heteronuclear multiple and single quantum coherence spectra,

^1H - ^{13}C HMQC and HSQC spectra provide important correlations. Owing to the usually large dispersion of the ^{13}C signals, these spectra have a great value for the interpretation. Often in the HMQC spectra the one-bond correlations are

filtered out. In our laboratory this filter is hardly used, because omitting the filter gives many one-bond correlations as well, which are convenient in the assignment process. The HMQC spectrum of compound 1 is shown in Fig. 4 (12).

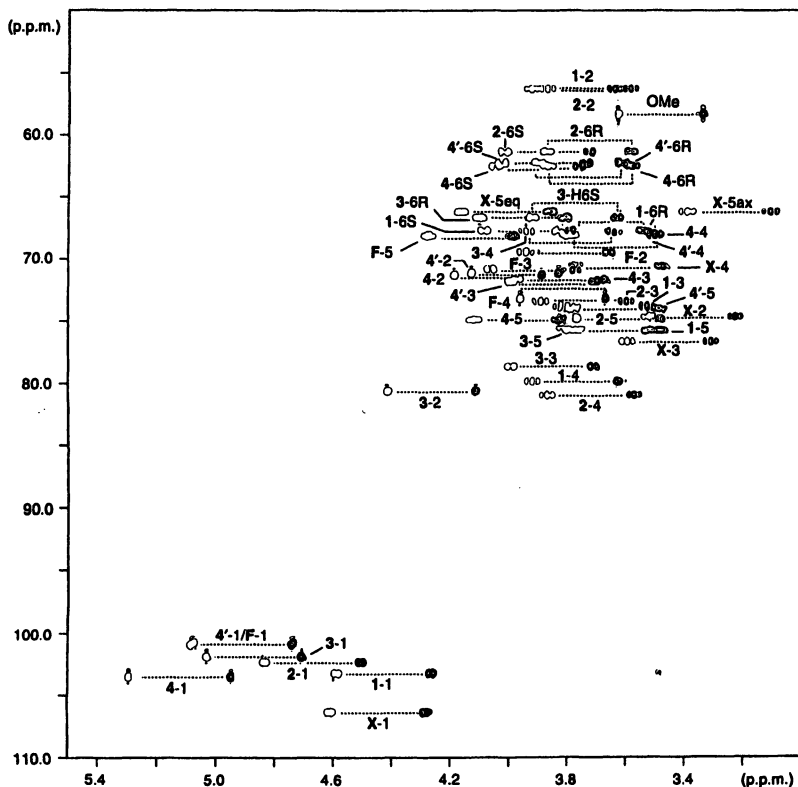


Figure 4. ^1H - ^{13}C HMQC spectrum of compound 1, 300 K, abbr. according to the legend of Fig. 1, second number refers to the ^1H and ^{13}C positions in the monosaccharide.

In determination of the structure of the repeating unit of polysaccharides HSQC spectra are very helpful. This is demonstrated in Fig. 5 for the repeating unit of the exopolysaccharide from *Streptococcus thermophilus* S3 (13). A complication in the structure stems from the partial acetylation of constituent B. So far it was not feasible to elucidate the distribution of the acetyl groups over the polysaccharide chain.

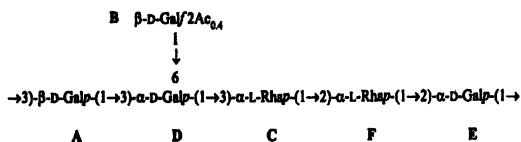


Figure 5. Structure of the repeating unit of the exopolysaccharide of *Streptococcus thermophilus* S3.

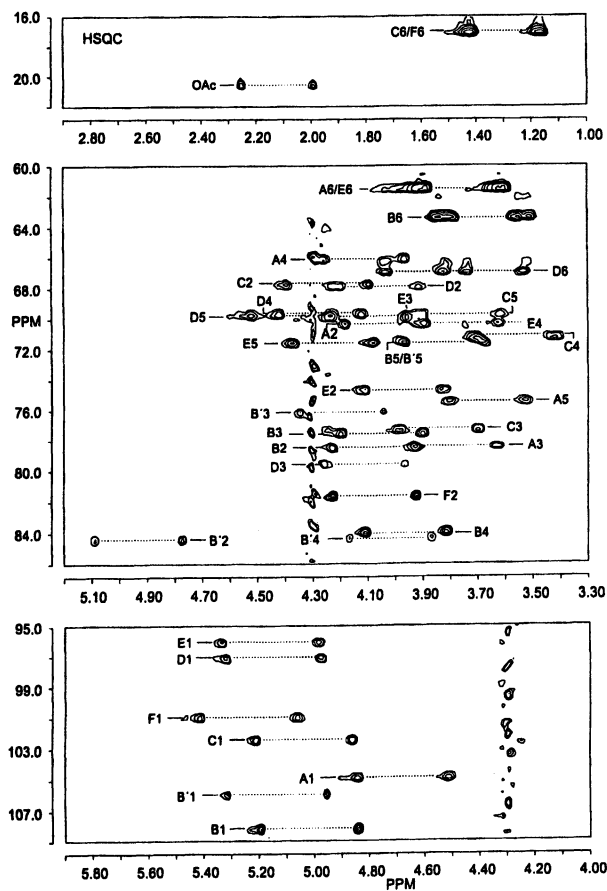


Figure 6. 500-MHz $2D$ 1H - ^{13}C undecoupled HSQC spectrum of the native exopolysaccharide of *Streptococcus thermophilus* S3 at 340 K (13).

The heteronuclear multiple-bond spectroscopy (HMBC) is basically a HMQC experiment tuned at small (1-10 Hz) heteronuclear couplings. For sequence information the HMBC experiment is useful, since it yields through bond correlations. The spectrum provides cross peaks between each anomeric

carbon atom and the proton of the adjacent sugar residue as well as between the anomeric proton and the ring carbon atom of the adjacent sugar residue.

In particular in assigning the structure of the repeating units in polysaccharides, these spectra are very helpful. This is illustrated for the repeating unit of the exopolysaccharide of *Lactobacillus delbrueckii subspecies bulgaricus* (Fig. 7) in the HMBC spectrum shown in Fig. 8 (14).

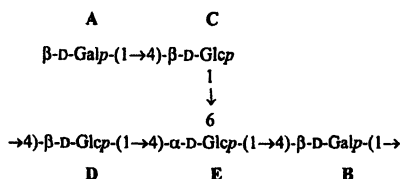


Figure 7. Structure of the repeating unit of the exopolysaccharide of *Lactobacillus delbrueckii subspecies bulgaricus*.

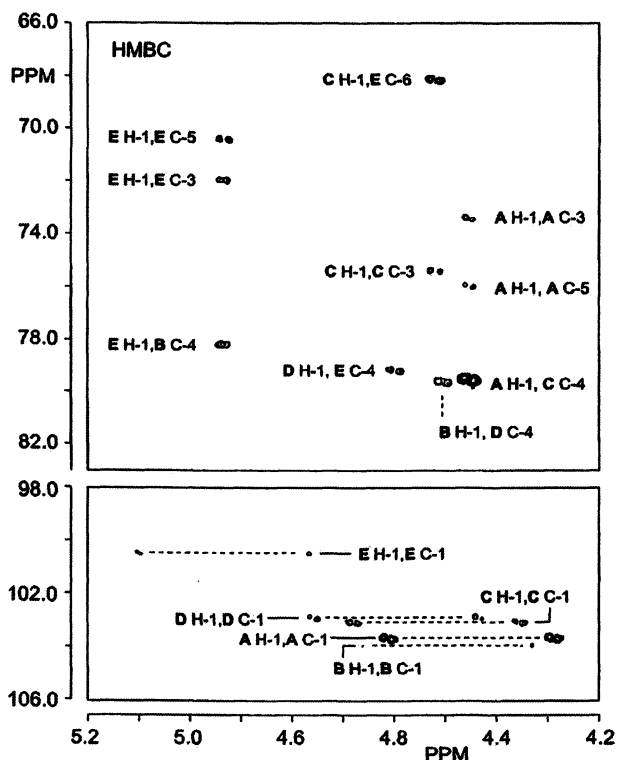


Figure 8. Partial 500-MHz ^1H - ^{13}C undecoupled HBQC spectrum of the polysaccharide from *Lactobacillus delbrueckii subspecies bulgaricus* at 353 K.

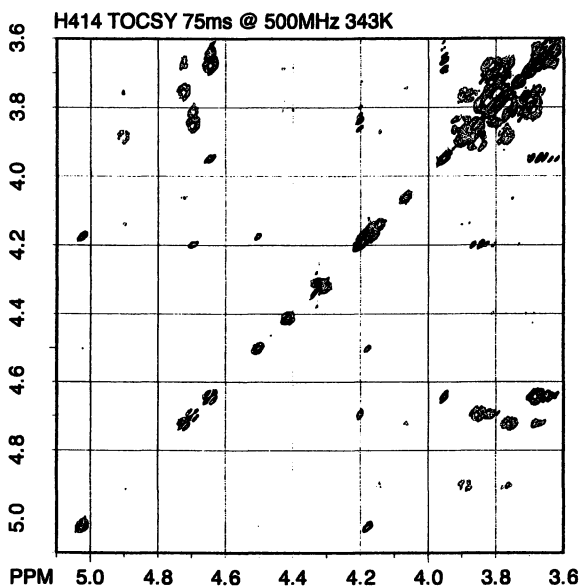


Figure 11. TOCSY spectrum of the exopolysaccharide at 500 MHz.

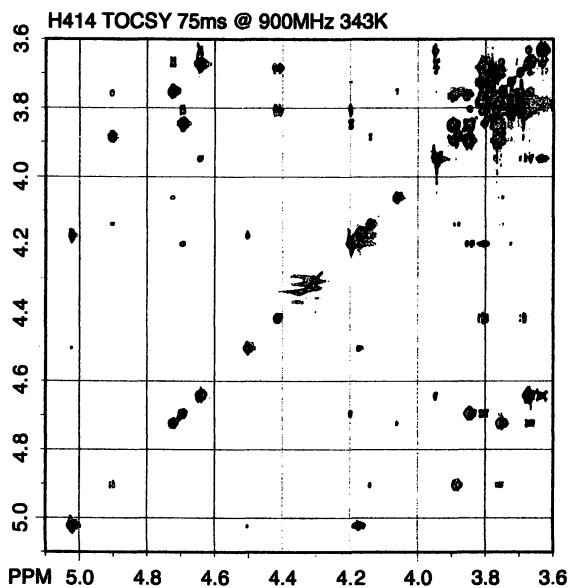


Figure 12. TOCSY spectrum of the exopolysaccharide at 900 MHz.

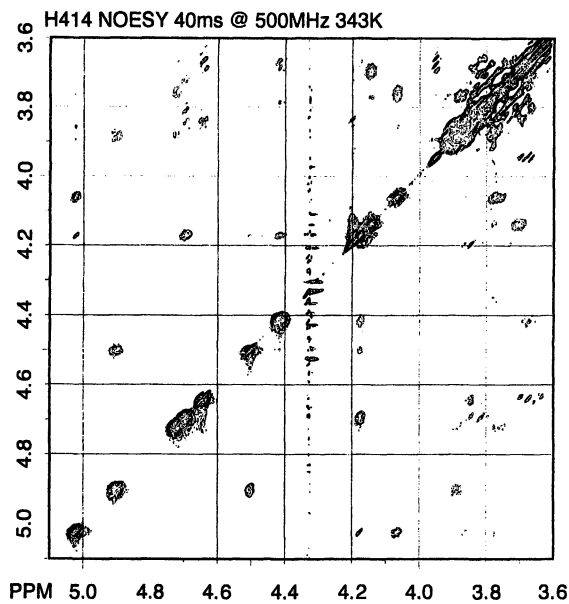


Figure 13. NOESY spectrum of the exopolysaccharide at 500 MHz.

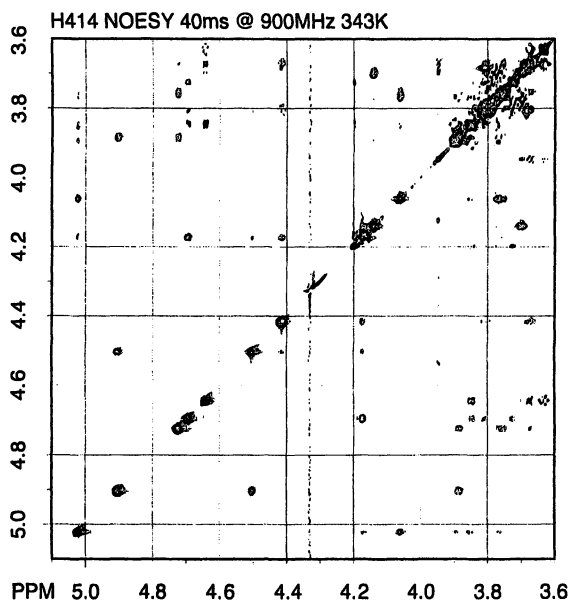


Figure 14. NOESY spectrum of the exopolysaccharide at 900 MHz.

3-D NMR spectra

Owing to the small chemical shift dispersion in carbohydrate NMR spectra these molecules are well suited for 3-D NMR.

In principle any two 2-D methods can be combined into a 3-D experiment. Non-selective homonuclear 3-D NOESY-TOCSY was the first method applied to an oligosaccharide (16, 17) Suitable 2-D alternatives have been developed since then (18, 19), which are more efficient in the data collection.

Conformation

The characterization of the spatial structure in solution of glycans, free, in a complex with complementary molecules, or covalently bound to other molecules like proteins or lipids, is relevant for gaining insight into the functioning of these molecules and for their recognition by complementary molecules. In the last case, it should be emphasized that by such interactions the conformational equilibrium may undergo alterations in comparison to the free state, since carbohydrates have an inherent flexibility and can adapt to the binding site. Similarly, in glycoconjugates the glycan conformation can be affected by interaction with non-glycan part. Estimation of the overall solution structure requires determination of the conformation of the constituting monosaccharides and of all glycosidic linkages. An oligosaccharide chain manifests flexibility throughout the whole molecule on a *short* time scale by fast vibrations at bonds and angles and on a *longer* time scale by changes of the dihedral angles. The conformationally relevant dihedral angles ϕ , ψ and ω are indicated in Fig. 15.

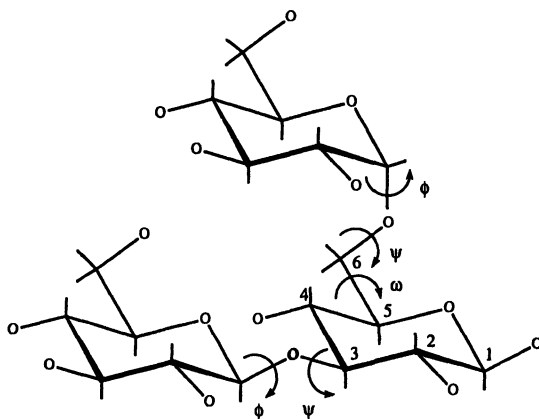


Figure 15. Conformational angles ϕ , ψ , and ω according to IUPAC-IUB (20).

Information on the rate of tumbling and the internal flexibilities can be deduced from the longitudinal relaxation time T_1 .

^1H - ^1H coupling constants are important for exploring the ring conformations of the constituents and for an estimation of the dihedral angle ω for the glycosidic linkages via an exocyclic hydroxymethyl group.

The skeleton protons of the constituting monosaccharide can be studied by homonuclear ^1H -NOE experiments. The smallest NOEs that can be measured correspond to a distance between the protons of about 5 Å. Several *intra*-residual ^1H - ^1H NOEs can be observed, but only one or two *inter*-residual NOEs can be measured for consecutive monosaccharides. Only the latter are relevant for an overall 3-D structure. NOEs between further remote constituents are rare in non-rigid chains.

Heteronuclear ^1H - ^{13}C NOEs are useful by giving access to the calculation of distances between carbon atoms and protons. The three-bond ^1H - ^{13}C NOEs are suitable for the determination of the glycosidic linkage conformation.

Complications in the analysis of NOEs arise from internal motions and flexibility of the carbohydrate chain. When in comparison to the NOE build up rate, fast internal motions occur within a chain, they will affect the NOE build up rates and intensities. Since the observed NMR parameters are averages of the different coexisting conformations, unrealistic results may be obtained by translating the measured NOEs directly into distances.

Therefore, complementary approaches are necessary to identify the distinct conformations and to determine their abundance, *e.g.* through a combination of Molecular Mechanics (MM) and Molecular Dynamics (MD) calculations. The MM calculations are used to search for energy minima *in vacuo* in the conformational space. The MD calculations are based on a derivative of the potential energies to calculate the accelerations of the individual atoms using Newton's law: $F = m \times a$.

Velocities of the atoms are then calculated and after integration of a time step of typically 1 - 2 fs, the molecular system gets new coordinates. This process is repeated as long as the MD simulation lasts.

The obtained trajectories of a molecule through space in time make it possible to extract time-dependent parameters like correlation time, diffusion constants and free energies. During the MD simulation time-averaged NOE constraints are applied. A time averaged interval of the MD trajectory is evaluated with the NOE constraints, so the molecular system gets the opportunity to adapt its conformation.

In our laboratory, the MD simulations are carried out in water by means of the GROMOS program to probe the solution-structure (21).

Other approaches that yield independent information on the conformations involved can be obtained from NMR residual dipolar couplings in aqueous, dilute liquid crystalline media (see chapters by Prestegård and Widmalm in this volume) and from uniform or specific ^{13}C -labeling of the constituents.

Conformational study compound 1

In the framework of our investigations of the solution structure of intact glycoproteins, we studied the conformational behavior of several glycan chains. As an illustration of the approach presented in the preceding section of this chapter, the results will be briefly summarized for compound 1; for its structure see Fig. 1 (12):

After assignment of the ^1H and ^{13}C NMR signals, MD simulations in water and NOESY and ROESY experiments were performed to study the conformational behavior of compound 1. The measured nuclear Overhauser effects were compared to calculated values derived from the MD trajectories by applying the CROSSREL program (22) This afforded insight into occupied conformations of the glycosidic linkages, as expressed by the generalized order parameters and internal rotation correlation times.

In more detail: HSEA energy calculations were carried out with the GEGOP program (23). The HSEA ϕ , ψ iso-energy contour plots for the glycosidic linkages were obtained in steps of 5° in ϕ , ψ of the constituting disaccharides and plotted at intervals of 1.5 kcal/mol.

The MD simulations in water and comparisons of calculated with observed NOE and ROE intensities data indicate that the part consisting of:



is relatively rigid; each linkage occupies a single region of the ϕ , ψ space. In case conformational transitions would occur, then they would take place in a longer time period than that of the overall molecular rotation time (>1 ns).

For the other glycosidic linkages:

$\text{Man}\alpha 1-6\text{Man}\beta 1-$;
 $\text{Man}\alpha 1-3\text{Man}\beta 1-$; and
 $\text{Fuc}\alpha 1-6\text{GlcNAc}\beta 1-$,

a clear flexibility was found. Single conformations are not compatible with the experimental NOE and ROE data, in contrast to ensembles of conformations for each of these linkages.

For the $\text{Man}\alpha 1-3\text{Man}\beta 1-$ linkage the slow-internal rotational correlation time $\tau_{s,k}$ indicate that the rate of large conformational transitions is relatively small, but for $\text{Man}\alpha 1-6\text{Man}\beta 1-$ and $\text{Fuc}\alpha 1-6\text{GlcNAc}\beta 1-$ the transition rates are approximately 7 times higher. The (1-6) linkages have a high mobility as could be derived from the order parameter S_f^2 .

The $\text{Man}\alpha 1-6$ and $\text{Fuc}\alpha 1-6$ residues explore large parts of the conformational space. The superimposed conformations of compound 1 are depicted in Fig. 16.

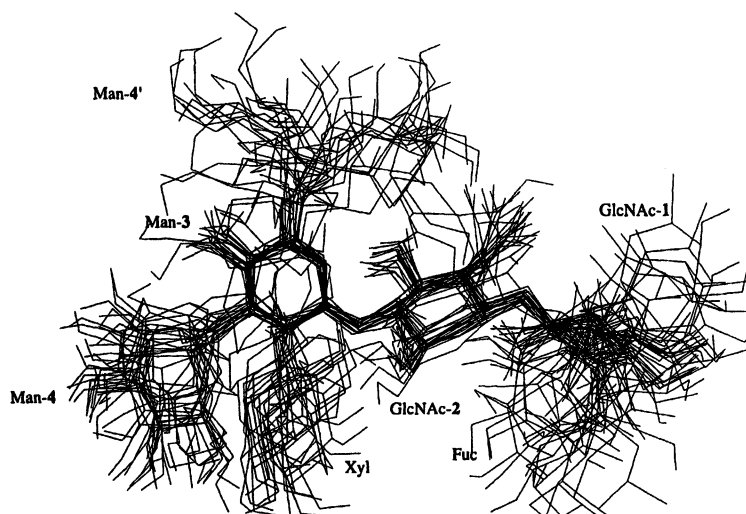


Figure 16. Superimposed conformations of compound 1 obtained from MD simulation.

Pineapple Stem Bromelain: Glycopeptide versus Glycoprotein

The glycoprotein pineapple stem bromelain is a proteolytic enzyme carrying a single N-glycosylation site, only. From bromelain a glycopeptide was prepared by exhaustive pronase digestion (24) (see Fig. 17).

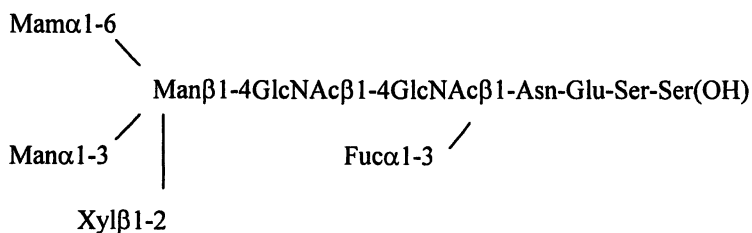


Figure 17. Structure of the glycopeptide derived from pineapple stem bromelain.

The conformational behavior of the glycopeptide was investigated by using a combination of MD simulations in water with NOESY ^1H NMR spectroscopy (24, 25). Theoretical NOESY peak intensities were calculated on the basis of models obtained by MD simulations through the CROSSREL program (22). The peptide moiety shows a large flexibility and extends in a larger area of the conformational space than the glycan part. There are no strong interactions between the glycan and peptide chain, although in the MD simulations of the glycopeptide several hydrogen bonds were indicated. In particular, hydrogen bonds were found for Asn and GlcNAc-1 for 5.5%. ($\text{N-N}_\delta \cdots \text{G1-O7}$) and between Ser-1 and GlcNAc-1 ($\text{S1-O}_\gamma \cdots \text{G1-O6}$) for 13.9% of the simulated time. The presence of temporary hydrogen bonds illustrates the proximity of the peptide and glycan parts, almost without inducing preferred conformations for the peptide. In the glycan the $\text{Xyl}\beta\text{1-2Man}\beta\text{1-4GlcNAc}\beta\text{1-4GlcNAc}$ part is rather rigid, occurring in mainly one conformation. The $\text{Man}\alpha\text{1-6Man}$ and the $\text{Fuc}\alpha\text{1-3GlcNAc}$ linkages are more flexible as can be concluded from the MD simulation and the S_f^2 values. The latter linkage occurs in two conformations at average values of $\phi, \psi = -145, 100$ and $\phi, \psi = -140, -135$. Transitions of the glycosidic linkages occur at a time scale of about 1 – 100 ns. In Fig. 18 superimposing the various conformations summarizes the results.

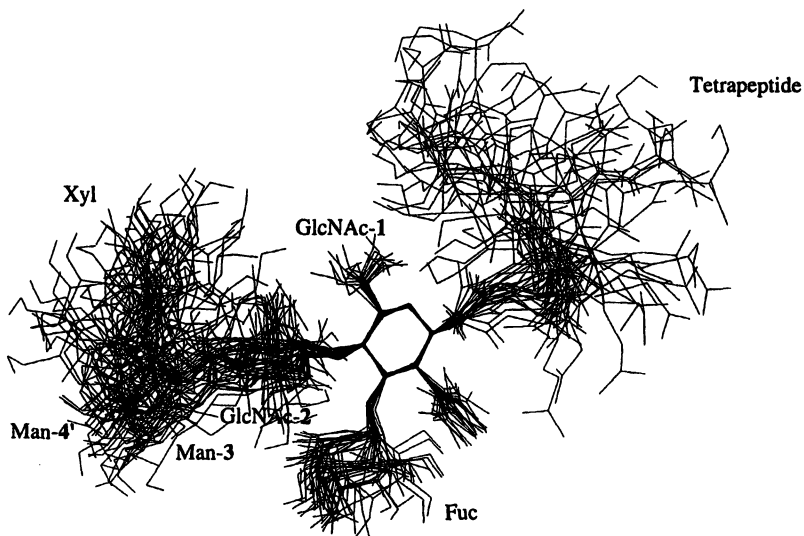


Figure 18. Snapshots of structures of the glycopeptide taken at each 20 ps of the simulated trajectory. The conformations are superimposed on the center of mass the GlcNAc-1 residue. For the $\text{Man}\alpha\text{1-6Man}$ linkage 6 gg and 7 gt conformations are included.

Interestingly, investigations of the intact bromelain glycoprotein showed a number of significant differences in the glycan conformation (26). All glycosidic linkages showed a greatly reduced flexibility. This holds in particular for the Man α 1-6Man and Fuc α 1-3GlcNAc linkages, probably due to interactions with the protein. For the Fuc α 1-3GlcNAc linkage also the distribution between the populations of the two main conformations is affected.

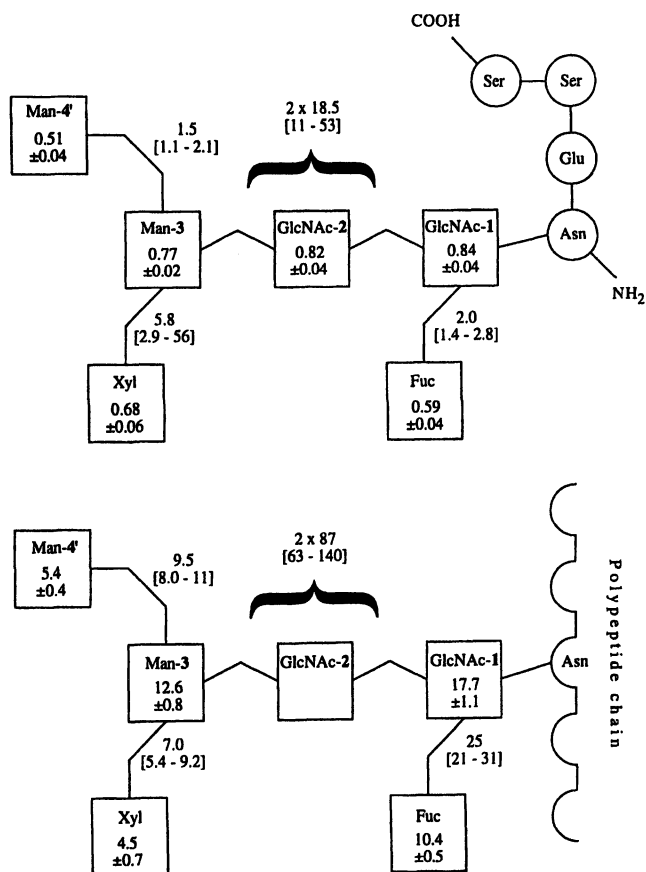


Figure 19. Effective rotation correlation times of the monosaccharide residues and the slow internal rotation correlation times of the glycosidic linkages in the glycopeptide (upper part) and the glycoprotein (lower part).

Comparison of the effective rotation correlation times of the constituting monosaccharides and of the slow-internal rotation correlation times of the glycosidic linkages demonstrate the additivity of motions in a glycan chain going from the center of mass to the terminus/termini of the chain and the effect of a large protein versus a small oligopeptide. The rigidity of the *N,N*-diacetylchitobiose unit and the site of interaction (α or β site, or end-on positioning) of the Asn-linked GlcNAc residue with the protein characterize to a large part the presentation of the inner part of the glycan. The additivity of the motions of the individual residues leads to a relatively large flexibility at the outer part(s) of the glycan in a glycoprotein. This feature has important implications for the interaction with complementary molecules and recognition phenomena as well as on the accessibility for enzymes. The comparison of the conformations of free glycans and attached to small peptides with those of glycans in glycoproteins show clearly that the location of the center of mass and the interactions with the protein chain have a great influence. By consequence, conformational data obtained for free glycans or glycans attached to small peptides should not be generalized to glycoproteins without carrying out further experimental studies.

In summary: NMR has proven to be an invaluable tool in determining primary and three-dimensional structures of carbohydrates.

References

1. *NMR spectroscopy of Glycoconjugates*; Jimenez-Barbero, J., Peters, T., Eds.; Wiley-VCH Weinheim, Germany, 2003.
2. Simon, B., & Sattler, M. *Angew. Chem., Int Ed Engl* 2004, 43, 782-786.
3. Hård, K. & Vliegthart, J.F.G.; *Glycobiology, a practical approach*; Fukuda, M. & Kobata, A. Eds.; IRL Press at Oxford University Press, Oxford, U.K., 1993, 223-242.
4. Leeftang, B.R. & Vliegthart, J.F.G. *Encyclopedia of Analytical Chemistry*; Meyers, R.A, Ed.; John Wiley & Sons Ltd. Chichester, USA, 2000, 821-195.
5. Vliegthart, J.F.G., van Halbeek, H., & Dorland, L. *Pure and Applied Chemistry* 1981, 53, 45-77.
6. Vliegthart, J.F.G., Dorland, L. & van Halbeek, H. *Adv. Carbohydr. Chem. Biochem.* 1983, 41, 209-374.
7. Kamerling, J.P. & Vliegthart, J.F.G.; *Biol. Magn. Res.*; Berliner, L.J. & Reuben, J. Eds.; Plenum Corporation, New York, U.S. 1992, 1-194.
8. Van Kuijk, J.A. & Vliegthart, J.F.G., *Trends Glycosc. Glycotechnol.* 1991, 3, 115-122.
9. Van Kuijk, J.A., Hård, K. & Vliegthart, J.F.G., *Carbohydr. Res.* 1992, 235, 53-68.

10. <http://www.boc.chem.uu.nl/static/sugabase/sugabase.html>
11. Van Kuijk, J.A., van Halbeek, H., Kamerling, J.P. & Vliegthart, J.F.G. *J. Biol. Chem.* **1985**, *260*, 13984-13988.
12. Lommerse, J.P.M., van Rooijen, J.J.M., Kroon-Batenburg, L.M.J., Kamerling, J.P. & Vliegthart, J.F.G., *Carbohydr. Res.* **2002**, *337*, 2279-2299.
13. Faber, E.J., van der Haak, M.J., Kamerling, J.P. & Vliegthart, J.F.G. *Carbohydr. Res* **2001**, *331*, 173-182.
14. Faber, E.J., Kamerling, J.P. & Vliegthart, J.F.G. *Carbohydr. Res* **2001**, *331*, 183-194.
15. Gruter, M., Leeftang, B.R., Kuiper, J.P. Kamerling, J.P. & Vliegthart, J.F.G., *Carbohydr. Res.* **1992**, *231*, 273-291.
16. Vuister, G.W., de Waard, P., Boelens, R., Vliegthart, J.F.G. & Kaptein, R. *J. Am. Chem. Soc.* **1989**, *111*, 263-270.
17. De Waard, P., Boelens, R., Vuister, G.W. & Vliegthart, J.F.G., *J. Am. Chem. Soc.* **1990**, *112*, 3232-3234.
18. De Waard, P., Leeftang, B.R., Vliegthart, J.F.G., Boelens, R. Vuister, G.W. & Kaptein, R., *J. Biomol. NMR* **1992**, *2*, 211-226.
19. De Beer, T., van Zuylen, C.W.E.M., Hård, K., Boelens. R., Kaptein, R. Kamerling, J.P. & Vliegthart, J.F.G. *FEBS Lett.* **1994**, *348*, 1-6.
20. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), *Eur. J. Biochem.* **1983**, *131*, 5-7.
21. <http://www.igc.ethz.ch/gromos-docs/index.html>
22. Leeftang, B.R. & Kroon-Batenburg, L.M.J. *J. Biomol. NMR* **1992**, *2*, 495-518.
23. Struik-Prill, R. & Meyer, B. *Eur. J. Biochem.* **1990**, *194*, 903-919.
24. Bouwstra, J.B., Spoelstra, E.C., de Waard, P., Leeftang, B.R., Kamerling, J.P. & Vliegthart, J.F.G. *Eur. J. Biochem.* **1990**, *190*, 113-122.
25. Lommerse, J.P.M., Kroon-Batenburg, L.M.J., Kamerling, J.P. & Vliegthart, J.F.G. *J. Biomol. NMR* **1995**, *5*, 318-330.
26. Lommerse, J.P.M., Kroon-Batenburg, L.M.J., Kamerling, J.P. & Vliegthart, J.F.G. *Biochemistry* **1995**, *34*, 8196-8206.

Chapter 2

Conformational Dynamics of Oligosaccharides: NMR Techniques and Computer Simulations

Jennie L. M. Jansson^{1,2}, Arnold Maliniak², and Göran Widmalm¹

¹Department of Organic Chemistry and ²Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91, Stockholm, Sweden

NMR spectroscopy techniques in conjunction with molecular dynamics simulations facilitate description of conformation and dynamics of oligosaccharides in solution. Herein we describe approaches based on hetero-nuclear carbon-proton spin-spin coupling constants useful for assessing conformational preferences at the glycosidic linkage, exemplified for α -cyclodextrin. Furthermore, we utilize hetero-nuclear carbon-proton residual dipolar couplings together with molecular dynamics simulations in the analysis of the conformational dynamics of the milk oligosaccharide Lacto-*N*-neotetraose.

Introduction

Carbohydrates play many important roles in biological systems where they often are found as glycoconjugates, e.g., in the form of glycolipids or glycoproteins as part of a biological membrane (Figure 1). In glycoproteins the glycan part, being of significant size, can shield regions of the surface of the protein thereby providing protection from proteases. Furthermore, the state of glycosylation is important in different processes such as trafficking and protein

folding (1,2). Likewise, the extent of glycosylation also serves as a regulatory modification and is important for signal transduction cascades. In man, genetic defects that are due to alterations in the biosynthetic pathways of *N*-linked glycoproteins may lead to severe medical consequences and these are referred to as congenital disorders of glycosylation (3). Recognition of carbohydrates is performed by a special group of proteins, known as lectins, and these interactions are of major importance in, e.g., cell-cell interactions (4).

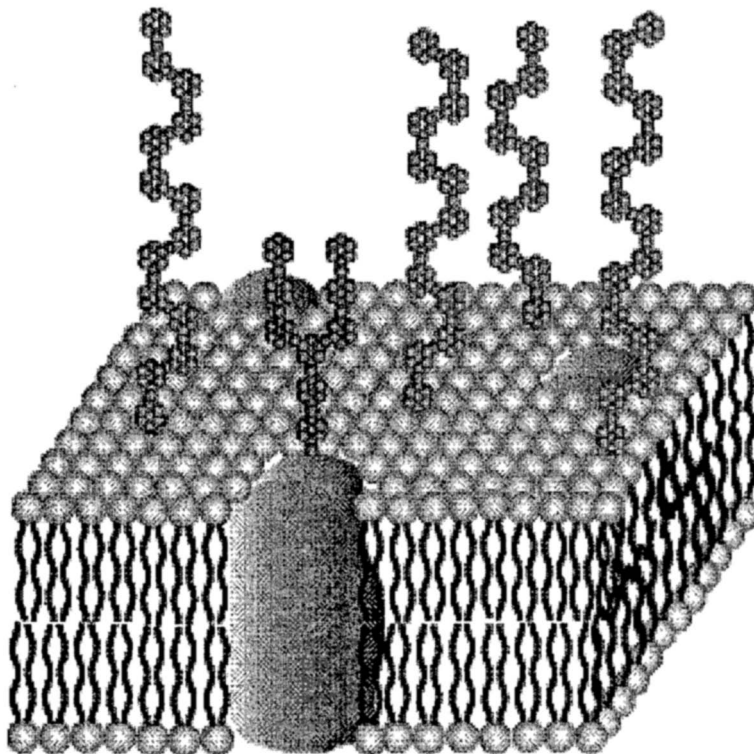


Figure 1. Schematic of a biological membrane containing, inter alia, phospholipids, glycolipids and glycoproteins.

A number of techniques exist to study carbohydrate structure, interaction and function. One of the most important is undoubtedly NMR spectroscopy and a great many approaches have been developed to this end (5-7). In these studies, closely related at least from an interpretation point of view, various modeling techniques and in particular molecular dynamics (MD) simulations are of importance (8-12). In our hands, four NMR approaches are particularly useful in studying carbohydrate conformation and dynamics, viz., (i) ^1H , ^1H nuclear

Overhauser effects (NOEs); (ii) ^{13}C nuclear spin relaxation; (iii) $^3J_{\text{C,H}}$ trans-glycosidic spin-spin coupling constants; and (iv) ^1H , ^{13}C and ^1H , ^1H residual dipolar couplings (RDCs). In the following we will highlight the two latter techniques and in addition describe some use of MD simulations in the interpretation of results from NMR spectroscopy experiments.

Hetero-nuclear carbon-proton couplings

Heteronuclear carbon-proton coupling constants are important in carbohydrate conformational analysis since they can shed light on conformational preferences at the glycosidic linkage via Karplus-type relationships. There are several ways to determine long-range carbon-proton coupling constants (13-15). A robust method applicable to oligosaccharides is our extension (16,17) of the scheme proposed by Freeman and coworkers (18) in which the hetero-nuclear carbon-proton coupling over two or more bonds are detected as an anti-phase splitting of the resonance in the ^1H NMR spectrum. The pulse sequence is shown in Figure 2.

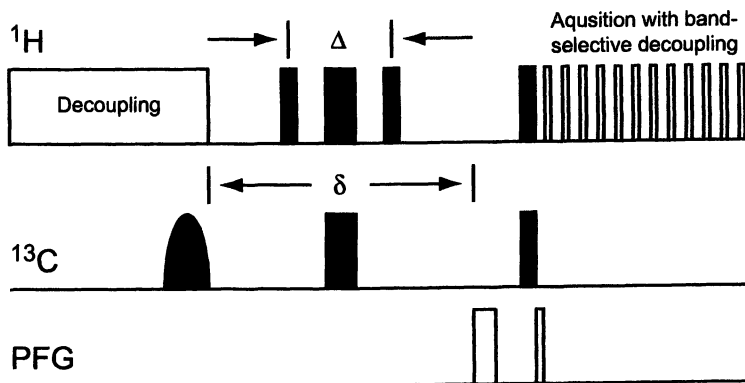


Figure 2. NMR pulse sequence used to determine long-range carbon-proton coupling constants, where $\delta = 1/2 {}^nJ_{\text{C,H}}$ and $\Delta = 1/{}^1J_{\text{C,H}}$.

Selective excitation under proton decoupling conditions is often performed on several ^{13}C resonances using Hadamard matrix encoding from which individual traces subsequently can be obtained. The delay $\delta = 1/2 {}^nJ_{\text{C,H}}$ is usually set to a shorter time than that of the nominal value calculated from the anticipated long-range coupling constant. The BIRD module with $\Delta = 1/{}^1J_{\text{C,H}}$

at the mid-point of the δ interval is used to refocus the hetero-nuclear one-bond splittings. At the end of the pulse sequence magnetization is transferred to the protons for detection. The use of pulsed field gradients ensures artifact-free spectra (17). In addition, in-phase homo-nuclear proton couplings can be eliminated by band-selective decoupling during acquisition.

The NMR signal in the time domain of an anti-phase doublet can be described by:

$$S(t) = \sin(\pi J_{C,H}t) \exp(-\lambda t) \quad (1)$$

where λ describes the decay of the signal. After Fourier transformation (FT) a typical pattern can be obtained as shown in Figure 3a. In this spectrum ${}^nJ_{C,H}$ is readily determined from the anti-phase splitting. However, line-width increase in combination with complex proton multiplicity and similar magnitude of homo- and hetero-nuclear couplings render a direct estimate of ${}^nJ_{C,H}$ problematic. One way to extract the couplings is to apply the J doubling procedure (19-21).

In the time domain an anti-phase double doublet is described by:

$$S(t) = \sin(\pi J_{C,H}t) \exp(-\lambda t) \cos(\pi J^*t). \quad (2)$$

This is also equivalent to J doubling one stage, where J^* is a trial coupling constant. The result after FT of this signal with $J = J^*/2$ is shown in Figure 3b. The J doubling procedure can be applied several times as described by:

$$S(t) = \sin(\pi J_{C,H}t) \exp(-\lambda t) \prod_{n=0}^k \cos(2^n \pi J^*t). \quad (3)$$

Three stages of J doubling with $J = J^*/2$ separates the peaks, but the correct peak pattern is absent after FT (Figure 3c). It should be noted that when $J = J^*/2$, the intermediate contributions cancel (compare spectra below) (19).

When $J^* = J$, FT shows the appropriate peak pattern (Figure 3d), with a separation of $|J + XJ^*|$, where X is dependent on the number of stages of J doubling.

Proton multiplicity renders the peak pattern more complex. The NMR signal in the time domain is then described by:

$$S(t) = \cos(\pi J_{H,H}t) \sin(\pi J_{C,H}t) \exp(-\lambda t) \quad (4)$$

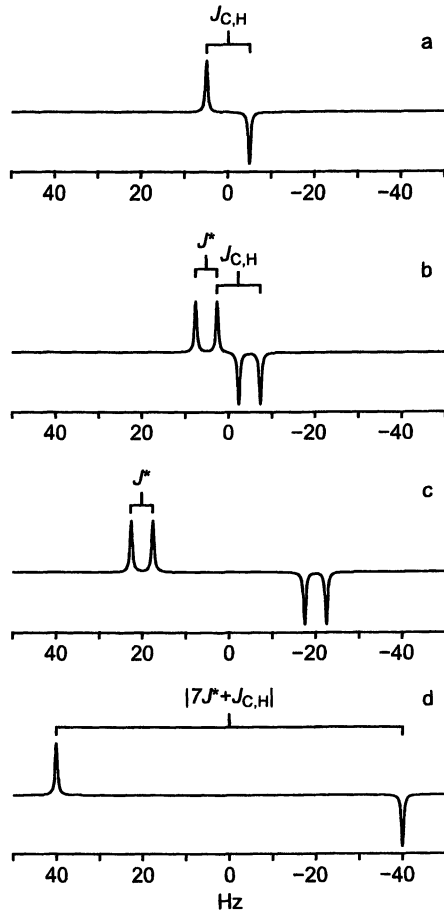


Figure 3. Simulated NMR spectra: (a) Original anti-phase doublet with $J_{C,H} = 10$ Hz; Spectrum after J doubling: (b) one stage with $J^* = J/2$; (c) three stages, with $J^* = J/2$; (d) three stages with $J^* = J$.

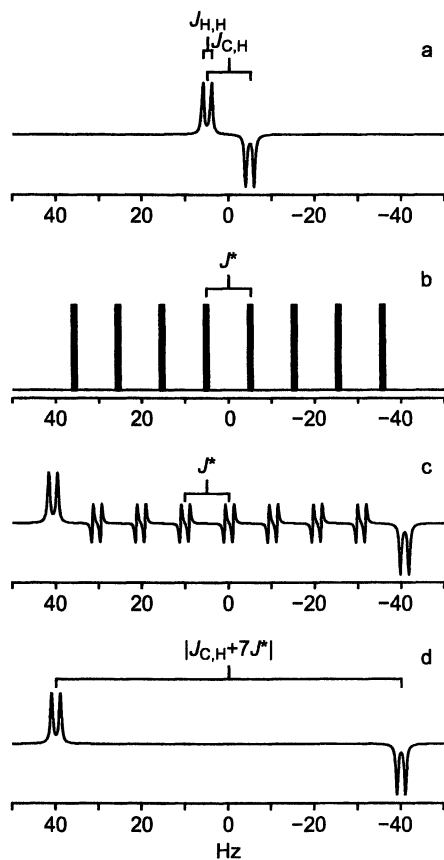


Figure 4. Simulated NMR spectra: (a) Original in-phase doublet with $J_{H,H} = 3$ Hz and anti-phase doublet with $J_{C,H} = 10$ Hz; (b) A set of eight delta functions used as an equivalent to J doubling in three stages; Spectrum after convolution: (c) with $J^ = 1.02J$; (d) with $J^* = J$.*

After FT one obtains a more complex multiplet (Figure 4a). The J doubling procedure may also be performed in the frequency domain based on the convolution theorem (22):

$$h(\nu') = \int_{-\infty}^{\infty} f(\nu)g(\nu' - \nu) d\nu. \quad (5)$$

Convolution of the original multiplet with a set of delta functions (Figure 4b) leads to an equivalent result, i.e., a spectrum in which the anti-phase splitting is increased depending on the number of J doublings applied and the magnitude of J^* as shown in Figure 4c, where $J^* = 1.02J$. The inner anti-phase multiplets have a splitting of $|{}^nJ_{C,H} - J^*|$. When $J^* = J$, the correct peak pattern, with a separation of $|J + XJ^*|$ is present and all contributions in the ordinate axis cancel (Figure 4d). Consequently, we have determined the unknown hetero-nuclear ${}^nJ_{C,H}$ that was our goal.

We will now demonstrate how this technique can be used to obtain transglycosidic ${}^3J_{C,H}$ for an oligosaccharide using three approaches for its estimate, namely, (i) J doubling in the presence of the intact homo-nuclear multiplet, (ii) selective deuteration as a means to reduce proton multiplicity, and (iii) band-selective decoupling during the acquisition period which leads to elimination of in-phase proton couplings and simplification due to reduced multiplicity. As an example we have chosen α -cyclodextrin (α -CD) (23), a cyclic α -(1 \rightarrow 4)-linked glucose hexasaccharide (Figure 5) with the corresponding ${}^1\text{H}$ NMR spectrum (Figure 6).

In the first approach the multiplet structure of the anomeric proton due to its three-bond heteronuclear coupling to C4 and its intra-ring three-bond homo-nuclear coupling to H2 is shown in Figure 7a. The partial cancellation present in the peak indicates that the magnitude of the ${}^3J_{H1,H2}$ and ${}^3J_{C4,H1}$ couplings are similar. Application of the J doubling procedure facilitates determination of the coupling constant related to the ϕ torsion angle. Thus, $J_\phi = 5.17$ Hz (Figure 7b). The multiplet of the H4 resonance is shown in Figure 7d. It consists of ${}^3J_{C1,H4}$, ${}^3J_{H3,H4}$, and ${}^3J_{H4,H5}$ in which the homo-nuclear couplings are of almost equal magnitude (~ 9 Hz). The J doubling procedure gives $J_\psi = 5.04$ Hz. In the procedure used the unknown coupling is usually also investigated by finding a minimum in the summation of all ordinates in the spectrum (Figure 8). However, for the latter coupling the minimum is not well defined and additional experiments are needed to obtain reliable data (*vide infra*).

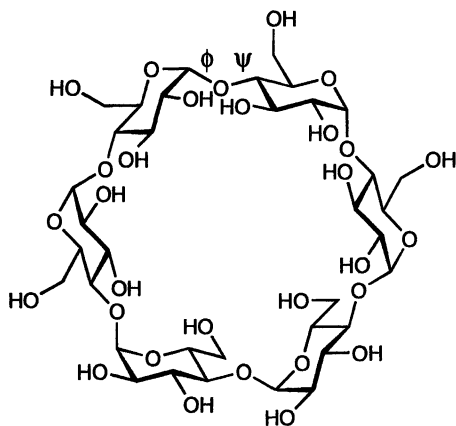


Figure 5. Schematic of α -cyclodextrin (α -CD), in which the torsion angles at glycosidic linkages are denoted by ϕ and ψ .

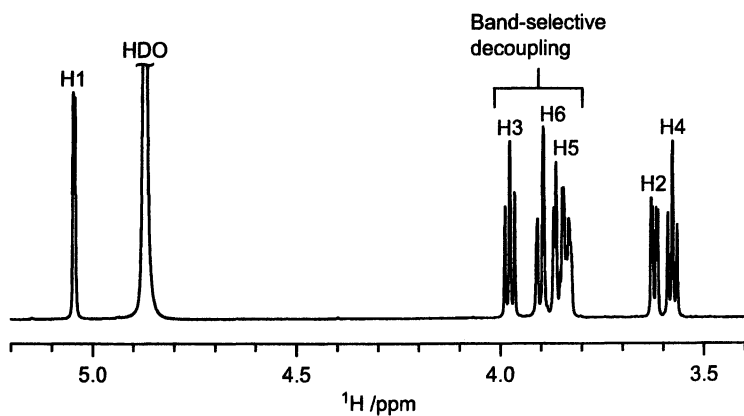


Figure 6. ^1H NMR spectrum of α -CD in D_2O at 15°C . The spectral region is indicated where band-selective decoupling was performed during the acquisition period (cf. Figures 7 and 8).

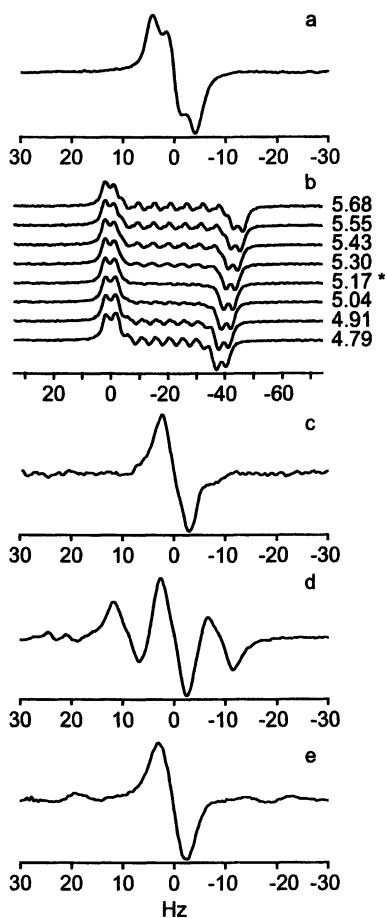


Figure 7. Application of the pulse sequence shown in Figure 2; (a) H1 resonance in α -CD; (b) Application of the J doubling procedure in search of J_ϕ , using different trial values of J^ ; (c) H1 resonance in α -CD- d_{18} ; (d) H4 resonance in α -CD; (e) H4 resonance in α -CD after band-selective decoupling of the H3, H5 and H6 resonances.*

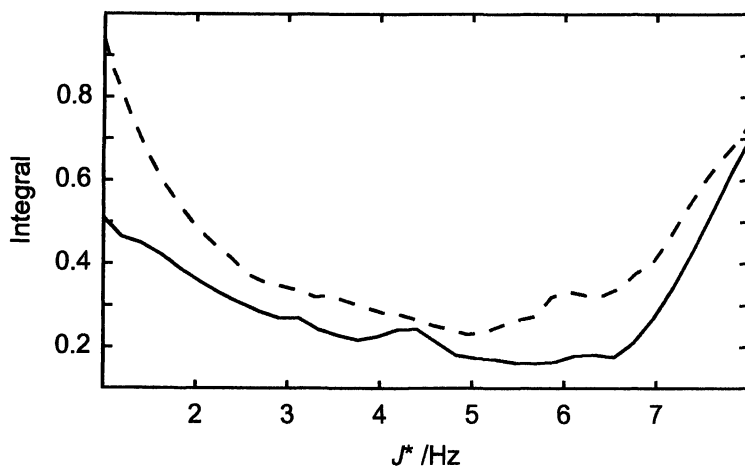


Figure 8. Output from the J doubling procedure (Integral vs. J^*) for the H4 resonance without band-selective decoupling (solid line), which gives a broad ill-defined minimum, and with band-selective decoupling (dashed line), which facilitates determination of ${}^3J_{C4,H1}$.

In the second approach we chemically modify the α -CD molecule by exchanging hydrogen atoms on hydroxyl bearing carbons with deuterium by reflux in D_2O in the presence of a Raney-Nickel catalyst (24). In the resulting molecule the H2, H3 and H6 protons have been exchanged for deuterons. Part of a schematic structure is shown in Figure 9, in which only the remaining protons are present. Measurement of J_ϕ at the H1 resonance is now significantly simplified since it is only an anti-phase doublet (Figure 7c). In a similar manner J_ψ could be determined at the H4 resonance, which only has a significant ${}^3J_{H4,H5}$ coupling left.

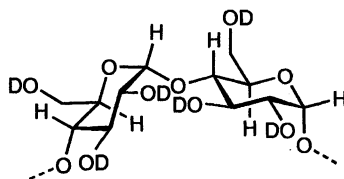


Figure 9. Schematic of a part of α -cyclodextrin in which hydrogen atoms on hydroxyl bearing carbons have been exchanged with deuterium (α -CD- d_{18}). Only the remaining H1, H4 and H5 atoms are depicted.

The third approach is to band-selectively decouple (25) some of the protons during the acquisition period (Figures 2 and 6). Band-selective decoupling of the H3, H5 and H6 resonances eliminates the homo-nuclear couplings to H4 (Figure 7e). Subsequent J doubling gives J_ψ . Importantly, the integral minimum is now well defined (Figure 8). In all the different measurements we were able to determine the ${}^3J_{C,H}$ couplings to a high precision, on the order of ± 0.1 Hz. The technique is furthermore believed to be very accurate (19,22). This is essential in using J couplings in Karplus-type relationships (26) for conformational and dynamics studies of oligosaccharides. Summarizing these measurements we report $J_\phi = 5.2$ Hz and $J_\psi = 5.0$ Hz. Our result for J_ϕ is, however, at variance with that reported by Forgo and D'Souza where $J_\phi = 6.4$ Hz (27). The conformationally dependent trans-glycosidic coupling constants of α -CD will be used in future studies of the molecule.

Residual Dipolar Couplings

The through-space magnetic dipole-dipole couplings have proven to be of great importance for investigation of molecular conformations (28). These interactions depend on the spin-spin distances and on the orientations of the inter-nuclear vectors with respect to the external magnetic field. This means that the residual dipolar coupling (RDC) is a valuable probe of long-range order and molecular structure. For oligosaccharides, the technique promises great advantages because valuable restraints can be added to the otherwise scarce experimental information. The general motivation for these investigations is to describe the conformational preferences at the glycosidic linkage.

To extract useful information about conformation from the experimental RDCs in a flexible molecule, we need a model that takes into account both molecular tumbling and internal degrees of freedom. Several approaches have been considered for the interpretation of RDCs (29). The simplest possible model assumes that the molecule can be described by an average conformation and thus a single ordering matrix. The molecular frames, required for the analyses, are constructed from the moment of inertia tensor (30,31) or the gyration tensor (32). On increasing the sophistication of the model we have a situation where only a small set of minimum energy structures is populated (the rotational isomeric state or RIS approximation) (33). Finally, the analysis may involve more realistic models, which allow for continuous bond rotations. Two such protocols that have been frequently used for interpretations of RDCs in bulk liquid crystals are the additive potential (AP) model (34) and the maximum entropy (ME) approach (35). Recently, we proposed a new approach for the

analysis of NMR parameters in general and dipolar couplings in particular (36,37). This procedure, valid in the low-order limit, is a combination of the AP and ME methods and was denoted APME.

The general expression for the carbon-proton RDCs is given by:

$$D_{CH} = -\frac{\mu_0 \gamma_C \gamma_H \hbar}{8\pi^2} \frac{1}{2} \left\langle (3 \cos^2 \theta - 1) r_{CH}^{-3} \right\rangle \quad (6)$$

where r_{CH} the spin-spin distance and θ is the angle between the spin-spin vector and the magnetic field. As orienting media we have hitherto used DMPC:DHPC (38), DMPC:DHPC:CTAB (39), CPCl/*n*-hexanol/brine (40), C8E5/*n*-octanol (41), and most recently the mineral lyotropic phase V_2O_5 (42). The alignment of the carbohydrate molecules in these dilute liquid media is small, on the order of 10^{-3} , and the presence of an ordered phase is usually ascertained by the quadrupolar splitting of the D_2O resonance in the sample (Figure 10). Consequently, only a fraction of the molecules contribute to the anisotropic alignment. However, high-resolution NMR spectra contain in such case valuable information on molecular structure and ordering.

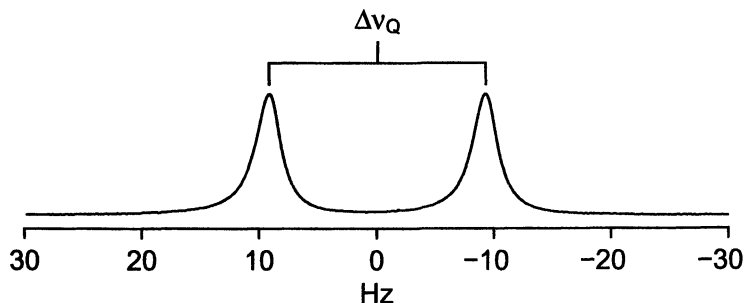


Figure 10. Deuterium NMR spectrum at 20 °C and 14.1 T in a V_2O_5 preparation showing the quadrupolar splitting of the D_2O resonance.

The expression in Equation 6 is general and valid for dipolar interactions in solid and liquid crystalline phases. It is, however, practically useless for analyses of RDCs in partially ordered systems. The molecular orientation with respect to the magnetic field, defined by the angle θ , is in general not known. We need therefore to rewrite Equation 6 by using three distinct transformations, according to:

$$D_{CH} = -\frac{\mu_0}{16\pi^2} \frac{\gamma_C \gamma_H h}{r_{CH}^3} \left[3 \cos^2 \eta - 1 \right] \quad (7)$$

$$\times \left[S_{zz} \left(3 \cos^2 \theta_{CH}^z - 1 \right) + \left(S_{xx} - S_{yy} \right) \left(\cos^2 \theta_{CH}^x - \cos^2 \theta_{CH}^y \right) \right]$$

where θ_{CH}^α ($\alpha = x, y, z$) are the angles between the spin-spin vector and the molecular coordinate frame. The order parameters, $S_{\alpha\alpha}$, describe the transformation from the molecular axis system to the phase director. Finally, the angle η defines the orientation of the director with respect to the magnetic field.

In order to specify the angles θ_{CH}^α for every RDC in the molecule we need to define the molecular coordinate frame, that coincides with the ordering tensor, characterized by the order parameters $S_{\alpha\alpha}$.

At this point we introduce the molecular dynamics (MD) computer simulation, employing a recently developed force field for carbohydrates (43). The purpose of carrying out the MD simulations is twofold: (i) to aid the interpretation of the RDCs, and (ii) to perform conformational analysis of oligosaccharides. The power of the MD technique will be demonstrated using results obtained in a computer simulation of the human milk oligosaccharide Lacto-*N*-neotetraose (LNnT) (41,44), Figure 11.

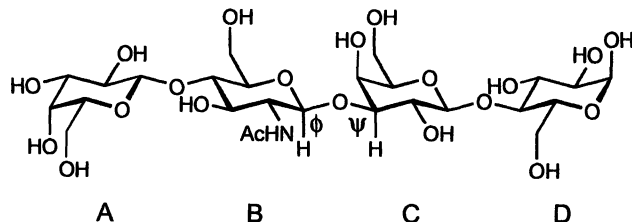


Figure 11. Schematic of the tetrasaccharide Lacto-*N*-neotetraose β -D-Galp-(1 \rightarrow 4)- β -D-GlcpNAc-(1 \rightarrow 3)- β -D-Galp-(1 \rightarrow 4)- α -D-Glcp (LNnT) where the torsion angles at glycosidic linkages are denoted by ϕ and ψ followed by a subscript related to the sugar residues labeled A through D.

The three glycosidic linkages, i.e., six torsion angles, constitute the important degrees of freedom in the molecule. To obtain an idea of the conformational dynamics present in the oligosaccharide, an MD simulation with explicit water molecules as solvent was performed. The results as scatter plots are shown in Figure 12. At the β -(1 \rightarrow 4)-linkages (a and c in Figure 12) one may regard the conformation to be described by dynamics in a single well on the time scale of

the simulation (4 ns). However, for the β -(1 \rightarrow 3)-linkage two dynamically interconverting states are populated (Figure 12b). This observation shows that at least two conformational states should be considered in description of the molecule.

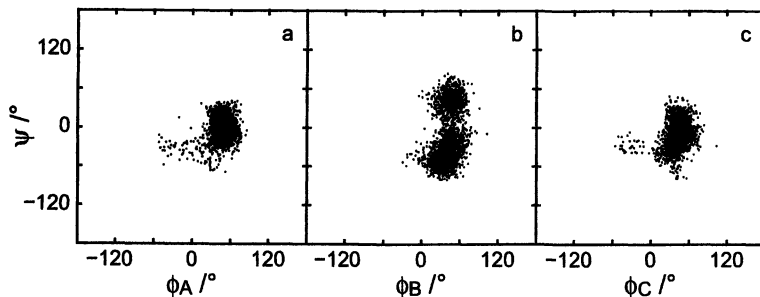


Figure 12. Scatter plots from the MD simulation of LNnT. (a) β -(1 \rightarrow 4)-linkage at the terminal end; (b) β -(1 \rightarrow 3)-linkage; (c) β -(1 \rightarrow 4)-linkage at the reducing end.

Like for $^3J_{C,H}$ there are many ways to measure and extract RDCs (40,41,45). The hetero-nuclear one-bond carbon-proton RDCs of LNnT in a 5% aqueous C8E5/*n*-octanol liquid crystalline medium (40) are shown in Figure 13, where A1 represents $^1D_{C,H}$ of the anomeric carbon-proton interaction in the terminal galactosyl residue A, etc. It is directly evident that there are differences in magnitude and particularly in the sign of different $^1D_{C,H}$ in the four sugar residues.

In the interpretation we assume that the moment of inertia frame, calculated from the trajectory defines the ordering tensor. This is the simplest approach of those mentioned above. It is used herein due to the limited number of RDCs available so far in the analysis and the fact that these RDCs are not explicitly conformation dependent. Two procedures were considered for determination of the ordering tensor necessary for the analyses of the RDCs. The order parameters, $S_{\alpha\alpha}$, were calculated from the eigenvalues of: (i) the inertia tensor, $I_{\alpha\alpha}$, and (ii) the gyration tensor $\rho_{\alpha\alpha}$. Note that both tensors share the common principal frames, while the order parameters are slightly different. The former method was proposed for thermotropic liquid crystals (30) and recently applied for weakly ordered phases (31).

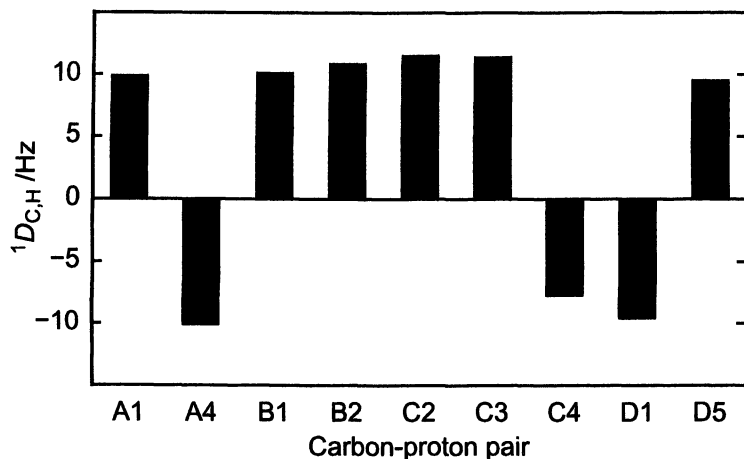


Figure 13. Residual dipolar couplings, $^1D_{C,H}$, measured for LNT.

In this approach the order parameters are related to the molecular shape through the ellipsoid semiaxes, A_α , according to

$$\begin{aligned}
 S_{zz} &= 1 - (A_x + A_y) / 2A_z \\
 S_{yy} &= -1/2 + A_y / 2A_z \\
 S_{xx} &= -1/2 + A_x / 2A_z
 \end{aligned} \tag{8}$$

with

$$A_\alpha = \left[(I_{\beta\beta} + I_{\gamma\gamma} - I_{\alpha\alpha}) 5 / 2m \right]^{1/2} \tag{9}$$

where m is the molecular mass. The conformational dependence of the (hypothetic) order parameters, $S_{\alpha\alpha}$, is shown in Figure 14a. The moment of inertia tensor used for the calculation of $S_{\alpha\alpha}$ was derived from the trajectory generated in the MD simulation. In fact, the moment of inertia tensor and therefore the order parameters, in particular S_{zz} , reflect the transitions observed for the ψ_B torsion angle (not shown).

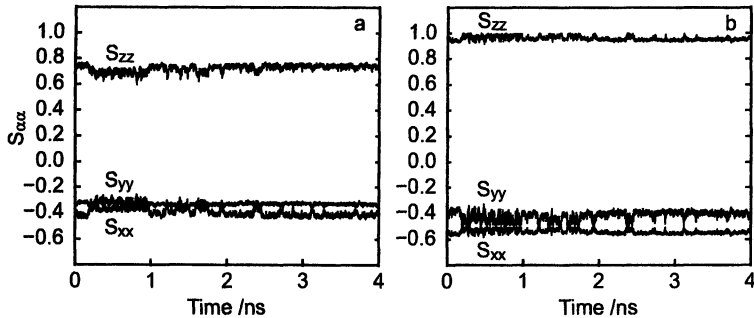


Figure 14. The time dependence of the (hypothetic) order parameters, $S_{\alpha\alpha}$, based on the inertia tensor (a), or the gyration tensor (b).

The second method for calculation of the ordering tensor is similar and based on the description of the molecular shape using the atomic distribution (32) where the order parameters are calculated using the eigenvalues of the gyration tensor

$$\begin{aligned}
 S_{zz} &= 1 - \frac{1}{2} \delta \\
 S_{yy} &= -\frac{1}{2} - \frac{1}{2} \delta \\
 S_{xx} &= \delta - \frac{1}{2}
 \end{aligned} \tag{10}$$

with the dimensionless scalar δ given by

$$\delta = \frac{(\rho_x - \rho_y)}{(\rho_z - \rho_y)} \tag{11}$$

where ρ_x , ρ_y , and ρ_z are the square roots of the eigenvalues of the gyration tensor. The conformational dependence of these order parameters, $S_{\alpha\alpha}$, is shown in Figure 14b. Note that a single, average ordering matrix is assumed in these procedures, which is indeed an approximation. We have, however, investigated the validity of this assumption by analyzing the RDCs with the ordering matrixes calculated for every step in the trajectory, i.e., for every rigid conformation. The results were essentially identical compared to the analysis performed using the average order parameters.

For the description of LNnT the analyses were carried out on the entire MD trajectory and on two fractions, each of 0.5 ns, corresponding to the two ψ_B states: ψ_B^+ and ψ_B^- (cf. Figure 12b). The correlation is significantly better for the ψ_B^+ state. However, from the $^1\text{H}, ^1\text{H}$ T-ROESY build-up curves the experimentally determined proton-proton distance, from H1B to H4C, cannot be explained using the ψ_B^+ state only; a mixture of the two states ψ_B^+ and ψ_B^- is required. The calculation of order parameters for LNnT correspond to ellipsoid models and these are shown in Figure 15, for the two conformational states populated in the MD simulation.

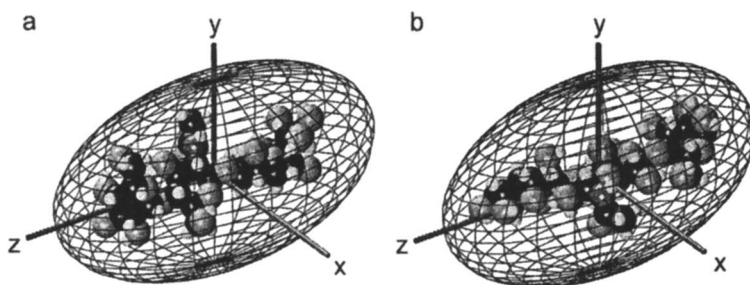


Figure 15. The ellipsoid model for calculation of the order parameters showing a lower biaxiality for ψ_B^+ (a) than for ψ_B^- (b).

Concluding Remarks

The use of experimental NMR techniques in combination with molecular simulations offers a very powerful approach to the study of conformational dynamics of molecules such as oligosaccharides. The recently developed force fields for carbohydrate molecular dynamics simulations have reached a level of accuracy where the overall description is good, but subtle differences in conformational equilibria still need to be improved. The continuous development of improved NMR pulse sequences will facilitate both faster and more accurate determination of parameters that can be measured experimentally and used in the interpretation related to conformation and dynamics of these oligosaccharides. In the near future we foresee the use of NMR active stable isotopes as an important research tool for oligosaccharides and their interactions with other molecules, in particular lectins.

Acknowledgment

This work was supported by grants from the Swedish Research Council (VR), the Carl Trygger Foundation and the Swedish International Development Cooperation Agency (SIDA).

References

1. Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364-2369.
2. Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. *Science* **2001**, *291*, 2370-2376.
3. Freeze, H. H. *Glycobiology* **2001**, *11*, 129R-143R.
4. Sharon, N.; Lis, H. *Glycobiology* **2001**, *14*, 53R-62R.
5. Bubb, W. A. *Concepts Magn. Reson.* **2003**, *19A*, 1-19.
6. Bagno, A.; Rastrelli, F.; Scorrano, G. *J. Magn. Reson.* **2004**, *167*, 31-35.
7. Ravindranathan, S.; Mallet, J.-M.; Sinay, P.; Bodenhausen, G. *J. Magn. Reson.* **2003**, *163*, 199-207.
8. Woods, R. J. *Rev. Comput. Chem.* **1996**, *9*, 129-165.
9. Rosen, J.; Robobi, A.; Nyholm, P.-G. *Carbohydr. Res.* **2004**, *339*, 961-966.
10. Yu, H.; Amann, M.; Hansson, T.; Köhler, J.; Wich, G.; van Gunsteren, W. F. *Carbohydr. Res.* **2004**, *339*, 1697-1709.
11. Kony, D.; Damm, W.; Stoll, S.; Hünenberger, P. H. *J. Phys. Chem. B* **2004**, *108*, 5815-5826.
12. Engelsens, S. B.; Monterio, C.; Hervé de Penhoat, C.; Pérez, S. *Biophys. Chem.* **2001**, *93*, 103-127.
13. Morat, C.; Taravel, F. R.; Vignon, M. R. *Magn. Reson. Chem.* **1994**, *26*, 264-270.
14. Krishnamurthy, V. V. *J. Magn. Reson. Ser. A* **1996**, *121*, 33-41.
15. Uhrin, D.; Batta, G.; Hruby, V. J.; Barlow, P. N.; Kövér, K. E. *J. Magn. Reson.* **1998**, *130*, 155-161.
16. Nishida, T.; Widmalm, G.; Sándor, P. *Magn. Reson. Chem.* **1995**, *33*, 596-599.
17. Nishida, T.; Widmalm, G.; Sándor, P. *Magn. Reson. Chem.* **1996**, *34*, 377-382.
18. Blechta, V.; del Río-Portilla, F.; Freeman, R. *Magn. Reson. Chem.* **1994**, *32*, 134-137.
19. McIntyre, L.; Freeman, R. *J. Magn. Reson.* **1992**, *96*, 425-431.
20. Freeman, R.; McIntyre, L. *Israel J. Chem.* **1992**, *32*, 231-244.

21. del Río-Portilla, F.; Blechta, V.; Freeman, R. *J. Magn. Reson. Ser. A* **1994**, *111*, 132-135.
22. Garza-García, A.; Ponzanelli-Velázquez, G.; del Río-Portilla, F. *J. Magn. Reson.* **2001**, *148*, 214-219.
23. Naidoo, K. J.; Yu-Jen Chen, J.; Jansson, J. L. M.; Widmalm, G.; Maliniak, A. *J. Phys. Chem. B* **2004**, *108*, 4236-4238.
24. Hällgren, C.; Widmalm, G. *J. Carbohydr. Chem.* **1993**, *12*, 309-333.
25. Kupče, Ě.; Freeman, R. *J. Magn. Reson. Ser. A* **1993**, *102*, 364-369.
26. Cloran, F.; Carmichael, I.; Serianni, A. S. *J. Am. Chem. Soc.* **1999**, *121*, 9843-9851.
27. Forgo, P.; D'Souza, V. T. *Magn. Reson. Chem.* **1999**, *37*, 48-52.
28. Emsley, J. W.; Lindon, J. C. *NMR Spectroscopy Using Liquid Crystal Solvents*; Pergamon Press: Oxford, 1975.
29. Emsley, J. W. *Liquid Crystalline Samples: Structure of Nonrigid Molecules*; In *Encyclopedia of NMR*; Grant, D. M.; Harris, D. M., Eds.; Wiley: New York, 1996; pp 2781-2787.
30. Samulski, E. T.; Dong, R. Y. *J. Chem. Phys.* **1982**, *77*, 5090-5096.
31. Azurmendi, H. F.; Bush, C. A. *J. Am. Chem. Soc.* **2002**, *124*, 2426-2427.
32. Almond, A.; Axelsen, J. B. *J. Am. Chem. Soc.* **2002**, *124*, 9986-9987.
33. Sinton, S. W.; Zax, D. B.; Murdoch, J. B.; Pines, A. *Mol. Phys.* **1984**, *53*, 333-362.
34. Emsley, J. W.; Luckhurst, G. R.; Stockley, C. P. *Proc. R. Soc. Lond. A* **1982**, *381*, 117-138.
35. Catalano, D.; Di Bari, L.; Veracini, C. A.; Shilstone, G. N.; Zannoni, C. *J. Chem. Phys.* **1991**, *94*, 3928-3935.
36. Stevansson, B.; Landersjö, C.; Widmalm, G.; Maliniak, A. *J. Am. Chem. Soc.* **2002**, *124*, 5946-5947.
37. Stevansson, B.; Sandström, D.; Maliniak, A. *J. Chem. Phys.* **2003**, *119*, 2738-2746.
38. Rundlöf, T.; Landersjö, C.; Lycknert, K.; Maliniak, A.; Widmalm, G. *Magn. Reson. Chem.* **1998**, *36*, 773-776.
39. Staaf, M.; Höög, C.; Stevansson, B.; Maliniak, A.; Widmalm, G. *Biochemistry* **2001**, *40*, 3623-3628.
40. Landersjö, C.; Höög, C.; Maliniak, A.; Widmalm, G. *J. Phys. Chem. B* **2000**, *104*, 5618-5624.
41. Landersjö, C.; Jansson, J. L. M.; Maliniak, A.; Widmalm, G. manuscript submitted
42. Desvaux, H.; Gabriel, J.-C. P.; Berthault, P.; Camerel, F. *Angew. Chem. Int. Ed.* **2001**, *40*, 373-376.

43. Eklund, R.; Widmalm, G. *Carbohydr. Res.* **2003**, *338*, 393-398.
44. Newburg, D. S.; Ruiz-Palacios, G. M.; Altaye, M.; Chaturvedi, P.; Meinzen-Derr, J.; de Lourdes Guerrero, M.; Morrow, A. L. *Glycobiology* **2004**, *14*, 253-263.
45. Tian, F.; Al-Hashimi, H. M.; Craighead, J. L.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 485-492.

Chapter 3

Structure and Dynamics of Carbohydrates Using Residual Dipolar Couplings

James H. Prestegard and Xiaobing Yi

Complex Carbohydrate Research Center, University of Georgia,
Athens, GA 30602

A method to analyze the dynamics of carbohydrates by combining residual dipolar couplings and molecular dynamics simulations is presented. Previously reported data acquired in bicelle and phage alignment media for methyl-3,6-di-*O*-(α -D-mannopyranosyl)- α -D-mannopyranoside is used to illustrate this procedure. Molecular dynamic simulations of this model compound suggest that the dynamics of the two types of glycosidic linkages, 1–3 and 1–6, are quite different. The 1-3 linkage is fairly rigid whereas the 1–6 linkage is very flexible sampling three major conformers at (ϕ , φ , ω): **S1**(80.1, 176.3, 52.8°), **S2**(77.2, 174.1, -159.9°) and **S3**(74.2, 172.0, -84.4°). The correct populations of the three conforms cannot be obtained from molecular dynamics because the transition time scale is very long. Analysis of the populations for the three states were instead performed using RDC data. Our studies reveal that **S1** (*gauche-gauche*, *gg*) and **S2** (*gauche-trans*, *gt*) are nearly equally populated whereas the **S3** (*trans-gauche*, *tg*) state has a low population (< 20%).

Introduction

The structural analysis of cell-surface carbohydrates is an important step in understanding a variety of biologically important protein-carbohydrate interactions (1). Lectins interacting with specific oligosaccharides influence proliferation of both normal and malignant cells (2,3), glycosyltransferases and glycosidases interacting with donor and acceptor molecules govern the synthesis of cell-specific carbohydrates during development (4), and toxins and adhesion proteins from pathogens interacting with carbohydrates often control infectivity (5). The design of new molecules to compete with carbohydrates in controlling these processes benefits from knowledge of the detailed three-dimensional structures of the native carbohydrates. However, structure analysis of carbohydrates is complicated by the fact that at least in solution there is substantial mobility about some of the glycosidic bonds that connect sugar residues. Structure must therefore be described as a distribution of thermally accessible conformers sampled by glycosidic torsional motions. Any one of these conformers can be important in protein interaction events. It is our goal here to present some new NMR based methodology for determination of these conformational distributions.

Motion in carbohydrates has been considered previously using a number of approaches. The presence of motion can be qualitatively detected though inconsistencies in NOE data used to determine solution structures of carbohydrates (6). Here, the very steep dependence of the NOE ($1/r^6$) will sometimes lead to significant transfers of magnetization from a given proton to two protons that are widely separated in space. Because of bonding constraints, such NOE data cannot be satisfied by a single conformer. The alternative is that two states are populated, each with close approach to one proton. Motion can also be assessed from spin relaxation rates (7,8). Here, internal motion can dramatically decrease spin relaxation rates for nuclei in some parts of a carbohydrate indicating the presence of internal motion. But, both the timescales and the amplitudes of motion can contribute to the decrease (9), making it very difficult to separate contributions from differences in structures sampled and the timescales at which they are sampled. Here we consider a particularly useful set of observables, residual dipolar couplings (RDCs) that can greatly facilitate our ability to separately define structures and motions. We also use molecular dynamics simulations to improve our physical description of states sampled.

RDC measurement and analysis

Residual dipolar couplings come from the same through-space dipole-dipole interaction that gives rise to NOEs and spin relaxation (10). For a pair of weakly

coupled spin one-half nuclei this interaction can be represented as in Equation 1.

$$D_{ij} = -\frac{h\mu_0\gamma_i\gamma_j}{(2\pi r)^3} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle \quad (1)$$

Here r is the distance between nuclei i and j , γ_i, γ_j are the magnetogyric ratios for the nuclei, μ_0 is the permeability of space, h is Planck's constant, and θ is the angle between the inter-nuclear vector and the direction of magnetic field. For structural applications, NOEs and other spin relaxation measurements capitalize on the internuclear distance dependence; the angular dependence averages to zero as molecules tumble in solution and only provides the time dependent fluctuations needed to induce spin transitions. RDCs typically capitalize on the angular dependence using pairs of nuclei having their internuclear distance fixed by bonding geometry; the angular dependent term is prevented from averaging to zero by partially ordering molecules in liquid crystal media (11).

The methods of measuring RDCs also differ from those used with NOEs and other relaxation parameters. RDCs are often measured directly as splittings of resonances in NMR spectra as opposed to the time dependence of resonance intensities. In fact, Equation 1 has been written in a form that gives the RDC contribution to splitting directly in Hz when all constants are given in SI units. The spin operator part of the interaction (not shown) is identical to that for first order scalar coupling. So, RDCs add to scalar couplings when they are present, and splittings must be measured under aligned and isotropic conditions to extract the RDCs from the difference in splittings.

Some of the RDCs that can be measured for a simple carbohydrate are illustrated in Figure 1. One bond $^{13}\text{C}-^1\text{H}$ RDCs are measured from splittings in coupled HSQC spectra (12). Usually coupling in the indirect (^{13}C) dimension as opposed to the direct (^1H) dimension is used because natural line widths are smaller and there are fewer distortions due to cross-correlation effects. Multiple bond $^1\text{H}-^1\text{H}$ couplings can be measured by analysis of cross-peak intensities in constant time $^1\text{H}-^1\text{H}$ COSY experiments (12) or simply by fitting couplings in regular COSY experiments (13). The distances over which these couplings can be measured is limited to about 5Å by the $1/r^3$ dependence in Equation 1. For the $^1\text{H}-^1\text{H}$ couplings illustrated in Figure 1 we also assume that fluctuations in pyranose ring geometry are sufficiently small to allow r to be extracted from low energy representations of this ring structure. Experiments have also been devised for the measurement of longer range $^{13}\text{C}-^1\text{H}$ RDCs (14-16).

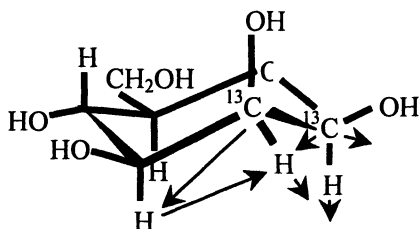


Figure 1. Illustration of some measurable RDCs in a pyranoside sugar residue

Analysis of measured RDCs in terms of structure can take a number of forms (10,17). However, we prefer to use an order matrix analysis (18). Equation 1 can be rewritten in terms of elements of a 3x3 order matrix, S_{kl} (Equation 2) (19). Here, $D_{max\ ij}$ is the coupling that would result for a pair of nuclei at a 1.0 Å separation with their internuclear vector along the magnetic field, and $\cos(\phi_{k,l})$ are the direction cosines relating an inter-nuclear vector of interest to the axes (x,y,z) of an arbitrarily chosen molecular fragment frame.

$$D_{ij} = \frac{D_{max\ ij}}{r^3} \sum_{k,l} S_{kl} \cos(\phi_k) \cos(\phi_l) \quad (2)$$

Because the order matrix is traceless and symmetric, only five order matrix elements are independent in the expression for each RDC. For a pyranose ring of assumed geometry the $\cos(\phi_{k,l})$ are known, making a set of equations of the form of Equation 2 solvable for any five or more independent RDC measurements. Singular value decomposition can be used to give a best least squares solution for the order parameters under these circumstances (18). The resulting matrix can be diagonalized to yield order parameters in a principal order (or alignment) frame and a set of Euler angles that can be used to transform coordinates in the initial molecular frame to the principal order frame. Structures for oligosaccharides can be derived by realizing that elements of a rigid structure must share the same order frame (except for a four-fold degeneracy that can be removed by collecting data in two or more different alignment media). Individual sugar rings are therefore each rotated into their principal order frame and rings are translated to make glycosidic bonds. This approach has recently been facilitated by the development of a program named REDCAT (RESidual Dipolar Coupling Analysis Tool) (20). We use this program in the example that follows.

When internal motion exists order parameters are affected directly through selective reductions in their values. Order parameters are defined as in Equation

3 where θ_i are the angles between the magnetic field and the molecular frame axes and the bar denotes averaging over molecular reorientation.

$$S_{ij} = \frac{\overline{3\cos\theta_i\cos\theta_j} - \delta_{ij}}{2} \quad (3)$$

When expressed in the principal frame only diagonal elements are finite and the S_{zz} value becomes formally the equivalent of the order parameter frequently used in analysis of spin relaxation data ($S_{zz} = -2S_{yy} = -2S_{xx} = \frac{1}{2}(3\cos^2\theta_{zz} - 1)$) (21). The primary differences are that we do not assume axial symmetry ($S_{xx} \neq S_{yy}$), the averaging is over all motions on time scales shorter than the reciprocal of the RDC as opposed to all motions on time scales shorter than the molecular tumbling correlation time. Our order parameters also include the effects of both overall and internal motion. In either case, if the system were rigid and z was along the magnetic field, S_{zz} would be 1, and if motional amplitude approached completely isotropic motion S_{zz} would be zero.

Separating the effects of overall and internal motion on measured RDC order parameters is formally possible in the limit where overall order is dominated by the properties of one residue. This is illustrated in Figure 2 where the order parameters measured for ring I (S_{ij}) are significantly larger than those from ring II (S'_{ij}). We can formally describe the order tensor for ring II in terms of that for ring I by transforming the matrix from one coordinate system to the other using a rotation matrix, R , composed of direction cosines relating principal axes for ring I to principal axes for ring II, shown in Equation 4.

$$[S'_{ij}] = [R_{ij}] \times [S_{ij}] \times [R_{ij}^{-1}] \quad (4)$$

The resulting expressions for elements, S'_{ij} , are then clearly related to S_{ij} by products of direction cosines that are averaged by internal motion. When collected for each axis of S'_{ij} , these averages are formally equivalent to those in the expression for order parameters in Equation 3 above, except they now pertain to internal motion as opposed to overall motion. A simple example is one in which the two order matrices share an average z axis and internal motion results in equal amplitude oscillations about x and y axes. It is then easily shown that S'_{zz} equals the average of $(3\cos^2\theta - 1)/2$ times S_{zz} where θ is the instantaneous angle between the z axis in the two frames, and the average of $(3\cos^2\theta - 1)/2$ is the order parameter for internal motion. We have recently produced a system that approximates this situation by attaching a short alkyl chain to the reducing end of a disaccharide and allowing this chain to anchor that ring to bicelle particles of an ordered liquid crystal matrix (22). Differences in order parameters could be interpreted in terms of average angles of oscillation.

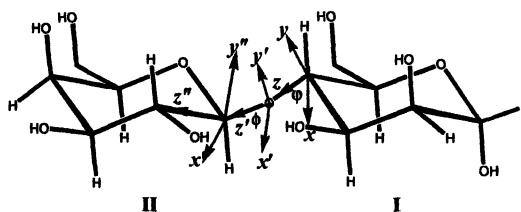


Figure 2. Separation of the contributions of overall tumbling and internal motions by transforming the Saupe order tensors by rotations about glycosidic bonds.

This type of interpretation has been possible with small amplitude oscillations about 1-3 and 1-4 linkages in di- and trisaccharides (23,24), but for larger oscillations, where axes of average order frames will not in general superimpose an alternate approach may be more appropriate.

One approach is to specifically allow sampling of a discrete set of conformers. These conformers can be selected from those represented in molecular dynamics (MD) trajectories. Significant advances have been made in optimization of force fields for oligosaccharides and proteins (25,26). The conformers that are frequently sampled, particularly when run as systems completely solvated with water, are very likely to be good representations of low energy states. It is much more difficult to evaluate populations of various states from these trajectories. This is first because transitions between significantly different conformers can occur at time scales too long to allow MD simulations to adequately sample conformations, and second because the energy differences between conformers, as derived from the force field, are unlikely to have the 0.4 kcal precision necessary to properly predict even a 2:1 population distribution. However, combining models for low energy states from MD with experimental data, such as RDCs, can test the accuracy of low energy state models, and give reasonable population estimates.

A tool for testing models based on sampling of multiple conformational states has recently been implemented in REDCAT (20). This works by entering multiple sets of coordinates used to define the $\cos\phi$ angles in Equation 2, averaging the products of these $\cos\phi$ terms directly, and treating the experimental RDC as an average value. The presumption is that one knows how to orient the two conformers in a way that allows them to share an order matrix. When the conformational differences are small, or the group undergoing conformational change is a minor portion of the molecule (exocyclic methylene rotation in a pyranose ring, for example), superposition of conserved portions by an RMSD overlay of conformers is very likely adequate. Where the origin of order is known, as in our alkylated disaccharide (22), a simple overlay the most ordered portion of the molecule is valid. In the case considered here, one of three sugar rings moves significantly. While this is not a minor portion of the

molecule, we will assume we can approximate a shared overall order matrix using a RMSD overlay of the entire molecule. We will illustrate the application of the REDCAT averaging procedure using this overlay in what follows.

A molecular example of conformational averaging.

Well before the advent of the REDCAT averaging procedure we collected and analyzed RDC data on a simple oligosaccharide free in solution, the trimannoside core found in all N-linked oligosaccharides of glycoproteins, methyl-3,6-di-*O*-(α -D-mannopyranosyl)- α -D-mannopyranoside (12,27). This structure is depicted in Figure 3. RDC data were collected in two different media, DMPC/DHPC bicells and bacteriophage (12). Both one bond ^1H - ^{13}C and multiple bond ^1H - ^1H couplings were collected yielding 60 RDCs distributed over the three sugar rings (12), as shown in Table 1.

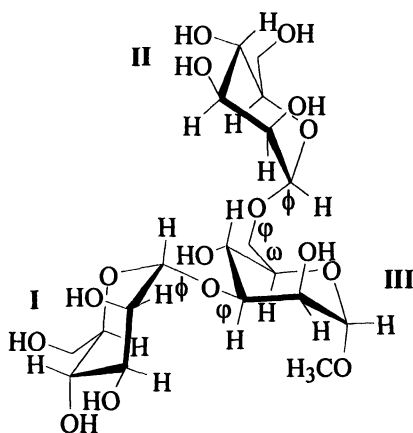


Figure 3. Structure of methyl 3,6-di-*O*-(α -D-mannopyranosyl)- α -D-mannopyranoside. Rings are labeled with Roman numerals, and glycosidic dihedral angles are labeled ϕ , φ and ω . The IUPAC nomenclature is used, ϕ : $O5(i)-C1(i)-O_n(i-1)-C_n(i-1)-C(n-1)(i-1)$, φ : $C1(i)-O_n(i-1)-C_n(i-1)-C(n-1)(i-1)$ ω : $O6(i)-C6(i)-C5(i)-C4(i)$, where (i) is a given residue and (n) is the number of the ring position.

Table 1. RDCs for trimannoside in aligned media (12)

| RDCs(Hz) | Residue Media | I | | II | | III | |
|----------|------------------|---------|-------|---------|-------|---------|-------|
| | | Bicells | Phage | Bicells | Phage | Bicells | Phage |
| C1H1 | | 9.3 | 2.3 | 4.0 | 1.0 | 16.5 | 5.5 |
| C2H2 | | -5.8 | 0.3 | 3.5 | 3.0 | -17.4 | -4.1 |
| C3H3 | | -0.2 | -0.4 | -10.9 | -2.6 | 10.9 | 1.9 |
| C4H4 | | | | | | 7.4 | 2.2 |
| C5H5 | | -3.0 | -3.9 | -12.9 | -3.3 | 11.2 | 2.6 |
| H1H2 | | 1.9 | 0.5 | 2.86 | 0.4 | -2.6 | -1.7 |
| H2H3 | | 3.65 | 1.37 | 1.3 | 0.86 | 4.1 | 1.7 |
| H1H3 | | 0.0 | 0.0 | 0.0 | 0.0 | | |
| H1H4 | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| H1H5 | | | | | | 0.0 | 0.0 |
| C1H2 | | | 0.1 | | 0.6 | | -0.7 |
| C2H1 | | | 0.0 | | 0.2 | | 0.0 |
| C3H2 | | | -0.1 | | | | 0.6 |
| C3H4 | | | | | | | -0.1 |

SOURCE: reproduced from reference 12. Copyright 2001, American Chemical Society

It was possible to determine the orientation of the principal order frame as seen from the point of view of each ring and do so separately for each medium. Rotating about glycosidic bonds to align order frames produced a structure for each medium. In principle, these structures should be identical, and they were for rings I and III (the 1-3 linked pair) (12). The glycosidic angles for this linkage were, in fact, very close to those observed in crystal structures and predicted by energy minimization (67° , -103° for ϕ and φ). However, the glycosidic conformation of the II–III pair (the 1-6 linked pair) seen in the two media appeared quite different. It is tempting to assume some media induced conformational preference resulted in this observation. However, at the very low levels of order being induced (less than 1 part in 1000 away from an isotropic distribution), it is unlikely that average conformations are really different. It is possible that one could strongly orient a minor member of a preexisting distribution of conformers in one medium and observe RDCs heavily weighted by the properties of this conformer, but it is more likely that all conformers are ordered to similar extents, and the different directions for interaction vectors sampled by motion in the two media leads to the observation. Here we explore the latter option using a new molecular dynamics simulation to identify conformers that could be sampled and the tools now introduced into the REDCAT program to see whether a single averaging process is consistent with RDCs measured in both media.

A new analysis of trimannoside data

Probable conformations for the trimannoside core have been investigated previously using molecular dynamics simulations to sample conformational space (28,29). However, force field refinements and the ability to do longer simulations with more powerful computers makes it useful to re-run these simulations. We have done this starting with a conformation suggested by previous simulations, ($\phi_{13} = 67$, $\varphi_{13} = -103$, $\phi_{16} = 64$, $\varphi_{16} = 180$, $\omega_{16} = 60^\circ$) (28). The molecule was solvated in a cubic box with explicit water molecules under periodic boundary conditions. The molecular dynamics (MD) simulation was carried out using the program AMBER 7.0 (30) with the GLYCAM_04 (31) force field on an HP Linux cluster using 4 dual processor nodes. The system was first well relaxed by applying a short energy minimization followed a MD simulation of 1.0 ns at 300 K. A 20 ns MD simulation was then performed at 300 K with an integration time step of 1 fs, storing snapshots every 500 fs.

The MD trajectory is shown in Figure 4. A few significant points can be noted. First, conformations about the 1-3 linkage are well clustered with ϕ_{13} and φ_{13} fluctuating by small oscillations (standard deviations of 25°) about values of 75 and -113° . The conformation about the 1-3 linkage can clearly be well represented by a single average conformation. Second, ϕ_{16} is also fairly well constrained to 75° with a standard deviation of 16° . Third, φ_{16} prefers to be near 180° , although it makes occasional excursions to both $+60$ and -60° . Because these transitions are on time scales of about 1 ns (short compared to the MD simulation of 20 ns) and the excursions are short-lived, it is reasonable to use the average value of 172.0° for φ_{16} to analyze the measured RDC data. Finally, ω_{16} , aside from an occasional excursion to -60° remained at 180 for a long period of time before making a single transition to $+60^\circ$. ω_{16} then remained in the $+60$ conformer for the duration of the MD run. It is clear that we cannot do a meaningful sampling of the ω_{16} angle in MD trajectories. However, as we can see from Figure 4, the ω_{16} is rather dynamic and the two major sampled conformers ($\omega_{16} = 180$ and $\omega_{16} = +60^\circ$) do provide a basis for modeling of experimental RDC data.

We will use the previously reported RDC data as given in Table 1, and the REDCAT averaging utility as described in the literature (20). The standard input file is as depicted in Figure 5. The columns normally are Cartesian coordinates for pairs of atoms displaying an RDC ($x_1, y_1, z_1, x_2, y_2, z_2$), followed by the experimental RDC value, the maximum coupling for nuclei of a particular pair if they were separated by a 1\AA , and an error estimate for the RDC. When an averaging model is appropriate, "AVG" is entered in place of an experimental RDC and the following line contains another set of coordinates for the same pair of atoms, but now from the alternate conformer. The RDC entry in the new line can be the experimental value if only two conformers exist and these are equally

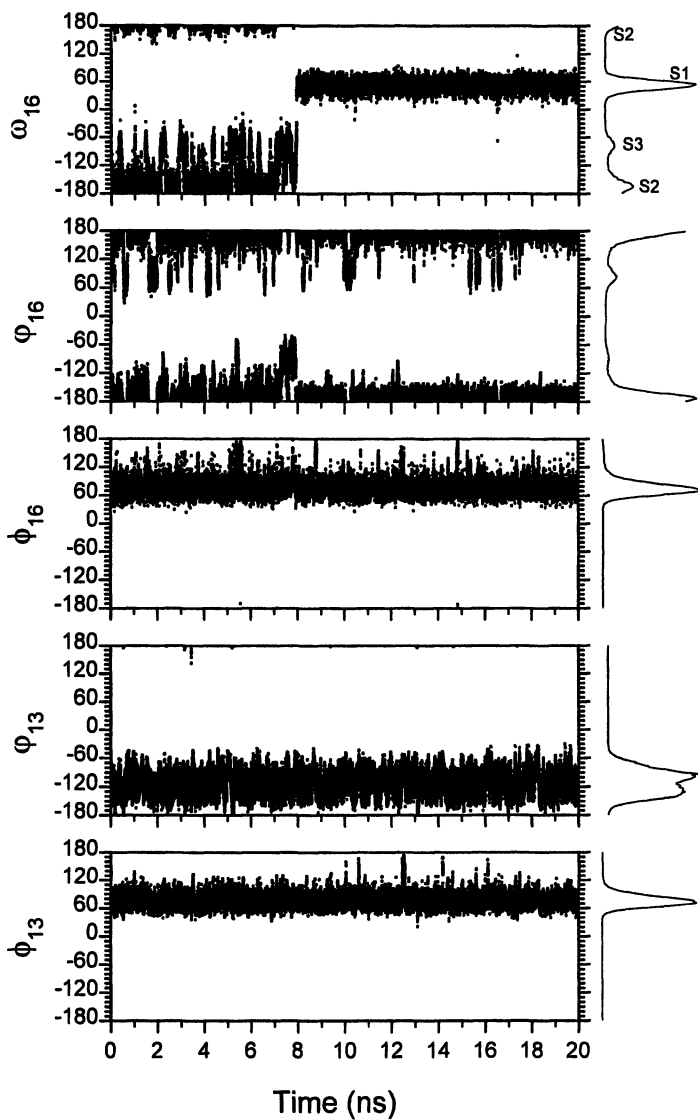


Figure 4. MD trajectories for glycosidic torsion angles of trimannoside. The MD simulation was carried out with explicit water molecules at 300 K.

populated, or it can be “AVG” if a third conformer is involved. Coordinate lines can be duplicated to effectively alter the population distribution between conformers. The example shown in Figure 5 is the one for which there are two conformers with a 3:2 population distribution.

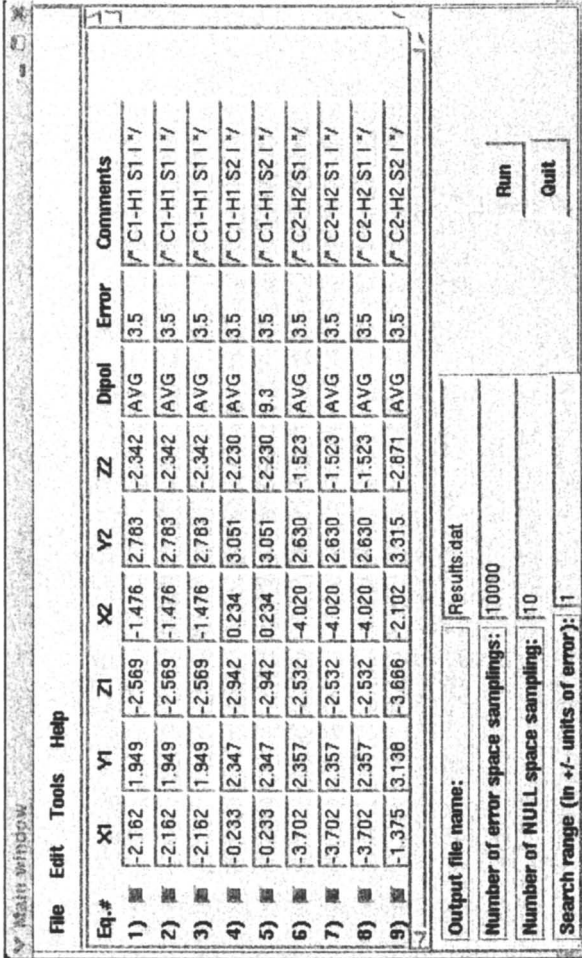


Figure 5. The main interface of REDCAT, with an input example for averaging RDCs over two states (S1 and S2 with a population of 3:2).

The conformers we choose for our model are derived from our MD trajectory by clustering and averaging glycosidic torsion angles in a RMSD sense. The conformers and their torsion angles are listed in Table 2.

Table 2. Glycosidic torsion angles for the three conformers

| | $(\phi_{13}, \varphi_{13})$ | $(\phi_{16}, \varphi_{16}, \omega_{16})$ |
|-------------------------------|-----------------------------|--|
| S1 (gauche-gauche, gg) | (75.4, -112.5) | (80.1, 176.3, 52.8) |
| S2 (gauche-trans, gt) | (75.1, -113.7) | (77.2, 174.1, -159.9) |
| S3 (trans-gauche, tg) | (74.5, -110.8) | (74.2, 172.0, -84.4) |

Notes: IUPAC nomenclature; units in degree.

The interpretation of RDCs in terms of averaging over multiple states could be challenging because of the need to choose a proper alignment for conformers. Molecules are tumbling in a liquid crystal solution and each conformer could be aligned in a different way. In our previous work we knew that the reducing end residue dominated the overall order because the alkyl chain attached to this sugar anchored the molecule to the bicelle membrane through this residue. The atoms of this residue were, therefore, overlaid in each conformer involved in an average (22). For the trimannoside there is no such clear basis for aligning conformers.

Figure 4 and Table 2 show that the 1–3 linked residues in mannoside do have a relatively well defined glycosidic conformation. We might try to use these as a point of reference assuming they, as opposed to the single 1-6 linked sugar, would dominate overall ordering in the alignment media. This type of size-based assumption has been successfully used in analyzing motion of ligands bound to proteins and the motion of small domains of larger proteins (32,33). The three states of the trimannoside were reorientated in space by RMS fitting the coordinates of heavy atoms on ring I and ring III and the results are shown in Figure 6. Based on this figure, this assumption is clearly suspect, because the dramatic change in overall shape for the S1 state in particular.

An alternate approach is to attempt a calculation of order matrix axis directions for each conformer and then align the order matrix axes. Reasonable results have obtained from Allen Bush's group, assuming that the principal axes of the moment of inertia tensor of the molecule to be the same as those of the principal alignment frame of the order matrix (34). Andrew Almond and Jacob B. Axelsen suggested the gyration tensor may be another good way to evaluate the alignment of neutral molecules (35). The two methods, in fact, give the same orientation because the two tensors share eigenvectors with only eigenvalues being different (35). Figure 7 shows the three conformers aligned in space according to their moment of inertia tensor. If inertia tensors adequately represent alignment tensors, the combined effects of overall orientation and

internal motional averaging would be well modeled. However, the relationship between moment of inertia tensors or gyration tensors and liquid crystal alignment tensors remains in question.

Here we choose to follow a somewhat simpler procedure based on an RMSD overlay of the entire molecule. An alignment based on an RMSD overlay is closely related to well established steric obstruction model (PALES) that has been used to successfully predict alignment of biological molecules in neutral aqueous liquid crystalline media like bicelles, and rod-like media like phage (36). Figure 8 shows the above three conformers reorientated to the best overlay of the atomic coordinates (RMSD fitting) of all the three residues, except those flexible hydroxyl protons and hydroxyls of methylene groups. The alignments of the conformers are very similar to those reorientated to align moment of inertia tensors. For example, an exact superposition of the S2 and S3 conformers, having aligned the S1 conformers for the two methods can be achieved with Euler angle rotations (z - y '- z '') of (224.5, 5.8, 106.5 °) and (256.0, 4.0, 130.0°), respectively. The orientation differences are about $\pm 30^\circ$, as rotations along y -axis are near zero and the accumulative rotations about z -axis are around $\pm 30^\circ$.

Simulation of the experimental RDC data was performed in REDCAT, using the three RMSD reorientated conformers with different population ratios. We begin our analysis considering the fit with just a single conformer first, and then focus on the averaging RDCs using pairs of conformers, and later add a third conformer as needed. We have set error limits for RDCs to 3.5 Hz for one bond C-H couplings and 1.5 Hz for other couplings in bicelle media and to 2.5 Hz for one bond C-H couplings and 1.0 Hz for other couplings in phage media. Only solutions having all couplings meet these criteria are accepted. RDCs were back calculated using order parameters from the best solution in an RMSD sense or using order parameters averaged over all allowed solutions. Linear regression of the back calculated results was done and the squared linear correlation coefficients (R^2) are reported. The number of solutions and the squared linear correlation coefficients are then used as criteria to judge the quality of the averaging models. The results are summarized in Table 3.

Several conclusions can be drawn from Table 3. First, no single conformer can represent the experiment data within the levels of error we have specified; this indicates that significant internal motion must exist in the trimannoside. We can however find a best solution regardless of error in order to evaluate motion in a single conformer model. For each ring, as mentioned in the introductory paragraphs, the order parameters should be the same if the residues are rigidly connected, whereas differences in order parameters can reflect the level of internal motion. It is convenient to combine all order parameters into a general degree of order (GDO) for this assessment. The GDOs are 4.4×10^{-4} , 3.4×10^{-4} and 5.2×10^{-4} in bicelles, and 1.5×10^{-4} , 1.3×10^{-4} and 1.8×10^{-4} in phage for sugar residues I, II and III, respectively (12). The difference in GDO values are

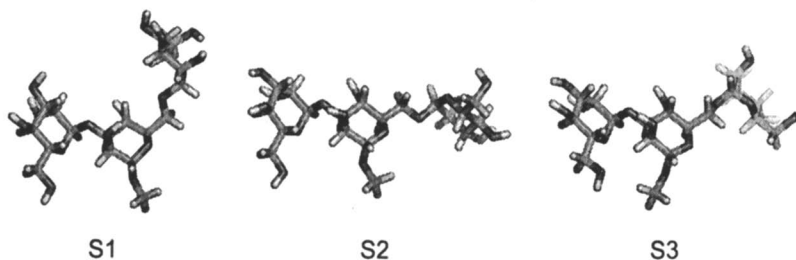


Figure 6. Alignment of conformers assuming their orientations in liquid crystal are dominated by the orientations of rings I and III.

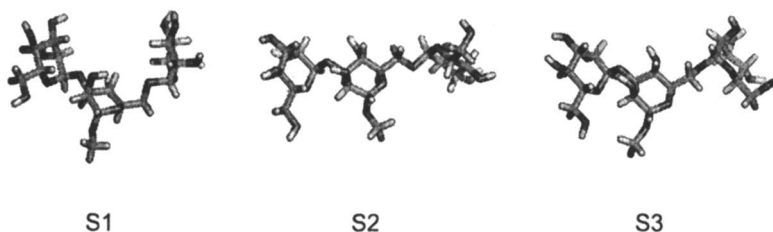


Figure 7. Alignment of conformers assuming their orientations in liquid crystal are dictated by a moment of inertia tensor.

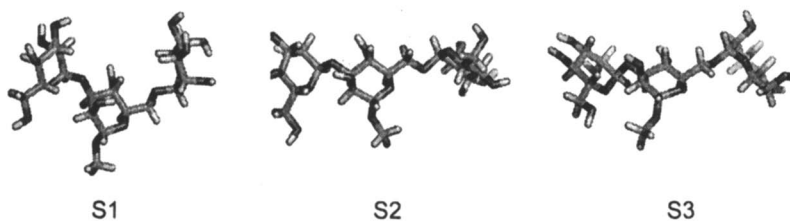


Figure 8. Alignment of conformers assuming orientations in liquid crystal are dictated by shape. The reorientation was accomplished by RMS fitting of the coordinates of atoms on all the three sugar residues.

larger for residue II and III (1–6 linkage) than for residue I with III (1–3 linkage) in both alignment media, suggesting the glycosidic linkage between II and III is more flexible. This difference in flexibility for the 1–6 linkage and the 1–3 linkage agrees with our above MD simulation.

Second, when multiple states are allowed, we do get solutions, but in all cases S1 and S2 are the major states. Combinations of S1 and S2 in which populations are nearly equal give large numbers of solutions and the back calculated RDCs show good correlation with experiment observables, both in bicelles and in phage media. Although combinations of S1 and S3 give some solutions for the data obtained in phage medium with generous errors (2.5 Hz for C–H one bond couplings and 1.0 Hz for others, (50% of the experimental RDCs range), the number of solutions is much smaller and the correlation coefficients between the back calculated and experiment measurements are lower than those for the S1 and S2 combinations. Moreover, the combinations of S1 and S3 give no solution for bicells data. No solutions were observed from the combinations of S2 and S3 conformers, also suggesting that S1 and S2 are the major populated states. These two major states were also observed in our previous studies using NOESY data and MD simulations (28).

Third, the conformers of S1 and S2 are almost equally sampled, because pretty good and very similar results are observed for S1 and S2 with a population ratios of 2:3 (40%:60%), 1:1(50%:50%) and 3:2(60%:40%), in both alignment media. The linear correlation coefficients for these three combinations are 0.964, 0.968, 0.959 in bicells, and 0.868, 0.903 and 0.905 in phage, respectively. The Sauson-Flamsteed projections of the alignment tensors for the solutions are presented in Figure 9. While the order tensors are different between bicelle and phage, few differences in direction of axes for the average order tensors for a given medium are seen for these three populations ratios either in bicelles or in phage media.

Finally, S3 may be populated to a small extent (< 20%) in the presence of larger amounts of S1 and S2. Correlation coefficients for bicelles and number of solutions for phage improve somewhat for small percentages of S3 (9% (5:5:1) and 14% (3:3:1) as opposed to no S3. The number of solutions and correlation coefficients for bicelles and for phage decrease somewhat as the population of S3 increase to 20% (2:2:1). No solutions are observed when the population of S3 is 33% (1:1:1).

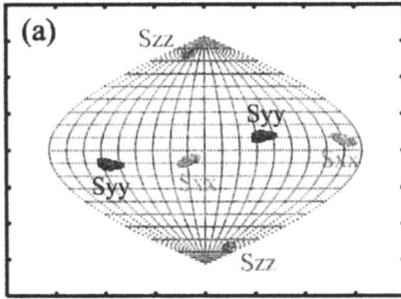
Discussion

The data presented above suggest that the origin deviations in structures initially found for single conformer analysis of the trimannoside in two orientation media (12) was in fact due to conformational averaging. Fits to RDC data from both media are possible with two state and three state models that

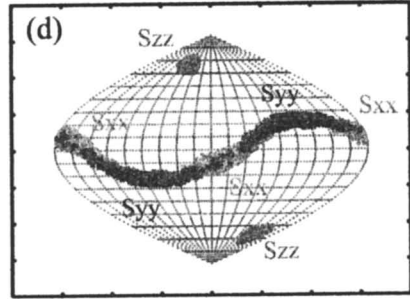
Table 3. Averaging RDCs from multi-conformers for trimannoside

| ratios | | | <i>Bicelle</i> ^a | | <i>Phage</i> ^b | |
|-----------|-----------|-----------|-------------------------------|-------------------------|-------------------------------|-------------------------|
| <i>S1</i> | <i>S2</i> | <i>S3</i> | <i>Solutions</i> ^c | <i>R</i> ^{2 d} | <i>Solutions</i> ^c | <i>R</i> ^{2 d} |
| 1 | | | 0 | | 0 | |
| | 1 | | 0 | | 0 | |
| | | 1 | 0 | | 0 | |
| 1 | 4 | | 0 | | 0 | |
| 2 | 3 | | 492 | 0.964 | 2353 | 0.868 |
| 1 | 1 | | 2284 | 0.968 | 2371 | 0.903 |
| 3 | 2 | | 1077 | 0.959 | 2073 | 0.905 |
| 4 | 1 | | 0 | | 910 | 0.837 |
| 1 | | 4 | 0 | | 0 | |
| 2 | | 3 | 0 | | 18 | 0.687 |
| 1 | | 1 | 0 | | 436 | 0.768 |
| 3 | | 2 | 0 | | 971 | 0.805 |
| 4 | | 1 | 0 | | 295 | 0.763 |
| | 1 | 4 | 0 | | 0 | |
| | 2 | 3 | 0 | | 0 | |
| | 1 | 1 | 0 | | 0 | |
| | 3 | 2 | 0 | | 0 | |
| | 4 | 1 | 0 | | 0 | |
| 1 | 1 | 1 | 0 | | 0 | |
| 2 | 2 | 1 | 3 | 0.952 | 2260 | 0.881 |
| 3 | 3 | 1 | 2028 | 0.971 | 2722 | 0.893 |
| 4 | 4 | 1 | 2371 | 0.972 | 2652 | 0.897 |
| 5 | 5 | 1 | 2458 | 0.972 | 2585 | 0.899 |

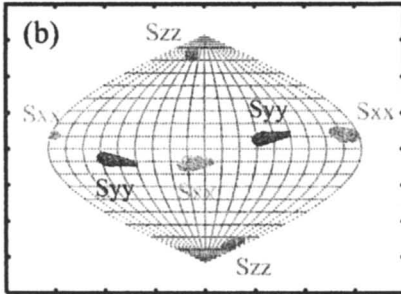
Notes: ^a 3.5 Hz errors for one bond C–H RDCs and 1.5 Hz for others are used; ^b 2.5 Hz errors for one bond C–H RDCs and 1.0 Hz for others are used; ^c REDCAT uses a statistic method to sample the solutions within experiment errors. A total number of 10000 samplings was examined; ^d The linear correlation coefficients between experiment RDCs and back calculated values either from the REDCAT best solution or from the average of the solutions.



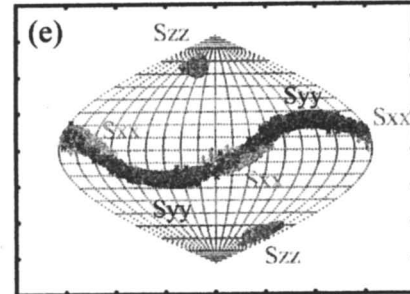
S1:S2=2:3, in Bicells



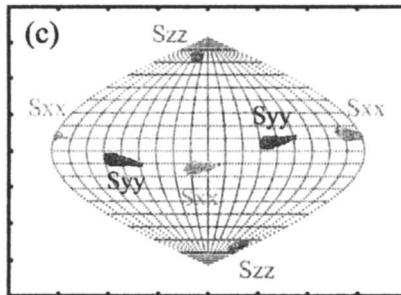
S1:S2=2:3, in Phage



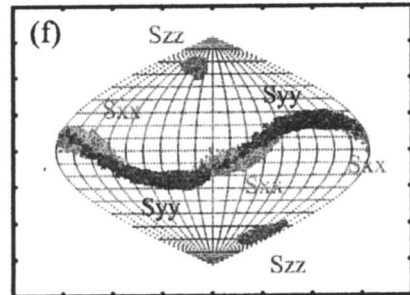
S1:S2=1:1, in Bicells



S1:S2=1:1, in Phage



S1:S2=3:2, in Bicells



S1:S2=3:2, in Phage

Figure 9. Sauson-Flamsteed projections of the alignment tensors in bicelles (a, b and c) and in phage (d, e and f) for different ratios of conformers.

incorporate large percentages of two of the conformers most frequently sampled in our MD trajectory (S1 and S2). The agreement also speaks well for the ability of current force fields to accurately predict stable conformers of oligosaccharides. The population ratio cannot be compared to the MD predictions simply because even at 20 ns, the trajectory is too short to adequately sample populations. The utility of averaging tools incorporated in the REDCAT program is nicely illustrated through its ability to evaluate populations.

The fact that the conformation of the 1–3 linkage is well represented by a single state, and that the 1–6 linkage is much more dynamic is not unexpected. Other authors have reached this conclusion for similar molecules (37). The ability to model the actual conformations and estimate populations of various conformations is, however, a valuable addition. Knowing accessible conformations will be useful in synthesizing constrained mimics of natural carbohydrates that represent single accessible states that might bind to protein receptors for these carbohydrates.

The acquisition of RDC data and the ability to extract conformational models is, of course, not limited to studies of carbohydrates free in solution. RDCs for carbohydrates bound to proteins can also be measured (38–40). One does not expect as extensive averaging due to internal motion when bound to protein as when free in solution. In fact, a single conformer near to our S1 state was found to give adequate fits to transferred NOE data (29) and to RDC data (38) for the trimannoside bound to mannose binding protein. However, exploring the possibility of motion when bound to proteins remains important. This is not only because such motions may exist in bound states, but because proper averaging may allow use of data that would otherwise be incompatible with a single state analysis. For carbohydrates, for example, data from the exocyclic methylenes in pyranose rings is a good example. These methylenes are known to sample various rotamers and C–H couplings from these sites can only be used if a proper motional model can be applied. We therefore expect significant improvements in use of RDCs for conformational analysis with the addition of tools such as that described here.

Acknowledgment

This work is supported by NIH Grants GM33225 and RR05351. We would like to thank Dr. Woods R. J. for providing a platform to run molecular dynamics simulations.

References

1. Dwek, R. A. *Biochem. Soc. Trans.* **1995**, *23*, 1.
2. Helenius, A.; Aebi, M. *Annu. Rev. Biochem.* **2004**, *73*, 1019.

3. Gabor, F.; Bogner, E.; Weissenboeck, A.; Wirth, M. *Adv. Drug Deliver Rev.*, **2004**, *56*, 459.
4. Mu, J.; Roach, P. J. *J. Biol. Chem.*, **1998**, *273*, 34850.
5. Armstrong, G. D. *Curr. Opin. Drug Di. De.* **2000**, *3*, 191.
6. Schleucher J.; Wijmenga S. S. *J. Am. Chem. Soc.* **2002**, *124*, 5881.
7. Daragan V. A.; Mayo K. H. *Prog. Nucl. Mag. Res. Sp.* **1997**, *31*, 63.
8. Frueh D. *Prog. Nucl. Mag. Res. Sp.* **2002**, *41*, 305.
9. Larsson, G.; Martinez, G.; Schleucher, J.; Wijmenga, S. S. *J. Biomol. NMR* **2003**, *27*, 291.
10. Prestegard J. H.; Al-Hashimi H. M.; Tolman J. R. *Q. Rev. Biophys.* **2000**, *33*, 371.
11. Bax, A.; Kontaxis, G.; Tjandra, N. *Method. Enzymol.*, **2001**, *B 339*, 127.
12. Tian F; Al-Hashimi H. M.; Craighead J. L.; Prestegard J. H. *J. Am. Chem. Soc.* **2001**, *123*, 485.
13. Delaglio F.; Wu Z.; Bax A.; *J. Magn. Reson.* **2001**, *149*, 276.
14. Marquez, B. L.; Gerwick, W. H.; Williamson, R. T. *Magn. Reson. Chem.* **2001**, *39*, 499.
15. Meissner, A.; Sorensen, O. W. *Magn. Reson. Chem.* **2001**, *39*, 49.
16. Krishnamurthy, V. V. *J. Magn. Reson.*, **1996**, *A 121*, 33.
17. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.*, **2003**, *160*, 65.
18. Losonczi J. A.; Andrec M.; Fischer M. W. F.; Prestegard J. H. *J. Magn. Reson.* **1999**, *138*, 334.
19. Saupe, A. *Angew. Chem., Int. Ed. Engl.* **1968**, *7*, 97.
20. Valafar H.; Prestegard J. H. *J. Magn. Reson.* **2004**, *167*, 228.
21. Palmer, A. G., III. *Chem. Rev.* **2004**, *104*, 3623.
22. Yi, X.; Venot, A.; Glushka, J.; Prestegard, J. H. *J. Am. Chem. Soc.* **2004**, *126*, 13636.
23. Martin-Pastor M.; Bush C. A. *Carbohydr. Res.* **2000**, *323*, 147.
24. Lycknert K.; Maliniak A.; Widmalm G. *J. Phys. Chem.* **2001**, *A 105*, 5119.
25. Woods, R. J.; Dwek, R. A.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832.
26. Sakae, Y.; Okamoto, Y. *J. Theor. Comput. Chem.* **2004**, *3*, 339.
27. Dam, T. K.; Cavada, B. S.; Grangeiro, T. B.; Santos, C. F.; De Sousa, F. A. M.; Oscarson, S.; Brewer, C. F. *J. Biol. Chem.* **1998**, *273*, 12082.
28. Sayers, E. W.; Prestegard, J. H. *Biophys. J.* **2002**, *82*, 2683.
29. Sayers, E. W.; Prestegard, J. H. *Biophys. J.* **2000**, *79*, 3313.
30. Case D. A.; Pearlman D. A.; Caldwell J. W.; Cheatham III T. E.; Wang J.; Ross W. S.; Simmerling C. L.; Darden T. A.; Merz K. M.; Stanton R. V.; Cheng A. L.; Vincent J. J.; Crowley M.; Tsui V.; Gohlke H.; Radmer R. J.; Duan Y.; Pitera J.; Massova I.; Seibel G. L.; Singh U. C.; Weiner P. K.; Kollman P. A.; *AMBER Revision 7.0*, **2002**, University of California, San Francisco.

31. Woods R. J. *GLYCAM_04*, <http://glycam.ccruc.uga.edu/index.html>.
32. Umemoto, K.; Leffler, H.; Venot, A.; Valafar, H.; Prestegard, J. H. *Biochemistry* **2003**, *42*, 3688.
33. Fischer, M. W. F.; Losonczi, J. A.; Weaver, J. L.; Prestegard, J. H. *Biochemistry* **1999**, *38*, 9013.
34. Azurmendi, H. F.; Bush, C. A. *J. Am. Chem. Soc.* **2002**, *124*, 2426.
35. Almond, A.; Axelsen, J. B. *J. Am. Chem. Soc.* **2002**, *124*, 9986.
36. Zweckstetter M.; Bax A. *J. Am. Chem. Soc.* **2000**, *122*, 3791-3792.
37. Bush A. C.; Martin-Pastor M.; Imberty, A. *Annu. Rev. Bioph. Biom.* **1999**, *28*, 269.
38. Jain, N. U.; Noble, S.; Prestegard, J. H. *J. Mol. Biol.* **2003**, *328*, 451.
39. Shimizu, H. *Trends Glycosci. Glyc.* **2003**, *15*, 221.
40. Thompson, G. S.; Shimizu, H.; Homans, S. W.; Donohue-Rolfe, A. *Biochemistry* **2000**, *39*, 13153.

Chapter 4

Protein–Carbohydrate Interactions: A Combined Theoretical and NMR Experimental Approach on Carbohydrate–Aromatic Interactions and on Pyranose Ring Distortion

Jesús Jiménez-Barbero^{1,*}, F. Javier Cañada¹, Gabriel Cuevas²,
Juan L. Asensio¹, Nuria Aboitiz¹, Angeles Canales¹,
M. Isabel Chávez^{1,2}, M. Carmen Fernández-Alonso¹,
Alicia García-Herrero¹, Silvia Mari¹, and Paloma Vidal¹

¹Centro Investigaciones Biológicas, C.S.I.C., Ramiro de Maeztu 9, 28006 Madrid, Spain

²Instituto de Química, Universidad Nacional Autónoma de México, Circuito Exterior, Coyoacán, México, D.F.

Abstract

Carbohydrate recognition by lectins and enzymes requires the consideration of a variety of structural and conformational details. The relevance of sugar-aromatic interactions for molecular recognition of carbohydrates by lectins has been studied by a combined protocol using NMR and theoretical calculations. In addition, and in relation to saccharide recognition by glycosidase enzymes, we also present experimental and theoretical evidences on the importance of the chemical nature of the sugar or glycomimetic to suffer ring distortion. Two different systems (lectin and glycosidase) have been evaluated. As lectin system to test the importance of sugar-aromatic interactions, a chitin binding domain (hevein) has been employed. For the analysis of pyranoid ring distortion, the interaction of *E. coli* β -galactosidase with lactose and glycomimetics thereof has been explored. For the theoretical studies, simple models of both the sugar and protein systems have been studied.

Introduction

Unraveling the mechanisms that rule how sugar molecules are bound by lectins, antibodies, and enzymes is currently a topic of major interest (1). A detailed understanding of the structural, thermodynamics and kinetic features of the molecular complex when carbohydrates are bound to receptors is indeed relevant (2), and that requires the use of a multidisciplinary approach to study this phenomenon in depth (3). The concerted use of a variety of techniques and the access to synthetically prepared saccharides and glycomimetics (4), as well as to natural and "designed" protein domains is of great importance to this aim (5). X-ray, NMR, and other biophysical methods have been more widely used to access detailed structural and thermodynamic information (6).

Due to the amphipatic character of saccharide molecules, different types of forces may be involved in its recognition by receptors (7). The presence of the hydroxyl groups makes possible their involvement in intermolecular hydrogen bonds to side-chains of polar amino acids (8). Moreover, it has also been hypothesized that water provides the driving force for the complex formation (9). Nevertheless, not only these polar interactions are involved in carbohydrate recognition. Experimental data have shown that depending on the stereochemistry of the saccharide, the presence of a number of rather apolar C-H groups indeed constitute patches, that interact with the protein side chains. Although the exact nature and origin of this interaction is still upon investigation (10), it can be proposed that the mutual shielding of the non-polar surfaces from bulk water is entropically favorable (11), and that the electrostatic interaction between the positive net charge of the C-H groups and that the quadrupole created by the π -system of the aromatic ring makes a favorable enthalpic contribution (12). As in other protein/ligand complexes, probably, the polarizability of the aromatic electrons and the polarizing nature of the C-H vector may lead to an attractive force. Indeed, similar features have been proposed to account for an important portion of the driving force in ligand-accommodating mechanisms (13).

Herein we present theoretical and experimental evidences on the importance of the chemical nature of the aromatic ring for the sugar interaction process. Moreover, we also present a theoretical study of the physicochemical origin of the intermolecular interaction between sugar and aromatic rings, employing different levels of theory.

Regarding carbohydrate processing enzymes, structural studies on molecular recognition of oligosaccharides and analogues by glycosidase and glycosyltransferase enzymes have shown that distortion of the groundstate conformation of the saccharide ligand may take place on numerous occasions (14). The attained degree of distortion may resemble the proposed oxocarbenium-type transition state of glycoside hydrolysis (15). The determination of the energy associated with the deformation in unsubstituted rings is a first step towards understanding the behavior of complex systems. This

knowledge is of paramount importance to comprehend and predict the outcome of the recognition process and to rationally design inhibitors of biomedical processes in which these enzymes are involved (16). On this basis, we present also herein experimental and theoretical evidences on the importance of the chemical nature of the sugar or glycomimetic to suffer ring distortion. Moreover, we evaluate the inversion and the conformational exchange processes of oxane and thiane (17), and place them within the context of a global inversion scheme.

Two different systems have been evaluated. As lectin system to test the importance of sugar-aromatic interactions, hevein and related chitin binding domains have been employed. For the analysis of pyranoid ring distortion, the interaction of *E. coli* β -galactosidase with lactose and glycomimetics thereof has been evaluated. For the theoretical studies, simple models of both the sugar and protein systems have been studied.

Hevein-chitooligosaccharide interactions

From the lectin recognition viewpoint, it is today known that among the biological processes in which carbohydrates are involved, many plants respond to pathogenic attack by producing defense proteins able to bind reversibly to chitin (18), the ubiquitous β (1-4)-linked *N*-acetylglucosamine (GlcNAc) polysaccharide. This natural biopolymer is a key structural component of the cell wall of fungi, and of the exoskeleton of invertebrates, such as insects and nematodes. Indeed, anti-fungal activity of plant chitinases is largely restricted to those chitinases that contain a noncatalytic, plant-specific chitin-binding domain (also dubbed hevein domain, 19). This domain displays a common structural motif of 30-43 residues rich in glycines and cysteines in highly conserved positions and organized around a four disulfide core (20). The hevein domain is present in hevein itself (21) and many other lectins (22).

NMR has been widely employed to deduce the 3D architecture of several of these hevein domains. A variety of studies on latex hevein (23-27), five-disulfide containing heveins (28, 29), a 32 amino acids truncated hevein (30), pseudohevein (31), the B domain of WGA (32), as well as for natural AcAMP2 (33), and related Ac-AMP2 (34) peptides have been described. Basically, the 3D structures of these small proteins are again very similar among them and also basically identical to their respective solid state structures described by X-ray. Using this technique, the structures of WGA (35), UDA (36-38), and hevein itself (39,40) have been elucidated.

Except for the first X-ray structure of hevein (39), the polypeptide backbone of all these structures keeps a similar architecture, with rmsd deviations below 2 Å among the different structures.

A variety of NMR-based techniques have been used to study the binding of chitooligosaccharides to hevein domains. It is worthy to mention the use of

titration NMR (41), based on chemical shift perturbation analysis (42) of the free versus the bound species, diffusion ordered spectroscopy (DOSY) (43), the analysis of interprotein and intermolecular sugar-protein NOEs (44), and laser induced photo-cidnp methods (45). These techniques, complemented by differential scanning (46) or isothermal titration microcalorimetry (47), fluorescence (48), infrared spectroscopy (49), analytical ultracentrifugation (50), and molecular modeling (51) and X-ray protocols have permitted to analyze several complexes of hevein domains bound to oligosaccharide ligands at atomic resolution and to analyze their binding energy features.

Importantly, all the structures show a cluster of three aromatic residues and one serine that form the basic carbohydrate-binding domain (52, 53). Indeed, the location of the chitoooligosaccharide binding-site of hevein was deduced by chemical shift perturbation analysis and further confirmed by the presence of a number of unambiguous intermolecular protein-carbohydrate NOEs. Laser photo-CIDNP methods also allowed demonstrating the presence of Tyr and Trp residues in sugar binding (54).

For instance, the $\Delta\delta$ (complex-free) of protein protons permit to deduce that the binding site for chitoooligosaccharides is located between residues 18 to 30, since their corresponding ¹H NMR signals show the major changes in chemical shifts upon interaction with the sugars. Moreover, several hundreds of protein-protein NOEs as well as several protein-sugar NOEs were deduced and used as upper bound constraints in a simulated annealing protocol to determine the three dimensional structure of the hevein-sugar complexes. In addition, a hydrogen bond between the hydroxyl group of Ser19 and the carbonyl group of one of the GlcNAc residues could always be detected and included as additional constraint.

The resulting three-dimensional structures of the protein/carbohydrate complexes are fairly well defined (Fig. 1 and 2). Indeed, the experimental data indicate that the protein experiences only slight changes in its conformation, when interacting with the saccharides. Indeed, with regard to the NMR structure of the free proteins, no important changes in the protein NOEs were observed, indicating that carbohydrate-induced conformational changes are small (Fig. 3). As example, the average backbone rmsd of 20 refined NMR structures for the hevein/(GlcNAc)₅ complex was 0.6 Å, while the heavy atom rmsd was 1.2 Å.

Generally speaking, the hevein domain backbone maintains the same topology in the free and bound states, and minor movements are observed in the lateral chains of the amino acids, which form the binding site. The study of the 3D structure of hevein with (GlcNAc)₂ indicated that both GlcNAc residues make interactions with several protein² lateral chains: The non-reducing end makes stacking to the aromatic Trp23 moiety. Additionally, the non-reducing acetamido methyl group displays non-polar contacts to the aromatic Tyr30 residue, and, in addition, there are key hydrogen bonds which confer stability to the complex: one between the non-reducing sugar acetamido group and Ser19 and a second one involving C3-OH and Tyr30. An additional interaction is

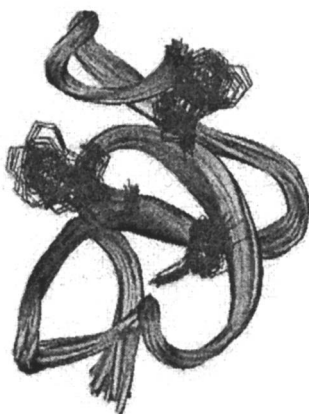


Figure 1. Collection of 23 NMR structures for hevein bound to (GlcNAc)₅. The backbone fit RMSD between residues (3-41) is 0.57 Å; between residues (16-1): 0.39Å and for the key lateral chains holds S19: 0.19; W21: 0.28; W23: 0.32; Y30: 0.30.

(See page 1 in color insert in this chapter.)

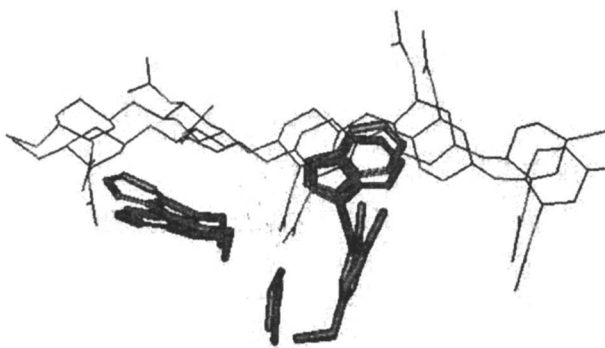


Figure 2. Superimposition of two NMR structures for hevein bound to (GlcNAc)₅. The orientation of the key lateral chains is emphasized.

(See page 1 in color insert in this chapter.)



Figure 3. Superimposition of the NMR structures for free and (GlcNAc)₅-bound hevein. The orientation of the key lateral chains is emphasized.

(See page 2 in color insert in this chapter.)

observed between the less polar α -face of the reducing GlcNAc moiety and the plane of the indole ring of Trp21. Additional evidences for the existence of this stacking interaction between the lateral chain of Trp21 with the reducing end came from the observation of strong shielding of several protons of the indole ring of Trp21 in the complex of hevein with *p*-nitrophenyl-GlcNAc, and from the upfield shifting of the *O*-methyl group of methyl chitobioside in the presence of hevein with respect to that measured for the free sugar.

This combination between van der Waals, CH- π , and hydrogen bond interactions is perfectly suitable to provide the basic features of the interaction between hevein and chitoooligosaccharides.

Hevein domains have also provided a suitable model to verify the minimum chitoooligosaccharide binding domain. Based on the structure of the natural AcAMP-2 antifungal polypeptide, a 32-residue truncated hevein lacking eleven C-terminal amino acids (HEV32), was prepared and correctly folded with three disulfide bridge pairs (30). The NMR structure of ligand-bound HEV32 in aqueous solution was determined to be highly similar to the NMR structure of ligand-bound hevein. HEV32 might provide a simpler molecular model for studying protein-carbohydrate interactions and for understanding the physiological relevance of small native hevein domains lacking C-terminal residues.

Moreover, hevein may also accommodate ManNAc and GalNAc acetamidoglycans at certain positions of the chitin chain (55). NMR data assisted by modeling protocols indicate that modifications at either the reducing end (with ManNAc instead of GlcNAc), or at the non-reducing end (with GalNAc instead of GlcNAc) do not modify the mode of binding of the saccharide to

hevein. Nevertheless, the association constant values indicate that chitotriose binding is better than that of its ManNAc or GalNAc analogues.

By taking into consideration these geometries, and determining the binding affinities and the thermodynamic parameters for the interaction, it may be easily deduced that both non-polar and polar interactions contribute to the complexation process, stabilizing the orientation of the sugar rings through the formation of hydrogen bonds and stacking interactions with aromatic side chains. Therefore, the structural view obtained in solution perfectly agrees with the deductions from the equilibrium thermodynamic parameters. The variations in binding constants can be explained in structural terms: The minimum sugar size which can be bound by hevein is the GlcNAc monosaccharide, which is stabilised by non-polar forces involving Trp23 and Tyr30, and by hydrogen bonds involving Ser19 and the hydroxyl group of Tyr30. The binding constant of chitobiose is one order of magnitude higher. The new reducing end makes additional non-polar contacts to Trp21. Moreover, with methyl β -chitobioside, the binding is improved. Both the fixing of the anomeric configuration in a favored β -orientation, and the presence of additional non-polar interactions between both the reducing end and the O-methyl group and the extended surface of Trp21 are probably the key factors in this case. The higher affinities deduced for the beta-linked disaccharide with respect to the GlcNAc monosaccharide and to GlcNAc(α 1 \rightarrow 6)-Man can only be explained by favorable stacking of the second beta-linked GlcNAc moiety and Trp21 (56).

The binding constant found for chitotriose (GlcNAc)₃ is one order of magnitude higher than that to (GlcNAc)₂, due to the better van der Waals contacts which are established between the extended surface of Trp21 indole ring and the pyranoid chair. Thus, using these simple models, there is approximately 1.5 kcal/mol increase in the free energy of binding per additional sugar unit, when passing from the monosaccharide to the disaccharide and to the trisaccharide. This gain in free energy of binding is basically accounted for by additional van der Waals/stacking interactions between the sugar and the properly oriented aromatic rings in the protein structure. Indeed, no additional intermolecular hydrogen bonds to those observed for the monosaccharide are taking place for the di- and trisaccharide in the hevein binding-site. The gain in enthalpy amounts to about 2 kcal/mol per additional sugar/aromatic interaction, with a concomitant loss of entropy of binding, as typically observed in lectin-sugar interactions. For higher oligosaccharides, multivalency effects start to appear, and several hevein domains are bound to the same polysaccharide chain (57). Interestingly, the chemical nature of the aromatic ring at relative position 21 (Trp 21 in hevein) modulates the binding energy and enthalpy of the sugar-protein interaction. Taking a truncated hevein of the AcAMP2 family as scaffold, it can be observed that the larger the aromatic ring size, the higher the free energy and enthalpy of binding. In particular, AcAMP2 domains with either phenylalanine, tryptophan, or non-natural naphthylalanine side chains have been

studied (58). These domains show an increasing affinity to (GlcNAc)₃ that parallels the increase in the surface of the aromatic ring. Moreover,³ the deactivation of the aromatic rings at relative positions 21 and 23 produce a two-fold decrease in the binding ability of the same AcAMP2 scaffold to (GlcNAc)₃. These data strongly indicate the importance of sugar-aromatic interactions for the molecular recognition process (34).

With these experimental data at hand, we decided to theoretically study the origin of this interaction. A simple sugar-aromatic model was chosen, with just a benzene ring and a 6-deoxy sugar moiety to avoid the additional complications of several rotamers around the C5-C6 ring (Fig. 4).

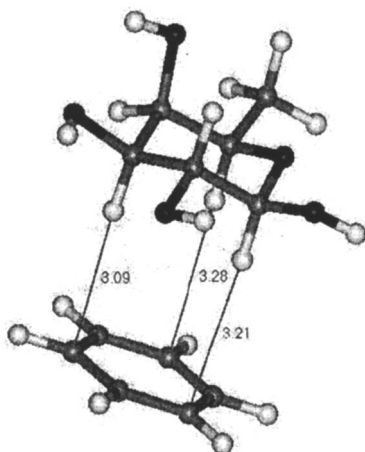


Figure 4. The molecular complex used to study the sugar-aromatic interaction. The result at the b3lyp/6-31G(d,p) level of theory is used, with the corresponding key intermolecular distances

The full geometry optimization of the sugar-benzene complex has been performed with Gaussian98 and Gaussian03 (59, 60) at different levels of theory, using Density Functional Theory (DFT) and Møller-Plesset calculations. In particular, MP2/6-31G(d, p) (method A), b3lyp/6-31G(d, p) (method B), MP2/6-31G(d, p) corrected by the counter poise (CP) method (61) (method C), and MP2/6-31G(d, p)// b3lyp/6-31G(d, p) (method D).

For the calculations employing the MP2/6-31G(d, p)// b3lyp/6-31G(d,p) level of theory, the resulting structure from the b3lyp/6-31G(d, p) optimization was employed, and just a single point (SP) calculation was performed. Vibrational frequency calculations were also carried out in order to confirm that the obtained structures were indeed minima, at levels A, B and C.

In addition, Potential Energy Curves were calculated for the complex. Thus, single point calculations were performed with Gaussian 98, varying the distance between a given C-H bond and the closest corresponding carbon atom. To

determine the interaction energy, the Basis Set Superposition Error Correction (BSSE) was calculated (62). The Counterpoise method was used, and thus the proper correction by the change of geometry of the components of the complex was considered. The method proposed by Sotiris and Xantheas (63), that permits optimization of the geometry by considering the basis set superposition error, as implemented in Gaussian03, was also employed.

Finally, the Atoms in Molecules Theory (AIM, 64) was also employed. With AIM, it is possible to define structure and stability in terms of the gradient field vector, $\nabla\rho$, associated with a scalar field, such as ρ . Thus, critical points, where $\nabla\rho = 0$, can be deduced. There is a set of trajectories associated with each critical point that define an interatomic surface, which separates the basins of neighboring atoms. Bond, ring and cage critical points were deduced for the supramolecule at each level of theory, using the AIM-PAC software (65).

The comparison of the obtained results for the different levels of theory indeed establish the importance of considering the dispersion energy for evaluating the complexation energies involved in aromatic-sugar interactions. The best results in terms of obtaining a well-defined potential energy curve and a geometry in agreement with X-ray experimental data was obtained when the MP2/6-31G(d, p) corrected by the counterpoise (CP) method C was employed. The obtained geometry is outlined in Fig. 5, which also shows the outcome of the AIM analysis of the complex, with the bond, ring, and cage critical points. In this case, the interaction energy of fucose-benzene is about 3.0 kcal/mol, somewhat higher than that established for single carbohydrate-aromatic interactions, for instance on the hevein complexes (about 1.5 kcal/mol). However, entropy factors and the role of water may influence the actual value for the aromatic-sugar interaction. Nevertheless, the calculated energy is of the same order of magnitude. These values show the efficiency of the CH/ π interaction in the stabilization process of the sugar-aromatic supramolecule. Since three CH/ π interactions participate in the formation of the complex, it can be estimated that the stabilization produced by each interaction is of ~ 1 kcal/mol. Moreover, the results obtained herein indicate that the carbohydrate-aromatic interactions are stabilizing interactions with an important dispersive component.

Indeed, an intramolecular carbohydrate-aromatic interaction has been recently invoked to provide the stabilizing energy that defines the observed conformation of a phenyllactic-containing GM1 glycomimetic (Fig. 6).

The observed intramolecular benzene-GalNAc stacking provides a preorganized conformation to efficiently interact with the lectin B-chain moiety of cholera toxin (CTB, 8), as deduced by TR-NOE experiments. The replacement of the phenyl ring for a cyclohexane moiety provokes a drastic conformational rearrangement in which the cyclohexane ring points away from the GalNAc ring. Thus, the requirement for a phenyl ring to make the proper interactions with the sugar moiety was further demonstrated (8).

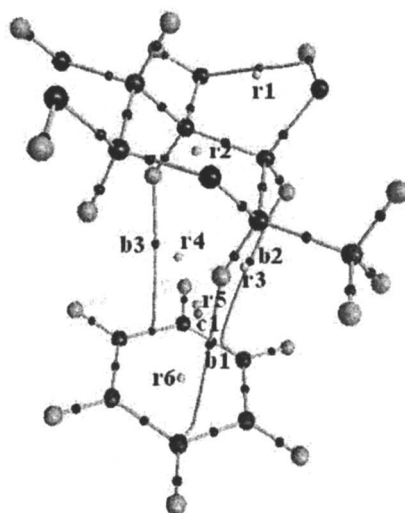


Figure 5. The geometry of the benzene-sugar complex at the MP2/6-31G(d, p) level corrected by the counter poise (CP) method.

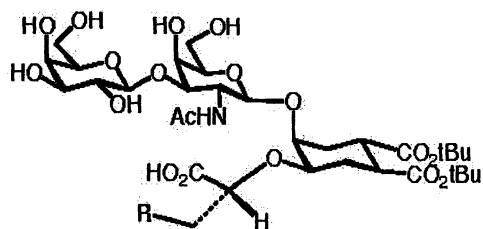


Figure 6. This GM1 glycomimetic interacts with CTB with micromolar affinity. The R is a phenyl moiety that stacks with the GalNAc moiety to give a well defined conformation (8).

The interaction of lactose with *E. coli* β -galactosidase

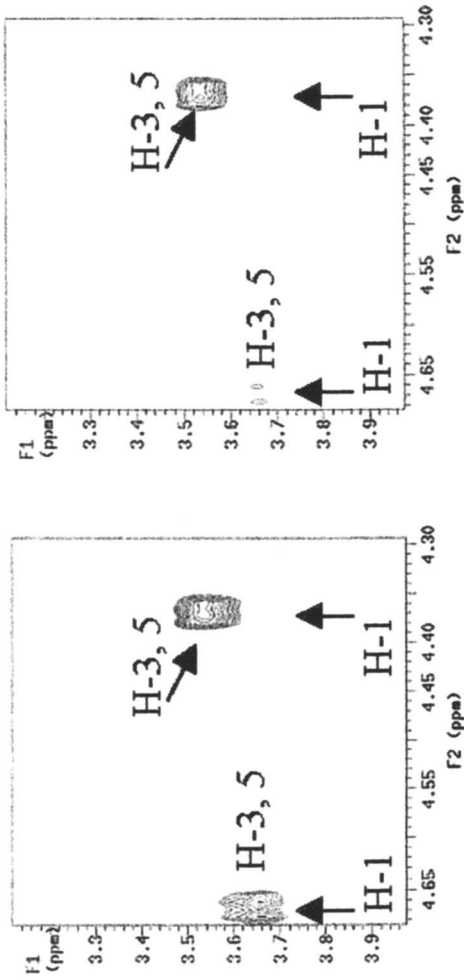
It is generally considered that the best inhibition of the processes catalyzed by enzymes is obtained when compounds that resemble the transition-state of the catalytic process are used (66). This fact is probably due to the energy gain provided by the binding of a high-energy conformer in the pathway towards the transition state, and in the particular case of glycosidase enzymes, molecules with oxocarbenium-like shape would have the best chance to give the best results (67). In this context, the knowledge of the enzyme-bound conformation of potential inhibitors is of paramount importance. Obviously, the deduction of the molecular features of an enzyme/substrate complex is a difficult task, since the substrate is readily transformed into products. In principle, two ways to deduce such properties for a given molecular complex can be envisaged: studying complexes of either wild type (WT) enzyme/inhibitor (68) or inactive mutated enzyme/substrate (69-73). We have used both methods to study the *E. coli* β -galactosidase bound conformation of lactose and several of its derivatives, namely, 2'-deoxy lactose, allolactose, with a (β 1 \rightarrow 6)- linkage, as well as their C- and S-glycosyl analogues. For the study of the interaction of the three O-glycosides, an inactive enzyme, for which the nucleophilic aminoacid of *E. coli* β -galactosidase, glutamic acid 537, has been mutated to glutamine, E537Q, was used. For the recognition of the C- and S-glycosyl analogues, the wild type enzyme has been employed. In the particular case discussed here, NMR and *ab initio* results showed that the 3D-shapes of the substrate (lactose, 2'-deoxy lactose and allolactose), or inhibitor (C- and S-glycosyl compounds) within the enzyme binding site depend on the chemical nature of the molecules. In fact, they depend on the relative size of the stereoelectronic barriers for chair deformation or for rotation around Φ glycosidic linkage. TR-NOE experiments (44) permitted to deduce that for both C- and S-lactose, their high energy anti- Φ /syn- Ψ conformers (with $\Delta E=1.8$ or 1.5 kcal/mol with respect to the global minimum for the C- and S-glycosyl compounds, respectively) are bound by this enzyme. All the observed transferred NOEs are in agreement with the exclusive recognition of this conformer (68, 73) for both glycomimetics.

In principle, these results for the C- and S-glycomimetics cannot be directly extrapolated to O-lactose. Due to the presence of the exo-anomeric effect in the natural sugar (74), the energy barriers around Φ and the relative energies of the energy minima will have different values to those of C- and S-analogues. Thus, we decided to investigate the conformational features of the actual enzyme substrates, the three O-glycosides mentioned above, bound to an inactive enzyme, E537Q. As a first test, the study of the TR-NOE NMR spectra recorded for the complexes of C- and S-lactose to E537Q provided identical results to

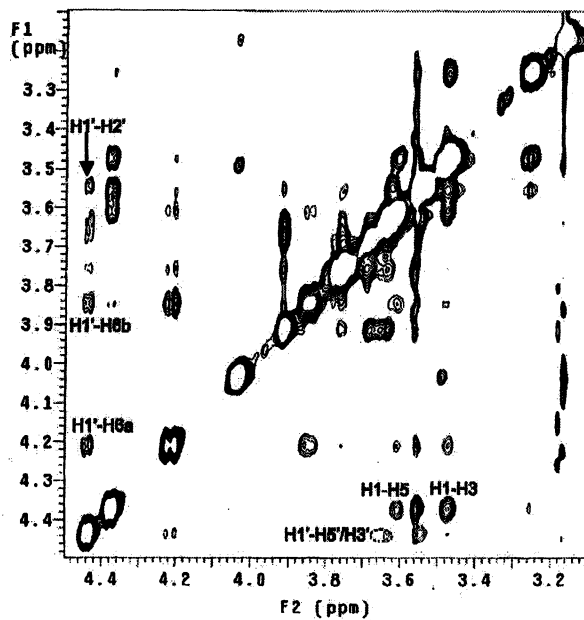
those for the WT enzyme: the anti- Φ /syn- Ψ conformation of both analogues is the only one that is bound by the mutated enzyme. Different results were obtained for the binding of the O-glycosides to E537Q. For lactose, 2'-deoxy lactose and allolactose, and in opposition to the observations for the complexes of C- and S-lactose with the WT and E537Q enzymes, the H-1'/H-3' and H-1'/H-5' cross peaks for the galactose moieties were very weak, much weaker than the analogous cross peaks (H-1/H-3 and H-1/H-5) for the neighboring Glc rings (Fig. 7). Thus, the bound galactose moieties of the lactose O-glycosides do not hold 1C_4 chair conformations.

Apart of further demonstrating that NMR may be used to get insights on the bound conformation on substrates/inhibitors at glycosidase catalytic sites, these experimental results indicate that *E. coli* β -galactosidase (WT or E537Q) selects a high-energy conformer of the natural substrate or the C-, S-glycomimetic. This conformational distortion takes place before the catalytic reaction has started. For C- and S-lactose, (with no exo-anomeric effect), the high-energy conformer is produced upon rotation around the glycosidic Φ -angle. In contrast, for O-glycosides, the enzyme produces distortion of the chair. Nevertheless, the three dimensional shapes of both types of conformers is fairly similar, with the aglycon adopting almost the same spatial orientation with respect to the glycon Gal moiety, and thus the observed bound conformations of C- and S-lactose are reminiscent of the deformations in the reaction pathway. Modeling studies showed that both types of conformers may be bound without significant distortion of the enzyme binding site. Interestingly, only the Gal rings of O-glycosides are bound in a conformation, somehow resembling the oxo-carbonium transition state of glycosidase-mediated hydrolysis.

It has been reported (75) that the rotation from syn to anti- Φ of 2-methoxy pyrane requires a energy cost (HF/6-31G*) of 2.9 kcal/mol, while that for the 2-ethyl (C-glycosyl analogue and 2-thiomethyl (S-glycosyl analogue) analogues are only 1.2-1.3 kcal/mol, respectively. B3LYP/6-31G* calculations carried out for the three O-, C-, and S-models indicated that the required energies to reach a flattening consistent with the NMR data) are higher (ca. 5.0 kcal/mol). Thus, at this level of theory, the cost of the deformation is more than 2.0 kcal/mol higher than that required to rotate around Φ , even considering the exo-anomeric contribution (76). However, one additional point has to be taken into consideration. For retaining glycosidase enzymes, such as *E. coli* β -galactosidase, and on the way towards the oxocarbenium-like transition state, one of the proposed mechanisms involves protonation of the glycosidic oxygen, prior to the nucleophilic attack. Indeed, theoretical studies on model compounds have shown that interglycosidic oxygen protonation facilitates precisely the



*Figure 7. Portion of the Tr-NOE spectrum measured for the β -methyl glycoside of 2'-deoxy lactose in the presence of *E. coli* β -galactosidase. The different intensities of the key cross peaks for the Gal and Glc rings are evident at both intensity levels. They indicate that the Gal moiety is not a chair conformer.*



*Figure 8. Tr-NOE spectrum measured for the β -methyl glycoside of allolactose in the presence of *E. coli* β -galactosidase. The different intensity of the key cross peaks for the Gal and Glc rings are evident. They indicate that the Gal moiety does not show a chair conformation.*

flattening of the chair (77). Therefore, when *ab initio* calculations (at the B3LYP/6-31G* level) were carried out on the model compounds, but with protonation of either lone pair at the interglycosidic atom, the relative energy costs for deformation of the protonated models were significantly smaller than for 2-ethyl pyrane, for which no protonation is possible. Now, it was 1.5 kcal/mol smaller than the energy required for rotating around Φ . In contrast, for the C- and S-glycosyl analogues, the required energies for flattening were higher than those mandatory for Φ -rotation.

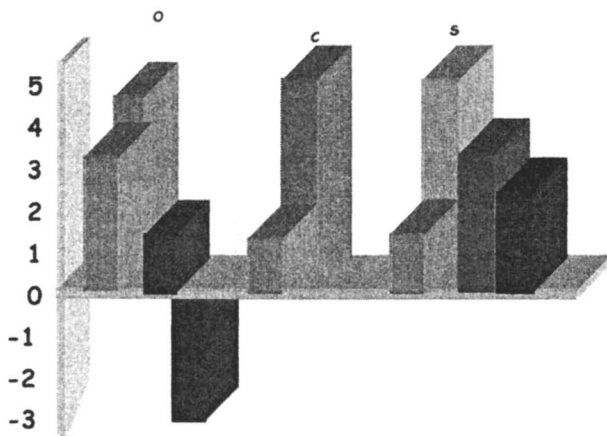


Figure 9. The required energies for rotating around F or for deforming the chair for O- (left), C- (middle), and S-glycosides (right). The green bar represents the required energy for rotation around Φ . The highest bars in all cases around 5 kcal/mol represent the required energies for deformation of the chair to the observed bound conformation. Protonation of the exo-cyclic oxygen lone pairs (additional two bars at the right handside of the O- and S-glycosides) strongly decreases the required energy for chair deformation, especially for the O-compound.

(See page 2 in color insert in this chapter.)

Therefore, at this level of theory, rotation around Φ is favored for C- and S-glycosides, while deformation of the chair is favored for O-glycosides. These results agree with the NMR experimental observations, and permit to explain, on stereoelectronic grounds, the distinct experimentally observed conformational selection of O-, C-, and S-glycosyl compounds by *E. coli* β -galactosidase, which depend on the chemical nature of the observed molecule.

Thus, for glycosides and glycomimetics, distortion takes place upon recognition by *E. coli* β -galactosidase. For C- and S-glycosides, with absent or reduced exo-anomeric effect, the distortion is easily accomplished by rotation

around Φ . However, for O-glycosides, with a strong exo-anomeric effect, and due to the higher stereoelectronic barrier for rotation around Φ , the energy required for such rotation is much higher than that necessary for the C- and S-analogues. Therefore, the enzyme distorts the chair. Although the enzyme binding site has been designed to accomplish such a distortion, there is no need for that in the case of glycomimetics. In these cases, the less energy demanding rotation around Φ provides a similar geometry of the glycomimetics in the enzyme catalytic site.

At this point, it seemed also relevant to study the associated energy for the complete interconversion pathway of oxane, the simplest pyranoid ring, as a previous step to analyze the energy requirements for deformation and inversion of carbohydrate six-membered rings (17). The theoretical calculations were performed at different levels of theory, namely, the MPW1K/6-311++G(2d,2p) level, using the modified Perdew-Wang 1 parameter functional (78), the B3LYP/6-311++G(2d,2p) level (79), the MP2/6-31G(d,p) level (80), and the Gaussian 2 Theory (G2, 81). Since the two extreme chairs that are interconverted are isoenergetic, the representation of the process is fairly simple. There are only two transition states associated with the inversion, that have different energy, in contrast to cyclohexane, where there are six isoenergetic states. Depending on the level of theory, the chair is between 10.1 and 10.6 kcal/mol under both transition states, in agreement with the experimental values of $\Delta G = 10.3$ kcal/mol at 212K (82). After these transition states, the boats can be reached, which are at 4.8 and 4.4 kcal/mol from the corresponding transition states, respectively. The corresponding process for thiane has also been evaluated, and shows strong similarities. Thus, on the potential energy surface of these compounds, there are two conformational processes that are independent, but somehow interconnected: the inversion process that allows the annular inversion and the low energy topo-isomerization process, that allows conformational exchange between the skew boats. Although these energies are only approximated in relation to a galactose chair, they represent the first step towards the quantitation of the energy that a given enzyme has to provide for attaining deformation of the corresponding pyranoid ring.

Conclusions

The combination of NMR experimental data and theoretical calculations allow to understand and to establish on chemical grounds some of the geometrical and structural requirements for molecular recognition of carbohydrates by protein receptors. Further exploration of other systems are currently underway following similar approaches.

Acknowledgments

We are grateful to Ministry of Education and Science of (Spain) via grant BQU2003-03550-C03-01. We also acknowledge the CSIC and UNAM-México for the allowance of an exchange programme between our two laboratories. We also acknowledge the discussions and collaborations with Drs. A. Bernardi (Milano), M. Muraki (Tsukuba), D. Andreu (Barcelona), H.-J. Gabius and H.-C. Siebert (München), and G. Asensio (Valencia).

References

- (1) Gabius, H.-J.; Siebert, H.-C.; Andre, S.; Jimenez-Barbero, J.; Rudiger, H. *ChemBioChem* 2004, 5, 740-764
- (2) Lis, H.; Sharon, N. *Chem. Rev.*, 1998, 98, 637-674.
- (3) Ratner, D. M.; Adams, E. W.; Disney, M. D.; Seeberger, P. H. *ChemBioChem*, 2004, 5, 1375-1383.
- (4) Bernardi, A.; Potenza, D.; Capelli, A. M.; Garcia-Herrero, A.; Cañada, J.; Jimenez-Barbero, J. *Chem. Eur. J.*, 2002, 8, 4598-4612.
- (5) Jeyaprakash, A. A.; Srivastav, A.; Surolia, A.; Vijayan, M., *J. Mol. Biol.*, 2004, 338, 757-770.
- (6) Rüdiger, H.; Siebert, H.-C.; Solís, D.; Jiménez-Barbero, J.; Romero, A.; von der Lieth, C. -W.; Diaz-Mauriño, T.; Gabius, H.-J. *Curr. Med. Chem.*, 2000, 7, 389-416.
- (7) Lemieux, R. U. *Chem. Soc. Rev.*, 1989, 18, 347-374.
- (8) Quioco, F. A. *Pure Appl. Chem.*, 1989, 61, 1293-1306.
- (8) Bernardi, A.; Arosio, D.; Potenza, D.; Sanchez-Medina, I.; Mari, S.; Cañada, F. J.; Jimenez-Barbero, J. *Chem. Eur. J.*, 2004, 10, 4395-4405.
- (9) Lemieux, R. U. *Acc. Chem. Res.* 1996, 29, 373-380.
- (10) For latest results, see Spiwok, V.; Lipovova, P.; Skalova, T.; Buchtelova, E.; Hasek, J.; Kralova, B. *Carbohydr Res.*, 2004, 339, 2275-2280.
- (11) See, for instance Biot, C.; Wintjens, R.; Rooman, M. *J. Am. Chem Soc.*, 2004, 126, 6220-6221.
- (12) Dougherty, D. A. *Science* 1996, 271, 163-166
- (13) Hunter, C. A. *Angew Chem Int Ed Engl.* 2004, 43, 5310-5324.
- (14) Heightam, T. D.; Vasella, A. T. *Angew. Chem. Int. Ed.*, 1999, 38, 750.
- (15) Legler, G. *Adv. Carbohydr. Chem. Biochem.* 1990, 48, 319.
- (16) Sinnott, M. L. *Chem. Rev.* 1990, 90, 1171
- (17) Fernández Alonso, C.; Asensio, J. L.; Cañada, F. J.; Jiménez-Barbero, J.; Cuevas, G. *ChemPhysChem*, 2003, 4, 748-753.

- (18a) For a review, see Rüdiger, H.; Gabius, H.-J. *Glycoconjugate J.*, 2001, 18, 589-613; for recent applications of chitin, see, for instance, (18b) Kato, Y.; Onishi, H.; Machida, Y. *Curr. Pharm. Biotechnol.*, 2003, 4, 303-309.
- (19) Drenth, J.; Low, B. W.; Richardson, J. S.; Wright, C. S. *J. Biol. Chem.*, 1980, 255, 2652-2655.
- (20) Soedjanaatmadja, U. M.; Hofsteenge, J.; Jeronimus-Stratingh, C. M.; Bruins, A. P.; Beintema, J. J. *Biochim Biophys Acta.*, 1994, 1209, 144-148.
- (21) Beintema, J. J.; Peumans, W. J. *FEBS Lett.*, 1992, 299, 131-134.
- (22) Wright, H. T.; Brooks, D. M.; Wright, C. S. *J. Mol. Evol.*, 1984-85, 21, 133-138.
- (23) Andersen, N. H.; Cao, B.; Rodriguez-Romero, A.; Arreguin, B. *Biochemistry*, 1993, 32, 1407-1422.
- (24) Asensio, J. L.; Cañada, F. J.; Bruix, M. Rodriguez-Romero, ; A.; Jiménez-Barbero, J. *Eur. J. Biochem.*, 1995, 230, 621-633.
- (25) Asensio, J. L.; Cañada, F. J.; Bruix, M.; Gonzalez, C.; Khier, N.; Rodriguez-Romero, A.; J. Jiménez-Barbero, J. *Glycobiology*, 1998, 8, 569-577.
- (26) Siebert, H.-C.; André, S.; Asensio, J. L.; Cañada, F. J.; Dong, X.; Espinosa, J. F.; Frank, M.; Gilleron, M.; Kaltner, H.; Kozar, T.; Bovin, N. V.; von der Lieth, C. W.; Vliegthart, J. F. G.; Jiménez-Barbero, J.; Gabius, H.-J. *ChemBioChem*, 2000, 1, 181-195.
- (27) Asensio, J. L.; Cañada, F. J.; Siebert, H.-C.; Laynez, J.; Poveda, A.; Nieto, P. M.; Soedjanaatmadja, U. M.; Gabius, H.-J.; Jiménez-Barbero, J. *Chem. Biol.*, 2000, 7, 529-543.
- (28) Huang, R. H.; Xiang, Y.; Tu, G. Z.; Zhang, Y.; Wang, D. C. *Biochemistry*, 2004, 43, 6005-6012.
- (29) Van den Bergh, K. P.; Proost, P.; Van Damme, J.; Coosemans, J.; Van Damme, E. J.; Peumans, W. J. *FEBS Lett.*, 2002, 530, 181-185.
- (30) Aboitiz, N.; Vila-Perello, M.; Groves, P.; Asensio, J. L.; Andreu, D.; Cañada, F. J.; Jiménez-Barbero, J. *ChemBioChem*, 2004, 5, 1245-1255.
- (31) Asensio, J. L.; Siebert, H.-C.; von der Lieth, C.-W.; Laynez, J.; Bruix, M.; Soedjanaatmadja, U. M.; Beintema, J. J.; Cañada, F. J.; Gabius, H. J.; Jimenez-Barbero, J. *Proteins* 2000, 40, 218-236.
- (32) Espinosa, J. F.; Asensio, J. L.; Garcia, J. L.; Laynez, J.; Bruix, M.; Wright, C.; Siebert, H. C.; Gabius, H.-J.; Cañada, F. J.; Jiménez-Barbero, J. *Eur. J. Biochem.*, 2000, 267, 3965-3978.
- (33) Martins, J. C.; Maes, D.; Loris, R.; Pepermans, H. A.; Wyns, L.; Willem, R.; Verheyden, P. *J. Mol. Biol.* 1996, 258, 322-333.
- (34) Chavez, M. I.; Vidal, P.; Aboitiz, N.; Freire, F.; Andreu, C.; Asensio, G.; Muraki, M.; Asensio, J. L.; Cañada, F. J.; Jiménez-Barbero, J., in preparation.
- (35) Wright, C. S.; Kellogg, G. E. *Protein Sci.*, 1996, 5, 1466-1476.
- (36) Harata, K.; Muraki, M. *J. Mol. Biol.*, 2000, 297, 673-681.

- (37) Saul, F. A.; Rovira, P.; Boulot, G.; Damme, E. J.; Peumans, W. J.; Truffa-Bachi, P.; Bentley, G. A. *Structure Fold. Des.*, 2000, 8, 593-603.
- (38) Harata, K.; Schubert, W. D.; Muraki, M. *Acta Crystallogr. D*, 2001, 57, 1513-1517.
- (39) Rodriguez-Romero, A.; Ravichandran, K. G.; Soriano-Garcia, M. *FEBS Lett.*, 1991, 291, 307-309.
- (40) Reyes-Lopez, C. A.; Hernandez-Santoyo, A.; Pedraza-Escalona, M.; Mendoza, G.; Hernandez-Arana, A.; Rodriguez-Romero, A. *Biochem. Biophys. Res. Commun.*, 2004, 314, 123-130.
- (41) See, for instance, Jahnke, W.; Florsheimer, A.; Blommers, M. J. J.; Paris, C. G.; Nalin, C. M.; Pérez, L. B. *Curr. Topics Med. Chem.*, 2003, 3, 69-80.
- (42) Fielding, L. *Curr. Topics Med. Chem.*, 2003, 3, 39-53.
- (43) Stilbs, P. *Anal. Chem.*, 1981, 53, 2135-2137.
- (44) For applications in the carbohydrate field, see, for instance, *NMR spectroscopy of glycoconjugates*, Jiménez-Barbero, J. & Peters, T. (Eds), Wiley-VCH, 2002, Weinheim.
- (45) Siebert, H.-C.; Adar, R.; Arango, R.; Burchert, M.; Kaltner, H.; Kayser, G.; Tajkhorshid, E.; von der Lieth, C.-W.; Kaptein, R.; Sharon, N.; Vliegthart, J. F. G.; Gabius, H. J. *Eur. J. Biochem.*, 1997, 249, 27-38.
- (46) Hernandez-Arana, A.; Rojo-Dominguez, A.; Soriano-Garcia, M.; Rodriguez-Romero, A. *Eur. J. Biochem.*, 1995, 228, 649-652.
- (47) Garcia-Hernandez, E.; Zubillaga, R. A.; Rojo-Dominguez, A.; Rodriguez-Romero, A.; Hernandez-Arana, A. *Proteins*, 1997, 29, 467-477.
- (48) Lee, R. T.; Gabius, H.-J.; Lee, Y. C. *Glycoconjugate J.* 1998, 15, 649-655.
- (49) Bonnin, S.; Besson, F.; Gelhausen, M.; Chierici, S.; Roux, B. *FEBS Lett.* 1999, 456, 361-364.
- (50) He, L.; André, S.; Siebert, H.-C.; Helmholz, H.; Niemeyer, B.; Gabius, H.-J. *J. Biophys. J.* 2003, 85, 511-524.
- (51) Imberty, A.; Pérez, S. *Chem. Rev.*, 2000, 100, 4567-4588.
- (52) Wright, C. S.; Jaeger, J. *J. Mol. Biol.*, 1993, 232, 620-638.
- (53) Wright, C. S. *J. Biol. Chem.*, 1992, 267, 14345-14352.
- (54) Siebert, H.-C.; von der Lieth, C.-W.; Kaptein, R.; Beintema, J. J.; Dijkstra, K.; van Nuland, N.; Soedjanaatmadja, U. M.; Rice, A.; Vliegthart, J. F. G.; Wright, C. S.; Gabius, H.-J. *Proteins*, 1997, 28, 268-284.
- (55) Aboitiz, N.; Cañada, F. J.; Husakova, L.; Kuzma, M.; Kren, V.; Jiménez-Barbero, J. *Org. Biomol. Chem.*, 2004, 2, 1987-1994.
- (56) Colombo, G.; Meli, M.; Cañada, F. J.; Asensio, J. L.; Jiménez-Barbero, J. *Carbohydr. Res.*, 2004, 339, 985-994.
- (57) Groves, P.; Rasmussen, M.; Molero, D.; Samain, E.; Cañada, F. J.; Driguez, H.; Jiménez-Barbero, J. *Glycobiology*, 2004, 14, 451-456.
- (58) Muraki, M.; Morii, H.; Harata K., *Protein Eng.*, 2000, 13, 385-389.
- (59) Gaussian 98 (Revision A.11.3) Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman J.R.; Zakrzewski, V.G.;

- Montgomery, J.A.; Stratmann, R.E.; Burant, J.C.; Dapprich, S.; Millam, J.M.; Daniels, A.D.; Kudin, K.N.; Strain, M.C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G.A.; Ayala, P.Y.; Cui, Q.; Morokuma, K.; Malick, D.K.; Rabuck, A.D.; Raghavachari, K.; Foresman, J.B.; Cioslowski, J.; Ortiz, J.V.; Stefanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R.L.; Fox, D.J.; Keith, T.; Al-Laham, M.A.; Peng, C.Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P.M.W.; Johnson, B.G.; Chen, W.; Wong, M.W.; Andres, J.L.; Gead-Gordon, M.; Replogle, E.S. Pople, J.A. Gaussian, Inc., Pittsburgh, PA, 1998
- (60) Gaussian 03, (Revision A.1), Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Pittsburgh PA, 2003
- (61) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.*, 1996, *105*, 11024-11031.
- (62) Boys, S. F.; Bernardi, F. *Mol. Phys.*, 1970, *19*, 553.
- (63) Sotiris S.; Xantheas, *J. Chem. Phys.*, 1996, *104*, 8821-8824.
- (64) Bader, R. F. W. *Atoms in Molecules in Encyclopedia of computational Chemistry*, Schleyer, P.v.R. editor, 1998, John Wiley & Sons, p. 64.
- (65) Biegler-Koing, F. W.; Bader, R. F. W.; Tang, T. H. *J. Comput. Chem.*, 1982, *3*, 317-328.
- (66) Neufeld, E. F. *Annu. Rev. Biochem.* 1991, *60*, 257.
- (67) von Itzstein, M.; Wu, w. Y.; Kok, G. B.; Pegg, M. S.; Dyanson, J. C.; Jin, B. P. T. V.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. *Nature* 1993, *363*, 1332
- (68) Espinosa, J. F.; Montero, E.; Vian, A.; Garcia, J. L.; Dietrich, H.; Schmidt, R. R.; Martin-Lomas, M.; Imberty, A.; Cañada, F. J.; Jimenez-Barbero, J. *J. Am. Chem. Soc.* 1998, *120*, 1309.

- (69) Mitchell, E. P.; Withers, S. G.; Ermert, P.; Vasella, A. T.; Garman, E. F. Oikonomakos, N. G.; Johnson, L. N. *Biochemistry* 1996, 35, 7341.
- (70) Tews, I. ; Perrakis, A.; Oppenheim, A.; Dauter, Z.; Wilson, K. S.; Vorgias, C.E. *Nature Struct. Biol.* 1996, 3, 638.
- (71) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schülein M.; Davies, G. J. *Biochemistry* 1996, 35, 15280.
- (72) Zou, J.; Kleywegt, G. J.; Stahlberg, J.; Driguez, H.; Nerinckx, W.; Claeysens, M.; Koivula, A.; Teeri, T. T.; Jones, T. A. *Structure Fold. Des.* 1999, 7, 1035.
- (73) García-Herrero, A.; Montero, E.; Muñoz, J. L.; Espinosa, J. F.; Vián, A.; García, J. L.; Asensio, J. L.; Cañada, J.; Jiménez-Barbero, J. *J. Am. Chem. Soc.*, 2002, 124, 4804-4810
- (74) Lemieux, R. U. ; Koto, S. ; Voisin, D. *Am. Chem. Soc. Symp. Ser.* 1979, 87, 17
- (75) Tvaroska, I.; Carver, J. P. *J. Phys. Chem.* 1996, 100, 11305.
- (76) The stereoelectronic component of the exo-anomeric effect has been experimentally deduced in water solution: (a) Asensio, J. L.; García-Herrero, A.; Murillo, M. T.; Fernandez-Mayoralas, A.; Cañada, F. J.; Johnson, C. R.; Jiménez-Barbero, J. *J. Am. Chem. Soc.* 1999, 121, 11318. (b) Asensio, J. L.; Cañada, F. J.; Kahn, N.; Cheng, X.; Mootoo, D. A.; Jiménez-Barbero, J. *Chem. Eur. J.* 2000, 6, 1035.
- (77) Andrews, C. W. ; Fraser-Reid, B. ; Bowen, J. P. *J. Am. Chem. Soc.* 1991, 113, 8293
- (78) Lynch, B.J.; Fast, P.L.; Harris, M.; Truhlar, D.G. *J. Phys. Chem. A.* 2000, 104, 4811-4815.
- (79) Becke, A.D. *J. Chem. Phys.* 1993, 98 5648-5652.
- (80) Moller, C.; Plesset, M.S., *Phys. Rev.* 1934, 46, 618-622. Frish, M.J.; Head-Gordon, M.; Pople, J.A. *Chem. Phys. Lett.* 1990, 166, 281-289.
- (81) Curtiss, L.A.; Raghavachari, K.; Trucks, G.W.; Pople, J.A. *J. Chem. Phys.* 1991, 94, 7221-7230.
- (82) Lambert, J.B.; Mixon, C.E.; Johnson, D.H. *J. Am. Chem. Soc.* 1973, 95, 4634-4639.

Chapter 5

Laser Photo CIDNP Technique as a Versatile Tool for Structural Analysis of Inter- and Intramolecular Protein–Carbohydrate Interactions

Hans-Christian Siebert¹, Emad Tajkhorshid²,
Johannes F. G. Vliegthart³, Claus-Wilhelm von der Lieth⁴,
Sabine André¹, and Hans-Joachim Gabius¹

¹Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität, Veterinärstraße 13, 80539 München, Germany,

²Theoretical and Computational Biophysics Group, Beckman Institute, University of Illinois at Urbana-Champaign, 405 North Mathews, Urbana, IL 61801

³Bijvoet Center for Biomolecular Research, University of Utrecht, P.O. Box 80075, 3508 TB Utrecht, The Netherlands

⁴Zentrale Spektroskopie, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

Abstract

Crystallographic data on proteins are generally considered as definitive information on the structure. However, the problem remains that the defined structure is only valid for the crystallized protein, since crystal packing might favor distinct molecular arrangements. Also dynamic movements will be frozen at a certain position, and the often non-physiologic conditions used to induce/promote crystal growth may lead to favoring structures different from the population under physiological conditions. Thus it is helpful to team up

crystallography with other experimental approaches to monitor protein attributes in solution. For this purpose, we herein focus on a NMR-spectroscopical technique, *i.e.* the laser photochemically induced dynamic nuclear polarization (CIDNP) method. Its scope of application is to assess surface accessibilities of histidine (His), tyrosine (Tyr), and/ or tryptophan (Trp) residues in a protein. To illustrate the value of this technique, we demonstrate by CIDNP spectra of plant/animal lectins recorded in the absence and in the presence of their ligands, how to gain pertinent information about structural properties of the binding pocket in combination with modeling data. Next, glycoproteins are analyzed successfully with this method in order to delineate characteristic differences in the CIDNP spectra when the structure of the glycan chain is deliberately altered. Therefore, it is possible to address two major questions in glycosciences with the CIDNP method: the structural impact of a) oligosaccharide ligands in lectin-oligosaccharide complexes and b) covalently linked glycan chains present in glycoproteins.

Introduction

The concept of the sugar code provides the basis to assign functions to the glycan chains of cellular glycoconjugates (1-6). In the interplay with lectins they can act as sugar-encoded biochemical signals. For example, common substitutions as the bisecting GlcNAc residue have been delineated as modulators of lectin affinity, giving a functional meaning to substitutions by altering glycan structures (7-9). Basically, bisecting GlcNAc gives rise to a change in primary structure. On top of that there is a change in conformation and maybe in presentation of peripheral epitopes. Glycans also influence surface properties of the protein backbone to which they are attached. To gauge such an influence and later track down details, the precision of X-ray-derived spatial parameters of the protein in the crystal is exploited.

X-ray crystallographic data are used as starting structures of molecular dynamics (MD) - simulations. This process is especially helpful in the case of glycoproteins. Unfortunately, only a limited number of them could so far be crystallized including the glycan part. The inherent flexibility of glycans thus poses a problem that can be addressed by NMR spectroscopic methods in conjunction with molecular modeling (such as MD simulations). This approach is especially called for cases wherein the glycoprotein or the lectin-carbohydrate complex under study is beyond the current limitations for a complete NMR analysis (as is the case for the glycoprotein serum amyloid P component (SAP), *v. i.*) or only distinct structural properties are of interest to be examined. With focus on surface properties of a protein where internal mobility of both protein- and glycan-parts can well be of eminent physiological importance, using the laser photo CIDNP - technique has particular merit, because it enables to quantitate surface (dye - accessible) positioning of the CIDNP-sensitive amino acid residues. The generation of signals in this method depends on the surface accessibility of the ring systems of Tyr, Trp and His (10-19). The signals then are the experimental input to challenge modeling-derived predictions for structural aspects of a protein in solution.

The technical side of laser photo CIDNP

The laser light suited for laser photo CIDNP experiments is normally generated by a continuous-wave argon ion laser, which operates in the multi-line mode with principal emission wavelengths of 488.0 and 514.5 nm, close to the edge of the 450 nm absorption band of the used dye. We routinely direct the laser light to the NMR tube by an optical fibre. Mechanical chopping controlled by the spectrometer will prevent heating of the lectin-containing solution, precluding denaturation of the protein. The CIDNP characteristic radical reaction is initiated by the flavin mononucleotide as laser-responsive dye. The irradiation leads to the generation of protein-dye radical pairs by the surface-exposed Tyr-, Trp-, and His-ring systems. Nuclear spin-polarization is achieved from back-reactions of the radical pairs illustrated together with the NMR responses in Figures 1-3.

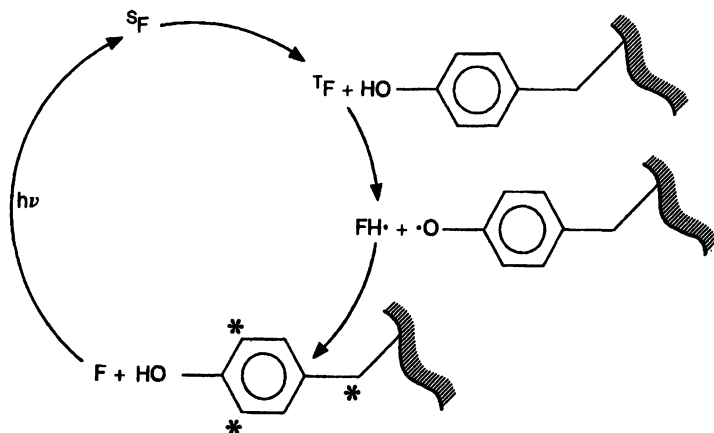


Figure 1. Cyclic reaction scheme for generation of laser photo CIDNP in tyrosine residues of a protein. The dye F (flavin) is irradiated by the light of an argon-ion laser and excited to the single state S^1F , which converts rapidly into the triplet state T^3F . This compound reacts with tyrosine's aromatic ring by transfer of a H-atom. The H-atom is returned in a recombination process of the radical pair followed by a polarization of the tyrosine moiety at the marked (*) positions (11).

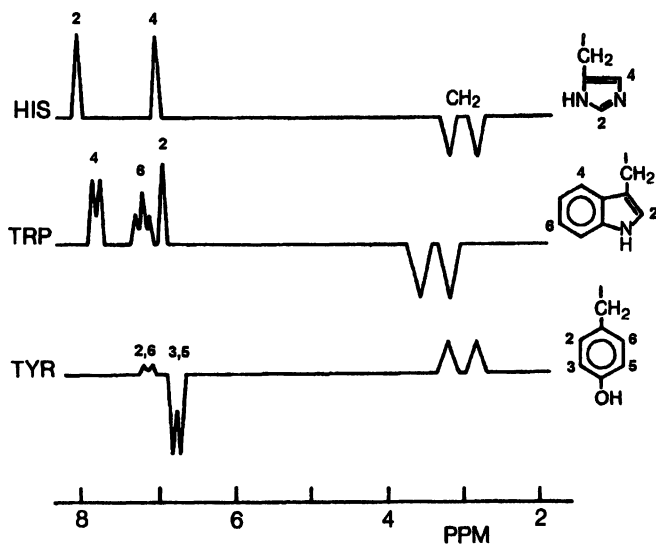


Figure 2. Overview of the laser photo CIDNP-effects for the reactive amino acids His, Trp and Tyr (theoretically generated spectra). The multiplet structure is given schematically (11).

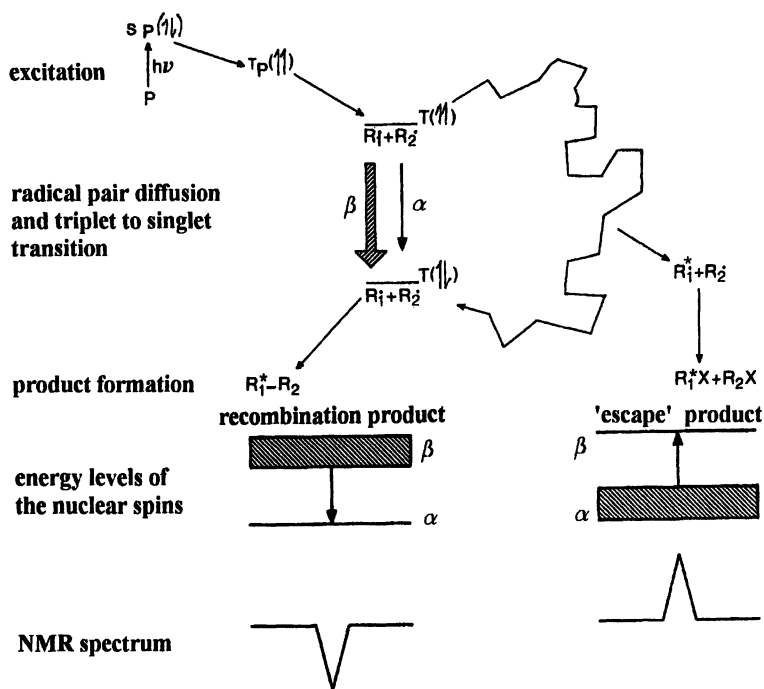


Figure 3. Radical pair mechanism establishing laser photo CIDNP. The scheme visualizes the generation of laser photo CIDNP (11).

Chemical shifts can be assessed relative to acetone (2.225 ppm) and HDO (4.76 ppm dependent on a defined temperature and pH/pD-value). In practice, two spectra are recorded in each laser photo CIDNP-experiment. The resulting light spectrum will be subtracted from the dark spectrum, thereby establishing the laser photo CIDNP difference spectrum only showing signals of polarized residues, as illustrated in Figure 2. This figure also presents the information on the involvement of protons in the dye-reactive amino acid groups: the tyrosine-dependent laser photo CIDNP effect represents the spin-density distribution of the intermediate phenoxy radical with strong negative signals of the 3, 5 protons and less intense positive signals for the 2, 6 protons. The CIDNP signals of tryptophan are generated by an intermediate radical with strong spin density at the 2, 4 and 6 protons, which invariably yield positive CIDNP signals, and the CIDNP responses of protons 2 and 4 of histidine occur as positive singlets (Figure 2). For further methodological details see the

references given above. We next turn to outline examples for practical application of this technique to answer questions in structural glycomics.

Applying the Laser Photo CIDNP Technique as Control for Computational Data

The development of algorithms to run homology-based modeling and the ease to implement such methods require techniques to validate computational results. Notably, the success of knowledge-based homology modeling is critically dependent on the predictive potency of the structure-based calculations, which attempt to translate homologous sequences into three-dimensional structures, as well as on the actual relevance of the crystal structure for the protein topology in solution. We illustrate an example for application of laser photo CIDNP using the crystal structure of the sialidase of *Salmonella typhimurium* as basis for model building of the homologue from *Clostridium perfringens* to derive a set of energy-minimized conformers (14). As parameter to be experimentally tested we calculate the surface presentations of Tyr, Trp and His residues. Tyr-residues occur especially abundantly on the surfaces of this enzyme, as shown in Figure 4.

Therefore, we put our data set obtained by *in silico* work to the experimental test. Although the number of reactive residues is in this case too large for an unequivocal assignment of individual residues to signals, the size and shape of the signals in the spectrum of *Clostridium perfringens* sialidase are in full agreement with their model-derived expectation. Next, we looked at single-site mutants of the enzyme. It was found that the effect of Tyr/Phe- and Cys/Ser-mutations on the surface accessibility of the CIDNP-reactive amino acid side chains in mutant forms of this sialidase are detectable by laser photo CIDNP methods (14). Thus, the effect of apparently rather conservative amino acids substitutions on a distinct conformational aspect of this protein even at distant sites should not be underestimated, as also documented for a plant lectin (20). These proteins home in on distinct carbohydrate epitopes to translate the message of the glycan code into cellular responses so that it is pertinent to document examples of respective applications (6, 21-24). Of note, these lectins are thus instrumental to map the glycomic profile of cells with functional relevance (24-27).

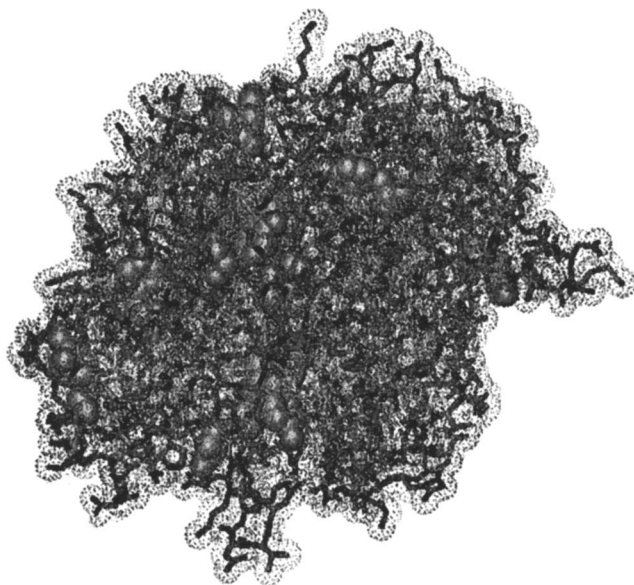


Figure 4. X-Ray structure of the sialidase of Salmonella typhimurium. Tyr.-residues at the surface are highlighted by the van der Waals representation. (See color page 1 in chapter 5.)

Applying the laser photo CIDNP technique in lectinology

Lectins from bacteria, plants and animals including mammals have been shown to harbor Tyr, Trp or His residues as essential parts of the carbohydrate-binding site, herein forming hydrophobic / CH- π electron contacts with their ligands in the binding pocket (28-31). As pointed out above the areas of surface accessibility for the laser photo CIDNP reactive Tyr, Trp and His residues can be calculated from energy-minimum conformations, as obtained by MD simulations of the lectin structure. The initial (start-) structures can routinely be obtained from the Brookhaven Protein Data Bank <http://www.rcsb.org>. Because area values above 60 Å² are generally assumed to be required for generation of a signal for Trp or Tyr, respective responses will be expected in the spectrum. We have performed these calculations for a member of the growth/ adhesion - regulatory family of the galectins, *i. e.* galectin-1, which acts as negative growth regulator on activated T-cells or neuroblastoma cells and as promotor of tissue invasion in glioblastoma (32-35). As compiled in Tables 1-3, the Trp indolyl ring is of special relevance, as it constitutes a key feature of the interaction site of galectins.

We looked, at the pH-dependence of the laser CIDNP-effect of Trp, Tyr and His. Because the distribution of CIDNP-inert (ligand residues) to CIDNP-sensitive (receptor) amino acids was favourable, we proceed to run the experiments to set CIDNP data in relation to results of modeling. Since the Trp68 residue contacts the B-face of the galactose unit by C-H/ π -interaction, we also could predict dependence of the signal by presence of the ligand. In addition to lactose we tested a complex (natural) ligand, *i. e.* the pentasaccharide of ganglioside GM₁. This interaction on neuroblastoma cell surface is the antiproliferative trigger (32). In line with the topology of the binding site, we indeed detected predicted signals and their alterations (Figure 5a, b, c).

Table 1. Areas of surface accessibility (SA) of laser photo CIDNP-reactive amino acid residues based on the X-ray structure of galectin-1 from bovine spleen (75) (monomer A, monomer B): a. without ligand and b. in the presence of the ligand

a.

| | SA [\AA^2] | SA [\AA^2] |
|---------|-----------------------|-----------------------|
| | A | B |
| His 45 | 30.4 | 31.8 |
| His 53 | 90.4 | 104.3 |
| Trp 69 | 97.3 | 99.4 |
| Tyr 104 | 69.4 | 25.3* |
| Tyr 119 | 36.1 | 34.4 |

b.

| | SA [\AA^2] | SA [\AA^2] |
|---------|-----------------------|-----------------------|
| | A | B |
| His 45 | 15.6 | 6.7 |
| His 53 | 92.2 | 91.8 |
| Trp 69 | 71.6 | 73.0 |
| Tyr 104 | 69.4 | 25.3* |
| Tyr 119 | 36.1 | 34.4 |

Dot density : 1; test sphere radius : 1.5 \AA ; gap in the structural data of the side chain of this residue (75)

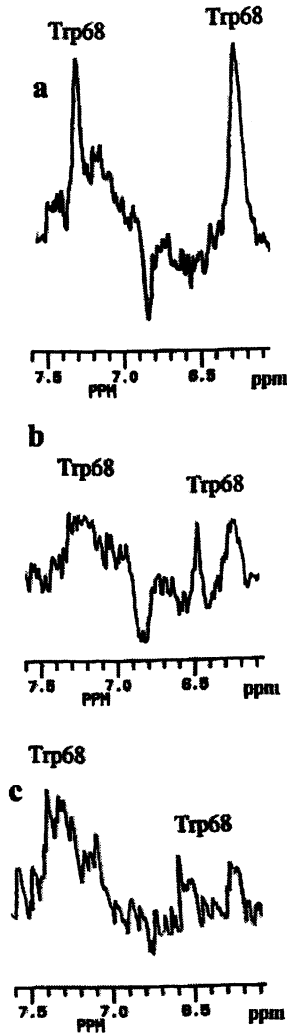
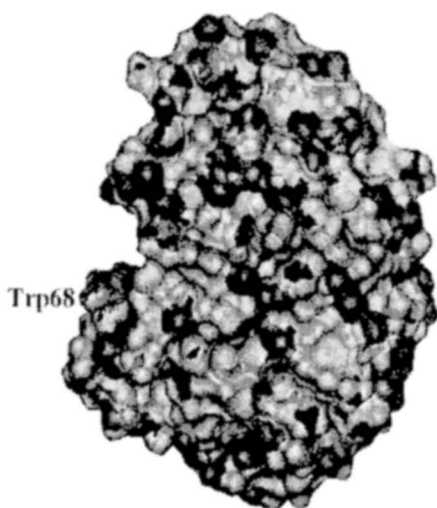


Figure 5. Laser photo-CIDNP difference spectra of human galectin-1 (aromatic part) in (a) the absence of a carbohydrate ligand showing the extent of dye accessibility of the indolyl ring of Trp68 side chain in the carbohydrate-binding site without ligand and (b) in the presence of 5 mM lactose to determine the effect of ligand binding on dye access to Trp68.

(c) Incubation of the lectin with ganglioside GM₁ presented in mixed micelles with dodecyl-phosphocholine illustrates the marked occupation of the binding site by this interaction. (d) Monomer of human galectin-1. (e) Monomer of human galectin-1 in the presence of the pentasaccharide chain of ganglioside GM₁.

d



e

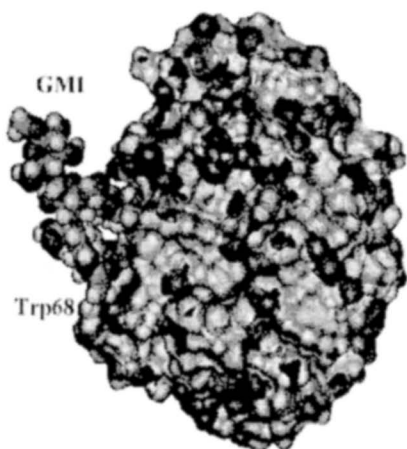


Figure 5. Continued. (See color page 2 in chapter 5.)

Table 2. Areas of surface accessibility (SA) of laser photo CIDNP-reactive amino acid residues based on the X-ray structure of galectin-1 from bovine heart (76) (monomer A, monomer B): a. without ligand and b. in the presence of the ligand

a.

| | SA [\AA^2] | SA [\AA^2] | SA [\AA^2] | SA [\AA^2] | SA [\AA^2] | SA [\AA^2] |
|---------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | A ₁ | B ₁ | A ₂ | C ₂ | A ₃ | B ₃ |
| His 45 | 35.3 | 32.0 | 32.2 | 26.1 | 23.1 | 29.4 |
| His 53 | 85.0 | 92.0 | 96.7 | 102.2 | 98.3 | 84.1 |
| Trp 69 | 99.6 | 96.0 | 105.4 | 100.7 | 108.1 | 102.1 |
| Tyr 104 | 67.7 | 80.2 | 57.8 | 80.5 | 90.5 | 85.8 |
| Tyr 119 | 48.8 | 45.1 | 34.2 | 37.8 | 44.7 | 49.9 |

b.

| | SA [\AA^2] | SA [\AA^2] | SA [\AA^2] | SA [\AA^2] | SA [\AA^2] | SA [\AA^2] |
|---------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | A ₁ | B ₁ | A ₂ | C ₂ | A ₃ | B ₃ |
| His 45 | 13.3 | 32.0 | 11.1 | 11.7 | 6.9 | 12.8 |
| His 53 | 69.2 | 92.0 | 78.4 | 90.9 | 72.0 | 67.2 |
| Trp 69 | 81.7 | 96.0 | 82.0 | 90.2 | 96.0 | 79.5 |
| Tyr 104 | 67.7 | 80.2 | 57.8 | 80.5 | 90.5 | 85.8 |
| Tyr 119 | 48.8 | 45.1 | 34.2 | 37.8 | 44.7 | 35.6 |

Dot density : 1; test sphere radius : 1.5 \AA ; the galectin has been crystallized and complexed with the ligand in three different crystal forms (hexagonal, trigonal, monoclinic): A₁B₁, A₂B₂C₂D₂, A₃B₃C₃D₃ (76)

Table 3. Molecular modeling-derived areas of surface accessibility (SA) of laser photo CIDNP-reactive amino acid residues of galectin-1 from human lung obtained from two different molecular dynamics procedures: (MD1 - tethered, no minimization; MD2 - relaxed, no minimization)

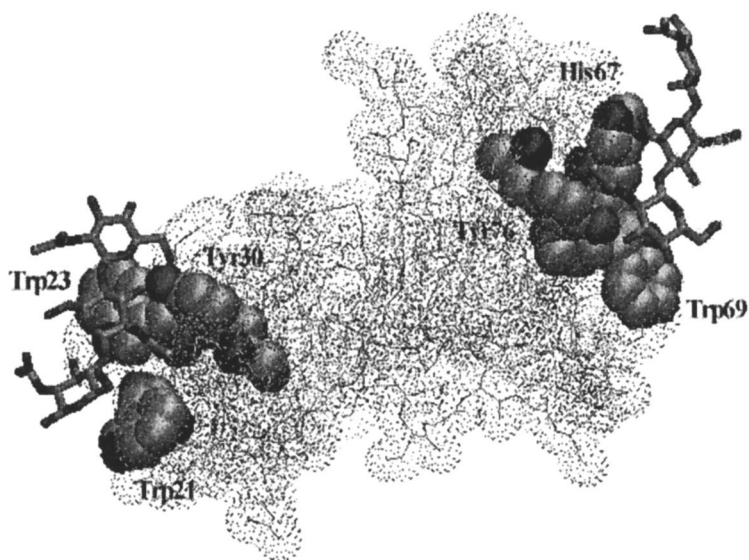
| | SA [\AA^2] | SD | SA [\AA^2] | SD |
|---------|-----------------------|------|-----------------------|------|
| | MD1 | | MD2 | |
| His 45 | 26.4 | 5.3 | 33.4 | 9.8 |
| His 53 | 113.4 | 6.8 | 130.4 | 10.9 |
| Trp 69 | 109.5 | 9.6 | 112.4 | 15.1 |
| Tyr 104 | 77.2 | 10.9 | 83.2 | 4.8 |
| Tyr 119 | 51.0 | 5.1 | 46.1 | 8.3 |

Dot density : 1; test sphere radius : 1.5 \AA ;
SD = standard deviation

The intense and sharp proton signals of Trp68, the only Trp present in galectin-1's sequence, in ligand-free galectin-1 arising from the free access of dye to this site, were markedly reduced in intensity in the presence of a specific ligand. Monitoring Trp signals when galectin-1 interacted with ganglioside GM₁-containing micelles produced definitive evidence for a physical interaction between the surface-presented pentasaccharide and galectin-1 (Figure 5d, e).

Furthermore, the rather similar alterations of signals by lactose and the more complex natural ligand indicated that the relative positioning of a galactose residue in the binding site, the key factor for the extent of dye access to Trp68, was not drastically altered. The CH- π stacking between the B-face of the hexopyranose ring of galactose and the indole of Trp68, characteristic for any galectin studied so far, appeared to be preserved in other cases too (not shown).

For the natural ligand detailed docking into the binding site was fully in agreement with the experimental data, highlighting the potential of this approach to control the reliability of the computational data (20, 36). The suitability of this method in detecting the proximity of a ligand to sensitive aromatic side chains has also been validated for hevein-domain-containing plant agglutinins (37). To refine the description of the binding site, close contacts to the protein can be exploited by saturation transfer difference or transferred nuclear Overhauser enhancement spectroscopy, as illustrated for the analysis of the positioning of the pentasaccharide into the binding site of human galectin-1 (36). The recruitment of hydroxyl groups as topological sensors deserve attention, because lectins can maintain their activity in the presence of aprotic solvents (38-40). Equally important, the monitoring of surface presentations of certain amino acids is also a sensitive tool to detect ligand-induced conformational changes. Galectin-1 is subject to such a change (20), later further analyzed by small angle neutron scattering (41). This phenomenon, to be reckoned with when considering crystallographic data, was also observed in another case in our studies: *Urtica dioica* agglutinin (UDA) from the stinging nettle which is a member of the chitin-binding family of plant lectins was subject to this alteration (37). It contains a His-residue instead of a Trp-residue in one of its two binding pockets (Figure 6).



*Figure 6. Model of the stinging nettle agglutinin (*Urtica dioica* agglutinin - UDA). The two binding sites are occupied by the potent ligand *N,N',N''*-triacetylchitotriose. (See color page 3 in chapter 5.)*

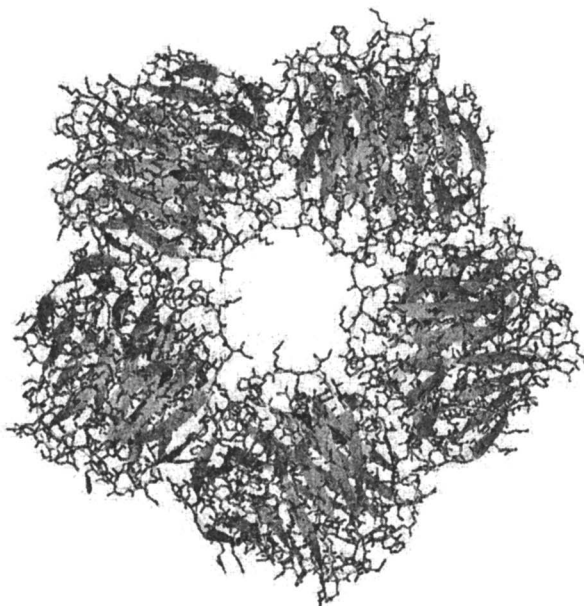
UDA consists of two hevein-like domains and has a single chain of 89 amino acids, comprising two binding sites with different affinities for oligomers of *N*-acetylglucosamine (GlcNAc). Besides its putative role in defense mechanisms in the plant (24), it is interesting to note that it acts as superantigen (42). Positional alignment of the UDA structure with other hevein-like lectins reveals that Trp21, Trp23 and Tyr30 as well as His67, Trp69 and Tyr76 are the functional aromatic amino acids in the binding pocket. Trp16, Trp40, His47 and Tyr84 are other CIDNP-reactive amino acid residues of UDA, which are located in exposed surface regions other than the GlcNAc-binding sites. Laser photo CIDNP spectra of hevein-, pseudohevein-GlcNAc complexes and the complex with GlcNAc containing a domain of wheat germ agglutinin confirm the expectation of a reduced signal intensity in relation to the CIDNP spectra of ligand-free

protein (37). In contrast, the laser photo CIDNP spectrum of UDA with its two overlapping Tyr signals is not strongly affected by ligand addition (37). This result is interpreted as evidence for a reduced extent of shielding of the binding-site residues of UDA by GlcNAc-oligomers in comparison to that observed for the three hevein-domain monomers (37).

Applying the Laser photo CIDNP Technique in Structural Glycoproteomics

Serum amyloid P component (SAP): As part of the systemic reaction to inflammation, significant changes in the synthetic output of the liver, termed the acute-phase response, result in a dramatical alteration of plasma concentrations of particular proteins. These proteins are termed APRs (acute-phase reactants) and in mammals include serum amyloid A protein, CRP (C-reactive protein) and SAP (serum amyloid P component). CRP is the major APR in humans, whereas SAP, which is the major APR in mice, shows only a slight increase in its plasma level. SAP is, however, generally associated with amyloid deposits. Both CRP and SAP are principal members of the pentraxin family of proteins, which are characterized by a planar disc arrangement of five non-covalently associated subunits. The physiological functions of these pentraxins involve Ca^{2+} -dependent ligand binding, and SAP has been shown to bind chromatin and DNA of apoptotic and necrotic cells, inhibiting the formation of antibodies against chromatin fragments, thus preventing antinuclear autoimmunity. Besides acting as scavenger, SAP also binds to bacteria, resulting in a strong anti-opsonic effect, which thereby results in enhanced virulence of the infectious agent (43-45). Structurally, the pentraxin SAP is arranged in crystals as a planar, non-covalently linked face-to-face dimer of two disc-shaped pentamers, constituting a decameric complex (49). Further details concerning SAP's protein and glycan parts have been described (49-52). The X-ray structure of SAP provides information about the protein part only (Figure 7).

The invariant complex-type diantennary glycan of each monomer at Asn32, located on a β -strand under the single α -helix on the protomer surface opposite of the Ca^{2+} - and ligand-binding sites (51) is not be



*Figure 7. X-Ray structure of the pentameric serum amyloid P component.
(See color page 3 in chapter 5.)*

described by the X-ray crystallographic data. For this reason we initiated a modeling study on the glycan's structure. To answer the question whether removal of the two sialic acid moieties from the single *N*-linked saccharide chain per protomer may cause a change in surface exposure of aromatic amino acids, laser photo CIDNP monitoring can - again - prove its suitability.

Due to its molecular weight the proton signals of the protein part of SAP are not resolved. However, it was possible to assign distinct CIDNP-signals due to their alteration after addition of sialidase (53). For ensuing analysis the glycan chain was completely removed with peptide- N^4 -(*N*-acetyl- β -glucosaminyl)asparagine amidase F (PNGase-F) from *Flavobacterium meningosepticum* (EC 3.5.1.52), (Figure 8).

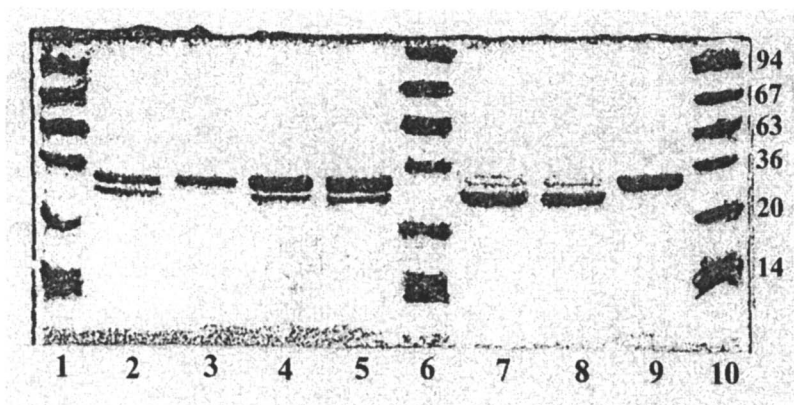


Figure 8. Analysis of molecular mass of SAP which had been enzymatically deglycosylated and subjected to mock treatment as control to assess extent of deglycosylation and the resulting shift in molecular mass.

As demonstrated in Figure 9a, b, the increase in intensity of the Tyr-CIDNP signals moving from the sialylated to desialylated form is already remarkable. An additional increase in Tyr-CIDNP signal intensity is observed, when comparing the desialylated with the completely deglycosylated form (Figure 9c).

In combination with molecular modeling data these laser photo CIDNP-derived results can structurally be interpreted as follows: a Tyr-CIDNP signal is already picked up, when analyzing the naturally, fully sialylated SAP. On removal of sialic acid residues, the surface accessibility of at least one Tyr residue is increased as a consequence. This is also the case, when the N-glycans are completely removed. A series of MD simulations on glycosylated, desialylated and deglycosylated SAP focused on dimers support the data of laser photo CIDNP experiments. The most suitable models for *in silico* work consist of a dimeric part from the pentameric SAP in which only one monomer bears a carbohydrate chain. In the case of the naturally glycosylated and the desialylated forms the carbohydrate chain is attached to Asn32 of only one monomer (Figure 10a). In this case the interaction of the glycan at monomer A with the surface of monomer B can be studied in an optimal way. The MD simulations were carried out in explicit water (Figure 10b).

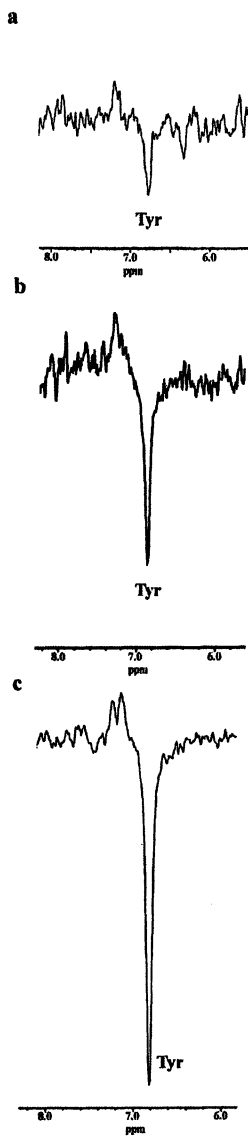


Figure 9. Aromatic parts of laser photo CIDNP-difference spectra of (a) natural SAP, (b) its desialylated and (c) its deglycosylated form.

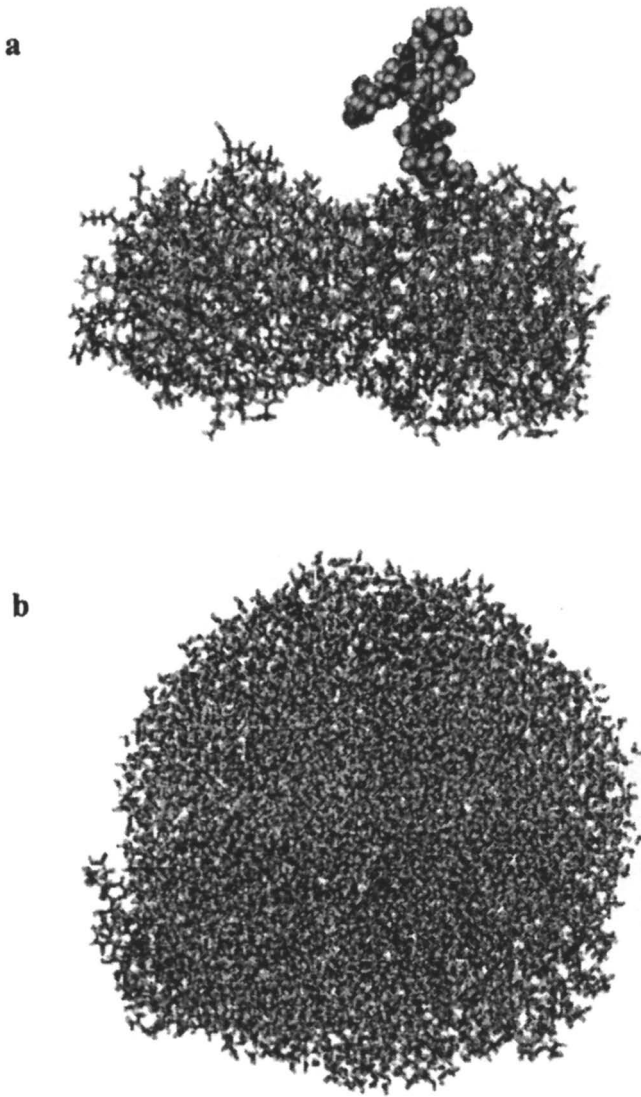


Figure 10. Illustration of the structure of a dimer from pentameric SAP (a) including the N-glycan at one subunit and (b) surrounding water molecules. (See color page 4 in chapter 5.)

Candidates for being responsible for the results measured in the CIDNP experiments are Tyr190 and Tyr195 (see also Table 4), because the terminus of the flexible carbohydrate chain comes rather close to this part of the protein (Figure 11a, b, c, d).

Table 4. Surface accessibility values for laser photo CIDNP-reactive of an SAP units obtained from the X-ray structure of the SAP pentamer (ABCDE) (49). Surface accessibility values which differ due to their position at a monomer-monomer interface are underlined.

| Amino acid | SA [\AA^2] in A | SA [\AA^2] in AB | SA [\AA^2] in AE |
|------------|----------------------------|-----------------------------|-----------------------------|
| His 1 | 118.4 | 118.4 | 118.4 |
| His 78 | 70.3 | 70.3 | 70.3 |
| Trp 108 | 57.4 | 57.4 | <u>51.1</u> |
| Trp 203 | 43.3 | <u>30.8</u> | 43.3 |
| Tyr 40 | 70.9 | <u>22.9</u> | 70.9 |
| Tyr 123 | 43.8 | 43.8 | 43.8 |
| Tyr 140 | 85.2 | 85.2 | 85.2 |
| Tyr 173 | 63.3 | 63.3 | 63.3 |
| Tyr 190 | 56.5 | 56.5 | 56.5 |
| Tyr 195 | 83.9 | <u>47.4</u> | 83.9 |

Dot density : 1; test sphere radius : 1.5 \AA ; SD = standard deviation

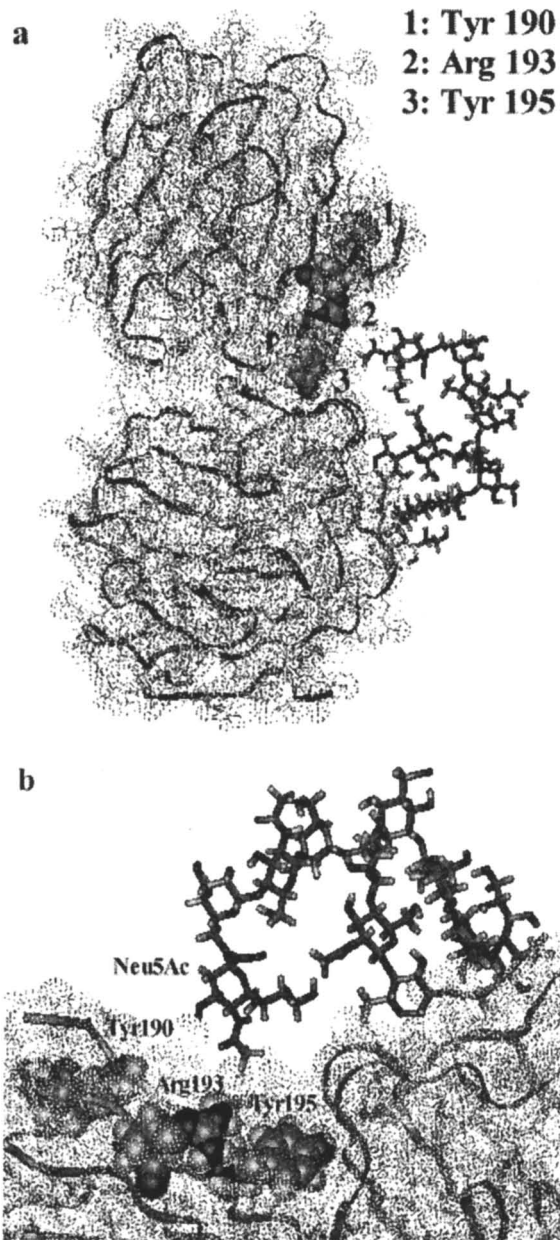


Figure 11. Illustration of the structure of a dimer from pentameric SAP (a) including the unmodified glycan at one subunit (b) with emphasis on the contact region (c) including the desialylated N-glycan of one subunit and (d) with emphasis on the contact region of this form. (See color page 5 in chapter 5.)

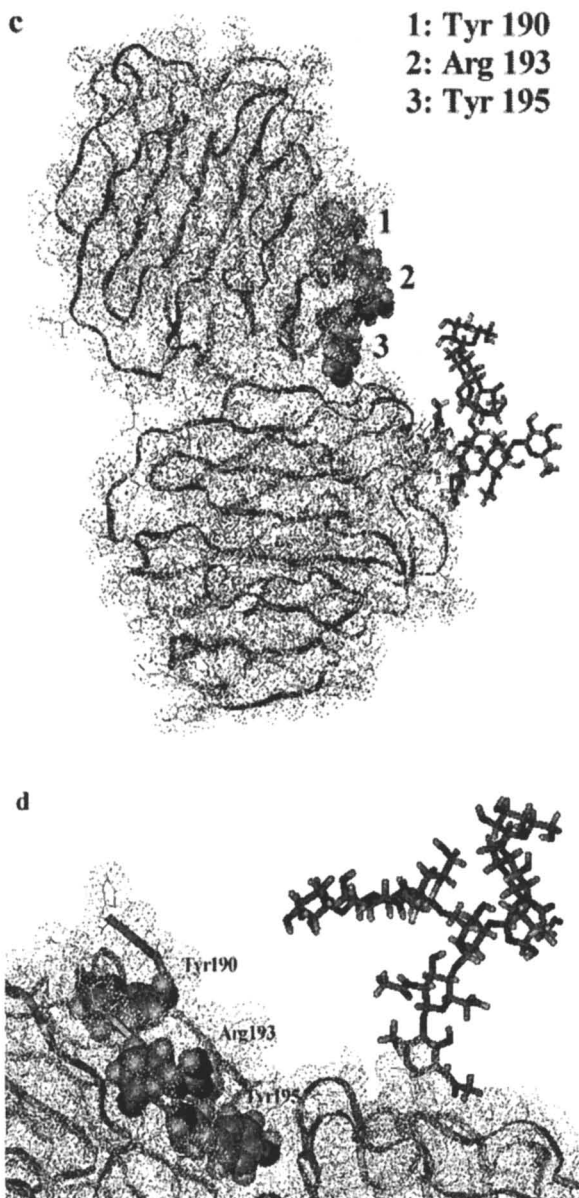


Figure 11. Continued. (See color page 6 in chapter 5.)

Thus, inherent flexibility of the N-glycan antennae leads to contacts to the protein at sites distant to the actual attachment position. The sialic acid may be attracted to this part of the surface due to the presence of the positively charged side chain of Arg193. The drastic effect after sialic acid removal argues in favor of a structural alteration, into the direction of an increase of in size. Indeed, the calculated surface accessibilities (using the Accessible Molecular Surface program <http://swift.cmbi.kun.nl/WIWWWI/> show an increase of the value describing the average surface accessibilities of Tyr190 and Tyr195 when comparing the MD-derived data of simulations carried out for the SAP dimer with an intact carbohydrate chain at one monomer vs. the desialylated carbohydrate chain at one monomer. In detail, with focus on Tyr190/195 yielded an increase in surface accessibility of 79 % and 71 %, respectively. For the complete set of the twelve Tyr-residues an average increase of 27 % concerning their average surface accessibility values is calculated when no sialic acid is present in the N-glycan. We can thus conclude that already desialylation notably affects glycan-protein backbone interactions, with the interplay between computational and experimental work providing detailed insights. As second example, we scrutinize structural aspects of another clinically relevant glycoprotein.

Human chorionic gonadotropin (hCG): The physiological functions of this placental glycoprotein hormone have been described (54-61). Human chorionic gonadotropin (hCG) is a heterodimer consisting of two non-covalently associated, glycosylated subunits, *i. e.* α (α hCG) and β (β hCG). The α -subunit consists of 92 amino acids and is N-glycosylated at Asn52 and Asn78. Both subunits of hCG contain a cystine knot motif formed by three disulfide bridges as a central structural element (57, 58). In the α -subunit of hCG, the two glycosylation sites at Asn52 and Asn78, respectively, have remarkably different properties with respect to the stability of the subunit, as shown by glyco-engineering site-directed mutagenesis (62). The absence of glycosylation at Asn52 by delineating its sequon did not affect folding and stability of the α -subunit. In contrast, mutant α -subunits lacking glycosylation at Asn78 are poorly secreted and rapidly degraded, although the precise molecular origin of this instability is unknown. Structural aspects concerning the glycan chains of the α -subunit

of hCG and their potential conformational impact on the protein part of this hormone have been discussed (63-66). In order to gain further insight into the effects originating from glycosylation in solution, laser photo CIDNP experiments were also carried out. We asked the question as to whether there are any detectable alterations in surface accessibility of distinct amino acid residues in the glycoprotein hormone in the absence and in the presence of its glycan chain. The laser photo CIDNP-sensitive amino acid residue Tyr proved to be a sensitive sensor when comparing the spectra of glycosylated with those of the deglycosylated variant of the isolated α -subunit of hCG. However, results are not as unequivocal as those for SAP. A Tyr-CIDNP signal is detectable for the fully deglycosylated form (Figure 12a) and with comparable size also for the glycosylated form.

The structures presented in Figure 12b give an answer to the question why the laser photo CIDNP results for α hCG cannot be as definite as those for SAP, which is significantly larger than the glycoprotein hormone. Due to the higher degree of flexibility and conformational freedom – as revealed by NMR-supported MD simulations – the surface accessibility values of Tyr residues of α hCG and its overall degree of compactness is not so strongly influenced by the presence or the absence of the carbohydrate chain as is the case for SAP. These data thus delineate properties of the system under investigation, which have a marked influence on the quality of the results.

Conclusions

The suitability of the laser photo CIDNP technique in glycosciences has been documented by a number of examples starting with lectin aspects (binding-site topology as well as ligand-dependent changes in conformation) and extending to intramolecular contacts of the glycan with the protein part's surface. It provides information salient for the interpretation of the conformational analysis of oligosaccharides (6, 67-70), because description of both parts (ligand conformation and the

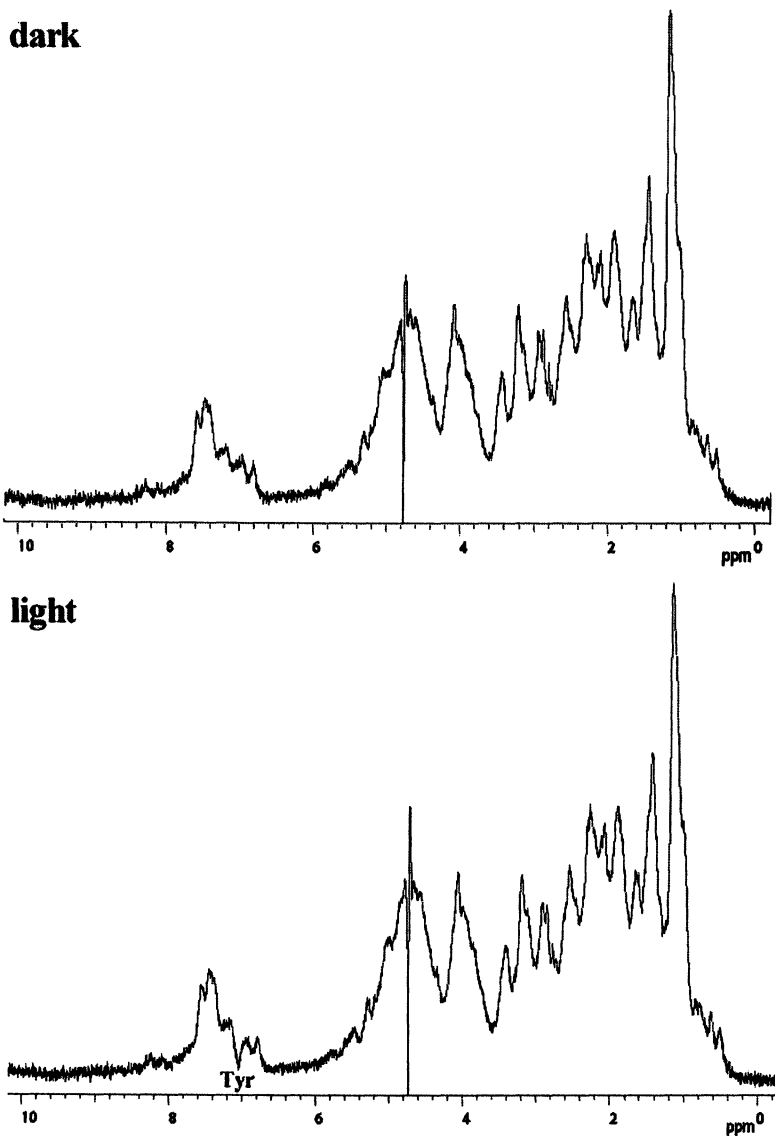


Figure 12. (a) Laser photo CIDNP spectra of deglycosylated ahCG: dark spectrum, light spectrum and difference CIDNP spectrum of deglycosylated ahCG.

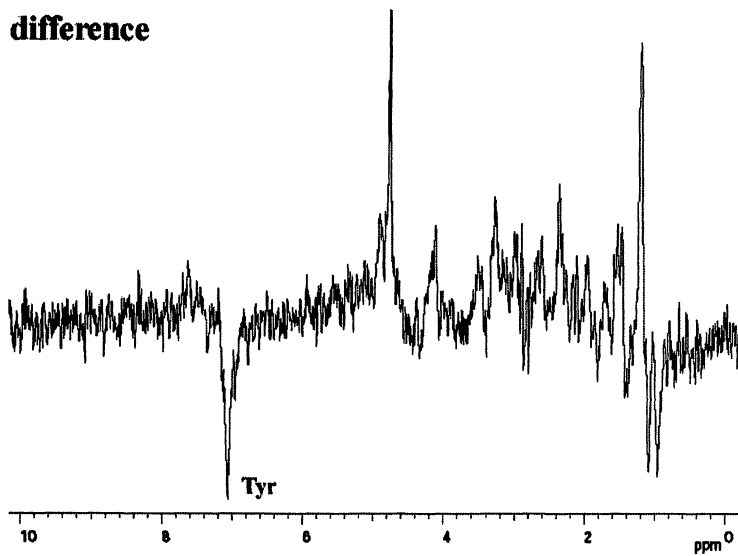
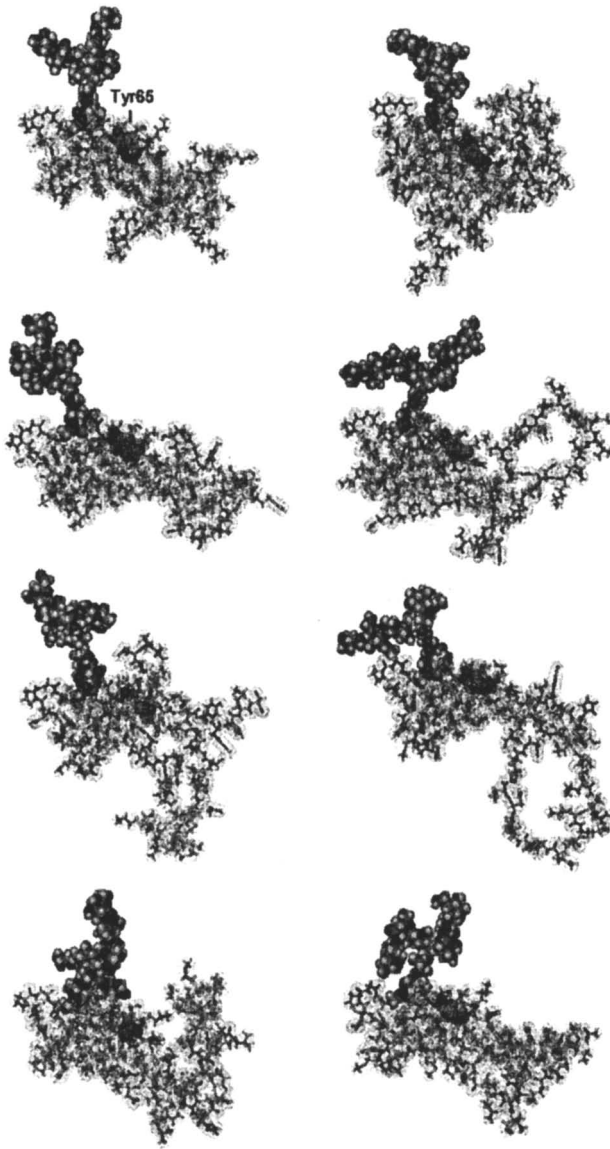


Figure 12. (a) Continued.



*Figure 12. (b) Energy minimum conformations of ahCG glycosylated at Asn78 – (pdb-entry: 1HD4). The CIDNP-relevant Tyr-residue (Tyr65) is highlighted by its van der Waals representation.
(See color page 7 in chapter 5.)*

architecture of the receptor's binding pocket) are essential to structurally understand the interplay of glycans with plant/ animal lectins even with perspective for drug design (36, 71, 72). As described in this chapter, laser photo CIDNP experiments provide information on the extent of surface accessibility of Tyr, Trp and His residues of lectins and other carbohydrate-binding proteins such as enzymes (e. g. sialidases). Since these three CIDNP sensitive amino residues are very often located in the binding pocket of a lectin, it is possible to precisely and conveniently monitor ligand binding by reduction of the corresponding CIDNP signal, as demonstrated for plant lectins such as hevein-like lectins and human β -galactoside-binding lectins. The obtained results are in line with detailed structural analysis by NMR spectroscopy for the hevein family (73, 74) or crystallography - including modeling - with mammalian galectin-1 (75-78). Furthermore, it is also technically feasible and informative to study the contact of a carbohydrate chain or even parts of such an oligosaccharide chain to the protein part by the laser photo CIDNP technique. The discussed laser photo CIDNP data obtained for the glycoproteins SAP and hCG with an intact, truncated or completely deleted carbohydrate part are illustrative examples to underline this point. As pointed out, the obtained data are a useful quality control for data sets obtained by molecular modeling. We exemplify the value of combined analysis in the case study SAP. When comparing SAP with an intact and SAP with a partially or fully truncated oligosaccharide chain using distinct surface-exposed Tyr residues as sensors, significant changes of the signal intensity of the Tyr CIDNP signal were detectable. The increase of the signal intensity of the Tyr signal after cleavage of the sialic acid or the complete carbohydrate chain combined with the computational data indicated that the density of the protein surface structure is likely reduced. With the examples illustrated herein, is thus justified to count laser photo CIDNP analysis as a useful tool in structural glycosciences.

Acknowledgements: The authors thank Dr. Johannes J. M. van Rooijen and Drs. Vincent F. van Miegem for helpful technical assistance.

References

1. Gabius, H.-J.; Gabius, S. (eds.) *Glycosciences: Status and Perspectives*. 1997. Chapman & Hall, Weinheim-London.
2. von der Lieth, C.-W.; Siebert, H.-C.; Kozár, T.; Burchert, M.; Frank, M.; Gilleron, M.; Kaltner, H.; Kayser, G.; Tajkhorshid, E.; Bovin, N. V.; Vliegenthart, J. F. G.; Gabius, H.-J. *Acta Anat.* 1998, 161, 91.

3. Reuter, G.; Gabius, H.-J. *Cell. Mol. Life Sci.* **1999**, *59*, 368.
4. Wormald, M. R.; Dwek, R. A. *Structure* **1999**, *7*, 155.
5. Gabius, H.-J.; André, S.; Kaltner, H.; Siebert, H.-C. *Biochim. Biophys. Acta* **2002**, *1572*, 165.
6. Gabius, H.-J.; Siebert, H.-C.; André, S.; Jiménez-Barbero, J.; Rüdiger H. *ChemBioChem* **2004**, *5*, 740.
7. André, S.; Unverzagt, C.; Kojima, S.; Dong, X.; Fink, C.; Kayser, K.; Gabius, H.-J. *Bioconjugate Chem.* **1997**, *8*, 845.
8. Unverzagt, C.; André, S.; Seifert, J.; Kojima, S.; Fink, C.; Srikrishna, G.; Freeze, H. H.; Kayser, K. & Gabius, H.-J. *J. Med. Chem.* **2002**, *45*, 478.
9. André, S.; Unverzagt, C.; Kojima, S.; Frank, M.; Seifert, J.; Fink, C.; Kayser, K.; von der Lieth, C.-W.; Gabius, H.-J. *Eur. J. Biochem.* **2004**, *271*, 118.
10. Kaptein, R.; Dijkstra, K.; Nicolay, K. *Nature* **1978**, *274*, 293.
11. Kaptein, R. In: "Biological Magnetic Resonance." **1982**, *4* (Berliner, L. J. ed.) New York, Plenum Press., 145.
12. Broadhurst, R. W.; Dobson, C. M.; Hore, P. J.; Radford, S. E.; Rees, M. L. *Biochemistry* **1991**, *30*, 405.
13. Hore, P. J.; Broadhurst, R. W. *Progr. NMR Spectrosc.* **1993**, *25*, 345.
14. Siebert, H.-C.; Tajkhorshid, E.; von der Lieth, C.-W.; Kleineidam, R. G.; Kruse, S.; Schauer, R.; Kaptein, R.; Gabius, H.-J.; Vliegthart, J. F. G. *J. Mol. Model.* **1996**, *2*, 446.
15. Tajkhorshid, E.; Siebert, H.-C.; Burchert, M.; Kaltner, H.; Kayser, G.; von der Lieth, C.-W.; Kaptein, R.; Vliegthart, J. F. G.; Gabius, H.-J. *J. Mol. Model.* **1997**, *3*, 325.
16. Siebert, H.-C.; Lu, S.-Y.; Frank, M.; Kramer, J.; Wechselberger, R.; Joosten, J.; André, S.; Rittenhouse-Olson, K.; Roy, R.; von der Lieth, C.-W.; Kaptein, R.; Vliegthart, J. F. G.; Heck, A. J. R.; Gabius, H.-J. *Biochemistry* **2002**, *41*, 9707.
17. Arnusch, C.-J.; André, S.; Valentini, P.; Lensch, M.; Russwurm, R.; Siebert, H.-C.; Fischer, M. J. E.; Gabius, H.-J.; Pieters, R. J. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1437.
18. Mok, K. H.; Hore, P. J. *Methods* **2004**, *34*, 75.
19. Morozova, O. B.; Yurkovskaya, A. V.; Sagdeev, R. Z.; Mok, K. H.; Hore, P. J. *J. Phys. Chem. B* **2004**, *108*, 15355.
20. Siebert, H.-C.; Adar, R.; Arango, R.; Burchert, M.; Kaltner, H.; Kayser, G.; Tajkhorshid, E.; von der Lieth, C.-W.; Kaptein, R.; Sharon, N.; Vliegthart, J. F. G.; Gabius, H.-J. *Eur. J. Biochem.* **1997**, *249*, 27.
21. Villalobo, A.; Gabius, H.-J. *Acta Anat.* **1998**, *161*, 110.
22. Kaltner, H.; Stiersdorfer, B. *Acta Anat.* **1998**, *161*, 162.
23. Gabius, H.-J. *Anat. Histol. Embryol.* **2001**, *30*, 3.
24. Rüdiger, H.; Gabius, H.-J. *Glycoconjugate J.* **2001**, *18*, 589.

25. Gabius, H.-J., Gabius, S. (eds.) *Lectins and Glycobiology 1993*; Springer Verlag, Heidelberg - New York.
26. Gabius, H.-J. *Biochimie* **2001**, *83*, 659.
27. Manning, J. C.; Seyrek, K.; Kaltner, H.; André, S.; Sinowatz, F.; Gabius, H.-J. *Histol. Histopathol.* **2004**, *19*, 1043.
28. Gabius, H.-J. *Eur. J. Biochem.* **1997**, *243*, 543.
29. Solís, D.; Jiménez-Barbero, J.; Kaltner, H.; Romero, A.; Siebert, H.-C.; von der Lieth, C.-W.; Gabius, H.-J. *Cells Tissues Organs* **2001**, *168*, 5.
30. Shionyu-Mitsuyama, C.; Shirai, T.; Ishida, H.; Yamane, T. *Protein Eng.* **2003**, *16*, 467.
31. Sujatha, M. S.; Balaji, P. V. *Proteins* **2004**, *55*, 44.
32. Kopitz, J.; von Reitzenstein, C.; André, S.; Kaltner, H.; Uhl, J.; Ehemann, V.; Cantz, M.; Gabius, H.-J. *J. Biol. Chem.* **2001**, *276*, 35917.
33. Camby, I.; Belot, N.; Lefranc, F.; Sadeghi, N.; de Launoit, Y., Kaltner, H., Musette, S.; Darro, F.; Danguy, A.; Salmon, I.; Gabius, H.-J.; Kiss, R. *J. Neuropathol. Exp. Neurol.* **2002**, *61*, 585.
34. Rappl, G.; Abken, H.; Mucbe, J. M.; Sterry, W.; Tilgen, W.; André, S.; Kaltner, H.; Ugurel, S.; Gabius, H.-J.; Reinhold, U. *Leukemia* **2002**, *16*, 840.
35. Lahm, H.; André, S.; Hoeflich, A.; Kaltner, H.; Siebert, H.-C.; Sordat, B.; von der Lieth, C.-W.; Wolf, E.; Gabius, H.-J. *Glycoconjugate J.* **2004**, *20*, 227.
36. Siebert, H.-C.; André, S.; Lu, S.-Y.; Frank, M.; Kaltner; van Kuik, J.-A.; Korchagina, E.-Y.; Bovin, N.; Tajkhorshid, E.; Kaptein, R.; Vliegenthart, J. F. G.; von der Lieth, C.-W.; Jiménez-Barbero, J.; Kopitz, J.; Gabius, H.-J. *Biochemistry* **2003**, *42*, 14762.
37. Siebert, H.-C.; von der Lieth, C.-W.; Kaptein, R.; Beintema, J. J.; Dijkstra, K.; van Nuland, N.; Soedjanaatmadja, U. M. S.; Rice, A.; Vliegenthart, J. F. G.; Wright, C. S.; Gabius, H.-J. *Proteins* **1997**, *28*, 268.
38. Siebert, H.-C.; André, S.; Asensio, J. L.; Cañada, F. J.; Dong, X.; Espinosa, J.-F.; Frank, M.; Gilleron, M.; Kaltner, H.; Kozár, T.; Bovin, N. V.; von der Lieth, C.-W.; Vliegenthart, J. F. G.; Jiménez-Barbero, J.; Gabius, H.-J. *ChemBioChem* **2000**, *1*, 181.
39. Siebert, H.-C.; André, S.; Vliegenthart, J. F. G.; Gabius, H.-J.; Minch, M. J. *J. Biomol. NMR* **2003**, *25*, 197.
40. Siebert, H.-C.; Jiménez-Barbero, J.; André, S.; Kaltner, H.; Gabius, H.-J. *Meth. Enzymol.* **2003**, *362*, 449.
41. He, L.; André, S.; Siebert, H.-C.; Helmholz, H.; Niemeyer, B.; Gabius, H.-J. *Biophys. J.* **2003**, *85*, 511.
42. Galelli, A.; Truffa-Bachi, P. *J. Immunol.* **1993**, *151*, 1821.

43. Coker, A. R.; Purvis, A.; Baker, D.; Pepys, M. B.; Wood, S. P. *FEBS Lett.* **2000**, *473*, 199.
44. Noursadeghi, M.; Bickerstaff, M. C. M.; Gallimore, J. R.; Herbert, J.; Cohen, J.; Pepys, M. B. *Proc. Natl. Acad. Sci. U S A.* **2000**, *97*, 14584.
45. Gillmore, J. D.; Winston L. Hutchinson, W. L.; Herbert, J.; Bybee, A.; Mitchell, D. A.; Hasserjian, R. P.; Yamamura, K.-I.; Suzuki, M.; Sabin, C. A.; Pepys, M. B. *Immunology*, **2004**, *112*, 255.
46. Kilpatrick, D. C. *Biochim. Biophys. Acta* **2002**, *1572*, 401.
47. Cambi, A.; Figdor, C. G. *Curr. Opin. Cell Biol.* **2003**, *15*, 539.
48. van Kooyk, Y.; Geijtenbeek, T. B. H. *Nature Rev. Immunol.* **2003**, *3*, 697.
49. Emsley, J.; White, H. E.; O'Hara, B. P.; Oliva, G.; Srinivasan, N.; Tickle, I. J.; Blundell, T. L.; Pepys, M. B.; Wood, S. P. *Nature* **1994**, *367*, 338.
50. Loveless, R. W.; Floyd-O'Sullivan, G.; Raynes, J.-G.; Yuen, C.-T.; Feizi, T. *EMBO J.* **1992**, *11*, 813.
51. Pepys, M. B.; Rademacher, T. W.; Amatayakul-Chantler, S.; Williams, P.; Noble, G. E.; Hutchinson, W. L.; Hawkins, P. N.; Nelson, S. R.; Gallimore, J. R.; Herbert, J.; Hutton, T.; Dwek, R. A. *Proc Natl Acad Sci USA* **1994**, *91*, 5602.
52. Gabius, H.-J. *Adv. Drug Deliv. Rev.* **2004**, *56*, 421.
53. Siebert, H.-C.; André, S.; Reuter, G.; Gabius, H.-J.; Kaptein, R.; Vliegthart, J. F. G. *FEBS Lett.* **1995**, *371*, 13.
54. Erbel, P. J. A.; Karimi-Nejad, Y.; De Beer, T.; Boelens, R.; Kamerling, J. P.; Vliegthart, J. F. G. *Eur. J. Biochem.* **1999**, *260*, 490.
55. Erbel, P. J. A., Karimi-Nejad, Y., Van Kuik, J. A., Boelens, R., Kamerling, J. P., Vliegthart, J. F. G. *Biochemistry* **2000**, *39*, 6012.
56. Erbel, P. J.; Haseley, S. R.; Kamerling, J. P.; Vliegthart; J. F. G. *Biochem. J.* **2002**, *364*, 485.
57. Laphorn, A. J.; Harris, D. C.; Littlejohn, A.; Lustbader, J. W.; Canfield, R. E.; Machin, K. J.; Morgan, F. J.; Isaacs, N. W. *Nature* **1994**, *369*, 455.
58. Wu, H.; Lustbader, J. W.; Liu, Y.; Canfield, R. E.; Hendrickson, W. A. *Structure* **1994**, *2*, 545.
59. Bliithe, D. L.; Richards, R. G.; Skarulis, M. C. *Endocrinology* **1991**, *129*, 2257.
60. Bliithe, D. L.; Iles, R. K. *Endocrinology* **1995**, *136*, 903.
61. Wyss, D. F.; Wagner, G. *Curr. Opin. Biotechnol.* **1996**, *7*, 409.
62. Matzuk, M. M.; Boime, I. *J. Cell. Biol.* **1988**, *106*, 1049.
63. van Zuylen, C. W. E. M.; De Beer, T.; Rademaker, G. J.; Haverkamp, J.; Thomas-Oates, J. E.; Hård, K.; Kamerling, J. P.; Vliegthart, J. F. G. *Eur. J. Biochem.* **1995**, *231*, 754.

64. De Beer, T.; Van Zuylen, C. W. E. M.; Leeftang, B. R.; Hård, K.; Boelens, R.; Kaptein, R.; Kamerling, J. P.; Vliegthart, J. F. G. *Eur. J. Biochem.* **1996**, *241*, 229.
65. van Zuylen, C. W. E. M.; Kamerling, J. P.; Vliegthart, J. F. G. *Biochem. Biophys. Res. Commun.* **1997**, *232*, 117.
66. Thijssen-van Zuylen, C. W. E. M.; De Beer, T.; Leeftang, B. R.; Boelens, R.; Kaptein, R.; Kamerling, J. P.; Vliegthart, J. F. G. *Biochemistry* **1998**, *37*, 1933.
67. Siebert, H.-C.; Reuter, G.; Schauer, R.; von der Lieth, C.-W.; Dabrowski, J. *Biochemistry* **1992**, *31*, 6962.
68. Siebert, H.-C.; Gilleron, M.; Kaltner, H.; von der Lieth, C.-W.; Kozár, T.; Bovin, N. V.; Korchagina, E. Y.; Vliegthart, J. F. G.; Gabius, H.-J. *Biochem. Biophys. Res. Commun.* **1996**, *219*, 205.
69. Siebert, H.-C.; von der Lieth, C.-W.; Dong, X.; Reuter, G.; Schauer, R.; Gabius, H.-J.; Vliegthart, J. F. G. *Glycobiology* **1996**, *6*, 561.
70. Gilleron, M.; Siebert, H.-C.; Kaltner, H.; von der Lieth, C.-W.; Kozár, T.; Halkes, K. M.; Korchagina, E. Y.; Bovin, N. V.; Gabius, H.-J.; Vliegthart, J. F. G. *Eur. J. Biochem.* **1998**, *252*, 416.
71. Siebert, H.-C.; Kaptein, R.; Beintema, J. J.; Soedjanaatmadja, U. M. S.; Wright, C. S.; Rice, A.; Kleineidam, R. G.; Kruse, S.; Schauer, R.; Pouwels, P. J. W.; Kamerling, J. P.; Gabius, H.-J.; Vliegthart, J. F. G. *Glycoconjugate J.* **1997**, *14*, 531.
72. Rüdiger, H.; Siebert, H.-C.; Solís, D.; Jiménez-Barbero, J.; Romero, A.; von der Lieth, C.-W.; Díaz-Mauriño, T.; Gabius, H.-J. *Curr. Med. Chem.* **2000**, *7*, 389.
73. Asensio, J. L.; Siebert, H.-C.; von der Lieth, C.-W.; Laynes, J.; Bruix, M.; Soedjanaatmadja, U. M. S.; Beintema, J. J.; Cañada, F. J.; Gabius, H.-J.; Jiménez-Barbero, J. *Proteins* **2000**, *40*, 218.
74. Asensio, J. L.; Cañada, F. J.; Siebert, H.-C.; Laynez, J.; Poveda, A.; Nieto, P. M.; Soedjanaatmadja, U. M. S.; Gabius, H.-J.; Jiménez-Barbero, J. *Chem. Biol.* **2000**, *7*, 529.
75. Liao, D.; Kapadia, G.; Ahmed, H.; Vasta, G. R.; Herzberg, O. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1428.
76. Bourne, Y.; Bolgiano, B.; Liao, D.; Strecker, G.; Cantau, P.; Herzberg, O.; Feizi, T.; Cambillau, C. *Nature Struct. Biol.* **1994**, *1*, 863.
77. López-Lucendo, M. F.; Solís, D.; André, S.; Hirabayashi, J.; Kasai, K.; Kaltner, H.; Gabius, H.-J.; Romero, A. *J. Mol. Biol.* **2004**, *343*, 957.
78. André, S.; Kaltner, H.; Lensch, M.; Russwurm, R.; Siebert, H.-C.; Tajkhorshid, E.; Heck, A. J. R.; von Knebel-Doerberitz, M.; Gabius, H.-J.; Kopitz J. *Int. J. Cancer*, **2005**, *114*, 46.

Chapter 6

Hydroxy Protons in Structural Studies of Carbohydrates by NMR Spectroscopy

Corine Sandström and Lennart Kenne

Department of Chemistry, Swedish University of Agricultural Sciences,
P.O. Box 7015, SE-750 07 Uppsala, Sweden

The use of hydroxy protons in conformation, structure and interaction studies of carbohydrates in aqueous solution by NMR spectroscopy has increased in importance. In this paper we present the experimental conditions that make it possible to obtain ^1H NMR spectra of hydroxy protons of samples in aqueous solutions and procedures for assignments of these signals. The most important NMR parameters obtained for the hydroxy protons are chemical shifts, vicinal coupling constants ($^3J_{\text{HC,OH}}$), temperature coefficients and exchange rates. These parameters give information on hydrogen bond interactions and hydration. Interresidue NOEs and chemical exchanges observed by NOESY and ROESY experiments increase the number of distance restraints used in conformational analysis.

Introduction

Oligosaccharide chains, free or chemically bound in glycoconjugates, are known to mediate cell-cell recognition, including the infection of cells by bacteria and viruses, moderate the behavior of enzymes and other proteins, and fulfill various functions in the immune response. The recognition of carbohydrates is also important for carbohydrate metabolism and for their transport across cell membranes. They can function in plants as molecular signals involved in the regulation of growth, development and defence. Understanding the conformational properties of carbohydrates is essential for elucidation of their mechanisms of action, and this may aid in the design of carbohydrate-based vaccines, antiviral drugs and other therapeutic agents but also in the production of tailor made polysaccharide materials with improved functions.

For these reasons, a significant amount of research is devoted to studies of the three-dimensional structures and dynamics of oligosaccharides and polysaccharides. However, unlike proteins and nucleic acids, in solution oligo- and polysaccharides do not contain secondary structural motifs or well-organized tertiary structures. Instead oligosaccharides often exist as an ensemble of multiple conformational families, and the structural characterization of carbohydrates in aqueous solution is thus challenging. The number of conformational restraints provided by ^1H homonuclear NOEs and ^1H - ^{13}C spin-coupling constants is usually insufficient to define the preferred conformations around the glycosidic linkage. Isotope labeling of the oligosaccharide with ^{13}C enables the observation of additional conformational parameters across the glycosidic linkage, such as heteronuclear ^1H - ^{13}C NOEs (1) and homonuclear ^{13}C - ^{13}C spin-spin coupling constants (2). Recently it was shown that residual long-range dipolar couplings can give long-range structural information and be used to characterize amplitudes of motion (3).

Another factor that has to be taken into account in studies of carbohydrates is that the structure and dynamics of carbohydrates are significantly influenced by localized interactions with surrounding water molecules. The majority of the hydrogen-bonding interactions with water occur through hydroxyl groups and carbohydrates affect the surrounding water structure. In return, the water affects the structure of the carbohydrate molecules. However, due to a lack of experimental methods to study these interactions at the molecular level, the understanding of the hydration of carbohydrates remains fragmentary.

A mean of increasing the number of conformational constraints and of gaining information on the hydration of carbohydrates is to use the hydroxy protons. Their observation by NMR offers several important features for the study of carbohydrates in aqueous solution: (i) hydroxy protons are abundant; (ii) their signals have a relatively large dispersion of chemical shifts and appear in a well-isolated region in the NMR spectrum downfield from the aliphatic protons; (iii) they have several NMR parameters that are sensitive to the

conformation. Basically, five NMR parameters can be obtained from the detection of the hydroxy protons: The chemical shifts of the OH proton signals (δ), the vicinal coupling constants ($^3J_{\text{H,OH}}$), the temperature coefficients ($d\delta/dT$) obtained from the variation of chemical shifts with temperature, the rate of exchange with water (k_{ex}), and NOEs and chemical exchanges. From these NMR parameters, additional structural information in terms of hydrogen bonding, distances and hydration can be obtained. Because of rapid exchange between protons of hydroxyl groups and protons of water, the use of hydroxy protons for structural studies in aqueous solution by NMR has in the past presented experimental challenges. To our knowledge, the first observation by NMR of hydroxy proton signals of a carbohydrate was reported in 1976 (4).

In this chapter, we will first describe the experimental conditions allowing observation and assignment of hydroxy proton signals. The use of hydroxy protons for conformational and hydration studies of free carbohydrates with a special emphasis on the chemical shifts will be discussed. Finally, the use of hydroxy protons to study the formation of inclusion complexes of cyclodextrins and carbohydrate-protein interactions will be presented.

Experimental conditions

A careful sample preparation is usually required to allow the observation of hydroxy protons of carbohydrates in water solution by ^1H NMR spectroscopy. Typically, it is necessary to adjust the pH to 6-7 and remove all traces of metal ions that could accelerate the exchange of hydroxy protons with water. For this reason the compounds are passed through an Amberlite MB-3 mixed ion-exchange resin. The pH is carefully adjusted by addition of minute amount of concentrated HCl or NaOH but also buffered solutions are used. To minimize release of impurities from glassware the NMR sample tubes are soaked for more than 1 h in a 50 mM phosphate buffer (pH 7) (5). Lowering the sample temperature below 0 °C can further reduce the rate of exchange. Sub-zero sample temperatures of aqueous samples have been achieved by using supercooling (6) or highly concentrated solutions (7). The most common strategy to avoid freezing of the sample, however, is to add 10-15% of an organic solvent such as acetone or methanol. This mixture permits lowering the temperature to around -15 °C without freezing.

The huge water peak can be efficiently suppressed with pulse sequences such as Watergate (8,9) that is easily implemented in 2D NMR experiments.

Acetone- d_6 is a relatively weak hydrogen bond acceptor (10). A study on the effect of stereochemistry on hydroxy proton chemical shifts of simple carbohydrates revealed that the chemical shifts of hydroxy proton signals were relatively insensitive to sample conditions, such as pH and concentration of acetone in water (11). Another study on hydrogen bonding in dicarboxylic acids showed that even in 90% $(\text{CD}_3)_2\text{CO}/10\%$ H_2O the water amount was sufficient to

allow full solvation of the intramolecularly hydrogen bonded species (12). In β -cyclodextrin, both the chemical shifts of the hydroxy proton signals and the observed coupling constants were the same in 95% $\text{H}_2\text{O}/5\%$ D_2O and in 85% $\text{H}_2\text{O}/15\%$ $(\text{CD}_3)_2\text{CO}$ (13). These studies indicate that acetone has no significant effects on the hydroxy protons.

Assignments

The first step in studies of hydroxy protons by NMR spectroscopy involves the assignment of the hydroxy proton resonances. It can be done on the basis of scalar connectivities to the aliphatic protons by two-dimensional chemical shift correlation experiments such as COSY and TOCSY. In case of severe spectral overlap in the ^1H NMR spectra, two-dimensional heteronuclear chemical shift correlation experiments such as ^1H - ^{13}C HMBC can be used.

As the hydroxy protons resonate in a well-isolated region (δ 5–7.5 ppm) their signals can aid in the assignment of ring proton signals. For example, the assignment of the six glucose proton resonances in mono-*altro*- β -cyclodextrin was greatly facilitated from the observation of the hydroxy proton signals (14). The six proton spin systems could be identified in the TOCSY spectrum starting from the chemical shift of the O(2)H signals (Fig. 1).

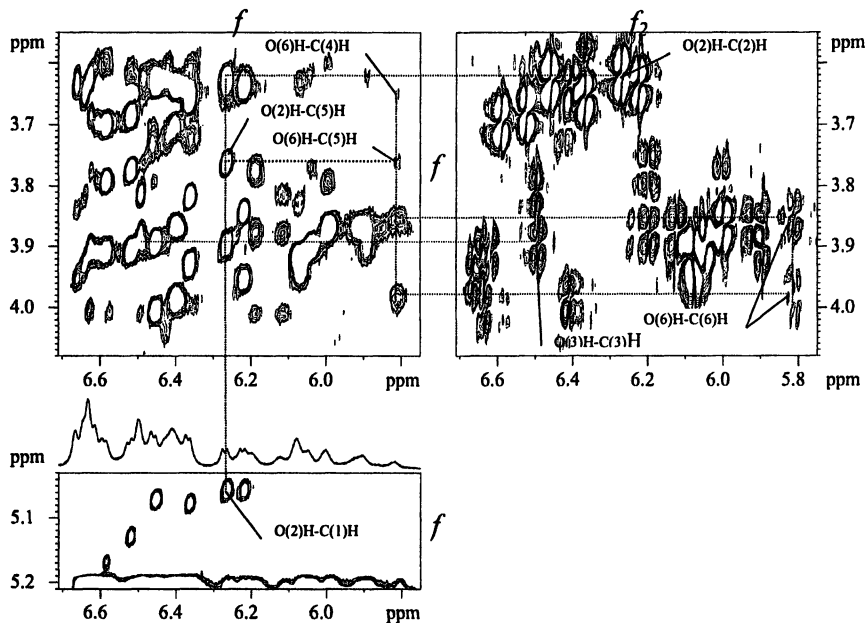


Figure 1. Hydroxy proton region of 2D TOCSY (left) and DQF-COSY (right) spectra of mono-*altro*- β -cyclodextrin. The assignment of the spin system of one residue is indicated by lines.

Hydroxy protons have also been used to determine the glycosyl linkage positions in oligosaccharides (15). The method is based on the absence of a hydroxyl group on the carbon atom that participates in the glycosidic linkage.

In cases of rapid chemical exchange, where the resonances of solvent-exchangeable protons cannot be distinguished from the solvent resonance, the water stripe in the f2 dimension of a TOCSY spectrum has been used to identify the chemical shifts of protons that are scalar coupled to these exchangeable protons (16).

Hydroxy protons in conformational analysis

Hydrogen bond interaction

It is often believed that hydrogen bonds usually do not persist in aqueous solution, due to the dominating solvation of the hydroxy protons with the solvent water molecules (17). When a hydrogen bond exists in aqueous solution, it would only reflect the presence of a stable conformation, in which the hydroxyl groups are located by chance within hydrogen-bonding distance. It is, however, also possible that hydrogen bonds play a deciding role in the selection between two or more conformations of otherwise similar energies (18).

Protons involved in hydrogen bonds are deshielded and the magnitude of the deshielding is dependent on the strength of the hydrogen bond (19). There are however only few reports on persistence of strong intramolecular hydrogen bonds in aqueous solution evidenced by a downfield shift and these are mainly for aromatic compounds (20). To our knowledge, the only example of large downfield shift (1 ppm) of hydroxy protons measured in aqueous solution of carbohydrate related compounds is for the intramolecular hydrogen-bonding between hydroxyl and phosphate groups in *myo*-inositol compounds (21). Thus, the chemical shifts of hydroxy protons have so far played little role in conformational studies of carbohydrates by NMR and instead coupling constants, temperature coefficients, exchange rates and deuterium-induced differential isotope shift (22) have been used to detect the existence of hydrogen bonding in aqueous solution of carbohydrates.

According to the Karplus equation derived for hydroxy protons (23), vicinal coupling of the order of 5.5 ± 0.5 Hz indicates conformational averaging with free rotation of the hydroxyl group around the C-O bond. A hydrogen bond that enforces some particular angle could be reflected in a deviation of the coupling constant for that hydroxy proton from the rotationally averaged value. The chemical shifts of hydroxy protons that are hydrogen bonded to the solvent have marked temperature dependence due to changes in mobility of the solvent molecules. The chemical shifts of the resonances of hydrogen-bonded hydroxyl groups are less affected by temperature, due to the decreased interaction with solvent. Thus, temperature coefficients of <3 ppb/°C have been measured for

intramolecularly hydrogen-bonded protons (24). Protons engaged in strong intramolecular hydrogen bonds should also exchange more slowly with the solvent. Solvent accessibility could be reduced by participation in a hydrogen bond, but other causes are also possible (25). Exchange rates are very sensitive to pH, solvent composition and catalysis by small traces of impurities, and comparison between different compounds should be avoided. It has also been suggested that the existence of chemical exchange between hydroxyl groups may be an important parameter in probing weaker, transient hydrogen bonds in aqueous solution (26). Thus, if a hydroxy proton is involved in hydrogen bonding, it should have at least one NMR parameter that differs from that of the other hydroxy protons (27- 33). In α -D-Galp-(1 \rightarrow 4)- β -D-Galp, an efficient inhibitor of the bacterial pilus adhesin PapG₁₉₆, the presence of a transient hydrogen bond between O(2')H and O(6)H was evidenced from the existence of a chemical exchange between the two hydroxy protons together with the large $^3J_{\text{CH,OH}}$ -value of 9.1 Hz measured for O(2')H (34). The average values of coupling constants and the absence of chemical exchange cross peaks indicated that the hydrogen bond did not exist in the S-linked 4-thiodisaccharide analogue. This compound was also a much less efficient inhibitor.

As the anomeric configuration of the reducing end of an oligosaccharide usually has a negligible influence on the conformation, the signals from the protons of the non-reducing unit have similar chemical shifts for both the α - and β -form and only those from the reducing unit are different. However, while 12 hydroxy proton signals were expected for maltose, the DQF-COSY spectrum (Figure 2) showed the presence of 13 hydroxy proton signals (13). Two O(2')H signals with different 3J -values (8 Hz in α and 6.6 Hz in β) were observed. An exchange cross peak was observed in the ROESY spectrum between the O(3)H and O(2')H signals of β -maltose but not of α -maltose. This result demonstrates the influence of the anomeric configuration on the interresidual O(2')H-O(3)H interaction.

The O(2')H of the 3-linked sugar in the 3,4-di-*O*-glycopyranosyl-substituted methyl α -D-galactopyranosides **1** – **3** (Figure 3) had NMR parameters indicating the participation in persistent hydrogen bond interactions (35). While most hydroxy protons had $^3J_{\text{H,OH}}$ -values of 5.5 ± 0.5 Hz, representing rotational averaging, the O(2')H in **1**, **2** and **3** had J -values indicating restricted rotation. In **1**, the large $^3J_{\text{H,OH}}$ of 10.5 Hz showed that the O(2')H proton adopted a *trans* conformation with respect to C(2')H. In **2** and **3**, the small $^3J_{\text{H,OH}}$ of 1.8 and 3 Hz indicated a preferred *syn* orientation. The O(2')H had small temperature coefficient values (<5 ppb. $^{\circ}\text{C}^{-1}$), indicating that these protons were strongly protected from exchange with the solvent. The exchange of O(2')H with water in the three trisaccharides was much slower than that for the other hydroxy protons. NOESY and ROESY experiments together with HSEA (36) and MM2 calculations (35) indicated that the most probable hydrogen bonding partner was the O(5'') of the 4-linked sugar.

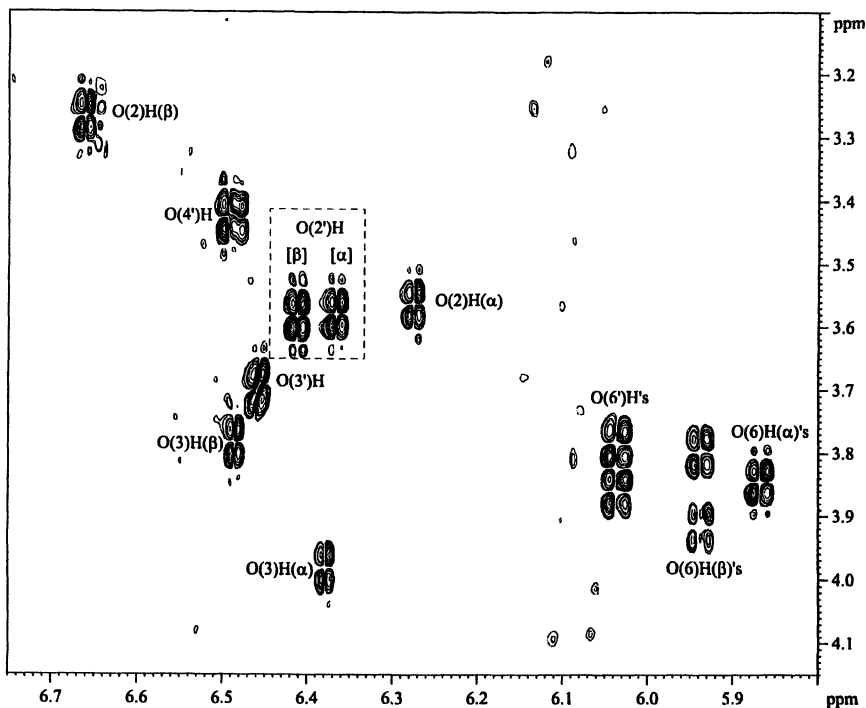


Figure 2. Part of the 2D DQF-COSY spectrum of maltose with two cross peaks for O(2')H resulting from the different anomeric configurations of the reducing end. (Reproduced from reference 13. Copyright 2003. American Chemical Society.)

The NMR data ($^3J_{\text{H,OH}}$, $d\delta/dT$ and k_{ex}) indicated that the hydrogen bond between O(2')H and O(5'') was stronger in 1 and 2 than in 3. A chemical exchange interaction between O(2') and O(6''), that might further stabilize the conformation in which O(2') and O(5'') are hydrogen bonded, was found in 1 and 2 but not in 3. However, despite the fact that the NMR data suggested that O(2')H protons were involved in a persistent hydrogen bonding, the protons were shielded, the O(2)H hydroxy proton involved in the strongest hydrogen bond being the least shielded.

The existence of a hydrogen bond network between the secondary hydroxyl groups in α -, β - and γ -cyclodextrins (CD), well established in the crystal structures (37) and in DMSO solutions (38-40), was demonstrated to also be present in aqueous solution (13).

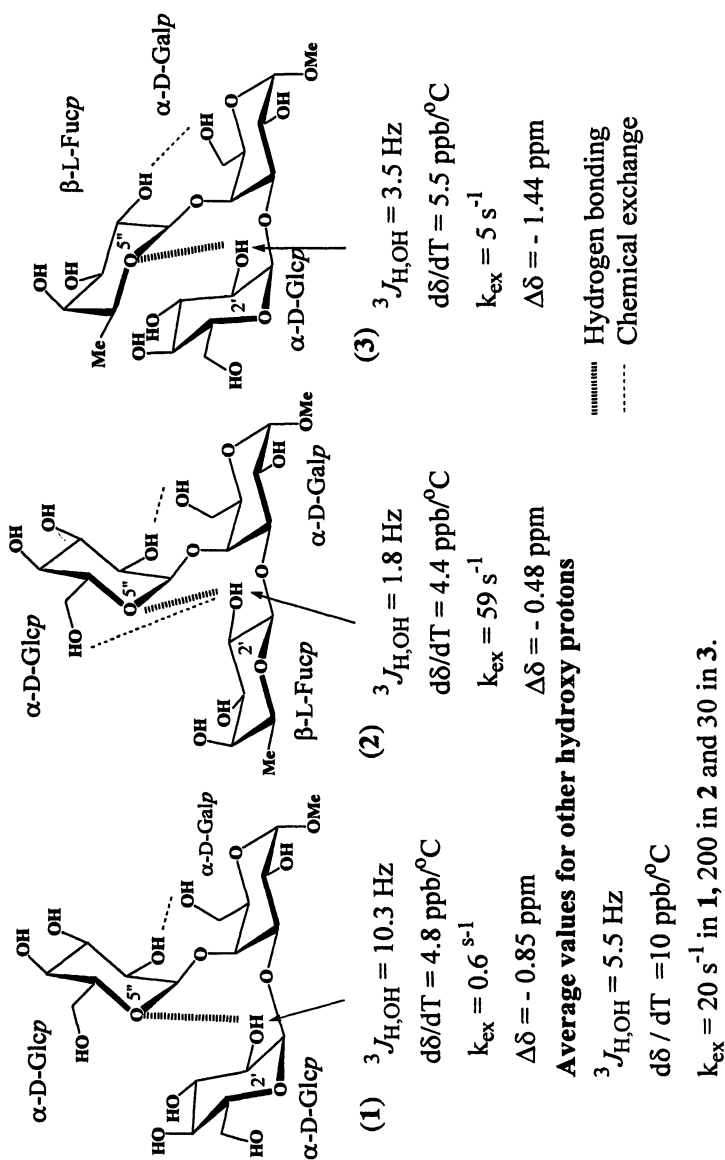


Figure 3. Schematic representation of trisaccharides **1**, **2** and **3** showing the $O(2')\text{H} - O(5'')$ interaction and the NMR parameters for the hydroxy protons.

NOE and chemical exchange

The observation of hydroxy protons by NMR allows the detection of an increased number of inter-residue NOEs. Both NOESY and ROESY are usually recorded to discriminate between cross peaks as a result of dipolar relaxation and cross peaks due to chemical exchange. In NOESY, both exchange and dipolar relaxation contributions have the same sign, and their relative contributions to a given cross peak cannot be distinguished. In ROESY, cross peaks as a result of dipolar relaxation are negative, whereas cross peaks from proton exchange are positive (the same sign as the diagonal peaks). The additional NOEs involving exchangeable protons are very important for conformational analysis since often only few inter-residue NOEs involving non-exchangeable protons are observed. Such NOEs often involve protons located very close to the glycosidic linkage that are not very sensitive to torsional fluctuations of the glycosidic linkage. On the other hand, NOEs involving hydroxy protons are not always close to the glycosidic linkage, and are therefore much more sensitive to conformational changes. In this way, the number of conformational restraints is substantially increased. For example, in a study of uniformly ^{13}C -enriched oligosaccharides (41) a gradient HSQC based technique for the resonance assignment and for the measurement of NOEs/ROEs (NOESY or ROESY-HSQC) was used. A total of 35 ROEs involving exchangeable protons were detected and assigned. The quantitative analysis of these ROEs showed that substantial torsional fluctuations exist around the glycosidic linkage, information not obtainable from observation of ring protons only.

Chemical shifts and hydration

Glycosidic shift

Series of disaccharides (42, 43) and trisaccharides (35, 44) have been investigated to determine how the chemical shifts of hydroxy proton signals could be used as probes for conformational studies. In carbohydrates, the chemical shifts of the non-exchangeable protons are mainly dependent on the type of sugar and glycosidic linkage and less on conformational effects. The effect of glycosylation on the aliphatic protons is typically a deshielding of the protons across the glycosidic bond as well as of the protons at the two neighbouring sites of the aglycon. The magnitude of the deshielding depends on type of monosaccharide, anomeric linkage, and conformation around the glycosidic bond. The main causes for this deshielding are the steric repulsion between hydrogens and the positions of oxygen lone pairs close in space to the hydrogens in question. The measurement of the chemical shifts of the hydroxy proton signals in the series of di- and trisaccharides did not demonstrate such glycosylation effects. Most of the hydroxy protons showed only minor effects by the glycosylation and the signals had $\Delta\delta$ -values (chemical shifts of the hydroxy

proton signals in the oligosaccharide minus those in the corresponding monosaccharide methyl glycoside) < 0.2 ppm. The largest $\Delta\delta$ -values were mostly measured for the hydroxyl groups adjacent to the linkage positions and the smallest for the flexible (1–6)-linked disaccharides (Table I). The hydroxy protons that occurred close to the ring oxygen of the non-reducing sugar were shielded whereas those close to another hydroxyl group were deshielded (Figure 4).

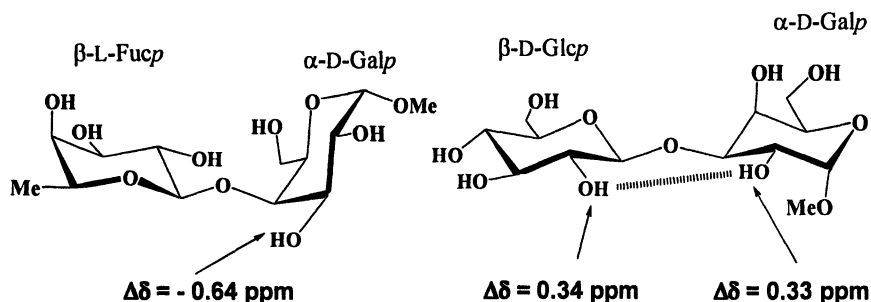


Figure 4. Schematic representations of disaccharides showing the $O(3)H - O(5')$ and $O(2)H - O(2')H$ interactions. The $\Delta\delta$ -values for the hydroxy proton signals involved in the interaction are given

In branched trisaccharides, hydroxy protons close to a neighbouring ring oxygen or close to two glycosidic oxygens were shielded (35). Since these also showed a slower exchange with water protons, the shielding was attributed to an alteration of the hydrogen bond network between the hydroxyl groups and water. To verify this hypothesis, the chemical shifts of hydroxy proton signals from three disaccharides were studied (45). NMR experiments showed (42) that in each of the disaccharides, one hydroxy proton had a large negative $\Delta\delta$ -value attributed to the proximity to the ring oxygen of the neighbouring sugar. *Ab initio* calculations (45), performed in vacuum, showed that for the minimum energy conformations, these hydroxy protons were involved in hydrogen bonding with the neighbouring ring oxygen and as expected from the dependency of the chemical shift on the delocalization of the electrons, the hydrogen-bonded hydroxy protons had large positive $\Delta\delta$ -values, indicating a deshielding relative to the monosaccharide. Since the *ab initio* calculations were performed on molecules in vacuum, the effects of the solvent on the chemical shifts were not taken into account, and the discrepancy between the sign of the $\Delta\delta$ -values obtained from calculations and experiments was attributed to interactions with the solvent.

To verify that the shielding was effectively caused by disruption of the hydrogen bond network between hydroxyl groups and water, a series of binary model systems was analyzed by ^1H NMR spectroscopy (45) (Figure 5). The solvent

Table I: $^1\text{H-NMR}$ chemical shifts (δ) and chemical shift differences $\Delta\delta$ (in parenthesis) in some 1,3-, 1,4- and 1,6-linked disaccharides.

| | O(2)H | O(3)H | O(4)H | O(6)H |
|--|----------------------------|------------------------------|------------------------------|-----------------------------|
| α -L-Fucp-(1 \rightarrow 3)- α -D-Galp-OMe | 6.01 (-0.1) 6.41 (0.25) | 5.99 (0.07) | 5.99 (0.03) 6.15 (0.21) | 6.06 (0.03) |
| α -L-Fucp-(1 \rightarrow 3)- α -D-Glcp-OMe | 5.92 (0.20) 6.19 (0.13) | 5.96 (0.05) | 5.95 (-0.01) 6.43 (0.06) | 6.01 (0.06) |
| α -L-Fucp-(1 \rightarrow 3)- α -D-Manp-OMe | 6.14 (0.03) 6.38 (0.13) | 5.95 (0.03) | 5.94 (-0.02) 6.40 (0.07) | 5.97 (-0.05) |
| β -L-Fucp-(1 \rightarrow 3)- α -D-Galp-OMe | 6.49 (0.03) 6.00 (0.15) | 6.05 (0.04) | 5.95 (0.09) 5.93 (0.01) | 6.05 (0.03) |
| β -L-Fucp-(1 \rightarrow 3)- α -D-Glcp-OMe | 6.42 (0.03) 6.00 (0.31) | 6.10 (0.08) | 6.00 (0.13) 6.49 (0.12) | 6.03 (0.07) |
| α -D-Glcp-(1 \rightarrow 3)- α -D-Galp-OMe | 6.34 (0.03) 6.19 (0.04) | 6.39 (0.04) | 6.38 (0.01) 6.00 (0.06) | 5.88 (-0.11) 6.06 (0.04) |
| β -D-Glcp-(1 \rightarrow 3)- α -D-Galp-OMe | 6.66 (0.34) 6.48 (0.33) | 6.53 (0.01) | 6.41 (-0.04) 5.82 (-0.12) | 6.04 (0.02) 6.05 (0.03) |
| β -D-Galp-(1 \rightarrow 3)- α -D-GalpNAcOMe | 6.20(-0.39) | 6.14 (-0.03) | 5.98 (0.04) 6.06 (-0.01) | 6.26 (0.08) 6.15 (0.08) |
| α -L-Fucp-(1 \rightarrow 4)- α -D-Galp-OMe | 6.48 (0.36) 6.30 (0.14) | 5.95 (0.04) 6.14 (0.17) | 5.99 (0.03) | 6.09 (0.06) |
| α -L-Fucp-(1 \rightarrow 4)- α -D-Glcp-OMe | 6.04 (0.08) 6.36 (0.04) | 5.94 (0.03) 6.22 (-0.13) | 5.94 (-0.02) | 5.91 (-0.05) |
| β -L-Fucp-(1 \rightarrow 4)- α -D-Galp-OMe | 6.49 (0.04) 6.25 (0.10) | 6.09 (0.08) 5.33 (-0.64) | 5.99 (0.13) | 6.06 (0.03) |
| β -L-Fucp-(1 \rightarrow 4)- α -D-Glcp-OMe | 6.37 (0.08) 6.41 (0.10) | 6.06 (0.05) 6.38 (0.03) | 5.92 (0.05) | 5.48 (-0.48) |
| α -D-Galp-(1 \rightarrow 4)- β -D-Galp-OMe | 6.20 (0.08) 6.31(-0.17) | 6.04 (0.09) 5.76 (-0.31) | 5.91 (0.00) | 5.80 (-0.2) 6.08 (0.01) |
| β -D-GlcpNAc-(1 \rightarrow 4)- β - D-GlcpNAc-OMe | | 6.61 (0.13) 5.94 (-0.46) | 6.60 (0.05) | 6.13 (0.05) 6.094 (0.08) |
| α -D-Glcp-(1 \rightarrow 6)- α -D-Glcp-OMe | 6.27 (0.04) 6.38 (0.06) | 6.39 (0.04) 6.37 (0.02) | 6.43 (0.06) 6.41 (0.04) | 5.90 (-0.06) |
| β -D-Glcp-(1 \rightarrow 6)- β -D-Glcp-OMe | 6.57 (0.08) 6.54 (0.11) | 6.40 (-0.11) 6.43 (-0.08) | 6.35 (-0.11) 6.47 (0.02) | 5.89 (0.13) |
| β -D-Glcp-(1 \rightarrow 6)- β -D-Galp-OMe | 6.54 (0.11) 6.44 (0.04) | 6.44 (-0.07) 6.04 (-0.03) | 6.37 (-0.08) 5.86 (0.01) | 5.91 (-0.10) |
| β -L-Fucp-(1 \rightarrow 6)- α -D-Galp-OMe | 6.45 (0.01) 6.19 (0.04) | 6.04 (0.03) 6.01 (0.04) | 5.91 (0.05) 6.05 (0.11) | |

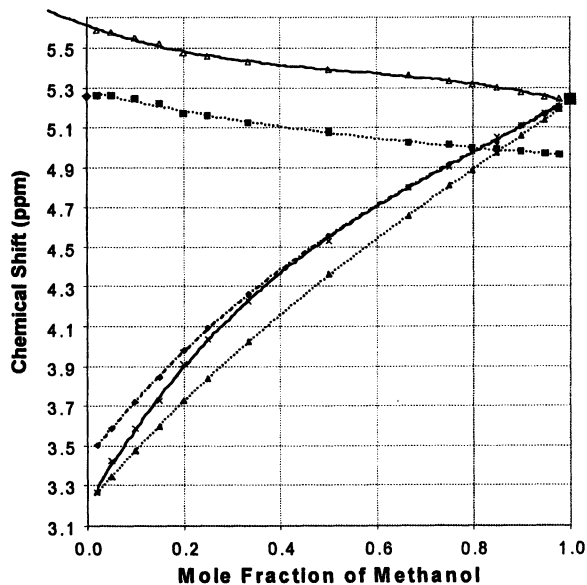


Figure 5. Chemical shift of the hydroxy proton signal of methanol as a function of the mole fraction of methanol in water (Δ), diethyl ether (\times), THF (\diamond) and dioxane (\blacktriangle). Chemical shift of the water proton signal (\blacksquare) as a function of the mole fraction of methanol. (Reproduced from reference 45. Copyright 2004.)

systems were methanol/water, methanol/diethyl ether, methanol/tetrahydrofuran and methanol/dioxane. These solvents were chosen to represent the interactions of a hydroxy proton with water and acetal oxygen. Shielding of the hydroxy proton of methanol was observed for increased ether concentrations, whereas deshielding was observed for increased concentrations of water. The effect of the water content on the chemical shift of the hydroxy proton of methanol was less pronounced as compared to that of ethers, ~ 0.3 ppm downfield shift with water compared to ~ 2 ppm upfield shift with the ethers. The magnitude of the shifts, being related to the strength of hydrogen bonding, showed that hydrogen bonding between water and methanol was more efficient than hydrogen bonding between ether and methanol. The chemical shifts of C-H signals were not significantly changed upon changes in concentration of ethers or water. Thus, the chemical shift changes measured for the OH signals were predominantly due to alteration in hydrogen bonding interactions with solvent (Figure 6).

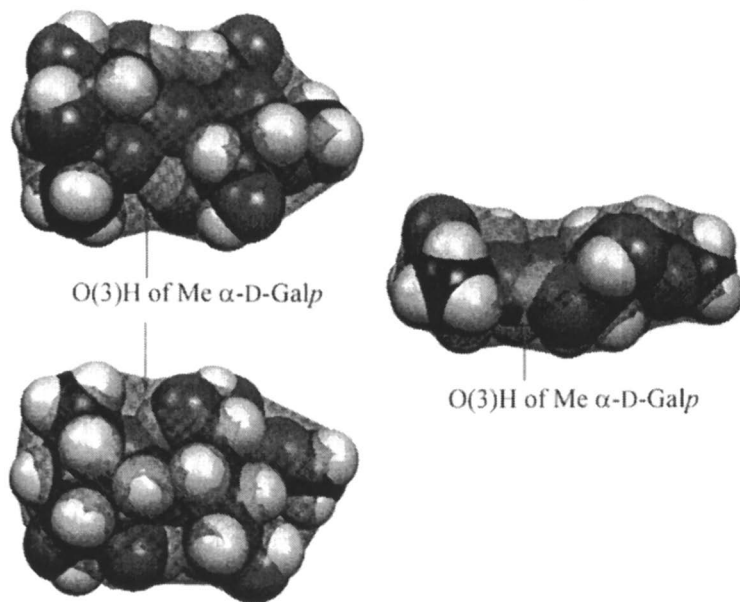


Figure 6. Space-filling representation of β -L-Fucp-(1 \rightarrow 4)- α -D-Galp-OMe in three orientations showing the location of O(3)H within the solvent inaccessible surface (slightly shaded). The $\Delta\delta$ -value of -0.64 ppm was attributed to reduced hydration. (Reproduced from reference 45. Copyright 2004.)

(See page 1 in color insert in this chapter.)

The dependence of the chemical shifts of hydroxy proton signals on hydration was found to be a common feature for disaccharides and trisaccharides with similar stereochemistry around the glycosidic bond. For example, lactose,

cellobiose and chitobiose are 1,4-linked disaccharides with similar stereochemistry at the glycosidic linkage. In these compounds, there is a short distance between O(3)H of the reducing end and O(5') of the non-reducing end, causing a shielding of ~ 0.45 ppm for the O(3)H proton (6, 43, 46).

Lewis b and Y oligosaccharides

In Lewis b (Le^b), O(3)H and O(4)H of galactose and O(4)H of fucose provide the key polar interactions with the lectin IV glycoprotein of *Griffonia simplicifolia* (47). The same hydroxyl groups were also found in the epitope of Lewis Y (Le^y) to which the monoclonal antibody BR55-2 binds (48). The NMR study of Le^b (49) and Le^y (50) showed that these hydroxy protons had large negative $\Delta\delta$ -values (Figure 7). These results were explained from the orientation of the hydroxyl groups relative to the hydrophobic and hydrophilic faces of the oligosaccharide. The hydroxy protons located at the surface of the oligosaccharide and consequently most exposed to the bulk water had small $\Delta\delta$ -values. Their interactions with the bulk water were similar to those of the monosaccharides, and accordingly so were their chemical shifts. Hydroxyl groups with large negative $\Delta\delta$ -values were situated in the vicinity of the amphiphilic region formed by the hydrophobic face of Fucp and Galp. More specifically, O(3)H and O(4)H of Galp, positioned under the hydrophobic face of Fucp, experienced the largest upfield shifts. The O(2)H of Fucp situated on the far side of the amphiphilic region was also influenced by this effect, but to a smaller extent, as seen from the $\Delta\delta$ -values.

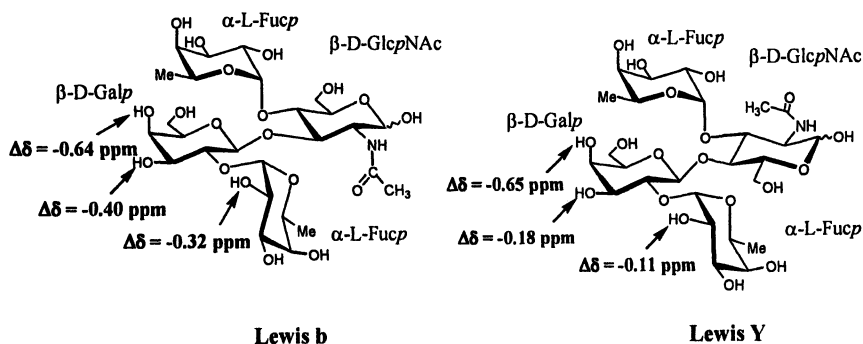


Figure 7. Schematic representations of Le^b and Le^y oligosaccharides showing the hydroxy protons with large $\Delta\delta$ -values.

Hydroxy protons in interaction studies

Cyclodextrin interactions

CDs have the ability to form inclusion complexes with specific guest molecules and are thereby potential drug carriers. They are also used as model systems in studies of enzyme-substrate interactions. The driving forces to form inclusion complexes of CDs are electrostatic, van der Waals and hydrophobic interactions and hydrogen bonding. As for intramolecular interactions, there is some doubt about the formation of intermolecular hydrogen bonds in aqueous solution because of strong hydration to hydrogen-bond sites of both the host and guest molecules. Biological systems however clearly suggest that hydrogen bonds can be formed in water when the hydrogen-bonding sites are located in a microscopically hydrophobic environment and/or situated very close to each other. Although examples have been reported about inclusion complexes of cyclodextrins where hydrogen bonds could participate in complexation, no direct evidence for formation of hydrogen bonds in water has been obtained. Usually complexation-induced $^1\text{H-NMR}$ shifts and NOEs are used to determine the geometry of the complex, but hydroxy protons can also be used as probes to study the structural changes occurring upon inclusion of a guest into the hydrophobic cavity. For example, the addition of D- or L-Trp to β -CD resulted in a downfield shift and broadening of the O(2)H and O(3)H signals, whereas the O(6)H signal was unaffected (13). These spectral modifications were attributed to the formation of hydrogen bonds between β -CD and Trp. It is indeed known (51) that cooperative hydrogen bonding causes significant stabilization of hydrogen-bonded complexes, and two-centered hydrogen bonds could be formed between O(2)H and O(3)H of β -CD and the NH_3^+ and COO^- groups of Trp.

The hydrophobic cavity of cyclodextrin and the low water solubility of adamantanes make inclusion complexes favourable in aqueous solution. α -CD was found to form 1:1 inclusion complexes with 1-adamantanol and 1-(hydroxymethyl)-adamantane (52). NMR data of hydroxy protons indicated an interaction between O(2)H of α -CD and the guest molecule. With adamantane and 2-adamantanol, that are poorly soluble in water, an additional O(3)H signal originating from the bound form of α -CD was observed. This signal was narrower, shifted upfield by more than 0.2 ppm and had a very small temperature coefficient (-1.2 ppb/ $^\circ\text{C}$ in 2-adamantanol) indicative of the reduced hydration of O(3)H upon formation of the inclusion complex. The different responses of the O(2)H and O(3)H signals to the formation of inclusion complexes was attributed to the different positions of O(2)H and O(3)H on the rim of the α -CD cavity (52).

Protein-carbohydrate interactions

The recognition of carbohydrates by proteins usually involves the formation of hydrogen bonds between hydroxyl groups of the ligand and polar groups of the receptor, and water is believed to play a central role in these associations. One way to study these intermolecular interactions by NMR should be to study the hydroxy protons. However, due to the high rate of exchange of hydroxy protons with the water protons only the ring protons are usually used for interactions studies. Also the experimental conditions used for observation of hydroxy protons of free carbohydrates are far from physiological and have been questioned for studies involving proteins.

An approach (53) based on the combined use of selective water excitation and ^{13}C -filtering has been presented. This method allows data specific to water-exchangeable protons of the bound carbohydrate ligand to be obtained via cross-relaxation to the non-exchangeable ^{13}C -coupled ring protons. By using a water-selective NOESY-HSQC experiment, the cross-relaxation between the hydroxy protons and the non-exchangeable ring protons in the complex of methyl α -D- $^{13}\text{C}_6$ -mannopyranoside with a recombinant rat mannose-binding protein was observed.

Siebert *et al.* (54) proposed the use of binary mixtures of water and DMSO to facilitate the observation of hydroxy protons and at the same time retain the protein activities and structures observed in aqueous solution. The interaction between lactose and the mistletoe lectin was chosen as a model to investigate the conformation of the bound ligand by trNOESY. Since preferential solvation affects the neighborhood of the solute, it can also affect its conformation and reactivity. Thus, the investigation of carbohydrate-protein interaction in mixtures of DMSO-water will probably require that each system is carefully tested to ensure that both the carbohydrate and the protein will retain the structure and binding capacity existing under physiological conditions.

Conclusion

The usefulness of the vicinal coupling constants, temperature coefficients and rate of exchange of hydroxy protons to probe the existence of hydrogen bonding and of their NOEs to increase the number of conformational constraints is now well established. The chemical shifts of hydroxy proton signals, so far very little used, are shown to be potential experimental probes to study the hydration properties of carbohydrates free in solution or when involved in intermolecular interactions. By combing these experimental data with results from molecular dynamics simulation of oligosaccharides in water, a better description of the preferred conformations and their interaction with water might be obtained.

It has been suggested that in carbohydrate-protein interactions, hydration can affect the affinity of restrained molecules. If it is easier to displace water molecules from the less polar regions, the chemical shifts could become conformational probes to be used to identify the hydroxy protons of a carbohydrate that could be recognized by proteins.

References

- (1) (a) Kidle, G.R.; Harris, R.; Homans, S.W. *J. Biomol. NMR* **1998**, *11*, 289-294; (b) Milton, M.J.; Harris, R.; Probert, M.A.; Field, R.A.; Homans, S.W. *Glycobiology* **1998**, *8*, 157-153.
- (2) Bose, B.; Zhao, S.; Stenutz, R.; Cloran, F.; Bondo, P.B.; Bondo, G.; Hertz, B.; Carmichael, I.; Serianni, A.S. *J. Am. Chem. Soc.* **1998**, *120*, 11158-11173.
- (3) Tiam, F.; Bolon, P.J.; Prestegard, J.H. *J. Am. Chem. Soc.* **1999**, *121*, 7712-7713.
- (4) Harvey, J.M.; Symons, M.C.R. *Nature* **1976**, *261*, 435-436.
- (5) Adams, B.; Lerner, L. *J. Am. Chem. Soc.* **1992**, *114*, 4827-4829.
- (6) Poppe, L.; van Halbeek, H. *Nature Struct. Biol.* **1994**, *1*, 215-216.
- (7) Batta, G.; Kövér, K. E. *Carbohydr. Res.* **1999**, *320*, 267-272.
- (8) Piotto, M.; Saudek, V.; Sklenár, V. *J. Biomol. NMR* **1992**, *2*, 661-665.
- (9) Sklenár, V.; Piotto, M.; Leppik, R.; Saudek, V. *J. Magn. Reson. A* **1993**, *102*, 241-245.
- (10) Freitas, L.C.G.; Cordeiro, J.M.M.; Garbujo, F.L.L. *J. Molecular Liquids* **1999**, *79*, 1-15.
- (11) Adams, B.; Lerner, L.E. *Magn. Reson. Chem.* **1994**, *32*, 225-230.
- (12) Lin, J.; Frey, P.A. *J. Am. Chem. Soc.* **2000**, *122*, 11258-11259.
- (13) Bekiroglu, S.; Kenne, L.; Sandström, C. *J. Org. Chem.* **2003**, *68*, 1671-1678.
- (14) Bendeby, B.; Immel, S.; Kenne, L.; Sandström, C. *Carbohydr. Res.* **2005**, In press.
- (15) Sheng, S.; Cherniak, R.; van Halbeek, H. *Analyt. Biochem.* **1998**, *256*, 63-66.
- (16) Nakashima, T.T.; McClung, R.E.D.; Kotovych, G. *J. Magn. Reson.* **1998**, *133*, 222-226.
- (17) Jeffrey, G.A.; Saenger, W. *Hydrogen Bonding in Biological Structures*. Springer: Berlin 1991.
- (18) Woods, R.J. *Curr. Opin. Struct. Biol.* **1995**, *5*, 591-598.
- (19) Hinton, J.F.; Wolinski, K.; in D. Hadzi (Ed), *Theoretical Treatments of Hydrogen Bonding*, Wiley, Chichester, 1997 (p 75).
- (20) Exarchou, V.; Troganis, A.; Gerotheranassis, I.P.; Tsimidou, M.; Boskou, D. *Tetrahedron* **2002**, *58*, 7423-7429.
- (21) Felemez, M.; Spiess, B. *J. Am. Chem. Soc.* **2003**, *125*, 7768-7769.

- (22) Christofides, J.C.; Davies, D.B. *J. Am. Chem. Soc.* **1983**, *105*, 5099-5105.
- (23) Fraser, R.R.; Kaufman, M.; Morand, P.; Govil, G. *Can. J. Chem.* **1969**, *47*, 403-409.
- (24) Poppe, L.; van Halbeek, H. *J. Am. Chem. Soc.* **1991**, *113*, 363-365.
- (25) Kroon, J.; Kroon-Batenburg, L.M.J.; Leeftang, B.R.; Vliegthart, J.F.G. *J. Mol. Struct.* **1994**, *322*, 27-31.
- (26) Sheng, S.; van Halbeek, H. *Biochem. Biophys. Res. Commun.* **1995**, *215*, 504-510.
- (27) Poppe, L.; Stuike-Prill, R.; Meyer, B.; van Halbeek, H. *J. Biomol. NMR* **1992**, *2*, 109-136.
- (28) van Halbeek, H.; Poppe, L. *Magn. Reson. Chem.* **1992**, *30*, S74-S86.
- (29) Brisson, J.R.; Uhrinova, S.; Woods, R.J.; van der Zwan, M.; Jarrell, H.C.; Paoletti, L.C.; Kasper, D.L.; Jennings, H.J. *Biochemistry* **1997**, *36*, 3278-3292.
- (30) Hanessian, S.; Hori, H.; Tu, Y.; Boulanger, Y. *Tetrahedron*, **1994**, *50*, 77-92.
- (31) Vanhaverbeke, C.; Heyraud, A.; Mazeau, K. *Biopolymers* **2003**, *69*, 480-497.
- (32) Sicinska, W. Adams, B.; Lerner, L. *Carbohydr. Res.* **1993**, *242*, 29-51.
- (33) Bundle, D.R.; Baumann, H.; Brisson, J.-R.; Gagné, S.M.; Zdanov, A.; Cygler, M. *Biochemistry* **1994**, *33*, 5183-5192.
- (34) Sandström, C.; Magnusson, G.; Nilsson, U.; Kenne, L., *Carbohydr. Res.* **1999**, *322*, 46-56.
- (35) Sandström, C.; Baumann, H.; Kenne, L., *J. Chem. Soc. Perkin Trans. 2* **1998**, 2385-2393.
- (36) Baumann, H.; Erbing, B.; Jansson, P.-E.; Kenne, L. *J. Chem. Soc. Perkin Trans 1* **1989**, 2153-2165.
- (37) Schneider, H.-J.; Hacket, F.; Rüdiger, V. *Chem. Rev.* **1998**, *98*, 1755-1785.
- (38) Christofides, J.C.; Davies, D.B. *J. Chem. Soc., Chem. Commun.* **1982**, 560-562.
- (39) Onda, M.; Yamamoto, Y.; Inoue, Y.; Chujo, R. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 4015-4021.
- (40) St-Jacques, M.; Sundararajan, P.R.; Taylor, K.; Marchessault, R.H. *J. Am. Chem. Soc.* **1976**, *98*, 4386-4391.
- (41) Harris, R.; Rutherford, T.J.; Milton, M.J.; Homans, S.W. *J. Biomol. NMR* **1997**, *9*, 47-54.
- (42) Ivarsson, I.; Sandström, C.; Sandström, A.; Kenne, L., *J. Chem. Soc. Perkin Trans. 2* **2000**, 2147-2152.
- (43) Kindahl, L.; Sandström, C.; Norberg, T.; Kenne, L., *J. Carbohydr. Chem.* **2000**, *19*, 1291-1303.
- (44) Sandström, C.; Baumann, H.; Kenne, L., *J. Chem. Soc. Perkin Trans. 2* **1998**, 809-815.

- (45) Bekiroglu, S.; Sandström, C.; Kenne, L.; Sandström, C. *Org. Biomol. Chem.* **2004**, *2*, 200-205.
- (46) Leeflang, B.R.; Vliegthart, J.F.G.; Kroon-Batenburg, L.M.J. van Eijck, B.P.; Kroon, J. *Carbohydr. Res.* **1992**, *230*, 41-61.
- (47) Lemieux, R.U. *Chem. Soc. Rev.* **1989**, *18*, 347-374.
- (48) BlaszczykThurin, M.; Murali, R.; Westerink, M.A.J.; Stepkowski, Z.; Co, M.S.; KieberEmmons, T. *Protein Engineering*, **1996**, *9*, 447-459.
- (49) Bekiroglu, S.; Sandström, C.; Norberg, T.; Kenne, L., *Carbohydr. Res.* **2000**, *328*, 409-418.
- (50) Bekiroglu, S.; Kenne, L.; Sandström, C. *Carbohydr. Res.* **2004**, *339*, 2465-2468.
- (51) Hayashi, T.; Miyahara, T.; Koide, N.; Kato, Y.; Masuda, H.; Ogoshi, H. *J. Am. Chem. Soc.* **1997**, *119*, 7281-7290.
- (52) Bendeby, B.; Kenne, L.; Sandström, C. *J. Incl. Phenom. Macrocyc. Chem.* **2004**, *50*, 173-181.
- (53) Sayers, E.W.; Weaver, J.L.; Prestegard, J.H. *J. Biomol. NMR* **1998**, *12*, 209-222.
- (54) Siebert, H.-C.; Andre, S.; Vliegthart, J.F.G.; Gabius, H.-J.; Minch, M.J. *J. Biomol. NMR* **2003**, *25*, 197-215.

Chapter 7

Oligosaccharides and Cellulose Crystal Surfaces: Computer Simulations

Bas R. Leeﬂang, J. Albert van Kuik[†], Loes M. J. Kroon-Bratenburg

Bijvoet Center, Department of Chemistry, Utrecht University, Utrecht,
The Netherlands

[†]J. Albert van Kuik passed away on August 18, 2003

In the present study we focus on the interaction of cellulose with hemicelluloses. As a model system the interaction of oligosaccharides derived from cellulose, mannan and xylan at a cellulose crystal surface have been studied using molecular dynamics modelling and NMR spectroscopy.

The strength of wood largely depends upon the interactions of (micro)crystalline cellulose and hemicelluloses, lignins and tanins. Native cellulose occurs in two crystallographic morphologies. Lower plants and bacteria pack the cellulose chains according to the triclinic I α phase, whereas higher plants go into the two-chain monoclinic phase referred to as I β . The latter is energetically more stable than I α . Cellulose I chains are oriented parallel.

Hemicelluloses are structural regulators in the aggregation of native cellulose (1). Various macromolecular crystalline assemblies with cellulose have been reported: Mannan crystals have been shown to grow on cellulose (2). Cellulose II grows on cellulose I (3,4,5).

The packing of the cellulose I β is such that the chains in the (200) planes are connected by hydrogen bonds and form sheets. In the bulk of the crystal, two distinct (200) planes exist: I β /even and I β /odd phase. The odd (200) planes are

shifted $1/4c$ relative to the even (200) planes. Chains in the even (200) plane are parallel to that plane, while chains in the odd planes are rotated by 11.5 deg relative to the (200) plane. However at the edge of the crystal where the polymers interact with water, this difference disappears (6,7,8). Cellulose exposes two crystal surfaces, the 110 and the 1-10 surface. Binding of hemicellulose at these surfaces of cellulose is known to occur with DP4 and larger.

In the present study we report the evaluation of the interaction with the 110 surface of cellulose with different sizes of cello-, manno- and xylo oligosaccharides using molecular modeling techniques and NMR spectroscopy (NMR data will be published elsewhere). Furthermore, we have studied the influence of parallel, antiparallel and other orientations.

Methods

The monoclinic (110) surface system was constructed as in (6,7,8). Six cellulose layers, each consisting of six chains of four cellobiose units, were placed in the center of a monoclinic periodic box with dimensions, $L_x = 3.602$, $L_y = 4.152$ nm, $L_z = 6.187$ nm, with the normal of the monoclinic (110 or 1-10) surface parallel to the z-axis; the crystallographic c-axis was chosen parallel to the y-axis. The angle between the L_x and L_z sides is then equal to the angle between the (110) and (1-10) vectors, *i.e.* $\alpha_y = 87.767^\circ$. The chains in the (200) planes are closely interconnected through hydrogen bonds and are referred to as a 'sheet' (6). The (hemi)cellulose tetrasaccharides were placed above the (110) plane in such a way that they form an extension of the (200) plane (Figure 1). The rest of the box was filled with SPC water molecules according to orientation of lattices of 'shish-kebab' structures (4).

Molecular dynamics calculations - Molecular dynamics simulations were performed using the GROMOS87 program package (9) and the improved force field for carbohydrates (10), on Linux computers. The oligosaccharide was placed at the cellulose surface. A layer of SPC (11) water molecules was added on top of the cellulose-oligosaccharide assembly to yield $L_z = 6.187$ nm. All bond lengths were kept fixed using the SHAKE procedure (12). Non-bonded interactions were calculated using the twin-range cutoff procedure with radii of 0.9 and 1.2 nm, respectively, and a time step of 2 fs. Simulations were performed with loose coupling to a pressure bath at 1 atm and a temperature bath at 300 K (13) with time constants of 2.0 and 0.1 ps, respectively. Configurations were saved every 0.2 ps. A hydrogen bond is considered to be present if $O-H...O < 2.5$ Å, and if $O-H...O$ angle $> 120^\circ$.

Inter-glycosidic torsion angles are defined as φ (Oring-C1'-On-Cn) and ψ (C1'-On-Cn-C(n-1)) and ω (On-Cn-C(n-1)-C(n-2)) according to IUPAC convention (14).

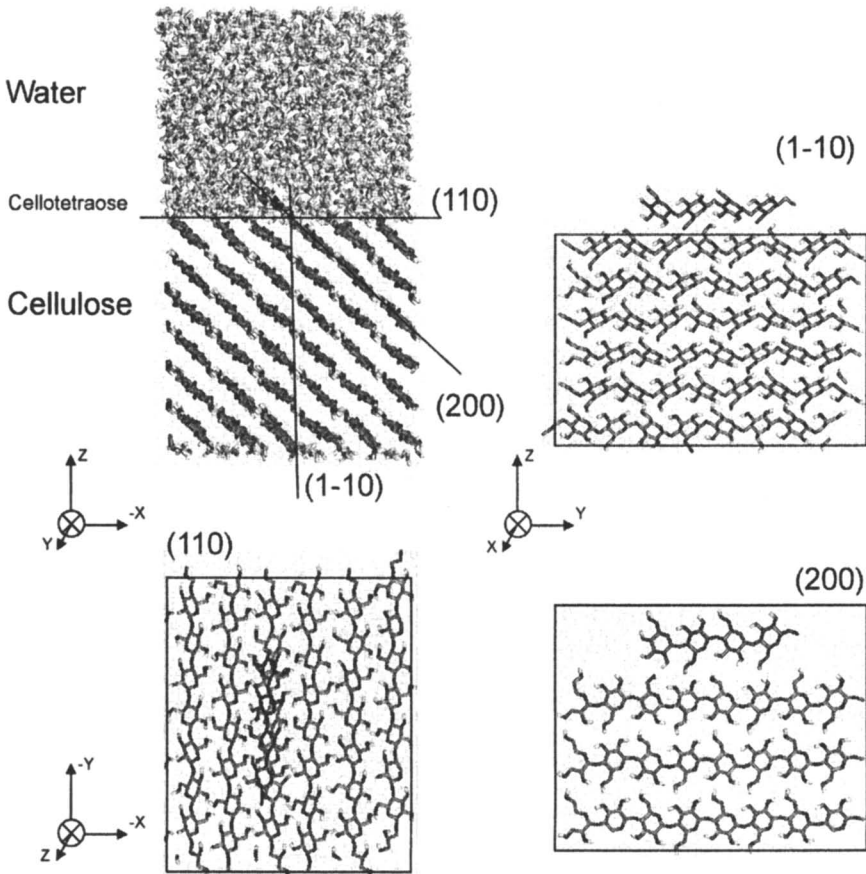


Figure 1. Cellotetraose positioned on top of the cellulose (110) plane and surrounded by a box of water (top-left) as viewed from the reducing end along the y axis, when looking perpendicular on the (1-10) plane (top-right), when looking perpendicular on the (110) plane bottom-left), and when looking perpendicular on the (200) plane (bottom-right).

(See page 1 in color insert in this chapter.)

Results

A series of MD experiments was performed with the disaccharides cellobiose, mannobiose and xylobiose that were placed initially in parallel orientation above the (110) plane of cellulose. None of these disaccharides kept this, or any other, position during a 2 ns MD simulation. Xylobiose did not even stay at the cellulose surface, but moved into the solution. From these experiments, it was concluded that disaccharides are too small to bind to the cellulose surface.

New series of experiments were performed with the tetrasaccharides cellotetraose, mannotetraose and xylotetraose. For each of these tetrasaccharides, MD experiments were performed from starting positions which have the chains oriented in parallel, antiparallel, or rotated positions, relative to the chain axes in the (110) plane of cellulose. In the series of parallel experiments, the tetrasaccharides were placed above the (110) surface of cellulose in such a way that the tetrasaccharide mimics an extension of the (200) plane (4) (Figure 1). (MD simulations are denoted CP for cellotetraose, MP for mannotetraose and XP for xylotetraose.) In the antiparallel series, the tetrasaccharide chain was placed in antiparallel fashion (it is rotated 180° relative to the parallel orientation) and MD simulations are denoted CA for cellotetraose, MA for mannotetraose and XA for xylotetraose. Finally, a series of experiments was performed where the tetrasaccharide was rotated around the axis perpendicular to the 110 surface away from its parallel orientation (See Figure 2 for cellotetraose), denoted CR for cellotetraose, MR for mannotetraose and XR for xylotetraose. This angle was initially set to 30° and was increased by 30° for each new MD experiment. Starting positions with an angle of 360° (parallel) or 180° (antiparallel) were skipped.

Experiments were run for 500 ps or 2 ns. Binding of the tetrasaccharide to the cellulose surface was considered to take place when the chain did not move from its initial (or another) position, and the van der Waals and Coulombic interaction-energy profiles were 'constant' (data not shown) during the simulation. Specific binding was considered when specific hydrogen bonds were formed during most of the simulation time, by more than one monosaccharide residue. For each of these hemicelluloses, MD simulations of 500 ps were performed. For starting positions of the parallel and antiparallel MD simulations, the tetrasaccharides were placed 6 times above the (110) surface in parallel fashion, 3 times as 'extension' of an odd (200) plane and 3 times as 'extension' of an even (200) plane. The 3 runs in one (200) plane differ in the initial position of the tetrasaccharide (i.e. 0c, 1/2c and 3/4c). MD simulations were performed for 500 ps. For at least one parallel and one antiparallel configuration, a 2 ns simulation was performed.

Cellotetraose

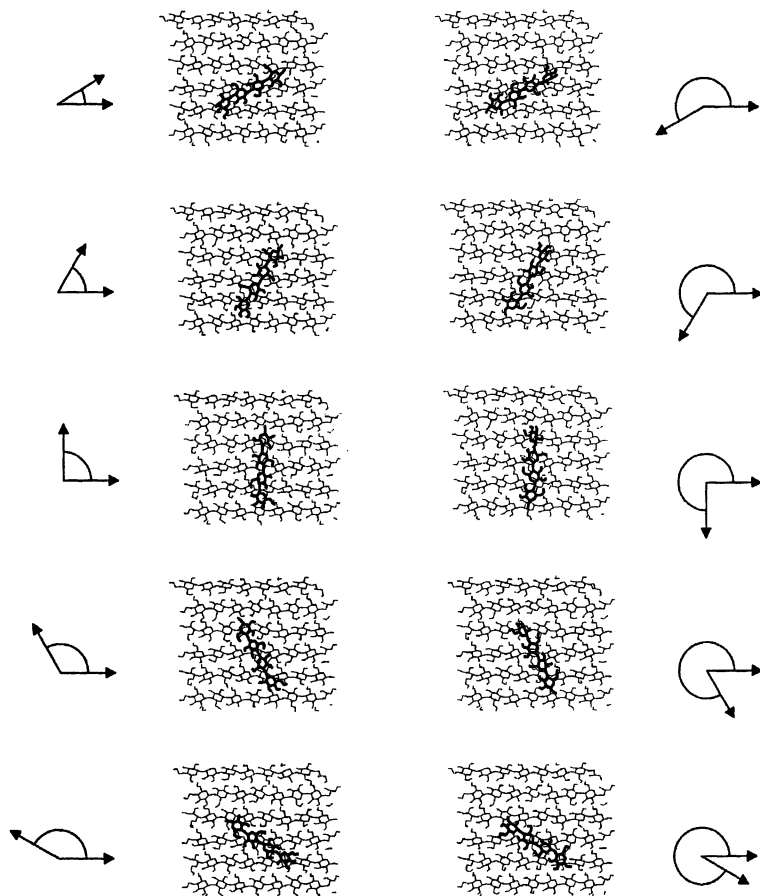


Figure 2. The orientation of cellotetraose on top of the cellulose (110) plane in the starting conformations of MD simulations CR30 – CR330.

Hydrogen bonds are considered 'specific' when formed from both tetrasaccharide-residues 1 and 3, or both 2 and 4, to cellulose, and exist for more than 50% of the simulation time. Hydroxyl groups can act both as donor or as acceptor, and in this scheme C-O-H...O-C and C-O...H-O-C are considered one hydrogen bond.

Cellulose modeling

The cellulose crystalline polymorphs fall into two types, the parallel-type of cellulose I (15) and the antiparallel-type of cellulose II (16,17), that has a packing comparable to that of the crystal structure of β -D-cellotetraose hemihydrate (18). Cellulose II crystals can grow on the cellulose I microfibril axes (4).

Cellotetraose

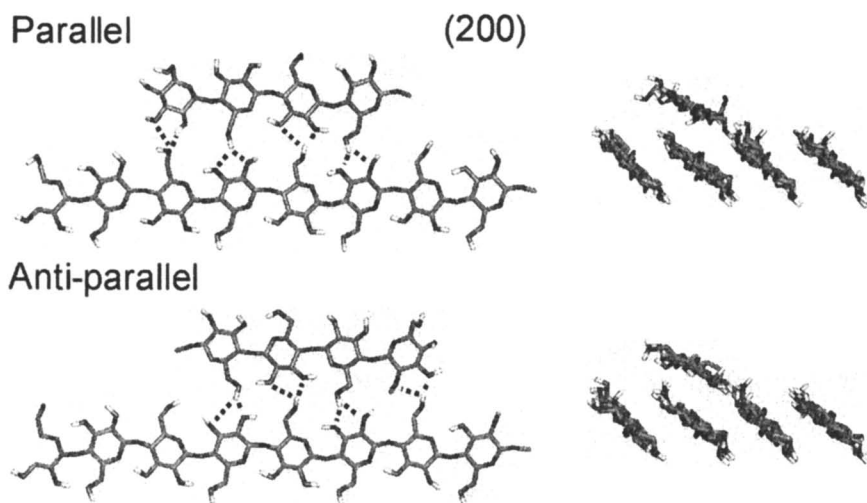


Figure 3. The position of cellotetraose in parallel (top) and antiparallel (bottom) orientation. Hydrogen bonds between cellotetraose and cellulose are depicted by dotted lines.

In parallel-binding orientation cellotetraose is positioned similarly as the cellulose chains in cellulose. In cellulose, the inter-chain hydrogen bonds are O6-H6...O2 and O2...H6-O6. Evaluating MD simulation CP shows that the inter-chain hydrogen bonds between cellotetraose and cellulose are more flexible than between cellulose chains themselves (Table 1). Inter-chain hydrogen bonds were observed from cellotetraose O6 to cellulose O2 and O3, and from cellotetraose O2 and O3 to cellulose O6. The hydrogen bonds that were formed of O2, O3, and O6 of either cellotetraose or cellulose can act either as donor or as acceptor (Figure 3) In addition, some non-specific hydrogen bonds with a neighboring cellulose chain were formed. In that case cellotetraose turns towards the (110) plane and can form hydrogen bonds with two different (both an even and an odd) cellulose chains from different (200) planes. Of the 6

(See page 2 in color insert in this chapter.)

Table 1. Hydrogen bonds between cellotetraose and cellulose.

| Cellotetraose residue | | <i>Hydrogen bond occupancy</i> | | | | |
|-----------------------|--------|--------------------------------|----------------|----------------|----------------|--------------|
| | | <i>O6...O2</i> | <i>O6...O3</i> | <i>O2...O6</i> | <i>O3...O6</i> | <i>Total</i> |
| CP | _Glc-4 | | | 0.64 | 0.34 | 0.98 |
| | _Glc-3 | 0.58 | 0.20 | | | 0.78 |
| | _Glc-2 | | | 0.71 | 0.29 | 1.00 |
| | _Glc-1 | 0.50 | 0.35 | | | 0.85 |
| CP2 | _Glc-4 | | | 0.61 | 0.43 | 1.03 |
| | _Glc-3 | 0.77 | 0.40 | | | 1.17 |
| | _Glc-2 | | | 0.28 | 0.66 | 0.94 |
| | _Glc-1 | 0.52 | 0 | | | 0.52 |

Table 2. Potential energies for the cellulose-cellotetraose interaction.

| | | | | |
|--|-----------|------------|------------|------------|
| Energy contrib. (kJ mol ⁻¹) | <i>CP</i> | <i>CP1</i> | <i>CP2</i> | <i>CP3</i> |
| Van der Waals | -239.3 | -245.7 | -234.4 | -220.4 |
| Coulombic | -106.8 | -99.5 | -99.4 | -81.7 |
| Energy contrib. (kJ mol ⁻¹) | | <i>CP4</i> | <i>CP5</i> | <i>CP6</i> |
| Van der Waals | | -219.8 | -234.8 | -193.2 |
| Coulombic | | -102.0 | -134.7 | -88.8 |

parallel simulations CP1-CP6, cellotetraose in CP2 showed this behavior the least and essentially resembled MD simulation CP.

Evaluating the 6 parallel-oriented MD simulations, denoted CP1-CP6, and considering the interaction energies, the MD simulations CP1, CP2, and CP5 showed tetrasaccharides that were binding to the (110) plane during the simulation. This is reflected by the low average energies as shown in Table 2. Due to non-specific binding to the cellulose (110) surface, changing from one binding position to another is a slow process. Therefore, it is difficult to judge a binding position from the van der Waals interaction energies only, as non-specific binding positions also can generate low van der Waals interaction

energies. The MD simulation CP2 is the single one where cellotetraose forms stable hydrogen bonds (Table 1). It is therefore the only MD simulation where the tetrasaccharide can find a parallel binding position within 500 ps. This shows that it is advantageous to use several different starting configurations. For cellotetraose, just one type of binding configuration was found for parallel orientations. CP is an extended (2ns) simulation for this configuration. From these simulation we conclude that CP2 (and CP) is the only configuration where cellotetraose binds specifically to cellulose.

Table 3. Hydrogen bonds between cellotetraose and cellulose.

| | | <i>Hydrogen bond occupancy</i> | | | | <i>Total</i> |
|------------------|--------|--------------------------------|----------------|----------------|----------------|--------------|
| | | <i>O6...O2</i> | <i>O6...O3</i> | <i>O2...O6</i> | <i>O3...O6</i> | |
| CA ^a | _Glc-4 | | | 0.21 | 0.77 | 0.98 |
| | _Glc-3 | 0.25 | 0.81 | | | 1.06 |
| | _Glc-2 | | | 0.25 | 0.83 | 1.08 |
| | _Glc-1 | | 0.11 | | | 1.02 |
| CA1 ^a | _Glc-4 | | | 0.45 | 0.78 | 1.23 |
| | _Glc-3 | 0.48 | 0.56 | | | 1.04 |
| | _Glc-2 | | | 0.47 | 0.70 | 1.17 |
| | _Glc-1 | 0.64 | 0.41 | | | 1.05 |
| CA2 ^a | _Glc-4 | | | 0.33 | 0.52 | 0.85 |
| | _Glc-3 | 0.29 | 0.53 | | | 0.82 |
| | _Glc-2 | | | 0.65 | 0.24 | 0.89 |
| | _Glc-1 | 0.35 | 0.47 | | | 0.82 |
| CA3 ^a | _Glc-4 | | | 0.33 | 0.68 | 1.01 |
| | _Glc-3 | 0.56 | 0.44 | | | 1.00 |
| | _Glc-2 | | | 0.27 | 0.91 | 1.18 |
| | _Glc-1 | 0.42 | 0.79 | | | 1.21 |
| CA4 ^b | _Glc-4 | | | 0.77 | 0.24 | 1.01 |
| | _Glc-3 | 0.82 | 0.20 | | | 1.02 |
| | _Glc-2 | | | 0.82 | 0.30 | 1.12 |
| | _Glc-1 | 0.85 | 0.13 | | | 0.98 |
| CA6 ^b | _Glc-4 | | | 0.87 | 0.16 | 1.03 |
| | _Glc-3 | 0.57 | 0.60 | | | 1.17 |
| | _Glc-2 | | | 0.75 | 0.35 | 1.10 |
| | _Glc-1 | 0.94 | 0.06 | | | 1.00 |

^aodd, ^beven

For the MD simulations of cellotetraose in antiparallel orientation, the hydrogen bonds are compiled in Table 3 and the average van der Waals

interaction energy values are accumulated in Table 4. Looking at the van der Waals interaction energy trajectory and the hydrogen bonds of the 2 ns simulation CA and the six 500 ps simulations CA1-CA6, cellotetraose binds in all MD simulations, with the exception of simulation CA5. It finds its position on the cellulose surface very quickly. All binding positions are essentially the same. As for the parallel orientation, hydrogen bonds go from cellotetraose O₆ to cellulose O₃ and O₂, and from cellotetraose O₂ and O₃ to cellulose O₆ (Figure 3) All O's can act both as hydrogen donor as well as receptor.

Table 4. Potential energies for the cellulose-cellotetraose interaction (CA).

| Energy contrib. (kJ mol ⁻¹) | CA | CA1 | CA2 | CA3 |
|--|--------|--------|--------|--------|
| van der Waals | -243.4 | -238.7 | -238.6 | -239.8 |
| Coulombic | -120.8 | -109.3 | -101.7 | -129.1 |
| Energy contrib. (kJ mol ⁻¹) | CA4 | CA5 | CA6 | |
| van der Waals | -212.9 | -180.9 | -224.0 | |
| Coulombic | -113.4 | -59.1 | -111.9 | |

Table 5. Potential energies for the cellulose-cellotetraose interaction (CR).

| Energy contrib. (kJ mol ⁻¹) | CR30 30° | CR60 60° | CR90 90° | CR120 120° | CR150 150° |
|--|----------------|----------------|---------------|---------------|---------------|
| van der Waals | -232.5 | -198.4 | -215.4 | -251.2 | -211.7 |
| Coulomb | -132.6 | -96.7 | -99.0 | -96.8 | -105.4 |
| Energy contrib. (kJ mol ⁻¹) | CR210 -150° | CR240 -120° | CR270 -90° | CR300 -60° | CR330 -30° |
| van der Waals | -253.0 | -41.6 | -145.4 | -166.2 | -243.2 |
| Coulomb | -102.6 | -50.1 | -66.0 | -104.7 | -120.4 |

To investigate whether cellotetraose could find a binding position other than with parallel or antiparallel orientation, a series of 10 MD simulations was performed with cellotetraose that had been rotated relative to its parallel orientation (Figure 2) Using a low-energy van der Waals interaction energy

trajectory as first binding assessment, possible binding occurs in MD simulations CR30 (30°), CR120 (120°), CR210 (-150°), and CR330 (-30°). In CR30 the cellotetraose rotated and finally assumed an almost parallel orientation. In CR120 cellotetraose kept its position during the first 450 ps, after which two glucose residues started to shift. During CR210 cellotetraose did rotate to an antiparallel orientation. The hydrogen bonds in the last 80 ps of these simulations corresponded to the antiparallel pattern, as described above. In simulation CR330 the tetrasaccharide attached to the cellulose surface in one (-30°) position, but did not show a specific hydrogen bond pattern. Cellotetraose in simulation CR90 did not keep its initial 90° orientation, but first briefly adopted a 60° orientation before shifting and then rotating to an orientation near 120°. Although this simulation did not show specific binding, three of the four glucose residues in cellotetraose were in line with this orientation. In simulation CR270 (-90°) cellotetraose did not find a binding position but assumed on average an orientation of -120°. In simulations CR60 (60°), CR150 (150°), CR240 (-120°) and CR300 (-60°), the tetrasaccharide maintained its original orientation without sticking to one position on the (110) surface. In most of these simulations, the plane through the cellotetraose carbohydrate rings is parallel to the (110) surface, resulting in a low average van der Waals interaction energies. However, in simulation CR300, the plane through the cellotetraose carbohydrate rings is perpendicular to the (110) surface, leading to higher van der Waals interaction energies. Nevertheless, it maintained its orientation very persistently. Therefore, van der Waals interaction energies cannot be the only factor for binding, but also the Coulomb interactions and the related hydrogen bonds.

Cellotetraose binds good in both a parallel and an antiparallel orientation, although the antiparallel orientation seems to bind better. This is corroborated by the Coulombic interaction energies and hydrogen bonds that are more favorable in the antiparallel orientation than in the parallel orientation.

Mannose modeling

The polysaccharide (β 1 \rightarrow 4)-D-mannan exists in a 2-fold helical conformation in the crystalline structural component of cell walls. Similar to cellulose, mannan is present in two polymorphs: granular mannan I and microfibrillar mannan II. Electron diffraction data showed that, unlike for cellulose, both polymorphs have antiparallel packing (19,20). Mannan I crystals can grow on cellulose I microfibrils (2).

The difference between glucose and mannose is the configuration of C2-OH where the hydroxyl is equatorial in glucose and axial in mannose. During the 2 ns MD simulation of mannotetraose in parallel orientation (MP), this property

causes the mannotetraose hydroxyl at C2 to form hydrogen bonds with two adjacent cellulose chains. To O6 of one cellulose chain and to O3 of an adjacent chain (Table 6, Figure 4). At the same time, the percentage of hydrogen bonds between mannose O6 and cellulose O3 drops below 10%, which is less than those in the MD simulations of cellotetraose in parallel orientation. The interaction with adjacent cellulose chains leads to a more horizontal positioning of the mannotetraose chain on the cellulose (110) surface, compared to CP (Figures 3 and 4) This is also reflected by the more jumpy profile of the van der Waals interaction energies of MP (data not shown).

Table 6. Hydrogen bonds between mannotetraose and cellulose.

| Mannotetraose residue | <i>Hydrogen bond occupancy</i> | | | | | | |
|-----------------------|--------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|--------------|------|
| | <i>O6...O2^a</i> | <i>O6...O3^a</i> | <i>O2...O6^a</i> | <i>O3...O6^a</i> | <i>O2...O3^b</i> | <i>Total</i> | |
| MP | -Man-4 | | | 0.30 | 0.18 | 0.66 | 1.14 |
| | -Man-3 | 0.89 | 0.08 | | | | 0.97 |
| | -Man-2 | | | 0.53 | 0.21 | 0.59 | 1.33 |
| | -Man-1 | 0.56 | 0.04 | | | | 0.60 |
| MP2 | -Man-4 | | | 0.24 | 0.09 | 0.62 | 0.95 |
| | -Man-3 | 0.16 | 0.01 | | | | 0.17 |
| | -Man-2 | | | 0.93 | 0.11 | 0.95 | 1.98 |
| | -Man-1 | 0.90 | 0.08 | | | | 0.98 |

^aTo 11 odd; ^bTo 29 even

Of the six MD simulations of mannotetraose, started from different parallel positions on the cellulose surface and with a duration of 500 ps, MP2 was the only simulation wherein mannotetraose kept its parallel orientation and gave a stable hydrogen bond profile, similar to that of MP (Table 6). During the MD simulations MP1, MP3, MP4 and MP6, mannotetraose lost its parallel orientation and starts rotating to varying degrees. Throughout simulation MP5, mannotetraose stayed in parallel orientation, but did not form stable hydrogen bond patterns, which is reflected by the relatively unfavourable Coulombic interaction energies (Table 7). Once, a chain has been rotated, it did not return to parallel orientation during the rest of the MD simulation. The simulations MP4 and MP6 were extended to 2 ns (denoted MP4e and MP6e, respectively) to see if a binding position could be found within this simulation time. For the period of 0.5 to 2 ns of the simulation of MP4e, the tetrasaccharide stayed in the same (-60°) orientation, although only -Man-4 binds to the cellulose surface (Figure 4). Simulation MP6e resulted in a -30° orientation of the tetrasaccharide (Figure 4).

Consequently, only during the MD simulations MP and MP2, mannotetraose binds to the cellulose surface in parallel orientation.

Table 7. Potential energies for the cellulose-mannotetraose interaction.

| Energy contrib. (kJ mol ⁻¹) | <i>MP</i> | <i>MP1</i> | <i>MP2</i> | <i>MP3</i> | <i>MP4</i> |
|--|-----------|-------------|------------|------------|-------------|
| Van der Waals | -233.7 | -206.9 | -218.5 | -200.4 | -186.5 |
| Coulomb | -103.7 | -74.8 | -112.4 | -75.6 | -109.6 |
| Energy contrib. (kJ mol ⁻¹) | <i>MP</i> | <i>MP4e</i> | <i>MP5</i> | <i>MP6</i> | <i>MP6e</i> |
| Van der Waals | -233.7 | -109.6 | -210.0 | -203.4 | -179.8 |
| Coulomb | -103.7 | -186.5 | -97.9 | -106.8 | -69.9 |

The binding to two adjacent cellulose chains might make it more difficult for mannotetraose to find a binding position on the cellulose surface.

A MD simulation of 2 ns with mannotetraose in antiparallel orientation to the cellulose surface (MA) showed a continuous reorientation of the chain, and ended with the chain rotated approximately 45° around the chain axis, relative to its antiparallel starting position. Of the six 500 ps simulations started from antiparallel orientation, MA1-MA6, only MA2 and MA3 showed some binding of mannotetraose according to the van der Waals interaction energy-trajectory (data not shown, average values in Table 9). These simulations were extended to 2 ns, denoted MA2e and MA3e, respectively, to investigate the persistence of this binding. In the other simulations, the tetrasaccharide did not stay in antiparallel orientation, nor kept its position at the cellulose surface. During simulation MA2e, the tetrasaccharide shifted along the parallel axis and found a stable position after 400 ps of simulation time. During the entire simulation MA3e, mannotetraose stayed bound at its original position. The final positions of mannotetraoses at the end of simulations MA2e and MA3e have shifted 2 glucose units, relative to each other and thus are identical. Similar to the parallel orientation, the hydroxyl at C2 formed hydrogen bonds with adjacent cellulose chains (Table 7, Figure 4) Simulation MA6, which showed non-binding behavior, was also extended to 2 ns (MA6e) to see whether within this time span mannotetraose would find a stable binding position. This was not the case, as the tetrasaccharide rotated between 1000 ps and 1500 ps, and for the rest of the simulation time formed many different hydrogen bonds. Because the odd and even (200) layers are slightly different, the oligosaccharide binding to them may as well be somewhat different. The three simulations that were started on an even cellulose chain did not result in antiparallel binding of mannotetraose.

Mannotetraose

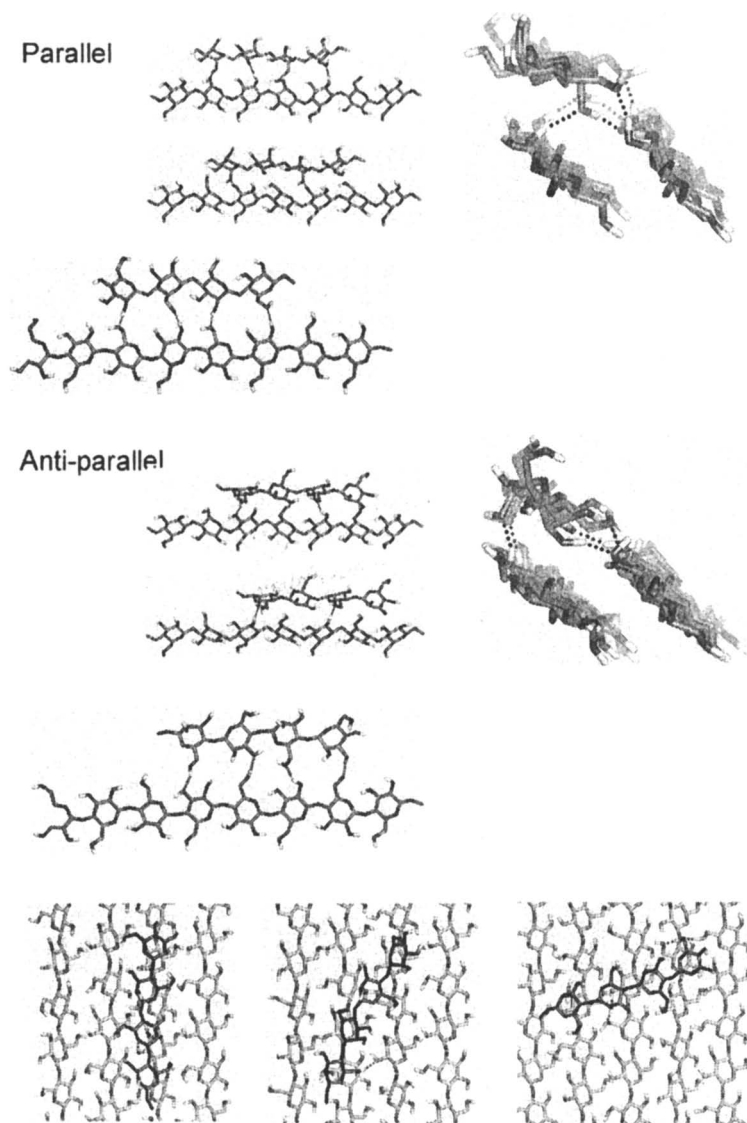


Figure 4. The position of mannotetraose in parallel (top) and antiparallel (middle) orientation. Hydrogen bonds between mannotetraose and cellulose are depicted by dotted lines. (bottom) Mannotetraose on top of the (110) plane of cellulose in antiparallel starting position (left) and after rotation of the tetrasaccharide (right) (1.5 ns simulation time) in MD simulation MA6e.

(See page 3 in color insert in this chapter.)

Table 8. Hydrogen bonds between mannotetraose and cellulose

| Mannotetraose residue | Hydrogen bond occupancy | | | | | |
|-----------------------|-------------------------|----------------------------|----------------------------|----------------------------|----------------------------|--------------|
| | | <i>O6...O2^a</i> | <i>O6...O3^a</i> | <i>O3...O6^a</i> | <i>O2...O2^b</i> | <i>Total</i> |
| MA2e ^c | -Man-4 | | | 0.32 | | 0.32 |
| | -Man-3 | 0.85 | 0.03 | | 0.88 | 1.76 |
| | -Man-2 | | | 0.28 | | 0.28 |
| | -Man-1 | 0.73 | 0.20 | | 0.75 | 1.68 |
| MA3e | -Man-4 | | | 0.80 | | 0.80 |
| | -Man-3 | 0.32 | 0.69 | | 0.82 | 1.83 |
| | -Man-2 | | | 0.46 | | 0.46 |
| | -Man-1 | 0.51 | 0.55 | | 0.45 | 1.51 |

^aTo 11 odd; ^bTo 29 even; ^cAfter 400 ps.

Table 9. Potential energies for the cellulose-mannotetraose interaction.

| | | | | | |
|---|-------------|------------|------------|-------------|-------------|
| Energy contrib. (kJ mol ⁻¹) | <i>MA</i> | <i>MA1</i> | <i>MA2</i> | <i>MA2e</i> | <i>MA3</i> |
| Van der Waals | -243.7 | -177.9 | -236.5 | -243.6 | -238.5 |
| Coulomb | -114.5 | -81.5 | -112.8 | -118.1 | -140.9 |
| Energy contrib. (kJ mol ⁻¹) | <i>MA3e</i> | <i>MA4</i> | <i>MA5</i> | <i>MA6</i> | <i>MA6e</i> |
| Van der Waals | -234.8 | -177.2 | -175.9 | -188.4 | -168.3 |
| Coulomb | -129.6 | -106.6 | -75.2 | -106.7 | -98.6 |

The ten MD simulations that were started from rotated orientations of mannotetraose (MR30 – MR330) resulted mostly in final orientations that were close to their starting orientation. The exceptions were MR60 that went from 60° to 90°, MR270 that went from -90° to almost -60°, and MR330 that went from -30° to -60°. As a result, there were three simulations (-90°, -60°, -30°) that finished with mannotetraose in -60° orientation, an orientation that was similar to the final orientation in simulation MA4e. Mannotetraose in some orientations was rather stable; though in other orientations keeps shifting, moving and rotating. The MD simulations that produced the most stable orientations of mannotetraose also had the lowest average van der Waals interaction energies (See Table 10). Those were MR30, MR60 that first went from 60° to 90°, MR90 that stayed at 90°, and MR270 that first went from -90° to almost -60°. These four MD simulations gave a low and also constant van der Waals interaction energy profile (data not shown). The rest of the MD

simulations kept essentially their starting orientations. No specific hydrogen bonds were detected for any of these simulations, which is reflected as well in the relatively weak Coulomb energies.

Table 10. Potential energies for the cellulose-mannotetraose interaction.

| | | | | | |
|--|-----------------------|-----------------------|----------------------|----------------------|----------------------|
| Energy contrib. (kJ mol ⁻¹) | <i>MR30</i> 30° | <i>MR60</i> 60° | <i>MR90</i> 90° | <i>MR120</i> 120° | <i>MR150</i> 150° |
| Van der Waals | -222.6 | -232.8 | -260.4 | -88.0 | -183.3 |
| Coulomb | -97.5 | -130.7 | -88.3 | -12.9 | -107.3 |
| Energy contrib. (kJ mol ⁻¹) | <i>MR210</i> -150° | <i>MR240</i> -120° | <i>MR270</i> -90° | <i>MR300</i> -60° | <i>MR330</i> -30° |
| Van der Waals | -169.5 | -189.2 | -33.2 | -194.5 | -184.7 |
| Coulomb | -57.3 | -90.7 | -138.3 | -74.8 | -77.9 |

Rotated mannotetraose aligns with some monosaccharides in the cellulose chains, but do not really bind in a regular manner. The hydrogen bonds are usually different for all four mannose residues, in type as well as frequency. Thus it is non-specific binding.

Conclusions: Mannotetraose binds less well to the surface than cellotetraose. When it binds, it forms hydrogen bonds with 2 adjacent cellulose chains. It seems to bind in various orientations, but starts to rotate easily with a preference for 60° (unspecific binding). However, as a general observation the antiparallel orientation of mannotetraose on cellulose has stronger specific binding than parallel.

Xylose modeling

Xylose has all its hydroxyl groups in equatorial configuration and therefore can be considered glucose without C6. Considering this property, one would presume a binding to the cellulose (110) surface similar to that of cellotetraose, except for the hydrogen bonds between the cellotetraose hydroxymethyl and cellulose. The absence of the hydroxymethyl in xylo-tetraose implies that with this model 50% of the xylose residues is not able to form hydrogen bonds with the cellulose surface. However, the preferred conformation of polysaccharide (β 1 \rightarrow 4)-D-xylan chains is different from that of glucose or mannose analogues. X-ray fiber diffraction analysis shows that xylan has a 3-fold helical conformation (21), instead of the 2-fold screw axis found for cellulose

(cellotetraose and mannotetraose). Xylopentaose binds in an approximately 3-fold helical structure to the carbohydrate-binding module CBM15 (22), and to *Pseudomonas fluorescens* xylanase A (23). In a free MD simulation with MM2Carb of xylobiose, the lowest energy conformation corresponds to 3-fold symmetry conformation with glycosidic linkage dihedral angles φ , $\psi = -69^\circ$, 152° , respectively (24). Xylotetraose in 3-fold conformation does not fit as well on the cellulose surface as in 2-fold conformation. Nevertheless, X-ray studies have revealed that xylan crystals grow on crystalline cellulose with chains in the 3-fold screw axial symmetry (25).

One MD simulation was started with xylotetraose with 2-fold screw-axial symmetry in parallel orientation on the cellulose (110) surface (XP). During the 2 ns MD run the tetrasaccharide adopted (partially) 3-fold screw-axial symmetry. Considering the van der Waals interaction energies (Table 12) and the hydrogen bonds, it did not bind specifically to one position in this simulation. Six MD runs (XP1-XP6) of 500 ps were started with xylotetraose in 2-fold screw-axial symmetry in parallel orientation on the cellulose surface. Of these simulations, only XP2 and (to a lesser extent) XP3 bound according to the van der Waals interaction energies. The MD simulation of XP2 showed 2-fold screw-axial conformation throughout the 500 ps simulation time. This simulation was extended to 2 ns (XP2e) and showed that xylotetraose bound in a 2-fold screw-axial conformation and produced a stable hydrogen bond pattern during the 2 ns simulation time. However, the hydrogen bonds (Table 11) differ from those that were observed for cellotetraose in parallel orientation (Table 1). For xylotetraose, the most prominent hydrogen bonds are from -Xyl-4 and -Xyl-2 to four adjacent glucose residues in one cellulose chain (Fig. V, top-left). Each of these xylose residues had hydrogen bonds between xylose O₂ and glucose O₂ and O₃ and between xylose O₃ to a neighboring glucose O₆ in a cellulose chain. Thus the xylan chain is shifted 1/4c (Figure 5) The MD simulations XP1, XP4, XP5 and XP6 showed for xylotetraose the (partial) formation of 3-fold symmetry and the forming of hydrogen bonds with other cellulose chains.

A MD simulation with a duration of 2 ns was started with xylotetraose in 2-fold screw-axial symmetry conformation and antiparallel orientation on the cellulose surface (XA). During the entire simulation the tetrasaccharide maintained its 2-fold symmetry and bound to the cellulose surface according to the van der Waals interaction energies (Table 14) and the hydrogen bonds (Table 13). The hydrogen bonds that were formed were different from those observed for cellotetraose in MD simulation CA. Due to the absence of a hydroxymethyl in xylose, only 2 out of 4 xylose-residues were involved in the hydrogen bond formation. Like the binding of xylotetraose in parallel orientation (XP), hydrogen bonds from one xylose residue to two glucose residues were observed (Figure 5). This hydrogen bonding pattern differs from that observed for cellotetraose in CA, where xylose O₂ binds to glucose O₃ and xylose O₃ to glucose O₂ of one glucose residue and hydrogen bond xylose O₂ to glucose O₆ of the neighboring glucose in cellulose.

Table 11. Hydrogen bonds between cellulose and xylotetraose.

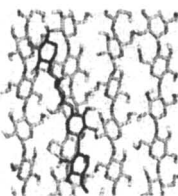
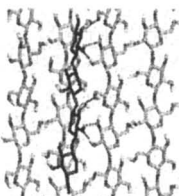
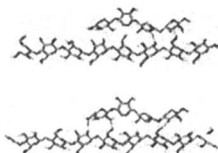
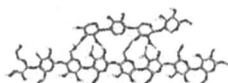
| Xylotetraose residue | <i>Hydrogen bond occupancy</i> | | | | | <i>Total</i> |
|----------------------|--------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|--------------|
| | <i>O2...O3^a</i> | <i>O3...O3^a</i> | <i>O2...O2^a</i> | <i>O3...O6^a</i> | <i>O3...O6^b</i> | |
| XP2e -Xyl-4 | 0.23 | 0.09 | 0.70 | 0.24 | | 1.26 |
| -Xyl-3 | | | | | | |
| -Xyl-2 | 0.50 | 0.13 | 0.88 | 0.50 | | 2.01 |
| -Xyl-1 | | | | | 0.60 | 0.60 |

^aTo 11 odd^bTo 29 even**Table 12. Potential energies for the cellulose-xylotetraose interaction.**

| Energy contrib. (kJ mol ⁻¹) | <i>XP</i> | <i>XP1</i> | <i>XP2</i> | <i>XP2e</i> |
|---|------------|------------|------------|-------------|
| Van der Waals | -154.4 | -192.6 | -207.1 | -199.0 |
| Coulomb | -70.3 | -67.3 | -91.6 | -98.8 |
| Energy contrib. (kJ mol ⁻¹) | <i>XP3</i> | <i>XP4</i> | <i>XP5</i> | <i>XP6</i> |
| van der Waals | -208.6 | -174.3 | -163.3 | -175.5 |
| Coulomb | -94.6 | -71.5 | -69.8 | -98.0 |

Xylotetraose

Parallel



Anti-parallel

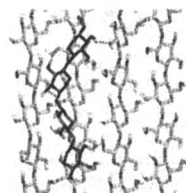
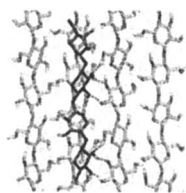
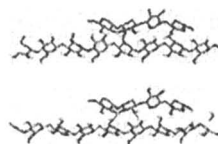
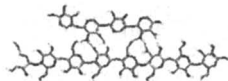


Figure 5. The position of xylotetraose in parallel (top) and antiparallel (bottom) orientation. Hydrogen bonds between xylotetraose and cellulose are depicted by dotted lines. The left column shows the 2-fold symmetry configuration (XP2e top and XA bottom) and the right column depicts the (partial) 3-fold symmetry configuration (XP1 top and XA3e bottom).

(See page 4 in color insert in this chapter.)

The result of the behavior that only -Xyl-4 and -Xyl-2 bind to the cellulose surface when in 2-fold symmetry conformation is that -Xyl-1 is very flexible, but -Xyl-3 is fixed by -Xyl-4 and -Xyl-2. Six short MD simulations (500 ps) were started with xylotetraose in antiparallel position and with 2-fold symmetry conformation, at different positions on the cellulose surface (XA1-XA6). Only during simulation XA2 the tetrasaccharide bound according to the van der Waals interaction energies (Table 14) while retaining its 2-fold symmetry conformation. Hydrogen bonds were similar to those observed for XA. Through the first 350 ps of MD simulation XA1, xylotetraose bound, but then lost its 2-fold conformation. In the other MD simulations, the tetrasaccharide quickly adopted a mix of 2-fold and 3-fold conformations. MD simulation XA3 ended with xylotetraose in 3-fold conformation as is depicted in Figure 5.

Ten MD simulations were performed with xylotetraose in rotated orientations on the cellulose surface. The rotation angle was started at 30° and increased each time by 30°, and the runs were denoted XR30 to XR330. The positions that corresponded to parallel and antiparallel binding have been presented above as XP and XA, respectively. In MD simulations XR30 (30°) and XR210 (-150°) xylotetraose did rotate to 0° (parallel) and -180° (antiparallel) orientation, respectively, but with adopting a partial 3-fold symmetry conformation. During simulation XR240 (-120°) the tetrasaccharide rotated to a position with orientation between -90° and -60°. Throughout the other simulations, xylotetraose more or less maintained its starting orientation. Xylotetraose in XR30 (to parallel), XR60, XR120, XR210 (to antiparallel), and XR150 showed binding during most of the MD simulation, judged from the van

Table 13. Hydrogen bonds between cellulose and xylotetraose

| Xylotetraose residue | | <i>Hydrogen bond occupancy</i> | | | | <i>Total</i> |
|----------------------|--------|--------------------------------|----------------------------|----------------------------|----------------------------|--------------|
| | | <i>O2...O3^a</i> | <i>O2...O6^a</i> | <i>O3...O3^a</i> | <i>O3...O2^a</i> | |
| XA | -Xyl-4 | 0.60 | 0.90 | 0.17 | 0.91 | 2.58 |
| | -Xyl-3 | | | | | |
| | -Xyl-2 | 0.80 | 0.49 | | 0.92 | 2.21 |
| | -Xyl-1 | | | | | |
| XA2 | -Xyl-4 | 0.73 | 0.63 | 0.04 | 0.94 | 2.34 |
| | -Xyl-3 | | | | | |
| | -Xyl-2 | 0.79 | 0.92 | 0.04 | 0.92 | 2.67 |
| | -Xyl-1 | | | | | |

^aTo 11 odd

^bTo 29 even

Table 14. Potential energies for the cellulose-xylotetraose interaction.

| Energy contrib. (kJ mol ⁻¹) | <i>XA</i> | <i>XA1</i> | <i>XA2</i> | <i>XA3</i> |
|--|------------|------------|------------|------------|
| van der Waals | -208.8 | -201.2 | -211.4 | -154.5 |
| Coulomb | -105.1 | -118.9 | -112.8 | -51.7 |
| Energy contrib. (kJ mol ⁻¹) | <i>XA4</i> | <i>XA5</i> | <i>XA6</i> | |
| van der Waals | -157.3 | -127.2 | -132.0 | |
| Coulomb | -54.8 | -88.6 | -39.6 | |

Table 15. Potential energies for the cellulose-xylotetraose interaction.

| Energy contrib. (kJ mol ⁻¹) | <i>XR30</i> | <i>XR60</i> | <i>XR90</i> | <i>XR120</i> | <i>XR150</i> |
|--|--------------|--------------|--------------|--------------|--------------|
| | 30° | 60° | 90° | 120° | 150° |
| Van der Waals | -182.5 | -176.9 | -157.9 | -184.1 | -208.7 |
| Coulomb | -64.2 | -94.9 | -57.8 | -80.8 | -80.1 |
| Energy contrib. (kJ mol ⁻¹) | <i>XR210</i> | <i>XR240</i> | <i>XR270</i> | <i>XR300</i> | <i>XR330</i> |
| | -150° | -120° | -90° | -60° | -30° |
| Van der Waals | -194.0 | -114.6 | -131.9 | -153.4 | -132.3 |
| Coulomb | -119 | -52.6 | -60.0 | -46.5 | -65.5 |

der Waals interaction energy profiles (data not shown, average values in Table 15). In all these simulations, xylotetraose adopts conformations with (partly) the 3-fold screw axial symmetry.

Conclusions

In the present study we investigated the interaction of three tetrasaccharides, cellotetraose, mannotetraose, xylotetraose, with the cellulose 110 surface. The binding affinity, based on the modeling study is largest for cellotetraose, followed by mannotetraose and xylotetraose, respectively.

The simulations were started in a wide variety of positions and orientations. In the case of cellotetraose the antiparallel orientations bind better than the parallel ones for cellotetraose. This preference, although less profound, is also observed for mannotetraose in interaction with cellulose. However, mannose oligosaccharides prefer to rotate around their chain axis and bind in a flat orientation to the cellulose surface

Xylose hemicellulose oligosaccharides in solution adopt a three fold screw-axial symmetry. Xylotetraose does not have a consistent binding mode. It can bind in both the two- and three-fold screw axial conformations. The latter has a significantly lower interaction affinity. There is a clear competition between the interaction energies and xylan internal energies. Simulations with the hexamer could remove the bias for a 2-fold interaction.

Although tetrasaccharides seem too small to bind in a regular fashion, it can be concluded that chains can run both parallel or antiparallel.

A single mannotetraose molecule lies flat on the cellulose surface. However, in practice crystalline domains of hemicelluloses will bind to the cellulose surface. So one should consider binding of more than one chain simultaneously. This could affect the binding mode of *e.g.* mannan that does not need to bind to two cellulose chains, because it has binding partners in its own crystal.

References

1. R. H. Atalla, J. M. Hackney, I. Uhlin, N. S. Thompson. Hemicelluloses as structure regulators in the aggregation of native cellulose. *Int. J. Biol. Macromol.* 15 (1993) 109-112.
2. Atkins E. D. T. (1992) in *Xylan and Xylanases: Progress in Biotechnology* (Visser, J., Beldman, G., van Kusters, S and Vorhagen, A. G. L., eds) Vol. 7, pp. 39-50, Elsevier Science Publishers B.V., Amsterdam.

3. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNiola, J. R. Haak. Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, 81 (1984) 3684-3690.
4. H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma. The missing term in effective pair potentials, *J. Phys. Chem.*, 91 (1987) 6269-6271.
5. A. Buléon, H. Chanzy, E. Roche. Epitaxial crystallization of cellulose II on *Valonia* cellulose. *J. Polym. Sci., Polym. Phys. Ed.* 14 (1976) 1913-1916.
6. A. Buléon, H. Chanzy, E. Roche. Shish kebab-like structures of cellulose. *Polym. Lett.* 15 (1977) 265-270.
7. A. Buléon, H. Chanzy. Single crystals of cellulose IV_{II}: preparation and properties. *J. Polym. Sci., Polym. Phys. Ed.* 18 (1980) 1209-1217.
8. H. Chanzy, M. Dube, R. H. Marchessault, J. F. Revol. Single crystals and oriented crystallization of ivory nut mannan. *Biopolymers*, 18 (1979) 887-898.
9. H. Chanzy, S. Pérez, D. P. Miller, G. Paradossi, W. T. Winter. An electron diffraction study of mannan I. Crystal and molecular structure. *Macromol.* 20 (1987) 2407-2413.
10. A. J. Duben, M. Hricovini, I. Tvaroška. The conformational analysis of methyl β -xylobioside: effect of choice of potential functions. *Carbohydr. Res.* 247 (1993) 71-81.
11. K. Geßler, N. Krauß, T. Steiner, C. Betzel, C. Sandmann, W. Saenger. Crystal structure of β -D-cellobiose hemihydrate with implications for the structure of cellulose II. *Science* 266 (1994) 1027-1029.
12. van Gunsteren, W. F.; Berendsen, H. J. C. GROMOS-87: Groningen Molecular Simulation Program Package. University of Groningen, The Netherlands, Groningen, 1987.
13. Heiner A. P., Sugiyama J., and Teleman, O. Crystalline cellulose I α and I β studied by molecular dynamics simulation. *Carbohydr. Res.* 273 (1995) 207-223.
14. Heiner, A. P., Teleman, O. Interface between monoclinic crystalline cellulose and water: breakdown of the odd/even duplicity. *Langmuir* 13 (1997) 511-518.
15. Heiner A. P., Kuutti, L., Teleman O. Comparison of the interface between water and four surfaces of native crystalline cellulose by molecular dynamics simulation. *Carbohydr. Res.* 306 (1998) 205-220.
16. IUPAC-IUB Joint commission on Biochemical Nomenclature (JCBN) *Eur. J. Biochem.*, 1983, 131, 5-7.
17. F. J. Kolpak, J. Blackwell. Determination of the structure of cellulose II. *Macromol.* 9 (1976) 273-278.
18. L. L. Leggio, J. Jenkins, G. W. Harris, R. W. Pickersgill. X-Ray Crystallographic Study of Xylopentaose Binding to *Pseudomonas Fluorescens* Xylanase A. *Proteins: Struct., Funct., Genet.* 41 (2000) 362-373.

19. J. P. Ryckaert, G. Giccotti, H. J. C. Berendsen. Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of N-alkanes, *J. Comput. Phys.*, 23 (1977) 327-341.
20. Sarko, R. Muggli. Packing analysis of carbohydrates and polysaccharides. III. *Valonia* cellulose and cellulose II. *Macromol.* 7 (1974) 486-494.
21. J. Stipanovic, A. Sarko. Packing analysis of carbohydrates and polysaccharides. 6. Molecular and crystal structure of regenerated cellulose II. *Macromol.* 9 (19976) 851-857.
22. L. Szabó, S. Jamal, H. Xie, S. J. Charnock, D. N. Bolam, H. J. Gilbert, G. J. Davies. Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation., *J. Biol. Chem.*, 276 (2001) 49061-49065.
23. T. Yui, K. Ogawa, A. Sarko. Molecular and crystal structure of konjac glucomannan in the mannan II polymorphic form. *Carbohydr. Res.* 229 (1992) 41-55.
24. S.A.H. Spieser, J.A. van Kuik, L.M.J. Kroon-Batenburg and J. Kroon, Improved carbohydrate force field for GROMOS: ring and hydroxymethyl group conformations and exo-anomeric effect, *Carbohydr. Res.* 322 (1999) 264-273.
25. L. M. J. Kroon-Batenburg , B. R. Leeftang , J. A. van Kuik , J. Kroon, Book of Abstracts, 223rd ACS National Meeting (Orlando), CELL / Xylans, Mannans, and Other Hemicelluloses: Biology, Chemistry, and Technology, Paper 148 : (Oral) Interaction of mannan and xylan with cellulose microfibrils: An X-ray, NMR, and modeling study.

Chapter 8

Biomolecular Dynamics: Testing Microscopic Predictions against Macroscopic Experiments

Andrew Almond

Department of Biochemistry, University of Oxford, South Parks Road,
Oxford OX1 3QU, United Kingdom

Structural biology has largely concentrated on producing static models, which may not be sufficient to explain many important biological phenomena. This article describes how dynamic models of carbohydrates can be constructed and investigated with a combined theoretical (prediction making) and experimental (prediction testing) approach, using computer modeling and nuclear magnetic resonance (NMR).

Introduction

Our present understanding of biomolecular three-dimensional structure is based on fifty years of experimentation. In particular, X-ray crystallography has played a key role. It was used by Watson and Crick to determine the structure of DNA in 1953, and by Perutz and Kendrew to obtain the first three-dimensional views of the proteins myoglobin and haemoglobin in 1960. Amazingly, the resultant static protein structures contained regular arrangements of α -helices and β -sheets that had been predicted by Pauling almost ten years earlier. The determination of the structure of the first enzyme by Phillips in 1965 confirmed the relationship between geometrical shape and biological activity, in accordance with Fischer's "lock and key" hypothesis. Since these pioneering studies many structures have been determined by X-ray crystallography (*1*),

driven by the notion that the geometrical arrangement of atoms can be used to understand function at the molecular level, which is a foundation of modern biochemistry. Although these ordered static structures have an appealing perfection, they do not provide a full story of behavior at the molecular level. This is realized when nuclear magnetic resonance (NMR) is used to study molecular structure (2). Although generally in agreement with X-ray crystallography, NMR often reveals that biomolecules are not static, but have local and global dynamics; in many cases this behavior is central to their function.

Complex carbohydrates, for example, are hydrophilic molecules that have a strong interaction with solvent water and little tendency to form cooperative tertiary structures. Therefore, they are often highly dynamic molecules and serve as good model systems for studying biomolecular dynamics. In mammals, they are found outside of the cell, and are normally covalently attached to other molecules. One major subclass are the glycosaminoglycans, which are polymeric (and often sulfated) sugar chains that can be found free or covalently attached to proteins; examples of these polysaccharides are hyaluronan and chondroitin sulfate. Another subclass comprise oligosaccharides that are post-translational modifications to proteins and lipids. These modifications can significantly change physical properties of the molecules they are attached to, as in the case of mucins, which are highly glycosylated proteins. The oligosaccharides can be attached to proteins through serine or threonine (O-linked) or asparagines (N-linked), and are often highly branched. Figure 1 shows a schematic representation of a typical N-linked glycan. Their cores are often elaborated with α -linked mannose residues, which can be highly branched allowing these oligosaccharides to have many termini. At the extremities sugars such as galactose and fucose, are added to generate specific epitopes that are recognized by proteins.

In the following sections methodology for investigating the dynamics of oligosaccharides is described, using a combination of NMR and aqueous molecular dynamics simulations. The application of this methodology to various fragments from oligosaccharides is detailed.

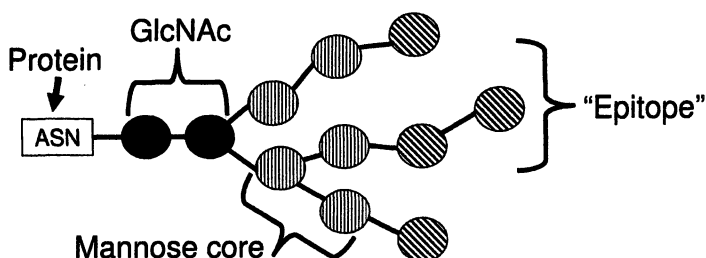


Figure 1. Sugar composition of a typical N-linked oligosaccharide from a glycoprotein, highlighting the branched mannose core and terminal sugars.

Predicting residual dipolar couplings

The dipolar coupling between nuclei is relatively strong but in solution the isotropic tumbling averages the interaction to zero. However, addition of an ordered co-solute to a molecule of interest can result in weak alignment and measurable residual dipolar couplings (RDCs) (3). These couplings can provide global structural information of vectors between nuclei (\mathbf{x}^n) relative to a molecular frame (4, 5). According to the theory of weak alignment (6), the molecular frame is correlated to the laboratory frame through an alignment tensor (\mathbf{S}_{ij}), which can be derived either by considering the distribution of experimental data (7), or theoretically using information such as geometric molecular shape (8, 9). Equation (1) shows the relationship between measured residual dipolar couplings (D^n) and their maximal value D_{\max} , which is dependent on the magnitude of alignment, the type of nuclei involved and the distance between them.

$$D^n = D_{\max} \mathbf{S}_{ij} x_i^n x_j^n \quad (1)$$

$$\mathbf{S}_{ij} = \frac{1}{2} \langle \cos \alpha_i \cos \alpha_j - \delta_{ij} \rangle \quad (2)$$

Calculation of the alignment tensor is therefore a requirement for relating residual dipolar couplings to internal molecular geometry. In this section a method is described that estimates the alignment tensor from overall molecular shape (9). The alignment tensor (more generically known as the Saupe order matrix) represents the degree of correlation between the molecular frame and the laboratory frame (direction of the magnetic field). This is shown in equation (2), where α_i is the angle between the i^{th} molecular axis and the magnetic field, and $\delta_{ij}=1$ when $i=j$, otherwise it is 0.

$$\mathbf{R}_{ij}^2 = \frac{1}{N} \sum_{r=1}^N x_i^{(r)} x_j^{(r)} \xrightarrow{\text{Diagonalize to } \mathbf{R}_D} \mathbf{V}^T \mathbf{R}^2 \mathbf{V} = \begin{pmatrix} \rho_1^2 & 0 & 0 \\ 0 & \rho_2^2 & 0 \\ 0 & 0 & \rho_3^2 \end{pmatrix} \quad \rho_1 > \rho_2 > \rho_3$$

↓

$$\text{Alignment tensor } \mathbf{S} = \mathbf{V} \mathbf{S}_D \mathbf{V}^T \longleftarrow \mathbf{S}_D = \frac{1}{2} \begin{pmatrix} 2-\delta & 0 & 0 \\ 0 & 2\delta-1 & 0 \\ 0 & 0 & -1-\delta \end{pmatrix} \longleftarrow \delta = \frac{(\rho_2 - \rho_3)}{(\rho_1 - \rho_3)}$$

Figure 2. Schematic showing how the Saupe order matrix (alignment tensor) can be estimated from geometric shape. The method involves calculation of the tensor \mathbf{R}^2 in the molecular frame, and matrix diagonalization.

Figure 2 shows a schematic outlining the method for calculating the order matrix \mathbf{S} from geometric shape. As detailed in the figure, the process can be broken down into a number of steps. The tensor \mathbf{R}^2 is used to describe the shape

of a molecule from its coordinates, and calculate the characteristic lengths of an equivalent ellipsoid. These lengths are the eigenvectors of \mathbf{R}^2 (the entries that are left when \mathbf{R}^2 is diagonalized); let these values be ρ_1^2 , ρ_2^2 and ρ_3^2 in descending order of numerical value. Simulations show that the theoretical alignment of an ellipsoid can be related to these characteristic lengths (9), through a parameter δ (see figure 2), which allows prediction of the alignment tensor in the diagonal frame of \mathbf{R}^2 , referred to as S_D in figure 2. If \mathbf{R}_D^2 is the diagonal form of \mathbf{R}^2 , then $\mathbf{R}_D^2 = \mathbf{V}^T \mathbf{R}^2 \mathbf{V}$ for some matrix \mathbf{V} (containing the eigenvectors of \mathbf{R}^2), using basic linear algebra. Consequently, the order matrix in the frame of the original molecular coordinates is given by $\mathbf{S} = \mathbf{V} \mathbf{S}_D \mathbf{V}^T$.

Example: structure and dynamics of mannose containing oligosaccharides

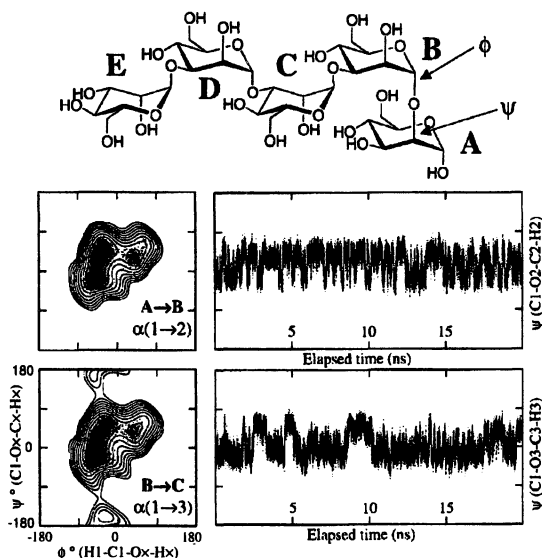


Figure 3. Chemical structure of the simulated mannose pentasaccharide and exploration of representative $\alpha(1 \rightarrow 2)$ and $\alpha(1 \rightarrow 3)$ linkages. (Adapted from (10) © 2001 American Chemical Society)

As indicated above, residual dipolar couplings provide a useful test of theoretical models of conformation. These couplings were measured in the pentasaccharide shown in figure 3 (10) using the following methodology. Half of the oligosaccharide sample was immersed in a solution of Pfl phages, which

align along the direction of the magnetic field and induce alignment along the long axis of the molecule; see figure 4 for an example conformation of the pentasaccharide. The other half of the sample was made up identically but without phages. In the aligned sample, residual dipolar couplings exhibit themselves as enhancements to the ordinary J -couplings between atomic nuclei. Figure 4 shows how typical slices from ^1H - ^{13}C HSQC NMR experiments (at natural abundance) may appear, with and without weak alignment. Subtraction of the couplings measured under these two conditions results in the residual dipolar couplings (figure 4). Measurements can be made directly from HSQC experiments, and the like, although complex lineshapes due to the plethora of proton-proton couplings and strong coupling may make specifically tailored techniques more appropriate (4). In the mannose pentasaccharide both ^1H - ^{13}C and ^1H - ^1H residual dipolar couplings were measured experimentally (10).

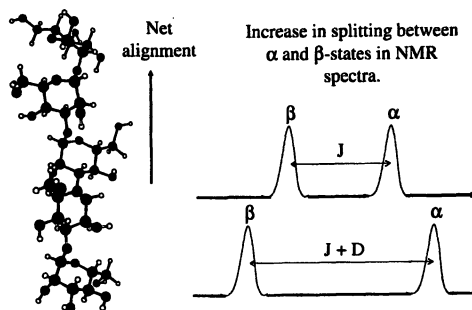


Figure 4. Left: shows an example conformation of the mannose pentasaccharide, showing the predicted net alignment. Right: two example spectra, one in free solution (top), and one in partially aligned solution (bottom).

A simulation was performed (in aqueous solution) on the linear mannose pentasaccharide (10) to obtain a prediction of its structure and dynamics. Figure 3 shows the exploration of typical $\alpha(1\rightarrow2)$ and $\alpha(1\rightarrow3)$ linkages, consistent with a reasonable amount of molecular flexibility. During this 20ns simulation no major changes of conformation were observed. Using the theory of residual dipolar couplings shown above it is possible to make predictions from the simulations that can be compared with experimental data. Simulations in aqueous solution on the nanosecond timescale can produce thousands of possible conformers, which (assuming the simulations have been performed for long enough) represent a correctly weighted reflection of the potential energy surface. Predictions of residual dipolar couplings can be made from such a simulation by calculating them for each point in the simulation, using the scheme in figure 2 and equation (1), and arithmetically averaging over the whole simulation. The result is a set of predictions that can be compared directly with those from experiments. This comparison has been performed in the graph

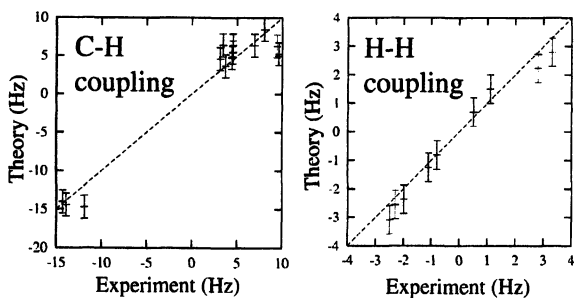


Figure 5. Correlation between experimentally measured residual dipolar couplings and those calculated from molecular dynamics simulations. (Reproduced from reference (10) © 2001 American Chemical Society).

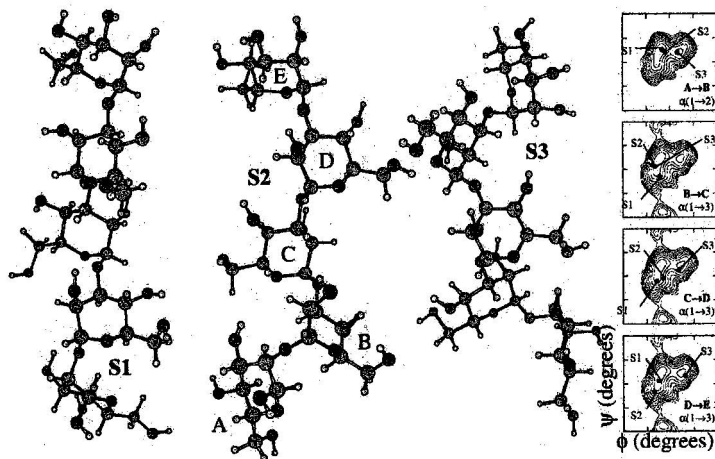


Figure 6. Example static structures for a mannose pentasaccharide, S1, S2 and S3. The plots on right show the glycosidic linkage positions on the adiabatic energy surface for each mannose linkage. (Reproduced from (10) © 2001 American Chemical Society).

of figure 5. It was found that when all linkages are in the conformation shown in figure 3, then the agreement with experiment is good. Simulations in other conformations had a lower agreement with experimental data (10).

The agreement with the dynamic model is more surprising when the residual dipolar couplings for individual conformers are compared. Figure 6 shows some examples together with the dihedral angles at each of their

glycosidic linkages. While each conformer is within the ensemble of possible structures, only S1 is in agreement with the experimental data. Therefore, although there is a static structure (a "mean structure") that is in agreement with the experimental data, slight perturbations away from this can be highly inconsistent with the data. The interpretation of its three-dimensional structure in terms of a single conformation is unlikely, and will be readdressed in a later section when this simulation is compared against relaxation data.

Example: mixed conformations - branched mannose trisaccharide

Analysis of simulations of the mannose pentasaccharide led to the conclusion that only one locally dynamic structure exists in solution. However, in simulations of other molecules it is often the case that multiple, distinct metastable conformers are predicted to coexist. In this section the methodology described above is extended to such situations. The mannose trisaccharide shown in figure 7 is an example of a molecule predicted by simulations to exist in multiple conformations (11). Figure 7 also shows the exploration of the ω -angle in the constituent $\alpha(1\rightarrow6)$ linkage during 50ns molecular dynamics simulations in aqueous solution.

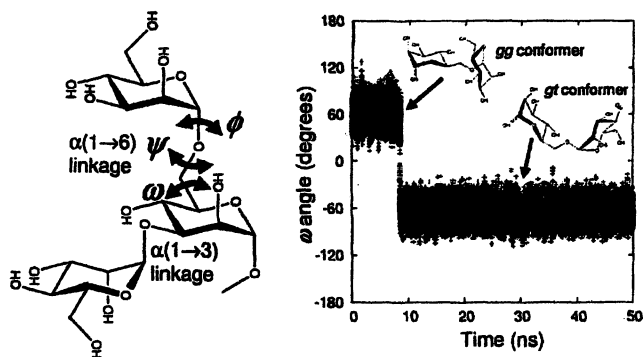


Figure 7. Left: chemical structure of the branched mannose trisaccharide with the $\alpha(1\rightarrow6)$ linkage labeled. Right: exploration of the $\alpha(1\rightarrow6)$ ω angle.

Natural abundance one-bond ^1H - ^{13}C and geminal ^1H - ^1H residual dipolar couplings were measured for the trisaccharide in figure 7. In particular, the NMR experiments were designed to obtain as many couplings as possible around the $\alpha(1\rightarrow6)$ linkage. For example, a geminal and two one-bond couplings can be measured at the 1 \rightarrow 6 linkage methylene group. As mentioned above the simulation provided evidence for two conformers at the $\alpha(1\rightarrow6)$ linkage, termed *gg* and *gt* (11); the alternate *tg* conformer was predicted to be

unstable. The overall geometric shape of these conformers is different and therefore they should align differently; the *gg* conformer is more extended than the *gt* and thus will align more strongly, and contribute disproportionately to the residual dipolar couplings (figure 8).

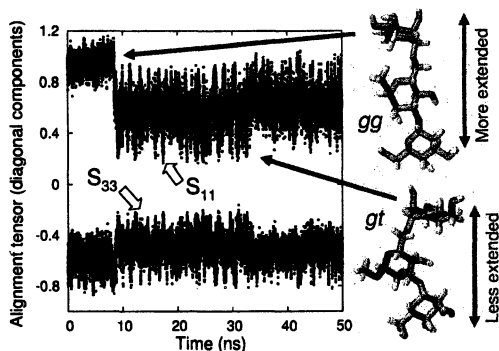


Figure 8. Calculated alignment tensor for conformations extracted from a 50ns molecular dynamics simulation. The two conformations explored during the simulation are shown.

Predictions of residual dipolar couplings were made for each conformer independently based on subsections of the simulation shown in figure 7. It was found that the two independent conformers had a low correlation with the experimental data. However, simulation data in figure 7 indicates that the two conformers may be in exchange on the nanosecond timescale; although the simulation was not long enough (with one transition) to provide an estimate for the relative population of the two conformers. Therefore, the two conformers were mixed in different relative populations and the predictions were compared against the experimental data. It was found that a mixture of 55% *gg* and 45% *gt* was the best agreement to the experimental data, and significantly better than the individual conformers (11). The utility of residual dipolar couplings for characterizing conformer populations is not widely established, and consequently comparison with a previous technique is important. Another possible means of calculating the rotamer populations at the $\alpha(1\rightarrow6)$ linkage is through vicinal coupling constants, between H5 and the two methylene protons in the linkage. Relations between the coupling constants and torsional angle have been published previously (12), and accurate measurements of these coupling constants could be obtained from one-dimensional proton spectra. Using the simulation to yield dynamic information for each of the individual conformers and these relations again allowed the rotamer populations to be estimated. Although an exact fit could not be found, a population of 36% *gg* and 64% *gt* was predicted. This is not in major disagreement with those predicted on

the basis of residual dipolar couplings; the two techniques differing by only 20% in population prediction. It should be noted, though, that the equations used for calculating coupling constants are empirical and may contain significant errors.

Predicting NMR relaxation data

Relaxation data from NMR experiments can also be used as a test of theoretical models. In this section a theory is described that allows predictions of NOESY cross-peaks to be made from simulations (13); prediction of other types of relaxation follows relatively straightforwardly. The cross-relaxation rate matrix (Γ), sometimes referred to as the relaxation matrix, that determines relaxation for N protons can be described in terms of a spectral density function $J(\omega)$, as shown in equation (3), where ω is the angular frequency.

$$\begin{aligned}\Gamma_{ij} &= \frac{1}{4} d_{ij}^2 \{6J(\omega_i + \omega_j) - J(\omega_i - \omega_j)\} & i \neq j \\ \Gamma_{ij} &= \frac{1}{4} \sum_{k \neq i}^N d_{ik}^2 \{J(\omega_i - \omega_k) + 3J(\omega_i) + 6J(\omega_i + \omega_k)\} & i = j \\ d_{ij} &= \left(\frac{\mu_0}{4\pi} \right) \gamma_i \gamma_j \hbar \langle r_{ij}^{-3} \rangle & (3)\end{aligned}$$

The relaxation matrix describes relaxation in a complex system of interacting protons. In contrast to other methods, such as the independent spin-pair approximation, this methodology is capable of predicting spin-diffusion and explicitly takes account of multiple relaxation pathways. Spin-diffusion is likely to occur in carbohydrate NOESY spectra because long mixing times are often used to provide measurable cross-peaks, thus it is worthwhile using a full relaxation matrix. The predicted NOESY intensities are contained in the solution to the differential equation (4), as described below. The solution is clearly an exponential, but because Γ is a matrix it must be solved by some tricks of linear algebra. First, the relaxation matrix is diagonalized to Γ^D , and the diagonal components are used to calculate the diagonal matrix Ξ^D , equation (5). This is back-transformed to obtain the matrix Ξ , equation (6).

$$\frac{d\Xi}{dt} = -\Gamma\Xi \quad (4)$$

$$\Xi_{ii}^D = \exp(-\Gamma_{ii}^D \tau_{\text{mix}}) \quad \text{where } \Gamma^D = \mathbf{V}^T \Gamma \mathbf{V} \quad (5)$$

$$\Xi = \mathbf{V} \Xi^D \mathbf{V}^T \quad (6)$$

The matrix Ξ , equation (6), contains the fractional cross-peak intensities, assuming the diagonal had a value of 1.0 at zero mixing time. The off-diagonal

terms in Ξ contain the cross-peak intensities, and the diagonal terms in Ξ contain the NOESY diagonal intensities. All that remains is to calculate values for $J(\omega)$, which depends on the motion of the vector joining the two vectors. Although it is the Fourier transform of the time correlation function for this vector, this can rarely be calculated because it is too complicated. Instead, Lipari and Szabo suggested that the spectral density function could be approximated from two correlation times, internal (τ_i) and overall (τ_c), and an order parameter (S^2) describing the mixture of the two (14).

$$J(\omega) = \frac{2}{5} \left(\frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1 - S^2) \tau_i}{1 + \omega^2 \tau_i^2} \right)$$

$$\tau^{-1} = \tau_c^{-1} + \tau_i^{-1} \quad (7)$$

Equation (7) shows the approximation, which assumes isotropic tumbling, among other things; its limitations, however, will not be discussed further in this article. With these relations, calculation of NOESY spectra can be reduced to estimation of the two correlation times in equation (7) and the distances in equation (3). Often it is the case that the true percentage intensity of a cross-peak is not determined experimentally. A correct methodology would be to measure the diagonal intensity at zero mixing time, and use this as 100%. In the absence of such a measure it suffices to calculate the ratio between the cross-peak intensity and the diagonal. This value can also be obtained from theoretical calculations because the diagonal intensity is also given in the matrix Ξ .

Example: studying a trisaccharide with relaxation

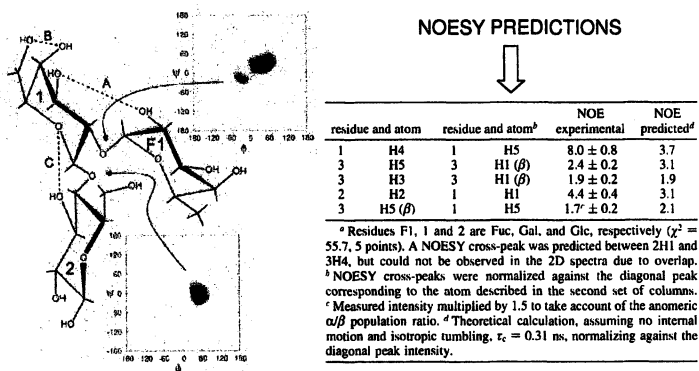


Figure 9. Left: chemical structure of the fucosylated trisaccharide and exploration in molecular dynamics. Right: the calculated NOESY cross-peak intensities. (Reproduced from (15) © 2004 American Chemical Society).

A simulation was performed on a fucosylated trisaccharide, shown in figure 9, in aqueous solution for 50ns (15). The data extracted from the simulation was used as a basis for calculation of the NOESY spectra, using the methodology described above. In this case, due to lack of additional information, the order parameter was set at 1.0. Simultaneously, NOESY spectra were recorded on the same compound with a mixing time of 800ms. The theoretical and experimental NOESY cross-peak ratios are shown in figure 9. It is evident that, even with the approximations used here, the NOESY data is in trend agreement. In particular, a cross-peak is predicted between H5 protons on residues at the two extremes of the molecule, which is also seen experimentally, and is a good conformational indicator. This is evidence that the conformation predicted theoretically is similar to that existing in solution.

Combined use of residual dipolar couplings and relaxation

| Relaxation | | | | | Residual dipolar couplings ⇔ | | | | | | |
|------------------|---------------------------|---------------|-----------------------|--------------------|------------------------------|------------|---------------|-------|-------|----------------|------|
| | | | | | assignment | exptl data | static models | | | dynamic models | |
| | | | | | | | S1 | S2 | S3 | M1 | M2 |
| NOESY cross-peak | simulation S ² | τ_c (ps) | predictn (% diagonal) | exptl (% diagonal) | | | | | | | |
| B H1-A H2 | 0.57 | 35 | 2.0 | 1.9 | A C1-H1 | 4.5 | 2.9 | -9.3 | 7.6 | 6.1 | 5.0 |
| C H1-B H3 | 0.76 | 36 | 2.6 | 2.3 | A C2-H2 | 6.9 | 9.7 | 1.0 | -17.4 | 6.0 | 4.0 |
| D H1-C H3 | 0.76 | 33 | 2.7 | 3.4 | A C3-H3 | -14.3 | -13.7 | 7.6 | 10.0 | -13.3 | -6.0 |
| E H1-D H3 | 0.70 | 41 | 1.8 | 2.6 | A C4-H4 | -13.9 | -12.2 | 6.6 | 9.6 | -13.7 | -6.1 |
| A H1-B H5 | 0.53 | 35 | 0.8 | 0.7 | A C5-H5 | -11.9 | -12.3 | 8.4 | 9.3 | -14.0 | -6.3 |
| B H2-C H5 | 0.72 | 35 | 1.3 | 1.4 | B C1-H1 | 9.6 | 8.4 | 9.9 | 9.5 | 5.0 | 4.9 |
| C H2-D H5 | 0.87 | 26 | 1.9 | 1.5 | B C3-H3 | 3.4 | 3.1 | 5.9 | -1.7 | 6.1 | 3.5 |
| C H1-C H2 | 0.90 | 30 | 1.9 | 2.0 | C C1-H1 | 4.4 | 3.8 | 6.8 | -3.0 | 4.5 | 3.2 |
| D H1-D H2 | 0.84 | 28 | 1.8 | 2.0 | D C1-H1 | 9.4 | 9.1 | 2.3 | -5.7 | 6.0 | 3.7 |
| E H1-E H2 | 0.57 | 33 | 1.0 | 1.1 | E C2-H2 | 8.0 | 8.4 | 9.8 | -2.5 | 8.0 | -2.6 |
| | | | | | E C3-H3 | 4.5 | 2.9 | -9.6 | 9.5 | 5.2 | -7.0 |
| | | | | | E C4-H4 | 3.1 | 2.7 | -14.5 | 9.9 | 4.4 | -7.3 |
| | | | | | A H1-H2 | 2.8 | 3.1 | -2.1 | -1.6 | 2.0 | 0.7 |
| | | | | | A H4-H5 | -2.3 | -2.3 | -0.5 | 0.8 | -2.4 | -1.2 |
| | | | | | B H1-H2 | -2.5 | -1.4 | -2.6 | -0.8 | -2.9 | -1.3 |
| | | | | | E H1-H2 | -2.0 | -1.5 | 2.3 | 3.2 | -2.2 | 1.2 |
| | | | | | F H2-H3 | 3.3 | 3.2 | 2.6 | -6.0 | 2.6 | 1.4 |
| | | | | | E H3-H4 | 1.1 | 1.1 | -0.2 | 1.4 | 1.4 | -1.2 |
| | | | | | E H4-H5 | -1.1 | -2.1 | -3.4 | 1.8 | -1.2 | -0.7 |
| | | | | | E H2-H4 | 0.5 | 0.4 | 0.7 | -0.8 | 0.6 | -0.3 |
| | | | | | A H1-D H1 | -0.8 | -3.4 | -6.0 | -3.0 | -0.8 | -0.5 |

Figure 10. Experimental data and theoretical predictions for NOESY cross-peaks and residual dipolar couplings. (Reproduced from (10) © 2001 American Chemical Society).

In contrast to static models, the internal parameters that constitute a dynamic molecular model frequently outnumber the observables that can be obtained from a single experiment. Therefore, it unlikely that a single experimental technique will be sufficient to comprehensively test a dynamic model, and it is preferable to use multiple experimental techniques when investigating dynamics. Ideally, each experimental technique should be complementary to the others and hence provide significant new information. Two such techniques are NMR residual dipolar couplings and relaxation, which detect global and local dynamics respectively.

Simulations of the mannose pentasaccharide (figure 3), discussed previously, were found to be in agreement with residual dipolar couplings (figure 10, dynamic model M1). Although a static structure was also found to be

in agreement with this experimental data (figures 6 and 10, model S1), it proved incapable of predicting the NOESY experimental data using the theory described above. However, the simulation was effective in predicting the NOESY data, while simultaneously being able to predict the residual dipolar coupling data, as shown in figure 10. Therefore, although a static structure may be capable of describing the data produced by a single experimental technique, it may not be able to predict the results from multiple types of experiments. This is because each experiment produces a different average measure of dynamics. On the one hand, residual dipolar couplings sense average changes in overall shape, and local bond motions. On the other hand, NOESY senses average local distances and dynamics, as well as the overall molecular tumbling. To agree with both sets of data a dynamic model is necessary, which is averaged according to the relevant theory. According to the experimental data presented here the mannose pentasaccharide has a dynamic, rather than static, structure and the molecular dynamics simulation appears to be a reasonable model of this.

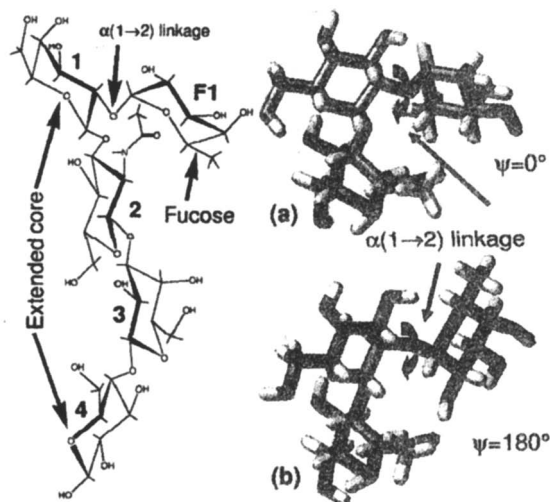


Figure 11. Left: the fucose pentasaccharide, with the sugar labeling nomenclature. Right: two conformations of the $\alpha(1 \rightarrow 2)$ fucose linkage.

It is apparent that some experimental techniques may be better at characterizing particular types of molecular conformation and dynamics than others. This can be exemplified by studying the fucosylated oligosaccharide (shown in figure 11) with both residual dipolar couplings and NOESY. With the aim of characterizing the fucose conformation, residual dipolar ^1H - ^{13}C one-bond couplings were measured for this molecule (again with the aid Pfl1 phage). Simulations showed that the oligosaccharide core (the sugar residues other than

the fucose) were likely to form a rather extended conformation. Therefore, according to the theory described previously, this linear, elongated core should dominate the alignment. Comparison of the simulation with the residual dipolar couplings showed this to be the case (figure 12). However, with this alignment, two likely conformations of the fucose residue (see figure 11) are indistinguishable from measurements of residual dipolar couplings (RDCs). To resolve this, molecular dynamics simulations were performed on the fucosylated pentasaccharide in these two conformations, and cross-relaxation was measured from NOESY spectra. Analysis of the theoretical data shows that only one of the two conformers shown in figure 11, labeled (a) with $\psi=0^\circ$, is consistent with both sets of experimental data. Consequently, the simulation of this conformer is in agreement with the combined experimental data, as found for the mannose pentasaccharide described above. However, in this case, the residual dipolar couplings alone are not sufficient to pin down the likely conformation, and again shows the advantages of using multiple experimental techniques. The result is a realistic dynamic model of the molecule, where the fucose termini is in a relatively ordered conformation, consistent with its function as a recognition motif for proteins.

| residue & atom | residue & atom ^a | H1 | NOE experimental | | NOE theoretical ^r | | residue & atom | residue & atom | RDC experiment | RDC theory | | |
|----------------|-----------------------------|----|------------------|--------------------|------------------------------|--------------------|----------------|----------------|----------------|------------|-------|-------|
| | | | $\psi = 0^\circ$ | $\psi = 180^\circ$ | $\psi = 0^\circ$ | $\psi = 180^\circ$ | | | | | | |
| F1 | H2 | F1 | H1 | 5.5 ± 0.6 | 6.9 | 8.7 | F1 | C1 | F1 | H1 | -20.4 | -20.7 |
| F1 | H5 | F1 | H1 | 0.6 ± 0.1 | 0.4 | 0.6 | F1 | C2 | F1 | H2 | 1.6 | 4.8 |
| F1 | H3 | F1 | H5 | 4.9 ± 0.5 | 6.5 | 6.1 | F1 | C3 | F1 | H3 | 2.7 | 4.5 |
| F1 | H4 | F1 | H5 | 5.8 ± 0.6 | 7.5 | 6.9 | F1 | C4 | F1 | H4 | -19.3 | -18.8 |
| 1 | H5 | 1 | H1 | 6.2 ± 0.6 | 6.8 | 7.9 | F1 | C5 | F1 | H5 | 6.6 | 5.9 |
| 1 | H2 | 1 | H1 | 2.5 ± 0.3 | 1.8 | 2.4 | 1 | C1 | 1 | H1 | 1.1 | -1.6 |
| 1 | H3 | 1 | H1 | 5.7 ± 0.6 | 5.0 | 4.9 | 1 | C2 | 1 | H2 | 4.8 | 1.4 |
| 1 | H4 | 1 | H5 | 5.8 ± 0.6 | 7.2 | 6.6 | 1 | C3 | 1 | H3 | 3.1 | -0.4 |
| 2 | H2 | 2 | H1 | 2.2 ± 0.2 | 1.8 | 1.9 | 1 | C4 | 1 | H4 | 15.5 | 5.5 |
| 2 | H3 | 2 | H1 | 7.0 ± 0.7 | 2.1 | 3.5 | 2 | C1 | 2 | H1 | 19.0 | 14.8 |
| 2 | H5 | 2 | H1 | 7.7 ± 0.8 | 9.9 | 9.0 | 2 | C2 | 2 | H2 | 21.3 | 17.0 |
| 4 | H3 | 4 | H1 | 3.0 ± 0.3 | 4.0 | 4.0 | 2 | C3 | 2 | H3 | 22.0 | 19.5 |
| 4 | H5 | 4 | H1 | 7.8 ± 0.8 | 6.4 | 7.0 | 2 | C4 | 2 | H4 | 20.4 | 17.6 |
| 1 | H2 | F1 | H1 | 6.8 ± 0.7 | 5.8 | 0.9 | 2 | C5 | 2 | H5 | 17.7 | 17.1 |
| 1 | H3 | F1 | H1 | 0.5 ± 0.1 | 0.7 | 4.3 | 3 | C1 | 3 | H1 | 13.0 | 9.7 |
| 1 | H2 | F1 | H5 | 1.3 ± 0.1 | 2.4 | 0.5 | 3 | C2 | 3 | H2 | 10.9 | 10.1 |
| 2 | H2 | F1 | H5 | 6.8 ± 0.7 | 4.7 | 0.0 | 3 | C3 | 3 | H3 | 13.0 | 11.6 |
| 2 | H4 | F1 | H5 | 0.6 ± 0.1 | 2.5 | 0.0 | 3 | C4 | 3 | H4 | -0.1 | 2.5 |
| 2 | H3 | 1 | H1 | 5.7 ± 0.6 | 5.6 | 4.3 | 3 | C5 | 3 | H5 | 8.2 | 10.5 |
| 3 | H1 | 2 | H1 | 0.4 ± 0.1 | 0.3 | 0.3 | 4 | C2 | 4 | H2 | 0.2 | 1.8 |
| 3 | H3 | 2 | H1 | 11.5 ± 1.2 | 9.8 | 9.5 | 4 | C3 | 4 | H3 | 2.9 | 2.2 |
| 3 | H4 | 2 | H1 | 0.8 ± 0.1 | 1.1 | 1.1 | 4 | C4 | 4 | H4 | 3.6 | 4.7 |
| 3 | H3 | 2 | H5 | 0.5 ± 0.1 | 0.9 | 0.7 | 4 | C5 | 4 | H5 | 6.5 | 3.4 |
| 4 | H4 | 3 | H1 | 11.0 ± 1.1 | 11.3 | 9.6 | 4 | | | | | |

Figure 12. Comparison of experimental data and theoretical predictions for the oligosaccharide in figure 11; NOESY (left) for $\psi=0^\circ, 180^\circ$ and RDCs (right) for $\psi=0^\circ$ (adapted from (15) © 2004 American Chemical Society).

Summary

This article describes methodologies for calculation of experimental parameters from molecular dynamics simulations. In particular, it is shown how

a simple model of alignment can be used to predict residual dipolar couplings and how a suitably calculated relaxation matrix can be used to make predictions of NOESY spectra. These predictions can be used to test dynamic models against experimental measurements. However, individual experimental techniques can have limitations and idiosyncrasies. Residual dipolar couplings, for example, provide global molecular information, but may not provide an unambiguous characterization of conformation. Relaxation, on the other hand, provides only local information, and may provide not information for certain regions of the molecule if NMR active groups are sparse. Dynamic models are also inherently complicated and possess many internal parameters, and hence it is advantageous to use multiple experimental techniques to test them. Suitable application of these techniques allowed dynamic models of oligosaccharides containing mannose and fucose to be tested, providing new views of these molecules that can be used to infer and interpret their biological functions.

References

1. Campbell, I. D., *Nat Rev Mol Cell Biol* **2002**, *3*, 377-81.
2. Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E., *Biochemistry* **1994**, *33*, 5984-6003.
3. Tjandra, N.; Bax, A., *Science* **1997**, *278*, 1111-1114.
4. Prestegard, J. H.; Bougault, C. M.; Kishore, A. I., *Chem. Rev.* **2004**, *104*, 3519-3540.
5. Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J. R., *Q. Rev. Biophys.* **2000**, *33*, 371-424.
6. Saupé, A.; Englert, G., *Phys. Rev. Lett.* **1963**, *11*, 462-465.
7. Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H., *J. Magn. Reson.* **1999**, *138*, 334-342.
8. Zweckstetter, M.; Bax, A., *J. Am. Chem. Soc.* **2000**, *122*, 3791-3792.
9. Almond, A.; Axelsen, J. B., *J. Am. Chem. Soc.* **2002**, *124*, 9986-9987.
10. Almond, A.; Bunkenborg, J.; Franch, T.; Gotfredsen, C. H.; Duus, J. Ø. *J. Am. Chem. Soc.* **2001**, *123*, 4792-4802.
11. Almond, A.; Duus, J. Ø. *J. Biomol. NMR* **2001**, *20*, 351-363.
12. Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C., *Tetrahedron* **1980**, *36*, 2783-2792.
13. Macura, S.; Ernst, R. R., *Mol. Phys.* **1980**, *41*, 95-117.
14. Lipari, G.; Szabo, A., *J. Am. Chem. Soc.* **1982**, *104*, 4546-4559.
15. Almond, A.; Petersen, B. O.; Duus, J. Ø. *Biochemistry* **2004**, *43*, 5853-5863.

Chapter 9

Structural Insights: Toward a Molecular Comprehension of the GH13 Amylase Specificity

Gwénaëlle André-Leroux

INRA-Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

Despite the fact that the α -amylase family (GH13) displays a similar $(\alpha/\beta)_8$ architecture and a consensus α -retaining mechanism on the α -D-glucose polymer, the whole family of α -amylase enzymes displays a wide range of reactions. Our hypothesis is that hydrolysis *vs* transglycosylation could be governed by specific amino-acid residues located at subsites -1 and +1 but this hypothesis lacks i) structural comprehension at a molecular level and ii) mutational support (2). This chapter is dedicated to structural characterization and binding studies of the starch / glycogen branching and debranching enzymes. A relevant exploration of the binding of the two enzymes that used linear and branched oligosaccharides has identified two signature motifs which are highly conserved within each subfamily. They could act as finger prints that could create an extra sugar binding space at the catalytic subsites -1 and +1 of the branching enzyme. Finally, natural but pivotal mutations reveal the wide array of catalytic specificities within family GH13.

Introduction

The α -amylases family 13 (1) (<http://afmb.cnrs-mrs.fr/CAZY/>) is a large and highly studied family of enzymes that displays a conserved $(\alpha/\beta)_8$ architecture and an α -retaining catalytic mechanism on the α -D-glucose polymer. Among them, the starch/glycogen branching enzyme (α -(1,4)-glucan: α -(1,4)-glucan 6-glucosyl-transferase; EC 2.4.1.18) (BE)¹ and the starch/glycogen debranching enzyme (glycogen 6-glucanohydrolase; EC 3.1.68) (DBE)¹ provide the structure and solubility of starch in plants or of glycogen in animals and bacteria. Accumulation of insoluble glycogen in the cell is known as glycogen storage disease type IV (GSD IV) and is caused by mutations in the BE gene (3). Mutations of the impaired enzyme prevent the formation of the branch point thus leading to an insoluble polymer in the liver, the muscular tissues and the peripheral and central nervous system. The branching enzyme excises the terminal section of a pre-existing α -(1,4) linked chain and transfers the resulting glucan fragment to an acceptor glucan chain, creating an α -(1,6) glycosidic linkage. The debranching enzyme, also called isoamylase, catalyzes the hydrolysis of α -(1,6) glycosidic linkages, specific to amylopectin and glycogen. Both enzymes are among a few members within the glycosyl-hydrolase family that are able to accommodate sugars in the α -(1,6) position.

In general, the first step of catalysis consists in the cleavage of the glycosidic bond between subsites -1 and +1, leading to a covalent intermediate at subsite -1 and the departure of the leaving group. In the branching enzyme, the cleaved glycosidic bond is α -(1,4), whereas in the isoamylase it is an α -(1,6) bond. The sugar chain of the covalent intermediate is called the donor sugar and is bound at the donor site, starting at subsite -1 to -n. In the next reaction step, an acceptor molecule replaces the leaving group and occupies the acceptor site from +1 to +m. The acceptor is a water molecule in the hydrolysis reaction and a sugar molecule in the transglycosylation reaction respectively. On the basis of this reaction mechanism, the α -amylase specificity can be ascribed to donor/acceptor site affinities. For example, CGTases have acceptor specificity since they display a ~100 fold higher affinity for free sugars compared to a water molecule, thus favoring the transglycosylation or disproportionation reactions *vs* hydrolysis (4). Similarly, the branching enzyme that favors sugar as acceptor displays transferase activity whilst isoamylase that favors the water molecule thus shows hydrolytic activity.

Branching and debranching enzymes are folded into modular domains, one or more NH2 terminal domain, a carboxyl-terminal domain and a central $(\alpha/\beta)_8$

catalytic domain. To date, native structures of one branching and one debranching enzyme have been solved by X-ray diffraction (5, 6). The *E. coli* branching enzyme contains 728 residues. The crystallographic structure displays a truncation at amino acid 112 and two disordered regions: 361-373 and 414-429. From chimeric studies, the COOH-terminal domain is involved in the substrate preference and catalytic capacity, while the NH₂-terminal domain is deeply involved in the length of the chain transferred, ranging from 5 to 23 glucose units (7). The 3D structure of *Pseudomonas amyloclavata* isoamylase has 750 amino-acid residues and an additional novel domain in the N-terminal region. Despite the fact that the two enzymes are structurally similar as they both miss the $\alpha 5$ helix and show additional turn in helices $\alpha 1a$, $\alpha 6a$, $\alpha 7a$ compared to the regular catalytic barrel, they evidence differences in the loops surrounding the active site. These loops, significantly shorter in BE, could be responsible for the distinct catalytic properties between the two enzymes. The catalytic triad in *E. coli* BE is numbered D_{BE}405, E_{BE}458 and D_{BE}526, its isoamylase homologue is numbered D_{DBE}375, E_{DBE}435 and D_{DBE}510 respectively (numbering without the promoter region).

To gain molecular understanding of α -(1,4)/ α -(1,6) hydrolysis vs. α -(1,4) transglycosylation specificity, we combined 1D sequence analysis and extensive 3D computational docking. Various substrates from short linear to branched long sugar chains were optimally docked in *E. coli* BE and *Ps. amyloclavata* enzyme. Signature motifs that are structurally involved in the binding were identified from the sequences and mapped onto the structures. Interestingly, both variations in substrate docking and differences in sequence patterns could be ascribed to the subsite specificity. The chapter will focus on understanding the specificity of branching and debranching enzymes in order to improve the comprehension of the wide range specificities of amylases, neopullulanases, or CGTases in general.

Computational modeling

Bioinformatics

For retrieval of primary sequence, we used SRS at the EBI (8) (<http://srs.ebi.ac.uk>). The GLGB_ECOLI sequence was used as a query for a Blast exploration of the whole Swall databank. The default parameters were used: score matrix Blosum62 and E value 10,000. Similarly, the ISOA_PSEAY

sequence was used for a Blast query. The two queries were combined as a merge with no redundancy. A clustalW analysis was performed on the set using the following parameters: aligned sequences order, slow pair wise algorithm, 10-gap penalty and 0.05 gap extension. From the set of aligned sequence families, clusters of α -amylases appeared with highly conserved regions. We analyzed the most similar sequences inside these clusters.

Molecular modeling

Molecular modeling was carried out on Silicon Graphics computers with the Accelrys packages (Accelrys. Inc, San Diego, CA, USA). Molecular displays and energy minimizations were performed with Insight II, using the Biopolymer and Discover modules respectively. For all calculations, the CFF91 force field and the steepest descent algorithm were selected. This force field is adapted to the protein / polysaccharide interaction studies and has been widely used (9 -11).

Coordinates of enzymes

X-ray coordinates are very useful in attempting to deduce the catalytic mechanism of enzymes so they have been used as the starting point of our modeling study. The crystal structures of the *E. coli* branching enzyme and the *Ps. amyloclavata* debranching enzyme were solved by X-ray crystallography diffraction at 2.3 Å (5) (RCSB id: 1m7x) and 2.2 Å (6) (RCSB id 1bf2) respectively. To save cpu time, the two enzymes were truncated at their C-terminal domain. The 1m7x.pdb structure displays two truncations, due to crystallographic disorder, at loops 3 and 4, from S360 to T372 and from R413-G428 respectively. However they are located in the catalytic $(\alpha/\beta)_8$ barrel and the best_fit (GCG) provides a nice alignment with the isoamylase bf2.pdb whose 3D structure is well defined in those regions (data not shown). The missing parts of the truncated loops could be homology modeled. They were then refined by successive rounds of minimization.

Coordinates of complexes

Maltose and isomaltose were positioned to span subsites -1 and +1 of the active sites (11-13). Those oligosaccharides are known from the literature to display four and six low energy conformations respectively (14, 15). These conformations were tested in the environment of the active site of the two enzymes. The best docking for each dimer was kept. Then, stepwise binding of glucose moiety, linked in α -(1,4), was performed. To reduce the number of solutions (in theory $4 \times 4^{n-1}$ for a linear chain with n residues added), only the

conformations that could combine hydrogen bonds, van der Waals interactions and no big steric clashes with the cleft of the enzymes were kept. During minimization, the backbones of the proteins were fixed to preserve their integrity whereas the side chains and the glucosyl residues were completely free to adapt. The potential energies of enzyme and substrate were evaluated and their interaction energy too. Eventually through this set of energetic parameters, the complexes were selected, the amino-acid residues involved in the subsites were listed and the subsite affinities were measured. We ended up with a complete filling of the catalytic crevice with a final amylose or glycogen fragment.

Discussion

Homology modeling of BE loops 3 and 4

BE (1m7x) loops 3 and 4 that lack 11 and 15 amino acid residues respectively display a folding in respect to their template fragments of DBE loops 3 and 4. At this stage, both enzymes share a complete $(\alpha/\beta)_8$ catalytic barrel with a superimposed catalytic triade as shown on Figure 1(a) and (b).

Maltose and isomaltose docking

Maltose and isomaltose represent the shortest units of linear amylose and branched glycogen homopolymers respectively. Their docking in the catalytic site, spanning subsites -1 to +1 of BE and DBE, shows a binding that is in agreement with the experimental data. The branching enzyme displays a slightly stronger docking energy for the maltose than for the isomaltose. Similarly, the isoamylase favors a better docking energy for isomaltose than for maltose. However in both cases, the energy is not high enough to assure the affinity required for a real substrate, showing that maltose and isomaltose are poor ligands, if not weak inhibitors. Besides, the interaction energy difference between the two enzymes and their ligands is not discriminating enough to be relevant. The catalytic residues plus the amino acids involved in the hydrogen bond and van der Waals network display similar side-chain positioning. Only two exceptions were noticed for the highly conserved histidine residues H_{BE}340, H_{BE}525 and H_{DBE}297 and H_{DBE}509 for BE and DBE enzymes respectively. At this stage, it is delicate to hypothesize on the specificity of hydrolysis vs transglycosylation. Could a longer substrate explain this specificity?

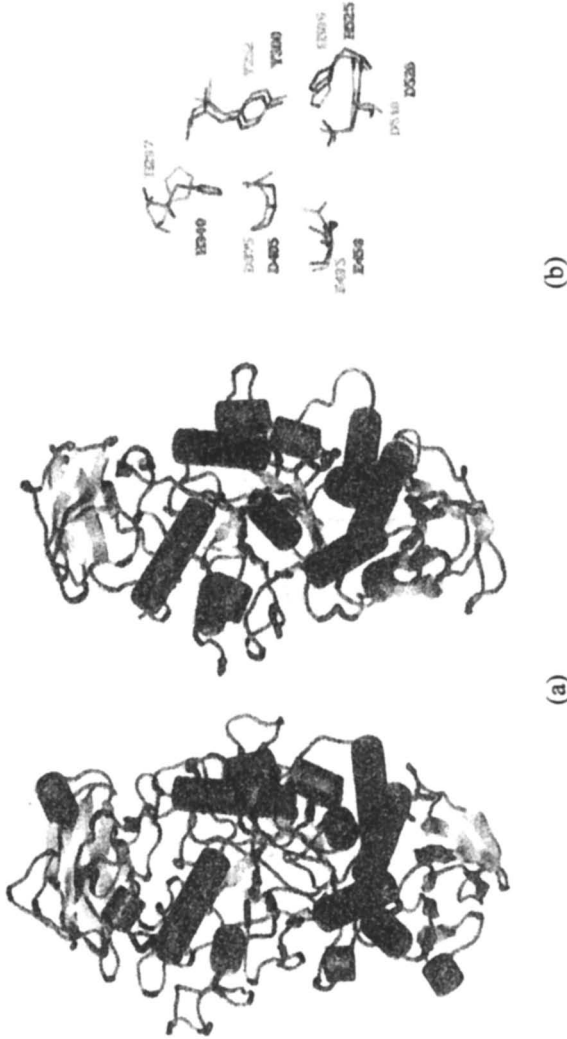


Figure 1. Topologies of *Ps. amyloclavata* debranching and *E. coli* branching enzymes (a). Details of their superimposed active sites (b). (See color page 1 in chapter.)

Docking of G10' and G12

In *Ps.amyloclavata* DBE, the substrate was stepwise α -(1,4) elongated apart from the initial isomaltose, in strict accordance with the allowed (ϕ , ψ) values. A G10' that occupies subsites -5 to +5 was bound. It showed no ring distortion and a good docking energy. Strong π hydrophobic patches buoy the binding: namely Y_{DBE}444 at subsite +3, W_{DBE}437 at subsite +2, Y_{DBE}250 which is strictly conserved at subsite -1, Y_{DBE}595 at subsite -2 and W_{DBE}304 at subsite -5, see Figures 2(a), 2(b). Tyrosine, tryptophan and phenylalanine residues are known to act as strong sugar binders (9, 16-18). Moreover, they are significantly more numerous than in other α -amylases (9, 18). Y_{DBE}250, W_{DBE}437 and Y_{DBE}444 belong to strictly conserved regions in DBE. Besides the van der Waals interactions, several hydrogen bonds can bind the substrate in the crevice. It is not the first time that such a long substrate is shown to bind at a subsite located far from the α -amylase active site (18). It has also been shown that a remote subsite like subsite -6 in CGTases can control the transferase/hydrolysis activity of the enzyme (19). Here, beyond subsite -5, the catalytic cleft could not accept a non-reducing end around DP 6-9 as it would not accommodate the crevice and would result in strong steric clashes. Particularly, W_{DBE}304 could act as a flap that could border out the donor site. For a much longer chain, the substrate could be flexible enough to accommodate, in part, in the crevice. On the other hand, the propagation at the reducing end reaches the end of the barrel with no residue collision.

Similarly in the *E.coli* BE, the substrate was stepwise α -(1,4) elongated apart from the initial maltose, in strict accordance with the allowed (ϕ , ψ) values. Our strategy revealed a possible binding of a G12 that contains no ring distortion and spans from subsites -7 to +5 (see Figure 4). Hydrophobic π patches that are located exclusively on the donor site strongly fix the maltododecaose at its non-reducing side. The donor site can bind at least 7 glucose moieties, in agreement with the experimentally evaluated chain transfer of DP 5-16 (20) or DP 8-14 (21). An even stronger experimental approval is the inhibiting effect of BAY e4609, an acarbose like inhibitor with a non-reducing end of 7 glucose residues (22). The donor site contains the strictly conserved Y_{BE}300 at subsite -1, F_{BE}294 at subsite -4 and F_{BE}346 at subsite -7. Beyond this point, the elongation could reach the N-terminal domain with no steric collision (see figure 3). This propagation direction is once more consistent with the experimental data as the N-terminal domain shows a role in the DP length transfer. Indeed, the truncation of 112 amino acids at the N-terminus changes the chain transfer pattern of this glycogen branching enzyme (7, 16).

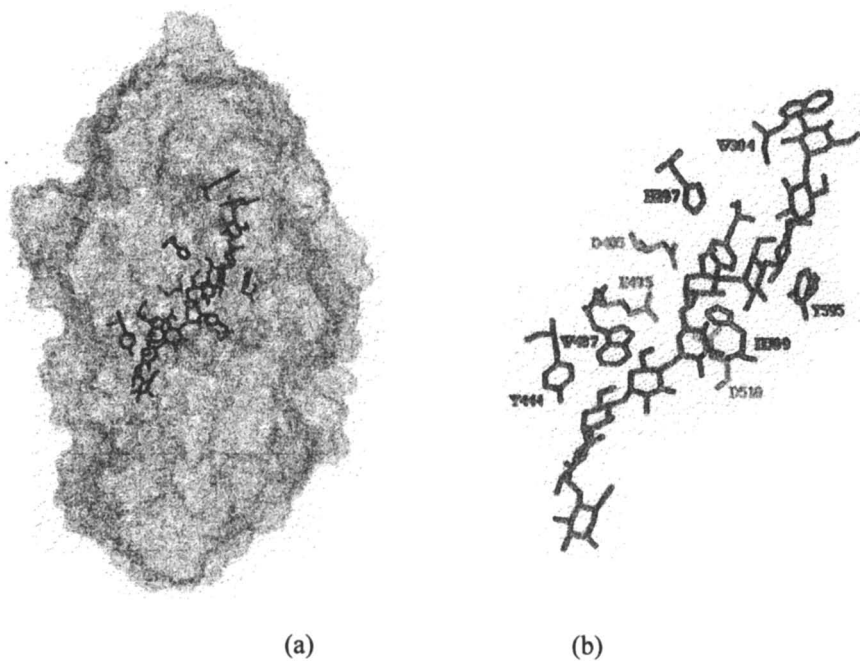
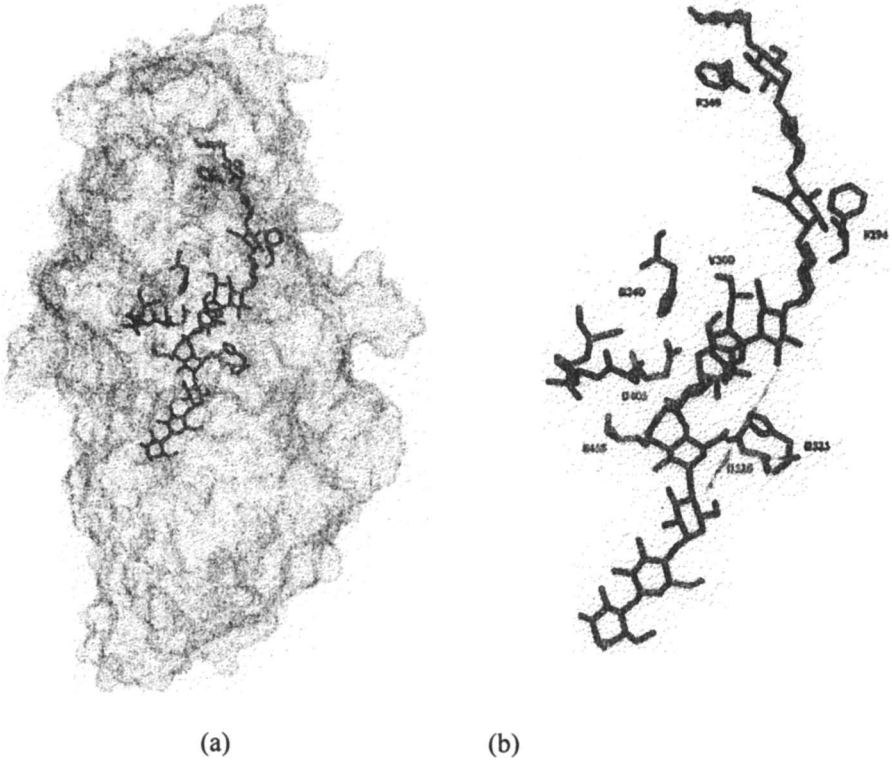


Figure 2. Docking of the G10' in DBE (a). Details of the π stackings (b). (See color page 1 in chapter.)



*Figure 3. Docking of the G12 in BE (a). Details of the π stackings (b).
(See color page 2 in chapter.)*

To summarize, the *E.coli* BE exhibits a two subsites longer donor site [-1; -7] compared to the *Ps.amylocladus* one [-1; -5]. Both enzymes display an equal length for their acceptor site. However, DBE shows several π stacking interactions at the acceptor site whereas BE lacks such interactions, actually rendering the acceptor site slightly less strong. It can be reasonably hypothesized that the π stacking patches which are numerous along the crevice of the DBE could sensitively stabilize the binding of the more flexible α -(1,6) substrate compared to α -(1,4) substrate.

At this stage, two questions remain to fully comprehend the hydrolysis vs. transglycosylation specificity:

- i In DBE, the donor site is fully occupied by a glycogen like branch (G6); however, the docking of the linear α -(1,4) main chain remains to be modeled. What could favor the entrance of a water molecule to act as the nucleophile?
- ii In BE, what would favor a sugar acceptor over a water molecule to act as a nucleophile and transfer the O₆ of its non-reducing end to the C₁ of the donor oligosaccharide so that a branch is created?

Role of subsites -1 and +1, subsite +1' ending

Starting from the isomaltodecaose G10', a α -(1,4) elongated DP6 was docked in the catalytic cleft, in tight accordance with the (ϕ , ψ) values and the branching point constraints. The two non-reducing ends of the branched substrate actually occupy two grooves, covering almost the whole (α/β)₈ barrel. The first groove is the known catalytic cleft with the branch chain in the donor site. A second groove is identified for the first time. It shows numerous hydrogen bonds but no aromatic π stacking patches. The substrate avoids local glucose ring distortion and keeps the native helical chain shape. Hence, the binding shows agreement between the conformation of the native substrate and the structural features of the twin cleft (see Figure 4 (a)).

The glycogen fragment accommodates its branch α -(1,4)/ α -(1,6) at subsites -1 and +1. No other connection seems energetically possible due to the glycogen branch constraints (11, 12). The Figure 4 (b) highlights the so-called "junction region" that describes subsites -1, +1 and +2. Interestingly, it shows this region as composed of two of the four segments highly conserved within α -amylases. The regions that are known in α -amylase as region 1 and 2 are respectively D₂₉₃VVYNH₂₉₇T and R₃₇₃FD₃₇₅LASV, in *Ps. amylocladus* numbering, where D₃₇₅ is the catalytic nucleophile. Site-directed mutations of the latter motif which is conserved as RXDAVKH in CGTases showed activity change between cyclization over hydrolysis (23). Similarly the mutation E332Q (*Thermus* maltogenic amylase numbering) of the equivalent region, RLDVANE in maltogenic amylases, neopullulanases and cyclodextrinases, results in an activity

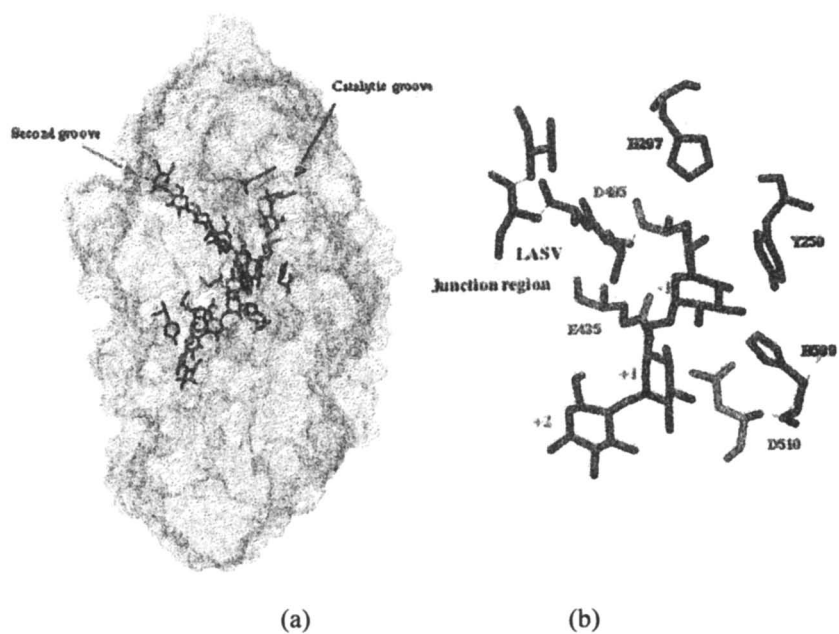


Figure 4. Docking of glycogen fragment in BE (a). The “junction region” (b). (See color page 2 in chapter.)

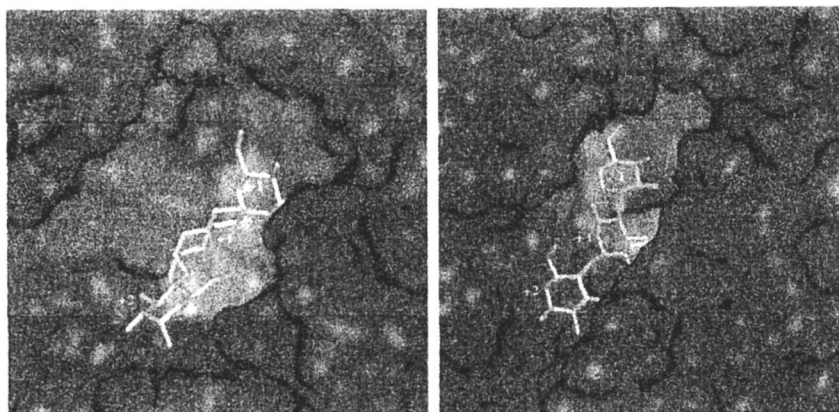
change from trans-glycosylation to hydrolysis (24). The authors suspected that an acidic side-chain located at this position could play a pivotal role in the formation and accumulation of transfer sugar and could thus modulate the ratio of hydrolysis vs. transglycosylation. Besides, they observed that the side chain of this residue is located in a pocket partly formed by region 2. The acidic side-chain could be involved in aligning the acceptor molecule to compete with a water molecule in the nucleophilic attack of the glycosyl-enzyme intermediate. Moreover, a second extra sugar-binding site has also been identified in *Bacillus stearothermophilus* neopullulanase. It is located at subsites -1 and +1 and involves the conserved region 2. In this enzyme, this residue is also highly suspected to participate to the acceptor site (25).

A homologous region is also highly conserved in bacterial branching enzymes where the motif RVD₄₀₃AVAS describes region 2. It corresponds to R₄₀₃VD₄₀₅AVAS in *E. coli* numbering. From a structural point of view, this region corresponds to a large pocket in *E. coli* BE, located within hydrogen bonding distance to subsites -1 and +1 (see Figure 5 (a)). As highlighted in figure 5 (b), the corresponding sequence in DBE is not a pocket but a wall.

The cavity exploration through our extensive docking studies showed that maltose and maltotriose moieties could accommodate this space with no ring or glycosidic linkage distortions. The docking is well maintained through hydrogen bond and electrostatic interactions. The acceptor site could extend to the interior of the pocket with extra subsites numbered +1', +2' and +3'. This could occur without weakening the binding of the substrate at the donor site. More interestingly, the lowest energy conformation evidenced the O₆ atom of the glucose at subsite +1' as pointing directly towards the C₁ of the donor sugar and located within 2 Å (see Figure 6). From the active site topology and our molecular modeling docking, we hypothesized that this extra space could favor the repositioning of the leaving group while preventing a water molecule from entering in the cleft and act as the nucleophile. The leaving group would sequentially accommodate the extra pocket and become the neo-incoming sugar acceptor. By acting as the nucleophile, it would favor the transglycosylation to obtain an intra molecular α -(1,6) bond. This pocket which is partly conserved in CGTases and neopullulanases would be a finger print of BE.

Conclusion

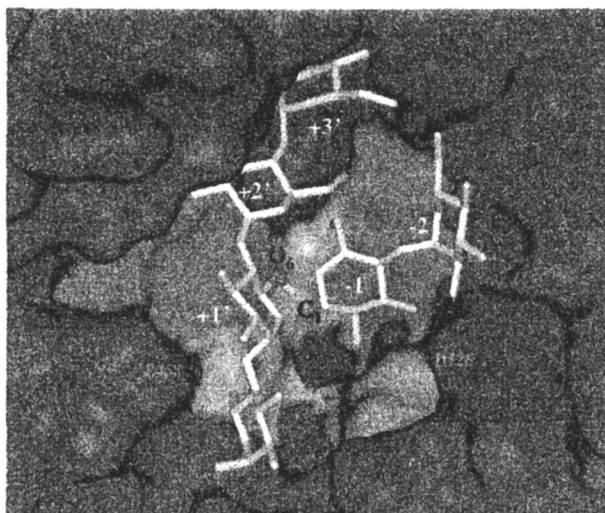
Besides the binding affinities at the donor and the acceptor sites, the α -(1,4) hydrolysis vs. α -(1,6) transglycosylation could be due to a subtle docking energy balance during the catalytic event. The energy variations between the two binding sites could drive first the docking of an α -(1,4) substrate and its



(a)

(b)

*Figure 5. DAVAS pocket in BE (a). DLASV wall in DBE (b).
(See color page 3 in chapter.)*



*Figure 6. Subsites +1' and -1 of BE designed for a α -(1,6) transglycosylation?
(See color page 3 in chapter.)*

“classical” cleavage at subsites -1 and +1. Then the leaving group could reposition in the extra pocket, particularly at subsite +1'. Finally, it could participate to the formation of the intra molecular α -(1,6) bond with the sugar donor covalently bound at the donor subsite.

Neopullulanase or maltogenic amylases are known to have abilities to hydrolyze and transglycosylate a large panel of bonds from α -(1,3) to α -(1,6). Moreover, they are very active on rather short substrates like cyclodextrins (26). They share a conserved sequence in region 2 and interestingly display a pocket that belongs to the close vicinity of the catalytic subsites -1 and +1. But their so-called pocket has a major difference compared with the one of the branching enzymes. This difference is not only due to sequence and topology discrepancies but mainly to their oligomerization state. Indeed, the extra space could be formed in part during the *in vivo* homodimerization of the enzymes (25).

The α -amylases that strictly bind and hydrolyze α -(1,4) linkage do not display such extra space. They do not show evidence of multimerization either. Their α -(1,4) disproportionation capacity would be due to a high concentration of substrate or inhibitor. In that case, such high concentration that is much higher than physiological condition would reverse the enzymatic reaction (27). Isoamylases that strictly hydrolyze the α -(1,6) linkages do not display such extra space either despite some experimental data referencing multimeric organization (28). Their ability to bind an α -(1,6) linkage could be greatly due to their Y-shaped groove that is adapted to accommodate a branched substrate. Actually, the two depressions which connect themselves at the active site plus the hydrophobic patches that are disseminated all along the enzyme surface could participate to the binding of this flexible substrate.

Among the regions that are conserved in α -amylases, region 2 is the differential factor that could be critical to modulate the hydrolysis over transglycosylation reactions. The DAVAS sequence of BE deserves rational mutation to get to DLASV, signature pattern of DBE. Ideally, one can expect the catalytic characteristics of DBE. Similarly, the rational design of DAVAS to obtain DVANE which is the signature of neopullulanases region 2 could favor α -(1,6) hydrolysis over transglycosylation. All those rational mutations are already under process in our laboratory.

Acknowledgment

I would like to deeply thank Dr. Cécile Albenne, Pr. Steven Ball and Pr. Pedro Alzari for helpful discussion.

References

1. Henrissat B.; Davies G. J. *Curr. Op. Struct. Biol.* **1997**, *7*, 637-644.
2. Holm L.; Koivula A. K.; Lehtovaara P. M.; Hemminki A.; Knowles J. K. C. *Protein Eng.* **1990**, *3*, 181-190
3. DiMauro S.; Lamperti C. *Muscle & Nerve.* **2001**, *24*, 984-999
4. Uitdehaag J. C. M.; van Alebeek G.J. W. M.; van der Veen B.; Dijkhuizen L.; Dijkstra B. W. *Biochemistry.* **2000**, *39*, 7772-7780
5. Abad C. M.; Binderup K.; Rios-Steiner J.; Arni R. K.; Preiss J.; Geiger J. H. *J. Biol. Chem.* **2002**, *277*, 42164-42170
6. Katsuya Y. ; Mezaki Y.; Kubota M.; Matsuura Y. *J. Biol. Chem.* **1998**, *281*, 885-897
7. Devillers C. H.; Piper M. E.; Ballicora M. A.; Preiss J. *ABB.* **2003**, *418*, 34-38
8. Zdobnov E. M.; Lopez R.; Apweiler R.; Etzold T. *The EBI SRS server-recent developments, Bioinformatics.* **2002**, *18*, 368-373.
9. André G.; Buléon A.; Haser R.; Tran V. *Biopolymers*, **1999**, *50*, 751-762.
10. André G.; Tran V. *Biopolymers*, **2004**, *75*, 95-108
11. Kadziola A.; Søgaard M. ; Svensson B.; Haser R. *J. Mol. Biol*, **1998**, *278*, 205-217
12. Skov L. K.; Mirza O.; Sprogøe D.; Dar I.; Remaud-Simeon M.; Albenne C.; Monsan P.; Gajhede M. *J. Biol. Chem.* **2002**, *277*, 47741-47747.
13. Dauter Z.; Dauter M.; Brzozowski A. M.; Christensen S. Borchert T. V.; Beier L.; Wilson K. S. Davies G. J. *Biochemistry*, **1999**, *38*, 8385-8392
14. André G.; Buléon A.; Vallee F.; Juy M.; Haser R.; Tran V. *Biopolymers*, **1996**, *39*, 737-751.
15. Dowd M. K.; Reilly P. J.; French A. D. *Biopolymers*, **1994**, *34*, 625-638
16. Quijcho F.A. *Pure Appl. Chem.* **1989**, *61*, 1293-1306
17. Vyas N.K. *Curr. Opin. Struct. Biol.* **1991**, *1*, 732-740
18. Bak-Jensen K. S.; André G.; Gottschalk T. E.; Paës G.; Tran V.; Svensson B. *J. Biol. Chem.* **2004**, *279*, 10093-10102
19. Leemhuis H.; Uitdehaa J. C. M.; Rozeboom H. J.; Dijkstra B. W.; Dijkhuizen L. *J. Biol. Chem.* **2002**, *277*, 1113-1119
20. Guan H.; Li P.; Impari-Radosevic J.; Preiss J.; Keeling P. *Arch. Biochem. Biophys.* **1997**, *342*, 92-98
21. Binderup K.; Mikkelsen R.; Preiss J. *Arch. Biochem. Biophys.* **2002**, *397*, 279-285
22. Binderup K.; Libessart N.; Preiss J. *Arch. Biochem. Biophys.* **2000**, *374*, 73-78

23. Van der Veen B. A.; Leemhuis H.; Kralj S.; Uitdehaag J. C. M.; Dijkstra B. W.; Dijkhuizen L. *J. Biol. Chem.* **2001**, *276*, 44557-44562
24. Kim T. J.; Park C. S.; Cho H. Y.; Cha S. S.; Kim J. S.; Lee S. B.; Moon T. W.; Kim J. W.; Oh B. H.; Park K. H. *Biochemistry*, **2000**, *39*, 6773-6780
25. Hondoh H.; Kuriki T.; Matsuura Y. *J. Mol. Biol.* **2003**, *326*, 177-188
26. Dauter Z.; Dauter M.; Brzozowski A. M.; Christensen S.; Borchert T. V.; Beier L. Wilson K. S.; Davies G. *Biochemistry*, **1999**, *38*, 8385-8392

Chapter 10

QM/MM Simulations of Carbohydrates

Abdul-Mueed Muslim¹, Jonathan P. McNamara², Hoda Abdel-Aal¹,
Ian H. Hillier², and Richard A. Bryce^{1,*}

¹School of Pharmacy and Pharmaceutical Sciences,
University of Manchester, Manchester M13 9PL, United Kingdom

²Department of Chemistry, University of Manchester, Manchester
M13 9PL, United Kingdom

Hybrid quantum mechanical (QM)/molecular mechanical (MM) molecular dynamics simulations were used to investigate disaccharide conformation in aqueous solution. *In vacuo* and aqueous solution conformational free energy surfaces were constructed from potential of mean force calculations, using weighted histogram analysis of combined QM/MM molecular dynamics simulations of 8.5 ns and 13.5 ns respectively. Calculations indicated the presence of direct and water-bridged intersaccharide hydrogen bonds, the latter consistent with a broad range of $\phi\psi$ space. To improve the accuracy of the description of carbohydrates by semi-empirical QM methods, we also detail our work on reparameterization of the PM3 Hamiltonian. This is based on fitting to 1,2-ethanediol structures and energies. Application of the resulting model, PM3CARB-1, to modeling of glucose is discussed. Improvement in energetic ranking of ⁴C₁ and ¹C₄ conformations was found. QM/MM dynamics simulations of a disaccharide using PM3CARB-1 did not exhibit transitions from ⁴C₁ to ¹C₄ structures.

With advances in glycobiology, the complex and significant role of carbohydrates is being elucidated at the physiological and molecular levels. Carbohydrates are found as oligosaccharides, polysaccharides, proteoglycans, glycoproteins and glycolipids, and are implicated in a wide range of processes, many involving cell-cell interactions. For example, regulation of intercellular signaling pathways for embryonic cell fate decisions involves fucose-specific GlcNAc-transferases, known as “Fringe” proteins. The complexity of carbohydrates is reflected by the number of constitutional isomers. A trisaccharide, for example, can form 119,736 isomers from a pool of nine possible monosaccharides; a tripeptide, with 20 possible amino acid constituents, can form only 8000 isomers.¹ In actuality, there are greater than 100 known monosaccharides.² Beyond constitutional isomerism, carbohydrates are also conformationally flexible. For a linear polysaccharide of n monomers and assuming three staggered orientations for rotatable single bonds, the number of potential rotamers increases as 3^{6n} . Thus, for maltose, there are over half a million possible conformations; for a trisaccharide, the number of conformers potentially contributing to the equilibrium population increases to greater than a third of a billion. It is perhaps unsurprising that given this constitutional and conformational flexibility, carbohydrates have been proposed as informational bridges,³ spanning at a molecular level the acknowledged gap in complexity of the genome relative to the human brain.

Carbohydrate flexibility presents a considerable challenge to experimental and theoretical approaches to structural characterization. This may in part be responsible for the arguable lag in computational modelling approaches applied to carbohydrates relative to nucleic acids and proteins.⁴ To permit evaluation of the many accessible conformers, conformational analysis demands a computationally efficient potential energy function. In this regard, classical force fields employing fixed charges have been the mainstay of biomolecular modelling. In the case of carbohydrates, however, describing the correct physical behaviour of these polar molecules, which incorporate stereoelectronic subtleties (anomeric, exo-anomeric and gauche effects), has been somewhat problematic at the molecular mechanical level; for example, this is evidenced by several reparametrizations of the widely-used force fields, AMBER⁵⁻¹⁰, CHARMM^{11,12} and OPLS.^{13,14}

Hybrid QM/MM calculations on carbohydrates

Ideally, carbohydrates should be modelled in electronic detail, incorporating both intrinsic and external influences on electronic structure, and thus, their effect on molecular conformation and condensed phase behaviour. For example, calculations have estimated that electric polarization due to aqueous solvent contributes 10-20% of the solute-solvent interaction energy.¹⁵ With many hydrogen bond donor and acceptor groups,

carbohydrates experience directional polar interactions with environments such as aqueous solvent or protein clefts. To enable appropriate coupling to this environment, it is possible to employ hybrid potential energy functions, combining an inner region described at a quantum mechanical (QM) level of theory (for example, the carbohydrate) with an outer region (for example, solvent) modelled at a molecular mechanical (MM) force field. One type of hybrid QM/MM approach is afforded by the ONIOM framework of Morokuma.¹⁶ Here, the total energy of the system, $E_{tot}^{ONIOM(QM:MM)}$, is obtained from three separate computations, two of which are performed at the less intensive MM level of theory:

$$E_{tot}^{ONIOM(QM:MM)} = E_{QM}^{inner} + (E^{all} - E^{inner})_{MM} \quad (1)$$

where E_{QM}^{inner} is the energy of the inner region, treated at a QM level of theory. The second and third terms refer to the total MM energy of the entire system and the inner region respectively. In this scheme, generally called 'mechanical embedding', there is no polarization of the inner QM region by the outer MM one, the interaction between the two regions being evaluated at the MM level.

Although the ONIOM scheme has not been directly applied to carbohydrates in a QM/MM context, it has been applied to calculation of NMR chemical shifts in β -D-glucose, using a combination of QM levels of theory.¹⁷ An interesting variant of ONIOM has been applied by French *et al.* to calculation of carbohydrate $\phi\psi$ maps in aqueous solution¹⁸ and to computation of protein-bound distortion energies of carbohydrates.¹⁹ Here, the inner region is based on the sugar backbone. For example, for sucrose, an analogue based on linked tetrahydropyran and tetrahydrofuran moieties is used. For each $\phi\psi$ calculated for the analogue at the QM level, many MM evaluations are performed to explore the hydroxyl and primary hydroxymethyl conformations of the sugar. The total energy of the sugar as a function of $\phi\psi$ is then obtained *post facto* from Eq.1. Thus, the approach allows QM treatment of the sugar backbone, accounting for the overlapping anomeric effect found in sucrose. Although the myriad permutations of the OH and CH₂OH groups are considered, the interaction of these groups with the QM ring is considered only at the MM level. Thus, explicit polarization of the ring due to its polar pendant groups is omitted.

An alternative hybrid QM/MM approach is to couple directly the QM inner region with the MM outer environment.²⁰ Here, the total energy, E_{tot} , can be written as the sum of the energy of the inner QM region, the outer MM region and the interaction between the two:

$$E_{tot} = E_{QM}^{inner} + E_{QM/MM}^{inter} + E_{MM}^{outer} \quad (2)$$

For non-covalent interactions between QM and MM regions, the $E_{QM/MM}^{inter}$ coupling energy is given by:

$$E_{QM/MM}^{inter} = - \sum_{\mu, \nu, s} P_{\mu\nu} q_s (\mu | r_s^{-1} | \nu) + \sum_m \sum_s q_s \frac{Z_m}{r_{ms}} + \sum_m \sum_s \left(\frac{A_{ms}}{r_{ms}^{12}} - \frac{B_{ms}}{r_{ms}^6} \right) \quad (3)$$

The first two terms involve the interaction of the MM point charges, q_s , with the QM charge density and QM nuclei, Z_m , respectively, the final term being the van der Waals interaction between the QM and MM atoms. If we write $E_{QM/MM}^{inter} = (E^{all} - E^{inner} - E^{outer})_{MM}$, with the interaction of inner and outer regions at a purely MM level, then the ONIOM scheme (Eq.1) is recovered.

A number of groups have applied the direct coupling approach to characterization of enzyme reaction mechanisms involving carbohydrates: examples include xylose isomerase,²¹⁻²³ neuraminidase,^{24,25} human aldose reductase,²⁶ uracil-DNA glycosylase,²⁷ triose phosphate isomerase,²⁸ and glucose oxidase.²⁹ Here, we describe aspects of our current work on the conformational behaviour of carbohydrates employing a fully electronic description of the sugar *via* a directly coupled QM/MM approach.³⁰

QM/MM free energy surface for a disaccharide in aqueous solution

To investigate the conformation of a carbohydrate solute using quantum mechanics in an aqueous environment which is described *via* an appropriate molecular mechanical force field, we employed a disaccharide model, 4-O- α -D-xylopyranosyl- α -D-xylopyranose (Figure 1). This α -(1 \rightarrow 4)-linked analogue of maltose, lacking primary hydroxymethyl groups, is subsequently denoted "dixylose" after Naidoo and Brady.³¹

A semi-empirical PM3 Hamiltonian was selected to describe the disaccharide. PM3 exhibits an improved ability to model hydrogen bonding in organic systems relative to AM1,³² and has had some success in predicting energy differences of hydroxyl rotamers of glucose³³ and methanediol.³⁴ However, the PM3 potential underestimates stability of the glucose ring 4C_1 conformation, relative to 1C_4 .³⁵ Consequently, the dixylose rings were constrained to the 4C_1 conformation during dynamics calculations. Using this computationally tractable QM/MM potential, the conformational free energy surface of the disaccharide as a function of the glycosidic angles φ and ψ

was determined. Here, we define φ and ψ as the dihedrals H1-C1-O1-C4' and C1-O1-C4'-H4' respectively (Figure 1). A free energy surface for dixylose *in vacuo* and in aqueous solution was obtained using the potential of mean force approach.³⁶ Here, the Helmholtz free energy difference of the system along the coordinates φ and ψ , $\Delta A(\varphi, \psi)$, is given by, $\Delta A(\varphi, \psi) = -RT \ln \rho(\varphi, \psi)$, where $\rho(\varphi, \psi)$ is the distribution of states. Sampling was enhanced by use of a suitable umbrella potential. The 2D potential of mean force was subsequently constructed using the weighted histogram analysis method.³⁷ Condensed phase simulations considered the QM dixylose in the presence of a cubic 25.1^3 \AA^3 box of 492 TIP3P water molecules. *In vacuo*, each $\varphi\psi$ simulation window comprised a 5 ps equilibration preceding 50 ps production dynamics; for solution, this was 25 ps and 80 ps respectively. This led to a total of 8.5 ns and 13.5 ns combined QM/MM molecular dynamics simulations from which respective *in vacuo* and aqueous solution conformational free energy surfaces were calculated. Full computational details are given elsewhere.³⁰

The resulting free energy surface in aqueous solution (Figure 2) exhibited a central, low energy region, qualitatively similar both to the adiabatic and *in vacuo* free energy surfaces,³⁰ and to the previous MM studies.³¹ Two energy wells are formed, X₁ and X₂, centred around $(-20^\circ, 16^\circ)$ and $(38^\circ, 48^\circ)$ respectively. These minima lie in proximity to the crystal structures of α -maltose³⁸ at (φ, ψ) of $(-2^\circ, 2^\circ)$ and the neutron diffraction structure of the β -maltose monohydrate³⁹ at $(4^\circ, 12^\circ)$. No such structure has yet been determined for dixylose.

MM studies³¹ identified the global minimum at $(-37^\circ, 52^\circ)$, where the conformer appears to be stabilized by a solvent-mediated interaction between O2 and O3'. Analysis of 2 ns of extended QM/MM molecular dynamics simulations indicated that 34.4% of the dixylose conformations possessed solvent-mediated interactions at the PM3/TIP3P level of theory.³⁰ Representative snapshots taken from the extended simulations (Figure 3) indicate that all four possible hydrogen bonding arrangements are adopted.

Most of the water-bridged conformers sampled during simulations involved either (i) the water oxygen atom as an acceptor of hydrogen bonds from O2-H and O3'-H (Figure 3c), or (ii) a O2-H \cdots O_w-H_w \cdots O3' interaction (Figure 3a), analogous to the interaction observed from neutron scattering (Table I). A second observation was that water-bridged conformers appear to be consistent with a wide range of $\varphi\psi$ values, evident from the snapshot conformations (Figure 3). Different bridging interactions correspond to different regions of the map, as indicated by quadrant analysis of the bridging interactions (Table I). In particular, water as the double acceptor is found mainly for $(-\varphi, +\psi)$ values, whereas the O2-H \cdots O_w-H_w \cdots O3' interaction generally corresponds to the $(+\varphi, +\psi)$ region. However, we note that 60.4% of the calculated disaccharide ensemble lacked bridged or direct intersaccharide interactions.³⁰

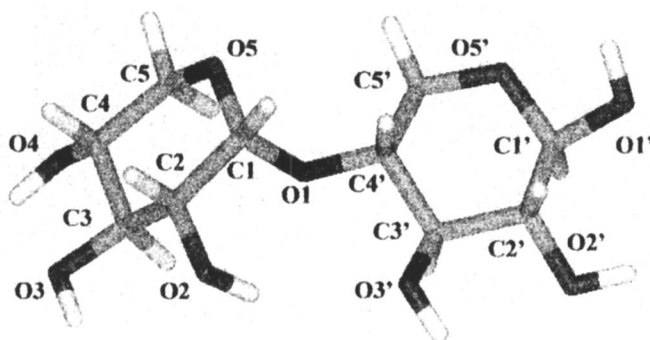


Figure 1. 4-O- α -D-xylopyranosyl- α -D-xylopyranose

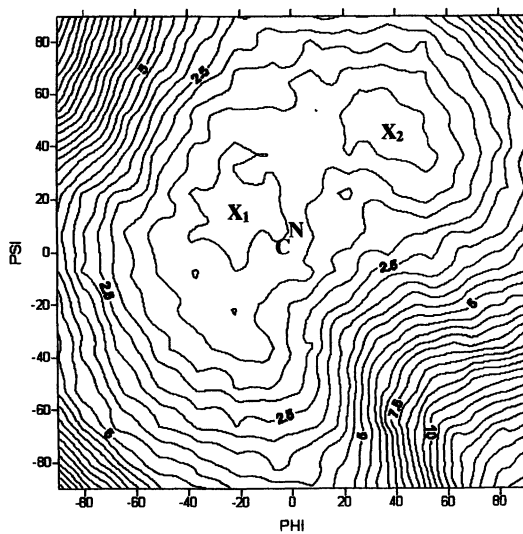


Figure 2. Aqueous solution PMF for dicylose at QM/MM level, with minimum energy (X_1, X_2), maltose crystal (C) and neutron scattering (N) conformations (0.5 kcal/mol energy contours.)

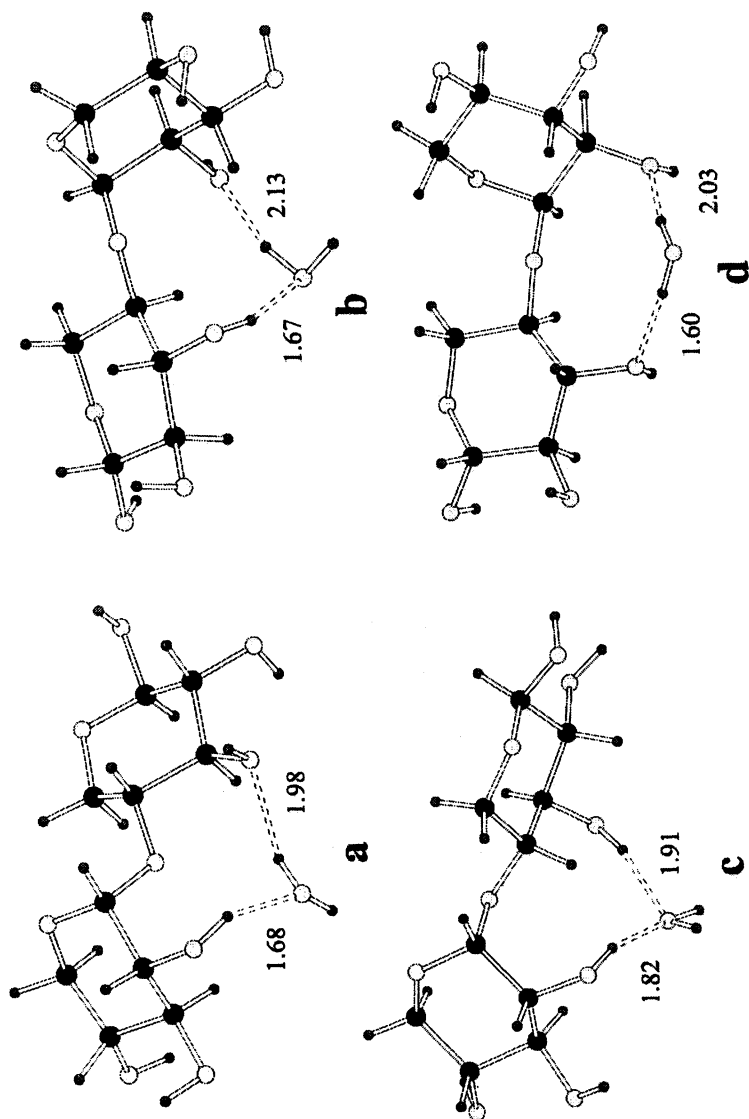


Figure 3. Snapshot of dioxyllose in solution at (ϕ, ψ) values of (a) $(-49^\circ, 52^\circ)$ (b) $(42^\circ, 50^\circ)$ (c) $(42^\circ, 55^\circ)$ (d) $(-14^\circ, 24^\circ)$. Distances in Å

Table I. Analysis of water-bridging populations (%) from extended aqueous MD simulations, including division into $\phi\psi$ quadrants

| <i>bridging interaction</i> | <i>total</i> | (ϕ, ψ) | $(-\phi, \psi)$ | $(-\phi, -\psi)$ | $(\phi, -\psi)$ |
|--|--------------|----------------|-----------------|------------------|-----------------|
| H ₂ O double acceptor | 38.5 | 3.1 | 32.3 | 3.0 | 0.1 |
| H ₂ O double donor | 5.7 | 0.9 | 3.1 | 1.6 | 0.0 |
| O2-H \cdots O _w -H _{w\cdotsO3'} | 39.5 | 26.1 | 11.2 | 1.8 | 0.3 |
| O2 \cdots H _w -O _w \cdots H-O3' | 16.4 | 0.3 | 8.3 | 7.7 | 0.1 |

As discussed earlier, employing a QM/MM potential permits polarisation of the carbohydrate solute wavefunction due to the solvent partial charges. The extent of carbohydrate polarisation was estimated *via* calculation of atom-centred Mulliken charges for dixylose over the extended solution and *in vacuo* trajectories. Population analysis of structures from the PM3/TIP3P trajectory indicated increased polarity (17%) of the pendant hydroxyl groups on solvation.³⁰ This is commensurate with the 10-20% overestimation of molecular polarity at the HF/6-31G* level of theory, commonly used to implicitly account for the effect of condensed phase electronic polarization when deriving atomic partial charges *in vacuo*.¹⁴ During QM/MM molecular dynamics, the standard deviations about mean partial charge for ring, glycosidic and hydroxyl oxygens in dixylose, were non-negligible, increasing by approximately a factor of two in aqueous solution compared to the gas phase.³⁰ The largest fluctuation in charge was exhibited by atoms O2' and O3' on the reducing sugar; the O3' Mulliken charge ranges from -0.27 e to -0.50 e (Figure 4). However, even the less polarisable ring O5' oxygen charge fluctuated over a range of 0.22 e (Figure 4).

A quantum mechanical force field approach for carbohydrates

Although useful insights were obtained from these QM/MM free energy calculations, for more extensive exploration of the conformational hypersurface of carbohydrates, the limitations of the PM3 method, particularly the underestimation of the stability of ⁴C₁ conformation requires redress. The approach we adopted was to reparametrize the PM3 Hamiltonian, in a fashion analogous to fitting of an empirical force field, basing our strategy on small molecule carbohydrate analogues.⁴⁰ We then explored the transferability of the model to the monosaccharide, glucose, considering in particular ¹C₄ and ⁴C₁ conformations. Our parametrization strategy follows the specific reaction parameter (SRP) approach of Rossi and Truhlar,⁴¹ whereby selected parameters of a semi-empirical molecular orbital

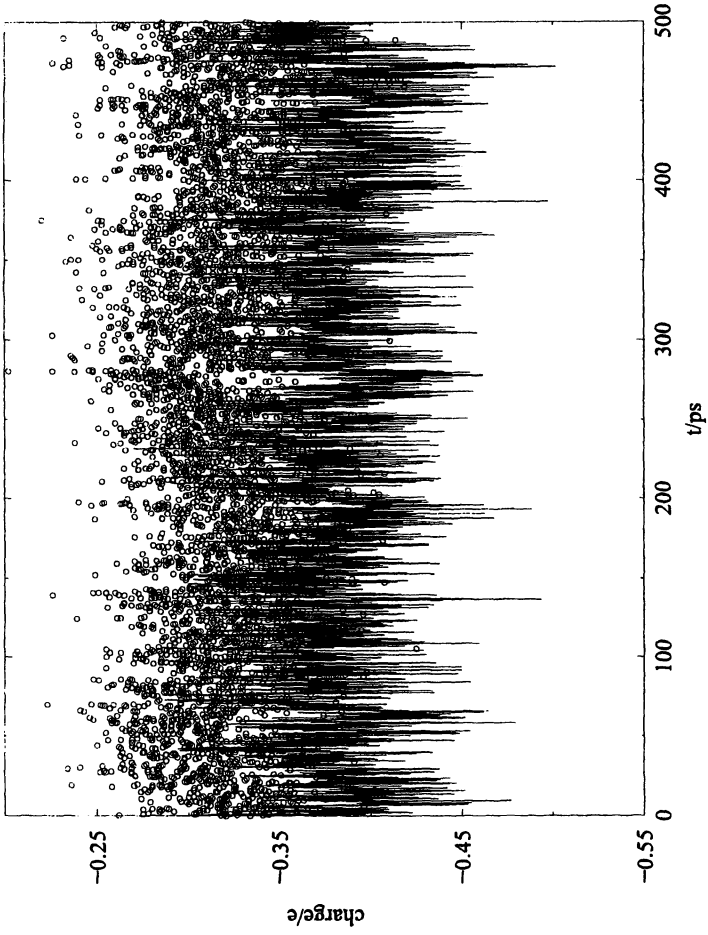


Figure 4. Charge dynamics at 03' (circles) and 05' (line)

method are adjusted to fit *ab initio* data for a specific reaction (or a small range of reactions). In our work, we use a genetic algorithm to optimise both PM3 hydrogen (U_{ss} , β_s , α) and oxygen parameters (U_{ss} , U_{pp} , β_s , β_p , α), with the remaining PM3 parameters unchanged. In the language of atomistic modeling, we term the resulting model a “quantum mechanical force field”.

A parametrization of the PM3 Hamiltonian was obtained,⁴⁰ based on ten 1,2-ethanediol conformers evaluated at the MP2/cc-pVTZ level of theory, and is denoted PM3CARB-1 (Table II). The ten conformers are labelled according to three torsion angles, $H_1O_1C_1C_2$, $O_1C_1C_2O_2$ and $C_1C_2O_2H_2$.

We first compare the ability of the PM3 and PM3CARB-1 models to reproduce the *ab initio* QM energetics and structures of the 1,2-ethanediol training set. The general PM3 potential fails to predict tGg⁻ as the lowest energy 1,2-ethanediol structure, identifying instead gGg⁻, which is another intramolecularly hydrogen-bonded structure (Table II). The AM1 model incorrectly predicts the gGg conformer to be the global minimum (Table II)⁴³. A measure of intramolecular hydrogen bond strength is afforded by the difference in stability of tGg and tGg⁻. Interestingly, PM3 predicts an energy difference of 1.8 kcal/mol (Table II), in closer agreement with MP2 (3.8 kcal/mol) than obtained *via* the OPLS-AA force field¹³ (8.2 kcal/mol). This overestimation of intramolecular hydrogen bond strength by the force field is commensurate with the use of fixed partial charges designed to represent the average effect of solute polarization in the condensed-phase. Use of a 1-5 scaling factor to yield the OPLS-AA-SEI potential led to improved agreement (energy difference of 4.8 kcal/mol).¹⁴ The fitted PM3CARB-1 potential predicts tGg⁻ as the correct lowest energy conformer, and finds an energy difference compared to tGg of 3.6 kcal/mol, in good agreement with the *ab initio* value of 3.8 kcal/mol (Table II). From an overall rms error in energies of 2.6 kcal/mol at PM3 for the ten diol conformations, PM3CARB-1 reduces this error to 0.6 kcal/mol. With regard to diol geometry, the PM3CARB-1 model reduces the rms error for the two CCOH torsions from values of 13.7° and 15.2° at the PM3 level, to 4.7° and 4.9°, respectively.⁴⁰ However, the rms error in the OCCO angle increases at the PM3CARB-1 level from 5.9° to 9.8°.⁴⁰ The tGt conformer, not identified on the PM3 surface, is now found to be stable *via* the PM3CARB-1 model. The PM3CARB-1 Hamiltonian thus appears to exhibit overall improvement in prediction of diol energetics and structure.

The transferability of PM3CARB-1 to molecules which include the characteristic OCO motif was considered, including glucose.⁴⁰ We consider here the prediction of the relative energetics and structure of a set of fourteen glucose conformations (Table III). Using the convention of Kony *et al.*,¹⁴ glucose structures are labelled according to conformation of ring, anomer type, hydrogen bond network direction and primary hydroxymethyl group torsion angles.

Table II. Relative energies (kcal/mol) of 1,2-ethanediol conformations, and rms error with respect to MP2 level of theory. Conformations are labeled according to torsion angles, $H_1O_1C_1C_2$, $O_1C_1C_2O_2$ and $C_1C_2O_2H_2$.

| <i>Conformer</i> | <i>MP2^a</i> | <i>HF</i> | <i>AM1</i> | <i>PM3</i> | <i>PM3CARB-1</i> |
|------------------|------------------------|-----------|------------|--------------------|------------------|
| g^-Gg^- | 1.19 | 1.03 | -0.98 | -0.78 | 1.45 |
| gGg | 3.15 | 3.04 | -0.98 | -0.82 | 1.75 |
| gGg^- | 0.30 | 0.69 | -0.88 | -1.39 | 1.10 |
| gTg | 3.00 | 2.59 | 0.39 | -0.43 | 2.87 |
| gTg^- | 2.81 | 2.28 | 0.20 | -0.67 | 2.50 |
| tGg | 3.81 | 3.46 | 2.56 | 1.76 | 3.63 |
| tGg^- | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| tGt | 3.48 | 2.90 | 3.19 | -0.78 ^b | 2.56 |
| tTg | 2.92 | 2.08 | 1.55 | 1.07 | 2.40 |
| tTt | 2.87 | 1.69 | 2.98 | 2.56 | 2.23 |
| rms | | 0.51 | 1.75 | 2.56 | 0.57 |

^a *Ab initio* QM levels of theory for 1,2-ethanediol are MP2/cc-pVTZ¹⁴ and HF/6-311+G(2d,2p)⁴²; ^b unstable – converts to g^-Gg^-

Table III. Relative energies (kcal/mol) of glucose conformers, and rms error with respect to *ab initio* QM level of theory.

| <i>Conformer</i> | <i>Ab initio^a</i> | <i>PM3</i> | <i>PM3CARB-1</i> |
|-------------------------------------|------------------------------|------------|------------------|
| ⁴ $C_1/\alpha/cc/g^-g^b$ | 0.00 | 0.34 | 0.43 |
| ⁴ $C_1/\alpha/cc/gg^-$ | 0.01 | -0.11 | 0.35 |
| ⁴ $C_1/\alpha/cc/tg$ | 0.00 | 0.00 | 0.00 |
| ⁴ $C_1/\alpha/cl/g^-g$ | 0.91 | 0.40 | 2.07 |
| ⁴ $C_1/\alpha/cl/tt$ | 1.14 | 2.08 | 3.31 |
| ⁴ $C_1/\alpha/cl/tt$ | 1.61 | 2.08 | 3.31 |
| ⁴ $C_1/\alpha/cl/tt$ | 1.24 | 2.08 | 3.31 |
| ⁴ $C_1/\beta/cc/g^-g$ | 0.74 | 1.22 | 1.88 |
| ⁴ $C_1/\beta/cc/gg^-$ | 0.63 | 0.78 | 2.10 |
| ⁴ $C_1/\beta/cc/tg$ | 0.85 | -0.67 | 1.34 |
| ⁴ $C_1/\beta/cc/tg$ | 1.77 | -0.67 | 1.34 |
| ¹ $C_4/\beta/cc/gg$ | 9.48 | 1.44 | 6.52 |
| ¹ $C_4/\beta/cc/g^-g^-$ | 8.77 | 0.82 | 6.31 |
| ¹ $C_4/\beta/cl/g^-g$ | 8.69 | -2.43 | 4.73 |
| rms | | 2.69 | 1.60 |

^a B3LYP/6-311+G**//HF/6-31G*¹⁴

^b Adapted from Kony et al.¹⁴, notation for glucose conformers: orientation of ring/anomer/hydrogen bond network/CH₂OH group (O₅C₅C₆O₆ and C₅C₆O₆H)

The first eleven conformations in Table III have been studied previously in the fitting of a molecular mechanical force field to *ab initio* QM energies and structures.^{13,14} The final three glucose structures in Table III have a 1C_4 ring conformation. PM3 previously was shown to predict a 1C_4 structure as more stable than the lowest energy 4C_1 conformation, contrasting with crystallographic evidence and high level *ab initio* calculations, which predict 1C_4 geometries to be of the order of at least 6 kcal/mol less stable than the 4C_1 geometries.³⁵

For the fourteen glucose conformers, the PM3CARB-1 model has a rms error in relative energies of 1.6 kcal/mol compared to those at the B3LYP/6-311+G**//HF/6-31G* level (Table III). The PM3 model has a corresponding error of 2.7 kcal/mol. For the eleven 4C_1 conformations, the error is 1.1 kcal/mol *via* the PM3CARB-1 model, which although an improvement over PM3, is not yet as accurate as the errors of 0.6 kcal/mol for the OPLS-AA, and 0.5 kcal/mol for the OPLS-AA-SEI force fields.¹⁴ Significantly, 4C_1 conformations are predicted to be more stable than 1C_4 conformations *via* the PM3CARB-1 model. For example, the ${}^1C_4/\beta/cl/g^-g$ conformation is predicted by PM3 to be the global minimum for the fourteen conformers. At the PM3CARB-1 level, this conformer is 4.7 kcal/mol higher in energy than the correct lowest energy 4C_1 conformer, in significantly improved agreement with the predictions of density functional theory (Table III). This improvement in performance over PM3 is related to the increased ability of the reparametrized potential to model the interactions between neighbouring OH groups; more specifically, underestimation of the H-H repulsion, a known limitation in the gaussian core functions of PM3,⁴⁴ appears to be corrected by the PM3CARB-1 model. Thus, from a O2-H \cdots H-O4 distance of 1.81 Å for the ${}^1C_4/\beta/cl/g^-g$ conformation at the PM3 level, a H \cdots H distance of 3.44 Å is obtained at PM3CARB-1, in improved agreement with a value of 2.77 Å at the HF/6-31G* level of theory (Figure 5). The correct O2-H \cdots O4 hydrogen-bonding interaction is now formed (Figure 5).

Inspection of the fourteen geometries indicates that clockwise and counter-clockwise hydrogen bond networks agree well with *ab initio* geometries.⁴⁰ Unfortunately, however, PM3CARB-1 is unable to model some of the subtle differences in conformation of O₅C₁O₁H exhibited at the density functional level of theory. For example, the three ${}^4C_1/\alpha/cl/tt$ glucose conformers converge to the same O₅C₁O₁H angle, both at the PM3CARB-1 (185.6°) and PM3 (188.7°) levels.⁴⁰ Thus, the semi-empirical description of the anomeric OCO motif may need further development. Nevertheless, ring

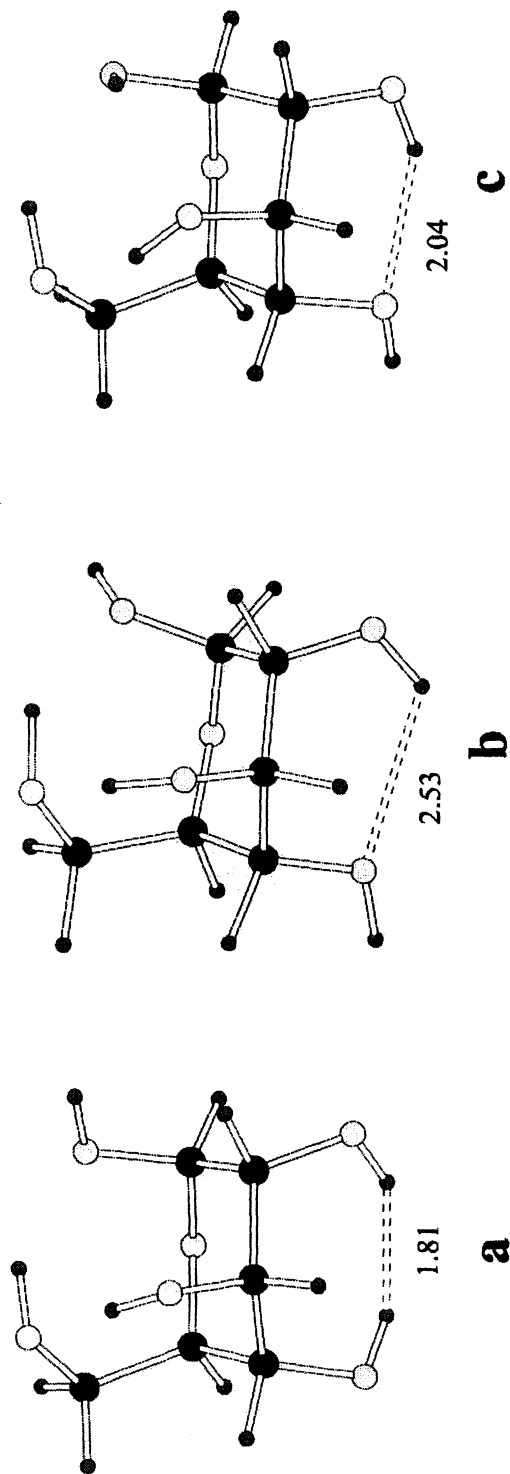


Figure 5. Minimum energy structure of glucose conformer ${}^1C_4/\beta/cl/g$ at (a) PM3 (b) PM3CARB-1 and (c) HF/6-31G* level of theory. Distances in Å.

constraints employed previously in PM3/TIP3P molecular dynamics simulations to exclude unphysical transitions to 1C_4 conformations are now obviated using the new PM3CARB-1/TIP3P potential (Figure 6).

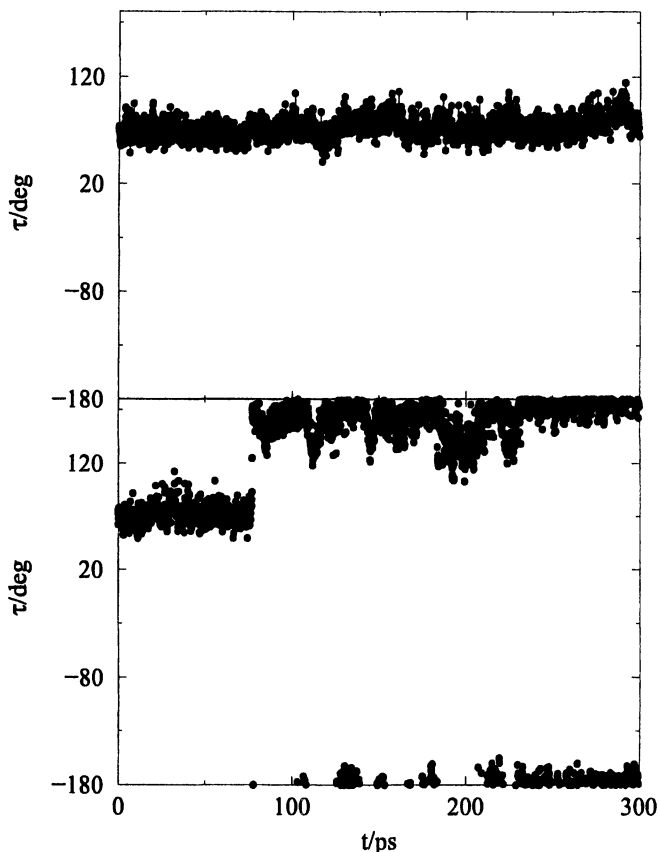


Figure 6. Time series of $C_5O_5C_1O_1$ angle (τ) of non-reducing saccharide residue of 4-O- α -D-xylopyranosyl- α -D-xylopyranose from molecular dynamics using (bottom) a PM3/TIP3P potential, and (top) a PM3CARB-1/TIP3P potential. Angles in degrees.

Conclusions

In this Chapter, we have described the ability of QM/MM methods to incorporate electronic detail into condensed phase carbohydrate dynamics, including the effects of electric polarization. Calculation of QM/MM free energies remains computationally intensive, requiring use of a semi-empirical QM Hamiltonian. Initial results of fitting a semi-empirical QM Hamiltonian to specifically improve modelling of carbohydrates shows

promise. We describe the resulting model as a quantum mechanical force field. This is distinct from classical force fields fitted to results of QM calculations,⁴⁵ or to functional fitting to obtain accurate harmonic force constants from QM calculations.⁴⁶ Although the SRP approach has been increasingly explored in the context of mechanistic studies, there are relatively few examples of its application to non-covalent interactions, examples being a study of glycine betaine in water⁴⁷ and nucleic acid base pairing.⁴⁸ We note related approaches, where replacement or supplement of the gaussian correction functions employed in semi-empirical Hamiltonians has been pursued, as in the PIF⁴⁹ and PDDG⁵⁰ models respectively. Clearly, there is scope for the application of quantum mechanical force fields to model a wide range of condensed phase processes, at the level of both conformation and chemistry. In the 1990 ACS Symposium Series on "Computer Modelling of Carbohydrates Molecules", Madsen *et al.*⁵¹ speculated that "presumably, molecular dynamics simulations of carbohydrates will soon become as commonplace as are conformational energy studies of these molecules". As these prophetic words have come true for classical MD approaches, they may soon be realised for quantum and hybrid QM/MM molecular simulations as routine tools for characterization of covalent and non-covalent interactions of carbohydrates in the condensed phase.

Acknowledgements

Figures 1 and 3 are reprinted from *Chemical Physics Letters*, 388, A.-M. Muslim and R. A. Bryce "Carbohydrate conformation in aqueous solution: calculation of a QM/MM potential of mean force" p473-478, Copyright (2004) with permission from Elsevier. Figure 5 is reprinted from *Chemical Physics Letters*, 394, J. P. McNamara, A.-M. Muslim, H. Abdel-Aal, H. Wang, M. Mohr, I. H. Hillier and R. A. Bryce "Towards a quantum mechanical force field for carbohydrates: a reparametrized semi-empirical MO approach" p429-436, Copyright (2004) with permission from Elsevier.

References

1. Nagai, Y. *Glycoconjugate J.* **2003**, *19*, 161-163.
2. Imberty, A.; Perez, S. *Chem. Rev.* **2000**, *100*, 4567-4588.
3. Roseman, S. *J. Biol. Chem.* **2001**, *276*, 41527-41542.
4. Wilson, E. K. *Chem. Eng. News* **2004**, *82*, 36-39.
5. Homans, S. W. *Biochemistry* **1990**, *29*, 9110-9118.
6. Glennon, T. M.; Zheng, Y.-J.; LeGrand, S. M.; Shutzberg, B. A.; Merz, K. M. *J. Comput. Chem.* **1994**, *15*, 1019-1040.
7. Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832-3846.

8. Pathiaseril, A.; Woods, R. J. *J. Am. Chem. Soc.* **2000**, *117*, 5179-5197.
9. Momany, F. A.; Willett, J. L. *Carbohydr. Res.* **2000**, *326*, 194.
10. Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10541-10545.
11. Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohydr. Res.* **1988**, *180*, 207-221.
12. Reiling, S.; Schlenkrich, M.; Brickmann, J. *J. Comput. Chem.* **1996**, *10*, 75.
13. Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1997**, *18*, 1955-1970.
14. Kony, D.; Damm, W.; Stoll, S.; van Gunsteren, W. F. *J. Comput. Chem.* **2002**, *23*, 1416-1429.
15. Gao, J. L. *J. Comput. Chem.* **1997**, *18*, 1061-1071.
16. Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
17. Rickard, G. A.; Karadakov, P. B.; Webb, G. A.; Morokuma, K. *J. Phys. Chem. A* **2003**, *107*, 292-300.
18. French, A. D.; Kelterer, A. M.; Cramer, C. J.; Johnson, G. P.; Dowd, M. K. *Carbohydr. Res.* **2000**, *326*, 305-322.
19. French, A. D.; Johnson, G. P.; Kelterer, A. M.; Dowd, M. K.; Cramer, C. J. *J. Comput. Chem.* **2001**, *84*, 416-425.
20. Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
21. Hu, H.; Liu, H. Y.; Shi, Y. Y. *Proteins: Struct. Funct. Genet.* **1997**, *27*, 545-555.
22. Nicoll, R. M.; Hindle, S. A.; MacKenzie, G.; Hillier, I. H.; Burton, N. A. *Theor. Chem. Acc.* **2001**, *106*, 105-112.
23. Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. L. *J. Comput. Chem.* **2003**, *24*, 177-190.
24. Tomas, A.; Jourand, D.; Bret, C.; Amara, P.; Field, M. J. *J. Am. Chem. Soc.* **1999**, *121*, 9693-9702.
25. Barnes, J. A.; Williams, I. H. *Biochem. Soc. Trans.* **1996**, *24*, 263-268.
26. Varnai, P.; Richards, W. G.; Lyne, P. D. *Proteins: Struct. Funct. Genet.* **1999**, *37*, 218-227.
27. Dinner, A. R.; Blackburn, G. M.; Karplus, M. *Nature* **2001**, *413*, 752-755.
28. Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 3093-3124.
29. Meyer, M.; Wohlfahrt, G.; Knablein, J.; Schomburg, D. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 425-440.
30. Muslim, A. M.; Bryce, R. A. *Chem. Phys. Lett.* **2004**, *388*, 473-478.
31. Naidoo, K. J.; Brady, J. W. *J. Am. Chem. Soc.* **1999**, *121*, 2244-2252.
32. Zheng, Y.-J.; Merz Jr., K. M. *J. Comput. Chem.* **1992**, *13*, 1151.

33. Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1992**, *18*, 1088-1097.
34. Woods, R. J.; Szarek, W. A.; Smith, V. H. *Chem. Commun.* **1991**, 334-337.
35. Barrows, S. E.; Dulles, F. J.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *Carbohydr. Res.* **1995**, *276*, 219-251.
36. Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187-199.
37. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011-1021.
38. Takusagawa, F.; Jacobson, R. A. *Acta Cryst.* **1978**, *B34*, 213-218.
39. Gress, M. E.; Jeffrey, G. A. *Acta Cryst.* **1977**, *B33*, 2490-2495.
40. McNamara, J. P.; Muslim, A. M.; Abdel-Aal, H.; Wang, H.; Mohr, M.; Hillier, I. H.; Bryce, R. A. *Chem. Phys. Lett.* **2004**, *394*, 429-436.
41. Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231-236.
42. Reiling, S.; Schlenkrich, M.; Brickmann, J. *J. Comput. Chem.* **1996**, *17*, 450-468.
43. Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1994**, *116*, 3892-3900.
44. Csonka, G. I.; Angyan, J. *J. Mol. Struct. (THEOCHEM)* **1997**, *393*, 31.
45. Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J.-R.; Hwang, M.-J.; Li, H. L. C.; Maple, J.; Peng, Z.; Stockfish, T. P.; Thatcher, T. S.; Yan, L.; Ni, X.; Hagler, A. T. *J. Comput. Chem.* **2001**, *22*, 1782-1800.
46. Pulay, P.; Fogarasi, G.; Pongor, G.; Boggs, J. E.; Vargha, A. *J. Am. Chem. Soc.* **1983**, *105*, 7037-7047.
47. Sironi, M.; Fornili, A.; Fornili, S. L. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1081-1085.
48. Sherer, E. C.; York, D. M.; Cramer, C. J. *J. Comput. Chem.* **2003**, *24*, 57-67.
49. Bernal-Uruchurtu, M. I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-Lopez, M. F. *J. Comput. Chem.* **2000**, *21*, 572-581.
50. Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601-1622.
51. Madsen, L. J.; Ha, S. N.; Tran, V. H.; Brady, J. W. In *Computer Modelling of Carbohydrate Molecules*; French, A. D., Brady, J. W., Eds.; ACS: 1990; pp. 69-90.

Chapter 11

Predicting the Three-Dimensional Structures of Anti-Carbohydrate Antibodies: Combining Comparative Modeling and MD Simulations

Jane Dannow Dyekjaer and Robert J. Woods*

**Complex Carbohydrate Research Center, University of Georgia,
315 Riverbend Road, Athens, GA 30602**

Knowledge of the 3D structures of antibody carbohydrate complexes would be invaluable in characterizing the structural aspects of the immune response to many bacterial and fungal pathogens. Experimental x-ray crystallographic or NMR data are rarely available for these large and frequently heterogeneous systems. If, however, the primary sequences have been determined for the antibody variable domains, an accurate 3D structure of the antigen binding Fv fragment can be derived by combining comparative modeling and molecular dynamics simulations. It is then possible in principle to model the immune complexes using molecular docking techniques.

Introduction

Understanding the interactions between complexes formed by proteins and carbohydrates is of great importance, due to the many potential applications in the design of therapeutics and in the furthering of biological insight. The role of carbohydrates in cell recognition processes is well established (1-8). In this function, the oligosaccharides are generally presented as glycoconjugates, i.e. as glycoproteins and glycolipids (9, 10). Oligo- and polysaccharides from bacterial and fungal pathogens may be employed in vaccine development and consequently their interactions with antibodies are of significant interest (11-14). Characterization of the structures of oligosaccharide antigen-antibody complexes using traditional experimental techniques, such as NMR spectroscopy and x-ray diffraction can provide detailed atomic level insight, however, these studies are typically limited to systems involving antibody fragments, such as the antigen binding fragment (Fab) or variable fragment (Fv), and to small oligosaccharides. These experimental techniques have yet to be able to characterize the structure of complexes involving large oligo- or polysaccharides, such as the capsular polysaccharides from bacterial surfaces. Frequently a large body of data pertinent to polysaccharide antigenicity is available, for example, as related to the apparent size of the epitope or to the variation of antigenicity with oligosaccharide sequence, but in the absence of a structural model little may be deduced as to the origin of the observed behavior. We report here a synopsis of computational techniques that may be combined to provide a medium resolution structural model, which nevertheless provides a basis from which to interpret the immunological data.

Antibody Structure

Antibodies are characterized by having highly conserved and well-defined structures. An antibody of the IgG isotype consists of three domains arranged schematically in a Y-shape. The "stem" is called the Fc fragment, where the letter "c" refers to the ease with which this fragment may be crystallized, while each of the identical antigen binding "arms" are referred to as Fab fragments. A single polypeptide chain forms one half of the Fc and Fab domains. To form a complete Fc thus requires two of these "heavy" chains, conjoined by a disulfide bond at the vertex between the Fc and Fab regions. Each Fab structure is augmented by a shorter polypeptide, known as the "light" chain, which is linked through a disulfide bond to the heavy chain (see Figure 1). The heavy and light chains in each Fab are arranged into four distinct domains, two constant regions, C_{H/L}, and two hypervariable regions, V_{H/L}. While the overall structure of the Fab is highly conserved, the antigen binding region, located at the interface of the

V_{HL} domains, consists of sequentially and structurally flexible amino acid loops. Within the light chain these hypervariable loops are called L1, L2, and L3; similarly, the three loops contributed by the heavy chain are called H1, H2, and H3. Because of their direct role in binding antigen, the loops are also referred to as the complementarity determining regions (CDRs). The minimal structure capable of displaying antigen binding ability is the variable fragment (Fv) which consists of the two terminal hypervariable domains V_H and V_L in the Fab. To be structurally stable, a synthetic linker is generally introduced between the V_H and V_L chains, producing a chimeric single chain Fv (scFv) fragment. Structural studies of antigen binding, both experimental and theoretical, typically focus on the Fab or Fv - antigen interaction. In this chapter we focus primarily on approaches to modeling Fv domains.

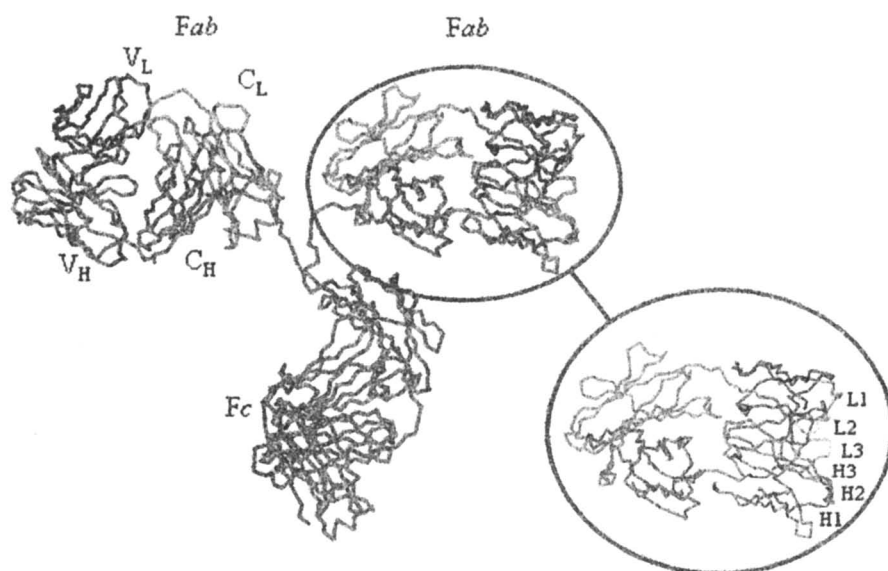


Figure 1. IgG antibody structure indicating the variable and constant domains. The hypervariable loops associated with the light (L1-3) and heavy (H1-3) chains in the Fv are indicated. The figure was created using Pymol(15) from the PDB file entry 1IG, the first structure of an intact IgG to be reported (16, 17). (See color page 1 in chapter.)

Various schemes have been proposed to characterize the amino acid sequence of CDRs. A useful technique for characterizing hypervariable sequences is based on the fact that a disulfide bond is conserved in both the light heavy chains. A review of conserved residues is provided in by Chothia *et al.*

(18). To ensure that the conserved cysteine residues always have the same number in the sequence, when compared to other antibodies, Kabat introduced a scheme that accounts for insertions and deletions in the CDRs (19). The Kabat numbering scheme uses a sequential numbering of the residues, but where the possibility of insertions occurs, a letter follows the numbers. According to this numbering scheme, the three loops arising from the heavy chain are approximately within the amino acid ranges 31-35b, 50-65 and 95-102 for H1, H2 and H3, respectively. Similarly, the loops formed by the light chains arise from the range of amino acid residues 24-34, 50-56 and 89-97. The Kabat scheme is still widely in use, although alternative schemes have been proposed (20-22).

An alternative to characterizing the hypervariable loops by chain length alone is to employ a scheme that recognizes that despite variations in sequence, the CDRs adopt only a limited number of conformations, called canonical structures. Canonical structures depend on the number and nature of the amino acids present in each loop (20, 21, 23, 24). Thus, a combination of the length and sequence can be used to classify and characterize the majority of the loop structures.

Antibody Comparative (Homology) Modeling

Given the relatively large number of x-ray diffraction structures for antibody Fab and Fv fragments, as well as the constancy of much of the 3-D framework, and the ability to allocate canonical structures to the CDRs, Fab and Fv fragments are well suited to comparative modeling. Below we present a synopsis of comparative modeling, with a particular focus on application to antibody fragments, however the reader is directed to the appropriate references for in depth discussions of the technique and related issues. While the terms homology modeling and comparative modeling are frequently used interchangeably, homology modeling implies a reliance solely on similarities between primary amino acid sequences. A general rule of thumb is that successful homology modeling requires a sequence identity greater than 70% (25). A significantly lower identity may be acceptable when it is known for example that the target and the template proteins share functional similarities, belong to the same family, etc. In this case, the approach is often described as comparative, rather than homology modeling. Thus, as concerns antibody modeling, as all of the template structures are themselves antibody fragments, the latter term will be employed in this discussion.

The first step in antibody modeling is to identify one or more experimental structures Fab or Fv fragments having adequate sequence similarity with the

target sequence that will serve as structural templates for the model. There are several options when determining the alignment and consequently the similarity and identity of two or more sequences. A straightforward approach is to perform a pair wise alignment using a fast heuristic alignment tool, such as FASTA (26) or BLAST (27, 28). A number of web-based databases are available for alignment searches (29-32). In addition to having the primary sequence of the target protein, it is necessary to select an appropriate alignment scoring matrix. This matrix provides a sophisticated method for evaluating amino acid similarity, i.e. not only does it determine whether two amino acids share functional characteristics, it also accounts for steric, electronic and hydrophobic similarity. Permutations and substitutions of amino acids to others with similar properties are therefore allowed and accounted for (25). Commonly used scoring matrices are PAM and BLOSUM. Matches can be altered depending on which scoring matrix is used, some are appropriate for searches of closely related amino acid sequences, e.g. PAM30, whereas others, like the PAM250, are more suitable when searching for matches for more distantly related amino acids (25). In addition to the identification of similar residues, the sequence alignment identifies any insertions or deletions in the sequences; a feature that is somewhat dependent on the nature of the alignment algorithm, and which is critical to the modeling of antibody CDR loops.

Once selected, the experimental structures are then retrieved, for example from the RCSB Protein Data Bank (33). In the case of identical residues, the 3D coordinates for the entire residue are transferred to the model, otherwise the backbone atoms are transferred and the positions of the side chains estimated. A variety of methods exist, for example those employed in the freeware program SwissPDB tool (34, 35), or commercial software such as InsightII (36) or MOE (37), to add the side chain atoms in appropriate conformations. For example, it is common to transfer the side chain orientations from a library of known conformations, which may then be subjected to energy-based refinement.

Treatment of the Complementary Determining Regions

Since antigenicity (affinity for antigen) depends on the interactions between the antigen and the hypervariable loops, it is critically important that these loops be modeled in a way that is in agreement with the expected canonical structures, and that the side chains adopt favorable orientations. The structure of the H3 CDR loop is less readily characterized (20, 21, 24) than the others, and careful refinement of the comparative model is required. Over the past decade much effort has been put into approaches to enhancing the accuracy of the modeled structures of hypervariable loops (23, 38-43). Some of the methods are based on

computational molecular dynamics (MD) simulations, sequence threading, fragment matching, or *ab initio* techniques (25, 43).

A combination of more than one method may be required in order to adequately model the CDRs. For example, once the peptide motifs in the hypervariable loops have been identified, a search for sequence matches can be performed, in a similar manner as used during homology modeling. However, a scoring matrix suitable for closely related amino acids should be used. In addition, it will have to be ensured that the resulting matches actually are from CDRs, otherwise, they are unlikely to fit into the classes of canonical structures. Each loop can then be grafted into a relatively generic Fab or Fv template by superimposing three to four residues to either side of the CDR loop on the corresponding template residues, as described by Morea et al. (40). In addition, it is likely that more than one potential template structure will be identified for each loop, differing in both sequence homology and 3D structure. Thus, it is possible to develop several initial models for the Fv that differ in the conformations of the CDR loops. Choosing the most representative structure may be challenging, however, access to more than one model provides multiple initial structures for MD refinement. In principle, within the limitations of the force field and the simulational time, MD simulations will converge to a common structure independent of initial conformation, thereby providing a level of confidence in the final model.

An alternative to identifying and grafting loop sequences into a template, is to identify independent matches for the entire, V_H and V_L chains. Particular attention must nevertheless be given to the canonical structures and homologies of the CDRs, and further to the presence of any insertions or deletions. Whether it is preferable to locate a high identity match that has an insertion or deletion, or a match with lower identity, but with the same number of residues is a matter of choice. In this chapter we present the results from applying this latter approach, that is, we develop models based on matches to the V_H and V_L chains.

Generation of the Fv fragment

In the event that models for the heavy and light domains were generated from different experimental structures, they must be oriented relative to each other to obtain the complete antibody Fv fragment. We have approached this important issue by aligning the modeled chains on Fab or Fv template structures whose V_H and V_L domains have high *structural similarity* with those obtained from the modeling. That is, after initial models for the V_H and V_L domains are generated, the protein structure database is searched, with for example the web tools DALI (44) or MSDFold (45), for a structural match to each of the two individual variable domains. A single Fab or Fv is then selected that has high

structural similarity (as determined by a low root mean squared deviation (RMSD) in the backbone C α positions) with each of the variable domains. The modeled V_H and V_L domains are then mutually aligned by superimposing them onto the Fab or Fv template. Thus, a structural model for the Fv is obtained by employing three different structures; one that serves as a template for each of the V_H and V_L domains, selected on the basis of sequence (overall and CDR) similarity, and one used as a structural framework onto which the modeled V_H and V_L domains are superimposed.

Comparison of Fv models with experimental 3D structures

To illustrate the accuracy with which carbohydrate-binding Fv domains may be predicted, we have generated models for two representative fragments, and then compared the resultant models with the reported experimental structures for each. The first Fv is from an antibody (mAb 735) against the capsular polysaccharide from the pathogenic bacterium *Neisseria meningitidis* type B and displays a groove-like binding site, characteristic of an antibody that binds a polysaccharide in a lengthwise manner. The second is from an antibody (mAb Se155-4) against the O-antigen from *Salmonella paratyphi* B and displays a cavity-like binding site, typical of a structure that binds a terminal cluster of carbohydrate residues. Lastly, we predict the structure of a Fv for the antibody against the capsular polysaccharide from type III group B *Streptococcus*, for which the experimental structure is still unknown.

Fv from mAb 735: The 2.8 Å x-ray diffraction structure for the Fab fragment for this monoclonal antibody (pdb id: 1PLG) was reported by Evans et al. in 1995 (46) and was not in any way employed when developing the comparative of the Fv for mAb 735 (Fv 735). A Blast search for sequences similar to the V_H and V_L domains in Fv 735 revealed that the light chain in Fab 2PCP (47) matched the V_L in Fv 735 with 93% identity, without any gaps in the CDRs, and that the heavy chain in Fab 1AE6 (48) had a 77% match for V_H, with a 1% gap. Preliminary models for the V_H and V_L domains of Fv 735 were then generated by residue transfer using SwissPDB (34, 35). The side chains of non-identical residues generated using the rotamer library available in SwissPDB. In this particular program, the most preferable rotamer is chosen by a score, which is based on the obtaining the smallest number of bad steric contacts with existing atoms. The score also depends on the number of hydrogen bonds for each rotamer. A subsequent structural search for the best Fab or scFv fragment on which to align the two domains identified scFv 1KTR (49). Superimposing the modeled V_L domain onto the C α atoms in 1KTR resulted in a value for the RMSD of 0.92 Å, and for the V_H domain a value of 0.90 Å.

The hypervariable loops in Fv 735 (1PLG) belong to canonical structures 3, 1, and 1 for CDR-L1, CDR-L2 and CDR-L3, respectively; and 1 and 2 for CDR-H1 and CDR-H2 in the heavy chain. The same canonical structures are present in the template structures used to construct our model, that is, in 2PCP and 1AE6. CDR-H3 adopts a kinked conformation (24, 50).

Superimposing the modeled Fv domains with the experimental x-ray diffraction structure (1PLG) gave very low RMSD values for the backbone C α atoms of 0.44 Å (see Table 2). When all non-hydrogen atoms of the side chains were included in the alignment of the two structures, the RMSD increased to 0.79 Å (see Figure 2 and Table 2).

Fv from mAb Se155-4: The 2.1 Å x-ray diffraction structure for the Fab from this antibody (pdb id: 1MFD) was reported in 1994 by Bundle et al. (51). A BLAST search for the V_H and V_L domains identified Fab 1NGQ (52) with 98% identity with the light chain, and Fab 1KEG (53) which for the heavy chain had 80% identity, and 2% gaps. Again using SwissPDB, initial models were generated for the V_H and V_L domains in Se155-4. A structural search for a suitable template identified Fv 1A6W (54). Overlaying the light and heavy chains of the modeled domains onto the Fv template gave a value for the RMSD in the C α positions of 0.31 Å for V_L and 0.79 Å for V_H.

The hypervariable loops of Fv Se155-4 can be assigned to the canonical classes 7, 1, and 4 for CDR-L1, L2 and L3. The heavy chain has canonical structure 1 for CDR-H1 and 2 for CDR-H2 (23, 55). CDR-H3 has a kinked conformation (24, 50, 56).

Superimposing the modeled Fv domains on the experimental x-ray diffraction structure (1MFD) again gave a low RMSD value for the backbone C α atoms of 0.62 Å. When the non-hydrogen atoms of the side chains were included in the alignment of the two structures, the RMSD increased to 0.73 Å (see Figure 2 and Table 2).

Fv from 1B1: Finally, a comparative model has been generated for an antibody against the surface polysaccharide from type III group B *Streptococcus*, which has no previously reported structure. The best alignment for the light chain was achieved with Fab 1QFU (57), with a match of 95% and no gaps. Similarly, the heavy chain aligned well with Fab 1NLD (58), having one gap out of 113 residues, and a 90% match overall. The most favorable structural match for the individual chains was Fv 1G7H(59), with RMSD values of 0.62 Å for the light chain and 0.83 Å for the heavy chain.

The canonical structures of the light chain are 4 for CDR-L1, and 1 for both CDR L2 and L3. CDR-H1 and H2 both have canonical structure 1. Although CDR-H3 is very short, only 4 residues, it adopts a kinked structure, similar to that present in both Fv 735 and Se155-4.



*Figure 2. Overlay of comparative FV models with x-ray diffraction structures (crystal structure is blue, comparative model light chain red and the heavy chain cyan). Model **a** superimposes onto 1PLG with a Ca RMSD of 0.44 Å, while **b** superimposes on 1MFD with a value 0.62 Å. The predicted homology model **c** for Fv 1B1 is shown. (See color page 1 in chapter.)*

Refinement of Comparative Models using MD Simulation

Despite the good agreement between the backbone conformations predicted by comparative modeling and those determined by x-ray diffraction, the positions of the side chains in the modeled Fv generally need to be further refined. Energy minimization with a classical force field can be employed to remove unfavorable steric interactions, however, to improve rotamer distributions a technique such as molecular dynamics simulation is required.

MD simulations enable the positions of the atoms in a molecule to vary as a function of the force on the atoms at a particular temperature, usually 300 K. In contrast to energy minimization, the nonzero temperature in MD simulations contributes kinetic energy, which enables the crossing of energy barriers, such as during bond rotation. Thus, MD simulations result in a time-series or trajectory for the atomic coordinates, rather than a single structure. MD simulations provide an estimate of the contribution of dynamics to the molecular structure. To analyze the positional data from a simulation requires that they be treated as an ensemble of conformations. The equations defining atomic motion are

classical in origin and begin with the familiar Taylor series expansion of position (x) in terms of velocity (v) and acceleration (a).

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(\Delta t)^2 + \dots, \quad [1]$$

Equation 1 may be reformulated as the Verlet algorithm (25, 60) in terms only of positions and accelerations as:

$$x(t + \Delta t) \approx 2x(t) - x(t - \Delta t) + a(\Delta t)^2 \quad [2]$$

Using Equation 2 to compute the position of an atom at some future time ($t + \Delta t$) requires only a knowledge of the atom's current and previous positions and its acceleration. The atomic accelerations are readily derived from the relationship between force on the atom (F), atomic mass (m) and potential energy (V) using Newton's second law:

$$F_i = m_i a_i(t) = -\frac{\partial V}{\partial x} \quad [3]$$

The energy (V) is that defined by the classical mechanical force field as:

$$V = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} K_{\phi,n} [1 + \cos(n\phi + \gamma)] \\ + \sum_{\text{vanderWaals}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] + \sum_{\text{Electrostatics}} \left[\frac{q_{ij}}{\epsilon R_{ij}} \right] \quad [4]$$

which is the same mathematical definition of V employed during energy minimization. For insight into the particular parameters (K , θ , ϕ , q etc.) that make up the force field, see for example (61). For molecular simulations the magnitude of the time step (Δt) is limited by the highest frequency motion, and is typically between 1-2 fs. Thus, many millions of iterative cycles are required in order to simulate time scales appropriate for the motion of side chains and loops.

Importantly, for biomolecular systems, and particularly for carbohydrates, water will have to be accounted for, preferably explicitly, which means that thousands of water molecules are typically included in the simulation.

Application of MD simulations to Fv fragments

In applying MD simulations to Fab and Fv fragments, complete solvation in a periodic box or large droplet of explicit water molecules may result in a system that is extremely computationally demanding. While complete solvation may be required, as here, in order to allow the entire antibody fragment to relax, it may be sufficient to solvate only the region covering the CDRs (11). This is a useful approximation when only the motion of the CDRs is under investigation, as due to the reduction in the number of participating water molecules, it allows significantly more sampling of the CDR conformations, for a given real time of simulation. A localized solvation droplet is justifiable since structure of the constant domains in the Fab is highly conserved.

Each of the homology modeled Fv domains was subjected to 3 ns of MD simulation using explicit water at 300 K with the PARM99 force field in Amber 7 (62). Prior to the MD simulations, all three systems were energy minimized and then heated from 5 K to 300 K over 50 ps using a 2 fs timestep. For each Fv two MD simulations were performed; one in which the atoms in the backbone were restrained (with a weak harmonic potential), allowing only the side chains to move, and one in which all atoms were allowed full motional freedom. Restraining the backbone is generally necessary when an incomplete treatment of solvation is employed, or then some uncertainty exists as to the stability of the antibody fragment, as with Fv domains. In each MD simulation the Fv domains were explicitly solvated by a 40 Å radius droplet of water molecules placed at the geometric center of each Fv. This size of droplet fully encapsulated the protein leaving a boundary of approximately 10 Å between the surface of the protein and the edge of the droplet. In each system approximately 6900 water molecules were present, see Table 1 for a summary of the simulation details.

Generally, a number of bad steric contacts are introduced during the homology modeling and these were reduced but not eliminated during energy minimization (see Table 2). Further, energy minimization alone is unable to correct poor rotamer orientations for the side chains, nor can it correct any problems associated with the backbone, which may be of particular importance in the CDRs. Bad steric contacts may be effectively removed by a short MD simulation of approximately 1-5 ns, which is also adequate to significantly improve side chain rotamer distributions (see Table 2). Simulations in excess of 10 ns would provide further insight into the degree of flexibility associated with the loop regions, but are proportionately more demanding.

Results of the 3 ns MD simulations

After the MD simulations, in which the backbones were restrained, the RMSD values for the side chains (not including hydrogen atoms) were 1.34 Å

and 1.25 Å for Fv 735 and Fv Se155-4, respectively. For the MD simulations in which the entire Fv was permitted to move, the RMSD for the Ca positions in Fv 735 with respect to the original crystal structure, was a very modest 0.96 Å, while the side chain RMSD increased to 1.60 Å (see Table 2). Similarly, under the same MD conditions, for the Fv from Se155-4 the RMSD values were 1.00 and 1.65 Å, for the Ca and side chain positions, respectively. These are larger numbers than for the homology models themselves, however, a moderate RMSD is expected (11), given that the side chains are now moving in an aqueous environment at room temperature, rather than populating single static rotamers in a crystal lattice. Further insight into the quality of the models is provided by an analysis of the backbone torsion angles and side chain rotamer distributions.

Table 1. Computational details for MD simulation of Fv-fragments.

| | Fv 735 ^a | Fv Se155-4 ^b | Fv 1B1 ^c |
|----------------------|----------------------|-------------------------|----------------------|
| Residues (Fv) | 226 | 224 | 222 |
| Atoms (Fv) | 3466 | 3345 | 3398 |
| Waters | 6891 | 6924 | 6941 |
| Atoms (Total) | 24122 | 24117 | 24221 |
| Disulfides | 23 - 93 133 - 207 | 22 - 90 130 - 204 | 23 - 93 134 - 207 |
| V _L Range | 1-111 | 1-108 | 1-112 |
| V _H Range | 112-226 | 109-224 | 113-222 |
| Net Charge | 3 | 1 | -1 |

^aV_H: QIQLQSGPELVRPGASVKISCKASGYTFDYYIHVWKQRPGEGLWIGWIYPGSGNTKYNEKFKG
KATLTVDTSSSTAYMQLSSLTSEDSAVYFCARGGKGFAMDYWGQGTSTVTV;

V_L: DVVMTQTPLSLPVSLGDAQASISCRSSQSLVHSGNTYLYWYLQKPGQSPKPLIYRVSINRFSGVDPDR
FSGSGSGTDFTLKISRVEAEDLGVYFCFQGTHTVPYTFGGGTGLEIK

^bV_H: EVQVQQSGTVLARPGASVKMSCKASGYTFNYYMHVWKQRPQGGLWIGWAIYPGNSATFYNHKF
RAKTKLTAVTSTITAYMELSSLTNEEDSAVYYCTRGHGYGDIYWGQASLTV;

V_L: QAVVTQESALTTSPGETVTLTCSRSTGTVTSGNHANWVQEKPDHLFTGLIGDNTNRRAPGVPA
RFSGSLIGDKAALTITGAQPEDEAIYFCALWCNNHWIFGGGTGLTV

^cV_H: QVQLLESQGLVQPSQSLTCTVSGFSLTSYGVHWRQSPGKLEWLGVWGGGSDYNAAFIS
RLSISRDFSRSSQVFFKMNLSLQADDTAIYCVRNWDYWGQGTTLTV;

V_L: DVVMTQTPLSLPVSLGDAQASISCRSSQSLSDSNGNSYLNWYLQKPGQSPQLLIYRVSINRFSGVLDLDR
FSGSGSGTDFTLKISRVEAEDLGVYFCQLTHTVPYTFGGGTGLEIK

Ramachandran analyses of the x-ray diffraction structures as well as the homology modeled and MD refined structures for the Fv domains of mAb 735 and Se155-4, as well as for the theoretical model for Fv 1B1 are presented in Table 2. Inspection of the data shows that the MD simulations can allow the residues to move into disallowed regions, however, loop structures are not well represented in a Ramachandran statistical analysis, due to their flexibility. This flexibility frequently prevents the loop regions from being resolved experimentally and therefore their conformations are under represented in crystallographic data bases. Perhaps more significant is the fact that in the

Table 2. Structural characterization of x-ray (*a*) and theoretical Fv models: *b*) Homology Model; *c*) Minimized with backbone restrained; *d*) Minimized without restraints; *e*) MD with backbone restrained; and *f*) MD without restraints

| mAb | Ramachandran Backbone Analysis (%) (63, 64) | | | | | | | | | |
|----------|---|------|---------|--------------------|------------|------------------------|---------------------------|------------|------|------|
| | RMSD ^a (Å) (15) | Core | Allowed | Generously Allowed | Disallowed | Number of Bad Contacts | Bad Rotamers (%) (65, 66) | Side Chain | All | Ca |
| <i>a</i> | - | 88.2 | 11.3 | 0.0 | 0.5 | 0 | 14.1 | - | - | - |
| <i>b</i> | 0.44 | 83.2 | 14.7 | 1.6 | 0.5 | 12 | 12.0 | 1.33 | 0.79 | 1.05 |
| <i>c</i> | 0.44 | 82.3 | 15.1 | 1.6 | 1.1 | 8 | 12.5 | 1.33 | 0.79 | 1.14 |
| <i>d</i> | 0.44 | 82.3 | 15.1 | 1.6 | 1.1 | 8 | 12.5 | 1.34 | 0.80 | 1.14 |
| <i>e</i> | 0.46 | 86.6 | 11.3 | 1.1 | 1.1 | 0 | 9.4 | 1.60 | 1.28 | 1.25 |
| <i>f</i> | 0.99 | 84.4 | 14.5 | 0.5 | 0.5 | 0 | 11.5 | 1.65 | 1.29 | 1.25 |
| <i>a</i> | - | 89.8 | 9.6 | 0.0 | 0.5 | 0 | 9.0 | - | - | - |
| <i>b</i> | 0.62 | 86.9 | 11.9 | 0.6 | 0.6 | 17 | 12.1 | 1.05 | 0.74 | 1.05 |
| <i>c</i> | 0.48 | 83.6 | 13.8 | 1.3 | 1.3 | 11 | 10.1 | 1.14 | 0.73 | 1.14 |
| <i>d</i> | 0.48 | 83.0 | 13.8 | 1.9 | 1.3 | 11 | 10.1 | 1.14 | 0.73 | 1.14 |
| <i>e</i> | 0.46 | 81.8 | 15.1 | 1.9 | 1.3 | 0 | 5.9 | 1.25 | 0.85 | 1.25 |
| <i>f</i> | 1.00 | 76.7 | 21.4 | 1.3 | 0.6 | 0 | 7.1 | 1.65 | 1.29 | 1.65 |
| <i>a</i> | - | - | - | - | - | - | - | - | - | - |
| <i>b</i> | - | 80.5 | 18.3 | 0.0 | 1.2 | 22 | 15.0 | - | - | - |
| <i>c</i> | - | 83.4 | 15.4 | 0.0 | 1.2 | 12 | 15.5 | - | - | - |
| <i>d</i> | - | 83.4 | 15.4 | 0.0 | 1.2 | 12 | 15.5 | - | - | - |
| <i>e</i> | - | 86.1 | 12.3 | 1.1 | 0.5 | 0 | 15.5 | - | - | - |
| <i>f</i> | - | 76.3 | 21.9 | 1.2 | 0.6 | 0 | 7.0 | - | - | - |

^aRelative to x-ray diffraction structure (*a*).

^bPDB id given in parentheses.

absence of restraints, the backbones appear to relax away from the core region of the Ramachandran surface. Despite this, the average RMSD values for the backbones were only approximately 1 Å away from the experimental structures. Thus, the MD simulations are not disturbing the protein fold noticeably, but rather are introducing local fluctuations around the backbone torsion angles. It is unclear to what extent this results from the normal motions associated with a protein under physiological conditions or from a weakness in the force field parameters related to the backbone. In contrast to the effects of the unrestrained MD simulations, the backbone conformations of the comparative models are each consistent with expectations based on the crystallographic data. This feature is not surprising given that the backbone conformations in the comparative models were all transferred from x-ray crystal structures. A significant benefit of the MD simulations becomes apparent in terms of the nearly complete removal of bad steric contacts and in terms of the improvement in rotamer distributions. Relaxation of the backbone facilitated rotamer transitions, as indicated by their improved distributions, however, even with the backbone restrained the MD results led to an improvement in rotamer distributions relative to both the comparative models and the experimental structures.

Summary

A protocol to construct 3D models of the antigen-binding domains of anti-carbohydrate antibodies is illustrated. The method requires only the primary sequence of the heavy and light chains, and employs established comparative modeling and MD techniques. While the results for the systems examined are encouraging, it should be emphasized that the models are only as good as the data they are based upon. Errors in primary sequence may have catastrophic effects, as they may remain largely unrecognized. Similarly, inaccuracies in force field parameters, or in the treatment of solvation, or poor selection of template structures may degrade the resolution of the models. Nonetheless, the combination of methodologies presented here provides a remarkably accurate approach to construct and refine 3D Fv structures. Given the relative ease of obtaining the primary sequences for Fv domains, in contrast to the difficulties associated with determining the experimental 3D structures of these domains, this computational approach may prove particularly useful.

References

1. C. Taroni, S. Jones, and J. M. Thornton. (2000) *Prot. Eng.* 13, 89-98.
2. S. T. Iobst, and K. Drickamer. (1996) *J. Biol. Chem.* 271, 6686-6693.
3. F. R. Carbone, and P. A. Gleeson. (1997) *Glycobiology* 7, 725-730.
4. A. R. Kolatkar, A. K. Leung, R. Isecke, R. Brossmer, K. Drickamer, and W. I. Weis. (1998) *J. Biol. Chem.* 273, 19502-19508.
5. G. Cunto-Amesty, P. Luo, B. Monzavi-Karbassi, A. Lees, and T. Kieber-Emmons. (2001) *Vaccine* 19, 2361-2368.
6. N. K. Vyas, M. N. Vyas, M. C. Chervenak, D. R. Bundle, B. Mario Pinto, and F. A. Quijcho. (2003) *Proc. Natl. Acad. Sci. USA* 100, 15023-15028.
7. J. H. Naismith, and R. A. Field. (1996) *J. Biol. Chem.* 271, 972-976.
8. C. Clarke, R. J. Woods, J. Gluska, A. Cooper, M. A. Nutley, and G.-J. Boons. (2001) *J. Am. Chem. Soc.* 123, 12238-12247.
9. E. I. Park, S. M. Manzella, and J. U. Baenziger. (2003) *J. Biol. Chem.* 278, 4597-4602.
10. E. Liwsowska. (2002) *Cell. Mol. Life Sci.* 59, 445-455.
11. A. Pathiaseril, and R. J. Woods. (2000) *J. Am. Chem. Soc.* 122, 331-338.
12. D. R. Rose, M. Przybylska, R. J. To, R. P. Oomen, E. Vorberg, N. M. Young, and D. R. Bundle. (1993) *Prot. Sci.* 2, 1106-1113.
13. Y. Han, T. Kanbe, R. Charniak, and J. E. Cutler. (1997) *Infect. Immun.* 65, 4100-4107.
14. M. Nitz, C.-C. Ling, A. Otter, J. E. Cutler, and D. R. Bundle. (2002) *J. Biol. Chem.* 277, 3440-3446.
15. W. L. DeLano. DeLano Scientific LLC, San Carlos, CA, USA.
16. L. J. Harris, S. B. Larson, K. W. Hasel, J. Day, A. Greenwood, and A. McPherson. (1992) *Nature* 360, 369.
17. L. J. Harris, S. B. Larson, K. W. Hasel, and A. McPherson. (1997) *Biochemistry* 36, 1581.
18. C. Chothia, I. Gelfand, and A. Kister. (1998) *J. Mol. Biol.* 278, 457-479.
19. E. A. Kabat, T. Te Wu, H. M. Perry, K. S. Gottesman, and C. Foeller (1991) *Sequences of Proteins of Immunological Interest*, 5th ed., U.S. Dept of Health and Human Services.
20. C. Chothia, A. M. Lesk, A. Tramontano, M. Levitt, S. J. Smith-Gill, G. Air, S. Sheriff, E. A. Padlan, D. Davies, W. R. Tulip, P. M. Colman, S. Spinelli, P. M. Alzari, and R. J. Poljak. (1989) *Nature* 342, 877-883.
21. B. Al-Lazikani, A. M. Lesk, and C. Chothia. (1997) *J. Mol. Biol.* 273, 927-948.
22. A. Honegger, and A. Plückthun. (2001) *J. Mol. Biol.* 309, 657-670.
23. A. C. R. Martin, and J. M. Thornton. (1996) *J. Mol. Biol.* 263, 800-815.
24. H. Shirai, A. Kidera, and H. Nakamura. (1996) *FEBS Letters* 399, 1-8.

25. A. R. Leach (2001) *Molecular Modelling. Principles and Applications*, Second Edition ed.
26. W. R. Pearson, and D. J. Lipman. (1990) *Proc. Natl. Acad. Sci. USA* 85, 2444-2448.
27. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. (1990) *J. Mol. Biol.* 215, 403-410.
28. S. F. Altschul, T. L. Madden, A. A. Schaeffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. (1997) *Nucl. Acid Res.* 25, 3389-3402.
29. www.ebi.ac.uk/Tools.
30. <http://www.ncbi.nlm.nih.gov/BLAST/>.
31. <http://searchlauncher.bcm.tmc.edu/>.
32. <http://genopole.toulouse.inra.fr/blast/megablast.html>.
33. www.rcsb.org/pdb/.
34. N. Guex, and M. C. Peitsch. (1997) *Electrophoresis* 18, 27 14-2723.
35. <http://www.expasy.org/spdbv/>.
36. Biosym Technologies. (1998), San Diego.
37. Chemical Computing Group Inc. (1997-2004).
38. T. A. Jones, and S. Thirup. (1986) *EMBO J.* 5, 819-822.
39. V. Morea, A. Tramontano, M. Rustici, C. Chothia, and A. M. Lesk. (1997) *Biophys. Chem.* 68, 9-16.
40. V. Morea, A. M. Lesk, and A. Tramontano. (2000) *Methods* 20, 267-279.
41. R. E. Bruccoleri, E. Haber, and J. Novotný. (1988) *Nature* 335, 564-568.
42. A. C. R. Martin, J. C. Cheetham, and A. R. Rees. (1989) *Proc. Natl. Acad. Sci. USA* 86, 9268-9272.
43. V. Hornak, and C. Simmerling. (2003) *Proteins* 51, 577-590.
44. <http://www.ebi.ac.uk/dali/Interactive.html>.
45. <http://www.ebi.ac.uk/lmsd-srv/ssm/ssmcite.html>.
46. S. V. Evans, B. W. Sigurskjold, H. J. Jennings, J.-R. Brisson, R. To, W. C. Tse, E. Altman, M. Frosch, C. Weisgerber, H. Kratzin, S. Klebert, M. Vaesen, D. Bitter-Suermann, D. R. Rose, N. M. Young, and D. R. Bundle. (1995) *Biochemistry* 34, 6737-6744.
47. K. Lim, S. M. Owens, L. Arnold, J. C. Sacchettini, and D. S. Linthicum. (1998) *J. Biol. Chem.* 273, 28576-28582.
48. M. J. Banfield, D. J. King, A. Mountain, and R. L. Brady. (1997) *Proteins: Struct. Funct. Genet.* 29, 161-171.
49. M. Kaufmann, P. Lindner, A. Honegger, K. Blank, M. Tschopp, G. Capitani, A. Plueckthun, and M. G. Grütter. (2002) *J. Mol. Biol.* 318, 135-147.
50. V. Morea, A. Tramontano, M. Rustici, C. Chothia, and A. M. Lesk. (1998) *J. Mol. Biol.* 275, 269-294.
51. D. R. Bundle, H. Baumann, J.-R. Brisson, S. Gagne, A. Zdanov, and M. Cygler. (1994) *Biochemistry* 33, 5183-5192.
52. R. Mizutani, K. Miura, T. Nakayama, I. Shimada, Y. Arata, and Y. Satow. (1995) *J. Mol. Biol.* 254, 208-222.

53. H. Yokoyama, R. Mizutani, Y. Satow, K. Sato, Y. Komatsu, E. Ohtsuka, and O. Nikaido. *To be Published*.
54. T. Simon, K. Henrick, M. Hirshberg, and G. Winter. *To be Published*.
55. <http://www.bioinf.org.uk/abs/>.
56. <http://antibody.bath.ac.uk/H3struct.html>.
57. D. Fleury, B. Barrere, T. Bizebard, R. S. Daniels, J. J. Skehel, and M. Knossow. (1999) *Nat. Struct. Biol.* 6, 530-534.
58. C. Davies, J. C. Beauchamp, D. Emery, A. Rawas, and H. Muirhead. (1997) *Acta Cryst. D53*, 186-194.
59. E. J. Sundberg, M. Urrutia, B. C. Braden, J. Isern, D. Tsuchiya, B. A. Fields, E. L. Malchiodi, J. Torno, F. P. Schwarz, and R. A. Mariuzza. (2000) *Biochemistry* 39, 15375-15387.
60. D. Frenkel, and B. Smit (2001) *Understanding Molecular Simulation. From Algorithms to Applications*, Academic Press.
61. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. (1995) *J. Am. Chem. Soc.* 117, 5179-5197.
62. D. A. Case, D. A. Pearlman, J. W. Caldwell, T. F. Cheatham III, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsurii, H. Gohlke, R. J. Radmer, Y. Duan, J. Pitera, I. Massovo, G. L. Seibel, U. C. Singh, P. K. Wiener, and P. A. Kollman. (2002), University of California, San Francisco.
63. R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton. (1993) *J. Appl. Cryst.* 26, 283-291.
64. R. A. Laskowski, M. W. MacArthur, D. K. Smith, D. T. Jones, E. G. Hutchinson, A. L. Morris, D. Naylor, D. S. Moss, and J. M. Thornton. (1994).
65. S. C. Lovell, I. W. Davis, W. B. Arendall III, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, and Richardson, D. C. (2003) *Proteins: Struct. Funct. Genet.* 50, 437-450.
66. <http://kinemage.biochem.duke.edu/molprobity/main.php>.

Chapter 12

Carbohydrate Structure from NMR Residual Dipolar Couplings: Is There a Correlation between Lactose's Anomeric Configuration and Its Three-Dimensional Structure?

Darón I. Freedberg, Susan O. Ano, Scott E. Norris,
and Richard M. Venable

Department of Economics, Haverford College, Haverford, PA 19041

The three-dimensional structures of α - and β -lactose were analyzed using residual dipolar couplings in two liquid crystalline media, ^1H - ^1H NOEs and force field calculations. Previously unreported long-range proton-carbon J values were measured in natural abundance as part of RDC data acquisition and are included here. The slow anomerization rate produced two subspectra, and permitted a direct structural comparison of α - and β -lactose. Fits of the data to rotamers of α -lactose define glycosidic torsion angle ranges $150^\circ \leq \phi \leq 180^\circ$ and $150^\circ \leq \psi \leq 180^\circ$, while those for β -lactose $150^\circ \leq \phi \leq 210^\circ$ and $210^\circ \leq \psi \leq 270^\circ$. In addition, chemical shift data (500 MHz) imply that there are differences between the structures of α - and β -lactose. Natural abundance HSQC-NOESY spectra measured at 700 MHz suggest that the H1-H4' α internuclear distance is similar to the H1-H4' β distance, but does not preclude differences in glycosidic torsions. Altogether, the experimental results support the notion that α -lactose and β -lactose have different three-dimensional structures in aqueous solution.

Introduction

Carbohydrates are ubiquitous in nature and participate in a wide variety of biological processes. They play roles in cell-cell interactions such as immune responses, fertilization, inflammation, and cell growth (1), influence protein folding and stability (2), and may be involved in signal transduction (3,4). Consequently, they are used as vaccines, potential drug delivery systems, and antibiotics (3,5,6). Carbohydrates often function in these processes as glycoconjugates with proteins or lipids (2,7). Given the variety of monosaccharides, linkage types, and functional group modifications, oligosaccharides alone have potential structural complexity unsurpassed by other classes of biomolecules. Conjugation to a protein or lipid moiety, as in conjugate vaccines or glycoconjugate mimics, further increases this diversity. Yet enzymes and other proteins are able to distinguish carbohydrates that deviate in very subtle ways. Some of these subtle differences such as stereochemical modifications result in morphological changes. Characterization of three-dimensional carbohydrate structure should aid in establishing structure-function relationships, facilitate development of carbohydrate-based therapies, and provide insight into shape-based protein-carbohydrate interactions.

Six-membered chair-like pyranose monosaccharide units (e.g. glucose, galactose, mannose, and fucose) are building blocks of many oligosaccharides. Three-dimensional carbohydrate structure is defined by the glycosidic torsion angles ϕ and ψ , which describe the relative orientations of the individual rings. Along with the length of the oligosaccharide, these angles can lead to global structure motifs similar to those found in proteins and nucleic acids. For example, di- and trisaccharides of the rigid mannose residue in 1,2-linked β -mannans form antigens, and longer oligosaccharides of the same building block form compact helical motifs (8).

NMR (nuclear magnetic resonance) methods for carbohydrate structure determination include NOEs (nuclear Overhauser enhancements), J couplings to obtain information regarding torsion angles, and RDCs (residual dipolar couplings). NOEs and J couplings are limited to local geometry. RDCs, however, are related to overall molecular structure and consequently shape. RDCs are scaled dipolar couplings relating the angle of an internuclear vector to the static magnetic field through the equation:

$$D = \left(\frac{\hbar \gamma_H \gamma_C}{r^3} \right) (3 \cos^2 \theta - 1)$$

Here, D is the RDC, \hbar is Planck's constant divided by 2π , γ_H and γ_C are the magnetogyric ratios of ^1H and ^{13}C , respectively, r is the internuclear distance, and θ is the average angle subtended by the internuclear vector and the magnetic field.

The utility of RDCs in structure determination of biological systems was recognized when the magnetic susceptibility of certain molecules caused slight alignment at high magnetic fields (9). Many biomolecules have magnetic susceptibility anisotropy that is too small for detection of preferential alignment, *i.e.*, the distribution of their orientations in the magnetic field is nearly isotropic. In these instances, the angular term in the above equation averages to almost zero, and direct measurement of RDCs in unsubstituted carbohydrates is impossible. Though these molecules will not spontaneously align, tunable, non-isotropic orientation distributions can be induced in these systems through the use of weakly aligning aqueous liquid crystalline media (10-18), or other methods (19,20). For these systems, it is possible to measure RDCs for almost all classes of biomolecules while still maintaining many of the features of a high-resolution solution NMR spectrum (10,21-27).

RDCs are complementary to NOEs and trans-glycosidic J couplings in three-dimensional carbohydrate structure determination because they provide multiple simultaneous constraints and allow the relation of multiple internuclear vectors through one common alignment frame. Thus, RDCs are a source of long-range constraints previously unavailable from OEs, trans-glycosidic J values or even the combination of the two techniques. RDCs can be used to find the relative orientation of rigid subunits making up oligo- and polysaccharides. Structure analysis using multiple RDC constraints requires simultaneous fitting to all values, ensuring the best possible structures. One challenge in studying carbohydrate structure using this method is the need for five independent measurements for each rigid fragment. This is because many of the bond vectors in pyranose rings are nearly parallel and therefore do not provide independent constraints (12). This has been overcome by a recently developed method that has added ${}^nD_{CH}$ (long-range proton-carbon RDCs; n = number of bonds) values to the structure information (21,28). These RDCs, along with one-bond (${}^1D_{CH}$) and homonuclear (${}^3D_{HH}$) data, provide the necessary input for carbohydrate structure determination.

Structure determination using RDCs is accomplished through an order matrix analysis (29,30). To ensure an overdetermined fit, the order matrix elements, S_{kl} (where k and l are x, y, or z), must be obtained from simultaneous solution of five independent equations for experimentally measured residual dipolar couplings for each rigid subunit. Because interresidue motion may be present in carbohydrates it is prudent to determine the order parameters for each ring independently. In addition to the minimum number of measurements, this approach requires specification of molecular geometry in a defined molecular coordinate system, such as an x-ray or theoretically calculated structure. Subsequent diagonalization of this Saupe matrix yields Euler angles, α , β and γ ,



Figure 1. Equilibrium between α - and β -lactose and numbering schemes for α -lactose and β -lactose. Most OH groups are omitted for clarity. Note that H's (and their attached C's) on the glucose ring are labeled with a prime ϕ and ψ are torsional angles defined by $\phi = \text{C2-C1-O4-C4}'\alpha/\beta$, and $\psi = \text{C1-O4-C4}'\alpha/\beta\text{-C5}'\alpha/\beta$.

which relate the molecular coordinate system to the alignment frame (or principal order tensor frame), and order parameters S_{zz} , S_{xx} and S_{yy} . Determinations of Saupe matrices for rigid molecular subunits allow analysis of RDCs in terms of molecular structure, but these can also be used to understand dynamics (21,26,28).

Lactose (Figure 1) is a component of many larger oligo- and polysaccharides. This disaccharide and many closely related derivatives have been shown to affect tumor colonization and growth (31,32). Furthermore, lactose is of fundamental interest as a representative disaccharide for studying carbohydrate β -1-4 linkages. Additionally, α - and β -lactose are in equilibrium in solution ($K_{\text{eq}} \approx 0.55$) and have mostly well-resolved signals, which allows these anomers to be studied independently.

Previous studies of unmodified (*e.g.* unmethylated at the anomeric carbon) α - and β -lactose in either the solid or solution state are sparse (33-39). Lactose structure is characterized by the two linkage torsions, for which there are multiple conventions; in this chapter, we use $\phi = \text{C2-C1-O4-C4}'\alpha/\beta$, and $\psi = \text{C1-O4-C4}'\alpha/\beta\text{-C5}'\alpha/\beta$. X-ray crystallographic studies showed that α -lactose crystallized in the presence of either CaCl_2 or CaBr_2 and β -lactose in the monohydrate form had similar ϕ , ψ values around 170° , -135° (33,34,36). However, ϕ changes to $\sim 140^\circ$ in the monohydrate form of α -lactose (35). Previous non-NMR based solution studies of lactose (38,39) generally did not consider the impact of multiple anomers on data interpretation. NMR studies, which were able to detect the multiple anomers, determined structure on the basis of trans-glycosidic J values (37).

Here we present results from NMR three-dimensional structural studies of lactose (75 mM in 20 mM Na_2PO_4 , $\text{pD}^* = 6.8$) using RDCs obtained in two alignment media. Fitting RDCs reveals evidence for two different structures. New measurements showing differences in ^{13}C chemical shifts support this notion. We further probe these structures using RDCs and HSQC-NOESY spectra in natural abundance. Together, traditional techniques and RDCs can be

used to provide a complete picture of carbohydrate structure. We also show evidence for structural differences from 1D NMR data. For lactose, we find that the change in anomeric carbon configuration of the reducing sugar correlates with changes in remote glycosidic torsion angles.

NMR Residual Dipolar Couplings, NOEs and Calculations for Lactose

RDC measurements are straightforward because the techniques used to measure them are similar to J coupling measurements. In this report, we use ^1H - ^{13}C HSQC, long-range ^1H - ^{13}C HSQC and ^1H - ^1H COSY NMR spectra to measure coupling constants in isotropic solution and splittings in the oriented phase. The resulting $^1D_{\text{CH}}$, $^2D_{\text{CH}}$, $^3D_{\text{CH}}$ and $^3D_{\text{HH}}$ are listed in Table I. Figure 2 displays two natural abundance HSQC spectra where the ^1H - ^{13}C splitting in the isotropic phase is different from that in the oriented phase. This figure also illustrates that the two anomeric forms of the glucose ring are easily detected.

We generated 5,329 structures in 5° increments of ϕ and ψ using the CSFF (40) and Accelrys (41) molecular mechanics force fields. Our intent in generating the structures with CSFF and Accelrys is two fold. First, it is important that consistency between force fields is verified. Second, we have observed that in sucrose, structures resulting from different force field calculations sometimes yield conflicting structural results (28).

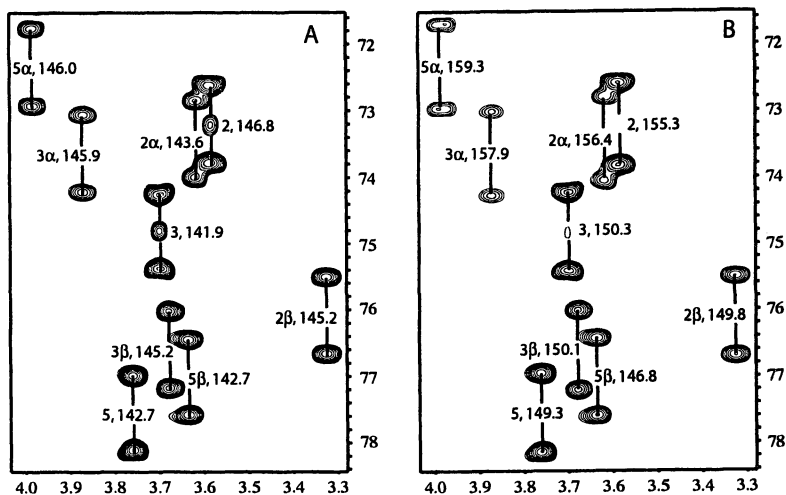


Figure 2. Natural abundance HSQC spectra of lactose displaying signals for the 2, 3 and 5 atoms of the galactose, α -glucose and β -glucose in isotropic solution (A), and in a solution of CTAB bicelles (B).

Table I. D values for α - and β -lactose in two liquid crystalline media

| <i>Atom pair</i> | D (DMPC/DHPC) | D (DMPC/DHPC/CTAB) |
|--|------------------------------------|-------------------------------------|
| H1-C1 | 5.49 \pm 0.20 | 8.84 \pm 0.20 |
| H2-C2 ^b | 5.33 \pm 0.20 | 8.52 \pm 0.20 |
| H3-C3 ^b | 5.43 \pm 0.20 | 8.40 \pm 0.20 |
| H4-C4 | -1.88 \pm 0.20 | -7.10 \pm 0.20 |
| H5-C5 | 4.96 \pm 0.20 | 8.22 \pm 0.20 |
| H1'α-C1'α | -8.55 \pm 0.20 | -12.88 \pm 0.20 |
| H2'α-C2'α | 7.90 \pm 0.20 | 12.90 \pm 0.20 |
| H3'α-C3'α | 7.63 \pm 0.20 | 11.97 \pm 0.20 |
| H5'α-C5'α | 9.84 \pm 0.20 | 13.33 \pm 0.20 |
| H1'β-C1'β | 3.41 \pm 0.20 | 3.69 \pm 0.20 |
| H2'β-C2'β | 3.53 \pm 0.20 | 4.56 \pm 0.20 |
| H3'β-C3'β | 3.94 \pm 0.20 | 4.89 \pm 0.20 |
| H5'β-C5'β | 3.59 \pm 0.20 | 4.16 \pm 0.20 |
| H2-C1 ^b | | 3.29 \pm 0.23 |
| H4-C2 | 0.63 \pm 0.13 | 0.15 \pm 0.08 |
| H4-C3 | -0.86 \pm 0.31 | -0.65 \pm 0.29 |
| H4-C6 | 0.01 \pm 0.19 | 0.06 \pm 0.11 |
| H1' α -C3' α | -0.94 \pm 1.35 | -0.34 \pm 0.63 |
| H1' α -C5' α | -0.07 \pm 0.75 | 0.05 \pm 0.58 |
| H1' β -C2' β | -0.25 \pm 0.15 | 0.08 \pm 0.15 |
| H1' β -C3' β | -0.17 \pm 0.13 | -0.48 \pm 0.06 |
| H1' β -C5' β | -0.06 \pm 0.15 | -0.30 \pm 0.15 |
| H2' β -C1' β | -0.55 \pm 0.76 | -0.40 \pm 0.67 |
| H2' β -C3' β | -0.51 \pm 0.28 | 0.48 \pm 0.47 |
| H2' β -C4' β | | 0.04 \pm 0.25 |
| H3' β -C2' β ^b | 0.62 \pm 0.85 | -0.52 \pm 0.52 |
| H1-H2 | 0.93 \pm 0.30 | 1.92 \pm 0.79 |
| H2-H3 | 0.56 \pm 0.56 | 0.60 \pm 0.40 |
| H3-H4 | -0.11 \pm 0.34 | 0.21 \pm 0.21 |
| H1' α -H2' α | 2.39 \pm 0.11 | 2.29 \pm 0.12 |
| H2' α -H3' α | 0.18 \pm 0.11 | 0.55 \pm 0.18 |
| H3' α -H4' α ^b | 1.04 \pm 0.22 | 2.12 \pm 1.01 |
| H4' α -H5' α | 2.67 \pm 0.60 | 3.01 \pm 0.60 |
| H1' β -H2' β | 0.32 \pm 0.08 | 0.13 \pm 0.09 |
| H2' β -H3' β ^b | -0.01 \pm 0.57 | 0.13 \pm 0.65 |
| H4' β -H5' β ^b | 0.07 \pm 0.57 | 1.26 \pm 0.65 |
| H1-C4' α | 0.48 \pm 0.66 | 0.92 \pm 0.21 |
| H1-C4' β | -0.71 \pm 0.26 | -0.20 \pm 0.08 |

Table II. ϕ, ψ Coordinates for Conformational Energy Minima

| Anomer | CSFF | Accelrys |
|-------------------|------------|------------|
| α -lactose | 160°, 240° | 185°, 235° |
| | 150°, 65° | 150°, 190° |
| | 150°, 210° | 280°, 250° |
| | 290°, 245° | 185°, 85° |
| β -lactose | 160°, 240 | 185°, 235° |
| | 170°, 85 | 150°, 190° |
| | 140°, 210 | 280°, 250° |
| | 290°, 245 | 150°, 70° |

Note: listed in order of increasing relative energy.

Comparison with experimental results can be used to improve force fields. The conformational energies are drawn in Figure 3 as contour lines. Overall, the CSFF and Accelrys force fields agree on the features of the potential energy

We used the structures resulting from the two force fields in an exhaustive conformational search in fitting RDCs obtained from measurements in two liquid crystalline media, DMPC/DHPC bicelles (11) and DMPC/DHPC/CTAB bicelles (42) (Table I). The best fits for both anomers yield a range of acceptable conformations (Figure 3A – 3D). Fits for α -lactose do not coincide with the X-ray structure, $(\phi, \psi) = (120^\circ, 270^\circ)$, while those for β -lactose agree with the X-ray structure fitting best to a structure with $(\phi, \psi) = (180^\circ, 240^\circ)$. Nevertheless, both anomers have additional acceptable fits that are in regions indicated as minimum energy regions from CSFF and Accelrys force fields, which do not coincide with the X-ray structures. The intersection of force field calculations and RDC fits suggests that some of the anti- ϕ conformation is present in solution for the α -anomer, but not for the β -anomer.

Close examination of the RDCs for the glucose ring, highlighted in bold in Table I, reveals that RDCs for α -glucose differ from those of β -glucose by a factor of two to three, depending on the alignment medium. Thus, RDC values alone for the glucose ring imply that α - and β -lactose have different three-dimensional shapes.

To further scrutinize the results of the RDC fittings, we measured ^{13}C -edited ^1H - ^1H NOESY (HSQC-NOESY) spectra of lactose in natural abundance. These measurements are facilitated by the approximately 0.12 ppm chemical shift difference between C4' α and C4' β (C4 α -Glucose and C4 β -Glucose atoms, respectively). Well-resolved, cross-peaks between H1 (H1 α / β -Gal) and the H4' α -(α / β) peaks are observed in this two-dimensional, ^{13}C -edited experiment.

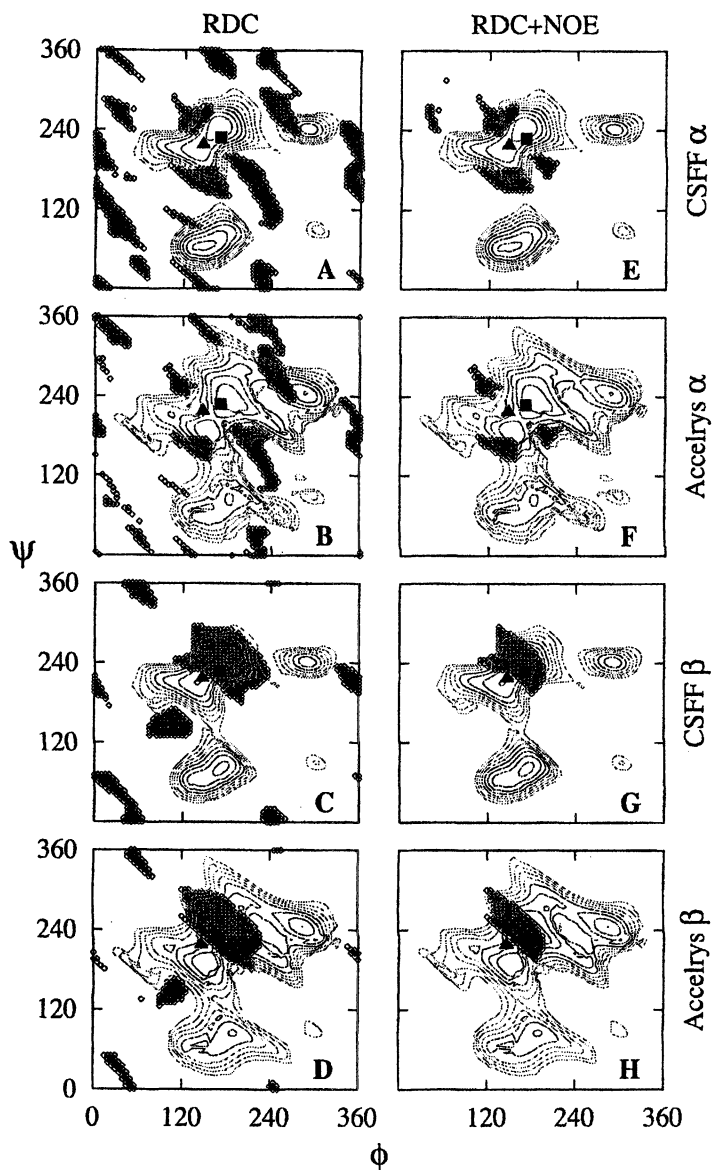


Figure 3. ϕ, ψ maps displaying energy contours and RDC fits (blue diamonds) of α - and β -lactose to structures calculated by CSFF and Accelrys (panels A-D). Panels E-H show similar maps with added NOE distance constraints. X-ray structures are indicated by a triangle for α -lactose and a square for β -lactose.

Table III. NMR Derived Internuclear Distances

| Atom Pair | Internuclear Distance (Å) | Reference |
|-----------------|---------------------------|----------------------------|
| H1-H4' α | 2.91 \pm 0.14 | H1' α -H2' α |
| H1-H4' β | 2.71 \pm 0.12 | H1' β -H3' β |
| H1-H4' β | 2.56 \pm 0.06 | H1' β -H5' β |

Note: Measured from NOE buildup curve. NOE mixing times used for the data range from 100 ms to 200 ms. Errors are at 95% confidence limits.

The intensity of these cross peaks was tracked as a function of NOE mixing time in the linear portion of the NOE buildup, to yield internuclear distances for the two anomers (Table III) (43). To obtain internuclear distances, the slope of the buildup curves for the peaks of interest (H1-H4' α/β) was divided by the slope of the buildup curve for a well-resolved peak whose H-H internuclear distance is known from *ab-initio* quantum mechanical calculations as a reference distance (44). Depending on the reference peak used two possible internuclear distances are calculated for H1-H4' β , which indicates that β -lactose tumbles anisotropically. The discrepancy arises from a difference between the angle made by internuclear vectors of H1' β -H3' β and H1' β -H5' β and the principal diffusion axis. Though these values are within experimental error of one another, the value determined from referencing to the H1' β -H5' β internuclear distance differs significantly from H1-H4' α . However, when referenced to the H1' β -H3' β distance the H1-H4' β internuclear distance is within experimental error to that of surfaces of H1-H4' α . The vacuum calculations (dielectric constant, $\epsilon = 5$) yield surfaces with energy minima for four values in ϕ, ψ space (Table II). These regions indicate that lactose may be interconverting amongst a few structures. The calculated potential energy surfaces for α -lactose are similar to those for β -lactose. This introduces an ambiguity that can only be properly dealt with if the overall molecular shape is known.

Intersection of the force field potential energy surfaces, RDC fits and NOEs (Figures 3E-3H) unambiguously indicates that the solution conformation of β -lactose is in the same region as the X-ray structure, $150^\circ \leq \phi \leq 210^\circ$ and $210^\circ \leq \psi \leq 270^\circ$. On the other hand, α -lactose deviates from the X-ray structure with $150^\circ \leq \phi \leq 180^\circ$ and $150^\circ \leq \psi \leq 180^\circ$.

Detection of Previously Unassigned Chemical Shifts in Lactose

The preceding results suggested that it might be possible to detect two sets of chemical shifts: one for the galactose ring attached α -glucose and another for

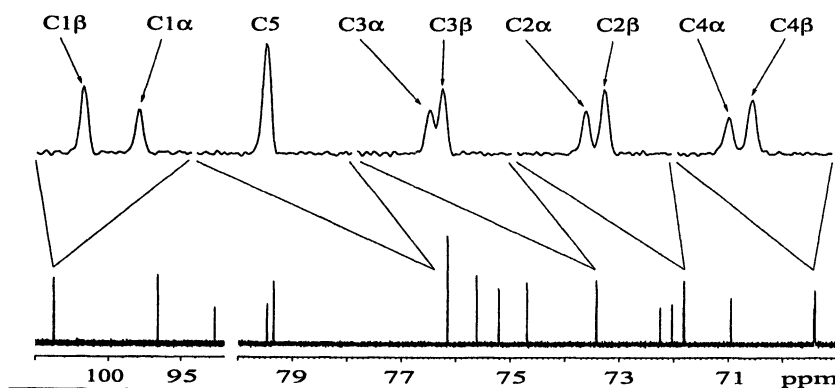


Figure 4. 125 MHz, one-dimensional ^{13}C spectra showing that peaks for the galactose ring carbon atoms in the two anomers can be differentiated.

galactose ring attached to β -glucose. To further explore this possibility, we measured high-resolution one-dimensional ^{13}C NMR spectra (Figure 4). These spectra show two sets of ^{13}C chemical shifts for every atom in the galactose ring, except for C5. One set for the β -galactose ring attached to α -glucose, and a second set for the β -galactose ring attached to β -glucose. Though C5 chemical shifts are not differentiated, the presence of two unresolved chemical shifts is suggested by the asymmetric lineshape of the C5 resonance and the relative peak intensity. Previous studies reported only one set of chemical shifts for the galactose carbon atoms. The peak intensity ratios of the newly detected peaks match the ratios of the α : β peaks in the glucose ring of lactose. We conclude that these shifts depend on whether the galactose ring is attached to α -glucose or β -glucose.

Implications of RDC Studies for Lactose Structure

RDCs for the glucose rings of α - and β -lactose are indeed different and imply that the two anomers have markedly distinct three-dimensional structures. RDC fits give a range of acceptable structures such as those seen in Figure 3 for the α -anomer. These conclusions are supported by chemical shift data showing two distinct sets of signals for the galactose ring. Even though they are ambiguous in showing the differences between α - and β -lactose, the NOE data complement the results of the RDC fits by significantly restricting conformational space for lactose. Alone, the NOE data can be misleading

because they show that the H1-H4' α and H1-H4' β internuclear distances (Table II) are comparable, which implies that their structures are likewise similar. However, the range of allowed structures deduced from NOE restraints alone is broad: $30^\circ \leq \phi \leq 210^\circ$, $150^\circ \leq \psi \leq 330^\circ$ for α -lactose and $60^\circ \leq \phi \leq 180^\circ$, $180^\circ \leq \psi \leq 300^\circ$ for β -lactose. These ranges are significantly reduced when overlap with low-energy regions calculated by molecular mechanics is considered. Yet, the possibility exists that while the distances are indistinguishable within experimental error, the torsional angles, ϕ and ψ are different and may yield detectably different three-dimensional structures: for example α -lactose structure can be defined by $(\phi, \psi) = (90^\circ, 210^\circ)$, while β -lactose is defined by $(\phi, \psi) = (150^\circ, 270^\circ)$. RDCs further aid in interpreting this data by reducing the likely conformations to $150^\circ \leq \phi \leq 180^\circ$, $150^\circ \leq \psi \leq 180^\circ$ for α -lactose and $150^\circ \leq \phi \leq 210^\circ$, $210^\circ \leq \psi \leq 270^\circ$ for β -lactose. The results presented here indicate that α -lactose has a different structure in solution than it does in the X-ray structure, while the structure of β -lactose is quite similar in solution and X-ray. The discrepancy for α -lactose may be due to crystal packing forces.

The above results signify that the shapes of α - and β -lactose are dissimilar (Figure 5). It is remarkable that a minor stereochemical difference such as α - vs β -lactose (axial or equatorial OH at the reducing ring) can result in significant remote changes in ϕ and ψ and ultimately in molecular shape. We are currently exploring possible explanations for this effect. One possibility is that an experimentally unobserved conformation is populated in α -lactose and not in β -lactose. This would imply that conformational interconversion is ongoing.



Figure 5. Likely structures of α - and β -lactose, $(\phi, \psi) = (150^\circ, 160^\circ)$ and $(170^\circ, 230^\circ)$, respectively from a combination of RDCs, NOEs and calculations. These structures highlight the different shapes of the two lactose anomers.

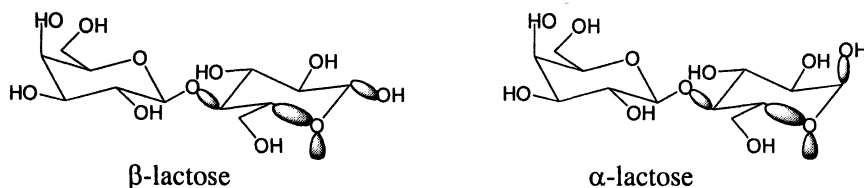


Figure 6. Differences in hyperconjugative interactions in β - and α -lactose

Since each molecular conformation has a different shape, the RDCs and NOEs for each of these shapes are expected to be different. If the conformational jumps were occurring faster than 10^{-6} s^{-1} , we would expect to observe averaging of RDCs, NOEs and chemical shifts for the unobserved, individual conformations. The results of the presence of more than one conformation on the NOEs would be a larger apparent H1-H4' internuclear distance. This effect would make it more difficult to clearly differentiate internuclear distances between α - and β -lactose. The effect of additional conformations on RDC-based analysis could also result in virtual conformations. We attempted to rule out virtual structures from a combination of force field calculations and NOEs.

It is also possible that the RDCs lead to underdetermined structures because many of the internuclear vectors are parallel. Thus, accurate conformations may be difficult to obtain. For small molecules, the mechanism of alignment, which leads to the observation of RDCs, is not well understood. Though it is not likely, the two anomers of lactose may interact differently with the orienting media. This would result in a conformational change compared to that in isotropic solution, and could explain the differences in alignment between α - and β -lactose. Nuclear relaxation measurements would be used to probe this potential difference in interactions.

Another possible source of shape differences between α - and β -lactose may be that their hydration spheres are different. An axial OH in the α - anomer may change the hydration sphere through changes in the water network relative to the equatorial OH in β -lactose. This difference in hydration may also result in a significant shape difference, and hence differences in molecular alignment. Understanding changes in hydration requires significant molecular modeling efforts in the presence of bound and surrounding water molecules.

Yet another possible source of disparate structures between α - and β -lactose is hyperconjugation (Figure 6). Both α - and β -lactose have the galactose ring attached to the glucose in a β 1-4 linkage, yet the torsional angles ϕ and ψ differ depending on whether the reducing ring is in the α or β form.

Extended orbital overlap is possible in β -lactose between O4-C4 Glc and C5-O5 Glc, and C5-O5 Glc and C1-O1 Glc. In α -lactose orbital overlap is still

present between O4-C4 Glc and C5-O5 Glc, but this network cannot be extended to C1-O1 Glc in the same fashion as in the β -anomer. In the α -anomer orbital overlap of the axial p-type lone-pair on O5 with the anti-periplanar axial C1-OH bond is much more prevalent than in the β -anomer. The donation of electron density from the O5 lone-pair into the axial C1-OH bond in the α -anomer would inductively withdraw electron density from the O5 making the overlap between O4-C4 and O5-C5 different from the overlap in the β -anomer. These overlap differences may cause the observed differences in three-dimensional structure of the two anomers. This possibility could be tested with ab-initio quantum mechanical calculations.

The RDC fits are complicated by the fact that the reported RDCs for the galactose ring in lactose are superpositions for β -galactose attached to α - and β -glucose. Measurement of $^1D_{CH}$ at high resolution in two-dimensional HSQC spectra is not trivial, and is complicated by long-range 1H - ^{13}C couplings, which are usually not resolved in a typical HSQC experiment. Measurement of long-range 1H - ^{13}C couplings is yet more challenging. Finally, three-dimensional NMR will be needed to measure 1H - 1H couplings.

Concluding Remarks

The reported data presented in this chapter demonstrate that the two lactose anomers have distinct structures. This conclusion is based on a combination of new NMR RDC measurements, interresidue NOEs and molecular mechanics calculations. The first indication that this is the case was observed from differences between RDCs for the two anomers of the glucose ring. One method we have successfully used to find the likeliest conformations in carbohydrates is to overlap the acceptably fit regions from RDCs with those from molecular mechanics calculations (21,28). This technique appears to be helpful in lactose as well. According to these fits, the structural differences manifest themselves in terms of different ϕ, ψ torsion angles and hence different shapes. Chemical shift differences in the galactose ring are further evidence that structural differences exist between the structures of α - and β -lactose.

The results for lactose appear to be true for cellobiose and maltose as well, and imply that α - anomers of these disaccharides have different structures from the β -anomers (45). These conclusions were first supported by RDC measurements, which are expected to be powerful structural aids in delineating carbohydrate structure in the future.

References

1. Dwek, R. A., *Chem. Rev.* **1996**, *96*, 683-720.
2. Bertozzi, C. R.; Kiessling, L. L., *Science* **2001**, *291*, 2357-2364.
3. Barchi Jr., J. J., *Curr. Pharm. Design* **2000**, *6*, 485-501.
4. Williams, S. J.; Davies, G. J., *Trends Biotechnol.* **2001**, *19*, 356-362.
5. Ritter, T. K.; Wong, C. H., *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 3509-3533.
6. Allen, J. R.; Harris, C. R.; Danishefsky, S. J., *J. Am. Chem. Soc.* **2001**, *123*, 1890-1897.
7. Vankar, Y. D.; Schmidt, R. R., *Chem. Soc. Rev.* **2000**, *29*, 201-216.
8. Nitz, M.; Ling, C.-C.; Otter, A.; Cutler, J. E.; Bundle, D. R., *J. Biol. Chem.* **2002**, *277*, 3440-3446.
9. Lisicki, M. A.; Mishra, P. K.; Bothner-By, A. A.; Lindsey, J. S., *J. Phys. Chem.* **1988**, *92*, 3400-3403.
10. Tjandra, N.; Bax, A., *Science* **1997**, *278*, 1111-1114.
11. Sanders, C. R.; Schwonek, J. P., *Biochemistry* **1992**, *31*, 8898-8905.
12. Losonczi, J. A.; Prestegard, J. H., *J. Biomol. NMR* **1998**, *12*, 447-451.
13. Fleming, K.; Gray, D.; Prasannan, S.; Matthews, S., *J. Am. Chem. Soc.* **2000**, *122*, 5524-5525.
14. Prosser, R. S.; Losonczi, J. A.; Shiyonovskaya, I. V., *J. Am. Chem. Soc.* **1998**, *120*, 11010-11011.
15. Barrientos, L. G.; Dolan, C.; Gronenborn, A. M., *Journal of Biomolecular NMR* **2000**, *16*, 329-337.
16. Ruckert, M.; Otting, G., *J. Am. Chem. Soc.* **2000**, *122*, 7793-7797.
17. Sass, J.; Cordier, F.; Hoffmann, A.; Rogowski, M.; Cousin, A.; Omichinski, J. G.; Lowen, H.; Grzesiek, S., *J. Am. Chem. Soc.* **1999**, *121*, 2047-2055.
18. Desvaux, H.; Gabriel, J. P.; Berthault, P.; Camerel, F., *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 373-376.
19. Sass, H. J.; Musco, G.; Stahl, S. J.; Wingfield, P. T.; Grzesiek, S., *Journal of Biomolecular NMR* **2000**, *18*, 303-309.
20. Tycko, R.; Blanco, F. J.; Ishii, Y., *J. Am. Chem. Soc.* **2000**, *122*, 9340-9341.
21. Freedberg, D. I., *J. Am. Chem. Soc.* **2002**, *124*, 2358-2362.
22. Homans, S. W., *Biochem. Soc. Trans.* **1998**, *26*, 551-560.
23. Manuel-Pastor, M.; Bush, C. A., *Carbohydr. Res.* **1999**, *323*, 147-155.
24. Rundlöf, T.; Landersjö, C.; Lycknert, K.; Maliniak, A.; Widmalm, G., *Magn. Reson. Chem.* **1998**, *36*, 773-776.
25. Skrynnikov, N. R.; Goto, N. K.; Yang, D.; Choy, W.-Y.; Tolman, J. R.; Mueller, G. A.; Kay, L. E., *J. Mol. Biol.* **2000**, *295*, 1265-1273.
26. Tian, F.; Al-Hashimi, H. M.; Craighead, J. L.; Prestegard, J. H., *J. Am. Chem. Soc.* **2001**, *123*, 485-492.

27. Wang, H.; Eberstadt, M.; Olejniczak, E. T.; Meadows, R. P.; Fesik, S. W., *J. Biomol. NMR* **1998**, *12*, 443-446.
28. Venable, R.; Delaglio, F.; Norris, S. E.; Freedberg, D. I., *Carbohydr. Res.* **2005**, *340*, 863-874.
29. Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H., *J. Magn. Reson.* **1999**, *138*, 334-342.
30. Saupe, A., *Angew. Chem., Int. Ed. Engl.* **1968**, *7*, 97-112.
31. Oguchi, H.; Toyokuni, T.; Dean, B.; Ito, H.; Otsuji, E.; Jones, V. L.; Sadozai, K. K.; Hoakomori, S., *Cancer Commun.* **1990**, *2*, 311-316.
32. Zhu, X. X.; Ding, P. Y.; Cai, M. S., *Carbohydr. Res.* **1996**, *296*, 229-235.
33. Cook, W. J.; Bugg, C. E., *Acta Crystallographica, Section B* **1973**, *29*, 907.
34. Bugg, C. E., *J. Am. Chem. Soc.* **1973**, *95*, 908-913.
35. Fries, D. C.; Rao, S. T.; Sundaralingam, M., *Acta Crystallographica, Section B* **1971**, *B27*, 994-1005.
36. Hirotsu, K.; Shimada, A., *Bull. Chem. Soc. Jpn.* **1974**, *47*, 1872-1879.
37. Nunez, H. A.; Barker, R., *Biochemistry* **1980**, *19*, 489-495.
38. Duda, C. A.; Stevens, E. S., *Carb. Res.* **1990**, *206*, 347-351.
39. Rees, D. A., *J. Chem. Soc. B* **1970**, 877-884.
40. Kuttel, M.; Brady, J. W.; Naidoo, K. J., *J. Comput. Chem.* **2002**, *23*, 1236-1243.
41. Momany, F. A.; Rone, R., *J. Comput. Chem.* **1992**, *13*, 888-900.
42. Ramirez, B. E.; Bax, A., *J. Am. Chem. Soc.* **1998**, *120*, 9106-9107.
43. Neuhaus, D.; Williamson, M., *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. VCH: New York, 1989.
44. Negin, R. S.; Venable, R.; Woodcock, L.; Freedberg, D. I., **2005**, manuscript in preparation.
45. Norris, S. E.; Venable, R.; Ano, S. O.; Freedberg, D. I., **2005**, manuscript in preparation.

Chapter 13

Incorporation of Carbohydrates into Macromolecular Force Fields: Development and Validation

Sarah M. Tschampel¹, Karl N. Kirschner², and Robert J. Woods¹

¹Complex Carbohydrate Research Center, 315 Riverbend Road,
Athens, GA 30602

²Chemistry Department, Hamilton College, 198 College Hill Road,
Clinton, NY 13323

Their inherent flexibility, combined with the multitude of linkages and functionalities possible, provides a fundamental challenge in the development of carbohydrate force fields for modeling oligosaccharides. Validation of carbohydrate force fields has been hampered by both the paucity of structural data for all but the simplest glycans, but the development has rapidly progressed due to advancements in *ab initio* methods. Condensed-phase properties of monosaccharides have been accurately modeled with current methodology, but in order to determine the ability of these carbohydrate force fields to accurately describe oligosaccharides, a thorough systematic analysis remains to be performed.

I. Introduction

i. Necessity of carbohydrate force fields

Carbohydrates are involved in many biological recognition processes, including immune response to bacterial pathogens, adhesion of bacteria and virus particles to host cells, in the hyperacute rejection of tissue transplants from nonhuman sources (xenograft transplantation) and in inflammatory response (1-6). Further, the covalent attachment of carbohydrates (glycosylation) is one of the most prevalent post-translational modifications of proteins. The heightened awareness of carbohydrate function has increased the desire and need to understand their three-dimensional structures. Traditionally, experimental techniques, such as X-ray diffraction and nuclear magnetic resonance (NMR) spectroscopy, have been used to study the conformations of carbohydrates and their interactions with proteins. However, these techniques yield static, or average, pictures of inherently dynamic systems. Computational chemistry has proven to be a valuable tool in studying the dynamics of protein and nucleic acid systems, and it has already been shown to be successful in studying systems involving carbohydrates (7-9). Molecular modeling tools enable the exploration of sequence-structure relationships in oligosaccharides, which are experimentally difficult to establish due to the challenges associated with generating modified carbohydrate sequences in sufficient quantity and purity. In addition, computational chemistry can provide accurate models for the interpretation of experimental data that are only indirectly related to structure, such as, spectroscopic data, but also from biological and immunological studies. As the interest in glycobiology grows, so grows the demand for improved theoretical methodologies.

Quantum mechanical methods are powerful tools for examining the conformational stabilities of molecules in the gas phase. The applicability of these *ab initio* methods to systems under biologically relevant conditions (*i.e.* in water) is limited to relatively small systems, whereas methods based on Newtonian physics, such as molecular mechanics and dynamics, alleviate this problem. Just as the accuracy of the *ab initio* method depends heavily on the choice of basis set, so, the accuracy of the classical method depends on the rigorous derivation of force field parameters. With a properly tuned force field one can model large carbohydrate and glycoprotein systems with a remarkably high degree of confidence in the obtained structures. Protein force fields have been refined for several decades with many research groups contributing to their development. The growing interest in structural glycobiology has led to the need to develop protein-consistent carbohydrate force fields. The application of

existing classical methods, such as molecular dynamics (MD), in combination with the more recently developed protein-consistent carbohydrate force fields can potentially become a key structural tool in the field of glycobiology. In section I, a brief summary of the challenges associated with carbohydrate simulation, with a focus on the development of appropriate force fields is presented. In the following section a description of specific implementations is given. Lastly, some general conclusions regarding the current status and future direction of carbohydrate force field development are reached.

ii. Challenges in carbohydrate force field development

Development of a carbohydrate force field is a challenging endeavor due to the inherent flexibility and lack of regular tertiary structure in oligosaccharides. Minimally, a carbohydrate force field must address the stability and structural properties of glycosyl rings, rotational properties of exocyclic bonds (*i.e.* the C5-C6 ω -angle in pyranoses), and the conformational properties associated with the O-C-O moiety (*i.e.* the anomeric and exo-anomeric effects) (10).

Amino acids polymerize in a single, linear fashion to form polypeptides. In contrast, carbohydrates polymerize by linking through any of several different hydroxyl positions to form both linear and branched structures, Figure 1. While only one peptide can be formed from two alanyl residues, up to eleven chemically unique disaccharides can arise from combining two glucopyranosyl residues through permutations of linkage position and anomeric configuration. Each of these disaccharides may adopt distinct conformational families about their linkage depending on the shape of the local potential energy surface.

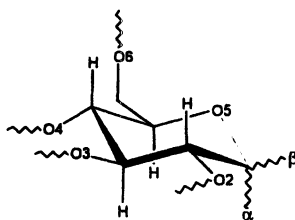


Figure 1. Possible linkage positions (2,3,4,6) and anomeric configurations (α , β) involved in a glucopyranosyl (Glc_p) residue in an oligosaccharide.

Due to the high number of hydroxyl groups, carbohydrates have the ability to form inter-residue and solute-solvent hydrogen bonds. Water can compete with internal hydrogen bonds, attenuating the extent of inter-residue interactions and increasing molecular flexibility (11-13). For these reasons, carbohydrates present a unique challenge in the development of a robust force field for modeling the physics of oligosaccharides.

In addition to purely carbohydrate systems, the ability to integrate the carbohydrate force field with a protein force field allows the former to be of broader scientific use (14). Typical quadratic biomolecular force fields, such as OPLS, AMBER, and CHARMM, include contributions from bond stretching (r), angle bending (θ), and non-bonded interactions (Coulomb and Lennard-Jones terms) (14-18). In order to account for non-classical behavior, such as the hyperconjugation involved in the rotation around the O-C-O-C torsion angle in carbohydrates, the addition of a torsion term is necessary in the potential energy, V_{total} , equation 1. It should be noted that the inclusion of torsion terms can be considered a correction that partially compensates for the deficiencies in the classical model, such as the omission of molecular polarizability and other quantum effects. Derivation of parameters that influence the overall three-dimensional structure of oligosaccharides, such as the partial atomic charges and torsion terms associated with the glycosidic linkages, is frequently the most crucial task in carbohydrate force field parameterization.

$$V_{total} = \frac{1}{2} \sum_{bonds} K_r (r - r_0)^2 + \frac{1}{2} \sum_{angle} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} \sum_n \frac{V_n}{2} [1 + \cos(n\phi - \gamma_n)] + \sum_{\substack{nonbonded \\ i < j}} S_{vdw} \left(\frac{B_{ij}}{R_{ij}^{12}} - \frac{A_{ij}}{R_{ij}^6} \right) + S_q \left(\frac{q_i q_j}{\epsilon r_{ij}} \right), \quad (1)$$

where $A_{ij} = e_{ij} R_{ij}^{12}$, $B_{ij} = 2e_{ij} R_{ij}^6$, $e_{ij} = \sqrt{e_i e_j}$, and $R_{ij} = R_i + R_j$

In order to achieve compatibility between carbohydrate and protein force fields, both models need to be derived for use with the same potential energy function, the same solvent model, and must employ compatible partial atomic charges (q). In short, mixing parameters from different force fields or from different derivation methods can lead to a critical loss of internal consistency.

In some cases, protein force fields employ a scaling factor to dampen 1,4 van der Waals (S_{vdw}) and electrostatic interactions (S_q) (15,16). This treatment is problematic when deriving carbohydrate parameters and is exemplified by looking at the ω -angle in α -D-Glcp-OMe, Figure 2. In the *gg* rotamer, the distances between O6...O4 and O6...O5 are essentially equivalent. However, the former is a 1,5 interaction, while the latter is a 1,4 and is artificially weakened by the application of the 1,4 scaling factor. Removal of the scaling factors or introduction of additional scaling factors has been shown to be necessary to obtain the best fit to rotational properties for this linkage, but may result in loss of compatibility with the protein force field (19-20).

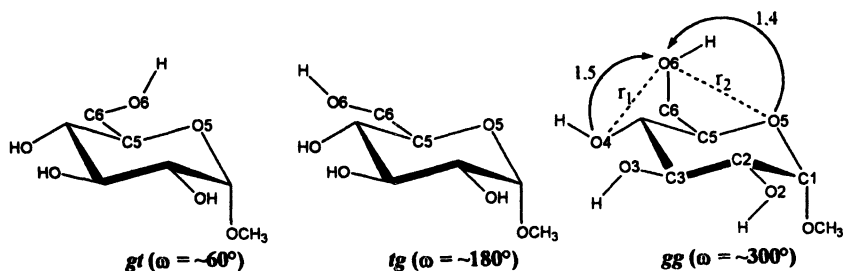


Figure 2. The ω -angle rotamers in α -D-Glcp-OMe, in which $r_1 \approx r_2$.

iii. Validation of carbohydrate force fields

In developing and validating force fields one of the primary considerations is to decide if the force field should reproduce gas-phase data, condensed-phase data, or both. Furthermore, one should decide if the force field should be developed using experimental data, theoretical data, or both. The quadratic force fields discussed here generally utilized a combination of experimental and theoretical data for parameter development, except for the torsion parameters, which were derived almost exclusively from *ab initio* gas-phase data. In some cases it is necessary to further modify the parameters to reproduce condensed-phase data. The decision as to how to appropriately modify the parameters is part of the “art” in parameterization. A recent example of this is the series of carbohydrate force fields by Brady and coworkers for use in CHARMM that were initially based on gas-phase *ab initio* calculations, but later modified to reproduce solution-phase NMR data (17).

Hemmingsen *et al.* recently performed a comparison of 20 modern carbohydrate force fields examining their ability to reproduce molecular properties in the gas phase (18). Quantum mechanical optimizations were performed on selected conformations of 5-deoxy- β -D-Xylf-OMe, α,β -D-Glcp-OMe, α,β -D-Galp-OMe, and a hydrogen-bonded complex between a water molecule and α -D-Glcp-OMe. From this study, it was concluded that the force fields evaluated, which included only the GLYCAM2000a, AMBER*, and the general CHARMM22 and AMBER (PARM94) parameters sets of those discussed here, have three major problems: 1) the inability to reproduce the gas-phase interaction energy between water and α -D-Glcp-OMe; 2) the inability to reproduce the relative stability of ω -angle conformers; and 3) the inability to predict the correct ranking of anomeric stabilities. However, a more relevant, if pragmatic, method of validation is to examine how the force fields reproduce condensed-phase properties (*i.e.* binding free energies, solution conformational properties, etc.).

Performing solvated MD simulations on carbohydrates and oligosaccharides allows the comparison of a force field's performance with respect to condensed-phase experimental data, such as NMR scalar J -couplings, NOE intensities, and residual dipolar couplings. Evaluation of force fields in this manner is more demanding and less frequently reported, but provides a potentially powerful method when coupled with gas-phase validation for determining the strengths and weakness of a given force field, as well as gaining insight into the physics behind the role that solvent plays in mediating carbohydrate conformations (11,19,20).

The ability to increase the timescale of MD simulations and to employ free-energy based methods, such as potential of mean force (PMF) calculations, has allowed a more thorough evaluation of condensed-phase conformational properties to be made (21-23). The ability to reproduce the experimental solution-phase rotamer populations for glycosidic linkages as well as the ω -angle provides an indication of the suitability of the force field for solution-phase MD simulations(17). Rotamers around the C5-C6 bond are classified into three states: *trans-gauche* (*tg*), *gauche-gauche* (*gg*), and *gauche-trans* (*gt*), based on the conformation of the O5-C5-C6-O6 and C4-C5-C6-O6 dihedral angles, respectively, and are illustrated in Figure 2. These populations have been determined using MD simulations, with a variety of force fields, and by NMR spectroscopy, with the resulting population ratios in qualitative agreement, as shown in Table I. Accurate comparisons require the simulations to be statistically converged; a simulation time with the magnitude of ≥ 100 ns is often necessary when employing explicit solvation, Figure 3. The rate of convergence appears to be different for each saccharide, being dependent on the barrier heights between the populated rotamers. Statistical convergence of simulations is a difficult task to achieve particularly for rotamers with long lifetimes, but is an important consideration when evaluating a force field and interpreting MD results for flexible molecules.

Extension of these condensed-phase methods to look at the population ratios for the ϕ - and ψ - angles, as well as the ω -angle in 1 \rightarrow 6 linkages, has been useful in determining the orientation of the two monosaccharides relative to each other (19,24). The difference in conformational properties in the gas phase versus solution phase is exemplified for α -D-Xylp(1 \rightarrow 4)- α -D-Xylp in Figure 4 (19). The global minimum is B in the gas phase, with C and A 0.10 and 0.89 kcal \cdot mol $^{-1}$ higher, respectively. In contrast, the global minimum is S in the condensed phase with the relative stability of conformers A and C reversed. Not surprisingly, inter-residue hydrogen bonding can be a dominant force in the gas phase as well as non-polar solvents, but not in polar solvents such as water (11,25).

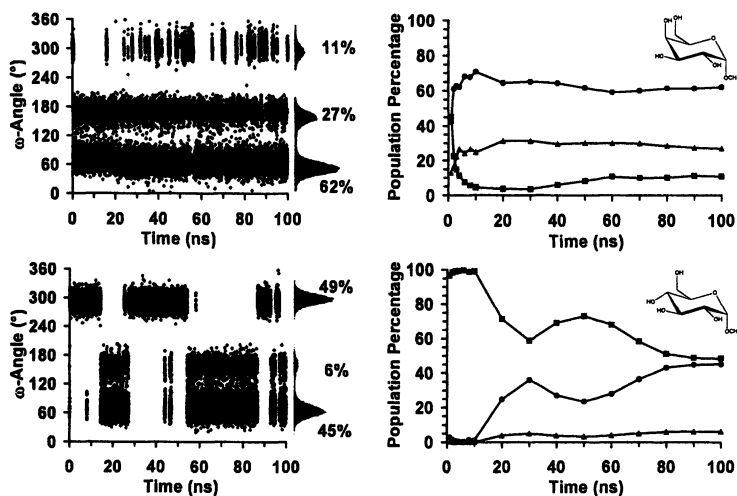


Figure 3. Population and convergence for the ω -angle conformers (*gt*(●), *tg*(▲), and *gg*(■)) in α -D-Galp-OMe (top) and α -D-Glcp-OMe (bottom) from explicitly solvated MD simulations using the GLYCAM04 parameter set(26).

Table I. Reproduction of the ω -angle populations for D-Glcp and D-Galp.

| Force Field | Method | D-Glcp | | | D-Galp | | |
|------------------------------------|-------------|-----------|-----------|-----------|-----------|-----------|-----------|
| | | <i>tg</i> | <i>gt</i> | <i>gg</i> | <i>tg</i> | <i>gt</i> | <i>gg</i> |
| Exp. | (27,28) | 7 | 41 | 52 | 32 | 56 | 12 |
| | (29,30) | 2 | 45 | 53 | 21 | 61 | 18 |
| PHLB (31) | PMF | 0 | 15 | 85 | 8 | 91 | 19 |
| | MD (10 ns) | - | - | - | 8 | 91 | 1 |
| CSFF (31) | PMF | 1 | 33 | 66 | 21 | 75 | 4 |
| | MD (10 ns) | 0 | 31 | 69 | 14 | 80 | 6 |
| GLYCAM2000a ^b (11,32) | MD (50 ns) | 6 | 40 | 54 | 28 | 64 | 8 |
| | PMF | 8 | 44 | 48 | 38 | 57 | 5 |
| GLYCAM04 ^b (26) | MD (100 ns) | 6 | 45 | 49 | 27 | 62 | 11 |
| GLYCAM2000a-LP ^{a,b} (32) | PMF | 5 | 49 | 46 | 28 | 65 | 7 |
| OPLS-AA (33) | MD (10 ns) | 4 | 51 | 45 | 46 | 52 | 2 |
| | MD (10 ns) | 11 | 77 | 12 | 33 | 64 | 3 |
| OPLS-AA-SEI ^c (33) | MD (10 ns) | 4 | 29 | 67 | 38 | 53 | 9 |
| | MD (10 ns) | 4 | 27 | 69 | 61 | 35 | 4 |

^a Explicitly solvated with TIP5P, while all others utilized the TIP3P model (34,35).

^b The methyl glycoside was utilized.

^c Simulations initiated with different starting conformations of the ω -angle.

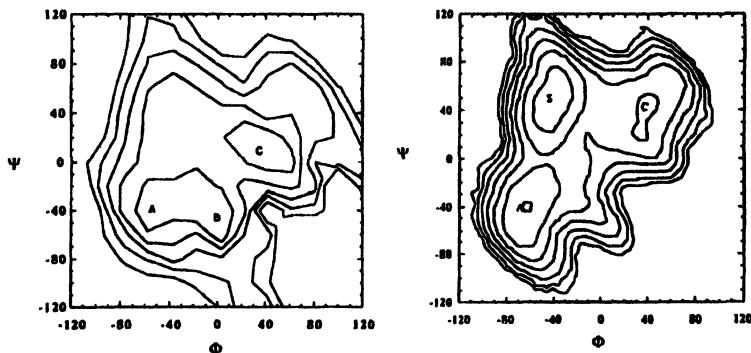


Figure 4. ϕ, ψ plots of the free energy for α -D-Xylp(1 \rightarrow 4)- α -D-Xylp in vacuum (left) and aqueous solution (right) computed with the HGFB parameter set as implemented in CHARMM (36-38). Reprinted with permission from Naidoo *et al.* (19). Copyright 1999 American Chemical Society.

Corzana *et al.* examined the condensed-phase conformational properties of methyl α -maltoside (α -D-Glcp(1 \rightarrow 4)- α -D-Glcp-OMe) and methyl α -isomaltoside (α -D-Glcp(1 \rightarrow 6)- α -D-Glcp-OMe) employing explicitly solvated 10 ns MD simulations with a variety of carbohydrate force fields (39). Generally, the force fields performed well with respect to conformational properties of the 1 \rightarrow 4 linked glycan, but extension to the more flexible 1 \rightarrow 6 linkage and examination of solute-solvent properties, indicated large variations in the results between different parameter sets.

II. Carbohydrate parameter sets implemented into the OPLS, CHARMM, and AMBER biomolecular force fields

i. Optimized Potentials for Liquid Simulations (OPLS (15,40))

Unlike the original OPLS force field, the successor OPLS-AA no longer merges the aliphatic hydrogen atoms with their attached carbons in a united-atom fashion (40). The majority of bond and angle terms were taken from the AMBER force field, while new torsion and non-bonded terms were derived (37). The torsion and non-bonded parameters were derived simultaneously to reproduce conformational energy profiles obtained at the HF/6-31G(d) level of

theory for over 50 organic molecules and ions as well as enthalpies of vaporization and densities for 34 pure organic liquids, respectively (41). The partial atomic charges were derived from pure liquid simulations of compounds containing the relevant functional groups. The group-based charges become transferable between molecules by requiring that each group has a net charge of zero. Analogous with early AMBER force fields, OPLS-AA employs a scaling factor of 1/2.0 for 1,4 Coulomb and Lennard-Jones interactions. In 2001, new Fourier coefficients were derived for the backbone and side chain torsion angles in the protein force field and small modifications were made to the non-bonded parameters in order to achieve agreement with data obtained at the LMP2/cc-pVTZ(-f)//HF/6-31G(d,p) level of theory (42-44). The resulting force field, termed OPLS-AA/L, achieved an overall better fit to a representative set of conformers for all the amino acids, with an average root mean square (RMS) deviation of 0.47 kcal·mol⁻¹, compared to 0.81 kcal·mol⁻¹ for OPLS-AA.

Extension of OPLS-AA to carbohydrates by Damm et al. (45)

To extend the OPLS-AA force field to carbohydrates, 144 conformers of α,β D-Glcp, D-Galp, D-Manp, D-Glcp-OMe, and D-Manp-OMe were optimized at the HF/6-31G(d) level of theory. Utilization of model compounds, such as 2-hydroxytetrahydropyran, to derive relevant carbohydrate torsion parameters did not lead to optimal results, which may be due to the use of a set of low energy conformers as opposed to the complete rotational profile for the ϕ -angle in each model compound. Of the 144 conformers obtained, the energies of those within 3 kcal·mol⁻¹ at the HF/6-31G(d) level were recomputed at the B3LYP/6-311+G(d,p) level and then utilized to derive the torsion parameters (46-48). All torsion parameters were derived simultaneously with the Fitpar fitting program (49). Since only local minima were utilized, no information was obtained regarding the barrier height between these minima. As a result, the transition frequency between accessible conformations in solution may be either too frequent or infrequent. A single new atom type was introduced for the anomeric carbon, which was used for both α and β anomers. Non-bonded parameters were derived in standard OPLS fashion, from liquid simulations of smaller, neutral molecular units and pieced together to form the desired molecule. The relevant non-bonded parameters were taken from alcohols, ethers, diols, and alkanes. Overall, the optimized parameters gave an RMS deviation of 0.75 kcal·mol⁻¹ for the relative energies of the 44 minima utilized during the derivation of the torsion parameters.

Application of the 1,4 scaling term in the standard OPLS-AA force field to the carbohydrate force field was found to be problematic. Correction of the resulting inequality created between 1,4 and 1,5 interactions associated with O6...O5 and O6...O4 interactions discussed above, could not be fully achieved solely through optimization of the relevant torsion parameters. The populations of the ω -angle conformers generated by OPLS-AA demonstrated that the *tg* conformer is underestimated by about 20% during two explicitly solvated 10 ns MD simulations of α -D-Galp, however no final assessment can be made for α -

D-Glcp due to the large variation observed between the populations for two MD simulations initiated from two different ω -angle orientations, Table I (33).

Scaling Electrostatic Interactions (OPLS-AA-SEI (33))

The problem arising from the use of a 1,4 scaling term in simulations of carbohydrates with the general OPLS-AA force field was alleviated by introduction of additional 1,5 and 1,6 scaling factors. The 1,5 electrostatic interactions between the hydroxyl oxygen and vicinal hydroxyl hydrogen groups were scaled by 1/1.26, while the 1,6 interactions for $O6\cdots HO4$ and $HO6\cdots O4$ were scaled 1/1.22, Figure 5. After the alteration of scaling factors, the relevant carbohydrate torsion terms were refit. Three separate $C_{sp^3}-C_{sp^3}-C_{sp^3}-O_{sp^3}$ terms were introduced to differentiate between the C-C-C-OH found throughout the pyranosyl ring, the C-C-C-O5, and the C3-C2-C1-O1 sequences, which necessitated the introduction of a unique atom type for the O1 oxygen atom.

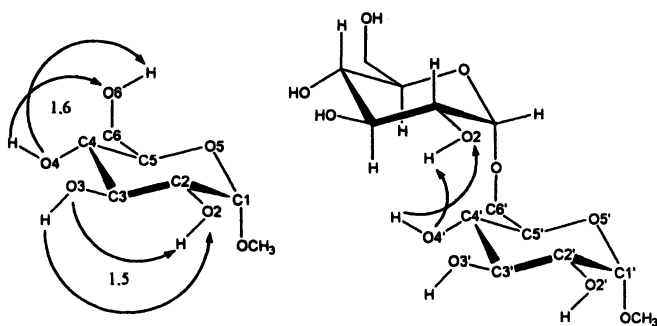


Figure 5. Illustration of the atoms involved in the 1,5 and 1,6 (left) and extended (right) scaling factors introduced in the OPLS-AA-SEI force field (33).

Owing to the presence of 1,6 scaling in the OPLS-AA-SEI force field, extension of these parameters to oligosaccharides containing 1,6 linkages is not straightforward. Therefore, the parameters relevant for the $O5-C5-C6-O6$ type torsion angle were reoptimized to reproduce the ω -angle potential energy profile for 6-O-(1'-methoxy-ethyl)- β -D-Glcp, which was obtained at the HF/6-31G(d) level, as well as the experimental ω -angle populations for β -gentiobiose (6-O- β -D-Glcp(1 \rightarrow 6)- β -D-Glcp) (34,50). An additional, extended scaling term between the C2 hydroxyl group and C4' hydroxyl group was needed to yield an adequate fit to both data sets (34,50). A value of 1/1.0 for the 1,6 and extended scaling

terms yielded an ω -angle population ratio for β -gentiobiose of 36:62:2, which is in agreement with the experimental data of 34:66:0, both corresponding to the *gg:gt:tg* ratio (51). Application to β -laminarabiose (D-Glcp(1 \rightarrow 3)- β -D-Glcp) revealed that no new parameters or scaling factors needed to be derived, only that the original OPLS-AA-SEI 1,5 and 1,6 scaling factors must be extended to all corresponding interactions in the molecule.

ii. Chemistry at HARvard Macromolecular Mechanics (CHARMM (17,18))

The focal point of development for the CHARMM parameter sets was to achieve the correct balance between solute-solute, solvent-solvent, and solute-solvent interactions, in which the solvent model employed was TIP3P (34). The Lennard-Jones parameters for the solute were derived via fitting to interaction energies and thermodynamic parameters obtained from experimental data and *ab initio* computations. A 1,4 Lennard-Jones scaling term was only found to be necessary to achieve the best fit in specific cases, such as between the peptide oxygen and nitrogen. Based on the formation of groups, which contained approximately 5 atoms, the charge of the whole group was set to zero and the individual partial charges were optimized to reproduce interaction energies and thermodynamic data from the aforementioned data. Similarly to the partial charge derivation in OPLS, this method has the advantage that functional groups can easily be added, *i.e.* to form derivatized amino acids that were not originally developed. There are no specific hydrogen bond parameters implemented.

Urey-Bradley terms were added to the potential energy function to improve the fit to the vibrational data for in-plane deformations and separation of symmetric and asymmetric bond stretching modes. Data from X-ray crystallography, gas-phase infrared and Raman spectroscopy, and HF/6-31G(d) *ab initio* computations were utilized to determine the Urey-Bradley terms, equation 2, where K_{UB} is the Urey-Bradley force constant and S is the distance between two covalent bonds (1,3 distance). The CHARMM force field was developed explicitly for proteins, nucleic acids, and common functional groups for use in a variety of molecular modeling applications.

$$U_{UB} = \sum K_{UB} (S - S_0)^2 \quad (2)$$

Extension of CHARMM to carbohydrates by Ha, Giammona, Field, and Brady (HGFB (37))

HGFB is an all-atom model for MD simulations of carbohydrates, which utilized the CHARMM methodology to determine partial charges. All parameters were optimized via fitting to experimental infrared and Raman

vibrational spectra as well as neutron-diffraction structural data for α -D-Glcp (52-54). Generalized torsions of the form X-A-B-X were utilized and there were no explicit hydrogen bonding terms. The initial study focused on applications in the gas phase and subsequently, crystal simulations were performed for both α -D-Glcp and its monohydrate, in which crystallographic symmetry was not imposed (55). Performing crystalline-phase MD simulations yielded a low RMS deviation (0.36 Å for both) to the experimental solid-phase structure (52,56). In contrast, explicitly solvated PMF simulations were performed to examine the ω -angle rotamer populations of β -D-Glcp. This analysis illustrated the inability of this force field to yield an experimentally consistent rotamer distribution, in which the experimentally least populated rotamer, *tg*, was computed to be present for approximately half of the simulation (17).

Extension of CHARMM to carbohydrates by Palma, Liang, and Brady (PHLB (57))

The PHLB force field was developed as an improvement over the HGFB force field, with a specific goal of reducing the frequency of hydroxyl rotations in solution. As in HGFB, experimental vibrational and structural properties were utilized to develop the bond, angle, and torsion parameters, but, in addition the MP2/6-31G(d) rotational energy profile around the C-C bond in ethylene glycol was included to revise the O-C-C-O torsion parameters. The van der Waals parameters and partial atomic charges (equivalent charges utilized for the α and β anomers) were taken from CHARMM and explicit torsion terms were implemented as an improvement over the generalized terms utilized in the HGFB parameter set. Overall, the resultant effect was a stiffening in the rotational freedom of the secondary hydroxyl groups. Explicitly solvated PMF calculations of α -D-Glcp indicated that the population of the *gg* conformer was overestimated at the expense of both other conformers, but is still an improvement over the HGFB results. The PHLB results for β -D-Galp were closer to experiment, with only a slight underestimation in the population of the *tg* rotamer.

Carbohydrate Solution Force Field (CSFF (31))

In 2002, the PHLB force field was empirically modified to reproduce the experimental ω -angle rotamer populations, resulting in the CSFF force field.

The computed populations were determined from explicitly solvated PMF calculations on α -D-Glcp and α -D-Galp. Based on these results, the O5-C5-C6-O6, O5-C5-C6-H6, and C4-C5-C6-O6 dihedral terms were reparameterized with introduction of a new atom type for C6 to make this term specific for the rotation of the hydroxymethyl group. In addition, the H-O-C-H dihedral force constants were reduced by 40% to increase the frequency of rotations of the secondary hydroxyl groups. All other parameters were left unadjusted from the PHLB force field, including the partial charges, shown in Table II. From both the explicitly solvated 10 ns MD simulations, as well as the explicitly solvated PMF computations, the CSFF model displayed a clear improvement over the HGFB and PHLB force fields with respect to the ω -angle rotamer populations (Table I) and the rotational timescale is on the order of a nanosecond.

Table II. Partial charges implemented for α -D-Glcp.

| | CSFF | AMBER* | GLYCAM | |
|-------|--------|--------|--------|--------------------|
| | | | 04 | 04-LP ^a |
| C1 | 0.200 | 0.513 | 0.509 | 0.292 |
| H1 | 0.090 | -0.011 | 0.000 | 0.000 |
| O1 | -0.660 | -0.673 | -0.639 | -0.440 |
| HO1 | 0.430 | 0.425 | 0.445 | 0.327 |
| C5 | 0.250 | 0.246 | 0.283 | 0.227 |
| O5 | -0.400 | -0.500 | -0.574 | -0.366 |
| C2 | 0.140 | 0.196 | 0.246 | 0.170 |
| O2 | -0.660 | -0.512 | -0.713 | -0.436 |
| HO2 | 0.430 | 0.316 | 0.437 | 0.281 |
| C3 | 0.140 | 0.196 | 0.286 | 0.109 |
| O3 | -0.660 | -0.512 | -0.699 | -0.430 |
| HO3 | 0.430 | 0.316 | 0.427 | 0.280 |
| C4 | 0.140 | 0.161 | 0.255 | 0.146 |
| O4 | -0.660 | -0.512 | -0.710 | -0.428 |
| HO4 | 0.430 | 0.316 | 0.436 | 0.278 |
| C6 | 0.050 | 0.175 | 0.277 | 0.138 |
| O6 | -0.660 | -0.425 | -0.682 | -0.420 |
| HO6 | 0.430 | 0.285 | 0.418 | 0.272 |
| H2-H6 | 0.090 | 0.000 | 0.000 | 0.000 |

^a The sum of the partial charges from the pair of lone-pairs centered on each oxygen atom. The actual charge at the oxygen position is zero.

PARM22/SU01 (58)

PARM22 was developed for CHARMM following the protocol previously employed in the development of GLYCAM for AMBER (59). Based on rotational profiles determined at the B3LYP/6-31G(d) for rotation around the ϕ angle for both axial and equatorial 2-methoxytetrahydropyran (2-MTHP), the C-C-O-C and O-C-O-C torsions were reparameterized to yield PARM22/SU01. In addition, the rotational profiles for the ω -angle in α -D-Galp-OMe and α -D-Glcp-OMe were computed with DFT without intramolecular hydrogen bonding in order to reparameterize the O5-C5-C6-O6 torsion. Close agreement was obtained between PARM22/SU01 and experimentally determined ^1H - ^1H distances and $^3J_{\text{C,H}}$ coupling constants for the explicitly solvated trisaccharide, β -D-Glcp(1 \rightarrow 2)[β -D-Glcp(1 \rightarrow 3)]- α -D-Glcp-OMe (60).

iii. Assisted Model Building and Energy Refinement (AMBER (14,15))

The key difference in AMBER, which was originally developed to model proteins and nucleic acids, versus the OPLS and CHARMM biomolecular force fields, is in the derivation of partial charges. The partial charges in AMBER are derived from a least squares fitting to *ab initio* derived molecular electrostatic potentials with the Merz-Kollman method at the HF/6-31G(d) level (61). Within this methodology, the most promising results were obtained when multiple conformations were utilized in the electrostatic potential fit and a hyperbolic restraint was employed, termed restrained electrostatic potential (RESP) fitting (62,63). Utilization of multiple conformations and a multi-stage RESP fitting resulted in more-consistent charge sets, reduced artifacts observed for the partial charges of buried atoms, and the need for a reduction in the 1,4 scaling factor for the Coulomb potential, from 1/2.0 to 1/1.2 (16).

Due to the success of the OPLS description of liquids, the OPLS van der Waals parameters were implemented in the Cornell *et al.* parameters for the AMBER force fields. For those atom types not currently available in OPLS, the effective radii (R_i) and well depth (e_i) were optimized in a similar fashion, via reproduction of densities and enthalpies of vaporization for liquids containing the pertinent atom types. This was in contrast to the original method employed in the Weiner *et al.* parameters in AMBER, which entailed fitting to crystal structures and lattice energies, which necessitated the use of a 10-12 term in the potential for atoms involved in hydrogen bonding, equation 3 (64). In both the Cornell *et al.* and Weiner *et al.* set of force fields the 1,4 van der Waals terms are scaled by a factor of 1/2.0.

$$U_{HB} = \sum_{H\text{-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right] \quad (3)$$

AMBER* (65,66)

The united atom AMBER* carbohydrate force field was modified to reproduce anomeric free energies of solvation for a variety of monosaccharides utilizing the Generalized Born/Surface Area (GB/SA) continuum model (67,68). Extension to oligosaccharides resulted in poor agreement with experimental NMR data and led to the development of the all-atom AMBER* force field (69-74). Partial charges were computed utilizing the electrostatic potential obtained with the CHELPG method at the PS-GVB/6-13G(d,p)/HF/6-31G(d,p) level (75,76). The final partial charges on exocyclic atoms were obtained via averaging the individual partial charges obtained from the axial and equatorial epimers, resulting in a single partial charge set for all pyranosides. In addition, the partial charges on aliphatic hydrogen atoms were summed into their attached carbon atom *posteriori*, yielding hydrogen atoms capable of only steric interactions. Only the aliphatic hydrogen attached to the anomeric carbon retained a partial charge, $q_{H1} = -0.011 e$, which is the same for both α and β sugars. Torsion parameters were determined using rotational profiles computed at either the PS-GVB or HF level with standard Pople basis sets for a series of model compounds extending to monosaccharides. The bond and angle parameters of uracil present in the AMBER force field were utilized and only parameters pertaining to the glycosidic and aglyconic linkages were derived from *ab initio* data (64,77).

The AMBER* all-atom model reproduced free energies of solvation for tetrahydropyran derivatives and monosaccharides utilizing the GB/SA continuum model with an average error of $0.32 \text{ kcal}\cdot\text{mol}^{-1}$. The extension of the AMBER* model to disaccharides was validated by performing a series of 10 ns hybrid Monte Carlo-Stochastic Dynamic (MC/SD) simulations and comparing the experimental and computed ${}^3J_{C,H}$ coupling constants across the glycosidic (H1-C1-O4-C4', ψ) and aglyconic (C1-O4-C4'-H4', ϕ) linkages for a series of 1 \rightarrow 4 linked disaccharides (78). Overall, for four disaccharides, methyl β -cellobioside, methyl β -maltoside, methyl β -xylobioside, and β -lactose, the ${}^3J_{C,H}(\psi)$ was computed to be within the experimental error of 1 Hz, while ${}^3J_{C,H}(\phi)$ was underestimated (79). The underestimation may be due to lack of sampling over 10 ns, which is in agreement with the observed increase in flexibility in the ϕ angle relative to the ψ angle.

AMB99C (80)

The AMBER force field was modified to reproduce B3LYP/6-31G(d) data for 10 low energy conformers of maltose. To achieve the best fit, different atom types were introduced for the ring oxygen, the hydrogen atom attached to the anomeric carbon, and the anomeric carbon in α and β sugars. A scaling factor of 1/2.0 was used for all 1,4 non-bonded interactions, which is consistent with the AMBER force fields developed prior to 1994, and the partial atomic charges were assigned from the MSI InsightII 4.0 database based on bond group charges. Initial testing focused on comparison of gas-phase properties to *ab initio* and crystal data. In addition, explicitly solvated (TIP3P) MD simulations were performed for less than 1 ns on both anomers of maltose as well as cyclomalto-oligosaccharides (81). In general, the results were in decent agreement with NMR and crystallographic data, but the short timescale makes the results inconclusive. In addition, the force field is not readily extendable to the multitude of linkages present in oligosaccharides since it was developed explicitly for 1 \rightarrow 4 linkages.

Glycosides and Glycoproteins with AMBER (GLYCAM (11,26,59,82))

In 1995, the Cornell *et al.* AMBER force field was modified to include carbohydrates by fitting torsion profiles determined at the HF/6-31G(d) and MP2/6-31G(d) level for model compounds, such as 2-MTHP (16). Three new atom types were introduced, one each for the α and β anomeric carbon and one for the glycosidic oxygen. The same 1,4 scaling factors for electrostatic and van der Waals interactions (1/1.2 and 1/2.0) were implemented as in the Cornell *et al.* force field. Partial charges were determined from fitting to the CHELPG-determined electrostatic potential at the HF/6-31G(d) level. Reproduction of experimentally determined unit-cell crystal parameters for α -D-Glcp validated the use of a 0.01 RESP weighting in determining partial charges for pyranosides (83). Additionally, in 2000 the OH-CT-CT-OS torsion term was refit to B3LYP/6-31++G(2d,2p)//HF/6-31G(d) torsion profiles for the ω -angle in α -D-Galp-OMe and α -D-Glcp-OMe. In order to accurately reproduce the torsion profiles, the 1,4 scaling factors were set to unity, thus alleviating the inequality this had caused. Furthermore, *e*RESP fitting was implemented, which involves utilizing RESP fitting to derive partial charges for an *ensemble* of 100-200 conformers taken from solvated MD simulations for each pyranoside, which were averaged *posteriori*, yielding unique partial charge sets for α and β glycosides (84). Explicitly solvated (TIP3P) 50 ns MD simulations utilizing the GLYCAM2000a force field reproduced the ω -angle rotamer populations for both α -D-Galp-OMe and α -D-Glcp-OMe, Table I (11). Underestimation of the interaction between pyranosides and the TIP3P water model during explicitly solvated MD simulations of two α -glycans was corrected by implementation of the OPLS van der Waals parameters for the hydroxyl oxygen atom, leading to GLYCAM2000b (39).

In 2004, a novel set of parameters were developed independently, but for use in conjunction with the Cornell *et al.* parameter sets in AMBER. Bond, angle, and torsion parameters were determined from B3LYP/6-31++G(2d,2p)//HF/6-31G(d) data for a training set of approximately 100 molecules that consisted of hydrocarbons, alcohols, ethers, amides, and carboxylate containing compounds, as well as molecules with mixed functional groups, while van der Waals parameters were taken from the AMBER force field. At this level of theory and basis set, a reduction of rotational barrier heights occurs in comparison to the traditional HF/6-31G(d) potential energy surfaces (Figure 6), which will lead to improved conformational sampling for oligosaccharides (85,86).

Unlike previous versions of GLYCAM, there are no generalized torsion parameters (*e.g.* X-CT-CT-X) present in GLYCAM04 or explicit hydrogen bonding terms. Partial charges were determined utilizing the same method implemented in GLYCAM2000a, but with the addition that the charges on the aliphatic hydrogen atoms were set to zero during the *e*RESP fitting (87). Similar partial charges are obtained for the ring hydroxyl groups and carbon atoms as the group-based method utilized in CSFF and OPLS-AA-SEI, with a notable difference for the anomeric carbon and the hydroxyl group at C1, Table II. The parameters were tested by their ability to reproduce the relative energies of selected carbohydrate conformations determined at the HF/6-31G(d) level of theory, as well as the ability to reproduce the experimentally determined solution-phase population ratios for the ω -, ϕ -, and ψ -angles during explicitly solvated MD simulations of mono- and disaccharides. Specifically, MD simulations of methyl α -maltoside and methyl α -isomaltoside have shown an improvement in the computed scalar *J*-couplings, Table III.

Glycosides and Glycoproteins with AMBER lone-pair model (GLYCAM-LP (32))

In 2004, point charges, which represent electron lone-pairs on the oxygen atom, were introduced into GLYCAM to develop a carbohydrate force field for use with the TIP5P water model (35). The distance of the lone-pair from the oxygen nucleus was varied from 0.0 to 1.0 Å in order to achieve the best fit to the electrostatic potential. For sp^3 type oxygen atoms, the best fit to the electrostatic potential was achieved at a distance of 0.70 Å between the oxygen atom and the lone-pair position. It is noteworthy that at this distance the partial charge on the oxygen atom converged to zero, which is consistent with the TIP5P model (35). This lone-pair inclusive charge set for carbohydrates was developed using the *e*RESP methodology utilizing the same conformations as GLYCAM04 and can be used in conjunction with either the GLYCAM2000a or GLYCAM04 parameter sets leading to GLYCAM2000a-LP and GLYCAM04-LP, respectively (84).

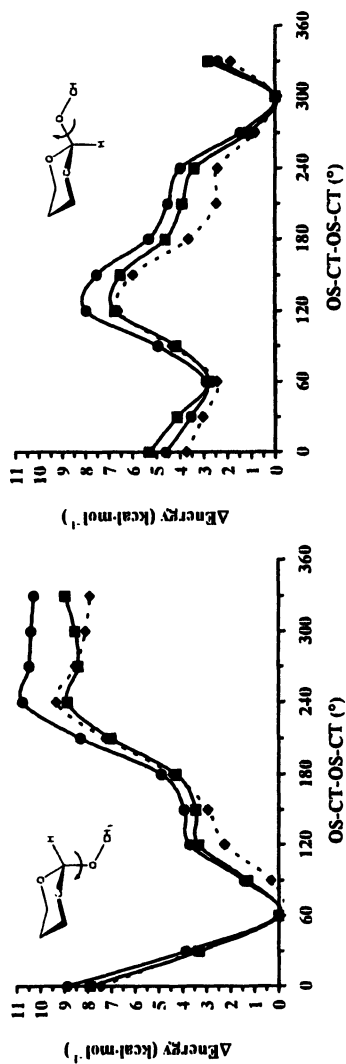


Figure 6. *Ab initio* (solid lines are HF/6-31G(d))//HF/6-31G(d): ● and B3LYP/6-311++G(2d,2p)//HF/6-31G(d): ■) and GLYCAM04 (dashed) rotational curves for the ϕ -angle in axial (left) and equatorial (right) 2-MTHP.

Table III. Scalar 3J -couplings for two α -glycans (26,32,39).

| | CHARMM | | | GLYCAM | |
|---|-------------|-------------|-------------|--------------|-----------------------|
| | <i>Exp.</i> | <i>HGFB</i> | <i>CSFF</i> | <i>2000a</i> | <i>04^a</i> |
| α -D-Glcp(1 \rightarrow 4)- α -D-Glcp-OMe | | | | | |
| $^3J_{C4'-H1}$ | 4.1 | 1.9 | 3.4 | 4.0 | 4.2 |
| $^3J_{H4'-C4}$ | 4.9 | 2.9 | 3.9 | 4.9 | 4.4 |
| $^3J_{H5H6R}$ | 5.0 | 4.1 | 4.1 | 5.8 | 3.9 |
| $^3J_{H5H6S}$ | 1.8 | 6.5 | 2.8 | 2.0 | 1.8 |
| $^3J_{H5'H6'R}$ | 5.0 | 3.4 | 2.1 | 2.3 | 4.5 |
| $^3J_{H5'-6'S}$ | 2.2 | 4.1 | 2.0 | 1.7 | 1.6 |
| $^3J_{C1'-H3}$ | 4.4 | 3.3 | 3.3 | 3.3 | - |
| α -D-Glcp(1 \rightarrow 6)- α -D-Glcp-OMe | | | | | |
| $^3J_{C6'-H1}$ | 3.2 | 2.5 | 2.7 | 2.4 | 3.1 |
| $^3J_{C1-H6'R}$ | 3.3 | 2.4 | 3.1 | 2.8 | 2.4 |
| $^3J_{C1-H6'S}$ | 3.8 | 2.1 | 2.0 | 2.0 | 1.7 |
| $^3J_{H5-H6R}$ | 5.0 | 4.5 | 3.8 | 6.5 | 3.7 |
| $^3J_{H5-H6S}$ | 2.3 | 7.3 | 2.8 | 1.9 | 1.7 |
| $^3J_{H5'-H6'R}$ | 4.5 | 3.5 | 9.3 | 5.7 | 2.5 |
| $^3J_{H5'-H6'S}$ | 2.0 | 4.5 | 1.4 | 1.9 | 1.5 |

^a Average standard deviation of 1.4 Hz.

III. Conclusion

In recent years, emphasis has shifted from reproduction of gas-phase properties to solution-phase properties for validation of carbohydrate force fields. The current generation of carbohydrate force fields have been shown to reproduce solution-phase properties, such as homo- and heteronuclear J -couplings, ω -angle rotamer populations (Tables I and III), and rotational correlation times (39). With the increased amount of experimental data for larger glycans, and developments in force field methodology, a systematic analysis of the ability of each parameter set to reproduce condensed-phase properties can be envisioned.

A complete carbohydrate parameter set must also include the possibility to model ionic sugars, such as sialic acid and uronic acids. These systems present unique developmental challenges and have not been included in force field comparisons. In addition, the majority of models mentioned here utilize three point charge water models, SPC/E or TIP3P, while a number of more accurate water models exist. Finally, as computer performance continues to increase, polarizable biomolecular MD simulations become a possibility. Overall, the accuracy of carbohydrate force fields has improved dramatically in the past decade, and the methodologies have been applied to a variety of biologically relevant problems from structural determinations to the calculation of carbohydrate-protein interaction energies (88-92). By requiring the force field to reproduce ensemble properties that are inherently dynamic in their origin, carbohydrate force fields are being held to more stringent criteria than their protein counterparts. This rigor has the benefit that MD simulations of oligosaccharides can be used predictively, and can provide a much needed basis for interpreting and augmenting limited experimental data.

References

1. Galili, U., Clark, M. R., *et al. Proc. Natl. Acad. Sci. USA* **1987**, *84*, 1369-1373.
2. Platt, J. L. and Bach, F. H. *Curr. Opin. Struct. Immun.* **1991**, *3*, 735-739.
3. Good, A. H., Cooper, D. K. C., *et al. Transplant. Proc.* **1992**, *24*, 559-562.
4. Sandrin, M. S., Vaughan, H. A., *et al. Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11391-11395.
5. Auchincloss, H. and Sachs, D. H. *Annu. Rev. Immunol.* **1998**, *16*, 433-470.
6. Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683-720.
7. Imberty, A. and Perez, S. *Chem. Rev.* **2000**, *100*, 4567-4588.
8. Wormald, M. R., Petrescu, A. J., *et al. Chem. Rev.* **2002**, *102*, 371-386.
9. Woods, R. J. *Glycoconjugate Journal* **1998**, *15*, 209-216.

10. Pinto, B. M. and Leung, R. Y. N. In *The anomeric effect and associated stereoelectronic effects*; G. R. J. Thatcher, Ed.; American Chemical Society: Washington, DC, 1993.
11. Kirschner, K. N. and Woods, R. J. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10541-10545.
12. Leroux, B., Bizot, H., *et al.* *Chem. Phys.* **1997**, *216*, 349-363.
13. Naidoo, K. J. and Chen, J. Y. J. *Mol. Phys.* **2003**, *101*, 2687-2694.
14. Ponder, J. W. and Case, D. A. In *Protein Simulations*, 2003; Vol. 66, pp 27-85.
15. Jorgensen, W. L., Maxwell, D. S. and TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.
16. Cornell, W. D., Cieplak, P., *et al.* *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
17. Kuttel, M., Brady, J. W. and Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236-1243.
18. Hemmingsen, L., Madsen, D. E., *et al.* *Carbohydr. Res.* **2004**, *339*, 937-948.
19. Naidoo, K. J. and Brady, J. W. *J. Am. Chem. Soc.* **1999**, *121*, 2244-2252.
20. Almond, A. *Carbohydr. Res.* **2005**, *340*, 907-920.
21. Kumar, S., Bouzida, D., *et al.* *J. Comput. Chem.* **1992**, *13*, 1011-1021.
22. Kumar, S., Rosenberg, J. M., *et al.* *J. Comput. Chem.* **1995**, *16*, 1339-1350.
23. Bartels, C. and Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450-1462.
24. Ueda, K., Ueda, T., *et al.* *Carbohydr. Res.* **2004**, *339*, 1953-1960.
25. Rockwell, G. D. and Grindley, T. B. *J. Am. Chem. Soc.* **1998**, *120*, 10953-10963.
26. Kirschner, K. N. and Woods, R. J., *manuscript in prep.*
27. Bock, K., Guzman, J. B. F. and Ogawa, S. *Carbohydr. Res.* **1988**, *174*, 354-359.
28. Bock, K. and Duus, J. O. *J. Carbohydr. Chem.* **1994**, *13*, 513-543.
29. Nishida, Y., Hori, H., *et al.* *J. Carbohydr. Chem.* **1988**, *7*, 239-250.
30. Nishida, Y., Ohru, H. and Meguro, H. *Tetrahedron Lett.* **1984**, *25*, 1575-1578.
31. Kuttel, M., Brady, J. W. and Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236-1243.
32. Tschampel, S. M., Kennerty, M. R. and Woods, R. J., *manuscript in prep.*
33. Kony, D., Damm, W., *et al.* *J. Comput. Chem.* **2002**, *23*, 1416-1429.
34. Jorgensen, W. L., Chandrasekhar, J., *et al.* *J. Chem. Phys.* **1983**, *79*, 926-935.
35. Mahoney, M. W. and Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910-8922.
36. Brooks, B. R., Bruccoleri, R. E., *et al.* *J. Comput. Chem.* **1983**, *4*, 187-217.
37. Ha, S. N., Giammona, A., *et al.* *Carbohydr. Res.* **1988**, *180*, 207-221.

38. Mackerell, A. D., Bashford, D., *et al.* *J. Phys. Chem. B* **1998**, *102*, 3586-3616.
39. Corzana, F., Motawia, M. S., *et al.* *J. Comput. Chem.* **2004**, *25*, 573-586.
40. Jorgensen, W. L., Madura, J. D. and Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638-6646.
41. Roothan, C. C. *J. Rev. Mod. Phys.* **1951**, *23*, 69-89.
42. Kaminski, G. A., Friesner, R. A., *et al.* *J. Phys. Chem. B* **2001**, *105*, 6474-6487.
43. Schutz, M., Rauhut, G. and Werner, H. J. *J. Phys. Chem. A* **1998**, *102*, 5997-6003.
44. Moller, C. and Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618-622.
45. Damm, W., Frontera, A., *et al.* *J. Comput. Chem.* **1997**, *18*, 1955-1970.
46. Miehlisch, B., Savin, A., *et al.* *Chem. Phys. Lett.* **1989**, *157*, 200-206.
47. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
48. Lee, C., Yang, W. and Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.
49. Maxwell, D. S. and Tirado-Rives, J.; 1.1.1 ed.; Yale University: New Haven, CT, 1994.
50. Kony, D., Damm, W., *et al.* *J. Phys. Chem. B* **2004**, *108*, 5815-5826.
51. Poppe, L. *J. Am. Chem. Soc.* **1993**, *115*, 8421-8426.
52. Brown, G. M. and Levy, H. A. *Acta Cryst. B* **1979**, *35*, 656-659.
53. Vasko, P. D., Blackwel.J and Koenig, J. L. *Carbohydr. Res.* **1971**, *19*, 297-310.
54. Vasko, P. D., Blackwel.J and Koenig, J. L. *Carbohydr. Res.* **1972**, *23*, 407-416.
55. Kouwijzer, M. L. C. E., van Eijck, B. P., *et al.* *J. Comput. Chem.* **1993**, *14*, 1281-1289.
56. Hough, E., Neidle, S., *et al.* *Acta Cryst. B* **1973**, *B 29*, 365-367.
57. Palma, R., Zuccato, P., *et al.* In *Glycosyl Hydrolases in Biomass Conversion*; M. E. Himmel, J. O. Baker and J. N. Saddler, Eds.; American Chemical Society: Washington, D.C., 2001; Vol. 769, pp 112-130.
58. Eklund, R. and Widmalm, G. *Carbohydr. Res.* **2003**, *338*, 393-398.
59. Woods, R. J., Edge, C. J., *et al.* In *Complex Carbohydrates in Drug Research*; K. Bock, H. Clausen, P. Krogsgaard-Larsen and H. Koford, Eds.; Munksgaard: Copenhagen, Denmark, 1993; Vol. 36, pp 15-36.
60. Rundlof, T., Eriksson, L. and Widmalm, G. *Chem-Eur J* **2001**, *7*, 1750-1758.
61. Singh, U. C. and Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129-145.
62. Bayly, C. I., Cieplak, P., *et al.* *J. Phys. Chem.* **1993**, *97*, 10269-10280.
63. Cornell, W. D., Cieplak, P., *et al.* *J. Am. Chem. Soc.* **1993**, *115*, 9620-9631.
64. Weiner, S. J., Kollman, P. A., *et al.* *J. Am. Chem. Soc.* **1984**, *106*, 765-784.
65. Senderowitz, H., Parish, C. and Still, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 8985-8985.

66. Senderowitz, H. and Still, W. C. *J Org Chem* **1997**, *62*, 1427-1438.
67. Still, W. C., Tempczyk, A., *et al.* *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.
68. Cramer, C. J. and Truhlar, D. G. *J. Am. Chem. Soc.* **1993**, *115*, 5745-5753.
69. Carver, J. P., Mandel, D., *et al.* In *Computer Modeling of Carbohydrate Molecules*; A. D. French and J. W. Brady, Eds.; American Chemical Society: Washington, DC, 1990.
70. Poppe, L. and Vanhalbeek, H. *J. Magn. Reson.* **1991**, *93*, 214-217.
71. Hricovini, M., Tvaroska, I. and Hirsch, J. *Carbohydr. Res.* **1990**, *198*, 193-203.
72. Shashkov, A. S., Lipkind, G. M. and Kochetkov, N. K. *Carbohydr. Res.* **1986**, *147*, 175-182.
73. Lipkind, G. M., Shashkov, A. S. and Kochetkov, N. K. *Carbohydr. Res.* **1985**, *141*, 191-197.
74. Hamer, G. K., Balza, F., *et al.* *Can J Chem* **1978**, *56*, 3109-3116.
75. Breneman, C. M. and Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361-373.
76. Ringnalda, M. N., Langlois, J. M., *et al.*; 2.01 ed.; Schrodinger, Inc., 1994.
77. Weiner, S. J., Kollman, P. A., *et al.* *J. Comput. Chem.* **1986**, *7*, 230-252.
78. Guarnieri, F. and Still, W. C. *J. Comput. Chem.* **1994**, *15*, 1302-1310.
79. Mulloy, B., Frenkiel, T. A. and Davies, D. B. *Carbohydr. Res.* **1988**, *184*, 39-46.
80. Momany, F. A. and Willett, J. L. *Carbohydr. Res.* **2000**, *326*, 194-209.
81. Momany, F. A. and Willett, J. L. *Carbohydr. Res.* **2000**, *326*, 210-226.
82. Woods, R. J., Dwek, R. A. and Edge, C. J. *J. Phys. Chem.* **1995**, *99*, 3832-3846.
83. Woods, R. J. and Chappelle, R. J. *J. Mol. Struct.* **2000**, *527*, 149-156.
84. Basma, M., Sundara, S., *et al.* *J. Comput. Chem.* **2001**, *22*, 1125-1137.
85. Lii, J. H., Ma, B. Y. and Allinger, N. L. *J. Comput. Chem.* **1999**, *20*, 1593-1603.
86. Kirschner, K. N. and Woods, R. J. *J. Phys. Chem. A* **2001**, *105*, 4150-4155.
87. Francl, M. M., Carey, C., *et al.* *J. Comput. Chem.* **1996**, *17*, 367-383.
88. Ford, M. G., Weimar, T., *et al.* *Proteins* **2003**, *53*, 229-240.
89. Gonzalez-Outeirino, J., Kadirvelraj, R. and Woods, R. J. *Carbohydr. Res.* **2005**, *340*, 1007-1018.
90. Bernardi, A., Potenza, D., *et al.* *Chem-Eur J* **2002**, *8*, 4598-4612.
91. Colombo, G., Meli, M., *et al.* *Carbohydr. Res.* **2004**, *339*, 985-994.
92. Lycknert, K., Rundlof, T. and Widmalm, G. *J. Phys. Chem. B* **2002**, *106*, 5275-5280.

Chapter 14

Docking Carbohydrates to Proteins: Empirical Free Energy Functions

Alain Laederach

Department of Genetics, Stanford University, Stanford, CA 94304
(email: alain@helix.stanford.edu, phone: 650-725-4484)

Abstract

Automated docking of small molecules to proteins is a powerful approach for both predicting and understanding the molecular mechanisms of ligand recognition and specificity. Carbohydrate recognition is a key component of many biological processes including innate immunity, metabolism, and immune response. Computational tools specific for studying the molecular basis of carbohydrate recognition have recently been developed that efficiently and accurately dock a flexible carbohydrate to a protein. Two aspects of carbohydrate docking are particularly challenging; parametrization of the potential energy function for the intra- and intermolecular forces and efficient conformational searching. The combination of a genetic conformational search algorithm coupled with a novel empirically derived free energy function specific for carbohydrates allows the prediction of free energies of formation to $1 \text{ kcal}\cdot\text{mol}^{-1}$. This chapter covers the basic principles of empirical free energy functions and their application to carbohydrates.

Introduction

Carbohydrate-protein interactions play critical roles in many biological processes. Specific recognition of carbohydrates by proteins is a key element of the innate immune response, metabolism, fertilization, embryogenesis, and neuronal development (1-3). Carbohydrates and carbohydrate like molecules are therefore beginning to play an important role in the basic and pharmaceutical sciences and the need for computational tools to accurately model them is increasing (4, 5). More specifically, tools that can predict the structure of protein-carbohydrate complexes are needed. Furthermore, given the structure of a carbohydrate-protein complex, tools that accurately predict the free energy of formation are needed for the rational design of specific small molecule inhibitors.

The rich structural diversity of carbohydrates has resulted in a wide variety of carbohydrate recognizing proteins (6-9). In general, carbohydrate-protein complexes are stabilized by a combination of specific hydrogen bonds and hydrophobic contacts (10, 11). Specificity is often governed by the formation of only a few critical hydrogen bonds (12, 13). Accurate prediction of protein-carbohydrate complex structures is therefore contingent upon two criteria. An accurate potential energy function (often called the scoring function) is needed, such that the minimum of the energy function corresponds to the correct structure. Secondly, an efficient search algorithm is needed to explore the conformational space of the carbohydrate-protein complex in order to find this global energy minimum.

Docking is an optimization problem (14, 15). An objective function is defined that evaluates the energy of a starting protein-carbohydrate conformation. The carbohydrate is then moved relative to the protein, and the energy re-evaluated. This process is repeated until an optimal conformation is found where the energy is lowest, or minimized. A docking strategy is defined by the combination of the potential energy function, search algorithm, and parameters used. This chapter describes some of the recent developments in carbohydrate docking strategies, and provides the reader with a basis for deciding which docking strategy is best suited for his/her particular problem. Indeed many different docking strategies have been developed, but some are better suited for carbohydrates than others (16).

Conformational search

The conformational search problem for carbohydrates is complicated by the inherent flexibility of these types of molecules. Unlike other small molecules with many double bonds, most carbohydrates are mainly comprised of tetrahedral carbons that afford the molecule significant flexibility. As a result the conformational search space is important, and only a judicious choice of search algorithm parameters can make the problem computationally tractable. Two

major types of search algorithms can be used, gradient driven minimization (or local minimization) and gradient free optimization. For most docking potential energy functions (see next section), an analytical derivative of the function exists and can be used for local minimization. Nonetheless, the problem of how to move the carbohydrate remains. Simply moving atoms in space is not satisfactory as atoms are bonded to each other, and atoms often move in a concerted fashion.

To capture concerted atomic motions in carbohydrates, moves can be accomplished in internal coordinate space, rather than Cartesian space. This requires a judicious choice of internal coordinates. Figure 1 shows a typical carbohydrate molecule (three rings).

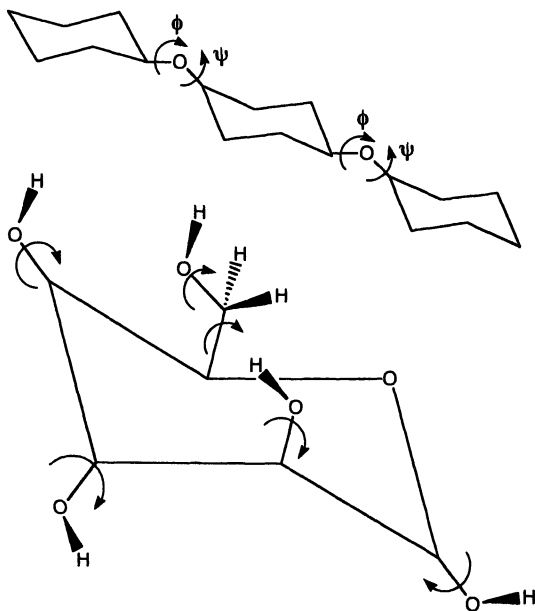


Figure 1. Illustration of internal coordinates used to confer flexibility to carbohydrate structures by rotation around a bond.

By allowing rotation about the glycosidic linkages, hydroxyls, and the C5-C6 bond, most of the carbohydrate's flexibility is captured. However, ring pucker is also an important contributor to carbohydrate flexibility. It is treated separately in most docking strategies. Multiple ring pucker conformations are docked independently. These docking simulations are independent and thus embarrassingly parallel, and can be run on large clusters of computers (17).

The state of the system is thus fully described by a series of variables. If a molecule has n rotatable bonds, then the n values of their respective angles, combined with the Cartesian coordinates of the center of mass and 4 quaternion

values define a unique conformation of the ligand. For each conformation an energy value can be computed, and the docking problem is simplified to finding the variables that minimize the energy. An analytical form of the gradients in this new internal coordinate space is significantly more difficult to derive than in Cartesian space. As a result non-gradient minimization algorithms (e.g. Monte Carlo or genetic algorithms) have provided better performance.

The advantages of using internal coordinates for docking greatly outweigh the disadvantages, most notable the fact that the analytical gradients are difficult to compute. Local minimization algorithms in Cartesian space can quickly converge on a local minimum, but a very large number of initial conditions must be tested in order to find the global minimum. Using the internal coordinate representation of the docking problem, moves are much more efficient. This is especially true if a genetic algorithm is used to find the global energy minimum. The principal of the genetic algorithm is illustrated in Figure 2.

Genetic algorithms are best adapted for docking carbohydrates, provided they are coupled with a local minimizer. In general we have found that applying local minimization every 100 generations to all members in the population yields optimal docking results. There are certain applications where using a Monte-Carlo/simulated annealing protocol (MC/SA) is also advantageous. Internal coordinate space must be used with MC/SA as random moves are much more likely to be accepted, given that they result in concerted atomic movement. MC/SA protocols guarantee a Boltzmann distribution of docked conformations. Analysis of the population of different conformations (as measured by structural clustering) can yield insight into the relative entropy of a particular conformation (15, 18). However, as will be shown in the next chapter, entropy of binding can be quantified by a specific term in the potential energy function, and as a result most docking strategies now use a genetic algorithm.

Potential Energy Functions

Accurate docking of carbohydrates to proteins draws on advances made in the fields of optimization and potential energy function parameterization. Successful docking strategies require combining a well-parameterized potential energy function with an efficient conformational search strategy. The efficiency of the search algorithm can greatly reduce the computational effort required to obtain a docked structure; however if the potential energy function is not accurate, the docking results will be erroneous. Similarly if the conformational search algorithm does not find the global minimum of the potential energy function, equally wrong results may be obtained. In general the form of most molecular mechanics potential energy function is relatively standard:

$$\Delta E = \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i,j} E(\theta) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \sum_{i,j} \left(\frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right) + \left[\sum_{\text{torsions}} K_\phi (1 + \cos(n\phi) - \delta) + \Delta H_{vdW}^{\text{ligand}} + \Delta H_{elec}^{\text{ligand}} + \Delta H_{hbond}^{\text{ligand}} \right] \quad (1)$$

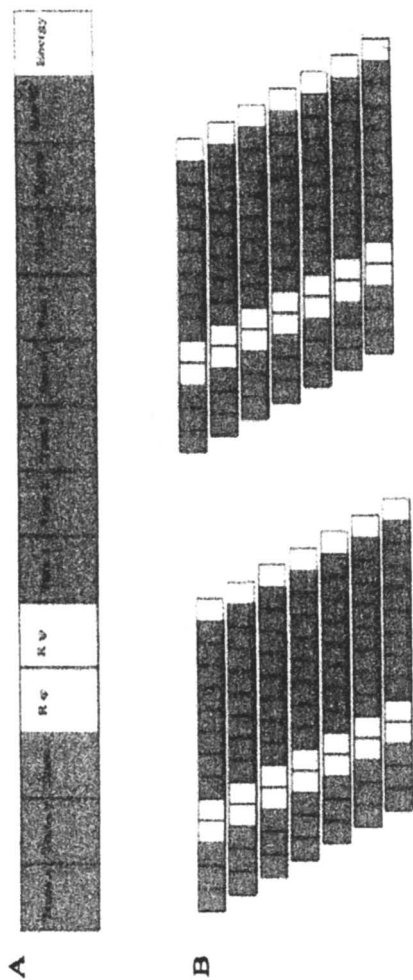


Figure 2. Illustration of the principle of the genetic algorithm docking strategy. A.) The state of the system is entirely defined by a series of variables: one value for each torsion, the coordinates of the center of mass (Trans X, Y and Z) and four quaternion values. B.) Multiple random states are generated as a starting population. The system is allowed to evolve by mutation and crossover. (See color page 1 in chapter.)

The potential energy function has two main parts. The first three terms represent the intermolecular potential, van der Waals, hydrogen bonding, and electrostatics, respectively. The four last terms evaluate the intra-molecular energy of the ligand. Certain binding modes, especially in enzymes that catalyze carbohydrate reactions, can often put internal strain on the ligand and these terms account for it (16, 18). The parameters ($A, B, C, D, q, K, \phi, \delta$) for these potential energy functions have been refined such that in general the minimum of the potential energy function is the correct structure (19). Unfortunately however, the potential energy function described above does not take into account effects of solvation or binding entropy.

To model solvent effects into a docking strategy, the principle of the thermodynamic cycle can be used (14, 16). This principle is illustrated in Figure 3.

The thermodynamic cycle is based on the fact that the change in free energy of a system is independent of the path taken. Therefore, binding of a ligand to a protein can be broken down into steps. First the ligand and protein are removed from solution. Binding then occurs in vacuum, and finally the complex is resolvated. This is thermodynamically equivalent to complex formation in solution. However breaking down the binding event into such steps greatly simplifies modeling. Indeed, solvation models exist based on changes in accessible hydrophobic surface area that can accurately predict free energies of solvation (14, 16). Furthermore, docking is greatly simplified when performed in vacuum as no explicit water is required.

Practically speaking, the implementation of the thermodynamic cycle is accomplished using an empirical free energy function:

$$\Delta G = f_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + f_{hbond} \left(\sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + E_{hbond} \right) + f_{elec} \sum_{i,j} \left(\frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right) + \Delta G_{tor} N_{tor} + f_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2 / 2\sigma^2)} \quad (2)$$

The intermolecular terms in Equation 2 (first three terms) are similar in form to Equation 1, however each term is multiplied by an empirical coefficient (f). The intra-molecular energy is treated completely differently. The fourth term in Equation 2 is an entropic penalty, and it is dependent on the number of rotatable bonds in the ligand (N_{tor}). The fifth term describes the change in hydrophobic solvent accessibility upon ligand binding. The fourth and fifth terms are multiplied by empirical coefficients, ΔG_{tor} and f_{sol} . Hydrogen bonding is treated separately in this model, as it is key to correctly model hydrogen bonds energetically and conformationally. E_{hbond} is added to the hydrogen bonding term to account for hydrogen bonding to solvent. When the carbohydrate binds to the protein, it forms hydrogen bonds to the protein. However, in order to form these bonds it must break the hydrogen bonds to water. E_{hbond} is therefore an adjustable parameter of the model that accounts for breaking of solvent hydrogen bonds upon binding.

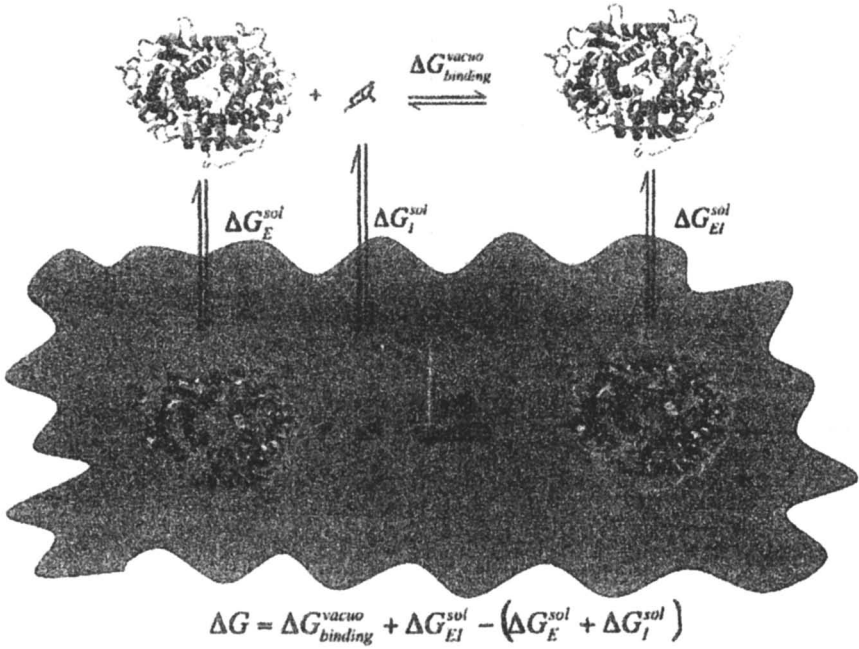


Figure 3. Illustration of the thermodynamic cycle principle. The blue shape represents solvent. The free energy change (ΔG) of the binding event in solution can be calculated as the sum of the free energy change in vacuum plus the difference in the free energy changes of solvation of the ligand, protein and protein-ligand complex. (See color page 1 in chapter.)

Methods

Calibration of the Free energy function

The empirical free energy function must be calibrated, and this is accomplished using a training set. The training set contains a series of protein-carbohydrate complexes of known structure and affinity, the details of which are described by Laederach and Reilly (16). The calibration process is straight forward. For each complex of known structure and affinity the ligand is docked (using a local minimizer) to the active site using as objective function Equation 1. By starting the docking simulation in the crystallographic (known) conformation and using a local minimizer, the ligand will be relaxed and atomic clashes resolved. The values of the five terms in Equation 2 can then be evaluated. The five empirical parameters (f_{vdw} , f_{hbond} , f_{elec} , f_{sol} , and ΔG_{Tor}) are then fit to the data set using linear regression. This is illustrated in Figure 4.

Several empirical models can be tested. For example, the value of E_{hbond} can be adjusted. In most potential energy functions the maximum energetic contribution of a hydrogen bond is 5 kcal-mol⁻¹. However in solution, rarely is the hydrogen bond geometry ideal; therefore it is likely that the energetic penalty for breaking a hydrogen bond to water will be less than 5 kcal-mol⁻¹. When calibrating the empirical free energy function, we used a total of 18 different models and selected the model that gave the lowest residual error. Table 1 lists the empirical coefficient values for the three best models that fit the data. Our work differs from previously reported empirical free energy functions in that the training set we used was comprised only of protein-carbohydrate complexes.

Model A (Table 1) has the lowest residual error and can be used as a potential energy function for docking of carbohydrates to proteins. It is interesting to compare the values of the empirical parameters for this model to a similar model developed for drugs and drug like molecules (14). In Model A, the weight of the van der Waals interactions is half that reported in Ref. 14. This is compensated by a three fold larger electrostatic contribution. This is not surprising given the importance of electrostatics in carbohydrate recognition.

Discussion

The empirical free energy model specific to carbohydrates described above has been implemented for use in AutoDock. AutoDock is a software package that allows users to perform docking simulations of small molecules to proteins. Many of the steps required for docking, such as definition of rotatable bonds, preparation of the ligand and protein files, are automated such that novice users can relatively quickly setup a docking simulation. Use of the carbohydrate specific model is described in the supplementary material available from <ftp://ftp.wiley.com/public/journals/jcc/suppmat/24/1748>.

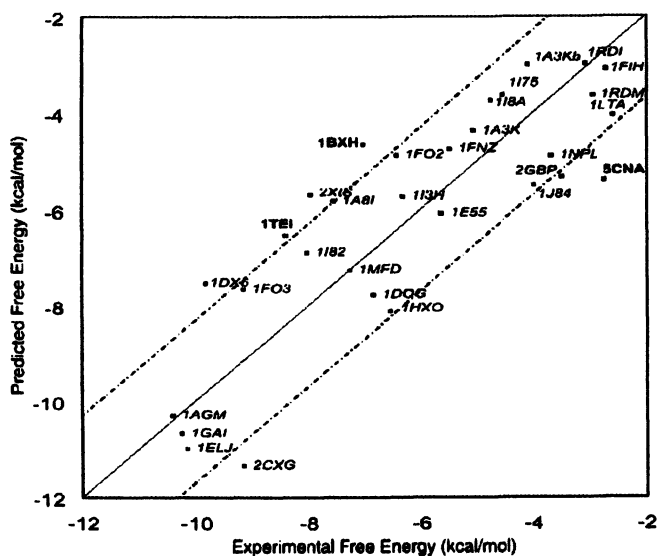


Figure 4. Predicted versus experimental free energy residuals after linear regression of empirical coefficients in Equation 2 to a training set of 30 protein-carbohydrate complexes of known structure and affinity. Residual error is $1.4 \text{ kcal}\cdot\text{mol}^{-1}$. Adapted from Laederach and Reilly (16). Copyright 2003 Wiley.

Table 1. Least Squares Fitting of the Top Three Models to the Training Set.

| Model | E_{nb} (kcal/mol) | Torsion model | Residual standard error (Å) | | f_{vib} (—) | f_{ibond} (—) | f_{elec} (—) | ΔG_{tor} (kcal/mol) | f_{cal} (—) |
|----------------------|------------------------|------------------|--------------------------------|-------------|----------------------------|--------------------|-------------------|--------------------------------|------------------|
| | | | Training set | Test set | | | | | |
| A | 1 ^a | Heavy | 1.403 | 1.101 | 0.0737±0.0344 ^b | 0.0566±0.0401 | 0.3330±0.0982 | 0.0497±0.0091 | 0.1186±0.0712 |
| B | 1 ^c | Heavy | 1.418 | 1.163 | 0.0801±0.0401 | 0.0526±0.0212 | 0.3447±0.1089 | 0.1138±0.0641 | 0.1401±0.0482 |
| C | 5 ^d | All | 2.708 | 1.738 | 0.0307±0.0450 | 0.0332±0.0406 | 0.2768±0.1321 | 0.0976±0.0772 | 0.0274±0.0991 |
| Ref. 14 ^d | — | Heavy | 2.177 | 2.070 | 0.1485±0.0237 | 0.0656±0.0558 | 0.1146±0.0238 | 0.3113±0.0910 | 0.1711±0.1035 |

^aEq. (3) was used as the hydrogen bonding model^bStandard error.^cEq. (4) was used as the hydrogen bonding model.^dThese parameters were derived using a different training set and are for a different hydrogen-bonding model.

AutoDock is available free of charge to academic users from <http://www.scripps.edu/mb/olson/doc/autodock/>.

As mentioned previously, successful docking of carbohydrates requires a good conformational search algorithm combined with an accurate potential energy function. It is critical to couple a local minimization algorithm to all global search strategies. AutoDock uses a Solis & Wets algorithm for local minimization of a structure. In our experience docking carbohydrates, far superior results are obtained if a large number (>3000) minimization steps are used at regular intervals whether a MC/SA or genetic global optimizer is used. If the structures are not minimized it is very unlikely that the global minimum will be found. This is due to the rugged nature of the energy landscape, especially when searching using internal coordinate space. Small changes in a single angle can cause relative large motions in the molecule resulting in major steric collisions and very large energies.

Although modifying search algorithms parameters can significantly improve docking results, there are cases where the minimum of the potential energy function does not correspond to the correct structure. This is often the case with proteins that bind carbohydrates on their surface, like lectins (9). In general, lectins have relative weak binding affinities for carbohydrates. They make up for this weak affinity by using avidity effects (12, 13). However, given that docking simulations are run on single molecules, these effects are not apparent in the models. As a result, docking protocols have significant problems finding the correct structure. This is further complicated when an intermolecular water mediated hydrogen bond is present in the complex, as is the case in Concanavalin A (Figure 5). In these cases, simple docking strategies are generally inapplicable and more involved simulations are required, such as explicit water molecular dynamics.

Another important consideration when docking carbohydrates to proteins is the use of the empirical free energy function as the objective function. In most docking protocols the empirical free energy function is used as the objective function. This can potentially lead to errors in the predicted structure of the protein-carbohydrate complex. The parameters of the empirical free energy function are adjusted so as to provide the correct free energy given a particular structure. The function's global minimum may therefore not correspond to the correct structure. In general, it is better to use as an objective function Equation 1, which also models conformational strain of the ligand, and then apply the free energy function to determine the free energy of binding. More details into the possible error introduced by using the empirical free energy as the objective function are provided by Laederach & Reilly (16).

Conclusions

Docking of carbohydrates to proteins can provide significant insight into the molecular details of the recognition process. The development of empirical free energy functions allows the models to predict experimental values, and this is a critical step towards validation of the models. Nonetheless, docking protocols are ultimately limited by the assumptions made during model development. A

non-explicit treatment of solvation limits the ability to correctly predict structures where water mediated hydrogen bonds occur. Currently, efforts are being made to include explicit waters in docking strategies, however these greatly complicate the conformational search problem. In general however, for screening purposes, automated docking remains an efficient way to identify potential high affinity binding molecules.

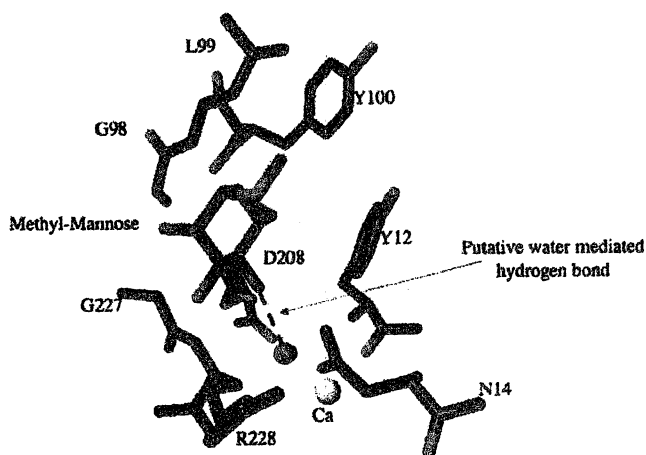


Figure 5. Binding site of Concavalin A (9). A water molecule serves as a hydrogen bond bridge between the mannose ring and the Arg 228 side chain. (See color page 1 in chapter.)

Acknowledgements

A.L. is grateful to the US National Science Foundation for a pre-doctoral fellowship (IGERT 9927653), and to the Damon Runyan Cancer Research Foundation post-doctoral fellowship program.

References

1. Guse, A.H. *Curr Mol Med* **2004**, 4, 239-248.
2. Guse, A.H. *Curr Med Chem* **2004**, 11, 847-855.

3. Santella, L., Lim, D., and Moccia, F. *Trends Biochem Sci* **2004**, 29, 400-408.
4. Barchi, J.J., Jr. *Curr Pharm Des* **2000**, 6, 485-501.
5. Witzcak, Z.J. *Curr Med Chem* **1999**, 6, 165-178.
6. Gabius, H.J., Siebert, H.C., Andre, S., Jimenez-Barbero, J., and Rudiger, H. *Chembiochem* **2004**, 5, 740-764.
7. Andre, S., Kaltner, H., Lensch, M., Russwurm, R., Siebert, H.C., Fallsehr, C., Tajkhorshid, E., Heck, A.J., von Knebel Doeberitz, M., Gabius, H.J., and Kopitz, J. *Int J Cancer* **2004**.
8. Gabius, H.J., Andre, S., Kaltner, H., and Siebert, H.C. *Biochim Biophys Acta* **2002**, 1572, 165-177.
9. McGreal, E.P., Martinez-Pomares, L., and Gordon, S. *Mol Immunol* **2004**, 41, 1109-1121.
10. Coutinho, P.M., Dowd, M.K., and Reilly, P.J. *Proteins* **1997**, 28, 162-173.
11. Coutinho, P.M., and Reilly, P.J. *Proteins* **1997**, 29, 334-347.
12. Allen, M.J., Laederach, A., Reilly, P.J., Mason, R.J., and Voelker, D.R. *Glycobiology* **2004**, 14, 693-700.
13. Allen, M.J., Laederach, A., Reilly, P.J., and Mason, R.J. *Biochemistry* **2001**, 40, 7789-7798.
14. Osterberg, F., Morris, G.M., Sanner, M.F., Olson, A.J., and Goodsell, D.S. *Proteins* **2002**, 46, 34-40.
15. Morris, G.M., Goodsell, D.S., Huey, R., and Olson, A.J. *J Comput Aided Mol Des* **1996**, 10, 293-304.
16. Laederach, A., and Reilly, P.J. *J Comput Chem* **2003**, 24, 1748-1757.
17. Mulakala, C., and Reilly, P.J. *Proteins* **2002**, 49, 125-134.
18. Laederach, A., Dowd, M.K., Coutinho, P.M., and Reilly, P.J. *Proteins* **1999**, 37, 166-175.
19. Woods, R.J., Dwek, R.A., Edge, C.J., and Fraser-Reid, B. *J Phys Chem* **1995**, 99, 3832-3846.

Chapter 15

Molecular Modeling of Carbohydrates with No Charges, No Hydrogen Bonds, and No Atoms

Valeria Molinero and William A. Goddard III

Materials and Process Simulation Center, MC 139-74, California Institute of Technology, Pasadena, CA 91125

We review the M3B coarse grain model for glucans and their aqueous mixtures that we show is able to describe the helical structures of oligo- and polysaccharides, experimental glass transition temperatures, and the dynamics of water in supercooled and glassy malto-oligosaccharide mixtures. M3B involves no electrostatic interactions, no hydrogen bonds, and no other directional nonbond interaction. Indeed M3B has no atoms, just three connected beads per monomer. M3B permits coarse grain simulations ~ 7000 faster times than atomistics, yet provides at the end a mapping back to the full atomistic description.

Introduction

Carbohydrates are polyhydroxylated compounds, whose conformations and molecular dynamics are dominated by hydrogen bond interactions. Among natural carbohydrates, glucose homopolymers abound as cellulose, starch, glycogen, and dextran. These polymers consist exclusively of glucose residues connected through 1 \rightarrow 4 and/or 1 \rightarrow 6 glycosidic linkages. The partial hydrolysis of starch renders amyloextrins syrups, mainly linear $\alpha(1\rightarrow4)$ glucans with degree of polymerization (DP) of 2 to 60. These syrups are ubiquitous in the food industry, being used to control the rheology, hygroscopicity, browning ability, and sweetness of the products. Their physical properties are determined

by the temperature and composition (their water content and distribution of glucans of different DP).

Molecular dynamics (MD) simulations provide an adequate level of detail to study the structure and dynamics of concentrated water-carbohydrate systems. Nevertheless, the use of atomistic MD to predict water distribution, glass transition temperatures and diffusion coefficients for water and carbohydrates in concentrated solutions has been limited to simple binary water-carbohydrate mixtures(1-4) due to the hampering costs of simulating complex systems for times well above the nanoseconds. In order to dramatically speedup the calculations, we have developed a *coarse grain model* in which the degrees of freedom are decreased substantially, allowing larger time steps while eliminating long range coulomb potentials. This coarse grain model provides a particle-based description of the system – just as for the atomistic one- where the particles (beads) correspond not to atoms but rather to groups of atoms, monomers or complete chains -depending on the degree of coarsening of the model.(5) These beads interact through effective potentials that ensure that the coarse grain system accurately reproduces certain behaviors of the system.

In this chapter we summarize the development of the M3B coarse grain model for glucans – the first of this kind for carbohydrates - and show that in spite of not having hydrogen bonds, nor electrostatic interactions, and not even atoms, M3B is able to reproduce the variety of helical forms characteristic of α -1 \rightarrow 4 glucans, the glass transition temperatures of pure glucose, glucose oligomers and their water mixtures, the microscopic distribution of water in concentrated carbohydrates. Indeed, by extending the simulations times to the order of a microsecond, it allows the study of the mechanism of water diffusion in supercooled and glassy carbohydrate mixtures, reproducing the experimental order of magnitude of water diffusion in the glass and its decoupling from carbohydrate diffusion.

M3B coarse grain model of glucans and their aqueous mixtures.

Since we want the coarse grain model to be based on first-principles, but we want to handle long length and time scales, we develop it from fundamental methods, ideally from atomistic force fields which were in turn based on quantum mechanical (QM) calculations. In the latter case, the electronic degrees of freedom in QM are averaged out for each set of nuclear coordinates (atomic positions) to obtain effective potentials for the interaction of atoms that reproduce the QM energies for some set of molecules and geometries. In the coarse graining of atoms into beads the procedure is analogous: we optimize effective potentials between the beads to reproduce the selected results (structural and thermodynamic data) of atomistic simulations of the system that are relevant to the target properties of interest.

The degree of coarsening of the model is determined by the target properties. The M3B coarse grain model for glucose oligosaccharides was developed to address the distribution of water in starch hydrolyses, the conformations and dynamics of oligosaccharide chains in these mixtures, and the dynamics and mechanisms of mobility of water and carbohydrates in the supercooled and glassy state.

The M3B model represents each glucose monomer by three beads (B1, B4, B6) bonded to each other. The three beads per glucose residue are mapped into the positions of the carbons C1, C4 and C6 of the atomistic model, as shown in Figure 1. This mapping allows the 1 \rightarrow 1, 1 \rightarrow 4, and 1 \rightarrow 6 glycosidic bonds in glucose oligomers and polymers to be properly described as simple bonds B1-B1', B1-B4', and B1-B6' of contiguous residues, respectively. The degree of coarsening of the M3B model is appropriate to reproduce the excluded volume (size and shape, see below) and segmental flexibility of the carbohydrate molecules around the glycosidic linkage.(6) To be consistent with the coarseness of the glucose, we represent the water molecule by a single particle, which we find is adequate for the microscopic study of the molecular mechanism of mobility in concentrated carbohydrate mixtures.(7)

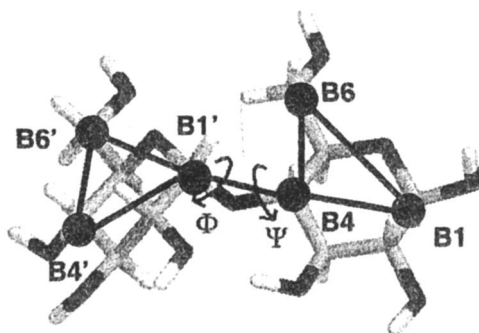


Figure 1. The beads (black balls) in the M3B coarse grain model of maltose are positioned on the carbons C1, C4 and C6 of the atomistic model. The intermonomer bond between B1' and B4 represents the 1 \rightarrow 4 glycosidic linkage. ϕ and ψ denote the glycosidic torsional angles of the atomistic model.

The relative rigidity of the glucose ring allows fast, reliable reconstruction of the atomistic coordinates from the position of the coarse grain particles. The positions of the carbons C1, C4 and C6 of a glucose residue completely define the orientation of the D-glucose residue. The only significant uncertainty while reconstructing the atomistic model is related to the rotation of the exocyclic group around C6. This inverse transformation is unique without need of costly

iterative procedures.(6) To estimate the fidelity of the reconstruction, we mapped an arbitrary configuration of a DP11 malto-oligomer onto the M3B model – losing all the atomistic information- and then reconstructed the coordinates of the 234 atoms from the 33 beads. The RMS difference between the original and reconstructed structure was just 0.34 Å, less than 1/10 of the radius of the smaller coarse grain particle. This uniqueness and invertability mapping from atomistic to M3B is essential for important uses of the M3B model in which fast coarse grain simulations are used to rapidly equilibrate the system, which is then mapped back into the atomistic model where specific properties (detailed structure, free volume or short time dynamics) are analyzed.

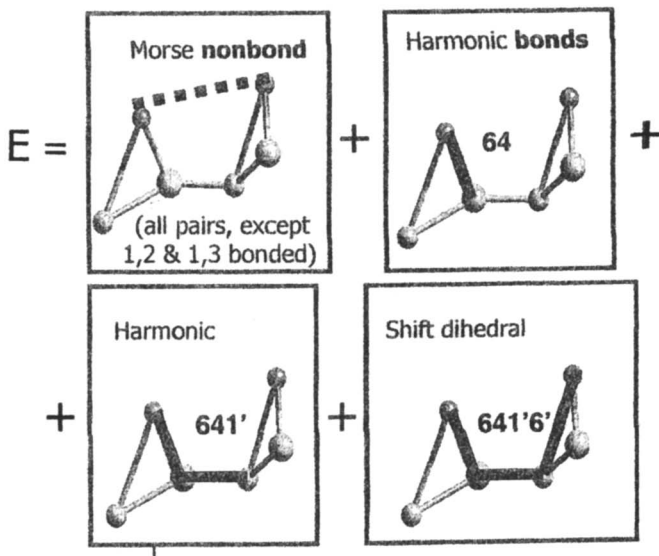


Figure 2. The M3B coarse grain model consists of a sum of Morse nonbonded interactions $E(R_{ij}) = D_o \{Y^2 - 2Y\}$, where $Y = \exp\{0.5\alpha(1 - r_{ij} / r_o)\}$, between all pairs of beads that are not involved in a bond or bending angle, plus bonded terms: harmonic bond terms $E_{bond}(r) = 0.5k_b(r - r_o)^2$ between pairs of bonded atoms, harmonic angle terms $E_{angle}(\theta) = 0.5k_\theta(\theta - \theta_o)^2$ between every three connected beads, and shifted the torsions interactions for all dihedrals φ (four linearly connected beads) $E_{torsion}(\varphi) = \sum 0.5B_j(1 - \cos(j\varphi - \varphi_o^j))$. Only one contribution to each term for M3B maltose is illustrated in the scheme.

The energy expression of the M3B force field uses only simple bonded and nonbonded potentials (Figure 2). The parameters were optimized to reproduce thermodynamic and structural parameters of bulk amorphous glucose and $\alpha(1\rightarrow4)$ D-glucans in a wide range of pressure, from -1 to 20 GPa. We found that fitting the equation of state was essential in achieving the proper balance in the compressibility of the valence bonds versus the nonbond terms. The M3B force field renders densities, cohesive energies, and structural measures (cell parameters, bonds, angles, torsions) within 2.5% of the atomistic model over the entire range of pressure used in the parameterization.(6) The atomistic simulations were all performed with a slightly modified version of Dreiding force field.(6,8) The three nonbond parameters for water were chosen to reproduce the experimental density, cohesive energy, and diffusion coefficient of water at 298 K.(6) The details of the procedure and a complete list of M3B parameters for any arbitrary $\alpha(1\rightarrow4)$ glucans are published elsewhere.(6)

In the atomistic model the hydrogen bonds contribute to a background cohesive energy and an essential stabilization of particular conformations of the oligomer chains, giving rise to helical structures as observed in crystalline forms of amylose.(9-11) In contrast the M3B force field has *no* electrostatic interactions (the beads are neutral) and there are *no* hydrogen bond interactions between them. Instead the M3B model exploits the existence of ubiquitous hydrogen bonds between all the species to provide a simple description of the nonbond interactions through non-directional Morse potentials. The Morse potential provides both the magnitude of the attraction between molecules, and the excluded volume of the molecules. Figure 3 shows that the shape and size of glucose as would be perceived from a probe with the size of water molecule is very well represented in the M3B model. Although M3B does not have explicit hydrogen bond or electrostatic terms, these are accounted for in the parameterization of the coarse grain force field. The directional effect of the hydrogen bonds is not absent in the M3B model: it is encoded in the bonded potentials, in particular in the coarse grain bending and torsional angles, as discussed in the next section.

The decrease in the number of particles (24 to 3 for glucose), the longer time steps accessible with M3B (10 fs, instead of the 1 fs needed to integrate the high frequency C-H and O-H bond vibrations of the atomistic model), and the lack of long-range electrostatic interactions that would require costly Ewald sums makes M3B simulations of carbohydrates more than 7000 times faster than atomistic simulations.(6)

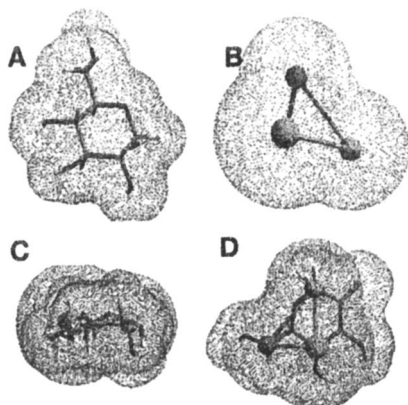


Figure 3. Connolly surfaces of α -glucose molecule: A: atomistic glucose, B: M3B glucose. Lower panels, C and D, show "side" and "top" view of the two superimposed models, with their respective Connolly surfaces. The solvent probe radius was 1.4 Å. (with permission from Ref. 6)

Conformations of $\alpha(1\rightarrow4)$ glucans

A remarkable feature of linear dextrans is their crystallization in helical structures, all of them left-handed. The existence of two different types of beads, B1 and B4, along the backbone chain of the M3B $\alpha(1\rightarrow4)$ glucans implies that left- and right-handed helical structures are distinct –not superimposable– M3B configurations. The handedness is indicated by the sign of the $141'4'$ torsions along the backbone of the chain. If all $141'4'$ torsions along the chain were zero the chain would be straight, if positive it would be right-handed, and if negative left-handed. Figure 4 shows the $141'4'$ distribution for α -maltose at 300 K from both coarse grain and atomistic simulations. The M3B succeeds in reproducing the relative energies of the atomistic dimer's conformers with negative and positive $141'4'$ torsion (Figure 4). Consistent with experimental findings, M3B and atomistic left-handed helices are more stable than right-handed ones. The energy difference per monomer between right-handed and left-handed single helices for DP24 $\alpha(1\rightarrow4)$ glucans in vacuum at 300 K is $12.8 \text{ kcal mol}^{-1}$ in the M3B model, in good agreement with the $7.8 \text{ kcal mol}^{-1}$ obtained with atomistic simulations of the same systems.(6)

The joint distributions for the intermonomer torsions $141'4'$ and $641'6'$ are shown in Figure 4 for maltose simulated with the atomistic and M3B models. These plots are analogous to Ramachandran plots showing the distribution of populations (or energy) as a function of the glycosidic torsional angles ϕ and ψ defined by $H1-C1-O-C4$ and $C1-O-C4-H4(12)$ (Figure 1). The $141'4'$ and $641'6'$ torsions are an alternative base to ϕ and ψ for the analysis of the conformational transitions in oligosaccharides.

The rise per monomer of the helix, h , in the M3B model is related to the value of the coarse grain torsion angle, $141'4'$, where the number of monomers per turn, n , depends also on the backbone angles $141'$ and $414'$. The configurational $[n, h]$ space available to the molecule in the M3B model depends, thus on the range of the backbone torsion and angles, and is tuned by the nonbonded interactions. Figure 5 shows two helical configurations of DP24 at 300 K in vacuum. The initial structure was constructed and minimized with M3B. The two helical structures differ in the number of turns (n) and pitch (p): the average n is 5.5 for structure A and slightly less than 7 for B. The average p is 8.3 Å for A and 7.2 Å for B. The experimental p of V_a and V_h amylose crystals(10,11) are 7.91 and 8.05 Å, respectively, and their $n=6$. The more compact B structure is stabilized by nonbond interactions, at the expense of increasing the torsional strain, that is lower for structure A.

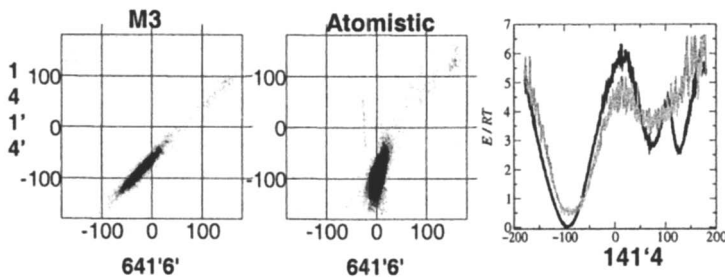


Figure 4. Left: Joint distribution of the coarse grain torsional angles $141'4'$ and $641'6'$ for M3B α -maltose at 300 K in vacuum. Center: Joint distribution of the torsional angles $141'4'$ (formed by $C1C4C1'C4'$) and $641'6'$ for atomistic α -maltose at 300 K in vacuum. Right: Potential of mean force as a function of the $141'4'$ torsional angle for atomistic (black) and M3B (gray) α -maltose at 300 K. (adapted from Ref. 6)

The local conformations of the chain are strongly affected by intermolecular interactions in concentrated solutions. Using M3B simulations, we find no evidence of preferential helical structures in DP12 chain in amorphous liquid mixtures with 6% and 12%wt of water. The $141'4'$ torsions between residues along the chain correspond as much to left-handed as to right-handed local twists. Equivalent results are found when comparing the potential of mean force for the torsion $141'4'$ in coarse grain DP2 in vacuum and in the bulk amorphous mixture with 10%wt water, both at 300 K: the bulk phase present the same two minima around -90° and 90° as observed in the right panel of Figure 3, but while $E[90^\circ]-E[-90^\circ]=1.8$ kcal/mole for DP2 in vacuum, it is only 0.45 kcal mol $^{-1}$ in the amorphous mixture. We expect a similar stabilization at the atomistic level,

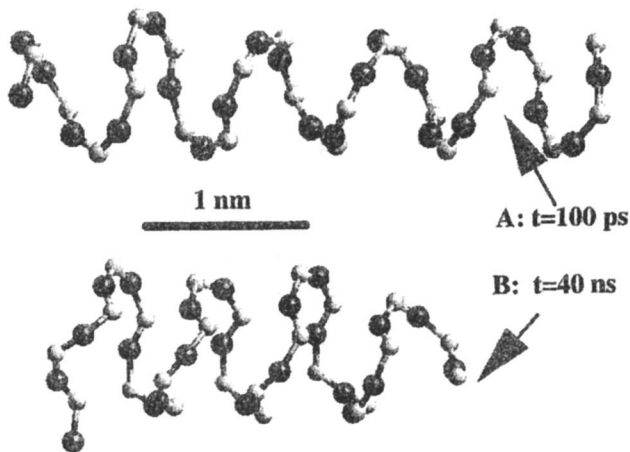


Figure 5: Helical structures obtained with M3B for of DP24 in vacuum at 300 K for two stages of the trajectory. Only the backbone (B1 and B4) beads are shown. (With permission from Ref. 6)

with the intermolecular hydrogen bonds (HB) competing with the intramolecular HB that favor the -90° local structure that gives rise to left-handed helices.

V_h -amylose is a crystalline form of hydrated amylose containing 4/3 of water molecules per glucose residue. The structure obtained from X-ray data indicates that each amylose chain forms a compact left-handed single helix, with water molecules inside the helix and in the inter-helical space.(10) We compared the experimental X-ray crystal structure of V_h -amylose(10) with the optimized structure obtained by the atomistic and M3B model, starting in both cases from the X-ray structure. The density agreement of the M3B model with the experiment is excellent; the coarse grain prediction is 100.6% of the experimental, while the atomistic prediction is 101.5% of the experimental. The agreement for the cell parameters is also very good, with predictions within 1-3% of the experimental value for the atomistic optimization and within 1-6% for the M3B model of the crystal.(6) The RMS between the M3B optimized structure of V_h -amylose and the experimental one is 2.3 Å, considerably less than the radius of the beads. The RMS between atomistic and experimental structures was 1.4 Å. The ability of the M3B model to reproduce the crystalline helical structure of amylose is surprising, considering that it was parameterized exclusively from amorphous glucans and that it has no directional nonbond interactions to stabilize the helical structures, nor the helical pattern of waters inside and outside the saccharide helices suggested by the X-ray data.

Amylose(9) and linear dextrans of $DP \geq 10$ (13) yield double stranded helical crystals. Although in principle double-stranded helices may exist in parallel or antiparallel configuration, the antiparallel configuration is not found for pure

$\alpha(1\rightarrow4)$ glucans. The X-ray spectra of starch polymorphs A and B show that the double stranded helices are almost twice as extended as the individual helices in the single stranded varieties.(9) While single stranded helices are stabilized by interactions between contiguous monomers and between monomers one turn apart, the structure of double helices is stabilized by the interactions between monomers of different chains. The M3B model is able to produce stable left-handed parallel and anti-parallel double helices for a DP12 oligomer in the crystalline phase at 300 K.(6) The two double helical structures have similar densities ($\rho_{\text{anti}}=1.448$ and $\rho_{\text{parallel}}=1.451$ g/cm³) and indistinguishable energies, in agreement with atomistic simulations of the same oligomer(6) and infinite double stranded helical structures.(14) These results suggest that the lack of anti-parallel double helices in nature would be related to the mechanism of biosynthesis and not an intrinsic instability of the structure.

Parallel and antiparallel DP12 duplexes are unstable in M3B vacuum simulations at 300 K, due to the unfavorable energy contribution of the coarse grain bending angles required for the extended configurations in the M3B model. Nevertheless, the vacuum double helices are local minima of the atomistic force field. This limitation of the M3B model is the price paid for selecting a smoothly varying energy function for the coarse grain force field that cannot reproduce the multiple shallow minima of the atomistic model's potential energy surface.

Water distribution in amorphous carbohydrates.

Water has a dramatic effect on the properties of carbohydrates. It acts as a strong plasticizer of carbohydrates, decreasing the glass transition temperature and the viscosity of the solutions. When the concentration of water is high, typically above 20%wt(15), ice separates from saccharide mixtures on cooling.

Atomistic MD simulations of binary water-saccharide solutions(2,7,8,16) indicate that water forms a spatially heterogeneous structure in the nanometer length-scale. Water molecules in carbohydrates form extended clusters that are percolated for water concentrations above ~18%wt (we defined two water molecules connected if they are closer than the first minimum of their radial distribution function). As expected from the similarity of the interactions between water and the carbohydrates, no signs of water pools or micelles are observed in concentrated carbohydrate solutions.(7,8) The results from a Zimm-Lundberg clustering function(17) analysis of experimental adsorption isotherms and densities of a maltodextrin syrup with variable water content(18) are in agreement with the structure predicted by the atomistic simulations.

It is interesting to consider whether a coarse grain model unable to form hydrogen bonds or any other directional nonbond interaction would succeed on predicting the structure of water in carbohydrates. To validate the predictions of

the M3B model, we studied the structure of water in glucose with 8-20%wt water using both atomistic and M3B simulations.(7) We observe that the M3B model is able to reproduce the structure of water in glucose, rendering clusters with the same percolation threshold as in the atomistic simulations, and predicting the same distribution of first water neighbors.(7) These distributions, shown in Figure 6, indicate the probability of finding a water molecule with a given water coordination, using the water connectivity criterion indicated above.

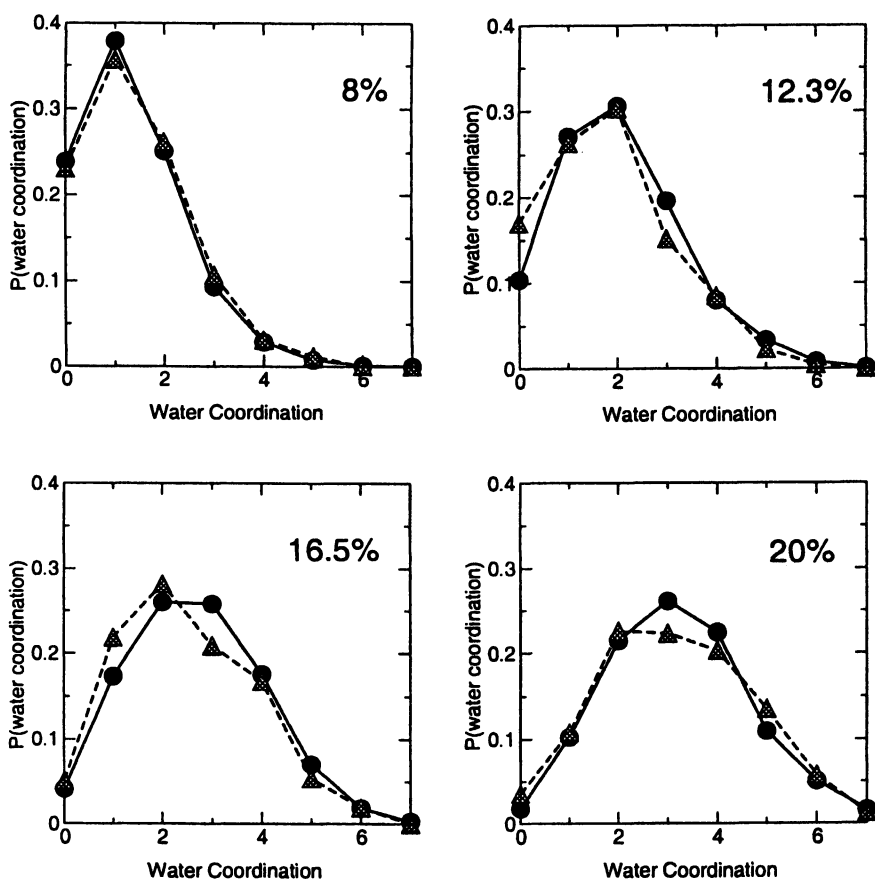


Figure 6. Water-water coordination distribution in glucose at different water contents at $T \sim 340$ K. The excellent agreement between the atomistic (triangles) and M3B (circles) results indicates that the packing of water in carbohydrates does not depend on the details of the hydrogen bond interactions.

The excellent agreement between the atomistic and coarse grain distribution for all the studied water contents, indicate that the distribution of water in carbohydrates does not depend on the details of the hydrogen bond interactions, but is rather determined by the packing of water and the carbohydrates, and the relative magnitude of the interaction energies. The M3B force field was parameterized to reproduce the cohesive energy of the pure components, and thus represents well the intermolecular interaction energies. The packing of water must depend critically on the description of the excluded volume (shape, size) and surface area of glucose as seen from a water molecule. This shape is very well reproduced by the M3B model, as can be seen from the Connolly surface of glucose in both representations, displayed in Figure 3.

Studies of diffusion of water in glucose using the M3B model reveal that the heterogeneous distribution of water in the carbohydrate matrix contribute to the observed nonexponential relaxation of water in supercooled carbohydrate solutions:(7) the more water neighbors a water has, the more mobile it is. The study of the dynamics of water in these supercooled conditions required simulation times on the order of a microsecond, practically unattainable in atomistic simulations with standard computing resources.

Glass transition and molecular mobility.

Pure carbohydrates and their low water content solutions do not crystallize if rapidly cooled, but instead they produce an amorphous solid phase, a glass. The glass transition temperature (T_g) signals the dynamical transition from the supercooled liquid to the glass phase. It can be localized from the change in the thermal expansion coefficient that accompanies the losing of degrees of freedom as the supercooled liquid becomes a solid. Using M3B coarse grain simulations, we have determined the T_g of pure and aqueous glucans at the inflection point of V vs. $1/T$. The details of the procedure are given elsewhere.(6) Table I summarizes the experimental and M3B glass transition temperatures for the studied systems. The coarse grain model succeeds not only in capturing the magnitude of the plasticisation effect of water, but also reproduces very well the absolute experimental T_g of these systems.

Table I. T_g predictions of the M3B model.

| Carbohydrate-water mixture | M3B T_g (K) | Experim. T_g (K) |
|---|---------------|--------------------|
| Glucose | 296±20 | 304 ^a |
| 12%wt water - glucose | 239±20 | 240 ^a |
| 12%wt water - DP12 $\alpha(1\rightarrow4)$ glucan | 341±17 | 325 ^b |

a) From Ref. (19). b) From Ref. (20)

Of the properties that have more impact on the processability of the materials and their stability under storage, the most important ones are the viscosity (η) and the molecular mobility (characterized by the diffusion coefficient D). The Stokes-Einstein (SE) equation $D=kT/6\pi\eta r$, relate viscosity and diffusivity in liquids. This relation breaks, however, in concentrated carbohydrate solutions-especially for the diffusion of water in the mixture, and when the system approaches the glass transition region.(21,22) At T_g , SE prediction for the water diffusion coefficient is $\sim 10^{-23} \text{ cm}^2 \text{ s}^{-1}$, while the experimental diffusivity of water in maltose at the glass transition was found to be nine orders of magnitude higher, $\sim 10^{-14} \text{ cm}^2 \text{ s}^{-1}$.(23) At a molecular level, this D implies that the characteristic time for water to diffuse a water diameter at T_g is $t \approx 1 \mu\text{s}$. These times are attainable with the M3B model. We studied the microscopic mechanism of water diffusion in 12%wt water-glucose in the supercooled liquid(6) and in the glass,(24) and find that water diffuses in glucose close and below the glass transition exclusively through jumps into neighbor positions occupied by water or glucose. The jumps into glucose positions couple the diffusion of water to the rotation of the sugar, also seen in this temperature range,(25)while the jumps into water positions couple the dynamics of water to water structure.(7,24) The diffusion coefficient for water in the glucose glass at 220 K estimated from the M3B molecular dynamics simulations(24) is $D=10^{-14} \text{ cm}^2 \text{ s}^{-1}$, the same order of magnitude observed in the experiments for the 10%wt water-maltose glass.(23)

Conclusions

Notwithstanding the extent and relevance of hydrogen bond interactions in carbohydrates and their aqueous mixtures, we have shown that thermodynamic, structural and dynamical properties of these systems can be reproduced with a coarse grain model, M3B that does not consider explicit hydrogen bonds or electrostatic interactions. Indeed, we represent water by a single particle and each glucose residue by just three beads.

Since the M3B energy expression consists of a sum of simple analytical energy functions of widespread use in atomistic simulations, is easy to implement using the machinery of atomistic Molecular Dynamics or Monte Carlo engines. The parameterization of the M3B force field was based on atomistic simulations of the pure components in a wide range of conditions (gas phase at different temperatures, amorphous bulk systems at different degrees of compression). The model was originally parameterized for water, glucose and $\alpha(1 \rightarrow 4)$ glucans,(6) but without any change in the mapping of the atomistic into the coarse grain particles, the parameterization can be extended to describe $\beta(1 \rightarrow 4)$, $(\alpha, \beta) 1 \rightarrow 6$, and $(\alpha, \beta) 1 \rightarrow 1$ glucans.

The success of M3B on describing the conformations and relative energies of helical structures of dextrans is based on a careful parameterization of the

coarse grain torsions. This torsional term –usually absent in coarse grain models of polymers- is key to reproduce the relative energies of the conformers.

The agreement on the microscopic water structure predicted by M3B and the atomistic model is indicative that in an amorphous environment where *all* the atomistic interactions are directional, such as the hydrogen bonds, their contribution to long-range interactions can be captured with a simplified potential that represents the short range repulsive effect (excluded volume, shape) and the overall medium range attraction energy. The good description in M3B of the shape of the glucose molecule (Figure 3) as would be perceived by a water molecule was surely decisive in attaining a good description of the packing of water in the sugar matrix.

The dynamical properties of carbohydrates solutions –difficult to study with atomistic simulations- are well predicted by the coarse grain model. M3B predicts the glass transition temperatures of carbohydrates and the plasticization effect of water in good agreement with experiment.(6) The coarse grain model shed light into the mechanism of water diffusion in supercooled carbohydrates(7) and into the microscopic origin of the decoupling of water and sugar translation that allows water diffusion even below the glass transition temperature.(24)

The coarse grain simulations of carbohydrates are ~7000 times faster than atomistic simulations of the same systems, opening the possibility to study systems with complex formulations (such as starch hydrolisates) for times on the order of a microsecond. The increase in the time scale is crucial for studies of the molecular mobility in these extremely viscous systems. Note that the conformational barriers of dextrans (Figure 4) imply that the characteristic times for the segmental motion of these molecules should be above a nanosecond even in dilute solutions; making it difficult to equilibrate the conformers with standard atomistic MD simulations.

The accurate and non-iterative reconstruction of the atomistic coordinates from the M3B model enables the use of the coarse grain model for the equilibration of systems that afterwards are mapped into the atomistic model. In this atomistic level shorter simulations can be performed to extract properties that require full atomistic detail, for example the interaction with other biopolymers.

References

1. Caffarena, E. R.; Grigera, J. R. *Carbohydrate Research* **1997**, *300*, 51-57.
2. Roberts, C. J.; Debenedetti, P. G. *Journal of Physical Chemistry B* **1999**, *103*, 7308-7318.
3. Ekdawi-Sever, N. C.; Conrad, P. B.; de Pablo, J. J. *Journal of Physical Chemistry A* **2001**, *105*, 734-742.

4. Ekdawi-Sever, N.; de Pablo, J. J.; Feick, E.; von Meerwall, E. *Journal of Physical Chemistry A* **2003**, *107*, 936-943.
5. Muller-Plathe, F. *Chemphyschem* **2002**, *3*, 754-769.
6. Molinero, V.; Goddard III, W. A. *J. Phys. Chem. B* **2004**, *108*, 1414-1427.
7. Molinero, V.; Cagin, T.; Goddard III, W. A. *J. Phys. Chem.* **2004**, *108*, 3699-3712.
8. Molinero, V.; Cagin, T.; Goddard III, W. A. *Chem. Phys. Lett.* **2003**, *377*, 469-474.
9. Imberty, A.; Chanzy, H.; Perez, S.; Buleon, A.; Tran, V. *Journal of Molecular Biology* **1988**, *201*, 365-378.
10. Brisson, J.; Chanzy, H.; Winter, W. T. *International Journal of Biological Macromolecules* **1991**, *13*, 31-39.
11. Winter, W. T.; Sarko, A. *Biopolymers* **1974**, *13*, 1447-1460.
12. Rao, V. S. R. *Conformation of carbohydrates*; Harwood Academic Publishers: Amsterdam, The Netherlands, 1998.
13. Gidley, M. J.; Bulpin, P. V. *Carbohydrate Research* **1987**, *161*, 291-300.
14. Schulz, W.; Sklenar, H.; Hinrichs, W.; Saenger, W. *Biopolymers* **1993**, *33*, 363-375.
15. Roos, Y.; Karel, M. *Cryo-Letters* **1991**, *12*, 367-376.
16. Caffarena, E. R.; Grigera, J. R. *Carbohydrate Research* **1999**, *315*, 63-69.
17. Zimm, B. H.; Lundberg, J. L. *Journal of Physical Chemistry* **1956**, *60*, 425-428.
18. Kilburn, D.; Claude, J.; Mezzenga, R.; Dlubek, G.; Alam, A.; Ubbink, J. *Journal of Physical Chemistry B* **2004**, *108*, 12436-12441.
19. Roos, Y. *Carbohydrate Research* **1993**, *238*, 39-48.
20. Orford, P. D.; Parker, R.; Ring, S. G.; Smith, A. C. *International Journal of Biological Macromolecules* **1989**, *11*, 91-96.
21. Champion, D.; Hervet, H.; Blond, G.; LeMeste, M.; Simatos, D. *Journal of Physical Chemistry B* **1997**, *101*, 10674-10679.
22. Rampp, M.; Buttersack, C.; Ludemann, H. D. *Carbohydrate Research* **2000**, *328*, 561-572.
23. Parker, R.; Ring, S. G. *Carbohydrate Research* **1995**, *273*, 147-155.
24. Molinero, V.; Goddard III, W. A. *Physical Review Letters*, in press.
25. Wachner, A. M.; Jeffrey, K. R. *Journal of Chemical Physics* **1999**, *111*, 10611-10616.

Chapter 16

Molecular Modeling of Retaining Glycosyltransferases

Igor Tvaroška

Institute of Chemistry, Slovak Academy of Sciences, 845 38 Bratislava,
Slovakia

The retaining glycosyltransferases catalyze the transfer of a glycosyl moiety from a specific donor to the acceptor with retaining the configuration at the anomeric centre. The catalytic mechanism of retaining glycosyltransferases is not yet completely understood but two possible mechanisms are considered. The first is a double-displacement mechanism via a covalent glycosyl-enzyme intermediate and the second is the one step S_Ni -like mechanism. Non-empirical quantum mechanical methods were employed to investigate a variety of reaction pathways for these mechanisms using structural models containing from 74 to 136 atoms. Different reaction mechanisms were characterized by means of potential energy surfaces calculated using non-empirical quantum chemical methods at different levels of theory. For both proposed reaction mechanisms the energetically favorable reaction pathways were indicated and the structures of all intermediates and transition state models were determined. These calculations provided new insights on the catalytic mechanism of retaining glycosyltransferases that are not available experimentally. Moreover, these results provide a rational basis for the design of new classes of glycosyltransferases inhibitors.

Introduction

Glycosyltransferases catalyze the biosynthesis of complex carbohydrates and glycoconjugates of considerable structural diversity that perform a variety of essential functions such as signaling, cell-cell communication, energy storage and cell-wall structure (1-6). These enzymes are responsible for the glycoside bond formation transferring a glycosyl residue from activated donor molecules to the hydroxyl group of specific acceptor molecules. Two stereochemical outcomes are possible for the glycosylation reaction: the configuration at the anomeric carbon C-1 of the product can either be retained or inverted with respect to the donor molecule. The overall reaction catalyzed by retaining glycosyltransferases is illustrated in Figure 1. Glycosyltransferases have been classified into 65 families on the basis of sequence analysis (7), of which about 17 led to the retention of the configuration at the anomeric carbon C-1. The mechanism of inverting glycosyltransferases is now rather well documented experimentally and theoretically, and it is thought to be a single nucleophilic displacement of the nucleotide by a hydroxyl group of the acceptor (8-11). On the other side, many aspects of the mechanism of retaining glycosyltransferases remain unclear and therefore raise a lot of interest (12).

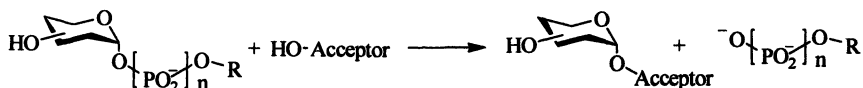


Figure 1. Schematic representation of overall reactions catalyzed by retaining glycosyltransferases. R represents a nucleoside.

The X-ray structures of retaining glycosyltransferases available to this date (13-19) suggest a possibility of two different mechanisms. In the first, the reaction catalyzed by retaining glycosyltransferases proceeds, by analogy with retaining glycosidases, via a double-displacement mechanism. In the second, the catalytic reaction proceeds through a single front-side displacement

reaction, also known as a S_Ni mechanism and is based on the crystal structure (13) of the lipopolysaccharyl α -1,4-galactosyltransferase C (LgtC). To date reported three-dimensional structures of glycosyltransferases have revealed an interesting difference from those of retaining glycosidases as in some of them (14-17) only a catalytic nucleophile has been observed. Moreover, in two structures (13,18) only an amino acid residue, which is unlikely to function as a nucleophile was found in the active site. This raises three questions to be answered. First, how can possibly the first step a double-displacement mechanism proceed without the contribution of a catalytic acid; second, how can the glycosyl transfer proceeds without a direct involvement of the enzyme amino acids via the S_Ni mechanism; and third, what is the most likely candidate that can act as a general base to facilitate the proton departure from the attacking group? Clearly, more data are needed to fully understand the mechanism of retaining glycosyltransferases and high-level quantum mechanical calculations were used to gain some insights into characteristics of the enzymatic reaction catalyzed by these enzymes (20,21). In these studies the calculations were carried out using the Jaguar program (22). The optimization of the geometry was performed using different non-empirical methods and energy was calculated using the B3LYP density functional method (23) with the basis sets up to the B3LYP/6-311++G** level. Theoretical analysis provides access to details on an atomic level of the mechanism that are not available from experiment and can lead to a deeper understanding of catalysis. The purpose of this paper is to summarize recent developments in molecular modeling of the catalytic mechanism of retaining glycosyltransferases.

Double displacement mechanism

Reactions catalyzed by retaining glycosidases are known to proceed via a double displacement mechanism involving the formation and the subsequent breakdown of a covalent glycosyl-enzyme intermediate. The analogy with retaining glycosidases has led to an assumption that retaining nucleoside diphosphate-utilizing glycosyltransferases proceed by the similar two-steps mechanism that is shown in Figure 2. In the first step, the enzyme nucleophile attacks the anomeric carbon C-1 of the donor resulting in a formation of a covalent glycosyl-enzyme intermediate. A second amino acid is involved in this step to facilitate the departure of the leaving group. The configuration at the anomeric carbon C-1 is inverted to the equatorial orientation in a covalent glycosyl-enzyme intermediate. In the second step, the covalent glycosyl-enzyme intermediate is broken by a direct displacement of the C-1-O glycosyl-enzyme bond with the reactive hydroxyl group of the acceptor and the configuration at the C-1 is inverted from the equatorial back to the axial. In this

the essential parts of the catalytic acid A and nucleophile B^- represented by acetic acid and acetate molecules. This model consisted of 74 atoms and had an overall charge of minus one.

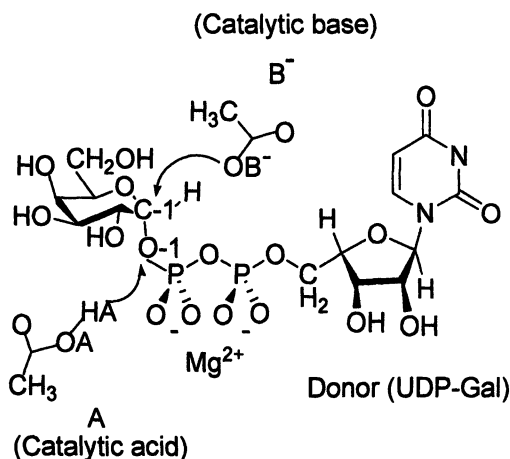


Figure 3. Schematic representation of the structural model for the examination of the first step of the double displacement mechanism of a retaining galactosyltransferase.

The first step of the investigated mechanism characterizes the attack of the nucleophile oxygen O_B on the anomeric carbon C-1 of the UDP-Gal. The proton transfer from a catalytic acid A to the glycosidic oxygen O-1 can be, in principle, used by an enzyme to facilitate a departure of the UDP leaving group. The two-dimensional contour map describing PES for various reaction pathways of this step is given in Figure 4. The $r(C-1-O_B)$ distances plotted along the x-axis of the contour map describe the nucleophilic attack of the catalytic base B^- on the anomeric carbon C-1 of galactopyranose. The y-axis defines the $r(O-1-H_A)$ distances and describes the proton (H_A) transfer from the catalytic acid A to the glycosidic oxygen O-1. The reaction proceeds from reactants (R) to the covalent glycosyl-enzyme intermediates (INT1 or INT3). These two intermediates differ only by the proton transferred from the catalytic acid A to the glycosidic oxygen O-1 and thus distinguish between the mechanisms that involve either one or two catalytic amino acids. The formation of the covalent glycosyl-enzyme intermediate INT1 *via* the reaction pathway $R \rightarrow TS1 \rightarrow INT1$ corresponds to the mechanism involving only one catalytic base in the reaction mechanism. In the case of a mechanism involving two catalytic residues, the $R \rightarrow TS3 \rightarrow INT2 \rightarrow TS4 \rightarrow INT3$ pathway leads to the

covalent glycosyl-enzyme intermediate INT3. Comparison of the energy barriers required to proceed along these two pathways (Figure 4) showed that the pathway starting with the nucleophilic attack is less energy demanding as compared to the pathway involving the proton transfer; 16.5 kcal/mol vs. 37.9 kcal/mol at the B3LYP/6-311++G**// B3LYP/6-31G* level. The activation energy of 16.5 kcal/mol calculated for the nucleophilic attack in the first step is slightly higher compared to the value of 13.4 kcal/mol calculated for the inverting *N*-acetylglucosaminyltransferases (24,25). This difference is probably a consequence of differences in the nucleophilic character of the attacking groups.

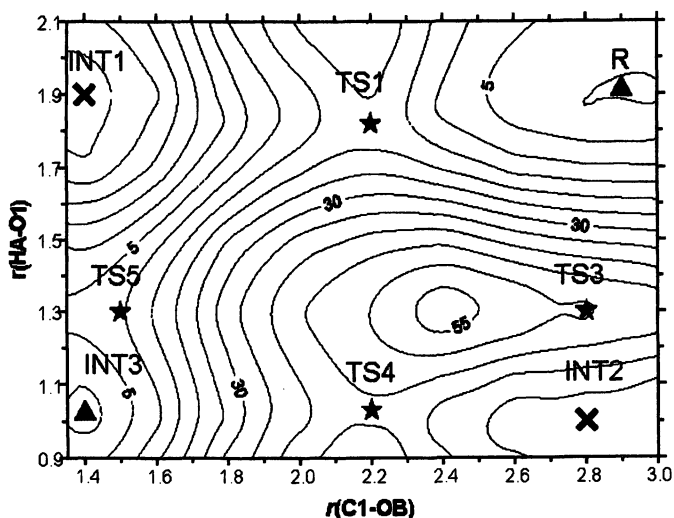


Figure 4. The calculated potential energy surface for the first step of the double displacement mechanism of a retaining galactosyltransferase (20). (Reproduced from reference 20. Copyright 2003.)

Similarly as for the previously calculated reaction mechanism of inverting *N*-acetylglucosaminyltransferases (9,11), these results suggested that the proton transfer from a catalytic acid into the glycosidic oxygen is not required for the breaking of the C-1-O-1 bond. Therefore, in contrast to glycosidases, it seems that glycosyltransferases do not need a second catalytic residue in their active site for the reaction to proceed. In fact, none of the solved crystal structure of glycosyltransferases, either inverting or retaining has so far revealed the presence of a second amino acid in the active site of the enzyme. One

explanation resides probably in the significantly better leaving group character of a nucleotide compared to a glycosyl residue.

Displacement step

In the second step, the sugar-acceptor displaces the enzymic carboxylate from the covalently bound glycosyl-enzyme intermediate. The nucleophilic attack of the acceptor oxygen at the anomeric carbon C-1 of the intermediate requires a general base to deprotonate the acceptor hydroxyl (O_a-H_a). Several possible functional groups were considered and investigated for this function. For enzymes without a second catalytic residue in the active site, one can consider the oxygen of UDP and the oxygen of the catalytic acid B as potential candidates for this function. The mechanism of this step involving the proton transfer to the oxygen atoms of the diphosphate group was evaluated using the structural model shown in Figure 5.

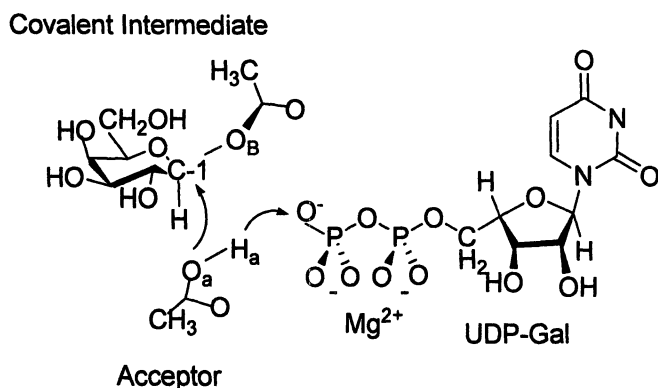


Figure 5. Schematic representation of the structural model for the examination of the second step of the double displacement mechanism of a retaining galactosyltransferase.

The reaction energy surface was calculated as a function of the reaction coordinates describing the proton transfer to the phosphate group $r(H_a-O)$ and the nucleophilic attack $r(O_a-C-1)$, respectively. The two-dimensional contour map representing the calculated PES is shown in Figure 6. The x-axis represents the nucleophilic attack on the anomeric carbon of the INT1. The y-axis describes the proton transfer from the acceptor hydroxyl to the UDP oxygen. The map demonstrates the reaction pathway in which the formation of

the new glycosidic C-1-O_a bond and the proton transfer from the acceptor hydroxyl to the UDP occur simultaneously. The calculated barrier for this reaction process at the B3LYP/6-311++G**//B3LYP/6-31G* level is about 8.5 kcal/mol.

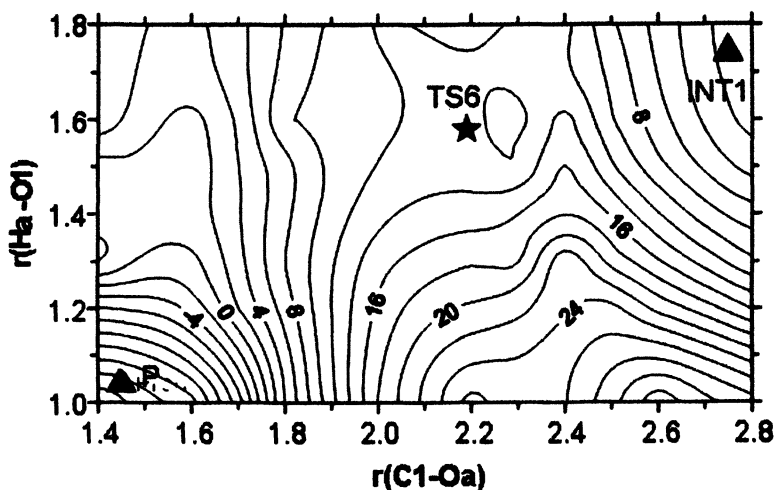


Figure 6. The calculated potential energy surface for the second step of the double displacement mechanism of a retaining galactosyltransferase with the UDP functioning as a general base (20). (Reproduced from reference 20. Copyright 2003.)

The mechanism using the catalytic acid as the potential nucleophile is quite interesting, however, the calculations clearly demonstrated that this mechanism is very unlikely. The energy barrier calculated for this pathway was over 65 kcal/mol what is considerably higher than the energy barrier calculated for the processes involving UDP as the potential general base and denotes a very unfavorable mechanism. Furthermore, the calculations predicted that this mechanism led to a glycoside product with the inverted configuration at the C-1 carbon, rather than the retained configuration expected by this type of enzymes. All these results support the participation of UDP as the catalytic base in the displacement step of a double-displacement mechanism of retaining glycosyltransferases. As a result, the following reaction mechanism illustrated in Figure 7 was proposed.

The first step consists of the nucleophilic attack of the enzyme nucleophile on the anomeric carbon and as a consequence the covalent glycosyl-enzyme

intermediate (INT1) with the inverted (β -) configuration at the C-1 carbon is formed. The first step proceeds via the oxocarbenium ion-like transition state

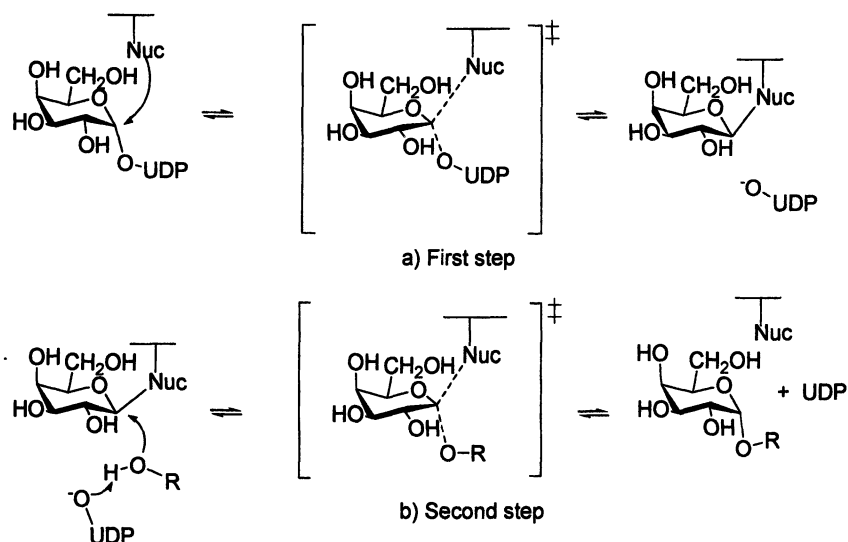


Figure 7. Schematic representation of the predicted double displacement mechanism for retaining glycosyltransferases from the calculated potential energy surfaces for various mechanisms of retaining galactosyltransferases. *R* represents a nucleoside.

(TS1) with the calculated barrier of 16.5 kcal/mol. In the next step, the acceptor hydroxyl attacks the INT1 while the leaving the UDP plays the role of a general base deprotonating the attacking hydroxyl. This step proceeds in a quasi-concerted manner with the second inversion at the anomeric center with the resulting glycoside having α -configuration. The calculated barrier for this step is about 8.5 kcal/mol. The overall reaction barrier 18 kcal/mol calculated for the proposed mechanism lies in the range of experimentally observed barriers for glycosyltransferases of 15-25 kcal/mol (24). This type of mechanism has been proposed recently for the retaining α -1,3-Galactosyltransferase (14,25). Kinetics measurements of some retaining glycosyltransferases indicated that UDP is not released before the transfer of the sugar to the acceptor (26-28). These data are consistent with this type of mechanism. Interestingly, a proton transfer of this type from the hydroxyl of the acceptor to a phosphate group has also been proposed for the reverse reaction of maltodextrin phosphorylase (29).

The S_{Ni} -like Mechanism

The second mechanism proposed for retaining glycosyltransferases was based on the X-ray structure of the lipopolysaccharyl α -1,4-galactosyltransferase C (LgtC) from *Neisseria meningitidis* (13). Three-dimensional structure of the LgtC complex with the stable donor and acceptor substrates revealed that in a catalytic site glutamine 189 was the only amino acid appropriately placed to function as a nucleophile. This amino acid is unlikely to form a covalent intermediate. Although several other electronegative groups in the vicinity of the UDP-Gal were considered as possible catalytic nucleophile (13,27) but the analysis of the structural and mutagenesis results was inconclusive. These results were pointing toward another mechanism that might not require any participation of a catalytic residue and led to the suggestion (13) that this enzyme might rather proceed through a single front-side displacement reaction, also known as the S_{Ni} mechanism and is illustrated in Figure 8. Recently, based on a comparison of the LgtC (13) and EXTL2 (18) crystal structures, the S_{Ni} -like mechanism for the retaining glycosyl transfer was also suggested (30) for the α -1,4-*N*-acetylglucosaminyltransferase EXTL2.

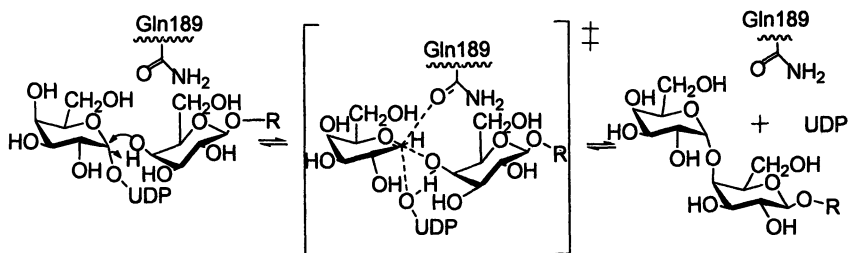


Figure 8. Schematic representation of the proposed S_{Ni} catalytic mechanism for the retaining LgtC. R represents UDP.

The LgtC catalyzes the transfer of a galactosyl residue from UDP-Gal to a galactose of the terminal lactose moiety and the catalytic mechanism of the LgtC was explored using a reaction site model generated from the crystal structure of the of LgtC determined in the presence of both donor and acceptor stable substrate analogues, UDP 2-deoxy-2-fluoro-galactose and 4'-deoxylactose (13). A schematic representation of the reaction site model is shown in Figure 9. This model contained 136 atoms consisting of the complete sugar-donor molecule, UDP-Gal; a galactose derivative representing the oligosaccharide-acceptor; the divalent metal cofactor Mg^{2+} fully coordinated by

aspartate D103, aspartate D105, and histidine H244, as found in the X-ray structure; a portion of glutamine Q189 presumed to be the nucleophile in one of assumed mechanisms; and the essential fragment of aspartate D8 interacting with the uridine part of the donor. Various reaction hypotheses for the galactosyl transfer were explored with this reaction model using non-empirical computational methods (21).

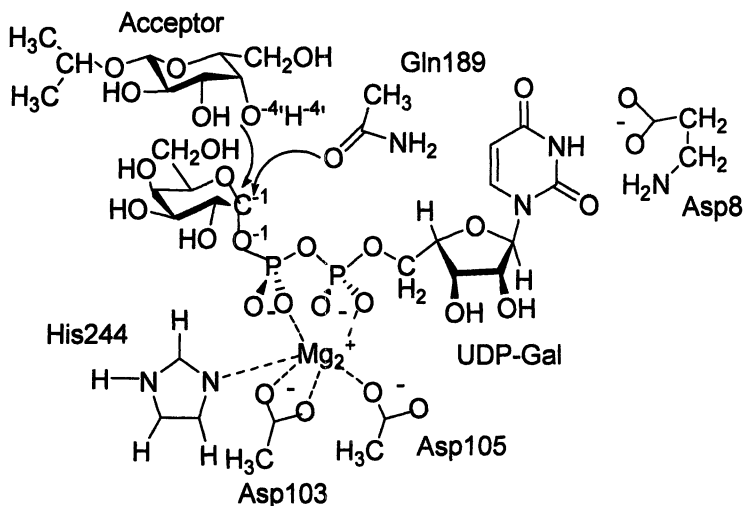


Figure 9. Schematic representation of the structural model used to investigate the reaction catalyzed by LgtC.

A two-dimensional map describing PES for various catalytic mechanism calculated at the B3LYP/6-31G* level is shown in Figure 10. The $r(\text{C}-\text{O}-4')$ distances plotted along the x-axis of the contour map describe the nucleophilic attack of the acceptor oxygen O-4' on the anomeric carbon of galactopyranose. The y-axis characterizes the $r(\text{C}-1-\text{O}_{\text{Gln}})$ distances that describe the formation of a glycosyl-enzyme intermediate. Each calculated point on the map corresponds to the optimized structure and the arrangement of the model for the given pair of values of the $r(\text{C}-1-\text{O}-4')$ and $r(\text{C}-1-\text{O}_{\text{Gln}})$ distances, respectively. During the optimization, all geometrical parameters were relaxed with the exception of those defining the location of the α -carbon of the enzyme amino acids. The map shows two pathways along axes. The pathway parallel to the y-axis characterizes the nucleophilic attack of the Gln189 oxygen on the anomeric C-1 of UDP-Gal and should represent the formation of the covalent

galactosyl-enzyme intermediate and could be considered as the first step of the double-displacement reaction mechanism. It turned out that the intermediate does not correspond to the real minimum. A refinement of this local minimum

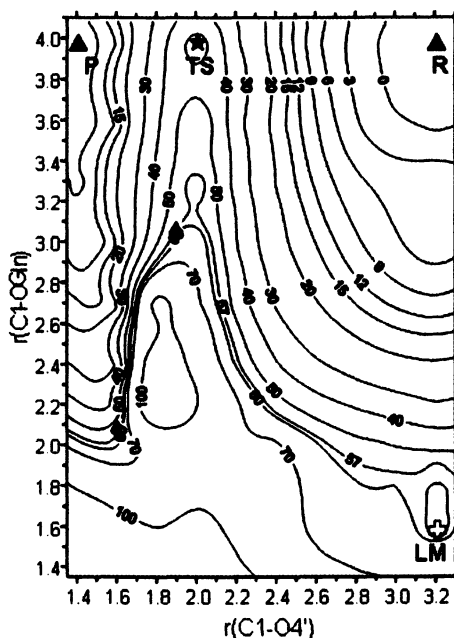


Figure 10. The B3LYP/6-31G* potential energy surface for the S_{Ni} mechanism of the LgtC calculated as a function of the $r_{C1-O4'}$ and $r_{C1-OGln}$ distances (21). (Reproduced from reference 21. Copyright 2004.)

(LM) without constraints on the $r(C-1-O-4')$ and $r(C-1-O_{Gln})$ coordinates led to the covalent galactosyl-LgtC structure with a relative energy about 33 kcal/mol higher compared to the reactants. Moreover, the analysis of the geometrical parameters showed that the acceptor residue moved during the optimization from the crystallographic orientation. It appeared that the acceptor orientation is a very important structural feature for the outcome of the catalytic reaction. In the crystal structure, this highly ordered orientation of the galactosyl residue in the active site is stabilized by the hydrogen bonds between O-3, O-4, and O-6 oxygens and side chains of the enzyme (13). Indeed, any

attempt to follow the nucleophilic attack of O-4', as a function of the $r(\text{C-1-O-4}')$ from the orientation in this local minimum led to a product either with an open galactose ring or with a β -glycosidic linkage. These findings are likely the result of the structural constraints of carbohydrate-binding domains occurring in the LgtC. These results also suggest that Gln189 does not play the role of an enzymatic nucleophile and might explain why all attempts to trap the covalent intermediate with the Gln189 were unsuccessful (27). However, a covalent intermediate with Asp190 that is located about 9 Å from the anomeric carbon C-1 was recently observed (31) for the Gln198Glu mutant. This revived discussion about the catalytic mechanism of LgtC but more experimental evidence is clearly needed to fully understand catalytic mechanism of retaining glycosyltransferases.

The second pathway described by the reaction coordinate plotted along the x-axis of the contour map proceeds in one step. The $\text{R} \rightarrow \text{TS} \rightarrow \text{P}$ pathway involves the nucleophilic attack of the acceptor hydroxyl O-4' onto the anomeric carbon C-1 of UDP-Gal and the formation of a new α -(1 \rightarrow 4) linkage with the H4' proton transferred from the attacking acceptor hydroxyl to the leaving UDP oxygen O-1. The reaction barrier of 31 kcal/mol was estimated for this pathway using the B3LYP/6-311++G**//B3LYP/6-31G* theory. Amongst the three stationary points observed along the evaluated pathway, the transition state represents the structure with the most separated charges. This character should be reflected in the most pronounced stabilization by surrounding environment effects. To estimate this effect qualitatively the electrostatic environment was treated as a dielectric continuum using the procedure implemented in Jaguar (22,32). Calculations were carried out for two different environment represented by cyclohexane ($\epsilon = 2$) and water ($\epsilon = 78$) using the 6-31G** basis set. Treated this way, the reaction barrier was decreased by 7 kcal/mol and 21 kcal/mol, respectively. After the correction for the electrostatic environment effects, the reaction barrier could be estimated around 10 - 24 kcal/mol. This should be compared to the experimental value of the catalysis of galactose transfer by the bacterial LgtC, which has been kinetically estimated (13,27) to be between of 12 and 14 kcal/mol. However, it is clear that these values are very approximate and to estimate effects of both the protein and solvent more realistically would require inclusion of the whole enzyme and surrounding water in the calculations.

The calculations predicted that preferred reaction pathway for the LgtC corresponds to the proposed $\text{S}_{\text{N}}\text{i}$ -like mechanism (13) in which the nucleophile hydroxyl from the acceptor attacks the anomeric carbon of the galactopyranose residue of the UDP-Gal donor from a side of the leaving group with the simultaneous proton transfer to a phosphate oxygen. This one-step mechanism differs from the other mechanisms proposed for glycosyltransferases in that it does not require the presence of any catalytic amino acid in the active site.

From the calculations, it is apparent that the active site constraints imposed by the enzyme on the substrates change the conformation of the donor UDP-Gal to the one that is suitable for the nucleophilic attack from the leaving group side.

Transition states

Structures of transition state models represent important outcomes of the calculated potential energy surfaces. A comparison of the transition state structures associated with the preferred reaction pathways for both the catalytic mechanisms revealed some interesting features. Although the structure of transition state models for these pathways differ, there are some general characteristics that are typical for all the transition state models of retaining glycosyltransferases: a) the ring conformation resembles a deformed chair conformation with the oxo-carbenium character at the anomeric carbon; b) the C-1-O-1 distance is longer than the standard C-O bond length; c) a new α -glycosidic bond is being created with the bond length longer than the standard bond length. The calculated features of the transition state models are illustrated by the values of selected bond lengths and angles in Table 1 and schematically shown in Figure 11.

Table 1. The B3LYP/6-31G* calculated selected geometric parameters of the transition state structures of the preferred mechanisms

| | <i>Parameter</i> | <i>TS1</i> | <i>TS2</i> | <i>TS</i> |
|----------------|----------------------|------------|------------|-----------|
| Bond Lengths | C-1-O _{Nuc} | 2.200 | 2.190 | 2.341 |
| | C-1-O-1 | 1.649 | 2.279 | 2.662 |
| | C-1-O-5 | 1.366 | 1.3.22 | 1.420 |
| Bond angles | C-1-O-5-C-5 | 111.9 | 121.3 | 121.2 |
| | C-2-C-1-O-5 | 112.2 | 112.6 | 123.7 |
| Torsion angles | C-2-C-1-O-5-C-5 | -53.2 | -52.1 | -22.2 |
| | C-1-O-5-C-5-C-4 | 47.6 | 48.7 | 40.3 |

The first step of the double displacement mechanism involves the formation of a covalently bound glycosyl-enzyme intermediate via transition state TS1. In this transition state model, the ring conformation can be described as a distorted ⁴C₁ chair and arrangement of substrates undergoes significant geometrical changes compared to the ground state structure. The C-1-O_B reaction coordinate has the value of 2.20 Å in the TS1, the C-1-O-1 bond length increases slightly by 0.15 Å from 1.493 Å to 1.640 Å, and the C-1-O-5

shortens from 1.382 Å to 1.366 Å. The structure of the transition state model TS2 for the second step of this mechanism is similar although a larger variation in the C-1–O-1 bond length as shown in Table 1. Analysis of geometrical changes revealed that the distance between the anomeric carbon and the leaving group elongates from 1.493 Å in ground state to 2.279 Å in the TS2 as the distance between the anomeric oxygen and the attacking catalytic base decreases to 2.190 Å.

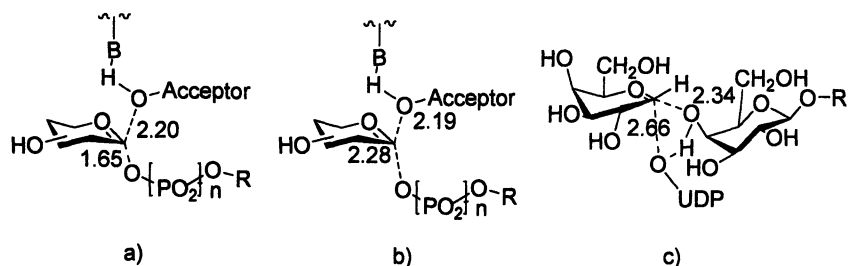


Figure 11. Schematic representation of the transition states for the (a) the first step, (b) the second step, of the double-displacement, and (c) S_Ni mechanism of retaining glycosyltransferases.

The analysis of the transition state structure for the S_Ni mechanism clearly showed that the geometry of the transition state model is unique and differs from the transition state structures determined for others inverting or retaining glycosyltransferases (9,11,20). This structure is characterized by the C-1–O-1 and C-1–O-4' distances of about 2.66 Å and 2.34 Å, which indicate a weak bond order to both the nucleophile and the leaving group. The main difference is that both the nucleophilic attack by the acceptor oxygen O4' and the leaving group departure occur on the same face of the transferred galactose. The ring conformation of the transferred galactose is a distorted 4E envelope. Another interesting and unique feature of this transition state is the location of the transferred proton. The nucleophile proton is sandwiched between the oxygen from the phosphate and the nucleophile. The H-4' proton is located at the O-4'–H-4' and O-1–H-4' distances of about 1.09 Å and 1.38 Å, respectively. The distance between the O-4' and O-1 oxygen in the TS is about 2.4 Å. This resembles the short distances found for low barrier hydrogen bonds, which have an energy of 15-20 kcal/mol and have been postulated to be important in enzyme catalysis (33). The estimated stabilization of the TS by this hydrogen bond is approximately 19 kcal/mol at the B3LYP/6-31G* level.

In conclusion, it is clearly evident that the “entire” transition state for retaining glycosyltransferases consists of a deformed sugar-nucleotide and acceptor oligosaccharide, metal cofactor, and amino acids involved in the catalytic reaction or in binding substrates. The resulting geometric and electronic models of the transition states provide blueprint for the design of the transition state analog inhibitors. The calculated transition state structures also emphasize that the acceptor contributes substantially to the nature of the transition state, and the development of stable analogues of the transition state as potent inhibitors of glycosyltransferases should therefore take these characteristics into account.

Acknowledgments

This work was supported by the grants from the Mizutani Foundation for Glycoscience No. 040013 and from the Science and Technology Assistance Agency under contract No. APVT-51-04204.

References

1. Lis, H.; Sharon, N. *Eur. J. Biochem.* **1993**, *218*, 1-27.
2. Varki, A. *Glycobiology*, **1993**, *3*, 97-130.
3. Kleene, R.; Berger, E. G. *Biochim. Biophys. Acta* **1993**, *1154*, 283-325.
4. Gilbeaut, D. M. *Plant Physiol. Biochem.* **2000**, *38*, 69-80.
5. Dennis, J. W.; Granowsky, M.; Warren, C. *BioEssays* **1999**, *21*, 412-421.
6. Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. *Science* **2001**, *291*, 2370-2376.
7. Coutinho, P. M.; Deleury, E.; Davis, G. J.; Henrissat, B. *J. Mol. Biol.* **2003**, *328*, 307-317.
8. Unligil, U. M.; Rini, J. M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 510-517.
9. Tvaroška, I.; André, I.; Carver, J. P. *J. Am. Chem. Soc.* **2000**, *122*, 8762-8776.
10. Tarbouriech, N.; Charnock, S. J.; Davies, G. J. *J. Mol. Biol.* **2001**, *314*, 655-661.
11. Tvaroška, I.; André, I.; Carver, J. P. *Glycobiology* **2003**, *13*, 559-566.
12. Lairson, L. L.; Withers, S. G. *Chem. Commun.* **2004**, 2243-2248.

13. Persson, K.; Ly, H. D.; Dieckelmann, M.; Wakarchuk, W. W.; Withers, S. G.; Strynadka, N. C. *J. Nat. Struct. Biol.* **2000**, *8*, 166-175.
14. Gastinel, L. N.; Bignon, C.; Misra, A. K.; Hindsgaul, O.; Shaper, J. H.; Joziassse, D. H. *EMBO J.* **2001**, *20*, 638-649.
15. Patenaude, S. I.; Seto, N. O.; Borisova, S. N.; Szpacenko, A.; Marcus, S.L.; Palcic, M. M.; Evans, S. V. *Nat. Struct. Biol.* **2002**, *9*, 685-690.
16. Gibbons, B. J.; Roach, O. J.; Hurley, T. D. *J. Mol. Biol.* **2002**, *319*, 463-477.
17. Gibson, R. P.; Turkenburg, J. P.; Charnock, S. J.; Lloyd, R.; Davies, G. *J. Chem Biol.* **2002**, *9*, 1337-1346.
18. Pedersen, L. C.; Dong, J.; Taniguchi, F.; Kitagawa, H.; Krahn, J. M.; Pedersen, L. G.; Sugahara, K.; Negishi, M. *J. Biol. Chem.* **2003**, *278*, 14420-14428.
19. Lairson, L. L.; Chiu, C. P. C.; Ly, H. D.; He, S.; Wakarchuk, W. W.; Strynadka, N. C. J.; Withers, S. G. *J. Biol. Chem.* **2004**, *279*, 28339-28344.
20. André, I.; Tvaroška, I.; Carver, J. P. *Carbohydr. Res.* **2003**, *338*, 865-877.
21. Tvaroška, I. *Carbohydr. Res.* **2004**, *339*, 1007-1014.
22. Jaguar 3.5, S., Inc., Portland, OR, 1998.
23. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
24. Seto, N. O.; Compston, C. A.; Evans, S. V.; Bundle, D. R.; Narang, S. A.; Palcic, M. M. *Eur. J. Biochem.* **1999**, *259*, 770-775.
25. Boix, E.; Zhang, Y.; Swaminathan, G. J., Brew, K.; Acharya, K. R. *J. Biol. Chem.* **2002**, *277*, 28310-28318.
26. Zhang, Y.; Wang, P. G.; Brew, K. *J. Biol. Chem.* **2001**, *276*, 11567-11574.
27. Ly, H. D.; Loughheed, B.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **2002**, *41*, 5075-5085.
28. Loughheed, B.; Ly, H. D.; Wakarchuk, W. W.; Withers, S. G. *J. Biol. Chem.* **1999**, *274*, 37717-37722.
29. Geremia, S.; Campagnolo, M.; Schinzel, R.; Johnson, L. N. *J. Mol. Biol.* **2002**, *322*, 413-423.
30. Negishi, M. C.; Dong, J.; Darden, T. A.; Pedersen, L. G.; Pedersen, L. C. *Biochem. Biophys. Res. Commun.* **2003**, *303*, 393-398.
31. Lairson, L. L.; Chiu, C. P. C.; Ly, H. D.; He, S.; Wakarchuk, W. W.; Strynadka, N. C. J.; Withers, S. G. *J. Biol. Chem.* **2004**, *279*, 28339-28344.
32. Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875-11882.
33. Cleland, W. W.; Frey, P. A.; Gerlt, J. A. *J. Biol. Chem.* **1998**, *273*, 25529-25532.

Chapter 17

Two Unexpected Effects Found with 2,3,4,6-Tetra-*O*-methyl-*D*-Gluco- and Mannopyranosyl Oxacarbenium Ions

An O-2 Pseudoequatorial Preference and a Large H-2-C-2--O-2-CH₃ *syn* Preference

Andrei Ionescu¹, LiJie Wang¹, Marek Z. Zgierski¹, Tomoo Nukada²,
and Dennis M. Whitfield^{3,*}

¹Stearie Institute for Molecular Sciences, National Research Council
Canada, 100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada

²The Institute for Physical and Chemical Research (RIKEN), Wako-shi,
351-01 Saitama, Japan

³Institute for Biological Sciences, National Research Council Canada,
100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada

Analysis of two of the stable conformations (**B0** and **B1**) of a number of differently configured glycopyranosyl oxacarbenium ions by Density Functional Theory (DFT) calculations found two unexpected phenomena. Firstly, a marked preference for O-2 to be pseudo-equatorial that most dramatically leads to the **B1** conformation of *D-gluco* configured cations taking a ⁵S₁ conformation. The second phenomenon is that the H-2-C-2--O-2-CH₃ torsion angle is found to prefer *syn* conformations over *anti* conformations by more than 10 kJ mol⁻¹. Preliminary analysis of these two effects by examining the geometric consequences, the LUMO's and by Natural Bond Order (NBO) analysis are presented.

Introduction

The glycopyranosyl oxacarbenium ion is central to the chemistry of the formation and degradation of complex oligosaccharides, see Figure 1 (1). The formation of the ketal or acetal bond of the glycosidic linkage ($R^2CO-CR^3-O-CR^1$) is typically formulated as nucleophilic attack by an alcohol (R^1C-OH) on an oxacarbenium ion ($R^2C-O-CR^3$)⁺ (2). Similarly, glycosidic bond cleavage is usually formulated as exocyclic protonation followed by dissociation to an alcohol and an oxacarbenium ion. A variety of experimental and theoretical studies of the enzymes notably glycosidases which break glycosidic linkages and glycosyltransferases which form glycosidic linkages suggest the importance of such ions as intermediates or transition states (TS). For example, the standard

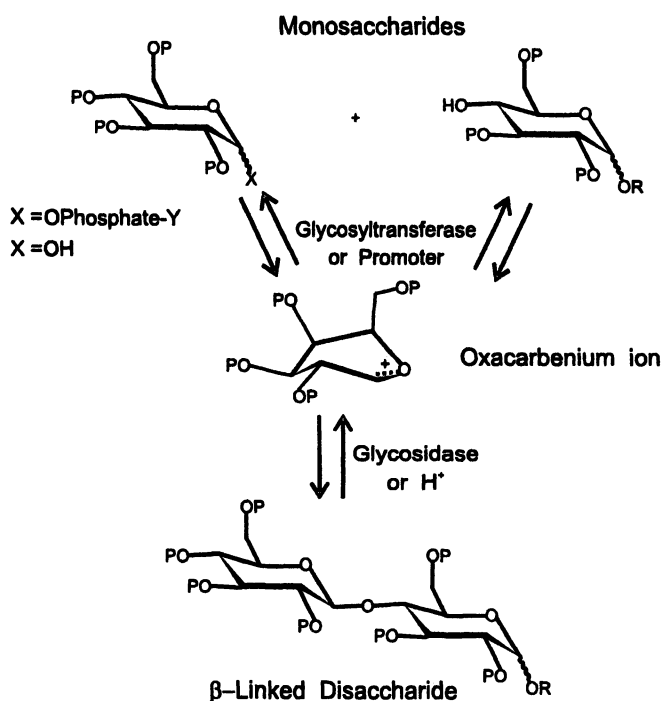


Figure 1. Formation or cleavage of glycosidic bonds is thought to proceed through oxacarbenium ion intermediates or TS in many cases. For enzyme catalysed bond formation $X = O\text{Phosphate-Y}$ where Y is usually either a nucleoside or a polyisoprene derivative. For enzyme reactions $P = H$ where the $OH'S$ are usually bound to the protein and for chemical reactions the $P =$ protecting groups in most cases. The example is for formation of a β -1,4-linked disaccharide formed from two monosaccharides but could apply to any glycosidic linkage.

mechanism of retaining glycosidases is thought to proceed through two TS's with appreciable oxacarbenium ion character (3). Similarly inverting glycosidases are believed to use a mechanism with considerable oxacarbenium ion character in the TS (4). The case with glycosyltransferases is much less studied but most current models invoke oxacarbenium ion intermediates or TS (5). For organic chemical reactions there is also considerable evidence that acid catalysed hydrolysis of glycosides go through oxacarbenium ion intermediates or TS (6). For "traditional" glycosylation reactions involving a donor with a leaving group at the anomeric carbon activated by electrophilic promoters reacting with alcohols are also believed to proceed through oxacarbenium ion intermediates or TS (7). For the special case of neighboring group participation where such ions are additionally stabilized by a second electronegative atom (usually the oxygen of a carbonyl) in the absence of nucleophiles there is unequivocal evidence for the formation of dioxolenium ions (8).

The importance of the conformation of glycopyranosyl oxacarbenium ions has been suggested many times (9). A historical example from the work of R. Lemieux concerns the experimental observation that the epimeric chlorides of 3,4,6-tri-*O*-acetyl-D-glucopyranosyl chloride exhibit a high degree of stereoselectivity under solvolysis conditions (α -Cl to β -OR and β -Cl to α -OR) yet exhibit S_N1 kinetics, see Figure 2 (10). In other words, although product analysis suggests a S_N2 reaction, kinetic analysis suggests a S_N1 reaction with an oxacarbenium ion intermediate. Prof. Lemieux's hypothesis was that each epimeric chloride ionized to a glycosyl oxacarbenium ion with different ring conformations, in this case a 4H_3 and 3H_4 pair. Furthermore, he hypothesized that these different ring conformations should exhibit facial selectivity leading to the observed solvolysis products.

For a number of years we have been studying the conformations of glycosyl oxacarbenium ions using Density Functional Theory (DFT) calculations (11) with the aim of assisting synthetic carbohydrate chemists like ourselves to develop better glycosylation reactions. Notably, we are interested in optimizing stereoselectivity and minimizing side reactions (12). To this end we have now studied a number of configurational isomers with different protecting groups. In the absence of trans fused rings all glycosyl oxacarbenium ions studied to date have been calculated to have at least two stable conformations. We call these conformations **B0** and **B1** where the conformation closest to that of the donor **A** is **B0**. In the Lemieux example above **B0** corresponds to the 4H_3 conformation and **B1** to the 3H_4 conformation. In many cases (see below) other conformations may be populated too but for simplicities sake we confine the discussion to two conformations. We call this idea the two conformer of glycosyl oxacarbenium ion hypothesis. Furthermore, by calculating the structures and energetics of complexes with the model nucleophile methanol by DFT of these pairs of

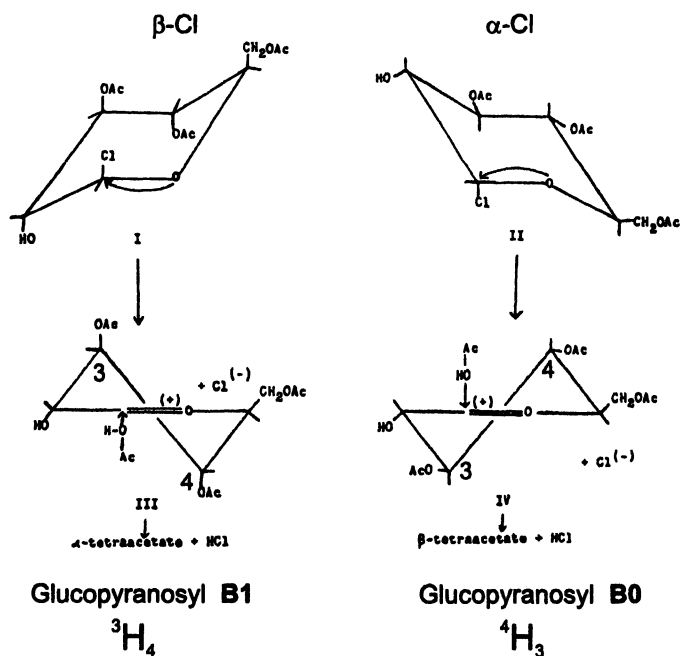
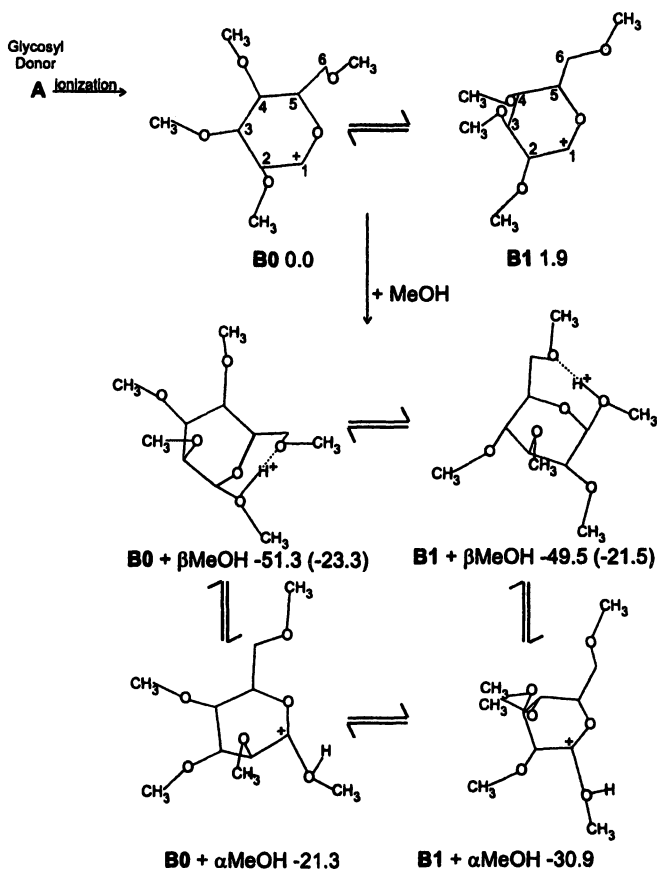


Figure 2. Under solvolysis conditions 3,4,6-tri-O-acetyl-β-D-glucopyranosyl chloride gives an α-acetate whereas the epimeric α-chloride gives a β-acetate. One hypothesis to account for these observations is that the β-chloride ionizes to a ³H₄ oxacarbenium ion whereas the α-chloride ionizes to a ⁴H₃ oxacarbenium ion. Furthermore, each conformation is hypothesized to show opposite facial selectivity. Structures scanned from the original publication with permission, see Reference 10.

cations we find facial selectivity. One example is shown in Scheme 1 for the case of 2,3,4,6-tetra-*O*-methyl-*D*-mannopyranosyl cations with methanol (13). Our analysis suggests that the key variables are the possibilities of intramolecular hydrogen bonding especially once the complexes obtain appreciable hydronium ion character and the consequences of changes of the τ_5 (C-5-O-5--C-1-C-2) ring torsion angle from the isolated cation to the respective α - or β -complexes. For the *D*-sugars considered so far this angle becomes more positive for β -attack and more negative for α -attack. These changes can be favorable or unfavorable depending on the starting conformation. These and other details are discussed in published reports and will not be elaborated on further here.

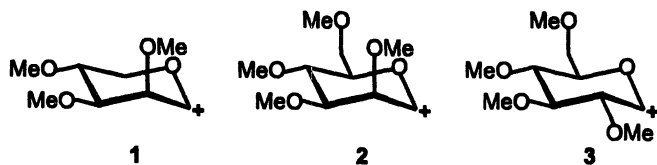


*Scheme 1. DFT (ADF) calculated facial selectivity of methanol complexes with 2 in its **B0** and **B1** conformations. Relative energies in kJ mol⁻¹ with **B0** + MeOH set to 0.0 and numbers in parentheses are corrected for H-bonding.*

Background - the Two Observations

In this communication we wish to concentrate on two unexpected observations that have arisen from our studies. The first concerns the orientation of O-2 with respect to the ring in these glycosyl oxocarbenium ions i.e. pseudo-axial or pseudo-equatorial, see Scheme 2 for the structures considered. The second observation concerns the rotational preferences about C-2--O-2 in some of these ions.

As in the example above the two conformers most often considered for glycosyl oxocarbenium ions are the half chairs \sim I-13 and 3 H₄ since these



Scheme 2. Structures of glycopyranosyl oxacarbenium ions 1 to 3.

conformations allow τ_5 to be almost planar with the C-1--O-5 bond exhibiting considerable double bond character and the positive charge largely localized on these two atoms. For *D-gluco* configured compounds the 4H_3 conformation allows the C-2, C-3, C-4 and C-5 substituents to be pseudo-equatorial whereas the 3H_4 conformation forces these substituents to be pseudo-axial. Conventional conformational wisdom penalizes such conformations by about $4 \times 4 = 16 \text{ kJ mol}^{-1}$ (14). The exact values depend on the electronegativity and the steric size of the substituents as well as other factors. For *D-manno* configured compounds the configuration at C-2 is inverted and the destabilisation should be smaller but still significant. Thus, it was quite surprising to us that the calculated energy difference between the **B0** (4H_3) and **B1** (3H_4) conformations of 2,3,4-tri-*O*-methyl-*D*-lyxopyranosyl oxacarbenium ions, **1**, was 12.6 kJ mol^{-1} in favour of **B1**, see Figure 3. Note that this 3H_4 conformation of **1** allows O-2 to be pseudo-equatorial.

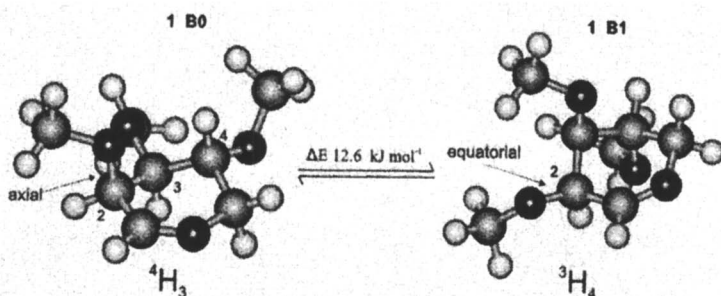


Figure 3. The DFT (ADF) calculated conformations of **B0** (4H_3) and **B1** (3H_4) oxacarbenium ions of 2,3,4-tri-*O*-methyl-*D*-lyxopyranosyl cation, **1**.

Subsequent calculations on 2,3,4,6-tetra-*O*-methyl-*D*-manno (**2**) and glucopyranosyl (**3**) oxacarbenium ions also found **B0** and **B1** pairs of conformations (13). For **2** the **B0** conformation was found to be 4H_3 with O-2 pseudo-axial and for **B1** the 3E envelope conformation which is adjacent in conformational space to 3H_4 with O-2 pseudo-equatorial, see Figure 4. Again

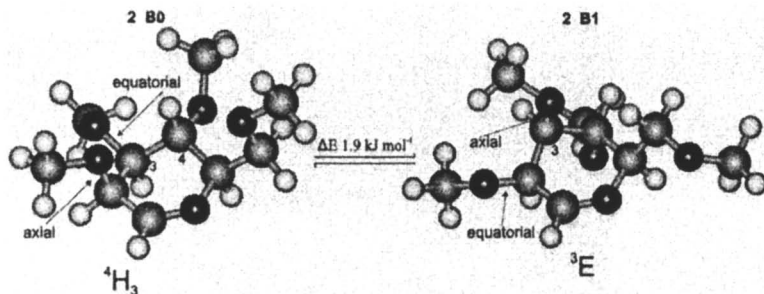
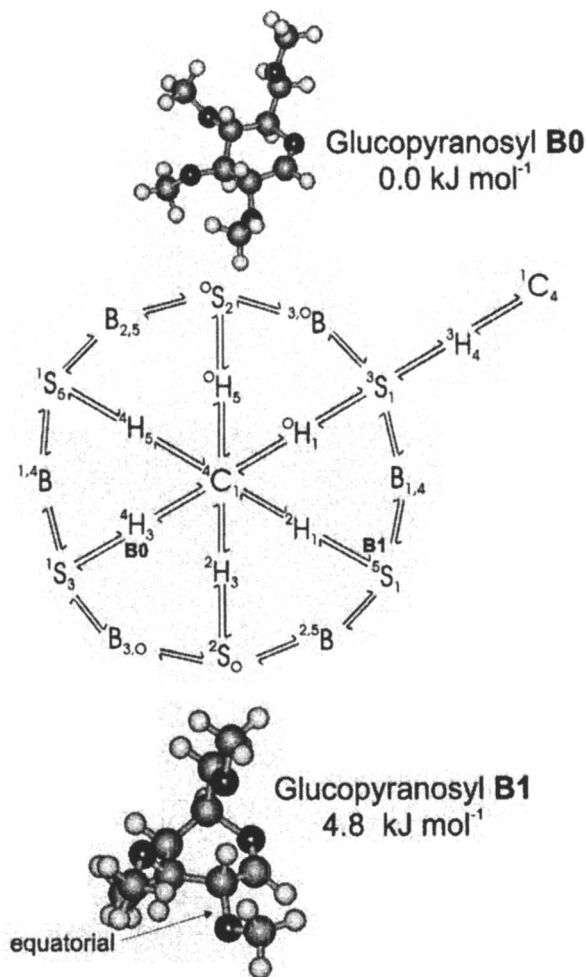


Figure 4. The DFT (ADF) calculated conformations of **B0** (4H_3) and **B1** (3E) oxocarbenium ions of 2,3,4,6-tetra-*O*-methyl-*D*-mannopyranosyl cation, **2**.

the energy difference was surprising and was found to be effectively 0. For **3** the **B0** conformation was indeed 4H_3 as expected but the **B1** conformation was found to be a twist boat 5S_1 far in conformational space from 3H_4 , see Figure 5. This conformation allows O-2 to be pseudo-equatorial. It is also adjacent in conformational space to the boat ${}^{2,5}B$ conformation that has been considered a possible conformation for the oxocarbenium ion TS of some glycosidase catalysed reactions (15). In this case **B0** is calculated to be 4.8 kJ mol^{-1} more stable than **B1**. Our conclusion from these and some related results is that glycopyranosyl oxocarbenium ions have a special preference for O-2 to be pseudo-equatorial with a very approximate stabilization of about 10 kJ mol^{-1} for this effect. The origin of this effect is not clear but could be related to the relative orientation of O-2 to the O-5--C-1 bond in these cations. For pseudo-equatorial orientations this puts O-2 in the same plane as O-5--C-1 whereas pseudo-axial orientations put O-2 approximately perpendicular to this plane.

The second effect was discovered during our studies of **2** and **3**. We had chosen *O*-methyl substituents for two reasons. One was to prevent intramolecular hydrogen bonding from obscuring other stereoelectronic effects on reactivity (16) and the second was that it was a prototype for our ultimate goal of studying the effects of protecting groups on reactivity. Previous studies of *O*-methyl substituted sugars had found that at least for secondary substituents conformations with the *O*-methyl groups *syn* to the corresponding sugar methines were the most stable (17). A similar *syn* preference has been noted for 1,2,3-trisubstituted 2-*O*-methyl cyclohexanes and was attributed to subtle steric effects (18). Examination of our minimum energy conformations for **1** to **3** shows that all secondary *O*-methyl substituents are *syn*. Calculations of other *O*-methyl conformations at positions other than *O*-2 showed that these were indeed the minima but the effects were small, typically $1\text{--}2 \text{ kJ mol}^{-1}$. Consequently when our initial studies of **3** **B1** found an *anti* conformation for H-2-C-2--O-2-CH₃ we were not concerned since frequency calculations showed this to indeed be a minimum on the Potential Energy Surface, PES. Our first interpretation was that this different conformation was caused by the 5S_1 conformation.



*Figure 5. The DFT (ADF) calculated conformations of **B0** (⁴H₃) and **B1** (⁶S₁) oxacarbenium ions of 2,3,4,6-tetra-O-methyl-D-glucopyranosyl cation, **3**. Also shown is the ¹C₄ hemisphere of the spherical representation of pyranose conformations. The extension on the upper right shows a portion of the ¹C₄ hemisphere. The labels **B0** and **B1** show the locations in conformational space of these conformations.*

Our attention was drawn back to this question when we embarked on a study of the PES for interconversion between the **B0** and **B1** conformations of **2** and **3**. Conventional for quantum mechanics studies using linear transit methods where a proposed TS is first found and then using the vibrational modes of the single negative frequency to find the minima forwards and backwards from the TS were performed. The results for **2** are shown in Figure 6. A pathway starting from ${}^3\text{E}$ (**B1**) passes through a first TS at ${}^3\text{H}_2$ half chair to a third minimum conformation ${}^0\text{S}_2$ followed by a second TS at ${}^1\text{S}_5$ and eventually to ${}^4\text{H}_3$ (**B0**). The barriers are about 19 kJ mol^{-1} between minima which is considerably lower than 34 kJ mol^{-1} calculated for a more rigid D-galactopyranosyl analogue (19). These barriers are low enough to allow equilibration if the species have sufficient lifetimes at normal reaction temperatures (20). One problem with this data is that the **B0** conformation is calculated to be higher in energy than **B1** by 9.2 kJ mol^{-1} . Part of this discrepancy is the different rotamers about C-5--C-6 for the **B0** and **B1** conformations which as shown by the dotted lines on the left of the Figure 6 accounts for 3.8 kJ mol^{-1} of the energy difference. The remaining energy difference is due to the orientation of O-2 as it is *anti* in the anomalous case.

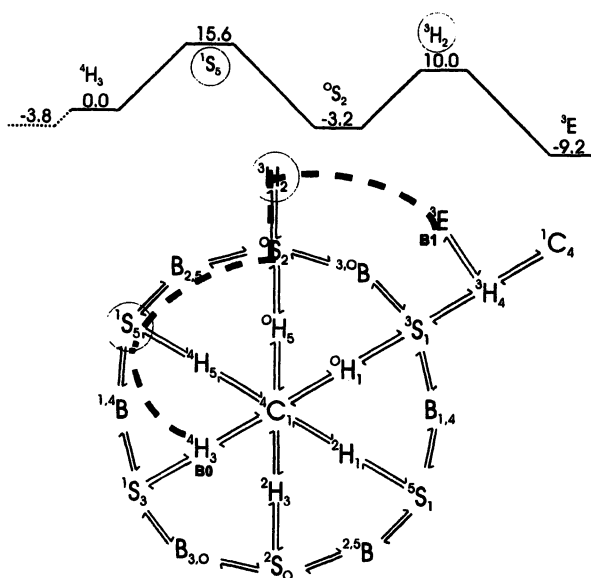


Figure 6. The DFT (Gaussian) calculated pathway for interconversion of **2** **B0** to **B1**. The dotted line is the pathway which is traced over the ${}^1\text{C}_1$ hemisphere of the spherical representation of pyranose conformations. The extension on the upper right shows a portion of the ${}^1\text{C}_4$ hemisphere. The bar graph at the top shows the relative energies. The dotted line on the left shows the correction for the C-5--C-6 torsion change, *gg* to *gt*.

This observation and similar observations with 3 prompted us to examine the O-2 orientation for all 4 cases with 2 and 3 i.e. **B0** and **B1**. The first result was that the *syn* conformation about O-2 for 3 **B1** was much lower in energy. In fact as can be seen in Table 1 the *syn anti* energy difference is about 15 kJ mol⁻¹ without the ring conformations changing appreciably for all four pairs. Thus, our original interpretation that the *anti* conformation was stabilized by the ⁵S₁ ring conformation was wrong and what we now consider the “normal” *syn* preference applies in this case too. This *syn* preference is much larger than the effect for “normal” O-methyl substituents on sugar or cyclohexane rings. The origin of this effect is not known but as is readily apparent from Newman projections about O-2--C-2 different lone pairs on O-2 are antiperiplanar to the p-type orbital on C-1 in the *syn* and *anti* conformations, see Figure 7. The exact values of the dihedral angles (C-1-C-2--O-2-CH₃) are given in the last row of Table 2.

Table 1 Relative Energies of *syn* and *anti* Conformers of 2 and 3.

| Species | DFT(1) E kJ mol ⁻¹ | DFT(2) E kJ mol ⁻¹ | Species | DFT(1) E kJ mol ⁻¹ | DFT(2) E kJ mol ⁻¹ |
|---------------------------|----------------------------------|----------------------------------|----------------------------|----------------------------------|----------------------------------|
| 2 B0 <i>syn</i> | 0.0 ^a | 0.0 ^b | 2 B0 <i>anti</i> | 16.28 | 20.83 |
| 2B1 <i>syn</i> | 1.91 | -0.87 | 2B1 <i>anti</i> | 14.12 | 16.24 |
| 3 B0 <i>syn</i> | 0.0 ^a | 0.0 ^b | 3 B0 <i>anti</i> | 14.79 | 19.25 |
| 3 B1 <i>syn</i> | 4.72 | 4.89 | 3 B1 <i>anti</i> | 13.62 | 14.60 |

a, b These energies set to 0.0 kJ mol⁻¹ for convenience.

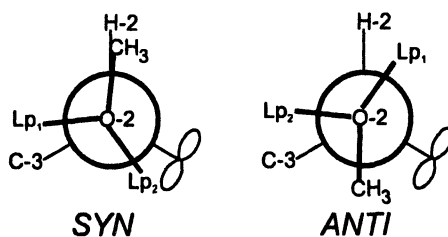


Figure 7. Schematic Newman projections about the O-2--C-2 bond showing the different orientations of lone electron pairs relative to the vacant p-type orbital on C-1.

Table 2. Geometric Variables for *syn* and *anti* 2 and 3 from DFT (ADF).

| Variable | 2 B0 <i>syn</i> | 2 B0 <i>anti</i> | 2 B1 <i>syn</i> | 2 B1 <i>anti</i> | 3 B0 <i>syn</i> | 3 B0 <i>anti</i> | 3 B1 <i>syn</i> | 3 B1 <i>anti</i> |
|--------------------------|--------------------|---------------------|--------------------|---------------------|--------------------|---------------------|--------------------|---------------------|
| C-5--O-5 | 1.538 | 1.531 | 1.522 | 1.518 | 1.526 | 1.531 | 1.521 | 1.523 |
| O-5--C-1 | 1.269 | 1.267 | 1.266 | 1.268 | 1.264 | 1.262 | 1.274 | 1.269 |
| C-1--C-2 | 1.482 | 1.479 | 1.47 | 1.467 | 1.48 | 1.48 | 1.467 | 1.472 |
| C-2--O-2 | 1.444 | 1.449 | 1.402 | 1.406 | 1.416 | 1.412 | 1.408 | 1.411 |
| O-2--CH ₃ | 1.453 | 1.456 | 1.446 | 1.457 | 1.45 | 1.449 | 1.443 | 1.448 |
| <C-5-O-5 | 121.6 | 121.4 | 122.7 | 122.7 | 121.9 | 121.4 | 117.6 | 117.8 |
| -C-1 | | | | | | | | |
| <O-5-C-1 | 124.5 | 124.7 | 125.1 | 124.3 | 126 | 126.1 | 125.9 | 124.9 |
| -C-2 | | | | | | | | |
| <C-1-C-2 | 92.4 | 96.5 | 107.9 | 114.8 | 107 | 110.7 | 109.9 | 111.9 |
| -O-2 | | | | | | | | |
| <C-2-O-2-CH ₃ | 115 | 118.7 | 114.2 | 115.5 | 113.6 | 113.5 | 113.5 | 114.8 |
| -C-5-O-5 | -9.4 | -18.9 | 3.8 | 5.1 | 0.4 | -6.2 | -4.6 | -6.2 |
| <O-5-C-1 | 94.6 | -105.1 | -157 | -163.9 | 138.8 | 152.9 | 160.3 | 174.0 |
| -C-2-O-2 | | | | | | | | |
| <C-1-C-2 | -116.9 | 81.8 | -143.1 | 48.5 | 119.9 | -70 | 92.4 | -71.9 |
| -O-2-CH ₃ | | | | | | | | |

Preliminary Analysis: Geometries, LUMO's and NBO's

Initial tests to clarify the origin of these two effects namely the unexpected preference for O-2 pseudo-equatorial and the unexpected magnitude of the *syn anti* energy difference about C-2--O₂CH₃ consist of three types. The first is to see if any geometric variables reflect these differences where the relevant data is collected in Table 2. The second is to plot the LUMO's associated with these 8 species (see Figure 8) and the third is to perform natural bond order NBO analysis to see if there are any trends in the apparent bonding patterns, see Figure 9.

Examination of Table 2 shows that for the species (2 B0 *syn* and *anti*) with O-2 pseudo-axial two geometric variables show some small differences. The C-2--O-2 bond length is about 0.04 Å longer than for the equatorial cases. The angle <C-1-C-2-O-2 is smaller than for the pseudo-equatorial cases. These observations are all consistent with some extra degree of electron delocalization from the C-2--O-2 bond into the O-5--C-1 bond in the axial cases. Perhaps related to these structural effects is the observation that the Lowest Unoccupied Molecular Orbitals (LUMO's) of the pseudo-axial species (see Figures 8ab) have electron density on O-2 which is not present in the pseudo-equatorial species (see Figures 8cd). The effect on geometric parameters for the *syn anti* pairs is less obvious and only two variables show any significant changes

besides the $\langle\text{C-1-C-2--O-2-CH}_3\rangle$ torsion angle that defines the difference. These two variables are the $\langle\text{C-2-O-2-CH}_3\rangle$ bond angle which is consistently larger in the *anti* isomers and the $\langle\text{O-5-C-1--C-2--O-2}\rangle$ torsion angle which is larger in magnitude for all *anti* isomers.

Natural Bond Order (NBO) analysis was performed for all 8 species to search for possible hyperconjugative interactions (21) that might explain the magnitude of the observed *syn* selectivity. NBO analysis is a method to separate localized energies, $E(\text{Lewis})$, from delocalization energies, $E(\text{Non-Lewis})$.

Based on DFT (DFT2, Gaussian 98) calculations the $E(\text{Lewis})$ favors the *syn* conformation of **3 B0** by 64.1 kJ mol^{-1} . Whereas the $E(\text{Non-Lewis})$ favors the *anti* conformation by 43.6 kJ mol^{-1} leading to a net energy difference of 19.3 kJ mol^{-1} , see Table 1. The absolute energies are much larger for $E(\text{Lewis})$

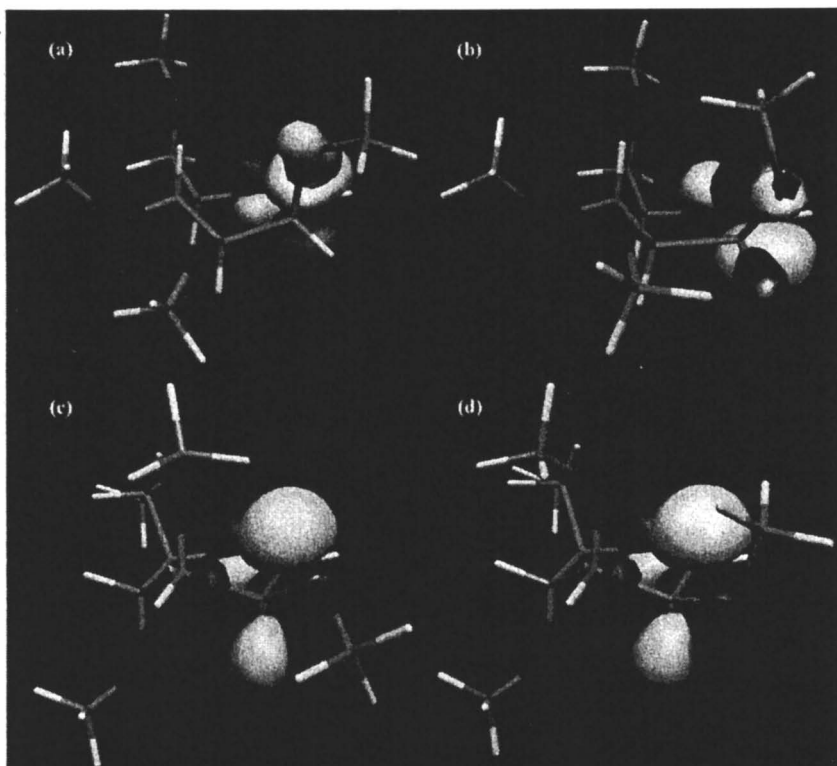


Figure 8. DFT optimized geometry (Gaussian) and spatial distribution of LUMO's for: (a) (2) B0 syn, (b) (2) B0 anti, (c) (2) B1 syn, and (d) (2) B1 anti. Note pseudo-axial orientation of O-2 in (a) and (b). (See color page 1 in chapter.)

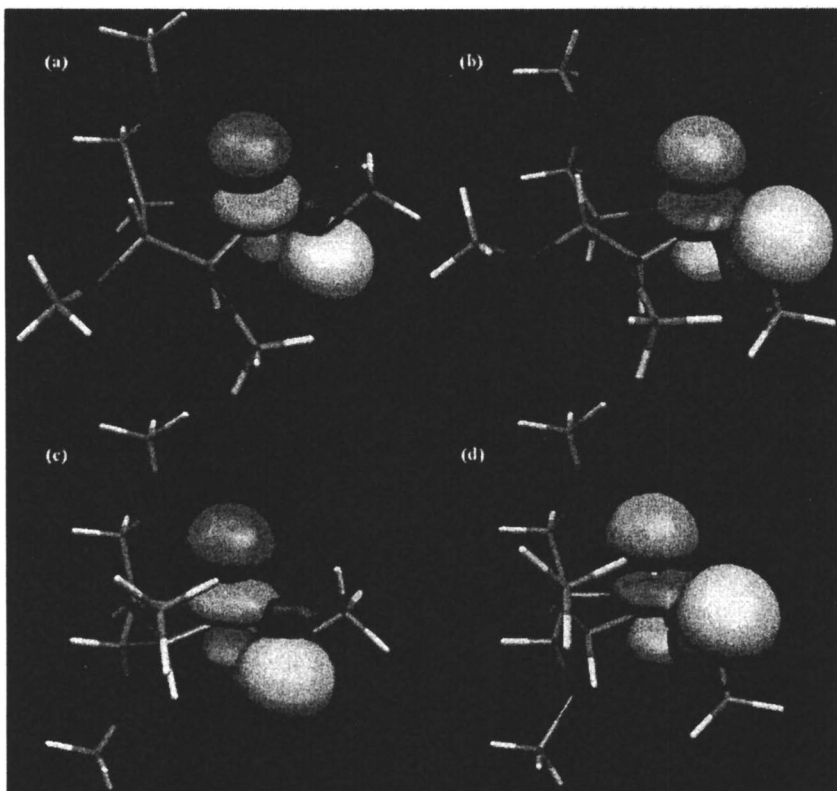


Figure 9. The $n_{O2} \rightarrow \sigma^* CH$ hyperconjugative interactions from NBO analysis for: (a) (3)B0 *syn*, $E = 19.6 \text{ kJ mol}^{-1}$ (b) (3)B0 *anti*, $E = 6.7 \text{ kJ mol}^{-1}$ (c) (3)B1 *syn*, $E = 18.8 \text{ kJ mol}^{-1}$ and (d) (3) B1 *anti*, $E = 6.7 \text{ kJ mol}^{-1}$.
(See color page 2 in chapter.)

representing 99.88 % of the total energy (99.879% for *syn* and 99.877% for *anti*). However comparing this to benzene which at a similar level of theory has an E(Lewis) contribution of 99.742% shows that delocalization energy is an important factor for glycopyranosyl oxacarbenium ions.

Further analysis of the delocalization contributions shows that the magnitude of the interaction of one of the lone pairs on O-2 with the σ^* orbital of CH-2 is much larger in all cases examined for the *syn* conformers. This interaction is shown in Figure 9 for 3. The difference in this case is that the interaction is antiperiplanar in the *syn* conformer but synperiplanar for the *anti* conformer. This is consistent with the Bohlmann torsional effect (22). For the corresponding σ orbital of CH-2 there is a large charge transfer to the π^* orbital of C-0--O-5.

To further elucidate this effect a torsion angle scan of H-2-C-2--O-2-CH₃ for **3 B0** was undertaken. The results confirmed that the *syn* and *anti* conformations are the only two minima with a barrier of about 40 kJ mol⁻¹ separating them. Further NBO analysis revealed that at the barrier the total E almost equals the E(Lewis) whereas at the *syn* minimum these energies are different and nearly all the energy difference is contributed by the delocalization E, about 20 kJ mol⁻¹. Thus the CH-2-C-2--O-2-CH₃ *syn* preference can be largely attributed to hyperconjugative effects.

Computational Methods

The DFT(1) calculations were carried out with the Amsterdam Density Functional (ADF) program system, ADF2000 (23). The atomic orbitals were described as an uncontracted double- ζ Slater function basis set with a single- ζ polarization function on all atoms which were taken from the ADF library. The 1 s electrons on carbon and oxygen were assigned to the core and treated by the frozen core approximation. A set of s, p, d, f, and g Slater functions centered on all nuclei were used to fit the electron density, and to evaluate the Coulomb and exchange potentials accurately in each SCF cycle. The local part of the V_{xc} potential (LDA) was described using the VWN parametrization (24), in combination with the gradient corrected (CGA) Becke's functional (25) for the exchange and Perdew's function for correlation (BP86) (26). The CGA approach was applied self-consistently in geometry optimizations. Second derivatives were evaluated numerically by a two point formula. The solvation parameters were dielectric constant $\epsilon=9.03$, ball radius =2.4 Å, with atomic radii of C=1.7, O=1.4 and H=1.2 Å. All reported energies include this solvation correction and Zero Point Vibration energies calculated from frequency calculations.

The DFT(2) calculations were carried out using the Gaussian 98 package (27) All structures were fully optimized at the b3lyp/6-31+g** level of theory. The diffuse orbitals augmented basis set (6-31+g**) was used to take into account the relative diffuse nature of the lone pairs. The DFT was used to take into account electron correlation. The NBO analysis was performed through the inclusion of population keyword in Gaussian 98 using the NBO 3.1 program (28). The filled NBOs of the natural Lewis structure describe covalency effects. The general transformation of canonical delocalized Hartree-Fock MOs to localized hybrid orbitals (NBOs) also leads to orbitals that are unoccupied in the formal Lewis structure (29). The interactions between filled (donor) and unfilled (acceptor) orbitals represent the deviation of the molecule from the Lewis structure and can be used as a measure of delocalization (30). The energy associated with the anti-bond orbitals can be numerically assessed by deletion of these orbitals from the basis set and recalculation of the total energy.

Possible Implications for Chemistry and Biology

There is very active interest in the area of glycobiology to make inhibitors of glycosyl transferases and glycosidases. Since many of these enzymes are thought to have reaction pathways that involve glycopyranosyl oxacarbenium ions one strategy towards such inhibitors is to make compounds that mimic them. The marked preference for *O*-2 to be in the same plane as *O*-5--*C*-1 suggests that such mimics should contain planar arrays of the general type *V*-*X*-*Y*-*Z* where the array has a net positive charge, the *V*-*X*--*Y*-*Z* torsion angle is near 180° and *Z* is an electronegative atom. Such an array mimics the *O*-5-*C*-1-*C*-2-*O*-2 grouping in pseudo-equatorial conformers of glycopyranosyl oxacarbenium ions.

The activating effect of electron donating versus the deactivating effect of electron withdrawing substituents at *O*-2 of glycopyranosyl donors in glycosylation reactions is well known. The observed *syn* preference about *C*-2--*O*-2 strongly suggests that protecting groups that highly populate this conformer should be more reactive. Similarly it seems probable that some enzymes may take advantage of this conformational preference and preferentially stabilize the *syn* conformations as the reactants proceed towards oxacarbenium ions. This will have a large structural impact on the protein. This is schematically shown in Figure 10 where selected atoms of the *syn* and *anti* **B1** conformations of 2-

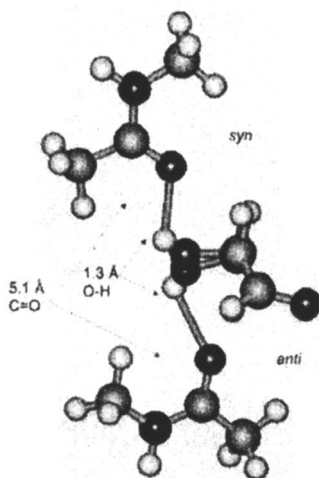


Figure 10. Superimposition of the *syn* and *anti* conformers of 2-hydroxy-3,4,6-tri-*O*-methyl-*D*-glucopyranosyl cation H-bonded to *N*-methylacetamide. The *C*-3, *C*-4, *C*-5 and *C*-6 atoms have been omitted for clarity.

hydroxy-3,4,6-tri-*O*-methyl-D-glucopyranosyl cation are shown with the hydroxyl proton donating a H-bond to N-methylacetamide. In this case the carbonyl oxygens are $>5 \text{ \AA}$ apart. Methods to test these hypotheses are under current investigation.

Acknowledgements

The authors gratefully acknowledge the use of the VPP770 Fujitsu parallel computer situated in the RIKEN Computer Center. This work was partly supported by the iHPC multiscale modeling initiative of the NRC. This is NRC paper #42501.

References

1. Stubbs, J.M.; Marx, D. *J. Am. Chem. Soc.* **2003**, *125*, 10960.
2. Kochetkov, N.E. *Studies in Natural Products Chemistry*, Rahman, A., Ed.; **1994**, *14*, p201.
3. Davies, G.J.; Ducros, V.M.-A.; Varrot, A.; Zechel, D.L. *Biochem. Soc. Trans.* **2003**, *523*, 523.
4. Guérin, D.M.A.; Lascombe, M.B.; Costabel, M.; Souchon, H.; Lamzin, V.; Béquin, P.; Alzari, P.M. *J. Mol. Biol.* **2002**, *316*, 1061.
5. André, I.; Tvaroška, I.; Carver, J.P. *Carbohydr. Res.* **2003**, *338*, 865.
6. (a) Hosie, L.; Sinnott, M.L. *Biochem. J.* **1985**, *226*, 437. (b) Ahmad, I.A.; Birkby, S.L.; Bullen, C.A.; Groves, P.D.; Lankau, T.; Lee, W.H.; Maskill, H.; Miatt, P.C.; Menneer, I.D.; Shaw, K. *J. Phys. Org. Chem.* **2004**, *17*, 560. (c) Lee, J.K.; Bain, A.D.; Beri, P.J. *J. Am. Chem. Soc.* **2004**, *126*, 3769.
7. (a) Whitfield, D.M.; Douglas, S.P. *Glycoconjugate J.* **1996**, *13*, 5. (b) Paulsen, H. *Angew. Chem. Int. Ed. Engl.* **1982**, *21*, 155. (c) Ogawa, T. *Chem. Soc. Rev.* **1994**, 397. (d) Demchenko, A.V. *Curr. Org. Chem.*, **2003**, *7*, 35. (e) Schmidt, R.R.; Kinzy, W. *Adv. Carbohydr. Chem. Biochem.*, **1994**, *50*, 21.
8. (a) Paulsen, H.; Schüttpelz, E. *Chem. Ber.* **1979**, *112*, 3214. (b) Crich, D.; Dai, Z.; Gastaldi, S. *J. Org. Chem.* **1999**, *64*, 5224.
9. Andrews, C.W.; Fraser-Reid, B.; Bowen, J.P. *J. Am. Chem. Soc.* **1991**, *113*, 8293.
10. Lemieux, R.U.; Huber, G. *Can. J. Chem.* **1955**, *33*, 128.
11. Bravo, F.; Viso, A.; Alcázar, E.; Molas, P.; Bo, C.; Castillón, S. *J. Org. Chem.* **2003**, *68*, 686.
12. (a) Nukada, T.; Bérces, A.; Zgierski, M.Z.; Whitfield, D.M. *J. Am. Chem. Soc.* **1998**, *120*, 13291. (b) Nukada, T.; Bérces, A.; Whitfield, D.M. *Carbohydr. Res.* **2002**, *337*, 765. (c) Bérces, A.; Whitfield, D.M.; Nukada,

- T.; do Santos, Z.I.; Obuchowska, A.; Krepinsky, J.J. *Can. J. Chem.* **2004**, *82*, 1157.
13. (a) Nukada, T.; Bérces, A.; Wang, L.J.; Zgierski, M.Z.; Whitfield, D.M. *Carbohydr. Res.* **2005**, *340*, 841. (b) Nukada, T.; Bérces, A.; Whitfield, D.M. *J. Org. Chem.* **1999**, *64*, 9030.
14. Dale, J. *Stereochemistry and Conformational Analysis*; Verlag Chemie: New York, 1978.
15. (a) Varrot, A.; Macdonald, J.; Stick, R.V.; Pell, G.; Gilbert, H.J.; Davies, G.J. *J. Chem. Soc. Chem. Commun.* **2003**, 946. (b) Blériot, Y.; Vadivel, S.K.; Herrera, A.J.; Greig, I.R.; Kirby, A.J.; Sinaÿ, P. *Tetrahedron* **2004**, *60*, 6813. (c) Guérin, D.M.A.; Lascombe, M.-B.; Costabel, M.; Souchon, H.; Lamzin, V.; Béguin, P.; Alzari, P.M. *J. Mol. Biol.* **2002**, *316*, 1061. (d) Sidhu, G.; Withers, S.; Nguyen, N.T.; McIntosh, L.P.; Ziser, L.; Brayer, G.D. *Biochemistry*, **1999**, *38*, 5346.
16. (a) Barrows, S.E.; Dulles, F.J.; Cramer, C.J.; French, A.D.; Truhlar, D.G. *Carbohydr. Res.* **1995**, *276*, 219. (b) Whitfield, D.M. *J. Molec. Struct. (Theochem)*, **1997**, *395*, 53. (c) Polavarapu, P.L.; Ewig, C.S. *J. Comput. Chem.* **1992**, *13*, 1255. (d) Jeffrey, G.A. *J. Molec. Struct.* **1990**, *237*, 75.
17. Mendonca, A.; Johnson, G.P.; French, A.D.; Lame, R.A. *J. Phys. Chem. A* **2002**, *106*, 4115.
18. Anderson, J.E.; Ijeh, A.I. *J. Chem. Soc. Perkin Trans 2* **1994**, 1965.
19. Bérces, A.; Nukada, T.; Whitfield, D.M. *J. Am. Chem. Soc.* **2001**, *123*, 5460.
20. (a) Banait, N.S.; Jencks, W.P. *J. Am. Chem. Soc.* **1991**, *113*, 7951. (b) Bennett, A.J. *J. Chem. Soc. Perkin Trans 2* **2002**, 1207. (c) Chiappe, C.; Lo Moro, G.; Munforte, P. *Tetrahedron*, **1997**, *53*, 10471.
21. (a) Ribeiro, D.S.; Rittner, R. *J. Org. Chem.* **2003**, *68*, 6780. (b) Alabugin, I.V. *J. Org. Chem.* **2000**, *65*, 3910.
22. Lii, J.H.; Chen, K.H.; Allinger, N.L. *J. Phys. Chem. A*, **2004**, *108*, 3006.
23. (a) Baerends, E.J.; Ellis, D.E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41. (b) te Velde, G.; Baerends, E.J.; *J. Comput. Phys.* **1992**, *99*, 84. (c) Fonseca Guerra, C.; Snijders, J.G.; te Velde, G.; Baerends, E.J. *Theor. Chim. Acta* **1998**, *99*, 391. (d) Versiuis, L. Ziegler, T. *J. Chem. Phys.* **1988**, *88*, 322. (e) Fan, L.; Ziegler, T. *J. Chem. Phys.* **1992**, *96*, 9005.
24. Vosko, S.H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.
25. Becke, A.D. *Phys. Rev. A* **1988**, *38*, 3098.
26. Perdew, J.P. *Phys. Rev. B* **1986**, *34*, 7506.
27. Gaussian 98, Revision A.11.3, Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Zakrzewski, V.G.; Montgomery Jr., J.A.; Stratmann, R.E.; Burant, J.C.; Dapprich, S.; Millam, J.M.; Daniels, A.D.; Kudin, K.N.; Strain, M.C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G.A.; Ayala, P.Y.; Cui, Q.;

- Morokuma, K.; Rega, N.; Salvador, P.; Dannenberg, J.J.; Malick, D.K.; Rabuck, A.D.; Raghavachari, K.; Foresman, J.B.; Cioslowski, J.; Ortiz, J.V.; Baboul, A.G.; Stefanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R.L.; Fox, D.J.; Keith, T.; Al-Laham, M.A.; Peng, C.Y.; Nanayakkara, A.; Challacombe, M.; Gill, P.M.W.; Johnson, B.; Chen, W.; Wong, M.W.; Andres, J.L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E.S.; Pople, J.A. Gaussian, Inc., Pittsburgh PA, 2002.
28. NBO Version 3.1, Glendening, ED.; Reed, A.E.; Carpenter, J.E.; Weinhold, F.
29. Foster, J.P.; Weinhold, F. *J. Amer. Chem. Soc.* **1980**, *102*, 7211.
30. (a) Reed, A.E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066. (b) Reed, A.E.; Weinstock, R.B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.

INDEXES

Author Index

- Abdel-Aal, Hoda, 186
Aboitiz, Nuria, 60
Almond, Andrew, 156
André, Sabine, 81
André-Leroux, Gwénaëlle, 170
Ano, Susan O., 220
Asensio, Juan L., 60
Bryce, Richard A., 186
Cañada, F. Javier, 60
Canales, Angeles, 60
Chávez, M. Isabel, 60
Cuevas, Gabriel, 60
Dyckjaer, Jane Dannow, 203
Fernández-Alonso, M. Carmen,
60
Freedberg, Darón I., 220
Gabius, Hans-Joachim, 81
García-Herrero, Alicia, 60
Goddard III, William A., 271
Hillier, Ian H., 186
Ionescu, Andrei, 302
Jansson, Jennie L. M., 20
Jiménez-Barbero, Jesús, 60
Kenne, Lennart, 114
Kirschner, Karl N., 235
Kroon-Bratenburg, Loes M. J.,
133
Laederach, Alain, 258
Leeflang, Bas R., 133
Maliniak, Arnold, 20
Mari, Silvia, 60
McNamara, Jonathan P.,
186
Molinero, Valeria, 271
Muslim, Abdul-Mueed, 186
Norris, Scott E., 220
Nukada, Tomoo, 302
Prestegard, James H., 40
Sandström, Corine, 114
Siebert, Hans-Christian, 81
Tajkhorshid, Emad, 81
Tschampel, Sarah M., 235
Tvaroška, Igor, 285
van Kuik, J. Albert, 133
Venable, Richard M., 220
Vidal, Paloma, 60
Vliegenthart, Johannes F. G.,
1, 81
von der Lieth, Claus-Wilhelm,
81
Wang, LiJie, 302
Whitfield, Dennis M., 302
Widmalm, Göran, 20
Woods, Robert J., 203, 235
Yi, Xiaobing, 40
Zgierski, Marek Z., 302

Subject Index

A

- AcAMP2 (*Amaranthus caudatus* Antimicrobial Peptide-2), 62, 66–67
See also Hevein domains; Protein–carbohydrate interactions
- Acute-phase reactants (APRs), 96
- AMB99C force field, 249
- AMBER (Assisted Model Building and Energy Refinement) force field, 248–253
- AMBER* force field, 249
- Amorphous carbohydrates, water distribution in, 279–281
- α -Amylase (GH13) family
bioinformatics, 172–173
branching enzyme (BE), from *E. coli*, 171, 172, 174, 175*f*
computational modeling methods, 172–174
coordinates of complexes, modeling, 173–174
debranching enzyme (DBE) or isoamylase, from *Pseudomonas amyloclavata*, 171, 172, 174, 175*f*
- G10' and G12 docking, 176, 177*f*, 178*f*, 179
- hydrolysis reactions, 171
- maltose and isomaltose docking, 173–174
- molecular modeling, 173–174
- overview, 171–172
- role of subsites –1, +1, and +1', 179, 180*f*, 181, 182*f*, 183
- structures and domains, 171–172
- transfer reactions, 171
- X-ray crystallography, 173
- V_h-amylose, M3B coarse grain model, 278–279
- Anomers
anomeric protons, NMR shifts, 3
lactose anomers, methods of detection, 223, 224
maltose DQF-COSY spectrum with 2 O(2')H signals from anomeric configuration, 120*f*
possible linkage positions and anomeric configurations (α,β) in glucopyranosyl oligosaccharides, 237*f*
- Anti-carbohydrate antibodies. *See* Antibodies
- Antibodies
amino acid sequence, characterization of, 205–206
1B1 antibody, 209–210
chimeric single chain Fv (scFv), 205
comparative modeling
alignment scoring matrices, 207
comparison of Fv models with experimental structures, 209–210
generation of Fv fragment, 208–209
refinement of models using MD simulation, 211–212, 211–216
template selection, 206–207
treatment of complementary determining regions, 207–208
complementarity determining regions (CDRs), 205, 207–208
constant regions (C_{H/L}), 204
Fab (antigen binding fragment), structure, 204–206
Fc (stem fragment), 204, 205*f*
Fv (variable fragment)
application of MD simulations to Fv fragments, 213–214, 215*t*, 216
chimeric single chain Fv (scFv), 205

- comparison of Fv models with experimental structures, 209–210
 - Fv from 1B1, 210, 211*f*, 215*t*
 - Fv from mAb 735, 209–210, 211*f*, 215*t*
 - Fv from mAb Se155-4, 210, 211*f*, 215*t*
 - structure, overview, 205
 - heavy chain hypervariable loops (H1, H2, H3), 205
 - homology modeling, definition, 206
 - hypervariable regions ($V_{H/L}$), 204
 - IgG isotype structure, 204–206
 - light chain hypervariable loops (L1, L2, L3), 205
 - mAb 735 antibody, 209–210, 211*f*, 215*t*
 - mAb Se155-4 antibody, 209, 210, 211*f*, 215*t*
 - overview, 203–204
 - structure, 204–206
- B**
- Biological membranes, glycoconjugates in, 20, 21*f*
 - Branching enzyme (BE), from *E. coli*, 171, 172, 174, 175*f*
 - See also* α -Amylase (GH13) family
 - Bulk signal, 3
- C**
- Cellotetraose, surface interactions with cellulose
 - hydrogen bonding, 137–138, 139*t*
 - on top of (110) plane, 135*f*, 136, 137*f*
 - parallel (CP), antiparallel (CA), and rotated (CR) orientations, 136, 138
 - potential energies of interactions, 139–142
 - Cellulose crystals, surface interactions
 - cellotetraose, hydrogen bonding, 137–138, 139*t*
 - cellotetraose, on top of (110) plane, 135*f*, 136, 137*f*
 - cellotetraose, parallel (CP), antiparallel (CA), and rotated (CR) orientations, 136, 138
 - cellotetraose, potential energies of interactions, 139–142
 - construction of monoclinic (110) surface system, 134
 - disaccharides, lack of binding, 136
 - mannotetraose, hydrogen bonding, 143, 145*f*, 146*t*
 - mannotetraose, on top of (110) plane, 144, 145*f*
 - mannotetraose, parallel (MP), antiparallel (MA), and rotated (MR) orientations, 136, 145*f*
 - mannotetraose, potential energies of interactions, 144, 146–147
 - molecular dynamics calculations, 134–135
 - morphology of cellulose crystals, 133–134, 138
 - tetrasaccharide MD experiments, description, 136–138
 - tetrasaccharide MD simulations, abbreviations, 136
 - xylotetraose, hydrogen bonding, 148, 149*t*, 150*f*, 151*t*
 - xylotetraose, parallel (XP), antiparallel (XA), and rotated (XR) orientations, 136, 150*f*
 - xylotetraose, potential energies of interactions, 148, 149*t*, 151, 152*t*
 - xylotetraose, two- and three-fold symmetry, 148, 150*f*, 153
 - CHARMM (Chemistry at HARvard Macromolecular Mechanics) force field, 245–248
 - Chitin (β (1-4)-linked *N*-acetylglucosamine; GlcNAc), 62–63, 64*f*, 65–67
 - Conclavin A, 268, 269*f*

Conformation

β -methyl glycosides, 70–75
 binding, effect on, 12
 conformer states (gg, gt, tg), 51–57, 240, 241*t*
 conformers from molecular dynamics (MD) trajectories, sampling, 45
 dihedral angles ϕ , ψ , and ω , 12*f*
 flexibility, effect on, 12, 41, 187, 237
 $\alpha(1\rightarrow4)$ glucans, 276–279
 glycans, 12–15
 glycosidic torsion angles (ϕ , φ , ω) for conformers, 51*t*
 glycosidic torsion angles (ϕ , φ , ω) in gas or solid phases, 240, 241*f*, 242*f*
 glycosidic torsion angles ϕ and ψ , 221
 GM1 glycomimetic, 67–68, 69*f*
Helix Pomatia α_D -hemocyanin glycan, 14–15
 internal coordinates for rotation around a bond, 260*f*
 β -methyl glycosides, 70–75
 molecular mechanics (MM) calculations, 13
 pineapple stem bromelain, 15–18
 possible linkage positions and anomeric configurations (α, β) in glucopyranosyl oligosaccharides, 237*f*
 rotamers in carbohydrates, ω -angles, 238, 239*f*, 240, 241*f*, 241*t*
 trimannoside 1–3 linkage, 47, 48, 51–54
 trimannoside 1–6 linkage, 47, 51–54
See also Molecular dynamics (MD) simulations and calculations
 COSY (COrelated Spectroscopy), 3
 Coupling constants. *See* Heteronuclear carbon-proton coupling constants ($J_{C,H}$); Residual dipole couplings (RCD)
 CSFF (Carbohydrate Solution Force Field), 246–247
 α -Cyclodextrin (α -CD)

deuteration to reduce proton multiplicity, 26
 ^1H NMR spectrum, 27*f*
J doubling procedure for obtaining coupling constant, 26, 28*f*, 29*f*
 structure, 27*f*

 α -cyclodextrin (α -CD)

interactions with hydroxy protons, 117, 128

D

Debranching enzyme (DBE) or isoamylase, from *Pseudomonas amyloclavata*, 171, 172, 174, 175*f*
See also α -Amylase (GH13) family
 Degree of polymerization (DP), 271
 Docking of carbohydrates to proteins
 Conclavin A binding, 268, 269*f*
 conformational search algorithms, 259–261
 entropy of binding, 261, 263
 flexibility, effects of, 259–260
 free energy change (ΔG), calculation, 263, 264*f*
 free energy function, AutoDock simulations, 265, 268
 free energy function, calibration with linear regression, 265, 266*f*, 267*t*
 G10' and G12, docking of, 176, 177*f*, 178*f*, 179
 genetic algorithm docking strategy, 261, 262*f*
 hydrogen bonding and, 259, 263, 268, 269*f*
 lectins, binding affinities, 268
 maltose and isomaltose docking, 173–174
 Monte-Carlo/simulated annealing protocol (MC/SA), 261
 overview, 259, 268–269
 potential energy functions, 261, 263, 264*f*, 265, 268

thermodynamic cycle principle, 263, 264*f*

E

Exopolysaccharides

Lactobacillus delbrueckii subspecies bulgaricus, 8*f*

Lactobacillus lactis cremoris, 9*f*–11*f*

Streptococcus thermophilus, 6, 7*f*

F

Field strength, effects of, 9, 9*f*–11*f*, 36

Force fields

AMB99C force field, 249

AMBER (Assisted Model Building and Energy Refinement) force field, 248–253

AMBER* force field, 249

carbohydrate force field development, challenges, 237–238, 239

carbohydrate force fields, necessity of, 236–237

carbohydrate force fields, validation of, 239–242

CHARMM (Chemistry at HARvard Macromolecular Mechanics) force field, 245–248

CSFF (Carbohydrate Solution Force Field), 246–247

field strength, effects of, 9, 9*f*–11*f*, 36

GLYCAM (Glycosides and Glycoproteins with AMBER) force field, 250–251, 252*f*

GLYCAM-LP (Glycosides and Glycoproteins with AMBER lone-pair model) force field, 251

glycosidic torsion angles (ϕ , φ , ω) in gas or solid phases, 240, 241*f*, 242*f*

HGFB (Ha, Giammona, Field, Brady CHARMM) force field, 245–246
integrated biomolecular force fields, advantages, 238

OPLS-AA and OPLS (Optimized Potentials for Liquid Simulations) force fields, 242–245

OPLS-AA-SEI (Scaling Electrostatic Interactions) force field, 244–245
parameterization of carbohydrate force fields, 238, 239

PARM22/SU01 force field, 248

partial charges for α -D-Glcp, 247*t*

PHLB (Palma, Liang, Brady CHARMM) force field, 246

protein and nucleic acid force fields, development, 236

quantum mechanical force field approach for carbohydrates, 193, 195–199

rotamers in carbohydrates, ω -angles, 238, 239*f*, 240, 241*f*, 241*t*

scalar J couplings for α -glycans, 253*t*

scaling factors, 238, 244

torsion terms, 238, 239

See also Hybrid quantum mechanical/molecular mechanical (QM/MM) simulations; M3B coarse grain model of glucans; Molecular dynamics (MD) simulations and calculations

Fucosylated trisaccharide conformer predictions from RDCs and relaxation, 167–168

NOESY predictions from molecular dynamics simulation, 165*f*, 166

structure, 165*f*, 167*f*

sugar labeling nomenclature, 167*f*

G

β -galactosidase (*E. coli*), 70–75
Galectin-1

- interaction with pentasaccharide of ganglioside GM1, 88, 91*f*, 94
 laser photo CIDNP spectra, 90*f*
 ligand-induced conformational changes, 91*f*, 94
 structure, 91*f*
 surface accessibility of laser photo CIDNP-reactive amino acid residues, 88, 89*t*, 92*t*–93*t*, 94
See also Laser photo CIDNP (chemically induced dynamic nuclear polarization) technique;
 Lectins
- GlcNAc. *See* Chitin
- Glucans
 conformations of $\alpha(1\rightarrow4)$ glucans, 276–279
 glass transition and molecular mobility, 281–282
 glycosidic torsion angles ϕ and ψ , 221
 helical structure, 276–279
 properties and uses, 271–272
 water distribution in amorphous carbohydrates, 279–281
See also Cellulose crystals, morphology
- Glucose, atomist and M3B models, 275, 276*f*
- Glucose homopolymers. *See* Glucans
- GLYCAM (Glycosides and Glycoproteins with AMBER) force field, 250–251, 252*f*
- GLYCAM-LP (Glycosides and Glycoproteins with AMBER lone-pair model) force field, 251
- Glycans
 binding, effect on conformation, 12
 conformation, 12–15
 dihedral angles ϕ , ψ , and ω , 12*f*
¹H NMR spectra and reporter group concept, 2–3
 molecular dynamics (MD) calculations, 13–14
- NOE (nuclear Overhauser effect) measurements, 13
 scalar *J* couplings for α -glycans, 253*t*
- Glycobiology, advances in, 186–187, 236–237, 259, 316–317
- Glycomimetics
 β -methyl glycosides, conformation and interactions, 70–71, 72*f*, 73*f*, 74
E. coli β -galactosidase and, 70–71, 74–75
 interactions with *E. coli* β -galactosidase, 70–71
 phenyllactic-containing GM1 glycomimetic, conformation and interactions, 67–68, 69*f*
- Glycoproteins
 epitopes, 157
 glycans, function of, 20–21
 N-linked glycans, structure, 157*f*
 oligosaccharides in, 157
- Glycopyranosyl oxacarbenium ion
 3,4,6-tri-*O*-acetyl-D-glucopyranosyl chloride and, 304, 305*f*
 B0 and B1 conformations, facial selectivity, 304–305, 306*f*
 chemical and biological implications, 316–317
- Density Functional Theory (DFT)
 calculations of conformation, 304, 307*f*, 308*f*, 309*f*, 310*f*
- Density Functional Theory (DFT)
 computational methods, 315
 energy difference between conformations, 307
 geometric variables for *syn* and *anti* conformations, 312, 312*t*
 in intermediates or transition states, 304
 lone electron pairs in *syn* and *anti* conformations, 311*f*
- Lowest Unoccupied Molecular Orbitals (LUMOs), 312–313, 313*f*

- Natural Bond Order (NBO) analysis, 313-314
 preference for O-2 to be pseudo-equatorial, 306-308
 relative energies of *syn* and *anti* conformers, 311*t*
 role in formation or cleavage of glycosidic bonds, 303
 structures of conformers, 307*f*
 torsion angle of H-2-C-2--O-2-CH₃, preference for *syn* over *anti*, 306, 308-311
- Glycosylation
 congenital disorders of, 21
 function of, 21, 236
 lectins and, 21, 268
- Glycosyltransferases. *See* Retaining glycosyltransferases
- H**
- Helix pomatia* α_D-hemocyanin glycan conformational study, 14-15
 conformations, 15*f*
 heptosaccharide methyl β-glycoside spectra, 4*f*, 5*f*, 6*f*
 heptosaccharide methyl β-glycoside structure, 3, 4*f*
- Hemicelluloses, 133
- Heteronuclear carbon-proton coupling constants ($J_{C,H}$)
 band-selective decoupling, 26
 deuteration to reduce proton multiplicity, 26
- J* doubling procedure
 convolution with delta functions, 25*f*, 26
- α-cyclodextrin (α-CD) coupling constant determination, 26, 28*f*, 29*f*
- NMR signal of anti-phase doublet, 23
 pulse sequence for determining, 22
 stages of doubling, 23, 24*f*
- ¹H-¹³C heteronuclear multiple bond coherence (HMBC) spectroscopy, 7-8
- ¹H-¹³C heteronuclear multiple quantum coherence (HMQC) spectroscopy, 5-6
- ¹H-¹³C heteronuclear single quantum coherence (HMSC) spectroscopy, 6-7
- Hevein domains
 AcAMP2, 62, 66-67 (*See also* Protein-carbohydrate interactions)
- Atoms in Molecules (AIM) analysis, 68, 69*f*
 benzene model of sugar-aromatic interactions, 67-68, 69*f*
 binding affinities and binding constants, 66
 binding of GlcNAc analogues, 65-66
 counterpoise method and, 67-68, 69*f*
 HEV32 model of truncated hevein, 65
 phenyllactic-containing GM1 glycomimetic, conformation and interactions, 67-68, 69*f*
 protein/carbohydrate complexes, structure, 63, 64*f*, 65
 stacking interactions, 63, 65-66
 structure, 62
 sugar-aromatic interactions, 63, 66-67, 68
 techniques used to study binding, 62-63
- HGFB (Ha, Giammona, Field, Brady CHARMM) force field, 245-246
- History of NMR spectroscopy, 1-2
- HMBC (¹H-¹³C heteronuclear multiple bond coherence) spectroscopy, 7-8
- HMQC (¹H-¹³C heteronuclear multiple quantum coherence) spectroscopy, 5-6
- HMSC (¹H-¹³C heteronuclear single quantum coherence) spectroscopy, 6-7
- Human chorionic gonadotropin (hCG)

- laser photo CIDNP (chemically induced dynamic nuclear polarization) technique, 105, 106*f*–108*f*
- structure and function, 104–105, 108*f*
- Hybrid quantum mechanical/molecular mechanical (QM/MM) simulations
- ab initio* quantum mechanical methods, 195, 196*t*, 197, 236
- calculations and equations, 187–189
- directly coupled QM/MM calculations, 188–189
- dixylose (4-O- α -D-xylopyranosyl- α -D-xylopyranose)
- conformational free energy surface, 189–190, 191*f*
- hydrogen bonding arrangements, 190, 192*f*
- PM3 Hamiltonian, 189, 190
- PM3CARB-1 model, 199*f*
- polarization and charge, 193, 194*f*
- water-bridging conformers, populations, 190, 192*f*, 193*t*
- 1,2-ethanediol conformers
- PM3 Hamiltonian, 193, 195
- PM3CARB-1 model, 195
- relative energies, 196*t*
- glucose conformers
- minimum energy structures, 197, 198*f*, 199
- relative energies, 195, 196*t*, 197
- ONIOM calculations, 188
- PM3CARB-1 model, 193, 195–199, 200
- quantum mechanical force field approach for carbohydrates, 193, 195–199
- semi-empirical QM Hamiltonian (PM3 model), 189–193, 195, 199–200
- specific reaction parameter (SRP) approach, 193, 200
- See also* Molecular dynamics (MD) simulations and calculations
- Hydroxy protons
- assignment of resonances, 117–118
- chemical shifts (δ), 116, 122–127
- cyclodextrin interactions, 117, 128
- experimental conditions for observation, 116–117
- glycosidic shift differences ($\Delta\delta$), 122–123, 124*t*
- glycosyl linkage positions and, 118
- hydration, effects on chemical shifts, 123, 125*f*, 126–127
- hydrogen bond interactions, 118–120, 237
- Lewis b (Le^b) and Lewis Y (Le^y) oligosaccharides, 127
- maltose DQF-COSY spectrum with 2 O(2')H signals from anomeric configuration, 120*f*
- mono- α -D-cyclodextrin TOCSY and DQF-COSY spectra, 117*f*
- NMR parameters, 115–116
- NOEs and chemical exchanges, 116, 119, 120, 122
- protein-carbohydrate interactions, 128–129
- rate of exchange with water (k_{ex}), 116, 118, 119, 121*f*
- temperature coefficients ($d\delta/dT$), 116, 118–119
- trisaccharide configuration and O(2')H–O(5'') interaction, 119–120, 121*f*
- vicinal coupling constants ($^3J_{H,OH}$), 116, 118, 119–120, 121*f*
- L**
- Lacto-N-neotetraose (LNnT)
- ellipsoid model for calculating order parameters, 34, 36
- molecular dynamics (MD) computer simulation of LNnT, 32–33, 36
- RDCs ($^1D_{C,H}$) of LNnT, 33, 34*f*

- Lactobacillus delbrueckii* subspecies *bulgaricus* exopolysaccharide, 8*f*
- Lactobacillus lactis cremoris* exopolysaccharide, 9*f*, 10*f*, 11*f*
- Lactose
- anomers, methods of detection, 223, 224
 - conformational energy minima, ϕ and ψ values, 226, 227*f*
 - differences in shape of α - and β -lactose from ^{13}C spectra, 229*f*, 230
 - equilibrium between α - and β -lactose, 223*f*, 230
 - HSQC NOESY spectra, 226, 228
 - hyperconjugation in α - and β -lactose, 231–232
 - interactions with *E. coli* β -galactosidase, 70–75
 - linkage torsion (ϕ and ψ) values, 223*f*
 - natural abundance HSQC spectra, 224*f*
 - NMR-derived internuclear distances, 228
 - numbering schemes for between α - and β -lactose, 223*f*
 - previously unassigned chemical shifts, 228–229
 - RDC and NOE differences for α - and β -lactose, 226, 227*f*, 228, 229–232
 - RDC values in liquid crystal media, 224, 225*t*
 - structures of α - and β -lactose, 230*f*
- Laser photo CIDNP (chemically induced dynamic nuclear polarization) technique
- confirmation of X-ray crystallography data, 86, 94
 - galectin-1, 88, 89*t*, 90*f*, 92*t*–93*t*
 - generation of laser photo CIDNP, radical pair mechanism, 85
 - human chorionic gonadotropin (hCG), 105, 106*f*–108*f*
 - ligand-induced conformational changes, 91*f*, 94–96
 - protein–dye radical pair formation, 83, 84*f*
 - serum amyloid P component (SAP), 96–104, 97–98, 99*f*, 100, 101*t*, 104
 - sialidase of *Clostridium perfringens*, 86
 - sialidase of *Salmonella typhimurium*, 87*f*
 - spectra of polarized amino acid residues, 84*f*, 85
 - technical overview, 83–86
 - Urtica dioica* agglutinin (UAD), 94–96
- Lectins
- B-chain moiety of cholera toxin (CTB), 68
 - binding affinities, 268
 - carbohydrate recognition, 61, 62, 86
 - cell proliferation and, 41, 88
 - CH- π interactions at binding sites, 68, 88, 94
 - glycosylation and, 21, 268
 - laser photo CIDNP-reactive amino acid residues, 88, 89*t*, 92*t*–93*t*, 109
 - See also* Galectin-1; Hevein domains; Protein–carbohydrate interactions
- LNnT. *See* Lacto-N-neotetraose; Lacto-N-neotetraose (LNnT)
- Lowest Unoccupied Molecular Orbitals (LUMOs), 312–313
- M**
- Maltose, 273*f*, 276, 277*f*
 - Mannans, 133, 142, 221
 - Mannose, 142
 - Mannose pentasaccharide
 - calculated and measured RDCs, 161–162
 - conformation, 160*f*
 - NMR spectra, 160*f*
 - RDC calculation, 159–162
 - structure and $\alpha(1\rightarrow2)$ and $\alpha(1\rightarrow3)$ linkages, 159*f*, 160, 161*f*

- structure predictions from RCDs and relaxation data, 166–167
- Mannose trisaccharide, branched
 alignment tensor, 163*f*
 calculation of populations at $\alpha(1\rightarrow6)$ linkage, 163–164
 chemical structure, 162*f*
 conformation and RDC calculations, 162–164
 conformers, 162–163
 molecular dynamics exploration of $\alpha(1\rightarrow6)$ ω angle, 162–163
- Mannotetraose, surface interactions with cellulose
 hydrogen bonding, 143, 145*f*, 146*t*
 on top of (110) plane, 144, 145*f*
 parallel (MP), antiparallel (MA), and rotated (MR) orientations, 136, 145*f*
 potential energies of interactions, 144, 146–147
- M3B coarse grain model of glucans
 bonded terms in energy expression, 274*f*
 computation speed, increased, 275
 conformations of $\alpha(1\rightarrow4)$ glucans, 276–279
 electrostatic interactions and charges, lack of, 275, 279–280
 energy expression, 274*f*
 glass transition and molecular mobility, 281–282
 glucose, atomist and M3B models, 275, 276*f*
 helical structures of glucans, 276–279
 invertability mapping from atomist to M3B models, 274
 maltose, atomist and M3B models, 273*f*, 276, 277*f*
 Morse nonbonded interactions, 274*f*, 275
 overview, 272–275
 torsion angles, 273*f*, 276, 277*f*
 V_h -amylose structure, 278–279
- water distribution in amorphous carbohydrates, 279–281
- MD. *See* Molecular dynamics (MD) simulations and calculations
- Methyl-3,6,di-*O*-(α -D-mannopyranosyl)- α -D-mannopyranoside
 alignment tensors in bicelles and phage, 54
 averaging RDCs from multi-conformers for trimannoside, 52–54, 55*t*
 conformation about 1–3 linkage, 47, 48, 51–54
 conformation about 1–6 linkage, 47, 51–54
 conformational averaging, 46–47
 conformer states (S1, S2, S3), 51–57
 conformers, alignment, 51–52, 53*f*
 conformers (gg, gt, tg) from molecular dynamics (MD) trajectories, sampling, 50–51
 glycosidic torsion angles (ϕ , φ , ω) for conformers, 51*t*
 molecular dynamics (MD) simulation, 48, 49*f*
 RDCs for trimannoside in aligned media, 46–47, 47*t*
 REDCAT averaging, 48–54, 50*f*, 53*f*
 structure, 46*f*
- Molecular dynamics (MD) simulations and calculations
 branched mannose trisaccharide, $\alpha(1\rightarrow6)$ ω angle, 162–163
 branched mannose trisaccharide $\alpha(1\rightarrow6)$ ω angle, 162–163
 cellulose crystals, surface interactions, 134–135
 cellulose–tetrasaccharide interactions, 136–138
 complex systems, limitations of MD method, 272
 conformers from molecular dynamics (MD) trajectories, sampling, 45, 50–51

fucosylated trisaccharide, NOESY predictions, 165*f*, 166
 glycans, 13–14
 lacto-N-neotetraose (LNnT), 32–33
 methyl-3,6,di-*O*-(α -D-mannopyranosyl)- α -D-mannopyranoside, 48, 49*f*, 50–51
 molecular dynamics (MD) computer simulation of LNnT, 32–33, 36
 refinement of comparative antibody models using MD simulation, 211–216
 rotamers in carbohydrates, ω -angles, 238, 239*f*, 240, 241*f*, 241*t*
 X-ray crystallography and, 83
See also Force fields; Hybrid quantum mechanical/molecular mechanical (QM/MM) simulations

N

Natural Bond Order (NBO) analysis, 313–314
 NOE (nuclear Overhauser effect)
 detection of motion in carbohydrates, 41
 glycans, 13
 lactose, 226, 227*f*, 228, 229–232
 structure, three-dimensional (3-D), 221
 NOESY (nuclear Overhauser enhancement spectroscopy)
 effect of field strength, 9, 11*f*
 lactose HSQC NOESY spectra, 226, 228
 prediction of cross-peaks from relaxation matrix, 164–166

O

Oligosaccharide chains
 biological functions, 115, 186–187, 221

in glycoproteins, 157
 interaction with water molecules, 115
 structure determination methods, 115, 221

See also Hydroxy protons
 OPLS-AA and OPLS (Optimized Potentials for Liquid Simulations) force fields, 242–245
 OPLS-AA-SEI (Scaling Electrostatic Interactions) force field, 244–245

P

PARM22/SU01 force field, 248
 PHLB (Palma, Liang, Brady CHARMM) force field, 246
 Pineapple stem bromelain
 conformational study, 15–18
 effective rotation correlation times of monosaccharides, 17*f*, 18
 glycopeptide structure, 15*f*, 16*f*
 hydrogen bonds, 18
 Protein-carbohydrate interactions
 B-chain moiety of cholera toxin (CTB), 68
 benzene model of sugar-aromatic interactions, 67–68, 69*f*
 cell surface carbohydrates, 41
 CH- π interactions, 68, 88, 94
 deformation of saccharide ring, 61–62
E. coli β -galactosidase and glycomimetics, 70–71, 72*f*, 73*f*, 74–75
E. coli β -galactosidase and lactose, 70–75
 forces involved, 61
 GM1 glycomimetic, conformation and interactions, 68, 69*f*
 hydroxy protons, role of, 128–129
 sugar-aromatic interactions, 61, 63, 66–67, 68
See also Chitin; Docking of carbohydrates to proteins; Hevein

domains; Recognition of carbohydrates by receptors

Q

QM/MM. *See* Hybrid quantum mechanical/molecular mechanical (QM/MM) simulations

Quantum mechanical force field simulations. *See* Hybrid quantum mechanical/molecular mechanical (QM/MM) simulations

R

RCD. *See* Residual dipole couplings (RCD)

Receptors, structure. *See* Docking of carbohydrates to proteins; Hevein domains; Recognition of carbohydrates by receptors

Recognition of carbohydrates by receptors

background, 259

benzene model of sugar-aromatic interactions, 67–68, 69*f*

deformation of saccharide ring, 61–62

forces involved in interaction, 61, 67–68

See also Antibodies; Docking of carbohydrates to proteins; Hevein domains; Lectins; Protein–carbohydrate interactions

REDCAT (RESidual Dipolar Coupling Analysis Tool)

methyl-3,6,di-*O*-(α -D-mannopyranosyl)- α -D-mannopyranoside, 48–54, 50*f*, 53*f*

residual dipole couplings (RCD), 43, 45–46

Relaxation. *See* Spin relaxation

Residual dipole couplings (RCD)

alignment tensors, 54, 158

analytical models, overview, 30–31

APME (additive potential-maximum entropy) method, 30–31

averaging RDCs from multi-conformers for trimannoside, 52–54, 55*t*

branched mannose trisaccharide, conformation and RDC calculations, 162–164

carbohydrate structure determination, usefulness in, 221–220

combined use of relaxation and RDCs, 166–168

conformer predictions from RDCs and relaxation, 167–168

conformers from molecular dynamics (MD) trajectories, sampling, 45, 50–51

dipole interactions ($D_{C,H}$), expressions for, 31, 32, 42

D₂O resonance, quadrupolar splitting, 31

ellipsoid model for calculating order parameters, 34, 36, 159

magnetic field, mathematical relation to, 221

mannose pentasaccharide, RDC calculation, 159–162

mannose pentasaccharide structure predictions, 166–167

measurable RDCs in a simple carbohydrate, 42, 43*f*

measurement methods, 42

molecular dynamics (MD) computer simulation of LNNt, 32–33, 36

order matrix analysis, 43–44, 158–159, 222

order parameters ($S_{\alpha\alpha}$ or S_{ij}), 33–35, 43–44, 158–158, 223

ordering tensor, from eigenvalues of gyration tensor, 35–36

ordering tensor, from eigenvalues of inertia tensor, 33–35

RDCs (${}^1D_{C,H}$) of LNNt, 33, 34*f*

- REDCAT (REsidual Dipolar Coupling Analysis Tool), 43, 45–46, 48–54
- Saupe order matrix (alignment tensor), 158–159, 222–223
- separation of effects of overall and internal motion, 44, 45*f*
- See also* Methyl-3,6,di-*O*-(α -D-mannopyranosyl)- α -D-mannopyranoside
- Retaining glycosidases, 287, 303, 304
- Retaining glycosyltransferases
- double-displacement mechanism
 - background, 287–288
 - covalent glycosyl-enzyme intermediate formation (first step), 287, 288–291, 289*f*
 - displacement (second) step, 287–288, 291–294, 291*f*
 - glycosyl-enzyme intermediate, structure, 289*f*
 - overview, 286–288
 - potential energy surface (PES) for second step, 291–292
 - potential energy surfaces (PES) for first step, 288, 289
 - schematic of two-step mechanism, 288*f*, 293*f*
 - transition state, 298–299, 304
 - oxacarbenium ion intermediate or transition state, 304
 - reactions catalyzed by glycosyltransferases, 286, 303
- S_Ni (single front-side displacement) mechanism
- background, 286–287, 294
 - LgtC complex (lipopolysaccharyl α -1,4-galactosyltransferase C), 294–295
 - potential energy surfaces (PES), 295–296
 - reaction pathways, 295–298
 - schematic of mechanism, 294*f*
 - transition state, 299
 - structure, 286–287
 - transition states
 - characteristics, 298
 - geometric parameters, 298*t*
- ROESY (rotating-frame Overhauser effect spectroscopy), 5*f*
- ## S
- SAP. *See* Serum amyloid P component
- Serum amyloid P component (SAP)
- deglycosylation, 97–98
 - laser photo CIDNP (chemically induced dynamic nuclear polarization) technique, 97–98, 99*f*, 100, 101*t*, 104
 - overview, 96
 - sialic acid moieties, 97–98, 99*f*, 104
 - structure, 96, 97*f*, 100*f*, 102*f*–103*f*
- Sialidase
- Clostridium perfringens*, 86
 - Salmonella typhimurium*, 86, 87*f*
- Spin relaxation
- combined use of relaxation and RDCs, 166–168
 - cross-relaxation rate matrix (Γ), 164
 - diagonal matrix (Ξ), 164–165
 - fucosylated trisaccharide, 165*f*
 - mannose pentasaccharide structure predictions, 166–167
 - prediction of NOESY cross-peaks from relaxation matrix, 164–166
 - spectral density function ($J(\omega)$), 164, 165
 - spin relaxation rates and motion in carbohydrates, 41
 - trisaccharide NOESY predictions from molecular dynamics simulation, 165*f*, 166
- Stinging nettle agglutinin (*Urtica dioica* agglutinin, UDA). *See Urtica dioica* agglutinin
- Streptococcus thermophilus*
- exopolysaccharide, 6, 7*f*

Structural reporter groups, 3
 Structure, three-dimensional (3-D)
 flexibility, effect on, 12, 41, 187,
 237
 glycosidic torsion angles ϕ and ψ ,
 221
J couplings and, 221, 253*t*
 methods of determination, 115, 221,
 236
 NMR analysis, overview, 12–15
 NOE (nuclear Overhauser effect)
 and, 221
 residual dipole couplings (RCD) and,
 221–221
 rotamers in carbohydrates,
 ω -angles, 238, 239*f*, 240, 241*f*,
 241*t*
 structural complexity of
 carbohydrates, 187, 221, 237
 three-dimensional (3-D) NMR
 spectra, overview
 overview, 2, 12–13
 three-dimensional (3-D) spectra, 2,
 12–13
See also Residual dipole couplings
 (RCD)

T

Three-dimensional (3-D) spectra, 2,
 12–13
 TOCSY (Total Correlation
 Spectroscopy)
 effect of field strength, 10*f*
 heptasaccharide methyl β -glycoside
 TOCSY spectrum, 4*f*
 mono-altro- β -cyclodextrin TOCSY
 and DQF-COSY spectra,
 117*f*
 Two-dimensional (2D) spectra,
 3–5

U

UDA. *See* *Urtica dioica* agglutinin
Urtica dioica agglutinin (UAD), 94–
 96, 95*f*

W

Water distribution in amorphous
 carbohydrates, 279–281

X

X-ray crystallography
 advantages and disadvantages, 82–83
 α -Amylase (GH13) family, 173
 biomolecular three-dimensional
 structures, 156–157
 galectin-1, 91*f*
 laser photo CIDNP technique and,
 86, 94
 molecular dynamics (MD)
 simulations and, 83
 serum amyloid P component (SAP),
 96, 97*f*
 sialidase of *Salmonella typhimurium*,
 87*f*
 Xylan, 147–148
 Xylose, 142
 Xylotetraose, surface interactions with
 cellulose
 hydrogen bonding, 148, 149*t*, 150*f*,
 151*t*
 parallel (XP), antiparallel (XA), and
 rotated (XR) orientations, 136,
 150*f*
 potential energies of interactions,
 148, 149*t*, 151, 152*t*
 two- and three-fold symmetry, 148,
 150*f*, 153



Figure 1. Collection of 23 NMR structures for hevein bound to $(\text{GlcNAc})_5$. The backbone fit RMSD between residues (3-41) is 0.57 \AA ; between residues (16-1): 0.39 \AA and for the key lateral chains holds S19: 0.19 ; W21: 0.28 ; W23: 0.32 ; Y30: 0.30 .

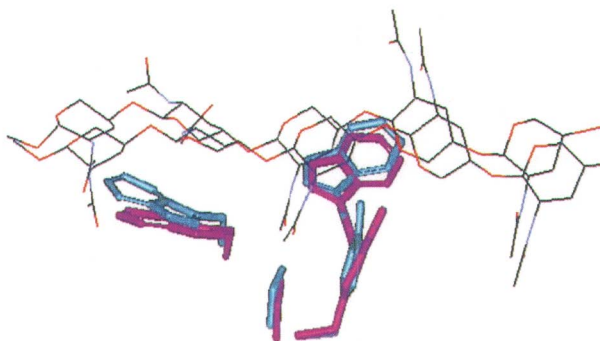


Figure 2. Superimposition of two NMR structures for hevein bound to $(\text{GlcNAc})_5$. The orientation of the key lateral chains is emphasized.



Figure 3. Superimposition of the NMR structures for free and (GlcNAc)₅-bound hevein. The orientation of the key lateral chains is emphasized.⁵

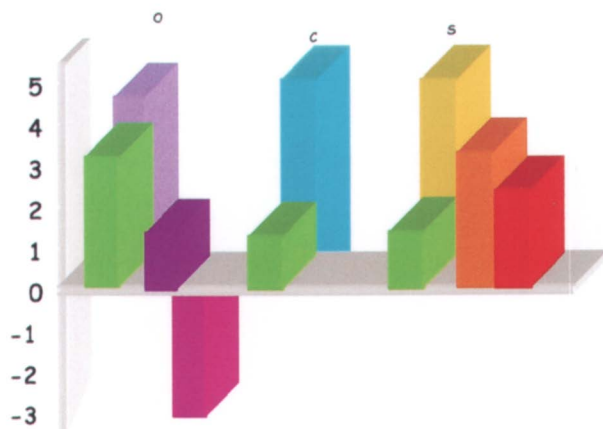


Figure 9. The required energies for rotating around Φ or for deforming the chair for O- (left), C- (middle), and S-glycosides (right). The green bar represents the required energy for rotation around Φ . The highest bars in all cases around 5 kcal/mol represent the required energies for deformation of the chair to the observed bound conformation. Protonation of the exo-cyclic oxygen lone pairs (additional two bars at the right handside of the O- and S-glycosides) strongly decreases the required energy for chair deformation, especially for the O-compound.

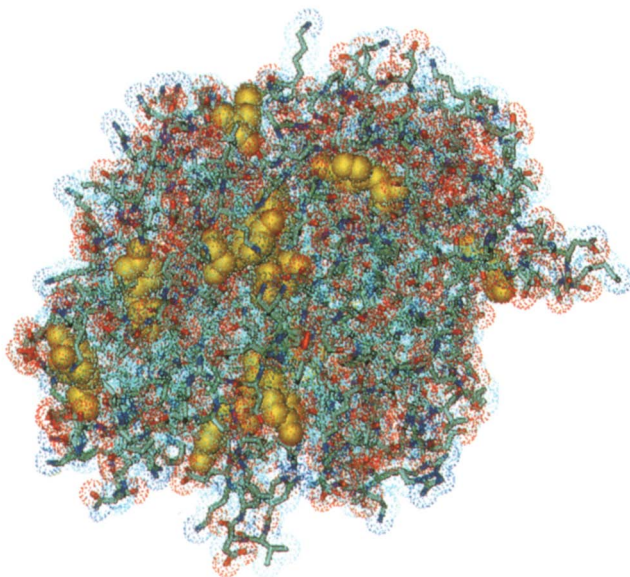
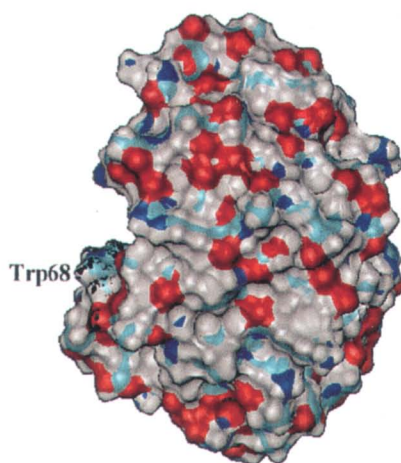


Figure 5.4. X-Ray structure of the sialidase of Salmonella typhimurium. Tyr-residues at the surface are highlighted by the van der Waals representation.

d



e

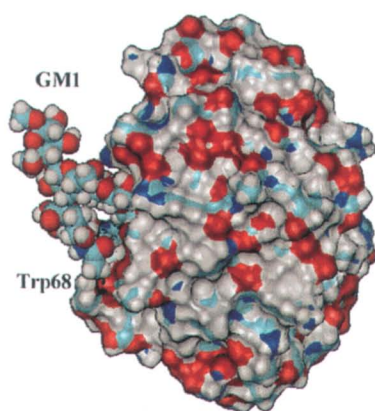
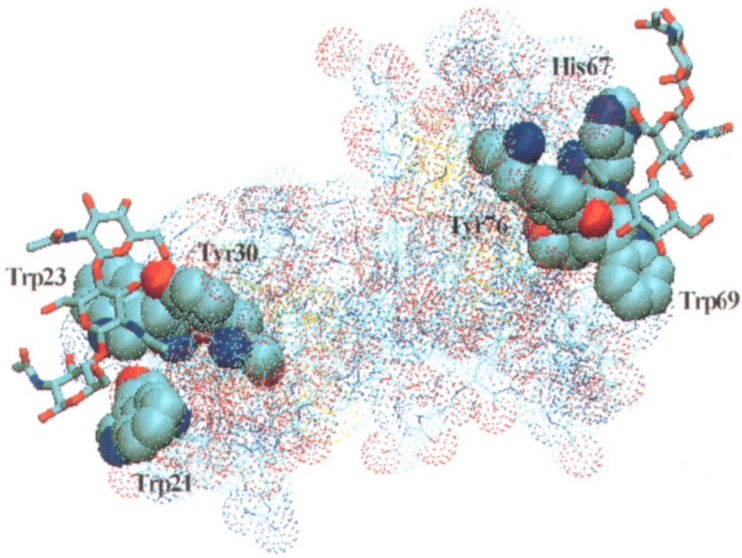


Figure 5.5. Laser photo-CIDNP difference spectra of human galectin-1 (aromatic part) in (d) Monomer of human galectin-1. (e) Monomer of human galectin-1 in the presence of the pentasaccharide chain of ganglioside GM₁.



*Figure 5.6. Model of the stinging nettle agglutinin (*Urtica dioica* agglutinin - UDA). The two binding sites are occupied by the potent ligand N,N',N''-triacylchitotriose.*

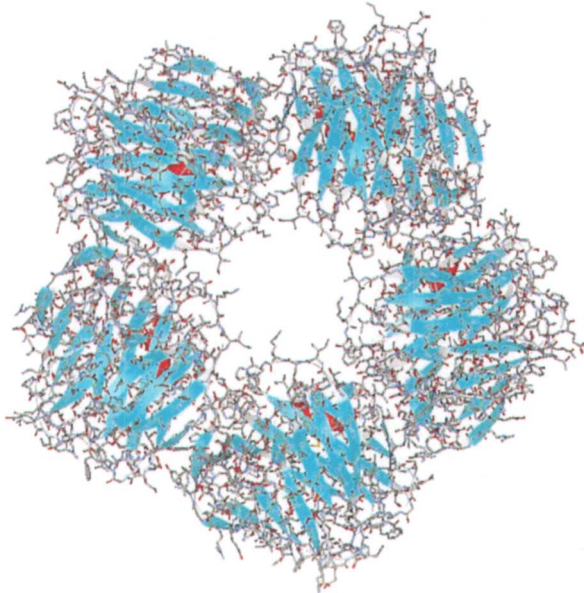


Figure 5.7. X-Ray structure of the pentameric serum amyloid P component.

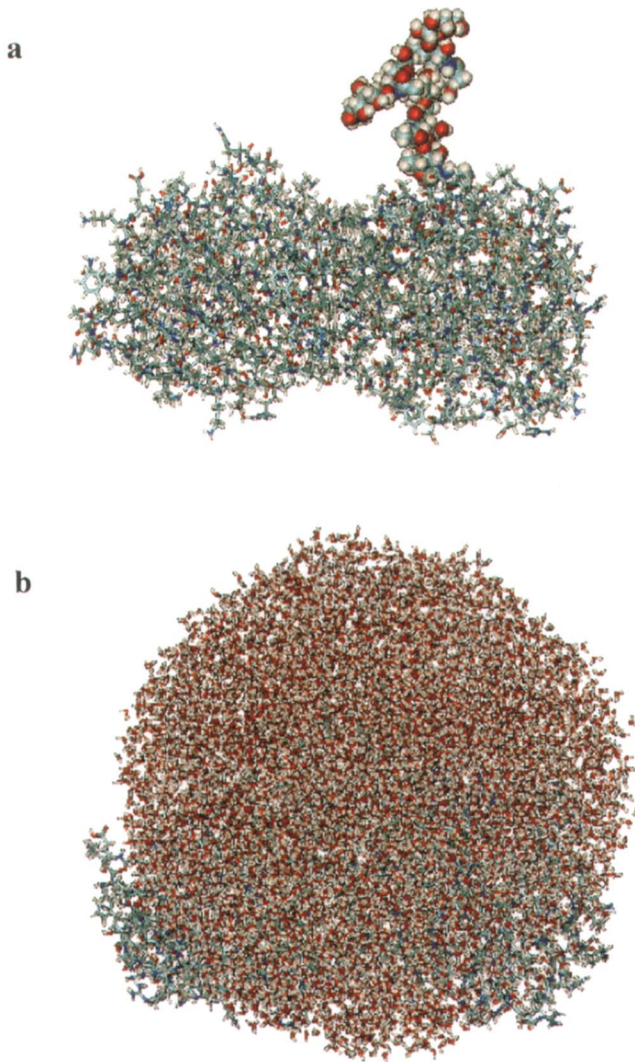


Figure 5.10. Illustration of the structure of a dimer from pentameric SAP (a) including the N-glycan at one subunit and (b) surrounding water molecules.

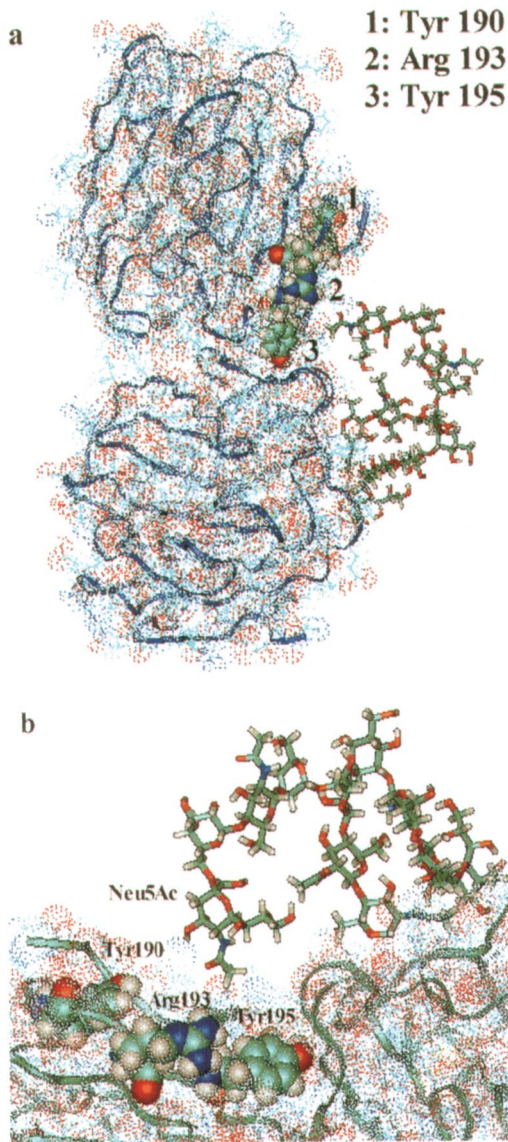


Figure 5.11. Illustration of the structure of a dimer from pentameric SAP (a) including the unmodified glycan at one subunit (b) with emphasis on the contact region (c) including the desialylated N-glycan of one subunit and (d) with emphasis on the contact region of this form.

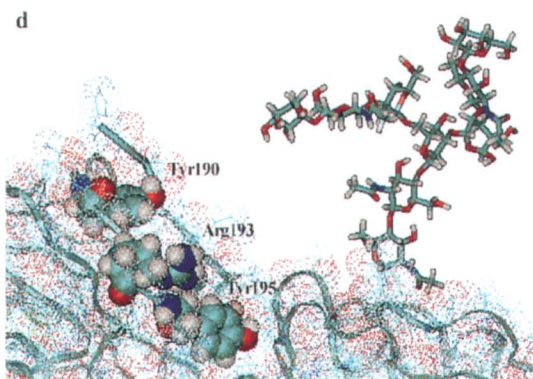
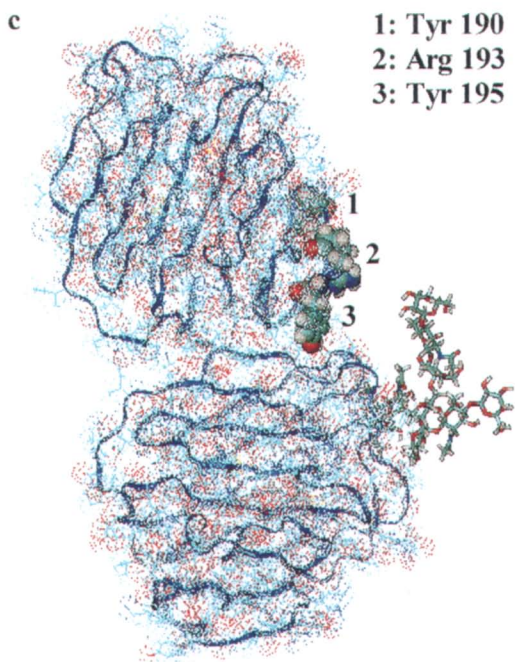


Figure 5.11. Continued.

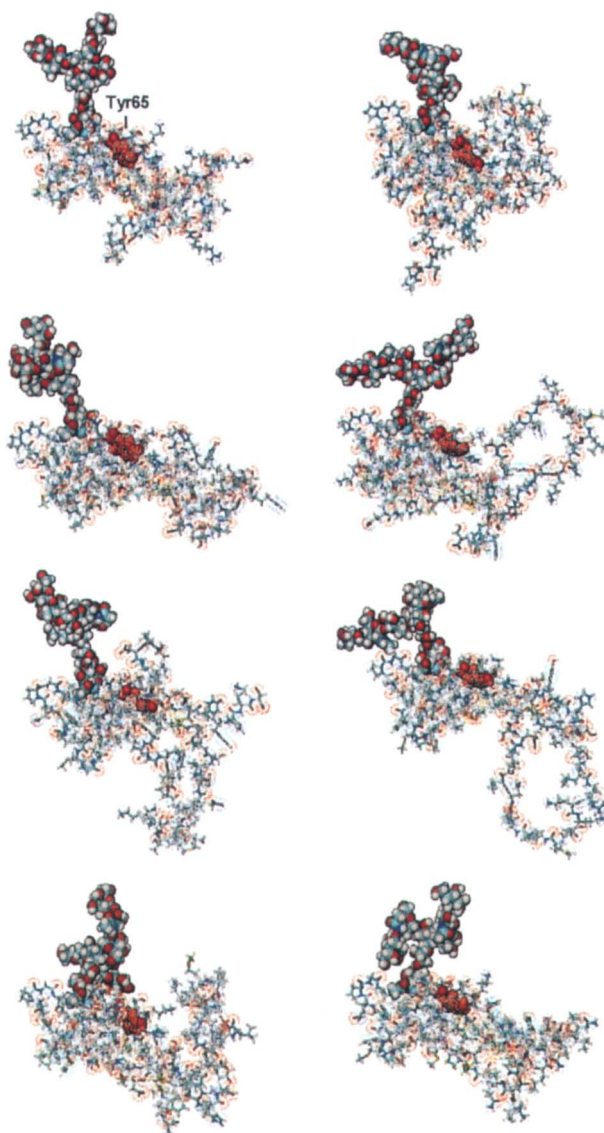


Figure 5.12. (b) Energy minimum conformations of ahCG glycosylated at Asn78 – (pdb-entry: 1HD4). The CIDNP-relevant Tyr-residue (Tyr65) is highlighted by its van der Waals representation.

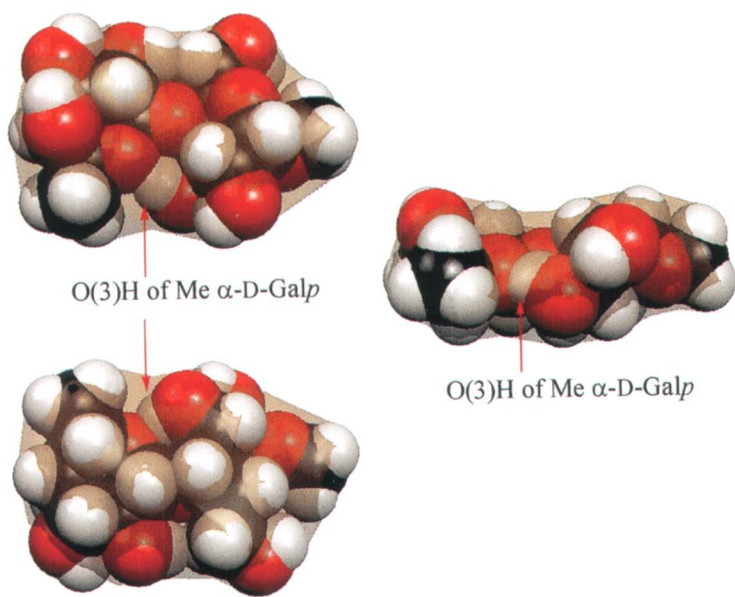


Figure 6. Space-filling representation of β -L-Fucp-(1 \rightarrow 4)- α -D-Galp-OMe in three orientations showing the location of O(3)H within the solvent inaccessible surface (slightly shaded). The $\Delta\delta$ -value of -0.64 ppm was attributed to reduced hydration. (Reproduced from reference 45. Copyright 2004.)

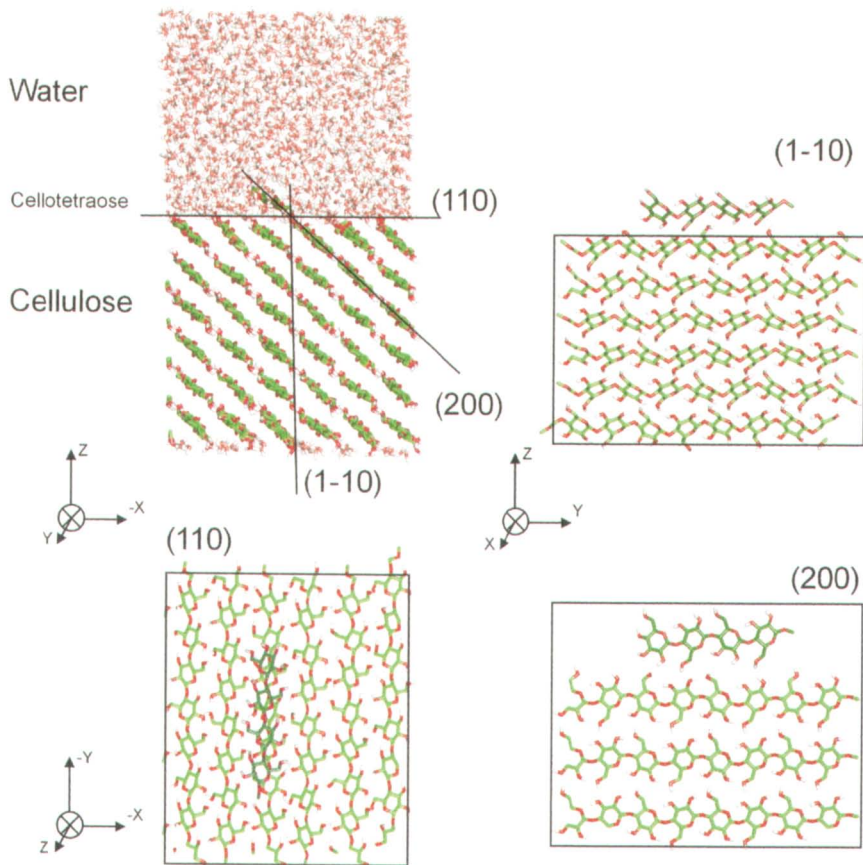
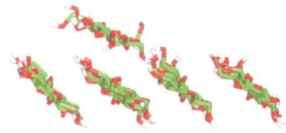
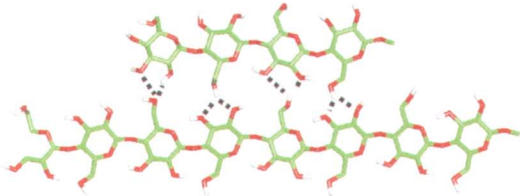


Figure 1. Cellotetraose positioned on top of the cellulose (110) plane and surrounded by a box of water (top-left) as viewed from the reducing end along the y axis, when looking perpendicular on the (1-10) plane (top-right), when looking perpendicular on the (110) plane bottom-left), and when looking perpendicular on the (200) plane (bottom-right).

Cellotetraose

Parallel

(200)



Anti-parallel

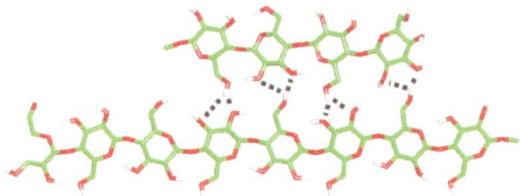
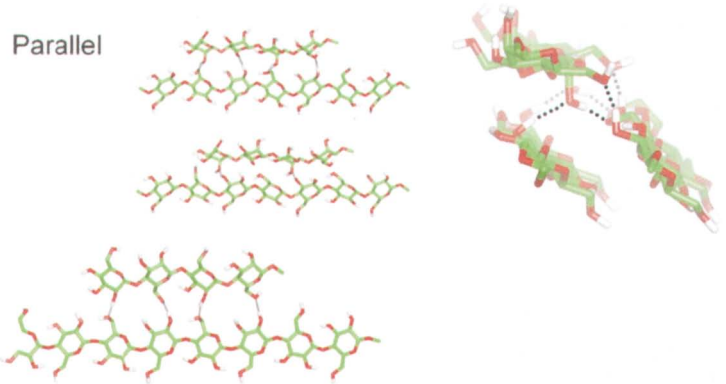


Figure 3. The position of cellotetraose in parallel (top) and antiparallel (bottom) orientation. Hydrogen bonds between cellotetraose and cellulose are depicted by dotted lines.

Mannotetraose

Parallel



Anti-parallel

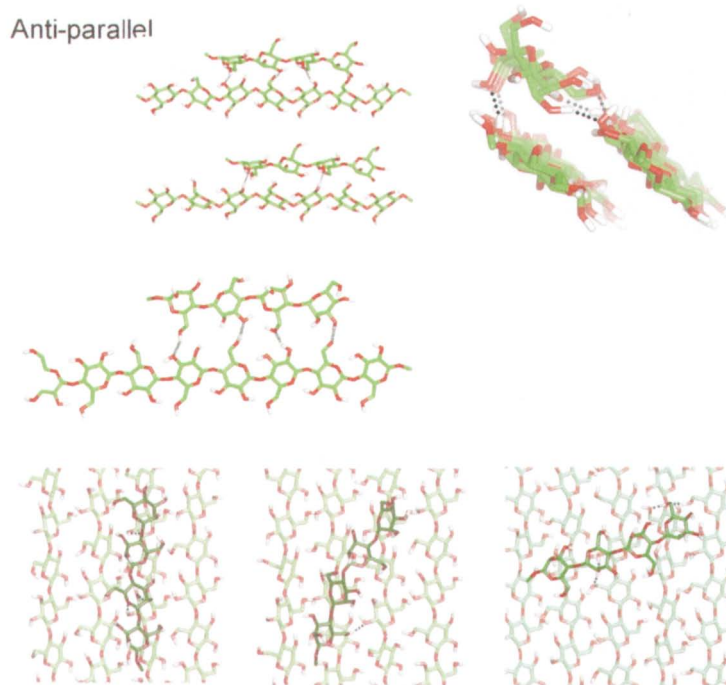
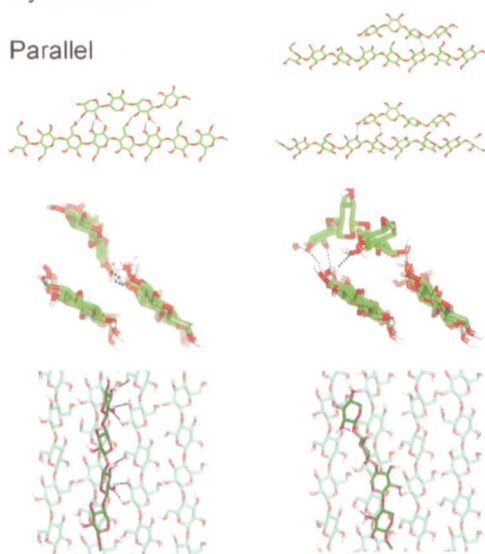


Figure 4. The position of mannotetraose in parallel (top) and antiparallel (middle) orientation. Hydrogen bonds between mannotetraose and cellulose are depicted by dotted lines. (bottom) Mannotetraose on top of the (110) plane of cellulose in antiparallel starting position (left) and after rotation of the tetrasaccharide (right) (1.5 ns simulation time) in MD simulation MA6e.

Xylotetraose

Parallel



Anti-parallel

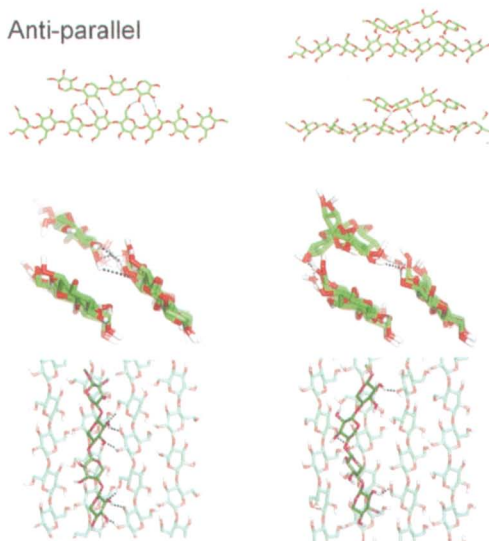


Figure 5. The position of xylotetraose in parallel (top) and antiparallel (bottom) orientation. Hydrogen bonds between xylotetraose and cellulose are depicted by dotted lines. The left column shows the 2-fold symmetry configuration (XP2e top and XA bottom) and the right column depicts the (partial) 3-fold symmetry configuration (XP1 top and XA3e bottom).

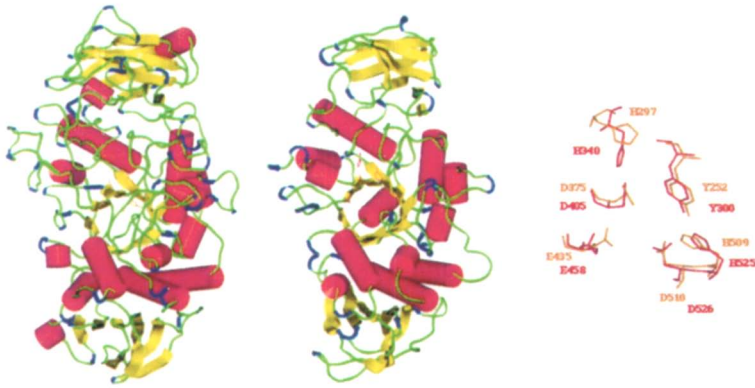


Figure 9.1. Topologies of *Ps. amyloclavata* debranching and *E. coli* branching enzymes (a). Details of their superimposed active sites (b).

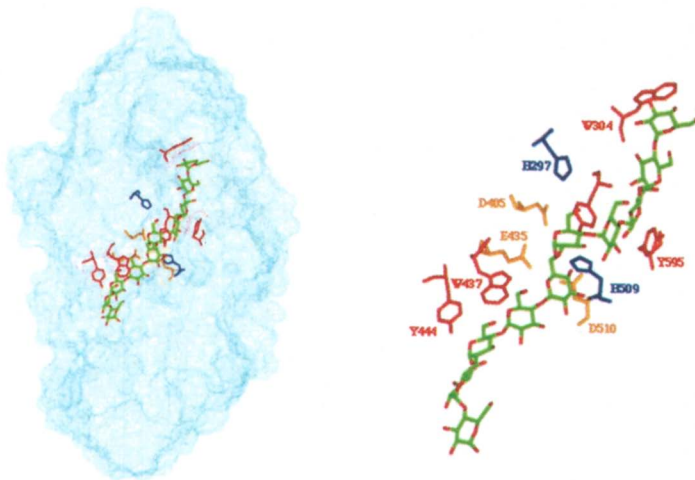


Figure 9.2. Docking of the G10' in DBE (a). Details of the π stackings (b).

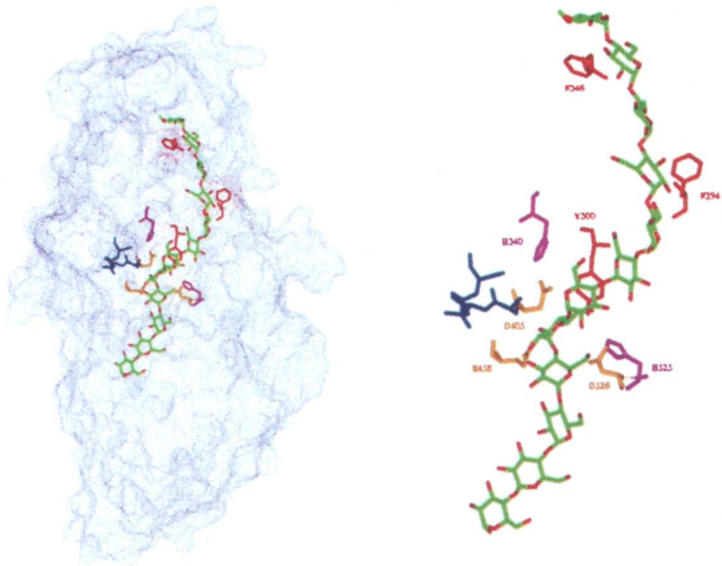


Figure 9.3. Docking of the G12 in BE (a). Details of the π stackings (b).

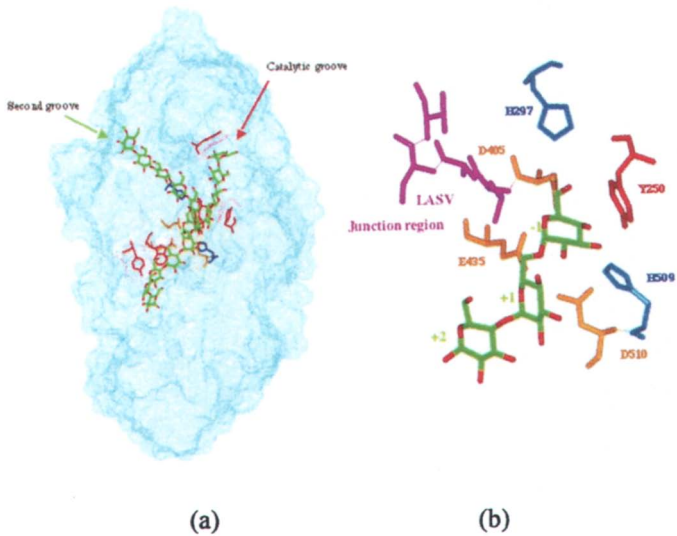


Figure 9.4. Docking of glycogen fragment in BE (a). The “junction region” (b).

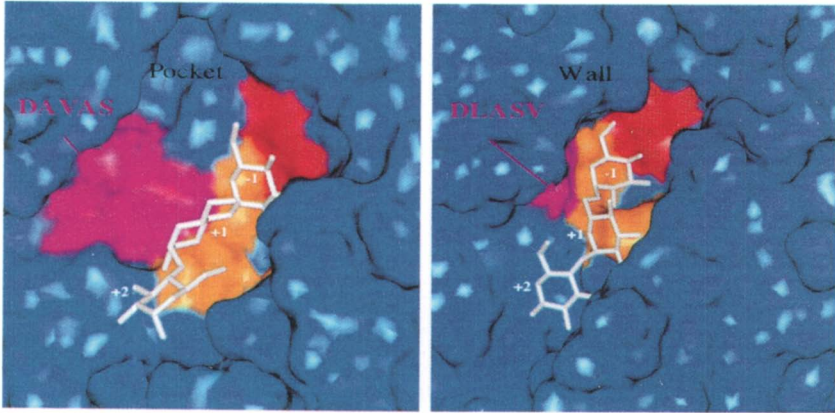


Figure 9.5. DAVAS pocket in BE (a). DLASV wall in DBE (b).

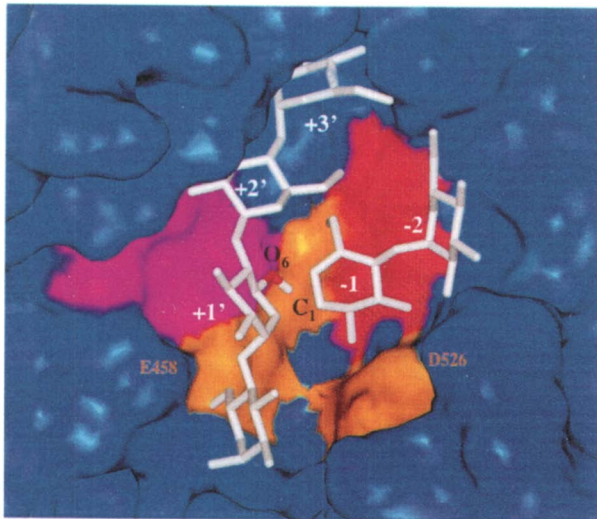


Figure 9.6. Subsites +1' and -1 of BE designed for a α -(1,6) transglycosylation?

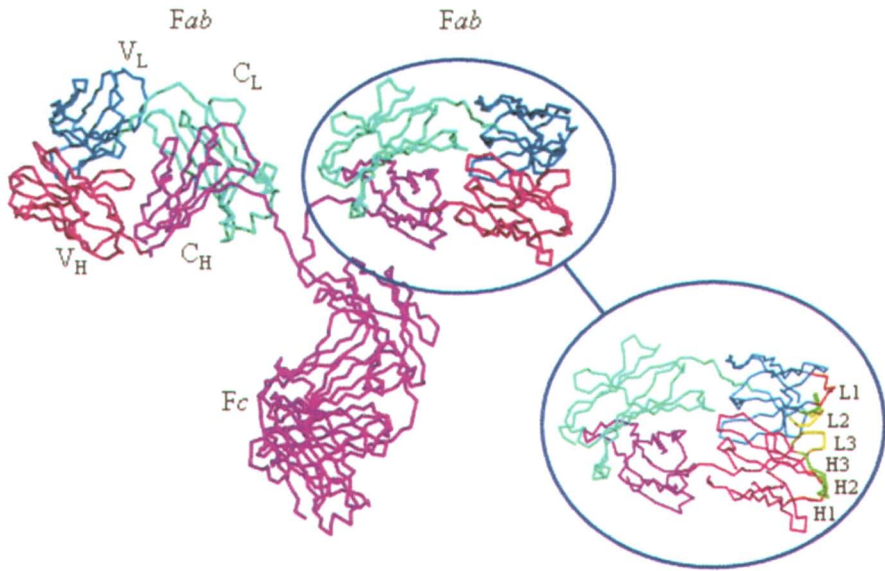
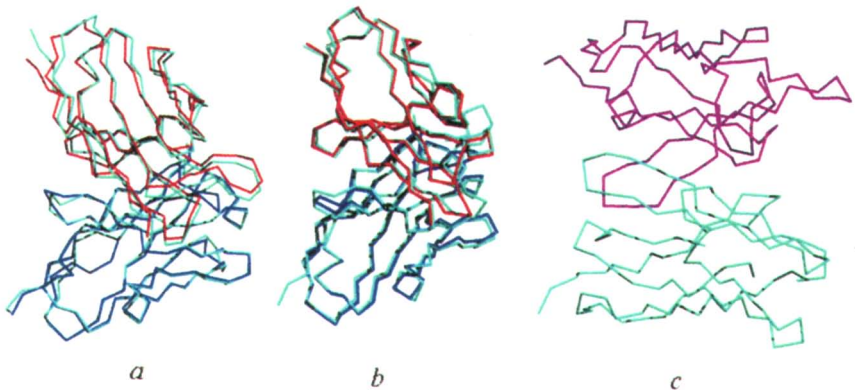


Figure 11.1. IgG antibody structure indicating the variable and constant domains. The hypervariable loops associated with the light (L1-3) and heavy (H1-3) chains in the Fv are indicated. The figure was created using Pymol(15) from the PDB file entry 1IG, the first structure of an intact IgG to be reported (16, 17).



*Figure 11.2. Overlay of comparative FV models with x-ray diffraction structures (crystal structure is blue, comparative model light chain red and the heavy chain cyan). Model **a** superimposes onto 1PLG with a Ca RMSD of 0.44 Å, while **b** superimposes on 1MFD with a value 0.62 Å. The predicted homology model **c** for Fv 1B1 is shown.*

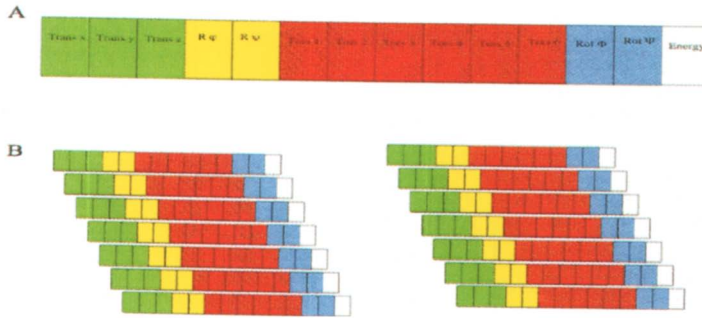


Figure 14.2. Illustration of the principle of the genetic algorithm docking strategy. A.) The state of the system is entirely defined by a series of variables: one value for each torsion, the coordinates of the center of mass (Trans X, Y and Z) and four quaternion values. B.) Multiple random states are generated as a starting population. The system is allowed to evolve by mutation and crossover.

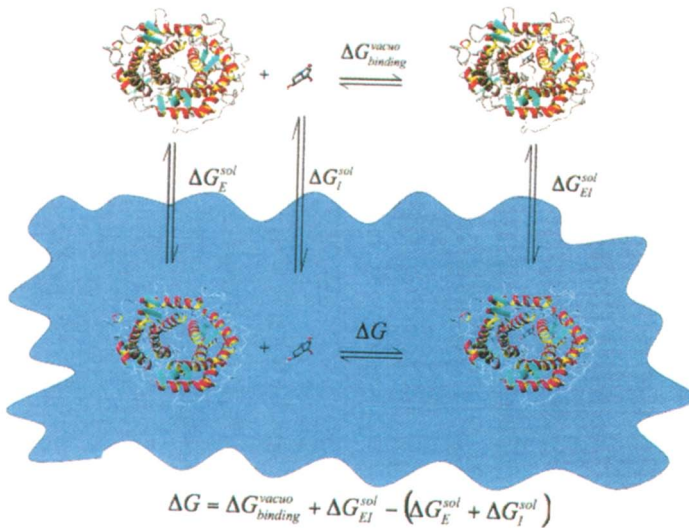


Figure 14.3. Illustration of the thermodynamic cycle principle. The blue shape represents solvent. The free energy change (ΔG) of the binding event in solution can be calculated as the sum of the free energy change in vacuum plus the difference in the free energy changes of solvation of the ligand, protein and protein-ligand complex.

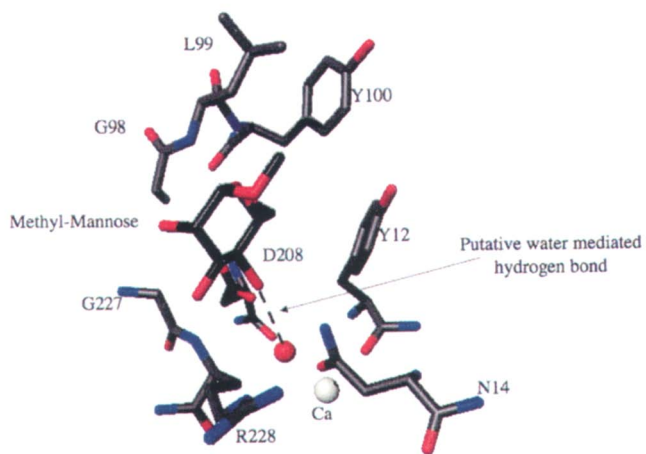
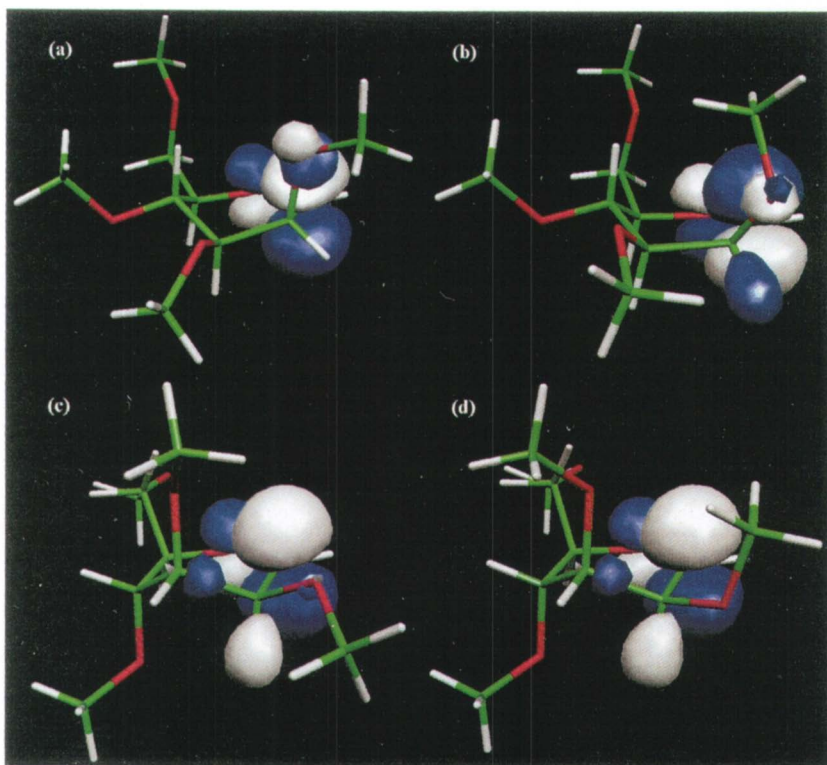


Figure 14.5. Binding site of Concanavalin A (9). A water molecule serves as a hydrogen bond bridge between the mannose ring and the Arg 228 side chain.



*Figure 17.8. DFT optimized geometry (Gaussian) and spatial distribution of LUMO's for: (a) (2) **B0** syn, (b) (2) **B0** anti, (c) (2) **B1** syn, and (d) (2) **B1** anti. Note pseudo-axial orientation of O-2 in (a) and (b).*

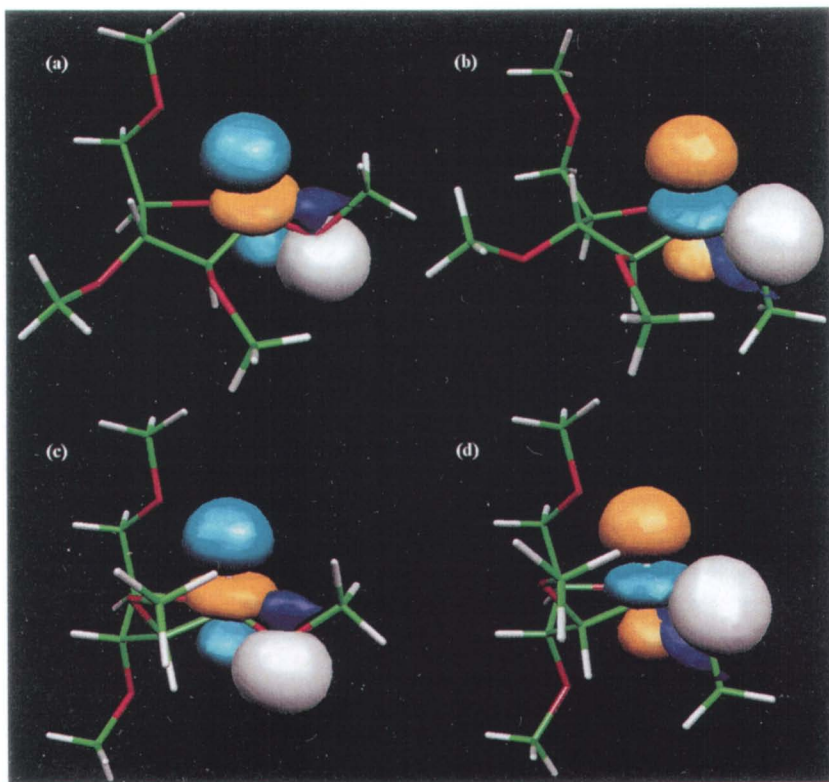


Figure 17.9. The $n_{O_2} \rightarrow \sigma^*_{CH}$ hyperconjugative interactions from NBO analysis for: (a) (3)**B0** syn, $E = 19.6 \text{ kJ mol}^{-1}$ (b) (3)**B0** anti, $E = 67 \text{ kJ mol}^{-1}$ (c) (3)**B1** syn, $E = 18.8 \text{ kJ mol}^{-1}$ and (d) (3) **B1** anti, $E = 6.7 \text{ kJ mol}^{-1}$.